Modern Education
and Computer Science
PRESS

Folding Bicycle Prospective Buyer Prediction Model

Trianggoro Wiradinata
Universitas Ciputra Surabaya, Surabaya, 60219, Indonesia
E-mail: twiradinata@ciputra.ac.id

Abstract: The trend of bicycle exercise during the pandemic has resulted in increased sales and even scarcity of bicycle stock in some shops. The phenomenon has raised attention from both the bicycle industry and government to provide necessary responses toward the trends. Even though it is a trend, many prospective buyers are still confused about their choices. The types of bicycles that sell the most on the market are folding bikes, mountain bikes, and racing bikes. The research data were collected from 242 bicycle users who came from various bicycle communities in major cities of Java Island, Indonesia. Some of the predictors used were age, gender, height, weight, and cycling speed. The target variable is the type of bicycle whose data is categorical. Predictor variables consist of nominal and ordinal variables, so preprocessing needs to be done using Python's Sklearn library. To test the accuracy of the model, the data was broken down into training data and test data with a test size of 20%. Several methods are used to form a classification model, including K-NN, Naive Bayes, Support Vector Machine, Decision Tree, and Random Forest. The results of the classification model evaluation show that the Support Vector Machine and Decision Tree have the highest accuracy of 90%, while Naive Bayes has the lowest accuracy of 73%. The model formed can be a predictive tool for potential bicycle buyers in order to be able to choose the right type of bicycle.

Index Terms: Bike, classification, data science, knn, naive bayes, support vector machine, decision tree, random forest

1. Introduction

Cycling as both sport and recreation has become more and more popular among the public during the coronavirus pandemic in 2020. There are many people who buy bicycles both through bicycle shops that sell offline and online bicycle shops that sell with various conveniences and at discounted prices. This was triggered by the trend during the Covid-19 pandemic, where people realized that maintaining body condition is very important so that they do not easily get infected by various types of diseases. For individuals, cycling either as exercise or recreation is an excellent form to develop physical fitness and minimize the risk of health problems [1].

Data shows that search interest in Google Trends for folding bikes has increased by 900% from March 1 to June 21, 2020. Further data showed, there were mountain bikes with an increase in search interest of up to 680% from the same period. Road bikes or racing bicycles is in the last rank with a 300% increase in searches over the same period [2]. From this data it can be observed that the most searches are triggered by folding bikes as keywords. People who are interested in folding bikes mostly have several considerations in mind, such as it is small, lightweight, and if folded it can be easily transported in other vehicles such as cars or buses for longer trips. Another benefit is that if there is a shortcut that cannot be passed by the vehicle (such as a pedestrian bridge), the folding bike can be lifted and continue the journey again afterwards.

Even though there are many types of bikes in the market, for beginners looking to buy a bicycle, choosing the best bike that fits their needs can be a difficult task. Physical condition, age, and propensity to ride a bike varies from person to person. Furthermore, with the increasing trend of purchasing bicycles, it is undeniable that the price of bicycles is also increasing; therefore it is necessary to be careful when buying a bicycle so as not to make the wrong choice which results in disappointment.

In general, before buying a bicycle, people will read reviews, but it is quite often for reviews to be found on the internet that causes confusion because they are sometimes written by people who are less experienced in choosing a bicycle or sometimes biased towards a certain type of bicycle. This study aims to contribute to a model that was formed inductively by using a machine learning algorithm from a dataset collected from experienced users from several bicycle communities so that the answer is based on expert judgment. Through this machine learning model, it is hoped that beginners who want to buy a bicycle can make careful decisions based on certain characteristics.

1.1. Bicycle Industry and Government Support

Global data during the pandemic shows that the bicycle industry around the world has experienced a drastic increase in the number of sales. The increase in demand for bicycles in the United States, from March to June 2020, increased by 40 percent. In the UK, personal bicycle use has increased by 33 percent. Meanwhile, bike sharing increased by 12 percent, likewise in France, which can be seen from the increasing budget for the provision of bicycle parking facilities. Initially, 20 million euros became 60 million euros in May 2020. These data show that bicycles are currently an alternative transportation option during a pandemic. Bicycles are now an option for people to carry out daily activities and maintain health by exercising [3].

Conditions in Indonesia show a more extreme condition where the demand for bicycles in Indonesia has increased by almost 1000 percent. This momentum was used by the Indonesian government to eliminate people's dependence on the use of motorized vehicles. Bicycles are not only a means of transportation but also as an economic driver for their daily needs. Many people even carry out selling activities using bicycles.

Government support through the Ministry of Transportation issued a Minister of Transportation Regulation Number 59 of 2020 concerning Safety of Cyclists on the Road. It is designed with the aim of providing protection to cyclists on the road, as has been done in many other countries.

1.2. Bicycle Types

In general, the types of bicycles sold in Indonesia consist of three types, namely road bikes, mountain bikes and folding bikes [4]. The following is a table of characteristics of the types of bicycles that are commonly sold without considering price variables. For each type of bicycle, the price varies from cheap to very expensive.

Table 1. Type of Bikes

Bike Type	Terrain	Speed	Mobility
Road Bike	On-road	Fast	Low
Mountain Bike	Off-road	Medium	Medium
Folding Bike	On-road	Slow	High

1.3. Problems in choosing bicycle types

In choosing the right type of bicycle, the consideration of potential buyers is not only cycling terrain, size, speed, or practicality, but often also related to the profile of the prospective buyer, which includes budget, age, gender, and body posture so that cycling can be comfortable and furthermore to avoid injury as suggested in [5, 6]. Many entry-level bicycle buyers do not understand the characteristics of choosing the right bicycle resulting in disappointment after buying.

1.4. The importance of predictive models to choose the right bicycle

Often when buying bicycles offline at bicycle shops, inexperienced prospective buyers face a difficult time due to choices being limited because the shop usually only displays a portion of the bikes that the shop owner wants to sell immediately. Meanwhile, buying a bicycle online makes potential buyers unable to try or feel whether a bicycle is right for their use.

For this reason, a machine learning-based prediction model is needed to estimate the suitability of the type of bicycle needed as a tool so that prospective bicycle buyers can estimate whether folding bicycles or other bicycle models are more suitable.

This prediction model can be applied through a web interface where prospective buyers can enter their personal data and several other preferences that affect the proper classification of the type of bicycle. The prediction model made needs to be compared with several algorithms that are commonly used for classification. The selection of the appropriate algorithm to model the classification equation is decided based on the highest accuracy value.

2. Literature Review

2.1. Problems in choosing bicycle types

Previous studies focusing on bicycles as research objects were numerous. Zaki et al. [7] describe a new approach to the binary classification of two-wheeled road users at heavily mixed traffic intersections. Classification is carried out to differentiate between motorized and non-motorized bicycles. One of the important features that are observed is speed. Performance analysis shows an accuracy of up to 93%. This research linkage contributes to one of the features used for the classification of folding bicycle selection. Jaya Prada et al. [8] also investigated the classification of bike buyer by categorizing into levels, and the model prediction was made utilizing different kinds of algorithms such as Logistic Regression, KNN (K-Nearest Neighbor), Support Vector Machine, Decision Tree, and Random Forest. Performance

evaluation was done based on different performance metrics, and the findings show that Random Forest Classifier resulted in the highest accuracy of 0.86. Another study by Singh et al. [9] comparing the performance of Random Forest, Naive Bayes, and K-Nearest Neighbors algorithms suggests their study could be extended by researchers who need to analyze using parametric methods.

Dill & McNeil [10] support the idea of labelling or categorizing cyclists that have occurred for more than a century for a variety of purposes. They succeeded in examining the typology developed whereby cyclists could be divided into four categories: Strong and Fearless, Enthused and Confident, Interested but Caring, and No Way No How. Unlike some other typologies, this broadly referenced typology is intended to apply to all adults, regardless of their current cycling behaviour. Adults were classified into four types based on their stated level of riding comfort at different types of facilities, their interest in cycling more for transportation, and their physical ability to ride. Nearly all sample populations clearly fall into one of four categories. The relationship with the classification in choosing the type of bicycle is also related to age and gender. It's just that in their study, women and the elderly were not represented proportionally.

Mistakes in choosing bicycle types can have negative impacts on user satisfaction. The negative effects range from less comfortable riding a bicycle to a more severe impact, such as injury. A study conducted by Van Hoof et al. [6] stated that bicycle users often experience low back pain injuries caused by various factors, including improper cycling position. Another study by Uden et al. [5] showed there were groups of cyclists who are at risk of experiencing foot pain. These groups are those who ride with an attached foot-pedal interface and those who are 26 years of age or older. Their study highlights the importance of addressing the current knowledge gap regarding foot pain and cycling. Furthermore, the need to investigate effective interventions for this problem.

2.2. Machine Learning Algorithm

Machine learning is part of the discipline of artificial intelligence (Artificial Intelligence) which uses statistical techniques to inductively produce a model from a set of data with various purposes such as predictions, recommendations, and so on. In general, the stages of forming a model derived from this data set are carried out in three stages, namely preprocessing data, training models, and estimation evaluation Nazli et al. [11]. The process of having computer programs study tasks and analyze situations has greatly reduced the amount of human intervention required [12].

There are many machine learning algorithms with different efficiency and case specifications, therefore in various studies it is necessary to make models using various algorithms so that their accuracy can be compared. As an instant example, Mukhopadhyay et al. [13] explored how to recognize customers who might pull off from completing business transactions. The cycle in which one user leaves transaction from a company and switches to another is called churn. Machine learning algorithms can be used to predict customer churn and help describe big data using data visualizations. Prediction churn plays an important role in various service-based industries such as telecommunications, banking, and hospitality. Many companies try to develop ways to predict prospective clients, especially in the telecommunications sector. Therefore, identifying the factors that increase customer churn is very important. In addition, the use of machine learning is ubiquitous from the classification of fake news [14] to the prediction of student performance [15], and it is not rare to be used for forecasting of the diabetes pandemic [16].

2.2.1. K-Nearest Neighbor (KNN)

K-Nearest Neighbour (KNN) is one of the simplest classification methods using a lazy learning algorithm approach which takes less time to train [9]. The KNN approach straightforwardly predicts new samples classification by using the K-closest values from the training dataset. Unlike other algorithms, KNN construction is merely based on the individual values taken from the training dataset. In order to predict classification of a new sample, KNN pinpoints the closest data in its perimeter based on how many K defined. Several distance formulas are available to be used with KNN to measure the nearest neighbouring data such as euclidean, manhattan, minkowski, and several others.

2.2.2. Naive Bayes (NB)

The Naive Bayes Algorithm is popular as machine learning algorithm due to its simplicity in involving all predictors variables towards the prediction of target variable. It predicts future probabilities based on past empirical experiences, so it is known as Bayes' Theorem. The main feature of this Nave Bayes Classifier is a very strong (naive) assumption of the independence of each condition/event [17].

2.2.3. Support Vector Machine (SVM)

Support Vector Machine can be utilized for both classification and regression problems. In classification modeling, SVM has a clear mathematical concept compared to other classification techniques. SVM can also be used for linear or non-linear relationships. SVM aims to find the best hyperplane by maximizing the distance between classes. Hyperplane is a function that can be used to separate data between classes. SVM is capable of producing good performance of classification accuracy [18].

2.2.4. Decision Tree (DT)

The decision tree is one of the most widely used classification methods, because of its ease to interpret with the help of visualization. Decision tree is a prediction model using a tree structure or hierarchical structure. The concept of a decision tree is to break down data and form a decision tree by applying decision rules. Decision Tree Classifiers are known to have good accuracy [19].

2.2.5. Random Forest (RF)

The Random Forest algorithm [20] is an extended method of Decision Tree algorithm. If in the Decision Tree a sample of data whose class is not yet known is classified into existing classes through a tree-shaped flow chart, then Random Forest combines several subsets of the Decision Tree feature randomly into one model trained by the bagging method. The trees are built, not pruned. The random selection of predictor variables produces less correlation among the trees and usually results in lower error rate [9].

2.3. Model Evaluation

The resulting model from the classification algorithm can be evaluated by creating a confusion matrix from python sklearn library to calculate the model's accuracy. Accuracy is a very common performance measurement for classification and it is formulated as a ratio between correctly classified prediction data against test data divided by the total number of test data [21]. Confusion Matrix consists of a table with 4 different combinations of the predicted data and the actual test data of the target variables. The four results of the classification process on the confusion matrix are True Positive, True Negative, False Positive, and False Negative as shown in Figure 1.

Fig. 1. Confusion Matrix.

3. Research Methodology

3.1. Data Collection and Preprocessing

Data were collected from several bicycle sports communities located in major cities of Java Island, Indonesia. The measurement instrument was formed based on the results of the interview which stated that the factors that influenced the choice of the type of bicycle were age, gender, height, weight, cycling speed, brand, and price. Considering that the brands and prices can vary and are not suitable as predictors, in the further analysis process these two predictors are not used. Subsequently, all variables, both predictors and targets, used in this study are nominal and ordinal, this is intended as data collection strategy to make it easier for respondents when filling in they only need to answer the multiple-choice form. To ensure accuracy and reliability, it needs be understood that it is likely that these bicycle sports enthusiasts fill in between their cycling time, therefore the instrument design must be made as easy as possible.

The data collection process was carried out by means of a self-administered survey to cycling sports enthusiasts who are members of the bicycle communities. The results of the collection were 242 data entered and all of them will be used in the analysis. Due to all variable types being nominal and ordinal, it is necessary to encode the data so that it becomes numeric and can be analyzed further.

3.2. Exploratory Data Analysis

Exploratory Data Analysis were done to better understand the characteristics of respondents who participated in self-administered data collection consists of 4 age group, majority are males with 70% of all respondents, predominantly within the height group of 161-170 centimeters and 61-70 kilograms, most of the respondents have medium speed preference as shown in table 2 below. An interesting fact found from the exploratory data analysis is that the age group of participants is between 21 to 60 years old, which is the productive age of most Indonesians.

Table 2. Characteristics of Respondents

	age_group	gender	height_group	weight_group	speed	bike_type
count	242	242	242	242	242	242
unique	4	2	5	7	3	3
top	41 – 60 years old	Male	161 – 170 cm	61 – 70 kg	Medium Speed	Mountain Bike
freq	119	170	111	60	140	109

The following pie charts in Figure 2 show a more detailed composition of respondents by age group, gender, height group, weight group, and cycling speed preferences.
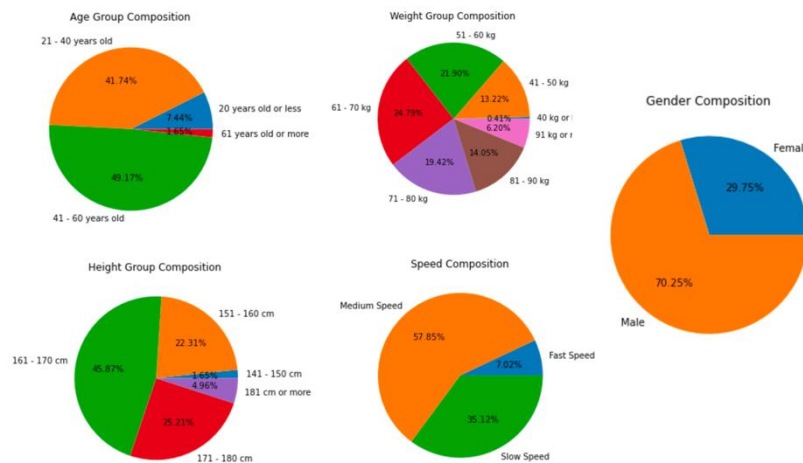

Fig 2. Composition of respondents' characteristics

The composition of respondents is as expected, specifically for cycling speed preference, by design slow speed refers to people who would choose a folding bicycle based on its size which is smaller and more suitable for recreational purposes, while medium and fast speed represents those who would choose either mountain bike or road bike. In this study target variable is designed as a categorical variable with two values namely folding bike or mountain/road bike.

3.3. Encoding and Standardizing

In machine learning data processing sometimes the data types are in nominal scale (categorical) which cannot be directly used for model building. This nominal (categorical) data must first be converted into numeric data so that it can be used for further analysis; the process of converting this data is called encoding. By using a python library called sklearn we can encode nominal data into numeric. For this research, the female gender is encoded to 0 while the male becomes 1. Subsequently, the target variable, namely bike-type was encoded to 0 for folding bike and 1 for mountain bike and road bike due to the focus of this research is about folding bike prediction.

Many machine learning algorithms (e.g., K-Nearest Neighbor) produce better performance when numeric variables are converted to standard scale. The process of standardizing the scale works by shifting the distribution of the data so that it has a mean of zero and a standard deviation of one. An example of its application is if we have several variables with different scales, for example liters, kilograms, and centimeters, using a standard scaler the data distribution will be changed and the scale will be made standard.

3.4. Model Building

During the model building stage, the dataset will be divided into two, namely training data and test data, where the training data will be used to build a mathematical equation model by utilizing several machine learning algorithms that are commonly used for classification problems. Test data will be used to test the accuracy of the predicted values. In this study, five algorithms were used for classification, namely K-Nearest Neighbor (KNN), Naive Bayes, Support Vector Machine, Decision Tree, and Random Forest as shown in Figure 3 below.

```
from sklearn.neighbors import KNeighborsClassifier

knn_classifier = KNeighborsClassifier(n_neighbors = 5, metric = "euclidean", p = 2)
knn_classifier.fit(X_train, y_train)
```

KNeighborsClassifier(metric='euclidean')

```
from sklearn.naive_bayes import GaussianNB

nb_classifier = GaussianNB()
nb_classifier.fit(X_train, y_train)
```

GaussianNB()

```
from sklearn.svm import SVC

svm_classifier = SVC(kernel = 'linear', random_state = 0)
svm_classifier.fit(X_train, y_train)
```

SVC(kernel='linear', random_state=0)

```
from sklearn.tree import DecisionTreeClassifier

dt_classifier = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
dt_classifier.fit(X_train, y_train)
```

DecisionTreeClassifier(criterion='entropy', random_state=0)

```
from sklearn.ensemble import RandomForestClassifier

rf_classifier = RandomForestClassifier(n_estimators = 10, criterion = 'entropy', random_state = 0)
rf_classifier.fit(X_train, y_train)
```

RandomForestClassifier(criterion='entropy', n_estimators=10, random_state=0)

Fig. 3. Model Building using Python

4. Results and Analysis

Model evaluation is the next step after model building to evaluate the performance of each mathematical model created by different kinds of algorithms. Utilizing confusion matrix class from python's sklearn metric library, predicted values resulted from data test are compared against target variable value from test data. There are four results of the classification process on the confusion matrix, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Accuracy score is (TP + TN) / (TP + FP + TN + FN). The higher the accuracy score the better.

The result of accuracy scores is listed in table 3 beneath this paragraph.

Table 3. Accuracy Score Comparison

Algorithm	True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)	Accuracy Score (TP+TN) / (TP+FP+TN+FN)
K-NN	7	3	3	36	0.8776
Naïve Bayes	5	5	8	31	0.7347
Support Vector Machine	6	4	1	38	0.8980
Decision Tree	10	0	5	34	0.8980
Random Forest	8	2	6	33	0.8367

All machine learning algorithms show good accuracy scores with Support Vector Machine and Decision Tree leading, however naive bayes showed average accuracy and the lowest among other four algorithms which suggest there may be dependence among the predictor variables.

5. Conclusion and Recommendation

This study uses a multi-purpose machine learning algorithm such as k-nearest neighbor, naive bayes, support vector machine, decision tree, and random forest. Data preprocessing is done with the help of the python programming language using the scikit learn (sklearn) library. The results of this study help prospective entry-level bicycle buyers to predict whether a folding bicycle is suitable for purchase or not.

Data as many as 242 respondents were collected from various communities spread across major cities on the island of Java, Indonesia. Comparatively, the best classification accuracy performance is shown by the Support Vector Machine and Decision Tree algorithms with an accuracy rate of close to 90%, while the poor performance is shown by the Naive Bayes algorithm.

In the future, data sets could be prepared with more features and using advanced machine learning techniques for model building could be used to build multi-class classification models to predict more than two types of bicycles. Another improvement is for data visualization can be done using a more sophisticated tool such as Tableau as suggested in [22].

Acknowledgment

References

[1] P. J. W. Van Den Noort, "Promoting Cycling for Public Health," *L. Use Transp. Res.*, pp. 105–132, 2016.

[2] T. K. Yunianto, "Tren Gowes Kerek Penjualan Sepeda Hingga 30% Selama Pandemi," *Katadata*, 2020. https://katadata.co.id/ekarina/berita/5f157dbd397ca/tren-gowes-kerek-penjualan-sepeda-hingga-30-selama-pandemi (accessed May 20, 2021).

[3] F. Pradolo, "Selama Pandemi Covid-19, Permintaan Sepeda Meroket 1.000 Persen," *Liputan 6*, 2020. https://www.liputan6.com/bisnis/read/4384886/selama-pandemi-covid-19-permintaan-sepeda-meroket-1000-persen.

[4] A. B. Tamtomo, "INFOGRAFIK: Jenis-jenis Sepeda dan Tips Membeli Sepeda," *kompas.com*, 2020. https://www.kompas.com/tren/read/2020/06/20/161659465/infografik-jenis-jenis-sepeda-dan-tips-membeli-sepeda (accessed May 20, 2021).

[5] H. Uden, S. Jones, and K. Grimmer, "Foot Pain and Cycling: a survey of frequency, type, location, associations, and amelioration of foot pain," *J. Sci. Cycl.*, vol. 1, no. 2, pp. 28–34, 2012, [Online]. Available: http://www.jsc-journal.com/ojs/index.php?journal=JSC&page=article&op=view&path[]=27%5Cnhttp://zadelpijn-expert.nl/data/documents/Footpain-and-Cycling.pdf.

[6] W. Van Hoof, K. Volkaerts, K. O'Sullivan, B. Malfait, S. Verschueren, and W. Dankaerts, "Low Back Pain in Cycling: Does It Matter How You Sit?," *Br. J. Sports Med.*, vol. 48, no. 7, pp. 609.2-609, 2014, doi: 10.1136/bjsports-2014-093494.134.

[7] M. H. Zaki, T. Sayed, and X. Wang, "Computer vision approach for the classification of bike type," *J. Adv. Transp.*, vol. 50, no. 3, pp. 348–362, 2016, doi: http://dx.doi.org/10.1002/atr.1327.

[8] S. Jaya Prada, A. Geetha Sri, B. Venkateswarlu, C. Vineesha, and P. Lakshmi Teja, "Bike Buyer Prediction," *Int. J. Comput. Eng. Technol.*, vol. 11, no. 3, pp. 45–51, 2020.

[9] A. Singh, M. N., and R. Lakshmiganthan, "Impact of Different Data Types on Classifier Performance of Random Forest, Naïve Bayes, and K-Nearest Neighbors Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 12, pp. 1–11, 2017, doi: 10.14569/ijacsa.2017.081201.

[10] J. Dill and N. McNeil, "Four types of cyclists: Examining a typology to better understand bicycling behavior and potential," *Transp. Res. Board*, vol. 500, pp. 1–18, 2013.

[11] B. Nazlı, G. Yasemin, and H. Altural, "Classification of Coronary Artery Disease Using Different Machine Learning Algorithms," *Int. J. Educ. Manag. Eng.*, vol. 10, no. 4, pp. 1–7, 2020.

[12] N. Kesav and M. G. Jibukumar, "Complexity Reduced Bi-channel CNN for Image Classification," *Lect. Notes Networks Syst.*, vol. 141, pp. 119–131, 2021, doi: 10.1007/978-981-15-7106-0_12.

[13] D. Mukhopadhyay, A. Malusare, A. Nandanwar, and S. Sakshi, "An Approach to Mitigate the Risk of Customer Churn Using Machine Learning Algorithms," *Lect. Notes Networks Syst.*, vol. 141, pp. 133–142, 2021, doi: 10.1007/978-981-15-7106-0_13.

[14] B. Akinyemi, O. Adewusi, and A. Oyebade, "An Improved Classification Model for Fake News Detection in Social Media," *Int. J. Inf. Technol. Comput. Sci.*, vol. 12, no. 1, pp. 34–43, 2020.

[15] M. T. Sathe and A. C. Adamuthe, "Comparative study of supervised algorithms for prediction of students' performance," *Int. J. Mod. Educ. Comput. Sci.*, vol. 13, no. 1, pp. 1–21, 2021.

[16] A. Adimabua Ojugo and E. Ekurume, "Predictive Intelligent Decision Support Model in Forecasting of the Diabetes Pandemic Using a Reinforcement Deep Learning Approach," *Int. J. Educ. Manag. Eng.*, vol. 11, no. 2, pp. 40–48, 2021.

[17] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *Int. J. Comput. Sci.*, vol. 4, no. 1, pp. 16–23, 2009, [Online]. Available: http://cogprints.org/6708/.

[18] D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine," *J. Theor. Appl. Inf. Technol.*, vol. 12, no. 1, pp. 1–7, 2010.

[19] M. Thamarai and S. P. Malarvizhi, "House Price Prediction Modeling Using Machine Learning," *Int. J. Inf. Eng. Electron. Bus.*, vol. 12, no. 2, pp. 15–20, 2020.

[20] L. (University of C. B. Breiman, "Random Forests," *Stat. Dep.*, 2001, [Online]. Available: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf.

[21] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/j.aci.2018.08.003.

[22] F. Murtagh and K. Devlin, "The development of data science: implications for education, employment, research, and the data revolution for sustainable development," *Big Data Cogn. Comput.*, vol. 2, no. 2, pp. 1–16, 2018, doi: 10.3390/bdcc2020014.

Authors' Profiles

Trianggoro Wiradinata is the Dean of School of Information Technology at Ciputra University. He received a bachelor's degree in electrical engineering (major in Computing) from Petra Christian University, Indonesia, and a master's degree in computer science from University of New South Wales, Australia and completed his Ph.D. program in Information Technology from Assumption University of Thailand. His main research interests are data science, software engineering, technology adoption, and technology-based entrepreneurship.

He was actively involved in many large-scale enterprise software development projects before joined as researcher at Ciputra University, Surabaya, Indonesia. Assoc. Prof. Dr. Wiradinata is currently a. member of Association for Computing Machinery (ACM) and Association for Information Systems Indonesia (AISINDO).