

Prediction of Student Entrepreneurship Future Work based on Entrepreneurship Course using the Naïve Bayes Classifier Model

Hanapi Hasan¹⁾, Asmar Yulastri^{2)*}, Ganefri³⁾, Tansa Trisna Astono Putri⁴⁾, Rizkayeni Marta⁵⁾

¹⁾Mechanical Engineering Education Department of Universitas Negeri Medan, Indonesia,

²⁾Catering Department of Universitas Negeri Padang, Indonesia,

^{3,5)}Electro Engineering Department of Universitas Negeri Padang, Indonesia,

⁴⁾Information Technology and Computer Education Study Program of Universitas Negeri Medan, Indonesia

¹⁾hanapi_hasan@unimed.ac.id, ^{2)*}asmaryulastri@pp.unp.ac.id, ³⁾ganefri@ft.unp.ac.id,
⁴⁾tansatrisna@unimed.ac.id, ⁵⁾rizkayeni.marta@ft.unp.ac.id

Submitted : Dec 25, 2023 | Accepted : Dec 29, 2023 | Published : Jan 10, 2024

Abstract: Entrepreneurs are critical to a country's economic progress and job creation. Few people felt schools have much to offer with business a generation ago. Students are expected to be an entrepreneur as the outcome of the course. The goal of this study is building a model to predict students' future employment, particularly in the field of entrepreneurship, using big data analysis and data mining. Various educational institutions can use data mining methodologies to identify hidden patterns in data contained in databases. The feature selection technique was utilised in this study to select and assess the significance of each element. The model was built using the final parameters determined by the feature selection technique (Correlation Based Feature Selection). Using the 10-fold cross validations for training and testing dataset distribution, the Naïve Bayes classifier was used to forecast the students' future of work. The dataset for the study was gathered from a student's performance report at Universitas Negeri Medan's engineering department. The effectiveness of using feature selection algorithms was compared to the effectiveness of not using feature selection algorithms, and the results are discussed. According to the findings of this study, the accuracy of Naïve Bayes with Correlation Based Feature Selection is 87.4%, which is higher than the model that did not use any feature selection. It was also discovered that the overall accuracy of the Correlation Based Feature Selection and Naïve Bayes Classifier models appears to be higher than that of the other treatments.

Keywords: Prediction model; students' future work; entrepreneurship; Naïve Bayes; Correlation Based Feature Selection

INTRODUCTION

Compared to many ASEAN nations, Indonesia continues to have a higher open unemployment rate. In 2019, Indonesia's open unemployment rate was at 5.3%. The unemployment rates in Singapore and Malaysia were 3.3% and 3.1 percent, respectively, in the same year. In the meanwhile, the Philippines' open unemployment rate, at 2.2%, is comparatively low. Thailand has the lowest unemployment rate of any country, at 0.7%. Given that Indonesia has a far higher population than the ASEAN nations, it makes sense. With respect to population, China, India, and the United States are followed by Indonesia. Entrepreneurship has the potential to positively improve Indonesia's financial growth. The more

*Asmar Yulastri



businesses there are, the more resilient the national economy will be. Entrepreneurs can offer an appealing solution to reducing, if not eliminating, the increasing number of unemployed people. Of course, as the amount of entrepreneurs grows, so will their resistance to the financial crisis. According to (Statistik, 2019), the open jobless rate in Indonesia is relatively high. Indonesia's general unemployment rate reached 7.05 million, or 5.28% of the total workforce of 133.56 million. Diploma I, II, and III graduates made up 42%, diploma I, II, and III students made up 5.99%, and university graduates (at least S-1) made up 5.67%. One strategy to lower the open jobless rate is to establish a business-minded mentality as promptly as feasible. Predicting students' outcome in the future has become more difficult in an educational system due to the large volume of data and imprecise data with fuzziness in educational databases.

Big data analytics methods are used to process large amounts of data. In traditional systems, processing data with different factors and parameters is difficult, big data analytics, on the other hand, can assist companies in better understanding the information contained inside the data. It also assists them in identifying the data that is most critical for prediction and future decision making. The goal of evaluating student's future work is to assist them in developing individual student professionalism, encouraging self-improvement, maintaining achievements, and providing them with a forewarning about their level of skills in placements. It is also crucial in increasing the number of entrepreneurship in Indonesia. According to Sagie and Elizur (Bhunia & Shome, 2023), the desire for success is an impetus drive in carrying out obligated responsibilities flawlessly. Individuals who have a great desire for accomplishment are inclined to take part in entrepreneurship. In accordance with (McClelland & Jorba, 2022) in the field of motivation theory, high achievers with a high need for achievement are also moderate risk takers.

Entrepreneurs play an important role in the growth of an economy and creation of employment. A generation ago, few people believed institutions had much in common about business. The earned autonomy of the higher education institution obtained by the famous German theorist, a researcher, and professor, Alexander von Humboldt, ensured not only that the university was freed of the indicates of both the state and the church, but also ensured the worth and priority of 'knowledge for its own sake,' instead of information since it offers some benefit for society or the economy. But circumstances change, and the university has developed from von Humboldt's paradigm to what some call the business school or the entrepreneurial society's university over the last quarter-century (Audretsch, 2017). There is a wide and extensive literature indicating the critical role that universities may play in promoting innovative activity through technology transfer. However, the goal of this study is to show how to forecast the student's future work whether it will be related to entrepreneurship or not, in order to develop the desire of students in business world.

LITERATURE REVIEW

Naive Bayes is a fundamental statistical predictor that relies upon the Bayesian theorem, according to Bryson (2007). It is founded on the data set's independent functions or properties. A class attribute is unaffected by the existence or lack of any other component. The key advantage of this predictor is that it depends on fewer inputs for training to calculate the variable's mean and variance for classification. Naive Bayes approaches have been employed by several scholars in the area of Educational Data Mining to forecast how well students perform, and they are additionally being used in various fields. The goal of their paper (Nageswari et al., 2019) is to forecast student progress utilising the idea of mining approaches.

A paper published in (2014) by Elsevier entitled "Educational Data Mining: A Survey and a Data Mining-Based Analysis of Recent Works" provided an analysis of published research from 2010 to 2013 and categorised Educational Data Mining methods through systems of education, fields of study, assignments, techniques, and algorithms. In accordance to the writer, every single Educational Data Mining approach can be divided into six features: student simulation, behavioural simulation, examination; student achievement simulations; student assistance and input versus the syllabus-knowledge-sequencing with an emphasis on academic achievement (Peña-Ayala, 2014).

A study by Ahmad (2015) provides a methodology for predicting academic achievement in a Computer Science course for first-year bachelor students. The variables in the study contained the

students' statistics, prior grades, and family background information. Decision Tree, Naive Bayes, and Rule-Based categorization approaches are used to develop the best learning result forecasting model for pupils.

Nakra & Duhan (2019) examined Bayes Net, Naive Bayes, and a combination of these two classifiers to determine which one generated the best results using data from the WEKA tool from diabetes patients. Merging Bayes Net and Naive Bayes delivers greater accuracy than either predictor alone, according to the results.

Another study by Da et al. (2011) built a learner classifier using naive Bayesian classification. Bayesian classification theories, in general, have been discovered to be valuable analysis forms in the distant education system for forecasting possible data trends and determining sound decisions.

The study of business has been proposed as an effective method of promoting and increasing interest in entrepreneurship among undergraduates. Students that are exposed to this information may develop an encouraging mindset towards entrepreneurship (Basu & Virick, 2008). In recent decades, there has been increased interest in bringing business into education (Lackéus & Middleton, 2015). Entrepreneurial in education not just inspires students to create their own firms, but it also makes them more creative, open to new opportunities, aggressive, and imaginative. Learning about entrepreneurship aims to improve students' capacity and readiness to make a difference in the lives of others. It is the basis for business ownership and a talent that all students must have regardless of professional path. Learning about entrepreneurship typically concentrates on building the abilities required to start a business, and the curriculum for entrepreneurship as a whole is frequently self-oriented in terms of increasing independence, self-worth, innovative thinking, action-taking, and direction (Mahieu, 2006; QAA, 2012). Entrepreneurship is socially significant because it is "the most potential economic power that the world has ever experienced" (Kuratko, 2005). This is consistent with (Mayhew et al., 2012), who argued that a nation's financial stability is more important than the efficient implementation of technologies, despite the fact that creativity is crucial for economic progress.

According to Matlay (Matlay, 2008), Entrepreneurship courses vary in their content. It explored the development of entrepreneurial education, which had previously been integrated primarily into standard business curricula. He further asserts as the business development programme is available in different levels and lengths. The objective of education in entrepreneurship varies to the extent that it has an impact on the generation of business skills; nevertheless, the ability to promote and learn entrepreneurship is heavily dependent on. Entrepreneurship coursework at universities attempts to connect entrepreneurial characteristics with business procedures and business behaviour by offering an effective EEP (Heinonen & Poikkijoki, 2006). Entrepreneurship education focuses not only on the transfer of business and management knowledge, but also on changing students' mindsets by cultivating new modes of thought, behaviour, skill, and personality (Gibb & Hannon, 2006; Henry et al., 2005; Sánchez, 2011). It proves that some parts of business ownership can be learned and studied, debunks the idea that businessmen are created instead of made, and calls into question the notion that businessmen are produced instead of formed (Kuratko, 2005; Sarasvathy & Venkataraman, 2011).

Many academics think that in order to promote learning and innovative potential, higher education on entrepreneurship should contain an experimental learning approach as well as some type of interactive pedagogy (Collins et al., 2006; Honig, 2004; Vinten & Alcock, 2004). Gibb (2006) presents a teaching strategy that focuses on the presentation procedure, user involvement in learning, learning from failures, customisable objectives for learning, and session adaptability and flexibility to integrate theoretical understanding with business behaviours. Previous study has concentrated on the expanding number of university-based entrepreneurship initiatives. In addition, there is a movement to investigate the process of learning of entrepreneurship education curriculum (Block et al., 2023) as well as the programme content (DeTienne & Chandler, 2004; Honig, 2004; Shepherd, 2004). In recent times, there has been an evolution of seeking to integrate concepts, methods, and real-world examples of what businessmen do and how they grow (Harmeling & Sarasvathy, 2013).

METHOD

The technique of collecting information from huge amounts of data is known as data mining. The mechanism operates on a large dataset in which the student performance in the entrepreneurship course is evaluated in the end-of-semester examination. The study's framework is based on (Khan et al., 2013) Knowledge Discovery Process (KDP). Selection, Preprocessing, Transformation, Data Mining, and Interpretation comprise the KDP process scheme. Initially, input data and target data were separated. To assure database trustworthiness, pre-processing and conversion were performed, with data extraction providing as the main type of investigation. The knowledge discovery process's outcomes were interpreted. This research created an outline based on the KDP concept. The first stage: Data Preparation required data collecting and the creation of datasets collected from the undergraduate student's data; Stage 2: Preprocessing included the elimination of noisy and unnecessary data, as well as further transformation into mining-ready formats. Stage 3: Classification Process - this involved developing a model employing the Naïve Bayes algorithm to forecast students' future of work; Phase 4: Result - this involved data interpretation and evaluation, as well as identifying and presenting to the user interesting patterns representing knowledge-based.

Data collection

From 2021 to 2023, student data were collected from the Universitas Negeri Medan using the sampling method of the engineering department. There were 144 records and thirteen (13) attributes in the dataset. The predictor variables were included in the dataset and were obtained from the entrepreneurship course registration. The variables that were chosen had a substantial impact on the investigations given in the associated articles. The variables in the student's information that generated from registration are shown in Table 1.

Table 1. Student's Attribute

No	Name	Possible Values
1	Gender	Male, Female
2	Class	A, B, C
3	Age	18, 19, 20
4	Living Location	Village, City
5	Where do you stay	Hostel, House with family
6	Number of Members in a Family	2, 3, >3
7	Parent's Annual Income	> Rp. 5.000.000, < Rp. 5.000.000
8	Father's qualification	No-Education, Elementary, Secondary, Graduate, Postgraduate, Doctorate, N/A
9	Mother's qualification	No-Education, Elementary, Secondary, Graduate, Postgraduate, Doctorate, N/A
10	Father's Occupation	Farmer, Business, Government, Retired, Not-Applicable
11	Mother's Occupation	Farmer, Business, Government, Retired, Not-Applicable
12	Final score of Entrepreneurship Course	A, B, C, E
13	Current Job	Entrepreneur, Other

Preprocessing - Correlation Based Feature Subset Selection

In this step, just the fields essential for the data mining process were chosen. The student's background information, class, final score course, gender, age, parent's income, and current job were used as attribute values for forecasts.

CFS is a method of correlation-based filtering. It rewards subsets with features that are significantly linked to the group's attribute yet unrelated to each other. Assume S is a variable subgroup with k

features, and rcf and rff represent the attribute's correlation to the group's variable and the variable intercorrelation, accordingly.

$$\text{meritS} = k \text{ rcf} / \sqrt{(k+k (k-1) \text{ rff})}$$

Naïve Bayes Classification

Classification is a data mining assignment that forecasts the group membership of data instances. Classification approaches are employed in this research to forecast a undergraduate student's future work. The Naïve Bayesian algorithm was used in this study as a classifier.

When the dimensionality of the inputs is high, the Naïve Bayes classifier technique is used. This is a simple algorithm that produces better results than others. This classifier predicts student's future work by assessing the likelihood of every input for a predicted outcome.

Step 1: Create tables of frequencies from the provided information.

Step 2: Create a Likelihood table by calculating the probability of the provided variables.

Step 3: Employ the Bayes theorem to compute the probability that follows..

WEKA was utilised to help us develop and evaluate our experiments. It is freely available open source data mining software that implements a vast range of mining methods. It supports a range of data formats and contains a converter. As a consequence, the student dataset was converted to a CSV file. We chose a ten-fold cross-validation approach under the "Test options" tab as our assessment method. The parts that follow look through all the different indicators of performance. The accuracy of the predictive model is determined with the true positive rate, false positive rate, precision, and accuracy. True Positive (TP) rate: The percentage of times a positive test result precisely describes what was tested for activity.

True positive (TP) occurs when the prediction result is P as well as the real value also happens to be P.

$$TP = TP/P \text{ where } P = (TP+FN)$$

FP rate (False positive): A false positive (FP) happens when the predicted value is p but the real number is n.

$$FP = FP / (FP+TN)$$

The fraction of relevant retrieved occurrences is referred to as precision.

$$\text{Precision} = TP / (TP+FP)$$

Remember that just a portion of the appropriate occurrences is returned.

$$\text{Recall} = TP / (TP+FN)$$

RESULT

Table 2. Student's Data Set

Gender	Age	Class	Living Location	Stay	Number of Members	Parent Income	Father Quality	Mother Quality	Father Occupation	Mother Occupation	Final Score	Job
F	19	A	Village	Hostel	three	below	SMA	SMA	farmer	NA	A	other
F	17	A	Village	Hostel	three	below	SMA	SMA	business	NA	A	other

F	18	A	Village	Hostel	two	above	SMA	SMA	government	NA	B	other
F	18	A	city	house	two	above	SMA	SMA	government	farmer	B	entrepreneur
F	18	A	city	house	two	above	SMA	S1	government	farmer	A	other
F	19	A	Village	Hostel	three	above	S1	S1	government	business	C	other
M	18	A	Village	Hostel	more than three	above	S1	SMA	farmer	other	A	other
M	19	A	city	house	more than three	above	SMA	SMA	farmer	other	A	other
M	17	A	Village	house	more than three	above	SMA	S1	other	other	A	other
M	18	A	city	house	more than three	below	S1	S1	farmer	NA	A	other

Table 2 illustrates a data set of students acquired through an archive and also a questionnaire of roughly 144 students from Universitas Negeri Medan's engineering major.

Table 3. Correlation Based Feature Selection Result

Rank	Score	Attribute
1	0.113	ParentIncome
2	0.077	FinalScore
3	0.0528	FatherOcc
4	0.0415	FatherQual

The current study uses feature selection techniques, cfsSubsetEval, This is significant and frequently utilised in data preprocessing in data mining. We may select the best qualities from an enormous variety of student variables that impact the student's future work using these attribute selection techniques. The Naïve Bayes classifier is used to achieve the results. From the data, we have 13 attributes and based on the feature selection, Correlation Based Feature Selection data result on Table 3, the four highest ranking attributes that can affect the prediction model are Parent's Income, Final Score of entrepreneurship course, father's occupation and father's qualification.

Table 4. Naive Bayes Classification Result

Feature Selection	TP-Rate	FP-Rate	Precision	Recall	F-Measure	Accuracy
no-cfsSubsetEval	86.0%	86.8%	75.1%	86.0%	80.2%	86.0%
cfsSubsetEval	87.3%	82.1%	89.0%	87.4%	82.2%	87.4%

In this experiment, The Naïve Bayes classifier was applied to the data set, and the results are shown in Table 4. It demonstrates that the classification results for Nave Bayes worked significantly using Correlation Based Feature Selection compared to a classifier model designed without any feature selection. For 10 fold cross validation, this model properly identifies roughly 87.4% of the time. In addition, the True Positive rate in this model is greater, at 87.3%.

DISCUSSIONS

The Naive Bayes model is a powerful statistical tool for data classification, prediction, interpretation, and manipulation with several potential applications in education research. The following benefits come from using the Naive Bayes model to describe research findings. This algorithm is fast and accurate in predicting the class of a test dataset. If feature independence is assumed to be correct, the Naive Bayes classifier beats different models with fewer training data. When compared to numerical parameters, the Naïve Bayes method excels with categorical input data.

An instance investigation in educational data mining is presented in this research article. The results suggest that by deleting irrelevant and redundant attributes, feature selection strategies can increase the classification algorithm's precision and efficacy. Its primary purpose was to improve student performance. The application for the CFS subset inspector is the most crucial feature. Naïve Bayes classifiers were employed on the selected features. When compared to no feature selection, the Correlation Based Feature Subset Evaluator improves the Nave Bayes classifier.

CONCLUSION

The categorization is utilised in this research in educational data to anticipate future work based on existing information. The Naive theorem is employed here because there are several techniques to knowledge classification. In order to anticipate the students' future field of work, knowledge such as the student's class, background, and achievement in the entrepreneurship course were obtained from the students' previous information. This research can assist students and professors in motivating students of all levels to achieve success. This study identifies kids who need special consideration and encourages students to be future entrepreneurs.

REFERENCES

- Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9(129), 6415–6426. <https://doi.org/10.12988/ams.2015.53289>
- Audretsch, D. B. (2017). Entrepreneurship and universities. In *Int. J. Entrepreneurship and Small Business* (Vol. 31, Issue 1).
- Basu, A., & Virick, M. (2008). *Assessing Entrepreneurial Intentions Amongst Students: A Comparative Study*.
- Bhunia, A. K., & Shome, M. K. (2023). INTERRELATION BETWEEN PERSONALITY CHARACTERISTICS AND SOCIAL CAPITAL IN FORMATION OF ENTREPRENEURIAL INTENTION: A CONCEPTUAL FRAMEWORK. *International Journal of Professional Business Review*, 8(3). <https://doi.org/10.26668/businessreview/2023.v8i3.1124>
- Block, J. H., Halberstadt, J., Högsdal, N., Kuckertz, A., & Neergaard, H. (2023). *Progress in Entrepreneurship Education and Training*.
- Bryson Welsh, J. (2007). *Identifying factors that predict student success in a community college online distance learning course*.
- Collins, L. A., Smith, A. J., & Hannon, P. D. (2006). Applying a synergistic learning approach in entrepreneurship education. *Management Learning*, 37(3), 335–354. <https://doi.org/10.1177/1350507606067171>
- Da, M., Wei, W., Hai-guang, H., & Jian-he, G. (2011). Modern Education and Computer Science. In *MECSJ. Modern Education and Computer Science* (Vol. 4). <http://www.mecs-press.org/>
- DeTienne, D. R., & Chandler, G. N. (2004). Opportunity Identification and Its Role in the Entrepreneurial Classroom: A Pedagogical Approach and Empirical Test. *Academy of Management Learning & Education*, 3(3), 242–257. <https://doi.org/10.5465/amle.2004.14242103>

- Gibb, A., & Hannon, P. (2006). *Towards the Entrepreneurial University?* www.ncge.org.uk
- Harmeling, S. S., & Sarasvathy, S. D. (2013). When Contingency Is a Resource: Educating Entrepreneurs in the Balkans, the Bronx, and Beyond. *Entrepreneurship: Theory and Practice*, 37(4), 713–744. <https://doi.org/10.1111/j.1540-6520.2011.00489.x>
- Heinonen, J., & Poikkijoki, S. A. (2006). An entrepreneurial-directed approach to entrepreneurship education: Mission impossible? *Journal of Management Development*, 25(1), 80–94. <https://doi.org/10.1108/02621710610637981>
- Henry, C., Hill, F., & Leitch, C. (2005). Entrepreneurship education and training: Can entrepreneurship be taught? Part I. *Education and Training*, 47(2), 98–111. <https://doi.org/10.1108/00400910510586524>
- Honig, B. (2004). Entrepreneurship Education: Toward a Model of Contingency-Based Business Planning. In *Source: Academy of Management Learning & Education* (Vol. 3, Issue 3).
- Khan, D. M., Mohamudally, N., & Babajee, D. K. R. (2013). A unified theoretical framework for data mining. *Procedia Computer Science*, 17, 104–113. <https://doi.org/10.1016/j.procs.2013.05.015>
- Kuratko, D. F. (2005). The Emergence of Entrepreneurship Education: Development, Trends, and Challenges. *Entrepreneurship Theory and Practice*, 29(5), 577–598.
- Lackéus, M., & Middleton, K. W. (2015). Venture creation programs: Bridging entrepreneurship education and technology transfer. *Education and Training*, 57(1), 48–73. <https://doi.org/10.1108/ET-02-2013-0013>
- Mahieu, R. (2006). *Agents of change and policies of scale : a policy study of entrepreneurship and enterprise in education*. Umeå universitet.
- Matlay, H. (2008). The impact of entrepreneurship education on entrepreneurial outcomes. *Journal of Small Business and Enterprise Development*, 15(2), 382–396. <https://doi.org/10.1108/14626000810871745>
- Mayhew, M. J., Simonoff, J. S., Baumol, W. J., Wiesenfeld, B. M., & Klein, M. W. (2012). Exploring Innovative Entrepreneurship and Its Ties to Higher Educational Experiences. *Research in Higher Education*, 53(8), 831–859. <https://doi.org/10.1007/s11162-012-9258-3>
- McClelland, T., & Jorba, M. (2022). Perceptual Motivation for Action. *Review of Philosophy and Psychology*, December 2021. <https://doi.org/10.1007/s13164-021-00608-8>
- Nageswari, S., Pallavi, D., Goel, M., & Divya, P. (2019). COMPARISON OF CLASSIFICATION TECHNIQUES ON DATA MINING. In *International Journal of Emerging Technology and Innovative Engineering* (Vol. 5, Issue 5). <https://ssrn.com/abstract=3375191>
- Nakra, A., & Duhan, M. (2019). Comparative Analysis of Bayes Net Classifier, Naive Bayes Classifier and Combination of both Classifiers using WEKA. *International Journal of Information Technology and Computer Science*, 11(3), 38–45. <https://doi.org/10.5815/ijitcs.2019.03.04>
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. In *Expert Systems with Applications* (Vol. 41, Issue 4 PART 1, pp. 1432–1462). <https://doi.org/10.1016/j.eswa.2013.08.042>
- QAA. (2012). *Enterprise and Entrepreneurship Education: Guidance for UK Higher Education Providers*.
- Sánchez, J. C. (2011). University training for entrepreneurial competencies: Its impact on intention of venture creation. *International Entrepreneurship and Management Journal*, 7(2), 239–254. <https://doi.org/10.1007/s11365-010-0156-x>
- Sarasvathy, S. D., & Venkataraman, S. (2011). Entrepreneurship as Method: Open Questions for an Entrepreneurial Future. *Entrepreneurship: Theory and Practice*, 35(1), 113–135. <https://doi.org/10.1111/j.1540-6520.2010.00425.x>
- Shepherd, D. A. (2004). Educating Entrepreneurship Students About Emotion and Learning From Failure. *Academy of Management Learning and Education*, 3(3), 274–287.
- Statistik. (2019). *Tingkat pengangguran terbuka (TPT) sebesar 5,01 persen*.
- Vinten, G., & Alcock, S. (2004). Entrepreneurship in education. *International Journal of Educational Management*, 18(3), 188–195. <https://doi.org/10.1108/09513540410527185>