

**ANALISIS SENTIMEN APLIKASI INDRIVE PADA ULASAN
GOOGLE PLAYSTORE DENGAN METODE NAIVE BAYES**

SKRIPSI

DISUSUN OLEH:

RAMADHAN PRATAMA

2009020096



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2024

LEMBAR PENGESAHAN

Judul Skripsi : ANALISIS SENTIMEN APLIKASI INDRIVE PADA
ULASAN GOOGLE PLAYSTORE DENGAN
METODE NAÏVE BAYES
Nama Mahasiswa : RAMADHAN PRATAMA
NPM : 2009020096
Program Studi : TEKNOLOGI INFORMASI

Menyetujui Dosen Pembimbing



(Amrullah, S.Kom, M.Kom)

NIDN. 0125118604

Ketua Program Studi



(Fatma Sari Hutagalung, S.Kom, M.Kom)
NIDN./0117019301

Dekan



(Dr. Al-Khowarizmi, S.Kom.,
M.Kom.) NIDN. 0127099201

PERNYATAAN ORISINALITAS

**ANALISIS SENTIMEN APLIKASI INDRIVE PADA ULASAN
GOOGLE PLAYSTORE DENGAN METODE NAIVE BAYES
SKRIPSI**

SKRIPSI

Saya menyatakan bahwa karya tulisan ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, Agustus 2024

Yang membuat pernyataan



Ramadhan Pratama

NPM. 2009020096

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Ramadhan Pratama
NPM : 2009020096
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**ANALISIS SENTIMEN APLIKASI INDRIVE PADA
ULASAN GOOGLE PLAYSTORE DENGAN METODE
NAÏVE BAYES**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, Agustus 2024

Yang membuat pernyataan



Ramadhan Pratama

NPM. 2009020096

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Ramadhan Pratama
Tempat dan Tanggal Lahir : Medan, 05 November 2002
Alamat Rumah : Medan
Telepon/Faks/HP : 081262006761
E-mail : ramadhanpratama185@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SD Swasta Nurul Fathimiyah TAMAT: 2014
SMP : SMPN 2 Percut Sei Tuan TAMAT: 2017
SMK : SMKN 4 Medan TAMAT: 2020

KATA PENGANTAR



Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Fatma Sari Hutagalung, S.Kom, M.Kom, Ketua Program Studi Teknologi Informasi
4. Bapak Mhd. Basri, S.Si, M.Kom Sekretaris Program Studi Teknologi Informasi
5. Pembimbing saya Bapak Amrullah, S.Kom., M.Kom.
6. Kepada ayah saya Achmad Syafriyandi saya ucapkan terimakasih banyak atas seluruh upaya dalam membiayai pendidikan penulis hingga sarjana.
7. Kepada ibu saya Ruminah saya ucapkan beribu terimakasih atas doa dan kasih sayangnya yang tulus, nasihat, motivasi, semangat dan doa yang tiada henti sehingga penulis bisa menyelesaikan studinya hingga sarjana, *I love you so much mom.*
8. Kepada kakak/abang sepupu, dan adik kandung saya yang sangat saya sayangi terimakasih telah memberikan motivasi dan perhatiannya.
9. Kepada Septiani Putri selaku pacar saya yang selalu selalu memberi semangat,dukungan,dan selalu mendengarkan keluh kesah si penulis agar bisa menyelesaikan skripsi ini, termakasih banyak saya ucapkan.
10. Kepada teman seperjuangan saya Doli,Dava,Ekky,Bintang saya ucapkan terima kasih karna selalu menemani dalam pengerjaan skripsi ini, diucapkan terima kasih.
11. Semua Pihak yang terlibat baik langsung maupun tidak langsung dalam pengerjaan Skripsi ini yang tidak penulis sebutkan satu persatu diucapkan terima kasih.

ABSTRAK

Perkembangan teknologi aplikasi pada smartphone berbasis Android saat ini terus berkembang sehingga mengubah cara dan perilaku masyarakat dalam beraktivitas sehari-hari terutama dalam berpergian kemana pun . Indrive adalah suatu aplikasi ojek online yang sangat populer dalam belakangan ini. Para penumpang memilih aplikasi ini dikarenakan aplikasi ini sangat mudah digunakan terutama para penumpang bisa memilih para drivernya,serta bisa memilih kendaraannya. Indrive merupakan salah satu aplikasi ojek online yang harga ongkosnya relatif murah dibandingkan aplikasi ojek lainnya. Sangkin populernya aplikasi ini, Aplikasi indrive memiliki banyak ulasan di google playstore dari para penggunanya dikarenakan aplikasi ini sering terjadi update secara terus menerus yang menyebabkan titik penjemputan pada maps nya sering error atau tidak akurat. Maka dari itu dilakukan lah penelitian untuk membantu menjaga kepopuleran aplikasi indrive. penelitian ini melakukan pengumpulan data dengan teknik web scrapping menggunakan Algoritma Naïve Bayes. Metode ini dipilih karena kemampuannya yang efektif dalam menentukan ulasan negatif serta ulasan positif. Penelitian ini menggunakan data ulasan yang diambil dari google playstore untuk memisahkan kategori ulasan negatif dan positif. Hasil penelitian diharapkan dapat memberikan informasi yang akurat kepada masyarakat tentang aplikasi indrive, serta membantu manajemen indrive dalam menjaga kepopuleran aplikasi ini.

Kata kunci : Indrive, Naive Bayes, Ulasan positif, Ulasan negatif, Web Scrapping

ABSTRACT

The development of application technology on Android-based smartphones is currently continuing to grow, changing the way and behavior of people in their daily activities, especially in traveling anywhere. Indrive is an online motorcycle taxi application that is very popular lately. Passengers choose this application because this application is very easy to use, especially passengers can choose their drivers, and can choose their vehicles. Indrive is one of the online motorcycle taxi applications whose fares are relatively cheap compared to other motorcycle taxi applications. Because of the popularity of this application, the Indrive application has many reviews on the Google Playstore from its users because this application often experiences continuous updates which cause the pick-up point on its maps to often error or be inaccurate. Therefore, research was conducted to help maintain the popularity of the Indrive application. This study collected data using the web scraping technique using the Naïve Bayes Algorithm. This method was chosen because of its effective ability to determine negative reviews and positive reviews. This study uses review data taken from the Google Playstore to separate the categories of negative and positive reviews. The results of the study are expected to provide accurate information to the public about the InDrive application, as well as help InDrive management in maintaining the popularity of this application.

Keywords: Indrive, Naive Bayes, Positive reviews, Negative reviews, Web Scrapping

DAFTAR ISI

DAFTAR ISI	iv
DAFTAR TABEL.....	vi
DAFTAR GAMBAR	vii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	5
BAB II LANDASAN TEORI	6
2.1 InDrive.....	6
2.2 Analisis Sentimen	8
2.3 Metode Naive Bayes	8
2.4 Google Collab	9
2.5 Pyhton.....	10
BAB III METODOLOGI PENELITIAN	11
3.1. Metode Penelitian.....	11
3.2. Waktu dan Tempat Penelitian.....	17
3.2.1 Waktu Penelitian	17
3.2.2 Tempat Penelitian.....	17
3.3. Perangkat Penelitian	17
3.4. Flowchart Sistem Klasifikasi	18

BAB IV HASIL DAN PEMBAHASAN	19
4.1 Hasil Web scrapping	19
4.2 Hasil Pelabelan Data	21
4.3 Hasil Preprocessing Data	21
4.4 Hasil Pembobotan TF IDF	26
4.5 Hasil Naive Bayes	31
4.6 Hasil Evaluasi	31
4.7 Hasil visualisasi	37
4.8 Hasil Pengujian	38
BAB V KESIMPULAN DAN SARAN	36
5.1 Kesimpulan	36
5.2 Saran	36
DAFTAR PUSTAKA	37

DAFTAR TABEL

Tabel 3.1 Alur penelitian.....	11
Tabel 3.1.2 Data Ulasan	12
Tabel 3.2. 2 Jadwal Penelitian.....	17
Tabel 3.2. 1 Jadwal Penelitian	18
Tabel 3.3 Perangkat penelitian	18
Tabel 3.4 Flowchart Sistem Klasifikasi	18
Tabel 4.5 Hasil evaluasi.....	22

DAFTAR GAMBAR

Flowchart 3.1.4 Tahap Preprocessing data	18
Gambar 4.1 Web scrapping pada pyhton.....	20
Gambar 4.2 Hasil Web scrapping.....	20
Gambar 4.3 Hasil Pelabelan Data.....	21
Tabel 4.4. Data Case Folding	22
Gambar 4.5 Hasil Case Folding	22
Gambar 4.6 Hasil tokenizing	23
Gambar 4.7 Stopword Pada google colab.....	25
Gambar 4.8 Hasil Stopword	25
Gambar 4.9 Hasil Stemming	26
Gambar 4.10 Volabulary TF IDF	27
Gambar 4.11 Skor chi-square untuk tiap fitur TF IDF	28
Gambar 4.12 Fitur Kedalam DataFrame skor.....	29
Gambar 4.13 Fitur terbaik berdasarkan nilai chi square.....	30
Gambar 4.14 Jumlah fitur yang ada	30
Gambar 4.15 Data train dan Data test	31
Tabel 4.16 Hasil evaluasi.....	31
Gambar 4.17 Hasil evaluasi	32
Gambar 4.18 Sentimen positif	33
Gambar 4.19 Sentimen negatif	34
Gambar 4.20 Grafik hasil pengujian.....	35

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi aplikasi pada smartphone berbasis Android saat ini terus berkembang sehingga mengubah cara dan perilaku masyarakat dalam beraktivitas sehari-hari. Masyarakat saat ini dimudahkan dengan adanya transportasi online berbasis aplikasi yang dapat diakses dimanapun dan kapanpun secara cepat (Abrilia&Sudarwanto,2020). Transportasi adalah kegiatan pemindahan barang (muatan) dan penumpang dari suatu tempat ke tempat lain. Dalam transportasi ada dua unsur yang terpenting yaitu pemindahan/pergerakan (movement) dan secara fisik mengubah tempat dari barang (comoditi) dan penumpang ke tempat lain. Transportasi sangatlah erat kaitannya dengan kehidupan sehari-hari, baik itu transportasi roda dua maupun roda empat. Namun semakin banyaknya masyarakat yang menggunakan transportasi dapat menyebabkan tingkat kemacetan di Indonesia, khususnya di kota-kota besar tengah mengalami peningkatan yang cukup signifikan. Maka dari itu kini pemerintah menghimbau masyarakat untuk menggunakan transportasi umum. Di zaman digitalisasi yang meningkat ini, transportasi umum sudah semakin berkembang, seperti terciptanya Transportasi Online atau yang biasa disebut Ojek Online. Transportasi Online merupakan bukti nyata dari perkembangan teknologi yang berbasis aplikasi. Jenis transportasi ini disambut dengan cukup baik mulai dari awal kemunculannya hingga saat ini dikarenakan masyarakat menganggap hal ini sebagai bentuk inovasi terbaik. Kini dengan adanya jasa ojek Online dapat

memberikan kemudahan untuk berpergian tanpa harus membawa kendaraan sendiri. Konsumen lebih memilih transportasi online karena dinilai lebih mudah dan cepat (Edo, et al 2019). Para pengguna jasa ojek tidak perlu lagi mencari driver ojek ke pangkalan, namun cukup dengan mengakses aplikasi jasa ojek online melalui smartphone kemudian melakukan proses order. Driver yang menerima order akan menjemput pengguna jasa tersebut sesuai dengan lokasi yang telah ditentukan pengguna jasa pada saat memesan layanan ojek online. Kemudahan ini lah yang menyebabkan ojek Online semakin ramai di minati saat ini. Kini semakin banyak juga perusahaan layanan ojek Online yang bermunculan. Salah satunya adalah Indrive.

Indrive merupakan perusahaan layanan jasa ojek Online yang berasal dari Rusia. inDrive didirikan pada tahun 2012 di Yakutsk. Ciri dari aplikasi ini yaitu berwarna hijau, begitu pula dengan warna jaket yang digunakan oleh para mitra drivernya. Indrive sudah dinikmati hampir diseluruh kota di indonesia. Indrive pun terkenal dengan harganya yang relatif murah. Seperti di sejumlah kota, yaitu Medan, Jakarta dan Bandung yang harga minimalnya berjumlahkan Rp.9000. dan bisa juga Rp.8000 tetapi jaraknya dari titik A ke titik B harus max 3.5km. Indrive salah satu aplikasi ojek online yang sangat mudah digunakan dimana para penumpang bisa memilih mau driver seperti apa, yang muda atau yang tua. Dan penumpang bisa juga memilih jarak atau waktu yang dekat maupun yang jauh. Maka untuk saat ini aplikasi indrive ialah aplikasi terpopuler setelah aplikasi gojek/grab. Indrive menempati posisi ulasan terbanyak diantara gojek/grab maupun aplikasi transportasi lainnya pada google playstore.

Google Playstore sebagai wadah pengunduh aplikasi ini memiliki fitur yang menarik yaitu sebuah fitur yang berisi ulasan dari para pengguna aplikasinya. Pada Google PlayStore terdapat ulasan pada setiap aplikasi.

Ulasan tersebut berisi opini pengguna mengenai kepuasan mereka terhadap aplikasi dan sering digunakan sebagai pertimbangan bagi calon pengguna dalam memilih aplikasi yang akan digunakan (Saputra dkk., 2019).

Google Play pada aplikasi indrive telah di download kurang lebih 100 juta unduhan dengan rating 4,5 dan tercatat 8 juta ulasan komentar pengguna aplikasi indrive pada situs google playstore. Dari sekian banyak nya ulasan, penulis akan melakukan penelitian analisis sentimen pada ulasan aplikasi indrive di Google PlayStore dengan menggunakan algoritma naive bayes. Diharapkan melalui penelitian ini bisa digunakan sebagai bahan evaluasi dan pertimbangan untuk pihak pengembang aplikasi dalam melakukan peningkatan, perbaikan dan penyempurnaan terhadap fitur-fitur serta layanan pada aplikasi Indrive

Naïve Bayes adalah sebuah metode klasifikasi yang memanfaatkan peluang sederhana dalam proses klasifikasinya (Susana et al., 2022). Naïve Bayes adalah metode yang sederhana dalam proses klasifikasi probabilitas dengan tingkat akurasi tinggi apabila diaplikasikan untuk data yang besar. Tingkat error (kesalahan) metode Naïve Bayes lebih kecil dibandingkan dengan metode klasifikasi yang lain. Metode Naive Bayes dipilih karena dinilai berpotensi baik untuk pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Tujuan dari Naive Bayes adalah

untuk menggambarkan atau menghitung secara langsung atribut data yang memiliki karakteristik kontinu. (Widiawati, F., et al., 2023).

1.2. Rumusan Masalah

Bagaimana cara mengklasifikasikan setiap opini pada ulasan aplikasi Indrive dari Google Play menjadi beberapa kategori sentimen yaitu positif dan negatif menggunakan metode naive bayes dengan *Google Collab*

1.3. Batasan Masalah

Dari uraian diatas dapat diambil beberapa batasan masalah seperti:

1. Data diambil berasal dari ulasan aplikasi indrive dengan melalui tahap web *scraping*
2. Data yang diolah hanya data teks dari ulasan yang berbahasa indonesia
3. Kelas sentimen yang digunakan hanya kelas sentimen positif dan negatif.

1.4. Tujuan Penelitian

Berdasarkan Rumusan Masalah diatas, adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Mengklasifikasi sentimen positif dan sentimen negatif pada *google play store* dengan menggunakan metode naive bayes
2. Untuk Mengetahui tingkat kepuasan pengguna aplikasi indrive terhadap fitur-fitur aplikasi didalamnya
3. Mengetahui cara pengimplementasian ulasan-ulasan indrive di google playstore dengan menggunakan *Google collab*.

1.5. Manfaat Penelitian

Hasil studi ini dinantikan dapat memperkaya ilmu pengetahuan terutama dalam menganalisis sentimen ulasan-ulasan pada aplikasi indrive di google playstore serta memberikan masukan kepada pengembang aplikasi yang nantinya masyarakat percaya bahwasanya aplikasi indrive cocok dan tepat dalam pemesanan transportasi online.

1. Bagi Peneliti

Hasil penelitian dapat membantu memperluas pengetahuan peneliti mengenal lebih jauh tentang bagaimana ulasan dari aplikasi indrive tersebut, serta memberi masukan terhadap pengembang aplikasi terhadap fitur-fitur yang harus di tambahkan atau di perbarui untuk kedepannya

2. Bagi Universitas

Hasil penelitian ini dapat menjadi referensi tambahan dipergustakaan dan perbandingan dalam penelitian-penelitian berikutnya.

3. Bagi Pengguna Indrive

Hasil penelitian ini dapat mengetahui seberapa banyaknya ulasan positif dan negatif dari pengguna aplikasi indrive. Apakah aplikasi indrive cocok dijadikan alat transportasi online dibandingkan dengan transportasi online lainnya.

BAB II

LANDASAN TEORI

2.1 InDrive

InDrive didirikan pada tahun 2012 di Yakutsk, salah satu kota terdingin di dunia. Layanan ini berawal ketika sekelompok mahasiswa lokal membentuk kelompok kolektif "pengemudi independen" (yaitu inDrivers) di media sosial sebagai respons terhadap kenaikan tajam harga taksi ketika suhu luar turun drastis. Anggota kelompok berbagi tujuan mereka dan harga yang bersedia mereka bayar, yang memungkinkan pengemudi untuk menghubungi pelanggan ini dan menegosiasikan harga. Setahun kemudian, kelompok tersebut pindah ke Perusahaan Sinet yang menciptakan sistem transportasi peer-to-peer berdasarkan konsep ini. Pada tahun 2013, inDriver resmi diluncurkan. Setelah inovasi Rusia ke Ukraina, inDrive melepas kepentingannya di pasar Rusia pada bulan Juni 2022, dan merelokasi 1.000 karyawan dari Rusia ke Kazakhstan dan Siprus.

inDrive kemudian meluncurkan layanan lain, termasuk perjalanan antar kota, kurir, pengiriman barang, dan inDrive.Services, yang memungkinkan orang menemukan penyedia layanan di berbagai bidang. inDrive.Services tersedia di Mesir, Kazakhstan, Meksiko, Chili, Ekuador, Kolombia, dan Brasil. Pada bulan Desember 2014, inDriver melakukan ekspansi internasional dengan meluncurkannya di Astana, Kazakhstan.

Hingga saat ini, layanan ini telah memperluas bisnisnya ke seluruh Amerika Latin, Afrika, India, dan Asia Barat Daya. Pada tahun 2018, perusahaan membuka kantor sementara di New York City, dan kemudian, mendirikan kantor pusatnya di Mountain View, California. Aplikasi ini juga memulai operasinya

di Pakistan pada tahun 2021 dan menjadi aplikasi pemesanan kendaraan yang paling banyak diunduh di negara tersebut dalam waktu satu tahun. Pada bulan Juni 2022, inDriver mengumumkan bahwa mereka memasuki pasar Australia, dengan peluncuran awal di Melbourne. Pada bulan Juli 2023, inDrive diluncurkan di AS, di Miami, Florida Selatan.

InDrive juga menempati peringkat keempat sebagai aplikasi perjalanan yang paling banyak diunduh. InDrive. inDrive mulai beroperasi di Indonesia pada 2021. Jumlah karyawan perusahaan ini di seluruh dunia lebih dari 2.800. Selain ride-hailing, inDrive menyediakan daftar layanan perkotaan seperti transportasi antar kota, pengiriman barang, bantuan tugas, kurir dan pencarian pekerjaan. InDrive memiliki keunggulan dibandingkan dengan transportasi lainnya diantaranya yaitu:

1. Menawarkan jangkauan layanan yang lebih luas, termasuk pengantaran ke luar kota. Pengguna dapat memesan mobil untuk mengantar mereka ke destinasi kota tujuan, dan pengemudi juga dapat menawarkan jasa mengantar ke luar kota dengan tarif yang ditentukan sendiri oleh pengguna dan pengemudi.

2. Membebaskan biaya jasa aplikasi sebesar 10% kepada mitra pengemudi, yang lebih rendah dibandingkan dengan aplikasi lain seperti Grab dan Gojek yang menerapkan biaya sebesar 15%.

Serta kekurangan aplikasi inDrive daripada transportasi lainnya ialah:

1. Tidak memiliki kantor seperti gojek dan grab. Dan apabila ada tindak kejahatan terhadap driver susah memberikan pengaduan.

2. Sering terjadinya update dalam kenaikan versi oleh karena itu penumpang yang ingin memesan di haruskan melakukan update aplikasi.

2.2 Analisis Sentimen

Pengertian Analisis Sentimen menurut Bing Liu (2020:1) adalah bidang studi yang menganalisis opini, sentimen, penilaian, sikap, dan emosi orang terhadap entitas dan atributnya yang diungkapkan dalam teks tertulis. Sentiment analysis merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentiment yang terkandung dalam sebuah opini (Sukawirasa et al., 2019). Pada Sentiment analysis terdiri dari 3 jenis opini, yaitu opini positif, opini negatif, dan opini netral sehingga dengan sentimen analysis perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui feedback masyarakat atau pun para ahli (Hafidz, 2021), (V. A. Safitri et al., 2019). Sentiment mengacu pada fokus topik tertentu, pernyataan pada suatu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subjek yang berbeda, oleh karena itu pada beberapa penelitian, terutama pada review ulasan aplikasi, pekerjaan didahului dengan menentukan elemen dari sebuah aplikasi yang sedang dibicarakan sebelum memulai proses Sentiment Analysis.

2.3 Naïve Bayes

Naïve Bayes merupakan salah satu algoritma dari machine learning. Dalam perkembangan database, *Naïve Bayes* termasuk supervised learning yaitu jenis machine learning yang membutuhkan sampel sebagai data latih yang memiliki label. Supervised learning dikelompokkan menjadi dua yaitu klasifikasi dan regresi. Klasifikasi pada saat variabel menjadi kategori seperti merah atau kuning,

penyakit atau tidak ada penyakit, dan sebagainya.

Metode ini merupakan bagian dari metode bayes yang digunakan dalam klasifikasi teks didasarkan pada model penyederhanaan bahwa nilai atribut secara kondisional bersifat bebas jika nilai output diberikan. Metode ini juga digunakan dalam penambangan teks pada analisis sentimen dengan cara memprediksi probabilitas berdasarkan data sebelumnya. Naïve Bayes tidak bisa mendeteksi gambar, tetapi hanya bisa mendeteksi teks dan numerik. Untuk menghitung probabilitas dalam metode ini menggunakan pendekatan teorema bayes (Prabowo & Wiguna, 2021).

Teorema bayes ditemukan oleh Thomas Bayes yaitu seorang pendeta dari Inggris pada tahun 1763 dan disempurnakan oleh Laplace. Teorema bayes adalah pengenalan pola melalui pendekatan statistik yang fundamental. Teorema bayes dapat dideskripsikan seperti probabilitas terjadinya hubungan A dengan syarat hubungan B sudah terjadi, begitupun sebaliknya. Dalam bidang kedokteran modern, teorema bayes sering digunakan. Teorema bayes berperan dalam memperbaiki hitungan probabilitas dengan memanfaatkan data informasi tambahan (Retnosari, 2021).

2.4 Google Collab

Google Colab adalah layanan dari Google untuk menjalankan dan mengelola notebook Jupyter berbasis cloud secara gratis. Ini memungkinkan pengguna untuk menulis dan menjalankan kode Python, menyimpan dan berbagi proyek (Haekal Hilmi Zain, dkk 2023). Google Colab adalah produk yang dapat menjadi pilihan tepat jika Anda memerlukan coding environment dengan format

notebook yang userfriendly. Google Colab merupakan produk dari Google Internal Research. Google Colaboratory, atau yang lebih umum dikenal sebagai dengan Google Colab adalah compiler dan editor bahasa pemrograman Python secara online. Dengan Google Colab, kita tidak perlu lagi menginstal perangkat lunak tambahan. Untuk menggunakan Google Colab, kita cukup mempunyai akun Google untuk sebagai syarat login dan tempat menyimpan file Python di Google Drive.

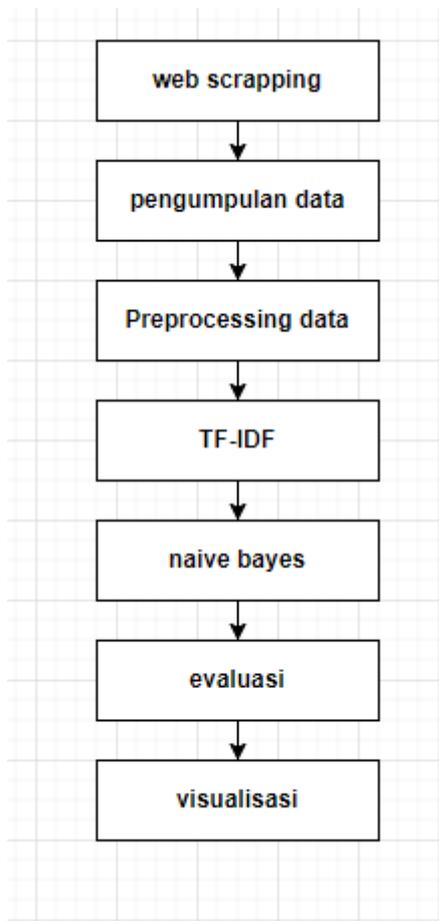
2.5 Python

Python adalah bahasa pemrograman tinggi yang bisa melakukan eksekusi sejumlah instruksi multi guna secara langsung (interpretatif) dengan metode Object Oriented Programming dan juga menggunakan semantik dinamis untuk memberikan tingkat keterbacaan syntax. Sebagai bahasa pemrograman tinggi, python dapat dipelajari dengan mudah karena telah dilengkapi dengan manajemen memori otomatis (Simanjuntak., et al., 2022). bahasa pemrograman populer yang dapat digunakan untuk berbagai hal, mulai dari membangun website, menganalisis data, scripting, hingga pembuatan game. Python adalah bahasa pemrograman yang bersifat open source, sehingga Anda dapat menggunakannya secara gratis. Bahkan, Anda pun bisa ikut berkolaborasi untuk pengembangannya. Sebagian besar programmer tentu setuju dengan pernyataan bahwa Python adalah bahasa pemrograman yang interpretatif dan multiguna. Sintaksnya mudah dibaca dan dipahami, sehingga pemula pun bisa dengan mudah mempelajarinya.

BAB III METODOLOGI PENELITIAN

3.1. Alur Penelitian

Algoritma Naive bayes digunakan dalam penelitian ini untuk menganalisis sentimen pengguna aplikasi Indrive berdasarkan ulasan yang ada di Google PlayStore dan menerapkannya pada software Googlecolab. Berikut gambar 1 yang menjadi alur proses penelitian kali ini:



Tabel 3.1 Alur penelitian

3.1.1 Web scrapping

Web scraping adalah salah satu metode paling efisien untuk mengekstrak data dari sebuah website. Data atau informasi yang diambil biasanya dalam jumlah besar, dan digunakan untuk keperluan riset atau analisis. Pada tahap scrapping ini, diambil data ulasan pada aplikasi indrive di google playstore yang kemudian data tersebut di download ke microsoft excel melalui google colab.

3.1.2 Pengumpulan data

Setelah proses scrapping tadi, dapatlah data ulasan tersebut yang dimana data tersebut berisikan 1000 data sebagai berikut:

No	UserName	Review	Rating	Sentimen
1	Ulfah Hmytl	Min mau tanya apakah keluar kota aman pake indrive contohnya dari bandung ke jakarta,soalnya takut penipuan malah rampok atau terjadi hal2 yg tidak diinginkan.??? Tolong jawab min	5	Negatif
2	Dimas arief	OK	5	Positif
3	Krismanto Aris	aplikasi yg sangat membantu	5	Positif
4	Zildjian Idris	Mantap	5	Positif
5	Zhao zilong	Uang saya nyangkut diaplikasi orderan pertama,cs nya respon lambat membantunya	2	Positif
6	Op. Choki Siallagan	Suka hoaks...tidak bisa koneksi	2	Positif
.....

995	Reza Akbar	Ga bisa diinstall	1	Positif
996	Fuji fermana	Mantaaaap	5	Positif
997	Yusman Handika	Seharusnya klo sudah upload dokumen dan sudah dibuat untuk narik penumpang gak dimintain dokumen lagi	1	Negatif
998	Muh riyanto	Goodjob	5	Positif
999	Liberti Manalu	mantap	5	Positif
1000	SU TIMAH	Tolong d perbaharui aplikasi indrive versi 5.94.0 buat penumpang?	5	Positif

Tabel 3.1.2 Data Ulasan

Data data di atas di jelaskan bahwa:

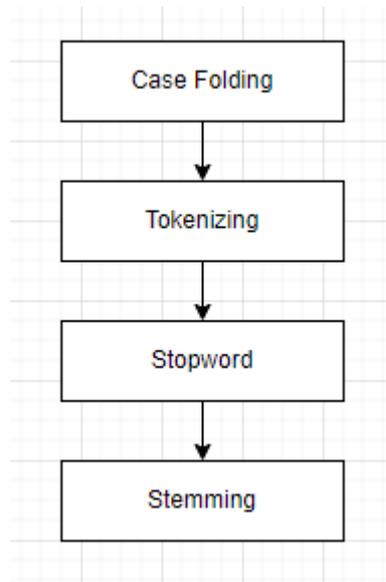
- **Username**: Nama yang memberikan komentar pada ulasan aplikasi indrive
- **Review** : Berisikan keluhan/ulasan dari pengguna aplikasi indrive
- **Rating** : Yang dimaksud score disini ialah bintang yang diberikan
- **Sentimen** : Berisikan hasil antara positif dan negatif

3.1.3 Pelabelan data

Pelabelan data pada Naive Bayes adalah proses memberikan nilai kebenaran atau label pada dataset yang akan digunakan sebagai acuan untuk pengklasifikasian. Dalam Naive Bayes, dataset harus diberi label terlebih dahulu sebelum dapat digunakan untuk pelatihan model. Label ini digunakan untuk menentukan kelas atau kategori dari setiap data point, sehingga model dapat memprediksi kelas atau kategori baru berdasarkan probabilitas yang dihitung.

3.1.4 Preprocessing data

Tahap selanjutnya adalah preprocessing data, ditahap ini data yang discraping kemudian dikumpulkan melalui microsoft excel akan di sama ratakan bentuk dan formatnya agar data dapat diproses pada tahap selanjutnya. Tahap preprocessing text ini terdiri dari 4 tahapan yaitu case folding, tokenizing, filtering, dan stemming yang dapat dilihat pada gambar dibawah ini:



Flowchart 3.1.4 Tahap Preprocessing data

1. Case Folding

Tahap case folding ini merupakan proses penyamataan huruf yang sebelumnya terdapat huruf kapital diubah menjadi huruf kecil semua.

2. Tokenizing

Tahap selanjutnya adalah tahap tokenizing, tahapan ini merupakan proses pemotongan atau pemisahan kalimat menjadi kata per kata.

3. Stopword

Tahap selanjutnya setelah tokenizing adalah tahap Stopword, tahap ini akan menghapus semua kata yang dianggap tidak mempunyai makna, seperti kata sambung

4. Stemming

Tahapan selanjutnya yaitu tahap stemming yaitu untuk mengubah kata yang memiliki imbuhan menjadi kata dasar.

3.1.5 Pembobotan TF-IDF

Setelah dilakukan tahap preprocessing data, tahapan selanjutnya yang harus dilakukan

adalah tahap TF-IDF. Tahap ini melakukan proses tf(term weighting) yang setelahnya akan diberikan bobot setiap kata yang dimana bobot tersebut mengindikasikan pentingnya sebuah term pada sebuah dokumen. Perhitungan bobot tiap term dicari keseluruhan pada setiap dokumen bertujuan untuk mengetahui dan mendapat jumlah kata yang sering muncul pada dokumen dan semakin sering kata tersebut muncul maka semakin tinggi pula bobot atau nilai term tersebut.

3.1.6 Naive Bayes

Tahapan setelah TF-IDF yaitu tahapan klasifikasi model menggunakan algoritma Naive Bayes. Naive Bayes merupakan metode klasifikasi sederhana yang bisa menghitung semua kemungkinan dengan cara menggabungkan jumlah kombinasi dan kerapatan nilai dari suatu dataset yang sudah diperoleh (Rachman and Handayani, 2021). Algoritma klasifikasi naive bayes merupakan model algoritma yang sudah sering digunakan untuk melakukan klasifikasi data dengan menggunakan perhitungan probabilitas. Perhitungan probabilitas untuk setiap kata pada dokumen atau dataset dapat menghasilkan 2 klasifikasi yaitu sentimen positif dan negatif dari proses komputasi yang sebelumnya pada pembobotan kata TF-IDF (Chatrina, Siregar, Ruli, A, Siregar and Yoga, Distra, Sudirman,2020). Adapun kelebihan menggunakan metode ini yaitu membutuhkan sedikit data training untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi. Karena diasumsikan sebagai variabel yang bebas (Alam, Resmi and Masripah, 2022).

3.1.7 Evaluasi

Tahap selanjutnya ialah tahap evaluasi menggunakan confusion matrix, confusion matrix adalah metode evaluasi yang digunakan untuk mengukur kinerja metode

klasifikasi. Terdapat empat istilah yang merepresentasikan hasil dari proses klasifikasi, yaitu TP, TN, FP, dan FN.

Setelah mendapatkan nilai dari keempat istilah tersebut, selanjutnya dapat dihitung nilai akurasi, nilai presisi, dan juga recall menggunakan rumus berikut :

$$\text{Accuracy} = \frac{TP+TN}{TOTAL DATA}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

Diketahui bahwa:

- 1. True Positive (TP):** Jumlah data yang bernilai Positif dan diprediksi benar sebagai Positif.
- 2. False Positive (FP):** Jumlah data yang bernilai Negatif tetapi diprediksi sebagai Positif.
- 3. False Negative (FN):** Jumlah data yang bernilai Positif tetapi diprediksi sebagai Negatif.
- 4. True Negative (TN):** Jumlah data yang bernilai Negatif dan diprediksi benar sebagai Negatif

3.1.8 Visualisasi data

Tahap selanjutnya setelah dilakukan evaluasi yaitu tahap visualisasi menggunakan wordcloud. Tahap ini digambarkan dengan word cloud pada seluruh sentimen keseluruhan seluruh dokumen ulasan positif dan negative tentang aplikasi Indriver. Tujuan utama dari visualisasi data adalah untuk membantu pengguna memahami makna dan struktur data, serta mendukung pengambilan keputusan yang lebih baik

dengan mengidentifikasi pola-pola yang tersembunyi.

3.2. Waktu dan Tempat Penelitian

3.2.1 Waktu Penelitian

NO	KEGIATAN PENELITIAN	BULAN								
		FEB	MAR	APR	MEI	JUNI	JULI	AGU	SEP	OKT
1.	Persiapan Penelitian									
	a. Pengajuan Judul				■					
	b. Pengajuan SK Pembimbing				■					
	c. Observasi				■	■				
	d. Penyusunan Proposal					■				
	e. Seminar Proposal					■				
2.	Implementasi & Pengumpulan Data									
	a. Pembuatan Sistem						■	■		
	b. Pengumpulan Data						■	■		
3.	Pemrosesan Data & Pelaporan									
	a. Validasi & Hasil							■	■	
	b. Penyusunan Laporan Skripsi							■	■	

Tabel 3.2. 2 Jadwal Penelitian

3.2.2 Tempat Penelitian

Penelitian ini dilakukan pada Google Playstore menggunakan aplikasi Google Colab dengan scrapping data.

3.3. Perangkat Penelitian

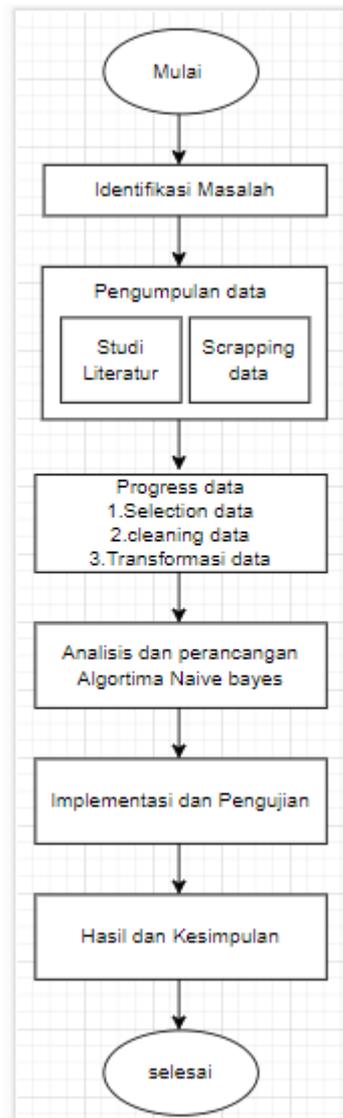
Perangkat pada penelitian ini berupa Hardware dan software:

Hardware	Laptop HP	AMD Ryzen 5
		8GB RAM
		HDD
		MONITOR
Software	Sistem Operasi	WINDOWS11
Tools	Google playstore	GOOGLE COLAB

Tabel 3.3 Perangkat penelitian

3.4 Flowchart Sistem Klasifikasi

Flowchart memiliki fungsi untuk menggunakan, menyederhanakan suatu rangkaian proses atau prosedur sehingga dapat dengan mudah dipahami dan mudah dilihat berdasarkan urutan Langkah dari suatu proses. Berikut ini adalah flowchart dari keseluruhan yang digunakan sebagai acuan bahan penelitian



Tabel 3.4 Flowchart Sistem Klasifikasi

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Web scrapping

Pengambilan data dilakukan dengan metode scraping menggunakan Google Colaboratory dengan menginstall google-play-scraper untuk mendapatkan data ulasan pada google playstore. Data yang didapatkan berupa review id, username, ulasan, rating, tanggal ulasan seperti pada gambar dibawah ini:

+ Kode + Teks

```
0d ✓ ▶ from google_play_scraper import Sort, reviews
import pandas as pd
from tqdm import tqdm

# Fungsi untuk mengambil ulasan dari Google Play Store
def get_reviews(app_id, lang='id', country='id', max_reviews=1000):
    all_reviews = []
    result, _ = reviews(
        app_id,
        lang=lang,
        country=country,
        sort=Sort.NEWEST,
        count=max_reviews,
        filter_score_with=None
    )
    for review in result:
        all_reviews.append({
            'review_id': review['reviewId'],
            'user_name': review['userName'],
            'rating': review['score'],
            'review': review['content'],
            'at': review['at']
        })
    return pd.DataFrame(all_reviews)
```

```
# ID aplikasi InDrive di Google Play Store
app_id = 'sinet.startup.inDriver'

# Mengambil ulasan
data = get_reviews(app_id, max_reviews=1000)
data.head()
```

Gambar 4.1 Web scrapping pada pyhton

	A	B	C	D	E
1	review_id	user_name	rating	review	at
2	0bb4caa5-	Ulfah Hmytl	5	Min mau tanya apakah keluar	24/09/2024 19:28
3	471b3fe3-	dimas arief	5	Ok	24/09/2024 18:59
4	38d7b786-	Krismanto Aris	5	aplikasi yg sangat membantu	24/09/2024 17:19
5	3e810d07-	Zildjian Idris	5	Mantap	24/09/2024 17:14
6	a9edf254-	Seftian	2	Uang saya nyangkut diaplikas	24/09/2024 17:05
7	20248fd2-	Op. Choki Siallagan	2	Suka hoaks...tidak bisa konek	24/09/2024 17:05
8	1876fe86-	Rolly Pupun	5	Klu bisa aplikasi nya di perbai	24/09/2024 16:55
9	bc8ab174-	Arif Gunawan	1	makin sini makin susah dapat	24/09/2024 16:03
0	926061b7-	Fadly Syahputra.	1	hilangkan nego argo dong ma	24/09/2024 15:57
1	1510ba4c-	Ali Rahman	5	Istimewa	24/09/2024 15:57
2	37fb9ea1-	Adit ae	1	Teman saya habis kenak romp	24/09/2024 15:23
3	5bab8452-	Miftachul Mufid	1	Butuh waktu berapa lama bag	24/09/2024 15:18
4	8538f236-	Dodol Garut	5	Baik.. Bagus..ðŸ˜ƒ	24/09/2024 15:01
5	805c0a04-	Yuanta	1	Pengemudi nya sangat meng	24/09/2024 14:38
6	28e0ce44-	Hendarto Sachro	5	good service and excellent	24/09/2024 14:21
7	d3e49ae4-	Eka adila Adila	1	Autobitnya kurang support,m	24/09/2024 14:19
8	0576c8e4-	á´á´ceá´á´ce á´á´É™á´†	3	Sungguu saya sangat kesal & r	24/09/2024 14:10
9	afc27dd3-	joko Sugiarto	1	Aplikasi parah buruk...autobi	24/09/2024 14:04
0	364d10f7-	Ikhsan Jhon	5	aku mendukung semua apk p	24/09/2024 13:50

Gambar 4.2 Hasil Web scrapping

4.2 Hasil Pelabelan Data

Hasil labelan data ini berbentuk file csv seperti tabel dibawah ini:

	A	B	C	D	E	F	G	H	I
1	review_id	user_name	rating	review	at	clean_rev	sentimen	score	sentimen
2	0bb4caa5-	Ulfah Hmy	5	Min mau t	24/09/2024 19:28	min mau t	1 star	0.4082152	Negatif
3	471b3fe3-	dimas arie	5	Ok	24/09/2024 18:59	ok	3 stars	0.4951144	Positif
4	38d7b786-	Krismantc	5	aplikasi yg	24/09/2024 17:19	aplikasi yg	5 stars	0.6137180	Positif
5	3e810d07-	Zildjian Id	5	Mantap	24/09/2024 17:14	mantap	3 stars	0.2485850	Positif
6	a9edf254-	Seftian	2	Uang saya	24/09/2024 17:05	uang saya	5 stars	0.2480365	Positif
7	20248fd2-	Op. Choki	2	Suka hoak	24/09/2024 17:05	suka hoak	4 stars	0.4092351	Positif
8	1876fe86-	Rolly Pupu	5	Klu bisa a	24/09/2024 16:55	klu bisa a	3 stars	0.4416875	Positif
9	bc8ab174-	Arif Gunav	1	makin sini	24/09/2024 16:03	makin sini	2 stars	0.3235483	Negatif
10	926061b7-	Fadly Sya	1	hilangkan	24/09/2024 15:57	hilangkan	1 star	0.6488752	Negatif
11	1510ba4c-	Ali Rahma	5	Istimewa	24/09/2024 15:57	istimewa	5 stars	0.4993018	Positif
12	37fb9ea1-	Adit ae	1	Teman say	24/09/2024 15:23	teman say	1 star	0.7554755	Negatif
13	5bab8452-	Miftachul	1	Butuh wal	24/09/2024 15:18	butuh wal	2 stars	0.3931011	Negatif
14	8538f236-	Dodol Gar	5	Baik.. Bag	24/09/2024 15:01	baik bagu	5 stars	0.4305090	Positif
15	805c0a04-	Yuanta	1	Pengemu	24/09/2024 14:38	pengemu	1 star	0.4448235	Negatif
16	28e0ce44-	Hendarto	5	good serv	24/09/2024 14:21	good serv	5 stars	0.7624075	Positif
17	d3e49ae4-	Eka adila A	1	Autobitny	24/09/2024 14:19	autobitny	2 stars	0.4807372	Negatif
18	0576c8e4-	á'á'ceá'á'	3	Sungguu s	24/09/2024 14:10	sungguu s	2 stars	0.5156812	Negatif
19	afc27dd3-	joko Sugie	1	Aplikasi p	24/09/2024 14:04	aplikasi p	1 star	0.5804567	Negatif
20	364d10f7-	Ikhsan Jhc	5	aku mend	24/09/2024 13:50	aku mend	5 stars	0.5652607	Positif

Gambar 4.3 Hasil Pelabelan Data

4.3 Hasil Preprocessing Data

Preprocessing data merupakan tahapan selanjutnya setelah tahap web scraping, pada tahap ini data akan di samakan seluruh bentuk dan format nya agar pada tahap selanjutnya data dapat diproses dan diolah. Tahap preprocessing text terdiri dari 4 tahapan yaitu, case folding, tokenizing, stopword, dan stemming.

Case Folding

Pada tabel 4.3 baris ke 7, menunjukkan hasil dari proses case folding dimana data sebelum di case folding masih terdapat huruf kapital lalu setelah proses case folding semua huruf kapital diubah menjadi huruf kecil semua

4	38d7b786-	Krismanto Aris	5	aplikasi yg sangat membantu
5	3e810d07-	Zildjian Idris	5	Mantap
6	a9edf254-	Seftian	2	Uang saya nyangkut diaplikasi orderan pertama,
7	20248fd2-	Op. Choki Siallagan	2	Suka hoaks...tidak bisa koneksi
8	1876fe86-	Rolly Pupun	5	Klu bisa aplikasi nya di perbaiki lagi.karena titik

Tabel 4.4. Data Case Folding

```

import re

# Fungsi case folding untuk teks
def casefolding(text):
    if isinstance(text, str): # Pastikan 'text' adalah string
        text = text.lower() # Mengubah teks menjadi huruf kecil
        text = re.sub(r'https?://\S+|www\.\S+', '', text) # Menghilangkan URL
        text = re.sub(r'[-+]?[0-9]+', '', text) # Menghilangkan angka
        text = re.sub(r'^\w\s', '', text) # Menghilangkan tanda baca
        return text
    else:
        return text # Jika bukan teks, kembalikan apa adanya

# Ambil sampel dari kolom teks, misalnya 'content' atau 'label'
raw_sample = data['review'].iloc[5] # Asumsi 'content' adalah kolom yang berisi teks tweet
case_folding = casefolding(raw_sample)

print('Raw data\t: ', raw_sample)
print('Case folding\t: ', case_folding)

```

Raw data : Suka hoaks...tidak bisa koneksi
Case folding : suka hoakstidak bisa koneksi

Gambar 4.5 Hasil Case Folding

Tokenizing

Setelah proses case folding, data akan dilakukan proses tokenizing yaitu memecah kalimat ulasan menjadi kata per kata seperti yang dapat dilihat pada gambar 2

```

✓ id
▶
↳
sent_tokenize \
0 Min mau tanya apakah keluar kota aman pake ind...
1 Ok
2 aplikasi yg sangat membantu
3 Mantap
4 Uang saya nyangkut diaplikasi orderan pertama ...

word_tokenize \
0 [Min, mau, tanya, apakah, keluar, kota, aman, ...
1 [Ok]
2 [aplikasi, yg, sangat, membantu]
3 [Mantap]
4 [Uang, saya, nyangkut, diaplikasi, orderan, pe...

spanish_tokenize \
0 Min mau tanya apakah keluar kota aman pake ind...
1 Ok
2 aplikasi yg sangat membantu
3 Mantap
4 Uang saya nyangkut diaplikasi orderan pertama ...

gensim_tokenize
0 Min mau tanya apakah keluar kota aman pake ind...
1 Ok
2 aplikasi yg sangat membantu
3 Mantap
4 Uang saya nyangkut diaplikasi orderan pertama ...

```

Gambar 4.6 Hasil tokenizing

1. sent_tokenize

sent_tokenize adalah fungsi dari pustaka nltk (Natural Language Toolkit) yang digunakan untuk membagi teks menjadi kalimat-kalimat terpisah. Ini berguna untuk analisis teks karena sering kali kita ingin mengidentifikasi struktur kalimat sebelum melakukan analisis lebih lanjut.

2. word_tokenize

word_tokenize adalah fungsi dari pustaka nltk yang membagi teks menjadi kata-kata individual. Ini berguna untuk analisis kata, seperti analisis frekuensi atau pembuatan model berbasis kata.

3. `spanish_tokenize`

`spanish_tokenize` tidak secara eksplisit ada di pustaka standar seperti `nlk`. Namun, beberapa pustaka atau algoritma mungkin memiliki fungsi serupa yang didesain khusus untuk tokenisasi teks dalam bahasa Spanyol. Dalam konteks yang umum, ini berarti tokenisasi teks dalam bahasa Spanyol, memperhatikan aturan gramatikal dan penggunaan tanda baca bahasa tersebut.

4. `gensim_tokenize`

`gensim_tokenize` adalah metode tokenisasi dari pustaka `gensim`, yang dikenal dengan alat pemodelan topik dan pemrosesan teks. `Gensim` menyediakan fungsi tokenisasi yang sering digunakan dalam konteks pemodelan topik dan pemrosesan teks skala besar.

Stopword

Setelah tahap tokenizing data akan masuk kedalam tahap filtering yaitu memfilter atau menghilangkan kata-kata yang tidak diperlukan dalam tahapan berikutnya seperti yang dapat dilihat pada tabel 3.

```

✓ [87] stopwords_ind = stopwords.words('indonesian')
0d
↳ [nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

✓ [88] from nltk.tokenize import sent_tokenize, word_tokenize
0d
from nltk.corpus import stopwords

stopwords_ind = stopwords.words('indonesian')

✓ [90] len(stopwords_ind)
0d
▶ 758

✓ [90] stopwords_ind
0d
↳ 'sebetulnya',
'sebisanya',
'sebuah',
'sebut',
'sebutlah',
'sebutnya',
'secara',
'secukupnya',
'sedang',
'sedangkan'.

```

Gambar 4.7 Stopword Pada google colab

```

↳ Raw data : Suka hoaks...tidak bisa koneksi
Case folding : suka hoakstidak bisa koneksi
Stopword removal : suka hoakstidak koneksi

```

Gambar 4.8 Hasil Stopword

Stemming

Tahapan selanjutnya adalah proses stemming, yaitu data yang sudah diproses pada tahap sebelumnya akan diubah kata imbuhan nya menjadi kata dasar. Untuk proses stemming ini diperlukan library Sastrawi untuk proses

selanjutnya dilakukan. Untuk hasil proses stemming dapat dilihat pada tabel 4

```
▶ raw_sample = data['review'].iloc[5]
  case_folding = str(casefolding(raw_sample))
  stopword_removal = remove_stop_words(case_folding)
  text_stemming = stemming(stopword_removal)

  print('Raw data\t\t: ', raw_sample)
  print('Case folding\t\t: ', case_folding)
  print('Stopword removal\t: ', stopword_removal)
  print('Stemming\t\t: ', text_stemming)
```

⇒ Raw data : Suka hoaks...tidak bisa koneksi
Case folding : suka hoakstidak bisa koneksi
Stopword removal : suka hoakstidak koneksi
Stemming : suka hoakstidak koneksi

Gambar 4.9 Hasil Stemming

4.4 Hasil Pembobotan TF IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode yang digunakan untuk menilai pentingnya suatu kata dalam sebuah dokumen relatif terhadap koleksi dokumen (corpus). Dalam konteks algoritma Naive Bayes, pembobotan ini membantu dalam mengukur kontribusi setiap kata terhadap klasifikasi suatu dokumen. Berikut hasil nya pada google colab:

```
# Transformasi fitur menggunakan TF-IDF
x_tf_idf = vec_TF_IDF.transform(X)

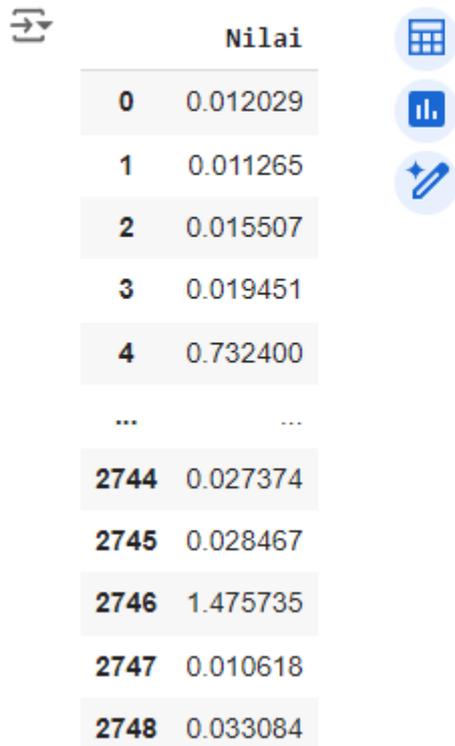
# Menyimpan vocabulary dari TF-IDF menggunakan pickle
pickle.dump(vec_TF_IDF.vocabulary_, open("feature_tf-idf.pkl", "wb"))

# Melihat vocabulary dari TF-IDF
vec_TF_IDF.vocabulary_

{'min': 1593,
 'mau': 1419,
 'tanya': 2475,
 'apakah': 95,
 'keluar': 1113,
 'kota': 1218,
 'aman': 57,
 'pake': 1871,
 'indrive': 919,
 'contohnya': 469,
 'dari': 515,
 'bandung': 189,
 'ke': 1076,
 'jakartasoalnya': 963,
 'takut': 2453,
 'penipuan': 1980,
 'malah': 1352,
 'rampok': 2118,
 'atau': 127,
```

Gambar 4.10 Volabulary TF IDF

```
[105] # Menampilkan skor chi-square untuk tiap fitur
df_scores = pd.DataFrame(chi2_features.scores_, columns=['Nilai'])
df_scores
```



	Nilai
0	0.012029
1	0.011265
2	0.015507
3	0.019451
4	0.732400
...	...
2744	0.027374
2745	0.028467
2746	1.475735
2747	0.010618
2748	0.033084

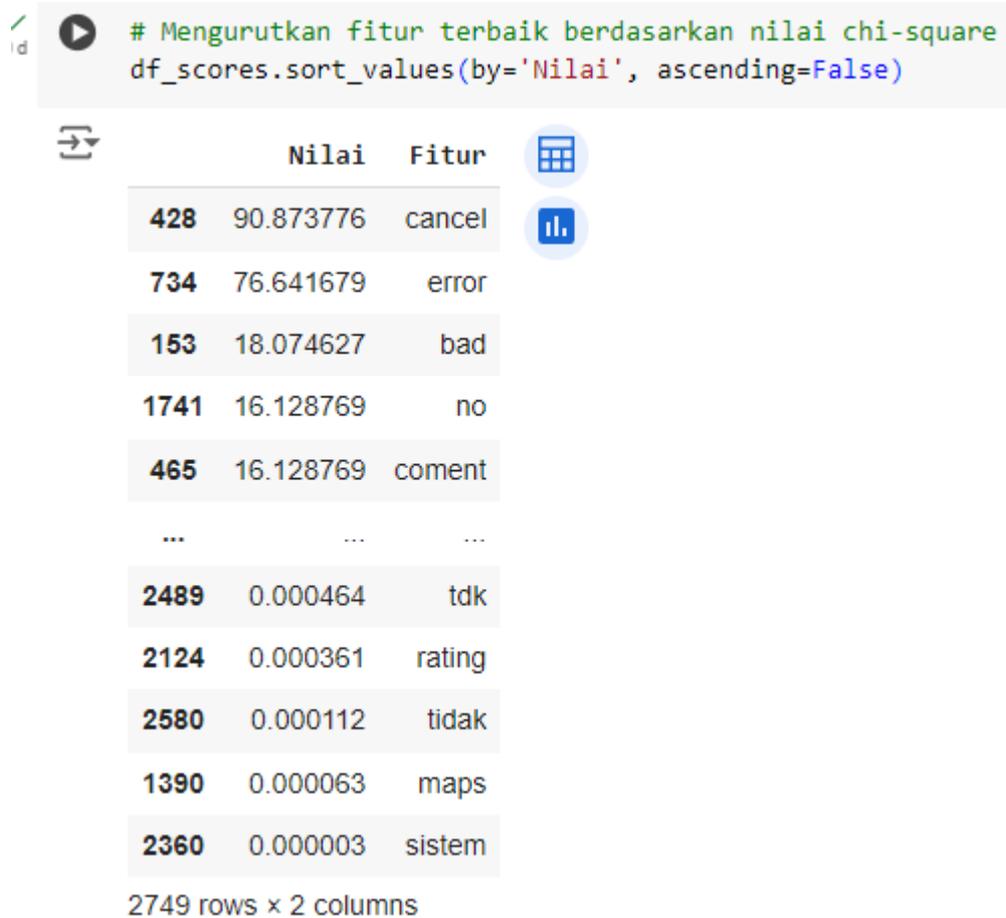
Gambar 4.11 Skor chi-square untuk tiap fitur TF IDF

```
# Menambahkan nama fitur ke dalam DataFrame skor
feature = vec_TF_IDF.get_feature_names_out()
df_scores['Fitur'] = feature
df_scores
```

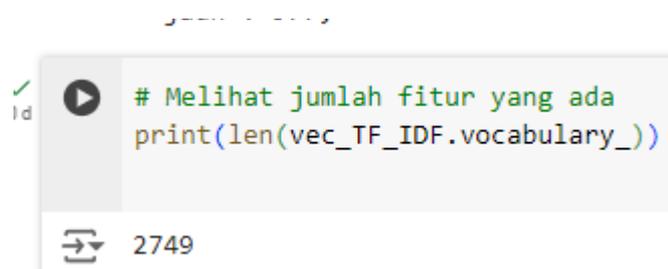


	Nilai	Fitur
0	0.012029	abis
1	0.011265	acc
2	0.015507	actionnyatindak
3	0.019451	ad
4	0.732400	ada
...
2744	0.027374	yh
2745	0.028467	yng
2746	1.475735	you
2747	0.010618	your
2748	0.033084	yya

Gambar 4.12 Fitur Kedalam DataFrame skor



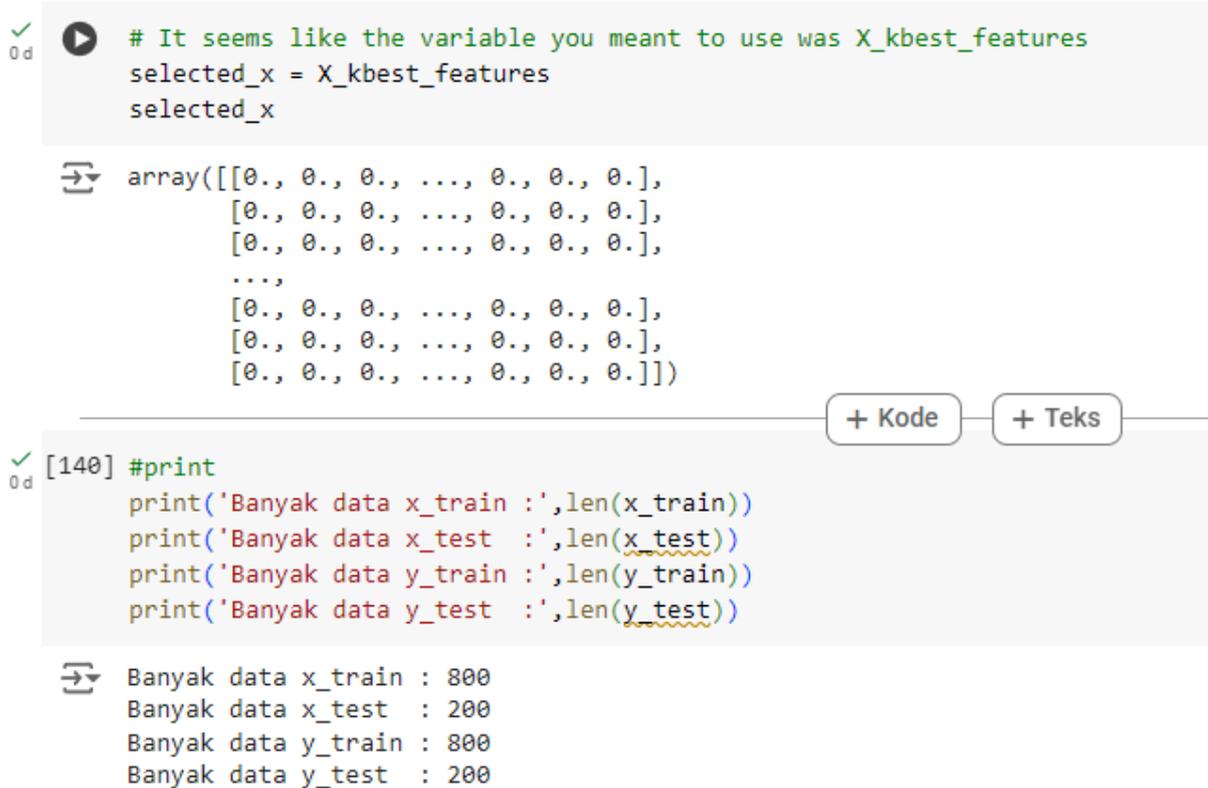
Gambar 4.13 Fitur terbaik berdasarkan nilai chi square



Gambar 4.14 Jumlah fitur yang ada

4.5 Hasil Naive Bayes

Pada naive bayes akan di cari data test dan data train dengan google colab



```
0d # It seems like the variable you meant to use was X_kbest_features
selected_x = X_kbest_features
selected_x

array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

+ Kode + Teks

```
0d [140] #print
print('Banyak data x_train :',len(x_train))
print('Banyak data x_test :',len(x_test))
print('Banyak data y_train :',len(y_train))
print('Banyak data y_test :',len(y_test))

Banyak data x_train : 800
Banyak data x_test : 200
Banyak data y_train : 800
Banyak data y_test : 200
```

Gambar 4.15 Data train dan Data test

4.6 Hasil evaluasi

Pada hasil evaluasi ini menggunakan confusion matrix dengan menggunakan data y_{test} pada hasil naive bayes.

	Prediksi positif	Prediksi negatif	total
Aktual positif	84(TP)	24(FN)	108
Aktual negatif	16(FP)	76(TN)	92
Total	100	100	200

Tabel 4.16 Hasil evaluasi

$$1. \text{Akurasi} = \frac{TP+TN}{TOTAL DATA} = \frac{84+76}{200} = \frac{160}{200} = 0,8 = 80\%$$

$$2. \text{Presisi} = \frac{TP}{TP+FP} = \frac{84}{84+16} = \frac{84}{100} = 0,84 = 84,0\%$$

$$3. \text{Recall} = \frac{TP}{TP+FN} = \frac{84}{84+24} = \frac{84}{108} = 0,7777777777777777 = 77.77777777779\%$$

```

⇒ prediksi benar: 160 data
   prediksi salah: 40 data
   Akurasi Algoritme: 80.0 %
   TRUE NEGATIVE (TN): 76
   FALSE NEGATIVE (FN): 24
   TRUE POSITIVE (TP): 84
   FALSE POSITIVE (FP): 16
   PRECISION: 84.0 %
   RECALL: 77.7777777777779 %

```

	precision	recall	f1-score	support
Negatif	0.76	0.83	0.79	92
Positif	0.84	0.78	0.81	108
accuracy			0.80	200
macro avg	0.80	0.80	0.80	200
weighted avg	0.80	0.80	0.80	200

Gambar 4.17 Hasil evaluasi

4.7 Hasil visualisasi

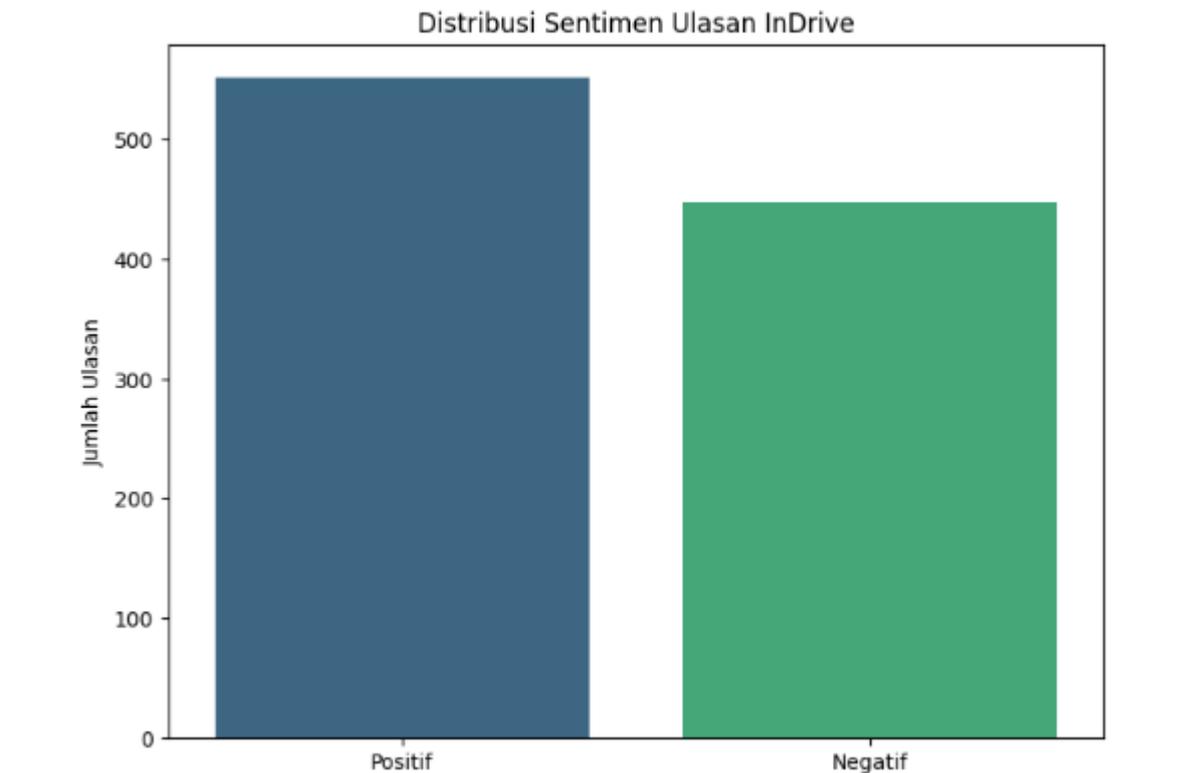
Tahapan terakhir yaitu adalah tahap visualisasi data menggunakan wordcloud untuk mengetahui kata yang sering muncul pada setiap sentimen. Berikut adalah hasil wordcloud pada sentimen positif yang dapat dilihat pada gambar dibawah ini:



Gambar 4.18 Sentimen positif

4.8 Hasil Pengujian

Pada hasil pengujiannya ini menghasilkan perbandingan model klasifikasi yang terbaik. Perbandingan tersebut ditampilkan dengan grafik pada gambar 11 dibawah ini:



Gambar 4.20 Grafik hasil pengujian

Dari gambar 4.20 di atas dapat dilihat jumlah ulasan yang digunakan adalah 1000 data ulasan. Dengan jumlah ulasan Positif sebanyak 550 ulasan dengan kalimat yang sering muncul pada ulasan positif yaitu: Aplikasi, Orderan, Mantap, Bagus, Membantu, dan lain lain. Sedangkan pada jumlah ulasan Negatif sebanyak 450 ulasan dengan kalimat yang sering muncul pada ulasan negatif yaitu: Aplikasi, Orderan, driver, tolong, penumpang, akun dan lain lain.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil penelitian diatas yakni mengenai analisis sentiment ulasan pengguna Aplikasi Indrive pada Google PlayStore yang berjumlah 1000 data dan sudah melewati tahap preprocess text yang terdiri dari transformation, tokenization, stopwords, stemming dan klasifikasi menggunakan algoritma Naïve Bayes serta evaluasi data menggunakan Confusion Matrix menggunakan Google Colaboratory dapat disimpulkan bahwa tanggapan / ulasan Aplikasi indrive tergolong **positif** lebih besar. untuk nilai akurasi (accuracy), 80% untuk nilai precision 84%, dan recall 77%. Nilai ini menunjukkan bahwa klasifikasi algoritma Naïve Bayes dinilai cukup baik dalam pemrosesan data ulasan pada google play store.

5.2 Saran

1. Untuk Peneliti selanjtnya bisa mencoba atau menggunakan algoritma yang lainnya dalam menganalisis sentiment seperti algoritma K-Nearest Neighbor, SVM, ataupun algoritma lainnya.
2. Dapat memperluas dataset yang lebih besar dan lebih beragam agar dapat membantu meningkatkan kinerja model analisis sentiment. Melibatkan data dari berbagai sumber, industri, atau Bahasa agar dapat memberikan wawasan yang lebih luas tentang sentimen.

DAFTAR PUSTAKA

- E. Wahyusetyawati, “Dilema pengaturan transportasi online,” *Jurnal RechtsVinding*. ISSN, pp. 2089–9009, 2017.
- F. V. Sari and A. Wibowo, “Analisis Sentimen Pelanggan Toko Onlin Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 10, no.2, pp. 681–686, 2019
- A. Z. Amrullah, A. S. Anas, and M. A. J. Hidayat, “Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square,” *Jurnal Bumigora Information Technology (BITe)*, vol. 2, no. 1, pp. 40–44, 2020.
- A. Afgiansyah, “Artificial Intelligence Neutrality: Framing Analysis of GPT Powered-Bing Chat and Google Bard,” *Jurnal Riset Komunikasi*, vol. 6, no.2, pp.179193,2023, Accessed: Feb.28,2024. [Online]. Available: <https://jurnalrisetkomunikasi.org/index.php/jrk/index>.
- A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, Accessed: Feb. 28, 2024. [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/index>
- U. Kusnia and F. Kurniawan, “Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (SVM) Dan Naive Bayes,” *Explore IT: Jurnal Keilmuan dan Aplikasi Teknik Informatika*, vol. 14, no. 1, pp. 24–28, 2022.
- M. Kholilullah, M. Martanto, and U. Hayati, “ANALISIS SENTIMEN PENGGUNA TWITTER (X) TENTANG PIALA DUNIA USIA 17 MENGGUNAKAN METODE NAIVE BAYES,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 392–398, 2024, Accessed: Feb. 28, 2024. [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/index>
- E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, “Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc,” *Faktor Exacta*, vol. 14, no. 2, pp. 64–71, 2021, Accessed: Feb.28,2024. [Online]. Available: <https://journal.lppmunindra.ac.id/index>.

- Z. Annisa and B. S. S. Ulama, “Analisis Sentimen Data Ulasan Pengguna Aplikasi ‘PeduliLindungi’ pada Google Play Store Menggunakan Metode Naïve Bayes Classifier Model Multinomial,” *Jurnal Sains dan Seni ITS*, vol. 11, no. 6, pp. D464–D471, 2023, Accessed: Apr. 02, 2024. [Online]. Available: https://ejournal.its.ac.id/index.php/sains_seni/index
- A. Putri et al., “Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir: Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 20–26, 2023, Accessed: Apr. 02, 2024. [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/index>
- D. Sandi, E. Utami, and K. Kusnawi, “Analisis Sentimen Publik Terhadap Elektabilitas Ganjar Pranowo di Tahun Politik 2024 di Twitter dengan Algoritma KNN dan Naïve Bayes,” *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1097–1108, 2023, Accessed: Feb. 28, 2024. [Online]. Available: <https://ejournal.stmik-budidarma.ac.id/index.php/mib/index>
- I. Verawati and B. S. Audit, “Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap ProviderBy. u,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, pp. 1411–1417, 2022, Accessed: Feb. 28, 2024. [Online]. Available: <http://www.ejournal.stmik-budidarma.ac.id/index.php/mib/index>
- Indrayanto, C.G., Ratnawati, D.E. and Rahayudi, B. (2023) ‘Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertaminadi Indonesia pada Google Play Store menggunakan Metode Random Forest’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(3), pp. 1131–1139. Available at: <http://j-ptiik.ub.ac.id>.