Research Article

Prediction of New Customer Segmentation Classification Using Artificial Intelligence Project Cycle Orange Data Mining

1,*Fransiska Febriyanti Kosat, 2Yasinta Oktaviana Legu Rema, 3Hevi Herlina Ullu ⬤iD

1,2,3 Department of Information Technology, Universitas Timor (Unimor), North Central Timor District - East Nusa Tenggara Province, Indonesia

* Corresponding Author: febriyantikosat@gmail.com

Abstract: This research aims to predict the right segmentation group or classification of new customers to become a classification comparison data carried out by the sales team to determine the strategy used to enter the market, whether it can be said to be feasible or not. This article discusses the basis of the method used, i.e., Machine Learning, discussed in detail about Artificial Intelligence (AI). Also discusses what is Classification, Segmentation, Data Mining, Neural Networks, Naive Bayes, Decision Trees, Random Forest (RF), and Support Vector Machine (SVM). This article discusses comprehensively the method used, and the development of Modeling, in the results and analysis section, comprehensively shows the prediction analysis of new customer segmentation classification, algorithm performance results of several methods, and distributions analysis. With the percentage prediction of new potential customer segmentation using the Neural Network method, the percentage prediction of Segmentation A is 25.21%, the percentage prediction of Segmentation B is 21.77%, the percentage prediction of Segmentation C is 23.49%, the percentage prediction of Segmentation D is 29.53%. The percentage of segmentation that has been calculated by the company is the percentage of Segmentation A of 32.13%, the percentage of Segmentation B of 20.89%, the percentage of Segmentation C of 17.69%, and the percentage of prediction of Segmentation D of 29.29%.

Keywords: Artificial intelligence; Machine Learning; Classification; Segmentation; Prediction; Data Mining.

1. Introduction

Currently, the marketing system has shifted a lot with the current technological developments, many business systems are changing towards technology that continues to change day by day, the sales system already uses a lot of internet-based sales platforms and also a very dynamic website platform on a mobile basis. Companies cannot stay silent with only conventional systems. They must be able to move in harmony with technological developments.

The goal is the addition of customers. The spread of customers or the expansion of the diversity of buyers is not only regional or local but also international markets. Marketing strategies need to be changed towards a more dynamic direction and can confidently reach many customers to increase sales. A car company wanted to expand into a new market. After conducting more in-depth research to understand the needs of its customers, the company's sales team concluded that the needs of customers in the new location were similar to the company's old marketing. From the old marketing, the company's sales team had classified all the old customers into four segments based on their similarities. And the sales team has a plan to use the same strategy for the new marketing with the new customers. So the author is asked to help predict the exact segmentation group/classification of new customers so that it can be compared with the

classification that has been done by the sales team to determine whether the strategy can be used when entering a new market or not.

New methods need to be applied in achieving the goal of increasing customers and also predicting to be able to get new customers, including Machine Learning, Artificial Intelligence (AI), Classification, Segmentation, Data Mining, Neural Networks, Naive Bayes, Decision Tree, Random Forest (RF), and Support Vector Machine (SVM). The details of these methods will be discussed in this research.

2. Theory

2.1 Machine Learning

According to Shalev-Shwartz and Ben-David, Machine Learning is the study of algorithms to learn how to perform certain tasks that are performed automatically by people. Machine learning is a part of artificial intelligence (AI) that helps computers or teaching machines learn from previous data and make intelligent decisions (Fathurohman, 2021). Machine Learning is part of the field of artificial intelligence that focuses on applying specialized algorithms and methods for prediction, pattern recognition, and classification (Muntiari & Hanif, 2022).

Based on some of the above definitions, it can be defined that machine learning is a branch of artificial intelligence (AI) that allows machines to learn independently without instructions or directions from users.

Machine Learning is divided into two kinds of learning concepts:

- Supervised Learning

  Supervised Learning is a machine learning technique that creates a function based on existing training data, in this case, it can be said that for this technique there is already detailed and well-verified training data available which will be used as a data model when the test process is carried out with new test data and produces output results that are as expected before based on existing training data.

- Unsupervised learning

  Unsupervised learning is a machine learning technique that tries to represent the pattern of an input that comes from training data and one of the differences with supervised learning is that there is no classification of the input data (Fathurohman, 2021).

2.2 Artificial Intelligence (AI)

Artificial Intelligence is a computer program that has algorithms that function to be able to learn data and use it to be able to carry out the process of thinking and acting like humans (Supriyadi & Asih, 2020). Artificial intelligence better known as artificial intelligence (AI) is a "machine" that can do various things that are considered to require intelligence when humans operate them (Disemadi, 2021).

Artificial intelligence (AI) is a study that includes computational procedures to perform actions that humans perform and require a certain amount of intelligence. Communication and learning are ways in which humans demonstrate their intelligence. Both ways require mental skills such as synthesis, analysis, decision-making, judgment, organization, reasoning, hypothesis, interpretation, and others. AI involves knowledge representation, search, perception, and inference. To produce results, AI must define and read data in a form that will enable representation and processing (Abyan Ramadhana & Hussein. Ananda Sabil, 2022). It can be concluded that artificial intelligence (AI) is a field of technology that studies and develops computer systems to perform human-like actions and intelligence.

Moreover, the benefits and goals of creating AI are as follows:

- a. Benefits of Artificial Intelligence (AI)

  There are several benefits of AI created:

- Available all the time or 24 hours.
- Save energy and time.
- Minimizes human error.
- Supporting the learning process at school or university.

b. Purpose of Artificial Intelligence (AI)
There are also several purposes for which AI was created:
- Identifying potential threats and preventing cyber attacks.
- Managing large amounts of data quickly and accurately.
- Assist humans in doing their daily tasks.
- Increase efficiency in various industries or companies.

2.3 Classification

Classification is the job of assessing data objects to put them into certain classes from several available classes (Putro et al., 2020). Meanwhile, according to (Febriani & Sulistiani, 2021) Classification is the process of finding a model or function that explains or distinguishes concepts or data classes, intending to be able to estimate the class of an object whose label is unknown. So classification is a term used to categorize something. More specifically, classification can be interpreted as the process of classifying an object based on certain characteristics and criteria as well as similarities and differences. For example, when classifying waste based on the type of material, it is divided into organic waste and inorganic waste.

2.4 Segmentation

Market segmentation is a way for business managers to look at the market creatively. Segmenting the market into groups of buyers according to the types of products being marketed (Laia et al., 2021).
Market segmentation is divided into several groups, among others:
a. Geographic Segmentation
Segmentation based on geography is usually used in market segmentation because considering the market potential for a company's products can be influenced by market location where the operating cost factor and the amount of demand from each region are different. e.g., countries, states, regions, provinces, districts, and cities.
b. Demographic Segmentation
Demographic segmentation is a market divided into groups based on demographic variables such as age, family size, family life cycle, gender, income, occupation, religion, race, nationality generation, and social class.
c. Psychographic Segments
The psychographic segment turns out to be able to influence the improvement of sales performance so that the rise and fall of a product marketed by the company can be determined by this. With this segment, it is easier for companies to see the response of consumers who are in certain demographic segments to the marketing stimuli of a product that will or has been marketed (Luis et al., 2019).

2.5 Data Mining

Data mining is the process of discovering meaningful correlations, patterns, and new trends by sifting through large amounts of data stored in repositories, using pattern recognition technology and statistical and mathematical techniques. Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. The term data mining has several views, such as knowledge discovery or pattern recognition. The term knowledge discovery is appropriate because the main purpose of data mining is to obtain knowledge that is still hidden in chunks of data. Meanwhile, the term for pattern recognition or pattern recognition is appropriate to use because it is to find patterns hidden in chunks of data (Nabila et al., 2021).
Meanwhile, according to (Nugraha et al., 2022) data mining is a term used to find knowledge hidden in the database. So it can be concluded that data mining is a semi-automated process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful potential and useful knowledge information stored in large databases.

2.6 Data Mining

Neural Network is a branch of artificial intelligence, which applies the ability of the human brain to receive information and the information is processed to produce an output. Neural networks can be used for pattern classification and prediction (Prasetyo et al., 2022). Meanwhile, according to (Almais et al., 2022) the learning machine in the Artificial Neural Network model used is an aspect of learning that mimics an experience to be able to predict future events. The neural network method is an information processing system whose characteristics are similar to the way the human brain works. The neural network architecture is determined by three layers, namely the input layer, hidden layer, and output layer. Neural networks are divided into two models:

a. Artificial Neural Network (ANN)
    Artificial Neural Network (ANN) is an information processing method that mimics the way the biological nervous system works, especially humans. Neural networks work by breaking down a particular problem and then forming several small structured instructions. These instructions will be converted into computer programs and then into machine code that will be used to detect more complex trends.
b. Convolutional Neural Network (CNN)
    Convolutional Neural Network (CNN) is a machine learning technique developed from Multi-Layer Perceptron (MLP) designed to process two-dimensional data. The way CNN works is similar to MLP, but in CNN each neuron is represented in two dimensions, while in MLP each neuron has only one dimension (Prasetyo et al., 2022).

Some of the stages in neural networks are:
a. The input layer is the value of the amount of data entered for learning. The value is no more than the number of variables or the amount of data.
b. The hidden layer is the neurons in the network. If the amount of data is insufficient, it causes underfitting, which makes the network less able to recognize signals and patterns in the data set. When the amount is excessive, overfitting occurs, which means there is not enough training dataset to train all the neurons. Each best value is used to determine the learning rate, momentum, training cycle, and input layer.
c. The output layer is how many outputs are from the neural network algorithm learning process (Safitri et al., 2023).

2.7 Naïve Bayes

Naive Bayes is one method that can be used to classify data. Bayesian classification is a statistical classification that can be used to predict the probability of class membership (Handoko & Neneng, 2021). Meanwhile, according to (Dewi et al., 2021) naive Bayes classifier is a problem-solving algorithm that is included in the classification method in data mining. Naive Bayes Classifier adopts statistical science, namely by using probability theory to solve a supervised learning case, meaning that in the data set, there is a label, class, or target as a reference or desert.

So it can be concluded that naive Bayes is a suitable technique for binary and multiclass classification. This method includes data classification methods based on probabilities that can occur in the future and this method also includes supervised learning which is used to distinguish objects based on certain features.

The steps of the naïve Bayes method are as follows:
a. Data collection.
b. Data preprocessing: clean the data, handle missing values, and transform features if necessary.
c. Probability calculation:
    *   Calculate the prior probability for each class.
    *   Calculate the conditional probability of each feature given a class.

d. Application of Bayes theorem: use Bayes theorem to calculate the highest posterior probability as prediction.
e. Classification: determine the class with the highest posterior probability as the prediction.
f. Model evaluation: use metrics such as precision, ACU, CA, F1, and recall to evaluate model performance.

2.8 Decision tree

The decision tree is a commonly used algorithm for decision-making. Decision Tree will find a solution to the problem by looking for criteria as interconnected nodes forming a tree-like structure. Decision tree is a prediction model for a decision using a hierarchical or tree structure. Each tree has branches, branches have an attribute that must be met to go to the next branch until it ends at the leaf. The concept of data in a decision tree is data expressed in the form of a table consisting of attributes and records (Pamuji & Ramadhan, 2021).

Meanwhile, according to (Mu'Alim & Hidayayati, 2022) a decision tree is a flow chart that is shaped like a tree structure in which each internal node states a test of an attribute, each branch states the output of the test and the leaf node states the classes. The topmost node, referred to as the root node, has several outgoing edges, whereas a leaf node will only have one incoming edge with no outgoing edges. Leaf nodes are the final result to represent the class label of the attribute combination that becomes the rule. So in conclusion, a decision tree is a machine-learning algorithm used for classification tasks. It creates a mode in the form of a tree-like structure that predicts a target value by organizing a set of decision rules based on input features.

Here are the steps in creating a decision tree:
a. Calculating the *gini impurity* value of each feature
b. Calculating the *gini impurity* value of each split condition
c. Calculating the average *gini impurity* value of each split condition
d. Calculating the *information gain* value of each split condition
e. Finding the highest *information gain* value to be the root
f. Repeat steps 2-4 up to the *tree* (Larasati et al., 2022).

2.9 Random Forest (RF)

Random Forest (RF) is an algorithm that uses a recursive binary splitting method to reach the final node in a tree structure based on classification and regression trees. The random forest algorithm shows several advantages including being able to produce relatively low errors, good performance in classification, can handle large amounts of training data efficiently, and is an effective method for estimating missing data. The random forest produces many independent trees with subsets randomly selected through bootstrapping from training samples and input variables at each node (Pamuji & Ramadhan, 2021). Meanwhile, random forest (RF) is a classification consisting of several decision trees. Each decision tree is created using a random vector. In general, random vectors are inserted in tree formation by selecting a random F value, such as F input attributes (features) to be shared at each node in the decision tree to be formed. By choosing a random value of F, it does not have to check all the existing attributes and see the F value of the selected attribute (Mu'Alim & Hidayati, 2022). Finally, Random forest (RF) is a machine-learning algorithm used for classification and regression tasks. Random forests work by combining a large number of relatively simple decision trees to produce more accurate decisions. Each decision tree is randomly generated from training data. This process is done to reduce overfitting and improve the generalization ability of the model.

The steps in the random forest method are as follows:
a. Data collection.
b. Data pre-processing: clean the data and encode the category variables if necessary.
c. Model building:
    - Decision tree building: build multiple decision trees on randomly selected subsets of training data with replacement (bootstrapping).
    - Random features: at each tree split node, select a random subset of features to determine the best split.
d. Aggregation: combine the predictions from all trees (via majority voting for classification).
e. Testing: test the model on the test set.
f. Evaluation: assess the model using metrics such as precision, ACU, CA, F1, and recall.

2.10 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning algorithm that applies a hyperplane function to the data so that regions of each class are formed. Hyperplane itself is a function that is used as a separator between existing classes. In predicting a class of data, SVM will label it based on which class area the data belongs to (Fikri et al., 2020). Meanwhile, according to (Amrozi et al., 2022) the support vector machine (SVM) method is a classification method in problem solving for two classes. SVM is usually used in classifying data that can be linearly separated, for data that cannot be linearly separated, a kernel function is used whose purpose is to map input data to feature data.

SVM is a machine learning method that aims to find the best hyperplane that separates two classes in the input space. The SVM classification algorithm uses training data to form a classification model, the model formed is used as a prediction of new data classes that have never existed before called testing data (Suryati et al., 2023). So it is concluded that SVM is one of the methods of machine learning that is included in supervised learning. This method is used for classification or grouping and analyzing data into one of two categories.

The steps of the support vector machine (SVM) method are as follows:
a. Data collection.
b. Data pre-processing: clean the data and perform feature normalization or standardization.
c. Model building:
    - Kernel function: choose a kernel function (linear, polynomial, RBF, etc.)
    - Margin optimization: find the hyperplane that maximizes the margin between different classes.
    - Regularization parameter: determine the parameter C that controls the trade-off between margin maximization and classification error.
d. Training: train the SVM model using the training data.
e. Testing: test the model on the test set.
f. Evaluation: score the model using metrics such as precision, ACU, CA, F1, and recall.

3. Method

There are several steps used for modeling development starting from data collection, data cleaning, modeling, model evaluation, and model presentation which can be seen in Figure 1.
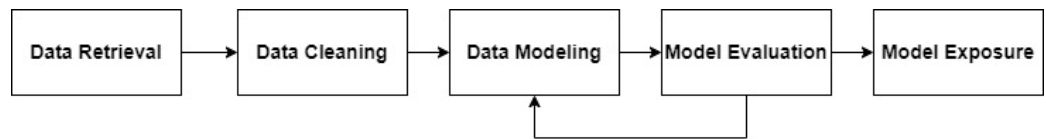

Figure 1. Modeling Development Steps

The following is an explanation of the modeling development steps:

a. Data Retrieval

At this stage, the author takes a dataset for modeling needs, with the topic "Customer Segmentation Classification". This dataset consists of two data, namely Train data and Test data. This train data contains 8,068 old customer data and test data contains 2,627 new customer data, with 11 features including ID, Gender, Ever_Married, Age, Graduated, Profession, Work_Experience, Spending_Score, Family_Size, Var_1, Segmentation.

b. Data Cleaning

After data collection, at the data cleaning or preprocessing stage, it will clean the data stored more than one data (Double), using the Preprocessing Widget for new customer data (test data), test data which was originally 2,627 data becomes 2,154 data and train data (old customer data) is not preprocessed because this data is only used for classification methods that will be used to predict new customer segmentation.

c. Data Modeling

After cleaning the data, the next step is modeling. Modeling here means the appearance of the interface, the classification results of the five methods the evaluation results using the confusion Matrix, and the results of predicting new customer segmentation.

d. Model Evaluation

At this stage, evaluation is needed to find out whether the performance of this modeling is following what is expected or not. If it is appropriate then, the modeling is ready to be presented or presented. If not, repeat the modeling stage.

e. Model Presentation

Modeling that is appropriate and approved by the mentor is ready to be presented or presented to other audiences.

4. Results and Analysis

The results of the interface created using the Orange3 application are as follows the Figure 2. Figure 2 is a display that contains the performance results of algorithms from several methods and the results of predicting the classification of new customer segmentation.
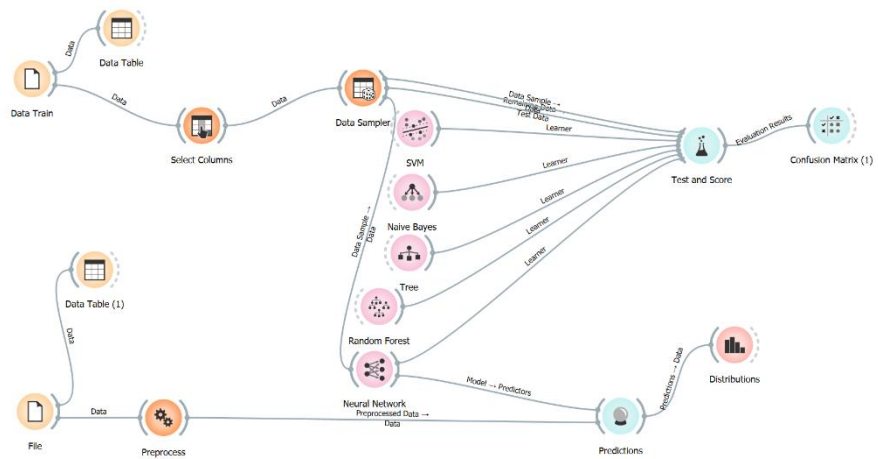
Figure 2. Orange3 application: Algorithm performance results of several methods and prediction results of new customer segmentation classification
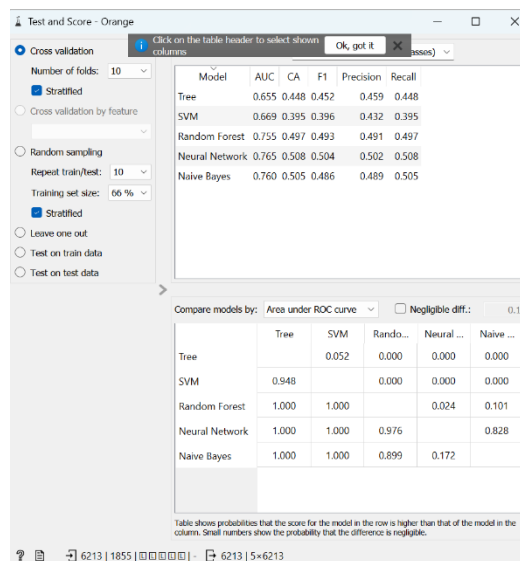
Moreover, Figure 3 is the result of the performance of five methods to get the highest value of AUC, CA, F1, Precision, and Recall. Judging from the picture below, it turns out that the Neural Network method has an AUC, CA, F1, Precision, and Recall value greater than the other four methods, namely AUC of 0.765, CA of 0.508, F1 of 0.504, Precision of 0.502 and Recall 0.508. With a greater value, the Neural Network method is the method used to predict the classification of new customer segmentation.

Figure 3. The result of the performance of five methods to get the highest value of AUC, CA, F1, Precision, and Recall

Furthermore, this is the result of predicting the segmentation of new potential customers using the Neural Network method and comparing it with the segmentation of the company.

Figure 4. Customer Segmentation Classification Prediction Results

Furthermore, the distribution here shows the number of predictions of new potential customers to segmentation A, segmentation B, segmentation C, and segmentation D using the Neural Network method. The number of new potential customers entering Segmentation A prediction is 543 customers, the number of new potential customers entering Segmentation B prediction is 469 customers, the number of new potential customers entering Segmentation C predictions is 506, and the number of new potential customers entering segmentation D predictions is 636 customers.
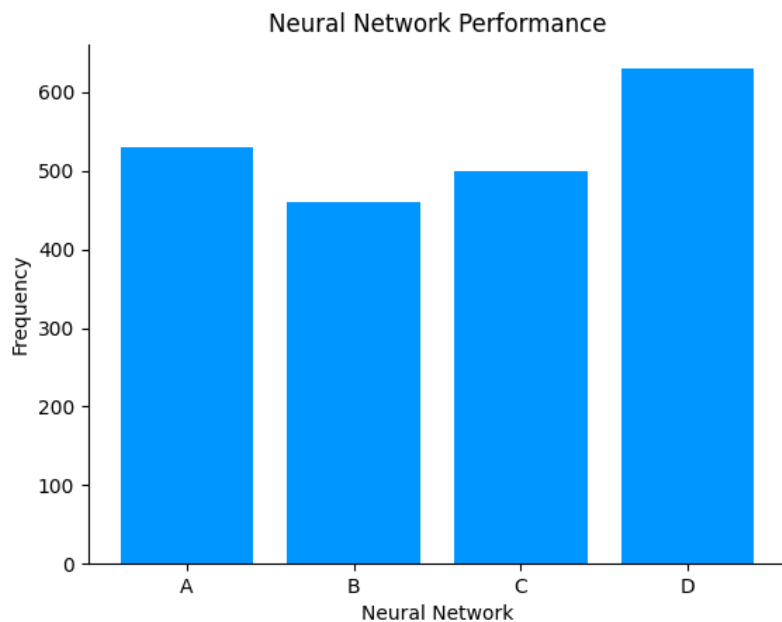

Figure 5. Neural Network Performance

As for the segmentation that has been carried out by the company is as follows, the number of new customers entering segmentation A is 692 customers, the number of new customers entering segmentation B is 450 customers, the number of new customers entering segmentation C predictions is 381 customers, the number of new customers entering segmentation D predictions is 631 customers.
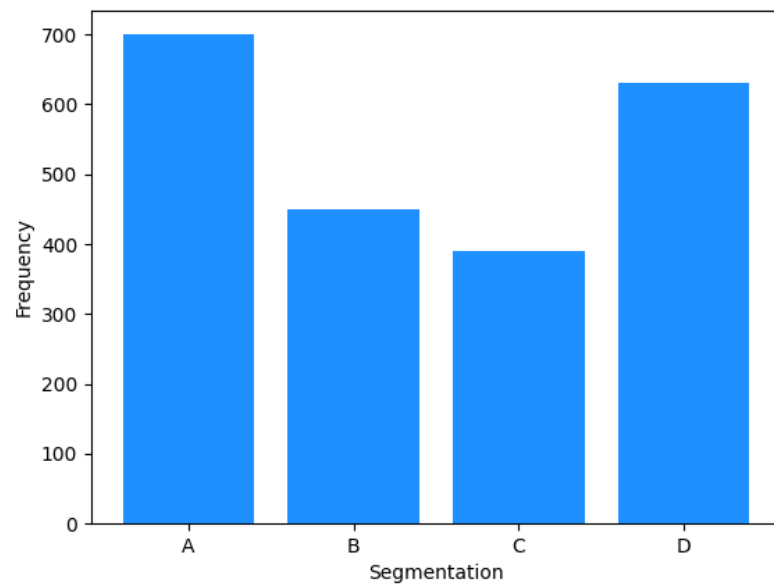
Figure 6. Segmentation

After being seen from the distributions, the prediction of new potential customer segmentation using the Neural Network method is not much different from the segmentation that has been carried out by the company. With the percentage prediction of new potential customer segmentation using the Neural Network method, the percentage prediction of Segmentation A is 25.21%, the percentage prediction of Segmentation B is 21.77%, the percentage prediction of Segmentation C is 23.49%, the percentage prediction of Segmentation D is 29.53%. The percentage of segmentation that has been calculated by the company is the percentage of Segmentation A of 32.13%, the percentage of Segmentation B of 20.89%, the percentage of Segmentation C of 17.69%, and the percentage of prediction of Segmentation D of 29.29%.

4. Conclusion
Based on research on the prediction of car customer segmentation classification, it can be concluded that The resulting interface contains the performance of the algorithm, which is the best method that will be used to predict the classification of new customers. The Neural Network method is considered more suitable for predicting new car customer segmentation, this is because the Neural network method has an AUC of 0.765, CA of 0.508, F1 of 0.504, precession of 0.502, and Recall of 0.508, higher than the SVM (Support Vector Machine), Naive Bayes, decision Tree, and Random Forest methods which are considered less suitable for predicting new car customer segmentation. With the prediction results, segmentation A is 25.21% or 543 customers, segmentation B is 21.77% or 469 customers, Segmentation C is 23.49% or 506 customers, and segmentation D is 29.53% or 636 customers.

5. Suggestion
The following are some of the author's suggestions for consideration in future similar research, It is expected to use other methods, for classification and get a better method, than the current method used to predict the segmentation of car customers. In future research, it is hoped that the prediction of new car customer segmentation classifications using the AI project cycle using the orange3 application can be developed or upgraded using the Jupyter Notebook application.

Availability of data and Materials: All data are available from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Additional Information: No Additional Information from the authors.

References

1. Abyan Ramadhana, M., & Hussein.Ananda Sabil. (2022). Pengaruh Penggunaan Artificial Intelligence Terhadap Customer Loyalty Dengan Citra Merek Sebagai Variabel Mediasi. Manajemen Pemasaran Dan Perilaku Konsumen, 1(4), 522–531. http://dx.doi.org/10.21776/jmppk.2022.01.4.11.

2. Almais, A. T. W., Crysdian, C., Holle, K. F. H ., & Roihan, A. (2022). Smart Assessment Menggunakan Backpropagation Neural Network. Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer, 21(3), 525-538.

3. Amrozi, Y., Yuliati, D., Susilo, A., Novianto, N., & Ramadhan, R. (2022). Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM. Jurnal Sisfokom (Sistem Informasi Dan Komputer), 11(3), 394–399. https://doi.org/10.32736/sisfokom.v11i3.1502

4. Ardi, R. B., Nastiti, F. E., & Sumarlinda, S. (2023). Algoritma K-Means Clustering Untuk Segmentasi Pelanggan (Studi Kasus : Fashion Viral Solo). 9(1), 124–131.

5. Dewi, P. S., Sastradipraja, C. K., & Gustian, D. (2021). Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naïve Bayes Classifier. Jurnal Teknologi Dan Informasi, 11(1), 66–80. https://doi.org/10.34010/jati.v11i1.3593

6. Dhery, H., Assyam, A., & Hasan, F. N. (2023). Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining. 4(1), 341–349. https://doi.org/10.30865/klik.v4i1.957

7. Disemadi, H. S. (2021). Urgensi Regulasi Khusus dan Pemanfaatan Artificial Intelligence dalam Mewujudkan Perlindungan Data Pribadi di Indonesia. Jurnal Wawasan Yuridika, 5(2), 177. https://doi.org/10.25072/jwy.v5i2.460

8. Fathurohman, A. (2021). Machine Learning Untuk Pendidikan: Mengapa Dan Bagaimana. Jurnal Informatika Dan Teknologi Komputer (JITEK), 1(3), 57–62. https://journal.amikveteran.ac.id/index.php/jitek/article/view/306

9. Febriani, S., & Sulistiani, H. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5. 89Jurnal Teknologi Dan Sistem Informasi (JTSI), 2(4), 89–95.

10. Fikri, M. I., Sabrila, T. S., Azhar, Y., & Malang, U. M. (2020). Comparison of the Naïve Bayes Method and Support Vector Machine on Twitter Sentiment Analysis. SMATIKA Jurnal: STIKI Informatika Jurnal, 10(2), 71–76.

11. Handoko, M. R., & Neneng. (2021). Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web. Jurnal Teknologi Dan Sistem Informasi (JTSI), 2(1), 50–58. http://jim.teknokrat.ac.id/index.php/JTSI

12. Hikmah, I. R., & Yasa, R. N. (2021). Perbandingan Hasil Prediksi Diagnosis pada Indian Liver Patient Dataset (ILPD) dengan Teknik Supervised Learning Menggunakan Software Orange. Jurnal Telematika, 16(2), 69–76.

13. Ibrohim, M., Selvia Lauryn, M., & Dhanan Jaya, R. (2019). Rancang Bangun Sistem Kehadiran Karyawan Berbasis Radio Frequency Identification (Rfid). Jurnal PROSISKO, 6(1), 43–52.

14. Laia, S., Dakhi, Y., & Dakhi, P. (2021). Pengaruh Segmentasi Pasar Terhadap Volume Penjualan di UD.LIS Teluk Dalam Kabupaten Nias Selatan. PARETO: Jurnal Riset Manajemen Dan Bisnis, 6(2), 36–46.

15. Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest. … Teknologi Informasi Dan …, 6(9), 4305–4313. http://j-ptiik.ub.ac.id

16. Lemantara, J. (2022). Penerapan Algoritma Naïve Bayes dan ID3 untuk Memprediksi Segmentasi Pelanggan pada Penjualan Mobil. 4(1), 31–40. https://doi.org/10.37802/joti.v4i1.265

17. Luis, P. C. U., Mandey, S., & Jorie, R. J. (2019). Analisis Pengaruh Strategi Segmentasi Pasar Terhadap Peningkatan Kinerja Penjualan Pada Pt. Astragraphia Tbk Cabang Manado. Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi, 7(2), 2581–2590.

18. Mu'Alim, F., & Hiday, R. (2022). Implementasi Metode Random Forest Untuk Penjurusan Siswa Di Madrasah Aliyah Negeri Sintang. Jupiter, 14(1), 116–125. https://www.neliti.com/publications/441871/implementasi-metode-random-forest-untuk-penjurusan-siswa-di-madrasah-aliyah-nege#cite

19. Muntiari, N. R., & Hanif, K. H. (2022). Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning. 3(1), 1–6.

20. Nabila, Z., Isnain, A. R., Permata, & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. Jurnal Teknologi Dan Sistem Informasi (JTSI), 2(2), 100. http://jim.teknokrat.ac.id/index.php/JTSI

21. Nugraha, A., Nurdiawan, O., & Dwilestari, G. (2022). Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport. JATI (Jurnal Mahasiswa Teknik Informatika), 6(2), 849–855. https://doi.org/10.36040/jati.v6i2.5755

22. Pamuji, F. Y., & Ramadhan, V. P. (2021). Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotheraphy. Jurnal Teknologi Dan Manajemen Informatika, 7(1), 46–50. https://doi.org/10.26905/jtmi.v7i1.5982

23. Prasetyo, V. R., Mercifia, M., Averina, A., Sunyoto, L., & Budiarjo, B. (2022). Prediksi Rating Film Pada Website Imdb Menggunakan Metode Neural Network. Network Engineering Research Operation, 7(1), 1. https://doi.org/10.21107/nero.v7i1.268

24. Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN), 8(2). https://doi.org/10.30646/tikomsin.v8i2.500

25. Reliovani, R., Nadia, N., Husein, S., Abdurrafi, K. Z., Husni, C. R. A., Khowarizmi, M. A. (2021). Algoritma Mean Shift untuk Menentukan Segmentasi Pelanggan pada Penjualan Toko Online. 3, 1–7.

26. Rosaly, R., & Prasetyo, A. (2020). Flowchart Beserta Fungsi dan Simbol-Simbol. Journal of Chemical Information and Modeling, 2(3), 5–7.

27. Safitri, D., Hilabi, S. S., & Nurapriani, F. (2023). Analisis Penggunaan Algoritma Klasifikasi Dalam Prediksi Kelulusan Menggunakan Orange Data Mining. Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab, 8(1), 75–81. https://doi.org/10.36341/rabit.v8i1.3009

28. Silitonga, P. D. ., & Purba, D. E. R. (2021). Implementasi System Development Life Cycle Pada Rancang Bangun Sistem. Jurnal Sistem Informasi Kaputama (JSIK), 5(2), 196–203.

29. Supriyadi, E. I., & Asih, D. B. (2020). Artificial intelligence (AI). 2(2), 12–23.

30. Suryati, E., Styawati, & Ari Aldino, A. (2023). Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM). Jurnal Teknologi Dan Sistem Informasi, 4(1), 96–106.

31. Tuasamu, Z., Lewaru, N. A. I. M., Idris, M. R., Syafaat, A. B. N., Faradilla, F., Fadlan, M., Nadiva, P., & Efendi, R. (2023). Analisis Sistem Informasi Akuntansi Siklus Pendapatan Menggunakan DFD Dan Flowchart Pada Bisnis Porobico. Jurnal Bisnis dan Manajemen(JURBISMAN), 1(2), 495-510.

32. Wibowo, A., & Handoko, A. R. (2020). Metode Data Mining Klasterisasi Dengan Analisis Recency Frequency Monetary (RFM) Termodifikasi Segmentation Of Customers Of Drug Pharmaceutical Product Retail Using Clasterization Mining Data Method Using Modified Monetary Recency Frequency (RFM) Anal. 7(3), 573–580. https://doi.org/10.25126/jtiik.202072925

33. Widyanto, A., & Witanti, A. (2020). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global). 204–215.