BAB II

LANDASAN TEORI

2.1 Akreditasi

Akreditasi menurut Kamus Besar Bahasa Indonesia adalah pengakuan terhadap lembaga pendidikan yang diberikan oleh badan yang berwenang setelah dinilai bahwa lembaga itu memenuhi syarat kebakuan atau kriteria tertentu. Menurut Peraturan Menteri Pendidikan dan Kebudayaan Nomor 13 tahun 2018 tentang Badan Akreditasi Nasional Sekolah/Madrasah dan Badan Akreditasi Nasional Pendidikan Anak Usia Dini dan Pendidikan Nonformal, pasal 1, bahwa akreditasi adalah suatu kegiatan penilaian kelayakan satuan Pendidikan dasar dan menengah, dan satuan Pendidikan anak usia dini dan Pendidikan nonformal berdasarkan kriteria yang telah ditetapkan untuk memberikan penjaminan mutu pendidkan.

2.2 Data Mining

2.2.1 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 computer (machine learning) untuk menganalisis dan mengekstrasi 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 penelitan 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 sebaga 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 mengelompokan data yang memiliki kemiripan.

4) *Estimation*

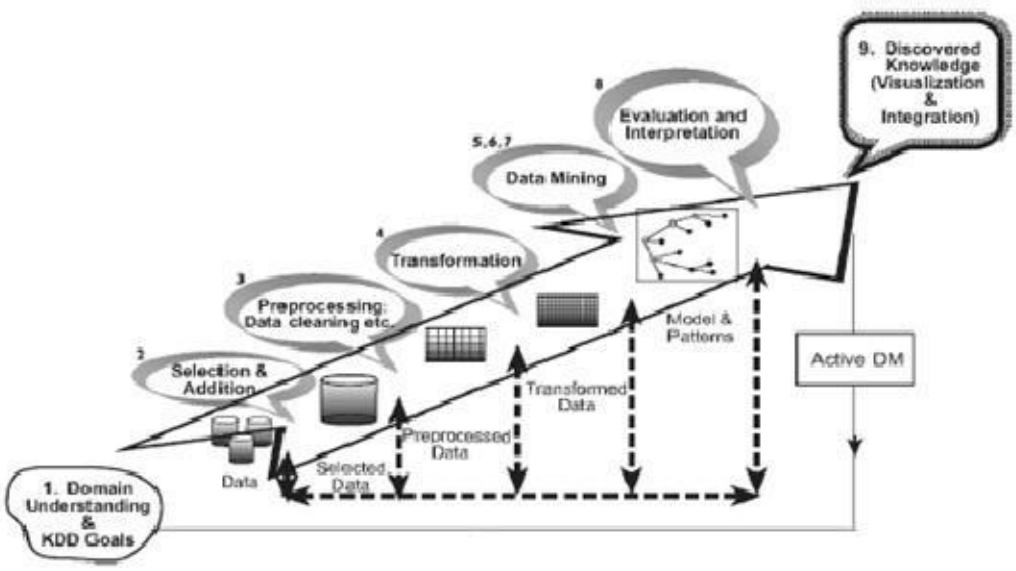
Metode *Estimation* sama halnya dengan metode *Classification*, namun variabel tujuan pada metode *estimation* lebih ke arah numerik daripada kategori.

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

2.3.1 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.

2.3 Klasifikasi

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Algoritma klasifikasi yang banyak digunakan secara luas , yaitu *Decision/classification trees, Bayesian classifiers/ Naïve Bayes classifier, Neural Network, Analisa Statisti, Algoritma Genetika, Rough Sets, K-Nearest Neighbor, Metode Rule Based, Memory based Reasoning, dan Support Vector Machine* (Annur, 2018).

Klasifikasi adalah proses analisis data untuk melakukan pengelompokan data agar menghasilkan kelas baru berdasarkan aturan pengelompokan yang telah dibuat (Sanjaya, Setiyati, & Budianto, 2020)

2.4 Algoritma Support Vector Machine (SVM)

SVM dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory. Konsep dasar SVM adalah kombinasi dari teori-teori komputasi yang telah ada puluhan tahun sebelumnya, seperti *margin hyperlane, kernel,* dan konsep-konsep pendukung yang lain. Akan tetapi hingga tahun 1992, belum pernah ada upaya merangkaikan komponen-komponen tersebut.

SVM adalah metode *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperlane* terbaik yang memisahkan dua buah class pada input space.

Prinsip dasar SVM adalah *linear classifier,* dan selanjutnya dikembangkan agar dapat bekerja pada *problem non-linear* dengan konsep kernel *trick ipada ruang kerja berdimensi tinggi.* hal ini meningkatkan minat penelitian pada bidang *pattern recognition* untuk mengetahui potensi kemampuan SVM secara teoritis maupun aplikasi.

Menurut (Perdana & Furqon, 2018) , *Support Vector Machine* adalah algoritma *supervised* yang berupa klasifikasi dengan cara membagi data menjadi dua kelas menggunakan garis vector yang disebut hiperlane.

Hyperlane pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur margin *hyperlane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperlane* tersebut dengan data terdekat dari masing-masing *class*. Subnet dari data *training set* yang paling dekat ini disebut sebagai *support vector* (Nugroho, 2008).

Data yang tersedia dinotasikan sebagai $\vec{x} \in$ Rd sedangkan label masing-masing dinotasikan yi $\in$ {-1 + 1} untuk i = 1,2,..., l, yang mana l adalah banyaknya data. Diasumsikan kedua class -1 dan +1 dapat terpisah secara sempurna oleh hyeprlane berdimensi d, yang didefinisikan :

$\vec{w} . \vec{x} + b = 0$…….. (1)

w = Bidang normal

b = Posisi bidang relatif terhadap pusat koordinat

Sebuah *pattern* $\vec{x}i$ yang termasuk class -1 (sampel negatif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan :

$\vec{w} . \vec{x}i + b \leq -1$……... (2)

Sedangkan pattern $\vec{x}$ yang termasuk *class* +1 (sample positif) :

$\vec{w} . \vec{x}I + b \geq +1$…….. (3)

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara hyperlane dan titik terdekatnya, yaitu $1/\|\vec{w}\|$. Hal ini dapat dirumuskan sebagai Quadratic Programming (QP) problem, yaitu mencari

titik minimal persamaan (4), dengan memperhatikan constraint persamaan (5).

$$\frac{min}{w}\tau(w) = \frac{1}{2} \parallel w \parallel \ldots\ldots\ldots\ldots \text{(4)}$$

$$y\text{i}(\vec{\text{x}}\text{i} \,.\, \vec{w} + b) - 1 \geq 0, \forall i \ldots\ldots\ldots\ldots\ldots\ldots \text{(5)}$$

problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya Language Multipler sebagaimana ditunjukan pada persamaan(6).

$$L(w, b, a) = \frac{1}{2} \parallel w \parallel^2 - \sum_{i=1}^{1} a_i \left(y_i \left((x_i \,.W + b) - 1\right)\right) \quad (i = 1,2, \ldots l) \text{ (6)}$$

αi adalah Lagrange multipliers, yang bernilai nol atau positif ($\alpha i \geq 0$). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan L terhadap $\vec{w}$ dan b, dan memaksimalkan L terhadap αi . Dengan memperhatikan sifat bahwa pada titik optimal gradient $L = 0$, persamaan (6) dimodifikasi sebagai maksimalisasi problem yang hanya mengandung saja αi sebagaimana persamaan (7).

Maximize:

$$\sum_{i=1}^{l} ai - \frac{1}{2} \sum_{i,j=1}^{l} a_i \;\; a_j y_i y_j x_i x_j \ldots\ldots\ldots\ldots \text{(7)}$$

Subject to :

$$a_i \geq 0 \; (i = 1,2, \ldots, l) \; \sum_{i=1}^{l} a_i y_i = 0 \ldots\ldots\ldots \text{(8)}$$

Hasil perhitungan ini diperoleh αi yang kebanyakan bernilai positif. Data yang berkolerasi dengan αi positif inilah yang disebut sebagai support vector (Nugroho, 2008).

2.5 RapidMiner

Rapid Miner adalah salah satu aplikasi *opensource* yang dapat digunakan untuk melakukan proses data mining. Salah satu metode data mining adalah menggunakan regresi linier. Rapid Miner menyeiakan prosedur data mining dan *machine learning* termasuk : ETL (*Extraction, Transformation, Loading*), data *preprocessing, visualisasi, modelling,* dan evaluasi (Imelda & Muhammad, 2015).

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.

2.6 *State Of The Art*

Penelitian terdahulu ini menjadi salah satu acuan dalam melakukan penelitian sehingga dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan. Ulasan penelitian terkait, dilakukan dengan maksud untuk menganalisis penelitian yang telah dilakukan sebelumnya. Judul penelitian terdahulu dapat dilihat sebagai berikut :

1. **Perbandingan Analisis Klasifikasi Menggunakan Metode K-Nearst Neighbor (K-NN) dan Multivariate Adaptive Regression Spline (MARS) Pada Data Akreditasi Sekolah Dasar Negeri Di Kota Semarang**

   Nama penulis dari jurnal tersebut adalah Bisri Merluarini, Diah Safitri, dan Abdul Hoyyi dengan terbitan Jurnal Gaussian, Volume 3, Nomor 3, Tahun 2014.

   Penelitian yang dilakukan pada jurnal tersebut adalah untuk membandingkan metode K-NN dan MARS dalam pengklasifikasian data

akreditasi Sekolah Dasar (SD) yang ada di Kota Semarang. Data yang digunakan didalam penelitian ini bersumber dari situs resmi Badan Akreditasi Nasional http//:www.ban-sm.or.id.

Kemudian dilakukan tahapan analisis data dengan menggunakan metode K-Nearst Neighbor (K-NN) dan metode MARS. Setelah hasil klasifikasi dari metode K-NN dan MARS didapatkan, langkah selanjutnya adalah membuat matriks konfusi serta menghitung akurasi klasifikasi dari kedua metode tersebut dengan menggunakan uji statistik Press's Q, APER, *sensitivity, dan specificity* untuk menguji metode manakah yang lebih baik digunakan dalam klasifikasi data akreditasi Sekolah Dasar Negeri di Kota Semarang.

Hasil dari penelitiaan ini berdasarkan perhitungan statistik *Press's Q* untuk metode K-NN adalah 133,929 dan metode MARS adalah 152,381 lebih besar dari $X^2\,(0{,}05;\,1) = 3{,}84$ maka dapat dikatakan pengklasifikasian SDN di Kota Semarang berdasarkan akreditasi menggunakan metode MARS dan Metode K-NN signifikan secara statistik. Sedangkan berdasarkan perhitungan APER, *Specificity*, dan *sensitivity* menunjukan bahwa pengklasifikasian menggunakan metode MARS lebih baik dibandingkan dengan Metode K-NN dalam mengklasifikasikan Sekolah Dasar Negeri Di Kota Semarang berdasarkan akreditasinya.

2. Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet

Nama penulis dari jurnal tersebut adalah Imelda A.Muis dan Muhammad Affandes, M.T dengan tahun terbitan jurnal yaitu Jurnal Sains, Teknologi dan Industri,Vol. 12, No. 2, Juni 2015.

Penelitian ini membahas tentang klasifikasi data *tweet* dengan menggunakan metode *Support Vector Machine* agar *tweet* yang ada tidak bercampur dengan iklan dan tidak iklan. Proses penelitian yang dilakukan berifat *non linear* maka dalam penelitian ini menggunakan *kerbel* RBF *(Radial Basic Function)* dimana parameter yang digunakan adalah nilai C dan y. Data yang digunakan pada penelitian ini bersumber dari *tweet* di tweeter, pada twitter tersedia *Aplication Programming Interface* (API) sebagai alat untuk mendapatkan data *tweet* di twitter. Kemudian data yang sudah diambil diimplementasikan ke dalam aplikasi RapidMiner agar dapat dijadikan solusi untuk klasifikasi data *tweet* agar tidak tercampur. Hasil dari penelitian ini mendapatkan nilai akurasi yang tinggi dalam melakukan klasifikasi, sehingga dapat diterapkan agar memberikan bantuan kepada pengguna dalam mengelola *tweet*, terutama *tweet* iklan, keberhasilan dengan mendapatkan nilai akurasi tertinggi 97,54% untuk data yang belum dilakukan penambahan *feature*, sedangkan untuk data yang sudah dilakukan pemilihan terhadap *feature* mencapai nilai akurasi tertinggi 99,12%.

3. Optimasi Klasifikasi Penilaian Akreditasi Lembaga Kursus Menggunakan K-NN dan Naïve Bayes

Nama penulis dari jurnal tersebut adalah Muhammad Amin, S.Kom, M.Kom dengan tahun terbitan jurnal Technologia Vol 7, No. 3, Juli 2016.

Penelitian dalam jurnal ini membahas tentang klasifikasi penilaian Akreditasi Lembaga Kursus dengan menerapkan metode K-NN dan Naïve Bayes. Data yang digunakan merupakan Data Kursus tahun 2015 yang kemudian data diolah dengan menggunakan RapidMiner, sebanyak 90% data akan digunakan untuk membangun struktur keputusan melalui metode K-NN dan Naïve Bayes. Hasil dari penerapan seleksi atribut dapat meningkatkan akurasi *K- Nearest Neighbor* menjadi lebih baik meskipun kenaikan yang dihasilkan tidak terlalu besar namun secara umum hasil penerapan K –NN dengan seleksi atribut menggunakan K lebih baik dari pada K – NN tanpa menggunakan seleksi atribut.

4. **Penerapan Metode Decision Tree Untuk Memprediksi Prestasi Siswa Kelas XII Dilihat dari Nilai Akhir Semester di SMK Negeri 1 Selong Tahun Pelajaran 2017/2018**

   Nama penulis dari jurnal tersebut adalah Martua Hamonangan Nasution, Muh. Iqbal Sofyan, dengan tahun terbit Jurnal Informatika dan Teknologi Vol. 3, No. 1, Januari 2020.

   Penelitian ini melakukan pengolahan terhadap nilai akhir siswa kelas XII yang ada di SMK Negeri 1 Selong untuk tahun pelajaran 2018/2019 dengan memanfaatkan Rapidminer sebagai tool dan menerapkan metode Decision Tree (C.45) untuk mengetahui tingkat prestasi siswa yang dilihat dari nilai raport. Penelitian ini merupakan salah satu tindakan sekolah yang

ditujukan guna meningkatkan prestasi siswa SMK Negeri 1 Selong. Dari penelitian yang telah dilakukan dengan Decision Tree Algoritma C4.5, hasil akurasi yang diperoleh sangat baik atau sempurna yaitu 97,22% Sedangkan nilai AUC yang diperoleh dari Decision Tree Algoritma C4.5 adalah 0.889 % dengan good classification.

5. Implementasi Algoritma Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa

Nama penulis dari jurnal tersebut adalah Arif Pratama, Randy Cahya Wihandika, dan Dian Eka Ratnawati dengan tahun terbitan Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 2, No. 4, April 2018.

Penelitian yang dilakukan pada jurnal tersebut adalah memprediksi ketepatan waktu kelulusan mahasiswa dengan menggunakan algoritma *Support Vector Machine.* Pada pengujiannya menggunakan data latih perbandingan yang berbeda, jumlah data latih yang digunakan antara laim 94, 113, 132, 150, dan 170 data. Sedangkan jumlah data uji yang digunakan untuk perbandingan antara lain 94, 75, 56, 38, dan 18 data. Hasil dari pengujian yang telah dilakukan dalam penelitian ini adalah rata-rata akurasi terbaik sebesar 80,55% menggunakan *kernel Gussian* RBF. Dengan hasil akurasi yang didapat dari metode SVM ini makan penelitian ini sudah cukup baik.

6. **Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang**

Nama penulis dari jurnal tersebut adalah Pusphita Anna Octaviani, Yuciana Wilandari, dan Dwi Ispriyanti dengan tahun terbitan Jurnal Gaussian, Volume 3, No. 4, Tahun 2014.

Peneilitan yang dilakukan pada jurnal tersebut adalah pengklasifikasian data akreditasi Sekolah Dasar di Kabupaten Magelang. Data yang digunakan dalam peneilitan ini adalah data tentang nilai akreditasi Sekolah Dasar (SD) di Kabupaten Magelang mulai tahun penetapan 2011-2013 yang diperoleh dari website resmi BAN-SM (Badan Akreditasi Nasional) dengan status akreditasi A,B, dan C. Setelah itu data yang diperoleh dianalisis dengan menggunakan software *MATLAB*. Hasil dari analisis yang telah dilakukan dalam penelitian ini yaitu klasifikasi Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang menggunakan metode SVM dengan data training yang diujikan sebanyak 337 data memiliki akurasi klasifikasi sebesar 98.810% dengan menggunakan fungsi kernel *Polynominal*, sedangkan dengan menggunakan fungsi kernel RBF memiliki akurasi klasifikasi 100%.

7. **Optimasi Data Mining Menggunakan Algoritma Naïve Bayes dan C4.5 Untuk klasifikasi Kelulusan Mahasiswa**

Nama penulis dari jurnal tersebut adalah Ni Luh Ratniasih dengan tahun terbitan jurnal yaitu Jurnal Teknologi Informasi dan Komputer, Volume 5, Nomor 1, Januari 2019.

Penelitian ini membahas tentang penyajian data apabila data yang tersedia dalam jumlah yang besar sehingga perlu menggunakan disipilin ilmu data mining agar informasi yang dihasilkan dapat ditampilkan secara rinci dan mudah untuk dimengerti. Dalam penelitian tersebut peneliti menggunakan 2 metode sebagai komparasi yaitu naïve bayes dengan C4.5 yang merupakan sebuah metode untuk melakukan teknik klasifikasi serta menggunakan salah satu tools data mining yaitu Rapid Miner. Hasil analisa dengan menggunakan metode C4.5 dengan menggunakan 50 record menghasilkan 5 aturan dalam pohon keputusan sehingga aturan tersebut dapat digunakan dalam penentuan kelulusan tepat waktu pada mahasiswa STIMIK STIKOM Bali, sedangkan menggunan metode naïve bayes menghasilkan akurasi data sebesar 89,27% dimana hasil performance akurasi tersebut menunjukkan kelulusan tepat waktu sebanyak 40 dan tidak tepat 10 orang.

Tabel 2.2 *State of The Art (SoTA)*

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
1	Theopilus Bayu Sangsoko	Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA)	Jurnal Teknik Informatika dan Sistem Informasi Volume 2 Nomor 2 Agustus 2016 e-ISSN : 2443-2229	SVM memiliki keterbatasan ketika jumlah atribut yang digunakan relatif banyak yang mengakibatkan beban komputasi menjadi berat dan akurasi menjadi kurang akurat.	pendekatan hybrid yang mengkombinasikan antara support vector machine (SVM) klasifier dengan metode particle swarm optimization (PSO) dengan jumlah dataset sebanyak 281 records dan 11 atribut.	pembentukan model klasifikasi yaitu nilai peserta didik saat mendaftar di salah satu SMA di Jawa Tengah pada tahun pelajaran 20132014 yang meliputi nama, nilai UN pada jenjang sebelumnya, nilai rata-rata raport, dan nilai tes psikologi peminatan	Tingkat akurasi yang paling besar sebesar 99.30% diperoleh ketika menerapkan PSO-SVM dengan parameter C (pinalti) sebesar 0.0 dengan menggunakan kernel anova.	Hasil uji coba pemilihan kernel yang sesuai dapat digantikan dengan automatic kernel selection sehingga tidak memerlukan waktu yang banyak untuk pemilihan kernel yang tepat.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
2	Ni Luh Ratiniasih	Optimasi Data Mining Menggunakan Algoritma Naïve Bayes dan C.45 Untuk Klasifikasi Kelulusan Mahasiswa	Teknologi Informasi dan Komputer, Volume 5, Nomor 1, Januari 2019.	Penyajian data yang tersedia dalam jumlah besar sehingga perlu menggunakan disiplin ilmu data minig agar informasi yang dihasilkan dapat rinci dan mudah untuk dimengerti.	Membangun sautu sistem pendukung keputusan untuk membantu pengklasifikasian kelulusan mahasiswa dengan menggunakan optimasi data mining menggunakan algoritma Naïve Bayes dan C4.5	Data yang digunakan merupakan dataset mahasiswa STIMIL SITKOM Bali.	Hasil dari Analisa dengan menggunakan metode C4.5 dengan menggunakan 50 record menghasilkan 5 aturan dalam pohon Keputusan. sedangkan menggunan metode naïve bayes menghasilkan akurasi data sebesar 89,27% dimana hasil performance akurasi tersebut menunjukkan kelulusan tepat	Pengembangan lebih lanjut dapat dilakukan dengan menggunakan algoritma yang lain untuk menyempurnakan atau menambah Tingkat akurasi pada klasifikasi kelulusan mahasiswa.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
3	Satrio Yudho Pangestu, Yudi Astuti, Lilis Dwi Farida	Algoritma Support Vector Machine Untuk Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia	Jurnal Mantik Penusa Volume 3, No1 Juni 2019. e-ISSN 2580-9741 p-ISSN2088-3943	Masyarakat bisa peropini melalui media social untuk mengekspresikan pemikiran secara bebas. Analisis sentiment dilakukan dengan memilah data dari twitter yang merupakan opini Masyarakat terhadap partai politik dan calon esekutif.	Klasifikasi data mining menggunakan Algoritma SVM untuk klasifikasi sikap politik terhadap partai politik indonesia	Sampel data yang bersumber dari twitter dengan total data 900.	Pengujian algoritma SVM untuk klasifikasi data Tweet Berbahasa Indonesia dengan total data 900 dan di peroleh rata-rata akurasi sebesar 71% error rate sebesar 29%, precission sevesar 73% recall sebesar 70% dan f_score sebesar 71%.	Dapat dilakukan komparasi dari beberapa algoritma klasifikasi data mining, missal Algoritma C4.5 dengan *Naïve Bayes* atau yang lainnya, sehingga dapat diketahui algoritma yang terbaik untuk pengklasifikasian sikap politik terhadap partai politik di Indonesia

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
4	Arif Pratama, Randy Cahya	Implementasi Algoritma SVM untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa.	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 2, No. 4, April 2018.	Kurang tepatnya Tingkat prediksi kelulusan Mahasiswa.	Menggunakan algoritma SVM untuk memprediksi Tingkat ketepatan waktu kelulusan mahasiswa	Menggunakan data latih dengan perbandingan berdeda.	menunjukan bahwa algoritma SVM mencapai Tingkat akurasi sebesar 80,55% menggunakan kernel gussin RBF. Dengan hasil ini akrasi yang didapat dari metode SVM ini membuat penelitian ini sudah cukup baik.	Mencoba membandingkan dengan algoritma lain untuk memprediksi Tingkat ketepatan waktu kelulusan mahasiswa.

5	Joy Nashar Utamajaya, Andi Mentari A.P, Siti Masnunah	Penerapan Algoritma *Naïve Bayes* 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 *Naïve Bayes* dalam penelitian ini menunjukan 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 *Naïve Bayes.*	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 Sulistyo 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 *precission,*	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 *Naïve Bayes,* Algoritma ID3 dan Regresi Linear.	atribut yang relevan, diantaranya: nilai rata-rata, ekstrakulikuler, semester, jumlah tanggungan orang tua dan penghasilan orang tua	*recall* dan *accucracy.* Berdasarkan nilai *precission*, algoritma ID3 memiliki nilai lebih baik daripada algoritma yang lain, sedangkan berdasarkan *recall* dan *accuracy* 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 mengetahu 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 *K-Fold Cross Validation* dna nilai keakuratan minat sebesar 90,23%	Pengujian menggunakan penggabungan beberapa metode untuk membandingkan tangkat 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	Al 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 *(http://archive.ics.uci.edu)*	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 siwa 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 *K-Means*	Metode *K-Means* memiliki tingkat

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
		Meningkatkan Mutu Pembelajaran pada Intansi Perguruan Tinggi menggunakan metode *KMeans Clustering*	03, Desember 2016	dihasilkan dapat memberikan sebuah pengetahuan baru terhadap intansi Akademi Komunitas Solok Selatan dalam meningkatkan mutu pendidikan	metode *clustetring* dalam meningkatkan mutu pembelajaran.	penerapan metode *KMeans Clustering*, diantaranya: nilai IP, jarak tempuh, jumlah kehadiran dan penghasilan orang tua.	*Clustering* sangat membantu pengelompokan data mutu pembelajaran siswa. Metode ini menunjukan bahwa karakteristik siswa dari *cluster* nilai IP dan kehadiran tergolong sedang dengan	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
10	Bisri Merluarlini, Diah Safitri, Abdul Hoyyi	Perbandingan Analisis Klasifikasi Metode K-Nearst Neighbor dan MARS Pada Data Akreditasi Sekolah Dasar Negeri Di Kota Semarang	Jurnal Gaussian, Volume 3, Nomor 3, Tahun 2014.	Data akreditasi yang belum terklasifikasi sesuai dengan kelas nya	Dilakukan perbandingan metode K-NN dan MARS dalam pengklasifikasian data akreditasi SD.	Data yang digunakan bersumber dari situs resmi badan akreditasi nasional.	Hasil dari penelitian ini untuk metode K-NN adalah 133,020 dan metode MARS adalah 152,381. Berdasarkan perhitungan APER, Specificty, dan sensivity menunjukan bahwa pengklasifikasian menggunakan metode MARS lebih baik	Ketersediaan data sangat mempengaruhi tingkat akurasi dalam melakukan pengklasifikasian data. Disarankan menggunakan agoritma yang lain untuk mengukur tingakt akurasi dari pada hasil klasifikasi yang telah didapat.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
11	Fithri Selva Jumeilah	Penerapan SVM untuk Pengkategorian Penelitian	Jurnal Resti Vol.1 No. 1 2017. ISSN Media Elektronik : 2580-0760	Penyusunan penelitian hendaknya harus perkategori agar mempermudah pencarian bagi yang membutuhkan.	Pengimplementasian algoritma SVM untuk mengkategorikan penelitian.	Untuk mengkategorikan penelitian dibutuhkan banyak penelitian yang membahas tentang algoritma text mining. Selain itu, juga dibutuhkan kumpulan abstrak penelitian yang akan digunaan sebagai data testing dan data training.	Algoritma Support Vector Machine dapat diimplementasikan dalam pembangunan Sistem Kategorisasi Topik penelitian. Sistem menghasilkan tingkat akurasi sebesar 90% dari 40 data pengujian untuk 4 kategori.	Saran untuk pengembangan lebih lanjut dari penelitian ini diharapkan penelitian ini mampu diimplementasikan pada beberapa perguruan tinggi atau dilakukan pengembangan terhadap masalah pengkategorian penelitian dengan pertimbangan yang lebih komplek dengan metode yang lain.

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
							PSO akurasinya 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 *e-ISSN: 27461335 Vol. 2, July 2021*	Apotek belum mengelompokan penjualan obat yang sering terjual sehingga laba dari penjualan obat tidak terserap dengan baik	mengklasifikasikan barangbarang yang sering terjual berdasarkan data dari tahun 2017- 2019 dengan penerapan Particle Swarm Optimization	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	*Improved Particle Swarm Optimization* untuk menyelesaikan permasalahan *Part Type Selection* dan	Brawijaya University, August 2015 (*page : 1003-1008)*	*Part type selection* dan *machine loading* 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
		Machine Learning pada *Flexible Manufacturing System (FMS)*		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 dangan melakukan hibridisasi dengan metode heuristik lain seperti simulated annealing, tabu search, dan variable neighborhood search (VNS)

No	Author	Judul	Jurnal/ Konferensi	Masalah	Solusi	Batasan	Hasil	Saran
14	Unler & Murat	*A discrete particle swarm optimization method for feature selection in binary classification problems*	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 membandingkannya dengan metode lain. Sepuluh dari 11 kumpulan data	Algoritma PSO yang dimodifikasi untuk masalah pemilihan subset fitur adalah algoritma PSO *diskrit* 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.

15							Tingkat akurasi yang paling besar sebesar 99.30%	Hasil uji coba pemilihan kernel yang sesuai dapat digantikan dengan
No							**Hasil**	**Saran**

							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.

Tabel 2.1 Matriks Penelitian

No.	Peneliti	Metode								*Tools*		Objek Penelitian		
		Naïve Bayes	MARS	Neural Network	Linear Regression	Support Vector Machine	Gaussian Process	C4.5	K-Nearest Neighbor	Rapidminer	Matlab	SD	SMP	SMA/SMK
1.	Bisri Merluarini, dkk (2014)		✔						✔			✔		
2.	Imelda A.Muis, dkk (2015)					✔				✔		-	-	-
3.	Muhammad Amin S.Kom,. M.Kom (2016)	✔							✔	✔		-	-	-

4.	Martua Hamonangan Nasution, dkk (2020)							✔		✔				✔
5.	Arif Pratama, dkk (2014)							✔		✔			✔	
6.	Pusphita Anna Octaviani, dkk (2014)					✔					✔	✔		
7.	Ni Luh Ratniasih (2019)	✔						✔		✔				