

BAB II

LANDASAN TEORI

2.1 *Data mining*

Secara singkat, *data mining* berarti menggali atau menemukan informasi dari sejumlah besar data dan secara luasnya, *data mining* adalah proses menemukan *interesting knowledge* dari sejumlah besar data yang tersimpan dalam *database*, *data warehouse*, atau media penyimpanan lainnya. Sedangkan berdasarkan *Gartner Group*, *data mining* adalah sebuah proses untuk menemukan hubungan, pola, dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam media penyimpanan, menggunakan teknologi pengenalan pola serta statistika dan matematika (Chatterjee et al. 2019).

Dengan berdasarkan pengertian di atas, maka dapat dikatakan jika *data mining* adalah sebuah proses mencari informasi dari data yang berjumlah besar dan tersimpan dalam media penyimpanan. *Data mining* bisa tercipta dikarenakan adanya keinginan untuk mencari pengetahuan atau informasi dari data yang tersimpan dalam jumlah yang banyak. Selain hal tersebut, dalam penerapannya terdapat beberapa faktor yang mendukung pengembangan *data mining*, antara lain:

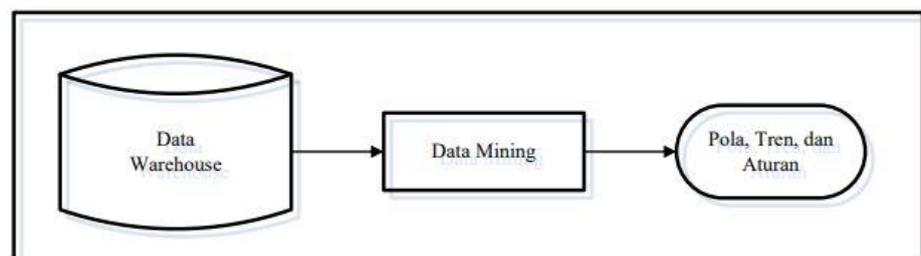
1. Pertumbuhan yang besar pada *Data mining*, dalam penerapannya akan mengakses file yang besar. Hal ini dikarenakan hasil dari *data mining* akan menjadi dasar dalam pengambilan keputusan untuk melakukan tindakan tertentu, sehingga tingkat kebenaran informasi yang dihasilkan sangat diperlukan untuk mencapai tingkat kebenaran informasi yang tinggi, maka diperlukan data yang sangat besar untuk diolah. Seiring dengan perkembangan media penyimpanan, maka perkembangan jumlah data yang disimpan bergerak naik sangat cepat. Akan tetapi, jika terjadinya minim pengetahuan untuk mengolah data tersebut maka akan tercipta kondisi “*data rich but information poor*” dimana tidak diketahuinya informasi dari data yang tersimpan tersebut.

2. Pertumbuhan yang luar biasa dalam daya komputasi dan kapasitas penyimpanan oleh perangkat keras. Penerapan *data mining* pada dasarnya akan membutuhkan kebutuhan sumber daya dan kemampuan komputasi yang sangat besar. Dengan perkembangan teknologi yang terjadi, mengakibatkan perkembangan yang sangat pesat bagi kemampuan komputasi dan kapasitas penyimpanan oleh perangkat keras. Sehingga dengan ketersediaan perangkat yang memiliki kemampuan tersebut, dapat menjadikan proses *data mining* menjadi cukup layak untuk dilakukan secara komersial maupun non-komersial.
3. Peningkatan kecepatan dan keakuratan untuk membuat atau menentukan keputusan. Sebuah lembaga atau institusi, organisasi atau perusahaan dituntut untuk selalu dapat berinovasi dan memberikan pelayanan-pelayanan agar mencapai hasil yang ditargetkan. *Data mining* dalam peranannya dapat memberikan pengolahan informasi yang dapat menjadi pertimbangan dalam melakukan sebuah tindakan.
4. Peningkatan terhadap akses ke data dengan melalui navigasi Web atau intranet. Perkembangan teknologi jaringan sampai sekarang, mengakibatkan kemudahan akses oleh data baik itu melalui jaringan internet atau intranet. Dengan kemudahan akses tersebut, maka mengakibatkan data transaksi mengenai akses data yang sangat besar. Sehingga dengan *data mining*, diharapkan dapat membantu dalam mengolah data transaksi tersebut agar dapat menjadi informasi yang berguna bagi perusahaan atau organisasi.
5. Pengembangan terhadap perangkat lunak untuk *data mining*. Seiring dengan perkembangan dan kebutuhan akan *data mining*, mengakibatkan pertumbuhan perangkat lunak yang dapat mendukung *data mining* tersebut. Perangkat lunak yang dapat membantu dalam *data mining*, antara lain Orange, RapidMiner, Weka, JhepWork, KNIME, dan lain sebagainya. Dimana pada masing-masing perangkat lunak memiliki perbedaan masing-masing seperti bahasa pemrograman yang

digunakan.

Dalam penerapannya, datamining dapat diterapkan dalam berbagai penyimpanan data. Media penyimpanan yang disebut antara lain *relational database, data warehouses, transactional database, advanced database system, flat files, data streams*, dan *World Wide Web*. Dan dalam penerapannya pun, teknik dan algoritma yang digunakan untuk *data mining* berbeda satu sama lain antar setiap media penyimpanan dan fungsi dari *data mining* tersebut. Algoritma *data mining* yang ada sekarang bermacam-macam, namun berdasarkan jurnal disebutkan 10 algoritma terbaik untuk *data mining*, antar lain C4.5, algoritma k-means, *Support Vector Machine*, algoritma apriori, algoritma EM, PageRank, AdaBoost, kNN (*K-Nearest Neighbor*), Naive Bayes, CART (*Classification and Regression Tree*). Sedangkan untuk setiap algoritma menyesuaikan dengan fungsi dari *data mining*, yaitu fungsi karakterisasi, fungsi diskriminasi, fungsi asosiasi, fungsi prediksi, fungsi klasifikasi, dan fungsi *cluster*.

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 seperti gambar 2.1 dibawah ini.



Gambar 2.1. Prinsip *Data mining*

Data mining terdiri dari *extract, transform*, dan memuat data transaksi ke sistem *data warehouse*, disimpan dan mengelola data dalam sistem database 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 (Srivastava and Khare 2017).

Istilah *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. *Description*

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. *Estimation*

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.

3. *Prediction*

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.

4. *Classification*

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.

5. *Clustering*

Clustering 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 klastering dengan klasifikasi adalah klastering tidak memiliki *target/class/label*, jadi termasuk *unsupervised learning*. Klastering sering digunakan sebagai tahap awal dalam proses *data mining*, dengan hasil klaster yang terbentuk akan menjadi input dari algoritma berikutnya yang digunakan.

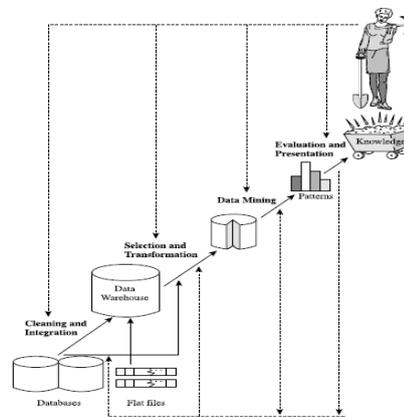
6. *Association*

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*.

2.2 KDD

KDD merupakan singkatan dari *Knowledge Discovery from Data*. KDD mulai dikembangkan pada era awal 1990-an. Fayyad pada 1996 menggagas proses model KDD dan menetapkan langkah untuk proyek DM. Pada metodologi KDD terdapat 9 tahap sesuai dengan gambar 2.2 di bawah ini.



Gambar 2.2. Tahapan Metodologi KDD

Berdasarkan gambar II.2 tahapan metodologi KDD dapat dijelaskan sebagai berikut.

1. *Data Cleaning*

Pada tahapan ini melakukan pemilihan data yang relevan dari database dengan melakukan pemisahan terhadap data yang tidak konsisten dan data yang tidak relevan.

2. *Data Integration*

Pada tahapan ini dilakukan integrasi terhadap data yang ada dengan cara menggabungkan berbagai sumber data menjadi satu sumber.

3. *Data Selection*

Pada tahapan ini melakukan pemilihan terhadap data yang relevan dengan analisa yang akan dilakukan pada database.

4. *Data Transformation*

Pada tahapan ini dilakukan perubahan terhadap format data yang ada menjadi format data yang sesuai untuk diproses dalam *data mining*.

5. *Data mining*

Pada tahapan ini dilakukan proses *data mining*, dengan menerapkan metode tertentu untuk mendapatkan informasi yang tersembunyi dari data yang ada.

6. *Pattern Evaluation*

Pada tahapan ini dilakukan identifikasi terhadap pola-pola yang menarik yang didapat dari hasil *data mining*, untuk kemudian direpresentasikan.

7. *Knowledge Presentation*

Pada tahapan ini dilakukan visualisasi dan penyajian terhadap pengetahuan mengenai teknik yang digunakan untuk memperoleh pengetahuan yang diperoleh user.

2.3 **Klasifikasi**

Klasifikasi adalah sebuah proses untuk mencari model atau fungsi yang menjelaskan dan membedakan kelas atau konsep dari data, dengan tujuan untuk menggunakan model dan melakukan prediksi dari kelas suatu objek dimana tidak diketahui label dari kelas tersebut. Model yang ada berasal dari analisis dari kumpulan training data (objek data dimana kelas dari label diketahui). (Ardi,2014)

Klasifikasi 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 indikator untuk dapat memprediksi kelas dari data yang ingin diprediksi. 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 seperti *naïve bayes*, *neural network*, *logistic regression*, *Support*

Vector Machine dan *decision tree*. Tujuan dari algoritma klasifikasi adalah untuk menemukan relasi antara beberapa variabel yang tergolong dalam kelas yang sama. Relasi tersebut akan digambarkan dengan aturan-aturan agar dapat memprediksi kelas dari data yang atributnya sudah diketahui (*Marconi et al. 2019*).

Proses pembuatan model klasifikasi dibedakan menjadi tiga tahap besar:

a. Tahap Pembelajaran

Tahap dimana algoritma klasifikasi diterapkan kedalam data contoh untuk mendapatkan relasi data dalam setiap kelasnya. Tahap ini akan membentuk model yang berisikan aturan aturan atribut dalam menentukan kelas data.

b. Tahap Pengujian

Tahap pengujian merupakan tahap penerapan aturan aturan yang sudah terbentuk pada tahap pembelajaran kedalam contoh data yang tidak termasuk dalam data pembelajaran. Dalam Tahap ini, aturan model yang dimiliki akan diterapkan pada setiap atribut dalam data pengujian dan dilihat kecocokan antara kelas yang diprediksi dengan kelas data sesungguhnya.

c. Tahap Prediksi

Tahap ini merupakan tahap dimana model yang dihasilkan benar benar diterapkan pada data yang belum diketahui kelasnya.

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.

2.4 Algoritma C4.5

Algoritma C4.5 yaitu pohon keputusan yang merupakan metode

klasifikasi dan prediksi yang sangat kuat dan kental. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel target. Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, dia sangat bagus sebagai langkah awal dalam proses pemodelan (Karpathy A, 2019).

Quinlan (1993) mengemukakan bahwa C4.5 adalah algoritma yang digunakan untuk klasifikasi data yang dapat mengolah data/atribut numerik, algoritma ini dapat mengatasi nilai atribut yang hilang, dan dapat mengatasi data kontinyu dan *pruning*/penyederhanaan. Hasil dari proses klasifikasi berupa aturan yang dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 dan secara umum digunakan untuk membangun pohon keputusan dengan melakukan tahapan sebagai berikut: pilih atribut sebagai akar (*root*), buat cabang untuk tiap-tiap nilai, bagi kasus dalam cabang, kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 yaitu :

1. Mempersiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus 2.1 dibawah ini :

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i \quad (2.1)$$

Keterangan :

S= Himpunan kasus

n = jumlah partisi S

P_i = proporsi S_i terhadap S

3. Kemudian hitung nilai gain menggunakan rumus 2.2 dibawah ini :

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

Keterangan :

S = Himpunan Kasus

A = Fitur

n = jumlah partisi atribut A

$|S_i|$ = Proporsi S_i terhadap S

$|S|$ = jumlah kasus dalam S

4. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat :
- semua *record* dalam simpul N mendapat kelas yang sama.
 - Tidak ada atribut didalam *record* yang dipartisi lagi
 - Tidak ada *record* didalam cabang yang kosong.
6. Dalam algoritma C4.5, pemilihan atribut dilakukan dengan menggunakan gain ratio. Atribut dengan nilai gain ratio tertinggi dipilih sebagai atribut test untuk simpul. Untuk menghitung gain ratio perlu diketahui suatu term baru yang disebut dengan split information. Split information dihitung dengan menggunakan rumus 2.3 sebagai berikut :

$$SplitInformation(S, A) = - \sum_{t=1}^c \frac{S_t}{S} \log_2 \frac{S_t}{S} \quad (2.3)$$

Keterangan:

S_1 sampai $S_c = c$ subset yang dihasilkan dari pemecahan S dengan n menggunakan atribut A yang mempunyai sebanyak c nilai.

7. Untuk menghitung gain ratio digunakan rumus 2.4 sebagai berikut :

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (2.4)$$

8. Untuk menghitung *Pessimistic error rate* digunakan rumus 2.5 berikut.

$$e = \frac{r + \frac{z^2}{2x} + z \sqrt{\frac{r}{n} - \frac{r^2}{n} + \frac{z^2}{4n^2}}}{2a} \quad (2.5)$$

Jika $c = 25\%$ (default untuk C4.5) maka $z = 0,69$ (dari distribusi normal)

f = nilai perbandingan error rate.

n = total sample.

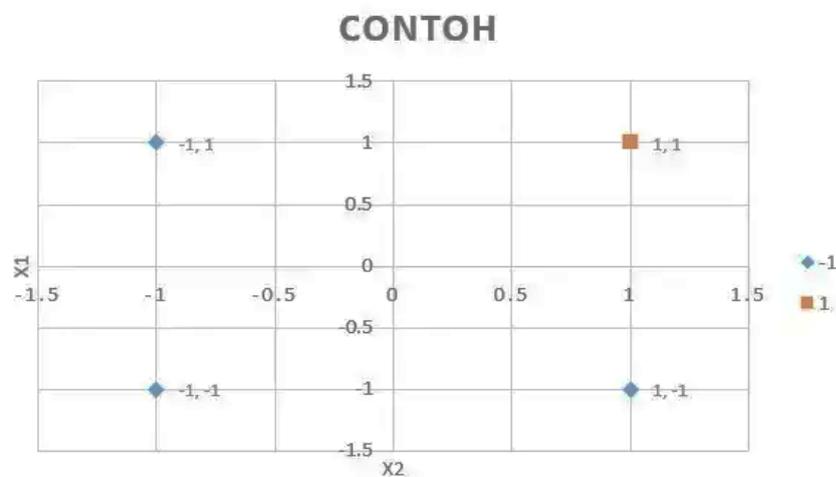
2.5 Support Vector Machine

Support Vector Machine adalah model pembelajaran terawasi dengan algoritme pembelajaran terkait yang menganalisis data untuk klasifikasi dan analisis regresi . Dikembangkan di AT&T Bell Laboratories oleh Vladimir Vapnik bersama rekan-rekannya (Boser et al., 1992, Guyon et al., 1993, Cortes and Vapnik , 1995, Vapnik et al., 1997 [rujukan?]) SVM adalah salah satu metode prediksi yang paling kuat, yang didasarkan pada kerangka pembelajaran statistik atau teori VC yang diusulkan oleh Vapnik (1982, 1995) dan Chervonenkis (1974). Diberikan satu set contoh pelatihan, masing-masing ditandai sebagai milik salah satu dari dua kategori, algoritma pelatihan SVM membangun sebuah model yang memberikan contoh baru untuk satu kategori atau yang lain, menjadikannya sebagai pengklasifikasi linier biner non- probabilistik (walaupun metode seperti Platt penskalaan ada untuk menggunakan SVM dalam pengaturan klasifikasi probabilistik). SVM memetakan contoh-contoh pelatihan ke titik-titik dalam ruang untuk memaksimalkan lebar celah antara kedua kategori tersebut. Contoh-contoh baru kemudian dipetakan ke dalam ruang yang sama dan diprediksi termasuk dalam kategori berdasarkan di sisi celah mana mereka jatuh.

Selain melakukan klasifikasi linier, SVM dapat secara efisien melakukan klasifikasi non-linier menggunakan apa yang disebut trik kernel , yang secara implisit memetakan inputnya ke dalam ruang fitur berdimensi tinggi. (Kantardzic, 2019).

Ketika data tidak diberi label, pembelajaran yang diawasi tidak dimungkinkan, dan diperlukan pendekatan pembelajaran yang tidak diawasi

, yang berupaya menemukan pengelompokan alami data ke dalam kelompok, dan kemudian memetakan data baru ke kelompok yang terbentuk ini. Algoritma pengelompokan vektor dukungan, dibuat oleh Hava Siegelmann dan Vladimir Vapnik, menerapkan statistik vektor dukungan, yang dikembangkan dalam algoritme mesin vektor dukungan, untuk mengkategorikan data yang tidak berlabel. Mengingat sifatnya yang kritis, hanya support vector inilah yang diperhitungkan untuk menemukan hyperplane yang paling optimal oleh SVM.



Gambar 2.3 Plot Contoh Data

Pada Gambar 2 menjelaskan bahwa terdapat 2 kelas yang terdiri dari -1 ditunjukkan dengan warna biru dan 1 ditunjukkan dengan warna orange. Pada masing-masing titik tersebut digunakan untuk mencari pemisah antara data positif dan data negatif. Penyelesaian dengan rumus 2.6 dan 2.7 sebagai berikut :

$$\frac{\min}{w} \tau(w) = \frac{1}{2} \|w\|^2 \quad (2.6)$$

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2) \quad (2.7)$$

Dengan syarat :

$$y_i(x_i \cdot w + b) - 1 \geq 0, \quad i = 1, 2, 3, \dots, n$$

$$y_i(x_1 \cdot w_1 + x_2 \cdot w_2 + b) \geq 1$$

sehingga ditemukan persamaan sebagai berikut:

$$(w_1 + w_2 + b) \geq 1 \text{ untuk } y_1 = 1, x_1=1, x_2=1$$

$$(-w_1 + w_2 - b) \geq 1 \text{ untuk } y_2 = -1, x_1=1, x_2=-1$$

$$(w_1 - w_2 - b) \geq 1 \text{ untuk } y_3 = -1, x_1=-1, x_2=1$$

$$(w_1 + w_2 - b) \geq 1 \text{ untuk } y_4 = -1, x_1=-1, x_2=-1$$

2.6 Evaluasi dan Validasi Metode Klasifikasi *Data mining*

Evaluasi dan validasi hasil klasifikasi dengan *data mining* pada penelitian ini digunakan metode *Cross Validation*, *Confusion matrix*, dan kurva ROC (*Receiver Operating Characteristic*).

1. *Confusion matrix*

Metode ini hanya menggunakan tabel matriks seperti pada Tabel 2.1, jika dataset hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007) Evaluasi dengan *Confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*. *Accuracy* dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi (Han dan Kamber, 2006). Sedangkan *precision* atau *confidence* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* atau *sensitivity* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar (Powers, 2011).

Tabel 2.1. Model *Confusion matrix*

<i>Correct Classification</i>	<i>Classified as</i>	
	+	-
+	<i>True positives</i>	<i>False negatives</i>
-	<i>False positives</i>	<i>True negatives</i>

Sumber: Han dan Kamber, 2006

True Positive adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah *record negative* yang diklasifikasikan sebagai positif, *false negative* adalah jumlah *record* positif yang diklasifikasikan sebagai negative, *true negative* adalah

jumlah *record negative* yang diklasifikasikan sebagai *negative*, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *Confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *Specifity*, *precision*, dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah t_{pos} terhadap jumlah *record* yang positif sedangkan *Specifity*, *precision* adalah perbandingan jumlah t_{neg} terhadap jumlah *record* yang *negative*. Untuk menghitung digunakan rumus 2.8, 2.9, 2.10 dan 2.11 di bawah ini (Han dan Kamber, 2006):

$$\text{Sensitifity} = \frac{t_{pos}}{pos} \quad (2.8)$$

$$\text{Specifity} = \frac{t_{neg}}{neg} \quad (2.9)$$

$$\text{Precision} = \frac{t_{pos}}{t_{pos}+f_{pos}} \quad (2.10)$$

$$\text{accuracy} = \text{Sensitivity} \frac{pos}{(pos+neg)} + \text{Specifity} \frac{neg}{(pos+neg)} \quad (2.11)$$

Keterangan :

t_{pos} = Jumlah *true positives*

t_{neg} = Jumlah *true negative*

p = Jumlah *record positives*

n = Jumlah *tupel negatives*

f_{pos} = Jumlah *false positives*

2. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *Confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positive* sebagai garis vertikal (Gorunescu, 2011). *The area under curve (AUC)* dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC digunakan dengan menggunakan rumus 2.12 dibawah ini:

$$\theta^r = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(xt^r, xj^r) \quad (2.12)$$

Dimana :

$$(X,Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases} \quad (2.13)$$

Keterangan :

X = Output positif

Y = Output negatif

2.7 Tinjauan Studi

Berikut adalah ringkasan dari beberapa penelitian sebelumnya yang terkait dengan klasifikasi :

1. Penelitian yang dilakukan oleh (Pritom, Sabab, Munshi, & Shihab, 2016) yang berjudul “*Predicting Breast Cancer Recurrence Using Effective Classification and Feature Selection Technique*”, peneliti menggunakan perbandingan 3 metode klasifikasi untuk mengukur tingkat akurasi yaitu *SVM*, *Naïve Bayes* dan *Decision Tree C4.5* dalam memprediksi kanker. Hasil menunjukkan bahwa dari ketiga metode klasifikasi, *SVM* memiliki tingkat akurasi yang lebih tinggi dari metode yang lain sebesar 75.75%, diikuti oleh *Decision Tree C.45* sebesar 73.73% dan *Naïve Bayes* sebesar 67.17%
2. *Jose Augusto Duarte Guterres* melakukan penelitian dengan judul : “Kelayakan Algoritma C45 Sebagai Pendukung Keputusan Dalam Pengajuan Penerima Beasiswa”. Tujuan dari penelitian ini untuk menguji kelayakan algoritma C4.5 sebagai pendukung keputusan dalam pengajuan penerima beasiswa pada STIKOM Artha Buana. Metode yang digunakan adalah metode klasifikasi menggunakan algoritma C45 dengan kriteria antara lain Tidak memiliki kendaraan pribadi, jarak tempat tinggal jauh ± 2 Km dari kampus, IPK minimal 2,75, belum pernah mendapatkan beasiswa dan penghasilan orang tua < 1 juta. Hasil dari penelitian adalah menentukan mahasiswa sebagai calon yang seharusnya diajukan oleh Kampus Stikom Artha Buana Kupang untuk

menerima beasiswa pendidikan berdasarkan hasil perhitungan algoritma C45.

3. Ika Menarianti melakukan penelitian dengan judul : “Klasifikasi *Data mining* Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi”. Metode yang digunakan dalam penelitian ini adalah *Cross Validation*, *confussion matrix*, *ROC curve* dan *T-test*. Hal ini dilakukan untuk melihat sejauh mana perbedaan data setelah dan sebelum dilakukan preprocessing data. Untuk menentukan klasifikasi yang digunakan pada suatu masalah diperlukan cara sistematis untuk mengevaluasi bagaimana metode yang bekerja dan membandingkannya dengan yang lain. Klasifikasi *data mining* yang digunakan adalah *Logistic Regression*, *Discriminant Analys*, *K-Nearest Neighbour*, *Naive Bayes*, *Decision Tree*, *Neural Network* dan *Support Vector Machine*. Dari hasil analisis komparasi dengan menggunakan *Cross Validation*, *Confusion matrix*, *ROC curve* dan *T-test* pada beberapa algoritma klasifikasi *data mining* dapat disimpulkan bahwa algoritma yang paling akurat adalah algoritma *Logistic Regression*. karena memiliki nilai akurasi tertinggi yaitu 87,41% dengan uji *T-test* paling dominan terhadap algoritma lainnya, dengan nilai AUC paling tinggi yaitu 1.000. *Algoritma Neural Network* walaupun nilai AUC-nya kecil yaitu 0.565 tetapi setelah dilakukan uji T- Test, algoritma ini memiliki sifat yang dominan dengan nilai akurasi cukup tinggi yaitu 86,73% sehingga dapat disimpulkan bahwa algoritma ini cukup akurat. *Algoritma Discriminant Analysis* dan *Decision tree* meskipun memiliki akurasi dan nilai AUC yang tinggi yaitu 87,41 % dan 1.000, tetapi berdasarkan uji *T-test* bukan merupakan algoritma yang dominan namun masih cukup baik untuk kasus penentuan pemberian kredit. Algoritma yang memiliki kinerja kurang memuaskan adalah *Support Vector Machine*, meskipun nilai akurasinya sebesar 86,39% dan *Naive Bayes* dengan tingkat akurasi sebesar 83,56%. Sedangkan *K-Nearest Neighbor* merupakan algoritma dengan nilai paling rendah yaitu dengan tingkat akurasi sebesar 76,71%.

4. Panny Agustia Rahayuningsih melakukan penelitian dengan judul : “Komparasi Algoritma Klasifikasi *Data mining* Untuk Memprediksi Tingkat Kematian Dini Kanker Dengan *Dataset Early Death Cancer*” Penelitian ini, penulis menggunakan dataset dari *Public Health England, Health Profile Datasheets 2013*. Penulis peroleh data tersebut dari <http://www.apho.org.uk/resource/browse.aspx?RID=126684>. Public Health England merupakan suatu lembaga yang bergerak dibidang kesehatan.

Dari hasil penelitian dengan menggunakan alat bantu rapid miner ini menunjukkan algoritma Neural Network dengan performa terbaik yang mendapat nilai akurasi sebesar 98.35%. Algoritma *Random Forest* dengan nilai sebesar akurasi 96.42%, algoritma *Naïve Bayes* dengan nilai sebesar akurasi 86.23%. algoritma C4.5 dengan nilai akurasi sebesar 69.97%. sedangkan untuk algoritma K-NN mendapatkan nilai akurasi sebesar 43.80%.

5. Harry Dhika melakukan penelitian dengan judul : Kajian Komparasi Penerapan Algoritma C4.5, *Naïve Bayes*, Dan *Neural Network* Dalam Pemilihan Mitra Kerja Penyedia Jasa Transportasi: Studi Kasus Cv. Viradi Global Pratama. Membahas mengenai pemilihan mitra kerja penyedia jasa transportasi menggunakan algoritma C4.5, *Naïve Bayes* dan *Neural Network* untuk memilih mitra kerja yang layak. Dari ketiga algoritma dilakukan pengujian dengan CRISP-DM dan hasilnya dilakukan komparasi untuk mengetahui model mana yang tingkat akurasinya paling tinggi untuk menentukan pemilihan mitra kerja penyedia jasa transportasi. Untuk mengukur ketiga kinerja algoritma tersebut digunakan *Confusion matrix* dan Kurva ROC dengan hasil bahwa algoritma yang memiliki nilai *accuracy* dan ROC paling tinggi adalah *Naïve Bayes* kemudian C4.5 dan Neural Network yang memiliki nilai sama tingkat *accuracy* dan ROC.

Ringkasan tinjauan studi dari beberapa penelitian sebelumnya bisa dilihat pada table 2.2 dibawah ini.

Tabel 2.2. Ringkasan Tinjauan Studi

No	Judul	Metode	Tujuan	Hasil
1.	“ <i>Predicting Breast Cancer Recurrence using effective Classification and Feature Selection technique</i> ” yang dilakukan oleh (Pritom, Sabab, Munshi, & Shihab, 2016).[3]	SVM, Naïve Bayes dan Decision Tree C4.5	Untuk memprediksi kambuhnya kanker payu dara	Hasil menunjukkan bahwa dari ketiga metode klasifikasi, SVM memiliki tingkat akurasi yang lebih tinggi dari metode yang lain sebesar 75.75%, diikuti oleh <i>Decision Tree C.45</i> sebesar 73.73% dan <i>Naïve Bayes</i> sebesar 67.17%
2.	“Kelayakan Algoritma C45 Sebagai Pendukung Keputusan Dalam Pengajuan Penerima Beasiswa” yang dilakukan oleh (Jose Augusto Duarte Guterres , 2015)[2]	Algoritma C4.5	Sebagai pendukung keputusan dalam pengajuan penerima beasiswa pada STIKOM Artha Buana.	Hasilnya adalah bahwa Algoritma C4.5 layak dijadikan sebagai pendukung keputusan dalam proses pengajuan penerima beasiswa pada Stikom Artah Buana Kupang karena hasil persentasi dari Algoritma C4.5 yang lebih tinggi daripada metode klasifikasi yang lainnya.
3.	“Klasifikasi <i>Data mining</i> Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi” yang dilakukan oleh (Ika Menarianti , 2015)[4]	<i>Cross Validation</i> , confussion matrix, ROC curve dan <i>T-test</i> . <i>Logistic Regression</i> , <i>Discriminant Analys</i> , <i>K-Nearest Neighbour</i> , <i>Naive Bayes</i> , <i>Decision Tree</i> , <i>Neural Network</i> dan <i>Support Vector Machine</i> .	Untuk membantu para analis kredit dalam menentukan pemberian kredit pada nasabah	Dari hasil analisis komparasi dengan menggunakan <i>Cross Validation</i> , <i>Confusion matrix</i> , <i>ROC curve</i> dan <i>T-test</i> pada beberapa algoritma klasifikasi <i>data mining</i> dapat disimpulkan bahwa algoritma yang paling akurat adalah algoritma <i>Logistic Regression</i> . karena memiliki nilai akurasi tertinggi yaitu 87,41% dengan uji <i>T-test</i> paling dominan terhadap algortima lainnya, dengan nilai AUC paling tinggi yaitu 1.000. Algoritma Neural Network walaupun nilai AUC-nya kecil yaitu 0.565 tetapi setelah dilakukan uji <i>T-test</i> , algoritma ini memiliki sifat yang dominan dengan

				<p>nilai akurasi cukup tinggi yaitu 86,73% sehingga dapat disimpulkan bahwa algoritma ini cukup akurat. Algoritma <i>Discriminant Analysis</i> dan <i>Decision tree</i> meskipun memiliki akurasi dan nilai AUC yang tinggi yaitu 87,41 % dan 1.000, tetapi berdasarkan uji <i>T-test</i> bukan merupakan algoritma yang dominan namun masih cukup baik untuk kasus penentuan pemberian kredit. Algoritma yang memiliki kinerja kurang memuaskan adalah <i>Support Vector Machine</i>, meskipun nilai akurasinya sebesar 86,39% dan Naive Bayes dengan tingkat akurasi sebesar 83,56%. Sedangkan <i>K-Nearest Neighbor</i> merupakan algoritma dengan nilai paling rendah yaitu dengan tingkat akurasi sebesar 76,71%.</p>
4.	<p>“Komparasi Algoritma Klasifikasi <i>Data mining</i> Untuk Memprediksi Tingkat Kematian Dini Kanker Dengan <i>Dataset Early Death Cancer</i>” Yang Dilakukan Oleh (Panny Agustia Rahayuningsih, 2019)[5]</p>	<p><i>Desecion Tree, Naïve Bayes, k-Nearset Neighbour, Random Forest</i> dan <i>Neural Network</i></p>	<p>Untuk memprediksikan tingkat kematian dini kanker pada penduduk Eropa</p>	<p>Hasil penelitian dari pengujian metode menggunakan <i>Confusion matrix</i> menghasilkan performe terbaik dengan nilai akurasi sebesar 98,35% adalah <i>algoritma Neural Network</i>.</p> <p>Urutan algoritma yang dihasilkan pada pengujian menggunakan uji beda <i>t-test</i> mendapatkan hasil model algoritma yang terbaik adalah <i>Random Forest dan Neural Network, algoritma Naïve Bayes</i> lumayan baik, algoritma C4.5 cukup baik sedangkan untuk algoritma K-NN algoritma yang kurang baik digunakan pada dataset ini.</p>
5.	<p>“Kajian Komparasi Penerapan Algoritma C4.5, <i>Naïve Bayes</i>, Dan <i>Neural Network</i> Dalam Pemilihan Mitra Kerja Penyedia Jasa Transportasi: Studi Kasus CV. Viradi Global Pratama” dilakukan oleh (Harry Dhika, 2015)[6]</p>	<p>Algoritma C4.5, <i>Naïve Bayes</i>, dan <i>Neural Network</i></p>	<p>Untuk mencari mitra perusahaan dengan prediksi <i>very good, good</i> dan <i>bad</i>.</p>	<p>Hasil bahwa algoritma yang memiliki nilai <i>accuracy</i> dan ROC paling tinggi adalah Algoritma C4.5 kemudian <i>Naïve Bayes</i> dan terakhir diikuti oleh <i>Neural Network</i>. Dengan begitu dapat disimpulkan bahwa metode yang paling baik adalah algoritma C4.5 dalam mengklasifikasikan data.</p>

2.8 Tinjauan Obyek Penelitian

Universitas Muhammadiyah Pringsewu (UMPRI) Lampung merupakan lembaga pendidikan tinggi yang berdiri berkat proses penggabungan tiga Perguruan Tinggi Muhammadiyah di Kabupaten Pringsewu yaitu STKIP Muhammadiyah Pringsewu Lampung, STIE Muhammadiyah Pringsewu Lampung, dan STIKes Muhammadiyah Pringsewu Lampung. Penggabungan tiga perguruan tinggi Muhammadiyah yaitu STKIP Muhammadiyah Pringsewu Lampung, STIE Muhammadiyah Pringsewu Lampung, dan STIKes Muhammadiyah Pringsewu Lampung menjadi Universitas Muhammadiyah Pringsewu merupakan salah satu bentuk amanah yang dinyatakan pada acara Musyawarah Wilayah Muhammadiyah Tahun 2010. Upaya untuk memperlancar proses persiapan penggabungan ini dibentuklah tim kepanitiaan yang ketua dan anggotanya berasal dari ketiga perguruan tinggi Muhammadiyah. Tim kepanitiaan yang terbentuk mengalami fase tiga kali perubahan komposisi, hal ini terjadi dikarenakan progres capaian yang diharapkan tidak tercapai, diawali Kepanitiaan tahun 2010, diperbaiki kembali dengan SK Pimpinan Wilayah Muhammadiyah Provinsi Lampung Nomor 21/KEP/II.0/D/2015 tentang Panitia Pendirian Universitas Muhammadiyah Pringsewu Lampung tanggal 21 Oktober 2015. Karena progress capaian belum membuahkan hasil maka munculah SK Pimpinan Wilayah Muhammadiyah Provinsi Lampung Nomor 22/KEP/II.0/D/2017 tentang Perubahan Panitia Pendirian Universitas Muhammadiyah Pringsewu Lampung tanggal 4 Desember 2017, kepanitiaan ini dipimpin oleh bapak Drs. Wanawir, Am., M.M., M.Pd. Surat Keputusan ini baru terealisasi pelaksanaannya di awal tahun 2018 dengan tim inti yang terlibat dalam proses penyusunan dan pemrosesan dokumen instrumen usulan pengajuan Program Studi baru serta dokumen pendukung lainnya yaitu Bapak Drs. H. Wanawir, Am., M.M., M.Pd., ibu Dr. Tri Yuni Hendrowati, M.Pd., Bapak Nurfaizal, M.Pd., dan Ibu Rahma Faelasofi, M.Sc., dibersamai oleh para

Ketua Perguruan Tinggi Muhammadiyah pada saat itu Bapak Drs. H. A. Rahman, M.M, M.Pd., Bapak Sapto Yuwono, S.Pd., M.M., Ibu Ns. Arena Lestari, M.Kep., Sp. Kep.J. , dan seluruh kepanitiaan yang ada baik secara langsung maupun tidak langsung.

Tujuan penggabungan dan perubahan bentuk ini adalah: (1) agar pengelolaan lebih efektif dan efisien, karena secara geografis terletak dalam satu kompleks; (2) ketiga perguruan tinggi tersebut adalah milik persyarikatan muhammadiyah, sehingga sangat memungkinkan berada dalam satu manajemen universitas; (3) untuk meningkatkan kualitas dosen dan output lulusan, karena berada dalam satu sistem pengelolaan; (4) untuk memperkuat sistem pengelolaan keuangan yang kuat dan sehat; (5) menjamin keberlangsungan institusi yang kuat dan berdaya saing global; serta (6) memperkuat kelembagaan dari adanya persaingan perguruan tinggi asing yang beroperasi di NKRI. Atas dasar tersebut, maka sesuai dengan aturan dalam Permenristekdikti No. 100 Tahun 2016 tentang Pendirian, Perubahan, Pembubaran PTN, dan Pendirian, Perubahan, Pencabutan izin PTS. Penggabungan perguruan tinggi ini menjadi Universitas Muhammadiyah Pringsewu Lampung, sesuai dengan ketentuan pada Persyaratan dan Prosedur Penggabungan dan Penyatuan Perguruan Tinggi Swasta, maka kami merencanakan penambahan Program Studi yang berada dalam rumpun keilmuan Sains, Teknologi, Engenering, dan Matematika (STEM) yang meliputi: (1) Program Studi peternakan (S1); (2) Program Studi kehutanan (S1); (3) Program Studi agronomi (S1); (4) Program Studi teknologi hasil pertanian (S1); (5) Program Studi akuakultur (S1); (6) Program Studi matematika (S1); (7) dan Program Studi pendidikan IPA (S1).

Pengajuan Program Studi tersebut tentunya didasari atas beberapa alasan diantaranya: (1) ada salah satu Program Studi yang diajukan belum memiliki kompetitor Program Studi yang sejenis pada perguruan tinggi negeri di Lampung; (2) kuota daya tampung mahasiswa baru pada perguruan tinggi negeri yang terbatas namun jumlah peminat yang tinggi terhadap Program Studi

tersebut; (3) jumlah peminat yang tinggi pada Program Studi tersebut diharapkan dapat membuka peluang bagi Program Studi baru yang diajukan untuk mendapatkan peminat dari para calon mahasiswa baru; serta (4) lokasi perguruan tinggi yang strategis pada Kabupaten Pringsewu diharapkan mendapatkan nilai lebih yang dapat dipertimbangkan oleh calon mahasiswa baru terutama yang berdomisili di wilayah sekitar Kabupaten Pesawaran, Lampung Tengah, Tanggamus, Pesisir Barat, Lampung Barat, Way Kanan, serta luar Provinsi Lampung.

Berikut ini data yang terkait dengan penetapan pengelolaan penggabungan tiga perguruan tinggi Muhammadiyah:

1. STKIP Muhammadiyah Pringsewu Lampung (STKIP MPL), perijinannya tertera pada Piagam Pendirian Perguruan Tinggi Muhammadiyah Nomor 024/III.LP.75/79 pada tahun 1975. Perkembangan Program Studi STKIP MPL sampai di tahun 2017 memiliki 6 Program Studi aktif keguruan dengan ijin operasional sebagai berikut:
 - a. Program Studi Bimbingan dan Konseling, perijinan Program Studi tersebut tertera sesuai hasil surat keputusan Kementerian Pendidikan dan Kebudayaan Koordinasi Perguruan Tinggi Swasta Wilayah II Nomor: 12181/D/T/K-II/2012 perihal perpanjangan ijin Program Studi Bimbingan dan Konseling Jenjang S-1. Selain itu, Program Studi Bimbingan dan Konseling saat ini terakreditasi B, sesuai surat keputusan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Nomor: 1145/SK/BAN-PT/Akred/S/XI/2015.
 - b. Program studi Pendidikan Bahasa dan Sastra Indonesia, perijinan program studi tersebut tertera sesuai hasil surat keputusan Kementerian Pendidikan dan Kebudayaan Koordinasi Perguruan Tinggi Swasta Wilayah II Nomor: 13427/D/T/K-

- II/2012 perihal perpanjangan ijin Program Studi Pendidikan Bahasa dan Sastra Indonesia Jenjang S-1. Selain itu, Program Studi Pendidikan Bahasa dan Sastra Indonesia saat ini terakreditasi B, sesuai surat keputusan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Nomor: 0883/SK/BAN-PT/Akred/S/III/2017.
- c. Program Studi pendidikan matematika, perijinan Program Studi tersebut tertera sesuai hasil surat keputusan Kementerian Pendidikan dan Kebudayaan Koordinasi Perguruan Tinggi Swasta Wilayah II Nomor: 11431/D/T/K-II/2012 perihal perpanjangan ijin Program Studi Pendidikan Matematika Jenjang S-1. Selain itu, Program Studi Pendidikan Matematika saat ini terakreditasi B, sesuai surat keputusan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Nomor: 0442/SK/BAN-PT/Akred/S/V/2016.
 - d. Program Studi pendidikan bahasa Inggris, perijinan Program Studi tersebut tertera sesuai hasil surat keputusan Kementerian Pendidikan dan Kebudayaan Koordinasi Perguruan Tinggi Swasta Wilayah II Nomor: 8087/D/T/K-II/2011 perihal perpanjangan ijin Program Studi Pendidikan Bahasa Inggris Jenjang S-1. Selain itu, Program Studi Pendidikan Bahasa Inggris saat ini terakreditasi B, sesuai surat keputusan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Nomor: 0441/SK/BAN-PT/Akred/S/V/2016.
 - e. Program Studi pendidikan guru dan sekolah dasar, perijinan Program Studi tersebut tertera sesuai hasil surat keputusan Kementerian Riset, Teknologi, dan Pendidikan Tinggi Nomor: 4566/A4/HK/2016 perihal pembukaan Program Studi pendidikan guru sekolah dasar program sarjana.
 - f. Program Studi pendidikan tari, perijinan Program Studi tersebut

tertera sesuai hasil surat keputusan Kementerian Riset, Teknologi, dan Pendidikan Tinggi Nomor : 1029/A4/HK/2017 perihal pembukaan Program Studi pendidikan tari program sarjana. Pada saat penggabungan dilakukan STKIP MPL terakreditasi B, sesuai dengan surat keputusan BAN-PT Nomor: 4047/SK/BAN-PT/Akred/PT/X/2017.

2. STIKES Muhammadiyah Pringsewu Lampung, perizinannya tertera pada surat keputusan Menteri Pendidikan Republik Indonesia Nomor: 254/D/O/2008 pada tahun 2008, surat keputusan tersebut menyatakan pemberian ijin penyelenggaraan Program Studi keperawatan (S1) dan penggabungan Akademi Keperawatan Muhammadiyah Pringsewu dengan Akademi Kebidanan Muhammadiyah Pringsewu di Lampung menjadi Sekolah Tinggi Ilmu Kesehatan Muhammadiyah Pringsewu di Lampung diselenggarakan oleh BPH-PT Muhammadiyah Pringsewu Lampung. Perkembangan Program Studi STIKES Muhammadiyah Pringsewu Lampung sampai di tahun 2017 memiliki 3 Program Studi aktif dengan ijin operasional sebagai berikut:
 - a. Program Studi Keperawatan (S1), perijinan Program Studi tersebut tertera sesuai hasil surat keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor : 254/D/O/2008 perihal perpanjangan ijin Program Studi Keperawatan Jenjang S-1.
 - b. Program Studi Keperawatan (DIII), perijinan Program Studi tersebut tertera sesuai hasil surat keputusan Menteri Kesehatan Republik Indonesia Nomor: HK.00.06.1.3.3768 tentang izin penyelenggaraan Akademi Keperawatan Muhammadiyah Pringsewu Provinsi Lampung dengan menetapkan penyelenggaraan program Diploma III Keperawatan.
 - c. Program Studi Kebidanan (DIII), perijinan Program Studi tersebut

tertera sesuai hasil surat keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor : 59/D/O/2007 perihal pemberian ijin penyelenggaraan Program Studi Kebidanan jenjang program Diploma III (DIII).

- d. Program Studi Profesi Ners, perijinan Program Studi tersebut tertera sesuai hasil surat keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor :81/E/O/2013 perihal ijin penyelenggaraan Program Studi Profesi Ners.
3. STIE Muhammadiyah Pringsewu Lampung, pada perguruan tinggi ini memiliki satu Program Studi aktif yaitu Program Studi manajemen dengan izin operasional sesuai surat keputusan Kementerian Pendidikan dan Kebudayaan Koordinasi Perguruan Tinggi Swasta Wilayah II Nomor; 12644/D/T/K-II/2012 mengenai perpanjangan ijin Program Studi manajemen jenjang S-1 di Sekolah Tinggi Ilmu Ekonomi Muhammadiyah Pringsewu Lampung (STIE MPL). Program Studi Manajemen terakreditasi B sesuai sertifikat akreditasi BAN-PT Nomor: 0705/SK/BAN-PT/Akred/S/III/2017. Sedangkan untuk institusi, STIE MPL terakreditasi dengan peringkat C sesuai sertifikat akreditasi BAN-PT Nomor: 872/SK/BAN-PT/Akred/PT/VIII/2015.

Selanjutnya dalam proses pengurusan penggabungan, dari tim inti sudah merampungkan tahap 1 yang upload berkas persyaratan tahap 1 pada bulan April 2018. Selanjutnya hasil dari upload berkas tahap 1 ada beberapa berkas perbaikan yang disarankan oleh Kemenristekdikti untuk diperbaiki, salah satunya terkait dengan perubahan akta kepemilikan tanah menjadi kepemilikan persyarikatan. Untuk perbaikan yang disarankan sudah diperbaiki oleh tim inti.

Pada 21 September 2018 tim panitia persiapan penggabungan mendapatkan undangan dari Kelembagaan Kemenristekdikti untuk menghadiri bimbingan teknis terkait kemajuan penyusunan borang Program Studi baru untuk mempersiapkan upload di tahap 2 yang meminta borang penggabungan

perguruan tinggi (institusi) dan borang Program Studi baru diupload. Sampai tahapan pengupload tahap dua ini masih berlangsung proses pengupload data yang diminta terutama terkait dengan borang dan lampirannya, proses tersebut masih berlangsung sambil menunggu hasil perbaikan borang dan lampiran oleh tim selesai dikerjakan.

Pada tanggal 12 Oktober 2018, tim inti persiapan penggabungan dan para pimpinan tiga PTM mendapat undangan untuk pemaparan/presentasi pengajuan penggabungan perguruan tinggi dan penambahan Program Studi baru oleh L2-Dikti, hanya saja terdapat kendala yang dihadapi akibat adanya SK Pengangkatan Rektor Universitas Muhammadiyah Pringsewu Lampung yang dianggap menyalahi prosedur oleh L2-Dikti, L2 Dikti berpendapat bahwa seyogyanya SK Rektor tersebut belum ada karena lembaga/universitasnya belum ada, sedang berproses. Hal ini mengakibatkan proses pemaparan/presentasi pengajuan penggabungan perguruan tinggi dan penambahan Program Studi baru ditunda sampai L2-Dikti mendapatkan surat keterangan pencabutan dari SK Pengangkatan Rektor Universitas Muhammadiyah Pringsewu Lampung yang dinilai keliru tersebut. Setelah proses pencabutan SK Rektor Universitas Muhammadiyah Pringsewu Lampung selesai dilakukan, L2-Dikti kembali berproses dengan memanggil tim untuk memaparkan kembali proses pengajuan penggabungan perguruan tinggi dan penambahan Program Studi baru dengan mengundang tim yaitu BPH UMPRI, Pimpinan Perguruan Tinggi Muhammadiyah, dan tim inti penggabungan UMPRI. Proses pemaparan penggabungan perguruan tinggi Muhammadiyah ini berhasil dilakukan sehingga di awal Januari 2019 terbitlah Surat Rekomendasi dari L2 -Dikti Nomor 104/L2/KL/2019 tentang Rekomendasi Penggabungan, Perubahan Bentuk PTS, dan Penambahan Program Studi tanggal 17 Januari 2019 yang berisikan rekomendasi dilakukannya penggabungan STKIP-STIKes-STIE Muhammadiyah Pringsewu Lampung menjadi Universitas Muhammadiyah Pringsewu Lampung.

Setelah Surat Rekomendasi diperoleh, tim inti mulai berproses dari awal kembali dengan mengikuti acuan peraturan penggabungan perguruan tinggi terbaru versi tahun 2019. Pada awal bulan Februari tim inti mulai mengupload tujuh instrumen pengajuan Program Studi baru yaitu Prodi Akuakultur, Prodi Matematika, Prodi Agronomi, Prodi Kehutanan, Prodi Peternakan, Prodi Teknologi Hasil Pertanian, dan Prodi pendidikan IPA serta instrumen penggabungan menjadi universitas beserta kelengkapan dokumen lainnya, lima belas hari kerja proses penantian verifikasi dari tim L2-Dikti dan tim Kemristekdikti hasilnya pada tahapan ini semua pengajuan 7 Prodi baru masih ditolak karena ada ketidaksesuaian SDM dosen yang lulusannya tidak linear dengan Prodi yang diajukan serta masih adanya ketidaktepatan dalam pengisian instrumen borang. Dari hasil tersebut tim melanjutkan revisi dokumen instrumen borang dan kembali mencari SDM dosen yang linear sesuai catatan dari tim reviewer.

Pada bulan Maret 2019, tim mulai mengupload ulang hasil revisi tujuh instrumen pengajuan Program Studi baru beserta kelengkapan lampiran pendukung sesuai dengan Program Studinya. Proses ini pun selesai di tanggal 24 April 2019, hingga 15 hari kerja sudah mendapat balasan dari tim reviewer L2-Dikti dan Kemristekdikti yang menyatakan dari tujuh Program Studi tersebut ada 2 Program Studi yang sepenuhnya disetujui dengan kriteria SDM dosen di Prodi terpenuhi dengan linear dibidang keilmuannya dan ketepatan dalam instrumen pengajuan Prodi baru, 2 Program Studi tersebut yaitu Prodi Akuakultur dan Prodi Matematika.

Proses disetujuinya dua Prodi tersebut, berakibat mempersiapkan tahapan selanjutnya yaitu visitasi lapangan oleh tim visit dari Kemristekdikti, untuk mempersiapkan visitasi, tim panitia mempersiapkan kelengkapan fasilitas Prodi akuakultur dan Prodi matematika, mulai dari ruang perkuliahan, ruang Prodi, ruang laboratorium, ruang perpustakaan Prodi, banner visi misi perguruan tinggi dan Prodi baru. Ketersiapan ini dilakukan untuk pelaksanaan visitasi

yang berlangsung di tanggal 22 dan 23 Juni 2019. Setelah proses visitasi selesai dilaksanakan, tim panitia yang dipimpin oleh Bapak Drs. H. Wanawir, Am., M.M., M.Pd. menanti hasil SK Penggabungan STKIP-STIE-STIKes Muhammadiyah Pringsewu Lampung menjadi Universitas Muhammadiyah Pringsewu Lampung.

Alhamdulillah berkat dukungan semua pihak yaitu Pimpinan Pusat Muhammadiyah, L2-Dikti Wilayah II, Pimpinan Wilayah Muhammadiyah Provinsi Lampung, Pimpinan Daerah Muhammadiyah Kabupaten Pringsewu, Ketua STKIP-STIE-STIKes Muhammadiyah Pringsewu Lampung, tim panitia persiapan penggabungan Universitas Muhammadiyah Pringsewu Lampung, dan sivitas akademik STKIP-STIE-STIKes Muhammadiyah Pringsewu Lampung, terbit SK Kemristekdikti Nomor 654/KPT/I.2019 tentang penggabungan STKIP-STIE-STIKes Muhammadiyah Pringsewu Lampung menjadi Universitas Muhammadiyah Pringsewu Lampung tanggal 29 Juli 2019 yang ditandatangani oleh Pejabat Kemristekdikti Bapak Ainun Na'im.

1. Visi

“Terwujudnya universitas unggul, islami, dinamis, dan berorientasi global pada tahun 2035”.

2. Misi

- a) Menyelenggarakan pendidikan, penelitian, dan pengabdian kepada masyarakat yang berkualitas.
- b) Mengaktualisasikan nilai-nilai Al Islam Kemuhammadiyah dalam kehidupan bermasyarakat.
- c) Melaksanakan kerjasama dalam bidang ilmu pengetahuan, teknologi, seni, dan/atau olahraga.
- d) Melaksanakan tata kelola universitas sesuai prinsip *Good University Governance* (GUG)

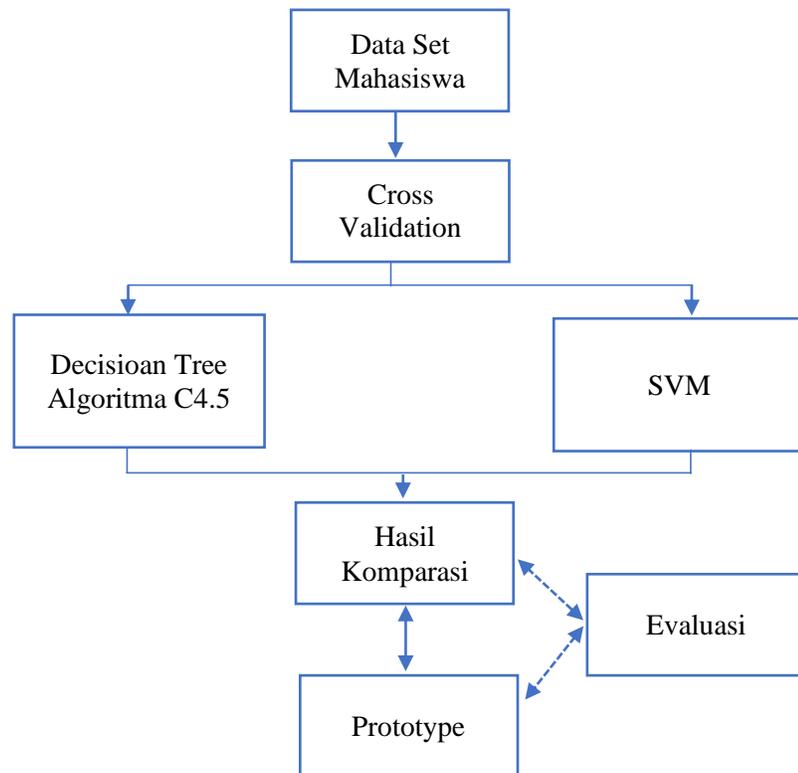
2.9 Beasiswa

Pemberian beasiswa merupakan sebagian program kerja rutin yang ada di setiap institusi pendidikan yang pada umumnya berupa beasiswa prestasi dan beasiswa kurang mampu. Program ini dapat diselenggarakan oleh pihak pemerintah maupun swasta yaitu dengan berupa pemberian sejumlah dana bantuan untuk menunjang biaya atau ongkos yang harus dikeluarkan oleh mahasiswa selama menempuh masa pendidikan. Dengan adanya bantuan ini, diharapkan mahasiswa dapat menyelesaikan pendidikannya tanpa ada gangguan terutama yang berhubungan dengan keuangan mahasiswa hingga tuntas atau lulus di jenjang pendidikan.

Berdasarkan UU No 12/2012[8] tentang Pendidikan Tinggi, Pemerintah Indonesia berkewajiban meningkatkan akses dan kesempatan belajar di Perguruan Tinggi serta menyiapkan insan Indonesia yang cerdas dan kompetitif. Oleh karena itu Pemerintah akan selalu berupaya untuk menjamin, bahwa anak Indonesia yang kurang mampu terutama yang memiliki prestasi akan dapat terus menempuh pendidikan hingga jenjang kuliah melalui Program Indonesia Pintar (PIP).

2.10 Kerangka Konsep / Pola Pikir Pemecahan

Pola pikir dapat dibuat dalam kerangka seperti pada gambar 2.4 sebagai berikut:



Gambar 2.4. Kerangka Konsep/Pola Pikir

Pola pikir yang tergambar di atas dapat dijelaskan sebagai berikut :

1. Pada proses input, data yang digunakan adalah dataset beasiswa KIP.
2. Pada bagian *Cross Validation*, dataset dibagi menjadi data *testing* dan data *training* yang selanjutnya akan dilakukan proses *modeling* dan proses *evaluasi*.
3. Pada proses *modeling*, data *training* akan diolah sesuai hasil komparasi algoritma C4.5 dengan *Support Vector Machine (SVM)*. Kemudian akan dibagi menjadi dua kelompok yaitu layak dan tidak layak.
4. Untuk mengukur peningkatan akurasi dari hasil masing-masing validasi, digunakan *Confusion matrix*, *Confusion matrix* ini merupakan matrik 2

dimensi yang menggambarkan perbandingan antara hasil prediksi dengan kenyataan.

5. Untuk penerapannya, akan dibuat sebuah aplikasi dengan rule berdasarkan hasil komparasi *algoritma C4.5* dengan *Support Vector Machine (SVM)*. Untuk menentukan keputusan pemberian beasiswa.

2.11 Hipotesis

Dari kedua klasifikasi *data mining* yang diuji yaitu *algoritma C4.5* dengan *Support Vector Machine (SVM)*. Dapat diambil hipotesis sebagai berikut :

1. Dengan menggunakan *Algoritma C4.5* dan *Support Vector Machine (SVM)*, diduga dapat digunakan untuk penentuan penerimaan beasiswa KIP-Kuliah di Universitas Muhamadiyah Pringsewu.
2. Diduga *algoritma C4.5* memiliki nilai akurasi yang lebih tinggi dibandingkan metode *Support Vector Machine (SVM)*.