

Article

Not peer-reviewed version

---

# The Adoption of a Machine Learning Approach in a Big Data Concept to Predict Project Cost Budgeting in the Thai Auction Process of Procurement Management for a Construction Project

---

[Wuttipong Kusonkhum](#) , [Korb Srinavin](#) <sup>\*</sup> , [Tanayut Chaitongrat](#)

Posted Date: 7 June 2023

doi: 10.20944/preprints202306.0541.v1

Keywords: Artificial neural network; Procurement management; Construction budgeting; Machine learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# The Adoption of a Machine Learning Approach in a Big Data Concept to Predict Project Cost Budgeting in the Thai Auction Process of Procurement Management for a Construction Project

Wuttipong Kusonkhum <sup>1</sup>, Korb Srinavin <sup>2,\*</sup> and Tanayut Chaitongrat <sup>3</sup>

<sup>1</sup> Department of Civil Engineering, Northeastern University, Khon Kaen, Thailand; w.kusonkhum@gmail.com

<sup>2</sup> Department of Civil Engineering, Khon Kaen University, Khon Kaen, Thailand

<sup>3</sup> Construction and Project Management Center, Faculty of Architecture, Urban Design and Creative Arts, Mahasarakham University, Maha Sarakham, Thailand; tanayut.c@msu.ac.th

\* Correspondence: korbsri@kku.ac.th; Tel.: +66894177357

**Abstract:** Big Data Technologies is one of the disruptive technologies that influence every business, including the construction industry. The Thai government is attempting to use machine learning technique from part of analytic by Big Data Technologies to forecast costs for public building projects. However, it was never developed, and They did not implement it with traditional data. In this study, traditional data is processed to predict the behavior of Thai government construction projects by using a machine learning model. Additionally, the data was collected from the government procurement system in 2019. There are eight input data including departmental groupings, project types, procurement methods, length, winning price over standard price, even criteria were examined, including winning price over budget and standard price above budget. Additionally, a range of classification techniques, including an artificial neural network (ANN), Decision tree (DC), K-nearest neighbor (KNN), were used in this study (ANN). According to the results, ANN has the greatest predicting accuracy with 78.9 percent after hyperparameter tuning. The study confirms that data from the Thai Government Procurement System can be usefully investigated using machine learning techniques from Big Data technologies.

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

## 1. Introduction

Big data or data management technology has been used by many sectors to their advantage by utilizing historical data. This trend has had an impact on the construction sector [1,2]. To assist lessen issues with building projects, the Thai government has been at-tempting to enhance its procurement system [3–5]. Risk avoidance, risk transfer, and risk reduction are all ways for decreasing risks in project management, but conflicts still arise [5] and they lead to a lack of openness in how government officers do their business [7]. A common technique that has been used in data management technology analysis is machine learning [2,8]. It has been used to improve the effectiveness of construction management by concentrating on analyzing historical data to generate new information or attempting to comprehend construction management behavior. [9–11].

The Thai government has recognized the advantages and potential of implementing machine learning in the Thai construction sector [1]. Although there may be some issues with the working process for Thai government agencies, the government will continue to look for ways to overcome them as well as ways to research [12] the procurement system and comprehend how such a process has behaved in the past [7]. One crucial aspect of the procurement process, in addition to streamlining the procedure, is predicting expenses before putting them up for auction [6,13] and establishing a standard price that adheres to the budget. That would aid in regulating the bidders' pricing [14]. Consequently, this study intends to advance the understanding gained through the actions of

procurement systems in 2019. Additionally, data was gathered from all government construction projects, including all kinds of building projects, in order to investigate and highlight the winning bidder's cost difference. to observe the behavior of the budget following the use of machine learning

Following this examination, the machine learning model was able to accurately anticipate their conduct. If they enhance their data-related goals, it may show how conventional process data collecting may help with data management.

## **2. Big Data in Thai Government**

Nowadays, the world is driven by several forms of data [15]. The increased usage of data has an impact on the business environment and is causing many firms to shift as competition for development intensifies. As a result, firms are enhancing their operations via the use of information technology (IT) [16]. However, the proliferation of data within businesses has an impact on traditional analytic tools and necessitates software suppliers offering new analytical tools to manage huge volumes of data, commonly known as Big Data [17]. Big Data (BIG DATA) refers to massive volumes of complicated data, both organized and unstructured, that typical analytical and algorithmic approaches cannot handle. The goals of BIG DATA technology are to expose hidden patterns or knowledge in vast amounts of data, which 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 alongside the revolution. The industry is faced with the acquisition of substantial volumes of data as a result of project execution [20]. However, data from this business is difficult to use since it comes from several sources and is in varied formats. Data security has been a source of worry, and understanding is limited [17]. However, most studies on Big Data technology only look at the benefits of employing accessible analytics data for their business rather than the readiness of technology for businesses [16,21]. The aim 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 is collected at a high velocity [22].

Many researchers have defined that massive factors influence Big Data technology readiness. However, like any technology, these factors may differ from one organization and industry to another. [3,8,9], including; Scalability, ICT infrastructure, security information, Machine learning management Support, size of organization, availability of finance, competitive pressure, organization demand, and application as well as analytic tools [17,23].

Big Data technology offers several options. This technology has numerous applications, such as waste minimization via design is the future of waste management research [24] and other concepts to adopt from this technology, such as Big Data with BIM, clash detection and resolution, performance prediction, and so on [25]. The building sector has been enabling a digital revolution to take place. They are implementing Big Data technology in Building Information Modelling (BIM) by handling building project data. BIM data is often 3D geometrically encoded, computationally intensively compressed, in a variety of proprietary formats, and interconnected. Accordingly, the data are enriched gradually and persisted, although project life cycle that make BIM files can quickly get voluminous, with the building model easily achieving fifty gigabytes in size [26,27].

## **3. Big data in Thai government**

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, among other things. Government procurement is said to be the primary source of funding for public services and government entities [28]. As a result, the sole factor restricting government spending is the quality of the procurement analysis. However, effective procurement does not always imply paying the lowest price. Instead, the acquisition is motivated mostly by the desire to advance the nation's technology and industry. [29]. Procurement has expanded rapidly in many countries to support the rise of national and international economies. Procurement is important because it assists the essential and enabling parts of the public and commercial sectors. As a result, in order to become

more flexible, the public procurement system must rapidly develop and adapt. The evolving missions of the public, corporate, and civil society sectors have resulted in 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 provide assistance to the private sector so that both sectors can benefit the country and progress indefinitely. As a result, in this day and age, information technology is essential in both the public and private sectors, and it 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 the annual building expenditure is approximately \$2 billion [32]. However, present government procurement methods fall well short of adequately meeting the demands of stakeholders. According to research and the media, it is the result of a variety of procurement issues, including corruption in government construction procurement projects, procurement costs that are 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 outline a detailed scope of work in every area [28,34], and the project can also be handled by hiring contractors 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 is in charge of developing standard pricing for all government construction projects in order to enter the data into the electronic government procurement (e-GP) system [14]. The government would handle the procurement process for each project as the project owner after setting the budget based on the proposal presentation [14]. The incorrect project cost is also such an important piece of information that it enhances the risk of a disagreement [5,13,35]. As a result, cost estimate is critical for government personnel who are involved in operations [6]. According to government data, it is still impossible to achieve a price that is greater than the project budget. Furthermore, the project's usual cost surpasses the project budget, affecting how government funds are administered. It is the cause of the excess and lost value gaps [1,7]. As a result, machine learning classification approaches may be utilized to create solutions to understand the behavior and investigate the influence of disagreement in standard pricing estimates [2,36].

#### 4. Algorithm of machine learning in this research.

The purpose of machine learning (ML), a branch of artificial intelligence (AI), is to enable computer systems to learn about a certain job automatically from data. A number of 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 methodology [39].

##### 4.1 Artificial neural network algorithm (ANN)

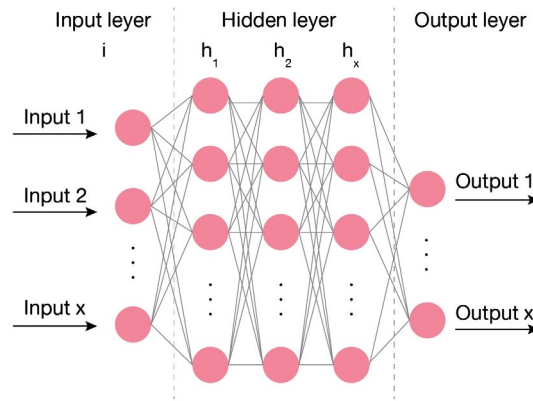
There are various types of artificial neural networks (ANNs). Classification and function estimation are ideal applications for Artificial Neural Networks (ANNs). These algorithms have been widely employed to solve difficult industrial issues since their inception. The most common type of ANN is the multi-layer perceptron (MLP). An ANN is made up of three layers: the input layer, the hidden (intermediate) layer, and the output layer. Through deep learning, ANN algorithms have recently changed machine learning. All the ANN applications used in the construction sector demand special attention, and new ANN algorithms are being developed to learn from data with a huge dimensionality (i.e., Big Data) [40]. Equation 1-2 should be used to refer to the hidden units in the neural network model that are shown in Figure 1 [41].

$$f(x) = \beta_0 + \sum_{k=1}^K \beta_k h_k(X)$$

$$= \beta_0 + \sum_{k=1}^K \beta_k g(w_{k0} + \sum_{j=1}^p w_{kj} X_j) \quad (1)$$

It is built up here in two steps. First the  $K$  activations  $A_k$ ,  $k = 1, \dots, K$ , in the hidden layer are computed as functions of the input features  $X_1, \dots, X_p$ ,

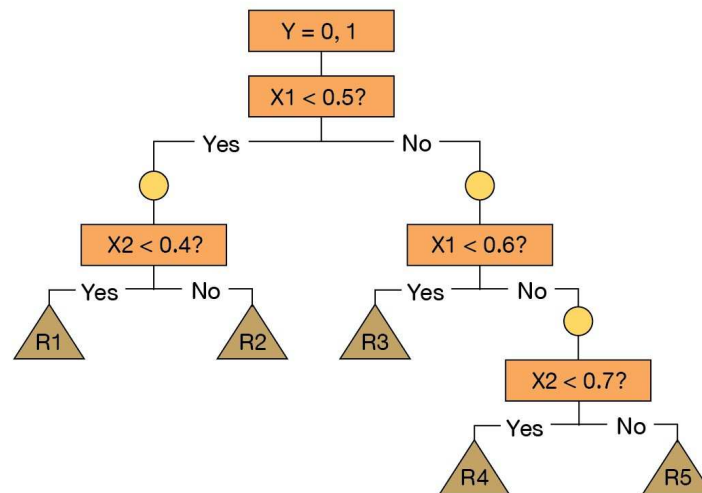
$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^p w_{kj} X_j) \quad (2)$$



**Figure 1.** the structure of ANN.

#### 4.2. Decision tree algorithm

Frameworks for making decisions. The latest machine learning approach for predicting qualitative and quantitative goal variables is decision trees (DTs). The creation of DT begins with the discovery of the decision node. The nodes are then separated recursively until no further divisions are feasible. Two metrics used to test the robustness of DT, which is dependent on the logic used to split nodes, are information gain (IG) and entropy reduction [40]. exemplifies a simple decision tree model with a single binary goal variable  $Y$  (0 or 1) and two continuous variables  $x_1$  and  $x_2$ , all of which span from 0 to 1. Figure 2 depicts the basic components of a decision tree model: nodes and branches. [41], and splitting, pausing, and pruning are the key modeling procedures.



**Figure 2.** Sample decision trees based on binary target variable  $Y$ .

#### 4.3. K-Nearest Neighbors

KNN is a non-parametric approach used in regression and classification. KNN has been used successfully for classification in a variety of applications. The separation between each item in a training set and each item in a test set is calculated using this approach; keep in mind that the  $K$  items



in the training set are the KNN. 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 further distant than the Kth nearest neighbors, resulting in a vote total larger than K.) A few scientists have considered the most precise way of measurement. They have employed global and local measurements in their initiatives, although these indicators are problem specific. Until now, the most widely used metric has been 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) between [42] and [43].

## 5. Algorithm of machine learning in this research.

Artificial intelligence (AI) is one of the major technologies driving 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 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. Machine learning, then, is a system's capacity to learn from data. It can make judgments based on previous experience in comparable situations, deal with ambiguity and limited data, and more [45]. A number of research on the application of machine learning in construction and project management have been conducted in recent years, although they are still quite restricted. They employed machine learning to classify construction papers 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 present construction information management system's quality. The proposed technique may be used in projects to assist diverse stakeholders in identifying and mitigating the risk of legal change conflicts [46,47]. The cost and timeline of building projects may be calculated using an artificial neural network (ANN) and other models [48]. According to the findings, early planning is critical to the success of a project. They created a text classification approach based on machine learning to classify contract general conditions provisions. This method makes it easier to assess the automatic 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. Their 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 has a procurement process management philosophy for building projects [51]. Furthermore, the system may enable developing technologies, which have a proclivity to learn from past events in order to address current challenges and improve enterprises [52]. Data collecting from government research efforts has also yielded some unusual findings. To verify whether the data obtained is acceptable for ML, the Thai government agency in charge of procurement must produce and discover data anomalies [53].

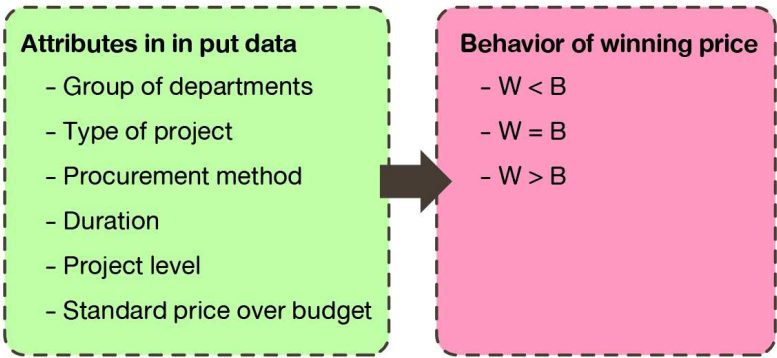
The previous research acquired data from Thailand's conventional government system. The four data characteristics are department name, site location, procurement method, and project type. The authors developed an ML-based model for anticipating over-budget projects. The produced model, which had an accuracy of 0.86, was built using the KNN approach. This study demonstrated that some of the Thai government's obsolete data may be used to develop cutting-edge technologies such as machine learning. ML has a progress goal, even if it is a component of big data for processing massive amounts of data. Finally, the study could show that big data and machine learning methodologies can successfully use even a little quantity of data from the Thai government [54], and this study used a large amount of data from the e-GP system. The information was gleaned from more than 200,000 construction projects in Thailand. Hyperparameter.

The optimization of hyperparameters can be simplified by determining how many function evaluations will be performed on each optimization in order to identify the optimal hyperparameter in that model. Furthermore, optimization may be defined as "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 the hyperparameter, the author will accept input as a function to the hyperparameter model and output as a measurement of model performance [56]. This is the rate of miscalculation or mistake. The hyperparameter space contains all of the potential values that are often established as acceptable boundaries for each hyperparameter, and the number of hyperparameters equals the function's dimension [58].

According to prior research, adjusting the hyperparameter necessitates understanding of the link between the settings and model performance. It 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 optimizing is to reduce the number of trials on hyperparameters while identifying the best model [59]. As a result, the author might regard the process as sequential rather than parallel.

6. Conceptual framework

The conceptual underpinning of this study is the procurement practices of the Thai government in a government building project. The 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 3. While W meaning winning price and B meaning project budget.



*Note: W meaning winning price and B meaning project budget.*

Figure 3. Conceptual framework.

7. Research methodology

The computerized government procurement system was used to collect the data for this investigation (e-GP). The data may be accessed with authorization from the comptroller general's office so that it could be gathered. The required data includes all 2019 government construction projects, with a total of about 283,000 projects, including the three scenarios that were the subject of this study: standard price over project budget, winning price over standard price, and winning price over project budget. Da-ta was separated into two categories based on how it was collected: 20% of the data was used for model validation, and 80% of the data was utilized for model training. Additionally, an Artificial Neural Network (ANN) was used to assess the training da-ta, and the results were validated using a confusion matrix [2,59]

7.1. Application of machine learning

The collected data had to be structured in a CSV file to ensure there were no blank values or categories with unknown contents. ANN also requires the use of computer software. The buried layer of the ANN is 100 [60] in size. Anaconda software was used to run the Python-based computer application.

### 7.2. Verifying the Model

Building a confusion matrix and using the following equations (3-5) allowed the ANN model's accuracy, precision, and recall to be checked [61]. where TP, TN, FP, and FN represent the true, false, and positive states, respectively.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (3)$$

$$\text{Precision} = TP/(TP+FP) \quad (4)$$

$$\text{Recall} = TP/(TP+FN) \quad (5)$$

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 to obtain consistent results from a variety of measurements. In information retrieval, random error is a sort of observational error that results in differences between precise values. [59].

### 7.3. Hyperparameter optimization with random search

Machine learning models contain hyperparameters that must be set in order for the model to be customized to your dataset. The general effects of hyperparameters on a model are frequently understood, but determining how to optimally set 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 different values for model hyperparameters and select a subset that results in a 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. The result of a hyperparameter optimization is a single set of well-performing hyperparameters that you can use to configure your model [56].

Hyperparameters are points of choice or configuration that allow a machine learning model to be tailored to a given job or dataset. Model configuration argument 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; 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 being the values that the hyperparameter can take on, such as real-valued, integer-valued, or categorical. The Random Search defines a search space as a bounded domain of hyperparameter values that is randomly sampled [56].

### 7.4. Data Collection

Seven characteristics of the input data gathered from the conventional Thai government system are displayed in Table 1 below. According to its definition, each parameter has a unique factor and the first parameter is the groups of department name attribute, which includes 13 groups of departments in Thailand including University School Hospital Irrigation Public works and town and country planning Highway Rural road Finance Local of administration Justice Police Soldier and other.

Furthermore, the project owner is one of the parties involved in building project conflicts [13,15]. Furthermore, these are 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 project cost estimation, budgeting, and procurement [21]. The Thai procurement method characteristic, as shown in Table 3, consists of three components, including the chosen bidding method and the specific methodology. Furthermore, each country has its own procurement system, but the final aim is the same: the abolition of corruption [18].

It has a considerable impact on a contractor's cost when using the procurement technique [14,18]. Project levels also come in five levels. According to the price ranges, the levels are divided as illustrated in Table 4 [3,62]. The length of a construction project as specified by the contract is referred to as "duration." In the case of project qualities, the question of whether a project is under or over budget considering Thai government policy is considered. It is made up of two components.

**Table 1.** Attributes in input data

No.	Attributes	Factor
1	Project owner epartments	13
2	Type of construciton project	3
3	Bidding method	3
4	Duration	5
5	Project level	5
6	Standard price over budget	3
7	Winning price over budget	3

**Table 2.** Thai procurement method.

Method of procurement	Detail
Bidding	Every firm was welcome to join and evaluate the initiatives.
Chosen	Only the qualifying firm could submit a proposal with specific project requirements.
Specific	The contractors might be chosen by the proprietors on their own.

**Table 3.** Project level.

Project level	Detail (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

## 8. Result

Three components make up the study's findings: general information, a machine learning model, and a model that has been verified.

### 8.1. General Information.

The specific approach, which accounts for 77.8% of all procurement methods, is the one that the Thai government prefers the most. The largest group of departments is classified as "Other," which accounts for 67.5 percent of Thai government data in 2019. The departments with the highest budgets, however, are local government and highways, as seen in Figure 4. Roads and buildings are the main construction projects undertaken by the Thai government. As seen in Figure 5, one of the numerous concerns that the Thai government is most worried about for the Thai people is infrastructure [7,14]. This study classified the budget situation into two categories, standard price over budget and winning price over budget, as shown in Figures 6 and 7. This study aims to predict the winning price

in a budgetary circumstance. A few successful projects have prices that are greater than anticipated due to the overall formation. This is an important aspect of the study, though, as police are prone to corruption [7]. The government's cost evaluation process is defective since there are 700 projects where the winning price is higher than the standard price [5,6]

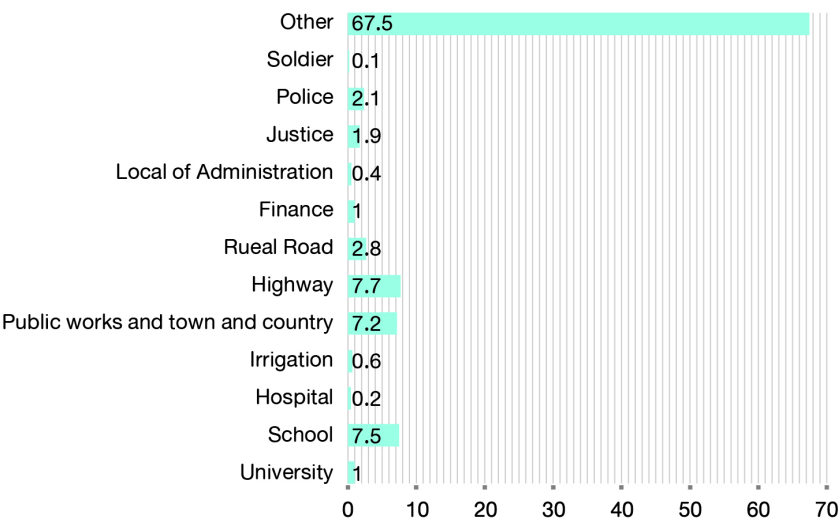


Figure 4. Group of departments in Thailand.

Table 4. Value of departments in Thailand.

Method of procurement	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
Highway	2,761,913,958	20.78
Rural road	1,131,877,654	8.51
Finance	15,551,811	0.12
Local of administration	2,946,100,502	22.16
Justice	560,697,143	4.22
Police	290,890,893	2.19
Soldier	336,971,946	2.53
Other	2,195,136,360	5.63
Sum	13,294,211,568	100.0

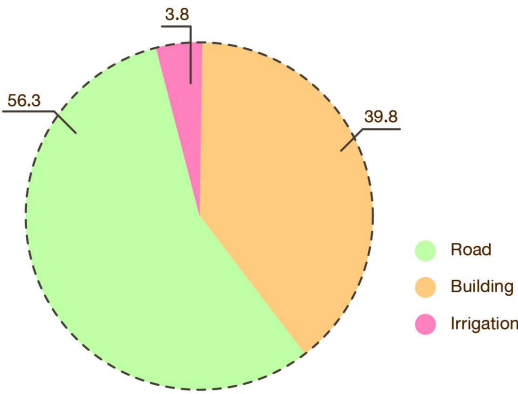


Figure 5. Percentage of project type.

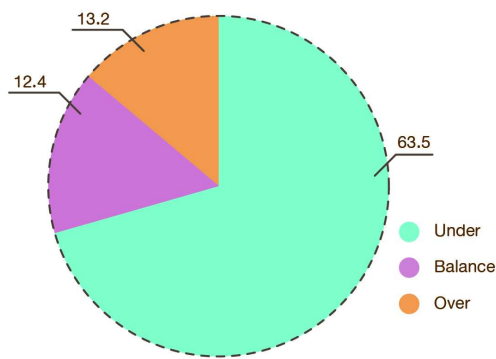


Figure 6. Standard price over budgeted.

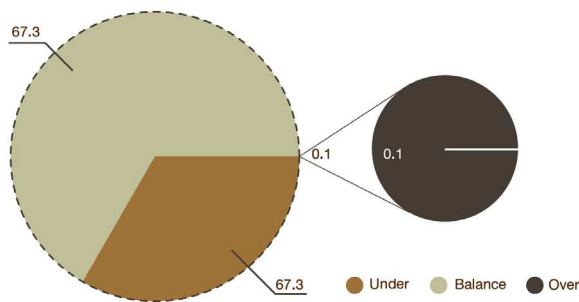


Figure 7. Winning price over budgeted.

8.2. Machine learning model

There are 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, while the ANN algorithm has the greatest percentage at 77.60. According to accuracy for beneficial, it must be greater than 75% [63].

Table 5. Percentage of project type.

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

8.3. Validating data with confusion matrix

The confusion matrix [64] is used to calculate the classification accuracy of the model. The ANN model matrix revealed that the model properly predicted 44,028 out of 56,705 instances. As a result, as shown in Figure 8, the grey box is misclassified and the white box is correctly classified, and the number zero in the confusion matrix table indicates that the model did not make a mistake in forecasting for each scenario. Similarly, the accuracy of the ANN model may be determined using the confusion matrix. As indicated in Table 6, accuracy may be split into three categories of bidding behavior (under, balance, and above). For the first case, under cluster, the model achieved a precision of 83%. Balance cluster, the model achieved a precision of 77%. Over cluster, the model did not achieve a precision score. as shown in Table 6.

**Table 6.** Confusion matrix of ANN model.

	<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>support</b>
Under	0.83	0.41	0.55	18,704
Balance	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

		<b>Actual class</b>		
		<b>Under</b>	<b>Balance</b>	<b>Over</b>
<b>Predicted class</b>	<b>Under</b>	7,632	11,072	0
	<b>Balance</b>	1,561	36,393	0
	<b>Over</b>	7	40	0

**Figure 8.** Confusion matrix table of ANN model.

The matrix of decision tree model showed that the model correctly predicted for 44,050 from 56,705 cases. Therefore, the grey box is misclassified and the white box is correctly classified as show that in Figure 9 and the number zero in confusion matrix table means that the model didn't make a mistake in predicting for each case. Similarly, the decision tree model's precision can also be calculated by using the confusion matrix. The precision can be divided into three cases of bidding's behaviours (i.e., Under, Balance, and Over. For the first case, Under cluster, the model achieved a precision of 81%. Balance cluster, the model achieved a precision of 77%. Over cluster, the model did not achieve a precision score. as shown in the Table 7.

**Table 7.** Confusion matrix of Decision tree model.

	<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>support</b>
Under	0.81	0.41	0.54	18,595
Balance	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

		<b>Actual class</b>		
		<b>Under</b>	<b>Balance</b>	<b>Over</b>
<b>Predicted class</b>	<b>Under</b>	7,572	11,022	1
	<b>Balance</b>	1,762	36,298	0
	<b>Over</b>	16	34	0

**Figure 9.** Confusion matrix table of Decision tree model.

The matrix of KNN model showed that the model correctly predicted for 42,549 from 56,705 cases. Therefore, the grey box is misclassified and the white box is correctly classified as show that in Figure 10 and the number zero in confusion matrix table means that the model didn't make a mistake in predicting for each case. Similarly, the KNN model's precision can also be calculated by using the confusion matrix. The precision can be divided into three cases of bidding's behaviours (i.e., Under,

Balance, and Over). For the first case, Under cluster, the model achieved a precision of 66%. Balance cluster, the model achieved a precision of 78%. Over cluster, the model did not achieve a precision score. as shown in the Table 8.

The precision of confusion matrix has shown that the ANN algorithm could show the highest accuracy in all the case however KNN could perform high efficiency with ANN for Balance case. this point could prove that the traditional data have a relation for application data technology [2]. The performance of classification algorithms is typically assessed by evaluating the accuracy of the classification artificial neural networks may be utilized and good results can be achieved from classification algorithms [65]. as shown in Table 9. However, the over case cannot be processes because of the data that have a little and this case can happen in a few times. According to this, the procurement in Thailand is efficiency process.

**Table 8.** Confusion matrix of KNN model.

	<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>support</b>
Under	0.66	0.50	0.57	18,564
Balance	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

		Actual class		
		Under	Balance	Over
Predicted class	Under	9,370	9,194	0
	Balance	4,909	33,179	0
	Over	17	36	0

**Figure 10.** Confusion matrix table of KNN model.

**Table 9.** Precision accuracy of each algorithm.

Cases	Machine learning algorithm		
	ANN	Decision Tree	KNN
Under	83%	81%	66%
Balance	77%	77%	78%
Over	0%	0%	0%

#### 8.4. After hyperparameters tuning.

In the final experiments, underperforming hyperparameters were removed. The Random search technique randomly samples the hyperparameter space. According to [66], random search has more advantages than grid search in terms of application that can be used even if the computer cluster fails. It enables practitioners to adjust the "resolution" on the fly, as well as add additional trials to the set or even disregard the fail test. At the same time, the random search procedure may be stopped at any time, forming a full experiment that can be carried out concurrently [67]. Furthermore, if more computers become available, a new trial may be added to the experiment without compromising it [68]. The following are the primary parameters for each model. The main parameters for each model are as follows. ANN model : random\_state = 42, hidden layer sizes = 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 show that in the the 11.

Table 10. 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%

5. Discussion

The method of procurement attribute in the input data is crucial for increasing the model's predictive performance [14]. One of the three methods of procurement used by the Thai government was created specifically for small-scale projects like the reinforcing of concrete roadways and tiny buildings. The special technique was created by the government for preferred contractors, whose ability to finish projects quickly could be guaranteed by the government. This approach has the advantage of delivering built amenities to people more quickly than the bidding technique. Government officials can therefore participate in a variety of activities during a particular method as one of the elements influencing the procurement process, provided that their actions are not monitored and recorded [69]. The procurement regulation does, however, include a shortcoming that hurts the government's reputation. This procurement approach has a fault in that it is only appropriate for modest projects. Furthermore, only government agencies have the authority to choose contractors in a direct, straightforward manner. Project auditing is a department that exists, although it cannot audit every single project. Another study indicates that a key contributing factor to corruption is a government agency [3].

The application's goal, such as storage to support data technology to collect data, depends on the data technology infrastructure [12]. This study makes sure that the Thai government's data collection procedures are effective and of high quality. The building industry, however, might gain greatly from the adoption of data technology and Big Data technologies, and this would be a developed nation [2]. They should, however, design their digital data collection in such a way that it supports the Big data concept and is a developed country with current technology.

9. Conclusions

This study demonstrates that the ANN's accuracy rate was 78.9 percent. In the current analysis, data from Thailand's conventional government system were employed. Group of Departments, Project Type, Procurement Method, Duration, Standard Price Over Budget, Winning Price Over Standard Price, Standard Price Over Budget, and Winning Price Over Standard Price are among the eight characteristics. The effectiveness of four algorithms showed that the data may be utilized to accurately forecast behaviour.

The budgeting process is one of important for government construction project management that need to be finish before auction process. The auction in procurement management in government define project budget early and a budget for a construction project is crucial since it may assist to save time and money over the course of the project. You will be able to take the necessary actions to guarantee that the building project stays within the budget after you have properly planned for it.

Finally, this study can demonstrate that the Thai conventional data gathering technique can be used with machine learning from Big data. According to this, the data used to create machine learning models is raw data obtained from the procurement system. However, even though we did not change or tweak the settings, the outcome was still better. If they enforce their policy to enhance data gathering techniques, the data will be more efficient and productive.

**Author Contributions:** Conceptualization, W.K. and K.S.; methodology, W.K. and K.S.; software, W.K.; validation, W.K.; formal analysis, W.K.; investigation, W.K.; resources, K.S.; data curation, K.S.; writing original

draft preparation, W.K. and K.S.; writing review and editing, T.C; visualization.; supervision, K.S.; project administration, K.S. and T.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data used to support the findings of this study are available from the corresponding author upon request.

**Acknowledgments:** We would like to sincerely thank the faculty of engineering, Khon Kaen University for their funding support. In addition, we would like to thank the comptroller general's department for allowing their data to be used in this study. Without any of them, our research would not be accomplished.

**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. *Journal of Advances in Information Technology* 2021, 12, 1–5, doi:10.12720/jait.12.1.1-5.
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. *Advanced Engineering Informatics* 2016, 30, 500–521, doi:10.1016/j.aei.2016.07.001.
3. Chaitongrat, T.; Leungbootnak, N.; Kusonkhum, W.; W, D.; S, L.; Srinavin, K. Measurement Model of Good Governance in Government Procurement. *IOP Conference Series: Materials Science and Engineering* 2019, doi:10.1088/1757-899x/639/1/012024.
4. Soni, S.; Pandey, M.K.; Agrawal, S. Conflicts and Disputes in Construction Projects: An Overview. *International Journal of Engineering Research and Applications* 2017, 07, 40–42, doi:10.9790/9622-0706074042.
5. Jaffar, N.; Tharim, A.H.A.; Shuib, M.N. Factors of Conflict in Construction Industry: A Literature Review. *Procedia Engineering* 2011, 20, 193–202, doi:10.1016/j.proeng.2011.11.156.
6. Rose, K. A Guide to the Project Management Body of Knowledge (PMBOK® Guide)-Fifth Edition. *Project Management Journal* 2013, 44, e1, doi:10.1002/pmj.21345.
7. Chaitongrat, T. CAUSAL RELATIONSHIP MODEL OF PROBLEMS IN PUBLIC SECTOR PROCUREMENT. *International Journal of GEOMATE: Geotechnique, Construction Materials and Environment* 2021, 20, doi:10.21660/2021.80.6266.
8. Hurwitz, J.; and Kirsch, D.; *Machine Learning for Dummies*, IBM Limited Edition. 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. *Automation in Construction* 2019, 105, 102833, doi:10.1016/j.autcon.2019.102833.
10. Naganathan, H.; Chong, W.C.; Chen, X.-W. Building Energy Modeling (BEM) Using Clustering Algorithms and Semi-Supervised Machine Learning Approaches. *Automation in Construction* 2016, 72, 187–194, doi:10.1016/j.autcon.2016.08.002.
11. Poh, C.Q.X.; Ubeynarayana, C.U.; Goh, Y.M. Safety Leading Indicators for Construction Sites: A Machine Learning Approach. *Automation in Construction* 2018, 93, 375–386, doi:10.1016/j.autcon.2018.03.022.
12. Chonpitakwong, B.; Kusonkhum, W.; Chaitongrat, T.; Srinavin, K.; Charnwasununth, P. Hindrance of Applying Big Data Technology for Construction Management in Thai Government. *Journal of Advances in Information Technology* 2021, doi:10.12720/jait.12.2.159-163.
13. Jervis, B.M.; Levin, P.T. *Construction Law, Principles and Practice*; 1988;
14. The Comptroller General's Department. *The Government Procurement and Supplies Management Act B.E. 2560*. The Comptroller General's Department. 2017.
15. Deal, J.L. *Information Shift Big Data: 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 Affairs* 2014, 33, 1300, doi:10.1377/hlthaff.2014.0581.
16. Michael, K.; Miller, K. Big Data: New Opportunities and New Challenges [Guest Editors' Introduction]. *IEEE Computer* 2013, 46, 22–24, doi:10.1109/mc.2013.196.
17. *Big Data Analytics*; Springer Science+Business Media, 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*, 2023; pp. 1309-1317.
19. Anand, R.J.D., March. More Data Usually Beats Better Algorithms. 2008, 24.
20. Eadie, R.; Browne, M.; Odeyinka, H.; McKeown, C.; McNiff, S.J.A.i.c. BIM implementation throughout the UK construction project lifecycle: An analysis. 2013, 36, 145-151.
21. Insights, F.J.R.f.w.f.c.f. *Betting on Big Data: How the right culture, strategy and investments can help you leapfrog the competition*. 2015.

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, 2013; pp. 995-1004.
23. Wielki, J. Implementation of the big data concept in organizations-possibilities, impediments and challenges. In Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, 2013; pp. 985-989.
24. Osmani, M.; Glass, J.; Price, A. Architect and contractor attitudes to waste minimisation. In Proceedings of the Institution of Civil Engineers-waste and resource management, 2006; pp. 65-72.
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. In Computing in Civil Engineering (2013); 2013; pp. 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 Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2014; pp. 283-288.
27. Lin, J.R.; Hu, Z.Z.; Zhang, J.P.; Yu, F.Q.J.C.A.C.; Engineering, I. A natural-language-based approach to intelligent data retrieval and representation for cloud BIM. 2016, 31, 18-33.
28. Dzuke, A.; Naude, M.J.J.o.T.; Management, S.C. Procurement challenges in the Zimbabwean public sector: A preliminary study. 2015, 9, 1-9.
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, 2011; pp. 1364-1368.
30. Mark McKevitt, D.; Davis, P.J.I.J.o.P.S.M. Supplier development and public procurement: allies, coaches and bedfellows. 2014, 27, 550-563.
31. Leungbootnak, N.; Chaithongrat, T.; Aksorn, P. An exploratory factor analysis of government construction procurement problems. In Proceedings of the MATEC Web of Conferences, 2018; p. 02057.
32. Tanayut, C.; Narong, L.; Preenithi, A.; Patrick, M.J.I.T.J.o.E., Management; Sciences, A.; Technologies. Application of Confirmatory Factor Analysis in Government Construction Procurement Problems in Thailand. 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, 2007; pp. 1337-1341.
34. Burke, R. Project management: planning and control techniques; John Wiley & Sons: 2013.
35. Chitkara, K. Construction Project Management-Planning, Scheduling and Controlling, Tata McGraw Hills. 2011.
36. Maemura, Y.; Kim, E.; Ozawa, K.J.J.o.C.E.; Management. Root causes of recurring contractual conflicts in international construction projects: Five case studies from Vietnam. 2018, 144, 05018008.
37. Diekmann, J.E.; Kruppenbacher, T.A.J.J.o.C.E.; Management. Claims analysis and computer reasoning. 1984, 110, 391-408.
38. Kim, M.P. US Army Corps Engineers construction contract claims guidance system. In Proceedings of the Utilization of Ocean Waves—Wave to Energy Conversion, 1989; pp. 203-209.
39. Chau, K.-W. Prediction of construction litigation outcome—a case-based reasoning approach. In Proceedings of the Advances in Applied Artificial Intelligence: 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2006, Annecy, France, June 27-30, 2006. Proceedings 19, 2006; pp. 548-553.
40. Atuahene, B.T.; Kanjanabootra, S.; Gajendran, T.J.C.M.; Economics. Transformative role of big data through enabling capability recognition in construction. 2023, 41, 208-231.
41. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An introduction to statistical learning; Springer: 2013; Volume 112.
42. Breiman, L. Random Forests--Random Features. - 1999.
43. Zhang, Z.J.A.o.t.m. Introduction to machine learning: k-nearest neighbors. 2016, 4.
44. Canhoto, A.I.; Clear, F.J.B.H. Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. 2020, 63, 183-193.
45. Chen, J.-H.J.A.i.C. KNN based knowledge-sharing model for severe change order disputes in construction. 2008, 17, 773-779.
46. Xie, S.; Fang, J.J.I.J.o.I.S.; Management, C. Prediction of construction cost index based on multi variable grey neural network model. 2018, 10, 209-226.
47. Salama, D.M.; El-Gohary, N.M.J.J.o.C.i.C.E. Semantic text classification for supporting automated compliance checking in construction. 2016, 30, 04014106.
48. Elfahham, Y.J.A.E.J. Estimation and prediction of construction cost index using neural networks, time series, and regression. 2019, 58, 499-506.
49. NGUYEN, P.T.; NGUYEN, Q.L.H.T.T.J.T.J.o.A.F., Economics; Business. Critical factors affecting construction price index: An integrated fuzzy logic and analytical hierarchy process. 2020, 7, 197-204.

50. Lin, W.-C.; Ke, S.-W.; Tsai, C.-F.J.K. Top 10 data mining techniques in business applications: a brief survey. 2017.
51. Cheng, M.-Y.; Peng, H.-S.; Wu, Y.-W.; Chen, T.-L.J.A.i.C. Estimate at completion for construction projects using evolutionary support vector machine inference model. 2010, 19, 619-629.
52. Cost, S.; Salzberg, S.J.M.I. A weighted nearest neighbor algorithm for learning with symbolic features. 1993, 10, 57-78.
53. Roy, R.; Low, M.; Waller, J.J.C.M.; Economics. Documentation, standardization and improvement of the construction process in house building. 2005, 23, 57-67.
54. Kusonkhum, W.; Srinavin, K.; Leungbootnak, N.; Aksorn, P.; Chaitongrat, T.J.J.o.A.i.I.T.V. Government construction project budget prediction using machine learning. 2022, 13.
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), 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 Proceedings of the 28th International Conference on International Conference on Machine Learning, 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), 2015; pp. 1-10.
58. Hazan, E.; Klivans, A.; Yuan, Y.J.a.p.a. Hyperparameter optimization: A spectral approach. 2017.
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), 2015; pp. 7234-7237.
60. Allen, K.; Berry, M.M.; Luehrs Jr, F.U.; Perry, J.W.J.A.D. Machine literature searching VIII. Operational criteria for designing information retrieval systems. 1955, 6, 93.
61. Gondia, A.; Siam, A.; El-Dakhkhni, W.; Nassar, A.H.J.J.o.C.E.; Management. Machine learning algorithms for construction projects delay risk prediction. 2020, 146, 04019085.
62. Suntharanurak, S. Screening for bid rigging in rural road procurement of Thailand. Doctoral dissertation, National Institute of Development Administration, 2012.
63. Samui, P.; Roy, S.S.; Balas, V.E. Handbook of neural computation; Academic Press: 2017.
64. Lu, B.; Hardin, J. Constructing prediction intervals for random forests. Ph. D. thesis, Pomona College, 2017.
65. Tang, L.; Zhao, Y.; Cabrera, J.; Ma, J.; Tsui, K.L.J.I.T.o.I.T.S. Forecasting short-term passenger flow: An empirical study on shenzhen metro. 2018, 20, 3613-3622.
66. Bergstra, J.; Bengio, Y.J.J.o.m.l.r. Random search for hyper-parameter optimization. 2012, 13.
67. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B.J.A.i.n.i.p.s. Algorithms for hyper-parameter optimization. 2011, 24.
68. 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, 2013; pp. 115-123.
69. Olson, D.L.; Delen, D. Advanced data mining techniques; Springer Science & Business Media: 2008.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.