Journal of Science and Social Research ISSN 2615 – 4307 (Print)
August 2024, VII (3): 1091 – 1096 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

PREDICTING JOBS OR COLLEGES WITH CLASSIFICATION ALGORITHMS USING

Ilham Wahyudi Siadi[1], Windu Gata[2], Cicih Sri Rahayu[3]
Nusa Mandiri University
e-mail: [1]14207026@nusamandiri.ac.id, [2]windu@nusamandiri.ac.id,
[3]cicihsrirahayu.smkn1@gmail.com

Abstract: *Education is the main foundation in shaping the future of students. In this era of technological development and global competition, it is important for schools to provide quality education, and also help students prepare for their next steps after graduating from school. The question of whether students will continue to higher education or immediately enter the world of work is a very important one. The research objects are Lagger Score, Family Status, KIP Data and Number of Siblings. Then classification identification was carried out from the data using the Random Forest, SVM, Naïve Bayes, Decision Tree, Neural Network algorithms in the Orange application. Based on the results of 433 research data that have been tested, precision, recall and accuracy calculation results are obtained for each model. the highest accuracy of Naïve Bayes and Random Forest is 95% (0,956). The results of this research show that the performance of Naïve Bayes and Random Forest is superior to SVM, Decision Tree and Neural Network. Decision Tree: (0.887), and SVM: (0.949) and Neural Network: (0.942).*

Keywords: *Education, Prediction, Algorithms, Classification, Software Engineering*

Abstrak: Pendidikan merupakan landasan utama dalam membentuk masa depan peserta didik. Di era perkembangan teknologi dan persaingan global ini, penting bagi sekolah untuk memberikan pendidikan yang berkualitas, dan juga membantu siswa mempersiapkan langkah selanjutnya setelah lulus sekolah. Pertanyaan apakah siswa akan melanjutkan ke pendidikan tinggi atau segera memasuki dunia kerja merupakan suatu hal yang sangat penting. Objek penelitiannya adalah Lagger Score, Status Keluarga, Data KIP dan Jumlah Saudara. Kemudian dilakukan identifikasi klasifikasi dari data tersebut dengan menggunakan algoritma Random Forest, SVM, Naïve Bayes, Decision Tree, Neural Network pada aplikasi Orange. Berdasarkan hasil 433 data penelitian yang telah diuji diperoleh hasil perhitungan presisi, recall dan akurasi untuk masing-masing model. akurasi tertinggi Naïve Bayes dan Random Forest adalah 95% (0,956). Hasil penelitian ini menunjukkan bahwa kinerja Naïve Bayes dan Random Forest lebih unggul dibandingkan SVM, Decision Tree dan Neural Network. Pohon Keputusan: (0,887), dan SVM: (0,949) dan Jaringan Syaraf Tiruan: (0,942).

Kata Kunci: Pendidikan, Prediksi, Algoritma, Klasifikasi, Rekayasa Perangkat Lunak

INTRODUCTION

Education is the main foundation in shaping the future of students (Mamytbayeva et al., 2024). In the current era of technological development and global competition, it is important for schools to not only provide quality education, but also help students prepare for their next steps after graduating from vocational high schools (SMK). In this context, the question of whether students will continue to college or go straight into the world of work becomes very important (Hora, 2019).

Taruna Bhakti Vocational School is a vocational secondary education institution that has a good reputation for providing vocational education to its students (Putra & Aini, 2023). However, to better understand the factors that

influence students' decisions to continue their studies or enter the world of work, further analysis needs to be carried out.

With the development of data mining and machine learning techniques, data-based predictions are becoming more possible (Yağcı, 2022). The use of classification algorithms such as those provided by the Orange platform provides the ability to identify patterns in historical student data (Adekitan & Noma-Osaghae, 2019). By leveraging data such as academic achievement, participation in extracurricular activities, interests and perhaps other factors, classification algorithms can provide an indication of whether a student is likely to go to college or work directly after graduation.

Through this research, we can provide significant benefits. First, the prediction results from classification algorithms can help schools design education and counseling programs that better suit students' career goals. Second, this information can also be a guide for students and parents in making more informed decisions regarding their next choices after graduating.

In this context, research on job or study predictions for Taruna Bhakti Vocational School students with classification algorithms using the Orange platform can provide a valuable contribution in improving the quality of education, assisting decision making, and preparing students for a more successful future.

METHODOLOGHY

An explanation of the research stages can be described in determining the research object by determining the problems to be studied in this research where the object of this research is the Lagger Value of Taruna Bhakti Vocational School, Family Status, KIP Data and Number of Siblings. Then, classification identification is carried out from the data.

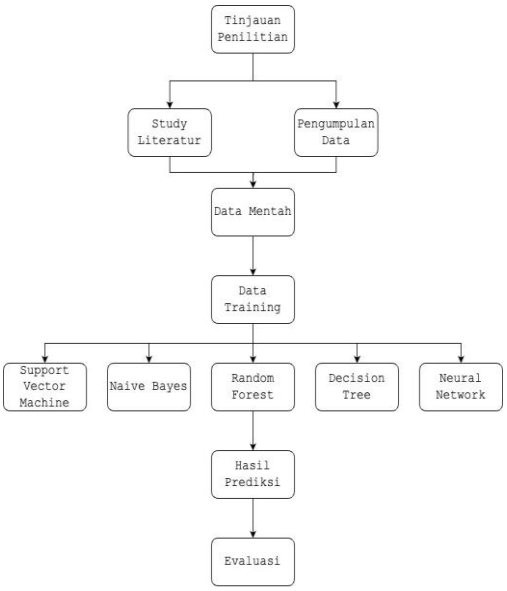

Figure 1. Research Design

After the dataset was collected, the data was processed using Orange Data Mining with the Support Vending Machine (SVM), Naïve Bayes, Random Forest, Decision Tree, Neural Network algorithms, in order to get the best classification from several of these algorithm methods (Urooj et al., 2022).

RESULTS AND DISCUSSION

Collection of data from Taruna Bhakti Vocational School Leger scores, Family Status, KIP Data and Number of Siblings.

Table 1. Value Attribute

No	Atribut	Type	Deskripsi
1	No	*Numeric*	Number
2	NISN	*Numeric*	National Student Identification Number
3	Nama	*Text*	Student Name
4	Kelas	*Categorical*	Student Class
5	Pendidikan Agama	*Numeric*	Subject Grades
6	PPKN	*Numeric*	Subject Grades
7	Bahasa	*Numeric*	Subject

	Indonesia		Grades
8	Matematika	*Numeric*	Subject Grades
9	Sejarah Indonesia	*Numeric*	Subject Grades
10	Bahasa Inggris	*Numeric*	Subject Grades
11	Seni Budaya	*Numeric*	Subject Grades
12	PJOK	*Numeric*	Subject Grades
13	Simulasi Digital	*Numeric*	Subject Grades
14	Fisika	*Numeric*	Subject Grades
15	Kimia	*Numeric*	Subject Grades
16	Bahasa Sunda	*Numeric*	Subject Grades
17	Kejuruan 1	*Numeric*	Subject Grades
18	Kejuruan 2	*Numeric*	Subject Grades
19	Kejuruan 3	*Numeric*	Subject Grades
20	Kejuruan 4	*Numeric*	Subject Grades
21	Kejuruan 5	*Numeric*	Subject Grades
22	Kejuruan 6	*Numeric*	Subject Grades
23	Kejuruan 7	*Numeric*	Subject Grades
24	Kejuruan 8	*Numeric*	Subject Grades
25	Data KIP	*Categorical*	Status KIP ada/tidak
26	Status Keluarga	*Categorical*	Status Keluarga Cukup/Kurang
27	Jumlah Saudara	*Numeric*	Jumlah Saudara Kandung
28	Status Kelulusan	*Categorical*	Status Bekerja atau Kuliah

Data Selection Process / Prepocessing

During the preprocessing process the dataset values are input into the file and pulled into preprocessing in the orange application (Kathuria et al., 2021).

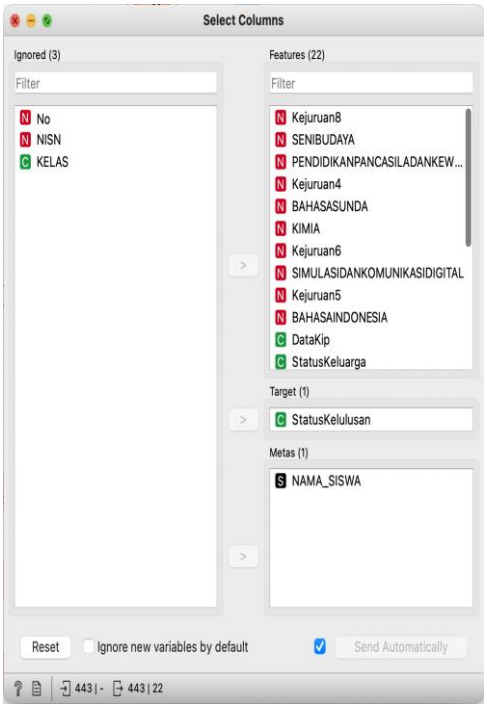

Figure 2. Select Column

Classification Model Testing Process

The process of testing the classification model is by entering the Random Forest, SVM, Naïve Bayes, Decision Tree, Neural Network algorithms in the orange application (Mohi, 2020).

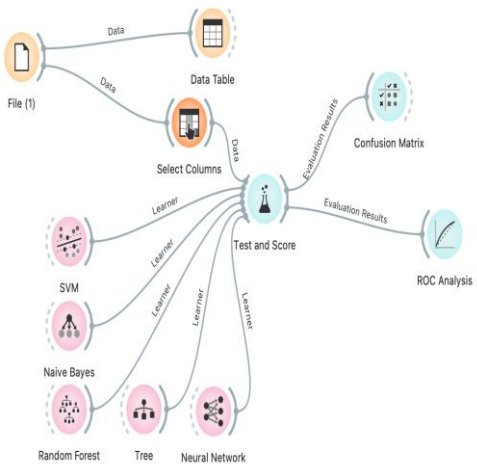

Figure 3. Algorithm Method In The Orange Application

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
August 2024, VII (3): 1091 – 1096 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

Simulation Results of 5 Classification Models

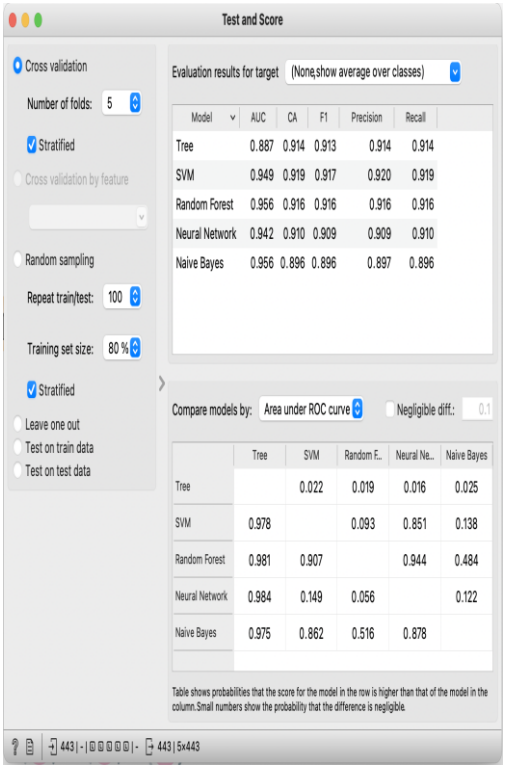

Figure 4. Results On Test And Score

Based on the results of 433 research data that have been tested, the calculation results of Precision, Recall, Accuracy for each SVM, Random Forest, Naïve Bayes, Decision Tree, Neural Network classification model show that the accuracy results of Naïve Bayes and Random Forest are the highest. i.e. 95%. Based on Figure 4.4, the comparison value of 5 AUC models shows that the value of Naïve Bayes and Random Forest is the highest, namely 0.956. AUC is used to measure discriminatory performance by estimating the probability of output from randomly selected examples from a positive or negative population. The greater the AUC, the better the classification results used.

Evaluation Results with Confusion Matrix

Confusion Matrix is a performance measurement for machine learning classification problems where the output can be in the form of 2 or more classes. Confusion Matrix is a table with 4 different mixtures of predicted values and actual values (Hasnain et al., 2020).

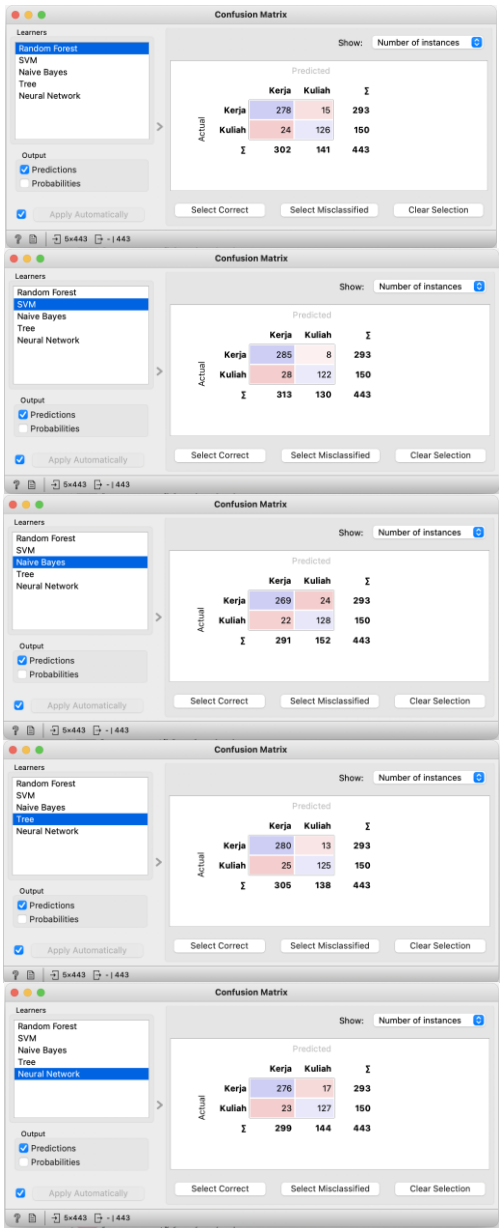

Figure 5. Results of Confusion Matrix Random Forest, SVM, Naïve Bayes, Decision Tree, Neural Network

Evaluation Results with ROC Curve

Manual accuracy values can be done by looking at the ROC curve comparison visualized from the Confusion Matrix. Model viewing ROC curves are the most easily visible way to graphically compare the accuracy values

of each classification model (Kamath & Liu, 2021).

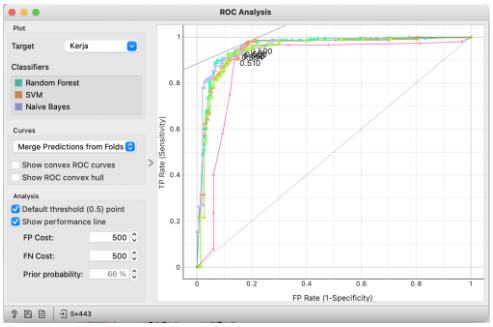

Figure 6. ROC Analysis Results For Work Targets

In Figure 4.6 it is explained that the results of ROC analysis for the work target are Random Forest: 0.500, SVM: 0.471, Naïve Bayes: 0.491, Decision Tree: 0.500 and Neural Network: 0.510. Therefore, for this case study the model that has the best accuracy value is the Neural Network because the curve is close to the 0.1 point.

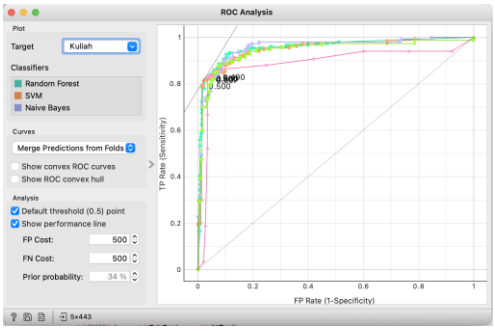

Figure 7. ROC Analysis Results For College Targets

CONCLUTION

The results of the research show that after using the Random Forest, SVM, Naïve Bayes, Decision Tree, Neural Network models to classify work or study predictions for students at Taruna Bhakti Vocational School, Depok, the results showed that the performance of Naïve Bayes and Random Forest was superior to SVM, Decision Tree, and Neural networks. It is proven from 443 test data used that Naive Bayes and Random Forest have an accuracy value of 95% (0.956), while Decision Tree has an accuracy value of 88% (0.887), and SVM has an accuracy value of 94% (0.949) and Neural Network has accuracy value 94% (0.942). The contribution of this research can be utilized by the management of Taruna Bhakti Depok Vocational School to detect students' initial conditions so that their graduation can be in line with their work or study targets.

REFERENCE

Adekitan, A. I., & Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies*, *24*, 1527–1543.

Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. *Ieee Access*, *8*, 90847–90861.

Hora, M. T. (2019). *Beyond the skills gap: Preparing college students for life and work*. Harvard Education Press.

Kamath, U., & Liu, J. (2021). Model Visualization Techniques and Traditional Interpretable Algorithms. In *Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning* (pp. 79–120). Springer.

Kathuria, A., Gupta, A., & Singla, R. K. (2021). A review of tools and techniques for preprocessing of textual data. *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1*, 407–422.

Mamytbayeva, Z. A., Orynbetova, E. A., Kyyakbayeva, U. K., Yeralin, K. Y., & Yeralina, A. K. (2024). Foundations for shaping the research culture of future teachers-educators in higher education institutions. *Bordón: Revista de Pedagogía*,

76(1), 99–117.

Mohi, Z. R. (2020). Orange Data Mining as a tool to compare Classification Algorithms. *Dijlah Journal of Sciences and Engineering*, *3*(3), 13–23.

Putra, A., & Aini, S. (2023). Successful Strategies Utilized by "Senat Taruna Kabinet Komando I" at Jakarta Technical University of Fisheries. *Jurnal Integrasi Sumber Daya Manusia*, *2*(1), 12–30.

Urooj, B., Shah, M. A., Maple, C., Abbasi, M. K., & Riasat, S. (2022). Malware detection: a framework for reverse engineered android applications through machine learning algorithms. *IEEE Access*, *10*, 89031–89050.

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.