

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Beberapa penelitian terdahulu yang berkaitan dengan penelitian ini, dan dijadikan acuan dalam penelitian ini dituangkan dalam tabel 2.1 dibawah ini.

Tabel 2.1 Penelitian Terdahulu

No	Judul /Penulis /Tahun	Dataset	Metode	Hasil
1	Implementasi Algoritma C4.5 Untuk Prediksi Kelulusan Tepat Waktu (Studi Kasus: Program Studi Teknik Informatika) Sarah Novia 2019	Dataset yang digunakan peneliti bersumber dari data mahasiswa lama dalam bentuk tabel yang berisi beberapa atribut meliputi IP semester 1-7 dan nilai BTQ di semester 6, diklasifikasikan berdasarkan target yang akan ditentukan	Setelah diklasifikasikan, data target dihitung menggunakan metode Decision Tree yaitu algoritma C4.5 untuk mencari nilai <i>Entropy</i> , <i>Information Gain</i> , <i>Gain Ratio</i> , dan <i>Split Information</i> , setelah proses perhitungan selesai maka akan menghasilkan rule atau kondisi yang digunakan dalam penentuan keputusan pada proses prediksi. Output dari sistem ini yaitu keterangan lulus tepat waktu atau tidaknya tiap mahasiswa yang diprediksi.	Dari hasil ujiacoba menggunakan 100 data sampel dari mahasiswa lulusan Teknik Informatika tahun 2015-2018, pola yang dibentuk mempunyai akurasi kecocokan sebesar 68.42 %, nilai presisi sebesar 42.86 %, dan penarikan sebesar 60%. Semakin banyak data sampel yang digunakan maka semakin kuat pula kebenaran prediksinya

No	Judul /Penulis /Tahun	Dataset	Metode	Hasil
2	Implementasi Algoritme <i>Support Vector Machine</i> (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa Arif Pratama 2018	Dataset yang digunakan peneliti bersumber dari data prediksi ketepatan waktu kelulusan mahasiswa ini menggunakan <i>dataset</i> dari bagian akademik Filkom yang berjumlah 188 data.	Terdapat beberapa macam pengujian, yaitu: pengujian menggunakan jumlah data latih yang berbeda, pengaruh nilai variabel atau parameter pada Sequential Training SVM, pengujian pengaruh penggunaan kernel. Pada pengujian menggunakan data latih perbandingan yang berbeda. Jumlah data latih yang digunakan antara lain 94, 113, 132, 150, dan 170 data. Sedangkan jumlah data uji yang digunakan untuk perbandingan antara lain 94, 75, 56, 38, dan 18 data. Nilai parameter Sequential Training yang digunakan tetap yaitu $\lambda = 0,5$, $\gamma = 0,01$, $c = 1$, IterasiMax = 100, dan $\epsilon = 0,0001$, dan menggunakan kernel RBF.	Dari 3 jenis kernel yang digunakan yaitu <i>Polynomial Degree</i> , <i>Gaussian RBF</i> , dan <i>Linier</i> Pengaruh Penggunaan Kernel pada Algoritme <i>Support Vector Machine</i> (SVM) yang paling baik yaitu menggunakan kernel <i>Gaussian RBF</i> karena kernel dengan rata-rata akurasi tertinggi adalah kernel <i>Gaussian RBF</i> yaitu 80,55%
3	Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode <i>Decision Tree</i> dan <i>Artificial Neural Network</i> Eko Prasetyo Rohmawan 2018	<i>Dataset</i> yang digunakan peneliti bersumber dari data mahasiswa Universitas Dehasen Bengkulu Program Studi Teknik Informatika tahun akademik 2009, 2010 dan 2011	Proses pengklasifikasian data menggunakan metode <i>Decision Tree</i> , Metode kedua yang digunakan adalah <i>Artificial Neural Network</i> . Dan algoritma yang digunakan adalah <i>Backpropagation</i> . Atribut masukan yang digunakan sama dengan atribut pada <i>Decision Tree</i>	Metode <i>Artificial Neural Network</i> memiliki tingkat akurasi sedikit lebih tinggi jika dibandingkan dengan metode <i>Decision Tree</i> , dengan akurasi sebesar 74,51% dan <i>Artificial Neural Network</i> sebesar 79,74%.

No	Judul /Penulis /Tahun	Dataset	Metode	Hasil
4	Komparasi <i>Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor</i> untuk Memprediksi Mahasiswa Lulus Tepat Waktu Agus Budiyantara 2020	Penelitian ini menggunakan data primer yang diperoleh dari arsip data BAAK STMIK Widuri yaitu data mahasiswa tahun akademik 2010/2011, 2011/2012, 2012/2013, dan 2013/2014 berjumlah 342 data kemudian sebanyak 242 data digunakan untuk data training dan 100 data digunakan untuk data testing	Penelitian ini dilakukan dengan mengkomparasi 3 metode yaitu <i>Decision Tree, Naive Bayes dan K-Nearest Neighbor</i> . Proses yang dilakukan adalah <i>training</i> dan <i>testing dataset</i> dengan menggunakan metode <i>Decision Tree, Naive Bayes dan K-Nearest Neighbor</i> untuk menghasilkan akurasi. Akurasi yang dihasilkan oleh ketiga metode tersebut kemudian dikomparasi untuk mendapatkan akurasi tertinggi	Dari hasil proses pengujian dengan tools <i>Rapid Miner</i> Menggunakan tiga metode yang telah dilakukan. <i>Decision Tree</i> (C4.5) memperoleh hasil akurasi tertinggi sebesar 98.04% pada pengujian ke 3. Metode <i>Naive Bayes</i> memperoleh hasil akurasi tertinggi sebesar 96.00% pada pengujian ke 4, dan Metode <i>K-Nearest Neighbor (K-NN)</i> memperoleh hasil akurasi tertinggi sebesar 90.00% pada pengujian ke 2.

Berdasarkan keempat penelitian di atas, maka yang menjadi keunikan dari penelitian ini terletak pada gaya penerapan metodologi, kasus dan hasil penelitian yang berbeda. Dimana pada dua dari empat penelitian diatas masih menggunakan satu metode, jumlah *dataset* yang lebih sedikit (dibawah 500) menggunakan fitur IPS dan IPK semester 1 sampai semester 7 dan nilai BTQ, sedangkan pada penelitian ini melakukan komparasi metode Algoritma C4.5 dan *Naive Bayes* menggunakan *dataset* diatas 500, menggunakan fitur IPK semester 5 dan menambahkan fitur status perkawinan dan pekerjaan untuk memprediksi kelulusan tepat waktu mahasiswa di Universitas Muhammadiyah Pringsewu Lampung.

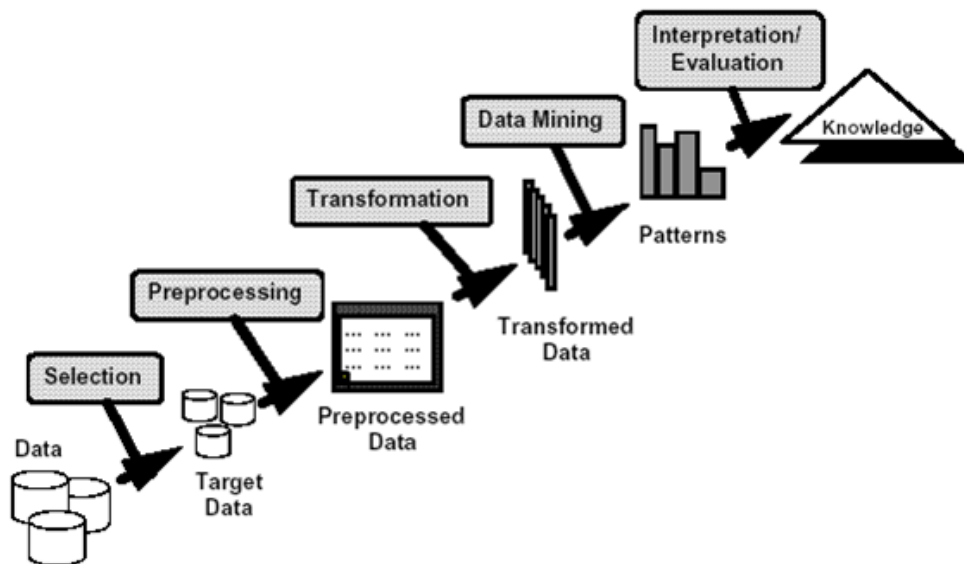
2.2 Data Mining

Menurut Daniel T. Larose (2004) *Data mining* adalah proses menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah sejumlah besar data

yang disimpan di repositori, menggunakan pengenalan pola teknologi serta teknik statistik dan matematika.

Menurut Efrain Turban (2005) *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, Aronson, and Liang 2005)

Menurut Kusri (2009) istilah *data mining* dan *knowledge discovery in database* sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Pemahaman kedua istilah tersebut memiliki konsep yang berbeda, akan tetapi berkaitan satu sama lain. Salah satu tahapan dalam keseluruhan proses *knowledge discovery in database* adalah *data mining*. *Knowledge discovery in database* secara garis besar dapat dijelaskan pada gambar 2.1 sebagai berikut.



Gambar 2.1 Tahapan *Knowledge Discovery in Databases*

1. *Data Selection*

Pemilihan (seleksi) dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *knowledge data discovery* (KDD) dimulai. Data

hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas terpisah dari basis data operasional.

2. *Preprocessing* atau *Cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus *knowledge data discovery*. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak juga dilakukan proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk *knowledge data discovery*, seperti data informasi.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam *knowledge data discovery* merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses *knowledge data discovery* secara keseluruhan.

5. *Interpretation* atau *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 *knowledge data discovery* yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada pada sebelumnya.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose 2005):

a. Deskripsi

Terkadang penelitian analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

b. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

c. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi di masa depan).

d. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategori. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

e. *Clustering*

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran.

f. Asosiasi

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

2.3 Klasifikasi

Klasifikasi berasal dari bahasa latin yaitu *classis* yang artinya pengelompokan benda yang sama serta memisahkan benda yang tidak sama. Secara harfiah arti klasifikasi adalah penggolongan, pengelompokan. Dalam kaitannya di dunia perpustakaan klasifikasi diartikan sebagai kegiatan pengelompokan bahan pustaka berdasarkan ciri-ciri yang sama, misalnya pengarang, fisik, isi dan sebagainya (Parteek Bhatia, 2019). Klasifikasi adalah suatu proses memilih dan mengelompokkan buku-buku perpustakaan atau bahan pustaka lainnya atas dasar tertentu serta diletakkannya secara bersama-sama disuatu tempat. Menurut istilah klasifikasi adalah proses membagi objek atau konsep secara logika kedalam klas-klas hirarki, subklas, dan sub-subklas berdasarkan kesamaan yang mereka miliki secara umum dan yang membedakannya. Klasifikasi secara umum juga diartikan sebagai kegiatan penataan pengetahuan secara universal kedalam beberapa susunan sistematis.

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 persamaan 2.1 sebagai berikut:

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

Keterangan :

S= Himpunan kasus

n = jumlah partisi S

P_i = proporsi S_i terhadap S

3. Kemudian hitung nilai gain menggunakan persamaan 2.2 sebagai berikut :

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

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 :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam *record* yang dipartisi lagi

- c. 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 persamaan 2.3 sebagai berikut :

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

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 persamaan 2.4 sebagai berikut :

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

8. Untuk menghitung *Pessimistic error rate* digunakan persamaan 2.5 sebagai berikut.

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

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

f = nilai perbandingan error rate.

n = total sample.

2.5 Algoritma Naive Bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class yang ditemukan oleh ilmuwan Inggris Thomas Bayes. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi dapat memprediksi berupa peluang di masa yang datang berdasarkan pengalaman di masa sebelumnya yang dikenal dengan Teorema Bayes. *Bayesian classification* terbukti memiliki akurasi dan kecepatan

yang tinggi saat diaplikasikan kedalam *database* dengan data yang besar (Andi Ainun Khaerunnisyah Qodrat, 2017).

Naive Bayes adalah algoritma klasifikasi yang cukup sederhana dan mudah diimplementasikan sehingga algoritma ini sangat efektif ketika diuji dengan *dataset* yang benar, terutama jika *Naive Bayes* dikombinasikan dengan pemilihan fungsi, sehingga *Naive Bayes* dapat mengurangi *redundant* pada data, selain itu *Naive Bayes* menunjukkan hasil yang bagus ketika digabungkan dengan metode *clustering*. *Naive Bayes* terbukti memiliki akurasi yang tinggi dibandingkan dengan *support vector machine* (Forsyth, 2018).

Naive Bayes Classifier termasuk ke dalam pembelajaran supervised, *Naive Bayes* mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label y , Asumsi independen bersyarat dapat dinyatakan dalam bentuk berikut (Suyanto,2017).

Adapun perhitungan *Naive Bayes* dengan persamaan berikut:

Naive Bayes Klasifikasi yaitu klasifikasi statistik dalam memprediksi kelas suatu anggota probabilitas. Adapun probabilitas atau peluang bersyarat dinyatakan dengan persamaan 2.6 sebagai berikut (Moolayil, 2019).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.6)$$

Maka X yaitu bukti, H yaitu Hipotesis, $P(H|X)$ yaitu probabilitas adalah hipotesis H benar bukti X atau pada $P(H|X)$ ialah propabilitas *posterior* H dengan syarat X , $P(X|H)$ yaitu probabilitas adalah bukti X benar atau hipotesis H atau probabilitas *Posterior* X sama syarat H , $P(H)$ yaitu probabilitas *prior* hipotesis H , dan $P(X)$ ialah probabilitas *prior* bukti X persamaan 2.7 sebagai berikut.

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1.F_n|C)}{P(F_1\dots F_n)} \quad (2.7)$$

Maka Variabel C menjelaskan kelas, sedangkan variabel $F_1 \dots F_n$ menjelaskan karakter petunjuk dalam melakukan klasifikasi. Dimana persamaan ini menerangkan peluang yang sampelnya masuk karakter khusus pada kelas C (*Posterior*) yaitu

peluang keluar kelas C (sebelum masuknya sampelnya, banyak dibuat *prior*), dikali pada kemungkinan muncul karakter sampel kelas C (disebut juga *likelihood*), dibagi berdasarkan kemungkinan kemunculan karakter contoh secara global (disebut juga *evidence*). Persamaan diatas bisa dibuat secara sederhana dengan persamaan 2.8 sebagai berikut

$$Posterior = \frac{Prior \times likelihood}{evidence} \quad (2.8)$$

Klasifikasi data kontinyu digunakan persamaan *Densitas Gauss* 2.9 sebagai berikut:

$$P(X_i = X_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_{ij}^2}} \quad (2.9)$$

Dimana : P : Peluang

X_i : Atribut ke i

x_i : Nilai atribut ke i

Y : Kelas dicari

y_i : Sub kelas Y dicari

μ : mean, menjelaskan rata – rata dari seluruh atribut

σ : Deviasi standar, menjelaskan varian

di seluruh atribut. (Negara, Sutabri, Suryatno, & Setiadi, 2018)

2.6 Evaluasi

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 (Masters, 2018). 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 (Kubat, 2017). 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 (Keita, 2017) lebih jelasnya bisa dilihat pada table 2.2 dibawah ini.

Tabel 2.2. 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 persamaan 2.10 sampai 2.13 di bawah ini (Caelen, 2017):

$$Sensitivity = \frac{t_pos}{pos} \quad (2.10)$$

$$Specifity = \frac{t_neg}{neg} \quad (2.11)$$

$$Precision = \frac{t_pos}{t_pos+f_pos} \quad (2.12)$$

$$Accuracy = Sensitivity \frac{pos}{(pos+neg)} + Specifity \frac{neg}{(pos+neg)} \quad (2.13)$$

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 persamaan 2.14 sampai 2.15 dibawah ini:

$$\theta^r = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(xt^r, xj^r) \quad (2.14)$$

Dimana :

$$(X, Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases} \quad (2.15)$$

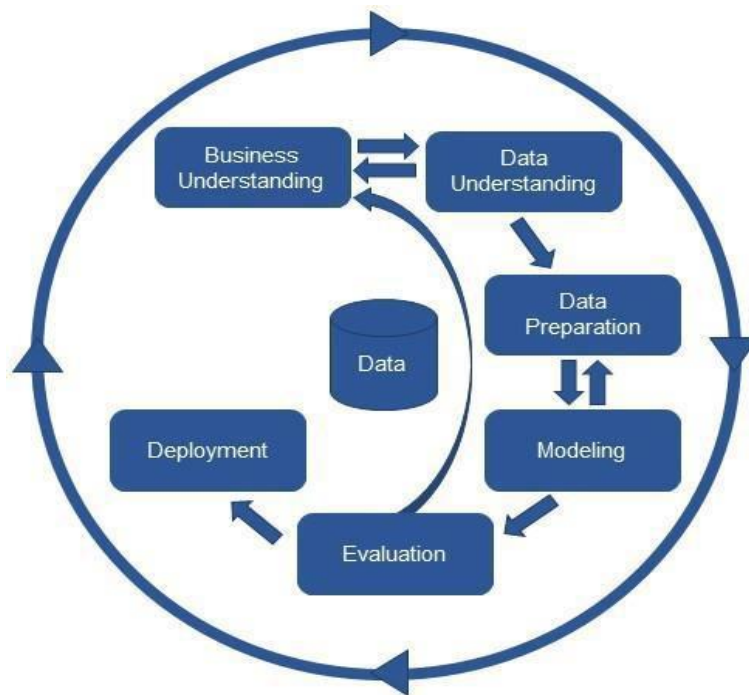
Keterangan :

X = Output positif

Y = Output negatif

2.7 CRIPS-DM

Cross-Industry Standard Process for Data mining (CRISP-DM) yang dikembangkan pada tahun 1996 oleh analis dari beberapa industri seperti *Daimler Chrysler*, SPSS dan NCR. CRISP-DM menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM, sebuah proyek *data mining* memiliki proses yang terbagi dalam enam fase. Keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung pada keluaran dari fase sebelumnya untuk lebih jelasnya dapat dilihat pada gambar 2.2 dibawah ini.(Yizhou Sun, 2017).



Gambar 2.2 Model Fase *Data Mining* dalam CRISP-DM

Berikut enam fase proses *data mining* berdasarkan CRISP-DM:

1. Fase Pemahaman Bisnis (*Business Understanding Phase*), yaitu fase untuk memahami tujuan dan kebutuhan dalam lingkup bisnis, kemudian menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*. Selanjutnya akan ditentukan strategi untuk mencapai tujuan
2. Fase Pemahaman Data (*Data Understanding Phase*), yaitu fase pengumpulan data yang kemudian akan dilakukan analisis penyelidikan untuk mengenali data lebih lanjut dan pencarian pengetahuan awal, mengidentifikasi kualitas data serta jika diinginkan, memilih sebagian kecil grup data yang mungkin mengandung pola dari permasalahan.
3. Fase Pengolahan Data (*Data Preparation Phase*), dimana fase ini meliputi pengumpulan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini juga mencakup pemilihan variabel yang ingin di analisis, melakukan perubahan pada beberapa variabel jika dibutuhkan serta menyiapkan data awal untuk kemudian dijadikan masukan dalam fase pemodelan.

4. Fase Pemodelan (*Modeling Phase*), dimana akan dilakukan pemilihan dan pengaplikasian berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mengoptimalkan hasil. Secara khusus, beberapa teknik yang berbeda dapat digunakan pada permasalahan *data mining* yang sama. Dan untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *data mining* tertentu, pada fase ini proses dapat kembali ke fase sebelumnya (pengolahan data).
5. Fase Evaluasi (*Evaluation Phase*), dimana pada fase ini dilakukan evaluasi terhadap model yang telah terbentuk untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan. Pada fase ini pula ditetapkan apakah terdapat model yang memenuhi tujuan pada fase awal dan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik. Di akhir fase ini kemudian diambil keputusan berkaitan dengan penggunaan hasil dari *data mining* .
6. Fase Penyebaran (*Deployment Phase*), dimana pada fase ini model yang dihasilkan telah dapat digunakan. Contoh sederhana pada fase penyebaran yakni pembuatan laporan, sedangkan contoh kompleks fase penyebaran yakni penerapan proses *data mining* secara paralel pada departemen lain (Istyfaiyah et al. 2017)

2.8 Gambaran Umum Universitas Muhammadiyah Pringsewu Lampung

Universitas Muhammadiyah Pringsewu merupakan perguruan tinggi swasta yang diselenggarakan oleh Persyarikatan Muhammadiyah untuk melaksanakan pendidikan akademik, vokasi, dan profesi dalam sejumlah disiplin ilmu, teknologi, olahraga, dan seni. Pendirian universitas diawali dengan pendirian Perguruan Tinggi Muhammadiyah Pringsewu dalam bentuk Sekolah Tinggi.

1. Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Muhammadiyah Pringsewu

Pada tahun 1979 didirikan Sekolah Tinggi Ilmu Pendidikan (STIP) Muhammadiyah Pringsewu sebagai filial Sekolah Tinggi Ilmu Pendidikan (STIP) Muhammadiyah Metro di Pringsewu. Pada tahun 1983 STIP Muhammadiyah Pringsewu secara resmi berdiri sendiri dan berubah nama menjadi Sekolah Tinggi Ilmu Keguruan dan Ilmu Pendidikan (STKIP) Muhammadiyah Pringsewu dengan izin operasional yang diperoleh pada tahun 1984 berdasarkan Surat Keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia nomor: 0629/O/1984 tanggal 09 Maret 1984 dengan empat (4) program studi, yaitu (1) Pendidikan Luar Sekolah; (2) Psikologi Pendidikan dan Bimbingan; (3) Pendidikan Matematika; dan (4) Pendidikan Bahasa dan Sastra Indonesia.

Status “Terdaftar” diperoleh pada tahun 1986 berdasarkan Surat Keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia nomor: 0512/O/1986 tanggal 31 Juli 1986 untuk program studi (1) Pendidikan Luar Sekolah; (2) Psikologi Pendidikan dan Bimbingan; (3) Pendidikan Matematika; dan (4) Pendidikan Bahasa dan Sastra Indonesia.

Status “Diakui” diperoleh pada tahun 1990 berdasarkan Surat Keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia nomor 0772/O/1990 tanggal 31 Desember 1990 untuk program studi (1) Pendidikan Luar Sekolah; (2) Psikologi Pendidikan dan Bimbingan; (3) Pendidikan Matematika; dan (4) Pendidikan Bahasa dan Sastra Indonesia, sekaligus membuka program jenjang diploma tiga (D3) untuk program studi Bahasa Inggris, dan Pendidikan Akuntansi/Dunia Usaha. Tahun 1992, berdasarkan Surat Keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia nomor 042/O/1992 tanggal 23 Januari 1992, kembali diperoleh status “Terdaftar” untuk program pendidikan Diploma Tiga (D3) program studi Pendidikan Bahasa dan Sastra Indonesia. Tahun 1993, dilakukan penutupan (passing out) seluruh program pendidikan jenjang Diploma Tiga (D3) berdasarkan

surat Dirjen Pendidikan Tinggi nomor 305/D4/XII/1993 dan Surat Kopertis Wilayah II nomor 2023/KL.03.02/Kop.II/1994 tanggal 22 Desember 1994. Pada tahun 1996, STKIP Muhammadiyah Pringsewu mengembangkan amal usaha muhammadiyah dengan mendirikan Akademi Keperawatan dan Sekolah Tinggi Ilmu Ekonomi.

Pada tahun 2009 STKIP Muhammadiyah Pringsewu membuka program studi Pendidikan Bahasa Inggris dengan ijin operasional berdasarkan Surat Keputusan Direktur Jenderal Pendidikan Tinggi nomor 456/D/T/2009 tanggal 06 April 2009. Tahun 2016 diperoleh ijin operasional pembukaan Program studi Pendidikan Guru Sekolah Dasar (PGSD) berdasarkan Surat Keputusan Kementerian Riset Teknologi dan Pendidikan Tinggi nomor 4566/A4/HK/2016 dan ijin operasional pembukaan program studi Pendidikan Tari diperoleh tahun 2017 berdasarkan Surat Keputusan Kementerian Riset Teknologi dan Pendidikan Tinggi nomor 1029/A4/HK/2017.

Sampai dengan tahun 2019 STKIP Muhammadiyah Pringsewu mengelola program pendidikan akademik jenjang Sarjana (S1) dengan data program studi sebagai berikut:

- (1) 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 Terakreditasi B, sesuai surat keputusan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Nomor: 1145/SK/BAN-PT/Akred/S/XI/2015.
- (2) 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 Terakreditasi B, sesuai surat keputusan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Nomor: 0883/SK/BAN-PT/Akred/S/III/2017.

- (3) 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 Terakreditasi B, sesuai surat keputusan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Nomor: 0442/SK/BAN-PT/Akred/S/V/2016.
- (4) 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 Terakreditasi B, sesuai surat keputusan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) Nomor: 0441/SK/BAN-PT/Akred/S/V/2016.
- (5) Program Studi Pendidikan Guru 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.
- (6) 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 Muhammadiyah Pringsewu Terakreditasi B, sesuai dengan surat keputusan BAN-PT Nomor: 4047/SK/BAN-PT/Akred/PT/X/2017.

2. Sekolah Tinggi Ilmu Kesehatan (STIKes) Muhammadiyah Pringsewu

Akademi Keperawatan (AKPER) resmi berdiri berdasarkan Surat Keputusan Menteri Kesehatan Republik Indonesia, nomor HK.00.06.1.1.560 tanggal 27 Februari 1996 dengan memiliki satu program studi Diploma Tiga (D3) Keperawatan. Selanjutnya, pada tahun 2007, Akper melakukan pengembangan dengan mendirikan Akademi Kebidanan (AKBID) berdasarkan Surat Keputusan Menteri Kesehatan Republik Indonesia 59/D/O/2007 tanggal 25 April 2007 dengan program studi Diploma Tiga (D3) Kebidanan.

Untuk memperkuat manajemen dan meningkatkan kualitas sumberdaya bidang kesehatan didirikan Sekolah Tinggi Ilmu Kesehatan (STIKes) Muhammadiyah Pringsewu dengan melakukan penggabungan (merger) antara AKPER dan AKBID Muhammadiyah Pringsewu. Ijin operasional pendirian Sekolah Tinggi Ilmu Kesehatan (STIKes) Muhammadiyah Pringsewu diperoleh berdasarkan Surat Keputusan Menteri Pendidikan dan Kebudayaan, nomor: 254/D/O/2008 tanggal 23 Desember 2008 dengan menambah satu (1) program studi jenjang Sarjana (S1) program studi Ilmu Keperawatan. 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 4 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 izin 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 :281/E/O/2013 perihal izin penyelenggaraan program studi Profesi Ners.

3. Sekolah Tinggi Ilmu Ekonomi (STIE) Muhammadiyah Pringsewu

Sekolah Tinggi Ilmu Ekonomi (STIE) Muhammadiyah Pringsewu diawali dengan membuka kelas filial Sekolah Tinggi Ilmu Ekonomi (STIE) Muhammadiyah Kalianda di Pringsewu dengan program studi Manajemen jenjang sarjana (S1). Selanjutnya, pada tahun 1996 STIE Muhammadiyah Pringsewu mendapatkan izin operasional berdasarkan Surat Keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia nomor 058/D/1996 tanggal 08 Agustus 1996 tentang pemberian status terdaftar bagi Sekolah Tinggi Ilmu Ekonomi (STIE) Muhammadiyah Pringsewu dengan program studi Manajemen jenjang sarjana (S1).

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 izin 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.

Sampai dengan saat ini Universitas Muhammadiyah Pringsewu memiliki 5 (lima) Fakultas, 1 (satu) program pasca sarjana dan 18 (delapan belas) program studi sebagai berikut:

- a. Fakultas Kesehatan
 - 1) Program Studi D3 Keperawatan
 - 2) Program Studi D3 Kebidanan
 - 3) Program Studi D4 Manajemen Informasi Kesehatan
 - 4) Program Studi D4 Teknologi Laboratorium Medis
 - 5) Program Studi S1 Keperawatan
 - 6) Program Studi S1 Kebidanan
 - 7) Program Studi S1 Fisioterapi
 - 8) Program Studi Pendidikan Profesi Ners
 - 9) Program Studi Pendidikan Profesi Bidan
- b. Fakultas Ekonomi dan Bisnis
 - 1) Program Studi S1 Manajemen
- c. Fakultas Keguruan dan Ilmu Pendidikan
 - 1) Program Studi S1 Bimbingan dan Konseling
 - 2) Program Studi S1 Pendidikan Matematika
 - 3) Program Studi S1 Pendidikan Bahasa dan Sastra Indonesia
 - 4) Program Studi S1 Pendidikan Bahasa Inggris
 - 5) Program Studi S1 Pendidikan Guru Sekolah Dasar
- d. Fakultas Syariah
 - 1) Program Studi S1 Hukum Ekonomi Syariah / Muamalah
- e. Fakultas Sains dan Teknologi
 - 1) Program Studi S1 Sistem dan Teknologi Informasi
- f. Pasca Sarjana
 - 1) Program Studi S2 Administrasi Pendidikan

Secara geografis Universitas Muhammadiyah Pringsewu Pringsewu terletak di antara 104.9723053" BT dan -5.3478598" LS yang berlokasi di Jl. KH. Ahmad Dahlan No. 112 Pringsewu Kabupaten Pringsewu Provinsi Lampung.