

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 Pengertian Konsentrasi

Konsentrasi merupakan salah satu faktor yang mempengaruhi proses belajar dan mengajar. Konsentrasi adalah memfokuskan pikiran terhadap suatu objek tertentu dengan menyampingkan hal hal yang tidak berhubungan dengan proses belajar dan mengajar yang dilakukan (Sinatrya & Pinilih, 2016)

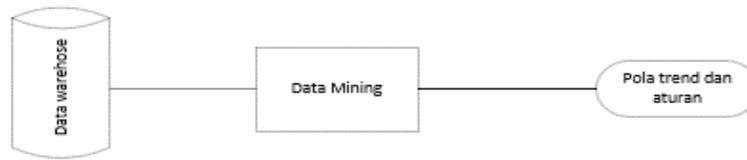
Ada beberapa kriteria dalam penilaian kredit kepada nasabah yang harus dilakukan atau prinsip utama yang berkaitan dengan kondisi calon nasabah. Prinsip itu dikenal dengan istilah 5C(Afifuddin & Nurjanah, 2019)(Fadillah & Hardiyana, 2018)

Konsentrasi adalah pemusatan fungsi jiwa terhadap suatu objek seperti konsentrasi pikiran, perhatian dan sebagainya (Azmi, 2017)

mengungkapkan konsentrasi dalam belajar merupakan pemusatan perhatian terhadap mata pelajaran dengan mengenyampingkan semua hal yang tidak berhubungan dengan pelajaran. Maka dari itu konsentrasi merupakan salah satu aspek yang mendukung siswa untuk mencapai prestasi yang baik dan apabila konsentrasi ini berkurang maka dalam mengikuti pelajaran di kelas maupun belajar secara pribadi akan terganggu.(Sukmayasa et al., 2016)

2.1.2 Data Mining

Data mining atau penambangan data adalah perangkat lunak yang digunakan untuk menemukan pola tersembunyi, tren, maupun aturan-aturan yang terdapat dalam basis berukuran besar dan menghasilkan aturan-aturan yang digunakan untuk memperkirakan perilaku di masa mendatang(Hidayati & Widi Nugroho, n.d.)



Gambar 2.1 Prinsip *Data Mining*(Afifuddin & Nurjanah, 2019)

Data mining terdiri dari *extract*, *transform*, dan memuat data transaksi ke sistem *data warehouse*, disimpan dan mengelola data dalam sistem database.

2.1.3 ETL

ETL (Extract, Transform, Load) adalah sekumpulan proses untuk mengambil dan memproses data dari satu atau banyak sumber menjadi sumber baru, misalnya mengolah database OLTP menjadi database OLAP. Sumber data yang diolah ETL bisa dari berbagai sumber data, tidak hanya dari database OLTP saja, tetapi bisa juga dari website, file teks, spreadsheet, database, email, dan lain sebagainya. ETL terdiri dari 3 bagian utama, setiap bagian memiliki fungsi sesuai namanya.

1. Extract, semua proses yang diperlukan untuk terhubung dengan beragam sumber data, dan membuat data tersebut tersedia bagi proses-proses selanjutnya.

Contoh :

1. Membaca file Microsoft Excel
2. Mengambil data dari database
3. Mengambil data kurs mata uang dari situs sebuah bank
4. Dan lain-lain

2. Transform, bagian ini mengacu pada fungsi apa saja yang berfungsi untuk mengubah data yang masuk menjadi data yang dikehendaki.

Fungsi-fungsi dapat berupa :

1. Pemindahan data
2. Validasi data sesuai aturan yang ditetapkan
3. Modifikasi isi, tipe atau struktur data

4. Integrasi atau penggabungan data dari sumber-sumber lain
5. Perhitungan, dan lain-lain

Contoh :

1. Mengubah tipe data dari String menjadi Date
2. Memeriksa apakah data nomor kartu kredit sesuai format kartu kredit yang baku atau tidak
3. Melakukan lookup untuk mendapatkan nilai tertentu berdasarkan kunci
4. dan lain-lain

3. Load, semua proses yang diperlukan untuk mengisi data ke target.

Contoh :

1. Hasil dari proses sebelumnya disimpan ke dalam file Microsoft Exce
2. Hasil dari proses sebelumnya disimpan ke dalam database OLAP
3. dan lain-lain (Data Warehouse. Penulis. Neni Purwati)

multidimensi, menyediakan akses data untuk analisis bisnis dan profesional teknologi informasi, menganalisis data oleh aplikasi perangkat lunak, menyajikan data dengan format yang berguna, seperti grafik atau tabel *Data mining* sering dikatakan berurusan dengan "penemuan pengetahuan" dalam basis data. Hal yang menarik, *data mining* menjadi perangkat yang membantu para pemakai untuk menemukan jawaban atas pertanyaan-pertanyaan yang tidak pernah mereka pikirkan sebelumnya. (Purwati et al., 2020)

Istilah(Chafid Nurul, 2018) *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Ada beberapa peran utama dalam *data mining* antara lain:

1. Deskripsi

Cara yang digunakan untuk menggambarkan sekumpulan data secara ringkas. Banyak cara yang digunakan dalam memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya, yaitu Deskripsi Grafis, Deskripsi Lokasi dan Deskripsi keragaman.

2. Estimasi

Algoritma estimasi yang biasa digunakan adalah: *Linear Regression*, *Neural Network*, *Support Vector Machine*. Algoritma estimasi mirip dengan algoritma klasifikasi, tapi variabel target adalah berupa bilangan numerik dan bukan kategorikal (nominal). Model dibangun dari data dengan *record* yang lengkap, yang menyediakan nilai dari variabel sebagai prediktor, kemudian estimasi nilai dari variabel target ditentukan berdasarkan nilai dari variabel prediktor. Penentuan kebijakan atau suatu nilai pada proses yang akan dilakukan. Estimasi dapat dilakukan dari data-data lama yang akan diolah.

1. Prediksi

Algoritma prediksi sama dengan algoritma estimasi dimana label/target/class bertipe numerik, bedanya adalah data yang digunakan merupakan data rentetan waktu (*data time series*). Sifat prediksi bisa

menghasilkan *class* berdasarkan berbagai atribut yang kita sediakan. Penentuan hasil dari proses yang sedang berlangsung. Data-data yang digunakan untuk prediksi berasal dari data yang ada saat proses sedang berlangsung. Istilah prediksi kadang digunakan juga untuk klasifikasi, tidak hanya untuk prediksi *time series*, karena sifatnya yang bisa menghasilkan *class* berdasarkan berbagai atribut yang kita sediakan.

2. Klasifikasi

Algoritma yang menggunakan data dengan *target/class/label* berupa nilai kategorikal (nominal). Pengelompokan data-data yang ada menjadi dalam kelompok yang sudah ditentukan nama kelompoknya. Metode yang cocok untuk klasifikasi, yakni: *Naïve Bayes*, *K-Nearest Neighbor*, *C4.5*, *ID3*, *CART*, *Linear Discriminant Analysis*, dan yang lainnya. Contoh, apabila *target/class/label* adalah pendapatan, maka bisa digunakan nilai nominal (kategorikal) pendapatan besar, menengah, kecil.

3. Cluster

Klastering adalah pengelompokkan data, hasil observasi dan kasus ke dalam *class* yang mirip. Suatu klaster (*cluster*) adalah koleksi data yang mirip antara satu dengan yang lain, dan memiliki perbedaan bila dibandingkan dengan data dari klaster lain. Metode yang cocok untuk klastering, yakni: *K-Means*, *K-Medoids*, *Self-Organizing Map (SOM)*, *Fuzzy*

C-Means, dan yang lainnya. Perbedaan utama algoritma klustering dengan klasifikasi adalah klustering tidak memiliki *target/class/label*, jadi termasuk *unsupervised learning*. Klustering sering digunakan sebagai tahap awal dalam proses *data mining*, dengan hasil kluster yang terbentuk akan menjadi input dari algoritma berikutnya yang digunakan.

4. Asosiasi

Algoritma *association rule* (aturan asosiasi) adalah algoritma yang menemukan atribut yang “jalan bersamaan”. Dalam dunia bisnis, sering disebut dengan *affinity analysis* atau *market basket analysis*. Algoritma *association rules* berangkat dari pola “*If antecedent, then consequent,*” bersamaan dengan pengukuran *support (coverage)* dan *confidence*

(*accuracy*) yang terasosiasi dalam aturan. Algoritma *association rule* diantaranya adalah: *Apriori algorithm, FP-Growth algorithm, GRI algorithm.*

Berikut beberapa alasan *data mining* menjadi penting saat ini :

1. Data yang tersedia sangat besar:

Selama *decade* terakhir harga *hardware* terutama harga *hardisk* telah menurun drastis. Bersama dengan ini, perusahaan telah menyimpan data yang sangat besar melalui banyak aplikasi. Dengan semua data ini untuk mengeksplorasi, perusahaan menginginkan dapat menemukan pola dan informasi yang tersembunyi sehingga dapat membantu mengarahkan strategi bisnis perusahaan menjadi lebih baik.

2. Meningkatkan kompetisi:

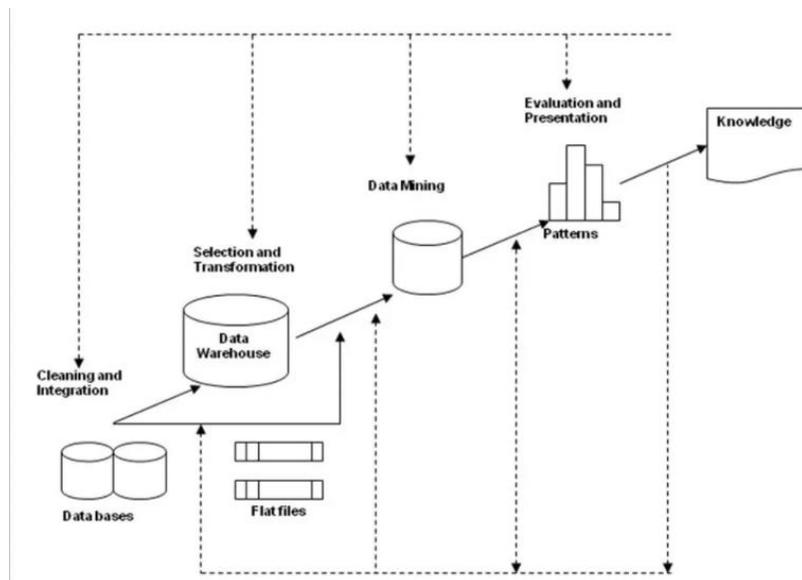
Kompetisi yang tinggi sebagai dampak dari modernisasi pasar dan distribusi seperti *internet* dan telekomunikasi. Perusahaan-perusahaan di seluruh dunia menghadapi persaingan dan kunci untuk keberhasilan bisnis adalah kemampuan untuk mempertahankan pelanggan dan memperoleh pelanggan baru. *Data mining* adalah teknologi yang memungkinkan perusahaan untuk menganalisis faktor-faktor yang mempengaruhi masalah ini.

3. Teknologi yang telah siap:

Teknologi *data mining* yang sebelumnya hanya ada dalam wilayah akademik, tapi sekarang banyak dari teknologi tersebut telah matang dan siap diterapkan dalam perusahaan. Algoritma yang lebih akurat, efektif dan dapat menangani data yang semakin banyak dan rumit. Selain itu pemrograman aplikasi antar muka *data mining* telah distandarisasi yang

akan memungkinkan para pengembang untuk membangun aplikasi *data mining* akan lebih baik.

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada gambar II-2.



Gambar 2.2 Proses *Data Mining*(Fadillah & Hardiyana, 2018)

Pada gambar 2.2 digambarkan proses *data mining* adalah data dipilih, dibersihkan, dan dilakukan *preprocessing* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan *data internal* dan *eksternal* ke dalam tinjauan organisasi secara menyeluruh. Penggunaan algoritma *data mining* dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai. Keluaran dari *data mining* dievaluasi untuk melihat apakah *knowledge domain* ditemukan dalam bentuk rule yang telah diekstraksi dari jaringan.

Adapun proses dari KDD(*Knowledge Discovery in Database*)sebagai berikut (Mandala & Putri, 2018):

1. *Data Selelection*

Pemilihan data dari sekumpulan data operasional perlu dilakukan tahap penggalian informasi sebelum KDD di mulai.

2. *Preporcessing*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*,

yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

Proses *coding* pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam *database*

4. *Data Mining*

Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu

5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.

Evaluasi adalah kunci ketika membuat aplikasi berbasis *data mining*. Ada berbagai macam cara melakukan evaluasi. Evaluasi ternyata tidak sesederhana yang kita bayangkan. Jika kita memiliki data yang kita gunakan dalam proses pelatihan, maka tidak serta merta kita menjadikan data tersebut sebagai indikator keberhasilan aplikasi yang kita buat. Oleh karena itu, kita membutuhkan metode tertentu guna memprediksi performa berdasarkan eksperimen untuk berbagai macam data selain *data training* tersebut.

Pada umumnya data yang cukup banyak dapat dimanfaatkan untuk pengujian. Hanya saja masalah yang kerap dijumpai adalah datanya. Oleh karena itu, kita harus memastikan data yang akan kita gunakan baik untuk pelatihan maupun untuk pengujian merupakan data yang berkualitas.

2.1.3 Clustering

Clustering dalam *data mining* merupakan metode pembelajaran data untuk memprediksi nilai dari sekelompok atribut. Algoritma klasifikasi akan menghasilkan sekumpulan aturan yang disebut *rule* yang akan digunakan sebagai *indicator* untuk dapat memprediksi kelas dari data yang ingin diprediksi (Mandala & Putri, 2018). Klasifikasi digunakan dalam banyak sekali bidang, dan secara teori algoritma klasifikasi sama seperti otak manusia.

Otak manusia mampu mengolah data yang sudah ada sebagai pengalaman dalam bertindak. Dalam *data mining* ada beberapa algoritma klasifikasi yang banyak digunakan dalam masyarakat atau dalam penelitian secara luas diantaranya adalah *Decision / classification trees*, *Bayesian classifier / Naïve Bayes classifiers*, *Neural networks*, Analisa Statistik, Algoritma Genetika, *Rough sets*, *K-Nearest Neighbor*, metode *Rule Based*, *Memory Based reasoning*, dan *Support Vector Machine (SVM)*.

Tujuan dari algoritma klasifikasi adalah untuk menemukan relasi antara beberapa variable yang tergolong dalam kelas yang sama. Relasi tersebut akan digambarkan dengan aturan-aturan agar dapat memprediksi kelas dari data yang *attribute* nya sudah diketahui.

Penilaian algoritma klasifikasi biasanya dilihat dari akurasi model. Akurasi model merupakan ketepatan model dalam memprediksi kelas data. Selain akurasi kecepatan pembentukan model, kemampuan algoritma dalam mengatasi data yang tidak relevan atau bahkan data yang tidak lengkap, serta kemampuan algoritma ketika diterapkan pada data jumlah besar maupun kecil.

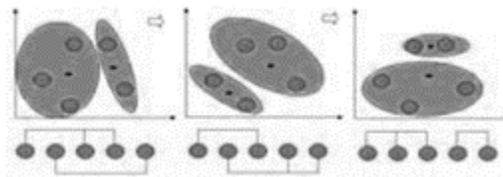
Algoritma K-Means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Namun, K-Means mempunyai kelemahan yang diakibatkan oleh penentuan pusat awal cluster. Hasil cluster yang terbentuk dari metode K-Means ini sangatlah tergantung pada inisiasi nilai pusat awal cluster yang diberikan (Sani, 2014)

2.1.4 K-Means

K-Means merupakan teknik pengelompokan yang bekerja berdasarkan Partitioned Clustering. Prinsip kerja dari pengelompokan Hierarchical Clustering dilakukan secara bertahap (Wahyuningsih & Utari, 2018).

Beberapa teknik klastering yang paling sederhana dan umum adalah klastering K-means. Secara detail teknik ini menggunakan ukuran ketidakmiripan untuk mengelompokkan obyek. Ketidakmiripan dapat diterjemahkan dalam konsep jarak. Dua obyek dikatakan mirip jika jarak dua objek tersebut dekat. Semakin tinggi nilai jarak, semakin tinggi nilai ketidakmiripannya (Verawati & Hasibuan, n.d.). K-Means Clustering merupakan metode yang termasuk ke dalam golongan algoritma Partitioning Clustering. Langkah-langkah dari metode K-Means adalah sebagai berikut : (Dwirohayati et al., n.d.)

1. Tentukan nilai k sebagai jumlah cluster yang ingin dibentuk.
2. Bangkitkan k centroid (titik pusat cluster) awal secara acak.
3. Hitung jarak setiap data ke masing-masing centroid menggunakan rumus korelasi antar dua objek (Euclidean Distance).
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
5. Tentukan posisi centroid baru ($k C$) dengan cara menghitung nilai rata-rata dari data yang ada pada centroid yang sama. Dimana n_k adalah jumlah dokumen dalam cluster k dan d_i adalah dokumen dalam cluster k.
6. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama, tidak sama.



Gambar 2.3 Langkah- Langkah metode k-means(Sani, 2014)

2.1.5 Pemilihan Variabel

Pemilihan variabel yang juga disebut sebagai pemilihan atribut, digunakan pada *dataset* untuk menemukan pola yang penting dalam *data mining*. Pemilihan variabel digunakan untuk pengurangan dimensi pada *dataset*. Pemilihan variabel digunakan untuk melakukan eliminasi variabel yang tidak *relevan* dan *redundan*, yang dapat menyebabkan kebingungan dalam penggunaan variable.

Pemilihan variabel dapat mengurangi dimensi data, hal ini memungkinkan lebih efektif dalam operasi agar lebih cepat dari beberapa algoritma data mining. Dengan adanya pemilihan variabel membuat algoritma data mining lebih cepat dan lebih efektif.

Penggunaan pemilihan variabel pada *dataset* yang menggunakan variabel bebas dapat meningkatkan performa model. Pemilihan variabel juga merupakan proses yang cukup memakan biaya, dan juga berten.tangan dalam asumsi awal, bahwa semua informasi diperlukan untuk mencapai akurasi yang maksimal.

Metode yang dapat digunakan untuk pemilihan variabel antara lain *Backward Elimination*, *Forward Selection*, *Genetic Algorithm*, dan yang lainnya. Metode-metode tersebut digunakan dalam penelitian *data mining* agar dapat menghasilkan variabel yang relevan dalam penelitian. Pemilihan variabel dengan filter model ini lebih murah dalam komputasi karena tidak melibatkan induksi algoritma dalam prosesnya.

2.1.6 Pengujian Akurasi dan Validasi Metode Klasifikasi Data Mining

2.6 Sampel

Sebelum melakukan tahapan pengujian data, dilakukan pengambilan sampel terlebih dahulu untuk melakukan *testing* data. Sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi (Sugiyono, 2016). Pengambilan sampel dalam penelitian ini dengan teknik probability sampling yaitu proportionate stratified random sampling dengan menggunakan rumus slovin. Probability sampling adalah teknik pengambilan sampel yang memberikan peluang yang sama bagi setiap unsur (anggota) populasi untuk dipilih menjadi anggota sampel. Besarnya sampel dalam penelitian ini ditentukan dengan menggunakan persamaan Slovin sebagai berikut :

$$n = \frac{N}{1 + Ne^2}$$

Keterangan :

n = jumlah elemen / anggota sampel

N = jumlah elemen / anggota populasi

e = error level (tingkat kesalahan)

(catatan: umumnya digunakan 1 % (0,01), 5 % (0,05), dan 10 % (0,1))

Selanjutnya dilakukan pengambilan sampling dengan menggunakan perhitungan *Proportionate Stratified Random* untuk menghitung proporsi jumlah sampel, yang menurut (Rulirianto, 2018) teknik sampling ini digunakan untuk populasi yang mempunyai anggota/unsur yang tidak homogen dan berstrata secara proporsional. Hal ini dilakukan karena atribut mempunyai anggota atau unsur yang tidak homogen dan berstrata secara proporsional, dengan rumus perhitungannya sebagai berikut:

$$n_i = \frac{N_i}{N} \times n$$

Keterangan :

n_i = jumlah anggota sampel menurut stratum (tingkatan)

n = jumlah anggota sampel seluruhnya (berdasarkan hasil Slovin)

N_i = jumlah anggota populasi menurut stratum

N = jumlah anggota populasi seluruhnya

Kurva ROC (*Receiver Operating Characteristic*)

Kurva ROC atau *Receiver Operating Characteristic* menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertical (Fadillah & Hardiyana, 2018)

2.2 Tinjauan Studi

Berikut adalah ringkasan dari beberapa penelitian sebelumnya yang terkait dengan klasifikasi *data mining*.

Table 2.1 Penelitian Terdahulu (Kurniawan & Sistem Informasi, 2016)

Judul	Metode	Tujuan	Hasil
-------	--------	--------	-------

<p>PENERAPAN METODE K- MEANS CLUSTERING PADA PERUSAHAAN</p>	<p>Pengujian menggunakan Kmeans</p>	<p>mencari partisi yang maksimum dari data dengan meminimalkan kriteria jumlah kesalahan kuadrat dengan prosedur iterasi yang optimal. Variabel yang digunakan pada penelitian ini adalah kode produk, jumlah transaksi, harga produk.</p>	<p>didapatkan pengelompokan jenis artikel dengan jumlah stok terbanyak dan sedikit.</p>
<p>PENERAPAN DATA MINING DALAM PENERAPAN DATA MINING DALAM MENINGKATKAN MUTU PEMBELAJARAN PADA MENINGKATKAN MUTU PEMBELAJARAN PADA INSTANSI PERGURUAN TINGGI</p>	<p>Pengujian menggunakan Kmeans</p>	<p>Meningkatkan mutu pem nakan metode clustering untuk Meningkatkan mutu pembelajaran belajaran pada Instansi Perguruan Tinggi di P pada Instansi Perguruan Tinggi di P rogram Studi TKJ Akademi Komunitas omunitas omunitas Solok Selatan</p>	<p>, jumlah kehadiran dan penghasilan orang tua. Di mana akan dan penghasilan orang tua. Di mana akan . Di mana akan mempresentasikan data mempresentasikan data mahasiswa dengan mutu pembelajaran sangat baik, mahasiswa dengan mutu pembelajaran sangat</p>

MENGGUNAKAN METODE K-MEANS CLUSTERING CLUSTERING CLUSTERING (STUDI KASUS DI (STUDI KASUS DI PROGRAM STUDI TKJ AKADEMI KOMUNITAS SOLOKSELATAN) SELATAN)			baik,baik, cukup baik, dan kurang baik. baik, cukup baik, dan kurang baik.
IMPLEMENTASI DATA MINING UNTUK CLUSTERING DAERAH PENYEBARAN PENYAKIT DEMAM BERDARAH DI KOTA TANGERANG SELATAN MENGGUNAKAN ALGORITMA K-MEANS	Pengujian menggunakan Kmeans	Untuk melihat daerah penyebarannya perlu dibuat sebuah pengelompokan penyebaran agar dapat memperoleh pusat titik penyebaran	Hasil pengelompokan C1 ada 20 Kelurahan, C2 ada 13 Kelurahan dan C3 ada 19 dari total Kelurahan 52 akan menjadi bahan untuk melakukan penyehatan lingkungan sesuai dengan kelompok yang akan dikerjakan oleh Dinas Kesehatan Tangerang Selatan.
PERANCANGAN APLIKASI K-MEANS SEBAGAI	Pengujian menggunakan Kmeans	memberikan sumber informasi penentu kelas	clustering data mahasiswa 2011 dengan algoritma k-

<p>PENENTU KONSENTRASI BAGI MAHASISWA INFORMATIKA UMS</p>		<p>konsentrasi bagi mahasiswa. Variabel yang digunakan nilai mata kuliah yang dominan pada masingmasing konsentrasi dari semester 1 sampai 4</p>	<p>means diperoleh data cluster 1 dengan titik centroid (3,333 ; 3,548 ; 3,098) dan cluster 2 dengan titik centroid (0,915 ; 1,110 ; 0,773) serta cluster 3 dengan titik centroid (2,682 ; 3,221 ; 1,880). Pada anggota cluster 1 maka akan direkomendasikan mengambil konsentrasi Jaringan Komputer dan Multimedia, anggota cluster 2 direkomendasikan untuk konsentrasi Rekayasa Perangkat Lunak dan Animasi sedangkan anggota cluster 3 direkomendasikan mengambil konsentrasi Sistem Informasi dan Enterprise.</p>
---	--	--	--

SISTEM REKOMENDASI PEMILIHAN MATA KULIAH PEMINATAN MENGUNAKAN ALGORITMA KMEANS DAN APRIORI (STUDI KASUS: JURUSAN S1 TEKNIK INFORMATIKA FAKULTAS INFORMATIKA	Pengujian menggunakan Kmeans	melakukan studi dan implementasi metoda K-means dan Apriori untuk memberikan rekomendasi pemilihan mata kuliah pilihan sesuai dengan kelompok keahlian	. Hasil eksperimen menunjukkan algoritma K-means dapat mengelompokkan mahasiswa kedalam kelompok keahlian dan Apriori dapat menghasilkan rule yang dapat digunakan sebagai rekomedasi pemilihan mata kuliah pilihan.
---	------------------------------------	---	---

Tabel II-2 (lanjutan)

Pada tabel 2.2. Ringkasan tinjauan studi, hampir semua penelitian menggunakan metode klasifikasi *data mining*. Dalam hal ini penulis mencoba membandingkan dua metode klasifikasi *data mining* yaitu *K-Means* untuk melihat akurasi yang tepat dari penghitungan metode tersebut penulis menggunakan pengujian *k-folds cross validation* dengan $k=10$ agar dapat mengetahui tingkat akurasinya terhadap prediksi peminatan konsentrasi matakuliah . Nantinya hasil dari pengujian tersebut akan dan kemudian akan dimodifikasi modelnya sehingga diharapkan dapat menaikkan nilai pengujian baik akurasi, precision, recall dan lain-lain.