Mantik Journal, 7 (4) (2024), ISSN 2685-4236 (Online)

Published by: Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Classifying student academic performance: C4.5 and SVM Methods in ITTP's Computer Engineering Program

Edelin Gultom[1], Yohani Setiya Rafika Nur[2], Rima Dias Ramadhani[3]

[1,2,3]Department of Informatics Engineering, Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Purwokerto

ARTICLE INFO

Article history:

Received Jan 08, 2024
Revised Feb 09, 2024
Accepted Feb 23, 2024

Keywords:

Classification Model;
CRISP-DM;
Data Mining;
Graduation Performance;
Support Vector Machine (SVM).

ABSTRACT

The decline in the number of graduates from the Computer Engineering Program, based on the graduation percentages of the 2017 and 2018 cohorts, coupled with the imbalance between on-time and delayed graduates, poses various challenges. These challenges include suboptimal program accreditation and an excessive number of active students. This research aims to develop a classification model for student performance categorization into On-Time Graduation (LTW) and Delayed Graduation (LTTW) classes using SVM and C4.5 algorithms. The C4.5 algorithm will handle attribute selection, while SVM will be responsible for building the prediction model. The classification results will be visualized on a website using the Flask framework, allowing users to input relevant data. The classification accuracy on the test set, reaching 77.74%, indicates the model's precision in predicting student performance categories.

Corresponding Author:

Yohani Setiya Rafika Nur,
Yohani Setiya Rafika Nur,
Institut Teknologi Telkom Purwokerto, Purwokerto,
Jl. DI Panjaitan No.128, Karangreja, Purwokerto Kidul, Kec. Purwokerto Sel., KabupatenBanyumas, Jawa Tengah 53147.
Email: yohani@ittelkom-pwt.ac.id

1. NTRODUCTION

In academic institutions, students serve as the cornerstone of program success, with their performance and progress significantly impacting program evaluations. Timely graduation, avoidance of dropouts, and sustained attendance are pivotal markers of optimal student performance (Fatma Ayu Rahman et al., 2020). High graduation rates not only reflect the efficacy of educational programs but also underscore an institution's commitment to delivering quality education (Suhaimi et al., 2019) (Wang & Jia, 2020). Regulated by the Ministry of Education and Culture, Directorate General of Higher Education (Janan & Ghosh, 2021), undergraduate programs are expected to be completed within a specified timeframe of 4 years, encompassing a minimum of 144 credit hours.

However, within the Computer Engineering undergraduate program at the Telkom University of Purwokerto (ITTP), there exists a concerning trend. Despite a consistent increase in annual admissions, the number of students graduating on time has witnessed a decline over recent cohorts (Sabna, 2021). For instance, the cohorts from

2017 to 2019 have experienced significant discrepancies between the number of admitted students and those graduating on time (Mengash, 2020). The graduation percentages for these cohorts stand at approximately 69.79%, 66.09%, and 60.4%, respectively. This decline in graduation rates not only raises concerns about program efficacy but also poses challenges that may impact program accreditation (Pambudi et al., 2019). Several previous studies using models to predict performance have been carried out (Qisthiano et al., 2021), Iwan Nurhidayat & Fatrianto, 2021 research using Sequential Minimal Optimization (SMO) which is used to train an SVM model related to predicting student performance (Thaniket & Taufik Luthf, 2020).

In response to this challenge, research into student performance classification is the right solution to form patterns and provide inside information classify categories or classes of student abilities that have not been passed (Tahta et al., 2023). Historical data on students who have graduated is used for classify students who have not graduated into two Pass classes On Time (LTW) and Passing Not on Time (LTTW) (Kurniawan et al., 2020). Guided by the CRISP-DM methodology (Rotty et al., 2022), this research aims to employ data mining techniques to address this issue comprehensively.

In particular, this study will focus on utilizing Support Vector Machine (SVM) and C4.5 algorithms for classification purposes (Dwi Sripamuji et al., 2022). SVM, known for its versatility and high performance in predictive analytics, will be employed to create a predictive model for student performance classification (Widaningsih, 2019) (Nalatissifa et al., 2021). Meanwhile, C4.5 will be utilized for attribute selection (Lestari, 2020), leveraging its strengths in handling various data types and ensuring model stability.

The primary objective of this research is to develop an accurate classification model for evaluating the performance of actively enrolled students in the Computer Engineering program at ITTP. By doing so, this study aims to provide insights that can inform interventions to improve graduation rates and enhance program effectiveness.

2. RESEARCH METHOD

The stages of the research method to be conducted in this study are illustrated in Figure 1.

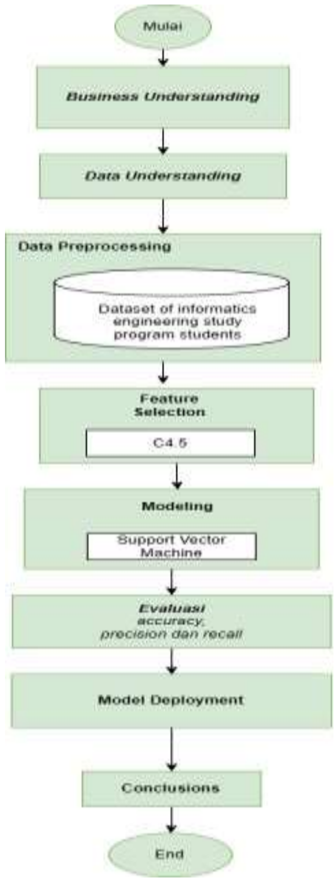

Figure 1. Research Method Diagram

The research follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. The flowchart consists of several stages (Schröer et al., 2021):

a. Business Understanding:
In this stage, the identification of the problem is conducted to understand the issues that need to be addressed. The identified problem is the classification of student performance into two categories: On-Time Graduation (LTW) and Not On-Time Graduation (LTTW).

b. Data Understanding:
This stage involves the collection of data about students enrolled in the Computer Engineering program at ITTP. The data needs to be analyzed to understand its characteristics, such as attribute types and distribution of values. Data for this study is obtained from two main sources: ITTP's academic records and the academic records of the Informatics Program. The dataset includes information on students from three different cohorts: 2017, 2018, and 2019, totaling 784 records.

c. Data Preprocessing:
Data cleaning, handling missing data, outlier treatment, and data format conversion are performed in this stage.

d. Feature Selection:

The C4.5 algorithm is used in this stage to select attributes for inclusion in the predictive model formation (Alfanz et al., 2023) (Yuliansyah et al., 2021). This process is crucial for obtaining a more efficient and accurate predictive model by eliminating or disregarding less relevant or redundant attributes.

e. Modeling:

Modeling is conducted using the Support Vector Machine (SVM) algorithm.

f. Evaluation:

Evaluation is performed to assess the model's performance. Metrics such as accuracy, precision, and recall are used for evaluation (Görtler et al., 2022).

g. Model Deployment:

In this study, the model will be implemented using the Flask framework to create a web-based platform.

h. Conclusion:

The research concludes by summarizing the findings obtained from the analysis and discussing the implications for improving student graduation rates within the Computer Engineering program at ITTP.

3. RESULTS AND DISCUSSIONS

3.1 Data Collection

The study gathered data from two primary sources: the academic records of ITTP and the academic records of the Computer Science Program. The dataset compiled encompasses information on students from three different cohorts, namely 2017, 2018, and 2019. The entirety of this dataset comprises 784 records, providing a comprehensive overview of the academic performance and progress of students over the specified academic years.

3.2 Preprocessing

In this phase of the study, several preprocessing steps were undertaken to ensure the integrity and reliability of the dataset.

a. Removal of Attributes with High Missing Values

A critical preprocessing step involved the removal of data with a target variable that did not align with the graduation status, distinguishing between "Lulus Tidak Tepat Waktu" (LTTW) and "Lulus Tepat Waktu" (LTW). Rows in the DataFrame with a 'Target Variable' not equal to 'LTW' or 'LTTW' were eliminated. This step ensures that the dataset is focused on instances directly related to timely or untimely graduation. The next preprocessing step addressed missing values in the "NILAI PANCASILA DAN KEWARGANEGARAAN (INDEX POINT)" column. Initially, the mean of the existing values in this column was calculated using the mean() function. Subsequently, the empty values in the column were replaced with the computed mean. Figure 3.1 visually depicts the absence of missing values in each column of the dataset. A value of 0 in each column signifies the absence of missing values.

```
In [22]: print(data_bersih.isnull().sum())
NIM                                              0
NAMA                                             0
SKS                                              0
IPK                                              0
nilai D                                          0
nilai E                                          0
IPS1                                             0
IPS2                                             0
IPS3                                             0
IPS4                                             0
IPS5                                             0
IPS6                                             0
NILAI BAHASA INDONESIA (INDEX POINT)             0
NILAI AGAMA (INDEX POINT)                        0
NILAI PANCASILA DAN KEWARGANEGARAAN (INDEX POINT)  0
Variabel Target                                  0
dtype: int64
```

Figure 3.1: Missing Values Handling

Columns encompassing student information such as NIM, NAMA, SKS, IPK, grades (D and E), IPS1 to IPS6, as well as index points for Bahasa Indonesia, Agama, and Pendidikan Pancasila dan Kewarganegaraan, all exhibit a complete dataset with no missing values. These preprocessing steps lay the foundation for a clean and reliable dataset, ensuring that subsequent analyses are based on accurate and complete information. The visual representation in Figure 3.1 provides a transparent overview, reinforcing the dataset's quality after preprocessing.

b. Data Duplication Check

A crucial step in ensuring the integrity of the dataset involved checking for and handling duplicate data. The process involved calculating the number of duplicate entries based on the "NIM" column in the dataset. The output of this analysis revealed that there were no duplicate entries based on "NIM," with the count of duplicates being 0. Figure 3.2 visually represents the Data Duplication Check.

```
In [23]: # Menghitung jumlah data duplikat berdasarkan kolom NIM
         jumlah_duplikat = data_bersih.duplicated(subset='NIM').sum()

         # Menampilkan jumlah data duplikat
         print("Jumlah data duplikat berdasarkan NIM:", jumlah_duplikat)

         Jumlah data duplikat berdasarkan NIM: 0
```

Figure 2: Data Duplication Check - No Duplicate Entries

The absence of duplicate data based on the "NIM" column confirms the uniqueness and reliability of individual student records within the dataset. This is essential for maintaining the accuracy of subsequent analyses and interpretations. To streamline the analysis and enhance model compatibility, the initial target variable, identified as "LTW" (Lulus Tepat Waktu) and "LTTW" (Lulus Tidak Tepat Waktu), was transformed into a binary format. The new representation uses "1" to denote "LTW" and "0" to denote "LTTW". This binary format simplifies the classification task, facilitating a more efficient and standardized approach in subsequent modeling and analysis. These preprocessing steps collectively contribute to a refined and well-structured dataset, laying the groundwork for meaningful and accurate insights into graduation patterns and factors influencing timely or untimely completion.

c. Feature Selection

The process of feature selection was employed using the C4.5 algorithm to enhance model performance (Maqfiroh & Mujiyono, 2022) . In the initial stage, attributes and the target variable were separated to facilitate the attribute selection process. The variable X stored the attributes, while the variable y contained the target variable.

```
In [99]: # Memisahkan atribut dan variabel target
         X = df_c4_5.drop(columns=['Variabel Target'], axis=1)
         y = df_c4_5['Variabel Target']
```

Figure 3: Separating Attributes and Target Variable

Next, the dataset was divided into two subsets: the training data and the test data, utilizing the train_test_split function from the scikit-learn library. This division aimed to evaluate the model's performance on data that it had not encountered during training.

```
In [101]: from sklearn.model_selection import train_test_split

          X_train, X_test, y_train, y_test = train_test_split(X,
                                                              y, test_size=0.2,
                                                              random_state=1)

          print(f'X train: {X_train.shape}')
          print(f'X test: {X_test.shape}')
          print(f'y train: {y_train.shape}')
          print(f'y test: {y_test.shape}')

          X train: (548, 13)
          X test: (137, 13)
          y train: (548,)
          y test: (137,)
```

Figure 4: Dataset Splitting

Subsequently, using the C4.5 algorithm, attribute selection was performed by calculating scores or rankings for each feature. In this stage, SelectKBest was employed with the mutual information method.

```
In [102]: selector = SelectKBest(score_func=mutual_info_classif, k='all')
          X_selected = selector.fit_transform(X, y)
```

Figure 5: Attribute Selection

```
Peringkat atribut:
SKS: 0.21466172698528485
IPS5: 0.13542128208416382
IPK: 0.13264704752457468
IPS2: 0.09594663872679443
IPS4: 0.07841754370991993
IPS6: 0.06777298643413698
nilai D: 0.06029237975323598
IPS1: 0.05677964041517214
nilai E: 0.053841482669269025
IPS3: 0.049698860095539255
NILAI AGAMA (INDEX POINT): 0.030472772737452303
NILAI PANCASILA DAN KEWARGANEGARAAN (INDEX POINT): 0.0176645656755563
NILAI BAHASA INDONESIA (INDEX POINT): 0.013992186088959802
```

Figure 6: Results of Attribute Selection

Figure 3.6 displays the outcomes of the attribute selection process using the C4.5 algorithm with the mutual information method. The scores indicate the strength of the relationship or correlation between each attribute and the target variable. Higher scores signify stronger correlations. Some features, such as scores for Bahasa Indonesia, Agama, and Pancasila dan Kewarganegaraan, exhibit low positive correlations with the target variable. Therefore, in an effort to streamline the model and improve analytical efficiency, a decision was made to eliminate these features. This decision was based on the consideration that the contributions of these features to the target variable were not statistically significant. After the cleaning and feature selection processes, the dataset was refined to consist of 685 data points with retained features: SKS, IPK, grades (D and

E), IPS1 to IPS6, and the Target Variable. These features were deemed as the most relevant for further analysis and modeling.

3.3 Modeling

In this research, the chosen algorithm for classification is Support Vector Machine (SVM) (Cardona & Cudney, 2019). The SVM algorithm is applied to classify data points of students from the Computer Science Program into the categories of "Lulus Tepat Waktu" (LTW) or "Lulus Tidak Tepat Waktu" (LTTW).

3.4 Evaluation Results

After modeling with Support Vector Machine (SVM) on the training data, the evaluation results of the model are as follows: Accuracy measures how well the model can correctly classify instances in the training dataset. In this case, the model achieves an accuracy of 82.97%, indicating its ability to recognize and classify training data correctly. Precision measures the proportion of instances predicted as positive by the model that are actually positive. In this context, the model achieves a precision of 81.54%, indicating its ability to accurately identify positive instances. Recall (sensitivity or true positive rate) measures the proportion of actual positive instances that can be correctly identified by the model. In this case, recall reaches 90.60%, showing the model's ability to capture most of the actual positive instances.

```
Akurasi pada X_train (set pelatihan): 0.8297
Presisi pada X_train (set pelatihan): 0.8154
Recall pada X_train (set pelatihan): 0.9060

Classification Report (Train):
              precision    recall  f1-score   support

           0       0.85      0.73      0.79       177
           1       0.82      0.91      0.86       234

    accuracy                           0.83       411
   macro avg       0.83      0.82      0.82       411
weighted avg       0.83      0.83      0.83       411
```

Figure 7: Classification Report (Train)

Figure 3.7 presents key metrics assessing the training set model performance, with an accuracy of 82.97%, precision of 81.54%, and recall of 90.60%. The Classification Report indicates the model's preference for better classification in class 1 compared to class 0, evidenced by higher precision, recall, and f1-score values for class 1. Subsequently, an evaluation on the test data yields the following results:

```
Akurasi pada X_test (set pengujian): 0.7774
Presisi pada X_test (set pengujian): 0.7624
Recall pada X_test (set pengujian): 0.8846

Classification Report (Test):
              precision    recall  f1-score   support

           0       0.81      0.64      0.71       118
           1       0.76      0.88      0.82       156

    accuracy                           0.78       274
   macro avg       0.78      0.76      0.76       274
weighted avg       0.78      0.78      0.77       274
```

Figure 3.8: Classification Report (Test)

Accuracy on the test set reaches 77.74%, indicating that the model can generally classify instances correctly in the test dataset. Precision on the test set is 76.24%, showing the model's ability to accurately identify actual positive instances in the test dataset. Recall on the test set reaches 88.46%, reflecting the model's ability to capture most of the actual positive instances in the test dataset.

Overall, the evaluation results on the test set indicate that the SVM model performs well in classifying both classes, with a focus on class 1 as the positive class. The model demonstrates a high ability to recognize true positive instances (TP) and is very minimal in making mistakes with no instances falsely predicted as negative (FN) or falsely predicted as positive (FP).

3.5 Result Analysis

Based on the results of the SVM algorithm modeling on the dataset of Computer Science Program students at ITTP, performance evaluation was conducted using accuracy, precision, and recall metrics. On the training set (X_train), the model achieved an accuracy rate of 82.97%, precision of 81.54%, and recall of 90.60%. The classification accuracy on the test set of 77.74% indicates the model's precision in predicting student performance categories. Previous research (Bangun et al., 2022) highlights that SVM has the highest accuracy rate in predicting student graduation time. This finding is consistent with recent research results, where the SVM model on the training set achieves an accuracy rate of 82.97%. Although the accuracy figure in recent research is lower than in previous studies, both provide support for the effectiveness of SVM in the context of predicting student performance. The statement that the RBF kernel is the best choice in modeling results to predict the performance of Computer Science Program students at ITTP aligns with previous research findings. In previous studies, it has been recognized that the success of SVM training processes is not without the contribution of linear, RBF, and polynomial kernels (Erny Herwindiati & Hendryli, 2023) (Iwan Nurhidayat & Fatrianto, 2021). Although each kernel contributes, the difference in accuracy results between kernels reflects different characteristics and distributions of the data. From the modeling results with various kernels in Table 4.6, the RBF kernel becomes the best choice because it provides good accuracy on both training and test datasets, and it is suitable for the non-linear distribution characteristics of the data shown in Figure 4.14. The choice of the RBF kernel can be justified as the most optimal approach for predicting the performance of Computer Science Program students at ITTP.

4. CONCLUSION

The development of an SVM classification model has addressed the challenge of declining graduation rates within ITTP's Computer Engineering program. The model exhibited strong performance metrics, achieving 82.97% accuracy, 81.54% precision, and 90.60% recall on the training set. Although slightly lower, the testing set results remained robust, with 77.74% accuracy, 76.24% precision, and 88.46% recall. Notably, the model effectively distinguished positive instances, indicating meeting performance criteria, while maintaining reliability in avoiding prediction errors. This research underscores the potential application of SVM in student performance management. Future research can go deeper into model evaluation by using additional evaluation metrics such as F1-score, area under the ROC curve (AUC-ROC), and Matthews correlation coefficient (MCC) to get a comprehensive picture of model performance.

REFERENCES

Alfanz, R., Hendrianto, R. K., & Siagian, A. H. A. M. (2023). Predicting Student Performance Through Data Mining: A Case Study in Sultan Ageng Tirtayasa University. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *27*(6), 1159–1167. https://doi.org/10.20965/jaciii.2023.p1159

Bangun, O., Mawengkang, H., & Efendi, S. (2022). Metode Algoritma Support Vector Machine (SVM) Linier Dalam Memprediksi Kelulusan Mahasiswa. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *6*(4), 2006. https://doi.org/10.30865/mib.v6i4.4572

Cardona, T. A., & Cudney, E. A. (2019). Predicting student retention using support vector machines. *Procedia Manufacturing*, *39*, 1827–1833. https://doi.org/10.1016/j.promfg.2020.01.256

Dwi Sripamuji, A., Ramadhanti, I., Riski Amalia, R., Saputra, J., & Prihatnowo, B. (2022). Penerapan Algoritma Support Vector Machine Dan Multi-Layer Perceptron Pada Klasifikasi Topik Berita. *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, *11*(2), 84–91. https://doi.org/10.23887/janapati.v11i2.44151

Erny Herwindiati, D., & Hendryli, J. (2023). Performa Support Vector Machine Pada Klasifikasi Lahan dan Air Tanah. *Jurnal Media Informatika Budidarma*, *7*(1), 231–241. https://doi.org/10.30865/mib.v7i1.5279

Fatma Ayu Rahman, A., Wartulas, S., Raya Pagojengan, J. K., & Brebes, P. (2020). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus Di Universitas Peradaban). *Ade Fatma Ayu Rahman IJIR*, *1*(2), 70–77.

Iwan Nurhidayat, A., & Fatrianto, D. (2021). Prediksi Kinerja Akademik Mahasiswa Menggunakan Machine Learning dengan Sequential Minimal Optimization untuk Pengelola Program Studi. *Journal Information Engineering and Educational Technology*, *5*(2), 84–91.

Janan, F., & Ghosh, S. K. (2021). Prediction of student's performance using support vector machine classifier. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 7078–7088. https://doi.org/10.46254/an11.20211237

Kurniawan, D., Anggrawan, A., & Hairani, H. (2020). Graduation Prediction System On Students Using C4.5 Algorithm. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *19*(2), 358–365. https://doi.org/10.30812/matrik.v19i2.685

Lestari, A. (2020). Increasing Accuracy of C4.5 Algorithm Using Information Gain Ratio and Adaboost for Classification of Chronic Kidney Disease. *Journal of Soft Computing Exploration* , *1*(1), 32–38.

Maqfiroh, F., & Mujiyono, S. (2022). Penerapan Klasifikasi Algoritma Data Mining C4.5 Untuk Memprediksi Tingkat Kelulusan Siswa Di Lembaga Pelatihan Kerja Shinju Semarang. *JAMASTIKA*, *1*(2), 35–49.

Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access*, *8*, 55462–55470. https://doi.org/10.1109/ACCESS.2020.2981905

Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2021). Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja. *Jurnal Informatika Universitas Pamulang*, *5*(4), 578. https://doi.org/10.32493/informatika.v5i4.7575

Görtler, J., Hohman, F., Moritz, D., Wongsuphasawat, K., Ren, D., Nair, R., Kirchner, M., & Patel, K. (2022). Neo: Generalizing Confusion Matrix Visualization to Hierarchical and Multi-Output Labels. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13.

Pambudi, R. D., Supianto, A. A., & Setiawan, N. Y. (2019). *Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya* (Vol. 3, Issue 3). http://j-ptiik.ub.ac.id

Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., & Akbar, M. (2021). Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(3), 987. https://doi.org/10.30865/mib.v5i3.3030

Rotty, A. K., Dewayana, T. S., & Habyba, A. N. (2022). Cross-Industry Standard Process for Data Mining (CRISP-DM) Approach in Determining the Most Significant Employee Engagement Drivers to Sales at X Car Dealership. *Proceedings of the 3rd Asia Pacific International Conference on Industrial Engineering and Operations Management*, 3368–3379.

Sabna, E. (2021). DATA MINING DENGAN 2 (DUA) MODEL KLASIFIKASI UNTUK PREDIKSI KINERJA MAHASISWA. *Jurnal Ilmu Komputer*, *10*(2), 128–131. https://doi.org/10.33060/jik/2021/vol10.iss2.229

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Suhaimi, N. M., Abdul-Rahman, S., Mutalib, S., Hamid, N. H. A., & Malik, A. M. A. (2019). Review on Predicting Students' Graduation Time Using Machine Learning Algorithms. *International Journal of Modern Education and Computer Science*, *11*(7), 1–13. https://doi.org/10.5815/ijmecs.2019.07.01

Tahta, A., Karim, I., & Sudianto, S. (2023). Dominant Requirements for Student Graduation in the Faculty of Informatics using the C4.5 Algorithm. *Journal of Dinda Data Science, Information Technology, and Data Analytics*, *3*(2), 50–58. http://journal.ittelkom-pwt.ac.id/index.php/dinda

Thaniket, R., & Taufik Luthf, E. (2020). PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. In *Jurnal Teknologi dan Rekayasa* (Vol. 5, Issue 2).

Wang, P., & Jia, Y. (2020). *Research on Prediction of College Students' Performance Based on Support Vector Machine*.

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, KNN Dan SVM. *Jurnal Tekno Insentif*, *13*(1), 16–25. https://doi.org/10.36787/jti.v13i1.78

Yuliansyah, H., Imaniati, R. A. P., Wirasto, A., & Wibowo, M. (2021). Predicting Students Graduate on Time Using C4.5 Algorithm. *Journal of Information Systems Engineering and Business Intelligence*, *7*(1), 67. https://doi.org/10.20473/jisebi.7.1.67-73

.