Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

Integration of SMOTE and Ensemble Models for Predicting Airline Passenger Satisfaction

Bagus Laksono[1], Ika Kurniawati[2], Aditya Budi Sriwiyanta[3], Zen Zen Zaenudin[4], Johan Afrian Ramadha[5], Desri Alfian[6]

[123456]*Information System, Universitas Bina Sarana Informatika, Jakarta, 10450, Indonesia*

[1]*mr.bagus.laksono@gmail.com,* [2]*ika.iki@bsi.ac.id,* [3]*dityabudisriwiyanta344@gmail.com,* [4]*zenzenzaenudin@gmail.com,* [5]*johan.afrian@artajasa.co.id,* [6]*alpiandesri@datacenter.my.id.*

ARTICLE INFORMATION

Article History:
Received: December 23, 2024
Last Revision: April 26, 2025
Published *Online*: April 30, 2025

KEYWORDS

AdaBoost,
Customer satisfaction prediction,
Data mining,
Ensemble learning,
Imbalanced data,
SMOTE

CORRESPONDENCE

Phone: 083898189929
E-mail: ika.iki@bsi.ac.id

ABSTRACT

The increasing public interest in air transportation has intensified competition among airlines. To maintain their presence in the market, airlines continuously strive to improve service quality. One of the efforts undertaken is conducting passenger satisfaction surveys. However, the resulting survey data often face various challenges, such as class imbalance, missing values, data noise, difficulty in identifying significant patterns, and bias. Imbalanced classes tend to cause the classification results to favor the majority class, which can reduce the predictive performance of the model. This study proposes the integration of Synthetic Minority Over-sampling Technique (SMOTE) and AdaBoost ensemble method with Naive Bayes and Decision Tree classification algorithms to enhance classification performance. Experimental results show that the DT+SMOTE+AdaBoost and DT+SMOTE model achieved the highest accuracy of 91.88%. Meanwhile, NB+AdaBoost achieved the best overall results, with the highest accuracy of 87.62%, compared to NB+SMOTE+AdaBoost at 87.17% and the baseline NB at 82.14%. The integration of SMOTE and AdaBoost proved effective in addressing data imbalance and improving model performance compared to traditional machine learning classifiers. The results of this study demonstrate the superiority of our proposed method, a robust ensemble learning compared to traditional machine learning classifiers. This approach offers significant potential as a reference for airlines and contributes to business growth and the development of a machine learning-based customer satisfaction evaluation system.

1. INTRODUCTION

As one of the highly competitive sectors, airline companies compete to improve the quality of their services to attract more passengers and company profits. During increasingly tight competition, one important aspect that is the key to the success of airline companies is improving the quality of service. Airlines have a big challenge in meeting public expectations. Public interest and factors that influence airline passenger satisfaction are increasingly diverse in the use of airlines and produce airline passenger satisfaction data. Passenger satisfaction with excellent service not only contributes to passenger loyalty but also has a direct impact on image and reputation. Continuous improvement of service quality is essential to maximize passenger satisfaction. The dynamic and customer-centric airline industry requires constant innovation in operational services to meet the evolving expectations of passengers. Airlines that focus on improving service quality can achieve higher levels of customer satisfaction and loyalty [1]. In the highly competitive airline industry, improving passenger satisfaction is essential to maintain customer loyalty and market competitiveness. Identifying the key factors that influence passenger satisfaction and developing predictive models can help airline management make informed strategic decision [2]. Airlines need to understand the strategic value of quality, as ongoing quality improvement is not costly over time instead, it serves as an investment that can generate higher long-term returns [3].

Accurate measurement and understanding of passenger satisfaction is essential for the sustainability and growth of airlines. One way to measure and understand passenger satisfaction is through passenger satisfaction survey data. The survey data generated is not small and complex, therefore it requires the right data processing method such as machine learning with classification techniques to analyze passenger satisfaction in more depth by studying the patterns in the dataset. Identifying passenger satisfaction classification is quite a challenge for companies. The passenger satisfaction survey data collected has several problems such as imbalanced data, bias, missing values, noise, and difficulty finding significant patterns.

An imbalanced dataset can hinder the effectiveness of a predictive model. This imbalance often stems from the data collection process. It arises when one class the minority class is significantly under-represented, while the other the majority class is disproportionately dominant in the dataset. The Synthetic Minority Over-Sampling Technique (SMOTE) is widely regarded as one of the most effective methods for addressing imbalanced datasets [4]. SMOTE demonstrated superior performance compared to CTGAN in handling customer data imbalance based on various evaluation metrics [5]. Another discussion also indicated that SMOTE generally showed better performance than ADASYN and GNUS in addressing class imbalance in clinical datasets [6].

In machine learning, ensemble learning is a meta-learning technique that combines the results of several models to increase forecast accuracy. This method works particularly well for improving the efficiency of traditional machine learning algorithms, especially when dealing with issues of class imbalance. By combining the strengths of various models, training techniques, and loss functions, ensemble learning improves predictive accuracy. Its flexibility in designing new architectures makes it a promising approach for addressing imbalanced datasets. Ensemble learning is generally categorized into three main types: bagging, boosting, and stacking [7]. Boosting algorithms focus on the differences between weak classifiers by dynamically adjusting sample weights, especially on misclassified samples, so that the classification ability continues to improve. Some popular boosting algorithms include Ada Boost, GB (Gradient boosting) and XGB (Extreme Gradient Boosting). Compared with bagging, boosting is more effective in reducing sample classification bias and improving accuracy [8].

AdaBoost has several benefits over other boosting algorithms like XGB and GBDT, such as fewer parameters, simple implementation, greater versatility, and ease of comprehension. Additionally, Ada Boost provides great flexibility in choosing weak classifiers, allowing the use of both linear and nonlinear classifiers during the training process [9]. Common classification techniques used to predict binary data include: Decision Tree, Naive Bayes, NN, K-NN, SVM, Random Forest. Decision trees have demonstrated strong potential and efficiency across numerous domains. Their capability to interpret complex data and uncover patterns and relationships makes them a valuable tool in machine learning [10]. Naïve Bayes (NB) is a well-known classification algorithm commonly used in data mining. It predicts the likelihood of a new instance belonging to a particular class by assuming that all features are conditionally independent given the class label. Its effectiveness stems from the assumption that attributes are independent, even though this assumption may not always hold true in real-world datasets [11]. Naive Bayes achieves strong performance relative to other classifiers when its assumptions are satisfied. Moreover, it remains robust, maintaining high accuracy even when those assumptions are violated [12].

In this research, proposed to integrate SMOTE and Ada Boost with machine learning models, Decision Trees and Naive Bayes on unbalanced airline passenger satisfaction data, evaluated through four performance metrics accuracy, recall, precision and AUC. This study aims to provide benefits for airlines in predicting the level of passenger satisfaction with the services provided, identifying the services that have the most influence on passenger satisfaction, so that airlines can improve the quality of these services in the future, which can ultimately support increased company profits.

2. RELATED WORK

A work by [5] used CTGAN and SMOTE to generate synthetic data to solve the problem of class imbalance in machine learning models. The results showed that SMOTE performed better than CTGAN on several performance parameters. The dataset in question had imbalanced class distributions, which can lead to models favoring the majority class. Among the methods tested, SMOTE achieved the highest accuracy score of 94.06. Research conducted by [6] compared SMOTE, ADASYN, and GNUS across a wide range of datasets and models and was the first to focus on clinical data and clinical decision-making. All three data augmentation techniques significantly enhanced classification performance. In most datasets, GNUS performed comparably to SMOTE and ADASYN. Consistent with findings from several other studies, SMOTE generally outperformed ADASYN.

According to additional research by [13] the experiment showed that the SMOTE algorithm's use successfully addressed class imbalances and enhanced intrusion detection models' performance. The models' resilience in managing unbalanced data was demonstrated by the findings, which demonstrated that they retained high accuracy levels even after oversampling. In experiments carried out by [14] demonstrated that the AdaBoost classification algorithm performs effectively in managing high feature redundancy, as reflected in the strong prediction and classification outcomes. AdaBoost is a valuable method for feature classification in machine learning. During the training process, it adjusts the weights of the samples raising them when errors increase and lowering them when errors decrease.

Research that adopts an ensemble approach to detect cyber threats conducted by [15], this ensemble approach has proven effective as a solution to improve cyber security against phishing threats. This finding reinforces the importance of the ensemble approach in machine learning to improve detection or prediction capabilities. The SMOTE algorithm, Logistic Model Tree, and Ada Boost were combined in an ensemble method study by [16] that shown satisfactory to strong performance in identifying the nutritional and chlorine status of adult oil palms based on the spatial scale of leaves and canopies. The study's findings suggest a trade-off between model accuracy and spatial scale. Proving that ensemble learning is more powerful than conventional learning classifier methods.

Conversely, this study proposes a hybrid model that integrates sophisticated class balancing methods to overcome the shortcomings identified in earlier research. The model integrates machine learning methods with AdaBoost algorithms to predict airline passenger satisfaction based on survey data. Unlike previous classification studied that relied on traditional statistical approaches, this study fills the gap by using a combined methodology.

3. METHODOLOGY

The methodology outlines the technical steps that will be implemented during the research phase. This study proposes an approach to classification with the integration of SMOTE and AdaBoost algorithms to classify airline passenger satisfaction with an imbalanced dataset. The preprocessing phase plays a vital role as it directly impacts the model's performance due to the significant influence of data quality. Initially, the data is cleaned and organized to ensure it is structured and ready for processing. Following this, the data is balanced using the SMOTE oversampling technique, and various ensemble experiments are conducted using boosting and classification models. The models are then evaluated based on four performance metrics: accuracy, recall, precision, and AUC. The stages of case categorization of the airline passenger satisfaction unbalanced dataset are depicted in Fig. 1, which also demonstrates the suggested strategy. This study preceded by data collection, data pre-processing, data balancing, proposed method process, and evaluation.

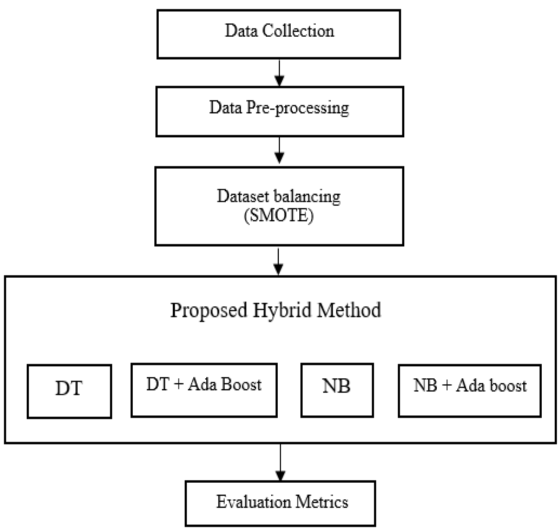

FIGURE 1. PROPOSED METHOD

3.1 Data Collection

The publicly available [23] kaggle.com site provided the dataset used in this investigation. The dataset has an xls extension and consists of 103,904 records. It has 20 attributes including 1 label attribute. This data was obtained in 2022 from the results of a passenger satisfaction survey of Citilink Indonesia airlines with a gradual approach at Soekarno-Hatta Airport. The distribution of this dataset consists of 58,879 records for the neutral or dissatisfied class and 45,025 records for the satisfied class. This dataset of customer satisfaction with Citilink Indonesia is taken from Kaggle. Each row represents a passenger and contains information about their experience, such as seat comfort, food and beverage service, check-in process, in-flight entertainment, and cleanliness. On the data provider page there is no information on the dataset from international or domestic flights. However, based on data analysis, there is a Flight Distance description attribute, which indicates the flight distance in kilometers. The maximum distance recorded is 4,983 km. This suggests that there may be some international flights included, as the distance may include routes to nearby countries. However, since Citilink operates mostly domestic routes, most of the data is likely to represent domestic flights. We focus on this entire dataset which is useful for studying what factors influence customer satisfaction with airline services. An explanation of the variables and descriptions of the dataset used in this study can be seen in Table 1.

TABLE 1. VARIABLE OF AIRLINE PASSENGER SATISFACTION DATASET

No	Attributes	Data Type	Parameters
1	Gender	Binominal	Male, Female
2	Subscription	Binominal	subscribed, unsubscribed
3	Age	Integer	(7-85) of years
4	Class	Polynomial	Eco, Business, First class
5	Flight Distance	Integer	(31-4983) km
6	Inflight Wi-Fi service	Integer	(0-5)
7	Departure/Arrival time	Integer	(0-5)
8	Ease of Online booking	Integer	(0-5)
9	Gate location	Integer	(0-5)
10	Food and drink	Integer	(0-5)
11	Online boarding	Integer	(0-5)
12	Seat comfort	Integer	(0-5)
13	Inflight entertainment	Integer	(0-5)
14	On-board service	Integer	(0-5)
15	Leg room service	Integer	(0-5)
16	Baggage handling	Integer	(1-5)
17	Checking service	Integer	(0-5)
18	Inflight service	Integer	(0-5)
19	Cleanliness	Integer	(0-5)
20	Satisfaction	Binominal	Satisfied, dissatisfied

3.2 Preprocessing

The dataset used in this study is clean and contains no missing values. However, at this stage, we eliminated irrelevant attributes such as age and gender to avoid bias, reduce model complexity, and facilitate easier interpretation of the results. Removing demographic attributes like age and gender also helps ensure that the model is not directly or indirectly biased toward certain groups. In this research, our focus is on operational attributes or passenger behaviors such as baggage handling, cleanliness, inflight Wi-Fi service, online boarding, legroom service, and others so demographic variables like age and gender are considered irrelevant to the study's objective, which is to identify the services that most influence passenger satisfaction to help airlines improve them. Based on test results show that age and gender had no discernible impact on model performance, these characteristics were excluded.

The next stage is transformation. Steps included in the transformation are attribute transformation, dimension reduction, data aggregation, and others. In the target variable: neutral or dissatisfied are combined into dissatisfied. The number of entries in the class doesn't necessarily indicate dissatisfaction; it may also reflect passengers who are neutral about their flight experience.

3.3 Data Balancing

Figure 2 presents the distribution details of the airline passenger satisfaction dataset. It clearly shows an imbalance in the data.

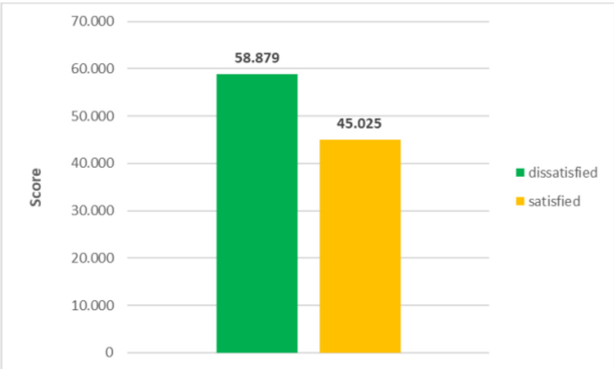

FIGURE 2. NUMBER OF DISTRIBUTION CLASS SATISFIED AND DISSATISFIED

The dataset consisted of 58,879 dissatisfied passengers and 45,025 satisfied passengers, indicating a class imbalance. To address this issue, we applied the SMOTE method to generate synthetic samples and balance the class distribution before training the model.

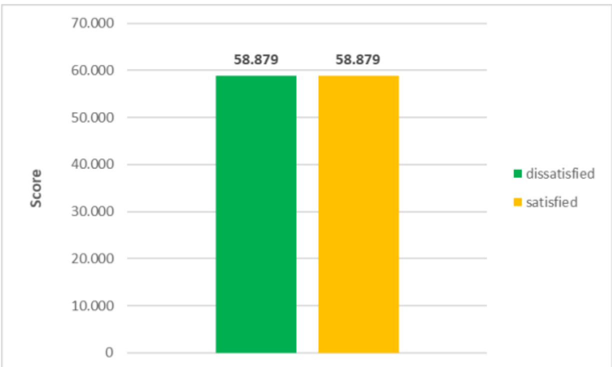

FIGURE 3. NUMBER OF DISTRIBUTION CLASS SATISFIED AND DISSATISFIED AFTER APPLIED SMOTE

After applied the SMOTE oversampling technique it becomes balanced with the proportions of each 58.879 for satisfied and 58.879 for dissatisfied.

3.4 Proposed Model

In this study, RapidMiner AI Studio is used as a tool to classify airline passenger satisfaction. Classification uses binary assignment which has two class labels. The classification algorithms used in this study are Decision Tree and Naive Bayes which are integrated with SMOTE and Ada Boost.

3.4.1 Decision Tree

The decision tree learning process consists of several steps where the dataset is divided into homogeneous groups, as illustrated in Figure 4. It begins at the root node, which represents the complete dataset. The algorithm selects the most suitable feature and threshold for splitting the data based on a specific evaluation metric. This splitting process is recursive, with each resulting subset being further divided at the child nodes. The process continues until a stopping criterion is met typically when the nodes become pure (i.e., all instances belong to the same class) or when a maximum tree depth is reached. The final nodes, called leaf or terminal nodes, indicate the predicted class labels. Splitting decisions at each node are guided by mathematical measures such as information gain, Gini impurity, or variance reduction [10].

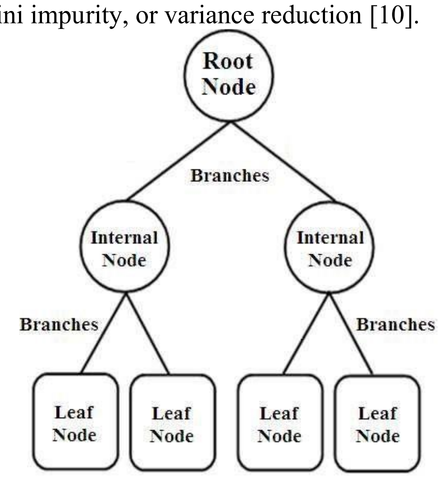

FIGURE 4. A DECISION TREE EXAMPLE

Decision trees are widely appreciated for their natural interpretability, which makes them especially useful in fields where it is important to understand how decisions are made [17]. Decision trees provide clarity by outlining the decision-making process through a set of easily comprehensible principles, in contrast to many machine learning algorithms that operate as black-box models. The path from the root to a leaf node reflects a sequence of judgments influenced by feature values, whereas each node represents a distinct feature and its decision threshold. Stakeholders can readily follow and understand how the model makes its predictions thanks to its clear and straightforward structure [10].

3.4.2 Naïve Bayes

Naive Bayes Classifier is considered one of the most efficient classifiers, as its predictive accuracy rivals that of many advanced, state-of-the-art models. This classifier uses the class label C from the training data to estimate the conditional probability of each attribute Ai. The class with the highest resultant posterior probability is then predicted using Bayes' theorem to determine the likelihood of C given a particular instance described by $A1,\ldots,An$. Given the class C, the model assumes that all characteristics Ai are conditionally independent of one another to streamline this calculation. When we talk about independence, we're talking about probabilistic independence, which means that every time $Pr(C) > 0$ and $Pr(A \mid B,C) = Pr(A \mid C)$ for any conceivable value of A,B, and C, A is independent of B given C [18].

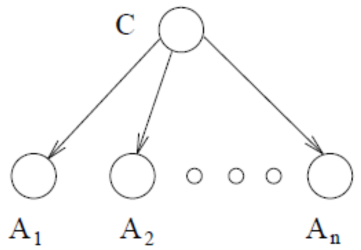

FIGURE 5. STRUCTURE OF THE NAIVE BAYES NETWORK

When visualized as a Bayesian network, Naive Bayes takes on the straightforward structure shown in Figure 5. This network captures the main assumption behind the naïve Bayesian classifier, namely, that every attribute (every leaf in the network) is independent from the rest of the attributes, given the state of the class variable (the root in the network). Naive Bayes is well-suited for classification problems, especially when the dataset has many features and limited instances, making it computationally efficient and easy to implement. Naive Bayes explicitly calculates prior probabilities for each class, which allows it to account for class imbalance. Even if one class is underrepresented, its prior probability ensures that it is not entirely ignored in the classification process. The algorithm is computationally lightweight and requires minimal training time, making it effective for large-scale datasets that may include class imbalances. It is particularly advantageous when combined with preprocessing techniques like oversampling or under sampling. Unlike more complex models, Naive Bayes is less prone to overfitting, especially in small or imbalanced datasets. This robustness stems from its simplicity and reliance on probabilities rather than complex decision boundaries [19].

3.4.3 ADABoost

One of the most widely studied boosting algorithms is Ada Boost, which has been applied across various fields. As a meta-algorithm, Ada Boost builds an ensemble by integrating multiple low-accuracy models to form a more reliable and accurate overall model. With parameters like the number of estimators, a learning rate (where 0 indicates no learning and 1 gives full weight to the most recent input), and a fixed seed for the random number generator, it usually uses a tree-based technique for base predictions [14]. It operates by repeatedly training a base classifier on different subsets of the data, placing greater emphasis on instances that were misclassified in previous rounds. The core idea behind Ada Boost is to adjust the weight of each training observation in every iteration, allowing the model to focus more on the difficult-to-classify samples [5]. The weight of an observation, denoted as wi, is updated using the following equation.

$$wi = (0{,}5)\ x\ 1n \frac{1- error_i}{0}\ error_i \quad (1)$$

$error_i$ is the base classifier's iteration misclassification rate.

After updating the weights for each observation, the base classifier is retrained using these adjusted weights. The final prediction is then generated by aggregating the outputs of all base classifiers, with each classifier's influence weighted according to its accuracy [5]. The formal expression for this aggregation determines the final prediction.

$$f(x) = sign\left(\sum i = \text{In}\alpha_i \text{h(x)}\right) \quad (2)$$

where h(x) is the ith classifier's prediction, and I is the ith classifier's weight.

ADABoost is a powerful ensemble technique commonly applied across numerous fields such as computer vision, NLP, and bioinformatics. It enhances the performance of weak classifiers by reducing both bias and variance. Moreover, Ada Boost is computationally efficient and easy to implement. However, due to its sensitivity to noise and outliers, careful data pre-processing is crucial before applying the algorithm [5].

3.4.4 SMOTE

The SMOTE over-sampling algorithm, introduced by Chawla in 2002 [20] , generates synthetic training samples for the minority class using a linear interpolation technique. A popular technique for creating synthetic data to rectify class imbalance in datasets is SMOTE. It operates by using preexisting instances of the minority class to create new ones. SMOTE does this by using the k-nearest neighbors' technique to find minority class samples in the feature space that are near to one another. It then produces synthetic data points by interpolating between a given minority instance and its nearest neighbors.

This interpolation is performed using the following equation [21].

$$New\ Sample = MI + (RM \times (NN - MI)) \quad (3)$$

Where:
MI = Minority Instance;
RM = Random Number;
NN = Nearest Number.

Unlike simple duplication, SMOTE helps prevent overfitting by creating new, diverse samples rather than replicating existing ones. The approach computes the difference between a chosen minority instance's feature vector and that of one of its closest neighbors. A new synthetic sample is then created by scaling this difference by a random number between 0 and 1 [13].

3.4.5 Evaluation Metrics

In this section, the outcomes are presented based on a series of experimental evaluations. 10-fold cross-validation was used to gauge how well the machine learning models performed. This method partitions the dataset into equal subsets to enable repeated and reliable testing. The models' effectiveness was assessed using standard evaluation metrics such as AUC, accuracy, recall, and precision. These evaluation techniques play a crucial role in thoroughly validating machine learning models across diverse application domains.

This stage is an activity of the process of extracting (mining) patterns from existing data using appropriate algorithms and testing the accuracy and visualization of the extracted data. Classification performance in machine learning can be calculated using classification performance, namely, TP, FP, TN, and FN. Binary misunderstanding occurs when analysis results are appropriately divided into two classes: True Positive (TP) and True Negative (TN), which are negative classes. False Positive (FP) data are those from the negative class that are mistakenly categorized as positive, and False Negative (FN) data are those from the positive class that are mistakenly classed as negative.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

The classification levels on the data for the algorithm are as follows. Excellent = 0.90 - 1.00, Good = 0.80 - 0.90, Fair = 0.70 - 0.80, Low = 0.60 - 0.70.

4. RESULT AND DISCUSSION

The testing stage of the classification model on the airline passenger satisfaction dataset by integrating SMOTE and AdaBoost. The dataset consists of 103,904 records, the number of distributions for each class in the satisfaction attribute in this dataset consists of 58,879 records for the neutral or dissatisfied class and 45,025 records for the satisfied class, the attributes used are 20 attributes included 1 label.

The evaluation uses k-fold cross-validation on the dataset. Based on research conducted by Liu dan Cocea, 10-fold cross-validation is the ideal choice to get accurate estimates. 10-fold cross-validation is a robust model evaluation technique used in machine learning to assess a model's performance and generalizability. The data is randomly divided into 10 equal parts, or "folds." For each iteration, the model is trained on 9 of these folds and tested on the remaining fold, ensuring that each fold serves as the test set once. This process is repeated 10 times, and the overall performance is averaged across all iterations. This technique helps to mitigate overfitting, providing a more reliable estimate of how well the model will perform on unseen data. It is particularly useful for smaller datasets, where using a single train-test split might not provide a sufficient understanding of the model's performance [22].

In Decision Tree (DT), dataset is tested using cross validation to ensure the machine learning model runs optimally. Furthermore, the integration of oversampling techniques with SMOTE and AdaBoost ensemble or a combination of both is carried out. In DT, the attribute that most influences airline passenger satisfaction is inflight Wi-Fi service. However, after the SMOTE oversampling technique was carried out, the attribute that most influences airline passenger satisfaction is online boarding. In traditional DT models, data imbalanced can may lead the model to favor the majority class, resulting in biased predictions. This results in attributes that are more frequently present in the majority class being selected more often as significant features because the model focuses on predicting that class. Attributes that are more relevant to the minority class may emerge as less frequent in unbalanced data distributions.

After applied SMOTE, the minority class data becomes more balanced, giving the model the opportunity to "notice" more patterns relevant to this class. This allows attributes such as "online boarding," which may have previously been underrepresented in the analysis, to become more significant. This attribute shift occurs because the baseline DT model tends to select only the features with the strongest influence on a lopsided dataset, which can lead to bias towards certain attributes. However, after the integration of SMOTE and Ada Boost, the model's sensitivity to minority attributes increases. This reflects that the optimized model can detect patterns that were previously hidden due to the imbalanced data.

The accuracy results of DT integrated with SMOTE and AdaBoost are proven to increase by 0.25%, which is 91.88% compared to DT without SMOTE and Adabost integration of 91.63%, while DT+AdaBoost obtains an accuracy of 91.82%.

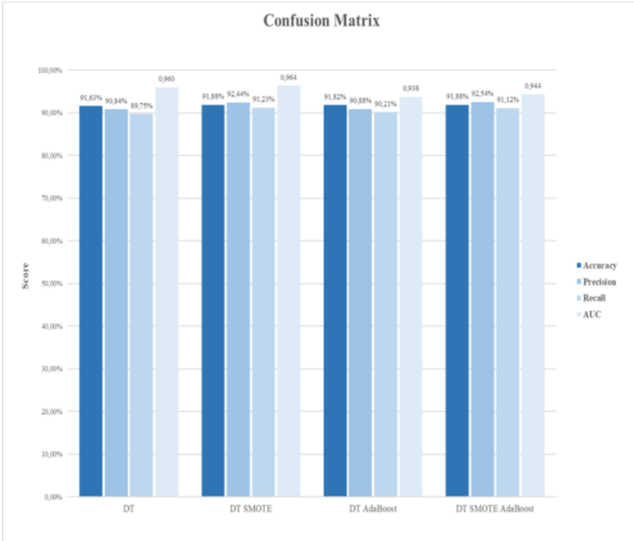

FIGURE 6. COMPARISON OF PROPOSED MODEL DECISION TREE

The AUC results on the ROC Curve obtained the highest value in the DT+SMOTE model with a value of 0.964 compared to other models. The AUC value ranges from 0.90 - 1.00, including the Excellent classification. Visualization of the comparison of model testing results can be seen in Figure 6. Naïve Bayes is easy to create, does not require a complicated iteration parameter estimation scheme. The accuracy results of the Naïve Bayes model integrated with SMOTE, and the Ada Boost ensemble have been shown to increase by 5.03% from the previous 82.14% (NB) which is 87.17%. The highest accuracy was obtained in the NB + AdaBoost model with an accuracy of 87.62%. While the integration of the NB+SMOTE model decreased the accuracy to 81.63, this is because before oversampling was carried out, there were many biases that ignored the minority class.

The AUC results on the ROC Curve obtained the highest value in the model NB+AdaBoost with a value of 0.939 compared to other models. The AUC value ranges from 0.90 - 1.00, including the Excellent classification. Visualization of the comparison of model testing results can be seen in Figure 7. This high AUC value indicates the model's strong discriminative ability in distinguishing between positive and negative classes, confirming its robustness. The combination of Naïve Bayes with AdaBoost effectively enhances classification performance by reducing bias and variance, contributing to a more reliable and generalizable predictive model.

The effectiveness of this method is validated through rigorous experiments, which show that the integration of

SMOTE and ADABoost ensemble with DT and NB classification algorithms offers a more robust solution for classifying airline passenger satisfaction.

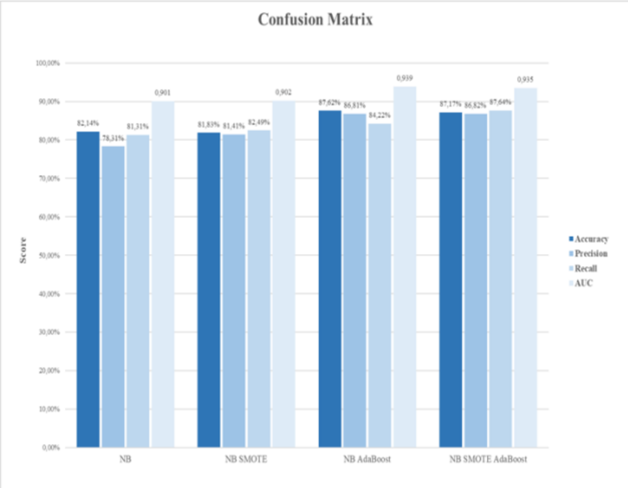

FIGURE 7. COMPARISON OF PROPOSED MODEL NAÏVE BAYES

This hybrid approach has the potential to address real-world challenges by improving the accuracy and efficiency of predictive models, especially in areas involving imbalanced, complex, noisy, or high-dimensional and bias data.

TABLE 2. PERFORMANCE COMPARISON OF THE PROPOSED METHODS

Methods	Accuracy	Precision	Recall	AUC
DT	91.63	90.84	89.75	0.960
DT + SMOTE	**91.88**	92.44	91.23	**0.964**
DT + AdaBoost	91.82	90.88	90.21	0.938
DT + SMOTE+ AdaBoost	**91.88**	92.54	91.12	0.944
NB	82.14	78.31	81.31	0.901
NB + SMOTE	81.83	81.41	82.49	0.902
NB + AdaBoost	**87.62**	86.81	84.22	**0.939**
NB + SMOTE+ AdaBoost	87.17	86.82	87.64	0.935

Based on the performance on table 2, ensemble models based on DT demonstrated relatively high and stable performance. However, the combination of DT + SMOTE produced the best overall results, particularly with the highest AUC (0.964) and accuracy (91.99 %), indicating the proposed model's strong ability to classify with both accuracy and balance. The performance comparison results of the integration of SMOTE and Ada boost on DT are proven to be able to overcome imbalanced data and improve accuracy compared to DT. The proposed method indicate that the AdaBoost combined with Naive Bayes (NB) outperforms the model without the boosting technique. Based on the performance table, the Naive Bayes model combined with AdaBoost (NB + AdaBoost) achieved the best overall results, with the highest accuracy (87.62%) and AUC (0.939). While the NB + SMOTE + AdaBoost model showed a slightly higher recall (87.64%), it did not outperform NB + AdaBoost in terms of accuracy and AUC. This indicates that adding SMOTE did not provide significant improvement when combined with AdaBoost. Therefore, NB + AdaBoost can be considered the most effective ensemble model in this experiment.

Integrating SMOTE with Naive Bayes can sometimes lead to reduced accuracy because SMOTE synthetically generates new samples by interpolating between minority class instances, which might introduce noise or create samples that don't align well with Naive Bayes' assumption of feature independence. Naive Bayes performs better when the data adheres closely to its probabilistic model, and the added synthetic data may distort this balance, leading to suboptimal decision boundaries. Additionally, if the original dataset is only mildly imbalanced, SMOTE might overcompensate, increasing overlap between classes and degrading model performance [23].

Naive Bayes performs well on datasets with many features. Since it assumes feature independence, it can handle high-dimensional data efficiently, even when the data is imbalanced. Naive Bayes can be combined with Ada Boost to further improve its performance on imbalanced datasets. These methods enhance the representation of minority classes while retaining the simplicity of Naive Bayes. Naive Bayes provides probabilistic outputs that indicate the likelihood of each class, which can be useful for threshold tuning to optimize performance on imbalanced datasets. This enables flexible decision-making depending on the application, such as emphasizing the minority class in critical scenarios. Naive Bayes is a strong candidate for handling imbalanced data when paired with proper data preprocessing and model optimization techniques such as Ada Boost. The results obtained in this study underline the significant potential of utilizing a combined bagging and boosting ensemble as a model that can help predict airline passenger satisfaction dataset. These results demonstrate the superiority of using a powerful combined bagging and boosting ensemble compared to traditional machine learning classifiers.

5. CONCLUSIONS

This research introduces an enhanced AdaBoost classification framework that incorporates weighted feature selection and noise-aware confidence levels to improve predictive performance. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied, effectively balancing data distribution, reducing model bias toward majority classes, and enhancing overall classification accuracy. Additionally, the study tackles the challenges of high feature redundancy, outliers, noise, and missing values, all of which can adversely affect prediction reliability. The proposed hybrid approach integrates SMOTE with the AdaBoost ensemble technique and a decision tree (DT) classifier. Experimental results demonstrate that this combination yields superior performance across key evaluation metrics. The DT+SMOTE+AdaBoost model achieved an accuracy of 91.88%, identical to the DT+SMOTE model, while the highest AUC score (0.964) was also recorded by DT+SMOTE. Notably, the application of AdaBoost to Naïve Bayes (NB) increased its accuracy by 5.03%, underscoring AdaBoost's effectiveness in mitigating high feature redundancy and enhancing classification robustness. Conversely, the combination of SMOTE and NB resulted in reduced accuracy. This decline is attributed to the inherent assumption of feature independence in NB, which is compromised by SMOTE's synthetic data generation. The introduction of interpolated samples may distort original data distributions, particularly when the synthetic data fails to accurately capture the characteristics of the minority class. The findings affirm the effectiveness of ensemble learning combined with oversampling in complex, imbalanced datasets, such as the airline passenger satisfaction dataset. Based on decision tree insights, actionable recommendations for airlines include enhancing online boarding services by optimizing educational content, simplifying user interfaces, integrating accessibility features and offering multilingual support. Ultimately, the proposed model not only improves predictive performance but also provides valuable guidance for operational enhancements and strategic decision-making in the airline industry.

REFERENCES

[1] S. Walia, D. Sharma, and A. Mathur, "The Impact of Service Quality on Passenger Satisfaction and Loyalty in the Indian Aviation Industry," *International Journal of Hospitality and Tourism Systems; Vol 14, No 2 (2021)*, Jul. 2022.

[2] W. S. Ismail, H. Bawazeer, and H. Almansori, "Predictive Analytics for Enhanced Passenger Satisfaction in the Airline Industry: Leveraging Machine Learning to Drive Strategic Decision-Making," in *2024 10th International Conference on Optimization and Applications (ICOA)*, 2024, pp. 1–6. DOI: 10.1109/ICOA62581.2024.10753807

[3] K. Hulliyah, "Predicting Airline Passenger Satisfaction with Classification Algorithms," *IJIIS: International Journal of Informatics and Information Systems*, vol. 4, no. 1, pp. 82–94, 2021. DOI: 10.47738/ijiis.v4i1.80

[4] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Machine Learning*, vol. 113, no. 7, pp. 4903–4923, 2024. DOI: 10.1007/s10994-022-06296-4

[5] N. Ahmad, M. J. Awan, H. Nobanee, A. M. Zain, A. Naseem, and A. Mahmoud, "Customer Personality Analysis for Churn Prediction Using Hybrid Ensemble Models and Class Balancing Techniques," *IEEE Access*, vol. 12, no. October 2023, pp. 1865–1879, 2024. DOI: 10.1109/ACCESS.2023.3334641

[6] J. Beinecke and D. Heider, "Gaussian noise up-sampling is better suited than SMOTE and ADASYN for clinical decision making," *BioData Mining*, vol. 14, no. 1, pp. 1–11, 2021. DOI: 10.1186/s13040-021-00283-6

[7] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Systems with Applications*, vol. 244, no. May 2023, p. 122778, 2024. DOI: 10.1016/j.eswa.2023.122778

[8] Y. Wang and L. Feng, "Improved Adaboost Algorithm for Classification Based on Noise Confidence Degree and Weighted Feature Selection," *IEEE Access*, vol. 8, pp. 153011–153026, 2020. DOI: 10.1109/ACCESS.2020.3017164

[9] Y. Wang and L. Feng, "An adaptive boosting algorithm based on weighted feature selection and category classification confidence," *Applied Intelligence*, vol. 51, no. 10, pp. 6837–6858, 2021. DOI: 10.1007/s10489-020-02184-3

[10] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, no. June, pp. 86716–86727, 2024. DOI: 10.1109/ACCESS.2024.3416838

[11] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Systems*, vol. 192, p. 105361, 2020. DOI: 10.1016/j.knosys.2019.105361

[12] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation," *Soft Computing*, vol. 25, no. 3, pp. 2277–2293, 2021. DOI: 10.1007/s00500-020-05297-6

[13] A. O. Widodo, B. Setiawan, and R. Indraswari, "Machine Learning-Based Intrusion Detection on Multi-Class Imbalanced Dataset Using SMOTE," *Procedia Computer Science*, vol. 234, pp. 578–583, 2024. DOI: 10.1016/j.procs.2024.03.042

[14] C. Karima and W. Anggraeni, "Performance Analysis of the Ada-Boost Algorithm For Classification of Hypertension Risk With Clinical Imbalanced Dataset," *Procedia Computer Science*, vol. 234, pp. 645–653, 2024. DOI: 10.1016/j.procs.2024.03.050

[15] D. Rofianto, E. Safitri, K. Amaliah, J. Fitra, and A. Hijriani, "Cyber Threat Detection Using an Ensemble Model Approach for Phishing Website Identification," *Innovation in Research of Informatics (INNOVATICS)*, vol. 2, pp. 81–89, 2024.

[16] A. D. Amirruddin, F. M. Muharam, M. H. Ismail, N. P. Tan, and M. F. Ismail, "Synthetic Minority Over-sampling TEchnique (SMOTE) and Logistic Model Tree (LMT)-Adaptive Boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (Elaeis guineensis) using spectroradiometers and u," *Computers and Electronics in Agriculture*, vol. 193, no. January 2021, 2022. DOI: 10.1016/j.compag.2021.106646

[17] V. G. Costa and C. E. Pedreira, "Recent advances in decision trees: An updated survey," *Artificial Intelligence Review*, vol. 56, no. 5, pp. 4765–4800, 2023.

[18] Y. Zhu, Y. Wang, L. Qin, B. Zhang, B.-C. Shia, and M. Chen, "Naïve Bayes classifier based on reliability measurement for datasets with noisy labels," *Annals of Operations Research*, 2023. DOI: 10.1007/s10479-023-05671-1

[19] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Constrained Naïve Bayes with application to unbalanced data classification," *Central European Journal of Operations Research*, vol. 30, no. 4, pp. 1403–1425, 2022. DOI: 10.1007/s10100-021-00782-1

[20] J. Wei, Z. Lu, K. Qiu, P. Li, and H. Sun, "Predicting drug risk level from adverse drug reactions using smote and machine learning approaches," *IEEE Access*, vol. 8, pp. 185761–185775, 2020. DOI: 10.1109/ACCESS.2020.3029446

[21] N. Ahmad, M. J. Awan, H. Nobanee, A. M. Zain, A. Naseem, and A. Mahmoud, "Customer Personality Analysis for Churn Prediction Using Hybrid Ensemble

Models and Class Balancing Techniques," *IEEE Access*, vol. 12, no. October 2023, pp. 1865–1879, 2024. DOI: 10.1109/ACCESS.2023.3334641

[22] H. Liu and M. Cocea, "Semi-random partitioning of data into training and test sets in granular computing context," *Granular Computing*, vol. 2, no. 4, pp. 357–386, 2017. DOI: 10.1007/s41066-017-0049-2

[23] N. S. Rahmi, N. W. S. Wardhani, M. B. Mitakda, R. S. Fauztina, and I. Salsabila, "SMOTE Classification and Random Oversampling Naive Bayes in Imbalanced Data : (Case Study of Early Detection of Cervical Cancer in Indonesia)," in *2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*, 2022, pp. 1–6. DOI: 10.1109/ICITDA55840.2022.9971421

AUTHORS

Bagus Laksono

He was completed his Diploma in 2014 and currently continuing his Bachelor's Program at Bina Sarana Informatika University. He also works as a Finance and Accounting Administration at Serpong Utara of Government Public Hospital.

Ika Kurniawati

She is a lecturer in the Department of Engineering and Informatics, Bina Sarana Informatika University and Nusa Mandiri University. Her current research interests include machine learning, deep learning, and information systems.

Aditya Budi Sriwiyanta

He was completed his Diploma in 2023 and currently continuing his Bachelor's Program at Bina Sarana Informatika University.

Zen Zen Zaenudin

He was completed his Diploma in 2015 and currently continuing his Bachelor's Program at Bina Sarana Informatika University.

Johan Afrian Ramadha

He was completed his Diploma in 2023 and currently continuing his Bachelor's Program at Bina Sarana Informatika University. He also works as a System Engineer at Artajasa Pembayaran Elektronis.

Desri Alfian

He was completed his Diploma Degree in 2013 and currently continuing his Bachelor's Program at Bina Sarana Informatika University. He is a certified Amil and Manager of Dakwah and Advocacy Programs at Laznas Dewan Dakwah.