

BAB II

LANDASAN TEORI

2.1 Data

Data adalah sekumpulan informasi atau juga keterangan – keterangan dari suatu hal yang diperoleh dengan melalui pengamatan atau juga pencarian ke sumber-sumber tertentu. Data yang diperoleh tersebut dapat menjadi sebuah anggapan atau fakta disebabkan karena belum diolah dengan lebih lanjut. Menurut Nuzulla Agustina (2018) , data adalah keterangan mengenai sesuatu hal yang sudah sering terjadi dan berupa berupa himpunan fakta, angka, grafik, tabel, gambar, lambang, kata, huruf-huruf yang menyatakan sesuatu pemikiran, objek, serta kondisi dan situasi.

2.2 Data Mining

Menurut Dr.Suyanto (2019:2), *data mining* merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar,meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistic*, dan *database system*.

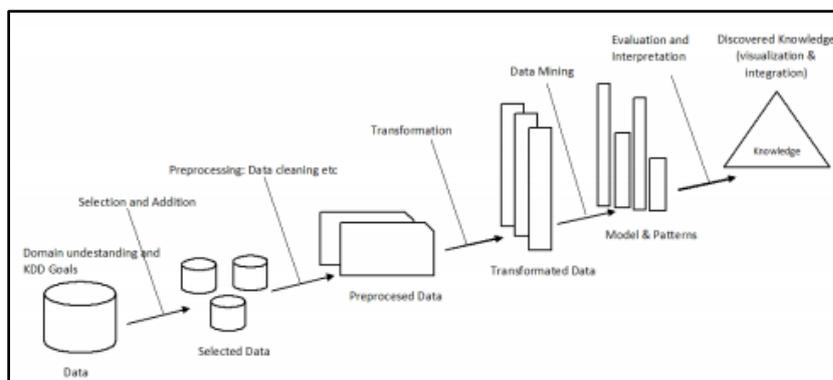
Data mining adalah proses proses yang menggunakan statistik, matematika, kecerdasan buatan dan teknik pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi yang berguna dan pengetahuan selanjutnya dari basis data besar (Wyatt & Taylor, 2008).

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai *database* besar (Turban et al., n.d.).

Berdasarkan definisi-definisi diatas tentang *Data Mining* dapat disimpulkan bahwa *data mining* adalah sebuah proses pencarian secara otomatis untuk menemukan pola atau model dari suatu *database* yang besar. Sederhananya definisi dari data mining adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar.

2.2.1 *Knowledge Discovery in Database*

Knowledge Discovery in Database (KDD) adalah proses menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui dan potensial bermanfaat (Amalia, 2018). *Data Mining* merupakan salah satu langkah dari serangkaian proses *Knowledge Discovery in Database* (KDD). Berikut ini tahapan dari proses *Knowledge Discovery in Database* (KDD) dapat dilihat pada Gambar 2.1.



Gambar 2.1 Tahapan dalam KDD (Nikmatun & Waspada, 2019)

Tahapan proses *Knowledge Discovery in Database* (KDD) adalah sebagai berikut:

- 1 *Data Selection*. Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada subset *variable* (sampel data) dimana penemuan akan dilakukan.
- 2 *Pre-Processing* dan *data cleaning*. *Pre-Processing* dan *data cleaning* dilakukan membuang data yang tidak konsisten, data *missing value*, *noise*, duplikasi data, memperbaiki kesalahan data, dan bisa diperkaya dengan data *eksternal* yang relevan.
- 3 *Transformation*. Proses ini dilakukan untuk mengubah bentuk data yang dapat memudahkan dalam proses *data*. Dalam proses *data transformation* bisa dilakukan dengan *centering*, *normalization*, dan *scaling* (Sulastri & Gufroni, 2017).
- 4 *Data Mining*. Proses *data mining* digunakan untuk mencari pola atau informasi dalam data dengan menggunakan teknik, metode atau algoritma tertentu sesuai dengan tujuan dari proses *Knowledge Discovery in Database*.
- 5 *Interpretation* / Evaluasi. Penerjemahan pola-pola yang dihasilkan dari *data mining*. Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

- 6 Presentasi pengetahuan (*knowledge presentation*). Proses ini merupakan visualisasi dari penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna(Heryana, 2019).

2.2.2 Klasifikasi

Klasifikasi merupakan salah satu teknik yang ada dalam data mining. Klasifikasi adalah suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan dan menjelaskan atau membedakan konsep atau kelas data, yang diperlukan untuk memperkirakan kelas dari suatu objek(Defiyanti, 2013). Klasifikasi membangun model berdasarkan data pelatihan yang ada, kemudian menggunakan model tersebut untuk mengklasifikasikan data baru.

2.2.3 Prediksi

Prediksi merupakan proses memperkirakan secara sistematis apa yang paling mungkin terjadi di masa depan, berdasarkan pengetahuan masa lalu dan sekarang(Rohmawati et al., 2017). Menurut (Kafil, 2019) suatu prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi.

2.2.4 K- Nearest Neighbor

Algoritma *K-Nearest Neighbor* merupakan suatu metode untuk melakukan pengklasifikasian terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Menurut (Saputra & Primadasa, 2018) salah satu teknik klasifikasi data yang kuat, dengan cara mencari kasus dengan menghitung jarak atau kedekatan antara kasus baru dengan kasus lama. Jarak data ditentukan oleh nilai k atau jumlah tetangga terdekat(Ainurrohmah, 2021).

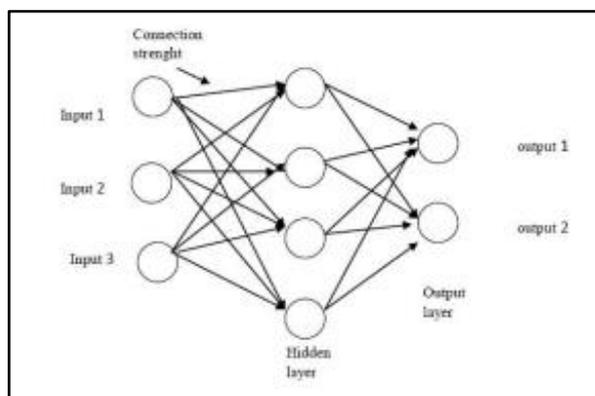
Banyak cara untuk mengukur jarak atau kedekatan antara data baru dan data lama yang salah satunya menggunakan persamaan *Euclidean distance*. *Euclidean distance* merupakan metode untuk mencari kedekatan dan kemiripan antara dua titik variabel, dimana semakin dekat dan mirip kedua titik tersebut maka semakin kecil jaraknya (Purwani et al., 2022). Rumus persamaan *euclidean distance* (Saputra & Primadasa, 2018) adalah sebagai berikut :

$$d_{Euclidean}(x, y) = \sqrt{\sum_i^n (x_{training} - y_{testing})^2} \quad (2.1)$$

2.2.5 Neural Network

Neural Network atau biasa disebut *artificial neural network* atau jaringan syaraf tiruan adalah algoritma yang meniru fungsi otak manusia yang didalamnya terdapat ribuan dan jutaan unit pengolahan kecil yang disebut neuron (R. Ridwan et al., 2020). Konsep *neural network* yaitu meniru sistem jaringan saraf yang ada pada tubuh manusia, dimana dibangun *node-node* yang saling terhubung melalui suatu *link* yang biasa disebut dengan istilah *weight*. Dari setiap neuron akan saling terhubung satu sama lain dan neuron tersebut akan memiliki input dan keluaran.

Menurut (Suleman & Palupi, 2023) *artificial neural network* merupakan sebuah model yang terinspirasi oleh sistem saraf otak manusia dalam melakukan klasifikasi data yang memiliki struktur jaringan yang terdiri dari layer *input* dan *output*. *Artificial neural network* memiliki 3 layer yaitu *input layer*, *hidden layer* dan *output layer*.



Gambar 2.2 Arsitektur *Neural Network* (Rohman & Rochcham, 2019)

Secara umum, *artificial neural network* terdapat 2 jenis yang sering digunakan yaitu *single layer perceptron* dan *multi layer perceptron*. *Artificial Neural Network* juga dibagi menjadi beberapa bagian yang lebih kecil, dimana masing–masing metode mempunyai karakteristiknya sendiri. Metode–metode tersebut diantaranya *Feedforward*, *Backpropagation*.

2.2.6 *Confusion Matrix*

Confusion matrix merupakan metode evaluasi model dari klasifikasi (Nikmatun & Waspada, 2019). *Confusion matrix* berbentuk sebuah tabel yang menunjukkan jumlah data yang diklasifikasikan dengan benar dan salah dalam masing-masing kategori atau kelas target. Tabel *confusion matrix* disajikan pada tabel 2.1 dibawah ini (Santoso & Kusumaningsih, 2018) :

Tabel 2.1 *Confusion Matrix*

	Prediction Class	
Actual Class	Positive	Negative
Positive	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Negative	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Berdasarkan Tabel 2.1 diatas :

1. *True Positive* (TP) adalah jumlah data yang bernilai positif dan diprediksi benar sebagai positif.
2. *False Positive* (FP) adalah jumlah data yang bernilai negatif tetapi diprediksi positif.
3. *True Negative* (TN) adalah jumlah data yang bernilai positif tetapi diprediksi sebagai negatif.
4. *False Negative* (FN) adalah jumlah data yang bernilai negatif dan diprediksi benar sebagai negatif.

Setelah diketahui hasil confusion matrix, data akan dihitung nilai accuracy, recall, precision dan f1-score dengan menggunakan persamaan sebagai berikut (Nikmatun & Waspada, 2019) :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2.2)$$

$$Precision = \frac{TP}{TP+FN} \times 100\% \quad (2.3)$$

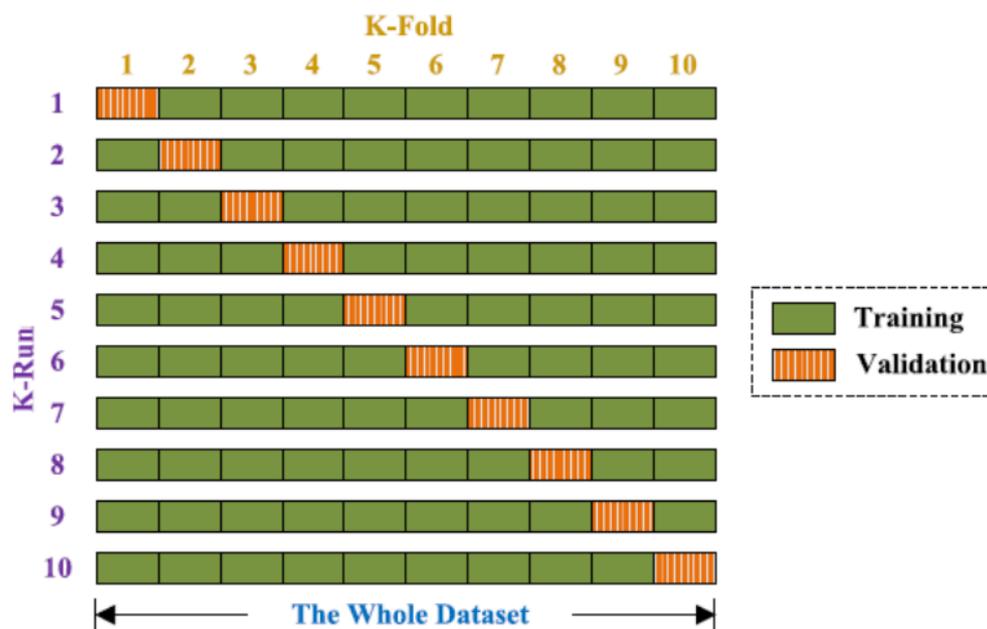
$$Recall = \frac{TP}{TP+FP} \times 100\% \quad (2.4)$$

$$F1-Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \times 100\% \quad (2.5)$$

2.2.7 K-Fold Cross Validation

K-Fold Cross Validation merupakan metode yang digunakan untuk mengukur keberhasilan model klasifikasi dengan membagi dataset secara acak menjadi k himpunan bagian (Nikmatun & Waspada, 2019). Setiap *fold* akan berfungsi sebagai data uji satu kali, sedangkan sisa *fold* digunakan sebagai data latih. Dalam setiap iterasi, model akan dilatih dengan menggunakan data latih dan

data uji menggunakan data uji yang berbeda setiap iterasi. Gambar 2.3 menunjukkan *K-Fold Cross Validation* dengan menggunakan nilai *k fold* 10.



Gambar 2.3 Diagram 10 *Fold Cross Validation*(Nikmatun & Waspada, 2019)

2.2.8 *Rapidminer*

Rapidminer adalah sebuah perangkat lunak yang bersifat *open source* yang digunakan untuk mengolah dan menganalisis data. *Rapidminer* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi(Luluk Elvitaria, 2017). Penggunaan *rapidminer* sangat dibutuhkan untuk menghitung data dalam jumlah besar menggunakan operator. Menurut (Ainurrohmah, 2021) terdapat lebih dari 500 operator yang terdapat pada *rapidminer* yang fungsi dari operator tersebut mencakup aspek utama dari analisis seperti pemuatan, transformasi data, *preprocessing*, visualisasi data, pemodelan dan evaluasi model.

2.2.9 Google Colab

Google Colab adalah *coding environment* berbasis bahasa pemrograman python. Format *google colab* hampir serupa dengan *Jupyter Notebook* berbasis *cloud computing*(Reynaldi, 2021).

Menurut (Gelar Guntara, 2023) *google colab* adalah sebuah IDE untuk pemrograman python dimana pemrosesan data akan dilakukan oleh server google yang memiliki perangkat keras dengan performa yang tinggi. *Google colab* telah menyediakan sebagian besar pustaka (*library*) yang dibutuhkan untuk analisis data yaitu diantaranya:

a. Pandas

Pandas merupakan *library* yang bersifat *open source* dan menyediakan struktur data tingkat tinggi yang fleksibel. Penggunaannya memudahkan analisis data, manipulasi data, dan pembersihan data.

b. Numpy

NumPy (Numerical Python) adalah *library python* yang digunakan untuk bekerja dengan array dan juga memiliki fungsi yang bekerja dalam domain aljabar linier, transformasi fourier, dan matriks.

c. Scikit-learn

Library scikit-learn menyediakan berbagai algoritma dan fungsi yang berguna untuk tugas-tugas seperti pemrosesan data, *preprocessing*, pembelajaran mesin dan evaluasi model.

d. Matplotlib

Matplotlib adalah salah satu library yang populer digunakan untuk visualisasi data di Python. Fungsi-fungsi dalam matplotlib memungkinkan pengguna untuk membuat berbagai jenis plot, grafik, dan visualisasi lainnya.

2.3 State of The Art

State of the art merupakan pengumpulan data penelitian sebelumnya dengan menentukan sumber informasi atau jurnal untuk dianalisa sebagai bahan pembahasan penelitian, serta membedakannya dengan penelitian yang sedang dilakukan. Jurnal tersebut bisa dilihat pada Tabel 2.1.

Tabel 2.2 *State of the art*

No	Nama	Judul	Metode	Keterangan
1	(Al-shehri et al., 2017)	Student Performance Prediction Using Support Vector Machine and K-Nearest Neighbor	Support Vector Machine and K-Nearest Neighbor	Penelitian ini membandingkan dua model algoritma untuk prediksi kinerja siswa dalam ujian akhir. Hasil studi empiris menunjukkan bahwa Support Vector Machine mencapai hasil yang sedikit lebih baik dengan koefisien korelasi 0,96, sedangkan K-Nearest Neighbor mencapai koefisien korelasi 0,95

No	Nama	Judul	Metode	Keterangan
2	(Azis & Fazriansyah, 2019)	Prediksi Tingkat Kelulusan Nilai Mahasiswa Terhadap Mata kuliah Web Programming Menggunakan Metode Neural Network	Neural Network	<p>Pada penelitian ini menggunakan metode Neural Network. Metode Neural Network memiliki kelebihan dalam data yang besar dan memberikan prediksi kelulusan nilai mahasiswa yang akurat agar mendapatkan akurasi yang tepat untuk melakukan klasifikasi kelulusan terhadap mata kuliah unggulan pada program studi Sistem Informasi. Hasil penelitian ini mendapatkan nilai accuracy sebesar 95.67% dan grafik ROC dengan nilai AUC (Area Under Curve) sebesar 0,939.</p>

No	Nama	Judul	Metode	Keterangan
3	(Salmu & Solichin, 2017)	Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta	Naïve Bayes	<p>Penelitian ini ditujukan untuk menemukan faktor – faktor yang mempengaruhi prediksi kelulusan mahasiswa berdasarkan data kinerja akademik mahasiswa pada semester satu sampai empat. Menggunakan model CRISP-DM dengan mengimplementasikan algoritma Naïve Bayes. Dihasilkan akurasi sebesar 80,17% dan disimpulkan model ini bisa dijadikan acuan dalam memprediksi kelulusan mahasiswa. Hasil ini dapat dimanfaatkan oleh pihak UIN sebagai langkah preventif guna menghindari penurunan tingkat kelulusan mahasiswa.</p>

No	Nama	Judul	Metode	Keterangan
4	(M. Ridwan et al., 2013)	Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier	<i>Naive Bayes</i>	Penelitian ini difokuskan untuk mengevaluasi kinerja akademik mahasiswa pada tahun ke-2 dan diklasifikasikan dalam kategori mahasiswa yang dapat lulus tepat waktu atau tidak. Dikembangkan suatu sistem yang dapat mengklasifikasi kelulusan, kemudian memberikan rekomendasi berdasarkan histori nilai yang telah ditempuh mahasiswa. Pengujian pada data mahasiswa angkatan 2005-2009 menggunakan algoritma NBC menghasilkan nilai precision, recall, dan accuracy masing-masing 83%, 50%, dan 70%.

No	Nama	Judul	Metode	Keterangan
5	(Kusumawati et al., 2015)	Prediksi Kelulusan Mahasiswa Menggunakan Metode Neural Network dan particle Swarm Optimization	<i>Neural Network dan Particle Swarm Optimization</i>	Penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa menggunakan metode Neural Network dan Particle Swarm Optimization. Variabel input yang digunakan dalam prediksi kelulusan mahasiswa ini menggunakan yaitu 30 nilai mata kuliah. Output yang dihasilkan adalah kelulusan Tepat Waktu atau Melebihi Batas Waktu. Pengujian dilakukan menggunakan Matlab R2010a. Berdasarkan hasil pengujian dengan 409 data nilai mahasiswa tahun akademik 2008 dan 2009, parameter terbaik dari

No	Nama	Judul	Metode	Keterangan
				Neural Network untuk prediksi kelulusan yaitu didapatkan tingkat akurasi sebesar 87.31%. Sedangkan pengaruh algoritma Particle Swarm Optimization menghasilkan kenaikan akurasi sebesar 9.05 % dari hasil uji menggunakan neural network dan mengurangi nilai error yang dihasilkan oleh Neural Network sebesar 0.022
6	(Prasetyo et al., 2016)	Prediksi Kelulusan Mahasiswa Pada Perguruan Tinggi Kabupaten Majalengka Berbasis	<i>backpropagation neural network dan k-nearest neighbor.</i>	bahwa metode algoritma k-nearest neighbor menghasilkan nilai akurasi tertinggi yang diperoleh dengan nilai jarak (k-100) serta menggunakan 7 variabel terhadap data baru yaitu sebesar 97,90%.

No	Nama	Judul	Metode	Keterangan
		Knowledge Based System		Penelitian ini untuk memprediksi kelulusan mahasiswa diambil dari pola algoritma backpropagation neural network dan algoritma k-nearest neighbor. Proses pengujian menggunakan confusion matrix diketahui
7	(Putri et al., 2019)	Analysis of Students Graduation Target Based on Academic Data Record Using C4.5 Algorithm Case Study: Information Systems Students of	Algoritma C4.5	Penelitian ini bertujuan untuk membuat model prediksi kelulusan mahasiswa yang dapat diperoleh dari proses klasifikasi data menggunakan decision tree dengan algoritma C4.5. Diimplementasikan pada catatan data akademik mahasiswa yang ada sehingga diperoleh dua kelompok mahasiswa, yaitu

No	Nama	Judul	Metode	Keterangan
		Telkom University		adalah mahasiswa yang diprediksi lulus tepat waktu dan mahasiswa lulus terlambat. Digunakan cross validation untuk mengevaluasi kinerja algoritma C4.5 dan hasilnya menunjukkan bahwa akurasi 82.24%.
8	(Maulana et al., 2019)	Prediksi Ketepatan Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi	<i>Decision Tree,</i> <i>Naïve Bayes,</i> <i>K-Nearest Neighbor,</i> <i>Rule Induction,</i> <i>Random Forest</i>	Penelitian ini menggunakan 5 metode algoritma data mining. Kemudian dilakukan komparasi yang bertujuan untuk menentukan metode yang paling optimal dalam menentukan ketepatan kelulusan mahasiswa dengan bantuan pengujian T-Test dan Rapidminer. Dataset yang digunakan sebanyak 349. Berdasarkan

No	Nama	Judul	Metode	Keterangan
				<p>perbandingan algoritma Decision Tree, Naïve Bayes, K- Nearest Neighbor, Rule Induction dan Random Forest dengan dataset kelulusan mahasiswa sebanyak 349, dapat disimpulkan bahwa dengan menggunakan uji beda T-Test algoritma Rule Induction dan C4.5 memiliki performa yang paling baik dalam memprediksi ketepatan kelulusan mahasiswa.</p>
9	(Rosita et al., 2020)	<p>Komparasi Data Mining Naive Bayes Dan Neural Network Memprediksi</p>	<p><i>Naïve Bayes</i> dan <i>Neural Network</i></p>	<p>Tujuan penelitian ini untuk memprediksi kelulusan mahasiswa sebagai acuan untuk menentukan kebijakan preventif terkait pencegahan dini kasus drop out.</p>

No	Nama	Judul	Metode	Keterangan
		Masa Studi Mahasiswa S1		<p>Menggunakan 2 metode algoritma data mining yaitu Naïve Bayes dan Neural Network. Data training dan data testing diambil dari data mahasiswa yang lulus dan drop-out pada tahun 2011 sampai 2019. Sedangkan angkatan 2015–2018 digunakan sebagai data target yang akan diprediksi masa studinya. Sebanyak 3229 mahasiswa, 1769 sebagai data training, 321 sebagai data testing, dan 1139 sebagai data target.</p> <p>Dari data testing diperoleh tingkat akurasi hanya 57,63%. Hasil dari penelitian menggunakan Naïve Bayes menunjukkan</p>

No	Nama	Judul	Metode	Keterangan
				banyaknya kelemahan, dikarenakan tingkat akurasi kevalidannya tergolong tidak terlalu tinggi. Sedangkan akurasi prediksi neural network adalah 72,58%, sehingga metode alternatif inilah yang lebih baik.
10	(Budiyantara et al., 2020)	Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu	Decision Tree, Naive Bayes Dan K-Nearest Neighbor	Prediksi mahasiswa lulus tepat waktu dibutuhkan oleh manajemen Perguruan Tinggi dalam menentukan kebijakan untuk pencegahan kasus Drop Out. Penelitian ini bertujuan untuk menentukan faktor akademis yang berpengaruh terhadap masa studi dengan menggunakan 3 algoritma data mining. Dari hasil evaluasi dan validasi yang

No	Nama	Judul	Metode	Keterangan
				<p>telah dilakukan menggunakan tools RapidMiner diperoleh hasil akurasi dari Metode Decision Tree (C4.5) sebesar 98.04% pada pengujian ke 3. akurasi Metode Naïve Bayes sebesar 96.00% pada pengujian ke 4. Dan akurasi Metode K-Nearest Neighbor (K-NN) sebesar 90.00% pada pengujian ke 2.</p>

7.	Arie Yandi Saputra, Yogi Primadasa(2018)		✓								✓						✓
8.	Andi Gita Novianti, Dian Prasetyo(2017)		✓								✓						✓
9.	Inna Alvi Nikmatun, Indra Waspada		✓								✓						✓
10.	Wilda Imama Sabilla, Tesa Eranti Putri(2017)		✓								✓						✓
11	Abdul Rohman(2015)		✓								✓					✓	✓

2.4 Penelitian Terkait

2.4.1 Abdul Rahman dan Muhammad Rochcham (2019). Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa. Penelitian ini dilakukan perbandingan metode *data mining* yaitu algoritma *Neural Network*, *K-Nearest Neighbor* dan *Decision Tree*. Hasil pengujian ketiga metode dengan mengukur kinerja menggunakan *cross validation*, *confusion matrix* dan kurva ROC, diketahui *Neural Network* memiliki nilai akurasi yang paling tinggi yakni sebesar 87,32%, *Decision Tree* dengan nilai akurasi 83.57%, *K-Nearest Neighbor* dengan akurasi 83,66%. Nilai AUC untuk *Neural Network* menunjukkan nilai tertinggi yakni 0,917 dan yang terendah adalah *Decision Tree* yaitu 0,844.

2.4.2 Sri Widaningsih (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, *Naïve Bayes*, KNN, dan SVM. Penelitian ini bertujuan untuk memprediksi tingkat kelulusan dan IPK dengan menggunakan algoritma *data mining* yaitu *Naïve bayes*, *C4.5*, *K-Nearest Neighbor*, *Support Vector Machine*. Variabel prediktor terdiri dari jenis kelamin dan indeks prestasi dari semester 3 hingga 6. Digunakan *software Rapidminer* untuk mengolah data. Hasil yang diperoleh bahwa algoritma *Naïve Bayes* merupakan algoritma terbaik untuk memprediksi kelulusan mahasiswa yang tepat waktu dan $IPK \geq 3$ dengan nilai *accuracy* (76,79%), *error* (23,17%), dan AUC (0,850).

2.4.3 Ika Kurniawati (2018). Komparasi Metode *Machine Learning* Pada Prediksi Kelulusan Mahasiswa. Tujuan penelitian ini untuk mengetahui kinerja dari 5 model *data mining* dalam memprediksi kelulusan mahasiswa dengan melakukan perbandingan nilai akurasi dari 5 model tersebut, yaitu algoritma C4.5, *K-Nearest Neighbor*, *Iterative Dichotomiser Three (ID3)*, *Logistic Regression* serta *Naive Bayes*. Hasil pengujian didapatkan nilai akurasi tertinggi untuk prediksi kelulusan mahasiswa adalah dengan model algoritma C4.5 dengan nilai akurasi 94.33% dan kurva ROC 0.930 sedangkan paling rendah tingkat akurasinya adalah model *K-Nearest Neighbor* dengan nilai akurasi 50.93% dan nilai kurva ROC 0.500.