Analisis Sentimen Ulasan Aplikasi Tiktok di Google Playstore Menggunakan Algoritma Naive Bayes

Disusun Oleh :

Laila Mustika Sari	11221002
Alpian Roymundus Siringo-ringo	11211009
Arief Reno Fathurrahman	11201014

PROGRAM STUDI INFORMATIKA

JURUSAN MATEMATIKA DAN TEKNOLOGI INFORMASI

INSTITUT TEKNOLOGI KALIMANTAN

2024

1. **Penentuan Model**

   Dalam mempelajari MK Penggalian Data (Data Mining), apa yang ingin kelompok anda capai dengan data mining? Isu apa yang ingin coba kelompok anda selesaikan dengan data mining, apakah itu berkaitan dengan klasifikasi, prediksi, deteksi anomali, atau lainnya? Jika sudah menentukan, selanjutnya;

   a. Identifikasi metode
   b. Justifikasi pilihan: tuliskan alasan mengapa kelompok anda memilih metode tersebut. Sertakan kelebihan dan kekurangan metode ini dibandingkan metode lain!

2. **Review Jurnal**

   a. **Pencarian Jurnal :** Cari dan pilih 3-5 jurnal yang relevan dengan metode/teknik penggalian data yang dipilih. Pastikan jurnal tersebut berkualitas dan terkini dalam 5 tahun belakang baik jurnal/prosiding internasional dan/atau jurnal nasional dengan level Sinta antara S1, S2, S3 dan S4.

   b. **Review Jurnal Mencakup :**
      - Masalah yang diangkat serta tujuan dari penelitian
      - Dataset yang digunakan, atribut yang digunakan
      - Metode/teknik penggalian data yang digunakan
      - Algoritma/flowchart/alur penyelesaian seperti apa
      - Hasil yang diperoleh pada penelitian
      - Identifikasi kelemahan dan keterbatasan (seperti metode yang digunakan, ukuran sampel yang diteliti, ruang lingkup penelitian, variable/data yang tidak dipertimbangkan, dsb)
      - Sertakan link masing-masing

3. **Penentuan GAP dan Ide Topik**

   a. **Menentukan GAP Penelitian**

      Berdasarkan analisis dari kelemahan dan keterbatasan masing-masing jurnal yang telah direview. tuliskan satu atau lebih gap penelitian yang dapat diidentifikasi. Gap ini bisa berupa area/data yang kurang diteliti, metode atau pendekatan yang belum diuji, konteks yang belum diperhatikan, dsb.

   b. **Ide/Topik Penelitian**

      Berdasarkan gap yang telah diidentifikasi, usulkan ide penelitian yang dapat menjadi fokus topik dari Proyek Tugas Besar kelompok anda. Ide ini boleh menggunakan isu/permasalahan/konteks lain diluar topik jurnal yang direview (data) tetapi menggunakan metode/pendekatan yang berkaitan dengan jurnal yang direview atau menggunakan dataset yang sama namun metode/pendekatan yang berbeda, ataupun tidak keduanya, ide baru dari kelompok anda dipersilahkan namun tetap menggunakan jurnal yang telah direview sebagai referensi utama. Kemudian buatlah mind map/ kerangka pemikiran yang mencakup:

- Isu yang ingin diangkat
- Judul/topik penelitian yang diusulkan
- Tujuan dan rumusan dari penelitian
- Metode yang diusulkan
- Luaran/output yang diharapkan

4. Pencarian Dataset

a. Pencarian Dataset

Setelah menentukan teknik/metode lalu kalian dapati gap serta ide/topik penelitian proyek kalian, carilah dataset yang relevan dan memenuhi syarat untuk metode yang digunakan dan permasalahan yang diangkat. Pastikan dataset tersebut berkualitas dan memiliki fitur yang diperlukan. Kalian bisa mencari di repositori data seperti Kaggle, UCI Machine Learning Repository, Google Scholar, Badan Pusat Statistik, Satu Data Indonesia atau sumber lain. Lalu buatlah deskripsi dataset, meliputi :

- Sumber Data (cantumkan link data atau sumber data)
- Format, jumlah data dan fitur
- Atribut pada dataset: tampilkan 10 kolom data beserta atributnya
- Analisis atribut (deskripsi, type data dan jumlah data)

b. Pra-Processing Dataset

Setelah mendapatkan dataset, lakukan langkah-langkah pra-processing yang sesuai untuk mempersiapkan data sebelum penerapan teknik penggalian data

- **Pembersihan Data:** lakukan pembersihan data, termasuk menghapus nilai yang hilang, menghapus duplikasi, dan memperbaiki kesalahan data.
- **Transformasi Data:** lakukan transformasi yang diperlukan, seperti normalisasi, standarisasi, encoding variabel kategorikal (konversi variabel kategorikal menjadi numerik jika diperlukan), dan pengurangan dimensi jika diperlukan.
- **Pembagian Data:** seperti memisahkan dataset menjadi set pelatihan, validasi, dan pengujian untuk evaluasi model yang lebih baik.
- **Penanganan Data Imbalance:** mengatasi ketidakseimbangan antara kelas dalam dataset, misalnya dengan menambah jumlah contoh dari kelas minoritas atau mengurangi kelas mayoritas.
- Dan proses pra-processing dataset lainnya sesuai kebutuhan dari penelitian kelompok anda.
- Buatlah visualisasi dari dataset dalam beberapa bentuk diagram seperti histogram, blox pot, dsb.

Buatlah dokumentasi proses dengan menambahkan catatan tentang setiap langkah pra-processing yang dilakukan. Jelaskan alasan di balik setiap langkah.

1. PENENTUAN MODEL

A. Identifikasi Metode

Kami akan menggunakan Data Mining untuk menyelesaikan isu analisis sentimen, yang berkaitan dengan bagaimana pengguna menilai dan memberikan opini tentang suatu aplikasi yang ada di Google Playstore atau bagaimana pengguna memberikan opini atau perasaan mereka di media sosial. Sentimen ini bisa berupa positif, negatif, atau netral, dan penting untuk memahami opini publik terkait aplikasi tersebut.

Kelompok kami ingin memahami bagaimana persepsi atau opini pengguna terhadap suatu aplikasi melalui ulasan atau komentar mereka. Secara spesifik, dengan melakukan analisis sentimen, kita bisa menyortir ulsan berdasarkan sentimen (misalnya, ulasan dengan sentimen negatif dapat digunakan untuk meningkatkan fitur aplikasi), mengidentifikasi pola umum dalam umpan balik pengguna (misalnya, pengguna yang merasa aplikasi lambat atau terdapat banyak iklan), kemudian bisa membuat prediksi tren berdasarkan ulasan pengguna, seperti penurunan popularitas aplikasi jika banyak ulasan negatif. Pendekatan yang digunakan akan berfokus pada klasifikasi sentimen (positif, negatif atau netral).

B. Justifikasi Pilihan

Kelompok kami akan menggunakan Naive Bayes Classifier untuk mengklasifikasikan sentimen dari ulasan dan komentar.

Justifikasi Pilihan:

1. Alasan memilih Naive Bayes:
    a. Efisiensi: Naive Bayes adalah metode yang cepat dan efisien untuk mengolah data teks yang besar. Ini sangat cocok untuk dataset ulasan yang bisa mencapai ribuan bahkan lebih.
    b. Simplicity: Naive Bayes sangat sederhana namun efektif untuk tugas klasifikasi teks seperti analisis sentimen, terutama saat bekerja dengan fitur yang diekstrak dari teks seperti kata-kata atau frasa.
    c. Performa yang baik dengan data teks: Metode ini bekerja dengan baik pada dataset teks dengan model distribusi sederhana yang berasumsi bahwa setiap fitur (kata) dalam teks independen satu sama lain (asumsi independensi).

2. Kelebihan Naive Bayes:
    a. Cepat dalam training dan prediksi: Algoritma ini relatif cepat dalam proses pelatihan dan sangat efisien untuk dataset besar.
    b. Baik digunakan untuk data dengan noise: Naive Bayes memiliki kemampuan yang baik dalam menangani data teks yang sering kali mengandung noise seperti kesalahan ejaan atau kata-kata tidak penting.
    c. Bekerja baik dengan data yang besar: Naive Bayes cenderung punya performa baik ketika dataset memiliki banyak contoh tetapi jumlah fitur (kata atau frasa) juga besar.
3. Kekurangan Naive Bayes:
    a. Asumsi independensi: Salah satu kelemahan utamanya adalah asumsi bahwa semua kata di dalam teks independen satu sama lain, padahal dalam kenyataannya, kata-kata seringkali memiliki hubungan kontekstual.
    b. Kurang akurat dibandingkan model yang lebih kompleks: Model yang lebih kompleks seperti Support Vector Machine (SVM) atau Neural Network dapat memberikan akurasi yang lebih baik, tetapi mereka juga lebih membutuhkan waktu dan sumber daya komputasi untuk pelatihan.
    c. Sensitif terhadap data tidak seimbang: Jika data ulasan lebih banyak mengandung ulasan positif atau negatif, Naive Bayes cenderung bias terhadap kelas dengan jumlah data yang lebih besar.

Kelebihan dan Kekurangan Dibandingkan dengan Metode Lain:

A. Support Vector Machine:
    a. Kelebihan SVM: Lebih akurat dalam menangani data yang tidak linier, terutama dalam kasus data teks dengan banyak dimensi.
    b. Kekurangan SVM: Lebih kompleks dan membutuhkan lebih banyak waktu komputasi dibandingkan Naive Bayes, sehingga tidak selalu efisien untuk dataset besar.
B. Logistic Regression:
    a. Kelebihan: Dapat digunakan sebagai model probabilistik untuk klasifikasi dan lebih mudah diinterpretasikan.

b. Kekurangan: Logistic Regression tidak selalu mempunyai performa lebih baik dibandingkan Naive Bayes, terutama ketika fitur teks tidak banyak.

C. Deep Learning (LSTM, CNN):

a. Kelebihan: Dapat menghasilkan akurasi yang sangat baik dengan menangkap hubungan antar kata dalam teks (misalnya urutan kata).

b. Kekurangan: Memerlukan sumber daya komputasi yang besar dan waktu pelatihan yang lebih lama, tidak cocok untuk proyek sederhana atau dataset yang lebih kecil.

Kelompok kami memilih Naive Bayes untuk analisis sentimen karena kemampuannya yang cepat dan efisien dalam menangani data teks yang besar.

2. REVIEW JURNAL

Jurnal 1 (Sinta 2, Tahun 2021)

Judul Jurnal	Analisis Sentimen Dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayesdan Latent Dirichlet Allocation
Nama Penulis	1. Ni Luh Putu Merawati 2. Ahmad Zuli Amrullah 3. Ismarmiaty
Masalah dan Tujuan Penelitian	Masalah: Pulau Lombok menjadi tujuan wisata populer di Indonesia, menarik banyak wisatawan yang membagikan opini mereka di media sosial, khususnya di Twitter. Karena jumlah opini sangat besar dan bervariasi, identifikasi sentimen publik dan topik pembicaraan secara manual sulit dilakukan. Tujuan Penelitian: 1. Mengklasifikasikan opini wisatawan tentang pariwisata di Lombok menjadi dua kelas, yaitu sentimen positif dan negatif, menggunakan metode Naive Bayes. 2. Melakukan pemodelan topik pada masing-masing kelas sentimen (positif dan negatif) menggunakan metode Latent Dirichlet Allocation (LDA) untuk mengetahui topik utama yang sering dibicarakan pada tiap kelas sentimen.

Dataset dan Atribut	Dataset: ● Dataset berisi **12.971 tweet** yang dikumpulkan menggunakan API Twitter dengan tagar terkait pariwisata Lombok (misalnya, #gilitrawangan, #wonderfulllombok, #pantaikutamandalika). Data ini mencakup rentang waktu dari tahun **2014 hingga 2019**. ● Setelah proses pembersihan dan pemilahan, **9.496 tweet** digunakan untuk analisis, yang terdiri dari **8.996 tweet positif** dan **500 tweet negatif**. Karena terjadi ketidakseimbangan data, diterapkan **teknik undersampling** untuk menyeimbangkan kelas, sehingga digunakan **500 tweet positif** dan **500 tweet negatif**. Atribut: ● **Tweet**: Isi teks tweet yang berisi opini wisatawan terkait pariwisata Lombok. ● **Sentimen**: Label yang mengklasifikasikan sentimen tweet, yaitu **positif** atau **negatif**. ● **Fitur teks**: Fitur-fitur teks yang dihasilkan dari proses **text transformation** (case folding, tokenisasi, stopword removal, lemmatisasi, dan pembobotan TF-IDF).
Metode	Menggunakan dua metode utama untuk analisis sentimen dan pemodelan topik pada ulasan pariwisata Lombok di Twitter, yaitu Naive Bayes untuk klasifikasi sentimen dan Latent Dirichlet Allocation (LDA) untuk pemodelan topik.
Algoritma/flowchart/alur penyelesaian	1. Pengumpulan Data dan Pelabelan: Mengambil data dari Twitter dan melakukan pelabelan sentimen untuk digunakan dalam analisis. 2. Preprocessing: Membersihkan teks dan melakukan transformasi agar siap untuk proses klasifikasi dan pemodelan topik. 3. Klasifikasi dengan Naive Bayes: Menentukan sentimen (positif/negatif) untuk tiap tweet. 4. Pemodelan Topik dengan LDA: Mengidentifikasi topik yang sering dibicarakan dalam kelas positif dan negatif. 5. Evaluasi dan Visualisasi: Menginterpretasikan hasil klasifikasi dan pemodelan topik serta memvisualisasikan hasil untuk analisis lebih lanjut.
Hasil Penelitian	Hasil pemodelan topik dengan metode LDA pada masing-masing kelas positif dan negatif dapat dilihat dari nilai koherensi yaitu semakin tinggi nilai koherensi suatu topik maka semakin mudah topik tersebut diinterpretasikan oleh manusia. Nilai koherensi tertinggi untuk kelas positif diperoleh pada topik

	ke 8 dengan nilai sebesar 0,613dan untuk kelas negatif pada topik ke 12 dengan nilai sebesar 0,528.
Kelemahan dan Keterbatasan	1. Ukuran Dataset yang Terbatas: Meskipun jumlah 9.496 tweet terlihat cukup, namun jumlah tersebut mungkin masih kurang representatif untuk mencakup seluruh perspektif wisatawan terhadap pariwisata Lombok. 2. Kualitas Data: Data yang diperoleh dari media sosial seperti Twitter dapat mengandung noise (kebisingan) seperti spam, informasi yang tidak relevan, atau tweet yang tidak terkait langsung dengan pariwisata. Ini dapat mempengaruhi akurasi analisis sentimen.
Link Jurnal	https://www.jurnal.iaii.or.id/index.php/RESTI/article/view/2587/375

Jurnal 2 (Sinta 3, Tahun 2021)

Judul Jurnal	Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter
Nama Penulis	Dedy Atmajaya, Annisa Febrianti, Herdianti Darwis
Masalah dan Tujuan Penelitian	Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma machine learning, Support Vector Machine (SVM) dan Naive Bayes, dalam menganalisis sentimen pengguna Twitter terkait ChatGPT, model bahasa canggih dari OpenAI. Analisis dilakukan untuk memahami sentimen publik terhadap ChatGPT.
Dataset dan Atribut	Dataset berjumlah 1000 tweet terkait ChatGPT yang diambil dari Kaggle. Dataset ini mencakup atribut "date," "tweet," dan "username."
Metode	Menggunakan pendekatan preprocessing data seperti tokenisasi dan pembersihan teks, serta pelabelan data dengan model Vader dan RoBERTa. Algoritma yang digunakan adalah SVM dan Naive Bayes, dengan evaluasi melalui metrik akurasi, presisi, recall, dan F1-score.
Algoritma/flowchart/alur penyelesaian	Alur penyelesaian penelitian meliputi pengumpulan data, preprocessing data, pelabelan sentimen menggunakan Vader dan RoBERTa, ekstraksi fitur menggunakan TF-IDF, dan klasifikasi dengan SVM dan Naive Bayes.

Hasil Penelitian	SVM menunjukkan kinerja lebih baik dibandingkan Naive Bayes dalam analisis sentimen. Dengan Vader, SVM mencapai akurasi, presisi, dan recall sebesar 59%, sedangkan RoBERTa dengan SVM mencapai 55%. Naive Bayes menunjukkan performa yang lebih rendah dengan akurasi 47% (Vader) dan 43% (RoBERTa).
Kelemahan dan Keterbatasan	Naive Bayes menunjukkan kinerja yang lebih rendah dibandingkan SVM dalam analisis sentimen, terutama pada penggunaan RoBERTa. Keterbatasan penelitian ini juga mencakup ketergantungan pada label sentimen yang memengaruhi hasil klasifikasi.
Link Jurnal	http://ijcs.net/ijcs/index.php/ijcs/article/view/3341

Jurnal 3 (Sinta 4, Tahun 2020)

Judul Jurnal	Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ)
Nama Penulis	- Naomi Chatrina Siregar - Riki Ruli A. Siregar - M. Yoga Distra Sudirman
Masalah dan Tujuan Penelitian	Masalah: Pembelajaran jarak jauh (PJJ) di Indonesia menghasilkan banyak komentar yang sulit dikelola secara manual, sehingga perlu sistem untuk mengklasifikasikan komentar tersebut. Tujuan: Mengembangkan sistem yang dapat mengelola dan mengklasifikasikan komentar PJJ menjadi kategori positif, negatif, atau netral menggunakan algoritma Naive Bayes Classifier.
Dataset dan Atribut	Dataset: Dataset terdiri dari komentar yang diberikan oleh warga sekolah mengenai pelaksanaan Pembelajaran Jarak Jauh (PJJ). Komentar ini mencakup berbagai pendapat dan reaksi terhadap sistem pembelajaran yang diterapkan. Atribut: - Komentar: Teks yang berisi pendapat atau feedback dari pengguna terkait PJJ. Komentar ini dapat berupa kalimat atau frasa yang menggambarkan pengalaman belajar. - Kategori: Hasil klasifikasi komentar, yang dikelompokkan

	ke dalam tiga kategori: - Positif: Komentar yang menunjukkan kepuasan atau pandangan baik terhadap PJJ. - Negatif: Komentar yang menunjukkan ketidakpuasan atau kritik terhadap PJJ. - Netral: Komentar yang tidak menunjukkan pendapat yang kuat, baik positif maupun negatif.
Metode	Penelitian menggunakan metode text mining dengan langkah-langkah: - Text Preprocessing (case folding, tokenizing, filtering, stemming) - Pembobotan TF-IDF - Klasifikasi menggunakan Naïve Bayes Classifier
Algoritma/flowchart/alur penyelesaian	- Input Komentar: Dimana pengguna memasukkan komentar ke dalam sistem - Text Processing Steps: Komentar yang dimasukkan ke dalam sistem akan melalui serangkaian tahapan: - Case Folding: dimana semua huruf diubah menjadi kecil - Tokenizing: dimana teks dipecah menjadi kata-kata - Filtering: untuk menghilangkan kata-kata tidak penting - Stemming: untuk mengubah kata menjadi bentuk dasar - Pembobotan TF-IDF: Setelah preprocessing, setiap kata dalam komentar dihitung bobotnya menggunakan metode TF-IDF. - Klasifikasi dengan Naive Bayes: Menggunakan bobot kata yang telah dihitung, algoritma Naïve Bayes menghitung probabilitas untuk setiap kategori (positif, negatif, netral). - Output: Sistem akan menampilkan hasil klasifikasi komentar, menunjukkan apakah komentar tersebut termasuk dalam kategori positif, negatif, atau netral.
Hasil Penelitian	Sistem yang dibangun dapat menganalisis komentar dengan lebih tepat dibandingkan pengolahan manual. Diperoleh tingkat akurasi sebesar 68% dari pengujian yang dilakukan.
Kelemahan dan Keterbatasan	- Klasifikasi komentar dipengaruhi oleh kualitas data; komentar yang tidak sesuai dengan kaidah bahasa Indonesia dapat mengurangi akurasi. - Penggunaan data dalam jumlah kecil dapat mempengaruhi hasil klasifikasi. - Ketergantungan pada kamus bahasa Indonesia yang mungkin tidak mencakup semua variasi kalimat.
Link Jurnal	https://aperti.e-journal.id/teknologia/article/download/67/45/

Jurnal 4 (Sinta 4, Tahun 2022)

Judul Jurnal	Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet
Nama Penulis	- Heliyanti Susana - Nana Suarna - Fathurrohman - Kaslani
Masalah dan Tujuan Penelitian	Masalah: Penggunaan akses internet di kalangan siswa dapat berdampak pada perilaku dan kenakalan siswa. Penelitian ini bertujuan untuk menganalisis hak akses internet siswa dengan menggunakan metode klasifikasi. Tujuan: Menerapkan algoritma Naive Bayes untuk mengklasifikasikan penggunaan akses internet siswa dan mengukur akurasinya.
Dataset dan Atribut	Dataset: Dataset yang digunakan terdiri dari data kuisioner yang diisi oleh siswa SMA N 1 Plumbon dengan jumlah total sampel yang digunakan dalam penelitian adalah 270 siswa, yang diambil dari kelas X, XI, dan XII. Atribut: - Umur: Usia siswa. - Jenis Kelamin: Laki-laki atau perempuan. - Kelas: Kelas yang diikuti siswa (X, XI, XII). - Tempat Tinggal: Apakah siswa tinggal di daerah perkotaan atau pedesaan. - Gunakan HP: Apakah siswa menggunakan handphone dalam 3 bulan terakhir. - Gunakan Laptop: Apakah siswa menggunakan laptop dalam 3 bulan terakhir. - Akses Internet: Apakah siswa mengakses internet.
Metode	Melalui survei kuisioner dan analisis deskriptif untuk data kuantitatif, beserta tahapan KDD sebagai berikut: - Data Collection - Data Cleaning - Data Transformation - Data Mining - Evaluation

	- Knowledge Extraction
Algoritma/flowchart/alur penyelesaian	- Pengumpulan Data: Melakukan survei dengan mengedarkan kuisioner kepada siswa. - Data Cleaning: Menghapus data yang tidak relevan, nilai null, atau duplikat. - Data Transformation: Mengubah nilai nominal menjadi nilai numerik untuk mempermudah analisis. - Data Mining: Menerapkan algoritma Naive Bayes untuk mengklasifikasikan data berdasarkan atribut yang ada. - Evaluasi Model: Menghitung akurasi model dengan hasil prediksi dibandingkan dengan data aktual. Dalam penelitian ini, akurasi yang didapat adalah 89.83%. - Interpretasi Hasil: Menganalisis hasil prediksi dan mendeskripsikan karakteristik siswa berdasarkan hasil klasifikasi.
Hasil Penelitian	Akurasi penelitian mencapai 89.83%. Rincian hasil klasifikasi: - Hasil Prediksi Ya dan ternyata Ya: 34 - Hasil Prediksi Ya dan ternyata Tidak: 6 - Hasil Prediksi Tidak dan ternyata Ya: 0 - Hasil Prediksi Tidak dan ternyata Tidak: 19
Kelemahan dan Keterbatasan	- Keterbatasan dalam jumlah data yang digunakan, sehingga hasil mungkin tidak sepenuhnya representatif. - Potensi bias akibat jawaban kuisioner yang mungkin tidak akurat. - Penelitian hanya fokus pada satu sekolah, yang membatasi generalisasi hasil ke populasi yang lebih luas.
Link Jurnal	https://jursistekni.nusaputra.ac.id/article/download/96/59/

3. PENENTUAN GAP dan IDE TOPIK

A. Menentukan Gap Penelitian

Berdasarkan dari jurnal yang telah di review, terdapat beberapa kelemahan dan keterbatasan dari masing-masing jurnal. Dari analisis tersebut, dapat diidentifikasikan beberapa gap penelitian yang dirangkum menjadi sebagai berikut:

1. Keterwakilan Data

   Jurnal 1 dan Jurnal 4 mengindikasikan bahwa ukuran dataset yang digunakan mungkin tidak sepenuhnya representatif. Pengambilan data dari platform media sosial seperti Twitter (Jurnal 1) dan dari satu institusi pendidikan (Jurnal 4) berpotensi menghasilkan bias. Oleh karena itu, terdapat kebutuhan untuk melakukan penelitian yang melibatkan dataset yang lebih besar dan beragam, mencakup berbagai sumber media sosial serta lokasi geografis yang berbeda.

2. Perbandingan Metode Analisis

   Jurnal 2 menunjukkan bahwa algoritma Support Vector Machine (SVM) menunjukkan kinerja yang lebih baik dibandingkan Naive Bayes. Namun, tidak ada penelitian yang melakukan perbandingan terhadap kinerja algoritma machine learning lainnya, seperti Logistic Regression atau Decision Trees, dalam konteks analisis sentimen yang sama. Penelitian lebih lanjut sebaiknya mengeksplorasi algoritma-algoritma tersebut untuk mengevaluasi potensi peningkatan akurasi dalam situasi yang berbeda.

3. Penggunaan Metode Pembersihan Data yang Lebih Canggih

   Semua jurnal menyebutkan adanya masalah terkait kualitas data, tetapi tidak ada yang mengkaji penggunaan teknik pembersihan data yang lebih canggih, seperti deep learning atau teknik pemrosesan bahasa alami yang lebih mutakhir untuk mengurangi noise dalam dataset. Penelitian yang memanfaatkan metode-metode ini dapat memberikan wawasan yang lebih mendalam mengenai sentimen publik.

4. Analisis Sentimen Multikonteks

   Jurnal-jurnal yang diteliti berfokus pada konteks tertentu, seperti pariwisata, pendidikan, dan teknologi; namun, belum ada penelitian yang menganalisis sentimen dalam konteks yang lebih luas, seperti sentimen terhadap kebijakan pemerintah atau masalah sosial yang lebih luas. Penelitian yang menggabungkan berbagai konteks ini dapat memberikan pemahaman yang lebih luas.

5. Aspek Temporal dalam Analisis Sentimen

   Tidak terdapat penelitian yang mempertimbangkan perubahan sentimen seiring waktu, terutama dalam konteks tren sosial dan teknologi yang berkembang pesat. Penelitian yang menganalisis data dalam rentang waktu tertentu dapat membantu dalam memahami bagaimana opini publik berkembang dan dipengaruhi oleh peristiwa-peristiwa tertentu.

6. Keterbatasan dalam Penggunaan Alat Analisis Sentimen

   Jurnal 2 hanya menggunakan dua alat analisis sentimen, yaitu Vader dan RoBERTa. Penelitian lebih lanjut sebaiknya menguji kombinasi berbagai alat analisis sentimen untuk meningkatkan akurasi, serta mengeksplorasi bagaimana alat-alat tersebut dapat saling melengkapi.

B. Ide/Topik Penelitian

Berdasarkan gap yang telah diidentifikasi, kelompok kami mengusulkan sebuah ide penelitian yang berjudul "**Analisis Sentimen Ulasan Aplikasi Tiktok di Google Playstore Menggunakan Algoritma Naive Bayes**" berdasarkan dari jurnal-jurnal yang telah di review diatas menggunakan teknik yang sama yaitu algoritma Naive Bayes Classifier dimana kami gunakan pada ulasan-ulasan pengguna aplikasi TikTok untuk menyortir dan memisahkan ulasan positif dan negatif dari seluruh pengguna aplikasi TikTok.

LINK MIRO: Link Miro Kelompok 8