Data-Driven Prediction of Contract Failure of Public-Private Partnership Projects

Yongqi Wang[1]; Zhe Shao[2]; and Robert L. K. Tiong[3]

Abstract: The public-private partnership (PPP) has been adopted by many governments in developing countries to provide better public services. However, PPP projects have a high risk of contract failure. To proactively predict PPP contract failure and obtain the most significant failure factors from a quantitative perspective, this research compared the performance of different combinations of machine learning models and data-balancing techniques. Forty-three project-specific and country-specific factors were examined, and the top 15 were chosen for the transportation, water and sewer, and energy sectors. The results show that the selected model can forecast contract failure with a recall of 75.9%, 73.3%, and 76.2%, respectively. This study showed the effectiveness and applicability of machine learning in predicting PPP contract failure. The results can facilitate decision making by forecasting the probability of PPP contract failure in the early planning stage.
DOI: 10.1061/(ASCE)CO.1943-7862.0002124.

Author keywords: Public-private partnership (PPP); Contract failure; Failure factors; Machine learning; Decision-making process.

Introduction

Public-private partnership (PPP) procurement has been increasingly adopted by emerging markets to deliver sustainable and advanced infrastructure because of the fiscal limitations of governments and increasing needs for public services (Osei-Kyei and Chan 2019). Unlike traditional forms of public financing, the private-sector partner signs a long-term contract with a public agency to share responsibilities and profits in a PPP project. Consequently, the involvement of the private sector can help governments achieve better public services by introducing practical management experience, state-of-the-art technologies, innovative ideas, and ample funding (Jiang et al. 2019). At the same time, the private sector may open new international markets (Marques 2018; Ye et al. 2018). Despite these benefits, there is a potential risk of contract failure in PPP projects due to long contract periods and large-scale investments (Song et al. 2018).

PPP contract failure is defined as the early termination or distress of a project contract (Marcelo et al. 2017). It usually happens after renegotiation between public and private parties and may have a profound impact on them. For example, public agencies may bear an unavoidable, unexpected financial burden by redeeming private shares, and private-sector partners may receive a bad reputation (Demirel et al. 2019; Zhang and Xiong 2015). It is therefore crucial to have the ability to predict the outcome of a PPP project before its implementation. Many studies have concentrated critical success factors (CSFs) to ease the feasibility analysis of a PPP project. CSFs can include proper selection of projects, domestic government support, transparent bidding, and risk allocation (Ahmadabadi and Heravi 2019; Hsueh and Chang 2017; Tiong 1996). According to Díaz (2020), these CSFs may be different from those associated with contract failure, and the extent of their influence can vary. Nevertheless, literature on the determinants of PPP contract failure is scant. Knowing the key factors triggering contract failure and how they influence the final status of a PPP project can help PPP promoters devise sound strategies to avoid initiation of a project that has a high probability of contract failure.

PPP projects are intricate because of their unique characteristics, such as lengthy contracts and complex relationships between private-sector companies and government agencies (Abdullah and Khadaroo 2020). From the perspective of contract form, a PPP project adopts a build-operate-transfer (BOT) contract, a lease contract, or a build-own-operate (BOO) contract. From the perspective of project type, a PPP project can be a greenfield or a brownfield. From the perspective of disciplines, a PPP project can involve transport, energy, or water and sewer. Project-specific factors make each PPP project unique. In other words, the probability of contract failure should depend not only on its investment environment but also on its specific properties.

The main objective of this study was to determine the primary failure triggers of the PPP contract and predict the contract's final status using the classification machine learning model, which contains a series of algorithms as well as modeling tools for data processing and analysis. Three questions must be answered: (1) how to build a machine learning model to forecast the final status of a PPP contract, (2) what factors lead to PPP contract failure, and (3) how political, institutional, macroeconomic, and project-specific parameters affect the outcome of a PPP project.

Literature Review

PPP models are becoming popular worldwide, particularly in developing countries, for promoting urban development. However, most developing countries lack the knowledge to initiate and manage such projects. Moreover, they do not have a mature framework for predicting project outcomes in advance. As a result, contract

[1]Ph.D. Student, School of Civil and Environmental Engineering, Nanyang Technological Univ., 639798, Singapore. Email: yongqi001@e.ntu.edu.sg

[2]General Manager, Urban Dot Solution Pte Ltd., 1 lrving Place, #08-01, 369546, Singapore. Email: alex@urbandotsolution.com

[3]Associate Professor, School of Civil and Environmental Engineering, Nanyang Technological Univ., 639798, Singapore (corresponding author). Email: clktiong@ntu.edu.sg

Note. This manuscript was submitted on November 11, 2020; approved on March 29, 2021; published online on June 2, 2021. Discussion period open until November 2, 2021; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, © ASCE, ISSN 0733-9364.

failure may occur. It has been argued that contract failure is increasing due to rising costs and reductions in project financing caused by the global financial crisis (Harris and Pratap 2009).

In this context, some studies have examined the primary PPP failure factors. For example, Song et al. (2018) found 11 primary factors, based on case studies, behind early termination of Chinese PPP contracts, such as improper operation and government payment default. Tariq and Zhang (2020),reviewing 35 projects, revealed 60 critical failure determinants in water PPP projects from a global perspective. Dolla and Laishram (2020), using questionnaires and interviews, found 13 negative factors for Indian municipal solid waste PPP projects, such as lengthy delays due to politics. Most current research is on specific sectors, regions, and countries. Failure factors in one place may differ from those in other places, which may mislead decision making (Kwofie et al. 2016). Moreover, the studies just mentioned employed case studies and surveys and so the results may contain errors due to mistakes made during the surveys as well as bias introduced by respondents.

In order to resolve such issues, some studies have identified factors driving PPP contract failure based on real-world data sets. For example, Harris and Pratap (2009) revealed key factors that influence project cancellation by analyzing data from the World Bank's Private Participation in Infrastructure (PPI) database, such as previous experience of the host country and project sectors. Mansaray (2018) obtained the main risks associated with PPP contract failure, such as corruption and governance variables, using the PPI database. Although these studies determined PPP failure factors, they could not use them to predict the final status of an existing project or a project to be initiated given the combinations of variables. When applying these criteria to the feasibility of a PPP project, the main challenges are analyzing and matching them with the project background. The assessment results tend to be subjective and are highly dependent on the knowledge and even interests of the decision makers.

In other words, rather than knowing the exact probability of contract failure, decision makers can obtain a qualitative result only by referring to these factors. As a result, they may implement a project that is more likely to fail while ignoring projects that might be promising. To date, there are only limited quantitative studies on the prediction of final status. For example, Ng et al. (2010) used the structural equation model to predict whether PPP projects in Hong Kong would meet the requirements of different stakeholders. Akbari et al. (2020) proposed a decision support system to predict the success of infrastructure projects based on sustainable success factors. However, such models still need prior knowledge from users such as success criteria, which are neither sufficient nor accurate.

Machine learning is a potential solution to these problems which has been extensively applied to the PPP domain. For instance, Chou and Lin (2013) used ensemble machine learning models to predict disputes in PPP projects before they actually happen. Owolabi et al. (2018) forecast potential delays in PPP projects using machine learning models such as regression tree, support vector machine, and deep neural network. Wan and Fang (2020) established the risk evaluation index for PPP expressway projects by adopting the support vector machine model.

When disputes, cost overruns, and construction delays happen during the contract period, negotiations might be held between the public and private partners. However, if the negotiations fail, contract failure occurs (Tariq and Zhang 2021). The result of PPP contract failure is therefore worse than that of disputes and time and cost overruns. Although current studies imply the feasibility of machine learning models in the PPP domain, no studies have been carried out on the prediction of PPP contract success or failure.

When compared with conventional statistical methods, the machine learning model focuses more on predicting the result rather than on determining the relationships between variables (Rajula et al. 2020). Unlike traditional statistical techniques that have to follow specific rules, the model learns from the data by providing inputs and outputs. In this study, various machine learning models, such as logistic regression (LR), support vector machine (SVM), and random forest (RF), were used to predict the final status of a PPP project and determine contract failure determinants considering both project-specific variables and the host country investment environment. Consequently, this is a classification problem that predicts the discrete class label of the output, such as success or failure.

Research Methodology

To provide decision support to both public and private partners, it is of high importance to have a good understanding of the relationships between failure factors and the final status of a PPP project. Various models were used to predict the possibility of contract failure based on different variables. This section introduces the proposed methodology, which comprised four modules: (1) data preparation, (2) machine learning, (3) performance evaluation, and (4) model interpretation.

Data Preparation

Collection

Generally, factors that determine the final status of a PPP project can be placed in four categories: project-specific (e.g., project type), institutional (e.g., corruption control), legal (e.g., rule of law), and macroeconomic (e.g., gross domestic product, consumer price index) (Díaz 2020; Ng et al. 2010; Zhang et al. 2016).

Project-specific factors were extracted from the PPI database, which includes 7,096 PPP projects dating from 1990 to 2019. These projects are from 121 developing countries in regions such as the Middle East and North Africa (MENA), South Asia (SAR), and East Asia and the Pacific (EAP). They belong to different sectors, including transportation, water and sewer, and energy. Each project presents a series of variables, such as financial closure year, project type, project host country, and project status. Also, provided is a detailed description of each project, such as project structure and timeline and sponsor. In the context of machine learning, each project in the data set is referred to as an observation and the project variables are referred to as attributes.

Institutional and legal factors were obtained from the World Bank's Worldwide Governance Indicators (WGI) database, which provides six indicators of governance quality based on survey results from enterprises, citizens, and experts. These indicators are voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption. The macroeconomic factors are from the World Bank's World Development Indicators (WDI) database, which contains 1,600 time series indicators for 217 economies, some of which go back 50 years. Nine distinct indicators were used, such as GDP deflator, GDP per capita, and tax as GDP.

The WGI and WDI databases are not complete. For instance, contract period and PPP project type may be missing and have to be supplied so they can be fed into a data-mining model.

Missing Values

Handling missing values is imperative to improve the accuracy of the data mining, which relies on rich, complete data sets (Al-Janabi and Alkaim 2020). Three common methods can be used to handle

J. Constr. Eng. Manage., 2021, 147(8): 04021089

missing values. First, observations that contain missing values in any attribute can be removed. An attribute is still valid when the missing values appear only in a small number of observations. However, when there are missing values in a large number of observations, those observations should be dropped. The disadvantage of this method is that the useful information in the data set may be ignored. Second, the data set can be manually checked by domain experts who determine and enter reasonable and feasible values based on their knowledge or the literature. This is time-consuming in the case of large data sets. Third, missing value imputation can be carried out automatically. Missing value imputation means replacing missing values with estimates based on nonmissing values in other observations in the same attribute. Most often the missing values are replaced with a global constant, the mean or most frequent value of the current attribute, or clustering (Raja and Thangavel 2020). This research handled the missing values by combining clustering and domain knowledge.

Clustering groups similar observations using unsupervised algorithms; the missing values can be filled in based on the cluster information (Raja and Thangavel 2020). Rather than clustering the data set by algorithms, this research used prior knowledge and literature review. For example, it was assumed that the contract period of a PPP project depends on the country that develops the project and the contract type. Therefore, the missing contract period of a BOT project in one country could be replaced with the average contract period of all BOT projects in that country.

Categorical Variables

Categorical variables have more than one possible value, classifying each observation into a particular group. A categorical variable should be transformed before using the machine learning model for analysis. One of the most popular ways to do this is conversion into numerical values and labels (Liu et al. 2019), which applies to ordinal variables with the natural order. For example, the income level of a country is an ordinal variable, which can be low, low-medium, upper-medium, and high.

The second way to transform a categorical variable is suitable for nominal variables. For instance, the primary source of revenue of a PPP project is a nominal variable with several values, such as purchase agreement and user fees.

Values in an ordinal variable are replaced with different numerical values according to their order or rank. For example, four levels in the income variable can be assigned 1, 2, 3, and 4, respectively. The nominal variable can be replaced by a combination of binary variables that take a value of 0 or 1 depending on whether a specific category is present in an observation. For example, project type contains greenfield, brownfield, and management and lease values based on which the Boolean variables "PF18-Greenfield," "PF18-Brownfield," and "PF18-Management & lease" can be generated. If one project is greenfield, the value for "PF18-Greenfield" is 1, while the value for "PF18-Brownfield" and "PF18-Management & lease" is 0.

Data Set Balancing

Data sets may be considerably imbalanced when the number of observations of one class, the majority class, significantly outnumbers that of other classes, the minority classes (Krawczyk et al. 2019). Examples containing imbalanced data sets are credit card fraud detection and medical sciences (Fotouhi et al. 2019; Makki et al. 2019). Identification of minority samples is often of high priority in these problems. For example, because most PPP projects end successfully, the distribution of contract success and failure is imbalanced. Unlike balanced data sets, imbalanced data sets are challenging for almost all machine learning models because the models may neglect the minority samples (Ye et al. 2019). Therefore, these data sets have to be balanced before being analyzed. Currently, the most popular data-level method to handle an imbalanced data set is resampling it into a balanced data set.

In binary classification problems, resampling can be achieved by oversampling and undersampling. Oversampling increases minority observations. Its most representative example is the synthetic minority oversampling technique (SMOTE), which creates new observations by interpolating neighboring minority observations. The main advantage of SMOTE is that it can overcome overfitting to a large extent (Cheng et al. 2019). Undersampling reduces the number of majority samples, but may discard useful information in the majority classes (Zhu et al. 2020).

Data balancing is conducted only on the training data set to avoid data leakage (Schlögl et al. 2019; Zhang et al. 2017). In other words, the testing data set is still imbalanced.

Dimensionality Reduction

Dimensionality reduction is useful when preprocessing the data set with a large number of attributes and observations. It changes the data set from high-dimensional to low-dimensional as a way improve the performance of a machine learning model by eliminating insignificant attributes (Reddy et al. 2020). Principal component analysis (PCA), one of the most popular dimensionality reduction techniques, reduces the dimension of the data set by linearly transforming the interrelated attributes into a new set of uncorrelated principal components (PCs) while retaining as much of the variation and information of the original data set as possible (Abdi and Williams 2010). PCs are obtained by

$$PC_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \tag{1}$$

where i = ith principal component; x_n = nth attribute in the data set; and a_{in} = loading of variable x_n for PC_i.

The importance of each PC depends on the original data set's variation that it retains. As the order of the principal components decrease, retention of the variance by each principal component also decreases. Therefore, the first principal component has the maximum variance and is of the highest importance in determining the result of the dependent variable (Nobre and Neves 2019).

Because the results of PCA can be considerably affected by scale, StandardScaler was used to normalize the attributes before the PCA was fitted.

Machine Learning Model

Logistic Regression

The logistic regression (LR) model predicts the probability of an event (Tien Bui et al. 2019). This probability can be calculated by a sigmoid function, as shown in Eq. (2)

$$P(t) = \frac{e^t}{1 + e^t} \tag{2}$$

where t = linear combination of independent variables of the data set, which can be calculated by

$$t = b_0 + b_1x_1 + \cdots + b_nx_n \tag{3}$$

where x_n = nth attribute in the data set; and b_n = estimated coefficient from the data.

The main advantages of the LR model are that it has no requirement for distribution of the variables in the data set and it is easier and quicker to implement (Tien Bui et al. 2019).

J. Constr. Eng. Manage., 2021, 147(8): 04021089

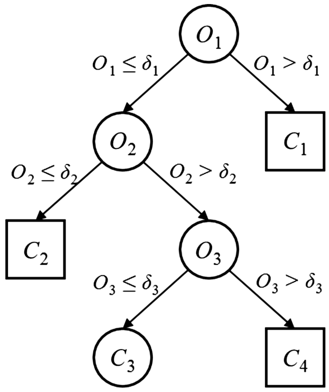

Fig. 1. Decision tree.

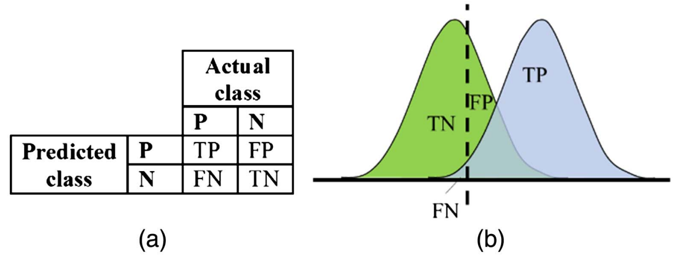

Fig. 2. Confusion matrix and distribution of predicted result (N = negative; and P = positive): (a) confusion matrix; and (b) class distribution.

Support Vector Machine

The super vector machine (SVM) model maps input training data to points in high-dimensional space to maximize the width of the gap between various classes (Ahmad et al. 2018). In a data set with two different classes, denoted (x_n, y_n), the maximum-margin hyperplane will be detected to separate the points x_n into two groups according to their labels. The advantage of the SVM model is that it can handle both a linear and a nonlinear classification problem.

Random Forest

Random forest (RF) is an assembling method for classification and regression that aggregates the results of a series of decision trees (Zhao et al. 2020). A decision tree is composed of inner nodes and leaf nodes, which represent decision thresholds and prediction results, respectively (Liang et al. 2019). Fig. 1 is an example of a decision tree with a depth of 3, where O_1–O_3 are the attributes in the data set, C_1–C_4 are leaf nodes denoting predicted classes, and δ_1–δ_3 are thresholds.

RF has inherited the advantages of decision trees because it grows each tree by selecting random samples for replacement from the whole training data set, but can provide a more precise and stable prediction. RF is robust in handling outliers in the data set and captures the nonlinearities and relationships among high-dimensional variables (Maskey et al. 2020). Optimal hyperparameters are needed to obtain better prediction from a machine learning model. The most critical hyperparameters of the RF model include number and depth of decision trees and minimum number of samples required to split an internal node.

Performance Evaluation

Evaluation metrics are needed to evaluate the performance of machine learning models. For a classification problem, model performance can be assessed by the receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC), recall, precision, the precision-recall (PR) curve, accuracy, and the F_β score. These metrics can be obtained from the confusion matrix, which shows the result of a supervised machine learning model. Fig. 2(a) is a confusion matrix for binary classification, where the rows and columns denote the observations for predicted and actual classes, respectively. P (positive) and N (negative) represent the number of real positive and real negative observations in the data set. TP (true positive) and TN (true negative) represent the number of correctly detected observations in P and N. FP (false positive) represents the number of negative observations incorrectly predicted as positive, and FN (false negative) represents the number of positive observations incorrectly predicted as negative (Chou and Lin 2013).

To separate the positive class from the negative class in binary classification task, a decision threshold is defined. If the predicted probability of an observation is greater than the threshold, the observation is classified as positive and vice versa. Fig. 2(b) shows the distribution of the predicted results, where the dashed line represents the decision threshold. For each threshold, a different confusion matrix is generated. The ROC curve is obtained by plotting the true positive rate (TPR) against the false positive rate (FPR) under different decision thresholds, which can be calculated by Eqs. (4) and (5), respectively (Lin et al. 2017). The area under the ROC curve, AUC, indicates the ability of a model to distinguish different classes. It ranges from 0 to 1, and the higher the AUC, the better the model. In this research, a high AUC meant good performance in differentiating successful and failed PPP contracts. Accuracy, denoting the ratio of correct predictions to total predictions, can be calculated by Eq. (6). However, neither the ROC nor the AUC is sensitive enough to determine class distribution in an imbalanced data set, which may be misleading (Ye et al. 2019)

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

To evaluate the classification performance of imbalanced data sets, recall, precision, the PR curve, and the F_β score may outperform the AUC. The PR curve is obtained by plotting precision against recall under different thresholds. Recall is the same as TPR. Precision is the ratio of correctly predicted positive samples to predicted positive samples, which can be calculated by Eq. (7) (Chou and Lin 2013). In practical situations, a trade-off should be made between precision and recall. The F_β score can combine recall and precision with different weights, which can be calculated by Eq. (8) (Lin et al. 2017). When recall and precision are assigned the same weight, β equals 1.0. Finally, the F_β score turns into the F_1 score, which was used in this research

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$F_\beta = (1 + \beta^2)\frac{precision \times recall}{\beta^2 \times precision + recall} \tag{8}$$

Model Interpretation

The relationship between an independent variable and the dependent variable can be obtained by the correlation coefficient:

Table 1. Attributes of PPI, WGI, and WDI databases and descriptive statistics

No.	Attribute	Data type	Missing value	Source	Minimum	Mean	Maximum
PF1	Contract period (years)	Numeric	Yes	PPI	1.0	25.0	99.0
PF2	Cross-border project	Numeric	No	PPI	1	—	8
PF3	Percentage of private participation	Numeric	Yes	PPI	0	93.05	100.0
PF4	Physical assets ($ millions)	Numeric	Yes	PPI	0	195.9	14,800.0
PF5	Project capacity	Numeric	Yes	PPI	Sector-dependent	Sector-dependent	Sector-dependent
PF6	Project year	Numeric	No	PPI	1990	—	2019
PF7	Total investment ($ millions)	Numeric	Yes	PPI	0.03	228.1	35,586.5
PF8	Bilateral support	Categorical	Yes	PPI	Two values: yes or no	Two values: yes or no	Two values: yes or no
PF9	Government granting contract	Categorical	Yes	PPI	Three values: federal, local, or state/provincial	Three values: federal, local, or state/provincial	Three values: federal, local, or state/provincial
PF10	Lending type	Categorical	Yes	PPI	Four values: IBD, IDX, IDB, or LNX	Four values: IBD, IDX, IDB, or LNX	Four values: IBD, IDX, IDB, or LNX
PF11	Multilateral support	Categorical	No	PPI	Two values: yes or no	Two values: yes or no	Two values: yes or no
PF12	Public disclosure	Categorical	No	PPI	Two values: yes or no	Two values: yes or no	Two values: yes or no
PF13	Renewable	Categorical	No	PPI	Two values: conventional or renewable	Two values: conventional or renewable	Two values: conventional or renewable
PF14	Subsector	Categorical	No	PPI	Twelve values: electricity, etc.	Twelve values: electricity, etc.	Twelve values: electricity, etc.
PF15	Segment	Categorical	No	PPI	Thirty-three values: highway, etc.	Thirty-three values: highway, etc.	Thirty-three values: highway, etc.
PF16	Revenue source	Categorical	Yes	PPI	Five values: user fees, etc.	Five values: user fees, etc.	Five values: user fees, etc.
PF17	Project subtype	Categorical	No	PPI	Ten values: BOT, etc.	Ten values: BOT, etc.	Ten values: BOT, etc.
PF18	Project type	Categorical	No	PPI	Three values greenfield, etc.	Three values greenfield, etc.	Three values greenfield, etc.
PF19	Unsolicited proposal	Categorical	No	PPI	Two values: yes or no	Two values: yes or no	Two values: yes or no
PF20	Sponsor number	Numeric	Yes	PPI	1	2	13
PF21	Foreign countries	Categorical	Yes	PPI	Two values: yes or no	Two values: yes or no	Two values: yes or no
CF1	Overall country experience	Numeric	No	PPI	0	360	1,582
CF2	Country sector experience	Numeric	No	PPI	0	127	611
CF3	CPI 2019	Numeric	No	PPI	1.00	1.22	1.92
CF4	Country IDA status	Categorical	No	PPI	Three values: IDA, blended, or non-IDA	Three values: IDA, blended, or non-IDA	Three values: IDA, blended, or non-IDA
CF5	Region by income level	Categorical	No	PPI	Six values: such as LAC, etc.	Six values: LAC, etc.	Six values: LAC, etc.
CF6	Administrative region	Categorical	Yes	PPI	Six values: MENA, etc.	Six values: MENA, etc.	Six values: MENA, etc.
CF7	Income level	Categorical	Yes	PPI	Four values: high income, etc.	Four values: high income, etc.	Four values: high income, etc.
CF8	Control of corruption	Numeric	Yes	WGI	−1.72	−0.38	1.03
CF9	Voice and accountability	Numeric	Yes	WGI	−2.20	−0.38	1.14
CF10	Rule of law	Numeric	Yes	WGI	−2.42	−0.32	1.08
CF11	Regulatory quality	Numeric	Yes	WGI	−2.45	−0.15	1.13
CF12	Political stability & absence of violence	Numeric	Yes	WGI	−2.86	−0.60	1.28
CF13	Government effectiveness	Numeric	Yes	WGI	−2.22	−0.09	1.27
CF14	Real effective exchange	Numeric	Yes	WDI	31.7	96.7	388.1
CF15	Tax as percentage of GDP	Numeric	Yes	WDI	0.89	12.29	110.18
CF16	Exports as percentage of GDP	Numeric	Yes	WDI	0.18	25.93	156.54
CF17	Imports as percentage of GDP	Numeric	Yes	WDI	0.10	27.65	191.46
CF18	Domestic credit to private sector percentage of GDP)	Numeric	Yes	WDI	0.78	65.40	158.50
CF19	Electric power consumption (kWh per capita)	Numeric	Yes	WDI	13.5	1,825.7	6,687.7
CF20	Life expectancy (years)	Numeric	Yes	WDI	46.2	73.3	82.0
CF21	GDP deflator	Numeric	Yes	WDI	−26.3	30.0	26,765.9
CF22	GDP per capita ($)	Numeric	Yes	WDI	102.6	4,510.3	15,974.6

Note: IDA = International Development Association; LAC = Latin America and Caribbean; MENA = Middle East and North Africa; and CPI = consumer price index.

a positive correlation coefficient means that the dependent variable tends to increase when one independent variable increases (Xu and Deng 2017). When predicting PPP contract failure, if failed projects are labeled 1 and successful ones are labeled 0, the increase in the dependent variable means that the status of the project changes from 0 (successful) to 1 (fail). Therefore, if an independent variable is positively related to contract failure, it is categorized as a failure factor.

To find the independent variables that make a considerable contribution to the differentiation of various classes in the dependent variable, feature importance can be used as a criterion. Some studies have determined feature importance by correlation coefficients, where the independent variables having little correlation with the dependent variable are neglected. However, the correlation coefficient only looks for linear correlations without capturing the combined effect of various factors on the target (Mangal and Holm 2018). A low correlation coefficient only means that there is no obvious linear relationship between the independent variable and the dependent variable; it does not indicate that they are not related (Schober et al. 2018). Therefore, the feature importance of an independent variable is measured based on its contribution to prediction, which can be obtained according to its loading in the first principal component of the PCA (Holland 2008). In contract failure prediction, the importance of each feature is interpreted as the ranking of the failure factor. The higher the variable importance of a failure factor, the more important the factor.

Case Study

Data Preprocessing

Three public databases were analyzed to extract the factors influencing the final status of a PPP project. Twenty-one project-specific and 22 country-specific factors were extracted. Table 1 summarizes the original attributes and their descriptive statistics, where "Missing value" denotes whether the attribute contains missing values and "Data type" gives the value type of the attribute. In this research, project-specific and country-specific factors were denoted "PF" and "CF," respectively. These were the independent variables, and the final status of the project was the dependent variable. Table 2 shows the distribution of project status before and after class balancing in the training data set and in the testing data set.

Table 1 indicates that 21 project-specific factors and 7 country-specific factors were extracted from the PPI database after removing duplicate features and features with more than 30% missing values. These factors were divided into numerical and categorical types. PF1–PF19 and CF3–CF7 were extracted from the PPI database directly, while PF20, PF21, CF1, and CF2 were calculated by feature engineering based on other factors. They were obtained via the following procedure:

- Sponsor number (PF20) and foreign countries (PF21) were extracted from the variable "description" in the PPI database. PF20 was the number of project sponsors, and PF21 denoted foreign country involvement.
- Overall country experience (CF1) and country sector experience (CF2) were calculated based on the yearly number of PPP projects. In this research, overall country experience in each year referred to the number of PPP projects completed by that country. Similarly, yearly country sector experience was PPP experience in the different sectors in that country.

The number of factors obtained from the WGI and WDI databases was 6 and 9, respectively. Moreover, 13 factors from the PPI database and all factors from the WGI and WDI databases had missing values, which had to be imputed before analysis. As previously mentioned, the numerical attributes were replaced by the mean of existing values in the same cluster while the missing values in the categorical attributes were replaced by the mode in the same cluster. The specific methods were these:

- Imputing missing numerical values. PF1 and PF5 were assumed to depend on country, region by income level (CF5), and project subtype (PF17). PF3, PF4, and PF7 were assumed to depend on country, subsector (PF14), and project subtype (PF17). PF20 was assumed to depend on country and subsector (PF14). CF10 and CF14–CF19 were assumed to depend on region by income level (CF5).
- Imputing missing categorical values. PF8–PF10, PF16, and PF21 were assumed to depend on country and subsector (PF14). CF6 and CF7 were assumed to depend on region by income level (CF5).

Final project status from the PPI database was active, concluded, canceled, or distressed. Active and concluded projects were ongoing and finished, respectively. Following previous research, both canceled and distressed contracts were treated as failed while others were treated as successful (Harris and Pratap 2009). In this research, contract failure in the three PPP sectors that contained failed projects—transportation, water and sewer, and energy—, was predicted. The data set for each sector was randomly split into training (75%) and testing (25%) data sets Table 2 shows that the ratio between number of contract failures and total projects in the training data set for each sector was 5.99%, 5.83%, and 2.26%, respectively, indicating the large skewness in class distribution. As the main objective of this study was to identify failed projects, these projects were labeled 1 (positive) and successful projects were labeled 0 (negative). Moreover, data balancing, data normalization, and PCA were used to convert categorical data into labels.

Model Results and Performance

The LR, SVM, and RF models trained the training data set for each sector via different data-balancing techniques, as suggested by Horta and Camanho (2013). They were verified by the testing data set and the best combination for each sector was chosen. Table 3 lists the AUC and the area under the PR curve (AUPRC) for different combinations, with boldface indicating the best performing AUC and underscoring indicating the best performing AUPRC. Figs. 3–5 show, respectively, the ROC curve, the PR curve where the corresponding y value (dashed line) is the ratio of actual positive samples to actual negative samples, and the confusion matrix of the best model. Table 4 gives the primary evaluation metrics.

Table 2. Project status distribution by PPP sector

	Training data set (original)			Training data set (after SMOTE)			Testing data set		
Sector	Total	Failed	Ratio (%)	Total	Failed	Ratio (%)	Total	Failed	Ratio (%)
Transportation	1,419	85	5.99	2,668	1,334	50	474	29	6.12
Water and sewer	789	46	5.83	1,486	743	50	263	15	5.70
Energy	2,699	61	2.26	5,276	2,638	50	900	21	2.33

J. Constr. Eng. Manage., 2021, 147(8): 04021089

Table 3. Model performance by sector

Sector	Data-balancing technique	LR model AUC	LR model AUPRC	SVM model AUC	SVM model AUPRC	RF model AUC	RF model AUPRC
Transportation	SMOTE	0.81	0.170	0.73	0.153	**0.91**	0.470
Transportation	Undersampling	0.83	0.204	0.34	0.044	0.85	0.376
Water and sewer	SMOTE	0.82	0.324	0.80	0.186	**0.85**	0.456
Water and sewer	Undersampling	0.74	0.110	0.83	0.330	0.80	0.178
Energy	SMOTE	0.86	0.165	0.85	0.109	0.85	0.122
Energy	Undersampling	0.85	0.108	0.76	0.154	**0.87**	0.106

Note: Bold and underlined values are the largest AUC and the area under the PR curve for each sector. They correspond to the best combination from the perspective of AUC and the area under the PR curve.

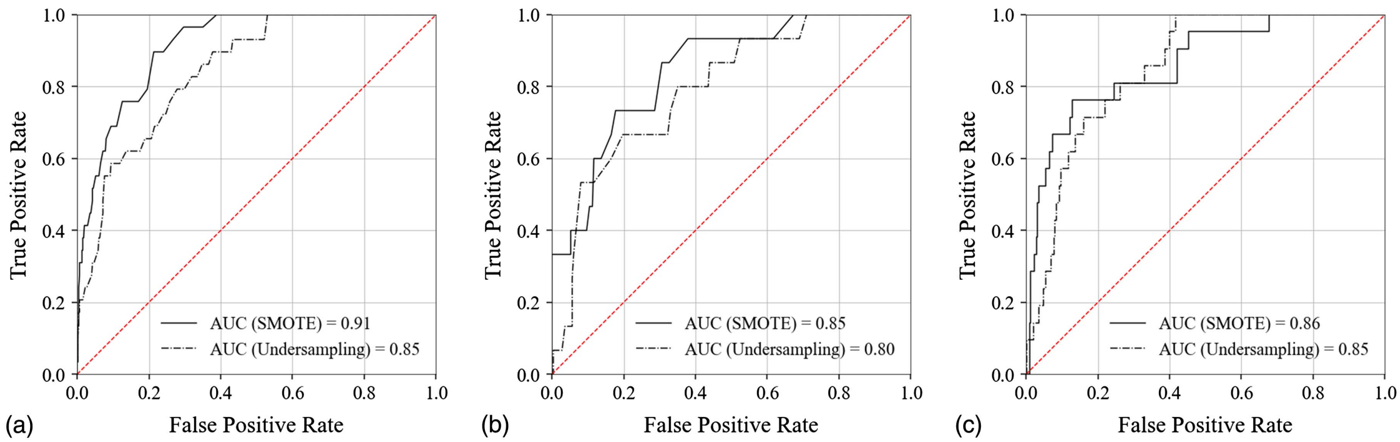

Fig. 3. ROC curve and AUC for selected model using different balancing techniques: (a) transportation sector; (b) water & sewage sector; and (c) energy sector.

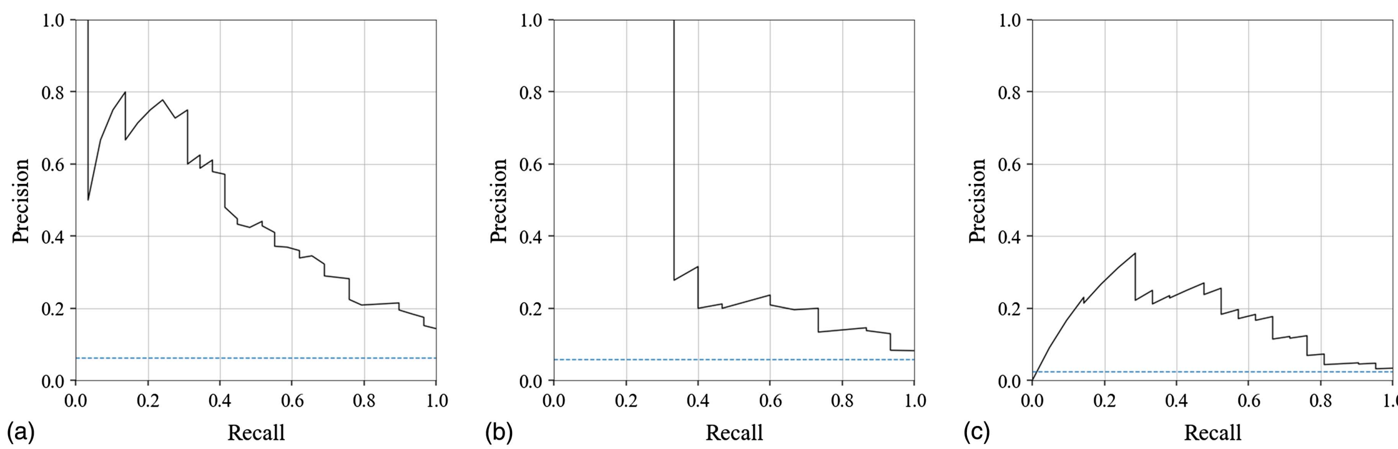

Fig. 4. PR curve for selected model: (a) transportation sector; (b) water & sewage sector; and (c) energy sector.

The performance of a model can vary greatly in different contexts. Table 3 indicates that the best combination for the transportation, water and sewer, and energy sectors was, respectively RF +SMOTE, RF+SMOTE, and RF+Undersampling from the AUC perspective, and, respectively, RF+SMOTE, RF+SMOTE, and LR+SMOTE from AUPRC perspective. However, as previously mentioned, because the ROC may be misleading in an imbalanced data set, the AUPRC is used to select the best combination in this context. The primary hyperparameters of each selected model were (1) 100 for "n_estimators," 2 for "min_samples_leaf," and "gini" for "criterion"; (2) 120 for the "n_estimators," 3 for "min_samples_leaf," and "gini" for "criterion"; and (3) 100 for "max_iter" and "l2" for "penalty."

Fig. 3 shows that the AUC of the selected model was 0.91, 0.85, and 0.86, respectively. According to Yang and Berdine (2017), an AUC larger than 0.8 has outstanding discrimination. Bauder and Khoshgoftaar (2018) reported the AUC for an imbalanced data set with 95:5 class distribution as 0.87. These findings indicate that the model can separate failed and successful projects appropriately. Fig. 4 indicates that an increase in recall can lead to a drop in precision, which depends on the position of the decision threshold. Low recall means that most contract failures cannot be correctly forecast, and low precision means that some successful projects are identified as failed. Therefore, an appropriate threshold should be selected based on the practical problem. As the main aim of the proposed model was to facilitate the feasibility study of a PPP

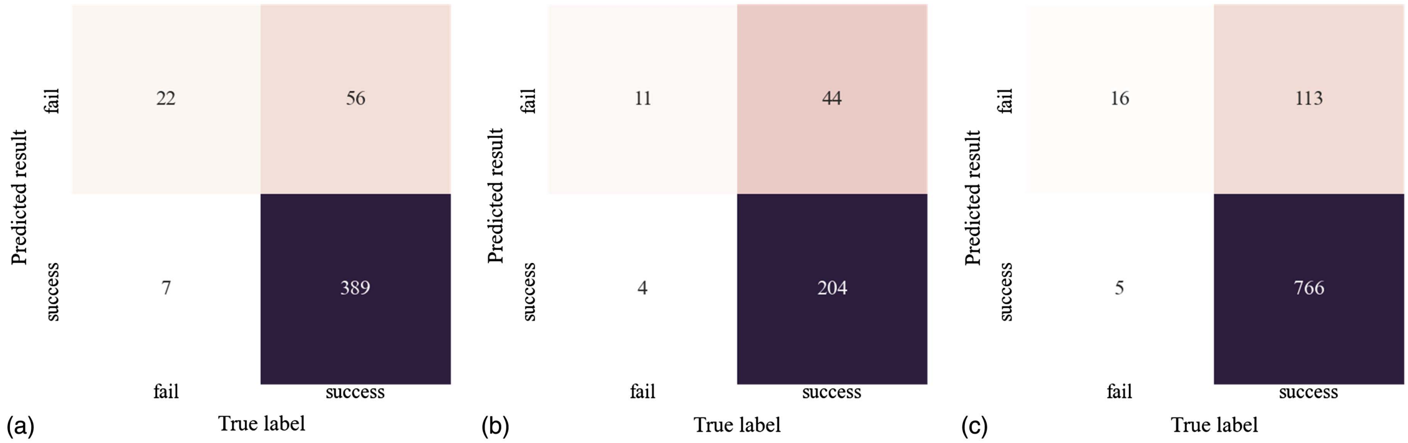

Fig. 5. Confusion matrix for selected model: (a) transportation sector; (b) water & sewage sector; and (c) energy sector.

Table 4. Performance metrics for selected model by sector

Sector	AUC	Recall (%)	Accuracy (%)
Transportation	0.91	75.9	86.7
Water and sewer	0.85	73.3	81.7
Energy	0.86	76.2	86.9

project, as many projects with a high probability of failure as possible should be detected to avoid investment failure. In other words, high recall was expected in this context. After fine-tuning the model, the cut-off point for the three sectors was located at 21%, 21%, and 60%, respectively. That is to say, any project having a failure probability larger than these thresholds would be predicted to fail. From the perspective of project selection, these projects were not recommended for private sectors with low risk-bearing ability and limited resources (Armaneri et al. 2010).

As the imbalance ratio rises, the performance of a model is negatively affected. The best AUC in the transportation sector was the highest (0.91) but in the water and sewer sector was the lowest (0.85). As the imbalance ratio of these two sectors was similar, the reason for the low AUC in the water and sewer sector may have been a smaller sample size. Although the energy sector had the largest number of projects, the best AUPRC was only 0.165, which was much less than that of the other sectors (more than 0.45). This may have been because of its lower imbalance ratio (2.26%) when compared with the other sectors (more than 5.8%). Therefore, to improve the performance of the machine learning model in this context, more projects would need to be collected and those with too many missed variables would need to be removed to increase the ratio of failed projects.

Failure Factors

To obtain a clear view of how different factors influence the final status of a PPP project, the correlation coefficients between each independent variable and the dependent variable were calculated. Table 5 shows the top 15 failure factors with the largest feature importance and their correlation with contract failure. It can be seen that there were some common important failure factors in the three PPP sectors, which could be explained as follows:

- CPI 2019 (CF3). CPI was used to measure the risk of inflation or deflation; a high CPI meant low services and goods purchasing power (Sun et al. 2019). Inflation can erode the real value of debt and thus impact PPP projects by increasing material, labor, equipment hiring, and consultant services costs. It may lead to a high risk of cost overruns during the construction phase (Musarat et al. 2020). Moreover, for projects that rely on end-user fees, such as water infrastructure, inflation may lower the private partner's operating margins in the operation phase. As it

Table 5. Failure factor importance by sector

No.	Transportation		Water and sewer		Energy	
	Failure factor	Correlation	Failure factor	Correlation	Failure factor	Correlation
1	CF21	0.058	CF17	0.232	CF21	0.043
2	CF14	0.047	CF16	0.192	CF3	0.596
3	PF3	0.075	CF3	0.606	PF20	0.243
4	PF20	0.067	PF7	0.400	PF1	0.123
5	PF4	0.061	PF16-other	0.374	CF7	0.084
6	PF5	0.008	CF14	0.065	PF5	0.353
7	PF1	0.063	PF1	0.025	PF7	0.270
8	CF17	0.186	PF5	0.429	PF9-federal	0.041
9	CF3	0.339	PF4	0.189	PF16-other	0.368
10	CF16	0.137	CF15	0.033	PF17-BROT	0.141
11	CF7	0.018	CF5-AFR	0.016	CF6-SSA	0.013
12	PF16-other	0.127	CF6-SSA	0.016	CF5-AFR	0.010
13	CF15	0.065	PF11-yes	0.131	PF19_No	0.095
14	PF9-federal	0.179	PF9-state/provincial	0.072	PF8_No	0.085
15	CF5-LAC	0.080	PF20	0.288	PF12_No	0.129

is less likely for the public partner to increase user fees to fill the service price gap due to inflation, a project may have to be terminated.

- Sponsor number (PF20). Sponsors of a PPP project can bring various technologies and knowledge to a project, which can be important to project success. However, the rise in sponsors, priorities, responsibilities, and stakeholder capabilities may become a source of disputes and conflicts of interest (Shrestha et al. 2017). If interests are not well balanced, enhanced cooperation among sponsors becomes unachievable, increasing the chance of contract failure. A larger number of sponsors also indicates more transaction costs related to coordination.
- Total investment (PF7) and project capacity (PF5). An increase in project scale may make a project more complicated and risky by increasing the strain on resources and private partner experience (Schleifer et al. 2014).
- Contract period (PF1). A long contract period can make a PPP project riskier. This is in contrast to the claims of some studies that a longer PPP contract correlates with better performance (Klijn and Koppenjan 2016), but in keeping with the findings of others (Wang et al. 2018). A long-term contract can be inflexible in the face a variety of unforeseen and unexpected events that may have a negative influence on value for money, such as changes in user demands, competition from new and cheaper products, and fluctuatinginflation rates. These events can have severe repercussions for both public and private partners, particularly in less experienced countries, if provisions for change have not been properly set in the contract.

Each sector has its own failure factors. The water and sewer and energy sectors, for example, are more sensitive to country income level., Projects based in Sub-Saharan Africa (the AFR region), which is low-income, are more likely to fail.

Limitations and Future Work

There were some limitations to this research. First, ongoing projects were treated as successful, even though their final status could not be known. Second, model performance needs to be improved. Third, the PPI database only includes large projects, excluding those in health, education, prisons, and simple government buildings (Loxley 2013). This may have influenced the importance of failure factors.

Future work should improve the results of this research. First, neural networks will be used to predict project status considering the time factor. Second, more PPP projects and variables will be collected from other databases to increase model power. Third, cost-sensitive learning will be used to lower the influence of class imbalance.

Conclusions

Engaging the private sector in public infrastructure can lower the financial burdens of government entities and provide better services. Although PPP models are used around the world, PPP contract failure is still a problem, especially in developing countries, which are faced with unpredictable environments and shortages of management experience (Shrestha et al. 2017). To enable PPP practitioners to analyze the viability of a PPP project, this research used machine learning models to predict the final success or failure of a PPP project in different sectors. The conclusions of this research are as follows:

- Different combinations of machine learning models and balancing techniques perform differently in different PPP sectors. From the AUPRC perspective, RF+SMOTE, RF+SMOTE, and LR+SMOTE rank highest, respectively, in transportation, water and sewer, and energy.
- Recall of 75.9%, 73.3%, and 76.2% was achieved by setting respective decision thresholds of 21%, 21%, and 60% for transportation, water and sewer, and energy.
- Results for the relationship between each factor and contract status show that factors correlate differently with PPP contract failure. Factors positively that correlate are identified as failure factors, such as CPI, total investment, and project capacity.
- The importance of each failure factor is based on variable importance in the PCA. Top-ranked failure factors vary with PPP sector.
- A high correlation coefficient between an independent variable and the dependent variable does not necessarily indicate high importance of a failure factor. The variable can be important even though the correlation coefficient is low.

Decision makers can evaluate the initial feasibility of a project by predicting its final status using the appropriate machine learning model. First, by referring to the revealed failure factors, they can adjust project variables to minimize the probability of failure. Second, understanding the importance of the failure factors can give them a clearer view of risks, which will make their bargaining position better able to avoid overpromised guarantees and supports offered by private partners. In this way, the interests of taxpayers are protected and they gain value for money. Finally, guidelines for risk mitigation and monitoring can be established to improve domestic project performance and encourage more private investment in PPP projects.

The private sector can select projects based on detected failure factors, such as projects sponsored by nonfederal governments. This may lower the probability of early termination, especially in unfamiliar countries. On the other hand, sound measures can be taken to mitigate the risks introduced by high-ranked failure factors. For example, inflation risk can be neutralized by indexation. Finally, the private partners can determine whether to invest in projects with different failure probabilities according to their risk-bearing ability.

The findings of this study contribute to the field of construction management in three ways. First, PPP contract failure has been shown to be predictable. It usually occurs after negotiation between public and private partners. Although disputes (Chou and Lin 2013) and possible risks (Owolabi et al. 2018; Wan and Fang 2020) can be predicted, until now outright failure, which can have worse repercussions, was not predictable based on the outcome of previous projects. Second, the importance of individual failure factors is now better understood. Although some studies have tried to find the critical failure factors for PPP projects, only qualitative methods such as questionnaires and surveys have been used (Dolla and Laishram 2020; Song et al. 2018; Tariq and Zhang 2020). This research adopted a quantitative perspective based on real-world projects. Finally, the performance of machine learning models in predicting PPP project outcomes in different sectors has been compared, providing a pathway to selection of the most appropriate machine learning models.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

References

Abdi, H., and L. Williams. 2010. "Principal component analysis." *Wiley Interdiscip. Rev. Comput. Stat.* 2 (4): 433–459. https://doi.org/10.1002/wics.101.

Abdullah, A., and I. Khadaroo. 2020. "The trust-control nexus in public private partnership (PPP) contracts." *J. Accounting Public Policy* 39 (6): 106768. https://doi.org/10.1016/j.jaccpubpol.2020.106768.

Ahmad, I., M. Basheri, M. J. Iqbal, and A. Rahim. 2018. "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection." *IEEE Access* 6 (Jul): 33789–33795. https://doi.org/10.1109/ACCESS.2018.2841987.

Ahmadabadi, A. A., and G. Heravi. 2019. "The effect of critical success factors on project success in public-private partnership projects: A case study of highway projects in Iran." *Transp. Policy* 73 (Jan): 152–161. https://doi.org/10.1016/j.tranpol.2018.07.004.

Akbari, S., F. P. Rahimian, M. Sheikhkhoshkar, S. Banihashemi, and M. Khanzadi. 2020. "Dynamic sustainable success prediction model for infrastructure projects: A rough set based fuzzy inference system." *Constr. Innovation* 20 (4): 545–567. https://doi.org/10.1108/CI-04-2019-0034.

Al-Janabi, S., and A. F. Alkaim. 2020. "A nifty collaborative analysis to predict a novel tool (DRFLLS) for missing values estimation." *Soft Comput.* 24 (1): 555–569. https://doi.org/10.1007/s00500-019-03972-x.

Armaneri, Ö., G. Özdağoğlu, and Ö. Yalçınkaya. 2010. "An integrated decision support approach for project investors in risky and uncertain environments." *J. Comput. Appl. Math.* 234 (8): 2530–2542. https://doi.org/10.1016/j.cam.2010.03.018.

Bauder, R., and T. Khoshgoftaar. 2018. "Medicare fraud detection using random forest with class imbalanced big data." In *Proc., 2018 IEEE Int. Conf. on Information Reuse and Integration (IRI)*, 80–87. New York: IEEE.

Cheng, K., C. Zhang, H. Yu, X. Yang, H. Zou, and S. Gao. 2019. "Grouped SMOTE with noise filtering mechanism for classifying imbalanced data." *IEEE Access* 7 (Dec): 170668–170681. https://doi.org/10.1109/ACCESS.2019.2955086.

Chou, J.-S., and C. Lin. 2013. "Predicting disputes in public-private partnership projects: Classification and ensemble models." *J. Comput. Civ. Eng.* 27 (1): 51–60. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000197.

Demirel, H. Ç., L. Volker, W. Leendertse, and M. Hertogh. 2019. "Dealing with contract variations in PPPs: Social mechanisms and contract management in infrastructure projects." *J. Constr. Eng. Manage.* 145 (11): 04019073. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001714.

Díaz, G. R. 2020. "What drives the failure of private participation in infrastructure projects?" *Int. J. Managing Projects Bus.* 13 (6): 1167–1185. https://doi.org/10.1108/IJMPB-12-2019-0298.

Dolla, T., and B. Laishram. 2020. "Factors affecting public-private partnership preference in Indian municipal waste sector." *Int. J. Constr. Manage.* 20 (6): 567–584. https://doi.org/10.1080/15623599.2019.1703085.

Fotouhi, S., S. Asadi, and M. W. Kattan. 2019. "A comprehensive data level analysis for cancer diagnosis on imbalanced data." *J. Biomed. Inf.* 90 (Feb): 103089. https://doi.org/10.1016/j.jbi.2018.12.003.

Harris, C., and K. V. Pratap. 2009. "What drives private sector exit from infrastructure? Economic crises and other factors in the cancellation of private infrastructure projects in developing countries." Accessed May 23, 2020. https://openknowledge.worldbank.org/handle/10986/10569.

Holland, S. M. 2008. *Principal components analysis (PCA)*. Athens, GA: Univ. of Georgia.

Horta, I. M., and A. S. Camanho. 2013. "Company failure prediction in the construction industry." *Expert Syst. Appl.* 40 (16): 6253–6257. https://doi.org/10.1016/j.eswa.2013.05.045.

Hsueh, C.-M., and L.-M. Chang. 2017. "Critical success factors for PPP infrastructure: Perspective from Taiwan." *J. Chin. Inst. Eng.* 40 (5): 370–377. https://doi.org/10.1080/02533839.2017.1335619.

Jiang, X., K. Lu, B. Xia, Y. Liu, and C. Cui. 2019. "Identifying significant risks and analyzing risk relationship for construction PPP projects in China using integrated FISM-MICMAC approach." *Sustainability* 11 (19): 5206. https://doi.org/10.3390/su11195206.

Klijn, E. H., and J. Koppenjan. 2016. "The impact of contract characteristics on the performance of public–private partnerships (PPPs)." *Public Money Manage.* 36 (6): 455–462. https://doi.org/10.1080/09540962.2016.1206756.

Krawczyk, B., M. Koziarski, and M. Woźniak. 2019. "Radial-based oversampling for multiclass imbalanced data classification." *IEEE Trans. Neural Networks Learn. Syst.* 31 (8): 2818–2831. https://doi.org/10.1109/TNNLS.2019.2913673.

Kwofie, T. E., S. Afram, and E. Botchway. 2016. "A critical success model for PPP public housing delivery in Ghana." *Built Environ. Project Asset Manage.* 6 (1): 58–73. https://doi.org/10.1108/BEPAM-04-2014-0026.

Liang, J., Z. Qin, S. Xiao, L. Ou, and X. Lin. 2019. "Efficient and secure decision tree classification for cloud-assisted online diagnosis services." *IEEE Trans. Dependable Secure Comput.* https://doi.org/10.1109/TDSC.2019.2922958.

Lin, W., Z. Wu, L. Lin, A. Wen, and J. Li. 2017. "An ensemble random forest algorithm for insurance big data analysis." *IEEE Access* 5 (Aug): 16568–16575. https://doi.org/10.1109/ACCESS.2017.2738069.

Liu, X., H. Huang, Z. Tang, T. Zhang, and B. Yang. 2019. "Sparse block regression (SBR) for big data with categorical variables." In *Proc., 2019 IEEE Int. Conf. on Big Data (Big Data)*, 221–226. New York: IEEE.

Loxley, J. 2013. "Are public-private partnerships (PPPs) the answer to Africa's infrastructure needs?" *Rev. Afr. Political Econ.* 40 (137): 485–495. https://doi.org/10.1080/03056244.2013.817091.

Makki, S., Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid, and H. Zeineddine. 2019. "An experimental study with imbalanced classification approaches for credit card fraud detection." *IEEE Access* 7 (Jul): 93010–93022. https://doi.org/10.1109/ACCESS.2019.2927266.

Mangal, A., and E. A. Holm. 2018. "A comparative study of feature selection methods for stress hotspot classification in materials." *Integrating Mater. Manuf. Innovation* 7 (3): 87–95. https://doi.org/10.1007/s40192-018-0109-8.

Mansaray, A. A. 2018. *Public-private partnership: Countries' attractiveness and the risk of project failure*. Loughborough, UK: Loughborough Univ.

Marcelo, D., S. House, C. Mandri-Perrott, and J. Schwartz. 2017. *Do countries learn from experience in infrastructure PPP? PPP practice and contract cancellation*. Washington, DC: The World Bank.

Marques, R. C. 2018. "Is arbitration the right way to settle conflicts in PPP arrangements?" *J. Manage. Eng.* 34 (1): 05017007. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000564.

Maskey, S. R., S. Badsha, S. Sengupta, and I. Khalil. 2020. "Bits: Blockchain based intelligent transportation system with outlier detection for smart city." In *Proc., 2020 IEEE Int. Conf. on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 1–6. New York: IEEE.

Musarat, M. A., W. S. Alaloul, and M. S. Liew. 2020. "Impact of inflation rate on construction projects budget: A review." *Ain Shams Eng. J.* 12 (1): 407–414. https://doi.org/10.1016/j.asej.2020.04.009.

Ng, S. T., Y. M. Wong, and J. M. Wong. 2010. "A structural equation model of feasibility evaluation and project success for public–private partnerships in Hong Kong." *IEEE Trans. Eng. Manage.* 57 (2): 310–322. https://doi.org/10.1109/TEM.2009.2037142.

Nobre, J., and R. F. Neves. 2019. "Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets." *Expert Syst. Appl.* 125 (Jul): 181–194. https://doi.org/10.1016/j.eswa.2019.01.083.

Osei-Kyei, R., and A. P. Chan. 2019. "Model for predicting the success of public-private partnership infrastructure projects in developing countries: A case of Ghana." *Archit. Eng. Des. Manage.* 15 (3): 213–232. https://doi.org/10.1080/17452007.2018.1545632.

Owolabi, H. A., M. Bilal, L. O. Oyedele, H. A. Alaka, S. O. Ajayi, and O. O. Akinade. 2018. "Predicting completion risk in PPP projects using big data analytics." *IEEE Trans. Eng. Manage.* 67 (2): 430–453. https://doi.org/10.1109/TEM.2018.2876321.

Raja, P. S., and K. Thangavel. 2020. "Missing value imputation using unsupervised machine learning techniques." *Soft Comput.* 24 (6): 4361–4392. https://doi.org/10.1007/s00500-019-04199-6.

Rajula, H. S. R., G. Verlato, M. Manchia, N. Antonucci, and V. Fanos. 2020. "Comparison of conventional statistical methods with machine learning in medicine: Diagnosis, drug development, and treatment." *Medicina (Kaunas)* 56 (9): 455. https://doi.org/10.3390/medicina56090455.

Reddy, G. T., M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker. 2020. "Analysis of dimensionality reduction techniques on big data." *IEEE Access* 8 (Mar): 54776–54788. https://doi.org/10.1109/ACCESS.2020.2980942.

Schleifer, T., K. Sullivan, and J. Murdough. 2014. *Increase in project size*. Hoboken, NJ: Wiley.

Schlögl, M., R. Stütz, G. Laaha, and M. Melcher. 2019. "A comparison of statistical learning methods for deriving determining factors of accident occurrence from an imbalanced high resolution dataset." *Accid. Anal. Prev.* 127 (Jun): 134–149. https://doi.org/10.1016/j.aap.2019.02.008.

Schober, P., C. Boer, and L. A. Schwarte. 2018. "Correlation coefficients: Appropriate use and interpretation." *Anesth. Analg.* 126 (5): 1763–1768. https://doi.org/10.1213/ANE.0000000000002864.

Shrestha, A., T.-K. Chan, A. A. Aibinu, C. Chen, and I. Martek. 2017. "Risks in PPP water projects in China: Perspective of local governments." *J. Constr. Eng. Manage.* 143 (7): 05017006. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001313.

Song, J., Y. Hu, and Z. Feng. 2018. "Factors influencing early termination of PPP projects in China." *J. Manage. Eng.* 34 (1): 05017008. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000572.

Sun, Q., X. Gao, Z. Wang, S. Liu, S. Guo, and Y. Li. 2019. "Quantifying the risk of price fluctuations based on weighted Granger causality networks of consumer price indices: Evidence from G7 countries." *J. Econ. Interaction Coordination* 15 (4): 821–844. https://doi.org/10.1007/s11403-019-00273-2.

Tariq, S., and X. Zhang. 2020. "Critical failure drivers in international water PPP projects." *J. Infrastruct. Syst.* 26 (4): 04020038. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000581.

Tariq, S., and X. Zhang. 2021. "Critical analysis of the private sector roles in water PPP Failures." *J. Constr. Eng. Manage.* 147 (4): 04021015. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002020.

Tien Bui, D., K. Khosravi, H. Shahabi, P. Daggupati, J. F. Adamowski, A. M. Melesse, B. Thai Pham, H. R. Pourghasemi, M. Mahmoudi, and S. Bahrami. 2019. "Flood spatial modeling in northern Iran using remote sensing and GIS: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model." *Rem. Sens.* 11 (13): 1589. https://doi.org/10.3390/rs11131589.

Tiong, R. L. 1996. "CSFs in competitive tendering and negotiation model for BOT projects." *J. Constr. Eng. Manage.* 122 (3): 205–211. https://doi.org/10.1061/(ASCE)0733-9364(1996)122:3(205).

Wan, A., and J. Fang. 2020. "Risk prediction of expressway PPP project based on PSO-SVM algorithm." In *Proc., ICCREM 2020: Intelligent Construction and Sustainable Buildings*, 55–63. Reston, VA: ASCE.

Wang, H., B. Chen, W. Xiong, and G. Wu. 2018. "Commercial investment in public-private partnerships: The impact of contract characteristics." *Policy Polit.* 46 (4): 589–606. https://doi.org/10.1332/030557318X15200933925414.

Xu, H., and Y. Deng. 2017. "Dependent evidence combination based on Shearman coefficient and Pearson coefficient." *IEEE Access* 6 (Dec): 11634–11640. https://doi.org/10.1109/ACCESS.2017.2783320.

Yang, S., and G. Berdine. 2017. "The receiver operating characteristic (ROC) curve." *Southwest Respir. Crit. Care Chron.* 5 (19): 34–36. https://doi.org/10.12746/swrccc.v5i19.391.

Ye, W., Y. Lin, M. Li, Q. Liu, and D. Z. Pan. 2019. "LithoROC: Lithography hotspot detection with explicit ROC optimization." In *Proc., 24th Asia and South Pacific Design Automation Conf.*, 292–298. New York: Association for Computing Machinery.

Ye, X., S. Shi, H.-Y. Chong, X. Fu, L. Liu, and Q. He. 2018. "Empirical analysis of firms' willingness to participate in infrastructure PPP projects." *J. Constr. Eng. Manage.* 144 (1): 04017092. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001404.

Zhang, L., Z. Sheng, and N. A. Tao. 2016. "Financial system risk tolerance capacity and economic growth: Evidence from a cross-country analysis." *Global Econ. Rev.* 45 (2): 97–115. https://doi.org/10.1080/1226508X.2015.1124343.

Zhang, X., Z. Fang, Y. Wen, Z. Li, and Y. Qiao. 2017. "Range loss for deep face recognition with long-tailed training data." In *Proc., IEEE Int. Conf. on Computer Vision*, 5409–5418.

Zhang, X., and W. Xiong. 2015. "Renegotiation and early-termination in public-private partnerships." *Int. J. Archit. Eng. Constr.* 4 (4): 204–213. https://doi.org/10.7492/IJAEC.2015.021.

Zhao, X., R. Lovreglio, and D. Nilsson. 2020. "Modelling and interpreting pre-evacuation decision-making using machine learning." *Autom. Constr.* 113 (May): 103140. https://doi.org/10.1016/j.autcon.2020.103140.

Zhu, T., Y. Lin, and Y. Liu. 2020. "Improving interpolation-based oversampling for imbalanced data learning." *Knowl.-Based Syst.* 187 (Jan): 104826. https://doi.org/10.1016/j.knosys.2019.06.034.