

Predicting GPA in Entrepreneurship Study Program by Using Data Mining Technique

Ivan Diryana Sudirman^{*}, Iston Dwija Utama, Sherlyn

Department of Entrepreneurship, BINUS Business School Undergraduate Program, Bina Nusantara University, Bandung Campus, Bandung, Indonesia

Received April 22, 2020; Revised May 14, 2020; Accepted June 23, 2020

Copyright ©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Entrepreneurship education in universities tries to produce entrepreneurs through formal education. This relatively new study program faces many challenges in creating graduates who will become a successful entrepreneur. It is expected that with a good GPA, graduates of this study program have a higher chance of becoming successful entrepreneurs. This study tries to predict the GPA from existing datasets using data mining techniques with classification methods. There are several variables involved to predict GPA, namely, attendance, gender, school origin, motivation score, capability score, and observation score. For the last three variables, the data is taken from interview when the student went through an admission selection process. There are three algorithms tested, namely Naive Bayes, Decision Tree, and Deep Learning. Using RapidMiner, the result from model validation shows that Naïve Bayes algorithm has the highest accuracy for this dataset. The model shows that although the number of female students is smaller but the number of female students whose GPA is above 3.00 is higher than male student because females are more willing to attend lectures. Despite having good capability and attitude, if the students are reluctant to attend the class then it is likely that the GPA students will be below 3.00. This study provides an overview of the use of data mining techniques to predict GPA using the Naive Bayes algorithm. In addition, it is expected that educational institutions can pay more attention to the attendance of their students in class, create teaching strategies that are able to make students present in class, especially for male students.

Keywords Entrepreneurship, Education, GPA, Data Mining

1. Introduction

Indonesia is a developing country that currently needs more entrepreneurs. According to Bikse (Bikse et al., 2014), Entrepreneurship is a potential for economic development and the growth of modern society. According to McClelland, Ideally, meeting people's welfare requires at least 2% of the entrepreneurial population. The requirement for a developed country is to have an entrepreneur of at least 2 percent of the total population. Although the number of entrepreneurs in Indonesia reaches 3.1 percent of the total population, it is still inferior to Malaysia and Singapore (Post, n.d.). Therefore the Indonesian government encourages informal and formal entrepreneurship education. The programs have been instilled in various tertiary institutions. Lately, entrepreneurship education has been introduced in elementary schools.

In Bandung, Indonesia, Bina Nusantara University, Bandung campus is one of the institutions that have opened an entrepreneurship undergraduate program. The program called Creativepreneurship is supported by several courses and practicums that are believed to be able to produce new entrepreneurs. Entrepreneurship as a study program in universities is relatively new in Indonesia, so various challenges must be faced.

In Bandung campus, the efforts to achieve the goal of creating new entrepreneurs began from the beginning. From the registration process, interviews were conducted with the prospective students as part of the selection process. So not only using a test as a selection but also in-depth interviews. Prospective students who do not pass the interview cannot be accepted as students.

There is evidence that entrepreneurial orientation is positively related to entrepreneurial intention (So et al., 2017). So interviews were conducted face to face, one on one with the lecturer. Three things were explored, namely motivation, capability, and attitude. Interviews and

examinations are conducted to ensure that prospective students will be able to follow the program while in college and become an entrepreneur. In this study, the variables asked in the interview will be tested against the student's GPA, along with various other variables such as absence. This research is done to see what variables can affect student GPA from the dataset available. It is expected that students with a high GPA will be able to become successful entrepreneurs after graduating later.

However, many factors affect a student's GPA, and this study tries to uncover several factors that might affect the GPA. That way, it is expected to provide insight for campuses, especially campuses, with entrepreneurship study programs in educating their students. Data processing will be performed using data mining. There are several models to be tested, namely Naive Bayes, Decision Tree, Regression, and Deep learning. In order to get the model, which is the ablest to explain the available datasets. The results of the model that are considered the most suitable will then be studied further.

2. Literature Study and Method

The Office of Advocacy of Small Business Administration surveyed nearly 5,600 undergraduate and graduated active students and graduates from five U.S. educational institutions. To classify the styles of classes, instructional methods, and learning resources, which were most suitable for groundbreaking postgraduate business practices (Summit Consulting, 2009). The study concluded that, while there was "no noticeable association" among roles in entrepreneurship and education exams and assessments (calculated by GPA and SAT scores), some students who took an entrepreneurship program were much more creative. Research by Bae, Qian, Miao, and Fiet (Bae et al., 2014) emphasizes the slight connection between entrepreneurship and entrepreneurship education. Their survey highlights the need for entrepreneurship education and analysis to recognize the factors that determine the success of company start-ups.

Innovation in the classroom must deter students from continuing to struggle with an entrepreneurial aim. There is no single approach or mechanism to successfully grow entrepreneurs. Entrepreneurship development is very complicated in the fields of pedagogy, goals, college-based, and services. (Fayolle et al., 2006). The programs vary widely between places and universities. The quality and objective of the curriculum can also affect the output of entrepreneurial programs (Hannon, 2006).

Entrepreneurship education is considered necessary in the development of new entrepreneurs. Because an entrepreneur with a company that will last a long time need more creative thinking to adapt with the business environment. Thus entrepreneur with entrepreneurship education background have more creativity, thus have

more capability to survive the competition. Although the GPA score is not related to the success of being an entrepreneur, it is generally known the relationship between GPA and the performance of a person in the company. That is why one of the requirements for working in a company is that a person must pass a certain GPA score. Good performance can also be expected to occur to entrepreneurs who have high GPA scores. While success or failure of an entrepreneur is more dependent on many things and also the standard of the success itself. A person's performance and success depend on many factors, perhaps including a person's educational background which can be measured by his GPA score.

While the GPA is not linked to the success of an entrepreneur, there are many studies that try to look for factors that affect GPA. A study by Pepe (2012) explained that higher study skills also present in the students with higher GPA, therefore, there is a linear parallelism between study skills and GPA. Another study shows that high school GPA impacts the university GPA greater than pre-university tests, and enrolled year has a surprising effect on the university GPA (Hassan & Al-Razgan, 2016). Another study shows that there is a positive correlation between the musical aptitude exam scores and freshman year GPAs of the students (Demirbatir et al., 2015).

The development of information technology, integrated networks, and organizations has seen an immense rise in data over the last few centuries. This trend is also aligned with a rapid decline in data storage and processing costs. These structures, such as online businesses, social media, and mobile apps, show a wide variety of data that are waiting for analysis. Data mining is a way of handling vast volumes of data by using highly advanced algorithms to test data models with several parameters. Data mining is typically a way to produce insights from data gathered, compiled, and analyzed by an organization (i.e., valuable information or knowledge). Organizations use a wide variety of data analytics approaches to understand their buyers and activities better and understand complex business problems (Sharda et al., 2018)

Data mining generally starts with data ranging from a simple range with few numbers to a massive array containing thousands of metrics with trillions of observations. Many powerful computer techniques are employed to analyze data for the detection of applicable and usable data structures. These computing approaches have been developed from statistics, machine learning, and artificial intelligence. Extracting relevant information deals with various associated disciplines, including database systems, cleaning of information, displays, data testing, and performance analysis (Kotu & Deshpande, 2015).

The problem of data mining may typically be categorized as managed or unsupervised models of training. Supervised or directed data mining aims to infer a function or link to new, unlabeled data based on labeled training data. Supervised models predict performance values relying on multiple input variables. A model by a training

dataset with previously specified input and output parameters is designed for this objective. The model extrapolates the relationship between input and output variables and suggests that only input variables for the data gathering are known. The predicted output attribute is often referred to as a class or goal. To learn the model from the data under supervision through data mining, a significant amount of labeled data are necessary. Unsupervised or uncontrollable data mining reveals unlabeled algorithms of data. Through unsupervised data mining, indeed, no output factors can be predicted. This class of techniques intends to identify data behaviors based on relationships between datasets themselves. Any implementation involves both supervised and unsupervised learners. Classification, regression, search for correlations, identification of anomalies, time series, and text mining ventures are also challenges in data mining. (Kotu & Deshpande, 2015).

Data mining deals with recent analysis, recognition, and organizational data from interactions and trends. Data mining is, therefore, defined as a method of detecting data behaviors that could produce unknown pattern predictions (Witten et al., 2011). Data mining and basic statistical modeling are utilized in multiple fields. Conventional statistical approaches mainly focus on assessing reported hypotheses, and data exploration explores specific potential hypotheses which are often unknown (Witten et al., 2016)

By incorporating the statistical and data mining approach, we increase our comprehension of the ever-increasing number of digital data. In years ahead, data mining and computational approaches must be embedded, not only for the study of different various data sets as well as for the integration of subjects and methodologies, including information processing, knowledge building blocks, artificial intelligence, and machine learning algorithms.(2011).

A wide range of techniques is accessible for conducting data mining projects, namely classification, regression, clustering, and association. Many data mining analytical tools are using more than one technique (or algorithm) for each of these approaches. In this study, the data mining technique to be used is classification, and classification is undoubtedly the most commonly utilized data mining approach for real-world problems. As a standard component of the machine learning category of approaches, classification recognizes behaviors from past data to place new instances (with unknown labels) in their respective groups or classes. Classification explores the relation between other features of objects and the participation across a supervised training mechanism where all categories of factors (input and output) are introduced to the algorithm. (Sharda et al., 2018).

When evaluating the model, multiple variables are addressed, such as the following.

- **Speed.** The computational costs associated throughout the development and the use of the model, whereby quicker is perceived to be better.

- **Robustness.** The capacity of the algorithm to render relatively accurate projections based on incomplete data or data containing missing and incorrect elements.
- **Scalability.** The capacity to build a predictive model efficiently provided a relatively massive amount of data.
- **Interpretability.** The degree of comprehension and knowledge given by the model. For example, how and/or what the model ends with such predictions.

When examining the predictive performance of several or even more approaches, the analysts would use a technique called k-fold cross-validation to eliminate the bias linked with a sampling of training and holdout data sets. A complete data set is randomly divided into k mutually unique sets of data of equivalent size in k-fold cross-validation, often termed rotation estimation. K times are trained and evaluated in the classification approach. It is trained on all but one fold every cycle, and then measured on the next one fold. As seen in the corresponding equation, the cross-validation of an entire model's accuracy is determined clearly by multiplying the single measurement accuracy k:

$$CVA = \frac{1}{k} \sum_{i=1}^k A_i$$

For which CVA accounts for cross-validation accuracy, k is the total of folds being used, A is the precision value for each fold (e.g., hit rate, sensitivity, specificity) (Sharda et al., 2018).

Many algorithms can be used in data mining projects. Some of them are Naive Bayes, Decision Tree, Deep Learning, Neural Network, etc. In this research, the algorithm that will be tested is Naive Bayes, Decision Tree, and Deep Learning.

Analysts and professionals have introduced various methods (workflow or simple, stage-by-stage mechanisms) to maximize their likelihood of success in the production of data mining activities. An improper method of data mining was established in the early nineties by a European Group of companies, CRISP-DM. (Sharda et al., 2018). CRISP-DM might be the standard approach most frequently used. It is a six-phased framework that starts with a comprehension of the business and the need for a project in data mining. Because the sequences overlap, much obfuscation typically occurs. Data mining can be quite iterative since it is a method of understanding and evaluation based on the researcher's issue and understanding. More measures should be taken to ensure that the entire process does not go badly since more projects rely on the consequences of past acts.

1. Business Understanding

The critical aspect of designing a data mining thesis is to understand what the research project is all about. Responses to the need for management to acquire new

knowledge and a clear business overview of the study to be carried out will help to resolve this issue.

2. Data Understanding

The use of data mining is focused on single business operation, and several business operations need specific data sets. The primary task of data mining is to recognize data from different credible sources in the context of company acknowledgment. Many vital points need to be taken into consideration in the phase of data identification and compilation. In order to better understand the most relevant information, the analyst should be bright as to the scope of the data mining task.

In this study 384 data were used originating from the first years and second years students. The data is taken from the results of the interview when a student first registered as a prospective student. Then the dataset is combined with data after students have done lectures such as attendance data and GPA data. When the sample students enrolled in the past, they went through an interview process to be assessed in terms of their motivation, behavior and capabilities. This interview assessment data along with attendance data and GPA are then processed using data mining techniques.

3. Data Preparation

The purpose of the data preparation is to analyze the data ready for data mining and assessment, as mentioned in the prior section. Compared to other CRISP-DM data pre-processing steps, researchers believe that this constitutes approximately 80 percent of the cumulative data mining activity. This intense effort is evident by the fact that real-world data is not accurate (importance attribute deficiencies or aggregate data only) and unclear (comprising inconsistency or different attributes). As explained earlier, the data are grouped into 2 GPA groups, namely above 3.00 and below 3.00. The amount of data from the two groups is quite balanced so that it is good for classification

4. Model Building

Each method chooses and implements multiple modeling approaches for the given collection of data to fulfill the particular organization needs of the project. The numerous models created during the model development process are frequently assessed and examined. Despite the comprehensive understanding of data mining practical application, an experimentation and assessment methodology should be applied to achieve the "appropriate" solution in a particular manner for any specific reason. In this study, tree algorithms are used to analyze existing datasets. The three algorithms are Decision Tree, Naive Bayes and Deep Learning.

5. Testing and Evaluation

The established models are validated and evaluated for

accuracy and reliability. This stage evaluates how well and to what degree the chosen model (or pattern) matches the goals and objectives (i.e., further models should have to be created and analyzed?). Some alternative is to evaluate the model(s) that are built into a real-world scenario if resources and budget constraints allow. In this study, the three algorithms are evaluated using k-fold cross-validation. Several parameters that we look for this process is the model accuracy score, classification error, and AUC.

6. Deployment

The phase of deployment is as easy as reporting and sometimes even data mining in real-time inside the organization; it depends on the requirements. Consumers are doing the deployment stage for some instances, not the data analyst.

RapidMiner will be used in this study to analyze the data. Rapidminer is sophisticated data science, machine learning, deep learning, text mining, and predictive analysis platform created by the Company of the same brand. This promotes various aspects of machine learning, namely data processing, outcomes assessment, model validation, and enhancement, and extends to software and industry implementations as well as science, teaching, rapid prototyping and creation of software (Hofmann & Klinkenberg, 2013). For convenience and convenience, researchers will use the auto-model function. The use of these functions will facilitate the comparison of algorithms.

3. Result and Discussion

The first step is to understand the business or the organization. As explained earlier, entrepreneurship education has a role in creating new entrepreneurs. In the Creativepreneurship study program, Bina Nusantara University, Bandung campus, efforts to create new entrepreneurs started from the selection process. Admission selection not only uses written tests but also conducts an interview test to explore the prospective students. This interview is based on the perspective that if a student graduate from the Creativepreneurship study program, it requires not only analytical skills but also the capability as an entrepreneur, motivation to become an entrepreneur, and also behavior. These three things are explored through three interview variables, namely competency, motivation, and observation.

Through various entrance selection tests, it is expected that future students can graduate well from Creativepreneurship and become successful entrepreneurs. During lectures, students are equipped with various courses needed to become a successful entrepreneur. Some of these courses include an introduction to entrepreneurship, critical thinking, design thinking, marketing, business statistics, and so on.

Although the admission selection and interview were conducted, there were still students whose GPA scores were not good enough. By using available datasets and data mining techniques, we try to uncover factors that might contribute to a person's GPA score.

The next step is understanding the data. Data attributes consist of numbers, binusian, student ID number, binusian ID, temporary GPA, GPA, the origin of the school, the type of school, gender, capabilities score, dead score, observation score, total score, and absence.

Capability score is obtained from interviews, the factors measured in this capability variable are the courage to take risks, hard work, and discipline, confidence, financial management, having planning, the ability to adapt, the ability to analyze the market. The weight of this capability score is 20%. Motivation scores were also obtained from interviews; for this variable, there was only one question item, namely how much the desire of prospective students to become entrepreneurs and the reasons for wanting to become entrepreneurs. The weight for this score is 30%. As for the observation score, measured through several factors, namely, eye contact, communication, not aggressive, good interaction, can receive opinions, pleasant manner, polite. The weight for this score is 50%.

After understanding the data, the next step is to prepare the data. Regarding the purpose of the research, some

attributes will not proceed to the next process, namely the name of the student, student ID number, binusian ID number, temporary GPA, and total score. The type of school was excluded because there is not enough public type in the data.

For GPA, it is divided into two groups, those with a GPA below 3.00 and those with GPA above 3.00. The GPA 3.00 limit was taken because the minimum GPA limit for hiring in good companies is 3.00 ("New Meeting Form in the Digital Working Environment Era," 2019). Although the Creativepreneurship study program is more focused on creating entrepreneurs, by following the recruitment standards in large companies, it is expected that students with GPA above 3.00 have the potential to succeed in their careers either as intrapreneurs or entrepreneurs.

After the data is ready, then the data is imported into RapidMiner, and by using the retrieve operator and select attribute operator like in the figure 2, we can see a description of the data to be processed.

From figure 2, we can see the names of the attributes that will be processed further. There is no missing value, and there are three attributes with polynomial data types. For GPA attribute > 3.00, which is used as a label, the amount of data recorded as "No" and "Yes" can be said to be balanced, only 10 records differences.

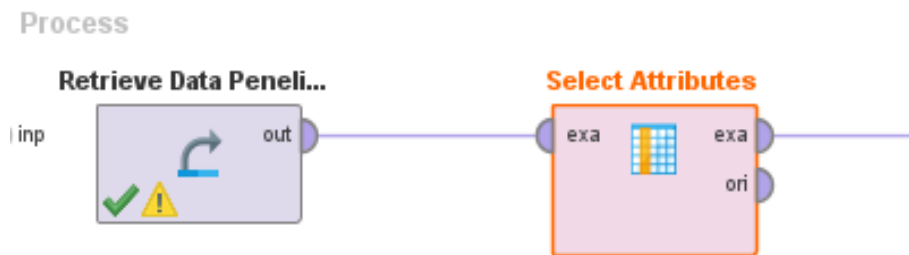


Figure 1. RapidMiner Operators

Name	Type	Missing	Statistics
Asal Sekolah	Polynomial	0	Least: SMU KARTIKA III-1 (1) Most: SMAK 2 B [...] dung (34)
Gender	Polynomial	0	Least: Female (180) Most: Male (204)
Score Capabilities	Real	0	Min: 0.107 Max: 0.371
Score Motivation	Real	0	Min: 0.200 Max: 0.400
Score Pengamatan	Real	0	Min: 0.200 Max: 0.500
Absence	Integer	0	Min: 0 Max: 7
GPA>3.00	Polynomial	0	Least: No (187) Most: Yes (197)

Figure 2. Data Description

For the modeling, testing, and evaluation stages, we use the auto-model function for ease and comfort. The function offered by RapidMiner is one of the advantages of the application. There are three models to be evaluated, namely Naive Bayes, Decision Tree, and Deep Learning, because they all can handle Polynomial data quite well.

Classification Error

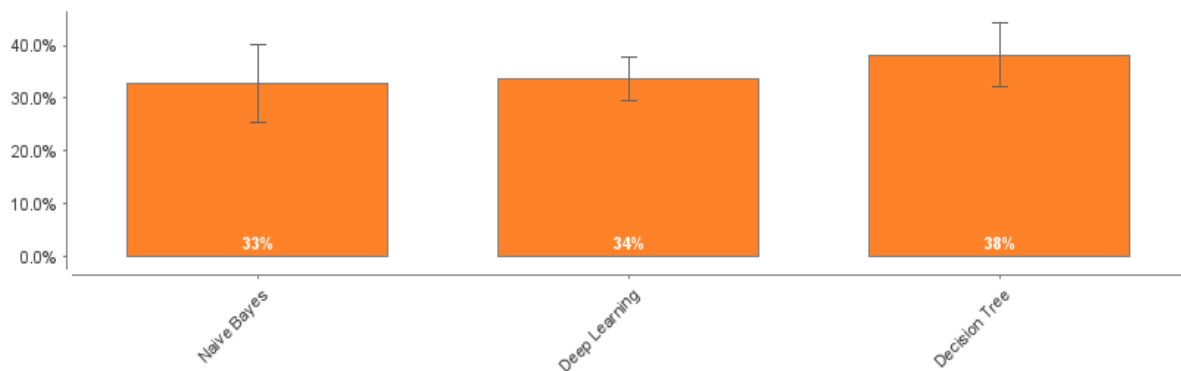


Figure 3. Model Evaluation

As shown in Figure 3, among the three models tested, Naive Bayes has the lowest classification error when compared to Deep Learning and Decision Tree. Low classification error also means the highest accuracy. Accuracy is generally used as a reference to determine the best model, as well as in this study. Therefore, further results from the Naive Bayes model will be explored further.

Table 1. Naive Bayes Performance

Performances

Criterion	Value	Standard Deviation
Accuracy	67.3%	± 7.5%
Classification Error	32.7%	± 7.5%
AUC	71.6%	± 11.2%
Precision	65.0%	± 12.2%
Recall	71.4%	± 9.9%
F Measure	67.5%	± 8.8%
Sensitivity	71.4%	± 9.9%
Specificity	63.7%	± 14.0%

Table 1 provides information about the performance of Naive Bayes. Accuracy is the roughly comparable amount of successfully identified examples or, in other phrases, the proportion of accurate predictions. (*Performance (Classification) - RapidMiner Documentation*, n.d.). An accuracy score of 67.3% can be considered quite right because the object of research is human behavior. For

recall, it means the same as sensitivity, it shows a good enough score. Both of them use the same formula (true positive predictions) / (number of positive examples). This shows that the model is able to predict true positive values accurately. For specificity, the value is also good enough where specificity is the ratio between true negative predictions and the number of negative examples.

As can be seen in table 2, the confusion matrix. The model predicts "Below 3.00" and turns out to be true "Below 3.00" as many as 49, while the model predicts "Above 3.00" and turns out to be "Above 3.00" as many as 23. The model predicts "Below 3.00," but it should "Above 3.00" as many as 30, while the model predicts "Above 3.00," but it turns out "Below 3.00" as many as 9.

Table 2. Confusion Matrix

	True Below 3.00	True Above 3.00	Class prediction
Pred Below 3.00	36	15	71.59%
Pred Above 3.00	21	38	64.41%
Class recall	63.16%	71.70%	

Figure 4 shows that the density for females above 3.00 is higher when compared to below 3.00. For males, the GPA below 3.00 is higher when compared to GPA above 3.00. This result becomes interesting, especially if we pay attention to the data used in this research, the female sample is smaller than male. However, the number of women with a GPA above 3.00 is higher. This result raises the question, is the teaching system being conducted more suitable for women? We need further research related to this.

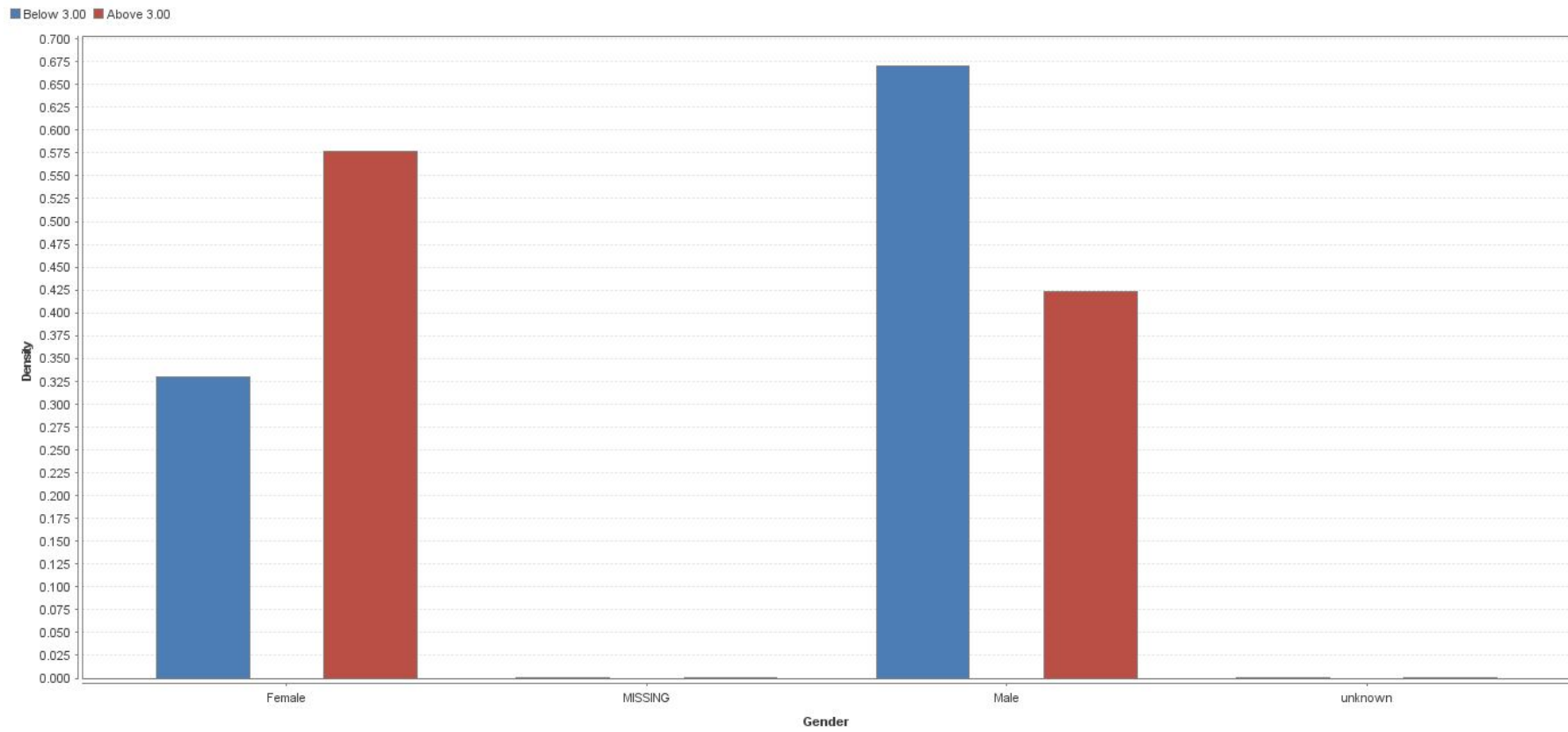


Figure 4. Gender-Density Model

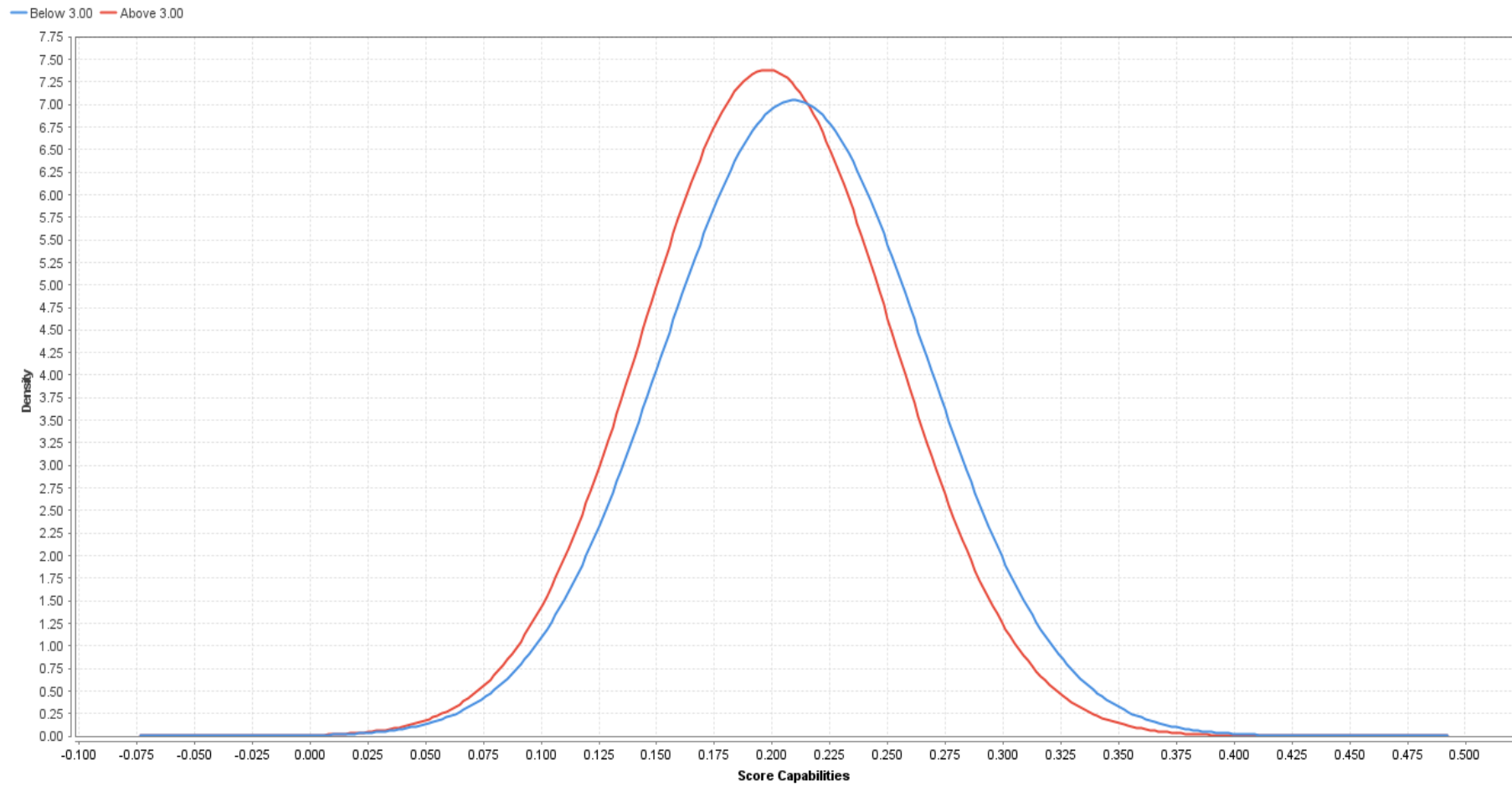


Figure 5. Density – Capability Model

For capability, the distribution can be seen in figure 5. Those who have a GPA above 3.00 have a slightly lower score capability than those with a GPA below 3.00. The Above curve is 3.00 peak at 0.2 in score capability, while the Below curve is 3.00 peak between 0.2 and 0.225.

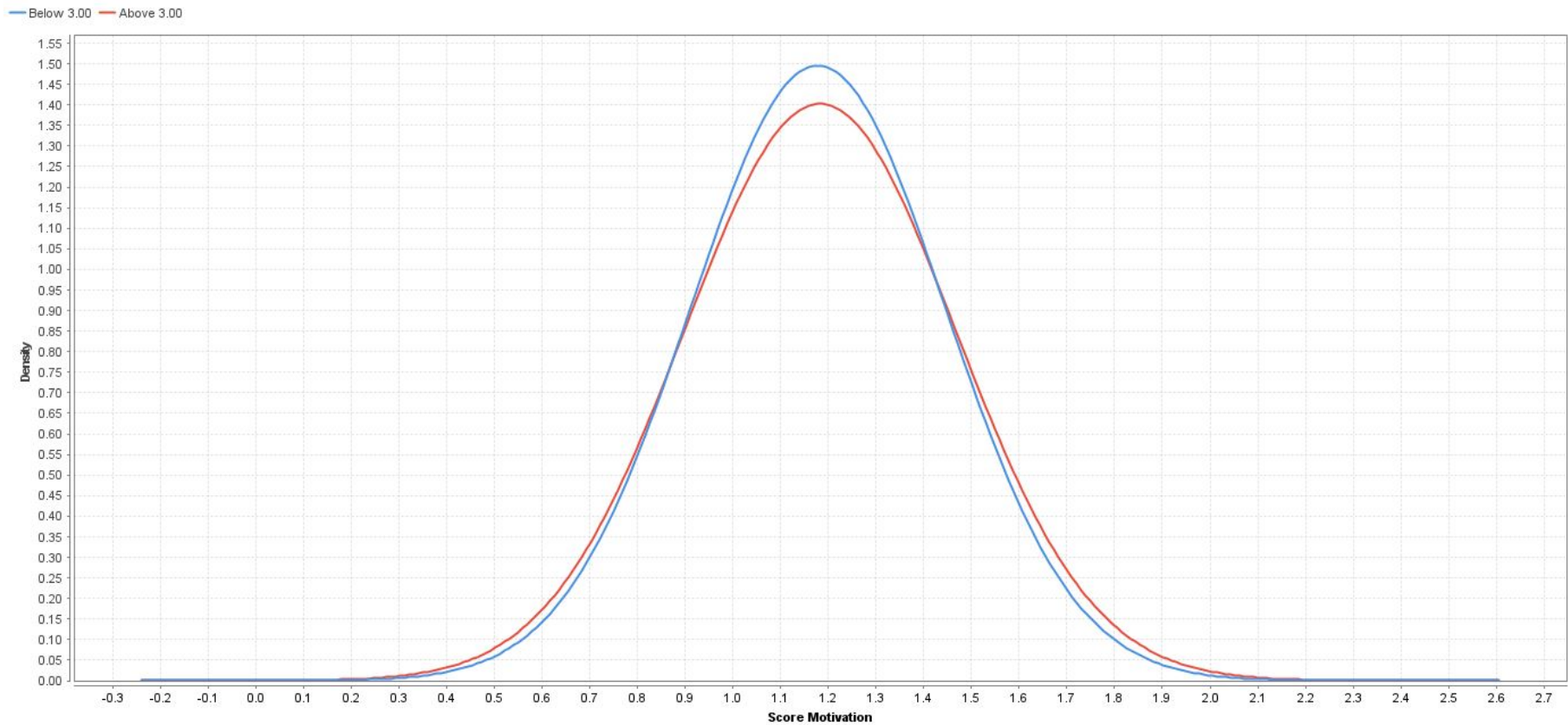


Figure 6. Density – Motivation Model

For motivation, the distribution can be seen in Figure 6. Both groups, Below 3.00 and Above 3.00, have almost the same distribution. With the peak of the Below 3.00 curve slightly higher than the Above 3.00 curve

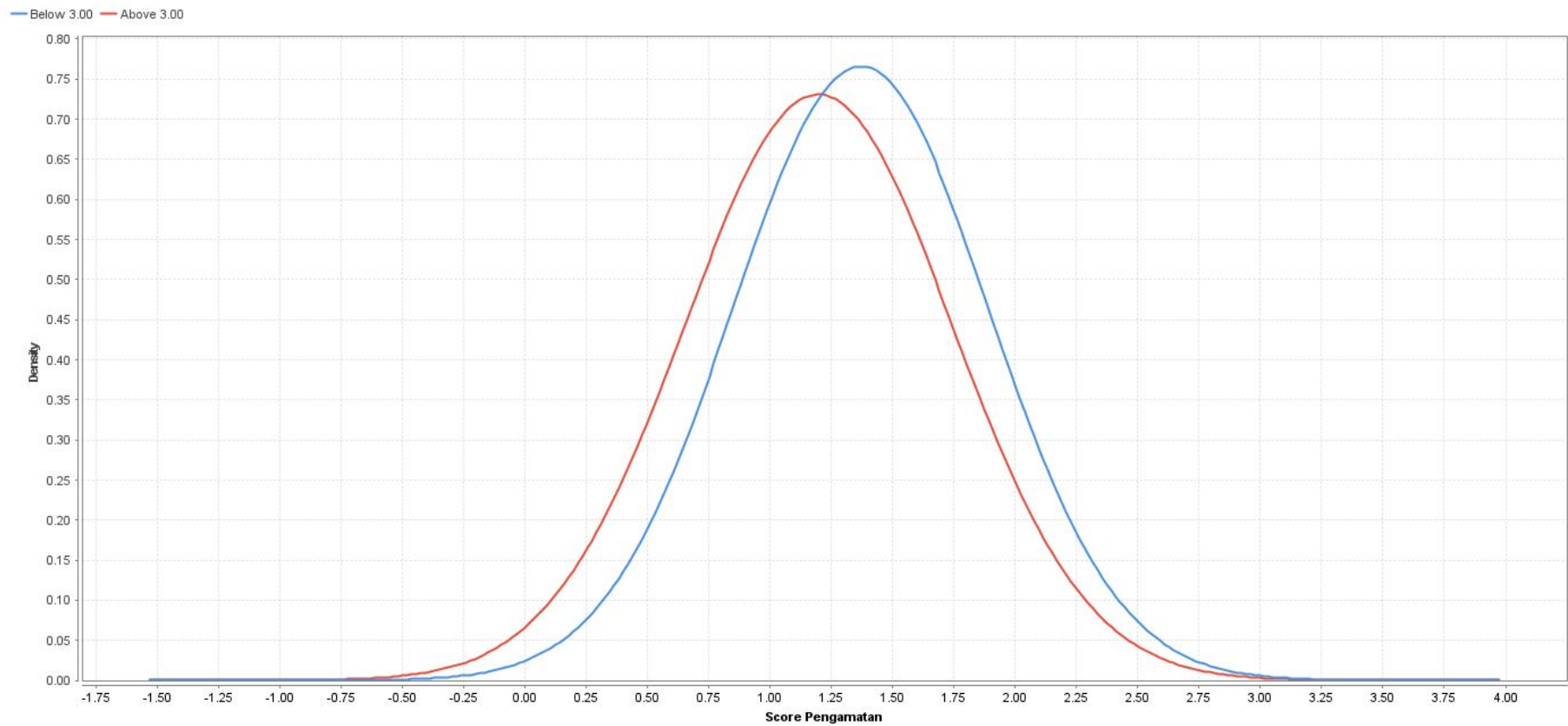


Figure 7. Density – Observation Model

For observation, the distribution can be seen in Figure 7. The Above 3.00 distribution is slightly lower at the peak of the curve than Below 3.00. For observation, the distribution of Above 3.00 is slightly lower than Below 3.00.

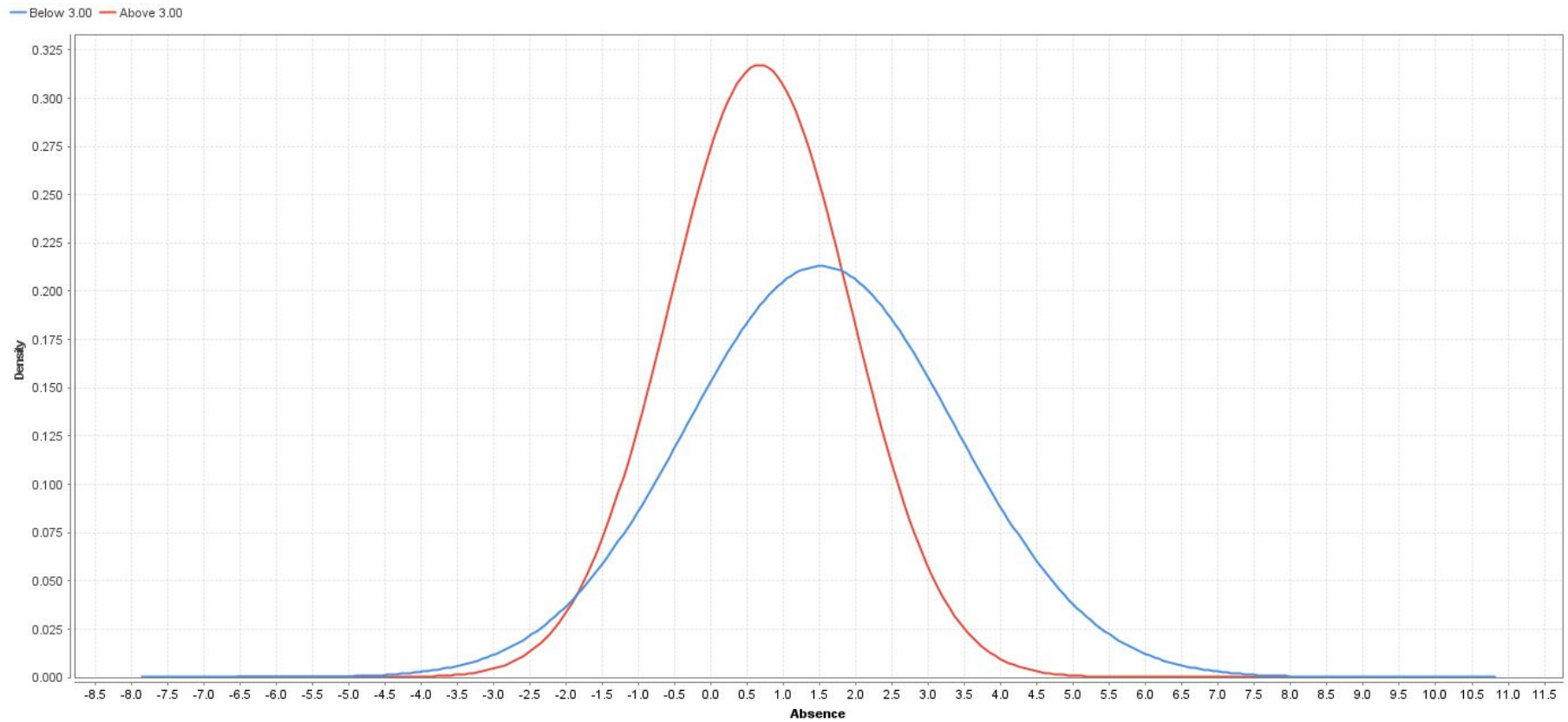


Figure 8. Density – Absence Model

For the distribution of attendance from the two groups, it can be seen in Figure 8. The difference is quite striking between the Below 3.00 group and Above 3.00 for the percentage of attendance. Above 3.00 is more likely present in class during lectures than below 3.00.

Attribute	Weight
Gender	0.503
Absence	0.204
Score Capabilities	0.100
Score Pengamatan	0.050
Score Motivation	0.028

Figure 9. Attribute Weight.

Figure 9 shows that the highest weight that determines a student is included in one group is the gender attribute, then the absence attribute, then the capability attribute, then the observation attribute, and finally, the motivation attribute.

Naive Bayes - Simulator

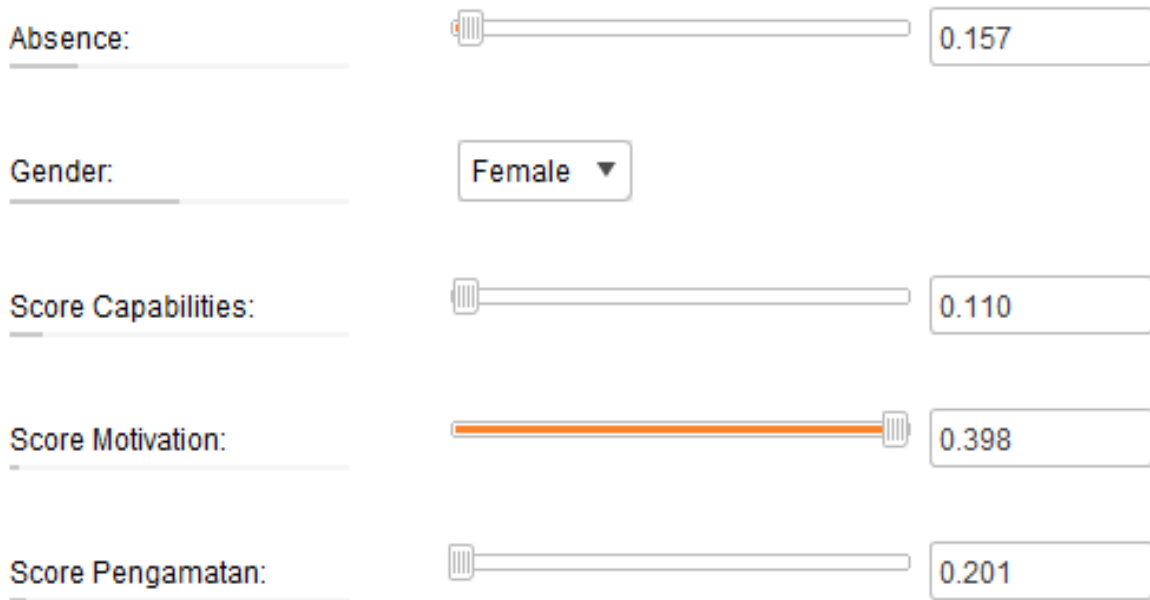


Figure 10. Simulator

One of the advantages of RapidMiner is the ease of simulating the model. Figure 10 shows the results of the Naive Bayes simulation model for the data used. Optimization simulations have been carried out for the Above 3.00 group. The results show that to increase the likelihood of obtaining a GPA above 3.00, and then the absence attribute is at least 0.157. If multiplied by the number of 13 meetings, then it is equal to at least two times absent. The gender of the students is female, and the capability score is 0.11, the minimum motivation score is 0.39 while the observation score is 0.201.

Most Likely: Above 3.00

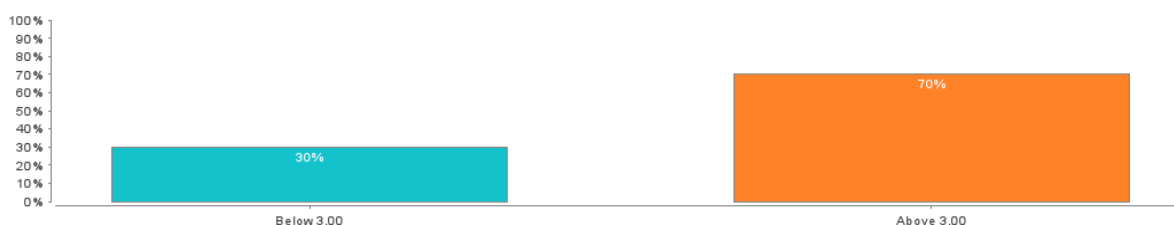


Figure 11. Simulation Result for Most Likely Above 3.00

Important Factors for Above 3.00

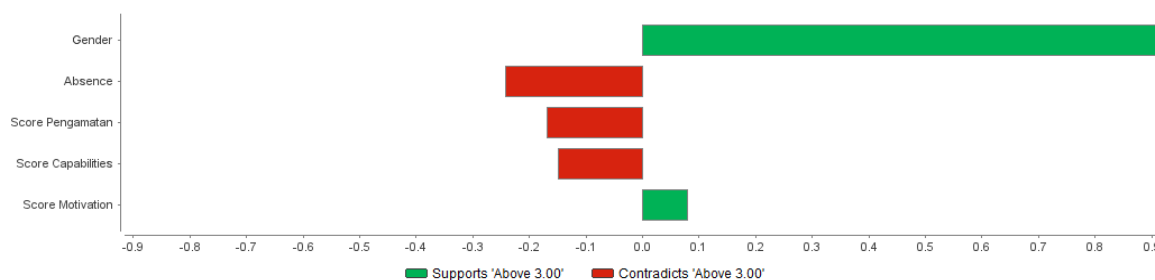


Figure 12. Important Factor for Above 3.00

So based on the important factors in figure 12, the Gender and Motivation Score are supportive attributes for the Above 3.00 group while absence in class is clearly in contradiction with that group. Because in the dataset, 0 means that the student always attend the class, 1 means the student did not attend 1 lecture, 2 means miss 2 weeks, and so on, thus the smaller the number, the greater the chance of being classified into groups above 3.00. For motivation, it is important for the interviewer to pay attention and assess student motivation when they enter undergraduate program.

However, what is interesting is the observation score, and the capability score of figure 12 turned out that both of the variable are contradict to the Above 3.00 group. This result can occur because of several possibilities, namely the measurements used can not explore the subject of the interview properly, the assessment made by the interviewer is not proper, difficulties when making an assessment, and many more.

4. Conclusions

This study tries to explore the datasets related to students studying in entrepreneurship study programs. When they first registered as prospective students at Creativepreneurship, they must through the interview stage. Then, after passing the interview process, the prospective students is accepted as Creativepreneurship student. They undergo the program and get a GPA score. Data from the interviews and data during their study years are then combined and expected to get insights from these data. GPA is divided into two groups, namely Above 3.00 and

Below 3.00. Using several models, researchers try to find the best prediction model. For the dataset, Naive Bayes has better accuracy than other models, so the model is studied more deeply. Naive Bayes model shows that for the data examined, gender and attendance are important factors in determining groups. Females with high-class attendance have a higher chance of getting a GPA above 3.00. It is possible the reason why there are more female with GPA above 3.00 is that they are more willing to attend lectures than male students.

Three variables are tried to be measured using interviews, namely motivation, capability, and observation, or attitude. While we do not use exam test, this interview also can be considered like an collage entrance test like in the Hassan and Al-Razgan study (2016). Of these three variables, motivation scores can support predictions for the Above 3.00 group, while the other two scores contradict. Even though capabilities and attitudes surely affect student success during lectures. However, in this study both of the variable does not match with the Pepe (Pepe, 2012) study, because these two variables are actually included in study skills.

There are two possibilities, first, these results might indicate that despite having good potential in terms of capability and attitude, if the students are reluctant to attend the class then it is likely that the students GPA will be below 3.00. Another possibility is, this study shows the difficulty of measuring capabilities and attitudes using limited interviews. So the interviewer need to have a good judgment when measuring those variable. This study provides an overview of the use of data mining techniques to predict GPA using the Naive Bayes algorithm. In addition, it is expected that educational institutions can pay

more attention to the attendance of their students in class, create teaching strategies that are able to make students present in class, especially for male students. The use of interviews to explore the potential of prospective students has a good intention. However, it may be necessary to consider the ability and experience of the interviewer to assess the potential of prospective students.

REFERENCES

- [1] Bae, T. J., Qian, S., Miao, C., & Fiet, J. O. (2014). The Relationship Between Entrepreneurship Education and Entrepreneurial Intentions: A Meta-Analytic Review. *Entrepreneurship: Theory and Practice*, 38(2), 217–254. <https://doi.org/10.1111/etap.12095>
- [2] Bikse, V., Riemere, I., & Rivza, B. (2014). The Improvement of Entrepreneurship Education Management in Latvia. *Procedia - Social and Behavioral Sciences*, 140, 69–76. <https://doi.org/10.1016/j.sbspro.2014.04.388>
- [3] Demirbatir, R. E., Celiktas, H., & Engur, D. (2015). A Comparative Examination of the 2013 Musical Aptitude Test Scores of Music Education Students and their 1st Academic year GPAs. *Procedia - Social and Behavioral Sciences*, 197, 815–820. <https://doi.org/10.1016/j.sbspro.2015.07.193>
- [4] Fayolle, A., Gailly, B., & Lassas-Clerc, N. (2006). Assessing the impact of entrepreneurship education programmes: A new methodology. *Journal of European Industrial Training*, 30(9), 701–720. <https://doi.org/10.1108/03090590610715022>
- [5] Hannon, P. D. (2006). Teaching pigeons to dance: Sense and meaning in entrepreneurship education. *Education + Training*, 48(5), 296–308. <https://doi.org/10.1108/00400910610677018>
- [6] Hassan, S. M., & Al-Razgan, M. S. (2016). Pre-University Exams Effect on Students GPA: A Case Study in IT Department. *Procedia Computer Science*, 82, 127–131. <https://doi.org/10.1016/j.procs.2016.04.018>
- [7] Hofmann, M., & Klinkenberg, R. (Eds.). (2013). *RapidMiner: Data Mining Use Cases and Business Analytics Applications* (1 edition). Chapman and Hall/CRC.
- [8] Kotu, V., & Deshpande, B. (2015). *Predictive analytics and data mining: Concepts and practice with RapidMiner*. Elsevier/Morgan Kaufmann, Morgan Kaufmann is an imprint of Elsevier.
- [9] New Meeting Form in the Digital Working Environment Era. (2019). *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 2700–2705. <https://doi.org/10.35940/ijitee.A4849.119119>
- [10] Pepe, K. (2012). A Research of the Relationship Between Study Skills of Students and their GPA. *Procedia - Social and Behavioral Sciences*, 47, 1048–1057. <https://doi.org/10.1016/j.sbspro.2012.06.777>
- [11] *Performance (Classification)—RapidMiner Documentation*. (n.d.). Retrieved April 19, 2020, from https://docs.rapidminer.com/latest/studio/operators/validateon/performance/predictive/performance_classification.html
- [12] Post, T. J. (n.d.). *Ministry targets entrepreneurship rate to hit 4% this year*. The Jakarta Post. Retrieved April 16, 2020, from <https://www.thejakartapost.com/news/2017/03/22/ministry-targets-entrepreneurship-rate-to-hit-4-this-year.html>
- [13] Sharda, R., Delen, D., & Turban, E. (2018). *Business intelligence, analytics, and data science: A managerial perspective* (Fourth edition). Pearson.
- [14] So, I. G., Ridwan, A., & Aryanto, R. (2017). *Confirming Entrepreneurial Orientation Dimensions and Linking It With Entrepreneurial Intention Among Business Students in Indonesia*. 23.
- [15] Summit Consulting, L. L. C. (2009). Toward effective education of innovative entrepreneurs in small business: Initial results from a survey of college students and graduates. Retrieved September, 20, 2017.
- [16] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3 edition). Morgan Kaufmann.
- [17] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4 edition). Morgan Kaufmann.