

## **BAB II**

### **TINJAUAN PUSTAKA DAN DASAR TEORI**

#### **2.1 Tinjauan Pustaka**

Terdapat beberapa penelitian yang sudah ada dilakukan untuk mengklasifikasi terhadap *sentiment* review yang tersedia secara *online* yaitu;

Analisis Sentimen terhadap Review Aplikasi *PUBG Mobile* di *Google Play Store*. Pada penelitian ini menganalisis *sentiment* penguji kinerja BERT menggunakan dua model *pre-training* dengan data yang di gunakan yaitu ulasan pengguna untuk aplikasi *PUBG Mobile* di *Google Play Store* (Braja & Kodar, 2023).

Analisis Kemenangan Pemain Pada Permainan Player Unknown Battle Ground (*PUBG*) Menggunakan Metode CRISP-DM. Penelitian ini mengambil dataset *PUBG* dari *Kaggle*, dalam Dataset menggunakan metode CRISP-DM untuk melakukan pengolahan data hingga mendapatkan sebuah strategi menang dalam *game PUBG* (Ermanto, et al., 2022).

Menurut Wahid Desta Prastowo pada tahun 2023 untuk melakukan Analisis Sentimen Kepuasan Pemain *PUBG* Berdasarkan Komentar di Media Sosial. Penelitian ini Analisis *sentiment* dari penggemar serta pemain *game PUBG Mobile* melalui komentar di *fanpage Facebook*. Menggunakan statistik yang mendasar dalam pengenalan pola (*pattern recognition*).

Analisis *Sentiment* Terhadap *Game Multiplayer Online Ebattleroyale* Menggunakan Algoritma *K-Means* dan *Levensthein Distance*. Pada penelitian ini melakukan analisis *sentiment* dari komentar masyarakat terhadap fatwa dari MPU aceh. Data yang dianalisis yaitu data yang diambil dari kemenpora, komentar

Instagram, komentar X, dan komentar youtube (Pradipta, 2021).

Dalam Klasifikasi Sentimen Ulasan Aplikasi *Sausage Man* Menggunakan *VADER Lexicon* dan *Naïve Bayes Classifier*. Penelitian ini melakukan analisis *sentiment* untuk melihat pendapat positif, netral atau negatif dari pengguna *game Sausage Man*. Dengan pengumpulan data menggunakan *web scraping*, pelabelan data, *text preprocessing*, pembobotan dokumen, klasifikasi, dan evaluasi (Maulana, 2023).

Analisis yang akan dibuat mampu mengklasifikasi data *tweet* dari setiap topik di media sosial X. Dan menganalisis sentimennya menjadi 2 kategori yaitu positif dan negatif. Topik berita yang diambil adalah berita yang terkait tentang *PUBG Mobile* dan *Free Fire* di Indonesia.

**Tabel 2. 1 Perbandingan Penelitian**

Penulis	Topik	Metode	Hasil
(Braja & Kodar., 2023)	Implementasi <i>Fine-Tuning BERT</i> untuk analisis Sentimen terhadap <i>Review Aplikasi PUBG Mobile</i> di <i>Google Play Store</i> .	BERT	Peneliti ini melakukan analisis data yang digunakan adalah ulasan pengguna untuk aplikasi <i>PUBG Mobile</i> di <i>Google Play Store</i> . Dengan pelabelan berbasis skor dan pelabelan berbasis <i>Text blob</i> untuk menentukan efisiensi model. Hasil percobaan menunjukkan model <i>fine-tuned indoBERT</i> memiliki akurasi yang lebih baik dalam pelabelan data berbasis <i>Text blob</i> dengan akurasi tertinggi 94% pada <i>learning rate</i> 0.000002, <i>batch size</i> 32, jumlah <i>epoch</i> 5, dan waktu pelatihan 12 menit.
(Ermanto, et al., 2022)	Analisis Kemenangan Pemain pada Permainan <i>Player Unknown Battle Grounds (PUBG)</i> Menggunakan Metode <i>CRISP-DM</i> .	CRISP-DM	Penelitian menganalisis sentimen menggunakan <i>CRISP-DM</i> untuk memprediksi tingkat kemenangan pemain dalam pertandingan <i>PUBG</i> . Data yang diambil dari <i>Kaggle</i> yaitu dataset <i>PUBG</i> bertujuan mencari atribut-atribut pengaruh tingkat kemenangan pertandingan yang kemungkinan 90%.

Lanjutan Tabel 2.1 Perbandingan Penelitian

Penulis	Topik	Metode	Hasil
(Prastowo, 2023)	Analisis Sentimen Kepuasan Pemain <i>PUBG</i> Berdasarkan Komentar di Media Sosial.	Algoritma <i>Naïve Bayes Classifier</i>	Penelitian dilakukan dengan pertimbangan antara menghitung <i>trade-off</i> antara berbagai keputusan yang menggunakan, dan probabilitas yang muncul dalam keputusan tersebut. Hasil pengujian dengan percobaan pertama diperoleh sebesar 41,7% dan pada percobaan kedua dengan jumlah data yang lebih banyak meningkat menjadi 43,429%.
(Pradipta, 2021)	Analisis Sentimen Terhadap <i>Game Multiplayer Online Ebattleroyale</i> Menggunakan Algoritma <i>K-Means</i> dan <i>Levensthein Distance</i> .	Algoritma <i>K-Means</i> dan <i>Levensthein Distance</i> .	Hasil dari penelitian ini yaitu <i>sentiment</i> masyarakat terhadap fatwa lebih banyak yang negatif dari pada positif, dari tingkat akurasi yang didapat dari penggunaan algoritma <i>K-Means</i> yaitu 85% dengan nilai akurasi rendah dibandingkan dengan kombinasi algoritma diatas yaitu 96%.
(Maulana, 2023)	Klasifikasi <i>Sentiment</i> Ulasan Aplikasi <i>Sausage Man</i> Menggunakan <i>VADER Lexicon</i> dan <i>Naïve Bayes Classifier</i> .	<i>Naïve Bayes Classifier</i> .	Hasil dari penelitian ini menggunakan <i>VADER Lexicon</i> didapatkan hasil <i>sentiment</i> positif sebanyak 1089 ulasan (36.3%), untuk <i>sentiment</i> netral sebanyak 912 ulasan (30.4%), dan untuk <i>sentiment</i> negatif sebanyak 999 ulasan (33.3%). Evaluasi menggunakan pembagian 90% data latih 10% data uji menghasilkan akurasi sebanyak 75% <i>precision</i> 79% dan <i>recall</i> 75%. Untuk pembagian 80% data latih 20% data uji menghasilkan akurasi sebesar 73% <i>precision</i> 76% dan <i>recall</i> 73%. Kalimat positif lebih banyak ditemukan, namun akurasi masih dibawah 80%.
(Nuari, 2024)	Analisis <i>Sentiment</i> pada <i>X</i> Mengenai <i>Game Online PUBG Mobile</i> di Indonesia dan <i>Free Fire</i> . Menggunakan Metode <i>Naïve Bayes Classifier</i> .	<i>Naïve Bayes Classifier</i>	Hasil penelitian ini melakukan analisis <i>sentiment</i> positif dan negatif dari <i>game PUBG</i> dan <i>Free Fire</i> . Dengan tahapan pengumpulan data menggunakan <i>web scrapping auth token</i> dengan menggunakan <i>tweet Harvest</i> , pelabelan data, <i>text preprocessing</i> , klasifikasi, dan evaluasi. Hasil pelabelan data memakai <i>Vader Lexicon</i> , Klasifikasi memakai <i>Naïve Bayes classifier</i> . Model evaluasi memakai <i>Confusion matrix</i> .

## 2.2 Landasan Teori

### 2.2.1 *PUBG Mobile*

*PUBG Mobile* adalah sebuah permainan dengan *genre battle royal* yang memberikan izin 100 pemain dalam satu kali pertempuran (*match*) dengan koneksi Internet. Pemain memiliki pilihan individu (solo) pemain dapat untuk bermain sendiri melawan pemain 99 pemain lainnya, dua orang (duo) pemain dapat membentuk satu tim dengan rekannya untuk bertempur sebagai pasangan melawan pemain lain dan empat orang (skuat) pemain dapat membentuk sebuah tim empat orang untuk bertempur sebagai sebuah skuat. Kemudian pertandingan acak, pemain dapat bermain dengan orang di seluruh dunia yang di acak oleh system. Dan tersedia beberapa pelatihan seperti *arcade* dan sebagainya (Pane & Mahendra, 2023). *PUBG* permainan video dengan genre battle royale yang dikembangkan oleh *PUBG Corporation* dan *Tencent Games*. Memiliki grafis yang lebih realistis dan detail, dengan desain peta yang cenderung lebih besar dan luas. Tersedia di berbagai platform termasuk PC, konsol, dan ponsel.

### 2.2.2 *Free Fire*

Garena *Free Fire* adalah permainan battle royale yang dikembangkan oleh 111 Dots Studio dan diterbitkan oleh Garena untuk platform Android dan iOS. *Game* ini dirilis pada tahun 2017 dan sejak itu telah meraih popularitas yang signifikan di kalangan pemain *Mobile*. *Game* ini sudah dimainkan *gamer* pemula maupun professional, *game Free Fire* ini memiliki peminat rata-rata anak-anak yang berusia 15 tahun kebawah karena *game* ini memiliki permainan yang mudah dan fiturnya untuk anak-anak (Fadilah, 2023). *Free Fire* dengan pertandingan yang

melibatkan 50 pemain di peta yang semakin menyusut, *game* ini juga menawarkan mode permainan tambahan seperti *Clash Squad* yang merupakan mode 4v4 dengan focus strategi tim, Mengusung grafis yang lebih sederhana, namun tetap menarik, dengan desain peta yang lebih kecil untuk pertandingan yang lebih cepat. *Game* ini lebih di arahkan ke platform ponsel pintar dan tablet, dengan fokus pada pengalaman *Mobile*.

### **2.2.3 X**

X adalah layanan jejaring sosial memungkinkan penggunanya untuk mengirim dan membaca pesan *text* sampai 280 karakter dikenal dengan sebutan kicauan (*tweet*). X telah menjadi satu dari sepuluh situs paling sering dikunjungi di internet, dan dijuluki sebagai ‘pesan singkat dari internet’(Rante Tana, 2023).

### **2.2.4 Tweet Harvest**

*Tweet Harves* adalah sebuah *\_command-line tool* menggunakan *playwright* untuk mendapatkan *tweet* dari hasil pencarian X sesuai kata kunci tanggal yang ditentukan. *Tweet* yang didapatkan akan disimpan dalam file berbentuk CSV. Untuk menggunakan *tweet Harvest* membutuhkan *authorization token* yang didapatkan dengan *login* ke akun X kemudian mengcopy *authorization token* (Vincent, 2023).

### **2.2.5 Text Mining**

*Text mining* merupakan *Text mining* merupakan proses mendapatkan informasi dari *text*, informasi yang diperoleh berupa pola dan trend pembelajaran dan pola statistik. Bertujuan untuk mengestrak informasi yang berguna dari sekumpulan dokumen untuk mendukung proses penemuan pengetahuan dari kumpulan dokumen yang besar (Sukriadi, et al 2023). *Text mining* digunakan untuk menganalisis data dalam jumlah besar, *text mining* adalah proses menemukan

pengetahuan yang tidak diketahui melalui ekstraksi informasi otomatis dari sejumlah *text* yang tidak terstruktur ( (Nota, et al., 2022)).

### 2.2.6 Pre-processing

*Preprocessing* meliputi proses *Case Folding*, *Tokenizing Klining*, *Stop Removal*, *Stemming*. Tahap ini merupakan tahap untuk mempersiapkan data yang akan digunakan untuk proses klasifikasi. Tahap yang dilakukan meliputi proses *case folding* mengubah kata dari kapital menjadi huruf kecil, tahap berikutnya proses *tokenizing klining* melakukan pemotongan kalimat menjadi satu kata tunggal kemudian diberi tanda ‘,’ untuk pemisah kata dan melakukan *klining* membersihkan kalimat yang mengandung symbol url. Tahap proses selanjutnya *stopword removal* untuk menyeleksi kata ganti, kata sambung, jadi kata tersebut akan dihilangkan. Kemudian tahap terakhir proses *Stemming* yaitu mengubah suatu kata menjadi kata dasar atau menghapus imbuhan kata (Ananta, 2022).

#### 1. Case Folding

*Case Folding* yaitu merubah kata yang semula huruf kapital diubah menjadi *lowercase* agar seragam (Gifari, et al., 2022). Pada gambar 2.1 dan 2.2 merupakan contoh hasil case folding.

	Sentiment	Text Tweet
0	0	Jelek filmnya... apalagi si ernest tidak bermu...
1	0	Film king Arthur ini film paling jelek dari se...
2	0	Sepanjang film aku berkata kasar terus pada ba...
3	0	aku tidak suka fast and furious..menurutku kok...
4	0	Aku tidak tahu film nya, kamu bilang perang pe...

**Gambar 2. 1 Sebelum Case Folding (Gifari, et al., 2022).**

	Sentiment	Text Tweet
0	0	jelek filmnya apalagi si ernest tidak bermu...
1	0	film king arthur ini film paling jelek dari se...
2	0	sepanjang film aku berkata kasar terus pada ba...
3	0	aku tidak suka fast and furious menurutku kok...
4	0	aku tidak tahu film nya kamu bilang perang pe...

**Gambar 2. 2 Sesudah *Case Folding* (Gifari, et al., 2022).**

## 2. *Tokenizing*

Tahap ini adalah tahap dimana suatu kalimat dipecah menjadi kata-kata yang terpisah, biasanya yang menjadi acuan pemisah antar token adalah spasi atau tanda baca (Hendra & Fitriyani, 2020). Pada gambar 2.3 merupakan contoh hasil tokenizing.

Sebelum tokenisasi	Setelah tokenisasi
cegah <i>covid</i> ajak rajin cuci tangan	cegah – <i>covid</i> – ajak – rajin – cuci – tangan

**Gambar 2. 3 Sebelum dan Sesudah Tokenizing (Kaparang, et al., 2021)**

## 3. *Stopword*

Menghapus stopword seperti di, ke, dari, serta, atau, berikan, kalau, akan, dan lain - lain (Gifari, et al., 2022). Pada gambar 2.4 merupakan contoh hasil stopword.

Sebelum penghapusan <i>stopwords</i>	Setelah penghapusan <i>stopwords</i>
cegah covid anggota polsek laung tuhup ajak warga rajin cuci tangan depan mako	cegah covid ajak rajin cuci tangan
tindak pelanggar protokol kesehatan tidak pakai masker	tindak pelanggar protokol kesehatan tidak pakai masker

**Gambar 2. 4** Sebelum dan Sesudah Stopword (Kaparang, et al., 2021)

#### 4. Stemming

Proses mencari kata dasar dari sebuah kata. Seperti kata” menurut” setelah di stemming menjadi “turut” (Gifari, et al., 2022). Pada gambar 2.5 dan 2.6 merupakan contoh hasil stemming.

	Sentiment	Text Tweet
0	0	jelek filmnya si ernest bermutu sekali acti...
1	0	film king arthur film paling jelek seluruh cer...
2	0	sepanjang film aku berkata kasar terus bapaknya
3	0	aku suka fast and furious menurutku kok jelek...
4	0	aku tahu film nya kamu bilang perang perangan...

**Gambar 2. 5** Sebelum Stemming (Gifari, et al., 2022).

	Sentiment	Text Tweet
0	0	jelek film si ernest mutu sekali actingnya fil...
1	0	film king arthur film paling jelek seluruh cer...
2	0	panjang film aku kata kasar terus bapak
3	0	aku suka fast and furious turut kok jelek itu ...
4	0	aku tahu film nya kamu bilang perang perang pe...

**Gambar 2. 6** Sesudah Stemming (Gifari, et al., 2022).



### 2.2.7 Vader Sentimen

*Vader Sentimen* merupakan metode analisis *lexicon-based* berbasis *rule-based sentiment* analisis. Yang proses pelabelannya menggunakan *Vader Sentimen* dengan menentukan *compound score* dari setiap *tweet* tersebut (Indraprasta Ananta, 2022).

### 2.2.8 Python

*Python* adalah salah satu bahasa pemrograman yang dapat melakukan eksekusi sejumlah instruksi multi guna secara langsung (interpretatif) dengan metode orientasi objek. *Python* adalah bahasa pemrograman yang paling mudah dipahami. Python dibuat oleh programmer Belanda bernama Guido Van Rossum (Artaye, et al., 2022). *Python* yaitu bahasa pemrograman yang memiliki banyak *library* yang bisa digunakan untuk berbagai bidang ilmu, *python* juga merupakan bahasa yang dapat digunakan untuk berbagai aplikasi yaitu desktop, *Mobile*, *web*, analisis dan berbagai editor (Retnoningsih & Pramudita, 2020). Dukungan perpustakaan Python yang ditingkatkan (terutama *pandas*) telah membuatnya menjadi alternatif yang kuat untuk tugas analisis data (Sodik et al., 2020).

Perpustakaan *Pandas*, yang telah dikembangkan sejak tahun 2008, memiliki tujuan untuk mengatasi kesenjangan dalam ketersediaan alat analisis antara Python, sebagai bahasa komputasi umum dan ilmiah, dengan berbagai platform khusus dalam komputasi statistik, domain tertentu, dan bahasa basis data. Fokus *pandas* ini tidak hanya pada penyediaan fungsionalitas setara, tetapi juga pada implementasi fitur-fitur inovatif, seperti penyelarasan data otomatis dan indeks hierarki, yang tidak mudah dicapai secara terintegrasi dalam perpustakaan lain atau lingkungan

komputasi serupa. Meskipun awalnya dikembangkan untuk analisis data keuangan, tujuannya adalah agar pandas dapat membuat Python ilmiah lebih menarik dan menjadi lingkungan komputasi statistik yang praktis baik untuk akademisi maupun praktisi industri. Nama perpustakaan ini diambil dari "data panel," yang merupakan istilah umum untuk set data multidimensi yang sering digunakan dalam statistik dan ekonometrika (Wes Mckinney, 2011).

### 2.2.9 Sentiment Analysis

*Sentiment Analysis* adalah proses pengumpulan, pengolahan, dan evaluasi opini dari pendapat public atau produk tertentu (Akbar & Sugiharto, 2023) *Naïve Bayes Classifier*

*Naïve Bayes classifier* adalah sebuah metode algoritma untuk melakukan klasifikasi yang didasarkan pada konsep probabilitas. Metode ini merujuk pada teorema bayes sebagai dasar konsepnya. Keunggulan terletak pada kemampuan sederhana yang menghasilkan kinerja klasifikasi *text* yang baik (Muflih, et al., 2023). Rumus Naïve Bayes dapat dilihat pada rumus 2.

Persamaan dari teorema Bayes yakni:

Keterangan:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

X : Data dengan kelas tidak dikenal

H : Hipotesis dengan kelas X adalah kelas khusus

$P(H | X)$  : Probabilitas hipotesis H didasarkan pada kondisis X (*postering probability*)

$P(H)$  : Probabilitas hipótesis H (*prior probability*)

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

$P(X)$  : Probabilitas X

### 2.2.10 Confusion matrix

Confusion matrix adalah suatu teknik yang digunakan untuk menggambarkan kinerja algoritma klasifikasi, yang berbentuk matriks  $n \times n$  yang menunjukkan klasifikasi yang diprediksi dan aktual, dengan  $n$  merupakan jumlah kelas yang berbeda. Beberapa istilah yang terdapat dalam confusion matrix:

1. True Positif (TP), Jumlah prediksi yang benar untuk kelas positif.
2. False Negatif (FN), Jumlah prediksi yang salah untuk kelas positif.
3. True Negatif (TN), Jumlah prediksi yang benar untuk kelas negatif.
4. False Positif (FP), Jumlah prediksi yang salah untuk kelas negatif.

Akurasi mengukur seberapa tepat model dapat mengklasifikasi dengan benar, dihitung dengan membagi jumlah prediksi yang benar dengan jumlah total sampel.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.2)$$

Presisi menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model, dengan nilai presisi yang tinggi menunjukkan rendahnya nilai false positif. Presisi =  $\frac{TP}{TP+FP}$  (2.3)

Recall adalah rasio observasi positif yang diprediksi dengan benar dengan semua observasi di kelas yang sebenarnya, dan nilai recall yang tinggi menunjukkan rendahnya nilai false negative. Menggabungkan nilai presisi dan recall akan menghasilkan metrik tunggal yang dikenal sebagai F1-score, yang merupakan rata-rata harmonik tertimbang dari presisi dan recall (Noviana & Rasul, 2023).

$$F1 = \frac{2 \times R \times P}{R+P} \quad (2.4)$$

Precision adalah ukuran keakuratan antara informasi yang diminta oleh pengguna dan jawaban yang diberikan oleh sistem. Untuk menghitung nilai precision. Recall adalah metrik yang mengukur tingkat keberhasilan sistem dalam

menemukan kembali informasi  $Recall = \frac{TP}{TP+FN}$  (2.5)