BAB II

LANDASAN TEORI

A. Tinjauan Pustaka

1. Informasi

Menurut Kenneth C.Laudon dan Jane P. Laudon (2022:46) definisi informasi yaitu :

"By information we mean data that have been shaped into a form that is meaningful and useful to human beings."

Informasi dimaksud dengan data yang telah terbentuk menjadi sebuah yang sangat berarti bagi manusia.

2. Sistem Informasi

Menurut Kenneth C.Laudon dan Jane P. Laudon (2022:46) definisi sistem informasi yaitu :

"An Information System can be defined technically as a set of interrelated components that collect (or retrieve),process,store and distribute information to support decision making and control in an organization."

Sistem Informasi bisa didefinisikan secara teknis sebagai sekumpulan komponen yang saling terkait yang mengumpulkan (atau mendapatkan kembali), memproses, menyimpan, dan mendistribusikan informasi untuk mendukung pengambilan keputusan dan pengendalian dalam suatu organisasi.

3. Data

Data adalah sebuah informasi yang didapatkan berdasarkan observasi, wawancara dan komunikasi yang akan dibutuhkan untuk kebutuhan tertentu, data bisa bermacam-

macam variasi contohnya seperti : data sampel, data observasi, data primer dan data sekunder.

Pembagian data mempunyai 4 kategori yaitu : sifat, sumber data, cara memperolehnya dan waktu pengumpulannya. Berikut ini kategori tersebut akan dijelaskan masing-masing.

Menurut Syafrizal Helmi Situmorang (2014:2), ada dua macam sifat data yaitu :

a. **Kualitatif**

   Data yang tidak berbentuk angka, dan mempunyai ciri tidak bisa dilakukan operasi matematika seperti penambahan, pengurangan, perkalian dan pembagian. Contohnya adalah kuesioner pertanyaan, kualitas pelayanan sebuah dealer mobil.

b. **Kuantitatif**

   Data yang berbentuk angka yang bisa dilakukan dengan operasi matematika. Contohnya seperti pendapatan per kapita, jumlah orang mempunyai mobil.

Menurut Syafrizal Helmi Situmorang (2014:3), ada dua macam sumber data yaitu :

a. ***Internal***

   Data dari dalam suatu organisasi yang menggambarkan keadaan organisasi tersebut. Contohnya seperti jumlah karyawan, jumlah pendapatan.

b. **Eksternal**

   Data dari luar suatu organisasi yang dapat menggambarkan faktor-faktor yang mempengaruhi hasil kerja suatu organisasi. Contohnya seperti daya beli masyarakat dalam mempengaruhi hasil penjualan suatu perusahaan.

Menurut Syafrizal Helmi Situmorang (2014:3), ada dua macam cara memperoleh data yaitu :

a. **Primer**

Data yang dikumpulkan sendiri oleh perorangan/organisasi secara langsung dari objek yang diteliti dan untuk kepentingan studi yang bersangkutan dapat berupa *interview* dan observasi.

b. **Sekunder**

Data yang dikumpulkan dari studi-studi sebelumnya atau diterbitkan oleh instansi lain yang dapat berupa data dokumentasi dan arsip-arsip yang disimpan dalam penyimpanan di website.

Menurut Syafrizal Helmi Situmorang (2014:3), ada dua macam waktu pengumpulan data yaitu :

a. ***Cross Section***

Data yang dikumpulkan pada suatu waktu tertentu untuk menggambarkan keadaan dan kegiatan pada waktu tersebut. Contohnya seperti data penelitian yang menggunakan kuesioner.

b. **Berkala(*time series data*)**

Data yang dikumpulkan dari waktu ke waktu untuk melihat perkembangan kejadian/kegiatan selama periode tersebut. Misalnya, perkembangan uang beredar, harga 9 macam bahan pokok, penduduk.

4. *Dataset*

Menurut Pan Ning Tan, et.al (2019:46) definisi dari *dataset* yaitu :

"*A data set can often be viewed as a collection of data objects. Other names for a data object are record, point, vector, pattern, event, case, sample, instance, observation, or entity.*"

Dataset bisa dilihat sebagai kumpulan objek data. Nama lain dari objek data adalah *record*, titik, vector, pola, kejadian, kasus, sampel, contoh, observasi atau entitas.

Menurut Pan Ning Tan, et.al (2019:54), ada tiga karaterisktik dari *dataset* yaitu:

a. ***Dimensionality***

   Dimensi dari *dataset* adalah jumlah atribut yang dimiliki objek dalam kumpulan data.

b. ***Distribution***

   Distribusi dari *dataset* adalah frekuensi kejadian dari berbagai nilai atau kumpulan nilai untuk atribut yang terdiri dari objek data.

c. ***Resolution***

   Resolusi dari *dataset* adalah suatu tingkat dimana resolusi tinggi menghasilkan data secara mendalam sedangkan resolusi rendah menghasilkan data yang sudah digeneralisasi atau ringkasan. Resolusi tersebut tergantung kebutuhan data yang digunakan.

5. *Data Mining*

Menurut Pan-Ning Tan, et.al (2019:24) definisi *data mining* yaitu :

"Data mining is the process of automatically discovering useful information in large data repositories"

Data mining adalah sebuah proses otomatis mencari informasi yang berguna dalam penyimpanan data yang besar.

Data mining juga bagian dari *Knowledge Discovery In Databases* (KDD) yang prosesnya mengkonversi data mentah menjadi informasi berguna, proses yang dilakukan dari data *preprocessing* sampai *postprocessing* hasil dari *data mining*. Berikut gambar yang menjelaskan proses tersebut

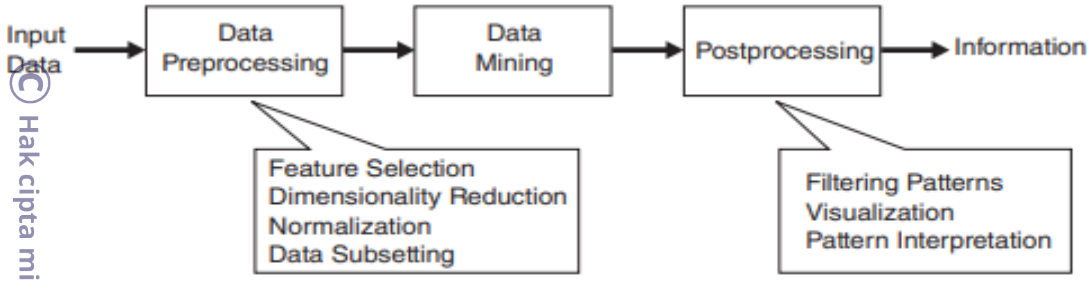


Figure 1.1. The process of knowledge discovery in databases (KDD).

Gambar 2.1
Proses *Data Mining*
Sumber : Pan Ning-Tan (2019:24)

Menurut Pan Ning Tan et al (2019:25) gambar 2.1 diatas dijelaskan sebagai berikut yaitu:

a. ***Data preprocessing*** bertujuan untuk mengubah data mentah menjadi format yang digunakan untuk analisis, tahapan dari *data preprocessing* termasuk menggabungkan data dari beberapa sumber, membersihkan data untuk menghapus observasi duplikat dan *noise*, dan menyeleksi catatan dan fitur yang relevan dalam *data mining*.
b. ***Data postprocessing*** bertujuan untuk memastikan bahwa hasil valid dan berguna yang akan diintergrasikan pada pengambilan keputusan seperti filter pola, visualisasi dan interpretasi pola.

Menurut Pan Ning Tan, et.al (2019:29), tugas dari *data mining* dibagi menjadi dua kategori yaitu:

a. ***Predictive Tasks***

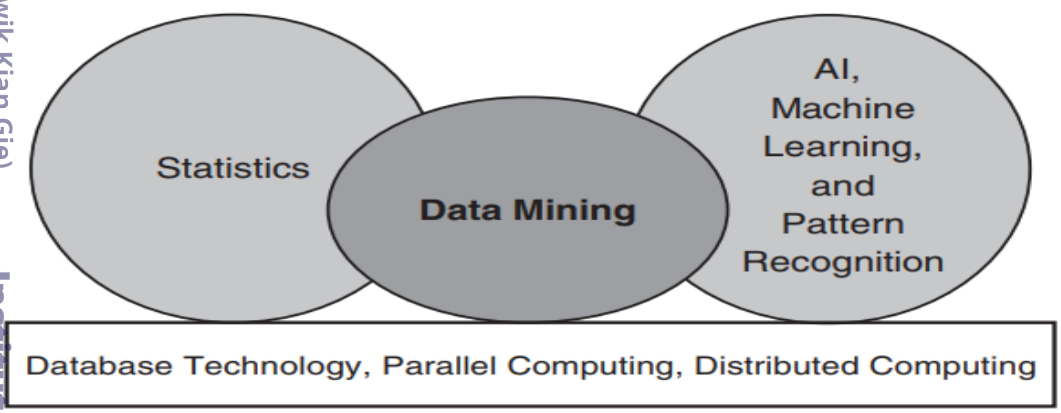
   Tujuan dari tugas tersebut untuk memprediksi nilai dari atribut tertentu berdasarkan nilai atribut lainnya. Atribut yang akan diprediksi biasa disebut dengan variabel tidak bebas atau target

sedangkan atribut yang dilakukan untuk prediksi disebut dengan variabel bebas atau bersifat penjelasan.

b. ***Descriptive Tasks***

Tujuan dari tugas tersebut untuk mendapatkan pola (korelasi, tren, klaster, lintasan dan anomali) yang merangkum hubungan yang mendasari data. Untuk tugas deskriptif bersifat eksploratif dan sering membutuhkan teknik *postprocessing* untuk memvalidasi dan menjelaskan hasilnya.

Data mining membantu dalam bidang lainnya seperti statistik dengan pengujian hipotesis dan estimasi sampel, untuk *artificial intelligence* seperti algoritma pencarian, teknik *modeling*, *machine learning* dan pengenalan pola. Berikut gambar yang menunjukan relasi antar bidang dari *data mining* :



Gambar 2.2
Relasi *data mining* antar bidang
Sumber : Pan Ning-Tan (2019:27)

Dalam pembelajaran *data mining* memiliki dua tipe yaitu:

a. ***Supervised Learning***

Menurut Taeho Jo (2021:11), definisi *supervised learning* adalah paradigma pembelajaran dimana *parameter* yang dioptimisasi untuk minimalisir perbedaan antara *output target* dan output yang berbasis komputer. Contoh

model algoritma yang digunakan dalam klasifikasi dan regresi yaitu : KNN (*K Nearest Neighbour*), *Naïve Bayes* dan *Decision Tree*.

b. ***Unsupervised Learning***

Menurut Taeho Jo (2021:12), definisi *unsupervised learning* adalah proses dalam optimisasi prototype kluster tergantung pada persamaan dalam contoh data pelatihan. Contoh model algoritma yang digunakan dalam klustering yaitu : K-Means, *Fuzzy C-Means*.

Menurut Pan Ning Tan, et.al (2019:29), pengolahan *data mining* memiliki beberapa metode pengolahan yaitu:

a. **Prediksi** (*Predictive*).

Teknik prediksi digunakan apabila suatu nilai memiliki atribut yang berbeda, contoh algoritmanya seperti Algoritma *Linear Regression, Neural Network* dan lain-lain.

b. **Asosiasi** (*Association)*

Teknik Asosiasi digunakan untuk hubungan antar data, contoh algoritmanya seperti Algoritma Apriori.

c. **Klastering** (*Clustering*)

Teknik *clustering* digunakan untuk pengelompokan data dalam suatu kelompok tertentu, contoh algoritmanya seperti *K-Means, K-Medoids, Self Organization Map(SOM), Fuzzy C-Means.*

d. **Klasifikasi** (*Classification*)

Teknik klasifikasi mengelompokkan data yang mempunyai variabel tertentu berbeda dengan *clustering* yang tidak memiliki variabel yang dependen, contoh algoritma seperti *ID3* dan *K Nearest Neighbour.*

6. *Naïve Bayes*

Menurut Jiawei Han, et al (2023:259) definisi *Naive Bayes* atau *bayesian classifier* yaitu :

"Bayesian classifier are statistical classifier. They can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class"

Naive Bayes adalah klasifikasi statistika. Mereka bisa memprediksi kelas keanggotaan probabilitas seperti probabilitas yang diberikan tupel milik pada kelas tertentu. Contoh yang diberikan oleh Ethem Alpaydin (2020:88) seperti berikut :

"Tossing a coin is a random process because we cannot predict at any toss whether the outcome will be heads or tails, that is why we toss coins, or buy lottery tickets, or get insurance. We can only talk about the probability that the outcome of the next toss will be heads or tails."

Melempar koin adalah proses acak karena kita tidak memprediksi pada lemparan apapun apakah hasilnya kepala atau ekor, itulah sebabnya kita melempar koin, atau membeli tiket lotre, atau mendapatkan asuransi, kami hanya dapat berbicara tentang kemungkinan bahwa hasil lemparan berikutnya adalah kepala atau ekor.

Teori *Bayes* dapat dijelaskan sebagai berikut menurut Pan Ning-Tan, et al (2019:416) yaitu :

Misalkan P(Y|X) menunjukan probabilitas bersyarat untuk mengamati acak variabel Y setiap kali variabel X mengambil nilai tertentu. P(Y|X) sering dibaca sebagai probabilitas mengamati Y yang dikondisikan pada hasilnya dari X. Probabilitas bersyarat dapat digunakan untuk menjawab pertanyaan seperti "cuaca di luar, berapa kemungkinan saya akan pergi ke sekolah." Probabilitas bersyarat dari X dan Y terkait dengan kedua probabilitas tersebut seperti ini :

$$
\begin{aligned}
P(Y|X) &= \frac{P(X,Y)}{P(X)}, \quad \text{which implies} \\
P(X,Y) &= P(Y|X) \times P(X) \\
&= P(X|Y) \times P(Y).
\end{aligned}
$$

Gambar 2.3
Formula Probabilitas X dan Y
Sumber : Pan Ning-Tan (2019:417)

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}.$$

Gambar 2.4
Formula Teori Bayes
Sumber : Pan Ning-Tan (2019:417)

Teori Bayes memberikan hubungan antara probabilitas bersyarat P(Y|X) dan P(X|Y), dengan catatan denominator yang melibatkan margin probabilitas dari X yang dapat direpresentasikan seperti gambar berikut.

$$P(X) = \sum_{i=1}^{k} P(X, y_i) = \sum_{i=1}^{k} P(X|y_i) \times P(y_i).$$

Gambar 2.5
Hubungan Teori Bayes dengan Probabilitas X
Sumber : Pan Ning-Tan (2019:417))

Dengan expresi di gambar 2.5, kita mendapatkan persamaan untuk P(Y|X) dalam hal P(X|Y) dan P(Y) seperti gambar berikut :

$$P(Y|X) = \frac{P(X|Y)P(Y)}{\sum_{i=1}^{k} P(X|y_i)P(y_i)}.$$

Gambar 2.6
Formula Teori Bayes antara P(X|Y) dan P(Y)
Sumber : Pan Ning-Tan (2019:417)

7. *Decision Tree*

Menurut Andreas C. Muller & Sarah Guido (2017:72) definisi *decision tree* yaitu :

"Decision trees are widely used models for classification and regression tasks. Essentially, they learn a hierarchy of if/else question, leading to a decision."

Decision Tree merupakan model pembelajaran untuk digunakan dalam tugas klasifikasi dan regresi yang memiliki hierarki dari pertanyaan jika atau lain, yang mengarah ke keputusan. Struktur pohon yang terdiri dari *root node,* cabang, *internal node* dan *leaf node*.

Decision tree memiliki 3 tipe *nodes* menurut Pan Ning-Tan et.al (2019: 140) yaitu:

a. ***Root Node*** adalah node yang tidak memiliki tautan masuk atau disebut nol atau banyak tautan keluar.
b. ***Internal Node*** adalah *node* yang masing-masing memiliki satu tautan masuk dan dua atau banyak tautan keluar.
c. ***Leaf Node*** adalah *node* yang mempunyai tautan masuk dan tidak ada tautan keluar.

8. *Random Forest*

Menurut Andreas C.Muller & Sarah Guido (2017:85) definisi *random forest* yaitu:

"Random forest is essentially a collection of decision trees, where each tree is slightly different from the other. The idea behind random forests is that each tree might do a relatively good job of predicting but will likely overfit on part of the data. If we build many trees, all of which work well and overfit in different ways, we can reduce the amount of overfitting by averaging their results. This reduction in overfitting, while retaining the predictive power of the trees, can be shown using rigorous mathematics."

Random Forest pada dasarnya adalah kumpulan *decision tree,* di mana setiap pohon sedikit berbeda dari yang lain. Gagasan di balik *random forest* adalah bahwa setiap pohon mungkin melakukan pekerjaan prediksi yang relatif baik tetapi kemungkinan

besar akan menyesuaikan sebagian data. Jika kita membangun banyak pohon, yang semuanya bekerja dengan baik dan *overfit* dengan cara yang berbeda, kita dapat mengurangi jumlah overfitting dengan merata-ratakan hasilnya. Pengurangan *overfitting* ini, sambil mempertahankan kekuatan prediksi pohon, dapat ditunjukan dengan menggunakan matematika yang ketat.

Menurut Pan Ning-Tan (2019:512) fungsi dari *random forest* yaitu :

"Random forests attempt to improve the generalization performance by constructing an ensemble of decorrelated decision trees. Random forests build on the idea of bagging to use a different bootstrap sample of the training data for learning decision trees."

Random forests mencoba untuk meningkatkan kinerja generalisasi dengan membangun sekumpulan dekolerasi *decision trees*. *Random forests* dibangun atas gagasan untuk menggunakan sampel *bootstrap* yang berbeda dari data pelatihan untuk mempelajari *decision tree*.

9. *Entropy*

Menurut Pan Ning-Tan(2019:372) definisi *entropy* adalah :

"The degree to which each cluster consists of objects of a single class."

Entropy adalah dimana setiap kluster terdiri dari objek dari satu kelas. Rumus dasar dari *entropy* seperti berikut :

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

Gambar 2.7
Rumus Dasar *Entropy*
Sumber : Abdul Muiz Khalimi

Penjelasan keterangan rumus *entropy* yaitu :

S = Himpunan Kasus

n = Jumlah Partisi S

p_i = Probabilitas yang didapat dari jumlah kelas dibagi total kasus

10. *Gini Index*

Menurut Oded Maimon dan Lior Rokach (2010:153) definisi *gini index* adalah :

"Gini index is an impurity-based criterion that measures the divergences between the probability distributions of the target attribute's values."

Gini index adalah kriteria berbasis *impurity* yang mengukur perbedaan antara distribusi probabilitas dari nilai atribut target.

Berikut ini adalah rumus dari *gini index* :

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

Gambar 2.8
Rumus Dasar *Gini Index*
Sumber : Abdul Muiz Khalimi

Penjelasan keterangan rumus *gini index* :

C = Jumlah masing-masing atribut

Pi = Jumlah atribut dari masing-masing kelas atau label

11. Penelitian

Menurut Umesh Kumar dan D.P. Kothari (2022:1) definisi penelitian yaitu :

"Research is a process through which new knowledge is discovered. Research helps us to organize this new information into a coherent body, a set of related ideas that explain events that have occurred, and predict events that may happen."

Penelitian adalah proses dimana pengetahuan baru ditemukan. Riset membantu kita mengatur informasi baru ini ke dalam tubuh yang koheren, sekumpulan gagasan terkait yang menjelaskan peristiwa yang telah terjadi, dan memprediksi peristiwa yang mungkin terjadi.

12. Penelitian Kuantitatif

Menurut Umesh Kumar dan D.P. Kothari (2022:6) definisi penelitian kuantitatif yaitu :

"Quantitative research is variables based whereas qualitative research is attributes based. Quantitative research is based on measurement or quantification of the phenomenon under study. That is , it is data based and hence more objective and popular."

Penelitian kuantitatif berbasis variabel sedangkan penelitian kualitatif berbasis atribut. Penelitian kuantitatif didasarkan pada pengukuran atau kuantifikasi dari fenomena yang diteliti. Artinya, ini berbasis data dan karena lebih objektif dan populer.

13. *Machine Learning*

Menurut Andreas C.Muller dan Sarah Guido (2017:1) definisi *machine learning* yaitu :

"Machine learning is about extracting knowledge from data. It is a research field at the intersection of statistics, artificial intelligence, and computer science and is also known as predictive analytics or statistical learning."

Pembelajaran mesin adalah tentang mengekstraksi pengetahuan dari data. Ini adalah bidang penelitian di persimpangan statistik, kecerdasan buatan, dan ilmu komputer dan juga dikenal sebagai analitik prediktif atau pembelajaran statistik.

Menurut A.C. Faul (2020:6) definisi *machine learning* yaitu :

"Machine learning is all about transferring the various modes of learning we have identified here to machines."

Pembelajaran mesin tentang mentransfer berbagai mode pembelajaran yang telah kami identifikasi di sini ke mesin.

14. *Python*

Menurut William Wizner (2020:28) definisi *Python* yaitu :

"Python is an object-oriented and interpretive computer program language. Its syntax is simple and contains a set of standard libraries with complete functions, which can easily accomplish many common tasks."

Python adalah bahasa program komputer yang berorientasi objek dan interpretatif. Sintaksnya sederhana dan berisi sekumpulan pustaka standar dengan fungsi lengkap, yang dapat dengan mudah menyelesaikan banyak tugas umum.

Menurut Andreas C.Muller dan Sarah Guido (2017:5) mengapa harus menggunakan *Python* yaitu :

"Python has become the lingua franca for many data science applications. It combines the power of general-purpose programming languages with the ease of use of domain-specific scripting languages like MATLAB or R. Python has libraries for data loading, visualization, statistics, natural language processing, image processing, and more. This vast toolbox provides data scientists with a large array of general- and special-purpose functionality. One of the main advantages of using Python is the ability to interact directly with the code, using a terminal or other tools like the Jupyter Notebook. Machine learning and data analysis are fundamentally iterative processes, in which the data drives the analysis. It is essential for these processes to have tools that allow quick iteration and easy interaction."

Python telah menjadi bahasa perantara untuk banyak aplikasi ilmu data. Ini menggabungkan kekuatan bahasa pemrograman tujuan umum dengan kemudahan penggunaan bahasa skrip khusus domain seperti *MATLAB* atau *R*. Python memiliki perpustakaan untuk pemuatan data, visualisasi, statistik, pemrosesan bahasa alami, pemrosesan gambar, dan banyak lagi. *Toolbox* yang luas memberi para ilmuwan data sejumlah besar fungsi tujuan umum dan khusus. Salah satu keuntungan utama dalam menggunakan *python* adalah kemampuannya untuk berinteraksi langsung dengan kode, menggunakan terminal atau alat lain seperti *Jupiter Notebook*. Pembelajaran mesin dan analisis data pada dasarnya adalah proses berulang, dimana data mendorong analisis.

Penting bagi proses ini untuk memiliki alat yang memungkinkan iterasi cepat dan interaksi yang mudah.

15. *Orange*

Orange adalah sebuah *platform data mining* untuk menjalankan analisis data dan visualisasi. Tujuan dari pemanfaatan *Orange* adalah menyediakan platform model prediksi dan sistem rekomendasi. Contoh penggunaan *Orange* pada pembelajaran selain *data mining* adalah bioinformatika, penelitian *genomic* dan lain-lain. *Orange* menggunakan bahasa pemrograman C++, Python dan Cython.

B. Peneliti Terdahulu

1. Penerapan Data Mining Dalam Menentukan Masa Studi Mahasiswa Menggunakan Algoritma C4.5

Hasil penelitian yang dilakukan oleh Wirdana Rauda Ningsih, bahwa kesimpulannya :

a. Pada penerapan *Data Mining* dengan metode yang digunakan Algoritma *C4.5* yang menghasilkan sebuah informasi untuk prediksi masa studi secara tepat waktu di universitas STMIK Budi Darma pada Program Studi Manajemen Informatika (D3) hingga tingkat akurasi yang dihasilkan 755 dengan jumlah 56 data tes.

b. Penerapan *data mining* dalam menentukan masa studi mahasiswa menggunakan algoritma *C4.5* di STMIK Budi Darma Program Studi Manajemen Informatika (D3)sehingga mempermudah pihak kampus dalam mengatur strategi manajemen untuk kedepannya agar mahasiswa bisa lulus tepat waktu.

c. Penerapan dari hasil *data mining* yang ditentukan dari masa Studi Mahasiswa sehingga dapat kita ketahui mahasiswa "Lulus secara Tepat Waktu" untuk mahasiswa yang mempunyai nilai IPK tinggi dan "Lulus Secara Tidak Tepat Waktu" mahasiswa memiliki nilai IPK sedang dengan konsentrasi jurusan komputer akuntansi dan lulusan SMK berdasarkan Algoritma *C4.5* yang dilakukan oleh peneliti.

2. Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbour, Decision Tree Serta Naive Bayes

Hasil Penelitian yang dilakukan oleh Hozairi, Anwari, Syariful Alim bahwa kesimpulannya :

Hasil penelitian ini menunjukkan bahwa setelah menggunakan model *K-Nearest Neighbor*, *Decision Tree* serta *Naive Bayes* untuk mengklasifikasi status kelulusan mahasiswa Teknik Informatika Universitas Islam Madura diperoleh hasil bahwa kinerja *Naive Bayes* lebih unggul dari *K-Nearest Neighbor* serta *Decision Tree*. Terbukti bahwa dari 35 data uji yang digunakan *Naive Bayes* memiliki nilai akurasi 89%, presisi 88% sedangkan *K-Nearest Neighbor* memiliki nilai akurasi 77% , presisi 76% dan *Decision Tree* memiliki nilai akurasi 74% dan presisi 84%. Kontribusi riset ini bisa digunakan oleh manajemen Prodi Teknik Informatika Universitas Islam Madura untuk mendeteksi sejak awal kondisi mahasiswa supaya tidak kelulusannya tidak terlambat dan mempengaruhi nilai akreditasi Program Studi Teknik Informatika Universitas Islam Madura.

3. Sistem Klasifikasi Kelulusan Mahasiswa Dengan Algoritma Random Forest

Hasil Penelitian yang dilakukan oleh Jaya S.Saleh, Angelia M. Adrian dan Junaidy B.Sanger, bahwa kesimpulannya :

a. Sistem klasifikasi kelulusan mahasiswa menggunakan algoritma *Random Forest* telah berhasil dibangun dan berfungsi sesuai dengan fitur-fitur yang telah ditentukan yaitu sistem dapat memasukkan data latih, memasukkan data uji, me-*reset data* latih dan data uji, memasukkan jumlah pohon, mengklasifikasi data uji, dan dapat mengklasifikasi *dataset* tunggal.

b. Nilai akurasi dari hasil klasifikasi *dataset* kelulusan mahasiswa menggunakan algoritma *Random Forest* meningkat seiring dengan menaikkan penggunaan jumlah pohon yaitu dengan nilai akurasi 90.00% menggunakan 50 pohon.

4. Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Dengan Metode Naive Bayes

Hasil penelitian yang dilakukan oleh Tias Mugi Rahayu, Besse Arnawisuda Ningsi, Ismurani dan Irvana Arofah, bahwa kesimpulannya :

a. Data kelulusan mahasiswa S1 Fakultas Ekonomi Universitas Pamulang tahun ajaran 2018/2019 menunjukan 61,9% yang lulus tepat waktu dan 38,1% yang lulus tidak tepat waktu.

b. Pada 225 data testing yang diuji menggunakan perhitungan *naïve bayes* menunjukan hasil bahwa data testing tersebut terklasifikasi dengan benar sebanyak 156 data.

c. Hasil perhitungan *naïve bayes* pada 156 data yang terklasifikasi dengan benar menunjukan tingkat akurasi sebesar 69,33%.

5. Implementasi Decision Tree untuk Prediksi Kelulusan Mahasiswa Tepat Waktu

Hasil penelitian yang dilakukan oleh Christin Nandari Dengen, Kusrini dan Emha Taufiq Luthfi, bahwa kesimpulannya :

a. *Decision tree* dapat diterapkan untuk prediksi kelulusan mahasiswa tepat waktu di Program Studi Teknik Informatika Universitas Mulawarman dengan tingkat akurasi sebesar 60%.

b. Penelitian ini dapat membantu pihak program studi untuk memprediksi kelulusan mahasiswa tepat waktu.