

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

1. Klasifikasi

Klasifikasi merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data baru untuk diklasifikasi. Teknik ini menggunakan supervised induction yang memanfaatkan kumpulan pengujian dari data set yang terklasifikasi. Dalam definisi yang lain klasifikasi dalam data mining dapat didefinisikan sebagai cara menganalisis data dengan menggunakan *training* data untuk membentuk sebuah model yang digunakan untuk memprediksi kelas label dari sebuah data (Aradea, A, Z, & A, 2011).

Pengukuran kinerja klasifikasi dilakukan dengan mengevaluasi hasil pengujian menggunakan *Confusion Matrix*. *Confusion Matrix* merupakan metode untuk mengevaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah (Gorunescu, 2011).

2. Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya dapat diperkecil (Farida dkk., 1978).

Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Orozco dkk., 2013)

Menurut Kamus Besar Bahasa Indonesia (<https://kbbi.web.id/prediksi>), prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi sama dengan ramalan atau perkiraan yang akan terjadi pada suatu objek disuatu saat atau suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan mendatang.

3. Data Mining

Data Mining merupakan proses pengestraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data (R & Hirnawal, 2013). *Data Mining* adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya (Mabrur & Lubis, 2012). *Data Mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar (Mujiasih, 2011). Maka dari beberapa definisi di atas penulis dapat menyimpulkan bahwa *data mining* merupakan proses atau kegiatan untuk mengumpulkan data yang berukuran besar, kemudian mengekstaksi data tersebut menjadi informasi-informasi yang nantinya

dapat digunakan. *Data Mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Kartini, 2017).

Menurut (Ridwan, 2013) sebagai suatu rangkaian proses, *Data Mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*. Tahap-tahap *Data Mining* adalah sebagai berikut:

a. Pembersihan data (*Data Cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.

b. Integrasi data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.

c. Seleksi data (*Data Selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

d. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *Data Mining*.

e. Proses *Mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan *Data Mining*.

f. Evaluasi pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

g. Presentasi pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

4. Sekolah Menengah Atas (SMA)

Sekolah Menengah Atas (SMA) merupakan jenjang pendidikan dasar pada pendidikan formal di Indonesia setelah lulus Sekolah Menengah Pertama (SMP) atau sederajat. Sekolah menengah atas ditempuh dalam waktu 3 tahun, mulai dari kelas 10 sampai kelas 12. Sebagai keberlanjutan dari program Wajib Belajar 9 Tahun, pada tahun 2012 ini Pemerintah Pusat mencanangkan program Wajib Belajar 12 Tahun atau yang lebih dikenal dengan nama Pendidikan Menengah Universal (PMU). Adapun payung hukum untuk program PMU ini yaitu Peraturan Menteri Pendidikan dan Kebudayaan No.80 Tahun 2013. Program ini dimaksudkan untuk menjaga kesinambungan keberhasilan pelaksanaan program Wajib Belajar Pendidikan Dasar 9 Tahun sekaligus menyiapkan generasi emas Indonesia 2045 (Hermiyanty, Wandira Ayu Bertin, 2017).

Lulusan SMA dapat melanjutkan pendidikan ke perguruan tinggi. Pelajar SMA umumnya berusia 16-18 tahun. Setiap warga negara Indonesia berusia 7-15 tahun wajib mengikuti pendidikan dasar, yakni sekolah dasar atau sederajat selama 6 tahun, sekolah menengah pertama atau sederajat selama 3 tahun dan sekolah menengah atas atau sederajat selama 3 tahun.

Sekolah menengah atas diselenggarakan oleh pemerintah maupun swasta. Sejak diberlakukannya otonomi daerah pada tahun 2001, pengelolaan sekolah menengah atas negeri di Indonesia yang sebelumnya berada di bawah Kementerian Pendidikan Nasional, kini menjadi tanggung jawab pemerintah provinsi, sedangkan Kementerian Pendidikan Nasional hanya berperan sebagai regulator dalam bidang standar nasional pendidikan. Secara struktural, sekolah menengah atas negeri merupakan unit pelaksana teknis dinas pendidikan provinsi.

5. *Naïve Bayes*

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Patil & Sherekar, 2013). Definisi lain mengatakan *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya (Bustami, 2013).

Naïve Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Bustami, 2013). Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Pattekari & Parveen, 2012).

Persamaan dari teorema Bayes adalah (Bustami, 2013):

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (\text{II-1})$$

Keterangan :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

P(H|X) : Probabilitas hipotesis *H* berdasar kondisi *X* (posteriori probabilitas)

P(H) : Probabilitas hipotesis *H* (prior probabilitas)

P(X|H) : Probabilitas *X* berdasarkan kondisi pada hipotesis *H*

P(X) : Probabilitas *X*

Untuk menjelaskan metode *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naïve Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C) \cdot P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (\text{II-2})$$

Di mana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (II-3)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss*:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (II-4)$$

Keterangan :

P : Peluang

X_i : Atribut ke i

x_i : Nilai atribut ke i

Y : Kelas yang dicari

y_i : Sub kelas Y yang dicari

μ : *mean*, menyatakan rata – rata dari seluruh atribut

σ : *Deviasi standar*, menyatakan varian dari seluruh atribut.

6. C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk pohon keputusan. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu.

Algoritma Pohon Keputusan C4.5 atau Classification version 4.5 adalah pengembangan dari algoritma ID3. Oleh karena pengembangan tersebut, algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3. Secara umum, proses algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk tiap-tiap nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

Secara khusus, algoritma C4.5 Decision Tree menggunakan kriteria split yang telah dimodifikasi yang dinamakan Gain Ratio dalam proses pemilihan split atribut. Split atribut merupakan proses utama dalam pembentukan pohon keputusan (Decision Tree) di C4.5.

Tahapan dari algoritma C4.5 adalah sebagai berikut.

- a. Menghitung nilai *Entropy*

- b. Menghitung nilai Gain Ratio untuk masing-masing atribut
- c. Atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi akar (*root*) dan atribut yang memiliki nilai *Gain Ratio* lebih rendah dari akar (*root*) dipilih menjadi cabang (*branches*)
- d. Menghitung lagi nilai *Gain Ratio* tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (*root*) di tahap sebelumnya
- e. Atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi cabang (*branches*)
- f. Mengulangi langkah ke-4 dan ke-5 sampai dengan dihasilkan nilai *Gain* = 0 untuk semua atribut yang tersisa. Untuk menghitung nilai *Entropy* dapat dihitung dengan persamaan :

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (II-5)$$

Keterangan :

S = himpunan kasus

n = jumlah partisi S

Pi= proporsi dari terhadap S

Sementara itu nilai *information gain* (*Gain*) dapat dihitung menggunakan persamaan :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (II-6)$$

Keterangan :

S = himpunan kasus

A = atribut

n = jumlah partisi atribut A

$||$ = jumlah kasus pada partisi ke-i

$|S|$ = jumlah kasus dalam S

Selanjutnya nilai Split Info dapat dihitung dengan persamaan :

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (II-7)$$

Keterangan :

S = himpunan kasus

A = atribut

S_i = jumlah sampel untuk atribut i

Maka nilai *Gain Ratio* yang menentukan sebuah atribut dapat dijadikan akar maupun cabang suatu pohon keputusan dapat dihitung dengan persamaan:

$$GainRatio(S, A) = \frac{Gain(S,A)}{SplitIn(S,A)} \quad (II-8)$$

Keterangan :

S = himpunan kasus

A = atribut

Gain(S,A) = info gain pada atribut A

SplitInfo(S,A) = split info pada atribut A

7. Rapidminer

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna

sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. Berdasarkan Gartner Magic Quadrant for Data Science and Machine Learning Platforms 2021 RapidMiner masuk kedalam kategori *Visioners*.

RapidMiner sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. *RapidMiner* didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan *RapidMiner* di lebih dari 40 negara. *RapidMiner* sebagai *software open source* untuk *data mining* tidak perlu diragukan lagi karena *software* ini sudah terkemuka di dunia. *RapidMiner* menempati peringkat pertama sebagai *Software data mining* pada polling oleh KDnuggets, sebuah portal *data-mining* pada 2010-2011.

RapidMiner menyediakan GUI (*Graphic User Interface*) untuk merancang sebuah *pipeline analitis*. GUI ini akan menghasilkan file XML (*Extensible Markup Language*) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh *RapidMiner* untuk menjalankan analisis secara otomatis.

2.2 Penelian Terkait dan Keterbaruan Penelitian

Penelitian Terkait merupakan proses membaca dan memahami penelitian-penelitian sebelumnya yang berkaitan dengan penelitian yang dilakukan. Adapun penelitian-penelitian terkait yang dijadikan acuan terdapat pada Tabel 2.1.

Tabel 2. 1 Penelitian Terkait

Judul	Penulis	Tahun Terbit	Metode
IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI KELULUSAN SISWA DENGAN METODE NAÏVE BAYES	Karolina Sinaga, Koko Handoko	JURNAL COMASIE - VOL. 04 NO. 06 (2021) ISSN (Online) 27156265	Data yang diseleksi yaitu berupa data-data siswa pada Sekolah Menengah Kejuruan (SMK) Putra Jaya School dari tahun 2017 hingga tahun 2019. Selanjutnya melakukan pengolahan data yang telah diseleksi, selanjutnya diolah kedalam metode naïve bayes dari teknik data mining. Langkah terakhir yaitu melakukan pengujian menggunakan aplikasi WEKA versi 3.9 dalam memprediksi tingkat kelulusan siswa pada Sekolah Menengah Kejuruan.

Judul	Penulis	Tahun Terbit	Metode
Analisis Kinerja Algoritma <i>C4.5</i> dan <i>Naïve Bayes</i> Dalam Memprediksi Keberhasilan Sekolah Menghadapi UN	Yeni Angraini, Siti Fauziah, Jordi Lasmana Putra	Jurnal ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER VOL. 5 NO. 2 FEBRUARI 2020 P- ISSN : 2685 – 8223 E – ISSN : 2587 – 4864.	Penelitian ini membahas tentang memprediksi keberhasilan sekolah yang akan menghadapi UN (Ujian Nasional) dengan data mining dan menggunakan algoritma <i>C4.5</i> dan <i>Naïve Bayes</i> sebagai perbandingan dari akurasi kedua algoritma tersebut dengan menggunakan tools rapidminer. Hasil yang didapat dari beberapa perbandingan yang dilakukan diantaranya menggunakan <i>Confusion Matrix</i> , kurva ROC dan t-Tes pada dataset diketahui bahwa <i>Naïve Bayes</i> memiliki nilai akurasi 95% sedangkan <i>C4.5</i> mendapat nilai akurasi 78,50%. Data yang digunakan berasal dari rata-rata nilai UN di Banda Aceh yang didapat dari situs data.bandaacehkota.go.id
Komparasi Algoritma Decion Tree, <i>Naïve Bayes</i> dan K-Nearest Neighbor Untuk	Agus Budiyantra, Irwansyah, Egi Prengki,	JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER VOL. 5	Penelitian ini membahas tentang prediksi mahasiswa lulus tepat waktu atau tidak dengan menggunakan data mining dan 3 algoritma sebagai komparasi dari akurasi kinerja yang ingin diketahui. Atribut yang digunakan

Judul	Penulis	Tahun Terbit	Metode
Memprediksi Mahasiswa Lulus Tepat Waktu	Pandi Ahmad Pratama, Ninuk Wiliani	NO. 2 FEBRUARI 2020 p-ISSN : 2685 – 8223 E-ISSN : 2527 – 4864.	berjumlah 11 yaitu NPM, jenis kelamin, usia, jurusan, kelas, pekerjaan, indek prestasi semester 1 sampai 4, dan keterangan sebagai atribut hasil serta menggunakan aplikasi rapidminer. Penelitian ini menggunakan metode kuantitatif dan data yang didapat dari bagian administrasi adalah data mahasiswa teknik informatika tahun akademik 2010/2011, 2011/2012, 2012/2013, 2014 berjumlah 342 data kemudian 242 sebagai data training dan 100 data digunakan untuk data testing. Hasil yang didapat <i>Decision Tree (C4.5)</i> memperoleh hasil akurasi tertinggi sebesar 98,04% pada pengujian ke 3. Metode <i>Naïve Bayes</i> memperoleh 96,00% pada pengujian ke 4, dan K-Nearest Neighbor memperoleh akurasi tertinggi 90,00% pada pengujian ke 2.

Judul	Penulis	Tahun Terbit	Metode
<p>PREDIKSI KEBERHASILAN IMMUNOTHERAPY PADA PENYAKIT KUTIL DENGAN MENGGUNAKAN ALGORITMA <i>NAÏVE BAYES</i></p>	<p>Toni Arifin, Siti Syalwah</p>	<p>JURNAL RESPONSIF, Vol. 2 No.1 Februari 2020, pp. 38~43 E-ISSN: 2685-6964</p>	<p>Penelitian ini menerapkan algoritma klasifikasi <i>Naïve Bayes</i> pada Immunotherapy Data Set menggunakan metode k-fold validasi performa klasifikasi.</p>
<p>MENENTUKAN PREDIKSI KELULUSAN SISWA DENGAN MEMBANDINGKAN ALGORITMA <i>C4.5</i> DAN <i>NAIVE BAYES</i> STUDI KASUS SMKN. 1 CIKARANG SELATAN</p>	<p>M Makmun Effendi, Arie Setiawan</p>	<p>Volume 10 Nomor 3 Maret 2020 ISSN : 2407-3903</p>	<p>Penelitian ini menggunakan dua metode data mining, yaitu algoritma <i>C4.5</i> dan <i>Naïve bayes</i>. Penggunaan dua metode pada penelitian ini bertujuan untuk membandingkan kinerja dari kedua algoritma dalam memprediksi kelulusan siswa berdasarkan tingkat accuracy, precision, dan recall yang didapatkan.</p>
<p>Perbandingan Kinerja 6 Algoritma Klasifikasi Data Mining Untuk Prediksi Masa Studi Mahasiswa</p>	<p>Mariana Windarti, Agustinus Suradi.</p>	<p>1 No. 1 Februari 2019 ISSN : 1979 – 925X e-ISSN : 2442 – 4528.</p>	<p>Penelitian ini membahas tentang perbandingan kinerja dari 6 algoritma klasifikasi data mining yaitu Decision Tree (DT) <i>C4.5</i>, Bayesian Network (BN), K-Nearest Neighbors (KNN), <i>Naïve Bayes</i> (NB), Neural Network (NN) dan SVM (Support Vector Machine).</p>

Judul	Penulis	Tahun Terbit	Metode
			<p>Hasil dari penelitian ini menunjukkan bahwa ke-6 algoritma tersebut mampu melakukan prediksi masa studi mahasiswa UNWIDHA kedalam empat kelas. Algoritma Bayesian Network memiliki kinerja paling baik dibanding lainnya baik berdasarkan nilai akurasi, presisi, recall dan nilai AUC. Nilai akurasi BN sebesar 80.615%, nilai presisi dan recall sebesar 0.785 & 0.806, sedangkan untuk nilai AUC sebesar 0.837. Sedangkan untuk kinerja diposisi terendah adalah Decision Tree <i>C4.5</i> .</p>
<p>Optimasi Data Mining Menggunakan Algoritma <i>Naive Bayes</i> dan <i>C4.5</i> Untuk Klasifikasi Kelulusan Mahasiswa</p>	<p>Ni Luh Ratniasih</p>	<p>Jurnal Teknologi Informasi dan Komputer, Volume 5, Nomor 1, Januari 2019.</p>	<p>Penelitian menggunakan 2 metode sebagai komparasi yaitu <i>Naive Bayes</i> dengan <i>C4.5</i> yang merupakan sebuah metode untuk melakukan teknik klasifikasi serta menggunakan salah satu tools data mining yaitu Rapid Miner. Data set yang digunakan berjumlah 50 record dan 4 atribut predicto 1 atribut target.</p>

Judul	Penulis	Tahun Terbit	Metode
<p>Impelementasi Data Mining Algoritma C4.5 Untuk Memprediksi Kelulusan Uji Kompetensi SMK Teknik Komputer Dan Jaringan (TKJ) (Study Kasus : SMK Pembangunan Daerah Lubuk Pakam)</p>	<p>Angga Kurniawan</p>	<p>Majalah Ilmiah INTI, Volume 4, Nomer 3, September 2019 ISSN 2339 – 210X.</p>	<p>Penelitian ini membahas tentang penerapan data mining dengan menggunakan algoritma C4.5 dengan menggunakan aplikasi rapidminer untuk memprediksi kelulusan uji kompetensi SMK. Hasil dari penelitian ini didapatkan kecocokan antara menggunakan cara manual dengan menggunakan data mining</p>
<p>Data Mining Di Bidang Pendidikan Untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi Pada STIMIK Jabar</p>	<p>Galih</p>	<p>Jurnal Teknologi Sistem Informasi dan Aplikasi Vol,2 No 1, Januari 2019 ISSN 2654 – 3788 e-ISSN 2654 – 4229</p>	<p>Pada penelitian ini algoritma yang digunakan adalah <i>Naive Bayes</i> classifier dan C4.5. Hasil dari pengujian 2 algoritma tersebut dengan menggunakan ratio data training untuk nilai akurasi NBC adalah 86,83% dengan ratio data training 80% sedangkan untuk algoritma C4.5 menghasilkan nilai akurasi 88,10% dengan ratio data training 90%. Dilihat dari hasil tersebut algoritma C4.5 memperoleh nilai akurasi terbaik.</p>

Judul	Penulis	Tahun Terbit	Metode
Penerapan Metode Neural Network Untuk Prediksi Nilai Ujian Nasional (Study Kasus Di SMK Muhammadiyah Slawi)	Nur Ariesanto Ramadhan	Systax Literatur : Jurnal Ilmiah Indonesia p-ISSN : 2541 – 0849 e-ISSN : 2548 – 1398 Vol. 4, No. 4 April 2019.	Jurnal ini membahas tentang penggunaan metode neural network untuk memprediksi nilai ujian nasional (UN). Variabel input yang digunakan dalam penelitian ini menggunakan laporan evaluasi pada semester 1, 2, 3, 4,5 dan variable ouput Ujian Nasional dalam proses pembelajaran tahun 2012/2013. Jumlah data yang digunakan berjumlah 193 siswa. Hasil prediksi diambil dari prediksi dengan akurasi terbaik dan waktu komputasi terpendek. Hasil dari setiap mata pelajaran adalah Indonesia dengan menggunakan root mean squared error (RMSE) (1,129 +/- 0,090), Bahasa Inggris (0,835 +/- 0,102), dan Matematika (1,238 +/- 0,133).
Analisa Kinerja Algoritma C4.5 Dalam Memprediksi Hasil Belajar	Muhammad Ardiansyah Sembiring, dkk.	Journal of Science and Social Research February 2018, I (1): 73 – 79 ISSN 2615 – 4307 ISSN 2615 – 3262.	Penelitian ini menghasilkan 7 buah pola aturan (rule) sebagai landasan dalam memprediksi pencapaian hasil belajar mahasiswa. Tools yang digunakan untuk mempermudah dalam penelitian ini adalah Rapidminer 5.

Judul	Penulis	Tahun Terbit	Metode
SISTEM PREDIKSI KESIAPAN SISWA SMP DALAM MENGHADAPI UJIAN NASIONAL MENGGUNAKAN METODE NAÏVE BAYES	Medika Purnama Putri , Fatah Yasin Al Irsyadi	JMP Online Vol 2, No. 5, 469-481. © 2018 Kresna BIP. e-ISSN 2550-0481 p-ISSN 2614-7254	Metode yang digunakan dalam penelitian ini menggunakan metode <i>Naïve Bayes</i> yang bisa memprediksi kesiapan siswa dalam menghadapi ujian nasional dengan dibantu sistem sederhana yang dibuat.
Penerapan Algoritma <i>C4.5</i> Dalam Memprediksi Kesiapan Siswa SMP IT PAPB Semarang Menghadapi Ujian Nasional	Nadia Zulfa Rahma, Andik Setyono	Vol. 8, No. 1, Januari 2018	Metode klasifikasi dapat diterapkan untuk memprediksi kesiapan siswa SMP Islam Terpadu PAPB Semarang menghadapi ujian nasional. Hal tersebut berdasarkan pengujian yang telah dijelaskan pada bab sebelumnya, pemodelan algoritma <i>C4.5</i> menghasilkan akurasi sebesar 99,48%.

Judul	Penulis	Tahun Terbit	Metode
<p>Penerapan Data Mining dengan Algoritma Neural Network (Backpropagation) Untuk Prediksi Lama Studi Mahasiswa</p>	<p>Dwi Kartini</p>	<p>Prosiding 2017 ISSN 2597 – 3584 2017.</p>	<p>Jurnal ini membahas tentang data mining yang merupakan proses untuk menggali nilai tambah berupa pengetahuan baru yang selama ini tidak diketahui dari sekumpulan data. Prediksi ini bertujuan untuk menentukan factor akademis kedepannya. Variabel input yang digunakan ada 4 yaitu IP semester 1, semester 2, semester 3, semester 4 dengan output yang dihasilkan berupa lama studi mahasiswa dengan kriteria lama studi kurang dari sama dengan 4.5 tahun = tepat dan lama studi lebih dari 4.5 tahun = lambat. Hasil dari pengujian dengan menggunakan algoritma neural network ini menghasilkan tingkat akurasi 99%.</p>

Judul	Penulis	Tahun Terbit	Metode
<p>KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN METODE ALGORITMA C4.5 BERBASIS WEB</p>	<p>Erricha Paramitha Dewi, Ade eviyanti S.kom</p>	<p>Universitas Muhammadiyah Sidoarjo (2015)</p>	<p>Algoritma C45 dapat digunakan untuk memudahkan dalam pengambilan keputusan dengan memproyeksikan data-data yang ada ke dalam bentuk pohon keputusan, berdasarkan nilai entropy dan gain yang dimiliki masing-masing atribut data.</p>
<p>Prediksi Nilai Ujian Nasional Produktif Sekolah Menengah Kejuruan Menggunakan Metode Neural Network</p>	<p>Winarto Utomo</p>	<p>Techno.COM, Vol. 14, No.1, Februari 2015 : 33 – 41.</p>	<p>Jurnal ini membahas tentang tujuan untuk menerapkan algoritma neural network untuk memprediksi nilai Ujian Nasional mata pelajaran Teori produktif di SMK sehingga dapat digunakan sebagai acuan pengayaan pada mata pelajaran Teori Produktif. Manfaat dari penelitian di jurnal ini sendiri adalah diharapkan dapat digunakan oleh pihak sekolah dalam menentukan kebijakan dalam melakukan pengayaan selain itu penelitian ini dapat memprediksi hasil nilai UN mata pelajaran teori produktif di SMK. Hasil dari penelitian</p>

Judul	Penulis	Tahun Terbit	Metode
			di jurnal ini adalah nilai RMS terkecil 0.169+/-016 dengan number of validation 5, hidden layer 3, learning rate 0.2, momentum 0,1, training of cycles 500

Tabel 2. 2 Tabel Relefansi Penelitian

No	Peneliti	Metode	Algoritma	Tools	Objek Penelitian
1.	Mariana Windarti, dkk (2019)	Klasifikasi	C.45, Support Vector Machine, Bayesian Network, Neural Network, K-Nearest Neighbors, Naïve Bayes	RapidMiner, Weka	Mahasiswa
2.	Ni Luh Ratniasih (2019)	Klasifikasi	C.45, Naïve Bayes	RapidMiner	Mahasiswa
3.	Galih (2019)	Klasifikasi	C.45, Naïve Bayes Classifier	RapidMiner	Mahasiswa
4.	Muhammad A.S, dkk (2018)	Klasifikasi	C.45	RapidMiner	Mahasiswa
5.	Angga Kurniawan (2019)	Klasifikasi	C.45	RapidMiner	SMA/SMK
6.	Yeni Angraini, dkk (2020)	Klasifikasi	C.45, Naïve Bayes	RapidMiner	SMA/SMK
7.	Agus, dkk (2020)	Klasifikasi	Decision Tree, K-Nearest Neighbors, Naïve Bayes	RapidMiner	Mahasiswa
8.	Winarto Utomo (2015)	Klasifikasi	Neural Network	RapidMiner	SMA/SMK

Tabel 2. 2 Tabel Relefansi Penelitian

No	Peneliti	Metode	Algoritma	Tools	Objek Penelitian
9.	Nur Ariesanto R (2019)	Klasifikasi	Neural Network	RapidMiner	Mahasiswa
10.	Dwi Kartini (2017)	Klasifikasi	Neural Network	RapidMiner	Mahasiswa
11.	Medika,dkk (2018)	Klasifikasi	Naïve Bayes	SISTEM	SMP
12.	Nadia, dkk (2018)	Klasifikasi	C.45	SISTEM	SMP
13.	Errica Paramitha ,dkk (2015)	Klasifikasi	C.45	SISTEM	
14.	Toni Arifin, dkk (2020)	Klasifikasi	Naïve Bayes	RapidMiner	
15.	Sinaga, k., dkk. (2021)	Klasifikasi	Naïve Bayes	Weka	SMA/SMK
16.	Ari Rizqi Mardhotillah, (2022)	Klasifikasi	C4.5, Naïve Bayes, 45 atribut	RapidMiner	SMA

Tabel 2. 3 Keterbaruan Penelitian

Peneliti	(Sinaga, k. dkk., 2021)	(Ari Rizqi Mardhotillah, 2022)
Judul	Implementasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naïve Bayes	ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN NAÏVE BAYES PADA KELULUSAN SISWA SMA NEGERI 1 SINGAPARNA
Masalah Penelitian	Membuat aplikasi pendukung untuk perhitungan tingkat kelulusan siswa dalam penerapan data mining untuk siswa yang lulus dan tidak lulus dengan algoritma <i>Naive bayes</i> dengan menghitung nilai Ujian Akhir Sekolah (UAS), Ujian Nasional (UN) dan nilai Akhir (NA)	Membuat model perhitungan dengan algoritma <i>C4.5</i> dan <i>Naive bayes</i> dalam menentukan siswa yang berpotensi lulus dan tidak lulus berdasarkan nilai raport dan nilai Ujian Sekolah (US)
Objek Penelitian	Klasifikasi	Klasifikasi
Algoritma/ Metode	<i>Observasi dan wawancara, Naive bayes</i>	<i>Observasi dan wawancara, 45 atribut, C4.5, Naive bayes</i>
Implementasi	Klasifikasi dilakukan dengan <i>Tools WEKA</i>	Prediksi dilakukan dengan <i>Tools RapidMiner</i>