

**SKRIPSI**

**ANALISIS SENTIMEN PENGGUNA APLIKASI *MOBILE*  
*BANKING* BANK SULSELBAR PADA *GOOGLE PLAY STORE*  
DENGAN METODE *DECISION TREE* BERBASIS SMOTE**

*Sentiment Analysis Of Mobile Banking Application Users Of  
Bank Sulselbar On Google Play Store Using Decision Tree  
Method Based On Smote*



**Disusun oleh:**

**MAMNUN AULIA S**

**D0221342**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS SULAWESI BARAT**

**MAJENE**

**2025**

# LEMBAR PERSETUJUAN

## SKRIPSI

### ANALISIS SENTIMEN PENGGUNA APLIKASI *MOBILE BANKING* BANK SULSELBAR PADA *GOOGLE PLAY STORE* DENGAN METODE *DECISION TREE* BERBASIS SMOTE

Telah dipersiapkan dan disusun oleh

**MAMNUN AULIA S**

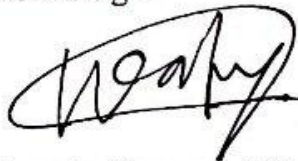
**D0221342**

Telah dipertahankan di depan Tim Penguji

Pada tanggal 30 Oktober 2025

Susunan Tim Penguji

Pembimbing 1



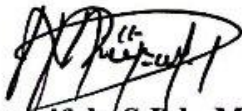
Heliawaty Hamrul, S.Kom., M.Kom  
NIP. 198710152019032008

Penguji 1



Farid Wajidi, S.Kom., M.T  
NIP. 198904182019031018

Pembimbing II



Musyriifah, S.Pd., M.Pd  
NIDN: 00144119302

Penguji II



Siti Aulia Rachmini, S.T., M.T  
NIP. 198207062008042003

Penguji III



Nurani Natsir, S.Si., M.Si  
NIDN. 9990549367

## LEMBAR PENGESAHAN

### ANALISIS SENTIMEN PENGGUNA APLIKASI *MOBILE BANKING* BANK SULSELBAR PADA *GOOGLE PLAY STORE* DENGAN METODE *DECISION TREE* BERBASIS SMOTE

#### SKRIPSI

Untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer

Disusun oleh:

**MAMNUN AULIA S**  
**D0221342**

Skripsi ini telah diuji dan dinyatakan lulus  
pada Tanggal 30 Oktober 2025  
Telah diperiksa dan disetujui oleh:

Pembimbing I



Heliawaty Hamrul, S.Kom., M.Kom  
NIP. 198710152019032008

Dekan Fakultas Teknik,  
Universitas Sulawesi Barat



Prof. Dr. Ir. Hafsah Nirwana, M.T  
NIP. 1964040519900322002

Pembimbing II



Musyrifah, S.Pd., M.Pd  
NIDN. 00144119302

Ketua Program Studi  
Informatika,



Muh. Ram Rasyid, S.Kom., M.T  
NIP. 198808182022031006

## PERNYATAAN OROSINALITAS

Saya menyatakan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis di sitasi dalam naskah ini dan disebutkan dalam daftar referensi.

Apabila ternyata di dalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Majene, 16 Oktober 2025



Mamnun Aulia S  
NIM: D0221342



## ABSTRAK

**Mamnun Aulia S.** Analisis Sentimen Pengguna Aplikasi *Mobile Banking* Bank Sulselbar Pada *Google Play Store* Dengan Metode *Decision Tree* Berbasis *SMOTE* (dibimbing oleh **Heliawaty Hamrul** dan **Musyrifah**).

Perkembangan layanan *mobile banking* menghadirkan tantangan baru dalam memahami kepuasan dan ketidakpuasan pengguna melalui ulasan di platform digital. Penelitian ini bertujuan untuk menganalisis sentimen pengguna aplikasi *mobile banking* Bank Sulselbar dengan menggunakan metode klasifikasi *Decision Tree* berbasis *SMOTE*. Ulasan yang diperoleh dari *google play store* dengan jumlah keseluruhan setelah dilakukan pembersihan nilai *NaN* pada pelabelan berjumlah 1.626 data. Hasil evaluasi model *Decision Tree*, pada rasio 80:20 sebelum penerapan metode *SMOTE* diperoleh akurasi sebesar 81%, namun nilai *f1-score* keseluruhan kelas menunjukkan ketidakseimbangan, di mana model cenderung lebih akurat dalam mengenali kelas mayoritas. Setelah penerapan *SMOTE*, akurasi menurun menjadi 74%, tetapi nilai *f1-score* meningkat berada pada rentang 40% hingga 85%, yang menunjukkan peningkatan keseimbangan kinerja model dalam mengenali setiap kelas sentimen. Sedangkan pada *10-fold cross validation*, akurasi yang diperoleh sebelum *SMOTE* mencapai 82%, meskipun nilai *f1-score* keseluruhan masih belum seimbang. Setelah penerapan *SMOTE* akurasi menurun menjadi 77% dengan peningkatan *f1-score* pada rentang 43% hingga 86%, yang mengindikasikan hasil yang lebih stabil dan seimbang antar kelas. Dengan demikian, penerapan *SMOTE* pada rasio 80:20 dan *10-fold cross validation* terbukti mampu meningkatkan keseimbangan antar kelas serta memperbaiki kemampuan model dalam mengenali kelas minoritas. Selain itu, evaluasi menggunakan *10-fold cross validation* menunjukkan kinerja model yang lebih stabil dan representatif terhadap keseluruhan data.

**Kata Kunci:** Analisis Sentimen, *Mobile Banking*, Bank Sulselbar, , *Decision Tree*, *SMOTE*.

## ***ABSTRACT***

***Mamnun Aulia S.*** *Sentiment Analysis Of Mobile Banking Application Users Of Bank Sulselbar On Google Play Store Using Decision Tree Method Based On SMOTE.* (supervised by **Heliawaty Hamrul** and **Musyrifah**).

*The development of mobile banking services presents new challenges in understanding user satisfaction and dissatisfaction through reviews on digital platforms. This study aims to analyze user sentiment toward the Bank Sulselbar mobile banking application using a Decision Tree classification method based on SMOTE. Reviews obtained from the Google Play Store with a total number of 1,626 data points after cleaning NaN values in the labeling. Evaluation results of the Decision Tree model, with an 80:20 ratio before applying the SMOTE method, the accuracy reached 81%. However, the overall F1-score across classes indicated an imbalance, showing that the model was more accurate in recognizing the majority class. After applying SMOTE, accuracy decreased to 74%, but the F1-score improved, ranging from 40% to 85%, indicating better balance in the model's performance across sentiment classes. In the 10-fold cross-validation evaluation, the model achieved an accuracy of 82% before applying SMOTE, although the overall F1-score remained unbalanced. After applying SMOTE, accuracy decreased to 77%, but the F1-score increased, ranging from 43% to 86%, suggesting more stable and balanced results across classes. Thus, applying SMOTE in both the 80:20 ratio and 10-fold cross-validation proved effective in improving class balance and enhancing the model's ability to recognize minority classes. Furthermore, the 10-fold cross-validation evaluation demonstrated a more stable and representative model performance across the entire dataset.*

***Keywords:*** *Sentiment Analysis, Mobile Banking, Bank Sulselbar, , Decision Tree, SMOTE*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi yang pesat telah membawa perubahan yang signifikan dalam berbagai aspek kehidupan, termasuk di sektor perbankan yang kini lebih cenderung ke arah digitalisasi. Karena teknologi yang memberikan kemudahan dan efisiensi, sehingga industri perbankan kemudian berinovasi untuk memberikan layanan secara digital kepada masyarakat (Benito et al., 2022). Menurut Peraturan Otoritas Jasa Keuangan (PJOK) No. 12/PJOK.03/2018 tentang penyelenggaraan layanan perbankan digital oleh bank umum, perbankan digital atau digital *banking* didefinisikan sebagai layanan perbankan elektronik yang dikembangkan dengan mengoptimalkan penggunaan data nasabah (Sujadada et al., 2023). Salah satu layanan perbankan digital yang populer adalah *mobile banking*.

*Mobile banking* atau *m-banking* adalah layanan digital dengan memanfaatkan teknologi informasi untuk memudahkan nasabah dalam melakukan berbagai transaksi keuangan. Layanan ini dapat diakses melalui perangkat mobile seperti ponsel, sehingga mempermudah aktivitas keuangan kapan saja dan di mana saja. *M-banking* hadir sebagai solusi atas kebutuhan masyarakat yang semakin sering melakukan transaksi secara berkala, serta untuk mendukung kelancaran kegiatan bisnis agar lebih praktis, cepat, dan efisien (Sari et al., 2021). Riset yang dikemukakan Juniper (2021) memperkirakan pengguna digital *banking* global akan naik 53% dalam lima tahun, dari 2,5 miliar pada tahun 2021 menjadi lebih dari 4,2 miliar pada tahun 2026, atau 53% populasi dunia. Asia tenggara dengan lebih dari 600 juta nasabah diproyeksikan menjadi salah satu pusat ekonomi terbesar dunia (Nurhidayah & Puspawati, 2024). Hal ini menandakan bahwa layanan *mobile banking* telah menjadi tren, tetapi juga kebutuhan dalam dunia ekonomi digital. PT. Bank Sulselbar adalah badan usaha milik pemerintah provinsi Sulawesi selatan serta pemerintah provinsi Sulawesi barat. Sebagai lembaga keuangan perbankan, Bank Sulselbar merupakan salah satu bank milik pemerintah yang berhasil menjalankan kegiatan bisnisnya (Arifuddin, 2023).

Sejalan dengan tren digitalisasi perbankan, Bank Sulselbar telah mengembangkan layanan sulselbar mobile sebuah aplikasi *mobile banking*. Salah satu layanan yang disediakan oleh *google* adalah *play store*, yang menawarkan berbagai konten digital dari berbagai kategori, termasuk game, aplikasi, film, musik, dan buku elektronik. Salah satu fitur yang ada di *play store* adalah sistem ulasan atau komentar. (Insan et al., 2023) komentar berfokus pada pendapat di mana pengguna dapat menyampaikan pendapat mereka mengenai aplikasi yang telah mereka gunakan. Dengan terus meningkatnya jumlah pengguna aplikasi *mobile*, sangat penting untuk memahami tingkat kepuasan dan ketidakpuasan mereka saat menggunakan aplikasi tersebut (Nurian, 2023). Misalnya, pada aplikasi bank sulselbar, ulasan pengguna biasanya berisi saran positif juga keluhan yang sifatnya negatif, hal ini tentunya akan mempengaruhi ekspektasi dari calon pengguna (Arminda et al., 2023). Ulasan yang tersedia dalam jumlah besar ini bersifat tidak terstruktur, sehingga diperlukan teknik analisis untuk memahami sentimen pengguna secara otomatis (Hendriyanto et al., 2022).

Salah satu pendekatan yang dapat digunakan adalah analisis sentimen, yaitu proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis guna mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini (Arsi & waluyo,. 2021). Analisis sentimen atau bisa disebut juga *opinion mining*, adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya (Sujadi, 2022).

Metode klasifikasi yang digunakan dalam penelitian ini adalah *Decision Tree* karena merupakan salah satu algoritma yang populer dan sering digunakan oleh peneliti. Algoritma ini membentuk model prediksi berbentuk pohon atau struktur hierarki yang mudah dipahami. *Decision Tree* mengubah data menjadi serangkaian keputusan yang disusun secara sistematis. Keunggulan utamanya adalah menyederhanakan proses pengambilan keputusan yang kompleks, sehingga memudahkan dalam menemukan solusi suatu permasalahan (Nuryawan et al., 2023). Penelitian yang membandingkan algoritma *Support Vector Machine* dan *Decision Tree* dalam analisis sentimen terhadap kepuasan pengguna aplikasi Migo



E-Bike di menunjukkan bahwa algoritma *Decision Tree* memiliki kinerja lebih baik, dengan perolehan akurasi sebesar 76,39% (Al Azkiah et al., 2024).

Namun, permasalahan umum dalam analisis sentimen adalah ketidakseimbangan kelas (*class imbalance*), yaitu kondisi ketika jumlah data pada masing-masing kategori sentimen tidak seimbang. Ketidakseimbangan ini dapat menyebabkan model cenderung lebih akurat pada kelas dengan data yang dominan dan kurang optimal pada kelas lainnya. Untuk mengatasi hal ini, penelitian ini menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). Penelitian (Cahyaningtyas et al., 2021) pada ulasan aplikasi Shopee menunjukkan bahwa penggunaan SMOTE dengan *Decision Tree* mampu meningkatkan akurasi hingga 99,91% dan AUC 0,999, dibandingkan tanpa SMOTE yang menghasilkan akurasi 99,89% dan AUC 0,950. Selanjutnya, penelitian (Putri et al., 2025) menunjukkan bahwa penerapan metode *Decision Tree* berbasis SMOTE mampu menghasilkan akurasi sebesar 88%, dengan nilai *precision*, *recall*, dan *F1-score* masing-masing mencapai 91%. Di sisi lain, penelitian (Teguh, 2023) pada aplikasi PLN Mobile menggunakan *Decision Tree* menghasilkan akurasi tertinggi 96,5%, *precision* 87,2%, dan *recall* 83%.

Meskipun penelitian terdahulu menunjukkan hasil yang baik, terdapat beberapa kesenjangan yang menjadi dasar penelitian ini. Sebagian besar studi sebelumnya hanya mengklasifikasikan ulasan ke dalam dua kelas sentimen, yaitu positif dan negatif, sehingga ulasan netral belum banyak diperhatikan dan penelitian sebelumnya masih menggunakan tahapan *preprocessing* yang terbatas, sedangkan penelitian ini menambahkan tahap yang lebih komprehensif sehingga data yang digunakan untuk klasifikasi menjadi lebih representatif.

Berdasarkan uraian tersebut, penelitian ini berjudul “Analisis Sentimen Pengguna Aplikasi *Mobile Banking* Bank Sulselbar Pada *Google Play Store* Dengan Metode *Decision Tree* Berbasis SMOTE.” Penelitian ini diharapkan dapat memberikan kontribusi baru dengan menghasilkan analisis sentimen yang lebih akurat, seimbang, dan representatif terhadap opini pengguna aplikasi Mobile Bank Sulselbar.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dibahas sebelumnya maka rumusan masalah pada penelitian ini yaitu:

Bagaimana kinerja metode *Decision Tree* berbasis *SMOTE* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi mobile Bank Sulselbar berdasarkan data dari *google play store* ?

## 1.3 Tujuan Penelitian

Berdasarkan latar belakang yang telah dibahas sebelumnya maka tujuan dari penelitian ini yaitu:

Untuk menilai kinerja metode *Decision Tree* berbasis *SMOTE* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi mobile bank sulselbar berdasarkan data dari *google play store*.

## 1.4 Manfaat Penelitian

Berdasarkan dengan permasalahan dan tujuan penelitian, maka penulis mengharapkan penelitian ini dapat memberikan beberapa manfaat di bawah ini:

1. Memberikan informasi mengenai sentimen pengguna terhadap aplikasi mobile Bank Sulselbar.
2. Memberikan gambaran mengenai seberapa baik metode *Decision Tree* berbasis *SMOTE* dalam menghasilkan akurasi dan memperbaiki keseimbangan kinerja model pada proses klasifikasi.

## 1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan berupa ulasan pengguna aplikasi mobile Bank Sulselbar yang tersedia di *google play store*.
2. Sentimen yang dianalisis dikategorikan menjadi tiga kelas, yaitu positif, negatif, dan netral.
3. Penelitian ini tidak membahas performa aplikasi secara teknis, tetapi hanya berfokus pada ulasan berbasis teks dari pengguna.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Landasan Teori**

##### **2.1.1 *Mobile Banking***

*Mobile banking*, atau yang biasa disebut *m-banking*, adalah layanan yang memungkinkan nasabah melakukan berbagai aktivitas perbankan melalui perangkat seluler seperti ponsel. Melalui aplikasi yang diinstal di ponsel, pengguna dapat mengakses berbagai fitur perbankan secara praktis kapan pun dan di mana pun. Dibandingkan dengan *SMS banking*, layanan ini menawarkan fungsionalitas yang lebih lengkap dan modern. Perlu dibedakan bahwa *mobile banking* tidak sama dengan *SMS banking* maupun *internet banking*. *Mobile banking* memberikan pengalaman layanan yang hampir menyerupai mesin ATM, meskipun tidak dapat digunakan untuk penarikan tunai. Sementara itu, *SMS banking* hanya menyediakan layanan berbasis pesan singkat yang dikirim melalui jaringan seluler untuk mendapatkan informasi atau melakukan transaksi tertentu. Sedangkan *internet banking* memungkinkan nasabah mengakses layanan perbankan melalui website resmi bank dengan koneksi internet, yang mendukung berbagai aktivitas seperti pengecekan saldo, transfer dana, hingga pembayaran tagihan. (M.Sayuti, 2021).

##### **2.1.2 *Google Play Store***

*Google play store* adalah platform digital milik *Google* yang ditujukan untuk pengguna Android. Pengguna dapat mengakses berbagai aplikasi, game, film, musik, dan e-book, baik gratis maupun berbayar. Sebagai pusat distribusi utama di ekosistem Android, memudahkan pengguna dalam mencari dan mengunduh konten digital. Fitur ulasan dan penilaian pada aplikasi di *Play Store* memainkan peran penting dalam membentuk persepsi pengguna. Ulasan memberikan gambaran tentang pengalaman pengguna sebelumnya, sedangkan rating mencerminkan kualitas aplikasi secara umum. pengembang perlu merespons ulasan, memperbaiki kekurangan, serta mendorong pengguna untuk memberikan ulasan yang membangun guna meningkatkan reputasi aplikasi (Sebastian Vidal, 2023).

### 2.1.3 *Machine Learning*

*Machine learning* (ML) merupakan salah satu cabang dari kecerdasan buatan (*artificial intelligence/AI*) yang berfokus pada pengembangan algoritma yang mampu belajar secara otomatis dan meningkatkan performanya berdasarkan pengalaman tanpa harus diprogram secara eksplisit. Teknologi ini telah banyak dimanfaatkan untuk mendukung berbagai aktivitas manusia, terutama dalam melakukan prediksi. Keberhasilan penerapan *machine learning* dalam melakukan prediksi sangat bergantung pada kualitas data serta algoritma pembelajaran yang digunakan. Istilah *machine learning* dan *artificial intelligence* kini menjadi sangat populer di dunia teknologi. Secara umum, algoritma *machine learning* dibagi menjadi empat kategori utama, yaitu *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement learning* (Rasyid et al., 2022).

*Machine learning* adalah cabang ilmu komputer yang memanfaatkan pengalaman (peristiwa) pada masa lalu untuk dipelajari dan menggunakan pengetahuannya untuk membuat keputusan di masa depan. *Machine learning* merupakan irisan ilmu komputer, teknik dan statistika (Prasetya, 2022). Tujuannya untuk mengembangkan algoritma dan model komputer yang dapat mengidentifikasi pola dalam data dan membuat keputusan atau prediksi berdasarkan pola tersebut (Anugrah et al., 2023).

### 2.1.4 *Analisis Sentimen*

Analisis sentimen merupakan sebuah proses untuk menganalisis atau mengidentifikasi sebuah opini seseorang yang menunjukkan sikap terhadap suatu topik atau produk tertentu masuk ke dalam kategori positif, negatif, atau netral. Tidak seperti informasi faktual, opini dan sentimen memiliki karakteristik yang menunjukkan bahwa hal tersebut bersifat subjektif. Pemeriksaan sekumpulan opini dari banyak pihak diperlukan karena hal tersebut merupakan pandangan subjektif yang asalnya lebih dari satu orang sehingga diperlukan sebuah ringkasan untuk mewakili suatu pendapat (Parasati et al., 2020).

Analisis sentimen merupakan bidang penelitian yang berfokus untuk menganalisis opini, emosi, sikap, dan sentimen publik terhadap suatu entitas yang

mencakup produk, layanan, isu, organisasi, acara, atau fitur tertentu. Analisis sentimen dapat dilakukan dengan berbagai macam algoritma berbeda (Husen et al., 2023). Tujuan utama analisis sentimen adalah untuk mengolah, mengekstrak, meringkas, dan menganalisis informasi dalam sebuah teks menggunakan berbagai metode sehingga dapat menyimpulkan sudut pandang penulis dari teks dan berbagi informasi subjektif penulis tentang kecenderungan emosional teks di dalamnya (Kurniawan et al., 2022).

### 2.1.5 *Text Mining*

*Text mining* merupakan cabang dari *data mining* yang bertujuan untuk mengekstrak data untuk menemukan pola atau hubungan unik yang mencerminkan isi atau karakteristik khusus dari sebuah dokumen. Beberapa permasalahan yang dapat diatasi oleh *text mining* yaitu analisis sentimen, pembubaran dokumen, klasifikasi dokumen, ekstraksi informasi, pencarian informasi, dan pemrosesan. *Text mining* merupakan proses ekstraksi informasi dari data teks dan sumber data yang berasal dari dokumen yang memungkinkan analisis hubungan antar dokumen, tujuan *text mining* adalah untuk menemukan kata-kata yang mewakili isi dokumen (Salsabila et al., 2022).

### 2.1.6 *Preprocessing*

*Preprocessing* merupakan proses pengolahan dataset mentah menjadi dataset yang lebih bersih, terorganisir, dan mudah dipahami sehingga siap digunakan pada tahap analisis selanjutnya. Tahapan ini memiliki peran penting dalam memastikan kualitas data, karena data yang masih tidak terstruktur atau mengandung *noise* dapat menghambat kinerja model analisis maupun algoritma *machine learning*. Tujuan utama dari *preprocessing* adalah untuk mengubah data teks yang tidak terstruktur menjadi format yang lebih terstruktur dan terstandarisasi agar dapat diproses secara efektif. Secara umum, tahapan *preprocessing* melibatkan beberapa proses utama yang dilakukan secara sistematis yaitu (Ritonga et al., 2023).

#### 1. *Cleaning*

*Cleaning* merupakan tahap pembersihan kata-kata yang kurang penting seperti URL, dan nama pengguna (@username), Meskipun URL Format URL antara lain https atau www., selain itu tanda baca seperti titik(.) koma(,) dan

*emoticon*. Proses ini bertujuan untuk meminimalisir *noise* data yang tidak mendukung.

## 2. *Case folding*

*Case folding* merupakan proses untuk menukar huruf atau tulisan besar menjadi tulisan kecil. Tujuan dari tahapan ini untuk menyederhanakan proses selanjutnya.

## 3. *Tokenizing*

*Tokenizing* merupakan suatu tahapan pemecahan kalimat untuk membentuk suatu token atau potongan kata.

## 4. *Normalisasi*

*Normalisasi* merupakan tahapan untuk memperbaiki ejaan yang salah pada setiap token atau mengubah menjadi kata baku.

## 5. *Stopword removal*

*Stopword removal* merupakan proses penghilangan kata-kata yang dianggap tidak memiliki makna penting dalam sebuah teks sehingga hanya kata-kata yang relevan saja yang dipertahankan. Tahapan ini dapat dilakukan dengan menggunakan daftar kata yang harus diabaikan (*stoplist*) atau dengan memilih kata-kata yang relevan melalui daftar kata yang dibolehkan (*word list*).

## 6. *Stemming*

*Stemming* merupakan proses mengubah berbagai variasi bentuk kata menjadi satu bentuk dasar atau kata akar. Dengan *stemming*, kata-kata yang memiliki makna sama dipetakan ke bentuk baku sehingga perbedaan variasi kata dapat diminimalkan.

### 2.1.7 *Lexicon Based*

*Lexicon based* merupakan salah satu metode dalam analisis sentimen yang menggunakan kamus *leksikon* yang telah diakurasi secara manual. Kamus *leksikon* ini berisi daftar kata-kata atau frasa yang memiliki hubungan dengan sentimen tertentu, seperti positif, negatif, atau netral. Setiap kata atau frase dalam kamus diberi label sentimen yang sesuai (Prasetya et al., 2021). *Lexicon based* adalah sebuah fitur kata yang memiliki sentimen positif atau negatif berdasarkan kamus



atau *lexicon*. Proses pelabelan data dilakukan oleh kamus *lexicon based* dengan menghitung skor sentimen. Setelah kata-kata yang mengandung sentimen positif, negatif, dan netral diidentifikasi dalam sebuah kalimat, langkah selanjutnya adalah menghitung setiap kata yang mengandung sentimen dalam kalimat tersebut, dengan menjumlahkan nilai opini. (Manullang et al., 2023) Berikut persamaan yang digunakan dalam menentukan label *score* sentimen.

$$Sentence_{sentiment} \begin{cases} \text{positive if } S_{positive} > S_{negative} \\ \text{neutral if } S_{positive} = S_{negative} \\ \text{negative if } S_{positive} < S_{negative} \end{cases} \quad (2.1)$$

Keterangan:

Positif = jika kata positif lebih banyak dari kata negatif

Negatif = jika kata negatif lebih banyak dari kata positif

Netral = jika jumlah kata positif sama dengan kata negatif

### 2.1.8 Term Frequency-Inverse Document Frequency (TF-IDF)

Metode TF-IDF adalah teknik yang digunakan untuk menentukan bobot setiap kata yang sering digunakan dalam proses pengambilan informasi. Konsep utama TF-IDF adalah menghitung bobot berdasarkan hubungan antara suatu kata (*term*) dengan dokumen tertentu. Teknik ini berperan sebagai indikator statistik untuk menilai seberapa penting suatu kata dalam sebuah dokumen atau dalam Kumpulan dokumen (Ramlan et al., 2023). Metode TF-IDF memperhitungkan dua faktor penting yaitu (Kharisma et al., 2023).

1. *Term Frequency* (TF): mengukur seberapa sering suatu kata muncul dalam sebuah dokumen. untuk menghitung TF adalah dengan menghitung jumlah kemunculan kata tersebut dibagi dengan jumlah total kata dokumen. Dapat dihitung dengan persamaan (2.2) :

$$tf_{(t,d)} = \frac{\text{frekuensi kemunculan } t \text{ dalam dokumen } d}{\text{jumlah kata dalam dokumen } d} \quad (2.2)$$

2. *Inverse Document Frequency* (IDF): adalah metode untuk mengukur tingkat pentingnya sebuah kata dalam suatu konteks koleksi dokumen yang lebih besar. Konsep ini bekerja dengan prinsip semakin jarang sebuah kata muncul di seluruh dokumen, semakin tinggi nilai IDF-nya. Perhitungan IDF dilakukan dengan membandingkan jumlah total dokumen dengan banyaknya dokumen yang mengandung kata tersebut. Rumus perhitungannya adalah jumlah total dokumen yang dibagi dengan jumlah dokumen yang memuat kata tersebut, lalu hasilnya diambil algoritma untuk menormalkan skala nilai. Berikut persamaanya (2.3):

$$idf_{(t)} = \log\left(\frac{N}{df_{(t)}}\right) \quad (2.3)$$

*Term Frequency - Inverse Document Frequency* (TF-IDF) dihitung setiap kata dengan menggunakan persamaan (2.4) :

$$tf - idf_{(t,d)} = tf \times idf \quad (2.4)$$

Keterangan :

t = kata (term) yang terdapat dalam dokumen

d = dokumen

N = jumlah total dokumen dalam korpus

idf = nilai yang menunjukkan seberapa jarang term t muncul

df = banyak dokumen yang mengandung term t

### 2.1.9 *Synthetic Minority Over-Sampling Technique* (SMOTE)

Metode SMOTE pertama kali diperkenalkan oleh Nithes V. Chawla pada tahun 2002 sebagai salah satu pendekatan untuk menangani permasalahan distribusi data yang tidak seimbang. Pendekatan ini berbeda dari teknik *oversampling* yang telah diperkenalkan sebelumnya (Sofyan & Prasetyo, 2021). SMOTE merupakan sebuah teknik yang digunakan untuk mengimbangi jumlah distribusi sampel data pada kelas minoritas dengan memilih sampel data tersebut sehingga jumlahnya seimbang dengan jumlah sampel data kelas mayoritas. Dengan menggunakan

metode ini, data pada kelas minoritas dapat di-"oversampling" sehingga dataset menjadi lebih seimbang antara kelas mayoritas dan kelas minoritas. (Chiekal & Kurniasih, 2023). Berikut langkah penerapan *SMOTE* (Rika Widianita, 2023).

- Memilih data yang akan direplikasi ( $x_i$ ) dari kelas minoritas secara acak.
- Menentukan nilai  $k$  (jumlah tetangga terdekat) dan menghitung jarak antara data  $x_i$  dan data tetangga terdekat ( $x_{knn}$ ) yang berasal dari kelas minoritas yang sama.
- Untuk setiap  $x_{knn}$  yang dipilih, hitung selisih antara  $x_i$  dan  $x_{knn}$ , kemudian kalikan selisih tersebut dengan angka acak dalam rentang  $[0,1]$ , dan tambahkan hasilnya ke fitur yang sedang dianalisis.

$$x_{syn} = x_i + (x_{knn} - x_i) \times \delta \quad (2.5)$$

Keterangan

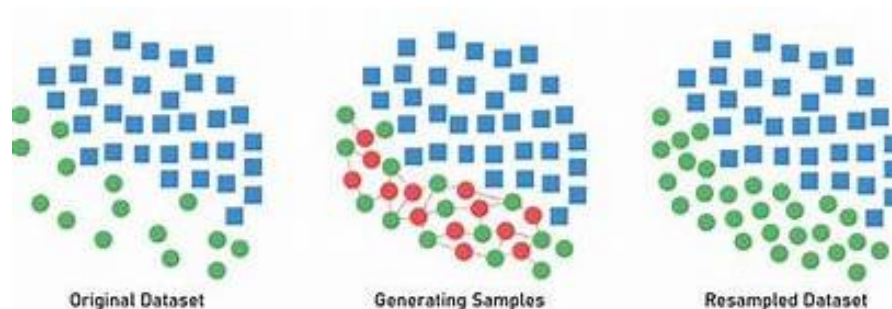
$x_{syn}$  = data sintetis yang akan diciptakan

$x_i$  = data yang akan direplikasi

$x_{knn}$  = data yang memiliki jarak terdekat dari  $x_i$

$\delta$  = nilai random dari  $[0,1]$

- Menggabungkan data asli dengan data yang dihasilkan.



**Gambar 2. 1** Ilustrasi SMOTE

<https://python.plainenglish.io/balancing-the-scales-how-SMOTE-transforms-machine-learning-with-imbalanced-data-a6d3367254cd>

Gambar 2.1 menunjukkan bagaimana teknik *Synthetic Minority Over-sampling Technique* (SMOTE) digunakan untuk menangani ketidakseimbangan kelas dalam dataset. Pada bagian kiri, terlihat kelas mayoritas (kotak biru) jauh lebih banyak dibanding kelas minoritas (lingkaran hijau), yang dapat menyebabkan

model bias terhadap kelas mayoritas. Bagian tengah menunjukkan proses pembuatan data sintetis (lingkaran merah) untuk kelas minoritas. Data ini dihasilkan bukan dengan menyalin data lama, tetapi melalui interpolasi dari data minoritas yang sudah ada. Hasil akhirnya, pada bagian kanan gambar, menunjukkan distribusi data yang lebih seimbang antara kedua kelas. Dengan demikian, model dapat belajar dengan lebih adil dan mengenali kedua kelas secara lebih akurat.

#### 2.1.10 *K-Fold Cros Validation*

*K-fold Cross Validation* merupakan teknik dalam data mining yang digunakan untuk memperoleh akurasi yang optimal dengan membagi dataset menjadi *subset* data latih dan data uji secara berulang. Teknik ini membagi dataset secara acak ke dalam K partisi. Pada setiap iterasi, satu partisi digunakan sebagai data uji, sementara partisi lainnya digunakan sebagai data latih. Proses ini diulang sebanyak K kali sesuai dengan nilai K yang ditentukan dalam penelitian (Hendriyanto et al., 2022). Teknik ini kerap digunakan karena terbukti mampu memastikan bahwa model tidak hanya bekerja baik pada satu *subset* data saja, tetapi juga memberikan performa yang stabil pada seluruh dataset. (Khoerunnisa et al., 2025).

#### 2.1.11 *Decision Tree*

*Decision Tree* adalah salah satu algoritma yang umum digunakan untuk mengambil sebuah keputusan. Algoritma ini baik digunakan untuk klasifikasi atau prediksi. Pohon keputusan merupakan sebuah metode klasifikasi yang bekerja dengan cara membentuk sebuah struktur seperti pohon, di mana setiap keputusan diwakili oleh sebuah *node*, dan cabang-cabangnya menghubungkan *node* tersebut dari titik awal (akar) hingga ke bagian akhir (daun). Model *Decision Tree* berbentuk sebuah pohon yang terdiri dari beberapa bagian yaitu *root node*, *internal node*, dan *terminal node*. *Data query* menelusuri *root node* dan *internal node* mencapai *terminal node* merupakan proses melakukan klasifikasi pada metode *Decision Tree* ini (Tangkelayuk, 2022).

*Decision Tree* merupakan sebuah representasi terstruktur dari sekumpulan atribut yang digunakan untuk melakukan pengujian dengan tujuan memprediksi *outputnya*. Pohon keputusan terdiri dari simpul-simpul, di mana simpul teratas disebut simpul akar (Cahyaningtyas et al., 2021).

1. menghitung *entropy* untuk mengukur ketidakmurnian atau ketidakpastian setiap atribut menggunakan persamaan (2.6)

$$Entropy (S) = \sum_{i=0}^n - p_i * \log^2 p_i \quad (2.6)$$

Keterangan:

S : Himpunan kasus.

n : Jumlah partisi S

$p_i$  : proporsi yang didapat dari jumlah kelas dibagi total kasus

2. Menghitung *Information gain* mengukur seberapa banyak ketidakpastian *entropy* dari setiap atribut menggunakan persamaan (2.7)

$$Gain (S, A) = Entropy (S) - \sum_i^n \frac{(s_i)}{(s)} * Entropy (S_i) \quad (2.7)$$

Keterangan:

s : Kasus.

A : Atribut

n : Jumlah partisi atribut A

$S_i$  : Jumlah kasus pada partisi ke-i.

S : Jumlah kasus s

### 2.1.12 Evaluasi

Evaluasi model adalah suatu proses menilai suatu performa model yang telah dihasilkan. Performa model akan diukur menggunakan *confusion matrix* (Naseer et al., 2024). Menurut (Fibriyanti Arminda et al., 2023). *Confusion matrix* adalah metode dalam data mining yang digunakan untuk menilai suatu algoritma atau model klasifikasi, dengan memberikan gambaran mengenai *accuracy*,

*precision*, *recall*. menunjukkan kemungkinan hasil klasifikasi dengan *matrix* jumlah data uji untuk kelas sebenarnya.

Dalam penelitian ini digunakan format 3x3 karena klasifikasi sentimen terbagi ke dalam tiga kategori: positif, negatif, dan netral (Vincent et al., 2024).

**Tabel 2. 1** Rumus *Multiclass Confusion Matrix*

		Nilai Akurasi		
		Positif	Netral	Negatif
Nilai Prediksi	Positif	True Positive (TP)	False Positive (FP)	False Positive (FP)
	Netral	False Neutral (FL)	True Neutral (TL)	False Neutral (FL)
	Negatif	False Negative (FN)	False Negative (FN)	True Negative (TN)

True Positive(TP) : Data dengan aktual positif dan diprediksi positif.

False Positive (FP) : Data dengan aktual bukan positif yang diprediksi positif.

True Negatif(TN) : Data dengan aktual negatif yang diprediksi negatif

False Negatif(FN) : Data dengan aktual bukan negatif yang diprediksi negatif.

True Neutral (TL) : Data dengan aktual netral yang diprediksi netral.

False Neutral (FL) : Data dengan aktual bukan netral yang diprediksi netral.

Melalui *confusion matrix* dapat dilakukan perhitungan berbagai matrik evaluasi kinerja model di antaranya *accuracy*, *precision*, *recall* dan *f1-Score*. Berikut penjelasan tersebut. Pengujian menggunakan *confusion matrix* menghasilkan empat jenis keluaran utama yaitu:



1. *accuracy* digunakan untuk menilai sejauh mana suatu model mampu melakukan klasifikasi data dengan tepat secara keseluruhan. Berikut rumus yang digunakan untuk menghitung nilai *accuracy*.

$$accuracy = \frac{TP+TL+TN}{TP+TL+TN+FN+FP+FL} \quad (2.8)$$

2. *Precesion* digunakan untuk menilai banyaknya prediksi positif yang benar, serta mengurangi kemungkinan *false* positif. Berikut adalah rumus untuk menghitung nilai *precesion* :

$$precesion = \frac{TP}{TP+FP+FN} \quad (2.9)$$

3. *Recall* digunakan untuk mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total data positif yang sebenarnya, serta mengurangi kemungkinan *false* negatif. Berikut rumus untuk menghitung nilai *recall* :

$$recall = \frac{TP}{TP+FL+FN} \quad (2.10)$$

4. *F1-score* merupakan rata-rata harmonik dari *precesion* dan *recall* yang digunakan untuk menilai kinerja model dengan *precesion* dan *recall* yang tinggi. Berikut rumus untuk menghitung nilai skor *F1*:

$$f1\ score = 2 \times \frac{Recall*Precesion}{Recall+Precision} \quad (2.11)$$

## 2.2 Penelitian Terdahulu

Dalam penelitian ini, kajian terhadap penelitian terdahulu menjadi landasan penting untuk memperluas wawasan dan memperoleh informasi yang relevan. Tujuannya adalah agar penelitian ini dapat dikembangkan dengan lebih baik. Oleh karena itu, digunakan sejumlah referensi dari penelitian sebelumnya yang memiliki keterkaitan dengan permasalahan yang diteliti.

**Tabel 2. 2** Penelitian terdahulu

No	peneliti	Judul Penelitian	Hasil Penelitian	Perbedaan	Persamaan
1.	(Dewi, 2025)	Analisis Sentimen terhadap Ulasan Aplikasi Canva di <i>Play Store</i> dengan Menggunakan Pendekatan <i>Lexicon</i> dan Algoritma <i>Decision Tree</i> .	hasil analisis menunjukkan model klasifikasi, algoritma <i>Decision Tree</i> menunjukkan performa yang sangat baik dengan akurasi sebesar 87%. Kinerja terbaik tercatat pada kelas “2”, dengan nilai <i>precision</i> sebesar 90%, <i>recall</i> 89%, dan <i>f1-score</i> 89%.	Penelitian yang diusulkan menerapkan metode pembobotan TF-IDF, dan kategori kelas yang digunakan yakni positif, negatif, dan netral sedangkan penelitian sebelumnya tidak melibatkan tahapan tersebut dalam proses pengolahan datanya dan kategori kelas yang digunakan	Penelitian terbaru menggunakan metode <i>Decision Tree</i> dan pelabelan <i>lexicon</i> untuk analisis sentimen di <i>google play store</i> yang sama.

				hanya dua yakni positif dan negatif	
2.	(Al Azkiah et al., 2024)	Perbandingan Algoritma SVM dan <i>Decision Tree</i> Dalam Klasifikasi Kepuasan Pengguna Aplikasi Migo E-Bike di <i>play store</i>	Penelitian tentang aplikasi Migo E-Bike menunjukkan Hasil penelitian bahwa algoritma <i>Decision Tree</i> lebih unggul dibandingkan dengan <i>Support Vector Machine</i> (SVM) dalam mengklasifikasikan kepuasan pengguna berdasarkan rating di <i>Play store</i> . <i>Decision Tree</i> mencapai akurasi 76,39% dengan keseimbangan <i>precision</i> dan <i>recall</i> yang baik, sementara SVM hanya 51,25%, dengan ketidakseimbangan signifikan, terutama dalam kategori "tidak puas". <i>Dataset</i> terbagi antara 181 rating "tidak puas" dan 180	pada penelitian sebelumnya hanya menggunakan empat tahap <i>preprocessing</i> yakni <i>Case folding</i> , <i>Tokenizing</i> , <i>Filtering</i> , <i>Stemming</i> dan kategori sentimen yang digunakan dua yakni positif dan negatif. Sedangkan pada penelitian yang diusulkan menggunakan enam tahap <i>preprocessing</i> yakni <i>Cleaning</i> , <i>Case Folding</i> , <i>Tokenizing</i> , <i>Normalisasi</i> , <i>Stopword Removal</i> , <i>Stemming</i> dan kategori sentimen yang digunakan tiga yakni positif, negatif dan netral	Penelitian terbaru menerapkan metode yang sama yaitu <i>Decision Tree</i> dalam analisis sentimen

			"sangat puas", menunjukkan distribusi yang seimbang.		
3.	(Putri et al., 2025)	Analisis sentimen peran artificial intelligence terhadap kreativitas dan efektivitas mahasiswa dalam penyelesaian tugas akhir menggunakan <i>Decision Tree</i> berbasis SMOTE	Hasil analisis menunjukkan dengan metode <i>Decision Tree</i> berbasis SMOTE, diperoleh akurasi sebesar 88%, dengan nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> mencapai 91% dengan hasil analisis sentimen menunjukkan bahwa 69,7% mahasiswa memberikan sentimen positif terhadap penggunaan AI, sementara 30,3% negatif.	Penelitian yang diusulkan dalam memperoleh datanya dengan cara <i>scrapping</i> menggunakan python, dan tahap <i>preprocessing</i> menggunakan enam tahap dan pelabelan yang digunakan yaitu <i>lexicon based</i> sedangkan penelitian terdahulu proses pengumpulan datanya menggunakan kuesioner dan menggunakan tiga tahap <i>preprocessing</i> dan pelabelan menggunakan berbasis skala Likert.	Penelitian terbaru menerapkan metode yang sama yaitu <i>Decision Tree</i> berbasis SMOTE dalam analisis sentimen
4.	(Fatkhudin et al., 2024)	<i>Decision Tree</i> berbasis <i>SMOTE</i> dalam analisis	Hasil dari penelitian	Pada penelitian yang diusulkan akan berjalan	Penelitian terbaru menerapkan

		sentimen pengguna <i>artificial iniligence</i> untuk skripsi	yang dilakukan didapatkan bahwa <i>Decision Tree C.45</i> berbasis <i>SMOTE</i> mendapatkan hasil akurasi sebesar 66% yang mengindikasikan ketepatan algoritma <i>Decision Tree C.45</i> berbasis <i>SMOTE</i> dalam mengklasifikasikan sentimen. Melalui klasifikasi sentimen ditemukan bahwa 84,4% dari opini yang muncul memiliki nilai sentimen yang negatif, sedangkan pada sentimen positif ditemukan sebanyak 15,6%.	menggunakan aplikasi IDLE <i>Python</i> dalam melakukan pemrosesan data sedangkan penelitian terdahulu menggunakan RapidMiner sebagai perangkat utama untuk melakukan klasifikasi dan evaluasi.	metode yang sama yaitu <i>Decision Tree</i> berbasis <i>SMOTE</i> untuk menganalisis sentimen
5.	(Teguh, 2023)	Analisis sentimen pada PLN mobile	hasil penelitian yang telah dilakukan, mengenai analisis	Penelitian sebelumnya hanya menggunakan lima tahap	Penelitian terbaru menerapkan

		menggunakan metode <i>Decision Tree</i>	sentimen pada PLN mobile menggunakan metode <i>Decision Tree</i> menunjukkan hasil dilakukan dengan akurasi tertinggi sebesar 96,5%, <i>precision</i> 87,2%, dan <i>recall</i> 83%.	<i>preprocessing</i> dan kategori sentimen yang digunakan (positif dan negatif) sedangkan penelitian yang diusulkan menggunakan enam tahap <i>preprocessing</i> dan tiga kategori sentimen (Positif, negatif dan netral).	metode yang sama yaitu <i>Decision Tree</i> dalam analisis sentimen.
6.	(Adi et al., 2024)	Analisis Sentimen Masyarakat Terhadap Progress Ikn Menggunakan Model <i>Decision Tree</i>	Hasil analisis yang dilakukan, ditemukan bahwa tanggapan masyarakat terhadap IKN dominan memiliki sentimen positif yaitu sebesar 42.3% diikuti dengan sentimen netral sebesar 30.1%, dan paling sedikit sentimen negatif sebesar 27.6%. Kemudian ditemukan juga bahwa model <i>Decision Tree</i> yang digunakan memiliki	Penelitian sebelumnya menggunakan <i>preprocessing</i> hanya empat tahap sedangkan pada penelitian yang diusulkan menggunakan enam tahap <i>preprocessing</i>	Penelitian terbaru menerapkan metode yang sama yaitu <i>Decision Tree</i> dan tiga kelas (positif, negatif, dan netral)



			tingkat akurasi sekitar 81% dengan 10-fold Cross Validation.		
7.	(Zulfahmi, 2023)	Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode <i>Decision Tree</i>	Hasil pengujian dan evaluasi menggunakan algoritma <i>Decision Tree</i> menunjukkan akurasi sebesar 96%, rata-rata <i>precision</i> sebesar 91%, rata-rata <i>recall</i> sebesar 96%, dan rata-rata skor <i>f1</i> sebesar 93%.	Penelitian yang diusulkan menggunakan metode SMOTE untuk menyeimbangkan data yang tidak seimbang sedangkan pada penelitian sebelumnya tidak menggunakan.	Penelitian terbaru menerapkan metode yang sama yaitu <i>Decision Tree</i> dalam menganalisis sentimen
8.	(Munir et al., 2024)	Perbandingan Akurasi Algoritma <i>Naive Bayes</i> dan Algoritma <i>Decision Tree</i> dalam Pengklasifikasian Penyakit Kanker Payudara	hasil akhir penelitian menunjukkan <i>Decision Tree</i> memiliki akurasi yang lebih tinggi sebesar 92.04% dan <i>naïve bayes</i> memiliki akurasi 91.15% Hasil ini membuktikan bahwa <i>Decision Tree</i> lebih unggul dalam pengklasifikasian penyakit kanker payudara	Pada penelitian sebelumnya hanya menggunakan dua kelas yaitu positif dan negatif sedangkan pada penelitian yang diusulkan menggunakan tiga kelas yaitu positif, negatif, dan netral.	Penelitian terbaru menerapkan metode yang sama yaitu <i>Decision Tree</i> dalam analisis sentimen.

			dibandingkan dengan <i>naïve bayes</i> .		
9.	(Cahyaningtyas et al., 2021)	Analisis sentimen pada rating aplikasi Shopee menggunakan metode <i>Decision Tree</i> berbasis SMOTE	Hasil penelitian mengenai analisis sentimen pada rating aplikasi Shopee menggunakan metode <i>Decision Tree</i> berbasis <i>SMOTE</i> menunjukkan tingkat akurasi yang sangat tinggi. Penggunaan <i>SMOTE</i> menghasilkan nilai akurasi sebesar 99,91%, dengan AUC ( <i>Area Under Curve</i> ) mencapai 0,999, <i>recall</i> 99,88%, dan <i>precision</i> 99,98%. Sementara itu, tanpa menggunakan <i>SMOTE</i> , akurasi yang diperoleh adalah 99,89%, dengan AUC sebesar 0,950, <i>recall</i> 99,88%, dan <i>precision</i> 99,98%.	Namun pada penelitian sebelumnya hanya menggunakan empat tahap <i>Preprocessing Data</i> yakni <i>tokenize, filter token, Stopword, Transform Case</i> dan menggunakan RapidMiner sebagai perangkat utama untuk melakukan klasifikasi dan evaluasi, Sedangkan pada penelitian yang diusulkan menggunakan enam tahap <i>Preprocessing Data</i> yakni <i>Cleaning, Case Folding, Tokenizing, Word Normalisasi, Stopword Removal, Stemming</i> dan Pada penelitian yang	Penelitian terbaru menerapkan metode yang sama yaitu metode <i>Decision Tree</i> dan teknik SMOTE dalam analisis sentimen

				disulkan menggunakan aplikasi IDLE Python ( <i>Google Colab</i> ) dalam melakukan pemrosesan data.	
10.	(Prasasti & Amarudin, 2025)	Analisis Sentimen Pengguna Twitter terhadap Konten Edukasi Dokter Detektif dengan Metode Pohon Keputusan.	Hasil penelitian menunjukkan bahwa <i>Decision Tree</i> menghasilkan akurasi sebesar 84% dan menunjukkan performa yang lebih stabil dibandingkan Naive Bayes berdasarkan evaluasi <i>precision</i> , <i>recall</i> , <i>f1-score</i> , dan <i>confusion matrix</i> . Temuan ini menunjukkan bahwa <i>Decision Tree</i> lebih efektif dalam menganalisis sentimen teks terkait konten edukatif di media sosial, khususnya dalam domain kesehatan.	Penelitian terdahulu membandingkan algoritma <i>Naive Bayes</i> dan <i>Decision Tree</i> , sedangkan pada penelitian terdahulu menggunakan kelas sentimen positif dan negatif, sedangkan pada yang diusulkan hanya berfokus satu algoritma yaitu <i>Decision Tree</i> , sedangkan menggunakan tiga klas sentimen yaitu positif, negatif, dan netral	Penelitian terbaru menerapkan metode yang sama yaitu metode <i>Decision Tree</i> dan teknik SMOTE dalam analisis sentimen

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang diperoleh dari ulasan pengguna aplikasi *mobile banking* Bank Sulselbar. Dataset yang digunakan berjumlah 1.626 data setelah dilakukan pembersihan nilai *NaN* pada tahap pelabelan. Berdasarkan hasil evaluasi model *Decision Tree*, pada skenario pembagian data dengan rasio 80:20 sebelum penerapan metode SMOTE diperoleh akurasi sebesar 81%, namun nilai *f1-score* keseluruhan kelas menunjukkan ketidakseimbangan, di mana model cenderung lebih akurat dalam mengenali kelas mayoritas. Setelah penerapan SMOTE, akurasi menurun menjadi 74%, tetapi nilai *f1-score* meningkat berada pada rentang 40% hingga 85%, yang menunjukkan peningkatan keseimbangan kinerja model dalam mengenali setiap kelas sentimen. Sedangkan pada *10-fold cross validation*, akurasi yang diperoleh sebelum SMOTE mencapai 82%, meskipun nilai *f1-score* keseluruhan masih belum seimbang. Setelah penerapan SMOTE akurasi menurun menjadi 77% dengan peningkatan *f1-score* pada rentang 43% hingga 86%, yang mengindikasikan hasil yang lebih stabil dan seimbang antar kelas. Dengan demikian, penerapan SMOTE pada rasio 80:20 dan *10-fold cross validation* terbukti mampu meningkatkan keseimbangan antar kelas serta memperbaiki kemampuan model dalam mengenali kelas minoritas. Selain itu, evaluasi menggunakan *10-fold cross validation* menunjukkan kinerja model yang lebih stabil dan representatif terhadap keseluruhan data.

#### 5.2 Saran

Berdasarkan temuan dari penelitian ini, terdapat beberapa rekomendasi yang dapat diberikan untuk pengembangan penelitian di masa mendatang agar hasil yang diperoleh menjadi lebih komprehensif, akurat, dan dapat memberikan kontribusi yang lebih signifikan terhadap bidang analisis sentimen.

1. Penelitian selanjutnya disarankan untuk menggunakan jumlah data yang lebih besar dengan memanfaatkan sumber data lain.
2. Karena penelitian ini menggunakan metode *lexicon based* yang memiliki keterbatasan dalam memahami konteks bahasa, maka penelitian berikutnya disarankan untuk mencoba pelabelan manual berbasis *machine learning* agar hasil klasifikasi lebih akurat.
3. Penelitian selanjutnya dapat membandingkan metode *Decision Tree* dengan metode lain untuk mendapatkan hasil yang lebih optimal.

## DAFTAR PUSTAKA

- Adi, S. I. R., Bakkara, B., Zega, K. A., Vielita, F. N., & Rakhmawati, N. A. (2024). Analisis Sentimen Masyarakat Terhadap Progress Ikn Menggunakan Model Decision Tree. *JIKA (Jurnal Informatika)*, 8(1), 57. <https://doi.org/10.31000/jika.v8i1.9803>
- Al Azkiah, D. S., Erizal, E., & Hikmah, F. N. (2024). Perbandingan Algoritma SVM dan *Decision Tree* Dalam Klasifikasi Kepuasan Pengguna Aplikasi Migo E-Bike di Playstore. *Building of Informatics, Technology and Science (BITS)*, 6(1), 158–167. <https://doi.org/10.47065/bits.v6i1.5344>
- Alun Sujjadaa, Somantri, Juwita Nurfazri Novianti, & Indra Griha Tofik Isa. (2023). Analisis Sentimen Terhadap Review Bank Digital Pada *Google Play Store* Menggunakan Metode Support Vector Machine (Svm). *Jurnal Rekayasa Teknologi Nusa Putra*, 9(2), 122–135. <https://doi.org/10.52005/rekayasa.v9i2.345>
- Anugrah, G. B., Shafa, R. N. Z., & Kurniawan, A. (2023). Perbandingan Algoritma Machine Learning Untuk Klasifikasi Dry Bean Dataset. *Jurnal Responsif: Riset Sains Dan Informatika*, 5(2), 407–415. <https://doi.org/10.51977/jti.v5i2.1292>
- Arifuddin, A. A. A. I. (2023). Pengaruh Kompensasi dan Promosi Jabatan. *Jumansi Stindo*, 1(1), 32–38.
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Astri Prasasti, I., & Amarudin, A. (2025). Analisis Sentimen Pengguna Twitter terhadap Konten Edukasi Dokter Detektif dengan Metode Pohon Keputusan. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(7), 2005–2015. <https://doi.org/10.52436/1.jpti.893>
- Beno, J., Silen, A. ., & Yanti, M. (2022). Pengaruh Transaksi Digital *Banking* Terhadap Fee Based Income Pada Bank Umum Syariah. *Braz Dent J.*, 33(1), 1–12.
- Cahyaningtyas, C., Nataliani, Y., & Widiarsari, I. R. (2021). Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode *Decision Tree* Berbasis *SMOTE*. *Aiti*, 18(2), 173–184. <https://doi.org/10.24246/aiti.v18i2.173-184>
- Chieka, M., & Kurniasih, A. (2023). Teknik *SMOTE* untuk Mengatasi Imbalance Class dalam Klasifikasi Bank Customer Churn Menggunakan Algoritma



- Naïve bayes dan Logistic Regression. *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer Dan Aplikasinya*, 4(2), 552–559.
- Dewi, Y. P. (2025). *Analisis Sentimen terhadap Ulasan Aplikasi Canva di Play Store dengan Menggunakan Pendekatan Lexicon dan Algoritma Decision Tree*. 13, 57–63.
- Fatkhudin, A., Artanto, F. A., Safli, N. A., & ... (2024). *Decision Tree Berbasis SMOTE Dalam Analisis Sentimen Penggunaan Artificial Intelligence Untuk Skripsi*. *REMIK: Riset Dan E ...*, 8(April), 494–505. <https://www.jurnal.polgan.ac.id/index.php/remik/article/view/13531%0Ahttps://www.jurnal.polgan.ac.id/index.php/remik/article/download/13531/2453>
- Fibriyanti Arminda, N., Sulistiyowati, N., & Nur Padilah, T. (2023). Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1817–1822. <https://doi.org/10.36040/jati.v7i3.7012>
- Hendriyanto, M. D., Ridha, A. A., & Enri, U. (2022). Analisis Sentimen Ulasan Aplikasi Mola Pada *Google Play Store* Menggunakan Algoritma Support Vector Machine. *INTECOMS: Journal of Information Technology and Computer Science*, 5(1), 1–7. <https://doi.org/10.31539/intecom.v5i1.3708>
- Husen, R. A., Astuti, R., Marlia, L., Rahmadden, R., & Efrizoni, L. (2023). Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 211–218. <https://doi.org/10.57152/malcom.v3i2.901>
- Ihsan Zulfahmi. (2023). Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decission Tree. *Jurnal Penelitian Rumpun Ilmu Teknik*, 3(1), 11–21. <https://doi.org/10.55606/juprit.v3i1.3096>
- Kharisma, I. L., Septiani, D. A., Fergina, A., & Kamdan, K. (2023). Penerapan Algoritma *Decision Tree* untuk Ulasan Aplikasi Vidio di Google Play. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(2), 218–226. <https://doi.org/10.25077/teknosi.v9i2.2023.218-226>
- Khoerunnisa, S., Shiddieq, D. F., & Nurhayati, D. (2025). Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 566–577. <https://doi.org/10.57152/malcom.v5i2.1852>
- Khoirul Insan, M. K., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan

- Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 478–483. <https://doi.org/10.36040/jati.v7i1.6373>
- Kurniawan, B., Ari Aldino, A., & Rahman Isnain, A. (2022). Sentimen Analisis terhadap Kebijakan Penyelenggara Sistem Elektronik (PSE) Menggunakan Algoritma Bidirectional Encoder Representations from Transformers (Bert). *Jurnal Teknologi Dan Sistem Informasi*, 3(4), 98–106. <http://jim.teknokrat.ac.id/index.php/JTSI>
- Lilis Nurhidayah1\*, D. P. (2024). Al-Kharaj : Jurnal Ekonomi , Keuangan & Bisnis Syariah Al-Kharaj : Jurnal Ekonomi , Keuangan & Bisnis Syariah. *Al-Kharaj : Jurnal Ekonomi , Keuangan & Bisnis Syariah*, 6(2), 2547–2562. <https://doi.org/10.47467/alkharaj.v7i1.6535>
- M.Sayuti. (2021). Pengaruh Penggunaan *Mobile Banking* Terhadap Profitabilitas Bank Aceh Syariah Kantor Pusat Operasional Banda Aceh Disusun. *Pharmacognosy Magazine*, 75(17), 399–405.
- Manullang, O., Prianto, C., & Harani, N. H. (2023). Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest. *Jurnal Ilmiah Informatika*, 11(02), 159–169. <https://doi.org/10.33884/jif.v11i02.7987>
- Munir, A. S., Saputra, A. B., Aziz, A., & Barata, M. A. (2024). Perbandingan Akurasi Algoritma Naive Bayes dan Algoritma *Decision Tree* dalam Pengklasifikasian Penyakit Kanker Payudara. *Jurnal Ilmiah Informatika Global*, 15(1), 23–29. <https://doi.org/10.36982/jiig.v15i1.3578>
- Naseer, W. A., Sarwido, S., & Wahono, B. B. (2024). Gradient Boosting Optimization with Pruning Technique for Prediction of Bmt Al-hikmah Permata Customer Data. *Jinteks*, 6(3), 719–727.
- Nooryuda Prasetya, Y., & Winarso, D. (2021). Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19. *Jurnal Fasikom*, 11(2), 97–103. <https://ejurnal.umri.ac.id/index.php/JIK/article/view/2772/1566>
- Nurian, A. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3s1), 829–835. <https://doi.org/10.23960/jitet.v11i3s1.3348>
- Octa Nuryawan, A. T., Hasbullah, M., Rizal, M., Rajab, M. F., & Agustina, N. (2023). Algoritma *Decision Tree* Untuk Analisis Sentimen Public Terhadap Marketplace Diindonesia. *Naratif: Jurnal Nasional Riset, Aplikasi Dan Teknik Informatika*, 5(1), 18–25. <https://doi.org/10.53580/naratif.v5i1.186>
- Parasati, W., Abdurrachman Bachtiar, F., & Setiawan, N. Y. (2020). Analisis

- Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(4), 1–10. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7134>
- Prasetya, J. (2022). Penerapan Klasifikasi Naive Bayes dengan Algoritma Random Oversampling dan Random Undersampling pada Data Tidak Seimbang Cervical Cancer Risk Factors. *Leibniz: Jurnal Matematika*, 2(2), 11–22. <https://doi.org/10.59632/leibniz.v2i2.173>
- Putri, I. A., Firmansyah, A., Suprianto, A., Bangsa, U. P., Selatan, C., Bekasi, K., Barat, J., Tree, D., Analisis, S., & Informasi, T. (2025). Analisis Sentimen Peran Artificial Intelligence Terhadap Kreativitas Dan Efektivitas Mahasiswa Dalam Penyelesaian Tugas Akhir Menggunakan *Decision Tree* Berbasis SMOTE. 9(3), 4909–4916.
- Ramlan, R., Satyahadewi, N., & Andani, W. (2023). Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM. *Jambura Journal of Mathematics*, 5(2), 431–445. <https://doi.org/10.34312/jjom.v5i2.20860>
- Rasyid, M. F., Djafar, I., & Mahersatillah, A. A. (2022). Prediksi penyebaran Sub Varian omicron di Indonesia menggunakan Machine Learning. *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, XI(1), 1–7.
- Rika Widianita, D. (2023). Evaluasi Performa Synthetic Minority Oversampling Technique (SMOTE) Untuk Mengatasi Klasifikasi Data Tidak Seimbang Pada Metode *K-Nearest Neighbor* (Knn) Dan *Support Vector Machine* (SVM). *AT-TAWASSUTH: Jurnal Ekonomi Islam*, VIII(I), 1–19.
- Ritonga, S. W., . Y., Fikry, M., & Cynthia, E. P. (2023). Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode Naïve Bayes Classifier. *Building of Informatics, Technology and Science (BITS)*, 5(1). <https://doi.org/10.47065/bits.v5i1.3535>
- Salsabila, S. M., Alim Murtopo, A., & Fadhilah, N. (2022). Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier. *Jurnal Minfo Polgan*, 11(2), 30–35. <https://doi.org/10.33395/jmp.v11i2.11640>
- Sari, D. M., Fasa, M. I., & Suharto, S. (2021). Fitur-Fitur Aplikasi *Mobile Banking* Bank Syariah. *Al-Infaq: Jurnal Ekonomi Islam*, 12(2), 170.
- Sebastian Vidal. (2023). *Apa itu Google Play Store?* TECNOBITS. <https://tecnobits.com/id/apa-itu-google-playstore/>
- Sofyan, S., & Prasetyo, A. (2021). Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Data Tidak Seimbang Pada Tingkat

- Pendapatan Pekerja Informal Di Provinsi D.I. Yogyakarta Tahun 2019. *Seminar Nasional Official Statistics*, 2021(1), 868–877. <https://doi.org/10.34123/semnasoffstat.v2021i1.1081>
- Sujadi, H. (2022). Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine. *INFOTECH Journal*, 8(1), 22–27. <https://doi.org/10.31949/infotech.v8i1.1883>
- Tangkelayuk, A. (2022). The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109–1119. <https://doi.org/10.35957/jatisi.v9i2.2048>
- Teguh. (2023). *Analisis Sentimen Pada Pln Mobile Menggunakan Metode Decision Tree*. 42. <https://repositori.uma.ac.id/handle/123456789/20244>
- Vincent, R., Maulana, I., & Komarudin, O. (2024). Perbandingan Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Analisis Sentimen Dengan Multiclass Di Twitter. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2496–2505. <https://doi.org/10.36040/jati.v7i4.7152>