# Predicting Interest in Working Abroad using Naïve Bayes and Decision Tree Algorithms

Nur Widiastuti
Master Program of Information Technology
Universitas Teknologi Yogyakarta
Jl. Siliwangi Jl. Ring Road Utara, Jombor Lor, Sendangadi, Kec. Mlati, Kab. Sleman, Daerah Istimewa Yogyakarta 55285

Arif Pramudwiatmoko
Master Program of Information Technology
Universitas Teknologi Yogyakarta
Jl. Siliwangi Jl. Ring Road Utara, Jombor Lor, Sendangadi, Kec. Mlati, Kab. Sleman, Daerah Istimewa Yogyakarta 55285

## ABSTRACT
In order to reduce the unemployment rate, especially in Central Java Province, the Provincial Manpower Office Central Java created an innovation, namely the Central Java e-Makaryo application or https://bursakerja.jatengprov.go.id/. This application is useful for connecting job seekers and employers. However, the application has not been able to analyze how many job seekers are interested in working abroad and job seekers who are not interested in working abroad. Meanwhile, this is very much needed by the government to prepare job vacancies as needed. In this regard, the data of job seekers who are interested in working abroad are analyzed using the RapidMiner application with the Naïve Bayes algorithm classification method and the Decision Tree algorithm. The number of data used is 11,464. The accuracy results after testing on job seeker data classification using the Decision Tree algorithm has accuracy rate of 77,35%. +/_ 0,10% (micro average: 77.35%), higher than the Naïve Bayes algorithm which has an accuracy rate of 75.99%. Thus, the accuracy performance using the Decision Tree algorithm is better than Naïve Bayes.

## General Terms
Machine Learning, Supervised Learning, Classification

## Keywords
Algorithm, classification, machine learning, Naïve Bayes, Decision Tree, data mining, unemployment, job seekers.

## 1. INTRODUCTION
One of the problems faced by developing countries including Indonesia is unemployment and poverty. Every year there are new graduates from school to university level followed by an increasing unemployment rate. The high unemployment rate causes the poverty rate to increase. Coronavirus also causes a global pandemic that has adversely affected public health, the economy, including every life aspect. To manage the spread, innumerable measurements are gathered. [1] The current problem with unemployment is that if the population continues to grow each year, it will become unbalanced in the world of work. Because the existing population is not proportional to employment. [2]

While the social impact of this type of unemployment is relatively greater and there are many negative effects from this, one of which is the crime rate in each region also increases due to economic encouragement. [3]

Unemployment also showed problems inefficiencies in the use of production factors, causing the level of prosperity of society do not reach the maximum potential. Besides that, unemployment can also use as a measure in assessing a government's performance.[4]

Unemployment is also caused by the number of job seekers that is not proportional to the number of jobs available, in other words, the number of job seekers is more than the number of jobs. The lack of government in providing jobs when the number of workers is getting higher results in a lot of unemployment. [5] The Central Java Statistics Agency (2022) recorded the unemployment rate in Central Java Province in 2022 at an open unemployment rate of 5.57 percent. The imbalance between employment and the number of existing workers results in fierce competition for job seekers. [6]

The government, in this case the Central Java Provincial Manpower and Transmigration Office, has the responsibility to provide facilities to job seekers in order to facilitate the workforce to get decent work. In response to this, the Provincial Manpower and Transmigration Office. Central Java created an Online Job Exchange application or called e-Makaryo Jateng. Job exchange is equated with the notion of labor market information, which is to bring together and facilitate meetings between job seekers and employers online. [7] However, the shortcomings of the e-Makaryo application have not been able to classify job seeker data based on worker interests, especially interest in working domestically or abroad. This is important because the mismatch of interests and jobs informed by the government to job seekers can reduce interest in applying for work. Thus, a solution is needed to overcome this problem.

## 2. LITERATURE REVIEW
One of the studies that has been conducted on labor issues is data analysis research of Indonesian Migrant Workers using the C4.5 Algorithm and Naïve Bayes by processing variables of age, gender, education, marriage status, education, country_destination, status_PMI, sector_work. The experiment was conducted with 1802 training data and 772 testing data, resulting in the highest accuracy value for both algorithms. The C4.5 algorithm is able to predict better with an accuracy rate of 84.84% while the Naïve Bayes algorithm produces an accuracy value of 58.29%.[8].

In connection with the shortcomings in previous studies, namely the low accuracy of the Naïve Bayes algorithm, this study will classify primary data, namely job seekers recorded at the Central Java Province Manpower and Transmigration Office based on the interest of job seekers abroad and domestically using the Datamining Classification method with the Naïve Bayes algorithm and Decision Tree algorithms by applying feature selection Cross Validation. With this research, it is hoped that it will produce classified job seeker data based

on interest in working abroad and domestically, thus facilitating the functional performance of job introducers at the Central Java Province Manpower and Transmigration Office to prepare job vacancies according to the interests of job seekers.

Research on Classification of concentrations that can be chosen by fourth semester students including Computerized Accounting, Computer Administration, and Multimedia using Naïve Bayes and Decision Tree J48 using WEKA tools. The dataset used in this study is 111 with a split test percentage mode of 75% and 25% training data. The accuracy result of Naïve Bayes is 71.4% while Decision Tree J48 has an accuracy of 64.3%. In Decision Tree J48 there are 4 patterns or rules formed to determine the concentration selection so that the academic department can help students in determining the selection of concentration. [9]

Research on Decision Support Model for Employee Recruitment Using Data Mining (DSM) Classification, namely model created for the needs of companies in recruiting employees objectively. There are four data mining classification algorithms compared in this case C4.5 Decision Tree, Naïve Bayes, Support Vector Machine, and RandomForest. The result is that C4.5 Decision Tree has the highest accuracy of 88.24%, specificity of 88.10% and sensitivity of 100% sensitivity.[10]

# 3. RESEARCH METHODOLOGY

In this study, data was taken from the Central Java e-Makaryo application or bursakerja. In this research, the performance of classification algorithm on the data of job seekers will be assessed. The method used is classification using the Naïve Bayes and Decision Tree algorithms with feature selection Correlation Matrix while the software used is RapidMiner. The research stages are as follows:

a. Prepare the dataset,
b. Import into the RapidMiner application,
c. Select attribute,
d. Label assignment,
e. Divide data into training data and testing data
f. Analysis process using Naïve Bayes and Decision Tree operators,
g. Apply the model and performance so that the accuracy performance results are obtained. [11]

## 3.1 Data Collection

In this study, data was taken from the Central Java e-Makaryo website address bursakerja.jatengprov.go.id from the Central Java Provincial Manpower and Transmigration Office, then the data was grouped into 11 attributes with a total amount of 11,464 data.

## 3.2 Dataset

Datasets are collection of related data corresponding to the content of a particular database table where each column represents a variable (feature). [11]. The dataset used in this research is the original data, obtained directly from the Central Java Province Online Job Exchange or e-Makaryo. This research uses a dataset of job seekers from the Central Java e-Makaryo website, where the web-based application is an innovation of the Provincial Manpower Office Central Java for connecting job seekers and job providers and facilitating job seekers to get jobs both domestically and abroad.

## 3.3 Research Variables (Attributes)

In this study, the number of variables to be used consists of 11 attributes selected from 16 attributes in the dataset. These 11 attributes have been selected based on the most important attributes because there are some attributes that can represent several other attributes such as name, gender, date of birth and district origin. Of the 11 attributes selected as a label or class is LN Job Interest, while the other 10 attributes are ordinary attributes. The attributes are shown in Table 1.

**Table 1. Job seeker attribute table**

| Attribute | Data Type | Class |
|---|---|---|
| Interest in Working Abroad | Binominal | Class/Label |
| Place of Birth | Polynominal | Attribute |
| Date of Birth | Date_time | Attribute |
| Age | Integer | Attribute |
| Regency | Polynominal | Attribute |
| District | Polynominal | Attribute |
| Gender | Polynominal | Attribute |
| Status | Polynominal | Attribute |
| Last Education | Polynominal | Attribute |
| Major | Polynominal | Attribute |
| Year Graduated | Date_time | Attribute |

## 3.4 Data Mining

Data mining algorithms can be used efficiently in prediction and classification. [12]. Data mining is a process that has the aim of finding an automatic or semi-automatic pattern from the data that already have in the database is used to solve a problem. [13]. Data mining can be interpreted as a series of processes to obtain knowledge or patterns from data sets. The purpose of data mining is to classify, cluster, find association patterns to do predicting [14]. Data mining involves the searching of large information of the data or records to discover patterns and utilize these patterns in the prediction the future events. [15] Data mining is to have a role as prediction, description. Not only that, its role is classification association analysis and predicting cluster analysis, outlier analysis, trend analysis and evolution. [16] Data mining itself is one of the sciences in the field of computer science which can be in the form of grouping to be used in solving problems such as predictions, such as classification or regression. [17] Important information is carried out using several methods, including statistical methods, mathematic-section, and artificial intelligence technology. Data mining is described more explicitly as tools and programs that employ statistical data analysis and filters to save as much data as possible. [18]

## 3.5 Classification

Classification is one of the methods in data mining for categorizing a particular group of items to targeted groups. [15]. Classification is a technique in data mining to form a model from a predetermined data set. Data mining techniques are the choices that can overcome in solving this problem. [19] Classification is also one of the techniques in data processing that works by dividing the object used into classes with the desired number of classes. Classification can create a pattern that can separate each class of data that aims to determine which objects belong to a particular category based on the behavior and attributes of the group that has been defined [8].

## 3.6 Naïve Bayes

Naïve Bayes is an algorithm method that works on how to calculate the frequency of each term in the document. [20]. Naïve Bayes is one of the algorithms found in classification techniques, which predicts future chances in the future based on previous experience, which is known as Bayes' theorem.[21] Naïve Bayes Classifier is a classification method based on the

Bayes theorem for predicting opportunities based on existing experiences using probability and statistical methods proposed by a British scientist named Thomas Bayes [22]. The Naïve Bayes Classifier algorithm model has a very minimum error rate and is known for its simple, fast, and highly accurate calculations [23]. The Naïve Bayes classification method is often referred to as Naïve Bayes Classifier (NBC) in which this technique is known as the best technique in computation time compared to other data mining algorithms. The advantage of using Naïve Bayes is this method requires only a small amount of training data to determine the estimated parameter required in the classification process. Naïve Bayes often works much better in most real-world situations that are more complex than expected [24].

## 3.7 Decision Tree

The Decision Tree method is one of many classification methods for building a predicting system. The Decision Tree method converts a large number of facts into a Decision Tree that represents the rules. The rules that are developed must be simple to understand in everyday language. [25]

The Decision Tree is a supervised simple classification tool that can separate data records into designated categories by applying specific conditions in the decision-making process. It is an established tool, and one of the most powerful with relatively small learning curves for interpretability, and is regularly applied in numerous settings such as image processing, ML, data mining and identifications of patterns [26]. DT classifier has a form of tree structure, each node in the tree is either a decision node or a leaf node, the decision node specifies a rule or test that carried out on a single attribute value, while the leaf node indicates the class of a data sample, and the decision node may directly followed by a class, or it will have a sub-tree for each possible outcome of test carried out in the node.[27] The tree is built in the first phase by recursively splitting the training set based on local optimal criteria until all or most of the records belonging to each of the partitions bearing the same class label. The tree may overfit the data. The pruning phase handles the problem of over fitting the data in the Decision Tree. Pruning phase accesses only the fully grown tree. The growth phase requires multiple passes over the training data. The time needed for pruning the Decision Tree is very less compared to build the Decision Tree. [28]

## 3.8 Cross Validation

Cross Validation is a validation technique by dividing the data randomly into parts and each part will be classified. This method divides the data into two parts, namely training data and testing data [8]. The technique of cross validation is k-fold cross validation, which breaks the data into 'k' parts of the dataset with the same size. The use of k-fold cross validation to eliminate bias in the data. Training data and testing data are carried out as many times as the number k specified [9]. The technique of cross validation is k-fold cross validation, which breaks the data into 'k' parts of the dataset of the same size. Training data and testing data are performed as many times as the number k specified. [29] This research will classify data on job seekers who are interested in working abroad using the Naïve Bayes and Decision Tree algorithms by applying feature selection Cross Validation. With this research, it is hoped that it will produce classified job seeker data based on interest in working abroad and domestically, thus facilitating the functional performance of job introducers at the Central Java Province Manpower and Transmigration Office to prepare job vacancies according to the interests of job seekers. The next process is data processing which begins with normalization, replacing missing values, selecting attributes and determining labels on the dataset. The dataset will be divided into training data and testing data using Cross Validation, which is to divide the training data and test data randomly, with a k-fold of 12 folds. Then the data will be tested using the algorithm and apply mode.

# 4. CASE STUDIES

The results of this study are to test the accuracy of data analysis of job seekers interested in working abroad using Naïve Bayes and Decision Tree algorithms. Then the results of model testing will be compared to find out the best algorithm for classification for data analysis of job seekers who are interested in working abroad.

## 4.1 Model Testing

In this study, the RapidMiner tool or software is used to test the algorithm model used. The model will be used to process data and then compare the performance of the results of each algorithm that has been tested on data of job seekers who are interested in working abroad.

### *4.1.1 Testing using the Decision Tree Algorithm*

The stages of testing using the Decision Tree algorithm are data retrieval, then normalizing the data and replacing missing values, followed by selecting the attributes to be used using the select attributes operator. Next, determine the label (set role) then add a cross validation operator to sort the data into training data and testing data.

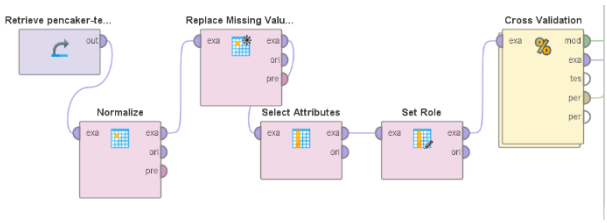


**Fig 1: Data analysis process using the Decision Tree Algorithm**

In Fig 1 above, in Cross Validation there is a process of modeling and testing using the Decision Tree algorithm, so that later it will produce accuracy from testing data on job seekers who are interested in working abroad.

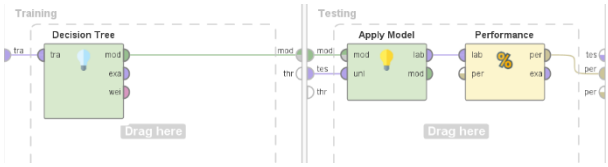


**Fig 2: Process of applying the data model with the Decision Tree Algorithm**

In the Fig 2 above is the process of measuring the accuracy level using the Decision Tree algorithm using the Cross Validation method with k-fold of 12 folds, so that the best accuracy results will be obtained.

### *4.1.2 Testing using the Naïve Bayes Algorithm*

The stages of testing using the Naïve Bayes algorithm are data retrieval, then selecting the attributes to be used using the select attributes operator and normalizing the data. The next stage is to add a replace missing value operator then continue to determine the label using the set role operator. After that, sorting the data into training data and testing data and testing using the Cross Validation operator.

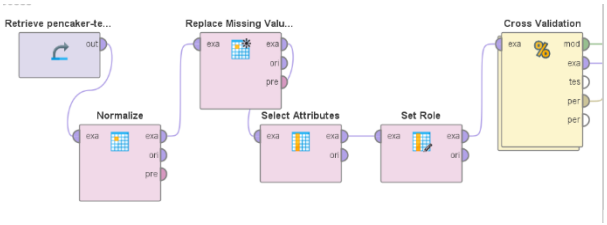


**Fig 3: Process of data analysis with the Naïve Bayes Algorithm**

In Cross Validation there is a process of modeling and testing using the Naïve Bayes algorithm, so that later it will produce accuracy from testing the data of job seekers who are interested in working abroad, with a k-fold of 12 folds, so that the best accuracy results will be obtained.

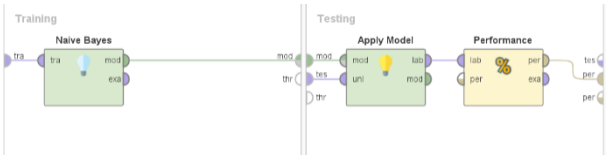


**Fig 4: Process of applying the data model with the Naïve Bayes Algorithm**

In Fig 4 above, based on the model that has been formed, testing is then carried out to measure the accuracy level of testing using the Naïve Bayes algorithm, then adding the apply model and performance operators to determine the percentage of accuracy.

## 4.2 Research Result

### *4.2.1 Decision Tree Algorithm Result*

After the testing process is carried out, the following are the test results using the Decision Tree algorithm, which can be seen in Table 2.

**Table 2. Decision Tree Algorithm Test Results**

| Description | True Yes | True No | Class precision |
|---|---|---|---|
| Pred. Yes | 1 | 8 | 11,11 % |
| Pred. No | 2589 | 8866 | 77,40% |
| Class recall | 0,04% | 99,91% | |

Based on Table 2, it can be seen that the Decision Tree algorithm produces an accuracy rate of 77,35%. +/_ 0,10% (micro average: 77.35%). The class precision level of prediction yes is 40.00%% and class precision prediction no is 77.40%, while class recall true yes is 0.04% and class recall true no is 99.91%. Data on job seekers who are interested in going abroad based on yes predictions are 2589, and no predictions are 8866.
In testing using the Decision Tree algorithm, the Decision Tree scheme can be seen as follows as shown in Figure 1.

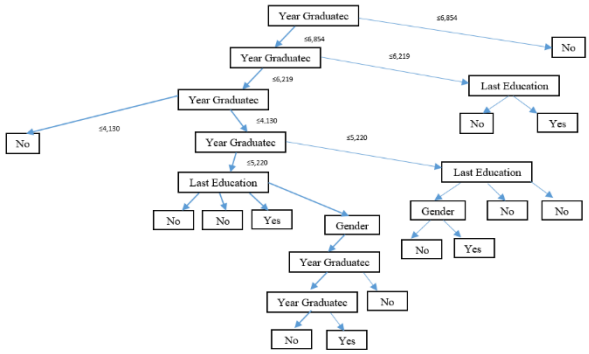


**Fig 5: Decision Tree graph results**

In Fig 5 above, based on the Decision Tree for data classification model for job seekers interested in going abroad. It can be seen that the attributes that have the main influence in testing are the regency attributes and gender.

### *4.2.1 Naïve Bayes Algorithm Result*

After the testing process is carried out, the following are the test results using the Naïve Bayes algorithm, which can be seen in Table 3.

**Table 3. Naïve Bayes Algorithm Test Results**

| Description | True yes | True no | Class precision |
|---|---|---|---|
| Pred. Yes | 601 | 741 | 44,78 % |
| Pred. No | 1982 | 8123 | 80,35% |
| Class recall | 23,25% | 91,54% | |

Based on Table 3, it can be seen that the Naïve Bayes algorithm produces an accuracy rate of 76.10% +/_ 0.63% (micro average: 76.10%). The class precision level of yes prediction is : 44.78% and class precision prediction is not 80.35%, while class recall true yes is 23.25% and class recall true no is 91.54%. Data on job seekers who are interested in going abroad based on yes predictions are 1982, and no predictions are 8123. The results of testing using the Naïve Bayes algorithm show that the most significant attribute that determines interest in working abroad is marital status (Fig 6), followed by the last education (Fig 7). On the other hand, the results show that gender only makes a slightly significant contribution with men more interested in working abroad than women (Fig 8).

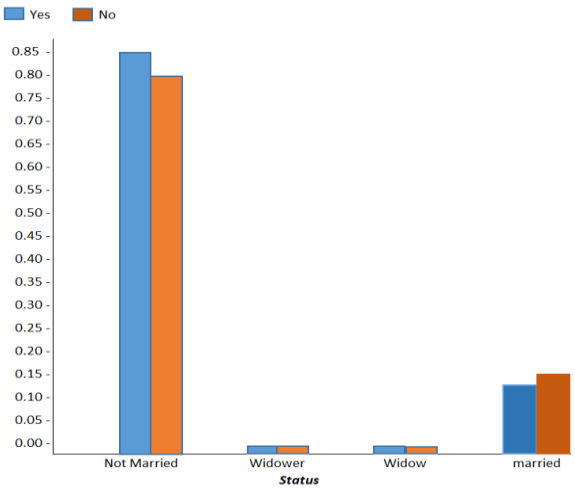


**Fig 6: Naïve Bayes graph results by status**

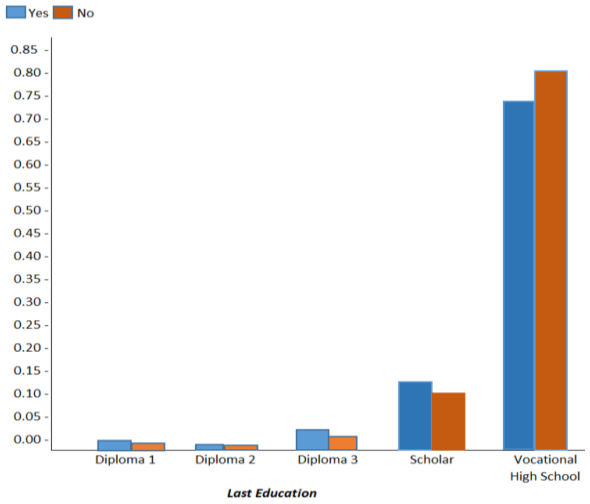


**Fig 7: Naïve Bayes graph results by last education**

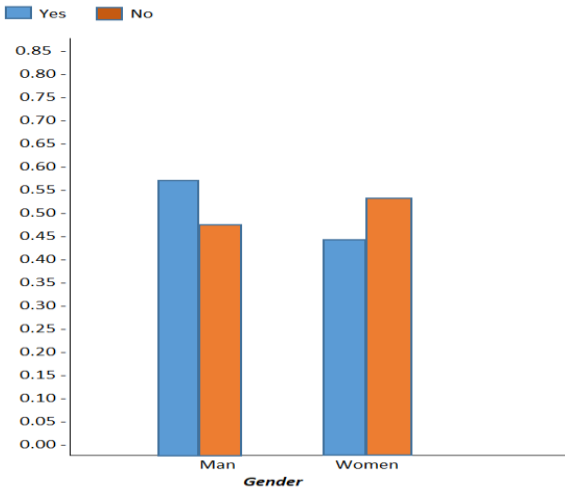


**Fig 8: Naïve Bayes graph results by gender**

## 5. CONCLUSION

Classification results for data analysis of job seekers interested in working abroad using Decision Tree with Cross Validation and k-fold 12 fold testing features produce an accuracy rate of 77.40%. +/_ 0.08% (micro average: 77.40%). Meanwhile, the test results using the Naïve Bayes algorithm with the Cross Validation and k-fold 12 fold testing features produced an accuracy level of 76.07%. +/_ 0.67% (micro average: 76.07%). Based on the results of research using the Naïve Bayes algorithm, it produces data on those interested in working abroad based on gender, more men are interested, based on the last education, more vocational high schools are interested and based on status more are not married who are interested in working abroad. By looking at the accuracy results above, it can be seen that the Naïve Bayes algorithm has less performance when compared to the Decision Tree algorithm, the results are that there are not as many job seekers who are interested in going abroad as there are job seekers who are interested in working domestically. The difference in accuracy between the Naïve Bayes algorithm and the Decision Tree is 1.37%. It can be concluded that the performance of the Decision Tree algorithm in researching data on job seekers who are interested in going abroad is better than the Naïve Bayes algorithm.

## 6. REFERENCES

[1] S. H. Ing, A. A. Abdullah, M. Y. Mashor, Z. A. Mohamed-Hussein, Z. Mohamed, and W. C. Ang, "Exploration of hybrid deep learning algorithms for covid-19 mRNA vaccine degradation prediction system," *Int. J. Adv. Intell. Informatics*, vol. 8, no. 3, pp. 404–416, 2022, doi: 10.26555/ijain.v8i3.950.

[2] A. Azaluddin and L. Hanifa, "Effect of Inflation and Economic Growth on The Rate of Unemployment," *Sang Pencerah J. Ilm. Univ. Muhammadiyah But.*, vol. 7, no. 4, pp. 609–617, 2021, doi: 10.35326/pencerah.v7i4.1559.

[3] L. Yanthiani, "The Impact of Unemployment on the Economy in Indonesia," *J. Islam. Econ. Bus.*, vol. 2, no. 2, pp. 112–130, 2023, doi: 10.15575/jieb.v2i2.21310.

[4] E. K. Oktafianto, N. A. Achsani, and T. Irawan, "The Determinant of Regional Unemployment in Indonesia: The Spatial Durbin Models," *Signifikan J. Ilmu Ekon.*, vol. 8, no. 2, pp. 179–194, 2019, doi: 10.15408/sjie.v8i2.10124.

[5] M. L. B. Ginting, "Perluasan Kesempatan Kerja Bagi Freshgraduate di Masa Pandemi Covid-19, Apa Peran Pemerintah?," *J. Ketenagakerjaan*, vol. 16, no. 2, 2021, doi: 10.47198/naker.v16i2.106.

[6] BPS Kebumen, "Kabupaten Kebumen dalam Angka 2022," p. 292, 2022.

[7] K. Wulandari, D. Hariani, and S. Sulandri, "Analisis E-Service Bursa Kerja Online," *J. Public Policy Manag. Rev.*, vol. 10, no. 2, pp. 1–19, 2021.

[8] F. D. S. Harahap, "Dampak Pandemi Covid-19 Terhadap Masyarakat Khususnya Dunia Ketenagakerjaan," *OSF.*, vol. 2019, 2020.

[9] B. Budiman, R. Nursyanti, R. Y. R. Alamsyah, and I. Akbar, "Data Mining Implementation Using Naïve Bayes Algorithm and Decision Tree J48 In Determining Concentration Selection," *Int. J. Quant. Res. Model.*, vol. 1, no. 3, pp. 123–134, 2020, doi: 10.46336/ijqrm.v1i3.72.

[10] C. E. Amos Pah, "Decision Support Model for Employee Recruitment Using Data Mining Classification," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, pp. 1511–1516, 2020, doi: 10.30534/ijeter/2020/06852020.

[11] A. Renear, S. Sacchi, and K. Wickett, "Definitions of dataset in the scientific and technical literature," *Proc. Am. Soc. Inf. Sci. Technol.*, vol. 47, pp. 1–4, Nov. 2010, doi: 10.1002/meet.14504701240.

[12] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[13] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.

[14] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.35329/jiik.v7i2.203.

[15] A. Oluwaseun and M. S. Chaubey, "Data Mining Classification Techniques on the analysis of student

performance," *Glob. Sci. J.*, vol. 7, no. April, pp. 79–95, 2019.

[16] Y. T. Utami and E. Elisa, "Prediksi Kinerja Karyawan Berdasarkan Proses Trainer Menggunakan Data Mining," *J. Comasie*, vol. 04, 2022.

[17] J. T. Samudra, R. Rosnelly, and Z. Situmorang, "Comparative Analysis of Support Vector Machine and Perceptron In The Classification of Subsidized Fuel Receipts," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 652–656, 2023, doi: 10.29207/resti.v7i3.4731.

[18] M. Muhasshanah, M. Tohir, D. A. Ningsih, N. Y. Susanti, A. Umiyah, and L. Fitria, "Comparison of the Performance Results of C4.5 and Random Forest Algorithm in Data Mining to Predict Childbirth Process," *CommIT (Communication Inf. Technol. J.*, vol. 17, no. 1, pp. 51–59, 2023, doi: 10.21512/commit.v17i1.8236.

[19] D. Syafira, S. Suwilo, and P. Sihombing, "Analysis of Attribute Reduction Effectiveness on the Naive Bayes Classifier Method," *J. Phys. Conf. Ser.*, vol. 1566, no. 1, 2020, doi: 10.1088/1742-6596/1566/1/012060.

[20] B. S. Prakoso, D. Rosiyadi, H. S. Utama, and D. Aridarma, "Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifer Dengan Seleksi Fitur Dan Boosting," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 227–232, 2019, doi: 10.29207/resti.v3i2.1042.

[21] M. Hadikristanto Wahyu ; Suprayogi, "SIGMA - Jurnal Teknologi Pelita Bangsa SIGMA - Jurnal Teknologi Pelita Bangsa," *SIGMA - J. Teknol. Pelita Bangsa 167*, vol. 10, no. September, pp. 167–172, 2019.

[22] R. Adrian, M. A. J. S. Perdana, A. Asroni, and S. Riyadi, "Applying the Naive Bayes Algorithm to Predict the Student Final Grade," *Emerg. Inf. Sci. Technol.*, vol. 1, no. 2, pp. 49–57, 2020, doi: 10.18196/eist.127.

[23] Y. F. Safri, R. Arifudin, and M. A. Muslim, "K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor," *Sci. J. Informatics*, vol. 5, no. 1, p. 18, 2018, doi: 10.15294/sji.v5i1.12057.

[24] R. Achmad and A. S. Girsang, "Implementation of naive bayes classifier algorithm in classification of civil servants," *J. Phys. Conf. Ser.*, vol. 1485, no. 1, 2020, doi: 10.1088/1742-6596/1485/1/012018.

[25] R. Sistem, R. Prabaswara, J. Lemantara, and J. Jusak, "Classification of Secondary School Destination for Inclusive Students," *Jurnal Resti.*, vol. 5, no. 158, pp. 1009–1019, 2023.

[26] C. S. Lee, P. Y. S. Cheang, and M. Moslehpour, "Predictive Analytics in Business Analytics: Decision Tree," *Adv. Decis. Sci.*, vol. 26, no. 1, pp. 1–29, 2022, doi: 10.47654/V26Y2022I1P1-30.

[27] I. M. Nasser and A. H. Alzaanin, "Machine Learning and Job Posting Classification: A Comparative Study," *Int. J. Eng. Inf. Syst.*, vol. 4, no. 9, pp. 6–14, 2020, [Online]. Available: www.ijeais.org

[28] D. Lavanya and K. U. Rani, "Performance Evaluation of Decision Tree Classifiers on Medical Datasets," *Int. J. Comput. Appl.*, vol. 26, no. 4, pp. 1–4, 2011, doi: 10.5120/3095-4247.

[29] H. Husni, "Kajian Literatur Mengenai Klasifikasi Blog," *J. Simantec*, vol. 8, no. 2, pp. 63–77, 2020, doi: 10.21107/simantec.v8i2.7223.