Article

The Adoption of a Big Data Approach Using Machine Learning to Predict Bidding Behavior in Procurement Management for a Construction Project

Wuttipong Kusonkhum [1], Korb Srinavin [2,*] and Tanayut Chaitongrat [3]

[1] Department of Civil Engineering, Northeastern University, Khon Kaen 40000, Thailand; wuttipong.kus@neu.ac.th
[2] Department of Civil Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
[3] Construction and Project Management Center, Faculty of Architecture, Urban Design and Creative Arts, Mahasarakham University, Maha Sarakham 44150, Thailand; tanayut.c@msu.ac.th
[*] Correspondence: korbsri@kku.ac.th; Tel.: +66-894177357

Abstract: Big data technologies are disruptive technologies that affect every business, including those in the construction industry. The Thai government has also been affected and attempted to use machine learning techniques with the analytics of big data technologies to predict which construction projects have a winning price over the project budget. However, this technology was never developed, and the government did not implement it because they had data obtained via a traditional data collection process. In this study, traditional data were processed to predict the behavior in Thai government construction projects using a machine learning model. The data were collected from the government procurement system in 2019. There were seven input data, including the project owner department, type of construction project, bidding method, project duration, project scale, winning price overestimated price, and winning price over budget. A range of classification techniques, including an artificial neural network (ANN), a decision tree (DC), and a K-nearest neighbor (KNN), were used in this study. According to the results, after hyperparameter tuning, the ANN had the greatest prediction accuracy of 78.9 percent. This study confirms that the data from the Thai government procurement system can be investigated using machine learning techniques from big data technologies.

Keywords: artificial neural network; procurement management; construction budgeting; machine learning; big data

1. Introduction

Big data, or data management technology, has been used in many sectors by employing historical data. This trend has had an impact on the construction sector [1,2]. To assist in mitigating issues with building projects, the Thai government has been attempting to enhance its procurement system [3,4]. Risk avoidance is a method of decreasing risks in construction projects, but conflicts still arise if mistakes are made [5,6]. In government projects, such mistakes will lead to a lack of openness regarding how government officers perform their jobs. Technologies have enabled the development of tools for helping government officers avoid risks in areas, such as procurement management, where corruption should be avoided [7]. A common analysis technique in data management is machine learning [2,8]. It has been used to improve the effectiveness of construction management by focusing on analyzing historical data to generate new information or attempting to understand construction management behavior. Thus, many opportunities are provided by these technologies [9–11].

The benefits and possibilities of implementing machine learning in the Thai construction industry have been acknowledged by the Thai government [1]. Although difficulties

might arise due to the ways in which Thai government agencies operate, the government will keep looking for solutions and methods of studying [12] the procurement system to understand how it has operated in the past [7]. Along with simplifying the procedure, estimating costs ahead of time and creating a budget-compliant estimate are important components of the procurement process [6,13]. In doing so, the pricing of bids can be regulated [14]. Consequently, this study intends to advance the understanding gained through the actions of procurement systems in 2019 (Electronic Government Procurement: e-GP). Moreover, data were gathered from all kinds of government construction projects, including building projects, to investigate and highlight the difference between the winning bid and the project budget and to observe the budgeting behavior following the use of machine learning. It could be helpful to avoid winning bids over the project's budget because if it happens, the Thai government should complete the project in the initiation phase again. As a result, the people will lose the benefit from this.

The main contribution of this study is that it developed a machine learning model that is able to accurately predict budgeting behavior. If we succeed in achieving our data-related objectives, we might demonstrate that traditional data collection techniques aid in data management. Three algorithms were applied to the data to demonstrate that the model possesses the accuracy that big data technologies require. Our prediction of the behavior in the auction process is that the winning price will be over the budget.

2. Big Data

Today, the world is driven by various forms of data [15]. The increased usage of data has impacted the business environment and is causing many firms to shift as competition for development intensifies. As a result, firms are enhancing their operations using information technology (IT) [16]. However, the proliferation of data within businesses has an impact on traditional analytic tools and requires software suppliers to offer new analytical tools to manage huge volumes of data, commonly known as big data [17]. Big data refers to massive volumes of complicated data, both organized and unstructured, that cannot be handled using typical analytical and algorithmic approaches. The goal of this technology is to expose hidden patterns or knowledge in vast amounts of data, which has led to the creation of data-driven science [18].

Different algorithmic strategies are used to boost productivity in different sectors [19]. The building industry has grown along with the digital revolution and faces the acquisition of substantial volumes of data because of project execution [20]. However, data from this business are difficult to use since they are obtained from various sources and are in varied formats. Data security has been a source of worry, and our understanding of it is limited [17]. However, most studies on big data technology only look at the benefits of employing accessible analytical data for their businesses rather than the readiness of technology for businesses [16,21]. The requirements of implementing big data in businesses should be evident in that it necessitates numerous abilities, such as the gathering, processing, and analysis of vast volumes of data that may arise when data from multiple sources are collected at a high velocity [22].

Researchers have defined several factors that influence big data technology readiness. However, like any technology, these factors may differ from one organization or industry to another [3,8,9] and include scalability, ICT infrastructure, information security, machine learning management, the availability of finance, competitive pressure, organization demand, and applications, as well as analytic tools [17,23].

Big data technology has numerous applications, such as waste minimization via design, which represents the future of waste management research [24], and other concepts, such as big data with BIM, clash detection and resolution, performance prediction, etc. [25]. The building sector has been undergoing a digital revolution and is implementing big data technology into building information modeling (BIM) to handle building project data. BIM data are often 3D geometrically encoded and computationally intensively compressed, exist in a variety of proprietary formats, and are interconnected. Accordingly, the data are

gradually enriched, although the project life cycle that makes BIM files can quickly become voluminous with building models easily reaching fifty gigabytes in size [26,27].

3. Thai Government Procurement

Government procurement is crucial to a country's growth since it requires the government to spend budget funds on supplies for public services, such as education, safety, security, and facilities [28]. As a result, the sole factor that restricts government spending is the quality of procurement analysis. However, effective procurement does not always involve paying the lowest price. Instead, acquisitions are motivated mostly by the desire to advance the nation's technology and industry [29]. Government procurement has expanded rapidly in many countries to support the rise in national and international economies. Procurement is important because it assists in improving the most essential and valuable parts of the public and commercial sectors. Therefore, to become more flexible, the public procurement system must rapidly develop and adapt. The evolving aims of the public, corporate, and civil service sectors have resulted in the establishment of more efficient, transparent, and effective public and private procedures. The administrative system is shrinking, but it is also becoming more adaptive and efficient [30]. The government may aid the private sector so that both sectors can benefit the country and progress indefinitely. As a result, information technology is essential in both the public and private sectors and is critical to every organization. This is because having rapid access to a large amount of current information makes work more efficient [31]. Government expenditure accounts for 10% to 15% of each country's gross domestic product (GDP). Furthermore, it is estimated that annual building expenditure accounts for approximately USD 2 billion [32]. However, the present government procurement methods fall short of adequately meeting the demands of stakeholders. According to research and the media, this is the result of a variety of procurement issues, including corruption in government procurement construction projects, procurement costs being higher than actual costs, cronyism that results in subpar work, additional costs or budget losses, and/or subpar, overpriced materials [33].

The project owner or person in control makes the decision to start a construction project. The project owner must thoroughly specify the scope of work for each task [28,34], and contractors can also be hired for each component of the task [6]. When selecting a contractor, the typical cost of the project is also an important factor to consider [13,14]. The Thai government oversees the development of price estimations for all government construction projects and enters the data into the electronic government procurement (e-GP) system [14]. The process by which the public sector purchases products and services has been transformed by information technology. E-procurement is a web-based technology that can help speed up the procurement process. The internet is used by the government to offer services and connect with residents and organizations in the digital age. To improve procurement control and eliminate corruption, the Thai government has developed e-government procurement (e-GP). Good governance refers to processes and structures that ensure effective resource management [3]. Transparency and the maximization of benefits to the country, people, and society are continuously and appropriately prioritized in good governance of public sector management. This includes the establishment of clear principles, public engagement, responsibility, the rule of law, efficacy, efficiency with equity, and accountability. E-government procurement (e-GP) has been developed by the Thai government for auctions. There are five types of contracts that are auctioned in Thailand, one of the most important being for construction projects. Additionally, the method of selecting companies to carry out work includes bidding, selection, and specific. These auctions invite contractors who have general qualifications and meet the specified conditions to submit an offer. The selection method only invites contractors who meet the specified qualifications, and the department issues invitations based on their judgment of which contractors are appropriate for their project, which must not be fewer than three firms. This specific method invites contractors who meet the requirements to submit proposals or negotiate prices with government agencies directly according to the conditions

outlined in the Act [3]. So, the e-GP system contains data on all of the construction projects that are undertaken for Thai procurement, as shown in Figure 1.
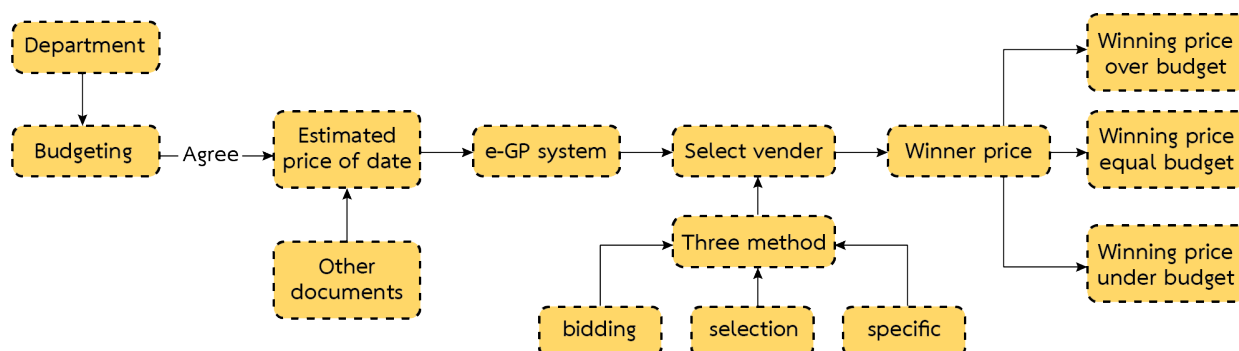
**Figure 1.** Thai procurement data collection process.

As the project owner, the government would handle the procurement process for each project after setting a budget based on the proposal presentation [14]. The winning price being over budget will make problems for government to reprocess again. It will waste time to re-establish the process for a long time, especially in government procurement [5,13,35]. As a result, price estimation is critical for government personnel who are involved in operations [6]. According to government data, it is still impossible to achieve a budget that is greater than the winning price. Furthermore, for some projects, the estimated price contains a mistake; this can affect government officers' control and is a cause of excess and lost value [1,7]. Thus, machine learning classification approaches may be utilized to create solutions to understand budgeting behavior and investigate the influence of disagreement on pricing estimates [2,36].

4. Machine Leaning Algorithm Used in This Research

The purpose of machine learning (ML), a branch of artificial intelligence (AI), is to enable computer systems to automatically learn about a certain job using data. Several methodologies are used to model judicial reasoning and forecast litigation outcomes, including rule-based learning strategies [37], artificial neural network techniques [38], case-based reasoning tactics, and hybrid methodologies [39].

4.1. Artificial Neural Network Algorithm (ANN)

There are various types of artificial neural networks (ANNs). Classification and function estimation are ideal applications for artificial neural networks (ANNs). These algorithms have been widely employed to solve difficult industrial issues since their inception. The most common type of ANN is a multi-layer perceptron (MLP). An ANN is composed of three layers: the input layer, the hidden (intermediate) layer, and the output layer. All the ANN applications used in the construction sector demand special attention, and new ANN algorithms are being developed to learn from data with huge dimensionality (i.e., big data) [40]. Figure 2 refers to the hidden units in the neural network model [41].
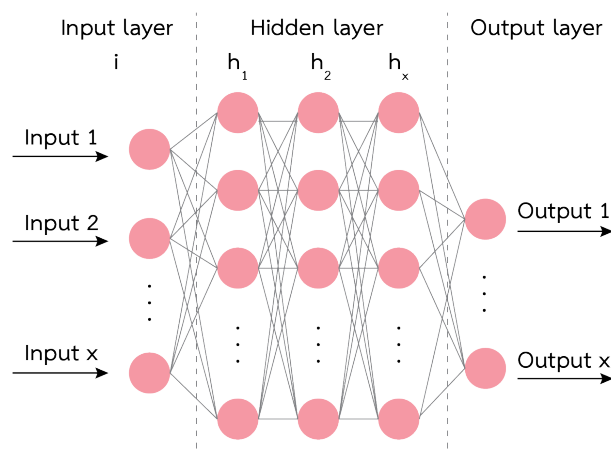
Figure 2. The structure of an ANN.

4.2. Decision Tree (DC) Algorithm

The creation of a DT begins with the discovery of decision nodes. The nodes are then separated recursively until no further divisions are feasible. Two metrics are used to test the robustness of a DT, which is dependent on the logic used to split the nodes; these metrics are information gain (IG) and entropy reduction [40]. Figure 3 exemplifies a simple decision tree model with a single binary goal variable, Y (0 or 1), and two continuous variables, x1 and x2, all of which span from 0 to 1, and depicts the basic components of a decision tree model: nodes and branches [41]. Splitting, pausing, and pruning are the key modeling procedures.
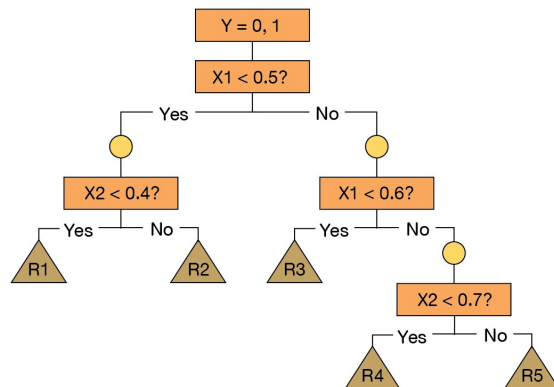
Figure 3. Sample decision tree based on binary target variable Y.

4.3. K-Nearest Neighbors (KNNs)

A KNN is a non-parametric approach used in regression and classification. A KNN has been successfully used for classification in a variety of applications. The separation between each item in a training set and between each item in a test set is calculated using this approach; keep in mind that the K items in the training set are the KNNs. The test set items are then classified based on the most frequent class in the KNN, with each neighbor having the opportunity to vote. (If there is a tie, the voting procedure includes any training set items that are not more distant than the Kth nearest neighbors, resulting in a vote total that is larger than K). A few scientists have considered this to be the most precise method of measurement. They have employed global and local measures in their initiatives, although these indicators are problem-specific. Until now, the most widely used metric has been the Euclidean distance, which calculates distances between two points by calculating the square root of the sum of the squared distances across each coordinate (possibly weighted), as shown in [42,43].

5. Machine Learning in Construction Research

Artificial intelligence (AI) is one of the major technologies driving the Industrial Revolution 4.0. It is a type of artificial intelligence that has been programmed into computers to assist in the automatic replication of intelligent actions comparable to those of humans. In other words, AI uses machines, namely computer systems, to imitate human reasoning and learning processes. This method includes comprehension (the accumulation of knowledge and the implementation of the rules that govern its application), reasoning (the application of rules to arrive at approximations), decision-making, and self-correction. "Machine learning" is a subfield of artificial intelligence (AI) technology [44]. Classification, grouping, and regression problems are tackled in this multidisciplinary area by integrating statistics, computer science, and optimization skills. Thus, machine learning is characterized by a system's capacity to learn from data. Using this technology, a system can make judgments based on previous experience in comparable situations, deal with ambiguity and limited data, and more [45]. Various studies have been conducted on the application of machine learning in construction and project management in recent years, although they are still quite scarce. These studies employ machine learning to classify construction documents based on project components.

The KNN classification algorithm was used to establish a model of awareness and information sharing, as well as to improve the quality of the present construction information management system. The proposed technique may be used in projects to assist diverse stakeholders in identifying and mitigating the risk of conflicts related to legal changes [46,47]. The authors of [48] determined that the cost and timeline of building projects may be calculated using an artificial neural network (ANN) and other models. According to their findings, early planning is critical to the success of a project. They created a text classification approach based on machine learning to classify general contract conditions and provisions. This method makes it easier to automatically assess the compliance of text-based construction contracts [49]. Using neural networks and linear regression, a construction cost index for concrete structures was created based on historical records of main construction costs. The authors' key contribution was to provide stakeholders with a credible method of forecasting pricing for prospective project developments [50].

The Thai government's procurement process has been particularly effective in obtaining and storing data in an electronic system [31], and it also includes a procurement process management philosophy for building projects [51]. Furthermore, the system may enable the development of technologies that have a proclivity to learn from past events to address current challenges and improve enterprises [52]. Data collected from government research efforts have also yielded some unusual findings. To verify whether the data obtained is acceptable for machine learning, the government agency in charge of procurement must produce and monitor the behavior of their data [53].

Previous research acquired data from Thailand's conventional government system. The four data characteristics included the department name, site location, procurement method, and project type. The authors developed a machine learning model for anticipating over-budget projects. The produced model, which had an accuracy of 0.86, was built using the KNN approach, but it used a few of the project's data [54].

6. Hyperparameter Tuning

The optimization of hyperparameters can be simplified by determining how many function evaluations will be performed on each optimization to identify the optimal hyperparameter in that model. Furthermore, optimization may be defined as follows: "given a function that accepts inputs and returns a numerical output, how can it efficiently find the inputs, or parameters, that maximize the function's output?" [55]. As a result, while tuning or optimizing a hyperparameter, the author will accept the input as a function of the hyperparameter model and the output as a measurement of the model's performance [56]. This represents the rate of miscalculation or mistakes. The hyperparameter space contains

all the potential values that are often established as acceptable boundaries for each hyperparameter, and the number of hyperparameters equals the function's dimension [57,58].

According to prior research, adjusting a hyperparameter necessitates an understanding of the link between the settings and the model's performance. The model will first conduct a trial to collect performance data on several settings and then make an inference to choose which configuration will be used next. The goal of optimization is to reduce the number of hyperparameter trials while identifying the best model [59]. As a result, the author might regard the process as sequential rather than parallel.

6.1. The Hyperparameters of ANN

The number of neurons in each hidden layer is the first hyperparameter that must be adjusted. The number of neurons in each layer is specified to be the same in this scenario. This can also be established in a variety of ways. The number of neurons should be proportional to the complexity of the answer. To forecast at a higher degree of complexity, more neurons are required. The number of neurons is specified to range between 10 and 100. Each layer has an activation function as a parameter. The input data are delivered to the input layer, followed by the hidden layers, and finally, the output layer. The output value is stored in the output layer. The activation function causes the input values to change as they go from one layer to the next. The activation function determines how to convert a layer's input values into output values. The output values of one layer are then transferred as input values to the following layer. These values are subsequently computed to obtain the output values for the following layer. To tune into this presentation, there are nine activation functions. To compute the input values, each activation function has its own formula (and graph). The neural network's layers are assembled, and an optimizer is assigned. The optimizer oversees the alteration of the learning rate and the weights of neurons in the neural network to obtain the lowest loss function.

The optimizer is critical for achieving the maximum possible accuracy or minimizing loss. There are seven different optimizers to choose from, and each is based on a distinct idea. The learning rate is one of the optimizer's hyperparameters. The learning rate governs the step size used by a model to achieve the smallest loss function. A greater learning rate allows the model to learn more quickly but may cause it to miss the minimal loss function and merely reach functions in its immediate surroundings. A lower learning rate increases the likelihood of finding a minimal loss function. A lower learning rate necessitates longer epochs or more time and memory capacity resources. The model will take longer to develop if the training dataset has too many observations [60,61].

6.2. The Hyperparameters of DC

The process of calibrating our model by finding the right hyperparameters to generalize it is called hyperparameter tuning. We will discuss a few of these hyperparameters in this paper. This argument represents the maximum depth of a tree. If not specified, the tree is expanded until the last leaf nodes contain a single value. Hence, by reducing this value, we can preclude the tree from learning all of the training samples, preventing over-fitting [61,62].

6.3. The Hyperparameters of KNN

Hyperparameter tuning is achieved by performing an exhaustive search of all possible combinations of the KNN parameters. This helps to achieve better accuracy by searching for the best combination of parameters for training [61,63]. The parameters eligible for KNN algorithms are as follows:

- Distance functions: the search for hyperparameters among distance functions.
- The distance weighting might be equal, inverse, or squared inverse. The exhaustive hyperparameter search guarantees that all three distance weighting functions are tried.
- Number of neighbors: This hyperparameter search ranges from 1 to N.

- Data standardization: Standardization is the process of standardizing data to guarantee that the training data fall within the range of [0, 1]. Cross-validation of the hyperparameters is used to select the optimal training parameters. More information on cross-validation may be found in the following section.

7. Conceptual Framework

The conceptual underpinnings of this study are the procurement practices of the Thai government in a government building project. This research will enter conventional data for analysis using a machine learning technique to predict the winning price and compare it to the project budget based on Figure 4.
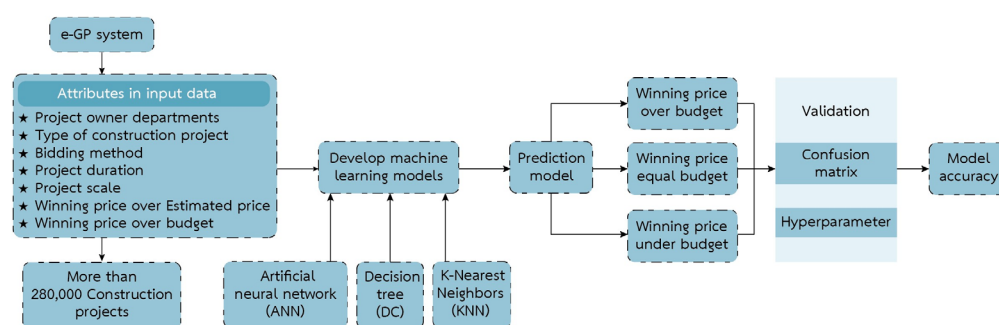

Figure 4. Conceptual framework of this study.

8. Research Methodology

The computerized government procurement system (e-GP) was used to collect the data for this investigation, which may be accessed with authorization from the comptroller general's office. The data encompass all government construction projects carried out in 2019 with a total of about 283,000, as well as the 3 scenarios investigated in this study: a winning price under the project budget, a winning price equal to the project budget, and a winning price over the project budget. The data were separated into two categories based on how they were collected: 20% of the data were used for model validation and 80% of the data were utilized for model training. Moreover, an artificial neural network (ANN) was used to assess the training data, and the results were validated using a confusion matrix [2,59].

8.1. Application of Machine Learning

The collected data had to be formatted as a CSV file to ensure that there were no blank values or categories with unknown contents. The ANN also required the use of computer software. The buried layer of the ANN was 100 [64] in size. Anaconda software 2.4.0 was used to run the Python-based computer application.

8.2. Verifying the Model

Building a confusion matrix and using the following Equations (1)–(3) allowed the ANN model's accuracy, precision, and recall to be checked [65] where true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) represent the true, false, positive, and negative states, as shown in Figure 5.

$$\text{Accuracy} = (TP - TN)/TP - TN - FP - FN) \tag{1}$$

$$\text{Precision} = TP/(TP - FP) \tag{2}$$

$$\text{Recall} = TP/(TP - FN) \tag{3}$$

Figure 5. Table of confusion matrix.

The proportion of total accurate classifications to total predicted classifications is used to measure a model's accuracy. Another definition of precision is the ability of a model to obtain consistent results from a variety of measurements. In information retrieval, a random error is a type of observational error that results in differences between precise values [59].

### 8.3. Hyperparameter Optimization with Random Search

Machine learning models contain hyperparameters that must be set in order for the model to be customized to a dataset. The general effects of hyperparameters on a model are well understood, but determining how to optimize a hyperparameter and combinations of interacting hyperparameters for a given dataset can be difficult. For configuring hyperparameters, there are frequently broad heuristics or rules of thumb. A better technique would be to objectively search for different values of model hyperparameters and select a subset that results in the model that performs best on a particular dataset. This is known as hyperparameter optimization or hyperparameter tuning, and it is supported by the scikit-learn Python machine learning toolkit. Hyperparameter optimization results in a single set of hyperparameters with good performance that can be used to configure a model [56].

Hyperparameters are points of choice or configurations that allow a machine learning model to be tailored to a given job or dataset. The model configuration argument is given by the developer to guide the learning process for a specific dataset. Machine learning models also have parameters, which are the internal coefficients determined by training or tuning the model using a training dataset. Parameters differ from hyperparameters; parameters are learned automatically, while hyperparameters are set manually to aid in the learning process. In general, a hyperparameter has a known effect on a model, but it is unclear how to optimally configure a hyperparameter for a given dataset. Furthermore, many machine learning models feature a variety of hyperparameters that might interact nonlinearly. As a result, it is frequently necessary to look for a set of hyperparameters that result in the greatest performance of a model on a dataset. This is known as hyperparameter optimization, tweaking, or hyperparameter search [58].

A search space is defined as part of an optimization technique. This can be visualized as an n-dimensional volume with each hyperparameter representing a separate dimension and the scale of the dimension represented by the values that the hyperparameter can take on, such as real-valued, integer-valued, or categorical. A random search defines a search space as a bounded domain of hyperparameter values that is randomly sampled [56].

8.4. Data Collection

The process by which the data were collected from the government is specified in Figure 6, and the seven characteristics of the input data gathered from the conventional Thai government system are displayed in Table 1. Each parameter has a unique attribute according to its definition, and the first parameter has the department group name attribute, which includes 13 groups of departments in Thailand.
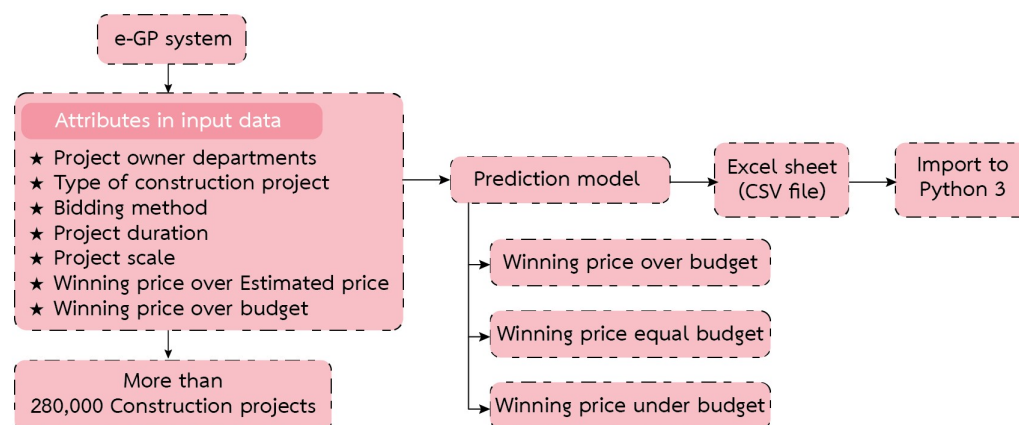
Figure 6. Data collection process of this study.

Table 1. Attributes of input data.

No.	Attributes	Number of Categories
1	Project owner department	13
2	Type of construction project	3
3	Bidding method	3
4	Project duration	5
5	Project scale	5
6	Winning price overestimated price	3
7	Winning price over budget	3

The project owner is one of the parties involved in building project conflicts [13,15], which are one of the key aspects that influence cost estimation. The project type characteristic is made up of three components, including attempts to develop roads and install irrigation systems [14]. This trait has a considerable influence on price estimation, budgeting, and procurement [21]. The Thai procurement method characteristic consists of three components, including the bidding method, selection method, and specific method, as shown in Table 2. Furthermore, each country has its own procurement system, but the final aim is the same: the abolition of corruption [18,66].

Table 2. Thai procurement method.

Method of Procurement	Details
Bidding	Every firm was welcome to join and evaluate the initiatives.
Chosen	Only the qualifying firms could submit a proposal with specific project requirements.
Specific	The contractors might be chosen by the proprietors alone.

Using the procurement technique has a considerable impact on a contractor's cost [14,18]. In this paper, the project scales were also separated into five levels as show that in Table 3 [54]. The departments were divided according to their budgets in 2019 in Table 4. The project duration was specified according to the contract. The two factors that were considered in this study included a comparison between the winning price with an estimated price and the winning price with a budget [54]. However, comparing the budget and the winning price was the goal of the prediction model.

Table 3. Project scale.

Project Scale	Details (Million USD)
L1	<140,000
L2	140,001–280,000
L3	280,001–1,400,000
L4	1,400,001–7,000,000
L5	>7,000,001

Table 4. Value of departments in Thailand.

Group of Departments	Price (USD)	%
University	747,838,774	5.63
School	477,369,208	3.59
Hospital	332,634,917	2.50
Irrigation	680,402,982	5.12
Public works and town and country planning	29,724,273,461	6.14
Highways	2,761,913,958	20.78
Rural roads	1,131,877,654	8.51
Finance	15,551,811	0.12
Local administration	2,946,100,502	22.16
Justice	560,697,143	4.22
Police	290,890,893	2.19
Military	336,971,946	2.53
Other	2,195,136,360	5.63
Sum	13,294,211,568	100.0

9. Results

The results of this study, obtained using several analysis tools and techniques, are split into the following three parts: general information, the machine learning model, and the validation of data using a confusion matrix.

9.1. General Information

The specific approach that accounts for 77.8% of all procurement methods is the one preferred by the Thai government. The largest group of departments is classified as "Other," which accounts for 67.5 percent of Thai department groups in 2019. The departments with the highest budgets, however, are the local government and highways, as seen in Figure 7. Roads and buildings represent the main construction projects undertaken by the Thai government. As seen in Figure 8, one of the most pressing concerns for the Thai government in terms of its effect on the Thai people is infrastructure [7,14]. This study classified the budget situation into two categories, estimated price over budget and winning price over budget, as shown in Figures 9 and 10. This study aimed to predict the winning price in a budgetary context. A few successful projects have higher prices than anticipated due to their overall structure. This is an important aspect of this study, as police are prone to corruption [7]. The government's cost evaluation process is inadequate since there are 700 projects for which the winning price is higher than the estimated price [5,6].
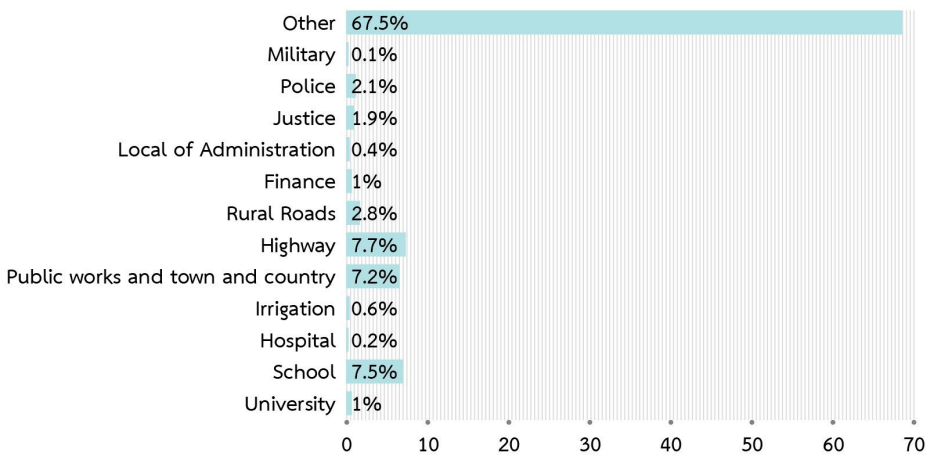
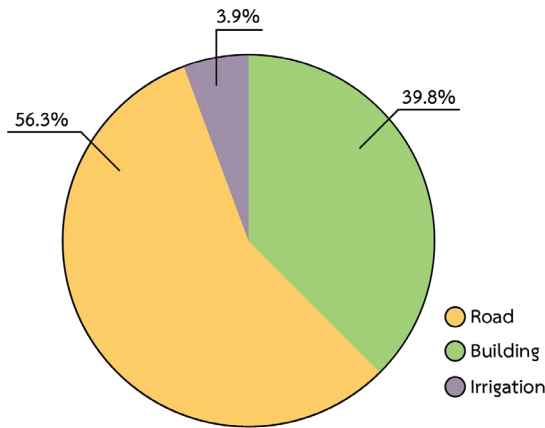**Figure 7.** Groups of departments in Thailand.

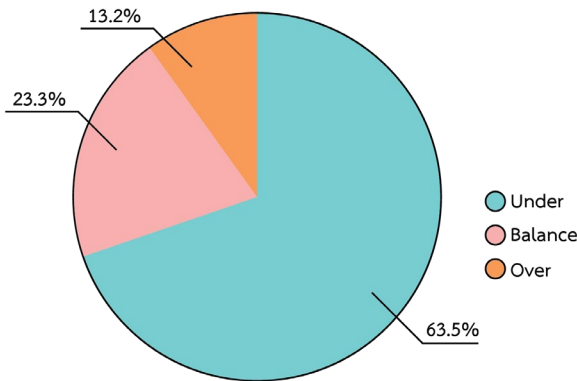
Figure 8. Percentages of project types.

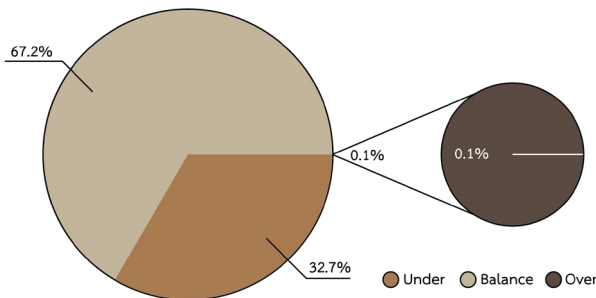**Figure 9.** Estimated price over budget.

Figure 10. Winning price over budget.

9.2. Machine Learning Model

There were three algorithms used to generate the model for categorizing the behavior of Thai government construction project bidders: ANN, decision tree, and KNN. Table 5 shows that the accuracy of the algorithms is approximately equal. The ANN algorithm has the greatest percentage at 77.60 and has greater efficiency compared to previous studies that use the same analytical techniques [67].

Table 5. Accuracy of each algorithm.

Algorithm	Accuracy
ANN	77.60%
Decision tree	77.30%
KNN	75.00%

9.3. Validating Data Using Confusion Matrix

A confusion matrix [68] was used to calculate the classification accuracy of the model. The ANN model matrix reveals that the model correctly predicted 44,028 out of 56,705 cases. In Table 6, the grey box represents misclassified cases, and the white box represents correctly classified ones, and the number zero in the confusion matrix indicates that the model did not make a prediction error. Similarly, the accuracy of the ANN model may be determined using the confusion matrix. As indicated in Table 7, accuracy may be split into three categories of bidding behavior (under, equal, and above). For the under-cluster, the model achieved a precision of 83%. For the equal cluster, the model achieved a precision of 77%. For the over-cluster, the model did not achieve a precision score, as shown in Table 7.

Table 6. Confusion matrix table of ANN model.

		Actual Class		
		Under	**Equal**	**Over**
	Under	7632	11,072	0
Predicted class	Equal	1561	36,393	0
	Over	7	40	0

Table 7. Confusion matrix of ANN model.

	Precision	Recall	f1-Score	Support
Under	0.83	0.41	0.55	18,704
Equal	0.77	0.96	0.85	37,954
Over	0.00	0.00	0.00	47
Accuracy			0.78	56,705
Macro avg	0.53	0.46	0.47	56,705
Weighted avg	0.79	0.78	0.75	56,705

The matrix of the decision tree model shows that the model correctly predicted 44,050 of 56,705 cases. The grey box represents misclassified cases, and the white box represents correctly classified ones, as shown in Table 8, and the number zero in the confusion matrix indicates that the model did not make a prediction error. Similarly, the decision tree model's precision can also be calculated using the confusion matrix. The precision can be divided into three types of bidding behavior (i.e., under, equal, and over). For the under-cluster, the model achieved a precision of 81%. For the equal cluster, the model achieved a precision of 77%. For the over-cluster, the model did not achieve a precision score, as shown in Table 9.

Table 8. Confusion matrix table of decision tree model.

		Actual Class		
		Under	**Equal**	**Over**
	Under	7572	11,022	1
Predicted class	Equal	1762	36,298	0
	Over	16	34	0

Table 9. Confusion matrix of decision tree model.

	Precision	**Recall**	**f1-Score**	**Support**
Under	0.81	0.41	0.54	18,595
Equal	0.77	0.95	0.85	38,060
Over	0.00	0.00	0.00	50
Accuracy			0.77	56,705
Macro avg	0.53	0.45	0.46	56,705
Weighted avg	0.78	0.77	0.75	56,705

The matrix of the KNN model shows that the model correctly predicted 42,549 of 56,705 cases. The grey box represents misclassified cases, and the white box represents correctly classified ones, as shown in Table 10, and the number zero in the confusion matrix indicates that the model did not make a prediction error. Similarly, the KNN model's precision can also be calculated using the confusion matrix. The precision can be divided into three bidding behaviors (i.e., under, equal, and over). For the under-cluster, the model achieved a precision of 66%. For the equal cluster, the model achieved a precision of 78%. For the over-cluster, the model did not achieve a precision score, as shown in Table 11.

Table 10. Confusion matrix table of KNN model.

		Actual Class		
		Under	**Equal**	**Over**
	Under	9370	9194	0
Predicted class	Equal	4909	33,179	0
	Over	17	36	0

Table 11. Confusion matrix of KNN model.

	Precision	**Recall**	**f1-Score**	**Support**
Under	0.66	0.50	0.57	18,564
Equal	0.78	0.87	0.82	38,088
Over	0.00	0.00	0.00	53
Accuracy			0.75	56,705
Macro avg	0.48	0.46	0.46	56,705
Weighted avg	0.74	0.75	0.74	56,705

The precision of the confusion matrix shows that the ANN algorithm had the highest accuracy in all cases; however, the KNN had high efficiency with the ANN in the equal case. This could prove that traditional data have the potential for application in data technology [2]. The performance of classification algorithms is typically assessed by evaluating the accuracy of classification using artificial neural networks, and good results can be achieved [69], as shown in Table 12. However, the over case cannot be processed if the dataset is too small, and in this case, this problem occurred a few times. According to this, procurement in Thailand is an efficient process.

Table 12. Precision accuracy of each algorithm.

Cases	Machine Learning Algorithm		
	ANN	**Decision Tree**	**KNN**
Under	83%	81%	66%
Equal	77%	77%	78%
Over	0%	0%	0%

9.4. After Hyperparameter Tuning

In the final experiments, underperforming hyperparameters were removed. The random search technique randomly samples the hyperparameter space. According to [70], a random search has more advantages than a grid search in terms of its applications even if the computer cluster fails. It enables practitioners to adjust the "resolution" on the fly and to add additional trials to the set or even disregard the failure test. At the same time, the random search procedure may be stopped at any time, enabling a full experiment to be carried out concurrently [71]. Furthermore, if more computers become available, a new trial may be added to the experiment without compromising it [72]. The following are the primary parameters for each model: ANN model: random_state = 42, hidden layer size = 20, alpha = 0.001, and activation = tanh; decision tree model: random_state = 42, min_samples_leaf = 4, max_depth = 10, and n_iter = 10; KNN model: 'kneighborsclassifier__weights': 'distance', random_state = 42, min_samples_leaf = 10, n_neighbors = 7, n_iter = 10, and algorithm = kd_tree, as shown in Table 13.

Table 13. Accuracy of hyperparameter.

Algorithm	Accuracy before Hyperparameter	Accuracy after Hyperparameter
ANN	77.6%	78.9%
Decision tree	77.3%	78.8%
KNN	75.0%	77.7%

10. Discussion

All of the attributes in the input data are crucial for increasing the model's predictive performance [14]. The specific method of the procurement method used by the Thai government was created specifically for small-scale projects like the reinforcing of concrete roadways and tiny buildings. This special technique was created by the government for preferred contractors whose ability to finish projects quickly can be guaranteed. This approach has the advantage of enabling the delivery of built amenities to people more quickly than the bidding technique. Government officials can therefore participate in a variety of activities when using a particular method that influences the procurement process provided that their actions are not monitored and recorded [73]. Procurement regulation does, however, include a shortcoming that can damage the government's reputation. This procurement approach has a fault in that it is only appropriate for small projects. Furthermore, only government agencies have the authority to choose contractors in a direct, straightforward manner. There is a project auditing department, although it cannot audit every single project. Another study indicates that a key contributing factor to corruption is government agencies [3].

The goal of this system is to facilitate employers and contractors in the construction industry of the public sector. The system can store data but cannot be analyzed or applied immediately in accordance with the research that in the past, government agencies have been working on, which is in the early stages [12]. The Thai government's data collection procedures are effective and of high quality. The building industry, however, could greatly benefit from the adoption of data technology and big data technologies, thus leading it to

become a developed nation [2]. They should, however, design their digital data collection method in such a way that it supports the big data concept and uses current technology.

The e-GP system can form part of big data analysis for the Thai government if they use the model of prediction proposed in this study. Additionally, the highest result of the model was for the ANN whose accuracy rate was 78.9 percent. It underscores that this ANN classification model offers high accuracy compared to other algorithms on the same data set [74]. In recent studies, it has been proven that data correlation is important for big data technology. A high yield of the model implies a high correlation of the data [75]. Big data will prove efficient and successful if the government makes efforts to plan a process of enabling its use. Adopting these technologies will result in success and be highly beneficial [1].

11. Conclusions

This study demonstrates that the ANN's accuracy rate was 78.9 percent. Data from Thailand's conventional government system were employed in the current analysis. The project owner department, type of construction project, bidding method, project duration, project scale, winning price over the estimated price, and winning price over the budget represent the seven characteristics studied. The effectiveness of the three algorithms showed that data may be utilized to accurately forecast budgeting behavior. The budgeting process is one of the most important aspects of government construction project management and needs to be carried out before the auction process. Government procurement management enables the early definition of a project's budget, and a budget for a construction project is crucial since it may save time and money over the course of the project. Through procurement management, the government will be able to take the actions necessary to guarantee that construction projects stay within budget after finishing the bidding process; without this process, the government would have more work to perform and re-bidding would be necessary. This is a waste of time for the officer and emphasizes the importance of proper planning. As a result, this study provides an opportunity for the government to apply the data from the traditional procurement system to help with their work. They can use this prototype to develop their technologies in the future to recheck tools for their budget. Machine learning algorithms can be adopted, and they will be more efficient if the government plans and develops its goals using several technologies. This will greatly benefit people and increase the performance transparency of technologies.

Finally, this study demonstrates that the Thai conventional data-gathering technique can be used with machine learning from big data. The data used to create the machine learning models represent the raw data obtained from the procurement system. Additionally, if the government enforces its policy to enhance data-gathering techniques, data collection will be more efficient and productive.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Srinavin, K.; Kusonkhum, W.; Chonpitakwong, B.; Chaitongrat, T.; Leungbootnak, N.; Charnwasununth, P. Readiness of Applying Big Data Technology for Construction Management in Thai Public Sector. *J. Adv. Inf. Technol.* **2021**, *12*, 1–5. [CrossRef]
2. Bilal, M.; Oyedele, L.O.; Qadir, J.; Munir, K.; Ajayi, S.O.; Akinade, O.O.; Alaka, H.; Pasha, M. Big Data in the Construction Industry: A Review of Present Status, Opportunities, and Future Trends. *Adv. Eng. Inform.* **2016**, *30*, 500–521. [CrossRef]
3. Chaitongrat, T.; Leungbootnak, N.; Kusonkhum, W.; Deewong, W.; Liwthaisong, S.; Srinavin, K. Measurement Model of Good Governance in Government Procurement. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *639*, 012024. [CrossRef]
4. Soni, S.; Pandey, M.K.; Agrawal, S. Conflicts and Disputes in Construction Projects: An Overview. *Int. J. Eng. Res. Appl.* **2017**, *07*, 40–42. [CrossRef]
5. Jaffar, N.; Tharim, A.H.A.; Shuib, M.N. Factors of Conflict in Construction Industry: A Literature Review. *Procedia Eng.* **2011**, *20*, 193–202. [CrossRef]
6. Rose, K. A Guide to the Project Management Body of Knowledge (PMBOK®Guide)—Fifth Edition. *Proj. Manag. J.* **2013**, *44*, e1. [CrossRef]
7. Chaitongrat, T. Causal relationship model of problems in public sector procurement. *Int. J. Geomate* **2021**, *20*, 52–58. [CrossRef]
8. Hurwitz, J.; Kirsch, D. *Machine Learning for Dummies*; IBM Limited Edition; IBM: Armonk, NY, USA, 2018; p. 75.
9. Bai, S.; Li, H.; Kong, R.; Han, S.; Li, H.; Qin, L. Data Mining Approach to Construction Productivity Prediction for Cutter Suction Dredgers. *Autom. Constr.* **2019**, *105*, 102833. [CrossRef]
10. Naganathan, H.; Chong, W.C.; Chen, X.-W. Building Energy Modeling (BEM) Using Clustering Algorithms and Semi-Supervised Machine Learning Approaches. *Autom. Constr.* **2016**, *72*, 187–194. [CrossRef]
11. Poh, C.Q.X.; Ubeynarayana, C.U.; Goh, Y.M. Safety Leading Indicators for Construction Sites: A Machine Learning Approach. *Autom. Constr.* **2018**, *93*, 375–386. [CrossRef]
12. Chonpitakwong, B.; Kusonkhum, W.; Chaitongrat, T.; Srinavin, K.; Charnwasununth, P. Hindrance of Applying Big Data Technology for Construction Management in Thai Government. *J. Adv. Inf. Technol.* **2021**, *12*, 159–163. [CrossRef]
13. Jervis, B.M.; Levin, P.T. *Construction Law, Principles and Practice*; McGraw-Hill College: New York, NY, USA, 1988.
14. The Comptroller General's Department. *The Government Procurement and Supplies Management Act B.E. 2560*; The Comptroller General's Department: Bangkok, Thailand, 2017.
15. Deal, J.L. Information: A Revolution That Will Transform How We Live, Work, And Think by Mayer-Schonberger Viktor Cukier Kenneth New York (NY): Houghton Mifflin Harcourt, 2013, 242 Pp., $27.00. *Health Aff.* **2014**, *33*, 1300. [CrossRef]
16. Michael, K.; Miller, K. Big: New Opportunities and New Challenges [Guest Editors' Introduction]. *IEEE Comput.* **2013**, *46*, 22–24. [CrossRef]
17. *Big Data Analytics*; Springer: Berlin, Germany, 2018.
18. Creely, E.; Henriksen, D.; Henderson, M. Artificial intelligence, creativity, and education: Critical questions for researchers and educators. In Proceedings of the Society for Information Technology & Teacher Education International Conference, New Orleans, LA, USA, 13 March 2023; pp. 1309–1317.
19. Anand, R. More Data Usually Beats Better Algorithms. *DataWocky* **2008**.
20. Eadie, R.; Browne, M.; Odeyinka, H.; McKeown, C.; McNiff, S. BIM implementation throughout the UK construction project lifecycle: An analysis. *Autom. Constr.* **2013**, *36*, 145–151. [CrossRef]
21. Betting on Big Data: How the Right Culture, Strategy and Investments Can Help You Leapfrog the Competition. *Forbes Insights*. 2015. Available online: https://www.forbes.com/forbesinsights/teradata_big_data/index.html (accessed on 2 July 2023).
22. Kaisler, S.; Armour, F.; Espinosa, J.A.; Money, W. Big data: Issues and challenges moving forward. In Proceedings of the 2013 46th Hawaii International Conference on System Sciences, Wailea, HI, USA, 7–10 January 2013; pp. 995–1004.
23. Wielki, J. Implementation of the concept in organizations-possibilities, impediments and challenges. In Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, Krakow, Poland, 8–11 September 2013; pp. 985–989.
24. Osmani, M.; Glass, J.; Price, A. Architect and contractor attitudes to waste minimisation. *Proc. Inst. Civ. Eng. Waste Resour. Manag.* **2006**, *169*, 65–72. [CrossRef]
25. Wang, L.; Leite, F. Knowledge discovery of spatial conflict resolution philosophies in BIM-enabled MEP design coordination using data mining techniques: A proof-of-concept. *Comput. Civ. Eng.* **2013**, *2013*, 419–426.
26. Jiao, Y.; Zhang, S.; Li, Y.; Wang, Y.; Yang, B.; Wang, L. An augmented MapReduce framework for building information modeling applications. In Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Hsinchu, Taiwan, 21–23 May 2014; pp. 283–288.
27. Lin, J.R.; Hu, Z.Z.; Zhang, J.P.; Yu, F.Q. A natural-language-based approach to intelligent data retrieval and representation for cloud BIM. *Comput. Civ. Infrastruct. Eng.* **2016**, *31*, 18–33. [CrossRef]
28. Dzuke, A.; Naude, M.J. Procurement challenges in the Zimbabwean public sector: A preliminary study. *J. Transp. Supply Chain. Manag.* **2015**, *9*, a166. [CrossRef]
29. Hazra, J.; Mahadevan, B. A procurement model in an electronic market with coordination costs. In Proceedings of the 2011 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, 6–9 December 2011; pp. 1364–1368.

30. Mark McKevitt, D.; Davis, P.J. Supplier development and public procurement: Allies, coaches and bedfellows. *Int. J. Public Sect. Manag.* **2014**, *27*, 550–563. [CrossRef]

31. Leungbootnak, N.; Chaithongrat, T.; Aksorn, P. An exploratory factor analysis of government construction procurement problems. *MATEC Web Conf.* **2018**, *192*, 02057. [CrossRef]

32. Tanayut, C.; Narong, L.; Preenithi, A.; Patrick, M.J. Application of Confirmatory Factor Analysis in Government Construction Procurement Problems in Thailand. *Int. Trans. J. Eng. Manag. Appl. Sci. Technol.* **2017**, *8*, 22.

33. Du, J.; Jiao, Y.-Y.; Jiao, R.J.; Kumar, A.; Ma, M. A case study of obsolete part procurement process reengineering. In Proceedings of the 2007 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, 2–7 December 2007; pp. 1337–1341.

34. Burke, R. *Project Management: Planning and Control Techniques*; John Wiley & Sons: Hoboken, NJ, USA, 2013.

35. Chitkara, K. *Construction Project Management-Planning, Scheduling and Controlling*; Tata McGraw Hills: New York, NY, USA, 2011.

36. Maemura, Y.; Kim, E.; Ozawa, K. Root causes of recurring contractual conflicts in international construction projects: Five case studies from Vietnam. *J. Constr. Eng. Manag.* **2018**, *144*, 05018008. [CrossRef]

37. Diekmann, J.E.; Kruppenbacher, T.A. Claims analysis and computer reasoning. *J. Constr. Eng. Manag.* **1984**, *110*, 391–408. [CrossRef]

38. Kim, M.P. US Army Corps Engineers construction contract claims guidance system. In *Utilization of Ocean Waves—Wave to Energy Conversion*; ASCE: Reston, VA, USA, 1989; pp. 203–209.

39. Chau, K.-W. Prediction of construction litigation outcome—A case-based reasoning approach. In *Advances in Applied Artificial Intelligence: 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Annecy, France, 27–30 June 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 548–553.

40. Atuahene, B.T.; Kanjanabootra, S.; Gajendran, T. Transformative role of through enabling capability recognition in construction. *Constr. Manag. Econ.* **2023**, *41*, 208–231. [CrossRef]

41. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 112.

42. Breiman, L. *Random Forests—Random Features*; University of California: Berkley, CA, USA, 1999.

43. Zhang, Z. Introduction to machine learning: K-nearest neighbors. *Ann. Transl. Med.* **2016**, *4*, 218. [CrossRef] [PubMed]

44. Canhoto, A.I.; Clear, F. Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Bus. Horizons* **2020**, *63*, 183–193. [CrossRef]

45. Chen, J.-H. KNN based knowledge-sharing model for severe change order disputes in construction. *Autom. Constr.* **2008**, *17*, 773–779. [CrossRef]

46. Xie, S.; Fang, J. Prediction of construction cost index based on multi variable grey neural network model. *Int. J. Inf. Syst. Chang. Manag.* **2018**, *10*, 209–226. [CrossRef]

47. Salama, D.M.; El-Gohary, N.M. Semantic text classification for supporting automated compliance checking in construction. *J. Comput. Civ. Eng.* **2016**, *30*, 04014106. [CrossRef]

48. Elfahham, Y. Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alex. Eng. J.* **2019**, *58*, 499–506. [CrossRef]

49. Nguyen, P.T.; Nguyen, Q.L.H.T.T. Critical factors affecting construction price index: An integrated fuzzy logic and analytical hierarchy process. *J. Asian Financ. Econ. Bus.* **2020**, *7*, 197–204. [CrossRef]

50. Lin, W.-C.; Ke, S.-W.; Tsai, C.-F. Top 10 data mining techniques in business applications: A brief survey. *Kybernetes* **2017**, *46*, 1158–1170. [CrossRef]

51. Cheng, M.-Y.; Peng, H.-S.; Wu, Y.-W.; Chen, T.-L. Estimate at completion for construction projects using evolutionary support vector machine inference model. *Autom. Constr.* **2010**, *19*, 619–629. [CrossRef]

52. Cost, S.; Salzberg, S.J. A weighted nearest neighbor algorithm for learning with symbolic features. *Mach. Learn.* **1993**, *10*, 57–78. [CrossRef]

53. Roy, R.; Low, M.; Waller, J.J. Documentation, standardization and improvement of the construction process in house building. *Constr. Manag. Econ.* **2005**, *23*, 57–67. [CrossRef]

54. Kusonkhum, W.; Srinavin, K.; Leungbootnak, N.; Aksorn, P.; Chaitongrat, T.J. Government construction project budget prediction using machine learning. *J. Adv. Inf. Technol.* **2022**, *13*, 29–35. [CrossRef]

55. Wistuba, M.; Schilling, N.; Schmidt-Thieme, L. Hyperparameter optimization machines. In Proceedings of the 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, Canada, 17–19 October 2016; pp. 41–50.

56. Le, Q.V.; Ngiam, J.; Coates, A.; Lahiri, A.; Prochnow, B.; Ng, A.Y. On optimization methods for deep learning. In Proceedings of the 28th International Conference on International Conference on Machine Learning, Bellevue, WA, USA, 28 28–2 July 2011; pp. 265–272.

57. Wistuba, M.; Schilling, N.; Schmidt-Thieme, L. Learning hyperparameter optimization initializations. In Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, France, 19–21 October 2015; pp. 1–10.

58. Hazan, E.; Klivans, A.; Yuan, Y. Hyperparameter optimization: A spectral approach. *arXiv* **2017**, arXiv:1706.00764.

59. Hernández-Torruco, J.; Canul-Reich, J.; Frausto-Solis, J.; Méndez-Castillo, J.J. Towards a predictive model for Guillain-Barré syndrome. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 7234–7237.

60. Menapace, A.; Zanfei, A.; Righetti, M. Tuning ANN hyperparameters for forecasting drinking water demand. *Appl. Sci.* **2021**, *11*, 4290. [CrossRef]
61. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [CrossRef]
62. Mantovani, R.G.; Horváth, T.; Cerri, R.; Junior, S.B.; Vanschoren, J.; de Carvalho, A. An empirical study on hyperparameter tuning of decision trees. *arXiv* **2018**, arXiv:1812.02207.
63. Wazirali, R. An improved intrusion detection system based on KNN hyperparameter tuning and cross-validation. *Arab. J. Sci. Eng.* **2020**, *45*, 10859–10873. [CrossRef]
64. Allen, K.; Berry, M.M.; Luehrs, F.U., Jr.; Perry, J.W. Machine literature searching VIII. Operational criteria for designing information retrieval systems. *Am. Doc.* **1955**, *6*, 93.
65. Gondia, A.; Siam, A.; El-Dakhakhni, W.; Nassar, A.H. Machine learning algorithms for construction projects delay risk prediction. *J. Constr. Eng. Manag.* **2020**, *146*, 04019085. [CrossRef]
66. Suntharanurak, S. Screening for Bid Rigging in Rural Road Procurement of Thailand. Ph.D. Thesis, National Institute of Development Administration, Bangkok, Thailand, 2012.
67. Samui, P.; Roy, S.S.; Balas, V.E. *Handbook of Neural Computation*; Academic Press: Cambridge, MA, USA, 2017.
68. Lu, B.; Hardin, J. Constructing Prediction Intervals for Random Forests. Ph.D. Thesis, Pomona College, Claremont, CA, USA, 2017.
69. Tang, L.; Zhao, Y.; Cabrera, J.; Ma, J.; Tsui, K.L. Forecasting short-term passenger flow: An empirical study on Shenzhen metro. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 3613–3622. [CrossRef]
70. Bergstra, J.; Bengio, Y.J. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
71. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for hyper-parameter optimization. In Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS 2011), Granada, Spain, 12–15 December 2011; Volume 24.
72. Bergstra, J.; Yamins, D.; Cox, D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 115–123.
73. Olson, D.L.; Delen, D. *Advanced Data Mining Techniques*; Springer Science & Business Media: Berlin, Germany, 2008; p. 279.
74. Kusonkhum, W.; Srinavin, K.; Leungbootnak, N.; Chaitongrat, T. Using a Machine Learning Approach to Predict the Thailand Underground Train's Passenger. *J. Adv. Transp.* **2022**, *2022*, 8789067. [CrossRef]
75. Batty, M. Big Data, smart cities and city planning. *Dialogues Hum. Geogr.* **2013**, *3*, 274–279. [CrossRef]