

**SKRIPSI**

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK ANALISIS  
SENTIMEN PENGGUNAAN APLIKASI *PERPLEXITY***

***IMPLEMENTATION OF RANDOM FOREST ALGORITHM FOR  
SENTIMENT ANALYSIS OF PERPLEXITY APPLICATION USAGE***

Diajukan untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer



**NOPRIANTY**

**D0221113**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS SULAWESI BARAT  
MAJENE  
2025**

**LEMBAR PERSETUJUAN**

**SKRIPSI**

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK ANALISIS  
SENTIMEN PENGGUNAAN APLIKASI *PERPLEXITY***

Telah dipersiapkan dan disusun oleh

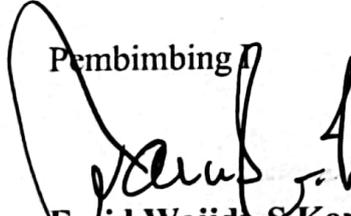
**NOPRIANTY  
NIM. D0221079**

Telah dipertahankan di depan Tim Penguji

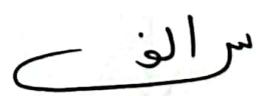
Pada tanggal 14 Mei 2025

Susunan Tim Penguji

Pembimbing I

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

Penguji I

  
**Dr. Eng. Sulfayanti, S.Si., M.T**  
NIP: 198903172020122011

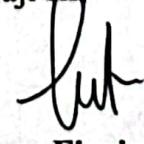
Pembimbing II

  
**Nurhikma Arifin, S.Kom., M.T**  
NIP: 199304252022032011

Penguji II

  
**Arnita Irianti, S.Si., M.Si**  
NIP: 198708062018032001

Penguji III

  
**Wawan Firgiawan, S.T., M.Kom.**  
NIDK. 8948080023

**LEMBAR PENGESAHAN**

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK ANALISIS  
SENTIMEN PENGGUNAAN APLIKASI *PERPLEXITY***

**SKRIPSI**

Untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer

Disusun oleh:

**NOPRIANTY  
NIM. D0221113**

Skripsi ini telah diuji dan dinyatakan lulus  
Pada 14 / 05/ 2025

Telah diperiksa dan disetujui oleh:

Pembimbing I

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

Pembimbing II

Nurhidaya Arifin, S.Kom., M.T  
NIP: 199304252022032011

Dekan Fakultas Teknik,  
Universitas Sulawesi Barat

Dria Hafsah Nirwana., M.T  
NIP: 196404051990032002

Ketua Program Studi  
Informatika,

Muh. Rully Rasyid, S.Kom., M.T  
NIP: 198808182022031006

## PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi 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 disitasi dalam naskah ini dan disebutkan dalam daftar referensi.

Apabila ternyata didalam 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, 14 Mei 2025



Noprianty

NIM: D0221113

## ABSTRAK

**Noprianty.** Implementasi Algoritma *Random Forest* untuk analisis sentimen penggunaan aplikasi *perplexity*. (Dibimbing oleh **Farid Wajidi** dan **Nurhikma Arifin**).

Penggunaan aplikasi *Perplexity* telah berkembang pesat, namun kajian mengenai sentimen pengguna masih terbatas. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi *Perplexity* dengan menerapkan algoritma *Random Forest*. Data sebanyak 1.400 ulasan dikumpulkan melalui teknik *web scraping* dari *Google Play Store*. Selanjutnya, data melalui tahap *preprocessing* untuk meningkatkan kualitas teks. Untuk menangani ketidakseimbangan kelas, diterapkan teknik *Random Oversampling* yang menyeimbangkan distribusi kelas. Hasil penelitian menunjukkan bahwa tanpa *Random Oversampling*, performa model terbaik diperoleh pada  $N\_estimator = 20$  dengan akurasi 91.86%. Dengan *Random Oversampling*, akurasi meningkat menjadi 96.68% pada  $N\_estimator = 30$ , disertai peningkatan presisi 96.76%, *recall* 96.68%, dan *F1-score* 96.68%. Konfigurasi model terbaik mencakup  $Max\_features = 0.8$ , *criterion = Gini*, serta *Max\_Depth* yang tidak dibatasi untuk pertumbuhan pohon yang optimal. Pengujian menggunakan *K-Fold Cross Validation* (K=5) memastikan kestabilan model. Secara keseluruhan, penerapan *Random Oversampling* terbukti efektif dalam meningkatkan akurasi dan keseimbangan model *Random Forest*, sehingga menghasilkan performa klasifikasi sentimen yang lebih baik.

**Kata kunci:** *Random Forest*, *Random Oversampling*, Klasifikasi *Perplexity*.

## **ABSTRACT**

**Noprianty.** *Implementation of Random Forest Algorithm for sentiment analysis of perplexity application usage. (Supervised by Farid Wajidi and Nurhikma Arifin).*

*The use of the Perplexity application has grown rapidly, but studies on user sentiment are still limited. This study aims to analyze user sentiment towards the Perplexity application by applying the Random Forest algorithm. Data of 1,400 reviews were collected through web scraping techniques from the Google Play Store. Furthermore, the data went through a preprocessing stage to improve text quality. To handle class synchronization, the Random Oversampling technique was applied to balance the class distribution. The results showed that without Random Oversampling, the best model performance was obtained at  $N\_estimator = 20$  with an accuracy of 91.86%. With Random Oversampling, the accuracy increased to 96.68% at  $N\_estimator = 30$ , accompanied by an increase in precision of 96.76%, recall of 96.68%, and F1-score of 96.68%. The best model configuration includes  $Max\_features = 0.8$ ,  $criterion = Gini$ , and unlimited  $Max\_Depth$  for optimal tree growth. Testing using K-Fold Cross Validation ( $K = 5$ ) ensures model stability. Overall, the application of Random Oversampling proved effective in improving the accuracy and balance of the Random Forest model, resulting in better sentiment classification performance.*

**Keywords:** *Random Forest, Random Oversampling, Perplexity Classification.*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi dalam bidang kecerdasan buatan *AI (Artificial Intelligence)*, terus membawa perubahan besar dalam berbagai aspek kehidupan manusia, terutama dalam interaksi manusia dengan mesin. Dengan terus berkembangnya teknologi sangat memudahkan manusia untuk melakukan aktivitas dari berbagai bidang, baik dalam bidang keamanan, kesehatan, transportasi, hiburan, komunikasi, pembelajaran dan sebagainya. Dengan adanya bantuan teknologi akan meminimalkan kesalahan yang dibuat oleh manusia saat bekerja atau melakukan aktivitas keseharian. Dengan begitu dapat disimpulkan bahwa teknologi memiliki peran penting dalam kehidupan manusia. Salah satu inovasi *AI* yang mendapat perhatian luas adalah *Perplexity*. *Perplexity* adalah sebuah mesin pencari dan chatbot yang menggunakan teknologi (*AI*) untuk menyajikan informasi kepada pengguna. Mesin ini memanfaatkan pemrosesan bahasa alami (*Natural Language Processing* atau *NLP*) untuk memberikan jawaban yang akurat dan komprehensif atas pertanyaan pengguna (DetikInet, 2023). Dengan kemampuan untuk memberikan referensi ke sumber asli, *Perplexity* memungkinkan pengguna untuk memverifikasi informasi dan memperdalam pemahaman. Alat ini menawarkan cara pencarian yang lebih efisien dibandingkan mesin pencari tradisional, memfasilitasi interaksi percakapan yang natural, dan dalam beberapa versi, mampu memproses informasi dalam format teks dan gambar. Dengan semua kemampuan ini, *Perplexity* menjadi sumber yang berharga untuk riset, belajar, dan pencarian informasi yang lebih interaktif dan mudah dipahami.

Pada penelitian yang dilakukan (Kirana et al., 2023) *perplexity* memiliki peningkatan, penggunaan *perplexity* telah memicu berbagai tanggapan dan sentimen dari pengguna, baik yang positif maupun yang negatif, terkait pengalaman mereka dalam menggunakan teknologi ini. Sentimen pengguna merupakan salah satu indikator penting dalam menilai keberhasilan dan dampak pada suatu produk teknologi. Dalam hal ini analisis sentimen menjadi alat yang

efektif untuk menggali opini pengguna dari teks ulasan atau percakapan. Analisis sentimen merupakan proses untuk mengidentifikasi dan mengevaluasi perasaan, emosi, serta opini yang disampaikan seseorang melalui teks. Bidang ini juga berkaitan dengan studi yang bertujuan memahami pendapat, evaluasi, sikap, dan emosi individu yang dituangkan dalam bentuk tulisan. Secara umum, analisis sentimen dilakukan dengan mengelompokkan teks, baik dalam bentuk dokumen, kalimat, maupun opini, untuk menentukan apakah muatan dari teks tersebut bersifat positif atau negatif (Aldean et al., 2022).

Dalam penelitian yang membahas tentang Komparasi Algoritma *Support Vector Machine* dan *Random Forest* pada analisis sentimen Metaverse Secara keseluruhan memberikan pemahaman tentang komparasi kinerja algoritma *SVM* dan *Random Forest*, dalam eksperimen ini *Random Forest*, terbukti unggul dengan nilai akurasi 91% sedangkan *SVM* mencapai 90% (Sari et al., 2024).

Penelitian selanjutnya Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naive Bayes*, *Random Forest*, Dan *Support Vector Machine*. Pada pengujian tersebut menggunakan beberapa model algoritma klasifikasi untuk menentukan nilai akurasi tertinggi tiap model serta menghasilkan *AUC (Area Under Curve)* untuk mengetahui hasil akurasi, dataset yang sudah di *scraping* akan melalui tahapan *text preprocessing* setelah data diolah masing-masing pengujian menggunakan metode *Cross Validation* dengan nilai *K Fold* 10 untuk mengetahui performa yang terbaik dari masing-masing algoritma. Adapun hasil menunjukkan bahwa pada pengujian diatas didapat nilai tertinggi pada pengujian algoritma *Random Forest*, dengan jumlah akurasi sebesar 97,16% serta nilai *AUC* 0,996 kemudian disusul oleh algoritma *SVM* yang menghasilkan akurasi sebesar 96,01% dengan nilai *AUC* 0,543, dan nilai terendah pada pengujian algoritma ini adalah metode *naive bayes* dengan nilai akurasi sebesar 94,16% nilai *AUC* 0,999. Dengan demikian dapat disimpulkan bahwa hasil performa dari pengujian di atas menunjukkan *Random Forest* menghasilkan nilai akurasi tertinggi dari kedua algoritma lainnya (Fitri et al., 2020).

Berdasarkan latar belakang yang telah diuraikan, penulis mengangkat penelitian dengan judul “**Implementasi Algoritma *Random Forest* Untuk**

**Penggunaan Aplikasi *Perplexity***” Penggunaan aplikasi *Perplexity* telah mengalami peningkatan yang signifikan, namun belum banyak kajian yang dilakukan untuk memahami bagaimana pengalaman pengguna diekspresikan dalam bentuk sentimen, baik positif maupun negatif. Masalah ini menjadi penting, mengingat pemahaman mendalam mengenai sentimen pengguna dapat membantu pengembang meningkatkan performa dan kepuasan pengguna terhadap aplikasi tersebut. Oleh karena itu, penelitian ini bertujuan untuk menyelesaikan masalah tersebut dengan menerapkan metode *Random Forest*. Pada penelitian ini penulis memilih menggunakan metode *Random Forest* karena berdasarkan penelitian sebelumnya, metode ini terbukti memiliki tingkat akurasi yang tinggi dalam analisis data. *Random Forest*, sebagai algoritma *ensemble*, mampu mengatasi permasalahan *overfitting* dan menghasilkan prediksi yang lebih *robust*. Oleh karena itu, metode ini dipandang sebagai pilihan yang tepat untuk mencapai hasil analisis yang lebih akurat dan andal dalam konteks penelitian ini.

## 1.2 Rumusan Masalah

Rumusan masalah ini berfokus pada penggunaan metode *Random Forest* untuk klasifikasi sentimen ulasan pengguna *Perplexity*. Fokus utama adalah untuk menilai performa *Random Forest* dalam menganalisis sentimen serta mendapatkan gambaran sentimen positif, dan negatif dari pengguna *Perplexity* berdasarkan data dari *Google Play Store*.

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disusun, tujuan penelitian ini adalah menilai performa model *Random Forest* dalam klasifikasi sentimen ulasan pengguna *Perplexity*, untuk mengetahui hasil klasifikasi menggunakan algoritma *Random Forest*.

## 1.4 Batasan masalah

1. Data sumber penelitian ini hanya menggunakan data ulasan pengguna *perplexity* yang diambil dari *Google Play Store*. Data ulasan dari platform lain atau media sosial lain tidak akan disertakan.

2. Metode klasifikasi penelitian berfokus pada metode *Random Forest* untuk analisis sentimen.
3. Jenis sentimen analisis dibatasi pada dua kategori utama, yaitu positif dan negatif.
4. Menggunakan bahasa pemrograman *python*.

### **1.5 Manfaat**

Manfaat penelitian ini adalah memberikan wawasan mendalam terkait sentimen pengguna terhadap aplikasi *Perplexity*, sehingga dapat menjadi acuan bagi pengembang aplikasi untuk memahami umpan balik pengguna secara lebih akurat. Dengan menggunakan algoritma *Random Forest*, hasil analisis sentimen diharapkan dapat membantu meningkatkan kualitas layanan aplikasi. Selain itu, penelitian ini juga berkontribusi dalam pengembangan metode analisis sentimen yang dapat diaplikasikan pada platform lain di masa mendatang.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Data Mining**

Data *Mining* merupakan sebuah cara atau sebuah proses yang dilakukan untuk mendapatkan sebuah informasi dari data yang besar. Hasil data yang didapatkan akan digunakan untuk pengambilan sebuah keputusan jika diperlukan suatu saat. Adapun tahapan dari proses data *mining* yaitu *selection* yang dimana Di lakukan untuk pemilihan data dari serangkaian data yang ada, lalu dilakukantahapan *preprocessing* untuk membersihkan data seperti data duplikat, perbaikan data dan proses lainnya. setelah data diproses maka dilakukan *transformation* pada data dengan menggunakan coding. masuk pada proses data *mining* yang mencari pola pada data menggunakan metode atau teknik tertentu, kemudian tahapan terakhir yaitu *interpretation/evaluation* proses memahami pola atau informasi untuk dilakukan *iterasi*. (Larasati et al., 2022). *Data Mining* melibatkan data terstruktur. (Fauziyyah et al., 2020). Menemukan pola atau informasi penting dalam kumpulan data yang dipilih, menggunakan teknik, metode, atau algoritma tertentu yang sesuai dengan tujuan yang telah ditetapkan (Herjanto et al, 2024). *Data Mining* dapat dikatakan sebuah proses untuk memperoleh informasi berharga dari kumpulan data besar yang dapat digunakan dalam pengambilan keputusan di masa depan. Proses ini melibatkan beberapa tahapan, seperti pemilihan data (*selection*), pemrosesan data (*preprocessing*), transformasi data, pencarian pola dengan teknik tertentu (*data mining*), dan evaluasi hasil (*interpretation/evaluation*). *Data Mining* umumnya bekerja dengan data terstruktur dan menggunakan berbagai metode atau algoritma untuk menemukan pola yang relevan sesuai dengan tujuan yang telah ditetapkan.

#### **2.2 Google Play Store**

*Google Play Store* merupakan sebuah aplikasi khusus yang hanya ada pada sistem android yang di kembangkan oleh *google* yang digunakan oleh pengguna untuk mengunduh aplikasi (Larasati et al., 2022).Ulasan pengguna aplikasi pada *platform google play store* telah menjadi sumber data atau informasi yang populer, dengan begitu *Google Play store* menjadi sumber informasi yang

signifikan dalam melakukan analisis sentimen berdasarkan ulasan pengguna. (Huda et al., 2023). *Google Play Store* tidak hanya berfungsi sebagai tempat bagi pengguna untuk mengunduh aplikasi, tetapi juga sebagai sumber data yang signifikan sehingga ulasan pengguna di platform ini telah menjadi sumber informasi yang populer dan penting, terutama untuk keperluan analisis sentimen, karena ulasan-ulasan tersebut mencerminkan opini pengguna terhadap aplikasi yang digunakan, sehingga pada penelitian ini *google play store* menjadi platform pengambilan data ulasan aplikasi *Perplexity*.

### 2.3 Perplexity

*Perplexity* merupakan sebuah platform yang dirancang khusus untuk membantu pengguna dalam mencari jawaban atas pertanyaan mereka melalui teknologi *AI*. Platform ini menggunakan algoritma canggih untuk mengidentifikasi dan memberikan jawaban terbaik berdasarkan konteks pertanyaan kelebihan utama *Perplexity* pada kemampuannya untuk memahami dan respon pertanyaan dengan cara yang lebih intuitif dan relevan bagi penggunanya (Syafiuddin et al., 2024).

*Perplexity* sebagai alat yang mempercepat penyusunan artikel dan melakukan parafrase guna menghindari plagiarisme (Rofiki et al., 2024). *Perplexity* platform *AI* yang efektif dalam membantu pengguna menemukan jawaban yang relevan atas pertanyaan mereka, dengan kemampuan pemahaman yang intuitif. Selain itu, platform ini juga bermanfaat dalam mempercepat penyusunan artikel dan mencegah plagiarisme melalui parafrase.

### 2.4 Text Mining

*Text Mining* adalah proses mengekstraksi informasi dari data sumber untuk dianalisis dan mengelompokkan informasi berdasarkan kata-kata, sehingga hubungan dengan data sumber lainnya dapat diidentifikasi. Adapun tugas khusus *text mining* adalah melakukan pengkategorian teks (*text categorization*) atau pengelompokan teks (*text clustering*). teks yang sudah melalui tahapan ini akan dipecah kalimat menjadi kata (*Tokenization*) lalu huruf besar menjadi huruf kecil (*case folding*) dilakukan perubahan kata yang berimbuhan menjadi kata dasar (*stemming*) perhitungan dan pengelompokan data (*filtering*) penerapan *text mining*

biasanya pada analisis sentimen, preferensi pelanggan, penyaringan, spam dan lainnya (Larasati et al., 2022). *Text Mining* bertujuan untuk mengekstraksi serta mengidentifikasi sebuah informasi yang berguna untuk memenuhi kebutuhan analisis yang dilakukan (Sari et al., 2024). *Text Mining* berkaitan dengan fitur-fitur tertentu yang perlu untuk di *preprocessing*, *text Mining* berkaitan dengan *Natural Language Processing (NLP)*. Memiliki tujuan yang sama dengan data mining, yaitu untuk mendapatkan informasi dan pengetahuan yang bernilai dari sejumlah besar data. Namun, terdapat perbedaan di antara keduanya, terutama dalam hal jenis data yang digunakan (Huda et al., 2023).

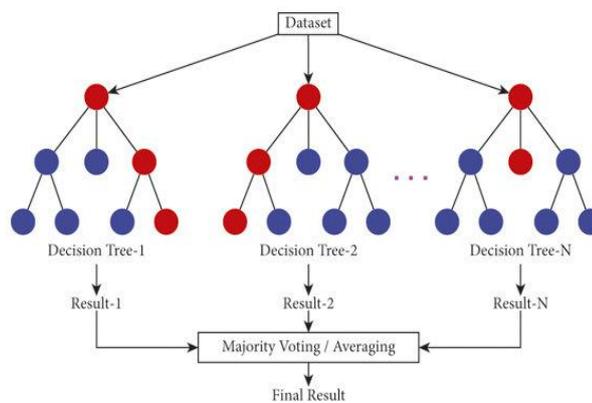
*Text Mining* menjadi proses penting dalam ekstraksi dan analisis informasi dari sumber data, yang melibatkan berbagai tahapan seperti pengkategorian, *tokenization*, *case folding*, dan *stemming*. Meskipun memiliki tujuan yang serupa dengan data mining dalam hal memperoleh informasi berharga, *text mining* lebih fokus pada data tidak terstruktur dan berhubungan erat dengan *Natural Language Processing (NLP)*. Penerapannya mencakup analisis sentimen dan penyaringan data, menunjukkan kemampuannya untuk mendukung berbagai kebutuhan analisis.

## 2.5 Random Forest

*Random Forest* merupakan metode yang digunakan dalam mengklasifikasikan dan regresi dengan menggunakan pohon keputusan sebagai *base classifier* yang dibangun lalu dikombinasikan. Beberapa aspek penting dalam metode ini diantaranya melakukan *bootstrap sampling* untuk membangun pohon prediksi, masing-masing pohon keputusan memprediksi dengan prediktor acak, dan melakukan prediksi dengan cara *majority vote* untuk klasifikasi dan rata-rata untuk regresi (Fitri et al., 2020). Pohon keputusan dengan suara terbanyak akan menjadi penentu dalam pemilihan kelas pada klasifikasi (Huda et al., 2023). Adapun langkah-langkah dalam membuat *decision tree* atau pohon keputusan (Larasati et al., 2022):

Berikut ini langkah-langkah dalam membuat *Decision tree* atau pohon keputusan: (Larasati et al., 2022).

- Menghitung nilai *gini impurity* pada setiap fitur.
- Menghitung nilai *gini impurity* dari setiap kondisi *split*.
- Menghitung nilai *average gini impurity* dari setiap kondisi *split*.
- Menghitung nilai *information gain* dari setiap kondisi *split*.
- Mencari nilai *Information gain* tertinggi untuk dijadikan *root*.
- Mengulang langkah 2-4 sampai k tree



**Gambar 2. 1** Proses Algoritma *Random Forest*

Berikut cara kerja *Random Forest*(Sumber : Researchgate.net) :

- Pemilihan data secara acak:

Pemilihan data secara acak ini menggunakan *bootstrapping*. *Bootstrapping* adalah teknik sampling dengan penggantian yang digunakan untuk membentuk dataset baru dari dataset asli. Proses ini dilakukan dengan mengambil sampel secara acak sebanyak ukuran dataset asli, di mana beberapa data dapat terpilih lebih dari sekali.

- Membangun pohon keputusan

Pilih fitur *subset* sejumlah fitur secara acak, biasanya lebih sedikit dari total fitur. Jumlahnya dapat diatur sebagai "auto", "sqrt", "log2", atau angka tertentu. Kemudian membagi data menggunakan fitur terbaik yang telah dipilih untuk membagi data pada *node* saat ini menjadi *subset* yang lebih kecil. Setiap *subset* akan menjadi cabang dari *node* saat ini dalam pohon keputusan.

- c. Ulangi langkah a-b sebanyak  $n$  kali untuk mendapatkan  $n$  pohon keputusan.
- d. Prediksi dengan menggunakan *ensemble*.

Setelah semua pohon keputusan dibangun, prediksi dilakukan dengan cara melakukan *voting* (menggunakan mayoritas suara untuk klasifikasi dan rata-rata untuk regresi).

## 2.6 Web Scraping

*Web Scraping* merupakan sebuah metode pengambilan data atau informasi secara otomatis dari situs *web* dengan cara diekstraksi (Larasati et al., 2022). Penggunaan *web scraping* digunakan pada berbagai pekerjaan yang berkaitan dengan data (Huda et al., 2023) Tahapan umum *web scraping* meliputi sebagai berikut (Larasati et al., 2022):

- a. Meminta *HTTP* mengambil data yang di inginkan.
- b. Selanjutnya akan diproses oleh server dengan format URL berisi sebuah permintaan *GET/HTTP* yang berisi *POST*.
- c. Setelah diterima data yang diminta akan diambil lalu dikirim kembali dengan format yang beragam.

Pada penelitian ini penulis akan menggunakan *Google Play Scraper* dalam pengambilan data.

## 2.7 Analisis Sentimen

Analisis sentimen adalah sebuah proses menganalisis sebuah data sentimen atau opini dari data yang telah dikumpulkan dari berbagai sumber seperti ulasan pada sebuah aplikasi, produk, ataupun komentar pada laman sosial media. Adapun tujuan dari analisis sentimen adalah memahami pandangan atau perasaan pada sebuah teks apakah itu positif atau negatif (Sari et al., 2024). Selain itu analisis sentimen sering juga disebut *Opinion Mining* yang bertujuan untuk menganalisa pendapat, sentimen, evaluasi, sikap, penilaian, emosi seorang, subjektifitas, penilaian, atau pandangan yang terdapat dalam suatu teks apakah pembicara atau penulis berkaitan dengan suatu topik, produk layanan ataupun kegiatan tertentu (Basar et al., 2022). Data yang sudah dikumpulkan akan dianalisis agar mendapatkan pandangan sehingga dapat menghasilkan sebuah kesimpulan dengan hasil sentimen positif dan negatif (Larasati et al.,

2022). Adapun langkah analisis sentimen yang dapat diterapkan sebagai berikut (Fauziyyah et al., 2020):

- 1) Mencari kalimat sentimen negatif dan positif
- 2) Menganalisa sentimen seluruh dokumen sebagai negatif dan positif.
- 3) Pengelompokan level dimana semua atribut yang memiliki sentimen sama
- 4) Pengguna memanfaatkan visualisasi data pada analisis sentimen untuk berinteraksi dengan pengguna *social*.

Analisis sentimen ini berfokus pada pendapat seseorang pada aplikasi *Perplexity* yang mengungkapkan atau menyiratkan sentimen positif atau negatif. Sebagian besar analisis sentimen berkaitan dengan orang-orang di media sosial. Selain polaritas positif dan negatif, terkadang polaritas dianggap sebagai spektrum, di mana sebuah dokumen dapat berisi pernyataan dengan polaritas campuran.

## **2.8 Text preprocessing**

*Text preprocessing* merupakan langkah untuk membersihkan teks agar berkualitas dan sesuai dengan persyaratan yang dibutuhkan untuk diolah oleh suatu algoritma. *Text Preprocessing* digunakan untuk mengubah data yang tidak terstruktur (Alita et al., 2023) (Larasati et al., 2022). Ada beberapa tahapan *text preprocessing* yaitu :

- 1) *Case folding* mengonversi teks dalam dokumen data menjadi *lowercase*.
- 2) *Tokenizing* memecah teks menjadi per kata.
- 3) *Stop removal* menghilangkan kata tidak penting.
- 4) *Stemming* penghapusan imbuhan dan akhiran

## **2.9 Pelabelan(IndoBERT)**

IndoBERT merupakan model *pretrained* yang diperkenalkan oleh dua tim riset independen, yaitu tim IndoLEM dan tim IndoNLU. Model ini merupakan adaptasi dari BERT yang dirancang khusus untuk Bahasa Indonesia. IndoBERT dilatih menggunakan korpus berbahasa Indonesia yang sangat besar, sekitar 4 miliar kata, yang dikumpulkan dari berbagai sumber seperti berita daring, media sosial, Wikipedia, artikel online, subtitle video, dan *parallel dataset*. Korpus ini

kemudian dikenal sebagai Indo4B (Jayadianti et al., 2022). IndoBERT bekerja dengan memanfaatkan arsitektur transformer bidirectional untuk memahami konteks kata dalam kalimat secara menyeluruh. Dalam tugas pelabelan seperti Named Entity Recognition (NER) atau analisis sentimen, IndoBERT terlebih dahulu memecah teks menjadi token sub-kata, lalu mengubah token tersebut menjadi representasi vektor melalui embedding. Representasi ini diproses oleh lapisan encoder untuk menangkap hubungan antar-token dalam dua arah (kiri dan kanan). Hasil pemrosesan ini digunakan untuk memberikan label pada setiap token (token-level) atau pada keseluruhan teks (sequence-level), sesuai dengan jenis tugas. Dengan proses fine-tuning menggunakan data berbahasa Indonesia, IndoBERT dapat melakukan pelabelan secara akurat dan kontekstual.

### 2.10 *Fitur Extraction (TF IDF)*

*Feature Extraction* merupakan salah satu teknik dalam mengurangi kompleksitas data dimana data besar diubah menjadi fitur-fitur yang lebih sederhana tanpa mengurangi kinerja model. Kata yang sering muncul dalam dokumen dibandingkan dengan keseluruhan jumlah kata pada dokumen (*TF*) dan *inverse document frequency (IDF)* digunakan (Sari et al., 2024). Algoritma *TF-IDF* digunakan untuk memberi bobot pada kata kunci di setiap dokumen dan menghitung frekuensi kemunculannya dalam setiap dokumen (Jepi Suprianto et al., 2023). Rumus *TF-IDF* pada persamaan 2.1, 2.2, dan 2.3 berikut (Larasati et al., 2022)

$$TF(t, d) = \frac{f_{t,d}}{N_d} \quad (2.1)$$

$f_{t,d}$  = jumlah kemunculan term  $t$  dalam dokumen  $d$

$N_d$  = total jumlah kata dalam dokumen

$$IDF(t) = \log\left(\frac{N}{n_t}\right) \quad (2.2)$$

$N$  = jumlah total dokumen

$n_t$  = jumlah dokumen yang mengandung term  $t$

$$IDF(t, d) = TF(t, d).IDF(t) \quad (2.3)$$

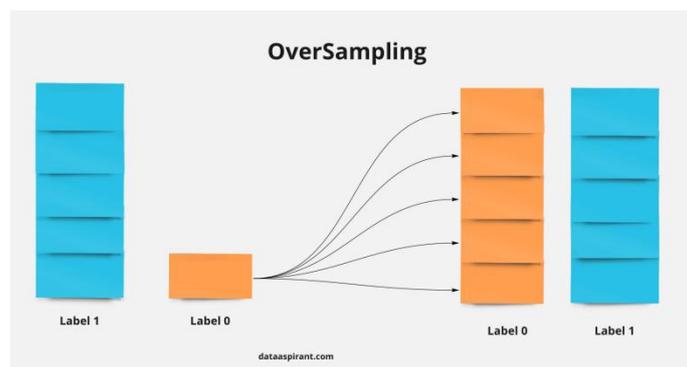
Untuk mendapatkan hasil *TF-IDF*, hasil *TF* dikalikan dengan hasil *IDF*.

### 2.11 *K-Fold Cross Validation*

*K-Fold* merupakan teknik dalam data mining yang digunakan untuk memperoleh akurasi optimal dengan membagi dataset menjadi beberapa subset data uji dan latih secara bergantian. (Herjanto et al., 2024). Teknik ini membagi dataset menjadi sejumlah partisi acak sebanyak nilai  $K$  yang telah ditentukan. Pada setiap iterasi, satu partisi digunakan sebagai data uji, sementara sisanya berfungsi sebagai data latih. Proses ini diulangi sebanyak  $K$  kali hingga seluruh partisi telah digunakan sesuai dengan jumlah  $K$  yang diterapkan dalam penelitian (Asri et al., 2021). Metode ini sering digunakan oleh para peneliti karena terbukti mampu mengurangi bias dalam proses pengambilan sampel. *K-fold cross validation* secara berkelanjutan membagi data menjadi data latih (*train*) dan data uji (*test*), sehingga setiap data memiliki kesempatan untuk menjadi data uji.  $K$  merujuk pada jumlah pembagian data yang digunakan untuk memisahkan data latih dan uji.

### 2.12 *Ketidakseimbangan Data (Data Imbalance)*

Proses klasifikasi pada *Dataset* yang memiliki ketidakseimbangan kelas merupakan tantangan utama dalam bidang *machine learning* dan data mining. Ketidakseimbangan kelas, atau *imbalanced data*, terjadi ketika distribusi kelas dalam *Dataset* tidak seimbang, dengan jumlah data *mayoritas* (*positif*) lebih banyak daripada jumlah data *minoritas* (*negatif*). Ketidakseimbangan ini dapat menyebabkan masalah *misclassification*, di mana *classifier* cenderung memprediksi kelas *mayoritas* dan menganggap data *minoritas* sebagai *noise* atau *outlier*, yang dapat mengurangi kinerja *classifier*. Untuk mengatasi masalah ini dilakukan teknik *Random Oversampling* dan . di mana data dari kelas *minoritas* ditambahkan ke dalam data *training* secara acak sampai jumlah data kelas *minoritas* sama dengan jumlah kelas *mayoritas*. Proses ini dimulai dengan menghitung selisih antara jumlah kelas *mayoritas* dan kelas *minoritas*, lalu dilakukan perulangan untuk menambahkan data kelas *minoritas* secara acak ke dalam data *training* sampai jumlahnya seimbang dengan kelas *mayoritas* (Ferdita Nugraha et al., 2022). Ilustrasi proses *Random Oversampling* dapat dilihat pada Gambar 2.2 berikut.



**Gambar 2. 2** Ilustrasi Proses *Random Oversampling*

Gambar 2.2 diatas menggambarkan konsep *oversampling* dalam penanganan ketidakseimbangan kelas pada dataset. Pada sisi kiri, terlihat Label 1 (ditandai dengan blok biru) yang jumlahnya lebih banyak dibandingkan Label 0 (blok *orange*). Dalam teknik *oversampling*, Label 0 akan diperbanyak dengan cara menyalin contoh yang ada, sehingga jumlahnya menjadi lebih seimbang dengan Label 1. Proses ini ditunjukkan di tengah gambar, di mana blok *orange* yang sedikit akan digandakan hingga jumlahnya mirip dengan blok biru. Hasil akhirnya, seperti yang terlihat di sisi kanan, adalah distribusi yang lebih seimbang antara kedua label, Label 1 dan Label 0. Teknik *oversampling* ini membantu model untuk belajar dengan lebih baik dan mengurangi bias yang disebabkan oleh ketidakseimbangan kelas dalam data.

### 2.13 *Confusion matriks*

Evaluasi klasifikasi dalam penelitian ini dilakukan dengan menggunakan metode *Confusion Matrix*. Metode ini bekerja dengan membandingkan matriks prediksi dengan kelas asli, yang berisi data sebenarnya dan prediksi nilai klasifikasi. Setelah sistem berhasil mengklasifikasikan *tweet*, diperlukan ukuran untuk mengukur sejauh mana akurasi atau ketepatan klasifikasi yang dihasilkan oleh sistem (Styawati et al., 2021). *Confusion Matrix* adalah metode untuk menggambarkan hasil akurasi dari model yang telah dibuat. Matriks ini meringkas kinerja model dalam mengklasifikasikan data ke dalam berbagai kategori sesuai dengan nilai sebenarnya dari kelas objek yang diprediksi. Berikut ini *confusion matrix* dengan dua kelas sentimen seperti pada Tabel 2.1 (Aldean et al., 2022):

**Tabel 2. 1** *Confusion matriks*

<i>classification</i>		<i>Actual</i>	
		<i>TRUE</i>	<i>FALSE</i>
<i>prediction</i>	<i>TRUE</i>	<i>True Positif (TP)</i>	<i>False Positif(FP)</i>
	<i>FALSE</i>	<i>False Negatif(FN)</i>	<i>True Negatif (TN)</i>

Penjelasannya:

- True Positif (TP)* = memiliki arti bahwa banyak data yang aktual kelasnya positif, kemudian model juga memprediksi positif
- True Negatif (TN)* = memiliki arti bahwa banyak data yang aktual kelasnya negative dan model juga memprediksi negatif
- False Positif (FP)* = memiliki arti banyak data yang aktual kelasnya negatif akan tetapi model memprediksi positif
- False Negatif (FN)* = memiliki arti banyak data yang aktual kelasnya positif, akan tetapi model memprediksi negatif

Penjelasan Rumus sebagai berikut (Aldean et al., 2022):

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.5)$$

$$Presisi = \frac{True\ positif\ (TP)}{True\ positif\ (TP)+False\ Positif\ (FP)} \quad (2.6)$$

$$Recall = \frac{True\ Positif\ (TP)}{True\ positif\ (TP)+False\ negatif\ (FN)} \quad (2.7)$$

$$F1\ score = 2 * \frac{Recal*Precision}{Recal+Precision} \quad (2.8)$$

## 2.14 Penelitian Sebelumnya

Adapun penelitian terdahulu yang relevan dengan topik penelitian ini dapat dilihat pada Tabel 2.2. Beberapa penelitian sebelumnya telah menggunakan algoritma *Random Forest* untuk analisis sentimen di berbagai bidang, termasuk aplikasi keuangan, kesehatan, serta platform media sosial. Namun, penelitian-penelitian tersebut umumnya berfokus pada objek analisis yang berbeda.

**Tabel 2. 2** Penelitian Sebelumnya

NO	Judul Penelitian	Penulis	Hasil	Perbedaan Penelitian
1.	<i>Comparative Analysis Of Phishing Website Prediction Classification Algorithm Using Logistic Regression, Decision Tree, And Random Forest (2022).</i>	Fanka Angelina Larasati et al., 2022	Akurasi 84%, <i>precision</i> 84%, recall 84%, F1-Score 84%, kedalaman tree 65, jumlah tree 400.	Penelitian ini menganalisis sentimen aplikasi <i>Perplexity</i> dengan objek pengguna aplikasi ini dan berfokus pada pengembangan model evaluasi untuk aplikasi <i>AI</i> , bukan aplikasi keuangan
2.	Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan <i>Random Forest</i>	Muhammad Yusril Aldean et al., 2022	Akurasi 79%, <i>Precision</i> 85%, <i>recall</i> 90%, F1-Score 88%.	Penelitian ini menggunakan ulasan aplikasi <i>Perplexity</i> dari platform berbeda ( <i>Google Play Store</i> ), dengan tujuan analisis sentimen pada aplikasi <i>AI</i> daripada

	<i>Classifier (Sinovac).</i>			topik kesehatan.
3.	Analisis Sentimen Ulasan Pengguna Aplikasi SIREKAP di Play Store Menggunakan Algoritma <i>Random Forest Classifier</i>	Muhamad Fajar Yudhistira Herjanto et al., 2024	Akurasi 74%, <i>precision</i> 75%, <i>Recall</i> 74, <i>F1-Score</i> 74%.	algoritma dalam aplikasi <i>AI</i> yang berbeda konteks, yaitu aplikasi pendidikan berbasis <i>AI</i> .
4.	Analisis Sentimen Terhadap Ulasan Film Menggunakan Algoritma <i>Random Forest</i>	Muhammad Asjad Adna Jihad et al.	Akurasi 75.76%, menggunakan <i>skip-gram Word2Vec</i> 300 dimensi dan Adaptive Boosting untuk model.	Penelitian ini tidak menggunakan teknik <i>Word2Vec</i> atau <i>Adaptive Boosting</i> tetapi berfokus pada parameter <i>tuning Random Forest</i> untuk aplikasi <i>AI</i> .
5.	Analisis Komparasi Penggunaan Teknologi Kecerdasan Buatan dalam Penyusunan Karya Ilmiah:	Hermila A. et al., 2024	Nilai untuk <i>ChatGPT</i> : 16.78, <i>Perplexity</i> : 12.69, <i>Consensus</i> : 14.7.	Penelitian ini menganalisis sentimen terhadap ulasan aplikasi <i>Perplexity</i> dari pengguna umum,

	<i>ChatGPT, Perplexity, Consensus</i>			bukan dari segmen mahasiswa saja, serta lebih fokus pada pengukuran performa algoritma <i>Random Forest</i> daripada perbandingan <i>AI</i> secara statistik.
6.	Implementasi Text Mining Terhadap Analisis Sentimen Dunia di Twitter tentang Kota Medan ( <i>K-Fold Cross Validation dan Naïve Bayes Classifier</i> )	Tengku Ridwan syah et al., 2022	Total tweet positif 80.2%, negatif 19.8%.	Penelitian ini menggunakan algoritma <i>Random Forest</i> pada ulasan pengguna aplikasi <i>Perplexity</i> , tidak berfokus pada <i>tweet</i> atau topik pariwisata, melainkan analisis kinerja aplikasi <i>AI</i> .
7.	Analisis Sentimen Pandemi Covid-19 pada Twitter dengan Text Mining Python	Anni Karimatul Fauziyyah et al., 2020	Sentimen netral paling tinggi: 58.94 % untuk Covid- 19, 55.10% untuk	Penelitian ini berfokus pada analisis ulasan aplikasi <i>AI</i> , bukan sentimen terkait pandemi, dan menggunakan algoritma <i>Random</i>

			variabel coronavi rus.	<i>Forest</i> untuk mengukur performa pada aplikasi <i>Perplexity</i> .
8.	Analisis Sentimen Aplikasi <i>E-Government</i> pada Google Play dengan Algoritma <i>Naive Bayes</i>	Artanti Inez Tanggra eni et al., 2022	Akurasi 89%, precision 83%, recall 87% pada data ulasan aplikasi Sentuh Tanahku	Penelitian ini menggunakan algoritma <i>Random Forest</i> dengan objek aplikasi <i>AI Perplexity</i> , bukan algoritma <i>Naive Bayes</i> atau aplikasi <i>e- Government</i> .
9.	Analisis Sentimen Kalimat Depresi pada Twitter dengan <i>Naive Bayes</i> , <i>SVM</i> , dan <i>Random Forest</i>	Moham mad Fachriz a et al., 2023	Akurasi tertinggi Random Forest: 83.33%. Naive Bayes: 80.5%, SVM: 80.4%.	Penelitian ini hanya berfokus pada algoritma <i>Random Forest</i> untuk sentimen aplikasi <i>Perplexity</i> di Google Play, tanpa menggunakan <i>Naive Bayes</i>
10.	Komparasi Algoritma <i>SVM</i> dan <i>Random Forest</i> untuk Analisis Sentimen Metaverse	Putri Kumala Sari et al., 2024	Random Forest (dengan SMOTE) akurasi 91%, SVM akurasi 90%.	Penelitian ini mengkaji sentimen pada aplikasi <i>Perplexity</i> menggunakan <i>Random Forest</i> tanpa metode

				<i>SMOTE</i> atau algoritma <i>SVM</i> .
11.	Penerapan N-Gram dengan Algoritma <i>Random Forest</i> dan <i>Naïve Bayes</i> pada Analisis Sentimen Kebijakan PPKM 2021	Putri Pratama Erika Indarbensyah et al., 2021	Unigram Random Forest akurasi 99.5%. <i>Naïve Bayes</i> akurasi 97.9%.	Penelitian ini hanya menggunakan <i>Random Forest</i> tanpa teknik <i>N-Gram</i> atau <i>Naïve Bayes</i> untuk analisis aplikasi AI Perplexity.
12.	Perbandingan Algoritma <i>Random Forest</i> , <i>Naïve Bayes</i> , dan <i>SVM</i> untuk Analisis Sentimen Penghapusan Tenaga Honorer	Akhmad Miftahu salam et al., 2022	<i>Random Forest</i> (dengan <i>random oversampling</i> ) akurasi tertinggi 66.67%.	Penelitian ini menggunakan <i>Random Forest</i> pada ulasan aplikasi <i>Perplexity</i> tanpa teknik <i>oversampling</i> atau <i>Naïve Bayes</i> dan <i>SVM</i> .

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan, dapat disimpulkan bahwa implementasi algoritma *Random Forest* untuk analisis sentimen pengguna aplikasi *Perplexity* menunjukkan hasil yang cukup baik. Data dikumpulkan melalui teknik *web scraping* dari *Google Play Store*, dengan total 1.400 komentar pengguna terkait aplikasi *Perplexity*, kemudian melalui tahap *preprocessing* yang mencakup *cleaning*, *stopword*, *case folding*, *stemming*, normalisasi, dan tokenisasi untuk meningkatkan kualitas data. Untuk mengatasi ketidakseimbangan kelas dalam dataset, digunakan teknik *Random Oversampling* yang terbukti mampu meningkatkan akurasi model dengan mengurangi bias terhadap kelas mayoritas dan memberikan representasi lebih seimbang untuk kelas minoritas. Berdasarkan pengujian, pemilihan parameter terbaik menunjukkan bahwa model tanpa *Random Oversampling* mencapai performa optimal dengan jumlah estimator ( $N\_estimator$ ) sebanyak 20, sedangkan model dengan *Random Oversampling* memberikan performa terbaik pada  $N\_estimator = 30$ . Kedua model menggunakan  $max\_features = 0.8$ ,  $criterion = Gini$ , dan  $max\_depth = None$ , serta diuji menggunakan teknik *Stratified K-Fold Cross-Validation* dengan 5 *fold* untuk memastikan evaluasi yang adil. Model tanpa *Random Oversampling* menghasilkan akurasi tertinggi sebesar 89.24% pada  $N\_estimator = 60$ , sedangkan model dengan *Random Oversampling* memberikan akurasi lebih tinggi sebesar 93.96% pada  $N\_estimator = 40$ . Secara keseluruhan, implementasi *Random Forest* dengan teknik *Random Oversampling* terbukti lebih efektif dalam menangani ketidakseimbangan kelas dan meningkatkan performa model dalam analisis sentimen pengguna aplikasi *Perplexity*. Oleh karena itu, penggunaan metode *balancing* seperti *Random Oversampling* sangat direkomendasikan dalam pengolahan data yang tidak seimbang agar hasil klasifikasi lebih akurat dan adil terhadap semua kelas.

## 5.2 Saran

Penelitian selanjutnya dapat memperluas sumber data dengan mengumpulkan ulasan pengguna dari berbagai platform lain, seperti *Twitter*, *Tiktok* atau *YouTube*, untuk mendapatkan perspektif sentimen yang lebih luas terhadap aplikasi *Perplexity*. Dengan menambahkan lebih banyak data dari berbagai sumber, model dapat diuji dalam lingkungan yang lebih beragam, sehingga meningkatkan generalisasi dan mengurangi potensi bias dari satu platform tertentu. Selain itu, pendekatan analisis dapat dikembangkan dengan mengombinasikan algoritma *Random Forest* dengan teknik lain yang lebih canggih untuk meningkatkan akurasi model.

## DAFTAR PUSTAKA

- Adriana, N. M. T. O., Suarjaya, I. M. A. D., & Githa, D. P. (2023). Analisis Sentimen Publik Terhadap Aksi Demonstrasi di Indonesia Menggunakan Support Vector Machine Dan Random Forest. *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 257-267.
- Aldean, M. Y., Paradise, P., & Nugraha, N. A. S. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac). *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 4(2), 64-72.
- Basar, T. F., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(3), 1426-1433.
- Buntoro, G. A. (2016). Analisis sentimen hatespeech pada twitter dengan metode naïve bayes classifier dan support vector machine. *Jurnal Dinamika Informatika*, 5(2), 1-12.
- Fauziyyah, A. K. (2020). Analisis sentimen pandemi Covid19 pada streaming Twitter dengan text mining Python. *Jurnal Ilmiah SINUS*, 18(2), 31-42.
- Fachriza, M., & Munawar, M. (2023). Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest. *KOMPUTEK*, 7(2), 49-58.
- Fitri, E. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *Jurnal Transformatika*, 18(1), 71-80.
- Firsttama, R. A., Arifiyanti, A. A., & Kartika, D. S. Y. (2024). Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 6(2), 282-285.

- Hermila, A., Pakaja, J. A., Bau, R. T. R., & Farman, I. (2024). Analisis Komparasi Penggunaan Teknologi Kecerdasan Buatan Dalam Penyusunan Karya Ilmiah Oleh Mahasiswa: ChatGPT, Perplexity Dan Consensus. *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, 7(3), 6687-6693.
- Herjanto, M. F. Y., & Carudin, C. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(2).
- Indarbensyah, P. P. E., & Rochmawati, N. (2021). Penerapan N-Gram menggunakan Algoritma Random Forest dan Naïve Bayes Classifier pada Analisis Sentimen Kebijakan PPKM 2021. *Journal of Informatics and Computer Science (JINACS)*, 2(04), 235-244.
- Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. *ILKOM Jurnal Ilmiah*, 14(3), 348-354.
- Jihad, M. A. A., Adiwijaya, A., & Astut, W. (2021). Analisis sentimen terhadap ulasan film menggunakan algoritma random forest. *eProceedings of Engineering*, 8(5).
- Kurniawan, D. A. A., Utami, E., & Al Fatta, H. (2023). Analisis Sentimen Pada Opini Pengguna Aplikasi Qasir Menggunakan Support Vector Machine Dan Random Forest. *TEKNIMEDIA: Teknologi Informasi dan Multimedia*, 4(1), 1-8.
- Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(9), 4305-4313.
- Miftahusalam, A., Nuraini, A. F., Khoirunisa, A. A., & Pratiwi, H. (2022, November). Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini

Masyarakat Terhadap Penghapusan Tenaga Honorer. In *Seminar Nasional Official Statistics* (Vol. 2022, No. 1, pp. 563-572).

Ridwansyah, T. (2022). Implementasi text mining terhadap analisis sentimen masyarakat dunia di twitter terhadap Kota Medan menggunakan k-fold cross validation dan naïve bayes classifier. *KLIK Kaji. Ilm. Inform. dan Komput*, 2(5), 178-185.

Ramadhansyah, D., Asrofiq, A., & Yunefri, Y. (2024). Analisis Sentimen Ulasan Penumpang Maskapai Penerbangan Di Indonesia Dengan Algoritma Random Forest Dan KNN. *ZONasi: Jurnal Sistem Informasi*, 6(2), 287-297.

Sari, P. K., & Suryono, R. R. (2024). Komparasi algoritma Support Vector Machine dan Random Forest untuk analisis sentimen metaverse. *Jurnal Mnemonic*, 7(1), 31-39.

Saepudin, A., Faqih, A., & Dwilestari, G. (2024). Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee. *J. Tekno Kompak*, 18(1), 178-192.

Tanggraeni, A. I., & Sitokdana, M. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), 785-795.