

# Teknik Prediksi Data Mining pada Perguruan Tinggi sebagai Kajian Literatur

Robby Yuli Endra <sup>1\*</sup>, Usman Rizal <sup>2</sup>

<sup>1</sup> Informatika, Fakultas Ilmu Komputer, Universitas Bandar Lampung, Bandar Lampung, Indonesia

<sup>2</sup> Sistem informasi, Fakultas Ilmu Komputer, Universitas Satu Nusa Lampung, Bandar Lampung, Indonesia

<sup>1\*</sup> robbiyuliendra@ubl.ac.id, <sup>2</sup> usmanrizal@usn-lampung.ac.id

**ABSTRACT** – Higher education is an interesting research object because there are complex research topics that can be used by students, lecturers or researchers. One of the research topics that uses university research objects is related to data mining. Data mining is a technique that can be used to extract data into knowledge and functions to find patterns from large data. There are currently 5 data mining roles used, namely Estimation, Prediction, Classification, Clustering and Association. One of the techniques used is Prediction. The aim of this research is to conduct systematic literature review research related to data mining prediction techniques with 3 aspects, namely Algorithms, Frameworks/Methods and Research Topics. The results of this research are the Algorithm Trends used in the research, namely the Support Vector Machines (SVM), Random Forest, Decision tree (DT), Artificial Neural Networks (ANN), Naive Bayes (NB), Neural Network (NN), and K- Nearest Neighbor (KNN) Framework used is Data Mining, Educational data mining and Machine Learning and the trending topic is related to Prediction, namely measuring Student performance.

**Keywords:** Algorithm; Data Mining; Forecast; SLR

**ABSTRAK** – Perguruan tinggi menjadi salah satu objek penelitian yang menarik karena terdapat topik-topik penelitian yang kompleks yang dapat digunakan oleh mahasiswa, dosen ataupun peneliti. Salah satu topik penelitian yang menggunakan objek penelitian perguruan tinggi adalah terkait dengan data mining. Data mining merupakan salah satu teknik yang dapat digunakan untuk mengekstraksi data menjadi pengetahuan dan berfungsi untuk menemukan pola dari suatu data yang besar. Peran data mining saat ini yang digunakan ada 5 yaitu Estimasi, Prediksi, Klasifikasi, Klusterisasi dan Asosiasi. Salah satu yang teknik yang digunakan adalah Prediksi. Tujuan dari penelitian ini peneliti melakukan penelitian Systematic Literatur Review terkait dengan Teknik Prediksi data mining dengan 3 aspek yaitu Algoritma, Framework/Metode dan Topik Penelitian. Hasil penelitian ini yaitu Tren Algoritma yang digunakan pada penelitian yaitu Algoritma Support Vector Machines (SVM), Random Forest, Decision tree (DT), Artificial Neural Networks (ANN), Naive Bayes (NB), Neural Network (NN), and K-nearest Neighbour (KNN) Framework yang digunakan yaitu Data Mining, Educational Data Mining dan Machine Learning dan topik yang menjadi tren yaitu terkait dengan prediksi yaitu mengukur student performance.

**Kata Kunci:** Algoritma; Data Mining; Prediksi; SLR

## 1. PENDAHULUAN

Manusia merupakan penghasil data yang besar dan sangat bervariasi di era digital saat ini [1]. Bagi organisasi data merupakan aset yang penting selain infrastruktur dan Sumber daya manusia. Data yang sangat melimpah di organisasi tentunya menjadi aspek yang dapat digali untuk menjadikan informasi dari data tersebut dapat berguna dan menghasilkan pengetahuan yang baru. Sistem informasi Akademik merupakan penghasil data yang banyak pada Universitas, setiap semester tentu data yang berhasil disimpan akan terus bertambah seperti (1) data mahasiswa baru, (2) data profil mahasiswa (3) data dosen, (4) data nilai, (5) data *feedback* mahasiswa, (6) data publikasi dan data lainnya. Hal tersebut akan menjadi sangat disayangkan apabila data tersebut hanya tersimpan di sistem basis data, seharusnya data tersebut dapat

digunakan lebih lanjut untuk mendapatkan informasi yang bermanfaat untuk universitas.

*Data mining* merupakan salah satu teknik yang dapat digunakan untuk mengekstraksi data menjadi pengetahuan dan berfungsi untuk menemukan pola dari suatu data yang besar [2]. *Data mining* merupakan proses interaktif dan berulang-ulang dari berbagai sub bagian tugas dan keputusan yang sering disebut dengan *Knowledge Discovery from Data (KDD)*. Data yang ditemukan kemudian diproses menjadi pengetahuan yang dapat berguna untuk pengambilan keputusan disebut dengan *data mining* [3]. *Data mining* memiliki 5 peran utama [4] (1) *Estimation* (2) *Forecasting* (3) *Classification* (4) *Clustering* (5) *Assosiation*. Masing-masing peran tersebut memiliki algoritma untuk menyelesaikan sebuah permasalahan seperti algoritma untuk (1) *Estimation* dapat menggunakan algoritma *Linear Regression (LR)*, *Neural Network (NN)*, *Deep Learning (DL)*, *Support Vector Machine (SVM)*,



*Generalized Linear Model (GLM)*, (2) *Forecasting* dapat menggunakan algoritma *Linear Regression (LR)*, *Neural Network (NN)*, *Deep Learning (DL)*, *Support Vector Machine (SVM)*, *Generalized Linear Model (GLM)*, (3) *Classification* menggunakan Algoritma *Decision Tree (CART, ID3, C4.5, Credal DT, Credal C4.5, Adaptive Credal C4.5)*, *Naive Bayes (NB)*, *K-Nearest Neighbor (kNN)*, *Linear Discriminant Analysis (LDA)*, *Logistic Regression (LogR)*, (4) *Clustering* dapat menggunakan Algoritma *K-Means*, *K-Medoids*, *Self-Organizing Map (SOM)*, *Fuzzy C-Means (FCM)* serta (5) *Association* dapat menggunakan Algoritma *FP-Growth*, *A Priori*, *Coefficient of Correlation*, *Chi Square* [4].

Konteks *role* model utama pada penelitian ini khususnya di objek Perguruan tinggi atau universitas adalah terkait dengan Teknik Prediksi yang dilakukan oleh peneliti untuk melakukan penelitian. Tujuan dari penelitian ini adalah melakukan penelitian *Systematic Literatur Review* terkait dengan Teknik Prediksi *data mining* dengan 3 aspek yaitu Algoritma, Data Set dan Topik Penelitian. Dari penjelasan tersebut peneliti membuat pertanyaan penelitian yaitu: (1) Apa saja Topik atau Pembahasan yang digunakan pada penelitian tersebut? (2) Apa saja Algoritma yang digunakan pada penelitian tersebut? (3) Apa saja *Framework* atau metode yang digunakan pada penelitian tersebut?

Pertanyaan penelitian tersebut bagian terkait dengan judul penelitian ini yaitu terkait dengan teknik prediksi *data mining* yang digunakan pada perguruan tinggi atau universitas. Hal tersebut penting untuk dilakukan penelitian sebagai referensi penelitian selanjutnya.

## 2. METODOLOGI

Metode penelitian yang dilakukan adalah menggunakan *Systematic Literatur Review (SLR)*, dengan menggunakan tahap-tahap Proses pada penelitian ini dikelompokkan dengan beberapa tahap penelitian yaitu menentukan sumber atau *database* penelitian, menentukan kata kunci untuk proses pencarian disumber penelitian dengan kesesuaian topik penelitian, menentukan kriteria data yaitu kriteria inklusi dan kriteria eksklusif, tahap selanjutnya memproses data yang akan disesuaikan dengan menyaring artikel yang akan digunakan dan tahap terakhir adalah tahap menganalisis dan menarik kesimpulan untuk menjawab pertanyaan penelitian.

### Tahap Pertama Proses Pencarian Paper

Proses pertama yang dilakukan pada tahap *Systematic Literatur Review (SLR)*, pada penelitian ini dilakukan tahap pencarian yang terdiri dari 6 Sumber yaitu:

- Scopus (<https://www.scopus.com/>)
- IEEE Xplore (<https://ieeexplore.ieee.org/Xplore/home.jsp>)
- Science Direct (<https://www.sciencedirect.com/>)
- Emerald Insight (<https://www.emerald.com/insight/>)

- Taylor & Francis Online <https://www-tandfonline-com>)
- Clarivate (Web of Science) (<https://www-webofscience-com>)

Pola Pencarian kata kunci untuk pencarian makalah yang berkaitan dengan jawaban pertanyaan penelitian dibentuk dengan menggunakan operator *boolean* untuk proses penyaringan data, dengan model pencarian tersebut peneliti mendapatkan prioritas untuk makalah penelitian dengan simbol-simbol yang digunakan dalam pencarian tersebut.

Adapun Operator-operator yang digunakan dalam pencarian tersebut menggunakan operator *And*, *Or* dan *Not*, kombinasi operator dapat dilihat dibawah berikut ini:

- forecasting AND data AND mining AND techniques AND higher AND education*
- predictions AND data AND mining AND higher AND education*
- ((("Forecasting" OR "predictions") AND ("Data mining" OR "Big Data" OR "Data Science"))) AND ("Higher Education" OR "University")*
- "Forecasting" OR "predictions" And Data Mining*
- data AND mining AND forecasting AND techniques AND in AND higher AND education*
- data AND mining AND higher AND education*

Pada proses pencarian artikel dilakukan dengan menggunakan 3 proses pencarian yang pertama *Studies Found* yaitu proses pencarian yang terkait dengan artikel yang muncul pada sumber pencarian berdasarkan kata kunci dan kriteria inklusi yang ditetapkan oleh peneliti. Adapun kriteria inklusi yang ditetapkan yaitu:

- Tahun publikasi artikel dimulai tahun 2020-2023
- Tipe Dokumen dan sumber yang dipilih berasal dari Jurnal dan *Conference*
- Tipe dari Artikel terstruktur dan lengkap
- Artikel yang ditulis menggunakan Bahasa Inggris

Pada Proses yang kedua peneliti menyebutnya adalah Kandidat studi yaitu artikel yang memiliki kesesuaian antara Judul, Abstrak dan *keyword* sesuai dengan pertanyaan pada penelitian ini. Kemudian untuk bagian ke tiga atau bagian terakhir untuk penyaringan artikel ini adalah tahap *Selected Studies* merupakan tahap dimana makalah bisa menjawab pertanyaan dari penelitian ini.

### Ekstraksi Data

Pada penelitian *Systematic Literatur Review* ini didapatkan 495 artikel yang berasal dari semua sumber yang didapatkan dari hasil pencarian dengan menggunakan kata kunci. Dari 495 artikel yang didapatkan kemudian dicari dengan menyesuaikan judul dan abstrak untuk dijadikan kandidat studi, *paper* yang didapatkan sebanyak 87 artikel. Kemudian setelah dianalisis dan ditelaah dengan seksama maka terdapat 39 Artikel yang terpilih untuk penelitian ini.

Tabel 1 Ekstrasi Data dengan Kriteria Inklusi

Source	Studies Found	Candidate Studies	Selected Studies
Scopus	55	35	25
IEEE Explore	22	4	3
Science Direct	161	26	6
Taylor & Francis Online	72	11	2
Clarivate (Web of Science)	18	9	3
Emerald	167	2	0
Total	495	87	39

### 3. HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini adalah mencari informasi terkait dengan teknik prediksi *data mining* yang digunakan oleh para peneliti di dunia yang menjadikan perguruan tinggi menjadi bagian dari objek atau topik penelitian. Dengan mulai kompleksnya data yang ada di perguruan tinggi, teknik *data mining* dapat menjadi model atau pengetahuan untuk pengolah data tersebut sehingga data tersebut dapat menghasilkan informasi yang lebih bermanfaat untuk kemajuan perguruan tinggi. Peran *data mining* dibagi menjadi 5 yang telah dijelaskan pada poin 1,

dan pada penelitian ini peneliti hanya mengambil topik satu peran yaitu Prediksi. Berdasarkan hal tersebut peneliti ingin melakukan identifikasi artikel-artikel yang membahas tentang teknik prediksi pada *data mining* yang meliputi algoritma yang digunakan dalam penelitian, Topik Penelitian, *Software* yang digunakan untuk pengembangan serta jika ada *framework* yang digunakan pada penelitian yang didapat. Selain itu artikel ini menyediakan karakteristik demografi dan tren literatur pada *paper* yang terpilih. Pada Tabel 2 dijelaskan terkait informasi artikel yakni judul jurnal, tahun, dan tipe *paper*, berasal dari *conference* atau jurnal.

Tabel 2 Sumber Publikasi

No	Title	Years	Type
1.	Student Performance Prediction Model ...[5]	2020	C
2.	Forecasting Graduation Schedule Model of Higher Education...[6]	2023	J
3.	Comparison of Predictive Models with Balanced Classes...[7]	2021	J
4.	Predicting Student Retention Among a Homogeneous...[8]	2020	J
5.	An Overview on the Use of Educational Data Mining...[9]	2021	C
6.	Predicting Academic Performance of Immigrant Students Using... [10]	2022	C
7.	Prediction of educationist's performance using regression model [11]	2020	C
8.	Prediction of Student Performance in Higher Academy... [12]	2022	C
9.	Developing web-based support systems for predicting poor-performing students ...[13]	2020	J
10.	Application of Data Mining in Predicting College...[14]	2021	C
11.	A review on Student Performance Prediction using Educational Data mining...[15]	2021	C
12.	Early Predicting of Students Performance...[16]	2023	J
13.	An effective recommendation system to forecast the best educational program...[17]	2020	J
14.	Drop-Out Prediction in Higher Education...[18]	2020	J
15.	Churn prediction in digital game-based learning using data mining techniques...[19]	2022	J
16.	Analyzing Applicant's Pre-Admission data and Predicting Applicant's Performance...[20]	2022	C
17.	Educational Data Mining Utilization to Support the Admission Process in Higher Education...[21]	2023	C
18.	Using data mining techniques to predict student performance...[22]	2020	J
19.	Scholarship Recipients Prediction Model using k-Nearest Neighbor Algorithm...[23]	2022	C
20.	A Regression Model to Predict...[24]	2022	C
21.	Simulation of Machine Learning Techniques to Predict Academic Performance [25]	2022	C
22.	Performance Evaluation of Machine Learning Techniques for Prediction of Graduating Students...[26]	2020	C
23.	Review of Data Mining Techniques in Performance Prediction...[27]	2023	C
24.	Data Mining Students' performance in a Higher...[28]	2021	C
25.	Prediction of Study Period and Information...[29]	2022	C
26.	Prediction of students' assessment readiness in online learning...[30]	2020	C
27.	Predicting the On-The-Job Training Performance of IT Students...[31]	2022	C
28.	Predicting Student Retention in Higher Education Using Data Mining...[32]	2023	C
29.	Predicting Students Academic Performance using...[33]	2020	C
30.	Student Dropout Prediction: A KMUTT Case...[34]	2021	C
31.	A Recommendation System for Selecting the Appropriate Undergraduate Program at Higher Education...[35]	2022	J
32.	Determining the Factors Influencing the Academic Accomplishment of Students...[36]	2023	C
33.	Analysis of machine learning strategies for prediction of passing...[37]	2022	C
34.	Student Performance Prediction on Primary and Secondary Schools...[38]	2022	J

No	Title	Years	Type
35.	Machine learning in the enrolment management process: a case study of using GANs in postgraduate students'...[39]	2022	J
36.	Learning behaviours data in programming education...[40]	2022	J
37.	Prediction of middle school students' programming...[41]	2020	J
38.	Predicting the academic progression in student's...[42]	2022	J
39.	Enhancing data pipelines for forecasting student performance...[43]	2021	C

Klasifikasi penulis berdasarkan Disiplin Ilmu Kompter sebesar (33,61%), Teknologi informasi (31,09 %), Sistem Informasi (21,01%), Matematika dan Statistik (5,04%), Manajemen dan Pemasaran (2,52%), Ilmu Pendidikan dan Teknik Komputer (1,68 %), ilmu teknik,

Ilmu kimia, Ilmu Administrasi dan Terapan sebesar (0,84). Secara detail dapat dilihat Pada Table 3 terkait keilmuan yang melakukan penelitian terkait dengan teknik prediksi dengan menggunakan *data mining*.

**Tabel 3** Daftar Penulis Sesuai Bidang Keilmuan

No	Displin	Jumlah	Persentasi (%)
1.	Ilmu Komputer	40	33,61
2.	Teknologi Informasi	37	31,09
3.	Sistem Informasi	25	21,01
4.	Matematika dan Statistik	6	5,04
5.	Manajemen dan Pemasaran	3	2,52
6.	Ilmu Pendidikan	2	1,68
7.	Teknik Komputer	2	1,68
8.	Ilmu Teknik	1	0,84
9.	Ilmu Kimia	1	0,84
10.	Ilmu Administrasi	1	0,84
11.	Ilmu Terapan	1	0,84
	<b>Total</b>	<b>119</b>	<b>100</b>

**Tabel 4** Klasifikasi Jawaban Penelitian

Algoritma	Framework/ Metode	Topik	Judul
Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Sequential Minimal Optimisation and Neural Network.	Supervised machine learning algorithms	Student Performance Prediction	Student Performance Prediction Model ...[5]
Tree algorithm, forward selection analysis, cross-validation techniques, and confusion matrix performance	Data mining techniques, including CRISP-DM	factors affecting students' academic achievement	Forecasting Graduation Schedule Model of Higher Education...[6]
Random Forest Model	SMOTE method	Forecast of Student Dropout	Comparison of Predictive Models with Balanced Classes... [6]
Classification algorithms	Data Mining	Student Retention Among a Homogeneous Population	Predicting Student Retention Among a Homogeneous...[8]
Regression Model	Educational Data Mining methodologies	Mitigate Retention Student	An Overview on the Use of Educational Data Mining...[9]
Prediction Algorithm	XGBoost	Academic Performance of Immigrant Students	Predicting Academic Performance of Immigrant Students Using... [10]
Linear Regression, Logistics Regression	Educational data mining	Prediction of educationist's performance	Prediction of educationist's performance using regression model [11]
K-Nearest Neighbor, Decision Tree, Gaussian Naive Bayes, Support Vector Machine, Logistic Regression, and Random Forest.	Educational data mining	Prediction of Student Performance in Higher Academy	Prediction of Student Performance in Higher Academy... [12]
Hybrid random forest	Educational data mining	Predicting poor-performing students	Developing web-based support systems for predicting poor-performing students ...[13]
Random forest algorithm	Data mining	Predicting College Graduates Employment	Application of Data Mining in Predicting College...[14]
Decision Tree, Random Forest, Genetic Algorithm, Artificial Neural Networks	Educational data mining	Predicting student's performance	A review on Student Performance Prediction using Educational Data mining...[15]
Logistic Recognition, Random forest, KNN, SVM, GNB,and XGB	Data mining	Predicting student's performance	Early Predicting of Students Performance...[16]

Algoritma	Framework/ Metode	Topik	Judul
Random Forest algorithm rec	Educational data mining and machine learning	Predicting Student's academic performance	An effective recommendation system to forecast the best educational program...[17]
Random Forest, Decision tree (DT), Artificial Neural Network (ANN)	Machine learning	Student Drop-out prediction	Drop-Out Prediction in Higher Education...[18]
Logistic regression, decision tree, and random forest	Data mining	Churn prediction in digital game-based learning	Churn prediction in digital game-based learning using data mining techniques...[19]
Naïve bayes, KNN and Logistic Regression	Educational data mining	Predicting Applicant's Performance	Analyzing Applicant's Pre-Admission data and Predicting Applicant's Performance...[20]
Decision Trees (DT), Logistic Regression (LR), Artificial Neural Networks (ANN), Naive Bayes (NB), Support Vector Machines (SVM), K-Nearest Neighbor (KNN)	Educational Data Mining	Support the Admission Process in Higher Education	Educational Data Mining Utilization to Support the Admission Process in Higher Education...[21]
Support vector machines (SVM)	Educational Data Mining	Predicting Applicant's Performance	Using data mining techniques to predict student performance...[22]
k-Nearest Neighbor	Educational Data Mining	Scholarship Recipients Prediction Model	Scholarship Recipients Prediction Model using k-Nearest Neighbor Algorithm...[23]
Linear regression	Data Mining	Predict Key Performance Indicators in Higher Education	A Regression Model to Predict...[24]
Decision Tree (DT), Random Forest (RF), K-Nearest Neighbour (KNN), Logistic Regression, Naïve Bayes (NB), Gradient Boost Tree (GBT), Multi-Linear Perceptron (MLP), Support Vector Machine (SVM)	Educational Data Mining	Predict Academic Performance	Simulation of Machine Learning Techniques to Predict Academic Performance [25]
Artificial Neural Networks, Naive Bayes	Education data mining	Prediction of Graduating Students	Performance Evaluation of Machine Learning Techniques for Prediction of Graduating Students...[26]
Decision Tree (DT),K-Nearest Neighbour (KNN), Logistic Regression, Naïve Bayes (NB),Support Vector Machine (SVM), Artificial Neural Networks (ANN)	Data Mining	Performance Prediction	Review of Data Mining Techniques in Performance Prediction...[27]
Decision Tree (DT), Naïve Bayes (NB), Artificial Neural Networks (ANN)	Data Mining	Students' performance	Data Mining Students' performance in a Higher...[28]
C5.0, Decision Tree	Data Mining	Prediction of Study Period and Information on Judiciary in Higher Education	Prediction of Study Period and Information...[29]
Long Short-Term Memory Algorithm	Deep Learning	Prediction of students' assessment readiness	Prediction of students' assessment readiness in online learning...[30]
Naive Bayes	Data mining	Predicting the On-The-Job Training Performance of IT Students	Predicting the On-The-Job Training Performance of IT Students...[31]
Decision tree (DT), random forest (RF), logistic regression (LR), neural network (NN), AdaBoost, XGBoost, and long short-term memory (LSTM) algorithms	Data Mining	Predicting Student Retention	Predicting Student Retention in Higher Education Using Data Mining...[32]
Naive Bayes, Bagging, Boosting and Random Forest	Education data mining	Predicting Students Academic Performance	Predicting Students Academic Performance using...[33]
Decision Tree, Random Forest	Education Data Mining	Student Dropout Prediction	Student Dropout Prediction: A KMUTT Case...[34]
Decision Tree, Random Forest	Education Data Mining	A Recommendation System for Selecting the Appropriate Undergraduate Program	A Recommendation System for Selecting the Appropriate Undergraduate Program at Higher Education...[35]
Random Forest (RF), Support Vector Machines (SVM), Naive Bayes (NB), Neural Network (NN), and K-nearest Neighbour (KNN)	Education Data Mining	Predicting Students Success	Determining the Factors Influencing the Academic Accomplishment of Students...[36]
Support Vector Machines (SVM)	Machine Learning	Prediction of passing undergraduate admission test	Analysis of machine learning strategies for prediction of passing...[37]
Random Forest, Decision tree (DT)	Machine Learning	Student performance prediction	Student Performance Prediction on Primary and Secondary Schools...[38]
Random Forest	Machine learning	Student performance prediction	Machine learning in the enrolment management process: a case study of using GANs in postgraduate students'...[39]

Algoritma	Framework/ Metode	Topik	Judul
Random Matrix Theory	Educational data mining	prediction of the students' learning performance	Learning behaviours data in programming education...[40]
ANN	Machine learning	Prediction of middle school students' programming talent	Prediction of middle school students' programming...[41]
Support Vector Machines (SVM)	Machine learning	Predicting the academic progression in student's standpoint	Predicting the academic progression in student's...[42]
DMMs LR and GLMNET	Educational data mining	Forecasting student performance	Enhancing data pipelines for forecasting student performance...[43]

#### 4. KESIMPULAN

Penelitian ini memiliki implikasi terdapat 2 bagian yaitu Implikasi Teori terkait dengan hasil penelitian *Systematic Literatur Review (SLR)* adalah secara teori hasil penelitian ini dapat menjadi acuan peneliti selanjutnya untuk melanjutkan penelitian terkait dibidang *Data Mining*. Secara praktik peneliti selanjutnya dapat mengetahui Teknik Prediksi yang banyak digunakan sebagai topik penelitian adalah terkait tentang *Student Performance*. Untuk algoritma yang banyak digunakan pada penelitian ini sangat bervariasi dari Algoritma *Support Vector Machines (SVM)*, *Random Forest*, *Decision tree (DT)*, *Artificial Neural Networks (ANN)*, *Naive Bayes (NB)*, *Neural Network (NN)*, *K-nearest Neighbour (KNN)* dan lainnya yang dapat dilihat pada Tabel 4. Terkait metode dan *framework* yang digunakan pada penelitian ini adalah *Data Mining*, *Educational data mining*, dan *Machine Learning*.

Dengan data yang ditemukan tersebut peneliti selanjutnya dapat mengetahui topik penelitian apa saja yang sudah dilakukan oleh peneliti sebelumnya, misalnya terkait dengan siswa berprestasi, evaluasi pengajaran dan lainnya. Selain itu dari hasil penelitian ini algoritma dan metode apa saja yang digunakan dalam penelitian tersebut sehingga peneliti dapat membandingkan hasil penelitian sebelum dan sesudah, atau dapat mencoba algoritma atau metode yang belum banyak digunakan oleh peneliti sebelumnya.

#### 5. BATASAN DAN PENELITIAN SELANJUTNYA

Pada penelitian ini komponen diteliti berdasarkan algoritma yang digunakan, kerangka kerja atau *framework* yang digunakan, serta topik penelitian yang digunakan untuk penelitian pada objek penelitian di perguruan tinggi atau universitas. Komponen yang didapat bervariasi tentunya dapat dipertimbangkan guna melakukan penelitian selanjutnya, misalnya terkait algoritma yang cocok untuk melakukan prediksi atau *framework/metode* apa yang digunakan dalam *data mining* pada penelitian tersebut. Komponen tersebut merupakan komponen yang didapat dari artikel-artikel yang memenuhi kriteria yang dilakukan pada penelitian ini. Tentunya tidak mudah untuk melakukan proses organisasi dari komponen-komponen tersebut, meskipun penelitian yang sesuai kata kunci banyak tetapi akses *database* penelitian sangat terbatas untuk menggambarkan fakta-fakta penelitian secara detail dan lengkap.

#### DAFTAR PUSTAKA

- [1] J. Liu, "The Impact of Data Mining on Management and Digital Marketing in the Age of Big Data," *Comput. Aided. Des. Appl.*, vol. 21, no. S4, pp. 229–247, 2024, doi: 10.14733/cadaps.2024.S4.229-247.
- [2] A. Motaghifard, M. Omidvari, and A. Kazemi, "Forecasting of safe-green buildings using decision tree algorithm: data mining approach," *Environ. Dev. Sustain.*, vol. 25, no. 9, pp. 10323–10350, 2022, doi: 10.1007/s10668-022-02491-4.
- [3] F. Lemke and J.-A. Müller, "Self-Organising Data Mining," *Syst. Anal. Model. Simul.*, vol. 43, no. 2, pp. 231–240, Feb. 2003, doi: 10.1080/0232929031000136135.
- [4] D. T. Larose and Chantal D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. USA: John Wiley & Sons, Inc, 2014.
- [5] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 928, no. 3, doi: 10.1088/1757-899X/928/3/032019.
- [6] W. S. Nuankaew, T. Sittiwong, S. Bussaman, P. Nasa-Ngium, and P. Nuankaew, "Forecasting Graduation Schedule Model of Higher Education Learners using Feature Selection Techniques," *Int. J. Eng. Trends Technol.*, vol. 71, no. 4, pp. 354–358, 2023, doi: 10.14445/22315381/IJETT-V71I4P231.
- [7] V. Flores, S. Heras, and V. Julián, "Comparison of Predictive Models with Balanced Classes for the Forecast of Student Dropout in Higher Education," in *Communications in Computer and Information Science*, 2021, vol. 1472 CCIS, pp. 139–152, doi: 10.1007/978-3-030-85710-3\_12.
- [8] G. Bilquise, S. Abdallah, and T. Kobbaey, "Predicting Student Retention Among a Homogeneous Population Using Data Mining," in *Advances in Intelligent Systems and Computing*, 2020, vol. 1058, pp. 35–46, doi: 10.1007/978-3-030-31129-2\_4.
- [9] T. N. De Oliveira, F. Bernardini, and J. Viterbo, "An Overview on the Use of Educational Data Mining for Constructing Recommendation Systems to Mitigate Retention in Higher Education," in *Proceedings - Frontiers in Education Conference, FIE*, 2021, vol. 2021-Octob, doi: 10.1109/FIE49875.2021.9637207.



- [10] S. Jeganathan, A. R. Lakshminarayanan, N. Ramachandran, and G. B. Tunze, "Predicting Academic Performance of Immigrant Students Using XGBoost Regressor," *Int. J. Inf. Technol. Web Eng.*, vol. 17, no. 1, 2022, doi: 10.4018/IJITWE.304052.
- [11] S. Arora, M. Agarwal, and R. Kawatra, "Prediction of educationist's performance using regression model," in *Proceedings of the 7th International Conference on Computing for Sustainable Global Development, INDIACom 2020*, 2020, pp. 88–93, doi: 10.23919/INDIACom49435.2020.9083708.
- [12] A. T. Rahman and A. Raihan, "Prediction of Student Performance in Higher Academy Through Different Data Mining Techniques," in *4th International Conference on Electrical, Computer and Telecommunication Engineering, ICECTE 2022*, 2022, doi: 10.1109/ICECTE57896.2022.10114531.
- [13] P. Sokkhey and T. Okazaki, "Developing web-based support systems for predicting poor-performing students using educational data mining techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, pp. 23–32, 2020, doi: 10.14569/IJACSA.2020.0110704.
- [14] S. He, X. Li, and J. Chen, "Application of Data Mining in Predicting College Graduates Employment," in *2021 4th International Conference on Artificial Intelligence and Big Data, ICAIBD 2021*, 2021, pp. 65–69, doi: 10.1109/ICAIBD51990.2021.9459039.
- [15] P. S. Pawar and R. Jain, "A review on Student Performance Prediction using Educational Data mining and Artificial Intelligence," in *2021 IEEE 2nd International Conference on Technology, Engineering, Management for Societal Impact using Marketing, Entrepreneurship and Talent, TEMSMET 2021*, 2021, doi: 10.1109/TEMSMET53515.2021.9768773.
- [16] E. Alhazmi and A. Sheneamer, "Early Predicting of Students Performance in Higher Education," *IEEE Access*, vol. 11, pp. 27579–27589, 2023, doi: 10.1109/ACCESS.2023.3250702.
- [17] J. Dhar and A. K. Jodder, "An effective recommendation system to forecast the best educational program using machine learning classification algorithms," *Ing. des. Syst. d'Information*, vol. 25, no. 5, pp. 559–568, 2020, doi: 10.18280/ISI.250502.
- [18] N. S. Sani, A. F. M. Nafuri, Z. A. Othman, M. Z. A. Nazri, and K. Nadiyah Mohamad, "Drop-Out Prediction in Higher Education Among B40 Students," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, pp. 550–559, 2020, doi: 10.14569/IJACSA.2020.0111169.
- [19] M. Kiguchi, W. Saeed, and I. Medi, "Churn prediction in digital game-based learning using data mining techniques: Logistic regression, decision tree, and random forest," *Appl. Soft Comput.*, vol. 118, 2022, doi: 10.1016/j.asoc.2022.108491.
- [20] E. N. Iqbal and R. Asif, "Analyzing Applicant's Pre-Admission data and Predicting Applicant's Performance in Pre-Admission Test using Data Mining Techniques," in *3rd International Conference on Innovations in Computer Science and Software Engineering, ICONICS 2022*, 2022, doi: 10.1109/ICONICS56716.2022.10100546.
- [21] A. I. Al-Alawi, M. A. A. Alfateh, and A. M. Alrayes, "Educational Data Mining Utilization to Support the Admission Process in Higher Education Institutions: A Systematic Literature Review," in *2023 International Conference on Cyber Management and Engineering, CyMaEn 2023*, 2023, pp. 332–339, doi: 10.1109/CyMaEn57228.2023.10051077.
- [22] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020, doi: 10.1109/ACCESS.2020.2981905.
- [23] D. Kurniadi, F. Nuraeni, N. Abania, L. Fitriani, A. Mulyani, and Y. H. Agustin, "Scholarship Recipients Prediction Model using k-Nearest Neighbor Algorithm and Synthetic Minority Over-sampling Technique," in *12th International Conference on System Engineering and Technology, ICSET 2022 - Proceeding*, 2022, pp. 89–94, doi: 10.1109/ICSET57543.2022.10010947.
- [24] A. Abdelhadi, S. Zainudin, and N. S. Sani, "A Regression Model to Predict Key Performance Indicators in Higher Education Enrollments," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 1, pp. 454–460, 2022, doi: 10.14569/IJACSA.2022.0130156.
- [25] P. Chakrapani and D. Chitradevi, "Simulation of Machine Learning Techniques to Predict Academic Performance," in *Proceedings of the 2022 International Conference on Electronic Systems and Intelligent Computing, ICESIC 2022*, 2022, pp. 329–334, doi: 10.1109/ICESIC53714.2022.9783487.
- [26] A. M. Olalekan, O. S. Egwuche, and S. O. Olatunji, "Performance Evaluation of Machine Learning Techniques for Prediction of Graduating Students in Tertiary Institution," in *2020 International Conference in Mathematics, Computer Engineering and Computer Science, ICMCECS 2020*, 2020, doi: 10.1109/ICMCECS47690.2020.240888.
- [27] J. K. Alanzi and A. I. Al-Alawi, "Review of Data Mining Techniques in Performance Prediction for Medical Schools," in *2023 International Conference on Cyber Management and Engineering, CyMaEn 2023*, 2023, pp. 266–270, doi: 10.1109/CyMaEn57228.2023.10050935.
- [28] R. Nand, A. Chand, and E. Reddy, "Data Mining Students' performance in a Higher Learning Environment," in *NILES 2021 - 3rd Novel Intelligent and Leading Emerging Sciences Conference, Proceedings*, 2021, pp. 241–245, doi: 10.1109/NILES53778.2021.9600504.
- [29] R. Septiana, F. Renaldi, and I. Santikarama, "Prediction of Study Period and Information on Judiciary in Higher Education Using the C5.0 Algorithm," in *2022 International Conference on Science and Technology, ICOSTECH 2022*, 2022, doi: 10.1109/ICOSTECH54296.2022.9829103.

- [30] D. Malekian, J. Bailey, and G. Kennedy, "Prediction of students' assessment readiness in online learning environments: The sequence matters," in *ACM International Conference Proceeding Series*, 2020, pp. 382–391, doi: 10.1145/3375462.3375468.
- [31] A. C. Tolentino and T. D. Palaoag, "Predicting the On-The-Job Training Performance of IT Students on Virtual Internship Using Naïve Bayes," in *6th International Conference on Information Technology, InCIT 2022*, 2022, pp. 289–292, doi: 10.1109/InCIT56086.2022.10067296.
- [32] W. M. Attiya and M. B. Shams, "Predicting Student Retention in Higher Education Using Data Mining Techniques: A Literature Review," in *2023 International Conference on Cyber Management and Engineering, CyMaEn 2023*, 2023, pp. 171–177, doi: 10.1109/CyMaEn57228.2023.10051056.
- [33] S. Jayaprakash, S. Krishnan, and J. Jaiganesh, "Predicting Students Academic Performance using an Improved Random Forest Classifier," in *2020 International Conference on Emerging Smart Computing and Informatics, ESCI 2020*, 2020, pp. 238–243, doi: 10.1109/ESCI48226.2020.9167547.
- [34] W. Tenpipat and K. Akkarajitsakul, "Student Dropout Prediction: A KMUTT Case Study," in *2020 1st International Conference on Big Data Analytics and Practices, IBDAP 2020*, 2020, doi: 10.1109/IBDAP50342.2020.9245457.
- [35] Y. Zayed, Y. Salman, and A. Hasasneh, "A Recommendation System for Selecting the Appropriate Undergraduate Program at Higher Education Institutions Using Graduate Student Data," *Appl. Sci.*, vol. 12, no. 24, 2022, doi: 10.3390/app122412525.
- [36] R. K. Kavitha, C. Rajan Krupa, and J. I. Menezes, "Determining the Factors Influencing the Academic Accomplishment of Students and Predicting their Success Using Machine Learning Techniques," in *2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2023*, 2023, doi: 10.1109/ICAECA56562.2023.10200741.
- [37] M. A. A. Walid, S. M. M. Ahmed, M. Zeyad, S. M. S. Galib, and M. Nesa, "Analysis of machine learning strategies for prediction of passing undergraduate admission test," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100111, 2022, doi: <https://doi.org/10.1016/j.jjime.2022.100111>.
- [38] L. S. Rodrigues, M. dos Santos, I. Costa, and M. A. L. Moreira, "Student Performance Prediction on Primary and Secondary Schools-A Systematic Literature Review," *Procedia Comput. Sci.*, vol. 214, pp. 680–687, 2022, doi: <https://doi.org/10.1016/j.procs.2022.11.229>.
- [39] B. Nita *et al.*, "Machine learning in the enrolment management process: a case study of using GANs in postgraduate students' structure prediction," *Procedia Comput. Sci.*, vol. 207, pp. 1350–1359, 2022, doi: <https://doi.org/10.1016/j.procs.2022.09.191>.
- [40] T. T. Mai, M. Bezbradica, and M. Crane, "Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data," *Futur. Gener. Comput. Syst.*, vol. 127, pp. 42–55, 2022, doi: 10.1016/j.future.2021.08.026.
- [41] A. Çetinkaya and Ö. K. Baykan, "Prediction of middle school students' programming talent using artificial neural networks," *Eng. Sci. Technol. an Int. J.*, vol. 23, no. 6, pp. 1301–1307, 2020, doi: <https://doi.org/10.1016/j.jjestch.2020.07.005>.
- [42] M. S. Sassirekha and S. Vijayalakshmi, "Predicting the academic progression in student's standpoint using machine learning," *Automatika*, vol. 63, no. 4, pp. 605–617, Dec. 2022, doi: 10.1080/00051144.2022.2060652.
- [43] R. Bertolini, S. J. Finch, and R. H. Nehm, "Enhancing data pipelines for forecasting student performance: integrating feature selection with cross-validation," *Int. J. Educ. Technol. High. Educ.*, vol. 18, no. 1, 2021, doi: 10.1186/s41239-021-00279-6 WE - Social Science Citation Index (SSCI).