Machine Learning with Applications 6 (2021) 100166

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Applied artificial intelligence for predicting construction projects delay

Christian Nnaemeka Egwim [a], Hafiz Alaka [a,*], Luqman Olalekan Toriola-Coker [b], Habeeb Balogun [a], Funlade Sunmola [c]

[a] *Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, United Kingdom*
[b] *School of Built Environment, University of Salford, Manchester, M5 4WT, United Kingdom*
[c] *School of Physics, Engineering & Computer Science, University of Hertfordshire, Hatfield, AL10 9AB, United Kingdom*

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Ensemble of ensembles
Ensemble learning
Project delay
Predictive analytics

ABSTRACT

This study presents evidence of a developed ensemble of ensembles predictive model for delay prediction – a global phenomenon that has continued to strangle the construction sector despite considerable mitigation efforts. At first, a review of the existing body of knowledge on influencing factors of construction project delay was used to survey experts to approach its quantitative data collection. Secondly, data cleaning, feature selection, and engineering, hyperparameter optimization, and algorithm evaluation were carried out using the quantitative data to train ensemble machine learning algorithms (EMLA) – bagging, boosting, and naïve bayes, which in turn was used to develop hyperparameter optimized predictive models: Decision Tree, Random Forest, Bagging, Extremely Randomized Trees, Adaptive Boosting (CART), Gradient Boosting Machine, Extreme Gradient Boosting, Bernoulli Naive Bayes, Multinomial Naive Bayes, and Gaussian Naive Bayes. Finally, a multilayer high performant ensemble of ensembles (stacking) predictive model was developed to maximize the overall performance of the EMLA combined. Results from the evaluation metrics: accuracy score, confusion matrix, precision, recall, f1 score, and Compute Area Under the Receiver Operating Characteristic Curve (ROC AUC) indeed proved that ensemble algorithms are capable of improving the predictive force relative to the use of a single algorithm in predicting construction projects delay.

1. Introduction

Construction sector is considered a major contributor to the global economy — represents 13% of the global gross domestic product (GDP) with a promising 85% to $15.5 billion globally by the year 2030 with three leading countries – China, the United States and India – contributing 57% of its global demand (Robinson, 2015). Furthermore, Woetzel et al. (2017) estimates global infrastructure spending at $3.4 trillion annually from 2013 to 2030, which is roughly 4% of total GDP. The sector is also considered a major backbone of any country's economy — represents 3% of the total economic output of Nigeria (Egwim et al., 2021), 4.3% of the total economic output of Germany (European Comission, 2017), 6% of the total economic output of United Kingdom (UK) (Rshodes, 2019), 4.1% and 6.8% of the total economic output of the United States of America (USA) and China respectively (Wang, 2018, 2019) etc.

However, despite its importance the construction industry has continued to underperform. According to Egan (2018) the construction industry is under-achieving as evident in its low profitability, capital investment, research and development generally caused by delay of construction projects, resulting in great dissatisfaction from the industry's clients on its overall performance. Some research publications, for instance, Flyvbjerg (2014) and Rhodes (2019) indicated that 9 out of 10 global mega projects encounter delay, which usually results in excess cost overruns. Delay is the main factor in the general completion of every construction project as it raises overflow costs (Haq et al., 2017). Delay is described as an increase in time outside the stakeholder's negotiated timeline of project completion or after a date of the termination of a lawful contract. For the client, delay connote loss of revenue or investments at the end of agreed time, while to the contractor, a delay can imply an increase in overhead cost (Assaf & Al-Hejji, 2006). Also, (Bartholomew, 2001) makes an important point arguing that delay is a deceleration of some part of a construction project without a complete halt.

Investigation by several researchers have shown that delay of construction projects has adverse effect on the reputation of the construction industry's contribution to the global economy. With reference to Abdul-Rahman et al. (2011), the effects of construction delay can be evaluated with respect to its national footprints which with prejudice sway the industry's subsidy to the economy; at an industry level, where

* Corresponding author.
E-mail addresses: c.n.egwim@herts.ac.uk (C.N. Egwim), h.alaka@herts.ac.uk (H. Alaka), o.toriola-coker@edu.salford.ac.uk (L.O. Toriola-Coker), h.balogun@herts.ac.uk (H. Balogun), f.sunmola@herts.ac.uk (F. Sunmola).

https://doi.org/10.1016/j.mlwa.2021.100166
Received 19 February 2021; Received in revised form 28 May 2021; Accepted 13 September 2021
Available online 25 September 2021

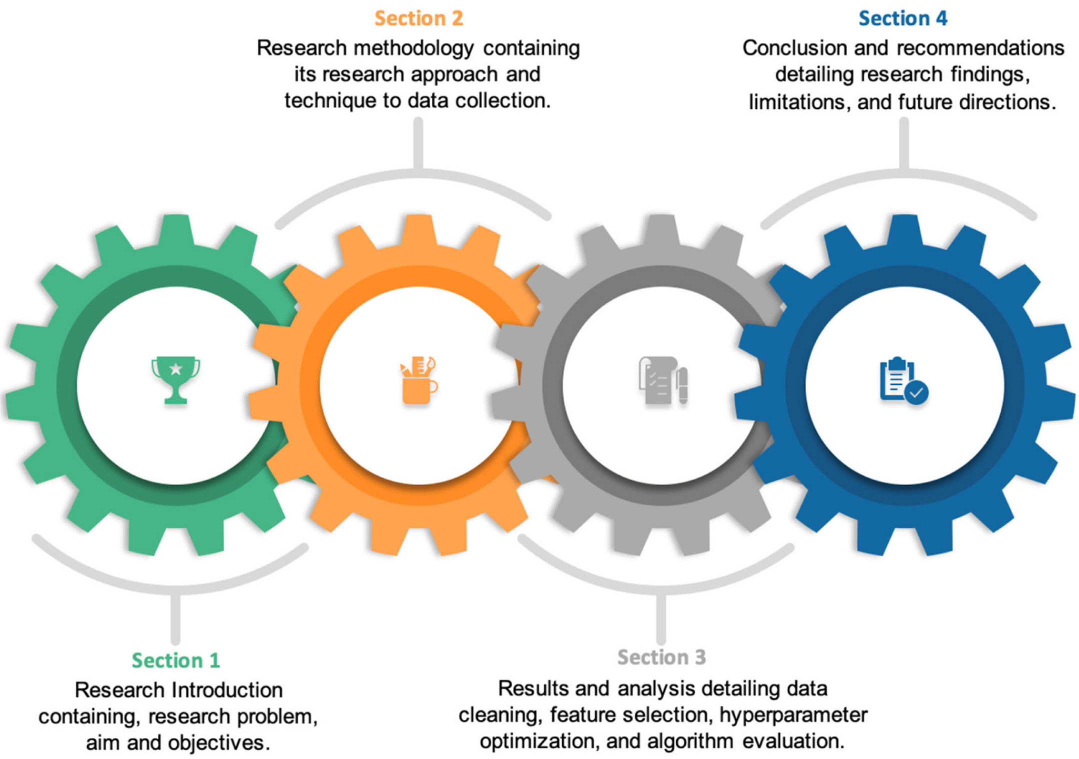

Fig. 1. This study's organogram.

delays impact profitability and productivity negatively; and at a project level where delay foster industry client's dissatisfaction on its overall performance, cessation of contracts by the owner, and unprofitability for contractor(s). Furthermore it has been argued (Kumar, 2016) that delay often lead to project cost overruns, insolvency of organization, loss of opportunity of future projects, dispute among project stakeholders. Major delay factors have been identified by several researchers as evident in vast body of international literature (from amongst the oldest articles to the most recent), e.g. bad weather and jurisdictional/contractual disputes in both United States of America (USA) and UK by Baldwin, Manthei, Rothbart, and Harris (1971) and Sullivan and Harris (1986) respectively; variation orders in both Nigeria and United Arab Emirates (UAE) by Motaleb and Kishk (2010) and Odeyinka and Adebayo (1997) respectively; planning and scheduling deficiencies in Australia, delay in payment certificates in Ghana and poor site management in Malaysia by Shah (2016); ground problems and inefficient structural connections for prefabricated components in both the UK and India by (Agyekum-Mensah & Knight, 2017; Ji et al., 2018) respectively and finally shortage of adequate equipment and poor communication among contracting parties in China by Chen et al. (2019).

Several research approaches and guidelines for mitigating delay of construction projects have been established over the decades. For instance, Sullivan and Harris (1986) suggested more teamwork especially at the early stages of project planning. According to Alaghbari and Sultan (2018), Assaf, Al-Khalil, and Al-Hazmi (1995), Enshassi, Al Najjar, and Kumaraswamy (2009) and Owalabi et al. (2014) clients should adhere to timely payment of progress fee and consider funding levels at the planning stage of project. Furthermore, the survey by Gondia et al. (2020), Yaseen et al. (2020) recommended the use of predictive models to mitigate delay risks and time claim in construction projects. Despite all these delay factors and recommendations towards mitigating delay in construction, delay still strives in the industry, hence the first motivation of this study. Interestingly, only a few studies have taken the advantage of the contemporary analysis method which best explains the factors that can be affecting a phenomenon like delay based on its predictive capabilities. This analysis method is the Artificial Intelligence (AI)/Machine Learning (ML) which has been widely adopted across other industries, but construction industry is slow to adopt (Blanco et al., 2018). The adoption of AI/ML algorithms in construction is relatively evolving, especially when compared to other industries like healthcare: guiding in the choice of treatment; education: virtual lectures; and transportation: autonomous vehicles, as it currently uses lots of methods that were used in the centuries past (Marks, 2017).

As a commonplace the industry produces massive amount of data daily on every project. For example data produced from images captured from smart devices, IoT sensors, Building Information Modelling (BIM is defined as structured model of data that represents building elements with its usage spanning beyond the pre-construction phase to the post-construction phase (Ameziane, 2000)) etc, presents a window of opportunity for the industry and its customers to examine and gain profits from insights generated from past construction data through the aid of AI and ML. AI is defined as a collection of state-of-the-art technologies that permit machines or any computer programme to sense, comprehend, act and learn (Goyal, 2019). ML on the other hand is a branch of AI that allows computers to learn by a direct route from examples, data and experience replacing the traditional approaches to programming that relied on hardcoded step by step rules (Royal Society, 2017). Several ML algorithms such as Genetic Algorithm, Neural Networks, Linear Regression, Logistic Regression, Nearest-Neighbour Mapping, Decision Trees, K-Means Clustering, Random Forests, and Support Vector Machines exist for ML model implementation. Which ML algorithm to use depends on lot of factors, e.g., ease of use, accuracy, training time, etc. Few researchers have attempted the use of AI and ML algorithms in some aspect of construction. Poh, Ubeynarayana, and Goh (2018) used five popular ML algorithms to predict accident occurrence and severity of construction sites in Singapore; Zou and Ergan (2019) Leveraged on three ML techniques to predict the influence of construction projects on urban quality of life; Arditi and Pulket (2005) and Mahfouz and Kandil (2012) used only one and three ML models respectively to forecast end results of construction litigation all in the USA.

Table 1
List of features and target.

Section ID	Factor ID	Factors
A	F1	Project size
	F2	Equipment breakdown/ Management
	F3	Inflation or sudden increase in good/commodities
	F4	Labour dispute or strikes
	F5	Effective or poor communication among stakeholders
	F6	Inclement or bad weather
	F7	Contractor's financial difficulties
	F8	Structural design variations
	F9	Late deliveries of materials/equipments
	F10	Changed orders/ discrepancies in contract documents
	F11	Price fluctuation
	F12	Contract management
	F13	Decision making
	F14	Cash flow during construction
	F15	Government regulations
	F16	Material procurement
	F17	Site conditions
	F18	Political Influence
B	F19	Project schedule/program of work
C	F20	Site accident
	F21	Project quality control
D	F22	Late payment
	F23	Proportion of unskilled labourers
	F24	Late delivery of materials by supplier
E	F25	Project delay

Only a hand full of literature have attempted the adoption of AI or ML to mitigate construction delay. For example, Gondia et al. (2020) used two ML models — Decision Tree and Naïve Bayesian Classifiers (with accuracy value of 74.5% and 78.4% respectively) towards expediting precise project delay risk assessments and forecast in building project in Egypt. Also, Asadi, Alsubaey, and Makatsoris (2015) used two ML approach (with accuracy value of 79.41% and 73.52% for decision tree and Naive Bayes model respectively) to predict delays in construction logistics in Qatar. Furthermore, Yaseen et al. (2020) developed a hybrid artificial intelligence model (a combination of Random Forest and Genetic Algorithm) and achieved an accuracy value of 91.67% for delay problem prediction in Iraq. Evidently, no specific literature to the best of our knowledge at the time of this study have attempted to use Ensemble Machine Learning Algorithms (EMLA) to predict delays of construction projects, hence the final motivation of this study. EMLA utilize a group of algorithms where the cumulative outcome from them is almost always greater in terms of predictive accuracy relative to the use of a single model as it integrate decisions from different algorithms to maximize the overall performance (Badawi et al., 2019; Dietterich, 2000; Hastie, Tibshirani, & Friedman, 2009). Consequently, this study aims to develop a multilayer high performant ensemble of ensembles predictive model using hyperparameter optimized EMLA to predict delay of construction projects. The following objectives will be used to achieve this aim:

1. Carry out literature review towards gathering the most common factors affecting delay of construction projects and use it to conduct survey of experts to establish the most applicable factors affecting delay of construction projects.
2. Utilize established factors in objective 1 as independent variables for EMLA (bagging and boosting) to develop hyperparameter optimized predictive models.
3. Combine the best predictive models from objective 2 to develop a multilayer high performant ensemble of ensembles (stacking) predictive model.

Bagging is an ensemble machine learning technique where multiple models of the same algorithm are used, however with different subsets of data selected randomly (Opitz & Maclin, 1999). Boosting is a repetitive technique that adapts the weight of the observation to the last grading. If an observation has been falsely categorized, the weight of this observation would be raised and conversely (Dietterich, 2000). Naive Bayes — an effective and efficient inductive ensemble methods also referred to as conditional independence, is the most basic type of Bayesian network, in which all characteristics are independent of the class variable's value (Zhang, 2004). Different from the bagging, boosting and naïve bayes ensemble machine learning techniques, stacking often considers heterogeneous week learners by combining the base algorithms using a meta-model rather than some averaging processes (Seni & Elder, 2010). To achieve these objectives this study will proceed to its research methodological approach to data collection and exploration in the next section. Section 3 will follow detailing its results and analysis of how the high performant ensemble of ensembles predictive model was developed. Finally, Section 4 will detail its conclusion and recommendation (see Fig. 1).

Table 2
Reliability statistics.

Cronbach's alpha	Cronbach's alpha based on standardized factors	N of factors
0.938	0.935	24

2. Research methodology

A review of existing literatures on influencing factors of construction projects delay was used to establish the most applicable factors there by fulfilling part of the first objective. Twenty-four applicable factors (see Table 1) were consolidated at the end of the review which was pre-empted as search results became repetitive. These factors were used to design a survey in form of questionnaire to fulfil the remaining part of the first objective.

The questionnaire was divided into five sections such that each section deals with a specific feature of event under investigation (delay factors). Section A asked the responders to rate how eighteen factors affected the duration of the project. Where a project does not have an official schedule/programme of work indicating the duration of the project, they were asked to use an assumed duration that such a project would have taken, or the duration based on an agreed date of completion with the client. Section B enquired to what level of detail one factor had, and Section C asked for frequency of occurrence of two factors, Section D enquired what percentage a responder would give to

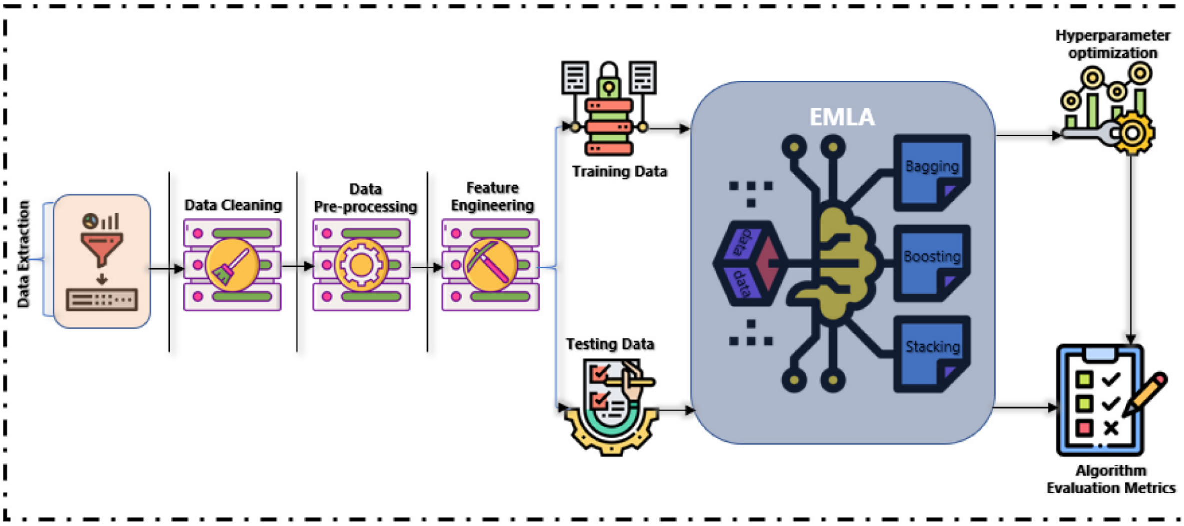

Fig. 2. EMLA Prediction Architecture.

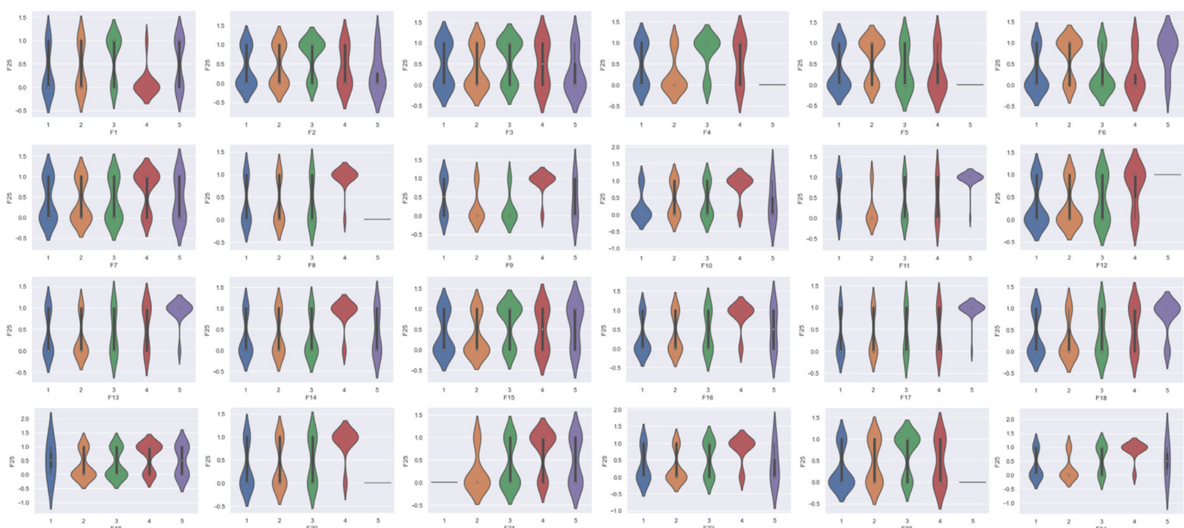

Fig. 3. Shape of Dataset's Distribution.

three factors. All these made a total of twenty-four factors as features (independent variables). Also, the responders were asked to rate how long the entire project delayed for in the final Section E; this represents the target (dependent variable) for the EMLA implementation. In the end, a total of 302 questionnaire were distributed. A total of 120 responses were received and since sampling cannot be done in isolation as there are no special right decision for determining sample size for a research Flick (2014), this number of responses are considered satisfactory (Delice, 2001; Durbarry, 2019; James, Joe, & Chadwick, 2001).

The questions were designed on a Likert scale with a scale of one to five. Although questions in each section were analysed individually, they were also linked together in such a way that their respective answers accumulatively helped to arrive at a finding (delay). The use of questionnaire research signifies independent observation — implies the questionnaire will be completed in the absence of the researcher, and since one of the objectives of this study is set out to establish the true (most) applicable factor to construction delay makes it a positivist research. A positivist researcher is usually independent (of the subject) as an observer, reduces a phenomenon to simpler measurable factors (causes of delay in construction projects deduced from several literature was reduced using Likert scale.), explains the elements with regards to how they affect the phenomenon (cause and effect) and often uses large samples (Easterby-Smith et al., 2008; Morgan, 1980).

Prior to distribution of the questionnaire, pilot testing was conducted by asking group of experts in construction to comment on the representativeness and suitability of the questions. This was done to ensure thorough understanding of the questions by the responders and to avoid errors when recording data, to assess questions validity and the likely reliability of data to be collected (Saunders et al., 2009, p.425). The responders of the questionnaire were experienced stakeholders from the construction industries in Nigeria. They were instructed to have in mind any project of choice they have worked on in the past while answering the questions. Since this study aims to develop a multilayer high performant ensemble of ensembles predictive model using hyperparameter optimized EMLA to predict delay of construction projects makes it a deductive research which further reinforces its positivism. Convenient sampling method was selected due to its ease of accessibility, geographical proximity and affordability which satisfied this research. Convenient sampling (also called haphazard/accidental sampling) is a typical nonprobability/non-random sampling where a researcher considers the most convenient object(s) and time, effort and money for conducting data collection (Matthews, Ross, & Ellison, 2010).

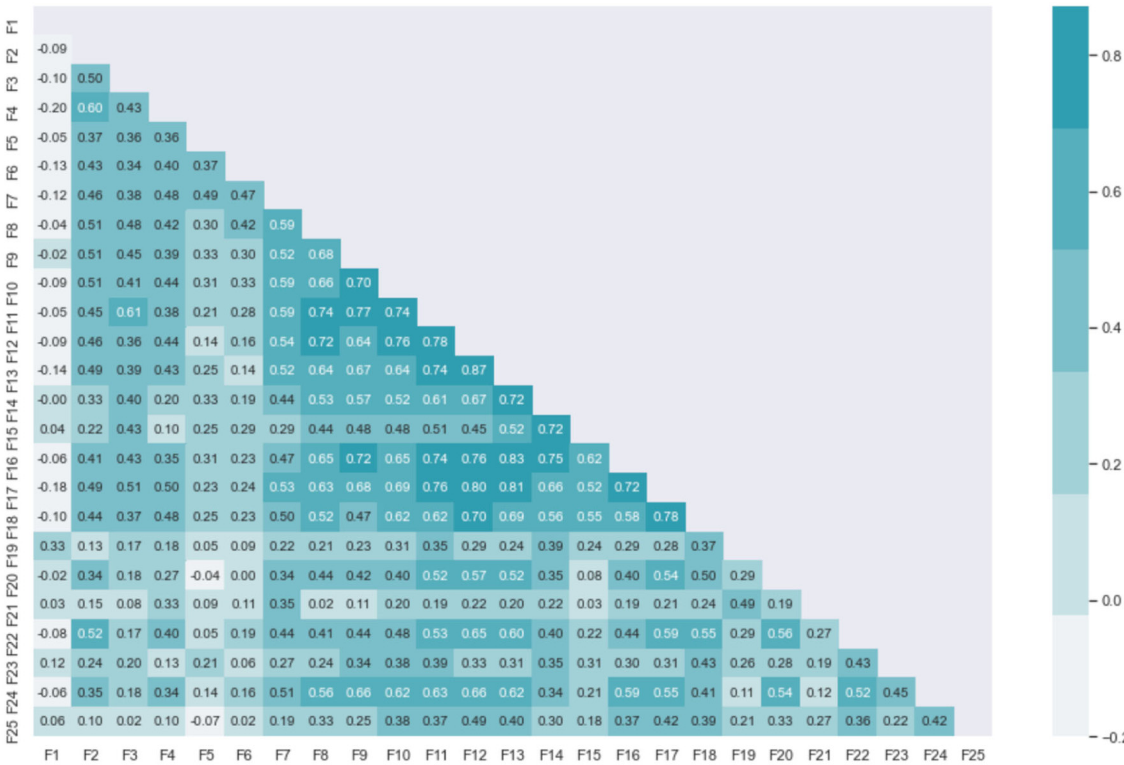

Fig. 4. Correlation Matrix Plot.

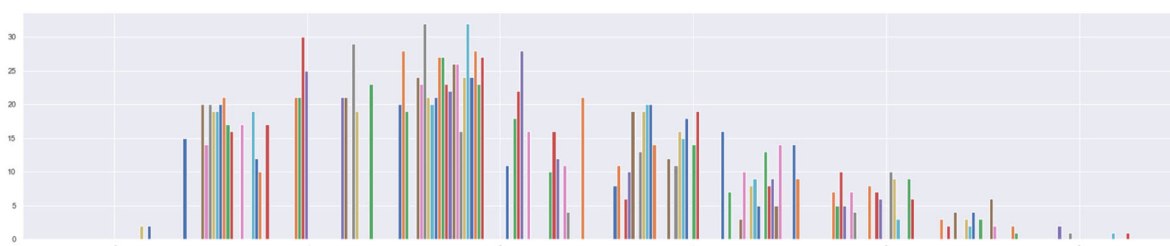

Fig. 5. Standardized Dataset Distribution.

The collected data received via Google forms were extracted and converted into a comma-separated values file. To achieve the second objectives, this raw data was pre-processed into a clean and an analysable dataset by carrying out data imputation and outlier detection. Scaling and encoding feature engineering techniques were employed to enable the selection of features or input variables to increase the predictive power (hyperparameter optimization) of the EMLA. The resulting clean, pre-processed and feature engineered dataset was split randomly into two in a ratio of 60% to 40% of training dataset and testing dataset respectively. EMLA were imported into a running instance of Jupiter Notebook using Scikit-learn — an integral Python programming language module with a broad spectrum of state-of-the-art algorithms for supervised and unsupervised medium-scale problems (Pedregosa et al., 2011). Since EMLA fit input variables (delay factors) to a known output variable (delay) supervised modelling taxonomy was undoubtedly chosen. The training dataset (60% of total dataset) was used to fit different EMLA while their knobs were optimized during successive runs to further improve the performance for making predictions on unseen test dataset (40% of total dataset). The resulting best performing EMLA selected via the hard and soft voting rule were used as new input variables which produced a multilayer high performant ensemble of ensembles algorithm (to achieve the third objective). Finally, Accuracy, Confusion Matrix, Precision, Recall, F_1-Score and ROC curve modelling evaluation metrics were employed to measure the new model and EMLA performance on the testing dataset as shown in Fig. 2

3. Results and analysis

3.1. Reliability analysis of survey outcomes

A reliability analysis from the Alpha Test of Cronbach was carried out to test the reliability of the respondents ' answers for all 24 factors. Alpha of Cronbach (α) can be written as:

$$\alpha = \frac{N.c}{\upsilon + (N-1).c}$$

where, N is the number of factors, c the average covariance between factor-pairs and υ the average variance. The main purpose of α was to assess how accurate the data obtained from the survey were, by evaluating the internal consistency coefficient of data. In addition, it was important to decide whether the combined factors help to calculate the same construct (delay).

While there is no lower bound, the higher the alpha coefficient of Cronbach is to 1 , the greater the internal accuracy of the factors (Gliem & Gliem, 2003). An α of 0.7 or higher is known to be symptomatic of strong inner harmony of the factors in determining the reliability

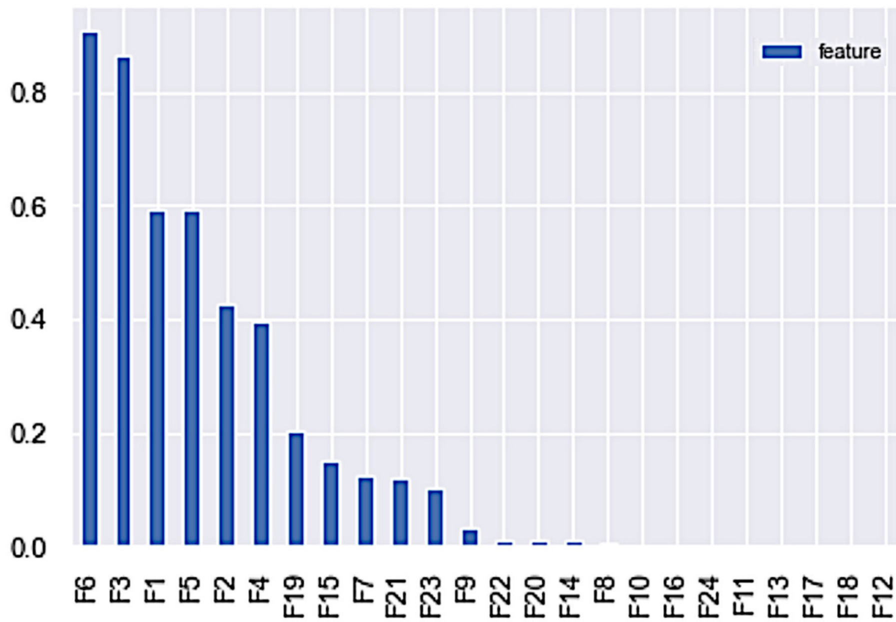

Fig. 6. Chi-squared Test.

Table 3
Algorithms and libraries.

S/N	Algorithms	Scikit-learn Library
1	Decision Tree (DT)	DecisionTreeClassifier
2	Random Forest (RF) (Ensemble of DT)	RandomForestClassifier
3	Bagging Ensemble	BaggingClassifier
4	Extra-Trees (Extremely Randomized Trees)	ExtraTreesClassifier
5	AdaBoost	AdaBoostClassifier
6	Gradient Boosting / Gradient Boosting Machine (GBM)	GradientBoostingClassifier
7	XGBoost (Extreme Gradient Boosting)	XGBClassifier
8	Bernoulli Naive Bayes (BNB)	BernoulliNB
9	Multinomial Naive Bayes (NNB)	MultinomialNB
10	Gaussian Naive Bayes (GNB)	GaussianNB
11	Ensemble of Ensembles	Mlens

Table 4
Confusion matrix.

		Prediction	
		Negative (delay < threshold limit) = 0	Positive (delay > threshold limit) = 1
Actual	Negative (delay < threshold limit) = 0	*True negative*	*False positive*
	Positive (delay > threshold limit) = 1	*False negative*	*True positive*

of the construct (Bhatnagar, Kim, & E. Many, 2014). However, It is the viewpoint of Nunnally (1978) that the α should surpass 0.8 for fundamental science to consider accurate responses to a factor. The findings of this study on the 24 factors in this analysis show strong inner stability α of 0.938 as shown in Table 2.

3.2. Data pre-processing

An initial investigation on the data through Exploratory Data Analysis (EDA) showed that the data is a two-dimensional array with 120 rows and 25 columns where the 1st to the 24th columns (F1–F24 factor IDs) represent the features/independent variables and the 25th column (F25) represent the target/dependent variable as shown in Table 1. Descriptive statistics of these columns showed they contain discrete categorical data with ordinal values from 1–5. Furthermore, Fig. 3 displays a summary of the central tendency, dispersion and shape of the dataset's distribution, as relates to its mean, median and standard deviation (std). For instance, F3 has a mean of 2.48, median of 3, and std of 1.11 — implies that on the average, during the course of most of the project on which each respondent answers were based, Inflation or sudden increase in good/commodities (F3) was medium.

After EDA, correlation analysis was done to identify multicollinearity among predictors (features vs target) using their respective correlation coefficient values (See Fig. 4). The correlation matrix plot in Fig. 4

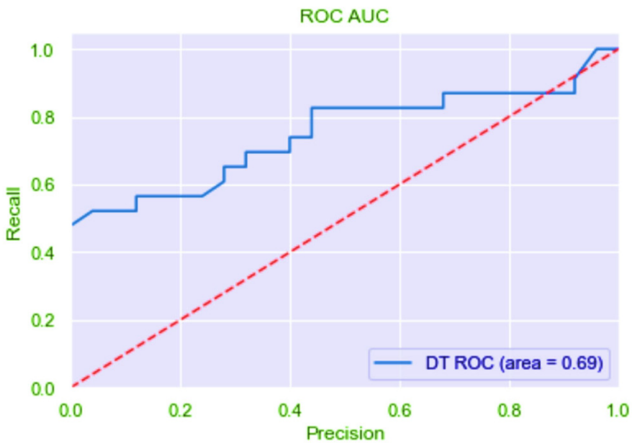

Fig. 7a. DT ROC AUC Plot.

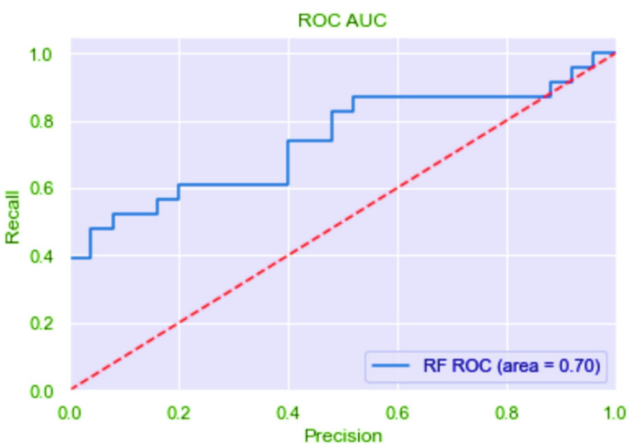

Fig. 7b. RF ROC AUC Plot.

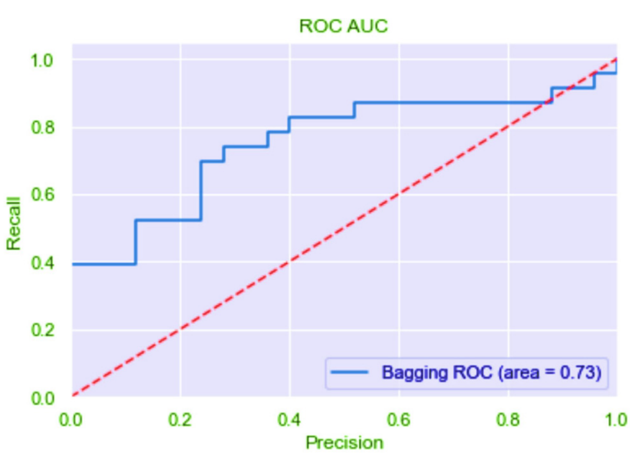

Fig. 7c. Bagging ROC AUC Plot.

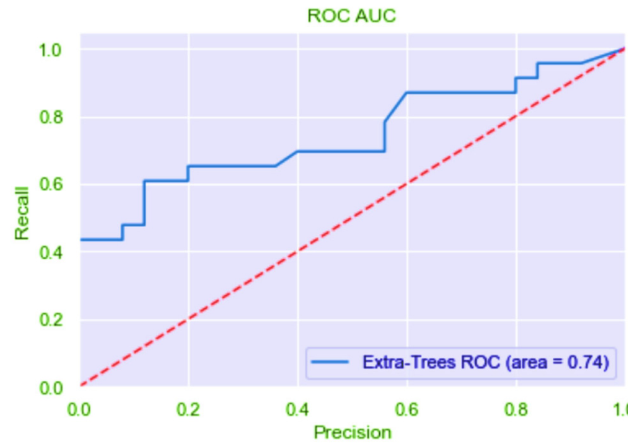

Fig. 7d. Extra Tree ROC AUC Plot.

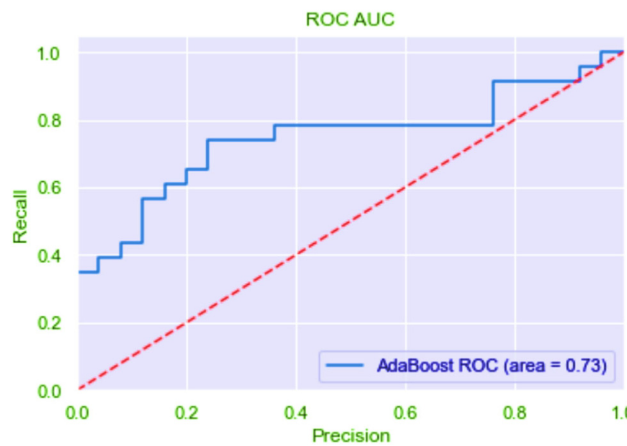

Fig. 7e. AdaBoost ROC AUC Plot.

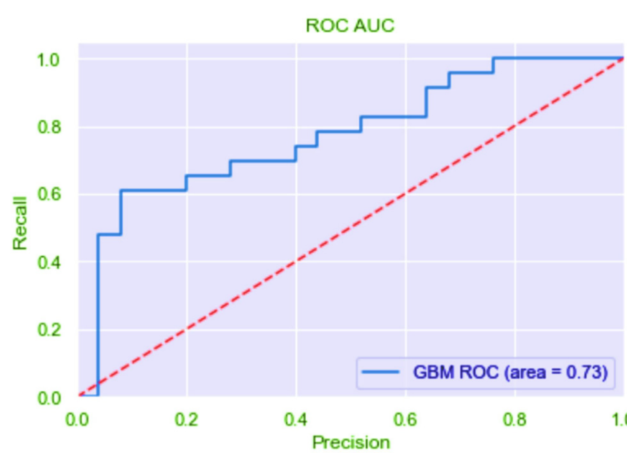

Fig. 7f. GMB ROC AUC Plot.

shows the cross correlation between each feature (F1–F24) and the target (F25). For example, F13 has a positive correlation of 0.4 to F25, F5 has a negative correlation of −0.07 to F25 and so on. In general, the existence of multicollinearity implies an absolute correlation coefficient >0.7 among two or more predictors (Dormann et al., 2013). Evidently, there exist multicollinearity between F12 and 13, F13 and F14, F13 and F16 etc as shown below.

3.3. Feature engineering

As a habitual requirement for most ML estimators owing to their underlying assumptions of any given dataset to be normally distributed, with zero mean and unit variance (Pedregosa et al., 2011), this study utilized standardization feature scaling method to meet this requirement by subtracting the mean from each feature observation and

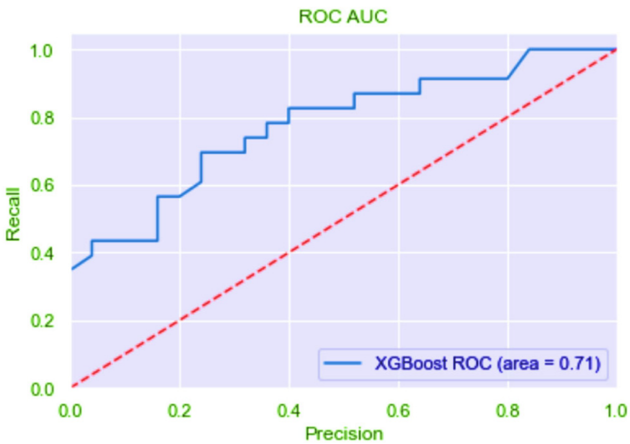

Fig. 7g. XGBoost ROC AUC Plot.

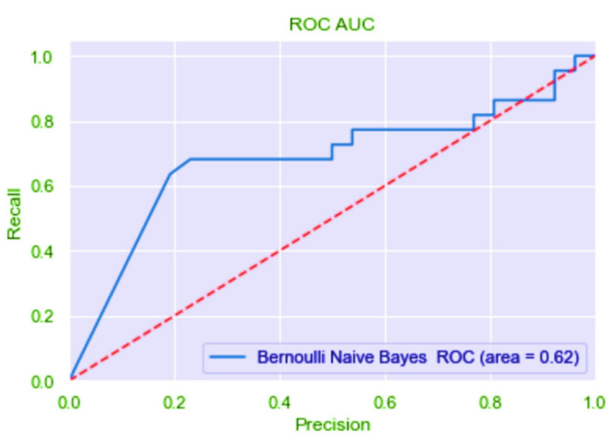

Fig. 7h. Bernoulli ROC AUC Plot.

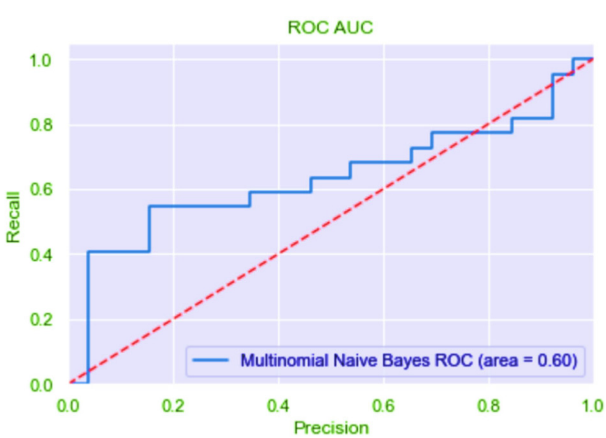

Fig. 7i. Gaussian ROC AUC Plot.

dividing by the standard deviation as shown in the equation below:

$$X' = \frac{X - \bar{\mathrm{x}}}{\sigma}$$

Where X' represents the standardized value; X a given feature observation; $\bar{\mathrm{x}}$ the mean and σ the standard deviation. Hence our resulting feature scaled dataset has its variance at 1, centred its mean at 0 with a varying min max value as shown in Fig. 5.

Furthermore, as a final transformation on the dataset, One-hot encoding (*k-1 variant*) a categorical encoding technique was done on

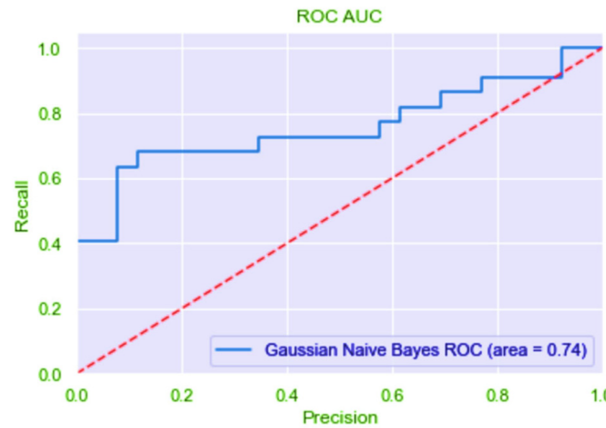

Fig. 7j. Multinomial ROC AUC Plot.

the target (F25). Consequently, data from column F25 were encoded into 0 (no delay) for its ordinal values containing any number <= 3 and 1(delay) for its ordinal values containing 4 and 5 such that on each occurrence, this value (0 or 1) can then help to show if a category (delay) is present or not. Finally, the dataset was split using Scikit-learn's train_test_split function at a ratio of 60:40 for training (72 data points) and testing (48 data points) respectively.

3.4. Feature selection

In typical machine learning pipeline, feature selection is a crucial mechanism designed to eliminate obsolete, redundant, and noisy characteristics and retain a limited subset of features from the primary feature space (Kira & Rendell, 1992; Wei et al., 2020). In relation to this study, a multivariate filter-based feature selection method called Chi-squared was chosen to eliminate obsolete, redundant and noisy features, boost model accuracy, improve model interpretability, lower computational complexity and enhance generalizability by mitigating overfitting. This Chi-squared method was inevitably chosen since our data contains categorical features (frequencies) and binary target variable. Fig. 6 shows the outcome of the Chi-squared test with the varying minimal degree of association of each feature and the target.

Consequently, these irrelevant features with the coloured bars (F6, F3, F1, F5, F2, F4, F19, F15, F7, F21, F23, F9, F22, F20, and F14) were removed before fitting different EMLA on the training dataset (60% of total dataset) and comparing their respective parameter settings on unseen test dataset (40% of total dataset) as a bias trade-off for individual EMLA to further improve their performance for making predictions. Hence only the remain 9 important feature factors out of the initial 24 factors was subsequently used.

3.5. Ensemble machine learning technique

This technique involves the use of multiple algorithms where the cumulative outcome from them is almost always greater in terms of predictive accuracy relative to the use of a single algorithm as it integrate decisions from different algorithms to maximize the overall performance (Badawi et al., 2019; Dietterich, 2000; Hastie et al., 2009). All examples of ensemble learning techniques available in Scikit-learn version 0.23.2 were used for EMLA experimentations in this study. They include Bagging (Bootstrap Aggregating), Boosting (Hypothesis Boosting), Naive Bayes, and Stacking (see Table 3). Their respective algorithms and libraries used are as follows in Table 3:

To further understand the underlying principles behind the proposed approach of this study we represent these ensemble methods mathematically by the following formulae:

Table 5
Default parameter ensemble's performance metrics report.

Algorithms	Delay Status	Precision (%)	Recall (%)	F_1 Score (%)	Accuracy Score (%)	Cross Validation Score (%)	ROC AUC
DT	0	65%	56%	60%	58.33%	63.50%	0.5873
	1	52%	62%	57%			
RF	0	67%	59%	63%	60.42%	71.50%	0.6058
	1	54%	62%	58%			
Bagging	0	62%	48%	54%	54.17%	75.50%	0.5503
	1	48%	62%	54%			
Extra-Trees	0	67%	59%	63%	60.42%	73.00%	0.6058
	1	54%	62%	58%			
AdaBoost	0	73%	59%	65%	64.58%	75.50%	0.6534
	1	58%	71%	64%			
GBM	0	60%	44%	51%	52.08%	68.50%	0.5317
	1	46%	62%	53%			
XGBoost	0	63%	44%	52%	54.17%	66.50%	0.5556
	1	48%	67%	56%			
BNB	0	69%	42%	52%	58.33%	60.00%	0.5979
	1	53%	77%	63%			
MNB	0	53%	35%	42%	47.92%	68.50%	0.4913
	1	45%	64%	53%			
GNB	0	74%	77%	75%	72.92%	74.00%	0.7255
	1	71%	68%	70%			

Bagging is mathematically expressed by the following formula:

$$f_{bag} = f_1(x) + f_2(x) + \cdots + f_b(x)$$

where the term on the left, f_{bag} is the bagged prediction, and $f_1(x)$ to $f_b(x)$ the actual learners (Random Forest, Bagging and Extra-Trees used in this study) are the term on the right. b represents the cumulative number of learners.

Three key steps were used to experiment the **boosting** ensemble technique. First, the target variable (projects delay) is predicted using an initial model f_0 with a residual (y – f_0). Secondly, a new model h_1 is fit to the previous step's residuals. Finally, f_0 and h_1 are merged to produce f_1, the boosted variant of f_0 as shown below:

$$f_1(x) < -f_0(x) + h_1(x)$$

To boost f_1's results, we built a new model f_m based on f_1's residuals repeated for 'm' iterations until the residuals are as low as possible as shown below:

$$f_m(x) < -f_{m-1}(x) + h_m(x)$$

Naive Bayes ensemble methods we used follows the Bayes' theorem which establishes the link between dependent variable y and related independent variables vector x_1 to x_n as shown below:

$$P\left(y|x_1,\ldots,x_n\right) = \frac{P(y)\,P\left(x_1,\ldots,x_n|y\right)}{P(x_1\ldots,x_n)} \propto P(y)\prod_{i=1}^{n} P\left(x_i|y\right)$$

Thus,

$$\hat{y} = \arg\max_{y} P(y)\prod_{i=1}^{n} P\left(x_i|y\right)$$

Where $\hat{y}$ and P are predicted class and probability of occurrence respectively.

Unlike bagging, boosting, and naive bayes, **stacking,** also known as stacked generalization whose base estimator(s) e.g., DT algorithm used in this study are trained on heterogeneous EMLA such that base estimator's outputs are combined using a meta-classifier as shown below:

$$\min_f \sum_{i=1}^{n} l\left(f\left(x_i\right), y_i\right) + \lambda r(f)$$

Where the first term in the above equation is the empirical risk which is defined by a loss function S, that evaluates the effectiveness of the function f. The second term is the regularization term, and it evaluates the complexity of the function f, which is normally a norm of function f or its derivatives. Consequently, we proceed to the performance metrics of EMLA in the next sub section.

3.6. Algorithms performance metrics

Typical performance metrics for evaluating classification-based problems like the one for this study are, accuracy score, confusion matrix, precision, recall, f_1 score and Compute Area Under the Receiver Operating Characteristic Curve (ROC AUC).

Table 6
Hyperparameter optimized ensemble's performance metrics report.

	Algorithms	Delay Status	Precision (%)	Recall (%)	F_1 Score (%)	Accuracy Score (%)	Cross Validation Score (%)	ROC AUC
BASE ESTIMATOR	DT	0	69%	72%	71%	68.75%	67.50%	0.6861
		1	68%	65%	67%			
BAGGING	RF	0	68%	84%	75%	70.83%	72.50%	0.7026
		1	76%	57%	65%			
	Bagging	0	73%	76%	75%	72.92%	78.50%	0.7278
		1	73%	70%	71%			
	Extra-Trees	0	71%	88%	79%	75.00%	76.00%	0.7443
		1	82%	61%	70%			
BOOSTING	AdaBoost	0	75%	72%	73%	72.92%	74.50%	0.7296
		1	71%	74%	72%			
	GBM	0	71%	80%	75%	72.92%	75.00%	0.7261
		1	75%	65%	70%			
	XGBoost	0	72%	72%	72%	70.83%	76.50%	0.7078
		1	70%	70%	70%			
NAIVE BAYES	Bernoulli	0	71%	46%	56%	60.42%	60.00%	0.6171
		1	55%	77%	64%			
	Multinomial	0	64%	62%	63%	60.42%	75.00%	0.6031
		1	57%	59%	58%			
	Gaussian	0	75%	81%	78%	75.00%	76.00%	0.7448
		1	75%	68%	71%			
STACKING	Ensemble of Ensembles	0	78%	95%	86%	80.65%	85.50%	0.7654
		1	88%	58%	70%			

Accuracy score is the sum of accurate estimates that have been made and separated into one per cent by the overall number of predictions that have been made. Accuracy score is generally not the only preferred metrics to use for classifiers especially with skewed datasets. The formula used to calculate accuracy score for this study is:

$$\text{Accuracy score} = \frac{\text{True Negatives} + \text{True Positives}}{\text{Number of Predictions}} x100$$

Confusion matrix is a 2 x 2 matrix description of the number of accurate and inaccurate predictions made by a classifier. The confusion matrix result used for EMLA experimentations is shown in Table 4.

Precision (false positive rate) measures the accuracy of positive predictions. Hence, in this study it is the accurately predicted ratio of cases with delay $\leq$ threshold limit to be less than or equal to the threshold limit to the total number of cases with delay $\leq$ threshold limit in the test data. It is expressed mathematically as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall (sensitivity or true positive rate) is the ratio of positive instances that are correctly detected by the classifier. It is the ratio of cases with delay > threshold accurately predicted to surpass the total number of cases with delay > threshold in the test data of this study. It is expressed mathematically as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F_1 Score is the harmonic mean of precision and recall. As regular mean gives equal weight to all values, harmonic mean gives more weight to low values. The F_1 score favours classifiers that have similar precision and recall in this study. It is expressed mathematically as:

$$F_1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = 2x\frac{\text{Precision } x \text{ Recall}}{\text{Precision} + \text{Recall}}$$

The Receiver Operator Characteristic Curve (ROC) is a recall plot of the y-axis against precision of the x-axis. In this study, the threshold of the algorithm, which ranges from zero to one with a scale of 0.1, is seen on the vertical axis to the right of the plot and on the curve as well. Area under the curve (AUC) is the area under the ROC curve that is generally recognized as the best indicator of the overall performance

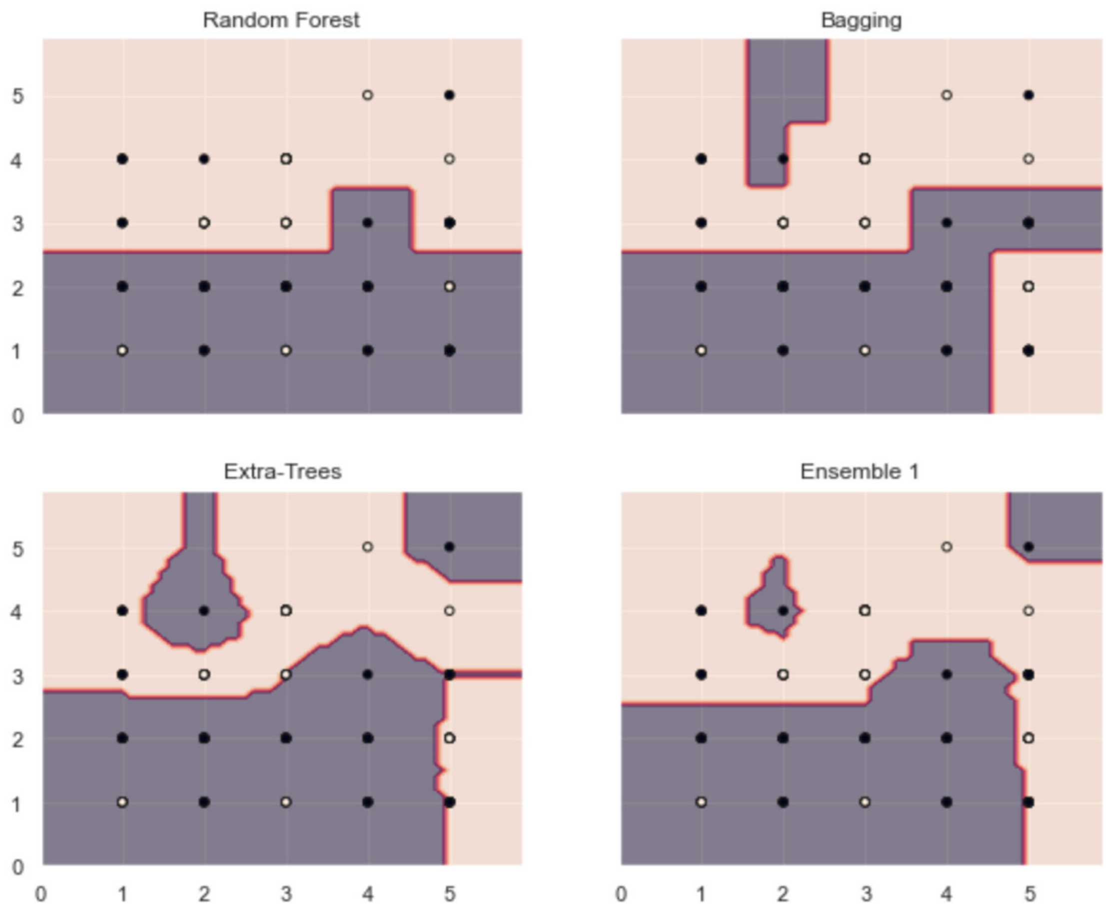

Fig. 8. Bagging Decision Boundaries.

of a classifier. Since the maximum value of the precision and recall of the x and y axes of the ROC curve are 1, the maximum AUC value, indicating excellent accuracy, is 1. The minimum AUC value, however, is 0.5.

Cross validation is a resampling technique for evaluating machine learning models on a small dataset. Since the estimators (see Table 3) for this study are classifiers and the target variable is binary, stratified k-fold a variant of k-fold that returns stratified folds containing about the same proportion of target class as the initial dataset is used to evaluate cross validation scores. As a result, the variation between the estimates is minimized, and the average error estimation is more accurate, hence mitigates potential overfitting. To obtain the cross validation score in this study, we took the mean of stratifield10-fold (i.e., where k = 10) for each EMLA.

The main parameters for each model are as follows. **(i) DT**: random_state = 42, min_samples_leaf = 5, criterion = 'gini', min_samples_split = 4, n_jobs = -1; **(ii) RF**: n_estimators = 100, n_jobs = -1, random_state = 42, bootstrap = True, warm_start = False; **(iii) Bagging**: n_estimators = 100, bootstrap = True, n_jobs = -1, random_state = 42, min_samples_leaf = 2, min_samples_split = 3, verbose = 1; **(iv) Extra-Trees**: min_samples_split = 4, random_state = 42, criterion = 'entropy', n_jobs = -1, min_samples_leaf = 2, n_estimators = 100; **(v) AdaBoost**: max_depth = 1, n_estimators = 100, learning_rate = 3, min_samples_split = 3, n_jobs = -1, random_state = 42; **(vi) GBM**: loss: 'deviance', n_jobs = -1, random_state = 42, learning_rate = 2, min_samples_split = 3, min_samples_leaf = 3; **(vii) XGBoost:** base_score = 1, booster = 'gbtree', colsample_bylevel = 1.9, colsample_bynode = 1, colsample_bytree = 1, gamma = 0, gpu_id = -1, importance_type = 'gain', learning_rate = 1, max_delta_step = 2, max_depth = 6, min_child_weight = 1, n_estimators = 100, n_jobs = -1, num_parallel_tree = 3, random_state = 42, tree_method = 'exact'; **(viii) BNB:** binarize = True, alpha = 0, fit_prior = True, class_prior = None; **(ix) MNB:** alpha = 100, fit_prior = False, class_prior = None; **(x) GNB:** priors = None, var_smoothing = 1e-01.

3.7. Algorithms performance evaluation

At first the EMLA experimentations was performed on the first 7 algorithms (see Table 3) with their default parameters (without hyperparameter optimization) on the unseen test dataset (40% of total dataset) after training the EMLA (model fitting on training dataset). Table 5 shows their respective evaluation metrics report on test dataset. A close attention to column 9 of Table 5 shows that the best of them only had a 15% increase in the minimum AUC value. Hyperparameter optimization became more necessary, hence we proceeded with it and decided not to continue the model training and testing experiment with the default parameter for the last algorithm due to computational cost and time.

Interestingly, we obtained almost double (27%) in value after hyperparameter optimization on the EMLA when compared to the initial EMLA experimentations with their default parameters on test dataset (40% of total dataset) as shown in Table 6.

Table 6 presents the evaluation metrics report of the models developed with the 11 algorithms used for EMLA experimentations on the test data (40% of total dataset) for a weak learner DT as the base estimator. Comparing the ensemble of ensembles (stacking) approach proposed with existing naïve bayes method used to potentially predict construction project delay risk by Gondia et al. (2020), this report clearly shows and confirms that although naïve bayes which is considered as one of the most effective and efficient inductive EMLA due to the conditional independence assumption on which it is theoretically built, is however rarely valid in a real-world applications such as

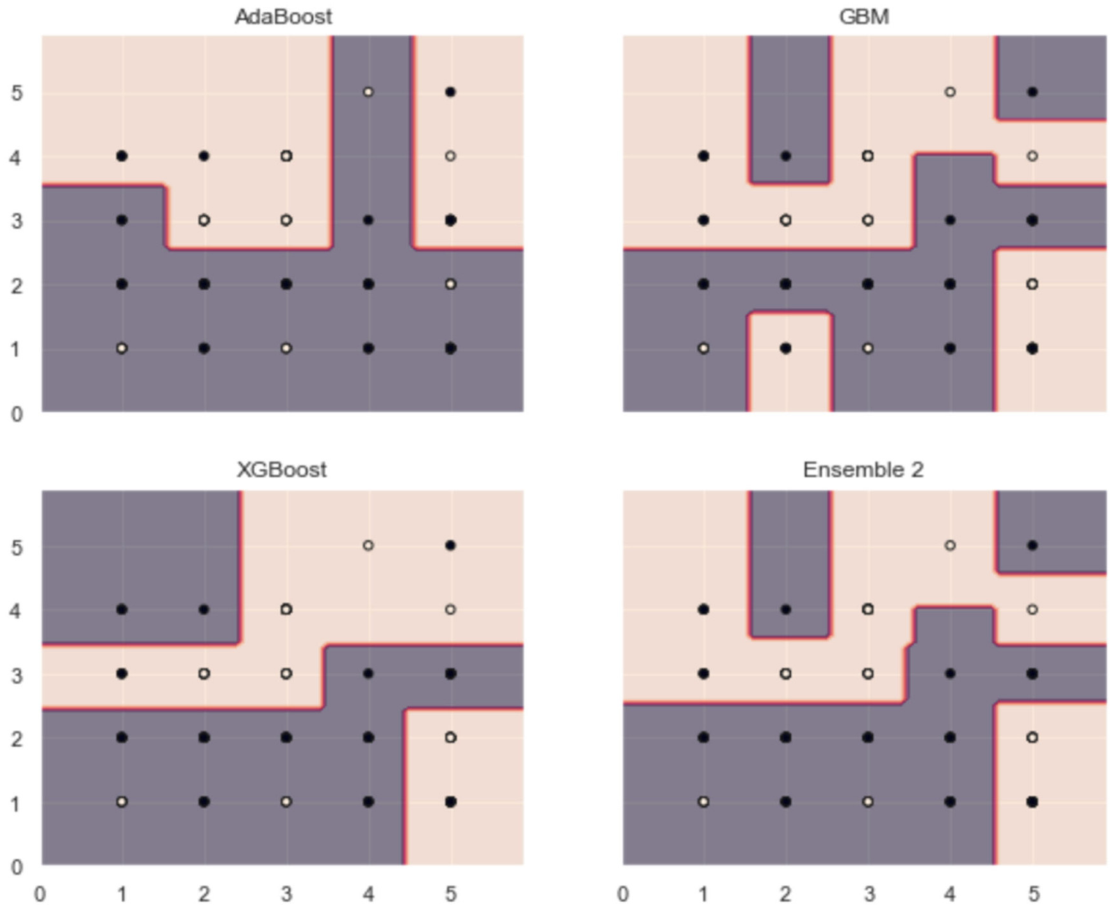

Fig. 9. Boosting Decision Boundaries.

the case in this study (Zhang, 2004). Furthermore, the better performance (75%, 76%, and 0.7448 of accuracy score, cross validation score and ROC AUC respectively) of Gaussian naïve bayes over Bernoulli naïve bayes (60.42%, 60% and 0.6171 of accuracy score, cross validation score and ROC AUC respectively) and multinomial naïve bayes (60.42%, 75% and 0.6031 of accuracy score, cross validation score and ROC AUC respectively) owes to the existence of dependences between its features since our dataset was optimally standardized to normally distributed features (see Section 3.3). Fig. 11 (see Appendix A) shows the tree structure of this base estimator while Figs. 7a to 7j present the multiple ROC AUC plot for the algorithms used.

For a start, we benched marked our EMLA on DT (typical unstable model) as their base estimator, then we proceeded the experimentation by trying to gain stability for the base estimator using 3 bagging ensemble algorithms namely optimized RF (a natural ensemble of DT), Bagging and Extremely Randomized Trees. As expected, they all performed better than the base estimator based on all evaluation metrics (see Table 6), the challenge, however, was how to identify the best ensemble algorithm. To avoid bias and to enhance generalizability, we casted a vote with the bagging ensembles using the hard and soft voting rule in Scikit-learn's VotingClassifier.

A resulting performant model (higher accuracy score and low variance) undoubtedly emerged. The experiment was repeated with the base estimator but this time using 3 boosting ensemble algorithms: Adaptive Boosting (CART), Gradient Boosting Machine, and Extreme Gradient Boosting and not surprisingly similar to the bagging ensembles, they all performed better than the base estimator (see Table 6). We again casted votes amongst them which yielded yet another performant model. Finally, we again repeated the experiment with the base estimator using 3 naïve bayes ensembles: Bernoulli, multinomial and Gaussian, however although one of them (multinomial) did not perform as much as the base estimator, we proceeded with vote casting, which ultimately yielded a third performant model. Figs. 8–10 shows how these ensembles made their respective decisions interdependently.

Where Ensemble 1, 2, and 3 are the resulting new performant models obtained during vote casting in ensemble bagging, boosting, and naïve bayes respectively. All in all, we conclusively proceeded to use these aggregated predictions from Ensemble 1, 2, and 3 to train a new algorithm called Ensemble of Ensembles using Scikit-learn's Mlens library. Consequently, resulting to a more performant model with a much more better accuracy score, confusion matrix, precision, recall, f1 score, and Compute Area Under the Receiver Operating Characteristic Curve (ROC AUC) as shown in Table 6.

4. Conclusion & recommendations

The perpetual occurrence of a global phenomenon — delay in construction sector despite considerable mitigation efforts remains a huge concern to its policy makers. Interestingly, this sector which produces massive amount of data from IoT sensors, building information modelling, on most of its projects daily is slow in taking the advantage of the contemporary analysis method — artificial intelligence/machine learning which best explains the factors that can affect a phenomenon like delay based on its predictive capabilities haven been widely adopted across other sectors. In this study therefore, a premise to use ensemble machine learning algorithms (EMLA) for predicting delay of construction projects was architected, built and presented. First a review of existing body of knowledge on influencing factors of construction projects delay was used to conduct survey of experts as an approach to its data collection and exploration.

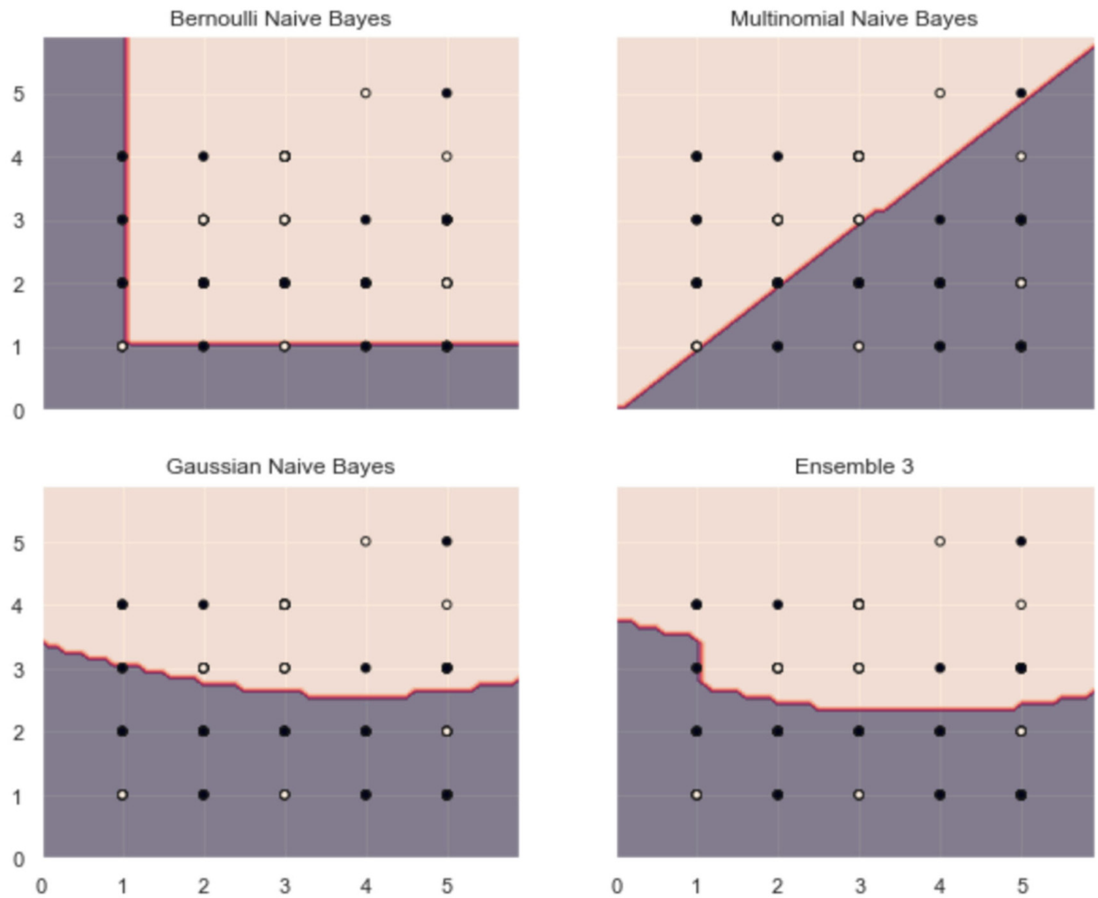

Fig. 10. Naïve Bayes Decision Boundaries.

The resulting dataset applied to EMLA was used to develop hyperparameter optimized predictive models: Decision Tree, Random Forest, Bagging, Extremely Randomized Trees, Adaptive Boosting (CART), Gradient Boosting Machine, and Extreme Gradient Boosting. Finally, a multilayer high performant ensemble of ensembles predictive model was developed to maximize the overall performance of the EMLA combined.

Results from the algorithm evaluation metrics: accuracy score, confusion matrix, precision, recall, F1, and ROC AUC indeed proved that EMLA are capable of improving the predictive force relative to the use of a single algorithm in predicting construction projects delay. By developing a multilayer high performant ensemble of ensembles predictive model, the current research contributes to the effort of improving time efficiency of construction projects – a key performance indicator for successful projects. Ultimately, this model can subsequently be integrated into construction information system to promote evidence-based decision-making, thereby enabling constructive project risk management initiatives. As compared to existing numerical or statistical approaches, which used pure mathematical techniques such as the arithmetic mean, standard deviation, hypothesis testing, etc. to draw inference from data, our predictive analytics approach used known results (input variables), statical methods and advance ML algorithms to develop a novel multilayer high performant ensemble of ensembles predictive model to forecast futuristic delay values for complex and new data of typical construction projects. Thus, will help improve the quality of decisions and risks to be taken by several construction sector stakeholders on their present or future construction projects which as a result will foster trust, increase in productivity and revenue and more importantly yield timely delivery of construction projects in the sector. While the proposed contemporary method of analysis is assumed to be applicable in mitigating delay of any construction project within the sector, the unique data transformation employed in this study may not, as typical of any data driven model, be transferable to the data from other regions. Nevertheless, other region's project datasets can be applied to the processes described in this study. Also, the sample size of the respondents of this study may not be representative of the total population size of the region. In order to produce improved classification outcomes, future studies should be targeted at extending the algorithms either by further parameter optimization or feature engineering. Other methods used in the creation of ensemble models, apart from bagging, boosting, naïve bayes and stacking, should also be considered for predicting construction projects delay.

CRediT authorship contribution statement

Christian Nnaemeka Egwim: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Hafiz Alaka:** Supervision, Writing – review & editing, Investigation, Visualization. **Luqman Olalekan Toriola-Coker:** Data curation. **Habeeb Balogun:** Resources, Data curation. **Funlade Sunmola:** Supervision, Writing – review & editing.

Appendix A

See Fig. 11.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.mlwa.2021.100166.

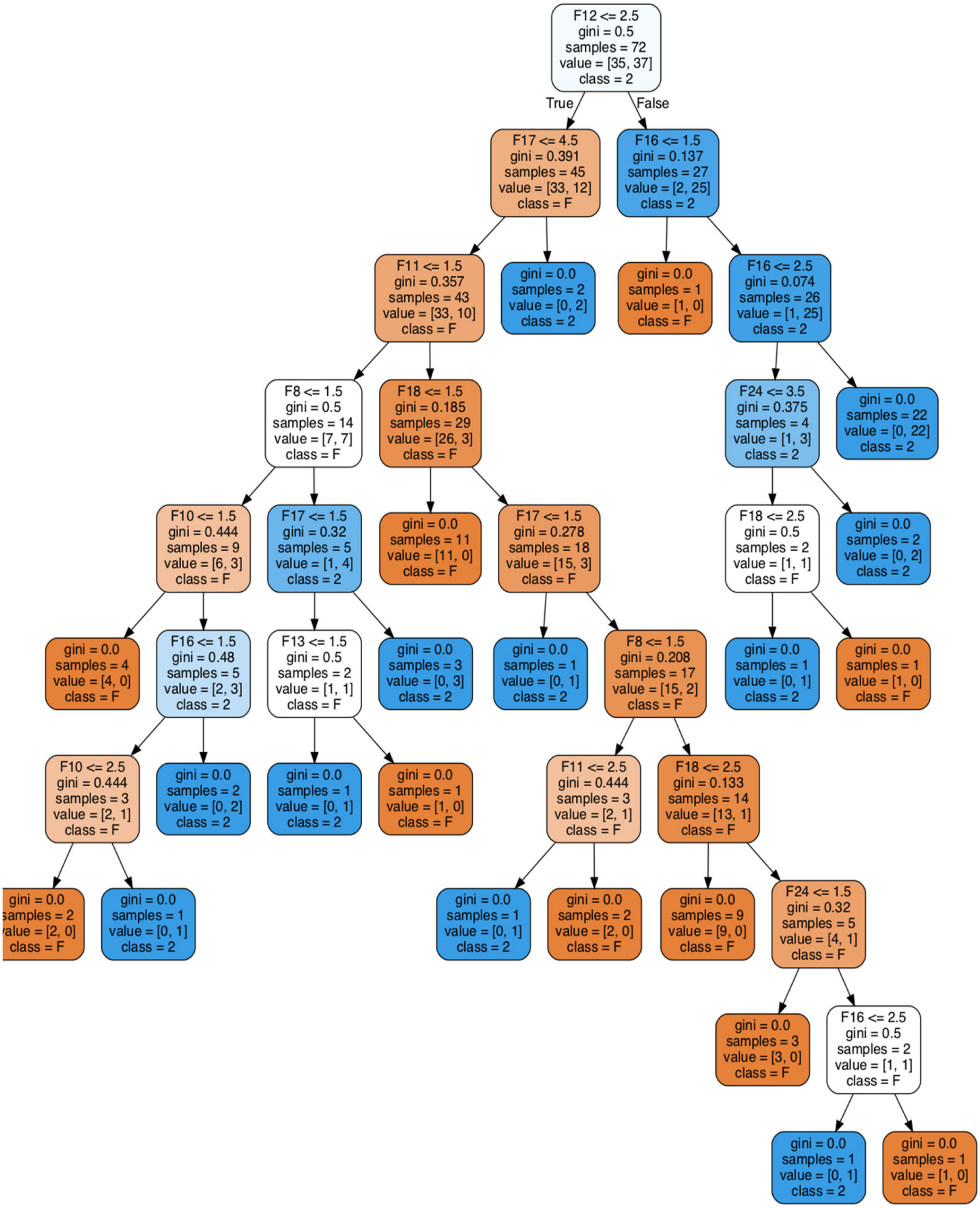

Fig. 11. DT as base estimator.

References

Abdul-Rahman, H., et al. (2011). Project schedule influenced by financial issues: Evidence in construction industry. *Scientific Research and Essays*, *6*(1), 205–212. http://dx.doi.org/10.5897/SRE10.989.

Agyekum-Mensah, G., & Knight, A. D. (2017). The professionals' perspective on the causes of project delay in the construction industry. *Engineering, Construction and Architectural Management*, *24*(5), 828–841. http://dx.doi.org/10.1108/ECAM-03-2016-0085.

Alaghbari, W., & Sultan, B. (2018). Delay factors impacting construction projects in Sana ' a -Yemen 1. *PM World Journal*, *VII*(December), 1–28, Available at: https://www.researchgate.net/publication/329656460%0ADelay.

Ameziane, F. (2000). Information system for building production management. *International Journal of Production Economics*, *64*(1), 345–358. http://dx.doi.org/10.1016/S0925-5273(99)00071-7, Elsevier Science B.V..

Arditi, D., & Pulket, T. (2005). Predicting the outcome of construction litigation using boosted decision trees. *Journal of Computing in Civil Engineering*, *19*(4), 387–393. http://dx.doi.org/10.1061/(ASCE)0887-3801(2005)19:4(387), American Society of Civil Engineers.

Asadi, A., Alsubaey, M., & Makatsoris, C. (2015). A machine learning approach for predicting delays in construction logistics. *International Journal of Advanced Logistics*, *4*(2), 115–130. http://dx.doi.org/10.1080/2287108x.2015.1059920, Informa UK Limited.

Assaf, S. A., & Al-Hejji, S. (2006). Causes of delay in large construction projects. *International Journal of Project Management. Pergamon*, *24*(4), 349–357. http://dx.doi.org/10.1016/j.ijproman.2005.11.010.

Assaf, S. A., Al-Khalil, M., & Al-Hazmi, M. (1995). Causes of delay in large building construction projects. *Journal of Management in Engineering. American Society of Civil Engineers*, *11*(2), 45–50. http://dx.doi.org/10.1061/(ASCE)0742-597X(1995)11:2(45).

Badawi, H., et al. (2019). Use of ensemble methods for indirect test of RF circuits: Can it bring benefits? In *LATS 2019-20th IEEE Latin American test symposium (no. 1)*. http://dx.doi.org/10.1109/LATW.2019.8704641.

Baldwin, J. R., Manthei, J. M., Rothbart, H., & Harris, R. B. (1971). Causes of delay in the construction industry. *Journal of the Construction Division*, 177–187, Available at: https://cedb.asce.org/CEDBsearch/record.jsp?dockey=0018302. (Accessed 23 April 2020).

Bartholomew, S. H. (2001). *Construction contracting: business and legal principles*. Prentice Hall.

Bhatnagar, R., Kim, J., & E. Many, J. (2014). Candidate surveys on program evaluation: Examining instrument reliability, validity and program effectiveness. *American Journal of Educational Research. Science and Education Publishing Co. Ltd*, *2*(8), 683–690. http://dx.doi.org/10.12691/education-2-8-18.

Blanco, J. L., et al. (2018). *Artificial intelligence : Construction technology ' s next frontier*. Mckinsey & Company, (April), 1–8. Available at: https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/artificial-intelligence-construction-technologys-next-frontier.

Chen, G. X., et al. (2019). Investigating the causes of delay in grain bin construction projects: the case of China. *International Journal of Construction Management. Taylor and Francis Ltd.*, *19*(1), 1–14. http://dx.doi.org/10.1080/15623599.2017.1354514.

Delice, A. (2001). The sampling issues in quantitative research. *Educational Sciences: Theory & Practices*, *10*(4), 2001–2019.

Dietterich, Thomas G. (2000). Ensemble methods in machine learning. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, (pp. 1–15). http://dx.doi.org/10.1007/3-540-45014-9_1.

Dormann, C. F., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography. Blackwell Publishing Ltd*, *36*(1), 27–46. http://dx.doi.org/10.1111/j.1600-0587.2012.07348.

Durbarry, R. (2019). Quantitative research. *Research Methods for Tourism Students*, (August), 98–113. http://dx.doi.org/10.4324/9780203703588-9.

Easterby-Smith, M., et al. (2008). Management research: theory and practice. cloudgene.uibk.ac.at. Available at: https://cloudgene.uibk.ac.at/l2xq3pav9aoy/12-alysha-lowe/read-9781847871770-management-research-theory-and-practice-paperbac.pdf. (Accessed 24 September 2020).

Egan, S. J. (2018). *Rethinking the Report of the Construction Task Force*, Construction Task Force, Available at: http://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.

Egwim, C. N., et al. (2021). Extraction of underlying factors causing construction projects delay in Nigeria. *Journal of Engineering, Design and Technology*, http://dx.doi.org/10.1108/jedt-04-2021-0211, ahead-of-p(ahead-of-print).

Enshassi, A., Al Najjar, J., & Kumaraswamy, M. (2009). Delays and cost overruns in the construction projects in the Gaza Strip. *Journal of Financial Management of Property and Construction*, *14*(2), 126–151. http://dx.doi.org/10.1108/13664380910977592.

European Comission (2017). European construction sector observatory. (p. 27). (June).

Flick, U. (2014). *An introduction to qualitative research*. Available at: https://us.sagepub.com/en-us/nam/an-introduction-to-qualitative-research/book240398.

Flyvbjerg, B. (2014). What you should know about megaprojects and why: An overview. *Project Management Journal*, *45*(2), 6–19. http://dx.doi.org/10.1002/pmj.21409.

Gliem, J. a., & Gliem, R. R. (2003). Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales. In *2003 Midwest research to practice conference in adult, continuing, and community education* (1992), (pp. 82–88). http://dx.doi.org/10.1109/PROC.1975.9792.

Gondia, A., et al. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management. American Society of Civil Engineers (ASCE)*, *146*(1), Article 04019085. http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0001736.

Goyal, M. (2019). Artificial intelligence : A tool for hyper personalization. *International Journal of 360 Management Review*, *07*, 2320–7132.

Haq, S., et al. (2017). Effects of delay in construction projects of punjab-Pakistan : An empirical study effects of delay in construction projects of punjab-Pakistan. *Journal of Basic and Applied Scientific Research*, *3*(January 2014), 87–96.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Ensemble learning. (pp. 1–20). http://dx.doi.org/10.1007/b94608_16.

James, E., Joe, W., & Chadwick, C. (2001). Organizational research : Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, *19*(1), 43–50.

Ji, Y., et al. (2018). Assessing and prioritising delay factors of prefabricated concrete building projects in China. *Applied Sciences*, *8*(11), 2324. http://dx.doi.org/10.3390/app8112324, (Switzerland). MDPI AG.

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Machine learning proceedings 1992* (pp. 249–256). Elsevier, http://dx.doi.org/10.1016/b978-1-55860-247-2.50037-1.

Kumar, R. D. (2016). Causes and effects of delays in construction industry. *International Research Journal of Engineering and Technology*, *3*(4), 1831–1837, Available at: www.irjet.net.

Mahfouz, T., & Kandil, A. (2012). Litigation outcome prediction of differing site condition disputes through machine learning models. *Journal of Computing in Civil Engineering*, *26*(3), 298–308. http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000148.

Marks, M. (2017). Construction: The next great tech transformation Voices Michael Marks.

Matthews, B., Ross, L., & Ellison, N. (2010). RESEARCH METHODS : A great starting point for students and would-be social researchers. Available at: www.pearsoned.co.uk/matthews. (Accessed 24 September 2020).

Morgan, D. H. J. (1980). Sociological paradigms and organisational analysis. *Sociology*, *14*(2), 332–333. http://dx.doi.org/10.1177/003803858001400219.

Motaleb, O., & Kishk, M. (2010). An investigation into causes and effects of construction delays in UAE. In *Association of Researchers in Construction Management, ARCOM 2010 - Proceedings of the 26th Annual Conference* (pp. 1149–1157). Available at: https://www.researchgate.net/publication/266174953. (Accessed 24 April 2020).

Nunnally, J. C. (1978). *Phychometric theory* (p. 640).

Odeyinka, H., & Adebayo, Y. (1997). The causes and effects of construction delays on completion cost of housing projects in Nigeria. *Journal of Financial Management of Property and Construction*, *2*(3), 31–44, Available at: https://www.researchgate.net/publication/249643683. (Accessed 24 April 2020).

Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research. Morgan Kaufmann Publishers*, *11*, 169–198. http://dx.doi.org/10.1613/jair.614.

Owalabi, J. D., et al. (2014). Causes and effects of delay on project construction delivery time. *International Journal of Education and Research*, *2*(4), 197–208, Available at: www.ijern.com. (Accessed 24 April 2020).

Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830, Available at: http://scikit-learn.sourceforge.net. (Accessed 7 January 2021).

Poh, C. Q. X., Ubeynarayana, C. U., & Goh, Y. M. (2018). Safety leading indicators for construction sites: A machine learning approach. *Automation in Construction*, *93*, 375–386. http://dx.doi.org/10.1016/j.autcon.2018.03.022, Elsevier B.V..

Rhodes, C. (2019). *Construction industry: statistics and policy* (01432), (pp. 1–13). House of Commons Library.

Robinson, T. G. (2015). Global construction market to grow $ 8 trillion by 2030 : driven by China, US and India. *Global Construction*, *44*.

Royal Society, T. (2017). Machine learning: The power and promise of computers that learn by example.

Saunders, M., et al. (2009). *Research methods for business students* (5th ed.). Pearson Education Limited, Available at www.pearsoned.co.uk. (Accessed 24 September 2020).

Seni, G., & Elder, J. F. (2010). *Synthesis lectures on data mining and knowledge discovery*: *vol. 2*, *Ensemble methods in data mining: improving accuracy through combining predictions* (1), (pp. 1–126). Morgan & Claypool Publishers LLC, http://dx.doi.org/10.2200/s00240ed1v01y200912dmk002.

Shah, R. K. (2016). An exploration of causes for delay and cost overruns in construction projects: Case study of Australia, Malaysia & Ghana. *Journal of Advanced College of Engineering and Management*, *2*, 41. http://dx.doi.org/10.3126/jacem.v2i0.16097, Nepal Journals Online (JOL).

Sullivan, A., & Harris, F. C. (1986). Delays on large construction projects. *International Journal of Operations & Production Management*, *6*(1), 25–33. http://dx.doi.org/10.1108/eb054752.

Wang, T. (2018). China: construction industry's contribution share to GDP 2018–2021 | statista. Statista. Available at: https://www.statista.com/statistics/1068213/china-construction-industry-gdp-contribution-share/. (Accessed 18 April 2020).

Wang, T. (2019). Value added of U.S. construction as a percentage of GDP 2018. Statista. Available at: https://www.statista.com/statistics/192049/value-added-by-us-construction-as-a-percentage-of-gdp-since-2007/. (Accessed 18 April 2020).

Wei, G., et al. (2020). A novel hybrid feature selection method based on dynamic feature importance. *Applied Soft Computing*, *93*, Article 106337. http://dx.doi.org/10.1016/j.asoc.2020.106337.

Woetzel, J., et al. (2017). Bridging infrastructure gaps has the world made progress?. McKinsey Global Institute, (October), p. 10. Available at: https://www.mckinsey.com/~/media/McKinsey/Industries/CapitalProjectsandInfrastructure/OurInsights/Bridginginfrastructuregapshastheworldmadeprogress/BridginginfrastructuregapsHowhastheworldmadeprogressv2/MGI-Bridging-infrastructure-gaps.

Yaseen, Z. M., et al. (2020). Prediction of risk delay in construction projects using a hybrid artificial intelligence model. *Sustainability*, *12*(4), 1514. http://dx.doi.org/10.3390/su12041514.

Zhang, H. (2004). The optimality of naive Bayes naive Bayes and augmented naive Bayes. *Aa*, *1*(2), 3, Available at: www.aaai.org. (Accessed 27 May 2021).

Zou, Z., & Ergan, S. (2019). Leveraging data driven approaches to quantify the impact of construction projects on urban quality of life. (pp. 1–31). arXiv preprint arXiv:1901.09084. Available at: https://arxiv.org/abs/1901.09084. (Accessed 8 May 2020).