APPLICATION OF MACHINE LEARNING TO LIMITED DATASETS: PREDICTION OF PROJECT SUCCESS

Sofie Bang, PhD Candidate,
Norwegian University of Science and Technology (NTNU);
sofie.bang@ntnu.no (corresponding author)

Magnus O. Aarvold, Researcher,
Norwegian University of Science and Technology (NTNU)

Wilhelm J. Hartvig, Researcher,
Norwegian University of Science and Technology (NTNU)

Nils O. E. Olsson, Professor,
Norwegian University of Science and Technology (NTNU)

Antoine Rauzy, Professor,
Norwegian University of Science and Technology (NTNU)

SUMMARY: Much research is conducted on the importance of success factors. This study contributes to the body of knowledge by using artificial intelligence (AI), specifically machine learning (ML), to analyse success factors through data from construction projects. Previously conducted studies have explored the use of AI to predict project success and identify important success factors in projects; however, to the extent of the authors' knowledge, no studies have implemented the same method as this study. This study conducts quantitative analysis on a sample of 160 Norwegian construction projects, with data obtained from a detailed questionnaire delivered to relevant project team members. The method utilises ML through a Random Forest Classifier (RFC). The findings obtained from the analysis show that it is possible to use AI and ML on a limited dataset. Furthermore, the findings show that it is possible to identify the most important success factors for the projects in question with the developed model. The findings suggest that a group of selected processes is more important than others to achieve success. The identified success factors support the theoretically acknowledged importance of thorough and early planning and analysis, complexity throughout the project, leadership involvement, and processes supporting project success.

1. INTRODUCTION

Over the recent years, AI has made a significant impact in the industries where it is applied; including manufacturing (Lee et al., 2018), energy (Sozontov, Ivanova and Gibadullin, 2019), agriculture (Misra et al., 2020), and petroleum (Rahmanifard and Plaksina, 2019), among others. The construction industry has increasingly applied new technology to digitise and digitalise the workflow but remains in a nascent stage (Oliver Wyman, 2018). This study explores how AI can be utilised to analyse a selection of project data to identify important success factors in a project and addresses the following two research questions:

- RQ1: How can AI, specifically ML, be applied to analyse limited datasets from project evaluations?

- RQ2: Based on such an analysis, what are the most important factors for project success?

Project success is fundamental to the competitiveness of a company. Multiple definitions of project success exist, and there are different types of success within one project (Hussein, 2016). Despite the potential that is demonstrated to lie within AI, evidence shows that the construction industry lags behind other sectors, both in terms of productivity and the adoption of new technology (McKinsey Global Institute, 2017). New technology and tools, along with new areas of applications are constantly delivered to the market, and AI-based technology has recently regained momentum (Loureiro, Guerreiro and Tussyadiah, 2020). The industry operates with small margins, and the need to implement new, smart technology to accommodate the market is recognised (Deloitte AI Institute, 2020). Research suggests that technology and areas of applications becoming more common could contribute to the adoption in the industry, as well as increased digital maturity (Cubric, 2020).

Success factors relate to different aspects of a project - certain success factors relate to organisational complexity, others to the experience level of the project manager, coordination, or productivity (Chua et al., 1997; dos Santos et al., 2019).

Both academics and practitioners are exploring the use of AI to predict project success and identify critical success factors. Several techniques are utilised in previous (Magaña and Fernández Rodríguez, 2015) including neural networks (Chua et al., 1997; Wang, Yu and Chan, 2012) and regression analysis (Dvir et al., 2006).The body of knowledge on project success and the use of AI in the construction industry is growing. This study will build on the existing body of knowledge to explore the application of ML to a limited dataset and how it can be used to identify critical success factors.

The paper is divided into the following sections. First, the theoretical framework is presented, covering relevant aspects of the three topics of project management, project success and AI in construction. The following section describes the methodology of the study, including an analysis of the utilised dataset, insights, cleaning, splitting of data, and ultimately implementation of ML. Subsequently, the findings are presented, followed by a discussion of the model itself and its findings. Limitations of the study are evaluated, and suggestions for further research are presented. The last section concludes with an assessment of all previous sections.

2. THEORY

2.1 Project Management

A Guide to the Project Management Body of Knowledge (PMBOK) defines a series of knowledge areas that should be inherent in a project (Project Management Institute, 2017): the management of integration, scope, time, cost, quality, human resources, communication, risk, procurement, and stakeholders. Hwang and Ng (2012) identify schedule management and planning, cost management, quality management, human resources management, and communication management as the most important areas. At the same time, the field is constantly developing, and the knowledge requirements for project managers are changing with it, along with fundamental roles and functions in the project team (Russel, Jaelskis and Lawrence, 1997; Edum-Fotwe and McCaffer, 2000). A shift can be seen from the traditional responsibility of technical content of the project, the reliability of the facility and within-cost performance to include additional responsibility in non-engineering knowledge to meet expectations and demands for professionalism and expertise.

The majority of projects experience cost and time overruns to some extent, despite the availability of project control techniques and the increased utilisation of digital tools (KPMG, 2015; Project Management Institute, 2018). A report from McKinsey Global Institute (2017) indicates that the rate of productivity in the construction

industry has been stagnant and thus remained at the same level for decades. Nationally, the Norwegian construction industry has seen a 10% decrease in productivity from 2000 to 2016, whereas the total productivity in mainland Norway has increased by 30% in the same period (Todsen, 2018). This evidence supports the need to elevate the efficiency of these sectors, and research suggests the field and the industry is ready for disruption (Agarwal et al., 2016; Assaad, El-Adaway and Abotaleb, 2020).

Increased digitalisation and introductions of new technology are already making waves in the industry (Vikan, 2018; Brekkhus, 2017). Adapting to new conditions and circumstances is crucial to maintaining a lasting and sustainable industry.

2.2 Project Success

According to Ika (2009), the research on project success can be divided into project success criteria or critical success factors (CSFs). The findings suggest that the definition of project success has evolved. Definitions have traditionally been based on the iron triangle, including time, cost, and quality. Later definitions are seen to include more dimensions of the projects, such as their relation to stakeholders, project team, and end-user, as well as strategic objectives.

Hussein (2016) suggests a difference between the factors necessary to achieve project management success, project success, and long-term strategic success. The same distinction between project management success and strategic success seems to be supported by the literature in general, among others, Samset and Volden (2016). Project management success is generally seen to relate to the fulfilment of project objectives (de Wit, 1988) and traditional measurements of time, cost, and quality (Radujkovic and Sjekavica, 2017). These are easily quantifiable. Therefore, project management success (hereby referred to as 'project success') constitutes the foundation for this study.

A success factor is, by definition, a condition, event, or circumstance that contributes to project success. Certain success factors are attributed to specific project characteristics (Hussein, 2016); for instance, if there is organisational complexity in the project structure, the project will need (1) a good flow of information, (2) clear roles and responsibilities, and (3) project manager authority in order to achieve success. Chua et al. (1997) identified eight significant success factors for predicting success, in descending order of significance:

- Number of organisational levels between project manager and craftsmen

- Percentage of the detailed design completed at construction start

- Frequency of control meetings during the construction phase

- Number of budget updates per year

- Implementation of constructability program

- Project-team-turnover rate

- Control system budget

- Project manager's technical experience in a project with similar technology

2.3 AI in Projects

The concept of AI has been around for decades (Russell and Norvig, 2003), often associated with science fiction and human-like robots; this has created an inaccurate picture of what AI is. Numerous definitions exist, recent ones including 'the science and engineering of making intelligent machines' (ScienceDaily, 2020) and 'the field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem-solving and pattern recognition (Marr, 2020). The field experienced a renaissance around 2000 and has since sparked the debate on whether the increased interest is a 'hype' or a necessary step for businesses to maintain a competitive advantage (Walch, 2020). In the construction context, AI systems can be grouped into four categories: machine learning, knowledge-based systems, evolutionary algorithms and hybrid systems (Akinade, 2017).

Automated project management (APM) is the automation of software development tasks, typically organised as software projects (Campbell and Terwilliger, 1986). In general terms, APM contains all approaches for

automating project management tasks and activities (Auth, Jokisch and Dürk, 2019). The expectations of what AI can do still exceed the current possibilities that lie within the technology, and the broad and dynamic field of tasks of a project manager can currently only be automated in limited, clearly defined areas. Niu et al. (2019) highlight the potential of using AI for project managers to be more accurate, precise, and swift, and argue that smart construction objects can be effective tools for data collection, information processing, and decision support. In addition to characteristics that differ between individual projects, such as planning and reporting, the project manager relies on knowledge from previous projects. This information can be categorised as tacit knowledge. To utilise such knowledge in an AI context, the information contained needs to be made explicit. Kowalski et al. (2012) explore the use of AI as a tool for decision-making with input of know-how in the form of natural language.

Among the major challenges seen in overrun construction projects is delay risk, the time overrun from the date agreed upon for delivery (Assaf and Al-Hejji, 2005). Yaseen, Salih and Al-Ansari (2005) analysed prediction of risk delay using a hybrid AI model, using genetic algorithms and a Random Forest model. The model was proved to handle the nonlinearity and complexity of data used and demonstrated that such models can be utilised in the construction industry. Another demonstration is provided by Worldsensing (2020), connecting civil infrastructures to the Internet of Things (IoT) to continuously monitor assets and analyse risks. Project managers and decision-makers can receive insights into local operations, track relevant key indicators and use gathered information for analyses. Ultimately, these insights can be used to detect anomalies or anticipate needs. GHD (2020) has successfully applied ML on information collected from projects, to provide a dashboard of key measures for the project manager.

3. MATERIALS AND METHODS

This study is based upon a quantitative analysis of data obtained from construction projects, through the tool CII 10-10. The database is built through the project team members' submission of a questionnaire after chosen project phases. The theoretical framework presented in the previous section formed the basis for the preparation of the dataset, to ensure that no data was lost in the process.

The dataset was then loaded into a Python script, where the libraries Pandas, SKLearn, and NumPy were used. When a dataset is loaded into Pandas, it is called a data frame (DF). Figure 1 illustrates the steps of the analysis. First, the original dataset is processed through an exploratory data analysis (EDA) and preliminary cleaning, resulting in an initial DF. This DF is then split into nine purposed DFs before the next steps are carried out in order: main cleaning, labelling, train-test split, scale, train and fit, classification, and lastly analysis and plot of the results.
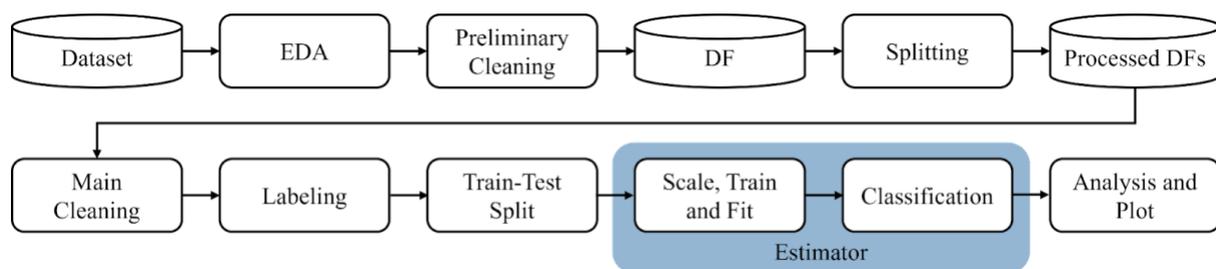

FIG. 1: Flow chart of the utilised methods in this study.
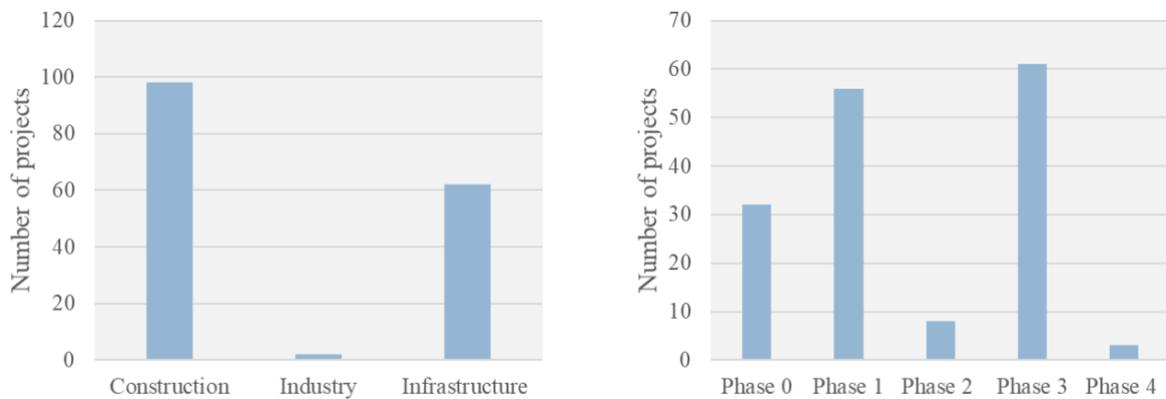
3.1 Dataset

The model was built on data from the CII Nordic 10-10 database. CII 10-10 is a tool for project benchmarking to develop and enhance processes continuously. It is developed and provided by the Construction Industry Institute at the University of Texas and has later been translated to fit the Norwegian construction industry, resulting in the Nordic 10-10 initiative. The tool provides the users with a report that evaluates their project and compares it to relevant projects in the database (Nordic 10-10, 2020). It is ultimately providing a report serving as a foundation for further discussion and improvements, for individual projects, and for the organisation as a whole. It has been proven that participating companies perform better than the industry average (Prosjekt Norge, 2017). The questionnaire used to obtain the data constituting the 10-10 datasets is upon input specified by sector

(construction, industry, or infrastructure) and project phase (phases 0 through 4). Consequently, some data points are only relevant to certain sectors or phases. To maximise the number of useful columns within each DF, it was decided to split up the DF.

The 10-10 dataset contains several different features, including the four categories of General descriptive data (G), Output ratings (O), Question scores (Q), and Project ratings (I). The Q-attributes are distinct, and closely related to the project sector and phase. Furthermore, they are divided into two categories, those under 40 and those over 100. The sub-40 questions are binary, while the above 100 questions are ranked on a scale from 1 to 5. For each given Q-attribute, they may only relate to one specific sector or phase. However, as there is more than one respondent for each project, the sub-40 Q-attributes will appear in the database as the average of the respondents' answers, resulting in a scale from 0 to 1.

3.2 Exploratory Data Analysis

A preliminary EDA confirms that the sample of projects comes from three sectors: construction, industry, and infrastructure. The EDA also shows that there were only two projects registered from the industry sector, illustrated in Figure 2(a). This is not enough data points for a meaningful analysis, and the projects contained in the category will consequently be discarded. What remains is the distribution of the remaining 160 projects and their phases, illustrated in Figure 2(b).

(a) Sector count plot. Before discarding industry projects.

(b) Phase count plot. Sectors combined.

FIG. 2: Count plots, aggregated on sector and phase.

3.3 Preliminary Cleaning

Algorithms for ML only appraise information as numbers. Consequently, columns and rows with a high percentage of missing data must be discarded. The dataset contains nominal numbering, for instance, the number corresponding to the respective phase and sector. In the original dataset, the construction sector is assigned the value '0' and the infrastructure sector is assigned the value '2' in the column called 'G1_Project-Category'. To avoid the inherent sense of scale, being that '2' is bigger than '0', dummy variables were introduced. The mentioned column would be split, where all projects originally assigned '0' would be assigned '1' in a new column called 'G1_Construction'; correspondingly, the projects originally assigned '2' would be assigned '1' in a column called 'G1_Infrastructure'. This procedure is illustrated in Figure 3.

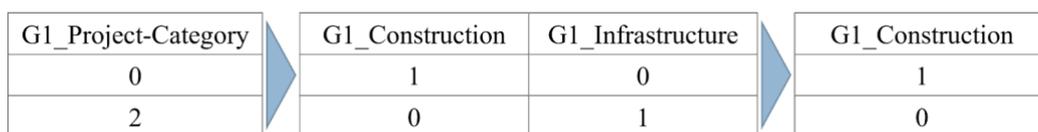| G1_Project-Category		G1_Construction	G1_Infrastructure		G1_Construction
0		1	0		1
2		0	1		0

FIG. 3: Illustration of one-hot-encoding.

The two new columns will contain the same information, only in reverse. This allows for deleting the second of the two columns, while keeping the information contained; this process is called one-hot-encoding. The same

procedure is done on the columns corresponding to the phases of the project. As a part of the preliminary cleaning, a few columns were discarded for further analysis, this included columns with a particularly high percentage of missing data, columns containing nominal data types, or deemed irrelevant for the analysis.

3.4 Splitting

The dataset was split between sectors and phases. More precisely, the split first made a copy of the sector DF and split it into each of the project phases. This way, it was not necessary to fill in the missing values, not available (NA) or not a number (NaN). This produced 12 DFs, one for each sector, and one for each of the five phases of each sector. The process is illustrated in Figure 4.
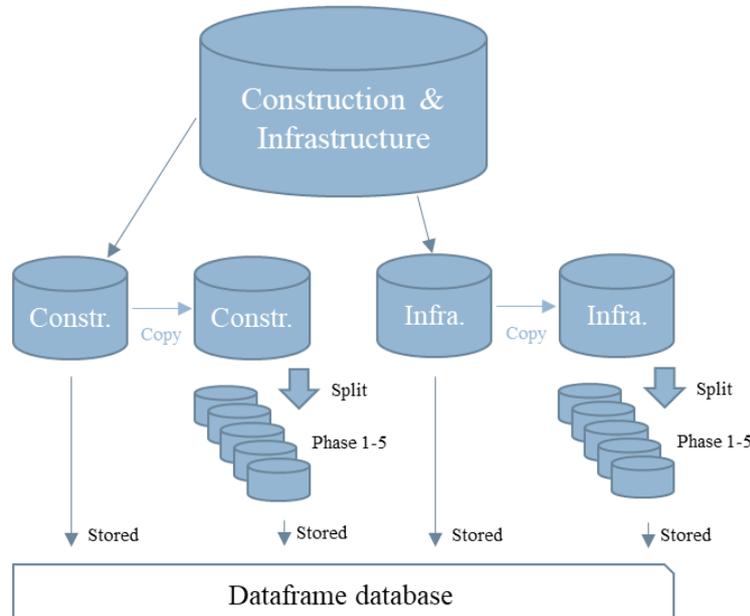

FIG. 4: Splitting and storing of the dataset illustrated.

Subsequently, various combinations of the DFs were evaluated. For instance, one combination was that the same phase from different sectors was joined together, or that phases 1 and 3 within a sector were combined. To keep a low number of DF and sort out the least relevant, only some combinations were further assessed. For example, if a DF had too few projects, or only successes or failures, they were discarded.

3.5 Main Cleaning

Several values were still missing in the DFs, and the next step consisted of investigating the percentage of missing values in each column of each sub-DF. It became apparent that some DFs had one projects missing a substantial number of columns, ultimately polluting the whole DF. A clean DF is one where all cells of the table are filled with legal values. If one cell is missing, the cell can be filled or the whole row or column can be removed. Three options were considered. Firstly, discarding the occurrence that polluted the columns; secondly, discarding the polluted columns; thirdly, filling the gap with an educated guess for the missing value. The third option was undesirable, as it would mean to temper with the datapoint on a limited foundation. Since a big number of different DFs were to be generated from the dataset, the second option was chosen for this model.

Then, a function was made to look for columns where all entries were of the same value. These columns would have provided no value to the estimator; thus, analysing these would mean wasting processing power and time. Therefore, if the function found one or more of these columns, it would remove these from the DF. The model approach to outliers is particularly important, especially for outliers classified as failures. The outliers represent projects that have gone far beyond budget or estimated time.

3.6 Defining Success and Labelling

A scoring system was established. Since project success is being predicted and evaluated, this feature must be explicitly quantified. The theoretical framework suggests that project success as defined in this study is three-

fold and based upon the three dimensions of 'the iron triangle': time, budget and quality, or specifications. The dimensions reflect whether the project is delivered within the set time frame, and similarly within the set budget and agreed-upon project specifications. The time and cost dimensions are well documented in the 10-10 dataset, in the columns O_01 and O_02 respectively. The values correspond to the percentage increase in cost and time for the given project, summarised in Equations 1 and 2 respectively.

$$\frac{G20\_PhaseRealCost\_MNOK}{G19\_PhaseEstCost\_MNOK} - 1 = O\_01 \tag{1}$$

$$\frac{G22\_RealVarFase}{G21\_EstVarFase} - 1 = O\_02 \tag{2}$$

The resulting output columns will be positive if the real value succeeds the estimated value, and negative otherwise. To quantify the specifications, and whether they were met, the column in the dataset 'Q149' was used. This column reflects the level of customer satisfaction regarding the deliveries of the specified phase on a scale from 1 to 5, as submitted by the questionnaire respondents. This feature was chosen based on the PMI definition of quality, which sees quality as 'how the inherent characteristics actually fulfil the set requirements, and to which degree this occurs' (Project Management Institute, 2017). Customer satisfaction is related to meeting specifications; however, it is not necessarily equivalent, as there could exist scenarios where the customer fails to specify exactly what they need.

The dataset also contains a feature labelled 'I7_Pr', in the CII system denoted 'Quality', which is deducted from other available features. However, the resulting 'Quality' feature is a conglomeration and can therefore be seen as less precise than the 'Q149', as different projects might utilise different combinations of features to determine the 'I7_Pr', even within the same sector and phase. The reason for this choice is further elaborated on in Section 3.7.

To make the scoring of the customer satisfaction compatible with the other two dimensions, the scoring had to be standardised. Therefore, the mean of the column was subtracted from each row, and then divided by the maximum score, which was 5. Equation 3 was used for all rows, in which *i* represents a single row.

$$\frac{Q149_i - Q149_{AVG}}{Q149_{MAX}} = customer\_satisfaction_i \tag{3}$$

The next step was to decide how the three dimensions should be combined to reflect project management success. Three solutions were considered, labelled A, A_fillNA and B.

3.6.1 Project success definition – Solution A

Solution A would quite simply be a summation of the three dimensions. The values of the three dimensions would at this point be of the same magnitude and could therefore be summated. However, positive values of the first two dimensions would negatively impact project success, as they reflect overruns in time and cost. The summation approach stems from the idea that if a project lasted 15% longer than estimated, and the cost was 15% less than estimated, the deviations would cancel each other out. Since positive cost and time dimensions imply longer and more costly projects than estimated, these were summed as a negative value. If the value for customer satisfaction was high, i.e., a value above 3 on the scale from 1 to 5, a good score would be positive after standardisation. Therefore, for the quality dimension, the value was kept positive for the summation. Solution A is illustrated in Equation 4.

$$-O\_01_i - O\_02_i + customer\_satisfaction_i = success\_score\_A_i \tag{4}$$

The next step would be to make the score binary. If the score was higher than 0, the binary score would become 1. Otherwise, it would become 0. A weakness of this method lies within the fact that if one of these features were missing from the dataset, the summation would become NaN, and thus useless. Consequently, many projects would have to be removed if one or more values were missing. One way to combat this would be by the use of the 'fillNa'-function in Pandas. The 'fillNa' function replaces the NaN with a value, so the project does not have to be discarded. Possible values to replace the NaN with are the overall mean, the mean of similar projects, or simply 0. For this examination, the latter was chosen. The 'fillNa' was not taken any further in this study but constitutes a potential for future studies and research. For this model, another solution was chosen.

3.6.2 Project success definition – Solution B

Solution B was classifying the projects through a two out of three (2oo3) voting system. To do this, a function that takes in variables for voting had to be implemented. This function took three arguments: a DF, a list of wanted columns, and a limit. First, the columns of interest, the three aforementioned dimensions, were located in the DF. Second, the function counted the number of not-NaN values in each column. Then, the value of cost and time was compared to the limit value. Different values for the limit were tested and the resulting success and failure counts of each DF were inspected; ultimately, 0 was chosen as the most objective and balanced limit. Successful projects had values equal to or higher than the set value. For the last column, the customer satisfaction, the value was compared with the weighted mean, 3. In this column, the successful projects would have a value equal to or higher than the resulting limit. The next step was to identify the outliers. To do this, the 'empirical two-sigma rule' was utilised, as illustrated in Equation 5.

$$Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545$$

[5]

In short, this rule says that an interval containing two standard deviations, σ, away from the mean, μ, covers approximately 95% of the distribution. Thus, the confidence interval will be $\bar{X} \pm 2 \sqrt{\sigma} n$ where n is the sample size which yield the average $\bar{X}$. So, if either the cost or time dimensions value deviated more than 2σ over the mean, the project was classified as an 'outlier - failure'. The second classification was is a 2oo3; if two or more of the dimensions have satisfactory values, the project is classified as a success. After this point, the unassigned projects either have two or more NaNs, or one NaN; this would lead to a tie. If two or more NaN values were found, the project was classified as a '2 or more nan'. These projects werediscarded, as they did not provide enough data points for a 2oo3 voting system to be implemented.

Furthermore, if the function found a NaN value, it investigated how to rule the tie. It checked which dimension were NaN, and the values of the two remaining dimensions. If the NaN value was customer satisfaction, and the remaining two values were of different signs, i.e., '+' and '-' it was classified as a 'tie - 1v1'. Since this is inconclusive, the project was discarded in further analysis. However, if the NaN value was only one of the dimensions, the function investigated the other dimension and compared this to the set limit. If the present dimension was considered satisfactory, meaning negative or 0, the project was classified as a success. If not, the project was classified as a 'tie - 1v1 - failure'. This classification was regarded as a 'failure' later in the function. Lastly, if two or more dimensions are higher than the limit, the project will be classified as a 'failure'. The count plots of all the categories are illustrated in Figure 5.
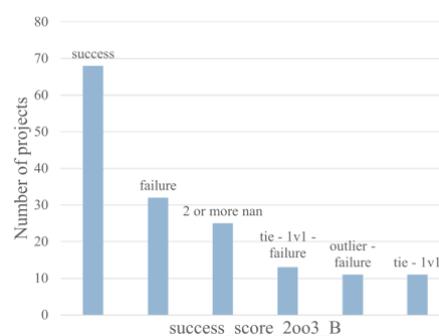

FIG. 5: Before discarding categories '2 or more NaN' and 'tie - 1v1'.

Subsequently, another function inspected the columns produced by solution A, A_fillNA and B. The function translated the classifications into a binary system where '1' denotes a successful project and '0' a failed project. Binary classification was chosen, as the dataset was small and preliminary analysis using regression yielded less than wanted accuracy. Table 1 summarises the count of the remaining projects of each solution. Solution A_fillNA and Solution B has the most remaining projects. Figure 6 illustrates this, showing the 41 projects that are retained within Solution B, but discarded in Solution A due to NaN values. In a small dataset, every project matters, and contributes to providing the model a more stable foundation for training and testing.

TAB. 1: Labels of the different solutions.

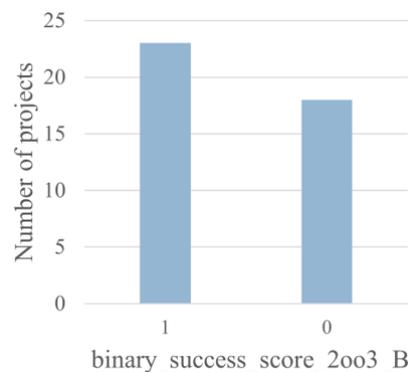| Solution	Success	Failure	Total
A	51	30	81
A_fillNA	75	49	124
B	68	56	124

FIG. 6: Count plot of the projects that were 'saved' by switching to Solution B.

A confusion matrix (CM) is plotted in Figure 7. The matrix corresponds to a sensitivity analysis where the two different solutions of labelling are shown. On one axis, the labels from Solution A are plotted; on the other axis, Solution B is plotted. On the main diagonal are the number of projects the two solutions labelled the same. High numbers on the diagonal implies that the solutions agree, which strengthens the reliability and validity of the models. The top right square (1,2) shows the number of projects that are deemed a success by Solution A, and a failure by Solution B; a false positive. The bottom left (2,1) square shows the opposite, false negatives. Solution B appears to be stricter than Solution A and Solution A_fillNA. However, this may not be entirely true, as the matrix only displays the projects that Solution A actually did label.
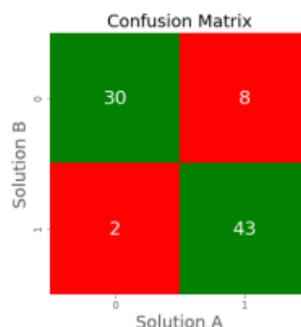

FIG. 7: CM of projects labelled by both solutions A and B.

3.7 Train-Test Split

Following the preceding steps, the DF was cleaned and ready for further analysis. First, the projects were shuffled to remove a possible bias in the original ordering. For the analysis itself, the Python library, SciKit

Learn (SKLearn), was utilised. The set of columns describing the DF labels were discarded, because they are mutually correlated, including the three success dimension columns, the resulting success column, all columns I1-I10 and the four columns on which O_01 (cost) and O_02 (time) are based on.

Already having a small DF to base the model on, the split between train and test data is of even bigger importance. The split process divides the labelled data in two: training and testing. After preliminary testing, an 80-20 split was chosen. It is desirable to retain as much data as possible to train the model, while leaving the model with enough data for testing, and scoring.

There are more successes than failures in the dataset, making the DF unbalanced - it is therefore necessary to stratify the data split. Stratifying ensures that the split of each set is approximately the same as the split of the complete set; if the complete set contains 20% of class 0, both the training and test split will have about 20% samples of class 0.

3.8 Estimator

The choice of an estimator for the model depends on whether the issue at hand is considered a regression or classification problem. Determining if a project is a success or not is a classification problem. The argument could be made that success is a subjective and continuous characteristic, but this model defines success based on the iron triangle, and thus as a binary factor of fail or success.

3.8.1 Scale, train and fit

Since the columns of the data set were of different magnitudes, a scaler was used to scale all columns. For this model, the MinMaxScaler was utilised. The MinMaxScaler scales all features sequentially, to a number between 0 and 1. The scaler was fit based on the training set, and subsequently transformed the test set.

3.8.2 Classification

For classification, several classifiers were tested, including LinearSVC, KNeighbors Classifier, MLPClassifier and Random Forest Classifier (RFC), which was ultimately chosen. A RF model uses multiple decision trees (DT) as the base learner. An inherent attribute of a DT is low bias but high variance. However, as the model aggregates over several DTs, and proceeds to calculate the mean of the DTs, the input variance decreases. The R2 score, the accuracy score, tends to overfit on the training data, yielding a score of 0.90 and higher. It is not desirable for the model to overfit, as this reduces the generalising properties of the model. RF is an ensemble method, meaning that the overfitting is reduced with a higher number of estimators. In this model, 100 DTs are used in each iteration. Additionally, the RFC provides an insight into the attributes of highest importance for the model to find the proposed label, increasing the transparency of the model. This enables an investigation of importance of the individual attribute, on a scale from 0 to 1.

In modelling, the simulated results become more accurate to the true result if the model is run a high number of times (Schwarz, 2015). It was therefore decided to run Monte Carlo Simulations (MCS) on this classifier. MCS introduces randomness to the variables, as well as a high number of iterations to create a nominal distribution of results (Oberle, 2015). From this distribution, a mean can be calculated. A higher number of iterations yields higher quality in the results, ultimately resulting in a higher quality of the mean. The model iterated 10 000 times over each DF. The law of large numbers (Kent State University, 2021) then states that the measured accuracy trends toward a number that is sufficient to use as the true value.

To balance the initially unbalanced datasets, selected functions in the SKLearn library were utilised. First, the built-in parameter called class_weight was set to 'balanced'. Next, the code implemented the built-in function of random search and grid search with cross-validation to find the best hyper parameters, such as max depth and number of estimators. No random state was set since this would counterweigh the effect of the MCS.

Then, the process of fitting was initiated. The fit function further contributed to decreasing the effects of an unbalanced dataset through sample weight. The argument for the sample weight of a function is another function using the training values; this is done to find a balanced class, and thus, sample weight. Upon completion of the fit process, the predicting could commence.

Predictions were stored as the variable 'y_pred' for further analysis. Both the f1-scoring method as well as the CMs use this variable. The built-in method RFC.score() function does not; thus, it does not catch, for instance,

true and false-positive predictions. RFC.feature_importances_ were utilised to retrieve the importance score of the features, and then stored in an appropriate format as a new DF. Consequently, this DF was sorted and sliced. Contributors with an importance score below 0.01 were discarded. Based on the f1-score of the prediction, the top five entries from this DF were stored in different tiers of lists. More precisely, if the f1-score was higher than 0.5, it was appended to a specified list. Similarly, if the score was higher than 0.7, 0.8 and 0.9, it was appended to other, respectively specified lists. If the score was higher than 0.8, the CM was also appended , into a list called 'cm_over_80'. When one MCS had reached its set number of iterations, all the lists were saved into another list as a list of top entries. This list, containing up to 10 000 entries, was stored as a single element in a new list; this originated the wording list of lists, as seen in Figure 8. Other lists were also established, summarised in Figure 8.
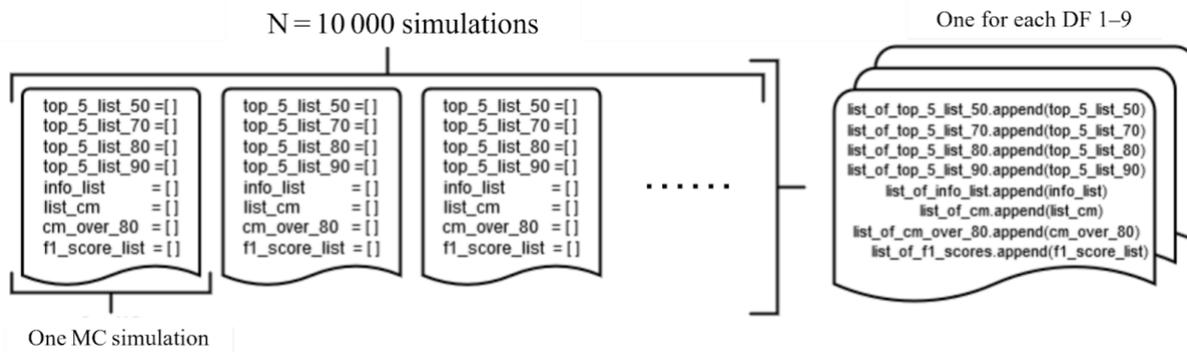

FIG. 8: Illustration of how lists, and lists of lists, are made.

Figure 9 outlines the method in its entirety. Every DF is simulated 10 000 times. In each of these 10 000 iterations, 100 DTs were made. The most accurate tree was used for further analysis, to determine whether the f1-score, the predictive performance of the model, was sufficiently high.

4. RESULTS

Important findings and characteristics of the models are presented in Table 2.

TAB. 2: DF splits and respective characteristics.

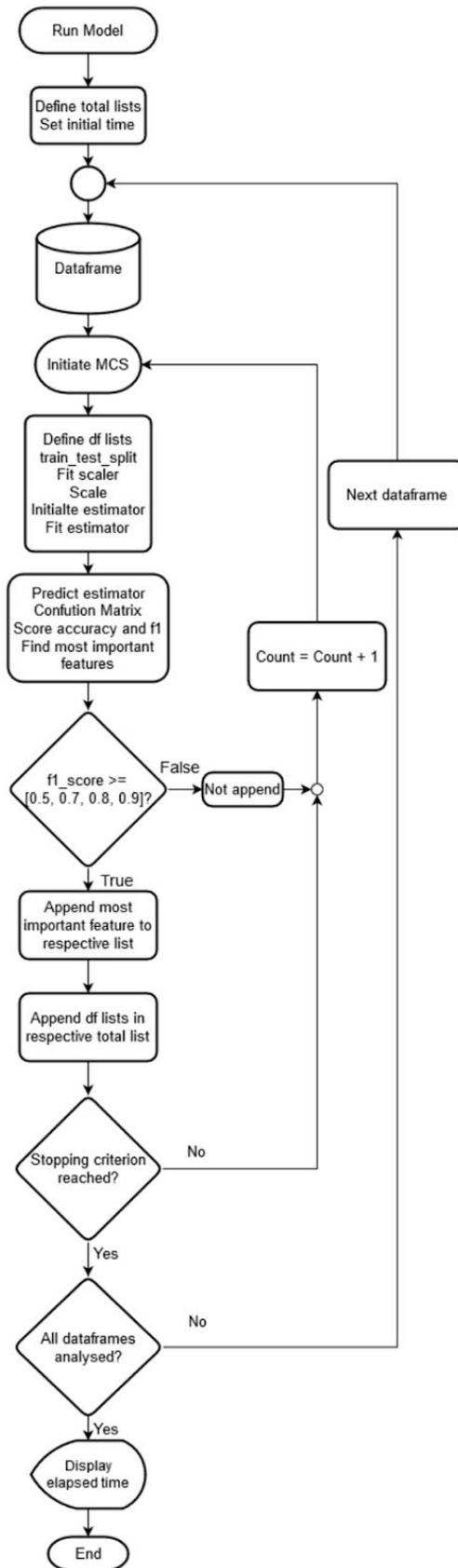| DF#	Split	Shape (row,col)	Shape after NaN discards (row,col)	# success	% success	Mean accuracy	Mean f1-score	Mean CM
0	Whole construction sector	(47,296)	(47,51)	28	60	0.54	0.62	[[1.5 2.5] [2.1 3.9]]
1	Whole infrastructure sector	(77,296)	(77,38)	41	53	0.7	0.72	[[4.8 2.2] [2.66 6.34]]
2	Construction phase 1	(22,296)	(22,86)	12	55	0.55	0.62	[[0.75 1.25] [1.01 1.99]]
3	Construction phase 3	(16,296)	(16,105)	12	75	0.69	0.8	[[0.01 0.99] [0.26 2.74]]
4	Infrastructure phase 1	(32,296)	(32,79)	14	44	0.73	0.64	[[3.32 0.68] [1.17 1.83]]
5	Infrastructure phase 3	(28,296)	(28,99)	20	71	0.73	0.81	[[0.83 1.17] [0.45 3.55]]
6	Both sectors only phase 0	(17,296)	(17,80)	7	41	0.48	0.05	[[1.82 0.18] [1.92 0.08]]
7	Both sectors only phase 1	(54,296)	(54,296)	26	48	0.67	0.64	[[3.93 2.07] [1.6 3.4]]
8	Both sectors only phase 3	(44,296)	(44,92)	33	75	0.67	0.78	[[0.45 1.55] [1.44 5.56]]

FIG. 9: Flow chart of the MCS and classifier.

The analysis will primarily focus on DFs 1, 4 and 7. These DFs were the ones that yielded the best results from the simulations, as described in the previous section, and are highlighted in Table 2. Since the dataset was limited, it is reasonable to assume not all DFs would be correctly predicted by the base model, even after implementing remedies such as built-in functions like sample weights, stratify and choice of classifier.

As illustrated in Table 2, DF 1, 4 and 7 were the only DFs in which 'Mean F1-score' are relatively large as compared to '% success'. This is further illustrated in Figure 10, where the subtraction of 'Mean F1-score' and '% success' in decimal form is illustrated. The '% success' column shows what a baseline classifier would get as accuracy, if the predictions were purely based on guesses; this means that the proficiency of the developed model will be implied by the Delta, the difference, between these two columns.

Figure 10 illustrates the difference between '% success' and 'Mean f1-score', the Delta. The Delta is the difference between the respective Mean f1-scores and %-success. Worth noting, the Delta score of DF 6 is -0.36, but is cropped out to illustrate the differences of the Delta scores more accurately in the remaining DFs.
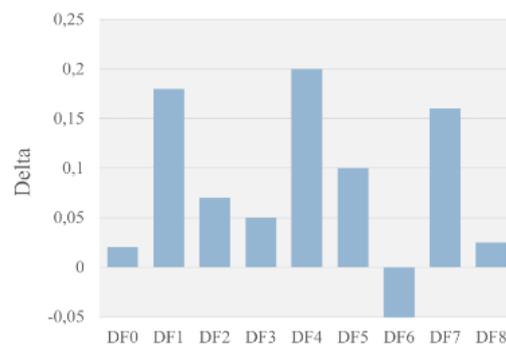

FIG. 10: Bar plot of Delta, converted to decimal value.

The first metrics that were analysed further were the CMs. As mentioned, a high number in the main diagonal is desirable. Element (1,1) is the true negative location in the matrices, and element (2,2) is the true positive location. In the off-diagonal, an as-low-as-possible value is preferable. Only the matrices of DFs 1, 4 and 7 showed a clear connection with this principle and were therefore selected for further analysis. The DFs are illustrated in Figures 11(a)-(c).
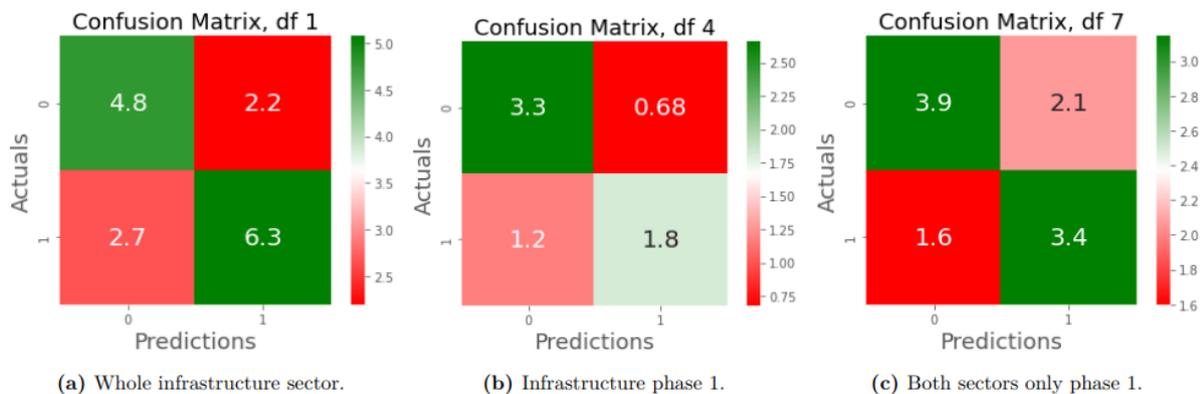

FIG. 11: CMs for DFs 1, 4 and 7.

Count- and density plots of the DFs are presented in Figure 12(a)-(c). It becomes apparent that the mean score is quite high. Moreover, the distribution of the bars looks to resemble a bell curve, referring to the inherent characteristics of an MCS (Oberle, 2015). From Table 2, we know that these DFs had a '% success' score of 50 ± 6% as their base score.
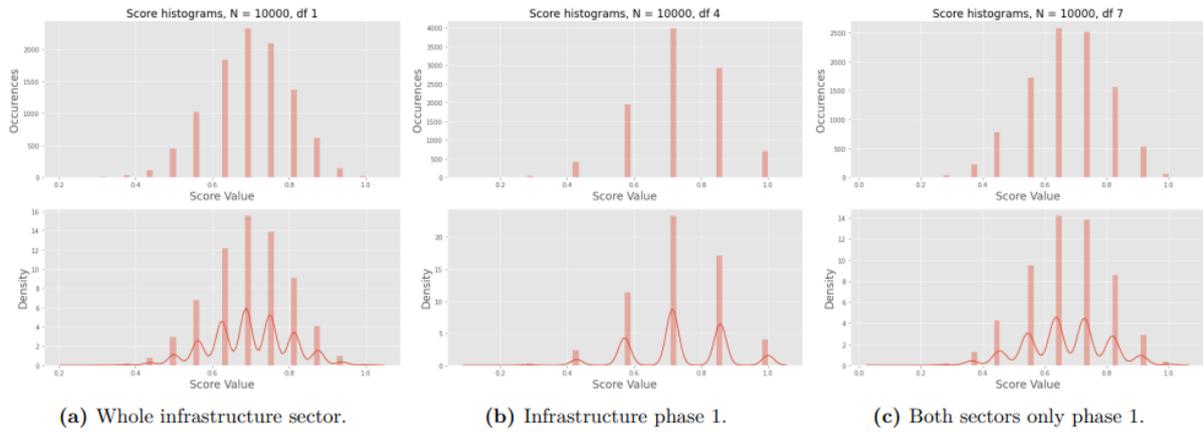
FIG. 12: Count and density plot of DFs 1, 4 and 7.

Figure 13(a)-(c) shows the top five most important features for the three DFs, only collected if the f1-score for the feature was above 80. The count along the x-axis provides an insight into how frequently this occurred during the iterations. For instance, Figure 13(b) shows that both features 'Q146' (planning) and 'Q112' (planning) was in the top five more than 2000 out of 10 000 times. Similarly, in Figure 14, the highest count is that of Figure 14(b). For these, the threshold for features to be appended is a f1-score of 90; this increase in the threshold limit results in a drastic decrease in the count, approximately five times as low as for DF 4.
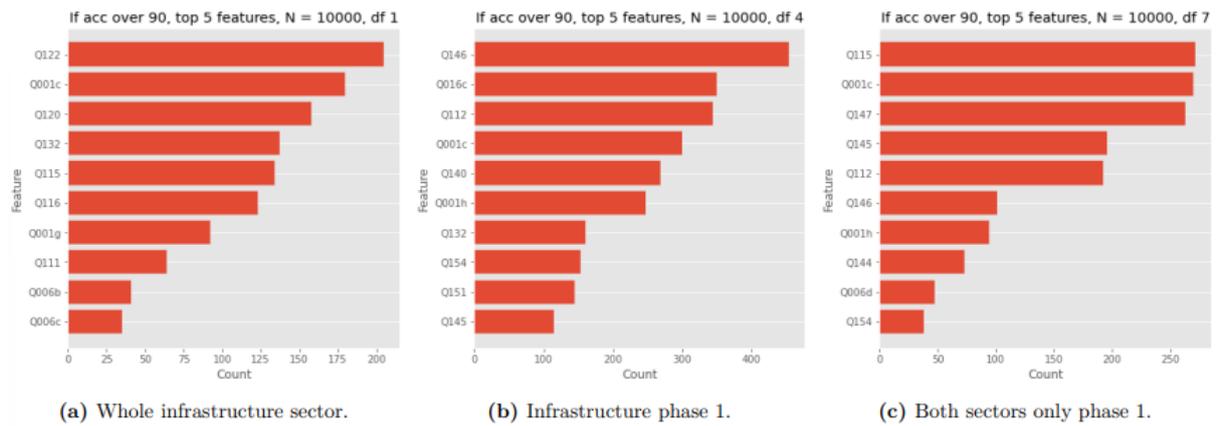

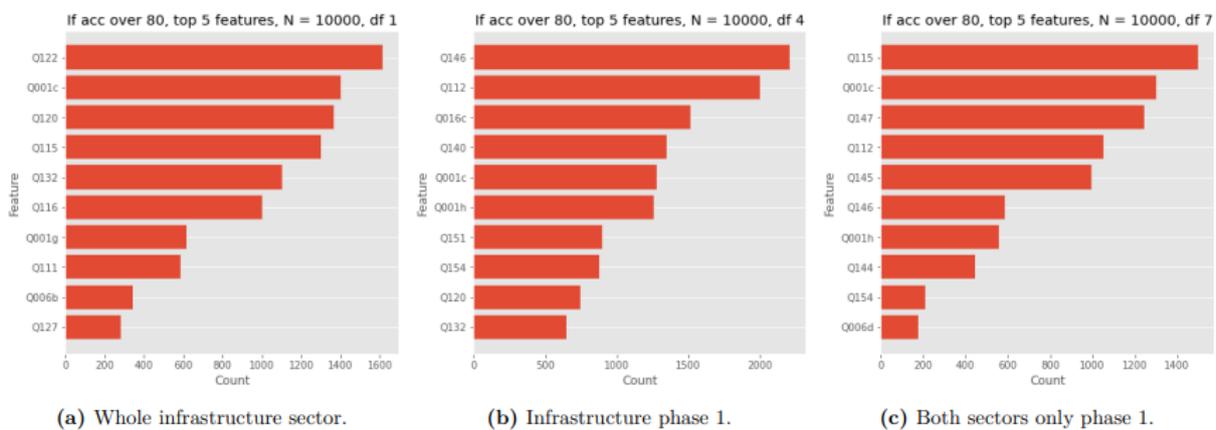*FIG. 13: Top occurring features from MCS, f1-score over 80 for DFs 1, 4 and 7.*

FIG. 14: Top occurring features from MCS, f1-score over 90 for DFs 1, 4 and 7.

For Figure 15, the top 10 features of all nine DFs were aggregated and plotted against the number of times the respective feature appeared in all the DFs. The blue bar indicates the number of times the feature appeared in the top 10, and the red bar how many possible times the same feature could have been chosen. The relationship between the two bars is of importance. For instance, the ratio between the bars of 'Q001c' (complexity) is the same as the ratio of 'Q146' (planning), 'Q017a' (measure progression) and 'Q047' (cost of quality). Therefore, one could argue that the better features are located on the left side of the plot.
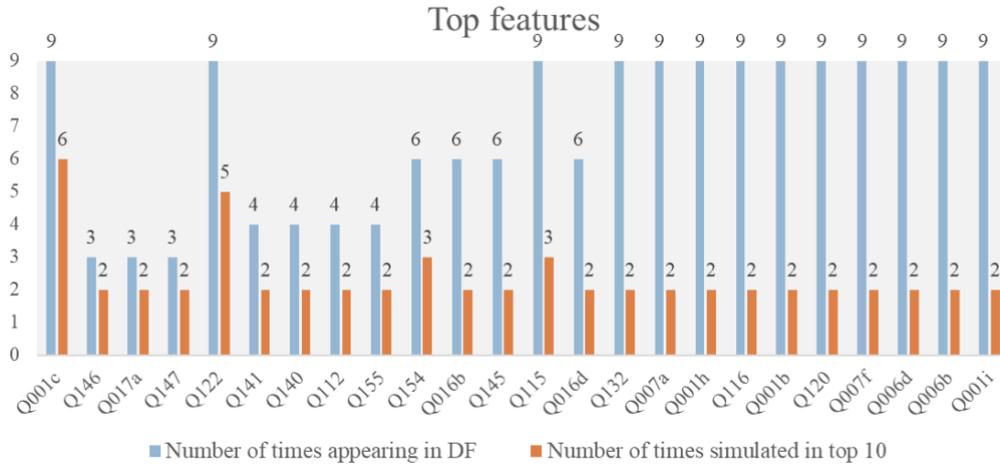

FIG. 15: Top features of top 10, sorted on the ratio between the bars.

Figure 15 illustrates the correlation (Pearson's r score) between the 24 most occurring features, meaning the features that occurred more than twice in the top 10 in all DFs. In this plot, red indicates a strong positive correlation, while blue indicates a strong negative correlation between the two. The dimmer the colour, the closer the absolute value is to 0, meaning no correlation in either direction. A small correlation is defined as an r score between 0.1 and 0.3 in absolute value. Similarly, a medium correlation is defined between 0.3 and 0.5, and a large correlation over 0.5.

Figure 16 illustrates an example of a DT. This specific tree is collected from one of the many trees in the RF when trying to model DF 1. As the dataset is relatively small, the model can only produce a small tree before the gini value becomes 0.
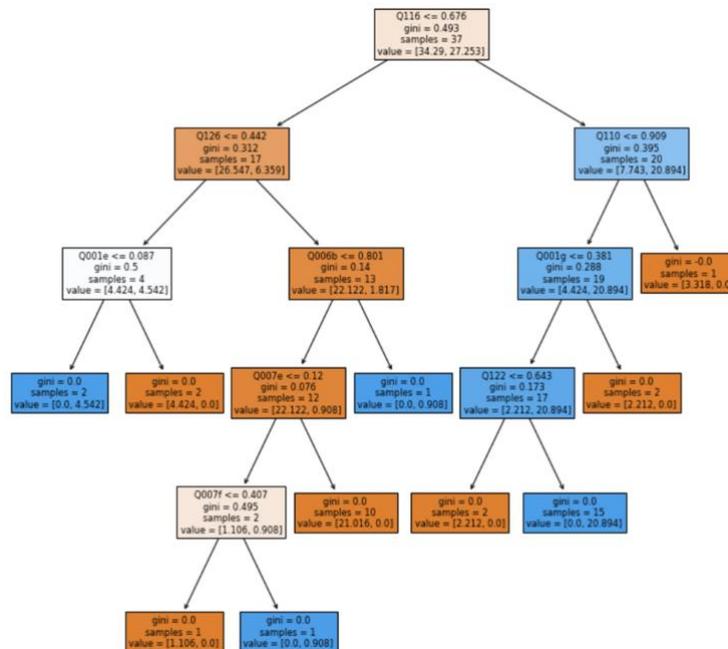

FIG. 16: A single decision tree from the RF for DF 1. Colours indicate which class is in majority of the leaf.

5. DISCUSSION

5.1 ML Model Development

As mentioned, DFs 1, 4 and 7 shows better results in terms of accuracy, representing the infrastructure sector, infrastructure phase 1, and both sectors in phase 1, respectively. Table 2 shows that these DFs are closest to an equilibrium between the number of successes and failures. Worth noting is that infrastructure phase 1 appears in all the top-performing DFs. Therefore, the two other DFs could possibly perform well because they also contain infrastructure 1. However, by inspecting Figure 13, it becomes evident that the most frequently appearing Q-attributes have some differences. For instance, 'Q146' (constructability) is the single most appearing in infrastructure phase 1 but does not appear among the top 10 features in infrastructure as a whole. The same applies in reverse; the most occurring in infrastructure as a whole does not appear in infrastructure phase 1. The top feature for both sectors in phase 1 is 'Q115' (uncertainty analysis); this feature is not among the top 10 in infrastructure phase 1 but appears as the fourth feature for infrastructure as a whole.

Choosing Solution B over Solution A_fillNA may have affected the results. Solution B labelled fewer projects as successful. This could mean that this was a stricter solution. At the same time, this solution labelled 12 projects that would have been discarded by Solution A_fillNA as failures.

The CM of the two solutions A_fillNA and B has been plotted in Figure 17. Comparing this to the CM in Figure 7, it becomes apparent that the two solutions A_fillNA and B share characteristics. Of the 41 gained projects by filling in the NaN, only five are labelled differently. This is found by subtracting the numbers in the off-diagonal, top-right to bottom-left, in the two CMs in question; (11-8) + (4-2) = 5.
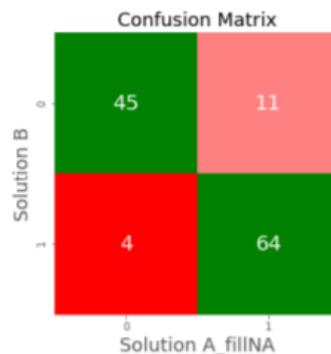

FIG. 17: CM of projects labelled by both solutions A_fillNA and B.

5.2 Identified Success Factors

5.2.1 Most occurring features

Table 3 presents the top features – and by extension, success factors – from DFs 1, 4 and 7. Several features are appearing in two or more DFs. All DFs contain five features of the ten listed. This suggests that certain success factors are of importance both across different project phases and different sectors. For instance, the schedule ('Q001c') leads to high complexity in the engineering phase in both infrastructure and construction, and this appears to be a problem in the infrastructure as a whole. The results in Table 3 illustrates that the top features in DF 7, also appears in DF 1, 4 or in both - the exception is 'Q147' (cost of quality). This could be because the data points in DF 7, as mentioned, also appear in DF 1 and 4. DF 7 could therefore be argued to be a duplicate of the two others. Alternatively, it may indicate that the top features for the engineering phase across all sectors are the same as the top features for engineering in infrastructure, and for infrastructure as a whole.

All features presented in Table 3 seem reasonable in regard to the theory presented in Section 2.2.2. Similarities can be seen in factors addressing involvement from leadership, early planning, structured risk-handling, and implementation of a constructability program. The similarities indicate that it is possible to use ML to obtain the most important success factors, and that the model is performing well.

Multiple listed features relate to the early phases of the project, such as planning, analysis and engineering. This suggests that it could be possible to predict success at an early phase in the project by measuring, reporting, and

assessing these features at early stages. Choosing another definition for project success could have yielded different results – and the inability of an owner or a customer to specify their wants and needs explicitly and correctly poses a potential challenge. Additionally, it could be argued that the project quality is in fact the ex-post value created (Haddadi and Johansen, 2019).

TAB. 3: Top features from the best performing DFs with their conceptual meaning.

Feature	Description	Concept	DF
Q001c	The complexity was remarkably high due to the schedule	Complexity	1,4,7
Q016c	The project had a large number of changes in the list of main components	Changes	4
Q112	The tender plan was developed and communicated to the project team during the engineering phase	Planning	4,7
Q115	All necessary and relevant members of the project team were involved in the process of uncertainty analysis	Uncertainty	1,7
Q120	Involvement from the project owner was appropriate	Leadership involvement	1,4
Q122	The project processes and systems support project success	Project owner's process	1
Q132	Members of the project team participated in adequate professional training directly associated with their engineering work	Training	1
Q146	A rich supply of suggestions for improved constructability was evaluated and integrated during the engineering phase	Planning	4,7
Q147	Cost to fix potential faults were considered during the engineering phase	Cost of quality	7

5.2.2 Correlation matrix

Upon inspection of the bivariate correlation matrix in Figure 18, a few observations can be made. The most important features can be compared with the correlation score r to determine if it is a positive or negative attribute, a '+' or '-' correlation. 'Sol B' in this plot is an abbreviation of 'binary_success_score_2oo3_B'. Feature 'Sol B' is seen to have two medium correlations, with the remaining classified as small correlations, if numbers are rounded down (Kent State University, 2021). The features 'Q001c', 'Q016c' and 'Q016e' reflect complexity and uncertainty and are negatively correlated to the 'Sol B' label. This seems reasonable, as a high value of one of these features, like 1, usually means that the 'Sol B' is low, like 0, and therefore classified as a failure. Similarly, features 'Q112', 'Q132', 'Q146' and 'Q147' reflect adequate early analysis and processes and show a positive correlation with the label feature. The same holds true for 'Q120', reflecting leadership involvement, and 'Q122', relating to the extent to which the work processes in the project supports project success.

Figure 18 further illustrates how 'O_01', known as the cost growth, is slightly positively correlated with 'Q001c' (complexity) and 'Q016c' (changes) respectively. Upon investigating 'Q115' (uncertainty) and 'Q120' (leadership involvement) there is a medium-large correlation with the cost growth and customer satisfaction score, 'Q149'. Both are negatively correlated with the cost, which deducts that inclusion of key personnel and project owner aided the project to keep its budget. Furthermore, both features are positively correlated with the customer satisfaction score, suggesting that the customer was happier with the result if these inclusions were present.

Similarly, both 'Q122' (project owner success) and 'Q132' (training) are correlated positively to both cost and customer satisfaction. 'Q122' (project owner success) only has a correlation score of 0.17 with 'Sol B'. This could indicate that the extra cost this causes, deducted from the positive cost growth, does not do as much for the overall project success as defined for the framework of this study. However, it becomes apparent that this affects the customer satisfaction score, with a correlation score of 0.47. The same argument can be made for the 'Q132' feature, which relates to the training of the project team before the engineering phase.
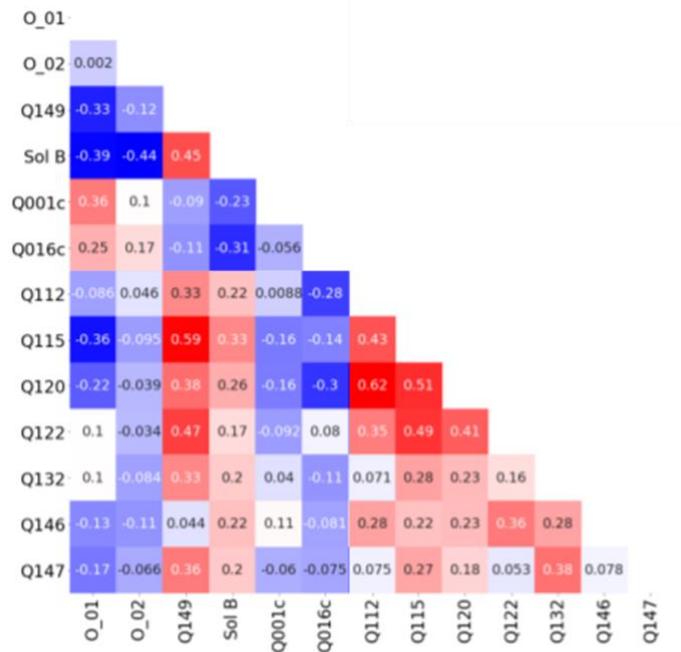
FIG. 18: Correlation matrix of top three features from DFs 1, 4 and 7.

5.2.3 Features between phases

As there is some overlap in the DFs, it is interesting to compare infrastructure as a whole with the single, separate phases of 1 and 3, which infrastructure as a whole is based upon. The correlation of top features between the infrastructure DFs is presented in Table 4.

TAB. 4: Correlation of top features between DFs.

Feature	Infrastructure	Phase 1	Phase 3	Concept
Q001	x	x	x	Complexity
Q122	x	-	x	Project owner's process
Q132	x	x	-	Training
Q115	x	-	-	Uncertainty
Q120	x	x	-	Leadership involvement
Q016	-	x	x	Changes
Q154	-	x	x	Deliver on time
Q146	-	x	-	Planning
Q112	-	x	-	Planning

Some similarities are seen between phases 1 and 3. 'Q154', 'Q001' and 'Q016' are appearing in both top features; however, the dataset shows that they contain different aspects of 'Q001' and 'Q016'. In phase 1, the high complexity is due to the progression plan and diversity in the project team. In phase 3, the complexity is mainly linked to the ability of the supplier to deliver on time.

Five of the top 10 features in phase 3 mainly relate to three features but concern different aspects, namely 'Q016', 'Q001' and 'Q014'. Of these, 'Q016' (numerous deviation reports) occurs over 1000 more times than the runner up. This implies that the quantity of deviation reports is indubitably more important than other features during the building phase of infrastructure projects. Considering that phase 1 represents engineering, it seems reasonable that engineering to a bigger extent is more associated with complexity due to schedule, team

diversity, and changes of main components. As phase 3 represents the building phase, it seems reasonable that it is associated with complexity due to the ability of the supplier to deliver on time, along with numerous deviation reports.

As illustrated in Table 4, the top 10 features are mostly reflected in phases 1 and 3, but some are solely in the infrastructure as a whole. These features include 'Q115' (uncertainty), 'Q116' (changes), 'Q001g' (complex scope), 'Q111' (trust), 'Q006b' (effective meetings), and 'Q127' (team aware of goals). In short, these regard the uncertainty analysis, trust and respect across the team, and an adequate flow of information in the project. Even though these only appear in the infrastructure as a whole, they are more conceptual in nature, which is reasonable when analysing multiple project phases. Keeping in mind that 60 of 77 projects in infrastructure are from phases 1 and 3, one could expect more of the same features. However, four of the top five features in infrastructure as a whole also appear in phases 1 and 3. The three most occurring features in infrastructure as a whole are 'Q122', 'Q001c' and 'Q120', representing processes that support project success, complexity due to the schedule, and involvement of the project owner.

Another observation that can be made is that certain top features in phases 1 and 3 do not appear in infrastructure as a whole. As explained, this could be due to the fact that infrastructure as a whole to a bigger degree contains features that are wider in scope.

5.2.4 Theoretical features compared to model findings

The 10-10 dataset contains more than 100 questions touching on the many aspects of project management. Based on the literature addressing project success and success factors, certain questions and features were expected to be among the factors identified as the most important for project success. Ultimately, some of these did not appear as success factors in any results, including:

- Q013a-c: Did the main goal of the project change during engineering/procurement/construction?

- Q103: The project team was aware of the project goals, requirements, and project owner expectations.

- Q105: Communication with key personnel was handled in a satisfactory manner.

- Q111: There was a high degree of trust, respect, and transparency between the actors in the project.

- Q113: The execution plan supports the goal of the project.

- Q114: Key members of the team understood the owner's goal and scope of this project.

- Q126: The leadership communicated strategic goals, project goals in an effective manner.

- Q139: Key personnel were identified and adequately included in an early phase.

Among the listed features, only 'Q013', 'Q105', 'Q111', 'Q114' and 'Q126' had a sufficiently low percentage of missing datapoints to be used in this analysis. Although these success factors are not emphasised by the model, they appear to be important for success in the sample projects. One possible explanation for this is that the concepts they represent are reflected in other, appearing features. For instance, 'Q113', 'Q114' and 'Q122' (processes support success) all relate to project success, but only 'Q122' appears as an important success factor. The same holds true for 'Q013', 'Q105', 'Q111' and 'Q126', as they can be related – some more strongly than others – to the most important features. This means that the low occurrence of certain features not necessarily implies that the features are of smaller importance, but that they are reflected in other features that are seen to occur more frequently.

5.3 Construction Project Datasets

In construction projects, dimensions such as time, cost, quality, scope, benefits, and risk are all indicators of primary importance for classifying and quantifying project success. Construction projects data can be of high resolution and domain specific, such as plans for large projects. This study is based on what can be described as low-resolution data, as they are based on qualitative evaluations done by the project organisations themselves. This has advantages; the data describe what the projects experienced themselves, for instance. Disadvantages include a risk of bias by the staff reporting the scores. However, we believe that the 10-10 data are interesting. Future analyses would benefit from more consistent registrations of the questions and parameters, a common issue in machine learning and other quantitative analyses.

A model or approximation will only ever be as reliable as the data it is based upon. Currently, no standards exist for collection and utilisation of data in construction projects. To a certain extent, this is understandable, because all projects are unique. However, it would greatly benefit this type of analysis if some standardisation of data structures would emerge. Some industry-specific standards exist for structuring of data, such as for Building information Models (BIM) and standards for data coding such as NORSOK in the Norwegian oil and gas industry.

Data that can be consistently compared and tracked between projects has the potential to improve project-based benchmarking, support project success prediction, and perhaps most importantly, serve as early warning systems that can identify potential issues in time when it still possible to do something about it.

6. LIMITATIONS AND FUTURE RESEARCH

The 10-10 data is based on reports from members of the project team in respective projects. This means that there is a possibility that some of the data points are biased or imprecise; consequently, a value can have been put in the wrong place or provide an inaccurate or biased image of the actual situation.

6.1 Handling Missing Values

When developing the model in this study, several solutions were tested. The model did not implement a function to remove dirty projects within a selected phase; in mixed phase DFs, this would not have worked, while it could have in single phase DFs. The idea is that a single-phase DF, in theory, should include all the same features. This means that no missing values unless all the projects in the DF are missing the same values. With the chosen method, if one project was missing a value in a column the entire column would be discarded.

Analyses showed that if the DF has some missing values where it should not be, it is often one or two projects that are the cause of this. One method to keep more information in the DFs could be to fill missing values in the cleaning; this was deemed undesirable, as it would mean to temper with the available data, inserting values that could be wrong, and ultimately yield imprecise results. A complete DF is always preferred.

Alternatively, a method to keep more information in the DFs could be to discard the projects with missing data, instead of discarding entire columns with missing data.

An alternative sensitivity analysis was performed on DF 4 (infrastructure phase 1), by using a model that discarded the polluted projects. One project in particular had several missing data points. Originally, DF 4 had 32 projects, and 79 non-NaN columns after cleaning and discarding. By discarding the project in question, 31 projects and 109 non-NaN columns remained, leaving 30 more columns for the model to analyse. One project constitutes 3% of the DF, meaning one contaminated project would contaminate the entire DF. As illustrated in Figure 19, only one out of these 30 features show up among the most important features. This feature is 'Q002b', concerning the classification of the level of difficulty of the project. Also worth noting, is that none of the additional features regarding BIM, 'Q031' (BIM used), 'Q032' (who used BIM), and 'Q033' (reason BIM was used), is among the most important features; this is shown in Figure 19(a).

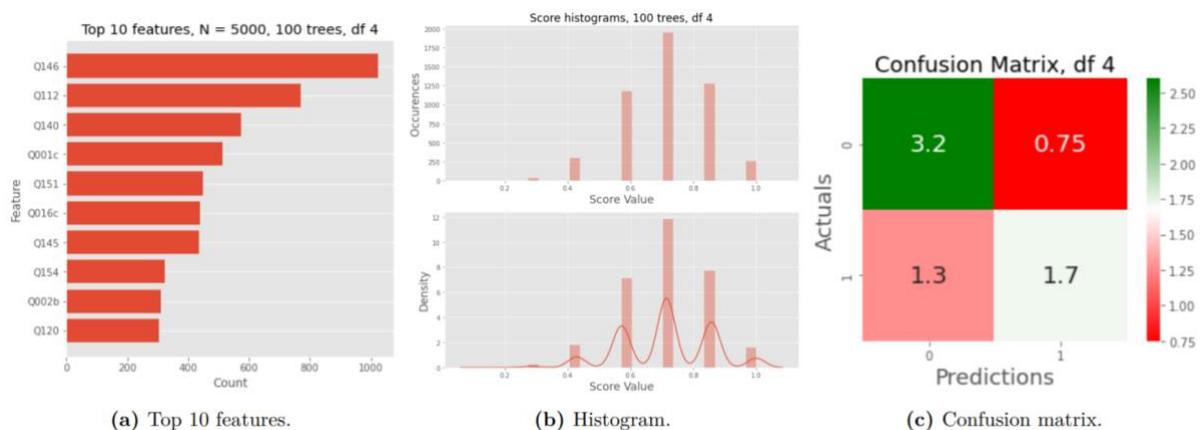(a) Top 10 features.   (b) Histogram.   (c) Confusion matrix.

FIG. 19: After discarding the polluted project. N = 5000, 100 estimators/trees for DF 4.

The same DF was tested with a higher value for the number of estimators, meaning DTs in the RF. However, no correlation between higher estimator count and f1-score was found. The result is illustrated in Figure 19 and can be compared with the CM in Figures 11 and 19.
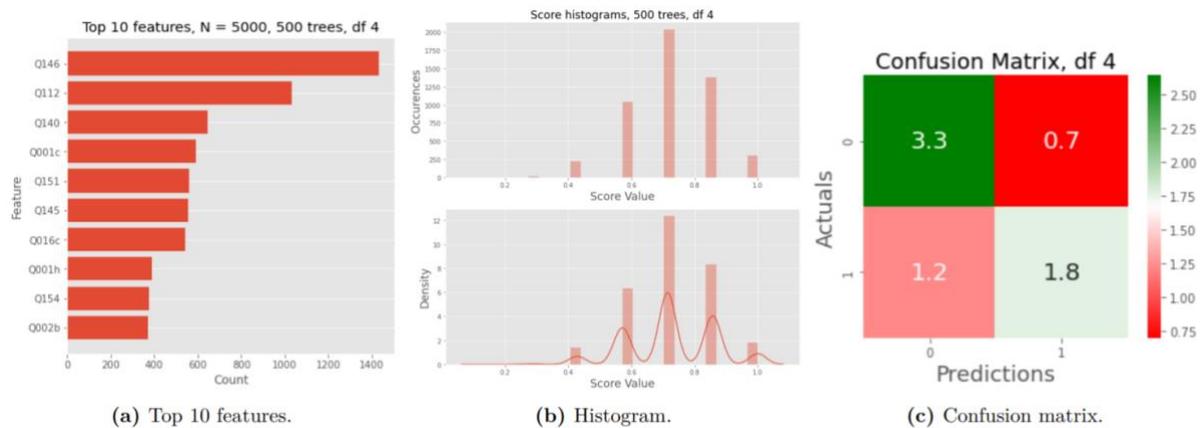

(a) Top 10 features.   (b) Histogram.   (c) Confusion matrix.

FIG. 20: After discarding the polluted project. N = 5000, 500 estimators/trees for DF 4.

6.2 Weighting of DFs

The weighting of DFs posed a challenge in constructing the model. Using Python built-in parameters and functions, such as stratify, class_weight, and sample_weight, the model became more equipped to handle the troubles of a small, unbalanced DF. Alternatively, the parameters could have been weighted manually and individually; this could have yielded a different result. Certain DFs could potentially have performed differently with a different split than the 80/20 split chosen in this model.

6.3 Tuning of Hyper Parameters

This study was intended as a pilot analysis of the Nordic 10-10 dataset, limiting the allocated time and scope for the development of the model. The tuning of DFs was done by searching for global best parameters; another potential approach for future studies would be to analyse one DF at a time and subsequently tune hyper parameters through RandomSearchCV or GridSearchCV functions in SK Learn. Recommended hyper parameters for further analysis and assessment are ccp_alpha, class_weight and sample_weight. This model only utilised the built-in 'balanced' arguments in the two latter.

Furthermore, for a corresponding model, the test size for each DF can be explored further, along with the different paths in the cleaning procedure. Another potential lies within the assessment and comparison of the performance of different ML algorithms on the same dataset.

6.4 Classes

For this model, two classes were defined: success or failure. Further work could look into the possibility of using additional categories, for instance success, failure, and outlier failure. An outlier failure category could provide interesting insights into the identification of the most important features for these projects. Alternatively, classes such as success outliers, neutral projects and failure outliers could be defined. As previously discussed, choosing a non-binary approach would heighten the importance of an unambiguous definition of project success.

Due to the small size of the dataset, manual inspection of the individual projects, specifically outliers, could be yet another option - such inspections could provide unique insights, and prove valuable for further categorisation and classification.

7. CONCLUSIONS

The first research question, regarding how AI, and ML specifically, can be applied to analyse limited datasets from project evaluation has been answered through the description and demonstration of the developed model. However, as the results indicate, only a few DFs display high enough accuracy to facilitate a constructive discussion of the identified features. This indicates that the dataset may have been too limited to provide high-

quality statements. Results provided by the DFs displaying high accuracy suggests that the proposed method is indeed useful for limited datasets.

The second research question was answered through the demonstration of the developed model. The model presented top features for each sector, and for each phase in the two sectors. Among the DFs displaying the highest accuracy, the top features identified align with established success factors in project management theory. Ten features appear more frequently than the others. These features relate to complexity, number of design changes, adequate training and knowledge in the project team, early planning including uncertainty analyses, involvement from top management, and whether or not the processes in the project are perceived to support project success. At the same time, success factors highlighted in literature did not appear as significant in this analysis, and the reasons for this have been discussed.

Ultimately, the ML model demonstrates the ability to discover important factors for project success. Such analyses can be used in early phases of a project to predict project success in later phases, or in the project as a whole, and could prove to be a useful tool in order to eventually achieve more project success.

REFERENCES

Agarwal, R., Chandrasekaran, S. and Sridhar, M. (2016) Imagining construction's digital future. Accessed from https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/imagining-constructions-digital-future on 14th of August 2021.

Akinade, O. O. (2017) *BIM-based software for construction waste analytics using artificial intelligence hybrid models*. Doctoral dissertation. University of the West of England.

Assaad, R., El-Adaway, I. H. and Abotaleb, I. S. (2020) Predicting Project Performance in the Construction Industry. Journal of Construction Engineering and Management. 146(5). Doi: 10.1061/(ASCE)CO.1943-7862.0001797.

Assaf, S. and Al-Hejji, S. (2005) Causes of delay in large construction projects. International Journal of Project Management. 24(4): 349-357. Doi: 10.1016/j.ijproman.2005.11.010.

Auth, G., Jokisch, O. and Dürk, C. (2019) Revisiting automated project management in the digital age - a survey of AI approaches. Online Journal of Applied Knowledge Management. 7(1). Doi: 10.36965/OJAKM.2019.7(1)27-39.

Brekkhus, A. (2017) Den fulldigitale byggenæringen kommer stadig nærmere. Accessed from http://www.bygg.no/article/1323049 on 13th of October 2020.

Campbell, R. and Terwilliger, R. (1986) The SAGA Approach to Automated Project Management. *Computer Science*. 244. Doi: 10.1007/3-540-17189-4_95.

Chua, D., Loh, P., Kog, Y. and Jaelskis, E. (1997) Neural Networks for Construction Project Success. Expert Systems With Applications. 13(4): 317-328. Doi: 10.1016/S0957-4174(97)00046-8.

Cubric, M. (2020) Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. Technology in Society. 62. Doi: 10.1016/j.techsoc.2020.101257.

de Wit, A. (1988). Measurement of project success. Project Management Journal. 6(3). 164–170. Doi: 10.1016/0263-7863(88)90043-9.

Deloitte AI Institute (2020) Deloitte Survey: State of AI in the Enterprise, Third Edition: Thriving in the Era of Pervasive AI. Accessed from https://www2.deloitte.com/us/en/pages/about-deloitte/articles/press-releases/deloitte-survey-state-of-ai-in-the-enterprise-third-edition.html on 15th of September 2020.

dos Santos, I., Barriga, G., Jugend, D. and Cauchick-Miguel, P. (2019) Organizational factors influencing project success: an assessment in the automotive industry. Production. 29(5). Doi: 10.1590/0103-6513.20180108.

Dvir, D., Ben-David, A., Sadeh, A., and Shenhar, A. J. (2006) Critical managerial factors affecting defense projects success: A comparison between neural network and regression analysis. *Engineering Applications of Artificial Intelligence*. 19 (5): 535-543. Doi: 10.1016/j.engappai.2005.12.002

GHD (2020) Using instant intelligent insights to revolutionise major infrastructure construction. Accessed from https://www.ghd.com/en/projects/using-instant-intelligent-insights-to-revolutionise-major-infrastructure-construction.aspx on 27th of October 2020.

Ika, L. (2009) Project Success as a Topic in Project Management Journals. Project Management Journal. 40(4). 6–19. Doi: 10.1002/pmj.20137.

Haddadi, A. and Johansen, A. (2019) Value Proposition in Different Types of Buildings- Characteristics and Means. The Journal of Modern Project Management. 6(3). Doi: 10.19255/JMPM01805.

Hussein, B. (2016) Veien til suksess. Fagbokforlaget: Bergen.

Hwang, B. and Ng, W. (2012) Project management knowledge and skills for green construction: Overcoming Challenges. International Journal of Project Management. 31(2): 272-284. Doi: 10.1016/j.ijproman.2012.05.004.

Kent State University (2021) SPSS Tutorials: Pearson Correlation. University Libraries. Accessed from https://libguides.library.kent.edu/SPSS/PearsonCorr on 11th of December 2021.

Kowalski, M., Zelewski, S., Bergenrodt, D. and Klüpfel, H. (2012) Application of new techniques of artificial intelligence in logistics: An ontology-driven case-based reasoning approach. The European Modelling and Simulation Conference. 323-328.

KPMG (2015) Global construction survey 2015: Climbing the curve. KPMG international. Accessed from https://assets.kpmg/content/dam/kpmg/pdf/2015/05/construction-survey-201502.pdf on 23rd of September 2021.

Lee, J., Davari, H., Singh, J. and Pandhare, V. (2018) Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. Manufacturing Letters. 18: 20-23. Doi: 10.1016/j.mfglet.2018.09.002.

Loureiro, S., Guerreiro, J. and Tussyadiah, I. (2020) Artificial intelligence in business: State of the art and future research agenda. Journals of Business Research. 129: 911-926. Doi: 10.1016/j.jbusres.2020.11.001.

Magaña, D. and Fernández Rodríguez, J. C. (2015). Artificial Intelligence Applied to Project Success: A Literature Review. *International Journal of Artificial Intelligence and Interactive Multimedia*. 3: 77-84. Doi: 10.9781/ijimai.2015.3510.

Marr, B. (2020) The key definitions of artificial intelligence (AI) that explain its importance. Accessed from https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/?sh=14dcaf9f4f5d on 9th of October 2020.

McKinsey Global Institute. (2017) Reinventing Construction: a Route to Higher Productivity. Executive Summary. McKinsey & Company.

Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M.S, Upadhyay, R. and Martynenko, A. (2020) IoT, big data and artificial intelligence in agriculture and food industry. IEEE Internet of Things Journal. Doi: 10.1109/JIOT.2020.2998584.

Niu, Y., Lu, W., Chen, K., Huang, G. and C. Anumba (2019) Smart Construction Objects. Journal of Computing in Civil Engineering. 30(4). Doi: 10.1061/(ASCE)CP.1943-5487.0000550.

Nordic 10-10 (2020) *Hvordan bruke CII 10-10*. How to use CII 10-10. Accessed from https://nordic10-10.org/hvordan-bruke-cii-10-10/ on 11th of November 2020.

Oberle, W. (2015) Monte Carlo Simulations: Number of Iterations and Accuracy. Technical note. US Army Research Laboratory.

Oliver Wyman (2018) Digitalization of the Construction Industry: the Revolution is Underway. Marsh & McLennan Companies. Oliver Wyman.

Project Management Institute (2017) A guide to the Project Management Body of Knowledge. PMBOK® GUIDE. Sixth Edition. Project Management Institute.

Project Management Institute (2018) Success in disruptive time: Expanding the value delivery landscape to address the high cost of law performance. Accessed from https://www.pmi.org/-

/media/pmi/documents/public/pdf/learning/thought-leadership/pulse/pulse-of-the-profession-2018.pdf on 14th of August 2021.

Prosjekt Norge (2017) Bli med på å fremme benchmarking i prosjektledelse som gulloppskrift for BAE-bransjen. Contribute to promotion of benchmarking in project management as the golden standard for the AEC-industry. Accessed from https://www.bnl.no/siteassets/dokumenter/rapporter/invitasjon_prosjektnorge_screen.pdf on 11th of November 2020.

Radujkovic, M. and Sjekavica, M. (2017) Project Management Success Factors. Procedia Engineering. 196: 607–615. Creative Construction Conference 2017. Doi: 10.1016/j.proeng.2017.08.048.

Rahmanifard, H. and Plaksina, T. (2019) Application of artificial intelligence techniques in the petroleum industry: a review. Artificial Intelligence Review. 52. 2295-2318. Doi: 10.1007/s10462-018-9612-8.

Ramos, M., Tacla, C., Sato, G., Paraiso, E. and Barthès, J. (2010) Dialog Construction in a Collaborative Project Management Environment. International Conference on Computer Supported Cooperative Work in Design. Doi: 10.1109/CSCWD.2010.5471963.

Russell, S. J. and Norvig, P. (2003) Artificial Intelligence: A Modern Approach. Second Edition. Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2.

Samset, K. and Volden, G. (2016) Front-end definitions of projects: Ten paradoxes and some reflections regarding project management and project governance. International Journal of Project Management. 34(2): 297–313. Doi: 10.1016/j.ijproman.2015.01.014.

Schwarz, J. (2015) Implementation of artificial intelligence into risk management decision-making processes in construction projects. Universität der Bundeswehr München. Institut für Baubetrieb.

ScienceDaily (2020) Artificial intelligence. Accessed from https://www.sciencedaily.com/terms/artificial_intelligence.htm on 9th of October 2020.

Sozontov, A., Ivanova, M. and Gibadullin, A. (2019) Implementation of artificial intelligence in the electric power industry. International Conference of Young Scientists Energy Systems Research 2019. 114. Doi: 10.1051/e3sconf/201911401009.

The Economist (2020) An understanding of AI's limitations is starting to sink in. Accessed from https://www.economist.com/technology-quarterly/2020/06/11/an-understanding-of-ais-limitations-is-starting-to-sink-in on 9th of October 2020.

Todsen, S. (2018) Produktivitetsfall i bygg og anlegg. Statistisk Sentralbyrå.

Vikan, J. (2018) Digitalisering gir store muligheter - og store utfordringer. Accessed from http://www.bygg.no/article/1365237 on 19th of October 2020.

Walch, K. (2020) Is AI overhyped? Accessed from https://www.forbes.com/sites/cognitiveworld/2020/06/04/is-ai-overhyped/?sh=1a5e35fb63ee on 9th of October 2020.

Wang, Y., Yu, C. and Chan, H. (2011) Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models. International Journal of Project Management. 30(4): 470–478. Doi: 10.1016/j.ijproman.2011.09.002.

Witten, I. H. (2016) Data Mining: Practical Machine Learning Tools and Techniques. Elsevier Science & Technology.

Worldsensing (2020) Smart construction powered by onemind. Accessed from https://www.worldsensing.com/one-mind-smart-construction/ on 27th of October 2020.