NTNU

Norwegian University of
Science and Technology

Enhanced lithological description of the Jurassic sequence in the Viking Graben and Oseberg Terrace using machine learning

Harald Wie Bøe

Abstract

Automatic classifications of well logs using machine learning techniques has gained improved attention within the last couple of years among others for increasing the accuracy and speed of lithofacies prediction on wireline logs. A supervised machine learning methodology written in R combines initial wireline log signatures and computer assisted interpretations (CPI's) to automatize lithofacies predictions. The method utilizes the XGBoost algorithm, a gradient boosting library with emphasis on computational speed and model accuracy. The study area is the Jurassic sequence of the prolific quadrant 30 in the Norwegian Northern North Sea, with 26 available wells that consists of wells from structural Highs, Grabens and Terraces.

Raw well log data are preprocessed in order to obtain a consistent database and the well log signatures from all 26 wells are analyzed concurrently with the use of boxplots. Seven different lithoclasses are generated based on the CPI's which are used as training data and validation data. Three different machine learning models (initial, normalized and enhanced) are created and optimalized. The initial model is created using only the raw logs without modifications and the normalized model is created using standardized logs. The enhanced model is created with several features and modifications added to increase prediction accuracy. The models are applied to 11 wells and key differences between the models in terms of accuracy are presented.

The results show that the initial model provides a decent baseline accuracy while the normalized model shows signs of underfitting. The enhanced model shows great promise in accuracy. The number of well logs available within a well has a direct influence on the accuracy of the models. For an acceptable predicting power the wells should minimum contain the gamma ray, density and neutron logs. To distinguish between water-bearing and hydrocarbon-bearing zones in sandstones the resistivity logs should be present. The choice of petrophysical filters when creating the lithoclasses has potential and will have more predicting power with improved refinement according to the aim of the study. When implemented on a larger scale caution need to be taken when introducing more geologic complexity in the input data.

Machine Learning shows good promise in assessing wireline logs and the ML design in terms of training data should account for lateral variations such as depositional environment and varying burial depth to achieve optimal prediction accuracy.

Sammendrag

Automatisk klassifisering av brønnlogger ved bruk av maskinlæringsteknikker har fått økt oppmerksomhet de siste par årene blant annet for å øke nøyaktigheten og hastigheten til litofacies prediksjoner på wireline logger. En veiledet maskinlæringsmetode skrevet i R kombinerer wireline signaturer og datamaskinassisterte tolkninger (CPI'er) til automatiske litofacies prediksjoner. Metoden benytter XGBoost-algoritmen, et "gradient boosting" bibliotek med særlig vekt på beregningshastighet og modellnøyaktighet. Studien er utført i Jurassic-sekvensen i den produktive kvadrant 30 i Nordlige Nordsjøen, med 26 tilgjengelige brønner hvor Jura-intervallet befinner seg på strukturelle graben, høyder og terrasser.

Råbrønnloggdata er forhåndsbehandlet for å oppnå en konsistent database og brønnloggsignaturer fra alle 26 brønner analyseres samtidig med bruk av boksplot. Syv forskjellige litoklasser genereres basert på CPI'ene og brukes som treningsdata og valideringsdata. Tre forskjellige maskinlæringsmodeller (initiell, normalisert og forbedret) genereres og optimaliseres. Den opprinnelige modellen er generert kun ved å bruke de rå loggene og den normaliserte modellen er generert ved hjelp av standardiserte logger. Den forbedrede modellen er generert med ekstra egenskaper og modifikasjoner lagt til for å øke prediksjonsnøyaktigheten. Modellene er anvendt på 11 brønner og viktige forskjeller mellom modellene med hensyn til nøyaktighet presenteres.

Resultatene viser at den opprinnelige modellen gir en adekvat innledende nøyaktighet, den normaliserte modellen viser tegn på "underfitting" mens den forbedrede modellen viser en kraftig økning i nøyaktighet. Antall brønnlogger tilgjengelig i en brønn har direkte innflytelse på nøyaktigheten av modellene. For en akseptabel prediksjonskraft bør brønnene minst inneholde gammastråling, tetthets- og nøytronlogger. For å skille mellom vannbærende og oljebærende soner i sandstein bør resistivitetsloggene være tilstede. Valget av petrofysiske filtre når litoklassene blir laget har potensiale og prediksjonene vil ha økt prediksjonskraft etter en raffinering i henhold til formålet med studien. Når metoden blir implementert på større skala må dette gjøres med varsomhet med tanke på at dette kan introdusere økt geologisk kompleksitet i inputdataene. Maskinlæring viser godt potensial i å anvendes på logger og designet av ML modellen i hensyn til treningsdata bør ta høyde for laterale variasjoner i avsetningsmiljø og varierende begravningsdyp for å oppnå optimal prediksjonsnøyaktighet.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The oil industry has a tradition of quickly embracing up-and-coming technologies, if they are seen as beneficial to the industry. Machine learning (ML) is emerging as a new tool which can be applied to both seismic data and welldata to speed up routine or repetetive tasks. Within seismics, recent work has focused on automatizing salt classifications (Waldeland and Solberg 2017) and horizon/structure interpretations (geoforskning (2017), GEOExPro (2017)) by using deep learning. ML has also been used for seismic petrofacies characterization in the SW Barents Sea (Honoré et al. 2018). For wells the focus has been on automatizing lithofacies classifications using well logs with a competition by SEG in 2016 (Hall (2016), (Hall and Hall 2017)) generating large amounts of publicity. Additionally, ML has been used to automate porosity and permeability estimations based on well logs (Larsen (2018), Total (2018)). CGG released in June 2018 a new version of it's powerlog software which contains machine learning capabilities (CGG 2018) such as facies predictions shown in Figure 1.1. The release proves that ML is highly relevant in the oil industry today.

The aim of this thesis is to create a model for predicting lithofacies within the Jurassic sequence on specific wells in quadrant 30 by using supervised machine learning. For this purpose well log data from 26 wells together with Computer Processed Interpretations (CPI's) of the wells performed by Exploro are used. I show how to analyze the well log distributions from all 26 wells concurrently with the use of boxplots and generate seven different lithoclasses based on the CPI's. I then create three different machine learning models (initial, normalized and enhanced) by using the XGBoost algorithm and apply the models on nine wells. Only raw well logs and additional modifications of them are used when training and applying the models, the use of the CPI is limited to lithoclasses generation and model validation.

A brief introduction to ML and its uses within the oil industry together with an overview of gradient boosting is presented in the first part. The study area, well database and geologic setting is presented in the following chapters. How the seven different lithoclasses are created

based on a well selection from Exploro's Toolkit part one (Marello et al. 2016) is then shown followed by a filtering routine to remove bad data and the use of boxplots. The filtering routine and the use of boxplots are both included to gain additional knowledge of the training and test data, which is vital for generating a good ML model. Key differences between the ML models are then presented along with how they are created and optimized. The results in the following chapter focus on model accuracy and how the models differ in terms of which lithoclasses they predict. 100 meters of logging interval from well 30/9-14 and well 30/11-8 A is presented to show how the ML models differ on a smaller scale.
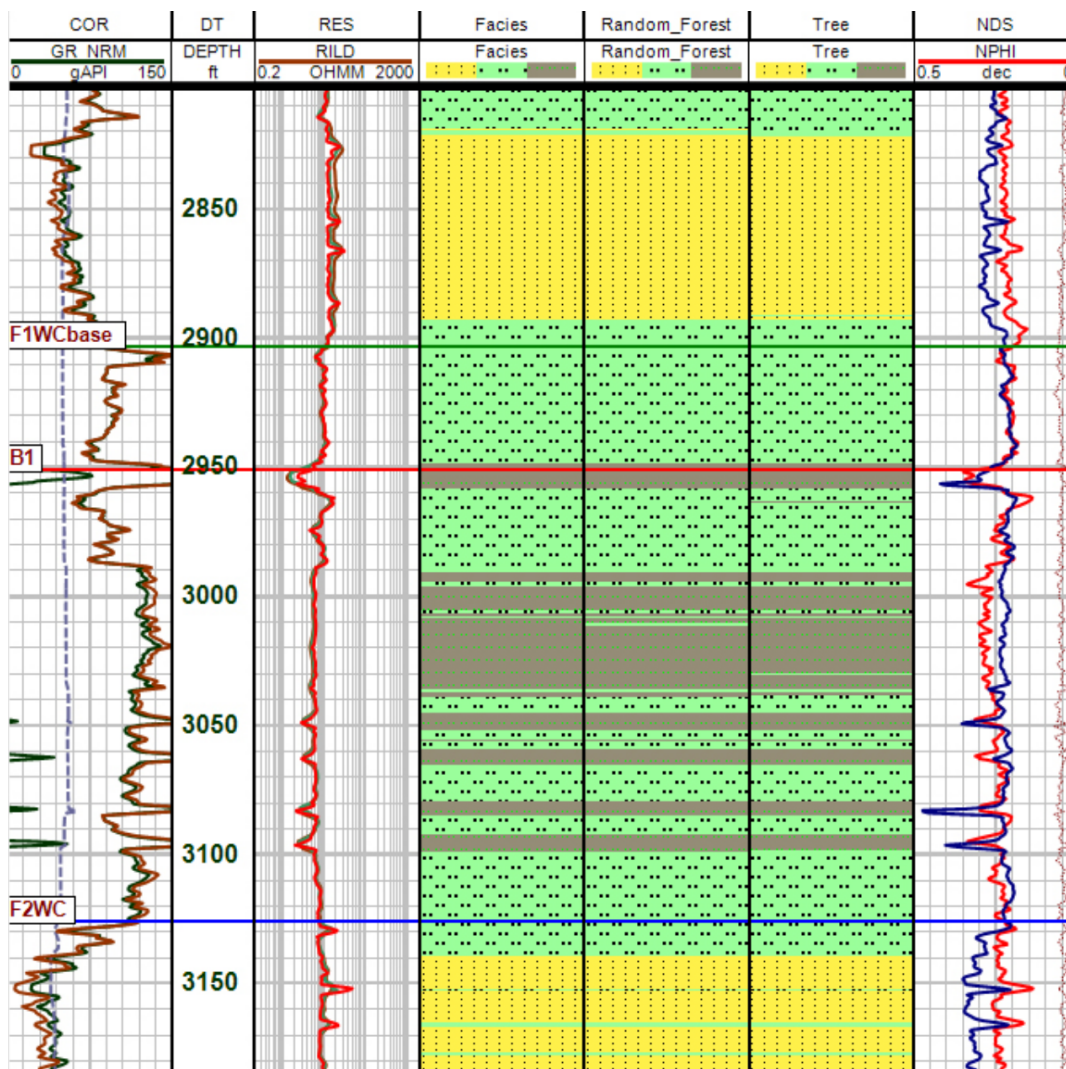
Figure 1.1: CGG advertisement photo of its software update to Techlog in June 2018 which introduces Machine learning-based facies predictions. Image courtesy of CGG GeoSoftware (CGG 2018).

Chapter 2

Theoretical Background

2.1   A brief introduction to Machine learning

Machine learning (ML) is not something new. Algorithms used in machine learning today such as logistic regression (Cox 1958), k-nearest-neighbour (Cover and Hart 1967), artificial neural networks (Werbos 1974), decision trees (Quinlan 1986) and random decision forest (Ho 1995) has been around for several decades.

Despite the availability of algorithms which can be used in ML, the popularity and usage of ML has in the past been relatively low. However, in recent years this has changed dramatically. Major E&P companies such as Total, ConocoPhillips, Equinor and AkerBP is investing heavily into digitalization (Worldoil (2018), Equinor (2018), O&G-Journal (2017), Cognite (2018)), and ML is a major part of this digitalization effort.

The rise of ML in recent years among the oil companies is therefore a product of a broader digitalization shift within the oil industry, and there are several influental factors stemming from the digitalization effort which has led to the recent increase in popularity of ML:

- **Processing power** is no longer as expensive and limited as it once was.

- **Big data** is more accessible than ever. Together with the increase in computing power analysis can quickly be performed on the full dataset, not only on samples as done in the past. The quality of the data has also increased, leading to better models and better results from ML (Bryant et al. (2008), Frankel and Reid (2008)).

- **Internet of Things (IoT)** has emerged in recent years and resulted in a major increase of data generation from sensors and measurement tools. New tools and analytical methods are needed to interpret and make decisions based on the data and ML is one of these methods (Gubbi et al. 2013).

- **High-dimensional challenges** requires new methods and tools to get optimal results. While humans can only visualize and interpret up to three dimensions, a dataset can have thousands, even millions of variables. Analysis of such large amounts of data by a human is not a realistic approach.

- **High influx of open-source software and libraries** in the recent years has led to ML becoming more accessible to the masses.

One of the most important aspects when building an ML model is the data used as training data. The data available has in the recent years increased dramatically, and this has been a significant contribution to the emergence of ML in recent years.

There are generally three main types of machine learning (Christopher 2016): Supervised, unsupervised and reinforcement learning. Supervised and unsupervised learning will be expanded upon in the following sections while reinforcement learning has seen little use within the oil & gas industry and is therefore omitted.

2.1.1 Supervised learning

In supervised ML the user supplies the system with training data which contains multiple features and is labeled. The system learns patterns from the training data and classifies/labels unseen data in a best guess based on the training data.
Supervised learning is therefore task driven, meaning it is given a single task which it performs.

Supervised ML is categorized into two subcategories, regression and classification problems. Regression problems are where the output is continous, such as given several wireline logs (e.g. sonic, density and resistivity), an ML model may be tasked to predict the porosity or permeability within an interval. Classification problems is where the output is discrete, e.g. given the same wireline logs as in a regression problem, now an ML model may be tasked to classify/identify the facies within an interval.

A recently presented regional reservoir quality distribution (facies classification) in the Barents Sea by using ML shows the merit in applying this methodology in the geoscience industry (Larsen 2018) The results allowed them to utilize the large volume of data available, and gave them 'More freedom and time to be creative and collaborate across disciplines' (Larsen 2018).

2.1.2 Unsupervised learning

In unsupervised learning, the machine must realize by itself how to group and categorize clusters of data, where it attempts to find patterns and structure in a unlabeled dataset. The structure is derived by clustering the data based on relationships among the variables within the data (Christopher 2016). An advantage with this approach is that it is possible to approach problems with little or no idea what the results should look like. Unsupervised learning is therefore data driven unlike supervised learning which is task driven. It lets the data speaks for itself, and has no clear objective on how it should classify or cluster the data.

An example of a unsupervised learning algorithm is a self-organizing map (SoM, Kohonen (1982)). A SoM is a type of artificial neural network which returns grouped/clustered data based on identified similarities within the dataset.

Statoil has applied SoM on seismic from the Oseberg field within quadrant 30 in the Northern North Sea, where the goal was to find disconnected sandstone channels within the Ness formation (Thurmond 2018). According to John Thurmond, instead of finding the sandstone channels they actually found the oil water contact within the formation. He also mentioned that one of the challenges with this approach is that the resulting data clusters still required interpretation (Thurmond 2018).

2.2 Gradient boosting

Gradient boosting (Breiman 1997) is a machine learning technique for regression and classification problems which builds a prediction model based on several weak prediction models. Building a ML model upon several ML models is called an ensemble method or ensemble model, and the idea behind an ensemble method is that by combining multiple learning algorithms the result is a model which has higher prediction/classification accuracy than a single model (Opitz and Maclin 1999).

The term boosting comes from the fact that the algorithm converts weak learners to strong learners. A weak learner is defined as beeing slightly better than random guessing when classifying unseen data, while a strong learner is seen as performing considerably better than a weak learner when exposed to unseen data (Zhou 2012). In terms of statistics, a weak learner can be seen as having "some" correlation with the true classification, while a strong learner can be seen as having a good correlation with the true classification thus having a higher prediction accuracy (Zhou 2012).

This thesis utilizes XGBoost (Chen and Guestrin 2016), a gradient tree boosting system as the machine learning method. An introduction to key elements of the algorithm is therefore in order and they will be outlined in the following section.

2.2.1 Decision trees and gradient tree boosting

There are two main types of decision trees in terms of machine learning: Classification trees and regression trees.

- Classification trees are models where the prediction variable is within a discrete set of values (e.g. deciding if a point within a well log is sandstone or shale).

- Regression trees are models where the prediction variable is continous, such as predicting the porosity on a certain point within a well based on the values of the well logs.

A collective term for the two types of decision trees are Classification And Regression Tree (CART) (Breiman 2017).

Gradient tree boosting, also known as gradient boosting machine (GBM) is a highly effective and widely used machine learning method. It utilizes multiple decision trees in the form of an ensemble model to make predictions about a target variable. The following compressed mathematical introduction to gradient tree boosting is modified from (Chen and Guestrin 2016).

For a given data set with n examples and m features $\mathscr{D} = \{(\mathbf{x}_i, y_i)\}$ ($|\mathscr{D}| = n, \mathbf{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R}$), a tree ensemble model uses K additive functions to predict the target variable

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \ f_k \in F, \tag{2.1}$$

where

$$F = \{f(\mathbf{x}) = w_{q(\mathbf{x})}\}(q : \mathbb{R} \to T, w \in \mathbb{R}^T) \tag{2.2}$$

is the space or set of all possible CART's. $\hat{y}_i$ is the predicted target variable, q is a representation of the structure of each tree that maps an example to its respective leaf index and T is the number of leaves in the tree. Each f_k corresponds to an independent tree structure q and leaf weights w.

The tree ensemble model corrects what has been learned, then adds one new tree at a time. To decide the set of functions f_k the ensemble model should consist of, the following *regularized* objective is minimized:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{2.3}$$

where

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2. \tag{2.4}$$

l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i. Ω punishes the complexity of the model, γ is the regularization parameter and also the minimum the loss reduction to make a further partition on a leaf node of the tree. Increasing γ will make the algorithm more conservative and if γ is set to 0, no regularization occurs. λ is a $L2$ regularization term on weights, increasing this value will make the model more conservative (less prone to overfitting).

Eq 2.3 is minimized to learn the set of functions used in the model. In general the optimal model is a model which is both simple and predictive. By optimalizing one of the terms, the

other term is increased. This is also known as the bias-variance tradeoff in ML.

Because the tree ensemble model shown in Eq. 2.1 contains functions as parameters they can not be optimized using traditional optimization methods in Euclidean space (Chen and Guestrin 2016). The model is instead trained in an additive manner by using a second order approximation to optimize the objective in Eq. 2.3. The mathematical steps involved can be seen in Appendix B.

Chapter 3

Geologic setting

The study area is located in the Northern North Sea and the stratigraphical time of interest is the Jurassic. Figure 3.1 shows a structural map of the Northern North Sea at base Brent group comprising of both relative structural highs (Terraces) and lows (Grabens), the study area (quadrant 30) is marked in red. The information within this chapter is mainly based on the article by Sneider et al. (1995).

The North Sea basin has been formed during two major rift phases: The Permian-Triassic and the Middle Jurassic to Early Cretaceous (Ziegler 1982). In Figure 3.2 the Permian/Triassic rift can be seen within the Horda Platform and the Middle to Upper Jurassic rift visible in the North Viking Graben. The Permian/Triassic rift can be seen within the Horda Platform in Figure 3.2 with the Middle to Upper Jurassic rift in the North Viking Graben.

The study area comprises of the Viking Graben which is a north-south oriented rift system and relative structural highs located to the east and west of it (Fig 3.1). East of the Viking Graben the narrow Oseberg Terrace is located. The E. Shetland Terrace, which is west of the Viking Graben consists of rotated fault blocks. Further to the east is the E. Shetland Platform. Most of the wells used in this study are drilled on the Oseberg Terrace in relation to the Oseberg Discovery, whereas few of the wells are drilled within the North Viking Graben. As seen in Fig. 3.2 the depth of the Jurassic interval is deeper within the North Viking Graben in relation to the Oseberg Terrace. This difference in burial depth is expected to have an impact on the differences between the wells in regards to logging data such as density and porosity.

The location of two wells, 30/9-14 and 30/11-8A is shown on Figure 3.1. Logging sections with machine learning results from these wells will be presented in detail. Well 30/9-14 is drilled where the Jurassic interval is at a structural high (Oseberg Terrace) while well 30/11-8A is drilled where the Jurassic interval is at a structural low (Viking Graben).

Figure 3.3 shows the lithostratigraphic chart for the Norwegian North Sea of Jurassic age (NPD 2018). The Lower Jurassic strata consists predominantly of marine shales while the Middle

Jurassic strata consists of sandy marginal marine to deltaic sediments (Brent Group) overlain by marine shales in the Heather Formation of the Viking Group. The Upper Jurassic consists mostly of marine shales with coarser clastic sequences developing locally. The Brent Group consists of five formations from bottom to top: Broom Formation, Rannoch Formation, Etive Formation, Ness Formation and the Tarbert Formation. The sedimentary rocks of the Brent Group are considered to be deposits of a major river-delta system (Helland-Hansen et al. 1992).

Major oil & gas discoveries have been made on both sides of the Viking Graben and are shown in Figure 4.1 in the following chapter.
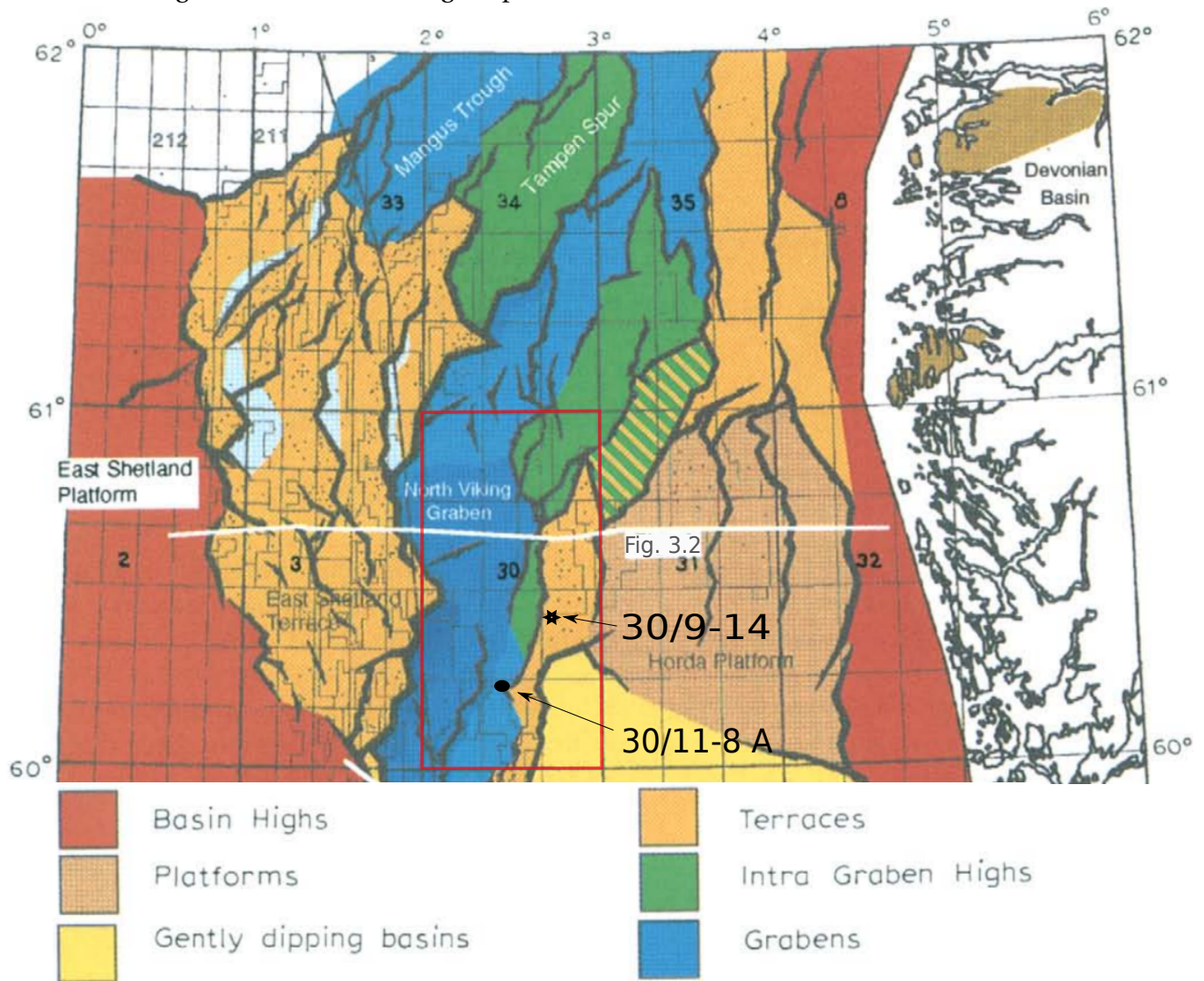
Figure 3.1: Structure map of Northern North Sea between $60-62$ degrees at base Brent Group with study area marked in red (quadrant 30). A W-E line indicates the location of a regional seismic line in Fig. 3.2. Logging sections from wells 30/9-14 and 30/11-8A is presented in the results. Figure modified from Sneider et al. (1995).
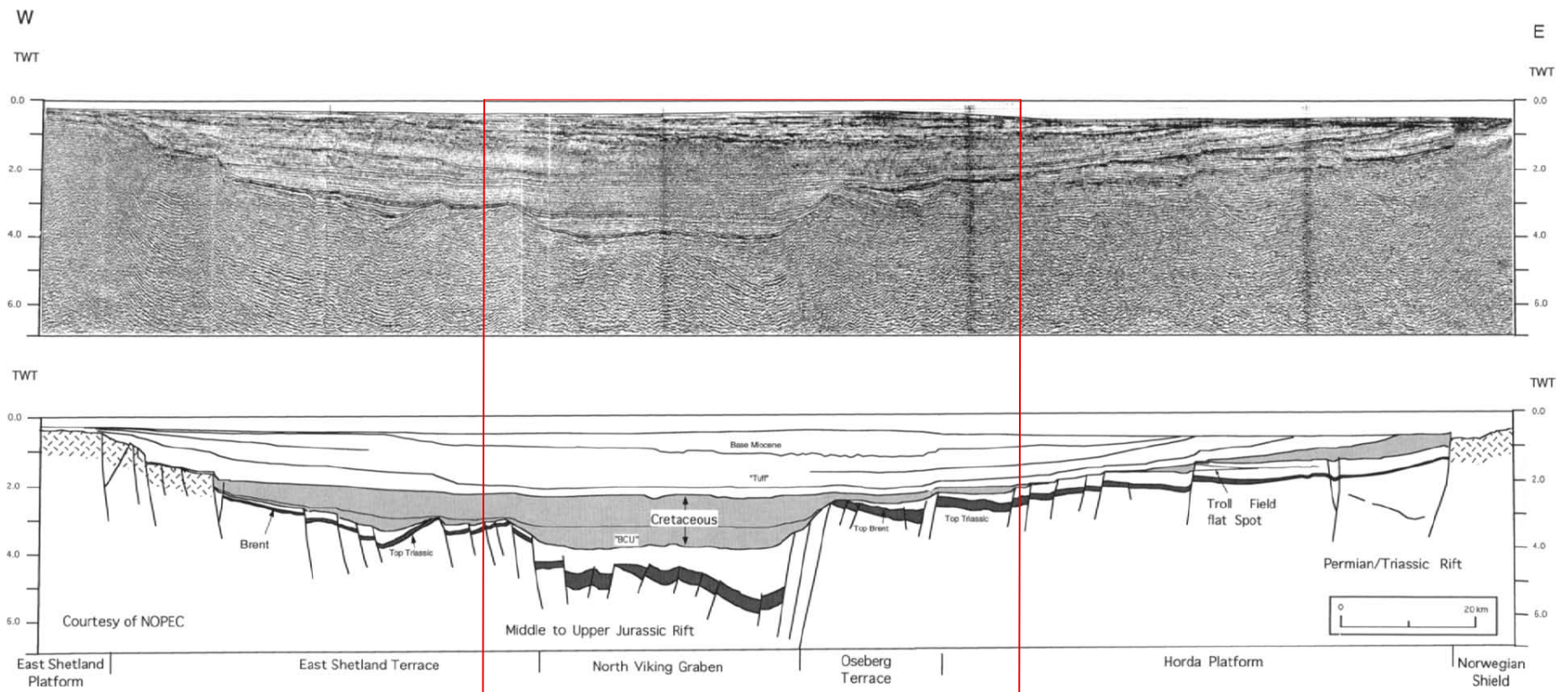
Figure 3.2: Regional W-E seismic line across the northern part of the Viking Graben as shown in Fig. 3.1. The study area (quadrant 30) is within the red box. Many of the wells used in this study are from the Oseberg Terrace, with a minority of the wells within the North Viking Graben itself. Figure modified from Sneider et al. (1995)

.

LITHOSTRATIGRAPHIC CHART
NORWEGIAN NORTH SEA

NPD

2014
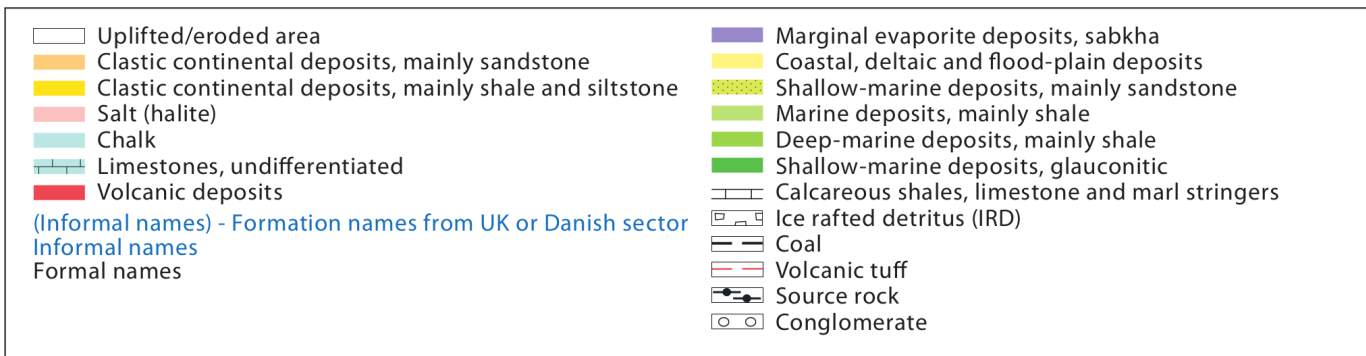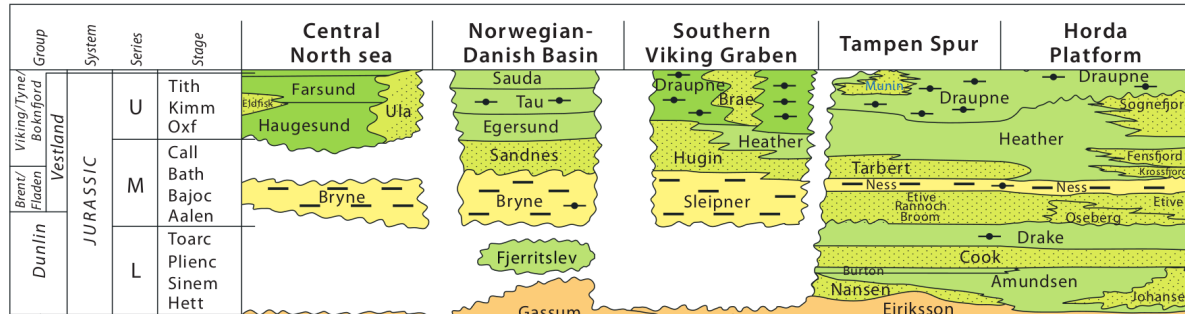
Figure 3.3: Lithostratigraphic chart of the Norwegian North Sea showing the geological Groups and Formations of Jurassic age. The geological Formations in the study area is below Tampen Spur. The chart is modified from NPD (2018).

Chapter 4

Database

4.1 Data available

For this project 26 wells with Computer Processed Interpretations (CPI) from quadrant 30 in the Northern North Sea are made available for the study. The wells are extracted from and are part of Exploro's North Sea Toolkit - part 1 (Marello et al. 2016).

Figure 4.1 shows the positions of the wells within quadrant 30 in the Northern North Sea, along with the fields and discoveries in the vicinity. The blue diamonds are wells with CPI performed by Exploro. A large proportion of the wells are drilled within the Oseberg and Oseberg south fields. NOAKA (North of Alvheim Krafla Askja) is a group of multiple smaller discoveries currently in development phase by AkerBP and Statoil and is marked by the red circles, about half of these discoveries are within quadrant 30.

Table 4.1 contains information about the CPI wells, such as the Jurassic interval for each well and when the well was completed. The data is from the factpages of the Norwegian Petroleum Directorate (NPD 2018), and the top and bottom Jurassic interval as well as the jurassic thickness has been calculated using the data available. The top and bottom Jurassic is calculated by selecting the Viking GP, Brent GP, Dunlin GP and Vestland GP. The next step is then to take the maximum and minimum value within these groups for each well. The thickness is the difference between maximum and minimum.

The table shows that most of these 26 analyzed wells contained hydrocarbons and only three wells are dry wells. The top Jurassic depth and the total Jurassic thickness vary greatly between the wells and is reflecting the lateral geologic variations outlined in figures 3.1 and 3.2. There is also a large time difference between the first and last completed date, with the first well completed in 1975 and the last well completed in 2011, a time difference of 36 years.
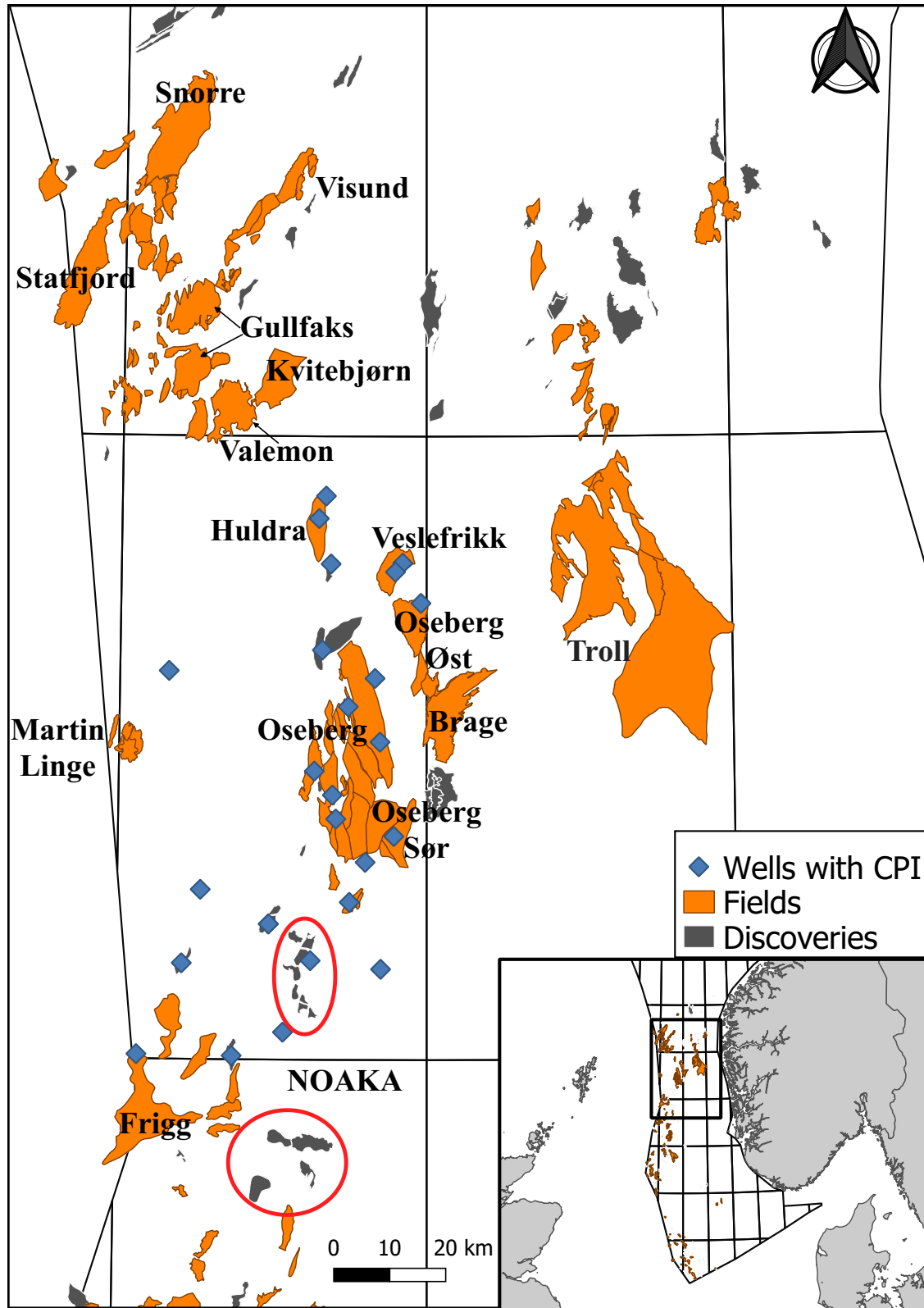
Figure 4.1: Map of Northern North Sea outside Norway where the study is performed, with most of the significant fields named. Wells with CPI provided by Exploro and used in this study is shown in blue. NOAKA is a group of multiple discoveries currently in development phase and is marked in red. Fields and discoveries polygons from NPD (2018).

Table 4.1: Wells with Computer-processed interpretation (CPI) available for the study. The depths and thicknesses are in meters and is given by measured depth (MD).

ID	Name	Top Jurassic	Btm Jurassic	Thickness	Content	Completed
1	30/2-1	3636	4110	474	gas/condensate	12.10.1982
2	30/3-1 R	3729	4227	498	shows	26.04.1982
3	30/3-2 R	2714	3228	514	oil/gas	16.02.1981
4	30/3-7 B	4261	5970	1709	oil/gas	04.08.1998
5	30/3-10 S	3708	4034	326	gas/condensate	29.04.2009
6	30/4-1	4387	5454	1067	dry	14.05.1979
7	30/5-3 A	4650	4746	96	gas	30.05.2009
8	30/6-4	2415	2942	527	oil/gas	11.05.1981
9	30/6-6	2587	3225	638	dry	24.03.1982
10	30/6-22	2497	3290	793	oil	13.07.1988
11	30/6-28 S	2376	2483	107	oil	29.03.2012
12	30/7-7	3878	4884	1006	gas shows	01.07.1979
13	30/8-1 S	3361	4688	1327	gas/condensate	01.03.1995
14	30/9-5 S	2236	2604	368	gas	19.07.1985
15	30/9-14	2968	3680	712	oil/gas	14.05.1993
16	30/9-16	2652	3457	805	oil/gas	08.08.1994
17	30/9-20 S	2750	3124	374	oil	11.02.2002
18	30/9-21 S	3597	4090	493	oil/gas	30.04.2008
19	30/10-5	4162	5049	887	oil/gas	01.05.1975
20	30/10-6	4344	5248	904	gas	09.11.1992
21	30/11-3	3285	4637	1352	oil/gas shows	14.03.1983
22	30/11-6 S	2961	3550	589	shows	02.07.2004
23	30/11-7	3722	4067	345	gas/condensate	03.02.2009
24	30/11-8 A	3671	4475	804	oil/gas/condensate	03.07.2011
25	30/11-8 S	3474	4043	569	oil/gas/condensate	20.05.2011
26	30/12-1	2795	3596	801	dry	07.03.1994

4.2   Data preparation

The wells will be separated into two groups where one group will be used as training data for the algorithm, and the other group will be used as test data. All practical work is done with the programming language R (R Core Team 2018) in combination with Rstudio (RStudio Team 2018), a free and open-source integrated development environment (IDE) for R. Several additional libraries is also used in R, and a complete list of libraries used can be found in Appendix A.

The first step is now to load the CPI files into the R environment and merge the wells into a single, consistent dataset. All wells are in the .las ASCII format, and the internal overall structure of the las files are consistent with a header of variable length depending on the number of variables followed by the well log data. By utilizing the consistent internal structure of the .las files a script is written which automatically loads all files within a single folder into a list and assigns the correct column names for each well.

Additional data from NPD factpages (NPD 2018) such as casing info, welltops (lithostratigraphic information) and mud type used is also merged into the dataset. The Jurassic interval is extracted from each well by using the welltops provided by NPD in the same way as Table 4.1 was created. The casing info from NPD is originally in a string format displaying inches, and needs to be converted into a numeric form, which is done by using multiple regular expressions (Thompson 1968). This information can then be used as an additional large-scale quality control by checking the casing diameter against the caliper log.

Altough the structure of each file are consistent, the column/log names are not. The technology used when logging has changed dramatically since the pioneering days of the 1960's-1970's, and this is reflected in the log names. The various oil service companies which has performed the logging has used different names for the same measurements, and as the methods of logging has changed, so has the log naming with it. It is therefore necessary to merge together columns which has the same measurements, before it is possible to use the dataset in a Machine Learning setting. The assumption for performing the merge is that tools which measure the same property but has different names give approximately the same measurements and that the difference between them is insignificant. This is a necessary database simplification in order to

use the wells for Machine learning.

Table 4.2 shows the different tools available, the abbreviation used for each tool and which tools that will be used for ML. Since the companies performing the well logging uses slightly different tools, they have also developed their own log abbreviations. During loading and the subsequent merging of well log data this is a challenge since the resulting merged dataset consists of 30 − 40 additional redundant columns. This has been accounted for during the preprocessing by identifying different column names for the same measurement and then merging the data from the different abbreviation columns into the main respective column. The redundant columns are removed from the dataset.

The result of the merge is that the number of variables are drastically reduced without losing any information/data. The resulting dataset has a higher degree of consistency and is compact in terms of columns. The different tools have consistent names and the next step in the preprocessing can begin. The full dataset consists of 115.536 rows.

Table 4.2: Logs available in the study, abbreviations and which logs are beeing used for ML purposes. Additional columns with other abbreviation names are renamed in order to get a consistent dataset with less variables.

Log	Abbreviation	Used in ML
Caliper	CALI	
Gamma Ray	GR	x
Spontaneous Potential	SP	x
Bulk Density	DEN	x
Density Correction	DENC	
Delta-T Compressional (sonic)	AC	x
Delta-T Shear (ShearSonic)	ACS	x
Neutron Porosity	NEU	x
Formation Photoelectric Factor	PEF	x
Shallow Laterolog Resistivity	LLS	x
Medium Resistivity	RMED	x
Deep Resistivity	RDEP	x
Nominal Bit Size	BS	
Rate of penetration	ROP	

Chapter 5

Methodology

5.1 Defining lithoclasses

The CPI performed on the wells contains information about the lithologies, e.g. percent sandstone, silt, wet clay, as well as information about the total and effective porosity and the water saturation. Based on the lithology percentages it is possible to create lithoclasses by dividing the percentages into several intervals, where each interval is assigned a lithoclass.

The starting point for defining the lithoclasses is the Classification of Fine-grained Sedimentary rocks from Picard (1971). Volume wet clay (VWCL) is used together with water saturation and coal flags to define the lithoclasses. This assumes that the Jurassic interval consists of different fractions of sand and mudstone as well as coal. Table 5.1 shows descriptions of the different lithoclasses chosen along with their abbreviations and which logs are used in their determination. Volume wet clay is used as a classifier for all classes except coal. This is done in order to prevent classifying the same row multiple times, and to maintain consistency. To distinguish between hydrocarbon and water filled sandstones water saturation (SW) is used as an additional classifier. Moving from sandstone with commercial hydrocarbons (Table 5.1) down to mudstone the clay content is gradually increased and three different petrophysical filters is tested before deciding on a model. Mud is used as a general term for clay and silt content. Deciding on the petrophysical filters is a balancing act. From a reservoir perspective a lower VWCL is optimal as it normally gives better reservoir properties, while from a sedimentological perspective shifting the percentage higher for additional generalization is better.

Since the well interval is limited to the Jurassic age in the Northern North Sea there is a low amount of carbonate sediments (Ramberg 2008). Thus excluding a carbonate class is not expected to make a significant difference. An additional group, carbonate "stingers" is discovered in the data, defined by having a density at 3.0 g/cm^3 or more and GR around 50. However there were so few datapoints (approximately 70 meaning $\approx$ 10m cumulative thickness) that grouping

them into a distinct group is not viable.

The CPI lithology interpretations is made up of different percentages of the lithologies which sums up to 1. Effective porosity is included this percentage, and the lithology columns must therefore first be normalized to get correct lithology intervals. The normalization is done with the following equation

$$lithology_{nm} = \frac{lithology}{1 - \phi_e},$$

(5.1)

where $lithology_{nm}$ is the normalized lithology and ϕ_e is the effective porosity. Eq. 5.1 is applied on the VWCL and the resulting normalized VWCL together with water saturation (SW) is used to generate the lithoclasses.

Three different petrophysical filters is tested before deciding on a model, and the different limits are shown in Table 5.2. The main difference between the models is the upper limit for VWCL content in sandstones, with maximum 15% for model A, 20% in model B and 25% in model C. The VWCL content in the other classes are adjusted accordingly, to have a large enough percentage interval for each class. The upper limit for the sandstone containing uncommercial amounts of hydrocarbons is set relatively conservative to $SW = 0.9$ in order to prevent false classifications.

Figure 5.1 shows the lithoclass distribution within the dataset based on the petrophysical filters in Table 5.2. The distribution differences between the models can be seen in the different sandstone classes and the sandy mudstone class. The waterfilled sandstone is doubled from around 4000 points to approximately 8000 points when moving the upper limit from 15% to 25%

Table 5.1: Overview of the different lithofacies classes implemented by using data from the CPI. Note that most the classes use volume wet clay in order to maintain consistency and prevent assigning multiple classes to a single row.

ID	Lithofacies description	Abbreviation	CPI used
1	Sandstone	SS	VWCL
2	Sandstone with uncommercial hydrocarbons	SSUHC	VWCL & SW
3	Sandstone with commercial hydrocarbons	SSCHC	VWCL & SW
4	Muddy sandstone	MUDSS	VWCL
5	Sandy mudstone	SMUD	VWCL
6	Mudstone	MUDST	VWCL
7	Coal	COAL	Coal flag

and the hydrocarbon filled sandstone classes also see a major increase. The higher VWCL limits for the sandstone classes will cause the treshold for the algorithm to classify the different sandstone intervals instead of a mudstone interval to go down. The result of this lowered treshold is that the ML model will have increased leniency in its sandstone classifications and the number of sandstone related class predictions is expected to increase. This is an important decision that will directly influence how the ML model should behave. By using a low upper limit such as 15% there is a risk of missing intervals containing HC, while setting the limit too high can cause the algorithm to falsely classify uneconomic intervals as economic. A high limit is chosen because it is expected to be easier to correct for uneconomic intervals than to actively look for intervals which has not been picked up when reviewing the results. The model should be used as a tool to speed up lithoclass decisions so that it is not necessary to create it from scratch but rather to review the results from the ML model.

Table 5.2: Petrophysical filters used when dividing the dataset into lithological classes. Coal is not included here as the CPI gives a coal flag (1/0 for each row with data).

Model	CPI	SS	SSUHC	SSCHC	MUDSS	SMUDST	MUDST
A	VWCL	$< 15\%$	$< 15\%$	$< 15\%$	$15 - 30\%$	30-75%	$> 75\%$
A	SW		$60 - 90\%$	$< 60\%$			
B	VWCL	$< 20\%$	$< 20\%$	$< 20\%$	$20 - 35\%$	$35 - 75\%$	$> 75\%$
B	SW		$60 - 90\%$	$< 60\%$			
C	VWCL	$< 25\%$	$< 25\%$	$< 25\%$	$25 - 40\%$	$40 - 75\%$	$> 75\%$
C	SW		$60 - 90\%$	$< 60\%$			
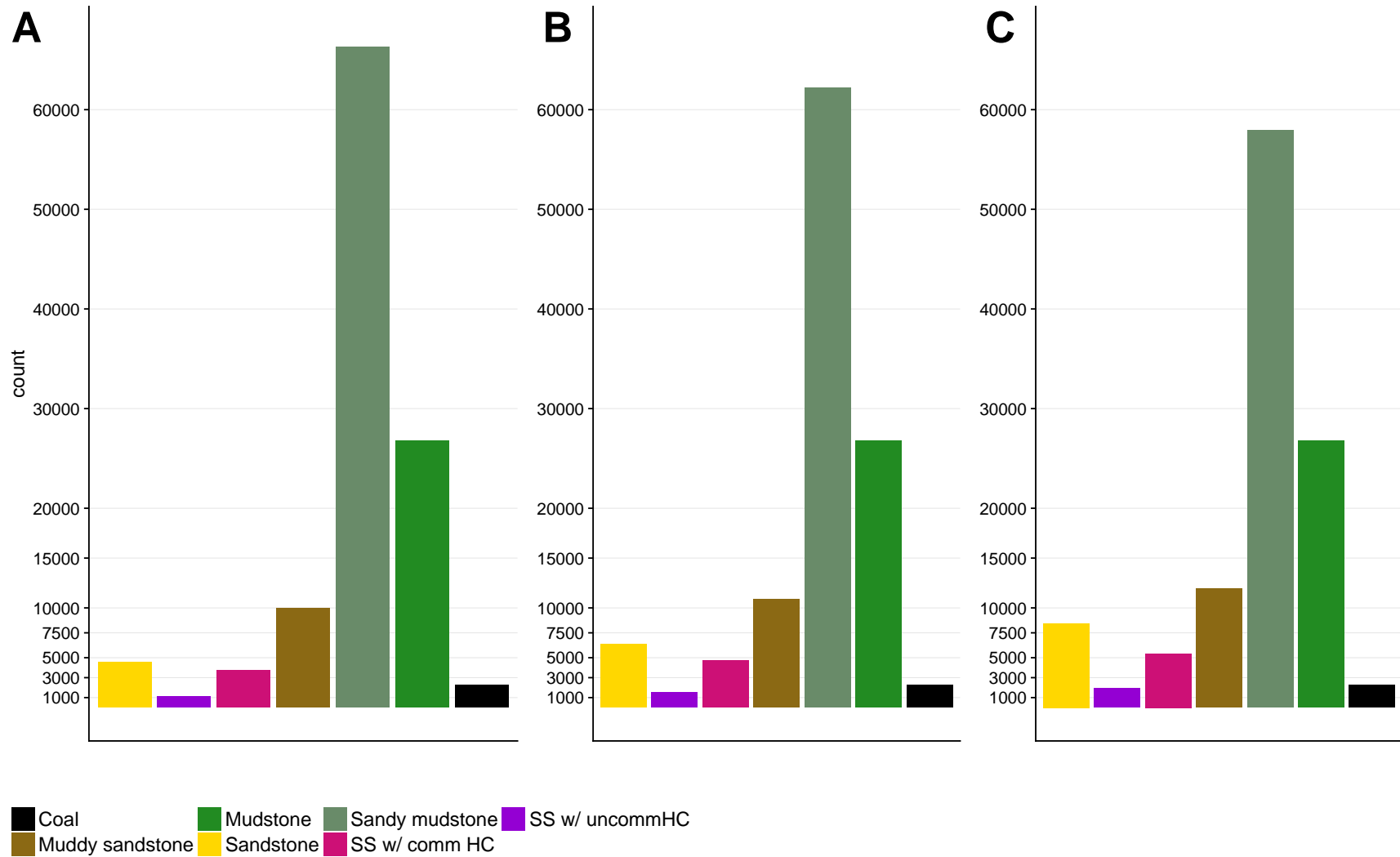
Figure 5.1: Varying the upper limit for Volume Wet Clay (VWCL) when classifying lithologies has great impact on the group distribution. Subfigures A-C shows the effect of changing the upper limit of VWCL from 15% (A) to 20% (B) and to 25% (C) in sandstones with or without HC. Table 5.2 displays how the other parameters were changed when modifying VWCL limits for sandstones. The petrophysical filters applied in subfigure C is chosen to be used for ML.

5.2 Filtering the dataset

Before applying ML on the dataset, identifying and removing or filtering out bad data is an essential step. If this is not done, the weighting can be skewed and the results will be suboptimal. From a well logging perspective, there are several reasons for why the filtering routine is necessary. Some common examples for wrong measurements are:

- If the sonic tool(s) are not vertically aligned they may give false readings because the measured travel time from the source to receiver array is wrong (Ellis and Singer 2007).

- Washout zones, which are essentially cavities in the wellbore can cause the density tool to give unrealistic low readings because the depth of investigation (DoI) is only about 15 cm (Gluyas and Swarbrick 2013). The low DoI causes the tool to instead measure the drilling mud.

- Areas close to where there is a casing change (reduction in borehole diameter) tends to cause problems for the logger and may cause several tools to give abnormal values (Ellis and Singer 2007).

- The borehole environment can damage well logging components while the logging is performed and cause the tools to give false measurements (Ellis and Singer 2007).

A filter is therefore created and applied for each log. The upper and lower limit for each log is a balancing act. If the filter is too strict the result may be that relevant information is lost and if the filter is too mild false readings may be included and reduce the accuracy of the ML model. The limits are therefore decided with a combination of visual inspection and from expected values for each log from Rider (2011). The filter is relatively mild, and its main purpose is to serve as an initial filter to remove the worst outliers which are expected to be false readings.

Table 5.3 shows the lower and upper limit of the filter for each log. Rows with data outside these intervals are removed from the dataset. The filtering is applied to the whole dataset (both training and test wells) and the filters reduce the number of rows within the dataset from 115536 to 114733. This is a difference of 803 rows, which is a 0.7% reduction of the original dataset.

Figure 5.2 shows the filtered dataset with Gamma Ray (GR) plotted on the x-axis against density (A), neutron (B), sonic compressional (C), sonic shear (D), photoelectric factor (E) and SP (F). Coal is trivial to distinguish by its low density in subfigure A, high neutron and sonic signatures in subfigures B, C and D and generally low PEF values in subfigure E except some spikes in the data above $PEF = 10$. The different types of mudstone classes can be partially separated by the GR log, however there is significant overlap in the data which is expected when plotting the whole dataset. The different sandstone classes is clustered together within the same GR interval of approximately 10-80, and there is some overlap with the mudstone classes. Interestingly the SP log in Figure 5.2F is divided into three different main plot areas. A potential cause of this may be that different companies use different tools, which has their own distinct signature.

Table 5.3: The filters applied to the dataset. Rows containing values outside specified intervals are removed.

Log	Interval
GR	0 – 250
PEF	0 – 20
AC	40 -200
ACS	70 – 400
DEN	1 – 3.5
NEU	0 – 0.8
SP	0 – inf
RDEP	0 – 2000
RMED	0 – 2000
LLS	0 – 2000

Figure 5.3 shows the different resistivity logs; deep, medium and shallow, with linear scale on the left and logarithmic scale on the right. Data from the deep resistivity logs (Figure 5.3A,B) shows that there is some separation between the sandstone classes with and without commercial amounts of hydrocarbons (HC). The medium and shallow resistivity logs do not show the same amount of separation.

Several outliers with coal classification can be seen in all resistivity logs. A possible geologic explanation for these abnormal values is that coal within deltaic sequences can give characteristically strong readings (Mondol 2015), and this may be what is observed.

Figure 5.2: Gamma Ray (GR) plotted against the standard logs. Coal is easy to identify in several of the logs due to its distinct characteristics (e.g. low density, high neutron, high AC and ACS).
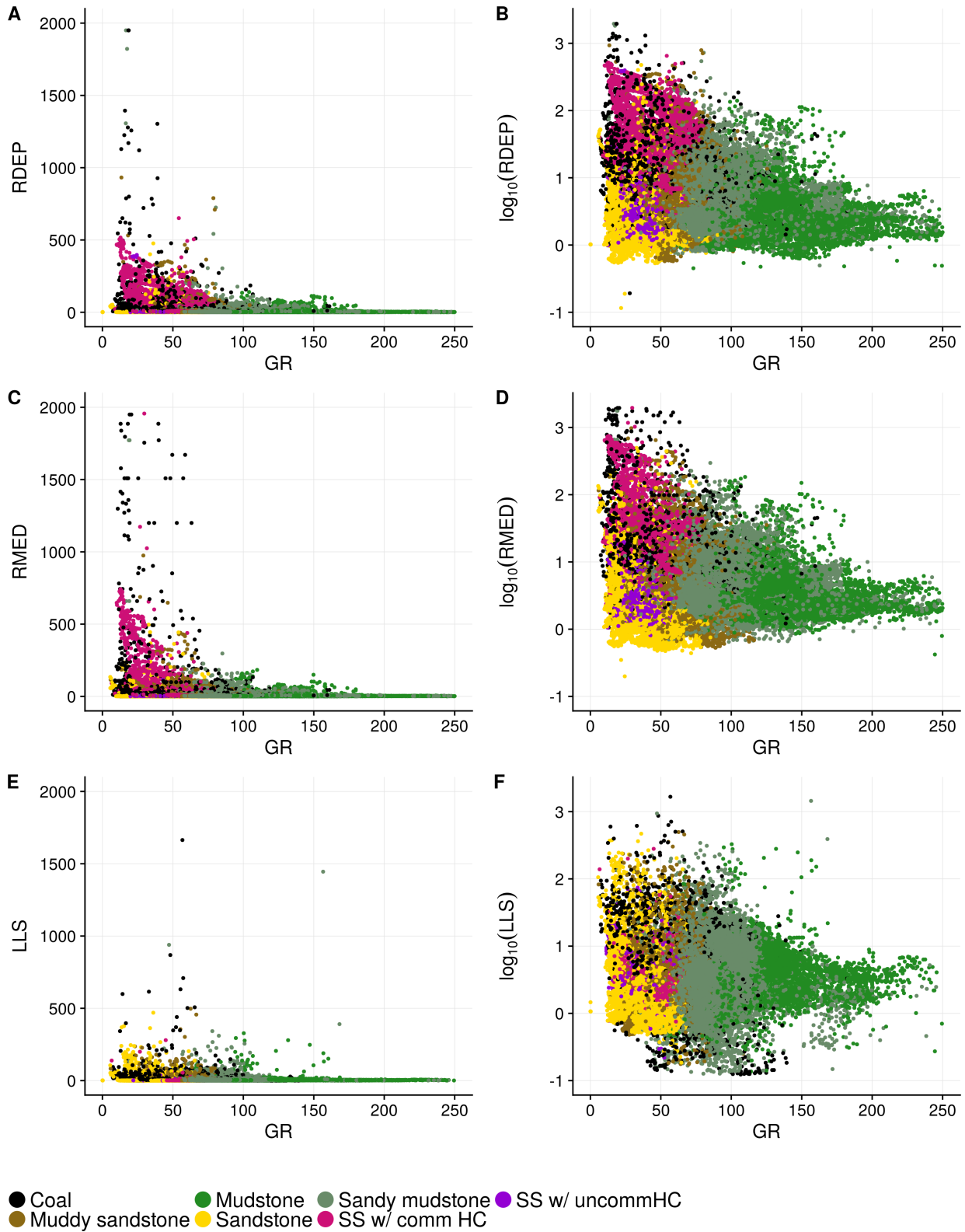
Figure 5.3: Gamma Ray (GR) plotted against the resistivity logs. Linear scale on the left and logarithmic scale on the right. In subfig. B note the increased separation width on the logarithmic scale when comparing waterfilled sandstone to sandstone filled with hydrocarbons.

5.3 Data analysis

An important step in a ML workflow is to visually and statistically inspect the data. This is done to identify outliers and features which may have a large impact on the prediction accuracy (James et al. 2013). In this study, each well and its respective logs can be seen as a separate dataset which is merged together. The ideal approach would then be to analyze each well separately and compare the wells with each other to look for potential outliers and significant differences between the wells.

The Box-and-Whisker plot (also called boxplot) was introduced by John W. Tukey (Tukey 1977) and provides a compact summary of a distribution. It is particulary useful for comparing distributions across groups (Wickham and Stryjewski 2011). Multiple boxplots are therefore an effective way to plot all logs from all the wells individually and compare them. It also gives a great overview of each well, and shows which logs are available for each well, without the need of analyzing each well separately. Figure 5.4 shows the key elements of a boxplot with the median

Figure 5.4: Key elements of a Box-and-Whisker plot. Figure courtesy of flowingdata (2008)

seen as a line within the box. 50% of the data are above the median and 50% are below. Also note that as the variation increases within a distribution, the box becomes stretched.

Figures 5.5, 5.6 and 5.7 shows boxplots for all the CPI wells, with one plot for each individual log. Gamma ray (GR), density and neutron is shown in Figure 5.5 while sonic compressional and shear as well as formation photoelectric factor (PEF) and spontaneous potential (SP) are shown in Figure 5.6. The shallow, medium and deep resistivity logs is shown in Figure 5.4.

Well 30/6-28S is from the boxplots observed to only contain one log (PEF) and it is therefore removed from the study.

Figure 5.5: Boxplot of all CPI wells showing the distribution of Gamma Ray (GR), Density (DEN) and Neutron (NEU) logs. The GR log has high variation in distribution between wells, while the density log readings are centered around 2.5 with significant outliers. The outliers in the neutron log are mostly in the upper region.
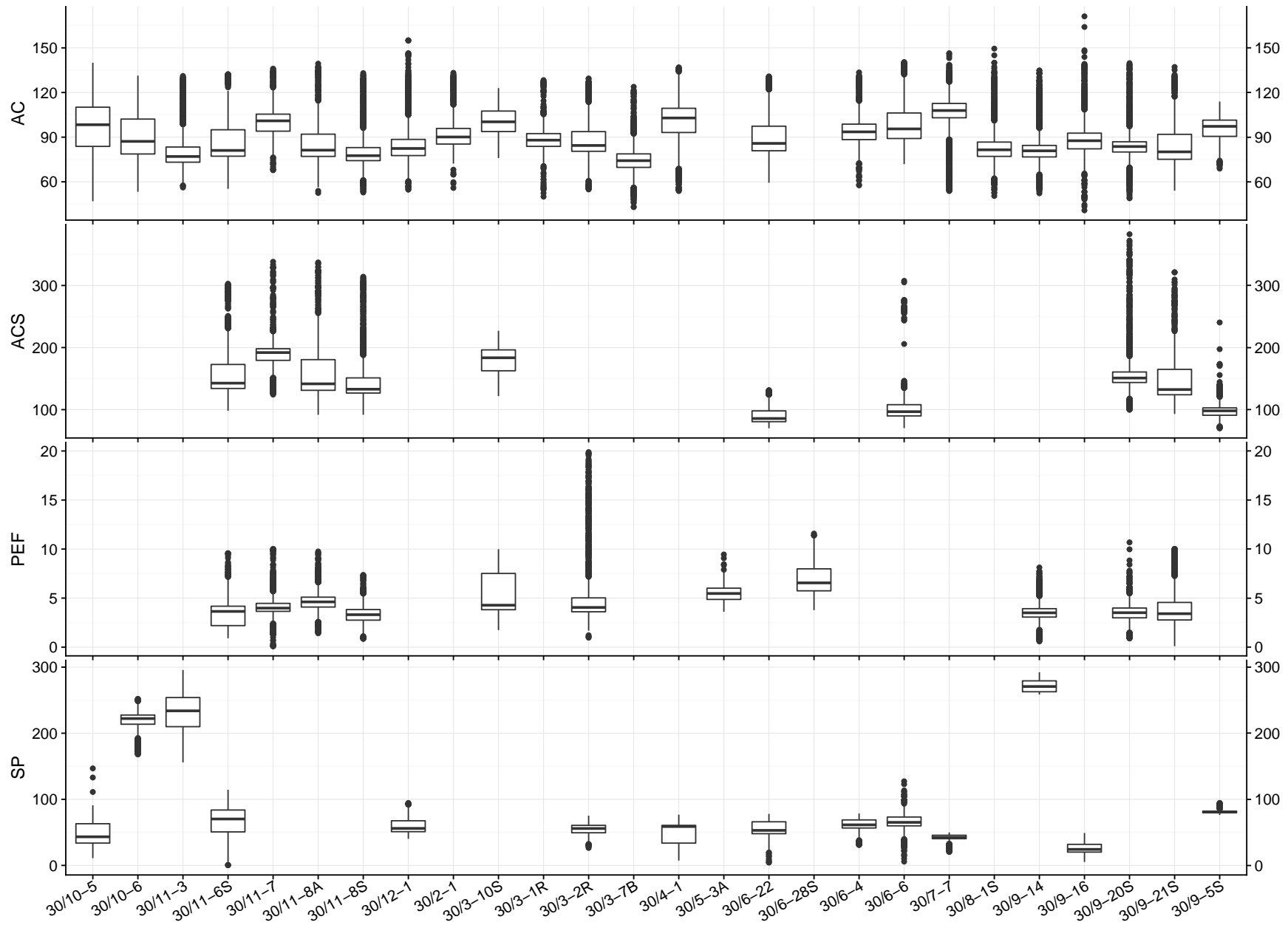
Figure 5.6: Boxplot of all CPI wells showing the distribution of sonic compressional (AC) and shear (ACS) as well as photoelectric factor (PEF) and spontaneous potential (SP). Note how the SP logs in wells 30/10-6, 30/11-3 and 30/9-14 are significantly different from the other wells, those logs are removed from the dataset.
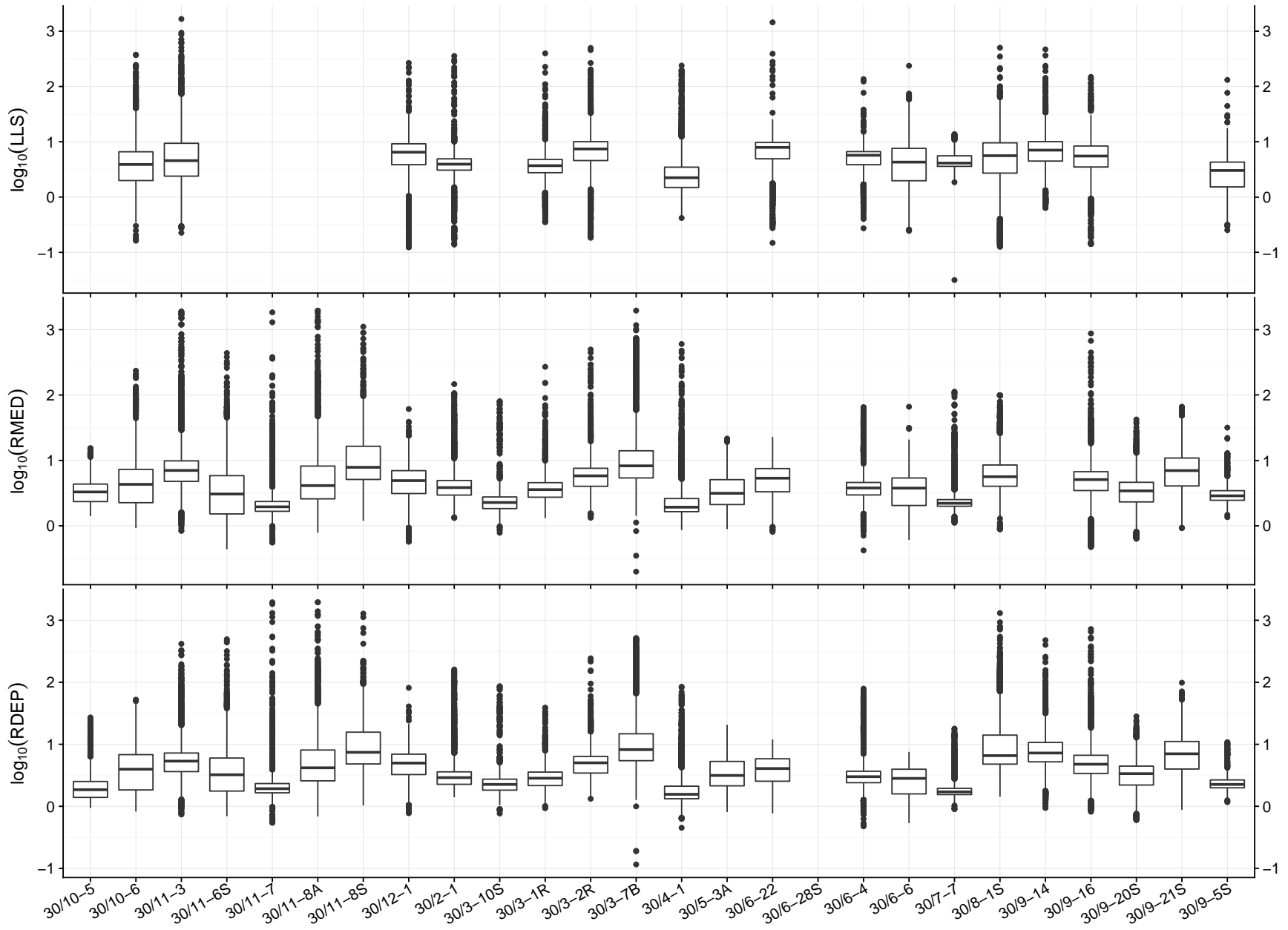
Figure 5.7: Logarithmic boxplot of the shallow, medium and deep resistivity logs from all wells with CPI made available by Exploro.

From Fig. 5.5 the GR distribution is seen to have significant variations between the wells. The most probable cause of this variation is that the clay content is different from well to well, where wells with higher amounts of clay has a higher median GR thus moving the boxes upwards. Different tools and different logging companies may also have an influence as the reference point for GR values may differ between tools. There is no source on the GR tool, but the detector which is measuring the radioactivity may be calibrated differently between wells (Rider 2011). Using barite or KCL in the drilling mud will also influence the GR log. Barium is efficent at absorbing low-energy radiation, while potassium in the KCL mud is a source of radioactivity (Ellis and Singer 2007). The timespan of when the wells was drilled is likely to also be a factor of the variability in GR. Technology advances within these years may have led to different readings and measurements troughout the years.

The density plot (Fig. 5.5) is the most stable in terms of variability, with all of the wells having a median density of around 2.5. Almost all wells have significant outliers and the low density outliers is probably caused by coal which has a low density as shown in Fig. 5.2 or the tool may have read the casing or drilling mud (Rider 2011). Hydrocarbons will also cause the density readings to go down slightly (Ellis and Singer 2007). Causes for the high density outliers may be highly cemented zones or an invasion together with using barite mud as drilling fluid, since barite has a density of 4.5 g/cm^3 (Bertozzi et al. 1981).

The neutron logs (Fig. 5.4) shows more variation from well to well than the density log and also has some outliers, mainly with high readings. As shown in the point plots from Fig. 5.2, coal is the culprit for many of the high neutron readings. Additionally, the neutron log measures the amount of hydrogen, and since hydrocarbons has long chains of hydrogen the readings go up when measuring formations containing hydrocarbons (Rider 2011). The increase differ between gas and oil, with gas giving the highest increase (Rider 2011).

From Fig. 5.6 it is evident that only two wells are missing the compressional log (AC), while 10/26 of the wells has shear log measurements. The AC distribution show variance between wells and most of the wells has outliers in both the upper and lower region. Fig. 5.2 shows that some of the higher outliers can be identified as Coal. This is logical as lower density causes slower travel time (Ellis and Singer 2007), and the AC log is the inverted travel time.

The PEF and SP log shown in Fig 5.6 are all represented in about one third of all wells. The general PEF median is around 5, with many of the wells having significant outliers. Especially well 30/3-2 R is extraordinary with values up to 20, the validity of those high values are questionable since PEF values usuallly are below 10 with the exception beeing if the well has been drilled with mud containing barite which will dominate the log response (Rider 2011).

The SP distribution is seen to be divided into three subgroups, with 30/9-14 having the highest readings, followed by 30/10-6 and 30/11-3. The rest of the wells have readings below 100, except a few readings and outliers going slightly above. The subgrouping of the SP log indicates that completely different tools has been used for the three wells with median value above 200. This will cause problems for the ML model and the SP log in wells 30/10-6, 30/11-3 and 30/9-14 are therefore removed from the database.

The resistivity logs in Fig. 5.7 show that all wells except the removed 30/6-28S well has the deep resistivity log and only well 30/9-14 is missing the medium resistivity in addition to the removed well. 15/26 wells has the shallow resistivity log available and the median value of the LLS log is the most stable in terms of variation and the deep and medium resistivity logs show higher variability between the wells.

5.4 Splitting the dataset

Before the ML model can be created, it is necessary to decide which wells are to be used as training data, and which wells are to be used as test data. It is expected that to obtain the best results good regional coverage is necessary. Well 30/6-28S was removed because the data analysis showed it consisted of one log. The remaining wells are therefore split in a way that optimalize regional coverage for both the training data and test data set.

Table 5.4 shows how the wells are split, and how many rows of data each well contains. Fourteen wells are used as training wells and eleven wells are used as test wells. All rows represent the Jurassic sequence for its respective well. Well 30/3-7B is a sidestep well and has not been vertically drilled (NPD 2018), which is the reason for it having an exceptional amount of rows compared to the other wells. For an ML model to be successfull, it is necessary to supply it with enough training data, and this is why a majority of the wells are chosen as training wells. The test wells are removed from the dataset and the ML model can not see the CPI logs or the lithoclasses when it is applied.

Table 5.4: The split between training wells and test wells, and how many rows of data each well contains. The sum shows that a higher fraction of the data is used to train the model than to apply it.

	Training wells	Rows	Test wells	Rows
1	30/10-5	5820	30/10-6	5895
2	30/11-3	8813	30/11-6S	3550
3	30/11-7	2095	30/11-8A	5248
4	30/12-1	5166	30/11-8S	3699
5	30/3-10S	2079	30/2-1	3093
6	30/3-2R	3275	30/3-1R	3267
7	30/4-1	7001	30/3-7B	11083
8	30/6-22	5135	30/5-3A	351
9	30/6-4	3379	30/6-6	4053
10	30/7-7	6598	30/9-14	4611
11	30/8-1S	7356	30/9-21S	2561
12	30/9-16	5229		
13	30/9-20S	2346		
14	30/9-5S	2328		
	Sum	66620		47411

5.5 Generating models

Three different models are developed, applied and compared against each other. Fig. 5.8 shows a simplified workflow on how the models are generated. The first step is to select the training wells from the well database, perform model specific enhancements or modifications, then optimalize and train the model. Finally the model is applied on the test wells.
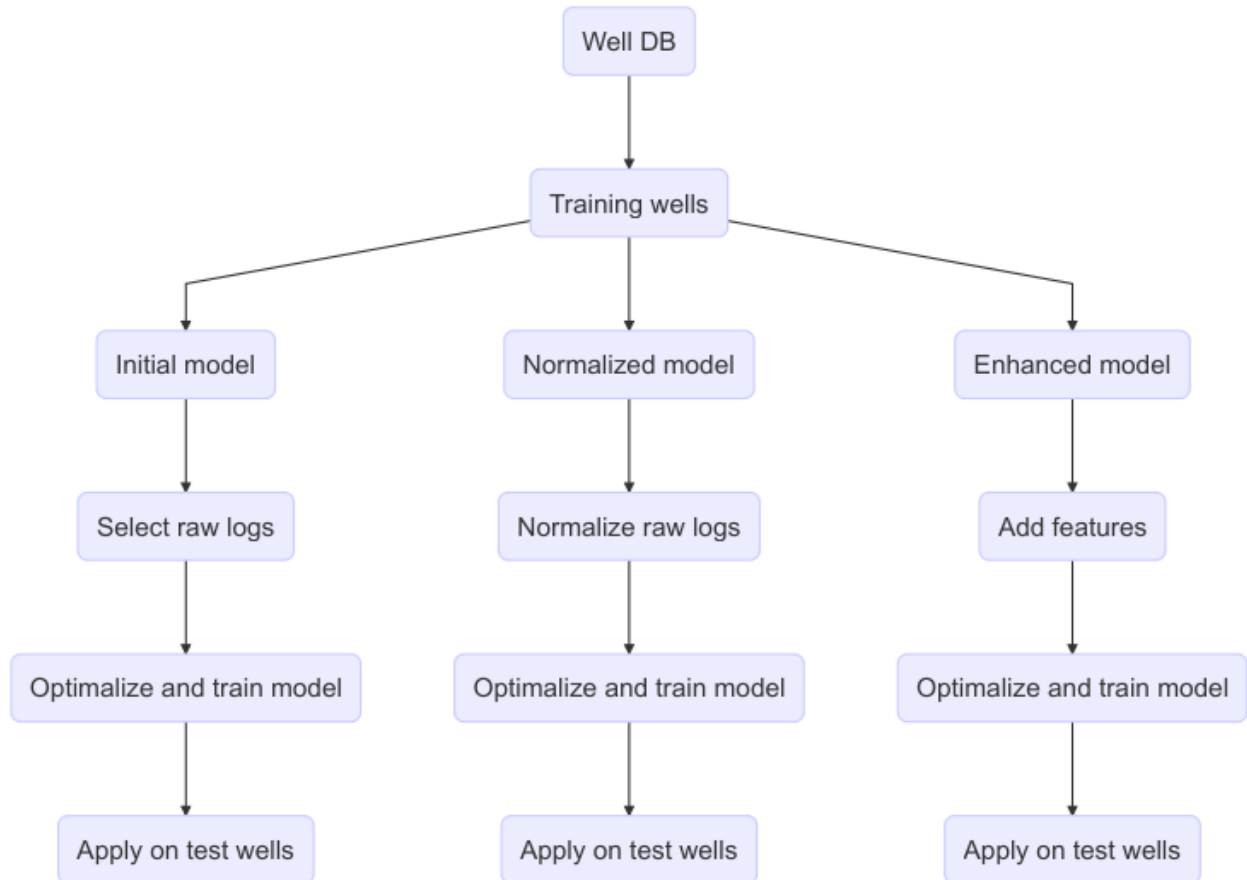
Figure 5.8: Workflow on how the different machine learning models are created and applied

5.5.1 Initial model

The first model is created to act as a reference point for the other models and to study how well the algorithm performs by limiting it to the raw logs without adding additional features or modifying the logs.

Before applying the model the hyperparameters are optimalized with a grid search, where the root-mean-square error (RMSE) (Tibshirani et al. 2013) is minimalized. Based on the documentation from the creators of xgboost (XGBoost 2016), important parameters to optimalize are max_depth, min_child_weight and the learning rate. Max depth controls how close the training data is connected to the model. Increasing max depth will increase the chances of overfitting the data causing the model to perform poorly on unseen data. Reducing max depth will make the model increasingly general however this runs the risk of underfitting, where the model is not able to see key differences between the lithoclasses. The learning rate decides how many steps the xgboost algorithm needs before reaching optimum. Increasing the learning rate will make the computations go faster (since fewer steps/rounds are needed) (Laurae 2016) but may cause the model to not be fully optimalized, if the model is complex.

The parameters are optimalized by performing a grid search where different combinations are trained and tested against a unseen part of the training dataset. To further speed up the process the grid search is parallelized with the parallel and DoParallel R packages. Figure 5.9 shows the results of the parameter optimalization. A learning rate of 0.1, max depth of 6 and minimum child weight of 6 is found to be the optimal parameters for this model and is used when the model is trained and applied to the test set.

An additional model were tested with max depth of 12. However this caused overfitting and the model performed poorly compared to the model with max depth of 6.
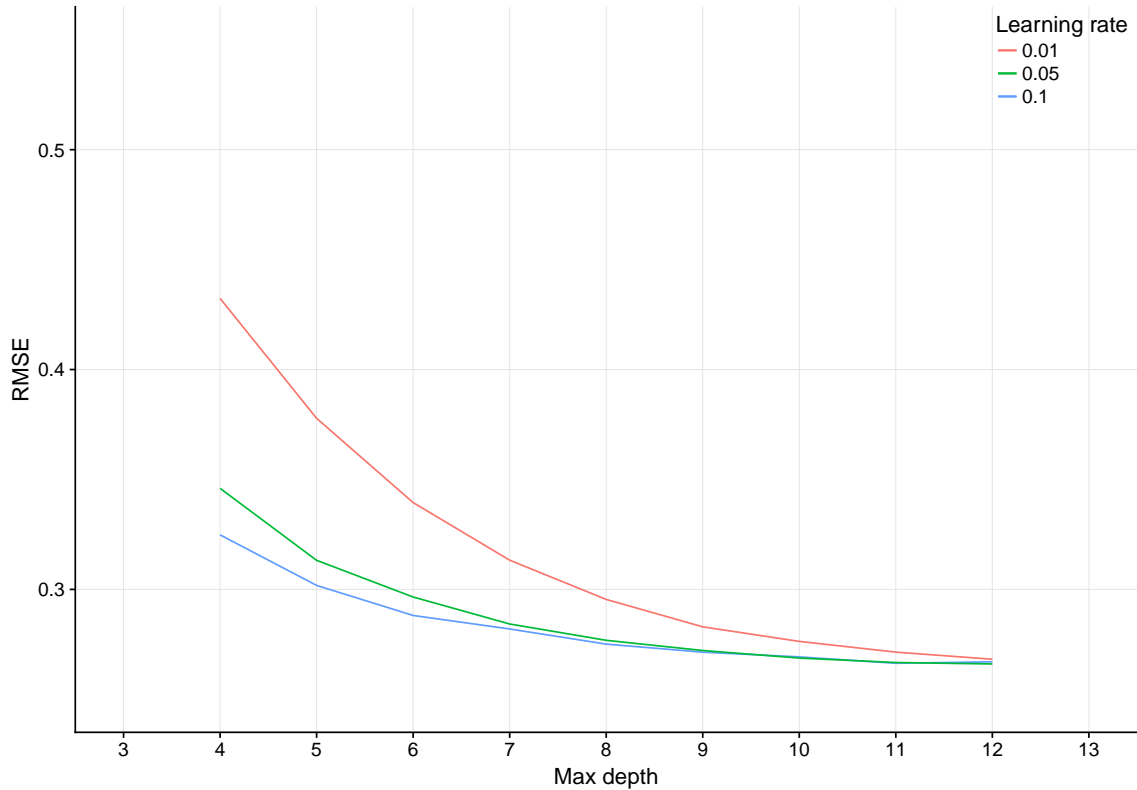
Figure 5.9: The effects of optimalizing hyperparameters on root mean square error (RMSE). Note how the RMSE flattens out after max depth of 6 for the learning rates of 0.1 and 0.05.

5.5.2 Normalized model

The normalized model is generated by transforming all the well logs into normalized logs. Within statistics this is also called standardization ([Kreyszig 2010]). The background for generating this model is to look at the implications of weighting the variables involved differently. The expected result of normalization is that outlying data is given reduced importance. The normalization equation is given by

$$log_{nm} = \frac{log - \mu_{log}}{\sigma_{log}}, \tag{5.2}$$

where log_{nm} is the normalized log, log is the original log, μ_{log} is the mean and σ_{log} is the standard deviation. μ_{log} and σ_{log} are well specific and is calculated individually for each well within the Jurassic sequence.

Figure 5.10 shows the hyperparameter optimalization where each line is with a different learning rate, either 0.1, 0.05 or 0.01. The best parameters from the optimalization is a max

depth of 12, learning rate = 0.1 and minimum child weight = 7.
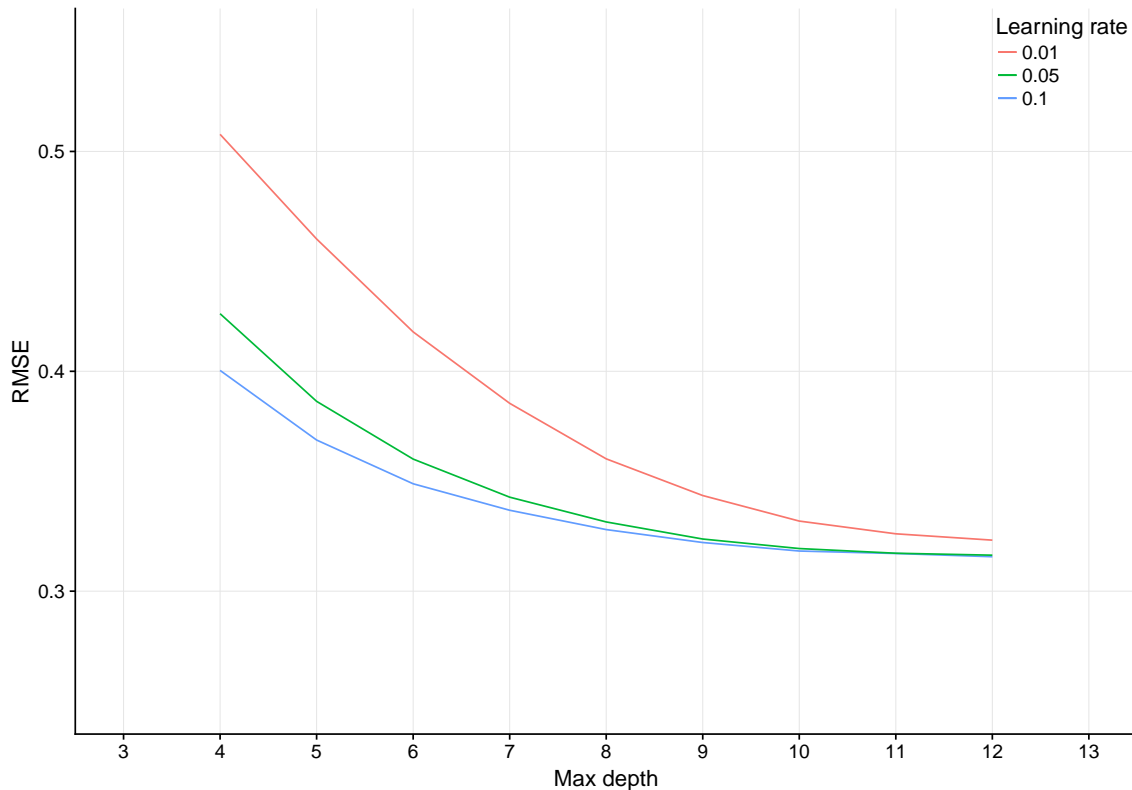
Figure 5.10: Hyperparameter optimalization of the normalized model.

5.5.3 Enhanced model

To further improve the accuracy of the model several additonal features are added to address multiple challenges related to the welldata:

- The study area is large and the training wells are spread apart up to several tens of kilometers.

- The geology on such a scale is expected to have natural variations such as the mineral composition. This will affect the log signatures.

- The well logging has been performed over a large timespan and by different contractors. The tools used and the measurements done is highly likely to have changed with time, e.g. the detector used in GR measurements may differ.

The result of these variations between wells is that the log signatures may be different and can create problems for the ML model when it is attempting to classify intervals. The introduction

of specific additional features are designed to mitigate this problem by generalizing the log signatures from each well.

First, the resistivity logs are transformed from a linear scale to a logarithmic scale with base 10. The differences between the linear and logarithmic scale can be seen in Figure 5.3. The large variations in the resistivity logs is expected to cause reduced prediction accuracy. Transforming the resistivity logs to a logarithmic scale will make the distribution more linear and may have a positive effect.

Next, the P99 quantile is calculated for each log in each well, and the log is then divided by the P99 value. The choice of using P99 instead of maximum value is that the P99 value gives a more accurate representation of a high value within the log and is not affected by outliers in the data in the same way as the maximum value. Even after the filtering is applied the likelihood for outliers is still present and the P99 value has a lower likelihood of beeing influenced by the outliers. It represents a more accurate depiction of the highest value within the log.

Figure 5.11 shows an extreme case of the difference between the P99 value and maximum value for the GR log of well 30/6-4. The plot is flipped from the traditional way of plotting a well log with depth on the y-axis to emphasize the GR values. The model is optimized in the same way as the initial and normalized model which is shown in Fig. 5.12. The RMSE value in the enhanced model is lower than both the intial and normalized models shown in Figures 5.9 and 5.10. The chosen parameters based on the optimization routine is a max depth of 12, learning rate of 0.05 and minimum child weight of 8.
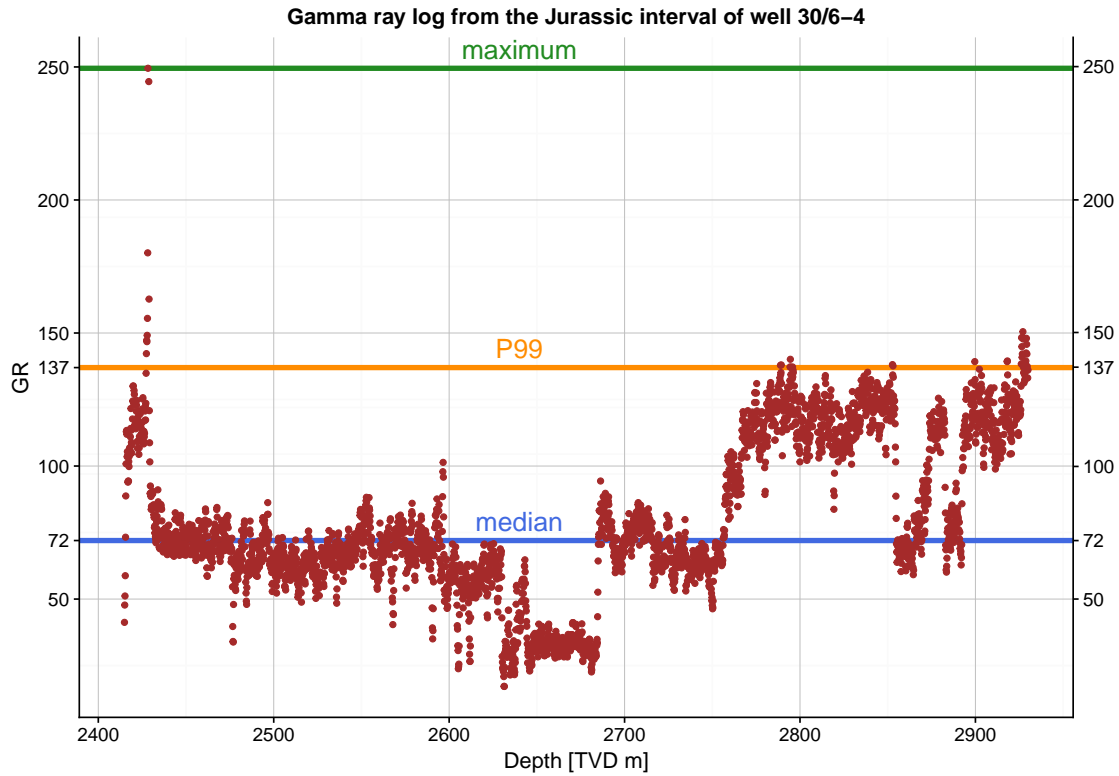
Figure 5.11: Gamma ray log from well 30/6-4 showing the median, P99 and maximum value of the Jurassic interval. The P99 value is used when generating additional relative features.
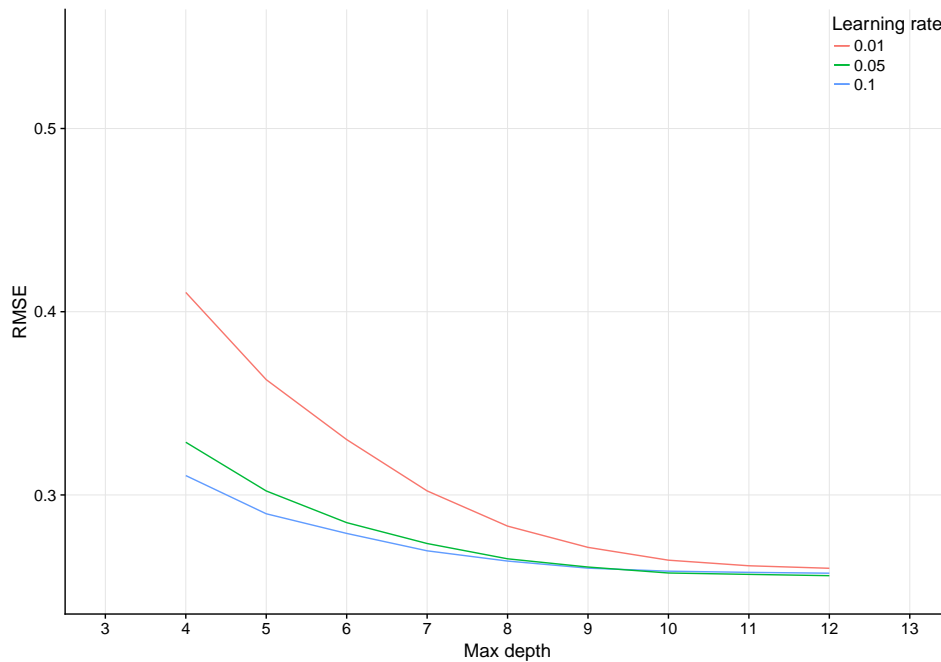
Figure 5.12: Hyperparameter optimalization of the enhanced model. The RMSE value is lower than the RMSE values in the intial and normalized models shown in Figures 5.9 and 5.10.

Chapter 6

Results

6.1 Initial model

Figure 6.1 shows a confusion matrix (A) and a feature importance plot (B) for the initial model. Within the matrix the true labels for each classification is along the vertical axis, and the predictions are along the horizontal axis. The correct predictions are going along the diagonal from the upper left corner to the lower right corner. The accuracy is calculated by taking the number of correct predictions for a class and dividing it by the total sum of the lithoclass. E.g. the SS accuracy is calculated in the following way

$$Accuracy_{SS} = \frac{3208}{3208 + 194 + 0 + 396 + 148 + 0 + 2} * 100\% = 81.3\%.$$

From Fig. 6.1A waterfilled sandstone (SS) has a high degree of accuracy with 81.3% while the sandstone with uncommercial hydrocarbons (SSUHC) is challenging to predict with the lowest degree of accuracy (22.5%). The sandstone with commercial amounts of hydrocarbons (SSCHC) has a higher accuracy with 59.5%, while the predictions of muddy sandstone has an accuracy of 40.5%. When the mud content increases further the accuracy goes up significantly with 70.7% and 71.4% for sandy mudstone and mudstone respectively. Coal has the highest degree of accuracy with 96.6%.

In the feature importance plot (Fig 6.1B) GR has the highest influence on class predictions with almost 0.3, followed by the neutron and density log which are both around 0.2. There is a significant jump in terms of importance from GR, density and neutron down to the rest of the logs.

A

Reference

	SS	SSUHC	SSCHC	MUDSS	SMUDST	MUDST	COAL	Accuracy
SS	3208	1004	369	1718	124	0	3	81.3
SSUHC	194	323	964	257	0	0	0	22.5
SSCHC	0	31	2224	333	34	0	6	59.5
MUDSS	396	65	103	2799	3057	0	3	40.5
SMUDST	148	14	68	1747	16915	1826	24	70.7
MUDST	0	0	0	40	3754	4556	1	71.4
COAL	2	0	11	11	29	0	1050	96.6

Prediction

B
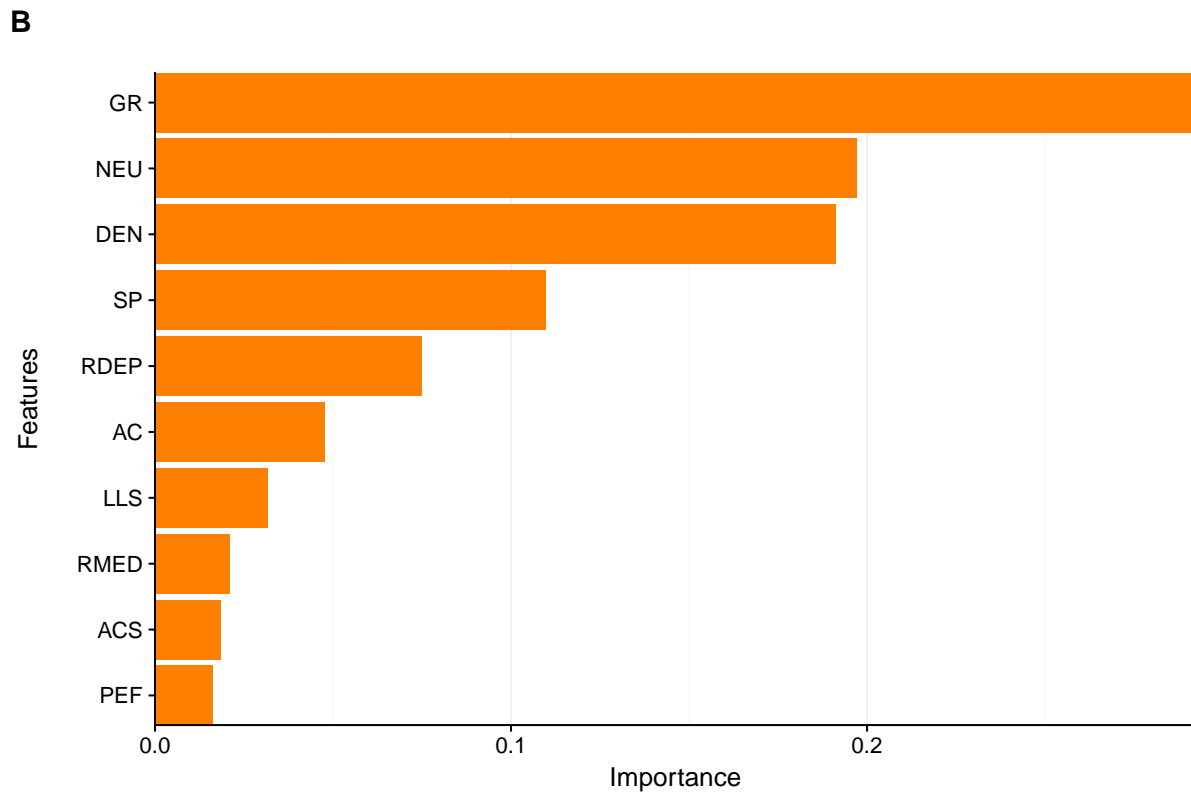
Figure 6.1: Confusion matrix (A) and feature importance plot (B) from baseline model where only the raw logs were used. Correct predictions are on the diagonal from upper left down to lower right, while the other values on each row are false positives.

6.2 Normalized model

Figure 6.2 shows the confusion matrix (A) and feature importance (B) for the normalized model. Comparing the confusion matrix from the normalized model with the initial model shown in Fig. 6.1 it is evident that the accuracy is reduced substantially for all lithoclasses except sandy mudstone which has gone from 70.7% to 83.3%. By normalizing the log responses outliers and unique tendencies becomes less weighted, and this is reflected in the results. Coal, which can easily be identified in the logplots from Fig. 5.2 is a good example of this behavior. The accuracy has been reduced from 96.6% in the initial model to 74.6% in the normalized model.

The feature importance plot shown in Fig. 6.2B has changed slightly when comparing it to the initial model. Gamma ray is still the most important feature and the density log has become more important than the neutron log. The shallow resistivity log (LLS) is now more important than the deep resistivity log (RDEP).

6.3 Enhanced model

Figure 6.3 shows the confusion matrix and feature importance plot of the enhanced model. Compared to the initial model there is an accuracy increase in sandstone, muddy sandstone, sandy mudstone and coal predictions. A subtle difference is that the false predictions placements are different than from the initial model. For example, the SSUHC false predictions has moved from MUDSS and SMUDST to primarily SS. The same pattern can be seen in the SSCHC lithoclass.

Compared to the normalized model (Fig. 6.2) there is a significant accuracy increase across all lithoclasses except SMUDST which has a slight decrease.

The confusion matrix in Fig. 6.3B shows the impact from the implementation of relative features. Especially the relative features of GR, density and neutron has made a large impact.

A

		Reference						
		SS	SSUHC	SSCHC	MUDSS	SMUDST	MUDST	COAL	Accuracy
Prediction	SS	2537	851	667	936	56	0	5	64.3
	SSUHC	118	80	188	170	56	0	1	5.6
	SSCHC	16	35	1376	149	37	0	4	36.8
	MUDSS	726	355	1241	2521	2243	1	21	36.5
	SMUDST	548	115	212	3110	19918	4242	233	83.3
	MUDST	0	0	0	4	1518	2129	12	33.4
	COAL	3	1	55	15	85	10	811	74.6

B
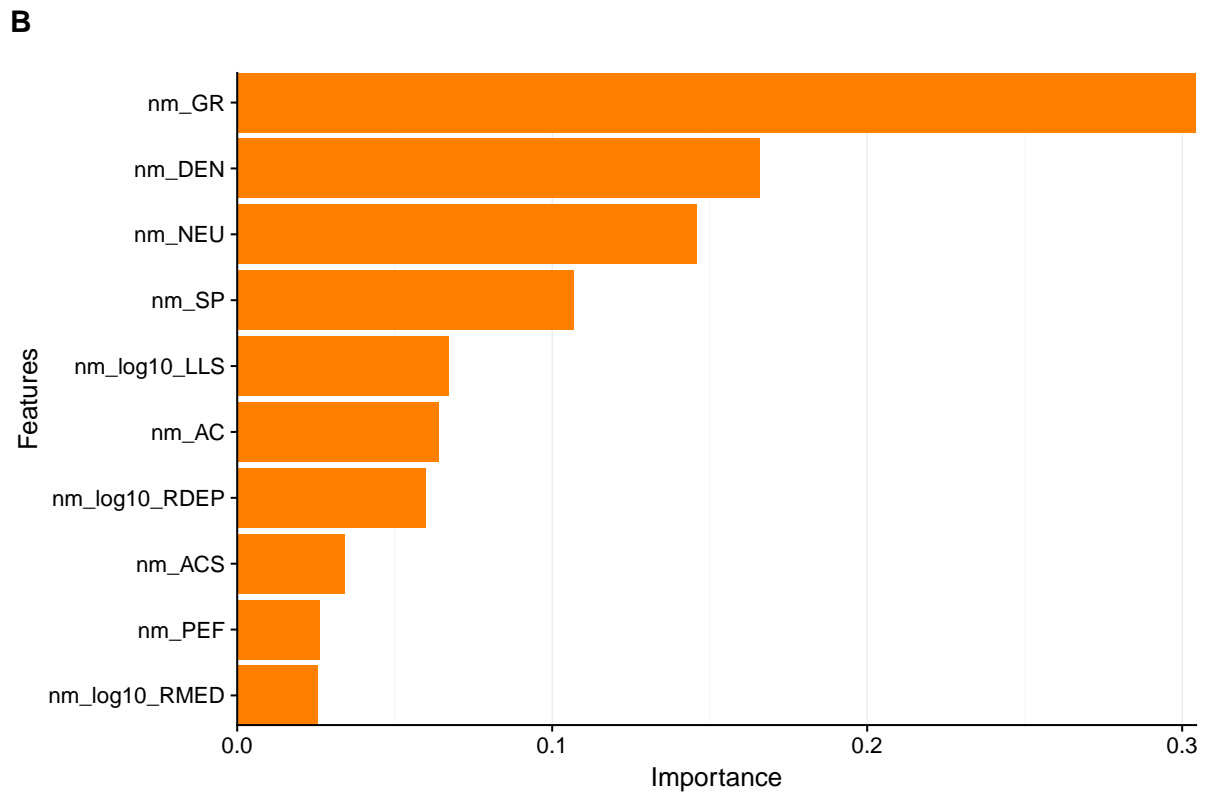
Figure 6.2: Confusion matrix (A) and feature importance plot (B) of the normalized model. Compared to the initial model, accuracy in predictions has been significantly reduced for all classes except sandy mudstone.

A

Prediction	Reference							Accuracy
	SS	SSUHC	SSCHC	MUDSS	SMUDST	MUDST	COAL	
SS	3445	1191	1011	1198	14	0	1	87.3
SSUHC	35	150	322	123	0	0	0	10.4
SSCHC	0	48	2181	297	29	0	4	58.3
MUDSS	342	44	129	4092	2518	1	3	59.3
SMUDST	124	4	88	1161	19041	2978	13	79.6
MUDST	0	0	0	22	2266	3403	0	53.3
COAL	2	0	8	12	45	0	1066	98.1

B
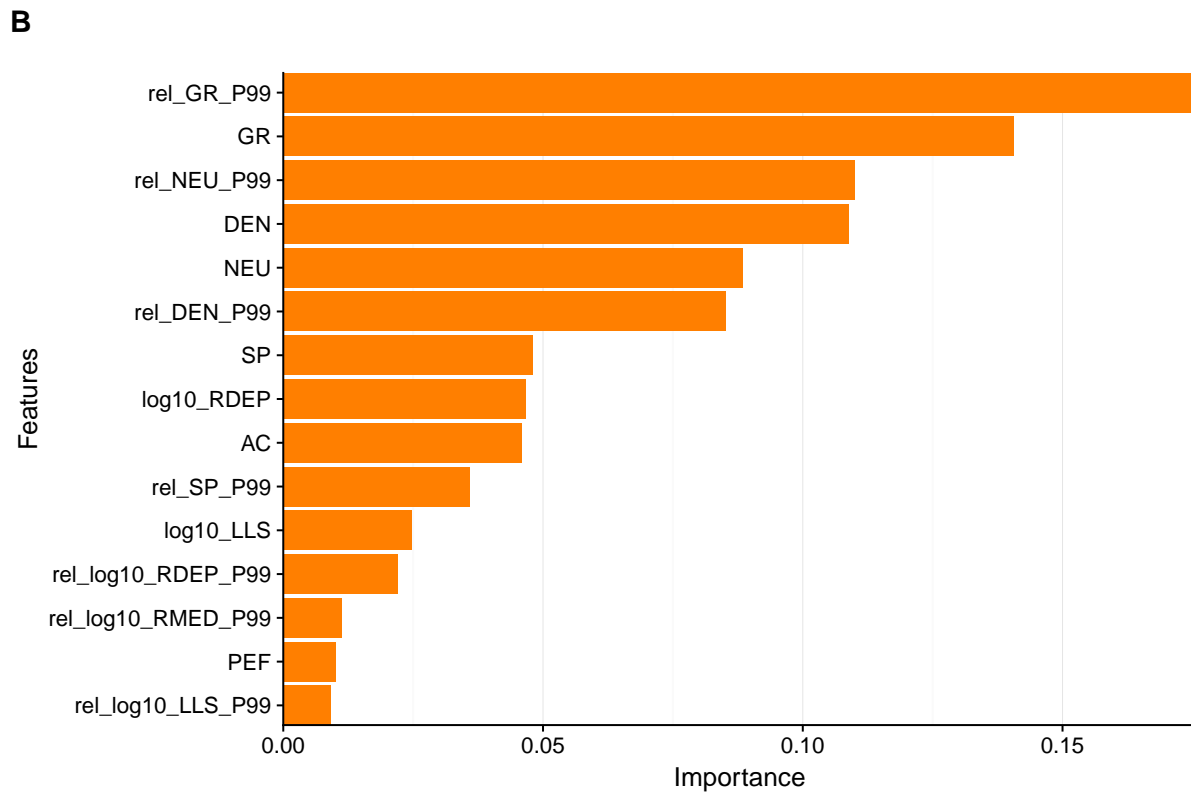
Figure 6.3: Confusion matrix (A) and feature importance plot (B) of the enhanced model. The feature importance plot shows the 15 most important features.

6.4   Model accuracies and lithoclass distribution

Table 6.1 shows the well prediction accuracy for the three different models and how many logs are available in each of the test wells. Figure 6.4 shows a bar graph of the accuracy distribution for the wells. For the initial and enhanced model, the two wells that scores poorest in terms of accuracy are 30/2-1 and 30/3-1R. The two wells has five logs available in the Jurassic interval and from the boxplots in Figures 5.5, 5.6 and 5.7 the logs available are GR, AC and the three resistivity logs RDEP, RMED and LLS. In the normalized model, well 30/3-1R has doubled its accuracy compared to the initial and enhanced models and well 30/2-1 also has increased accuracy compared to the other models. Well 30/9-14 and 30/11-6S are the two wells with highest accuracy, where the enhanced model scores highest. The two wells also has a high amount of logs available, with 7 and 9 for 30/9-14 and 30/11-6S respectively. The general trend is that a higher amount of logs gives a higher degree of accuracy for the initial and enhanced model.

Figure 6.5 shows the total number of lithoclass predictions for each model and the reference model. All models are predicting a higher amount of sandstone compared to the correct number in the reference model. The models also underpredicts the number sandstone with commercial/uncommercial amounts of hydrocarbons.

Table 6.1: Well prediction accuracies for the three different ML models and how many logs were available in each well. Note how the number of logs available has an influence on the accuracy.

Wellname	Initial	Normalized	Enhanced	#Logs
30/10-6	57.4	74.4	72.3	7
30/11-6S	74.5	59.4	83.8	9
30/11-8A	69.6	45.1	69.6	8
30/11-8S	77.1	59.7	79.1	8
30/2-1	49.4	69	58.7	5
30/3-1R	34.5	70.6	35.7	5
30/3-7B	62.1	60.9	68.2	6
30/5-3A	73.8	70.4	75.8	6
30/6-6	81.2	67.3	76.4	9
30/9-14	74.5	62.7	88.1	7
30/9-21S	78.9	48.5	62.4	8
Avg acc.	65.5	62	70.4	
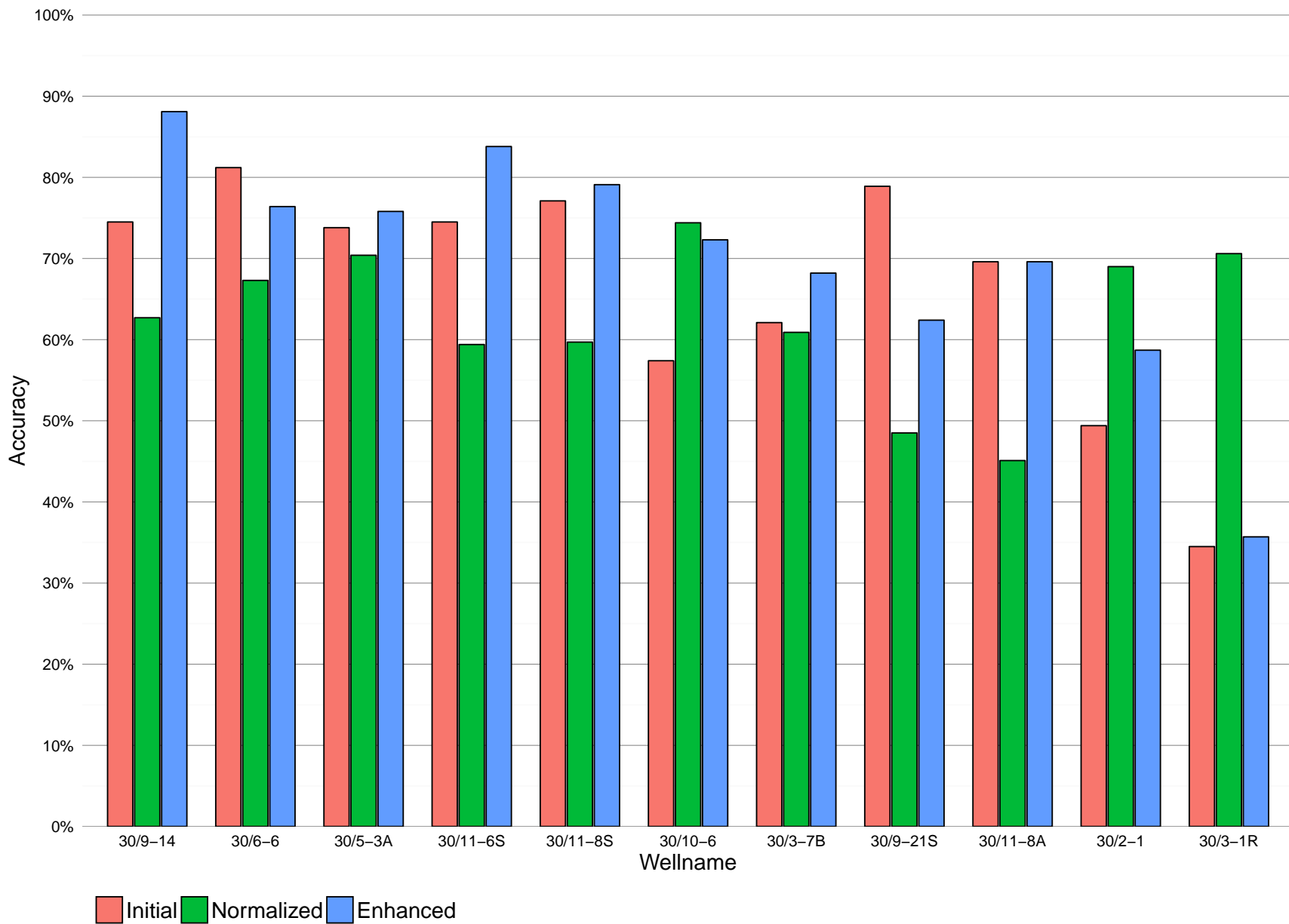
Figure 6.4: Bar plot of model accuracy for each well. The normalized model has the highest accuracy of the models on wells 30/2-1 and 30/3-1R which also has the lowest amount of logs available.
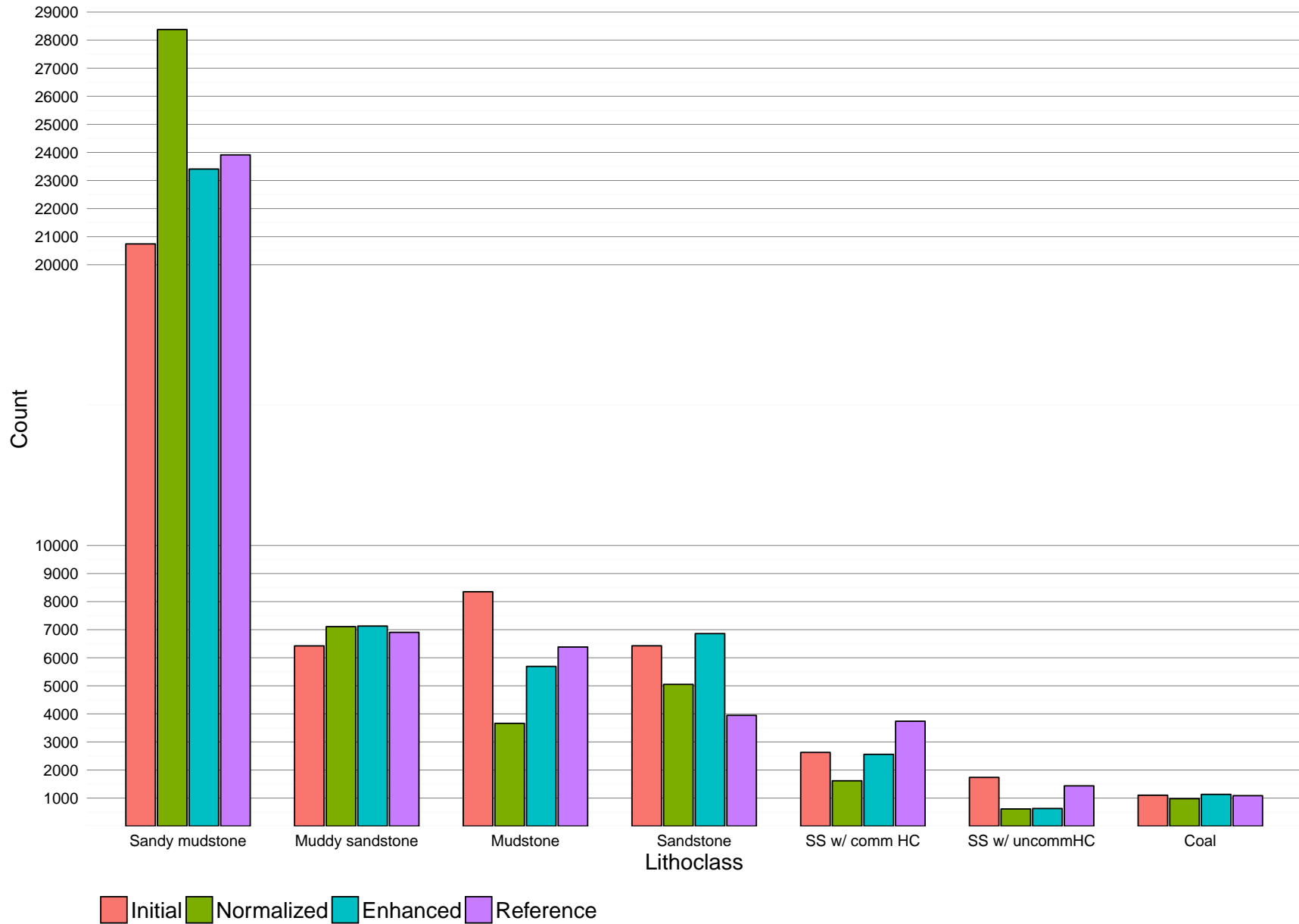
Figure 6.5: Number of total lithoclass predictions for each model. All ML models predict a higher amount of sandstone values than the reference model, and a lower amount of sandstone with commercial or uncommercial amounts of HC.

6.5 Well 30/9-14

Figure 6.6 displays gamma ray (GR), water saturation (SW), volume wet clay (VWCL) and the results from the three different models together with the correct classification in a section between 3100 and 3150 meter in the well. The SW and VWCL are from the CPI and is used for lithology classifications shown in the Reference column.

The initial model is seen to be interpreting mostly sandy mudstone instead of muddy sandstone several places in the interval. In addition it does not accurately recognize the hydrocarbon zones, instead classifying them as sandstone intervals. This is especially noticeable at approximately 3125m, where it predicts sandstone or sandstone with uncommercial amounts of HC where the correct classification is sandstone filled with commercial amounts of HC.

The predictive behavior of the normalized model is significantly different than the initial model. A general tendency to predict lithologies with lower clay content is evident, e.g. where the model predicts muddy sandstone instead of sandy mudstone or sandstone instead of muddy sandstone.

The enhanced model shows the highest degree of accuracy when compared to the reference model, and accurately predicts the main features of the interval (HC and sand zones). At approximately $3101 - 3103$m the enhanced model predicts an uncommercial HC zone while the reference model classifies it as a muddy sandstone. From the VWCL and SW values the reference classification is at the border between classifications (Table 5.2), and the differences between the lithologic classes are so small that it does not manage to distinguish between them. At approximately 3119 meter and between $3124 - 3128$ meters the enhanced model accurately predicts the HC zones, with few significant errors.

Figure 6.7 shows the interval between $3350 - 3400$ meters which has high presence of coal. Between 3354 and 3369 meters the initial model is seen to falsely predict mudstone instead of sandy mudstone, while the other models have a greater accuracy in their predictions in the same interval. The coal is predicted accurately in all models except at approximately 3384 meters where the normalized model fails to identify the coal layer. In general all three models accurately identifies the sandstone layers, with some minor errors such as interpreting muddy sandstone instead of sandstone in some parts.
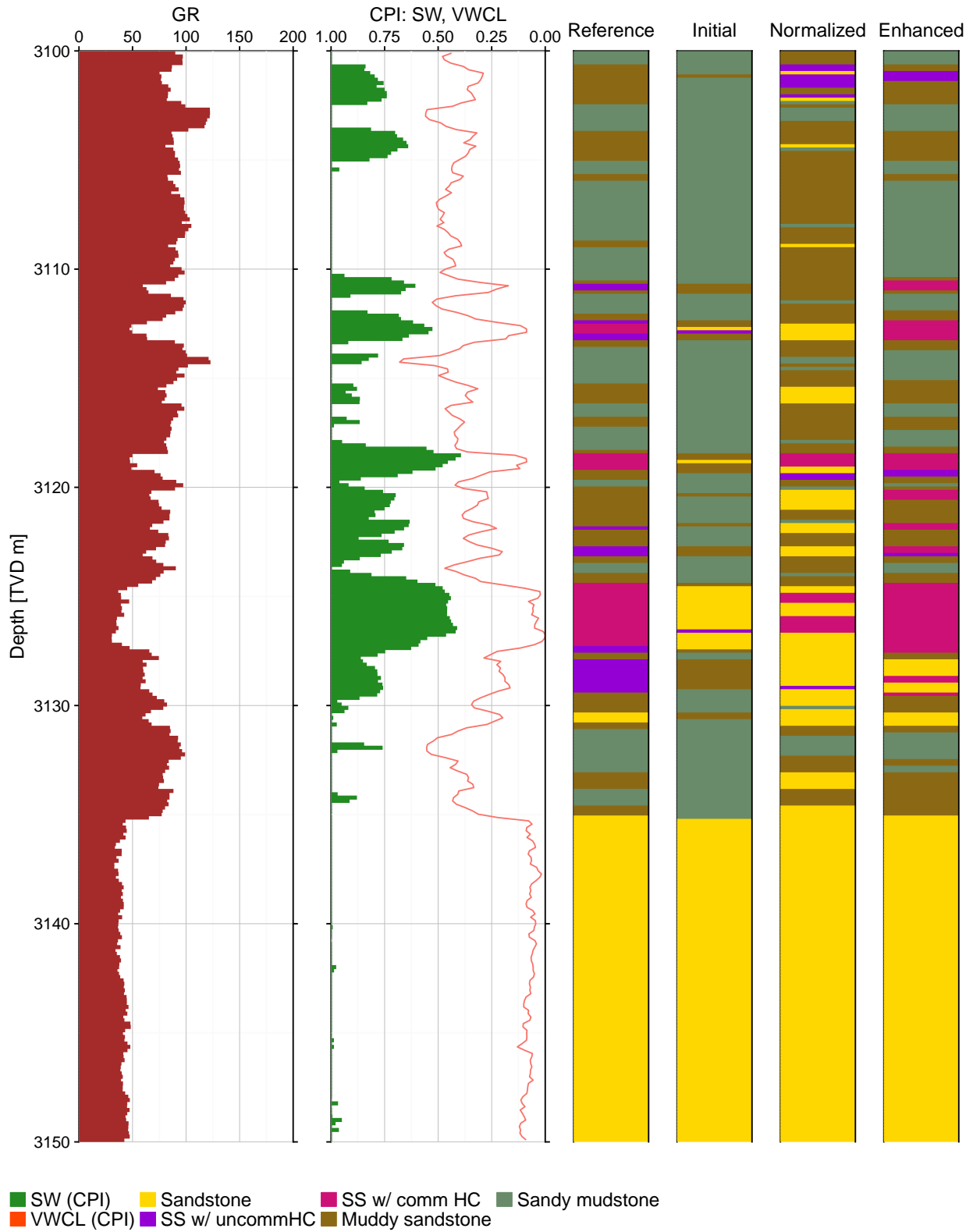
Figure 6.6: Log interval between 3100 to 3150 meters (Tarbert Fm.) from well 30/9-14 showing GR, SW, VWCL and a comparison of the different ML models. When water saturation drops below 100% it is shown in green and is a indication of HC. Coal is not present in this log interval.
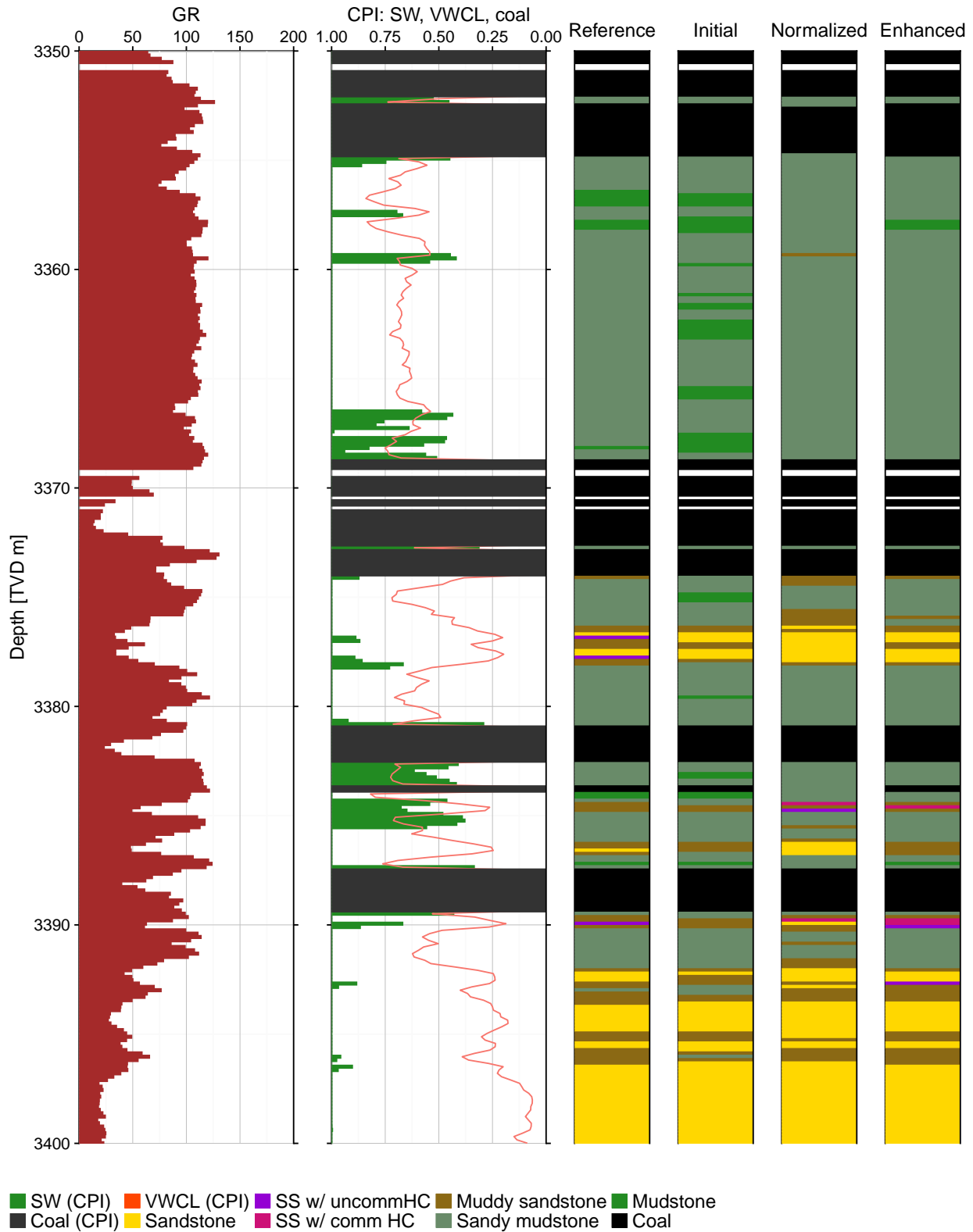
Figure 6.7: Log interval between 3350 to 3400 meters (Ness Fm.) from well 30/9-14. The white stripes in the log are intervals with missing data or because of filtering in the preprocessing. High amount of coal is present and all ML models accurately predict the presence of coal.

6.6   Well 30/11-8 A

Figure 6.8 shows the interval $3590 - 3640$ meters (Heather Fm.) in well 30/11-8 A. The upper part of the figure shows a mudstone acting as a barrier or trap for the hydrocarbons within the rock below. The VWCL values within the hydrocarbon interval is right at the border between muddy sandstone and a sandstone filled with hydrocarbons (VWCL >= 25% for muddy sandstone). This is reflected in the reference model where the classification changes between sandstone with commercial amounts of HC and a muddy sandstone. Major differences can also be seen in the classifications between the reference well and the three models within this interval. All three models classifies sandstone with commercial amounts of HC instead of muddy sandstone troughout the interval, and the normalized log is picking up many of the smaller intervals where the water saturation is reduced by only $10 - 15\%$.

Figure 6.9 shows the interval $4100 - 4150$ meters (Ness Fm.) in the well. Based on the reference log the lithology in the section consists mainly of sandy mudstone and mudstone along with some sandstone intervals. There is a small HC section at 4105 m, and a larger section at approximately $4116 - 4120$ meters. Coal is present troughout the section. The initial model accurately classifies most of the interval, except mixing the sandy mudstone and mudstone some places. There is also a falsely predicted HC interval at the bottom of the section. The normalized log generally predicts sandy mudstone instead of mudstone troughout the interval, and also has some errors in the main HC interval ($4116 - 4120$m). Additionally, the normalized log has some errors in its coal classifications, which can be seen at approximately 4115m and 4131m.

The classifications of the enhanced model are accurate troughout the section, with some errors in the HC interval where it classifies a commercial sandstone interval as a waterfilled sandstone interval.
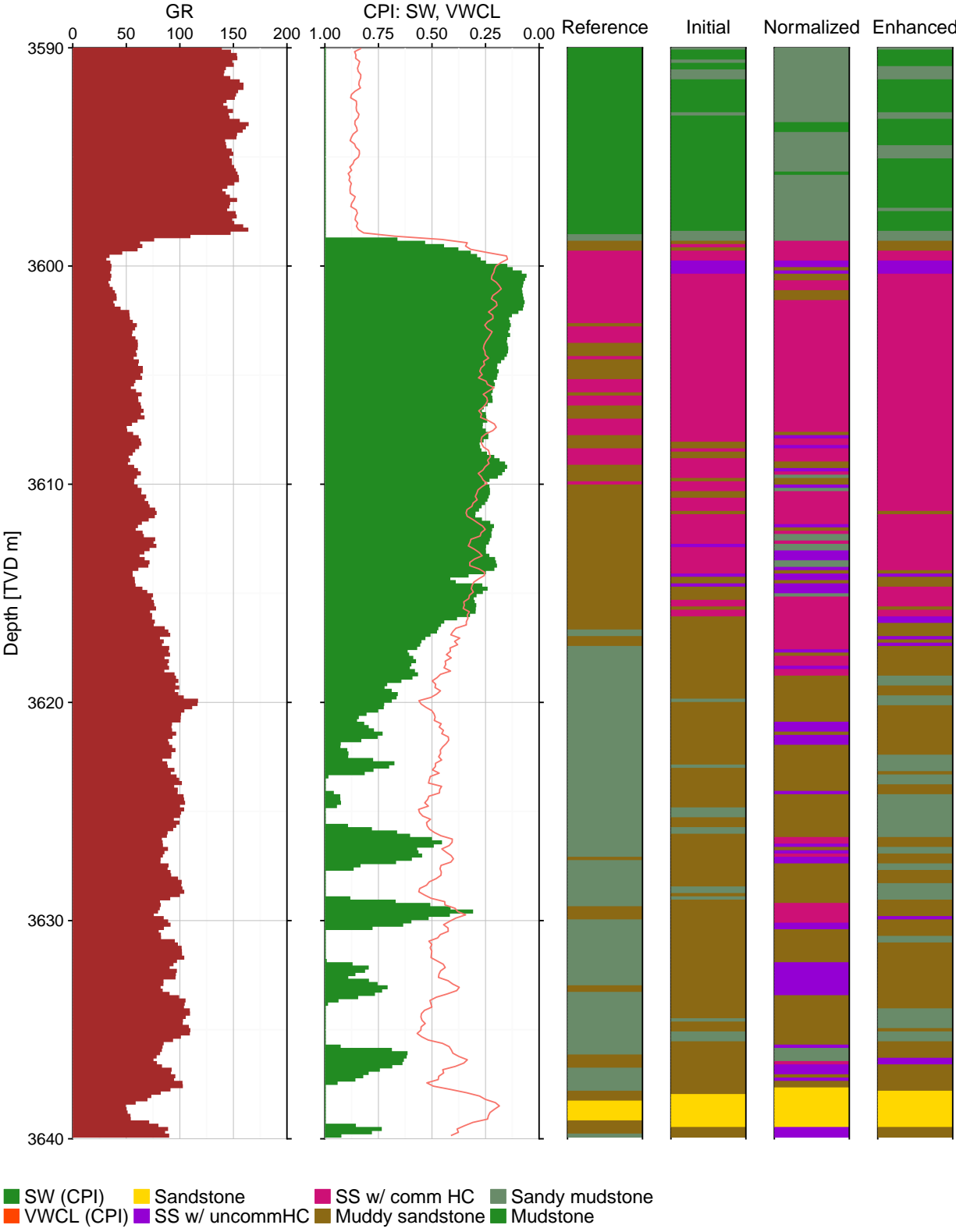
Figure 6.8: Log interval between 3590 − 3640 meters (Heather Fm.) from well 30/11-8 A . The upper part of the log shows a mudstone acting as a barrier or trap for the hydrocarbons within the interval 3598 − 3620m.
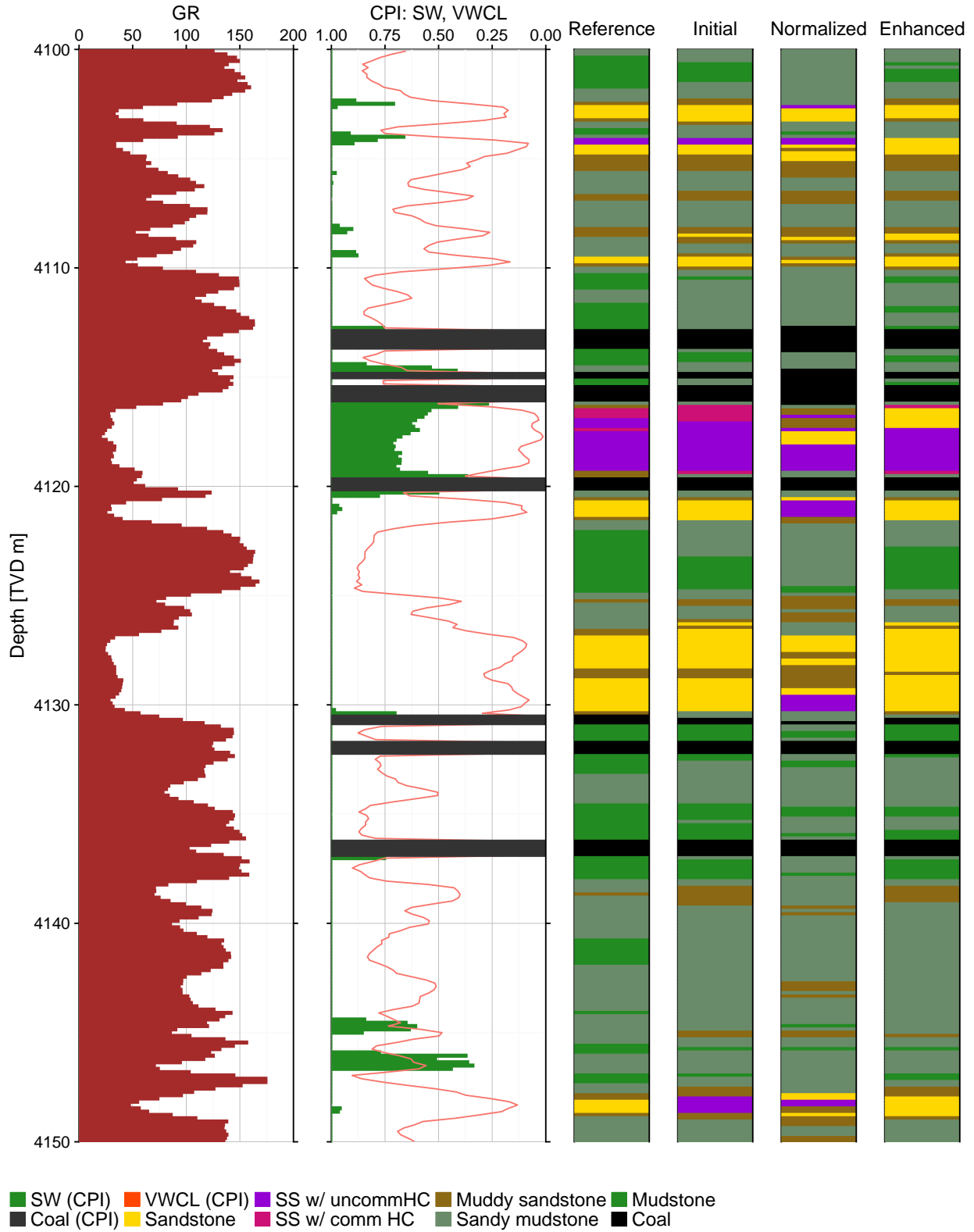
Figure 6.9: Log section (4100 − 4150m) from well 30/11-8 A in the Ness Fm. One major HC interval can be seen between 4116 − 4120m. The normalized model is the least accurate, while the initial and enhanced model fares better.

Chapter 7

Discussion

Lithoclasses and the model's ability to correctly classify them is first to be discussed followed by the prediction accuracy in individual wells with emphasis on the wells with the highest and lowest accuracies. Finally, a summary comparison between the models is discussed.

7.1   Lithoclasses and prediction ability

Based on the confusion matrixes shown in Figures 6.1A, 6.2A and 6.3A, a predominant observation is that the sandstone class is easy to predict with high accuracies in all models. The high accuracy can be related to the filters used when creating the lithoclasses shown in Table 5.2, with the sandstone class having between $0-25\%$ VWCL. This provide a good coverage of training data and makes it easier for the algorithm to identify it than if the upper limit was set 15 or 20%. From the point plots in Figures 5.2 and 5.3 the sandstone class is clearly distinguishable within the data, especially in the GR log which is the most important feature. This is expected to be the main reason for the accurate predictions.

When looking at the SSUHC and SSCHC classes, they are are less accurate than the SS class and especially SSUHC is difficult to correctly classify. There are several reasons which may explain the poor results. The first is the treshold values used when creating the classes. The SSUHC class is created with using SW between 90 and 60% and VWCL below 25%, and this distinction may not be good enough to separate it from a SSCHC. The SSCHC class was created with the SW between 60 and 0% SW, and this gives it a clear distinction, especially for the deep resistivity logs (Fig. 5.3) and I believe this is the main reason for it having significantly higher accuracy than the SSUHC.

The muddy sandstone class (MUDSS) is suffering from the same problem as the SSUHC class. The VWCL interval for this class is between $25-40\%$ giving it an interval of 15 percent points. This is significantly different from the SMUDST and MUDST class which have $40-75\%$

and 75 − 100% VWCL content respectively giving them both a 35 percent points interval. The higher interval width gives the algorithm additional prediction leniency, meaning that the data may vary more within the class. The false predictions within the MUDSS, SMUDST and the MUDST is mostly shared between the three different models. While the normal and enhanced models all have their false predictions fall mainly between the three classes, the normalized model is seen to predict in a more erratic fashion. This is likely related to the effects of normalization, where outliers does not get weighted as much.

Coal is in a class of its own in terms of accuracy, with the initial and enhanced model having 96.6% and 98.1% accuracy respectively. This high accuracy is interpreted to be related to its distinct features. From the logplots in Fig. 5.2, coal is easily distinguished by its low density and its high neutron and sonic readings which makes it easy for the XGBoost algorithm to correctly predict it. These distinct features are dampened/mitigated when normalizing the data and is reflected in the poor accuracy of 74.6% for the normalized model shown in Fig. 6.2.

7.2 Prediction accuracy in individual wells

Table 6.1 provides details on how the prediction accuracy is influenced by the number of well logs available. By using the enhanced model as reference, wells 30/3-1 R and 30/2-1 are the least accurate with 35.7% and 58.7% respectively. Coincidentally the wells has a high accuracy in the normalized model with 70.6% for 30/3-1R and 69% for 30/2-1. The normalized model has a tendency to predict a higher amount of sandy mudstone than the other models as shown by the total distribution in Figure 6.5. Figure 7.1 shows distribution plots of the two wells and the reference distribution shows that both wells has a high amount of sandy mudstone. The abnormally high accuracy by the normalized model in the two wells can therefore be explained by the lithoclass distribution within the wells.

The feature importance plots from Figures 6.1B, 6.2B and 6.3B show how the density and neutron logs play a major part in the lithology classification. Without the presence of those logs, the accuracy will suffer. Wells 30/9-14 and 30/11-6 S has the highest degree of accuracy with 88.1% and 83.8% with seven and nine logs available respectively. There is therefore a link related to how many logs available in a well, and the prediction accuracy of the ML model. The

higher amount of logs available, the more accurate the predictions can be expected to be. Additionally, with increasing amount of logs available, it may be possible to have additional classes and still expect the model accuracy to be within acceptable accuracy since there will be enough difference within the data to distinguish between the classes.

The poor accuracy may also stem from overlap within lithoclasses, they are not different enough within the data for the algorithm to correctly distinguish between two or three classes which are highly similar.
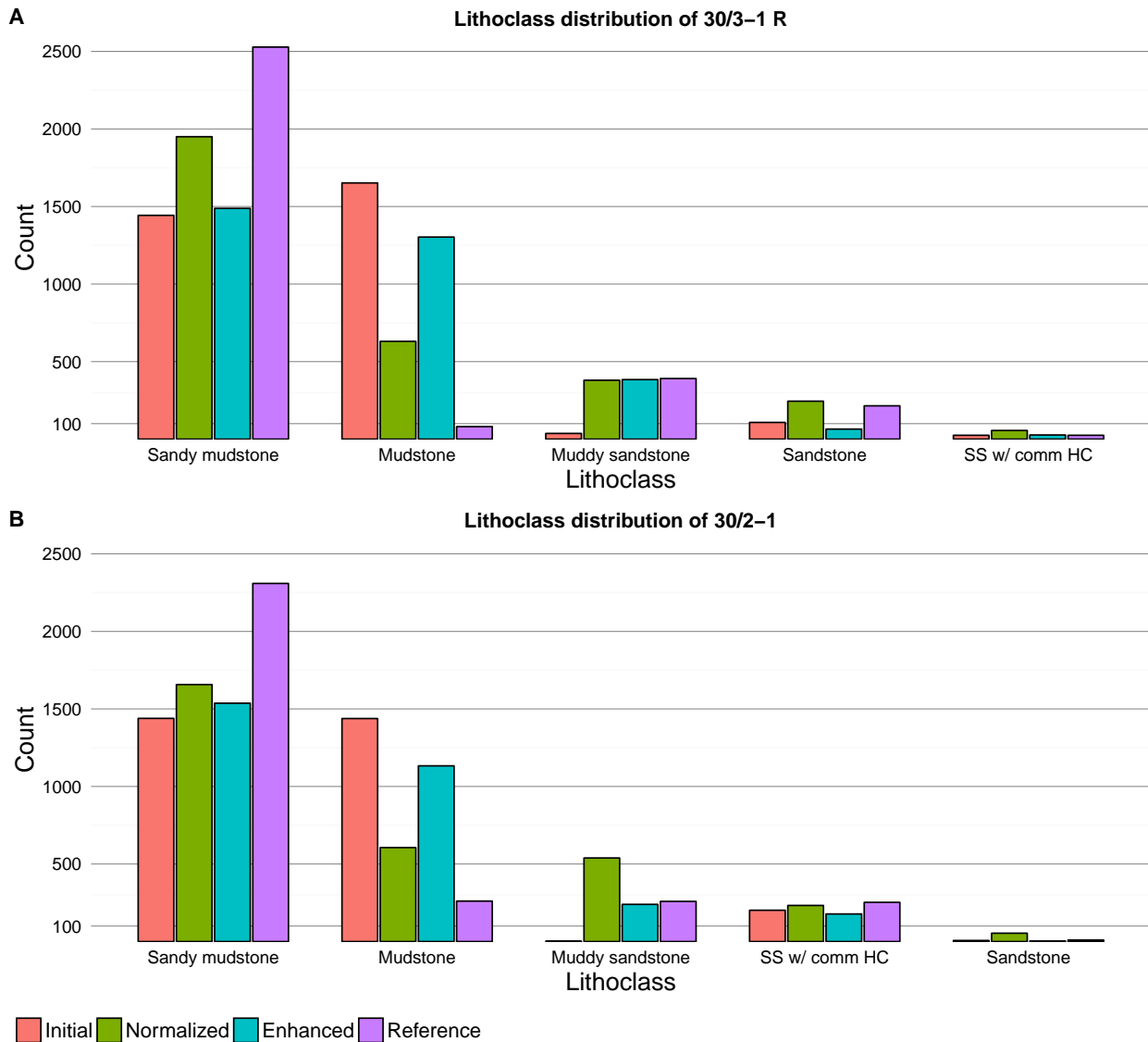
Figure 7.1: Distribution plots of well 30/3-1R (A) and 30/2-1 (B). The wells contains high amounts of sandy mudstone and combined with the normalized model's tendency to predict sandy mudstone the result is a falsely perceived high accuracy for the normalized model.

7.2.1   30/9-14

Well 30/9-14 is the well that scored highest in terms of accuracy with 88.1% using the enhanced model. There are some key properties of this well which may have impacted the accuracy in a positive way: The Jurassic interval in the well is seen in Figure 3.1 to be on a structural high and table 4.1 shows that the top Jurassic is at 2968 meters, while the bottom Jurassic is at 3680 meters (MD). The well is vertical and the differences between measured depth and true vertical depth are negligible. The Jurassic thickness is 712 meters and the well is drilled within the confines of the Oseberg Sør field (NPD 2018).

The fact that the well has been drilled on the Oseberg Terrace has given it a good amount of correletable data. Table 4.1 Figures 3.1 and 4.1 show that a majority of the wells included in this study has the Jurassic interval on a structural Terrace or Graben high. The implications of this is that the majority of the Jurassic interval in the wells used for training data and well 30/9-14 has low differences in burial depth. The low differences is expected to cause good correlation between the training data and well 30/9-14 as the burial depth has a direct influence on the porosity and density values (Ramm and Bjørlykke 1994).

From Fig. 6.6 the normalized model is seen to be predicting more sand than the other models in the interval $3100 - 3135$m. The GR values within this interval are slightly lower than similar intervals in well 30/11-8A and from the feature importance plot of the normalized model in Fig. 6.2B it is evident that the GR log plays a significant role when predicting the lithoclasses. This combination is likely the reason for the normalized model having such high amounts of sandstone predictions, while the other two models does not rely on the GR log as much in their predictions.

7.2.2   30/11-8A

Well 30/11-8A has equal accuracy in the initial and enhanced model with an accuracy of 69.6% (Table 6.1) and 45.1% for the normalized model. Figure 3.1 shows that the Jurassic interval is within a structural low (Viking Graben) and Table 4.1 shows that the top Jurassic is at 3671 meters and the bottom Jurassic is at 4475 meters (MD). The well is a deviation well and the equivalent top and bottom depths in TVD is $3472 - 4263$ meters (Marello et al. 2016). The difference of approximately 500 meters in burial depth between Top Jurassic of 30/9-14 and 30/11-8A may be a contributing factor for the reduction in accuracy, as it can cause lower correlation between the training set and the well in terms of log signatures. The porosity reduction trend for sandstone is quite steep steep between $2 - 3.5$ km (Blazevic Vucelic 2017) and the reduction will increase the density and reduce the neutron log signatures which can affect the prediction accuracy.

The interval between $3600 - 3620$m in well 30/11-8A (Fig. 6.8) shows the challenge with the petrophysical filters. The VWCL is defined at the border between sandstone with hydrocarbons and a muddy sandstone, and this is reflected in the reference log where it is alternating between two lithoclasses. From a geologic standpoint, the interval would be marked as a commercial zone, and the ML models do interpret a high percentage of the zone as a hydrocarbon zone. The primary cause of this is because the models does not see the VWCL and SW values their limits are less rigid than the reference model. Therefore the accuracies shown in the confusion matrixes (Figures 6.1A, 6.2A, 6.3A) and Table 6.1 should be interpreted as indicative, not absolute.

It is also evident that the petrophysical filters defined in Table 5.2 gives a floating transition, not a rigid one and in a geological view that can be interpreted as a good trait. Varying the limits of the petrophysical filters allows us to fine-tune how the zones should be defined, e.g. from an economic perspective or from a sedimentological perspective. In the interval between $3620 - 3640$ in Fig. 6.8 there are several smaller hydrocarbon zones which are not identified by the ML models, this may be further improved by fine-tuning the petrophysical filters.

7.3   Model comparisons

To generalize the models, the initial or baseline model shows the maturity and potential of applying ML on well logs, whereby only using the raw logs without any modifications it was able to achieve an average accuracy of 65.5% (Table 6.1).

The normalized model shows a general decrease in accuracy and based on it's accuracy show no applicability at all. The model seems to be a classic example of underfitting, where the model is too general to capture the underlying structure of the data. This is especially noticeable in the Coal lithoclass which should be easy to identify as seen for the other two models.

The enhanced model improves the accuracy significantly in wells where a high amount of well logs are available when compared to the initial model and is a good example on how generating additional features can improve the quality of a ML model. This accuracy is expected to be improved even further by generating additonal features and performing more rigorous optimalization of the algorithm parameters.

Chapter 8

Conclusions

This thesis has generated and applied three different models for predicting lithofacies within the Jurassic sequence on specific wells in quadrant 30 by using supervised machine learning. Well log data from 26 wells has been concurrently analyzed with the use of boxplots and seven different lithoclasses has been defined based on CPI's provided by Exploro. One well was removed due to lack of data in the Jurassic interval. The models have been trained by using well log data from 14 wells without the use of CPI's and they have been applied to 11 wells.

The key findings from this study are:

- Structured machine learning enables working on multiple wells and performing adjustments on data input, number of categories and number of well logs in a short period of time and in a fast and consistent manner. Lateral variations in terms of depositional environment and varying burial depth need to be addressed when deciding on which wells to include in the training data to achieve optimal prediction accuracy.

- The preprocessing steps involving loading and merging of well data, analysis and removal of bad data is the most critical part of a machine learning routine. A lackluster effort in this area will result in a poor ML model performance.

- The choice of petrophysical filters makes the lithoclass grouping rigid/strict and does not take into account the natural variability of geology. This affects the reference model and the perceived accuracy of the models, where from a geologic perspective the perceived errors are in many cases not significant. A refinement of the lithoclass definitions combined with manual quality control of the classes within the training data is expected to lead to better results.

- The number of well logs has a direct influence on the accuracy of the machine learning models, where a higher number of well logs available led to better results for the initial

67

and enhanced model. For an acceptable predicting power the wells should contain minimum the gamma ray, density and neutron log. To accurately differentiate between water-bearing and hydrocarbon intervals in sandstones the resistivity logs should be present.

- The enhanced model improves the accuracy significantly in a majority of the wells compared to the other models and is the best option for further use in additional wells. The addition of relative features by dividing multiple logs by their P99 value proved to be a valuable feature enhancement operation.

Chapter 9

Implications of work

The impact of this presented work has high potential in exploration, as high resolution information about lithological changes is a crucial knowledge in exploration. Detailed well knowledge need to be accounted for in order to reduce the exploration risk. In the mature province of Northern North Sea, the large amount of data is challenging as the time required to organize and quality ensure the data is huge. The use of ML in this thesis has shown that it is a system that provides quick and consistent high resolution lithology interpretations which represents a solution to utilize the enormous amount of data available. Additional refinement of this methodology with its improved lithology results can gain more predicting power for:

1. Localizing lithological heterogeneity and boundaries

2. Estimate borehole quality

3. Define zonation's of intervals with similarly petrophysical characteristics including hydrocarbon bearing intervals.

Chapter 10

Further work

- The petrophysical filters used are rigid and in some cases they generated overlapping problems, where multiple classes had similar data signatures. A refinement of the classes e.g. by using different petrophysical filters would probably lead to improved results. Additionally there are several ways to define the lithoclasses, developing classes in collaboration with a sedimentologist or petrophysicist may lead to a different group of lithoclasses.

- A natural extension to the lithofacies predictions in this thesis is predictions of depositional environment and transition zones between them. Bad hole zones may also be identified by adding another class.

- Both the normalized and the enhanced model used a max depth of 12 as hyperparameter setting. This may have been too high and can have caused overfitting of the models. The implications of applying a lower max depth should be assessed in the future.

- A significant portion of the work done in this thesis has focused on data quality and preparation. Further improvement is possible by performing more advanced data analysis to understand the process variables. One possibility is to use principial component analysis (PCA) as done in Brandsegg et al. (2010).

- The work done in this thesis has concentrated on using supervised ML. Unsupervised ML has shown good potential for grouping/clustering data with large variations or unstructured data (Thurmond 2018). The combination of unsupervised and supervised ML may result in a better understanding of patterns within the data which has not been observed in the supervised ML routine alone and vice versa.

References

Bertozzi, W., D. Ellis, and J. Wahl (1981). The physical foundation of formation lithology logging with gamma rays. *Geophysics*, 46(10):1439–1455.

Blazevic Vucelic, L. A. (2017). Modeling burial induced changes in physical sandstone properties-a case-study of north sea and norwegian sea sandstone formations. Master's thesis, NTNU.

Brandsegg, K. B., E. Hammer, and R. Sinding-Larsen (2010). A comparison of unstructured and structured principal component analyses and their interpretation. *Natural resources research*, 19(1):45–62.

Breiman, L. (1997). Arcing the edge. Technical report, Technical Report 486, Statistics Department, University of California at Berkeley.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Bryant, R., R. H. Katz, and E. D. Lazowska (2008). Big-data computing: creating revolutionary breakthroughs in commerce, science and society.

Bühlmann, P. and S. Van De Geer (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

CGG (2018). Cgg geosoftware releases powerlog 9.7.2 petrophysical software with machine learning capabilities. https://www.cgg.com/en/Media-and-Events/Media-Releases/2018/06/CGG-GeoSoftware-Releases-PowerLog-971-Petrophysical-Software-With-Machine-Learning-Capabilities. Accessed: June-2018.

Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, Pp. 785–794. ACM.

Christopher, M. B. (2016). *PATTERN RECOGNITION AND MACHINE LEARNING*. Springer-Verlag New York.

Cognite (2018). Cognite: Creating real-time models of Aker BP. `https://cloud.google.com/customers/cognite/`. Accessed: May-2018.

Cover, T. and P. Hart (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, Pp. 215–242.

Ellis, D. V. and J. M. Singer (2007). *Well logging for earth scientists*, volume 692. Springer.

Equinor (2018). Digitalisation in our dna. `https://www.equinor.com/en/how-and-why/digitalisation-in-our-dna.html`. Accessed: May-2018.

flowingdata (2008). how to read and use a box and whisker plot. `https://flowingdata.com/2008/02/15/how-to-read-and-use-a-box-and-whisker-plot/`. Accessed: May-2018.

Frankel, F. and R. Reid (2008). Big data: Distilling meaning from data. *Nature*, 455(7209):30.

Friedman, J., T. Hastie, R. Tibshirani, et al. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2):337–407.

GEOExPro (2017). seismic interpretation with machine learning. `https://www.geoexpro.com/articles/2017/01/seismic-interpretation-with-machine-learning`. Accessed: May-2018.

geoforskning (2017). ung, lovende, analytisk og lærevillig (norwegian). `http://geoforskning.no/nyheter/olje-og-gass/1621-ung-lovende-analytisk-og-laerevillig`. Accessed: May-2018.

Gluyas, J. and R. Swarbrick (2013). *Petroleum geoscience*. John Wiley & Sons.

Gubbi, J., R. Buyya, S. Marusic, and M. Palaniswami (2013). Internet of things (iot): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7):1645–1660.

Hall, B. (2016). Facies classification using machine learning. *The Leading Edge*, 35(10):906–909.

Hall, M. and B. Hall (2017). Distributed collaborative prediction: Results of the machine learning contest. *The Leading Edge*, 36(3):267–269.

Helland-Hansen, W., M. Ashton, L. Lømo, and R. Steel (1992). Advance and retreat of the brent delta: recent contributions to the depositional model. *Geological Society, London, Special Publications*, 61(1):109–127.

Ho, T. K. (1995). Random decision forests. In *Document analysis and recognition, 1995., proceedings of the third international conference on*, volume 1, Pp. 278–282. IEEE.

Honoré, Y., M. N. Haque, L. Isabelle, F. J. Inge, and L. Johan (2018). Integrating faciesbased bayesian inversion and supervised machine learning for petrofacies characterisation in the snadd formation of the goliat field, sw barents sea. *Geophysical Prospecting*.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An introduction to statistical learning*, volume 112. Springer.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69.

Kreyszig, E. (2010). *Advanced engineering mathematics*. John Wiley & Sons.

Larsen, E. (2018). Machine learning assisted petroleum geoscience. `https://youtu.be/xMy6dBbMwio`. Janet Watson Meeting 2018.

Laurae (2016). Let me learn the learning rate (eta) in xgboost! (or in anything using gradient descent optimization). `https://medium.com/data-design/let-me-learn-the-learning-rate-eta-in-xgboost-d9ad6ec78363`. Accessed: May-2018.

Marello, L., A. Antobreh, K. Brandsegg, A. Črne, C. Gyllenhammar, V. Kalashnikova, T. Kjennerud, B. Klein, J. Knudseth, J. Sandvik, G. Salvaggio, B. Seeland, J. Tetreault, H. Østby, and R. Øverås (2016). North sea toolkit - part 1. Technical report, Exploro AS and partners proprietary multi-client report.

Mondol, N. H. (2015). Well logging: Principles, applications and uncertainties. In *Petroleum Geoscience*, Pp. 385–425. Springer.

NPD (2018). Norwegian Petroleum Directorate factpages. `http://factpages.npd.no`. Accessed: January-2018.

O&G-Journal (2017). Unconventional resources, proving ground for applied analytics. `https://www.ogj.com/articles/print/volume-115/issue-7/special-report-applied-analytics/unconventional-resources-proving-ground-for-applied-analytics.html`. Accessed: May-2018.

Opitz, D. and R. Maclin (1999). Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, 11:169–198.

Picard, M. D. (1971). Classification of fine-grained sedimentary rocks. *Journal of Sedimentary Research*, 41(1).

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramberg, I. B. (2008). The making of a land: geology of norway. Geological Society of London.

Ramm, M. and K. Bjørlykke (1994). Porosity/depth trends in reservoir sandstones: Assessing the quantitative effects of varying pore-pressure, temperature history and mineralogy, norwegian shelf data. *Clay minerals*, 29(4):475–490.

Rider, M. H. (2011). The geological interpretation of well logs.

RStudio Team (2018). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

Sneider, J. S., P. de Clarens, and P. R. Vail (1995). Sequence stratigraphy of the middle to upper jurassic, viking graben, north sea. In *Norwegian Petroleum Society Special Publications*, volume 5, Pp. 167–197. Elsevier.

Thompson, K. (1968). Programming techniques: Regular expression search algorithm. *Communications of the ACM*, 11(6):419–422.

Thurmond, J. (2018). Keynote: When failure (a lot of failure) becomes an option - machine and deep learning on seismic data. `https://youtu.be/jjoz3sEK1Lc`. Janet Watson Meeting 2018.

Tibshirani, R., G. James, D. Witten, and T. Hastie (2013). An introduction to statistical learning-with applications in r.

Total (2018). Presentation of current research. Oral presentation in Rio de Janeiro, Brazil, march.

Tukey, J. W. (1977). *Exploratory data analysis*, volume 2. Reading, Mass.

Waldeland, A. and A. Solberg (2017). Salt classification using deep learning. In *79th EAGE Conference and Exhibition 2017*.

Werbos, P. (1974). Beyond regression: new fools for prediction and analysis in the behavioral sciences. *PhD thesis, Harvard University*.

Wickham, H. and L. Stryjewski (2011). 40 years of boxplots. *Am. Statistician*.

Worldoil (2018). Total partners with google cloud to develop artificial intelligence solutions for e&p. http://www.worldoil.com/news/2018/4/25/total-partners-with-google-cloud-to-develop-artificial-intelligence-solutions-for-ep. Accessed: May-2018.

XGBoost (2016). Notes on parameter tuning. `http://xgboost.readthedocs.io/en/latest/how_to/param_tuning.html`. Accessed: May-2018.

Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*. CRC press.

Ziegler, P. (1982). Geological atlas of central and western europe. *Shell International Petroleum, Maatschappij BV, Amsterdam.*

Appendix A

Libraries used in R

Table A.1 summarize the libraries used with R, sorted by topic.

Table A.1: R libraries used

Data management	
data.table	High-performance version of base R's data.frame.
xlsx	Read/write .xlsx files
ML related	
xgboost	R interface of the XGBoost algorithm
MLmetrics	Misc ML functions
caret	Hyperparameter optimalization interface (grid search)
parallel	Parallelization of grid search
doParallel	Parallelization of grid search
Data visualization	
ggplot2	Generation of figures
cowplot	Extension to ggplot2, adds custom theme
egg	Misc functions for ggplot2, used when generating logplots.
Misc	
maptools	Generating shapefile from well coordinates
rgdal	Shapefiles load/write and UTM zone conversions
stargazer	Automatic LaTeX table generation from data.table

Appendix B

Optimalization of tree ensemble model

The following derivation is based on the derivation from (Chen and Guestrin 2016).

In ML, *regularization* is a process where additional information is introduced in order to prevent overfitting or to solve a ill-posed problem (Bühlmann and Van De Geer 2011). Thus the goal of introducing regularization in supervised ML is to reduce model complexity which leads to a reduction in over-fitting. A common regularization approach to Boosted trees is to reduce the number of nodes (leaves) to make simpler trees. For a full overview, let's start by looking at how a tree ensemble model predicts the output:

Because the tree ensemble model shown in Eq. 2.1 contains functions as parameters they can not be optimized using traditional optimization methods in Euclidean space (Chen and Guestrin 2016). The model is instead trained in an additive manner by using a second order approximation to optimize the objective in Eq. 2.3. Let $y_i^{\hat{(t)}}$ be the prediction of the i-th instance at the t-th iteration. To minimize the following objective it is necessary to add f_t

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t). \tag{B.1}$$

The result is that the f_t which has the highest improvement to the model in Eq. 2.3 is *greedily* added. *Greedy* in this context means that the algorithm uses several iterations to compute the result. For each iteration, the algorithm assumes that the result is obtained by selecting the best result. Thus the global optimum is obtained by selecting the local optimum at each iteration.

To quickly optimise the objective, a second order approximation can be used (Friedman et al. 2000)

$$L^{(t)} \simeq \sum_{i=1}^{n} [l(y_i, \hat{y}^{(t-1)}) + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)] + \Omega(f_t), \tag{B.2}$$

where $g_i = \partial_{y^{(\hat{t-1})}} l(y_i, y^{(\hat{t-1})})$ and $h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}^{(t-1)})$ are first and second order gradient statistics on the loss function. By removing the constant terms the following simplified expression at

step t can be obtained

$$\widehat{L}^{(t)} = \sum_{i=1}^{n} [g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)] + \Omega(f_t). \tag{B.3}$$

Define $I_j = \{i | q(\mathbf{x}_i) = j\}$ as the instance set of leaf j. By expanding Ω in Eq B.1 it can be rewritten into

$$
\begin{aligned}
\widehat{L}^{(t)} &= \sum_{i=1}^{n} [g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w^2 \\
&= \sum_{j=1}^{T} [(\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] + \gamma T
\end{aligned}
\tag{B.4}
$$

For a fixed structure $q(\mathbf{x})$ the optimal weight w_j^* of leaf j can be computed with

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \tag{B.5}$$

and calculate the corresponding optimal value by

$$\widehat{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T. \tag{B.6}$$

Eq B.6 can then be used as a scoring function to measure the quality of a tree structure q. Due to the sheer size of of all the possible tree structures q it is usually impossible to enumerate all of them. Instead it is possible to use a greedy algorithm that starts from a single leaf and iteratively adds branches to the tree (Chen and Guestrin 2016). Assume that I_L and I_R are the instance sets of left and right nodes after the split. By letting $I = I_L \cup I_R$ the loss reduction after the split is given by

$$L_{split} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} + h_i + \lambda} \right] - \gamma \tag{B.7}$$

This equation is usually used in practice for evaluating the split candidates.