Journal of Electrical Engineering and Computer (JEECOM)
Vol. 7, No. 1 (2075), DOI: 10.33650/jeecom.v4i2
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Classification of LPG Gas Usage Satisfaction Level Using The Naïve Bayes Algorithm

Angelia Melani Adrian[1], Bella Alisia Patras[2], Junaidy B. Sanger[3]
[1,2,3] Program Studi Teknik Informatika, Fakultas Teknik, Universitas Katolik De La Salle, Manado, Indonesia

Article Info

Article history:

Received April 11, 2025
Revised April 14, 2025
Accepted April 16, 2025

Keywords:

Classification
Data Mining
LPG Gas
Satisfaction Level
Naïve Bayes Algorithm

ABSTRACT

LPG gas is a very important energy source in everyday life, especially for cooking activities. Although its importance in supporting daily life has been widely recognized, the issue of user satisfaction with LPG gas should not be ignored. Dissatisfaction can arise when products or services fail to meet consumer expectations, which may be due to low quality, prices that are not in line with the perceived value, or a mismatch with user expectations. Considering that LPG gas is used by a wide segment of society, understanding consumer satisfaction is crucial to improving service quality and consumer trust. This study aims to classify the level of satisfaction with LPG gas usage using the Naïve Bayes algorithm as a decision-support tool that can assist producers and distributors in identifying key satisfaction factors. The dataset consists of 250 survey responses assessed using five attributes: meeting needs, good quality, affordable prices, repurchasing intention, and willingness to recommend the product. The classification is based on two classes: satisfied and dissatisfied. The resulting model achieved varying accuracies depending on the training-to-test data ratio, with 89.3% for a 70:30 split, 91.2% for an 80:20 split, and 94.0% for a 60:40 split, indicating that the algorithm performs well and can be utilized to support customer satisfaction analysis and improvement strategies.

Corresponding Author:
Angelia Melani Adrian,
Program Studi Teknik Informatika, Fakultas Teknik, Universitas Katolik De La Salle, Manado, 95253, Indonesia
Email: madrian@unikadelasalle.ac.id

1. INTRODUCTION

Indonesia's increasing economic growth in recent years has encouraged the government to take active steps to improve national economic conditions. One of the significant policies implemented is the conversion from kerosene to LPG gas, aimed at replacing kerosene as a fuel source due to the projected depletion of kerosene reserves by 2025 [1]. Understanding public satisfaction with LPG gas usage becomes essential to ensure the effectiveness and sustainability of this conversion policy. Therefore, this research is conducted to analyze and classify the level of user satisfaction with LPG gas, which is crucial for supporting policy evaluation and future improvements.

The main objective of this study is to classify the level of satisfaction of LPG gas users by identifying key influencing factors. Although the importance of LPG gas as an alternative energy source is widely acknowledged, dissatisfaction can arise when the product fails to meet user expectations in terms of quality, price, or overall user experience. Satisfaction plays a key role in the continued use and public trust in LPG gas. Hence, this study seeks to contribute by providing a structured classification model that captures the satisfaction levels of users. The problem addressed in this research is the lack of specific studies focusing on classifying LPG gas user satisfaction using a machine learning approach, particularly the Naïve Bayes algorithm.

To achieve this objective, the Naïve Bayes algorithm is employed as the primary classification method. This algorithm is selected due to its proven accuracy, simplicity, and efficiency in handling both continuous and discrete data types [3]. It also performs well with relatively small datasets and is capable of generating probabilistic predictions, which makes it suitable for this classification task. This research uses survey data collected from LPG gas users at depots and implements the model through a web-based application titled "Classification of LPG Gas Satisfaction Levels Using the Naïve Bayes Algorithm," aiming to assist stakeholders in understanding and enhancing consumer satisfaction.

2. LITERATURE STUDY

1. LPG Gas

Liquefied Petroleum Gases (LPG) is the main gas fuel commonly used for cooking purposes. The composition in LPG cylinders consists of a mixture of propane (C_3H_8) and butane (C_4H_{10}), with a smaller amount of liquid pentane (C_5H_{12}) [4]. LPG is chosen as an alternative to replace kerosene because the production cost of LPG is more economical than kerosene. Therefore, the use of LPG is not only among urban residents, but also in rural areas [5]. LPG is generally used for cooking purposes, but besides that LPG can also be used as fuel for motor vehicles known as BBG (Gas Fuel) [2].

2. Satisfaction

Satisfaction is a state of fulfilling consumer needs, desires, and expectations. This is considered important because this satisfaction can produce a positive response and can help in determining whether users are loyal or not [6]. User satisfaction describes the match between individual expectations and the results received [7].

3. Data Mining

Data mining is the process of exploring extensive and complexly structured data to find patterns, relationships, or other interesting information [8]. Raw data is used in this process to produce information that can be applied to decision making or strategic action. Data mining is a data analysis process that can be done using various methods. In addition to collecting data, data mining also includes analysis and prediction of the information that will be displayed. The collected data will be stored in a database and then processed so that it can be used in making decisions to view relevant information [9]. Data mining has elements that are included in the Knowledge Discovery in Database (KDD) series. The stages of KDD are as follows [8]:

1) Data Selection
   Selecting relevant data and attributes from the operational data set. The selected data must meet the criteria required for further analysis. The selected data is stored in a separate file from the operational data.
2) Pre-processing (Cleaning)
   Cleaning the selected data by removing duplicates, handling missing data, and eliminating inconsistent data. This process is important to ensure data quality before the next stage.
3) Transformation
   Transforming the cleaned data into a format suitable for analysis. This process may involve normalization, aggregation, or other transformations as needed.
4) Data MiningApplying techniques and methods to find interesting patterns or information in the data. The method or algorithm chosen is very important because it will affect the final result.
5) Interpretation
   Presenting the patterns or information found in a format that can be easily understood by the parties involved.
6) Classification
   Classification is a step of finding patterns that show similarities in a group with the aim of predicting the class of an unrecognized object. Grouping in classification is based on the attributes possessed by an object [11].

In general, the classification process is divided into two parts: the learning process involving training data and the case classification process. The classification algorithm works through training data to create a model during the learning phase. After the model has been tested and declared valid, the model is used in the classification phase to estimate the class of new cases, thus supporting decision making [10].

4. Naïve Bayes

The naïve Bayes algorithm is an easy way to classify probability values, by adding a combination of values and frequencies from the available data set [12]. Bayes' theorem is used in this method which states that the class variable must be independent or independent when passed to the variable [12]. The use of naïve Bayes has the advantage of requiring a small amount of training data to determine the parameters needed for classification [13]. Training data is used as a reference in every algorithm calculation, while test data is used to evaluate the accuracy of predictions and decisions produced by the algorithm [13]. The following is the Bayes theory Equation [1]:

$$P(C|X) = \frac{P(x|c)P(c)}{P(x)} \tag{1}$$

Where:

- *X* : *Data with unknown class.*
- *C* : *Hypothesis that c is data with class.*
- *P(C|X)* : *Probability of C against hypothesis condition X (posterior probability).*
- *P(c)* : *Probability of hypothesis c (prior probability).*
- *P(x|c)* : *Probability of hypothesis x based on condition c.*
- *P(x)* : *Probability of observed x.*

6. Confusion Matrix

A confusion matrix is a table used to evaluate the performance of a classification model in machine learning. This table visualizes the prediction results of the model by comparing the predictions with the actual labels. The confusion matrix helps in identifying how well the model predicts certain classes and the possible errors. The following is a matrix table that can be seen in Table 1 [14]:

Table 1**.** Confusion Matrix

Actual	**Predicition Data**		
Data	**True**	**False**	**Total**
True	TP	FN	P
False	FP	TN	N
Total	P	N	P+N

Description:

- *TP* : *Number of samples that are truly positive and predicted as positive by the model.*
- *TN* : *Number of samples that are truly negative and predicted as negative by the model.*
- *FP* : *Number of samples that are truly negative but predicted as positive by the model (also known as Type I Error).*
- *FN* : *Number of samples that are truly positive but predicted as negative by the model (also known as Type II Error).*

Based on the matrix table, the accuracy was used to measure the performance of the matrix. Accuracy is the correct prediction value of all data. Accuracy is calculated using the Equation 2 [15]:

$$\text{Accuracy} = \frac{TP+TN}{P+N} \times 100\% \tag{2}$$

7. Waterfall Method

The Waterfall method is one of the most classic and traditional software development process models. This method is known as a linear sequential model because each stage in the software development process is carried out sequentially and there is no overlap between the stages [16]. With this method, the resulting application tends to have good quality because each stage is carried out sequentially without focusing on one particular stage [16]. The stages of the waterfall method are as follows:

1) Requirement Analysis
   At this stage, all the needs and requirements of the system to be developed are collected and analyzed. The result of this stage is a detailed and detailed requirements specification document. This document serves as a guide for the entire development process.
2) System Design
   Based on the requirements specifications that have been determined, system and software designs are created. This design includes system architecture, software components, user interfaces, and database designs. The goal is to design a solution that meets all the needs that have been identified.
3) Implementation
   At this stage, the design that has been created is translated into program code. Each unit or software component is developed separately and tested individually to ensure that each unit functions properly according to the design made.
4) Testing
   After all software components have been developed and tested individually, the components are integrated and tested as a whole. The goal of this stage is to ensure that the entire system works well and in accordance with the specifications that have been determined.

3. METHODOLOGY

3.1. Data

The data collection process was carried out by surveying gas users at the base. Survey questions were created by the author then printed and distributed at four LPG gas bases located in Pineleng District, Minahasa Regency, Sulawesi Utara. The data obtained from the survey results were 250 data.

Table 2. LPG Usage Data

No.	Meet the needs	Good Quality	Affordable Price	Buy Back	Product Recommendation	Category
1.	Agree	Agree	Agree	Agree	Agree	Satisfied
2.	Agree	Agree	Agree	Agree	Agree	Satisfied
3.	Agree	Partially Agree	Agree	Agree	Partially Agree	Disatisfied
4.	Disagree	Agree	Agree	Agree	Agree	Satisfied
5.	Disagree	Partially Agree	Agree	Agree	Agree	Satisfied
…	…	…	…	…	…	…
250.	Agree	Agree	Agree	Agree	Agree	Satisfied

3.2. Proposed Method: Naive Bayes Algorithm Calculation

Given Test Data:

X = (Meets needs = agree, good quality = doubtful, affordable price = agree, repurchase = agree, recommend product = doubtful).

1. Calculating class probability

Satisfaction Class	Probability
Disatisfied = 59	P(Dissatisfied) = 59/250 = 0,236
Satisfied = 191	P(Satisfied) = 191/250 = 0,764

2. Calculate the probability of each attribute

	Satisfied	Disatisfied
Meets Needs (Agree)	185/191 = 0,9685	40/59 = 0,6779
Good Quality (Undecided)	2/191 = 0,0104	24/59 = 0,4067
Affordable Price (Agree)	187/191 = 0,9790	19/59 = 0,3220
Return to Purchase (Agree)	186/191 = 0,9738	36/59 = 0,6101
Recommend Product (Undecided)	3/191 = 0,0157	16/59 = 0,2711

3. Calculate the final probability

Satisfied	Disatisfied
0,9685 × 0,0104 × 0,9790 × 0,9738 × 0,0157 = 0,0001	0,6779 × 0,4067 × 0,3220 × 0,6101 × 0,2711 = 0,0146

4. Comparing probability values

Satisfied	Disatisfied
0,0001*0764 = 0,0001	0,0146*0,236 = 0,0034

Thus, based on the calculation steps above, the result from the known test data show the Dissatisfied category, since the probability value is greater.

3.3. Modelling

System modeling involves the presentation of flowcharts and DFDs that describe the system modeling process.

1. Flowchart

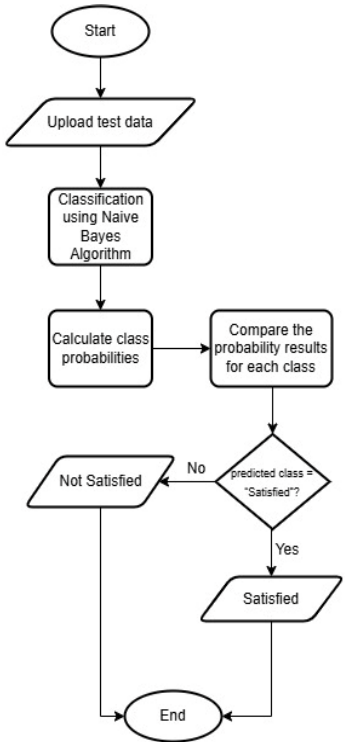

Figure 1. System Flowchart

This flowchart (Fig. 1) illustrates a step-by-step process for classifying user satisfaction using the Naive Bayes algorithm. The process begins with the upload of test data, which serves as input for the classification model. The Naive Bayes algorithm is then applied to this data to perform the classification task. Following this, the algorithm calculates the probability of the data belonging to each possible class. Once the class probabilities are obtained, they are compared to determine which class has the highest likelihood. A decision point then evaluates whether the predicted class is "Satisfied". If the result is "Satisfied", the process continues to a corresponding output. If the predicted class is anything other than "Satisfied", the classification result is marked as "Not Satisfied". The process concludes after the final classification decision is made. This flowchart provides a clear overview of how probabilistic classification is used to predict satisfaction levels based on input features.

2. Data Flow Diagram

This stage will discuss the description of the application of satisfaction level classification displayed in the data flow diagram. (Fig. 2) explains This DFD provides a simple yet effective overview of how user input (regarding LPG gas usage satisfaction) is processed using a machine learning model, specifically the Naive Bayes algorithm to classify and return the level of satisfaction. The system acts as an evaluator, giving feedback to the user based on their provided information.

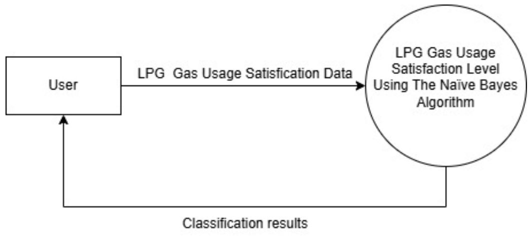

Figure 2. DFD Level 0

(Fig. 3) presents a detailed Data Flow Diagram (DFD) Level 1 of the classification process for evaluating LPG gas usage satisfaction using the Naive Bayes algorithm. The process begins with the user providing input in the form of Gas User Satisfaction Data. This data is first handled in the "Perform Data Partitioning" stage, where it is divided into two distinct sets: training data and testing data. The partitioned results are then returned to the user for transparency and further processing. Next, the training data is passed to the "Display Training Data" process, which visualizes and presents the training dataset along with its attributes or details. This ensures that the dataset is ready to be used for training purposes. Following this, the testing data is sent to the "Display Testing Data" process, where it is likewise displayed with complete testing data details, preparing it for use in model evaluation. Finally, the testing data flows into the "Apply Naive Bayes Algorithm" process. In this step, the trained model uses the Naive Bayes algorithm to classify the data and determine the user's LPG gas usage satisfaction level. The result of this classification is then delivered back to the user as the classification result, completing the flow.

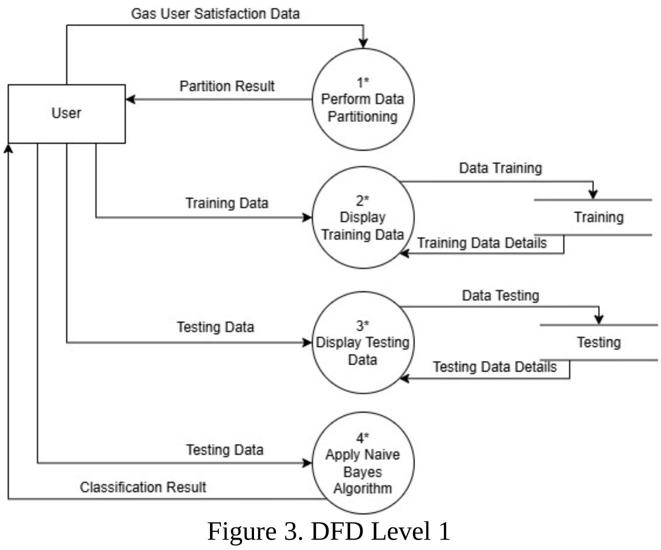

Figure 3. DFD Level 1

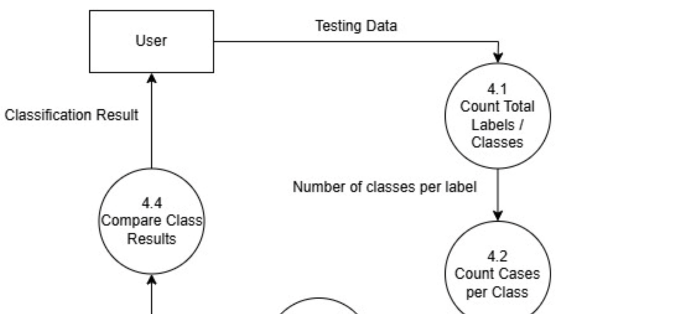

Figure 4. DFD Level 2 Proses 4

(Fig. 4) provides a detailed Data Flow Diagram (DFD) showing the inner workings of the Naive Bayes classification process used to determine LPG gas usage satisfaction. The process begins when the user provides testing data, which is then passed into a sequence of calculations to determine the most likely classification result. The first step, Count Total Labels / Classes, involves identifying the total number of class labels in the dataset such as "Satisfied" or "Not Satisfied". This establishes how many potential outcomes (or classes) the algorithm must consider. Next, Count Cases per Class, the system calculates the number of data cases (or instances) that fall under each class. This helps in determining the prior probabilities, which are essential in the Naive Bayes approach for making predictions. Following this, the process moves to Multiply All Class Variables, where all relevant class-related variables (such as probabilities of each attribute given a class) are multiplied together. This is done for each class, resulting in a final multiplication value that represents the likelihood of the testing data belonging to that specific class. Next to Compare Class Results, the multiplication results from the previous step are compared. The class with the highest result is selected as the predicted classification. This outcome is then sent back to the user as the classification result.

3.4. Application Interface

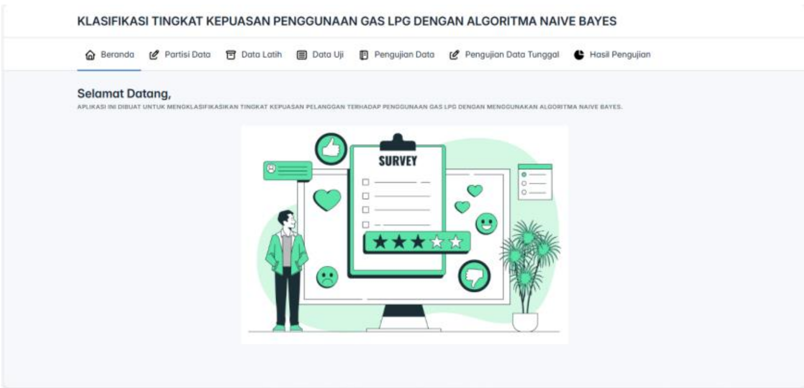

Figure 5. Application Home Page

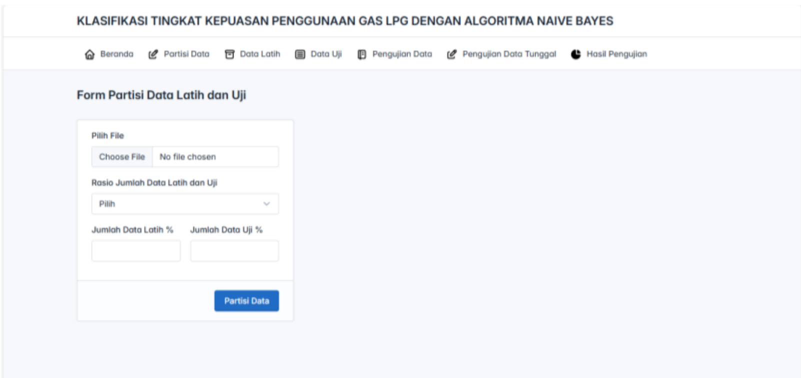

Figure 6. Data Partition Page

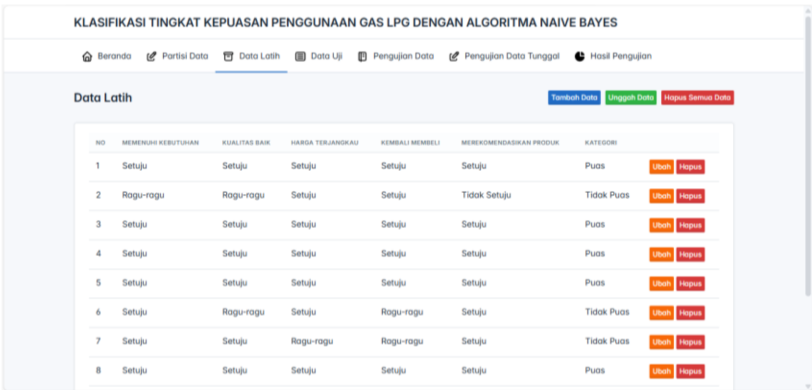

Figure 7. Training Data Page

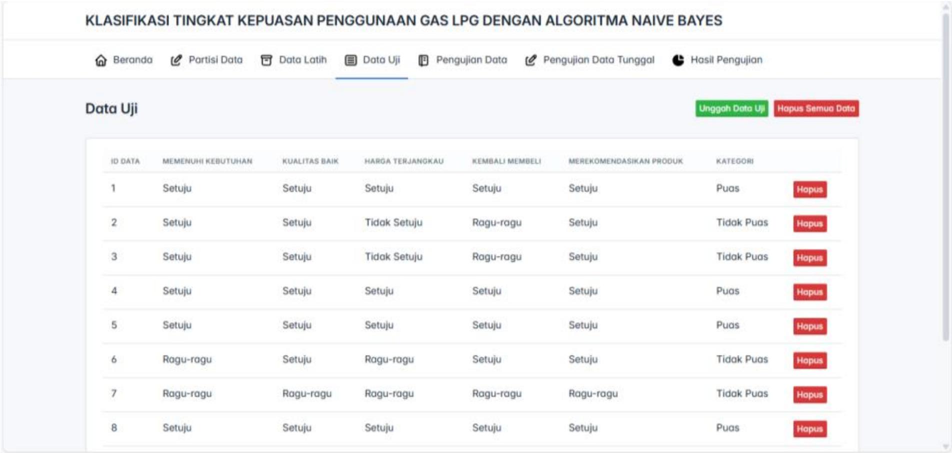

Figure 8. Testing Data Page

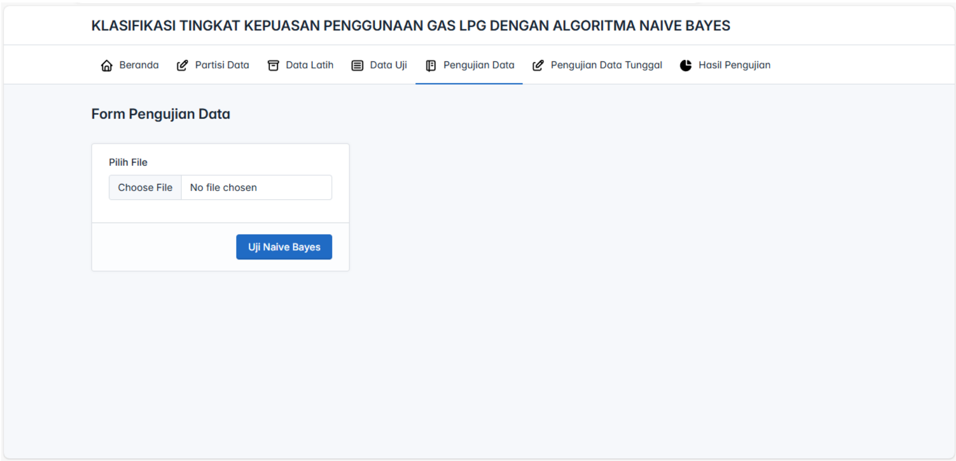

Figure 9. Figure 9. Data Testing Page Interface

Figure 10. Data Test Results Page

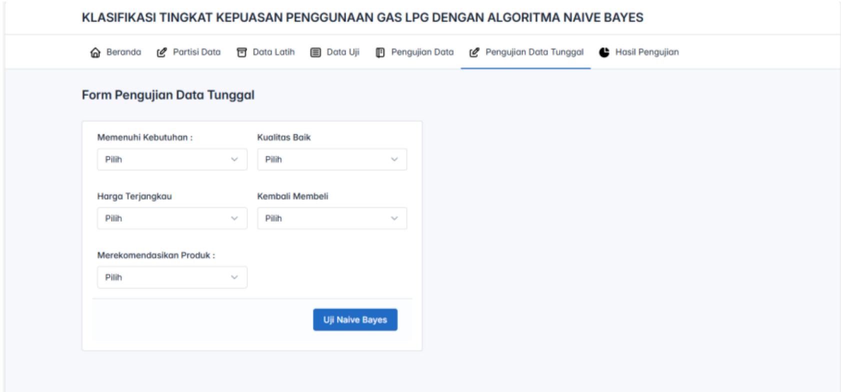

Figure 11. Single Data Test Page

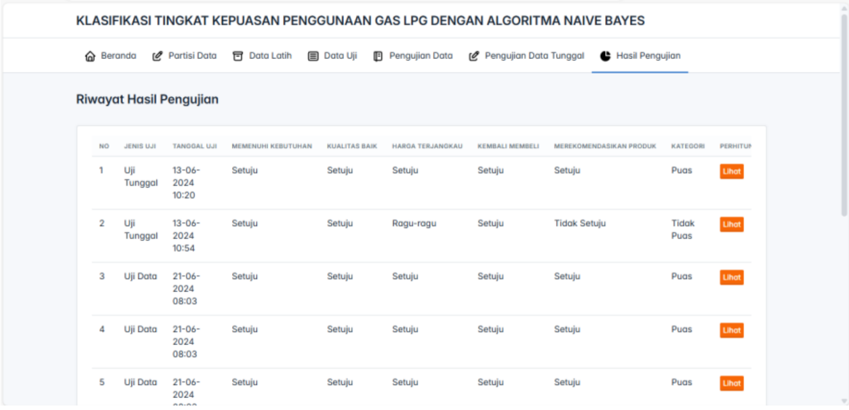

Figure 12. Test Result History Page

4. RESULTS AND DISCUSSION

Table 2 presents the experimental results, illustrating the performance of the Naïve Bayes algorithm in classifying LPG gas user satisfaction across different training and testing data splits. Generally, as the percentage of training data increases, the model learns better patterns, but accuracy can fluctuate due to variations in data distribution.

1. At 60% training and 40% testing, the model achieved the highest accuracy of 94.00%, suggesting that this split provided a good balance between learning and generalization.
2. With 70% training and 30% testing, the accuracy dropped to 89.30%, possibly due to overfitting or variations in the dataset.
3. When the training data increased to 80% (with 20% testing), the accuracy improved slightly to 91.20%, indicating that more training data helped the model learn better, though not as high as the 60-40 split.

These results highlight the trade-off between training data size and model accuracy, showing that the best performance was obtained with 60% training data, while accuracy fluctuated at other splits.

Table 2. Results

% Data Training	% Data Testing	Accuracy
60	40	94.00%
70	30	89.30 %
80	20	91.20 %

5. CONCLUSION

Based on the results obtained from the analysis and testing of the application "Classification of LPG Gas Satisfaction Levels Using the Naïve Bayes Algorithm", the following conclusions can be drawn:

1. Classification of LPG gas satisfaction levels using the Naïve Bayes algorithm has been successfully implemented.
2. Testing with a percentage of 70-30 got an accuracy value of 89.3%, a percentage of 80-20 got an accuracy value of 91.2%, and a percentage of 60-40 got an accuracy value of 94%.
3. The application was tested on 3 different web browsers, Microsoft Edge version 125.0, Google Chrome version 125.0, and Mozilla Firefox version 126.0, the results showed that all features could function properly.
4. The application meets the requirements needed by users.

For future research, several enhancements can be implemented:

1. Incorporating additional variables and expanding to five classification categories, such as: very satisfied, satisfied, moderately satisfied, dissatisfied, and very dissatisfied.
2. Conducting comparisons with other algorithms to evaluate and determine their performance.

ACKNOWLEDGEMENTS

The completion of this research would not have been possible without the support and guidance of many individuals and institutions. Special appreciation is given to the respondents who participated in the survey and shared their time and experiences, which served as the foundation of this study. This research is a small contribution to the field of data science and consumer satisfaction analysis, and the author hopes it will be beneficial for future studies and practical applications in energy usage and public policy development.

REFERENCES

[1] M. Adriansa, L. Yulianti, and L. Elfianty, “Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5,” *J. Tek. Inform. UNIKA St. Thomas*, vol. 07, no. 21, pp. 115–121, 2022.

[2] M. R. H. M. Saleh and I. W. Sudiarsa, “Rancang Bangun Pendeteksi Monitoring Gas LPG Berbasis Mikrokontroller dan Notifikasi Telegram Messenger,” *J. Krisnadana*, vol. 3, no. 1, pp. 310–319, 2023.

[3] M. Siddik, H. Hendri, R. N. Putri, Y. Desnelita, and G. Gustientiedina, “Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 162–166, 2020.

[4] R. Andriana and E. Prasetyaningrum, “Sistem Informasi Pendistribusian Gas Lpg Di Pt . Budi Bhakti Kalimantan Berbasis Web,” *J. Penelit. Dosen Fikom*, vol. 10, no. 1, pp. 1–11, 2019.

[5] S. Laitera, W. A. Dewa, and S. Arifin, “Penerapan Sistem Alarm Berbasis Arduino Uno Untuk Mendeteksi Kebocoran Gas LPG,” *J. Janitra Inform. dan Sist. Inf.*, vol. 2, no. 2, pp. 96–106, 2022.

[6] R. Alfiani and Y. Umaidah, “Prediksi Kepuasan Customer Terhadap Performance Terapis Baby Massage Menggunakan Algoritma Naive Bayes,” *Bina Insa. Ict J.*, vol. 9, no. 1, p. 83, 2022.

[7] P. L. Lokapitasari Belluano, I. Indrawati, H. Harlinda, F. A. . Tuasamu, and D. Lantara, “Analisis Tingkat Kepuasan Pengguna Sistem Informasi Perpustakaan Menggunakan Pieces Framework,” *Ilk. J. Ilm.*, vol. 11, no. 2, pp. 118–128, 2019.

[8] E. T. Naldy and A. Andri, “Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN,” *J. Nas. Ilmu Komput.*, vol. 2, no. 2, pp. 89–101, 2021.

[9] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, “Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest,” *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, 2022.

[10] S. Lestari, A. Akmaludin, and M. Badrul, “Implementasi Klasifikasi Naive Bayes Untuk Prediksi Kelayakan Pemberian Pinjaman Pada Koperasi Anugerah Bintang Cemerlang,” *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 7, no. 1, pp. 8–16, 2020.

[11] Sutisna and N. M. Yuniar, “Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve baiyes,” *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 243–246, 2023.

[12] I. Amillina and A. Qoiriah, “Penerapan Algoritma Naïve Bayes Dalam Klasifikasi Tingkat Kepuasan Siswa Terhadap Pembelajaran Daring,” *J. Ilm. Teknol. Inf. dan Robot.*, vol. 3, no. 2, pp. 16–23, 2021.

[13] N. Asmiati and Fatmawati, “Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Pengaruh Negatif Game Online Bagi Remaja Milenial,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 3, pp. 141–149, 2020.

[14] A. C. Khotimah and E. Utami, “Comparison Naive Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine in the classification of individual on twitter account,” *J. Tek. Inform.*, vol. 3, no. 3, pp. 673–680, 2022.

[15] A. Nugroho and Y. Religia, “Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 504–510, 2021.

[16] K. Kadarsih and S. Andrianto, “Membangun Website SMA PGRI Gunung Raya Ranau Menggunakan PHP dan MYSQL,” *JTIM J. Tek. Inform. Mahakarya*, vol. 03, no. 2, pp. 37–44, 2022.