



ANALISIS KINERJA DRIVER PADA APLIKASI CARPOOL MENGUNAKAN METODE C.45



Jefri Marbun
Nisa Hanum Harani

Penulis

:

JEFRINANDA IASPARTOGI MARBUN

NISA HANUM HARANI, S.Kom., M.T., CDSP, SPFC

Penyunting : NAMA PENGUJI SIDANG

Layout : JEFRINANDA IASPARTOGI MARBUN

Sanksi Pelanggaran Pasal 113

Undang-Undang Nomor 28 Tahun 2014

tentang Hak Cipta

- (1) Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
- (2) Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf e, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).
- (3) Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf c, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah).
- (4) Setiap Orang yang memenuhi unsur sebagaimana dimaksud pada ayat (3) yang dilakukan dalam bentuk pembajakan, dipidana dengan pidana penjara paling lama 10 (sepuluh) tahun dan/atau pidana denda paling banyak Rp4.000.000.000,00 (empat miliar rupiah).

Kata Sambutan

Kata Pengantar

Puji Syukur kepada Tuhan Yang Maha Esa, sehingga penulis mampu menyelesaikan proposal intership I yang berjudul “Analisis Kinerja Driver pada Aplikasi Carpool Menggunakan Metode C.45” tepat pada waktunya. Proposal ini disusun untuk memenuhi kelulusan Intership I. Penulis juga menyadari bahwa selama berlangsungnya penyusunan proposal ini tak lepas dari dukungan serta bantuan berbagai pihak. Saya menyadari dalam proposal ini masih banyak kekurangan, oleh karena itu perlu kritik dan saran untuk penyempurnaan laporan selanjutnya.

Bandung, Februari 2022

Penulis

Daftar Isi

KATA SAMBUTAN.....	2
KATA PENGANTAR	3
DAFTAR ISI.....	4
BAB 1.....	5
PENDAHULUAN	5
BAB 2.....	7
LANDASAN TEORI.....	7
BAB 3.....	79
METODELOGI PENELITIAN	79
BAB 4.....	92
ANALISIS DAN HASIL PEMBAHASAN	92
BAB 5.....	96
KESIMPULAN	96
DAFTAR PUSTAKA	100

Bab 1

Pendahuluan

Perkembangan teknologi dan informasi ternyata membawa perubahan dalam segala bidang termasuk kepada perkembangan dunia industri. Perkembangan teknologi yang pesat saat ini banyak menarik perhatian masyarakat luas. Perkembangan komputer yang dinamis diiringi perkembangan zaman dapat menghasilkan informasi yang cepat dan akurat dan tidak memerlukan waktu yang lama dalam menggali informasi yang dibutuhkan. Aplikasi Carpool merupakan aplikasi yang digunakan oleh PT. Perusahaan Pengelola Aset untuk melakukan pemesanan supir pribadi menggunakan driver yang sudah terdaftar di perusahaan.

Kinerja (performance) mengacu kepada kadar pencapaian tugas-tugas yang membentuk sebuah pekerjaan karyawan (Simamora, 2004). Penilaian kinerja adalah proses dimana organisasi mengawasi pelaksanaan kerja individu. Dalam penilaian kinerja dinilai kontribusi karyawan kepada organisasi selama periode tertentu. Umpan balik penilaian kinerja memungkinkan karyawan mengetahui seberapa baik mereka bekerja jika dibandingkan dengan standar organisasi (Simamora, 2004). Kinerja karyawan, yaitu hasil kerja atau taraf kesuksesannya dalam melaksanakan pekerjaan, ternyata dipengaruhi oleh beberapa faktor. Pada dasarnya kinerja karyawan merupakan hasil proses yang

kompleks, baik berasal dari diri pribadi karyawan (internal factor), keadaan lingkungan perusahaan (external factor) maupun upaya strategis dari perusahaan. Kinerja dari suatu pegawai dapat dinilai dengan bertitik tumpu pada data yang ada. Selama ini perusahaan memiliki sekumpulan data terkait data aplikasi Carpool yang paramaternya bisa digunakan untuk mengukur kinerja karyawan. Data tersebut dapat dimanfaatkan dengan baik. Melalui penerapan teknologi data mining, sekumpulan data yang belum dimanfaatkan tersebut akan diproses dan menghasilkan *new knowledge* (pengetahuan baru) yang menjadi sesuatu yang berhargadan berguna bagi perusahaan.

Maka dari itu peneliti tertarik untuk melakukan proses *data mining* pada data aplikasi Carpool untuk melihat kinerja karyawan yang berasal dari laporan table status pada basis data yang ada. guna mendapatkan informasi yang dapat digunakan untuk memprediksi nilai kinerja karyawan. Metode yang peneliti gunakan adalah metode Klasifikasi dengan *Algoritma C.45* yang merupakan cabang dalam *data mining*. Melalui metode klasifikasi *Algoritma C.45* dengan aplikasi Rapidminer 5 yang digunakan untuk mengukur akurasi daeri kinerja Algoritma C.45 dalam meghasilkan rule sebagai landasan dalam prediksi nilai kinerja karyawan.

Bab 2

Landasan Teori

2.1. Studi Literatur

Penelitian ini tidak terlepas dari hasil penelitian-penelitian terdahulu yang pernah dilakukan sebagai tolak ukur dan acuan. Adapun hasil penelitian-penelitian yang dijadikan tolak ukur dan acuan tidak terlepas dari topik yang dilakukan

2.1.1. ANALISA SENTIMEN PENGGUNA E-MONEY PADA TWITTER MENGGUNAKAN ALGORITMA C4.5 DAN NA'VE BAYES

Pengumpulan (crawling) data melalui API twitter dengan maksimal sekali iproses menghasilkan 100 data. Dan menghasilkan 300 data tweet dengan 70 untuk data Training dan 30 persen data latih. Metode Algoritma C4.5 dan Naive bayes dengan perbandingan menggunakan voting by majority dapat melakukan perbandingan secara Jurnal Informatika dan Sistem Informasi (JIFoSI) Vol.1,No.3 November 2020 ANALISIS SENTIMEN PENGGUNA E-MONEY PADA TWITTER MENGGUNAKAN ALGORITMA C4.5 DAN

NA'IVE BAYES — 1071 otomatis. Uji coba dilakukan dengan menggunakan data testing sebanyak 90 kalimat. Setiap data diklasifikasikan dengan label positif, negatif dan undefined. Performa dari hasil metode Na'ive Bayes dan C4.5 mencapai akurasi 80 persen dan 84 persen dan hasil voting sebesar 60 persen ini. Pembaca diharapkan dapat meningkatkan system dengan menerapkan metode algoritma lebih banyak agar meningkatkan kinerja dari meto devoting by majority, dan penambahan koleksi kamus pada database kata gaul akan meningkatkan kinerja pada pemilihan atribut.

2.1.2. THE APPLICATION OF DATA MINING IN DETERMINING TIMELY GRADUATION USING THE C45 ALGORITHM

Kelulusan tepat waktu merupakan salah satu unsur penilaian akreditasi perguruan tinggi. Dalam Strata 1, mahasiswa dinyatakan lulus tepat waktu jika dapat menyelesaikan studinya \leq delapan semester atau empat tahun. BAN-PT menetapkan standar kelulusan tepat waktu $\geq 50\%$. Jika standar tidak terpenuhi, maka akan menurunkan nilai akreditasi. Masalah-masalah ini mendorong Universitas Simalungun Pematangsiantar untuk melakukan evaluasi dan langkah strategis dalam upaya peningkatan angka kelulusan mahasiswa agar target BANPT dapat tercapai.

Untuk itu perlu diketahui terlebih dahulu polamahasiswa yang cenderung tidak lulus tepat waktu. Dalam penelitian ini,

Algoritma C4.5 diusulkan untuk memprediksi kelulusan siswa. Algoritma ini akan memproses dataset profil siswa sebanyak 150 data. Dataset ini memiliki label status kelulusan. Nilai label bersifat kategoris, yaitu, benar dan terlambat. Ciri atau atribut yang digunakan yaitu nama siswa, jenis kelamin, status mahasiswa, IPK. Hasil dari algoritma C4.5 berupa model pohon keputusan yang sangat mudah untuk dianalisis. Bahkan oleh orang biasa sekalipun. Ini model akan memetakan pola mahasiswa yang berpotensi lulus tepat waktu dan terlambat..

2.1.3. METODE C45 UNTUK MENGLARIFIKASI PELANGGAN PERUSAHAAN TELEKOMUNIKASI SELULER

Penelitian ini dilakukan untuk mengklasifikasikan pelanggan perusahaan telekomunikasi seluler guna menemukan decision tree dengan keputusan masih berlangganan atau berhenti berlangganan. Klasifikasi menggunakan algoritma C45 yang akan memproses 54.979 record menggunakan Rapidminer. Data dari perusahaan telekomunikasi yang berupa MS Excel dilakukan pembersihan dari record yang tak lengkap, data ganda, dan data yang salah. Variabel input yang akan diproses untuk klasifikasi adalah Umur Pelanggan, Lama Berlangganan, Penggunaan Data, Pemakaian telepon ke Internasional, Asal kota

(besar/kecil). Sedangkan variabel target adalah Berhenti Berlangganan atau Masih Berlangganan.

2.1.4. IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA C-45 PADA DATA MASYARAKAT KECAMATAN GAROGA UNTUK MENENTUKAN POLA PENERIMA BERAS RASKIN

Program bantuan sosial merupakan salah satu komponen program jaminan sosial yang menjadi bentuk ekspresi tanggung jawab pemerintah, pemerintah daerah yang sangat peduli terhadap kondisi masyarakat yang miskin dan terlantar di atas akar rumput (grass root level). Dengan menggunakan data mining, setiap kumpulan atau gudang data dapat memberikan pengetahuan penting yang menjadi informasi yang sangat berharga bagi suatu pemerintahan, seperti pada pemerintahan. Pada sebuah pemerintahan, suatu sistem informasi dapat digunakan untuk memperoleh informasi yang menunjang setiap kegiatan pada pengambilan suatu keputusan. Data mining algoritma C4.5 dapat digunakan untuk menyusun sistem yang mempunyai kemampuan melihat pola penerimaan beras raskin, untuk selanjutnya bisa menjadi strategi dalam proses pengambilan keputusan untuk penerimaan beras raskin. Pada penelitian ini peneliti menggunakan algoritma C4.5 dalam menentukan masyarakat yang mendapat beras raskin di Kecamatan Garoga.

2.1.5. DESIGN of CUSTOMER SATISFACTION APPLICATION at BCA KEP RENGASDENGKLOK USING C.45 ALGORITHM METHOD

Bank BCA selalu meningkatkan kualitas layanan sesuai dengan slogan “Senantiasa Disisi Anda”. Penilaian tersebut terdiri dari 4 atribut (waktu, akurat, fokus dan kepuasan). Setiap atribut memiliki bobot nilai angka 1 (sangat tidak puas) sampai 5 (sangat puas). Penilaian tersebut masih dilakukan secara manual (menggunakan kertas), oleh sebab itu penulis pada penelitian ini menggunakan algoritma C.45 dengan dilakukan 3 kali pengujian sehingga menghasilkan klasifikasi yang diperoleh bahwa nilai akurasi yaitu mencapai 88,75 persen dengan nilai AUC yaitu 0,744 dan pengujian pada aplikasi yang dibuat menghasilkan 0,722. Dapat disimpulkan bahwa penilaian layanan di BCA KCP Rengasdengklok termasuk kelompok klasifikasi yang cukup baik dikarenakan nilai AUC-nya antara 0.70-0.80. kesimpulan dari penelitian ini:

Aplikasi berbasis web yang dikembangkan dapat mempermudah evaluasi kinerja staf di Contact Center PT.XYZ. Metode perhitungan dalam aplikasi berikut ketentuan atau standar yang ada di internal organisasi, baik target yang ingin dicapai maupun referensi penilaian.

2.1.6. IMPLEMENTATION OF DATA MINING TO ANALYZE DRUG CASES USING C4.5 DECISION TREE

Data mining adalah proses menemukan informasi yang berguna dari sekumpulan database. Salah satu teknik yang ada dalam data mining adalah klasifikasi. Metode yang digunakan adalah metode pohon keputusan dan algoritma yang digunakan adalah algoritma C4.5. Metode pohon keputusan adalah metode yang mengubah fakta yang sangat besar menjadi pohon keputusan yang menyajikan aturan. Metode pohon keputusan berguna untuk mengeksplorasi data, serta menemukan yang tersembunyi hubungan antara sejumlah variabel input potensial dengan variabel target. Keputusan pohon dari algoritma C4.5 dibangun dengan beberapa tahapan diantaranya pemilihan atribut sebagai root, membuat cabang untuk setiap nilai dan membagi kasus ke dalam cabang.

Ini tahapan akan diulang untuk setiap cabang sampai semua kasus di cabang memiliki kelas yang sama. Dari solusi pohon keputusan akan ada beberapa aturan dari suatu kasus. Dalam hal ini peneliti mengklasifikasikan data narapidana di Lapas Labuhan Deli untuk mengetahui faktor-faktor narapidana yang melakukan tindak pidana narkoba. Dengan menerapkan algoritma C4.5 ini, maka Pengetahuan tersebut diperoleh sebagai

informasi untuk meminimalisir tindak pidana narkoba. Dari temuan penelitian, ditemukan bahwa faktor yang paling berpengaruh dari tahanan yang dilakukan tindak pidana narkoba berasal dari variabel.

2.1.7. PENERAPAN ALGORITMA KLASIFIKASI C4.5 DALAM REKOMENDASI PENERIMAAN MITRA PENJUALAN STUDI KASUS : PT ATRIA ARTHA PERSADA

Permasalahan yang sering muncul dalam bisnis pada penjualan dengan sistem pembayaran kredit tempo adalah antara lain seperti kredit macet, order fiktif dan penipuan. Jika tidak menggunakan prosedur yang benar dalam penerimaan mitra penjualan bukan tidak mungkin masalah tersebut akan timbul dalam proses bisnis. Untuk mengembangkan bisnis tentu perlu meningkatkan penjualan dan dibarengi dengan pembayaran yang lancar. Oleh karena itu, pada proses penerimaan mitra perlu dibuatkan prosedur yang baik agar dapat meminimalisir kemungkinan masalah yang ada. Pada prosedur penerimaan mitra penjualan perlu dilakukan proses analisa untuk menentukan calon mitra penjualan tersebut diterima atau ditolak. Dari data mitra yang ada dapat digunakan untuk dilakukan pengolahan data. Salah satu Teknik pengolahan data yang dapat digunakan dalam proses tersebut adalah klasifikasi.

Klasifikasi adalah teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan. Dan menggunakan Algoritma C4.5 dalam proses pengklasifikasiannya. Yang dapat menentukan apakah calon mitra dapat diterima sebagai mitra atau tidak. Kemudian, Hasil dari pengklasifikasiannya divalidasi dengan tenfold cross validation dengan tingkat akurasi 96,26 %, presisi 100% dan recall 71,43%

Penerapan algoritma klasifikasi C4.5 akan dapat diimplementasikan pada rekomendasi penerimaan mitra penjualan di PT. Atria Artha Persada, dilihat dari tingkat accuracy yang mencapai 96.26 % dan recall 71.43%., yang menyatakan bahwa perhitungan yang dilakukan akan mampu memprediksi dan merekomendasikan penerimaan mitra penjualan dengan baik.

2.1.9. PENERAPAN METODE C4.5 UNTUK KLASIFIKASI MAHASISWA BERPOTENSI DROP OUT. ILKOM

Kualitas pendidikan di perguruan tinggi dapat dilihat dari tingginya tingkat keberhasilan mahasiswa dan rendahnya kegagalan mahasiswa. Salah satu indikator kegagalan mahasiswa adalah kasus Drop Out (berhenti studi). Permasalahan Drop Out menjadi sesuatu yang menarik untuk diteliti, karena hal ini dapat mempengaruhi kualitas pendidikan tersebut. Fakultas Ekonomi UNISAN Gorontalo

merupakan Fakultas favorit di UNISAN Gorontalo sehingga memiliki jumlah mahasiswa kurang lebih 1000 mahasiswa sampai tahun 2017. Namun rasio jumlah mahasiswa lulus dan tidak lulus tidak seimbang. Sehingga menghasilkan jumlah mahasiswa Drop Out kurang lebih 200 mahasiswa pertahunnya. Untuk mengatasi masalah tersebut diusulkan model baru dengan memanfaatkan sebuah metode komputasi C4.5, agar menghasilkan pola berdasarkan hasil klasifikasi yang tepat dalam menentukan mahasiswa yang berpotensi Drop Out. Adapun hasil yang diperoleh dari penerapan metode C4.5 pada penelitian ini yaitu ditemukannya 17 rule yang dapat dijadikan sebagai pola untuk menentukan mahasiswa yang berpotensi Drop Out.

2.2. METODE C.45

Algoritma C.45 Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data. (Irmayani dan Haris, 2020)

Contoh Perhitungan C45

no	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Cloudy	Mild	High	False	Yes
5	Rainy	Cool	High	False	Yes
6	Rainy	Cool	Normal	True	Yes
7	Rainy	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	Yes
9	Sunny	Cool	Normal	False	No
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Cloudy	Mild	High	True	Yes
13	Cloudy	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

Rumus Entropy: S = Himpunan Kasus. n = Jumlah Partisi S . p_i = Proporsi dari S_i terhadap S .

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

Rumus Gain: S = Himpunan Kasus. A = Atribut. n = Jumlah Partisi Atribut A. | Si | = Jumlah Kasus pada partisi ke-i. | S | = Jumlah Kasus dalam S

Perhitungan Entropy dan Gain Akar

Entropy Total

$$\text{Entropy(Total)} = \left(-\frac{4}{14} * \log_2\left(\frac{4}{14}\right)\right) + \left(-\frac{10}{14} * \log_2\left(\frac{10}{14}\right)\right)$$
$$\text{Entropy(Total)} = 0.863120569$$

Entropy Outlook

$$\text{Entropy(Cloudy)} = \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) = 0.000000000$$
$$\text{Entropy(Rainy)} = \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) = 0.721928095$$
$$\text{Entropy(Sunny)} = \left(-\frac{3}{5} * \log_2\left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5} * \log_2\left(\frac{2}{5}\right)\right) = 0.970950594$$

Entropy Temperature

$$\text{Entropy(Cool)} = \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) = 0.000000000$$
$$\text{Entropy(Hot)} = \left(-\frac{2}{4} * \log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} * \log_2\left(\frac{2}{4}\right)\right) = 1.000000000$$
$$\text{Entropy(Mild)} = \left(-\frac{2}{6} * \log_2\left(\frac{2}{6}\right)\right) + \left(-\frac{4}{6} * \log_2\left(\frac{4}{6}\right)\right) = 0.918295834$$

Entropy Humidity

$$Gain(Total, Humidity) = Entropy(Total) - \sum_{i=1}^n \frac{|Humidity_i|}{|Total|} * Entropy(Humidity_i)$$

$$Gain(Total, Humidity) = 0.863120569 - \left(\left(\frac{7}{14} * 0.985228136 \right) + \left(\frac{7}{14} * 0.000000000 \right) \right)$$

$$Gain(Total, Humidity) = 0.370506501$$

Gain

$$Gain(Total, Windy) = Entropy(Total) - \sum_{i=1}^n \frac{|Windy_i|}{|Total|} * Entropy(Windy_i)$$

$$Gain(Total, Windy) = 0.863120569 - \left(\left(\frac{8}{14} * 0.811278124 \right) + \left(\frac{6}{14} * 0.918295834 \right) \right)$$

$$Gain(Total, Windy) = 0.005977711$$

NODE			Jml Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1	TOTAL		14	4	10	0,863120569	
	OUTLOOK						0,258521037
		CLOUDY	4	0	4	0,000000000	
		RAINY	5	1	4	0,721928095	
		SUNNY	5	3	2	0,970950594	
	TEMPERATURE						0,183850925
		COOL	4	0	4	0,000000000	
		HOT	4	2	2	1,000000000	
		MILD	6	2	4	0,918295834	
	HUMIDITY						0,370506501
		HIGH	7	4	3	0,985228136	
		NORMAL	7	0	7	0,000000000	
	WINDY						0,005977711
		FALSE	8	2	6	0,811278124	
		TRUE	6	4	2	0,918295834	

Sumber : Prayudi (2013:16)

2.2.1. Data Mining

Data mining, sering disebut juga sebagai Knowledge Discovery in Database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data-data yang berukuran besar (Santoso, 2007). Dan keluaran dari Data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah pattern recognition sekarang jarang digunakan karena sudah termasuk bagian dari Data mining.

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (Decision Tree). Pohon keputusan merupakan metode klasifikasi dan prediksi yang terkenal. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain : ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3, Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- Pilih atribut sebagai akar.
- Buat cabang untuk tiap-tiap nilai.

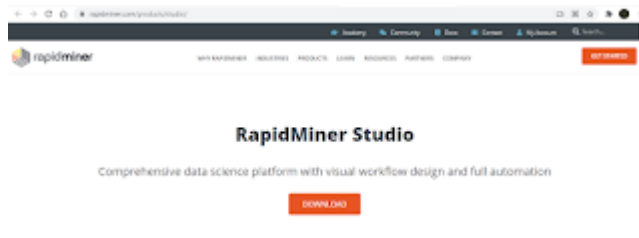
-
- Bagi kasus dalam cabang.
 - Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Konsep Entropy :

- Entropy (S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S.
- Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas.
- Entropy digunakan untuk mengukur ketidakaslisan S.

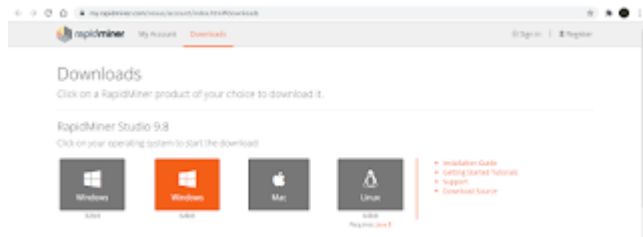
2.2.1.1. Langkah-langkah Download RapidMiner Studio

1. Kunjungi <https://rapidminer.com/products/studio/>, lalu klik tombol **download**.



Gambar 1 - Download RapidMiner Studio

2. Setelah itu Anda akan diarahkan ke <https://my.rapidminer.com/nexus/account/index.html#downloads>.

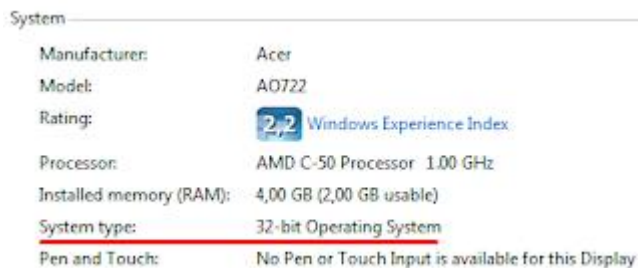


Gambar 2 - Memilih software sesuai dengan sistem operasi yang digunakan

3. Download sesuai dengan sistem operasi (dan arsitektur bit) yang dimiliki.

Bagi pengguna Windows yang belum tau bit pada sistem operasi yang digunakan, bisa dilihat melalui:

- Klik tombol keyboard **windows**.
- Klik kanan pada **Computer**, lalu **Properties**.
- Lihat pada "system type"



Gambar 3 - Tampilan arsitektur sistem operasi

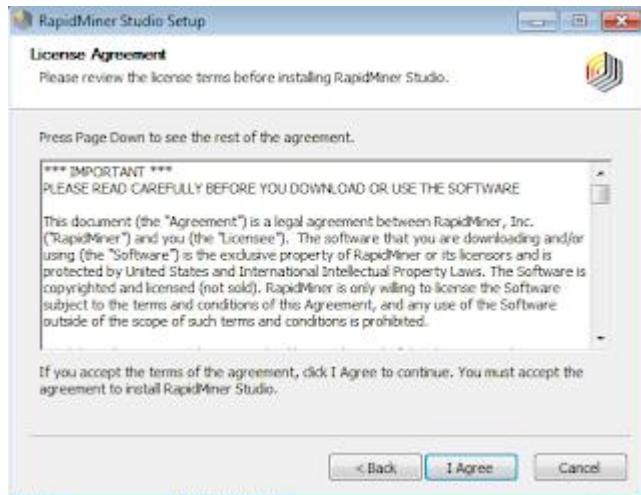
Tahapan Instalasi

1. Jalankan file **rapidminer-studio-*.*.*-install.exe** - **Next**.



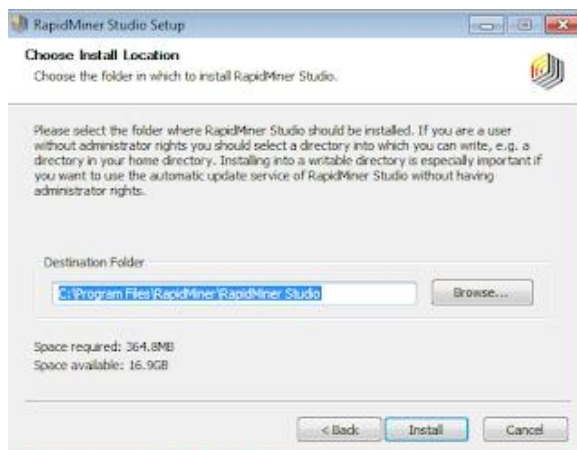
Gambar 4 - RapidMiner Setup Wizard

2. Pada **Licence Agreement**, klik **I Agree**.



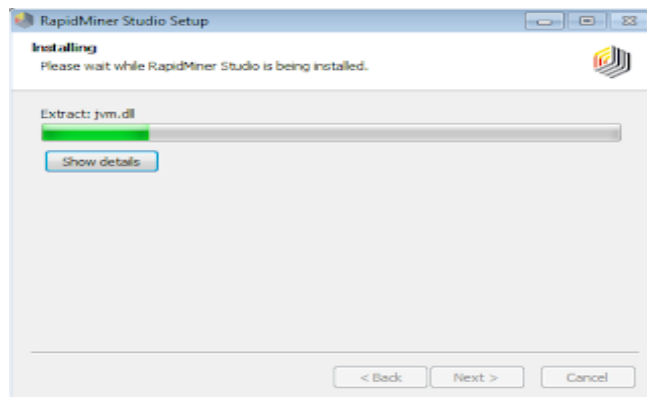
Gambar 5 - Licence Agreement

3. Atur lokasi tempat instalasi RapidMiner Studio (saran: ikuti defaultnya saja)



Gambar 6 - Memilih lokasi instalasi software

4. Tunggu hingga proses instalasi selesai.



Gambar 7 - Menunggu proses instalasi



Gambar 8 - Instalasi selesai!

3.1. PYTHON

Python adalah salah satu bahasa pemrograman yang dapat melakukan eksekusi sejumlah instruksi multi guna secara langsung (interpretatif) dengan metode orientasi objek (Object Oriented Programming) serta menggunakan semantik dinamis untuk memberikan tingkat keterbacaan syntax. Sebagian lain mengartikan Python sebagai bahasa yang kemampuan, menggabungkan kapabilitas, dan sintaksis kode yang sangat jelas, dan juga dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Walaupun Python tergolong bahasa pemrograman dengan level tinggi, nyatanya Python dirancang sedemikian rupa agar mudah dipelajari dan dipahami.

Python sendiri menampilkan fitur-fitur menarik sehingga layak untuk Anda pelajari. Pertama, Python memiliki tata bahasa dan script yang sangat mudah untuk dipelajari. Python juga memiliki sistem pengelolaan data dan memori otomatis. Selain itu modul pada Python selalu diupdate. Ditambah lagi, Python juga memiliki banyak fasilitas pendukung. Python banyak diaplikasikan pada berbagai sistem operasi seperti Linux, Microsoft Windows, Mac OS, Android, Symbian OS, Amiga, Palm dan lain-lain.

Python dibuat dan dikembangkan oleh Guido Van Rossum, yaitu seorang programmer yang berasal dari Belanda. Pembuatannya berlangsung di kota Amsterdam, Belanda pada tahun 1990. Pada tahun 1995 Python dikembangkan lagi agar lebih kompatibel oleh Guido Van Rossum. Selanjutnya pada awal tahun 2000, terdapat pembaharuan versi Python hingga mencapai Versi 3 sampai saat ini. Pemilihan nama Python sendiri diambil dari sebuah acara televisi yang lumayan terkenal yang bernama *Moth Python Flying Circus* yang merupakan acara sirkus favorit dari Guido van Rossum.

Python diciptakan oleh Guido van Rossum pertama kali di *Centrum Wiskunde & Informatica* (CWI) di Belanda pada awal tahun 1990-an. Bahasa python terinspirasi dari bahasa pemrograman ABC. Sampai sekarang, Guido masih menjadi penulis utama untuk python, meskipun bersifat *open source* sehingga ribuan orang juga berkontribusi dalam mengembangkannya.

Di tahun 1995, Guido melanjutkan pembuatan python di *Corporation for National Research Initiative* (CNRI) di Virginia Amerika, di mana dia merilis beberapa versi dari python.

Pada Mei 2000, Guido dan tim Python pindah ke BeOpen.com dan membentuk tim BeOpen PythonLabs. Di bulan Oktober pada tahun yang sama, tim python pindah ke Digital Creation (sekarang menjadi Perusahaan Zope). Pada tahun 2001,

dibentuklah Organisasi Python yaitu *Python Software Foundation* (PSF). PSF merupakan organisasi nirlaba yang dibuat khusus untuk semua hal yang berkaitan dengan hak intelektual Python. Perusahaan Zope menjadi anggota sponsor dari PSF.

Semua versi python yang dirilis bersifat open source. Dalam sejarahnya, hampir semua rilis python menggunakan lisensi GFL-compatible. Berikut adalah versi mayor dan minor python berikut tanggal rilisnya

Nama Python sendiri sebenarnya bukan diambil dari nama ular piton, namun nama Python tersebut diambil dari sebuah acara humor yang ditayangkan oleh channel BBC pada tahun 1980-an yang berjudul “Monty Python’s Flying Circus” dimana saat itu Monty Python merupakan sekelompok grup lawak yang membawakan acara tersebut dan Guido van Rossum sangat menyukai acara “Monty Python’s Flying Circus” tersebut.

Guido van Rossum saat membuat bahasa pemrograman Python sendiri terinspirasi dari bahasa pemrograman ABC. Dan sampai sekarang ini Guido masih menjadi seorang penulis utama bahasa pemrograman Python walaupun bahasa pemrograman Python tersebut bersifat open source sehingga memungkinkan ribuan orang untuk ikut berkontribusi mengembangkan bahasa pemrograman Python tersebut agar menjadi lebih baik.

Pada tahun 1994, akhirnya Python versi 1.0 dirilis kepada publik, selanjutnya pada tahun 1995 Guido melanjutkan

pembuatan Python di Corporation for National Research Initiative (CNRI) yang berada di Virginia Amerika Serikat. Selanjutnya pada tahun 2000 Python versi 2.0 dirilis kepada publik dan Python versi 3.0 dirilis pada tahun 2008.

Di Bulan Mei tahun 2000, Guido van Rossum bersama dengan tim Python pindah ke BeOpen.com untuk selanjutnya membuat sebuah tim BeOpen PythonLabs. Pada bulan Oktober tahun 2000, tim Python tersebut kembali pindah ke Digital Creation yang mana sekarang menjadi perusahaan Zope.

Selanjutnya pada tahun 2001, akhirnya dibuatlah sebuah organisasi Python Software Foundation (PSF) yang merupakan organisasi Nirlaba yang khusus dibuat untuk semua hal yang berhubungan dengan hal intelektual dari hasil pengembangan bahasa pemrograman Python.

- Python 1.0 – Januari 1994
- Python 1.2 – 10 April 1995
- Python 1.3 – 12 Oktober 1995
- Python 1.4 – 25 Oktober 1996
- Python 1.5 – 31 Desember 1997
- Python 1.6 – 5 September 2000
- Python 2.0 – 16 Oktober 2000
- Python 2.1 – 17 April 2001
- Python 2.2 – 21 Desember 2001
- Python 2.3 – 29 Juli 2003

-
- Python 2.4 – 30 Nopember 2004
 - Python 2.5 – 19 September 2006
 - Python 2.6 – 1 Oktober 2008
 - Python 2.7 – 3 Juli 2010
 - Python 3.0 – 3 Desember 2008
 - Python 3.1 – 27 Juni 2009
 - Python 3.2 – 20 Februari 2011
 - Python 3.3 – 29 September 2012
 - Python 3.4 – 16 Maret 2014
 - Python 3.5 – 13 September 2015
 - Python 3.6 – 23 Desember 2016
 - Python 3.7 – 27 Juni 2018
 - Python 3.8 – 14 Oktober 2019
 - Python 3.9 – 5 Oktober 2020

Banyak orang yang tertarik untuk menggunakan Python karena dianggap mudah untuk dipelajari, sekalipun oleh para pemula. Kode-kode yang ada didalamnya mudah dibaca dan dapat menjalankan banyak fungsi kompleks dengan mudah karena banyaknya standard library. Pengembangan program pada Python pun dapat dilakukan dengan cepat dan menggunakan lebih sedikit kode. Bahkan Python mampu menjadikan program dengan skala sangat rumit menjadi mudah. Python sendiri mendukung multi

platform dan multi system serta memiliki sistem pengelolaan memori otomatis seperti Java.

Sayangnya Python cukup lambat dijalankan. Untuk pengembangan platform Android dan IOS juga terbilang kurang support. Python juga memiliki keterbatasan dengan akses basis data. Selain itu Python tidak cocok untuk melakukan tugas-tugas intensif memori dan pekerjaan multi-core/ multi-processor.

Demikian sekilas informasi tentang bahasa pemrograman Python yang bisa Anda simak. Bagi pengembang atau programmer pemula sangat disarankan untuk mempelajari bahasa pemrograman tersebut karena dinilai mudah. Python juga dapat diterapkan dalam pembuatan software, aplikasi pada smartphone, program GUI, program CLI, Internet of Things, games dan lain-lainnya.

Bahasa pemrograman Python cocok digunakan sehari-hari, sebab dalam pengembangannya cukup singkat Open source, bisa dikembangkan oleh siapa saja Bahasa yang mudah dipahami dan intuitif sebanding dengan bahasa pemrograman lain Menggunakan kode bahasa Inggris biasa. Versi pertama Python saat pertama kali dirilis adalah versi 0.9.0. Kemudian pada tahun 2000, Python 2.0 dirilis dengan pengembangan pada daftar, Unicode, dan pengumpulan bug.

Python yang teranyar adalah versi 3.6.4 dengan pengembangan berbagai macam fitur mulai dari Syntax, Generator

Asynchronous, garis bawah di numeric literal dan format string literal.

3.2. DATA MINING

Data mining adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data. Proses pengumpulan dan ekstraksi informasi tersebut dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan statistika, matematika, ataupun teknologi *Artificial Intelligence* (AI). Data mining sering disebut juga *Knowledge Discovery in Database* (KDD) (Putria, 2018).

Data mining dilakukan untuk memenuhi beberapa tujuan tertentu. Berikut ini adalah tujuan-tujuan dari data mining:

- **Sebagai sarana menjelaskan (*Explanatory*)**

Data mining dapat digunakan sebagai sarana untuk menjelaskan suatu kondisi penelitian.

- **Sebagai sarana konfirmasi (*Confirmatory*)**

Data mining dapat digunakan sebagai sarana untuk memastikan sebuah pernyataan atau mempertegas suatu hipotesis.

- **Sebagai sarana eksplorasi (*Exploratory*)**

Data mining dapat digunakan sebagai sarana untuk mencari pola baru yang sebelumnya tidak terdeteksi.

Secara umum, terdapat beberapa metode yang digunakan untuk melakukan data mining. Berikut ini adalah metodenya:

(1) ***Association***

Teknik yang pertama adalah association. Association adalah metode berbasis aturan yang digunakan untuk menemukan asosiasi dan hubungan variabel dalam satu set data. Biasanya analisis ini terdiri dari pernyataan “if atau then” sederhana. Association banyak digunakan dalam mengidentifikasi korelasi produk dalam keranjang belanja untuk memahami kebiasaan konsumsi pelanggan. Sehingga, perusahaan dapat mengembangkan strategi penjualan dan membuat sistem rekomendasi yang lebih baik.

(2) ***Classification***

Selanjutnya classification, ia adalah metode yang paling umum digunakan dalam data mining. Classification adalah tindakan untuk memprediksi kelas suatu objek.

(3) ***Regression***

Regression adalah teknik yang menjelaskan variabel dependen melalui proses analisis variabel independen. Sebagai contoh, prediksi penjualan suatu produk berdasarkan korelasi antara harga produk dengan tingkat pendapatan rata-rata pelanggan.

(4) *Clustering*

Terakhir, metode clustering. Clustering digunakan dalam membagi kumpulan data menjadi beberapa kelompok berdasarkan kemiripan atribut yang dimiliki. Contoh kasusnya adalah Customer Segmentation. Ia membagi pelanggan ke dalam beberapa grup berdasarkan tingkat kemiripannya..

3.3. *DECISION TREE*

Pohon keputusan adalah alat pendukung keputusan yang menggunakan model keputusan seperti pohon dan kemungkinan konsekuensinya, termasuk hasil acara kebetulan, biaya sumber daya, dan utilitas. Ini adalah salah satu cara untuk menampilkan algoritma yang hanya berisi pernyataan kontrol bersyarat.

Decision tree adalah diagram yang dapat digunakan untuk memvisualisasikan proses pengambilan keputusan dengan memetakan berbagai tindakan dan potensi hasil yang ada. (Meythi, 2012).

Pada dasarnya diagram decision tree terdiri dari tiga elemen berbeda, yaitu:

- Node akar. Ini adalah bagian diagram paling atas yang mewakili tujuan akhir atau keputusan besar yang kamu sedang coba buat. (Yunianita *et.,al* 2018)

-
- Ranting. Ini adalah bagian cabang yang berasal dari akar dan mewakili opsi yang berbeda atau rangkaian tindakan yang tersedia saat membuat keputusan tertentu.
 - Simpul daun. Bagian ini berada pada ujung cabang yang mewakili kemungkinan hasil untuk setiap tindakan. Biasanya ada dua jenis simpul daun dalam decision tree. Daun berbentuk persegi menunjukkan keputusan lain yang harus dibuat, sementara daun berbentuk lingkaran menunjukkan hasil yang tidak diketahui atau kemungkinan perubahan peristiwa.

Ketika dibentuk, elemen-elemen ini akan menyerupai pohon. Itu sebabnya diagram ini lalu disebut sebagai decision tree atau pohon keputusan. Seperti yang disebutkan sebelumnya, decision tree dapat digunakan untuk membantu perusahaan atau manajer dalam proses pengambilan keputusan yang efektif. Lebih jelasnya, beberapa manfaat decision tree adalah:

1. Menawarkan kejelasan

Decision tree memperjelas pilihan, risiko, tujuan dan keuntungan untuk setiap pilihan. Sehingga kamu bisa memetakan berbagai kemungkinan dan pada akhirnya dapat menentukan tindakan mana yang memiliki kemungkinan sukses tertinggi.

Ini akan melindungi keputusan kamu dari hasil atau risiko yang tidak diinginkan.

2. Efisiensi

Decision tree menyajikan informasi dengan cara yang mudah dipahami sehingga dengan cepat dapat dianalisis dan membantu mengambil keputusan penting dengan efisien.

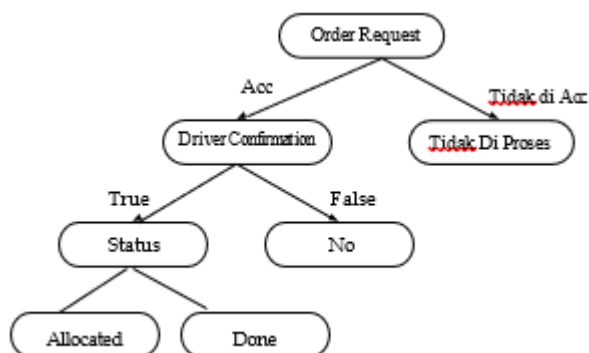
3. Kompatibilitas

Teknik decision tree dapat digunakan dengan metodologi manajemen proyek lainnya.

4. Menghindari bias

Walaupun berkonsultasi dengan orang lain saat membuat keputusan penting bisa membantu kamu, namun terlalu mengandalkan pendapat kolega atau orang lain bisa jadi berisiko. Sebab, saran yang diberikan turut dipengaruhi pada bias pribadi.

Sebaliknya decision tree memberikan pandangan yang seimbang karena didasarkan pada perhitungan risiko dan imbalan yang kemungkinan didapatkan pada setiap pilihan keputusan.



Setelah Atribut Order Request dijadikan akar dalam pohon keputusan, maka yang akan dijadikan node cabang adalah Status. Apabila Order Request bernilai Acc dan Status bernilai Done maka orderan bernilai Yes atau transaksi selesai. Apabila Order Request bernilai Acc dan Status bernilai Allocated maka order selesai bernilai No.

3.4. Database

Database merupakan kumpulan file-file yang saling berkaitan dan berinteraksi, relasi tersebut bila ditunjukkan dengan kunci dari tiap-tiap file yang ada. Satu database menunjukkan suatu kumpulan data yang dipakai dalam suatu lingkup perusahaan, instansi. Pengolahan database merupakan suatu cara yang dilakukan terhadap file-file yang berada di suatu instansi yang mana file tersebut dapat disusun, diurut, diambil sewaktu-waktu serta dapat ditampilkan dalam bentuk suatu laporan sehingga dapat mengolah file-file yang berisikan informasi tersebut secara rapi.

Database atau yang dikenal juga dengan istilah basis data adalah sekumpulan data yang dikelola dengan sedemikian rupa berdasarkan ketentuan tertentu yang saling berkaitan sehingga memudahkan dalam pengelolaannya.

Lewat pengelolaan itulah pengguna bisa mendapatkan kemudahan dalam mencari sebuah informasi, membuang informasi, maupun menyimpan informasi.

Selain pengertian database diatas, terdapat pengertian lain dari database yaitu sebuah sistem yang berguna untuk mengumpulkan file, arsip, atau tabel yang disimpan dan terhubung dalam berbagai media elektronik yang ada. (Eleanora & Wijanarko. 2022)

Jenis dan Fungsi Database

Hal itu sering disebut sebagai data management system (DBMS) atau sistem manajemen basis data. Database memiliki beberapa jenis dan fungsi yang beraneka macam. Selama ini database mempunyai 5 jenis database yang beroperasi pada perangkat.

Dari kelima jenis database tersebutlah, terdapat beberapa macam produk yang mempunyai fungsi berbeda-beda. Berikut jenis-jenis database beserta fungsinya. (Abdul Kadir. 2008)

1. Operational Database

Operational Database atau biasa dikenal dengan nama database On Line Transaction Processing memiliki fungsi sebagai suatu tempat untuk mengelola data dinamis secara langsung dan real-time. Jenis ini memungkinkan para penggunanya untuk bisa melihat, melakukan, dan memodifikasi data. Modifikasi data yang dimaksud yakni dengan cara menambah atau mengubah, ataupun

menghapus data secara langsung lewat suatu perangkat keras yang dipakai. (Wahana Komputer, 2010)

a. JSON

JSON atau JavaScript Object Notation merupakan format file yang memakai teks guna melakukan pengiriman data. Format tersebut biasa dipakai oleh pengguna untuk melakukan pertukaran data layaknya berinteraksi cepat melalui web browser maupun web server.

Sinkronisasi data dapat dikerjakannya secara real-time. Format JSON sendiri berasal dari pemrograman JavaScript. Melalui pembuatan tersebut JSON mempunyai sebuah format bahasa yang berbeda dengan lainnya. Perlu diketahui bahwa semua file JSON selalu memakai ekstensi khusus berupa “.json”

b. XML

XML atau Extensible Markup Language merupakan sebuah bahasa pada program markup yang mempunyai aturan guna memberikan dua kode dokumen yang berbeda agar dapat dibaca oleh manusia dan komputer.

Lewat XML tersebutlah, sistem ini akan menghasilkan sebuah format data yakni teks yang bisa dipakai untuk merepresentasikan sebuah struktur pada basis data. Selain itu, sinkronisasi data pada sistem ini bisa dilakukan secara real-time bagi para penggunanya.

Bahasa program ini sangat cocok dipakai untuk menangani basis data pada web browser maupun web server. Struktur yang digunakan XML dinilai banyak mempunyai kesamaan dengan format JSON.

1. Relational Database

Relational Database atau basis data relasional merupakan basis data yang mengorganisir berdasarkan pada model hubungan data. Banyak sekali perangkat lunak yang memakai sistem ini untuk mengatur dan memelihara basis data melalui hubungan setiap data.

Pada biasanya, semua sistem memakai Structured Query Language atau SQL sebagai bahasa pemrograman untuk pemeliharaan basis data dan query.

a. MySQL

MySQL adalah basis data sumber terbuka yang cukup populer dan biasa dipakai untuk aplikasi berbasis web contohnya website dinamis dan e-commerce. Tahun 2013 MySQL, ialah basis data kedua yang paling banyak dipakai di dunia. Serta yang pertama bagi basis data sumber terbuka.

MySQL yaitu sebuah sistem bagi manajemen basis data relasional. Banyak sekali produk-produk IT yang dibuat dengan memakai komponen utama MySQL.

Beberapa aplikasi tersebut diantaranya Google, WordPress, Flickr, Facebook, Youtube, Joomla, phpBB, MODx,

Drupal, dan lainnya yang memakai sistem ini guna manajemen basis data relasional mereka.

Penggunaan MySQL sebagai basis data utama bagi aplikasi web juga sering dikaitkan dengan PHP sebagai bahasa skrip yang berorientasi pada objek. MySQL merupakan salah satu komponen penting dari web service solution stack LAMP (Linux, MySQL, Apache, and PHP) yakni platform pengembangan web sumber terbuka. Linux sebagai sistem operasi, MySQL sebagai basis data, Apache sebagai Web Server, dan PHP sebagai bahasa skrip.

Manfaat Database

Dari berbagai jenis dan fungsi dari database, terdapat juga manfaat dari database bagi penggunaanya. Berikut adalah manfaatnya :

1. Kecepatan dan kemudahan

Sistem database memberikan sebuah kemampuan dalam menyeleksi data menjadi satu kelompok secara berurutan dengan cepat. Instrumen tersebut juga dapat menghasilkan pencarian informasi yang diperlukan bisa ditemukan secara cepat. Kecepatannya itu juga bisa dipengaruhi oleh jenis database yang dipakai. Setiap jenis dari database tersebut memberikan sebuah kemampuannya secara berbeda-beda.

2. Multi-user

Database memberikan sebuah kemudahan akses bagi para pengguna pada waktu yang bersamaan. Sistem ini juga memungkinkan akses pada suatu dokumen ke lebih dari satu pengguna. Sehingga kinerja mesin dan jaringan dimudahkan lewat multi-user sebab penyimpanan hanya memuat satu unit yang bisa diakses secara bersamaan.

3. Keamanan data

Sistem database melalui bahasa pemrogramannya telah disusun secara safety atau aman Melalui instrumen password membuat data tersebut hanya dapat diakses pada pihak yang diijinkan. Manajemen tersebut sudah menerapkannya pada hampir seluruh jenis sistem database. Sehingga menjadikan keamanan data adalah hal prioritas bagi layanan sistem database.

4. Penghematan biaya perangkat

Mempunyai satu database terpusat sudah cukup bagi para perusahaan besar yang memerlukan pengumpulan data secara ringkas. Hal itu menjadikan perusahaan tidak membutuhkan ruang penyimpanan di tiap tempat yang berbeda. Lewat jaringan internet, cabang perusahaan di daerah terpencil pun dapat melakukan akses data yang ada pada pusat.

5. Kontrol data terpusat

Database tidak membutuhkan server lebih dari satu dalam penggunaannya. Sebab, dengan cukup satu server terpusat untuk

melakukan penyimpanan data, memudahkan data tersebut dapat diakses oleh banyak pengguna. Hal itu memberikan harga yang murah bagi perusahaan untuk melakukan investasi ruang penyimpanan data penting perusahaan. Sebagai contoh adalah kantor perusahaan yang tidak perlu membuat suatu data di tiap divisinya. Setiap divisi dapat mengumpulkan data khusus melalui satu server yang telah ditentukan sehingga laporan untuk atasan dapat menjadi ringkas.

6. Mudah membuat aplikasi

Kaitannya terhadap perusahaan manakala perusahaan memerlukan aplikasi input data yang baru, programmer tidak perlu menyusun ulang struktur database. Dengan memakai struktur database yang telah disusun sebelumnya sudah cukup untuk mengenali aplikasi input data yang baru.

3.5. *PREPROCESSING DATAMINING*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal. *Data*

preprocessing adalah teknik yang digunakan untuk mengubah data mentah dalam format yang berguna dan efisien. Inisiatif ini diperlukan karena data mentah seringkali tidak lengkap dan memiliki format yang tidak konsisten. (Prianto *et al.*, 2020)

Kualitas data sendiri memiliki korelasi langsung dengan keberhasilan setiap proyek yang melibatkan analisis data. *Preprocessing* sendiri melibatkan validasi dan imputasi data. Tujuan dari validasi adalah untuk menilai tingkat kelengkapan dan akurasi data yang tersaring. Di sisi lain, tujuan imputasi adalah untuk memperbaiki kesalahan dan memasukkan nilai yang hilang, baik secara manual atau otomatis melalui program *business process automation* (BPA). *Data preprocessing* biasanya digunakan pada aplikasi *rule-based* dan yang berbasis *database*. Dalam *machine learning*, kegiatan ini sangat penting untuk memastikan bahwa *big data* sudah diformat dan informasi yang dikandungnya dapat ditafsirkan dan algoritma perusahaan.

3.5.1. Tahap Kerja Data Preprocessing

Sejatinya, *data preprocessing* adalah langkah awal yang wajib diterapkan sebelum perusahaan memulai penyaringan *insight*. Namun, supaya bisa berjalan secara optimal, proses kerjanya dibagi menjadi 4 tahap yang berbeda. Meskipun

berbeda, masing-masing tahap kerja saling berkaitan dan memiliki peran penting dalam pembentukan kualitas proses penyaringan data. Nah, supaya lebih jelas, berikut adalah keempat tahap kerja *data preprocessing* yang perlu kamu pelajari.

3.5.1. *Data Cleaning*

Melansir laman Techopedia, tahap kerja pertama dalam *data preprocessing* adalah *data cleaning*.

Dalam tahap kerja ini, data dibersihkan melalui beberapa proses seperti mengisi nilai yang hilang, menghaluskan *noisy data*, dan menyelesaikan inkonsistensi yang ditemukan.

Data juga bisa dibersihkan dengan dibagi menjadi segmen-segmen yang memiliki ukuran serupa lalu dihaluskan (*binning*). Kamu juga bisa menyesuaikannya dengan fungsi regresi linear atau berganda (*regression*), atau dengan mengelompokkannya ke dalam kelompok-kelompok data yang serupa (*grouping*).

3.5.2. *Data Integration*

Tahap kerja berikutnya dalam proses *data preprocessing* adalah *data integration*. Di sini, data dengan representasi yang berbeda disatukan dan semua konflik dalam di

dalamnya diselesaikan. Tahap kerja satu ini merupakan proses lanjutan dari *data cleansing* dengan tujuan untuk membuat data lebih halus.

3.5.3. *Data Transformation*

Data transformation adalah tahap kerja selanjutnya dalam proses *data preprocessing*. Pada tahap ini, data akan dinormalisasi dan digeneralisasikan. Normalisasi sendiri adalah sebuah proses di mana perusahaan memastikan bahwa tidak ada data yang berlebihan.

Semua data akan disimpan dalam satu tempat dan semua dependensinya haruslah logis. Langkah ini juga diambil untuk mentransformasikan data ke dalam bentuk yang sesuai untuk proses *mining*.

3.5.4. *Data Reduction*

Tahap kerja terakhir dalam proses kerja *data preprocessing* adalah *data reduction*. *Data mining* adalah sebuah teknik yang digunakan untuk menangani data dalam jumlah yang besar. Saat bekerja dengan volume data yang besar, proses analisis akan menjadi lebih sulit.

3.5.5. Manfaat Data Preprocessing

Setelah membaca definisinya, jelas bahwa *data preprocessing* adalah sebuah proses kerja yang sangat penting. Selain memperlancar proses *data mining*, langkah ini juga menawarkan beberapa manfaat lain untuk perusahaan.

Seperti apa manfaat lainnya yang ditawarkan *data preprocessing*? Berikut adalah penjelasannya, sesuai disebutkan oleh Science Direct.

- Membuat data lebih mudah untuk dibaca
- Mengurangi beban representasi dalam data
- Mengurangi durasi *data mining* secara signifikan
- Mempermudah proses analisis data dalam *machine learning*

3.6. *FLASK*

Flask adalah sebuah web framework yang ditulis dengan bahasa Python dan tergolong sebagai jenis microframework. Flask berfungsi sebagai kerangka kerja aplikasi dan tampilan dari suatu web. Dengan menggunakan Flask dan bahasa Python, pengembang dapat membuat sebuah web yang terstruktur dan dapat mengatur behaviour suatu web dengan lebih mudah. Flask termasuk pada jenis microframework karena tidak memerlukan suatu alat atau pustaka tertentu dalam penggunaannya. Sebagian besar fungsi dan komponen

umum seperti validasi form, database, dan sebagainya tidak terpasang secara default di Flask. Hal ini dikarenakan fungsi dan komponen-komponen tersebut sudah disediakan oleh pihak ketiga dan Flask dapat menggunakan ekstensi yang membuat fitur dan komponen-komponen tersebut seakan diimplementasikan oleh Flask sendiri. Selain itu, meskipun Flask disebut sebagai microframework, bukan berarti Flask mempunyai kekurangan dalam hal fungsionalitas. Microframework disini berarti bahwa Flask bermaksud untuk membuat core dari aplikasi ini sesederhana mungkin tapi tetap dapat dengan mudah ditambahkan. Dengan begitu, fleksibilitas serta skalabilitas dari Flask dapat dikatakan cukup tinggi dibandingkan dengan framework lainnya.

Flask adalah web framework dari bahasa python. Dia menyediakan, libraries dan kumpulan kode yang bisa kamu gunakan untuk membangun website, tanpa perlu melakukan semuanya dari nol. Karena fiturnya yang sederhana, flask akan lebih ringan dan tidak tergantung dengan banyak library luar yang perlu diperhatikan. Secara umum flask menyediakan 'Werkzeug' yang berguna untuk menerima request (url) dan memberikan respon dan ada Jinja2, template engine yang digunakan untuk menampilkan data dan menulis logika pada tampilannya.

Pada saat ini, teknologi, terutama internet, telah berkembang sangat pesat. Perkembangan yang pesat ini memunculkan

peningkatan yang sangat signifikan dari jumlah pengguna internet. Dari perkembangan teknologi dan peningkatan jumlah pengguna internet, timbul kebutuhankebutuhan dan keahlian-keahlian tentang internet, salah satunya adalah web development. Web development adalah sebuah kegiatan pengembangan halaman-halaman web yang ada di internet. Halaman web adalah barang utama yang ada di internet, dimana seluruh pengguna internet pasti akan mengakses suatu web tertentu yang ada di internet. Oleh sebab itu, kebutuhan akan web di internet akan terus bertambah dan ini menimbulkan keahlian tentang web development pun semakin dibutuhkan. Dengan perkembangan teknologi saat ini, terdapat banyak alat untuk membantu kegiatan web development, salah satunya adalah web framework yang dibuat menggunakan bahasa Python, yaitu Flask. Flask adalah sebuah alat yang membantu membuat kerangka untuk sebuah web dan dengan menggunakan Flask, pengembang pemula pun dapat menciptakan sebuah web yang bagus. Pada paper ini akan dijelaskan gambaran umum Flask sebagai sebuah web microframework serta alasan-alasan Flask cocok sebagai alat untuk web development bagi developer pemula.

Web Microframework Flask A. Flask Flask adalah sebuah web framework yang ditulis dengan bahasa Python dan tergolong sebagai jenis microframework. Flask berfungsi sebagai kerangka kerja aplikasi dan tampilan dari suatu web. Dengan menggunakan Flask dan bahasa Python, pengembang dapat membuat sebuah web

yang terstruktur dan dapat mengatur behaviour suatu web dengan lebih mudah. Flask termasuk pada jenis microframework karena tidak memerlukan suatu alat atau pustaka tertentu dalam penggunaannya. Sebagian besar fungsi dan komponen umum seperti validasi form, database, dan sebagainya tidak terpasang secara default di Flask. Hal ini dikarenakan fungsi dan komponen-komponen tersebut sudah disediakan oleh pihak ketiga dan Flask dapat menggunakan ekstensi yang membuat fitur dan komponen-komponen tersebut seakan diimplementasikan oleh Flask sendiri. Selain itu, meskipun Flask disebut sebagai microframework, bukan berarti Flask mempunyai kekurangan dalam hal fungsionalitas. Microframework disini berarti bahwa Flask bermaksud untuk membuat core dari aplikasi ini sesederhana mungkin tapi tetap dapat dengan mudah ditambahkan. Dengan begitu, fleksibilitas serta skalabilitas dari Flask dapat dikatakan cukup tinggi dibandingkan dengan framework lainnya.

- Menginstall Flask

Web framework Flask ditulis menggunakan bahasa Python, sehingga sebelum Flask dapat digunakan, maka developer harus menginstall Python pada perangkat yang akan digunakan. Oleh sebab itu, web developer yang akan Penggunaan Python Web Framework Flask Untuk Pemula Rahadian Irsyad Laboratorium Telematika, Sekolah Teknik Elektro & Informatika, Institut Teknologi Bandung, e-mail: rahadian.irsyad@gmail.com menggunakan Flask sebagai web

framework untuk web development harus setidaknya mempelajari bahasa pemrograman Python terlebih dahulu, sebelum dapat menggunakan Flask seutuhnya. Dalam melakukan instalasi Flask pada sebuah perangkat, dibutuhkan PIP yang biasanya sudah terinstall pada Python versi 3.4 ke atas. PIP adalah sebuah package management system yang biasa digunakan untuk mengatur dan menginstall package yang berisi modul-modul Python. PIP digunakan untuk menginstall Flask karena Flask ditulis dan dikembangkan dengan bahasa dan modulmodul pemrograman Python. Dengan menggunakan PIP, semua hal yang dibutuhkan untuk instalasi Flask akan diunduh dan dipasang dalam satu perintah. Sebelum melakukan instalasi, pengguna Flask sebaiknya menggunakan virtual environment dalam proses pengembangan web. Virtual environment adalah sebuah sistem untuk mengisolasi sebuah proyek Python agar tidak terjadi konflik dengan proyek Python lainnya. Dengan virtual environment, sebuah proyek Python—atau Flask pada kasus ini, seakan ada pada satu komputer sendiri dengan dependensi sendiri dan tidak berhubungan dengan proyek Python lainnya. Untuk menginstall virtual environment masukkan perintah ini di shell.

\$ pip install virtualenv

Setelah menginstall virtual environment di komputer, selanjutnya masuk ke direktori yang akan dijadikan tempat proyek Python berada.

Setelah itu, buat dan aktifkan virtual environment disana dengan memasukkan perintah, sesuai dengan versi Python yang digunakan.

python 2:

\$ virtualenv virenv

python 3:

\$ python3 -m venv virenv

Perintah di atas adalah perintah untuk membuat sebuah virtual environment dengan nama virenv. Untuk mengaktifkan virtual environment yang telah dibuat di direktori tujuan, masukkan perintah ini.

\$ source virenv/bin/activate

Setelah mengaktifkan virtual environment, shell pada komputer akan berubah menjadi seperti ini.

(virenv)\$

Hal ini menandakan bahwa virtual environment telah aktif dan proyek Python yang akan dibuat telah terisolasi dan mempunyai sumber dependensi tersendiri. Setelah mengaktifkan virtual environment, Flask dapat diinstall tanpa perlu mengkhawatirkan adanya konflik dependensi dengan proyek Python lainnya. Masukkan perintah ini setelah masuk dan mengaktifkan virtual environment untuk menginstall Flask.

\$ pip install Flask

Untuk memeriksa keberhasilan instalasi Flask, buat sebuah file berisi kode Python berikut dan beri nama file tersebut 'hello.py'.
from flask
import

```
Flask app = Flask(__name__)
```

```
@app.route("/")
```

```
def hello():
```

```
    return "Hello World!"
```

Selanjutnya, jalankan file tersebut menggunakan Flask dengan memasukkan perintah berikut. \$ FLASK_APP=hello.py flask run
Setelah itu akan muncul pemberitahuan bahwa Flask telah berjalan di localhost. * Running on http://localhost:5000/

- **Menginstall Virtualenv**

Virtualenv adalah alat yang berguna yang akan membuat lingkungan pengembangan Python yang terisolasi di mana kamu bisa mengerjakan semua pengembangan yang diperlukan.

Kita akan menggunakan virtualenv untuk menginstall Flask. Virtualenv adalah alat yang berguna yang akan membuat lingkungan pengembangan Python yang terisolasi di mana kamu bisa melakukan semua pengembangan yang diperlukan. Misalkan kamu menemukan library Python baru yang ingin kamu coba. Jika

kamu install pada sistem keseluruhan, ada resiko mengganggu library lain yang sudah pernah kamu install. Jadi kita gunakan `virtualenv` untuk membuat sandbox, di mana kamu bisa menginstall dan menggunakan library tersebut tanpa mempengaruhi sisa dari sistem yang kamu miliki. Kamu bisa tetap menggunakan sandbox ini untuk pekerjaan selanjutnya, atau kamu bisa menghapusnya begitu kamu selesai menggunakannya. Yang manapun yang kamu pilih, sistem kamu akan tetap rapi dan bersih dari kotoran.

Kemungkinan sudah ada `virtualenv` pada sistem yang kamu miliki. Buka command line, dan coba perintah berikut:

```
1 $ virtualenv --version
```

Jika kamu melihat nomor versi, kamu sudah siap dan bisa melewati bagian "menginstall Flask". Jika perintah tidak digunakan, gunakan `easy_install` atau `pip` untuk menginstall `virtualenv`. Jika kamu menjalankan Linux atau Mac OS X, salah satu perintah berikut seharusnya bisa digunakan:

```
1 $ sudo easy_install virtualenv
```

atau:

```
1 $ sudo pip install virtualenv
```

atau:

1 \$ sudo apt-get install python-virtualenv

Jika kamu tidak memiliki perintah-perintah tersebut pada sistem, ada beberapa tutorial online yang akan menunjukkan bagaimana menginstall virtualenv di sistem. Menginstall Flask.

Setelah menginstall `virtualenv`, kamu bisa membuat lingkungan pengembangan yang terisolir dengan perintah berikut:

1 \$ virtualenv flaskapp

Di sini, `virtualenv` membuat folder, `flaskapp/`, dan mengatur salinan bersih dari Python di dalamnya untuk kamu gunakan. Virtualenv juga menginstall package manager `pip`.

Masuk ke lingkungan pengembangan baru yang baru dibuat dan aktifkan agar bisa kamu gunakan untuk bekerja.

1 \$ cd flaskapp

2 \$. bin/activate

Sekarang kamu bisa menginstall Flask dengan aman:

1 \$ pip install Flask

Fitur-fitur Flask Seperti yang telah dijelaskan di poin A, Flask merupakan microframework yang berarti Flask mempunyai core

yang sangat sederhana dan kecil, namun tetap dapat ditambah dan tumbuh. Berdasarkan hal tersebut, fitur-fitur bawaan dari Flask sendiri termasuk sedikit jumlahnya, diantaranya adalah

- Built-in development server.
- Debugger cepat.
- Integrated support untuk pengetesan unit.
- Kompatibel dengan mesin aplikasi Google.
- RESTful request dispatching.
- Jinja2 templating.
- Mendukung secure cookies.
- Berbasis unicode.
- Mengikuti WSGI 1.0

Selain itu, Flask juga didukung dengan dokumentasi yang sangat baik dan banyak forum yang ada di internet untuk mendiskusikan masalah terkait Flask.

- Struktur Proyek

Kita buat beberapa folder dan file di dalam *flaskapp/* untuk membuat aplikasi web kita tetap rapi. Cara kerja Flask Setelah mengenal Flask dan menginstall Flask pada poin-poin sebelumnya, poin ini akan menjelaskan bagaimana cara

kerja Flask secara umum. Cara kerja Flask dapat dijelaskan melalui kode file bernama 'hey.py' berikut.

1. `from flask import Flask`

2.

3. `app = Flask(__name__)`

4.

5. `@app.route('/')`

6. `def home():`

7. `return "Hey there!"`

8.

9. `if __name__ == '__main__':`

10. `app.run(debug=True)`

Pada baris ke-1, kode tersebut mengambil kode-kode dasar yang dibutuhkan dalam pembuatan proyek. flask disini

merupakan kerangka kerja dan Flask adalah tipe data class Python. Dengan kata lain, Flask adalah prototype yang digunakan untuk membuat contoh aplikasi web. Selain itu, dengan menggunakan kode yang serupa dengan kode di baris-1, Flask dapat menggunakan berbagai fitur yang disediakan oleh pihak ketiga, seperti validasi form dengan WTForms dan sebagainya. Caranya cukup dengan mengunduh package dan module dari pihak ketiga tersebut dan menuliskan kode seperti ini. `from flask_wtf import FlaskForm` `from wtforms import StringField` `from wtforms.validators import DataRequired` Pada baris ke-3, kode tersebut membuat sebuah instance dari class Flask dengan nama `app`. Variabel `__name__` adalah sebuah variabel khusus yang akan mengambil string dari `'__main__'`. Pada baris ke-5 hingga 7, kode tersebut mendefinisikan sebuah fungsi dari aplikasi web ini. Kode `@app.route('/')` memetakan ke URL home atau `(localhost:5000/)` dan kode `def home():` mendefinisikan sebuah fungsi bernama `home` yang akan dipanggil ketika halaman `home` atau `'/'` diakses. Sehingga ketika halaman `home` atau `'/'` diakses akan mengembalikan sebuah string “Hey there!” pada halaman tersebut. Baris ke-5 hingga 7 adalah poin utama dari Flask. Ketika rute masukan diganti menjadi halaman lain, contohnya `'/profile'`, maka fungsi yang dijalankan adalah fungsi yang sudah disiapkan ketika web diakses dengan alamat `'localhost:5000/profile'`. Seperti pada kode contoh berikut `from`

flask import Flask app = Flask(__name__) @app.route('/') def home(): return "Hey there!" @app.route('/profile') def profile(): return profile_one Selanjutnya, pada baris ke-9 dan 10, ini adalah sebuah teknik yang digunakan para developer yang menggunakan bahasa pemrograman Python. Pada Python, script yang sedang dijalankan akan diberikan nama '__main__' dan jika script diambil dari script lain, maka script tersebut tetap menggunakan nama aslinya, contohnya 'hello.py'. Pada kasus ini, Flask menjalankan script ini, sehingga variabel __name__ mempunyai nilai '__main__'. Dengan begitu, kode pada baris ke-9 terpenuhi dan aplikasi web akan dijalankan dengan kode app.run(debug=True). Parameter pada method app.run(debug=True) berarti Python akan mencetak kesalahan-kesalahan yang mungkin terjadi pada halaman web tersebut.

01 .

02 .

03 |— app

04 | |— static

05 | | |— css

06 | | |— img

07			└─js
08			└─templates
09			└─routes.py
10			└─README.md

Di dalam *flaskapp/*, buat sebuah folder *app/* untuk menyimpan semua file-mu. Di dalam *app/*, buat folder *static/*, di sinilah kita akan simpan gambar, CSS, dan file JavaScript aplikasi web kita. Jadi buat folder untuk masing-masing jenis tersebut, seperti yang dicontohkan di atas. Begitu pula, buat folder lain bernama *templates/* untuk menyimpan template untuk aplikasi web ini.

Keunggulan Flask Berdasarkan penjelasan tentang apa itu Flask, cara menginstall Flask, fitur-fitur yang ditawarkan oleh Flask, dan cara kerja Flask, Flask memiliki keunggulan serta kekurangan dibandingkan dengan web framework berbasis Python lainnya, seperti Django, CherryPy, dan sebagainya. Keunggulan-keunggulan dari Flask adalah

- Ringan untuk dijalankan karena mempunyai core yang sederhana dan desain modular.
- Dapat menangani fungsi HTTP request dengan mudah.
- API yang baik dan koheren.

-
- Dokumentasi yang banyak dan terstruktur dengan baik, penuh dengan contoh yang dapat digunakan langsung.
 - Mudah untuk dipasang dan di-deploy untuk produksi.
 - Mudah untuk diperiksa secara menyeluruh. (Unit testability.)
 - Fleksibilitas tinggi, dengan konfigurasi yang sangat mudah diubah. Selain memiliki keunggulan-keunggulan yang telah disebutkan, Flask juga memiliki kekurangan, diantaranya adalah.
 - Tidak memiliki ORM dan database layer bawaan, sehingga harus menggunakan aplikasi dari pihak ketiga.
 - Tidak async-friendly. Karakteristik Flask yang sangat membutuhkan aplikasi pihak ketiga dalam menjalankan fitur-fiturnya menjadi pedang bermata dua yang bisa menjadi keunggulan maupun kekurangan dari Flask itu sendiri.

3.7. *HTML*

Hypertext Markup Language atau **HTML** **adalah** bahasa markup yang digunakan untuk membuat struktur halaman website. *HTML* atau *Hypertext Markup Language* adalah bahasa markup yang digunakan untuk membuat struktur halaman website agar dapat ditampilkan pada *web browser*. Jadi, *HTML* dapat dianalogikan sebagai pondasi awal dalam menyusun kerangka halaman web secara terstruktur sebelum membahas terkait tampilan desain dan sisi fungsionalitas.

Sejarah HTML pertama kali dibuat oleh Tim Berners-Lee yang merupakan fisikawan di lembaga penelitian CERN, Swiss. Berners-Lee mempunyai ide atau pemikiran mengenai sistem hypertext berbasis internet. Sekitar tahun 1991, Tim merilis versi HTML pertama yang di dalamnya terdapat sekitar 18 tag.

Untuk sekarang, HTML mengalami beberapa pengembangan dari sisi fitur serta informasi yang disajikan hingga versi terbaru, yaitu HTML5. Setiap tahunnya, bahasa ini mempunyai tingkat popularitas yang cukup tinggi sehingga dijadikan sebagai web standard resmi di dunia pemrograman.

Selanjutnya, HTML juga mengalami tahap perbaikan (*maintain*) dan dikembangkan oleh *World Wide Web Consortium* (W3C). Pada tahun 2014, merupakan waktu dimana HTML mengalami *upgrade* yang cukup signifikan dan terdapat fitur *semantic* untuk memudahkan pengembang website dalam menyusun kode dan memberitahukan makna dari konten tersebut, sebagai contoh **<article>**, **<footer>**, dan **<header>**.

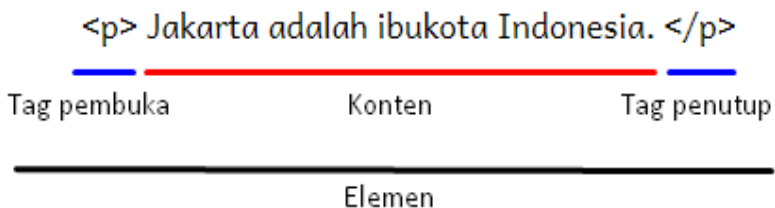
Setelah mengenal sejarah singkat dari bahasa markup HTML, selanjutnya kita juga harus mengetahui kegunaannya. Berikut ini beberapa fungsi dan tujuan dari penerapan HTML untuk pembuatan situs atau *web app*.

Fungsi utama penggunaan HTML sendiri adalah membangun tampilan website yang telah menerapkan metode

semantik untuk memudahkan setiap pengembang dalam proses *development* dan *maintenance*. Kemudian, HTML juga dapat dikolaborasikan dengan penggunaan bahasa CSS (*Cascade Style Sheet*) serta JavaScript.

Dimana, peran dari HTML5 berfungsi untuk menyusun kerangka dan struktur halaman website. Kemudian, CSS membantu dalam memberikan tampilan desain meliputi warna, font, outline, dan lain sebagainya. Dan Tugas dari bahasa pemrograman JavaScript adalah memberikan sentuhan interaksi untuk memberikan pengalaman yang berbeda kepada *user*.

HTML terdiri dari kombinasi teks dan simbol yang disimpan dalam sebuah file. Dalam membuat file HTML, terdapat standar atau format khusus yang harus diikuti. Format tersebut telah tertuang dalam standar kode internasional atau **ASCII (American Standard Code for Information Interchange)**.



Dengan adanya HTML, pengguna dapat membuat atau menyusun heading, paragraf, gambar, link, dan lainnya supaya dapat dilihat banyak orang melalui halaman website.

Untuk bisa diakses secara umum, pengguna perlu membukanya lewat aplikasi browser, seperti Internet Explorer, Chrome, atau Mozilla Firefox.

Setelah mengetahui pengertian HTML di atas, Anda mungkin penasaran bagaimana sejarah lahirnya HTML hingga saat ini? Untuk itu, mari simak terlebih dulu sejarah HTML berikut.

- Sejarah HTML

Tim Berners-Lee merupakan sosok ilmuwan dibalik lahirnya HTML. Di tahun 1991, ia awalnya menciptakan HTML sebagai solusi untuk memudahkan para ilmuwan dalam mengakses dokumen satu sama lain.

WWW. Keduanya saling bersinergi supaya informasi yang ingin disebarakan bisa diakses oleh banyak orang.

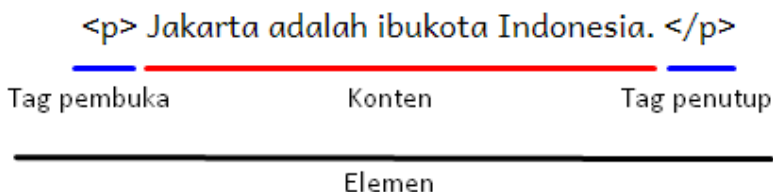
HTML telah berevolusi menjadi bahasa markup yang telah digunakan oleh lebih dari 92,3% persen website di dunia.

Selama perjalanannya, HTML tentu mengalami perkembangan yang cukup pesat. Seperti apa evolusinya? Simak timeline perjalanan HTML berikut:

Sampai saat ini HTML5 menjadi versi yang paling update dengan sistem yang paling canggih. Popularitas HTML versi terbaru ini pun sangat pesat. Bahkan, saat ini ada 87,4 persen website yang sudah menggunakan HTML5.

- Tag HTML, Elemen, dan Atribut

Setelah mengetahui bagaimana sejarah dari HTML, sekarang Anda perlu tahu komponen-komponen yang terdapat pada HTML. Secara umum HTML terdiri dari Tag, Elemen, dan Atribut.



3.7.1. TAG

Tag adalah tanda awalan dan akhiran dalam perintah HTML yang akan dibaca oleh web browser. Tag dibuat dengan menggunakan kurung siku `<...>`, di mana di dalamnya berisi nama tag.

Setiap tag memiliki fungsi perintah yang berbeda-beda. Mulai dari membuat judul, paragraf, heading, cetak tebal, miring, italic, dan lainnya.

Misalnya: **`<bold>`**. Tag tersebut akan meminta browser untuk menampilkan teks dengan format tebal.

Selain itu, tag ditulis secara berpasangan, yaitu **tag pembuka** dan **tag penutup**. Pada tag penutup ditambahkan garis miring (/) di depan nama tag.

Tag HTML sendiri awalnya hanya berjumlah 18. Hingga saat ini sudah ada lebih dari 250 tag.

2.9.2 ELEMEN

Elemen merupakan komponen HTML yang berupa keseluruhan kode dari tag pembuka hingga tag penutup. Elemen

terdiri dari teks dan simbol yang berupa tag pembuka, isi tag atau konten, dan tag penutup.

Contohnya: `<bold>Belajar HTML</bold>`.

Pada sebuah elemen juga bisa berisi elemen lainnya. Jadi elemen tersebut tak hanya berisi satu tag saja, tetapi banyak tag. Elemen itu disebut **nested elements**. Sebagai contoh, simak kode berikut:

```
<html>
<body>
<h2>Judul Heading</h2>
<p>Paragraf pertama artikel.</p>
</body>
</html>
```

Jika diperhatikan elemen HTML tersebut, dari tag pembuka `<html>` lalu tag `<body>` yang berisi elemen tag heading dan elemen tag paragraf.

2.9.3 ATRIBUT

Atribut adalah informasi atau perintah tambahan yang berada dalam elemen. Atribut ini berfungsi sebagai penjelas perintah tag pada elemen.

Misalnya, ``.

Tag `` memiliki atribut khusus yaitu (`src`) dan (`alt`) yang artinya browser harus menampilkan `gambar.jpg` dengan `alt` text "Bunga Matahari".

Selain atribut khusus, ada beberapa atribut yang digunakan untuk menentukan aksi pada dilakukan jika terjadi sesuatu pada elemen. Misalnya , (**onload**) saat loading selesai, (**onoffline**) saat tiba-tiba offline, (**onresize**) ukuran jendela. Atribut tersebut disebut dengan **atribut event**.

2.9.4 FUNGSI HTML

1. Membuat Halaman Website

Fungsi utama dari HTML adalah untuk membuat halaman website. Sehingga Anda dapat berbagi informasi pada banyak orang melalui browser di internet.

HTML juga mudah untuk diaplikasikan untuk seorang pemula, tak heran, hampir semua website pasti dibuat menggunakan HTML.

2. Menjadi Pondasi Website

HTML juga berfungsi sebagai pondasi utama pembuatan website. Beberapa bahasa pemrograman akan diaplikasikan pada HTML untuk dapat menampilkan website yang lebih menarik.

Sebagai contoh, tampilan website dapat lebih berwarna dengan menggunakan CSS. Lalu website menampilkan banyak fitur dengan menggunakan bahasa pemrograman Javascript, PHP, atau lainnya.

3. Menandai Teks dan Bagian pada Halaman Website

Fungsi lainnya dari HTML ialah menjadi penanda teks dan bagian dari laman website. Semisal untuk menandai kalimat cetak tebal dan miring hanya perlu diberi kode <bold>, <italic>.

Selain itu, HTML juga berfungsi untuk menandai bagian pada website. Misalnya bagian header, main, footer, dan navigation. Sehingga lebih memudahkan pembuatan website.

4. Menampilkan Multimedia di Website

Supaya website tidak membosankan, Anda tentu perlu menambahkan gambar, video, tabel, dan media lainnya, bukan? Nah dengan HTML, Anda bisa menampilkan berbagai media sesuai kebutuhan.

5. Mengarahkan Pengguna ke Halaman Website Lain

Seiring perkembangannya, HTML kini juga berfungsi untuk mengarahkan pengguna ke halaman atau website lain melalui teks tertentu. Fungsi tersebut juga bisa disebut **hyperlink**. Misalnya seperti contoh gambar berikut.

3.8. CSS

CSS adalah kepanjangan dari *Cascading Style Sheets* yang berguna untuk menyederhanakan proses pembuatan website dengan mengatur elemen yang tertulis di bahasa markup.

CSS dipakai untuk mendesain halaman depan atau tampilan website (*front end*). CSS menangani tampilan dan ‘rasa’ dari halaman website.

Ada banyak hal yang dapat Anda lakukan menggunakan CSS dibandingkan dengan bahasa pemrograman inti seperti HTML dan PHP. Ketika menggunakan CSS, Anda dapat mengatur warna teks, jenis font, baris antar paragraf, ukuran kolom, dan jenis *background* yang dipakai.

CSS sangat mudah dipelajari, tapi juga *powerful* karena dapat mengontrol penyajian tampilan dari dokumen HTML. Mulai dari yang simpel sampai kompleks. Tidak heran jika saat ini CSS hampir dipakai di berbagai website untuk dikombinasikan dengan HTML maupun PHP.

3.8.1. Cara Kerja CSS

Setelah mengetahui pengertian CSS dan fungsinya, pertanyaan selanjutnya ialah bagaimana cara kerja CSS? Jawabannya, CSS membuat style di dalam halaman web dapat berinteraksi dengan elemen HTML. Elemen merupakan komponen HTML dari sebuah halaman web. Berikut adalah salah satu contoh elemen di HTML:

```
<p>This is my paragraph!</p>
```

Jika Anda ingin membuat paragraf di atas muncul dengan warna yang lain dan bold, Anda bisa menambahkan baris CSS seperti di bawah ini:

```
p { color:pink; font-weight:bold; }
```

Pada contoh di atas, “**p**” (paragraf) merupakan bagian baris kode CSS untuk menspesifikasikan elemen HTML atau biasa dikenal dengan nama ‘**selector**’.

Di dalam CSS, selector ditulis di bagian kiri sebelum tanda kurung kurawal (*curly bracket*) pertama. Sedangkan informasi yang ada di dalam tanda kurung kurawal merupakan sebuah deklarasi yang di dalamnya mengandung properti dan nilai yang diaplikasikan ke dalam selector.

Contoh dari properti adalah ukuran font, warna, layout, format, dan lain sebagainya. Sedangkan nilai yang dimaksud adalah pengaturan yang diterapkan ke dalam properti. Misalnya dari contoh di atas, “**color**” dan “**font-weight**” merupakan properti, sedangkan “**pink**” dan “**bold**” merupakan sebuah nilai.

Contoh dari properti adalah ukuran font, warna, layout, format, dan lain sebagainya. Sedangkan nilai yang dimaksud adalah pengaturan yang diterapkan ke dalam properti. Misalnya dari contoh di atas, “**color**” dan “**font-weight**” merupakan properti, sedangkan “**pink**” dan “**bold**” merupakan sebuah nilai.

Jadi seluruh baris kode di dalam tanda kurung kurawal { color: pink; font-weight: bold; } merupakan sebuah deklarasi, sedangkan “p” (HTML paragraf) merupakan sebuah selector. Baik deklarasi dan selector merupakan prinsip dasar dari penggunaan CSS dan HTML. Prinsip ini juga dapat digunakan untuk mengatur ukuran font, warna background, format teks, dan lain sebagainya.

3.8.2. JENIS CSS

Inline CSS

Inline CSS adalah penempatan CSS yang langsung dilakukan di dalam konten HTML. Penempatan CSS secara inline

hanya akan mempengaruhi satu bagian baris kode. Berikut adalah contohnya.

```
<h1 style="font-size:30px;color:blue;">Cek berita utama ini!</h1>
```

CSS di atas berarti hanya akan mempengaruhi paragraf yang spesifik di dalam satu file .html dan muncul dengan warna biru (*blue*) ukuran font 30px.

3.8.1. EXTERNAL CSS

External CSS adalah penempatan baris kode CSS menggunakan file .css. Cara ini dapat digunakan untuk mendeklarasikan sebagian besar pengaturan tampilan website secara keseluruhan.

Cara ini lebih sederhana dan simpel daripada menambahkan baris kode di setiap elemen HTML yang ingin Anda atur tampilannya.

Jika menggunakan penempatan secara eksternal, Anda harus menambahkan header untuk memanggil file .css di dalam file HTML seperti contoh di bawah ini.

```
<head>
<link                rel="stylesheet"                type="text/css"
</head>
```

Baris kode di atas akan menghubungkan file .html dengan eksternal style sheet (CSS). Pada contoh di atas file CSS menggunakan nama 'fileCSSAnda.css' yang dapat Anda sesuaikan dengan letak file CSS di dalam direktori website. Ketika file HTML sudah terhubung dengan file CSS maka seluruh pengaturan CSS akan diterapkan di dalam kode HTML.

3.9. API

API adalah singkatan dari *Application Programming Interface* atau dapat diartikan sebagai perangkat lunak yang berfungsi menghubungkan suatu aplikasi dengan aplikasi lainnya. Dengan API, dua aplikasi yang awalnya sama sekali tidak berhubungan dapat terkoneksi dan terintegrasi sehingga mampu mengkomunikasikan fitur-fitur di antara keduanya.

Jika diibaratkan secara sederhana, API adalah jembatan antara dua daratan yang terpisah. Jembatan ini dibangun dengan sekumpulan kode yang menyediakan berbagai *access point* sehingga informasi dari aplikasi A dapat menyeberang ke aplikasi B secara mulus.

Dalam pengembangan aplikasi, API berperan penting karena memungkinkan suatu aplikasi dapat mengakses data, fitur, layanan, atau sistem operasi yang digunakan oleh aplikasi yang terhubung. Dengan begitu, API dapat menghemat waktu, tenaga, dan uang yang digunakan dalam pengembangan aplikasi. Dapat disimpulkan bahwa API merupakan ‘jalur tikus’ yang memudahkan penggunaannya untuk mencapai tujuan.

3.9.1. MANFAAT API

Penggunaan API dalam pengembangan aplikasi memiliki berbagai manfaat, di antaranya adalah:

1. Mengembangkan aplikasi lebih cepat dan efektif

Dari sisi pengembangan aplikasi, integrasi dua aplikasi dengan API jelas dapat membantu komunikasi tanpa perlu repot-repot menghubungkan keduanya secara manual. API juga akan menginformasikan pembaruan (*update*) fitur sehingga membuat pekerjaan menjadi lebih cepat dan efektif.

2. Menciptakan aplikasi yang fungsional

Tidak ada lagi memasukkan informasi secara manual dan melelahkan; semua teratasi apabila Anda memakai API. Berfokus

pada fungsi, API dapat menampilkan fitur yang sama dengan aplikasi tujuan.

3. Meringankan beban kerja server

Dengan menggunakan API, maka server tidak perlu lagi mencari dan menyimpan data yang sangat banyak jumlahnya. Jika sewaktu-waktu ada data yang diperlukan, maka cukup dengan meminta API untuk mengambilnya dari server asal.

3.9.2. JENIS API

1. Public API

Public API merupakan jenis API terbuka yang tersedia untuk digunakan oleh pengembang atau bisnis eksternal. Perusahaan yang mengembangkan strategi bisnis dengan pelibatan berbagai aplikasi dan data dengan bisnis lain akan mengembangkan dan menawarkan API kepada publik. API jenis ini umumnya melibatkan otentikasi dan otorisasi moderat. Perusahaan juga dapat berupaya memonetisasi API dengan mengenakan biaya per panggilan akses yang dipakai publik untuk memanfaatkan API.

2. Private API

Private API atau seringkali juga disebut sebagai API internal merupakan API yang digunakan di dalam suatu perusahaan secara

eksklusif untuk menghubungkan sistem dan data dalam bisnis. API internal ini misalnya dipakai untuk menghubungkan sistem penggajian dan pengorganisasian sumber daya manusia.

Private API secara khusus dapat menghadirkan keamanan dan otentikasi yang lemah atau bahkan tidak ada keamanan sama sekali. Hal ini disebabkan API internal ini ditujukan hanya untuk penggunaan internal dan asumsi kebijakan keamanannya lebih lemah. Namun belakangan, API internal pun kini memiliki beberapa tuntutan terkait strategi perusahaan yang membuatnya ditambahkan dengan kebijakan keamanan lebih.

3.9.3. Arsitektur API

REST

REST atau Representational State Transfer merupakan arsitektur API yang dapat dengan cepat menjadi model desain yang disukai untuk publik. Sebabnya, arsitektur API jenis ini bergantung pada protokol komunikasi stateless, dan juga HTTP (hypertext transfer protocol). Secara umum REST menyusun data tidak hanya dalam XML, tetapi juga YAML (yet another markup language) atau format lain yang dapat dibaca mesin seperti JSON (JavaScript Object Notation). REST mengikuti paradigma pemrograman berorientasi objek dari kata benda dan kata kerja. Dibandingkan dengan SOAP, REST amat bergantung pada fungsi atau pengodean tertentu.

3.9.4. PENGGUNAAN API

API telah menjadi aspek berharga bagi bisnis modern karena memungkinkan perusahaan untuk membuka akses ke sumber daya mereka sambil menjaga keamanan dan kontrol. Berikut adalah beberapa contoh populer API yang bisa dengan mudah kamu temui.

1. Login universal

Contoh API yang paling populer dan bisa kamu temui setiap saat adalah kemudahan untuk masuk ke situs menggunakan detail profil Facebook, Twitter, atau Google. Fitur ini sangat praktis karena membuat situs mana pun memanfaatkan API untuk mengautentikasi pengguna dengan cepat. Fitur ini juga menghemat waktu karena kita tidak perlu repot mengisi profil baru untuk setiap layanan situs atau keanggotaan baru.

2. Proses pembayaran third-party

Contoh penggunaan API lainnya adalah dalam proses pembayaran ke pihak ketiga, misalnya PayPal. Sekarang kita bisa leluasa membayar produk secara online tanpa harus mengungkapkan data pribadi atau memberikan akses kepada individu yang tidak berwenang.

3. Membandingkan pemesanan travel

Melalui API, pengguna kini bisa membandingkan opsi termurah untuk setiap tanggal dan tujuan di berbagai layanan travel booking.

4. Google Maps

Salah satu contoh paling umum dari API adalah Google Maps. Selain API menampilkan peta statis atau interaktif, aplikasi menggunakan API dan fitur lain untuk memberikan petunjuk arah atau tempat menarik kepada pengguna.

5. Twitter

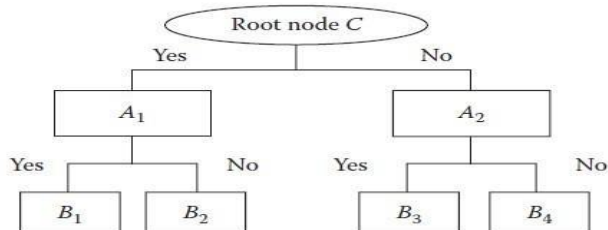
Setiap Tweet berisi berbagai macam info, mulai dari penulis, ID unik, pesan, waktu posting, hingga tempat.

Bab 3

Metodologi Penelitian

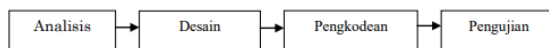
3.1 Alur Algoritma C.45

Pada akhir tahun 1970 sampai awal tahun 1980 J. Ross Quinlan, seorang peneliti di bidang machine learning, membuat sebuah algoritma decision tree yang dikenal dengan ID3 (Iterative Dichotomiser). Quinlan kemudian membuat algoritma C4.5 (sering disebut dengan pohon keputusan) yang merupakan pengembangan dari algoritma ID3. Algoritma ini memiliki kelebihan, yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan.



Gambar - 9 Alur Algoritma C.45

Algoritma C4.5 merupakan bagian metode klasifikasi pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama J. Rose Quinlan pada akhir tahun 1970-an. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun.



Gambar - 10 Kerangka Penelitian

Berikut langkah-langkah menjalankan algoritma C4.5 untuk membangun sebuah pohon keputusan adalah sebagai berikut:

1. Hitung jumlah data, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu. Untuk proses pertama syaratnya masih kosong.
2. Pilih atribut sebagai *Node*.
3. Bentuk cabang pada setiap anggota dari *Node*.

Node adalah atribut yang mempunyai nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai *gain* suatu atribut digunakan rumus seperti yang tertera dalam persamaan berikut :

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

Sementara itu, untuk menghitung nilai *Entropy* dapat dilihat pada persamaan berikut [2] :

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

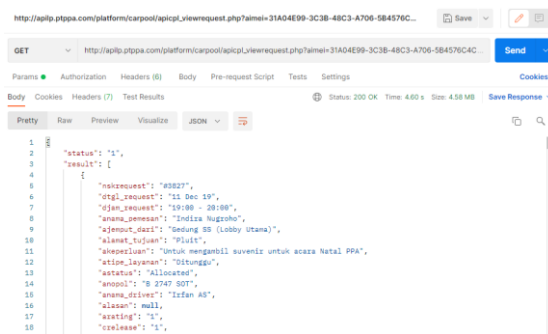
Contoh data yang diambil dari database :

no	Order Request	Driver Confirmation	Status	Poin
1	Acc	True	Allocated	No
2	Acc	True	Allocated	No
3	Tidak Acc	False	Allocated	No
5	Acc	False	Allocated	No
6	Tidak Acc	False	Allocated	No
7	Tidak Acc	False	Allocated	No
8	Acc	False	Done	Yes
9	Acc	True	Done	Yes
10	Tidak Acc	False	Allocated	No
11	Acc	True	Done	Yes

12	Tidak Acc	False	Allocated	No
13	Tidak Acc	False	Allocated	No
14	Acc	True	Done	Yes

3.2 Mengumpulkan Data API

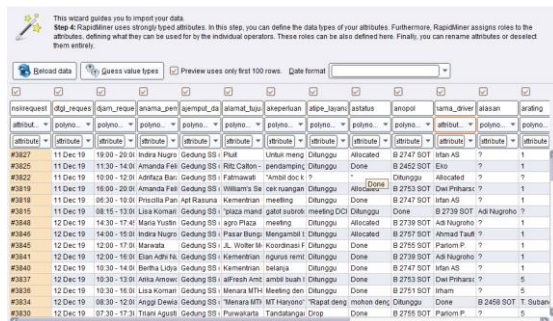
Mengumpulkan data spasial dan non spasial dengan cara menarik data dari API yang ada pada aplikasi Carpool dan memilih parameter. Berikut merupakan data API yang berhasil dipanggil melalui Postman



Gambar 11 Json Database

3.3 Menyusun Paramater Database

Database merupakan kumpulan file-file yang saling berkaitan dan berinteraksi, relasi tersebut bila ditunjukkan dengan kunci dari tiap-tiap file yang ada. Penyusunan parameter database atau struktur keputusan dilakukan untuk menentukan parameter atau alternatif keputusan yang teridentifikasi. Tabel dengan parameter yang sudah ditentukan dari setiap tabel dapat dilihat pada gambar



This wizard guides you to import your data. Step 4: Rapidminer uses strongly typed attributes. In this step, you can define the data types of your attributes. Furthermore, Rapidminer assigns roles to the attributes, defining what they can be used for by the individual operators. These roles can be also defined here. Finally, you can rename attributes or disable them entirely.

☒ Specify data ☒ Guess value types ☒ Preview uses only first 100 rows.

	idg_requests	idam_request	idama_pen	idampul_da	idamal_hajr	idapeperuan	idape_jasen	idastatus	idaport	idama_driver	idamasan	idastring
attribut	polymio	polymio	polymio	polymio	polymio	polymio	polymio	polymio	polymio	attribut	polymio	polymio
attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut
#3827	11 Dec 19	10:00 - 20:00	Indira Nugro	Gedung SS - P	Full	Unsur yang	Ditunggu	Allocated	B 2747 SOT	Idan AS	?	1
#3825	11 Dec 19	11:30 - 14:00	Amanda Feli	Gedung SS - Ritz Carlton	pendamping	Ditunggu	Done	B 2462 SOT	Elio	?	?	1
#3822	11 Dec 19	10:00 - 12:00	Aditiza Bari	Gedung SS - Falmawati	"Amil doc k	?	Ditunggu	Allocated	?	?	?	?
#3819	11 Dec 19	15:00 - 20:00	Amanda Feli	Gedung SS - Williams St	cek ruangan	Ditunggu	Allocated	B 2753 SOT	Ovi Prabasi	?	?	1
#3818	11 Dec 19	09:30 - 10:00	Priscilla Pan	api Rasmu	Kementerian	meeting	Ditunggu	Done	B 2747 SOT	Idan AS	?	1
#3815	11 Dec 19	08:15 - 13:00	Lisa Komari	Gedung SS - "gila mand	gatil subroki	meeting DCO	Ditunggu	Done	B 2739 SOT	Adi Nugroho	?	?
#3848	12 Dec 19	14:30 - 17:40	Maria Yustin	Gedung SS - agro Plaza	meeting	Ditunggu	Allocated	B 2739 SOT	Adi Nugroho	?	?	1
#3846	12 Dec 19	14:00 - 16:00	Indira Nugro	Gedung SS - Pesar Bungi	Mengamati	Ditunggu	Allocated	B 2757 SOT	Almad Tauli	?	?	1
#3845	12 Dec 19	12:00 - 17:00	Marwata	Gedung SS - J. Wolferli	Koordinasi F	Ditunggu	Done	B 2755 SOT	Parlom P	?	?	1
#3841	12 Dec 19	12:00 - 16:00	Elan Adhila	Gedung SS - Kementerian	ngurus remi	Ditunggu	Done	B 2739 SOT	Adi Nugroho	?	?	1
#3840	12 Dec 19	10:30 - 14:00	Bertha Liba	Gedung SS - Kementerian	belanja	Ditunggu	Done	B 2747 SOT	Idan AS	?	?	1
#3837	12 Dec 19	10:30 - 13:00	Akisa Kerner	Gedung SS - alifresh keni	arnel suah	Ditunggu	Done	B 2753 SOT	Ovi Prabasi	?	?	5
#3836	12 Dec 19	10:30 - 16:00	Lisa Komari	Gedung SS - Menara MTH	Meeting den	Ditunggu	Done	B 2751 SOT	Idham	?	?	5
#3834	12 Dec 19	08:30 - 12:00	Anggi Dewita	Gedung SS - "Menara MTH	MT Haryono	"Rapat deng	Ditunggu	Done	B 2468 SOT	T. Suban	?	?
#3830	12 Dec 19	07:30 - 17:30	Tisno Agusti	Gedung SS - Pura Wara	Tandatangan	Drop	Done	B 2750 SOT	Parlom P	?	?	5

Gambar 12 Paramater Database

- **Penentuan Parameter**

Untuk parameter yang digunakan dalam perhitungan berikut. Menggunakan mengambil parameter dengan label a_status dan a_driver.

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no"?>
2 <process version="5.3.000">
3   <context>
4     <input/>
5     <output/>
6     <macros/>
7   </context>
8   <operator activated="true" class="process" compatibility="5.3.000" expanded="true" name="Process">
9     <process expanded="true" height="280" width="547">
10      <operator activated="true" class="retrieve" compatibility="5.3.000" expanded="true" height="68" name="Retrieve carp">
11        <parameter key="repository_entry" value="//local Repository/carpool507"/>
12      </operator>
13      <operator activated="true" class="set_rule" compatibility="5.3.000" expanded="true" height="76" name="Set Role" wdd
14        <parameter key="name" value="anasa_driver"/>
15        <parameter key="target role" value="label"/>
16      </operator>
17    </process>
18  </operator>
19 </process>
20 </process>
21 </process>
22 </process>
23 </process>
24 </process>
25 </process>
26 </process>
27 </process>
28 </process>
29 </process>
30 </process>
31 </process>
32 </process>
33 </process>
34 </process>
35 </process>
36 </process>
37 </process>
38 </process>
39 </process>
40 </process>
41 </process>
42 </process>
43 </process>
44 </process>
45 </process>
46 </process>
47 </process>
48 </process>
49 </process>
50 </process>
51 </process>
52 </process>
53 </process>
54 </process>
55 </process>
56 </process>
57 </process>
58 </process>
59 </process>
60 </process>
61 </process>
62 </process>
63 </process>
64 </process>
65 </process>
66 </process>
67 </process>
68 </process>
69 </process>
70 </process>
71 </process>
72 </process>
73 </process>
74 </process>
75 </process>
76 </process>
77 </process>
78 </process>
79 </process>
80 </process>
81 </process>
82 </process>
83 </process>
84 </process>
85 </process>
86 </process>
87 </process>
88 </process>
89 </process>
90 </process>
91 </process>
92 </process>
93 </process>
94 </process>
95 </process>
96 </process>
97 </process>
98 </process>
99 </process>
100 </process>

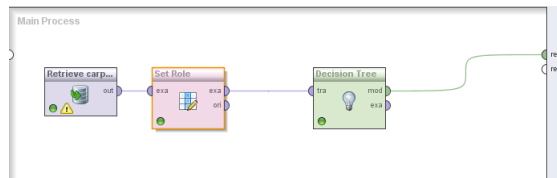
```

Gambar 12 Penentuan Parameter

3.4 Pengimplementasian Decision Tree

C.45

Semua elemen dikelompokkan akan diproses dengan penentuan role pada atribut yang ada dalam tabel. Maka akan tersusun decision tree sebagai berikut



Gambar 13 Design Decision Tree

3.5 Crisp DM

Cross-Industry Standard Process for Data Mining atau CRISP-DM adalah salah satu model proses datamining (*datamining framework*) yang awalnya (1996) dibangun oleh 5 perusahaan yaitu Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA. Framework ini kemudian dikembangkan oleh ratusan organisasi dan perusahaan di Eropa untuk dijadikan *methodology standard non-proprietary* bagi *data mining*. Versi pertama dari methodology ini dipresentasikan pada

4th CRISP-DM SIG Workshop di Brussels pada bulan Maret 1999 (Pete Chapman, 1999); dan langkah langkah proses datamining berdasarkan model ini di publikasikan pada tahun berikutnya (Pete Chapman,2000). Antara tahun 2006 dan 2008 terbentuklah grup CRISP-DM 2.0 SIG yang berkeinginan untuk mengupdate CRISP-DM process model (Colin Shearer, 2006). Namun produk akhir dari inisiatip ini tidak diketahui.

Banyak hasil penelitian yang mengungkapkan bahwa CRISP-DM adalah datamining model yang masih digunakan secara luas di kalangan industry, sebahagian dikarenakan keunggulannya dalam menyelesaikan banyak persoalan dalam proyek proyek data mining.

Mariscal, Marba dan Fernandez (Mariscal, Marban, and Fernandez 2010) menyatakan CRISP-DM sebagai *defacto* menjadi standar untuk pengembangan proyek *data mining* dan *knowledge discovery* karena paling banyak digunakan dalam pengembangan data mining. Hal tersebut dapat terlihat dari survei yang ditunjukkan pada Gambar 1 yang dilakukan terhadap penggunaan metodologi dalam proyek *data mining*.

Hasil survei “Penggunaan Metodologi dalam Proyek *Data Mining*”, memperlihatkan pengguna CRISP-DM di tahun 2002 mencapai 51%, kemudian menurun menuju 41% di tahun 2004. Meskipun persentasi penggunaan CRISP-DM

menurun 10%, jumlah pengguna metodologi ini masih terbilang lebih banyak daripada pengguna metodologi lain.

Model proses CRISP-DM memberikan gambaran tentang siklus hidup proyek data mining. CRISP-DM memiliki 6 tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* seperti ditunjukkan pada (Chapman et al. 2000).

Masing-masing tahapan tersebut dijelaskan sebagai berikut :

- ***Business Understanding***

Ini adalah tahap pertama dalam CRISP-DM dan termasuk bagian yang cukup vital. Pada tahap ini membutuhkan pengetahuan dari objek bisnis, bagaimana membangun atau mendapatkan data, dan bagaimana untuk mencocokkan tujuan pemodelan untuk tujuan bisnis sehingga model terbaik dapat dibangun. Kegiatan yang dilakukan antara lain: menentukan tujuan dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah data mining, dan selanjutnya mempersiapkan strategi awal untuk mencapai tujuan tersebut.

- ***Data Understanding***

Secara garis besar untuk memeriksa data, sehingga dapat mengidentifikasi masalah dalam data. Tahap ini memberikan

fondasi analitik untuk sebuah penelitian dengan membuat ringkasaan (*summary*) dan mengidentifikasi potensi masalah dalam data. Tahap ini juga harus dilakukan secara cermat dan tidak terburu-buru, seperti pada visualisasi data, yang terkadang *insight*-nya sangat sulit didapat jika dihubungkan dengan *summary data* nya. Jika ada masalah pada tahap ini yang belum terjawab, maka akan mengganggu pada tahap *modeling*.

Ringkasan atau *summary* dari data dapat berguna untuk mengkonfirmasi apakah data terdistribusi seperti yang diharapkan, atau mengungkapkan penyimpangan tak terduga yang perlu ditangani pada tahap selanjutnya, yaitu *Data Preperation*.

Masalah dalam data biasanya seperti nilai-nilai yang hilang, *outlier*, berdistribusi *spike*, berdistribusi bimodal harus diidentifikasi dan diukur sehingga dapat diperbaiki dalam *Data Preperation*.

- ***Data Preparation***

Secara garis besar untuk memperbaiki masalah dalam data, kemudian membuat *variabel derived*. Tahap ini jelas membutuhkan pemikiran yang cukup matang dan usaha yang cukup tinggi untuk memastikan data tepat untuk algoritma yang digunakan.

Bukan berarti saat *Data Preperation* pertama kali dimana masalah-masalah pada data sudah diselesaikan, data

sudah dapat digunakan hingga tahap terakhir. Tahap ini merupakan tahap yang sering ditinjau kembali saat menemukan masalah pada saat pembangunan model. Sehingga dilakukan iterasi sampai menemukan hal yang cocok dengan data.

Tahap *sampling* dapat dilakukan disini dan data secara umum dibagi menjadi dua, *data training* dan *data testing*.

Kegiatan yang dilakukan antara lain: memilih kasus dan parameter yang akan dianalisis (*Select Data*), melakukan transformasi terhadap parameter tertentu (*Transformation*), dan melakukan pembersihan data agar data siap untuk tahap *modeling* (*Cleaning*).

- ***Modeling***

Secara garis besar untuk membuat model prediktif atau deskriptif. Pada tahap ini dilakukan metode statistika dan *Machine Learning* untuk penentuan terhadap teknik *data mining*, alat bantu *data mining*, dan algoritma *data mining* yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap *data preparation*.

Beberapa modeling yang biasa dilakukan adalah *classification, scoring, ranking, clustering, finding relation*, dan *characterization*.

- ***Evaluation***

Melakukan interpretasi terhadap hasil dari data mining yang dihasilkan dalam proses pemodelan pada tahap sebelumnya. Evaluasi dilakukan terhadap model yang diterapkan pada tahap sebelumnya dengan tujuan agar model yang ditentukan dapat sesuai dengan tujuan yang ingin dicapai dalam tahap pertama.

- ***Deployment***

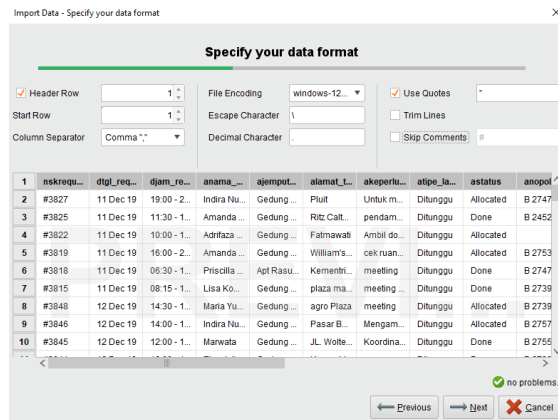
Tahap *deployment* atau rencana penggunaan model adalah tahap yang paling dihargai dari proses CRISP-DM. Perencanaan untuk *Deployment* dimulai selama *Business Understanding* dan harus menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional.

Pada akhirnya, rencana sistem *Deployment* mengakui bahwa tidak ada model yang statis. Model tersebut dibangun dari data yang diwakili data pada waktu tertentu, sehingga perubahan waktu dapat menyebabkan berubahnya karakteristik data.

Modelpun harus dipantau dan mungkin diganti dengan model yang sudah diperbaiki.

- **Menyusun Parameter Database**

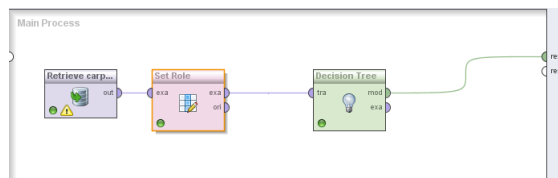
Output yang didapatkan dari dari penyusunan parameter yaitu pengelompokkan table yang akan dijadikan label dalam proses perhitungan seperti yang ditampilkan pada gambar 7.



Gambar 15. Paramater Database

- **Implementasi Metode C45**

Tabel yang sudah dikelompokkan akan menjadi titik acuan dalam perhitungan metode c45 seperti gambar design seperti pada gambar 8.



Gambar 16. Design Decision Tree

Proses perhitungan dari design pada figure 8 akan menghasilkan output, bahwa driver dengan kinerja tertinggi dengan tampilan visualisasi sebagai berikut.

4.2 Python Flask

- **Menyusun Parameter Database**

Dalam tahap penyusunan parameter database. Maka akan ditentukan parameter dari table yang akan digunakan untuk dilakukan analisis dan diimplementasikan C.45 berdasarkan data yang ditarik dari API. Maka akan terbentuk data seperti yang ditampilkan pada gambar 10 sebagai berikut.

Tabel Transaksi driver

	ID Transaksi	Driver	Status
1	#3827	Irfan AS	Allocated
2	#3825	Eko	Done
3	#3822	None	Allocated
4	#3819	Dwi Priharsono	Allocated
5	#3818	Irfan AS	Done
6	#3815	Adi Nugroho	Done
7	#3848	Adi Nugroho	Allocated
8	#3846	Ahmad Taufik	Allocated
9	#3845	Parlom P.	Done
10	#3841	Adi Nugroho	Done
11	#3840	Irfan AS	Done

Gambar 17. Tabel Data Transaksi

- **Tabel Implementasi Metode C.45 dalam Menghitung Nilai Kinerja Driver**

Berdasarkan pohon keputusan yang sudah didesain dan setelah dilakukannya preprocessing data maka akan tampil data nilai kinerja

pahlawan yang disusun dengan output tabel yang ditampilkan pada gambar 11.

Tabel Kinerja driver

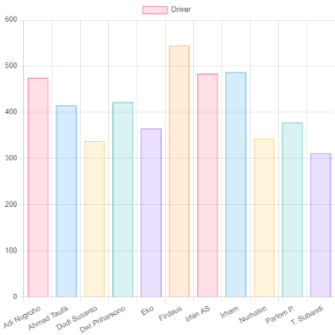
	Driver	Skor Rating
1	Adi Nugroho	477
2	Almad Taufik	415
3	Dodi Susanto	337
4	Dwi Prihansono	422
5	Eko	368
6	Firdaus	544
7	Irfan AS	490
8	Irfan	489
9	Nurhalim	344
10	Parlom P.	378
11	T. Subandi	311

Gambar 18. Hasil Akhir Kinerja Karyawan

- Graphic Kinerja Driver

Data akhir yang sudah dihasilkan dari perhitungan akan divisualisasikan. Data yang sudah divisualisasikan dapat dilihat pada gambar 12 sebagai berikut :

Grafik kinerja driver



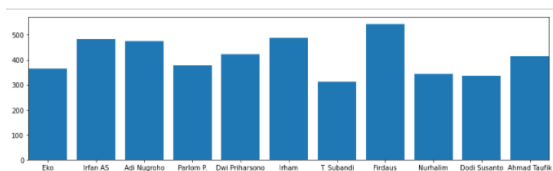
Gambar 19. Graphic Data Transaksi

BAB 5

Kesimpulan

5.1. Kesimpulan

Melalui proses data mining yang telah dilakukan menggunakan Algoritma C.45 dengan aplikasi *Rapidminer 5* diketahui informasi sebagai pengetahuan baru. Yaitu nilai kinerja tertinggi yaitu dengan poin 544.



Gambar 18 Graphic Data Kinerja

Disarankan kepada peneliti yang selanjutnya untuk melakukan perbandingan dengan 1 metode atau lebih selain algoritma C.45 yang merupakan bagian dari data mining metode klasifikasi guna

mendapatkan hasil yang berbeda dan lebih baik dengan tingkat akurasi yang lebih besar.

Disarankan kepada peneliti yang selanjutnya untuk melakukan perbandingan dengan 1 metode atau lebih selain algoritma C.45 yang merupakan bagian dari data mining metode klasifikasi guna mendapatkan hasil yang berbeda dan lebih baik dengan tingkat akurasi yang lebih besar. Dalam algoritma ini pemilihan atribut diproses menggunakan gain ration (Harani H N & Damayanti. 2021) Dari table yang sudah ditraining, maka akan terhitung tingkat. akurasinya seperti tabel berikut.

Model	Classification Error	Standard Deviation	Training Time (1,000 Rows)
Naive Bayes	0,4	0,0	47,2
Generalized Linear Model	0,1	0,0	136,3
Logistic Regression	0,1	0,0	56,9
Fast Large Margin	0,4	0,0	500,5
Deep Learning	0,2	0,0	395,5
Decision Tree	0,2	0,0	12,6
Random Forest	0,2	0,0	47,2
Gradient Boosted Trees	0,1	0,0	197,6

Tabel 1. Tabel Akurasi

Daftar Pustaka

- Umam, K. , D Puspitasari dan Achmad,N. 2020. Penerapan Algoritma C4.5 Untuk Prediksi Loyalitas Nasabah PT Erdika Elit Jakarta. Jurnal Media Informatika Budidarma. Volume 4, Nomor 1.
- Prianto, Cahyo, and Nisa Hanum Harani. "Sentiment Analysis of Covid-19 As A Social Media Pandemic." *IJISTECH (International Journal of Information System & Technology)* 4.1 (2020): 509-517.
- Irmayani, D dan Haris Musthafa. 2020. Sistem Informasi Pengelolaan Data Siswa pada SMA Negeri 02 Bilah Hulu Berbasis Web. Jurnal Ilmiah Fakultas Sains dan Teknologi. Universitas Labuhanbatu. Vol. 8, no.2.
- T.Novika, H.Okprana, A.P.Windarto, and H.Siahaan,“Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika,” vol.5, pp.9-17, 2021, doi: 10.30865/mib. v5i1.2498.
- K.F.Irmanda, D.Hartama, and A.P.Windarto, “Analisa Klasifikasi C4.5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi,” J.MediaInform. Budidarma, vol.5, no.1, pp.327–331, 2021, doi:10.30865/mib . v5i1 .2763
- Arifin, M.F. dan Firianah, Devi., “Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha

Persada” IncomTech, Jurnal Telekomunikasi dan Komputer, vol.8, no.2
Halaman 87-102, ISSN : 2085-4811, 2018.

Yunianita S, N Setiani dan Mulyati S. 2018. Prediksi Ketepatan Masa Studi Mahasiswa dengan Algoritma Pohon Keputusan C4.5. Universitas Islam Indonesia.

Nasrullah asmaul husnah, 2018 “Penerapan metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out” ILKOM Jurnal Ilmiah Volume 10 nomor 2 agustus 2018.

Sari, Febrina,, Saro, Davit 2018. “ Implementasi Algoritma C4.5 Dalam Menentukan Lokasi Prioritas Penyuluhan Program Keluarga Berencana Di Kecamatan Dumai Timur”, Jurnal Penelitian Pos dan Informatika Vol 8, No 1, 2018.

N. H. Harani and F. S. Damayanti, “Implementasi Algoritma C5.0 Untuk Menentukan Pelanggan Potensial Di Kantor Pos Cimahi,” J. SITECH Sist. Inf. dan Teknol. vol. 4, no. 1, pp. 69 – 76, 2021, doi : 10.24176/sitech.v4i1.6281

Putria, N.E. 2018. “Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori Pada Terminal Tiket Batam Tour & Travel”. Computer Based Information System Journal, Vol.06 (01), 29-38.

Rahmawati, F., Merlina, N. 2018. “Metode Data Mining Terhadap Dta Penjualan Sparepart Mesin Fotocopy Menggunakan Algoritma Apriori”. Jurnal Penelitian Ilmu Komputer, System Embedded & Logic, Vol.6 (1), 9-20.

Sulindawaty, Muhammad Zarlis, Niskarto Zendrato. 2018. “Strategi Peningkatan Penjualan Buku Pada PT. Tiga Serangkai Internasional Dengan Metode Clustering”. Seminar Nasional Teknologi Informasi dan Komunikasi STI&K, Vol 2. 102-110.

Fiona Y , Z Hassan dan Hamdani H (2021). *Macrozoobentos community structure in situ Ciburuy Padalarang West Bandung regency, West Java*. International Journal of Fisheries and Aquatic Studies 2021; 9(4): 280-289.

Eleanora, F. N., & Wijanarko, D. S. (2022). Jurnal internasional bereputasi (terindeks pada database internasional bereputasi dan berfaktor dampak) dengan Judul Child Friendly Principles Based On Justice In Child Development Institution Tangerang-Indonesia.

Pete Chapman (1999); The CRISP-DM User Guide.

Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth (2000); CRISP-DM 1.0 Step-by-step data mining guides.

Colin Shearer (2006); First CRISP-DM 2.0 Workshop Held

Mariscal, Gonzalo, Oscar Marban, and Covadonga Fernandez. 2010. “A Survey of Data Mining and Knowledge Discovery Process Models and Methodologies.” The Knowledge Engineering Review, Cambridge University Press 25:2(2010): 137–66.

David, Hansun, S. 2014. *Implementasi Data Mining Dengan Algoritma C4.5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa*, Tangerang.

Meythi dan Selvy Hartono. 2012. Pengaruh Informasi Laba dan Arus Kas Terhadap Harga Saham. Jurnal Akuntansi. 57– 59.