NTNU
Norwegian University of
Science and Technology

# Patient friendly Presentation of Electronic Patient Records

**Kjetil Stallemo**

Master of Science in Computer Science
Submission date: June 2008
Supervisor: Herindrasana Ramampiaro, IDI

Norwegian University of Science and Technology
Department of Computer and Information Science

# Problem Description

A prototype where medical terms in electronic patient records are automatically linked to accompanying explanations has been developed. The main goal of this system is to give the patient a better understanding of their health situation and treatment.

The work to be carried out with this thesis is a continuation of the specialization project executed fall 2007 where the main goal is to find out the effects of extension of vocabularies, improvements of algorithms, services in the system, and the user interface on the system quality, seen from both the health personnel and the patients. Implementation, testing, and evaluation of these improvements are also a part of the thesis.

Assignment given: 14. January 2008
Supervisor: Herindrasana Ramampiaro, IDI

# Abstract

Reading an electronic patient record (EPR) is a very challenging task because of the medical jargons, which are almost impossible to understand for the layman. This becomes a highly relevant challenge because of the more extensive use of the internet to get medical information. Also the Norwegian laws state that the patient has the right to read his or her own EPR. A master thesis executed in 2006, and a specialization project in 2007 addressed this subject and developed a prototype for adapting EPRs to a patient presentation.

This thesis continues this work and aims to extend the system with more functionality and improve the translation of the EPRs. The main issues discussed in the thesis are how disambiguating between Norwegian words and medical terms, provide summaries of EPRs, and supply the patient with external information about his or her health condition. In addition the refined user interface from the specialization project was implemented.

The conclusion of this thesis is that the Support Vector Machine classifier with character bigrams provides good and accurate disambiguation between Norwegian words and medical terms. The external information functionality provides correct and quality assured information from the patient hand book. There are still some issues, and possible improvements on providing only precise and relevant articles. Summarizing of EPRs is achieved through named entity extraction of ICD codes, and then presenting the codes together with their corresponding descriptions. This implementation seems to be accurate, correct, and precise.

# Preface

This thesis is written as a part of a master's degree taking place in the spring 2008 at the Department of Computer and Information Science, Norwegian University of Science and Technology in Trondheim, Norway. The main issue is to study an existing prototype of a patient friendly EPR system, expand, and improve this system. The thesis also includes a Specialization project executed the autumn 2007.

Thanks are given to the teaching supervisor Herindrasana Ramampiaro for his support and feedback during this project. In addition, feedback and information from Ilangko Balasingham, Nurse Karl Øyri, and Laura Slaughter at the Interventional Centre has been of invaluable utility. Also thanks are given to the text laboratory at the University of Oslo for providing a text corpus with Norwegian fiction literature that was used to train the text classifiers.

Kjetil Stallemo
Oslo, June 2008

# Contents

# Figure list

# Table list

# Part I Thesis Context

# 1 Part introduction

This chapter gives a summary of the purpose and scope of the thesis directive, and an overview of the different chapters.

## 1.1 Purpose

The purpose of this part is to give guidelines describing how the thesis should be executed. These guidelines will serve as a roadmap during the research, discussion, and evaluation.

## 1.2 Scope

The chapters in this part give the foundation for the thesis, like motivation, thesis context, problem definition, and outline for the report. The section also describes the scope of the thesis and what the focus of the thesis will be. This part will not present results of the work, only what is to be done along the way, and what kind of methods and processes are to be used.

## 1.3 Overview

This part of the thesis consists of the following chapters:

- Introduction: Describes the background, motivation, problem definition and context for the thesis.
- Research methods: Introduces and explains the research and study methods to use during the project work.
- Summary: Summarizes the thesis directive.

# 2 Introduction

This chapter will present the thesis and the context. The chapter describes the background of the thesis and the research surrounding it. The motivation for the work is also presented along with a definition of the problem, which aims to state clear and unambiguous research questions that should be answered during the work.

## 2.1 Background

When a patient is treated in a Norwegian hospital the patient health, treatment, and medication is stored in an Electronic Patient Record (EPR). According to the Norwegian law the patient has a fundamental right to inspect and get explanation about his or her own patient record [1]. In addition research has shown that a considerable part of patients are interested in reading his or her own patient record [2]. In this occasion a project group was assembled at the Interventional Centre at Rikshospitalet HF (RHF) to develop a web portal which offers heath information to patients [3]. One of the functions in this portal is to give the patient a full overview of his own EPR with explanations of the medical terms

Elena Ivanova's master thesis [3] was a part of this project, and a prototype was developed for presenting EPRs in a patient friendly environment. The prototype implements the Norwegian electronic medical handbook (NEL) as the main source for explanations to different medical terms. The prototype was tested with patient data producing varying results. During the following year Vegard Nossum worked at RHF developing a new architecture focusing on integration with the hospital EPR system, security, flexibility, and scalability. This system is described in [4].

During the fall semester 2007 a pre-study to this master thesis was executed, with the main goal to look at possible extensions and improvements to the system [5]. Some of the main issues discussed here are presented below.

*"This project aims to study different alternatives and options in this area, while the master thesis will continue this study, put some of these into life, evaluate, and suggest changes. The improvements on the information retrieval process have to be tested mainly against the precision concept. We know that the extensions will give more hits in the vocabularies, but the main issue is whether the descriptions are correct and accurate.*

*The vocabularies in this study would provide a valuable contribution to the system, and should be tested as an extension. The collocation and misspelling algorithm are also important aspects that will improve the system. The misspelling has to be considered against the risk for erroneous information.*

*The design and functionality is mainly developed as different suggestion that is meant to be compared further in the master thesis. The summary function is an important functionality that could* *have different area of application, like patients that want to keep track of their health history and usage to PHRs. When extending with extra information the patient handbook seems to be the best alternative, but this area needs further study."*

Issues from the project that are relevant to address in this master thesis are presented below.

- Searching techniques to get high precision on the retrieved articles
- Decide explicit design alternatives to compare and evaluate
- Evaluate the precision of the new information retrieval process
- Evaluate the summary function
- Evaluate the usability of the additional information
- Evaluate the result of extending the vocabulary

## 2.2 Motivation

The main motivation for this thesis can be derived from the results in the pre-study mentioned above. This study has shown that the system can be improved in many different areas. The system, as it is today, has never been tested with patients and has several weaknesses. Examples of this are the user interface, too small vocabulary, and problems separating Norwegian words and medical terms. These aspects together with other challenges like the fact that physicians and nurses use a combination of Norwegian, oral language, and medical terms, make the area challenging [6]. Medical terms are often mixes of Latin, Greek and Norwegian, making them highly complex to handle. There is a huge challenge with separating Norwegian words and medical terms, and therefore term disambiguation is a highly important task. The huge gap between the consumer's and professionals' language is an important obstacle for effective communication. This type of communication is very important since there is an increase in patients' interest in searching and reading health information on their own [2, 7].

All these challenges and needs, together with the fact that this system has a highly diverse user group, motivate to further work and research in this area.

## 2.3 Thesis context

The thesis is carried out as part of a master degree within the area of computer science at The Norwegian University of Science and Technology (NTNU). The subject TDT4900 Computer and Information Science is the context for this thesis. The assignment is given in corporation with the Interventional Centre at RHF and continues earlier work described in Section 2.1.

## 2.4 Problem definition

The problem definition is partly based upon the pre-study executed the fall 2007. Some of the same questions will be discussed further, while new aspects will be addressed. The disambiguation of word senses, whether they are Norwegian or medical terms, is an important challenge with the existing system and will therefore be the main topic in this thesis.

A prototype where medical terms in electronic patient records are automatically linked to accompanying explanations has been developed. The main goal of this system is to give the patient a better understanding of their health situation and treatment.

The work to be carried out with this thesis is a continuation of the specialization project executed fall 2007 where the main goal is to find out the effects of extension of vocabularies, improvements of algorithms, services in the system, and the user interface on the system quality, seen from both the health personnel and the patients. Implementation, testing, and evaluation of these improvements are also a part of the thesis.

The text above presents an English version of the thesis description. The main goals are presented here and it states a superior problem definition. To specify the definition with higher details there are developed some research questions (RQ) that will be answered during this thesis.

**RQ1** Is it possible to integrate external information sources into the EPR to provide secure, precise, and correct dynamic information to the patient?

**RQ2** Will extension of the information retrieval (IR) process, such as collocation, text mining, and spell suggestion give significant improvements to the system?

# 3 Research methods

This chapter presents the research methods and strategy used in this thesis to produce, and evaluate the results.

## 3.1 Methodology

The methodology used in this project will build on the method used in the pre-study, which involve using an agile and iterative method. The report writing, literature study, and implementation will be carried out iterative, ensuring that there will be a result of the work. After finishing the literature study and implementation of the prototype, the outcome will be evaluated through statistics and qualitative examples.

The study has to be divided into two parts, one focusing on the functionality of the system, and another part focusing on the IR process. The functionality will be evaluated by studying the outcome in testing, aiming at using case studies as the main method. Because of the lacking possibilities to test in large scale with patients, the study will be mainly qualitative based on fictive examples. A qualitative study collect data like observations, interviews, images, and analyzes it with methods without precise measurement [8]. In this thesis the functionality will be evaluated through observations of the system usage in some examples.

The IR process will be evaluated against regular evaluation criteria for IR and text mining like recall, precision, accuracy, true positive rate, false positive rate, and so on. This type of research is a quantitative experiment in a controlled setting where variables can be changed to produce the results [9]. The research will evaluate whether the improvements contributes with significant improvements to the system.

## 3.2 Research Strategy

The thesis research strategy will be based on the research questions stated in Section 2.4, whereas the result of the study will be a prototype and results from text mining experiments, validated by evaluation of these. The evaluation will be done in two parts, namely experiment and case studies based on examples [10].

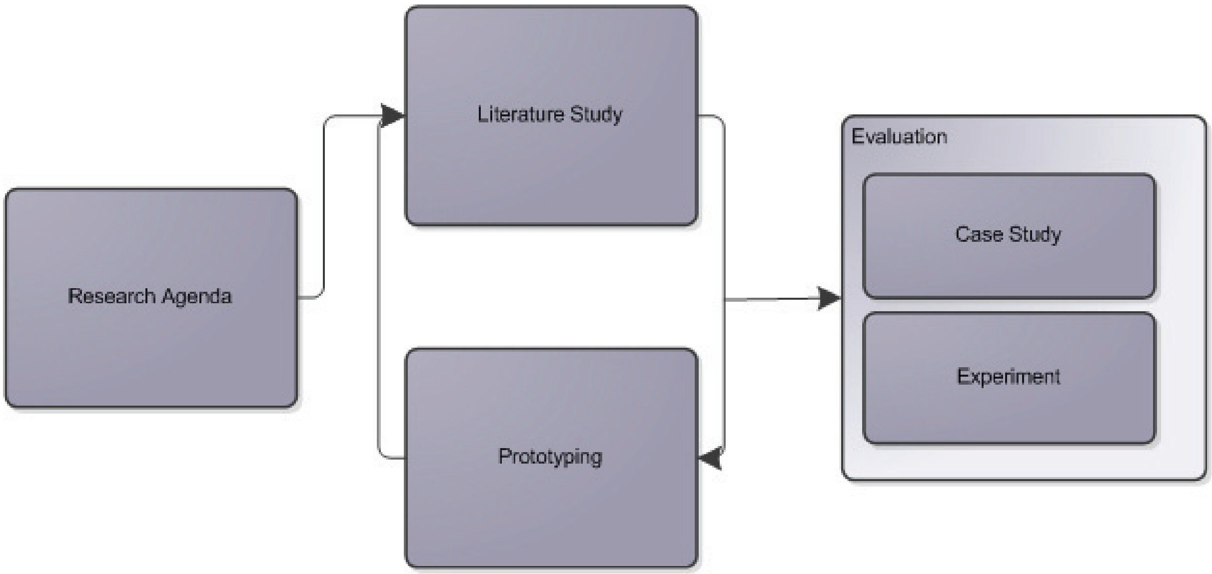


Figure 1 Research Strategy

Figure 1 describes the strategy as iterations between literature study and implementation of the prototype. When this phase is completed an evaluation of the system will take place as described above.

### 3.3 Challenges and obstacles

Working with this thesis one of the main challenges is getting access to data, and using real users as test persons. There are strict rules and constrictions regarding EPR's, and accomplishing case studies with patients. Because of this, some of the study has to be accomplished without user groups available, or with the required amount of data. If text mining is to be tested in large scale, access to a larger amount of EPR's is needed.

# 4 Summary

According to Norwegian laws the patient has a fundamental right to view his or her own EPR. In addition patients have a high interest in viewing own medical data, and the use of internet as a source for medical information is more available than before. The work already finished is not complete, and there are more issues like word sense disambiguation, extended functionality and so on to address.

The research will be based on quantitative evaluation of IR measures like precision and recall. While some of the functionality is evaluated and validated by qualitative examples of usage.

# Part II Analysis of theory and state of the art

# 1 Part Introduction

This chapter gives a summary of the purpose and scope in this part, and an overview of the different chapters.

## 1.1 Purpose

This part is a study of the subject area, and serves as a presentation of the existing prototype, research in the area, and different techniques and theory that are relevant to the study. Further the state of the art will be analyzed according to this thesis.

## 1.2 Scope

The chapters in this part will describe the work already done in this area. They will address the existing system, research done in this area which looks at possible IR techniques and functionality in a patient friendly EPR system, and other relevant areas in this thesis. Also theory relevant the different subject areas are presented.

## 1.3 Overview

This part of the thesis consists of the following chapters:

- Theory about IR techniques: Theory that provides background information for the different IR approaches.
- State of the art: Presents the existing system and other relevant systems.
- Summary: Summarizes this part.

# 2 Theory about IR techniques

The IR techniques are important parts of getting a good result when searching in text documents, for example EPRs. The pre-study [5] shows that there are several challenges in this area that could be addressed and tested further.

## 2.1 Collocation

When two or more terms together form an expression it is normal to call it a collocation. There are different ways to detect and index collocated terms like counting the number of times the expression occurs, or using statistical methods [11]. When searching for collocated terms in an EPR system we already have the expressions in the vocabulary so that detection of collocations are unnecessary [5]. As mentioned there are different types of techniques to detect collocated terms. Frequency counting, hypothesis testing with t-test, Pearson's chi-square test, or likelihood ratios, or mutual information are all techniques that are discussed [11]. Since detecting collocations are irrelevant for this thesis the techniques mentioned above will not be discussed further in this Section.

Collocated terms are defined by Choueka (1988) as "*a sequence of two or more consecutive words, that has characteristics of a syntactic and semantic unit, and whose exact and unambiguous meaning or connotation cannot be derived directly from the meaning or connotation of its components*" [11]. Collocated terms can also, in some cases, be words that are not adjacent to each other [11]. There are some typical criteria constituting collocated terms, these are described below.

**Non-compositionality,** the meaning of the collocation is not straight forward the meaning of the different words. Examples of this are "strong tea" and "to make up", which are collocations that not use the meaning of the different words straight forward.

**Non-substitutability,** describes that it is not possible to change the words in a collocation, even if they have the same meaning.

**Non-modifiability,** means that a collocation cannot be modified with lexical material or with grammatical transformations [11].

If the focus is shifted towards medical terms the collocations are terms that have a special meaning when they are placed together. An example of this is the term "dura mater" which refers to the outermost and thickest brain spinal marrow membrane, while dura means hard, and mater means membrane or mother. This collocation fulfills the criteria mentioned above, namely non-compositionality, non-substitutability, and non-modifiability.

## 2.2 Spell suggestion

When physicians or their assistants type EPRs with spelling errors, it can cause problems for our translation of words and terms. One of the most common errors is the difference between Norwegian, Latin, and Greek. The medical terms in Latin and Greek are norwegianized. An example of this is the word appendix which in Norwegian-Latin is spelled appendiks. This challenge was addressed in [3] and is not a subject in this thesis.

Spelling errors caused by other factors are addressed in this section. To detect these it is possible to use Levenshtein (edit) distance. This distance is the minimum number of operations to transform one string into another. The spell suggestion module could suggest the term in the index with the lowest distance as the translation.

## 2.3 Text mining

Text mining is a technique for extracting information and knowledge from unstructured text documents [12]. There are many interesting applications in this area that are possible to use in adaptation of EPRs. One important sub area of text mining is text categorization which is used to label documents in different categories [13]. Another application is to extract information like key phrases and relationships within the text [14].

### *2.3.1 Text categorization*

Text categorization is the activity of assigning different categories to documents [14]. There are different techniques to achieve this, namely Knowledge engineering (KE) or Machine Learning (ML). The former one was until the late 80s the most used approach while in the last years ML has taken this role [13].

KE uses manually defined rules to categorize the documents while ML uses a set of training documents to learn how to classify documents. The expertise needed with the latter approach is insignificant while you get approximately the same accuracy as KE [13]. But to use this technique there has to be a training set of documents available. Text classification can be used in many different domains and situations, some of them are presented below [13].

- Document organization
  Grouping documents into different categories, for example classifying ads in a newspaper.
- Text filtering
  Deal with the activity of categorization an incoming stream of documents.
- Word Sense disambiguation.
  Treat disambiguation of the word sense in different contexts.
- Hierarchical Categorization of web pages
  Deal with the classifying of different web sites into hierarchical categories.

Classifying documents can be achieved with many different constraints depending on the application. The documents can either be assigned to only one category, called single label categorization, or documents can be assigned to different and overlapping categories, called multi-label categorization. The categorization can be accomplished through ranking or with a hard decision. The latter describes a method which decides whether a document belongs to a certain category, while the former ranks different categories according to the likelihood of the document belonging to that category [13].

ML is an interesting way of implementing text categorization because there is no need for domain experts and the accuracy is fairly high. There are different approaches and algorithms to use during the text categorization, some of them are described below. There are some important aspects that separate this work from ordinary text categorization. In this thesis word sense disambiguation is an interesting field because this is a challenge in the existing system.

In the following sections some relevant classifiers are presented. The classifiers are taken from different groups, namely probabilistic, decision trees, and support vector machines which are a combination of linear models and instance based learning [15].

### Naïve Bayes

According to Witten and Frank this is one of the most used algorithms used for text classifying, mainly because of its speed and accuracy [15]. The algorithm is used in many applications and is fairly easy to implement, but there are some limitations. The algorithm's main weakness is that it assumes that all the attributes, and document lengths are independent [16]. There are several variants of this algorithm, and some of the most used are multinomial naïve bayes which accommodates word frequencies [16], and complement naïve bayes which takes skewed data into account [17].

Naïve Bayes is a probabilistic classifier that assumes that all attributes are independent. It combines the rules of statistical independence and bayes rule. The main thought behind the algorithm is to compute the likelihood that a document or word vector belongs to a class.

$$
\begin{gathered}
A \text{ and } B \text{ independent} \Leftrightarrow \\
\big(P(A|B) = P(A)\big) \wedge (P(B|A) = P(B)) \Rightarrow \\
P(A|B) = \frac{P(A \wedge B)}{P(B)} = P(A) \Rightarrow \\
P(A \wedge B) = P(A)P(B)
\end{gathered}
$$

The deduction above shows that if the variables are independent it is possible to calculate the probability for all the variables by multiplying them. If we know the probabilities of each attribute for a category, and assume the attributes are independent, then it is possible to compute the probability that given attributes belongs in given category. Bayes rule is presented in the formula below.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

This formula, together with the assumption that the attributes are independent, gives the possibility to calculate whether a vector A belongs in category B [18].

### Decision trees

Decision trees are another type of classifiers that are used in text categorization, and represents a rule based approach [19]. The C4.5 is a popular decision tree algorithm and is often called top-down induction of decision trees [15]. There are two main approaches for deciding which attribute to split on, namely information gain and gain ratio. The goal is to get as small trees as possible, and therefore get nodes that are as pure as possible.

Information gain measure the purity of the daughter nodes, and the information gain by the split. Continuing this until the leaf nodes are pure, which means that they only contain instances that have the same classification, is the ideal approach. If it is not possible to get pure leaf nodes, the process terminates when splitting is no longer possible [15]. Information gain has the weakness of often choosing the attributes with the largest number of values. This kind of branching is not suitable for classifying unknown instances, and does not present the decision structure in a good way. An alternative to this is to use gain ratio which divides the information gain on the information value of the attribute. This ensures that attributes that have a large number of possible values is not chosen as the root attribute [15].

The C4.5 has gotten a number of improvements, like the possibility to handle numeric attributes, dealing with noisy data, and missing values.

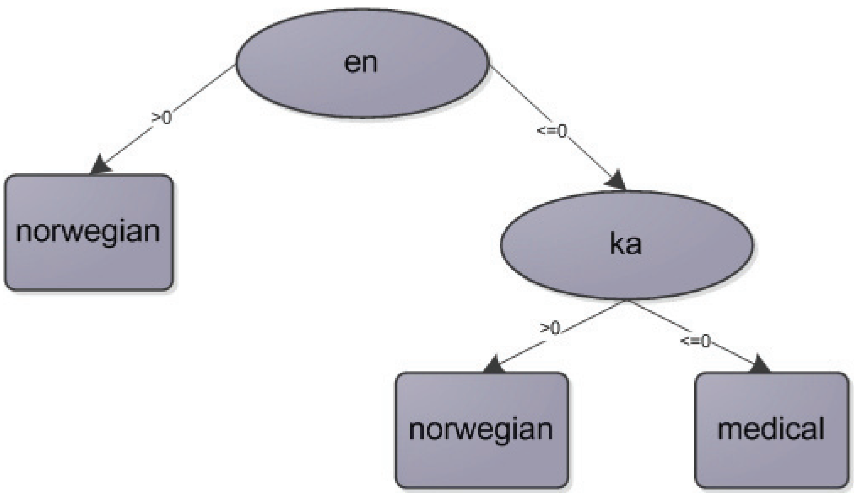


Figure 2 Example of a decision tree

Figure 2 shows an example of a decision tree classifying words with character bigrams, see Section 2.3.2. If the bigram en has a score higher than zero the word is classified as Norwegian, if not it looks at the bigram ka, and classifies to either a Norwegian word or a medical term.

**SVM**

Support Vector Machines (SVM) is an extension of linear models [15], and is a commonly used classifier in text categorization [16] [20]. SVMs are suitable as text classifier because of its high dimensionality input space, ability to handle datasets with few irrelevant features, and sparse document vectors. In addition most text classifying problems are linear separable [20].

Linear models biggest disadvantage is that they only can represent linear boundaries between different classes. SVMs solves this problem by using a nonlinear mapping, which transforms existing space into a new linear space [15]. A special linear model called the maximum margin hyper plane, presented as an example in Figure 3, gives the maximal separation of the classes.

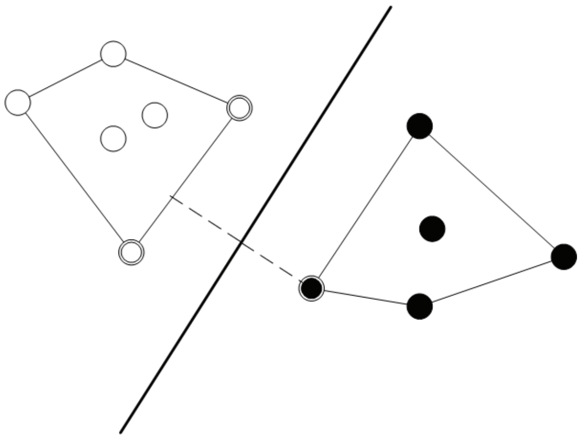

Figure 3 A maximum margin hyper plane

The line represents the maximum margin hyper plane which divides the two classes' outer lines. The dots marked with double circle represent the support vectors, in other words the instances that are closest to the maximum margin hyper plane [15].

There are different kernels that can be used to compute SVM, and some of the most suggested are radial basis function (RBF) kernel, and the sigmoid kernel. The results depend on the application and data, but it is important to note that there are seldom large differences in practice [15]. RBF is in many cases a good starting point [21].

### 2.3.2 Text preprocessing

Before applying text mining there are a lot of different approaches to preprocess the data to get better results. Stemming is an example of this, already used in the existing system [3, 13]. Other techniques that are relevant are weighting of terms with frequency, or expanded with idf score described in Section 2.3.4. Alternative to stemming is character n-gram tokenization which is language independent [22]. N-grams can consist of n subsequent words, or n subsequent characters. Character n-grams are substrings of words with length n.

To use text mining on strings it has to be preprocessed to a vector [15]. In our case the only medical data available are the medical vocabularies with terms, and no sentences with context. If the classifier were trained with these terms together with sentences of Norwegian literature the main part of words and sentences would be classified as ordinary Norwegian. Character n-grams is a good alternative when classifying languages [23]. Most of medical terms are Greek or Latin and therefore classifying these words are mostly the same as classifying languages.

If the training data would consist of Norwegian literature, which probably is a bigger dataset then the available medical terms, the data has to be balanced. There are different methods to achieve this, but since we have a rather big minority set (the smallest dataset) we will discuss random over and under sampling [24]. This will also be the best approach when thinking of computing expense.

**Over sampling** is when the minority class is expanded through random replication while **under sampling** is the opposite, namely to reduce the majority class through random elimination. It has been stated that over sampling may lead to over fitting since it copies already existing words while under sampling could discard useful and important words. In our setting, the Norwegian literature probably contains a lot of duplicates which minimizes the risk of discarding significant words.

### 2.3.3 Evaluating classifiers

The results of the different datasets and classifiers have to be evaluated. IR results are normally evaluated by precision and recall. These aspects are also important in text mining. In addition accuracy [13], also called success rate [15], and error rate are relevant measures. In the medical domain sensitivity and specificity are used in diagnostic tests. Sensitivity describes the people with the disease and a positive test result while specificity refers to the people without the disease and with a negative test result [15]. These measures are taken to a text mining context and described in the following section.

#### Measures

Before presenting any measures some text mining concepts have to be defined. True positive (TP) and true negative (TN) are correct classifications, while false positive (FP) is an incorrect positive prediction and false negative (FN) is an incorrect negative prediction [15].

The standard IR measures precision and recall are originally described as the formulas presented below [25].

$$recall = \frac{relevant\ documents\ retrieved}{total\ relevant\ documets}$$
$$precision = \frac{relevant\ documents\ retrieved}{total\ retrieved\ documets}$$

If the formulas are converted to text mining we get the following:

$$recall = \frac{TP}{TP + FN}$$
$$precision = \frac{TP}{TP + FP}$$

The success and error rate is another approach of measuring a classifier, but these measurements are not widely used in text classification [13]. The reason is that the denominator often has a large value, which leads to insensitivity to variations in the success rate (TP+TN).

$$success\ rate = \frac{TP + TN}{TP + TN + FP + FN}$$
$$error\ rate = 1 - success\ rate$$

Another interesting measure is the Kappa statistic which describes the agreement between predicted and observed results allowing for agreement that occurs by chance [15].

Sensitivity and specificity are taken from the medical domain and are calculated as presented in the following formulas.

$$sensitivity = \frac{TP}{TP + FN}$$
$$specificity = 1 - \left(\frac{FP}{FP + TN}\right)$$

### Testing classifiers

When testing classifiers ideally there should be a separate test set to run tests on. In this thesis this is not the case. The EPRs available are not tagged with word classes, and in other words it is impossible to do a large test with these. It is possible to mark a few records, and run a test on these, but because of the limited time available it is not possible to get a large test set.

An alternative solution to this problem is using the training data as test data with cross validation [15]. The first important part of this procedure is stratification which ensures that each class is properly represented in the dataset. Next the data is divided into a number of folds, or partitions of data. Each of these is use one at a time for testing while the rest is used for training. Different tests has shown that 10 is the right number of folds to get a good estimate of error, in addition this 10 fold cross validation should be run 10 times [15].

### Statistical tests

When the error estimate is calculated there has to be a measure of how sure we are that the estimate is the correct rate. In statistics the process of independent events that either successes or fails is called a Bernoulli trial [11, 15, 26]. The mean and variance in a Bernoulli trial are respectively $\mu = p$, and $S^2 = p\,(1-p)$.

If there are a large number of samples, the distribution is approximately normal distributed. The normal distribution has two tails and the probability that a random variable X is within a confidence range c is described as:

$$\mathrm{P}[-z \leq X \leq z] = c$$

The estimate has to have zero mean and unit variance to use the standard normal distribution table, which leads to the formula below. F is the estimated success divided on the number of samples, and N is the number of samples.

$$P\left[-z \leq \frac{f-p}{\sqrt{p(1-p)/N}} \leq z\right] = c$$

Since p is unknown the most reasonable way to use this to calculate how certain we are that result is true is to use confidence interval. To do this the formula above has to be expressed as equality for p.

$$p = \frac{\left(f + \frac{z^2}{2N} + z\sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}\right)}{1 + \frac{z^2}{N}}$$

Since the variable has zero mean and no variance a table for the normal distribution can be used to find the z value. The confidence value c is subtracted from 1, and then divided by 2 since the distribution is two tailed. The calculated value can then again be found in the table with confidence limits for the normal distribution together with a corresponding z value [15].

The method described above can be used to find if a text mining method is suitable for a certain dataset. If several classifiers is to be compared a statistical test has to be applied. Student t-test can be used for comparing if the means are significant different between two distributions. Since the variance is an estimate the normal distribution is no longer valid and the student-t distribution has to be used [15].

To decide whether the means are significant different, the test checks whether the difference between the means are zero, in other words the null hypothesis presented below.

*$H_0$: The means are not significantly different. $\mu_1 - \mu_2 = 0$*

Since the values are paired the more sensitive paired t-test is used. In this test the variance is calculated from difference between the samples. The t value is calculated through the following formula where σ is the estimate of the variance, and k is the number of means, in other words samples [15].

$$t = \frac{\bar{d}}{\sqrt{\sigma_d^2 / k}}$$

One aspect that is relevant for this thesis is if the assumption that the data is unlimited, and that there exists several independent datasets, is invalid. A corrected resampled t-test would work in this case [15]. This t value is calculated with the following expression where $n_1$ represents each time an instance is used for training, and $n_2$ each an instance is used for testing [15].

$$t = \frac{\bar{d}}{\sqrt{\left(\frac{1}{k} + \frac{n_2}{n_1}\right) \sigma_d^2}}$$

**Cost analysis**

The Receiver Operator Characteristics (ROC) space is originally a classic methodology from the signal detection theory. ROC graphs describe tradeoffs between hit rate, and false classification rate. The ROC space has a y axis with the true positive (TP) rate, and an x axis with the false positive (FP) rate. The TP and FP rate is described in the formulas below [15].

$$TP = \frac{TP}{TP + FP}$$

$$FP = \frac{FP}{FP + TN}$$

The convex hull is used to determine whether which classifiers are suboptimal, and which could be optimal for some conditions [27]. *"The convex hull of a set of points is the smallest convex set that contains the points" [27].* The classifiers on the convex hull are optimal for some conditions, and have to be considered together with the balance of the data. If one of the classes is more represented than another the choice of the optimal classifier is affected.

### 2.3.4 Information extraction

Information extraction is the application of extracting information from a text or data. One important application is to summarize articles or text documents. This Section will look at different techniques for applying this.

Summarizing information differs at a basic level if they either extract or abstract information [28]. An extract only takes the most important information from a text while an abstract may include a paraphrase and quotation. To extract information there are two main methods, also mentioned in Section 2.3.1, namely KE and ML.

To extract with KE there has to be defined some rules to extract information from texts. An approach to decide the weight of phrases presented below [28].

U = Textunit
Weight(U) = Location(U) + CuePhrase(U) + StatTerm(U) + AddTerm(U)

The location element (Location(U)) in the formula is based on the fact that sections that occur early in the text probably have a higher significance than later ones. The cue phrase addend (CuePhrase(U)) assigns higher weight to units that start with phrases that indicate higher significance. The statistical salience of the unit (StatTerm(U)) is based on metrics, for example TF-IDF. TF-IDF score is suitable for weighting a term in a document. The formula is shown below.

$$\mathrm{w}_{\mathrm{i,j}} = \frac{freq_{i,j}}{max_l freq_{l,j}} \times \log \frac{N}{n_i}$$

The term frequency (TF) score provides a measure of how well the term describes the text content. It is presented in the formula as $freq_{i,j}$, term $k_i$ in document $d_j$, divided on the frequency of the term that occurs most in the text, $max_l freq_{l,j}$. Dividing on the max term gives a normalization of the frequency. The inverse document frequency (IDF) measures the inverse of the frequency of a term in a document collection. N is the total number of documents, while $n_i$ is the number of documents $k_i$ occurs. The main goal of this measurement is that a term that occurs in few documents is more suitable at distinguishing documents. [25]. The last element in the formula refers to checking for additional terms (AddTerm(U)) in the unit that imply that the unit has higher significance. This could be terms that appear in the heading, abstract, and so on. This strategy could be used to find sections about diagnosis, medicines, important information, and so on in the EPR.

ML is another approach of extracting information from texts. In this case the system is trained by a training set instead of predefined rules. An illustration of how a system like this works is illustrated in Figure 4 [28].

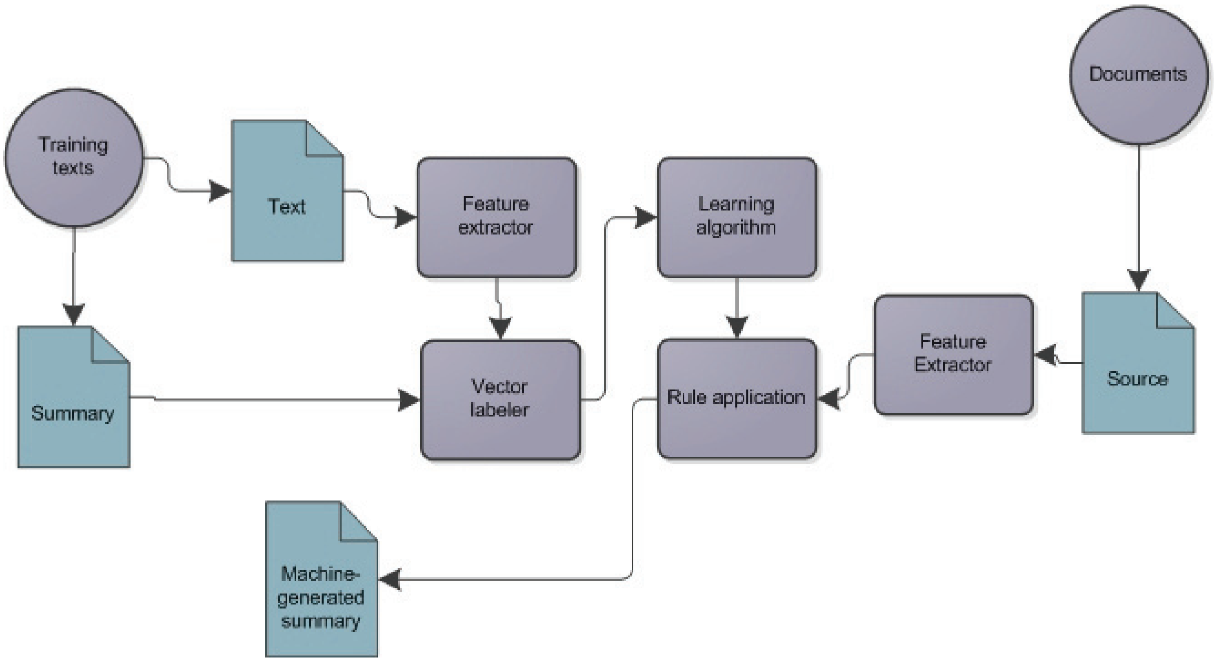


Figure 4 Summarizing through Machine Learning

Figure 4 describes two sets of documents, namely training texts and documents used for testing. The training texts have to have both a summary and a main text. The feature extractor assigns a vector to each sentence while the vector labeler compares the text and the summary, and then again the learning algorithm learns the rules for determining whether the sentence should be part of the summary or not.

### Named Entity Recognition

One application of information extraction is to extract entities, for example person names and connect them to the entity “Person”. There are many approaches in this area and it is possible to use both ML and KE. When lists of the different units belonging to an entity are available you do not need to use ML at all. This application is especially relevant for this thesis because of the possibility to extract, for example diagnosis codes.

### *2.3.5 Text mining tools*

This section will present different open source projects and tools that are possible to use during the implementation of text mining in the system.

### Lucene[1]

Lucene is a full text search engine which provides a tool for extracting sections based on keywords. This tool could be used in combination with LingPipe described below. As described in the specialization project Lucene provides functionality for giving spell suggestions based on edit distance [5].

Lucene provides features for highlighting special phrases. This is a functionality that might be used for summarizing EPRs. The class org.apache.lucene.search.highlight.Highlighter[2] gives the possibility to get a fragment from a text based on a score.

### Weka[3]

Waikato Environment for Knowledge Analysis (WEKA) is an open source project developed at the University at Waikato. The tool provides functionality for data mining either from an interface or directly from java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. Weka provides ML algorithms that are relevant for categorizing the terms and words in the EPR.

WEKA has been used with success in several other projects [29]. Berger and Merkl's approach is especially interesting since it is using n-grams for text classification [19]. WEKA provides a wrapper that support LibSVM[4] (a library for support vector machines), and a tokenizer for n-grams. Since WEKA is an open source project it is also possible to refine methods and algorithms if this is required.

The experimenter interface in WEKA gives the possibility to run experiments with different datasets and algorithms, and then compare the results with statistical methods.

### LingPipe[5]

LingPipe is another project in java that provides tools for linguistic analysis of human language. Some of the most relevant functionality is entity recognition, text classification, and correcting spelling based on a text. The tool suits to be combined with Lucene to provide a more complete functionality.

LingPipe provides relevant functions when it comes to entity recognition which can be useful to extract diagnosis codes, and perhaps other aspects in an EPR. In other areas like text classification

[1] http://lucene.apache.org/
[2] http://hudson.zones.apache.org/hudson/job/Lucenetrunk/javadoc//org/apache/lucene/search/highlight/Highlighter.html
[3] http://www.cs.waikato.ac.nz/ml/weka/
[4] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[5] http://alias-i.com/lingpipe/index.html

and word sense disambiguation LingPipe provides less classifiers then WEKA. In addition WEKA has a lot more possibilities for testing different datasets and classifiers in an easy way.

### Classifier4J[6]

Classifier4J is a java library for text classification. It only support naïve bayes and is therefore a poorer alternative than WEKA when it comes to classifying. The library provides ISummariser[7] which is an interface implemented by the class SimpleSummariser[8]. This functionality is an easy way of finding the words with the highest frequencies, and then presenting the first n sentences that contain these words.

[6] http://classifier4j.sourceforge.net/

[7]http://classifier4j.sourceforge.net/subprojects/core/apidocs/net/sf/classifier4J/summariser/ISummariser.html

[8]http://classifier4j.sourceforge.net/subprojects/core/apidocs/net/sf/classifier4J/summariser/SimpleSummariser.html

# 3 State of the art

This section describes the existing system developed at the RHF interventional centre and other similar existing systems that are relevant for this thesis. It is important to emphasize that the system developed in this thesis is a new approach and therefore there is no other equally existing system.

## 3.1 Existing system

The existing system is developed in two versions, namely Elena's master thesis prototype [3] and another system developed at RHF interventional centre. The latter is the basis for this thesis and is a server that the prototype developed in this thesis can use to translate terms. The server is treated as a "black box" and the architecture of the server is not discussed further. The thesis prototype by Elena [3] is developed with PHP and MySQL, and is described further in [3, 5].

The system used in this thesis consists of a database (MySQL) and a thesaurus server developed in C[9]. The system is documented by the developer and is available by contacting the author of this thesis. The Section gives an overall description of the existing system, and how the system will be integrated in the new prototype.

The thesaurus server receives XML requests and sends an XML in response with information about the translation. The request schema is illustrated in Figure 5.

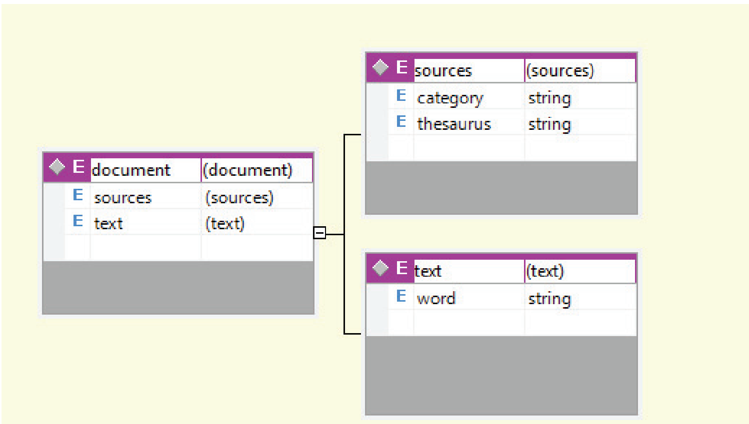


Figure 5 The request schema

[9] http://en.wikipedia.org/wiki/C_(programming_language)

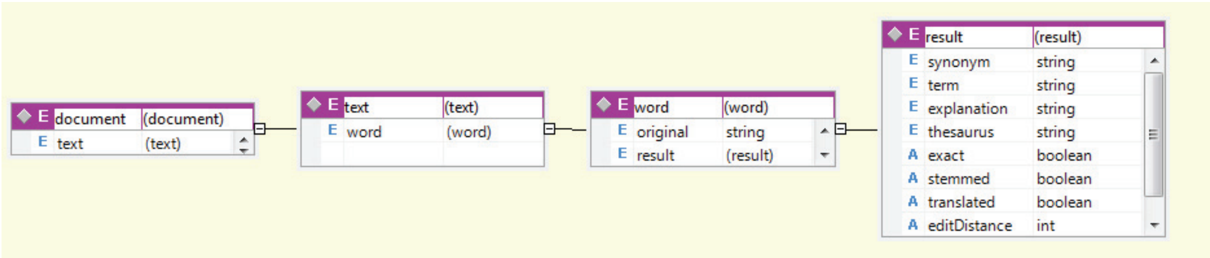


**Figure 6 Reponse schema**

Figure 6 describes the response from the thesaurus server. The prototype also has a client that could be used for translating EPRs directly from DocuLive, but as mentioned this is not relevant for this thesis.

When studying the existing thesaurus server it proved that the server already had functionality for giving spell suggestions based on edit distance. This functionality is interesting and is also a part of RQ2 based on the study executed in [5]. Since it unveiled that this feature was not totally complete and the thesaurus server implementation is outside the scope of this thesis it is not given further attention in this work.

## 3.2 Medical text mining applications

This section describes different applications that text mining has been applied to in other health informatics projects.

One approach investigates the value of diagnosis codes in EPRs [30]. The hypothesis is that diagnosis codes are set independently from EPR text, and not based on the text in the journal. The study concluded with a precision rate at 51.6% at the best, which is discussed further in the article. This discussion is not a topic in this thesis but states an example of using text mining in health informatics.

Another application described by Letrilliart uses string matching to automatic code reasons for hospital referral [31]. The system uses a look-up table referring to an International Classification of Primary Care (ICPC) code. The system was estimated to give 77% match rate, accuracy on 80% at code level. Røst and Nytrø also addresses the same issue using a document classifier trained with a set of manually coded EPRs [32]. This experiment gave an accuracy rate at 49.7%. Another interesting application looks at the possibility of categorization of medical text to improve the information retrieval [33]. When tested with queries related to diseases this gave gains as high as 84%.

Another relevant application of text mining is a study where the use of syntax was emphasized. Syntactic, lexical, and ontological information from UMLS where used with a semantic category recognizer to identify categories in discharge summaries [34].

A master thesis at NTNU [29] presents applications of text mining in EPRs and PHRs. The objective was to identify parts in EPRs, namely subjective, objective, and plan. When structuring information this way it may enhance information flow between EHRs and EPRs.

### 3.3 Health portals

This chapter describes the study of different functionality from health portals that could be implemented in the prototype. Some of this study has already been fulfilled in the specialization project [5], and this thesis will look further into one of these aspects. The prototype developed in this thesis is a part of the project minjournal.no[10] at RHF. This project aims at scheduling appointments and presenting EPRs to the patient.

Other similar projects are available, and some of them that are presented in this thesis are fetched from the specialization project [5]. Sundhed.dk[11] is a Danish project similar to minjournal.no which contains information about health and different conditions. There has also been adapted EPR to present them with relevant information according to International Classification of Primary Care (ICPC) codes. These codes are also used in Norwegian primary care and are a part of NEL. But many of these codes also contain reasons for encounter (the reason for the patient's visit to the physician) which could be confusing when using them in an adapted EPR [35].

This section is also fetched from the project [5] and describes MedlinePlus , the U.S. National Library of Medicine's health information website. This is another example of a health portal which contains medical information about diseases and other health related topics [36]. The system has over 700 topic pages and has linking to other sites among them a medical encyclopedia. An article from MedlinePlus is illustrated in Figure 7. The site also provides the national library of medicine's resources like basic information, learning, research papers, references, multimedia and other tools. Usability reviews of the portal have given some experiences that have to be taken to consideration. First of all linking to other sites is important to provide relevant information about a topic. There are many articles and information pages that can provide valuable information to the user. The challenge is to provide this information without taking the user away from the application, in this case MedlinePlus [36]. Pop-up windows with information about the fact that you are leaving the application or site could be annoying, and a problem especially related to pop-up blockers. Another important topic is how to fit dynamic information into a page display, and getting a consistent layout on different subjects with different information and modules, like search boxes etc [36].

Providing external information is an interesting extension to the patient friendly presentation of the EPR. There are different other relevant projects in this area [37, 38] that have tested the feature in different context. The following section is fetched from the specialization project [5].

---

[10] http://www.minjournal.no
[11] http://www.sunhed.dk

Patient Clinical Information System (PatCIS) [37] is a web-based system at the New York Presbyterian Hospital that patients use to view their own medical record and test results. The system has functionality that allows the patients to report data to the system, review information, and get education and advice. The interface has an “info button” that provides extra information that helps the patients in understanding the content. An example of medical information in the PatCIS system is presented in Figure 8. The experiences of the system were positive, and most of the patients agreed that the use of the system had improved their communication with their physicians. One of the positive elements that were discovered was that the system allows the patient an active role in his or her health care and also improves their understanding of their health.

Different approaches have been tested, and one article emphasizes the importance of using the whole context of diagnosis when searching for information [38]. The article looks at the aspects of providing information to the patient. Medical information provided by the health care professionals has high relevance and safety, but is not available at all times and has no confidentiality and selectivity. Information provided by the web has good availability, confidentiality, and selectivity but varying relevance for the patient, and no safety. The different websites also has to be classified ensuring correct information of high quality. In the project work the patient handbook, a part of NEL, was emphasized since the quality assurance is already taken care of. This addresses the issue of safety, because the information from this source as seen as safe. Together with information extraction, described in Section 2.3.4, the extracted information could be used while searching in the handbook.

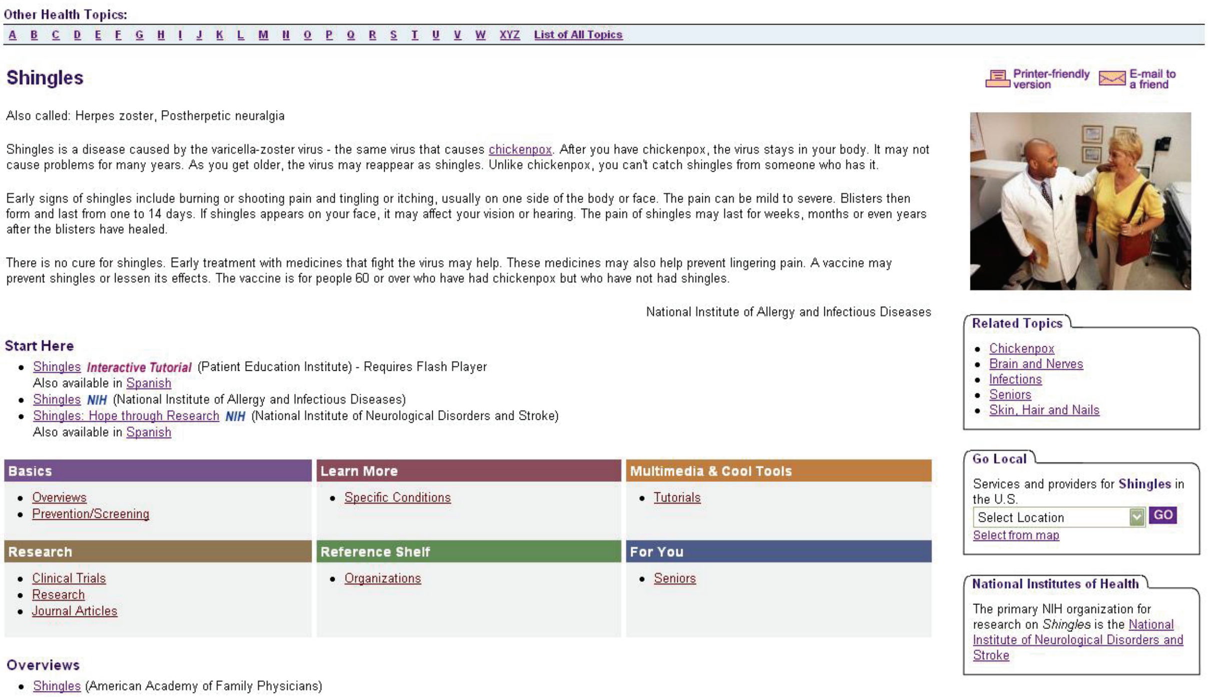

**Figure 7 Example of medlineplus article**

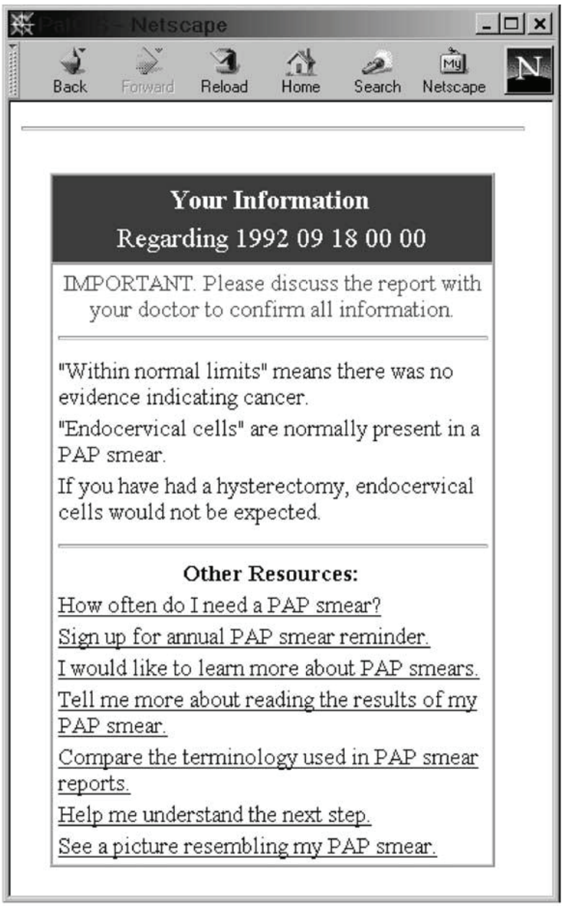


Figure 8 Patcis term explanation

### *3.3.1 Searching the patient handbook*

Initial searching showed that the patient handbook gives a lot of irrelevant hits when searching on long strings with different words. It seems that the search engine uses the “or operator” and therefore gives hits on documents only containing one of the keywords. Since the search engine ranks the different documents according to relevance it might be relevant to only present the documents with the highest ranking.

The patient handbook seems like a good source for information because of the provided quality assurance of information, and the fact that the articles are presented in Norwegian.

## 3.4 Personal Health Records

Personal health records (PHR) are highly relevant and are applications that provide the patient with a personal portal that presents medical information. The portals are available on the web and allow the patient to enter their own medical information and get an updated view and track of their own health and medical history [39].

The following section is taken from the specialization project [5].

Kim and Johnson accomplished a study and evaluation of PHR user-interfaces that was primarily focused on input methods. Different methods like free text, pick lists, radio buttons, check box, and dichotomous radio button were tested against thyroid patients. The results were varying, but there were some quite good indications that guidance of input was an important factor. The conclusion was that free text should be preprocessed to avoid the patient putting extra and uninteresting information [40].

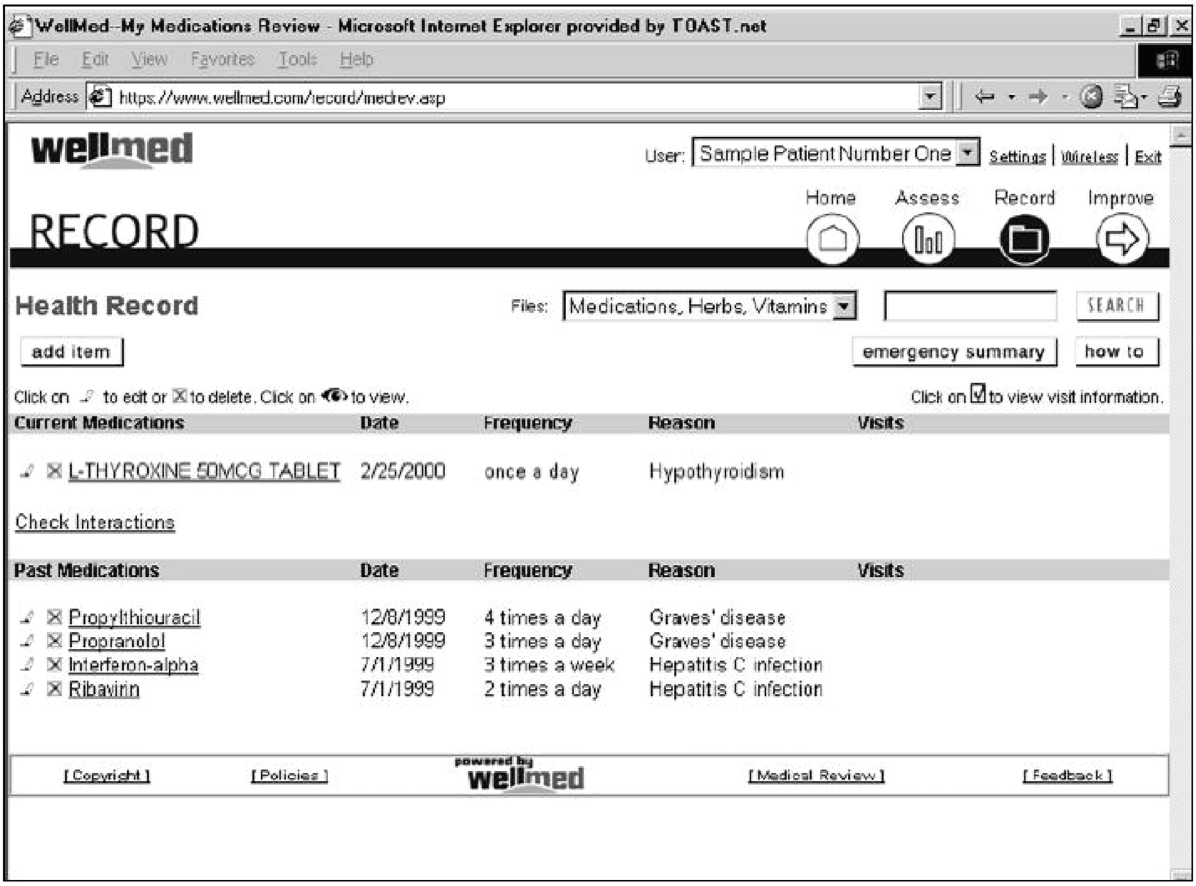

Figure 9 Wellmed PHR

An example of an online PHR is presented in Figure 9 and shows the current and past medications of one patient. A summary functionality that summarizes the patient’s EPR could be a useful feature for getting the correct information as input into a PHR system. As stated above the guidance of input is an important ensuring precise and correct input. A combination of guidance and predefined fields in a summary could help the patient finding the correct information. A summary would also save the patient time looking for the relevant information in long and extensive EPR texts.

## 3.5 ICD Codes

International Classification of Diseases (ICD)[12] codes are used in Norwegian hospitals to classify diseases and related health conditions. Examples of ICD codes are presented below in the EPIKRISE. There are different revisions of this system, but the 10th revision is the one used in Norwegian hospitals.

The textbox below gives an example of an EPIKRISE which is a record that is written after each hospitalization. This record is a kind of summary of the stay at the hospital and gives precise diagnosis according to the patient's health condition.

Diagn./pros.: H L309 Uspesifisert dermatitt
B L011 Impetiginisering av andre dermatoser
B D441 Svulst med usikkert/ukjent malignitetspotenisal i binyre
B I10 Essensiell (primær) hypertensjon
O TQX00 04.08.05 13:15 Hudbiopsi

These codes could be used further to summarize information about the patient, and then again provide external information.

[12] http://www.volven.no/produkt.asp?id=8&catID=3&subID=9&oid=

# 4 Summary

The classifiers naïve and complement bayes, decision trees, and support vector machine are relevant for categorizing documents. Using these classifiers on word sense disambiguation, techniques like character n-grams could be applied. This technique is usual when looking at language classification which is similar to disambiguating medical terms and Norwegian words. Evaluation of classifiers is usually done by looking at the percent correct classifications on a test set, namely accuracy. The rates are evaluated with a paired t-test to determine whether the results from the different classifiers are significant better or worse. Other measures are a comparison of correct and wrong classifications in one of the classes. This evaluation can be used to take costs into consideration.

The existing server, developed at the RHF interventional centre, uses XMLs messages to communicate with other systems, and send them translations. There are different approaches of making portals presenting EPRs and health information. One interesting approach is the PatientKB where external information from Google is presented. PatCIS is another system presenting EPRs with explanations for the laymen.

Text mining has been used in medical applications in different areas, but the main issues looked are automatically diagnosis coding of EPRs based on the text and studying whether the codes are set independently from the text or not. Other studies have looked at the structure of the EPR trying to easy information flow between EHRs and EPRs.

# Part III Implementation and results

# 1 Part Introduction

This chapter gives a summary of the purpose and scope in this part, and an overview of the different chapters.

## 1.1 Purpose

This part presents the implementation of the different aspects in this thesis. It will present an overall architecture and the different parts the system consists of. The purpose is to get an overview of the implementation, how the different parts connect, and reasons for some of the choices.

## 1.2 Scope

The chapters in this part describe the system parts, how they are implemented, and the main results of the implementation. There will be no extensive evaluation and discussion of the different results.

## 1.3 Overview

This part contains the following chapters:

- Overall system description: Presents the overall architecture of the prototype, and its user interface.
- Word sense disambiguation: Describes the text mining approach on separating Norwegian words from medical terms. The results of the different approaches are also presented here.
- Summarization: Gives an overall presentation of the implementation of EPR summarizing.
- External information: The architecture and approach of getting external patient information.
- Summary: Gives a summary of this part.

# 2 Overall system description

This section presents the overall system architecture with the new extensions of the system. The improvements of the user interface from [5] are also presented in this section although they are not evaluated in this thesis.

## 2.1 System architecture

The system architecture is based on the architecture presented in [5], and further developed in this thesis to the system described in Figure 10. The arrows describe the communication between the different components while the functionality included in the EPRPortal is placed within this box.

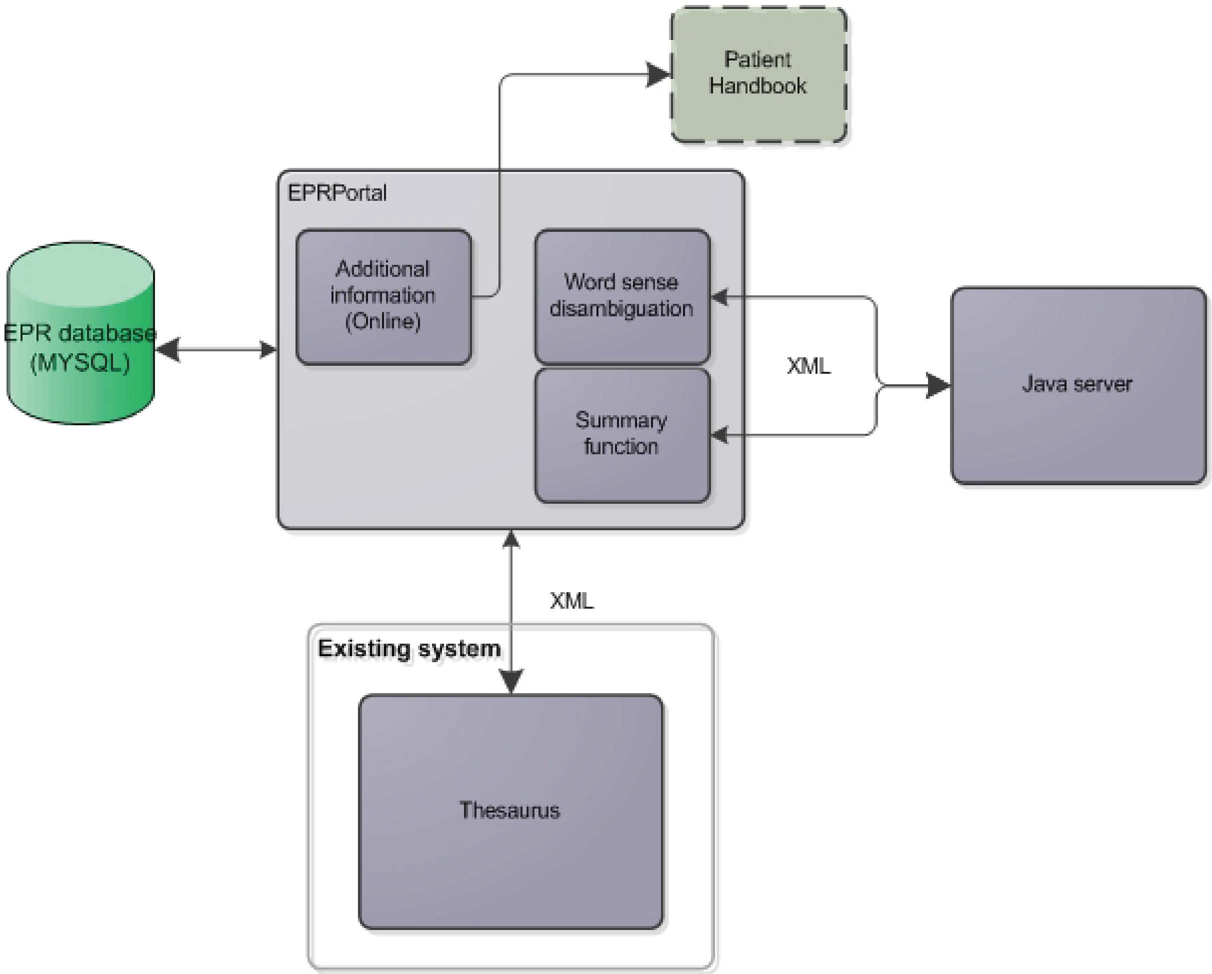


Figure 10 Overall system architecture

The architecture described here is based on the thesaurus server, see Section Part II 3.1, and is limited to the functionality relevant for this thesis. The EPRPortal is a web portal developed in PHP and its purpose is to present EPRs in a patient friendly way. This portal is presented further in Section 2.2. The java server offers web services to the portal, and makes it possible to combine java and PHP applications.

The text mining application is implemented in Java and therefore executed on the Java server. The same goes for the information extraction unit. The EPRPortal gets the EPRs from a MySQL[13] database containing some example records. The EPRs are translated by the thesaurus server and words not known for the server are tested by the text mining application. If the word still is unknown it is presented as a term without translation. The html parser is implemented in the portal using PHP and Simple HTML DOM Parser[14].

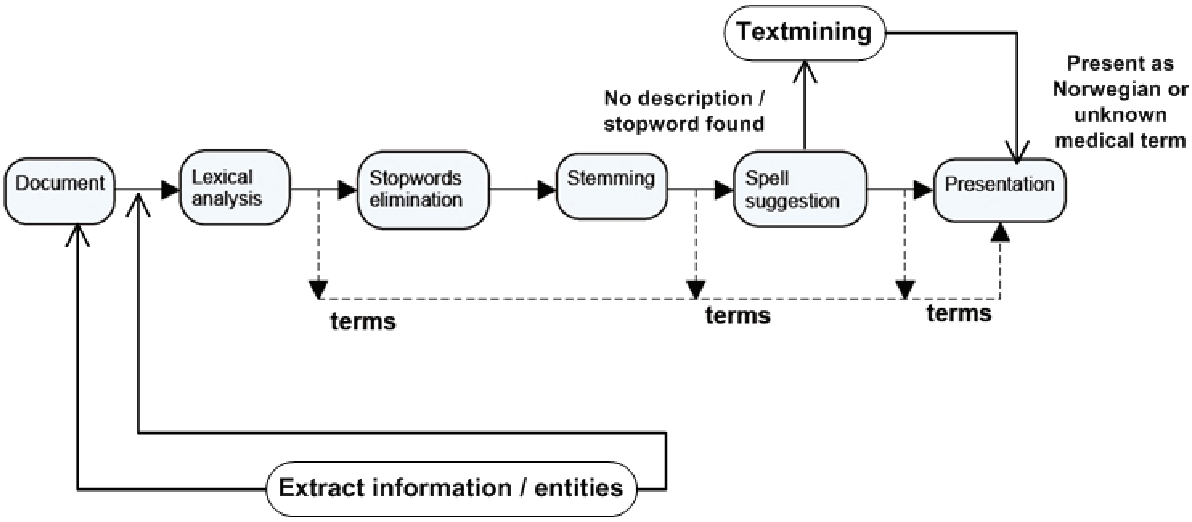


Figure 11 Flow diagram for translating EPRs

Figure 11 presents the translation process based on the IR process figure in [3]. The functionalities for extracting information to a summary and the word sense disambiguation functionality is added. When extracting information, illustrated by the arrow before "Lexical analysis", the summary is translated instead of the complete EPR.

Implementation of collocation will probably improve the system significantly, see Part II 2.1. Looking at the architecture, the implementation of this functionality should be done in the thesaurus server, see Part II 3.1. Since this is outside the scope of this thesis the implementation of this functionality is not fulfilled in this study

13 http://www.mysql.com
14 https://sourceforge.net/projects/simplehtmldom/

### 2.1.1 *Web services*

The communication between the EPRPortal and Java applications is implemented with Web services. The JAX-WS[15] framework is used to develop a Java web service, while PHP uses PHP Soap[16] to communicate with the service. A web service uses the SOAP[17] protocol to communicate. The PHP Soap framework uses the Web Service Description Language (WSDL[18]) to understand what services the server provides. SOAP is a communication protocol based on Extensible Markup Language (XML[19]). An example of a SOAP request is provided below.

```
POST /WebApplication1/SpellCheckerService HTTP/1.1
Host: 10.0.0.1:8080
Connection: Keep-Alive
User-Agent: PHP-SOAP/5.2.5
Content-Type: text/xml; charset=utf-8
SOAPAction: ""
Content-Length: 1199


<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:ns1="http://spell.me.org/"><SOAP-
ENV:Body><ns1:summary><journal>ALLERGIER:…..</journal></ns1:summary></SOAP-
ENV:Body></SOAP-ENV:Envelope>
```

[15] https://jax-ws.dev.java.net/
[16] http://ua.php.net/soap
[17] http://www.w3.org/TR/soap/
[18] http://www.w3.org/TR/wsdl
[19] http://www.w3.org/XML/

## 2.2 User interface and extended functionality

The user interface was refined taking the eight golden rules [41] into account and some of the results are presented in Figure 12.


**Figure 12 Refined user interface**

The interface is designed in PHP[20], HTML[21], and JavaScript[22]. The source code is enclosed in Appendix A.

Figure 13 shows the class diagram for the EPRPortal developed in PHP. The main class is the index file that presents EPR data to the user. The index file uses the classes Thesaurus for translating the EPR, and the Epr class for fetching EPRs from the database. The Thesaurus class is responsible for communicating with the thesaurus server while the EPR class communicates with the sql server and fetches the EPRs and ICD descriptions. The html_dom_parser parses the search results from the patient handbook and fetches the articles that will be presented in the portal.

[20] http://www.php.net/
[21] http://www.w3schools.com/html/default.asp
[22] http://www.w3schools.com/JS/default.asp

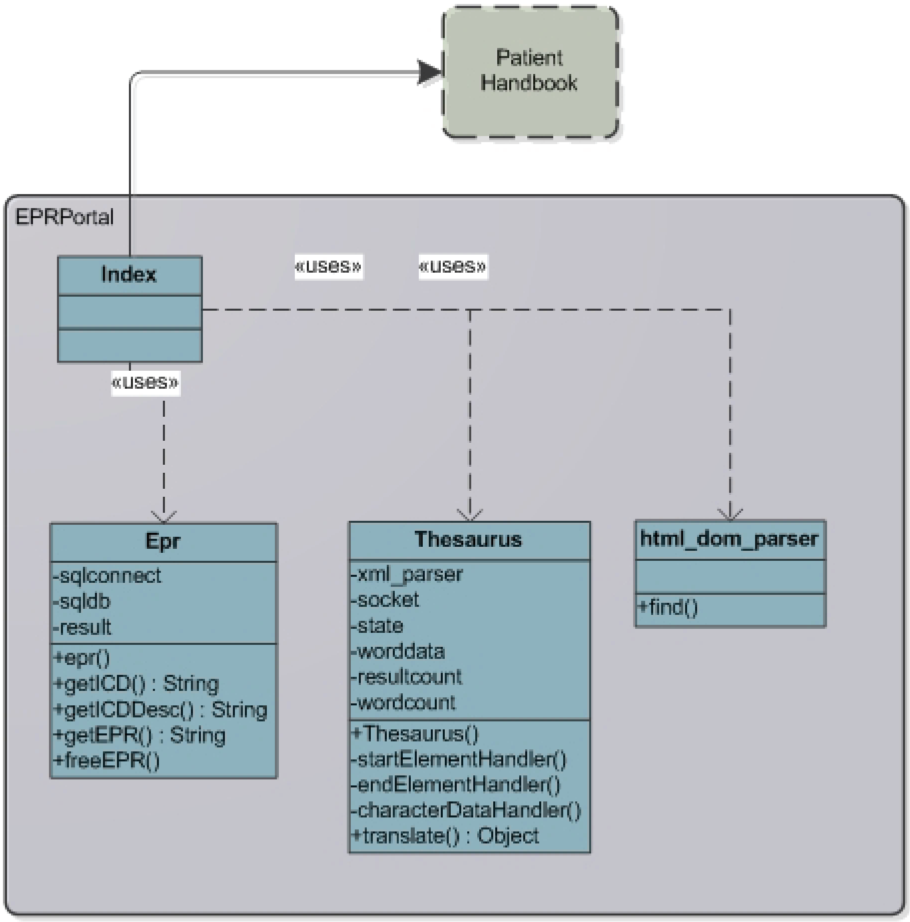


Figure 13 Class diagram

The main changes in the user interface are the development of a word list that displays all the terms with the translations at all times. This leads to a lower short time memory load on the user compared to the previous approach [41]. The list displays the terms and emphasizes the term that mouse is pointed on in the EPR text.

The user interface has two extensions that are relevant for this thesis which is external medical information and the summary functionality. The external information unit is illustrated in the screenshot; see Figure 14 and Figure 15.

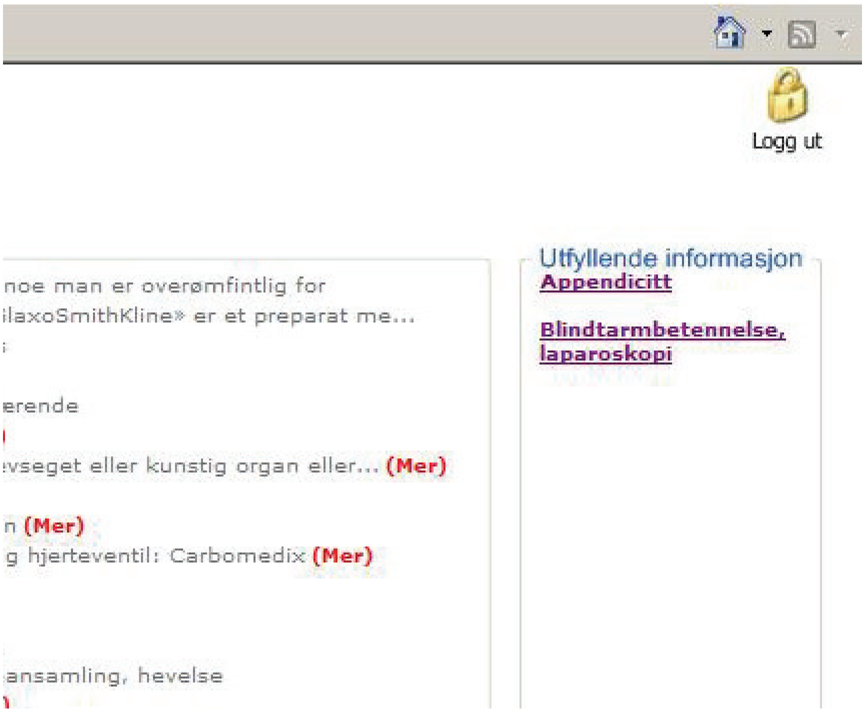


Figure 14 External information

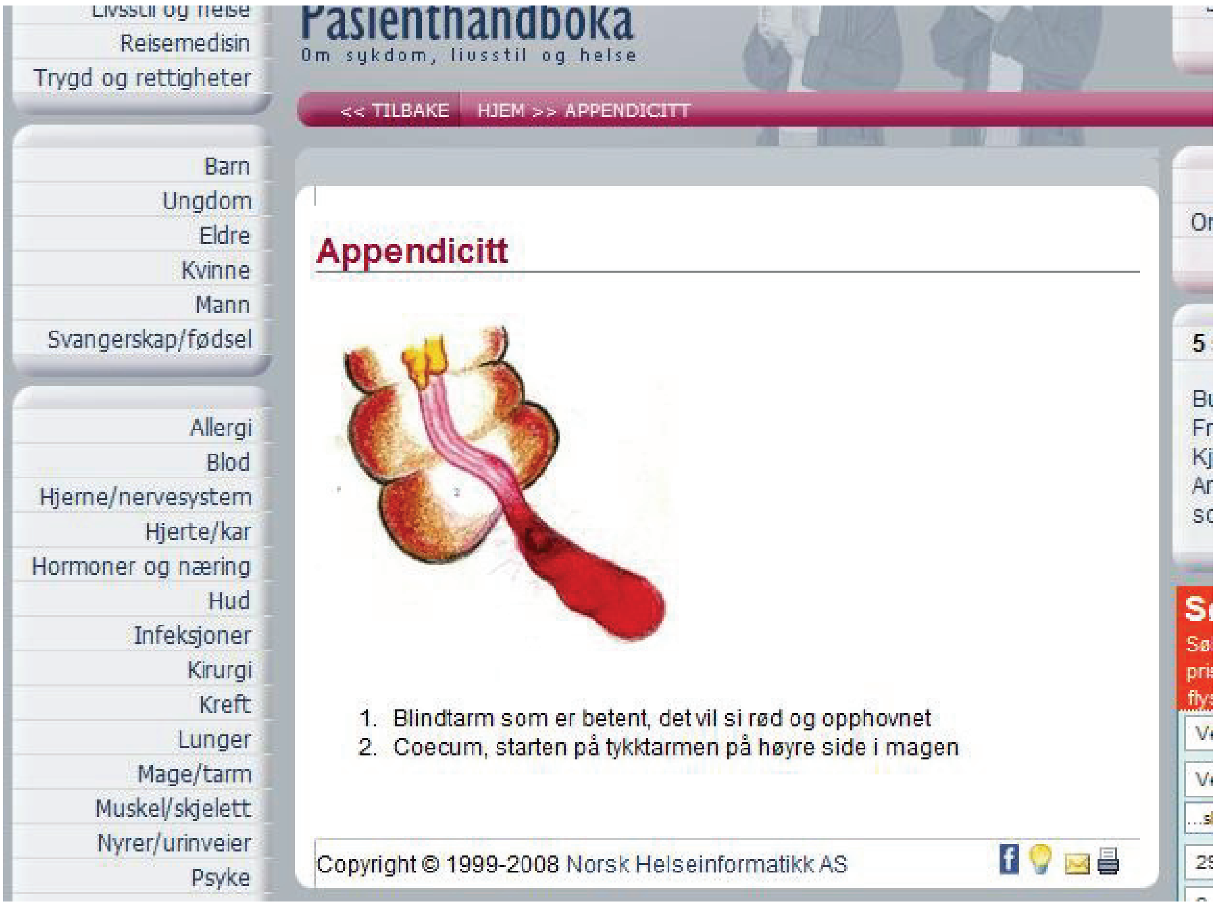


Figure 15 The patient handbook about Appendicitt

# 3 Word sense disambiguation

Word sense disambiguation was implemented using text mining. This section describes the implementation of the experiment which includes training, testing, and preparation of the results.

## 3.1 Implementation

Weka, which is presented in Part II, is chosen as the best suitable tool for executing this experiment. There many arguments for choosing this open source project, but one of the most important is the support of many different variants of classifiers. The tool also supports n-gram tokenization and has functionality that can be used to evaluate and analyze the results.

### *3.1.1 WEKA*

The setup of different classifiers and WEKA is described in this section. During the classifying and evaluation the graphical user interface was used. This choice was made because the interface supports comparing the results through statistical analysis. In addition the graphical user interface saves valuable implementation time.

#### Preprocessing

To use strings in text mining they have to be manipulated and one way of doing this is a word vector[5]. The StringToWordVector class in WEKA transforms the strings into vectors with numeric attributes. The vector can represent whether or not a word is present in a string, or the frequency of the different words. Other techniques like inverse documents frequency can also be used on the vector. These techniques are further presented in Part II 2.3.2 and Part II 2.3.4. The StringToWordVector class[23] gives the possibility to apply IDF scores on the record in addition to use a tokenizer called NGramTokenizer[24]. The combination of these two classes gives the possibility to create character n-grams in a vector representation. Table 1 presents an example vector with character bigrams where grams are represented with 1 or 0. The vector represents the word "dansen" with the 2-grams da, ns, and en.

| Da | Er | Ns | si | en | ha | Hu |
|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 1 | 0 | 0 |

**Table 1 Character 2-gram vector**

The StringToWordVector filter was initialized with the following parameters:

[23] http://weka.sourceforge.net/doc.dev/weka/filters/unsupervised/attribute/StringToWordVector.html
[24] http://weka.sourceforge.net/doc.dev/weka/core/tokenizers/NGramTokenizer.html

weka.filters.unsupervised.attribute.StringToWordVector -R first-last -W 1000 -I -N 0 -L -stemmer weka.core.stemmers.NullStemmer -M 1 -tokenizer "weka.core.tokenizers.NGramTokenizer -delimiters ; -max 3 -min 2"

This filter was applied with two different parameters, first allowing only bigrams, and secondly allowing both bi and trigrams.

Initial testing showed that idf scores gave higher accuracy because of its ability to emphasize grams that separates the instances from each other. The WEKA NGramTokenizer is originally aimed at word n-grams, and not character n-grams. This issue was solved by manipulating the training data, and separating each character instead of word. This is described further in Section 3.2.

The sample size of the Norwegian literature is likely to be larger than the medical dictionaries. In addition a too large dataset would lead to very long training time. To deal with these issues a combination of over and under sampling is used. The medical dataset is oversampled while the literature dataset is under sampled. This effect is achieved by using the Resample filter. The filter uses a combination of over and under sampling to balance datasets and reduce bias.

weka.filters.supervised.instance.Resample -B 1.0 -S 1 -Z 25.0

The filter is used with parameters specified above. It is set to resample the datasets to equal sizes, and reduce the size to 25% of the original size. The reason for reducing the sample size is to make it possible to handle in the available memory on the computer, and make the training time reasonable.

### Classifiers

The classifiers that were used in this thesis are naïve bayes[15, 17, 18], complement bayes[17], support vector machines[15, 20, 21], and the C45 decision tree[15, 19-21]. All these classifiers are implemented in WEKA, and could be reused during this work. The parameters of the different classifiers are set at standard values assuming this is the best approach without using time tuning each of them. This statement is further elaborated in the next section.

Naïve bayes is implemented through the class NaiveBayes[25], complement naïve bayes with ComplementNaiveBayes[26], and decision tree with the j48 package[27]. The configuration of complement bayes and j48 is stated below.

weka.classifiers.bayes.ComplementNaiveBayes -S 1.0
weka.classifiers.trees.J48 -C 0.25 -M 2

---

[25] http://weka.sourceforge.net/doc/weka/classifiers/bayes/NaiveBayes.html
[26] http://weka.sourceforge.net/doc/weka/classifiers/bayes/ComplementNaiveBayes.html
[27] http://weka.sourceforge.net/doc/weka/classifiers/trees/j48/package-frame.html

The support vector machine is implemented with a wrapper class LibSVM[28] which uses the LibSVM library for support vector machines[29]. The parameters used with this classifier are specified below.

weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.0010 -P 0.1

## Parameters

The different parameters could be tuned to achieve higher accuracy. WEKA provides a functionality called grid search[30] which performs a search for the best pair of parameters for the classification. The tuning of parameters together with cross validation testing could make the training computationally infeasible [42]. If we were to test each parameter with all combinations together with a 10 fold cross validation it could lead to as many as 10 million runs [42]. This could again lead to several weeks of training for some of the models in this thesis, and therefore it is not possible to complete it within the time limits of this thesis.

Since this is not the main focus of this thesis, the parameters will be set to standard WEKA values except for testing both bigrams and trigrams.

[28] http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/LibSVM.html
[29] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[30] http://weka.sourceforge.net/doc.dev/weka/classifiers/meta/GridSearch.html

## 3.2 Training

The training of the classifiers is achieved by using data from fiction literature and medical dictionaries. The data is transformed into n-grams, and then again into gram vectors as stated above. The Norwegian literature is fetched from the Oslo corpus[31] which is set of tagged literature and articles in Norwegian.

The Oslo corpus contains about 18.5 million words in the “Bokmål” variant of Norwegian. All these words are fetched from fiction, factual prose, newsletters and articles. These words had to be separated into single words and letters. The files were manipulated into WEKA standard format (arff) through scripts in textpad. The textbox below shows an example of an arff file.

```
@relation category
@attribute term string
@attribute class {1 2}
@data


A;D;V;A;R;S;E;L;E;N,2
A;L;F;R;E;D,2
A;N;S;I;K;T;E;T,2
A;N;T;O;N,2
A;V;L;Y;T;T;I;N;G,2
```

The example shows that the words are separated by lines, and the characters separated by semicolon. The N-gram tokenizer together with the StringToWordVector gives vectors with n-grams. The words are separated into two classes, namely Class II which is fiction, or ordinary Norwegian, and Class I which is medical terms.

Initial testing with the complete corpus gave problems with both memory size and training time. This is the main reason for limiting the data amount. The corpus was reduced to only using five fiction texts, namely:

- Hardy-guttene og den mystiske karavanen
- Davids bror av Kjell Askildsen
- Høst i mars av Georg Johannesen
- Sporet av en sti av Bernt Vestre

[31] http://www.tekstlab.uio.no/norsk/bokmaal/

In addition the resampling process described in Section 3.1.1 was executed to achieve reasonable training times. After running the resampling filter the set consisted of 22449 instances of medical terms, and 22324 instances of Norwegian fiction literature.

The training times on the C45 decision tree showed to be a problem, especially when running 10 fold cross validation. The training time, when only running this validation once, exceeded several days. This lead to the fact that this classifier is not validated as good as the other classifiers in this thesis.

## 3.3 Testing

The testing of classifiers is performed in two different ways, one using the training data, and the other using a small amount of EPRs. One of the main challenges in this part of the work is to get a reasonable amount of already classified data. The available EPRs are not tagged with classifications which led to a big amount of manual work. Because of the small amount of available testing data, cross validation with the training dataset has to be used. The EPRs that will be used in the testing has to be tagged manually with either class 1 or 2.

The testing was executed through the WEKA experimenter, and the datasets and classifiers were applied in two rounds. First the two datasets with different n-grams were trained and compared with both naïve and complement bayes. Then the support vector machines and complement bayes were compared with the two same datasets. Figure 16 shows the experimenter in WEKA ready to run training and testing with LibSVM and complement bayes.

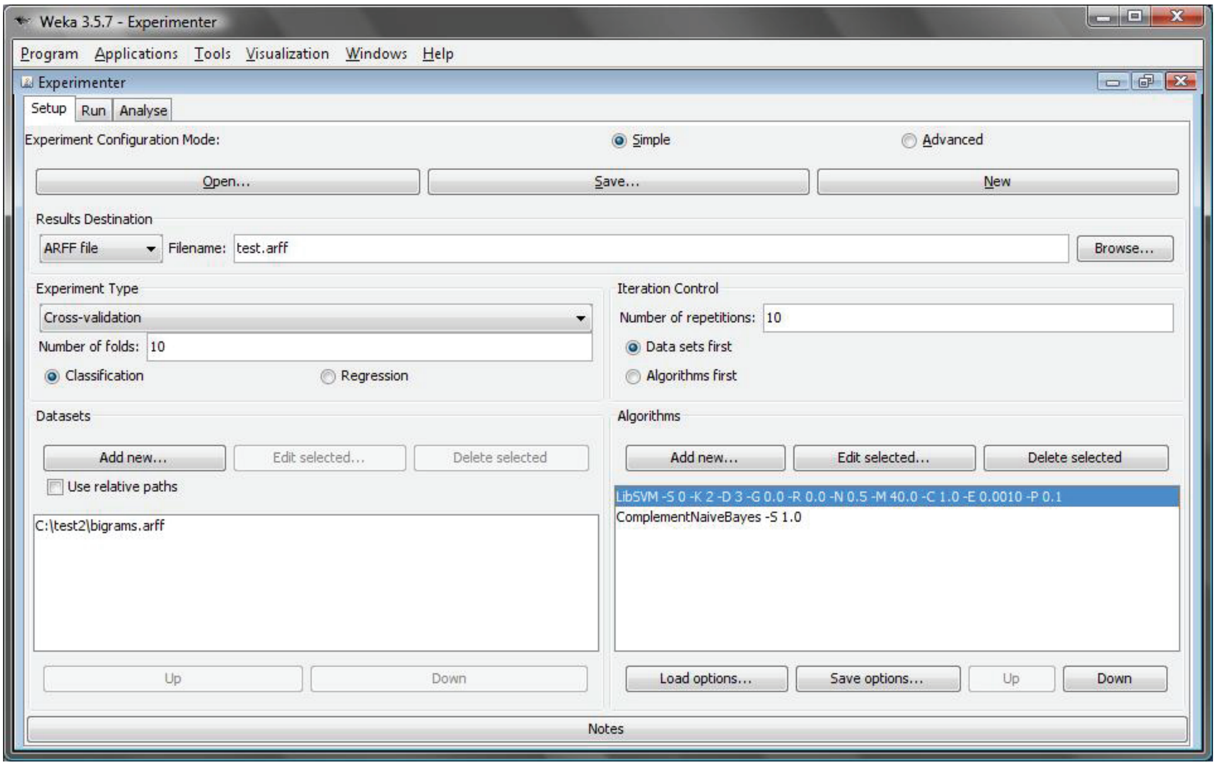

Figure 16 Experimenter comparing complement bayes and support vector machines

### 3.3.1 *Cross validation*

Based upon the studies in Part II 2.3.3 cross validation was executed with 10 folds, and repeated 10 times. The training corpus is divided into 10 folds, and the classifier is tested on each of the folds while the rest of them are used in the training. This process is illustrated in Figure 17. The cross validation is also stratified which is a process that ensures that each class is properly represented in the folds.

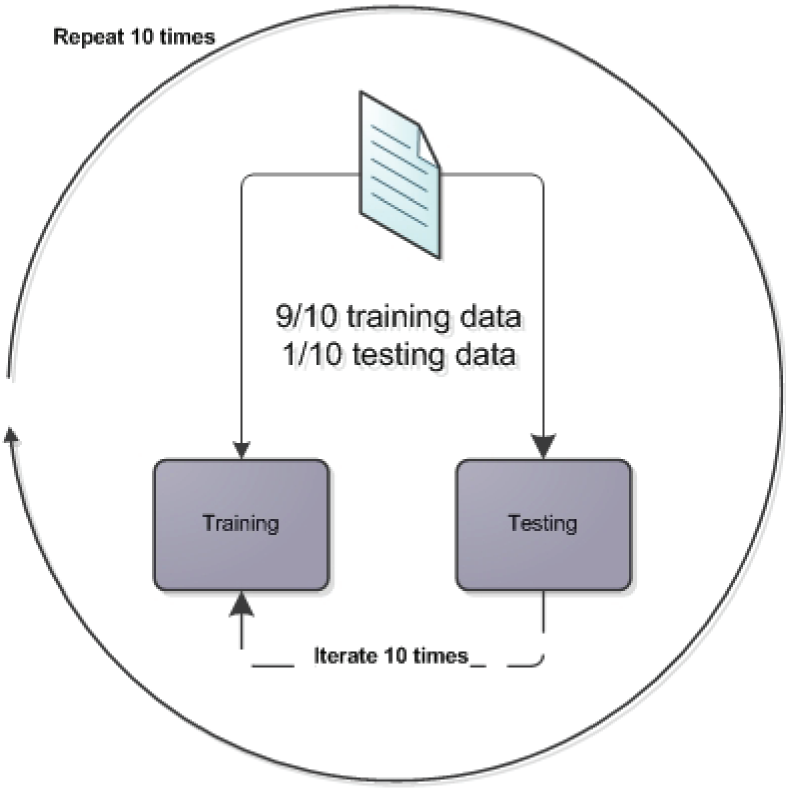


Figure 17 Flow diagram illustrating cross validation

Since the J48, an implementation of C45 decision tree, algorithm seems to be too time-consuming for this validation the classifier is only run once with 10 fold cross validation. This restriction results with a challenge when evaluating the results.

### *3.3.2 Testing on EPRs*

When testing on EPRs the text had to be manually classified, this resulted in a test set with about 900 words. The words were tested with all the classifiers mentioned above, but this time the test is only run once on the test set.

## 3.4 Results

This section will present the results from training and testing of different classifiers and datasets. The analysis of these results will be presented in Part IV.

The cross validation results are presented in the following section. The tests of classifiers are divided into different parts because of the computing complexity. The results are to be evaluated after the criteria presented in Part II 2.3.3, with focus on recall, precision, kappa-statistics, and accuracy. The different results will be evaluated and compared with t-tests and ROC plots.

### *3.4.1 Cross validation*

The first part of the cross validation is the testing of naïve and complement bayes. A complete overview of the results is attached in Appendix F. The accuracy results are presented in Table 2 and the kappa-statistic are presented in Table 3.

| | Naïve Bayes | Complement Bayes |
|---|---|---|
| Only bigrams | 78,31% | 85,62% |
| Bi/trigrams | 79,82% | 80,12% |

Table 2 Accuracy for naive and complement bayes

| | Naïve Bayes | Complement Bayes |
|---|---|---|
| Only bigrams | 0,57 | 0,71 |
| Bi/trigrams | 0,60 | 0,60 |

Table 3 Kappa statistics for naive and complement bayes

As it is possible to see complement bayes with bigrams has higher accuracy and kappa statistic than naïve bayes. Bigrams performs better than trigrams with complement bayes.

Other interesting measures are precision and recall which are presented in Table 4 and Table 5.

| | Naïve Bayes | Complement Bayes |
|---|---|---|
| Only bigrams | 0,74 | 0,84 |
| Bi/trigrams | 0,76 | 0,75 |

Table 4 Precision for naive and complement bayes

| | Naïve Bayes | Complement Bayes |
|---|---|---|
| Only bigrams | 0,88 | 0,88 |
| Bi/trigrams | 0,88 | 0,91 |

Table 5 Recall for naive and complement bayes

Complement bayes performs better on precision and equal on recall for bigrams while naïve bayes gives better precision but poorer recall on trigrams.

The second part of the cross validation is comparing complement bayes with support vector machines, and the results are presented in Table 6, Table 7, Table 8, and Table 9.

| | Complement Bayes | SVM |
|---|---|---|
| Only bigrams | 85,62% | 93,43% |
| Bi/trigrams | 80,12% | 93,23% |

Table 6 Accuracy for complement bayes and support vector machines

| | Complement Bayes | SVM |
|---|---|---|
| Only bigrams | 0,71 | 0,87 |
| Bi/trigrams | 0,60 | 0,86 |

Table 7 Kappa statistics for complement bayes and support vector machines

| | Complement Bayes | SVM |
|---|---|---|
| Only bigrams | 0,84 | 0,95 |
| Bi/trigrams | 0,75 | 0,96 |

Table 8 Precision for complement bayes and support vector machines

| | Complement Bayes | SVM |
|---|---|---|
| Only bigrams | 0,88 | 0,92 |
| Bi/trigrams | 0,91 | 0,90 |

Table 9 Recall for complement bayes and support vector machines

From these tables it obvious that SVM outer performs complement bayes on almost all the measures. SVM gives better accuracy, kappa statistic, precision, and recall on bigrams. Complement bayes gives higher recall for trigrams. Bigrams also seems to gives best results except for precision when it comes to the SVM classifier.

The J48 algorithm was as stated in Section 3.3.1 highly time consuming. A complete run on this algorithm was not possible to complete within the scope of this thesis. When trying to run the trigrams the training time increased so much that it was not possible to complete. Therefore the only result available are 10 fold cross validation on bigrams run once, presented in Table 10. The complete overview of the runs is attached in Appendix G.

| | J48 Decision tree |
|---|---|
| Accuracy | 95,42% |
| Kappa statistics | 0,91 |
| Precision | 0,96 |
| Recall | 0,95 |

Table 10 Accuracy, kappa, precision and recall for the J48 decision tree

The available results with J48 are better or equal on all measures compared to SVM.

### 3.4.2 EPR test

The results of testing with real EPRs are presented below, namely Table 11 and Table 12. A complete overview of the results is attached in Appendix E.

| | Naïve Bayes | | Complement Bayes | | SVM | | J48 | |
|---|---|---|---|---|---|---|---|---|
| Accuracy | 60,87% | | 74,36% | | 80,94% | | 80,27% | |
| Kappa-statistics | 0,2334 | | 0,4546 | | 0,5765 | | 0,551 | |
| | Class I | Class II | Class I | Class II | Class I | Class II | Class I | Class II |
| Precision | 0,396 | 0,843 | 0,53 | 0,906 | 0,618 | 0,929 | 0,617 | 0,908 |
| Recall | 0,735 | 0,559 | 0,81 | 0,717 | 0,846 | 0,795 | 0,791 | 0,807 |

Table 11 Results from tests with bigrams

| | Naïve Bayes | | Complement Bayes | | SVM | | J48 | |
|---|---|---|---|---|---|---|---|---|
| Accuracy | 62,215% | | 70,46% | | 79,26% | | 77,26% | |
| Kappa-statistics | 0,2596 | | 0,419 | | 0,5148 | | 0,4762 | |
| | Class I | Class II | Class I | Class II | Class I | Class II | Class I | Class II |
| Precision | 0,409 | 0,857 | 0,487 | 0,94 | 0,612 | 0,883 | 0,578 | 0,877 |
| Recall | 0,759 | 0,568 | 0,897 | 0,629 | 0,723 | 0,82 | 0,715 | 0,795 |

Table 12 results from tests with trigrams

The results present precision and recall with separate estimates for each class, class I represents the medical terms while class II represents Norwegian fiction literature. One important issue is the fact that a medical term classified as a normal Norwegian word is more serious error than the other way around, which leads to the fact that class II precision is more important in this context. This issue is further discussed in Part IV.

From these results it is possible to observe that SVM has higher accuracy and kappa statistic than the others. Precision and recall using bigrams are better with SVM except for J48 Class II recall. The results with bigrams also seems better than trigrams except for some exceptions, namely complement bayes Class II precision, and SVM Class II recall. In addition naïve bayes performs better with trigrams.

### 3.4.3 Combining the results

Looking at the results the first obvious observation is that SVM outer performs the other classifiers. The results when testing with real EPRs gives better results when using bigrams while the cross validation seems to give higher scores with trigrams. The differences between these measures are not that high, and it is important to notice that trigrams have higher computing complexity. The other classifiers seem to be performing better using bigrams than trigrams with some exceptions.

Complement bayes performs better compared to naïve bayes which is the one in these tests with the poorest performance. J48 performs almost on the same level as SVM except for the issue of training and testing time which are too high when it comes to large datasets. The precision measurement of Class II stated as an important issue when classifying in this subject area SVM has the best results except for complement bayes using trigrams. When using cross validation the Class II precision is not available because the precision in this experiment is calculated from Class I. Complement bayes good Class II precision is discussed further in Part IV, and compared to cross validation with True Negative and False Negative rates.

# 4 Summarization

This section will look at implementation of the summary functionality which provides the patient with a summary of his or her health condition. The EPRs have an EPIKRISE with diagnosis codes and descriptions of the health condition. These codes might be a good approach to get a summary. One of the main challenges is then to extract the information, in this case the diagnosis codes in the EPR. The diagnosis codes are described in Part II 3.4, and theory about information extraction in Part II 2.3.4. Another approach is to use the position of the sentence and the most frequent words to find sentences that summarizes the text.

The sections below present two alternative implementations of this application.

## 4.1 Sentence extraction

Extracting sentences according to their weight is a possible way of solving the summarization problem. The implementation of this functionality is achieved by using Classifier4J, see Part II 2.3.5. The source code is attached in Appendix C.

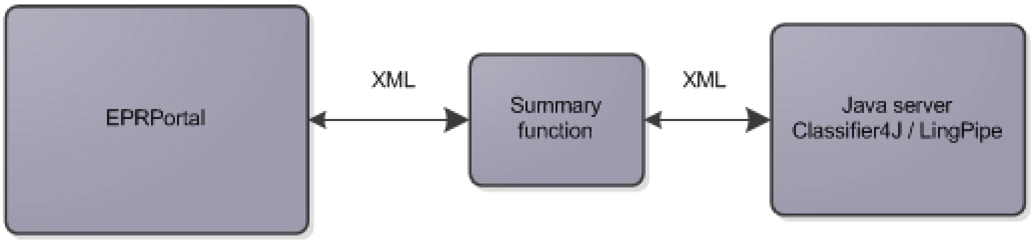


Figure 18 Summarizing functionality

Figure 18 illustrates how the EPRPortal will communicate with the java implementation. The sentence extraction server will receive the EPR and extract the four most relevant sentences based on the most frequent words. The system searches for the most frequent words and returns the first sentence that contains each of these words. In this case four sentences are returned. Using the first sentence is the same as giving the first sentences higher weight, thinking that the first part summarizes the document. This goes especially for articles where abstract often is the initial paragraph. It is important to take into consideration that this might not be the case with EPRs.

An example of a patient record with summary is presented in the textboxes below.

ALLERGIER: Inge kjente.
MEDIKASJON: Lanoxin mix. 50 mikrogr/ml, 0,7 ml x 2. Acetylsalicyl kapsl. 12,5 mg, 1 kapsl. x 3 pr. uke.

STATUS PRESENS 25.01.00 kl. 1600. Pasienten er 7 måneder gammel. Undersokes på mors fang. Han er våken og begynner å gråte så fort man forsoker å undersoke. Lengde: 64,5 cm. Vekt: 6255 gr. Blodtrykk er ikke målt. Blodtrykk målt okt. -99 var 91/51 ho. arm, 92/60 ve. arm, 105/72 ho. ue., 108/39 ve. ue. Puls: 150 slag/min, regelmessig. Respirasjon: Ubesværet. 48 pr. min. Ingen odemer eller icterus. Lett leppe/tungecyanose. Ingen generell glandelsvulst. Caput og collum: Ingen kliniske tegn til OLI. Thorax: Status etter sternumsplitt. Cor: Regelmessig aksjon, systolisk bilyd grad III med utstråling til rygg. Pulmones: Uten anmerkning. Abdomen: Hepar palperes noe usikkert ca 1 fingerbredde under hoyre costalbue.

Summary:

ALLERGIER: Inge kjente. MEDIKASJON: Lanoxin mix. Ingen odemer eller icterus. Caput og collum: Ingen kliniske tegn til OLI.

## 4.2 Named entity extraction

The other approach used in this thesis is to extract entities that are relevant to the patient. Figure 18 illustrates the architecture of this functionality. There are many features that could be interesting extracting from EPRs, but one of the most describing entities are diagnosis codes. When using this there is no need for machine learning since all the codes are known. The source code is enclosed in Appendix B. The application uses the LingPipe library which provides functionality for exactly these types of applications, see Part II 2.3.5. The different classes and interfaces used in this implementation are described in Table 13. The descriptions are taken from the LingPipe API[32].

| Class | Description |
|---|---|
| **com.aliasi.dict.ExactDictionaryChunker**[33] | An exact dictionary chunker extracts chunks based on exact matches of tokenized dictionary entries. |
| **com.aliasi.dict.MapDictionary**[34] | A MapDictionary uses an underlying map from |

[32] http://alias-i.com/lingpipe/docs/api/index.html
[33] http://alias-i.com/lingpipe/docs/api/com/aliasi/dict/ExactDictionaryChunker.html
[34] http://alias-i.com/lingpipe/docs/api/com/aliasi/dict/MapDictionary.html

| | |
|---|---|
| | phrases to their set of dictionary entries. |
| **com.aliasi.dict.DictionaryEntry**[35] | A DictionaryEntry provides a phrase as a string, an object-based category for the phrase, and a double-valued score. |
| **com.aliasi.chunk.Chunking**[36] | The Chunking interface specifies a set of chunks over a shared underlying character sequence. |
| **com.aliasi.chunk.Chunk**[37] | The Chunk interface specifies a slice of a character sequence, a chunk type and a chunk score. |

Table 13 LingPipe classes

The application creates a MapDictionary with DictionaryEntries specifying the different ICD-Codes. The ExactDictionaryChunker is then created with the already existing MapDictionary using parameters specifying that the chunker is not case sensitive, and not to find incidents where the entities overlap.

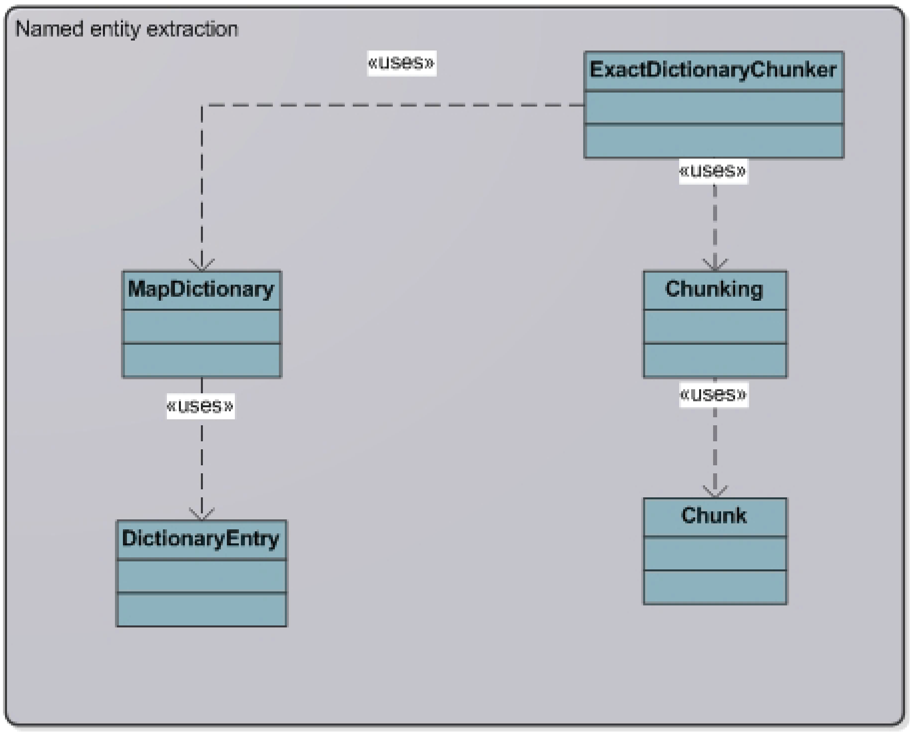


Figure 19 Class overview for the LingPipe implementation

Figure 19 illustrates the classes and how they are connected in the implementation. The ExactDictionaryChunker contains Chunking which again contains Chunks of EPR text. In addition the MapDictionary contains DictionaryEntry which is the dictionary with the ICD codes as entries. The ExactDictionaryChunker uses the dictionary to find chunks in the EPR text that matches the entries.

[35] http://alias-i.com/lingpipe/docs/api/com/aliasi/dict/DictionaryEntry.html
[36] http://alias-i.com/lingpipe/docs/api/com/aliasi/chunk/Chunking.html
[37] http://alias-i.com/lingpipe/docs/api/com/aliasi/chunk/Chunk.html

Looking at the test system it seems that all the EPRs in the prototype already has the ICD codes in a separate column. This column was used instead of extracting entities when implementing this functionality in the prototype, but the source code for extracting entities is attached in Appendix B. The codes where in many cases presented in different formats which lead to some challenges with separating the codes. The system removes all periods and commas that in some cases are used to separate the characters from the numbers in an ICD code. In addition all separators between codes have to be replaced with one global separator. After studying the existing codes it seemed that "/", space, and a combination of these are used as separators.

# 5 External information

This chapter describes the implementation of the module providing the patient with external information about his or her health condition. Because of the importance of presenting correct, high quality information, the patient handbook is chosen as the main external source. The information extraction unit extracts important information from the EPR, for example ICD codes and descriptions. This information gives the most precise description of the medical condition, and will therefore be used in the search for external supplementary information.

The system is implemented with the html parser described in Section 2.1. The description of the icd code is used to search in the patient handbook, and the result is presented in the EPRPortal through an HTML parser. Evaluation of this functionality has to be accomplished through case studies with both physicians and patients. Since this is outside the scope of the thesis some testing with example EPRs is accomplished to check if the patient handbook provides articles to different diagnosis. It is important to determine whether the information is relevant, correct, and gives the patient any valuable information. The work in this thesis is based on the fact that searching on different diagnoses in the patient hand book always will return information of acceptable quality.

Looking at some diagnosis texts from EPRs illustrates that the text is written with abbreviations, and as the rest of the EPRs with typing errors. To avoid this when searching for external information the diagnosis codes are used to get the correct descriptions from the ICD database. This description is then used to search in the patient handbook.

To avoid problems when searching on abbreviations the system removes all words ended with a period before sending the search string to the handbook. In addition only the three most relevant articles for each diagnosis are presented to avoid irrelevant articles.

The architecture is presented in Figure 20 and a screenshot of the system is provided in Figure 14.

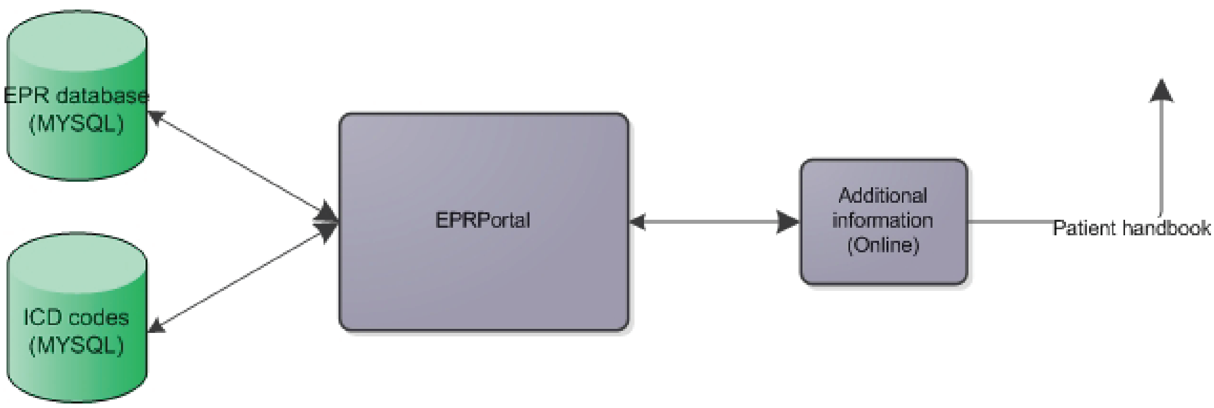


Figure 20 External information

# 6 Summary

The system consists of an EPRPortal presenting the EPR together with the extended functionality, namely summarizing and external information. The portal is developed with HTML and PHP using an XML and HTML parser. The thesaurus server already implemented is used for translating medical terms. The text mining is implemented, mainly as a test application, in Java using the WEKA library.

The text mining implementation and testing is done in java through the WEKA interface. The experimenter in WEKA is used for running the experiments, collect different measures, and compare them. The text mining is tested on the training data with cross validation, and also with some terms taken from an EPR.

# Part IV Discussion and Evaluation

# 1 Part Introduction

This chapter gives a summary of the purpose and scope in this part, and an overview of the different chapters.

## 1.1 Purpose

This part should evaluate, and give considerations around the results which will give a foundation for the conclusion.

## 1.2 Scope

The different implementations will be taken into consideration and the results will be presented with evaluations and discussions. The work that is subject for evaluation is text mining, the external information, and the summarization functionality.

## 1.3 Overview

- Fulfillment of research agenda: Discusses whether the goals of the research questions and agenda have been met.
- Discussion and evaluation of text mining: Evaluates the text mining results, compares the classifiers and datasets, and looks at statistical differences between the results.
- Discussion and evaluation of the extended functionality: Evaluates the functionality through qualitative examples.

## 2 Fulfillment of research agenda

This section will look at the research questions and problem definition, and discuss whether the aspects are fulfilled. Looking at the problem definition it is obvious that some aspects are not addressed in the work. This is mainly the focus on user interface and extension of vocabularies. Since the focus in this thesis was text mining the user interface was given less attention, even though the issues and prototype discussed in [5] are implemented in this thesis. The prototype is presented in Part III 2.2, and the theories behind the refined user interface are taken from the specialization project [5].

Another issue is extension of the vocabularies which was a part of the conclusion in the specialization project. The report concluded that the extension with both Clue and Ordnett would be a good contribution to the system. However, after discussing this extension with RHF it seems that the best solution is to only use the NEL vocabulary together with RHF's own vocabulary. This conclusion is taken on the basis of quality assurance of the medical translations.

The research questions are presented below:

**RQ1** Is it possible to integrate external information sources into the EPR to provide secure, precise, and correct dynamic information to the patient?

**RQ2** Will extension of the information retrieval (IR) process, such as collocation, text mining, and spell suggestion give significant improvements to the system?

RQ1 is fulfilled through the implementation described in Part III 5 where a prototype with the patient handbook as external source is implemented. The study of this functionality shows that the integration provides secure and correct dynamic information. The issue about precision is further discussed in Section 4.2, but it is obvious that there is more work to be done in getting more precise articles.

RQ2 is also fulfilled through the implementation presented in Part III 3. In this section there is a presentation of different text mining approaches for disambiguating between Norwegian words and medical terms. The spell suggestion feature using edit distance proved to already be a part of the thesaurus server. This feature was not documented, but a discussion with the developer of the server unveiled this feature. As mentioned in Part III 2.1 collocation is not implemented because this is feature that also belongs, and could be easily implemented, in the thesaurus server. The extension with text mining gives significant improvements to the system helping to disambiguate medical terms and Norwegian words.

# 3 Discussion and evaluation of text mining

This section presents a discussion and evaluation of the results described in Part III. The results have different classifiers and datasets that give us the possibility to compare them with each other to find the most suitable approach.

The discussion is divided into different parts of evaluation; the first part discusses the cross validation while the second looks at the further testing on EPRs.

## 3.1 Cross validation

The cross validation was executed because of the small amount of test data, see Part II 2.3.3. The test results are based on testing with the training data which could lead to bias in the results. To avoid this problem the test is run 10 fold 10 times. This method is described earlier. In addition it would be reasonable to take the fact that the training and test data are the same into account when evaluating the results.

The cross validation was executed with two different datasets, respectively bigrams and trigrams, see Part II 2.3.2. The following sections will evaluate, discuss, and compare the different results of the cross validation with bigrams, trigrams, and the different classifiers.

### *3.1.1 Bigrams vs. trigrams*

When looking at the results almost all of the tests give better results with bigrams then with trigrams. It has turned out that n-grams with n > 3 in many cases not are optimal, and might in some cases decrease the performance [43]. To compare the two approaches in this thesis, taking the issue about training data being used to testing, the paired corrected t-test is used. This test is described further in Part II 2.3.3.

During this test the alpha-level is set to 0.005 which is the weakest evidence normally excepted in the experimental sciences [11]. The degree of freedom is set to the number of validations run (k), minus one. In this case the tests are run 10 fold cross validation 10 times, in other words k=100 and the degree of freedom is set to 99.

One interesting measure is the accuracy which describes the success rate of the classifier. The values are taken from the results in Part III 3.4.1. The formula below calculates the t-value for difference between the means for bigrams and trigrams with the classifier complement bayes.

$$t = \frac{85{,}62 - 80{,}12}{\sqrt{\left(\left(\frac{1}{100} + \frac{0{,}1}{0{,}9}\right) 0{,}33088163\right)}} = 25{,}39809358$$

Looking at the critical t value for the chosen significance level 2,8713, and since 25,4 is larger than 2,8713 we reject the null hypothesis, which leads to the conclusion that bigrams are significantly better in this test.

The calculation below describes a corrected paired t-test for support vector machines. This test gives a result below 2,8713, which leads to the conclusion that bigrams does not perform significantly better or worse in this test.

$$t = \frac{93{,}43309312 - 93{,}23252694}{\sqrt{\left(\left(\frac{1}{100} + \frac{0{,}1}{0{,}9}\right) 0{,}092242612\right)}} = 1.409464014$$

| | Bigrams | Trigrams |
|---|---|---|
| | **Naïve Bayes** | *w=worse b=better* |
| **Accuracy** | 0,50 | 0,35b |
| **Precision** | 0,01 | 0,0b |
| **Recall** | 0,01 | 0,01 |
| | **Complement Bayes** | |
| **Accuracy** | 0,50 | 0,55w |
| **Precision** | 0,01 | 0,01w |
| **Recall** | 0,01 | 0,01b |

Table 14 Standard deviations and t-test results comparing datasets

| | Bigrams | Trigrams |
|---|---|---|
| | **Complement Bayes** | *w=worse b=better* |
| **Accuracy** | 0,50 | 0,55w |
| **Precision** | 0,01 | 0,01w |
| **Recall** | 0,01 | 0,01b |
| | **SVM** | |
| **Accuracy** | 0,35 | 0,35 |
| **Precision** | 0,0 | 0,0b |
| **Recall** | 0,01 | 0,01w |

Table 15 Standard deviations and t-test results comparing datasets

The same calculation for naïve bayes shows that it has better performance on trigrams then on bigrams. As it seems trigrams and bigrams performs quite similar, this is also in accordance with the results from a study using n-gram features for text categorization [43]. Looking at precision and recall, the results vary a little. Complement bayes has significant worse precision and better recall with trigrams, while support vector machines has significant worse recall, and better precision. Naïve bayes gives better precision, and the same recall with trigrams. Table 14 and Table 15 show all the standard deviations and the results of the paired corrected t-test. A significant worse result is marked by the character “w” while better results are marked by “b”.

The differences in the results between bigrams and trigrams are definitely largest with complement bayes as classifier. The training and testing times of the different datasets are measures that have to be taken into account. The times are presented in Table 16 and Table 17 and show that both testing and training time increase with trigrams. Because of this issue the trigrams dataset should perform significantly higher to be worth the increased computing times. Looking at the classification time with SVM bigrams it seems that each instance will demand 33,49/4478 = 0,0075 seconds classification time. If the EPR text contains 50 unclassified words the system would use 0,37 seconds to assign categories to them. This should not cause any problems when using this with real EPRs.

| | Naïve Bayes | Complement Bayes | SVM |
|---|---|---|---|
| **Only bigrams** | 182,01 | 0,16 | 471,94 |
| **Bi/trigrams** | 371,51 | 0,20 | 637,92 |

Table 16 Training times for 40295 instances

| | Naïve Bayes | Complement Bayes | SVM |
|---|---|---|---|
| **Only bigrams** | 17,37 | 0,05 | 33,49 |
| **Bi/trigrams** | 32,27 | 0,03 | 43,21 |

Table 17 Testing times for 4478 instances

The discussion of the different datasets will continue in Section 3.1.3 where the costs of the errors are taken in consideration.

### *3.1.2 Different classifiers*

In this section only naïve bayes, complement bayes, and SVM will be subject for discussion. The J48 classifier is discussed in Section 0 because this is the only test where results are available from all classifiers.

Comparing the results from the classifiers with the t-test gives pretty clear indications that SVM gives the best accuracy. SVM gives statistical better accuracy then both naïve and complement bayes, with both bigrams and trigrams. The precision and recall measures are both statically significant better with SVM than all the other except for recall where complement bayes performs better. A complete overview of the results can be found in Part III 3.4.1.

Table 18 and Table 19 presents all the t-test result where significant worse results are marked by “w” while better results are followed by ”b”.

| | Naïve bayes | Complement bayes |
|---|---|---|
| | **Bigrams** | *w=worse b=better* |
| Accuracy | 0,69 | 0,50b |
| Precision | 0,01 | 0,01b |
| Recall | 0,01 | 0,01 |
| | **Trigrams** | |
| Accuracy | 0,60 | 0,55 |
| Precision | 0,01 | 0,01 |
| Recall | 0,01 | 0,01b |

Table 18 Standard deviations and t-test results comparing classifiers with Naïve Bayes as baseline

| | Complement Bayes | SVM |
|---|---|---|
| | **Bigrams** | *w=worse b=better* |
| Accuracy | 0,50 | 0,35b |
| Precision | 0,01 | 0,0b |
| Recall | 0,01 | 0,01b |
| | **Trigrams** | |
| Accuracy | 0,55 | 0,35b |
| Precision | 0,01 | 0,0b |
| Recall | 0,01 | 0,01w |

Table 19 Standard deviations and t-test results comparing classifiers with Complement Bayes as baseline

Table 16 and Table 17 shows that the SVM classifier requires more computing time compared to other alternatives. As stated in 3.1.1 this should not be significant when classifying a journal with about 30-40 unknown words.

### *3.1.3 Cost analysis*

So far the discussion has focused on the results without taking consideration to the costs connected with wrong classifications. As mentioned earlier the most serious error is the one of classifying a medical word as fiction literature or ordinary Norwegian. When looking at accuracy as an evaluation measure there is an issue if the data is skewed. If an classifier scores 99,9% on a test sample that consists of 999 positive instances, and 1 negative instance. The classifier did not classify the negative instance correct, but got a high accuracy. If accuracy is used as the only evaluation comparing classifiers it could lead to invalid conclusions. To address these issues a receiver operating characteristic (ROC) plot is used [27], see Part II 2.3.3.

The tests executed earlier set medical terms as positive, while Norwegian literature as negative. To deal with the issue that classifying medical terms as Norwegian has higher cost the graphs presented here will use true negative (TN), and false negative (FN) rate.

$$TN = \frac{TN}{FP + TN}$$

$$FN = \frac{FN}{TP + FN}$$

The ROC plot is presented in Figure 21, and shows the different classifiers with the TN and FN rate. The values used in this plot are presented in Table 20.

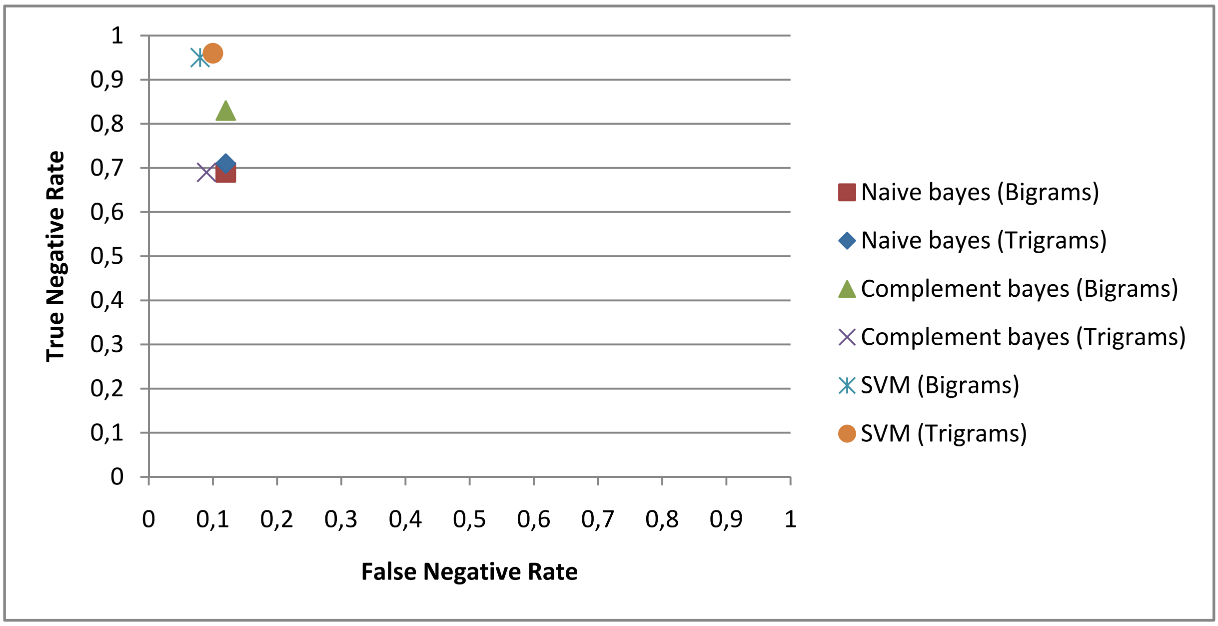


Figure 21 ROC plot

| Classifier | True Negative Rate | False Negative Rate |
|---|---|---|
| **Naïve Bayes (Bigrams)** | 0,69 | 0,12 |
| **Naïve Bayes (Trigrams** | 0,71 | 0,12 |
| **Complement Bayes (Bigrams)** | 0,83 | 0,12 |
| **Complement Bayes (Trigrams)** | 0,69 | 0,09 |
| **SVM (Bigrams)** | 0,95 | 0,08 |
| **SVM (Trigrams)** | 0,96 | 0,1 |

Table 20 Rates for the classifiers presented in the ROC plot

Figure 22 illustrates the convex hull in the ROC plot, and shows that all classifiers except SVM are suboptimal because they do not lie on the convex hull. The two points on the hull are respectively SVM with trigrams and bigrams, and therefore the optimal classifiers.

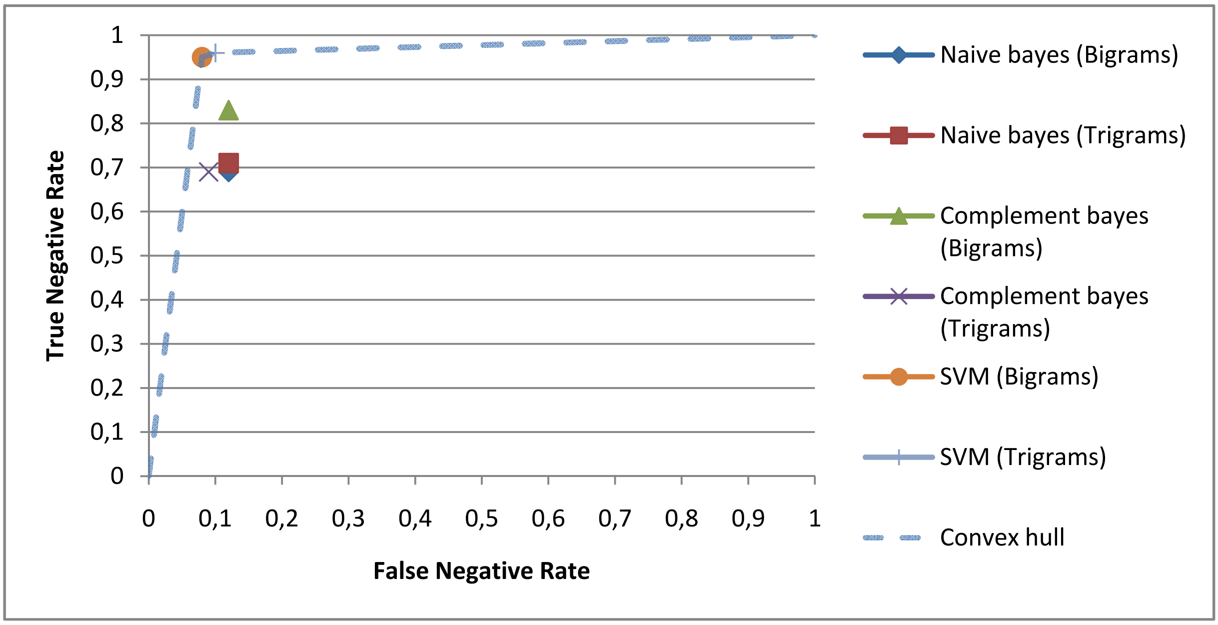


Figure 22 ROC plot with convex hull

The SVM classifier scores higher on TN rate with trigrams than with bigrams, but the FN rate is lower with bigrams. Taking into account the fact that there is an additional cost with classifying a positive word as negative it seems that SVM with bigrams gives the best result. In addition the training and testing times on trigrams are higher than bigrams which also strengthens the position of bigrams.

### 3.2 Testing with EPRs

The test with real EPRs should give a more valid result because of separate training and testing data [15]. The test set had to be classified manually, and therefore the size is limited. Looking at the accuracy in Part III 3.4.2 it is clear that this test has poorer results than the cross validation. This is natural since the cross validation uses the training data for testing.

Since we only have one result for each dataset and classifier a t-test does not make sense. There are not enough results to get statistically significant differences. Looking at the results it seems like both SVM and J48 with bigrams gives the best results. Since there are costs connected with wrong classifications a ROC plot seems to be one of the best ways to compare the classifiers.

Figure 23 and Table 21 shows the different classifiers and datasets with results. The best results are achieved by the SVM and J48 classifiers. When comparing J48 and SVM it is important to take into account the long training times with J48 described in Part III 3.3.1. Figure 24 illustrates the convex hull, and from this graph it is possible to conclude that only Complement Bayes (Trigrams), SVM (Bigrams), J48 (Bigrams), and SVM (Trigrams) are optimal classifiers.

The cost of misclassifications in this class (Norwegian literature) is high, and therefore it would be preferable achieving as low FN rate as possible. If this is set as the main criteria Complement Bayes with trigrams would be the best classifier in this test. But there are other issues that are important taking into consideration, namely the data balance, the statistical significance of this test, and the training and testing times [15, 27].

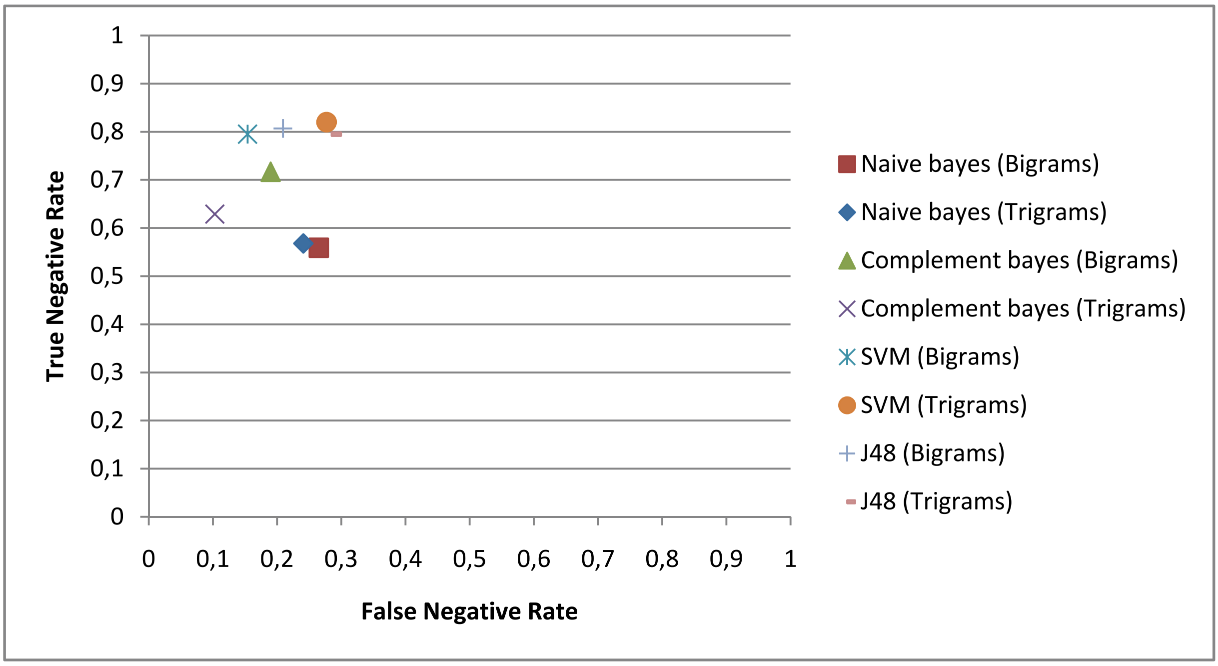


Figure 23 ROC plot

| Classifier | True Negative Rate | False Negative Rate |
|---|---|---|
| **Naïve Bayes (Bigrams)** | 0,559 | 0,265 |
| **Naïve Bayes (Trigrams** | 0,568 | 0,241 |
| **Complement Bayes (Bigrams)** | 0,717 | 0,19 |
| **Complement Bayes (Trigrams)** | 0,629 | 0,103 |
| **SVM (Bigrams)** | 0,795 | 0,154 |
| **SVM (Trigrams)** | 0,82 | 0,277 |
| **J48 (Bigrams)** | 0,807 | 0,209 |
| **J48 (Trigrams)** | 0,795 | 0,285 |

Table 21 Rates for the classifiers presented in the ROC plot

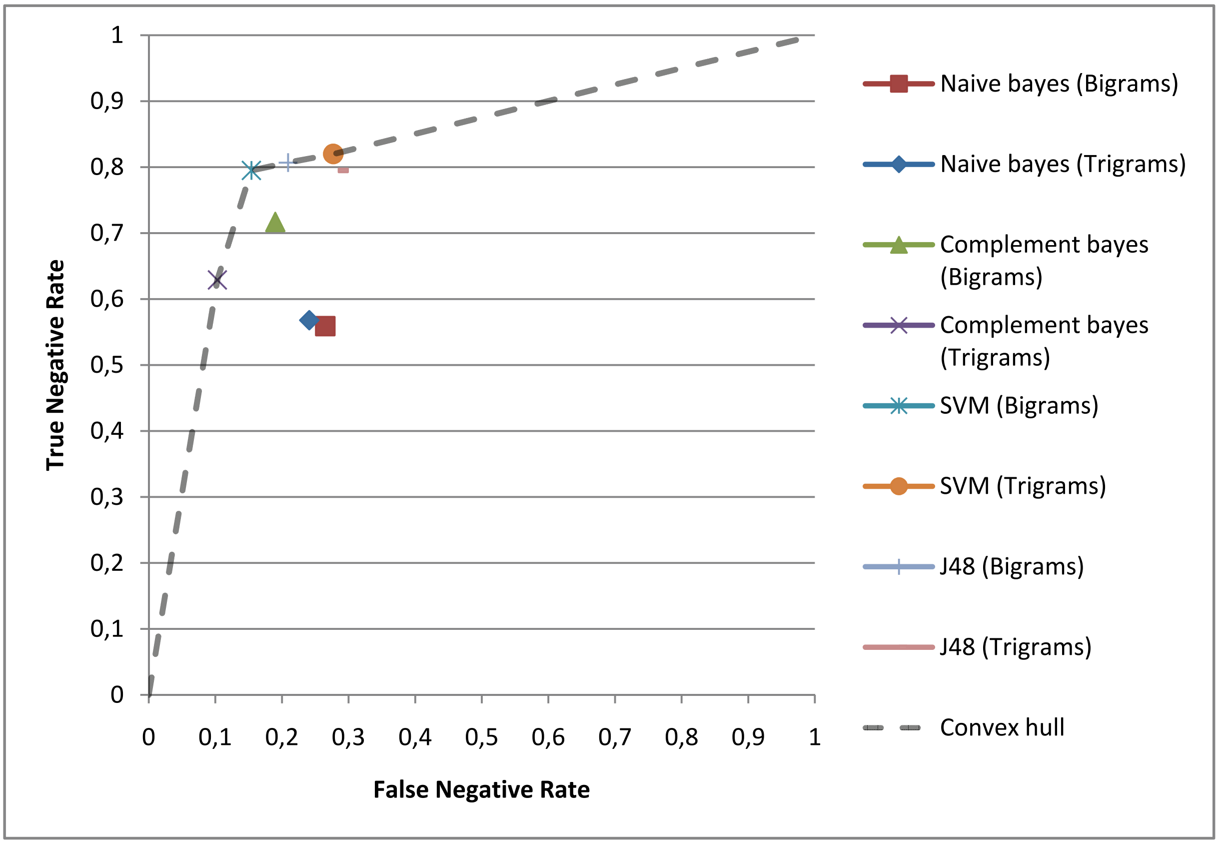


Figure 24 ROC plot with convex hull

Table 21 shows the actual rates for the classifiers and it is possible to conclude that SVM with bigrams gives a relatively good gain of TN rate without increasing FN rate too much. The other classifiers and SVM with trigrams give minimal improvements in the TN rate while increasing the FN rate to an unacceptable level. The J48 algorithm also has higher training and classifying times than all the others, which is an issue to take into consideration.

When using the classifier on real data it is reasonable to believe that Norwegian words are more frequent than medical terms. This fact strengthens the choice of the SVM classifier. In addition SVM with bigrams has higher accuracy and kappa-statistic than the other alternatives. All the results are presented in Part III 3.4.2. Precision in class II was emphasized earlier as an important measurement, mainly because of the misclassification costs in this class. The ROC plots presented above describes the same issue, and therefore precision in class II is not further discussed here. Looking at precision in Class I (medical terms) the results give SVM with bigrams as the best classifier.

A complete overview of the measures from the test is attached in Appendix E.

## 3.3 Combining the results

The tests presented in Section 3.1 and 3.2 both have their weaknesses. The cross validation experiment has several reruns, and is therefore statistical valid. The testing with real EPRs is more related to qualitative study because of the low number of test instances [9]. The cross validation test main weakness is because of using the same data for both training and testing. The corrected t-test takes the bias into consideration and to some extent avoids this problem.

The results from the different tests vary to some extent, but they are somewhat compatible. The cross validation results more or less conclude with SVM using bigrams as the best solution. The test with EPR data gives the alternatives SVM with both bigrams and trigrams, J48 with bigrams, and complement bayes with trigrams. If the other measures like accuracy, kappa statistic, class I precision, and recall from the EPR test are taken into consideration it seems that also this test concludes with SVM and bigrams as the preferred classifier. Complement Bayes with trigrams gives the best results when it comes to Class II precision in the EPR test. This result is caused by the low FN rate in this test. When looking at the more statistical valid cross validation the difference between SVM and complement bayes are smaller. In addition this classifier has a significant lower TN rate than SVM which leads to the conclusion that SVM is a better choice than complement bayes.

Both tests result in high computing time with the J48 decision tree, and together with the results it seems like decision trees do not provide an optimal solution for this application. Trigrams also result in long training and test time compared with bigrams. As stated earlier the results with trigrams should be significant better compared to bigrams if they should be worth the extended computing time.

Naïve bayes seems to be the classifier with the lowest performance in this test. This can be caused by the fact that this classifier assumes that the features are independent within a class which is not the case in this experiment. In addition this algorithm together with instance based classification has a issue with producing generalizations of data [44]. In many ways it is not surprising that the SVM classifier is the one with the best outcome. The algorithm has restrictive learning bias, which leads to the fact that it can handle high dimensionality [20, 45]. But it is also an issue if the dimensionality gets too high which leads to a lot of irrelevant features. This might be the case if trigrams are used, and especially using n-grams with $N<3$ [45]. In text categorization it is important to handle dense concepts, combine many different features, because mostly all the features in text categorization are relevant. But for each document mostly all of the entries in the document vector are zero except for a few. These kinds of issues are well suited for SVM classifiers [20].

# 4 Discussion and evaluation of the extended functionality

Evaluating the extended functionality of the patient portal is a different kind of issue than discussing the results in Section 2. In this case there are no measurements and experiment to compare and evaluate. The portal has to be evaluated with qualitative data that for example could be interviews with test persons. Because this is outside the scope of this thesis, and the fact that it was not possible to execute a study with patients, the evaluation is achieved through examples of usage and qualitative data taken from these examples [8, 10].

## 4.1 Summary functionality

The two alternatives of summary functionality are compared through examples of usage, and the evaluation is done without any medical expert knowledge. The first implementation, namely sentence extraction, is tested on different EPRs with varying results. Looking at EPRs it seems that they do not contain many words that are repeated several times. In some cases if there are repeated words these are stop words which are not relevant to the context of the EPR.

One example of this kind of summary is presented in the textbox in Part III 4.1. Below, another EPR and summary like this is presented:

NATURLIGE FUNKSJONER: Uten anmerkning.
ALLERGIER: Ingen kjente.
STIMULANTIA: Ingen.
MEDIKAMENTER: Selo-zok 50 mg x 2. Triatec 5 mg x 1. Digitoxin 0.05 mg hver annen dag.
Furosemid 40 mg 2+1/2+1+1/2. Spirix 25 mg x 1/2.

STATUS PRESENS 25.01.00: Blodtrykk: 100/60. Puls 84: Uregelmessig. Caput: Uten anmerkning.
Collum: Aa. carotider +/+. Thorax: Cor: 1. og 2. hjertelyd, systolisk bilyd grad III/VI p.m. v.4.
intercostalrom. Pulm: Uten anmerkning. Abdomen: Palpabel puls til tumor i epigastriet.
Underekstremiteter: Alle arterier positive. Ikke odemer.

NATURLIGE FUNKSJONER: Uten anmerkning. ALLERGIER: Ingen kjente. MEDIKAMENTER: Selo-zok 50 mg x 2. Triatec 5 mg x 1.

Examples like this illustrate the problems with this kind of summary. As it seems there are only repeating stop words like for example the word “uten” (without). This results in the first sentence in the summary consisting of a comment with the sentence “without marks”. It seems that these summaries do not give a good extraction of the information in the EPR. Another observation is that the first part of an EPR does not summarize the rest of the text as an introduction or abstract. This leads to the fact that it might not be reasonable to give the first part of the EPR higher weight which is a premise for using this method.

The other approach is to extract the diagnosis codes, and then get the descriptions of the different codes. The codes give a good overview of the case history that could be of high value to nurses, physicians, and patients. Discussions with Nurse Karl Øyri at the Interventional centre have shown that this is an interesting feature for both medical personnel and laymen. Figure 25 gives an example of the summary function. When using this type of summary you do not remove or add information to the EPR, only use the codes issued by the physician. Taking into account that the patient should not get a wrong comprehension of his or her health condition, the extraction of diagnosis codes seems like a better option than sentence extraction.

## 4.2 External information

The examples of presenting external information are evaluated through the relevance of the presented articles. Whether an article is relevant or not is based on whether the article has relevance to some of the words in the EPR, which means that the relevance not is evaluated by any medical expert.

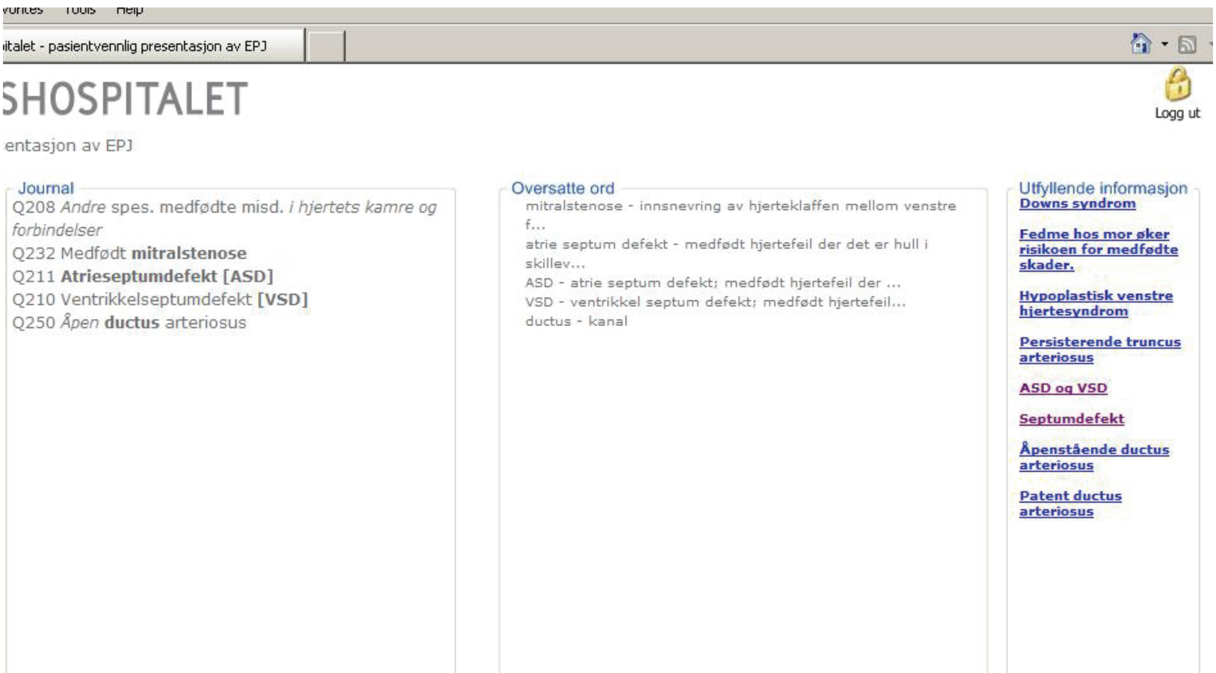


Figure 25 Example of summary with external information

Figure 25 shows an example of a summary with the external information presented in the right area of the window. The summary presents the text that the patient handbook gets as search strings.

| Q208 | Q232 | Q211 | Q210 | Q250 |
|---|---|---|---|---|
| Irrelevant | Irrelevant | Relevant | Relevant | Relevant |
| Irrelevant | Irrelevant | Relevant | Relevant | Relevant |

Table 22 Relevance of articles

Table 22 presents the articles from Figure 25 and describes whether or not they are relevant, based on a layman's evaluation. It seems that words without translation, ordinary Norwegian words, confuse the search engine. Since the search engine uses the or operator on all the words some of the articles are based on only one of the search words, for example "information", which then again results in an article about client confidentiality. The word "information" was only an irrelevant word in the diagnosis text about a heart condition. One solution could be to remove all words without translation or at least the words defined as stop words in the thesaurus server. If all stop words are removed from the search string, a restriction of the search results is achieved but valuable articles could be overlooked with this method. The word "cardiac infarction" is defined as a stop word, but removing this from the search would result in loosing relevant articles about this issue.

Another, and as it seems better, solution is to define a separate list of words that should be removed from the search string. Looking at ICD-codes it seems that the same irrelevant words are used over and over again. Examples of such words are presented in Table 23.

| Irrelevant search words |
|---|
| **Unspecified** |
| **Specified** |
| **Congenital** |
| **Other** |

Table 23 Examples of stop words

The examples in Table 23 are a small amount of the actual words. In addition normal stop words like "in", "on", "with", "without", and "and" are relevant in this case. A normal Norwegian stop word list provided with the snowball stemmer[38] could be used as a basis for the list used in this functionality. But as discussed above there are a lot of ICD code specific, irrelevant words, which has to be added to the list.

An example from Figure 25 illustrates the results of removing stop words. When the diagnosis "Medfødt mitralstenose" (Congenital mitral stenosis) is used in searching it gives two irrelevant hits about other issues, while when searching on "mitralstenose" (mitral stenosis) it gives the correct and relevant article presented in Figure 26. It also seems that the search engine in the patient handbook should be refined since the relevant article is not presented in the ten most relevant articles when using the search string without removing "congenital".

[38] http://snowball.tartarus.org/algorithms/norwegian/stop.txt

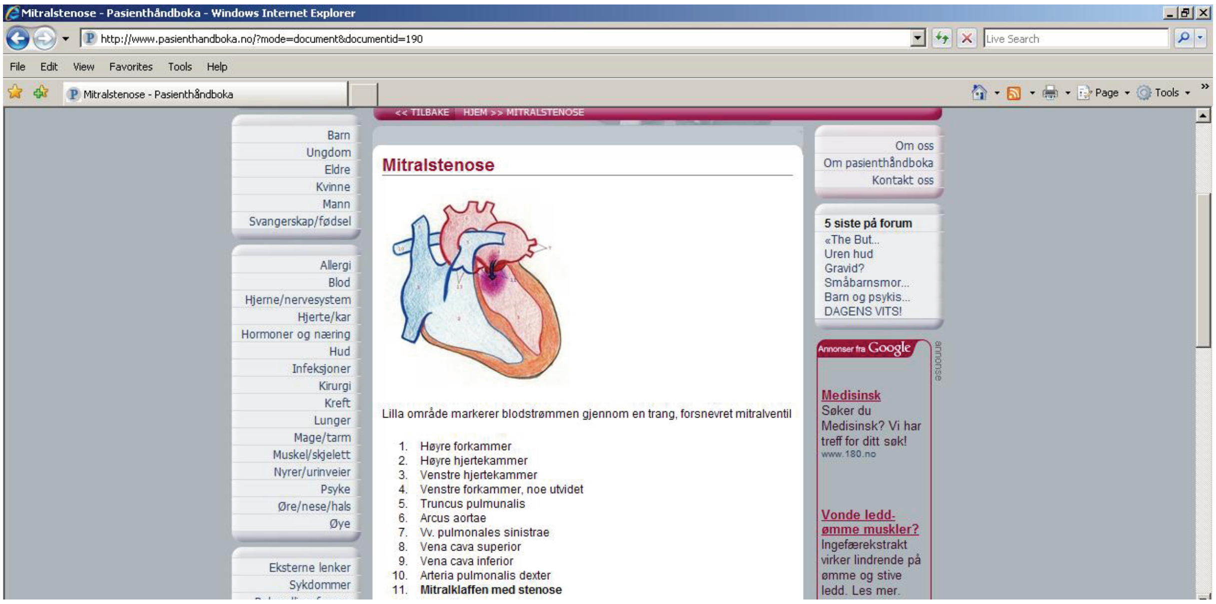

**Figure 26 Article from the patient handbook**

This evaluation gives clear indications that this functionality has to be tested and refined further. One article presents the possibility of using for example Google to present relevant information [38]. It also presents some criteria that could be achieved by providing patient friendly information. In this implementation there are achieved higher accessibility, confidentiality that the physician does not provide, and selectivity of information while keeping the safety medical personnel provides. One issue that could be a challenge in this prototype is as discussed above, the relevance of the articles. The feature of getting extended knowledge achieved when using Google as the search source is not that conspicuous in this implementation, but this is sacrificed to keep safety without an extensive classification of web sites.

Other articles have looked at the consequences of presenting the EPR to patient, and they conclude that the system must strive to present correct, precise and not frightening information [46, 47]. This could be the issue if, for example information about serious health conditions not relevant to the patient is presented in the EPR system. As mentioned in Part II 3.3, using Google has several weaknesses when it comes to quality assurance of the articles, but it seems that Google is a better search engine than the patient handbook. Since the quality assurance issue is important the patient handbook is a good choice for this prototype implementation.

# Part V Conclusion and Further Work

# 1 Part Introduction

This chapter gives a summary of the purpose and scope in this part, and an overview of the different chapters.

## 1.1 Purpose

This part should conclude the results of this thesis, what has been achieved through this study, and state further work.

## 1.2 Scope

Conclusions of the different evaluations, present the results of this thesis, and give elaborations about unsolved challenges and further work.

## 1.3 Overview

- Conclusion: Concludes the thesis.
- Further work: Presents unsolved issues, and challenges that need further work.

## 2 Conclusion

The development of a portal for patient friendly presentation of EPRs is a challenging subject area which demands several algorithms and techniques. Complex usage of different medical terms, shortenings, and a lot typing errors are among other things obstacles that has challenged the development of a patient friendly EPR system.

The challenges addressed in this thesis, and other issues, can be solved using text mining. Looking at the problem of disambiguating Norwegian words from medical terms the SVM classifier using bigrams proved to give acceptable results. This classifier has reasonable classification times that are manageable when it comes to using it in EPR translation. Assuming an accuracy of above 80% it seems that it would give significant contributions to the system. The validation of these results, achieved through statistical tests and cost analysis, support this conclusion. When looking at the issue of misclassifying medical terms as Norwegian words, which is one of the most serious errors, it seems that the SVM classifier seems to overcome this issue with reasonable error rates.

Presenting the EPR with relevant articles about the patient's health condition has shown to be interesting and valuable to the patient. The approach in this thesis has some good qualities and properties, namely quality assurance and safety of presenting articles with correct information. The articles of the patient handbook are a part of NEL which is assured by medical experts like physicians. What seems to be an issue with this implementation is the precision which results in sometime presenting articles not relevant to the patient.

Looking at the summary functionality it seems clear that this kind of information is valuable to both patients, and medical personnel. Comparing the two methods used in this thesis, the conclusion is that the extraction of diagnosis is the safest and best approach. Most of the repeated words in an EPR are not relevant stop words which lead to the conclusion that the sentences with the most frequent word do not summarize the content. The summary gives the patient, and possibly personnel not familiar to the patient a short but informative overview of the patients' health condition. This could also be an important feature helping the patient input information into PHRs which is highly relevant nowadays.

All the new functionality, namely summary and external information, presented in this thesis has been evaluated and validated through examples and qualitative considerations. Through this approach some of the weaknesses has been discovered and discussed, but to get an objective and extensive evaluation of these a case study with patients should be executed. A case study with patients was one of the planned activities in the work with this thesis. But during the study it has shown to be impossible because of hospital rules and the time required getting approval for a case study like this.

The prototype developed, and the results presented in this thesis, gives valuable contributions the work already done in this project. The refined user interface, primary studied in the specialization project and implemented in this thesis, together with the new and extended functionality gives a better starting point for executing a case study with patients. Looking at relevant articles it seems that providing the patient with information about their health condition could improve the communication with their physician, and lead to the patient taking a more active role in his or her health care.

# 3 Further work

The development of an electronic patient friendly presentation of EPRs is challenging and there are more issues to address in the future. Some of them have been discussed in this thesis, and other issues are new features and possible improvements.

In this thesis text mining has been used to address the issue of word sense disambiguation between Norwegian words and medical terms through character n-grams. Another issue is to use text mining to disambiguate words' senses through the context of the word. This issue is about words that can have both medical and ordinary Norwegian sense. An example of such a word is “mors” which can mean both death and mom. Using text mining and the context of the words is a way to address this challenge. In order to do this medical personnel must tag a significant amount of EPR text that can be used as training data. This seems to not be available at this time, and this was the main reason that the issue was not worked with in this thesis. Collocated terms are another improvement mentioned in this thesis. A lot of medical terms are collocated, and the system as it exists do not support these terms, and therefore translates them as two separate words. This functionality should be implemented in the thesaurus server and is considered to be a fairly simple improvement that could give significant improvements.

The work with providing external and extended information to the patient about his or her health condition was started in this thesis but there are several issues that have to be studied before getting a complete system. The existing search system has to be refined with a good stop word list based on words occurring in ICD descriptions. To give the articles even higher precision text mining could be used to learn the system which articles are relevant, and which is not. The system could also benefit from testing with other sources of information, for example using Google as an external source. Implementing Google presuppose that the quality assurance issues with this source is addressed.

The user interface and functionality has to be tested in an extensive study with patients. This is important to get a good basis for evaluating and improving the functionality. The evaluation in this thesis is based on some qualitative examples, but in order to get a better evaluation both quantitative and qualitative studies with real users has to be executed.

# Part VI Appendix

# A. Source code from the patient portal

**Index.php**

```
<html>
<?php
  if (!defined("CACHE"))
    define("CACHE", FALSE);

if (!defined("GLOBAL_CACHE"))
    define("GLOBAL_CACHE", false);

if (!defined("TTL"))
    define("TTL", 0);

if (!defined("PAGE_CACHE"))
    define("PAGE_CACHE", false);
header("Cache-Control: no-cache, must-revalidate"); // HTTP/1.1
header("Expires: Mon, 26 Jul 1997 05:00:00 GMT"); // Date in the past
header('Content-Type: text/html; charset=utf-8');
include_once('thesaurus.php');
include_once('epr.php');
include_once('html_dom_parser.php');

$eprno = $_GET['epr'];
$mode = $_GET['mode'];

$thesaurus = new Thesaurus();
$epr = new Epr();
$result="";
if($mode==0){
$result = $epr -> getEPR($eprno);}
$icd = $epr -> getICD($eprno);

$temp = array();
$i=0;
foreach($icd as $code){
		$codedesc = $epr->getICDDesc($code);
		if($mode == 1){
				$result = $result . $code . " " . $codedesc . "\n";
		}
		$codedesc = utf8_decode($codedesc);
		$codedesc = str_ireplace(array("[", "]", "uspesifisert ", "andre ", "og ", "spesifiserte ", "i ", "på ",
"medfØdt ", "medfødte ", "uten ", "opplysning ", "om "), "", $codedesc);
		$codedesc = utf8_encode($codedesc);
		$codedesc = preg_replace("/[A-Åa-å]*\./", "", $codedesc);
		$codedesc = trim($codedesc);
		$codedesc = str_replace(" ", "+", $codedesc);
		$dom = file_get_dom('http://www.pasienthandboka.no/default.asp?searchstring=' . $codedesc .
'&mode=search');

		foreach($dom->find('td#searchresult') as $node){
		   foreach($node->find('a') as $link){
				$temp[$i][$link->innertext][0] = $link->href;
						$temp[$i][$link->innertext][1] = $link->innertext;
```

```
            }
        }
    $i++;
}
$epr->freeEPR();




$terms = $thesaurus -> translate($result);
?>
<head><title>Rikshospitalet - pasientvennlig presentasjon av EPJ</title>
<link type="text/css" href="epj.css" rel="stylesheet">
</head>

<body>

<div id="container">
<div id="logo">
<img src="pictures/logo.gif" align="left">
<img src="pictures/logg.jpg" align="right"><br><br><br>
<p>Pasientvennlig presentasjon av EPJ</p>
</div>
<div id="menu">
<br>

<a href="?epr=8210&mode=1"><img src="pictures/summ.jpg"></a><br>
<a href="?epr=8210&mode=0"><img src="pictures/journal.jpg"></a><br>
</div>

<div id="epj">
<?php

$newEPR= explode("\n", $result);
$i = 0;
foreach($newEPR as $line){
   $newWords = preg_split("/[\s]+/", $line);
   foreach($newWords as $word){
   $stopword = false;
   $printed = false;
   $trans = false;
        for($k=0;$k<(count($terms[$i])-1);$k++){
                if(($terms[$i][$k]['thesaurus'] == 2)&&(strcmp($word,
"\r")!=0)&&strlen($word)!=0&&$terms[$i][$k]['explanation']==""){
                        $stopword = true;
                }
                else if((count($terms[$i][$k]) == 4)&&(strcmp($word, "\r")!=0)&&strlen($word)!=0){
                        if(!$stopword){
                        print ("<b><span title = '" . $terms[$i][$k]['term'] . " - " . $terms[$i][$k]['explanation']
. "' onmouseover=\"". strtolower(str_replace(array("-", " ", ".", ":", ","),"",$terms[$i]['original'])) .
"1.style.fontWeight = 'bold';\" onmouseout=\"". strtolower(str_replace(array("-", " ",".",":",
","),"",$terms[$i]['original'])) . "1.style.fontWeight = 'normal';\">" . $word . "</span></b> ");}
                        else{
                        print ("<b><span title = '" . $terms[$i]['original'] . " - norsk ord' onmouseover=\"".
strtolower(str_replace(array("-", " ", ".", ":", ","),"",$terms[$i]['original'])) . "1.style.fontWeight = 'bold';\"
onmouseout=\"". strtolower(str_replace(array("-", " ",".",":", ","),"",$terms[$i]['original'])) . "1.style.fontWeight
= 'normal';\">" . $word . "</span></b> ");}
```

```
                    $k=count($terms[$i]);
                    $printed = true;
                    $trans = true;
               }

          }
          if(!$printed){
                    if($stopword){
                         print("<i>" . $word . "</i> ");
                    }
                    else{
                         print($word . " ");}

               }
          if(strcmp($word, "\r")!=0&&strlen($word)!=0){
               $i++;
          }




  }
  print ("<br>");
}

?>
</div>
<div id="desc">
<?php
$print_array = array();
foreach($terms as $term){
          $stopword = "";
          foreach($term as $result){
               if(count($result)==4){
                    if ($print_array[strtoupper($term["original"])] != 1){
                    if($stopword==""){
                    print("<span title = '". $result["explanation"] ."' id='". strtolower(str_replace(array("-
", " ", ".",":",","),"",$term["original"])) . "1'>" . $result["term"]. " - " . substr($result["explanation"],0,45) .
"</span>");
                    if(strlen($result["explanation"])>45){print("...");}}
                    else{
print("<span title = '". $result["explanation"] ."' id='". strtolower(str_replace(array("-", " ",
".",":",","),"",$term["original"])) . "1'>" . $stopword. " - norsk ord</span>");
}
                    $desc = "";

                    if(count($term)>2){
                         foreach($term as $results){

                              if($results["thesaurus"] == 2 && $results["explanation"] == ""){
                              $desc = $desc . $results["term"] . " - norsk ord \n";}
                              else if ($results["thesaurus"] == 1){
                              $desc = $desc . $results["term"] . " - " . $results["explanation"] .
"\n";}
                              else if ($results["thesaurus"] == 2){
```

```
                                        $desc = $desc . $results["term"] . " - " . $results["explanation"] .
"\n";}
                                        else if ($results["thesaurus"] == 3){
                                        $desc = $desc . $results["term"] . " - " . $results["explanation"] .
"\n";}
                                }
                        print("<span title = '". $desc . "'><b><font color = 'red'> (Mer)</font></b></span>");
                        }
                        print ("<br>");
                        $print_array[strtoupper($term["original"])] = 1;
                        }

                }

                else if($result["thesaurus"] == 2 && $result["explanation"] == ""){
                        $stopword=$result["term"];

                }

        }

}

?>
</div>
<div id="info">
<?php
$print_array_link = array();
$count=0;
foreach($temp as $add){
foreach($add as $linking){
        if($count < 2){
        if($print_array_link[$linking[1]] != 1){
        print '<a href="http://www.pasienthandboka.no/' . $linking[0] . '" target="window">' . $linking[1] .
'</a><br><br>';
        $print_array_link[$linking[1]] = 1;}
        $count++;
        }
}

$count=0;
}
?>
</div>
<div id="footer">
<p class="footer">Systemet er utviklet for <a href="http://www.rikshospitalet.no/">Rikshospitalet HF</a> av
<a href="mailto:kjetil@stallemo.com">Kjetil Stallemo</a> i forbindelse med masteroppgave ved <a
href="http://www.ntnu.no/">NTNU</a> 20&copy;08</p>
</div>
</div>
</body>
</html>
```

**Epr.php**

```
<?php
Class Epr{
	private $sqlconnect;
	private $sqldb;
	private $result;

	function __construct(){
	$this->sqlconnect = mysql_connect('localhost', 'test', 'bb3176');
	$this->sqldb = mysql_select_db('datacor', $this->sqlconnect) or die("Unable to select database");
	mysql_query("SET NAMES 'utf8'");
	}

	function getICDDesc($icdcode){
			$this->result = mysql_query('select beskrivelse from icd where kode="' . $icdcode .
'";');
			while ($line = mysql_fetch_array($this->result, MYSQL_ASSOC)) {
			foreach ($line as $col_value) {
			return $col_value;
			}
			}

	}

	function getICD($eprno){
	$this->result = mysql_query('select diag_nr from record where id="' . $eprno . '";');
		while ($line = mysql_fetch_array($this->result, MYSQL_ASSOC)) {
		foreach ($line as $col_value) {
			$pass = preg_replace('/\.([0-9]+)/', '$1', $col_value);
			$pass = preg_replace('/\,([0-9]+)/', '$1', $pass);
			$pass = preg_replace('/ \/ ([A-Za-z]+)/', ';$1', $pass);
			$pass = preg_replace('/\/ ([A-Za-z]+)/', ';$1', $pass);
			$pass = preg_replace('/ \/([A-Za-z]+)/', ';$1', $pass);
			$pass = preg_replace('/\/([A-Za-z]+)/', ';$1', $pass);
			$pass = preg_replace('/ ([A-Za-z]+)/', ';', $pass);
			$pass = explode(';', $pass);
	return $pass;
		}
		}
	}

	function getEPR($eprno){
	$this->result = mysql_query('select innk_nota from record where id="' . $eprno . '";');
		while ($line = mysql_fetch_array($this->result, MYSQL_ASSOC)) {
		foreach ($line as $col_value) {
		return $col_value;
		}
		}
	}

	function freeEPR(){
		mysql_free_result($this->result);
		mysql_close($this->sqlconnect);
	}
}
```

**Thesaurus.php**

```
<?php
Class Thesaurus{

private $wordcount=0;
private $resultcount=0;
private $worddata=array();
private $state='';
private $socket;
private $xml_parser;

function __construct() {


if (!defined("CACHE"))
    define("CACHE", FALSE);

if (!defined("GLOBAL_CACHE"))
    define("GLOBAL_CACHE", false);

if (!defined("TTL"))
    define("TTL", 0);

if (!defined("PAGE_CACHE"))
    define("PAGE_CACHE", false);

$this->socket = pfsockopen("10.0.0.2", 49152);
stream_set_blocking ($this->socket , 1);
$this->xml_parser = xml_parser_create();
xml_set_object ($this->xml_parser, $this );
}

function startElementHandler ($parser, $name, $attrib){
switch ($name) {
case $name=="RESULT": {
$worddata[$this->wordcount][$this->resultcount]["exact"] = $attrib["EXACT"];
$worddata[$this->wordcount][$this->resultcount]["stemmed"] = $attrib["STEMMED"];
$worddata[$this->wordcount][$this->resultcount]["translated"] = $attrib["TRANSLATED"];
$worddata[$this->wordcount][$this->resultcount]["editDistance"] = $attrib["EDITDISTANCE"];
break;
}

default : {$this->state=$name;break;}
}
}

function endElementHandler ($parser, $name){

$state='';
if($name=="WORD"){$this->wordcount++;$this->resultcount=0;}
if($name=="RESULT"){$this->resultcount++;}
}

function characterDataHandler ($parser, $data) {
if (!$this->state) {return;}
if ($this->state=="ORIGINAL") { $this->worddata[$this->wordcount]["original"] = $this->worddata[$this-
>wordcount]["original"] . $data;}
```

```
if ($this->state=="SYNONYM") { $this->worddata[$this->wordcount][$this->resultcount]["synonym"] = $this-
>worddata[$this->wordcount][$this->resultcount]["synonym"] . $data;}
if ($this->state=="TERM") { $this->worddata[$this->wordcount][$this->resultcount]["term"] = $this-
>worddata[$this->wordcount][$this->resultcount]["term"] . $data;}
if ($this->state=="EXPLANATION"){$this->worddata[$this->wordcount][$this->resultcount]["explanation"] =
$this->worddata[$this->wordcount][$this->resultcount]["explanation"] . $data;}
if ($this->state=="THESAURUS") { $this->worddata[$this->wordcount][$this->resultcount]["thesaurus"] =
$data;}
}

function translate ($text){

$words= preg_split("/[\s]+/", $text);

$xml =
'<document><sources><thesaurus>1</thesaurus><thesaurus>2</thesaurus><thesaurus>3</thesaurus></sourc
es><text>';
$xmlEnd = ' </text></document>';


foreach($words as $word){
        if(strlen($word)>0)
        $xml = $xml . ' <word>' . $word . '</word>';
}

$xml = $xml . $xmlEnd;
$res = fwrite($this->socket, $xml);
  while (!feof($this->socket)) {
    $desc = $desc . fgets($this->socket, 100);
  }

fclose($this->socket);
xml_set_element_handler($this->xml_parser, "startElementHandler", "endElementHandler");
xml_set_character_data_handler($this->xml_parser, "characterDataHandler");
xml_parser_set_option($this->xml_parser,XML_OPTION_TARGET_ENCODING,"UTF-8");
if(!(xml_parse($this->xml_parser, $desc))){
  die("Error on line " . xml_get_current_line_number($this->xml_parser));
}

xml_parser_free($this->xml_parser);


return $this->worddata;
}


}

?>
```

## B. Source code from the Named Entity Extraction application

**Webservice**

```
/**
   * Web service operation
   */
  @WebMethod(operationName = "summary2")
   public String summary2(@WebParam(name = "journal")String journal) {
      try{
         FileInputStream fstream = new FileInputStream("C:/icd.txt");
         DataInputStream in = new DataInputStream(fstream);
         BufferedReader br = new BufferedReader(new InputStreamReader(in));
         com.aliasi.dict.MapDictionary dictionary = new com.aliasi.dict.MapDictionary();
         String strLine = "";
         while ((strLine = br.readLine()) != null){
            dictionary.addEntry(new com.aliasi.dict.DictionaryEntry(strLine,"ICD",1.0));
         }
         String njournal = stripGarbage(journal);
         System.out.println(njournal);
         com.aliasi.dict.ExactDictionaryChunker dictionaryChunkerTT = new
com.aliasi.dict.ExactDictionaryChunker(dictionary,com.aliasi.tokenizer.IndoEuropeanTokenizerFactory.FACTOR
Y,false,false);
         String retur = chunk(dictionaryChunkerTT,njournal);
         return retur;

      }
      catch(Exception e){

      }

      return null;
   }
```

**Methods**

```
static String chunk(com.aliasi.dict.ExactDictionaryChunker chunker, String text) {
   Chunking chunking = chunker.chunk(text);
   String retur = "";
   for (Chunk chunk : chunking.chunkSet()) {
      int start = chunk.start();
      int end = chunk.end();
      String type = chunk.type();
      double score = chunk.score();
      String phrase = text.substring(start,end);
     retur = retur + ("     phrase=|" + phrase + "|"
                    + " start=" + start
                    + " end=" + end
                    + " type=" + type
                    + " score=" + score);
   }
   return retur;
}

  public static String stripGarbage(String s) {
   String good =
     " abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ0123456789";
   String tokens = "/";
```

```
String result = "";
for ( int i = 0; i < s.length(); i++ ) {
 if ( good.indexOf(s.charAt(i)) >= 0 )
      result += s.charAt(i);
 if ( tokens.indexOf(s.charAt(i)) >= 0 )
    result += " ";
   }
return result;
}
```

# C. Source code from the sentence extraction application

## Code from the library

```
package net.sf.classifier4J.summariser;

import java.util.ArrayList;
import java.util.Collections;
import java.util.Comparator;
import java.util.Iterator;
import java.util.LinkedHashSet;
import java.util.List;
import java.util.Map;
import java.util.Set;

import net.sf.classifier4J.Utilities;

public class SimpleSummariser implements ISummariser {

  private Integer findMaxValue(List input) {
    Collections.sort(input);
    return (Integer) input.get(0);
  }


  protected Set getMostFrequentWords(int count, Map wordFrequencies) {
    return Utilities.getMostFrequentWords(count, wordFrequencies);
  }

  /***
   * @see net.sf.classifier4J.summariser.ISummariser#summarise(java.lang.String)
   */
  public String summarise(String input, int numSentences) {
    // get the frequency of each word in the input
    Map wordFrequencies = Utilities.getWordFrequency(input);

    // now create a set of the X most frequent words
    Set mostFrequentWords = getMostFrequentWords(100, wordFrequencies);

    // break the input up into sentences
    // workingSentences is used for the analysis, but
    // actualSentences is used in the results so that the
    // capitalisation will be correct.
    String[] workingSentences = Utilities.getSentences(input.toLowerCase());
    String[] actualSentences = Utilities.getSentences(input);

    // iterate over the most frequent words, and add the first sentence
    // that includes each word to the result
    Set outputSentences = new LinkedHashSet();
    Iterator it = mostFrequentWords.iterator();
    while (it.hasNext()) {
      String word = (String) it.next();
      for (int i = 0; i < workingSentences.length; i++) {
        if (workingSentences[i].indexOf(word) >= 0) {
          outputSentences.add(actualSentences[i]);
          break;
        }
        if (outputSentences.size() >= numSentences) {
          break;
        }
      }
      if (outputSentences.size() >= numSentences) {
        break;
```

```
        }

    }

    List reorderedOutputSentences = reorderSentences(outputSentences, input);

    StringBuffer result = new StringBuffer("");
    it = reorderedOutputSentences.iterator();
    while (it.hasNext()) {
        String sentence = (String) it.next();
        result.append(sentence);
        result.append("."); // This isn't always correct - perhaps it should be whatever symbol the sentence finished with
        if (it.hasNext()) {
            result.append(" ");
        }
    }

    return result.toString();
}

/***
 * @param outputSentences
 * @param input
 * @return
 */
private List reorderSentences(Set outputSentences, final String input) {
    // reorder the sentences to the order they were in the
    // original text
    ArrayList result = new ArrayList(outputSentences);

    Collections.sort(result, new Comparator() {
        public int compare(Object arg0, Object arg1) {
            String sentence1 = (String) arg0;
            String sentence2 = (String) arg1;

            int indexOfSentence1 = input.indexOf(sentence1.trim());
            int indexOfSentence2 = input.indexOf(sentence2.trim());
            int result = indexOfSentence1 - indexOfSentence2;

            return result;
        }

    });
    return result;
}

}
```

### Code from the webservice

```
/**
  * Web service operation
  */
 @WebMethod(operationName = "summary")
 public String summary(@WebParam(name = "journal") String journal) {

        net.sf.classifier4J.summariser.SimpleSummariser summariser = new
        net.sf.classifier4J.summariser.SimpleSummariser();
        String result = summariser.summarise(journal, 4);
        return result;

 }
```

# D. EPR words used when testing classifiers

```
@relation category
@attribute term string
@attribute class {1 2}
@data
N;a;t;u;r;l;i;g;e,2
f;u;n;k;s;j;o;n;e;r,2
U;a,1
A;l;l;e;r;g;i;e;r,2
I;n;g;e;n,2
k;j;e;n;t;e,2
S;t;i;m;u;l;a;n;t;i;a,1
P;a;s;i;e;n;t;e;n,2
r;ø;y;k;e;r,2
i;k;k;e,2
S;T;A;T;U;S,1
p;r;e;s;e;n;s,1
d;e;n,2
G;e;n;e;r;e;l;l,2
b;e;s;k;r;i;v;e;l;s;e,2
E;n,2
å;r,2
g;a;m;m;e;l,2
m;a;n;n,2
n;o;r;m;a;l;t,2
h;o;l;d,2
g;o;d,2
a;l;l;m;e;n;n;t;i;l;s;t;a;n;d,2
v;å;k;e;n,2
o;g,2
k;l;a;r,2
s;a;m;a;r;b;e;i;d;e;r,2
g;r;e;i;t,2
i;n;g;e;n,2
p;l;a;g;e;r,2
v;e;d,2
u;n;d;e;r;s;ø;k;e;l;s;e;n,2
B;T,1
H;o,1
s;i;d;e,2
v;e,1
s;i;d;e,2
P;u;l;s,1
r;e;g;e;l;m;e;s;s;i;g,2
C;o;l;l;u;m,1
I;n;g;e;n,2
t;e;g;n,2
t;i;l,2
h;a;l;s;v;e;n;e;s;t;u;v;n;i;n;
g,1
h;ø;r;e;r,2
i;n;g;e;n,2
s;t;e;n;o;s;e;l;y;d,1
o;v;e;r,2
c;a;r;o;t;i;d;e;r,1
C;o;r,1
R;e;g;e;l;m;e;s;s;i;g,2
a;k;s;j;o;n,2
h;ø;r;e;r,2
e;n,2
s;y;s;t;o;l;i;s;k,1
b;i;l;y;d,1
g;r;a;d,2
o;v;e;r,2
h;e;l;e,2
p;r;e;c;o;r;d;i;e;t,1
m;e;d,2
p;u;n;k;t;u;m,2
m;a;k;s;i;m;u;m,2
i;n;t;e;r;c;o;s;t;a;l;r;o;m,1
h;o,1
s;t;e;r;n;a;l;r;a;n;d,1
U;t;s;t;r;å;l;i;n;g,2
m;o;t,2
h;a;l;s;k;a;r,1
h;ø;r;e;r,2
o;g;s;å,2
e;n,2
d;i;a;s;t;o;l;i;s;k,1
d;i;s;h;l;y;d,1
l;a;n;g;s,2
v;e,1
s;t;e;r;n;a;l;r;a;n;d,1
P;u;l;m,1
V;e;s;i;k;u;l;ær,1
r;e;s;p;i;r;a;s;j;o;n;s;l;y;d,1
i;n;g;e;n,2
f;r;e;m;m;e;d;l;y;d;e;r,2
A;b;d;o;m;e;n,1
A;r;r,2
e;t;t;e;r,2
a;p;e;n;d;e;c;t;o;m;i,1
A;b;d;o;m;e;n,1
e;r,1
b;l;o;t,1
o;g,2
u;o;m,1
p;a;l;p;e;r;e;r,1
i;n;g;e;n,2
o;p;p;f;y;l;n;i;n;g;e;r,2
i;k;k;e,2
l;e;v;e;r,2
o;g,2
m;i;l;t,1
U;e;x,1
H;a;n,2
h;a;r,2
v;a;r;i;c;e;r,1
i,2
s;a;f;e;n;a,1
m;a;g;n;a,1
d;a;b;e;t;e;s,1
p;å,2
b;e;g;g;e,2
s;i;d;e;r,2
M;e;s;t,2
u;t;t;a;l;t,2
p;å,2
l;e;g;g;e;r,2
F;o;r,2
ø;v;r;i;g,2
s;l;a;n;k;e,2
u;e;x,1
m;e;d,2
g;o;d,2
p;e;r;i;f;e;r,1
p;u;l;s;a;s;j;o;n,1
R;e;s;y;m,1
o;g,2
v;u;r;d;e;r;i;n;g,2
M;a;n;n,2
m;e;d,2
c;o;r;o;n;a;r,1
s;y;k;d;o;m,2
a;o;r;t;a,1
i;n;s;u;f;f;i;s;i;e;n;s,1
g;r;a;d,2
s;a;m;t,2
a;o;r;t;a,2
a;s;c;e;n;d;e;n;s,1
u;t;v;i;d,2
t;i;l,2
c;m,2
L;e;g;g;e;s,2
n;å,2
i;n;n,2
f;o;r,2
A;V;R,1
A;C;B,1
o;p;e;r;a;s;j;o;n,2
s;a;m;t,2
e;v;t,2
s;u;p;e;r;a,1
c;o;r;o;n;a;r;t,1
g;r;a;f;t,1
M;e;d;i;k;a;m;e;n;t;e;r,2
I;n;g;e;n,2
f;a;s;t;e,2
A;l;l;e;r;g;i;e;r,2
I;n;g;e;n,2
k;j;e;n;t;e,2
S;t;i;m;u;l;a;n;t;i;a,1
P;a;s;i;e;n;t;e;n,2
h;a;r,2
s;l;u;t;t;e;t,2
å,2
r;ø;y;k;e,2
S;T;A;T;U;S,1
p;r;e;s;e;n;s,1
d;e;n,2
E;n,2
å;r,2
g;a;m;m;e,2
m;a;n;n,2
o;v;e;r,2
m;i;d;d;e;l;s,2
h;o;l;d,2
g;o;d,2
a;l;l;m;e;n;n;t;i;l;s;t;a;n;d,2
H;a;n,2
e;r,2
v;å;k;e;n,2
o;g,2
k;l;a;r,2
f;o;r;k;l;a;r;e;r,2
s;e;g,2
g;r;e;i;t,2
L;e;t;t,2
h;v;i;l;e;d;y;s;p;n;o;e,1
H;ø;y;d;e,2
c;m,2
V;e;k;t,2
k;g,2
B;T,1
P;u;l;s,2
r;e;g;e;l;m;e;s;s;i;g,2
C;o;r,1
R;e;g;e;l;m;e;s;s;i;g,2
a;k;s;j;o;n,2
r;e;n;e,2
t;o;n;e;r,2
m;e;n,2
t;r;o;l;i;g,2
s;p;l;i;t;t;e;t,2
P;u;l;m;o;n;e;s,1
V;e;s;i;k;u;l;ær,1
r;e;s;p;i;r;a;s;j;o;n;s;l;y;d,1
b;i;l;a;t;e;r;a;l;t,1
T;h;o;r;a;x,1
D;e;t,2
e;r,2
e;t,2
o;p;e;r;a;s;j;o;n;s;a;r;r,2
p;å,2
m;e;d;c;a;l;t,1
v;e,1
ø;v;r;e,2
t;h;o;r;a;x,1
A;b;d;o;m;e;n,1
S;y;m;e;t;r;i;s;k,2
b;l;ø;t,2
u;o;m,1
i;n;g;e;n,2
p;a;l;p;a;b;l;e,1
o;p;p;f;y;l;n;i;n;g;e;r,2
N;o;r;m;a;l;e,2
t;a;r;m;l;y;d;e;r,2
D;e;t,2
e;r,2
s;y;m;e;t;r;i;s;k,2
l;y;s;k;e;p;u;l;s,1
b;i;l;a;t;e;r;a;l;t,1
P;a;s;i;e;n;t;e;n,2
e;r,2
s;a;t;t,2
o;p;p,2
t;i;l,2
o;p;e;r;a;s;j;o;n,2
D;e;t,2
e;r,2
t;a;t;t,2
r;t;g,1
t;h;o;r;a;x,1
n;å,2
D;e;t,2
s;k;a;l,2
d;e;m;o;n;s;t;r;e;r;e;s,2
C;T,2
b;i;l;d;e;r,2
B;e;s;t;i;l;l;e;r,2
e;t;t;e;r;b;e;s;t;i;l;l;e;r,2
l;e;v;e;r;t;r;a;n;s;a;m;e;n;a
;s;e;r,1
B;e;s;t;i;l;l;e;r,2
o;g;s;å,2
E;K;G,1
p;u;l;s,2
s;k;a;l,2
t;a;s,2
g;a;n;g;e;r,2
p;r,2
v;a;k;t,2
f;o;r;s;t;e,2
d;ø;g;n,2
N;a;t;u;r;l;i;g;e,2
f;u;n;k;s;j;o;n;e;r,2
U;a,1
A;l;l;e;r;g;i;e;r,2
I;n;g;e;n,2
k;j;e;n;t;e,2
S;t;i;m;u;l;a;n;t;i;a,1
P;a;s;i;e;n;t;e;n,2
s;l;u;t;t;e;t,2
```

å,2
r;ø;y;k;e,2
å;r,2
g;a;m;m;e;l,2
F;a;s;t;e,2
m;e;d;i;s;i;n;e;r,2
M;a;r;e;v;a;n,1
e;t;t;e;r,2
I;N;R,2
S;T;A;T;U;S,1
p;r;e;s;e;n;s,1
d;e;n,2
G;e;n;e;r;e;l;l,2
b;e;s;k;r;i;v;e;l;s;e,2
E;n,2
å;r,2
g;a;m;m;e;l,2
m;a;n;n,2
n;o;r;m;a;l;t,2
h;o;l;d,2
g;o;d,2
a;l;l;m;e;n;n;t;i;l;s;t;a;n;d,2
v;å;k;e;n,2
o;g,2
k;l;a;r,2
s;a;m;a;r;b;e;i;d;e;r,2
g;r;e;i;t,2
i;n;g;e;n,2
p;l;a;g;e;r,2
v;e;d,2
u;n;d;e;r;s;o;k;e;l;s;e;n,2
B;T,2
H;o,1
s;i;d;e,2
v;e,1
s;i;d;e,2
P;u;l;s,2
r;e;g;e;l;m;e;s;s;i;g,2
C;o;l;l;u;m,1
I;n;g;e;n,2
t;e;g;n,2
t;i;l,2
h;a;l;s;v;e;n;e;s;t;u;v;n;i;n;
g,1
i;n;g;e;n,2
s;t;e;n;o;s;e;l;y;d,1
o;v;e;r,2
c;a;r;o;t;i;d;e;r,1
C;o;r,1
T;y;d;e;l;i;g,2
v;e;n;t;i;l;k;l;i;k;k,1
h;ø;r;e;r,2
i;n;g;e;n,2
b;i;l;y;d;e;r,2
r;e;g;e;l;m;e;s;s;i;g,2
a;k;s;j;o;n,1
P;u;l;m,2
U;a,1
A;b;d;o;m;e;n,2
U;a,1
T;h;o;r;a;x,2
S;T;A;T;U;S,1
e;t;t;e;r,2
s;t;e;r;n;o;t;o;m;i,1
U;e;x,1
G;o;d,2
p;u;l;s,2
a;r;t;e;r;i;a,1
t;i;b;i;a;l;i;s,1
p;o;s;t;e;r;i;o;r,1
b;i;l;a;t;e;r;a;l;t,1
I;n;g;e;n,2
a;n;k;e;l;o;d;e;m;e;r,1
R;e;s;y;m,2
o;g,2
v;u;r;d;e;r;i;n;g,2
M;a;n;n,2
o;p;e;r;e;r;t,2
A;V;R,1
k;o;m;m;e;r,2
n;å,2
t;i;l,2
å;r;s,2
k;o;n;t;r;o;l;l,2
E;k;k;o,2
v;i;s;e;r,2
l;i;t;e;n,2
p;a;r;a;v;a;l;v;u;l;æ;r,1
l;e;k;k;a;s;j;e,2
o;g,2
h;a;n,2
b;ø;r,2
d;e;r;f;o;r,2
k;o;n;t;r;o;l;l;e;r;e;s,2
i;g;j;e;n,2
o;m,2
å;r,2
M;E;D;I;K;A;M;E;N;T;E;R;,2
I;n;g;e;n,2
f;a;s;t;e,2
A;L;L;E;R;G;I;E;R,2
I;n;g;e;n,2
k;j;e;n;t,2
m;e;d;i;k;a;m;e;n;t;e;l;l,2
S;T;A;T;U;S,1
P;R;E;S;E;N;S,1
E;n,2
å;r,2
g;a;m;m;e;l,2
p;i;k;e,2
l;i;t;t,2
t;y;n;n,2
m;e;n,2
k;v;i;k;k,2
o;g,2
r;a;s;k,2
L;ø;p;e;r,2
r;u;n;d;t,2
i,2
r;o;m;m;e;t,2
o;g,2
l;e;k;e;r,2
H;ø;y;d;e,2
V;e;k;t,2
B;T,1
o;g,2
p;u;l;s,2
i;k;k;e,2
t;a;t;t,2
n;å,2
C;o;r,1
R;e;g;e;l;m;e;s;s;i;g,2
a;k;s;j;o;n,2
e;n,2
d;u;s;j;b;i;l;y;d,1
o;v;e;r,2
h;e;l;e,2
h;j;e;r;t;e;t,2
g;r;a;d,2
l;l;l,2
h;o;r;e;s,2
u;t;e;n,2
s;t;e;t;o;s;k;o;p,1
o;g;s;å,2
P;u;l;m,1
V;e;s;i;k;u;l;ær,1
r;e;s;p;i;r;a;s;j;o;n;s;l;y;d,1
b;i;l;a;t,1
A;b;d;o;m;e;n,1
S;y;m;m;e;t;r;i;s;k,1
b;l;ø;t,2
o;g,2
u;o;m,1
I;n;f;o;r;m;e;r;e;r,2
h;o;v;e;d;o;p;e;r;a;t;o;r,1
o;m,2
p;a;s;i;e;n;t;e;n,2
o;g,2
a;t,2
d;e;t,2
i;k;k;e,2
e;r,2
p;a;p;i;r;e;r,2
t;i;l;s;t;e;d;e,2
H;u;n,2
h;a;r,2
v;ær;t,2
t;i;l,2
t;i;l;s;y;n,2
p;å,2
b;a;r;n;e;m;e;d;i;s;i;n;s;k,2
o;g,2
s;k;a;l,2
t;r;o;l;i;g,2
t;a;s,2
o;p;p,2
p;å,2
m;ø;t;e,2
n;å,2
H;u;n,2
s;t;å;r,2
p;å,2
o;p;e;r;a;s;j;o;n;s;p;r;o;g;r;
a;m;m;e;t,2
t;i;l,2
i,2
m;o;r;g;e;n,2
D;e;t,2
e;r,2
t;a;t;t,2
n;e;s;e;p;r;ø;v;e;r,2
N;A;T;U;R;L;I;G;E,2
F;U;N;K;S;J;O;N;E;R,2
U;a,1
M;E;D;I;K;A;M;E;N;T;E;R,2
M;o;d;u;r;e;t;i;c,1
m;i;t;e,1
t;a;b;l,2
A;l;b;y;l;-E,1
m;g,1
S;T;I;M;U;L;A;N;T;I;A,1
S;l;u;t;t;e;t,2
å,2
r;ø;k;e,2
f;o;r,2
å;r,2
s;i;d;e;n,2
A;L;L;E;R;G;I;E;R,2
I;n;g;e;n,2
k;j;e;n;t;e,2
S;T;A;T;U;S,1
P;R;E;S;E;N;S,1
P;a;s;i;e;n;t;e;n,2
k;o;m;m;e;r,2
g;å;e;n;d;e,2
t;i;l,2
u;n;d;e;r;s;ø;k;e;l;s;e;n,2
H;u;n,2
v;i;r;k;e;r,2
t;i;l,2
å,2
v;ær;e,2
i,2
r;e;l;a;t;i;v;t,2
g;o;d,2
a;l;m;e;n;t;i;l;s;t;a;n;d,2
i,2
f;o;r;h;o;l;d,2
t;i;l,2
a;l;d;e;r,2
L;i;t;t,2
o;v;e;r,2
n;o;r;m;a;l;t,2
h;o;l;d,2
H;u;n,2
h;a;r,2
i;n;g;e;n,2
f;u;n;k;s;j;o;n;s;d;y;s;p;n;o;
e,1
a;v,2
b;e;t;y;d;n;i;n;g,2
K;l;a;r,2
o;g,2
o;r;i;e;n;t;e;r;t,2
o;g,2
s;a;m;a;r;b;e;i;d;e;r,2
g;r;e;i;t,2
I;n;g;e;n,2
i;c;t;e;r;u;s,1
c;y;a;n;o;s;e,1
e;l;l;e;r,2
g;e;n;e;r;e;l;l,2
g;l;a;n;d;e;l;s;v;u;l;s;t,1
H;u;n,2
h;a;r,2
b;e;t;y;d;e;l;i;g,2
v;a;r;i;k;ø;s;e,1
f;o;r;a;n;d;r;i;n;g;e;r,2
p;å,2
b;e;g;g;e,2
u;n;d;e;r;e;k;s;t;r,1
o;g,2
l;e;t;t;e,2
ø;d;e;m;e;r,1
B;T,1
h;o;a;r;m,1
l;i;g;g;e;n;d;e,2
P;u;l;s,2
r;e;g;e;l;m;e;s;s;i;g,2
D;o;p;p;l;e;r;t;r;y;k;k,1
h;o;u;n;d;e;r;e;k;s;t;r,1
u;n;d;e;r;e;k;s;t;r,1
C;o;l;l;u;m,1
A;r;r,2
e;t;t;e;r,2
k;a;r;o;t;i;s;k;i;r;u;r;g;i,1
p;å,2
v;e,1
s;i;d;e,2
S;t;e;n;o;s;l;y;d,1
h;e;r,2
P;å,2
h;o,1
s;i;d;e,2

e;r,2
d;e;r,2
a;n;g;i;v;e;l;i;g,2
o;k;k;l;u;d;e;r;t,1
k;a;r;o;t;i;s,1
i;f;ø;l;g;e,2
p;a;s;i;e;n;t;e;n;s,2
d;a;t;t;e;r,2
h;e;r,2
e;r,2
d;e;t,2
u;t;b;r;e;d;n;i;n;g,2
a;v,2
s;t;e;n;o;s;e;l;y;d,1
o;v;e;r,2
c;o;r,1
C;o;r,1
R;e;n;e,2
t;o;n;e;r,2
s;y;s;t;o;l;i;s;k,1
b;i;l;y;d,1
g;r;a;d,2
m;e;d,2
P;u;l;m,1
S;o;n;o;r,1
p;e;r;k;u;s;j;o;n;s;l;y;d,1
v;e;s;i;k;u;l;æ;r,1
r;e;s;p;i;r;a;s;j;o;n;s;l;y;d,1
i;n;g;e;n,2
f;r;e;m;m;e;d;l;y;d;e;r,2
A;b;d;o;m;e;n,1
B;l;ø;t,2
o;g,2
u;o;m,1
A;o;r;t;a,1
p;a;l;p;e;r;e;s,1
b;r;e;d;d;e;f;o;r;ø;k;e;t,2
c;a,2
c;m,2
i,2
e;p;i;g;a;s;t;r;i;e;t,1
O;g;s;å,2
t;y;d;e;l;i;g,2
p;a;l;p;a;b;e;l,1
s;v;a;r;e;n;d;e,2
t;i;l,2
a;o;r;t;a;b;i;f;u;r;k;a;t;u;r;e
;n,1
o;g,2
u;t,2
m;o;t,2
h;o,1
i;l;i;a;c;a,1
H;u;n,2
h;a;r,2
l;y;s;k;e;p;u;l;s,1
p;å,2
b;e;g;g;e,2
s;i;d;e;r,1
h;o,1
f;o;s;s;a,1
i;l;i;a;c;a,1
e;r,2
d;e;t,2
a;r;r,2
e;t;t;e;r,2
d;e;l;s,2
t;i;d;l;i;g;e;r;e,2
a;p;p;e;n;d;e;c;t;o;m;i,1
o;g,2
d;e;l;s,2

e;t;t;e;r,2
o;p;e;r;a;s;j;o;n,2
f;o;r,2
D;e;t,2
e;r,2
o;g;s;å,2
s;n;i;t;t,2
i,2
b;e;g;g;e,2
l;y;s;k;e;r,2
p;å,2
v;e,1
s;i;d;e,2
e;t;t;e;r,2
v;a;r;i;s;e;k;i;r;u;r;g;i,1
p;å,2
h;o,1
s;i;d;e,2
e;t;t;e;r,2
k;a;r;k;i;r;u;r;g;i,1
M;E;D;I;K;A;M;E;N;T;E;R,2
D;i;g;i;t;o;x;i;n,1
S;e;l;o;-z;o;k,1
m;g,1
R;e;n;i;t;e;c,1
m;g,1
A;d;a;l;a;t,1
O;r;o;s,1
m;g,1
A;L;L;E;R;G;I;E;R,2
I;n;g;e;n,2
k;j;e;n;t,2
m;e;d;i;k;a;m;e;n;t;e;l;l,2
S;T;I;M;U;L;A;N;T;I;A,1
P;a;s;i;e;n;t;e;n,2
r;ø;k;e;r,2
f;r;e;m;d;e;l;e;s,2
c;a,2
r;u;l;l;e;s;i;g;a;r;e;t;t;e;r,2
d;g;l,2
S;T;A;T;U;S,1
P;R;E;S;E;N;S,1
E;n,2
å;r,2
g;a;m;m;e;l,2
m;a;n;n,2
i,2
n;o;r;m;a;l;t,2
h;o;l;d,2
g;o;d,2
a;l;m;e;n;t;i;l;s;t;a;n;d,2
v;i;r;k;e;r,2
n;o;e,2
d;e;h;y;d;r;e;r;t,1
V;å;k;e;n,2
o;g,2
k;l;a;r,2
f;o;r;k;l;a;r;e;r,2
s;e;g,2
g;r;e;i;t,2
H;ø;y;d;e,2
c;m,2
V;e;k;t,2
k;g,2
B;T,1
P;u;l;s,2
l;i;t;t,2
u;r;e;g;e;l;m;e;s;s;i;g,2
C;o;r,1
L;i;t;t,2
u;r;e;g;e;l;m;e;s;s;i;g,2

a;k;s;j;o;n,2
m;e;n,2
i;n;g;e;n,2
p;u;l;s;d;e;f;i;s;i;t;t,1
s;p;o;r;s;m;å;l,2
b;i;l;y;d,1
g;r;a;d,2
m;e;d,2
v;e;d,2
a;p;e;x,1
D;e;t,2
e;r,2
n;o;e,2
s;p;l;i;t;t;e;t,2
I;n;g;e;n,2
p;u;l;s;d;e;f;i;s;i;t;t,1
P;u;l;m,1
L;e;t;t;e,2
i;n;t;e;r;c;o;s;t;a;l;e,1
i;n;n;d;r;a;g;n;i;n;g;e;r,1
v;e;d,2
r;e;s;p;i;r;a;s;j;o;n,1
H;ø;r;e;r,2
i;n;g;e;n,2
f;r;e;m;m;e;d;l;y;d;e;r,2
A;b;d;o;m;e;n,1
S;y;m;m;e;t;r;i;s;k,1
o;g,2
b;l;ø;t,2
k;j;e;n;n;e;r,2
t;y;d;e;l;i;g,2
b;u;k;a;o;r;t;a;a;n;e;u;r;i;s;
m;e,1
p;a;l;p;e;r;e;s,1
t;i;l,2
c;m,2
b;r;e;d;t,2
i;n;g;e;n,2
ø;m;h;e;t,2
o;v;e;r,2
d;e;t;t;e,2
E;l;l;e;r;s,2
i;n;g;e;n,2
o;p;p;f;y;l;n;i;n;g;e;r,1
O;p;e;r;a;s;j;o;n;s;a;r;r,2
h;o,1
l;y;s;k;e,2
p;r;o;x;i;m;a;l;e,1
e;n;d;e,2
a;v,2
d;e;t;t;e,2
e;r,2
d;e;t,2
e;r,2
l;i;t;e,2
D;e;t,2
e;r,2
s;y;m;m;e;t;r;i;s;k,1
l;y;s;k;e;p;u;l;s,1
P;a;l;p;a;b;e;l,1
p;u;l;s,2
d;i;s;t;a;l;t,1
o;g,2
i;n;g;e;n,2
a;n;k;e;l;ø;d;e;m;e;r,1
T;I;L;T;A;K,2
P;a;s;i;e;n;t;e;n,2
h;a;r,2
v;æ;r;t,2
t;i;l,2
C;T,1

t;h;o;r;a;x,1
o;g,2
s;k;a;l,2
o;g;s;å,2
t;i;l,2
e;k;k;o,1
c;o;r,1
V;a;n;l;i;g;e,2
b;l;o;d;p;r;ø;v;e;r,2
t;a;s,2
M;e;d;i;k;a;m;e;n;t;e;r,2
N;o;r;m;o;r;i;x,1
M;i;t;e,1
t;a;b;l,2
C;l;a;r;i;t;y;n,1
M;a;r;e;v;a;n,1
e;t;t;e;r,2
l;i;s;t;e,2
S;o;t;a;l;o;l,1
m;g,1
m;g,1
P;r;a;v;a;c;h;o;l,1
m;g,1
v;e;s;p,1
A;l;l;o;p;u;r,1
m;g,1
v;e;s;p,1
A;l;l;e;r;g;i;e;r,2
I;n;g;e;n,2
k;j;e;n;t;e,2
S;t;i;m;u;l;a;n;t;i;a,1
P;a;s;i;e;n;t;e;n,2
r;ø;y;k;e;r,2
i;k;k;e,2
S;T;A;T;U;S,1
p;r;e;s;e;n;s,1
E;n,2
å;r,2
g;a;m;m;e;l,2
k;v;i;n;n;e,2
o;g,2
l;i;t;t,2
o;v;e;r,2
m;i;d;d;e;l;s,2
h;o;l;d,2
g;o;d,2
a;l;l;m;e;n;n;t;i;l;s;t;a;n;d,2
H;u;n,2
e;r,2
v;å;k;e;n,2
o;g,2
k;l;a;r,2
o;g,2
f;o;r;k;l;a;r;e;r,2
s;e;g,2
g;r;e;i;t,2
H;ø;y;d;e,2
c;m,2
V;e;k;t,2
k;g,2
B;T,1
P;u;l;s,2
r;e;g;e;l;m;e;s;s;i;g,2
C;o;r,1
R;e;g;e;l;m;e;s;s;i;g,2
a;k;s;j;o;n,2
k;r;a;f;t;i;g,2
s;y;s;t;o;l;i;s;k,1
b;i;l;y;d,1
o;v;e;r,2
h;e;l;e,2

c;o;r,1
h;ø;r;e;s,2
i,2
g;r;u;n;n;e;n,2
b;e;s;t,2
i,2
h;o,1
t;h;o;r;a;x,1
g;r;a;d,2
H;ø;r;e;r,2
i;k;k;e,2
P;u;l;m;o;n;e;s,1
V;e;s;i;k;u;l;æ;r,1
r;e;s;p;i;r;a;s;j;o;n;s;l;y;d,1
b;i;l;a;t;e;r;a;l;t,1
A;b;d;o;m;e;n,1
S;y;m;e;t;r;i;s;k,1
b;l;ø;t,1
u;o;m,1
i;n;g;e;n,2
p;a;l;p;a;b;l;e,1
o;p;p;f;y;l;n;i;n;g;e;r,2
N;o;r;m;a;l;e,2
t;a;r;m;l;y;d;e;r,2
d;e;t,2
e;r,2
s;y;m;e;t;r;i;s;k,1
l;y;s;k;e;p;u;l;s,1
b;i;l;a;t;e;r;a;l;t,1
d;i;s;t;a;l;t,1
L;e;t;t;e,2
ø;d;e;m;e;r,1
G;o;d,2
o;g,2
v;a;r;m,2
K;a;n,2
i;k;k;e,2
m;e;d,2
s;i;k;k;e;r;h;e;t,2
s;i,2
j;e;g,2
k;j;e;n;n;e;r,2
p;u;l;s,2
i,2
d;i;s;t;a;l;e,1
c;o;r,1
P;a;s;i;e;n;t;e;n,2
h;a;r,2
v;æ;r;t,2
t;i;l,2
r;t;g,2
t;h;o;r;a;x,1
i,2
d;a;g,2
P;l;a;n;l;a;g;t,2
o;p;e;r;a;s;j;o;n,2
f;o;r,2
a;o;r;t;a,1
s;t;e;n;o;s;e,1
A;v;v;e;n;t;e;r,2
o;p;e;r;a;s;j;o;n;s;d;a;g,2
B;e;s;t;i;l;l;e;r,2
o;g;s;å,2
E;K;G,1
V;a;n;l;i;g,2
r;u;t;i;n;e;p;r;ø;v;e;r,2
t;a;s,2
M;e;d;i;k;a;m;e;n;t;e;r,2

# E. Results from text mining with EPRs

**Naïve Bayes with bigrams**

Time taken to build model: 183.34 seconds

=== Evaluation on test set ===
=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 546 | 60.8696 % |
| Incorrectly Classified Instances | 351 | 39.1304 % |
| Kappa statistic | 0.2334 | |
| Mean absolute error | 0.4094 | |
| Root mean squared error | 0.5755 | |
| Relative absolute error | 81.7779 % | |
| Root relative squared error | 114.9596 % | |
| Total Number of Instances | 897 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.735 | 0.441 | 0.396 | 0.735 | 0.515 | 0.723 | 1 |
| 0.559 | 0.265 | 0.843 | 0.559 | 0.672 | 0.723 | 2 |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 186 | 67 | \| a = 1 |
| 284 | 360 | \| b = 2 |

**Naive Bayes with trigrams**

Time taken to build model: 371.77 seconds

=== Evaluation on test set ===
=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 558 | 62.2074 % |
| Incorrectly Classified Instances | 339 | 37.7926 % |
| Kappa statistic | 0.2596 | |
| Mean absolute error | 0.4088 | |
| Root mean squared error | 0.5775 | |
| Relative absolute error | 81.6572 % | |
| Root relative squared error | 115.3581 % | |
| Total Number of Instances | 897 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.759 | 0.432 | 0.409 | 0.759 | 0.531 | 0.708 | 1 |
| 0.568 | 0.241 | 0.857 | 0.568 | 0.683 | 0.708 | 2 |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 192 | 61 | \| a = 1 |
| 278 | 366 | \| b = 2 |

**Complement bayes with bigrams**
Time taken to build model: 0.11 seconds

=== Evaluation on test set ===
=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 667 | 74.359 % |
| Incorrectly Classified Instances | 230 | 25.641 % |
| Kappa statistic | 0.4546 | |
| Mean absolute error | 0.2564 | |
| Root mean squared error | 0.5064 | |
| Relative absolute error | 51.2197 % | |
| Root relative squared error | 101.1505 % | |
| Total Number of Instances | 897 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.81 | 0.283 | 0.53 | 0.81 | 0.641 | 0.764 | 1 |
| 0.717 | 0.19 | 0.906 | 0.717 | 0.801 | 0.764 | 2 |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 205 | 48 | \| a = 1 |
| 182 | 462 | \| b = 2 |

**Complement Bayes with trigrams**
Time taken to build model: 0.13 seconds

=== Evaluation on test set ===
=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 632 | 70.4571 % |
| Incorrectly Classified Instances | 265 | 29.5429 % |
| Kappa statistic | 0.419 | |
| Mean absolute error | 0.2954 | |
| Root mean squared error | 0.5435 | |
| Relative absolute error | 59.014 % | |
| Root relative squared error | 108.5743 % | |
| Total Number of Instances | 897 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.897 | 0.371 | 0.487 | 0.897 | 0.631 | 0.763 | 1 |
| 0.629 | 0.103 | 0.94 | 0.629 | 0.753 | 0.763 | 2 |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 227 | 26 | \| a = 1 |
| 239 | 405 | \| b = 2 |

**J48 with bigrams**
Time taken to build model: 26134.67 seconds

=== Evaluation on test set ===
=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 720 | 80.2676 % |
| Incorrectly Classified Instances | 177 | 19.7324 % |
| Kappa statistic | 0.551 | |
| Mean absolute error | 0.2095 | |
| Root mean squared error | 0.4227 | |
| Relative absolute error | 41.845 % | |
| Root relative squared error | 84.4376 % | |
| Total Number of Instances | 897 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.791 | 0.193 | 0.617 | 0.791 | 0.693 | 0.848 | 1 |
| 0.807 | 0.209 | 0.908 | 0.807 | 0.855 | 0.848 | 2 |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 200 | 53 | \| a = 1 |
| 124 | 520 | \| b = 2 |

**J48 with trigrams**
Time taken to build model: 68695.26 seconds

=== Evaluation on test set ===
=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 693 | 77.2575 % |
| Incorrectly Classified Instances | 204 | 22.7425 % |
| Kappa statistic | 0.4762 | |
| Mean absolute error | 0.2426 | |
| Root mean squared error | 0.4531 | |
| Relative absolute error | 48.4654 % | |
| Root relative squared error | 90.5058 % | |
| Total Number of Instances | 897 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.715 | 0.205 | 0.578 | 0.715 | 0.64 | 0.81 | 1 |
| 0.795 | 0.285 | 0.877 | 0.795 | 0.834 | 0.81 | 2 |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 181 | 72 | \| a = 1 |
| 132 | 512 | \| b = 2 |

**LibSVM with bigrams**
Time taken to build model: 428.72 seconds

=== Evaluation on test set ===
=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 726 | 80.9365 % |
| Incorrectly Classified Instances | 171 | 19.0635 % |
| Kappa statistic | 0.5765 | |
| Mean absolute error | 0.1906 | |
| Root mean squared error | 0.4366 | |
| Relative absolute error | 38.0807 % | |
| Root relative squared error | 87.2172 % | |
| Total Number of Instances | 897 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.846 | 0.205 | 0.618 | 0.846 | 0.715 | 0.82 | 1 |
| 0.795 | 0.154 | 0.929 | 0.795 | 0.857 | 0.82 | 2 |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 214 | 39 | \| a = 1 |
| 132 | 512 | \| b = 2 |

**LibSVM with trigrams**
Time taken to build model: 696.4 seconds

=== Evaluation on test set ===
=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 711 | 79.2642 % |
| Incorrectly Classified Instances | 186 | 20.7358 % |
| Kappa statistic | 0.5148 | |
| Mean absolute error | 0.2074 | |
| Root mean squared error | 0.4554 | |
| Relative absolute error | 41.4212 % | |
| Root relative squared error | 90.9621 % | |
| Total Number of Instances | 897 | |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.723 | 0.18 | 0.612 | 0.723 | 0.663 | 0.772 | 1 |
| 0.82 | 0.277 | 0.883 | 0.82 | 0.85 | 0.772 | 2 |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 183 | 70 | \| a = 1 |
| 116 | 528 | \| b = 2 |

## F. Results from cross validation

| Set | # | Fold | Classifier | Train ins. | Test inst. | Correct | Incorrect | Correct % | Incorrect % | Kappa | TP rate | TP | FP rate | FP | TN rate | TN | FN rate | FN | Precision | Recall | Train time | Test time |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bi | 1 | 1 | Complement Bayes | 40295 | 4478 | 3831 | 647 | 85.551586 | 14.448414 | 0.711002 | 0.873942 | 1962 | 0.163009 | 364 | 0.836991 | 1869 | 0.126058 | 283 | 0.843508 | 0.873942 | 0.141 | 0.016 |
| Bi | 1 | 2 | Complement Bayes | 40295 | 4478 | 3811 | 667 | 85.104958 | 14.895042 | 0.702043 | 0.885078 | 1987 | 0.183162 | 409 | 0.816838 | 1824 | 0.114922 | 258 | 0.829299 | 0.885078 | 0.11 | 0.031 |
| Bi | 1 | 3 | Complement Bayes | 40295 | 4478 | 3797 | 681 | 84.792318 | 15.207682 | 0.695804 | 0.873051 | 1960 | 0.17734 | 396 | 0.82266 | 1837 | 0.126949 | 285 | 0.831919 | 0.873051 | 0.14 | 0.047 |
| Bi | 1 | 4 | Complement Bayes | 40296 | 4477 | 3821 | 656 | 85.347331 | 14.652669 | 0.706911 | 0.873497 | 1961 | 0.166667 | 372 | 0.833333 | 1860 | 0.126503 | 284 | 0.840549 | 0.873497 | 0.156 | 0.047 |
| Bi | 1 | 5 | Complement Bayes | 40296 | 4477 | 3891 | 586 | 86.910878 | 13.089122 | 0.738197 | 0.881514 | 1979 | 0.143369 | 320 | 0.856631 | 1912 | 0.118486 | 266 | 0.860809 | 0.881514 | 0.11 | 0.031 |
| Bi | 1 | 6 | Complement Bayes | 40296 | 4477 | 3826 | 651 | 85.459013 | 14.540987 | 0.70914 | 0.877506 | 1970 | 0.168459 | 376 | 0.831541 | 1856 | 0.122494 | 275 | 0.839727 | 0.877506 | 0.109 | 0.031 |
| Bi | 1 | 7 | Complement Bayes | 40296 | 4477 | 3865 | 612 | 86.330132 | 13.669868 | 0.726571 | 0.882405 | 1981 | 0.155914 | 348 | 0.844086 | 1884 | 0.117595 | 264 | 0.85058 | 0.882405 | 0.109 | 0.031 |
| Bi | 1 | 8 | Complement Bayes | 40296 | 4477 | 3839 | 638 | 85.749386 | 14.250614 | 0.714956 | 0.875724 | 1966 | 0.160842 | 359 | 0.839158 | 1873 | 0.124276 | 279 | 0.845591 | 0.875724 | 0.766 | 0.031 |
| Bi | 1 | 9 | Complement Bayes | 40296 | 4477 | 3836 | 641 | 85.682377 | 14.317623 | 0.71361 | 0.878396 | 1972 | 0.164875 | 368 | 0.835125 | 1864 | 0.121604 | 273 | 0.842735 | 0.878396 | 0.25 | 0.078 |
| Bi | 1 | 10 | Complement Bayes | 40296 | 4477 | 3833 | 644 | 85.615367 | 14.384633 | 0.712262 | 0.887255 | 1991 | 0.175101 | 391 | 0.824899 | 1842 | 0.112745 | 253 | 0.835852 | 0.887255 | 0.11 | 0.031 |
| Bi | 2 | 1 | Complement Bayes | 40295 | 4478 | 3840 | 638 | 85.752568 | 14.247432 | 0.715014 | 0.881069 | 1978 | 0.166144 | 371 | 0.833856 | 1862 | 0.118931 | 267 | 0.84206 | 0.881069 | 0.546 | 0.032 |
| Bi | 2 | 2 | Complement Bayes | 40295 | 4478 | 3813 | 665 | 85.14962 | 14.85038 | 0.702942 | 0.882405 | 1981 | 0.179579 | 401 | 0.820421 | 1832 | 0.117595 | 264 | 0.831654 | 0.882405 | 0.125 | 0.047 |
| Bi | 2 | 3 | Complement Bayes | 40295 | 4478 | 3798 | 680 | 84.814649 | 15.185351 | 0.696252 | 0.872606 | 1959 | 0.176444 | 394 | 0.823556 | 1839 | 0.127394 | 286 | 0.832554 | 0.872606 | 0.141 | 0.015 |
| Bi | 2 | 4 | Complement Bayes | 40296 | 4477 | 3809 | 668 | 85.079294 | 14.920706 | 0.701551 | 0.873886 | 1961 | 0.172414 | 385 | 0.827586 | 1848 | 0.126114 | 283 | 0.835891 | 0.873886 | 0.125 | 0.047 |
| Bi | 2 | 5 | Complement Bayes | 40296 | 4477 | 3847 | 630 | 85.928077 | 14.071923 | 0.718516 | 0.885969 | 1989 | 0.167563 | 374 | 0.832437 | 1858 | 0.114031 | 256 | 0.841727 | 0.885969 | 0.125 | 0.047 |
| Bi | 2 | 6 | Complement Bayes | 40296 | 4477 | 3829 | 648 | 85.526022 | 14.473978 | 0.71048 | 0.877951 | 1971 | 0.167563 | 374 | 0.832437 | 1858 | 0.122049 | 274 | 0.840512 | 0.877951 | 0.109 | 0.032 |
| Bi | 2 | 7 | Complement Bayes | 40296 | 4477 | 3829 | 648 | 85.526022 | 14.473978 | 0.710476 | 0.880624 | 1977 | 0.170251 | 380 | 0.829749 | 1852 | 0.119376 | 268 | 0.838778 | 0.880624 | 0.172 | 0.047 |
| Bi | 2 | 8 | Complement Bayes | 40296 | 4477 | 3833 | 644 | 85.615367 | 14.384633 | 0.712282 | 0.870379 | 1954 | 0.158154 | 353 | 0.841846 | 1879 | 0.129621 | 291 | 0.846987 | 0.870379 | 0.14 | 0.016 |
| Bi | 2 | 9 | Complement Bayes | 40296 | 4477 | 3866 | 611 | 86.352468 | 13.647532 | 0.727027 | 0.876615 | 1968 | 0.149642 | 334 | 0.850358 | 1898 | 0.123385 | 277 | 0.854909 | 0.876615 | 0.11 | 0.031 |
| Bi | 2 | 10 | Complement Bayes | 40296 | 4477 | 3864 | 613 | 86.307795 | 13.692205 | 0.726121 | 0.884187 | 1985 | 0.158154 | 353 | 0.841846 | 1879 | 0.115813 | 260 | 0.849016 | 0.884187 | 0.11 | 0.047 |
| Bi | 3 | 1 | Complement Bayes | 40295 | 4478 | 3802 | 676 | 84.903975 | 15.096025 | 0.698053 | 0.864588 | 1941 | 0.166592 | 372 | 0.833408 | 1861 | 0.135412 | 304 | 0.83917 | 0.864588 | 0.172 | 0.031 |
| Bi | 3 | 2 | Complement Bayes | 40295 | 4478 | 3846 | 632 | 85.886556 | 14.113444 | 0.717691 | 0.884633 | 1986 | 0.16704 | 373 | 0.83296 | 1860 | 0.115367 | 259 | 0.841882 | 0.884633 | 0.109 | 0.047 |
| Bi | 3 | 3 | Complement Bayes | 40295 | 4478 | 3849 | 629 | 85.953551 | 14.046449 | 0.719039 | 0.879733 | 1975 | 0.16077 | 359 | 0.83923 | 1874 | 0.120267 | 270 | 0.846187 | 0.879733 | 0.125 | 0.046 |
| Bi | 3 | 4 | Complement Bayes | 40296 | 4477 | 3846 | 631 | 85.90574 | 14.09426 | 0.71808 | 0.879287 | 1974 | 0.16129 | 360 | 0.83871 | 1872 | 0.120713 | 271 | 0.845758 | 0.879287 | 0.141 | 0.453 |
| Bi | 3 | 5 | Complement Bayes | 40296 | 4477 | 3837 | 640 | 85.704713 | 14.295287 | 0.714043 | 0.88686 | 1991 | 0.172939 | 386 | 0.827061 | 1846 | 0.11314 | 254 | 0.83761 | 0.88686 | 0.141 | 0.046 |
| Bi | 3 | 6 | Complement Bayes | 40296 | 4477 | 3826 | 651 | 85.459013 | 14.540987 | 0.709124 | 0.88686 | 1991 | 0.177867 | 397 | 0.822133 | 1835 | 0.11314 | 254 | 0.833752 | 0.88686 | 0.125 | 0.078 |
| Bi | 3 | 7 | Complement Bayes | 40296 | 4477 | 3820 | 657 | 85.324994 | 14.675006 | 0.706465 | 0.872606 | 1959 | 0.166219 | 371 | 0.833781 | 1861 | 0.127394 | 286 | 0.840773 | 0.872606 | 0.672 | 0.031 |
| Bi | 3 | 8 | Complement Bayes | 40296 | 4477 | 3900 | 577 | 87.111905 | 12.888095 | 0.742216 | 0.884633 | 1986 | 0.142473 | 318 | 0.857527 | 1914 | 0.115367 | 259 | 0.861979 | 0.884633 | 0.125 | 0.031 |
| Bi | 3 | 9 | Complement Bayes | 40296 | 4477 | 3770 | 707 | 84.208175 | 15.791825 | 0.68412 | 0.864588 | 1941 | 0.180556 | 403 | 0.819444 | 1829 | 0.135412 | 304 | 0.828072 | 0.864588 | 0.141 | 0.031 |
| Bi | 3 | 10 | Complement Bayes | 40296 | 4477 | 3855 | 622 | 86.106768 | 13.893232 | 0.722098 | 0.887255 | 1991 | 0.165249 | 369 | 0.834751 | 1864 | 0.112745 | 253 | 0.843644 | 0.887255 | 0.157 | 0.015 |
| Bi | 4 | 1 | Complement Bayes | 40295 | 4478 | 3804 | 674 | 84.948638 | 15.051362 | 0.698926 | 0.877506 | 1970 | 0.178683 | 399 | 0.821317 | 1834 | 0.122494 | 275 | 0.831575 | 0.877506 | 0.188 | 0.031 |
| Bi | 4 | 2 | Complement Bayes | 40295 | 4478 | 3853 | 625 | 86.042876 | 13.957124 | 0.720828 | 0.879287 | 1974 | 0.158531 | 354 | 0.841469 | 1879 | 0.120713 | 271 | 0.847938 | 0.879287 | 0.141 | 0.046 |
| Bi | 4 | 3 | Complement Bayes | 40295 | 4478 | 3858 | 620 | 86.154533 | 13.845467 | 0.723048 | 0.889087 | 1996 | 0.166144 | 371 | 0.833856 | 1862 | 0.110913 | 249 | 0.843262 | 0.889087 | 0.14 | 0.016 |
| Bi | 4 | 4 | Complement Bayes | 40296 | 4477 | 3837 | 640 | 85.704713 | 14.295287 | 0.714053 | 0.881069 | 1978 | 0.167115 | 373 | 0.832885 | 1859 | 0.118931 | 267 | 0.841344 | 0.881069 | 0.125 | 0.031 |
| Bi | 4 | 5 | Complement Bayes | 40296 | 4477 | 3832 | 645 | 85.593031 | 14.406969 | 0.711833 | 0.871269 | 1956 | 0.159498 | 356 | 0.840502 | 1876 | 0.128731 | 289 | 0.846021 | 0.871269 | 0.125 | 0.031 |
| Bi | 4 | 6 | Complement Bayes | 40296 | 4477 | 3832 | 645 | 85.593031 | 14.406969 | 0.711835 | 0.870379 | 1954 | 0.158602 | 354 | 0.841398 | 1878 | 0.129621 | 291 | 0.84662 | 0.870379 | 0.141 | 0.016 |
| Bi | 4 | 7 | Complement Bayes | 40296 | 4477 | 3864 | 613 | 86.307795 | 13.692205 | 0.726126 | 0.881069 | 1978 | 0.155018 | 346 | 0.844982 | 1886 | 0.118931 | 267 | 0.851119 | 0.881069 | 0.11 | 0.031 |
| Bi | 4 | 8 | Complement Bayes | 40296 | 4477 | 3834 | 643 | 85.637704 | 14.362296 | 0.712717 | 0.877506 | 1970 | 0.164875 | 368 | 0.835125 | 1864 | 0.122494 | 275 | 0.842601 | 0.877506 | 0.094 | 0.047 |
| Bi | 4 | 9 | Complement Bayes | 40296 | 4477 | 3796 | 681 | 84.788921 | 15.211079 | 0.695736 | 0.870824 | 1955 | 0.175179 | 391 | 0.824821 | 1841 | 0.129176 | 290 | 0.833333 | 0.870824 | 0.11 | 0.031 |
| Bi | 4 | 10 | Complement Bayes | 40296 | 4477 | 3835 | 642 | 85.66004 | 14.33996 | 0.713152 | 0.889929 | 1997 | 0.176892 | 395 | 0.823108 | 1838 | 0.110071 | 247 | 0.834866 | 0.889929 | 0.109 | 0.047 |
| Bi | 5 | 1 | Complement Bayes | 40295 | 4478 | 3856 | 622 | 86.10987 | 13.89013 | 0.722158 | 0.88686 | 1991 | 0.164801 | 368 | 0.835199 | 1865 | 0.11314 | 254 | 0.844002 | 0.88686 | 0.453 | 0.031 |
| Bi | 5 | 2 | Complement Bayes | 40295 | 4478 | 3860 | 618 | 86.199196 | 13.800804 | 0.723954 | 0.881514 | 1979 | 0.157635 | 352 | 0.842365 | 1881 | 0.118486 | 266 | 0.848992 | 0.881514 | 0.109 | 0.031 |
| Bi | 5 | 3 | Complement Bayes | 40295 | 4478 | 3858 | 620 | 86.154533 | 13.845467 | 0.723055 | 0.884633 | 1986 | 0.161666 | 361 | 0.838334 | 1872 | 0.115367 | 259 | 0.846187 | 0.884633 | 0.109 | 0.031 |
| Bi | 5 | 4 | Complement Bayes | 40296 | 4477 | 3841 | 636 | 85.794059 | 14.205941 | 0.715835 | 0.889929 | 1997 | 0.174205 | 389 | 0.825795 | 1844 | 0.110071 | 247 | 0.836966 | 0.889929 | 0.11 | 0.031 |
| Bi | 5 | 5 | Complement Bayes | 40296 | 4477 | 3820 | 657 | 85.324994 | 14.675006 | 0.706459 | 0.876169 | 1967 | 0.169803 | 379 | 0.830197 | 1853 | 0.123831 | 278 | 0.838448 | 0.876169 | 0.109 | 0.032 |
| Bi | 5 | 6 | Complement Bayes | 40296 | 4477 | 3824 | 653 | 85.41434 | 14.58566 | 0.708233 | 0.884633 | 1986 | 0.176523 | 394 | 0.823477 | 1838 | 0.115367 | 259 | 0.834454 | 0.884633 | 0.125 | 0.046 |
| Bi | 5 | 7 | Complement Bayes | 40296 | 4477 | 3816 | 661 | 85.235649 | 14.764351 | 0.704676 | 0.873051 | 1960 | 0.168459 | 376 | 0.831541 | 1856 | 0.126949 | 285 | 0.839041 | 0.873051 | 0.109 | 0.031 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bi | 5 | 8 | Complement Bayes | 40296 | 4477 | 3798 | 679 | 84.833594 | 15.166406 | 0.696631 | 0.870379 | 1954 | 0.173835 | 388 | 0.826165 | 1844 | 0.129621 | 291 | 0.83433 | 0.870379 | 0.156 | 0.438 |
| Bi | 5 | 9 | Complement Bayes | 40296 | 4477 | 3791 | 686 | 84.677239 | 15.322761 | 0.69351 | 0.865033 | 1942 | 0.171595 | 383 | 0.828405 | 1849 | 0.134967 | 303 | 0.835269 | 0.865033 | 0.5 | 0.016 |
| Bi | 5 | 10 | Complement Bayes | 40296 | 4477 | 3869 | 608 | 86.419477 | 13.580523 | 0.728362 | 0.880624 | 1977 | 0.15233 | 340 | 0.84767 | 1892 | 0.119376 | 268 | 0.853259 | 0.880624 | 0.141 | 0.047 |
| Bi | 6 | 1 | Complement Bayes | 40295 | 4478 | 3853 | 625 | 86.042876 | 13.957124 | 0.720815 | 0.887751 | 1993 | 0.16704 | 373 | 0.83296 | 1860 | 0.112249 | 252 | 0.84235 | 0.887751 | 0.141 | 0.015 |
| Bi | 6 | 2 | Complement Bayes | 40295 | 4478 | 3827 | 651 | 85.46226 | 14.53774 | 0.709211 | 0.875724 | 1966 | 0.166592 | 372 | 0.833408 | 1861 | 0.124276 | 279 | 0.84089 | 0.875724 | 0.125 | 0.391 |
| Bi | 6 | 3 | Complement Bayes | 40295 | 4478 | 3866 | 612 | 86.333184 | 13.666816 | 0.726635 | 0.88196 | 1980 | 0.155396 | 347 | 0.844604 | 1886 | 0.11804 | 265 | 0.850881 | 0.88196 | 0.109 | 0.032 |
| Bi | 6 | 4 | Complement Bayes | 40296 | 4477 | 3829 | 648 | 85.526022 | 14.473978 | 0.71049 | 0.87216 | 1958 | 0.161738 | 361 | 0.838262 | 1871 | 0.12784 | 287 | 0.844329 | 0.87216 | 0.172 | 0.047 |
| Bi | 6 | 5 | Complement Bayes | 40296 | 4477 | 3819 | 658 | 85.302658 | 14.697342 | 0.706009 | 0.877506 | 1970 | 0.171595 | 383 | 0.828405 | 1849 | 0.122494 | 275 | 0.837229 | 0.877506 | 0.187 | 0.016 |
| Bi | 6 | 6 | Complement Bayes | 40296 | 4477 | 3836 | 641 | 85.682377 | 14.317623 | 0.713599 | 0.885078 | 1987 | 0.171595 | 383 | 0.828405 | 1849 | 0.114922 | 258 | 0.838397 | 0.885078 | 0.125 | 0.031 |
| Bi | 6 | 7 | Complement Bayes | 40296 | 4477 | 3816 | 661 | 85.235649 | 14.764351 | 0.704665 | 0.879287 | 1974 | 0.174731 | 390 | 0.825269 | 1842 | 0.120713 | 271 | 0.835025 | 0.879287 | 0.11 | 0.031 |
| Bi | 6 | 8 | Complement Bayes | 40296 | 4477 | 3826 | 651 | 85.459013 | 14.540987 | 0.709128 | 0.884633 | 1986 | 0.175627 | 392 | 0.824373 | 1840 | 0.115367 | 259 | 0.835156 | 0.884633 | 0.5 | 0.015 |
| Bi | 6 | 9 | Complement Bayes | 40296 | 4477 | 3827 | 650 | 85.481349 | 14.518651 | 0.709585 | 0.878396 | 1972 | 0.168907 | 377 | 0.831093 | 1855 | 0.121604 | 273 | 0.839506 | 0.878396 | 0.125 | 0.031 |
| Bi | 6 | 10 | Complement Bayes | 40296 | 4477 | 3829 | 648 | 85.526022 | 14.473978 | 0.710502 | 0.867647 | 1947 | 0.157188 | 351 | 0.842812 | 1882 | 0.132353 | 297 | 0.847258 | 0.867647 | 0.109 | 0.016 |
| Bi | 7 | 1 | Complement Bayes | 40295 | 4478 | 3829 | 649 | 85.506923 | 14.493077 | 0.710108 | 0.873497 | 1961 | 0.163457 | 365 | 0.836543 | 1868 | 0.126503 | 284 | 0.843078 | 0.873497 | 0.172 | 0.016 |
| Bi | 7 | 2 | Complement Bayes | 40295 | 4478 | 3804 | 674 | 84.948638 | 15.051362 | 0.698931 | 0.874388 | 1963 | 0.175549 | 392 | 0.824451 | 1841 | 0.125612 | 282 | 0.833546 | 0.874388 | 0.109 | 0.047 |
| Bi | 7 | 3 | Complement Bayes | 40295 | 4478 | 3866 | 612 | 86.333184 | 13.666816 | 0.726626 | 0.887751 | 1993 | 0.161218 | 360 | 0.838782 | 1873 | 0.112249 | 252 | 0.847004 | 0.887751 | 0.125 | 0.046 |
| Bi | 7 | 4 | Complement Bayes | 40296 | 4477 | 3812 | 665 | 85.146303 | 14.853697 | 0.702899 | 0.868984 | 1950 | 0.166144 | 371 | 0.833856 | 1862 | 0.131016 | 294 | 0.840155 | 0.868984 | 0.109 | 0.031 |
| Bi | 7 | 5 | Complement Bayes | 40296 | 4477 | 3839 | 638 | 85.749386 | 14.250614 | 0.714952 | 0.877951 | 1971 | 0.163082 | 364 | 0.836918 | 1868 | 0.122049 | 274 | 0.844111 | 0.877951 | 0.141 | 0.031 |
| Bi | 7 | 6 | Complement Bayes | 40296 | 4477 | 3840 | 637 | 85.771722 | 14.228278 | 0.715382 | 0.888196 | 1994 | 0.172939 | 386 | 0.827061 | 1846 | 0.111804 | 251 | 0.837815 | 0.888196 | 0.14 | 0.031 |
| Bi | 7 | 7 | Complement Bayes | 40296 | 4477 | 3840 | 637 | 85.771722 | 14.228278 | 0.715395 | 0.880178 | 1976 | 0.164875 | 368 | 0.835125 | 1864 | 0.119822 | 269 | 0.843003 | 0.880178 | 0.11 | 0.047 |
| Bi | 7 | 8 | Complement Bayes | 40296 | 4477 | 3827 | 650 | 85.481349 | 14.518651 | 0.709602 | 0.868597 | 1950 | 0.15905 | 355 | 0.84095 | 1877 | 0.131403 | 295 | 0.845987 | 0.868597 | 0.11 | 0.015 |
| Bi | 7 | 9 | Complement Bayes | 40296 | 4477 | 3851 | 626 | 86.017422 | 13.982578 | 0.720298 | 0.889978 | 1998 | 0.169803 | 379 | 0.830197 | 1853 | 0.110022 | 247 | 0.840555 | 0.889978 | 0.125 | 0.031 |
| Bi | 7 | 10 | Complement Bayes | 40296 | 4477 | 3799 | 678 | 84.85593 | 15.14407 | 0.697077 | 0.871269 | 1956 | 0.174283 | 389 | 0.825717 | 1843 | 0.128731 | 289 | 0.834115 | 0.871269 | 0.125 | 0.031 |
| Bi | 8 | 1 | Complement Bayes | 40295 | 4478 | 3832 | 646 | 85.573917 | 14.426083 | 0.711427 | 0.887751 | 1993 | 0.176444 | 394 | 0.823556 | 1839 | 0.112249 | 252 | 0.834939 | 0.887751 | 0.188 | 0.015 |
| Bi | 8 | 2 | Complement Bayes | 40295 | 4478 | 3844 | 634 | 85.841894 | 14.158106 | 0.716808 | 0.87706 | 1969 | 0.160322 | 358 | 0.839678 | 1875 | 0.12294 | 276 | 0.846154 | 0.87706 | 0.125 | 0.032 |
| Bi | 8 | 3 | Complement Bayes | 40295 | 4478 | 3833 | 645 | 85.596248 | 14.403752 | 0.711894 | 0.875278 | 1965 | 0.163457 | 365 | 0.836543 | 1868 | 0.124722 | 280 | 0.843348 | 0.875278 | 0.125 | 0.391 |
| Bi | 8 | 4 | Complement Bayes | 40296 | 4477 | 3822 | 655 | 85.369667 | 14.630333 | 0.707348 | 0.879287 | 1974 | 0.172043 | 384 | 0.827957 | 1848 | 0.120713 | 271 | 0.83715 | 0.879287 | 0.11 | 0.031 |
| Bi | 8 | 5 | Complement Bayes | 40296 | 4477 | 3849 | 628 | 85.97275 | 14.02725 | 0.719404 | 0.889978 | 1998 | 0.170699 | 381 | 0.829301 | 1851 | 0.110022 | 247 | 0.839849 | 0.889978 | 0.141 | 0.015 |
| Bi | 8 | 6 | Complement Bayes | 40296 | 4477 | 3811 | 666 | 85.123967 | 14.876033 | 0.702436 | 0.875278 | 1965 | 0.172939 | 386 | 0.827061 | 1846 | 0.124722 | 280 | 0.835815 | 0.875278 | 0.11 | 0.047 |
| Bi | 8 | 7 | Complement Bayes | 40296 | 4477 | 3809 | 668 | 85.079294 | 14.920706 | 0.701553 | 0.868597 | 1950 | 0.167115 | 373 | 0.832885 | 1859 | 0.131403 | 295 | 0.839432 | 0.868597 | 0.125 | 0.032 |
| Bi | 8 | 8 | Complement Bayes | 40296 | 4477 | 3829 | 648 | 85.526022 | 14.473978 | 0.710483 | 0.876169 | 1967 | 0.165771 | 370 | 0.834229 | 1862 | 0.123831 | 278 | 0.841677 | 0.876169 | 0.5 | 0.015 |
| Bi | 8 | 9 | Complement Bayes | 40296 | 4477 | 3835 | 642 | 85.66004 | 14.33996 | 0.713169 | 0.874388 | 1963 | 0.16129 | 360 | 0.83871 | 1872 | 0.125612 | 282 | 0.845028 | 0.874388 | 0.125 | 0.031 |
| Bi | 8 | 10 | Complement Bayes | 40296 | 4477 | 3853 | 624 | 86.062095 | 13.937905 | 0.721208 | 0.884135 | 1984 | 0.163009 | 364 | 0.836991 | 1869 | 0.115865 | 260 | 0.844974 | 0.884135 | 0.172 | 0.047 |
| Bi | 9 | 1 | Complement Bayes | 40295 | 4478 | 3836 | 642 | 85.663243 | 14.336757 | 0.713232 | 0.87706 | 1969 | 0.163905 | 366 | 0.836095 | 1867 | 0.12294 | 276 | 0.843255 | 0.87706 | 0.157 | 0.343 |
| Bi | 9 | 2 | Complement Bayes | 40295 | 4478 | 3832 | 646 | 85.573917 | 14.426083 | 0.711421 | 0.891759 | 2002 | 0.180475 | 403 | 0.819525 | 1830 | 0.108241 | 243 | 0.832432 | 0.891759 | 0.11 | 0.031 |
| Bi | 9 | 3 | Complement Bayes | 40295 | 4478 | 3814 | 664 | 85.171952 | 14.828048 | 0.703402 | 0.873942 | 1962 | 0.170622 | 381 | 0.829378 | 1852 | 0.126058 | 283 | 0.837388 | 0.873942 | 0.125 | 0.031 |
| Bi | 9 | 4 | Complement Bayes | 40296 | 4477 | 3820 | 657 | 85.324994 | 14.675006 | 0.706462 | 0.874388 | 1963 | 0.168011 | 375 | 0.831989 | 1857 | 0.125612 | 282 | 0.839607 | 0.874388 | 0.141 | 0.031 |
| Bi | 9 | 5 | Complement Bayes | 40296 | 4477 | 3842 | 635 | 85.816395 | 14.183605 | 0.716297 | 0.875724 | 1966 | 0.159498 | 356 | 0.840502 | 1876 | 0.124276 | 279 | 0.846684 | 0.875724 | 0.109 | 0.047 |
| Bi | 9 | 6 | Complement Bayes | 40296 | 4477 | 3872 | 605 | 86.486486 | 13.513514 | 0.729706 | 0.878842 | 1973 | 0.149194 | 333 | 0.850806 | 1899 | 0.121158 | 272 | 0.855594 | 0.878842 | 0.125 | 0.031 |
| Bi | 9 | 7 | Complement Bayes | 40296 | 4477 | 3847 | 630 | 85.928077 | 14.071923 | 0.718526 | 0.880178 | 1976 | 0.161738 | 361 | 0.838262 | 1871 | 0.119822 | 269 | 0.845528 | 0.880178 | 0.125 | 0.047 |
| Bi | 9 | 8 | Complement Bayes | 40296 | 4477 | 3829 | 648 | 85.526022 | 14.473978 | 0.710473 | 0.882405 | 1981 | 0.172043 | 384 | 0.827957 | 1848 | 0.117595 | 264 | 0.837632 | 0.882405 | 0.141 | 0.015 |
| Bi | 9 | 9 | Complement Bayes | 40296 | 4477 | 3812 | 665 | 85.146303 | 14.853697 | 0.702879 | 0.877506 | 1970 | 0.174731 | 390 | 0.825269 | 1842 | 0.122494 | 275 | 0.834746 | 0.877506 | 0.109 | 0.031 |
| Bi | 9 | 10 | Complement Bayes | 40296 | 4477 | 3825 | 652 | 85.436676 | 14.563324 | 0.7087 | 0.877005 | 1968 | 0.168383 | 376 | 0.831617 | 1857 | 0.122995 | 276 | 0.83959 | 0.877005 | 0.11 | 0.031 |
| Bi | 10 | 1 | Complement Bayes | 40295 | 4478 | 3822 | 656 | 85.350603 | 14.649397 | 0.706985 | 0.869488 | 1952 | 0.162562 | 363 | 0.837438 | 1870 | 0.130512 | 293 | 0.843197 | 0.869488 | 0.484 | 0.016 |
| Bi | 10 | 2 | Complement Bayes | 40295 | 4478 | 3856 | 622 | 86.10987 | 13.89013 | 0.722148 | 0.893096 | 2005 | 0.17107 | 382 | 0.82893 | 1851 | 0.106904 | 240 | 0.839966 | 0.893096 | 0.109 | 0.032 |
| Bi | 10 | 3 | Complement Bayes | 40295 | 4478 | 3821 | 657 | 85.328272 | 14.671728 | 0.706542 | 0.867261 | 1947 | 0.16077 | 359 | 0.83923 | 1874 | 0.132739 | 298 | 0.844319 | 0.867261 | 0.157 | 0.031 |
| Bi | 10 | 4 | Complement Bayes | 40296 | 4477 | 3865 | 612 | 86.330132 | 13.669868 | 0.726583 | 0.874388 | 1963 | 0.147849 | 330 | 0.852151 | 1902 | 0.125612 | 282 | 0.856084 | 0.874388 | 0.109 | 0.047 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bi | 10 | 5 | Complement Bayes | 40296 | 4477 | 3813 | 664 | 85.16864 | 14.83136 | 0.703325 | 0.878396 | 1972 | 0.175179 | 391 | 0.824821 | 1841 | 0.121604 | 273 | 0.834532 | 0.878396 | 0.109 | 0.031 |
| Bi | 10 | 6 | Complement Bayes | 40296 | 4477 | 3827 | 650 | 85.481349 | 14.518651 | 0.709578 | 0.882851 | 1982 | 0.173387 | 387 | 0.826613 | 1845 | 0.117149 | 263 | 0.83664 | 0.882851 | 0.109 | 0.032 |
| Bi | 10 | 7 | Complement Bayes | 40296 | 4477 | 3836 | 641 | 85.682377 | 14.317623 | 0.713601 | 0.883742 | 1984 | 0.170251 | 380 | 0.829749 | 1852 | 0.116258 | 261 | 0.839255 | 0.883742 | 0.11 | 0.031 |
| Bi | 10 | 8 | Complement Bayes | 40296 | 4477 | 3791 | 686 | 84.677239 | 15.322761 | 0.693507 | 0.866815 | 1946 | 0.173387 | 387 | 0.826613 | 1845 | 0.133185 | 299 | 0.834119 | 0.866815 | 0.172 | 0.015 |
| Bi | 10 | 9 | Complement Bayes | 40296 | 4477 | 3850 | 627 | 85.995086 | 14.004914 | 0.719866 | 0.880624 | 1977 | 0.160842 | 359 | 0.839158 | 1873 | 0.119376 | 268 | 0.846318 | 0.880624 | 0.125 | 0.047 |
| Bi | 10 | 10 | Complement Bayes | 40296 | 4477 | 3860 | 617 | 86.21845 | 13.78155 | 0.724333 | 0.888146 | 1993 | 0.163905 | 366 | 0.836095 | 1867 | 0.111854 | 251 | 0.84485 | 0.888146 | 0.187 | 0.016 |
| Tri | 1 | 1 | Complement Bayes | 40295 | 4478 | 3597 | 881 | 80.326038 | 19.673962 | 0.606298 | 0.907795 | 2038 | 0.301836 | 674 | 0.698164 | 1559 | 0.092205 | 207 | 0.751475 | 0.907795 | 0.156 | 0.031 |
| Tri | 1 | 2 | Complement Bayes | 40295 | 4478 | 3536 | 942 | 78.963823 | 21.036177 | 0.579029 | 0.898441 | 2017 | 0.319749 | 714 | 0.680251 | 1519 | 0.101559 | 228 | 0.738557 | 0.898441 | 0.172 | 0.031 |
| Tri | 1 | 3 | Complement Bayes | 40295 | 4478 | 3564 | 914 | 79.589102 | 20.410898 | 0.59155 | 0.900668 | 2022 | 0.309449 | 691 | 0.690551 | 1542 | 0.099332 | 223 | 0.7453 | 0.900668 | 0.125 | 0.032 |
| Tri | 1 | 4 | Complement Bayes | 40296 | 4477 | 3586 | 891 | 80.09828 | 19.90172 | 0.601718 | 0.906904 | 2036 | 0.305556 | 682 | 0.694444 | 1550 | 0.093096 | 209 | 0.74908 | 0.906904 | 0.141 | 0.015 |
| Tri | 1 | 5 | Complement Bayes | 40296 | 4477 | 3597 | 880 | 80.34398 | 19.65602 | 0.606642 | 0.906459 | 2035 | 0.300179 | 670 | 0.699821 | 1562 | 0.093541 | 210 | 0.752311 | 0.906459 | 0.546 | 0.016 |
| Tri | 1 | 6 | Complement Bayes | 40296 | 4477 | 3611 | 866 | 80.65669 | 19.34331 | 0.612883 | 0.916704 | 2058 | 0.304211 | 679 | 0.695789 | 1553 | 0.083296 | 187 | 0.751918 | 0.916704 | 0.125 | 0.031 |
| Tri | 1 | 7 | Complement Bayes | 40296 | 4477 | 3583 | 894 | 80.031271 | 19.968729 | 0.600375 | 0.906904 | 2036 | 0.3069 | 685 | 0.6931 | 1547 | 0.093096 | 209 | 0.748254 | 0.906904 | 0.156 | 0.031 |
| Tri | 1 | 8 | Complement Bayes | 40296 | 4477 | 3623 | 854 | 80.924726 | 19.075274 | 0.61827 | 0.909131 | 2041 | 0.291219 | 650 | 0.708781 | 1582 | 0.090869 | 204 | 0.758454 | 0.909131 | 0.234 | 0.031 |
| Tri | 1 | 9 | Complement Bayes | 40296 | 4477 | 3600 | 877 | 80.41099 | 19.58901 | 0.607964 | 0.915367 | 2055 | 0.307796 | 687 | 0.692204 | 1545 | 0.084633 | 190 | 0.749453 | 0.915367 | 0.204 | 0.031 |
| Tri | 1 | 10 | Complement Bayes | 40296 | 4477 | 3580 | 897 | 79.964262 | 20.035738 | 0.599054 | 0.915775 | 2055 | 0.317062 | 708 | 0.682938 | 1525 | 0.084225 | 189 | 0.743757 | 0.915775 | 0.125 | 0.031 |
| Tri | 2 | 1 | Complement Bayes | 40295 | 4478 | 3563 | 915 | 79.566771 | 20.433229 | 0.591096 | 0.903786 | 2029 | 0.313032 | 699 | 0.686968 | 1534 | 0.096214 | 216 | 0.743768 | 0.903786 | 0.141 | 0.047 |
| Tri | 2 | 2 | Complement Bayes | 40295 | 4478 | 3603 | 875 | 80.460027 | 19.539973 | 0.608972 | 0.912695 | 2049 | 0.304075 | 679 | 0.695925 | 1554 | 0.087305 | 196 | 0.7511 | 0.912695 | 0.14 | 0.032 |
| Tri | 2 | 3 | Complement Bayes | 40295 | 4478 | 3597 | 881 | 80.326038 | 19.673962 | 0.606313 | 0.900668 | 2022 | 0.294671 | 658 | 0.705329 | 1575 | 0.099332 | 223 | 0.754478 | 0.900668 | 0.14 | 0.047 |
| Tri | 2 | 4 | Complement Bayes | 40296 | 4477 | 3600 | 877 | 80.41099 | 19.58901 | 0.608012 | 0.910873 | 2044 | 0.30318 | 677 | 0.69682 | 1556 | 0.089127 | 200 | 0.751194 | 0.910873 | 0.156 | 0.031 |
| Tri | 2 | 5 | Complement Bayes | 40296 | 4477 | 3606 | 871 | 80.545008 | 19.454992 | 0.610644 | 0.917595 | 2060 | 0.307348 | 686 | 0.692652 | 1546 | 0.082405 | 185 | 0.750182 | 0.917595 | 0.141 | 0.047 |
| Tri | 2 | 6 | Complement Bayes | 40296 | 4477 | 3555 | 922 | 79.405852 | 20.594148 | 0.587854 | 0.902895 | 2027 | 0.315412 | 704 | 0.684588 | 1528 | 0.097105 | 218 | 0.742219 | 0.902895 | 0.172 | 0.032 |
| Tri | 2 | 7 | Complement Bayes | 40296 | 4477 | 3578 | 899 | 79.919589 | 20.080411 | 0.598143 | 0.904677 | 2031 | 0.3069 | 685 | 0.6931 | 1547 | 0.095323 | 214 | 0.747791 | 0.904677 | 0.171 | 0.032 |
| Tri | 2 | 8 | Complement Bayes | 40296 | 4477 | 3588 | 889 | 80.142953 | 19.857047 | 0.6026 | 0.912695 | 2049 | 0.310484 | 693 | 0.689516 | 1539 | 0.087305 | 196 | 0.747265 | 0.912695 | 0.156 | 0.032 |
| Tri | 2 | 9 | Complement Bayes | 40296 | 4477 | 3587 | 890 | 80.120616 | 19.879384 | 0.602186 | 0.897996 | 2016 | 0.296147 | 661 | 0.703853 | 1571 | 0.102004 | 229 | 0.753082 | 0.897996 | 0.156 | 0.031 |
| Tri | 2 | 10 | Complement Bayes | 40296 | 4477 | 3599 | 878 | 80.388653 | 19.611347 | 0.607513 | 0.916704 | 2058 | 0.309588 | 691 | 0.690412 | 1541 | 0.083296 | 187 | 0.748636 | 0.916704 | 0.172 | 0.391 |
| Tri | 3 | 1 | Complement Bayes | 40295 | 4478 | 3598 | 880 | 80.34837 | 19.65163 | 0.606759 | 0.901114 | 2023 | 0.294671 | 658 | 0.705329 | 1575 | 0.098886 | 222 | 0.754569 | 0.901114 | 0.563 | 0.015 |
| Tri | 3 | 2 | Complement Bayes | 40295 | 4478 | 3604 | 874 | 80.482358 | 19.517642 | 0.609414 | 0.914922 | 2054 | 0.305867 | 683 | 0.694133 | 1550 | 0.085078 | 191 | 0.750457 | 0.914922 | 0.156 | 0.016 |
| Tri | 3 | 3 | Complement Bayes | 40295 | 4478 | 3605 | 873 | 80.50469 | 19.49531 | 0.60988 | 0.906013 | 2034 | 0.296462 | 662 | 0.703538 | 1571 | 0.093987 | 211 | 0.754451 | 0.906013 | 0.266 | 0.078 |
| Tri | 3 | 4 | Complement Bayes | 40296 | 4477 | 3566 | 911 | 79.651552 | 20.348448 | 0.592779 | 0.902004 | 2025 | 0.309588 | 691 | 0.690412 | 1541 | 0.097996 | 220 | 0.745582 | 0.902004 | 0.188 | 0.031 |
| Tri | 3 | 5 | Complement Bayes | 40296 | 4477 | 3585 | 892 | 80.075944 | 19.924056 | 0.601231 | 0.923831 | 2074 | 0.323029 | 721 | 0.676971 | 1511 | 0.076169 | 171 | 0.742039 | 0.923831 | 0.172 | 0.031 |
| Tri | 3 | 6 | Complement Bayes | 40296 | 4477 | 3607 | 870 | 80.567344 | 19.432656 | 0.611113 | 0.908241 | 2039 | 0.297491 | 664 | 0.702509 | 1568 | 0.091759 | 206 | 0.754347 | 0.908241 | 0.297 | 0.047 |
| Tri | 3 | 7 | Complement Bayes | 40296 | 4477 | 3564 | 913 | 79.60688 | 20.39312 | 0.591889 | 0.899777 | 2020 | 0.308244 | 688 | 0.691756 | 1544 | 0.100223 | 225 | 0.745938 | 0.899777 | 0.172 | 0.015 |
| Tri | 3 | 8 | Complement Bayes | 40296 | 4477 | 3585 | 892 | 80.075944 | 19.924056 | 0.601262 | 0.910468 | 2044 | 0.309588 | 691 | 0.690412 | 1541 | 0.089532 | 201 | 0.747349 | 0.910468 | 0.156 | 0.015 |
| Tri | 3 | 9 | Complement Bayes | 40296 | 4477 | 3536 | 941 | 78.981461 | 21.018539 | 0.579354 | 0.901114 | 2023 | 0.322133 | 719 | 0.677867 | 1513 | 0.098886 | 222 | 0.737783 | 0.901114 | 0.171 | 0.016 |
| Tri | 3 | 10 | Complement Bayes | 40296 | 4477 | 3616 | 861 | 80.768372 | 19.231628 | 0.615158 | 0.917558 | 2059 | 0.302732 | 676 | 0.697268 | 1557 | 0.082442 | 185 | 0.752834 | 0.917558 | 0.156 | 0.032 |
| Tri | 4 | 1 | Complement Bayes | 40295 | 4478 | 3554 | 924 | 79.365788 | 20.634212 | 0.587052 | 0.911804 | 2047 | 0.325123 | 726 | 0.674877 | 1507 | 0.088196 | 198 | 0.73819 | 0.911804 | 0.172 | 0.032 |
| Tri | 4 | 2 | Complement Bayes | 40295 | 4478 | 3608 | 870 | 80.571684 | 19.428316 | 0.611208 | 0.91314 | 2050 | 0.302284 | 675 | 0.697716 | 1558 | 0.08686 | 195 | 0.752294 | 0.91314 | 0.203 | 0.031 |
| Tri | 4 | 3 | Complement Bayes | 40295 | 4478 | 3610 | 868 | 80.616347 | 19.383653 | 0.612096 | 0.916258 | 2057 | 0.304523 | 680 | 0.695477 | 1553 | 0.083742 | 188 | 0.751553 | 0.916258 | 0.157 | 0.047 |
| Tri | 4 | 4 | Complement Bayes | 40296 | 4477 | 3618 | 859 | 80.813044 | 19.186956 | 0.616026 | 0.912249 | 2048 | 0.296595 | 662 | 0.703405 | 1570 | 0.087751 | 197 | 0.75572 | 0.912249 | 0.172 | 0.062 |
| Tri | 4 | 5 | Complement Bayes | 40296 | 4477 | 3572 | 905 | 79.785571 | 20.214429 | 0.595477 | 0.896659 | 2013 | 0.301523 | 673 | 0.698477 | 1559 | 0.103341 | 232 | 0.749442 | 0.896659 | 0.187 | 0.031 |
| Tri | 4 | 6 | Complement Bayes | 40296 | 4477 | 3596 | 881 | 80.321644 | 19.678356 | 0.606182 | 0.911804 | 2047 | 0.306004 | 683 | 0.693996 | 1549 | 0.088196 | 198 | 0.749817 | 0.911804 | 0.188 | 0.031 |
| Tri | 4 | 7 | Complement Bayes | 40296 | 4477 | 3587 | 890 | 80.120616 | 19.879384 | 0.602167 | 0.906013 | 2034 | 0.304211 | 679 | 0.695789 | 1553 | 0.093987 | 211 | 0.749724 | 0.906013 | 0.188 | 0.015 |
| Tri | 4 | 8 | Complement Bayes | 40296 | 4477 | 3540 | 937 | 79.070806 | 20.929194 | 0.581165 | 0.89265 | 2004 | 0.311828 | 696 | 0.688172 | 1536 | 0.10735 | 241 | 0.742222 | 0.89265 | 0.187 | 0.032 |
| Tri | 4 | 9 | Complement Bayes | 40296 | 4477 | 3559 | 918 | 79.495198 | 20.504802 | 0.58964 | 0.904677 | 2031 | 0.315412 | 704 | 0.684588 | 1528 | 0.095323 | 214 | 0.742596 | 0.904677 | 0.172 | 0.031 |
| Tri | 4 | 10 | Complement Bayes | 40296 | 4477 | 3633 | 844 | 81.14809 | 18.85191 | 0.622764 | 0.917112 | 2058 | 0.294671 | 658 | 0.705329 | 1575 | 0.082888 | 186 | 0.757732 | 0.917112 | 0.531 | 0.016 |
| Tri | 5 | 1 | Complement Bayes | 40295 | 4478 | 3618 | 860 | 80.794998 | 19.205002 | 0.615691 | 0.908241 | 2039 | 0.29288 | 654 | 0.70712 | 1579 | 0.091759 | 206 | 0.757148 | 0.908241 | 0.156 | 0.016 |

| | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tri | 5 | 2 | Complement Bayes | 40295 | 4478 | 3597 | 881 | 80.326038 | 19.673962 | 0.606301 | 0.906459 | 2035 | 0.300493 | 671 | 0.699507 | 1562 | 0.093541 | 210 | 0.752033 | 0.906459 | 0.171 | 0.016 |
| Tri | 5 | 3 | Complement Bayes | 40295 | 4478 | 3607 | 871 | 80.549352 | 19.450648 | 0.610749 | 0.918486 | 2062 | 0.308106 | 688 | 0.691894 | 1545 | 0.081514 | 183 | 0.749818 | 0.918486 | 0.171 | 0.016 |
| Tri | 5 | 4 | Complement Bayes | 40296 | 4477 | 3569 | 908 | 79.718562 | 20.281438 | 0.594131 | 0.916667 | 2057 | 0.322884 | 721 | 0.677116 | 1512 | 0.083333 | 187 | 0.740461 | 0.916667 | 0.172 | 0.015 |
| Tri | 5 | 5 | Complement Bayes | 40296 | 4477 | 3547 | 930 | 79.227161 | 20.772839 | 0.58427 | 0.904232 | 2030 | 0.320341 | 715 | 0.679659 | 1517 | 0.095768 | 215 | 0.739526 | 0.904232 | 0.187 | 0.016 |
| Tri | 5 | 6 | Complement Bayes | 40296 | 4477 | 3612 | 865 | 80.679026 | 19.320974 | 0.613335 | 0.914922 | 2054 | 0.301971 | 674 | 0.698029 | 1558 | 0.085078 | 191 | 0.752933 | 0.914922 | 0.187 | 0.016 |
| Tri | 5 | 7 | Complement Bayes | 40296 | 4477 | 3599 | 878 | 80.388653 | 19.611347 | 0.607557 | 0.89755 | 2015 | 0.290323 | 648 | 0.709677 | 1584 | 0.10245 | 230 | 0.756665 | 0.89755 | 0.188 | 0.031 |
| Tri | 5 | 8 | Complement Bayes | 40296 | 4477 | 3576 | 901 | 79.874916 | 20.125084 | 0.597236 | 0.909577 | 2042 | 0.312724 | 698 | 0.687276 | 1534 | 0.090423 | 203 | 0.745255 | 0.909577 | 0.172 | 0.032 |
| Tri | 5 | 9 | Complement Bayes | 40296 | 4477 | 3533 | 944 | 78.914452 | 21.085548 | 0.578019 | 0.897996 | 2016 | 0.320341 | 715 | 0.679659 | 1517 | 0.102004 | 229 | 0.738191 | 0.897996 | 0.141 | 0.015 |
| Tri | 5 | 10 | Complement Bayes | 40296 | 4477 | 3624 | 853 | 80.947063 | 19.052937 | 0.61872 | 0.908241 | 2039 | 0.289875 | 647 | 0.710125 | 1585 | 0.091759 | 206 | 0.759121 | 0.908241 | 0.156 | 0.032 |
| Tri | 6 | 1 | Complement Bayes | 40295 | 4478 | 3596 | 882 | 80.303707 | 19.696293 | 0.605868 | 0.899332 | 2019 | 0.293775 | 656 | 0.706225 | 1577 | 0.100668 | 226 | 0.754766 | 0.899332 | 0.14 | 0.016 |
| Tri | 6 | 2 | Complement Bayes | 40295 | 4478 | 3576 | 902 | 79.857079 | 20.142921 | 0.596927 | 0.896659 | 2013 | 0.300045 | 670 | 0.699955 | 1563 | 0.103341 | 232 | 0.75028 | 0.896659 | 0.125 | 0.031 |
| Tri | 6 | 3 | Complement Bayes | 40295 | 4478 | 3579 | 899 | 79.924073 | 20.075927 | 0.598226 | 0.916704 | 2058 | 0.318854 | 712 | 0.681146 | 1521 | 0.083296 | 187 | 0.74296 | 0.916704 | 0.25 | 0.031 |
| Tri | 6 | 4 | Complement Bayes | 40296 | 4477 | 3620 | 857 | 80.857717 | 19.142283 | 0.616921 | 0.912249 | 2048 | 0.295699 | 660 | 0.704301 | 1572 | 0.087751 | 197 | 0.756278 | 0.912249 | 0.14 | 0.016 |
| Tri | 6 | 5 | Complement Bayes | 40296 | 4477 | 3524 | 953 | 78.713424 | 21.286576 | 0.573976 | 0.904232 | 2030 | 0.330645 | 738 | 0.669355 | 1494 | 0.095768 | 215 | 0.733382 | 0.904232 | 0.171 | 0.016 |
| Tri | 6 | 6 | Complement Bayes | 40296 | 4477 | 3612 | 865 | 80.679026 | 19.320974 | 0.613342 | 0.911804 | 2047 | 0.298835 | 667 | 0.701165 | 1565 | 0.088196 | 198 | 0.754237 | 0.911804 | 0.141 | 0.047 |
| Tri | 6 | 7 | Complement Bayes | 40296 | 4477 | 3587 | 890 | 80.120616 | 19.879384 | 0.60215 | 0.913586 | 2051 | 0.311828 | 696 | 0.688172 | 1536 | 0.086414 | 194 | 0.746633 | 0.913586 | 0.172 | 0.016 |
| Tri | 6 | 8 | Complement Bayes | 40296 | 4477 | 3595 | 882 | 80.299308 | 19.700692 | 0.605719 | 0.918486 | 2062 | 0.313172 | 699 | 0.686828 | 1533 | 0.081514 | 183 | 0.746831 | 0.918486 | 0.172 | 0.016 |
| Tri | 6 | 9 | Complement Bayes | 40296 | 4477 | 3576 | 901 | 79.874916 | 20.125084 | 0.597244 | 0.906459 | 2035 | 0.309588 | 691 | 0.690412 | 1541 | 0.093541 | 210 | 0.746515 | 0.906459 | 0.172 | 0.047 |
| Tri | 6 | 10 | Complement Bayes | 40296 | 4477 | 3619 | 858 | 80.835381 | 19.164619 | 0.616518 | 0.9082 | 2038 | 0.291984 | 652 | 0.708016 | 1581 | 0.0918 | 206 | 0.757621 | 0.9082 | 0.171 | 0.016 |
| Tri | 7 | 1 | Complement Bayes | 40295 | 4478 | 3574 | 904 | 79.812416 | 20.187584 | 0.596007 | 0.908686 | 2040 | 0.313032 | 699 | 0.686968 | 1534 | 0.091314 | 205 | 0.744797 | 0.908686 | 0.14 | 0.032 |
| Tri | 7 | 2 | Complement Bayes | 40295 | 4478 | 3551 | 927 | 79.298794 | 20.701206 | 0.585715 | 0.909577 | 2042 | 0.324227 | 724 | 0.675773 | 1509 | 0.090423 | 203 | 0.73825 | 0.909577 | 0.156 | 0.031 |
| Tri | 7 | 3 | Complement Bayes | 40295 | 4478 | 3628 | 850 | 81.018312 | 18.981688 | 0.620137 | 0.921604 | 2069 | 0.301836 | 674 | 0.698164 | 1559 | 0.078396 | 176 | 0.754284 | 0.921604 | 0.172 | 0.015 |
| Tri | 7 | 4 | Complement Bayes | 40296 | 4477 | 3560 | 917 | 79.517534 | 20.482466 | 0.590137 | 0.900178 | 2020 | 0.310345 | 693 | 0.689655 | 1540 | 0.099822 | 224 | 0.744563 | 0.900178 | 0.563 | 0.016 |
| Tri | 7 | 5 | Complement Bayes | 40296 | 4477 | 3561 | 916 | 79.53987 | 20.46013 | 0.590522 | 0.910022 | 2043 | 0.319892 | 714 | 0.680108 | 1518 | 0.089978 | 202 | 0.741023 | 0.910022 | 0.359 | 0.032 |
| Tri | 7 | 6 | Complement Bayes | 40296 | 4477 | 3603 | 874 | 80.477999 | 19.522001 | 0.60931 | 0.913586 | 2051 | 0.304659 | 680 | 0.695341 | 1552 | 0.086414 | 194 | 0.751007 | 0.913586 | 0.735 | 0.015 |
| Tri | 7 | 7 | Complement Bayes | 40296 | 4477 | 3598 | 879 | 80.366317 | 19.633683 | 0.607099 | 0.902004 | 2025 | 0.295251 | 659 | 0.704749 | 1573 | 0.097996 | 220 | 0.754471 | 0.902004 | 0.156 | 0.032 |
| Tri | 7 | 8 | Complement Bayes | 40296 | 4477 | 3603 | 874 | 80.477999 | 19.522001 | 0.609332 | 0.904232 | 2030 | 0.295251 | 659 | 0.704749 | 1573 | 0.095768 | 215 | 0.754927 | 0.904232 | 0.219 | 0.031 |
| Tri | 7 | 9 | Complement Bayes | 40296 | 4477 | 3597 | 880 | 80.34398 | 19.65602 | 0.606635 | 0.909131 | 2041 | 0.302867 | 676 | 0.697133 | 1556 | 0.090869 | 204 | 0.751196 | 0.909131 | 0.172 | 0.016 |
| Tri | 7 | 10 | Complement Bayes | 40296 | 4477 | 3600 | 877 | 80.41099 | 19.58901 | 0.607992 | 0.902895 | 2027 | 0.295251 | 659 | 0.704749 | 1573 | 0.097105 | 218 | 0.754654 | 0.902895 | 0.172 | 0.031 |
| Tri | 8 | 1 | Complement Bayes | 40295 | 4478 | 3564 | 914 | 79.589102 | 20.410898 | 0.591544 | 0.903341 | 2028 | 0.312136 | 697 | 0.687864 | 1536 | 0.096659 | 217 | 0.74422 | 0.903341 | 0.156 | 0.016 |
| Tri | 8 | 2 | Complement Bayes | 40295 | 4478 | 3610 | 868 | 80.616347 | 19.383653 | 0.612107 | 0.910913 | 2045 | 0.299149 | 668 | 0.700851 | 1565 | 0.089087 | 200 | 0.753778 | 0.910913 | 0.187 | 0.032 |
| Tri | 8 | 3 | Complement Bayes | 40295 | 4478 | 3582 | 896 | 79.991067 | 20.008933 | 0.599577 | 0.912695 | 2049 | 0.31348 | 700 | 0.68652 | 1533 | 0.087305 | 196 | 0.745362 | 0.912695 | 0.156 | 0.015 |
| Tri | 8 | 4 | Complement Bayes | 40296 | 4477 | 3591 | 886 | 80.209962 | 19.790038 | 0.603952 | 0.908241 | 2039 | 0.304659 | 680 | 0.695341 | 1552 | 0.091759 | 206 | 0.749908 | 0.908241 | 0.656 | 0.016 |
| Tri | 8 | 5 | Complement Bayes | 40296 | 4477 | 3557 | 920 | 79.450525 | 20.549475 | 0.588736 | 0.908241 | 2039 | 0.319892 | 714 | 0.680108 | 1518 | 0.091759 | 206 | 0.740647 | 0.908241 | 0.156 | 0.031 |
| Tri | 8 | 6 | Complement Bayes | 40296 | 4477 | 3610 | 867 | 80.634353 | 19.365647 | 0.612442 | 0.914031 | 2052 | 0.301971 | 674 | 0.698029 | 1558 | 0.085969 | 193 | 0.752751 | 0.914031 | 0.172 | 0.015 |
| Tri | 8 | 7 | Complement Bayes | 40296 | 4477 | 3566 | 911 | 79.651552 | 20.348448 | 0.592789 | 0.89755 | 2015 | 0.305108 | 681 | 0.694892 | 1551 | 0.10245 | 230 | 0.747404 | 0.89755 | 0.172 | 0.031 |
| Tri | 8 | 8 | Complement Bayes | 40296 | 4477 | 3569 | 908 | 79.718562 | 20.281438 | 0.594106 | 0.908686 | 2040 | 0.314964 | 703 | 0.685036 | 1529 | 0.091314 | 205 | 0.743711 | 0.908686 | 0.172 | 0.015 |
| Tri | 8 | 9 | Complement Bayes | 40296 | 4477 | 3604 | 873 | 80.500335 | 19.499665 | 0.609767 | 0.909577 | 2042 | 0.300179 | 670 | 0.699821 | 1562 | 0.090423 | 203 | 0.75295 | 0.909577 | 0.172 | 0.016 |
| Tri | 8 | 10 | Complement Bayes | 40296 | 4477 | 3620 | 857 | 80.857717 | 19.142283 | 0.616966 | 0.907754 | 2037 | 0.291088 | 650 | 0.708912 | 1583 | 0.092246 | 207 | 0.758095 | 0.907754 | 0.172 | 0.016 |
| Tri | 9 | 1 | Complement Bayes | 40295 | 4478 | 3576 | 902 | 79.857079 | 20.142921 | 0.596915 | 0.90245 | 2026 | 0.305867 | 683 | 0.694133 | 1550 | 0.09755 | 219 | 0.747877 | 0.90245 | 0.141 | 0.031 |
| Tri | 9 | 2 | Complement Bayes | 40295 | 4478 | 3601 | 877 | 80.415364 | 19.584636 | 0.608069 | 0.916704 | 2058 | 0.309001 | 690 | 0.690999 | 1543 | 0.083296 | 187 | 0.748908 | 0.916704 | 0.172 | 0.015 |
| Tri | 9 | 3 | Complement Bayes | 40295 | 4478 | 3565 | 913 | 79.611434 | 20.388566 | 0.591991 | 0.903786 | 2029 | 0.312136 | 697 | 0.687864 | 1536 | 0.096214 | 216 | 0.744314 | 0.903786 | 0.172 | 0.015 |
| Tri | 9 | 4 | Complement Bayes | 40296 | 4477 | 3617 | 860 | 80.790708 | 19.209292 | 0.615602 | 0.901559 | 2024 | 0.28629 | 639 | 0.71371 | 1593 | 0.098441 | 221 | 0.760045 | 0.901559 | 0.188 | 0.016 |
| Tri | 9 | 5 | Complement Bayes | 40296 | 4477 | 3545 | 932 | 79.182488 | 20.817512 | 0.583386 | 0.899777 | 2020 | 0.316756 | 707 | 0.683244 | 1525 | 0.100223 | 225 | 0.740741 | 0.899777 | 0.172 | 0.016 |
| Tri | 9 | 6 | Complement Bayes | 40296 | 4477 | 3625 | 852 | 80.969399 | 19.030601 | 0.619159 | 0.911804 | 2047 | 0.293011 | 654 | 0.706989 | 1578 | 0.088196 | 198 | 0.757867 | 0.911804 | 0.156 | 0.032 |
| Tri | 9 | 7 | Complement Bayes | 40296 | 4477 | 3591 | 886 | 80.209962 | 19.790038 | 0.603941 | 0.91314 | 2050 | 0.309588 | 691 | 0.690412 | 1541 | 0.08686 | 195 | 0.747902 | 0.91314 | 0.157 | 0.031 |
| Tri | 9 | 8 | Complement Bayes | 40296 | 4477 | 3587 | 890 | 80.120616 | 19.879384 | 0.602151 | 0.91314 | 2050 | 0.31138 | 695 | 0.68862 | 1537 | 0.08686 | 195 | 0.746812 | 0.91314 | 0.141 | 0.031 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tri | 9 | 9 | Complement Bayes | 40296 | 4477 | 3579 | 898 | 79.941925 | 20.058075 | 0.59859 | 0.904677 | 2031 | 0.306452 | 684 | 0.693548 | 1548 | 0.095323 | 214 | 0.748066 | 0.904677 | 0.125 | 0.016 |
| Tri | 9 | 10 | Complement Bayes | 40296 | 4477 | 3562 | 915 | 79.562207 | 20.437793 | 0.591012 | 0.909982 | 2042 | 0.319301 | 713 | 0.680699 | 1520 | 0.090018 | 202 | 0.741198 | 0.909982 | 0.172 | 0.031 |
| Tri | 10 | 1 | Complement Bayes | 40295 | 4478 | 3559 | 919 | 79.477445 | 20.522555 | 0.589307 | 0.903786 | 2029 | 0.314823 | 703 | 0.685177 | 1530 | 0.096214 | 216 | 0.742679 | 0.903786 | 0.156 | 0.016 |
| Tri | 10 | 2 | Complement Bayes | 40295 | 4478 | 3610 | 868 | 80.616347 | 19.383653 | 0.612094 | 0.917149 | 2059 | 0.305419 | 682 | 0.694581 | 1551 | 0.082851 | 186 | 0.751186 | 0.917149 | 0.171 | 0.016 |
| Tri | 10 | 3 | Complement Bayes | 40295 | 4478 | 3564 | 914 | 79.589102 | 20.410898 | 0.591561 | 0.895768 | 2011 | 0.304523 | 680 | 0.695477 | 1553 | 0.104232 | 234 | 0.747306 | 0.895768 | 0.187 | 0.016 |
| Tri | 10 | 4 | Complement Bayes | 40296 | 4477 | 3553 | 924 | 79.361179 | 20.638821 | 0.586954 | 0.904677 | 2031 | 0.3181 | 710 | 0.6819 | 1522 | 0.095323 | 214 | 0.74097 | 0.904677 | 0.156 | 0.016 |
| Tri | 10 | 5 | Complement Bayes | 40296 | 4477 | 3559 | 918 | 79.495198 | 20.504802 | 0.589642 | 0.903786 | 2029 | 0.314516 | 702 | 0.685484 | 1530 | 0.096214 | 216 | 0.742951 | 0.903786 | 0.172 | 0.032 |
| Tri | 10 | 6 | Complement Bayes | 40296 | 4477 | 3617 | 860 | 80.790708 | 19.209292 | 0.615584 | 0.910022 | 2043 | 0.294803 | 658 | 0.705197 | 1574 | 0.089978 | 202 | 0.756387 | 0.910022 | 0.156 | 0.047 |
| Tri | 10 | 7 | Complement Bayes | 40296 | 4477 | 3601 | 876 | 80.433326 | 19.566674 | 0.608403 | 0.918931 | 2063 | 0.310932 | 694 | 0.689068 | 1538 | 0.081069 | 182 | 0.748277 | 0.918931 | 0.171 | 0.032 |
| Tri | 10 | 8 | Complement Bayes | 40296 | 4477 | 3577 | 900 | 79.897253 | 20.102747 | 0.597696 | 0.904232 | 2030 | 0.3069 | 685 | 0.6931 | 1547 | 0.095768 | 215 | 0.747698 | 0.904232 | 0.203 | 0.031 |
| Tri | 10 | 9 | Complement Bayes | 40296 | 4477 | 3633 | 844 | 81.14809 | 18.85191 | 0.622741 | 0.910913 | 2045 | 0.28853 | 644 | 0.71147 | 1588 | 0.089087 | 200 | 0.760506 | 0.910913 | 0.172 | 0.031 |
| Tri | 10 | 10 | Complement Bayes | 40296 | 4477 | 3607 | 870 | 80.567344 | 19.432656 | 0.611129 | 0.918449 | 2061 | 0.307658 | 687 | 0.692342 | 1546 | 0.081551 | 183 | 0.75 | 0.918449 | 0.172 | 0.015 |
| Bi | 1 | 1 | LibSVM | 40295 | 4478 | 4171 | 307 | 93.144261 | 6.855739 | 0.862895 | 0.916704 | 2058 | 0.053739 | 120 | 0.946261 | 2113 | 0.083296 | 187 | 0.944904 | 0.916704 | 480.848 | 35.698 |
| Bi | 1 | 2 | LibSVM | 40295 | 4478 | 4202 | 276 | 93.836534 | 6.163466 | 0.876738 | 0.925612 | 2078 | 0.048813 | 109 | 0.951187 | 2124 | 0.074388 | 167 | 0.95016 | 0.925612 | 535.279 | 41.386 |
| Bi | 1 | 3 | LibSVM | 40295 | 4478 | 4153 | 325 | 92.742296 | 7.257704 | 0.854855 | 0.914031 | 2052 | 0.059113 | 132 | 0.940887 | 2101 | 0.085969 | 193 | 0.93956 | 0.914031 | 584.023 | 39.635 |
| Bi | 1 | 4 | LibSVM | 40296 | 4477 | 4172 | 305 | 93.187402 | 6.812598 | 0.863756 | 0.920267 | 2066 | 0.056452 | 126 | 0.943548 | 2106 | 0.079733 | 179 | 0.942518 | 0.920267 | 571.681 | 51.275 |
| Bi | 1 | 5 | LibSVM | 40296 | 4477 | 4190 | 287 | 93.589457 | 6.410543 | 0.871798 | 0.922494 | 2071 | 0.050627 | 113 | 0.949373 | 2119 | 0.077506 | 174 | 0.94826 | 0.922494 | 534.154 | 30.981 |
| Bi | 1 | 6 | LibSVM | 40296 | 4477 | 4193 | 284 | 93.656466 | 6.343534 | 0.873139 | 0.922494 | 2071 | 0.049283 | 110 | 0.950717 | 2122 | 0.077506 | 174 | 0.949564 | 0.922494 | 453.788 | 32.59 |
| Bi | 1 | 7 | LibSVM | 40296 | 4477 | 4171 | 306 | 93.165066 | 6.834934 | 0.863312 | 0.917595 | 2060 | 0.054211 | 121 | 0.945789 | 2111 | 0.082405 | 185 | 0.944521 | 0.917595 | 519.609 | 43.698 |
| Bi | 1 | 8 | LibSVM | 40296 | 4477 | 4162 | 315 | 92.964038 | 7.035962 | 0.859293 | 0.91314 | 2050 | 0.053763 | 120 | 0.946237 | 2112 | 0.08686 | 195 | 0.9447 | 0.91314 | 608.864 | 55.024 |
| Bi | 1 | 9 | LibSVM | 40296 | 4477 | 4212 | 265 | 94.080858 | 5.919142 | 0.881627 | 0.925167 | 2077 | 0.043459 | 97 | 0.956541 | 2135 | 0.074833 | 168 | 0.955382 | 0.925167 | 618.19 | 39.308 |
| Bi | 1 | 10 | LibSVM | 40296 | 4477 | 4195 | 282 | 93.701139 | 6.298861 | 0.87403 | 0.923797 | 2073 | 0.049709 | 111 | 0.950291 | 2122 | 0.076203 | 171 | 0.949176 | 0.923797 | 524.858 | 42.245 |
| Bi | 2 | 1 | LibSVM | 40295 | 4478 | 4187 | 291 | 93.501563 | 6.498437 | 0.870041 | 0.919376 | 2064 | 0.049261 | 110 | 0.950739 | 2123 | 0.080624 | 181 | 0.949402 | 0.919376 | 447.742 | 33.152 |
| Bi | 2 | 2 | LibSVM | 40295 | 4478 | 4159 | 319 | 92.876284 | 7.123716 | 0.857535 | 0.915813 | 2056 | 0.058218 | 130 | 0.941782 | 2103 | 0.084187 | 189 | 0.940531 | 0.915813 | 455.882 | 34.558 |
| Bi | 2 | 3 | LibSVM | 40295 | 4478 | 4174 | 304 | 93.211255 | 6.788745 | 0.864237 | 0.914922 | 2054 | 0.050605 | 113 | 0.949395 | 2120 | 0.085078 | 191 | 0.947854 | 0.914922 | 526.498 | 36.433 |
| Bi | 2 | 4 | LibSVM | 40296 | 4477 | 4159 | 318 | 92.897029 | 7.102971 | 0.857947 | 0.91934 | 2063 | 0.061352 | 137 | 0.938648 | 2096 | 0.08066 | 181 | 0.937727 | 0.91934 | 522.265 | 37.511 |
| Bi | 2 | 5 | LibSVM | 40296 | 4477 | 4178 | 299 | 93.321421 | 6.678579 | 0.866439 | 0.918486 | 2062 | 0.051971 | 116 | 0.948029 | 2116 | 0.081514 | 183 | 0.94674 | 0.918486 | 528.623 | 31.621 |
| Bi | 2 | 6 | LibSVM | 40296 | 4477 | 4201 | 276 | 93.835157 | 6.164843 | 0.876718 | 0.915813 | 2056 | 0.038978 | 87 | 0.961022 | 2145 | 0.084187 | 189 | 0.959403 | 0.915813 | 475.864 | 31.074 |
| Bi | 2 | 7 | LibSVM | 40296 | 4477 | 4184 | 293 | 93.455439 | 6.544561 | 0.869119 | 0.919376 | 2064 | 0.050179 | 112 | 0.949821 | 2120 | 0.080624 | 181 | 0.948529 | 0.919376 | 490.518 | 40.402 |
| Bi | 2 | 8 | LibSVM | 40296 | 4477 | 4184 | 293 | 93.455439 | 6.544561 | 0.869113 | 0.928285 | 2084 | 0.05914 | 132 | 0.94086 | 2100 | 0.071715 | 161 | 0.940433 | 0.928285 | 428.209 | 29.617 |
| Bi | 2 | 9 | LibSVM | 40296 | 4477 | 4204 | 273 | 93.902167 | 6.097833 | 0.878051 | 0.926949 | 2081 | 0.048835 | 109 | 0.951165 | 2123 | 0.073051 | 164 | 0.950228 | 0.926949 | 408.229 | 33.477 |
| Bi | 2 | 10 | LibSVM | 40296 | 4477 | 4197 | 280 | 93.745812 | 6.254188 | 0.874924 | 0.925612 | 2078 | 0.050627 | 113 | 0.949373 | 2119 | 0.074388 | 167 | 0.948425 | 0.925612 | 517.172 | 34.916 |
| Bi | 3 | 1 | LibSVM | 40295 | 4478 | 4176 | 302 | 93.255918 | 6.744082 | 0.865134 | 0.909577 | 2042 | 0.044335 | 99 | 0.955665 | 2134 | 0.090423 | 203 | 0.95376 | 0.909577 | 507.003 | 33.115 |
| Bi | 3 | 2 | LibSVM | 40295 | 4478 | 4193 | 285 | 93.635552 | 6.364448 | 0.872718 | 0.924276 | 2075 | 0.0515 | 115 | 0.9485 | 2118 | 0.075724 | 170 | 0.947489 | 0.924276 | 529.146 | 35.463 |
| Bi | 3 | 3 | LibSVM | 40295 | 4478 | 4201 | 277 | 93.814203 | 6.185797 | 0.876292 | 0.924722 | 2076 | 0.048365 | 108 | 0.951635 | 2125 | 0.075278 | 169 | 0.950549 | 0.924722 | 479.262 | 31.637 |
| Bi | 3 | 4 | LibSVM | 40296 | 4477 | 4163 | 314 | 92.986375 | 7.013625 | 0.859741 | 0.912249 | 2048 | 0.052419 | 117 | 0.947581 | 2115 | 0.087751 | 197 | 0.945958 | 0.912249 | 424.29 | 29.199 |
| Bi | 3 | 5 | LibSVM | 40296 | 4477 | 4201 | 276 | 93.835157 | 6.164843 | 0.876708 | 0.930512 | 2089 | 0.053763 | 120 | 0.946237 | 2112 | 0.069488 | 156 | 0.945677 | 0.930512 | 420.667 | 28.122 |
| Bi | 3 | 6 | LibSVM | 40296 | 4477 | 4159 | 318 | 92.897029 | 7.102971 | 0.857953 | 0.912695 | 2049 | 0.054659 | 122 | 0.945341 | 2110 | 0.087305 | 196 | 0.943805 | 0.912695 | 402.002 | 26.099 |
| Bi | 3 | 7 | LibSVM | 40296 | 4477 | 4194 | 283 | 93.678803 | 6.321197 | 0.873588 | 0.918931 | 2063 | 0.045251 | 101 | 0.954749 | 2131 | 0.081069 | 182 | 0.953327 | 0.918931 | 418.315 | 28.34 |
| Bi | 3 | 8 | LibSVM | 40296 | 4477 | 4187 | 290 | 93.522448 | 6.477552 | 0.870456 | 0.925167 | 2077 | 0.054659 | 122 | 0.945341 | 2110 | 0.074833 | 168 | 0.94452 | 0.925167 | 506.909 | 34.742 |
| Bi | 3 | 9 | LibSVM | 40296 | 4477 | 4164 | 313 | 93.008711 | 6.991289 | 0.860181 | 0.920713 | 2067 | 0.060484 | 135 | 0.939516 | 2097 | 0.079287 | 178 | 0.938692 | 0.920713 | 487.995 | 34.022 |
| Bi | 3 | 10 | LibSVM | 40296 | 4477 | 4194 | 283 | 93.678803 | 6.321197 | 0.873585 | 0.921569 | 2068 | 0.047918 | 107 | 0.952082 | 2126 | 0.078431 | 176 | 0.950805 | 0.921569 | 472.901 | 32.31 |
| Bi | 4 | 1 | LibSVM | 40295 | 4478 | 4178 | 300 | 93.300581 | 6.699419 | 0.866019 | 0.921158 | 2068 | 0.055083 | 123 | 0.944917 | 2110 | 0.078842 | 177 | 0.943861 | 0.921158 | 476.273 | 36.088 |
| Bi | 4 | 2 | LibSVM | 40295 | 4478 | 4205 | 273 | 93.903528 | 6.096472 | 0.878076 | 0.929176 | 2086 | 0.051052 | 114 | 0.948948 | 2119 | 0.070824 | 159 | 0.948182 | 0.929176 | 400.061 | 27.3 |
| Bi | 4 | 3 | LibSVM | 40295 | 4478 | 4198 | 280 | 93.747209 | 6.252791 | 0.874951 | 0.925612 | 2078 | 0.050605 | 113 | 0.949395 | 2120 | 0.074388 | 167 | 0.948425 | 0.925612 | 427.237 | 26.675 |
| Bi | 4 | 4 | LibSVM | 40296 | 4477 | 4178 | 299 | 93.321421 | 6.678579 | 0.866441 | 0.915813 | 2056 | 0.049283 | 110 | 0.950717 | 2122 | 0.084187 | 189 | 0.949215 | 0.915813 | 405.93 | 27.644 |
| Bi | 4 | 5 | LibSVM | 40296 | 4477 | 4173 | 304 | 93.209739 | 6.790261 | 0.864209 | 0.912695 | 2049 | 0.048387 | 108 | 0.951613 | 2124 | 0.087305 | 196 | 0.94993 | 0.912695 | 384.633 | 29.568 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bi | 4 | 6 | LibSVM | 40296 | 4477 | 4191 | 286 | 93.611794 | 6.388206 | 0.872245 | 0.922049 | 2070 | 0.049731 | 111 | 0.950269 | 2121 | 0.077951 | 175 | 0.949106 | 0.922049 | 383.164 | 27.093 |
| Bi | 4 | 7 | LibSVM | 40296 | 4477 | 4183 | 294 | 93.433103 | 6.566897 | 0.868672 | 0.919822 | 2065 | 0.051075 | 114 | 0.948925 | 2118 | 0.080178 | 180 | 0.947682 | 0.919822 | 390.908 | 28.056 |
| Bi | 4 | 8 | LibSVM | 40296 | 4477 | 4169 | 308 | 93.120393 | 6.879607 | 0.862423 | 0.910913 | 2045 | 0.048387 | 108 | 0.951613 | 2124 | 0.089087 | 200 | 0.949837 | 0.910913 | 513.289 | 35.075 |
| Bi | 4 | 9 | LibSVM | 40296 | 4477 | 4168 | 309 | 93.098057 | 6.901943 | 0.86197 | 0.918931 | 2063 | 0.0569 | 127 | 0.9431 | 2105 | 0.081069 | 182 | 0.942009 | 0.918931 | 526.368 | 36.232 |
| Bi | 4 | 10 | LibSVM | 40296 | 4477 | 4160 | 317 | 92.919366 | 7.080634 | 0.858394 | 0.918895 | 2062 | 0.060457 | 135 | 0.939543 | 2098 | 0.081105 | 182 | 0.938553 | 0.918895 | 528.269 | 36.638 |
| Bi | 5 | 1 | LibSVM | 40295 | 4478 | 4206 | 272 | 93.92586 | 6.07414 | 0.878529 | 0.920713 | 2067 | 0.042096 | 94 | 0.957904 | 2139 | 0.079287 | 178 | 0.956502 | 0.920713 | 558.745 | 32.955 |
| Bi | 5 | 2 | LibSVM | 40295 | 4478 | 4182 | 296 | 93.389906 | 6.610094 | 0.867808 | 0.919376 | 2064 | 0.0515 | 115 | 0.9485 | 2118 | 0.080624 | 181 | 0.947223 | 0.919376 | 490.509 | 31.005 |
| Bi | 5 | 3 | LibSVM | 40295 | 4478 | 4206 | 272 | 93.92586 | 6.07414 | 0.878521 | 0.932294 | 2093 | 0.053739 | 120 | 0.946261 | 2113 | 0.067706 | 152 | 0.945775 | 0.932294 | 478.93 | 29.554 |
| Bi | 5 | 4 | LibSVM | 40296 | 4477 | 4171 | 306 | 93.165066 | 6.834934 | 0.863312 | 0.913993 | 2051 | 0.050605 | 113 | 0.949395 | 2120 | 0.086007 | 193 | 0.947782 | 0.913993 | 503.787 | 31.645 |
| Bi | 5 | 5 | LibSVM | 40296 | 4477 | 4173 | 304 | 93.209739 | 6.790261 | 0.864208 | 0.913586 | 2051 | 0.049283 | 110 | 0.950717 | 2122 | 0.086414 | 194 | 0.949098 | 0.913586 | 438.36 | 29.149 |
| Bi | 5 | 6 | LibSVM | 40296 | 4477 | 4173 | 304 | 93.209739 | 6.790261 | 0.864206 | 0.916704 | 2058 | 0.052419 | 117 | 0.947581 | 2115 | 0.083296 | 187 | 0.946207 | 0.916704 | 451.671 | 28.321 |
| Bi | 5 | 7 | LibSVM | 40296 | 4477 | 4177 | 300 | 93.299084 | 6.700916 | 0.865994 | 0.915813 | 2056 | 0.049731 | 111 | 0.950269 | 2121 | 0.084187 | 189 | 0.948777 | 0.915813 | 429.637 | 27.182 |
| Bi | 5 | 8 | LibSVM | 40296 | 4477 | 4186 | 291 | 93.500112 | 6.499888 | 0.870012 | 0.921158 | 2068 | 0.051075 | 114 | 0.948925 | 2118 | 0.078842 | 177 | 0.947754 | 0.921158 | 479.133 | 33.985 |
| Bi | 5 | 9 | LibSVM | 40296 | 4477 | 4159 | 318 | 92.897029 | 7.102971 | 0.85795 | 0.916258 | 2057 | 0.058244 | 130 | 0.941756 | 2102 | 0.083742 | 188 | 0.940558 | 0.916258 | 434.693 | 30.506 |
| Bi | 5 | 10 | LibSVM | 40296 | 4477 | 4200 | 277 | 93.812821 | 6.187179 | 0.876265 | 0.925167 | 2077 | 0.048835 | 109 | 0.951165 | 2123 | 0.074833 | 168 | 0.950137 | 0.925167 | 473.438 | 28.352 |
| Bi | 6 | 1 | LibSVM | 40295 | 4478 | 4187 | 291 | 93.501563 | 6.498437 | 0.870043 | 0.917149 | 2059 | 0.047022 | 105 | 0.952978 | 2128 | 0.082851 | 186 | 0.951479 | 0.917149 | 428.929 | 26.15 |
| Bi | 6 | 2 | LibSVM | 40295 | 4478 | 4181 | 297 | 93.367575 | 6.632425 | 0.867361 | 0.918931 | 2063 | 0.0515 | 115 | 0.9485 | 2118 | 0.081069 | 182 | 0.947199 | 0.918931 | 389.004 | 27.137 |
| Bi | 6 | 3 | LibSVM | 40295 | 4478 | 4203 | 275 | 93.858866 | 6.141134 | 0.877186 | 0.923831 | 2074 | 0.046574 | 104 | 0.953426 | 2129 | 0.076169 | 171 | 0.95225 | 0.923831 | 430.623 | 28.412 |
| Bi | 6 | 4 | LibSVM | 40296 | 4477 | 4169 | 308 | 93.120393 | 6.879607 | 0.86242 | 0.915367 | 2055 | 0.052867 | 118 | 0.947133 | 2114 | 0.084633 | 190 | 0.945697 | 0.915367 | 421.562 | 31.786 |
| Bi | 6 | 5 | LibSVM | 40296 | 4477 | 4182 | 295 | 93.410766 | 6.589234 | 0.868228 | 0.915813 | 2056 | 0.047491 | 106 | 0.952509 | 2126 | 0.084187 | 189 | 0.950971 | 0.915813 | 402.971 | 31.931 |
| Bi | 6 | 6 | LibSVM | 40296 | 4477 | 4218 | 259 | 94.214876 | 5.785124 | 0.884306 | 0.928731 | 2085 | 0.044355 | 99 | 0.955645 | 2133 | 0.071269 | 160 | 0.95467 | 0.928731 | 414.233 | 32.758 |
| Bi | 6 | 7 | LibSVM | 40296 | 4477 | 4193 | 284 | 93.656466 | 6.343534 | 0.873134 | 0.928731 | 2085 | 0.055556 | 124 | 0.944444 | 2108 | 0.071269 | 160 | 0.943866 | 0.928731 | 430.292 | 33.928 |
| Bi | 6 | 8 | LibSVM | 40296 | 4477 | 4156 | 321 | 92.83002 | 7.16998 | 0.856603 | 0.924276 | 2075 | 0.067652 | 151 | 0.932348 | 2081 | 0.075724 | 170 | 0.932165 | 0.924276 | 418.735 | 30.31 |
| Bi | 6 | 9 | LibSVM | 40296 | 4477 | 4170 | 307 | 93.14273 | 6.85727 | 0.862862 | 0.921158 | 2068 | 0.058244 | 130 | 0.941756 | 2102 | 0.078842 | 177 | 0.940855 | 0.921158 | 448.997 | 29.124 |
| Bi | 6 | 10 | LibSVM | 40296 | 4477 | 4178 | 299 | 93.321421 | 6.678579 | 0.866439 | 0.91533 | 2054 | 0.048813 | 109 | 0.951187 | 2124 | 0.08467 | 190 | 0.949607 | 0.91533 | 419.967 | 29.031 |
| Bi | 7 | 1 | LibSVM | 40295 | 4478 | 4175 | 303 | 93.233586 | 6.766414 | 0.864685 | 0.912695 | 2049 | 0.047918 | 107 | 0.952082 | 2126 | 0.087305 | 196 | 0.950371 | 0.912695 | 464.862 | 34.584 |
| Bi | 7 | 2 | LibSVM | 40295 | 4478 | 4185 | 293 | 93.4569 | 6.5431 | 0.869145 | 0.923831 | 2074 | 0.054635 | 122 | 0.945365 | 2111 | 0.076169 | 171 | 0.944444 | 0.923831 | 430.606 | 30.294 |
| Bi | 7 | 3 | LibSVM | 40295 | 4478 | 4190 | 288 | 93.568557 | 6.431443 | 0.871381 | 0.920267 | 2066 | 0.048813 | 109 | 0.951187 | 2124 | 0.079733 | 179 | 0.949885 | 0.920267 | 412.245 | 31.542 |
| Bi | 7 | 4 | LibSVM | 40296 | 4477 | 4175 | 302 | 93.254411 | 6.745589 | 0.865097 | 0.918449 | 2061 | 0.053292 | 119 | 0.946708 | 2114 | 0.081551 | 183 | 0.945413 | 0.918449 | 435.395 | 30.419 |
| Bi | 7 | 5 | LibSVM | 40296 | 4477 | 4201 | 276 | 93.835157 | 6.164843 | 0.876714 | 0.922494 | 2071 | 0.045699 | 102 | 0.954301 | 2130 | 0.077506 | 174 | 0.95306 | 0.922494 | 430.949 | 34.21 |
| Bi | 7 | 6 | LibSVM | 40296 | 4477 | 4182 | 295 | 93.410766 | 6.589234 | 0.868218 | 0.929621 | 2087 | 0.06138 | 137 | 0.93862 | 2095 | 0.070379 | 158 | 0.938399 | 0.929621 | 416.629 | 30.871 |
| Bi | 7 | 7 | LibSVM | 40296 | 4477 | 4165 | 312 | 93.031048 | 6.968952 | 0.860633 | 0.914477 | 2053 | 0.053763 | 120 | 0.946237 | 2112 | 0.085523 | 192 | 0.944777 | 0.914477 | 420.856 | 33.695 |
| Bi | 7 | 8 | LibSVM | 40296 | 4477 | 4169 | 308 | 93.120393 | 6.879607 | 0.86242 | 0.914922 | 2054 | 0.052419 | 117 | 0.947581 | 2115 | 0.085078 | 191 | 0.946108 | 0.914922 | 433.024 | 40.683 |
| Bi | 7 | 9 | LibSVM | 40296 | 4477 | 4211 | 266 | 94.058521 | 5.941479 | 0.881179 | 0.926503 | 2080 | 0.045251 | 101 | 0.954749 | 2131 | 0.073497 | 165 | 0.953691 | 0.926503 | 479.769 | 32 |
| Bi | 7 | 10 | LibSVM | 40296 | 4477 | 4184 | 293 | 93.455439 | 6.544561 | 0.86912 | 0.918486 | 2062 | 0.049283 | 110 | 0.950717 | 2122 | 0.081514 | 183 | 0.949355 | 0.918486 | 497.038 | 36.053 |
| Bi | 8 | 1 | LibSVM | 40295 | 4478 | 4191 | 287 | 93.590889 | 6.409111 | 0.871826 | 0.92294 | 2072 | 0.051052 | 114 | 0.948948 | 2119 | 0.07706 | 173 | 0.94785 | 0.92294 | 445.481 | 30.208 |
| Bi | 8 | 2 | LibSVM | 40295 | 4478 | 4184 | 294 | 93.434569 | 6.565431 | 0.868698 | 0.923385 | 2073 | 0.054635 | 122 | 0.945365 | 2111 | 0.076615 | 172 | 0.944419 | 0.923385 | 445.132 | 34.757 |
| Bi | 8 | 3 | LibSVM | 40295 | 4478 | 4170 | 308 | 93.121929 | 6.878071 | 0.862445 | 0.921604 | 2069 | 0.059113 | 132 | 0.940887 | 2101 | 0.078396 | 176 | 0.940027 | 0.921604 | 504.921 | 45.676 |
| Bi | 8 | 4 | LibSVM | 40296 | 4477 | 4201 | 276 | 93.835157 | 6.164843 | 0.876713 | 0.92294 | 2072 | 0.046147 | 103 | 0.953853 | 2129 | 0.07706 | 173 | 0.952644 | 0.92294 | 593.092 | 38.039 |
| Bi | 8 | 5 | LibSVM | 40296 | 4477 | 4194 | 283 | 93.678803 | 6.321197 | 0.87358 | 0.929621 | 2087 | 0.056004 | 125 | 0.943996 | 2107 | 0.070379 | 158 | 0.94349 | 0.929621 | 534.808 | 34.235 |
| Bi | 8 | 6 | LibSVM | 40296 | 4477 | 4209 | 268 | 94.013849 | 5.986151 | 0.880289 | 0.921158 | 2068 | 0.040771 | 91 | 0.959229 | 2141 | 0.078842 | 177 | 0.957851 | 0.921158 | 544.989 | 37.041 |
| Bi | 8 | 7 | LibSVM | 40296 | 4477 | 4146 | 331 | 92.606656 | 7.393344 | 0.85215 | 0.904677 | 2031 | 0.052419 | 117 | 0.947581 | 2115 | 0.095323 | 214 | 0.945531 | 0.904677 | 487.727 | 35.093 |
| Bi | 8 | 8 | LibSVM | 40296 | 4477 | 4180 | 297 | 93.366093 | 6.633907 | 0.867331 | 0.920713 | 2067 | 0.053315 | 119 | 0.946685 | 2113 | 0.079287 | 178 | 0.945563 | 0.920713 | 465.59 | 30.852 |
| Bi | 8 | 9 | LibSVM | 40296 | 4477 | 4182 | 295 | 93.410766 | 6.589234 | 0.868228 | 0.916258 | 2057 | 0.047939 | 107 | 0.952061 | 2125 | 0.083742 | 188 | 0.950555 | 0.916258 | 504.969 | 31.211 |
| Bi | 8 | 10 | LibSVM | 40296 | 4477 | 4195 | 282 | 93.701139 | 6.298861 | 0.874032 | 0.921569 | 2068 | 0.04747 | 106 | 0.95253 | 2127 | 0.078431 | 176 | 0.951242 | 0.921569 | 466.68 | 34.002 |
| Bi | 9 | 1 | LibSVM | 40295 | 4478 | 4188 | 290 | 93.523895 | 6.476105 | 0.870487 | 0.920713 | 2067 | 0.050157 | 112 | 0.949843 | 2121 | 0.079287 | 178 | 0.9486 | 0.920713 | 494.711 | 30.93 |
| Bi | 9 | 2 | LibSVM | 40295 | 4478 | 4199 | 279 | 93.76954 | 6.23046 | 0.875399 | 0.924722 | 2076 | 0.049261 | 110 | 0.950739 | 2123 | 0.075278 | 169 | 0.94968 | 0.924722 | 475.473 | 39.193 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bi | 9 | 3 | LibSVM | 40295 | 4478 | 4164 | 314 | 92.987941 | 7.012059 | 0.859766 | 0.918931 | 2063 | 0.059113 | 132 | 0.940887 | 2101 | 0.081069 | 182 | 0.939863 | 0.918931 | 530.287 | 40.471 |
| Bi | 9 | 4 | LibSVM | 40296 | 4477 | 4198 | 279 | 93.768148 | 6.231852 | 0.875377 | 0.917595 | 2060 | 0.042115 | 94 | 0.957885 | 2138 | 0.082405 | 185 | 0.95636 | 0.917595 | 530.012 | 36.785 |
| Bi | 9 | 5 | LibSVM | 40296 | 4477 | 4182 | 295 | 93.410766 | 6.589234 | 0.868227 | 0.91804 | 2061 | 0.049731 | 111 | 0.950269 | 2121 | 0.08196 | 184 | 0.948895 | 0.91804 | 515.702 | 39.631 |
| Bi | 9 | 6 | LibSVM | 40296 | 4477 | 4180 | 297 | 93.366093 | 6.633907 | 0.867332 | 0.919376 | 2064 | 0.051971 | 116 | 0.948029 | 2116 | 0.080624 | 181 | 0.946789 | 0.919376 | 532.494 | 34.359 |
| Bi | 9 | 7 | LibSVM | 40296 | 4477 | 4202 | 275 | 93.857494 | 6.142506 | 0.877161 | 0.922049 | 2070 | 0.044803 | 100 | 0.955197 | 2132 | 0.077951 | 175 | 0.953917 | 0.922049 | 573.281 | 37.792 |
| Bi | 9 | 8 | LibSVM | 40296 | 4477 | 4189 | 288 | 93.567121 | 6.432879 | 0.871343 | 0.933185 | 2095 | 0.061828 | 138 | 0.938172 | 2094 | 0.066815 | 150 | 0.9382 | 0.933185 | 564.773 | 42.718 |
| Bi | 9 | 9 | LibSVM | 40296 | 4477 | 4163 | 314 | 92.986375 | 7.013625 | 0.85974 | 0.913586 | 2051 | 0.053763 | 120 | 0.946237 | 2112 | 0.086414 | 194 | 0.944726 | 0.913586 | 512.151 | 31.536 |
| Bi | 9 | 10 | LibSVM | 40296 | 4477 | 4176 | 301 | 93.276748 | 6.723252 | 0.865546 | 0.914884 | 2053 | 0.049261 | 110 | 0.950739 | 2123 | 0.085116 | 191 | 0.949145 | 0.914884 | 484.621 | 30.553 |
| Bi | 10 | 1 | LibSVM | 40295 | 4478 | 4214 | 264 | 94.104511 | 5.895489 | 0.882102 | 0.921604 | 2069 | 0.039409 | 88 | 0.960591 | 2145 | 0.078396 | 176 | 0.959203 | 0.921604 | 462.396 | 34.655 |
| Bi | 10 | 2 | LibSVM | 40295 | 4478 | 4182 | 296 | 93.389906 | 6.610094 | 0.867807 | 0.919822 | 2065 | 0.051948 | 116 | 0.948052 | 2117 | 0.080178 | 180 | 0.946813 | 0.919822 | 464.72 | 30.475 |
| Bi | 10 | 3 | LibSVM | 40295 | 4478 | 4169 | 309 | 93.099598 | 6.900402 | 0.862003 | 0.914922 | 2054 | 0.052844 | 118 | 0.947156 | 2115 | 0.085078 | 191 | 0.945672 | 0.914922 | 439.08 | 35.809 |
| Bi | 10 | 4 | LibSVM | 40296 | 4477 | 4178 | 299 | 93.321421 | 6.678579 | 0.866442 | 0.914922 | 2054 | 0.048387 | 108 | 0.951613 | 2124 | 0.085078 | 191 | 0.950046 | 0.914922 | 448.702 | 34.717 |
| Bi | 10 | 5 | LibSVM | 40296 | 4477 | 4176 | 301 | 93.276748 | 6.723252 | 0.865545 | 0.918486 | 2062 | 0.052867 | 118 | 0.947133 | 2114 | 0.081514 | 183 | 0.945872 | 0.918486 | 451.651 | 32.003 |
| Bi | 10 | 6 | LibSVM | 40296 | 4477 | 4165 | 312 | 93.031048 | 6.968952 | 0.860634 | 0.913586 | 2051 | 0.052867 | 118 | 0.947133 | 2114 | 0.086414 | 194 | 0.945597 | 0.913586 | 437.021 | 28.073 |
| Bi | 10 | 7 | LibSVM | 40296 | 4477 | 4209 | 268 | 94.013849 | 5.986151 | 0.880283 | 0.930512 | 2089 | 0.050179 | 112 | 0.949821 | 2120 | 0.069488 | 156 | 0.949114 | 0.930512 | 417.822 | 32.628 |
| Bi | 10 | 8 | LibSVM | 40296 | 4477 | 4166 | 311 | 93.053384 | 6.946616 | 0.861081 | 0.912249 | 2048 | 0.051075 | 114 | 0.948925 | 2118 | 0.087751 | 197 | 0.947271 | 0.912249 | 390.311 | 31.785 |
| Bi | 10 | 9 | LibSVM | 40296 | 4477 | 4176 | 301 | 93.276748 | 6.723252 | 0.865545 | 0.918931 | 2063 | 0.053315 | 119 | 0.946685 | 2113 | 0.081069 | 182 | 0.945463 | 0.918931 | 393.679 | 27.995 |
| Bi | 10 | 10 | LibSVM | 40296 | 4477 | 4210 | 267 | 94.036185 | 5.963815 | 0.880726 | 0.935383 | 2099 | 0.054635 | 122 | 0.945365 | 2111 | 0.064617 | 145 | 0.94507 | 0.935383 | 399.543 | 30.662 |
| Tri | 1 | 1 | LibSVM | 40295 | 4478 | 4161 | 317 | 92.920947 | 7.079053 | 0.858442 | 0.89755 | 2015 | 0.038961 | 87 | 0.961039 | 2146 | 0.10245 | 230 | 0.958611 | 0.89755 | 527.885 | 34.389 |
| Tri | 1 | 2 | LibSVM | 40295 | 4478 | 4190 | 288 | 93.568557 | 6.431443 | 0.871393 | 0.903341 | 2028 | 0.031796 | 71 | 0.968204 | 2162 | 0.096659 | 217 | 0.966174 | 0.903341 | 512.818 | 40.488 |
| Tri | 1 | 3 | LibSVM | 40295 | 4478 | 4149 | 329 | 92.65297 | 7.34703 | 0.853082 | 0.897105 | 2014 | 0.043887 | 98 | 0.956113 | 2135 | 0.102895 | 231 | 0.953598 | 0.897105 | 557.611 | 37.556 |
| Tri | 1 | 4 | LibSVM | 40296 | 4477 | 4163 | 314 | 92.986375 | 7.013625 | 0.85975 | 0.901559 | 2024 | 0.041667 | 93 | 0.958333 | 2139 | 0.098441 | 221 | 0.95607 | 0.901559 | 535.558 | 35.076 |
| Tri | 1 | 5 | LibSVM | 40296 | 4477 | 4171 | 306 | 93.165066 | 6.834934 | 0.863327 | 0.897996 | 2016 | 0.034498 | 77 | 0.965502 | 2155 | 0.102004 | 229 | 0.963211 | 0.897996 | 535.043 | 34.951 |
| Tri | 1 | 6 | LibSVM | 40296 | 4477 | 4184 | 293 | 93.455439 | 6.544561 | 0.869132 | 0.902895 | 2027 | 0.033602 | 75 | 0.966398 | 2157 | 0.097105 | 218 | 0.96432 | 0.902895 | 551.965 | 34.905 |
| Tri | 1 | 7 | LibSVM | 40296 | 4477 | 4156 | 321 | 92.83002 | 7.16998 | 0.856627 | 0.895323 | 2010 | 0.03853 | 86 | 0.96147 | 2146 | 0.104677 | 235 | 0.958969 | 0.895323 | 488.848 | 36.276 |
| Tri | 1 | 8 | LibSVM | 40296 | 4477 | 4163 | 314 | 92.986375 | 7.013625 | 0.859754 | 0.896659 | 2013 | 0.036738 | 82 | 0.963262 | 2150 | 0.103341 | 232 | 0.960859 | 0.896659 | 554.569 | 34.733 |
| Tri | 1 | 9 | LibSVM | 40296 | 4477 | 4212 | 265 | 94.080858 | 5.919142 | 0.881639 | 0.908241 | 2039 | 0.026434 | 59 | 0.973566 | 2173 | 0.091759 | 206 | 0.971878 | 0.908241 | 557.315 | 37.415 |
| Tri | 1 | 10 | LibSVM | 40296 | 4477 | 4195 | 282 | 93.701139 | 6.298861 | 0.874038 | 0.910873 | 2044 | 0.036722 | 82 | 0.963278 | 2151 | 0.089127 | 200 | 0.96143 | 0.910873 | 545.633 | 34.623 |
| Tri | 2 | 1 | LibSVM | 40295 | 4478 | 4177 | 301 | 93.278249 | 6.721751 | 0.865589 | 0.898441 | 2017 | 0.032691 | 73 | 0.967309 | 2160 | 0.101559 | 228 | 0.965072 | 0.898441 | 558.266 | 37.976 |
| Tri | 2 | 2 | LibSVM | 40295 | 4478 | 4148 | 330 | 92.630639 | 7.369361 | 0.852638 | 0.893541 | 2006 | 0.040752 | 91 | 0.959248 | 2142 | 0.106459 | 239 | 0.956605 | 0.893541 | 512.506 | 34.437 |
| Tri | 2 | 3 | LibSVM | 40295 | 4478 | 4162 | 316 | 92.943278 | 7.056722 | 0.85889 | 0.895768 | 2011 | 0.036722 | 82 | 0.963278 | 2151 | 0.104232 | 234 | 0.960822 | 0.895768 | 537.867 | 37.508 |
| Tri | 2 | 4 | LibSVM | 40296 | 4477 | 4197 | 280 | 93.745812 | 6.254188 | 0.874933 | 0.909091 | 2040 | 0.034035 | 76 | 0.965965 | 2157 | 0.090909 | 204 | 0.964083 | 0.909091 | 561.183 | 34.592 |
| Tri | 2 | 5 | LibSVM | 40296 | 4477 | 4187 | 290 | 93.522448 | 6.477552 | 0.870468 | 0.908686 | 2040 | 0.038082 | 85 | 0.961918 | 2147 | 0.091314 | 205 | 0.96 | 0.908686 | 541.235 | 36.978 |
| Tri | 2 | 6 | LibSVM | 40296 | 4477 | 4181 | 296 | 93.38843 | 6.61157 | 0.867799 | 0.893541 | 2006 | 0.025538 | 57 | 0.974462 | 2175 | 0.106459 | 239 | 0.97237 | 0.893541 | 549.033 | 36.885 |
| Tri | 2 | 7 | LibSVM | 40296 | 4477 | 4162 | 315 | 92.964038 | 7.035962 | 0.859305 | 0.899332 | 2019 | 0.039875 | 89 | 0.960125 | 2143 | 0.100668 | 226 | 0.95778 | 0.899332 | 523.205 | 36.012 |
| Tri | 2 | 8 | LibSVM | 40296 | 4477 | 4170 | 307 | 93.14273 | 6.85727 | 0.862874 | 0.906459 | 2035 | 0.043459 | 97 | 0.956541 | 2135 | 0.093541 | 210 | 0.954503 | 0.906459 | 566.86 | 37.477 |
| Tri | 2 | 9 | LibSVM | 40296 | 4477 | 4169 | 308 | 93.120393 | 6.879607 | 0.862435 | 0.895768 | 2011 | 0.033154 | 74 | 0.966846 | 2158 | 0.104232 | 234 | 0.964508 | 0.895768 | 553.556 | 37.493 |
| Tri | 2 | 10 | LibSVM | 40296 | 4477 | 4187 | 290 | 93.522448 | 6.477552 | 0.870469 | 0.90735 | 2037 | 0.036738 | 82 | 0.963262 | 2150 | 0.09265 | 208 | 0.961303 | 0.90735 | 557.486 | 36.199 |
| Tri | 3 | 1 | LibSVM | 40295 | 4478 | 4170 | 308 | 93.121929 | 6.878071 | 0.862467 | 0.891759 | 2002 | 0.029109 | 65 | 0.970891 | 2168 | 0.108241 | 243 | 0.968553 | 0.891759 | 548.004 | 37.321 |
| Tri | 3 | 2 | LibSVM | 40295 | 4478 | 4175 | 303 | 93.233586 | 6.766414 | 0.864693 | 0.90245 | 2026 | 0.037618 | 84 | 0.962382 | 2149 | 0.09755 | 219 | 0.96019 | 0.90245 | 582.315 | 38.102 |
| Tri | 3 | 3 | LibSVM | 40295 | 4478 | 4201 | 277 | 93.814203 | 6.185797 | 0.876302 | 0.909577 | 2042 | 0.033139 | 74 | 0.966861 | 2159 | 0.090423 | 203 | 0.965028 | 0.909577 | 523.861 | 37.103 |
| Tri | 3 | 4 | LibSVM | 40296 | 4477 | 4147 | 330 | 92.628993 | 7.371007 | 0.852612 | 0.887751 | 1993 | 0.034946 | 78 | 0.965054 | 2154 | 0.112249 | 252 | 0.962337 | 0.887751 | 506.018 | 37.946 |
| Tri | 3 | 5 | LibSVM | 40296 | 4477 | 4196 | 281 | 93.723476 | 6.276524 | 0.874487 | 0.911804 | 2047 | 0.037186 | 83 | 0.962814 | 2149 | 0.088196 | 198 | 0.961033 | 0.911804 | 548.753 | 37.743 |
| Tri | 3 | 6 | LibSVM | 40296 | 4477 | 4157 | 320 | 92.852356 | 7.147644 | 0.85707 | 0.899332 | 2019 | 0.042115 | 94 | 0.957885 | 2138 | 0.100668 | 226 | 0.955513 | 0.899332 | 551.559 | 39.35 |
| Tri | 3 | 7 | LibSVM | 40296 | 4477 | 4191 | 286 | 93.611794 | 6.388206 | 0.872259 | 0.904232 | 2030 | 0.03181 | 71 | 0.96819 | 2161 | 0.095768 | 215 | 0.966207 | 0.904232 | 549.844 | 38.725 |
| Tri | 3 | 8 | LibSVM | 40296 | 4477 | 4179 | 298 | 93.343757 | 6.656243 | 0.866894 | 0.907795 | 2038 | 0.040771 | 91 | 0.959229 | 2141 | 0.092205 | 207 | 0.957257 | 0.907795 | 530.442 | 39.068 |
| Tri | 3 | 9 | LibSVM | 40296 | 4477 | 4153 | 324 | 92.763011 | 7.236989 | 0.855285 | 0.897105 | 2014 | 0.041667 | 93 | 0.958333 | 2139 | 0.102895 | 231 | 0.955861 | 0.897105 | 608.609 | 40.317 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tri | 3 | 10 | LibSVM | 40296 | 4477 | 4183 | 294 | 93.433103 | 6.566897 | 0.868682 | 0.902406 | 2025 | 0.033587 | 75 | 0.966413 | 2158 | 0.097594 | 219 | 0.964286 | 0.902406 | 582.378 | 38.939 |
| Tri | 4 | 1 | LibSVM | 40295 | 4478 | 4161 | 317 | 92.920947 | 7.079053 | 0.858442 | 0.897996 | 2016 | 0.039409 | 88 | 0.960591 | 2145 | 0.102004 | 229 | 0.958175 | 0.897996 | 586.755 | 42.318 |
| Tri | 4 | 2 | LibSVM | 40295 | 4478 | 4182 | 296 | 93.389906 | 6.610094 | 0.86782 | 0.902004 | 2025 | 0.034035 | 76 | 0.965965 | 2157 | 0.097996 | 220 | 0.963827 | 0.902004 | 590.083 | 39.913 |
| Tri | 4 | 3 | LibSVM | 40295 | 4478 | 4196 | 282 | 93.702546 | 6.297454 | 0.87407 | 0.908241 | 2039 | 0.034035 | 76 | 0.965965 | 2157 | 0.091759 | 206 | 0.964066 | 0.908241 | 513.156 | 41.845 |
| Tri | 4 | 4 | LibSVM | 40296 | 4477 | 4160 | 317 | 92.919366 | 7.080634 | 0.858416 | 0.893096 | 2005 | 0.034498 | 77 | 0.965502 | 2155 | 0.106904 | 240 | 0.963016 | 0.893096 | 617.813 | 40.69 |
| Tri | 4 | 5 | LibSVM | 40296 | 4477 | 4149 | 328 | 92.673665 | 7.326335 | 0.853504 | 0.889087 | 1996 | 0.035394 | 79 | 0.964606 | 2153 | 0.110913 | 249 | 0.961928 | 0.889087 | 551.403 | 38.601 |
| Tri | 4 | 6 | LibSVM | 40296 | 4477 | 4190 | 287 | 93.589457 | 6.410543 | 0.871808 | 0.909131 | 2041 | 0.037186 | 83 | 0.962814 | 2149 | 0.090869 | 204 | 0.960923 | 0.909131 | 573.535 | 36.822 |
| Tri | 4 | 7 | LibSVM | 40296 | 4477 | 4172 | 305 | 93.187402 | 6.812598 | 0.863772 | 0.900223 | 2021 | 0.03629 | 81 | 0.96371 | 2151 | 0.099777 | 224 | 0.961465 | 0.900223 | 536.977 | 38.773 |
| Tri | 4 | 8 | LibSVM | 40296 | 4477 | 4159 | 318 | 92.897029 | 7.102971 | 0.85797 | 0.891759 | 2002 | 0.033602 | 75 | 0.966398 | 2157 | 0.108241 | 243 | 0.96389 | 0.891759 | 531.659 | 42.515 |
| Tri | 4 | 9 | LibSVM | 40296 | 4477 | 4182 | 295 | 93.410766 | 6.589234 | 0.868236 | 0.906013 | 2034 | 0.037634 | 84 | 0.962366 | 2148 | 0.093987 | 211 | 0.96034 | 0.906013 | 562.337 | 40.113 |
| Tri | 4 | 10 | LibSVM | 40296 | 4477 | 4193 | 284 | 93.656466 | 6.343534 | 0.873144 | 0.912656 | 2048 | 0.039409 | 88 | 0.960591 | 2145 | 0.087344 | 196 | 0.958801 | 0.912656 | 585.652 | 40.052 |
| Tri | 5 | 1 | LibSVM | 40295 | 4478 | 4199 | 279 | 93.76954 | 6.23046 | 0.87541 | 0.907795 | 2038 | 0.032244 | 72 | 0.967756 | 2161 | 0.092205 | 207 | 0.965877 | 0.907795 | 526.045 | 40.675 |
| Tri | 5 | 2 | LibSVM | 40295 | 4478 | 4174 | 304 | 93.211255 | 6.788745 | 0.864248 | 0.900223 | 2021 | 0.035826 | 80 | 0.964174 | 2153 | 0.099777 | 224 | 0.961923 | 0.900223 | 589.973 | 42.001 |
| Tri | 5 | 3 | LibSVM | 40295 | 4478 | 4183 | 295 | 93.412238 | 6.587762 | 0.868264 | 0.906459 | 2035 | 0.038065 | 85 | 0.961935 | 2148 | 0.093541 | 210 | 0.959906 | 0.906459 | 613.445 | 40.223 |
| Tri | 5 | 4 | LibSVM | 40296 | 4477 | 4170 | 307 | 93.14273 | 6.85727 | 0.862875 | 0.900178 | 2020 | 0.03717 | 83 | 0.96283 | 2150 | 0.099822 | 224 | 0.960533 | 0.900178 | 584.998 | 39.739 |
| Tri | 5 | 5 | LibSVM | 40296 | 4477 | 4182 | 295 | 93.410766 | 6.589234 | 0.868241 | 0.899777 | 2020 | 0.031362 | 70 | 0.968638 | 2162 | 0.100223 | 225 | 0.966507 | 0.899777 | 599.424 | 36.246 |
| Tri | 5 | 6 | LibSVM | 40296 | 4477 | 4165 | 312 | 93.031048 | 6.968952 | 0.860648 | 0.895768 | 2011 | 0.034946 | 78 | 0.965054 | 2154 | 0.104232 | 234 | 0.962662 | 0.895768 | 604.696 | 38.039 |
| Tri | 5 | 7 | LibSVM | 40296 | 4477 | 4171 | 306 | 93.165066 | 6.834934 | 0.863325 | 0.900668 | 2022 | 0.037186 | 83 | 0.962814 | 2149 | 0.099332 | 223 | 0.96057 | 0.900668 | 546.304 | 39.162 |
| Tri | 5 | 8 | LibSVM | 40296 | 4477 | 4186 | 291 | 93.500112 | 6.499888 | 0.870022 | 0.90735 | 2037 | 0.037186 | 83 | 0.962814 | 2149 | 0.09265 | 208 | 0.960849 | 0.90735 | 589.849 | 38.148 |
| Tri | 5 | 9 | LibSVM | 40296 | 4477 | 4136 | 341 | 92.383292 | 7.616708 | 0.847697 | 0.887751 | 1993 | 0.039875 | 89 | 0.960125 | 2143 | 0.112249 | 252 | 0.957253 | 0.887751 | 609.765 | 42.811 |
| Tri | 5 | 10 | LibSVM | 40296 | 4477 | 4182 | 295 | 93.410766 | 6.589234 | 0.868236 | 0.905568 | 2033 | 0.037186 | 83 | 0.962814 | 2149 | 0.094432 | 212 | 0.960775 | 0.905568 | 654.854 | 47.974 |
| Tri | 6 | 1 | LibSVM | 40295 | 4478 | 4180 | 298 | 93.345243 | 6.654757 | 0.866929 | 0.898886 | 2018 | 0.031796 | 71 | 0.968204 | 2162 | 0.101114 | 227 | 0.966012 | 0.898886 | 607.472 | 38.679 |
| Tri | 6 | 2 | LibSVM | 40295 | 4478 | 4173 | 305 | 93.188924 | 6.811076 | 0.863802 | 0.897996 | 2016 | 0.034035 | 76 | 0.965965 | 2157 | 0.102004 | 229 | 0.963671 | 0.897996 | 610.17 | 46.726 |
| Tri | 6 | 3 | LibSVM | 40295 | 4478 | 4166 | 312 | 93.032604 | 6.967396 | 0.860676 | 0.89755 | 2015 | 0.036722 | 82 | 0.963278 | 2151 | 0.10245 | 230 | 0.960897 | 0.89755 | 605.257 | 39.115 |
| Tri | 6 | 4 | LibSVM | 40296 | 4477 | 4158 | 319 | 92.874693 | 7.125307 | 0.857519 | 0.896659 | 2013 | 0.038978 | 87 | 0.961022 | 2145 | 0.103341 | 232 | 0.958571 | 0.896659 | 590.082 | 41.283 |
| Tri | 6 | 5 | LibSVM | 40296 | 4477 | 4162 | 315 | 92.964038 | 7.035962 | 0.859307 | 0.896659 | 2013 | 0.037186 | 83 | 0.962814 | 2149 | 0.103341 | 232 | 0.960401 | 0.896659 | 676.58 | 40.9 |
| Tri | 6 | 6 | LibSVM | 40296 | 4477 | 4200 | 277 | 93.812821 | 6.187179 | 0.876278 | 0.906459 | 2035 | 0.030018 | 67 | 0.969982 | 2165 | 0.093541 | 210 | 0.968126 | 0.906459 | 670.14 | 54.401 |
| Tri | 6 | 7 | LibSVM | 40296 | 4477 | 4185 | 292 | 93.477775 | 6.522225 | 0.869575 | 0.90735 | 2037 | 0.037634 | 84 | 0.962366 | 2148 | 0.09265 | 208 | 0.960396 | 0.90735 | 720.603 | 41.917 |
| Tri | 6 | 8 | LibSVM | 40296 | 4477 | 4169 | 308 | 93.120393 | 6.879607 | 0.862424 | 0.909577 | 2042 | 0.047043 | 105 | 0.952957 | 2127 | 0.090423 | 203 | 0.951095 | 0.909577 | 625.531 | 39.634 |
| Tri | 6 | 9 | LibSVM | 40296 | 4477 | 4166 | 311 | 93.053384 | 6.946616 | 0.86109 | 0.901559 | 2024 | 0.040323 | 90 | 0.959677 | 2142 | 0.098441 | 221 | 0.957427 | 0.901559 | 637.01 | 41.396 |
| Tri | 6 | 10 | LibSVM | 40296 | 4477 | 4179 | 298 | 93.343757 | 6.656243 | 0.866898 | 0.89795 | 2015 | 0.0309 | 69 | 0.9691 | 2164 | 0.10205 | 229 | 0.966891 | 0.89795 | 634.571 | 40.351 |
| Tri | 7 | 1 | LibSVM | 40295 | 4478 | 4175 | 303 | 93.233586 | 6.766414 | 0.864697 | 0.895768 | 2011 | 0.0309 | 69 | 0.9691 | 2164 | 0.104232 | 234 | 0.966827 | 0.895768 | 628.392 | 43.253 |
| Tri | 7 | 2 | LibSVM | 40295 | 4478 | 4149 | 329 | 92.65297 | 7.34703 | 0.853082 | 0.896659 | 2013 | 0.043439 | 97 | 0.956561 | 2136 | 0.103341 | 232 | 0.954028 | 0.896659 | 594.922 | 51.945 |
| Tri | 7 | 3 | LibSVM | 40295 | 4478 | 4208 | 270 | 93.970523 | 6.029477 | 0.879426 | 0.914031 | 2052 | 0.034483 | 77 | 0.965517 | 2156 | 0.085969 | 193 | 0.963833 | 0.914031 | 633.37 | 46.296 |
| Tri | 7 | 4 | LibSVM | 40296 | 4477 | 4155 | 322 | 92.807684 | 7.192316 | 0.856177 | 0.894385 | 2007 | 0.038065 | 85 | 0.961935 | 2148 | 0.105615 | 237 | 0.959369 | 0.894385 | 680.087 | 46.655 |
| Tri | 7 | 5 | LibSVM | 40296 | 4477 | 4164 | 313 | 93.008711 | 6.991289 | 0.860202 | 0.895323 | 2010 | 0.034946 | 78 | 0.965054 | 2154 | 0.104677 | 235 | 0.962644 | 0.895323 | 944.789 | 65.254 |
| Tri | 7 | 6 | LibSVM | 40296 | 4477 | 4200 | 277 | 93.812821 | 6.187179 | 0.876274 | 0.912249 | 2048 | 0.035842 | 80 | 0.964158 | 2152 | 0.087751 | 197 | 0.962406 | 0.912249 | 947.144 | 61.307 |
| Tri | 7 | 7 | LibSVM | 40296 | 4477 | 4176 | 301 | 93.276748 | 6.723252 | 0.865556 | 0.904677 | 2031 | 0.038978 | 87 | 0.961022 | 2145 | 0.095323 | 214 | 0.958924 | 0.904677 | 972.797 | 74.258 |
| Tri | 7 | 8 | LibSVM | 40296 | 4477 | 4153 | 324 | 92.763011 | 7.236989 | 0.855287 | 0.894432 | 2008 | 0.038978 | 87 | 0.961022 | 2145 | 0.105568 | 237 | 0.958473 | 0.894432 | 998.075 | 59.419 |
| Tri | 7 | 9 | LibSVM | 40296 | 4477 | 4183 | 294 | 93.433103 | 6.566897 | 0.868687 | 0.900223 | 2021 | 0.031362 | 70 | 0.968638 | 2162 | 0.099777 | 224 | 0.966523 | 0.900223 | 813.483 | 46.998 |
| Tri | 7 | 10 | LibSVM | 40296 | 4477 | 4167 | 310 | 93.07572 | 6.92428 | 0.861541 | 0.896659 | 2013 | 0.034946 | 78 | 0.965054 | 2154 | 0.103341 | 232 | 0.962697 | 0.896659 | 709.095 | 41.24 |
| Tri | 8 | 1 | LibSVM | 40295 | 4478 | 4179 | 299 | 93.322912 | 6.677088 | 0.866478 | 0.904677 | 2031 | 0.038065 | 85 | 0.961935 | 2148 | 0.095323 | 214 | 0.95983 | 0.904677 | 665.903 | 42.879 |
| Tri | 8 | 2 | LibSVM | 40295 | 4478 | 4162 | 316 | 92.943278 | 7.056722 | 0.858889 | 0.89755 | 2015 | 0.038513 | 86 | 0.961487 | 2147 | 0.10245 | 230 | 0.959067 | 0.89755 | 645.65 | 41.007 |
| Tri | 8 | 3 | LibSVM | 40295 | 4478 | 4194 | 284 | 93.657883 | 6.342117 | 0.873174 | 0.910913 | 2045 | 0.037618 | 84 | 0.962382 | 2149 | 0.089087 | 200 | 0.960545 | 0.910913 | 610.745 | 41.334 |
| Tri | 8 | 4 | LibSVM | 40296 | 4477 | 4168 | 309 | 93.098057 | 6.901943 | 0.861986 | 0.898886 | 2018 | 0.036738 | 82 | 0.963262 | 2150 | 0.101114 | 227 | 0.960952 | 0.898886 | 844.784 | 53.459 |
| Tri | 8 | 5 | LibSVM | 40296 | 4477 | 4165 | 312 | 93.031048 | 6.968952 | 0.860642 | 0.902895 | 2027 | 0.042115 | 94 | 0.957885 | 2138 | 0.097105 | 218 | 0.955681 | 0.902895 | 841.383 | 52.568 |
| Tri | 8 | 6 | LibSVM | 40296 | 4477 | 4204 | 273 | 93.902167 | 6.097833 | 0.878068 | 0.903341 | 2028 | 0.02509 | 56 | 0.97491 | 2176 | 0.096659 | 217 | 0.973129 | 0.903341 | 788.633 | 40.039 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tri | 8 | 7 | LibSVM | 40296 | 4477 | 4144 | 333 | 92.561983 | 7.438017 | 0.851272 | 0.88686 | 1991 | 0.035394 | 79 | 0.964606 | 2153 | 0.11314 | 254 | 0.961836 | 0.88686 | 652.839 | 44.86 |
| Tri | 8 | 8 | LibSVM | 40296 | 4477 | 4158 | 319 | 92.874693 | 7.125307 | 0.85752 | 0.895768 | 2011 | 0.038082 | 85 | 0.961918 | 2147 | 0.104232 | 234 | 0.959447 | 0.895768 | 644.732 | 42.572 |
| Tri | 8 | 9 | LibSVM | 40296 | 4477 | 4189 | 288 | 93.567121 | 6.432879 | 0.871364 | 0.906013 | 2034 | 0.034498 | 77 | 0.965502 | 2155 | 0.093987 | 211 | 0.963524 | 0.906013 | 781.887 | 46.175 |
| Tri | 8 | 10 | LibSVM | 40296 | 4477 | 4190 | 287 | 93.589457 | 6.410543 | 0.871807 | 0.906417 | 2034 | 0.034483 | 77 | 0.965517 | 2156 | 0.093583 | 210 | 0.963524 | 0.906417 | 650.924 | 45.817 |
| Tri | 9 | 1 | LibSVM | 40295 | 4478 | 4178 | 300 | 93.300581 | 6.699419 | 0.866036 | 0.898441 | 2017 | 0.032244 | 72 | 0.967756 | 2161 | 0.101559 | 228 | 0.965534 | 0.898441 | 678.124 | 42.617 |
| Tri | 9 | 2 | LibSVM | 40295 | 4478 | 4185 | 293 | 93.4569 | 6.5431 | 0.869158 | 0.904677 | 2031 | 0.035378 | 79 | 0.964622 | 2154 | 0.095323 | 214 | 0.962559 | 0.904677 | 654.873 | 40.573 |
| Tri | 9 | 3 | LibSVM | 40295 | 4478 | 4166 | 312 | 93.032604 | 6.967396 | 0.860673 | 0.901559 | 2024 | 0.040752 | 91 | 0.959248 | 2142 | 0.098441 | 221 | 0.956974 | 0.901559 | 674.893 | 40.324 |
| Tri | 9 | 4 | LibSVM | 40296 | 4477 | 4177 | 300 | 93.299084 | 6.700916 | 0.866009 | 0.896659 | 2013 | 0.030466 | 68 | 0.969534 | 2164 | 0.103341 | 232 | 0.967323 | 0.896659 | 649.723 | 41.291 |
| Tri | 9 | 5 | LibSVM | 40296 | 4477 | 4166 | 311 | 93.053384 | 6.946616 | 0.861094 | 0.897105 | 2014 | 0.035842 | 80 | 0.964158 | 2152 | 0.102895 | 231 | 0.961796 | 0.897105 | 647.554 | 45.629 |
| Tri | 9 | 6 | LibSVM | 40296 | 4477 | 4163 | 314 | 92.986375 | 7.013625 | 0.85975 | 0.901559 | 2024 | 0.041667 | 93 | 0.958333 | 2139 | 0.098441 | 221 | 0.95607 | 0.901559 | 859.534 | 54.884 |
| Tri | 9 | 7 | LibSVM | 40296 | 4477 | 4181 | 296 | 93.38843 | 6.61157 | 0.867791 | 0.903786 | 2029 | 0.035842 | 80 | 0.964158 | 2152 | 0.096214 | 216 | 0.962067 | 0.903786 | 900.248 | 60.985 |
| Tri | 9 | 8 | LibSVM | 40296 | 4477 | 4183 | 294 | 93.433103 | 6.566897 | 0.868681 | 0.908241 | 2039 | 0.039427 | 88 | 0.960573 | 2144 | 0.091759 | 206 | 0.958627 | 0.908241 | 752.81 | 43.57 |
| Tri | 9 | 9 | LibSVM | 40296 | 4477 | 4163 | 314 | 92.986375 | 7.013625 | 0.859754 | 0.895768 | 2011 | 0.035842 | 80 | 0.964158 | 2152 | 0.104232 | 234 | 0.961741 | 0.895768 | 616.531 | 43.663 |
| Tri | 9 | 10 | LibSVM | 40296 | 4477 | 4182 | 295 | 93.410766 | 6.589234 | 0.868235 | 0.901961 | 2024 | 0.033587 | 75 | 0.966413 | 2158 | 0.098039 | 220 | 0.964269 | 0.901961 | 963.45 | 64.574 |
| Tri | 10 | 1 | LibSVM | 40295 | 4478 | 4195 | 283 | 93.680214 | 6.319786 | 0.873628 | 0.901114 | 2023 | 0.027318 | 61 | 0.972682 | 2172 | 0.098886 | 222 | 0.970729 | 0.901114 | 1161.619 | 40.839 |
| Tri | 10 | 2 | LibSVM | 40295 | 4478 | 4182 | 296 | 93.389906 | 6.610094 | 0.86782 | 0.901559 | 2024 | 0.033587 | 75 | 0.966413 | 2158 | 0.098441 | 221 | 0.964269 | 0.901559 | 321.977 | 75.848 |
| Tri | 10 | 3 | LibSVM | 40295 | 4478 | 4161 | 317 | 92.920947 | 7.079053 | 0.858441 | 0.898441 | 2017 | 0.039857 | 89 | 0.960143 | 2144 | 0.101559 | 228 | 0.95774 | 0.898441 | 652.231 | 37.028 |
| Tri | 10 | 4 | LibSVM | 40296 | 4477 | 4164 | 313 | 93.008711 | 6.991289 | 0.860203 | 0.893987 | 2007 | 0.033602 | 75 | 0.966398 | 2157 | 0.106013 | 238 | 0.963977 | 0.893987 | 631.628 | 59.368 |
| Tri | 10 | 5 | LibSVM | 40296 | 4477 | 4158 | 319 | 92.874693 | 7.125307 | 0.85752 | 0.895768 | 2011 | 0.038082 | 85 | 0.961918 | 2147 | 0.104232 | 234 | 0.959447 | 0.895768 | 982.637 | 67.172 |
| Tri | 10 | 6 | LibSVM | 40296 | 4477 | 4163 | 314 | 92.986375 | 7.013625 | 0.859753 | 0.897996 | 2016 | 0.038082 | 85 | 0.961918 | 2147 | 0.102004 | 229 | 0.959543 | 0.897996 | 781.615 | 68.719 |
| Tri | 10 | 7 | LibSVM | 40296 | 4477 | 4187 | 290 | 93.522448 | 6.477552 | 0.870468 | 0.908241 | 2039 | 0.037634 | 84 | 0.962366 | 2148 | 0.091759 | 206 | 0.960433 | 0.908241 | 581.988 | 37.937 |
| Tri | 10 | 8 | LibSVM | 40296 | 4477 | 4171 | 306 | 93.165066 | 6.834934 | 0.863324 | 0.90245 | 2026 | 0.038978 | 87 | 0.961022 | 2145 | 0.09755 | 219 | 0.958826 | 0.90245 | 585.163 | 43.779 |
| Tri | 10 | 9 | LibSVM | 40296 | 4477 | 4165 | 312 | 93.031048 | 6.968952 | 0.860647 | 0.896659 | 2013 | 0.035842 | 80 | 0.964158 | 2152 | 0.103341 | 232 | 0.961777 | 0.896659 | 647.25 | 41.676 |
| Tri | 10 | 10 | LibSVM | 40296 | 4477 | 4191 | 286 | 93.611794 | 6.388206 | 0.872251 | 0.911319 | 2045 | 0.038961 | 87 | 0.961039 | 2146 | 0.088681 | 199 | 0.959193 | 0.911319 | 692.643 | 41.637 |

| Set | # | Fold | Classifier | Train ins. | Test inst. | Correct | Incorrect | Correct % | Incorrect % | Kappa | TP rate | TP | FP rate | FP | TN rate | TN | FN rate | FN | Precision | Recall | Train time | Test time |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bi | 1 | 1 | Naive Bayes | 40295 | 4478 | 3486 | 992 | 77.847253 | 22.152747 | 0.556724 | 0.870379 | 1954 | 0.313927 | 701 | 0.686073 | 1532 | 0.129621 | 291 | 0.73597 | 0.870379 | 164.965 | 16.093 |
| Bi | 1 | 2 | Naive Bayes | 40295 | 4478 | 3532 | 946 | 78.874498 | 21.125502 | 0.57727 | 0.884633 | 1986 | 0.307658 | 687 | 0.692342 | 1546 | 0.115367 | 259 | 0.742985 | 0.884633 | 196.151 | 17.375 |
| Bi | 1 | 3 | Naive Bayes | 40295 | 4478 | 3419 | 1059 | 76.35105 | 23.64895 | 0.526708 | 0.885969 | 1989 | 0.359606 | 803 | 0.640394 | 1430 | 0.114031 | 256 | 0.712393 | 0.885969 | 224.291 | 20.984 |
| Bi | 1 | 4 | Naive Bayes | 40296 | 4477 | 3483 | 994 | 77.797632 | 22.202368 | 0.555676 | 0.884187 | 1985 | 0.328853 | 734 | 0.671147 | 1498 | 0.115813 | 260 | 0.730048 | 0.884187 | 235.838 | 20.312 |
| Bi | 1 | 5 | Naive Bayes | 40296 | 4477 | 3505 | 972 | 78.289033 | 21.710967 | 0.565532 | 0.880178 | 1976 | 0.314964 | 703 | 0.685036 | 1529 | 0.119822 | 269 | 0.737589 | 0.880178 | 206.932 | 18.609 |
| Bi | 1 | 6 | Naive Bayes | 40296 | 4477 | 3500 | 977 | 78.177351 | 21.822649 | 0.563297 | 0.879287 | 1974 | 0.316308 | 706 | 0.683692 | 1526 | 0.120713 | 271 | 0.736567 | 0.879287 | 224.135 | 19.046 |
| Bi | 1 | 7 | Naive Bayes | 40296 | 4477 | 3516 | 961 | 78.534733 | 21.465267 | 0.570465 | 0.876169 | 1967 | 0.306004 | 683 | 0.693996 | 1549 | 0.123831 | 278 | 0.742264 | 0.876169 | 205.604 | 18.687 |
| Bi | 1 | 8 | Naive Bayes | 40296 | 4477 | 3525 | 952 | 78.735761 | 21.264239 | 0.574505 | 0.871269 | 1956 | 0.297043 | 663 | 0.702957 | 1569 | 0.128731 | 289 | 0.74685 | 0.871269 | 218.979 | 19.39 |
| Bi | 1 | 9 | Naive Bayes | 40296 | 4477 | 3550 | 927 | 79.29417 | 20.70583 | 0.585671 | 0.880178 | 1976 | 0.294803 | 658 | 0.705197 | 1574 | 0.119822 | 269 | 0.75019 | 0.880178 | 213.338 | 20.031 |
| Bi | 1 | 10 | Naive Bayes | 40296 | 4477 | 3513 | 964 | 78.467724 | 21.532276 | 0.569139 | 0.885472 | 1987 | 0.316614 | 707 | 0.683386 | 1526 | 0.114528 | 257 | 0.737565 | 0.885472 | 220.089 | 20.218 |
| Bi | 2 | 1 | Naive Bayes | 40295 | 4478 | 3552 | 926 | 79.321126 | 20.678874 | 0.586226 | 0.880624 | 1977 | 0.294671 | 658 | 0.705329 | 1575 | 0.119376 | 268 | 0.750285 | 0.880624 | 215.573 | 19.812 |
| Bi | 2 | 2 | Naive Bayes | 40295 | 4478 | 3464 | 1014 | 77.355962 | 22.644038 | 0.546871 | 0.874833 | 1964 | 0.328258 | 733 | 0.671742 | 1500 | 0.125167 | 281 | 0.728217 | 0.874833 | 172.387 | 17.437 |
| Bi | 2 | 3 | Naive Bayes | 40295 | 4478 | 3465 | 1013 | 77.378294 | 22.621706 | 0.547328 | 0.870824 | 1955 | 0.32378 | 723 | 0.67622 | 1510 | 0.129176 | 290 | 0.730022 | 0.870824 | 176.089 | 17.453 |
| Bi | 2 | 4 | Naive Bayes | 40296 | 4477 | 3549 | 928 | 79.271834 | 20.728166 | 0.585248 | 0.884135 | 1984 | 0.299149 | 668 | 0.700851 | 1565 | 0.115865 | 260 | 0.748115 | 0.884135 | 175.886 | 16.968 |
| Bi | 2 | 5 | Naive Bayes | 40296 | 4477 | 3474 | 1003 | 77.596605 | 22.403395 | 0.55164 | 0.88686 | 1991 | 0.335573 | 749 | 0.664427 | 1483 | 0.11314 | 254 | 0.726642 | 0.88686 | 173.027 | 17.327 |
| Bi | 2 | 6 | Naive Bayes | 40296 | 4477 | 3560 | 917 | 79.517534 | 20.482466 | 0.590141 | 0.88196 | 1980 | 0.292115 | 652 | 0.707885 | 1580 | 0.11804 | 265 | 0.75228 | 0.88196 | 168.34 | 19.28 |
| Bi | 2 | 7 | Naive Bayes | 40296 | 4477 | 3424 | 1053 | 76.479786 | 23.520214 | 0.529288 | 0.876169 | 1967 | 0.347222 | 775 | 0.652778 | 1457 | 0.123831 | 278 | 0.71736 | 0.876169 | 192.542 | 17.875 |
| Bi | 2 | 8 | Naive Bayes | 40296 | 4477 | 3499 | 978 | 78.155015 | 21.844985 | 0.562839 | 0.883296 | 1983 | 0.320789 | 716 | 0.679211 | 1516 | 0.116704 | 262 | 0.734717 | 0.883296 | 182.152 | 17.812 |
| Bi | 2 | 9 | Naive Bayes | 40296 | 4477 | 3511 | 966 | 78.423051 | 21.576949 | 0.568229 | 0.875724 | 1966 | 0.307796 | 687 | 0.692204 | 1545 | 0.124276 | 279 | 0.741048 | 0.875724 | 186.526 | 16.406 |
| Bi | 2 | 10 | Naive Bayes | 40296 | 4477 | 3530 | 947 | 78.847442 | 21.152558 | 0.576708 | 0.885523 | 1988 | 0.30914 | 690 | 0.69086 | 1542 | 0.114477 | 257 | 0.742345 | 0.885523 | 187.355 | 18.093 |
| Bi | 3 | 1 | Naive Bayes | 40295 | 4478 | 3493 | 985 | 78.003573 | 21.996427 | 0.559853 | 0.871715 | 1957 | 0.312136 | 697 | 0.687864 | 1536 | 0.128285 | 288 | 0.737378 | 0.871715 | 187.214 | 21.64 |
| Bi | 3 | 2 | Naive Bayes | 40295 | 4478 | 3534 | 944 | 78.91916 | 21.08084 | 0.578185 | 0.875724 | 1966 | 0.297806 | 665 | 0.702194 | 1568 | 0.124276 | 279 | 0.747244 | 0.875724 | 187.339 | 15.89 |
| Bi | 3 | 3 | Naive Bayes | 40295 | 4478 | 3551 | 927 | 79.298794 | 20.701206 | 0.585767 | 0.885969 | 1989 | 0.300493 | 671 | 0.699507 | 1562 | 0.114031 | 256 | 0.747744 | 0.885969 | 185.386 | 19.578 |
| Bi | 3 | 4 | Naive Bayes | 40296 | 4477 | 3516 | 961 | 78.534733 | 21.465267 | 0.570485 | 0.868151 | 1949 | 0.297939 | 665 | 0.702061 | 1567 | 0.131849 | 296 | 0.745601 | 0.868151 | 195.417 | 19.672 |
| Bi | 3 | 5 | Naive Bayes | 40296 | 4477 | 3508 | 969 | 78.356042 | 21.643958 | 0.566859 | 0.886414 | 1990 | 0.319892 | 714 | 0.680108 | 1518 | 0.113586 | 255 | 0.735947 | 0.886414 | 195.401 | 21.844 |
| Bi | 3 | 6 | Naive Bayes | 40296 | 4477 | 3507 | 970 | 78.333706 | 21.666294 | 0.56644 | 0.875278 | 1965 | 0.30914 | 690 | 0.69086 | 1542 | 0.124722 | 280 | 0.740113 | 0.875278 | 192.371 | 22.483 |
| Bi | 3 | 7 | Naive Bayes | 40296 | 4477 | 3565 | 912 | 79.629216 | 20.370784 | 0.592344 | 0.896659 | 2013 | 0.304659 | 680 | 0.695341 | 1552 | 0.103341 | 232 | 0.747494 | 0.896659 | 197.339 | 19.874 |
| Bi | 3 | 8 | Naive Bayes | 40296 | 4477 | 3467 | 1010 | 77.44025 | 22.55975 | 0.548524 | 0.880624 | 1977 | 0.332437 | 742 | 0.667563 | 1490 | 0.119376 | 268 | 0.727106 | 0.880624 | 188.714 | 19.203 |
| Bi | 3 | 9 | Naive Bayes | 40296 | 4477 | 3461 | 1016 | 77.306232 | 22.693768 | 0.545843 | 0.878842 | 1973 | 0.333333 | 744 | 0.666667 | 1488 | 0.121158 | 272 | 0.726169 | 0.878842 | 200.807 | 20.265 |
| Bi | 3 | 10 | Naive Bayes | 40296 | 4477 | 3514 | 963 | 78.49006 | 21.50994 | 0.569594 | 0.881907 | 1979 | 0.312584 | 698 | 0.687416 | 1535 | 0.118093 | 265 | 0.73926 | 0.881907 | 194.667 | 18.827 |
| Bi | 4 | 1 | Naive Bayes | 40295 | 4478 | 3495 | 983 | 78.048236 | 21.951764 | 0.560733 | 0.877951 | 1971 | 0.31751 | 709 | 0.68249 | 1524 | 0.122049 | 274 | 0.735448 | 0.877951 | 200.041 | 24.015 |
| Bi | 4 | 2 | Naive Bayes | 40295 | 4478 | 3533 | 945 | 78.896829 | 21.103171 | 0.577719 | 0.884187 | 1985 | 0.306762 | 685 | 0.693238 | 1548 | 0.115813 | 260 | 0.743446 | 0.884187 | 204.605 | 17.28 |
| Bi | 4 | 3 | Naive Bayes | 40295 | 4478 | 3485 | 993 | 77.824922 | 22.175078 | 0.556265 | 0.875278 | 1965 | 0.319301 | 713 | 0.680699 | 1520 | 0.124722 | 280 | 0.733757 | 0.875278 | 173.792 | 17.968 |
| Bi | 4 | 4 | Naive Bayes | 40296 | 4477 | 3490 | 987 | 77.953987 | 22.046013 | 0.558824 | 0.878396 | 1972 | 0.319892 | 714 | 0.680108 | 1518 | 0.121604 | 273 | 0.734177 | 0.878396 | 174.136 | 17.266 |
| Bi | 4 | 5 | Naive Bayes | 40296 | 4477 | 3522 | 955 | 78.668751 | 21.331249 | 0.573142 | 0.879733 | 1975 | 0.3069 | 685 | 0.6931 | 1547 | 0.120267 | 270 | 0.742481 | 0.879733 | 171.933 | 16.922 |
| Bi | 4 | 6 | Naive Bayes | 40296 | 4477 | 3517 | 960 | 78.557069 | 21.442931 | 0.570886 | 0.88686 | 1991 | 0.316308 | 706 | 0.683692 | 1526 | 0.11314 | 254 | 0.738228 | 0.88686 | 173.465 | 16.89 |
| Bi | 4 | 7 | Naive Bayes | 40296 | 4477 | 3546 | 931 | 79.204825 | 20.795175 | 0.583875 | 0.882405 | 1981 | 0.298835 | 667 | 0.701165 | 1565 | 0.117595 | 264 | 0.748112 | 0.882405 | 172.824 | 17.843 |
| Bi | 4 | 8 | Naive Bayes | 40296 | 4477 | 3518 | 959 | 78.579406 | 21.420594 | 0.571369 | 0.872606 | 1959 | 0.301523 | 673 | 0.698477 | 1559 | 0.127394 | 286 | 0.744301 | 0.872606 | 172.496 | 17.124 |
| Bi | 4 | 9 | Naive Bayes | 40296 | 4477 | 3512 | 965 | 78.445388 | 21.554612 | 0.568681 | 0.873942 | 1962 | 0.305556 | 682 | 0.694444 | 1550 | 0.126058 | 283 | 0.742057 | 0.873942 | 173.995 | 17.391 |
| Bi | 4 | 10 | Naive Bayes | 40296 | 4477 | 3500 | 977 | 78.177351 | 21.822649 | 0.563319 | 0.887255 | 1991 | 0.324227 | 724 | 0.675773 | 1509 | 0.112745 | 253 | 0.733333 | 0.887255 | 173.542 | 16.922 |
| Bi | 5 | 1 | Naive Bayes | 40295 | 4478 | 3512 | 966 | 78.42787 | 21.57213 | 0.568317 | 0.887305 | 1992 | 0.319301 | 713 | 0.680699 | 1520 | 0.112695 | 253 | 0.736414 | 0.887305 | 170.418 | 16.859 |
| Bi | 5 | 2 | Naive Bayes | 40295 | 4478 | 3497 | 981 | 78.092899 | 21.907101 | 0.561631 | 0.876615 | 1968 | 0.315271 | 704 | 0.684729 | 1529 | 0.123385 | 277 | 0.736527 | 0.876615 | 169.214 | 18.218 |
| Bi | 5 | 3 | Naive Bayes | 40295 | 4478 | 3534 | 944 | 78.91916 | 21.08084 | 0.578177 | 0.879287 | 1974 | 0.301388 | 673 | 0.698612 | 1560 | 0.120713 | 271 | 0.74575 | 0.879287 | 185.136 | 15.734 |
| Bi | 5 | 4 | Naive Bayes | 40296 | 4477 | 3495 | 982 | 78.065669 | 21.934331 | 0.561093 | 0.881907 | 1979 | 0.321093 | 717 | 0.678907 | 1516 | 0.118093 | 265 | 0.73405 | 0.881907 | 178.636 | 18.047 |
| Bi | 5 | 5 | Naive Bayes | 40296 | 4477 | 3557 | 920 | 79.450525 | 20.549475 | 0.588796 | 0.883296 | 1983 | 0.294803 | 658 | 0.705197 | 1574 | 0.116704 | 262 | 0.750852 | 0.883296 | 195.432 | 18.234 |
| Bi | 5 | 6 | Naive Bayes | 40296 | 4477 | 3465 | 1012 | 77.395577 | 22.604423 | 0.547643 | 0.875278 | 1965 | 0.327957 | 732 | 0.672043 | 1500 | 0.124722 | 280 | 0.728587 | 0.875278 | 215.198 | 19.046 |
| Bi | 5 | 7 | Naive Bayes | 40296 | 4477 | 3487 | 990 | 77.886978 | 22.113022 | 0.557494 | 0.873497 | 1961 | 0.316308 | 706 | 0.683692 | 1526 | 0.126503 | 284 | 0.735283 | 0.873497 | 200.073 | 18.781 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bi | 5 | 8 | Naive Bayes | 40296 | 4477 | 3527 | 950 | 78.780433 | 21.219567 | 0.575353 | 0.890423 | 1999 | 0.315412 | 704 | 0.684588 | 1528 | 0.109577 | 246 | 0.739549 | 0.890423 | 202.261 | 19.187 |
| Bi | 5 | 9 | Naive Bayes | 40296 | 4477 | 3474 | 1003 | 77.596605 | 22.403395 | 0.551679 | 0.87216 | 1958 | 0.320789 | 716 | 0.679211 | 1516 | 0.12784 | 287 | 0.732236 | 0.87216 | 201.479 | 18.406 |
| Bi | 5 | 10 | Naive Bayes | 40296 | 4477 | 3506 | 971 | 78.311369 | 21.688631 | 0.565972 | 0.883296 | 1983 | 0.317652 | 709 | 0.682348 | 1523 | 0.116704 | 262 | 0.736627 | 0.883296 | 200.526 | 17.703 |
| Bi | 6 | 1 | Naive Bayes | 40295 | 4478 | 3523 | 955 | 78.673515 | 21.326485 | 0.573254 | 0.880178 | 1976 | 0.30721 | 686 | 0.69279 | 1547 | 0.119822 | 269 | 0.742299 | 0.880178 | 197.245 | 18.219 |
| Bi | 6 | 2 | Naive Bayes | 40295 | 4478 | 3533 | 945 | 78.896829 | 21.103171 | 0.577707 | 0.889532 | 1997 | 0.312136 | 697 | 0.687864 | 1536 | 0.110468 | 248 | 0.741277 | 0.889532 | 212.432 | 20.312 |
| Bi | 6 | 3 | Naive Bayes | 40295 | 4478 | 3546 | 932 | 79.187137 | 20.812863 | 0.58357 | 0.868151 | 1949 | 0.284819 | 636 | 0.715181 | 1597 | 0.131849 | 296 | 0.753965 | 0.868151 | 205.166 | 18.875 |
| Bi | 6 | 4 | Naive Bayes | 40296 | 4477 | 3509 | 968 | 78.378378 | 21.621622 | 0.567336 | 0.874833 | 1964 | 0.307796 | 687 | 0.692204 | 1545 | 0.125167 | 281 | 0.740853 | 0.874833 | 187.323 | 16.312 |
| Bi | 6 | 5 | Naive Bayes | 40296 | 4477 | 3466 | 1011 | 77.417914 | 22.582086 | 0.548101 | 0.871269 | 1956 | 0.323477 | 722 | 0.676523 | 1510 | 0.128731 | 289 | 0.730396 | 0.871269 | 169.027 | 16.906 |
| Bi | 6 | 6 | Naive Bayes | 40296 | 4477 | 3515 | 962 | 78.512397 | 21.487603 | 0.569994 | 0.885523 | 1988 | 0.31586 | 705 | 0.68414 | 1527 | 0.114477 | 257 | 0.73821 | 0.885523 | 165.761 | 15.797 |
| Bi | 6 | 7 | Naive Bayes | 40296 | 4477 | 3493 | 984 | 78.020996 | 21.979004 | 0.560154 | 0.883296 | 1983 | 0.323477 | 722 | 0.676523 | 1510 | 0.116704 | 262 | 0.733087 | 0.883296 | 166.683 | 16.546 |
| Bi | 6 | 8 | Naive Bayes | 40296 | 4477 | 3531 | 946 | 78.869779 | 21.130221 | 0.577145 | 0.889532 | 1997 | 0.312724 | 698 | 0.687276 | 1534 | 0.110468 | 248 | 0.741002 | 0.889532 | 167.246 | 16.109 |
| Bi | 6 | 9 | Naive Bayes | 40296 | 4477 | 3484 | 993 | 77.819969 | 22.180031 | 0.556129 | 0.88196 | 1980 | 0.326165 | 728 | 0.673835 | 1504 | 0.11804 | 265 | 0.731167 | 0.88196 | 165.418 | 16.312 |
| Bi | 6 | 10 | Naive Bayes | 40296 | 4477 | 3444 | 1033 | 76.926513 | 23.073487 | 0.538286 | 0.876114 | 1966 | 0.33811 | 755 | 0.66189 | 1478 | 0.123886 | 278 | 0.722528 | 0.876114 | 164.339 | 15.906 |
| Bi | 7 | 1 | Naive Bayes | 40295 | 4478 | 3505 | 973 | 78.27155 | 21.72845 | 0.565208 | 0.877506 | 1970 | 0.312584 | 698 | 0.687416 | 1535 | 0.122494 | 275 | 0.738381 | 0.877506 | 164.371 | 16.031 |
| Bi | 7 | 2 | Naive Bayes | 40295 | 4478 | 3520 | 958 | 78.606521 | 21.393479 | 0.571913 | 0.879733 | 1975 | 0.308106 | 688 | 0.691894 | 1545 | 0.120267 | 270 | 0.741645 | 0.879733 | 164.246 | 16.187 |
| Bi | 7 | 3 | Naive Bayes | 40295 | 4478 | 3508 | 970 | 78.338544 | 21.661456 | 0.566528 | 0.88686 | 1991 | 0.320645 | 716 | 0.679355 | 1517 | 0.11314 | 254 | 0.735501 | 0.88686 | 159.136 | 15.859 |
| Bi | 7 | 4 | Naive Bayes | 40296 | 4477 | 3522 | 955 | 78.668751 | 21.331249 | 0.573188 | 0.874777 | 1963 | 0.301836 | 674 | 0.698164 | 1559 | 0.125223 | 281 | 0.744407 | 0.874777 | 163.605 | 15.406 |
| Bi | 7 | 5 | Naive Bayes | 40296 | 4477 | 3474 | 1003 | 77.596605 | 22.403395 | 0.551657 | 0.880624 | 1977 | 0.329301 | 735 | 0.670699 | 1497 | 0.119376 | 268 | 0.728982 | 0.880624 | 161.027 | 15.484 |
| Bi | 7 | 6 | Naive Bayes | 40296 | 4477 | 3533 | 944 | 78.914452 | 21.085548 | 0.578041 | 0.889087 | 1996 | 0.31138 | 695 | 0.68862 | 1537 | 0.110913 | 249 | 0.741732 | 0.889087 | 162.418 | 16.531 |
| Bi | 7 | 7 | Naive Bayes | 40296 | 4477 | 3513 | 964 | 78.467724 | 21.532276 | 0.569078 | 0.893987 | 2007 | 0.325269 | 726 | 0.674731 | 1506 | 0.106013 | 238 | 0.734358 | 0.893987 | 163.496 | 15.344 |
| Bi | 7 | 8 | Naive Bayes | 40296 | 4477 | 3485 | 992 | 77.842305 | 22.157695 | 0.556608 | 0.869933 | 1953 | 0.31362 | 700 | 0.68638 | 1532 | 0.130067 | 292 | 0.736148 | 0.869933 | 206.964 | 18.64 |
| Bi | 7 | 9 | Naive Bayes | 40296 | 4477 | 3531 | 946 | 78.869779 | 21.130221 | 0.577186 | 0.873051 | 1960 | 0.296147 | 661 | 0.703853 | 1571 | 0.126949 | 285 | 0.747806 | 0.873051 | 183.292 | 16.765 |
| Bi | 7 | 10 | Naive Bayes | 40296 | 4477 | 3500 | 977 | 78.177351 | 21.822649 | 0.563308 | 0.874833 | 1964 | 0.311828 | 696 | 0.688172 | 1536 | 0.125167 | 281 | 0.738346 | 0.874833 | 166.73 | 15.374 |
| Bi | 8 | 1 | Naive Bayes | 40295 | 4478 | 3436 | 1042 | 76.730683 | 23.269317 | 0.534335 | 0.877951 | 1971 | 0.343932 | 768 | 0.656068 | 1465 | 0.122049 | 274 | 0.719606 | 0.877951 | 161.042 | 15.234 |
| Bi | 8 | 2 | Naive Bayes | 40295 | 4478 | 3520 | 958 | 78.606521 | 21.393479 | 0.571902 | 0.884633 | 1986 | 0.313032 | 699 | 0.686968 | 1534 | 0.115367 | 259 | 0.739665 | 0.884633 | 167.433 | 15.281 |
| Bi | 8 | 3 | Naive Bayes | 40295 | 4478 | 3544 | 934 | 79.142474 | 20.857526 | 0.582621 | 0.89265 | 2004 | 0.310345 | 693 | 0.689655 | 1540 | 0.10735 | 241 | 0.743048 | 0.89265 | 172.605 | 16.062 |
| Bi | 8 | 4 | Naive Bayes | 40296 | 4477 | 3465 | 1012 | 77.395577 | 22.604423 | 0.547621 | 0.883296 | 1983 | 0.336022 | 750 | 0.663978 | 1482 | 0.116704 | 262 | 0.725576 | 0.883296 | 164.683 | 15.281 |
| Bi | 8 | 5 | Naive Bayes | 40296 | 4477 | 3508 | 969 | 78.356042 | 21.643958 | 0.566866 | 0.883742 | 1984 | 0.317204 | 708 | 0.682796 | 1524 | 0.116258 | 261 | 0.736999 | 0.883742 | 165.934 | 15.64 |
| Bi | 8 | 6 | Naive Bayes | 40296 | 4477 | 3496 | 981 | 78.088005 | 21.911995 | 0.561515 | 0.876169 | 1967 | 0.314964 | 703 | 0.685036 | 1529 | 0.123831 | 278 | 0.736704 | 0.876169 | 165.183 | 15.515 |
| Bi | 8 | 7 | Naive Bayes | 40296 | 4477 | 3468 | 1009 | 77.462587 | 22.537413 | 0.548998 | 0.870379 | 1954 | 0.321685 | 718 | 0.678315 | 1514 | 0.129621 | 291 | 0.731287 | 0.870379 | 165.964 | 15.875 |
| Bi | 8 | 8 | Naive Bayes | 40296 | 4477 | 3513 | 964 | 78.467724 | 21.532276 | 0.569122 | 0.876615 | 1968 | 0.307796 | 687 | 0.692204 | 1545 | 0.123385 | 277 | 0.741243 | 0.876615 | 170.996 | 15.796 |
| Bi | 8 | 9 | Naive Bayes | 40296 | 4477 | 3531 | 946 | 78.869779 | 21.130221 | 0.577181 | 0.874833 | 1964 | 0.297939 | 665 | 0.702061 | 1567 | 0.125167 | 281 | 0.747052 | 0.874833 | 165.871 | 15.155 |
| Bi | 8 | 10 | Naive Bayes | 40296 | 4477 | 3533 | 944 | 78.914452 | 21.085548 | 0.578095 | 0.881907 | 1979 | 0.304075 | 679 | 0.695925 | 1554 | 0.118093 | 265 | 0.744545 | 0.881907 | 164.183 | 15.437 |
| Bi | 9 | 1 | Naive Bayes | 40295 | 4478 | 3504 | 974 | 78.249218 | 21.750782 | 0.56475 | 0.88196 | 1980 | 0.31751 | 709 | 0.68249 | 1524 | 0.11804 | 265 | 0.736333 | 0.88196 | 168.136 | 15.375 |
| Bi | 9 | 2 | Naive Bayes | 40295 | 4478 | 3517 | 961 | 78.539527 | 21.460473 | 0.570562 | 0.883742 | 1984 | 0.31348 | 700 | 0.68652 | 1533 | 0.116258 | 261 | 0.739195 | 0.883742 | 164.277 | 15.64 |
| Bi | 9 | 3 | Naive Bayes | 40295 | 4478 | 3518 | 960 | 78.561858 | 21.438142 | 0.571003 | 0.886414 | 1990 | 0.315719 | 705 | 0.684281 | 1528 | 0.113586 | 255 | 0.738404 | 0.886414 | 164.449 | 15.25 |
| Bi | 9 | 4 | Naive Bayes | 40296 | 4477 | 3450 | 1027 | 77.060532 | 22.939468 | 0.540933 | 0.873497 | 1961 | 0.332885 | 743 | 0.667115 | 1489 | 0.126503 | 284 | 0.725222 | 0.873497 | 188.948 | 22.141 |
| Bi | 9 | 5 | Naive Bayes | 40296 | 4477 | 3511 | 966 | 78.423051 | 21.576949 | 0.568221 | 0.878842 | 1973 | 0.310932 | 694 | 0.689068 | 1538 | 0.121158 | 272 | 0.739783 | 0.878842 | 212.182 | 16.141 |
| Bi | 9 | 6 | Naive Bayes | 40296 | 4477 | 3484 | 993 | 77.819969 | 22.180031 | 0.556139 | 0.877951 | 1971 | 0.322133 | 719 | 0.677867 | 1513 | 0.122049 | 274 | 0.732714 | 0.877951 | 171.839 | 15.765 |
| Bi | 9 | 7 | Naive Bayes | 40296 | 4477 | 3520 | 957 | 78.624079 | 21.375921 | 0.572238 | 0.883296 | 1983 | 0.31138 | 695 | 0.68862 | 1537 | 0.116704 | 262 | 0.740478 | 0.883296 | 171.308 | 15.624 |
| Bi | 9 | 8 | Naive Bayes | 40296 | 4477 | 3538 | 939 | 79.026134 | 20.973866 | 0.580279 | 0.889087 | 1996 | 0.30914 | 690 | 0.69086 | 1542 | 0.110913 | 249 | 0.743112 | 0.889087 | 173.355 | 15.671 |
| Bi | 9 | 9 | Naive Bayes | 40296 | 4477 | 3488 | 989 | 77.909314 | 22.090686 | 0.557964 | 0.864588 | 1941 | 0.3069 | 685 | 0.6931 | 1547 | 0.135412 | 304 | 0.739147 | 0.864588 | 166.465 | 16.171 |
| Bi | 9 | 10 | Naive Bayes | 40296 | 4477 | 3524 | 953 | 78.713424 | 21.286576 | 0.574073 | 0.879679 | 1974 | 0.305867 | 683 | 0.694133 | 1550 | 0.120321 | 270 | 0.742943 | 0.879679 | 192.526 | 17.968 |
| Bi | 10 | 1 | Naive Bayes | 40295 | 4478 | 3571 | 907 | 79.745422 | 20.254578 | 0.594707 | 0.889087 | 1996 | 0.294671 | 658 | 0.705329 | 1575 | 0.110913 | 249 | 0.752072 | 0.889087 | 184.019 | 15.213 |
| Bi | 10 | 2 | Naive Bayes | 40295 | 4478 | 3492 | 986 | 77.981242 | 22.018758 | 0.559381 | 0.88196 | 1980 | 0.322884 | 721 | 0.677116 | 1512 | 0.11804 | 265 | 0.733062 | 0.88196 | 174.74 | 15.369 |
| Bi | 10 | 3 | Naive Bayes | 40295 | 4478 | 3504 | 974 | 78.249218 | 21.750782 | 0.564756 | 0.879287 | 1974 | 0.314823 | 703 | 0.685177 | 1530 | 0.120713 | 271 | 0.737393 | 0.879287 | 164.845 | 15.571 |
| Bi | 10 | 4 | Naive Bayes | 40296 | 4477 | 3523 | 954 | 78.691088 | 21.308912 | 0.573602 | 0.874388 | 1963 | 0.301075 | 672 | 0.698925 | 1560 | 0.125612 | 282 | 0.744972 | 0.874388 | 167.081 | 15.415 |
| Bi | 10 | 5 | Naive Bayes | 40296 | 4477 | 3455 | 1022 | 77.172214 | 22.827786 | 0.543161 | 0.877506 | 1970 | 0.334677 | 747 | 0.665323 | 1485 | 0.122494 | 275 | 0.725064 | 0.877506 | 180.669 | 15.893 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bi | 10 | 6 | Naive Bayes | 40296 | 4477 | 3472 | 1005 | 77.551932 | 22.448068 | 0.550765 | 0.879287 | 1974 | 0.328853 | 734 | 0.671147 | 1498 | 0.120713 | 271 | 0.728951 | 0.879287 | 161.932 | 15.611 |
| Bi | 10 | 7 | Naive Bayes | 40296 | 4477 | 3528 | 949 | 78.80277 | 21.19723 | 0.575822 | 0.881514 | 1979 | 0.306004 | 683 | 0.693996 | 1549 | 0.118486 | 266 | 0.743426 | 0.881514 | 171.4 | 15.469 |
| Bi | 10 | 8 | Naive Bayes | 40296 | 4477 | 3481 | 996 | 77.75296 | 22.24704 | 0.5548 | 0.876615 | 1968 | 0.322133 | 719 | 0.677867 | 1513 | 0.123385 | 277 | 0.732415 | 0.876615 | 165.285 | 15.328 |
| Bi | 10 | 9 | Naive Bayes | 40296 | 4477 | 3486 | 991 | 77.864642 | 22.135358 | 0.557031 | 0.879287 | 1974 | 0.322581 | 720 | 0.677419 | 1512 | 0.120713 | 271 | 0.732739 | 0.879287 | 163.126 | 15.578 |
| Bi | 10 | 10 | Naive Bayes | 40296 | 4477 | 3563 | 914 | 79.584543 | 20.415457 | 0.591514 | 0.882799 | 1981 | 0.291536 | 651 | 0.708464 | 1582 | 0.117201 | 263 | 0.75266 | 0.882799 | 160.716 | 15.031 |
| Tri | 1 | 1 | Naive Bayes | 40295 | 4478 | 3535 | 943 | 78.941492 | 21.058508 | 0.578646 | 0.869933 | 1953 | 0.291536 | 651 | 0.708464 | 1582 | 0.130067 | 292 | 0.75 | 0.869933 | 402.743 | 35.796 |
| Tri | 1 | 2 | Naive Bayes | 40295 | 4478 | 3588 | 890 | 80.125056 | 19.874944 | 0.602319 | 0.885523 | 1988 | 0.283475 | 633 | 0.716525 | 1600 | 0.114477 | 257 | 0.758489 | 0.885523 | 398.709 | 35.655 |
| Tri | 1 | 3 | Naive Bayes | 40295 | 4478 | 3528 | 950 | 78.785172 | 21.214828 | 0.575482 | 0.884187 | 1985 | 0.309001 | 690 | 0.690999 | 1543 | 0.115813 | 260 | 0.742056 | 0.884187 | 448.082 | 39.249 |
| Tri | 1 | 4 | Naive Bayes | 40296 | 4477 | 3555 | 922 | 79.405852 | 20.594148 | 0.58789 | 0.887751 | 1993 | 0.300179 | 670 | 0.699821 | 1562 | 0.112249 | 252 | 0.748404 | 0.887751 | 436.364 | 39.843 |
| Tri | 1 | 5 | Naive Bayes | 40296 | 4477 | 3564 | 913 | 79.60688 | 20.39312 | 0.591932 | 0.881514 | 1979 | 0.289875 | 647 | 0.710125 | 1585 | 0.118486 | 266 | 0.753618 | 0.881514 | 445.691 | 36.234 |
| Tri | 1 | 6 | Naive Bayes | 40296 | 4477 | 3595 | 882 | 80.299308 | 19.700692 | 0.605807 | 0.880178 | 1976 | 0.274642 | 613 | 0.725358 | 1619 | 0.119822 | 269 | 0.763229 | 0.880178 | 443.099 | 38.795 |
| Tri | 1 | 7 | Naive Bayes | 40296 | 4477 | 3561 | 916 | 79.53987 | 20.46013 | 0.590595 | 0.879287 | 1974 | 0.288978 | 645 | 0.711022 | 1587 | 0.120713 | 271 | 0.753723 | 0.879287 | 451.754 | 39.812 |
| Tri | 1 | 8 | Naive Bayes | 40296 | 4477 | 3583 | 894 | 80.031271 | 19.968729 | 0.600452 | 0.873942 | 1962 | 0.273746 | 611 | 0.726254 | 1621 | 0.126058 | 283 | 0.762534 | 0.873942 | 442.989 | 33.358 |
| Tri | 1 | 9 | Naive Bayes | 40296 | 4477 | 3606 | 871 | 80.545008 | 19.454992 | 0.61073 | 0.879733 | 1975 | 0.269265 | 601 | 0.730735 | 1631 | 0.120267 | 270 | 0.766693 | 0.879733 | 401.709 | 38.327 |
| Tri | 1 | 10 | Naive Bayes | 40296 | 4477 | 3621 | 856 | 80.880054 | 19.119946 | 0.617442 | 0.892602 | 2003 | 0.275414 | 615 | 0.724586 | 1618 | 0.107398 | 241 | 0.765088 | 0.892602 | 399.536 | 34.796 |
| Tri | 2 | 1 | Naive Bayes | 40295 | 4478 | 3605 | 873 | 80.50469 | 19.49531 | 0.609931 | 0.88196 | 1980 | 0.272279 | 608 | 0.727721 | 1625 | 0.11804 | 265 | 0.76507 | 0.88196 | 399.943 | 34.437 |
| Tri | 2 | 2 | Naive Bayes | 40295 | 4478 | 3540 | 938 | 79.053149 | 20.946851 | 0.580869 | 0.875724 | 1966 | 0.295119 | 659 | 0.704881 | 1574 | 0.124276 | 279 | 0.748952 | 0.875724 | 434.02 | 34.39 |
| Tri | 2 | 3 | Naive Bayes | 40295 | 4478 | 3526 | 952 | 78.740509 | 21.259491 | 0.574615 | 0.87216 | 1958 | 0.297806 | 665 | 0.702194 | 1568 | 0.12784 | 287 | 0.746474 | 0.87216 | 426.77 | 35.28 |
| Tri | 2 | 4 | Naive Bayes | 40296 | 4477 | 3607 | 870 | 80.567344 | 19.432656 | 0.611192 | 0.885918 | 1988 | 0.274966 | 614 | 0.725034 | 1619 | 0.114082 | 256 | 0.764028 | 0.885918 | 394.693 | 33.936 |
| Tri | 2 | 5 | Naive Bayes | 40296 | 4477 | 3566 | 911 | 79.651552 | 20.348448 | 0.592811 | 0.888641 | 1995 | 0.296147 | 661 | 0.703853 | 1571 | 0.111359 | 250 | 0.75113 | 0.888641 | 393.021 | 33.578 |
| Tri | 2 | 6 | Naive Bayes | 40296 | 4477 | 3620 | 857 | 80.857717 | 19.142283 | 0.616993 | 0.880178 | 1976 | 0.263441 | 588 | 0.736559 | 1644 | 0.119822 | 269 | 0.770671 | 0.880178 | 426.38 | 36.155 |
| Tri | 2 | 7 | Naive Bayes | 40296 | 4477 | 3558 | 919 | 79.472861 | 20.527139 | 0.589256 | 0.877951 | 1971 | 0.288978 | 645 | 0.711022 | 1587 | 0.122049 | 274 | 0.75344 | 0.877951 | 426.052 | 34.483 |
| Tri | 2 | 8 | Naive Bayes | 40296 | 4477 | 3558 | 919 | 79.472861 | 20.527139 | 0.58924 | 0.884633 | 1986 | 0.295699 | 660 | 0.704301 | 1572 | 0.115367 | 259 | 0.750567 | 0.884633 | 402.365 | 35.546 |
| Tri | 2 | 9 | Naive Bayes | 40296 | 4477 | 3571 | 906 | 79.763234 | 20.236766 | 0.595083 | 0.873942 | 1962 | 0.279122 | 623 | 0.720878 | 1609 | 0.126058 | 283 | 0.758994 | 0.873942 | 434.927 | 37.499 |
| Tri | 2 | 10 | Naive Bayes | 40296 | 4477 | 3567 | 910 | 79.673889 | 20.326111 | 0.593259 | 0.888196 | 1994 | 0.295251 | 659 | 0.704749 | 1573 | 0.111804 | 251 | 0.751602 | 0.888196 | 432.395 | 56.108 |
| Tri | 3 | 1 | Naive Bayes | 40295 | 4478 | 3558 | 920 | 79.455114 | 20.544886 | 0.588923 | 0.874833 | 1964 | 0.286162 | 639 | 0.713838 | 1594 | 0.125167 | 281 | 0.754514 | 0.874833 | 455.723 | 38.171 |
| Tri | 3 | 2 | Naive Bayes | 40295 | 4478 | 3590 | 888 | 80.169719 | 19.830281 | 0.603232 | 0.87706 | 1969 | 0.274071 | 612 | 0.725929 | 1621 | 0.12294 | 276 | 0.762883 | 0.87706 | 434.957 | 35.25 |
| Tri | 3 | 3 | Naive Bayes | 40295 | 4478 | 3625 | 853 | 80.951318 | 19.048682 | 0.618863 | 0.888641 | 1995 | 0.27004 | 603 | 0.72996 | 1630 | 0.111359 | 250 | 0.767898 | 0.888641 | 426.005 | 35.233 |
| Tri | 3 | 4 | Naive Bayes | 40296 | 4477 | 3559 | 918 | 79.495198 | 20.504802 | 0.589731 | 0.86637 | 1945 | 0.276882 | 618 | 0.723118 | 1614 | 0.13363 | 300 | 0.758876 | 0.86637 | 431.176 | 34.983 |
| Tri | 3 | 5 | Naive Bayes | 40296 | 4477 | 3594 | 883 | 80.276971 | 19.723029 | 0.605332 | 0.892205 | 2003 | 0.287186 | 641 | 0.712814 | 1591 | 0.107795 | 242 | 0.757564 | 0.892205 | 387.021 | 32.718 |
| Tri | 3 | 6 | Naive Bayes | 40296 | 4477 | 3549 | 928 | 79.271834 | 20.728166 | 0.585238 | 0.873942 | 1962 | 0.288978 | 645 | 0.711022 | 1587 | 0.126058 | 283 | 0.752589 | 0.873942 | 397.974 | 35.859 |
| Tri | 3 | 7 | Naive Bayes | 40296 | 4477 | 3594 | 883 | 80.276971 | 19.723029 | 0.60532 | 0.89755 | 2015 | 0.292563 | 653 | 0.707437 | 1579 | 0.10245 | 230 | 0.755247 | 0.89755 | 385.912 | 32.983 |
| Tri | 3 | 8 | Naive Bayes | 40296 | 4477 | 3565 | 912 | 79.629216 | 20.370784 | 0.592377 | 0.882851 | 1982 | 0.290771 | 649 | 0.709229 | 1583 | 0.117149 | 263 | 0.753326 | 0.882851 | 365.475 | 30.781 |
| Tri | 3 | 9 | Naive Bayes | 40296 | 4477 | 3539 | 938 | 79.04847 | 20.95153 | 0.580745 | 0.881514 | 1979 | 0.301075 | 672 | 0.698925 | 1560 | 0.118486 | 266 | 0.746511 | 0.881514 | 360.1 | 31.046 |
| Tri | 3 | 10 | Naive Bayes | 40296 | 4477 | 3564 | 913 | 79.60688 | 20.39312 | 0.591969 | 0.879234 | 1973 | 0.287506 | 642 | 0.712494 | 1591 | 0.120766 | 271 | 0.754493 | 0.879234 | 364.897 | 31.046 |
| Tri | 4 | 1 | Naive Bayes | 40295 | 4478 | 3542 | 936 | 79.097812 | 20.902188 | 0.581765 | 0.875278 | 1965 | 0.293775 | 656 | 0.706225 | 1577 | 0.124722 | 280 | 0.749714 | 0.875278 | 375.303 | 30.608 |
| Tri | 4 | 2 | Naive Bayes | 40295 | 4478 | 3581 | 897 | 79.968736 | 20.031264 | 0.599191 | 0.884187 | 1985 | 0.285266 | 637 | 0.714734 | 1596 | 0.115813 | 260 | 0.757056 | 0.884187 | 359.568 | 30.734 |
| Tri | 4 | 3 | Naive Bayes | 40295 | 4478 | 3554 | 924 | 79.365788 | 20.634212 | 0.58713 | 0.876615 | 1968 | 0.289745 | 647 | 0.710255 | 1586 | 0.123385 | 277 | 0.752581 | 0.876615 | 384.819 | 31.311 |
| Tri | 4 | 4 | Naive Bayes | 40296 | 4477 | 3551 | 926 | 79.316507 | 20.683493 | 0.586123 | 0.878396 | 1972 | 0.292563 | 653 | 0.707437 | 1579 | 0.121604 | 273 | 0.751238 | 0.878396 | 372.709 | 30.905 |
| Tri | 4 | 5 | Naive Bayes | 40296 | 4477 | 3586 | 891 | 80.09828 | 19.90172 | 0.601783 | 0.878842 | 1973 | 0.27733 | 619 | 0.72267 | 1613 | 0.121158 | 272 | 0.761188 | 0.878842 | 400.818 | 35.468 |
| Tri | 4 | 6 | Naive Bayes | 40296 | 4477 | 3616 | 861 | 80.768372 | 19.231628 | 0.615177 | 0.891759 | 2002 | 0.276882 | 618 | 0.723118 | 1614 | 0.108241 | 243 | 0.764122 | 0.891759 | 371.335 | 31.139 |
| Tri | 4 | 7 | Naive Bayes | 40296 | 4477 | 3588 | 889 | 80.142953 | 19.857047 | 0.602669 | 0.882851 | 1982 | 0.280466 | 626 | 0.719534 | 1606 | 0.117149 | 263 | 0.759969 | 0.882851 | 351.709 | 30.265 |
| Tri | 4 | 8 | Naive Bayes | 40296 | 4477 | 3567 | 910 | 79.673889 | 20.326111 | 0.593294 | 0.873497 | 1961 | 0.280466 | 626 | 0.719534 | 1606 | 0.126503 | 284 | 0.758021 | 0.873497 | 349.366 | 30.124 |
| Tri | 4 | 9 | Naive Bayes | 40296 | 4477 | 3591 | 886 | 80.209962 | 19.790038 | 0.604028 | 0.875278 | 1965 | 0.271505 | 606 | 0.728495 | 1626 | 0.124722 | 280 | 0.764294 | 0.875278 | 374.865 | 34.437 |
| Tri | 4 | 10 | Naive Bayes | 40296 | 4477 | 3574 | 903 | 79.830243 | 20.169757 | 0.596419 | 0.891266 | 2000 | 0.295119 | 659 | 0.704881 | 1574 | 0.108734 | 244 | 0.752162 | 0.891266 | 398.115 | 33.718 |
| Tri | 5 | 1 | Naive Bayes | 40295 | 4478 | 3571 | 907 | 79.745422 | 20.254578 | 0.594703 | 0.890869 | 2000 | 0.296462 | 662 | 0.703538 | 1571 | 0.109131 | 245 | 0.751315 | 0.890869 | 399.427 | 30.218 |
| Tri | 5 | 2 | Naive Bayes | 40295 | 4478 | 3589 | 889 | 80.147387 | 19.852613 | 0.602778 | 0.880178 | 1976 | 0.277653 | 620 | 0.722347 | 1613 | 0.119822 | 269 | 0.761171 | 0.880178 | 340.272 | 30.171 |
| Tri | 5 | 3 | Naive Bayes | 40295 | 4478 | 3593 | 885 | 80.236713 | 19.763287 | 0.604563 | 0.88196 | 1980 | 0.277653 | 620 | 0.722347 | 1613 | 0.11804 | 265 | 0.761538 | 0.88196 | 353.491 | 30.389 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tri | 5 | 4 | Naive Bayes | 40296 | 4477 | 3550 | 927 | 79.29417 | 20.70583 | 0.585699 | 0.882353 | 1980 | 0.29691 | 663 | 0.70309 | 1570 | 0.117647 | 264 | 0.749149 | 0.882353 | 347.772 | 30.64 |
| Tri | 5 | 5 | Naive Bayes | 40296 | 4477 | 3602 | 875 | 80.455662 | 19.544338 | 0.608931 | 0.883742 | 1984 | 0.27509 | 614 | 0.72491 | 1618 | 0.116258 | 261 | 0.763664 | 0.883742 | 352.756 | 30.828 |
| Tri | 5 | 6 | Naive Bayes | 40296 | 4477 | 3529 | 948 | 78.825106 | 21.174894 | 0.576279 | 0.877951 | 1971 | 0.301971 | 674 | 0.698029 | 1558 | 0.122049 | 274 | 0.74518 | 0.877951 | 343.679 | 30.327 |
| Tri | 5 | 7 | Naive Bayes | 40296 | 4477 | 3543 | 934 | 79.137816 | 20.862184 | 0.582558 | 0.87216 | 1958 | 0.289875 | 647 | 0.710125 | 1585 | 0.12784 | 287 | 0.751631 | 0.87216 | 350.584 | 30.828 |
| Tri | 5 | 8 | Naive Bayes | 40296 | 4477 | 3602 | 875 | 80.455662 | 19.544338 | 0.608908 | 0.893987 | 2007 | 0.285394 | 637 | 0.714606 | 1595 | 0.106013 | 238 | 0.759077 | 0.893987 | 345.1 | 30.484 |
| Tri | 5 | 9 | Naive Bayes | 40296 | 4477 | 3552 | 925 | 79.338843 | 20.661157 | 0.586586 | 0.871715 | 1957 | 0.285394 | 637 | 0.714606 | 1595 | 0.128285 | 288 | 0.754433 | 0.871715 | 346.226 | 30.296 |
| Tri | 5 | 10 | Naive Bayes | 40296 | 4477 | 3560 | 917 | 79.517534 | 20.482466 | 0.590146 | 0.880178 | 1976 | 0.290323 | 648 | 0.709677 | 1584 | 0.119822 | 269 | 0.753049 | 0.880178 | 341.617 | 30.452 |
| Tri | 6 | 1 | Naive Bayes | 40295 | 4478 | 3580 | 898 | 79.946405 | 20.053595 | 0.598757 | 0.877951 | 1971 | 0.279445 | 624 | 0.720555 | 1609 | 0.122049 | 274 | 0.759538 | 0.877951 | 350.132 | 30.265 |
| Tri | 6 | 2 | Naive Bayes | 40295 | 4478 | 3601 | 877 | 80.415364 | 19.584636 | 0.608124 | 0.890423 | 1999 | 0.282579 | 631 | 0.717421 | 1602 | 0.109577 | 246 | 0.760076 | 0.890423 | 350.038 | 30.515 |
| Tri | 6 | 3 | Naive Bayes | 40295 | 4478 | 3606 | 872 | 80.527021 | 19.472979 | 0.610402 | 0.870379 | 1954 | 0.260188 | 581 | 0.739812 | 1652 | 0.129621 | 291 | 0.770809 | 0.870379 | 345.96 | 30.421 |
| Tri | 6 | 4 | Naive Bayes | 40296 | 4477 | 3588 | 889 | 80.142953 | 19.857047 | 0.602681 | 0.877506 | 1970 | 0.27509 | 614 | 0.72491 | 1618 | 0.122494 | 275 | 0.762384 | 0.877506 | 341.929 | 30.358 |
| Tri | 6 | 5 | Naive Bayes | 40296 | 4477 | 3560 | 917 | 79.517534 | 20.482466 | 0.590164 | 0.872606 | 1959 | 0.282706 | 631 | 0.717294 | 1601 | 0.127394 | 286 | 0.756371 | 0.872606 | 354.101 | 30.39 |
| Tri | 6 | 6 | Naive Bayes | 40296 | 4477 | 3609 | 868 | 80.612017 | 19.387983 | 0.612058 | 0.885969 | 1989 | 0.274194 | 612 | 0.725806 | 1620 | 0.114031 | 256 | 0.764706 | 0.885969 | 349.1 | 30.468 |
| Tri | 6 | 7 | Naive Bayes | 40296 | 4477 | 3558 | 919 | 79.472861 | 20.527139 | 0.589239 | 0.885078 | 1987 | 0.296147 | 661 | 0.703853 | 1571 | 0.114922 | 258 | 0.750378 | 0.885078 | 349.945 | 30.311 |
| Tri | 6 | 8 | Naive Bayes | 40296 | 4477 | 3584 | 893 | 80.053607 | 19.946393 | 0.600864 | 0.889087 | 1996 | 0.28853 | 644 | 0.71147 | 1588 | 0.110913 | 249 | 0.756061 | 0.889087 | 341.944 | 30.281 |
| Tri | 6 | 9 | Naive Bayes | 40296 | 4477 | 3541 | 936 | 79.093143 | 20.906857 | 0.581628 | 0.886414 | 1990 | 0.305108 | 681 | 0.694892 | 1551 | 0.113586 | 255 | 0.745039 | 0.886414 | 372.834 | 30.89 |
| Tri | 6 | 10 | Naive Bayes | 40296 | 4477 | 3546 | 931 | 79.204825 | 20.795175 | 0.583913 | 0.881016 | 1977 | 0.297358 | 664 | 0.702642 | 1569 | 0.118984 | 267 | 0.74858 | 0.881016 | 351.491 | 30.312 |
| Tri | 7 | 1 | Naive Bayes | 40295 | 4478 | 3544 | 934 | 79.142474 | 20.857526 | 0.582642 | 0.883296 | 1983 | 0.30094 | 672 | 0.69906 | 1561 | 0.116704 | 262 | 0.746893 | 0.883296 | 346.163 | 30.39 |
| Tri | 7 | 2 | Naive Bayes | 40295 | 4478 | 3584 | 894 | 80.03573 | 19.96427 | 0.600541 | 0.880624 | 1977 | 0.28034 | 626 | 0.71966 | 1607 | 0.119376 | 268 | 0.759508 | 0.880624 | 341.522 | 30.281 |
| Tri | 7 | 3 | Naive Bayes | 40295 | 4478 | 3567 | 911 | 79.656096 | 20.343904 | 0.592921 | 0.887751 | 1993 | 0.295119 | 659 | 0.704881 | 1574 | 0.112249 | 252 | 0.751508 | 0.887751 | 361.319 | 30.562 |
| Tri | 7 | 4 | Naive Bayes | 40296 | 4477 | 3585 | 892 | 80.075944 | 19.924056 | 0.601369 | 0.87656 | 1967 | 0.275414 | 615 | 0.724586 | 1618 | 0.12344 | 277 | 0.761813 | 0.87656 | 342.116 | 30.406 |
| Tri | 7 | 5 | Naive Bayes | 40296 | 4477 | 3578 | 899 | 79.919589 | 20.080411 | 0.598193 | 0.883296 | 1983 | 0.285394 | 637 | 0.714606 | 1595 | 0.116704 | 262 | 0.75687 | 0.883296 | 358.366 | 30.514 |
| Tri | 7 | 6 | Naive Bayes | 40296 | 4477 | 3632 | 845 | 81.125754 | 18.874246 | 0.622339 | 0.890423 | 1999 | 0.268369 | 599 | 0.731631 | 1633 | 0.109577 | 246 | 0.769438 | 0.890423 | 341.289 | 30.53 |
| Tri | 7 | 7 | Naive Bayes | 40296 | 4477 | 3564 | 913 | 79.60688 | 20.39312 | 0.59191 | 0.890869 | 2000 | 0.299283 | 668 | 0.700717 | 1564 | 0.109131 | 245 | 0.749625 | 0.890869 | 351.475 | 30.906 |
| Tri | 7 | 8 | Naive Bayes | 40296 | 4477 | 3547 | 930 | 79.227161 | 20.772839 | 0.584339 | 0.875724 | 1966 | 0.291667 | 651 | 0.708333 | 1581 | 0.124276 | 279 | 0.751242 | 0.875724 | 341.741 | 30.453 |
| Tri | 7 | 9 | Naive Bayes | 40296 | 4477 | 3577 | 900 | 79.897253 | 20.102747 | 0.597773 | 0.871715 | 1957 | 0.274194 | 612 | 0.725806 | 1620 | 0.128285 | 288 | 0.761775 | 0.871715 | 352.319 | 30.406 |
| Tri | 7 | 10 | Naive Bayes | 40296 | 4477 | 3558 | 919 | 79.472861 | 20.527139 | 0.589268 | 0.873051 | 1960 | 0.28405 | 634 | 0.71595 | 1598 | 0.126949 | 285 | 0.75559 | 0.873051 | 341.444 | 30.437 |
| Tri | 8 | 1 | Naive Bayes | 40295 | 4478 | 3565 | 913 | 79.611434 | 20.388566 | 0.592039 | 0.88196 | 1980 | 0.290193 | 648 | 0.709807 | 1585 | 0.11804 | 265 | 0.753425 | 0.88196 | 349.303 | 30.484 |
| Tri | 8 | 2 | Naive Bayes | 40295 | 4478 | 3574 | 904 | 79.812416 | 20.187584 | 0.596056 | 0.885969 | 1989 | 0.290193 | 648 | 0.709807 | 1585 | 0.114031 | 256 | 0.754266 | 0.885969 | 342.804 | 30.374 |
| Tri | 8 | 3 | Naive Bayes | 40295 | 4478 | 3603 | 875 | 80.460027 | 19.539973 | 0.609014 | 0.89265 | 2004 | 0.283923 | 634 | 0.716077 | 1599 | 0.10735 | 241 | 0.759666 | 0.89265 | 346.132 | 30.358 |
| Tri | 8 | 4 | Naive Bayes | 40296 | 4477 | 3587 | 890 | 80.120616 | 19.879384 | 0.602215 | 0.885523 | 1988 | 0.283602 | 633 | 0.716398 | 1599 | 0.114477 | 257 | 0.758489 | 0.885523 | 351.428 | 30.234 |
| Tri | 8 | 5 | Naive Bayes | 40296 | 4477 | 3552 | 925 | 79.338843 | 20.661157 | 0.586559 | 0.882851 | 1982 | 0.296595 | 662 | 0.703405 | 1570 | 0.117149 | 263 | 0.749622 | 0.882851 | 352.272 | 30.499 |
| Tri | 8 | 6 | Naive Bayes | 40296 | 4477 | 3543 | 934 | 79.137816 | 20.862184 | 0.582545 | 0.877506 | 1970 | 0.295251 | 659 | 0.704749 | 1573 | 0.122494 | 275 | 0.749334 | 0.877506 | 341.804 | 30.454 |
| Tri | 8 | 7 | Naive Bayes | 40296 | 4477 | 3557 | 920 | 79.450525 | 20.549475 | 0.58882 | 0.873051 | 1960 | 0.284498 | 635 | 0.715502 | 1597 | 0.126949 | 285 | 0.755299 | 0.873051 | 344.218 | 30.333 |
| Tri | 8 | 8 | Naive Bayes | 40296 | 4477 | 3571 | 906 | 79.763234 | 20.236766 | 0.595073 | 0.877951 | 1971 | 0.283154 | 632 | 0.716846 | 1600 | 0.122049 | 274 | 0.757203 | 0.877951 | 352.426 | 30.564 |
| Tri | 8 | 9 | Naive Bayes | 40296 | 4477 | 3599 | 878 | 80.388653 | 19.611347 | 0.607609 | 0.874833 | 1964 | 0.267473 | 597 | 0.732527 | 1635 | 0.125167 | 281 | 0.766888 | 0.874833 | 355.602 | 30.438 |
| Tri | 8 | 10 | Naive Bayes | 40296 | 4477 | 3590 | 887 | 80.187626 | 19.812374 | 0.60359 | 0.884135 | 1984 | 0.280788 | 627 | 0.719212 | 1606 | 0.115865 | 260 | 0.759862 | 0.884135 | 351.544 | 30.234 |
| Tri | 9 | 1 | Naive Bayes | 40295 | 4478 | 3609 | 869 | 80.594015 | 19.405985 | 0.611713 | 0.885078 | 1987 | 0.273623 | 611 | 0.726377 | 1622 | 0.114922 | 258 | 0.764819 | 0.885078 | 351.366 | 30.452 |
| Tri | 9 | 2 | Naive Bayes | 40295 | 4478 | 3585 | 893 | 80.058062 | 19.941938 | 0.600983 | 0.882851 | 1982 | 0.282132 | 630 | 0.717868 | 1603 | 0.117149 | 263 | 0.758806 | 0.882851 | 343.569 | 30.64 |
| Tri | 9 | 3 | Naive Bayes | 40295 | 4478 | 3575 | 903 | 79.834748 | 20.165252 | 0.596496 | 0.889087 | 1996 | 0.29288 | 654 | 0.70712 | 1579 | 0.110913 | 249 | 0.753208 | 0.889087 | 353.585 | 30.108 |
| Tri | 9 | 4 | Naive Bayes | 40296 | 4477 | 3535 | 942 | 78.959124 | 21.040876 | 0.578979 | 0.871715 | 1957 | 0.293011 | 654 | 0.706989 | 1578 | 0.128285 | 288 | 0.749521 | 0.871715 | 343.085 | 30.593 |
| Tri | 9 | 5 | Naive Bayes | 40296 | 4477 | 3606 | 871 | 80.545008 | 19.454992 | 0.610731 | 0.879287 | 1974 | 0.268817 | 600 | 0.731183 | 1632 | 0.120713 | 271 | 0.7669 | 0.879287 | 353.319 | 30.468 |
| Tri | 9 | 6 | Naive Bayes | 40296 | 4477 | 3532 | 945 | 78.892115 | 21.107885 | 0.577625 | 0.876615 | 1968 | 0.299283 | 668 | 0.700717 | 1564 | 0.123385 | 277 | 0.746586 | 0.876615 | 343.178 | 30.406 |
| Tri | 9 | 7 | Naive Bayes | 40296 | 4477 | 3599 | 878 | 80.388653 | 19.611347 | 0.607584 | 0.885523 | 1988 | 0.278226 | 621 | 0.721774 | 1611 | 0.114477 | 257 | 0.761978 | 0.885523 | 350.584 | 30.437 |
| Tri | 9 | 8 | Naive Bayes | 40296 | 4477 | 3580 | 897 | 79.964262 | 20.035738 | 0.599068 | 0.891759 | 2002 | 0.293011 | 654 | 0.706989 | 1578 | 0.108241 | 243 | 0.753765 | 0.891759 | 343.366 | 30.421 |
| Tri | 9 | 9 | Naive Bayes | 40296 | 4477 | 3541 | 936 | 79.093143 | 20.906857 | 0.581674 | 0.867706 | 1948 | 0.28629 | 639 | 0.71371 | 1593 | 0.132294 | 297 | 0.752996 | 0.867706 | 350.928 | 30.437 |
| Tri | 9 | 10 | Naive Bayes | 40296 | 4477 | 3592 | 885 | 80.232298 | 19.767702 | 0.604488 | 0.882799 | 1981 | 0.278549 | 622 | 0.721451 | 1611 | 0.117201 | 263 | 0.761045 | 0.882799 | 341.366 | 30.937 |
| Tri | 10 | 1 | Naive Bayes | 40295 | 4478 | 3636 | 842 | 81.196963 | 18.803037 | 0.62378 | 0.889978 | 1998 | 0.266458 | 595 | 0.733542 | 1638 | 0.110022 | 247 | 0.770536 | 0.889978 | 350.287 | 30.359 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tri | 10 | 2 | Naive Bayes | 40295 | 4478 | 3599 | 879 | 80.370701 | 19.629299 | 0.607246 | 0.882405 | 1981 | 0.275414 | 615 | 0.724586 | 1618 | 0.117595 | 264 | 0.763097 | 0.882405 | 348.116 | 30.234 |
| Tri | 10 | 3 | Naive Bayes | 40295 | 4478 | 3579 | 899 | 79.924073 | 20.075927 | 0.598303 | 0.881069 | 1978 | 0.283027 | 632 | 0.716973 | 1601 | 0.118931 | 267 | 0.757854 | 0.881069 | 346.944 | 30.405 |
| Tri | 10 | 4 | Naive Bayes | 40296 | 4477 | 3584 | 893 | 80.053607 | 19.946393 | 0.600886 | 0.879733 | 1975 | 0.279122 | 623 | 0.720878 | 1609 | 0.120267 | 270 | 0.7602 | 0.879733 | 351.272 | 30.39 |
| Tri | 10 | 5 | Naive Bayes | 40296 | 4477 | 3499 | 978 | 78.155015 | 21.844985 | 0.562847 | 0.880178 | 1976 | 0.317652 | 709 | 0.682348 | 1523 | 0.119822 | 269 | 0.73594 | 0.880178 | 350.835 | 30.359 |
| Tri | 10 | 6 | Naive Bayes | 40296 | 4477 | 3519 | 958 | 78.601742 | 21.398258 | 0.571796 | 0.881069 | 1978 | 0.309588 | 691 | 0.690412 | 1541 | 0.118931 | 267 | 0.741102 | 0.881069 | 342.007 | 30.624 |
| Tri | 10 | 7 | Naive Bayes | 40296 | 4477 | 3577 | 900 | 79.897253 | 20.102747 | 0.597749 | 0.88196 | 1980 | 0.284498 | 635 | 0.715502 | 1597 | 0.11804 | 265 | 0.75717 | 0.88196 | 350.976 | 30.53 |
| Tri | 10 | 8 | Naive Bayes | 40296 | 4477 | 3575 | 902 | 79.85258 | 20.14742 | 0.596856 | 0.881069 | 1978 | 0.284498 | 635 | 0.715502 | 1597 | 0.118931 | 267 | 0.756984 | 0.881069 | 356.976 | 30.421 |
| Tri | 10 | 9 | Naive Bayes | 40296 | 4477 | 3547 | 930 | 79.227161 | 20.772839 | 0.584339 | 0.875724 | 1966 | 0.291667 | 651 | 0.708333 | 1581 | 0.124276 | 279 | 0.751242 | 0.875724 | 345.679 | 30.421 |
| Tri | 10 | 10 | Naive Bayes | 40296 | 4477 | 3621 | 856 | 80.880054 | 19.119946 | 0.617464 | 0.881016 | 1977 | 0.263771 | 589 | 0.736229 | 1644 | 0.118984 | 267 | 0.77046 | 0.881016 | 350.444 | 30.327 |

## G. J48 cross validation

| Fold | Training inst. | Testing inst. | Correct | Incorrect | Correct % | Incorrect % | Kappa | TP rate | TP | FP rate |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 40295 | 4478 | 4270 | 208 | 95.355069 | 4.644931 | 0.907105 | 0.945212 | 2122 | 0.038065 |
| 2 | 40295 | 4478 | 4276 | 202 | 95.489058 | 4.510942 | 0.909781 | 0.953229 | 2140 | 0.043439 |
| 3 | 40295 | 4478 | 4254 | 224 | 94.997767 | 5.002233 | 0.899957 | 0.946548 | 2125 | 0.046574 |
| 4 | 40296 | 4477 | 4263 | 214 | 95.220013 | 4.779987 | 0.9044 | 0.950557 | 2134 | 0.046147 |
| 5 | 40296 | 4477 | 4278 | 199 | 95.555059 | 4.444941 | 0.911104 | 0.94833 | 2129 | 0.037186 |
| 6 | 40296 | 4477 | 4262 | 215 | 95.197677 | 4.802323 | 0.903956 | 0.945657 | 2123 | 0.041667 |
| 7 | 40296 | 4477 | 4260 | 217 | 95.153004 | 4.846996 | 0.903063 | 0.945657 | 2123 | 0.042563 |
| 8 | 40296 | 4477 | 4276 | 201 | 95.510386 | 4.489614 | 0.91021 | 0.949666 | 2132 | 0.039427 |
| 9 | 40296 | 4477 | 4288 | 189 | 95.778423 | 4.221577 | 0.915572 | 0.949666 | 2132 | 0.03405 |
| 10 | 40296 | 4477 | 4294 | 183 | 95.912441 | 4.087559 | 0.918248 | 0.959893 | 2154 | 0.041648 |

| FP | TN rate | TN | FN rate | FN | Precision | Recall | Training time | Testing time | Treesize | Leaves | Rules |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 85 | 0.961935 | 2148 | 0.054788 | 123 | 0.961486 | 0.945212 | 24553.23 | 0.343 | 1985 | 993 | 993 |
| 97 | 0.956561 | 2136 | 0.046771 | 105 | 0.956638 | 0.953229 | 25149.759 | 0.25 | 2051 | 1026 | 1026 |
| 104 | 0.953426 | 2129 | 0.053452 | 120 | 0.953342 | 0.946548 | 20112.858 | 0.218 | 1961 | 981 | 981 |
| 103 | 0.953853 | 2129 | 0.049443 | 111 | 0.953956 | 0.950557 | 25004.916 | 0.296 | 1947 | 974 | 974 |
| 83 | 0.962814 | 2149 | 0.05167 | 116 | 0.962477 | 0.94833 | 27514.809 | 0.188 | 2015 | 1008 | 1008 |
| 93 | 0.958333 | 2139 | 0.054343 | 122 | 0.958032 | 0.945657 | 24019.041 | 0.266 | 1993 | 997 | 997 |
| 95 | 0.957437 | 2137 | 0.054343 | 122 | 0.957169 | 0.945657 | 25175.055 | 0.281 | 2071 | 1036 | 1036 |
| 88 | 0.960573 | 2144 | 0.050334 | 113 | 0.96036 | 0.949666 | 28380.453 | 1 | 1933 | 967 | 967 |
| 76 | 0.96595 | 2156 | 0.050334 | 113 | 0.96558 | 0.949666 | 21027.328 | 0.188 | 1989 | 995 | 995 |
| 93 | 0.958352 | 2140 | 0.040107 | 90 | 0.958611 | 0.959893 | 18658.028 | 0.187 | 2003 | 1002 | 1002 |

## H. Abbrivations

| | |
|---|---|
| EPR | Electronic Patient Record |
| RHF | Rikshospitalet Helseforetak |
| NEL | Norsk Elektronisk Legehåndbok |
| IR | Information retrieval |
| RQ | Research Question |
| NTNU | The Norwegian University of Science and Technology |
| KE | Knowledge Engineering |
| ML | Machine Learning |
| TF | Term Frequency |
| IDF | Inverse Document Frequency |
| ICPC | International Classification of Primary Care |
| WEKA | Waikato Environment for Knowledge |
| SVM | Support Vector Machines |
| ROC | Receiver Operating characteristic |
| PHR | Personal Health Records |

# Bibliography


1. *Lov om pasientrettigheter (pasientrettighetsloven)*, in *LOV-1999-07-02-63* H.H.-o. omsorgsdepartementet), Editor. 1999.

2. Fowles, J.B., et al., *Patients' interest in reading their medical record: relation with clinical and sociodemographic characteristics and patients' approach to health care.* Arch Intern Med, 2004. **164**(7): p. 793-800.

3. Ivanova, E., *Automatic adaptation of information in electronic patient records for patients*, in *IDI*. 2006, NTNU: Trondheim.

4. Nossum, V., *Automatisk oversettelse av pasientjournaler*. 2007, Rikshospitalet HF.

5. Stallemo, K., *TDT4540, Patient Friendly presentation of Electronic Patient Records*. 2007, NTNU.

6. *Dokumentasjon av sykepleie i elektronisk pasientjournal*. 2007, [Oslo]: Norsk sykepleierforbund. 80 s.

7. Zeng, Q.T. and T. Tse, *Exploring and developing consumer health vocabularies.* J Am Med Inform Assoc, 2006. **13**(1): p. 24-9.

8. Sjøberg, D.I.K., T. Dybå, and M. Jørgensen, *The Future of Empirical Methods in Software Engineering Research.* Future of Software Engineering (FOSE '07), 2007.

9. Wang, A.I., R. Conradi, and ESERNET, *Empirical methods and studies in software engineering experiences from ESERNET*. Lecture notes in computer science 2765. 2003, Berlin: Springer. VIII, 278 s.

10. Shaw, M., *Writing Good Software Engineering Research Papers.* 25th International Conference on Software Engineering (ICSE'03) 2003.

11. Schütze, M., *Collocations*, in *Foundations of Statistical Natural Language Processing*.

12. Tan, A.-H., *Text Mining: The state of the art and the challenges.* 1999.

13. Sebastiani, D., *Machine Learning in Automated Text Categorization.* ACM Computing Surveys, 2002. **34**(1): p. 47.

14. Fan, W., et al., *Tapping into the Power of Text Mining.* ACM, 2005.

15. Witten, I.H. and E. Frank, *Data mining practical machine learning tools and techniques*. 2nd ed. The Morgan Kaufmann series in data management systems. 2005, Amsterdam: Elsevier. XXXI, 525 s.

16. Yong-feng, S. and Z. Yan-ping, *Comparison of text categorization algorithms.* Wuhan University Journal of Natural Sciences, 2004. **9**(5): p. 798-804.

17. Rennie, J., et al., *Tackling the poor assumptions of Naive Bayes text classifiers.* Rennie, J.D.M., Shih, L., Teevan, J., Karger, D.: Tackling the poor assumptions of Naive Bayes text classifiers. In Fawcett, T., Mishra, N., eds.: Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington, D.C., AAAI Press (2003) 616--623, 2003.

18. McCallum, A. and K. Nigam, *A comparison of event models for Naive Bayes text classification.* A. McCallum and K. Nigam. A comparison of event models for Naive Bayes text classification. In AAAI-98 Workshop on Learning for Text Categorization, 1998., 1998.

19. Berger, H. and D. Merkl, *A Comparison of Text-Categorization Methods Applied to N -Gram Frequency Statistics*, in *AI 2004: Advances in Artificial Intelligence*. 2005. p. 998-1003.

20. Joachims, T. *Text categorization with support vector machines: learning with many relevant features*. in *Proceedings of {ECML}-98, 10th European Conference on Machine Learning*. 1998.

21. Hsu, C.W., C.C. Chang, and C.J. Lin, *A practical guide to support vector classification*. 2003, Department of Computer Science and Information Engineering, National Taiwan University.

22. James, M. and M. Paul, *Single n-gram stemming*, in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval %@ 1-58113-646-3*. 2003, ACM: Toronto, Canada. p. 415-416.

23. Cavnar, W.B. and J.M. Trenkle, *N-Gram-Based Text Categorization.* Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval, 1994.

24. Batista, G., R. Prati, and M. Monard, *A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data.* Batista, G.E.A.P.A., Prati, R.C., Monard, M.C.: A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. SIGKDD Explorations 6 (2004) (to appear). 2004.

25. Baeza-Yates, R. and B. Ribeiro-Neto, *Modern information retrieval*. ACM Press books. 1999, New York: ACM Press. XX, 513 s.

26. Walpole, R.E., *Probability & statistics for engineers & scientists*. 7th ed. 2002, Upper Saddle River, N.J.: Prentice-Hall. XVI, 730 s.

27. Provost, F. and T. Fawcett, *Robust Classification for Imprecise Environments*. 2000.

28. Hahn, U. and I. Mani, *The challenges of Automatic Summarization.* IEEE Computer Society Press, 2000. **33**(11): p. 29-36.

29. Rose, Ø., *Text Mining in Health Records*, in *Department of Computer and Information Science*. 2007, Norwegian University of Science and Technology: Trondheim.

30. Røst, T.B., Ø. Nytrø, and A. Grimsmo, *Investigating the Value of Diagnosis Codes in the Primary Care Patient Record*, Department of Computer and Information Science and The Norwegian EHR Research Center, Norwegian University of Science and Technology: Trondheim.

31. Letrilliart, L., et al., *Automatic coding of reasons for hospital referral from general medicine free-text reports.* Proc AMIA Symp, 2000: p. 487-91.

32. Røst, T.B., Ø. Nytrø, and A. Grimsmo, *Classifying Encounter Notes in the Parimary Care Patient Record*, Department of Computer and Information Science and The Norwegian EHR Research Center, Norwegian University of Science and Technology: Trondheim.

33. Vale, R.F., et al., *Improving Text Retrieval in Medical Collections Through Automatic Categorization*, in *String Processing and Information Retrieval*. 2003. p. 197-1|0.

34. Sibanda, T., et al., *Syntactically-informed semantic category recognition in discharge summaries.* AMIA Annu Symp Proc, 2006: p. 714-8.

35. Musen, M.A., J.C. Helder, and J.H.v. Bemmel, *Handbook of medical informatics*. 1997, Heidelberg: Springer. XXXIV, 621 s., s. XXV-XL.

36. Marill, J.L., N. Miller, and P. Kitendaugh, *The MedlinePlus public user interface: studies of design challenges and opportunities.* J Med Libr Assoc, 2006. **94**(1): p. 30-40.

37. Cimino, J.J., V.L. Patel, and A.W. Kushniruk, *The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records.* Int J Med Inform, 2002. **68**(1-3): p. 113-27.

38. Al-Busaidi, A., A. Gray, and N. Fiddian, *Personalizing web information for patients: linking patient medical data with the web via a patient personal knowledge base.* Health Informatics J, 2006. **12**(1): p. 27-39.

39. Kim, M.I. and K.B. Johnson, *Personal Health Records: Evaluation of Functionality and Utility.* Journal of the American Informatics Association, 2002. **9**.

40. Kim, M.I. and K.B. Johnson, *Patient Entry of Information: Evaluation of User Interfaces.* Journal of Medical Internet Research, 2004. **6**(2).

41. Shneiderman, B. and C. Plaisant, *Designing the user interface: strategies for effective human-computer interaction*. 2005, Boston: Addison-Wesley. XVIII, 652 s.

42. Pomikalek, J. and R. Rehurek, *The Influence of preprocessing parameters on text categorization.* PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY, 2007. **21**.

43. Furnkranz, J., *A Study Using n -gram Features for Text Categorization*, Austrian Research Institute for Artificial Intelligence.

44. Georgios, P., et al., *Learning Rules for Large-Vocabulary Word Sense Disambiguation: A Comparison of Various Classifiers*, in *Proceedings of the Second International Conference on Natural Language Processing*. 2000, Springer-Verlag.

45. Mhamdi, F., R. Rakotomalala, and M. Elloumi, *A Compromise between N-gram Length and Classifier Characteristics for Protein Classification.* International Journal of Computer Science and Network Security, 2006. **6**(4).

46. Honeyman, A., B. Cox, and B. Fisher, *Potential impacts of patient access to their electronic care records.* Inform Prim Care, 2005. **13**(1): p. 55-60.

47. Pagliari, C., D. Detmer, and P. Singleton, *Potential of electronic personal health records.* BMJ, 2007. **335**(7615): p. 330-3.