



Received -
Accepted -
Published -

Open Access

DOI: -

Analisis Sentimen terhadap Ulasan Game PUBG dalam Data Pengguna Steam Tahun 2021: Pendekatan Big Data dengan PySpark

Evan Aprianto 1^a, M. Gilang Martiansyah M. 2^b,
Meinisa 3^c, Syifa Firnanda 4^d, Dara Cantika Dewi 5^e

^{a, b, c, d, e} Fakultas Sains, Program Studi Sains Data, Institut Teknologi Sumatera

* Corresponding E-mail: evan.121450024@student.itera.ac.id

Abstract: This research analyzes the sentiment of user reviews of the PlayerUnknown's Battlegrounds (PUBG) game on the Steam platform in 2021 with a big data approach using PySpark. Along with the rapid development of technology in the entertainment sector, especially games, Steam has become the most popular game distributor with 67 million active players per month and 7600 games that users can access. User reviews are an important source of data for understanding players' perceptions of game quality. Sentiment analysis, or opinion mining, is performed to group reviews into positive, negative, or neutral categories. By utilizing Spark, this research efficiently handles a data volume of 1644255 rows. The VADER (Valence Aware Dictionary and Sentiment Reasoner) method is used to create sentiment labels based on scores that reflect the level of positivity, negativity and neutrality of the text. These clustered labels are then used in the Naive Bayes classification model to make predictions about the most likely sentiment groups and Logistic Regression to model the relationship between text features and sentiment labels in making predictions. The research results show that the Logistic Regression model has higher accuracy (94%) compared to Naive Bayes (75%). This means that the performance provided by Logistic Regression is more able to correctly classify the sentiment of user reviews of the PUBG game compared to Naive Bayes so that Logistic Regression is more suitable for sentiment analysis in this context.

Keywords: *Sentiment analysis, PySpark, VADER, Naive Bayes, Logistic Regression*

Abstrak: Penelitian ini menganalisis sentimen ulasan pengguna terhadap game PlayerUnknown's Battlegrounds (PUBG) di platform Steam pada tahun 2021 dengan pendekatan big data menggunakan PySpark. Seiring dengan pesatnya perkembangan teknologi di bidang hiburan, khususnya game, Steam telah menjadi distributor game terpopuler dengan 67 juta pengguna aktif per bulan dan 7600 game yang dapat diakses pengguna. Ulasan pengguna menjadi salah satu sumber data penting untuk memahami persepsi pengguna mengenai kualitas game. Analisis sentimen, atau opinion mining, dilakukan untuk mengelompokkan ulasan menjadi kategori positif, negatif, atau netral. Dengan memanfaatkan Spark, penelitian ini menangani volume data sebanyak 1644255 baris secara efisien. Metode VADER (Valence Aware Dictionary and Sentiment Reasoner) digunakan untuk membuat label sentimen berdasarkan skor yang mencerminkan tingkat positivities, negativitas, dan netralitas teks. Label hasil klusterisasi ini kemudian digunakan dalam model klasifikasi Naive Bayes untuk membuat prediksi mengenai kelompok sentimen yang paling mungkin dan Regresi Logistik untuk memodelkan hubungan antara fitur-fitur teks dan label sentimen dalam membuat prediksi. Hasil penelitian menunjukkan bahwa model Regresi Logistik memiliki akurasi yang lebih tinggi (94%) dibandingkan dengan Naive Bayes (75%). Artinya, performa yang diberikan oleh Regresi Logistik lebih mampu mengklasifikasikan sentimen ulasan pengguna terhadap game PUBG dengan benar dibandingkan dengan Naive Bayes sehingga Regresi Logistik lebih cocok untuk analisis sentimen dalam konteks ini.

Kata Kunci : *Analisis sentimen, PySpark, VADER, Naive Bayes, Regresi Logistik*



Introduction / Pendahuluan

Latar Belakang

Teknologi khususnya dalam bidang hiburan sangat berkembang pesat. Hal ini ditandai dengan banyaknya platform hiburan berbasis online. Khususnya dalam dunia game. Game merupakan salah satu hiburan yang cocok bagi semua kalangan. Seiring berjalannya waktu dunia game sudah cukup canggih. Salah satu kemajuan game yang dapat kita rasakan adalah dengan game PC atau console mulai merambat ke mobile.

Distributor game yang cukup menarik perhatian adalah Steam. Saat ini steam merupakan distributor game yang paling banyak digunakan, hal ini ditandai dengan kenaikan jumlah pengguna yang mencapai 67 juta pengguna aktif perbulan [1], dan sekitar 7600 game dapat diakses pengguna melalui Steam [2].

Steam menyediakan konten video games yang dapat dengan mudah diakses. Untuk mengetahui bagaimana kualitas game tersebut, Steam menyediakan fitur ulasan yang dapat diisi oleh para pengguna game tersebut. ulasan ini berisi berbagai opini yang dapat berupa penilaian yang positif maupun negatif. Namun terdapat juga beberapa komentar yang tidak diketahui dengan jelas kategorinya.

Untuk mengetahui klasifikasi dari fitur ulasan dengan jelas, dapat dilakukan analisis sentimen. Analisis sentimen atau opinion mining adalah ilmu yang dapat memberikan analisis seputar pendapat, penilaian, emosi atau sentimen terhadap sebuah produk, layanan, atau topik [3].

Seiring dengan perkembangan teknologi dan meningkatnya jumlah pengguna internet, data yang dihasilkan oleh para pengguna semakin banyak dan beragam. Ulasan game di Steam, sebagai salah satu sumber data yang kaya, dapat memberikan wawasan yang berharga bagi pengembang game untuk meningkatkan kualitas dan kepuasan pengguna. Namun, jumlah data yang sangat besar dan beragam ini memerlukan pendekatan analisis yang cukup kompleks. Penelitian ini menggunakan Apache Spark yang dirancang untuk memproses data dalam skala besar dengan kecepatan tinggi. Dengan memanfaatkan Spark, penelitian ini dapat menangani volume data yang sangat besar secara efisien [4].

Untuk analisis sentimen pada ulasan Steam difokuskan pada ulasan game PlayerUnknown's Battlegrounds (PUBG) yang merupakan game dengan jumlah ulasan terbanyak dalam

ulasan Steam tahun 2021. Untuk melakukan analisis sentimen, digunakan metode VADER (Valence Aware Dictionary and Sentiment Reasoner) yang merupakan sebuah metode analisis sentimen dalam teks, terutama dalam konteks. VADER akan melakukan klusterisasi atau pembuatan label sentimen berdasarkan skor yang mencerminkan tingkat positività, negativitas, dan netralitas teks [5]. Label hasil klusterisasi ini kemudian digunakan dalam model klasifikasi Naive Bayes dan Regresi Logistik.

Model Naive Bayes menghitung probabilitas kata-kata untuk membuat prediksi mengenai kelompok sentimen yang paling mungkin, sementara Regresi Logistik memodelkan hubungan antara fitur-fitur teks dan label sentimen dalam membuat prediksi. Penelitian ini bertujuan untuk mengklasifikasikan persepsi pengguna berdasarkan distribusi sentimen dalam ulasan game PUBG dan mengevaluasi efektivitas serta akurasi penggunaan model Naive Bayes dan Regresi Logistik dalam PySpark. Distribusi sentimen ini diharapkan dapat membantu pengembang game PUBG dalam memahami lebih baik bagaimana permainan ini diterima oleh pengguna.

Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana perbandingan persepsi pengguna berdasarkan distribusi sentimen (positif, negatif, dan netral) dalam ulasan game PUBG di Steam tahun 2021 dengan metode VADER?
2. Bagaimana akurasi dari penerapan PySpark dan VADER dalam analisis sentimen ulasan game PUBG menggunakan model Naive Bayes dan Regresi Logistik?

Tujuan Penelitian

Tujuan dari penelitian ini meliputi:

1. Menganalisis hasil klasifikasi persepsi pengguna berdasarkan distribusi sentimen (positif, negatif, dan netral) dalam ulasan terhadap game PUBG di Steam tahun 2021 dengan metode VADER.
2. Mengevaluasi efektivitas penggunaan model Naive Bayes dan Regresi Logistik dalam PySpark untuk analisis sentimen ulasan game di Steam berdasarkan hasil akurasi.

Batasan Masalah

Penelitian ini memiliki beberapa batasan, yaitu:

1. Penelitian ini hanya fokus pada ulasan game PUBG di Steam tahun 2021 dan hanya akan menggunakan ulasan dalam bahasa Inggris.
2. Analisis sentimen dilakukan dengan metode VADER sebagai pembuat label distribusi sentimen (positif, negatif, dan netral), serta model Naive Bayes dan Regresi Logistik untuk membuat prediksi mengenai kelompok sentimen yang paling mungkin.

Theoretical Basis / Landasan Teori

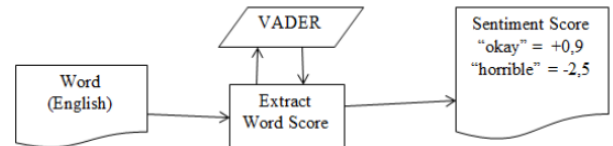
A. Analisis Sentimen

Analisis Sentimen adalah proses mengelompokkan polaritas dari teks dalam suatu dokumen, untuk menentukan apakah pendapat yang diungkapkan bersifat positif, negatif, atau netral. Analisis sentimen merupakan bidang yang telah berkembang sejak tahun 2003 dan menjadi bagian dari Text Mining. Fokus utamanya adalah melakukan analisis komputasi terhadap sentimen, emosi, opini, komentar, dan setiap bentuk ekspresi yang disampaikan melalui teks. Analisis sentimen terutama berfokus pada pengklasifikasian ulasan berdasarkan polaritas. Berdasarkan jenis klasifikasinya, analisis sentimen dibagi menjadi dua kategori utama: klasifikasi dokumen berdasarkan opini atau fakta, yang dikenal sebagai klasifikasi subjektivitas (subjectivity classification), dan klasifikasi dokumen menjadi positif atau negatif, yang dikenal sebagai analisis sentimen. Proses ini penting untuk mengidentifikasi dokumen yang mengandung opini dan menentukan apakah opini tersebut bersifat positif, negatif, atau netral [6].

B. Valence Aware Dictionary and Sentiment Reasoner

Valence Aware Dictionary and Sentiment Reasoner (VADER) adalah metode analisis sentimen yang dapat mengukur intensitas emosional berdasarkan kamus Lexicon yang tersedia. Diperkenalkan oleh C.J. Hutto dan Eric Gilbert pada tahun 2014, VADER menggunakan pendekatan human-centric yang menggabungkan analisis kualitatif dan validasi empiris dengan penilaian manusia [7]. Pendekatan leksikal menilai setiap kata dalam sebuah kalimat untuk menentukan keseluruhan skor sentimen [8]. Keunggulan utama VADER adalah tidak memerlukan pelatihan model dengan data berlabel. VADER menggunakan kamus yang berisi nilai setiap kata, menilai teks yang diproses berdasarkan

apakah kata tersebut positif, negatif, atau netral, dan menambahkan skor total (compound). VADER polarity detection telah menyediakan kamus berisi nilai dari setiap kata. Seperti yang ditampilkan pada **Gambar 1**, hasil Preprocess text akan dinilai berdasarkan lexicon apakah itu positif, negatif atau netral dan menambahkan skor total (compound).



Gambar 1. Alur penentuan polaritas dari suatu kalimat

C. Naive Bayes

Naive Bayes merupakan algoritma yang digunakan untuk menghitung probabilitas suatu kejadian berdasarkan informasi yang sudah diketahui sebelumnya [9]. Naive Bayes membuat asumsi bahwa semua fitur yang ada dalam data tidak saling bergantung atau independen. Algoritma ini menghitung probabilitas setiap kelas yang telah dihasilkan oleh metode VADER, yaitu positif, negatif, dan netral, untuk setiap ulasan. Kemudian Naive Bayes akan memilih kelas untuk ulasan tersebut berdasarkan probabilitas tertinggi sebagai prediksi. Algoritma yang didasarkan pada Teorema Bayes memiliki beberapa keuntungan seperti dapat dijalankan dengan komputasi yang sederhana sehingga mudah diimplementasikan dan efisien dalam waktu yang digunakan, dapat menangani data dalam jumlah besar serta tidak sensitif terhadap gangguan data [10].

D. Regresi Logistik

Regresi Logistik merupakan algoritma yang digunakan untuk memprediksi hasil dari suatu variabel berdasarkan satu atau lebih variabel independen [11]. Namun, berbeda dengan Naive Bayes, Regresi Logistik tidak membuat asumsi bahwa semua fitur tidak saling bergantung atau independen sehingga menghasilkan prediksi yang cenderung lebih baik. Beberapa jenis Regresi Logistik berdasarkan jumlah kategori dari variabel dependen yang ingin diprediksi adalah Regresi Logistik Biner (2 kategori), Regresi Logistik Multinomial (lebih dari 2 kategori), dan Regresi Logistik Ordinal (lebih dari 2 kategori dengan tingkatan) [12]. Berdasarkan jenis-jenis tersebut, penelitian ini menggunakan Regresi Logistik Multinomial dalam menangani hubungan antara variabel-variabel independen dalam analisis sentimen.

E. Apache Spark

Apache Spark adalah sebuah framework pemrosesan data terdistribusi yang dikembangkan oleh Apache Software Foundation. Spark dirancang untuk memproses data besar dengan menggunakan teknologi caching dalam memori dan eksekusi kueri yang dioptimalkan. Spark dapat digunakan untuk berbagai aplikasi, termasuk pemrosesan batch, kueri interaktif, analisis waktu nyata, dan machine learning. Spark dapat digunakan dalam berbagai bahasa pemrograman, termasuk Java, Scala, Python, dan R. Spark juga dapat dijalankan di berbagai platform, termasuk Hadoop YARN, Apache Mesos, Kubernetes, dan platform cloud seperti EC2 [13].

Pyspark adalah sebuah library Python yang digunakan untuk membangun aplikasi Apache Spark. Pyspark memungkinkan pengembang untuk menggunakan Python sebagai bahasa pemrograman untuk membangun aplikasi Spark. Pyspark juga memungkinkan pengembang untuk menggunakan berbagai fungsi dan metode yang disediakan oleh Spark, termasuk fungsi-fungsi pemrosesan data, fungsi-fungsi analisis, dan fungsi-fungsi machine learning [14].

Dalam konteks analisis sentimen, Pyspark dapat digunakan untuk menganalisis sentimen dalam teks menggunakan teknologi machine learning. Pyspark dapat digunakan untuk membangun model machine learning yang dapat memprediksi sentiment dalam teks berdasarkan analisis kata-kata dan frasa-frasa yang terkait dengan sentiment yang diinginkan. Pyspark juga dapat digunakan untuk menganalisis data besar dan memprosesnya secara cepat dan efisien menggunakan teknologi caching dalam memori dan eksekusi kueri yang dioptimalkan.

Method / Metode

a. Data Collection and Preparation

Dalam tahap pengumpulan data, dataset ulasan PUBG diunduh dari platform Steam. Dataset yang diunduh mencakup ulasan yang diberikan oleh pengguna PUBG di platform Steam pada tahun 2021. Dataset ini memiliki struktur yang kompleks dengan jumlah baris sebanyak 40.848.659 dan terdiri dari 23 kolom. Kolom-kolom ini mencakup berbagai informasi, seperti identifikasi aplikasi ('app_id' dan 'app_name'), ulasan pengguna ('review'), waktu pembuatan ulasan, rekomendasi

pengguna ('recommended'), dan informasi profil pengguna. Dengan jumlah baris yang signifikan, dataset ini mewakili sejumlah besar ulasan yang diberikan oleh pengguna PUBG di platform Steam. Ini memberikan sumber daya yang kaya untuk melakukan analisis sentimen yang mendalam dan menyeluruh terhadap pendapat pengguna terhadap permainan PUBG. Selanjutnya, data dimuat ke lingkungan Pyspark untuk pemrosesan yang efisien dalam skala besar. Dilakukan eksplorasi data untuk memahami struktur data, fitur yang tersedia, dan distribusi kelas sentimen. Proses pembersihan data juga dilakukan untuk memastikan kualitas data yang baik sebelum analisis lebih lanjut, termasuk penghapusan duplikat, penanganan nilai yang hilang, dan validasi data.

b. Implementasi PySpark

Selanjutnya, data dimuat ke lingkungan PySpark untuk pemrosesan yang efisien dalam skala besar. PySpark, sebagai bagian dari Apache Spark, menyediakan kemampuan untuk melakukan pemrosesan data terdistribusi yang memungkinkan analisis data besar dengan lebih cepat. Fungsi `spark.read.csv` digunakan untuk membaca dataset besar dengan efisien, dan operasi seperti filter, select, dan groupBy memudahkan manipulasi dan agregasi data. Dilakukan eksplorasi data untuk memahami struktur data, fitur yang tersedia, dan distribusi kelas sentimen. Proses pembersihan data juga dilakukan untuk memastikan kualitas data yang baik sebelum analisis lebih lanjut, termasuk penghapusan duplikat, penanganan nilai yang hilang, dan validasi data.

c. Text Preprocessing

Setelah memastikan kualitas data yang baik, langkah berikutnya adalah mempersiapkan teks mentah untuk analisis sentimen. Pertama, ulasan di-tokenisasi menjadi kata-kata individual. Selanjutnya, semua token diubah menjadi huruf kecil. Teks dinormalisasi dengan menghapus tanda baca dan karakter khusus. Kata-kata yang tidak penting dihapus untuk meningkatkan kualitas teks. Akhirnya, kata-kata diubah menjadi bentuk dasar atau dipangkas untuk konsistensi data.

d. Sentiment Analysis

Analisis sentimen dilakukan menggunakan library VADER, yang memiliki kamus dan aturan untuk menentukan sentimen positif, negatif, atau netral dari setiap ulasan.

e. Model Training

Dalam proses model training, dataset dibagi menjadi set pelatihan dan pengujian dengan proporsi tertentu, misalnya 80% untuk pelatihan dan 20% untuk pengujian. Proporsi ini memastikan bahwa model dilatih dengan data yang cukup untuk mempelajari pola sentimen dengan baik, sementara juga memberikan data pengujian yang cukup untuk mengevaluasi kinerja model dengan akurat. Setelah dataset dibagi menjadi set pelatihan dan pengujian untuk melatih dan menguji model, langkah selanjutnya adalah ekstraksi fitur dari teks ulasan menggunakan metode TF-IDF yang mengubah teks menjadi representasi numerik yang dapat diproses oleh model. Setelah itu, model klasifikasi sentimen, seperti Naive Bayes dan Logistic Regression, dilatih menggunakan data pelatihan untuk mempelajari pola sentimen dari fitur-fitur yang diekstraksi. Proses pelatihan ini memungkinkan model untuk memahami hubungan antara fitur-fitur numerik dan label sentimen, sehingga dapat digunakan untuk mengklasifikasikan sentimen ulasan yang belum pernah dilihat sebelumnya dengan akurasi yang baik.

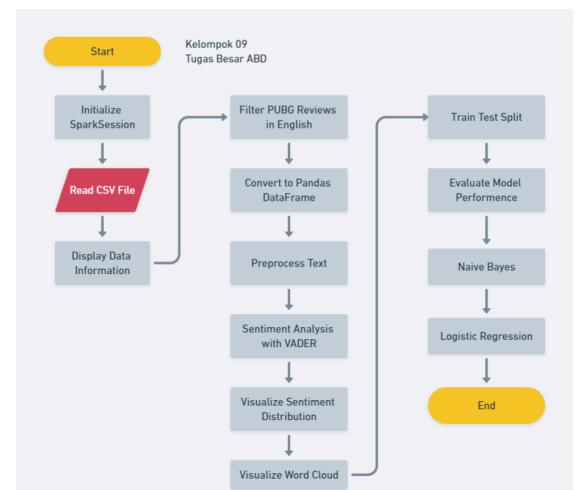
f. Model Performance Evaluation

Untuk mengevaluasi kinerja model, berbagai matrik evaluasi seperti akurasi, presisi, recall, dan F1-score digunakan. *Confusion matrix* juga dibuat untuk mengevaluasi klasifikasi model pada set pelatihan dan pengujian, yang membantu dalam memahami seberapa baik model dapat mengklasifikasikan label sentimen dengan benar. Selain itu, dilakukan validasi silang untuk memastikan konsistensi kinerja model dan mencegah overfitting. Proses cross-validation ini membagi dataset menjadi beberapa subset, di mana model dilatih dan diuji secara bergantian pada subset tersebut untuk memberikan estimasi kinerja yang lebih akurat. Dengan menggunakan berbagai metode evaluasi ini, dapat dipastikan bahwa model dapat memberikan prediksi sentimen yang konsisten dan akurat terhadap ulasan yang belum pernah dilihat sebelumnya.

g. Model Comparison

Dalam proses perbandingan dan pemilihan model, hasil kinerja antara model Naive Bayes dan Logistic Regression dibandingkan menggunakan matrik evaluasi yang telah ditetapkan, seperti akurasi atau F1-score. Tujuan dari perbandingan ini adalah untuk mengidentifikasi model yang paling sesuai dengan dataset dan dapat memberikan prediksi sentimen yang paling akurat. Setelah perbandingan dilakukan, model yang terbaik dipilih berdasarkan kinerjanya dan kemudian digunakan untuk menerapkan analisis sentimen pada ulasan PUBG. Dengan memilih model yang paling optimal, diharapkan hasil analisis sentimen dapat memberikan pemahaman yang lebih baik tentang sentimen pengguna terhadap permainan PUBG.

Hasil keseluruhan metode secara umum dalam bentuk diagram alir dapat dilihat pada **Gambar 2**.



Gambar 2. Flowchart Penelitian

Hasil dan Pembahasan

Dalam pengaturan eksperimental penelitian berjudul "Analisis Sentimen Terhadap Ulasan Game PUBG dalam Data Pengguna Steam Tahun 2021: Pendekatan Big Data dengan PySpark," terdapat beberapa langkah yang diikuti.

A. Dataset

Data yang digunakan pada penelitian ini adalah tipe data yang berasal dari scraping komentar para pengguna yang dilakukan di platform Steam. Namun, pada penelitian ini didapatkan data sekunder yang

bersumber dari website kaggle, sehingga kami sebagai pihak eksternal memperoleh data hasil scraping dari developer.

B. Dataset Preprocessing

Untuk membangun model pada penelitian ini dan melakukan perbandingan dengan evaluasi dengan teknik lain, langkah pertama akan melibatkan pra pemrosesan dataset. Tujuannya adalah untuk menghilangkan karakter yang tidak diperlukan dari teks asli serta menormalisasikan huruf yang seringkali ditulis secara bergantian yang dapat menyebabkan penyebaran data [15]. Langkah-langkah pra-pemrosesan pada penelitian ini sebagai berikut :

1. Memilih data yang berasal dari aplikasi "PLAYERUNKNOWN'S BATTLEGROUNDS" dengan ketentuan bahasa yang digunakan adalah bahasa Inggris dari aplikasi PUBG.
2. Mengubah huruf teks menjadi huruf kecil untuk memudahkan pembacaan dan memberikan standarisasi yang sama pada setiap kalimat.
3. Menghapus tanda baca, seperti tanda petik, dengan menggunakan fungsi punctuation.
4. Menghapus stopwords untuk menghilangkan kata-kata yang tidak penting, sehingga kalimat yang akan diproses menjadi lebih ringkas dan bernilai tinggi.

C. Split Dataset

Sebelum melakukan klasifikasi, data akan dibagi menjadi dua bagian utama, yaitu data latih dan data uji, yang masing-masing memiliki peran penting dalam pengembangan dan evaluasi model yang ilustrasinya dapat dilihat sebagai berikut :

```
X_train, X_test, y_train, y_test =
train_test_split(pubg_reviews_pd['review'],
pubg_reviews_pd['sentiment'], test_size=0.2, random_state=42)
```

Gambar 3. Split dataset

Pada **Gambar 3**. Didapatkan bahwa data latih biasanya terdiri dari sebagian besar dataset yang digunakan untuk melatih model agar dapat mengenali pola dan hubungan dalam data. Proses pelatihan ini melibatkan penggunaan algoritma pembelajaran mesin yang berupaya meminimalkan kesalahan prediksi dengan mengoptimalkan parameter model. Setelah model

dilatih, data akan diuji yang merupakan bagian terpisah dari dataset yang tidak digunakan selama proses pelatihan, digunakan untuk mengukur kinerja model. Data uji berfungsi sebagai evaluasi akhir untuk menilai kemampuan generalisasi model, yaitu seberapa baik model dapat membuat prediksi akurat pada data yang belum pernah dilihat sebelumnya. Pemisahan dataset ini penting untuk memastikan bahwa model yang dihasilkan tidak hanya bekerja baik pada data yang telah dilatih, tetapi juga dapat diterapkan secara efektif pada data baru.

D. Train and Test Model

Dalam penelitian ini, data latih digunakan untuk melatih dua model yang berbeda, yaitu model Multinomial Naive Bayes dan model Logistic Regression, yang ilustrasinya dapat dilihat sebagai berikut :

```
# Define the pipelines
pipeline_nb = Pipeline([
    ('tfidf', TfidfVectorizer(max_features=5000)),
    ('clf', MultinomialNB())
])

pipeline_lr = Pipeline([
    ('tfidf', TfidfVectorizer(max_features=5000)),
    ('clf', LogisticRegression(max_iter=1000))
])

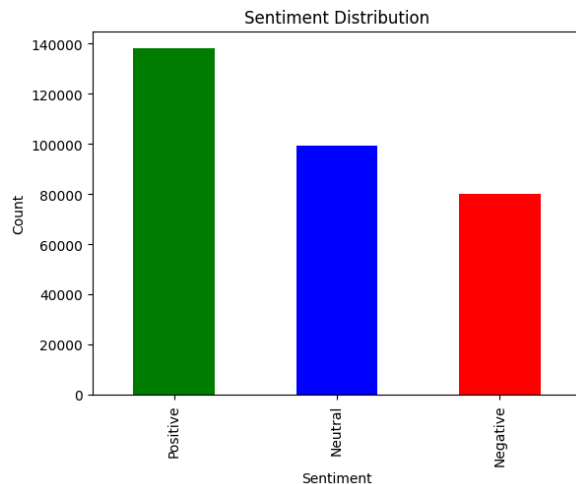
# List of pipelines for iteration
pipelines = [pipeline_nb, pipeline_lr]
pipeline_names = ['Naive Bayes', 'Logistic Regression']
```

Gambar 4. Train and Test Model

Pada **Gambar 4**. dapat dijelaskan bahwa model Multinomial Naive Bayes dan model Logistic Regression kemudian akan memprediksi atau menguji kinerja dengan data uji. Multinomial Naive Bayes dipilih karena memiliki keunggulan dalam menangani data dengan fitur atau ukuran yang besar, karena asumsi independensinya memungkinkan perhitungan yang efisien dan cepat, bahkan ketika jumlah fitur yang sangat banyak. Disisi lain, Logistic Regression dipilih karena cenderung memberikan hasil yang lebih baik dalam tugas klasifikasi teks dan probabilitas, karena dapat menangani hubungan linier antara fitur dan kelas target, serta memberikan interpretasi yang jelas tentang pengaruh setiap fitur terhadap probabilitas kelas. Dengan menerapkan kedua model ini, penelitian dapat membandingkan dan mengevaluasi performa masing-masing model dalam konteks klasifikasi teks, memastikan bahwa hasil yang diperoleh tidak hanya akurat tetapi juga efisien.

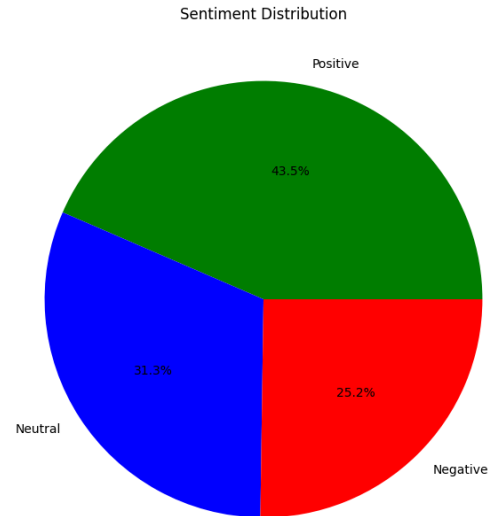
E. Diagram and Word Cloud

Pada bagian ini, akan dijelaskan secara rinci tentang hasil visualisasi data dalam bentuk word cloud dan diagram yang menggambarkan hasil ulasan aplikasi PUBG. Word cloud dan diagram digunakan untuk memberikan gambaran yang lebih jelas tentang pola dan tren yang ada dalam ulasan para pengguna aplikasi PUBG, untuk ilustrasinya dapat dilihat sebagai berikut :



Gambar 5. Diagram Barchart Distribution

Pada **Gambar 5**. didapatkan bahwa diagram barchart digunakan untuk menganalisis distribusi kata yang muncul dalam ulasan aplikasi PUBG berdasarkan kelas sentimen “Positif”, “Negatif”, dan “Netral”. Hasil tersebut menunjukkan bahwa kata-kata dengan sentimen positif lebih banyak muncul dalam ulasan dibandingkan dengan kata-kata sentimen negatif dan netral. Kata-kata sentimen positif muncul sekitar 138129 kata, menunjukkan bahwa banyak pengguna memberikan ulasan yang positif tentang aplikasi PUBG, menghasilkan kepuasan konsumen terhadap berbagai aspek dari aplikasi PUBG. Kata-kata sentimen Netral muncul sekitar 99379 kata, ulasan netral biasanya menghasilkan pendapat yang tidak condong ke arah yang sangat baik atau yang sangat buruk, mungkin dapat menunjukkan pengalaman pengguna yang biasa-biasa saja. Dan untuk kata-kata sentimen Negatif muncul sekitar 80135 kata, meskipun jumlahnya lebih sedikit, kata-kata negatif penting untuk dianalisis lebih lanjut untuk perbaikan atau peningkatan dari aplikasi PUBG.



Gambar 6. Pie Chart Distribution

Pada **Gambar 6**. didapatkan bahwa diagram pie chart memberikan representasi proporsi kata yang muncul dalam ulasan aplikasi PUBG dalam bentuk persentase, yang dibagi menjadi kelas “Positif”, “Negatif”, dan “Netral”. Pada hasil kata Positif didapatkan bahwa 43,5% dari total kata yang dianalisis, yang menunjukkan bahwa hampir setengah dari ulasan memiliki sentimen positif, yang dapat menghasilkan kepuasan yang cukup tinggi di antara pengguna aplikasi PUBG. Pada hasil kata Negatif menghasilkan 25,5% dari total kata, yang menunjukkan bahwa sekitar seperempat dari ulasan memiliki sentimen negatif, yang menyoroti area yang mungkin memerlukan perhatian dan perbaikan dari pengembang aplikasi. Dan untuk hasil kata Netral menghasilkan 31,3% dari total kata, yang menunjukkan bahwa sepertiga dari ulasan adalah netral yang biasanya mencerminkan pendapat yang objektif.

Tabel 1. Total sentimen

Positive	43.5%
Neutral	31.3%
Negative	25.2%



Gambar 7. Word cloud ulasan positif

Pada **Gambar 7**. didapatkan bahwa hasil word cloud bagian kata “Positif” terlihat bahwa kata-kata yang paling sering muncul adalah kata “best”, “great”, “better”, “fun”, dan “friend”. Kata “best” mengartikan bahwa pengguna menganggap sesuatu dalam game PUNGB sebagai yang terbaik baik itu fitur atau yang lainnya. Pada kata “great” mengartikan bahwa pengguna memiliki pengalaman yang sangat baik dengan aspek-aspek tertentu dalam game. Pada kata “better” mengartikan bahwa pengguna merasa ada peningkatan atau aspek yang lebih baik dari sebelumnya. Pada kata “fun” mengartikan bahwa game PUBG memberikan hiburan dan kesenangan bagi para pengguna. Pada kata “friend” mengartikan bahwa aspek sosial dari game seperti bermain dengan teman yang sangat dihargai oleh pengguna. Secara keseluruhan, kehadiran kata-kata dalam word cloud bagian kata positif menunjukkan bahwa banyak ulasan atau review yang dikirim oleh para pengguna menggunakan kata-kata tersebut untuk menyampaikan pengalaman positif mereka. Hal ini menunjukkan bahwa aspek terbaik, peningkatan, kesenangan, dan interaksi sosial adalah faktor-faktor yang sangat dihargai oleh pengguna PUBG. Pengembang dapat menggunakan informasi ini untuk terus meningkatkan dan mempertahankan elemen-elemen ini dalam game.



Gambar 8. Word cloud ulasan negatif

Pada **Gambar 8**. didapatkan bahwa dalam word cloud bagian kata “Negatif”, terlihat bahwa kata-kata yang paling sering muncul adalah “cheater”, “nt” (nice try), “lag”, “bug”, dan “crash”. Selain itu, terdapat juga kata-kata kasar yang digunakan dalam konteks negatif. Pada kata “cheater” mengartikan bahwa ketidakpuasan mereka terhadap pengguna yang curang, yang mengurangi kualitas dan keadilan dalam permainan. Pada kata “nt (nice try)” mengartikan bahwa menggambarkan fitur atau update yang tidak memenuhi harapan pengguna. Pada kata “lag” mengartikan bahwa masalah dengan performa jaringan yang mengganggu pengalaman bermain dan menyebabkan frustrasi di kalangan pengguna. Pada kata “bug” mengartikan bahwa pengguna sering mengeluh tentang bug atau kesalahan dalam game yang merusak pengalaman bermain mereka. Dan pada kata “crash” mengartikan bahwa permainan sering mengalami kegagalan atau berhenti secara tiba-tiba, yang sangat mengganggu dan menyebabkan ketidakpuasan. Kehadiran kata-kata ini dalam word cloud bagian kata negatif menunjukkan bahwa masalah teknis dan perilaku curang adalah sumber utama ketidakpuasan di antara para pengguna. Meskipun kata-kata ini digunakan untuk mengekspresikan keluhan, mereka juga memberikan umpan balik berharga bagi pengembang. Dengan fokus pada isu-isu yang diidentifikasi melalui kata-kata negatif ini, pengembang dapat melakukan evaluasi dan perbaikan yang diperlukan untuk meningkatkan kualitas dan performa game, serta meningkatkan kepuasan pengguna secara keseluruhan.



Gambar 9. Word cloud ulasan netral

Pada **Gambar 9**. didapatkan bahwa dalam word cloud bagian kata “Netral”, terlihat bahwa semua proporsi kata hampir sama, tanpa ada kata yang memiliki frekuensi kemunculan yang signifikan lebih besar daripada yang lain. Hal ini dapat diamati dari ukuran kata-kata dalam word cloud yang relatif seragam. Kata-kata netral dalam ulasan cenderung tidak

menunjukkan kecenderungan kuat ke arah positif atau negatif. Mereka sering bersifat objektif, memberikan pandangan yang seimbang atau tidak memihak. Kecenderungan netral mengartikan bahwa kalimat netral biasanya tidak memihak, tetapi dalam beberapa kasus, kata-kata negatif atau positif yang tidak terlalu kuat dapat masuk ke dalam kategori netral. Misalnya, sebuah ulasan mungkin menyebutkan adanya masalah kecil (negative) tetapi secara keseluruhan tetap puas (positif), yang membuat ulasan tersebut terlihat netral. Analisis ini menunjukkan bahwa ulasan netral memberikan pandangan yang lebih berimbang tentang pengalaman pengguna. Pengguna seringkali mencerminkan realitas yang lebih objektif dan dapat memberikan wawasan yang berguna bagi pengembang untuk memahami aspek-aspek yang mungkin tidak terlalu baik atau buruk tetapi tetap penting untuk diperhatikan. Dengan memahami pola dalam ulasan netral, pengembang dapat menangkap nuansa dan detail yang mungkin terlewatkan dalam ulasan yang sangat positif atau sangat negatif, sehingga dapat mengambil tindakan yang lebih tepat dalam meningkatkan aplikasi.

Dari hasil diagram bar chart, pie chart, word cloud kata positif, negatif dan juga netral dihasilkan mayoritas ulasan bersifat positif, tetapi analisis menunjukkan bahwa ada aspek-aspek tertentu yang masih perlu diperbaiki untuk meningkatkan pengalaman pengguna secara keseluruhan.

F. Akurasi Kedua Model evaluasi

Pada penelitian ini juga telah dibuat 2 model supervised learning yaitu naive bayes dan regresi logistik, setelah data telah dipartisi menjadi 80 persen menjadi data latih dan 20 persen data uji menghasilkan hasil seperti pada **Gambar 10**.

Naive Bayes Model				
Accuracy: 0.7196870720458374				
Classification Report:				
	precision	recall	f1-score	support
-1	0.83	0.76	0.79	16153
0	0.84	0.43	0.57	19789
1	0.65	0.91	0.75	27587
accuracy			0.72	63529
macro avg	0.77	0.70	0.70	63529
weighted avg	0.75	0.72	0.71	63529

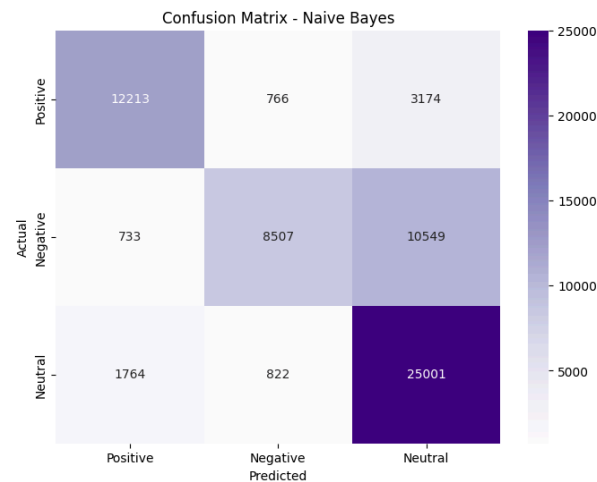
Logistic Regression Model				
Accuracy: 0.9488737427001842				
Classification Report:				
	precision	recall	f1-score	support
-1	0.93	0.91	0.92	16153
0	0.96	0.97	0.96	19789
1	0.96	0.96	0.96	27587
accuracy			0.95	63529
macro avg	0.95	0.95	0.95	63529
weighted avg	0.95	0.95	0.95	63529

Gambar 10. Akurasi dan evaluasi model

Akurasi yang dihasilkan pada **Gambar 10**, model logistik lebih tinggi 0.95 atau 95 persen dibandingkan dengan model Naive Bayes (0.72) atau 72 persen. Untuk semua kategori sentimen (Negative, Neutral, Positive), model Logistic Regression menunjukkan nilai precision, recall, dan f1-score yang lebih tinggi dibandingkan dengan model Naive Bayes. Terutama untuk sentimen Neutral, model Naive Bayes memiliki recall yang sangat rendah (0.43) dibandingkan dengan Logistic Regression (0.97). Nilai macro dan weighted averages untuk precision, recall, dan f1-score juga lebih tinggi pada model Logistic Regression, menunjukkan performa yang lebih konsisten di semua kategori sentimen.

G. Evaluation Metrics

Setelah dilakukan pengujian dataset, evaluasi akurasi dilakukan untuk mengukur seberapa tinggi keakuratan dan kesalahan dari model yang telah dibuat untuk klasifikasi teks, ilustrasi confusion matrix model Naive Bayes dapat dilihat sebagai berikut :



Gambar 11. Confusion Matrix Naive Bayes

Pada **Gambar 11.** dapat dijelaskan bahwa confusion matrix dari model Naive Bayes dapat digunakan untuk mengklasifikasikan data ke dalam tiga kelas, yaitu kelas “Positif”, “Negatif”, dan “Netral”. Berdasarkan confusion matrix, performa model Naive Bayes dapat dianalisis, yaitu untuk kelas “Positif”, model Naive Bayes dapat mengklasifikasikan kategori sebagai “True Positif” dengan jumlah 12.213 kata. Namun terdapat kategori “False Positif” dengan jumlah 733 kata. Dan klasifikasi dengan kategori “Netral Positif” dengan jumlah 1.764 kata. Ini menunjukkan bahwa meskipun model ini cukup efektif dalam mengenali kelas “Positif”, ada sejumlah besar kategori yang diklasifikasikan ke kategori lain.

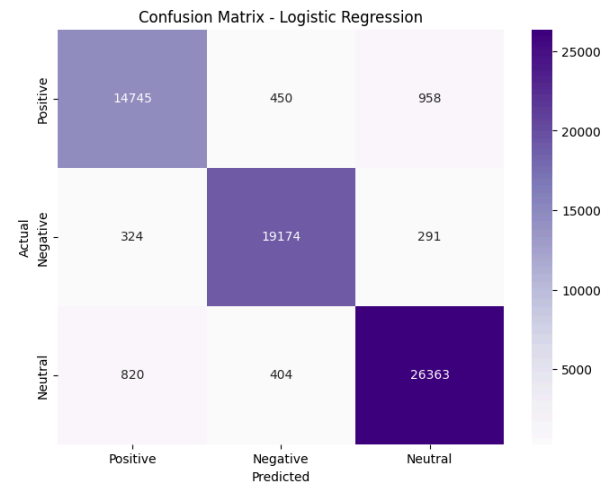
Untuk kelas “Negatif”, model Naive Bayes dapat mengklasifikasikan kategori “True Negatif” dengan jumlah 8.507 kata. Terdapat klasifikasi kategori “False Negatif” dengan jumlah 766 kata. Dan klasifikasi kategori “Netral Negatif” dengan jumlah 822 kata. Dari hasil tersebut dapat menandakan bahwa model ini masih memiliki kesulitan dalam membedakan antara kelas “Negatif” dan kategori lainnya.

Untuk kelas “Netral”, model Naive Bayes dapat mengklasifikasikan kategori “True Netral” dengan jumlah 25.001 kata. Terdapat klasifikasi dengan kategori “Positif Netral” dengan jumlah 3.174 kata. Dan klasifikasi kategori “Negatif Netral” dengan jumlah 10.549 kata. Dari hasil tersebut dapat menunjukkan performa yang cukup baik. Namun, masih ada sejumlah kategori yang mungkin salah diklasifikasikan.

Secara keseluruhan, model Naive Bayes menunjukkan hasil akurasi sebesar 75% untuk pengklasifikasian. Meskipun model ini memiliki kelebihan dalam efisiensi

dan kecepatan, terutama dengan dataset yang besar, hasil klasifikasinya menunjukkan bahwa masih ada ruang untuk perbaikan, terutama dalam mengurangi kesalahan klasifikasi antara kategori yang berbeda.

Selanjutnya dapat dilihat ilustrasi dari confusion matrix model Logistic Regression sebagai berikut :



Gambar 12. Confusion Matrix Logistic Regression

Pada **Gambar 12.** dapat dijelaskan bahwa model Logistic Regression dapat digunakan untuk mengklasifikasikan data dalam tiga kelas yang sama yaitu kelas “Positif”, “Negatif”, dan “Netral”. Berdasarkan confusion matrix, model Logistic Regression dapat dianalisis untuk per kelas nya. Untuk kelas “Positif”, model Logistic Regression dapat mengklasifikasikan sebagai kategori “True Positif” dengan jumlah 14.745 kata. Terdapat Klasifikasi dengan kategori “False Positif” dengan jumlah 324 kata. Dan klasifikasi dengan kategori “Netral Positif” dengan jumlah 820 kata. Dari hasil tersebut menunjukkan bahwa model ini sangat efektif dalam mengenali kelas “Positif” dengan kesalahan klasifikasi yang jauh lebih rendah dibandingkan dengan model Naive Bayes.

Untuk kelas “Negatif”, model Logistic Regression dapat mengklasifikasikan kategori “True Negatif” dengan jumlah 19.174 kata. Terdapat klasifikasi kategori “False Negatif” dengan jumlah 404 kata. Dan klasifikasi dengan kategori “Netral Negatif” dengan jumlah 366 kata. Dari hasil tersebut menunjukkan bahwa model ini sangat baik dalam membedakan kelas “Negatif” dari kategori lainnya.

Untuk kelas “Netral”, model Logistic Regression dapat mengklasifikasikan dengan kategori “True Netral”

dengan jumlah 26.363 kata. Terdapat klasifikasi dengan kategori “Positif Netral” dengan jumlah 958 kata. Dan klasifikasi dengan kategori “Negatif Netral” dengan jumlah 291 kata. Dari hasil tersebut menunjukkan bahwa model ini memiliki kemampuan yang kuat untuk mengidentifikasi kelas “Netral” dengan sangat akurat.

Secara keseluruhan, model Logistic Regression menunjukkan hasil akurasi sebesar 94% untuk pengklasifikasian. Model ini tidak hanya memiliki akurasi yang lebih tinggi dibandingkan dengan model Naive Bayes, tetapi juga menunjukkan performa superior dalam mengurangi kesalahan klasifikasi antara kelas berbeda. Meskipun model ini sudah efektif, tetapi selalu ada kemungkinan untuk meningkatkan performa lebih lanjut melalui teknik tambahan atau penyesuaian fitur.

Dengan mempertimbangkan kedua model dan analisis tersebut, didapatkan bahwa model Logistic Regression menghasilkan keunggulan yang lebih signifikan dalam klasifikasi teks dibandingkan dengan model Naive Bayes, meskipun setiap model memiliki kelebihan dan kekurangannya sendiri.

Conclusions / Kesimpulan

Penelitian ini menganalisis sentimen atau opini pengguna terhadap game PUBG berdasarkan ulasan yang ditulis di platform Steam pada tahun 2021. Berdasarkan hasil distribusi sentimen menunjukkan bahwa sebesar 43.5% ulasan tergolong dalam sentimen positif, diikuti oleh 31.3% sentimen netral dan 25.2% sentimen negatif. Dari hasil visualisasi data yang dihasilkan, mayoritas ulasan muncul dengan kata-kata bersifat “Positif” (138129 kata) dibandingkan “Netral” (99379 kata) dan “Negatif” (80135 kata). Selain itu, proporsi ulasan “Netral” dan “Negatif” yang signifikan juga menunjukkan bahwa terdapat ruang untuk peningkatan lebih lanjut dalam pengalaman pengguna. Pengembang aplikasi dapat menggunakan informasi ini untuk fokus pada aspek-aspek yang paling dihargai oleh pengguna, serta mengidentifikasi dan memperbaiki area yang menimbulkan ketidakpuasan.

Hasil evaluasi model menunjukkan bahwa Logistic Regression memiliki akurasi yang lebih tinggi, yaitu 94%, sedangkan Naive Bayes hanya 75%. Logistic Regression lebih efektif dalam mengklasifikasikan ulasan dengan kesalahan klasifikasi yang lebih rendah dibandingkan Naive Bayes. Model Naive Bayes kesulitan membedakan antara kelas “Negatif” dan kategori lainnya, sementara Logistic

Regression dapat membedakan ketiga kelas dengan lebih baik.

Logistic Regression menunjukkan performa yang lebih baik dalam klasifikasi teks dibandingkan Naive Bayes, dengan akurasi yang lebih tinggi dan kemampuan yang lebih baik dalam mengurangi kesalahan klasifikasi antara kelas yang berbeda. Oleh karena itu, penggunaan model Logistic Regression disarankan untuk analisis sentimen ulasan game PUBG di Steam agar dapat memberikan hasil yang lebih akurat dan efektif. Pengembang game dapat menggunakan hasil analisis ini untuk meningkatkan kualitas dan kepuasan pengguna terhadap game PUBG.

Conflicts of interest

Tidak ada konflik kepentingan yang perlu diungkapkan dalam penelitian ini.

Acknowledgements

Para penulis ingin mengucapkan terima kasih kepada Dosen Pengampu Mata Kuliah SD4202-Analisis Big Data, yaitu Ibu Luluk Muthoharoh, M. Si., Ibu Rizty Maulida Badri, M. Si., dan Bapak Ardika Satria, M. Si. atas bimbingannya yang sangat berharga sehingga penulis dapat menyelesaikan tugas besar ini dengan sebaik-baiknya.

References

- [1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135. <https://doi.org/10.1561/15000000011>
- [2] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster computing with working sets. *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing*, 10-10. <https://doi.org/10.5555/1863103.1863113>
- [3] S. Taylor, "Valve Reveal Steams Monthly Active User Count Game Sales Region," *geekwire*, 2017. [Online]. Available: <https://www.geekwire.com/2017/valve-reveals-steams-monthly-active-user-count-game-sales-region/>.
- [4] Marchand, A., & Hennig-Thurau, T. (2013). Value creation in the video game industry: Industry economics, consumer benefits, and research opportunities. *Journal of Interactive Marketing*, 27(3), 141-157. <https://doi.org/10.1016/j.intmar.2013.05.001>
- [5] Kumar, S., & Geethakumari, G. (2014). Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, 4(1), 14. <https://doi.org/10.1186/s13673-014-0014-x>
- [6] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report*, Stanford, 1-12. <https://doi.org/10.3115/1690219.1690271>
- [7] Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14)*, 216-225. <https://doi.org/10.1609/icwsm.v8i1.14550>

- [8] Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 Task 4: Sentiment Analysis in Twitter. Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 502-518. <https://doi.org/10.18653/v1/S17-2088>
- [9] R. Apriani and D. Gustian, "ANALISIS SENTIMEN DENGAN NAÏVE BAYES TERHADAP KOMENTAR APLIKASI TOKOPEDIA," Jurnal Rekayasa Teknologi Nusa Putra, vol. 6, pp. 54-62, Sep. 2019. [Online]. Available: <https://doi.org/10.52005/rekayasa.v6i1.86>
- [10] A. Wibawa et al., "Naïve Bayes Classifier for Journal Quartile Classification," International Journal of Recent Contributions from Engineering, Science & IT (IJES), vol. 7, no. 2, pp. 91-100, Jun. 2019. [Online]. Available: <https://doi.org/10.3991/ijes.v7i2.10659>
- [11] P. Reddy, D. Sri, C. Reddy, and Subhani Shaik, "Sentimental Analysis using Logistic Regression," International Journal of Engineering Research and Applications, vol. 11, no. 7, pp. 36-40, Jul. 2021. [Online]. Available: <https://doi.org/10.9790/9622-1107023640>
- [12] N. Lengkong, O. Safitri, S. Machsus, Y. Putra, A. Syahadati, and R. Nooraeni, "ANALISIS SENTIMEN PENERAPAN PSBB DI DKI JAKARTA DAN DAMPAKNYA TERHADAP PERGERAKAN IHSG," Jurnal Teknoinfo, vol. 15, no. 1, p. 20, Jan. 2021. [Online]. Available: <https://doi.org/10.33365/jti.v15i1.866>
- [13] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Perbandingan Naïve Bayes dan Logistic Regression dalam Analisis Sentimen pada Ulasan Marketplace Menggunakan Pelabelan Berbasis Rating," Journal of Information Systems and Informatics, vol. 5, no. 3, p. 915, Sep. 2023. [Online]. Available: <https://doi.org/10.51519/journalisi.v5i3.539>
- [14] A. Pangestu, Y. T. Arifin, and R. A. Safitri, "Analisis Sentimen Review Publik Pemain Game Online pada Platform Steam Menggunakan Algoritma Naïve Bayes," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 6, p. 3106, Dec. 2023. <https://doi.org/10.36040/jati.v7i6.8829>
- [15] Handrizal, F. Y. Manik, and H. M. A., "Analisis Sentimen Berdasarkan Aspek pengguna PUBG dari Ulasan Toko Aplikasi Menggunakan Bidirectional Encoder Representation from Transformer (BERT)," Journal of Theoretical and Applied Information Technology, vol. 102, no. 4, p. 1740, Feb. 2024.
- [16] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naïve Bayes dalam Menganalisis Sentimen Data Ulasan PLN Mobile," PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika, vol. 15, no. 2, p. 264, Sep. 2022. [Online]. Available: <https://doi.org/10.33322/petir.v15i2.1733>
- [17] F. Hidayat and Sugiyono, "Analisis Sentimen Masyarakat Terhadap Perekrutan PPPK Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," Jurnal Sains dan Teknologi, vol. 5, no. 2, pp. 665-672, Dec. 2023. [Online]. Available: <https://doi.org/10.55338/saintek.v5i2.1359>
- [18] D. Salsabillah, D. E. Ratnawati, and N. Setiawan, "Analisis Sentimen Ulasan Rumah Makan Menggunakan Perbandingan Algoritma Support Vector Machine dengan Naive Bayes (Studi Kasus: Ayam Goreng Nelongso Cabang Singosari, Malang)," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 11, pp. 107-116, Feb. 2024. [Online]. Available: <https://doi.org/10.25126/jtiik.20241117584>
- [19] Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," Jurnal Ilmiah Computer Science, vol. 2, no. 2, pp. 89-98, Jan. 2024. [Online]. Available: <https://doi.org/10.58602/iics.v2i2.24>
- [20] R. Prabowo, H. Sujaini, and T. Rismawan, "Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode Regresi Logistik Multinomial," Jurnal Sistem dan Teknologi Informasi (JustIN), vol. 11, no. 2, p. 366, Jul. 2023. [Online]. Available: <https://doi.org/10.26418/justin.v11i2.57449>
- [21] S. Rajagukguk, S. Mutmainah, and A. Satria, "Analisis Sentimen Pembelajaran Tatap Muka dengan Apache SPARK," Jurnal Rekayasa Teknologi Informasi (JURTI), vol. 6, no. 2, p. 159, Dec. 2022. [Online]. Available: <https://doi.org/10.30872/jurti.v6i2.8162>
- [22] R. Fauzianto and Supatman, "Analisis Sentimen Opini Masyarakat Terhadap Tech Winter Pada Twitter Menggunakan Natural Language Processing," Jurnal Syntax Admiration, vol. 3, no. 9, pp. 1577-1585, Sep. 2023. [Online]. Available: <https://doi.org/10.46799/jsa.v3i9.909>
- [23] A. Aulia and E. Ritonga, "Analisis Perilaku Agresivitas Komunikasi Terhadap Remaja pengguna Game Online PUBG (Playerunknown's Battlegrounds)," G-Couns: Jurnal Bimbingan dan Konseling, vol. 8, no. 2, pp. 1037-1047, Feb. 2024. [Online]. Available: <https://doi.org/10.31316/gcouns.v8i2.5907>
- [24] D. Abimanyu, E. Budianita, E. Cynthia, F. Yanto, and Yusra, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER," Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), vol. 5, no. 3, pp. 423-431, Jun. 2022. [Online]. Available: <https://doi.org/10.32672/jnkti.v5i3.4382>
- [25] F. Sarimole and Kudrat, "Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine: Analisis Sentimen," Jurnal Sains dan Teknologi, vol. 5, no. 3, pp. 783-790, Feb. 2024. [Online]. Available: <https://doi.org/10.55338/saintek.v5i3.2702>

Attachment / Lampiran

Folder Jurnal :  ANALISIS BIG DATA

Code :  CBL_Kel9_ABD_RB.ipynb

PPT : [CANVA](#)