Classification Auction Motorcycle and Car In South Jakarta District Attorney Using Naïve Bayes

1st Fransisca Kristina Mega
Department of Informatics Engineering
Universitas Trilogi
Jakarta, Indonesia
kristinamega@trilogi.ac.id

2nd Yaddarabullah
Department of Informatics Engineering
Universitas Trilogi
Jakarta, Indonesia
yaddarabullah@trilogi.ac.id

3rd Silvester Dian Handy Permana
Department of Informatics Engineering
Universitas Trilogi
Jakarta, Indonesia
handy@trilogi.ac.id

2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT) | 978-1-6654-3434-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/COMNETSAT53002.2021.9530827

***Abstract*—The Prosecutor's Office, as an institution, plays a role in enforcing the law, particularly in the field of prosecution. Goods used or resulting from a criminal act will be confiscated and then confiscated by the state. As a result of goods obtained as a result of a criminal act, said goods are to be auctioned off. According to the Minister of Finance Regulation Number 27 / PMK.06 / 2016 regarding auction implementation instructions, an auction is an open sale of goods to the public with the highest price bid in writing or orally, and it can be done online or offline. The South Jakarta District Prosecutor's Office must determine the grade of the items being auctioned. However, determining the quality to be categorized took a month. As a result, research is required to categorize the quality of the items being auctioned. The need to classify auctions, particularly motorcycles and cars, to reduce the time spent classifying and classifying vehicles. The standard condition is used to classify auction items. This grade applies to the categories of an engine, exterior, and interior car, vehicle year, letter in the form of SNTK, BPKB, and vehicle keys. The Naive Bayes method is used in this study to classify auction items at the South Jakarta District Prosecutor's Office. The accuracy value of this study is 74%, 44% precision, 50% recall, and 44% f1-score for motor vehicles and 57%, 44% precision, 50% recall, and 44% f1-score for car vehicles.**

Keywords— Naïve Bayes, Classification, Data Mining

I. INTRODUCTION

The process of buying and selling goods open to the public with the highest bid price is known as auctioning. Auctioning can be done directly or through electronic media (online). According to Minister of Finance Regulation Number 27/PMK.06/2016 regarding auction implementation instructions, an auction is an open sale of goods to the public with a written or oral price offer. The auction was held 282,441 times between 2015 until 2019, with a total principal value of auction transactions of Rp. 85.9 trillion. This demonstrates that there has been an increase in auction transactions recorded at the Indonesian Ministry of Finance over the last five years, from 2015 to 2019 [1]. Previous research, "Model Information System of Auction Minutes at the Office of State Wealth Services and Auction in Cirebon City with the C45 Algorithm Clustering Method," only created an auction information system for grouping specific criteria. This study's sole goal is to make the process of finding information easier. In this study, the reporting is still done in Microsoft Excel, so processing the data to obtain auction information takes a long time [2]. Several other studies, including "Combination of K-NN Algorithm and Manhattan Distance to Determine Auction Winners," discussed the feasibility of selecting the winner [3]. According to the study, online group buying has returned as a result of businesses taking advantage of social networking. Many online group buying platforms now include social networking features in order to entice customers to shop [4]. Taobao is one of China's largest online group purchasing platforms, offering instant messaging services to buyers to facilitate the exchange of auction information. The number of online groups purchasing customers has increased by 11% since 2009. Many people believe that online group-buying is a significant e-commerce business model with a lot of potentials. This study proposed several methods for determining prediction accuracy. This study investigates five financial risk, initiator trust, product performance, social influence, and quantity effect dimensions that have been shown to be significant in previous group purchasing studies. It also employs information retrieval and sentiment analysis. 894 auction data points were gathered from the most popular group buying platforms. The results demonstrated that the method's prediction accuracy was greater than 90%. There has been no research on auction classification at the South Jakarta District Prosecutor's Office using the Naive Bayes method, according to several previous research references. One of the methods required in data is naive Bayes classification. The Naive Bayes algorithm is a classification algorithm that predicts the future based on several parameters or criteria [5]. The satisfaction of auction participants can lead to an increase in auction transactions if auction items are classified. The Naive Bayes method was used in this study, which relied on the probability or opportunities that will occur based on the content of existing opportunities. Using the Naive Bayes method because it performs well when the training data does not contain all of the possibilities, allowing for a small amount of data[6]. A dataset is required for classification, and the dataset contains two conditions: general attributes (variable) and class. The Evidence and Seizure Management Section will record and examine the confiscated goods based on their condition before classifying or classifying the booty. The Conflict Goods Management Section requires one month to complete this task. Data obtained by 145 vehicles, with details of 100 motorbikes and 45 cars. In classifying booty based on the class that has been determined by the Prosecutor's Office. The class has been determined to be 3, namely class A grade, grade B, and grade C. The results of the auction classification will be carried out by a testing process to determine the accuracy percentage. The high accuracy value indicates that the auction classification at the South Jakarta District Prosecutor's

Office can use the Naive Bayes method. Based on this description, this research hopes to conduct auction classification at the South Jakarta District Prosecutor's Office using the Naive Bayes method. Aiming to expedite and simplify the Evidence and Confiscation Management Section in managing and classifying evidence or booty so that it can be auctioned immediately, as well as assisting auction participants in determining vehicle choices based on vehicle quality.

II. METHODOLOGY

The process begins with data collection, classification, testing, and evaluation, where data collection is done with the data sources obtained, classification is done to form a model from Naive Bayes, testing is done to test the models that have been created, and data evaluation is done to assess accuracy using the Naive Bayes method, which is used for auction classification at the South Jakarta District Prosecutor's Office.

A. Data Collection

The first step in this research is data collection; the data obtained is in the form of a vehicle auction list. Direct data was obtained from the South Jakarta District Attorney and the South Jakarta District Prosecutor's Office website. The vehicles used in this study are motorcycles and cars that have been confiscated and a decision made by the prosecutor's court so that the decision is made through auction. The information gathered is a description of the spoils or evidence. Come directly to the South Jakarta District Court to collect some of the data in the study was to see and examine the condition of the evidence. The grade determinant has been determined by the Evidence and Confiscation Management Section based on how much damage has been done to the vehicle's condition. This study made use of 145 vehicle auction data. There are various types of conditions in each vehicle, such as the state of the engine and physical or exterior appearance that require minor repairs to the condition of the vehicle that requires major repairs. This is one of the classification criteria for a vehicle. The classification process is divided into two stages in the Naive Bayes Classifier: training and testing. [7].

B. Classification

The model to be created employs the Naive Bayes method, or it can be said to create the Naive Bayes model. Naive Bayes is a classification method that uses statistical and probability methods to predict odds based on prior experience [8]. Classification is the process by which a function learns to map each set attribute to a predetermined class [9]. Classification is part of the supervised learning process; classification requires training data. When creating the Naive Bayes model, look for the class probabilities and conditional probabilities of each class. Class probabilities seek or calculate the odds of grades A, B, and C, whereas conditional probabilities seek or calculate the number of cases against the same class. The Naive Bayes model was used in the training and testing process. Vehicle data, such as information about the vehicle's condition, will be converted or assumed to be numerical. It is divided into 50 % : 50 % from 145 vehicle data, with 50% used for training and 50% used for testing. A Naive Bayes model will be constructed using the training data. Testing data will be used to validate the model that has been developed. The Naive Bayes algorithm can also handle datasets with many attributes. To create the Naive Bayes Classifier model, examine the training data by calculating the odds in each event class and the conditional probability, or by calculating each variable in each event class. The general form of Naïve Bayes is as follows:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

This equation is the general form of the formula for the Naïve Bayes algorithm. Because of the assumption that every word is not related to each other, the assumption of independence is very high (naive), with this assumption [10] the equation is as follows :

$$P(C|X_1, \ldots, X_n) = P(C) \prod_{i=1}^{n} P(X_i|C) \quad (2)$$

C. Testing

Testing data is used to determine the extent to which the Naive Bayes Classifier is successful in performing auction classification properly and correctly. The data being tested should be separated from the training data so that it is clear that the Naive Bayes Classifier model was successful in classifying the auction. 45 auction testing data is used in the testing. In testing, there are two approaches: multiplying all variables according to the class or label and then comparing the results of multiplying each variable according to the class. The prediction results will be used to create a confusion matrix, which will be used to evaluate the model's performance [11]. There are seven parameters used in the classification of auction vehicles: vehicle key, vehicle year, complete documents in the form of STNK and BPKB, engine condition, exterior, and interior (car). Following that, three classes are established: grade A, grade B, and grade C.

D. Evaluation

The results of the auction classification evaluation stage will be used to test and measure the accuracy of the performance of the Naive Bayes Classifier model. The evaluation stage can be seen using a confusion matrix table, which will show the classification accuracy by comparing the testing data that already has a prediction class and class. The confusion matrix will tell how well the Naïve Bayes model is used [12]. So this is very useful because the results of the classification generally cannot be described properly in just one number.

TABLE 1. CONFUSION MATRIX

		Predicted		
		Grade A	**Grade B**	**Grade C**
Class	Grade A	AA	AB	AC
	Grade B	BA	BB	BC
	Grade C	CA	CB	CC

Value accuracy, recall, precision, and F1-score can be generated in Confusion Matrix. The accuracy stage of the system built into the auction classification is the evaluation stage. Accuracy is commonly defined as the degree of similarity between the predicted and actual value. Precision refers to the accuracy with which an auctioned vehicle is classified as grade A, B, or C. Precision describes the relationship between the requested data and the predictive results of the model. The ratio

of true positive predictions to overall positive predictions is defined as precision. A recall is a successful information retrieval model. A recall is also referred to as the system's success rate in recovering data. The F1-Score is a comparison of average precision and recall. If the F1-Score has a value or a high score, it means that the auction classification model for motor vehicles and cars is precise and has a high recall.

III. RESULT AND DISCUSSION

It will explain the use of the Naive Bayes method in classifying motorbike and car auctions at the South Jakarta District Prosecutor's Office in the results and discussion stages. This chapter describes the design and implementation of a motorcycle and car auction classification system using 145 datasets covering 100 motor vehicle data and 45 car vehicle data. There are a total of 7 dataset variables for motorbikes and cars, as for the determining variables that will be used in classifying auction vehicles, namely vehicle keys, vehicle year, completeness of documents (BPKB and STNK), engine condition, exterior and interior (car). The Python programming language is used in the program design, which is built with the Visual Studio Code platform. The PySimpleGUI library is used for the program's display. The platform used can classify vehicle auctions at the South Jakarta District Prosecutor's Office using the Naive Bayes method. The user interface program is written in Python using the Visual Studio Code platform. The Pandas, Numpy, and Pysimplegui libraries were used to create the program. The procedure begins in the left column, where the combo box offers the option of motor vehicle data files or car data files in.csv format. If datasetmotor.csv is selected, the right-hand column will display 100 motorbike data. In this column, there are 5 windows, namely, sample data window, accuracy value, confusion matrix, classification, and prediction reports. In the sample, the data window displays vehicle data. The accuracy value window will display the accuracy value of the vehicle that has been processed. The confusion matrix window will display the correctness of the value in the classification. The recall, precision, and f1-score values are displayed in the classification report window. The prediction window is then used to determine the vehicle auction classification prediction. The following table depicts the confusion matrix.

TABLE 2. CONFUSION MATRIX OF MOTORCYCLE

		Predicted		
		Grade A	**Grade B**	**Grade C**
Class	Grade A	20	4	3
	Grade B	2	8	3
	Grade C	1	0	9

TABLE 3. CONFUSION MATRIX OF CAR

		Predicted		
		Grade A	**Grade B**	**Grade C**
Class	Grade A	10	0	0
	Grade B	6	0	1
	Grade C	2	1	3

Table 2 contains a 3x3 confusion matrix for determining the accuracy or inaccuracy of motor vehicle predictions. For each vehicle grade, the correct predicted vehicle values are 20, 8, and 9, for a total of 37 vehicles. A confusion matrix from a car vehicle is shown in Table 3, with 13 vehicles predicted to be correct. This value is calculated using half of the dataset.

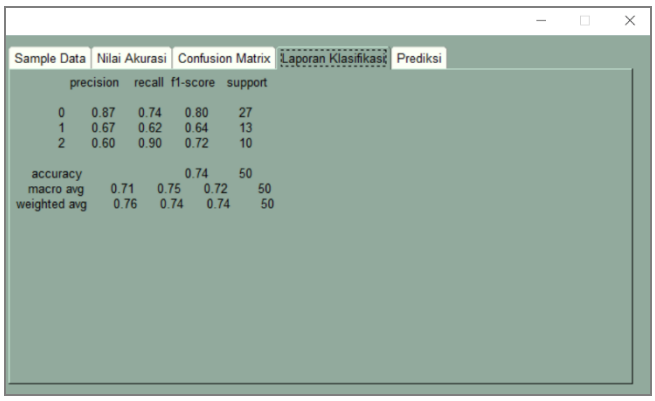

Fig. 1. Motor Vehicle Classification Report

Figure 1 depicts a motorcycle classification report program. The accuracy of the motor vehicle data in the report is 74%, with a total of 100 motor vehicle data. Figure 2 depicts the results of a classification report on a car.

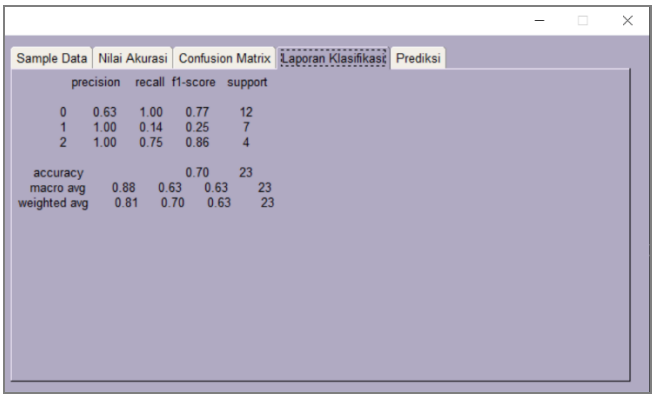

Fig. 2. Car Vehicle Classification Report

The program interfaces in Figure 2 displays the classification report results from the car vehicle dataset using the Naive Bayes method. The value of car vehicles on the data has a value of 57%. The accuracy value of the performance of the methods used in the classification of motorbike and car auctions at the South Jakarta Public Prosecutor's Office was 74% for the motor vehicle dataset and 57% for the car vehicle dataset, according to the findings of this study. Because the years used are from 2018 to 2020, the results of car vehicles have an accuracy value of 57 percent, and only a few car pieces of evidence are available at the South Jakarta state prosecutor. Regarding the accuracy value obtained from the performance of the Naive Bayes method, this system can be recommended to prosecutors, specifically the Management of Evidence and Seized Goods Section, in classifying or classifying motorbike and car auctions to shorten the time and avoid damage to vehicles that are stored in the warehouse attorney's evidence. To

assist the community in selecting a vehicle based on the desired conditions.

IV. CONCLUSIONS

Based on the results of the research that has been done, it can be said and concluded that the results of this study using Naïve Bayes can produce an accuracy rate of motor vehicles of 74% and an accuracy of car vehicles of 57%. This system can be recommended to prosecutors based on the accuracy value obtained from the Naive Bayes method performance. In this study, it is suggested that preprocessing be used to reduce process performance issues such as missing values, redundant data, and outliers. To improve prediction accuracy, deep learning should be used in future studies for dynamic or automated data training with the Naive Bayes approach to classification. Large amounts of data should be used to avoid low accuracy values.

REFERENCES

[1] F. N. Ulya, "DJKN Catat Transaksi Lelang Rp 85,9 Triliun pada 2015-2019," *www.kompas.com*, 2020. [Online]. Available: https://money.kompas.com/read/2020/02/28/163231226/djkn-catat-transaksi-lelang-rp-859-triliun-pada-2015-2019. [Accessed: 22-Jul-2020].

[2] R. Astuti and Nurhidayat, "Model Sistem Informasi Risalah Lelang di Kantor Pelayanan Kekayaan Negara dan Lelang Kota Cirebon dengan Metode Clustering Algoritma C45," *Media Inform.*, vol. 18, no. 3, pp. 107–121, 2019.

[3] K. Latifah, "Kombinasi Algorithma K-NN dan Manhattan Distance untuk Menentukan Pemenang Lelang," *J. Inform. Upgris*, pp. 49–58, 2015.

[4] C. C. Chen and M.-C. Chung, "Predicting the success of group buying auctions via classification," *Knowledge-Based Syst.*, vol. 89, pp. 627–640, 2015.

[5] R. Nadia and F. Nhita, "Analisis Dan Implementasi Algoritma Naïve Bayes Classifier Terhadapa Pemilihan Gubernur Jawa Barat 2018 Pada Media Online," vol. 5, no. 1, pp. 1678–1700, 2018.

[6] Varun, "Performance Comparison between Naïve Bayes, and Decision Tree?," 2019. .

[7] M. J. Islam, Q. M. J. Wu, M. Ahmadi, and M. A. Sid-Ahmed, "Investigating the Performance of Naive-Bayes Classifiers and K- Nearest Neighbor Classifiers," in *2007 International Conference on Convergence Information Technology (ICCIT 2007)*, 2007, pp. 1541–1546.

[8] M.-L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label naive Bayes classification," *Inf. Sci. (Ny).*, vol. 179, no. 19, pp. 3218–3229, 2009.

[9] M. Martinez-Arroyo and L. E. Sucar, "Learning an Optimal Naive Bayes Classifier," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, vol. 3, pp. 1236–1239.

[10] H. Rhomadhona and J. Permadi, "Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation," *J. Sains dan Inform.*, vol. 5, no. 2, p. 108, 2019.

[11] D. Bowes, T. Hall, and D. Gray, "Comparing the Performance of Fault Prediction Models Which Report Multiple Performance Measures: Recomputing the Confusion Matrix," in *Proceedings of the 8th International Conference on Predictive Models in Software Engineering*, 2012, pp. 109–118.

[12] M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 2, pp. 88–91, 2019.