Estimated Costs for Single Tuition Fee (STF) using Naïve Bayes Method

Juhriyansyah Dalle1a, Dwi Hastuti2b, Taufik Rahmain1d, A. Akrinim1e, Sri Erliani2c, Taufik Hidayat3f, Siska Devina4g, Agustina Lessari5h, B. Baharuddin6i, Hesti Fibrasari7j, Akhmad Murjani4k, Erika Lismayani8l, Ahmad Yusuf9m and Candra Kusuma Negara6n

1Department of Information Technology, Universitas Lambung Mangkurat, Jl. H. Hasan Basry, Banjarmasin, Indonesia
2Department of Electrical Engineering, Universitas PGRI Adi Buana, Jl. Ngagel Dadi III No.3B/37, Surabaya, Indonesia
3Department of Higher Education Management, Universitas Muhammadiyah Sumatera Utara, Jl. Kapten Mutchar Basri No.3, Medan, Indonesia
4Department of Health, Universitas Cuhaya Bangsa, Jl. A. Yani Km. 17, Banjarmasin, Indonesia
5Department of Management, Universitas Cuhaya Bangsa, Jl. A. Yani Km. 17, Banjarmasin, Indonesia
6Department of Nursing Sciences, Universitas Cuhaya Bangsa, Jl. A. Yani Km. 17, Banjarmasin, Indonesia
7Department of Electrical Engineering Education, Universitas Negeri Medan, Jl. Wiliem Iskandar Pasar V, Medan, Indonesia
8Department of Health, Universitas Negeri Medan, Jl. Wiliem Iskandar Pasar V, Medan, Indonesia
9Department of French Education, Universitas Negeri Medan, Jl. Wiliem Iskandar Pasar V, Medan, Indonesia
10Department of Information Technology, Universitas Cuhaya Bangsa, Jl. A. Yani Km. 17, Banjarmasin, Indonesia

Keywords: Decision Support System, STF, Naïve Bayes.

Abstract: The difference in the amount of single tuition fee (STF) paid by students with middle and upper economic backgrounds causes an injustice gap. This is partly due to the instability of the system used, especially in terms of STF determination methods. Other additional shortcomings include the criteria entered the system that is still not enough to be considered in determining STF. Therefore, this study aims to build a web based STF payment system using the Naïve Bayes, probability, and statistical methods for students to determine the cost of an institution's tuition fee easily. System testing is carried out by comparing the output and verification results. This showed that the estimated determination of the cost of STF payments is suitable, with 83.3%.

https://orcid.org/0000-0000-5700-2766
https://orcid.org/0000-0001-8716-7062
https://orcid.org/0000-0002-5553-2298
https://orcid.org/0000-0002-3984-9242
https://orcid.org/0000-0002-6562-3175
https://orcid.org/0000-0001-9659-4890
https://orcid.org/0000-0002-0755-6911
https://orcid.org/0000-0003-1180-9625
https://orcid.org/0000-0001-8334-3030
https://orcid.org/0000-0002-8009-4603
https://orcid.org/0000-0002-1162-2660
https://orcid.org/0000-0002-7963-4852
https://orcid.org/0000-0003-2383-9944
https://orcid.org/0000-0002-9831-7465

828


Estimated Costs for Single Tuition Fee (STF) using Naïve Bayes Method.
DOI: 10.5220/0010476702800289

In Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021) - Volume 1, pages 280-289
ISBN: 978-989-758-509-8

Copyright © 2021 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved
1 INTRODUCTION

The tuition fee of tertiary institutions is one of the main aspects considered by prospective students when choosing to further their studies. According to Gyamfi, Gyamfi, & Qi (2016), tuition fees are compared with the product, place, process, student quality, and graduate evidence. Ozekicioglu (2017) stated that living costs are also considered in the financial calculations of prospective students. In contrast, Bates & Kaye (2014) reported that changes in the amount of payment directly affect the number of applicants at tertiary institutions. Generally, when tuition fees are increased, the number of applicants tends to decrease, and vice versa (Sulistiyo & Soegoto, 2018). However, according to a study carried out by Burgess, Senior, & Moores (2018), the possibility of an increase in tuition fees can be minimized by improving the quality of service and providing a general view of universities. Lassila (2011) stated that providing financial assistance also enables students from low-income families to further their studies into tertiary institutions.

Therefore, a policy regarding the proportional tuition fee, which does not disrupt the student's family's financial flow, is needed (Surtiati, Siregar, & Andati, 2017). This led to the evolution of the STF policy, which was applied in accordance with the Minister of Education and Culture Regulation Number 55 of 2013. This was further amended to Minister of Education and Culture Regulation Number 73 of 2014, which was based on an evaluation of the implementation of single tuition fees for the 2013-2014 academic years at state universities. Additional changes occurred in 2015 and 2016 when the law was amended to No. 22 of 2015 and No. 39 of 2016, which stated that the costs borne by students need to be adjusted to the economic capabilities of their parents, or guidance. Furthermore, legal certainty in determining the costs borne by students need to be arranged regarding the cost of single tuition at state universities. In 2017 the law was amended by the Minister of Education and Culture Regulation to No. 39 of 2017 to regulate the costs borne by students according to their parents' economic capabilities or other parties that financed their education. This aimed at arranging a single tuition fee on state universities within the Ministry of Research, Technology, and Higher Education (Kemenristek-Dikti, 2017).

Dunga & Mncayi (2016) stated that improving the payment system is an absolute thing to do in order to assist students. However, this is not always directly proportional to enhancing the quality of higher education, made for the principles of accountability (Fortunata & Toni, 2020). This law was also amended to enable tertiary institutions to pay attention to aspects of parents' economic ability on an ongoing basis (Sumarno, Gimin, & Nas, 2017). This is also to prevent any form of protest likely to occur due to the unilateral determination by the stakeholders (Kajawo, 2019).

Implementing the STF system is to ease the financial burden of education funding on students (Fauzi, 2017). STF is the amount of fees that need to be paid by students each semester. It is divided into several groups, and each state university has 5 divisions, with some having more such as Indonesia Educational University. Universitas Lambung Mangkurat Banjarmasin adopted the STF System comprising of 5 groups in their 2013 academic year. With this system, tuition fees are paid every semester without building fees (Universitas Lambung Mangkurat, 2017). During this time, the university had a system to determine the class of STF. However, the required criteria were still lacking and did not use methods/algorithms to make calculations, therefore, the obtained results were inadequate in decision making. Besides that, with the existing system, prospective students are unable to determine the estimated cost of payment to the applied tertiary institution, which is very important (Galvin, Nieuwhuis, Phillips, Thain, & Kokkori, 2015).

Furthermore, Wardi, Abror, & Trinanda (2018), stated that the current STF system was inadequate and had a significant effect on students' desire to move to tertiary institutions. Therefore, for this reason, a web-based system for determining the cost of payment for STF was developed. This system uses a classification method known as the Naïve Bayes, which comprises probability and statistical methods to predict future opportunities based on past experiences. This is also known as the Bayes Theorem.

The use of the Naïve Bayes method evolved due to the predetermined use of data in various fields all over the world (Siledar & Chaudhary, 2017). Zhang and Gao (2011) used the Naïve Byes method in text classification based on the probability of students' ability in class. Furthermore, Safri, Ariifudin, & Muslim (2018) used it as a fundamental method to determine those that have the right to obtain a health insurance card provided by the state.

The Naïve Bayes method is also currently used in the field of economics and education. For example, Lagman et al. (2019) used this method to predict the speed and number of students graduating from college. Mirza (2019), also used this method to...
determine the right marketing and promotion strategies to attract new students. In economics, Kricehene proved that this method can be used to determine the probability of a bank's financial risk (2017). Therefore, the Naïve Bayes method can be used to determine the amount of single tuition for students in tertiary institutions.

With this system, prospective students can determine the estimated amount of Fee to be paid before the rector for proper verification.

2 METHODOLOGY

This research was carried out in five stages, namely planning, analysis, design, implementation, and system.

2.1 Planning

This is the earliest stage carried out to determine the purpose of making the system. This stage involves data collection, study libraries, interviews, and observations.

A literature study is carried out as a reference, source, and theories in research. The references in this study are journals, e-books, books, and the internet (electronic media).

Interviews stage is carried out with the directorate concerned on the STF that is the finance department. The topics discussed were, related to the definition of STF, how it is determined, and the various obstacles at Universitas Lambung Mangkurat.

After the interview stage, this process is carried out to determine the registration process and fees for prospective new students. Observation results were used to make the system, which is later used to form the data decision support system for STF.

2.2 Analysis

The analysis phase is carried out to analyze, understand, and provide a detailed document of the problems, solutions, and system requirements. It consists of a system, problem, and data analysis.

Systems analysis is an inseparable part of the study of business (Fotache, Olaru, & Iacoban, 2015), and software products. It is useful for an in-depth study of the needs of a product for suitable usage (Logunova et al., 2018). Analysis of the existing system enables Universitas Lambung Mangkurat to determine the class of STF. However, the criteria entered into the system are still lacking, and methods/algorithms are not used in making calculations, therefore, the results are still inadequate in making decisions.

The problem is analyzed to identify and solve the difficulties encountered (Annamalai, Kamaruddin, & Azid, 2013). Vizioli & Kaminski (2017) stated that understanding the problem also means providing in-depth descriptions. Innovative solutions are provided, assuming problem analysis is appropriate (Kim, Choi, Chang-Soo, & Park, 2018). Currently, prospective new students are unable to determine the amount of STF payment fees based on the criteria of the applied tertiary institution. Therefore, for this reason, a web-based STF fee payment estimation system was developed to determine the estimated STF fee for prospective new students. This system uses a classification method known as the Naive Bayes, which comprises the probability and statistical methods to predict future opportunities based on past experiences.

This stage is used to analyze the data needed in this research case study, such as parents' income, father's income, mother's income, PLN (electric power), number of families, number of dependents, and Indonesia Smart Card. These parameters are shown in Appendix 1.

A total of 400 data from the university and 40 from the manual calculation examples were used in this research (See appendix 2).

Data testing aims to determine the level of the presentation obtained by the Naive Bayes method in knowing how accurate the method is in solving STF problems. It was obtained from 30 data.

2.3 Design

At this stage, the tools used for system requirements, namely a computer/laptop and the software part is XAMPP, notepad, and web browser were determined. In addition, a flowchart or UML system was used to make the system.

This study's research material includes the results of surveys and observations carried out at Universitas Lambung Mangkurat. The tool used to implement the decision support system in this study is a computer, and the software part is XAMPP, notepad, and web browser.

The section comprises of planning or drawing a sketch that functions for the concept of making the system. The design can be carried out by developing a flowchart or UML system.
2.4 Implementation

This phase translates the design results in the previous stage into a system with a programming language. At this stage, a system is designed to estimate the cost of determining STF web-based payments for prospective new students.

2.5 System

At this stage, a system is formed, and testing is carried out to ensure it is made according to the study's objectives and needs. Furthermore, at this stage, a number of tests, including the Verifier Test, were carried out to determine the STF class accuracy.

Black box Testing is a functional test used to test software without knowing the internal structure of the code or program. Usability Testing is one way to determine whether users can easily use an application. It is also used to determine a program’s efficiency and effectiveness and help users achieve their goals.

3 RESULTS AND DISCUSSIONS

In the following example, it is assumed that the data is taken from a user and entered into the system. Table 1 shows the training and prospective student data selected based on the classification criteria of the STF class.

Approximately 400 training data were used to perform Naive Bayes calculations, as shown in table 1. The Naive Bayes method, used the following stages of calculation, namely: (1) counting the number of classes / labels, (2) count the same number of cases with the same class, (3) multiplying all results of group variables 1, 2, 3, 4, and 5. (4) comparing the high probability value. It found that the highest probability value is 0.005, which means it is included in group 3 (See appendix 3).

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Parents' incomes</td>
<td>2,000,000 - 2,500,000</td>
</tr>
<tr>
<td>2 Father's income</td>
<td>2,500,000</td>
</tr>
<tr>
<td>3 Mother's income</td>
<td>0</td>
</tr>
<tr>
<td>4 Electrical power</td>
<td>450 kwh</td>
</tr>
<tr>
<td>5 Number of families</td>
<td>5</td>
</tr>
<tr>
<td>6 Number of dependents</td>
<td>4</td>
</tr>
<tr>
<td>7 Indonesia smart card (kip)</td>
<td>None</td>
</tr>
</tbody>
</table>

Estimated group: ?

3.1 System Process

STF classification results were obtained from data collection and manual processing using the Naive Bayes method. This was followed by moving the database into the system in order to classify user entries.

The following is the process of calculating the system, starting with entering data according to criteria for proper classification in order to provide results in the form of STF classes.
3.2 Main Page

This page consists of 3 menus in the header section, namely STF, New STF, and Login. The STF menu functions by checking the results of the proof that has been processed, while the new STF menu serves as the input of STF data for prospective students of ULM to determine their STF group recommendations. The last is the login menu, which enables admins to log in to the back end of the website to manage STF data of registered prospective ULM students.

3.3 STF New Page

On the new STF page, prospective students are asked to fill in their data for the system to determine their estimated class.

First, the user fills in their personal data along with information on their "Parents' Income, Father's Income, Mother's Income, Electricity (PLN), Number of Families, Number of Dependents, and Smart Indonesia Cards."

After filling out the user selects the process button and classifies the data using the Naive Bayes algorithm as follows:
1. Count the number of classes/Labels.
2. Count the same number of cases with the same class.
3. Multiply all the results of Group 1, 2, 3, 4, and 5 variables.
4. Then compare and determine the highest probability value.
5. Store in a database and the system output in the form of STF results.

Figure 3 is a display of the results of the output system, which has carried out the Naive Bayes classification process.

3.4 Verification Test

Verification testing is carried out by comparing the results issued by the system with the verifier. The formulas used for testing system results are as follows:

\[
\text{Accuracy results} = \frac{\text{Total Correct}}{\text{Total Data}} \times 100\% = \text{accuracy results} \\
\text{(1)}
\]

The test results used to estimate the cost of STF payments using the Naive Bayes method at Universitas Lambung Mangkurat is calculated as follows:

\[
\text{Accuracy results} = \frac{25}{30} \times 100\% = 83.3\% \\
\text{(2)}
\]

Therefore, the percentage level of conformity on the system’s outputs with the verified results is 83.3%.

The explanation shows that several factors influence the calculation to determine the amount of single tuition. The first is the family’s economic background, which is the basis for determining the amount of a single tuition payment cluster and maintaining inequality to ensure the interest of going to college is maintained (Tang, Tang, & Tang, 2004). Previous studies have shown that cluster efficiency helps maintain the quantity and perception of students that register at a tertiary institution. Burer & Fethke (2016) stated that the distribution of payments in the cluster encourages students to pay for tuition.
Indirectly, the quantity of students’ impacts on improving the physical quality of the campus, leads to good maintenance and makes the instructors serious (Wanjala & Ali, 2017). Lin (2016) reported that the suitability of the amount of payment with students’ economic ability directly impacts their learning abilities and their desire to achieve academic success.

The next key factor is the use of the Naïve Bayes method, a key factor due to its ability to classify data in the appropriate cluster (Wibawa et al., 2019). Kaviani & Dhotre (2017) stated that the use of algorithms in this method overrides subjective assumptions and makes calculations more accurate. In addition, it has proven to be able to prevent corporations, institutions, and industries from financial risk due to incorrect calculation of funds received (Jang, Lee, Lee, & Han, 2015). Not only in the field of finance, the use of Naïve Bayes is also able to accurately predict the number of diligent students according to their performance (Shaziya, Zaeer, & G.Kavitha, 2015).

The benefits provided from the use of naïve Bayes have proven to be able to provide adequate assistance for the main educational institutions in calculating costs (Makhtar, Nawang, & Shamsuddin, 2017). In addition, the implementation of the Electronic information system facilitates students and operators in determining STF. Due to the current digital era, the use of information technology is a necessity in all fields (Dalle et al., 2015). Therefore, further research is needed on a broader study of the use of Naïve Bayes in educational institutions due to the limited scope of this study.

4 CONCLUSION

The following conclusions are drawn from the research carried out by observing and testing the various stages of the decision support systems:

1. Naïve Bayes method can be applied to estimate and determine the cost of STF payments at Universitas Lambung Mangkurat.
2. The classification process is carried out using the Naïve Bayes method, which works by classifying probabilities and statistics based on previous data.
3. The results issued or recommended by the system are suitable and in accordance with 83.3%.

ACKNOWLEDGEMENTS

We are thankful to employee of main office of Universitas Lambung Mangkurat who facilitate us in collecting data especially for single tuition fee. Also, we thank to Rector Universitas Cahaya Bangsa, Rector Universitas Negeri Medan, Dean of Faculty of Engineering ULM, and Head of Department of Information Technology who gave us more space for discussion of the paper.

REFERENCES


Jang, W., Lee, J. K., Lee, J., & Han, S. H., 2015. Naive bayesian classifier for selecting go00s4ad projects during the early stage of international construction bidding decisions. Mathematical Problems in Engineering, 1-12. doi: https://doi.org/10.1155/2015/830781


APPENDIX

Appendix 1: List of STF Parameters / Criteria.

<table>
<thead>
<tr>
<th>STF Parameters / Criteria</th>
<th>STF Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Parent Income</td>
<td></td>
</tr>
<tr>
<td>• 0-500.000</td>
<td>1</td>
</tr>
<tr>
<td>• 500-1,000.000</td>
<td>1</td>
</tr>
<tr>
<td>• 1,000-1,500.000</td>
<td>1</td>
</tr>
<tr>
<td>• 1,500-2,000.000</td>
<td>2</td>
</tr>
<tr>
<td>• 2,000-2,500.000</td>
<td>2</td>
</tr>
<tr>
<td>• 3,000-4,000.000</td>
<td>3</td>
</tr>
<tr>
<td>• 4,000-5,000.000</td>
<td>4</td>
</tr>
<tr>
<td>• 5,000-6,000.000</td>
<td>4</td>
</tr>
<tr>
<td>• 6,000-7,000.000</td>
<td>5</td>
</tr>
<tr>
<td>• 7,000-10,000.000</td>
<td>5</td>
</tr>
<tr>
<td>• 10,000-15,000.000</td>
<td>5</td>
</tr>
<tr>
<td>2. Father Income</td>
<td></td>
</tr>
<tr>
<td>• 1,000.000</td>
<td>1</td>
</tr>
<tr>
<td>• 1,500.000</td>
<td>2</td>
</tr>
<tr>
<td>• 2,000.000</td>
<td>2</td>
</tr>
<tr>
<td>• 2,500.000</td>
<td>2</td>
</tr>
<tr>
<td>• 3,000.000</td>
<td>2</td>
</tr>
<tr>
<td>• 4,000.000</td>
<td>3</td>
</tr>
<tr>
<td>• 5,000.000</td>
<td>4</td>
</tr>
<tr>
<td>• 6,000.000</td>
<td>4</td>
</tr>
<tr>
<td>• 7,000.000</td>
<td>5</td>
</tr>
<tr>
<td>• 9,000.000</td>
<td>5</td>
</tr>
<tr>
<td>• 10,000.000</td>
<td>5</td>
</tr>
<tr>
<td>3. Mother's Income</td>
<td></td>
</tr>
<tr>
<td>• 500.000</td>
<td>1</td>
</tr>
<tr>
<td>• 750.000</td>
<td>1</td>
</tr>
<tr>
<td>• 1,000.000</td>
<td>1</td>
</tr>
<tr>
<td>• 1,250.000</td>
<td>2</td>
</tr>
<tr>
<td>• 1,500.000</td>
<td>2</td>
</tr>
<tr>
<td>• 2,000.000</td>
<td>2</td>
</tr>
<tr>
<td>• 2,500.000</td>
<td>2</td>
</tr>
<tr>
<td>• 3,000.000</td>
<td>2</td>
</tr>
<tr>
<td>• 4,000.000</td>
<td>3</td>
</tr>
<tr>
<td>• 5,000.000</td>
<td>4</td>
</tr>
<tr>
<td>• 6,000.000</td>
<td>4</td>
</tr>
<tr>
<td>• 7,000.000</td>
<td>5</td>
</tr>
<tr>
<td>• 8,000.000</td>
<td>5</td>
</tr>
<tr>
<td>• 9,000.000</td>
<td>5</td>
</tr>
<tr>
<td>• 10,000.000</td>
<td>5</td>
</tr>
</tbody>
</table>

4. Electrical Power
   • 450kwh                 | 1          |
   • 500-1,000kwh           | 2          |
   • 1,000-1,500kwh         | 3          |
   • 1800-2200kwh           | 4          |
   • >2200kwh               | 5          |

5. Number of Families
   • >6                      | 1          |
   • 6                       | 2          |
   • 4                       | 3          |
   • 3                       | 4          |
   • 2                       | 4          |

6. Number of Dependents
   • >4                      | 1          |
   • 4                       | 2          |
   • 3                       | 3          |
   • 2                       | 4          |
   • 1                       | 5          |
Appendix 2: Data used as calculation experiments.

<table>
<thead>
<tr>
<th>No</th>
<th>Parents' Incomes</th>
<th>Father's income</th>
<th>Mother's income</th>
<th>Electrical power</th>
<th>Number of families</th>
<th>Number of dependents</th>
<th>Indonesia Smart Card</th>
<th>STF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,500.000 - 3,000.000</td>
<td>2,000.000</td>
<td>-</td>
<td>1300 kwh</td>
<td>3</td>
<td>1</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>2</td>
<td>500.000 - 1,000.000</td>
<td>6,000.000</td>
<td>-</td>
<td>900 kwh</td>
<td>5</td>
<td>2</td>
<td>None</td>
<td>Category V</td>
</tr>
<tr>
<td>3</td>
<td>2,000.000 - 2,500.000</td>
<td>2,500.000</td>
<td>-</td>
<td>900 kwh</td>
<td>5</td>
<td>4</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>4</td>
<td>1,500.000 - 2,000.000</td>
<td>2,000.000</td>
<td>-</td>
<td>900 kwh</td>
<td>4</td>
<td>3</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>5</td>
<td>500.000 - 1,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>6</td>
<td>500.000 - 1,000.000</td>
<td>-</td>
<td>1,000.000</td>
<td>450 kwh</td>
<td>5</td>
<td>&gt;4</td>
<td>None</td>
<td>Category II</td>
</tr>
<tr>
<td>7</td>
<td>500.000 - 1,000.000</td>
<td>500.000</td>
<td>-</td>
<td>450 kwh</td>
<td>&gt;6</td>
<td>3</td>
<td>None</td>
<td>Category I</td>
</tr>
<tr>
<td>8</td>
<td>0 - 500.000</td>
<td>-</td>
<td>500.000</td>
<td>450 kwh</td>
<td>6</td>
<td>3</td>
<td>None</td>
<td>Category I</td>
</tr>
<tr>
<td>9</td>
<td>0 - 500.000</td>
<td>1,000.000</td>
<td>-</td>
<td>450 kwh</td>
<td>6</td>
<td>3</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>10</td>
<td>4,000.000 - 5,000.000</td>
<td>3,000.000</td>
<td>-</td>
<td>900 kwh</td>
<td>6</td>
<td>&gt;4</td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>11</td>
<td>7,500.000 - 10,000.000</td>
<td>-</td>
<td>500.000</td>
<td>900 kwh</td>
<td>3</td>
<td>2</td>
<td>None</td>
<td>Category V</td>
</tr>
<tr>
<td>12</td>
<td>500.000 - 1,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>13</td>
<td>500.000 - 1,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>14</td>
<td>2,000.000 - 2,500.000</td>
<td>1,500.000</td>
<td>-</td>
<td>900 kwh</td>
<td>5</td>
<td>3</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>15</td>
<td>500.000 - 1,000.000</td>
<td>1,000.000</td>
<td>-</td>
<td>450 kwh</td>
<td>3</td>
<td>2</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>16</td>
<td>1,500.000 - 2,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>17</td>
<td>1,000.000 - 1,500.000</td>
<td>-</td>
<td>900 kwh</td>
<td>&gt;6</td>
<td>4</td>
<td>None</td>
<td>Category IV</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>500.000 - 1,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>19</td>
<td>1,000.000 - 1,500.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>20</td>
<td>1,500.000 - 2,000.000</td>
<td>1,500.000</td>
<td>-</td>
<td>900 kwh</td>
<td>6</td>
<td>4</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>21</td>
<td>1,500.000 - 2,000.000</td>
<td>1,500.000</td>
<td>-</td>
<td>900 kwh</td>
<td>3</td>
<td>1</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>22</td>
<td>0 - 500.000</td>
<td>-</td>
<td>1,500.000</td>
<td>1300 kwh</td>
<td>3</td>
<td>2</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>23</td>
<td>4,000.000 - 5,000.000</td>
<td>4,000.000</td>
<td>4,000.000</td>
<td>900 kwh</td>
<td>4</td>
<td>2</td>
<td>None</td>
<td>Category V</td>
</tr>
<tr>
<td>24</td>
<td>1,000.000 - 1,500.000</td>
<td>1,500.000</td>
<td>-</td>
<td>450 kwh</td>
<td>4</td>
<td>3</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>25</td>
<td>500.000 - 1,000.000</td>
<td>1,000.000</td>
<td>-</td>
<td>900 kwh</td>
<td>4</td>
<td>2</td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>26</td>
<td>2,500.000 - 3,000.000</td>
<td>3,000.000</td>
<td>-</td>
<td>450 kwh</td>
<td>4</td>
<td>3</td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>27</td>
<td>1,500.000 - 2,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>28</td>
<td>4,000.000 - 5,000.000</td>
<td>5,000.000</td>
<td>-</td>
<td>900 kwh</td>
<td>5</td>
<td>4</td>
<td>None</td>
<td>Category V</td>
</tr>
<tr>
<td>29</td>
<td>0 - 500.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>30</td>
<td>3,000.000 - 4,000.000</td>
<td>-</td>
<td>4,000.000</td>
<td>900 kwh</td>
<td>5</td>
<td>3</td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>31</td>
<td>1,500.000 - 2,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>32</td>
<td>0 - 500.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>33</td>
<td>1,000.000 - 1,500.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>34</td>
<td>1,500.000 - 2,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>35</td>
<td>1,000.000 - 1,500.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>36</td>
<td>500.000 - 1,000.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>None</td>
<td>Category IV</td>
</tr>
<tr>
<td>37</td>
<td>500.000 - 1,000.000</td>
<td>1,000.000</td>
<td>-</td>
<td>900 kwh</td>
<td>5</td>
<td>1</td>
<td>None</td>
<td>Category III</td>
</tr>
<tr>
<td>38</td>
<td>0 - 500.000</td>
<td>750.000</td>
<td>-</td>
<td>450 kwh</td>
<td>6</td>
<td>2</td>
<td>Yes</td>
<td>Category I</td>
</tr>
<tr>
<td>39</td>
<td>0 - 500.000</td>
<td>-</td>
<td>450 kwh</td>
<td>3</td>
<td>1</td>
<td>Yes</td>
<td>Category II</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>500.000 - 1,000.000</td>
<td>-</td>
<td>500.000</td>
<td>450 kwh</td>
<td>4</td>
<td>3</td>
<td>None</td>
<td>Category I</td>
</tr>
</tbody>
</table>

288
Appendix 3: Calculation Steps using Naive Bayes.

1. The first step is counting the number of classes/labels
   - \( P(Y = \text{Group 1}) = \frac{4}{40} \) "The amount of group 1 data in the training data is divided by the total amount of data"
   - \( P(Y = \text{Group 2}) = \frac{2}{40} \) "The amount of group 2 data in the training data is divided by the total amount of data"
   - \( P(Y = \text{Group 3}) = \frac{12}{40} \) "Total group 3 data in the training data is divided by the total amount of data"
   - \( P(Y = \text{Group 4}) = \frac{18}{40} \) "Amount of group 4 data in the training data is divided into the total amount of data"
   - \( P(Y = \text{Group 5}) = \frac{4}{40} \) "Amount of group 5 data in the training data is divided into the total amount of data"

2. Second step is to count the same number of cases with the same class
   - \( A1. P(\text{Parents' Income} = 2,000,000-2,500,000 \mid Y = \text{Group 1}) = \frac{0}{4} \)
   - \( B1. P(\text{Parents' Income} = 2,000,000-2,500,000 \mid Y = \text{Group 2}) = \frac{0}{2} \)
   - \( C1. P(\text{Parents' Income} = 2,000,000-2,500,000 \mid Y = \text{Group 3}) = \frac{2}{12} \)
   - \( D1. P(\text{Parents' Income} = 2,000,000-2,500,000 \mid Y = \text{Group 4}) = \frac{0}{18} \)
   - \( E1. P(\text{Parents' Income} = 2,000,000-2,500,000 \mid Y = \text{Group 5}) = \frac{0}{4} \)
   - \( A2. P(\text{father's income} = 2,500,000 \mid Y = \text{Group 1}) = \frac{0}{4} \)
   - \( B2. P(\text{father's income} = 2,500,000 \mid Y = \text{group 2}) = \frac{0}{2} \)
   - \( C2. P(\text{father's income} = 2,500,000 \mid Y = \text{group 3}) = \frac{1}{12} \)
   - \( D2. P(\text{father's income} = 2,500,000 \mid Y = \text{group 4}) = \frac{0}{18} \)
   - \( E2. P(\text{father's income} = 2,500,000 \mid Y = \text{group 5}) = \frac{0}{4} \)
   - \( A3. P(\text{Mother's income} = 0 \mid Y = \text{group 1}) = \frac{2}{4} \)
   - \( B3. P(\text{Mother's income} = 0 \mid Y = \text{group 2}) = \frac{1}{2} \)
   - \( C3. P(\text{Mother's income} = 0 \mid Y = \text{group 3}) = \frac{10}{12} \)
   - \( D3. P(\text{Mother's income} = 0 \mid Y = \text{group 4}) = \frac{16}{18} \)
   - \( E3. P(\text{Mother's income} = 0 \mid Y = \text{group 5}) = \frac{2}{4} \)
   - \( A4. P(\text{Electrical power} = 450kwh \mid Y = \text{Group 1}) = \frac{4}{4} \)
   - \( B4. P(\text{Electrical power} = 450kwh \mid Y = \text{Group 2}) = \frac{2}{2} \)
   - \( C4. P(\text{Electrical power} = 450kwh \mid Y = \text{Group 3}) = \frac{3}{12} \)
   - \( D4. P(\text{Electric power} = 450kwh \mid Y = \text{Group 4}) = \frac{1}{18} \)
   - \( E4. P(\text{Electric power} = 450kwh \mid Y = \text{Group 5}) = \frac{0}{4} \)
   - \( A5. P(\text{Number of Families} = 5 \mid Y = \text{Group 1}) = \frac{0}{4} \)
   - \( B5. P(\text{Number of Families} = 5 \mid Y = \text{Group 2}) = \frac{1}{5} \)
   - \( C5. P(\text{Number of Families} = 5 \mid Y = \text{Group 3}) = \frac{3}{12} \)
   - \( D5. P(\text{Number of Families} = 5 \mid Y = \text{Group 4}) = \frac{1}{18} \)
   - \( E5. P(\text{Number of Families} = 5 \mid Y = \text{Group 5}) = \frac{2}{4} \)
   - \( A6. P(\text{Number of Dependents} = 4 \mid Y = \text{Group 1}) = \frac{0}{4} \)
   - \( B6. P(\text{Number of Dependents} = 4 \mid Y = \text{Group 1}) = \frac{0}{2} \)
   - \( C6. P(\text{Number of Dependents} = 4 \mid Y = \text{Group 1}) = \frac{2}{12} \)
   - \( D6. P(\text{Number of Dependents} = 4 \mid Y = \text{Group 1}) = \frac{1}{18} \)
   - \( E6. P(\text{Number of Dependents} = 4 \mid Y = \text{Group 1}) = \frac{1}{4} \)

A7. \( P(\text{Smart Indonesia Card} = \text{None} \mid Y = \text{Group 1}) = \frac{3}{4} \)
B7. \( P(\text{Smart Indonesia Card} = \text{None} \mid Y = \text{Group 2}) = \frac{1}{5} \)
C7. \( P(\text{Smart Indonesia Card} = \text{None} \mid Y = \text{Group 3}) = \frac{2}{12} \)
D7. \( P(\text{Smart Indonesia Card} = \text{None} \mid Y = \text{Group 4}) = \frac{1}{18} \)
E7. \( P(\text{Smart Indonesia Card} = \text{None} \mid Y = \text{Group 5}) = \frac{4}{4} \)

3. Thirdly, multiply all the results of Group 1, 2, 3, 4, and 5 variables
   - \( A1*A2*A3*A4*A5*A6*A7 = \frac{0}{4}\times\frac{0}{4}\times\frac{2}{4}\times\frac{4}{4}\times\frac{0}{4}\times\frac{0}{4}\times\frac{3}{4} = 0 \)
   - \( B1*B2*B3*B4*B5*B6*B7 = \frac{0}{2}\times\frac{0}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{0}{2}\times\frac{1}{2}\times\frac{1}{2} = 0 \)
   - \( C1*C2*C3*C4*C5*C6*C7 = \frac{2}{12}\times\frac{1}{12}\times\frac{1}{12}\times\frac{3}{12}\times\frac{3}{12}\times\frac{2}{12}\times\frac{1}{12} = 0.005 \)
   - \( D1*D2*D3*D4*D5*D6*D7 = \frac{0}{18}\times\frac{0}{18}\times\frac{1}{18}\times\frac{1}{18}\times\frac{1}{18} = 0 \)
   - \( E1*E2*E3*E4*E5*E6*E7 = \frac{0}{4}\times\frac{0}{4}\times\frac{2}{4}\times\frac{0}{4}\times\frac{2}{4}\times\frac{1}{4}\times\frac{4}{4} = 0 \)

4. Then compare and identify the highest probability value with the results seen from the C value which is 0.005. This means it can be categorized in group 3.