

## BAB II TINJAUAN PUSTAKA

### 2.1. Penelitian Terkait

Penelitian terkait merujuk pada studi yang terkait dengan penelitian yang sedang berlangsung. Pentingnya penelitian sebelumnya sebagai pijakan pengetahuan serta fondasi bagi penelitian yang sedang dipelajari tidak dapat dipandang remeh. Berikut adalah penelitian sebelumnya yang menjadi acuan untuk penelitian “Implementasi Data Mining dalam Proses Analisa Sentimen Penggunaan Sunan (Sinau Temenan) *E-Learning* UMK sebagai Media Pembelajaran Menggunakan Metode *Naive Bayes Classifier*”.

Menurut (Ratnawati, 2018) dalam penelitiannya menggunakan algoritma *Naive Bayes Classifier* untuk menganalisis sentimen opini film Indonesia di Twitter. Data diambil dari 500 tweet yang dibagi menjadi sentimen positif dan negatif, dan model diuji menggunakan metode *5-fold cross-validation*. Hasil menunjukkan akurasi rata-rata 82,4%, dengan *precision* 83,8% dan *recall* 82,4%. Namun, penelitian ini memiliki kelemahan, seperti jumlah data training yang terbatas yang mempengaruhi akurasi model. Selain itu, adanya kata-kata yang sama dalam konteks positif dan negatif, seperti "lucu" dalam frasa "tidak lucu," menyebabkan kesalahan klasifikasi. Model juga mengalami kesulitan saat kalimat mengandung sentimen campuran, misalnya opini positif di awal yang diakhiri dengan opini negatif.

Menurut (Gunawan et al., 2018) dalam jurnalnya ini berfokus pada implementasi *Naive Bayes* untuk klasifikasi sentimen ulasan produk di *platform e-commerce*. Model klasifikasi melalui lima tahap, termasuk *crawling data*, *pre-processing*, dan kata yang dibobotkan menggunakan metode TF-IDF. Hasil pengujian menunjukkan bahwa model dengan data latih sebesar 90% dan data uji sebesar 10% menghasilkan akurasi sebesar 77.78% untuk klasifikasi tiga kelas (negatif, netral, dan positif). Di penelitiannya ini menyimpulkan bahwa algoritma *Naive Bayes* adalah algoritma yang sangat efektif untuk analisis sentimen pada ulasan produk online dalam skala besar. Namun, kelemahan penelitian ini berada pada nilai akurasi yang lebih rendah pada pengujian lima kelas sentimen, yang

hanya mencapai 59,33%, menunjukkan keterbatasan sistem pada klasifikasi lebih kompleks

Menurut (Atimi & Enda Esyudha Pratama, 2022) dalam jurnalnya ini bertujuan untuk mengembangkan model klasifikasi untuk sentimen *review* atau ulasan produk Lazada menggunakan algoritma *Multinomial Naïve Bayes*. Model yang dikembangkan kemudian dievaluasi memakai *confusion matrix* dan hasilnya menunjukkan kinerja yang baik dengan nilai akurasi, *precision*, *recall*, dan *f1-score* di atas 90%. Penelitian ini menyimpulkan bahwa algoritma *Multinomial Naïve Bayes* sangat efektif dalam proses klasifikasi sentimen ulasan produk aplikasi Lazada Indonesia ke dalam tiga kategori: positif, netral, dan negatif. Hasil ini menunjukkan bahwa metode ini cocok untuk diterapkan dalam analisis sentimen pada volume data yang besar seperti *e-commerce*. Namun, penelitian ini masih kurang memperhitungkan variabel dinamis seperti perubahan tren atau isu yang dapat mempengaruhi hasil analisis sentimen dalam jangka waktu panjang.

Menurut (Firdausy et al., 2023) penelitiannya ini membahas analisis sentimen terhadap reaksi peserta *e-learning* dengan menggunakan algoritma *machine learning*, yaitu *Naïve Bayes* (NB), *Support Vector Machine* (SVM), dan *Deep Learning* (DL). Proses analisis melibatkan ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah teks menjadi representasi *vektor numerik*, serta tahap *pre-processing*, *modeling*, dan evaluasi model. Data komentar peserta dibersihkan dan dibagi menjadi *data training* dan *data prediksi*. Penerapan model *Naïve Bayes* menghasilkan kinerja terbaik dengan akurasi 82,54%, diikuti oleh DL (79,95%) dan SVM (78,67%). Model NB berhasil mengklasifikasikan 70% komentar sebagai apresiasi, 16% sebagai saran, dan 14% sebagai kritik. Penelitian ini menyimpulkan bahwa *machine learning* efektif untuk analisis sentimen dalam evaluasi pelatihan dan merekomendasikan eksplorasi lebih lanjut dengan teknik lain seperti *ensemble learning* untuk meningkatkan akurasi. Selain itu, penelitian ini juga berkontribusi pada pemahaman umpan balik kualitatif dalam evaluasi pelatihan dan memperkaya literatur tentang *text mining* dalam analitik sumber daya manusia.

Menurut (Az-haari et al., 2024) dalam penelitiannya ini menggunakan algoritma *Naïve Bayes* untuk menganalisis sentimen tweet yang mengandung hashtag #BoikotStarbucks dan #BoikotProIsrael. Hasil penelitian menunjukkan bahwa metode *Naïve Bayes*, meskipun sederhana, sangat efektif dalam klasifikasi sentimen teks, menghasilkan akurasi tertinggi sebesar 71% pada model 90:10. Penelitian ini menyimpulkan jika algoritma *Naïve Bayes* dapat mengelompokkan tweet menjadi tiga kategori, yaitu positif, negatif, dan netral, dengan hasil klasifikasi yaitu cukup baik. Namun, penelitian ini akurasi masih cukup rendah, terutama pada kategori positif dan negatif yang memiliki presisi dan recall yang rendah.

Menurut (Apriani et al., 2024) dalam penelitiannya ini menganalisis sentimen pengguna terhadap TikTok sebagai alat pendidikan. Dengan 176 tanggapan dari kuesioner, algoritma *Naïve Bayes* digunakan untuk klasifikasi sentimen, menghasilkan akurasi 52,27%. *Precision* untuk ulasan positif adalah 83,52%, tetapi hanya 18,82% untuk ulasan negatif. Meskipun lebih efektif dalam mengidentifikasi ulasan positif, performa keseluruhan dianggap kurang memuaskan. Penelitian ini bertujuan memberikan wawasan bagi pengembang aplikasi pendidikan berbasis media sosial. Namun, penelitian ini memiliki kelemahan, yaitu akurasi model yang rendah sebesar 52,27%, menunjukkan keterbatasan keandalan model dalam klasifikasi sentimen. Selain itu, *precision* untuk sentimen negatif yang hanya mencapai 18,82% memperlihatkan bahwa model masih kesulitan dalam mengidentifikasi ulasan negatif secara akurat.

Menurut (Hadiwibowo et al., 2024) didalam penelitiannya ini menganalisis tingkat kepuasan pengguna Shopee dengan menggunakan metode *Naïve Bayes* pada 1.499 ulasan di *Google Play Store*. Dari analisis tersebut, ditemukan bahwa 40,3% pengguna memberikan rating positif terhadap Shopee, meskipun terdapat beberapa masalah yang perlu diperbaiki, terutama dalam pengalaman pengiriman. Metode yang dipakai dalam penelitiannya ini meliputi pengumpulan data, *pre-processing*, ekstraksi fitur, pelatihan model, dan pengujian model. Hasil analisis menunjukkan bahwa metode *Naïve Bayes* efektif dalam mengklasifikasikan ulasan menjadi positif dan negatif, dengan 604 ulasan positif dan 895 ulasan negatif, serta mencapai

akurasi sebesar 77%. Temuan ini diharapkan dapat memberikan wawasan bagi pengembang dalam meningkatkan kualitas layanan dan pengalaman berbelanja pengguna, serta membantu Shopee dalam merencanakan strategi pemasaran yang lebih efektif.

Menurut (Huda et al., 2024) dalam jurnalnya ini membahas penggunaan algoritma *Multinomial Naive Bayes* untuk klasifikasi sentimen dokumen, dengan fokus pada aplikasi *e-learning X*. Proses klasifikasi ini terdiri dari dua proses, yaitu pelatihan dan pengujian, di mana probabilitas setiap kelas dihitung berdasarkan data latih. Hasil analisis menunjukkan efektivitas algoritma ini dalam mengklasifikasikan dokumen, disertai dengan contoh perhitungan. Selain itu, artikel ini juga mengevaluasi *validitas* dan *reliabilitas* instrumen pengumpulan data menggunakan metode *Design Thinking*, serta menguji *prototipe* aplikasi *e-learning*. Penelitian menemukan bahwa desain antarmuka aplikasi kurang memadai, dengan rata-rata tingkat keberhasilan sebesar 9,2%. Rekomendasi diberikan kepada pengembang untuk lebih memahami kebutuhan pengguna dan menambahkan fitur seperti menu kontrol untuk orang tua, guna meningkatkan keterlibatan dalam proses belajar anak. Dengan menggabungkan analisis sentimen dan pendekatan *Design Thinking*, penelitian ini menunjukkan peningkatan signifikan dalam sentimen positif pengguna dan perbaikan pengalaman serta kepuasan pengguna, yang diharapkan dapat membantu pengembangan aplikasi *e-learning* di masa mendatang.

Menurut (Arofah et al., 2024), pada penelitiannya ini menganalisis sentimen terhadap sistem absensi berbasis web untuk pegawai ASN menggunakan algoritma *Naïve Bayes*. Model yang digunakan mengklasifikasikan ulasan menjadi positif dan negatif, dengan evaluasi kinerja yang menunjukkan akurasi sebesar 92,55%, presisi 98,13% untuk ulasan positif dan 81,65% untuk ulasan negatif, serta recall 91,30% untuk positif dan 95,70% untuk negatif. Hasil penelitian menunjukkan tingkat kepuasan yang tinggi di kalangan pegawai terhadap sistem absensi tersebut, meskipun terdapat potensi kesalahan dalam prediksi ulasan negatif yang perlu diperhatikan. Penelitian ini juga mengumpulkan data melalui kuesioner dari 500 responden, di mana 324 data berhasil diolah menggunakan *RapidMiner*. Temuan

ini memberikan wawasan mengenai pandangan pegawai ASN terhadap sistem absensi online dan menyarankan agar penelitian selanjutnya lebih fokus pada deteksi sentimen negatif serta pengumpulan umpan balik pengguna untuk meningkatkan kinerja model.

Menurut (Zain et al., 2024) dalam penelitiannya ini bertujuan untuk menilai dampak layanan sistem informasi terhadap kepuasan mahasiswa di Fakultas Teknik Universitas Nurul Jadid dengan menggunakan algoritma *Naïve Bayes*. Data dikumpulkan melalui kuesioner dari 316 mahasiswa di tiga program studi, dan hasil analisis menunjukkan akurasi model sebesar 94%, *precision* 92%, *recall* 95%, dan *f1-score* 93%. Temuan ini menunjukkan bahwa sistem informasi di fakultas tersebut sudah baik. Penelitian ini juga menyarankan agar universitas meningkatkan dan mengoptimalkan sistem informasi yang ada, serta memperhatikan atribut data dan pemilihan model yang tepat untuk penelitian selanjutnya. Metode penelitian yang digunakan meliputi identifikasi masalah, pengumpulan data, *pre-processing*, klasifikasi menggunakan *Naïve Bayes*, dan evaluasi model.

Penelitian ini akan berfokus pada implementasi algoritma *Naïve Bayes Classifier* untuk analisis sentimen pengguna terhadap *platform e-learning* Sunan (Sinau Temenan) yang dikembangkan sebagai media pembelajaran. Data akan diambil dari kuesioner dan wawancara dengan 700 responden dan diproses menggunakan *Google Collaboration* untuk efisiensi pemrosesan data dalam skala besar. Berbeda dengan penelitian terdahulu, seperti yang dilakukan oleh (Ratnawati, 2018), (Gunawan et al., 2018), dan (Apriani et al., 2024), penelitian ini akan mengatasi beberapa keterbatasan yang ada, termasuk keterbatasan *data training* yang mempengaruhi akurasi, kesalahan klasifikasi pada sentimen campuran, dan *precision* rendah pada sentimen negatif. Pembaruan metode dilakukan dengan mengoptimalkan teknik praproses data melalui *filtering* dan *stemming* serta memperbesar *dataset training* agar model lebih tangguh dalam menangani variasi sentimen. Dengan demikian, penelitian ini diharapkan mampu menghasilkan analisis sentimen yang lebih akurat dan efisien untuk memahami persepsi pengguna *e-learning* dalam konteks pendidikan.

## 2.2. Tabel Perbandingan Jurnal Penelitian Terkait

**Tabel 2. 1. Tabel Perbandingan Penelitian**

No	Judul	Metode	Data Set	Tools Google Collaboration	Akurasi	Presisi	Recall	F1-Score
1.	Implementasi Algoritma <i>Naive Bayes</i> untuk Analisis Sentimen Opini Film di Twitter ( Fajar Ratnawati, 2018)	<i>Naïve Bayes Classifier</i>	✓	x	✓	✓	✓	✓
2.	Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode <i>Naive Bayes</i> ( Gunawan et al., 2018)	<i>Naïve Bayes Classifier</i>	✓	x	✓	✓	✓	x
3.	Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia ( Atimi & Pratama, 2022)	<i>Multinomial Naive Bayes</i>	✓	x	✓	✓	✓	✓
4.	Analisis Sentimen Evaluasi Reaksi <i>E-Learning</i> Menggunakan Algoritma <i>Naive Bayes Support Vector Machine</i> dan <i>Deep Learning</i> (Firdausy et al., 2023)	<i>Naïve Bayes, SVM, Deep Learning</i>	✓	x	✓	✓	✓	x

(Bersambung)

Tabel 2.1. Lanjutan

5.	Analisis Sentimen Terhadap Boikot Brand Pro-Israel Pada Twitter Menggunakan Metode <i>Naïve Bayes</i> (Az-haari et al., 2024)	<i>Naïve Bayes Classifier</i>	✓	✓	✓	✓	✓	x
6.	<i>Sentiment Analysis of Using TikTok as a Learning Media Using the Naïve Bayes Classifiers Algorithm</i> (Elsa Apriani et al., 2024)	<i>Naïve Bayes Classifier</i>	✓	x	✓	✓	✓	x
7.	Analisa Tingkat Kepuasan Pengguna Shopee Berdasarkan Rating dan Ulasan <i>Google Play Store</i> Menggunakan <i>Naïve Bayes</i> (Hadiwibowo et al., 2024)	<i>Naïve Bayes Classifier</i>	✓	✓	✓	✓	✓	✓
8.	Analisa Sentimen <i>E-Learning X</i> Terhadap Antarmuka Pengguna Menggunakan Kombinasi <i>Multinomial Naive Bayes</i> dan Pendekatan <i>Design Thinking</i> (Huda et al., 2024)	<i>Multinomial Naive Bayes, Design Thinking</i>	✓	x	x	x	x	x
9.	Analisis Sentimen Pemakaian Sistem Absensi Berbasis <i>Web</i> Untuk Pegawai di Kabupaten Cirebon Menggunakan Metode Algoritma <i>Naïve Bayes</i> (Arofah et al., 2024)	<i>Naïve Bayes Classifier</i>	✓	x	✓	✓	✓	x

(Bersambung)

Tabel 2.1. Lanjutan

10	Klasifikasi Data Mining di Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Sistem Informasi Fakultas Teknik Universitas Nurul Jadid (Zain et al., 2024)	<i>Naïve Bayes Classifier</i>	✓	x	✓	✓	✓	✓
11.	Implementasi Data Mining dalam Proses Analisa Sentimen Penggunaan Sunan (Sinau Temenan) <i>E-Learning</i> UMK sebagai Media Pembelajaran Menggunakan Metode <i>Naive Bayes Classifier</i> (Alvina Gusti Pramita, 2024)	<i>Naïve Bayes Classifier</i>	✓	✓	✓	✓	✓	✓

## **2.3. Landasan Teori**

### **2.3.1 Pengertian Data Mining**

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Data mining merupakan proses iteratif dan interaktif untuk mengemukakan pola atau model baru, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar.

Data mining berisi pencarian tren atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya.

### **2.3.2 Pengertian *Sentiment Analysis***

Menurut Medhat et al (2014, 1093) analisis sentimen adalah suatu bidang yang berlangsung dalam penelitian berbasis teks. Analisis sentimen atau *opinion mining* adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas, dimana entitas tersebut dapat mewakili individu Wati (2016).

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif.

### **2.3.3 Pengertian *E-Learning***

*E-learning* adalah proses pembelajaran yang memanfaatkan teknologi digital, terutama internet, untuk memberikan fleksibilitas kepada siswa dalam mengakses materi pembelajaran kapan saja dan di mana saja. Menurut Husnussaadah (2021), *e-learning* memungkinkan siswa untuk belajar secara mandiri dengan dukungan bahan ajar yang dapat diakses melalui internet. *E-learning* memberikan manfaat seperti meningkatkan interaksi antara siswa dan

dosen, menyediakan sumber belajar yang tak terbatas, dan meningkatkan hasil pembelajaran. Selain itu, penelitian oleh Martín et al. (2021) menunjukkan bahwa penerapan e-learning selama pandemi COVID-19 berhasil memfasilitasi pembelajaran jarak jauh dengan berbagai model pedagogis yang efektif.

#### 2.3.4 Pengertian *Machine Learning*

*Machine learning* adalah sub bidang kecerdasan buatan (AI) yang bertujuan untuk memungkinkan komputer belajar dari data tanpa pemrograman eksplisit. Menurut Narkhede et al. (2021), machine learning melibatkan pengembangan algoritma yang menggunakan data historis untuk mengenali pola dan membuat prediksi. Contoh aplikasinya termasuk pengenalan gambar, analisis data kesehatan, dan sistem rekomendasi. Penelitian terbaru juga menyoroti bahwa machine learning berbasis neural network semakin populer dalam pengolahan data besar (big data), yang membuka peluang inovasi di berbagai bidang.

##### a. *Data Preprocessing*

Tahap *preprocessing* data merupakan proses dimana data akan diseleksi dan akan diproses. Setiap kata dalam ulasan akan diuraikan menjadi beberapa bagian yang lebih kecil agar maknanya menjadi lebih spesifik. Proses ini terdiri dari beberapa tahap yaitu, pada tahap preprocessing terdapat proses *case folding*, *filtering*, *correctiong slang word*, *stemming*, dan tokenizing.

##### b. *Term Frequency – Inverse Document Frequency (TF-IDF)*

*Term Frequency – Inverse Document Frequency (TF-IDF)* merupakan adalah metode untuk menghitung bobot kata dalam sebuah dokumen, melibatkan dua komponen utama yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Rumus untuk TF adalah:

$$tf_{ij} = \frac{f_a(i)}{\max f_a(j)}$$

Di sini,  $tf_{ij}$  adalah frekuensi kemunculan kata  $i$  pada dokumen  $j$  dibagi dengan total kata dalam dokumen tersebut. Perhitungan IDF pada

dasarnya adalah untuk mengukur seberapa unik suatu kata dalam seluruh dokumen, dengan rumus:

$$IDF_t = \log \left( \frac{D}{DF_t} \right)$$

Ket:

IDF = Hasil *Inversed Document Frecuency*

D = Jumlah dokumen

DF = Jumlah dokumen yang berisi term

Kemunculan suatu kata dalam sebuah dokumen menandakan tingkat kepentingannya dalam dokumen tersebut. Oleh karena itu, bobot antara kata tersebut dan dokumen akan semakin tinggi apabila jumlah dokumen lain yang mengandung kata tersebut sedikit dalam keseluruhan kumpulan dokumen. Setelah menghitung nilai TF dan IDF, langkah berikutnya adalah mengombinasikan kedua hasil ini untuk mendapatkan nilai TF-IDF. Adapun rumus TF-IDF adalah sebagai berikut:

$$TF - IDF = TF \times IDF$$

Secara umum, semakin sering kata muncul dalam dokumen, semakin tinggi nilainya, namun nilai ini disesuaikan agar kata yang sering muncul dalam banyak dokumen tidak memiliki bobot berlebihan.

### c. *Naïve Bayes Classifier*

Algoritma *Naive Bayes* adalah algoritma yang bekerja dengan mencari probabilitas tertinggi untuk mengklasifikasikan data ke dalam kategori tertentu. Algoritma ini dikenal akan kecepatannya dan efisiensinya dalam menangani dataset, baik dalam skala besar maupun kecil. Rumus untuk *Naive Bayes* adalah:

$$P_{(Y_k|X_a)} = \frac{P_{(Y_k)}P_{(X_a|Y_k)}}{P_{(X_a)}}$$

Keterangan:

$Y_k$  = Hipotesis data yang merupakan suatu kelas spesifik.

$X_a$  = Data dengan kelas yang belum diketahui.

$P_{(Y_k|X_a)}$  = Probabilitas hipotesis y berdasarkan kondisi a.

$P_{(Y_k)}$  = Probabilitas hipotesis y.

$P_{(X_a|Y_k)}$  = Probabilitas x berdasarkan kondisi pada hipotesis y.

$P_{(X_a)}$  = Probabilitas x

#### d. Evaluasi Model

*Confusion Matrix* adalah alat yang digunakan untuk mengevaluasi *accuracy*, *precisison*, dan *recall* dalam model, memastikan bahwa model bekerja sebagaimana yang diharapkan. Presisi menunjukkan seberapa tepat hasil yang diperoleh terhadap kebutuhan data, sementara recall menunjukkan keberhasilan sistem dalam menemukan data yang benar.

**Tabel 2. 2. Confusion Matrix**

	True	False
True (Positif)	TP (True Positif) <i>Correct result</i>	FP (False Positif) <i>Unexpected result</i>
False (Negative)	FN (False Negative) <i>Missing result</i>	TN (True Negative) <i>Correct absence of result</i>

Berikut adalah rumus dalam menghitung *precision*, *recall*, *accuracy*, dan *F1-Score*:

a. *Precision* :  $\frac{TP}{TP + FP}$

b. *Recall* :  $\frac{TP}{TP + FN}$

c. *Accuracy* :  $\frac{TP + TN}{TP + TN + FP + FN}$

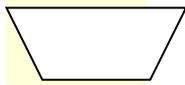
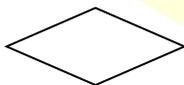
$$d. \text{ F1-Score} : \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Recall} + \text{Precision}}$$

Di sini, **TP** adalah jumlah data dengan nilai positif yang benar, **FP** adalah data negatif dengan nilai prediksi positif, **FN** adalah data positif dengan prediksi negatif, dan **TN** adalah data negatif yang diprediksi negatif. Model dikatakan baik jika memiliki nilai precision, recall, dan accuracy yang tinggi, mengindikasikan kemampuan model dalam memprediksi dengan akurat.

### 2.3.1 FOD (*Flow Of Document*)

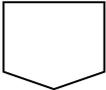
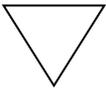
*Flow Of Document* (FOD) menggambarkan tentang gerakan dokumen yang dipakai di dalam suatu sistem. Bagan tersebut menunjukkan tentang dokumen apa saja yang bergerak di dalam suatu sistem, dan setiap kali dokumen tersebut sampai atau melalui suatu bagian tertentu akan dapat dilihat perlakuan apa saja yang telah diberikan terhadap dokumen tersebut. Adapun macam-macam simbolnya terlihat pada tabel 2.3.

**Tabel 2. 3. Simbol Flow of Document**

Simbol	Nama	Fungsi
	Dokumen	Dokumen input / output.
	Manual	Kegiatan manual.
	Connector	Berfungsi menyatakan sambungan dari suatu proses ke proses yang lainnya dalam halaman yang sama.
	Garis Alir	Menunjukkan arus dan proses.
	Logika/Decision	Suatu kondisi tertentu yang akan menghasilkan dua kemungkinan jawaban, ya dan tidak.
	Terminal	Menyatakan permulaan atau akhir suatu program.

(Bersambung)

Tabel 2.3. Lanjutan

	Offline Connector	Berfungsi seperti penghubung tetapi pada halaman yang berbeda.
	Offline Storage	Berfungsi untuk menunjukkan bahwa data dalam simbol ini akan disimpan ke suatu media tertentu.

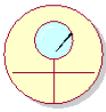
### 2.3.2 Unified Modelling Language (UML)

*Unified Modeling Language (UML)* adalah suatu bahasa permodelan untuk membangun perangkat lunak yang dibangun dengan menggunakan teknik pemrograman berorientasi objek. UML muncul karena adanya kebutuhan pemodelan visual untuk menspesifikasikan, menggambarkan, membangun dan dokumentasi dari sistem perangkat lunak (Muharni, 2021). Beberapa diagram grafis yang disediakan oleh UML, yaitu:

#### 1. *Bussiness Use Case Diagram*

Permodelan bisnis (*Bussiness Use Case Diagram*) merupakan suatu studi yang mempelajari tentang organisasi (Muharni, 2021). Saat melakukan pemodelan bisnis, kita menguji struktur organisasi, memperhatikan peranan-peranan di dalam organisasi, serta bagaimana mereka terhubung antara satu dengan lainnya. Juga menguji aliran kerja (*workflow*) dalam organisasi, proses utama di dalam organisasi, bagaimana mereka bekerja, seberapa efektif dan efisien cara kerja yang mereka lakukan. Demikian juga akan dilakukan pengujian entitas yang berada di luar organisasi, individu maupun perusahaan lain yang saling berhubungan dengan bisnis organisasi. Tabel 2.4. berikut adalah simbol atau notasi-notasi yang digunakan dalam diagram tersebut

**Tabel 2. 4. Notasi Bussiness Use Case Diagram**

Simbol	Nama	Keterangan
	Aktor Bisnis	Seseorang atau sesuatu yang ada di luar organisasi dan berinteraksi dengan organisasi yang terlibat dalam kegiatan bisnis organisasi.
	Pekerja Bisnis	Suatu peranan di dalam organisasi, bukan posisi. Seseorang boleh memainkan banyak peran tetapi memegang hanya satu posisi.
	<i>Use Case</i> Bisnis	Model yang digunakan untuk menggambarkan proses bisnis organisasi.

(Bersambung)

Tabel 2.4. Lanjutan

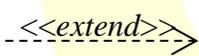
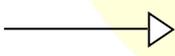
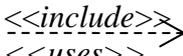
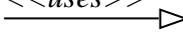
	Relasi Assosiasi	Relasi antara aktor bisnis atau pekerja bisnis dengan use case bisnis.
	Relasi Generalisasi	Digunakan ketika ada dua atau lebih aktor bisnis, pekerja bisnis, atau use case bisnis yang sangat serupa.

Sumber: (Muharni, 2021)

## 2. Use Case Diagram

*Use Case Diagram* mendeskripsikan interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibuat (Arwidiyarti et al., 2022). *Use case* digunakan untuk mengetahui fungsi apa saja yang ada di dalam sebuah sistem informasi dan siapa saja yang berhak menggunakan fungsi-fungsi itu. Notasi diagram tersebut dapat dilihat dalam tabel 2.5. berikut ini.

**Tabel 2. 5. Notasi Use Case Diagram**

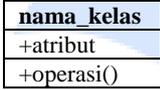
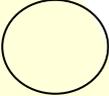
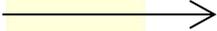
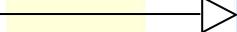
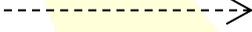
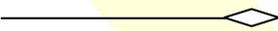
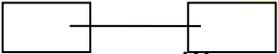
Simbol	Nama	Keterangan
 nama use case	<i>Use case</i>	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau aktor, biasanya dinyatakan dengan menggunakan kata kerja diawal nama frase nama use case.
 nama actor	Aktor / <i>actor</i>	Orang, proses, atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar itu <i>sendiri</i> , jadi meskipun aktor adalah gambar orang tapi aktor belum tentu merupakan orang, biasanya dinyatakan menggunakan kata benda awal frase nama aktor.
	Asosiasi / <i>association</i>	Komunikasi antara aktor dan <i>use case</i> yang berpartisipasi pada <i>use case</i> atau <i>use case</i> memiliki interaksi dengan aktor.
	Ekstensi / <i>Extend</i>	Relasi <i>use case</i> tambahan ke sebuah <i>use case</i> yang mana use case yang ditambahkan dapat berdiri sendiri meskipun tanpa <i>use case</i> tambahan itu
	<i>Generalization</i>	Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah <i>use case</i> . Fungsi yang satu adalah fungsi yang lebih umum dari lainnya.
 	<i>include / uses</i>	Relasi <i>use case</i> tambahan ke sebuah <i>use case</i> . <i>Use case</i> yang ditambahkan memerlukan <i>use case</i> ini untuk menjalankan fungsinya.

Sumber: (Arwidiyarti et al., 2022)

### 3. Class Diagram

Diagram kelas atau *class diagram* adalah gambaran visual tentang struktur sistem yang mencakup definisi kelas-kelas yang akan dibuat untuk membangun sistem (Muharni, 2021). *Class diagram* terdiri dari atribut dan operasi yang bertujuan untuk membantu pembuat program dalam membuat hubungan antara dokumentasi perancangan dan perangkat lunak yang sesuai. Dengan kata lain, *class diagram* membantu dalam merencanakan dan merancang sistem dengan menggambarkan kelas-kelas yang terlibat dan hubungan antara mereka. Simbol-simbol pada *class diagram* dapat dilihat pada tabel 2.6. berikut:

**Tabel 2. 6. Notasi Class Diagram**

Simbol	Nama	Deskripsi
	Kelas / <i>Class</i>	Kelas pada struktur sistem.
	Antarmuka / <i>Interface</i>	/ Sama dengan konsep <i>interface</i> dalam pemrograman berorientasi objek.
	Asosiasi / <i>Association</i>	/ Relasi antarkelas dengan makna umum, asosiasi biasanya juga disertai dengan <i>multiplicity</i> .
	Asosiasi berarah / <i>directed association</i>	Relasi antar kelas dengan makna kelas yang satu digunakan oleh kelas yang lain, asosiasi biasanya juga disertai dengan <i>multiplicity</i> .
	Generalisasi	Relasi antar kelas dengan makna generalisasi-spesialisasi (umum khusus).
	Kebergantungan / <i>dependency</i>	Relasi antarkelas dengan makna kebergantungan antarkelas.
	<i>Aggregation</i>	Relasi antar kelas dengan makna semua-bagian.
	Multiplisitas / <i>Multiplicity</i>	/ Jumlah suatu objek yang bisa berhubungan dengan objek lain.
0..1	Nol atau Satu	Bisa ada satu objek yang terkait atau tidak ada objek yang terkait.

(Bersambung)

Tabel 2.6. Lanjutan

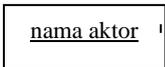
1	Hanya Satu	Ada satu objek dalam hubungan tersebut
0..*	Nol atau lebih	Tidak ada objek yang terkait atau ada banyak objek yang terkait.
1..*	Satu atau lebih	Setidaknya satu objek harus ada dalam hubungan tersebut. Dan bisa ada banyak objek yang terkait.
n	Hanya n (dengan $n > 1$ )	Jumlah objek dalam hubungan tersebut tidak spesifik, lebih dari satu.
0..n	Nol sampai n (dengan $n > 1$ )	Tidak ada objek yang terkait atau bisa ada banyak objek yang terkait. Jumlah objek yang terkait tidak terbatas.
1..n	Satu sampai n (dengan $n > 1$ )	Setidaknya satu objek harus ada dalam hubungan tersebut, dan bisa ada banyak objek yang terkait.

Sumber: (Muharni, 2021)

#### 4. Sequence Diagram

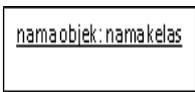
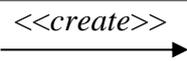
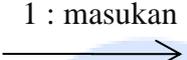
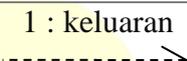
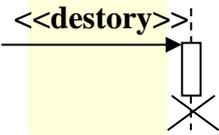
*Sequence diagram* menggambarkan bagaimana objek berinteraksi dalam suatu *use case* dengan memperlihatkan urutan waktu hidup objek serta pesan yang dikirim dan diterima di antara objek tersebut. *Sequence diagram* dibuat untuk setiap *use case* yang memiliki proses sendiri atau yang penting, dan semua interaksi pesan yang telah didefinisikan dalam *use case* tersebut sudah dicakup dalam *sequence diagram*. Semakin banyak *use case* yang didefinisikan, maka semakin banyak pula diagram urutan yang harus dibuat (Muharni, 2021). Adapun macam-macam notasi pada *Sequence Diagram* terlihat pada tabel 2.7. berikut.

Tabel 2. 7. Notasi Sequence Diagram

Simbol	Nama	Deskripsi
	Aktor	Orang, proses, atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi yang akan dibuat itu sendiri

(Bersambung)

### Tahap 2.7. Lanjutan

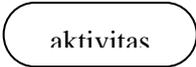
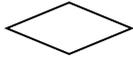
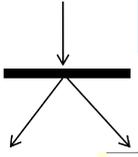
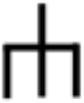
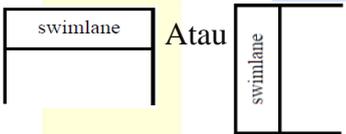
	Garis hidup / <i>lifeline</i>	Menyatakan kehidupan suatu objek
	Objek	Menyatakan objek yang berinteraksi pesan
	Waktu Aktif	Menyatakan objek dalam keadaan aktif dan berinteraksi pesan.
	Pesan tipe <i>creator</i>	Menyatakan suatu objek membuat objek lain, arah panah mengarah pada objek yang dibuat
	Pesan tipe <i>send</i>	Menyatakan bahwa suatu objek mengirimkan data / masukan / informasi ke objek lainnya, arah panah mengarah pada objek yang dikirim
	Pesan tipe <i>return</i>	Menyatakan bahwa suatu objek yang telah menjalankan suatu operasi atau metode menghasilkan suatu kembalian ke objek tertentu, arah panah mengarah pada objek yang menerima kembalian.
	Pesan tipe <i>destory</i>	Menyatakan suatu objek mengakhiri hidup objek yang lain, arah panah mengarah pada objek yang diakhiri.

Sumber: (Muharni, 2021)

### 5. Activity Diagram

Diagram aktivitas atau *activity diagram* menggambarkan aliran kerja atau aktivitas dari sebuah sistem atau proses bisnis atau menu yang ada pada perangkat lunak (Muharni, 2021). Adapun macam-macam notasi pada *Activity Diagram* terlihat pada tabel 2.8. berikut.

**Tabel 2. 8. Notasi *Activity Diagram***

Simbol	Nama	Deskripsi
	Status awal	Status awal aktivitas sistem, sebuah diagram aktivitas memiliki sebuah status awal.
	Aktivitas	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	Percabangan / <i>decision</i>	Asosiasi percabangan dimana jika ada pilihan aktivitas lebih dari satu.
	Penggabungan ( <i>Join</i> )	Asosiasi penggabungan yang mana lebih dari satu aktivitas digabungkan menjadi satu.
	Percabangan ( <i>Fork</i> )	Menunjukkan kegiatan yang dilakukan secara paralel.
	<i>Rake</i>	Merepresentasikan hierarki untuk menunjukkan perluasan aktivitas dalam diagram subaktivitas.
	Status Akhir	Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki sebuah status akhir.
	<i>Swimlane</i>	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas terjadi.

Sumber: (Muharni, 2021)

## 6. *Statechart Diagram*

*Statechart Diagram* adalah diagram yang menggambarkan urutan keadaan sesaat yang dilalui sebuah obyek, kejadian yang menyebabkan sebuah transisi dari satu *state* atau aktivitas kepada yang lainnya, dan aksi yang menyebabkan perubahan satu *state* atau aktivitas (Muharni, 2021). Adapun macam-macam notasi pada *Statechart Diagram* terlihat pada tabel 2.9 berikut.

**Tabel 2. 9. Notasi *Statechart Diagram***

Simbol	Nama	Deskripsi
	<i>Start (Initial State)</i>	<i>State</i> atau keadaan awal pada saat sistem mulai hidup
	<i>End (Final State)</i>	<i>State</i> keadaan akhir dari daur hidup suatu sistem
	<i>Event</i>	Kegiatan yang menyebabkan berubahnya status mesin
	<i>State</i>	Keadaan sistem pada waktu tertentu. <i>State</i> dapat berubah jika event tertentu yang memicu perubahan tersebut.

Sumber : (Muharni, 2021)

### 2.3.3 *Google Collaboration*

*Google Collaboration* adalah perangkat lunak yang umum digunakan untuk pengolahan data menggunakan *machine learning*. *Google Colaboratory*, atau *Google Collaboration*, merupakan *platform* berbasis *cloud* yang memungkinkan pengguna menjalankan kode *Python* dalam lingkungan *Jupyter Notebook* tanpa memerlukan konfigurasi perangkat keras lokal. Muflikhati et al. (2021) menyebutkan bahwa *Google Collaboration* sangat populer di kalangan peneliti dan pelajar karena dapat terintegrasi dengan *Google Drive*, menyediakan akses gratis ke GPU/TPU, dan mendukung kolaborasi *online* antar pengguna. *Platform* ini sering dimanfaatkan untuk eksperimen *machine learning*, analisis data, dan pelatihan model AI berkat kemampuannya menangani data dalam skala besar serta kompatibilitasnya dengan berbagai pustaka *Python*, seperti *TensorFlow* dan *PyTorch*.

### 2.3.4 *Jupyter Notebook*

*Jupyter Notebook* adalah perangkat lunak berbasis web yang digunakan untuk membuat dan berbagi dokumen yang berisi kode, visualisasi, dan teks naratif. *Platform* ini mendukung berbagai bahasa pemrograman, termasuk *Python*, dan sering digunakan untuk analisis data, pengembangan *machine learning*, dan dokumentasi kode. Menurut (Kluyver et al., 2016), *Jupyter Notebook* memungkinkan pengguna menjalankan kode secara interaktif dan melihat hasilnya langsung dalam satu antarmuka. Keunggulan lain dari *Jupyter Notebook* adalah kemampuannya untuk menampilkan visualisasi data secara langsung, membuatnya

menjadi alat yang sangat berguna dalam eksplorasi data. Selain itu, format *file* .ipynb yang digunakan oleh *Jupyter Notebook* memudahkan kolaborasi karena dokumen dapat dibagikan atau diunggah ke *platform* seperti GitHub.

### 2.3.5 *Streamlit*

*Streamlit* merupakan *framework open-source* yang dirancang untuk membangun aplikasi web interaktif berbasis data dengan menggunakan *Python*. *Streamlit* memungkinkan pengguna membuat aplikasi dengan cepat tanpa memerlukan pengetahuan mendalam tentang pengembangan web. *Streamlit* sangat populer di kalangan data *scientist* dan analis data karena kemudahannya dalam mengintegrasikan visualisasi data dan model *machine learning* ke dalam aplikasi web. Dengan hanya beberapa baris kode, pengguna dapat menghasilkan antarmuka yang intuitif untuk menampilkan grafik, tabel, atau masukan pengguna. *Streamlit* juga mendukung pembaruan data secara *real-time*, sehingga cocok digunakan untuk membangun *dashboard* atau *prototipe* analisis data.