International Journal Software Engineering and Computer Science (IJSECS)
4 (3), 2024, 882-895
Published Online December 2024 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v4i3.2948.

RESEARCH ARTICLE Open Access

Classification of Customer Satisfaction with the K-Nearest Neighbor Algorithm in Relation to Employee Performance at PT. Airkon Pratama

Ahmad Suprianto *
Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Corresponding Email: ahmadsuprianto20@gmail.com.

Untung Surapati
Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: kisuro2003@gmail.com.

Yuma Akbar
Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: yuma.pjj@gmail.com.

Aditya Zakaria Hidayat
Digital business Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.

Received: July 18, 2024; Accepted: October 20, 2024; Published: December 1, 2024.

Abstract: PT. Airkon Pratama is the technical consultancy company in the field of maintenance, repair, and operate system. Among its projects are a four-building, multi-story tax office complex. PT. Airkon Pratama experience obstacles to know how its customer satisfaction with their services that is was measured by a questionnaireobtained from work order form. The purpose of this study is to determine how well K-Nearest Neighbor data classification accurately classifies customer satisfaction based on employee performance by PT. Airkon Pratama. The data used in this study is from PT. Airkon Pratama with the data processing using RapidMiner with the K-Nearest Neighbor method which produces an accuracy of 96.53%. Among them four performance indicators were rated as "good", and two as "adequate". Of the 196 that were correctly predicted to be "good," three were "adequate." Most of the 04 respondents gave a positive response indicating their satisfaction with the management of tax office facilities provided by PT. Airkon Pratama in January 2024.

Keywords: PT. Airkon Pratama; RapidMiner Application; K-Nearest Neighbor Algorithm.

1. Introduction

PT. Airkon Pratama is a Technical & Industrial Consulting company that offers an integrated and coordinated form of managing your facilities. The company has deep expertise in Building Automation Systems, Security and Fire Alarm Systems, and HVAC systems, services and solutions. As part of its growth strategy, PT. Airkon Pratama is looking for high caliber individuals with a positive attitude, a passion for excellence, and customer satisfaction. The right candidate will have a service approach to their work, a deep understanding of quality management practices, and exceptional interpersonal communication and negotiation skills. Among the government buildings managed by PT. Airkon Pratama is a tax office located in Gatot Subroto, South Jakarta. The complex of the tax office includes four buildings: the first building Mar'ie Muhammad 29 floors, the second building A1 6 floors, the third B 17 floors and the fourth A2 6 floors. The purpose of this research is to determine the extent performance of the employees in the building of PT. Airkon Pratama connected with the objectives of improving the quality of service provided to the consumer in view of the fact that every day from PT, the climate control ideas are a lot of issues accomplished by the society or state tax office revealed to the building management. Four buildings and numerous floors, PT. Airkon Pratama struggle to evaluate the customer service satisfaction. That assessment is now based on a form created from work order forms. The forms are signed by clients after PT. Airkon Pratama technicians have handled the reported issue. Next, the work order forms get returned to the admin team to be entered in the system.

Data mining is the practice of automatically searching large stores of data to discover patterns and trends [1]. Data mining is needed because it is still difficult to see how customers assess employee performance PT. Airkon Pratama, the data mining used in this research is election Customer satisfaction prediction will be performed using K-Nearest Neighbor (K-NN) method. K-Nearest Neighbor (K-NN) Algorithm is a machine learning algorithm that is used for classification and regression problems [2]. This is a type of example-based learning, and the algorithm will not necessarily create an explicit internal internal "model" from the dataset, and instead will store the entire training dataset and make predictions based on the similarity of new cases to those already seen. To find out how well the accuracy of classifying customers pleases employees of PT. Airkon Pratama with K-Nearest Neighbor algorithm, this research aims to find a class of customer satisfaction with the services provided by PT. Airkon Pratama. This study is an analysis about the classification of level of customer satisfaction and to determine how accurate the K-Nearest Neighbor analysis of customer satisfaction classification.

The K-Nearest Neighbor (K-NN) classification algorithm has been used in many studies to classify customer satisfaction in the domain of various styles. Halim *et al.* (2023) used K-NN for different levels of customer satisfaction with e-commerce platforms. The study showed that K-NN was able to analyze user reviews and accurately predict satisfaction with respect to platform features [3]. Based on the studies, Ariani *et al* (2020) compared various data mining classification methods to predict consumer satisfaction with Telkomsel prepaid services in Indonesia using K-NN, Naive Bayes, Decision Tree and Random Forest; they discovered that K-NN has the highest accuracy in producing satisfaction classification results [4]. Diansyah (2022) was concerned with K-NN method for user satisfaction classification on technology services. He validated K-NN's ability to classify user satisfaction from survey data [5]. Similarly, Febrian *et al.* (2022) compared two methodologies of K-NN and Naïve Bayes algorithms and shown the performance of K-NN is better than Naïve Bayes techniques for identifying customer satisfaction among multiple products [6]. For classifying `Passenger Satisfaction in the Airline Industry', Setiono (2022) has compared few kind of machine learning algorithms, K-NN, Decision Trees, Random Forest. The K-NN was found to yield competitive results when classifying satisfaction using customer service attributes [7]. Supriyadi *et al.* (2022) utilized K-NN by combining it with the NPS to categorize the satisfaction and loyalty of users for e-learning systems. K-NN was used on this study to classify satisfaction and also to predict for user loyalty in educational contexts, and this study showed good results [8].

Sundari *et al.* It uses the K-NN to classify customer satisfaction in a textile store using data mining techniques (2022). Their work confirmed that K-NN was appropriate for customer satisfaction analysis, gaining useful information for retail management [9]. Similarly, Widiana *et al* (2023) utilized K-NN to forecast the sales of items at a thrift store, indirectly measuring customer satisfaction using sales data. This confirms that K-NN is multi-purpose and can be integrated into various areas of business from retail as well [10]. Nugraha and Hendry (2023) employed machine learning including K-NN to determine the impact of the Kai Access application for passengers of the state-owned railway company of Indonesia on customer satisfaction. It predictiveness for user satisfaction with the app was further validating its general use in all service industries [14]. Furthermore, Pradana *et al.* & Arfin *et al.* (2021) used K-Means and clustering techniques to cluster customers based on their level of satisfaction [12][13]. This study, however, showed that in more complex datasets it a combination of clustering with K-NN is more effective in classifying customer satisfaction even though K-Means was used focuses on clustering [12]. Febriyani *et al*. (2021) applied the C4. Classification of customer satisfaction with Informa products based on K-NN and other methods — comparing with 5 algorithm Their results suggested that whereas C4. Although K-NN was less efficient in comparison with the other methods with k = 5, it was still a well-performing classifier towards customer satisfaction with regard to both product related assessments [13]. Martiani *et al*. (2024) revised the machine learning approach of predicting customer satisfaction, including K-NN, on various services in GrabFood. They also proposed K-NN as a fitting algorithm for predicting customer satisfaction in food delivery services [11].

These studies stress the power of the K-Nearest Neighbor algorithm over a wide variety of domains, including e-commerce, telecommunications, education, retail, and transportation. K-NN has been used consistently as a strong method for customer satisfaction classification and this suggests there is potential for application in a wider variety of industries as well. This research is to enhance of research method of research previously by implementing the K-Nearest Neighbor algorithm to classify customer satisfaction toward the PT. Airkon Pratama services. In light of the performance that K-NN has provided with classification of customer satisfaction in wide broad of industries this paper is going to investigate K-NN algorithm performance in the building maintenance and facility management area. Aim: To analyze and predict customer satisfaction based on employee code for PT. Airkon Pratama through K-NN algorithm by using customer satisfaction data and work order data which had been previously collected.

2. Research Method

The application of methodology refers to a systematic sequence of steps based on established methods and principles, aimed at achieving specific objectives within a research study. The methodology used in this thesis is divided into several sections, as illustrated in the following research flow diagram.

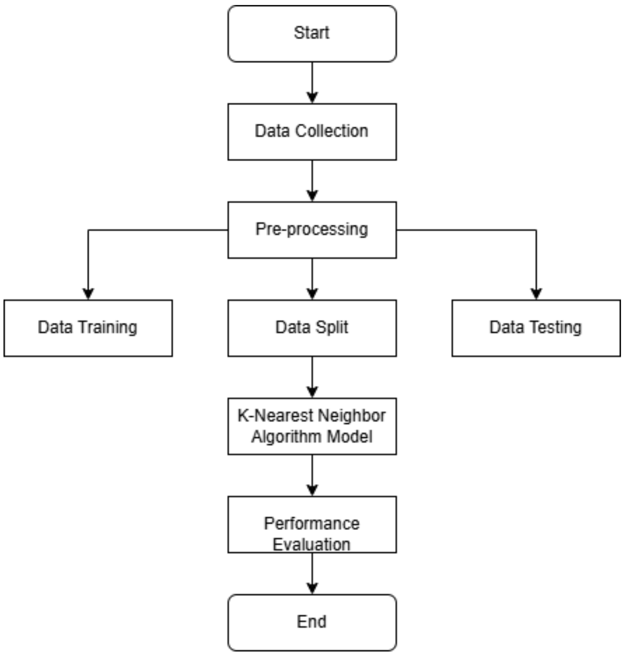

Figure 1. Research Framework

2.1 Data Collection

The dataset used in this research consists of work order reports from PT. Airkon Pratama. The data was collected from the work reports of January 2024. Below is a screenshot of the Excel file containing the work report data:

1	Penanggung Jawab	Keluhan	LOKASI	PIC	status pekerjaan	Respon	komunikasi	cek hasil penghuni
2	Pak Rano	Lampu mati di koridor lift 3 pcs bulb 12 watt	Lt. 6 MM	Security	selesai	cepat	baik dipahami	baik
3	Pak Rano	Lampu mati di koridor 4 pcs bulb 12 watt	Lt. 18 MM	Pak Fajar	selesai	cepat	kurang dipahami	cukup
4	Pak Rano	Lampu mati di toilet pria 1 pcs plc 18 watt	Lt. 10 B	Security	selesai	lambat	baik dipahami	cukup
5	Pak Hapiz	Kran wastafel toilet pria rusak 1 pcs & dispenser handsoap 1 pcs	Lt. 27 MM	Pak Rizal	selesai	cepat	baik dipahami	baik
6	Pak Rano	Lampu mati di koridor toilet 1 pcs bulb 12 watt	Lt. 27 MM	Security	selesai	cepat	baik dipahami	baik
7	Pak Rano	Lampu mati di koridor perpustakaan 3 pcs bulb 12 watt	Lt. 3 MM	Security	tunda	lambat	baik dipahami	baik
8	Pak Rano	Lampu mati arah tangga exit barat 1 pcs bulb 12 watt	Lt. 16 B	Pak Deril	selesai	cepat	baik dipahami	baik
9	Pak Rano	Lampu mati di toilet wanita 1 pcs bulb 12 watt	Lt. 9 MM	Pak Rizal	selesai	cepat	baik dipahami	baik
10	Pak Rano	Lampu mati di koridor lift 1 pcs bulb 12 watt	Lt. 23 MM	Pak Rizal	selesai	cepat	baik dipahami	baik
11	Pak Rano	Lampu mati di depan koridor lift executive & koridor toilet 2 pcs bulb 12 watt	Lt. 11 MM	Pak Rizal	selesai	cepat	kurang dipahami	baik
12	Pak Rano	Lampu mati di depan koridor lift 1 pcs bulb 12 watt	Lt. 2 MM	Pak Rizal	selesai	cepat	baik dipahami	baik
13	Pak Hapiz	Ganti jetshower 1 pcs di toilet pria	Lt. 24 MM	Pak Rizal	selesai	cepat	kurang dipahami	baik
14	Pak Rano	Lampu mati di koridor lift 2 pcs bulb 12 watt	Lt. 19 MM	Pak Kasno	selesai	lambat	baik dipahami	baik
15	Pak Anjar	lampu mati di ruang lift	Lt. 13 B	Pak Ritam	selesai	lambat	baik dipahami	baik
16	Pak Hapiz	Pintu bunyi keras saat di tutup, minta di setting ulang	Lt. 5 B	Pak Guno	selesai	lambat	baik dipahami	baik
17	Pak Hapiz	Perbaikan laci yang susah ditarik	Lt. 17 MM	Pak Hari	selesai	cepat	baik dipahami	baik
18	Pak Rano	Lampu lift direktur mati 1 pcs	Lt. 22 MM	Pak Doni	selesai	lambat	baik dipahami	baik
19	Pak Rano	Lampu kedip 1 pcs	Lt. 10 MM	Pak Krisna	selesai	lambat	baik dipahami	baik
20	Pak Hapiz	kunci laci rusak	Lt. 16 MM	Pak Lukas	selesai	lambat	baik dipahami	baik
21	Pak Rano	Lampu di depan AHU 3 (1 pcs)	Lt. 7 MM	Pak Alfian	tunda	lambat	baik dipahami	baik
22	Pak Hapiz	Ganti jetshower 1 pcs di toilet pria	Lt. 5 MM	Pak Rizal	selesai	lambat	baik dipahami	baik
23	Pak Hapiz	kunci laci rusak	Lt. 18 MM	Ibu Desi	selesai	cepat	baik dipahami	baik
24	Pak Hapiz	Penambahan stop kontak	Lt. 13 MM	Ibu Nina	selesai	cepat	baik dipahami	baik

Figure 2. Raw Data

2.2 Pre-Processing

Pre-processing is a crucial step for addressing issues that may affect data quality, such as inconsistencies in data formatting. This step involves several processes, including data cleaning, integration, transformation, and reduction. The purpose of pre-processing is to normalize or filter the data to ensure its relevance for analysis. The dataset consists of 673 records with 9 attributes, which were then normalized to 8 attributes.

2.3 Data Splitting

The data splitting operator is used to divide the dataset into subsets based on specified parameters. The number of subsets or partitions, as well as the relative size of each partition, is determined by partition parameters. The total ratio of all partitions must match the predefined specifications. The operator also ensures that the data is randomly distributed across the resulting partitions. In this study, 70% of the dataset is allocated for training data, while the remaining 30% is used for testing.

2.4 Modeling

In this phase of the research, the K-Nearest Neighbor (K-NN) algorithm is applied to the dataset to develop a model that classifies customer satisfaction based on the performance of employees at PT. Airkon Pratama. The K-NN algorithm is a supervised machine learning technique used primarily for classification tasks. It works by identifying the 'K' closest data points to a given test instance and classifying the instance based on the majority label of those neighbors. The dataset used in this study consists of customer feedback data, where the target variable is the level of customer satisfaction, and the feature variables include various performance indicators related to the employees' work. Each record in the dataset represents a customer's feedback, and the attributes provide insights into the quality of service, timeliness, and efficiency of the employee performance. The first step in the modeling process is to pre-process and normalize the data, ensuring that all features are on the same scale, which is crucial for the K-NN algorithm to perform effectively. Next, the model is trained using the training dataset, where the K-NN algorithm analyzes the relationship between customer satisfaction levels and the performance indicators. The number of neighbors (K) is a key parameter that influences the model's accuracy, and various values of K are tested to identify the optimal value that results in the highest classification accuracy. The trained model is then ready to classify new customer feedback data into predefined satisfaction categories, based on the patterns learned from the training data. This modeling stage also involves experimenting with different distance metrics (e.g., Euclidean distance, Manhattan distance) to assess which metric provides the best classification performance. The choice of distance metric is critical as it determines how the "closeness" of the data points is measured, which in turn affects the classification results. Therefore, evaluating multiple configurations of K and distance metrics is essential to optimizing the K-NN model.

2.5 Performance Evaluation

Once the K-NN model is developed, its performance must be evaluated to determine how well it classifies customer satisfaction based on employee performance. This is achieved by using a performance evaluation operator, which computes a set of performance metrics designed to assess the model's effectiveness in predicting customer satisfaction. These performance metrics include accuracy, precision, recall, F1 score, and confusion matrices, which provide insights into the model's ability to correctly classify instances and its overall reliability. The performance evaluation operator works by comparing the predicted labels generated by the model against the true labels from the test data. This comparison allows for the calculation of various statistical measures that quantify the model's performance. For classification tasks, accuracy is the most commonly used metric, but additional metrics such as precision and recall are also important, especially when the dataset contains imbalanced classes (e.g., more customers are satisfied than dissatisfied). Precision helps to determine the proportion of positive predictions that were actually correct, while recall measures how many actual positive instances were correctly identified by the model. The versatility of this evaluation operator lies in its ability to apply to different types of learning tasks, whether they are binary classification, multiclass classification, or even regression. By automatically identifying the type of learning task, the evaluation operator ensures that the most relevant performance metrics are selected, allowing for a comprehensive analysis of the model's strengths and weaknesses.

2.6 Accuracy

One main metric for the K-Nearest Neighbor model is accuracy, which is used to measure the performance of the algorithm in predicting customer satisfaction levels. Accuracy is the number of correct predictions made out of all predictions in the test set. The model is then evaluated based on how well it predicts customer satisfaction based on employee-related features using RapidMiner software. Accuracy is calculated by taking the total number of correct classifications over the total number of test instances. As such, Accuracy can be considered as a first measure for evaluating the performance of the model as a whole but it should not be relied upon since full classes are not always present in the dataset. Consider a dataset having 95% satisfied customers and only 5% dissatisfied customers, in which case the model can achieve high accuracy just by prediction of "satisfied" for most of the instances, as such it will give high accuracy even if it is not able to classify the dissatisfied customers correctly. That is why we also look for other metrics like precision, recall, and the f-score — to have a better insight. This is where the performance of the model gets evaluated with the unseen data from the training phase, called the test data. By using the test data, the model can be tested on completely new customer feedback and it will provide an unbiased score on the model's generalization capabilities. Based on employee performance metricsBBy using K-NN algorithm, the research objective is to measure the prediction accuracy by looking at the prediction of whether customers are satisfied with PT. Airkon Pratama services or not. Additionally, techniques like k-fold cross-validation can be used to provide a more robust accuracy measure by ensuring that the reported accuracy is not overly dependent on how we partition the dataset into training and testing. Cross-validation helps to prevent the risk of overfitting, and gives a more accurate estimate of the model's performance across different subsets of the data.

3. Result and Discussion

3.1 Results

3.1.1 Labeling Phase

The labelling phase in machine learning and data mining tasks is an integral part of preparing the data set for training and evaluation of the model. At labeling stage you define which attribute should be use as class (target variable) for prediction in combination of other input features. It helps to clarify which variable is being predicted, and keeps the model from predicting the wrong variable, which is what data at this phase is concerned with. The labeling phase of this research is to determine the predicted class where the predicted class is customer satisfaction level based on employees performance PT. Airkon Pratama. The diagram below shows how this label is attached.

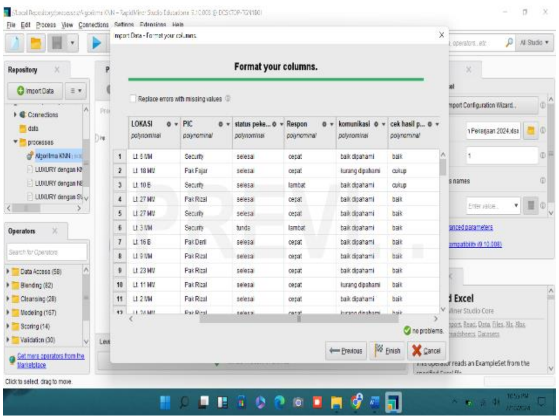

Figure 3. Labeling Phase

3.1.2 Set Role Phase

The Set Role operator is one of the key ones in this process, as it will prepare the dataset for machine learning by describing what role each attribute should play in subsequent operations The purpose of this code is to assign a usage to each attribute of the dataset, or, in other words, to give each attribute in the dataset a "role". For instance, some attributes may be marked as input features (features used for prediction), and some attributes may be used for the target variable(s) (output class). The attribute general role is regular, but there are roles such as label (the target variable) and weight (weighted examples). You want to make sure to assign the roles correctly so the model works as expected. If one or more attributes are assigned the same special role, those attributes will be treated as regular attributes (except for the last attribute assigned that particular special role). The next step is to configure the data for further processing after the roles have been defined. Set Role operator makes sure each attribute is assigned suitable role. The following diagram depicts the process by which roles are assigned to attributes.

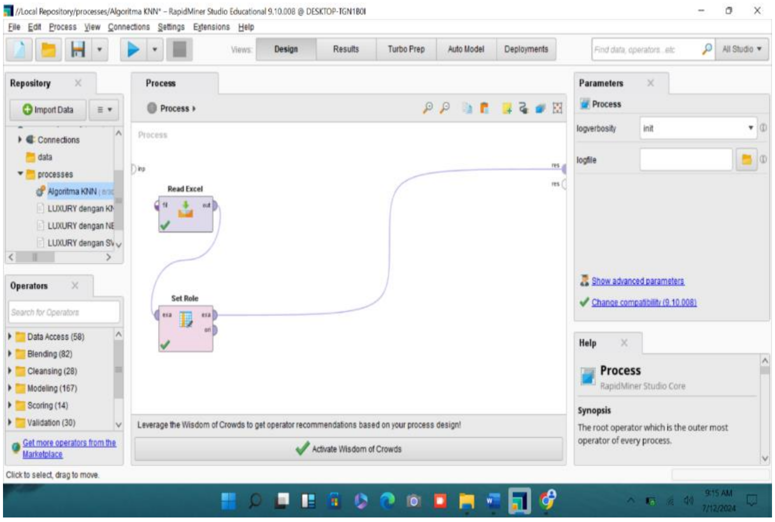

Figure 4. Set Role Phase

3.1.3 Data Split Phase

The Split Data operator is used to split the data into 2 specific data sets for training the model and testing the model. This is an important step in assessing the generalizability of the model and whether it can accurately predict customer satisfaction from unseen data. Data set is divided in a predefined number of subsets, each subset represents a fraction of the complete dataset. Partition parameters determine the overall size of each subset, and alert the operator to Randomly Shuffling the data to remove biases from separate training and testing sets. The Split Data operator in this study is set up to split the dataset into training and testing subset. Partitioning is performed, in which 70 percent of data is allocated for training and 30% for testing. This enables the model to be trained on an adequate amount of data while also being capable of measuring its performance on unseen data. Here is an image of splitting of data.

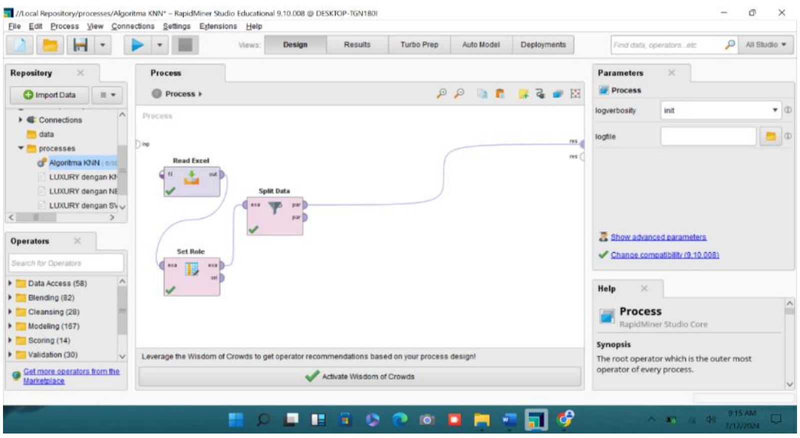

Figure 5. Data Split Phase

3.1.4 Training and Testing Data Setup

As I stated before, we apply Split Data operator to separate training data from testing data. To configure this, we modify the partition parameter such that the sum of the partition ratios is 1.0. The ratio for the training data should be greater than the ratio for the testing data for the model to perform optimally. The dataset is split into 70% for training and 30% for testing the prediction analysis of the model in this study. Splitting of Data, thing is that when you are training the model it is important to have evaluation of model on new5 unseen data for which we need split of data so that one part is used to train the model and other to test it. This way we estimate the accuracy of the model and evaluate the model generalizability beyond the training data. The following diagram demonstrates how the training and testing datasets are divided.

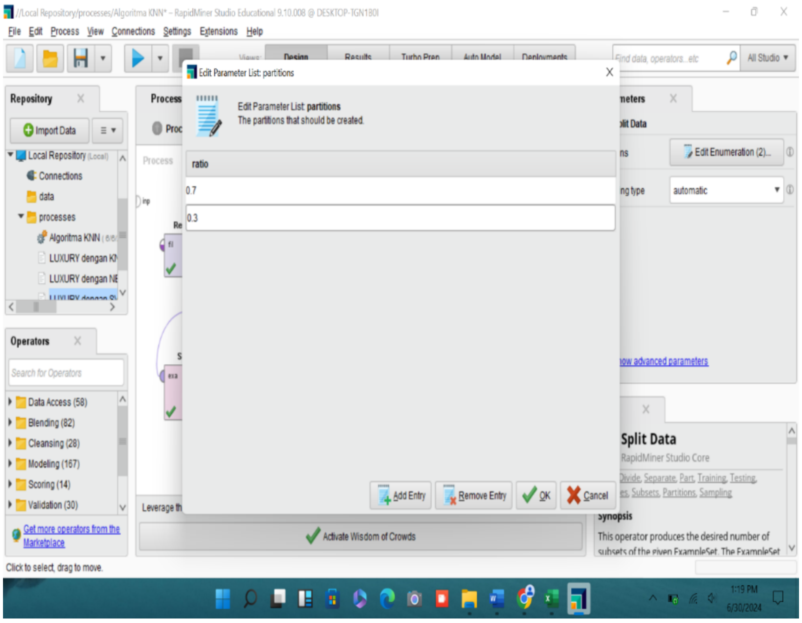

Figure 6. Training and Testing Data Setup

3.1.5 K-Nearest Neighbor Algorithm

After splitting the data appropriately into training and test sets, we can now use the K-Nearest Neighbor (K-NN) algorithm. K-NN is a very simple but powerful machine learning algorithms for classification and regression. In relation to this research, K-NN is used to make customer satisfaction classification from employee performance data. K-NN algorithm finds the "K" of the closest training examples to the entry. How close are points to each other, this is done using a distance metric; for example Euclidean distance which measures the shortest distance between points in the n dimensional space defined by the attributes. It is composed of K closest points in the datasets and its classification is labeled according to the label most common amongst its K closest neighbors. K-NN is a distance metric based algorithm, so we need to normalize the data before applying it. It must be performed as all attributes contributions to distance calculation should be equal The "K" as well as distance metric parameters are changed in the algorithm to get the optimal performance from the algorithm for the research. Below diagram shows the implementation of the K-Nearest Neighbor algorithm.

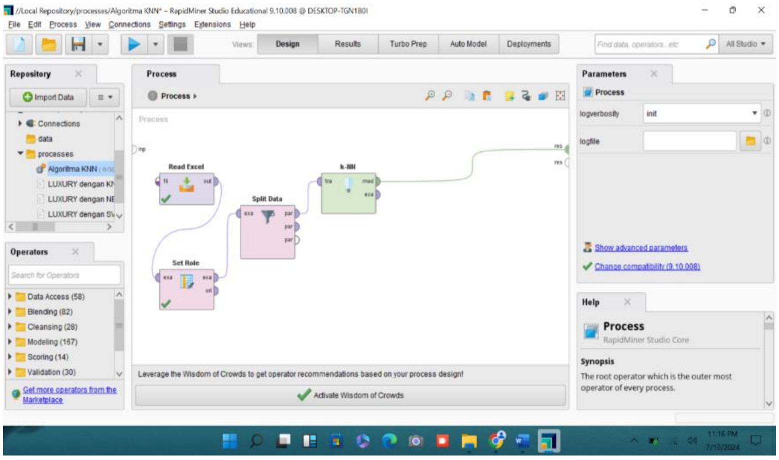

Figure 7. K-Nearest Neighbor Algorithm

3.1.6 Apply Model Phase

After the K-Nearest Neighbor algorithm has been trained on the training dataset, the Apply Model operator is used to apply the trained model to the test data. This phase is critical for obtaining predictions on new, unseen data, which is the ultimate goal of the model. The Apply Model operator takes the trained model and applies it to the testing dataset, generating predictions for each instance based on the features provided. For the model to work correctly during this phase, the test data must have the same number, type, and order of attributes as the training data. This ensures that the model's predictions are made using the same features that it was trained on. The purpose of this step is to evaluate how well the trained model generalizes to new data and to measure its predictive accuracy. The following diagram illustrates the Apply Model phase.

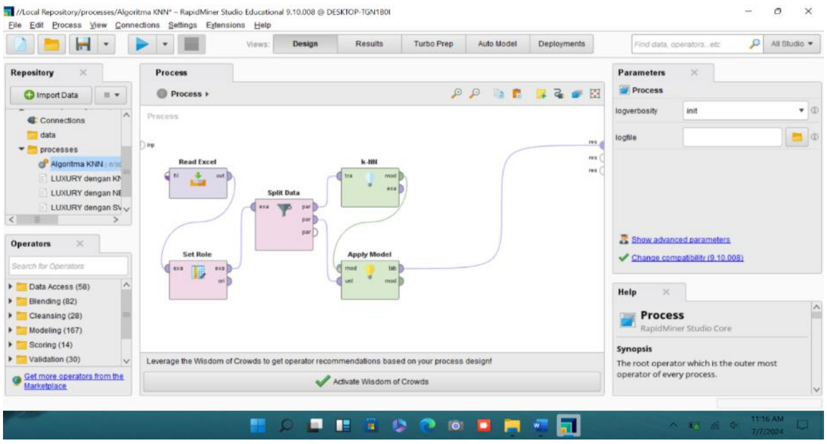

Figure 8. Apply Model Phase

3.1.7 K-Nearest Neighbor Results

The results of applying the K-Nearest Neighbor algorithm are then evaluated by comparing the model's predictions to the actual outcomes in the testing data. These predictions provide insights into the model's performance in classifying customer satisfaction based on employee performance. Key performance metrics, such as accuracy, precision, recall, and F1-score, are used to assess the model's effectiveness. The results from the K-NN algorithm show how well the model classified customer satisfaction into categories such as "Good" and "Adequate." The following diagram presents the results obtained from the K-Nearest Neighbor model.

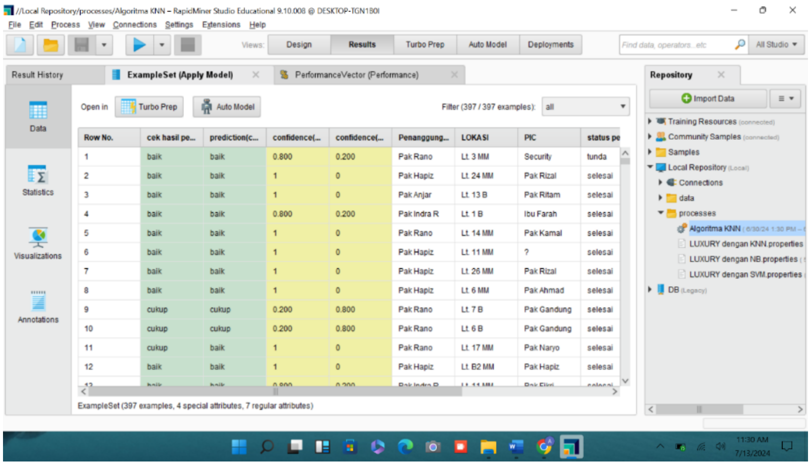

Figure 9. K-Nearest Neighbor Results

3.1.8 Performance Evaluation Phase

The Performance operator is used to evaluate the overall performance of the model. This operator computes various performance metrics based on the predicted and actual outcomes. These metrics help determine the accuracy of the model and its ability to classify customer satisfaction correctly. Common performance metrics for classification tasks include accuracy, precision, recall, and F1-score. Unlike other performance evaluation operators, which are limited to specific tasks like classification or regression, this operator can be used for all types of learning tasks. It automatically determines the appropriate evaluation metrics for the given task. The performance evaluation step provides valuable insights into how well the K-NN model performed in predicting customer satisfaction. The diagram below shows the performance evaluation phase.

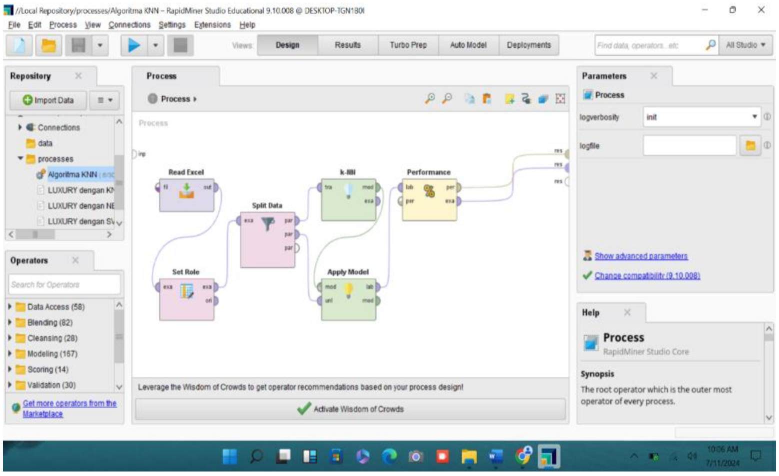

Figure 10. Performance Evaluation Phase

3.1.9 Performance Results

The performance results indicate that the K-Nearest Neighbor algorithm achieved an accuracy of 96.53%. The recall values for the classes "Good" and "Adequate" were 97.97% and 40.00%, respectively. The precision values were 98.47% for the "Good" class and 33.33% for the "Adequate" class. These results suggest that the model is particularly effective at identifying instances of high customer satisfaction ("Good") but less effective in identifying instances of moderate satisfaction ("Adequate"). The diagram below shows the performance results.

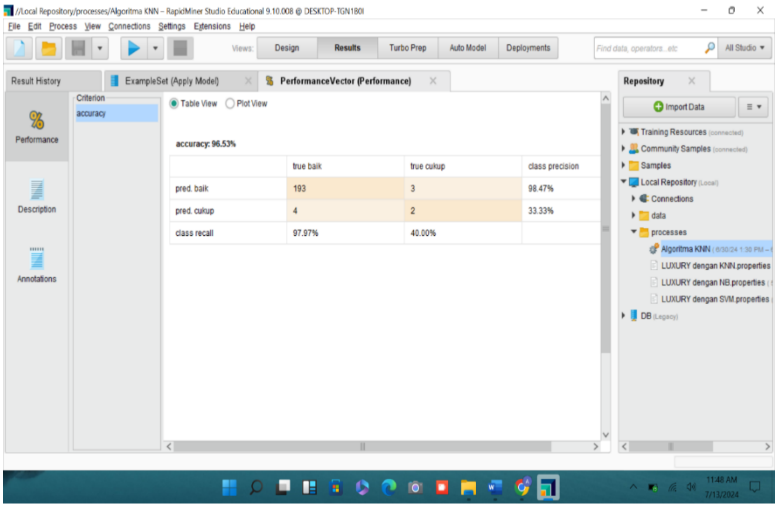

Figure 11. Performance Results

3.1.10 Final Testing Results

The final testing results demonstrate that the K-Nearest Neighbor algorithm successfully predicted customer satisfaction with an overall accuracy of 96.53%. The class recall for "Good" was 97.97%, while "Adequate" had a recall of 40.00%. The precision for the "Good" class was 98.47%, and for "Adequate," it was 33.33%. These findings confirm the algorithm's effectiveness in predicting customer satisfaction based on employee performance at PT. Airkon Pratama, although improvements in detecting moderate satisfaction levels may be needed.

Table 1. Performance Metrics for KNN Model

Metric	Class: Baik	Class: Cukup	Overall
Accuracy	-	-	96.53%
True Positives (TP)	193	2	-
False Positives (FP)	3	4	-
Precision	98.47%	33.33%	-
Recall	97.97%	40.00%	-

3.2 Discussion

The K-Nearest Neighbor (K-NN) algorithm was used to analyze this data to determine the level of customer satisfaction according to employee performance in this study at PT. Airkon Pratama. The accuracy which the developed model has achieved based of the test results was 96.53%, this means that K-NN, in the model, is an accurate method to determine customer satisfaction when the data works well. In this section, we will describe the obtained results in more detail, discuss the factors affecting the results, and compare the results with previous studies. ResultsThe accuracy of the model produced is 96.53%, which is an indication that the K-NN model can classify customer satisfaction well according to the data given. This value suggests that the majority of the model's predictions are aligned with reality, confirming that the only features the model used (indicators of employee performance) are indeed relevant for the model in predicting the level of customer satisfaction. But is it true that this means that we achieve high accuracy, and to be honest if you only take accuracy into account it does NOT guarantee to be the best model (e.g. If one of the data class is more imbalanced from the others). Here is all the information with values the "Fair" class has low recall value as Compared "Good" class also, which is 40.00% with 97.97% for "Good" class. This means the model is accurate in predicting very satisfied(good) customers, but struggles with detecting satisfied customers with some levels of satisfaction.medium(fair).

Recall and precision need to be considered separately for each class. The recall score for "Good" class is 97.97% — indicating the model is good at identifying customers that are satisfied with employee performance. On the downside, precision for "Good" class is 98.47%, meaning that most of predictions predicted as "Good" are truly satisfied customers. On the other hand, the recall for the "Fair" class (40.00%) is much poorer, suggesting that the model often does not find customers who show a moderate level of satisfaction. Possible reason could be — It is highly imbalanced data set between two classes "Good" and "Fair" where "Good" class is mostly present than "Fair" one. This is where again improvement can take place, for instance we can balance the dataset by using the over sampling or under sampling or can further use exploration with other algorithms which more efficiently manage the problem of class measurement. Also, the precision for "Fair" class is 33.33% which is also actually low. Which means that even if the model selected some of the data as "Fair" the model was wrong most of the time. This indicates that while the error rate of the classifier is low for the "Good" class, the misclassification for "Fair" class should still be dealt more cautiously.

The findings of this study are consistent with the previous studies using K-NN algorithms in classifying customer satisfaction. For instance, some research such as Halim *et al*. In the field of e-commerce, (Lazaro, 2023) used K-NN to classify customer satisfaction and showed that K-NN gave accurate results in predicting customer satisfaction based on product and service attributes [3]. In accordance with Ariani and Taufik, 2020, which compared several classification methods for predicting Telkomsel customer satisfaction, both of them found that K-NN as a method that gave better results compared to other methods [4]. However, while we achieved a high overall accuracy in this study, we experienced difficulties in detecting a moderate level of customer satisfaction ("Enough"), with these difficulties being even more pronounced than in these studies due to the imbalance of classes. These studies are typically performed on datasets where the classes were distributed more evenly which are unlikely to be as useful in datasets which have class imbalances seen in this study.

The results are indeed pretty good, yet there is still room for improvement in order to enhance the model performance, especially regarding the classification of the "Enough" class. Class balancing techniques, such as SMOTE (Synthetic Minority Over-sampling Technique) or undersampling can be considered to increase the representation of smaller classes (the "Enough" class in this situation). Also, using other ML algorithms, such as Decision Trees or Random Forest will allow a better comparison in the basis of accuracy and precision of the predictions amongst different classes. Cross-validation also helps to ensure that the model is not influenced by the unequal distribution of data between the training and test data. From this research it is successfully shown that K-NN algorithm can be implemented to classify customer satisfaction on employee performance at PT. Airkon Pratama The model accuracy is relatively high, yet there is always a room of improvement, particularly in detecting mid levels of customer satisfaction. These results are also consistent with previous studies that K-NN is a very effective method for customer satisfaction prediction applications. Still, class balancing for the skewed distribution and experimenting with other algorithms could now be the next steps to take to achieve better accuracy and performance from the model.

4. Related Work

K-Nearest Neighbor (K-NN) has been extensively used in many fields, such as classifying customer satisfaction and predicting product over-sales. K-NN is a simple yet powerful machine learning algorithm, and its effectiveness has been demonstrated in other classification studies, particularly in large datasets with many attributes and features. The K-NN domain includes some of the earliest studies, with Peterson (2009) providing a thorough explanation of the K-NN algorithm, including its advantages and disadvantages. According to Peterson (2009), K-NN is a non-parametric method, meaning it does not assume any specific distribution of data. This makes it suitable for numerous types of classification tasks, including those with complex and non-linear relationships [19]. Salsabilah *et al*. (2023) conducted a comparison between Naive Bayes and K-Nearest Neighbor (K-NN) algorithms in predicting user satisfaction with the TikTok Shop application. They found that K-NN outperformed Naive Bayes in terms of accuracy, suggesting that K-NN is more effective for classifying customer satisfaction in e-commerce applications [17]. In a similar study on customer satisfaction with product quality, Suwandi and Fauzi (2023) also used the Naive Bayes algorithm, highlighting the importance of selecting the right algorithm based on the problem domain and the data being used [16].

Abdy *et al*. (2023) applied K-NN in the automotive industry to predict the sale of the most frequent car spare parts for 2022. This study demonstrated the use of K-NN for predicting sales data based on historical data, showing its capability in handling time-series data and its usefulness for prediction tasks [15]. This approach aligns with our research, where we also predict customer satisfaction (behavior) based on historical data. It further illustrates why K-NN has generated widespread interest in modeling, prediction, and detection. Riadi *et al*. (2023) used K-NN to predict "when" students would graduate based on their knowledge and experience history. While the focus of this study was on education, it illustrates K-NN's versatility across different domains and its high accuracy in handling classification tasks involving categorical data, similar to our research on customer satisfaction prediction [18]. This supports the use of K-NN to accurately predict outcomes based on historical data, as demonstrated in our model predicting customer satisfaction based on employee performance.

Yuliarina & Hendry (2022) conducted a comparative analysis of several prediction models for GoFood service users, using K-NN and Naive Bayes. Their final analysis revealed that both algorithms were effective, but K-NN outperformed Naive Bayes for classifying user satisfaction in complex datasets with many features. This result confirms our previous research, where K-NN also showed similar performance in classifying customer satisfaction based on employee performance [20]. The application of K-NN to predict customer satisfaction and behavior has also been studied in e-commerce and service industries, as extracting useful patterns from users' data is crucial for these businesses. Salsabilah *et al.* (2023) demonstrated how K-NN can assess user satisfaction in online portals, which is directly relevant to our work predicting client satisfaction in a service environment using operational data. Although K-NN works well for many applications, it faces certain challenges, such as class imbalance and handling large datasets. As noted by Salsabilah *et al.* (2023) and Riadi *et al.* (2024), K-NN may perform poorly on datasets that are not pre-processed correctly or when there is a significant disparity in class distribution. In our dataset, we experienced the same issue, where the majority class (Good) was predicted with high recall, likely due to the imbalance between the Good and Adequate classes. K-NN has demonstrated its capability in classifying customer satisfaction based on different features across various domains, such as e-commerce, automotive sales, and educational data. Our research adds to this body of literature by examining K-NN for predicting customer satisfaction based on employee performance in a facility management environment at PT. Airkon Pratama. These insights contribute to the argument that K-NN is well-suited as an evaluation tool in service-based environments where performance metrics are common. Abdy *et al.* (2022) and Salsabilah *et al.* (2023) successfully used K-NN to predict customer preferences and satisfaction in e-commerce and the automotive sector. These studies serve as valuable references for understanding how K-NN can be generalized to industry-specific datasets, such as customer reviews or retail logs.

As demonstrated by the related works, K-Nearest Neighbor is an adaptable and powerful algorithm, optimized for a range of applications, from predicting consumer satisfaction in e-commerce to forecasting sales and academic performance. K-NN has been particularly effective for classification tasks where the relationships between features are non-linear, and the data is high-dimensional and categorical. To the best of our knowledge, this work extends existing research by testing and validating K-NN's success in predicting customer satisfaction based on employee performance, an area not previously explored. This provides further evidence of K-NN's usefulness in predicting customer satisfaction in service industries. The widespread use of K-NN across various fields is a testament to its strong ability in predictive modeling and understanding customer behavior. However, challenges such as imbalanced classes and performance in complex classification tasks suggest potential avenues for future research, including the use of ensemble learning or deep learning models to improve upon K-NN.

5. Conclusion

The results of this study indicate that the K-Nearest Neighbor (K-NN) method applied to classify customer satisfaction based on data obtained from PT. Airkon Pratama, and tested using the RapidMiner application, produces an accuracy of 96.53%. This model classifies 4 performance indicators as "good" and 2 performance indicators as "sufficient". From the prediction results, 193 data were predicted as "good" and 3 data were predicted as "sufficient". Based on these findings, it can be concluded that most customers are satisfied with the performance of PT. Airkon Pratama in managing tax building facilities during the January 2024 period. These results indicate that the company's overall performance has met customer expectations, although there is room for improvement in several aspects of service to further increase customer satisfaction levels.

References

[1] Saragih, D. A. R., Safii, M., & Suhendro, D. (2021). Penerapan Data Mining Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Sistem Informasi di Program Studi Sistem Informasi. Journal of Information System Research (JOSH), 2(2), 173-177.

[2] Hendriyanto, M. D., & Sari, B. N. (2022). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Judul Berita Hoax. Jurnal Ilmiah Informatika, 10(02), 80-84. https://doi.org/10.33884/jif.v10i02.5477

[3] Halim, T. N., Martin, R., Hutasoit, R. K. B., & Aisyah, S. (2023). Klasifikasi Kepuasan Pelanggan Terhadap Platform E-Commerce dengan Metode K–Nearest Neighbor (K-NN). Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika), 8(2), 512-523. http://dx.doi.org/10.30645/jurasik.v8i2.636

[4] Ariani, F., & Taufik, A. (2020). Perbandingan Metode Klasifikasi Data Mining untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar. SATIN-Sains dan Teknologi Informasi, 6(2), 46-55. https://doi.org/10.33372/stn.v6i2.666

[5] Diansyah, S. (2022). Klasifikasi tingkat kepuasan pengguna dengan menggunakan metode K-Nearest Neighbour (KNN). *Jurnal Sistim Informasi dan Teknologi*, 4, 7–12. https://doi.org/10.37034/jsisfotek.v4i1.114

[6] Febrian, A. K., Chrisnanto, Y. H., & Sabrina, P. N. (2022, April). Studi Komparasi Metode Klasifikasi K-Nearest Neighbor dan Naive Bayes Dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk. In Prosiding Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika (SNESTIK) (Vol. 1, No. 1, pp. 333-338). https://doi.org/10.31284/p.snestik.2022.2717

[7] Setiono, M. H. (2022). Komparasi algoritma Decision Tree, Random Forest, SVM dan K-NN dalam klasifikasi kepuasan penumpang maskapai penerbangan. *INTI Nusa Mandiri*, 17(1), 32–39. https://doi.org/10.33480/inti.v17i1.3420

[8] Supriyadi, D., Safitri, S. T., Amriza, R. N. S., & Kristiyanto, D. Y. (2022). Klasifikasi loyalitas pengguna sistem e-learning menggunakan Net Promoter Score dan machine learning. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 8(1), 38. https://doi.org/10.26418/jp.v8i1.49300

[9] Sundari, S., Windarto, A. P., Purba, Y. P., & Artikel, G. (2022). Penerapan data mining klasifikasi terhadap kepuasan pelanggan toko Bombay Textile. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, 1(4), 2828–9099. https://doi.org/10.55123/jomlai.v1i4.1672

[10] Widiana, S. A., Firdaus, I., Tenda, E., & Ketaren, E. (2023). Sistem Informasi Prediksi Penjualan Produk Thrift Di Toko Manado Menggunakan Algoritma K-Nearest Neighbor (K-Nn). Jurnal TIMES, 12(2), 52-57. https://doi.org/10.51351/jtm.12.2.2023708

[11] Martiani, E., Setyanto, A., & Nasiri, A. (2024). Studi literatur mengenai prediksi kepuasan GrabFood menggunakan machine learning. *Dielektrika*, 11(1), 22–31. https://doi.org/10.29303/dielektrika.v11i1.367

[12] Pradana, M. G., Nurcahyo, A. C., & Saputro, P. H. (2021). Penerapan metode K-Means klustering untuk menentukan kepuasan pelanggan. *Creative Information Technology Journal*, 7(1), 42. https://doi.org/10.24076/citec.2020v7i1.185

[13] Febriyani, A., Prayoga, G. K., & Nurdiawan, O. (2021). Index kepuasan pelanggan Informa dengan menggunakan algoritma C.45. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 330. https://doi.org/10.30865/jurikom.v8i6.3686

[14] Nugraha, F. T., & Hendry, H. (2023). Implementasi machine learning sebagai analisis kepuasan pelanggan terhadap penggunaan aplikasi Kai Access. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8(4), 1248–1255. https://doi.org/10.29100/jipi.v8i4.4185

[15] Abdy, S., Gultom, E. R. B., Ramadhany, S., & Afifudin, A. (2022). Prediksi penjualan sparepart mobil terlaris menggunakan metode K-Nearest Neighbor. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 2003. https://doi.org/10.30865/jurikom.v9i6.5189

[16] Suwandi, A., & Fauzi, R. (2023). Analisis kepuasan konsumen terhadap kualitas produk menggunakan algoritma Naive Bayes. *Computer and Science Industrial Engineering (COMASIE)*, 9(7), 30. https://doi.org/10.33884/comasiejournal.v9i7.7921

[17] Salsabilah, B. N., Kadek, I., & Nuryana, D. (2023). Komparasi algoritma Naive Bayes dan K Nearest Neighbor dalam kepuasan pengguna fitur Tiktok Shop. *Journal of Emerging Information System and Business Intelligence (JEISBI)*, 4(3), 31–39. Retrieved from https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/54211

[18] Riadi, I., & *et al*. (2024). Prediksi kelulusan tepat waktu berdasarkan riwayat akademik menggunakan metode K-Nearest Neighbor. *Jurnal Teknik Informatika*, 11(2), 249–256. https://doi.org/10.25126/jtiik.2024117330

[19] Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883. https://doi.org/10.4249/scholarpedia.1883

[20] Yuliarina, A. N., & Hendry, H. (2022). Comparison of prediction analysis of Gofood service users using the KNN & Naive Bayes algorithm with Rapidminer software. *Jurnal Teknik Informatika (Jutif)*, 3(4), 847–856. https://doi.org/10.20884/1.jutif.2022.3.4.294.

