

## **BAB II**

### **LANDASAN TEORI**

#### **2.1. Analisis Sentimen**

Fanissa et al (2018) mengatakan bahwa “analisis sentimen adalah bidang interdisipliner, di mana pendekatan pemecahan masalah menggunakan ulasan dari berbagai sudut pandang ilmiah yang digabungkan dengan cara yang tepat dan saling melengkapi.”

Proses atau cara memahami, mengekstraksi, dan menginput data disebut juga dengan analisis sentimen untuk dapat informasi sentimen yang mengandung dalam suatu kalimat pendapat. (Yusnitasari, ikasari, Pratiwi, Syahri, dan Ramdani, 2017).

#### **2.2. Data Mining**

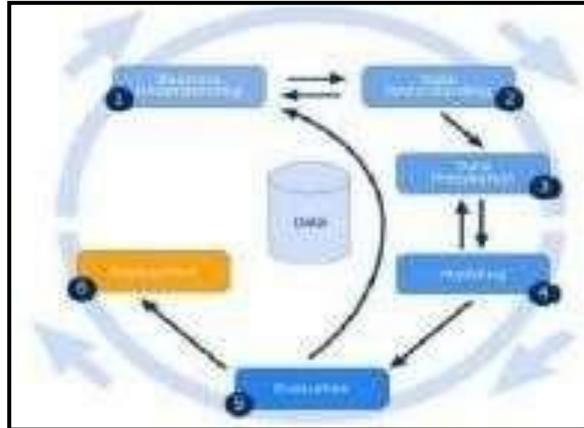
Sianturi, F. A. (2017) menyatakan “Data mining yaitu proses yang menggunakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk dilakukannya analisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah suatu ilmu berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang melakukan cara observasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Databases* (KDD) merupakan penerapan metode saintifik pada data mining. Pembahasan ini data mining merupakan satu langkah dari proses KDD”.

Menurut Suntoro (2019) yaitu pada buku yang berjudul “Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP” menjelaskan bahwa “Data mining ialah suatu proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu mengekstrasi agar menjadi informasi baru dan dapat membantu dalam pengambilan suatu keputusan”.

### **2.3. CRISP-DM**

CRISP-DM adalah model proses penambangan data (Data Mining Framework) yang awalnya dibuat pada 1996 oleh 5 perusahaan, yaitu Integral Solutions Ltd (SL), Teradata, Daimler AG, NCR Corporation dan ORHA. Framework ini telah dikembangkan oleh ratusan organisasi dan perusahaan di Eropa menjadi metodologi standar non-proprietary untuk data mining. Versi pertama dari metodologi ini dipresentasikan pada tahun keempat CRISP-DM SIG di Brussels pada Maret tahun 1999 (Pete Chapman, 1999) dan proses ekstraksi data langkah demi langkah berdasarkan model ini diterbitkan pada tahun berikutnya (Pete Chapman, 2000).

Dalam penelitian ini, penulis menggunakan metodologi data mining CRISP-DM sebagai pemecah masalah umum untuk bisnis dan penelitian. Metodologi ini terdiri dari enam langkah, yaitu: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment. Proses sistematis ini terdiri dari 6 langkah yang dapat dijelaskan di bawah ini.



**Gambar 2.1 Skema CRISP-DM**

1. Business Understanding, yaitu memiliki beberapa hal yang dilakukan pada tahap ini seperti memahami kebutuhan dan tujuan dari sudut pandang bisnis selanjutnya mengartikan pengetahuan ke dalam bentuk pendefinisian masalah pada data mining dan kemudian menentukan rencana serta strategi untuk mencapai tujuan.
2. Data Understanding, pada tahap ini diawali dengan mengumpulkan data, mendeskripsikan data, dan mengevaluasi kualitas data.
3. Data Preparation, merupakan tahapan yang membangun dataset akhir dari berupa data mentah. Beberapa hal yang akan dilakukan mencakup melakukan pembersihan data, melakukan pemilihan data, record dan atribut-atribut, dan juga melakukan transformasi terhadap data untuk dijadikan masukan dalam tahap pemodelan.
4. Modelling, pada tahap modelling ini secara langsung meliputi Machine Learning untuk penentuan teknik data mining, alat bantu data mining serta algoritma data mining.

5. Evaluation, pada tahapan ini melakukan dengan melihat tingkat performa dari pola yang dihasilkan oleh algoritma.
6. Deployment, pada tahapan ini melakukan dengan pembuatan laporan dan artikel jurnal menggunakan model yang didapatkan.

#### 2.4. *Naive Bayes*

Klasifikasi Naive Bayes adalah teknik klasifikasi berdasarkan teori Bayes. Metode klasifikasi dengan menggunakan probabilitas dan statistik ini pertama kali ditemukan oleh seorang ilmuwan Inggris bernama Thomas Bayes. Metode Naive Bayes Classifier dapat memprediksi apa yang akan terjadi di masa depan berdasarkan kejadian di masa lalu, oleh karena itu metode ini dikenal dengan teorema Bayes. Ciri utama dari metode ini adalah asumsi yang sangat kuat tentang keadaan dan peristiwa. (Mustafa, Moh & Angelina, 2018).

Rumus dasar yang digunakan oleh Bayes merupakan Teorema probabilitas bersyarat Bayes dalam (Muslehatin dkk, 2017) secara umum dapat di berikan sebagai berikut:

$$P(H|X) = P(X|H) P(H) / P(X) \quad (2,1)$$

Dimana :

X : Data kelas tidak diketahui

H : Hipotesis data X adalah spesifik kelas.

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior)

$P(H)$  : Probabilitas hipotesis H (probabilitas sebelumnya)

$P(X|H)$  : Probabilitas X didasarkan pada kondisi hipotesis H

$P(X)$  : Probabilitas X

Metode Naive Bayes, perhatikan bahwa proses klasifikasi diperlukan beberapa petunjuk untuk menentukan kelas mana yang cocok dengan sampel mana dianalisis. Oleh karena itu, metode *naive bayes* di atas dapat disederhanakan

Dengan:

$$P(H|X) = P(X|H)P(X)$$

## 2.5. **Support Vector Machine**

Vapnik meluncurkan Support Vector Machine (SVM) pertama, tepatnya pada tahun 1992, yang merupakan metode pembelajaran otomatis yang bekerja berdasarkan prinsip Structural Risk Reduction (SRM) yang ditujukan untuk menemukan jalur tinggi. Cara terbaik untuk memisahkan kedua kelas adalah dalam mode input. Metode ini menggunakan asumsi fungsi linier dalam ruang fitur berdimensi besar yang mengimplementasikan bias pembelajaran output dari teori pembelajaran statistik. SVM juga merupakan teknologi yang relatif baru dibandingkan dengan teknologi lain pada umumnya, sehingga memiliki kinerja yang lebih baik di berbagai bidang aplikasi. Namun keakuratan model yang dihasilkan oleh proses transisi SVM sangat bergantung pada fungsi kernel dan juga parameter yang digunakan. (Hilda Kusumahadi et al., 2019).

Support Vector Machine (SVM), yaitu teknologi yang relatif baru untuk melakukan prediksi baik dalam klasifikasi maupun regresi (Handayani, et al., 2020).

## **2.6. Twitter**

Twitter adalah salah satu sosial media yang banyak digunakan oleh sebagian besar pengguna internet. Pengguna harian twitter telah terhitung total mencapai 134 juta, naik dari tahun sebelumnya (Toy et al., 2021).

Twitter adalah layanan jejaring sosial *microblogging* gratis yang memungkinkan pengguna yang terdaftar menyiarkan pesan singkat yang disebut *tweets*. Anggota *Twitter* dapat menyiarkan *tweet* dan mengikuti *tweet* pengguna lain dengan menggunakan beberapa *platform* dan perangkat (Rustiana, dkk., 2017).

## **2.7. Website**

Menurut Sebok, Vermat dan tim (2018:70) “Website adalah kumpulan halaman yang saling terkait yang berisi berbagai elemen seperti dokumen dan gambar yang disimpan di web server. Aplikasi web adalah aplikasi di web server yang dapat diakses pengguna akses melalui browser. Paparan aplikasi Web sering mengumpulkan data pengguna dan informasi server.

Ruhi Abdullah (2018:1) menyatakan, “Website dapat diartikan sebagai kumpulan halaman yang berisi informasi data digital berupa teks, gambar, animasi, audio dan video, atau gabungan yang disediakan melalui koneksi internet yang dapat diakses. dan dilihat oleh semua orang di seluruh dunia. Dunia”.

## **2.8 PHP (*Hypertext Preprocessor*)**

"PHP berfungsi sebagai pemrosesan data di sisi server seperti yang diminta oleh klien menjadi informasi yang siap dilihat, serta menghubungkan aplikasi web ke database." Rawhi Abdullah (2018, hlm. 3).

Menurut Supono & Putratama (2018:1) "PHP adalah bahasa pemrograman yang digunakan untuk menerjemahkan basis kode program menjadikode yang dapat dimengerti mesin sisi server yang ditambahkan ke HTML."

## **2.9 Visual Studio Code (VS Code)**

Ummy Gusti Salamah (2021, h. 1) menjelaskan bahwa "Visual Studio Code (VS Code) adalah editor teks yang ringan dan andal yang diproduksi oleh Microsoft untuk sistem operasi lintas platform, yaitu tersedia juga untuk Linux, Mac dan Windows "VSCode Text Editor juga terbuka Sumbernya adalah tempat Anda dapat melihat kode sumber dan berkontribusi untuk pengembangannya. Ini juga menjadikan VSCode favorit bagi pengembang aplikasi karena pengembang aplikasi dapat berpartisipasi dalam proses pengembangan VSCode mulai sekarang."

## **2.10 XAMPP**

Darman Umagapi & Arisandy Ambarita (2018) menyatakan bahwa "XAMPP adalah freeware (perangkat lunak bebas), mendukung banyak sistem operasi, dan merupakan kumpulan dari perangkat lunak yang berbeda."

Mawadda dan Fawzi (2018) menyatakan bahwa "XAMPP adalah program yang didalamnya terdapat server MySQL, didukung dengan PHP sebagai bahasa

pemrograman untuk membuat website dinamis dan terdapat web server Apache yang dapat berjalan di berbagai platform seperti OS X, Windows, Linux, Mac dan Solaris”.

### **2.11 Database**

“Basis data adalah sistem terkomputerisasi yang tujuan utamanya adalah memelihara pemrosesan data atau informasi dan membuat informasi tersedia bila diperlukan.” Sukamto dan Saladin (2018: 43)

Menurut Hesananda et al. , (2017) “Database adalah wadah untuk menyimpan data dalam suatu sistem.Database juga dapat diartikan sebagai kumpulan data.Database disebut juga formal dan database juga dapat diartikan sebagai kumpulan data yang terintegrasi yang dapat cepat diproses, diambil dan dicari.

### **2.12 MYSQL**

Menurut Rusli, dkk (2019:5) “MySQL adalah sistem yang berguna untuk mengimplementasikan proses pengelolaan kumpulan struktur data (database), baik yang melibatkan proses pembuatan maupun pengelolaan database.”

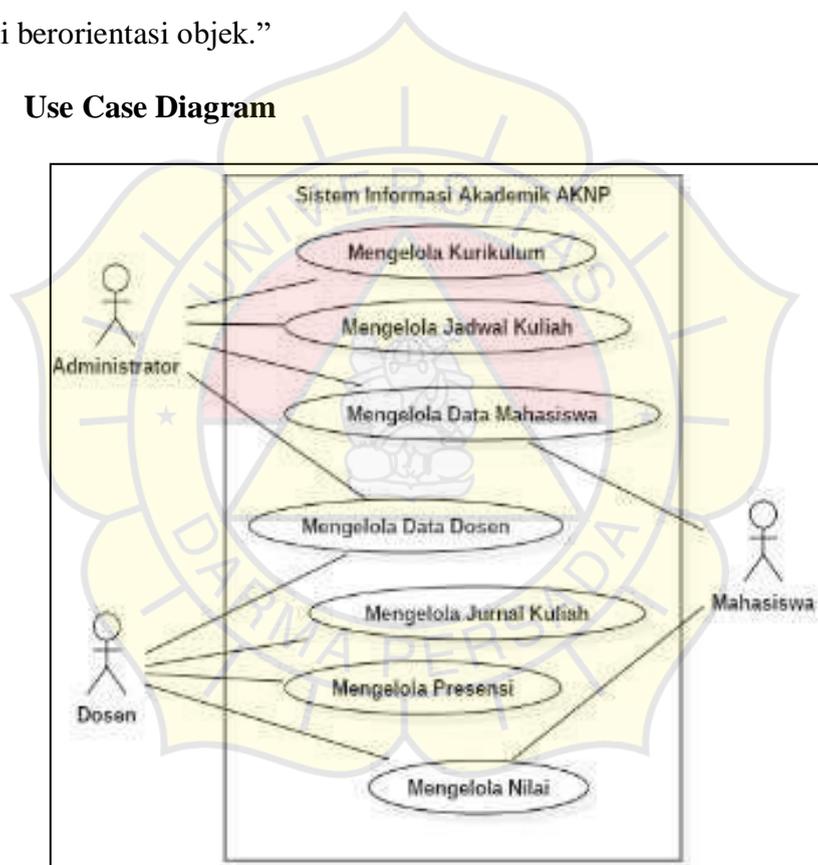
Ananda Putra Yoga, Sumijan & Mardison. (2019) membahas bahwa “MySQL merupakan implementasi dari Relational Database Management System (RDBMS) yang didistribusikan secara bebas di bawah GPL (General Public License), setiap pengguna dapat dengan bebas menggunakan MySQL namun dengan batasan bahwa software tersebut tidak boleh digunakan sebagai produk turunan komersial.

### 2.13 UML (*Unified Modeling Language*)

Menurut Sukamto dan Shalahuddin (2018:133), “UML adalah bahasa standar yang digunakan untuk menganalisis, mendesain, dan mendeskripsikan struktur program dalam pemrograman berorientasi objek.”

Menurut Kroenke et al. (2018) “UML adalah sekumpulan diagram, kerangka kerja, dan teknik untuk memodelkan dan merancang perangkat lunak dan aplikasi berorientasi objek.”

### 2.14 Use Case Diagram

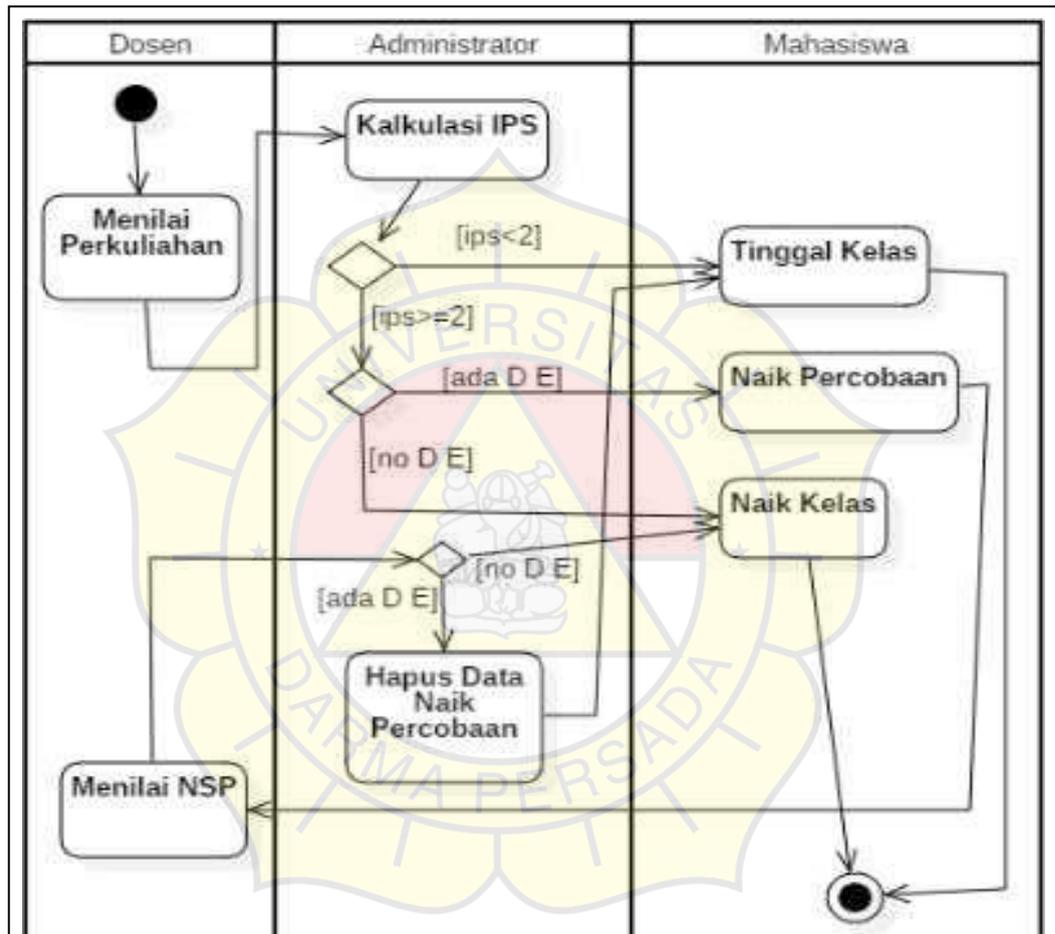


**Gambar 2. 2** Use Case Diagram  
(Sumber : Fu'adi, A., & Prianggono, A., 2022)

Menurut Sukamto dan Salahuddin (2018:155) “Use case diagram adalah model perilaku dari sistem informasi yang akan dibuat. Use case menggambarkan interaksi antara satu atau lebih aktor dengan sistem informasi yang akan

dibuat. Ujarnya use case itu digunakan untuk mengetahui fungsi-fungsi apa saja yang ada di dalam sistem informasi. Dialah yang berhak menggunakan fungsi-fungsi tersebut”.

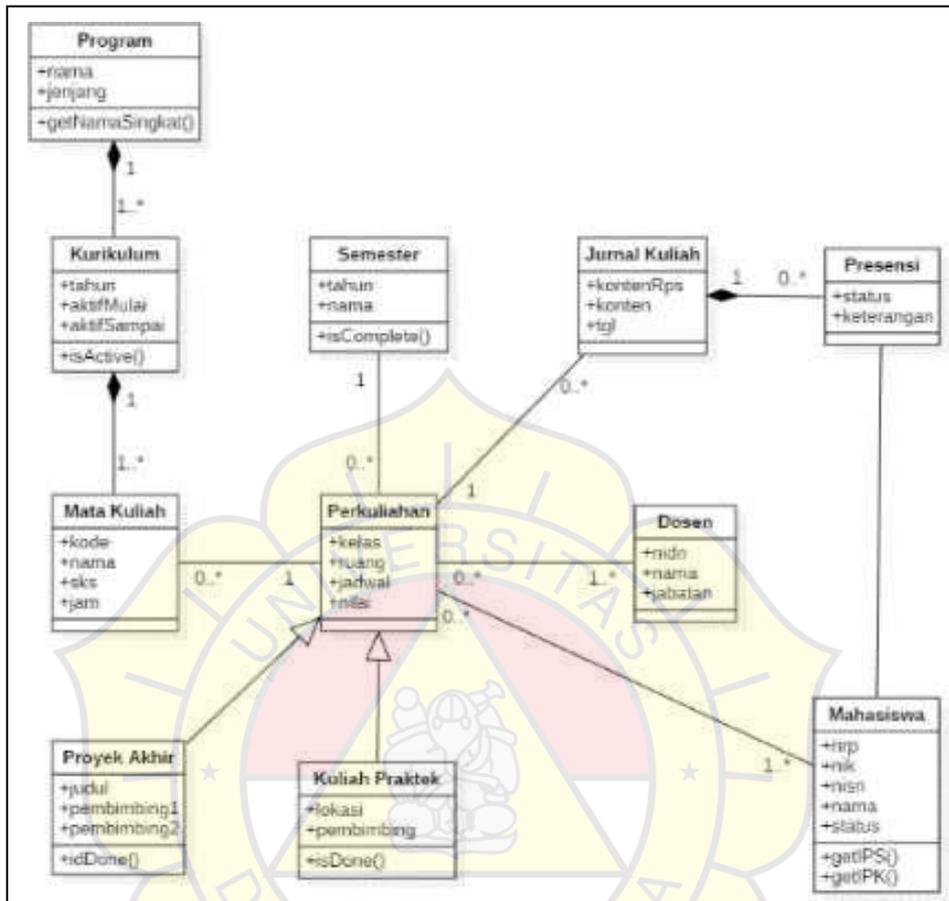
### 2.14.1. Activity Diagram



**Gambar 2. 3** Activity Diagram  
(Sumber : Fu'adi, A., & Prianggono, A., 2022)

Menurut Sukamto dan Shalahuddin (2018:161) “Activity diagram memvisualisasikan alur kerja atau aktivitas dari suatu sistem, proses bisnis, atau menu dalam suatu program”.

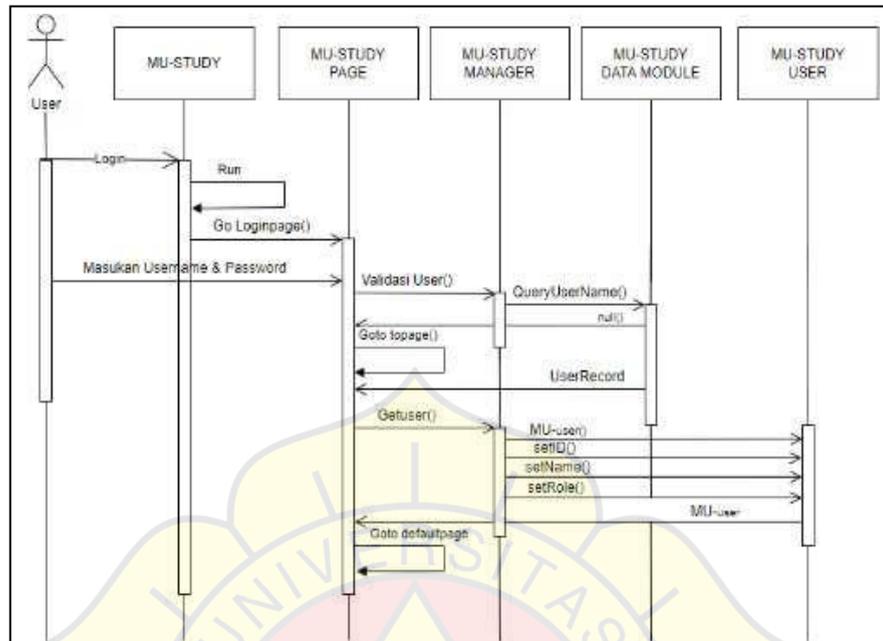
## 2.14.2. Class Diagram



**Gambar 2. 4** Class Diagram  
(Sumber : Fu'adi, A., & Prianggono, A., 2022)

Sukamto dan Shalauddin (2018:141) menyatakan bahwa “class diagram atau diagram kelas menggambarkan arsitektur suatu sistem dalam hal menentukan kelas mana yang akan dibuat untuk membangun sistem. Kelas berisi apa yang kita sebut atribut dan method atau operasi.

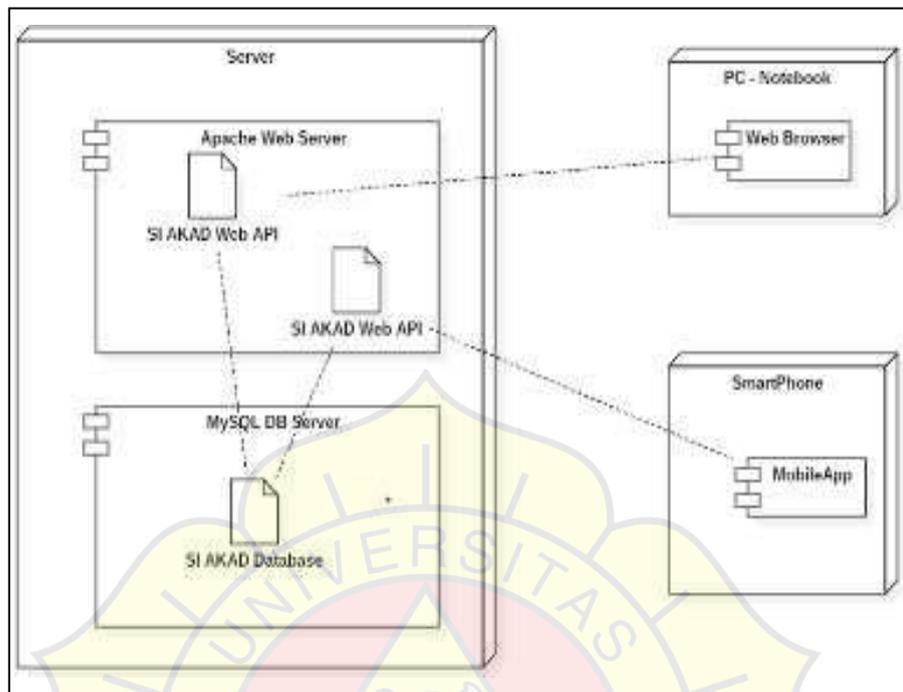
### 2.14.3. Sequence Diagram



**Gambar 2. 5** Sequence Diagram  
(Sumber : Fu'adi, A., & Prianggono, A., 2022)

Sukamto dan Shalahuddin (2018:165) mengatakan bahwa “Sequence diagram menggambarkan perilaku objek dalam use case dengan menggambarkan objek dan pesan yang dikirim dan diterima antar objek. Jadi untuk mendeskripsikan sequence diagram perlu diketahui objek-objek dalam use case beserta metode-metode yang dimilikinya ke dalam kelas yang dibuat dalam objek tersebut. Membuat diagram sequence juga diperlukan untuk melihat skenario pada use case”.

#### 2.14.4. Deployment Diagram



**Gambar 2. 6** Deployment Diagram  
(Sumber : Fu'adi, A., & Prianggono, A., 2022)

Yurindra (2017, h. 32) mengatakan bahwa “deployment diagram adalah gambaran umum dari berbagai proses dalam sistem operasi dan bagaimana di dalamnya”.

#### 2.14.5. Text Preprocessing

*Text pre-processing* merupakan tahap awal dalam *Text mining*, yaitu proses mempersiapkan teks agar teks dapat diproses lebih lanjut atau lebih dalam. *Text pre-processing* pada dasarnya mengubah informasi data menjadi bentuk atau format standar sebelum menerapkan berbagai metode pengambilan data ke dokumen yang akan diproses (Fitria, 2018).

## 2.15 Tinjauan Terhadap Penelitian Terdahulu

Pada Sub bab tinjauan terhadap penelitian terdahulu merupakan pemaparan hasil dari penelitian terdahulu yang dapat dijadikan acuan dalam penelitian yang penulis sedang lakukan. Penelitian terdahulu diharapkan dapat membantu penulis selama melakukan penelitian dan memberikan referensi bagi penulis. Berikut beberapa penelitian terdahulu yang menjadi referensi pada penulisan dan penelitian ini.

1. Pertama, Penelitian yang dilakukan oleh Evita Fitri, dkk. (2020) dengan judul penelitian “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine” yang mana penelitian ini mendapatkan hasil penelitian penggunaan algoritma Support Vector Machine yang memperoleh akurasi data sebesar 96,01% dan untuk algoritma Naive Bayes dengan nilai akurasi sebesar 94,16%, maka kesimpulan pada penelitian ini algoritma SPV lebih akurat dibandingkan dengan Naive Bayes.
2. Kedua, Penelitian yang dikarang oleh Mujaddid Izzul Fikri, dkk. (2020) yang berjudul “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter” memperoleh hasil penelitian yaitu pengaplikasian algoritma SPV mendapatkan tingkat akurasi sebesar 70,20% sedangkan jika dibandingkan dengan algoritma Naive Bayes akurasi yang diperoleh sebesar 73,65%, maka kesimpulannya adalah algoritma dengan akurasi tertinggi yaitu Naive Bayes.

3. Ketiga, Penelitian yang dilakukan oleh Tanthy Tawaqalia Widowati (2020) dengan penelitian “Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naive Bayes Dan Support Vector Machine” yang mana hasil penelitian disampaikan yakni metode Naive Bayes menghasilkan kinerja yang lebih baik dengan accuracy 91.48% dibandingkan dengan metode SPV 85,47%.
4. Keempat, Fajar Sodik Pamungkas (2021) melakukan penelitian yang berjudul “Analisis Sentimen dengan Naïve Bayes, SVM, dan KNN Untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 Pada Media Sosial Twitter” memperoleh hasil yaitu SVM memiliki akurasi tertinggi yaitu 90,01% sementara Naïve Bayes memperoleh nilai akurasi 79,20%.
5. Kelima, Ghulam Asrofi Buntoro (2021) Melakukan penelitian yang berjudul “Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine” Dalam penelitian ini juga dapat diketahui metode klasifikasi Support Vector Machine (SVM) lebih tinggi akurasinya untuk klasifikasi sentiment tweet HateSpeech Bahasa Indonesia dibandingkan metode klasifikasi NaïveBayes Classifier (NBC).
6. Keenam, Penelitian yang dilakukan Amanda Johnson, Michael Brown Comparative Study of Naive Bayes and Support Vector Machine for Sentiment Analysis on Twitter Data" (2016) dengan kesimpulan

menunjukkan bahwa SVM memiliki kinerja yang sedikit lebih baik daripada Naive Bayes dalam memprediksi sentimen pada data Twitter.

