

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Pendidikan**

Dalam Undang-Undang Republik Indonesia Nomor 20 tahun 2003 pasal 1 disebutkan bahwa pendidikan merupakan usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara (Habe & Ahiruddin, 2017).

Dikutip dalam jurnal (Pambudi & Setiawan, 2018) Pendidikan merupakan suatu proses yang sadar tujuan. Tujuan diartikan sebagai suatu usaha untuk memberikan rumusan hasil yang diharapkan oleh siswa setelah mendapatkan pengalaman belajar. Siswa dapat dibilang berhasil dilihat dari prestasi yang dicapai sebagai tujuan pengajaran. Dengan prestasi yang tinggi, para siswa mempunyai indikasi pengetahuan yang baik.

Motivasi menjadi salah satu faktor penting yang dapat mempengaruhi prestasi siswa di sekolah. Dengan adanya motivasi, siswa mendapat dukungan mental untuk belajar lebih keras, ulet, tekun dan memiliki konsentrasi penuh dalam proses pembelajaran. Dorongan motivasi perlu dilakukan dalam upaya pembelajaran di sekolah untuk menunjang prestasi siswa. Penting bagi siswa untuk mengenal prestasi belajarnya, karena dengan mengetahui hasil yang sudah dicapai maka siswa mendapat dorongan untuk meningkatkan prestasi yang telah didapatkan sebelumnya (Pambudi & Setiawan, 2018).

## **2.2 Beasiswa**

Dalam Undang-Undang Republik Indonesia Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional, Pasal 12 menyebutkan bahwa setiap peserta didik pada setiap satuan pendidikan berhak mendapatkan beasiswa bagi yang berprestasi dan bagi mereka yang orang tuanya tidak mampu membiayai pendidikannya (Habe & Ahiruddin, 2017).

Pengertian beasiswa menurut Sihotang adalah pembiayaan yang tidak bersumber dari pendanaan sendiri ataupun orangtua, akan tetapi yang diberikan oleh pemerintah, perusahaan swasta, kedutaan, universitas, serta lembaga pendidik atau peneliti lainnya untuk meningkatkan kapasitas sumber daya manusia melalui pendidikan. Sedangkan menurut Putra dan Hardiyanti Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan dengan tujuan untuk digunakan demi keberlangsungan pendidikan yang ditempuh (Saputra & Ayuningtias, 2016).

Sedangkan menurut (Baskoro et al., 2021) Beasiswa merupakan donasi dari pemerintah, lembaga atau yang diberikan kepada siswa yang masih menempuh pendidikan dalam rangka mencari ilmu pengetahuan. Donasi ini biasanya berbentuk dana untuk membayar biaya pendidikan yang dikeluarkan siswa dalam menempuh pendidikan di sekolah.

## **2.3 Data Mining**

### **2.3.1 Pengertian Data Mining**

Data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu

*database* yang sangat besar dengan memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan secara otomatis. Definisi lain diantaranya adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari (Kasus & Maal, 2018).

Dalam penelitian lain, menurut (Rahwali, Rasti Hansun, Seng Wiratama, 2017) menjelaskan bahwa Data Mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis yang menentukan keteraturan, pola dan hubungan dalam dataset berukuran besar. Maksud dari pengertian ini adalah proses pencarian informasi yang tidak diketahui sebelumnya dari sekumpulan data yang besar. Karakteristik data mining adalah sebagai berikut:

- a. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. Data mining menggunakan data yang besar. Data besar digunakan untuk membuat hasil lebih dipercaya
- c. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

Data mining juga merupakan serangkaian proses untuk mencari nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui sebelumnya. Secara khusus, simpanan metode yang disebut sebagai data mining memberikan metodologi data medis dari prediksi. Akurasi prediksi bergantung pada kualitas prediksi pada semua aplikasi dan data mining.

Ada beberapa peran utama data mining, yaitu sebagai berikut:

Data mining memiliki peranan utama dalam pengolahan data, diantaranya adalah sebagai berikut:

1) *Association*

Metode *Association* biasa disebut sebagai *Market Basket Analysis*. Metode ini biasanya digunakan untuk mencari hubungan sebuah variabel dalam mengatasi sebuah problem yang mengidentifikasi produk-produk yang sering diberi secara bersamaan oleh *customer*.

2) *Classification*

Metode *Classification* merupakan suatu tindakan untuk memberikan kelompok kepada setiap keadaan.

3) *Clustering*

Metode *Clustering* digunakan untuk mengidentifikasi kelompok alami dari kasus yang didasarkan pada sebuah kelompok atribut untuk mengelompokkan data yang memiliki kemiripan.

4) *Estimation*

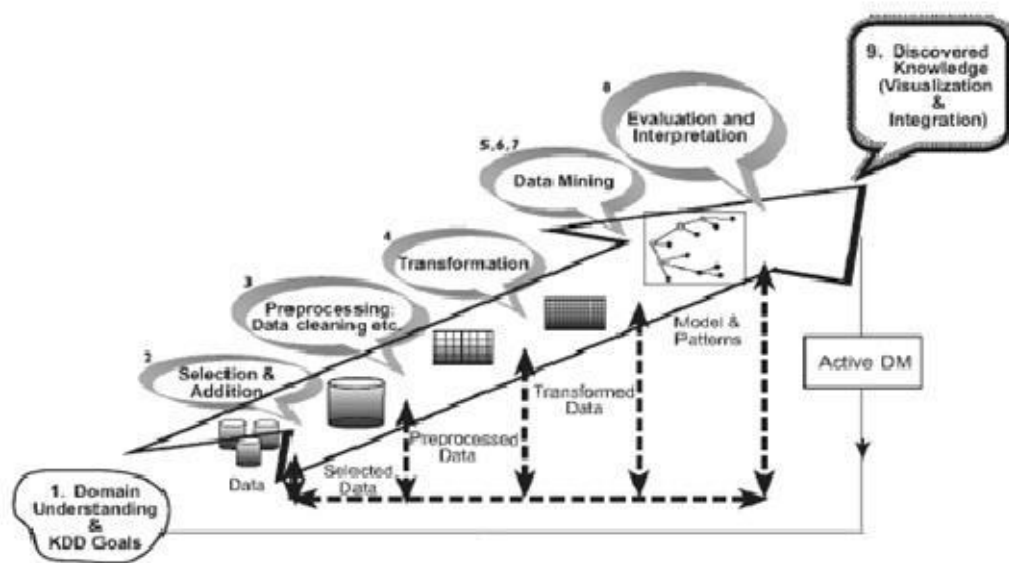
Metode *Estimation* sama halnya dengan metode *Classification*, namun variabel tujuan pada metode *estimation* lebih ke arah numerik daripada kategori.

5) *Prediction*

Metode *Prediction* adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang.

### 2.3.2 Tahapan Data Mining

Data mining mempunyai beberapa model proses yang digunakan, untuk mengarahkan pelaksanaan data mining, model proses data mining yang biasa digunakan yaitu *Knowledge Discovery Database (KDD)*, CRISP-DM, dan SEMMA.



Gambar 2.1 *Knowledge Discovery in Database*

Salah satu tahapan dalam keseluruhan proses KDD adalah data mining.

Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

- 1) *Data Selection* : pada tahapan ini, dilakukan penyeleksian, pembuatan kelompok data informasi dan target. Kemudian hasil dari penyeleksian data tersebut disimpan terpisah dari basis data operasional didalam berkas.
- 2) *Preprocessing dan Cleaning Data* : Meliputi pemilihan data, tugas KDD yaitu untuk mengambil data yang relevan dari database. *Data cleaning* berfungsi untuk menghilangkan *noise* dan data *double*, untuk menangani data yang

hilang. Dan data *integration* berfungsi untuk menyatukan data dari berbagai sumber

- 3) *Transformation* : Untuk mengubah data menjadi bentuk yang sesuai, yaitu untuk menemukan fitur yang berguna untuk contoh data.
- 4) *Data Mining* : Proses yang paling penting digunakan untuk mengekstrasi pola.
- 5) *Interpretation* : Operasi dasar yang termasuk pengidentifikasian pola yang menarik serta mewakili pengetahuan dan menyajikan pengetahuan yang digali dengan cara yang mudah di pahami.

#### **2.4 Prediksi**

Prediksi merupakan suatu proses memperkirakan sistematis mengenai sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti tentang apa yang akan terjadi melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Selain itu menurut Kamus Besar Bahasa Indonesia (KBBI), prediksi memiliki makna kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan (Niranjan Banik, Adam Koesoemadinata, Charles Wagner, Charles Inyang, 2013).

## 2.5 Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi, segmentasi atau pengelompokan dan bersifat prediktif. Dengan algoritma ini mesin (komputer) diberikan sekelompok data untuk dipelajari atau sering disebut sebagai learning dataset. Dasar dari algoritma C4.5 itu sendiri adalah pembentukan pohon keputusan (*Decision Tree*) (Algoritma et al., 2020).

Menurut (Hendrian, 2018) Algoritma C4.5 salah satu metode yang melibatkan konstruksi pohon keputusan. Algoritma C4.5 adalah algoritma yang mempunyai *input* berupa *training samples* berupa data contoh yang akan digunakan untuk membangun sebuah pohon yang telah diuji kebenarannya dan *samples* yang merupakan *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan prediksi data. Algoritma dasar dari C4.5 adalah sebagai berikut: (1) Pohon yang dihasilkan berupa pohon terbalik. (2) Pada tahap awal, semua contoh *training* adalah akar. (3) Atribut adalah kategori. (4) Contoh di partisi secara berulang berdasarkan atribut yang dipilih. (5) Atribut test dipilih dari pengukuran statistik. Langkah-langkah untuk membentuk pohon keputusan dengan Algoritma C4.5, yaitu: (1) Menyiapkan *data training* yang diambil dari data masa lampau yang dibuat kedalam kelas-kelas tertentu. (2) Memilih atribut akar yang didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada.

Untuk mendapatkan nilai *gain*, harus ditentukan terlebih dahulu nilai *Entropy* dengan rumus:

$$Entropy(S) = \sum_{i=0}^n -p_i * \log_2 p_i \quad (2.1)$$

*Keterangan:*

$S$  = Himpunan kasus  
 $n$  = Atribut  
 $P_i$  = Proporsi dari  $S_i$  terhadap  $S$

Rumus *Gain*:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2.2)$$

*Keterangan:*

$S$  = Himpunan kasus  
 $A$  = Atribut  
 $n$  = Jumlah partisi atribut  
 $|S_i|$  = Jumlah kasus pada partisi ke- $i$   
 $|S|$  = Jumlah kasus dalam  $S$

## 2.6 Decision Tree

*Decision Tree* sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat serta banyak diminati. Dalam *decision tree* ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan Bahasa alami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit. Walaupun tidak menutup kemungkinan dapat juga digunakan untuk kasus data dengan atribut numerik (Dewi, 2016).

Sedangkan penjelasan lain yang dikutip dari (Hendrian, 2018) bahwa *Decision Tree* atau Pohon Keputusan adalah struktur sederhana yang dapat digunakan sebagai pengklasifikasi. Referensi penting dalam pengerjaan aslinya adalah *Classification and Regression Tree* oleh Breiman et al. Pada pohon keputusan, masing-masing node internal (*non-leaf*) merepresentasikan sebuah variabel atribut (atribut prediksi atau fitur) dan masing-masing cabang



merepresentasikan satu keadaan dari variabel ini. Masing-masing dari tiga daun (*leaf*) menspesifikasikan nilai yang diharapkan dari kelas variabel (variabel yang akan di prediksi). Aspek penting dari prosedur untuk membangun pohon keputusan adalah pemisahan kriteria (*split criterion*) termasuk kriteria untuk membuat cabang dan kriteria terakhir (*stop criterion*), kriteria yang digunakan untuk menghentikan pencabangan. Pohon keputusan dibuat menggunakan himpunan dari data yang digunakan sebagai data pembelajaran (*training dataset*). Himpunan yang berbeda yang disebut *test dataset* digunakan untuk melakukan pengujian untuk mengecek model. Pohon keputusan menawarkan banyak keuntungan, antara lain:

1. Fleksibilitas untuk berbagai tugas data mining, seperti prediksi, klasifikasi, regresi, clustering dan seleksi fitur.
2. Cukup jelas dan mudah diikuti (ketika dipadatkan).
3. Fleksibilitas dalam menangani berbagai input data: nominal, numerik dan tekstual.
4. Adaptasi di *dataset* pengolahan yang mungkin memiliki kesalahan atau nilai-nilai yang hilang.
5. Kinerja prediktif tinggi untuk upaya komputasi yang relatif kecil.
6. Tersedia dalam berbagai paket data mining melalui berbagai platform
7. Berguna untuk *dataset* besar (dalam kerangka *ensemble*).

*Cross Validation* merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan *dataset* tertentu. Data yang digunakan dalam proses pembangunan model disebut data latih/training, sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai data test.

## 2.7 *Confusion Matrix*

*Confusion Matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. *Confusion Matrix* adalah media yang berguna untuk menganalisis seberapa baik *classifier* dapat mengenali tupel dari kelas yang berbeda (Sumiah & Mirantika, 2020.)

*Confusion Matrix* juga dikenal sebagai *error matrix* merupakan tata letak tabel khusus yang memungkinkan visualisasi kinerja suatu algoritma, biasanya pembelajaran yang diawasi (dalam pembelajaran tanpa pengawasan biasanya disebut matriks yang cocok). Setiap baris matriks mewakili instance di kelas aktual sementara setiap kolom mewakili instance di kelas yang diprediksi, atau sebaliknya kedua varian ditemukan dalam literatur. Nama berasal dari fakta bahwa memudahkan untuk melihat apakah sistem ini membingungkan dua kelas (yaitu biasanya salah memberi label satu sama lain). Dalam penelitian ini menggunakan perhitungan akurasi dimana jumlah data yang telah diklasifikasikan secara benar dan salah dibagi dengan total sampel data testing (Baskoro dkk., 2021).

Evaluasi menggunakan *Confusion Matrix* akan menghasilkan nilai *accuracy*, *precision*, dan *recall*. *Confusion matrix* merupakan tabel matrix yang terdiri dari dua kelas, yaitu kelas yang satu dianggap sebagai positif dan yang lainnya negative.

Klasifikasi yang benar		
	+	-
+	<i>True Positive</i>	<i>False Negative</i>
-	<i>False Positive</i>	<i>True Negative</i>

Keterangan :

1. True positive : jumlah record positif yang diklasifikasikan sebagai positif,
2. False positive : jumlah record negatif yang diklasifikasikan sebagai positif,
3. False negative : jumlah record positif yang diklasifikasikan sebagai negatif,
4. True negative : jumlah record negatif yang diklasifikasikan sebagai negatif.

Tabel 2.1 *Confusion Matrix*

Kelas	Prediksi Yes	Prediksi No	Total
Aktual Yes	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>	<i>Positive (P)</i>
Aktual No	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>	<i>Negative (N)</i>
Total	P'	N'	P+N

Rumus mencari Akurasi adalah:

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.3)$$

Rumus mencari *Sensitivity* atau *Recall* adalah:

$$sensitivity = \frac{TP}{TP+FN} \quad (2.4)$$

Rumus mencari nilai *Precision* adalah:

$$precision = \frac{TP}{TP+FP} \quad (2.5)$$

## 2.8 Particle Swarm Optimization (PSO)

Salah satu metode seleksi atribut yang cukup terkenal adalah *Particle Swarm Optimization* (PSO) algoritma berbasis populasi yang memiliki sejumlah  $n$  particle. Setiap hipotesis solusi dari permasalahan akan diwakili oleh partikel-

partikel tersebut. Masing-masing partikel mengubah posisi terhadap waktu. Pada sistem PSO, partikel terbang mengitar ruang pencarian multidimensional dan menyesuaikan posisinya berdasarkan pengalaman pribadinya dan pengalaman partikel kelompok (Firdaus Mahmudy, 2015).

*Particle Swarm Optimization* (PSO) merupakan suatu algoritma yang digunakan untuk menyelesaikan masalah optimasi. Kelebihan PSO mudah untuk diimplementasikan serta hanya membutuhkan sedikit parameter. Selain itu, PSO juga lebih efisien dalam hal komputasi dan PSO bersifat fleksibel dalam menjaga keseimbangan antara pencarian global dan lokal optimal (Sasongko, 2016).

PSO pertama kali dikenalkan oleh Kennedy dan Eberhart pada tahun 1995 yang merupakan teknik pencarian dengan proses algoritma yang diinspirasi oleh perilaku sosial Binatang seperti sekelompok burung atau ikan. PSO didasarkan pada penelusuran algoritma dan diawali dengan suatu populasi yang random atau disebut dengan *particle*. Berbeda dengan teknik komputasi *evolusioner* lainnya, setiap *particle* di dalam PSO juga berhubungan dengan suatu *velocity*. *Particle-particle* tersebut bergerak melalui penelusuran ruang dengan *velocity* yang dinamis yang disesuaikan menurut perilaku historisnya. Oleh karena itu, *particle-particle* mempunyai kecenderungan untuk bergerak ke area penelusuran yang lebih baik setelah melewati proses penelusuran (Unler & Murat, 2010).

*Particle Swarm Optimization* (PSO) mempunyai kesamaan dengan *genetic algorithm* yang mana dimulai dengan suatu populasi yang random dalam bentuk matriks. Namun PSO tidak memiliki operator evolusi yaitu *crossover* dan mutasi seperti yang ada pada *genetic algorithm*. Baris pada matriks disebut *particle* atau

dalam genetic algorithm sebagai kromosom yang terdiri dari nilai suatu variable. Setiap particle berpindah dari posisinya semula keposisi yang lebih baik dengan suatu *velocity*. Pada algoritma PSO vektor *velocity* diupdate untuk masing-masing particle kemudian menjumlahkan vektor *velocity* tersebut ke posisi particle. Update velocity dipengaruhi oleh kedua solusi yaitu *global best* yang berhubungan dengan biaya yang paling rendah yang pernah diperoleh dari suatu particle dan solusi *local best* yang berhubungan dengan biaya yang paling rendah pada populasi awal. Jika solusi local best mempunyai suatu biaya yang kurang dari biaya solusi global yang ada, maka solusi local best menggantikan solusi *global best* (Lestari & Nasir, 2021).

## 2.9 Rapidminer

Rapidminer yang dapat dijalankan pada berbagai sistem operasi dan bersifat terbuka (*open source*). Rapidminer dapat menjadi sebuah solusi 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 juga menempati peringkat pertama sebagai *software* data mining pada polling Kdnuggets, sebuah portal data mining pada 2010-2011.

Rapidminer memiliki beberapa fitur, diantaranya adalah:

1. Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis text.
2. Memiliki kurang lebih 500 operator data mining. Termasuk operator input, output, data preprocessing dan visualisasi.
3. Bentuk grafis yang canggih, seperti tree chart dan 3d scatter plots.
4. Mengintegrasikan proyek data mining Weka dan statistika R.

5. Operator-operator dalam proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI.
6. Menyediakan machine learning dan prosedur data mining.

### **2.10 *State of The Art (SoTA)***

Penelitian terdahulu sangat berperan penting pada penelitian, dimana dengan adanya penelitian terdahulu sangat berpengaruh dalam memperkaya teori dan referensi wawasan pengetahuan bagi penelitian ini. Ulasan dari penelitian terkait, dilakukan dengan maksud dan tujuan untuk menganalisis penelitian yang telah dilakukan sebelumnya. Penelitian terdahulu dapat dilihat pada tabel berikut:

Tabel 2.2 *State of The Art (SoTA)*

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
1	Baskoro, Sriyanto, Lilis Setya Rini	Prediksi Penerimaan Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu	Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021 ISSN: 2598- 0256, E-ISSN: 2598-0238	Masih adanya pemberian beasiswa yang kurang tepat, sehingga menyebabkan kurang maksimalnya penggunaan dana beasiswa tersebut.	Teknik data mining untuk melihat nilai akurasi yang lebih baik agar dapat membantu memprediksi penerimaan beasiswa dengan menggunakan 3 algoritma yaitu <i>Random Forests, Naïve Bayes dan K- NN.</i>	Data mahasiswa dari 4 fakultas yang nantinya akan berkaitan dengan penentuan penerimaan beasiswa.	Hasil akhir penelitian ini menunjukkan bahwa dari 3 algoritma yang digunakan, algoritma <i>Random Forest</i> termasuk algoritma yang akurat untuk memprediksi penerimaan beasiswa.	Untuk penelitian berikutnya diperlukan atribut yang mempengaruhi pengolahan data yang digunakan, serta menggunakan lebih banyak data dapat meningkatkan kualitas hasil penelitian.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
2	Rasi Rahwali, Seng Hansun, Yustinus Widya Wiratama	Prediksi Kelayakan Masuk Penjurusan IPA Siswa Sekolah Menengah Atas Menggunakan C4.5 (Studi Kasus: SMA Tarakanita Gading Serpong)	Jurnal Telematika Vol. 10 No. 2 Agustus 2017 ISSN: 1979 – 925X e-ISSN: 2442 – 4528	Hasil psikotes dan nilai rata-rata siswa tidak sinkron. Sering kali nilai Siswa kurang dari kriteria ketuntasan minimal, tapi memiliki bakat atau minat di Ilmu Pengetahuan Alam.	Membangun sautu sistem pendukung keputusan untuk membantu memprediksi kelayakan siswa dalam mengambil penjurusan Ilmu Pengetahuan Alam menggunakan algortima C4.5	Data yang digunakan merupakan dataset siswa, nilai akademik, hasil psikotest dan minat siswa.	Berdasarkan hasil uji coba validasi menggunakan <i>cross validation</i> , Algoritma C4.5 ini berhasil diimplementasik an dengan sangat baik. Selain itu hasil pengujian Algoritma C4.5 memiliki nilai tingkat error sebesar 16,65%	Pengembangan lebih lanjut dapat dilakukan untuk menyempurnakan sistem penjurusan ini dengan menambahkan atribut lain seperti Bahasa Indonesia atau Bahasa Inggris yang dapat memepengaruhi hasil penjurusan IPA di masa yang akan datang.



No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
3	Senna Hendrian	Algoritma Klasifikasi Data Mining untuk Memprediksi siswa dalam menerima Bantuan Dana Pendidikan	Faktor Excata 11 (3): 266- 274, 2018 p-ISSN:1979- 276X e-ISSN: 2502- 339X DOI:10.30998/ faktorexcata.v 11i3.2777	Ekonomi yang terbatas bagi sebagian orang tua menjadi faktor penghambat dalam kegiatan wajib belajar di sekolah.	Klasifikasi data mining menggunakan Decision Tree untuk memprediksi siswa dalam menerima bantuan dana pendidikan.	Sampel data yang bersumber dari Sekolah Bina Bangsa Mandiri dengan jumlah kasus sebanyak 254 <i>record</i> .	Hasil pengujian dilakukan test <i>Cros Validation</i> menghasilkan nilai akurasi sebesar 98,80%.	Dapat dilakukan komparasi dari beberapa algoritma klasifikasi data mining, missal Algoritma C4.5 dengan <i>Naïve</i> <i>Bayes</i> atau yang lainnya, sehingga dapat diketahui algoritma yang terbaik untuk memprediksi siswa menerima bantuan dana pendidikan.
4	M. A. Abdillah,	Implementasi Decision Tree	Vol. XV Nomor 2 Juli	Pendidikan karakter tidak hanya	Pengujian Decision Tree	Data siswa Universitas	Hasil akurasi sebesar 60,91%	Penentuan atribut yang memang

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
	Arief Setyanto, Sudarmawan	Algoritma C4.5 untuk Memprediksi Kesuksesan Pendidikan Karakter	2020 – Jurnal Teknologi Informasi ISSN: 1907 – 2430	berpotensi pada aspek kognitif saja, tetapi lebih berorientasi pada proses pembinaan potensi yang ada dalam peserta didik.	C4.5 untuk mendapatkan model prediksi kesuksesan pendidikan karakter.	Residence Yogyakarta terhitung dari tahun 2014 sampai 2015.	menunjukkan bahwa algoritma Decision Tree C4.5 layak digunakan untuk melakukan prediksi tingkat kesuksesan pada pendidikan karakter.	benar mempengaruhi kesuksesan pendidikan karakter sangat mempengaruhi tingkat akurasi yang lebih baik serta efisiensi.
5	Joy Nashar Utamajaya, Andi Mentari A.P, Siti Masnunah	Penerapan Algoritma <i>Naive Bayes</i> untuk Penentuan Calon Beasiswa PIP pada SDN 023 Penajam	Jurnal Sistem Informasi, Vol. 3 No. 1, November 2019 e-ISSN: 2597-3827	Program Bantuan Siswa Miskin dengan ditandai pemberian Kartu Indonesia Pintar kepada anak usia sekolah yang berasal	Analisis data seleksi kelulusan beasiswa melalui data siswa SD kemudian diolah	Kriteria penentuan penerimaan beasiswa antara lain: pekerjaan orang tua, penghasilan	Penerapann Algoritma <i>Naive Bayes</i> dalam penelitian ini menunjukkan tingkat akurasi sebesar 97,2%.	Dengan menambahkan variabel lain dalam proses mining dapat menambah tingkat akurasi yang lebih baik.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
				dari keluarga kurang mampu untuk menjamin agar seluruh anak usia sekolah dari keluarga kurang mampu terdaftar sebagai penerima bantuan sampai lulus jenjang pendidikan menengah.	dan di uji menggunakan Algoritma <i>Naïve Bayes</i> .	orang tua, jumlah tanggungan, nilai akademik, jarak dari rumah ke sekolah.	Hal tersebut dapat sangat membantu SDN 023 Penajam dalam menyeleksi siswa penerima bantuan Program Indonesia Pintar	
6	Varonica Andriyana & Yusuf Sulisty Nugroho, S.T., M.Eng	Perbandingan 3 Metode dalam Data Mining untuk Memprediksi Penerima	Naskah Publikasi, Universitas Muhammadiyah Surakarta – Maret 2015	Kurang tepatnya penyaluran beasiswa terhadap siswa.	Analisis perbandingan 3 metode dalam prediksi siswa penerima beasiswa,	Penentuan sampel dengan menyeleksi data keseluruhan untuk mendapatkan	Dari perhitungan 3 metode, menghasilkan nilai <i>precision</i> ,	Komparasi dari beberapa algoritma klasifikasi data mining, seperti Regresi Linear dengan Algoritma

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
		Beasiswa berdasarkan Prestasi di SMA Negeri 6 Surakarta			diantaranya <i>Naïve Bayes</i> , Algoritma ID3 dan Regresi Linear.	atribut yang relevan, diantaranya: nilai rata-rata, ekstrakurikuler, semester, jumlah tanggungan orang tua dan penghasilan orang tua	<i>recall</i> dan <i>accuracy</i> . Berdasarkan nilai <i>precision</i> , algoritma ID3 memiliki nilai lebih baik daripada algoritma yang lain, sedangkan berdasarkan <i>recall</i> dan <i>accuracy</i> metode Regresi Linear memiliki nilai lebih baik.	C4.5 atau yang lainnya, sehingga dapat diketahui algoritma yang terbaik untuk memprediksi siswa menerima bantuan dana pendidikan.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
7	Dita Merawati & Rino	Penerapan Data Mining Penentu Minat dan Bakat Siswa SMK dengan Metode C4.5	Jurnal Algor – Vol. 1 No. 1, 2019	Prestasi merupakan perwujudan dari bakat dan kemampuan. Salah satu cara mengetahui minat dan bakat adalah dengan melakukan psikotest.	Penerapan metode C4.5 sebagai pengambil keputusan bakat dan minat siswa berbasis android.	Data primer yang bersumber dari sekolah melalui pelaksanaan test minat dan bakat dengan peserta siswa kelas X (sepuluh).	Metode C4.5 memiliki keakuratan sebesar 82,7% dalam pengujian bakat menggunakan <i>K-Fold Cross Validation</i> dna nilai keakuratan minat sebesar 90,23%	Pengujian menggunakan penggabungan beberapa metode untuk membandingkan tingkat keakuratan, sehingga dapat diketahui metode mana yang lebih baik pada kasus yang sama.
8	Fathur Rahman & Muhammad Iqbal Firdaus	Penerapan Data Mining metode Naïve Bayes untuk Prediksi Hasil Belajar	AI Ulum Sains dan Teknologi Vol. 1 No. 2 Mei 2016	Sulitnya menentukan faktor atau variabel yang mempengaruhi hasil belajar siswa.	Penerapan metode Klasifikasi Naïve Bayes dalam	Data diambil dari repository dataset terbuka ( <a href="http://archive.ics.uci.edu">http://archive.ics.uci.edu</a> )	Hasil dari penelitian ini di uji menggunakan confusion	Banyaknya input yang tidak terlalu relevan, dan algoritma yang diujikan tidak

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
		Siswa Sekolah Menengah Pertama (SMP)			memprediksi hasil belajar siswa.	merupakan data sekunder yang di ambil dari dua sekolah yang berasal dari umpan balik siswa dengan beberapa faktor internal dan eksternal. Data tersebut memiliki 263 record dan terdiri dari 12 atribut.	matrix guna mendapatkan nilai akurasi. Pada metode Naïve Bayes menghasilkan nilai akurasi sebesar 56,79%, presisi 62,80% dan Recall 72,56%. Metode Naïve Bayes memiliki kinerja yang baik.	terlalu sensitif, penulis menyarankan agar peneliti selanjutnya dapat membandingkan Naïve Bayes dengan algoritma lainnya dalam kasus yang sama guna mendapatkan nilai akurasi, presisi dan recall yang lebih baik.
9	Koko Handoko	Penerapan Data Mining dalam	TEKNOSI, Vol. 02, No.	Bagaimana ekstrasi data mining yang	Penerapan data mining dengan	Variabel utama dalam pengujian	Algoritma <i>K-Means</i>	Metode <i>K-Means</i> memiliki tingkat

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
		Meningkatkan Mutu Pembelajaran pada Intansi Perguruan Tinggi menggunakan metode <i>K-Means Clustering</i>	03, Desember 2016	dihasilkan dapat memberikan sebuah pengetahuan baru terhadap intansi Akademi Komunitas Solok Selatan dalam meningkatkan mutu pendidikan	metode <i>clustetring</i> dalam meningkatkan mutu pembelajaran.	penerapan metode <i>K-Means Clustering</i> , diantaranya: nilai IP, jarak tempuh, jumlah kehadiran dan penghasilan orang tua.	<i>Clustering</i> sangat membantu pengelompokan data mutu pembelajaran siswa. Metode ini menunjukkan bahwa karakteristik siswa dari <i>cluster</i> nilai IP dan kehadiran tergolong sedang dengan nilai IP rata-rata 2,50 s.d 3,00	sensitive terhadap pemilihan pusat awal dan perhitunhn solusi local untuk mencapai kondisi optimal, sehinggal membutuhkan analisis cluster yang merupakan Teknik multivariat untuk membantu mengelompokan objek yang paling dekat kesamaanya dengan objek lainnya.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
							dan nilai kehadiran 85 s.d 90.	
10	Heri Susanto	Data Mining untuk Memprediksi Prestasi Siswa berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa Lalu	Jurnal Pendidikan Vokasi, Vol 4, Nomor 2, Juni 2014	Masih banyaknya peserta didik yang mementingkan kompetensi mata pelajaran produktif dari pada mata pelajaran adaptif dan normatif.	Prediksi prestasi belajar siswa menggunakan metode algoritma J48 yang kemudian hasil analisis di bandingkan dengan <i>CHAID</i> dan regresi ganda.	Penelitian ini menggunakan pendekatan kuantitatif kepada siswa kelas X SMK Negeri 4 Surakarta tahun pelajaran 2013/2014 sejumlah 416 siswa.	Analisis prediksi menggunakan algoritma J48 memiliki tingkat akurasi sebesar 95,7%, sedangkan metode <i>CHAID</i> memiliki tingkat akurasi sebesar 82,1% dan regresi ganda sebesar 90,6%.	Ketersediaan data sangat mempengaruhi tingkat akurasi dalam melakukan prediksi sehingga diperlukan penyimpanan data dalam waktu yang lama untuk menjamin ketersediaan data pada waktu yang diperlukan.



No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
11	Andi Taufik	Optimasi Particle Swarm Optimization (PSO) sebagai seleksi fitur pada analisis sentiment review hotel berbahasa Indonesia menggunakan algoritma Naïve Bayes	Jurnal Teknik Komputer, Vol. III, No.2, Agustus 2017	Semakin rendah peringkat hotel, semakin besar persentasi sentiment negative.	Peningkatan akursai Klasifikasi <i>Naïve Bayes</i> dengan <i>Particle Swarm Optimization</i> sebagai metode pemilihan fitur pada komentar dari review hotel berbahasa Indonesia.	Membantu mengambil keputusan saat ingin melakukan pemesanan hotel yang sesuai dengan keinginan agar lebih efisien dan efektif dibandingkan jika harus membaca review.	Data review hotel berbahasa Indonesia diklasifikasi dengan baik kedalam review positif maupun review negative. Akurasi model Naïve Bayes sebelum menggunakan PSO mencapai 90,50% sedangkan setelah menggunakan	Untuk mendapatkan model yang terbaik peneliti menyarankan penyesuaian dalam beberapa nilai parameter guna mendapatkan hasil akuras yang tinggi.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
							PSO akurasi-nya meningkat sebesar 6,42% menjadi 96,62%.	
12	Desi Lestari & Muhammad Nasir	Penerapan metode C4.5 berbasis Particle Swarm Optimization (PSO) untuk memprediksi penjualan obat pada apotek bunda Azka	Jurnal Pengembangan Sistem Informasi dan Informatika <i>e-ISSN: 2746-1335 Vol. 2, July 2021</i>	Apotek belum mengelompokan penjualan obat yang sering terjual sehingga laba dari penjualan obat tidak terserap dengan baik	mengklasifikasi kan barang-barang yang sering terjual berdasarkan data dari tahun 2017-2019 dengan penerapan Particle Swarm Optimization (PSO)	Sampel data yang digunakan adalah jumlah 65 record transaksi obat di apotek bunda azka.	Hasil pengujian di dapat akurasi sebesar 70.95% dan setelah menggunakan Particle Swarm Optimization meningkat menjadi 78.10%, untuk class recall 72.50%	Dengan menambah data record kemungkinan besar akan mempengaruhi nilai akurasi pada proses algoritma.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
							meningkat menjadi 78.33%, sedangkan precision 77.92% dan meningkat menjadi 80.33%. Dari	
13	Wayan Firdaus Mahmudy	<i>Improved Particle Swarm Optimization</i> untuk menyelesaikan permasalahan <i>Part Type Selection</i> dan	Brawijaya University, August 2015 ( <i>page : 1003-1008</i> )	<i>Part type selection</i> dan <i>machine loading</i> merupakan permasalahan utama dalam perencanaan produksi di industri manufaktur yang mengadopsi Flexible	Particle swarm optimization (PSO) dipilih untuk menyelesaikan kedua permasalahan tersebut karena	Perencanaan produksi pada FMS dilakukan untuk meningkatkan utilisasi sumber daya sistem yang meliputi	Optimasi permasalahan part type selection dan machine loading telah diselesaikan secara simultan	selain memaksimalkan throughput sistem dan menjaga keseimbangan beban mesin, IPSO juga harus meminimumkan

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
		<i>Machine Learning pada Flexible Manufacturing System (FMS)</i>		Manufacturing System (FMS)	telah terbukti berhasil digunakan pada berbagai permasalahan kombinatorial kompleks.	mesin dan tools, memaksimalkan kuantitas hasil produksi (throughput), dan menurunkan biaya produksi	dengan improved particle swarm optimization (IPSO)	total keterlambatan (tardiness) dari semua part type. Untuk menghasilkan solusi yang baik dari permasalahan kompleks seperti itu, IPSO yang lebih baik perlu dikembangkan, misalnya dengan melakukan hibridisasi dengan metode heuristik lain seperti simulated

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
								annealing, tabu search, dan variable neighborhood search (VNS)
14	Unler & Murat	<i>A discrete particle swarm optimization method for feature selection in binary classification problems</i>	European Journal of Operational Research (2010) 206(3) 528-539	Keakuratan prediksi subset yang dipilih bergantung pada ukuran subset serta fitur yang dipilih. Akurasi prediksi bukanlah fungsi monoton dari subset fitur sehubungan dengan penyertaan set.	Mengembangkan algoritma optimasi gerombolan partikel diskrit (PSO) yang dimodifikasi untuk masalah pemilihan subset fitur.	Mewujudkan prosedur pemilihan fitur adaptif yang secara dinamis memperhitungkan relevansi dan ketergantungan fitur yang termasuk dalam subset fitur	erangkaian percobaan yang dilakukan untuk mengevaluasi efektivitas metode yang diusulkan dan membandingkan nya dengan metode lain. Sepuluh dari 11 kumpulan data	Algoritma PSO yang dimodifikasi untuk masalah pemilihan subset fitur adalah algoritma PSO <i>diskrit</i> di mana subset fitur dikodekan dalam string biner. prosedur pemilihan fitur adaptif yang secara

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
							yang digunakan dalam eksperimen semuanya diperoleh dari repositori data pembelajaran mesin terkenal di Repositori Pembelajaran Mesin UCI Universitas California.	dinamis memperhitungkan relevansi dan ketergantungan fitur yang akan dimasukkan ke dalam subset fitur.
14	Theopilus Bayu Sangsoko	Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus	Jurnal Teknik Informatika dan Sistem Informasi	SVM memiliki keterbatasan ketika jumlah atribut yang digunakan relatif	pendekatan hybrid yang mengkombinasikan antara	pembentukan model klasifikasi yaitu nilai peserta	Tingkat akurasi yang paling besar sebesar 99.30%	Hasil uji coba pemilihan kernel yang sesuai dapat digantikan dengan

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
		Klasifikasi Jalur Minat SMA)	Volume 2 Nomor 2 Agustus 2016 e-ISSN : 2443-2229	banyak yang mengakibatkan beban komputasi menjadi berat dan akurasi menjadi kurang akurat.	support vector machine (SVM) klasifier dengan metode particle swarm optimization (PSO) dengan jumlah dataset sebanyak 281 records dan 11 atribut.	didik saat mendaftar di salah satu SMA di Jawa Tengah pada tahun pelajaran 2013-2014 yang meliputi nama, nilai UN pada jenjang sebelumnya, nilai rata-rata raport, dan nilai tes psikologi peminatan sebanyak 280 siswa	diperoleh ketika menerapkan PSO-SVM dengan parameter C (pinalti) sebesar 0.0 dengan menggunakan kernel anova.	automatic kernel selection sehingga tidak memerlukan waktu yang banyak untuk pemilihan kernel yang tepat.

### 2.9.1 Matriks Penelitian

Tabel 2.3 Matriks Penelitian

No	Nama Peneliti	Metode									Objek Penelitian				Optimalisasi Algoritma			
		Naïve Bayes	Random Forest	K-Neast Neighbor	Algoritma ID3	K-Means	Decision Tree	Algoritma C4.5	Regresi Linear	Algoritma J48	Lainnya	Beasiswa	Penjurusan Kelas	Prestasi	Minat dan Bakat	Lainnya	Particle Swarm Optimization (PSO)	Forward Selection
1	Baskoro, Sriyanto, Lilis Setya Rini (2021)	✓	✓	✓							✓							
2	Rasi Rahwali, Seng Hansun, Yustinus Widya Wiratama (2017)						✓					✓						
3	Senna Hendrian (2018)						✓				✓							
4	M. A. Abdillah, Arief Setyanto, Sudarmawan (2020)						✓								✓			
5	Joy Nashar Utamajaya, Andi Mentari A.P, Siti Masnunah (2019)	✓									✓							



No	Nama Peneliti	Metode									Objek Penelitian					Optimalisasi Algoritma		
		Naïve Bayes	Random Forest	K-Neast Neighbor	Algoritma ID3	K-Means	Decision Tree	Algoritma C4.5	Regresi Linear	Algoritma J48	Lainnya	Beasiswa	Penjurusan Kelas	Prestasi	Minat dan Bakat	Lainnya	Particle Swarm Optimization (PSO)	Forward Selection
6	Varonica Andriyana & Yusuf Sulisty Nugroho, S.T., M.Eng (2015)	✓			✓				✓			✓		✓				
7	Dita Merawati & Rino (2019)							✓					✓	✓				
8	Fathur Rahman & Muhammad Iqbal Firdaus (2016)	✓											✓					
9	Koko Handoko (2016)					✓							✓					
10	Heri Susanto (2014)								✓	✓			✓					
11	Andi Taufik (2017)	✓													✓	✓		
12	Desi Lestari & Muhammad Nasir (2021)							✓							✓	✓		

No	Nama Peneliti	Metode									Objek Penelitian				Optimalisasi Algoritma			
		Naïve Bayes	Random Forest	K-Neast Neighbor	Algoritma ID3	K-Means	Decision Tree	Algoritma C4.5	Regresi Linear	Algoritma J48	Lainnya	Beasiswa	Penjurusan Kelas	Prestasi	Minat dan Bakat	Lainnya	Particle Swarm Optimization (PSO)	Forward Selection
13	Unler & Murat (2010)														✓	✓		
14	Theopilus Bayu Sangsoko (2016)									✓		✓		✓		✓		
15	Fahmi Ahmad Fauzi, (2022)					✓	✓				✓					✓		