

Implementation of Naïve Bayes Algorithm in Predicting Alumni Waiting Time to Secure Employment (Case Study: Universitas Sriwijaya)

Shelly Putri, Ken Ditha Tania

Universitas Sriwijaya, Indonesia

Email: shellzaholic27@gmail.com, Kenya.tania@gmail.com

*Correspondence

ABSTRACT

Keywords: data mining; classification method; CRISP-DM, decision tree; K-NN, naive bayes classifier; random forest; waiting period category prediction

In education, alums' success in getting a job after graduation is a significant benchmark for educational institutions in assessing the quality of education they provide. This study aims to estimate the waiting period category of alums based on the ability of alums to graduate when they are related to the waiting period category and design software that can predict the waiting period category of alums by classification method. The method applied is CRISP-DM. The data used is tracer study data in 2021 with 4,734 records. With a significant level of 5% (0.05), it was found that the waiting period category had a positive and detrimental relationship with the variables of GPA, Waiting Period, First Work Province, First Income, Ethics, Expertise, and English language ability. In this study, 10-fold cross-validation was applied, which resulted in the accuracy of the decision tree algorithm of 84.33%, the K-NN algorithm of 75.45%, the Naive Bayes Classifier algorithm of 85.21%, and the Random Forest algorithm of 84.04%. Furthermore, a different test (T-Test) was carried out, which showed that the Naive Bayes Classifier algorithm was the most dominant algorithm among the other three algorithms so that it could classify and predict the waiting period category well. This study concludes that applying the Naïve Bayes algorithm can effectively predict the waiting period for alums to get a job. The implication of this study is the development of web-based software that educational institutions can use to analyze the waiting period of alumni, provide recommendations for educational policies, and assist students in planning better career strategies.



Introduction

In the global landscape, the employment rate among university graduates has become an essential indicator of the success of educational institutions. The increasing competitiveness in the job market and rapid technological advances have significantly impacted the employability skills of graduates (Iswahyudi et al., 2023). The ability to predict alums's waiting period for obtaining a job is a valuable asset for universities in

evaluating the effectiveness of their curriculum and teaching methods (Murniati et al., 2022).

In the world of education, the success rate of alums in obtaining jobs after graduation is an important indicator for educational institutions in evaluating the quality of education provided (Andrean & Utami, 2024; Brilliant et al., 2023; Widhiantoyo, 2021). Therefore, educational institutions such as Sriwijaya University must understand the factors affecting the waiting period for alums to obtain jobs. The Alumni Waiting Period in Getting a Job technique can be used to predict the length of time it will take for an alum to get a job after graduation. (Afifah et al., 2024).

Colleges play an important role in improving human education. The right curriculum and educational strategies can produce high-quality students. Colleges must also precisely determine students' career prospects. (Widhiantoyo, 2021) The ability to estimate alums's chances of getting a job can be valuable information for colleges. Universities can use this information to assess the effectiveness of the curriculum and teaching methods applied so far and produce graduates ready to compete in the world of work.

Tracer Study is one of the important methods universities use to supervise alums after they complete their education and enter the world of work. The National Accreditation Board for Higher Education (BAN-PT) makes *the Tracer Study* one of the assessments and is also used as one of the requirements for accreditation completeness; in the context of the accreditation process, data from alum tracking is often one of the parameters that are highly considered. (Rahmi, 2015) One of the parameters measured is the waiting period for alums to find a job. This data provides valuable insights into the extent to which educational institutions can prepare their graduates for the world of work. Therefore, implementing alum tracking and waiting period prediction is a strategic step in improving the quality of education and meeting accreditation standards.

Of course, with the help of many alums, this data can be processed to find new valuable knowledge for Sriwijaya University. This aligns with the definition of data mining, which is finding relationships or patterns from the many fields in a relational *database*.

The Naïve Bayes algorithm is a classification method based on Bayesian probability theory that predicts specific categories based on historical data. (Adnyana, 2020; Amalia & Wibowo, 2020) This algorithm is applied to the 2021 Sriwijaya University alum tracer study data to predict the waiting time for alums to get jobs. The classification process determines whether an alum will get a job within 6 months or less or more than 6 months after graduation. This algorithm works by assuming that each variable used in the prediction is independent of each other, so the calculation is more efficient even though the dataset used is quite large.

In this study, predictions were made based on several variables that affected the waiting period of alums, such as GPA, study period, province where they worked for the first time, first income, work ethics, skills, and English language skills. The main advantage of Naïve Bayes' algorithm in this study is its ability to process large amounts

of data quickly and efficiently. In addition, this algorithm has been proven to handle variables with a non-linear relationship and provide more accurate classification results than other methods tested. The implementation of Naïve Bayes in this prediction system is carried out by designing a web-based application where users can fill out questionnaires related to factors that affect the waiting period for work. Once the data is entered, the system processes the information using the Naïve Bayes algorithm and displays the prediction results in two categories of waiting periods (Rachmadiansyah et al., 2022).

According To Rachmawati & Miasary, (2024) Confusion matrix testing and accuracy value are methods to measure the effectiveness of the two algorithms. The test results show that the C5.0 algorithm performs better than the Naive Bayes algorithm, which can be seen from the accuracy value of the C5.0 algorithm, which is 94.12% compared to the Naive Bayes algorithm, which is 91.18%. This study uses data mining with the C5.0 algorithm and Nave Bayes to predict the pattern of graduates of the mathematics study program at UIN Walisongo Semarang in the category of graduating on time and not on time.

This research focuses on applying data mining techniques, specifically the Naïve Bayes classification method, to predict the waiting time for graduates to get jobs. The novelty of this research lies in the extensive use of alum data from the Career Development Center of Sriwijaya University (CDC Unsri) to develop a predictive model with high accuracy. Unlike previous studies that only examined the correlation between academic performance and employment, this study integrates a variety of factors, including regional employment trends and individual competencies, to improve predictability.

The increasing need for data-driven decision-making in higher education drives the urgency of this research. Universities must optimize the curriculum and career guidance services to improve graduates' employability skills. By implementing a predictive model, institutions can identify areas for improvement and better prepare students for the workforce.

This study aims to develop a system that can predict the waiting period for alums to get a job using the classification method. The data that will be used is alum data that has been successfully collected by the UPT Center for Character Development and Alumni Management of Sriwijaya University, which is often known as the *Career Development Center* of Sriwijaya University (CDC Unsri) through filling out the *Sriwijaya University tracer study* questionnaire in 2021 by Sriwijaya University graduates who have completed their studies in 2020. The *tracer study* questionnaire consists of many question items. Of the many questions in the questionnaire, this research only focuses on attributes related to the waiting period for alums to get a job.

The research aims to study and apply data mining with a naïve Bayes approach to predict the time for alums to get a job. It involves Supporting Sriwijaya University alums who have graduated to get an idea of the estimated waiting time for getting a job and designing software to predict the waiting time of alums to get a job using *the Naïve Bayes* method.

Method

This study used quantitative research with a data mining approach, especially the classification method using the Naïve Bayes algorithm. (Sinaga & Suwilo, 2020). In this study, the type of data used is sequential data. Secondary data is data obtained through intermediaries because Peranta has collected the data. The intermediary in this study is the CDC of Sriwijaya University. The data used and processed for this final project research amounted to approximately 4734 data records, namely the 2021 alum tracer study. The data used is in Excel (.xls). The data collection method carried out during the research process is to directly request research data from the authorities for the data, namely the CDC of Sriwijaya University. The data analysis technique was carried out by applying the Naïve Bayes Classifier algorithm to predict the category of alum waiting period in obtaining a job. The 10-fold validation method carried out the validation process, followed by a differential test (T-Test) to determine the best algorithm among Naïve Bayes, Decision Tree, K-NN, and Random Forest. The analysis results show that Naïve Bayes has the highest accuracy, at 85.21%, compared to other algorithms. This study produces a web-based system that can predict the waiting period for alums to obtain a job based on relevant competency variables.

Research Stages

There are framework stages, which are the steps in carrying out research. At this stage of research, a data mining development method was adopted, namely CRISP-DM (Cross-Industry Standard Process Model for Data Mining) (Aldowah, 2023; Firas, 2023; Plotnikova et al., 2022, 2023; Pratiwi & Saikhu, 2022). The stages of the research can be seen in the figure below:

Implementation of Naïve Bayes Algorithm in Predicting Alumni Waiting Time To Secure Employment (Case Study: Universitas Sriwijaya)

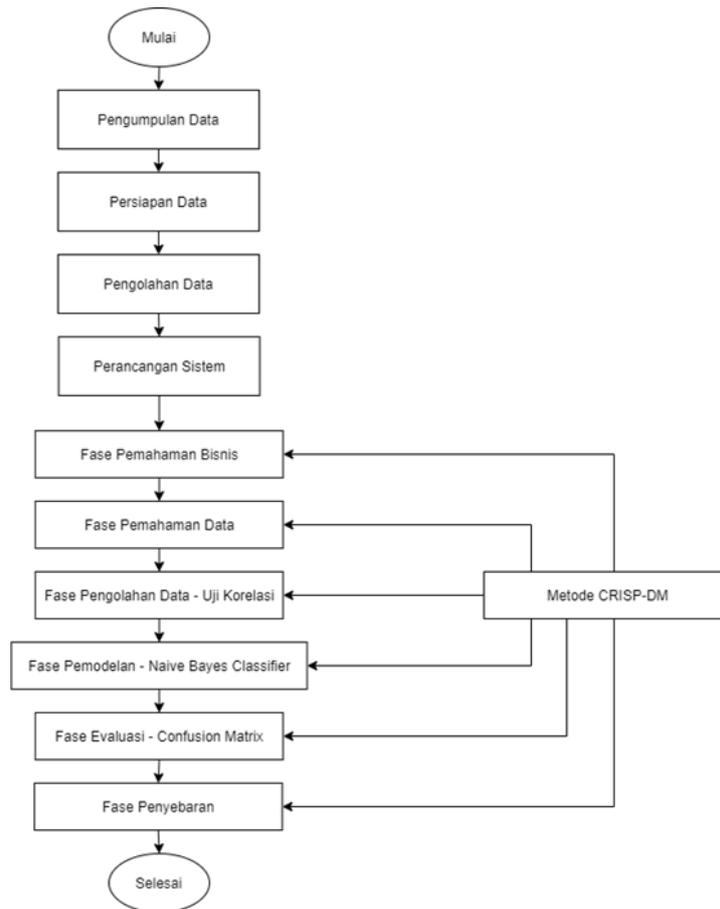


Figure 1. Research Stages

Results and Discussion

Result

The result of this study is the Application of *the Naïve Bayes* Algorithm to Estimate the Waiting Time for Alumni to get a Job (Case Study: Sriwijaya University), which was carried out through the analysis and design of an application system with the most superior appropriate algorithm for *the 2021* tracer study data. Several variables related to the waiting period level category have a very low correlation, namely the study duration and the type of first job.

Furthermore, a classification method was applied to forecast the waiting period category of students based on competency variables related to the category level. In this study, the applied algorithm has been tested with the 10-Fold cross-validation method and compared using a different test (T-Test) so that the naïve Bayes classifier algorithm is obtained as the most superior algorithm among the other three algorithms, namely *decision tree*, *K-NN*, and *random forest*, with an accuracy of 85.21%. The web-based application system that has been successfully developed can predict the Alumni Waiting Period Category, which consists of two levels: "Waiting Period of 6 months or less" and "Waiting Period of more than 6 months".

Discussion

Below is the interface page view of the Waiting Period Category prediction application system

a. Home



Figure 2. Future Page Wait

b. Waiting Period Category Classification Page

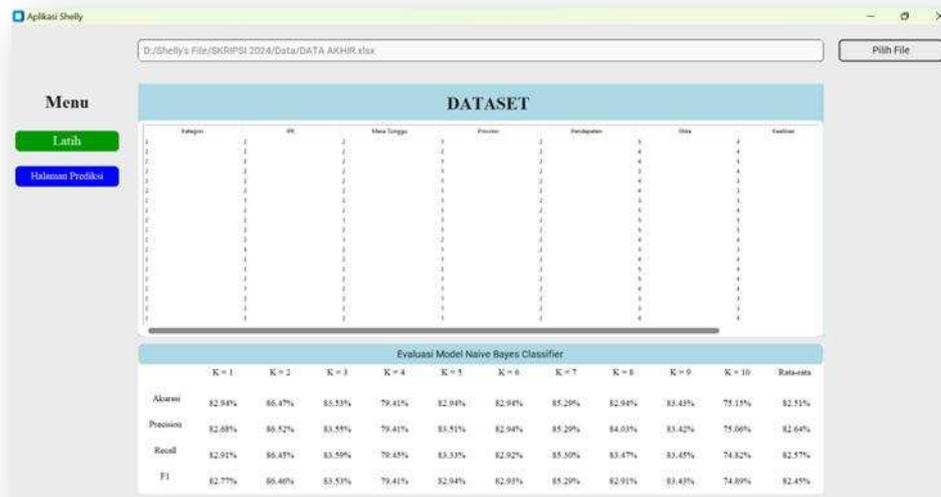


Figure 2. Waiting Period Category Classification Page

c. **Waiting Period Category Prediction Page**



Figure 3. Waiting Period Category Prediction Page

On the Waiting Period Category prediction page, users are asked to fill in the information and questionnaire provided based on the variables that affect the Waiting Period Category. Each question must be answered; if students ignore one question in the questionnaire, a notification will appear, as listed below.

After users fill out the questionnaire, they are then asked to click the "Data Prediction" button so that the system can determine the Waiting Period Category by utilizing the *naïve Bayes classifier* algorithm generated from the data mining modeling stage using the rapid miner application. After successfully carrying out the prediction process for the Waiting Period Category, the prediction results will appear as follows:



Figure 4. Waiting Period Prediction Page

d. Testing of the Waiting Period Category Prediction Application System

The Waiting Period Category prediction application system that has been designed must go through a test process before being implemented to discover errors in the application system and prevent potential failures.

Table 1. System Requirements Testing

No.	System Requirements	Conclusion
1	The system allows users, in this case students, to fill out a questionnaire that has been prepared and show the prediction results after pressing the Data Prediction button.	Succeed
2	The system can provide notifications if an error occurs while the system is being operated.	Succeed

Source: Data processed

Based on the test results, both system requirements were successfully met. This shows that the system functions as expected by completing questionnaires and providing error notifications.

Table 2. System Operational Testing

No.	Test Class	Action	Expected results	Conclusion
1	Selecting Dataset Page	The user selects a dataset from the document	The software successfully displays the selected dataset	Succeed
2	Classification Page	The user presses the classification/training button	The software displays the results of classification evaluations in the form of accuracy, precision, recall, and f1 score	Succeed
3	Prediction Page	The user answered all the questionnaires and clicked the "submit" button.	The software can display the results of data predictions in the form of a "Waiting period of 6 months or less" or "Waiting period of more than 6 months."	Succeed

Source: Data processed

Of the three operational test scenarios carried out, all were declared successful. This indicates that the system has fulfilled its operational functions to the designed specifications. The system can:

1. Select and view the dataset.
2. Evaluate the classification with relevant metrics.
3. Provides prediction results according to user input.

This success shows that the system is ready for use in an operational environment and performs as expected. However, further performance, scalability, and user experience testing can improve the system.

This study discusses the application of the Naïve Bayes algorithm in predicting the waiting period for Sriwijaya University alums to get a job. The results show that the Naïve Bayes algorithm is more accurate than other algorithms such as Decision Tree, K-NN, and Random Forest. With an accuracy of 85.21%, the algorithm can effectively classify alum waiting period categories into two groups: less than six months and more than six months.

In addition, this study identified several variables that affect the waiting period of alums, including GPA, waiting period, province of first place of employment, first income, ethics, expertise, and English proficiency. However, it was found that the variables of the study period and the type of first job had a very low correlation with the waiting period of graduation.

While this study shows that Naïve Bayes' algorithm excels at classifying alums' waiting periods, several research gaps can be explored further:

1. Variations of Factors Affecting Alumni Waiting Period This study only considers a few factors in predicting the alum waiting period. Further studies may include other factors such as work experience while in college, organizational activities, and internship involvement.
2. Although the study compared four algorithms, other algorithms, such as Support Vector Machine (SVM) or Deep Learning, can still be tested to see if they can produce more accurate prediction results.
3. Validation with Data from Other Universities This research data only comes from Sriwijaya University. Further studies can use data from other universities to test the reliability of the developed model.
4. The AI-based prediction Systems developed in this study are web-based, using the Naïve Bayes approach. Further research can develop artificial intelligence-based systems with more complex machine learning models to improve prediction accuracy.

Conclusion

Based on the analysis of the Application of *the Naïve Bayes* Algorithm in Predicting the Waiting Period for Alumni to Get a Job (Case Study: Sriwijaya University) it can be concluded that: 1) There are several variables related to the Alumni Waiting Period Category with a very low correlation level, including: Study Period and Type of First Job. 2) In this study, *the naïve Bayes classifier* algorithm is the most superior algorithm compared to *the decision tree, K-NN, and random forest algorithms* with an accuracy level of 85.21%. The *naïve bayes classsifier* algorithm is able to classify and predict the Waiting Period Category well.

Bibliography

- Adnyana, I. M. B. (2020). Implementasi Naïve Bayes untuk Memprediksi Waktu Tunggu Alumni dalam Memperoleh Pekerjaan. *Seminar Nasional Teknologi Komputer & Sains (Sainteks)*, 1(1), 131–134.
- Afifah, P. N., Rahmadhani, N., & Rifqi, S. (2024). Investigating Employment Delays: A Multiple Classification Analysis of Job Search Duration Among Graduates In Islamic Community Development. *Kubik: Jurnal Publikasi Ilmiah Matematika*, 9(2), 132–144.
- Aldowah, H. A. Y. (2023). *Modeling Factors Affecting Students' and Instructors' Behavior and Engagement In The Online Environment Using Educational Data Mining and Dematel Technique*.
- Amalia, R., & Wibowo, A. (2020). Prediction Of The Waiting Period For A Job Using The Naive Bayes Algorithm. *International Research Journal of Advanced Engineering And Science*, 5(2), 225–229.
- Andrean, L., & Utami, A. W. (2024). Rancang Bangun Sistem Informasi Prediksi Masa Tunggu Pekerjaan Alumni dengan Pendekatan Algoritma Naïve Bayes. *Journal of Emerging Information Systems and Business Intelligence (Jeisbi)*, 5(2), 148–158.
- Brilliant, M., Nurhasanah, I. A., & ... (2023). Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Waktu Tunggu Alumni dalam Memperoleh Pekerjaan (Study Kasus Smks Pgri 2 Pringsewu). *Seat: Journal Of ...*, 15–20. <https://doi.org/10.69769/Seat.V3i2.101>
- Firas, O. (2023). A Combination Of Semma & Crisp-Dm Models For Effectively Handling Big Data Using Formal Concept Analysis Based Knowledge Discovery: A Data Mining Approach. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 9–14.
- Iswahyudi, M. S., Irianto, I., Salong, A., Nurhasanah, N., Leuwol, F. S., Januaripin, M., & Harefa, E. (2023). *Kebijakan dan Inovasi Pendidikan: Arah Pendidikan di Masa Depan*. PT. Sonpedia Publishing Indonesia.
- Murniati, A. R., Usman, N., & Ulfah Irani, Z. (2022). *Who Wants To Be A Teacher? Menyiapkan Guru Profesional di Era Global Society 5.0*. Deepublish.
- Plotnikova, V., Dumas, M., & Milani, F. P. (2022). Applying The Crisp-Dm Data Mining Process in The Financial Services Industry: Elicitation of Adaptation Requirements. *Data & Knowledge Engineering*, 139, 102013. <https://doi.org/10.1016/j.datak.2022.102013>
- Plotnikova, V., Dumas, M., Nolte, A., & Milani, F. (2023). Designing A Data Mining Process for The Financial Services Domain. *Journal of Business Analytics*, 6(2), 140–166. <https://doi.org/10.1080/2573234x.2022.2088412>
- Pratiwi, P. A., & Saikhu, A. (2022). Classification Modeling Using Methods Cross Industry Standard Process for Data Mining to Improve Product Quality. *Jurnal Teknobisnis*, 8(2), 44–55. <https://doi.org/10.12962/J24609463.V8i2.1405>
- Rachmadiansyah, R., Rumlaklak, N. D., & Mauko, A. Y. (2022). Prediksi Masa Tunggu

- Kerja Alumni Menggunakan Naïve Bayes Classifier pada Program Studi Ilmu Komputer Universitas Nusa Cendana. *J-Icon: Jurnal Komputer dan Informatika*, 10(2), 143–150. <https://doi.org/10.35508/Jicon.V10i2.7426>
- Rachmawati, A. K., & Miasary, S. D. (2024). Performance Analysis Of C5. 0 And Naïve Bayes Classification Algorithm For Pattern Recognition Of Student Graduates. *Aip Conference Proceedings*, 3046(1). <https://doi.org/10.1063/5.0194629>
- Rahmi, A. N. (2015). Pengembangan Sistem Tracer Study Online Berbasis Website di Stmik Amikom Yogyakarta. *Csrid (Computer Science Research And Its Development Journal)*, 6(2), 108. <https://doi.org/10.22303/Csrid.6.2.2014.108-117>
- Sinaga, L. M., & Suwilo, S. (2020). Analysis Of Classification and Naïve Bayes Algorithm K-Nearest Neighbor In Data Mining. *Iop Conference Series: Materials Science And Engineering*, 725(1), 12106. [10.1088/1757-899x/725/1/012106](https://doi.org/10.1088/1757-899x/725/1/012106)
- Widhiantoyo, A. (2021). Penerapan Algoritma Naïve Bayes Dengan Backward Elimination Untuk Prediksi Waktu Tunggu Alumni Mendapatkan Pekerjaan. *Jiko (Jurnal Informatika Dan Komputer)*, 4(3), 145–151. <https://doi.org/10.33387/Jiko.V4i3.3272>