Malaysian Journal of Computing, 9 (1): 1673-1689, 2024

eISSN: 2600-8238

ANALYZING THE IMPACT OF FEATURE SELECTION USING INFORMATION GAIN FOR AIRLINES' CUSTOMER SATISFACTION

Farah Aqilah Bohani[1*], Farah Syazwani Mohamed Rashid[2], Yuzi Mahmud[3] and Sitti Rachmawati Yahya[4]

[1,2,3]*School of Computing Sciences, College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam*

[4]*Department of Information System, Asia Cyber University, South Jakarta, Indonesia*

[1*]farahaqilah@uitm.edu.my, [2]farahsyazwani@uitm.edu.my, [3]yuzi@uitm.edu.my, [4]sittiyahya@lecturer.unsia.ac.id

ABSTRACT

Feature selection has become a focus of research in many fields that deal with machine learning and data mining because it makes classifiers cost-effective, faster, and more accurate. In this paper, the impact of feature selection using filter methods such as Information Gain is shown. The impact of feature selection has been analyzed based on the accuracy of two classifiers: J48 and Naïve Bayes. The Airline Customer Satisfaction datasets have been used for comparing with and without applying Information Gain. As a result, J48 achieved 0.33% and 0.29% improvements in accuracy after applying Information Gain for 10-fold and 20-fold cross-validation, respectively compared to Naïve Bayes. Most of the precision and F1-score for J48 with Information Gain have also improved for both evaluation methods compared to Naïve Bayes. In conclusion, J48 seems to be the classifier that is most sensitive to feature selection and has shown improvements compared to Naïve Bayes.

Keywords: *Airline Customer Satisfaction***,** *J48, Naïve Bayes, Feature Selection, Information Gain*

Received for review: 04-12-2023; Accepted: 25-03-2024; Published: 01-04-2024
DOI: 10.24191/mjoc.v9i1.24163

1. Introduction

Feature selection is the process of selecting features from the original dataset that are redundant or irrelevant features. This method is unfeasible, costly in terms of computing, and results in a decrease in classification precision when applied to the whole input set (Mwadulo, 2016). Feature selection does not alter the original set of features. Instead of selecting a whole dataset, it selects a subset by removing any features whose existence in the dataset does not aid the learning model (Chandrashekar & Sahin, 2014).

The literature recognized the current variable feature selection strategies include filtering, embedding, and wrapping (Guyon et al., 2008). The filtering approach sorts the features in the preprocessing stage and is independent of the learning algorithm, thus the chosen features can be passed to any modelling procedure. The filter technique is further based on the filtering measures utilized, which include information, distance, dependency, consistency, similarity, and statistical metrics. The embedded technique incorporates feature selection within the modelling algorithm implementation. The features of method selection are determined by the learning algorithm. The wrapper technique trains the prediction model for candidate feature subsets using an independent algorithm and then employs greedy tactics such as forward or backward to select the optimum feature subset from all potential subsets throughout the learning process.

Filter technique adopt complete independence between the machine learning and the data. The technique assesses qualities based on the overall features of the data and functions independently of any classification algorithm. Filter techniques often require a non-iterative calculation on the data set, which runs much quicker than a classifier training session (Chandrashekar & Sahin, 2014; Liu & Yu, 2005). To extract the most relevant feature, in this paper, we focus on filter techniques such as Information Gain (IG), one of the successful attribute selection approaches in classification (Kurniabudi et al., 2021).

Classification is a supervised machine learning method that was used to identify the category of new observations based on training data. Classification can be used in a wide range of fields, including pavement performance prediction (Marcelino et al., 2021), medical field (Kaur & Kumari, 2020), employee promotion prediction (Shafie et al., 2023), finance (Malhotra et al., 2020), bank customer churn prediction (Hui et al., 2023), and flight delay classification (Yi et al., 2021). Various studies investigated customer satisfaction with airline service based on machine learning technology (Al-Qahtani, 2021; Amalia et al., 2022; Bellizzi et al., 2022; Hayadi et al., 2021; Keerthy & Mathew, 2022; Kumar & Zymbler, 2019; Leon & Martín, 2020; Rane & Kumar, 2018). Lack of study about feature selection using Information Gain based on Airline Customer Satisfaction datasets was conducted. Hence, this paper will discuss analyzing the impact of feature selection using the Information Gain approach for airline customer satisfaction in terms of accuracy, precision, recall, and F1-score.

The rest of the paper is structured as follows; several studies that investigated feature selection and supervised learning algorithms for the airline customer satisfaction prediction are reviewed in Section 2 (Related Work). The supervised learning techniques that were employed in the study are briefly explained in Section 3 (Overview of Feature Selection and Machine Learning Techniques). The pre-processing techniques and dataset are described in Section 4 (Methodology). The experimental findings and a comparative analysis of this study are then presented in Section 5 (Results and Discussion). Ultimately, Section 6 (Conclusion) presents the findings of the research.

2. Related Work

Every airline strives to give the highest level of customer service to guarantee that its customers are completely satisfied with its products and remain loyal. This allows the company to continue its expansion and maintain competitiveness in the industry. Customers are more inclined to rate airlines depending on the quality of their in-flight services (Saut & Song, 2022). As a result, improving onboard service quality becomes one of an airline's most crucial success criteria.

Customer satisfaction is one of the most significant characteristics to focus on when assessing consumer loyalty and repurchase intent, as well as increasing positive feedback and lowering acquisition costs (Roy et al., 2016). Businesses collect consumer feedback and assess their products and services by conducting customer satisfaction surveys. This data can be utilized to train machine learning algorithms, particularly those using supervised learning. Furthermore, the machine learning model can assist management in building future business goals and approaches for maintaining present clients and obtaining new ones.

There is still important research on predicting airline client satisfaction using survey findings. There are numerous algorithms, including K-Nearest Neighbors (KNN), Logistics Regression, Gaussian Naïve Bayes, Decision Tree, and Random Forest (RF) have been investigated for predicting airplane passengers' Wi-Fi service experience (Hayadi et al., 2021). Research by (Amalia et al. (2022) presents the Split Point and Attribute Reduced Classifier (SPAARC) to compare with existing algorithms for predicting airline customer satisfaction using a Likert scale of 1–5. The SPAARC approach has the advantage of reducing the computational cost associated with decision trees (Amalia et al., 2022).

One of the issues with airline customer satisfaction is the relatively high dimensions of the feature space. The dataset of airline customer satisfaction consists of 23 features. This is quite challenging for running machine learning algorithms. If reducing the set of features considered by the algorithm, the study can serve one purpose. As a result, it has the potential to significantly improve the overall accuracy of the model. Recently, several studies have addressed the issue of feature subset selection (Hayadi et al., 2021; Jiang et al., 2022). Information gain (IG) and chi-square test (CHI) were the most successful in aggressive term elimination while maintaining classification accuracy (Yang & Pedersen, 1997). The study observed a high correlation between IG and CHI ratings for a given phrase. Owing to that, in this study, IG will be used as a feature selection approach for airline customer satisfaction prediction.

As stated in (Huang, 2021), J48 offers the benefits of easily legible classification rules and excellent accuracy. Therefore, in this study, J48 is utilized for airline customer satisfaction prediction. There are two sets of features: one includes all of the features of datasets, while the other selects feature based on Information Gain. This study not only compares the performance evaluations of the two algorithms, but it also analyzes the impact of feature selection on the data.

Naive Bayes that was known as prominent learning method is frequently justified by assumptions of conditional independence or linked dependency (Cooper, 1991). This work uses Naïve Bayes to analyze feature selection effects and compare them to J48 classifier performance. This study compares two classification algorithms: J48 and Naïve Bayes.

3. Overview of Feature Selection and Machine Learning Algorithms

3.1 Feature Selection

Feature Selection can be described as reducing the features of dataset. The purpose of feature selection is to identify the minimum of attributes required to predict the class of individual data points with precision, while reducing the overall number of attributes required. This can facilitate the comprehension and interpretation of the patterns that are identified within the data. This approach evaluates the informational capacity of each attribute with respect to the data class and chooses those attributes that provide the greatest quantity of insight. By discerning the most pertinent attributes, one can enhance the precision of the predictions while decreasing the complexity of the data.

Utilized extensively, the IG feature selection method is a technique utilized to identify crucial features within a dataset. Fundamental concepts of information theory, including entropy and information gain is introduced in 1948 (Verdu, 1998). IG represents the quantity of information that an attribute imparts regarding the data class that is frequently employed in tandem with decision tree algorithms and proves to be highly beneficial in the process of feature selection for classification tasks. It is feasible to enhance the precision of the predictions and decrease the intricacy of the data by exclusively choosing the most critical characteristics (Verdu, 1998).

The quantity of information that each attribute provides about the class of data calculated to determine the most significant features in a dataset using the IG feature selection approach. A feature is included in the model and deemed relevant if its IG is greater than a certain threshold. As a result, it is said to be utilized for effective categorization in dimension reduction. Equation (1) from these researchers (Lee & Lee, 2006) is used to compute the information gain.

$$G\ (D,t) = -\sum_{i=1}^{m} P(C_i) logP\ (C_i) + P(t) \sum_{i=1}^{m} P(C_i|t) logP\ (C_i|t) + P(\bar{t}) \sum_{i=1}^{m} P(C_i|\bar{t}) logP\ (C_i|\bar{t}) \quad (1)$$

From the equation, C represents the document collection. $P(C_i)$ represents the probability of the ith category. $P(t)$ and $P(\bar{t})$ represents the probabilities that the term t appears or does not in the document respectively. $P(C_i|t)$and $P(C_i|\bar{t})$ represents conditional probability of the ith class value given the term t appears or does not appear in the document respectively Information Gain has widely been used as a feature selection method in many data mining task.

3.2 J48 Classifier

Decision Tree is an algorithm for creating a model that employs a tree-like structure to generate predictions based on feature of instances. The method utilizes the training data to establish each instance class, which it is able to predict for new, previously unseen examples instances (Korting, 2006). This method builds prediction rules for the target variable. The use of trees to categorize data simplifies understanding the critical distribution of information (Nadali et al., 2011).

J48 is a variation of the ID3 (Iterative Dichotomiser 3) decision tree method that includes extra features such as missing value management, decision tree pruning, and continuous attribute handling. J48 is also known as C4.5 and is part of the WEKA data mining tool. WEKA is a data mining tool that contains several methods for pruning decision trees. Pruning is a strategy for increasing the accuracy of a tree by reducing the risk of overfitting. In certain algorithms, such as decision trees, the classification process is repeated recursively until each leaf node in the tree is pure, which means it only includes data points from one class. Pruning serves to refine the tree and enhance its accuracy by removing branches that are not providing useful information. After applying it to data, this algorithm generates the rules that will be used to determine the identification of the data. The idea is to gradually generalize a decision tree so that it balances flexibility and accuracy.

3.3 Naïve Bayes

The Naive Bayes algorithm is widely employed in practical contexts to address classification challenges due to its straightforward construction and interpretation, in addition to its robust performance (Jiang et al., 2019). Naive Bayes is a supervised machine learning algorithm that operates under the assumption that the features or predictors in the dataset are independent of each other. Naive Bayes, founded upon the Bayes theorem, is frequently applied to classification tasks and is renowned for its efficiency and simplicity. To predict the class label of a given data point using a set of features, such as the presence of particular words in a document that are utilized to ascertain the document affiliation with a specific category or not, is one possible application of the algorithm. This suggests that the attributes of the class are not contingent upon the attributes of other classes.

The Naive Bayes classifier has been demonstrated to be suitable for both continuous and categorical data (Chen et al., 2020). The Bayes theorem, a mathematical formula utilized to compute the probability of an event (A) transpiring given some evidence or proof (B), serves as the foundation for the Naive Bayes algorithm. Typically, Equation (2) is used to represent the formula (Amra & Maghari, 2017).

$$P(A, B) = P(A)P(B) \tag{2}$$

Through equation (1) and using the concept of the Bayes theorem, the final equation of the Naïve Bayes algorithm is obtained as Equation (3).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

Based on Equation (2), the Naive Bayes algorithm is used to predict the value of a target variable (A) based on a set of predictor variables (B). In the context of the algorithm, A is often referred to as the dependent event or the predicted variable, while B represents the independent event or the predictor attribute. Based on the values of the predictor variables, the algorithm calculates the probability of each possible class (A) and chooses the class with the highest probability as the prediction. This process is repeated for each instance in the data, and the final predictions can be used to classify the data or make other decisions.

4. Methodology

This section will be explained experimental design for airlines customer satisfaction prediction with and without feature selection.

4.1 Experimental Design

This study utilized a system with the following specifications: AMD Ryzen 5 5500U with Radeon Graphics 2.10 GHz, 8.00 GB RAM, and Windows 10 64-bit. All tests were carried out with WEKA tools, which included missing values, outliers, feature selection, and classification routines. Figure 1 depicts the framework of airline customer satisfaction and the overall flow of model training, which is separated into the following phases. (a) Data collection, pre-processing, classification, and performance evaluation; and (b) Data collection, pre-processing, feature selection, classification, and performance evaluation. The difference between the frameworks in Figures 1 (a) and (b) is the feature selection procedure, which is relevant to the aims of this study to compare the classifier performance with and without feature selection. Each phase will be discussed in Sections 4.1.1 through 4.1.5.

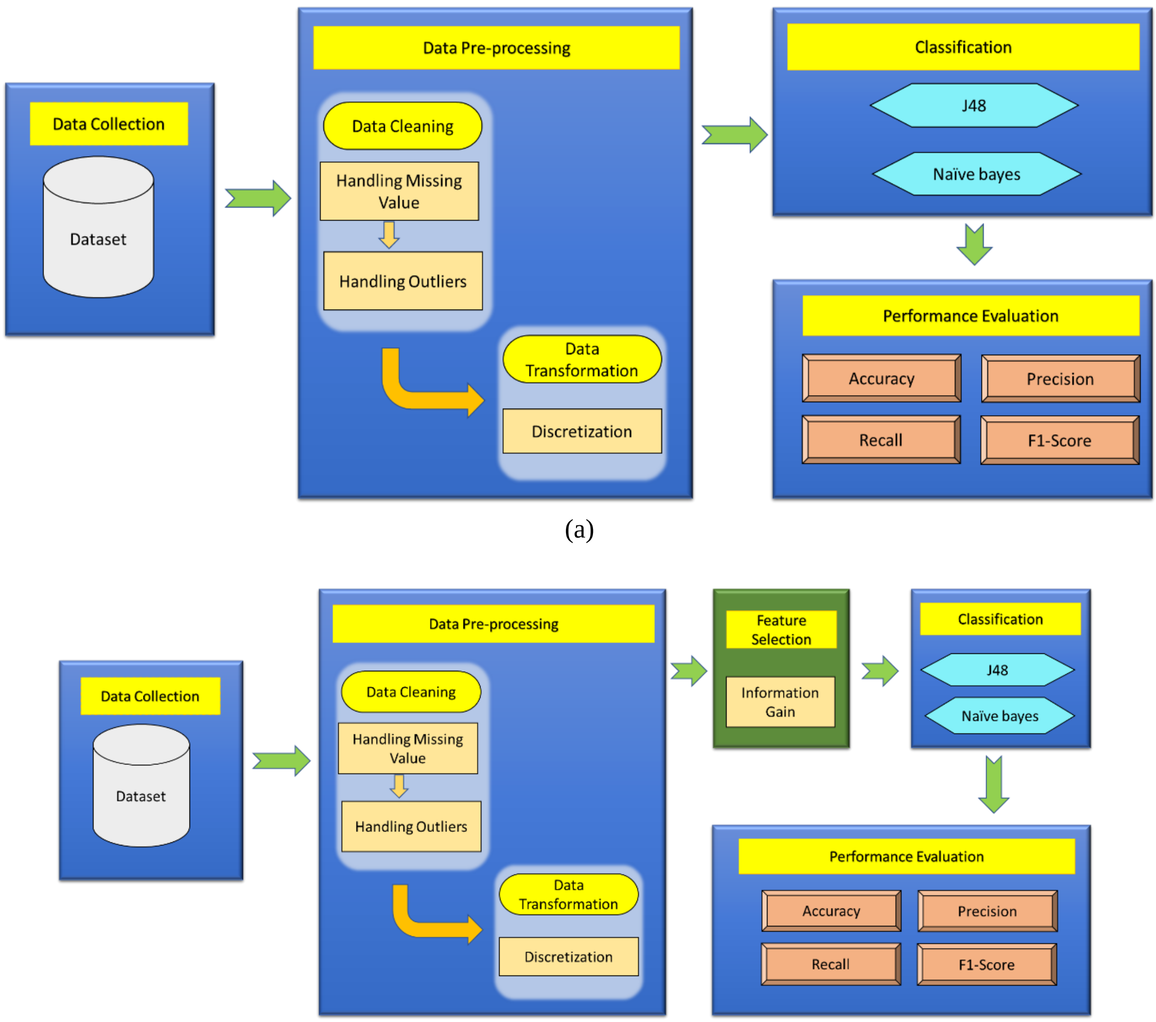


Figure 1. Framework of airlines customer satisfaction classification using J48 and Naïve Bayes (a) without feature selection, and (b) with feature selection using Information Gain.

4.1.1 Data Collection

Airline Passenger Satisfaction dataset from https://www.kaggle.com/datasets/ (Kaggle, 2020) was used in this work. This dataset can be considered as new dataset since it was added to www.kaggle.com in May 2020. The world's airline passengers were polled to get this dataset. The dataset consists of 23 features, 129882 instances, and two class labels: "satisfied" and "neutral or dissatisfied." The goal of this data is to find out what factors are strongly linked to flight passenger satisfaction. In addition, this information can be used to make a model for classification. Data on how satisfied airline passengers are with their service can be seen in Table 1. It shows each feature, type, and description. There are 16 factors that can be rated on a scale from 1 to 5 for how satisfied people are with the following: "Inflight Wi-Fi Service," "Departure/Arrival Time Convenient," "Food and Drink," "Online Boarding," "Seat Comfort," "Inflight Entertainment," "On-Board Service," "Leg Room Service," "Baggage Handling," "Check-In Service," "Inflight Service," "Cleanliness," "Departure Delay in Minutes," and "Arrival Delay In Minutes." Each of the satisfaction levels has been referred in this research (Chaudhury et al., 2019) is shown in Table 1.

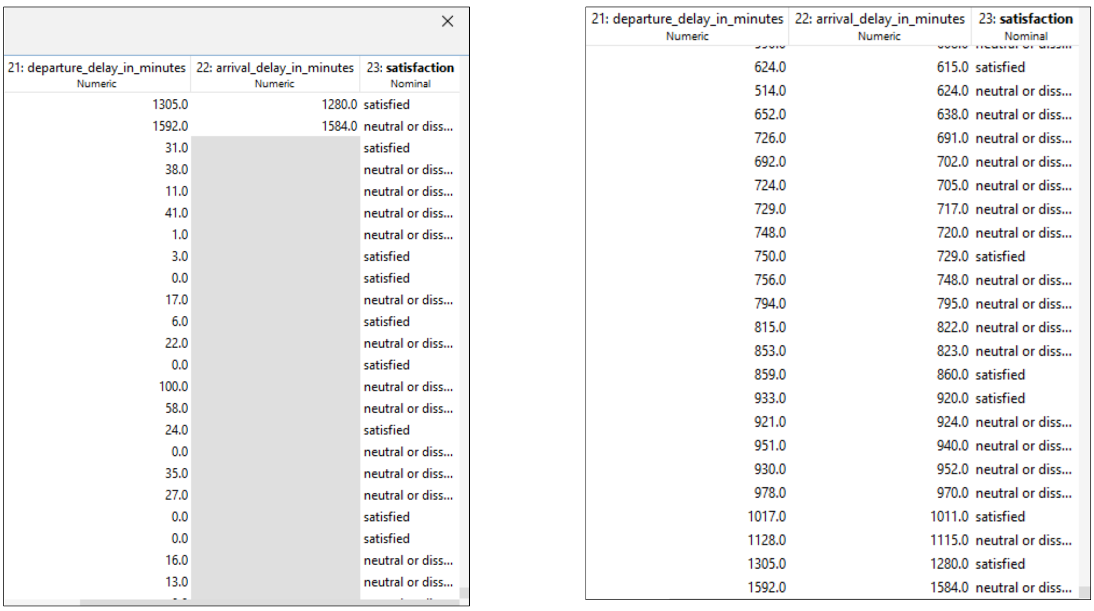
Table 1. Details of records of the airline passenger satisfaction dataset.

No	Attribute Name	Attribute Type	Description
1	Gender	Nominal	Gender of the passengers (Female, Male)
2	Customer Type	Nominal	The customer type (Loyal customer, disloyal customer)
3	Age	Numeric	The actual age of the passengers
4	Type Of Travel	Nominal	Purpose of the flight of the passengers (Personal Travel, Business Travel)
5	Customer Class	Nominal	Travel class in the plane of the passengers (Business, Eco, Eco Plus)
6	Flight Distance	Numeric	The flight distance of this journey
7	Inflight Wi-Fi Service	Ordinal	Satisfaction level of the inflight Wi-Fi service (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
8	Departure/Arrival Time Convenient	Ordinal	Satisfaction level of Departure/Arrival time convenient (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
9	Ease Of Online Booking	Ordinal	Satisfaction level of Online booking (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
10	Gate Location	Ordinal	Satisfaction level of Gate location (Rating: 0: “very inconvenient”, 1: “inconvenient”, 2: “need improvement”, 3: “manageable”, 4: “convenient”, 5: “very convenient”)
11	Food and Drink	Ordinal	Satisfaction level of Food and drink (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
12	Online Boarding	Ordinal	Satisfaction level of Online boarding (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
13	Seat Comfort	Ordinal	Satisfaction level of Seat comfort (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
14	Inflight Entertainment	Ordinal	Satisfaction level of inflight entertainment (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
15	On-Board Service	Ordinal	Satisfaction level of On-board service (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)

16	Leg Room Service	Ordinal	Satisfaction level of Leg room service (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
17	Baggage Handling	Ordinal	Satisfaction level of Baggage handling (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
18	Check-In Service	Ordinal	Satisfaction level of Check-in service (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
19	Inflight Service	Ordinal	Satisfaction level of Inflight service (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
20	Cleanliness	Ordinal	Satisfaction level of Cleanliness (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
21	Departure Delay In Minutes	Ordinal	Minutes delayed when departure (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
22	Arrival Delay In Minutes	Ordinal	Minutes delayed when Arrival (Rating: 0: “extremely poor”, 1: “poor”, 2: “need improvement”, 3: “acceptable”, 4: “good”, 5: “excellent”)
23	Satisfaction	Class Label	Airline satisfaction level (Satisfaction, Neutral or Dissatisfaction)

4.1.2 Data Pre-processing

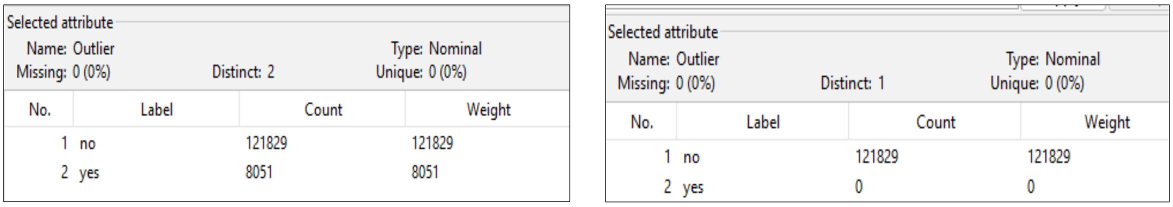
The next stage is data pre-processing, which improves data quality prior to processing. Data pre-processing includes two subset processes: data cleaning and data transformation. The data cleaning procedure is divided into two subsets: missing values and outliers. In general, handling missing data increases the classifiers prediction capability. In this research, addressing missing values improves airline customer satisfaction. In WEKA tools, a filter named "ReplaceMissingValues" is used to replace all missing values in a dataset with the average of each attribute. Figure 2 depicts one of the properties with missing data.



(a) (b)

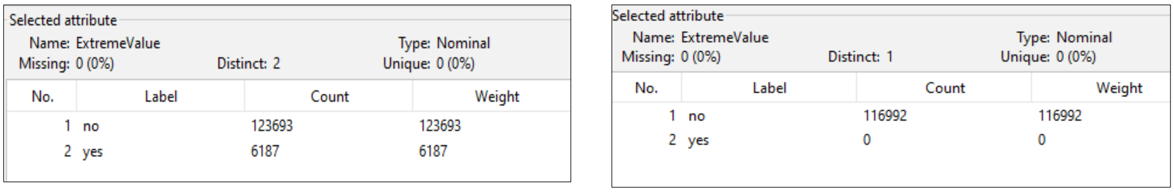
Figure 2. List of values for the ‘arrival_delay_in_minutes’ attribute (a) before handling missing values (b) after handling missing values.

After dealing with missing data, dealing with outliers is a key step in preventing skewing and speeding up the identification process. An outlier is significantly different from the other data points in which it appears (Dash et al., 2023). In this scenario, WEKA's "InterquartileRange" function is used, followed by a "RemovewithValues" filter to eliminate outliers. Figure 3 indicates that 8051 outliers have been eliminated from the dataset. This dataset contains both nominal data, such as satisfaction levels ranging from 1 to 5, and numeric attributes, such as 'departure_delay_in_minutes', which include extreme values. Extreme value arises when a data point deviates significantly from the other observations, such as '1305.0' compared to other values in the 'departure_delay_in_minutes' property. Figure 4 illustrates the number of extreme values that have been removed: 6187.



(a) (b)

Figure 3. Visualization in WEKA tools (a) before removing outliers (b) after removing outliers.



(a) (b)

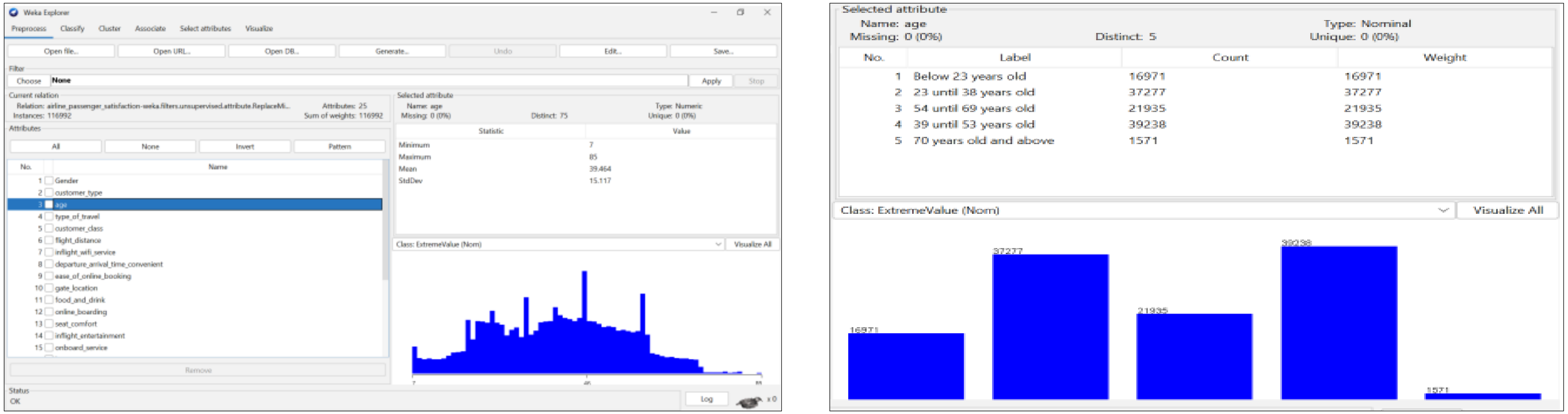
Figure 4. Visualization in WEKA tools (a) before removing extreme values (b) after removing extreme values.

Data transformation has been applied after removing outliers and extreme values. Data transformation techniques include discretization or grouping (Berka & Bruha, 1998). These stages should provide data in the format required by the data mining algorithms (Berka & Bruha, 1998). Thus, discretization is applied in this study. Features derived from raw data are often numerical (continuous). Some classification algorithms can only accept nominal data as inputs, such as those that discretize numeric data into nominal data. A tree-based algorithm, such as J48, is one of the algorithms that must be developed with nominal data. As a result, this research requires discretization as a preprocessing step. The data is partitioned from numeric into interval categories.

In this research, 'Age' and 'Flight_distance' were changed from numeric to nominal, as shown in Figures 5 and 6. The discretization procedure occurs exclusively for these characteristics since they are both numeric data, as indicated in Table 1, and must be turned into nominal data before J48 classification can be performed. Figure 5 illustrates that the 'Age' attribute has been discretized by classifying passengers as 'Below 23 years old', '23 until 38 years old', '54 until 69 years old', '39 until 53 years old', and '70 years old and above'. As shown in Figure 6, after executing the discretization process on the 'Flight_distance' property, the discrete has been marked as 'Below 1682 miles', '1682 till 3332 miles', and '3333 miles and above'.

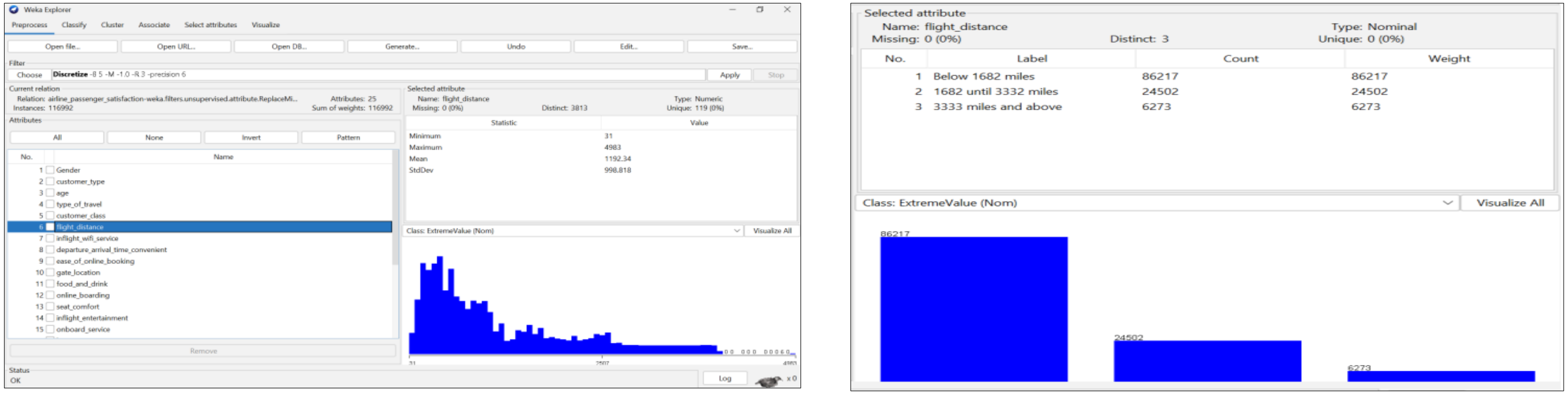
4.1.3 Feature Selection

In this stage, a feature selection strategy such as Information Gain was used. Figure 7 demonstrates that four of the least important qualities will be removed (red box). Figure 8 illustrates the list of properties before and after feature selection.



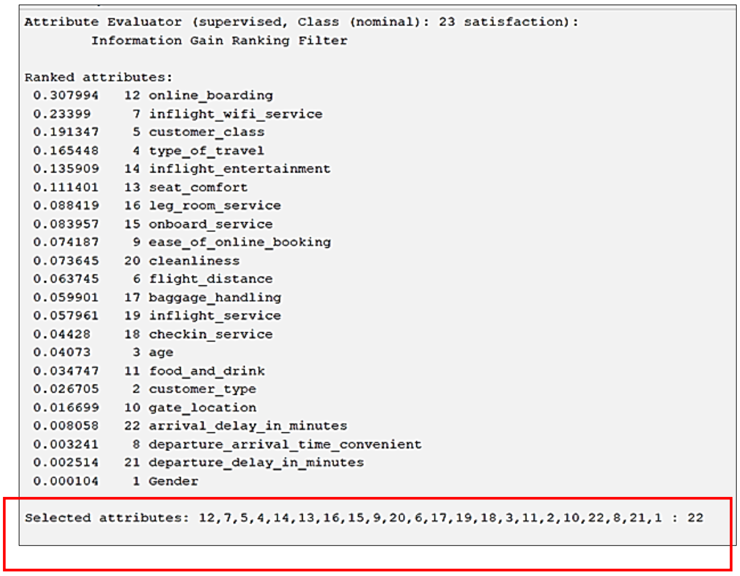
(a) (b)

Figure 5. Visualization in WEKA tools for 'Age' attribute (a) before discretization process (b) after the discretization process with 5 bins (groups).



(a) (b)

Figure 6. Visualization in WEKA tools for 'Flight_distance' attribute (a) before discretization process (b) after discretization process with 3 bins.

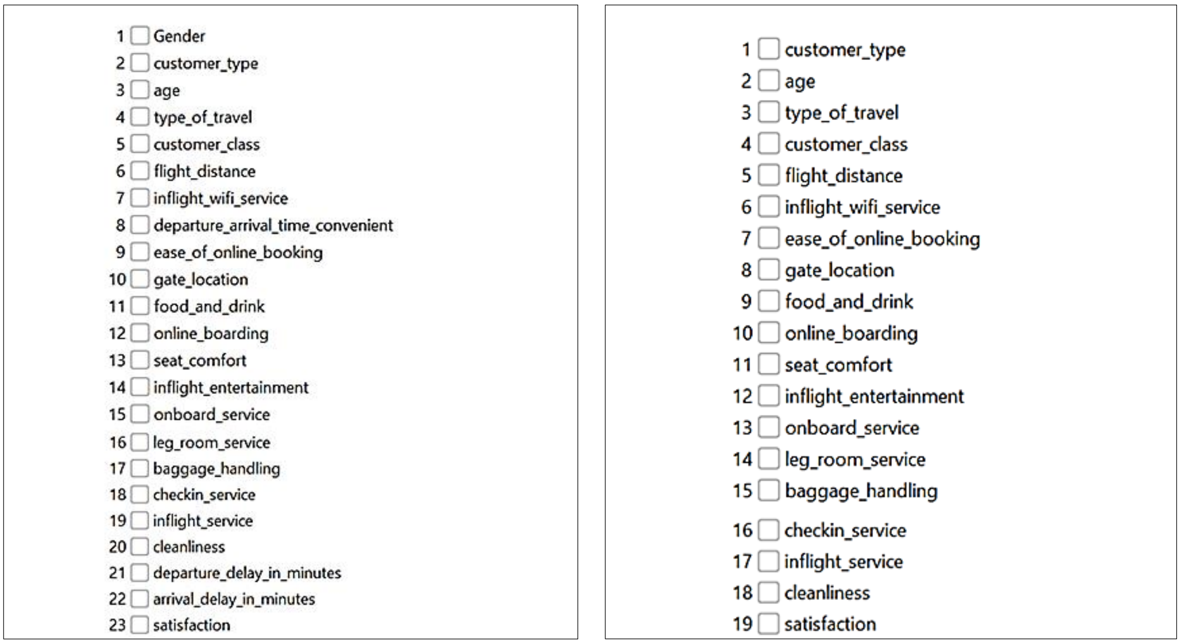


```
Attribute Evaluator (supervised, Class (nominal): 23 satisfaction):
        Information Gain Ranking Filter

Ranked attributes:
 0.307994   12 online_boarding
 0.23399     7 inflight_wifi_service
 0.191347    5 customer_class
 0.165448    4 type_of_travel
 0.135909   14 inflight_entertainment
 0.111401   13 seat_comfort
 0.088419   16 leg_room_service
 0.083957   15 onboard_service
 0.074187    9 ease_of_online_booking
 0.073645   20 cleanliness
 0.063745    6 flight_distance
 0.059901   17 baggage_handling
 0.057961   19 inflight_service
 0.04428    18 checkin_service
 0.04073     3 age
 0.034747   11 food_and_drink
 0.026705    2 customer_type
 0.016699   10 gate_location
 0.008058   22 arrival_delay_in_minutes
 0.003241    8 departure_arrival_time_convenient
 0.002514   21 departure_delay_in_minutes
 0.000104    1 Gender

Selected attributes: 12,7,5,4,14,13,16,15,9,20,6,17,19,18,3,11,2,10,22,8,21,1 : 22
```

Figure 7. List of features that have been ranked by Information Gain in WEKA tools.



(a) (b)

Figure 8. List of features in WEKA tools (a) before feature selection with 23 features (b) after feature selection with 19 features.

4.1.4. Classification

After data pre-processing, J48 and Naïve Bayes algorithms were employed to predict airline customer satisfaction levels. In this study, current hyperparameters in WEKA tools for both comparison algorithms are set to the default value. Table 2 displays the hyperparameters for J48 and Naïve Bayes.

Table 2. Method and Hyperparameter Used.

Method	Hyperparameter
J48	batchSize, binarySplits, collapseTree, confidenceFactor, debug, doNotCheckCapabilities, doNotMakeSplitPointActualValue, minNumObj, numDecimalPlaces, numFolds, reducedErrorPruning, saveInstanceData, seed, subtreeRaising, unpruned, useLaplace, useMDLcorrection.
Naïve Bayes	batchSize, debug, displayModelInOldFormat, numDecimalPlaces, useKernelEstimator, useSupervisedDiscretization.

4.1.5. Performance Evaluation

The section explained the performance assessment compares J48 with Naïve Bayes in terms of accuracy, precision, recall, and F1 measure. Each model assessment involves two set trials which are cross-validation and split percentage. Experiment 1 used algorithms without feature selection, whereas Experiment 2 examined methods with feature selection. All investigations considered the binary confusion matrix while calculating classification performance.

Cross-validation is a model validation approach used to assess the statistical results of a research that is meant to be generalized from its component data set (Hartmann & Carleo, 2019). Cross-validation is useful for assessing errors in forecasting or evaluating model performance (Tiwari et al., 2021). During cross-validation, rotation estimation is used to divide the data into k-subsets of about equal size. After then, training and testing are repeated k times; during each repetition, one set is utilised for test data, while the other k data subsets are used as training data for the next repetition. Then, up to k repetitions of training and testing are performed; in each repetition, one set is utilised for test data, while the remaining k data subsets are used for training data. If there is a limited quantity of data, the K-Fold approach can be used to evaluate the classifier's performance. The optimum implementation of the number of folds in the validity test employs 10-fold cross-validation in each model (Zhu & Liu, 2021). The utilization of cross-validation for algorithm evaluation is feasible as a result of the enhanced precision.

On the other hand, F1-score calculation considers the relative importance of the factor precision (the model's ability to accurately predict positive labels) and recalls (the amount of actual positive data that the model can capture using positive data labels). F1-score is utilized to ascertain the efficiency of the employed strategy. The F1-score is determined using Equation (4).

$$F1\ score = 2 \times \frac{Precision\ \times\ Recall}{Precision + Recall} \tag{4}$$

True positives (TP) in a binary confusion matrix relate to a class that has been accurately identified as positive; true negatives (TN) are correctly classed as negative. False negatives (FN) are cases of a positive class that are incorrectly classed as negative, whereas false positives are occurrences of a negative class that are incorrectly classified as positive. Classification performance indicators can be computed using the values of TP, FP, TN, and TP to represent how well the classifier detects a particular class (Ruuska et al., 2018). The most generally used markers are accuracy, precision, and recall (sensitivity), which can be expressed as the following equations, Equation (5), Equation (6), and Equation (7), respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (6)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (7)$$

Accuracy is the most basic and extensively used parameter for evaluating the performance of a classification model. In addition to accuracy, this research examines categorization performance criteria such as precision and recall. According to research in Juba and Le (2019). classification performance metrics based on accuracy, precision, and recall are appropriate for classifying unbalanced data. Despite the fact that the classification is based on a balanced dataset, accuracy, precision, and recall are utilized in this research.

5. Results And Discussion

This section compares experimental data to two algorithms, J48 and Naïve Bayes. There are two training datasets with a percentage split. One set of training datasets was partitioned into two parts: training (80%) and test data (20%); otherwise, training (70%) and test data (30%). J48 and Naïve Bayes models were used for training the dataset. Furthermore, this work employed the cross-validation approach for model evaluation to provide a diverse training dataset. Finally, this research evaluated and compared the classification models using four distinct types of evaluation metrics: accuracy, precision, recall, and F1.

The accuracy, precision, recall, and F1-scores of the two models were acquired to evaluate the models. Tables 3 and 4 provide more complete findings. Table 3 shows that J48 outperforms Naïve Bayes in terms of accuracy, precision, recall, and F1-score for k-folds = 10 and 20 (with and without feature selection). J48 outperforms Naïve Bayes in accuracy, precision, recall, and F1-score for many model assessment methods, including 70:30 and 80:20 split percentages.

The test results for k-folds = 10 and k-folds = 20 demonstrated that J48 performed better after feature selection than before, with 95.78% and 95.87%, respectively. This is most likely due to the Information Gain feature selection method, which included just four of the 23 features drawn from the Airline Passenger Satisfaction data. Following feature selection, split percentage of J48 produced 97% and 96.60% of accuracy at 70:30 and 80:20 ratios, respectively. However, the results show that the improvement is small. Similarly, F1-score produced by J48 has improved for split percentages of 70:30 and 80:20. Only the recall rate for J48 reduced from 94% to 93.60% after feature selection.

After feature selection and 10- and 20-fold cross-validation, Naïve Bayes accuracy decreases by 0.03%. By using feature selection approach such IG reduces recall and F1-Score in Naïve Bayes for k-folds of 10 and 20. Naïve Bayes maintained 83.50% accuracy before and after feature selection for k-folds=10 cross-validation. Results for after applying feature selection, accuracy of Naïve Bayes is declined at split percentages of 80:20 and 70:30. Feature selection reduces Naïve Bayes accuracy, recall, and F1-score. After feature selection, Naïve Bayes maintained 83.70% accuracy with a 70:30 split percentage.

Table 3. Accuracy of J48 and Naïve Bayes with different of ***k***-folds of cross validation before and after feature selection using Information Gain.

No of k-folds	**Evaluation Metrics**	**Accuracy (%)**		**Precision (%)**		**Recall (%)**		**F1-Score (%)**	
	Feature Selection/ Classifier	**J48**	**Naïve Bayes**	**J48**	**Naïve Bayes**	**J48**	**Naïve Bayes**	**J48**	**Naïve Bayes**
10	Before feature selection	95.45	85.11	95.90	83.50	**93.80**	**82.80**	94.80	**83.10**
	After feature selection	**95.78**	**85.08**	**96.90**	83.50	93.40	82.60	**95.20**	83.00
20	Before feature selection	95.58	85.11	96.00	83.40	**93.90**	**82.80**	95.00	83.10
	After feature selection	**95.87**	**85.08**	**97.00**	**83.60**	93.60	82.60	**95.20**	83.10

Table 4. Accuracy of J48 and Naïve Bayes with different of split percentage ratio before and after feature selection using Information Gain.

Split Percentage	**Evaluation Metrics**	**Accuracy (%)**		**Precision**		**Recall**		**F1-Score (%)**	
	Feature Selection/ Classifier	**J48**	**Naïve Bayes**	**J48**	**Naïve Bayes**	**J48**	**Naïve Bayes**	**J48**	**Naïve Bayes**
70:30	Before feature selection	95.65	**85.30**	96.10	83.70	**94.00**	**82.90**	95.00	**83.30**
	After feature selection	**95.90**	85.25	**97.00**	83.70	93.60	82.70	**95.30**	83.20
80:20	Before feature selection	95.46	**85.55**	95.80	**83.90**	93.80	**83.30**	94.80	**83.60**
	After feature selection	**95.85**	85.37	**96.60**	83.80	**93.90**	83.00	**95.30**	83.40

6. Conclusion

As conclusion, this study J48 and Naïve Bayes were evaluated using machine learning algorithms to identify the best classification model for passenger satisfaction. The study revealed that after using feature selection strategies such as Information Gain, J48 models performed somewhat better than models employing a whole feature set. J48 improved accuracy by 0.33% and 0.29% using Information Gain for 10-fold and 20-fold cross validation, compared to Naïve Bayes. The accuracy and F1-score for J48 with Information Gain increased for both evaluation techniques compared to Naïve Bayes. J48 is more sensitive to feature selection and performs better than Naïve Bayes.

The experimental findings show that using a feature selection approach such as Information Gain, airline customer satisfaction can be successfully classified even with fewer dataset features. It is possible to infer that by applying feature selection, both features and computing complexity can be reduced. Feature selection has enhanced the accuracy of prediction models. In the future, this J48 model with feature selection can be used to find aspects that satisfy consumers with airlines and features that disgruntled customers want improved.

Acknowledgment

The authors wish to thank to Fazrul Ahmad Fisol, Nurul Anis Zulaikha Mohamad Zolkipli, Wan Nur Nadhirah Wan Mohamad Hambali, Najah Munirah Abd Razak, and Nik Nur Afifah Suhaimi whom are Universiti Teknologi MARA (UiTM) Shah Alam students for completing this project.

Funding

The author(s) received no specific funding for this work.

Author Contribution

Author 1 prepared the literature review, wrote the content, prepared the research methodology, and provided results analysis and discussion. Author 2, Author 3, and Author 4 reviewed the literature review and provided constructive feedback on the manuscript. All authors have reviewed and approved the final version of the manuscript.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

Al-Qahtani, R. (2021). Predict sentiment of airline tweets using ML models. *(No. 5228). EasyChair*.

Amalia, S., Deborah, I., & Yulita, I. N. (2022). Comparative analysis of classification algorithm: Random Forest, SPAARC, and MLP for airlines customer satisfaction. *Sinergi*, *26*(2), 213-222.

Amra, I. A. A., & Maghari, A. Y. (2017). Students performance prediction using KNN and Naïve Bayesian. 2017 8th International Conference on Information Technology (ICIT), Bellizzi, M. G., Eboli, L., Mazzulla, G., & Postorino, M. N. (2022). Classification trees for analysing highly educated people satisfaction with airlines' services. *Transport Policy*, *116*, 199-211.

Berka, P., & Bruha, I. (1998). Discretization and grouping: Preprocessing steps for data mining. European symposium on principles of data mining and knowledge discovery,

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16-28.

Chaudhury, P. R., Harshadeep, B., Yadav, M. K., & Kolluru, S. (2019). *US - Falcon Airline Passenger Satisfaction*. Institute of Management (Global Mindset - Indian Robots).

Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems*, *192*, 105361.

Cooper, W. S. (1991). Some inconsistencies and misnomers in probabilistic information retrieval. Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval,

Dash, C. S. K., Behera, A. K., Dehuri, S., & Ghosh, A. (2023). An outliers detection and elimination framework in classification task of data mining. *Decision Analytics Journal*, *6*, 100164.

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2008). *Feature extraction: foundations and applications* (Vol. 207). Springer.

Hartmann, M. J., & Carleo, G. (2019). Neural-network approach to dissipative quantum many-body dynamics. *Physical review letters*, *122*(25), 250502.

Hayadi, B. H., Kim, J.-M., Hulliyah, K., & Sukmana, H. T. (2021). Predicting Airline Passenger Satisfaction with Classification Algorithms. *International Journal of Informatics and Information Systems*, *4*(1), 82-94.

Huang, F. (2021). Network Activities Recognition and Analysis Based on Supervised Machine Learning Classification Methods Using J48 and Naïve Bayes Algorithm. *arXiv e-prints*, arXiv: 2105.13698.

Hui, S. H., Khai, W. K., XinYing, C., & Wai, P. W. (2023). Prediction of customer churn for ABC Multistate Bank using machine learning algorithms/Hui Shan Hon...[et al.]. *Malaysian Journal of Computing (MJoC)*, *8*(2), 1602-1619.

Jiang, L., Zhang, L., Yu, L., & Wang, D. (2019). Class-specific attribute weighted naive Bayes. *Pattern recognition*, *88*, 321-330.

Jiang, X., Zhang, Y., Li, Y., & Zhang, B. (2022). Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model. *Scientific Reports*, *12*(1), 11174.

Juba, B., & Le, H. S. (2019). Precision-recall versus accuracy and the role of large data sets. Proceedings of the AAAI conference on artificial intelligence,

Kaggle. (2020). *Kaggle.com*. https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction.

Kaur, H., & Kumari, V. (2020). Predictive modelling and analytics for diabetes using a machine learning approach. *Applied computing and informatics*, *18*(1/2), 90-100.

Keerthy, A., & Mathew, H. S. (2022). Feature Analysis on Airline Passenger Satisfaction using Orange Tool. 2022 6th International Conference on Electronics, Communication and Aerospace

Technology,

Korting, T. S. (2006). C4. 5 algorithm and multivariate decision trees. *Image Processing Division, National Institute for Space Research–INPE Sao Jose dos Campos–SP, Brazil*, 22.

Kumar, S., & Zymbler, M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data*, *6*(1), 1-16.

Kurniabudi, K., Harris, A., & Mintaria, A. E. (2021). Komparasi Information Gain, Gain Ratio, CFs-Bestfirst dan CFs-PSO Search Terhadap Performa Deteksi Anomali. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(1), 332-343.

Lee, C., & Lee, G. G. (2006). Information gain and divergence-based feature selection for machine learning-based text categorization. *Information processing & management*, *42*(1), 155-165.

Leon, S., & Martín, J. C. (2020). A fuzzy segmentation analysis of airline passengers in the US based on service satisfaction. *Research in Transportation Business & Management*, *37*, 100550.

Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, *17*(4), 491-502.

Malhotra, D., Malhotra, K., & Malhotra, R. (2020). Evaluating consumer loans using machine learning techniques. In *Applications of Management Science* (pp. 59-69). Emerald Publishing Limited.

Marcelino, P., de Lurdes Antunes, M., Fortunato, E., & Gomes, M. C. (2021). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, *22*(3), 341-354.

Mwadulo, M. W. (2016). A review on feature selection methods for classification tasks.

Nadali, A., Kakhky, E. N., & Nosratabadi, H. E. (2011). Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system. 2011 3rd International Conference on Electronics Computer Technology,

Rane, A., & Kumar, A. (2018). Sentiment classification system of Twitter data for US airline service analysis. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC),

Roy, S. S., Kaul, D., Roy, R., Barna, C., Mehta, S., & Misra, A. (2016). Prediction of Customer Satisfaction Using Naive Bayes, MultiClass Classifier, K-Star and IBK. International Workshop Soft Computing Applications,

Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., & Mononen, J. (2018). Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behavioural processes*, *148*, 56-62.

Saut, M., & Song, V. (2022). Influences of airport service quality, satisfaction, and image on behavioral intention towards destination visit. *Urban, Planning and Transport Research*, *10*(1), 82-109.

Shafie, S., Soek, P. O., & Khai, W. K. (2023). Prediction of employee promotion using hybrid sampling method with machine learning architecture. *Malaysian Journal of Computing (MJoC)*, *8*(1), 1264-1286.

Tiwari, M. K., Deo, R. C., & Adamowski, J. F. (2021). Short-term flood forecasting using artificial neural networks, extreme learning machines, and M5 model tree. In *Advances in streamflow* *forecasting* (pp. 263-279). Elsevier.

Verdu, S. (1998). Fifty years of Shannon theory. *IEEE Transactions on information theory*, *44*(6), 2057-2078.

Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. Icml,

Yi, J., Zhang, H., Liu, H., Zhong, G., & Li, G. (2021). Flight delay classification prediction based on stacking algorithm. *Journal of Advanced Transportation*, *2021*, 1-10.

Zhu, B., & Liu, Y. (2021). General Approximate Cross Validation for Model Selection: Supervised, Semi-supervised and Pairwise Learning. Proceedings of the 29th ACM International Conference on Multimedia,