

BAB II

TINJAUAN PUSTAKA

2.1. Pembelajaran Mesin

Pembelajaran mesin atau *Machine learning* adalah cabang kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan sistem dan algoritma menggunakan teknik statistika yang memungkinkan komputer belajar dari data dan pengalaman masa lalu (Purnama and Sofana, 2021). Tujuan utama pembelajaran mesin adalah untuk mengembangkan model yang secara otomatis dapat meningkatkan kinerjanya ketika dihadapkan pada informasi baru, tanpa memerlukan pengkodean manual yang ekstensif.

Dalam pembelajaran mesin, algoritma dilengkapi dengan data latih yang mencakup contoh masalah dan solusinya berbentuk data masukan dan luaran (*input-output*). Algoritma ini menggunakan data untuk mengidentifikasi pola, hubungan, dan karakteristik penting dalam data. Setelah proses pelatihan selesai, model yang dihasilkan dapat digunakan untuk membuat prediksi atau keputusan berdasarkan data baru yang belum pernah dipelajari sebelumnya.

Terdapat beberapa karakteristik permasalahan model machine learning secara umum, yaitu (Cunningham, Cord and Delany, 2018.):

a. Supervised Learning

Supervised learning merupakan metode yang berfungsi memetakan data *input* dan *output* berdasarkan contoh yang tersedia pada kumpulan data atau biasa disebut dataset. Dataset terdiri dari data latih dan data uji, dimana data latih memiliki variabel keluaran yang perlu diprediksi kedalam bentuk klasifikasi. Pada prosesnya, metode ini melakukan proses prediksi dengan mempelajari pola dari dataset latih dan menerapkannya ke setiap dataset uji.

b. Unsupervised Learning

Unsupervised learning merupakan metode yang memiliki fungsi untuk melakukan pengelompokan data, tanpa adanya label awal dari dataset. Pada

prosesnya, metode ini melakukan eksplorasi dan menemukan struktur dalam data untuk mengenali kelas data.

c. *Semi-supervised Learning*

Semi-supervised learning merupakan kombinasi dari metode *supervised learning* dengan *unsupervised learning*. Data yang memiliki label dan tanpa label akan digunakan untuk melatih algoritma hingga dapat melakukan proses pembelajaran sendiri. Metode ini digunakan untuk data yang hanya memiliki sedikit label dibandingkan dengan jumlah data tanpa label.

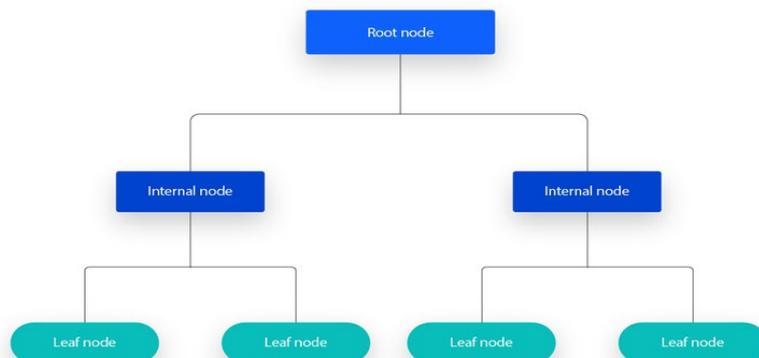
d. *Reinforcement Learning*

Reinforcement learning merupakan metode yang memiliki karakteristik pengambilan keputusan secara berurutan untuk memperoleh hasil yang terbaik. Pada prosesnya, metode ini mempelajari pola atau perilaku baru yang terus dimodifikasi secara mandiri secara bertahap untuk mendapatkan *reward* berdasarkan proses pelatihan *trial* dan *error*.

2.2. Algoritma Machine Learning – Supervised Learning Classification

2.2.1. Decision Tree (DT)

Decision tree (DT) merupakan algoritma berupa alat pendukung keputusan yang diwakili oleh grafik atau model keputusan seperti pohon yang mewakili hasil yang mungkin terjadi (Nowozin et al., 2011). Algoritma ini merupakan suatu pendekatan, yang berisi pernyataan kondisional dalam suatu algoritma. DT membuat sebuah pohon keputusan menggunakan kriteria pemisahan seperti Gini Impurity atau Entropi untuk menentukan pemisahan data yang akan dibentuk pada proses pelatihan. Setelah pohon keputusan terbentuk, DT melakukan prediksi pada data baru dengan melewati pohon dari akar ke daun berdasarkan nilai setiap fitur pada data yang digunakan. Ilustrasi dari proses pada algoritma DT dapat dilihat pada Gambar 2.1 sebagai berikut.



Gambar 2.1 Ilustrasi DT

Dalam bentuk matematis, prediksi hasil (Y) dapat direpresentasikan sebagai fungsi berulang yang melibatkan kondisi (Q) pada setiap simpul n pada pohon keputusan. Berikut perhitungan matematis dari prediksi yang dilakukan DT pada Formula (2.1).

$$Y(X) = \sum_{n=1}^N Q_n(X) \cdot c_n \quad (2.1)$$

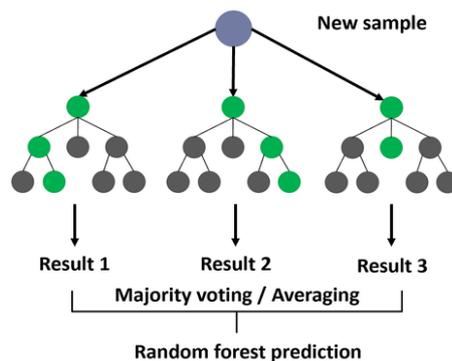
N : jumlah simpul dalam pohon

$Q_n(X)$: fungsi yang menghasilkan 1 jika sampel X memenuhi kondisi pada simpul n , dan 0 jika sebaliknya

c_n : nilai prediksi pada simpul n .

2.2.2. *Random Forest (RF)*

Random forest merupakan algoritma berupa ensemble learning atau algoritma gabungan untuk regresi, klasifikasi dan untuk tugas-tugas lainnya. Algoritma ini beroperasi dengan membuat banyak DT pada tugas pelatihan dan menghasilkan keluaran yang merupakan prediksi rata-rata (regresi) atau mode kelas (klasifikasi) dari pohon yang berbeda (Rigatti, 2017). Kelompok pohon keputusan hutan keputusan acak (dikenal sebagai “hutan”) melakukan praktik overfitting pada set pelatihannya. Setiap pohon di hutan berkontribusi pada suatu klasifikasi untuk mengklasifikasikan kasus baru berdasarkan atributnya. Ilustrasi dari proses yang terjadi pada RF ditunjukkan pada gambar 2.2.



Gambar 2.2. Ilustrasi RF

Model matematis dari RF dapat dilihat pada formula (2.2) sebagai berikut.

$$Y_{RF}(X) = \underset{c}{\operatorname{argmax}} \left(\sum_{n=1}^N 1(T_n(X) = c) \right) \quad (2.2)$$

N : jumlah pohon dalam RF

X : data input yang akan diprediksi

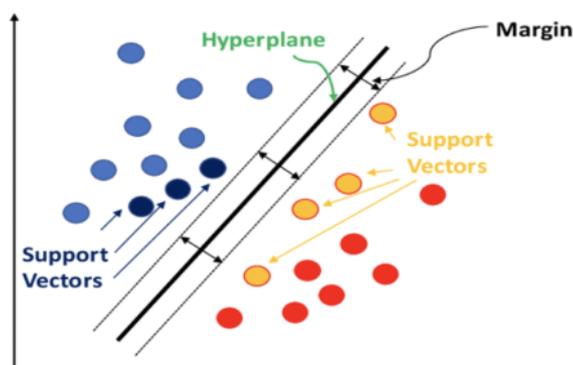
$T_i(X)$: prediksi dari pohon keputusan ke- i

c : kelas prediksi

$1(\cdot)$: fungsi indikator yang menghasilkan 1 jika pernyataan benar, dan 0 sebaliknya.

2.2.3. Support Vector Machine (SVM)

SVM merupakan sebuah algoritma klasifikasi dan regresi. Dalam SVM, setiap item data diplot dalam ruang berdimensi n , sejumlah dimensi setara dengan jumlah fitur atau atribut (Pisner and Schnyer, 2020). Dimana n mewakili jumlah atribut. Nilai setiap atribut merupakan nilai koordinat tertentu. Setelah seluruh item data diplot kemudian dilakukan klasifikasi dengan menggambar garis atau dengan mencari hyperplane optimal yang memisahkan dua kelas secara lengkap. Misalnya, jika kita memiliki dua ciri individu seperti Rambut dan Panjang Tinggi. Pertama, kita memplot kedua fitur ini dalam ruang dua dimensi. Dimana setiap titik mempunyai dua koordinat (koordinat ini disebut juga dengan Vektor Pendukung). Ilustrasi dari proses SVM dapat dilihat pada Gambar 2.3.



Gambar 2.3. Ilustrasi SVM

Fungsi matematis SVM ditunjukkan pada formula (2.3) berikut.

$$f(x) = w \cdot x + b \quad (2.3)$$

w : vektor bobot

x : vektor input

b : bias

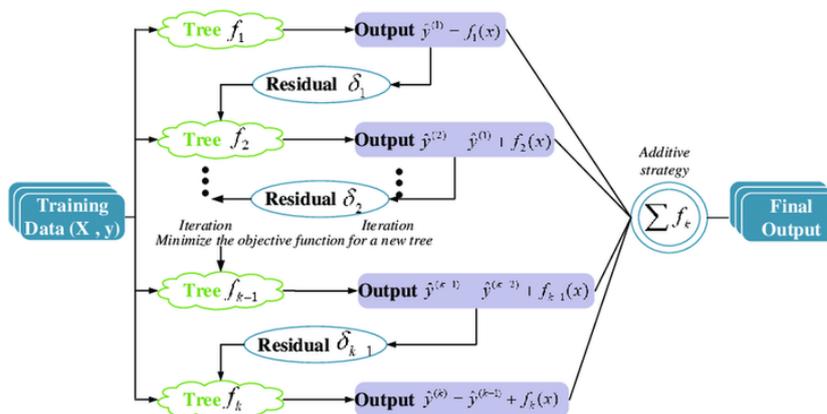
$w \cdot x + b$: dot produk antara w dan x .

Jika $f(x) > 0$ maka model memprediksi kelas 1

Jika $f(x) < 0$ maka model memprediksi kelas -1

2.2.4. *Extreme Gradient Boosting (XGBoost)*

XGBoost (Extreme Gradient Boosting) atau XGB adalah algoritma ensemble machine learning yang sangat populer dan kuat. Ini adalah pengembangan dari algoritma Gradient Boosting yang telah ada sebelumnya. Ini adalah implementasi yang sangat efisien dan efektif dari teknik ensemble boosting yang bertujuan untuk meningkatkan kinerja model prediksi, bekerja dengan cara berulang kali menambahkan pohon baru ke model sejauh pohon tersebut memperbaiki hasil prediksi secara berurutan (Velthoen et al., 2023). Setiap pohon yang ditambahkan mencoba untuk memperbaiki kesalahan prediksi model sebelumnya. Hal ini dilakukan dengan menghitung gradien (gradient) dari loss function terhadap prediksi model sebelumnya. Ilustrasi dari proses XGB dapat ditinjau pada gambar 2.4 sebagai berikut.



Gambar 2.4. Ilustrasi XGB

Fungsi matematis dari XGBoost dapat dilihat pada formula (2.4) berikut.

$$Objective = \sum_{i=1}^n Binary_{CrossEntropy}(y_i, \hat{y}_i) + \sum_{j=1}^T \Omega(f_j) \tag{2.4}$$

\hat{y}_i : prediksi model untuk sampel ke-i

$Binary_{CrossEntropy}(y_i, \hat{y}_i)$: fungsi untuk klasifikasi biner

$\Omega(f_j)$: fungsi regularisasi untuk pohon keputusan ke-i

T : jumlah pohon.

2.3. Parameter Evaluasi

Evaluasi yang dilakukan pada pengujian model menggunakan data uji dengan metode *confusion matrix* serta menghitung nilai akurasi, presisi, *recall*, dan *f1-score*. Ilustrasi dari *confusion matrix* dapat dilihat pada Gambar 2.5.

		Prediksi	
		Positive	Negative
Aktual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Gambar 2.5. Confusion Matrix

Persamaan TP (true positive) merupakan kelas suatu objek positif yang terklasifikasi benar, TN (true negative) merupakan kelas objek negatif yang terklasifikasi benar, FP (false positive) merupakan kelas suatu objek positif yang

tidak terklasifikasi benar, dan FN (false negative) merupakan kelas objek negatif yang tidak terklasifikasi benar. Perhitungan *confusion matrix* menghasilkan 2 nilai yang berbeda. Pertama adalah *Macro-Average* yang merupakan perhitungan rata-rata nilai evaluasi dari seluruh kelas. Kemudian yang kedua adalah *Weighted-Average* yang merupakan perhitungan rata rata total nilai TP dari seluruh kelas.

2.3.1. Accuracy

Accuracy atau akurasi pada proses evaluasi model merupakan nilai performa pada model berdasarkan tingkat kedekatan antara nilai prediksi dengan nilai aktual. Akurasi dapat diartikan sebagai tingkat ketepatan model dalam melakukan proses klasifikasi yang benar. Nilai akurasi dapat dihitung menggunakan formula (2.5), sebagai berikut:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2.5)$$

2.3.2. Precision

Precision atau presisi merupakan rasio jumlah informasi relevan yang dipilih oleh sistem dengan jumlah seluruh informasi yang terprediksi. Kecocokan antara informasi yang diminta oleh sistem dengan hasil prediksi yang diberikan oleh model dapat diartikan sebagai presisi. Nilai presisi dapat dihitung menggunakan menggunakan rumus yang ditunjukkan pada formula (2.6), sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \quad (2.6)$$

2.3.3. Recall

Recall merupakan rasio jumlah informasi relevan yang dipilih oleh sistem dengan jumlah seluruh informasi relevan pada dataset. Recall dapat digambarkan sebagai keberhasilan model dalam menemukan sebuah informasi yang relevan. Recall dapat dihitung menggunakan formula (2.7), sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \quad (2.7)$$

2.3.4. *F1-Score*

F1-Score merupakan metrik yang digunakan untuk mengukur performa model dengan melakukan perbandingan rata-rata nilai presisi dan recall. Perhitungan *F1-Score* dapat dilakukan menggunakan formula (2.8) sebagai berikut:

$$F1 - Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \quad (2.8)$$

2.4. Tools

2.4.1. Python

Python adalah bahasa pemrograman tingkat tinggi yang serbaguna dan sangat populer, dibuat untuk membantu pengembangan perangkat lunak dengan sintaksis yang jelas dan mudah dibaca. Dikenal karena dapat ditafsirkan, Python menawarkan kemudahan eksperimen dan eksperimen, dengan banyak komunitas pengembang yang mendukungnya. Fleksibilitasnya memungkinkannya digunakan di berbagai bidang seperti pengembangan web, analisis data, kecerdasan buatan, dan banyak lainnya. Python juga bersifat *open source*, memungkinkan pengguna untuk mengakses dan memodifikasinya sesuai kebutuhan mereka, menjadikannya alat yang ampuh yang tersedia bagi pengembang baru dan berpengalaman di mana pun.

2.4.2. Scikit-Learn (SkLearn)

Scikit-learn adalah salah satu perpustakaan bahasa pemrograman Python paling populer yang digunakan untuk pembelajaran mesin. SkLearn adalah perpustakaan sumber terbuka yang menyediakan alat dan algoritma canggih untuk tugas-tugas seperti klasifikasi, regresi, pengelompokan, pemrosesan data, dan evaluasi model (Pedregosa et al., 2011).

Scikit-learn menyediakan banyak algoritma pembelajaran mesin yang siap digunakan, termasuk Support Vector Machines (SVM), Decision Trees, Random Forests, Naive Bayes, dan banyak lagi. Selain itu, ia memiliki fungsi utilitas untuk pra-pemrosesan data, pemilihan fitur, dan evaluasi model, memungkinkan pengguna mengembangkan dan menguji model pembelajaran mesin dengan lebih mudah dan efisien.

Scikit-learn populer di kalangan praktisi pembelajaran mesin dan ilmuwan data karena dokumentasinya yang kaya, komunitas aktif, dan integrasi yang baik dengan alat lain di ekosistem Python, seperti NumPy, Pandas, dan Matplotlib.

2.4.3. Google Colaboratory

Google Colaboratory atau Google Colab biasa juga disebut dengan Colab, merupakan platform pengembangan proyek pembelajaran mesin (Bisong, 2019). Platform ini menyediakan lingkungan notebook Jupyter tanpa server untuk pengembangan interaktif berbasis awan. Untuk menjalankan fungsinya, Colab ditenagai oleh perangkat keras berperforma tinggi diantaranya Unit Pemrosesan Grafik (GPU) dan Unit Pemrosesan Tensor (TPU). Terdapat beberapa opsi yang dapat digunakan sesuai dengan paket pembelian diantaranya CPU, T4 GPU, A100 GPU, V100 GPU, dan TPU. Pada penelitian ini opsi yang digunakan adalah GPU T4.

2.5. Penelitian Terkait (State-Of-The-Art)

Salah satu tahap dalam metode penelitian yang dilakukan adalah studi literatur. Tahap ini dilakukan dengan mengumpulkan penelitian terdahulu dan pendalaman materi yang berkaitan dengan model pembelajaran mesin klasifikasi serta prediksi performa akademik yang kemudian akan menjadi acuan dalam melakukan penelitian ini. Berikut merupakan referensi penelitian terkait yang dapat ditinjau pada Tabel 2.1.

Tabel 2.1. *State of the Art* Penelitian

No	Judul Penelitian	Peneliti	Metode	Fokus Penelitian	Hasil
1	<i>A Hybrid Machine Learning Model for Grade Prediction in Online Engineering Education</i>	Zoe Kanetaki, Constantinos Stergiou, Georgios Bekas, Christos Troussas, Cleo Sgouropoulou (2022)	<i>Hybrid Model (Generalized Linear Model dan Neural Network)</i>	Membuat model untuk prediksi nilai pada pendidikan teknik secara online	Didapat model yang dapat memprediksi nilai siswa dengan metode Hybrid Model. Peneliti juga menemukan terdapat banyak sudut pandang literatur yang dapat menjadi aspek untuk menjadi parameter dalam prediksi nilai.
2	<i>Exploratory Data Analysis To Identify The Most Important Feature Of</i>	Acep Irham Gufroni, Purwanto, Farikhin, Adi	<i>Exploratory Data Analysis</i>	Analisa Feature Importance pada	Didapat persentase masing masing atribut dari data admisi sebagai tingkat

	<i>University Admission Test Criteria Using Random Forest And Neural Network Algorithm</i>	Wibowo, Budi Wasito (2021)		data admisi mahasiswa	kontribusi terhadap implementasi model sebagai percobaan.
3	<i>Multiclass Prediction Model for Student Grade Prediction Using Machine Learning</i>	Siti Dianah Abdul Bujang, Ali Selamat, Roliana Ibrahim, Ondrej Krejcar, Enrique Herrera-Viedma, Hamido Fujita, Dan Nor Azura Md. Ghani (2021)	Decision Tree (J48), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbor (kNN), Logistic Regression (LR) dan Random Forest (RF)	Membuat model dengan akurasi yang lebih baik dari penelitian sebelumnya dengan membandingkan 6 metode machine learning untuk dapat memprediksi nilai mahasiswa	Diketahui bahwa model yang menggunakan algoritma Random Forest dapat memberikan perkembangan yang signifikan dengan hasil tertinggi menggunakan pengukuran f-measure sebesar 99.5%
4	<i>Prediction of Student Performance Using Linear Regression</i>	Boddeti Sravani, Myneni Madhu Bala (2020)	Regresi Linear	Membuat model machine learning yang dapat	Didapat model yang dapat memprediksi performa mahasiswa. Selain itu juga

				memprediksi performa mahasiswa	diketahui bahwa implementasi machine learning untuk kasus tersebut dipengaruhi atribut latar belakang mahasiswa
5	Perbandingan Metode Data Mining Untuk Prediksi Nilai dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika dengan Algoritma C4.5, Naïve Bayes, Knn, Dan Svm	Sri Widaningsih (2019)	C4.5, NAÏVE BAYES, KNN, dan SVM	Membandingkan model data mining untuk prediksi nilai dan waktu kelulusan mahasiswa	Didapat hasil perbandingan dari keempat model yang diuji, dengan hasil terbaik menggunakan algoritma <i>Naïve Bayes</i> dengan akurasi 76,79%, error 23,17%, dan AUC 0,85
6	Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa	Irkham Widhi Saputro, Bety Wulan Sari (2019)	10-Fold Cross Validation, dan Confusion Matrix	Menguji performa model	Didapat hasil pengujian berupa nilai akurasi sebesar 68%, nilai precision sebesar 61,3%, nilai recall sebesar 65,3%, dan nilai fi-score sebesar 61%

7	Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika	Ratna Puspita Sari Putri, Indra Waspada (2018)	C4.5	Membuat model untuk prediksi kelulusan mahasiswa	Algoritma C4.5 dapat digunakan untuk prediksi kelulusan mahasiswa. Hasil yang didapat ditingkatkan kembali dengan memotong decision tree menggunakan metode error-based pruning dengan nilai confidence 0,4, sehingga didapat hasil precision sebesar 70,7%, recall 50,65%, dan akurasi 61,57%.
8	<i>Data Mining Approach To Predicting the Performance of First Year Student in a University Using the Admission Requirements</i>	Aderibigbe Israel Adekitan, Etinosa Noma-Osaghae (2018)	Analisa Regresi dan Artificial Neural network	Membuat analisa terhadap model untuk memprediksi performa tahun pertama mahasiswa	Diuji 6 algoritma data mining dan didapat hasil tertinggi dengan akurasi 50,23% dan 51,9%

9	<i>Study on Student Performance Estimation, Student Progress Analysis, and Student Potential Prediction Based On Data Mining</i>	Fan Yanga, Frederick W.B. Li (2018)	Student Attribute Matrix (SAM), dan Back Propagation Neural Network (BP-NN)	Analisa implementasi data mining yang digunakan untuk memprediksi performa mahasiswa, analisa progres, dan prediksi potensi	Didapat model yang dapat memprediksi performa mahasiswa, yang dapat dijadikan sistem rekomendasi untuk meningkatkan kualitas akademik mahasiswa
10	Prediksi Nilai Mata Kuliah Mahasiswa Menggunakan Algoritma K-Apriori	Lailil Muflikhah , W. Lisa Yunita, M. Tanzil Furqon (2017)	Wiener Transformation, K-Means, dan Apriori	Menghasilkan model yang dapat memprediksi nilai mata kuliah mahasiswa	Didapat model yang dapat memprediksi nilai mata kuliah mahasiswa dengan rata-rata error rate sebesar 11%
11	<i>Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm</i>	Hasmia Hamsa, Simi Indiradevi, Jubilant J. Kizhakkethottam (2016)	Decision Tree dan Fuzzy Genetic	Membuat model prediksi performa akademis mahasiswa Sarjana dan Pasca-sarjana	Didapat model yang dapat memprediksi performa mahasiswa dengan mengidentifikasinya menjadi aman dan beresiko. Hasil

				terhadap suatu matakuliah	dari algoritma decision tree memberikan hasil identifikasi beresiko lebih banyak dibandingkan algoritma Fuzzy Genetic
12	Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ)	Annisa Paramitha Fadillah (2015)	CRISP-DM, C4.5	Membuat model untuk prediksi pola dari kelulusan mahasiswa dalam pengambilan matakuliah	Diketahui bahwa metodologi CRISP-DM dapat melakukan prediksi terhadap kelulusan mahasiswa pada pengambilan mata kuliah. Simpulan tersebut diambil setelah memanfaatkan algoritma C4.5 yang dapat melakukan klasifikasi lulus dan tidak lulus.
13	Prediksi Lama Studi Mahasiswa dengan Metode Random Forest	I Made Budi Adnyana (2015)	Random Forest	Membuat model untuk prediksi lama studi mahasiswa	Didapat model klasifikasi dengan 2 kelas yakni lulus tepat waktu dan lulus batas

					waktu, dengan akurasi 83,54%
14	Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining	Paramita Mayadewi, Ely Rosely (2015)	ID3, CHAID, dan Naïve Bayes	Membuat rules yang dapat memprediksi nilai proyek akhir mahasiswa diploma	Model prediksi menggunakan algoritma ID3 memiliki akurasi 62,66%, CHAID 63,66%, dan Naïve Bayes 65,67%
15	<i>Mining Students' Data for Performance Prediction</i>	Tripti Mishra, Dr. Dharminder Kumar, Dr. Sangeeta Gupta (2014)	J48 (Implementasi dari C4.5), dan Random Tree	Membuat model prediksi performa akademis mahasiswa	Diketahui bahwa algoritma Random Tree memiliki hasil akurasi 94,418% yang lebih tinggi dibandingkan J48 dengan akurasi 88,372%
16	Perbandingan Akurasi k-NN dan Naive Bayes untuk Algoritma Sistem Prediksi Nilai Akhir Mahasiswa	Muhammad Ihsan Zul (2013)	k-NN dan Naive Bayes	Membandingkan akurasi model dalam memprediksi nilai akhir mahasiswa	Diketahui algoritma Naïve Bayes menghasilkan hasil lebih baik dibandingkan Naïve Bayes dengan nilai akurasi sebesar 95%

2.6. Keterbaruan Penelitian

Matriks perbedaan penelitian terdahulu dengan penelitian yang dilakukan, yang dapat dilihat pada tabel 2.2 sebagai berikut.

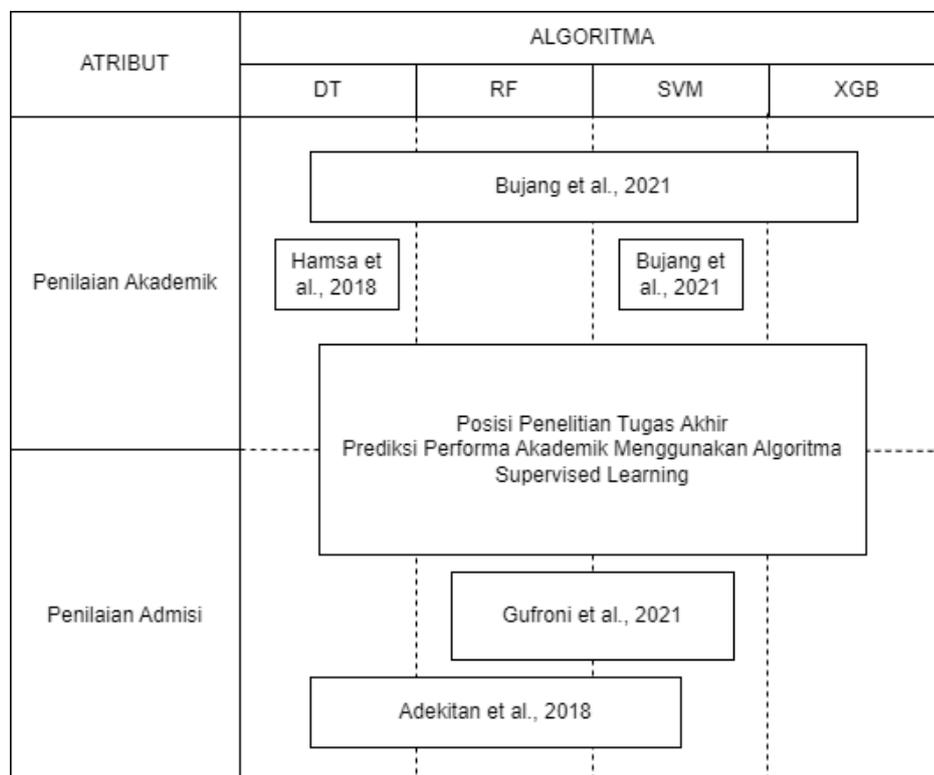
Tabel 2.2. Perbandingan Penelitian

No	Peneliti	Perbandingan penelitian
1	Zoe Kanetaki, Constantinos Stergiou, Georgios Bekas, Christos Troussas, Cleo Sgouropoulou (2022)	Terdahulu : Memprediksi dari beberapa mata pelajaran Baru: Memprediksi performa akademik secara menyeluruh lebih dini
2	Acep Irham Gufroni, Purwanto, Farikhin, Adi Wibowo, Budi Wasito (2021)	Terdahulu : Exploratory Data Analysis pada data admisi mahasiswa Baru: Pembuatan pembelajaran mesin untuk prediksi performa akademik menggunakan data admisi mahasiswa
3	Siti Dianah Abdul Bujang, Ali Selamat, Roliana Ibrahim, Ondrej Krejcar, Enrique Herrera-Viedma, Hamido Fujita, Dan Nor Azura Md. Ghani (2021)	Terdahulu : Memprediksi dari beberapa mata Pelajaran Baru: Memprediksi performa akademik secara menyeluruh lebih dini
4	Boddeti Sravani, Myneni Madhu Bala (2020)	Terdahulu : Prediksi performa akademik mahasiswa menggunakan Regresi Linear Baru: Prediksi performa akademik menggunakan 4 Algoritma (SVM,
5	Sri Widaningsih (2019)	Terdahulu : prediksi nilai dan waktu kelulusan Baru: Prediksi performa akademik
6	Irkham Widhi Saputro, Bety Wulan Sari (2019)	Terdahulu : Prediksi masa studi Baru: Prediksi performa akademik
7	Ratna Puspita Sari Putri, Indra Waspada (2018)	Terdahulu : Prediksi kelulusan mahasiswa Baru: Prediksi performa akademik mahasiswa

8	Aderibigbe Israel Adekitan, Etinosa Noma-Osaghae (2018)	Terdahulu : Prediksi performa akademik dengan data pre admisi menggunakan Regresi dan ANN Baru: Prediksi performa akademik dengan data pre admisi menggunakan DT,RF, XGB, SVM
9	Fan Yanga, Frederick W.B. Li (2018)	Terdahulu : Prediksi performa mahasiswa Baru: Prediksi performa mahasiswa dengan data yang lebih dini
10	Lailil Muflikhah , W. Lisa Yunita, M. Tanzil Furqon (2017)	Terdahulu : Prediksi nilai mata kuliah Baru: Prediksi performa akademik
11	Hasmia Hamsa, Simi Indiradevi, Jubilant J. Kizhakkethottam (2016)	Terdahulu : Prediksi performa akademik menggunakan DT dan Fuzzy Genetic Baru: Prediksi performa akademik menggunakan DT,RF, XGB, SVM
12	Annisa Paramitha Fadillah (2015)	Terdahulu: Implementasi crisp-dm Prediksi kelulusan pada matakuliah Baru: Prediksi performa akademik
13	I Made Budi Adnyana (2015)	Terdahulu : Prediksi lama studi menggunakan RF Baru: Prediksi performa akademik salah satunya menggunakan RF
14	Paramita Mayadewi, Ely Rosely (2015)	Terdahulu : Prediksi nilai proyek akhir mahasiswa Baru: Prediksi performa akademik dengan data admisi
15	Tripti Mishra, Dr. Dharminder Kumar, Dr. Sangeeta Gupta (2014)	Terdahulu : Prediksi performa akademik berdasarkan nilai sosial, akademik, dan uji emosi. Baru: Prediksi performa akademik menggunakan data admisi

16	Muhammad Ihsan Zul (2013)	<p>Terdahulu : Prediksi nilai akhir mata pelajaran untuk prediksi pencapaian masa depan anak</p> <p>Baru: Prediksi performa akademik mahasiswa sejak dini</p>
----	---------------------------	---

Berdasarkan matriks diatas, proses prediksi performa akademik mahasiswa masih belum maksimal dengan indikasi penggunaan jenis data yang kurang menyeluruh, model yang menghasilkan keluaran kurang relevan dengan penerapan dunia nyata, serta akurasi model yang belum maksimal. Oleh karena itu, pada penelitian ini akan membangun model pembelajaran mesin yang dapat memprediksi performa akademik mahasiswa. Untuk posisi penelitian ini, diilustrasikan pada gambar 2.6 sebagai berikut.



Gambar 2.6. Posisi Penelitian