

SENTIMEN ANALISIS ULASAN PENGGUNA LINKEDIN PADA APLIKASI GOOGLE PLAYSTORE DENGAN METODE DECISION TREE

Aditya Putra Prastyo, Ajif Yunizar Pratama Yusuf

Program Studi Ilmu Komputer, Universitas Bhayangkara Jakarta Raya
Jl. Raya Perjuangan, Bekasi Utara, Jawa Barat,
ubjaditya@gmail.com, ajif.yunizar@dsn.ubharajaya.ac.id

Abstract - LinkedIn is a social networking platform that focuses on professionals and careers. Success in business often involves understanding and properly addressing user satisfaction, trust, and comfort. Identifying problems from customer reviews is key, although reading and classifying each review manually is considered inefficient and difficult. Therefore, this study focuses on sentiment analysis of LinkedIn app user reviews to address this challenge. The data used comes from user reviews and ratings on the Google Play Store, collected through scraping techniques. The Decision Tree algorithm is used in this research, with the application of the CRISP-DM methodology to extract information from text data. Classification results using Decision Tree show an accuracy level of 85%. This research contributes by showing that sentiment analysis techniques can identify deficiencies in applications, which can be used as a basis for evaluation for optimizing products and services.

Keywords - Sentimen Analysis, LinkedIn, Decision Tree, Data Mining.

Abstrak - . LinkedIn merupakan salah satu platform jaringan sosial yang berfokus pada profesional dan karier. keberhasilan dalam bisnis sering kali melibatkan pemahaman dan penanganan dengan baik terhadap kepuasan, kepercayaan dan kenyamanan pengguna. Identifikasi masalah dari ulasan pelanggan menjadi kunci, meskipun membaca dan mengklasifikasikan setiap ulasan secara manual dianggap tidak efisien dan sulit. Oleh karena itu, penelitian ini fokus pada analisis sentimen ulasan pengguna aplikasi LinkedIn untuk mengatasi tantangan ini. Data yang digunakan berasal dari ulasan dan rating pengguna di Google Play Store, dikumpulkan melalui teknik Scraping. Algoritma Decision Tree digunakan dalam penelitian ini, dengan penerapan metodologi CRISP-DM untuk menggali informasi dari data teks. Hasil klasifikasi menggunakan Decision Tree menunjukkan tingkat akurasi sebesar 85%. Penelitian ini memberikan kontribusi dengan menunjukkan bahwa teknik analisis sentimen dapat mengidentifikasi kekurangan dalam aplikasi, yang dapat dijadikan dasar evaluasi untuk optimalisasi produk dan layanan.

Kata Kunci - Analisis Sentimen, LinkedIn, Pohon Keputusan, Penambangan Data.

I. PENDAHULUAN

Era yang dipenuhi oleh teknologi yang canggih dan sangat berkembang memungkinkan kita dapat mencari teman, koneksi profesional bahkan mencari lowongan pekerjaan hanya melalui media sosial. ada banyak media sosial yang dapat digunakan untuk itu seperti LinkedIn (2003), Facebook (2004), Twitter (2006), dan Instagram (2010). Media sosial merupakan platform daring di mana pengguna dapat berpartisipasi, berbagi, dan menciptakan konten termasuk blog, jejaring sosial, wiki, forum, dan dunia maya dengan mudah. Blog, jejaring sosial, dan wiki adalah bentuk-bentuk media sosial yang paling sering digunakan oleh masyarakat di seluruh dunia [1]

Salah satu media sosial yang berfokus pada pencarian kerja, pemberian lowongan pekerjaan, koneksi profesional, ataupun mendapatkan informasi mengenai hal yang pengguna ingin ketahui adalah LinkedIn. Aplikasi ini memungkinkan pengguna untuk menampilkan resume dan sertifikat keahlian untuk menunjukkan skill, ketertarikan dan informasi pengguna yang digunakan untuk pencari lowongan pekerjaan

ataupun mencari informasi dengan ketertarikan pengguna.

LinkedIn didirikan oleh Reid Hoffman pada tahun 2002 dan situs resminya diluncurkan pada tanggal 5 Mei 2003. Misi utama LinkedIn sejak awal adalah menghubungkan para profesional di seluruh dunia untuk meningkatkan produktivitas mereka. Pada akhir bulan pertama operasi penuh, LinkedIn berhasil mendapatkan 4.500 anggota, namun pada hari yang buruk hanya mendaftar 20 anggota. Awalnya, LinkedIn mengalami masa sulit [2]

Rating linkedin dari tahun 2018 – hingga 2023 mendapatkan lebih banyak ulasan negatif daripada positif pada *google playstore*, banyaknya ulasan negatif yang berarti banyaknya keluhan pengguna terhadap linkedin, apabila permasalahan ini tidak di perbaiki akan membuat pengguna merasa tidak puas, tidak percaya dan tidak nyaman terhadap penggunaan aplikasi linkedin. Masalah ini akan berdampak pada jumlah pengguna yang akan menurun dikarenakan aplikasi yang memiliki permasalahan.

Salah satu cara analisis dapat diselesaikan menggunakan Machine Learning, Machine Learning

(ML) adalah bagian dari Kecerdasan Buatan. Ini adalah bidang kecerdasan komputasi yang melibatkan analisis dan interpretasi struktur dan pola yang ada dalam data dan mesin mampu melakukan pembelajaran, pengambilan keputusan, dan penalaran dengan pola-pola ini dan mampu menarik solusi cerdas tanpa campur tangan manusia [3]. Salah satu algoritma yang dapat digunakan dalam analisis ini di Machine Learning yaitu Decision Tree.

Algoritma Decision Tree adalah salah satu komponen dalam teknik klasifikasi pada data mining. Algoritma ini digunakan untuk mengatasi permasalahan dalam mengelompokkan data ke kelas-kelasnya masing-masing. Output dari decision tree adalah aturan-aturan yang didasarkan pada struktur pohon keputusan yang terbentuk. Proses dalam algoritma decision tree didasarkan pada nilai Information Gain untuk membentuk akar pohon keputusan dari atribut yang dipilih. Pemilihan atribut didasarkan pada cara objek dipecahkan terkait tujuan klasifikasi. Atribut yang memiliki Information Gain tertinggi akan menjadi atribut akar dari pohon keputusan[4].

Penelitian ini menggunakan Decision Tree, algoritma klasifikasi Decision Tree yang diimplementasikan pada dataset sentiment ulasan pada aplikasi LinkedIn. Pada penelitian sebelumnya yang dilakukan oleh Ichwanul Muslim Karo Karo, Justaman Arifin Karo Karo, Yumianto, Hariyanto, Miftahul Falah dan Manan Ginting dengan judul “Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan Support Vector Machine” penelitian tersebut menggunakan 2500 ulasan mendapatkan akurasi sebesar 79% dengan data split 75:25 [5]. Pada penelitian yang dilakukan oleh Rifki Kosasih dan Anggi Alberto dengan judul “Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor” penelitian ini menggunakan 1000 data ulasan produk shopee dan membagi menjadi 700 data latih dan 300 data uji yang dimana mendapatkan akurasi model 79,33% [6].

Dalam penelitian ini, data ulasan pengguna platform LinkedIn dikumpulkan menggunakan Teknik scrapping dari sumber ulasan aplikasi yang berada pada Google Playstore.selanjutnya penulis melakukan analisis dengan mengklasifikasikan ulasan pengguna menjadi 2 kategori yaitu Positif dan negative. Pada penelitian ini ingin mengetahui pengguna aplikasi merasa puas (sentiment positif) atau tidak puas(sentiment negatif) dengan layanan yang telah ditawarkan, kedua sentiment tersebut dapat membantu perusahaan untuk memahami umpan balik pengguna secara umum dan mengambil Tindakan yang sesuai untuk memperbaiki pengalaman pengguna

Dalam Kasus diatas Penulis mendapatkan data yang telah discrapping dengan rentang waktu September 2018 hingga September 2023 dan mendapatkan Studi kasus dengan Judul yaitu : Analisis Sentiment Ulasan Pengguna LinkedIn pada Google Play Store dengan Metode Decision Tree

A. Analisis Sentimen

Analisis sentimen adalah proses otomatis yang melibatkan pengekplorasi sikap, opini, dan emosi yang terkandung dalam teks, ucapan, dan database melalui Natural Language Processing, Opini atau sentimen yang dinyatakan dalam bentuk emosi dikategorikan ke dalam kelas positif, negatif, atau netral.

Tugas utama dalam analisis sentimen mencakup klasifikasi polaritas pada berbagai tingkatan, termasuk tingkat dokumen, tingkat kalimat, dan tingkat aspek atau entitas. Pada tingkat dokumen, keseluruhan dokumen diklasifikasikan sebagai positif atau negatif. Klasifikasi sentimen pada tingkat kalimat membagi kalimat ke dalam kelas positif, negatif, atau netral [7]. Sentimen Analisis memiliki akurasi yang baik dan cepat dibanding Analisis manual, Pendekatan klasifikasi sentiment dapat diklasifikasikan menjadi 3 yaitu :

1. Machine Learning

Pendekatan Machine Learning digunakan untuk memprediksi polaritas sentiment berdasarkan kumpulan data yang dilatih dan diuji.

2. Lexicon based

Pendekatan berbasis Lexicon tidak memerlukan pelatihan sebelumnya untuk menambang data. Ini menggunakan daftar kata yang telah ditentukan sebelumnya, dimana pada tiap kata di katikan dengan sentiment tertentu

3. Hybrid

Hybrid kombinasi dari Machine Learning dan Lexicon based yang mempunyai potensi untuk meningkatkan kinerja klasifikasi sentimen

B. Data Mining

Penambangan data atau Data mining adalah suatu proses otomatis dalam menemukan informasi yang bernilai dari kumpulan data besar. Teknik ini digunakan untuk menyelidiki data yang luas dengan tujuan menemukan pola yang bermanfaat, bahkan yang mungkin belum diketahui sebelumnya. Selain itu, teknik penambangan data memberikan kemampuan untuk melakukan prediksi terhadap hasil pengamatan di masa mendatang [8].

Tujuan utama dari data mining adalah untuk mengidentifikasi hubungan yang tidak terlihat atau pola yang tidak terduga dalam data yang besar dan kompleks. Proses ini melibatkan penggunaan berbagai teknik matematika, statistika, dan kecerdasan buatan untuk menganalisis dan menggali informasi yang bermanfaat dari data

C. Cross-Industry Standard for Data Mining (CRISP-DM)

CRISP-DM (Cross Industry Standard Process for Data Mining) yang dikembangkan tahun 1996 oleh analisis dari beberapa industri seperti standarisasi Daimler Chrysler (Daimler-Benz), SPSS, NCR.

CRISP-DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian [9].

Struktur sistematisnya memberikan pendekatan yang terorganisir mulai dari pemahaman tujuan bisnis hingga implementasi model, memastikan tidak ada tahap yang terlewat. Metodologi ini bersifat fleksibel, memungkinkan adaptasi sesuai kebutuhan proyek tertentu, dan mendukung proses iteratif untuk penyempurnaan dan penyesuaian yang diperlukan.

D. Decision Tree

Algoritma Decision Tree adalah salah satu komponen dalam teknik klasifikasi pada data mining. Algoritma ini digunakan untuk mengatasi permasalahan dalam mengelompokkan data ke kelas-kelasnya masing-masing. Output dari decision tree adalah aturan-aturan yang didasarkan pada struktur pohon keputusan yang terbentuk. Proses dalam algoritma decision tree didasarkan pada nilai Information Gain untuk membentuk akar pohon keputusan dari atribut yang dipilih. Pemilihan atribut didasarkan pada cara objek dipecahkan terkait tujuan klasifikasi. Atribut yang memiliki Information Gain tertinggi akan menjadi atribut akar dari pohon keputusan. Rumus dan langkah-langkah yang digunakan dapat dilihat seperti berikut [4] :

$$Entropy(S) = -\sum_{i=1}^n p_i \times \log_2 p_i \quad (1)$$

Keterangan :

S : Himpunan Kasus

n : Jumlah Partisi S

p_i : Jumlah kasus pada partisi – I

Untuk Mendapatkan nilai gain :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \quad (2)$$

E. Evaluasi

Confusion matrix adalah sebuah tabel yang menggambarkan performa model atau algoritma dalam mengklasifikasikan data dengan rincian lebih lanjut. Setiap baris dari matrix tersebut merepresentasikan kelas aktual dari data, sementara setiap kolom merepresentasikan kelas prediksi dari data (atau sebaliknya). Dalam matrix ini, terdapat empat kemungkinan hasil: true positive (TP), true negative (TN), false positive (FP), dan false negative (FN), yang menggambarkan berbagai kombinasi prediksi dan aktual dari data. Matrix ini memberikan gambaran yang jelas tentang performa model dalam hal klasifikasi yang benar dan yang salah [10]

Selain Confusion Matrix Peneliti Menggunakan Akurasi, Precision, Recall dan juga f1-score.

Akurasi adalah perbandingan antara prediksi yang benar dengan keseluruhan prediksi yang dilakukan. Presisi mengukur seberapa tepat model dalam memprediksi kasus positif dari semua prediksi positif. Recall mengukur seberapa baik model dalam mendeteksi semua kasus positif yang sebenarnya. F1 Score merupakan perpaduan antara presisi dan recall, memberikan gambaran komprehensif tentang performa model dalam memprediksi kasus positif [11].

Setelah hasil nilai confusion matrix didapatkan maka performa matrix dapat diukur dengan Akurasi, Precision, Recall dan f1 score dengan rumus [12]:

1. Akurasi

Accuracy adalah ukuran berapa banyak prediksi benar yang dibuat model Anda untuk kumpulan data pengujian yang lengkap. Akurasi adalah metrik dasar yang baik untuk mengukur performa model. Dalam kumpulan data yang tidak seimbang akurasi menjadi metrik yang buruk. Berikut rumus akurasi :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2. Precision

Precision digunakan sebagai metrik kinerja ketika tujuannya adalah membatasi jumlah positif palsu. Berikut rumus Precision :

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3. Recall

Recall digunakan sebagai metrik kinerja ketika kita perlu mengidentifikasi semua sampel positif yaitu saat penting untuk menghindari negatif palsu. Berikut rumus Recall :

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

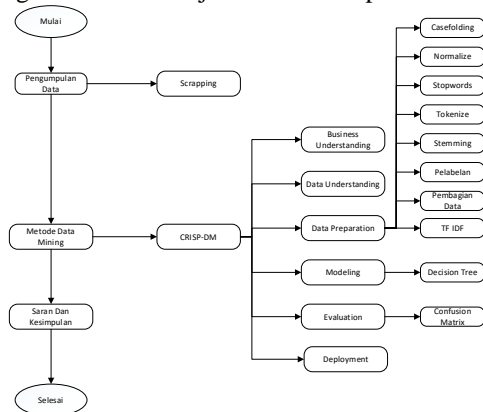
4. F1-Score

Meskipun Precision dan Recall adalah ukuran yang sangat penting, pertimbangkan hanya salah satu saja mereka tidak akan memberi Anda gambaran lengkap. Salah satu cara untuk meringkasnya adalah f-score atau f-measure, yaitu dengan mean harmonik presisi dan recall. Berikut rumus F1-Score :

$$F1-Score = \frac{2 (Precision \times Recall)}{Precision + Recall} \quad (6)$$

II. METODE PENELITIAN

Pada penelitian ini menggunakan proses proses pada gambar 1 akan dijelaskan dalam penelitian ini



Gambar 1 Proses Penelitian

1. Pengumpulan Data

Pada Tahap ini peneliti melakukan scrapping ulasan LinkedIn pada Google Playstore sebanyak 2000 data dengan rentang waktu September 2018 – September 2023.

2. CRISP DM

a. Bussiness Understanding

Pada tahap ini peneliti mencari arah dan tujuan serta startegi awal dalam penelitian ini. Akan ditentukan masalah bisnis dalam kasus penelitian ini seperti LinkedIn merupakan aplikasi jaringan profesional dan pencarian kerja yang dimana aplikasi ini adalah salah satu aplikasi terbesar di dunia tetapi 5 tahun belakangan ini linkedin mendapatkan lebih banyak komentar atau opini negatif daripada positif pada *Google Playstore*, peneliti menganalisis apa yang menyebabkan linkedin mendapatkan banyak ulasan negatif.

Peneliti menggunakan Teknik Analisis Sentiment dengan metode Decision Tree dan untuk mengetahui penyebab kenapa penggunaan LinkedIn di Indonesia memberikan banyaknya ulasan negatif pada aplikasi kemudian peneliti juga ingin menguak seberapa akurat Algoritma Decision tree dengan jumlah 2000 data ulasan LinkedIn

b. Data Understanding

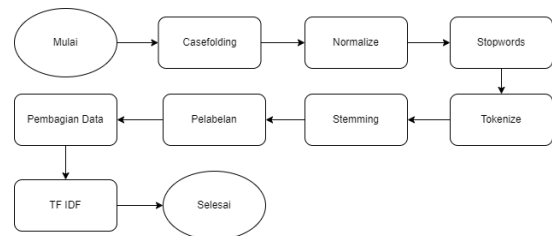
Pada tahap data understanding ini memiliki tujuan untuk menganalisis data yang telah di kumpulkan. Pada penelitian ini, peneliti mendapatkan data dari *Google Playstore* dengan dataset yang diambil ialah Ulasan pada aplikasi LinkedIn. Dengan total jumlah data 2000 ulasan dan 11 Atribut yang akan di bersihkan pada tahap selanjutnya.

Pada Proses ini dari 11 atribut penelitian ini hanya menggunakan 2 atribut saja yaitu Content yang berisikan ulasan pengguna terhadap aplikasi

linkedin *google playstore* dan *Score* yang berisikan rating yang diberikan oleh pengguna terhadap aplikasi linkedin pada google playstore sebagai opini. Tidak lupa pada proses ini data dicek apakah memiliki data kosong atau null.

c. Data Preparation

Pada tahap ini data akan dibersihkan yang bertujuan untuk mendapatkan data yang baik dan bersih sehingga akan menghasilkan hasil yang baik. Pada tahap ini peneliti menggunakan enam teknik Text Mining yaitu :



Gambar 2 Proses Data Preparation

Berikut adalah tabel dari penjelasan dari gambar 2 :

Proses	Penjelasan
Casefolding	Casefolding adalah proses pengubahan semua karakter dalam teks menjadi huruf kecil (<i>lowercase</i>) dan Menghapus tanda baca
Normalize	Pada tahap Normalisasi, proses dimana perbaikan kata kata typo , kata yang tidak memiliki spasi diperbaiki dan juga kata singkat
Stopwords	Pada tahap ini, kata-kata yang dianggap tidak perlu akan dihapus, seperti contohnya "namun", "tetapi", dan kata-kata yang tidak memiliki makna tersendiri saat dipisahkan dari kata lain, serta yang tidak terkait dengan kata sifat.

Tokenize	Tokenization adalah proses pemecahan kata menjadi unit lebih kecil yang disebut token
Stemming	Stemming adalah proses pemotongan atau penghapusan awalan atau akhiran kata (affixes) dari kata-kata dalam teks untuk menghasilkan bentuk dasar atau bentuk akar (root form) yang disebut "stem."
Pelabelan	Proses pelabelan adalah memberikan label pada data berdasarkan rating yang diterima, baik itu label positif atau negative
Pembagian Data	Pada proses ini data akan dibagi menjadi dua bagian yaitu data train atau data latih, data latih digunakan untuk melatih model dan data test atau data uji, data uji digunakan untuk menguji ketepatan model terhadap data.
TF IDF	Pada proses akhir pada Data Preparation ini kata diberikan bobot pada kata-kata dalam suatu dokumen berdasarkan seberapa sering kata tersebut muncul di dokumen tersebut (TF) dan seberapa penting kata tersebut secara umum (IDF).

Pada fase pemodelan dalam penelitian ini akan melibatkan Algoritma Decision Tree. Algoritma ini dipilih oleh peneliti dikarenakan pada penelitian terdahulu Algoritma ini dapat dengan baik memiliki akurasi yang baik.

e. Evaluasi
 Pada tahap ini, penggunaan data uji (testing data) digunakan untuk menguji kinerja model, dan Confusion Matrix digunakan sebagai alat evaluasi untuk proses data mining. Apabila model telah sesuai dengan tujuan yang telah dijelaskan pada tahap awal Business Understanding.

f. Deployment
 Pada tahap ini peneliti akan mempresentasikan hasil dari Penelitiannya pada Audience Dosen Penguji, Deployment tidak sepenuhnya dapat di implemetasi dikarenakan pada tahap ini peneliti hanya menggunakannya untuk Studi.

III. HASIL DAN PEMBAHASAN

Pada proses sentiment analisis ini menggunakan Bahasa pemrograman Python dengan ide DataSpell dan bantuan library Python seperti : *google-play-scraper*, *Pandas*, *Matplotlib*, *Numpy*, *Sklearn*, *Sastrawi* dan *NLTK*.

A. Data Preparation

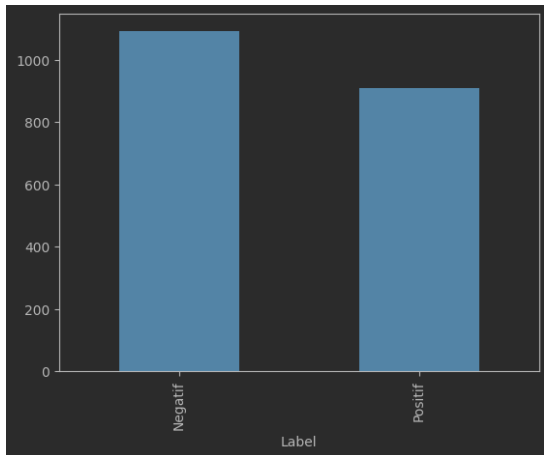
Jumlah data yang di scrapping yaitu sebanyak 2000 data yang diambil pada ulasan *google playstore* setelah itu data dipelajari dan dipahami oleh penulis setelah itu data di proses pada metode *Data Preparation*. Contoh dari penerapan *Data Preparation* ditunjukkan pada tabel 2.

Proses	Hasil
Data Asli	Bagus karena membantu mencari kerja yg sesuai harapan!!!
Casefolding	bagus karena membantu mencari kerja yg sesuai harapan
Normalize	bagus karena membantu mencari kerja yang sesuai harapan
Stopwords	bagus membantu mencari kerja sesuai harapan
Tokenize	['bagus', 'membantu', 'mencari', 'kerja', 'sesuai', 'harapan']
Stemming	bagus bantu cari kerja harap

Setelah data melalui proses pada tabel 2, selanjutnya data akan diberikan label. Pada proses pelabelan data akan diberikan label yang berupa Positif atau Negatif. Peneliti memberikan label Positif dengan rentang rating 4 – 5 pada rentang tersebut sangat banyak dan dominannya ulasan positif seperti pujian dan kepuasan pengguna, lalu pada rentang rating 1 – 3 diberikan label Negatif dikarenakan pada retang ini sangat banyak dan

d. Modelling

dominan ulasan Negatif seperti keluhan, amarah dan juga kritikan yang diberikan pengguna terhadap aplikasi linkedin pada google playstore.



Gambar 3 Jumlah total label pada data

Pada gambar 3 dapat dijelaskan bahwa ulasan negatif didapatkan sebanyak 1092 data dan positif 908. Pada pelabelan ini data juga secara manual pada rating 2 – 4 di cek oleh peneliti yang dimana data 4 yang mengandung ulasan negatif akan dipindahkan ke label 3 sedangkan data pada rating 2 dan 3 yang mengandung ulasan positif akan dipindahkan pada label 4. Lalu berikut pada tabel 3 adalah contoh data dengan label positif dan negatif.

Data	Label
Cari pekerjaan lebih mudah, & banyak relasi yg bisa diikuti supaya tambah wawasan	Positif
Verifikasi gagal terus tidak bisa masuk	Negatif

Setelah data diberikan label tahap selanjutnya adalah tahap TF IDF. Pada proses TF IDF ini Tiap kata akan di berikan bobot berdasarkan kemunculan kata kata tersebut, TF yang berarti seberapa sering kemunculan kata pada sebuah dokumen, IDF yang berarti Seberapa sering kemunculan kata pada Seluruh dokumen dan lalu TF IDF adalah gabungan antara seberapa sering kemunculan kata pada sebuah dokumen dan seberapa sering kemunculan kata pada seluruh dokumen Berikut adalah contoh pembobotan TF IDF

Tabel 1 TF IDF

Token	TF IDF
Bagus	0,106
Bantu	0,13
Cari	0,08
Kerja	0,08

Harapan	0,174
Verifikasi	0,477
Gagal	0,544

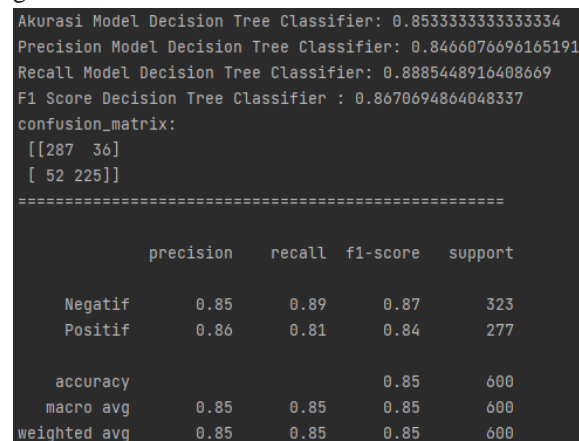
Setelah data dihitung nilai TF IDF maka selanjutnya data dapat dibagi menjadi 70 : 30 yang dimana pada pembagian ini 70% data akan menjadi data training dan 30% data akan menjadi data testing. Tujuan dari pembagian data yaitu untuk menghasilkan model yang dapat menggeneralisasi dengan baik pada data yang tidak pernah dilihat sebelumnya. Dan Pembagian 70:30 yang memberikan lebih banyak data untuk pelatihan membantu mencegah overfitting, yaitu ketika model terlalu sesuai dengan data pelatihan dan kehilangan kemampuan untuk menggeneralisasi pada data baru. Semakin banyak data yang digunakan untuk pelatihan, semakin baik model dapat belajar pola umum.

B. Modelling

Data yang telah melewati tahapan Data Preparation akan dibagi menjadi dua yaitu data training dan data testing. Data training bertujuan untuk mengajarkan model agar dapat membuat generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya. Model "melihat" data training selama proses pelatihan untuk menyesuaikan parameter dan mencari pola, sedangkan data testing bertujuan untuk menguji kinerja model pada data yang belum pernah dilihat sebelumnya. Model yang telah dilatih dengan data training diuji pada data test untuk mengukur sejauh mana model dapat membuat prediksi atau klasifikasi yang akurat dan umum. Pembentukan model algoritma Decision Tree dari data training menggunakan persamaan (1) dan 2.

C. Evaluasi

Pada Tahap ini data yang telah dimodelkan dengan Naïve Bayes pada tahap sebelumnya akan dievaluasi atau diukur kinerja. Kinerja Data dapat dilihat pada gambar 4



Gambar 4 Hasil Evaluasi Model

Pada gambar 4 evaluasi model mendapatkan nilai akurasi sebesar 85% yang dimana didapatkan dari hasil perhitungan dengan menggunakan persamaan (3), nilai precision sebesar 85% yang didapatkan dari hasil perhitungan dengan menggunakan persamaan (4), nilai recall didapatkan 89% yang didapatkan dari hasil persamaan (5) dan nilai F1-Score didapatkan 87% yang didapatkan dari hasil persamaan (6)

D. Visualisasi

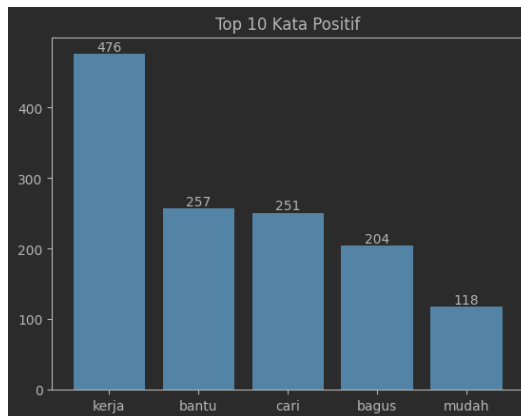
Visualisasi bertujuan untuk mengekstraksi informasi berupa topik topik yang sering dibicarakan atau diulas oleh pengguna LinkedIn, sehingga dari sekian banyak teks ulasan yang dapat diambil informasi penting yang terkandung. Dalam penelitian ini visualisasi hasil akan menggunakan Wordcloud.

Wordcloud adalah representasi dari suatu data yang menampilkan kumpulan kumpulan kata kata penting dan sering muncul dari data tersebut. Semakin besar kata pada wordcloud maka semakin sering kata itu muncul dalam data. Berikut pada gambar a dan b adalah bentuk visualisasi menggunakan wordcloud



Gambar 5 Wordcloud Kata Positif

Pada gambar 5 merupakan kata ulasan positif yang paling banyak diulas oleh pengguna linkedin pada Google playstore, dari kata kata diatas yang paling besar adalah kata kata yang memiliki ukuran yang besar. Berikut kata kata wordcloud yang sering muncul pada ulasan adalah 'Kerja', 'Bantu', 'Cari', 'Bagus' dan 'Mudah'. Dari kata yang paling banyak muncul tersebut dapat menunjukkan pengguna linkedin yang memberikan sentiment positif merasakan bahwa aplikasi tersebut dapat membantu mencari pekerjaan, aplikasi yang mudah digunakan dan juga aplikasi yang bagus dalam penggunaannya. Berikut adalah frekuensi ulasan positif kemunculan kata kata paling sering muncul pada wordcloud :



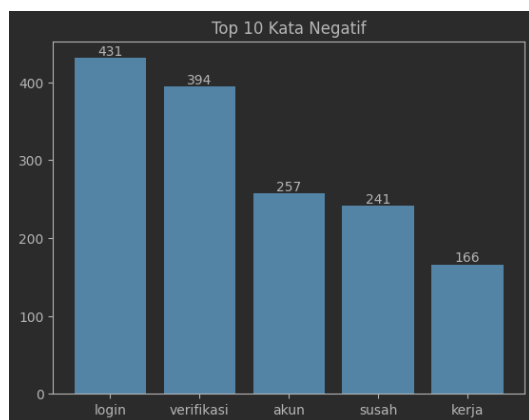
Gambar 6 Frekuensi Kata Positif

Pada gambar 6 adalah kemunculan kata kata yang paling sering muncul pada sentiment positif yang dimana kata 'Kerja' muncul sebanyak 476, kata 'Bantu' muncul sebanyak 257, kata 'Cari' muncul sebanyak 251, kata bagus muncul sebanyak '204', dan kata 'Mudah' muncul sebanyak 118.



Gambar 7 Wordcloud Kata Negatif

Pada gambar 4.28 merupakan kata ulasan negatif yang paling banyak diulas oleh pengguna linkedin pada Google Playstore, dari kata kata diatas didapatkan kata kata yang sering muncul yang memiliki ukuran besar pada wordcloud. Berikut adalah kata kata pada wordcloud yang sering muncul pada ulasan adalah 'Login', 'Verifikasi', 'Akun', 'Susah', dan 'Kerja'. Dari kata yang paling sering muncul dapat menunjukkan pengguna linkedin yang memberikan sentiment negatif merasakan bahwa adanya beberapa masalah terhadap aplikasi seperti pada login yang sulit bahkan ada yang tidak dapat login, lalu sulitnya verifikasi seperti adanya captcha atau mencocokkan gambar yang berulang ulang tetapi tetap gagal dalam memverifikasi akun, ada beberapa masalah pada akun pengguna seperti akun pengguna yang terkena suspend, ter banned, terhack dan juga para pengguna yang belum mendapatkan pekerjaan dari aplikasi linkedin. Berikut adalah frekuensi ulasan negatif kemunculan kata kata paling sering muncul pada wordcloud :



Gambar 8 Frekuensi Kata Negatif

Pada gambar 4.29 adalah kemunculan kata kata yang paling sering muncul pada sentiment negatif yang dimana kata 'Login' muncul sebanyak 431, kata 'Verifikasi' muncul sebanyak 394, kata 'Akun' muncul sebanyak 257, kata 'Susah' muncul sebanyak '241', dan kata 'Kerja' muncul sebanyak 166.

IV. KESIMPULAN

Hasil dari penelitian yang telah dilakukan dapat disimpulkan bahwa aplikasi linkedin memiliki banyaknya ulasan negatif yang dikarenakan oleh beberapa faktor yaitu : sulitnya login pada aplikasi, verifikasi yang sulit pada akun, beberapa pengguna belum mendapatkan kerja dan situs lowongan pekerjaan yang tidak di perbaharui oleh linkedin. Sentimen negatif mendapatkan 1092 data atau 54% ulasan negatif dari 2000 ulasan. Lalu hasil model Decision Tree mendapatkan nilai yang baik dalam menganalisis sentiment, dengan ini model dapat diterapkan pada jenjang selanjutnya. Model ini memperoleh akurasi sebesar 85%, Precision sebesar 85%, Recall sebesar 89% dan F1-Score sebesar 87%

DAFTAR PUSTAKA

- [1] Anang Sugeng Cahyoni, "PENGARUH MEDIA SOSIAL TERHADAP PERUBAHAN SOSIAL MASYARAKAT DI INDONESIA," *Publiciana*, vol. 9, pp. 140–157.
- [2] ZEN, "Inilah Sejarah LinkedIn, Platform Media Sosial Yang Mungkin Telah Berperan Besar Dalam Sukses Anda," *Bintang Bisnis*.
- [3] Deepti. Chopra and Roopal. Khurana, *Introduction to Machine Learning with Python*. Bentham Science Publishers, 2023.
- [4] I. Arfyanti, M. Fahmi, and P. Adytia, "Penerapan Algoritma Decision Tree Untuk Penentuan Pola Penerima Beasiswa KIP Kuliah," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2275.
- [5] I. Muslim Karo Karo *et al.*, "Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan Support Vector Machine," *Journal of Information System Research*, vol. 4, no. 4, pp. 1423–1430, 2023, doi: 10.47065/josh.v4i4.3943.
- [6] R. Kosasih and A. Alberto, "Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor," vol. 6, no. 1, 2021, doi: 10.30743/infotekjar.v6i1.3893.
- [7] Brij B. Gupta, Dragan Peraković, Ahmed A. Abd El-Latif, and Deepak Gupta, "Data Mining Approaches for Big Data and Sentiment Analysis in Social Media."
- [8] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to data mining*.
- [9] E. Damayanti and S. Kuswayati, "ANALISIS DAN IMPLEMENTASI FRAMEWORK CRISP-DM (CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING) UNTUK CLUSTERING PERGURUAN TINGGI SWASTA."
- [10] I. Widhi Saputro and B. Wulan Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction," *Citec Journal*, vol. 6, no. 1, 2019.

- [11] K. L. Kohsasih, Z. Situmorang, and I. Artikel, "Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular," *JURNAL INFORMATIKA*, vol. 9, no. 1, pp. 13–17, 2022, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [12] A. C. Müller and S. Guido, "Introduction to Machine Learning with Python A GUIDE FOR DATA SCIENTISTS Introduction to Machine Learning with Python."