KOMPARASI ALGORITMA DALAM ANALISIS SENTIMEN MEDIA SOSIAL TERKAIT APPLE X HOLIDAY 50 TAHUN BERBASIS MACHINE LEARNING

SKRIPSI

Diajukan sebagai syarat untuk menyelesaikan Pendididkan
Program Sarjana (s-1)
Pada Program Studi Informatika

oleh:
Erik kurniawan
2102020106

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU TEKNIK

UNIVERSITAS BINA INSAN

LUBUKLINGGAU

2024/2025

HALAMAN PENGESAHAN SKRIPSI

KOMPARASI ALGORITMA DALAM ANALISIS SENTIMEN MEDIA SOSIAL TERKAIT APPLE TAX HOLIDAY 50 TAHUN BERBASIS MACHINE LEARNING

Oleh:

ERIK KURNIAWAN

NIM:2102020106

Lubuklinggau, Januari 2025

Pembimbing I	Pembimbing II
Elmayati M.kom	Harma Oktafia Lingga Wijaya,M.Kom

Mengesahkan

Dekan Fakultas ilmu Teknik

Universitas Bina Insan,

(Dr. Rudi Kurniawan, S.T., M.Kom)

HALAMAN PERSETUJUAN TIM PENGUJI SKRIPSI

Pada hari Sabtu tanggal 25 bula tahun 2025 telah dilaksanakan sidang Skripsi oleh Program Studi Inf akultas Ilmu Teknik Universitas Bina Insan

Nama : Erik kurniawan

Nim : 2102020106

Judul Skirpsi :KOMPARASI ALGORITMA DALAM ANALISIS SENTIMEN MEDIA SOSIAL TERKAIT APPLE TAX HOLIDAY 50 TAHUN BERBASIS MACHINE LEARNING

Komisi Penguji

1. Ketua : Elmayati,M.kom ( )

2. Sekretaris : Harma oktafia lingga wijaya,M.kom ( )

3.Anggota : Budi santoso,M.kom ( )

Mengetahui,

Kepala Program Studi informatika

Fakultas Ilmu Teknik

Universitas Bina Insan

(Budi Santoso, M.kom)

HALAMAN MOTTO DAN PERSEMBAHAN

MOTTO: - **Tidak ada mimpi yang terlalu besar, selama kamu memiliki keberanian untuk mencobanya**

- **Kesulitan adalah batu pijakan menuju kesuksesan.**
- **Tidak ada kesuksessan di tempat tidur**
- **Setiap akhir adalah awal yang baru. Jangan menyerah di tengah jalan.**

Persembahan Kepada :

- **Ayah dan ibunda tercinta, yang telah banyak mendukungku dan memberikan do'a untuk keberhasilanku**
- **Teman-teman seperjuanganku**
- **Almamaterku**
- **Dan untuk diri ku sendiri terimah kasih sudah berjuang dalam penyusunan skripsi ini**

HALAMAN PERNYATAAN

Saya yang bertanda tangan diba

Nama : Erik Kurniawan

NIM : 2102020106

Program Studi : Informatika

Fakultas : Ilmu Teknik

Menyatakan dengan sesungguhnya bahwa penelitian dan penulisan Skripsi yang saya susun sebagai persyaratan untuk memperoleh gelar Sarjana (S-1) Universitas Bina Insan, merupakan hasil kerja saya sendiri dan tidak menyuruh orang lain yang mengerjakannya. Ada pun bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain dan telah saya tuliskan sumbernya secara jelas sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika dikemudian hari ternyata terbukti bahwa penelitian dan tugas akhir ini bukan hasil kerja saya sendiri atau plagiat dalam bagian-bagian tertentu, maka saya bersedia dikenakan sanksi sesuai dengan peraturan perundangan yang berlaku.

Lubuklinggau, Januari 2025

Penulis,

Erik Kurniawan

2102020106

Abstract

Social media platform Instagr ecome one of the most frequently used platforms for expressing opinio scussions on various topics, including the "Apple Tax Holiday" issue, which has garnered public attention. This study aims to analyze sentiments toward the issue using the Naive Bayes and Random Forest algorithms. The data was collected using the Instant Data Scraper extension on the Google Chrome browser from comments on two public accounts, @martapurapedia and @pembasmi.kehaluan.reall, with a total of 3,791 comments. A preprocessing step was conducted to remove noise, such as spam comments, to ensure data quality.The testing involved splitting the dataset into training and testing data, as well as applying random oversampling to address data imbalance. Naive Bayes achieved an accuracy of 92.79% on the training data and 88.07% on the testing data, while Random Forest achieved an accuracy of 89.05% on the training data and 87.03% on the testing data. The results indicate that Naive Bayes excels in data generalization, whereas Random Forest is more stable in handling neutral classes. This study demonstrates that Naive Bayes is more suitable for analyzing this policy due to its ability to capture overall sentiment patterns, which is crucial for understanding public opinion on the policy.

Keywords: ***Sentiment Analysis, Naive Bayes, Random Forest, Instagram, Apple Tax Holiday***

Abstrak

Media sosial Instagram menjad tu platform yang sering digunakan untuk menyampaikan opini dan disku nai berbagai topik, termasuk isu "Apple tax holiday" yang menjadi perbincangan publik. Penelitian ini bertujuan untuk menganalisis sentimen terhadap isu tersebut dengan menggunakan algoritma Naive Bayes dan Random Forest. Data dikumpulkan menggunakan ekstensi Instant Data Scraper pada browser Google Chrome dari komentar pada dua akun publik, yaitu @martapurapedia dan @pembasmi.kehaluan.reall, dengan total 3.791 komentar. Proses preprocessing dilakukan untuk menghapus noise seperti komentar spam guna memastikan kualitas data.Pengujian dilakukan dengan membagi dataset menjadi data training dan testing, serta menggunakan random oversampling untuk mengatasi ketidakseimbangan data. Naive Bayes menunjukkan akurasi 92.79% pada data training dan 88.07% pada data testing, sementara Random Forest memiliki akurasi 89.05% pada data training dan 87.03% pada data testing. Hasil menunjukkan bahwa Naive Bayes unggul dalam generalisasi data, sedangkan Random Forest lebih stabil dalam menangani kelas netral. Penelitian ini menunjukkan bahwa Naive Bayes lebih sesuai untuk analisis kebijakan ini karena kemampuannya dalam menangkap pola sentimen secara keseluruhan, yang penting untuk memahami opini masyarakat terhadap kebijakan.

Kata Kunci: Analisis Sentimen, Naive Bayes, Random Forest, Instagram, Apple Tax Holiday

KATA PENGANTAR

Alhamdulillah puji dan penulis ucapkan kepada Allah SWT atas segala rahmat dan karunia-Nya yang telah memberikan kekuatan dan kesempatan, sehingga penulis dapat menyelesaikan skripsi ini dengan maksimal, Untuk diajukan sebagai syarat menyelesaikan pendidikan program Sarjana (S-1) Pada Program Studi Informatika Fakultas Ilmu Teknik Universitas Bina Insan. Sholawat beserta salam semoga tetap tercurahkan kepada bagi Nabi Muhammad SAW, keluarga, sahabat, serta umatnya hingga akhir zaman.

Selama proses penulisan dan penyusunan skripsi ini, penulis telah berusaha sebaik-baiknya untuk dapat menyelesaikan skripsi ini baik tepat pada waktunya. Penulis menyadari bahwa skripsi ini tentunya masih jauh dari sempurna dan mungkin terdapat kesalahan baik sengaja maupun tidak sengaja. Oleh karena itu, kritik dan saran yang membangun tentunya sangat diharapkan dari berbagai pihak.

Penulis mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu selama proses penyelesaian skripsi ini diantaranya yaitu:

1. Bapak dan ibuku yang telah banyak memberikan dukungan dan bantuannya dalam penulisan Skripsi ini.
2. Bapak Dr. H. Sardiyo, M.M. selaku Rektor Universitas Bina Insan.
3. Bapak Dr. Muhamad Akbar, S.T., M.IT selaku Wakil Rektor I Universitas Bina Insan.
4. Bapak Wakhid Nur Mukhlis, M.Pd., M.M selaku Wakil Rektor II Universitas Bina Insan
5. Bapak Dr. Rudi Kurniawan,S.T., M.Kom selaku Dekan Fakultas Ilmu Teknik Universitas Bina Insan yang telah banyak memberikan bimbingan dan arah dalam penulisan skripsi ini.
6. Bapak Budi Santoso M.kom selaku Kepala Program Studi Informatika Fakultas Ilmu Teknik Universitas Bina Insan

yang telah banyak memberikan bimbingan dan arah dalam penulisan skripsi ini.

7. Ibu Elmayati M.kom selaku Pembimbing I yang telah banyak memberikan bimbingan dan arah dalam p[illegible] Skripsi ini.
8. Ibu Harma Oktafia Lingga [illegible]I.Kom selaku Pembimbing II yang telah banyak memberikan bimbing[illegible]ah dalam penulisan Skripsi ini.
9. Seluruh Staf Dosen dan Karyawan Universitas Bina Insan Lubuklinggau yang telah banyak memberikan ilmu pengetahuan dan bimbingan kepada penulis.
10. terima kasih kepada Tim Penguji Universitas Bina Insan Lubuklinggau atas waktu, arahan, dan masukan yang diberikan selama proses ujian.

Akhir kata semoga penelitian ini dapat bermanfaat bagi untuk peneliti selanjutnya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Lubuklinggau, January 2025

Penulis

Erik Kurniawan

DAFTAR RIWAYAT HIDUP

Biodata

Nama	: Erik Kurniawan
Tempat / Tanggal Lahir	: Lubuklinggau 19-12-2003
Jenis Kelamin	: Laki-laki
Agama	: Islam
Alamat	: Jl.Kelabat

Pendidikan

- SD	: Sd N 36 Lubuklinggau
- SMP/MTS Sederajat	: Smp N 3 Lubuklinggau
- SMA/MAN/SMK Sederajat	: Smk N 1 Lubuklinggau

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang Penelitian

Indonesia, sebagai negara berkembang, menghadapi tantangan besar dalam meningkatkan pembangunan ekonomi untuk mencapai kesejahteraan masyarakat. Salah satu langkah strategis yang ditempuh pemerintah adalah menarik investasi asing langsung *Foreign Direct Investment/FDI* melalui berbagai kebijakan fiskal, seperti pemberian *tax holiday*. Kebijakan ini dirancang untuk menarik minat investor asing dengan memberikan insentif pajak bagi perusahaan yang memenuhi kriteria tertentu. Dalam konteks ini, investasi memainkan peran penting dalam mendorong pertumbuhan ekonomi, khususnya di sektor-sektor strategis seperti teknologi dan infrastruktur.[1]

Permintaan Apple untuk *tax holiday* selama 50 tahun di Indonesia memicu perdebatan tentang manfaat dan dampaknya. Kebijakan ini diharapkan mampu menarik investasi asing langsung (*Foreign Direct Investment*), terutama di sektor teknologi, namun sering dianggap memberikan keistimewaan bagi perusahaan multinasional dengan mengorbankan pemerataan ekonomi dan keadilan sosial. Hal ini menimbulkan kekhawatiran terkait potensi ketimpangan manfaat yang dirasakan oleh masyarakat lokal. Dalam konteks ini, opini masyarakat terhadap kebijakan *tax holiday* menjadi penting untuk dipahami, karena dapat memengaruhi kepercayaan publik terhadap pemerintah dan keberhasilan kebijakan tersebut.[2] Analisis sentimen publik terhadap kebijakan tax holiday yang diminta Apple menjadi krusial untuk memahami opini masyarakat secara terukur, baik yang bersifat positif maupun negatif. Hal ini dapat dilakukan melalui penerapan algoritma machine learning yang terbukti mampu menangkap pola sentimen dari data teks secara efektif.

Berbagai algoritma machine learning telah banyak digunakan untuk analisis sentimen, termasuk Naive Bayes dan Random Forest. Penelitian sebelumnya menunjukkan bahwa Naïve Bayes Classifier merupakan teknik klasifikasi yang dapat bekerja dengan cepat dengan akurasi yang tinggi pada jumlah data yang besar[3], sementara itu Random forest memiliki hasil akurasi

yang bagus, kuat terhadap outliers dan noise, dan lebih cepat dibandingkan dengan bagging dan boosting. Sebagai contoh, analisis sentimen masyarakat terhadap kasus pembobolan dana nasabah Bank BSI di Twitter menggunakan metode Random Forest dan Naive Bayes menghasilkan akurasi 81% untuk Naive Bayes dan 78% untuk Random Forest.[5] Namun, pada penelitian perbandingan metode Naive Bayes Classifier dengan Random Forest dalam prediksi rating review drama Korea, Random Forest lebih unggul dengan akurasi sebesar 89% dibandingkan Naive Bayes yang mencapai 86%. Bahkan dalam prediksi rating lanjutan, Random Forest kembali lebih unggul dengan akurasi sebesar 41% dibandingkan 40% untuk Naive Bayes.[6] Namun, performa algoritma ini sangat bergantung pada karakteristik data yang dianalisis. Kekurangan dalam presisi dan efisiensi masing-masing algoritma menunjukkan perlunya penelitian lanjutan untuk menentukan algoritma yang paling optimal, khususnya pada konteks kebijakan tax holiday Apple.

Maka dari itu, penting untuk melakukan analisis sentimen publik mengenai permintaan tax holiday Apple melalui media sosial. Penelitian ini bertujuan untuk membandingkan efektivitas algoritma machine learning, yaitu Naive Bayes dan Random Forest, dalam mengklasifikasikan opini masyarakat terkait kebijakan tersebut. Hasil penelitian diharapkan dapat memberikan wawasan mendalam mengenai persepsi publik baik yang positif maupun negatif terhadap kebijakan ini serta mendukung pengambilan keputusan yang lebih transparan dan berbasis data.

1.2 Identifikasi Masalah

Permintaan Apple untuk tax holiday selama 50 tahun menimbulkan berbagai pertanyaan penting:

1. Persepsi masyarakat terhadap kebijakan ini belum dipetakan secara mendalam, baik yang bersifat positif maupun negatif. meskipun penting untuk menilai kepercayaan publik terhadap pemerintah.
2. Efektivitas algoritma machine learning, seperti naïve bayes dan random forest dalam menganalisis opini publik terhadap kebijakan ini belum diketahui

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah di atas, rumusan masalah yang dapat diajukan adalah:

a. Bagaimana opini masyarakat terkait permintaan tax holiday Apple di media sosial?
b. Algoritma machine learning mana yang paling efektif untuk menganalisis sentimen publik terhadap kebijakan tersebut?

1.4 BatasanMasalah

a. Penelitian ini hanya akan menganalisis sentimen publik berdasarkan data dari media sosial instgram
b. Hanya algoritma Naive Bayes, Random Forest yang akan digunakan dalam komparasi efektivitas model analisis sentimen.
c. Fokus penelitian terbatas pada konteks kebijakan tax holiday yang diminta Apple di Indonesia, tanpa membahas konteks global atau kebijakan perpajakan lainnya.

1.5 Tujuan dan Manfaat Penelitian

1.5.1 Tujuan Penelitian

1) Membandingkan efektivitas algoritma machine learning dalam menganalisis sentimen publik terkait kebijakan Apple tax holiday.
2) Memberikan wawasan tentang persepsi publik terhadap kebijakan tersebut.

1.5.2 Manfaat Penelitian

1) Memberikan kontribusi akademik pada pengembangan metode analisis sentimen berbasis machine learning.
2) Membantu pemerintah Indonesia dalam memahami opini publik untuk merancang kebijakan perpajakan yang lebih responsif dan transparan.

BAB II
K[illegible] PUSTAKA

2.1 Literatur

2.1.1. Analisis Sentimen

Analisis sentimen mengklasifikasikan sentimen berdasarkan polaritas teks dalam sebuah frasa, sehingga dapat ditentukan sebagai sentimen positif, negatif, atau netral. *Sentiment analysis* atau analisis sentimen dalam Bahasa Indonesia adalah sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan bagaimana sentimen tersebut dapat dikategorikan sebagai sentimen positif maupun sentimen negatif.[7]pengguna internet saat ini banyak yang menuliskan pendapat atau opini, pengalaman dan berbagai hal yang terjadi atau sekiranya menarik untuk mereka oleh karna itu dibutuhkan nya analisis sentimen untuk mengetahui pendapat dan opini publik [8]

2.1.2. Text Minning

Text mining adalah suatu langkah dari analisis teks yang secara otomatis dilakukan oleh komputer dengan tujuan menggali informasi yang berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen.[9] Text mining bertujuan untuk menganalisis opini, sentiment, evaluasi,penilaian, sikap serta emosi seseorang sehingga dapat diketahui kaitannya dengan suatu topik, layanan, organisasi,individu, atau kegiatan tertentu. Text mining dan data mining berbeda dalam artian melibatkan data yang terstruktur untuk data mining sedangkan untuk text mining erat kaitannya preprocessing untuk pengolahan data. Meskipun demikian keduanya memiliki konsep yang dalam persepsi ilmu algoritma yang sama. [10]

2.1.3. Machine Learning

Machine Learning adalah salah satu disiplin ilmu dari Computer Science yang mempelajari bagaimana membuat komputer atau mesin memiliki kecerdasan. Agar memiliki kecerdasan tersebut, komputer atau

mesin harus mampu belajar dari data. Dengan kata lain, Machine Learning adalah bidang keilmuan [illegible]erfokus pada pembelajaran komputer untuk memahami dan memp[illegible] guna menghasilkan keputusan cerdas. Salah satu pendekatan d[illegible]chine Learning adalah Concept Learning, sebuah metode yang membutuhkan data training dan mampu mengatasi data negatif maupun positif karena termasuk dalam kategori Supervised Learning,terdapat algoritma seperti Naive Bayes dan Random Forest sebagai bagian Supervised Learning, yang masing-masing memiliki keunggulan dan karakteristik unik dalam memproses data untuk menghasilkan analisis yang akurat.[11]

2.1.4. Naive Bayes

Naive Bayes adalah metode yang mudah diimplementasikan, memiliki struktur yang cukup sederhana, dan tingkat efektifitasnya yang tinggi sehingga banyak digunakan untuk kebutuhan data mining Metode Naïve Bayes tergolong supervised learning, karena membutuhkan data latih untuk mengelompokkan data kedalam sebuah kelas[12]. Gambar an umum klasifikasi Naive Bayes ditunjukkan pada Persamaan berikut:

$$P(C|W) = \mathrm{P}\frac{\mathrm{P(C)}P(W|C)}{\mathrm{P(W)}}$$

Keterangan :

$P(c|w)$: Posterior adalah peluang kategori $c$ketika terdapat kemunculan kata w

$P(w|c)$: Likelihood adalah peluang sebuah kata w yang masuk kategori c

$P(c)$: Prior adalah peluang munculnya kategori c

$P(w)$: Evidence adalah peluang kemunculan kata

Untuk perhitungan peluang kemunculan kelas atau prior dapat dilihat pada persamaan berikut :

$$P(c) = \frac{Nc}{\mathrm{N}}$$

Keterangan :

$P(c)$: peluang kemuncu kategori c

Nc : banyak dokumen la kategori c

N : total dokumen latih yang digunakan**.**

Pada persamaan likelihood, penggunaan Laplace Smoothing bertujuan untuk menghindari angka nol. Setiap kata pada metode ini diolah dengan menggunakan distribusi multinomial, karena urutan kejadian munculnya kata pada dokumen tidak terlalu diperdulikan Perhitungan nilai likelihood pada multinomial Naïve Bayes dapat dilihat pada Persamaan berikut :

$$p(w|c) = \frac{\text{count(w,c)}+1}{\text{count(c)}+|\text{v}|}$$

Keterangan :

$P(w|c)$: likelihood, peluang kata w masuk kategori c

$count(w, c)$: jumlah kata w yang ada pada suatu kategori c

$count(c)$: total kemunculan kata pada kategori c

$|V|$: jumlah kemunculan kata atau term unik

Untuk perhitungan nilai evidence ditunjukkan pada persamaan berikut

$$p(w) = \frac{|\text{w}|}{|\text{w}|}$$

Keterangan :

$P(w)$: Evidence adalah peluang kemunculan kata

$|N|$: jumlah kemunculan kata pada seluruh dokumen

Kelebihan Naïve Bayes[13] **:**

a) Bisa dipakai untuk data kuantitati maupun kualitatif
b) Tidak memerlukan jumlah data yang banyak
c) Tidak perlu melakukan data training yang banyak

d) Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan.

e) Perhitungannya cep[PDF]fisien di pahami.

2.1.5. Random Forest

Random Forest ada[PDF]embangan dari metode Decision Tree yang menggunakan beberapa Decision Tree, dimana setiap Decision Tree telah dilakukan pelatihan menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak. Random Forest memiliki beberapa kelebihan, yaitu dapat meningkatkan hasil akurasi jika terdapat data yang hilang, dan untuk resisting outliers, serta efisien untuk penyimpanan sebuah data. Selain itu, Random Forest mempunyai proses seleksi fitur dimana mampu mengambil fitur terbaik sehingga dapat meningkatkan performa terhadap model klasifikasi[14] Random Forest adalah metodep pelatihan yang berbasis ensemble learning, yang menggunakan algoritma Decision Tree. Prosesnya melibatkan pembuatan beberapa model Decision Tree dengan menggunakan data uji yang sama untuk setiap model. Selanjutnya, hasil prediksi dari setiap model Decision Tree digabungkan melalui proses majority voting dengan menggunakan metode modus. Hasil akhir dari proses ini menjadi kelas prediksi. Pada persamaan merupakan rumusuntuk menghitung Entropy, dan persamaan merupakan rumus untuk menghitung Gain pada metode Random Forest.

$$Entropy\ (E) = -\sum_{i=1}^{n} pi\ log2\ pi$$

Dimana n adalah jumlah kelas yang mungkin, dan pi adalah probabilitas dari kelas i.

$$Gain = Entropy_{\text{parent}} - Entropy_{\text{children}}$$

$Entropy_{\text{parent}}$ adalah entropi dari node induk dan $Entropy_{\text{children}}$ adalah entropi rata-rata dari node anak.[15]

Pada algoritma random forest terdapat banyak parameter yang digunakan untuk membuat pohon acak. Hanya saja, parameter yang paling berpengaruh terhadap hasil prediksi dan mencegah terjadinya overfitting adalah[4]:

1. N Estimators (digunakan untuk membentuk jumlah pohon yang ada pada hutan. Nilai n estimators diubah dari 10 sampai 100).
2. Max Depth (digunakan mengatur kedalaman pohon yang akan dibangun).
3. Criterion (digunakan untuk mengukur kualitas split. Kriteria yang didukung adalah "gini" untuk ketidakmurnian Gini dan "entropy" untuk perolehan informasi).
4. Minimal Samples Split (parameter untuk membentuk jumlah pengamatan minimum atau pemisahan yang diperlukan pada simpul yang diberikan untuk membagi hutan acak, nilai default parameter ini adalah 2).
5. Max Features (Jumlah fitur yang dipertimbangkan saat mencari pemisahan terbaik, nilai default-nya adalah auto, sqrt, dan log2).

2.1.6. Pembobotan Tf-idf

Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency) dapat didefinisikan sebagai metode untuk menentukan nilai frekuensi sebuah kata di dalam sebuah dokumen atau artikel. Perhitungan ini menentukan seberapa relevan sebuah kata di dalam sebuah dokumen. TF IDF merupakan hasil dari perhitungan antara TF (Term Frequency) dan IDF (Inverse Document Frequency). TF (Term Frequency) adalah frekuensi sebuah kata yang muncul dalam sebuah dokumen. IDF (Inverse Document Frequency) adalah pengukuran seberapa penting suatu kata Rumus dari TF-IDF adalah

TF-IDF = $TF \times IDF$……..…

$$\text{TF} = \frac{Frekuensi\ term\ dalam\ satu\ dokumen}{\text{Total kata dalam satu dokumen}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\text{TF} = \log \frac{Total\ dokumen+1}{\text{Frekuensi dokumen mangandung term}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

2.1.7. Media Sosial

Media sosial telah menjadi platform yang memiliki peranan yang sangat signifikan dalam kehidupan sehari-hari kita. Banyak individu yang

memanfaatkan media sosial untuk berbagi informasi, berinteraksi dengan kerabat dan teman, sert mempromosikan usaha mereka. Meskipun demikian, media sosial adi sumber informasi yang berharga bagi perusahaan dan orga tuk menganalisis pendapat pengguna terhadap merek, produk, atau layanan yang mereka tawarkan melalui analisis sentimen. Data untuk analis sentimen dapat beragam, bisa berupa positif, negatif, atau netral. Melalui analisis sentimen data, perusahaan dan organisasi dapat memperoleh pemahaman mengenai respons pengguna terhadap merek, produk, atau layanan yang mereka tawarkan melalui media sosial. Tahap awal dalam melakukan analisis sentimen adalah mengumpulkan data dari berbagai platform media sosial.[16]

2.1.8. Instagram

Instagram adalah platform media sosial yang memungkinkan pengguna untuk berbagi foto, video, dan cerita dalam format visual. Diluncurkan pada tahun 2010, Instagram awalnya fokus pada berbagi foto dengan filter artistik untuk meningkatkan penampilan gambar. Namun, seiring berjalannya waktu, platform ini berkembang menjadi salah satu media sosial terbesar di dunia dengan lebih dari satu miliar pengguna aktif bulanan pada tahun 2023[17]

2.1.9. KebijakanTaxHoliday

Tax holiday adalah kebijakan pemerintah yang memberikan pembebasan atau pengurangan pajak kepada sektor-sektor tertentu untuk periode waktu tertentu. Tax Holiday pertama kali diatur dalam Undang-Undang Nomor 1 Tahun 1967 yang membahas tentang Penanaman Modal Asing (PMA).[1] Tujuan utama dari *tax holiday* adalah untuk mendorong investasi dalam sektor-sektor tertentu, memicu pertumbuhan ekonomi, dan menciptakan lapangan kerja. Biasanya, kebijakan ini diterapkan pada sektor-sektor strategis atau industri-industri yang dianggap vital bagi pembangunan ekonomi negara. Di tengah dinamika perekonomian global, kebijakan fiskal seperti tax holiday telah menjadi salah satu strategi yang diandalkan oleh banyak negara untuk mendorong pertumbuhan ekonomi

dan investasi. Dalam konteks ini, tax holiday memberikan sejumlah manfaat yang signifikan. Pertama, tax holiday memberikan stimulasi kuat untuk aktivitas investasi perusahaan. Dengan beban pajak yang dikurangkan atau bahkan dihapuskan selama periode tertentu, perusahaan cenderung lebih termotivasi untuk melakukan investasi jangka panjang. Hal ini membuka pintu bagi ekspansi perusahaan, pembelian aset baru, dan peningkatan kapasitas produksi.[18]

2.1.10. Python

Python adalah salah satu bahasa pemrograman yang memiliki tujuan umum terbaik yang bisa digunakan diberbagai sistem operasi saat ini. Python juga merupakan bahasa pemrograman dengan tipe interpreted language yang dimana code tidak akan diubah menjadi sebuah code yang dapat dipahami oleh komputersebelum dijalankan, melainkan perubahan tersebut terjadi saat runtime. Python sendiri diciptakan oleh Guido Van Rossum. Aplikasi yang bisa dikembangkan dari bahasa pemrograman ini sangat beragam salah satunya dibidang Scientific dan numeric yang memiliki library cukup banyak seperti Spicy, Pandas, Numpy[17]

Python merupakan bahasa pemrograman yang mengalami peningkatan popularitas yang signifikan dalam beberapa tahun terakhir. Salah satu faktor keberhasilannya adalah perkembangan berbagai pustaka (library) yang semakin canggih, serta kontribusi dari komunitas pengembang yang sangat aktif. Hal ini menjadikan Python sebagai salah satu bahasa pemrograman yang stabil dan terus berkembang dengan cepat. Python menawarkan berbagai pustaka dan kerangka kerja (framework) yang sering digunakan dalam analisis data. Berikut adalah beberapa pustaka yang digunakan dalam penelitian ini:

1) Pandas adalah pustaka open-source yang menyediakan struktur data tingkat tinggi yang sangat fleksibel, serta berbagai alat untuk analisis data. Pustaka ini sering digunakan dalam pemrosesan data, termasuk analisis, manipulasi, dan pembersihan data.

2) NLTK adalah platform yang dirancang untuk mempermudah pengolahan data teks. Dikembangka Steven Bird dan Edward Loper, pustaka ini pertama kali dirilis pad 001 dan menjadi salah satu alat penting dalam analisis teks.
3) Sastrawi adalah pustaka yang dikembangkan khusus untuk melakukan proses stemming pada teks berbahasa Indonesia. Stemming adalah proses mengubah kata menjadi bentuk dasarnya. Sastrawi merupakan pengembangan dari proyek sebelumnya yang berbasis PHP.
4) Scikit-learn adalah pustaka Python yang sangat populer dalam dunia machine learning. Pustaka ini mendukung berbagai algoritma machine learning, baik yang diawasi (supervised) maupun tidak diawasi (unsupervised), seperti regresi linear, klasifikasi, dan clustering. Scikit-learn merupakan pustaka open-source yang dirancang untuk menangani data kompleks.
5) NumPy adalah pustaka Python yang digunakan untuk menangani array dan menyediakan berbagai fungsi dalam bidang aljabar linear, transformasi Fourier, dan operasi matriks. Pustaka ini dikembangkan pada tahun 2005 oleh Travis Oliphant dan tersedia secara open-source.
6) TfidfVectorizer adalah metode machine learning berbasis TF-IDF (Term Frequency-Inverse Document Frequency), yang dirancang khusus untuk mengolah teks dari dokumen. Pustaka ini berguna dalam menganalisis frekuensi dan relevansi kata dalam sebuah korpus.

2.2 Penelitian Relevan

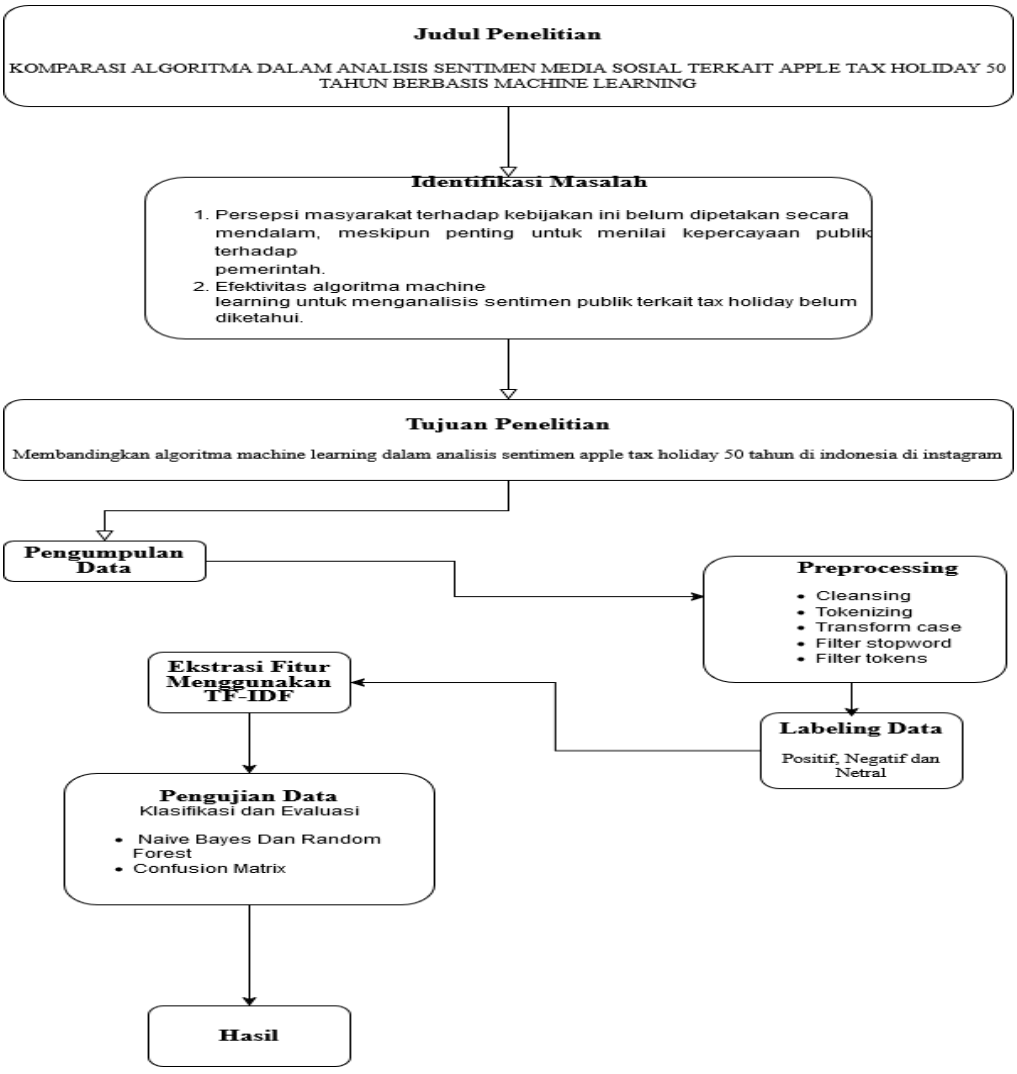
Tabel 2.1. Penelitian Terda

No	Author	Judu	Metode	Hasil Penelitian
1	Desti Mualfah, Ananda Prihatin, Rahmad Firdaus, Sunanto[5]	Anal Sent Masyarakat Terhadap Kasus Pembobolan Data Nasabah Bank BSI Pada Twitter	Naïve Bayes Dan random forest	Hasil penelitian ini menunjukkan bahwa metode Naive Bayes memiliki tingkat akurasi 81%, sedangkan metode Random Forest memiliki tingkat akurasi 78%. Kata
2	Ferisa Dwi Alfia Meisty, Dian Anggraeni, Mohamat Fatekurohman[6]	Perbandingan Metode Naïve Bayes Classifier dengan Metode Random Forest pada Prediksi Rating Review Drama Korea	Naïve Bayes dan Random forest	Pada prediksi review, metode random forest memperoleh nilai accuracy sebesar 89%, sedangkan metode naïve bayes classifier memperoleh nilai accuracy sebesar 86%. Pada prediksi rating, metode random forest memperoleh nilai accuracy sebesar 41%, sedangkan metode naïve bayes classifier memperoleh nilai accuracy sebesar 40%. Kesimpulan penelitian ini adalah metode random forest lebih unggul dan akurat dalam memprediksi rating review drama korea. Kata
3	Nicolaus Advendea Prakoso Indaryono , Rd.Rohmat Saedudin, Faqih Hamami[19]	analisa perbandingan algoritma random forest dan naïve bayes untuk klasifikasi curah hujan	Random Forest dan Naïve Bayes	Hasil penelitian, mendapatkan kesimpulan bahwa algoritma Random forest mendapatkan performa dan akurasi yang lebih baik daripada algoritma Naïve bayes dalam proses klasifikasi dataset

		berdasarkan iklim di indo		iklim di Indonesia. Algoritma Random forest mencapai akurasi sebesar 86.55%, sementara algoritma Naïve bayes hanya mencapai akurasi sebesar 36.61%.
4	Bustami Yusuf, Muthmainna Qalbi, Basrul, Ima Dwitawati, Malahayati5, Mega Ellyadi[20]	implementasi algoritma naive bayes dan random forest dalam memprediksi prestasi akademik mahasiswa universitas islam negeri ar-raniry banda aceh	Naïve Bayes dan Random Forest	Hasil yang di peroleh dari penelitian ini menunjukkan nilai korelasi tertinggi pada variabel IP awal sebesar r=0,783 dan variabel cuti memiliki tingkat korelasi sangat lemah sebesar r=0,054. Nilai keakuratan variabel algoritma naive bayes setelah di cleaning sebesar 78.0% dan variabel algoritma Random Forest sebesar 76,7%.
5	Ricky Leonardo, Janis Pratama, Chrisnatalis[21]	Perbandingan Metode Random Forest Dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing	Random Forest dan Naïve Bayes	dapat disimpulkan bahwa Algoritma Random Forest lebih tepat digunakan untukkasus prediksi keputusanklien. Hal ini terlihat dimana akurasiyang didapatkan adalah 90%, dimana lebih tinggi 5% dibandingkan algoritma Naïve Bayes. Nilai dari AUC dari algoritma Random Forest adalah 0.97 dimana lebih tinggi 1,3 dibandingkan algoritma Naïve Bayes.

2.3 Kerangka Berpikir

Kerangka berpikir atau [illegible]gka pemikiran adalah dasar pemikiran dari penelitian yang disinte[illegible]ri fakta-fakta, observasi dan kajian kepustakaan.[22] Kerangka [illegible]igunakan untuk menjelaskan pola a ntar teori dan objek dalam penelitian. Pemikiran dimulai dari latar belaka ng penelitian, identifikasi masalah, rumusan masalah, batasan masalah,tuj uan dan manfaat penelitian.[23] adapun alur dalam proses penelitian ini meliputi beberapa tahapan yang akan dijelaskan sebagai berikut:



Gambar 2.1. Kerangka Berfikir

BAB III
METO GI PENELITIAN

3.1 Analisa Sistem

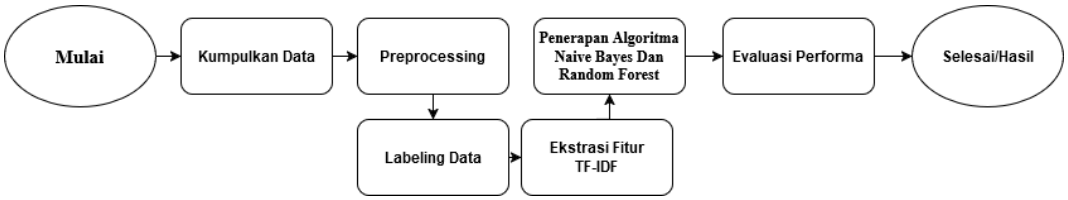
3.1.1 Analisa Sistem yang Ber

Penelitian ini dilakukan dengan menganalisis sistem yang berjalan terkait opini publik di media sosial mengenai permintaan Apple untuk tax holiday selama 50 tahun di Indonesia. Data opini diperoleh dari komentar di platform Instagram pada akun-akun yang relevan. Saat ini, belum ada sistem terintegrasi untuk mengumpulkan dan menganalisis opini publik terhadap kebijakan tersebut.

3.1.2 Alternatif Pemecahan Masalah

Alternatif solusi yang diajukan adalah penerapan algoritma machine learning, yaitu Naive Bayes dan Random Forest, untuk menganalisis data opini dari media sosial. Algoritma ini dipilih karena kemampuannya dalam mengklasifikasikan data teks secara efektif dan efisien.

3.1.3 Metode Analisa



Gambar 3.1. Flowchart Penelitian

Flowchart di atas menggambarkan tahapan dalam metode analisis yang dilakukan. Penjelasan dari setiap tahapan adalah sebagai berikut: Metode analisis yang digunakan dalam penelitian ini adalah algoritma Naive Bayes dan Random Forest. Algoritma-algoritma ini dipilih untuk membandingkan kinerja dalam klasifikasi opini publik berdasarkan sentimen. Tahapan analisisnya meliputi preprocessing data, penerapan algoritma, dan evaluasi performa algoritma berikut penjelasannya:

1) Pengumpulan Data

   Langkah pertama adalah mengumpulkan data dari media sosial yang

relevan dengan opini publik tentang tax holiday. Data yang dikumpulkan berupa teks mentah ya[illegible] opini pengguna media sosial.

2) Preprocessing Data

   Langkah ini mencakup [illegible]han data agar sesuai untuk analisis. Data yang dikumpulkan dari media sosial diolah dengan teknik seperti tokenisasi, stemming menggunakan library Sastrawi, dan penghapusan kata-kata tidak penting (stopwords). Langkah ini bertujuan untuk menghilangkan noise dalam data sehingga menghasilkan dataset yang relevan untuk pelatihan model.

3) Labeling Data

   Data yang dikumpulkan dari media sosial dan telah di preprocessing sebelumnya diberikan label berdasarkan kategori sentimen, seperti positif, negatif, atau netral. Proses ini dilakukan secara manual atau menggunakan alat bantu labeling untuk memastikan dataset sesuai dengan tujuan analisis.

4) Ekstraksi Fitur

   Setelah preprocessing, dilakukan pembobotan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Metode ini mengubah teks menjadi vektor numerik berdasarkan frekuensi kemunculan kata dan relevansinya dalam dokumen. Langkah ini penting untuk memastikan bahwa data dapat diproses oleh algoritma machine learning.

5) Penerapan Algoritma

   Naive Bayes digunakan untuk mengklasifikasikan opini publik berdasarkan pendekatan probabilistik, dengan menghitung kemungkinan suatu kategori berdasarkan kata-kata dalam teks.

   Random Forest digunakan untuk membangun model klasifikasi berbasis kumpulan pohon keputusan. Algoritma ini dipilih karena kemampuannya untuk menangani data yang kompleks dan mengurangi risiko overfitting.

6) Evaluasi Performa Algoritma

Performa kedua algo[illegible] dibandingkan dengan menggunakan metrik evaluasi seperti:

a. Accuracy untuk me[illegible]roporsi prediksi yang benar.

b. Precision untuk mengevaluasi akurasi prediksi positif.

c. Recall untuk mengevaluasi sejauh mana model dapat mendeteksi semua data positif.

d. F1-Score untuk memberikan keseimbangan antara precision dan recall.

3.2 Teknik Pemilihan Sampel

3.2.1 Teknik Pengumpulan Data

Data dikumpulkan menggunakan ekstensi Instant Data Scraper di browser Google Chrome. Data berupa komentar dari postingan Instagram terkait kebijakan Apple tax holiday pada akun publik seperti @martapurapedia dan @pembasmi.kehaluan. reall. Komentar yang tidak relevan dan spam dihapus untuk meningkatkan kualitas data.

3.2.2 Teknik Analisa Data

Proses analisis data mencakup langkah-langkah berikut:

1. Preprocessing Data: Data mentah dibersihkan dengan teknik tokenisasi, stemming menggunakan library Sastrawi, dan penghapusan stopwords.
2. Ekstraksi Fitur: Teks diubah menjadi representasi numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency).
3. Penerapan Algoritma: Data dianalisis menggunakan algoritma Naive Bayes dan Random Forest untuk mengklasifikasikan sentimen.
4. Evaluasi Kinerja: Kinerja algoritma diukur menggunakan metrik accuracy, precision, recall, dan F1-score.

3.3 Tempat dan Waktu Penelitian

a. Tempat

Penelitian ini dilakukan di rumah peneliti selama 2 bulan, mulai dari november hingga desember 2024

b. Waktu Penelitian

Penyusunan proposal skripsi ini telah dirancang rencana kegiatan yang akan dilaksanakan dala[illegible]s penelitian. Rencana tersebut disajikan dalam format tabel di b[illegible]

Tabel 3.1. Waktu Penelitian

NNO	Jenis Kegiatan	Waktu Kegiatan											
		November 2024				Desember 2024				Januari 2025			
		11	22	33	44	11	22	33	44	11	22	33	44
1	Pengajuan Judul												
2	Pengumpulan Data												
3	Penulisan Proposal												
4	Bimbingan Proposal												
5	Ujian Proposal												
6	Revisi Ujian Proposal												
7	Penulisan Skripsi												
8	Bimbingan Skripsi												
9	Ujian Skripsi												
10	Revisi Ujian Skripsi												

BAB IV

HASIL PENE[illegible]N DAN PEMBAHASAN

4.1 Gambaran Umum (Tempa[illegible]ian)

4.1.1 Gambaran Umum (Tem[illegible]itian)

Penelitian ini menggunakan data dari dua akun publik di Instagram, yaitu @martapurapedia dan @pembasmi.kehaluan.reall.

Akun @martapurapedia dibuat pada Desember 2021 dan memiliki 96,6 ribu pengikut dengan lebih dari 7.096 unggahan. Akun ini sering membagikan konten viral dari Indonesia maupun luar negeri, dengan fokus khusus pada berita viral di wilayah OKU Timur.

Sementara itu, akun @pembasmi.kehaluan.reall dibuat pada April 2016 dan saat ini memiliki 965 ribu pengikut dengan lebih dari 2.504 unggahan. Akun ini membagikan berbagai berita viral dari Indonesia dan luar negeri tanpa spesifikasi daerah tertentu.

Kedua akun ini dipilih karena memiliki relevansi dengan diskusi publik terkait "Apple Tax Holiday" dan menarik banyak interaksi dari pengguna Instagram. Data yang digunakan dalam penelitian ini berasal dari komentar pada postingan yang dibuat oleh akun-akun tersebut, dengan total 3.791 komentar yang relevan.

4.1.2 Struktur Organisasi (Tempat Penelitian)

Struktur data yang digunakan dalam penelitian ini berasal dari media sosial Instagram. Data terdiri dari beberapa elemen utama, yaitu:

1. Postingan: Sumber utama diskusi yang memuat informasi terkait "Apple tax holiday" yang diunggah oleh akun @martapurapedia pada tanggal 31 Oktober 2024 dan akun @pembasmi.kehaluan.reall pada tanggal 5 November 2024.
2. Komentar: Data utama penelitian berupa opini publik yang diambil dari komentar pengguna pada kedua postingan tersebut. Data ini mencakup berbagai sentimen, termasuk positif, netral, dan negatif.
3. Pengguna: Identitas pengguna yang memberikan komentar tidak dijadikan fokus penelitian, tetapi data mereka digunakan untuk analisis sentimen.

Data dikumpulkan menggunakan ekstensi Instant Data Scraper pada browser Google Chrome pada tangga[illegible]November 2024. Setelah data terkumpul, dilakukan proses filtrasi untuk [illegible]s noise, seperti komentar spam atau tidak relevan, sehingga memastikan k[illegible]ta untuk analisis lebih lanjut.

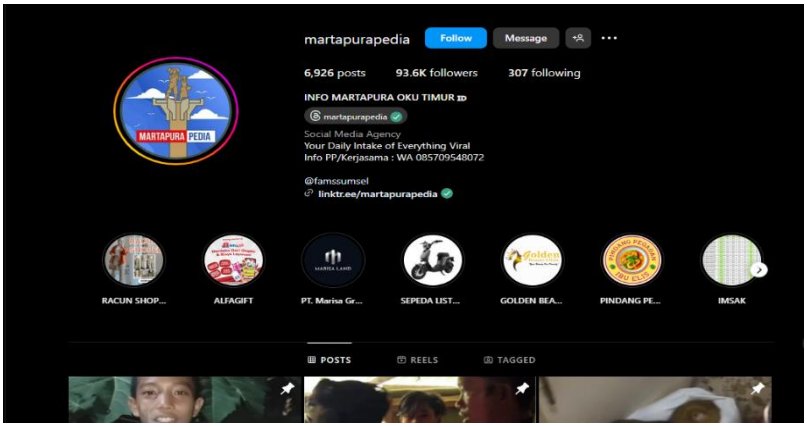
4.2 Hasil

4.2.1. Pengumpulan Data

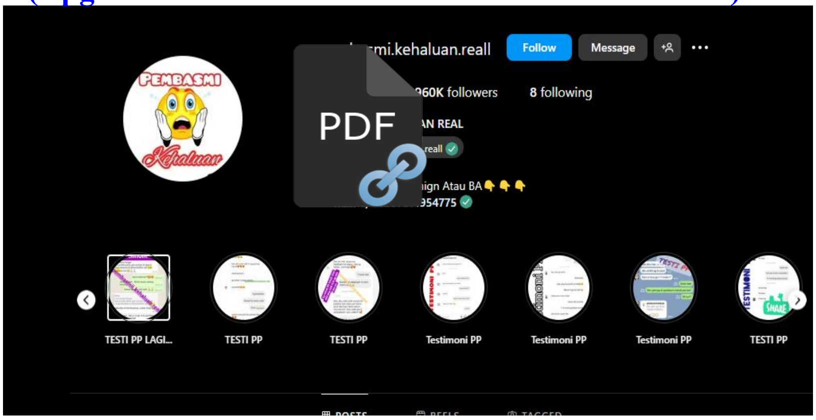
Data penelitian ini dikumpulkan menggunakan ekstensi Instant Data Scraper pada browser Google Chrome. Data yang diperoleh mencakup opini publik yang relevan dengan kata kunci tertentu, seperti *"Apple tax holiday"* atau *"kebijakan perpajakan"*. Data diambil dari komentar dua postingan akun publik Instagram, yaitu @martapurapedia dan @pembasmi.kehaluan.reall, dengan total 3.791 komentar.

Komentar diambil dari:

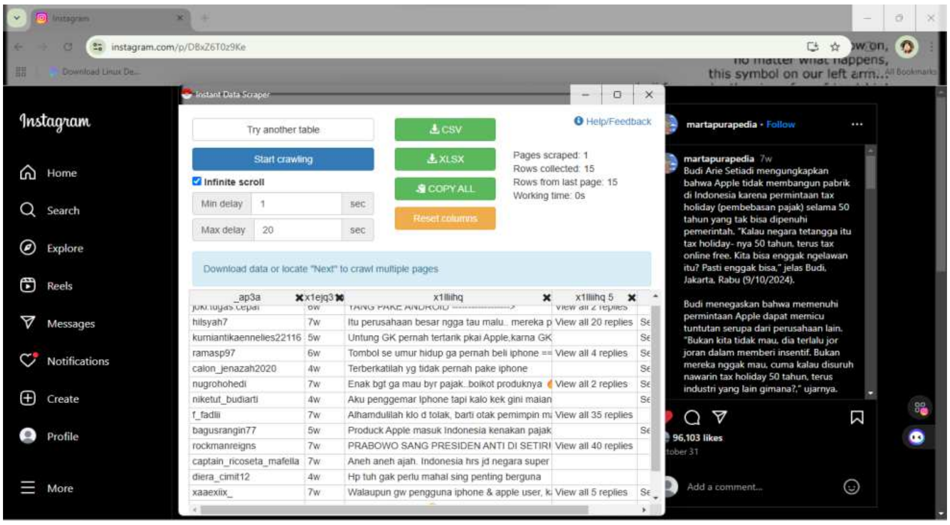
1. Postingan akun @pembasmi.kehaluan.reall yang diunggah pada 5 November 2024, diambil komentarnya pada 25 November 2024.
2. Postingan akun @martapurapedia yang diunggah pada 31 Oktober 2024, diambil komentarnya pada 25 November 2024.



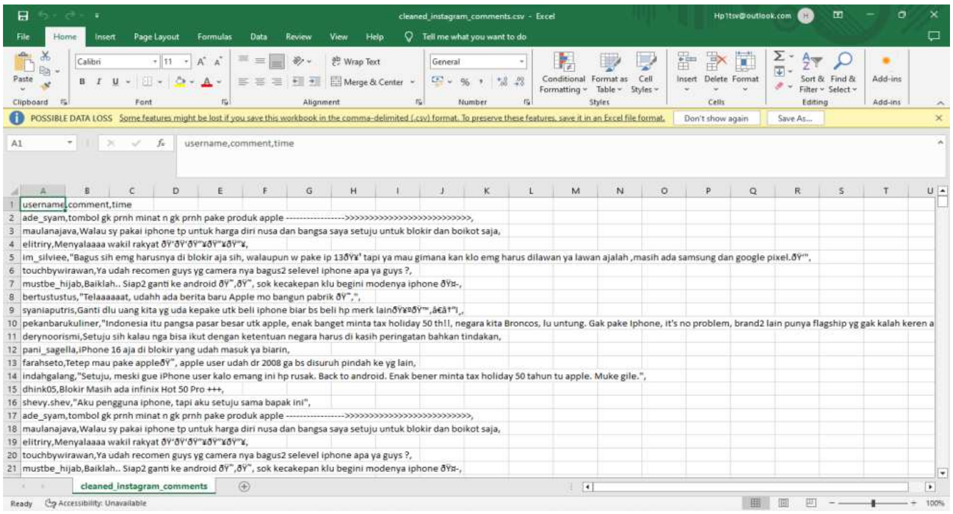
Gambar 4.1. @martapurapedia



Gambar 4.2 @pembasmi.kehaluan.reall



Gambar 4.3. Pengambilan data di instagram menggunakan instant scraper



Gambar 4.4. Hasil Pengambilan Data

Setelah data dikumpulkan, dilakukan proses filtrasi atau preprocesing untuk menghapus noise, seperti kon spam, komentar tidak relevan, atau data duplikat. Langkah ini bertujuan memastikan kualitas data yang digunakan dalam analisis lebih lanjut.

4.2.2. Teknik Preprocesing Data

Proses preprocessing data merupakan tahap penting dalam pengolahan teks untuk memastikan data yang digunakan dalam analisis memiliki kualitas yang baik dan relevan. Data mentah yang diperoleh dari proses pengumpulan sering kali mengandung informasi yang tidak diperlukan, seperti simbol, angka, karakter spesial, serta kata-kata yang tidak relevan. Oleh karena itu, langkah-langkah preprocessing dilakukan untuk mengubah data mentah menjadi data yang siap digunakan untuk proses analisis lebih lanjut.

Langkah-langkah preprocessing yang dilakukan dalam penelitian ini mencakup:

1. Pembersihan data

   Pada tahap ini, simbol, angka, dan karakter spesial yang tidak relevan dihapus dari teks. Langkah ini bertujuan untuk mengurangi noise pada data dan memastikan hanya informasi yang relevan yang dipertahankan. berikut kode program cleaned data pada gambar .

```python
import pandas as pd
import re

# Fungsi untuk membersihkan teks
def clean_text(text):
    # Periksa apakah input adalah string, jika bukan, ubah menjadi string kosong
    if not isinstance(text, str):
        return ""
    # Menghapus emoji dan karakter non-ASCII
    text = re.sub(r'[^\x00-\x7F]+', ' ', text)
    # Menghapus URL
    text = re.sub(r'https?://\S+|www\.\S+', '', text)
    # Menghapus simbol seperti --->>> atau ----
    text = re.sub(r'[-]+|[>]+', '', text)
    # Menghapus pola teks aneh seperti oYOY
    text = re.sub(r'o+Y+o+Y+', '', text, flags=re.IGNORECASE)
    # Menghapus tanda baca
    text = re.sub(r'[^\w\s]', '', text)
    # Menghapus kata sangat pendek (1-2 huruf)
    text = re.sub(r'\b\w{1,2}\b', '', text)
    # Menghapus spasi ganda
    text = re.sub(r'\s+', ' ', text)
    return text.strip()

# Contoh DataFrame (Pastikan Anda sudah memuat data Anda sebelumnya)

# Pastikan nilai NaN diatasi terlebih dahulu
data['comment'] = data['comment'].fillna("")

# Terapkan fungsi pembersihan
data['cleaned_comment'] = data['comment'].apply(clean_text)

# Menampilkan contoh hasil
print(data[['comment', 'cleaned_comment']].head())
```

Gambar 4.5. Kode Program tahap Cleaned Data

2. Tokenisasi

Data teks yang telah dib[illegible]n kemudian dipecah menjadi kata-kata individual (token). Proses [illegible]si ini bertujuan untuk mempermudah analisis pada tingkat kata. be[illegible]e program cleaned data pada gambar 12.

```
import nltk
nltk.download('punkt')

# Menyimpan data asli ke kolom baru agar dapat membandingkan sebelum dan sesudah tokenisasi
data['original_comment'] = data['comment']  # Simpan komentar asli sebelum dibersihkan
data['tokenized'] = data['cleaned_comment'].apply(nltk.word_tokenize)  # Tokenisasi

# Menyimpan hasil tokenisasi ke CSV, termasuk kolom before (original) dan after (cleaned + tokenized)
data.to_csv("tokenized_data_with_comparison.csv", index=False)

print("Hasil tokenisasi beserta perbandingan disimpan ke 'tokenized_data_with_comparison.csv'")
print(data[['original_comment', 'cleaned_comment', 'tokenized']].head())
```

Gambar 4.6. Kode Program tahap Tokenisasi Data

3. Stopword Removal

Kata-kata umum seperti "dan", "atau", "yang", yang tidak memiliki kontribusi signifikan terhadap analisis sentimen dihapus dari teks. Langkah ini membantu meningkatkan fokus analisis pada kata-kata bermakna. Berikut kode program Stopword Removal pada gambar 13.

```
# Memastikan kolom 'cleaned_comment' tidak mengandung NaN dan diubah ke format string
data['cleaned_comment'] = data['cleaned_comment'].fillna("").astype(str)

# Jika hasil tokenisasi dalam bentuk list (contoh: ['kata1', 'kata2']), ubah menjadi string
if data['cleaned_comment'].iloc[0].startswith("["):
    data['cleaned_comment'] = data['cleaned_comment'].apply(lambda x: " ".join(eval(x)))

# Menghapus stopwords
stopword_factory = StopWordRemoverFactory()
stopword_remover = stopword_factory.create_stop_word_remover()
data['cleaned_comment'] = data['cleaned_comment'].apply(stopword_remover.remove)

# Menyimpan hasil penghapusan stopwords ke file CSV
data.to_csv("stopwords_removed_data.csv", index=False)
print("Hasil stopword removal disimpan ke 'stopwords_removed_data.csv'")

# Menampilkan contoh hasil
print(data[['cleaned_comment']].head())
```

Gambar 4.7. Kode Program tahap Stopword Removal Data

4. Stemming

Untuk mengurangi variasi kata yang memiliki makna serupa, dilakukan proses stemming, yaitu mengubah kata-kata menjadi bentuk dasarnya. Stemming

membantu mengurangi redundansi dan meningkatkan konsistensi data. Berikut kode program Stemming pac ar 14.

```
# 4. Stemming
stemmer = StemmerFactory( temmer()
data['stemmed_comment'] = aned_comment'].apply(stemmer.stem)

# Menyimpan hasil stemming ke CSV
data.to_csv("stemmed_data.csv", index=False)
print("Hasil stemming disimpan ke 'stemmed_data.csv'")
print(data[['comment', 'stemmed_comment']].head())
```

Gambar 4.8. Kode Program tahap Stemming Data

Berikut contoh hasil proprecessing dari kode program di atas pada table 3

Table 4.1. Hasil contoh proprecessing

comment	Cleaned comment	tokenized	Stopword	Stemmed comment
tombol gk prnh minat n gk prnh pake produk apple ---------- >>>>>>>>>>>>>>>>>>>>>>>>	tombol prnh minat prnh pake produk apple	['tombol', 'prnh', 'minat', 'prnh', 'pake', 'produk', 'apple']	Gk,n	tombol prnh minat prnh pake produk apple
Blokir aj, lah ngalunjak bgt ipon, emg siape lu😂	Blokir lah ngalunjak bgt ipon emg siape	['Blokir', 'lah', 'ngalunjak', 'bgt', 'ipon', 'emg', 'siape']	aj, lah, bgt, emg, lu	blokir lah ngalunjak bgt ipon emg siape
Untung saya pake mito	Untung pake mito	['Untung', 'saya', 'pake', 'mito']	saya	untung pake mito
Bisa apa iphone tanpa indonesia 🔥	Bisa apa iphone indonesia	['Bisa', 'apa', 'iphone', 'tanpa', 'indonesia']	tanpa	bisa apa iphone indonesia
Gara2 iphone banyak yg jual ginjal🤣🤣🤣	Gara2 iphone banyak jual ginjal	['Gara2', 'iphone', 'banyak', 'jual', 'ginjal']	yg	gara2 iphone banyak jual ginjal

4.2.2. Pelabelan Data

Pelabelan data adalah lan penting dalam penelitian ini untuk mengklasifikasikan sentimen d ntar Instagram terhadap kebijakan *Apple tax holiday* selama 50 tahun. mentar yang telah dibersihkan atau di preprocesing yang sebelumnya sebanyak 3.791 data menjadi 3458 data diberi label sentimen berdasarkan isi teksnya. Sentimen ini dikategorikan menjadi tiga kelompok utama:

1. Positif (1): Komentar yang mendukung atau memberikan tanggapan yang setuju terhadap kebijakan tersebut.
2. Negatif (-1): Komentar yang menentang atau menunjukkan ketidaksetujuan terhadap kebijakan tersebut.
3. Netral (0): Komentar yang tidak secara langsung mendukung maupun menentang kebijakan tersebut, seperti pertanyaan atau komentar informatif.

Proses pelabelan dilakukan dengan menggunakan pendekatan berbasis kata kunci (keyword-based approach). Kata kunci yang relevan dengan masing-masing kategori sentimen telah didefinisikan sebelumnya. Misalnya:

1. Untuk sentimen positif, kata kunci seperti *"setuju"*, *"mantap"*, dan *"dukung"* digunakan.
2. Untuk sentimen negatif, kata kunci seperti *"boikot"*, *"memalukan"*, dan *"rugi"* digunakan.
3. Untuk sentimen netral, kata kunci seperti *"berapa harga"*, *"opsi lain"*, dan *"cek spek"* digunakan.

Proses ini menghasilkan dataset yang telah terlabeli, yang kemudian digunakan untuk tahap analisis lebih lanjut. Berikut program dari labeling yang digunakan pada gambar 15

```
import pandas as pd
# Membaca dataset
data = pd.read_csv("stemmed_data.

# Pastikan kolom 'comment' tidak          ilai NaN dan tipe datanya adalah string
data['comment'] = data['comment']       ).astype(str)

# Membuat kolom 'sentiment' dengan       fault 0 (Netral)
data['sentiment'] = 0
```

Gambar 4.9. Tahap awal kode program labeling

```
# Daftar kata kunci untuk sentimen (diperbarui sesuai data tambahan)
keywords_positive = [
    "setuju", "dukung", "mantap", "cinta produk indonesia", "menyala", "hebat", "bagus",
    "android terbaik", "tim android bangga", "produksi lokal", "cocok", "solid langkah",
    "ganteng kali bapak", "indonesia maju", "patriotisme", "tujuan baik", "bukti cinta indonesia",
    "adalah solusi", "rakyat kerja", "bangga indonesia", "samsung sejati", "produk lokal unggul",
    "kualitas indonesia", "peluang kerja", "tenaga lokal", "transfer teknologi", "bangga",
    "mantap jiwa", "puas", "luar biasa", "proud", "tetap maju", "se7", "prabowo my king",
    "cinta tanah air", "nyala abangkuh", "tuju aja aku", "solid", "senang", "mantap pak"
]

keywords_negative = [
    "blokir", "boikot", "hina", "pelecehan", "jahanam", "memalukan", "tendang apple",
    "ngadi ngadi", "anjir", "tidak mau bayar pajak", "rugi", "lemot", "negara diinjak",
    "jangan tunduk", "produk asing", "bebas pajak", "gila permintaan", "memalukan negara",
    "tuman", "peras rakyat", "kena pajak", "bebas pajak 50 tahun", "tendang dari indonesia",
    "tidak setuju", "usaha besar", "tidak bayar pajak", "enak banget", "harga diri negara",
    "menghina", "jangan beri izin", "boikot produk", "kebijakan buruk", "nggak sehat",
    "penjajahan baru", "blokiiir", "penindasan", "minta aneh-aneh", "bikin rugi", "lawak",
    "ngapain", "blokir kabeh", "boikot iphone", "anjir", "sakau", "pajak mahal", "kacau"
]

keywords_neutral = [
    "berapa harga", "alternatif", "opsi lain", "cek spek", "info produk lokal",
    "diskusi terbuka", "kebijakan pajak", "perbandingan produk", "diskusi netral",
    "menimbang opsi", "cek harga", "hubungan bea cukai", "ekosistem teknologi",
    "harga dan fitur", "analisis lokal", "info investasi", "lapangan kerja",
    "produksi dalam negeri", "sumber daya", "tenaga kerja", "industri teknologi",
    "produk lain", "dari sisi", "pilihan menarik", "komen netral", "opsi lain",
    "diskusi", "tidak tahu", "butuh spek", "bingung", "kira kira", "gpp 50th",
    "lapangan kerja", "kebijakan pajak", "cek spek"
]
```

Gambar 4.10. Keyword program Labeling

```
# Fungsi untuk melabeli komentar
def label_sentiment(comment):
    # Periksa apakah ada ka                    positif
    if any(keyword in comme                    for keyword in keywords_positive):
        return 1  # Positif
    # Periksa apakah ada ka                    gatif
    elif any(keyword in com                    ) for keyword in keywords_negative):
        return -1  # Negati
    # Periksa apakah ada kata kunci netral
    elif any(keyword in comment.lower() for keyword in keywords_neutral):
        return 0  # Netral
    else:
        return 0  # Netral jika tidak ada kata kunci yang cocok

# Terapkan fungsi labeling ke kolom komentar
data['sentiment'] = data['comment'].apply(label_sentiment)

# Simpan dataset yang telah diberi label ke file CSV
data.to_csv("labeled1_data.csv", index=False)

# Tampilkan distribusi sentimen
print("Dataset dengan sentimen berhasil disimpan ke 'labeled_data.csv'")
print("Distribusi Sentimen:")
print(data['sentiment'].value_counts())
```

Gambar 4.11. Tahap Terakhir kode program labeling

Pada penelitian ini, dilakukan proses pelabelan sentimen terhadap komentar-komentar Instagram yang berkaitan dengan kebijakan Apple Tax Holiday selama 50 tahun di Indonesia. Komentar-komentar tersebut diklasifikasikan menjadi tiga kategori sentimen, yaitu:

1. Sentimen Positif (1): Komentar yang mendukung atau memberikan tanggapan positif terhadap kebijakan tersebut.
2. Sentimen Negatif (-1): Komentar yang menolak atau memberikan tanggapan negatif terhadap kebijakan tersebut.
3. Sentimen Netral (0): Komentar yang bersifat netral, tidak menunjukkan dukungan atau penolakan secara langsung.

Berikut adalah tiga contoh data yang telah dilabeli dengan masing-masing kategori sentimen. Data ini dipilih secara representatif untuk menunjukkan hasil dari proses pelabelan:

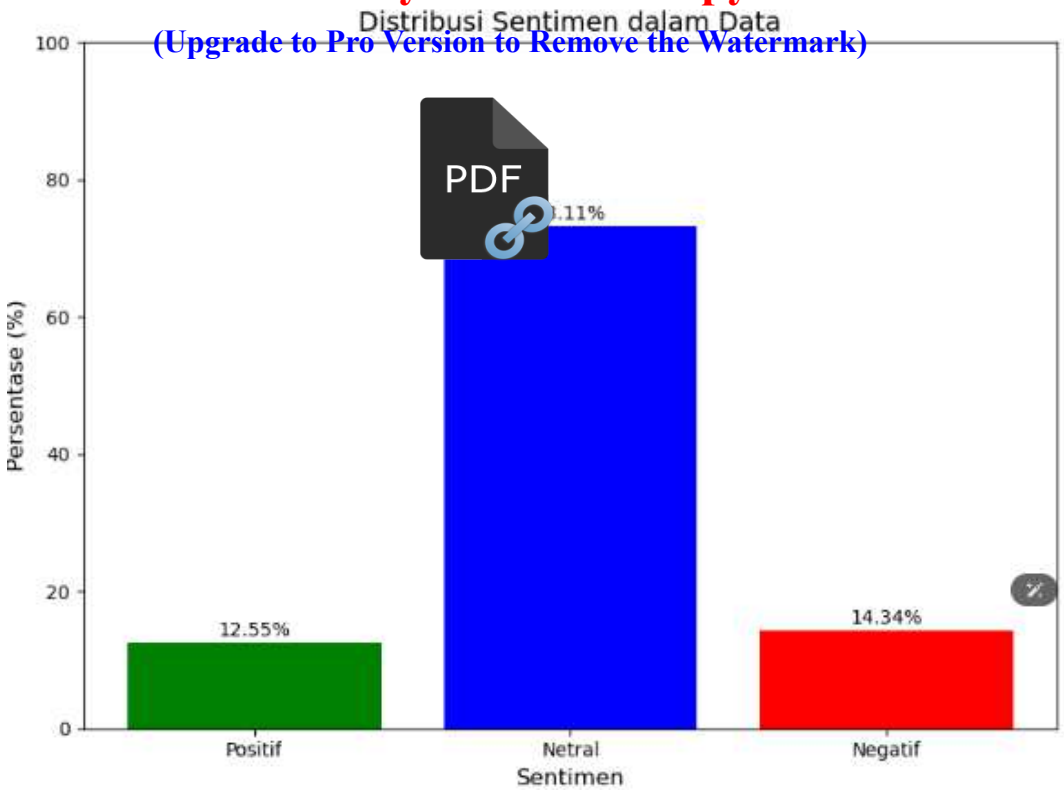
Table 4.2. Contoh labeling

Data Asli	Sentimen
Walau sy pakai iphone tp untuk diri nusa dan bangsa saya setuju blokir dan boikot saja	1
iPhone 16 aja di blokir yang udah masuk ya biarin	-1
tombol gk prnh minat n gk prnh pake produk apple ---------- >>>>>>>>>>>>>>>>>>>>>>>>>>>	0

Setelah pelabelan data selesai dilakukan, langkah berikutnya adalah menganalisis distribusi sentimen berdasarkan hasil labeling. Distribusi sentimen ini dihitung untuk mengetahui persentase komentar yang termasuk dalam kategori positif, netral, dan negatif. Hasil analisis ini memberikan gambaran mengenai proporsi sentimen dalam data penelitian, yang sangat penting untuk memahami pola sentimen secara keseluruhan. Selain itu, grafik distribusi sentimen juga digunakan untuk memberikan visualisasi yang lebih mudah dipahami mengenai frekuensi masing-masing sentimen.

Table 4.3. Persentase Komentar

Sentimen	Jumlah	Persentase
0 Positif	434	12.55%
1 Netral	2528	73.11%
2 Negatif	496	14.34%
3 Jumlah	3458	100%



Gambar 4.12. Grafik Persentase Komentar

Dari Tabel 5 dan Gambar 18 menunjukkan distribusi sentimen yang dihasilkan dari analisis komentar terkait kebijakan *Apple Tax Holiday* selama 50 tahun. Dari total 3.458 komentar yang sudah di preprocesing, sebagian besar menunjukkan sentimen netral dengan persentase sebesar 73,11%. Hal ini mengindikasikan bahwa sebagian besar pengguna memberikan respons yang bersifat informatif atau tidak menunjukkan dukungan maupun penolakan yang jelas terhadap kebijakan tersebut.

Sebaliknya, komentar dengan sentimen negatif tercatat sebesar 14,34%, yang mencerminkan kekhawatiran atau ketidaksetujuan terhadap kebijakan yang dianggap kontroversial ini. Sementara itu, hanya 12,55% komentar yang memiliki sentimen positif, menunjukkan adanya dukungan dari sebagian kecil pengguna terhadap kebijakan tersebut.

Hasil ini memberikan gambaran bahwa kebijakan *Apple Tax Holiday* telah memicu diskusi publik yang luas, meskipun banyak yang memilih untuk memberikan tanggapan netral.

Untuk memberikan gambaran yang lebih jelas mengenai distribusi kata yang sering muncul pada komen[illegible]gan label sentimen tertentu, dilakukan visualisasi menggunakan Word[illegible]sualisasi ini menampilkan kumpulan kata

yang paling sering muncul dalam setiap kategori sentimen positif, netral, dan negatif dengan ukuran kata mencerminkan frekuensi kemunculannya.

Gambar 4.13. Kumpulan data netral yang sering muncul

Gambar 4.14. Kumpulan data positif yang sering muncul

Gambar 4.15. Kumpulan data negatif yang sering muncul

4.3 Pembahasan

4.3.1 Penerapan Metode Anal Validitas Data

1. Random over sampling

   Salah satu tantangan yang dalam penelitian ini adalah distribusi data yang tidak seimbang, di mana jumlah komentar sentimen netral jauh lebih banyak dibandingkan sentimen positif dan negatif. Untuk mengatasi masalah ini, diterapkan teknik Random Over Sampling menggunakan pustaka imbalanced-learn. Teknik ini meningkatkan jumlah data pada kelas minoritas dengan menggandakan data secara acak hingga distribusi data menjadi lebih seimbang.

```python
import pandas as pd
from imblearn.over_sampling import RandomOverSampler
from collections import Counter

# Membaca dataset
data = pd.read_csv("labeled_dataset.csv")

# Memastikan kolom sentimen memiliki tipe data integer
data['sentiment'] = data['sentiment'].astype(int)

# Melihat distribusi awal
print("Distribusi Sebelum Oversampling:", Counter(data['sentiment']))

# Mengatur target distribusi
target_distribution = {0: 2576, 1: 1500, -1: 1200}

# Menggunakan Random Oversampler dengan target distribusi
ros = RandomOverSampler(sampling_strategy=target_distribution, random_state=42)

# Melakukan oversampling
X_resampled, y_resampled = ros.fit_resample(data[['comment']], data['sentiment'])

# Menggabungkan kembali data ke dalam DataFrame
oversampled_data = pd.DataFrame({'comment': X_resampled['comment'], 'sentiment': y_resampled})

# Melihat distribusi setelah oversampling
print("Distribusi Setelah Oversampling:", Counter(oversampled_data['sentiment']))

# Menyimpan hasil ke CSV
output_file = "oversampled_dataset.csv"
oversampled_data.to_csv(output_file, index=False)

print(f"Hasil oversampling disimpan ke file: {output_file}")
```

Gambar 4.16. Program Random Over Sampeling

Berikut hasil dari random over sampeling pada table 6

Table 4.4. Hasil Random Over Sampeling

Keterangan	Sebelum Over sampeling	Sesudah Over Sampeling
Netral	2576	2576
Positif	458	1500
Negatif	424	1200

Dengan penerapan Random Over Sampling, jumlah data dari setiap kelas menjadi seimbang, sehingga model da[illegible]lajar secara lebih baik tanpa terlalu bias terhadap kelas mayoritas.

2. TF-IDF

   Hasil transformasi data menggunakan TF-IDF memberikan representasi numerik yang mencerminkan pentingnya suatu kata dalam sebuah komentar dibandingkan dengan keseluruhan dataset. Setelah dilakukan Random Over Sampling untuk menyeimbangkan distribusi kelas sentimen, langkah TF-IDF ini menjadi sangat penting dalam mengubah data teks menjadi bentuk yang dapat dianalisis secara numerik oleh algoritma machine learning. Representasi ini memungkinkan model untuk memahami pola dan hubungan kata dalam setiap komentar berdasarkan frekuensi kemunculan dan relevansi kata tersebut terhadap keseluruhan dataset. Fitur-fitur dengan nilai TF-IDF tertinggi menunjukkan kata-kata yang memiliki peran signifikan dalam membedakan sentimen, sehingga dapat memberikan kontribusi yang besar dalam analisis sentimen terhadap kebijakan Apple terkait tax holiday 50 tahun.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer

# Membaca dataset yang sudah dilabel dan dioversample
data = pd.read_csv("oversampled_dataset.csv")

# Isi NaN pada kolom 'comment' dengan string kosong
data['comment'] = data['comment'].fillna("")

# Pastikan kolom komentar dan label tersedia
X = data['comment']
y = data['sentiment']

# Membagi data menjadi training dan testing (80:20)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

# Terapkan TF-IDF pada data teks
tfidf_vectorizer = TfidfVectorizer(max_features=5000)
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

print("TF-IDF diterapkan dengan max_features =", tfidf_vectorizer.max_features)
print("Jumlah data pada training set:", X_train_tfidf.shape[0])
print("Jumlah data pada testing set:", X_test_tfidf.shape[0])
```

Gambar 4.17. Program Tf-idf

Table 4.5. Contoh Tf-Idf

Kata - Kata	Nilai TF-IDF
wakil	0.577350
rakyat	0.577350
menyalaaaa	0.577350
dukung	0.500000
produk	0.500000

Hasil transformasi data menggunakan TF-IDF menunjukkan kata dengan nilai TF-IDF tertinggi pada contoh table 7, yang mencerminkan pentingnya kata-kata tersebut berdasarkan frekuensi kemunculan dan relevansi dalam dataset. Kata-kata ini memberikan wawasan tentang fitur-fitur utama yang digunakan dalam analisis sentimen, dengan nilai TF-IDF yang lebih tinggi menunjukkan kontribusi yang lebih besar terhadap representasi data. Tabel ini menyajikan kata-kata kunci yang paling signifikan, yang kemudian digunakan sebagai input untuk model Naive Bayes dan Random Forest dalam melakukan klasifikasi sentimen secara akurat.

4.3.2 Pengujian Hasil Analisa

Pengujian hasil analisa bertujuan untuk mengevaluasi kinerja model dalam mengklasifikasikan sentimen komentar terkait kebijakan *Apple Tax Holiday* selama 50 tahun. Pada tahap ini, model Naive Bayes dan Random Forest diuji menggunakan data latih dan data uji untuk menghasilkan metrik evaluasi seperti akurasi, presisi, *recall*, dan F1-score. Selain itu, dilakukan analisis menggunakan matriks kebingungan (*confusion matrix*) untuk memahami distribusi prediksi model terhadap data yang sebenarnya.

1. Pengujian dengan naïve bayes
   Naive Bayes adalah algoritma klasifikasi yang bekerja berdasarkan teorema Bayes dengan asumsi independensi antar fitur. Dalam analisis teks, algoritma ini menjadi pilihan populer karena kemampuannya menangani data berukuran besar dengan cepat dan efisien. Pada penelitian ini, Naive Bayes diterapkan pada dataset komentar untuk mengklasifikasikan sentimen positif, netral, dan negatif. Model ini mengandalkan probabilitas kondisi setiap kata dalam kelas tertentu untuk membuat prediksi, yang membuatnya ideal untuk dataset yang seimbang. Berikut pengujian nya :

```
# Melatih Model Naive Bayes
nb_model = MultinomialNB()
nb_model.fit(X_train_tfidf, y_t

# Evaluasi pada data training
nb_train_predictions = nb_model                train_tfidf)
print("Evaluasi Naive Bayes pad                ning:")
print("Accuracy:", accuracy_sco                 nb_train_predictions))
print("Classification Report:\n                 cation_report(y_train, nb_train_predictions))
print("Confusion Matrix:\n", confusion_matrix(y_train, nb_train_predictions))

# Evaluasi pada data testing
nb_test_predictions = nb_model.predict(X_test_tfidf)
print("\nEvaluasi Naive Bayes pada Data Testing:")
print("Accuracy:", accuracy_score(y_test, nb_test_predictions))
print("Classification Report:\n", classification_report(y_test, nb_test_predictions))
print("Confusion Matrix:\n", confusion_matrix(y_test, nb_test_predictions))
```

Gambar 4.18. Pemodelan dan evaluasi

Dari program tersebut didapatkan hasil seperti table berikut:

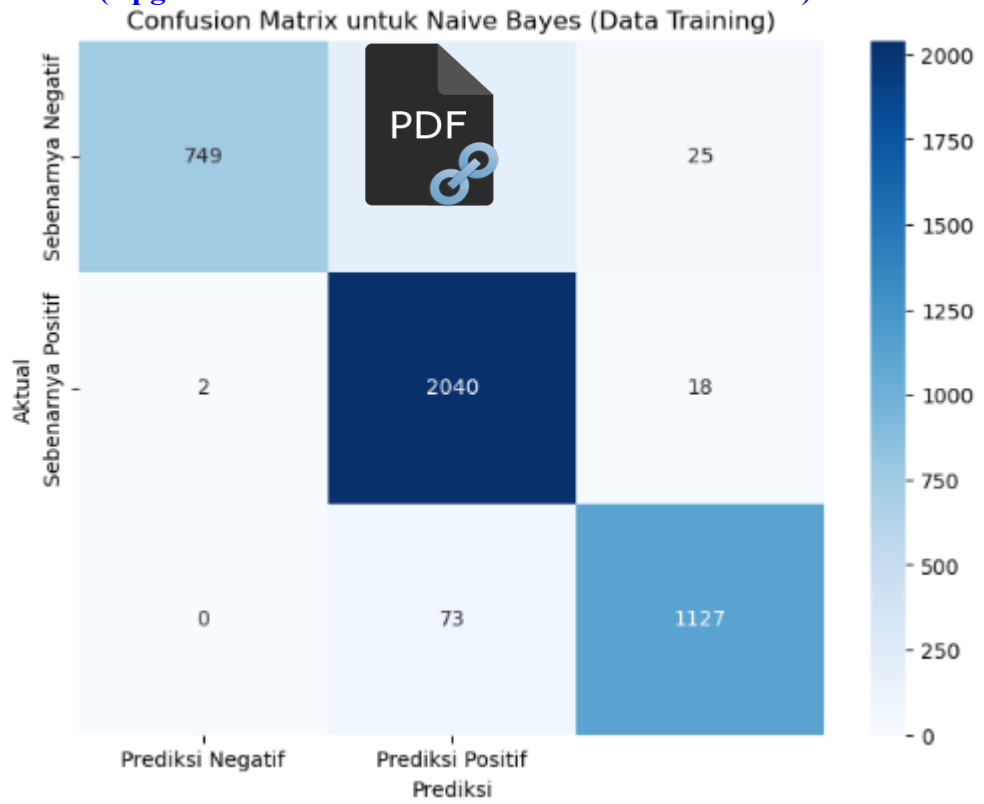
Table 4.6. Hasil pemodelan naïve bayes data training

Sentiment	Precision	Recall	F1-Score
-1	1.00	0.78	0.88
0	0.89	0.99	0.94
1	0.96	0.94	0.95
Accuracy	0.93		

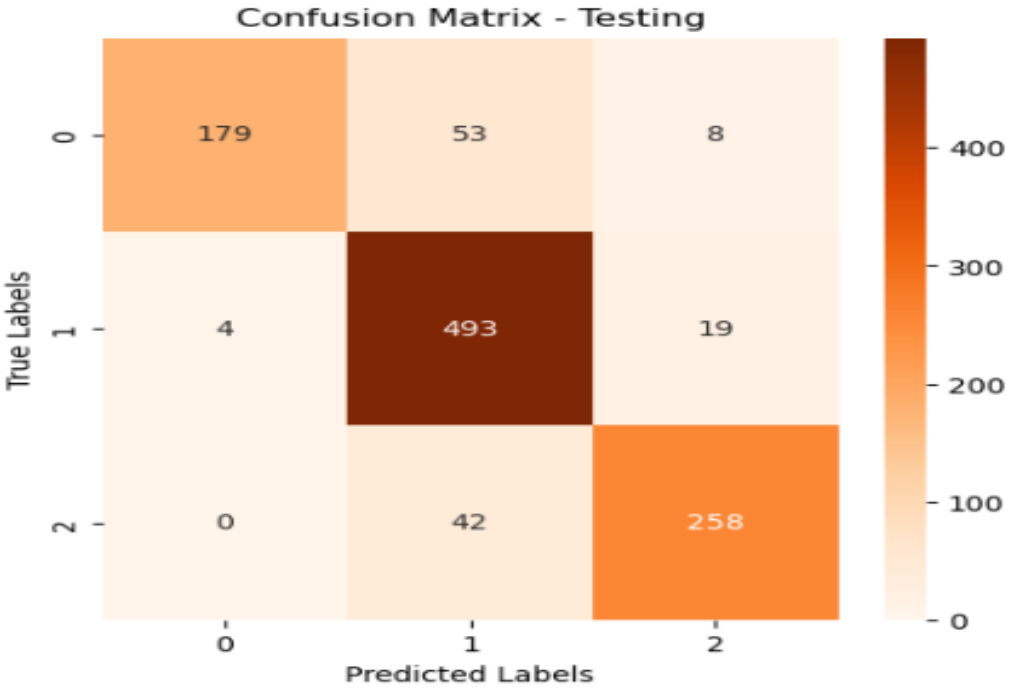
Table 4.7. Hasil pemodelan naïve bayes data testing

Sentiment	Precision	Recall	F1-Score
-1	0.98	0.75	0.85
0	0.84	0.96	0.89
1	0.91	0.86	0.88
Accuracy	0.88		

Dan untuk confusion matrix nya digambarkan pada gambar berikut:



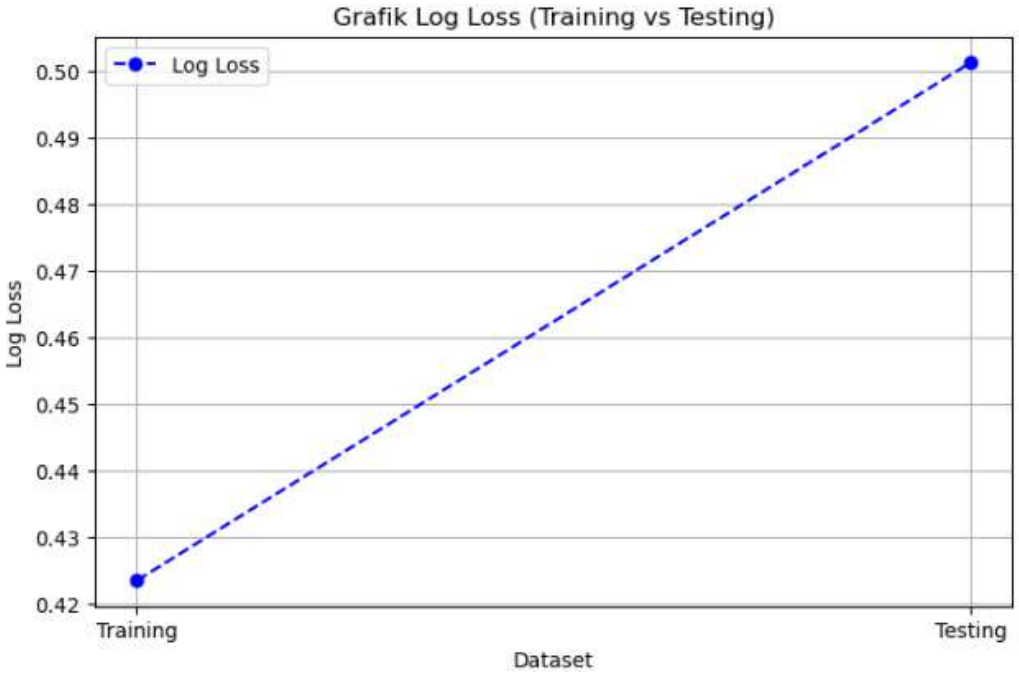
Gambar 4.19. Matrix confusion data training



Gambar 4.20. Matrix confusion data testing

Dari hasil evaluasi, model Naive Bayes menunjukkan performa yang sangat baik pada data training, dengan akurasi mencapai 92.8%. Model mampu mengenali hampir semua kelas dengan baik, terutama pada kelas 0 (netral) yang memiliki recall 0.99. Namun, pada kelas -1 (negatif), meskipun precision mencapai 1.00, recall-nya lebih rendah, yaitu 0.78, yang menunjukkan ada beberapa data negatif yang salah dikategorikan.

Pada data testing, model ini juga menunjukkan performa yang baik, dengan akurasi 88.1%. Meski demikian, recall untuk kelas -1 sedikit menurun menjadi 0.75 dibandingkan dengan data training, yang mungkin disebabkan oleh perbedaan distribusi antara data training dan testing. Untuk kelas 0 dan 1, model mampu mempertahankan recall yang tinggi, yaitu 0.96 dan 0.86, masing-masing. Secara keseluruhan, model Naive Bayes ini memberikan hasil yang memuaskan dengan performa yang cukup konsisten antara data training dan testing. Meskipun ada beberapa ruang untuk peningkatan, terutama pada kelas-kelas dengan data yang lebih tidak seimbang, model ini tetap memberikan hasil yang dapat diandalkan pada klasifikasi ini.



Gambar 4.21. Grafik log loss model naïve bayes

Pada gambar 4.20 model Naive Bayes, nilai Log Loss yang diperoleh adalah 0.4235 pada data training dan 0.5013 pada data testing. Log Loss mengukur seberapa baik probabilitas prediksi model terhadap kelas yang benar, di mana semakin kecil nilainya, semakin baik model dalam memberikan probabilitas yang mendekati nilai sebenarnya. Nilai 0.4235 pada data training menunjukkan bahwa model cukup baik dalam memprediksi probabilitas yang benar, karena mayoritas prediksi memiliki probabilitas tinggi terhadap kelas yang benar. Sementara itu, nilai 0.5013 pada data testing sedikit lebih tinggi, yang mengindikasikan adanya sedikit penurunan performa saat generalisasi ke data baru. Namun, selisihnya tidak terlalu besar, sehingga model tetap stabil dalam memprediksi data yang belum pernah dilihat sebelumnya. Dari hasil ini, dapat disimpulkan bahwa model Naive Bayes cukup cocok untuk digunakan dalam analisis ini karena mampu memberikan prediksi yang konsisten dan tidak mengalami overfitting yang signifikan.

2. Pengujian dengan random forest

   Random Forest adalah algoritma ensemble yang menggabungkan beberapa pohon keputusan (decision trees) untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting. Dalam analisis teks, Random Forest menjadi pilihan yang baik karena kemampuannya untuk menangani data yang kompleks dan banyak variabel secara efektif. Algoritma ini bekerja dengan membangun sejumlah besar pohon keputusan berdasarkan subset acak dari data dan fitur, lalu menggabungkan hasil dari semua pohon tersebut untuk membuat prediksi akhir.

   Pada penelitian ini, Random Forest diterapkan pada dataset komentar untuk mengklasifikasikan sentimen menjadi tiga kelas: positif, netral, dan negatif. Setiap pohon keputusan dalam Random Forest berkontribusi pada proses klasifikasi, dan prediksi akhir dihitung berdasarkan hasil mayoritas dari pohon-pohon tersebut. Dengan menggunakan teknik ensemble, Random Forest mampu mengurangi variabilitas model yang dapat muncul dari pohon keputusan tunggal, memberikan kestabilan dan akurasi yang lebih tinggi pada prediksi sentimen. Berikut pengujian nya:

```python
from sklearn.ensemble import RandomForestClassifier
from imblearn.over_sampling import SMOTE
from sklearn.calibration i    alibratedClassifierCV

# Inisialisasi Random Fore          tuning
rf_model = RandomForestCla
    random_state=42,
    n_estimators=200,    # Meningkatkan jumlah pohon untuk stabilitas
    max_depth=12,        # Mengurangi kedalaman agar tidak terlalu overfit
    min_samples_split=10, # Meningkatkan nilai untuk menghindari overfitting
    min_samples_leaf=10,  # Meningkatkan agar model tidak terlalu yakin
    max_features='sqrt',
    class_weight='balanced'  # Menangani kelas tidak seimbang
)

# Melatih model Random Forest
rf_model.fit(X_train_tfidf, y_train)

# Kalibrasi probabilitas dengan Platt Scaling
calibrated_rf = CalibratedClassifierCV(rf_model, method='sigmoid', cv=5)
calibrated_rf.fit(X_train_tfidf, y_train)

print("Model Random Forest dengan tuning dan kalibrasi telah dilatih!")
```

Gambar 4.22. Program pelatihan model dan probability calibration

```python
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, log_loss

# Prediksi probabilitas setelah kalibrasi
rf_train_proba = calibrated_rf.predict_proba(X_train_tfidf)
rf_test_proba = calibrated_rf.predict_proba(X_test_tfidf)

# Prediksi kelas
rf_train_preds = calibrated_rf.predict(X_train_tfidf)
rf_test_preds = calibrated_rf.predict(X_test_tfidf)

# Evaluasi pada Data Training
train_accuracy = accuracy_score(y_train, rf_train_preds)
train_log_loss = log_loss(y_train, rf_train_proba)
print("Evaluasi Random Forest setelah kalibrasi pada Data Training:")
print("Accuracy:", train_accuracy)
print("Log Loss:", train_log_loss)
print("Classification Report:\n", classification_report(y_train, rf_train_preds))
print("Confusion Matrix:\n", confusion_matrix(y_train, rf_train_preds))

# Evaluasi pada Data Testing
test_accuracy = accuracy_score(y_test, rf_test_preds)
test_log_loss = log_loss(y_test, rf_test_proba)
print("\nEvaluasi Random Forest setelah kalibrasi pada Data Testing:")
print("Accuracy:", test_accuracy)
print("Log Loss:", test_log_loss)
print("Classification Report:\n", classification_report(y_test, rf_test_preds))
print("Confusion Matrix:\n", confusion_matrix(y_test, rf_test_preds))

# ====== GRAFIK CONFUSION MATRIX ======
fig, ax = plt.subplots(1, 2, figsize=(12, 5))

# Confusion Matrix Training
sns.heatmap(confusion_matrix(y_train, rf_train_preds), annot=True, fmt='d', cmap='Blues', ax=ax[0])
ax[0].set_title('Confusion Matrix - Training')
ax[0].set_xlabel('Predicted Label')
ax[0].set_ylabel('True Label')

# Confusion Matrix Testing
sns.heatmap(confusion_matrix(y_test, rf_test_preds), annot=True, fmt='d', cmap='Oranges', ax=ax[1])
ax[1].set_title('Confusion Matrix - Testing')
ax[1].set_xlabel('Predicted Label')
ax[1].set_ylabel('True Label')

plt.show()

# ====== GRAFIK LOG LOSS ======
epochs = range(1, 3)  # Untuk tampilan sederhana 2 nilai (train & test)
log_loss_values = [train_log_loss, test_log_loss]

plt.figure(figsize=(6, 4))
plt.plot(epochs, log_loss_values, marker='o', linestyle='-', color='red', label='Log Loss')
plt.xticks([1, 2], ['Training', 'Testing'])
plt.xlabel('Dataset')
plt.ylabel('Log Loss')
plt.title('Log Loss Comparison')
plt.legend()
plt.grid(True)
plt.show()
```

Gambar 4.23. Program evaluasi dan visualisasi data

Dari program tersebut didapatkan hasil seperti table berikut:

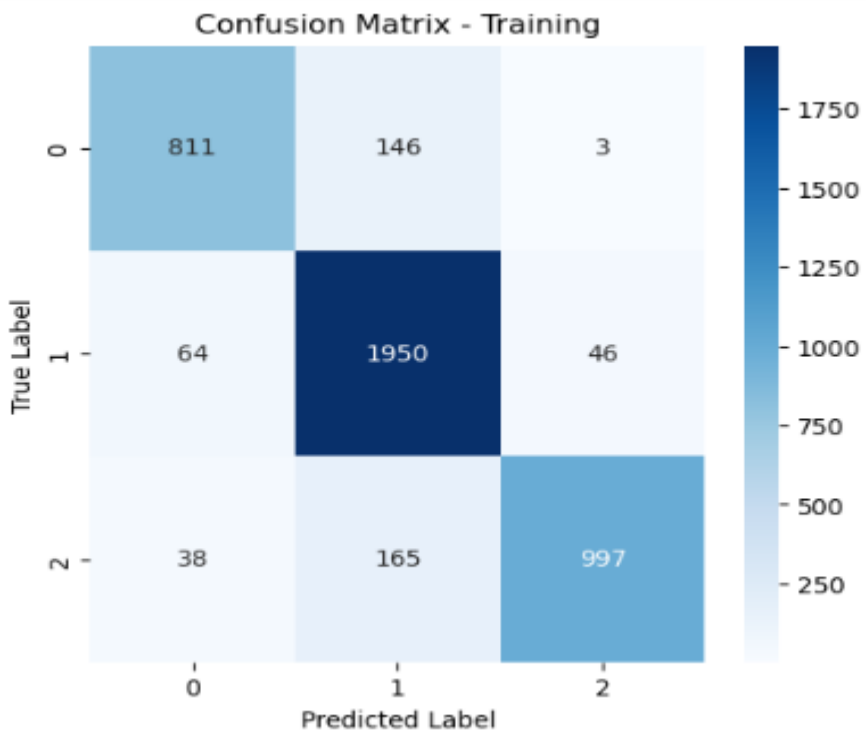
Table 4.8. Hasil pemodelan forest data training

Sentiment	Precisi	Recall	F1-Score
-1	0.89	0.84	0.87
0	0.86	0.95	0.90
1	0.95	0.83	0.89
Accuracy	0.89		

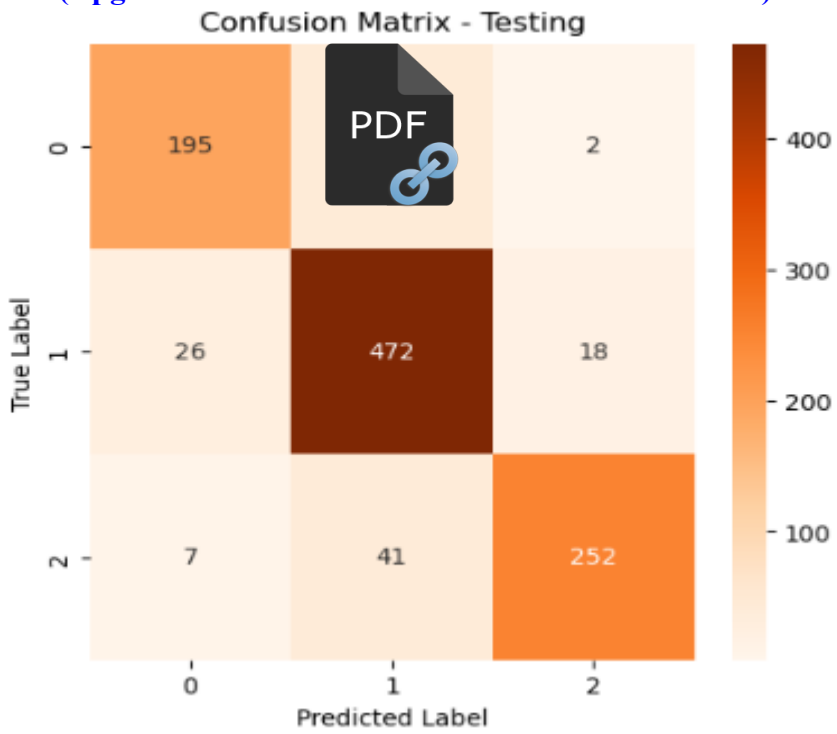
Table 4.9. Hasil pemodelan random forest data testing

Sentiment	Precision	Recall	F1-Score
-1	0.86	0.81	0.83
0	0.85	0.91	0.88
1	0.93	0.84	0.88
Accuracy	0.87		

Dan untuk confusion matrix nya digambarkan pada gambar berikut:



Gambar 4.24. Confusion matrix data training

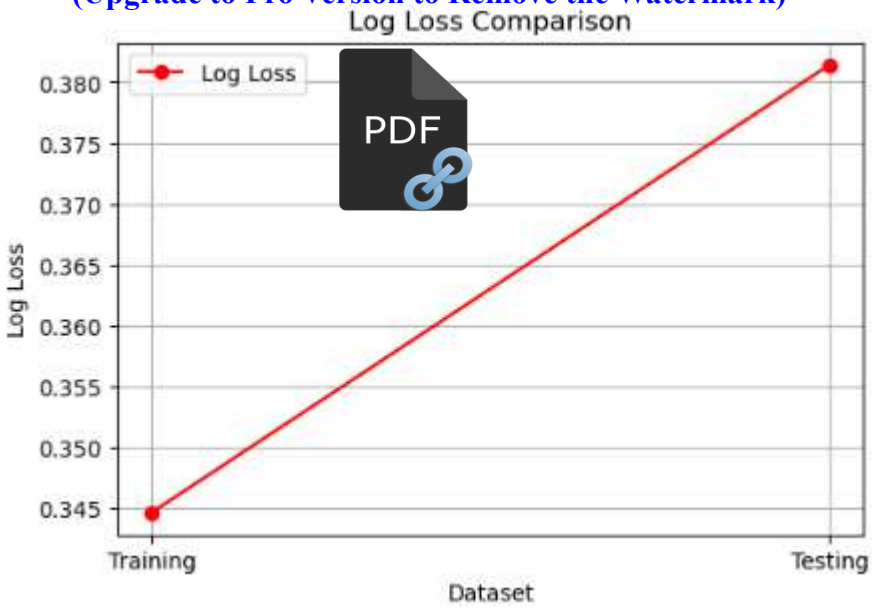


Gambar 4.25. Confusion matrix data testing

Berdasarkan hasil evaluasi, algoritma Random Forest menunjukkan kinerja yang cukup baik dengan akurasi sebesar 89.05%% pada data training dan 87.03% pada data testing. Model ini mampu mengklasifikasikan sentimen positif, netral, dan negatif dengan rata-rata precision, recall, dan F1-score yang tinggi pada kedua dataset.

Perbedaan akurasi antara data training dan data testing menunjukkan bahwa model memiliki generalisasi yang cukup baik tanpa indikasi overfitting yang signifikan. Namun, terdapat ruang untuk peningkatan performa, terutama pada kelas tertentu yang memiliki precision atau recall lebih rendah. Hal ini dapat diatasi dengan tuning parameter yang lebih mendalam atau penyesuaian teknik praproses pada data.

Secara keseluruhan, model Random Forest berhasil memberikan hasil yang konsisten dan dapat diandalkan untuk tugas klasifikasi sentimen pada dataset yang digunakan.



Gambar 4.26. Gambar log loss model random forest

Pada gambar 4.24 evaluasi model Random Forest setelah kalibrasi, nilai Log Loss yang diperoleh adalah 0.3446 pada data training dan 0.3814 pada data testing. Log Loss mengukur seberapa baik probabilitas prediksi model terhadap kelas yang benar, di mana semakin kecil nilainya, semakin baik model dalam memberikan probabilitas yang mendekati nilai sebenarnya. Nilai 0.3446 pada data training menunjukkan bahwa model mampu menghasilkan prediksi dengan probabilitas yang cukup baik terhadap kelas yang benar. Sementara itu, nilai 0.3814 pada data testing sedikit lebih tinggi, yang menunjukkan adanya sedikit penurunan performa saat diterapkan pada data baru. Namun, selisihnya tetap kecil, menandakan bahwa model memiliki generalisasi yang baik dan tidak mengalami overfitting yang signifikan. Dari hasil ini, dapat disimpulkan bahwa model Random Forest yang telah dikalibrasi cukup cocok untuk digunakan dalam analisis ini karena mampu memberikan prediksi yang stabil dan akurat dengan probabilitas yang lebih terkalibrasi.

3. Perbandingan algoritma naïve bayes dan random forest

Tabel 4.10. Perbanding[illegible]rmansi Naive Bayes dan Random Forest

Metrik	D[illegible]	Naive Bayes	Random Forest
Akurasi	Training	92.79%	89.05%
	Testing	88.07%	87.03%
Precision (-1)	Training	100%	89%
	Testing	98%	86%
Recall (-1)	Training	78%	84%
	Testing	75%	81%
Precision (0)	Training	89%	86%
	Testing	84%	85%
Recall (0)	Training	99%	95%
	Testing	96%	91%
Precision (1)	Training	96%	95%
	Testing	91%	93%
Recall (1)	Training	94%	83%
	Testing	86%	84%
F1-Score (-1)	Testing	85%	83%
F1-Score (0)	Testing	89%	88%
F1-Score (1)	Testing	88%	88%

4. Perbandingan Evaluasi

1. Akurasi:
   a. Training: Naive B ih unggul dengan akurasi 92.79%, dibandingkan Rando ang mencapai 89.05%.
   b. Testing: Pada data testing, Naive Bayes juga sedikit lebih baik dengan akurasi 88.07% dibandingkan Random Forest yang mendapatkan 87.03%. Ini menunjukkan bahwa Naive Bayes mungkin lebih mampu mengeneralisasi data pada testing set.
2. Precision (Kemampuan model untuk memprediksi kelas positif dengan benar):
   a. -1 (Negatif): Naive Bayes memiliki precision yang sempurna (100%) pada data training dan sangat baik pada data testing (98%). Sebaliknya, Random Forest memiliki precision lebih rendah, yaitu 89% untuk training dan 86% untuk testing.
   b. 0 (Netral): Naive Bayes juga sedikit lebih baik dengan precision 89% untuk training dan 84% untuk testing, sedangkan Random Forest berada di angka 86% untuk training dan 85% untuk testing.
   c. 1 (Positif): Precision untuk kelas positif juga menunjukkan keunggulan Naive Bayes, dengan 96% pada data training dan 91% pada data testing, dibandingkan dengan 95% pada training dan 93% pada testing untuk Random Forest.
3. Recall (Kemampuan model untuk menangkap semua kelas positif yang sebenarnya):
   a. -1 (Negatif): Di sini, Random Forest lebih unggul dalam recall, dengan 84% untuk training dan 81% untuk testing, dibandingkan Naive Bayes yang hanya memiliki 78% untuk training dan 75% untuk testing.
   b. 0 (Netral): Naive Bayes kembali unggul dalam recall, dengan 99% pada training dan 96% pada testing, sementara Random Forest hanya mencapai 95% pada training dan 91% pada testing.

c. 1 (Positif): Untuk recall kelas positif, Naive Bayes masih lebih baik dengan 94% untuk tr dan 86% untuk testing, sedangkan Random Forest memiliki 83% ning dan 84% pada testing.

4. F1-Score (Rata-rata ha ntara precision dan recall, memberikan gambaran keseluruhan performa model):
   a. -1 (Negatif): F1-score untuk kelas negatif menunjukkan sedikit perbedaan, dengan Naive Bayes mencapai 85% pada testing, sedangkan Random Forest hanya 83%.
   b. 0 (Netral): F1-score untuk kelas netral sedikit lebih unggul pada Naive Bayes, dengan 89% pada testing dibandingkan 88% pada Random Forest.
   c. 1 (Positif): Untuk kelas positif, kedua model memiliki nilai F1-score yang sama (88%) pada testing, menunjukkan performa yang hampir setara pada kelas ini.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan pada analisis sentimen menggunakan algoritma Naive Bayes dan Random Forest, dapat disimpulkan bahwa:

1. Naive Bayes menunjukkan akurasi yang sedikit lebih tinggi pada data training (92.79%) dibandingkan Random Forest (89.05%), tetapi hasilnya tetap kompetitif pada data testing dengan akurasi sebesar 88.07%. Hal ini mengindikasikan bahwa Naive Bayes mampu memanfaatkan pola probabilistik dengan baik, terutama pada dataset yang telah di-random oversampling.
2. Random Forest memberikan performa yang lebih stabil pada kelas netral (0) dalam metrik precision dan recall pada tabel perbandingan 4.10, setelah dilakukan probability calibration untuk meningkatkan keakuratan estimasi probabilitas, terutama pada data testing. Meskipun begitu, Naive Bayes tetap lebih unggul dalam beberapa metrik lainnya.
3. Naive Bayes lebih unggul dalam hal precision dan recall untuk kelas -1 (negatif) dan 1 (positif), dengan hasil yang lebih stabil pada kelas-kelas tersebut dibandingkan Random Forest. Meskipun Random Forest memiliki keunggulan dalam menangani kelas netral, secara keseluruhan Naive Bayes menunjukkan performa yang lebih baik dalam klasifikasi sentimen.
4. Teknik pemrosesan dataset, seperti random oversampling, sangat membantu dalam menyeimbangkan distribusi kelas dan meningkatkan representasi data untuk kelas minoritas.Ini berdampak positif pada performa Naive Bayes, terutama dalam menangani kelas minoritas yang sebelumnya sulit diprediksi dengan baik.
5. Naive Bayes lebih sederhana dalam implementasi dan lebih cepat dalam proses pelatihan, sementara Random Forest memerlukan probability calibration untuk meningkatkan keakuratan prediksi probabilitas. Meskipun demikian, kedua algoritma ini efektif untuk tugas analisis sentimen. Namun, Naive Bayes

memiliki keunggulan dalam akurasi dan konsistensi performa, terutama dalam kondisi dataset yang telah di[illegible] oversampling.

5.2 Saran

Berdasarkan hasil perbandingan[illegible]goritma Naive Bayes dan Random Forest dalam analisis sentimen, beberapa temuan penting dapat diambil. Meskipun kedua algoritma memberikan performa yang baik, terdapat area yang dapat diperbaiki dan dieksplorasi lebih lanjut untuk mencapai hasil yang lebih optimal. Oleh karena itu, beberapa saran berikut dapat dipertimbangkan untuk penelitian selanjutnya:

1. Dataset yang lebih besar dan distribusi kelas yang seimbang secara alami sangat disarankan untuk penelitian berikutnya. Meskipun random oversampling membantu menyeimbangkan kelas, dataset yang lebih besar dapat mengurangi risiko overfitting dan meningkatkan akurasi hasil analisis, serta memperkecil bias sampling pada model.
2. Eksperimen dengan lebih banyak teknik ensemble, seperti Gradient Boosting atau XGBoost, dapat dilakukan untuk Random Forest untuk mendapatkan performa yang lebih baik. Selain itu, probability calibration yang telah diterapkan pada model juga bisa diuji lebih lanjut dengan kombinasi hyperparameter lainnya untuk meningkatkan akurasi model pada data testing.
3. Naive Bayes sangat cocok untuk implementasi cepat pada dataset besar dengan distribusi data yang relatif sederhana. Namun, pada dataset yang lebih kompleks atau tidak terstruktur, model ini mungkin memerlukan teknik seleksi fitur atau pendekatan lebih lanjut seperti analisis mendalam terhadap distribusi data untuk memperoleh hasil yang optimal.
4. Penggunaan teknik preprocessing tambahan, seperti stemming, lemmatization, atau embedding (misalnya Word2Vec atau BERT), dapat meningkatkan representasi teks dan memperbaiki hasil analisis, terutama dalam menangani data teks yang lebih kompleks.

   Analisis ini bisa dikembangkan untuk diuji pada data dari berbagai platform sosial media. Ini akan membantu memvalidasi performa model di berbagai konteks data yang berbeda, dan memberikan gambaran tentang generalisasi model pada berbagai karakteristik dan pola bahasa yang ada di berbagai platfor

DAFTAR PUSTAKA

[1] S. R. Akbar and C. Kun[illegible]garuh Tax Holiday, Tax Allowance dan Inflasi terhadap Foreign Direct Investment di Indonesia,” *Innov. J. Soc. Sci. Res.*, vol. 3, no. 4, pp. 4741–4750, 2023, [Online]. Available: http://j-innovative.org/index.php/Innovative/article/view/3465%0Ahttps://j-innovative.org/index.php/Innovative/article/download/3465/2914

[2] C. A. Pohan, N. Rahmi, P. Arimbhi, I. Mawarni, M. Apriliani, and J. Tembaru, “Jurnal Reformasi Administrasi : Jurnal Ilmiah untuk Mewujudkan Masyarakat Madani Evaluasi Efektivitas Kebijakan Tax Holiday Dalam Meningkatkan Investasi di Indonesia,” vol. 8, no. 1, pp. 85–96, 2021, [Online]. Available: http://ojs.stiami.ac.id

[3] S. Kusuma Wardani and Y. Arum Sari, “Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5582–5590, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[4] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, “Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 122–130, 2022, [Online]. Available: http://ojs.serambimekkah.ac.id/jnkti/article/view/4004/pdf

[5] D. Mualfah, A. Prihatin, R. Firdaus, and Sunanto, “Analisis Sentimen Masyarakat Terhadap Kasus Pembobolan Data Nasabah Bank BSI Pada Twitter Menggunakan Metode Random Forest Dan Naïve Bayes,” *J. Fasilkom*, vol. 13, no. 3, pp. 614–620, 2024, doi: 10.37859/jf.v13i3.6478.

[6] F. D. A. Meisty, D. Anggraeni, and M. Fatekurohman, “Perbandingan Metode Naïve Bayes Classifier dengan Metode Random Forest pada Prediksi Rating Review Drama Korea,” *Estimasi J. Stat. Its Appl.*, vol. 5, no. 1, pp. 84–95, 2024, doi: 10.20956/ejsa.v5i1.26942.

[7] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritm dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaa istrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comp* vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.

[8] V. A. and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," *Int. J. Comput. Appl.*, vol. 139, no. 11, pp. 5–15, 2016, doi: 10.5120/ijca2016908625.

[9] Tommy Suhendra, B. Intan, and A. T. Martadinata, "Analisis Sentimen Pengguna Aplikasi Netflix Pada Ulasan Google Playstore Menggunakan Metode Naïve Bayes Tommy," *ESCAF 3rd*, vol. 2, pp. 1011–1022, 2024.

[10] A. Wandani, "Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 651–665, 2021.

[11] J. A. Pratama, Y. Suprijadi, and Z. Zulhanif, "The Analisis Sentimen Sosial Media Twitter Dengan Algoritma Machine Learning Menggunakan Software R," *J. Fourier*, vol. 6, no. 2, p. 85, 2017, doi: 10.14421/fourier.2017.62.85-89.

[12] A. A. A. Sumanjaya, Indriati, and A. Ridok, "Analisis Sentimen Data Tweets terhadap Penanganan Covid-19 di Indonesia menggunakan Metode Naïve Bayes dan Pemilihan Kata Bersentimen menggunakan Lexicon Based," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 1865–1872, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[13] Rayuwati, Husna Gemasih, and Irma Nizar, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.

[14] Marlina Haiza, Elmayati, Zulius Antoni, and Wijaya Harma Oktafia Lingga, "Penerapan Algoritma Random Forest Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo," *Penerapan Kecerdasan Buatan*, vol. 4, no. 2, pp. 138–143, 2023.

[15] M. Fachriza and Munawar, "Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Bayes, Support Vector Machine, Random Forest," *J. Tek. Univ. Muhammadiyah Ponorogo*, pp. 49–58, 2023, [Online]. Available: http://journal.umpo.ac.id/index.php/komputek

[16] M. C. Auliya Rahman Isnain, S. Kom., "opini cara menganalisis sentimen pada data di media sosial No Title," Universitas Teknokrat Indonesia. [Online]. Available: https://teknokrat.ac.id/opini-cara-menganalisis-sentimen-pada-data-di-media-sosial/

[17] R. Zulfiqri, B. N. Sari, T. N. Padilah, U. S. Karawang, and T. Timur, "Media Sosial Instagram Pada Situs Google Play Store Bayes," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 12, no. 3, 2024.

[18] Zidni Hudan Said Purnomo, "Mengenal Kebijakan Tax Holiday dan Tax Allowance," *Artik. Web*, 2021, [Online]. Available: https://pajak.go.id/index.php/id/artikel/mengenal-kebijakan-tax-holiday-dan-tax-allowance

[19] N. A. Prakoso Indaryono, "Analisa Perbandingan Algoritma Random Forest Dan Naïve Bayes Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 1, pp. 158–167, 2024, doi: 10.29100/jipi.v9i1.4421.

[20] B. Yusuf, M. Qalbi, B. Basrul, I. Dwitawati, M. Malahayati, and M. Ellyadi, "Implementasi Algoritma Naive Bayes Dan Random Forest Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh," *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 4, no. 1, p. 50, 2020, doi: 10.22373/cj.v4i1.7247.

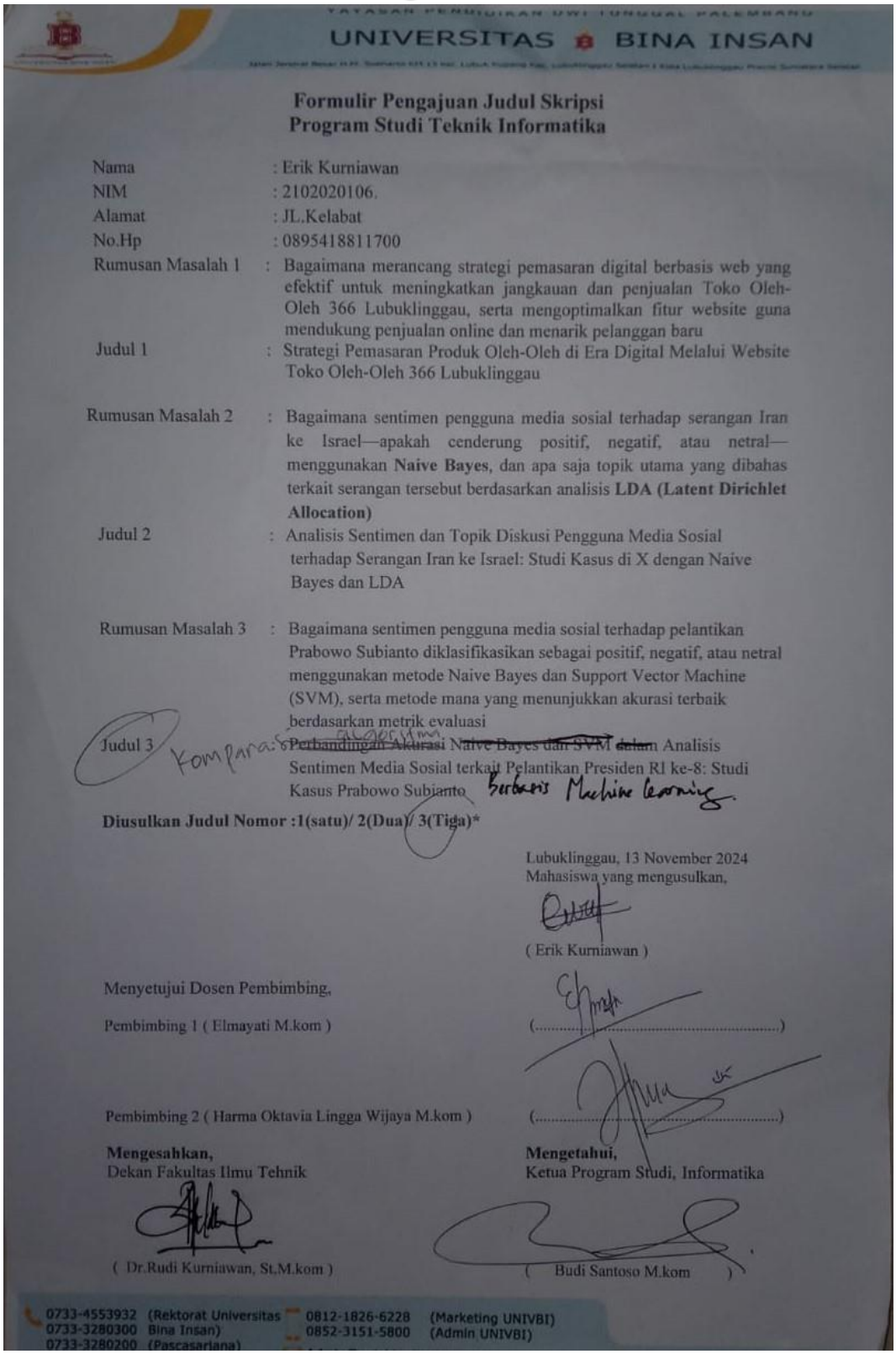
[21] R. Leonardo, J. Pratama, and C. Chrisnatalis, "Perbandingan Metode Random Forest Dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing," *J. Teknol. Dan Ilmu Komput. Prima*, vol. 3, no. 2, pp. 455–459, 2020, doi: 10.34012/jutikomp.v3i2.1321.

[22] A. Z. Syahputri, F. Della Fallenia, and R. Syafitri, "Kerangka berfikir penelitian kuantitatif," *Tarb. J. Ilmu Pendidik. dan Pengajaran*, vol. 2, no. 1, pp. 160–166, 2023.

[23] Satrianansyah, K. Adha, and N. Lestari, "Analisi Tingkat Keamanan Sistem Ams Pada Universitas I asan Lubuklinggau Menggunakan Cobit 5 Dengan Domain Dss05 enelitian ini akan menggunakan metode COBIT 5 sebagai standa keamanan teknologi informasi . Cobit 5 (Control Objective," *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 7, no. 1, pp. 47–59, 2022.

LAMPIRAN

Lampiran 1.Form pengajuan ju



UNIVERSITAS BINA INSAN

Formulir Pengajuan Judul Skripsi
Program Studi Teknik Informatika

Nama : Erik Kurniawan
NIM : 2102020106.
Alamat : JL.Kelabat
No.Hp : 0895418811700

Rumusan Masalah 1 : Bagaimana merancang strategi pemasaran digital berbasis web yang efektif untuk meningkatkan jangkauan dan penjualan Toko Oleh-Oleh 366 Lubuklinggau, serta mengoptimalkan fitur website guna mendukung penjualan online dan menarik pelanggan baru

Judul 1 : Strategi Pemasaran Produk Oleh-Oleh di Era Digital Melalui Website Toko Oleh-Oleh 366 Lubuklinggau

Rumusan Masalah 2 : Bagaimana sentimen pengguna media sosial terhadap serangan Iran ke Israel—apakah cenderung positif, negatif, atau netral—menggunakan **Naive Bayes**, dan apa saja topik utama yang dibahas terkait serangan tersebut berdasarkan analisis **LDA (Latent Dirichlet Allocation)**

Judul 2 : Analisis Sentimen dan Topik Diskusi Pengguna Media Sosial terhadap Serangan Iran ke Israel: Studi Kasus di X dengan Naive Bayes dan LDA

Rumusan Masalah 3 : Bagaimana sentimen pengguna media sosial terhadap pelantikan Prabowo Subianto diklasifikasikan sebagai positif, negatif, atau netral menggunakan metode Naive Bayes dan Support Vector Machine (SVM), serta metode mana yang menunjukkan akurasi terbaik berdasarkan metrik evaluasi

Judul 3 : Perbandingan Akurasi Naive Bayes dan SVM dalam Analisis Sentimen Media Sosial terkait Pelantikan Presiden RI ke-8: Studi Kasus Prabowo Subianto

Diusulkan Judul Nomor :1(satu)/ 2(Dua)/ 3(Tiga)*

Lubuklinggau, 13 November 2024
Mahasiswa yang mengusulkan,

(Erik Kurniawan)

Menyetujui Dosen Pembimbing,

Pembimbing 1 (Elmayati M.kom) (.....)

Pembimbing 2 (Harma Oktavia Lingga Wijaya M.kom) (.....)

Mengesahkan,
Dekan Fakultas Ilmu Tehnik

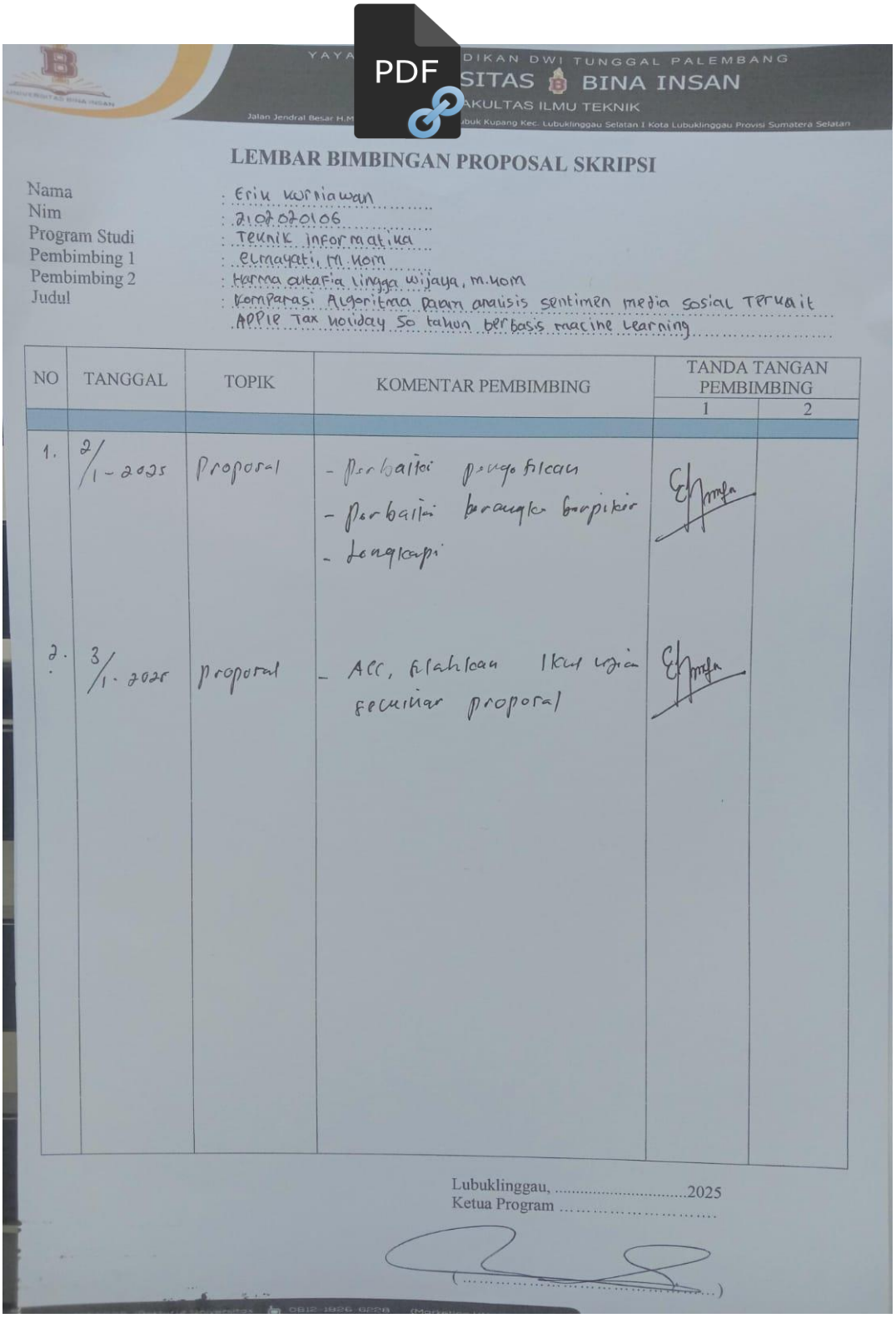
(Dr.Rudi Kurniawan, St,M.kom)

Mengetahui,
Ketua Program Studi, Informatika

(Budi Santoso M.kom)

0733-4553932 (Rektorat Universitas Bina Insan)
0733-3280300
0733-3280200 (Pascasarjana)
0812-1826-6228 (Marketing UNIVBI)
0852-3151-5800 (Admin UNIVBI)

Lampiran 2. Form bimbingan proposal skripsi acc p1



DIKAN DWI TUNGGAL PALEMBANG
SITAS BINA INSAN
AKULTAS ILMU TEKNIK
Jalan Jendral Besar H.M ... bukuk Kupang Kec. Lubuklinggau Selatan I Kota Lubuklinggau Provisi Sumatera Selatan

LEMBAR BIMBINGAN PROPOSAL SKRIPSI

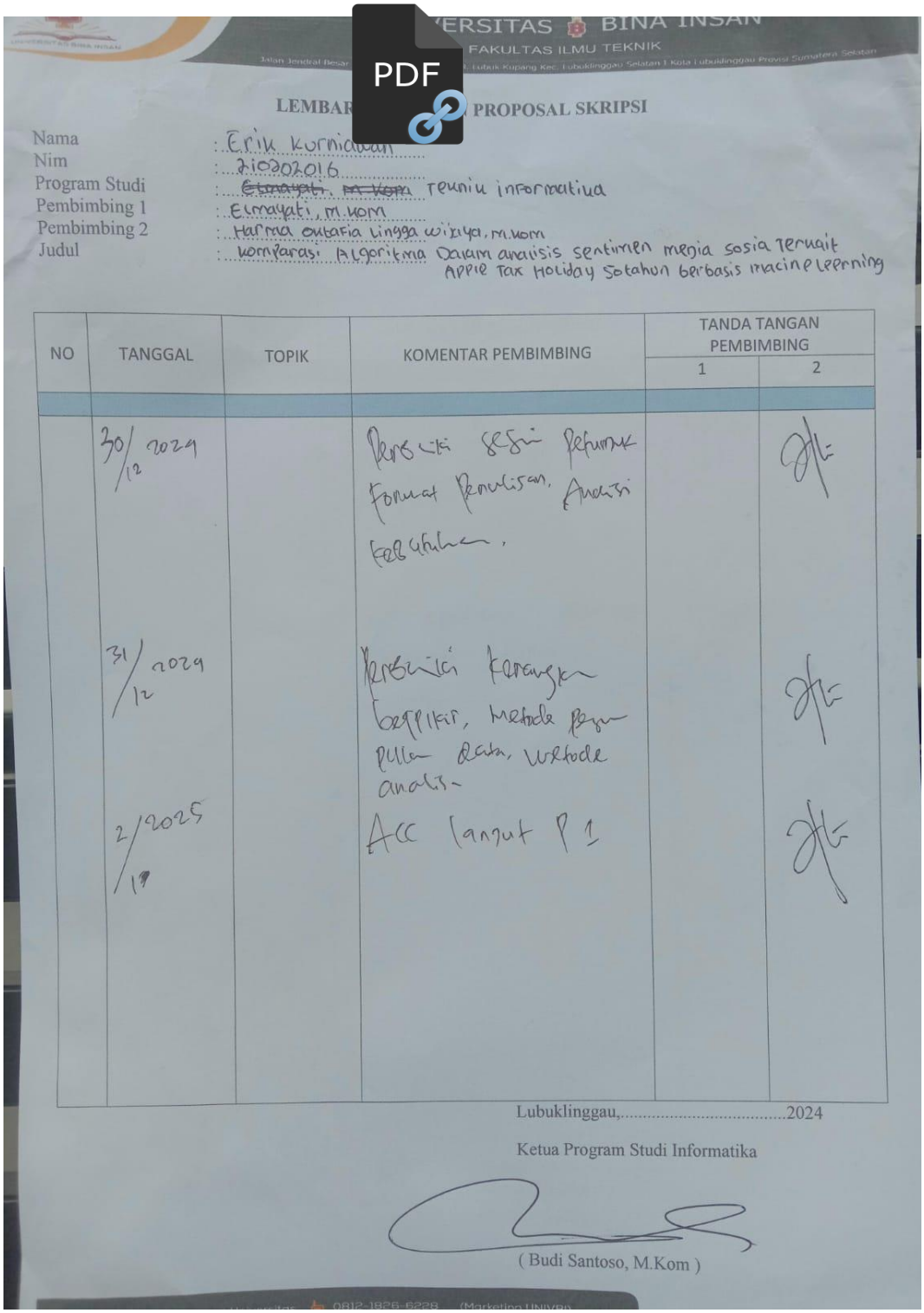
Nama : Erik kurniawan
Nim : 21070106
Program Studi : Teknik informatika
Pembimbing 1 : elmayati, M.kom
Pembimbing 2 : Karma cutafia lingga wijaya, m.kom
Judul : komparasi Algoritma dalam analisis sentimen media sosial terkait Apple Tax holiday 5o tahun berbasis macine Learning

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1.	2/1-2025	Proposal	- perbaiki penggolongan - perbaiki kerangka berpikir - lengkapi		
2.	3/1-2025	proposal	- ACC, silahkan lanjut ujian seminar proposal		

Lubuklinggau, ...2025
Ketua Program .....

(.....)

Lampiran 3. Form bimbingan proposal skripsi acc p2



ERSITAS BINA INSAN
FAKULTAS ILMU TEKNIK
Jalan Jendral Besar ... Kel. Lubuk Kupang Kec. Lubuklinggau Selatan I Kota Lubuklinggau Provinsi Sumatera Selatan

LEMBAR ... PROPOSAL SKRIPSI

Nama : Erik Kurniawan
Nim : 210202016
Program Studi : ~~Elmayati, M.Kom~~ Teknik Informatika
Pembimbing 1 : Elmayati, M.Kom
Pembimbing 2 : Harma Oktafia Lingga Wijaya, M.Kom
Judul : Komparasi Algoritma Dalam analisis sentimen media sosial terkait Apple Tax Holiday Setahun berbasis machine Learning

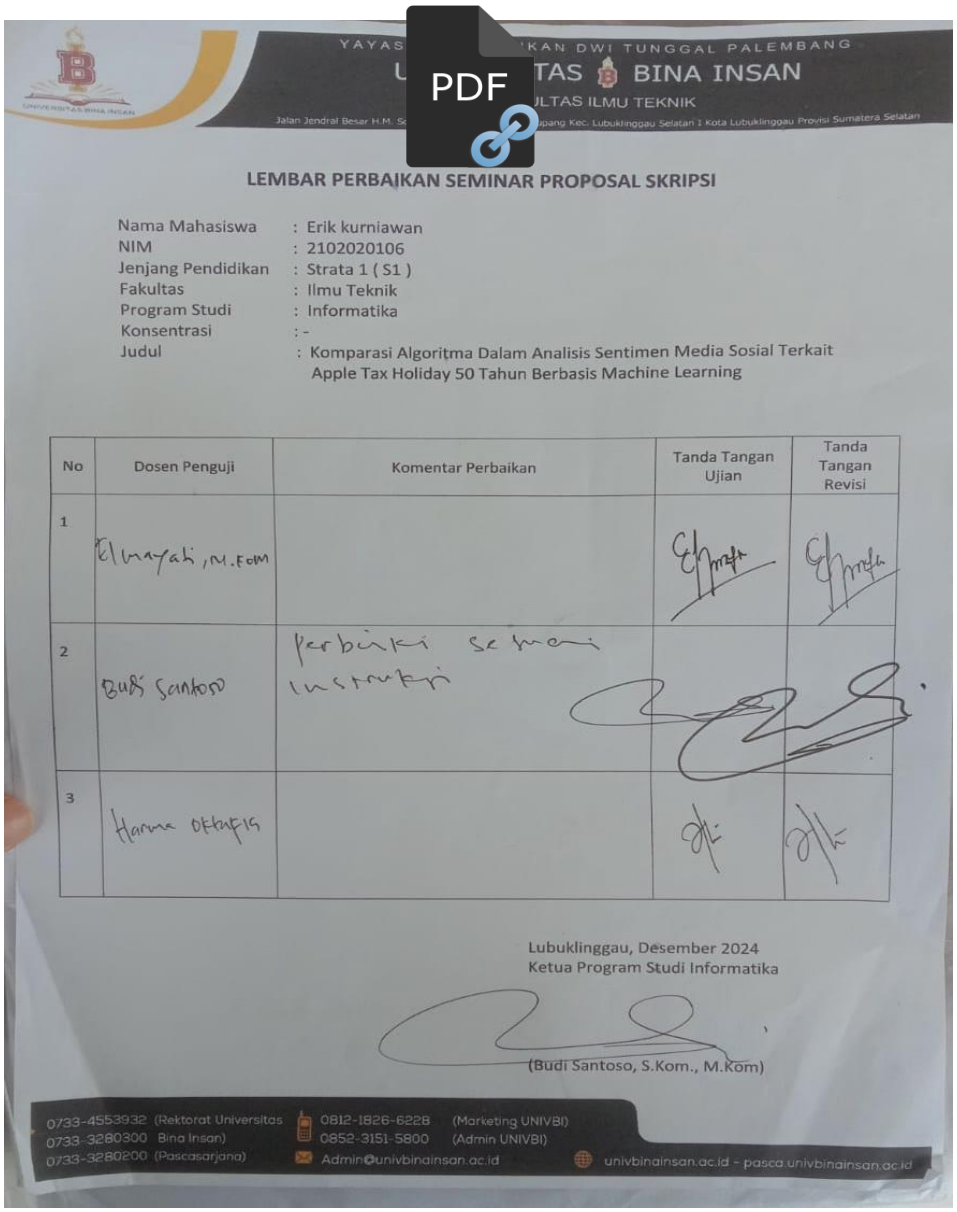
NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
	30/12 2024		Perbaiki sesuai referensi format penulisan, Analisi Fasilitas.		
	31/12 2024		Perbaiki kerangka berpikir, metode pengumpulan data, metode analisa.		
	2/2025		Acc lanjut P1		

Lubuklinggau,.....2024

Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

Lampiran 4. Lembar perbaikan seminar proposal



YAYASAN DWI TUNGGAL PALEMBANG
UNIVERSITAS BINA INSAN
FAKULTAS ILMU TEKNIK
Jalan Jendral Besar H.M. Soeharto Kelurahan Lubuklinggau Selatan I Kecamatan Lubuklinggau Selatan I Kota Lubuklinggau Provinsi Sumatera Selatan

LEMBAR PERBAIKAN SEMINAR PROPOSAL SKRIPSI

Nama Mahasiswa	: Erik kurniawan
NIM	: 2102020106
Jenjang Pendidikan	: Strata 1 (S1)
Fakultas	: Ilmu Teknik
Program Studi	: Informatika
Konsentrasi	: -
Judul	: Komparasi Algoritma Dalam Analisis Sentimen Media Sosial Terkait Apple Tax Holiday 50 Tahun Berbasis Machine Learning

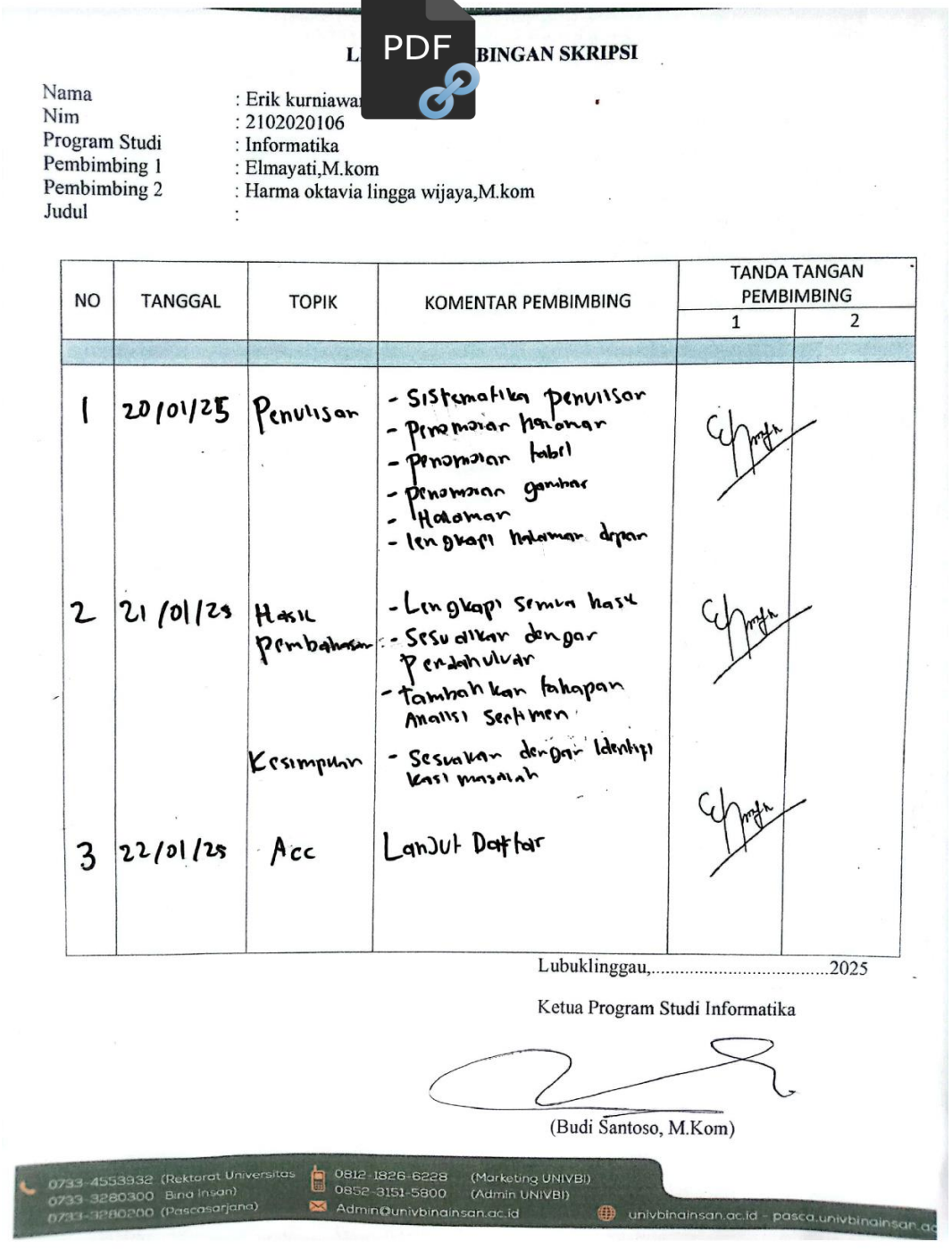
No	Dosen Penguji	Komentar Perbaikan	Tanda Tangan Ujian	Tanda Tangan Revisi
1	Elmayati, M.Kom			
2	Budi Santoso	perbaiki sesuai instruksi		
3	Hanna Oktafia			

Lubuklinggau, Desember 2024
Ketua Program Studi Informatika

(Budi Santoso, S.Kom., M.Kom)

0733-4553932 (Rektorat Universitas)
0733-3280300 Bina Insan)
0733-3280200 (Pascasarjana)
0812-1826-6228 (Marketing UNIVBI)
0852-3151-5800 (Admin UNIVBI)
Admin@univbinainsan.ac.id
univbinainsan.ac.id - pasca.univbinainsan.ac.id

Lampiran 5.Form bimbingan skripsi acc p1



L BINGAN SKRIPSI

Nama : Erik kurniawa
Nim : 2102020106
Program Studi : Informatika
Pembimbing 1 : Elmayati,M.kom
Pembimbing 2 : Harma oktavia lingga wijaya,M.kom
Judul :

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1	20/01/25	Penulisan	- Sistematika penulisan - Penomoran halaman - penomoran tabel - Penomoran gambar - Halaman - lengkapi halaman depan		
2	21/01/25	Hasil Pembahasan	- Lengkapi semua hasil - Sesuaikan dengan Perdahuluan - Tambahkan tahapan Analisi Sentimen		
		Kesimpulan	- Sesuaikan dengan Identifi kasi masalah		
3	22/01/25	Acc	Lanjut Daftar		

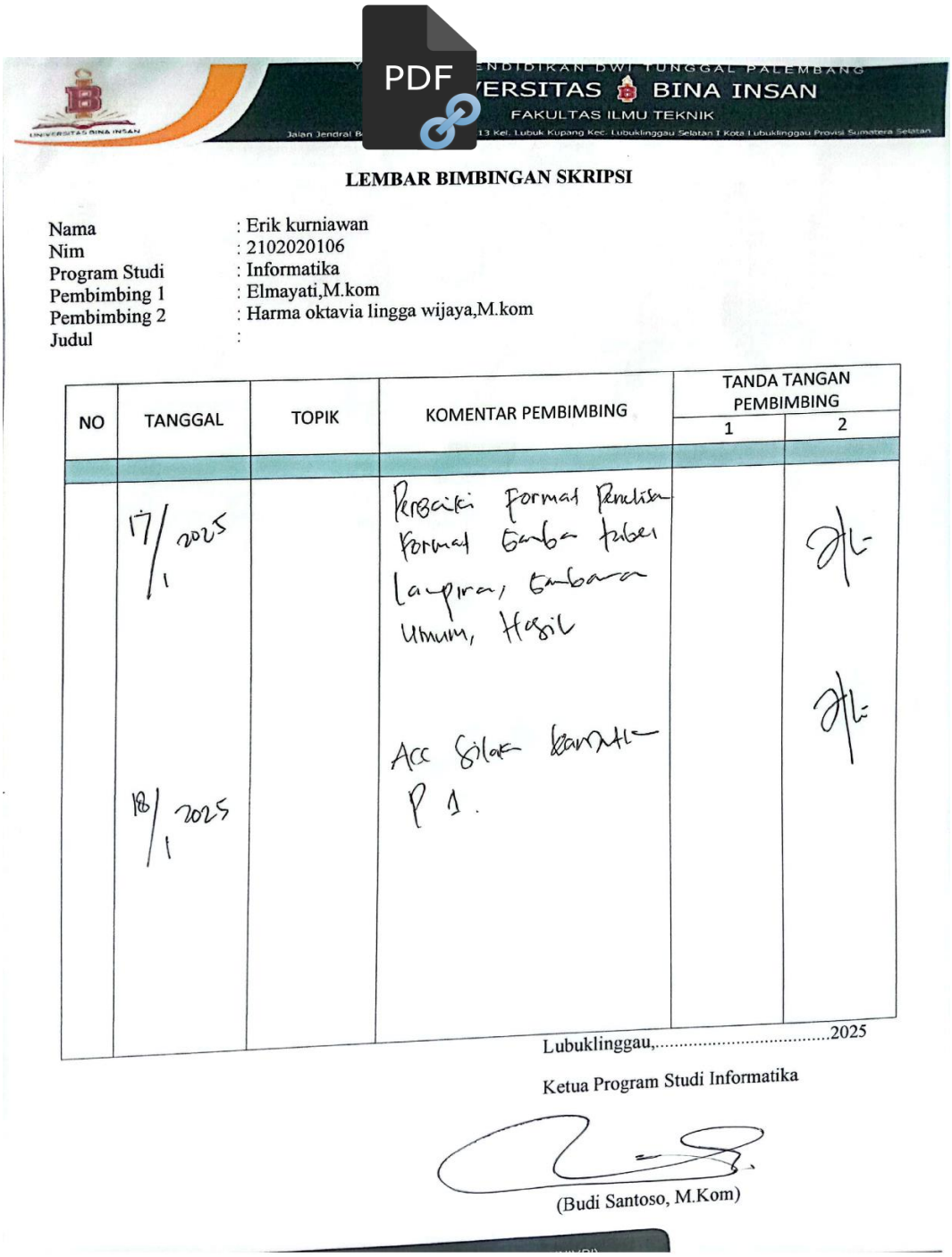
Lubuklinggau,.....2025

Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

0733-4553932 (Rektorat Universitas Bina Insan)
0733-3280300
0733-3280200 (Pascasarjana)
0812-1826-6228 (Marketing UNIVBI)
0852-3151-5800 (Admin UNIVBI)
Admin@univbinainsan.ac.id
univbinainsan.ac.id - pasca.univbinainsan.ac

Lampiran 6.Form bimbingan skripsi acc p2



PENDIDIKAN DWI TUNGGAL PALEMBANG
UNIVERSITAS BINA INSAN
FAKULTAS ILMU TEKNIK
Jalan Jendral B... 13 Kel. Lubuk Kupang Kec. Lubuklinggau Selatan I Kota Lubuklinggau Provinsi Sumatera Selatan

LEMBAR BIMBINGAN SKRIPSI

Nama : Erik kurniawan
Nim : 2102020106
Program Studi : Informatika
Pembimbing 1 : Elmayati,M.kom
Pembimbing 2 : Harma oktavia lingga wijaya,M.kom
Judul :

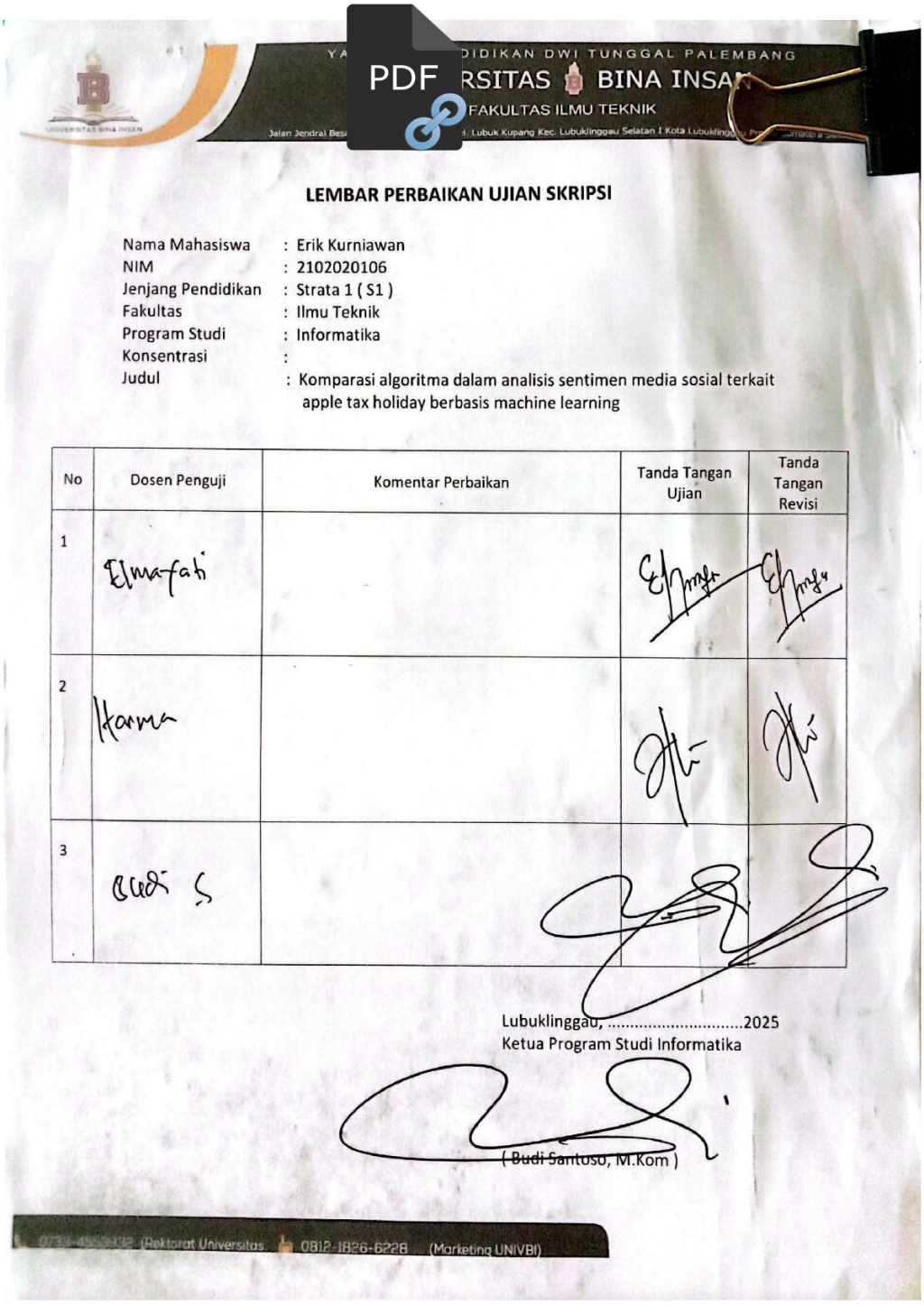
NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
	17/1/2025		Perbaiki format penulisan, format gambar tabel, lampiran, gambaran umum, hasil		
	18/1/2025		Acc Silakan bimbingan P 1.		

Lubuklinggau,.....2025

Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

Lampiran 7.Lembar perbaikan skripsi



DIDIKAN DWI TUNGGAL PALEMBANG
RSITAS BINA INSAN
FAKULTAS ILMU TEKNIK

LEMBAR PERBAIKAN UJIAN SKRIPSI

Nama Mahasiswa : Erik Kurniawan
NIM : 2102020106
Jenjang Pendidikan : Strata 1 (S1)
Fakultas : Ilmu Teknik
Program Studi : Informatika
Konsentrasi :
Judul : Komparasi algoritma dalam analisis sentimen media sosial terkait apple tax holiday berbasis machine learning

No	Dosen Penguji	Komentar Perbaikan	Tanda Tangan Ujian	Tanda Tangan Revisi
1	Elmafati			
2	Karma			
3	Budi S			

Lubuklinggau, ...2025
Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

(Rektorat Universitas) 0812-1826-6228 (Marketing UNIVBI)