INTERNATIONAL JOURNAL OF SOFTWARE ENGINEERING & COMPUTER SYSTEMS (IJSECS)
ISSN: 2289-8522 e-ISSN: 2180-0650
VOL. 8, ISSUE 2, 67 – 76
DOI: https://doi.org/10.15282/ijsecs.8.2.2022.7.0104

ORIGINAL ARTICLE

Ensemble Learning for the Prediction of Marketing Campaign Acceptance

F. Titiani[1,*], D. Riana[1]

[1]Computer Science Graduate Program , University Nusa Mandiri, Jakarta, Indonesia.

ABSTRACT – Artificial intelligence, commonly known as AI, has greatly influenced marketing strategies, including business models, sales processes, customer service options, and customer behaviour in receiving marketing campaigns. In a marketing campaign, all customers are targeted by advertising, including those who will not respond positively to the marketing campaign and reject the offer. This means that the company is inefficient; the marketing campaign is ineffective because customers are not segmented and targeted. As a result, costs increase and the company's profit decreases. Thence, this leads to the failure of the company's marketing campaigns. The purpose of this study is to experiment with using Ensemble Learning and tuning on the Marketing Campaign dataset by providing the classification methods. That classification method is called the Light Gradient Boosting Machine (LightGBM), Gradient Boosting Classifier (GBC), and AdaBoost Classifier (ADA), which have never been used in the classification of the Marketing Campaign dataset. The study results in the highest model with an accuracy value of 98.64%, AUC 0.9994, recall 95.77%, precision 95.77%, F1-score 95.77%, and kappa 94.98% when using the LightGBM for marketing campaign predictions.

ARTICLE HISTORY
Received: 18 February 2022
Revised: 2 July 2022
Accepted: 27 September 2022

KEYWORDS
Ensemble Learning
Marketing Campaign
LightGBM
Gradient Boosting
AdaBoost

INTRODUCTION

The high pace of business change and continuous digital transformation in the global environment has forced modern companies to remain agile and competitive by developing their business processes. It is based on the dynamic concept of the company's ability to maintain and strengthen its competitive advantage, especially in times of market uncertainty and intense competition in creating, updating, and managing resources and assets [1]. Machine Learning (ML) has been widely used in the field of digital marketing to communicate a company's products or services to existing or potential customers through advertising, in-product recommendations, and customer service [2].

Utilizing the internet, social media, and other online technologies to promote goods and brands is known as digital marketing. Digital marketing helps businesses connect with their audience, improve user experience, increase customer impact, and ultimately boost net profitability (ROI) [3]. A marketing campaign is a strategy for contracting out marketing work to a third-party organization in order to increase business finances, profits, and get an advantage against "peers" [4]. In a marketing campaign where all customers are targeted by advertising, including those who will not respond positively to the marketing campaign and those who will reject the offer. This means that the company does not work efficiently; its marketing campaign is not optimal because customers are not segmented and on target. As a result, the campaign costs will increase, and the company's profit will not be maximized. This can lead to the failure of the company's marketing campaign [3].

Ensemble learning and classification systems consist of several classifiers or subsystems. The goal of ensemble learning is to increase accuracy by combining the responses generated by these classifiers into a single response or output. In this framework, the final ensemble system's decisions are generally wrong when most of the sub-systems make the same error, because the final output is calculated by combining the outputs of the different classifiers [10].

LightGBM is a machine learning gradient augmentation framework. This framework uses tree-based decision making. Another algorithm tree is raised horizontally while LightGBM advances a tree that spreads vertically [5]. Gradient Boosting is a machine learning algorithm that can solve regression and classification problems. Gradient Boosting generates a predictive model consisting of an ensemble of weak prediction models in a decision tree [6]. The concept of a gradient enhancer decision tree is to combine a series of weak base classifiers into one strong one. GBDT is a traditional boosting method that weighs positive and negative samples. It builds a global convergence algorithm by following the negative gradient direction [7]. Adaptive Booster, or AdaBoost, is one of the algorithms in ML developed by Freund and Schapire. Adaboost gives more weight to weak classifications [8]. Adaboost uses different classifiers for the same training set, and then combines them to build the final strongest classifier. The algorithm itself is implemented by changing the weight distribution of D, which is initialised consistently and then changed to the next classifier [9].

The same dataset has been used in other studies [3, 10, 11] that employed Linear Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), and Multi-level Perceptron (MLP) methods. The study of these findings shows that the LR classifier produces the data's highest level of accuracy. It has the greatest average F1 score of 0.57, the highest average ROC AUC score of 0.88, the highest average accuracy of 0.87, and the best average

*CORRESPONDING AUTHOR | F. Titiani | ✉ 14002417@nusamandiri.ac.id

precision of 0.061. The findings of this investigation are not ideal, and the accuracy can still be increased. Customer data is preprocessed using feature engineering, input for missing values, and categorical to numeric conversion. To create datasets that are ready to be processed, data splits and SMOTE procedures are also carried out. Utilizing Ensemble Learning and the LightGBM, GBC, and ADA techniques, a classification model is built using customer data to get the best and most precise classification model with the tuning method to predict which customers will accept or reject the marketing campaign offer. Because of the ability to segment and target customers, this objective can optimize marketing strategies to ensure that the organization makes the most money possible.

However, an Esemble Learning technique exists that can increase classification accuracy [12]. This research will conduct an experiment using Esemble Learning and tuning on marketing campaign data [13]. Furthermore, this research will employ classification techniques that have never been used before, such as Light Gradient Boosting Machine (LightGBM), Gradient Boosting Classifier (GBC), and AdaBoost Classifier (ADA). The results of the model structure utilised in this research indicate an improvement in the value of the classification accuracy of which customers will accept and which customers will reject the marketing campaign.

RELATED WORK

We propose a categorization model using customer data based on the same dataset as in [3]. We use the LR, RF, SVM, NB, and MLP models to classify which customers will accept and which customers will reject the offer. The results showed that the LR classifier had the best model performance with the highest average accuracy of 0.87, the highest average precision of 0.61, the highest average F1 score of 0.57 and the highest ROC AUC score of 0.88.

Other research using marketing campaign dataset [10] discusses customer segmentation based on product/service responses where the segmentation is done using the RFM algorithm with segmentation class customer response having 2 classes. A class with a value of 1 means responding to products or services offered, while a class with a value of 0 means not responding to products or services offered. It is concluded that the best predictor model of the target variable is the RF model.

Ravi et al. [11] discuss clustering analysis using the K-means algorithm with a Differentially Private (DP) mechanism and adding Gaussion Noise with optimum covariance. The data used as a sample uses the Marketing Campaign dataset. From the results of the study, it is known that the DP mechanism maintains the distribution of calculations from various dataset metrics, thus leading to accurate conclusions (cluster results are taken from non-noise groupings).

J.Asare-Frempong, et al. [4] predict customer response to marketing campaigns. The experiment was carried out using the Waikato Environment for Knowledge Analysis (WEKA). The results showed that RF had an accuracy of 86.8% and an ROC AUC score of 92.7%, a Decision Tree Accuracy of 84.7% and an AUC ROC score of 87.7%, a Multi-layer Perceptron Neural Network Accuracy of 82.9% and an ROC AUC score of 90.0%, a LR Accuracy of 83.5%, and an AUC ROC score of 90.9%.

S.P. Singh, et al. [5] aiming at predicting the potential applicants who are likely to take admission is an important aspect of a college or university using LightGBM. The results of this study show that the LightGBM algorithm achieves an accuracy rate of 98.5%.

S. Lahmiri et al. [12] also used the ensemble learning method, which aims to assess the relative performance of the existing sophisticated classification system and ensemble learning with the application of corporate bankruptcy prediction and credit scoring. Additionally, S. Lahmiri et al. [12] used the ensemble learning approach, which intends to compare the effectiveness of the current sophisticated classification system and ensemble learning with the application of corporate bankruptcy prediction and credit scoring. AdaBoost, LogitBoost, RUSBoost, subspace, and ensemble bagging systems are among the ensemble systems under investigation. Five types of ensemble systems were tested on three different ones, including two data sets on corporate bankruptcy predictions and one data collection on credit scoring. By using the tenfold cross validation method, experimental results show that AdaBoost is effective in misclassification, has limited complexity, and has a short data processing time. Experimental results show that the ensemble classification system outperforms existing models that were recently validated on the same database.

METHODOLOGY

(i) Dataset Description

This marketing campaign dataset was created by Rodolfo Sadanha [13] last update two years ago. The purpose of this dataset is to predict who will respond to a product or service offer. The data contains marketing and features of various kinds that describe customer behavior and buying habits. The data set used in this study has 2240 instances and 29 features. Attributes and their descriptions and dataset characteristics from the Marketing Campaign dataset [13] can be seen in Table 1 and Table 2.

Table 1. Dataset attributes

Feature	Defenition
ID	identifier
YearBirth	client's birth year
DtCustomer	client's registration date in the firm
Education	client's educational degree
Marital_Status	client's relationship status
Kidhome	amount of kids in client's home
Teenhome	amount of teenagers in client's home
Income	client's annual salary
MntFishProducts	money spent on fish in the past 2 years
MntMeatProducts	money spent on meat in the past 2 years
MntFruits	money spent on fruits in the past 2 years
MntSweetProducts	money spent on sweets in the past 2 years
MntWines	money spent on wine in the past 2 years
MntGoldProducts	money spent on gold in the past 2 years
NumDealsPurchases	amount of deals conducted with deduction
NumCatalogPurchases	amount of deals conducted through the catalogue
NumStorePurchases	amount of deals conducted in shops
NumWebPurchases	amount of deals conducted via firm's web page in the past month
NumWebVisitsMonth	amount visits to the firm's web page in the past month
Recency	amount of days from the previous client's purchase
Complain	1 if a client complained in a past 2 years
AcceptedCmp1	1 if a client confirmed a marketing proposal in 1st campaign, if not 0
AcceptedCmp2	1 if a client confirmed a marketing proposal in 2nd campaign, if not 0
AcceptedCmp3	1 if a client confirmed a marketing proposal in 3rd campaign, if not 0
AcceptedCmp4	1 if a client confirmed a marketing proposal in 3rd campaign, if not 0
AcceptedCmp5	1 if a client confirmed a marketing proposal in 4th campaign, if not 0
Z_CostContact	contact fee
Z_Revenue	revenue
Response	1 if a client confirmed a marketing proposal in latest campaign, if not 0

Table 2. Dataset characteristics

No	Characteristics	Value
1	Missing values	24
2	Input variables	Numeric
3	Target variable	Numeric
4	No of instances	2240
5	No of features	29

The dataset consisting of 2240 instances has two class labels, namely accept is interpreted as 0 and refuses to be interpreted 1. Based on data 1906 customers rejected marketing campaigns, and 334 customers accepted marketing campaigns. Class distribution statistics can be seen in Figure 1.

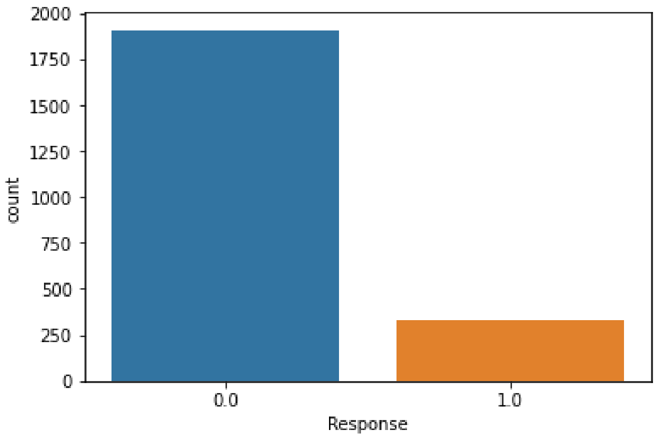

Figure 1. Distribution class

An overview of the methods carried out includes starting from the Marketing Campaign data preparation stage [13]. The next stage is doing data preprocessing which aims to get ready-to-use data for model testing, the next stage is testing the model on preprocessing data to get the best model in predicting whether the customer accept or reject product or service offers. The following is a marketing campaign prediction research method, which can be seen in Figure 2.

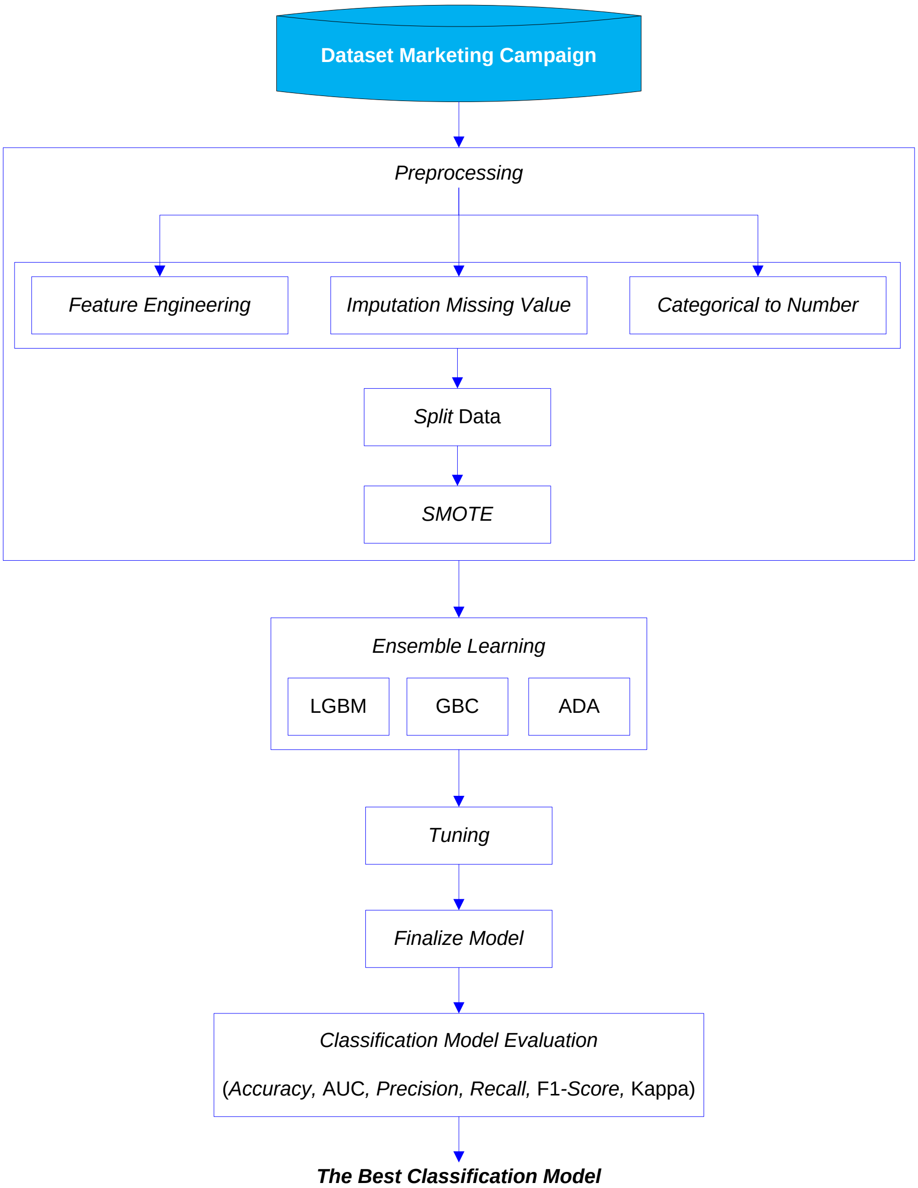

Figure 2. Research methodology

The general description stage of the method is the flow of the experiments carried out. An overview of the methods carried out, among others, starts from the stage of preparing the Marketing Campaign data [13]. The next stage is doing data preprocessing which aims to get ready-to-use data for model testing, the next stage is testing the model on preprocessing data to get the best model in predicting whether customers accept or reject product or service offers.

Data Preprocessing is a technique for preparing data in such cases as cleaning data from noise or changing data formats. Where there are five stages of data preprocessing, namely Categorical to Number which aims to categorize the number coding so that the data can be used by the model, Feature Engineering at this stage removes attributes that are considered not too influential and divides one attribute into 3 new attributes. Inputation Missing Value At this stage, the missing customer data is filled in so that the data can be used in the test model, the next stage is split validation which intends to share training and testing data, training data is used as training data for learning models and testing data is used as data. testing for model validation or evaluation. The last SMOTE oversampling was used to address unbalanced datasets. After the preprocessing stage, the next step is to test the proposed model. The next stage is to evaluate the optimized model. The optimized model is the final model that will be used for marketing campaign predictions.

(ii) Category to Number

This stage is carried out with the aim of categorizing by number coding so that the data can be read by the model. The Education attribute and the Marital_Status attribute have a data type in the form of text, so it is necessary to code them in the form of numbers. The Education attribute has five categories, namely 2n Cycle which is interpreted as 0, Graduation (2), Master (3), and PhD (4). The Marital_Status attribute has seven categories, namely Alone interpreted (1), Divorced (2), Married (3), Single (4), Together (5), Widow (6), and YOLO (7). The following is the display of the data after categorical to number can be seen in Table 3.

Table 3. Category to Number

ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Recency
0	1957	2	4	58138.0	0	0	58
1	1954	2	4	46344.0	1	1	38
2	1965	2	5	71613.0	0	0	26
3	1984	2	5	26646.0	1	0	26
4	1981	4	3	58293.0	1	0	94

(iii) Feature Selection

Feature selection is done by looking at the correlation between features. Correlation Coefficient value always lies between -1 to +1. If correlation coefficient value is positive, then there is a similar and identical relation between the two variables. Else it indicates the dissimilarity between the two variables. We perform analysis for features that have a significant correlation value. Figure 3 shows the pattern of relationships between features marked with colors that have values. Our decision not to use the ID feature because it does not have a strong enough correlation value with other variables. In addition, the ID feature is not taken because it only contains customer numbers which in our opinion have no direct effect on this marketing predictive activity. The following is the feature correlation can be seen in Figure 3.

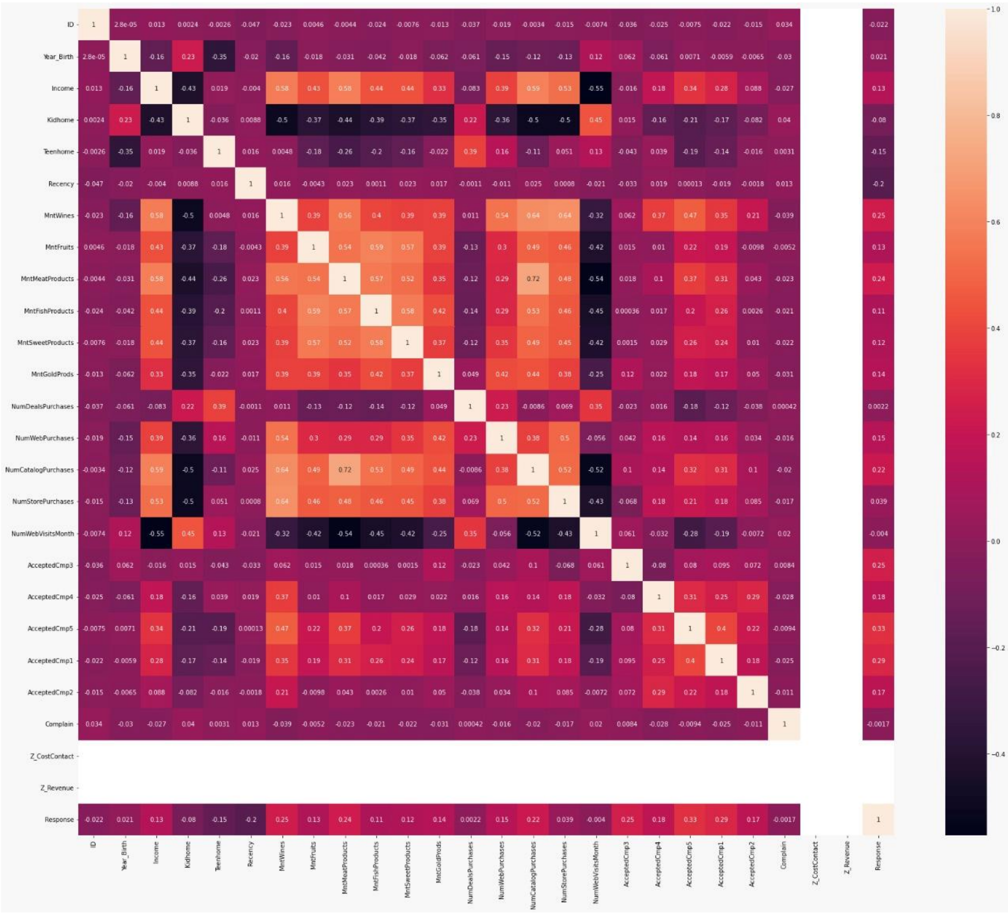

Figure 3. Feature Correlation

At this stage, the ID attribute is deleted which is considered not very influential and divides the D_Customer attribute into 3 new attributes. The D_Customer attribute has two different date formats, namely slash format (/) and strip format (-). The new attributes are Date_Customer, Mont_Customer, Year_Customer. Five data from the three new attributes can be seen in Table 4.

Table 4. Five data of three new attributes

Date_Customer	Mont_Customer	Year_Customer
21	8	2012
17	8	2012
6	2	2013
7	1	2013
13	6	2013

(iv) Imputation Missing Value

Imputation Missing Value used to fill in missing values in the data. In this dataset, there are 24 instances of the missing value in the Income attribute. At this stage, the KNN Imputer algorithm is used to find the nearest neighbor value of k=5. Attribute Income = 0 means that there is no missing value on the attribute. The following is the display of the dataset attributes after inputting the missing data using the KNN Imputer algorithm can be seen in Figure 3.

```
Imputing row 1/2240 with 0 missing, elapsed time: 1.277
Imputing row 101/2240 with 0 missing, elapsed time: 1.279
Imputing row 201/2240 with 0 missing, elapsed time: 1.280
Imputing row 301/2240 with 0 missing, elapsed time: 1.280
Imputing row 401/2240 with 0 missing, elapsed time: 1.281
Imputing row 501/2240 with 0 missing, elapsed time: 1.282
Imputing row 601/2240 with 0 missing, elapsed time: 1.282
Imputing row 701/2240 with 0 missing, elapsed time: 1.282
Imputing row 801/2240 with 0 missing, elapsed time: 1.283
Imputing row 901/2240 with 0 missing, elapsed time: 1.283
Imputing row 1001/2240 with 0 missing, elapsed time: 1.285
Imputing row 1101/2240 with 0 missing, elapsed time: 1.285
Imputing row 1201/2240 with 0 missing, elapsed time: 1.285
Imputing row 1301/2240 with 0 missing, elapsed time: 1.286
Imputing row 1401/2240 with 0 missing, elapsed time: 1.289
Imputing row 1501/2240 with 0 missing, elapsed time: 1.289
Imputing row 1601/2240 with 0 missing, elapsed time: 1.289
Imputing row 1701/2240 with 0 missing, elapsed time: 1.290
Imputing row 1801/2240 with 0 missing, elapsed time: 1.290
Imputing row 1901/2240 with 0 missing, elapsed time: 1.290
Imputing row 2001/2240 with 0 missing, elapsed time: 1.291
Imputing row 2101/2240 with 0 missing, elapsed time: 1.294
Imputing row 2201/2240 with 0 missing, elapsed time: 1.294
```

Figure 4. Imputation missing value

(v) Split Dataset

Split validation is done by sharing training and testing data, training data is used as training data for model learning, testing data is used as test data for model validation or evaluation. In this study, the dataset was split using a split percentage of 80:20. Split train data and test data can be seen in Figure 4.

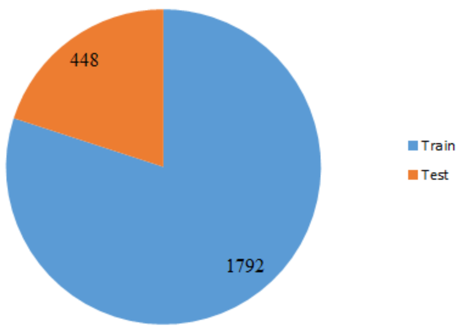

Figure 5. Split train data and test data

(vi) Oversampling SMOTE

After seeing the statistical analysis of the data on the unbalanced class distribution, then conducting experiments by adding a sampling technique with the SMOTE oversampling technique method. The number of classes after SMOTE oversampling is balanced, namely customers who reject (0) marketing campaigns are 1538 and those who receive (1) marketing campaigns are 1538. The following are the results of the sampling technique with SMOTE oversampling can be seen in Figure 5.

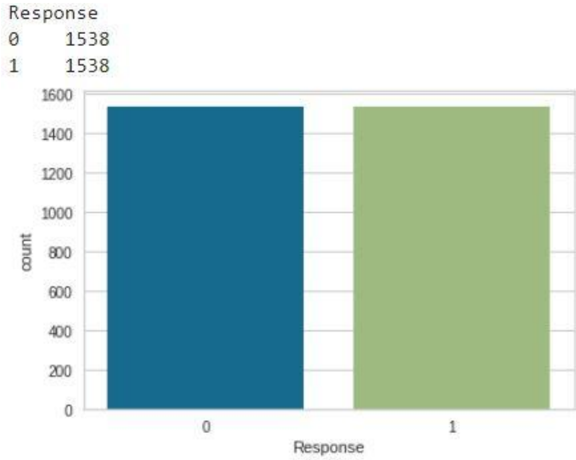

Figure 6. Oversampling SMOTE

RESULTS AND DISCUSSION

Experiments in this study using the Python3 programming language. Testing the model using the PyCaret package is done with the creat_model function, then optimization of the hyperparameter tuning is done with the tune_model function. After the tune_model is done, the next step is to finalize the model or complete the model with the finalize_model function. The model that has been tested is then evaluated using the confusion matrix, ROC curve. Confution matrix is used to determine the value of Accuracy, AUC/ROC, Precision, Recall, FI-score, and Kappa. The confusion matrix contains True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values.

$$\text{Accuracy}= (TP+TN)/(TP+TN+FP+FN) \quad (1)$$
$$\text{Precision}= TP/(TP+FP) \quad (2)$$
$$\text{Recall}= TP/(TP+FN) \quad (3)$$
$$\text{F1-score}= 2x(\text{Precision x Recall})(\text{Precision} + \text{Recall}) \quad (4)$$

Based on the results of tests and evaluations that have been described in the previous stage. Based on Table 4, it can be seen that the best model evaluation results were obtained on the LightGBM model with an accuracy value of 0.9866, AUC 0.9994, Recall 0.9577, Precision 0.9577, F1-score 0.9498, and Kappa 94.98%. Judging from the evaluation results on the LightGBM model, the FP and FN values are equivalent, so that the accuracy value is used as the main reference for performance metrics. Judging from the AUC value, it can be interpreted that the LightGBM model is included in the excellent classification. The summary results of the evaluation of the ensemble model, namely LightGBM, GBC, and ADA for classification in marketing campaign predictions, can be seen in Table 5.

Table 5. Model Ensemble Evaluation Summary

Ensemble Model	Accuracy	AUC	Recall	Precision	F1-Score	Kappa
LightGBM	0.9866	0.9994	0.9577	0.9577	0.9577	0.9498
GBC	0.8996	0.9530	0.4789	0.8095	0.6018	0.5486
ADA	0.8862	0.9322	0.3803	0.7941	0.5143	0.4587

EXPERIMENTAL RESULTS

Based on the evaluation results of model testing, it is known that the LightGBM model is the best classification model for marketing campaign predictions. In a previous study using the same dataset [11], the best model for predicting marketing campaigns was the LR model with an accuracy value of 0.8804. When compared with the results of this study, the LightGBM algorithm model has a better performance than the results of previous studies using the LR model. This stage is carried out to see the novelty or novelty of the experimental results from previous studies using the same dataset. The following is a comparison of the performance of the best model with previous research and this research can be seen in Table 6.

Table 6. Best Model Comparison

Authors	Model	Accuracy	AUC	Recall	Precision	F1-Score
David Zafirovski (2021)	LR	0.8804	0.8887	0.5302	0.6214	0.5656
Proposed Model	LightGBM	0.9866	0.9994	0.9577	0.9577	0.9577

CONCLUSION

One of the reasons why marketing efforts for businesses fail is that clients are not categorised and precisely targeted because all customers are targeted by advertising, including those who will not respond favourably to a marketing campaign and reject the offer. This may lead to higher expenses and lower business earnings. The goal of this research project is to conduct experiments utilising Ensemble Learning and tuning on the Marketing Campaign dataset by offering a classification method. Based on the challenges above, we require the best and most appropriate classification model to predict customers by preprocessing the customers' data needed for predictive marketing campaign analysis. Ensemble Learning has been utilised in the Marketing Campaign Dataset to provide the LightGBM, GBC, and ADA classifications, which can be used to predict the outcomes of marketing campaigns. This research has revealed something novel. Prior research related to marketing campaign datasets did not employ the LightGBM, GBC, or ADA classification algorithms. When the LightGBM classification method is combined with an ensemble learning model, the accuracy value is increased. Thus, the combination is able to increase the precision in predicting whether clients will accept or reject marketing campaign messages. Therefore, the result of this research contributes towards solving the prediction problems in the field of marketing campaigns.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their comments which helped in improving the article.

REFERENCES

[1] V. Eitle and P. Buxmann, "Business Analytics for Sales Pipeline Management in the Software Industry : A Machine Learning Perspective," vol. 6, pp. 1013–1022, 2019.

[2] Z. Zhao and M. Wang, "Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform".

[3] D. Zafirovski, "Analyse der Auswirkungen künstlicher Intelligenz im digitalen Marketing auf das personalisierte Kundenerlebnis," 2021.

[4] J. Asare-Frempong and M. Jayabalan, "Predicting customer response to bank direct telemarketing campaign," *2017 Int. Conf. Eng. Technol. Technopreneurship, ICE2T 2017*, vol. 2017-Janua, no. September, pp. 1–4, 2017, doi: 10.1109/ICE2T.2017.8215961.

[5] S. P. Singh, P. Singh, and A. Mishra, "Predicting Potential Applicants for any Private College using LightGBM," *2020 Int. Conf. Innov. Trends Inf. Technol. ICITIIT 2020*, 2020, doi: 10.1109/ICITIIT49094.2020.9071525.

[6] S. Bhanu Koduri, L. Gunisetti, C. Raja Ramesh, K. V. Mutyalu, and D. Ganesh, "Prediction of crop production using adaboost regression method," *J. Phys. Conf. Ser.*, vol. 1228, no. 1, 2019, doi: 10.1088/1742-6596/1228/1/012005.

[7] H. Rao *et al.*, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput. J.*, vol. 74, pp. 634–642, 2019, doi: 10.1016/j.asoc.2018.10.036.

[8] K. Nugroho *et al.*, "Improving random forest method to detect hatespeech and offensive word," *2019 Int. Conf. Inf. Commun. Technol. ICOIACT 2019*, pp. 514–518, 2019, doi: 10.1109/ICOIACT46704.2019.8938451.

[9] X. Feng, "Research of sentiment analysis based on adaboost algorithm," *Proc. - 2019 Int. Conf. Mach. Learn. Big Data Bus. Intell. MLBDBI 2019*, pp. 279–282, 2019, doi: 10.1109/MLBDBI48998.2019.00062.

[10] M. E. Tocco, "Trabajo de Fin de Máster TÍTULO : ' Análisis de datos para segmentar a los clientes y aumentar la eficiencia de las campañas de Marketing ' Alumno : Maria Emilia Tocco Tutor : Ramón Alberto Carrasco y Rocío González Martínez," 2021.

[11] N. Ravi, A. Scaglione, and S. Peisert, "Colored Noise Mechanism for Differentially Private Clustering," pp. 3–7, 2021, [Online]. Available: http://arxiv.org/abs/2111.07850

[12] S. Lahmiri, S. Bekiros, A. Giakoumelou, and F. Bezzina, "Performance assessment of ensemble learning systems

in financial data classification," *Intell. Syst. Accounting, Financ. Manag.*, vol. 27, no. 1, pp. 3–9, 2020, doi: 10.1002/isaf.1460.

[13] R. Saldanha, "Marketing Campaign." 2014. [Online]. Available: https://www.kaggle.com/rodsaldanha/arketing-campaign