TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 23, No. 4, August 2025, pp. 1020~1031
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v23i4.26341

Grid search vs Bayesian optimization for intensity scoring classification and channel recommendation prediction

Kelly Mae, Dinar Ajeng Kristiyanti
Department of Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia

Article Info

Article history:

Received May 7, 2024
Revised Mar 26, 2025
Accepted May 10, 2025

Keywords:

Bayesian optimization
Channel recommendation
Collection intensity scoring
Grid search
K-nearest neighbors
Random forest

ABSTRACT

Technological advancement has spurred financial technology growth, transforming traditional financial operations into digital. Peer-to-peer (P2P) lending is a key fintech solution offering online loans, though it struggles with repayment issues due to customer financial instability. To overcome these challenges, XYZ is a startup that focuses on enhancing the efficiency of collections and communication with customers. XYZ necessitates the implementation of a collection intensity scoring (CIS) model and a prediction model for interaction on recommended communication channels in order to optimize the collection process. This study evaluates the performance of grid search and Bayesian optimization on random forest (RF) classification models and K-nearest neighbors (KNN) regressor prediction models. RF and KNN regressor algorithms optimization are necessary to enhance their performance in CIS classification and channel recommendation prediction. The research stages follow the cross industry standard process-data mining (CRISP-DM) framework, which consists of business understanding, data understanding, data preparation, modeling, and evaluation. The model performance is assessed by accuracy and mean absolute error (MAE). The results of this study show that Bayesian optimization surpasses grid search, enhancing the accuracy of the RF model to 98.34% and reducing the MAE of the KNN regressor model to 0.24530.

Corresponding Author:

Dinar Ajeng Kristiyanti
Department of Information Systems, Faculty of Engineering and Informatics
Universitas Multimedia Nusantara
St. Scientia Boulevard, Gading Serpong, Kec. Kelapa Dua, Tangerang, Banten 15810, Indonesia
Email: dinar.kristiyanti@umn.ac.id

1. INTRODUCTION

Technological advances, including the rise of Industry 4.0, have transformed various industries, particularly through the integration of artificial intelligence (AI). The cognitive capabilities of AI are driving profound innovations in business, particularly in the development of new products and services [1]. The financial sector has seen significant changes with the emergence of financial technology (fintech) companies. Among them, peer-to-peer (P2P) lending companies provide online loans to meet the daily needs of customers. According to data from the Financial Services Authority for the period of September 2023, the number of fintech P2P lending users in Indonesia has reached 121.955 million, and the number of P2P lending fintech companies in Indonesia under the supervision of the Financial Services Authority has reached 101 companies [2]. However, P2P lending companies are facing challenges in loan repayment, which has prompted efforts to remind customers of due dates. These challenges are further compounded by the changing behavior of customers in the current economic environment, which places greater demands on resources, especially

customer support. To address these issues, XYZ's services are being utilized to enhance productivity in the collection and customer communication process.

XYZ is a technology startup that provides digital solutions to enhance customer relationships. The company employs a collection intensity scoring (CIS) system, analogous to credit scoring, to assess the intensity of customer collections. To optimize the collection process, a model is developed to classify collection intensity using customer, loan, and communication channel interaction data. The customer collection intensity is classified into three categories, namely low, medium, or high. However, applying this broadly across all channels without knowledge of the customer's interaction behavior can be inefficient and costly. Consequently, a predictive model is required to forecast the interactions that customers will respond to in their preferred channels. This enables XYZ to effectively target customers with tailored channel recommendations and predict the number of interactions.

The issues described above indicate a need for a data mining solution that will enable the implementation of a model to fulfill the classification needs of CIS and provide recommendations regarding channels to PT. XYZ. Conventional machine learning models are inadequate because they necessitate hyperparameter optimization. This is a process of identifying the most optimal hyperparameters for a machine learning model. Therefore, the utilization of hyperparameter optimization may result in an enhanced model with enhanced performance.

The machine learning algorithm utilized in this research is random forest (RF), applied to classify CIS in the context of credit scoring. Trivedi [3] achieved high accuracy with RF on German credit data using feature selection techniques, such as Chi-square (93.12%), Gain-Ratio (91.20%), and Info-Gain (90.90%). Furthermore, Xu *et al.* [4] demonstrated that RF achieved an accuracy of 98.4% for borrower default risk on a P2P platform in China, and Li and Chen [5] reported 81.05% on a lending club consumer credit loans dataset in Q4 2018. Additionally, the research employs the K-nearest neighbors (KNN) regressor for channel recommendations and customer interaction prediction. Patro *et al.* [6] introduced a hybrid action-related K-nearest neighbors (HAR-KNN) method for product recommendations based on Amazon data. This incorporated the KNN algorithm and achieved a mean absolute error (MAE) of 0.7165 with 10 neighbors. Jena *et al.* [7] demonstrated that KNN produced the best performance with MAEs of 0.0130 in e-Learning course recommendation. This research also incorporates hyperparameter optimization, specifically grid search and Bayesian optimization, to refine model accuracy. Previously, Li and Chen [5] applied grid search to an RF model for credit loans, achieving an accuracy of 81.05%. Shams *et al.* [8] applied it to a KNN regressor model for water quality in India, achieving a MAE of 0.0009. Furthermore, Wang *et al.* [9] optimized an RF model with Bayesian optimization, achieving an accuracy of 99.30% in cyber-physical systems, while Lee *et al.* [10] employed the same method to optimize a KNN model, predicting rainfall erosivity in Italy and Switzerland with a MAE of 20.576 MJ mm ha-1 h-1.

These research contributions include: (i) exploring the application of CIS classification and channel recommendation prediction, (ii) using customer data from a P2P lending company and communication channels provided by XYZ for the period July 2023 until December 2023, (iii) implementing classification and prediction models using the RF algorithm for CIS and KNN regressor for channel recommendation, (iv) optimizing the RF and KNN regressor models through hyperparameter optimization using grid search and Bayesian optimization, and (v) comparing the RF and KNN regressor models after grid search and Bayesian optimization implementations to determine the best optimization type for CIS and channel recommendation based on accuracy and MAE metrics.

2. METHOD

This research adheres to the stages delineated in the cross industry standard process-data mining (CRISP-DM) framework. The stages used include business understanding, data understanding, data preparation, modeling, and evaluation. However, the deployment stage is not employed, as this study is solely concerned with the comparison of optimization methods for the developed models. Furthermore, this research employs the extract, transform, load (ETL) process for the data preparation phase within the CRISP-DM framework. The workflow of the CIS and channel recommendation research is depicted in Figure 1.

2.1. Business understanding

In the current phase of business understanding, the objectives of XYZ include the identification of customer intensity categories in the CIS system and the provision of appropriate treatment based on the recommended communication channels. Additionally, the collection of unpaid loan payments from P2P lending customers is a key objective. XYZ has had the CIS system since 2022 to assess customer billing intensity in P2P lending companies. However, there are issues with inaccuracies in the classification of billing intensity for customers. Additionally, the use of communication channels that are rarely used by customers can lead to increased business communication costs. Therefore, the classification of customer billing intensity and

the prediction of customer interaction numbers are needed to generate insights for the implementation and evaluation of billing treatment types for customers through the appropriate communication channels.

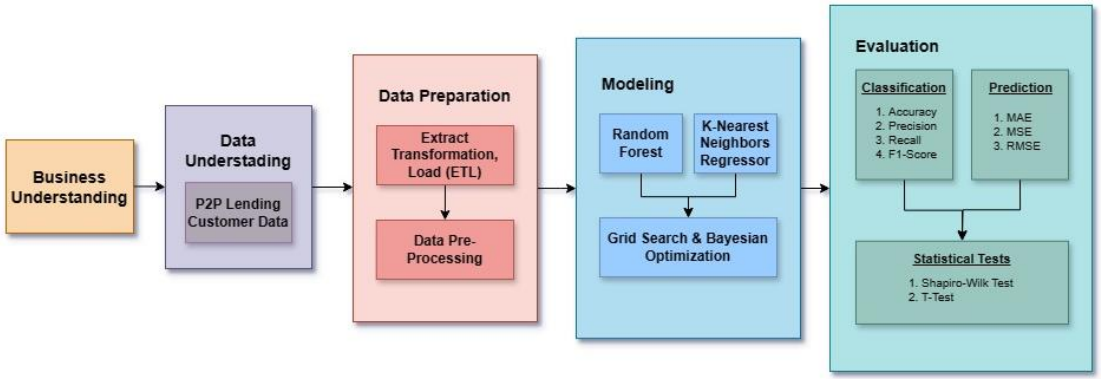

Figure 1. Research workflow

2.2. Data understanding

The data utilized in this study is comprised of secondary data from XYZ, which was provided in the form of tables spanning the period from July 2023 until December 2023. It predominantly encompasses detailed information regarding customer personal and loan information within a P2P lending company that uses the XYZ's service. In addition, the data includes the number of interactions conducted by customer service with customers through each communication channel, such as email, robocall, SMS, and telephony. The data is processed using an ETL process, which produces the final table of P2P lending customer data for data pre-processing. This process results in data consisting of thirteen features, as shown in Table 1.

Table 1. P2P lending customer data

User_id	Gender	Age	Email_count	Robo call_count	SMS_count	Tele phoy_count	Total_inter action	Last_interaction_type	Total_loan	Paid_loans	Avg_interacted_to_paid	Total_paid_amount
1	Male	26	4	0	1	0	5	SMS	1	2	0.0	1256000
2	Male	37	29	22	2	0	53	Robocall	3	7	0.0	4043000
3	Male	27	12	2	3	0	17	Email	1	1	0.0	623000

2.3. Data preparation

The data preparation stage represents the initial phase of the data processing pipeline, preceding the model development phase. This stage encompasses two primary processes, such as ETL and data pre-processing. The following sections provide an in-depth explanation of each data preparation process utilized in this research. These steps are essential to ensure that the input data is clean, structured, and suitable for effective model training and evaluation.

2.3.1. Extract, transformation, load

The ETL process is a sequence of structured operations, comprising the extract, transformation, and load phases, which are designed to facilitate the systematic processing of source data in order to transform it suitable for the intended use [11]. The extract step represents the initial stage of the ETL process, during which tables provided by XYZ in the form of comma-separated values (CSV) files are imported into the database. Transformation entails the combination of data from various sources, which generates multiple tables according to their respective data types. Data combination involves generating of unique identifiers such as user_id, which is incorporated into loan and interaction tables alongside data de-duplication. The subsequent

process is data filtering based on specified conditions, followed by data aggregation to generate a final table. This final table is then exported in a CSV format, which is loaded for the purpose of data pre-processing.

2.3.2. Data pre-processing

Data pre-processing involves the cleansing of data that has been loaded from the ETL process. This may entail the identification of missing data, the conversion of data types, and the standardization of data. Additionally, there is data pre-processing for the classification approach, which encompasses feature engineering to create intensity category as target variable and data encoding. Subsequent to that is data pre-processing for the prediction approach is conducted, which comprises data cleansing, data transformation, feature engineering to achieve recommended channel for each customer, data integration, data encoding, and data standardization. The pre-processed data has been divided into two parts, with 80% allocated for the training process, and 20% reserved for the testing process. This research employs the synthetic minority oversampling technique (SMOTE) after data splitting. SMOTE represents a technique for transforming imbalanced data into a balanced data set [4]. It employs oversampling to generalize the data distribution [12]. In the classification approach, SMOTE is applied to the training data in order to address the issue of imbalance within the dataset.

2.4. Modeling

Modeling phase is conducted using two distinct approaches. The first is a classification approach, which involves developing a model using the RF algorithm for classifying customer collection intensity to low, medium, and high. In addition, a predictive approach is employed, whereby a model is created using the KNN regressor algorithm to predict the number of customer interactions in recommended channels. This research employs hyperparameter optimization techniques, namely grid search and Bayesian optimization, for both approaches.

2.4.1. Random forest

RF is an ensemble learning approach based on decision trees that was proposed by Breiman in 2001 [13]. RF employs a supervised learning methodology, wherein information from a labeled dataset (training set) is utilized to make predictions and construct a model [14]. In the construction of the model, each decision tree selects a class for observations and predictions based on the class that receives the most votes [15].

RF classifier is a collection of structured tree classifiers, denoted by the set of functions $\{h(x,\Theta_k), k=1,\dots\}$, where $\{\Theta_k\}$ are identically distributed random vectors. Each tree casts one vote for the most popular class for the input x. The aggregation process in RF occurs after the creation of the random vectors. In this process, an ensemble of classifiers, $h_1(x), h_2(x), \dots, h_k(x)$, is created with training sets randomly drawn from the distribution of the random vectors Y, X [15]. A margin function is computed as in (1):

$$mg(\mathrm{X},Y) = av_k I(h_k(X) = Y) - max_{j \neq Y}\, av_k I(h_k(X) = j) \quad (1)$$

where $I(\cdot)$ is the indicator function and av_k is the average, $h_k(X) = Y$ is the result of the classification, and $h_k(X) = j$ is the classification outcome with j [16]. The margin function measures the extent to which the average votes in X, Y for the correct class exceed the average votes for other classes. The larger the margin obtained, the greater the confidence in the classification. The process of calculating generalization is performed after obtaining the margin function, where $P_{X,Y}$ indicates that the probability is above the space X, Y, and $mg(\mathrm{X},Y)$ is the margin function [15]. Generalization error is calculated as in (2):

$$PE^* = P_{X,Y}(mg(\mathrm{X},Y) < 0) \quad (2)$$

2.4.2. K-nearest neighbors

KNN is a non-parametric, supervised machine learning algorithm utilized for classification and regression [17]. KNN was initially introduced by Evelyn Fix and Joseph Hodges in 1951 as a method of supervised learning and was subsequently developed by Thomas Cover [18]. KNN algorithm operates by measuring the distance between the input data and the k closest data points in the training set [19]. This research focuses on the use of the KNN regressor to predict the number of interactions on a recommended channel using Cosine distance. Cosine distance is a commonly used measure in machine learning algorithms to calculate the distance or similarity between two data points [20]. Initially, the cosine similarity is calculated to measure the correlation between two vectors. Cosine similarity is calculated as in (3) [21].

$$\cos(x,\, y) = \frac{x \cdot y}{\|x\| \cdot \|y\|} \quad (3)$$

where (x, y) denotes the angle between two vectors. The symbol $x \cdot y$ denotes the dot product of the vectors *x* and *y*, while $\|x\| \cdot \|y\|$ denotes the length of the vectors *x* and *y*. The cosine distance is calculated to produce values between 0 and 1. A value of 0 indicates that the distance obtained is close, thus indicating a high degree of similarity. Conversely, a value of 1 indicates that the distance obtained is far, thereby indicating a low degree of similarity. Cosine distance is performed as in (4) [7]:

$$cosine\ distance\ =\ 1\ -\ cosine\ similarity \tag{4}$$

2.4.3. Grid search

Grid search is a distinctive method of hyperparameter optimization that involves a comprehensive search of a predefined search space [5]. Grid Search aims to exhaustively search all possible combinations of hyperparameters within a given range or set of values. This is done by first creating a grid of all possible combinations of hyperparameters, and then training and testing the model on a validation or cross-validation for each combination [8].

2.4.4. Bayesian optimization

Bayesian optimization is a method used to find the extrema of black-box functions [22]. It is useful for finding the optimal hyperparameters for a machine learning model. Bayesian optimization works well when the classification dataset is non-linear, complex, and noisy because the computation to identify costly hyperparameters can significantly impact model performance [23]. The Bayesian optimization process involves two main steps, such as estimating the black-box function from the data by a probabilistic surrogate model using Gaussian processes, known as the response surface, and maximizing an acquisition function to balance the exploration-exploitation tradeoff based on the uncertainty and optimality of the response surface [24]. Bayesian optimization incorporates prior information about the function f and updates posterior information that helps reduce loss and maximize model accuracy [23]. Gaussian process from Bayesian optimization is computed as in (5) [25]:

$$f(x) \sim GP\big(m(x), k(x, x')\big) \tag{5}$$

Gaussian process equation assumes a prior distribution as a multivariate normal distribution with a specific mean vector and covariance matrix. This equation includes *x* as a point in the input space, $m(x)$ as the mean function for the mean vector, and $k(x, x')$ as the covariance function or kernel for each pair of points *x* and x' for the covariance matrix. This kernel is chosen so that the points *x* and x' have a high positive correlation, indicating that they should have more similar function values than points further apart. Next, the posterior probability function is computed as a conditional distribution to evaluate the value of $f(x)$ and predict it at some new points x_*. The calculation of this posterior probability distribution includes the posterior mean and the posterior variance, which are computed as in (6)-(8) [25]:

$$f(x_*) \mid f(X) \sim GP\left(\mu_{post}(x_*), \sigma^2_{post}(x_*)\right) \tag{6}$$

$$\mu_{post}(x_*)\ =\ k(x_*, X)[k(X, X)\ +\ \sigma_n^2\, I]^{-1} Y \tag{7}$$

$$\sigma^2_{post}(x_*) =\ k(x_*, x_*)\ -\ k(x_*, X)[k(X, X)\ +\ \sigma_n^2\, I]^{-1} k(X, x_*) \tag{8}$$

As shown in (6) is the posterior probability distribution equation, where $\mu_{post}(x_*)$ is the posterior mean and $\sigma^2_{post}(x_*)$ is the posterior variance. The posterior mean in (7) is a weighted average between the prior mean $m(x)$ and the prediction based on the observed data $f(X)$, where the weights depend on the kernel. This equation includes $k(x_*, X)$ as the covariance between the prediction point and the observed points, $k(X, X)$ as the covariance between the observed points, and $\sigma_n^2 I$ as the noise variance adjusted with the covariance $k(X, X)$ and *Y* as the observed target values. The posterior variance in (8) is computed by taking the prior covariance $k(x, x')$ and adjusting it by subtracting the variance from the observed data $f(X)$.

The application of the acquisition function is then performed using the expected improvement to capture the expected value of the improvement and select *x* to maximize it [25]. The results of this application are used to update the Gaussian process statistical model [23]. Expected improvement is performed as in (9) [25].

$$EI(x)\ =\ \mathbb{E}\left[max(f(x)\ -\ f(x^+), 0)\right] \tag{9}$$

where $\mathbb{E}$ is the expected value taken from the posterior distribution of the score f, and x^+ is the point with the best observed value [25]. $EI(x)$ is positive if the predicted value is higher than the best observed value. In addition, $EI(x)$ is set to zero [9].

2.5. Evaluation metrics

Evaluation metrics include the use of different types of metrics in the classification approach and the prediction approach. The classification approach involves the use of accuracy, precision, recall, and f1-score metrics. Meanwhile, the prediction approach consists of MAE, mean square error (MSE), and root mean square error (RMSE). The following is will be used to explain the evaluation metrics for each approach used in this study.

2.5.1. Accuracy

Accuracy is the proportion of all data that is accurately classified. The higher the accuracy value obtained from a model, the better the performance of the model. Accuracy can be obtained in (10):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative [3].

2.5.2. Precision

Precision indicates how many samples are predicted to be positive or true positive. Precision consists of TP and FP. Precision can be obtained in (11) [4].

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

2.5.3. Recall

Recall, or sensitivity in binary classification, indicates how many positive values in the sample are correctly predicted. Recall is composed of TP and FN. Recall can be obtained from (12) as follows [4].

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

2.5.4. F1-score

F1-score is the harmonic mean of precision and recall [4]. In other words, f1-score is an effective way to measure the accuracy of a classification method by using the mean of precision and recall in its calculation [3]. F1-score can be calculated as in (13) [4].

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (13)$$

2.5.5. Mean absolute error

MAE is used to evaluate the measure between actual and predicted values. It indicates the difference between the target values and the predicted values [26]. MAE ranges from zero to infinity, with lower values defining better accuracy and infinity representing the maximum error in prediction rankings [6]. MAE can be computed as in (14).

$$MAE = \frac{\sum_{i=1}^{n} |X_i - \hat{X}_i}{n} \quad (14)$$

where $\hat{X}_i$ is predicted value, X_i is actual values, and n is the number of observations [26].

2.5.6. Mean square error

MSE is the square of the difference between the actual value and the evaluated value. Squaring is done to eliminate negative values. MSE can be calculated as in (15) [26].

$$MSE = \frac{\sum_{i=1}^{n} (X_i - \hat{X}_i)^2}{n} \quad (15)$$

2.5.7. Root mean square error

RMSE detects the level of error in a regression model and examines the size of the error relative to the size of the target values [26]. A lower RMSE value defines a better prediction accuracy of the results of the recommendation system (RS) [6]. RMSE is calculated as in (16) [26].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}} \tag{16}$$

2.6. Statistical tests

Statistical tests are used to analyze data that require assumptions for the results to be valid [27]. The primary purpose of statistical tests is to assess the validity of a claim or hypothesis about a model or parameter value based on the observed data [28]. In this study, statistical tests are divided into two types, such as Shapiro-Wilk test and T-Test.

2.6.1. Shapiro-wilk test

Shapiro-Wilk test is one of the most important methods for testing the normality assumption [29]. Shapiro-Wilk test is used to determine whether a sample from a population comes from a normally distributed population [30]. Shapiro-Wilk test is performed as in (17).

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{17}$$

where $x_{(i)}$ are the sample values ordered from the smallest to largest, a_i is constant obtained from the mean, variance, and covariance of the sequence statistic of a sample size n from a normal distribution, $\bar{x}$ is the sample mean, and n is the sample size [31].

2.6.2. T-test

T-test is a statistical test used to compare the means of two groups [32]. In this study, a paired t-test is used to demonstrate the presence of a significant difference in performance between the compared models based on accuracy metrics and MAE. The paired t-test is useful when analyzing the same data set measured under two different conditions, the measurement differences on the same subject before and after treatment, or the differences between two treatments given to the same subject [33]. The calculation of the paired t-test is performed using (18).

$$d_i = y_i - x_{\mathrm{i}} \tag{18}$$

As shown in (18) shows the paired difference equation, or the difference between the corresponding observations from two samples, denoted by y_i and x_{i}, where i represents one element in each pair of observations. The corresponding paired differences are treated as a variable, with the logic of the paired t-test and the one-sample t-test being identical. This equation is then included in the calculation of the mean of the difference and the standard deviation of the difference, as shown by (19) and (20) respectively [34].

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{19}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}} \tag{20}$$

Based on the calculations using (19) and (20), $\bar{d}$ is denoted as the mean, s_d as the standard deviation, and n as the sample size. The results of the calculations for the mean and standard deviation of the differences are used for the t-statistic, which involves dividing the mean of the difference ($\bar{d}$) by the standard error of the mean of the difference, denoted as $s_d\sqrt{n}$. The calculation of the t-statistic is shown in (21) [34].

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{21}$$

3. RESULTS AND DISCUSSION

The result stage presents the performance assessments of the conventional RF and KNN regressor models, as well as the models that have undergone the application of grid search and Bayesian optimization. The performance comparison of both models is based on specific evaluation metrics for each approach, with the objective of determining the best model. In order to ensure that the performance comparison of the models is conducted fairly, statistical tests such as the Shapiro-Wilk test and t-test are applied. This allows for the demonstration that there is a significant performance difference between the best model and the compared model.

3.1. Results of collection intensity scoring classification using random forest

A comprehensive performance evaluation was conducted for all RF models, employing a set of predefined evaluation metrics, including accuracy, precision, recall, and F1-score. Additionally, an additional metric was utilized to assess the training time, expressed in seconds. As demonstrated in Table 2, the RF x Bayesian optimization model exhibited the most optimal performance, with an accuracy of 98.34%, precision of 98.54%, recall of 98.34%, f1-score of 98.40%, and a training time of 6.690 seconds. The accuracy values for each RF model type are presented in the accuracy comparison plot, which is shown in Figure 2.

Table 2. Results of CIS classification using RF

Model	Accuracy	Precision	Recall	F1-score	Time (s)
RF	98.15	98.42	98.15	98.25	1.611
RF x grid search	98.21	98.42	98.21	98.28	7.718
RF x Bayesian optimization	98.34	98.54	98.34	98.40	6.690

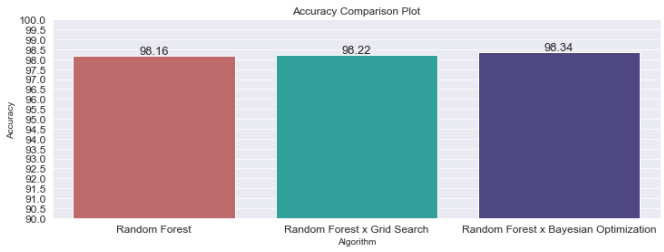

Figure 2. Accuracy comparison plot

3.2. Results of channel recommendation prediction using k-nearest neighbors regressor

The performance of all KNN regressor models was evaluated using a set of evaluation metrics, including MAE, MSE, and RMSE. In addition to these metrics, the training time in seconds was also considered. The results presented in Table 3 indicate that the KNN regressor x Bayesian optimization model exhibits the most optimal performance, with an MAE of 0.24530, an MSE of 0.12245, an RMSE of 0.34993, and a training time of 1.796 seconds. Furthermore, the MAE values for all KNN regressor models can be observed in the MAE comparison plot displayed in Figure 3.

Table 3. Results of channel recommendation prediction using KNN regressor

Model	MAE	MSE	RMSE	Time (s)
KNN regressor	0.25005	0.12394	0.35205	0.010967
KNN regressor x grid search	0.24754	0.12304	0.35077	9.737765
KNN regressor x Bayesian optimization	0.24530	0.12245	0.34993	1.796227

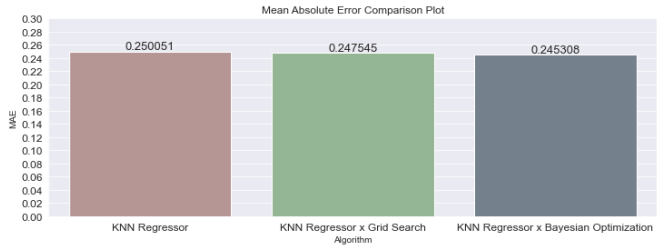

Figure 3. Mean absolute error comparison plot

3.3. Results of shapiro-wilk test

Shapiro-Wilk Test is employed in this study to assess the normality of the performance metric distribution of each model, with a focus on the accuracy of the RF model and the MAE of the KNN regressor model. The null hypothesis (H0) is used in the Shapiro-Wilk Test, which states that the data are normally distributed. In contrast, the alternative hypothesis (H1) posits that the data are not normally distributed. The interpretation of the test results is based on the p-value. If the p-value is greater than the significance level α of 0.05, then H0 is accepted. Conversely, if the p-value is less than α, then the H0 is rejected and the H1 is accepted. The results presented in the Table 4, indicate that the p-values of all RF and KNN regressor models are greater than the α value. This implies that there is insufficient evidence to reject the null hypothesis concerning the normality of the accuracy metric distribution of all RF models and the MAE of all KNN regressor models. The test results are in accordance with the hypothesis that the accuracy metrics of all RF models and the MAE of all KNN regressor models follow a normal distribution.

Table 4. Results of the shapiro-wilk test based on the accuracy metric from the rf model and the mae metric from the knn regressor model

Classification			Prediction		
Model	W	p-value	Model	W	p-value
RF	0.94345	0.6904	KNN regressor	0.96124	0.8166
RF x grid search	0.90731	0.4516	KNN regressor x grid search	0.88427	0.3291
RF x Bayesian optimization	0.96681	0.8544	KNN regressor x Bayesian optimization	0.97716	0.9189

3.4. Results of t-test

T-test in this study is conducted in paired manner to determine whether there was a significant difference in performance between the best model and other models. The two best models to be compared were the RF x Bayesian optimization and KNN regressor x Bayesian optimization, with the metrics of accuracy and MAE, respectively. The H0 is that there is no significant difference in performance between the best model and other models. The H1 is that there is a significant difference between the best model and other models. The H0 will be rejected if the p-value is less than α of 0.05, indicating that there is sufficient evidence to accept the H1 and declare that there is a significant performance difference. Conversely, if the p-value is greater than α, then there is insufficient evidence to reject the H0. The results of Table 5 indicate that p-values greater than the significance level α were found in the RF x Bayesian optimization and KNN regressor x Bayesian optimization models compared to other models. This suggests that there is insufficient evidence to reject the H0, which states that there are no significant differences in the accuracy metrics of RF x Bayesian optimization and the MAE of KNN regressor x Bayesian optimization compared to other models.

Table 5. Results of t-test based on the accuracy metric from the rf x bayesian optimization and the mae metric from knn regressor x bayesian optimization compared to other models

RF x Bayesian optimization			KNN x Bayesian optimization		
Compared model	t	p-value	Compared model	t	p-value
RF	0.94345	0.6904	KNN regressor	0.96124	0.8166
RF x grid search	0.90731	0.4516	KNN regressor x grid search	0.88427	0.3291

3.5. Discussion

This research employs the use of RF and KNN regressor algorithms, along with grid search and Bayesian optimization using customer data from a P2P lending company. This allows for the comparison of the results obtained with those of previous studies, as demonstrated in Table 6. In classification, the proposed RF model achieved 98.16% accuracy in 1.611 seconds, markedly better than the earlier 93.12% in 16.20 seconds [3]. Similarly, another study reached 98.40% accuracy [4], underscoring consistent high performance in the P2P lending sector. The proposed RF x grid search model demonstrated superior performance to previous standards, achieving 98.22% accuracy [5]. Furthermore, the proposed RF x Bayesian optimization model exhibited enhanced performance, with an accuracy of 98.34% in 7.718 seconds, surpassing the former 90.33% in 4.348 seconds [9]. In prediction, the proposed KNN regressor demonstrated an MAE of 0.250051 at k=10, which represents an improvement upon the earlier MAE of 0.7165 [6]. The proposed KNN regressor x grid search achieved an MAE of 0.247545 at k=12, which contrasts sharply with a prior MAE of 0.0009 at k=1 [8]. The proposed KNN regressor x Bayesian optimization demonstrated substantial improvements in channel recommendation prediction with an MAE of 0.245308 at k=11 in 1.796 seconds.

Table 6. Comparison between the related works and the proposed classification and prediction models

Classification			Prediction			
Model	Accuracy	Time (s)	Model	k	MAE	Time (s)
Trivedi [3]	93.12	16.20	Patro *et al*. [6]	10	0.7165	-
Xu *et al.* [4]	98.40	-	Jena *et al*. [7]	-	0.0130	-
Li and Chen [5]	81.05	0.8949	Shams *et al*. [8]	1	0.0009	-
Wang *et al*. [9]	90.33	4.348	Lee *et al*. [10]	10	20.576	264
Proposed RF	98.16	1.611	Proposed KNN regressor	10	0.25005	0.010
Proposed RF x grid search	98.22	7.718	Proposed KNN regressor x grid search	12	0.24754	9.737765
Proposed RF x Bayesian optimization	98.34	6.690	Proposed KNN Regressor x Bayesian optimization	11	0.24530	1.796227

4. CONCLUSION

This study compares the performance of RF and KNN regressor models, both in their conventional forms and after applying hyperparameter optimization with grid search and Bayesian optimization methods. The comparison focuses on two main evaluation metrics, such as accuracy for RF and MAE for KNN regressor. The results of this study conclude that both hyperparameter optimization methods enhance the performance of the conventional models. However, Bayesian optimization consistently yields superior results to grid search, both in improving accuracy for RF and in reducing MAE for KNN. Bayesian optimization application has increased the accuracy of RF to 98.34% and produced a lower MAE for KNN regressor at 0.245308. Therefore, the Bayesian optimization method is the recommended hyperparameter optimization for RF and KNN regressor models in CIS and channel recommendation research.

The findings of this research suggest several recommendations for future studies related to CIS and channel recommendation. Firstly, the incorporation of a greater quantity of data can facilitate the model's capacity to discern more generalized patterns, thereby enhancing its ability to generalize to novel data. Secondly, a more robust data pre-processing approach is required, including feature engineering, feature scaling, and feature selection, in order to reduce noise and complexity in the dataset. Third, an alternative approach to classification could be considered, such as the use of decision trees with pruning techniques or boosting algorithms, such as extreme gradient boosting (XGBoost) or light gradient boosting machine (LightGBM) to reduce the risk of overfitting by using regularization. Additionally, it is recommended that the model be refined by exploring a wider range of hyperparameter values from grid search and Bayesian optimization methods in order to identify the optimal hyperparameter combination for optimal model performance.

ACKNOWLEDGMENTS

We acknowledge the support received from Universitas Multimedia Nusantara for carrying out research work of this paper.

FUNDING INFORMATION

This research was funded by Universitas Multimedia Nusantara.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kelly Mae	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Dinar Ajeng Kristiyanti		✓		✓	✓	✓	✓	✓		✓		✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, [DAK]. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] A. F. S. Borges, F. J. B. Laurindo, M. M. Spínola, R. F. Gonçalves, and C. A. Mattos, "The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions," *International Journal of Information Management*, vol. 57, Apr. 2021, doi: 10.1016/j.ijinfomgt.2020.102225.

[2] "Statistik P2P lending periode September 2023," *Otoritas Jasa Keuangan*, 2023. https://www.ojk.go.id/id/kanal/iknb/data-dan-statistik/fintech/Pages/Statistik-P2P-Lending-Periode-September-2023.aspx (accessed Jan. 07, 2024).

[3] S. K. Trivedi, "A study on credit scoring modeling with different feature selection and machine learning approaches," *Technology in Society*, vol. 63, Nov. 2020, doi: 10.1016/j.techsoc.2020.101413.

[4] J. Xu, Z. Lu, and Y. Xie, "Loan default prediction of Chinese P2P market: a machine learning methodology," *Scientific Reports*, vol. 11, no. 1, Sep. 2021, doi: 10.1038/s41598-021-98361-6.

[5] Y. Li and W. Chen, "A comparative performance assessment of ensemble learning for credit scoring," *Mathematics*, vol. 8, no. 10, Oct. 2020, doi: 10.3390/math8101756.

[6] S. G. K. Patro *et al.*, "A hybrid action-related K-nearest neighbour (HAR-KNN) approach for recommendation systems," *IEEE Access*, vol. 8, pp. 90978–90991, 2020, doi: 10.1109/ACCESS.2020.2994056.

[7] K. K. Jena *et al.*, "E-learning course recommender system using collaborative filtering models," *Electronics*, vol. 12, no. 1, Dec. 2022, doi: 10.3390/electronics12010157.

[8] M. Y. Shams, A. M. Elshewey, E.-S. M. El-kenawy, A. Ibrahim, F. M. Talaat, and Z. Tarek, "Water quality prediction using machine learning models based on grid search method," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 35307–35334, Sep. 2023, doi: 10.1007/s11042-023-16737-4.

[9] T. Wang *et al.*, "Random forest-bayesian optimization for product quality prediction with large-scale dimensions in process industrial cyber-physical systems," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8641–8653, Sep. 2020, doi: 10.1109/JIOT.2020.2992811.

[10] S. Lee *et al.*, "Estimation of rainfall erosivity factor in Italy and Switzerland using Bayesian optimization based machine learning models," *CATENA*, vol. 211, Apr. 2022, doi: 10.1016/j.catena.2021.105957.

[11] R. Galici, L. Ordile, M. Marchesi, A. Pinna, and R. Tonelli, "Applying the ETL process to blockchain data. Prospect and findings," *Information*, vol. 11, no. 4, Apr. 2020, doi: 10.3390/info11040204.

[12] D. A. Kristiyanti, S. A. Sanjaya, V. C. Tjokro, and J. Suhali, "Dealing imbalance dataset problem in sentiment analysis of recession in Indonesia," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 2, pp. 2060–2072, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp2060-2072.

[13] Z. Tang, Z. Mei, W. Liu, and Y. Xia, "Identification of the key factors affecting Chinese carbon intensity and their historical trends using random forest algorithm," *Journal of Geographical Sciences*, vol. 30, no. 5, pp. 743–756, May 2020, doi: 10.1007/s11442-020-1753-4.

[14] A. Segnini and J. J. T. Motchoffo, "Random forests and text mining," *Academia*, pp. 1–5, 2019.

[15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[16] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 815–821, Apr. 2020, doi: 10.12928/telkomnika.v18i2.14785.

[17] G. E. Atteia, H. A. Mengash, and N. A. Samee, "Evaluation of using parametric and non-parametric machine learning algorithms for Covid-19 forecasting," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, 2021, doi: 10.14569/IJACSA.2021.0121071.

[18] I. Dănilă-Cernat, "Nearest neighbor estimates of Kaniadakis entropy," *Analele Universitatii "Ovidius" Constanta - Seria Matematica*, vol. 30, no. 1, pp. 171–189, Feb. 2022, doi: 10.2478/auom-2022-0010.

[19] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi algoritma Naive Bayes, support vector machine, dan K-nearest neighbors untuk analisa sentimen aplikasi Halodoc," *Faktor Exacta*, vol. 14, no. 2, pp. 64–71, Aug. 2021, doi: 10.30998/faktorexacta.v14i2.9697.

[20] S. Zhang, J. Li, and Y. Li, "Reachable distance function for KNN classification," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–15, 2022, doi: 10.1109/TKDE.2022.3185149.

[21] D. Jiang, J. Zhang, Z. Wang, C. Feng, K. Jiao, and R. Xu, "A Prediction model of blast furnace slag viscosity based on principal component analysis and K-nearest neighbor regression," *JOM*, vol. 72, no. 11, pp. 3908–3916, Nov. 2020, doi: 10.1007/s11837-020-04360-9.

[22] P. Mikkola, M. Todorovic, J. Jarvi, P. Rinke, and S. Kaski, "Projective preferential bayesian optimization," in *Proceedings of the 37th International Conference on Machine Learning*, 2020, pp. 6884–6892.

[23] A. H. Victoria and G. Maragatham, "Automatic tuning of hyperparameters using Bayesian optimization," *Evolving Systems*, vol. 12, no. 1, pp. 217–223, Mar. 2021, doi: 10.1007/s12530-020-09345-2.

[24] R. Moriconi, M. P. Deisenroth, and K. S. Sesh Kumar, "High-dimensional Bayesian optimization using low-dimensional feature spaces," *Machine Learning*, vol. 109, no. 9–10, pp. 1925–1943, Sep. 2020, doi: 10.1007/s10994-020-05899-z.

[25] P. I. Frazier, "Bayesian optimization," in *Recent Advances in Optimization and Modeling of Contemporary Problems*, INFORMS,

2018, pp. 255–278. doi: 10.1287/educ.2018.0188.
[26] Z. Shahbazi, D. Hazra, S. Park, and Y. C. Byun, "Toward improving the prediction accuracy of product recommendation system using extreme gradient boosting and encoding approaches," *Symmetry*, vol. 12, no. 9, Sep. 2020, doi: 10.3390/sym12091566.
[27] M. Tsagris and N. Pandis, "Normality test: Is it really necessary?," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 159, no. 4, pp. 548–549, Apr. 2021, doi: 10.1016/j.ajodo.2021.01.003.
[28] W. Q. Meeker, L. A. Escobar, and F. G. Pascual, *Statistical methods for reliability data, 2nd Edition*. Wiley & Sons, 2021.
[29] B. Gerald and T. Frank Patson, "Parametric and nonparametric tests: A brief review," *International Journal of Statistical Distributions and Applications*, vol. 7, no. 3, pp. 78–82, 2021, doi: 10.11648/j.ijsd.20210703.12.
[30] S. Yang and G. Berdine, "Normality tests," *The Southwest Respiratory and Critical Care Chronicles*, vol. 9, no. 37, pp. 87–90, Jan. 2021, doi: 10.12746/swrccc.v9i37.805.
[31] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3–4, pp. 591–611, Dec. 1965, doi: 10.1093/biomet/52.3-4.591.
[32] A. Adhikari, S. C. Gorain, P. Gayen, I. Pal, and S. Sen, "Studying the differences: A review on t-Test," *International Research Journal of Education and Technology*, vol. 5, no. 5, pp. 338–349, 2023.
[33] C. S. De Brito, D. E. W. Silva, L. G. de A. Aguiar, R. Abrãao, and N. A. de S. Sampaio, "Using the paired T-Test to compare suppliers," *Revista de Gestão e Secretariado (Management and Administrative Professional Review)*, vol. 14, no. 10, pp. 18565–18575, Oct. 2023, doi: 10.7769/gesec.v14i10.3068.
[34] H. M. Park, "Comparing group means: T-tests and one-way ANOVA Using Stata, SAS, R, and SPSS," 2009. http://www.indiana.edu/~statmath (accessed Apr. 27, 2024).

BIOGRAPHIES OF AUTHORS

Kelly Mae is currently an undergraduate student in the Information System Study Program at Universitas Multimedia Nusantara. She previously worked as a Growth Data intern at Astra Financial. She has served as a laboratory assistant in probability and statistics in 2022. Her specialization lies in big data, with research interests that include machine learning, data science, and data analysis. She can be contacted at email: kelly.mae@student.umn.ac.id.

Dinar Ajeng Kristiyanti received her Master's degree in Computer Science, and Bachelor degree in Information System from Universitas Nusa Mandiri, Indonesia. Currently, she is a Ph.D. candidate in Computer Science at IPB University, Indonesia. She is a lecturer at Information System Study Program in Universitas Multimedia Nusantara, Indonesia. Her research interests include sentiment analysis, text mining, feature selection, optimization, data science, and machine learning. She has published over 23 papers in international journals and conferences. She can be contacted at email: dinar.kristiyanti@umn.ac.id.