

**PERBANDINGAN METODE K – NEAREST NEIGHBOR (KNN) DAN
NAIVE BAYES TERHADAP ANALISIS SENTIMEN PADA PENGGUNA
E-WALLET APLIKASI DANA MENGGUNAKAN FITUR EKSTRAKSI
TF-IDF**

SKRIPSI

Diajukan Oleh :

Muhammad Rayhan Elfansyah

2011102441075



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR

JULI 2024

**PERBANDINGAN METODE K – NEAREST NEIGHBOR (KNN) DAN
NAIVE BAYES TERHADAP ANALISIS SENTIMEN PADA PENGGUNA
E-WALLET APLIKASI DANA MENGGUNAKAN FITUR EKSTRAKSI
TF-IDF**

SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar Sarjana Teknik
Informatika Fakultas Sains Dan Teknologi Universitas Muhammadiyah Kalimantan Timur

Diajukan Oleh :

Muhammad Rayhan Elfansyah

2011102441075



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR

JULI 2024

HALAMAN PERSETUJUAN

**PERBANDINGAN METODE K – NEAREST NEIGHBOR (KNN) DAN
NAIVE BAYES TERHADAP ANALISIS SENTIMENT PADA
PENGGUNA E-WALLET APLIKASI DANA MENGGUNAKAN FITUR
EKSTRAKSI TF-IDF**

SKRIPSI

Diajukan Oleh :

**Muhammad Rayhan Elfansyah
2011102441075**

Disetujui untuk diujikan

Pada tanggal 27 Juni 2024

Pembimbing



**Rudiman, S.Kom, M.Sc
NIDN 1105068202**

Mengetahui,

Koordinator Tugas Akhir/Skripsi/Tesis/Disertasi



**Abdul Rahim, S.Kom., M.Cs
NIDN 0009047901**

HALAMAN PENGESAHAN

**PERBANDINGAN METODE K – NEAREST NEIGHBOR (KNN) DAN
NAIVE BAYES TERHADAP ANALISIS SENTIMENT PADA
PENGGUNA E-WALLET APLIKASI DANA MENGGUNAKAN FITUR
EKSTRAKSI TF-IDF**

Tugas akhir/skripsi



Diajukan oleh:

Muhammad Rayhan Elfansyah

2011102441075

Diseminarkan dan Diujikan

Pada Tanggal 17 Juli 2024

Penguji I	Penguji II
 <u>Fendy Yulianto, S.Kom., M.Kom.</u> NIDN 1102079402	 <u>Rudiman, S.Kom, M.Sc</u> NIDN 1105068202

Mengetahui,

Ketua

Program Studi/Teknik Informatika



Arbansyah, S.Kom, M.TI

NIDN. 1118019203

PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Rayhan Elfansyah
NIM : 2011102441075
Program Studi : S1 Teknik Informatika
Judul Penelitian : Perbandingan Metode K-Nearest Neighbor (KNN) dan Naive Bayes Terhadap Analisis Sentiment Pada Pengguna Aplikasi DANA Menggunakan Fitur Ekstraksi TF-IDF

Menyatakan bahwa Skripsi yang saya tulis ini benar-benar hasil karya saya sendiri, dan bukan merupakan hasil plagiasi/falsifikasi/fabrikasi baik sebagian atau seluruhnya.

Atas pernyataan ini, saya siap menanggung resiko atau sanksi yang dijatuhkan kepada saya apa bila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam skripsi saya ini, atau klaim dari pihak lain terhadap keaslian karya saya ini.

Samarinda, 30 Juni 2024

Yang membuat pernyataan



Muhammad Rayhan Elfansyah

NIM : 2011102441075

ABSTRAK

Penelitian ini membandingkan akurasi metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam klasifikasi sentimen pengguna aplikasi *E-Wallet* DANA menggunakan fitur ekstraksi *Term Frequency-Inverse Document Frequency* (TF-IDF). Data ulasan pengguna dikumpulkan melalui teknik web scraping dan dilabeli oleh ahli bahasa dan model *lexicon*. Setelah melalui proses *pre-processing* seperti *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*, data diklasifikasikan menggunakan metode KNN dan Naive Bayes. Hasil penelitian menunjukkan bahwa pelabelan data oleh ahli bahasa signifikan dalam meningkatkan akurasi kedua metode klasifikasi tersebut. Selain itu, penggunaan TF-IDF sebagai metode pembobotan kata terbukti efektif dalam meningkatkan performa model klasifikasi sentimen. Analisis sentimen terhadap ulasan pengguna aplikasi DANA mengungkapkan berbagai keluhan dan masalah yang dihadapi oleh pengguna, memberikan informasi yang dapat digunakan untuk memperbaiki fitur dan layanan yang diberikan sehingga dapat meningkatkan kepuasan pengguna. Penelitian ini juga memberikan perbandingan antara metode KNN dan Naive Bayes yang dapat menjadi referensi bagi peneliti lain dalam memilih metode yang sesuai untuk analisis sentimen pada dataset serupa.

Kata kunci: Analisis Sentimen, Aplikasi DANA, K-Nearest Neighbor, Naive Bayes

ABTRACT

This research compares the accuracy of the K-Nearest Neighbor (KNN) and Naive Bayes methods in classifying user sentiment towards the DANA e-wallet application using Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction. User review data was collected through web scraping techniques and labeled by linguists and lexicon models. After undergoing pre-processing steps such as case folding, cleaning, tokenizing, stopword removal, and stemming, the data was classified using the KNN and Naive Bayes methods. The research results indicate that data labeling by linguists significantly improves the accuracy of both classification methods. Additionally, using TF-IDF as a word weighting method proves effective in enhancing the performance of sentiment classification models. Sentiment analysis of user reviews of the DANA application reveals various complaints and issues faced by users, providing information that can be used to improve the features and services offered, thereby increasing user satisfaction. This research also provides a comparison between the KNN and Naive Bayes methods, which can serve as a reference for other researchers in selecting appropriate methods for sentiment analysis on similar datasets.

Keywords: Sentiment Analysis, DANA Application, K-Nearest Neighbor, Naive Bayes

PRAKATA

Puji dan syukur saya panjatkan kepada Allah SWT. Atas ridho-Nya saya dapat menyelesaikan penyusunan skripsi ini. Adapun judul skripsi yang saya ajukan adalah “Perbandingan Metode K-Nearest Neighbor (KNN) Dan Naive Bayes Terhadap Analisis Sentimen Pada Pengguna E-Wallet Aplikasi DANA Menggunakan Fitur Ekstraksi TF-IDF ”

Skripsi ini diajukan untuk memenuhi syarat kelulusan mata kuliah Skripsi di Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur. Tidak dapat disangka bahwa butuh usaha yang keras dalam penyelesaian pengerjaan skripsi ini. Namun, karya ini tidak akan selesai tanpa orang-orang tercinta di sekeliling saya yang mendukung dan membantu. Terima kasih saya sampaikan kepada:

- 1) Kedua orang tua penulis, yaitu superhero dan panutan penulis Bapak Fadliansyah, terimakasih selalu berjuang dalam mengupayakan yang terbaik untuk kehidupan penulis, beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan, namun beliau mampu mendidik penulis, memotivasi, memberikan dukungan hingga penulis mampu menyelesaikan studinya sampai sarjana.
- 2) Pintu surga penulis, Ibunda Fatmawati, yang tidak henti-hentinya memberikan kasih sayang dengan penuh cinta dan selalu memberikan dukungan serta doa yang teramat tulus sehingga penulis mampu menyelesaikan studinya sampai sarjana.
- 3) Bapak Dr. Muhammad Musiyam, M.T. Selaku Rektor Univeritas Muhammadiyah Kalimantan Timur yang telah memberi kesempatan pada peneliti untuk menempuh pendidikan di Universitas Muhammadiyah Kalimantan Timur.

- 4) Bapak Arbansyah, Selaku Ketua Program Studi Teknik Informatika dan sebagai dosen pembimbing akademik penulis di Fakultas Sains dan Teknologi Universitas Muhammadiyah Kalimantan Timur.
- 5) Bapak Rudiman, M.Sc. Selaku Dosen Pembimbing penulis yang telah membagi ilmu, bimbingan dan arahan selama proses bimbingan.
- 6) Seorang wanita yang bernama Nadhira Sintha Aulia yang selalu ada disaat senang dan susah, terimakasih atas kesabaran, cinta, dan pengertian yang tiada batas, Kehadiranmu adalah anugerah terindah yang menguatkan langkahku setiap hari.
- 7) Sahabat penulis, Khoirul Huda, Ikhsan Nuttakwa, Ipan Hasmadi, Farhat Jundullah, Zulkarnaen, dan anggota kos lainnya terimakasih atas dukungan dan semangat yang diberikan kepada penulis selama ini.
- 8) Teman satu bimbingan dan teman kuliah penulis yang bernama Rendy Nurdiansyah, Anggiq Karisma, Ilham, Takhta Perlawanan, dll yang sudah membantu saya dan mensupport saya, bersama-sama menyelesaikan skripsi ini sampai akhir.

Semoga segala kebaikan dan pertolongan semuanya mendapat berkah dari Allah SWT. Dan akhirnya penulis menyadari bahwa skripsi ini tidak luput dari berbagai kekurangan, karena keterbatasan ilmu yang saya miliki. Untuk itu dengan kerendahan hati saya mengharapkan saran dan kritik yang sifatnya membangun dari semua pihak demi membangun laporan penelitian ini

Samarinda, 30 Juni 2024



Muhammad Rayhan Elfansyah

DAFTAR ISI

	Halaman
HALAMAN PERSETUJUAN	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN KEASLIAN PENELITIAN	iii
ABSTRAK	iv
<i>ABTRACT</i>	v
PRAKATA	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah.....	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	5
BAB II METODOLOGI PENELITIAN.....	6
2.1. Obyek Penelitian.....	6
2.2. Alat Dan Bahan	6
2.3. Prosedur Penelitian	7
2.3.1. Pengumpulan Data (<i>Scrapping</i>)	8
2.3.2. <i>Labelling Data</i>	9
2.3.3. <i>Pre-Processing</i>	10
2.3.4. Pembobotan Kata (TF-IDF)	10
2.3.5. Split Data.....	11
2.3.6. Klasifikasi.....	12
2.3.7. K-Nearest Neighbor (KNN).....	12
2.3.8. Naive Bayes.....	13
2.3.9. Evaluasi	14
2.3.10. Jadwal Penelitian.....	15
BAB III HASIL DAN PEMBAHASAN.....	17
3.1. Hasil Pengumpulan Data	17

3.2.	Hasil <i>Labelling</i> Data.....	18
3.3.	Hasil <i>Pre-Processing</i>	20
3.3.1.	<i>Case Folding</i>	20
3.3.2.	<i>Cleaning</i>	21
3.3.3.	<i>Tokenizing</i>	22
3.3.4.	<i>Stopword Removal</i>	23
3.3.5.	<i>Stemming</i>	24
3.4.	Hasil Pembobotan Kata (TF-IDF).....	25
3.5.	Hasil <i>Visualiasi</i>	28
3.6.	Hasil Evaluasi.....	29
3.6.1.	Metode K-Nearest Neighbor (KNN).....	31
3.6.2.	Metode Naive Bayes	36
BAB IV	PENUTUP	39
4.1.	Simpulan.....	39
4.2.	Implikasi	40
DAFTAR	RUJUKAN	42
LAMPIRAN	44
RIWAYAT	HIDUP	59

DAFTAR TABEL

Tabel	Halaman
2.1 Jadwal Penelitian.....	15
3.1 Hasil Pengumpulan Data.....	17
3.2 Hasil Labelling Data Expert.....	18
3.3 Hasil Labelling Data Lexicon.....	19
3.4 Hasil Case Folding.....	21
3.5 Hasil Cleaning.....	22
3.6 Hasil Tokenizing.....	23
3.7 Hasil Stopword Removal.....	24
3.8 Hasil Stemming.....	24
3.9 Pembagian Data KNN.....	30
3.10 Pembagian Data Naive Bayes.....	30
3.11 Hasil Data Uji KNN Expert.....	35
3.12 Hasil Data Uji KNN Lexicon.....	35
3.13 Hasil Data Uji Naive Bayes Expert.....	37
3.14 Hasil Data Uji Naive Bayes Lexicon.....	37

DAFTAR GAMBAR

Gambar	Halaman
2.1 Alur Penelitian.....	7
2.2 Hasil Pengumpulan Data	8
2.3 Pelabelan Data Expert	9
2.4 Pelabelan Data Lexicon.....	9
3.1 Pembobotan Kata (TF-IDF)	26
3.2 Skoring TF.....	27
3.3 WordCloud Positif.....	28
3.4 WordCloud Negatif.....	28
3.5 WordCloud Netral	29
3.6 Hasil Akurasi Nilai K.....	32
3.7 Grafik Akurasi Nilai K.....	33
3.8 Hasil Klasifikasi KNN Expert.....	33
3.9 Hasil Klasifikasi KNN Lexicon	34
3.10 Hasil Confusion Matriks KNN Expert	34
3.11 Hasil Confusion Matriks KNN Lexicon.....	34
3.12 Hasil Klasifikasi Naive Bayes Expert	36
3.13 Hasil Klasifikasi Naive Bayes Lexicon.....	36
3.14 Hasil Confusion Matriks Naive Bayes Expert.....	36
3.15 Hasil Confusion Matriks Naive Bayes Lexicon	37

DAFTAR LAMPIRAN

Lampiran	Halaman
L1 CV Expert Labelling Data	44
L2 Sertifikat Expert	46
L3 Surat Ijin Penelitian.....	48
L4 Code Pengumpulan Data.....	49
L5 Install Library	49
L6 Membaca Dataset.....	50
L7 Cek Informasi Dataset.....	50
L8 Cek Jumlah Sentimen Positif, Negatif, Dan Netral	50
L9 Membaca Dataset Kolom Review dan Sentimen.....	50
L10 Pre-Processing Data	50
L11 Pembobotan Kata (TF-IDF).....	52
L12 Hasil Akurasi Nilai K.....	52
L13 Grafik Akurasi Nilai K.....	53
L14 Split Data dan Evaluasi Confusion Matrix Metode KNN.....	53
L15 Split Data dan Evaluasi Confusion Matrix Metode Naive Bayes	54
L16 Visualisasi Wordcloud Positif, Negatif, Dan Netral	54
L17 Kartu Kendali Bimbingan	55

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Saat ini, teknologi informasi dan komunikasi sedang mengalami perkembangan pesat di era industri 4.0, terutama aktivitas bisnis dalam sektor niaga menjadi semakin kompleks. Kemudahan bertransaksi online telah mendorong banyak masyarakat beralih ke dalam penggunaan uang elektronik (*E-Money*) (Salehudin Basryah et al., 2021). Pembayaran non-tunai kini terus berkembang di Indonesia, terutama dengan munculnya berbagai produk *Financial Technology (Fintech)* dompet digital (*E-Wallet*) seperti Aplikasi GoPay, OVO, DANA, LinkAja, Paytren, dan lainnya (Badri, 2020). Setelah GoPay dan OVO, DANA menjadi salah satu aplikasi dompet digital yang populer.

Aplikasi DANA adalah layanan keuangan digital yang berbasis di Jakarta, yang berperan sebagai pembayaran digital. DANA berdiri pada tahun 2018 dan terdaftar di Bank Indonesia, Aplikasi DANA memungkinkan pengguna untuk melakukan transaksi keuangan tanpa uang tunai mulai dari pembayaran kebutuhan sehari-hari hingga penambahan saldo untuk investasi (Larasati et al., 2022). Meskipun demikian keamanan penggunaan uang elektronik masih menjadi titik lemah dan perlu ditingkatkan, Berbagai keluhan pengguna aplikasi DANA tentang layanan dan fitur yang diberikan memiliki opini negatif seperti sulitnya untuk mengupgrade aplikasi ke premium, seringnya terjadi kehilangan saldo, dan seringnya terjadi error saat ingin mau melakukan transaksi (Heti Palestina Yunani, 2024).

Tentunya dengan sejumlah fitur layanan yang disediakan oleh DANA, hal ini menciptakan beragam tanggapan dari pengguna yang terdapat dalam ulasan di Google Play Store. Sehingga sering menemukan ulasan yang bersifat positif, negatif, dan netral

baik itu dari segi fitur aplikasi, dari segi keamanan, dan lain-lain yang mungkin tidak sepenuhnya relevan dengan *rating* yang diberikan di Google Play Store (Athallah Muhammad et al., 2022). Pada tahun 2023, Aplikasi DANA sudah mencapai 170 juta pengguna atau meningkat 23 persen dari tahun sebelumnya. Data ulasan atau komentar pengguna aplikasi DANA akan diperoleh menggunakan metode teknik *web scraping* menggunakan bahasa pemrograman python yang bersumber dari kolom komentar pada Google Play Store (Shanty, 2024).

Maka dalam penelitian ini Google Play Store menjadi sumber data untuk menganalisis sentimen ulasan pengguna terhadap layanan dompet digital DANA. Analisis sentimen adalah ilmu yang melakukan evaluasi terhadap sentimen dan emosi yang terkait dengan suatu objek, produk, layanan, atau topik tertentu (Nurian & Nurina Sari, 2023). Analisis sentimen juga merupakan proses klasifikasi teks dalam kalimat atau dokumen dengan tujuan untuk menentukan apakah pendapat yang terungkap dalam teks tersebut bersifat positif, negatif atau netral (Amrullah et al., 2020).

Terdapat beberapa penelitian sebelumnya yang pertama dilakukan oleh Enggarbela Ogi Intan Pratiwi dan Wiyli Yustanti yang berjudul “Analisis Sentimen Kualitas Layanan Teknologi Pembayaran Elektronik pada Twitter (Studi Kasus OVO dan DANA)” peneliti menggunakan tiga model klasifikasi Naive Bayes Classifier, Support Vector Machine (SVM), dan K-Nearest Neighbor dengan fitur TF-IDF klasifikasi SVM mendapatkan akurasi paling tinggi yaitu 81,33% , Naive bayes 80,63% , dan KNN 75,83% (Ogi et al., 2021.) Penelitian kedua oleh M.Kabirur Rifa, Moch.Hafid Totohendarto, dan M.Rafi Muttaqin yang berjudul “Analisis Sentimen Pengguna E-Wallet DANA dan GoPay Pada Twitter Menggunakan Metode Support Vector Machine (SVM)” Peneliti menggunakan metode SVM dengan menggunakan fitur TF-IDF berhasil menganalisis sentimen dengan tingkat akurasi mencapai 92% (Rifa et al., 2023)

Data yang digunakan dalam penelitian ini terdiri dari 700 ulasan dalam bahasa Indonesia, Dalam pemilihan teknik atau algoritma yang tepat akan sangat bergantung pada tujuan yang diinginkan, dengan begitu, peneliti menggunakan dua metode dalam analisis sentimen untuk memperbandingkan tingkat akurasi dari kedua metode tersebut diantaranya adalah metode KNN Dan Naïve Bayes (Puspita & Widodo, 2021). Algoritma KNN adalah salah satu algoritma yang sangat populer, KNN termasuk dalam kelompok pembelajaran berbasis contoh. Metode KNN merupakan teknik pembelajaran yang lamban (*lazy learning*) (Cahyanti et al., 2020).

Metode K-Nearest Neighbor digunakan untuk proses klasifikasi dalam penelitian karena kesederhanaannya. Prosesnya didasarkan pada pendekatan pembobotan yang sederhana dan mudah diimplementasikan, diadaptasi, serta proses pembelajarannya. Selain itu, metode ini juga memiliki tingkat akurasi yang tinggi (Salim & Mayary, 2020). Selain menggunakan metode KNN, peneliti juga menerapkan metode Naïve Bayes. Naïve Bayes adalah sebuah metode *machine learning* yang berfokus pada probabilitas. Dengan kata lain, Naïve Bayes adalah metode untuk klasifikasi teks yang memiliki kecepatan pemrosesan yang tinggi, terutama saat digunakan dalam data yang besar (Nurul et al., 2019).

Berdasarkan uraian diatas, peneliti akan meneliti bagaimana melakukan analisis sentimen terhadap ulasan pengguna aplikasi DANA dengan menggunakan dua Metode yaitu KNN dan Naive Bayes, serta fitur TF-IDF untuk meningkatkan efisiensi proses klasifikasi. Dengan membagi sentimen menjadi kelas positif, negatif, dan netral. Data untuk penelitian ini diambil dari ulasan pengguna aplikasi DANA di Google PlayStore dengan menggunakan teknik *scraping* menggunakan bahasa pemograman pyhton. Tujuan dari penelitian ini adalah untuk memahami hasil analisis sentimen dalam kelas positif,

negatif dan netral serta memberikan evaluasi yang dapat digunakan untuk meningkatkan kualitas layanan pada aplikasi DANA.

1.2. Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah :

1. Bagaimana mengetahui permasalahan pengguna Aplikasi DANA berdasarkan ulasan dari kolom komentar Google Play Store Aplikasi DANA?
2. Bagaimana hasil perbandingan akurasi dari metode KNN Dan Naive Bayes menggunakan pelabelan *expert* dan model *lexicon based* berdasarkan data hasil ulasan dari Aplikasi DANA dengan menggunakan fitur ekstraksi TF-IDF?
3. Mengapa penting untuk membandingkan metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam analisis sentimen pada pengguna aplikasi E-Wallet DANA menggunakan fitur ekstraksi TF-IDF?

1.3. Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan umum :

1. Untuk mengetahui permasalahan pengguna Aplikasi DANA yang ada di kolom komentar Google Play Store.
2. Untuk mengetahui perbandingan akurasi metode K-Nearest Neighbor dan Naive Bayes dalam mengklasifikasikan opini pengguna di Google Play Store terhadap fitur dan pelayanan Aplikasi DANA ke dalam 3 kelas sentimen yaitu Positif, Negatif, dan Netral.
3. Untuk menganalisis dan membandingkan keakuratan serta efisiensi metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam melakukan analisis sentimen pada ulasan pengguna aplikasi *E-Wallet* DANA menggunakan fitur ekstraksi TF-IDF.

1.4. Manfaat Penelitian

Manfaat yang dapat diambil dari penelitian ini :

1. Penelitian ini diharapkan dapat meningkatkan kualitas fitur dan layanan di Aplikasi DANA yang diberikan terhadap pengguna yang Aplikasi DANA.
2. Memberikan informasi kepada masyarakat tentang sentimen pengguna terhadap Aplikasi DANA, sehingga dapat menjadi pertimbangan dalam memilih pembayaran elektronik (*E-Money*).
3. Memberikan pengalaman dalam melakukan penelitian yang melibatkan analisis sentimen menggunakan metode KNN dan Naive Bayes dengan Fitur Ekstraksi TF-IDF.

BAB II

METODOLOGI PENELITIAN

2.1. Obyek Penelitian

Objek penelitian dalam studi ini adalah ulasan pengguna aplikasi DANA yang tersedia di Google Play Store. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap layanan dompet digital DANA dengan menggunakan dua metode klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Naive Bayes. Data yang digunakan berasal dari teknik *web scraping* menggunakan bahasa pemrograman Python, yang mengumpulkan sebanyak 700 ulasan dalam bahasa Indonesia.

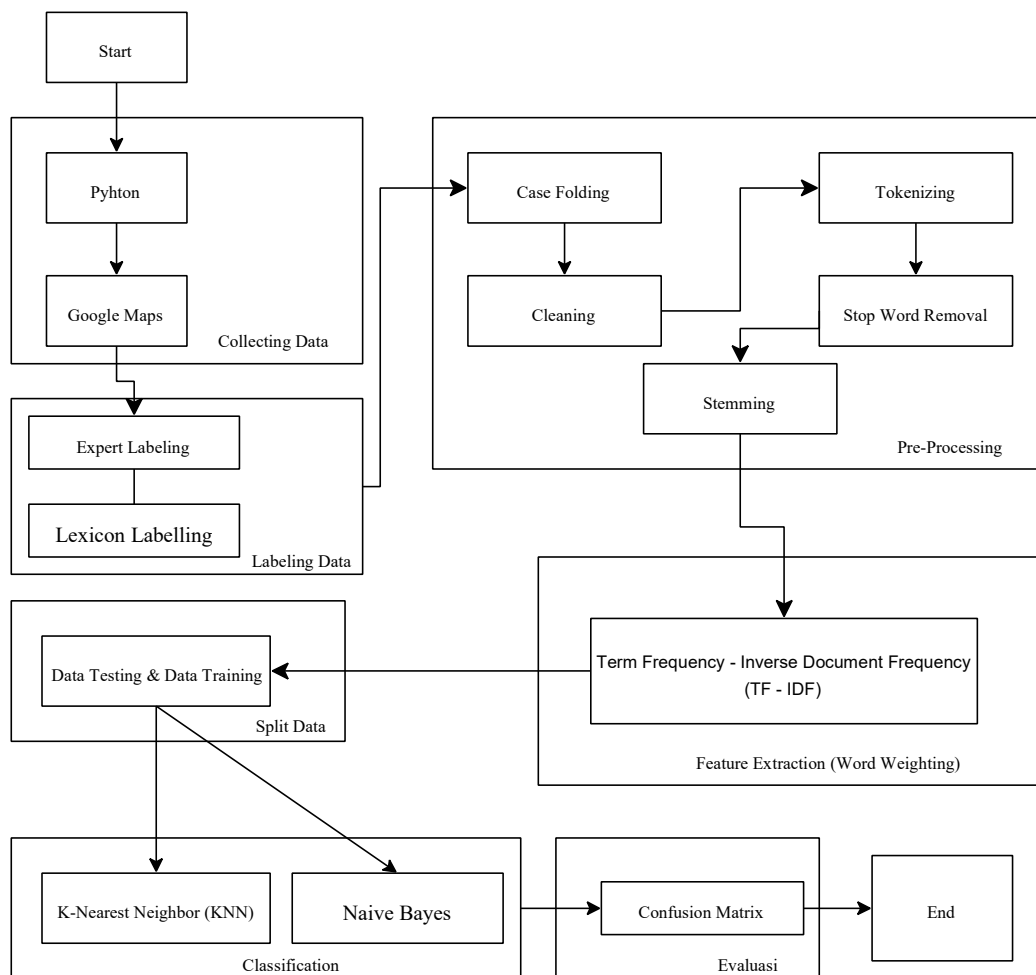
Teknik *web scraping* akan digunakan untuk mengumpulkan data ulasan pengguna, yang kemudian akan diproses melalui beberapa tahap seperti *Case Folding*, *Cleaning*, *Tokenizing*, *Stopword Removal*, dan *Stemming* sebelum dilakukan klasifikasi. Data yang telah diproses kemudian akan dianalisis untuk mengklasifikasikan sentimen menjadi kelas positif, negatif, dan netral. Evaluasi model akan dilakukan menggunakan metrik akurasi. Penelitian ini diharapkan dapat memberikan wawasan yang berguna bagi peningkatan kualitas layanan aplikasi DANA berdasarkan analisis sentimen dari ulasan pengguna.

2.2. Alat Dan Bahan

Dalam penelitian ini, alat dan bahan yang digunakan meliputi perangkat keras Laptop dengan spesifikasi prosesor intel CORE i5 Gen 10, RAM 4GB, dan penyimpanan SSD 512GB. Untuk perangkat lunak digunakan Google Colab versi 1.0.0 (<https://colab.research.google.com/>) dengan Python versi 3.8.10. Library Python yang digunakan mencakup Google-play-scraper-1.2.7, Pandas versi 2.0.3, NumPy versi 1.25.2, NLTK versi 3.8.1, Scikit-learn versi 1.2.2, Matplotlib versi 3.7.1, serta Sastrawi versi 1.0.1.

2.3. Prosedur Penelitian

Dalam penelitian ini, dilakukan teknik *web scraping* pada Google Play Store dengan menggunakan bahasa pemrograman Python kemudian mengambil link aplikasi DANA untuk proses analisis lebih lanjut. Metode yang dipakai dalam penelitian ini bertujuan untuk melakukan analisis sentimen terhadap data pengguna aplikasi DANA di Google Play Store dengan membandingkan dua metode, yaitu K-Nearest Neighbor dan Naive Bayes. Diagram alur penelitian pada Gambar 2.1.



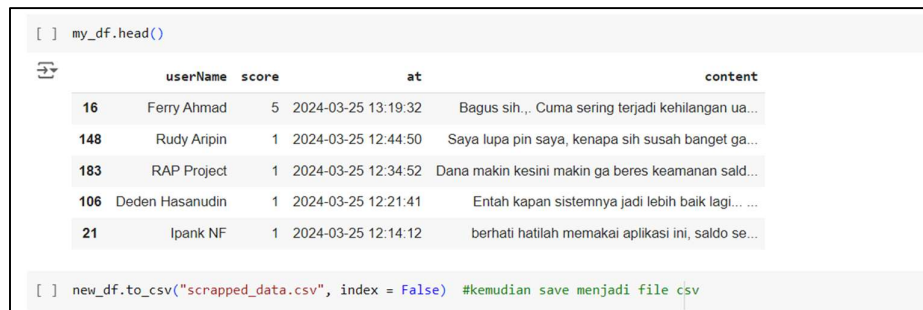
Gambar 2.1 Alur Penelitian

Pada Metode Penelitian terdiri dari beberapa tahapan utama, mulai dari Pengumpulan data melalui *scrapping* menggunakan Python, Labeling Data oleh ahli bahasa (*expert*)

dan menggunakan model *lexicon based*, Pre-processing, Pembobotan kata (TF-IDF), Pembagian data (*Split data*), Klasifikasi menggunakan metode KNN dan Naive Bayes, serta Evaluasi kinerja model menggunakan *confusion matriks*. Setiap tahapan akan dijelaskan secara rinci dengan memaparkan teori, konsep, dan algoritma yang digunakan.

2.3.1. Pengumpulan Data (*Scrapping*)

Dalam penelitian ini pengumpulan data dengan menggunakan metode teknik *web scraping* yang bersumber dari kolom komentar pada Google Play Store DANA. dengan menggunakan metode *web scraping* yaitu suatu kemampuan untuk pengambilan data dalam jumlah besar dari sebuah website. Proses *scrapping* dilakukan dengan mengambil label nama, waktu, komentar, dan jumlah data yang diambil sebanyak 700 data. Setelah proses *scrapping* selesai, Data yang berhasil diekstrak dapat disimpan dalam format file CSV atau Excel (Saputro et al., 2023) Hasil crawling data pada Gambar 2.2.



```
[ ] my_df.head()
```

	userName	score	at	content
16	Ferry Ahmad	5	2024-03-25 13:19:32	Bagus sih,.. Cuma sering terjadi kehilangan ua...
148	Rudy Aripin	1	2024-03-25 12:44:50	Saya lupa pin saya, kenapa sih susah banget ga...
183	RAP Project	1	2024-03-25 12:34:52	Dana makin kesini makin ga beres keamanan sald...
106	Deden Hasanudin	1	2024-03-25 12:21:41	Entah kapan sistemnya jadi lebih baik lagi... ..
21	lpank NF	1	2024-03-25 12:14:12	berhati hatilah memakai aplikasi ini, saldo se...

```
[ ] new_df.to_csv("scrapped_data.csv", index = False) #kemudian save menjadi file csv
```

Gambar 2.2 Hasil Pengumpulan Data

2.3.2. Labelling Data

Dalam proses klasifikasi teks pada data komentar untuk skripsi tugas akhir, peneliti memerlukan ahli bahasa yang berpengalaman dalam pelabelan data dan memiliki pengetahuan mendalam tentang bahasa Indonesia. Untuk itu, peneliti mengajukan permintaan di *website* project.co.id untuk mencari tenaga ahli yang memenuhi kriteria tersebut. Dalam pengajuan ini, peneliti menjelaskan bahwa diperlukan lulusan jurusan Bahasa Indonesia yang saat ini bekerja di bidang terkait, seperti guru/dosen bahasa Indonesia, penulis, atau ahli bahasa. Calon tenaga ahli diminta untuk mengajukan penawaran dengan menyertakan pekerjaan saat ini, pengalaman relevan dengan bahasa Indonesia, serta gelar akademik yang dimiliki. Hasil pelabelan data *expert* pada Gambar 2.3.

1	User Nama	Waktu	Komentar	SENTIMEN
2	Nitong99	2024-03-25 08:47:39	Buat APLIKASI DANA kebanyakan update, tapi masih lola lambat saat login m.	NEGATIF
3	Suci Afriany	2024-03-24 03:52:21	Aplikasinya membantu, tapi cape setiap mau transaksi system error, dan nyo	NEGATIF
4	M Hasan	2024-03-18 02:34:05	Lagi dan lagi, di saat pembayaran e-wallet ke shopeepay bermasalah lagi ales:	NEGATIF
5	DMS ALFARIZI	2024-03-15 03:32:58	Untuk fitur seperti transfer,nabung,inves,transaksi dll semua sudah oke men	POSITIF
6	Agung Nursetyav	2024-03-18 14:08:24	Brengsek isi saldo malah gak bisa dipake gara² minta upgrade premium. Coba	NEGATIF
7	Y	2024-03-22 17:54:04	Sebenarnya aplikasinya sudah bagus enak buat transaksi gak pake lama udah	NEGATIF
8	Vemi Love	2024-03-23 04:23:27	Sebelum update tidak ada masalah dan sangat membantu dalam bertransaks	NEGATIF

Gambar 2.3 Pelabelan Data Expert

Setelah mendapatkan hasil pelabelan data menggunakan *expert*, peneliti juga menggunakan model *lexicon based* untuk memberikan label secara otomatis. *Lexicon* adalah metode yang menggunakan pendekatan kamus untuk menganalisis sentimen. Kamus *lexicon* memproses bobot kata berdasarkan kamus yang memudahkan klasifikasi otomatis kalimat yang akan dianalisis. Oleh karena itu, *lexicon* dapat digunakan untuk melabeli data sebelum dilakukan pelatihan model (Fernanda & Fathoni, 2024). Hasil pelabelan data *lexicon* pada Gambar 2.4.

1	User Nama	Waktu	Komentar	Positif	Negatif	Netral	Compound sentiment
2	Nitong99	2024-03-25 08:47:39	Buat APLIKASI DANA	0,112	0	0,888	0,6981 Positif
3	Suci Afriany	2024-03-24 03:52:21	Aplikasinya membar	0	0,153	0,847	-0,6597 Negatif
4	M Hasan	2024-03-18 02:34:05	Lagi dan lagi, di saat	0	0	1	0 Netral
5	DMS ALFARIZI	2024-03-15 03:32:58	Untuk fitur seperti tr	0	0	1	0 Netral

Gambar 2.4 Pelabelan Data Lexicon

2.3.3. *Pre-Processing*

Sebelum melakukan *text mining* dokumen teks perlu disiapkan terlebih dahulu agar dapat digunakan dalam proses utama persiapan ini dikenal sebagai tahap *text preprocessing*, *Text preprocessing* berperan dalam mengubah data mentah yang tidak terstruktur menjadi terstruktur (Khairunnisa et al., 2021). Langkah-langkah yang dilakukan dalam tahap *preprocessing* (Dwiki et al., 2021) :

- a) *Case Folding* adalah langkah dalam pemrosesan teks yang bertujuan untuk mengubah semua karakter menjadi huruf kecil.
- b) *Cleaning* adalah proses membersihkan dokumen dan menyeleksi kata yang tidak diperlukan seperti menghapus HTML, emoticon, hashtag, mention, dan URL dari teks.
- c) *Tokenizing* adalah proses memecah teks menjadi token, yang bisa berupa kata-kata atau bahkan karakter tergantung pada kebutuhan. Selama proses tokenisasi, karakter pembatas seperti spasi dapat dihilangkan, angka dapat dihapus, dan tanda baca atau karakter non-huruf lainnya juga dapat dihapus, sesuai dengan kebutuhan pemrosesan teks yang diinginkan.
- d) *Stopwords* adalah proses menghapus kata-kata yang tidak memiliki makna atau relevansi penting.
- e) *Steaming* adalah Proses untuk menyaring kata-kata seperti kata sambung, kata ganti, dan kata depan, serta mengubahnya menjadi bentuk kata dasar dengan menghilangkan awalan atau akhiran.

2.3.4. **Pembobotan Kata (TF-IDF)**

Pada tahap ini, dilakukan pembobotan kata dalam teks menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Langkah ini bertujuan untuk menentukan bobot kata dalam sebuah dokumen atau seberapa sering kata tersebut

muncul dalam dokumen tersebut, TF-IDF adalah metode yang umum digunakan dalam analisis teks untuk mengevaluasi pentingnya sebuah kata dalam sebuah koleksi dokumen berdasarkan frekuensinya dalam dokumen dan dalam seluruh koleksi (Fadiyah Basar et al., 2022).

Penggunaan TF dapat menggunakan rumus pada Persamaan (2.1).

$$tf_{ij} = \frac{f_d(i)}{\max f_d(j)} \quad (2.1)$$

TF menunjukkan dokumen (d) seberapa banyak kata (t) yang muncul. Dan terkait untuk rumus IDF dapat dilihat pada Persamaan (2.2)

$$idf_t = \log \frac{N}{df_t} \quad (2.2)$$

N melambangkan jumlah kata dalam teks, df adalah jumlah teks yang memiliki kata t. Dengan menggabungkan TF dan IDF dalam pengerjaan dapat membantu meningkatkan performa. Terkait rumus pembobotan TF-IDF pada Persamaan (2.3).

$$W_{t,d} = tf_{d,t} \times idf_t \quad (2.3)$$

Keterangan :

t = Kata kunci, term

d = Dokumen

W_{d,t} = Bobot d terhadap t

Tf = Banyaknya t (kata) yang dicari dalam dokumen

Idf = Banyak t kebalikan dari kata yang dicari

2.3.5. Split Data

Pada tahap ini, split data adalah proses membagi dataset yang digunakan dalam penelitian menjadi dua bagian, dataset tersebut umumnya dibagi menjadi data latih

(*training*) dan data uji (*testing*). Data latih digunakan untuk melatih algoritma, sedangkan data uji digunakan untuk mengevaluasi kinerja algoritma tersebut (Putri et al., 2023). Dalam penelitian ini, dataset dibagi menjadi dua tahap dengan menggunakan rasio 90:10, 80:20, dan 70:30 untuk mengevaluasi pengaruh rasio tersebut terhadap kinerja model yang digunakan, yaitu Metode KNN dan Naïve Bayes dengan pembobotan fitur menggunakan TF-IDF. Hasil dari pembagian data ini menunjukkan bagaimana perbedaan rasio dapat mempengaruhi tingkat akurasi model dalam mengklasifikasikan sentimen.

2.3.6. Klasifikasi

Klasifikasi merupakan salah satu tahap penting dalam text mining yang bertujuan untuk mengelompokkan data atau objek baru ke dalam kelas atau label berdasarkan atribut-atribut tertentu. Proses ini melibatkan penggunaan teknik yang melihat variabel dari kelompok data yang sudah ada untuk menentukan aturan pengelompokan. Dengan mempelajari pola dari data yang sudah diberi label, klasifikasi memungkinkan kita untuk memprediksi kelas dari suatu objek yang belum diketahui sebelumnya (Azzahra Nasution et al., 2019)

2.3.7. K-Nearest Neighbor (KNN)

Dalam proses klasifikasi peneliti menggunakan dua metode yaitu KNN (K-Nearest Neighbors) dan Naive Bayes, KNN merupakan metode klasifikasi yang sederhana dan non-parametrik yang digunakan untuk mengklasifikasikan data. Meskipun algoritma ini sederhana, kinerjanya sangat baik dan menjadi parameter yang penting klasifikasi KNN membutuhkan metrik dan integer positif (K). Aturan KNN memegang posisi sampel pelatihan beserta kelas mereka ketika menghadapi data masuk baru, tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan nilai atribut dan data latih yang ada (Putra et al., 2022).

Berikut adalah langkah-langkah klasifikasi algoritma KNN:

- a) Tentukan parameter nilai k = banyaknya jumlah tetangga terdekat.
- b) Hitung jarak antara data *training* dan data *testing*, rumus dapat dilihat pada Persamaan (2.4).

$$euc = \sqrt{(\sum_{i=1}^n (p_i - q_i)^2)} \quad (2.4)$$

Keterangan :

p_i = sample data / data training

q_i = data uji / data testing

I = variabel data

n = dimensi data

- a) Urutkan jarak-jarak tersebut dan tetapkan tetangga terdekat berdasarkan berdasarkan jarak minimum hingga ke- k .
- b) Periksa kelas dari tetangga terdekat.
- c) Gunakan mayoritas sederhana dari kelas tetangga sebagai nilai prediksi untuk data baru.

2.3.8. Naive Bayes

Dan algoritma kedua yang digunakan untuk klasifikasi adalah Naive Bayes. Algoritma ini menggunakan perhitungan probabilitas dan statistik dalam pengoperasiannya. Metode klasifikasi Naive Bayes digunakan untuk mengambil keputusan dengan melakukan prediksi suatu kasus berdasarkan hasil dari klasifikasi yang telah diperoleh. Pada penelitian ini, metode Naive Bayes digunakan untuk menentukan sentimen dari data *Google Playstore* (Astari et al., 2020).

Berikut ini adalah langkah – langkah rumus Naive Bayes pada Persamaan (2.5).

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (2.5)$$

Keterangan :

A = Hipotesis data merupakan suatu *class* spesifik.

B = Data dengan kelas yang masih belum diketahui.

P(A|B) = Probabilitas hipotesis berdasar kondisi.

P(A) = Probabilitas hipotesis berdasar kondisi.

P(B|A) = Probabilitas berdasarkan kondisi pada hipotesis.

P(B) = Probabilitas B.

2.3.9. Evaluasi

Evaluasi digunakan untuk melihat kinerja dari model klasifikasi yang digunakan. Metode evaluasi yang digunakan adalah confusion matrix. Berdasarkan confusion matrix pengujian model dilakukan dengan menghitung nilai accuracy

Perhitungan untuk metode confusion matrix yang digunakan yaitu nilai skor Accuracy, Untuk mengukur seberapa sering model klasifikasi membuat prediksi yang benar (Khatib Sulaiman et al., 2023).

1. Akurasi

Akurasi dapat dijelaskan sebagai seberapa dekat nilai prediksi dengan nilai sebenarnya. Semakin tinggi *akurasi*, semakin baik proses klasifikasi tersebut.

Rumus untuk menghitung akurasi dalam Persamaan (2.6).

$$Akurasi = \frac{Juml \text{ Prediksi Benar}}{Total \text{ Prediksi}} \quad (2.6)$$

Pada rumus di atas, cara menentukan akurasi dari sebuah data dapat dilihat dengan menghitung jumlah prediksi yang benar dan membaginya dengan jumlah total prediksi yang dilakukan.

2.3.10. Jadwal Penelitian

Penelitian ini akan diawali dengan penentuan judul, identifikasi masalah, studi literatur, rancangan metode, pemilihan studi kasus, dan penyusunan proposal. Setelah tahap pra-penelitian selesai, penelitian akan melanjutkan ke pengumpulan data menggunakan metode *crawling* di *Google Play Store Aplikasi DANA*. Data yang terkumpul akan dilabeli oleh ahli Bahasa (*expert*) dan menggunakan model *lexicon* diproses sebelum pembobotan kata dengan metode TF-IDF. Selanjutnya data akan dibagi menjadi set pelatihan dan pengujian sebelum dilakukan klasifikasi menggunakan algoritma K – Nearest Neighbor (KNN) dan Naive Bayes. Tahap akhir adalah evaluasi dan analisis hasil untuk mengevaluasi efektivitas metode yang digunakan. Setelah selesai, akan dilakukan penyusunan laporan dan presentasi seminar untuk memperkenalkan hasil penelitian. Penelitian akan dilaksanakan sesuai jadwal terperinci pada Tabel 2.1 Jadwal Penelitian.

Tabel 2.1 Jadwal Penelitian

No	Kegiatan	Bulan/2024					
		Feb	Mar	Apr	Mei	Juni	Juli
Tahap Pra Penelitian							
1	Menentukan Judul						
2	Identifikasi Masalah						
3	Studi Literatur						
4	Rancangan Metode						
5	Pemilihan Studi Kasus						
6	Menyusun Proposal						
7	Review Desk Simpel						
Tahap Penelitian							
1	Pengumpulan Data (Crawling)						
2	Labelling Data						
3	Pre-Processing Data						
4	Pembobotan Kata (TF-IDF)						
5	Split Data						
6	Klasifikasi KNN dan Naive Bayes						
7	Evaluasi dan Analisis Hasil						
Tahap Akhir Penelitian							
1	Penyusunan Laporan						

No	Kegiatan	Bulan/2024					
		Feb	Mar	Apr	Mei	Juni	Juli
2	Seminar Hasil						

BAB III

HASIL DAN PEMBAHASAN

3.1. Hasil Pengumpulan Data

Dalam penelitian ini, data ulasan tentang hasil Aplikasi DANA yang diambil dari media Google PlayStore. Sebanyak 700 ulasan yang berhasil dikumpulkan dan disimpan dalam format CSV. Proses pengumpulan data dilakukan melalui metode *crawling* pada tanggal 27 Maret 2024. Hasil pengumpulan data pada Tabel 3.1.

Tabel 3.1 Hasil Pengumpulan Data

No.	User Nama	Waktu	Komentar
1.	Nitong99	2024-03-25 08:47:39	Buat APLIKASI DANA kebanyakan update, tapi masih lola lambat saat login maupun bertransaksi super lola, tolong lah update itu kan pembaruan, harusnya tingkatkan untuk kecepatan nya,, hapus fitur fitur yang memberatkan di Aplikasi Dana yang tidak Penting... Segera ditingkatkan!!!
2.	Suci Afriany	2024-03-24 03:52:21	Aplikasinya membantu, tapi cape setiap mau transaksi system error, dan nyoba buat upgrade baru bisa, ini masalahnya setiap mau transaksi yg error harus upgrade dulu. Masa iya upgrade trus trusan, tolong diperbaiki.
3.	M Hasan	2024-03-18 02:34:05	Lagi dan lagi, di saat pembayaran e-wallet ke shopeepay bermasalah lagi alesannya sistemnya bermasalah, terus waktu di tinggal sebentar malah "sesi kamu kadaluarsa, silahkan coba lagi" Saya coba ikutin cara developer di bawah ini ,sampai hapus data pun tidak bisa di gunakan tetap menyatakan sistem bermasalah, dari malam hingga pagi saya coba berulang kali kembali hasilnya tetap sama aja sistem bermasalah. Terakhir bisa minggu kemarin dan itu saya ulang" juga di jam 4 subuh

No.	User Nama	Waktu	Komentar
4.	DMS ALFARIZI	2024-03-15 03:32:58	Untuk fitur seperti transfer,nabung,inves,transaksi dll semua sudah oke menurutku, tapi tolong tambahkan fitur keamanan pas masuk ke apk dananya tolong buat fitur pin jadi pas masuk apk tidak langsung ke menu tampilan harus melalui pin untuk masuk terlebih dahulu supaya keamanannya makin oke, karna jika tidak nanti takutnya di hp kita otak atik sama org lain(hp dicuri) atau keluarga(anak) di apk dana kita
5.	Agung Nursetyawan	2024-03-18 14.08.24	Brengsek isi saldo malah gak bisa dipake gara ² minta upgrade premium. Coba upgrade malah gagal mulu padahal data sudah sesuai. Tolonglah kalo bikin aplikasi jangan kaya gini merugikan banyak orang. Permudah akses dan cara pemakaiannya gak perlu upgrade premium dulu baru bisa buat transaksi. Contoh noh aplikasi sebelah yang gak ribet harus upgrade premium segala.
Total			700

3.2. Hasil Labelling Data

Pelabelan data dilakukan oleh ahli bahasa (*expert*) dan model *lexicon based* untuk 700 data ulasan di Google Play Store. Dalam penelitian ini, ulasan dikategorikan menjadi tiga kelompok utama yaitu positif, negatif, dan netral. Dari data yang telah dilabeli *expert*, sebagian besar ulasan menunjukkan sentimen Negatif dan data yang diberi label *lexicon* sebagian besar ulasan menunjukkan Netral. Hasil pelabelan pada Tabel 3.2 dan 3.3.

Tabel 3.2 Hasil Labelling Data Expert

No.	Komentar	Sentimen
1.	Buat APLIKASI DANA kebanyakan update, tapi masih lola lambat saat login maupun bertransaksi super lola, tolong lah update itu kan pembaruan, harusnya tingkatkan untuk kecepatannya,, hapus fitur fitur yang memberatkan di Aplikasi Dana yang tidak Penting... Segera ditingkatkan!!	NEGATIF

No.	Komentar	Sentimen
2.	Aplikasinya membantu, tapi cape setiap mau transaksi system error, dan nyoba buat upgrade baru bisa, ini masalahnya setiap mau transaksi yg error harus upgrade dulu. Masa iya upgrade trusan, tolong diperbaiki.	NEGATIF
3.	Lagi dan lagi, di saat pembayaran e-wallet ke shopeepay bermasalah lagi alesannya sistemnya bermasalah, terus waktu di tinggal sebentar malah "sesi kamu kadaluarsa, silahkan coba lagi" Saya coba ikutin cara developer di bawah ini ,sampai hapus data pun tidak bisa di gunakan tetap menyatakan sistem bermasalah, dari malam hingga pagi saya coba berulang kali kembali hasilnya tetap sama aja sistem bermasalah. Terakhir bisa minggu kemarin dan itu saya ulang" juga di jam 4 subuh	NEGATIF
4.	Untuk fitur seperti transfer,nabung,inves,transaksi dll semua sudah oke menurutku, tapi tolong tambahkan fitur keamanan pas masuk ke apk dananya tolong buat fitur pin jadi pas masuk apk tidak langsung ke menu tampilan harus melalui pin untuk masuk terlebih dahulu supaya keamanannya makin oke, karna jika tidak nanti takutnya di hp kita otak atik sama org lain(hp dicuri) atau keluarga(anak) di apk dana kita	POSITIF
5.	Brengsek isi saldo malah gak bisa dipake gara ² minta upgrade premium. Coba upgrade malah gagal mulu padahal data sudah sesuai. Tolonglah kalo bikin aplikasi jangan kaya gini merugikan banyak orang. Permudah akses dan cara pemakaiannya gak perlu upgrade premium dulu baru bisa buat transaksi. Contoh noh aplikasi sebelah yang gak ribet harus upgrade premium segala.	NEGATIF
Total		700

Tabel 3.3 Hasil Labelling Data Lexicon

No.	Komentar	Sentimen
1.	Buat APLIKASI DANA kebanyakan update, tapi masih lola lambat saat login maupun bertransaksi super lola, tolong lah update itu kan pembaruan, harusnya tingkatkan untuk kecepatan nya,, hapus fitur fitur yang memberatkan di Aplikasi Dana yang tidak Penting... Segera ditingkatkan!!	Positif
2.	Aplikasinya membantu, tapi cape setiap mau transaksi system error, dan nyoba buat upgrade baru bisa, ini masalahnya setiap mau transaksi yg error harus upgrade dulu. Masa iya upgrade trusan, tolong diperbaiki.	Negatif
3.	Lagi dan lagi, di saat pembayaran e-wallet ke shopeepay bermasalah lagi alesannya sistemnya bermasalah, terus waktu di	Netral

No.	Komentar	Sentimen
	tinggal sebentar malah "sesi kamu kadaluarsa, silahkan coba lagi" Saya coba ikutin cara developer di bawah ini ,sampai hapus data pun tidak bisa di gunakan tetap menyatakan sistem bermasalah, dari malam hingga pagi saya coba berulang kali kembali hasilnya tetap sama aja sistem bermasalah. Terakhir bisa minggu kemarin dan itu saya ulang" juga di jam 4 subuh	
4.	Untuk fitur seperti transfer,nabung,inves,transaksi dll semua sudah oke menurutku, tapi tolong tambahkan fitur keamanan pas masuk ke apk dananya tolong buat fitur pin jadi pas masuk apk tidak langsung ke menu tampilan harus melalui pin untuk masuk terlebih dahulu supaya keamanannya makin oke, karna jika tidak nanti takutnya di hp kita otak atik sama orng lain(hp dicuri) atau keluarga(anak) di apk dana kita	Netral
5.	Brengsek isi saldo malah gak bisa dipake gara ² minta upgrade premium. Coba upgrade malah gagal mulu padahal data sudah sesuai. Tolonglah kalo bikin aplikasi jangan kaya gini merugikan banyak orang. Permudah akses dan cara pemakaiannya gak perlu upgrade premium dulu baru bisa buat transaksi. Contoh noh aplikasi sebelah yang gak ribet harus upgrade premium segala.	Netral
Total		700

3.3. Hasil *Pre-Processing*

Sebelum data ulasan Aplikasi DANA digunakan, tahap *Pre-processing* dilakukan untuk mendapatkan data bersih. Tahapan yang dilakukan antara lain *Cleaning*, *Tokenizing*, *Case Folding*, *Stopword Removal*, dan *Stemming*. Berikut ini adalah sebelum dan setelah dilakukan *Pre-processing*.

3.3.1. *Case Folding*

Tahap pertama dalam preprocessing adalah *case folding*. *Case folding* adalah proses mengubah teks menjadi huruf kecil (*lowercase*) untuk menghilangkan variasi antara huruf besar dan huruf kecil dalam analisis teks. Hasil dari data *Case Folding* pada Tabel 3.4.

Tabel 3.4 Hasil Case Folding

No.	Case Folding
1.	buat aplikasi dana kebanyakan update, tapi masih lola lambat saat login maupun bertransaksi super lola, tolong lah update itu kan pembaruan, harusnya tingkatkan untuk kecepatan nya,, hapus fitur fitur yang memberatkan di aplikasi dana yang tidak penting... segera ditingkatkan!!!
2.	aplikasinya membantu, tapi cape setiap mau transaksi system error, dan nyoba buat upgrade baru bisa, ini masalahnya setiap mau transaksi yg error harus upgrade dulu. masa iya upgrade trus trusan, tolong diperbaiki.
3.	lagi dan lagi, di saat pembayaran e-wallet ke shopeepay bermasalah lagi alesannya sistemnya bermasalah, terus waktu di tinggal sebentar malah ""sesi kamu kadaluarsa, silahkan coba lagi"" saya coba ikutin cara developer di bawah ini ,sampai hapus data pun tidak bisa di gunakan tetap menyatakan sistem bermasalah, dari malam hingga pagi saya coba berulang kali kembali hasilnya tetap sama aja sistem bermasalah. terakhir bisa minggu kemarin dan itu saya ulang" juga di jam 4 subuh
4.	untuk fitur seperti transfer,nabung,inves,transaksi dll semua sudah oke menurutku, tapi tolong tambahkan fitur keamanan pas masuk ke apk dananya tolong buat fitur pin jadi pas masuk apk tidak langsung ke menu tampilan harus melalui pin untuk masuk terlebih dahulu supaya keamanannya makin oke, karna jika tidak nanti takutnya di hp kita otak atik sama orng lain(hp dicuri) atau keluarga(anak) di apk dana kita
5.	brengsek isi saldo malah gak bisa dipake gara-gara minta upgrade premium. coba upgrade malah gagal mulu padahal data sudah sesuai. tolonglah kalo bikin aplikasi jangan kaya gini merugikan banyak orang. permudah akses dan cara pemakaiannya gak perlu upgrade premium dulu baru bisa buat transaksi. contoh noh aplikasi sebelah yang gak ribet harus upgrade premium segala
Total	700

3.3.2. *Cleaning*

Tahap berikutnya adalah menghapus karakter-karakter yang tidak diperlukan dari teks. Langkah ini bertujuan untuk membersihkan teks dari elemen-elemen yang tidak relevan dan bisa mengganggu proses analisis. Ulasan dalam bahasa Indonesia seringkali mengandung simbol, kalimat atau kata yang tidak baku, angka, hashtag, dan tautan URL. Hasil dari data *Cleaning* ini pada Tabel 3.5.

Tabel 3.5 Hasil Cleaning

No.	Cleaning
1.	buat aplikasi dana kebanyakan update tapi masih lola lambat saat login maupun bertransaksi super lola tolong lah update itu kan pembaruan harusnya tingkatkan untuk kecepatan nya hapus fitur fitur yang memberatkan di aplikasi dana yang tidak penting segera ditingkatkan
2.	aplikasinya membantu tapi cape setiap mau transaksi system error dan nyoba buat upgrade baru bisa ini masalahnya setiap mau transaksi yg error harus upgrade dulu masa iya upgrade trus trusan tolong diperbaiki
3.	lagi dan lagi di saat pembayaran wallet ke shopeepay bermasalah lagi alesannya sistemnya bermasalah terus waktu di tinggal sebentar malah sesi kamu kadaluarsa silahkan coba lagi saya coba ikutin cara developer di bawah ini sampai hapus data pun tidak bisa di gunakan tetap menyatakan sistem bermasalah dari malam hingga pagi saya coba berulang kali kembali hasilnya tetap sama aja sistem bermasalah terakhir bisa minggu kemarin dan itu saya ulang juga di jam subuh
4.	untuk fitur seperti transfer nabung inves transaksi dll semua sudah oke menurutku tapi tolong tambahkan fitur keamanan pas masuk ke apk dananya tolong buat fitur pin jadi pas masuk apk tidak langsung ke menu tampilan harus melalui pin untuk masuk terlebih dahulu supaya keamanannya makin oke karna jika tidak nanti takutnya di hp kita otak atik sama orng lain hp dicuri atau keluarga anak di apk dana kita
5.	brengsek isi saldo malah gak bisa dipake gara minta upgrade premium coba upgrade malah gagal mulu padahal data sudah sesuai tolonglah kalo bikin aplikasi jangan kaya gini merugikan banyak orang permudah akses dan cara pemakaiannya gak perlu upgrade premium dulu baru bisa buat transaksi contoh noh aplikasi sebelah yang gak ribet harus upgrade premium segala
Total	700

3.3.3. *Tokenizing*

Tahap berikutnya dalam *preprocessing* adalah *Tokenizing*. Proses ini melibatkan pemecahan kalimat menjadi kata-kata individual atau token. Tujuannya adalah untuk mempersiapkan teks agar bisa dianalisis pada tingkat kata. Hasil dari *tokenizing* pada Tabel 3.6.

Tabel 3.6 Hasil Tokenizing

No.	Tokenizing
1.	['buat', 'aplikasi', 'dana', 'kebanyakan', 'update', 'tapi', 'masih', 'lola', 'lambat', 'saat', 'login', 'maupun', 'bertransaksi', 'super', 'lola', 'tolong', 'lah', 'update', 'itu', 'kan', 'pembaruan', 'harusnya', 'tingkatkan', 'untuk', 'kecepatan', 'nya', 'hapus', 'fitur', 'fitur', 'yang', 'memberatkan', 'di', 'aplikasi', 'dana', 'yang', 'tidak', 'penting', 'segera', 'ditingkatkan']
2.	['aplikasinya', 'membantu', 'tapi', 'cape', 'setiap', 'mau', 'transaksi', 'system', 'error', 'dan', 'nyoba', 'buat', 'upgrade', 'baru', 'bisa', 'ini', 'masalahnya', 'setiap', 'mau', 'transaksi', 'yg', 'error', 'harus', 'upgrade', 'dulu', 'masa', 'iya', 'upgrade', 'trus', 'trusan', 'tolong', 'diperbaiki']
3.	['lagi', 'dan', 'lagi', 'di', 'saat', 'pembayaran', 'wallet', 'ke', 'shopeepay', 'bermasalah', 'lagi', 'alesannya', 'sistemnya', 'bermasalah', 'terus', 'waktu', 'di', 'tinggal', 'sebentar', 'malah', 'sesi', 'kamu', 'kadaluarsa', 'silahkan', 'coba', 'lagi', 'saya', 'coba', 'ikutin', 'cara', 'developer', 'di', 'bawah', 'ini', 'sampai', 'hapus', 'data', 'pun', 'tidak', 'bisa', 'di', 'gunakan', 'tetap', 'menyatakan', 'sistem', 'bermasalah', 'dari', 'malam', 'hingga', 'pagi', 'saya', 'coba', 'berulang', 'kali', 'kembali', 'hasilnya', 'tetap', 'sama', 'aja', 'sistem', 'bermasalah', 'terakhir', 'bisa', 'minggu', 'kemarin', 'dan', 'itu', 'saya', 'ulang', 'juga', 'di', 'jam', 'subuh']
4.	['untuk', 'fitur', 'seperti', 'transfer', 'nabung', 'inves', 'transaksi', 'dll', 'semua', 'sudah', 'oke', 'menurutku', 'tapi', 'tolong', 'tambahkan', 'fitur', 'keamanan', 'pas', 'masuk', 'ke', 'apk', 'dananya', 'tolong', 'buatkan', 'fitur', 'pin', 'jadi', 'pas', 'masuk', 'apk', 'tidak', 'langsung', 'ke', 'menu', 'tampilan', 'harus', 'melalui', 'pin', 'untuk', 'masuk', 'terlebih', 'dahulu', 'supaya', 'keamanannya', 'makin', 'oke', 'karna', 'jika', 'tidak', 'nanti', 'takutnya', 'di', 'hp', 'kita', 'otak', 'atik', 'sama', 'orng', 'lain', 'hp', 'dicuri', 'atau', 'keluarga', 'anak', 'di', 'apk', 'dana', 'kita']
5.	['brengek', 'isi', 'saldo', 'malah', 'gak', 'bisa', 'dipake', 'gara', 'minta', 'upgrade', 'premium', 'coba', 'upgrade', 'malah', 'gagal', 'mulu', 'padahal', 'data', 'sudah', 'sesuai', 'tolonglah', 'kalo', 'bikin', 'aplikasi', 'jangan', 'kaya', 'gini', 'merugikan', 'banyak', 'orang', 'permudah', 'akses', 'dan', 'cara', 'pemakaiannya', 'gak', 'perlu', 'upgrade', 'premium', 'dulu', 'baru', 'bisa', 'buat', 'transaksi', 'contoh', 'noh', 'aplikasi', 'sebelah', 'yang', 'gak', 'ribet', 'harus', 'upgrade', 'premium', 'segala']
Total	700

3.3.4. Stopword Removal

Tahap *Stopword Removal* digunakan untuk menghapus kata-kata yang tidak memiliki makna penting dalam teks, sehingga hanya kata-kata yang relevan dan

signifikan yang tersisa untuk dianalisis. Hasil dari data *Stopword Removal* pada Tabel 3.7

Tabel 3.7 Hasil Stopword Removal

No.	Stopword Removal
1.	['aplikasi', 'dana', 'kebanyakan', 'update', 'lola', 'lambat', 'login', 'bertransaksi', 'super', 'lola', 'tolong', 'update', 'pembaruan', 'tingkatkan', 'kecepatan', 'hapus', 'fitur', 'fitur', 'memberatkan', 'aplikasi', 'dana', 'ditingkatkan']
2.	['aplikasinya', 'membantu', 'cape', 'transaksi', 'system', 'error', 'nyoba', 'upgrade', 'transaksi', 'error', 'upgrade', 'iya', 'upgrade', 'trus', 'trusan', 'tolong', 'diperbaiki']
3.	['pembayaran', 'wallet', 'shopeepay', 'bermasalah', 'alesannya', 'sistemnya', 'bermasalah', 'tinggal', 'sebentar', 'sesi', 'kadaluarsa', 'silahkan', 'coba', 'coba', 'ikuti', 'developer', 'hapus', 'data', 'sistem', 'bermasalah', 'malam', 'pagi', 'coba', 'berulang', 'kali', 'hasilnya', 'sistem', 'bermasalah', 'minggu', 'kemarin', 'ulang', 'jam', 'subuh']
4.	['fitur', 'transfer', 'nabung', 'inves', 'transaksi', 'oke', 'menurutku', 'tolong', 'tambahkan', 'fitur', 'keamanan', 'pas', 'dananya', 'tolong', 'buatkan', 'fitur', 'pin', 'pas', 'langsung', 'menu', 'tampilan', 'pin', 'keamanannya', 'oke', 'karna', 'takutnya', 'hp', 'otak', 'atik', 'orng', 'hp', 'dicuri', 'keluarga', 'anak', 'dana']
5.	['brengek', 'isi', 'saldo', 'dipake', 'gara', 'upgrade', 'premium', 'coba', 'upgrade', 'gagal', 'mulu', 'data', 'sesuai', 'tolonglah', 'aplikasi', 'kaya', 'gini', 'merugikan', 'orang', 'permudah', 'akses', 'pemakaiannya', 'upgrade', 'premium', 'transaksi', 'contoh', 'noh', 'aplikasi', 'sebelah', 'ribet', 'upgrade', 'premium']
Total	700

3.3.5. Stemming

Tahap *stemming* adalah proses mengubah kata berimbuhan menjadi akar katanya. Pada tahap ini, akar kata dari setiap kata dicari menggunakan modul Sastrawi pada Python. Hasil dari *stemming* pada Tabel 3.8.

Tabel 3.8 Hasil Stemming

No	Stemming
1.	aplikasi dana banyak update lola lambat login transaksi super lola tolong update baru tingkat cepat hapus fitur fitur berat aplikasi dana tingkat

No.	Stemming
2.	<p>aplikasi bantu cape transaksi system error nyoba upgrade transaksi error</p> <p>upgrade iya upgrade trus trusan tolong baik</p> <p>bayar wallet shopeepay masalah alesannya sistem masalah tinggal sebentar sesi</p>
3.	<p>kadaluarsa silah coba coba ikutin developer hapus data sistem masalah malam</p> <p>pagi coba ulang kali hasil sistem masalah minggu kemarin ulang jam subuh</p> <p>fitur transfer nabung inves transaksi oke turut tolong tambah fitur aman pas</p>
4.	<p>dana tolong buat fitur pin pas langsung menu tampil pin aman oke karna takut</p> <p>hp otak atik orng hp curi keluarga anak dana</p> <p>brengsek isi saldo dipake gara upgrade premium coba upgrade gagal mulu data</p>
5.	<p>sesuai tolong aplikasi kaya gin rugi orang mudah akses pakai upgrade premium</p> <p>transaksi contoh noh aplikasi belah ribet upgrade premium</p>
Total	700

3.4. Hasil Pembobotan Kata (TF-IDF)

Pembobotan kata dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengidentifikasi kata-kata penting dalam ulasan aplikasi DANA di Google PlayStore. Dataset yang digunakan terdiri dari 700 baris dengan 2214 kolom, di mana setiap kolom mewakili kata unik dalam korpus data. Hasil yang ditampilkan dalam Gambar 3.1 TF-IDF menunjukkan bahwa sebagian besar kata memiliki bobot rendah dengan nilai 0, yang mengindikasikan bahwa kata-kata tersebut jarang muncul dalam dokumen.

	aaasssuuuuu	aaman	abai	abal	abdetan	abis	abistu	acc	account	\
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
..
695	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
696	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
697	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
698	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
699	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

	activate	...	yha	ynya	yotube	you	youtube	yt	ytub	yuk	zonk	\
0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
..
695	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
696	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
697	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
698	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
699	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

	Sentiment
0	Netral
1	Netral
2	Negatif
3	Positif
4	Negatif
..	...
695	Negatif
696	Negatif
697	Netral
698	Negatif
699	Positif

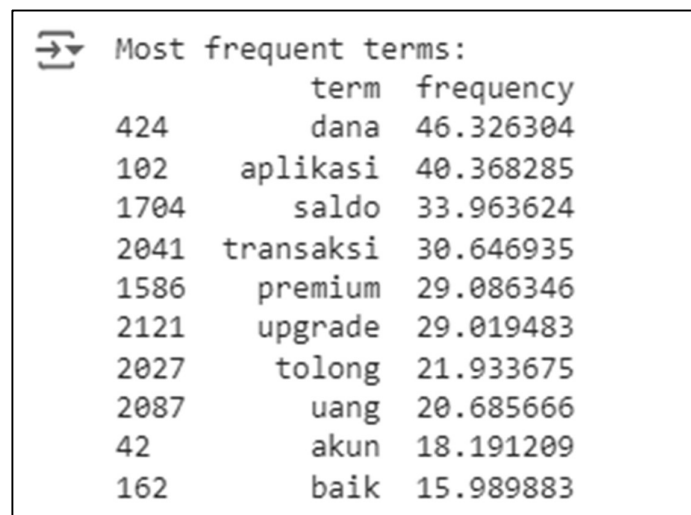
[700 rows x 2217 columns]

Gambar 3.1 Pembobotan Kata (TF-IDF)

Berdasarkan hasil skoring keseluruhan kata dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), dapat ditemukan kata-kata yang paling sering muncul dan memiliki bobot signifikan dalam korpus data. Gambar pada 3.2 menunjukkan bahwa kata "dana" memiliki *frekuensi* tertinggi dengan skor 46.326304, diikuti oleh kata "aplikasi" dengan skor 40.368285, dan "saldo" dengan skor 33.963624. Kata-kata lain yang juga memiliki bobot tinggi termasuk "transaksi" 30.646935, "premium" 29.086346, "upgrade" 29.019483, "tolong" 21.933675, "uang" 20.685666, "akun" 18.191209, dan "baik" 15.989883.

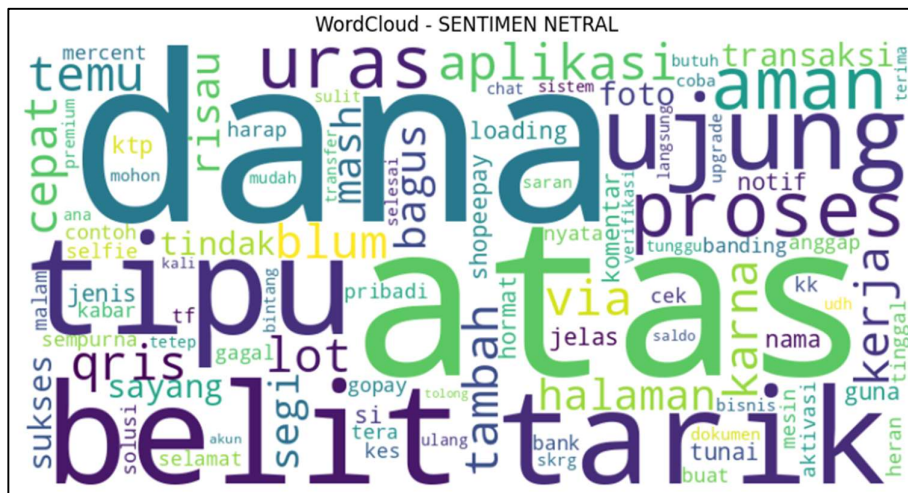
Frekuensi tinggi dari kata-kata ini menunjukkan topik yang dominan dan relevan dalam Google PlayStore ulasan pengguna aplikasi DANA. Kata "dana" sering muncul dalam dataset, menunjukkan bahwa topik pembicaraan banyak berkaitan dengan uang

atau sumber daya keuangan, kata "aplikasi" juga sering muncul, menunjukkan bahwa topik pembicaraan sering membahas tentang perangkat lunak atau aplikasi DANA, mungkin terkait dengan keuangan atau layanan teknologi, kemunculan kata "saldo" menunjukkan bahwa banyak diskusi berkaitan dengan saldo keuangan, kata "transaksi" menunjukkan bahwa ada banyak pembicaraan tentang aktivitas transaksi keuangan, seperti pembayaran, transfer uang, atau pembelian, kata "premium" dan "upgrade" dapat menunjukkan bahwa ada pembahasan tentang layanan atau produk premium yang mungkin melibatkan biaya tambahan atau fitur khusus, kata "tolong" mungkin muncul dalam konteks permintaan bantuan atau dukungan, menunjukkan bahwa pengguna sering meminta bantuan terkait layanan atau aplikasi, kemunculan kata "uang" menunjukkan bahwa banyak diskusi yang secara langsung berkaitan dengan uang, seperti pengelolaan uang atau masalah keuangan, kata "akun" sering muncul dalam konteks pembicaraan tentang akun pengguna, seperti pembuatan, pengelolaan, atau masalah dengan akun, kata "baik" bisa muncul dalam konteks evaluasi atau penilaian, seperti memberikan penilaian positif terhadap layanan atau aplikasi.



```
⇒ Most frequent terms:
      term frequency
424   dana 46.326304
102   aplikasi 40.368285
1704  saldo 33.963624
2041  transaksi 30.646935
1586  premium 29.086346
2121  upgrade 29.019483
2027  tolong 21.933675
2087  uang 20.685666
42    akun 18.191209
162   baik 15.989883
```

Gambar 3.2 Skoring TF



Gambar 3.5 WordCloud Netral

3.6. Hasil Evaluasi

Dalam tahap evaluasi peneliti menggunakan dua metode sebagai perbandingan. Evaluasi berfungsi untuk mengetahui akurasi dari model algoritma yang dibuat. Pada penelitian ini, evaluasi yang dilakukan adalah dengan menghitung akurasi. Data yang digunakan sebanyak 700 data. Sebelumnya peneliti melakukan *labelling* data menggunakan model *Lexicon Based*, Dari proses pelabelan tersebut didapat data kelas positif sebanyak 104, Data kelas negatif sebanyak 76 data, dan Data kelas netral sebanyak 520 data.

Kemudian peneliti juga melakukan *labelling* melalui *expert*, Data yang diberi label oleh ahli bahasa (*expert*) pada kolom sentimen terbagi menjadi 3 kelas dengan kelas positif, negatif, dan netral. Dari proses pelabelan tersebut didapat data kelas positif sebanyak 23, Data kelas negatif 672 data, dan Data kelas netral 5 data. Total dataset yang digunakan pada akhirnya berjumlah 700 data. Dataset tersebut akan dibagi menjadi data *training* dan data *testing* dengan menggunakan perbandingan rasio 90:10, 80:20, dan 70:30. Hasilnya di evaluasi menggunakan metrix evaluasi yaitu *accuracy* pada Tabel 3.9 dan 3.10.

Tabel 3.9 Pembagian Data KNN

Rasio	Akurasi
90:10	95%
80:20	95%
70:30	96%

Tabel 3.10 Pembagian Data Naive Bayes

Rasio	Akurasi
90:10	95%
80:20	95%
70:30	96%

Setelah dilakukan pengujian dengan 3 model rasio yaitu 90:10, 80:20, dan 70:30 didapatkan akurasi untuk masing masing rasio sebesar 95% untuk pengujian menggunakan 90:10, 95% untuk pengujian dengan rasio 80:20, dan 96% untuk pengujian dengan rasio 70:30. Dengan berbagai model rasio pengujian tersebut, menunjukkan bahwa model KNN dan Naive Bayes yang dilatih dengan teknik teks pre-processing yang benar serta pembobotan TF-IDF memberikan performa yang baik.

Hasil evaluasi menunjukkan bahwa model KNN dan Naive Bayes dengan rasio pengujian 70:30 mencapai akurasi sebesar 0.9619, Ini berarti model berhasil mengklasifikasikan sekitar 96.19% data dengan benar. Pada rasio pengujian 80:20, model menunjukkan performa dengan akurasi sebesar 0.9571, mengindikasikan bahwa sekitar 95.71% data diklasifikasikan dengan benar. Dan pada rasio pengujian 90:10, akurasi model menunjukkan akurasi yang sama 0.9571, yang berarti model mampu mengklasifikasikan sekitar 95.71% data dengan benar.

Perbedaan akurasi ini menunjukkan bahwa performa model sedikit bervariasi tergantung pada rasio pembagian data latih dan data uji. Pada rasio 70:30, model menunjukkan performa yang lebih baik dibandingkan dengan rasio 90:10 dan 80:20. Hal ini dapat disebabkan oleh berbagai faktor, seperti jumlah data uji yang lebih besar pada rasio 70:30, yang memungkinkan model untuk mempelajari lebih banyak pola dari data.

Sedangkan pada rasio 90:10 dan 80:20, jumlah data uji yang lebih sedikit mungkin tidak cukup untuk menangkap variasi yang ada dalam data, sehingga menyebabkan penurunan akurasi. Secara keseluruhan, hasil ini menunjukkan pemilihan rasio yang tepat dalam pembagian data latih dan data uji untuk mendapatkan performa model yang optimal. Pemilihan rasio yang baik dapat membantu model mempelajari pola dengan lebih baik dan menghasilkan prediksi yang lebih akurat (Kurniawan & Anubhakti, 2023) Dari ketiga rasio tersebut peneliti mengambil rasio yang terbaik yaitu 70:30.

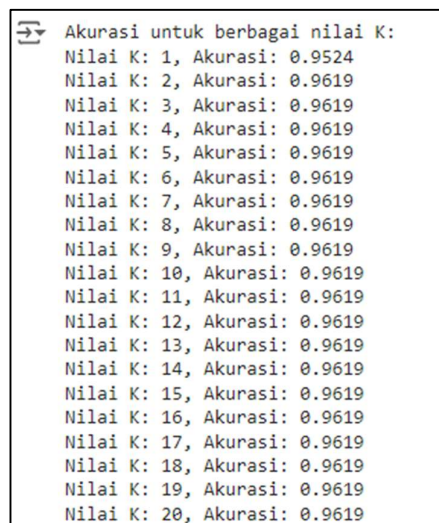
3.6.1. Metode K-Nearest Neighbor (KNN)

Pada metode KNN data yang digunakan yaitu sebanyak 700 data. Data tersebut dibagi menjadi dua kelompok yaitu data *training* (data latih) dan data *testing* (data uji) dengan menggunakan rasio 70:30 yang mendapatkan performa lebih baik. Sebanyak 490 data digunakan sebagai data *training* dan 210 data sebagai data *testing*.

Setelah menentukan rasio terbaik, dilakukan pengujian untuk mencari akurasi optimal pada berbagai nilai K dalam algoritma K-Nearest Neighbors (KNN). Jarak K dihitung menggunakan Euclidean Distance. Beberapa nilai K digunakan untuk membandingkan hasil akurasi dari masing-masing nilai. Uji ini dilakukan baik saat validasi maupun evaluasi akhir, untuk menentukan nilai K yang paling optimal di antara semua nilai yang diuji. Dengan membandingkan akurasi dari setiap nilai K, kita dapat menemukan nilai K yang memberikan kinerja terbaik untuk model tersebut (Herman et al., 2020)

Evaluasi akurasi model K-Nearest Neighbor (KNN) dengan berbagai nilai K, K1 menunjukkan akurasi 0.9524 dan K2 hingga K20 menunjukkan bahwa setiap nilai K menghasilkan akurasi yang sama, yaitu 0.9619. Hal ini menunjukkan bahwa model KNN memiliki performa yang sangat baik dan stabil pada dataset yang digunakan. Perubahan nilai K tidak mempengaruhi akurasi secara signifikan, karena

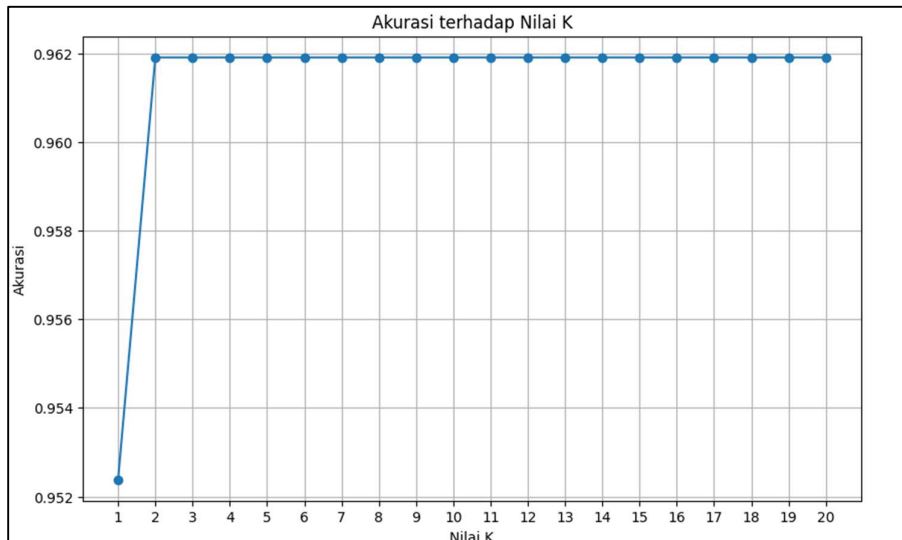
ketidakseimbangan kelas, di mana salah satu kelas memiliki jumlah sampel lebih banyak dibandingkan dengan kelas lainnya. Ketidakseimbangan ini bisa membuat model lebih mudah memprediksi kelas mayoritas dengan benar, sehingga meningkatkan akurasi keseluruhan tanpa benar-benar mencerminkan kemampuan model dalam memprediksi kelas minoritas dengan akurat (Nur Ariyanti & Cahya Wihandika, 2019) Hasilnya pada Gambar 3.6.



```
➔ Akurasi untuk berbagai nilai K:  
Nilai K: 1, Akurasi: 0.9524  
Nilai K: 2, Akurasi: 0.9619  
Nilai K: 3, Akurasi: 0.9619  
Nilai K: 4, Akurasi: 0.9619  
Nilai K: 5, Akurasi: 0.9619  
Nilai K: 6, Akurasi: 0.9619  
Nilai K: 7, Akurasi: 0.9619  
Nilai K: 8, Akurasi: 0.9619  
Nilai K: 9, Akurasi: 0.9619  
Nilai K: 10, Akurasi: 0.9619  
Nilai K: 11, Akurasi: 0.9619  
Nilai K: 12, Akurasi: 0.9619  
Nilai K: 13, Akurasi: 0.9619  
Nilai K: 14, Akurasi: 0.9619  
Nilai K: 15, Akurasi: 0.9619  
Nilai K: 16, Akurasi: 0.9619  
Nilai K: 17, Akurasi: 0.9619  
Nilai K: 18, Akurasi: 0.9619  
Nilai K: 19, Akurasi: 0.9619  
Nilai K: 20, Akurasi: 0.9619
```

Gambar 3.6 Hasil Akurasi Nilai K

Grafik yang menunjukkan hubungan antara nilai K dan akurasi dalam model K-Nearest Neighbor (KNN) menunjukkan bahwa sumbu horizontal mewakili nilai K, yang berkisar dari 1 hingga 20, sementara sumbu vertikal mewakili akurasi model. Dari grafik tersebut, terlihat K1 mendapatkan akurasi 0.952 dan K2 hingga K20 mendapatkan akurasi yang sama 0.962 untuk setiap nilai K yang diuji. Hasil ini pada Gambar 3.7.



Gambar 3.7 Grafik Akurasi Nilai K

Hasil evaluasi akurasi dari model K-Nearest Neighbor (KNN) menunjukkan bahwa K1 mencapai akurasi sebesar 0.952, sedangkan K2 hingga K20 menghasilkan akurasi yang sama, yaitu 0.962. Hal ini menunjukkan bahwa model KNN menunjukkan performa yang sangat baik dan stabil pada dataset yang digunakan, di mana perubahan nilai K tidak berpengaruh secara signifikan terhadap akurasi. Stabilitas ini mengindikasikan bahwa model KNN efektif dalam memprediksi label data uji dengan tingkat ketepatan yang konsisten. Meskipun terdapat perbedaan pada K1, nilai akurasi K2 hingga K20 tetap konsisten pada 0.962. Berdasarkan hasil ini, peneliti memilih nilai K=5 sebagai parameter final untuk model KNN.

Untuk mengevaluasi kinerja suatu model klasifikasi yang memprediksi sentimen dari teks. Evaluasi ini melibatkan perhitungan nilai-nilai seperti akurasi (*Accuracy*). Kemudian, dilakukan perhitungan persentase dari masing-masing kategori tersebut terhadap total jumlah data. *Akurasi* model dihitung berdasarkan prediksi yang dihasilkan pada data uji pada Gambar 3.8 dan 3.9.



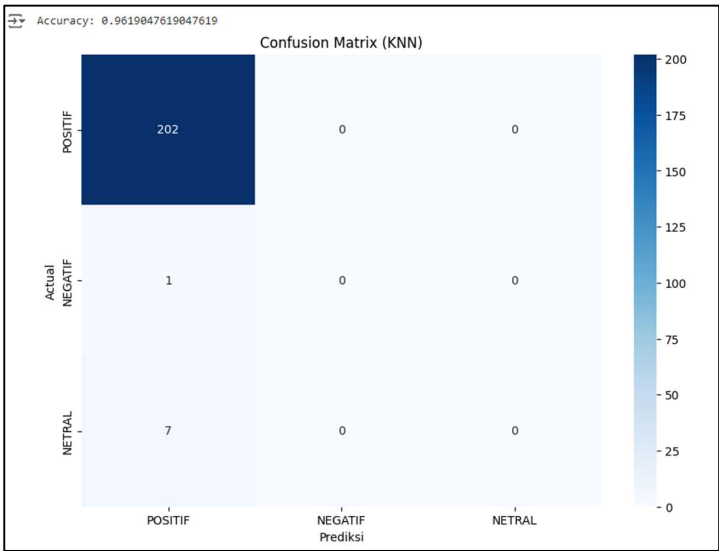
Gambar 3.8 Hasil Klasifikasi KNN Expert



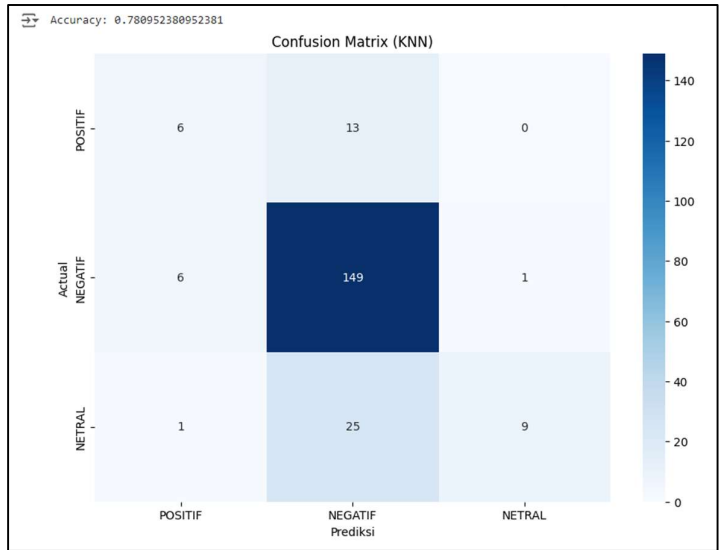
Gambar 3.9 Hasil Klasifikasi KNN Lexicon

Akurasi dari model diatas adalah akurasi yang digunakan oleh peneliti menggunakan data yang diberi label oleh (*expert*) sebesar 0.9619 atau 96%. Sebelumnya peneliti juga melakukan proses sampai tahap akurasi menggunakan data yang diberi label model *lexicon* dan mendapatkan akurasi sebesar 0.7809 atau 78%.

Untuk tampilan *Confusion Matrix* pada Gambar 3.10 dan 3.11.



Gambar 3.10 Hasil Confusion Matriks KNN Expert



Gambar 3.11 Hasil Confusion Matriks KNN Lexicon

Adapun perhitungan manual berdasarkan hasil Gambar 3.10 dan 3.11, pada Tabel 3.11 dan 3.12.

Tabel 3.11 Hasil Data Uji KNN Expert

Kategori	Prediksi Positif	Prediksi Negatif	Prediksi Netral
Aktual Positif	202	0	0
Aktual Negatif	1	0	0
Aktual Netral	7	0	0

Akurasi adalah rasio dari jumlah prediksi yang benar dan membaginya terhadap total jumlah prediksi.

Berikut adalah hasil perhitungan manual akurasi pada Persamaan (3.1).

$$Akurasi = \frac{202+0}{202+0+0+1+0+0+7+0+0} = \frac{202}{210} = 0.9619\% \quad (3.1)$$

Tabel 3.12 Hasil Data Uji KNN Lexicon

Kategori	Prediksi Positif	Prediksi Negatif	Prediksi Netral
Aktual Positif	6	13	0
Aktual Negatif	6	149	1
Aktual Netral	1	25	9

Berikut adalah hasil perhitungan manual akurasi pada Persamaan (3.2).

$$Akurasi = \frac{6+149+9}{6+13+0+6+149+1+1+25+9} = \frac{164}{210} = 0.7809\% \quad (3.2)$$

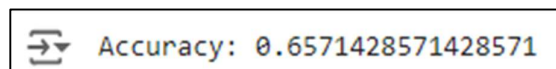
Hasil pengujian model menggunakan labelling *expert* menunjukkan akurasi sebesar 0.9619 atau 96% pada data uji, Berarti dari 210 data yang diuji model mampu memprediksi dengan benar 96 kasus. Dan hasil pengujian model menggunakan lexicon mendapatkan akurasi sebesar 0.7809 atau 78% pada data uji, Berati dari 210 data yang diuji model mampu memprediksi dengan benar 78 kasus. Akurasi ini dihitung berdasarkan jumlah prediksi yang benar dibagi dengan total jumlah kasus.

3.6.2. Metode Naive Bayes

Pada metode Naive Bayes data yang digunakan yaitu sebanyak 700 data. Data tersebut dibagi menjadi dua kelompok yaitu data *training* (data latih) dan data *testing* (data uji) dengan menggunakan rasio 70:30 yang mendapatkan performa lebih baik. Sebanyak 490 data digunakan sebagai data *training* dan 210 data sebagai data *testing*. *Akurasi* model dihitung berdasarkan prediksi yang dihasilkan pada data uji pada Gambar 3.12 dan 3.13.



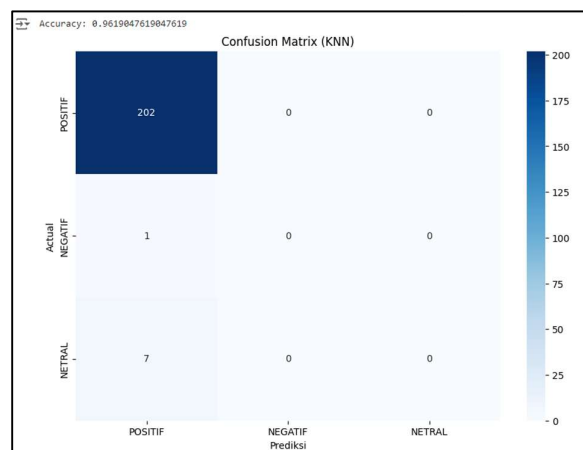
Gambar 3.12 Hasil Klasifikasi Naive Bayes Expert



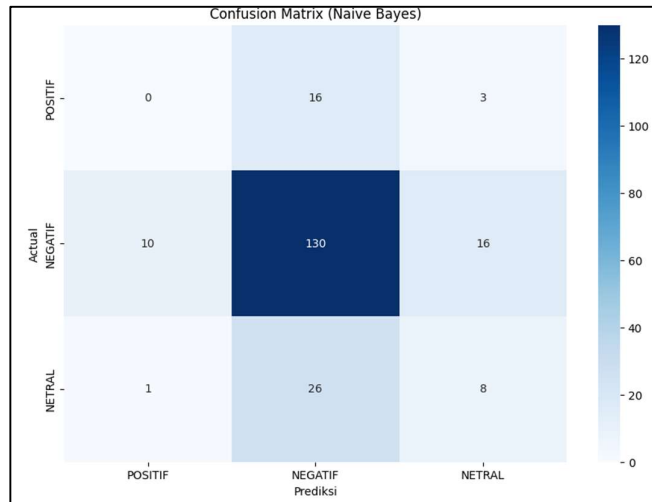
Gambar 3.13 Hasil Klasifikasi Naive Bayes Lexicon

Akurasi dari model diatas adalah akurasi yang digunakan oleh peneliti menggunakan data yang diberi label oleh (*expert*) sebesar 0.9619 atau 96%. Dan sebelumnya peneliti juga melakukan proses sampai tahap akurasi menggunakan data yang diberi label model *lexicon* dan mendapatkan akurasi sebesar 0.6571 atau 65%.

Untuk tampilan Confusion Matrix pada Gambar 3.14 dan 3.15.



Gambar 3.14 Hasil Confusion Matriks Naive Bayes Expert



Gambar 3.15 Hasil Confusion Matriks Naive Bayes Lexicon

Adapun perhitungan manual berdasarkan hasil Gambar 3.14 dan 3.15, pada Tabel 3.13 dan 3.14.

Tabel 3.13 Hasil Data Uji Naive Bayes Expert

Kategori	Prediksi Positif	Prediksi Negatif	Prediksi Netral
Aktual Positif	202	0	0
Aktual Negatif	1	0	0
Aktual Netral	7	0	0

Akurasi adalah rasio dari jumlah prediksi yang benar dan membagninya terhadap total jumlah prediksi.

Berikut adalah hasil perhitungan manual *akurasi* pada Persamaan (3.3).

$$Akurasi = \frac{202+0+0}{202+0+0+1+0+0+7+0+0} = \frac{202}{210} = 0.9619\% \quad (3.3)$$

Tabel 3.14 Hasil Data Uji Naive Bayes Lexicon

Kategori	Prediksi Positif	Prediksi Negatif	Prediksi Netral
Aktual Positif	0	16	3
Aktual Negatif	10	130	16
Aktual Netral	1	26	8

Berikut adalah hasil perhitungan manual *akurasi* pada Persamaan (3.4).

$$Akurasi = \frac{0+130+8}{0+16+3+10+130} = \frac{138}{210} = 0.6571\% \quad (3.4)$$

Hasil pengujian model menggunakan labelling *expert* menunjukkan *akurasi* sebesar 0.9619 atau 96% pada data uji, Berarti dari 210 data yang diuji model mampu memprediksi dengan benar 96 kasus. Dan hasil pengujian model menggunakan *lexicon* mendapatkan akurasi sebesar 0.6571 atau 65% pada data uji, Berati dari 210 data yang diuji model mampu memprediksi dengan benar 65 kasus. *Akurasi* ini dihitung berdasarkan jumlah prediksi yang benar dibagi dengan total jumlah kasus.

Dari kedua metode diatas terdapat perbedaan akurasi jika seorang peneliti menggunakan data yang sama tetapi menggunakan labelling yang berbeda. jika peneliti menggunakan data yang diberi labelling model *lexicon*, pada metode KNN mendapatkan akurasi sebesar 78% dan metode Naive Bayes mendapatkan akurasi sebesar 65%. Dan dapat dilihat terjadi perbedaan jika dibandingkan dengan menggunakan data yang diberi label oleh (*expert*) menggunakan metode KNN dan Naive Bayes mendapatkan akurasi yang sama sebesar 96%.

Perbedaan *akurasi* tersebut disebabkan karena terjadinya perubahan kelas sentimen yang diberi label oleh seorang ahli bahasa (*expert*) dan label menggunakan model *lexicon*. Terdapat penambahan pada kelas positif yang diberi label *lexicon*, dari 23 data kelas positif menjadi 104 data positif. Selanjutnya penambahan pada kelas netral dari 5 data kelas negatif menjadi 520 data netral. Dan menurunnya kelas negatif yang diberi label *expert*, dari 672 data kelas negatif menjadi 76 data netral. Hal inilah yang membuat terjadi perbedaan akurasi dari kedua metode tersebut.

BAB IV

PENUTUP

4.1. Simpulan

Penelitian ini membahas perbandingan akurasi metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam klasifikasi sentimen menggunakan data yang diberi label secara ahli bahasa (*expert*) dan menggunakan model *lexicon*. Pada proses *Pre-processing* melibatkan *Case Folding*, *Cleaning*, *Tokenizing*, *Stopword Removal*, dan *Stemming* terbukti efektif dalam mempersiapkan data teks untuk analisis. Penggunaan metode TF-IDF dalam pembobotan kata membantu mengidentifikasi kata-kata kunci yang relevan dari setiap ulasan, memungkinkan model untuk lebih memahami konteks dan makna di balik teks.

Hasil penelitian menunjukkan bahwa metode KNN dan Naive Bayes memiliki akurasi yang berbeda berdasarkan sumber label data. Pada data yang diberi label model *lexicon*, akurasi KNN mencapai 78% dan Naive Bayes 65%. Namun, ketika data diberi label oleh *expert*, akurasi kedua metode meningkat menjadi 96%. Perbedaan ini disebabkan oleh perubahan kelas sentimen saat pelabelan menggunakan model *lexicon* dibandingkan dengan pelabelan oleh *expert*, dimana data positif dan netral menurun tetapi pada data negatif mengalami peningkatan. Penelitian ini menyimpulkan bahwa kualitas pelabelan data sangat mempengaruhi akurasi metode klasifikasi yang digunakan, dan pelabelan oleh ahli bahasa (*expert*) dapat meningkatkan keakuratan dalam klasifikasi sentimen.

Selain itu, penelitian ini juga meneliti pengaruh rasio pembagian data dan nilai K pada kinerja model KNN. Hasilnya menunjukkan bahwa rasio pembagian data 70:30 memberikan akurasi terbaik sebesar 96%, dibandingkan dengan rasio 90:10 dan 80:20

yang masing-masing mencapai akurasi 95%. Dalam hal penggunaan nilai K, nilai K1 mencapai akurasi 95.24%, sementara nilai K2 hingga K20 mencapai akurasi yang sama yaitu 96.19%. Berdasarkan hasil ini, nilai K=5 dipilih sebagai parameter final karena memberikan kinerja optimal tanpa penurunan akurasi yang signifikan.

Penelitian ini menyimpulkan, bahwa pentingnya pemilihan metode, rasio data, dan parameter model yang tepat dalam meningkatkan akurasi klasifikasi sentimen. penggunaan pelabelan oleh *expert* dapat meningkatkan dan berkontribusi signifikan terhadap hasil yang diperoleh. dan memberikan informasi tentang efisiensi dan efektivitas dalam memilih metode klasifikasi yang lebih efektif dalam melakukan analisis sentimen yang sesuai dengan kebutuhan penelitian.

4.2.Implikasi

Penelitian ini memiliki beberapa implikasi penting. Pertama, penggunaan pelabelan data oleh ahli bahasa terbukti meningkatkan akurasi metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam klasifikasi sentimen, sehingga peneliti lain yang menggunakan analisis sentimen dapat mempertimbangkan pelibatan ahli bahasa untuk meningkatkan hasil klasifikasi. Hasil analisis sentimen pada ulasan pengguna aplikasi DANA memberikan wawasan mendalam mengenai keluhan dan masalah yang dihadapi pengguna, yang dapat digunakan untuk meningkatkan fitur dan layanan, sehingga meningkatkan kepuasan pengguna. Penelitian ini juga memberikan perbandingan antara metode KNN dan Naive Bayes dalam mengklasifikasikan sentimen, yang dapat menjadi referensi bagi peneliti lain dalam memilih metode yang sesuai untuk analisis sentimen pada dataset serupa. Penggunaan pembobotan kata dengan metode TF-IDF terbukti efektif dalam proses klasifikasi sentimen, menunjukkan bahwa metode ini dapat diterapkan dalam penelitian lain yang memerlukan analisis teks dengan akurasi tinggi. Akhirnya, pemahaman terhadap sentimen pengguna aplikasi DANA memungkinkan

masyarakat membuat keputusan yang lebih baik dalam memilih aplikasi pembayaran elektronik yang sesuai dengan kebutuhan mereka.

DAFTAR RUJUKAN

- Amrullah, A. Z., Sofyan Anas, A., Adrian, M., & Hidayat, J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, 2(1). <https://doi.org/10.30812/bite.v2i1.804>
- Astari, N. M. A. J., Dewa Gede Hendra Divayana, & Gede Indrawan. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27–29. <https://doi.org/10.30864/jsi.v15i1.332>
- Athallah Muhammad, A., Sandya Prasvita, D., Fatmawati Raya, J. R., & Labu, P. (2022). *ANALISIS SENTIMEN PENGGUNA APLIKASI DANA BERDASARKAN ULASAN PADA GOOGLE PLAY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE*.
- Azzahra Nasution, D., Khotimah, H. H., & Chamidah, N. (2019). *PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN* (Vol. 4, Issue 1).
- Badri, M. (2020). Adopsi Inovasi Aplikasi Dompot Digital di Kota Pekanbaru. In *Inovbiz: Jurnal Inovasi Bisnis* (Vol. 8). www.ejournal.polbeng.ac.id/index.php/IBP
- Cahyanti, D., Rahmayani, A., & Ainy Husniar, S. (2020). *Indonesian Journal of Data and Science Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara*. 1(2), 39–43.
- Dwiki, A., Putra, A., & Juanita, S. (2021). Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN. *JATISI*, 8(2). <http://jurnal.mdp.ac.id>
- Fadiyah Basar, T., Ratnawati, D. E., & Arwani, I. (2022). *Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest* (Vol. 6, Issue 3). <http://j-ptiik.ub.ac.id>
- Fernanda, M., & Fathoni, N. (2024). Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM. *Jurnal Informatika Dan Sains Teknologi*, 1(3), 62–76. <https://doi.org/10.62951/modem.v1i3.112>
- Herman, I. H., Widiyanto, D., & Ernawati, I. (2020). *PENGGUNAAN K-NEAREST NEIGHBOR (KNN) UNTUK MENGIDENTIFIKASI CITRA BATIK PEWARNA ALAMI DAN PEWARNA SINTETIS BERDASARKAN WARNA*.
- Heti Palestina Yunani. (2024, February 17). *4 Alasan Aman Pakai DANA, Memberikan Kenyamanan saat Bertransaksi*. Harian Disway.
- Khairunnisa, S., Adiwijaya, A., & Faraby, S. Al. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 406. <https://doi.org/10.30865/mib.v5i2.2835>
- Khatib Sulaiman, J., Dzakwan Ar Rosyid, M., & Artikel Abstrak, I. (2023). Klasifikasi Tingkat Risiko Kesehatan Ibu Hamil Menggunakan Algoritma Support Vectore Machine Universitas AMIKOM Yogyakarta. *Indonesian Journal of Computer Science Attribution*, 12(5), 2023–2798.
- Kurniawan, R., & Anubhakti, D. (2023). *Implementasi Algoritma Support Vector Machine Dalam Memprediksi Harga Saham PT. KRAKATAU STEEL TBK*. 2(2).
- Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). *Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest* (Vol. 6, Issue 9). <http://j-ptiik.ub.ac.id>
- Nur Ariyanti, R., & Cahya Wihandika, R. (2019). *Identifikasi Jenis Attention Deficit Hyperactivity Disorder (ADHD) Pada Anak Usia Dini Menggunakan Metode Modified K-Nearest Neighbor (MKNN)* (Vol. 3, Issue 1). <http://j-ptiik.ub.ac.id>
- Nurian, A., & Nurina Sari, B. (2023). *ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE PLAY MENGGUNAKAN NAÏVE BAYES*. *Jurnal Informatika*

- Dan Teknik Elektro Terapan*, 11(3), 2830–7062.
<https://doi.org/10.23960/jitet.v11i3%20s1.3348>
- Nurul, S., Fitriyyah, J., Safriadi, N., Esyudha, E., & #3, P. (2019). *Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes*. <http://dev.twitter.com>.
- Ogi, E., Pratiwi1, I., & Yustanti2, W. (2021). Analisis Sentimen Kualitas Layanan Teknologi Pembayaran Elektronik pada Twitter (Studi Kasus Ovo dan Dana). *JEISBI*, 02, 2021.
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naive Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646. <https://doi.org/10.32493/informatika.v5i4.7622>
- Putra, P., H Pardede, A. M., & Syahputra, S. (2022). ANALISIS METODE K-NEAREST NEIGHBOUR (KNN) DALAM KLASIFIKASI DATA IRIS BUNGA. *Jurnal Teknik Informatika Kaputama (JTIK)*, 6(1).
- Putri, A., Syaficha Hardiana, C., Novfujia, E., Try Puspa Siregar, F., Fatma, Y., & Wahyuni, R. (2023). Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir. *Institut Riset Dan Publikasi Indonesia (IRPI) MALCOM: Indonesian Journal of Machine Learning and Computer Science Journal Homepage*, 3(1), 20–26.
- Rifa, M. K., Totohendarto, M. H., & Muttaqin, M. R. (2023). Analisis Sentimen Pengguna E-Wallet Dana Dan Gopay Pada Twitter Menggunakan Metode Support Vector Machine (SVM). *IJCCS*, x, No.x, 1–5.
- Salehudin Basryah, E., Erfina, A., & Warman, C. (2021). *ANALISIS SENTIMEN APLIKASI DOMPET DIGITAL DI ERA 4.0 PADA MASA PENDEMI COVID-19 DI PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER*.
- Salim, S. S., & Mayary, J. (2020). ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP DOMPET ELEKTRONIK DENGAN METODE LEXICON BASED DAN K – NEAREST NEIGHBOR. *Jurnal Ilmiah Informatika Komputer*, 25(1), 1–17. <https://doi.org/10.35760/ik.2020.v25i1.2411>
- Saputro, W. E., Yuana, H., & Puspitasari, W. D. (2023). ANALISIS SENTIMEN PENGGUNA DOMPET DIGITAL DANA PADA KOLOM KOMENTAR GOOGLE PLAY STORE DENGAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 2).
- Shanty. (2024, February 5). *Capai Pertumbuhan Positif 2023 Hingga 23% DANA Sambut Optimis 2024*.

LAMPIRAN

L1 CV Expert Labelling Data

Irfan Abdul Hakim

Hirfan825@gmail.com/082135357602

PROFIL

Saya adalah seorang profesional yang memiliki keterampilan komunikasi yang baik, pemikiran terbuka, dan selalu bersemangat dalam mencari pengetahuan baru. Saya memiliki keinginan yang besar untuk terus belajar dan berkembang. Saya dengan cepat dapat menyesuaikan diri dengan perubahan dan memiliki kemampuan untuk beradaptasi dengan cepat dalam lingkungan kerja yang dinamis. Saya juga sangat berpengetahuan dalam penggunaan teknologi dan memiliki kemampuan untuk memanfaatkannya secara efektif. Saya mencari peluang untuk bekerja dalam lingkungan yang inovatif, di mana saya dapat menerapkan keterampilan analitis dan berpikir kritis saya untuk mencapai hasil yang optimal.

PENDIDIKAN

UNIVERSITAS GADJAH MADA

2015-2020

S1 Sosiologi, IPK: 3.40/4.00

S2 Sosiologi, IPK: 3.50/ 4.00

PENGALAMAN KERJA

SMA NURUL MUSLIM BATEALIT JEPARA

STAF PENGAJAR SOSIOLOGI

- Merancang Silabus pembelajaran di Awal Semester.
- Melakukan aktifitas belajar mengajar sesuai dengan rancangan awal silabus.
- Melakukan evaluasi hasil pembelajaran melalui penilaian harian, penilaian tengah semester dan penilaian akhir semester.

PANITIA PENGAWAS PEMILU (PANWASLU) KECAMATAN SEMBORO

2022 - sekarang

STAF PELAKSANA

- Membantu Tugas Pengawasan komisioner panwaslu kecamatan.
- Melakukan pendataan laporan dan temuan pelanggaran baik dari aduan masyarakat maupun temuan langsung Panwaslu.
- Melakukan Pengarsipan Surat Keluar Masuk Laporan dan Temuan Pelanggaran.

MTS ALI MAKSUM YOGYAKARTA

2021 - 2022

STAF PENGAJAR

- Merancang Silabus pembelajaran di Awal Semester.
- Melakukan aktifitas belajar mengajar sesuai dengan rancangan awal silabus.
- Melakukan evaluasi hasil pembelajaran melalui penilaian harian, penilaian tengah semester dan penilaian akhir semester.

PENGALAMAN ORGANISASI

GERAKAN MAHASISWA SATU BANGSA (GEMASABA) KAB. SLEMAN

2021 - 2022

WAKIL KETUA 1

- Membantu Ketua dalam Pembuatan, pelaksanaan serta pengawasan terhadap berjalannya program kerja.

- Menjalin hubungan baik dengan organisasi lain.
- Melakukan koordinasi dan inisiasi kerja sama dengan organisasi lain untuk kemajuan organisasi.

PENGURUS ASRAMA MTS PUTRA PONDOK PESANTREN KRAPYAK YAYASAN ALI MAKSUM 2015 - 2022

KOORDINATOR BENDAHARA

- Merancang, mengelola dan mendistribusikan Anggaran tahunan untuk tiga asrama
- Melakukan koordinasi dengan seluruh divisi terait kebutuhan anggaran dan pendistribusiannya.
- Melakukan koordinasi langsung dengan bendahara yayasan terkait dengan anggaran asrama.
- Melakukan Pelaporan rutin terhadap penggunaan anggaran kepada pengasuh pondok pesantren.

IKATAN ALUMNI MA ALI MAKSUM YOGYAKARTA ANGKATAN 2015

2018 - sekarang

KETUA

- Merancang serta merealisasikan program kerja tahunan.
- Menjaga hubungan baik dengan seluruh alumni.
- Melakukan koordinasi terkait agenda - agenda terdekat.

PERGERAKAN MAHASISWA ISLAM INDONESIA (PMII) GADJAH MADA

2018 - 2019

KOORDINATOR DIVISI KADERISASI

- Membuat dan menjalankan bersama tim program kerja yang telah disusun.
- Melakukan koordinasi dengan pengurus rayon divisi kaderisasi.
- Merancang strategi rekrutmen yang efektif dan tepat sasaran di awal tahun ajaran baru.

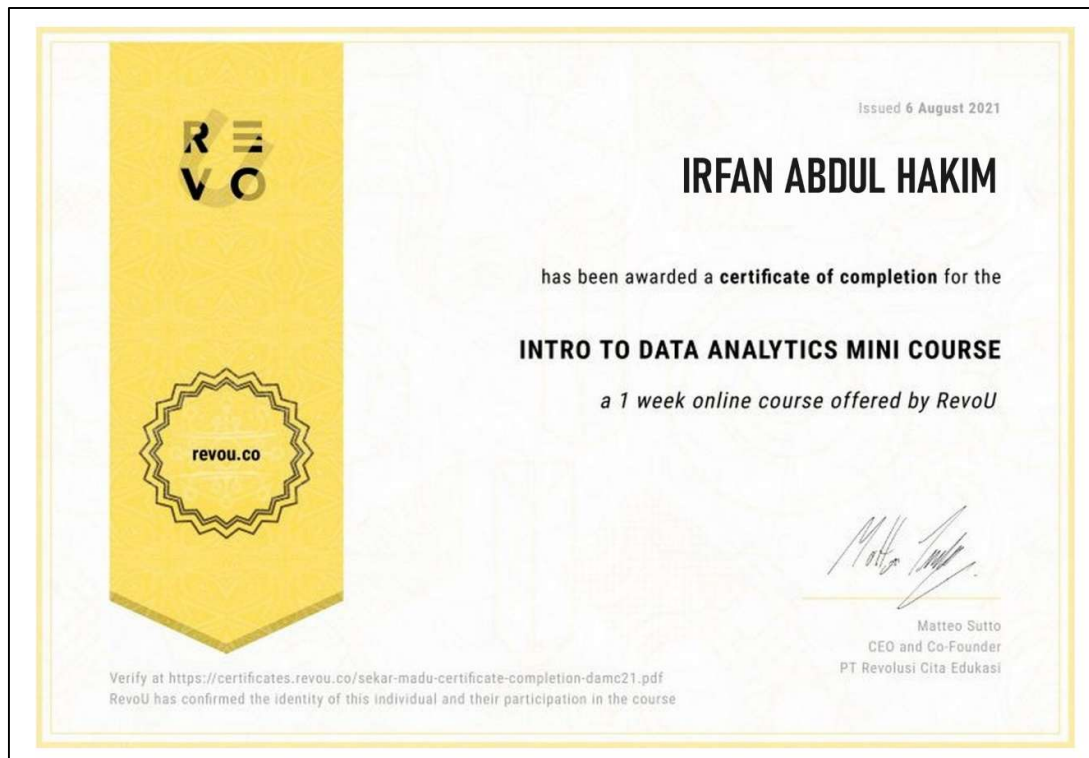
PENGALAMAN KESUKARELAAN

- **Survey Kepuasan Pelanggan PDAM Yogyakarta** - Yogyakarta (Februari 2022): Koordinator Enumerator
- **Seminar Nasional Menjadi Wirausahawan Sukses** - Yogyakarta (September 2020): Moderator
- **Penerimaan Santri Baru Yayasan Ali Maksu** - Yogyakarta (Januari - Februari 2020): Divisi Logistik
- **Riset Pusat Studi Pancasila Universitas Pembangunan Nasional Yogyakarta** - Yogyakarta (Agustus- September 2019): Enumerator
- **Pelatihan Kader Dasar PMII Gajah Mada** - (Maret 2018): Sterring Comitte
- **Pekan Raya Anak** - Yogyakarta (Desember 2017): Bendahara
- **Kulo Nuwun Party** - Yogyakarta (Maret 2016): Divisi Logistik

PENGALAMAN RISET

- Penulisan Buku Jusuf Kalla: Sang Pendamai Ulung
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci Kopi Sianida
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci IKN
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci Bitcoin Halving
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci Hak Angket
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci Quick Count

L2 Sertifikat Expert





#DQLABPRJ8 JIWHNQ

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

IRFAN ABDUL HAKIM

Has Completed in
Project Data Analysis for Finance: Performa Cabang

Sep 8, 2020

L3 Surat Ijin Penelitian



UMKKT
Program Studi
Teknik Informatika
Fakultas Sains dan Teknologi

Telp. 0541-748511 Fax. 0541-766832

Website <http://informatika.umkt.ac.id>

email: informatika@umkt.ac.id



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Nomor : 055-003/KET/FST.1/A/2024

Lampiran : -

Perihal : **Keterangan Pengambilan Data Sekunder**

Assalamu 'alaikum Warrahmatullahi Wabarrakatuh

Puji Syukur kepada Allah Subhanahu wa ta'ala yang senantiasa melimpahkan Rahmat-Nya kepada kita sekalian. Amin.

Dengan surat ini, kami menerangkan bahwa mahasiswa berikut:

Nama : Muhammad Rayhan Elfansyah

NIM : 2011102441075

Program Studi : Teknik Informatika

Melakukan penelitian dengan pengambilan data sekunder di Google Playstore data yang diambil yaitu data ulasan pengguna aplikasi DANA.

Demikian hal ini disampaikan, atas kerjasamanya kami ucapkan terima kasih.

Wassalamu 'alaikum Warrahmatullahi Wabarrakatuh

Samarinda, 20 Dzulhijjah 1445 H

27 Juni 2024 M

Dua Program Studi S1 Teknik Informatika



Elfansyah
Elfansyah, S.Kom., M.TI
IDN. 1118019203

Kampus 1 : Jl. Ir. H. Juanda, No.15, Samarinda
Kampus 2 : Jl. Pelita, Pesona Mahakam, Samarinda

L4 Code Pengumpulan Data

```
#Referensi: https://www.linkedin.com/pulse/how-scrape-google-play-reviews-4-simple-steps-using-python-kundi/
#download library google-play-scraper
!pip install google-play-scraper

Collecting google-play-scraper
  Downloading google_play_scraper-1.2.7-py3-none-any.whl (28 kB)
Installing collected packages: google-play-scraper
Successfully installed google-play-scraper-1.2.7

[ ] from google_play_scraper import app

import pandas as pd

import numpy as np

[ ] #scrape jumlah ulasan yang diinginkan
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'id.dana',
    lang='id', #disini kita mau men scrape data ulasan aplikasi shopee yang berada di google play store
    country='id', #kita setting bahasa nya menjadi bahasa indonesia
    sort=Sort.MOST_RELEVANT, # # kemudian kita gunakan most_relevan untuk mendapatkan ulasan yang paling relevant
    count=700, # disini jumlah ulasan yang mau kita ambil ada seribu
    filter_score_with=None # # kemudian di filter_score kita gunakan None untuk mengambil semua score atau rating bintang 1 sampai 5
)
```

```
[ ] df_busu = pd.DataFrame(np.array(result), columns=['review'])

df_busu = df_busu.join(pd.DataFrame(df_busu.pop('review').tolist()))

df_busu.head()
```

```
[ ] len(df_busu.index) #kemudian hitung kembali berapa jumlah data yg didapatkan

700

[ ] df_busu[['userName', 'score', 'at', 'content']].head() #preview userName, rating, date-time, and reviews only
```

```
[ ] #Run This Code to Sort the Data By Date

new_df = df_busu[['userName', 'score', 'at', 'content']]
sorted_df = new_df.sort_values(by='at', ascending=False) #Sort by Newst, change to True if you want to sort by Oldest.
sorted_df.head()
```

```
[ ] my_df = sorted_df[['userName', 'score', 'at', 'content']] #get userName, rating, date-time, and reviews only

[ ] my_df.head()
```

```
[ ] new_df.to_csv("scrapped_data.csv", index = False) #kemudian save menjadi file csv
```

L5 Install Library

```
import re # Modul untuk regular expression
import pandas as pd # Manipulasi data tabular dengan Pandas
import numpy as np # Operasi numerik dengan NumPy
import nltk # Pemrosesan teks dengan NLTK

!pip install Sastrawi # Paket Sastrawi untuk pemrosesan bahasa Indonesia

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory # Stemming bahasa Indonesia dari Sastrawi
from nltk.corpus import stopwords # Stopwords dari NLTK
from nltk.tokenize import word_tokenize # Tokenisasi teks dari NLTK

nltk.download('punkt') # Model tokenisasi teks dari NLTK
nltk.download('stopwords') # Daftar stopwords dari NLTK
```

L6 Membaca Dataset

```
[ ] data = pd.read_csv("Dana CSV.csv")
    data.head(700)
```

L7 Cek Informasi Dataset

```
▶ # Membaca dan memuat data dari file CSV "Dana CSV.csv"
data = pd.read_csv("Dana CSV.csv")
data.info() # Menampilkan informasi tentang struktur data
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 699
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   User Nama    700 non-null   object
1   Waktu        700 non-null   object
2   Komentar     700 non-null   object
3   SENTIMEN     700 non-null   object
dtypes: object(4)
memory usage: 22.0+ KB
```

L8 Cek Jumlah Sentimen Positif, Negatif, Dan Netral

```
[ ] sentiment_counts = data['SENTIMEN'].value_counts() # Menghitung Jumlah Kemunculan sentimen positif dan negatif

POSITIF = sentiment_counts.get('POSITIF', 0)
NEGATIF = sentiment_counts.get('NEGATIF', 0)
NETRAL = sentiment_counts.get('NETRAL', 0)

print(f"Total Sentimen POSITIF: {POSITIF}") # mencetak jumlah total sentimen positif.
print(f"Total Sentimen NEGATIF: {NEGATIF}") # mencetak jumlah total sentimen negatif.
print(f"Total Sentimen NETRAL: {NETRAL}") # mencetak jumlah total sentimen netral.
```

L9 Membaca Dataset Kolom Review dan Sentimen

```
[ ] columns = ['Komentar', 'SENTIMEN']
df = pd.DataFrame(data, columns=columns).iloc[0:700] # membatasi DataFrame hanya pada 700 baris pertama.
df.head(5) # Menampilkan 5 baris pertama dari DataFrame df untuk melihat preview data.
```

L10 Pre-Processing Data

```
[ ] def case_folding(text):
    # Mengubah teks menjadi huruf kecil
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text

df['case_folding'] = df['Komentar'].apply(case_folding)
df.head(5)
```



```
[ ] def clean_text(tweet):
    tweet = re.sub(r'https?:\/\/\/\S+', '', tweet)
    tweet = re.sub(r'^a-zA-Z', ' ', tweet)

    emoji_pattern = re.compile("[
        u"\U0001F600-\U0001F64F" # emoticons
        u"\U0001F300-\U0001F5FF" # symbols & pictographs
        u"\U0001F680-\U0001F6FF" # transport & map symbols
        u"\U0001F1E0-\U0001F1FF" # flags (iOS)
    ]+", flags=re.UNICODE)
    tweet = emoji_pattern.sub(r'', tweet)

    tweet = re.sub(r'\d+', '', tweet)
    tweet = re.sub(r'\b\w\b', '', tweet)
    tweet = re.sub(r'\s+', ' ', tweet).strip()

    return tweet

df['cleansing'] = df['case_folding'].apply(clean_text)
df.head(5)
```

```
[ ] def tokenize(text):
    return word_tokenize(text)

df['tokenizing'] = df['cleansing'].apply(tokenize)
df.head(5)
```

```
[ ] stop_words = stopwords.words('indonesian')

additional_stopwords = [
    'yg', 'b', 'd', 'g', 'nya', 'ya', 'sih', 'aku', 'kamu', 'dia', 'mereka',
    'kita', 'saya', 'kalian', 'ini', 'itu', 'loh', 'dong', 'kan', 'uh', 'aa',
    'uh', 'kak', 'ah', 'oh', 'hmm', 'eh', 'aduh', 'ad', 'deh', 'hehe', 'ehh',
    'gk', 'h', 'hh', 'huh', 'udah', 'mau', 'gw', 'lu', 'lo', 'loh', 'pake', 'ntah',
    'ma', 'km', 'kalo', 'kl', 'tuh', 'set', 'dll', 'mcm', 'mn', 'gan', 'bro', 'sis', 'de',
    'btw', 'kayak', 'spt', 'bilang', 'kata', 'tau', 'lihat', 'bikin', 'dan', 'atau',
    'setelah', 'juga', 'ts', 'btul', 'tg', 'tr', 'utk', 'kmrn', 'tdk', 'sy', 'sm', 'dm',
    'hrs', 'dg', 'dl', 'dt', 'sdh', 'tdk', 'tp', 'bsa', 'ga', 'gak', 'masuk',
    'aja', 'apk', 'buka'
]

stop_words.extend(additional_stopwords)

def remove_stopwords(text):
    return [word for word in text if word not in stop_words]

df['stopword_removal'] = df['tokenizing'].apply(lambda x: remove_stopwords(x))
df.head(5)
```

```
▶ factory = StemmerFactory()
  stemmer = factory.create_stemmer()

def stem_text(tokens):
    return [stemmer.stem(word) for word in tokens]

df['stemming'] = df['stopword_removal'].apply(stem_text)
df['stemming'] = df['stemming'].apply(lambda x: ' '.join(x))
df.head(5)
```

L11 Pembobotan Kata (TF-IDF)

```
[ ] import pandas as pd
    from sklearn.feature_extraction.text import TfidfVectorizer

    # Memuat data yang telah diproses sebelumnya
    documents = df['stemming'].tolist()
    N = len(documents)

    # Inisialisasi TF-IDF Vectorizer
    vectorizer = TfidfVectorizer()

    # Melakukan fit dan transformasi pada dokumen
    tfidf_matrix = vectorizer.fit_transform(documents)

    # Mendapatkan nama fitur (terms)
    terms = vectorizer.get_feature_names_out()

    # Mendapatkan nilai IDF
    idf_values = vectorizer.idf_

    # Mengonversi matriks TF-IDF menjadi DataFrame
    tfidf_df = pd.DataFrame(tfidf_matrix.toarray(), columns=terms)

    # Menggabungkan DataFrame TF-IDF dengan label sentimen
    final_df = pd.concat([tfidf_df, df['SENTIMEN']], axis=1)

    # Mencetak 700 baris pertama dari DataFrame akhir
    print(final_df.head(700))
```

```
▶ # Ambil daftar kata kunci (terms)
    terms = vectorizer.get_feature_names_out()

    # Menghitung frekuensi setiap istilah dalam seluruh dataset
    term_frequencies = tfidf_matrix.sum(axis=0).A1
    terms_frequencies_df = pd.DataFrame({'term': terms, 'frequency': term_frequencies})

    # Mengurutkan istilah berdasarkan frekuensi dalam urutan menurun
    sorted_terms = terms_frequencies_df.sort_values(by='frequency', ascending=False)

    # Menampilkan istilah yang paling sering muncul
    print("Most frequent terms:")
    print(sorted_terms.head(10))
```

L12 Hasil Akurasi Nilai K

```
▶ from sklearn.neighbors import KNeighborsClassifier
    from sklearn.metrics import accuracy_score

    # Mencari akurasi untuk berbagai nilai k
    k_values = range(1, 21)
    accuracies = []

    for k in k_values:
        knn_classifier = KNeighborsClassifier(n_neighbors=k)
        knn_classifier.fit(X_train, y_train)
        y_pred = knn_classifier.predict(X_test)
        accuracy = accuracy_score(y_test, y_pred)
        accuracies.append((k, accuracy))

    # Print hasil akurasi untuk setiap nilai k
    print("Akurasi untuk berbagai nilai k:")
    for k, accuracy in accuracies:
        print(f"Nilai K: {k}, Akurasi: {accuracy:.4f}")
```

L13 Grafik Akurasi Nilai K

```
[ ] # Mencari akurasi untuk berbagai nilai k
k_values = range(1, 21)
accuracies = []

for k in k_values:
    knn_classifier = KNeighborsClassifier(n_neighbors=k)
    knn_classifier.fit(X_train, y_train)
    y_pred = knn_classifier.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    accuracies.append(accuracy)

# Plotting akurasi terhadap nilai k
plt.figure(figsize=(10, 6))
plt.plot(k_values, accuracies, marker='o')
plt.title('Akurasi terhadap Nilai K')
plt.xlabel('Nilai K')
plt.ylabel('Akurasi')
plt.xticks(k_values)
plt.grid(True)
plt.show()
```

L14 Split Data dan Evaluasi Confusion Matrix Metode KNN

```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

# Membagi data menjadi set pelatihan dan pengujian (70% pelatihan, 30% pengujian)
X_train, X_test, y_train, y_test = train_test_split(final_df.drop('SENTIMEN', axis=1),
                                                    final_df['SENTIMEN'],
                                                    test_size=0.3,
                                                    random_state=42)

# Menginisialisasi classifier KNN dengan parameter yang disesuaikan
knn_classifier = KNeighborsClassifier(n_neighbors=5) # Sesuaikan jumlah neighbors sesuai kebutuhan

# Melatih classifier
knn_classifier.fit(X_train, y_train)

# Prediksi pada set pengujian
y_pred = knn_classifier.predict(X_test)

# Akurasi
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

# Menghitung dan memvisualisasikan confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(10, 7))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=['POSITIF', 'NEGATIF', 'NETRAL'], yticklabels=['POSITIF', 'NEGATIF', 'NETRAL'])
plt.xlabel('Prediksi')
plt.ylabel('Actual')
plt.title('Confusion Matrix (KNN)')
plt.show()
```

L15 Split Data dan Evaluasi Confusion Matrix Metode Naive Bayes

```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB # Import Gaussian Naive Bayes
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

# Membagi data menjadi set pelatihan dan pengujian (70% pelatihan, 30% pengujian)
X_train, X_test, y_train, y_test = train_test_split(final_df.drop('SENTIMEN', axis=1),
                                                final_df['SENTIMEN'],
                                                test_size=0.3,
                                                random_state=42)

# Menginisialisasi classifier Naive Bayes (Gaussian)
nb_classifier = GaussianNB()

# Melatih classifier
nb_classifier.fit(X_train, y_train)

# Prediksi pada set pengujian
y_pred = nb_classifier.predict(X_test)

# Akurasi
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

# Menghitung dan memvisualisasikan confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(10, 7))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=['POSITIF', 'NEGATIF', 'NETRAL'], yticklabels=['POSITIF', 'NEGATIF', 'NETRAL'])
plt.xlabel('Prediksi')
plt.ylabel('Actual')
plt.title('Confusion Matrix (Naive Bayes)')
plt.show()
```

L16 Visualisasi Wordcloud Positif, Negatif, Dan Netral

```
[ ] import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Menginisialisasi TF-IDF Vectorizer
vectorizer = TfidfVectorizer()

# Menyesuaikan dan mengubah dokumen
tfidf_matrix = vectorizer.fit_transform(documents)

# Mendapatkan nama fitur (istilah)
terms = vectorizer.get_feature_names_out()

# Membuat DataFrame dari matriks TF-IDF
tfidf_df = pd.DataFrame(tfidf_matrix.toarray(), columns=terms)

# Menggabungkan DataFrame TF-IDF dengan label sentimen
final_df = pd.concat([tfidf_df, df['SENTIMEN']], axis=1)

# Membuat DataFrame terpisah untuk setiap kategori sentimen
sentiments = df['SENTIMEN'].unique() # Mendapatkan kategori sentimen yang unik

# Membuat fungsi untuk menghasilkan dan menampilkan WordCloud
def generate_wordcloud(text, title):
    wordcloud = WordCloud(width=800, height=400, background_color='white').generate_from_frequencies(text)
    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.title(title)
    plt.show()

# Menghasilkan WordCloud untuk setiap kategori sentimen
for sentiment in sentiments:
    # Memilih baris berdasarkan sentimen
    subset_df = final_df[final_df['SENTIMEN'] == sentiment]

    # Menghitung frekuensi kata (jumlah nilai TF-IDF per kolom)
    word_freq = subset_df.drop(columns=['SENTIMEN']).sum(axis=0).to_dict()




    # Menghasilkan dan menampilkan WordCloud
    generate_wordcloud(word_freq, f"WordCloud - SENTIMEN {sentiment}")
```

L17 Kartu Kendali Bimbingan

KARTU KENDALI BIMBINGAN

Nama Mahasiswa : Muhammad Rayhan Elfansyah
 NIM : 2011102441075
 Nama Dosen Pembimbing : Rudiman, S.Kom, M.Sc
 Judul Penelitian : Perbandingan Metode K-Nearest Neighbor (KNN) dan Naïve Bayes Terhadap Analisis Sentiment Pada Pengguna Aplikasi DANA Menggunakan Fitur Ekstraksi TF-IDF

NO	TANGGAL	URAIAN PEMBIMBINGAN	PARAF DOSEN
1	7 Feb 2024	Bimbingan pertama mendiskusikan persetujuan bimbingan dengan dosen Bapak Rudiman S.Kom, M.Sc. dengan topik penelitian Data Sentiment Analyst	
2	14 Feb 2024	Mencari permasalahan topik yang akan diteliti untuk dijadikan judul	
3	22 Feb 2024	Melakukan tahap pencarian data menggunakan pyhton di Google Colab	
4	29 Feb 2024	Menentukan judul dan mulai perencanaan latar belakang dengan syarat ketentuan yang telah ditetapkan oleh dosen pembimbing, yaitu, menggunakan sumber jurnal paling sedikit 13 dan maksimal 15 jurnal untuk BAB 1 dan bersinta minimal sinta 4	
5	9 Maret 2024	Dalam pengerjaan latar belakang harus mengikuti aturan yang telah ditetapkan oleh dospem	
6	13 Maret 2024	Melakukan beberapa revisi terhadap mahasiswa yang dibimbing	
7	18 Maret 2024	Memberikan beberapa saran terhadap canvas pengajuan judul ke prodi	
8	27 Maret 2024	Memberikan beberapa revisi terhadap latar belakang rumusan masalah batasan masalah tujuan penelitian dan manfaat penelitian	

9	5 April 2024	Masuk BAB II menggunakan strategi yang telah diberikan oleh dosen pembimbing menggunakan pembobotan TF-IDF dalam penelitian yang dilakukan	
10	9 April 2024	Memberi masukan dan revisi dan bab II	
11	24 April 2024	Membahas target penelitian dan mencari sumber jurnal serta memberikan revisian pada BAB I dan BAB II	

Dosen Pembimbing


Rudiman, S.Kom, M.Sc
 NIDN. 1105068202

Mengetahui,
 Ketua Program Studi



Arbansyah, S.Kom, M.TI
 NIDN. 118019203

SKRIPSI MUHAMMAD RAYHAN ELFANSYAH

by Teknik Informatika Universitas Muhammadiyah Kalimantan Timur



Submission date: 25-Jul-2024 09:25AM (UTC+0800)

Submission ID: 2422037828

File name: SKRIPSI_MUHAMMAD_RAYHAN_ELFANSYAH.docx (900.82K)

Word count: 8125

Character count: 50749

SKRIPSI MUHAMMAD RAYHAN ELFANSYAH



ORIGINALITY REPORT

22%

SIMILARITY INDEX

19%

INTERNET SOURCES

13%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Brawijaya Student Paper	2%
2	ejournal.unesa.ac.id Internet Source	1%
3	repository.ub.ac.id Internet Source	1%
4	openjournal.unpam.ac.id Internet Source	1%
5	repository.uin-suska.ac.id Internet Source	1%
6	docplayer.info Internet Source	1%
7	123dok.com Internet Source	1%
8	Submitted to STMIK STIKOM Bali Student Paper	1%
9	doaj.org Internet Source	1%

RIWAYAT HIDUP



Muhammad Rayhan Elfansyah, Lahir di Samarinda pada Tanggal 07 Juli 2002. Merupakan anak pertama dari tiga bersaudara yang lahir dari pasangan Bapak Fadliansyah dan Ibu Fatmawati. memiliki saudara bernama Muhammad Naufal Rozansyah dan Muhammad Dzikri Romansyah. Pendidikan yang telah di tempuh oleh peneliti yaitu SD Negeri 028 Samarinda lulus pada tahun 2014. Kemudian dilanjutkan dengan menempuh Pendidikan di MTS DDI Tani Aman lulus pada tahun 2017. Setelah itu dilanjutkan dengan menempuh Pendidikan di SMA Negeri 7 Samarinda, lulus pada tahun 2020. Pada tahun 2020 peneliti menempuh Pendidikan Program Studi S1 Teknik Informatika di Universitas Muhammadiyah Kalimantan Timur sebagai syarat kelulusan dan memperoleh gelar Sarjana Komputer (S.Kom).