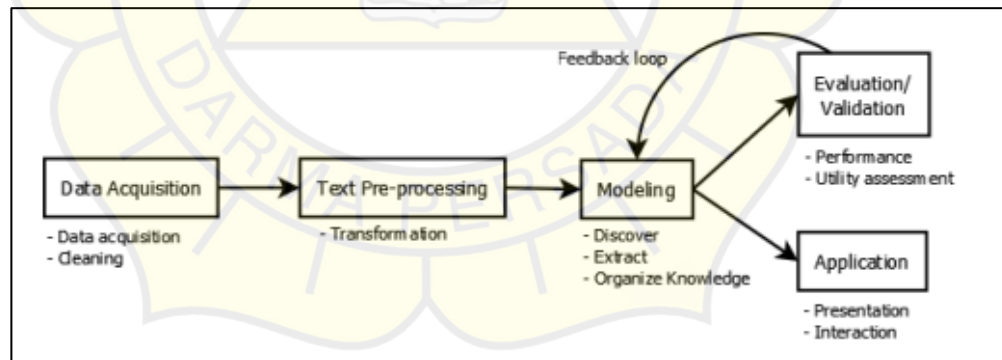


BAB II

LANDASAN TEORI

2.1 Text Mining

Onno W. Purbo (2019) memberikan pengertiannya mengenai *Text Mining* atau yang juga disebut sebagai proses analisis teks ialah upaya untuk memperoleh berbagai informasi dengan kualitas yang tinggi dari sebuah teks. Pengertian lainnya dari *Text Mining* yaitu prosedur yang mengharuskan penafsiran yang mendalam dan melibatkan individu yang berinteraksi dengan serangkaian dokumen seiring waktu, sambil menggunakan beragam alat analisis (Feldman & Sanger, 2007). Jadi, *text mining* ialah bidang yang sifatnya multi-disiplin berdasarkan eksplorasi informasi, penggalian informasi, *machine learning*, statistik, dan semantik komputasi.



Gambar 2. 1 Proses Data *Text Mining*

Sumber : Purbo W.Onno (2019)

Mirip dengan *data mining*, *text mining* memiliki tujuan untuk menarik informasi yang dapat digunakan dari asal mula data dengan melihat dan memeriksa pola-pola yang penting. Namun, pada suatu hal *text mining*, asal mula datanya ialah berupa koleksi dokumen serta berbagai pola-pola yang atraktif didapatkan bukan dari catatan *database* yang dilegalkan, melainkan berada di dalam data yang bersifat tekstual yang tidak sistematis di dalam dokumen-dokumen yang ada di dalam koleksi tersebut.

Teknik untuk *text mining* berkembang dan digunakan di berbagai sektor, mencakup sektor akademisi, *web application*, internet, serta sektor yang lainnya. *Text Mining* digunakan untuk melakukan *opinion mining*, *feature extraction*, *sentiment analysis*, prediksi, dan studi mengenai tren dalam suatu aplikasi seperti *search engine*, organisasi manajemen hubungan konsumen, penyaring pada email, studi mengenai saran produk, pendeteksi penipuan, dan studi mengenai media sosial.

2.2 Analisis Sentimen

Bing Liu (2012) membagikan pengertiannya mengenai analisis sentimen yaitu suatu disiplin ilmu yang meneliti tentang sentimen, pendapat, penilaian, sikap, dan emosi individu mengenai suatu subjek yang dapat berupa produk, layanan, institusi, pribadi, permasalahan, affair, topik, atau karakteristiknya. Nama dan fungsi lain untuk analisis sentimen yaitu *opinion mining*, *sentiment opinion*, ekstraksi opini, studi dari subjektivitas, studi mengenai pengaruh, studi emosi, penggalian mengenai ulasan, dan lainnya.

Pengertian lainnya mengenai analisis sentimen ialah proses otomatis dalam menafsirkan, melakukan ekstraksi, dan menggarap data yang bersifat tekstual untuk menghasilkan informasi mengenai sentimen yang tersemat pada ungkapan pendapat (Rozi, Pramono, & Dahlan, 2012). Pendekatan analisis sentimen ini digunakan untuk menentukan sikap maupun disposisi seseorang pada suatu permasalahan maupun objek terlepas dari bagaimana sudut pandang mereka positif atau negatif. Analisis sentimen memiliki beberapa manfaat dari berbagai sudut pandang, termasuk digunakan untuk memantau suatu produk, memprediksi penjualan, memantau politik, dan membantu pengambilan keputusan investor. Analisis sentimen ini juga dapat digunakan untuk mengetahui bagaimana respon dan perasaan masyarakat umum terhadap kualitas suatu layanan.

2.3 Media Sosial

Anang Sugeng Cahyono (2016) menjelaskan pengertiannya mengenai media sosial yaitu media platform yang bersifat daring yang memungkinkan pemakainya dapat mudah berperan serta, berbagi, dan menciptakan suatu konten, termasuk jejaring sosial, wiki, blog, forum, serta lingkungan virtual. Definisi lainnya, media sosial ialah sebuah platform daring yang mendukung adanya interaksi sosial serta memanfaatkan teknologi berbasis web untuk mengubah proses komunikasi menjadi suatu dialog yang bersifat interaktif. Media sosial dapat mengajak siapapun yang berminat serta terlibat dalam memberikan suatu kontribusi dan juga memberikan suatu *feedback* dalam

diskusi terbuka, berbagi ulasan serta menyebarkan informasi secara instan dan tanpa kendala.

2.4 Twitter

Twitter adalah sebuah layanan *microblogging* yaitu suatu perpaduan dari *blogging* dan juga pesan yang bersifat instan bagi pemakai yang telah terdaftar pada media sosial ini agar dapat berbagi, membuat postingan, menyukai, dan menanggapi postingan *tweet* dengan pesan yang singkat (Karami et al., 2020). Platform media sosial ini terkenal dengan menawarkan penggunaannya untuk menyiarkan postingan singkat yang disebut dengan *tweet* secara gratis. *Tweet* ini dapat berisi teks, video, foto, atau tautan.

Twitter biasanya digunakan oleh penggunaannya untuk tetap mengikuti berita dan promosi *brand*, terhubung dengan teman, dan juga untuk mempelajari apa yang dikatakan oleh para pebisnis, politisi, dan selebriti yang berpengaruh di media sosial. Selain itu, mereka juga menggunakannya untuk mengetahui informasi berita dan suatu acara. Twitter digunakan untuk berbagi informasi dengan cepat. Karena *tweet* dapat dikirimkan ke pengikut secara *real time*, *tweet* mungkin tampak seperti pesan yang bersifat instan (IM) bagi pengguna pemula. Namun tidak seperti pesan instan (IM) yang hilang saat penggunaannya menutup aplikasi tersebut, *tweet* juga diposting di situs *web* Twitter. *Tweet* memiliki sifat permanen, bisa dicari postingannya oleh penggunaannya, dan juga bersifat umum. Pengguna platform media sosial ini juga punya opsi atau pilihan untuk melindungi akun Twitter mereka sehingga hanya

dari pengikut akun pengguna tersebut saja yang dapat melihat *tweet* yang mereka posting.

Pada tanggal 23 Juli 2023, Twitter secara resmi memperbarui logo aplikasi Twitter dari burung biru menjadi "X" bergaris putih dengan latar belakang hitam. Perubahan ini sebenarnya sudah diantisipasi sejak lama sehubungan dengan pembelian "Twitter Inc" yang dilakukan oleh Elon Musk pada bulan Oktober 2022 senilai 44 miliar dolar, atau sekitar 643 triliun rupiah. Nama perusahaan "Twitter Inc." kini diubah menjadi "X Corp."

Twitter mempunyai beragam fitur yang dapat digunakan untuk berinteraksi dengan pengguna lainnya yaitu :

1. Tweet

Tweet adalah sebuah fitur yang dapat dimanfaatkan oleh penggunanya agar bisa mengekspresikan dirinya dengan cara membagikan konten berupa tulisan, foto, video, serta gambar *gif* ke bagian publik. Biasanya, *tweet* yang telah dipublikasikan di platform ini dapat dilihat oleh semua pemakai yang terdaftar atau juga yang tidak terdaftar, namun pemakai Twitter juga dapat menetapkan batasan pada postingannya agar dapat dilihat bagi *followers* terpilih saja.

Sesuai dengan kebijakan terbaru mengenai *rebranding* media sosial Twitter yang dilakukan oleh Elon Musk menjadi X, maka terdapat perubahan nama pada fitur kicauan (*tweet*) ini yaitu berganti nama menjadi Post.

2. Follow, Followers, dan Unfollowing

Sebutan pada *follow* yang memiliki arti pengikut, lalu sebutan *followers* yang artinya menjadi pengikut, serta sebutan *unfollowing* yang memiliki arti pembatalan pertemanan akan ditemukan oleh pengguna Twitter ketika menggunakan media sosial ini. Tujuan dari fitur ini adalah untuk memungkinkan penggunanya melakukan interaksi ke sesama pengguna di platform ini dan juga pengguna dapat menerima postingan *tweet* yang baru dari mereka secara *real-time* melalui *timeline* Twitter. Fitur blokir bisa digunakan oleh pemakainya untuk membatasi akses ke akun yang mereka tidak sukai atau yang mereka rasa dapat mengganggu mereka secara pribadi.

3. Tagar (*Hashtag*)

Tagar ialah suatu istilah atau rangkaian dari kata yang dimulai dengan tanda pagar "#" (*hashtag*). Pentingnya penggunaan simbol tersebut terletak pada fungsinya sebagai awal dari pembahasan topik yang tengah berlangsung di Twitter. Kemampuan untuk mengidentifikasi merek tertentu adalah keuntungan lain dari menggunakan fitur *hashtag* ini. Topik yang mendapatkan tingkat penggunaan *hashtag* yang tinggi juga memiliki potensi untuk masuk dalam tren di Twitter.

4. Username

Dalam pembuatan akun media sosial ini, para penggunanya pengguna diharuskan membuat nama pengguna (*username*) sesuai

dengan preferensi atau keinginan masing-masing dari pengguna itu sendiri. Setiap akun juga bisa diidentifikasi dari nama penggunanya dan juga peranannya memiliki signifikansi besar dalam membedakan satu pengguna dari yang lainnya. Pemakaian pada tanda “@” ini disertai dengan *username*.

5. **Retweet**

Istilah dari *retweet* ialah fitur yang diciptakan oleh Twitter yang merujuk pada proses penyebaran postingan *tweet* yang diterbitkan sebelumnya. Pengguna Twitter dapat memposting ulang *tweet* dari *tweet* yang dibuat di masa lalu, baik *tweet* itu milik mereka sendiri atau *tweet* yang berasal dari akun pengguna yang lainnya. Penggunaan fungsi ini berguna saat postingan *tweet* sesuai dengan situasi yang tengah berlangsung.

Berdasarkan kebijakan terbaru mengenai *rebranding* Twitter yang dilakukan oleh Elon Musk menjadi X, sama seperti fitur *tweet* yang berganti nama menjadi *post* maka fitur *retweet* juga melakukan perubahan nama yaitu *Repost*.

6. **URL Shortener**

Ketika berbagi tautan, Twitter secara otomatis akan melakukan proses memperpendek menjadi suatu tautan yang dimulai dengan <https://t.co/>, lalu diikuti oleh *URL Shortener*. Kini pengguna Twitter dapat berbagi *tweet* dengan batas 280 karakter tanpa khawatir dengan panjang tautan yang ingin mereka sertakan.

7. Trending Topics

Banyaknya pembicaraan atau obrolan yang dilakukan oleh pengguna Twitter dapat didasarkan pada tagar, kata, atau frasa tertentu ialah arti dari *Trending Topic*. Fitur *Trending Topics* pada Twitter ini sangat berguna untuk memberi tahu penggunanya mengenai apa yang sedang terjadi dan apa saja yang sedang populer saat ini.

8. Tampilan

Salah satu dari banyak fitur Twitter yang terbaik dan mutakhir adalah suatu opsi bagi penggunanya untuk memilih di antara tiga skema warna yang berbeda untuk mengubah tampilan *interface* yaitu Tampilan Lampu Mati, Bawaan serta Temaram. Pemakai akun Twitter tentunya bisa mengubah warna serta ukuran *font* hurufnya dari tampilan *interface* tersebut. Dengan adanya pembiasaan tampilan yang ditampilkan oleh media sosial ini akan meningkatkan impresi bagi penggunanya agar lebih nyaman ketika menggunakan media sosial ini.

2.5 Text Preprocessing

Text Preprocessing ialah tahap di mana aplikasi melakukan proses penyortiran data pada setiap dokumen teks (Hermawan & Bellanar Ismiati, 2020). Dalam proses menganalisis teks menggunakan *Text Mining*, *Natural Language Processing*, serta *Information Retrieval (IR)*, proses ini menjadi suatu aktivitas serta fase yang sangat penting. Pada *text mining*, *text preprocessing* digunakan untuk mengambil informasi yang menarik dan

signifikan dari data teks yang sifatnya tidak terstruktur. Berikut ini adalah beberapa tahapan dalam melakukan *text preprocessing* yaitu :

1. Cleansing

Cleansing adalah suatu proses pembersihan karakter atau tanda baca termasuk yang tidak diperlukan dalam kerangka pemikiran informasi yang ada. Misalnya dalam pengambilan data *tweet*, karakter semacam *url*, *username*, dan lain-lain akan dihilangkan pada proses *cleansing* ini.

2. Case Folding

Dalam metode dari *case folding* ialah melibatkan adanya perubahan yang terjadi pada setiap huruf kapital dalam dokumen teks yang berubah menjadi huruf kecil (*lowercase*) sehingga sistem dapat membandingkan setiap karakter dalam dokumen dengan setiap karakter lainnya.

3. Normalization

Normalization atau normalisasi adalah suatu tahapan dimana kalimat atau kata ejaan yang disingkat diperbaiki agar sesuai KBBI yang bertujuan untuk mempermudah dalam penelitian.

4. Tokenizing

Tokenizing adalah suatu proses pemotongan atau pemecahan teks menjadi suatu potongan yang lebih kecil atau biasa disebut dengan *token*.

5. Filtering / Stopword Removal

Filtering ialah proses menghilangkan kata-kata yang sifatnya tidak relevan dari proses *tokenizing*, contohnya pada kata ‘yang’, ‘di’, ‘ke’, dan lain sebagainya. Untuk membuatnya lebih mudah dalam proses membandingkan teks, kata-kata tertentu dihapus dari dokumen.

6. Stemming

Stemming ialah suatu langkah untuk menemukan akar dari suatu kata (kata dasar) dari kata-kata yang memiliki imbuhan yang berasal dari hasil penghapusan kata-kata yang tidak relevan (*filtering*). Pada tahapan ini dilakukan proses penormalan struktur kata agar menghasilkan representasi yang seragam.

2.6 Term Frequency-Inverse Document Frequency (TF-IDF)

Onno W. Purbo (2019) menjelaskan bahwa TF-IDF ialah nilai numerik yang dirancang untuk merepresentasikan seberapa signifikannya suatu kata dalam sebuah dokumen yang ada pada koleksi maupun pada korpus. Pengertian dari TF yang memiliki kepanjangan dari *Term Frequency* ialah suatu frekuensi yang menyatakan seberapa seringnya suatu kata dapat muncul dalam dokumen. Oleh sebab itu, setelah proses TF tersebut maka dihasilkan DF. Pengertian dari DF yang memiliki kepanjangan dari *document frequency* ialah sejumlah dokumen yang memuat kata tersebut. Nilai TF-IDF digunakan untuk menentukan bobot kata pada dokumen.

Hasil TF-IDF didapatkan dengan cara melakukan sebuah perkalian antara TF dan IDF, dimana IDF tersebut ialah hasil *invers* yang berasal dari DF.

Rumus dari TF-IDF dapat dijabarkan di bawah ini :

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \quad (1)$$

$$TF - IDF(w, d) = TF(w, d) \times IDF(w) \quad (2)$$

Keterangan:

IDF(w) : Bobot kata dalam seluruh dokumen

w : Sebuah kata

TF(w,d) : Frekuensi kemunculan kata w dalam dokumen d

IDF(w) : *Inverse* DF dari kata w

N : Jumlah seluruh dokumen

DF(w) : Jumlah dokumen yang mengandung kata w

Jika kita lakukan analisis dari perhitungan IDF dengan rumus pertama, jika N sama dengan DF(w), maka hasilnya akan menjadi 0 (nol). Untuk mengatasi situasi tersebut, dapat dimasukkan nilai 1 pada bagian IDF, sehingga perhitungan TF (w, d) akan menjadi seperti yang dijabarkan berikut ini :

$$TF - IDF(w, d) = TF(w, d) \times \left(\log\left(\frac{N}{DF(w)}\right) + 1\right) \quad (3)$$

Maka langkah selanjutnya, untuk mengubah nilai TF-IDF ke dalam rentang 0 hingga 1, proses normalisasi diperlukan. Rumus (3) dinormalisasi menggunakan rumus (4) seperti dijabarkan di bawah ini :

$$TF - IDF(w, d) = \frac{TF - IDF(w, d)}{\sqrt{\sum_{w=1}^n TF - IDF(w, d)^2}} \quad (4)$$

Dimana $\sum_{w=1}^n TF - IDF(w, d)$ ialah total dari nilai TF-IDF dari kata-kata pertama hingga kata ke n yang ada pada dokumen d.

2.7 CRISP-DM

Cross-Industry Standard Process for Data Mining ialah metode yang mampu diimplementasikan ke suatu pendekatan penyelesaian permasalahan umum serta metodologi yang menetapkan standar umum untuk proses penggalian data (Adhi Putra, 2021). Arti lain dari CRISP-DM yaitu suatu model dari proses industri mandiri yang biasanya dipakai untuk kepentingan dari penggalian data (Kurniawan & Yasir, 2022). Dalam hal menangani kebutuhan proyek industri selama siklus pembangunannya, CRISP-DM dianggap sebagai pendekatan penggalian data yang bersifat komprehensif dan kini menjadi yang paling sering digunakan dalam suatu proyek analitik, proses *data mining*, dan ilmu data. Terdapat enam langkah yang dijalankan dari CRISP-DM ini yaitu :

1. Pemahaman Bisnis

Pemahaman Bisnis memiliki arti suatu pemahaman mengenai tugas-tugas dari *data mining* yang harus dilakukan serta persyaratan dari sudut pandang bisnis. Tugas-tugasnya yaitu mulai dari merumuskan rencana strategi dan jadwal penelitian serta memutuskan tujuan atau sasaran perusahaan dan tujuan dari *data mining*.

2. Pemahaman Data

Tujuan dari fase ini ialah untuk melakukan akumulasi data awal, menganalisis data untuk memahami informasi yang akan digunakan, menemukan masalah dengan kualitas data, dan juga mengidentifikasi bagian data yang menarik untuk menarik suatu dugaan awal.

3. Persiapan Data

Proses ini selalu dikatakan seperti tahapan yang bersifat padat karya. Proses pemilihan tabel serta *field* yang akan digunakan untuk membuat program basis data baru (dataset mentah) untuk *data mining* adalah salah satu aktivitasnya.

4. Pemodelan

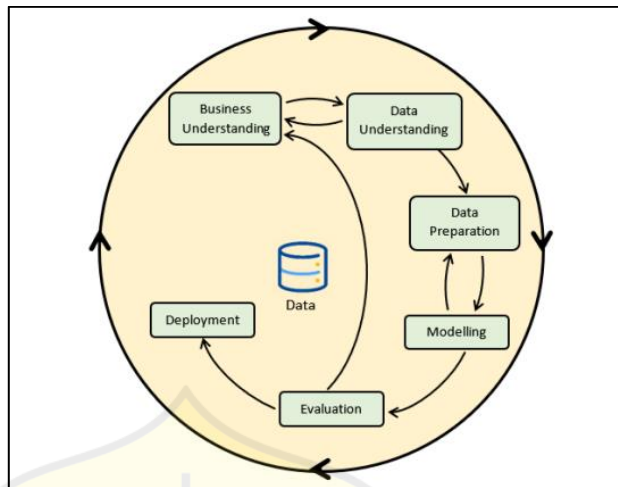
Pada fase pemodelan ini melibatkan pemilihan metode *data mining* yang akan digunakan, serta alat, teknik, algoritma, dan parameter dengan nilai terbaik.

5. Evaluasi

Tahapan ini ialah tahap di mana temuan-temuan *data mining* yang terungkap selama proses awal pemodelan diinterpretasikan. Evaluasi dilakukan secara cermat dengan maksud untuk mengadaptasi model yang telah diperoleh agar hasilnya sesuai dengan tujuan yang ingin dicapai pada langkah awal.

6. Penyebaran

Dalam tahapan ini dilakukan penyusunan laporan atau presentasi dari pengetahuan yang diperoleh dari evaluasi pada suatu proses *data mining*.

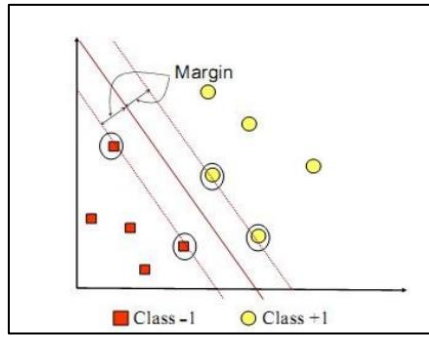


Gambar 2. 2 Tahapan CRISP-DM

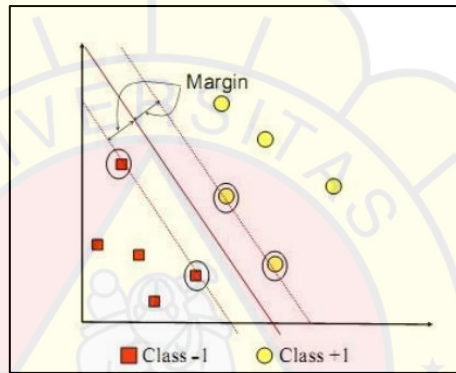
Sumber : Denni Kurniawan, Muhammad Yasir (2022)

2.8 Support Vector Machine (SVM)

Support Vector Machine (SVM) memiliki arti yaitu algoritma dari pembelajaran mesin yang mengimplementasikan metode *hyperplane* yang ada dari suatu data sampai terbentuklah ruang pada tiap kelas (Fikri et al., 2020). Pengertian dari *hyperplane* adalah sebuah fungsi yang berperan menjadi pemisah di antara kelas-kelas yang ada. Algoritma ini diciptakan oleh Vapnik pertama kalinya pada tahun 1992 sebagai susunan yang bersifat harmonis dari konsep-konsep utama pada bidang *pattern recognition* (Susilowati, Sabariah, & Gozali, 2015). Ciri dari algoritma ini yaitu memiliki kesalahan pembangkitan yang rendah (*low generation error*), tidak mahal secara komputasi, mudah untuk dimengerti, dan juga peka terhadap pemilihan kernel dan penyetelan parameter (Budiharto, 2016).



Gambar 2. 3 Ilustrasi SVM Saat Menemukan *Hyperplane* Terbaik
 Sumber : Susilowati, Sabariah, & Gozali (2015)



Gambar 2. 4 *Hyperplane* terbentuk di antara Class-1 dan Class+1
 Sumber : Susilowati, Sabariah, & Gozali (2015)

Gaussian RBF menjadi pilihan kernel yang akan dipakai dalam penelitian ini. Berikut ini akan dijabarkan rumus kernel yang terdapat dalam metode *Support Vector Machine* :

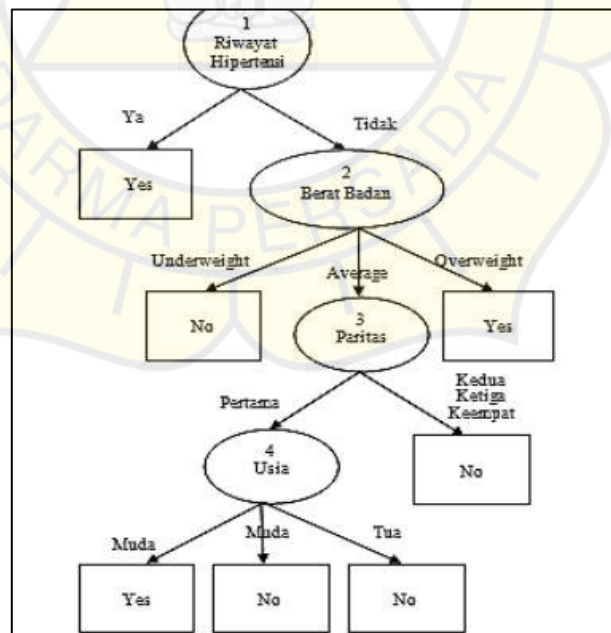
Tabel 2. 1 Rumus Kernel

Jenis Kernel	Model
<i>Polynomial</i>	$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i, \vec{x}_j + 1)^p$
<i>Gaussian RBF</i>	$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\frac{\ \vec{x}_i - \vec{x}_j\ ^2}{2\sigma^2}\right)$

<i>Sigmoid</i>	$K(\vec{x}_i, \vec{x}_j) = \tanh(\alpha \vec{x}_i, \vec{x}_j + \beta)$
<i>Linear</i>	$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^t \vec{x}_j$

2.9 Decision Tree

Decision Tree yaitu suatu metode klasifikasi serta prediksi yang sangat efektif dan terkenal, dimana metodenya dapat melakukan konversi dari kumpulan data yang sangat besar menjadi sistem pohon keputusan yang mempresentasikan aturan (Muzakir & Wulandari, 2016). Ketentuan dari pendekatan ini mudah dimengerti dalam bahasa yang sederhana serta bisa direpresentasikan dalam *database* menggunakan bahasa yang disebut dengan *Structure Query Language (SQL)* untuk mendapatkan *record* dalam sekumpulan data tertentu.



Gambar 2. 5 Model *Decision Tree*
 Sumber : Muzakir & Wulandari, 2016

Karakteristik dari *decision tree* disusun dari beberapa elemen yang dijabarkan berikut ini :

1. *Root node* memiliki ciri-ciri yaitu tak ada lengan masukan serta juga mempunyai lengan keluaran sejumlah 0 ataupun lebih
2. *Internal node* memiliki ciri-ciri yaitu node yang bukan daun yang bersifat *nonterminal* yang memiliki tepat satu lengan
3. Lengan memiliki ciri-ciri yaitu setiap cabang menunjukkan nilai dari hasil pengujian pada node bukan daun.
4. *Leaf node* (terminal) memiliki ciri-ciri yaitu node yang memiliki tepat satu lengan masukan serta tidak memiliki lengan keluaran. Node ini mewakili label kelas (keputusan).

2.9.1 Algoritma C4.5

Algoritma ini merupakan algoritma yang diciptakan oleh J.Ross Quinlan yang metodenya dikembangkan dari algoritma ID3 (*Iterative Dichotomiser*) (Muzakir & Wulandari, 2016). Algoritma C4.5 secara umum mengikuti langkah-langkah berikut saat membuat *decision tree* :

1. Pilih sebuah atribut untuk digunakan sebagai *root*
2. Di tiap nilai dibangun cabang
3. Dalam cabang dibagi kasus
4. Setelah tiap kasus dalam sebuah cabang punya kelas yang sama, ulangi prosedur untuk setiap cabang.

Berdasarkan nilai *gain* tertinggi dari atribut saat ini, satu atribut dipilih sebagai *root*. Dijabarkanlah rumus di bawah ini untuk melakukan perhitungan *gain* :

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

Keterangan :

S = Himpunan Kasus

A = Atribut

n = Jumlah partisi atribut A

|S_i| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

Sementara itu, untuk perhitungan nilai entropi dapat dilihat pada rumus berikut ini :

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (2)$$

Keterangan :

S = Ruang (data) sampel yang digunakan untuk pelatihan

n = Jumlah partisi S

P_i = Proporsi dari S_i terhadap S

2.10 Confusion Matrix

Confusion Matrix adalah suatu tabel yang menyajikan informasi yang isinya adalah perbandingan antara hasil klasifikasi prediksi sistem dengan hasil klasifikasi aktual yang sebenarnya (Fikri et al., 2020). Tabel pada *confusion matrix* menunjukkan total dari data uji yang hasil klasifikasinya benar dan total data uji yang hasil klasifikasinya salah. Berikut ini dijabarkan tabel di bawah ini adalah contoh *confusion matrix* dengan klasifikasi secara biner :

Tabel 2. 2 Confusion Matrix

		Prediksi	
		Positif	Negatif
Observasi / Aktual	Positif	TP	FN
	Negatif	FP	TN

Keterangan :

TP (*True Positive*) = Total keseluruhan data dari kelas positif yang hasil prediksinya benar sebagai kelas positif

TN (*True Negative*) = Total keseluruhan data dari kelas negatif yang hasil prediksinya benar sebagai kelas negatif

FP (*False Positive*) = Total keseluruhan data dari kelas negatif yang hasil prediksinya salah sebagai kelas positif

FN (*False Negative*) = Total keseluruhan data dari kelas positif yang hasil prediksinya salah sebagai kelas negatif

Berikut ini adalah rumus persamaan *confusion matrix* untuk menghitung *accuracy*, *precision*, *recall* serta *F Score* adalah sebagai berikut :

$$accuracy = \frac{TP+TN}{Total} \quad (1)$$

$$precision = \frac{TP}{TP+FP} \quad (2)$$

$$recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall+Precision} \quad (4)$$

2.11 Python

Abdul Kadir (2019) memberikan pengertiannya mengenai Python yaitu sebuah bahasa pemrograman rancangan dari Guido van Rossum yang bersifat serbaguna yang diciptakan pada tahun 1991. Bahasa pemrograman ini mempunyai tujuan agar dapat membuat kode mudah untuk dipahami. Untuk bentuk wujudnya, pemakaian spasi dalam menggarap indentansi kode sangat signifikan. Oleh karena itu tidak bisa sembarangan untuk menambahkan spasi ke kode secara acak.

Python bersifat *open source* dengan sebagian besar versinya menggunakan lisensi *GFL-compatible*. Kemampuan bahasa pemrograman ini untuk menjadi bahasa yang dinamis dengan fitur manajemen memori otomatis adalah salah satu fiturnya. Python bisa dipakai untuk banyak platform dari sistem operasi, termasuk Linux / Unix, Mac OS X serta Windows.

2.11.1 Flask

Pengertian dari Flask ialah *framework* yang dipakai untuk pengembangan web yang dirancang menggunakan bahasa pemrograman Python dan juga dapat digolongkan ke *microframework* karena tidak membutuhkan *tools* atau *library* khusus serta mempunyai *database* bawaan (Singh, Verma, Parasher, Chauhan, & Budhiraja, 2019). Pengertian lainnya dari Flask ini yaitu *micro framework* dari bahasa pemrograman Python yang menyediakan fungsionalitas dasar *web framework* dan memungkinkan lebih

banyak *plug-in* untuk ditambahkan sehingga fungsionalitas dan rangkaian fitur dapat diperluas ke tingkat yang baru.

Yang dimaksud dengan *micro framework* yaitu Flask bermaksud menjaga *core* aplikasi dengan sesederhana mungkin, namun tetap dapat dikembangkan. Dengan cara ini, Flask dapat menunjukkan fleksibilitas dan skalabilitas yang cukup besar jika dibandingkan dengan *framework* yang lain. Flask mendukung ekstensi. Ekstensi dapat dipakai untuk melakukan pemilahan relasi antar objek, memvalidasi formulir, menangani suatu unggahan, beraneka ragam teknologi autentikasi terbuka, dan masih banyak yang lainnya.

2.12 Javascript

Rohi Abdulloh (2018) menjelaskan pengertiannya mengenai Javascript yaitu sebuah bahasa pemrograman yang dipakai untuk *web* yang dijalankan prosesnya di sisi *client*. Dengan prosesnya yang berjalan di sisi *client*, Javascript bisa dijalankan hanya memakai *browser*. Saat hal-hal eksklusif berlangsung dalam laman *web* — baik hal yang dikerjakan oleh pengguna ataupun yang disebabkan oleh modifikasi pada halaman *web* — Javascript biasanya dijalankan.

2.13 CSS

Rohi Abdulloh (2018) menjelaskan pengertiannya mengenai CSS atau *Cascading Style Sheet* yaitu dokumen dari *web* yang fungsinya ialah untuk mengatur unsur dari HTML dengan beragam properti yang tersedia, jadi

website bisa ditampilkan dengan beragam gaya sesuai dengan keinginan. CSS adalah kumpulan kode untuk mengkarakterisasi rencana bahasa *markup* karena kata bahasa *markup* dalam arti CSS dan HTML sangat dekat. Jika didefinisikan secara luas, CSS berfungsi untuk membuat serta meningkatkan tampilan dari suatu laman.

2.14 Bootstrap

Rohi Abdulloh (2018) menjelaskan pengertiannya mengenai Bootstrap sebagai salah satu *framework* CSS yang sangat terkenal diantara banyaknya *framework* CSS yang telah ada. Dengan adanya bantuan Bootstrap, desain *website* dapat dibuat secara responsif sehingga terlihat baik di semua jenis perangkat. Bootstrap juga mempercepat proses konfigurasi desain dengan mengurangi kebutuhan untuk menulis banyak kode CSS, bahkan hanya dibutuhkan sedikit atau hampir tidak diperlukan, kecuali jika ada kebutuhan untuk menyesuaikan desain dengan gaya yang berbeda dari Bootstrap.

2.15 MySQL

Rahimi Fitri (2020) memberikan pengertiannya mengenai MySQL yaitu sebuah *server* basis data yang bebas biaya dan memiliki lisensi GNU *General Public License (GPL)* yang dapat dipakai oleh penggunanya untuk keperluan pribadi atau komersial. MySQL termasuk ke dalam tipe RDBMS (*Relational Database Management System*). Maka sebutan baris, kolom, dan tabel masuk dalam MySQL ini.

2.16 XAMPP

Rahimi Fitri (2020) memberikan pengertiannya tentang XAMPP yaitu sebuah aplikasi yang pengoperasiannya *server*-nya dikembangkan oleh Apache Friends serta bersifat *open source*. Karena XAMPP memiliki sifat *open source*, aplikasi ini dapat dipakai gratis oleh para penggunanya. XAMPP bersifat *cross platform* yang mempunyai arti bahwa aplikasi ini *support* terhadap beragam *platform* sistem operasi contohnya macOS, Linux serta Windows. XAMPP terdiri dari Apache, MariaDB, PHP dan juga Perl.

2.17 Visual Studio Code

Visual Studio Code adalah *source code editor* yang bersifat *free*, *open-source* dan *cross-platform* yang dikembangkan oleh Microsoft (Speight, 2021). Visual Studio Code menjadi *source code editor* yang bersifat ringan namun kuat, prosesnya dapat dijalankan di *desktop* dan sistem operasi contohnya adalah Linux, Windows, serta macOS. Teks editor ini mendukung langsung bahasa pemrograman layaknya Javascript, Typescript, serta Node.js. Contoh bahasa pemrograman lainnya yaitu bahasa pemrograman *support plugin* yang bisa dipasang dari *marketplace* aplikasi ini seperti C++, C#, Python, Go, Java, dan lainnya.

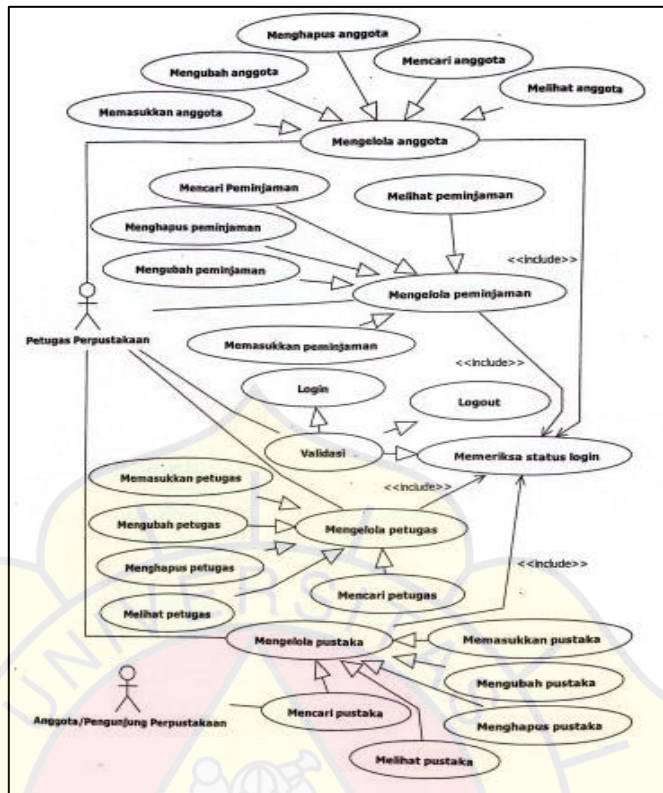
2.18 UML (Unified Modelling Language)

UML (*Unified Modelling Language*) adalah bahasa visual yang dipakai sebagai model serta bahan komunikasi tentang suatu sistem, dibangun dengan

menggunakan diagram serta teks pendukung (A.S & Shalahuddin, 2016). UML telah menjadi salah satu standar bahasa umum yang digunakan di sektor industri untuk mendeskripsikan suatu *requirement*, merancang analisis desain, beserta memaparkan arsitektur dalam pemrograman berorientasi objek. UML dikembangkan sebagai bahasa visual untuk berinteraksi secara saling berhubungan serta sebagai sistem pemodelan yang memanfaatkan teks dan diagram pendukung.

2.18.1 Use Case Diagram

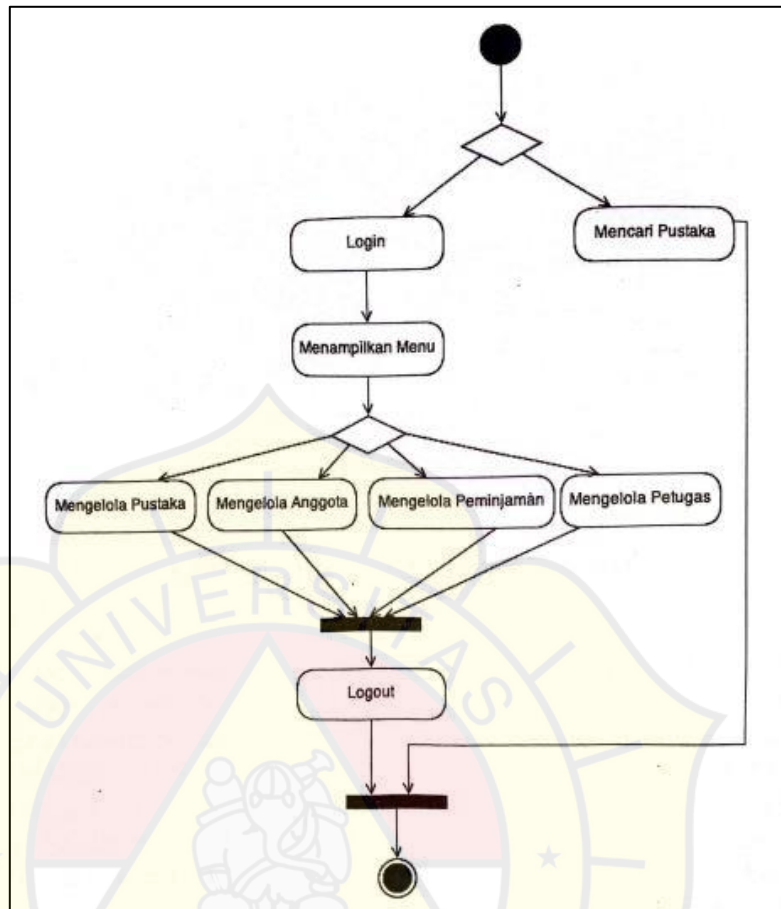
Use Case Diagram ialah suatu representasi dari model untuk suatu tingkah laku (*behavior*) dari sistem informasi yang ingin dirancang. Dari *use case diagram* ini menunjukkan suatu hubungan dari satu aktor maupun lebih dari satu aktor dan sistem informasi tertentu. Bisa dibilang bahwa penggunaan dari *use case* ini dipakai dalam mengidentifikasi kinerja sistem informasi dan pengguna yang berwenang untuk menggunakannya.



Gambar 2. 6 Use Case Diagram
 Sumber : A.S & Shalahuddin (2016)

2.18.2 Activity Diagram

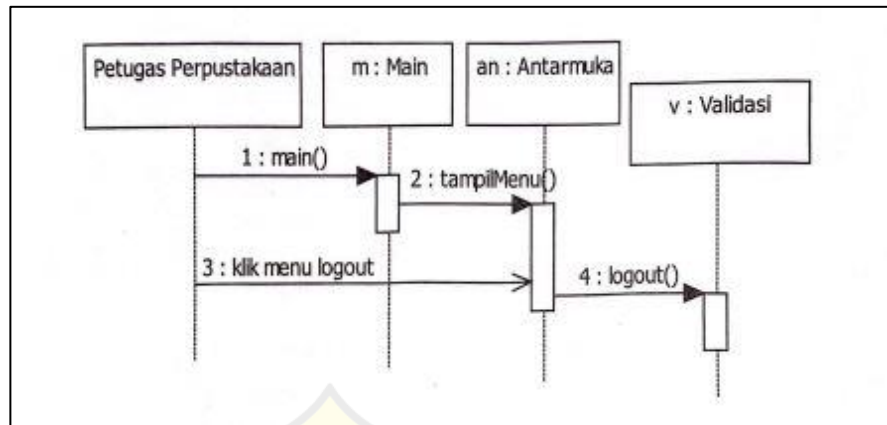
Activity Diagram memiliki fungsi yaitu mendeskripsikan aktivitas dari suatu sistem atau operasi dari bisnis serta dapat mendeskripsikan menu dalam *software*. Sangat penting untuk diingat bahwa diagram ini menunjukkan aktivitas sistem daripada tindakan yang diambil oleh aktor, atau aktivitas yang mampu dilakukan oleh sistem.



Gambar 2.7 Activity Diagram
 Sumber : A.S & Shalahuddin (2016)

2.18.3 Sequence Diagram

Diagram ini memaparkan objek pada *use case* dari pengguna dengan menjelaskan periode hidup objek serta pesan yang dikirim juga diterima diantara objek tersebut. Maka, objek yang berperan serta dalam *use case* dan operasi *class* yang digunakan ke dalam objek tersebut harus dipahami untuk menggambarkan diagram sekuen.



Gambar 2. 8 Sequence Diagram

Sumber : A.S & Shalahuddin (2016)

2.19 Web Hosting

Andy Kristianto, ST (2012) menjelaskan pengertiannya mengenai *Web Hosting* ialah suatu jasa sewa dari tempat mencadangkan data di internet yang dibutuhkan oleh suatu *website*. Tujuan yang ada dari *web hosting* ini supaya *website* dapat bekerja dalam mode *online* serta dapat diakses internet dari seluruh dunia. Dari segi fisik, komputer dan perangkat-perangkatnya pada *web hosting* berperan sebagai *server* internet. Kapasitas serta *bandwidth* menjadi ukuran yang dipakai pada sistem *web hosting*.

Kapasitas merujuk pada kemampuan suatu *web hosting* dalam menampung data dari *website*. Jika kapasitas *web hosting* semakin besar, maka semakin besar pula kapasitas dari penyimpanan data yang tersedia. *Bandwidth* ialah batas maksimal jumlah volume data yang diizinkan untuk diakses dari *web hosting* untuk tiap bulannya.

2.19.1 VPS

VPS (*Virtual Private Server*) ialah versi *web hosting* yang memakai *virtual* mesin. Jadi, tiap pemakai *web hosting* seolah-olah mempunyai komputer sendiri. Jika terjadi gangguan pada salah satu pemakai *hosting*, maka pemakai lainnya tidak akan terkena dampak dari gangguan tersebut.

2.19.2 Domain

Domain ialah identitas yang fungsinya untuk melakukan akses pada suatu *website*. Jika dijelaskan dengan bahasa sehari-hari, kita selalu menamakan *domain* dengan sebutan alamat *website*. Berikut ini contoh *domain* yaitu : **ugm.ac.id**, **google.com**, **indonesia.travel**, **wikipedia.org** dan lainnya.

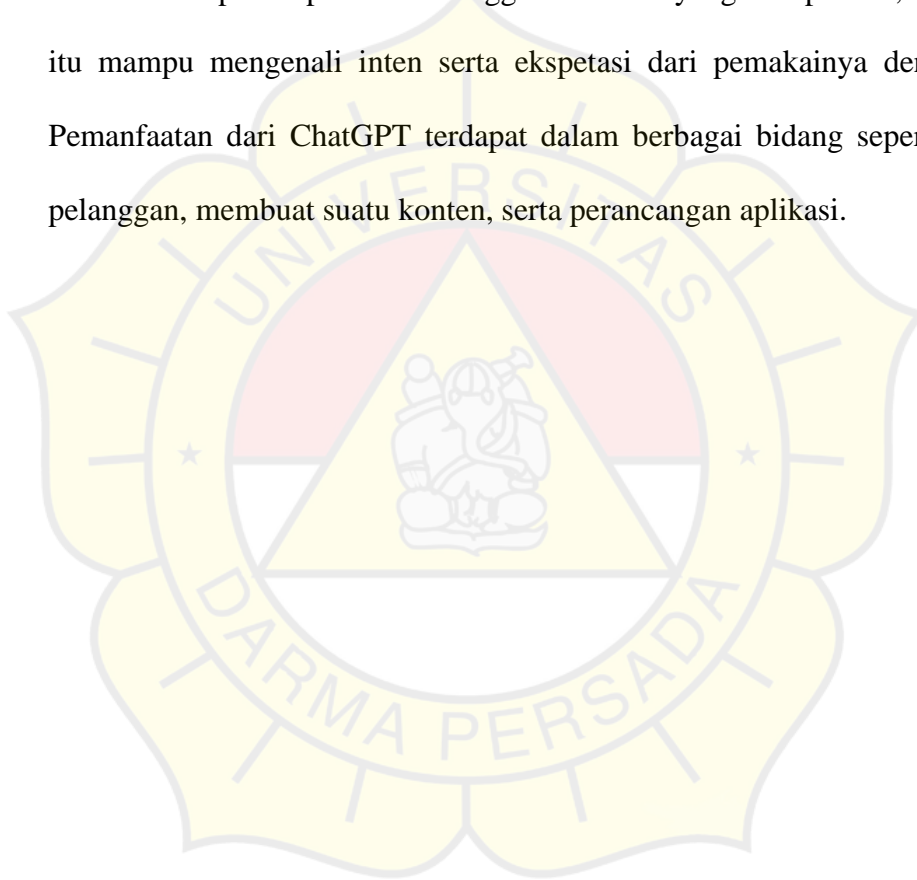
2.19.3 CPanel

CPanel ialah *software control panel* dengan mode *online* yang berfungsi untuk mengatur *website* dalam sebuah *web hosting* dan juga berperan menjadi alat utama untuk pemakai *web hosting*. Yang dapat dilakukan dalam CPanel yaitu dapat mengunggah data *website*, mengatur data *website*, melakukan instalasi pada *content management system*, dan lainnya.

2.20 ChatGPT

Pengertian dari ChatGPT (*Generative Pre-trained Transformer*) ialah suatu model dalam bidang *deep learning* yang dirancang di tahun 2018 oleh pihak OpenAI. (Setiawan & Luthfiyani, 2023). ChatGPT dirancang menjadi

model yang bersifat generatif yang bisa dipakai untuk mendapatkan teks yang sinkron dengan situasi percakapan. ChatGPT memakai arsitektur *Transformer* yang diketahui sebagai model *self-attention*. Dengan menggunakan arsitektur ini, ChatGPT dapat mengatasi percakapan yang kompleks dengan memperhatikan situasi dari percakapan sebelumnya. Di samping itu, ChatGPT telah melalui proses pelatihan menggunakan data yang cukup besar, oleh karena itu mampu mengenali inten serta ekspetasi dari pemakainya dengan baik. Pemanfaatan dari ChatGPT terdapat dalam berbagai bidang seperti layanan pelanggan, membuat suatu konten, serta perancangan aplikasi.





TEKNOLOGI INFORMASI
UNIVERSITAS DARMA PERSADA