Comparative Analysis Of Naïve Bayes And K-Nearest Neighbors Algorithms In Predicting Public Interest In Electric Motorcycles

Arief Syahputra[1*], Syaiful Zuhri Harahap[2], Marnis Nasution [3]

[1,2,3] Faculty of Science and Technology, Universitas Labuhanbatu, Sumatera Utara Indonesia
[*]Corresponding Author:
Email: ariefsyahputra001@gmail.com

Abstract.

Concerns about global warming and the need for sustainable transport solutions have led to the emergence of electric vehicles as an alternative to conventional vehicles. These vehicles offer cleaner and more efficient transportation, especially in urban areas. However, high costs in some countries, such as Indonesia, hinder their adoption. Electric motorcycle companies use previously recorded data to predict customer interest in purchasing their products. Based on these forecasts, business owners make decisions about the quantity of goods to supply. Researchers are increasingly using machine learning algorithms to study consumer behavior and predict demand for a variety of products, including electric motorcycles. The researchers have used this data-based method to analyze large amounts of consumer data, including online survey responses and reviews, as well as other sources. The study aims to perform a comparative analysis of the performance of the Naïve Bayes and K-Nearest Neighbors algorithms to predict public interest in electric motorcycles in Labuhanbatu district. We will perform a comparative analysis based on the performance evaluation matrix (accuracy, precision, recall, and f1 score) to determine the most suitable algorithm. The phases of the research methods included: data collection, exploratory data analysis, data preprocessing, splitting data, implementation of naïve bayes and k-nearest neighbors, and evaluation. The results showed that Naïve Bayes achieved an accuracy of 31.13%, precision of 83.33%, recall of 4.63%, and f1-score of 8.77%. In contrast, K-Nearest Neighbors attains an accuracy of 71.52% and a precision of 71.52%, recall 100%, and f1 score of 83,40%. According to the research results, K-nearest neighbors showed much better results in terms of accuracy, recall, and F1 scores. These results demonstrate that K-nearest neighbor is more effective in detecting public interest in electric motorcycles, and it strikes a better balance between identifying all positive cases and avoiding predictive errors.

Keywords: *Electric Motorcycles, K-Nearest Neighbors, Prediction, Machine Learning and Naïve Bayes.*

I. INTRODUCTION

With growing global concerns about environmental pollution and an urgent need for sustainable transport solutions, electric motorcycles have emerged as a promising alternative as a replacement for conventional petrol-powered vehicles [1]. Electric motors offer a much cleaner and more energy-efficient means of transportation, especially in densely populated urban areas, where emissions and traffic congestion are major problems [2]. However, the adoption of electric motorcycles is relatively low in some regions, such as Indonesia, due to a lack of supportive charging infrastructure, high initial purchase prices compared to conventional motors, and consumer perceptions of the performance and reliability of electrical motorcycle technology [3], [4]. Despite these challenges, the potential benefits of electric motorships, like reduced emissions, lower operating costs, and better energy efficiency, have prompted governments and policymakers to implement initiatives to promote their wider adoption [5]. Strategies to overcome obstacles to the adoption of electric motorcycles, such as investing in charging infrastructure, providing financial incentives, and raising consumer awareness, have been the focus of ongoing research and policy discussions [6], [7].Owners of electric motorcycle companies are trying to determine the interest of future consumers in buying their goods on the basis of previously recorded data [8]. These predictions greatly influence the decision of the company's owners to determine how much goods the company must provide in order to avoid the production of defective products or goods in large quantities, and it turns out that the sale of such goods only sells a few goods [9], [10].

Effective long-term and short-term planning depends on predicting demand for company products [11]. By incorporating these predictions into the production planning process, companies can effectively plan their production and achieve optimal results, thereby reducing the risk of planning errors [12], [13]. Companies need to develop marketing strategies to generate ready-to-use information that helps the marketing team make strategic decisions [14].The application of machine learning algorithms has become increasingly common in the study of consumer behavior and demand predictions for a variety of products, including electric motorcycles [15], [16], [17]. Researchers have used this data-based technique to analyze a large number of customer data, including response surveys, online reviews, and other sources, to gain a deeper insight into customer preferences, purchasing intentions, and the factors that influence the adoption of new technologies such as electric bikes [18], [19]. By applying advanced machine learning methods, researchers have been able to develop more sophisticated and accurate predictive models that can better predict market trends and guide strategic decision-making for companies that want to introduce or expand their electric bike offerings [20], [21]. This has proven invaluable for electric motorbike manufacturers, as they are trying to navigate complex market landscapes and develop marketing strategies that effectively drive increased consumer adoption of their products [22]. Several studies have shown the potential of machine learning algorithms for analyzing consumer attitudes and predicting demand for electric vehicles.

Researchers [23] used a machine learning model to predict whether people would "buy" or "don't buy" electric vehicles. Machine learning models show that age, gender, income, level of environmental concerns, vehicle costs, running costs, vehicle performance, driver range, and mass behavior are significant factors that influence consumer decisions to buy electric vehicles in India. The hybrid model (Hybrid LSTM with Two-Dimensional Attention and Residual Network) was better than other models using similar evaluation units at predicting the number of electric vehicles, with an acceptable 3.5% error rate on average [24]. The results of the study, conducted by [25], showed the use of two different algorithms: Naïve Bayes and Support Vector Machine. The support vector machine achieves the highest accuracy of 90%, whereas the naive bayes algorithm only achieves an accuracy of 88%. Thus, the support vector machine prevails in this study. Researchers [26] have used Naïve Bayes and K-Nearest Neighbors (KNN) methods to analyze customer opinions and knowledge about electric vehicles in India. The results of this analysis have provided valuable insights into the factors that influence consumer adoption of electric motorcycles, such as environmental awareness, perceived performance, and cost efficiency.The aim of this study is to provide a comparative analysis of the performance of Naïve Bayes and K-Nearest Neighbors algorithms in predicting public interest in electric motorcycles, especially in the Indonesian market in Labuhanbatu district. By analyzing consumer survey data conducted online, the study seeks to guide strategic decision-making and marketing for electric motorcycle manufacturers who want to drive increased adoption of their products in the Indonesian market, especially in Labuhanbatu district. We will determine the best algorithm by comparing the performance of the two applied algorithms using the performance evaluation matrix, which includes accuracy, precision, recall, and f1-score.

II. METHODS

The research methodology used in this study involved a comparative analysis of the machine learning algorithms Naïve Bayes and K-Nearest Neighbors (KNN) in identifying and understanding the public interest in electric motorcycles in Labuhanbatu district. All stages of this study use the Python programming language implemented in the Google Colab text editor. This study's research methods include the steps shown in Figure 1.

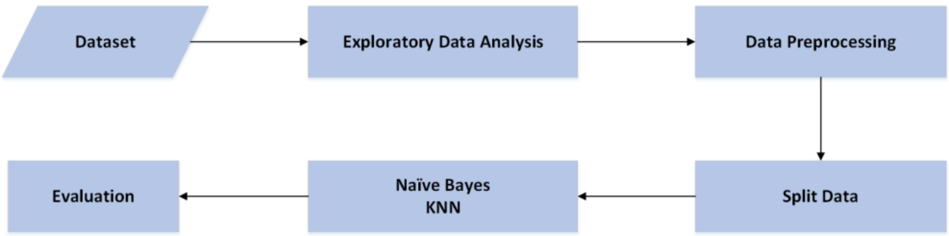

Fig 1. Research Stage

The first phase is the collection of datasets. The data sets used in this research are obtained from the results of the completion of questionnaires by the community of Labuhanbatu district regarding public interest in buying electric motorcycles. We collected the research data by disseminating the questionnaire via Google Forms. We obtained the research data by conducting a comprehensive survey among sample representatives of the population in Labuhanbatu district, Indonesia. Survey questions explore factors such as age, gender, marital status, type of job, income, motorcycle ownership, and interest in buying electric motorcycles.Exploratory Data Analysis (EDA) is the second important stage in this research method, with the aim of exploring patterns in the data. At this stage, the EDA process is divided into three structured parts: data identification, univariate analysis, and bivariate and multivariate analysis. The third stage is data preprocessing. This process is considered a critical step in the process of discovering knowledge, ensuring that decisions are based on high-quality data. This step minimizes data errors and overcomes systematic bias before analysis takes place. At the stage of data preprocessing: handling missing values, feature encoding, and data normalization. The fourth phase is the separation of the datasets into data sets for training and data sets for testing. This study applies training data division of 70% and test data division of as much as 30%. We will use Naïve Bayes and K-Nearest Neighbors (KNN) algorithms in the fifth phase to analyze survey data and predict public interest in electric motorcycles. The confusion matrix will measure each of these algorithms, evaluating the predictive model's performance by comparing the number of true and wrong predictions to the actual value. Then compare the performance by looking at the level of accuracy, precision, recall, and f1-score.

III. RESULT AND DISCUSSION

We will present the systematic results of applying research methods from the comparative analysis of Naïve Bayes and K-Nearest Neighbors algorithms to determine public interest in electric motorcycles, following the previously described phases of the research.

Table 1. Sample dataset

No	Age	Gender	Married	Job	Income	Motorcycle	Interest
1	19	Male	Single	Self-employed	<=5M	Yes	Interested
2	29	Male	Single	Private-employee	<=5M	Yes	Interested
3	18	Female	Single	Private-employee	<=5M	Yes	Interested
4	20	Male	Single	Private-employee	<=5M	Yes	Interested
5	23	Male	Single	Private-employee	<=5M	Yes	Interested

Table 1 shows a portion of the research data set obtained from the answers of respondents in the community in Labupatan Labuhanbatu through the distribution of questionnaires online through Google Form. The data set's total purity is 501 lines (respondents) and 7 columns. The age column represents the respondents' ages, ranging from 17 to 57 years old. The gender column consists of men and women. The married column displays the marital status of single, married, and divorced individuals. The job column is a job category that consists of students, private employees, public servants, and private employees. The income column is monthly income that consists of three categories: 0–5 million, 5–10 million, and above 10 million. A motorcycle column is the ownership of a motorcycle, whether it exists or not. The column of interest represents the public's interest in purchasing an electric motorcycle, regardless of their level of interest. The next step will display the results of the exploratory data analysis.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 501 entries, 0 to 500
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   age         501 non-null    int64
 1   gender      501 non-null    object
 2   married     501 non-null    object
 3   job         501 non-null    object
 4   income      501 non-null    object
 5   motorcycle  501 non-null    object
 6   interest    501 non-null    object
dtypes: int64(1), object(6)
memory usage: 27.5+ KB
```

Fig 2. Data Identification

Figure 2 shows the result of executing the df.info() command on a Python-defined DataFrame. This is a DataFrame from the Pandas library, which contains data from questionnaires about public interest in electric motorcycles. There are 501 lines, indexed from 0 to 500. There are 7 columns with different types of data. The data set consists of categories and numerical variables. There are six categories of object data variables consisting of columns: gender, married, job, income, motorcycle, and interest.

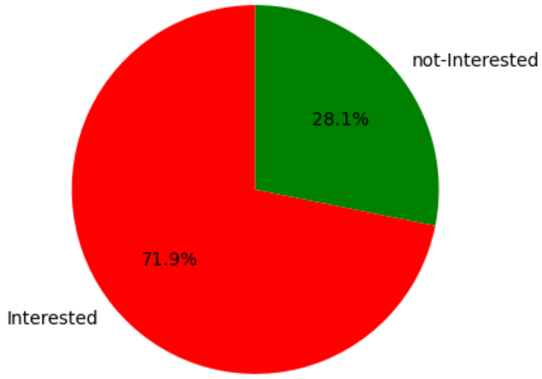

Fig 3. Distribution of Interest

Figure 3 shows a pie chart showing the distribution of public interest in electric motorcycles. This image is the result of univariate analysis, which aims to analyze the distribution of one variable, namely interest in the data set. This analysis helps to understand how many respondents have shown interest or indifference in electric motorcycles. The red portion indicates a significant portion, accounting for 71.9% of the 360 respondents who expressed interest in electric motorcycles.

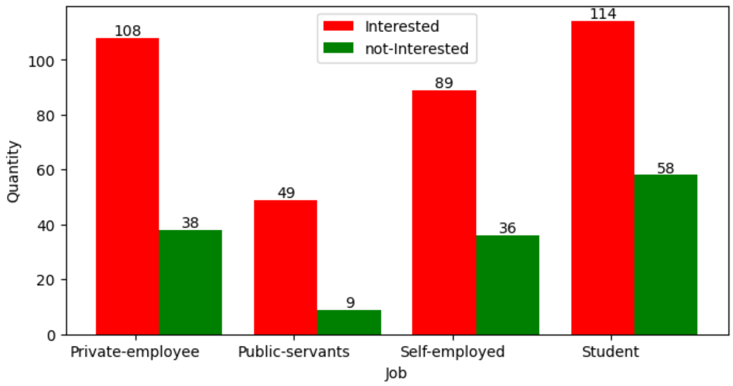

Fig 4. Interest status by Job

Figure 4 shows a grouped bar chart showing the relationship between jobs and interests in electric motorcycles. (interest). Axis X (horizontal) represents the job categories consisting of: private employees, public servants, self-employed, and students. Axis Y (vertical) represents the number of respondents (quantity) in each category. In the private-employee group, there were 108 interested respondents and 38 uninterested respondents. In the public-servants group, 49 and 9 were not interested. Of the self-employed group, 89 were interested, and 36 were not.

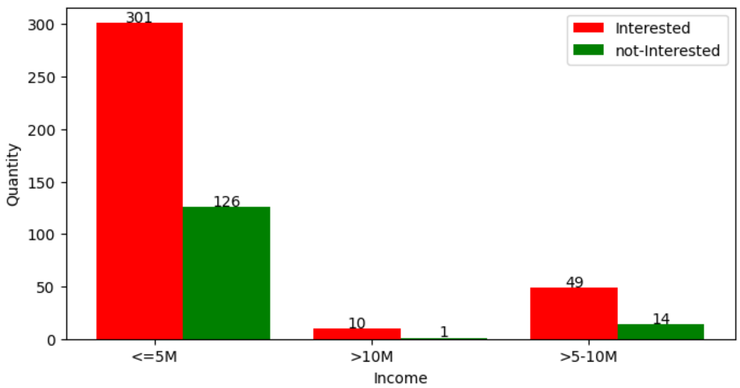

Fig 5. Interest status by Income

Figure 5 illustrates the relationship between income and people's interest in electric motorbikes. This graph shows the amount of interest and disinterest in electric motorbikes by income category. Income (Revenue): combines into three categories: <=5M is income less than or equal to 5 million, 5-10M is income more than 5 million to 10 million. >10 million is an income of more than 10 million. From the graph it can be seen that the majority of people with an income of less than or equal to 5 million are more interested (301 people) than those who are not interested (126 people). More people with incomes between 5 million and 10 million are also interested (49 people) than those who are not interested (14 people). People with an income of more than 10 million show very low interest (10 people are interested) and only 1 person is not interested.

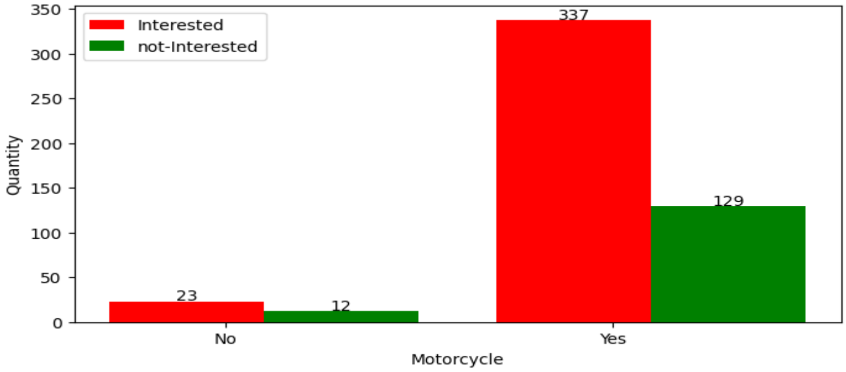

Fig 6. Interest status by Motorcycle

Figure 6 illustrates the relationship between ownership of conventional motorbikes (Motorcycle) and people's interest in electric motorbikes. This graph compares the number of interests and disinterests in electric motorbikes based on conventional motorbike ownership. The category of conventional motorbike ownership is grouped into two categories, namely, Yes: Owns a conventional motorbike, and No: Does not own a conventional motorbike. From the graph it can be seen that people who do not own a conventional motorbike show higher interest (23 people) than those who are not interested (12 people). Meanwhile, people who own conventional motorbikes show much higher interest (337 people) than those who are not interested (129 people). Next, the results of the preprocessing data are displayed.

```
[ ]  # Show features with percentage of missing values
     missing_values()
```

	Total of missing	Percentage of missing
age	0	0.0%
gender	0	0.0%
married	0	0.0%
job	0	0.0%
income	0	0.0%
motorcycle	0	0.0%
interest	0	0.0%

Fig 7. Percentage of missing values

Figure 7 displays the results of checking missing values on the dataset used to analyze public interest in electric motorbikes. This image is part of the preprocessing data. Total of missing is the number of missing values for each feature in the dataset. Percentage of missing is the percentage of missing values for each feature in the dataset. These results show that all features in this dataset have no missing values, which is indicated by the total missing values and the percentage of missing values which are both 0 for each feature.

	is_No	is_Yes	is_<=5M	is_>10M	is_>5-10M	is_Private-employee	is_Public-servants	is_Self-employed	is_Student	is_Divorced	is_Married	is_Single	is_Female	is_Male	age	interest
0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	19	1
1	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	29	1
2	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	18	1
3	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	20	1
4	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	23	1

Fig 8. The results of Feature Encoding

Figure 8 is the result of feature encoding at the data preprocessing stage of the public interest dataset for electric motorbikes. This encoding feature includes One-Hot Encoding for category features such as gender, married, job, income, and motorcycle, as well as Label Encoding for interest features. From these results, there are additional feature columns from the original 7 columns, increasing to 16 columns. The is_No feature is an indicator for "No" motorbike ownership (0 if no, 1 if yes), while is_Yes is an indicator for "Yes" motorbike ownership (0 if no, 1 if yes). Feature is_<=5M is Indicator for revenue "≤5M" (0 if no, 1 if yes), is_>5-10M: Indicator for revenue ">5-10M" (0 if no, 1 if yes), and is_ >10M is an indicator for revenue ">10M" (0 if no, 1 if yes), and. The feature is_Private-employee is an indicator for the job "Private-employee" (0 if no, 1 if yes), is_Public-servants is an indicator for the job "Public-servants" (0 if no, 1 if yes), is_Self-employed is an indicator for the job "Self-employed" (0 if no, 1 if yes), and is_Student is the indicator for the job "Student" (0 if no, 1 if yes). The feature is_Divorced is an indicator for marital status "Divorced" (0 if no, 1 if yes), is_Married is an indicator for marital status "Married" (0 if no, 1 if yes), and is_Single is an indicator for marital status "Single" (0 if no, 1 if yes). The is_Female feature is an indicator for the gender "Female" (0 if no, 1 if yes), and is_Male: An indicator for the gender "Male" (0 if no, 1 if yes). Age feature: respondent's age in original numeric form. Meanwhile, the interest feature is interest in electric motorbikes which has been encoded using Label Encoding (1 indicates interested, and 0 indicates not interested).

```
# Call the 'normalize' function on the 'df' DataFrame
df_final = normalize(df, dict_min, dict_max, decimal_places=3)
df_final.head()
```

	is_No	is_Yes	is_<=5M	is_>10M	is_>5-10M	is_Private-employee	is_Public-servants	is_Self-employed	is_Student	is_Divorced	is_Married	is_Single	is_Female	is_Male	age	interest
0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.073	1
1	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.317	1
2	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.049	1
3	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.098	1
4	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.171	1

Fig 9. The results of Data Normalization

Figure 9 shows the results of the data normalization process using the Min-Max Normalization method at the data preprocessing stage of the dataset. This process is applied to the age column, so that the value in the age column changes to a value in the range 0 to

```
# Train - test split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3, random_state = 45)
```

```
print("Training data shapes:")
print("X_train:", x_train.shape)
print("y_train:", y_train.shape)

print("\nTesting data shapes:")
print("X_test:", x_test.shape)
print("y_test:", y_test.shape)
```

```
Training data shapes:
X_train: (350, 15)
y_train: (350,)

Testing data shapes:
X_test: (151, 15)
y_test: (151,)
```

1.

Fig 10. Spliting into training and test data

Figure 10 shows the process of splitting data (split data) into training data and test data using the train_test_split function from the scikit-learn library in Python. The argument test_size=0.3 indicates that 30% of the dataset will be used as test data and 70% as training data. The splitting results show that the training data (x_train and y_train) has a size of (350, 15), which means there are 350 training samples with 15 features per sample. The test data (x_test and y_test) have a size of (151, 15), which means there are 151 test samples with 15 features per sample. After sharing this data, the training data will be used to train the model using the Naïve Bayes and K-Nearest Neighbors (KNN) algorithms. Test data will be used to test the performance of the trained model.

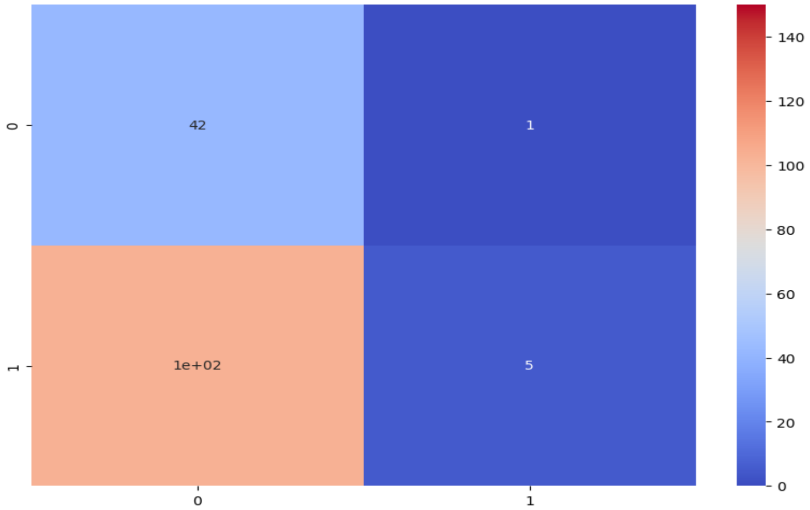

Fig 11. Confusion Matrix of Naïve Bayes

Figure 11 shows the confusion matrix results from testing the Naïve Bayes model on a dataset of public interest in electric motorbikes. In the image, the values displayed are TN (True Negatives): 42, FP (False Positives): 1, FN (False Negatives): 103, and TP (True Positives): 5. True Negatives (TN = 42) shows that the model correctly predicted 42 samples as negative (class 0). False Positives (FP = 1) indicates that the model incorrectly predicted 1 sample as positive (class 1) when it was actually negative (class 0). False Negatives (FN = 103) indicates that the model incorrectly predicted 103 samples as negative (class 0) when they were actually positive (class 1). True Positives (TP = 5) indicates that the model correctly predicted 5 samples as positive (class 1).

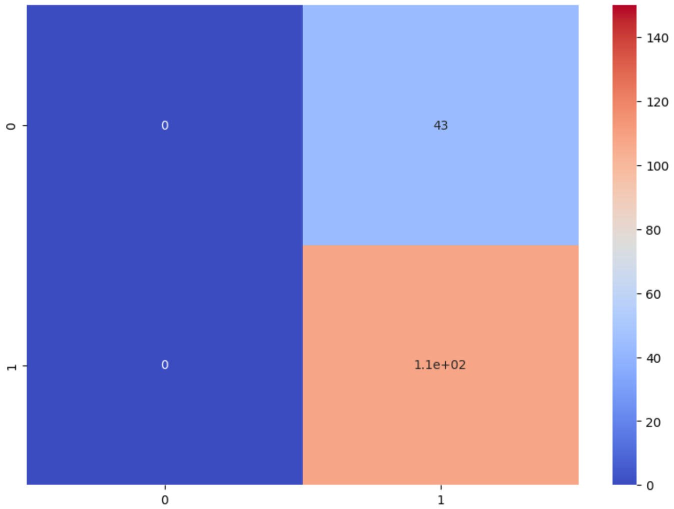

Fig 12. Confusion Matrix of KNN

Figure 12 shows the confusion matrix results from testing a dataset of public interest in electric motorbikes using the K-Nearest Neighbors (KNN) algorithm. TN with a result of 0, indicates that there are no negative cases predicted as negative. FP with a result of 43, shows that there were 43 negative cases that were predicted to be positive. FN with a result of 0, indicates that there are no positive cases predicted as negative. TP with a result of 108, shows that there are 108 positive cases that are predicted to be positive.

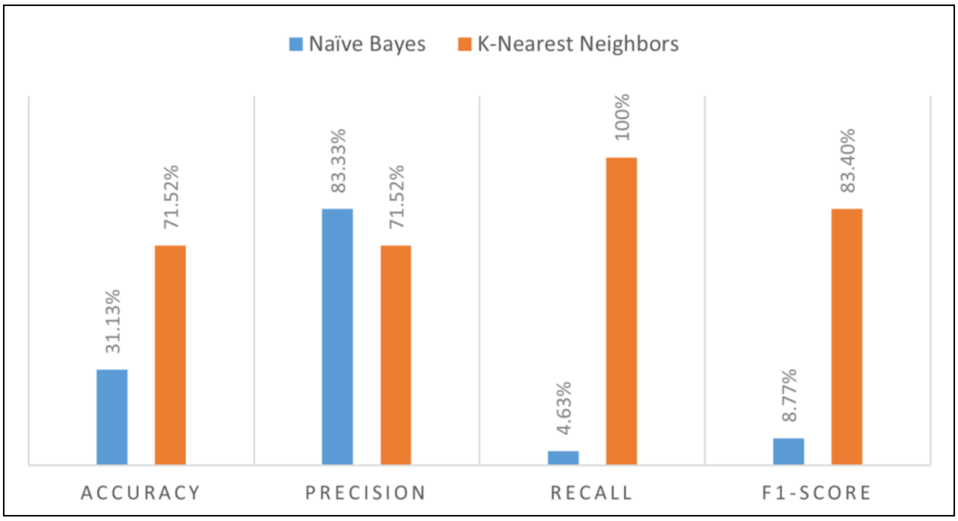

Fig 13. Comparison based on Evaluation Matrix

Figure 13 shows a comparison of the evaluation matrix results of two machine learning algorithms, namely Naïve Bayes and K-Nearest Neighbors, which were applied to a dataset of public interest in electric motorbikes. The evaluation matrix includes four main metrics: Accuracy, Precision, Recall, and F1-score. From this image, the accuracy results from Naïve Bayes were 31.13%, and K-Nearest Neighbors: 71.52%. The precision of Naïve Bayes is 83.33%, and K-Nearest Neighbors is 71.52%. Recall from Naïve Bayes 4.63%, and K-Nearest Neighbors 100%. F1-score from Naïve Bayes 8.77%, and K-Nearest Neighbors 83.40%.From the distribution of public interest in electric motorbikes based on the results of exploratory data analysis, this research shows that there is significant interest in electric motorbikes among respondents, which could be a positive indication for electric motorbike manufacturers in designing their marketing strategies. Overall, more respondents from various job categories showed interest in electric motorbikes than those who were not interested. Students show the highest interest in electric motorbikes followed by Private-employees, Self-employed, and Public-servants. These results can help identify market segmentations that have a high interest in electric motorbikes based on occupation, which can be valuable information for marketing strategies. Respondents with lower incomes (<=5M) showed much higher interest in electric motorbikes compared to other income categories.

People with high incomes (>10M) show the least interest. These results provide a clear picture that there is a relationship between income level and interest in electric motorbikes, where people with lower incomes tend to be more interested. The majority of respondents who own conventional motorbikes are more interested in electric motorbikes than those who are not interested. Although the number is smaller, respondents who do not own a conventional motorbike also show higher interest in electric motorbikes compared to those who are not interested. These results illustrate that ownership of a conventional motorbike influences interest in electric motorbikes, where people who own conventional motorbikes tend to be more interested in switching to electric motorbikes.Based on the data preprocessing results, the absence of missing values indicates that this dataset has good data quality in terms of completeness. No additional steps are required to handle missing values, such as imputation or data deletion. A complete dataset with no missing values makes further analysis easier, because all the necessary data is available and intact. These results provide important information that the stage of handling missing values in data preprocessing has been completed, and the results show that there are no missing values in the dataset used. Feature encoding has succeeded in converting category data into a numerical form that can be used by machine learning algorithms.

One-Hot Encoding is used to avoid order or priority assumptions in category data, while Label Encoding is used for label targets that only have two categories (Interested/Not interested). As a result of data normalization, it can be seen that the age column has changed to a value in the range 0 to 1, while the other columns are binary columns resulting from the encoding process. The interest column remains with a value of 1 because it does not undergo a normalization process. This data sharing process is important in machine learning to ensure that the model being built can generalize well on new data that was not seen during training.From the results of the confusion matrix, the Naïve Bayes model used has poor performance in detecting positive classes (1). This model tends to be wrong in predicting the positive class more often, which can be seen from the high number of False Negatives (FN). Further evaluation is needed to improve model performance, perhaps by optimizing parameters or using different classification algorithms. Meanwhile, the KNN model used in this test has a good level of accuracy and precision. This model has a very high recall (100%), which means all positive cases were successfully detected. There was a significant number of false positives (43), meaning there were many negative cases that were incorrectly predicted as positive. This model may be suitable if more concerned with positive detection, but needs to be refined to reduce false positives if this type of error has a large negative impact.

IV. CONCLUSION

Based on the results of the research that has been carried out, it can be concluded that K-Nearest Neighbors has much higher accuracy (71.52%) than Naïve Bayes (31.13%), indicating that K-Nearest Neighbors is better at classifying data correctly overall. Naïve Bayes had higher precision (83.33%) than K-Nearest Neighbors (71.52%), indicating that when Naïve Bayes makes positive predictions, more of those predictions are correct. K-Nearest Neighbors has perfect recall (100%), which means K-Nearest Neighbors is able to detect all positive cases without missing anything. In contrast, Naïve Bayes had a very low recall (4.63%), indicating that many positive cases were not detected.

K-Nearest Neighbors had a much higher F1-score (83.40%), indicating a better balance between precision and recall than Naive Bayes (8.77%). K-Nearest Neighbors shows much better performance in terms of accuracy, recall and F1-score. This shows that K-Nearest Neighbors is more reliable in detecting people's interest in electric motorbikes and provides a better balance between detecting all positive cases and avoiding prediction errors. Naïve Bayes has higher precision, but performs very poorly in terms of recall and F1-score. This suggests that although the positive prediction is more accurate, many positive cases are not detected, which greatly reduces its reliability in this context. Overall, K-Nearest Neighbors is more recommended for use on this dataset compared to Naïve Bayes, especially if it is important to detect all positive cases.

REFERENCES

[1] A. Habibie and W. Sutopo, “A Literature Review: Commercialization Study of Electric Motorcycle Conversion in Indonesia,” **IOP Conf. Ser. Mater. Sci. Eng**., vol. 943, no. 1, 2020, doi: 10.1088/1757-899X/943/1/012048.

[2] M. Weiss, P. Dekker, A. Moro, H. Scholz, and M. K. Patel, “On the electrification of road transportation – A review of the environmental, economic, and social performance of electric two-wheelers,” **Transp. Res. Part D Transp. Environ**., vol. 41, pp. 348–366, 2015, doi: https://doi.org/10.1016/j.trd.2015.09.007.

[3] S. Marwiyah and V. M. Rengganis, “Policy Formulation on Accelerating the Development of Electric Motor Vehicles,” **Proc. Int. Conf. Democr. Natl. Resil. (ICDNR 2021)**, vol. 620, no. Icdnr, pp. 83–86, 2022, doi: 10.2991/assehr.k.211221.015.

[4] D. S. Sulistyono, Yuniaristanto, W. Sutopo, and M. Hisjam, “Proposing Electric Motorcycle Adoption-Diffusion Model in Indonesia: A System Dynamics Approach,” ***J. Optimasi Sist. Ind***., vol. 20, no. 2, pp. 83–92, 2021, doi: 10.25077/josi.v20.n2.p83-92.2021.

[5] M. N. A. Jodinesa, W. Sutopo, and R. Zakaria, “Markov chain analysis to identify the market share prediction of new technology: A case study of electric conversion motorcycle in Surakarta, Indonesia,” **AIP Conf. Proc**., vol. 2217, no. 1, p. 30062, Apr. 2020, doi: 10.1063/5.0000817.

[6] P. Kerdlap and S. H. Gheewala, “Electric Motorcycles in Thailand: A Life Cycle Perspective,” ***J. Ind. Ecol***., vol. 20, no. 6, pp. 1399–1411, Dec. 2016, doi: https://doi.org/10.1111/jiec.12406.

[7] P. Chutima and V. Tiewmapobsuk, “The Feasibility Study on the Infrastructure of Swapping Battery Station for Electric Motorcycles in Thailand,” **IOP Conf. Ser. Earth Environ. Sci**., vol. 811, no. 1, 2021, doi: 10.1088/1755-1315/811/1/012014.

[8] L. Zhu, Q. Song, N. Sheng, and X. Zhou, “Exploring the determinants of consumers’ WTB and WTP for electric motorcycles using CVM method in Macau,” **Energy Policy**, vol. 127, pp. 64–72, 2019, doi: https://doi.org/10.1016/j.enpol.2018.12.004.

[9] N. D. Andini, S. Z. Harahap, and M. Nasution, "Penerapan Machine Learning Algoritma Regresi Linear untuk Memprediksi Sahan Bank BNI," **INFORMATIKA**, vol. 12, no. 2, pp. 162–169, 2024, doi: https://doi.org/10.36987/informatika.v12i2.5649.
[10] M. Munif, M. Mustain, and K. Yahya, "Analysis of the K-Nearest Neighbor Algorithm to Determine the Prediction of Tofu Production," ***Appl. Technol. Comput. Sci. J***., vol. 5, no. 1, pp. 57–64, 2022, doi: 10.33086/atcsj.v5i1.3677.
[11] S. A. Afolalu, B. B. Matthew, S. O. Ongbali, A. Abdulkareem, M. E. Emetere, and O. U. Iheanetu, "Overview impact of planning in production of a manufacturing sector," **IOP Conf. Ser. Mater. Sci. Eng**., vol. 1036, no. 1, p. 012060, 2021, doi: 10.1088/1757-899x/1036/1/012060.
[12] S. A. Afolalu *et al.*, "The Role of Production Planning in Enhancing an Efficient Manufacturing System - An Overview," **E3S Web Conf**., vol. 309, 2021, doi: 10.1051/e3sconf/202130901002.
[13] P.-S. Chen, F.-Y. Tseng, and J. C.-M. Chen, "Determining production capacity planning in the stochastic demand environment for the food processing manufacturers: A case study," J. Inf. Optim. Sci., vol. 42, no. 5, pp. 909–936, Jul. 2021, doi: 10.1080/02522667.2020.1800789.
[14] S. E. Dincer and E. Demir, "Determining the faulty and refund products in manufacturing system: application on a textile Firm," **Pressacademia**, vol. 7, no. 3, pp. 178–187, 2020, doi: 10.17261/pressacademia.2020.1281.
[15] X. Zhang, Y. Hu, J. Deng, H. Xu, and H. Wen, "Feature Engineering and Artificial Intelligence-Supported Approaches Used for Electric Powertrain Fault Diagnosis: A Review," **IEEE Access**, vol. 10, pp. 29069–29088, 2022, doi: 10.1109/ACCESS.2022.3157820.
[16] P. Ndikum, "Machine Learning Algorithms for Financial Asset Price Forecasting," pp. 1–16, 2020, [Online]. Available: **http://arxiv.org/abs/2004.01504**
[17] I. Kazancoglu, M. Ozbiltekin-Pala, S. K. Mangla, A. Kumar, and Y. Kazancoglu, "Using emerging technologies to improve the sustainability and resilience of supply chains in a fuzzy environment in the context of COVID-19," **Ann. Oper. Res**., vol. 322, no. 1, pp. 217–240, 2023, doi: 10.1007/s10479-022-04775-4.
[18] D. Proserpio *et al.*, "Soul and machine (learning)," **Mark. Lett**., vol. 31, no. 4, pp. 393–404, 2020, doi: 10.1007/s11002-020-09538-4.
[19] L. Ma and B. Sun, "Machine learning and AI in marketing – Connecting computing power to human insights," ***Int. J. Res. Mark***., vol. 37, no. 3, pp. 481–504, 2020, doi: https://doi.org/10.1016/j.ijresmar.2020.04.005.
[20] N. S. Pearre, W. Kempton, R. L. Guensler, and V. V Elango, "Electric vehicles: How much range is required for a day's driving?," **Transp. Res. Part C Emerg. Technol**., vol. 19, no. 6, pp. 1171–1184, 2011, doi: https://doi.org/10.1016/j.trc.2010.12.010.
[21] S. Modi, J. Bhattacharya, and P. Basak, "Convolutional neural network–bagged decision tree: a hybrid approach to reduce electric vehicle's driver's range anxiety by estimating energy consumption in real-time," **Soft Comput**., vol. 25, no. 3, pp. 2399–2416, 2021, doi: 10.1007/s00500-020-05310-y.
[22] M. W. Dela Utami, Yuniaristanto, and W. Sutopo, "Adoption Intention Model of Electric Vehicle in Indonesia," ***J. Optimasi Sist. Ind***., vol. 19, no. 1, pp. 70–81, 2020, doi: 10.25077/josi.v19.n1.p70-81.2020.
[23] S. K. Dixit and A. K. Singh, "Predicting Electric Vehicle (EV) Buyers in India: A Machine Learning Approach," **Rev. Socionetwork Strateg**., vol. 16, no. 2, pp. 221–238, 2022, doi: 10.1007/s12626-022-00109-9.
[24] S. Afandizadeh, D. Sharifi, N. Kalantari, and H. Mirzahossein, "Using machine learning methods to predict electric vehicles penetration in the automotive market," **Sci. Rep**., vol. 13, no. 1, p. 8345, 2023, doi: 10.1038/s41598-023-35366-3.
[25] S. Suryani, M. F. Fayyad, D. T. Savra, V. Kurniawan, and B. H. Estanto, "Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm," ***Public Res. J. Eng. Data Technol. Comput. Sci***., vol. 1, no. 1, pp. 1–9, 2023, doi: 10.57152/predatecs.v1i1.814.
[26] V. M. R., R. Ashok, and L. Nitha, "Electric Vehicles Acceptance and Knowledge Identification in India using Naive Bayes and k-Nearest Neighbor Classifiers," ***Int. J. Innov. Technol. Explor. Eng***., vol. 9, no. 5, pp. 1630–1633, 2020, doi: 10.35940/ijitee.e3008.039520.