

**ANALISIS SENTIMEN PENGGUNA X TERHADAP APLIKASI
iPusnas MENGGUNAKAN METODE NAÏVE BAYES**

SKRIPSI



Oleh :

DWI SEKAR NAGARI

NIM. 19680061

PROGRAM STUDI PERPUSTAKAAN DAN SAINS INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI MAULANA IBRAHIM

MALANG

2024

**ANALISIS SENTIMEN PENGGUNA X TERHADAP APLIKASI
iPusnas MENGGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

Oleh:

DWI SEKAR NAGARI

NIM. 19680061

Diajukan Kepada:

Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang

Untuk Memenuhi Salah Satu Persyaratan dalam

Memperoleh Gelar Sarjana Sains Informasi (S.S.I)

PROGRAM STUDI PERPUSTAKAAN DAN SAINS INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM

MALANG

2024

HALAMAN PERSETUJUAN

**ANALISIS SENTIMEN PENGGUNA X TERHADAP APLIKASI
iPusnas MENGGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

Oleh:

DWI SEKAR NAGARI

NIM. 19680061

Telah Diperiksa dan Disetujui:

Tanggal: 13 Desember 2024

Pembimbing I,

Fakhris Khusnu Reza Mahfud, M.Kom
NIP. 19900506 201903 1 007

Pembimbing II,

Firma Sahrul Bahtiar, M.Eng
NIP. 19850201 201903 1 009

Mengetahui,

**Ketua Program Studi Perpustakaan dan Sains Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang**



Dr. H. Mokhammad Amin Hariyadi, M.T.
NIP. 19660118 200501 1 001

HALAMAN PENGESAHAN

ANALISIS SENTIMEN PENGGUNA X TERHADAP APLIKASI iPusnas MENGUNAKAN METODE NAÏVE BAYES

SKRIPSI

Oleh:

DWI SEKAR NAGARI

NIM. 19680061

Telah Dipertahankan di Depan Dewan Penguji Skripsi dan
Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Sains Informasi (S.S.I)
Pada tanggal 13 Desember 2024

Susunan Dewan Penguji

- Ketua Penguji : Dr. Ir. Mokhamad Amin Hariyadi, M.T.
NIP. 19670118 200501 1 001
- Anggota Penguji I : Wahyu Hariyanto, M.M
NIP. 19890721 201903 1 007
- Anggota Penguji II : Fakhris Khusnu Reza Mahfud, M.Kom
NIP. 19900506 201903 1 007
- Anggota Penguji III : Firma Sahrul Bahtiar, M.Eng
NIP. 19850201 201903 1 009

Tanda Tangan



Mengetahui,
Ketua Program Studi Perpustakaan dan Sains Informasi
Fakultas Sains dan Teknologi
Universitas Maulana Malik Ibrahim Malang



Mokhamad Amin Hariyadi, M.T.
NIP. 19670118 200501 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Dwi Sekar Nagari
NIM : 19680061
Prodi : Perpustakaan dan Sains Informasi
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenar-benarnya bahwa skripsi yang saya tulis merupakan hasil karya saya sendiri melalui observasi dan penelitian yang telah dilakukan. Bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 13 Desember 2024
Yang membuat Pernyataan,

A blue official stamp with the text 'KEPERAAN' and 'TEMPER' is visible. The stamp also contains the number '30574AMX062841214'. The signature is written in black ink over the stamp.

Dwi Sekar Nagari
NIM. 19680061

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Syukur Alhamdulillah peneliti haturkan kehadiran Allah Subhanahu Wa Ta'ala atas limpahan rahmat dan karunia-Nya, sehingga peneliti dapat menyelesaikan studi di program studi Perpustakaan dan Sains Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan skripsi ini dengan baik.

Selanjutnya Peneliti menyadari bahwa banyak pihak yang telah mendukung dalam menyelesaikan penulisan skripsi ini. Oleh karena itu, peneliti mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. M. Amin Hariyadi, M.T, selaku Ketua Program Studi Perpustakaan dan Sains Informasi Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang, serta menjadi Dosen Penguji I yang telah mendampingi dan mengawasi setiap tahap pengerjaan skripsi hingga selesai.
2. Bapak Fakhris Khusnu Reza Mahfud, M.Kom selaku Dosen Wali sekaligus Dosen Pembimbing I yang telah memberikan waktu, bimbingan, nasehat, dan seluruh dukungannya selama proses pengerjaan skripsi, sehingga tulisan ini dapat terselesaikan dengan baik.
3. Bapak Firma Sahrul Bahtiar, M.Eng, selaku dosen pembimbing II yang telah berkenan meluangkan waktu dan tenaganya untuk membimbing peneliti hingga skripsi terselesaikan.
4. Bapak Wahyu Hariyanto, M.M selaku Dosen Penguji II yang telah mendampingi dan mengawasi setiap tahap pengerjaan skripsi hingga selesai.
5. Seluruh dosen dan staff Program Studi Perpustakaan dan Sains Informasi yang telah membagikan ilmu dan pengalamannya.
6. Kedua orang tua, kedua saudara, serta keluarga besar peneliti yang selalu memberikan doa beserta dukungan moral, materi, dan spiritual untuk menyelesaikan perkuliahan ini.

7. Teman peneliti yaitu Didik Cahya Indirman, yang telah memberikan motivasi serta dukungan dalam pengerjaan skripsi.

Peneliti menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan. Harapan peneliti semoga skripsi ini memberikan manfaat kepada pembaca khususnya kepada peneliti secara pribadi. Amin Ya Rabbal Alamin.

Wassalamu'alaikum Wr. Wb.

Malang, 13 Desember 2024

Peneliti

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	iv
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
ABSTRAK	xi
ABSTRACT.....	xii
مستخلص البحث.....	xiii
1.1 Latar Belakang Masalah	1
1.2 Identifikasi Masalah.....	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	4
1.6 Sistematika Penulisan	4
BAB II KAJIAN PUSTAKA	6
2.1 Hasil Penelitian Terdahulu	6
2.2 Landasan Teori	9
2.2.1 Ipusnas	9
2.2.2 X	10
2.2.3 Tweet Harvest.....	10
2.2.4 Python.....	11
2.2.5 Sastrawi.....	11
2.2.6 <i>Preprocessing</i>	11
2.2.7 <i>Tf-IDF</i>	12
2.2.8 Klasifikasi Naïve Bayes.....	13
2.2.9 <i>Performance Evaluation Measure (PEM)</i>	15
BAB III METODE PENELITIAN	18

3.1	Jenis Penelitian	18
3.2	Alur Penelitian	18
3.3	Sumber Data	25
3.4	Pengumpulan Data.....	25
BAB IV HASIL DAN PEMBAHASAN		26
4.1	Hasil Penelitian.....	26
4.2	Pembahasan	45
BAB V PENUTUP.....		49
5.1	Kesimpulan.....	49
5.2	Saran	49
DAFTAR PUSTAKA		50
LAMPIRAN.....		53

DAFTAR TABEL

Tabel 2.2 Confusion Matrix (Wang dkk., 2022).....	15
Tabel 3.1 Contoh proses penerapan Case Folding.....	21
Tabel 3.2.1 Contoh proses penerapan Filtering.....	21
Tabel 3.2.2 Contoh proses penerapan Tokenizing.....	22
Tabel 3.2.3 Contoh proses penerapan Stemming.....	23
Tabel 4.1 Tabel hasil pelabelan data.....	27
Tabel 4.2 Data Sebelum Preprocessing.....	27
Tabel 4.3 Hasil Case Folding.....	28
Tabel 4.4 Hasil Filtering.....	28
Tabel 4.5 Hasil Tokenizing.....	29
Tabel 4.6 Hasil Stemming.....	29
Tabel 4.7 Rasio yang digunakan dalam pembagian.....	31
Tabel 4.8 Hasil Klasifikasi rasio 90:10 dengan TF-IDF.....	32
Tabel 4.9 Hasil klasifikasi rasio 80:20 dengan TF-IDF.....	32
Tabel 4.10 Hasil Klasifikasi rasio 70:30 dengan TF-IDF.....	32
Tabel 4.11 Hasil prediksi kelas pada data uji.....	33

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian.....	18
Gambar 3.2 Desain Sistem.....	20
Gambar 3.3 <i>Source Code</i> Hasil <i>Crawl Data Tweet Harvest</i>	25
Gambar 4.1 Hasil Labeling Data.....	26
Gambar 4.2 Tampilan data ekstraksi fitur dengan TF-IDF.....	30
Gambar 4.3 Diagram Hasil Analisis Naïve Bayes Rasio 90:10.....	34
Gambar 4.4 Diagram Hasil Analisis Naïve Bayes Rasio 80:20.....	34
Gambar 4.5 Diagram Hasil Analisis Naïve Bayes Rasio 70:30.....	35
Gambar 4.6 <i>Confusion Matrix</i> Naïve Bayes rasio 90:10	36
Gambar 4.7 <i>Confusion Matrix</i> Naïve Bayes rasio 80:20	38
Gambar 4.8 <i>Confusion Matrix</i> Naïve Bayes rasio 70:30	40
Gambar 4.9 Diagram Hasil Evaluasi <i>Confusion Matrix</i> Rasio 90:10	43
Gambar 4.10 Diagram Hasil Evaluasi <i>Confusion Matrix</i> Rasio 80:20	43
Gambar 4.11 Diagram Hasil Evaluasi <i>Confusion Matrix</i> Rasio 70:30	44

ABSTRAK

Nagari, Dwi Sekar. 2024. **Analisis Sentimen Pengguna X terhadap Aplikasi iPusnas Menggunakan Metode Naïve Bayes. Program Studi Perpustakaan dan Sains Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.**

Pembimbing: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Firma Sahrul Bahtiar, M.Eng.

Kata kunci : Analisis Sentimen, iPusnas, Naïve Bayes, X, *Text Mining*

Aplikasi iPusnas memiliki banyak opini positif dan negatif, contohnya seperti aplikasi iPusnas mempunyai koleksi yang lengkap, pinjam buku di iPusnas gratis, atau iPusnas sering *error*, serta antrean iPusnas lama. Banyak pengguna yang memberikan tanggapan tentang iPusnas yang pernah mereka gunakan dalam bentuk opini berupa pengalaman baik maupun buruk. Opini dengan bahasa Indonesia di X tentang aplikasi iPusnas akan dijadikan sebagai data penelitian. Tujuan penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma *Multinomial Naïve Bayes* dalam analisis sentimen terhadap aplikasi iPusnas pada media sosial X. Metode yang digunakan pada penelitian ini adalah *Multinomial Naïve Bayes* yang termasuk dalam algoritma probabilitas, serta evaluasi kinerja modelnya menggunakan *Confusion Matrix*. Berdasarkan hasil dari evaluasi model, diketahui bahwa nilai tertinggi ada pada rasio 90:10 yang memiliki akurasi sebesar 82%, pada rasio 80:20 memiliki akurasi sebesar 79%, serta pada rasio 70:30 sebesar 74%. Pengujian rasio 90:10 memiliki akurasi 82%, dilihat dari kelas positif presisi sejumlah 80%, *recall* sejumlah 83%, *f1-score* sejumlah 82%. Sedangkan dilihat dari kelas negatif presisi sejumlah 85%, *recall* sejumlah 82%, *f1-score* 84%.

ABSTRACT

Nagari, Dwi Sekar. 2024. **The Analysis of X Users's Sentiments on iPusnas Application Using Naïve Bayes Method. Library and Information Science Study Program, Faculty of Science and Technology Universitas Islam Negeri Maulana Malik Ibrahim Malang.**

Advisor: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Firma Sahrul Bahtiar, M.Eng.

Keywords: *Sentiment Analysis, iPusnas, Naïve Bayes, X, Text Mining*

The iPusnas application receives many positive and negative opinions for example, it has a complete collection, offers free book borrowing, often experiences errors, and has a long queue. Many users give their opinions about the iPusnas, sharing their good and bad experiences. The research data consists of opinions in the Indonesian language on X dealing with the iPusnas application. The research aims to determine the accuracy level of the Multinomial Naïve Bayes algorithm in sentiment analysis on iPusnas application on X social media. The research employed Multinomial Naïve Bayes included in the probability algorithm, and the model performance evaluation used a Confusion Matrix. The model evaluation result shows that the highest score is for a ratio of 90:10, which has an accuracy level of 82%. The ratio of 80:20 has an accuracy level of 79%, and the ratio of 70:30 has an accuracy level of 74%. The testing of a 90:10 ratio has an accuracy level of 82%. The positive class shows a precision of 80%, a recall of 83%, and an f1-score of 82%. Meanwhile, the negative class shows a precision of 85%, a recall of 82%, and an f1-score of 84%.

مستخلص البحث

ناجاري، دوي سيكار. 2024. تحليل مشاعر المستخدم تطبيق X تجاه تطبيق iPusnas باستخدام طريقة المصنف البايزي الساذج. البحث الجامعي. قسم علوم المكتبات والمعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: فهرس حسن ريزا محفوظ، الماجستير. المشرف الثاني: فيرما شهر البختار، الماجستير.

الكلمات الرئيسية: تحليل مشاعر، iPusnas، مصنف بايزي ساذج، X، تحليل نصوص.

يحتوي تطبيق iPusnas على عدة الآراء الإيجابية والسلبية، على سبيل المثال يحتوي تطبيق iPusnas على مجموعة كاملة، أو استعارة الكتب في iPusnas مجانية، أو iPusnas غالباً ما تكون أخطاء، بالإضافة إلى قوائم انتظار iPusnas الطويلة. قدم العديد من المستخدمين ملاحظات حول iPusnas التي استخدموها في شكل آراء في صورة تجارب جيدة وسينة. سيتم استخدام الرأي باللغة الإندونيسية في X حول تطبيق iPusnas كبيانات بحثية. الهدف من هذا البحث هو معرفة قيمة دقة خوارزمية المصنف البايزي الساذج متعددة الحدود في تحليل المشاعر لتطبيق iPusnas على وسائل التواصل الاجتماعي X. الطريقة المستخدمة في هذا البحث هي Naïve Bayes 'Multinomial' والتي يتم تضمينها في خوارزمية الاحتمالات، بالإضافة إلى تقييم أداء النموذج باستخدام مصفوفة الارتباك. بناءً على نتائج تقييم النموذج، من المعروف أن أعلى درجة هي في نسبة 90:10 التي تبلغ دقتها 82٪، بينما تبلغ دقة نسبة 80:20 79٪، ونسبة 70:30 74٪. تمتع اختبار نسبة 90:10 بدقة 82٪، انطلاقاً من فئة الدقة الإيجابية بنسبة 80٪، والاستدعاء بنسبة 83٪، ودرجة ف1 بنسبة 82٪. وفي الوقت نفسه، انطلاقاً من فئة الدقة السلبية بنسبة 85٪، والاستدعاء بنسبة 82٪، ودرجة ف1 بنسبة 84٪.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perpustakaan Nasional Republik Indonesia (PNRI) membuat sebuah inovasi dengan menghadirkan *platform* perpustakaan digital berbasis aplikasi yaitu iPusnas. Pengguna iPusnas banyak yang memanfaatkan media sosial untuk berkomunikasi serta bertukar informasi tentang koleksi yang tersedia, tidak hanya itu pengguna juga memberikan tanggapan maupun opini mereka terhadap layanan iPusnas. Aplikasi iPusnas merupakan salah satu topik yang sangat menarik untuk dibicarakan. Ketika seseorang menulis opini tentang iPusnas, maka semua unsur yang ada di dalam iPusnas akan dituliskan. Contohnya seperti aplikasi iPusnas mempunyai koleksi yang lengkap, pinjam buku di iPusnas gratis, atau iPusnas sering *error*, serta antrean iPusnas lama. Banyak pengguna X yang memberikan tanggapan tentang iPusnas yang pernah mereka gunakan dalam bentuk opini berupa pengalaman baik maupun buruk. Hal tersebut menjadikan iPusnas memiliki berbagai macam topik yang informasinya bisa digali kembali. Oleh karena itu, opini berbahasa Indonesia di X yang berkaitan dengan aplikasi iPusnas akan dijadikan data pada penelitian ini.

Media sosial X digunakan sebagai tempat penambangan teks atau pengumpulan opini masyarakat atau pengguna terkait aplikasi iPusnas. X mempunyai tingkat aksesibilitas yang tinggi, kebaruan, serta format waktu yang jelas. Sehingga banyak pengguna terutama kalangan akademisi menggunakan X untuk mendapatkan informasi, saling berinteraksi, berpartisipasi dengan komunitas sesuai minat masing-masing, serta berbagi wawasan mengenai topik tertentu (Malik dkk., 2019). Oleh karena itu media sosial X dipilih oleh peneliti sebagai tempat penggalian data teks.

Analisis sentimen atau opini masyarakat saat ini semakin bertambah luas dan bebas diungkapkan diberbagai *platform* media sosial manapun. Sentimen dapat menjadi potensi besar bagi suatu perusahaan ataupun instansi terkait yang ingin mengetahui *feedback* masyarakat terhadap suatu produk atau layanan spesifik (Gunawan dkk., 2018). Analisis sentimen merupakan sebuah proses mengolah data

tekstual yang dilakukan secara otomatis sehingga diperoleh informasi untuk mengetahui kecenderungan opini pada suatu objek, terdiri dari opini positif atau negatif (Rozi dkk., 2012). Melalui analisis sentimen, iPusnas dapat memperoleh informasi tentang opini pengguna terhadap layanan pada aplikasi iPusnas, sehingga bisa menjadi bahan evaluasi untuk iPusnas kedepannya.

Analisis sentimen sangat penting untuk diperhatikan oleh seorang muslim tujuannya agar berhati-hati dalam menerima suatu opini sebelum diteliti apakah itu positif atau negatif. Hal ini berkaitan dengan Qur'an Surat Al-Hujurat ayat 6 yang berbunyi :

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنْ جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَنْ تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْحَبُوا عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ

Artinya :

“Wahai orang-orang yang beriman, jika seorang fasik datang kepadamu membawa berita penting, maka telitilah kebenarannya agar kamu tidak mencelakakan suatu kaum karena ketidaktahuan(-mu) yang berakibat kamu menyesali perbuatanmu itu” (Q.S Al-Hujurat ayat 6).

Ayat ini menjelaskan bahwa jika seseorang mendapatkan opini atau beita dari masyarakat yang belum jelas, lebih baik diteliti terlebih dahulu. Dijadikan sebagai pedoman bagi sekalian kaum mukmin supaya berhati-hati dalam menerima berita, terutama jika bersumber dari seorang yang fasik. Maksud yang terkandung dalam ayat ini adalah agar diadakan penelitian dahulu mengenai kebenarannya. Mempercayai suatu berita tanpa diselidiki kebenarannya, besar kemungkinan akan membawa korban jiwa dan harta yang sia-sia, yang hanya menimbulkan penyesalan belaka (*Qur'an Kemenag.*, 2023). Tafsir ayat tersebut dapat dipahami bahwa penting untuk meneliti terlebih dahulu opini yang belum jelas, salah satu upaya yang dilakukan adalah analisis sentimen iPusnas pada X untuk menentukan opini mana yang lebih cenderung positif atau negatif, sehingga mendapatkan hasil yang berguna untuk lembaga instansi terkait maupun untuk masyarakat.

Dalam penelitian ini, untuk melakukan analisis sentimen digunakan alogaritma Naïve Bayes *Classifier*. Alogaritma Naïve Bayes dipilih karena merupakan *supervised document classification* yang sederhana namun efisien

(Pujiyanto dkk., 2019). Metode Naïve Bayes mempunyai sejumlah keutamaan, termasuk kecepatan dan akurasi yang tinggi ketika diterapkan pada dataset yang besar dan beragam (Fauzan & Hikmah., 2022). Tak hanya itu, dibandingkan dengan metode klasifikasi lainnya, Naïve Bayes adalah metode yang relatif mudah dan hanya memerlukan sejumlah kecil data pelatihan (Septianingrum & Irawan., 2021).

Salah satu penelitian analisis sentimen X yang pernah dilakukan yaitu menganalisis sentimen terhadap pembelajaran daring pada masa pandemi Covid-19 menggunakan metode klasifikasi Naïve Bayes melalui proses *crawling data*, *preprocessing*, dan klasifikasi. Hasil penelitian ini menunjukkan bahwa pembelajaran daring memiliki 30% sentimen positif, 69% sentimen negatif, dan 1% netral dengan tingkat akurasi mencapai 97,15% (Samsir dkk., 2021).

Data yang digunakan adalah data *tweet* pada media sosial X tentang iPusnas yang diambil dengan menggunakan teknik *Crawling* data dikumpulkan dengan menggunakan *Twitter Harvest*, data tersebut berhasil dikumpulkan namun masih dalam bentuk teks utuh dari *tweet* pengguna. Selanjutnya, tahap *preprocessing* data digunakan untuk melakukan analisis sentimen yang terdiri dari proses *case folding*, *tokenizing*, *filtering*, dan *stemming*. Kemudian setelah *preprocessing* data, maka dilakukan ekstraksi fitur dengan metode *Tf-Idf* untuk melakukan perhitungan nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap fitur kata disetiap dokumen dalam dataset yang diharapkan mampu mendapatkan nilai akurasi yang lebih baik dengan pengklasifikasian pada algoritma Naïve Bayes *Classifier*. Setelah semua proses klasifikasi, maka dilakukan evaluasi di antaranya *accuracy*, *precision*, dan *recall* (Pintoko, 2018).

Penelitian ini menggunakan kata kunci “iPusnas” pada *tweet* dalam bahasa Indonesia yang diambil dari X. Berdasarkan latar belakang yang telah dipaparkan, maka menarik minat peneliti untuk mengambil judul “Analisis Sentimen Pengguna X Terhadap Aplikasi Ipusnas Menggunakan Metode Naïve Bayes”.

1.2 Identifikasi Masalah

Identifikasi masalah yang diangkat dalam penelitian ini yaitu seberapa besar akurasi Naïve Bayes *Classifier* dapat menganalisis sentimen X terhadap iPusnas?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengukur akurasi dari penerapan metode Naïve Bayes *Classifier* dalam menganalisa sentimen terhadap iPusnas pada X.

1.4 Manfaat Penelitian

Manfaat penelitian diharapkan dapat menjadi acuan atau referensi bagi penelitian selanjutnya apabila ingin mengembangkan dengan objek atau metode yang lainnya, serta dapat memberikan manfaat sebagai bahan evaluasi terhadap aplikasi iPusnas agar semakin baik.

1.5 Batasan Masalah

Batasan masalah dalam pembahasan penelitian ini adalah menggunakan *tweets* berbahasa Indonesia dengan kata kunci “iPusnas” yang di-*posting* mulai dari tanggal 01 Januari 2022 sampai 03 Oktober 2023.

1.6 Sistematika Penulisan

Penelitian ini terdiri dari lima bab dengan sistematika penulisan menggunakan pedoman penulisan skripsi UIN Maulana Malik Ibrahim Malang yang diuraikan sebagai berikut:

BAB I Pendahuluan

Bab I terdiri atas latar belakang masalah, identifikasi masalah, tujuan penelitian, manfaat penelitian, penerapan metode, batasan masalah, dan sistematika penulisan. Latar belakang masalah menjelaskan tentang perkembangan teknologi perpustakaan dan X, dan menambahkan hadist yang sesuai dengan topik pembahasan untuk melengkapi dan memperkuat gagasan. Bagian identifikasi masalah menjelaskan rumusan permasalahan yaitu seberapa besar Naïve Bayes *Classifier* dapat mempengaruhi akurasi dari hasil analisis sentimen terhadap iPusnas pada X. Kemudian tujuan yang akan dicapai pada penelitian dan kegunaan dari penelitian. Pada bagian manfaat penelitian menjelaskan manfaat teoritis dan manfaat praktis dari penelitian yang dilakukan. Penerapan metode juga dirumuskan sebagai gambaran sementara pada penelitian. Selain itu, adanya batasan masalah menggunakan *tweet*

berbahasa Indonesia dengan rentang tanggal data dari 01 Januari 2022 sampai 03 Oktober 2023.

BAB II Kajian Pustaka

Bab II berisi tentang penelitian terdahulu yang terkait, tinjauan pustaka dan landasan teori. Pada tinjauan pustaka membahas tentang penelitian terdahulu yang berkaitan dengan penelitian berupa artikel ilmiah serta memuat perbedaan pada objek dan subjek penelitian. Sedangkan pada dasar teori membahas tentang konsep, pengertian, dan teori Naïve Bayes *Classifier* yang mendukung penelitian.

BAB III Metode Penelitian

Bab III menjelaskan tentang sub bab yang terdiri dari jenis penelitian yaitu metode kuantitatif dengan pendekatan teknik *text mining*, alur penelitian yang menjelaskan tahap pelaksanaan penelitian, tempat dan waktu penelitian, subjek dan objek penelitian, sumber data, serta pengumpulan data

BAB IV Hasil dan Pembahasan

Bab IV membahas mengenai hasil penelitian tentang “Analisis Sentimen Pengguna X Terhadap Aplikasi iPusnas Menggunakan Metode Naïve Bayes”. Hasil dan pembahasan dilakukan dengan pengolahan *crawling* data X, selanjutnya dilakukan proses *preprocessing*, klasifikasi Naïve Bayes, serta evaluasi dan analisis data. Hasil akan dipaparkan dalam bentuk uraian, tabel, dan diagram.

BAB V Kesimpulan dan Saran

Bab V merupakan bab penutup dengan membahas kesimpulan dan saran. Kesimpulan dituliskan secara singkat dari hasil analisis data yang telah dilakukan untuk menjawab masalah. Sedangkan pada bagian saran diberikan berdasarkan pengalaman dan pertimbangan peneliti sebagai bahan evaluasi terhadap lembaga terkait, aplikasi iPusnas serta pengembangan peneliti lain dibidang sejenisnya.

BAB II KAJIAN PUSTAKA

2.1 Hasil Penelitian Terdahulu

Penelitian analisis sentimen twitter yang berjudul “Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter” penelitian ini menjelaskan lebih dari 25 juta orang terkontaminasi Covid-19 dan 1,2 juta orang meninggal dunia pada bulan November 2020. Meminimalisir agar virus tidak menyebar maka dilakukan pembatasan sosial. Tujuan penelitian yaitu mengimplementasikan algoritma Naïve Bayes *Classifier* terhadap analisis sentimen data opini film berbahasa Indonesia pada twitter. Penelitian menggunakan metode algoritma Naïve Bayes *Classifier* dengan pembagian datasetnya menggunakan *5-fold cross validation*. Penelitian ini di implementasikan menggunakan metode python dan *library* NLTK. NLTK digunakan untuk mendapatkan secara otomatis klasifikasi *tweet* sebagai sentimen *tweet* yang positif dan negatif. Data yang dipakai 200 *tweet* positif dan 200 *tweet* negatif untuk melatih *classifier* dan disimpan dalam file.txt sehingga total data untuk *classifier* adalah 400. Data dari hasil analisis yang dilakukan semakin banyak data training yang digunakan maka akan mempengaruhi kinerja dan sistem. Hasil akurasi akan semakin tinggi dan itu menandakan sistem berhasil melakukan klasifikasi pada fold kedua yaitu 90%, *precision* 92%, *Recall* 90% dan *f-measure* 90%. Hasil menunjukkan tingkat sentimen positif lebih besar dibandingkan dengan tingkat sentimen negatif (Ratnawati, 2018).

Selanjutnya, penelitian yang berjudul “Analisis Sentimen Jasa Transportasi *Online* pada Twitter Menggunakan Metode Naïve Bayes *Classifier*” masalah yang di angkat adalah analisis sentimen masyarakat terhadap jasa transportasi *online* yaitu Gojek dan Grab. Data yang digunakan harus merupakan data yang valid di ambil dari salah *platform* media sosial yaitu twitter. Data diambil dari gabungan kedua akun kemudian data tersebut dibagi menjadi data latih dan data uji secara acak. Pembagian data latih terdiri dari 750 data akun @gojekindonesia dan 750 data akun @grabID. Data uji terdiri dari 250 data dari setiap akun. Tujuan melakukan analisis sentimen dari masyarakat terhadap jasa transportasi *online* diantaranya

Gojek dan Grab untuk memberikan informasi tentang kepuasan yang berguna bagi perusahaan untuk mengetahui kepuasan pelanggan. Berdasarkan hasil analisis yang dilakukan terhadap jasa transportasi *online* dengan menggunakan metode Naïve Bayes untuk menentukan kelas prediksi, didapatkan hasil sentimen positif sebesar 88.60% dan sentimen negatif sebesar 11.40% dengan akurasi sebesar 86.80%. Hasil menunjukkan bahwa banyak masyarakat yang merasa puas dengan layanan jasa transportasi *online* di Indonesia berdasarkan data dari twitter (Pintoko, 2018b).

Penelitian tentang analisis sentimen twitter dengan judul “Analisis Sentimen Twitter menggunakan Metode Naïve Bayes dengan *Relevance Frequency Feature Selection* (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal)” membahas penyebaran virus meningkat sangat cepat di Indonesia yang membuat pemerintah mengambil keputusan untuk mengeluarkan kebijakan baru dalam pemberlakuan kebijakan tersebut menimbulkan banyak opini terjadi pada masyarakat. Opini masyarakat diklasifikasikan menjadi opini positif, negatif dan netral sehingga memerlukan analisis sentimen. Proses analisis sentimen terdiri dari *preprocessing* untuk pemrosesan opini, *Relevance Frequency Feature Selection* untuk mengurangi jumlah fitur dan klasifikasi menggunakan metode Naïve Bayes. Penelitian yang dilakukan ditujukan untuk mengetahui seberapa besar pengaruh metode *Relevance Frequency Feature Selection*, sehingga didapatkan hasil term baru yang akan digunakan untuk proses klasifikasi Naïve Bayes. Data latih yang digunakan pada pengujian saat ini adalah data *tweet* sebanyak 300 data. Hasil dari penelitian ini mendapatkan akurasi tertinggi sebesar 73,3% yang terjadi pada fold ke-4 dalam 5 kali pengujian dengan pembagian dataset menggunakan *5-fold cross validation* dan dihasilkan rata-rata akurasi sebesar 62,6%. Dari 5 kali pengujian yang dilakukan, terjadi peningkatan hasil akurasi disetiap *fold*-nya. Terjadi peningkatan nilai rata-rata akurasi sebesar 2,7%. Dengan hasil rata-rata akurasi yang didapatkan sebesar 65,3% (Toy dkk., 2021).

Penelitian serupa juga berjudul “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naïve Bayes *Classifier*” membahas sentimen masyarakat terhadap kinerja Dewan Perwakilan Rakyat (DPR) yang dianggap tidak berpihak kepada rakyat. DPR sebagai lembaga

yang mewakilkan rakyat memiliki tanggung jawab yang harus dipenuhi secara demokratis dan responsif untuk mengatasi permasalahan sosial yang ada pada masyarakat serta tidak memprioritaskan kepentingan partai politik. Penelitian ini menggunakan sebanyak 1546 data *tweet*. Tahapan analisis yang dilakukan yaitu pengumpulan data (*crawling*), *preporcessing* data yang terdiri dari proses *cleaning* data, *tokenization*, *stop remove* dan *case folding*, *splitting* data dan klasifikasi data menggunakan metode *Naïve Bayes Classifier*. Berdasarkan hasil penelitian menggunakan algoritma *Naïve Bayes* didapatkan *accuracy score* sebesar 0.8% atau 80% hal ini berarti sistem mampu memprediksi 80% secara akurat dari total data testing sebesar 20%. Sedangkan hasil penelitian berdasarkan analisis sistem *sentiment analyzer* mendapatkan klasifikasi *tweet* dari twitter mengenai DPR sebanyak 95 positif, 693 netral dan 758 negatif dari data hasil *crawling* sebanyak 1546 (Duei Putri dkk., 2022).

Penelitian lainnya yang berjudul “Analisis Sentimen Ulasan Pengguna Aplikasi Ipusnas Pada Google Play Store Dengan *Multinomial Naïve Bayes*” membahas tentang ulasan pengguna terkait dengan pengalaman menggunakan aplikasi tersebut pada kolom ulasan yang ada pada google play store. Namun, nilai bintang aplikasi *mobile* android iPusnas saat ini mencapai 3,9, sedangkan beberapa aplikasi sejenis lainnya mendapatkan nilai bintang di atas 4.1. Ulasan dapat dijadikan sebagai gambaran umum mengenai pendapat masyarakat sekaligus sebuah indikator kepuasan. Penelitian menggunakan 2 dataset, dataset pertama di ambil data ulasan dari tanggal 22 Desember 2022 sampai 1 Januari 2022 dan dataset kedua di ambil data ulasan dari tanggal 31 Desember 2021 sampai 1 Januari 2021. Data dikumpulkan dengan bantuan *package google-play-scraper* dengan sortir kategori ulasan terbaru, menghasilkan 1224 data ulasan pada dataset pertama dan 2332 data ulasan pada dataset kedua. Ulasan dapat dikelompokkan menjadi 2 yaitu ulasan positif dan negatif sehingga dapat dilakukan analisis sentimen menggunakan metode *Multinomial Naïve Bayes* pada data ulasan. untuk mengetahui tingkat akurasi dari metode *Multinomial Naïve Bayes* untuk analisis sentimen terhadap data ulasan aplikasi iPusnas pada Google Play Store. Hasil dari penelitian yang telah dilakukan, pada kedua dataset dapat diketahui bahwa mayoritas dari sentimen yang

ada lebih mengarah ke sentimen positif. Pada dataset pertama nilai akurasi, presisi, dan *recall* masing-masing sebesar 96,3%, 96,67%, dan 96,67%. Sedangkan dataset kedua nilai akurasi, presisi, dan *recall* masing-masing 98,31%, 96,92%, dan 100%. Hasil akurasi, presisi, dan *recall* dari metode *multinomial* Naïve Bayes dalam melakukan analisis sentimen memiliki perfoma yang sangat baik (Sejati & Pramusinto, 2023).

Terdapat persamaan dan perbedaan pada penelitian terdahulu dengan penelitian yang dilakukan. Persamaannya yaitu terfokus pada analisis sentimen. Sedangkan, perbedaan penelitian terdahulu yaitu analisis sentimen pengguna X terhadap aplikasi iPusnas menggunakan metode Naïve Bayes.

2.2 Landasan Teori

Proses penelitian mencakup teori ilmiah yang dijadikan sebagai pedoman untuk menjelaskan setiap variabel yang diteliti (Iba & Wardhana, 2023). Berikut landasan teori yang digunakan.

2.2.1 Ipusnas

Ipusnas adalah aplikasi perpustakaan berbasis digital yang dapat di akses di berbagai perangkat mulai dari desktop dan PC berbasis situs (*web-based*), *mobile* (*smartphone-based application*), *netbook* dan *tab based hybrid* (*tab-base application*) (Maulana, 2018). Ipusnas merupakan wujud konsep “*borderless library*” artinya, dengan adanya iPusnas perpustakaan tidak lagi dapat diakses terbatas tetapi sudah tidak terbatas.

Sistem iPusnas harus cukup fleksibel untuk menangani perubahan-perubahan yang terjadi, kepentingannya cukup beralasan dalam kondisi di mana sistem beroperasi atau dalam kebutuhan yang diwajibkan oleh organisasi (Fauzan & Suwanto, 2018). Aplikasi iPusnas sangat fleksibel karena bisa dimanfaatkan sebagai sarana temu balik informasi yang dapat memudahkan pengguna untuk mendapatkan informasi serta koleksi yang ada di perpustakaan tersebut di mana saja dan kapan saja sesuai dengan keinginan pengguna.

Sistem yang digunakan dalam pengoperasian aplikasi iPusnas haruslah mudah untuk dikontrol. Kemudahan pengontrolan digunakan agar pengguna dapat menemukan apa yang ingin dilakukan dengan menggunakan aplikasi tersebut. Dari

analisis sentimen X, pengguna dapat mengendalikan dan mengoperasikan aplikasi iPusnas sesuai dengan kebutuhan mereka meskipun terdapat beberapa kendala seperti *error* atau *bug*. Aplikasi iPusnas dapat mudah dipahami pengguna karena sudah terbiasa dalam pengoperasian aplikasi yang cukup sederhana, terstruktur, mudah dimengerti dan prosedurnya mudah diikuti (Jagiyanto, 2005).

2.2.2 X

Dari sekian banyak media sosial yang paling cepat perkembangannya adalah X, media sosial X sangat cepat mengambil perhatian masyarakat Indonesia dari berbagai kalangan. X merupakan platform yang berfokus pada *real-time*, artinya informasi dan peristiwa baru dapat dengan cepat menjadi topik hangat dan menjadi *trending topic*. Pengguna X dapat berkomunikasi menggunakan pesan singkat yang disebut dengan '*tweet*' atau kicauan hingga 280 karakter (Mustaqililah dkk., 2023).

X sering digunakan untuk menyalurkan opini mengenai topik-topik tertentu sebagai kampanye digital yang dapat mempengaruhi opini bahkan perilaku masyarakat (Mahfud dkk., 2020). Salah satu kampanye digital tersebut adalah kampanye pilpres tahun 2014, X digunakan secara masif bukan hanya oleh netizen namun juga oleh kandidat capres dan para pendukungnya. Penggunaan format *twitwar* dan *kultwit* dalam kicauan akun X kedua pendukung kandidat capres dipandang sebagai permainan kata-kata untuk mengkonstruksi realitas politik dengan tujuan membentuk opini publik. Setiap ekspresi, sikap, gagasan, ide maupun pandangan pendukung capres yang telah terwujud dalam format unggahan kicauan mereka, merupakan bentuk dari permainan pesan-pesan politik yang bertujuan untuk mempengaruhi opini publik. Layaknya percakapan lisan, format *kultwit* dan *twitwar* dipandang mampu meningkatkan kuantitas pesan politik yang bersirkulasi selama masa kampanye pilpres 2014 (Firmansyah dkk., 2018).

2.2.3 Tweet Harvest

Tweet Harvest merupakan *command-line tool* yang diciptakan oleh Helmi Satria dengan menggunakan *playwright* untuk mendapatkan *crawl* data pada X berdasarkan kata kunci dan rentang tanggal tertentu lalu. Dalam memudahkan pengguna menggunakan alat ini, pencipta membuat *project* google colab yang di

dalamnya sudah menggunakan *tweet-harvest* untuk melakukan *crawl* data. Tweet Harvest ini hanya boleh digunakan untuk tujuan pendidikan saja, dan untuk penggunaannya harus memiliki akun X yang valid untuk mendapatkan token akses, yang dapat diperoleh dengan masuk ke X di browser dan mengekstrak *cookie auth_token*. Data *tweet* yang sudah didapatkan disimpan dalam file *CSV* (Rahman, 2023).

2.2.4 Python

Python merupakan salah satu bahasa pemrograman populer yang sering digunakan untuk mengembangkan berbagai macam aplikasi dekstop, web, dan mobile. Selain itu python juga digunakan untuk mengotomatiskan tugas serta melakukan analisis data (Romzi & Budi, 2020).

2.2.5 Sastrawi

Sastrawi merupakan *library stemming* untuk Bahasa Indonesia. *Library* ini menggunakan Algoritma *Stemming* dari Nazief dan Adriani untuk bahasa Indonesia. Kata dasar yang dipakai berdasarkan data kamus kata dasar dari *kateglo.com* dengan beberapa modifikasi digunakan untuk mengurangi kata-kata berinfleksi dalam bahasa Indonesia ke bentuk dasarnya (*stem*) (Budiman & Widjaja, 2020).

2.2.6 Preprocessing

Preprocessing merupakan merupakan tahapan sebelum proses pengklasifikasian yang diperlukan untuk membersihkan, menghilangkan, mengubah sumber data, baik itu berupa karakter non-alfabet maupun kata-kata yang tidak diperlukan tujuannya agar data yang digunakan lebih optimal (Muttaqin, 2016). *Preprocessing* mempunyai tugas dan langkah penting dalam *text mining*. Data *preprocessing* digunakan untuk mengekstraksi pengetahuan yang menarik dan penting serta dari data teks yang tidak terstruktur. Berdasarkan ketidak teraturan struktur dalam data teks, maka proses sistem pada *text mining* memerlukan beberapa tahap awal agar teks dapat di ubah menjadi lebih terstruktur (Hermawan & Bellaniar Ismiati, 2020). Proses *preprocessing* ini meliputi:

1. *Case Folding*

Pada tahap ini yang dilakukan untuk menyeragamkan karakter pada data yaitu mengubah seluruh *Text tweet* yang sudah di-*crawling* sebelumnya menjadi huruf kecil atau *lowercase*. Tujuannya adalah untuk mengurangi kesalahan acak (*noise*) pada data.

2. *Filtering*

Tahap ini merupakan proses menghapus simbol dan karakter yang tidak dibutuhkan dalam sebuah teks seperti *hashtag* (#), *mentions* (@), angka, tanda baca, emoji, menghapus spasi yang berlebih serta menghapus kata-kata yang sangat umum.

3. *Tokenizing*

Proses memisahkan kata perkata dalam setiap kalimat kemudian mendapatkan hasil akhir berupa serangkaian token pada kalimat dokumen *tweet*. Token berupa kata, frasa, atau elemen lain yang memiliki makna atau representasi tertentu dalam bahasa.

4. *Stemming*

Proses mengubah suatu kata menjadi bentuk dasarnya dengan menghilangkan imbuhan. Dalam penelitian ini proses *stemming* dilakukan dengan menggunakan *package* Sastrawi yang merupakan modul *stemmer* bahasa Indonesia yang dibuat dengan bahasa pemrograman python yang memiliki fungsi untuk melakukan *stemming* pada kata-kata dalam bahasa Indonesia.

2.2.7 *Tf-IDF*

TF-IDF adalah merupakan cara untuk memberikan bobot hubungan suatu kata terhadap dokumen. Pada metode ini proses perhitungan bobot teks dilakukan dengan menghitung frekuensi kemunculan kata dalam dokumen dan *inverse* frekuensi dari kata tersebut. Semakin tinggi frekuensi suatu teks terhadap suatu dokumen menunjukkan bahwa hubungan kata terhadap suatu dokumen juga semakin tinggi (Hendra Saputra dkk., 2021). *Term frequency* (TF) adalah rasio dari jumlah kemunculan kata dalam dokumen dan *Inverse Document Frequency* (IDF) adalah

kemunculan kata terhadap keseluruhan dokumen dalam data. Adapun rumus perhitungan *Tf-Idf* dapat dilihat sebagai berikut.

- a. Rumus perhitungan TF

$$tf_{t,d} = \frac{n_{t,d}}{N} \quad (2.1)$$

Keterangan:

$n_{t,d}$ = Nilai istilah yang muncul

N = Total dokumen

tf = Frekuensi kemunculan kata pada sebuah dokumen

- b. Rumus perhitungan IDF

$$idf_d = \log \left(\frac{N}{df} \right) \quad (2.2)$$

Keterangan:

N = Total semua dokumen

df = Banyak dokumen yang mengandung term tersebut

- c. Rumus perhitungan TF-IDF

$$tfidf_{t,d} = tf_{t,d} \times idf_{t,d} \quad (2.3)$$

Keterangan:

$tf - idf$ = *Term Frequency-Inverse Document Frequency*

tf = Nilai TF

idf = Nilai IDF

2.2.8 Klasifikasi Naïve Bayes

Naïve Bayes merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan data dalam analisis sentimen dengan model statistik untuk menghitung peluang dari suatu kelas yang memiliki masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal. Naïve Bayes banyak digunakan dalam teknik klasifikasi terutama X menggunakan beberapa metode seperti *Unigram Naïve Bayes*, *Multinomial Naïve Bayes*, dan *Maximum Entropy Classification* (Parikh dan Movassate, 2009).

Perhitungan kategori probabilitas dalam menggunakan pendekatan algoritma Naïve Bayes dengan persamaan :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.4)$$

Keterangan :

P : Sampel data yang memiliki *class* (label) yang tidak diketahui.

A : Hipotesis bahwa P adalah data *class* (label).

P(B) : Probabilitas hipotesis B

P(A) : Peluang dari data sampel yang di amati (probabilitas A).

P(B|A) : Probabilitas berdasarkan kondisi pada hipotesis.

Dalam klasifikasi Naïve Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis ditunjukkan pada persamaan (2.5) sebagai berikut :

$$P(C|F1 \dots Fn) = \frac{P(F1 \dots Fn|C)p(C)}{P(F1 \dots Fn)} \quad (2.5)$$

Sebagaimana variabel C mempresentasikan kelas, sementara variabel F1Fn mempresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*), dikali dengan peluang kemunculan karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana pada persamaan (2.6) sebagai berikut:

$$Posterior = \frac{Prior \times likelihood}{evidence} \quad (2.6)$$

Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan (Siswandi & Fitriana, 2019).

$$P(V1|C) = \frac{CountTerms(v1,docsv(c))+1}{AllTerms(docs(C))+|V|} \quad (2.7)$$

Di mana |V| menunjuk pada jumlah semua kata dalam *tweet* yang ada didataset. Untuk dapat mengklasifikasikan suatu *tweet*, dapat dilakukan dengan persamaan:

$$P(V1|C) = \frac{CountTerms(v1,docsv(c))}{AllTerms(docs(C))} \quad (2.8)$$

Di mana v_1 dalam persamaan di atas adalah satu kata tertentu dalam *tweet*, sedangkan $CountTerms(v_1, docs(c))$ menunjukkan pada jumlah kemunculan suatu kata berlabel c (“positif” atau “negatif”). $AllTerms(docs(c))$ menunjuk pada jumlah semua kata berlabel c yang ada pada dataset.

2.2.9 Performance Evaluation Measure (PEM)

Performance Evaluation Measure (PEM) atau dalam Bahasa Indonesia bisa disebut pengukuran evaluasi performa adalah satu bundel tahapan yang digunakan untuk mengukur performa suatu sistem. PEM dalam banyak kasus digunakan dalam training data, tujuannya untuk mengevaluasi model yang sudah dibuat (Amaniputra dkk., 2023). *Performance Evaluation Measure* (PEM) biasanya digambarkan dalam *confusion matrix*. *Confusion matrix* yang merupakan sebuah alat yang digunakan untuk menentukan performa model Naïve Bayes untuk mengklasifikasikan data, yang diterapkan untuk mengevaluasi hasil pengujian pada sistem dalam analisis sentimen (Apriani & Gustian, 2019). *Confusion matrix* sebagai metode pengukuran akurasi dan presisi model, meliputi model pada *deep learning*, *supervised learning*, dan *unsupervised learning*. *Confusion matrix* merupakan sebuah tabel yang memiliki 4 (empat) kombinasi nilai prediksi (*predicted values*) dan nilai aktual (*actual values*) seperti gambar berikut (Wang dkk., 2022).

Tabel 2.1 *Confusion Matrix* (Wang dkk., 2022)

		<i>Predicted</i>	
		<i>Negative</i>	<i>Positive</i>
<i>Actual</i>	<i>Negative</i>	<i>True Negative</i> (TN)	<i>False Positive</i> (FP)
	<i>Positive</i>	<i>False Negative</i> (FN)	<i>True Positive</i> (TP)

Keterangan :

TP (*True Positive*) : Data dalam kelas positif diprediksi benar sebagai kelas positif (hasil prediksi positif dari *machine learning* sama dengan data opini yang berlabel positif).

FN (*False Negative*) : Data dalam kelas negatif diprediksi bukan sebagai kelas negatif (hasil prediksi negatif dari *machine learning*, tetapi data opini menunjukkan label positif)

TN (*True Negative*) : Data dalam kelas negatif diprediksi benar sebagai kelas negatif (hasil prediksi negatif dari *machine learning* sama dengan data opini yang berlabel negatif)

FP (*False Positive*) : Data dalam kelas positif diprediksi bukan sebagai kelas positif (hasil prediksi positif dari *machine learning*, tetapi data opini menunjukkan label negatif)

Setelah data uji dimasukkan ke dalam *confussion matrix*, selanjutnya menghitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah akurasi, presisi, *recall* dan *f1-score*. Presisi digunakan untuk membandingkan jumlah *false positive* terhadap jumlah *record positive*. Sedangkan *recall* perbandingan jumlah *false negative* terhadap jumlah *record negative*. Untuk menghitung akurasi digunakan persamaan di bawah ini :

A. Akurasi

Perbandingan antara jumlah data yang di prediksi benar positif dan negatif dengan keseluruhan data.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.9)$$

Keterangan :

TP : Jumlah dokumen yang benar dari label positif

TN : Jumlah dokumen yang benar dari label negatif

FP : Jumlah dokumen yang salah dari label positif

FN : Jumlah dokumen yang salah dari label negatif

B. Presisi

Tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan oleh sistem persamaan. Presisi ini digunakan untuk mengukur kejadian *false positive*, apabila presisi rendah berarti tingkat kejadian *false positive* tinggi dan juga sebaliknya.

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (2.10)$$

Keterangan :

TP : Jumlah dokumen yang benar dari label positif

FP : Jumlah dokumen yang salah dari label positif

C. *Recall*

Keberhasilan sistem dalam menemukan kembali sebuah informasi persamaan. *Recall* digunakan untuk mengukur kejadian *false negative*, apabila *recall* rendah berarti tingkat kejadian *false negative* tinggi dan juga sebaliknya.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.11)$$

Keterangan :

TP : Jumlah dokumen yang benar dari label positif

FN : Jumlah dokumen yang salah dari label negatif

D. *F1-Score*

F1-score merupakan nilai perbandingan rata – rata antara *recall* dan *precision*. Rumus *F1-Score* :

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.12)$$

Keterangan :

Precision : Jumlah perhitungan presisi

Recall : Jumlah perhitungan *recall*

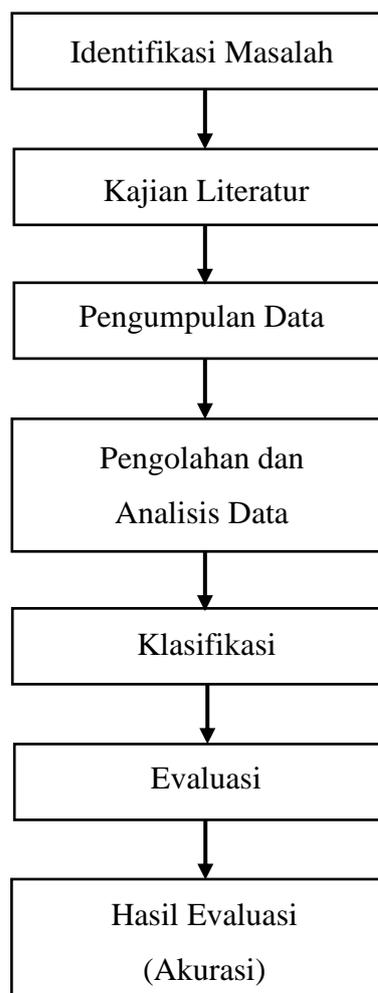
BAB III METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini dilakukan menggunakan metode kuantitatif dengan pendekatan *text mining*. *Text mining* merupakan pendekatan yang digunakan untuk menangani permasalahan klasifikasi, *clustering*, *information extraction* dan *information retrieval* (Firdaus & Firdaus, 2021). *Text mining* bertujuan untuk menemukan informasi berharga yang tersembunyi baik dari sumber informasi terstruktur dan tidak terstruktur.

3.2 Alur Penelitian

Berikut ini merupakan alur penelitian yang dibuat oleh peneliti :



Gambar 3.1 : Alur Penelitian

1. Identifikasi Masalah

Sebelum melakukan penelitian, langkah pertama yang dilakukan penulis yaitu perlu untuk mengidentifikasi, memilih, dan merumuskan masalah yang ada. Identifikasi masalah ini dilakukan dengan cara observasi atau pengamatan dan studi literatur. Observasi atau pengamatan dilakukan penulis secara langsung dan terencana. Sedangkan studi literatur dilakukan dengan mengumpulkan dan membaca jurnal-jurnal penelitian sebelumnya yang berkaitan.

2. Kajian Literatur

Kajian literatur dilakukan agar penulis mengetahui apakah topik yang dipilih belum pernah diteliti oleh peneliti sebelumnya. Selain itu, mengumpulkan penelitian yang sudah dilaksanakan oleh peneliti lain untuk mendapatkan ide-ide baru dan memperkaya pengetahuan.

3. Pengumpulan Data

Pada penelitian ini data yang digunakan adalah teks *tweet* berbahasa Indonesia mengenai respon atau opini terhadap iPusnas dikumpulkan dari 01 Januari 2022 sampai 03 Oktober 2023. Data yang dikumpulkan berupa data teks yang di ambil dengan teknik *crawling* data menggunakan *Twitter Harvest*. Hasil *crawling* data disimpan dalam file format *excel (csv)*. Variabel metadata hasil *crawling* data yang akan diambil adalah *username*, teks *tweet*, tanggal dan waktu *tweet*.

4. Pengolahan dan Analisis Data

Proses mengubah dataset mentah ke dalam bentuk teks yang lebih terstruktur. Tujuan dari tahap ini adalah untuk membersihkan dataset komentar hasil *crawling* X sehingga menjadi data yang bersih dan siap dilakukan pengolahan dalam analisis data. Langkah-langkah *preprocessing* yang dilakukan disesuaikan berdasarkan hasil data yang telah di-*crawling* mengenai komentar atau opini masyarakat terhadap aplikasi iPusnas.

5. Klasifikasi

Dalam melakukan proses klasifikasi menggunakan algoritma Naïve Bayes, dataset komentar pada penelitian ini harus melalui tahap pembagian dataset terlebih dahulu menjadi data latih dan data uji.

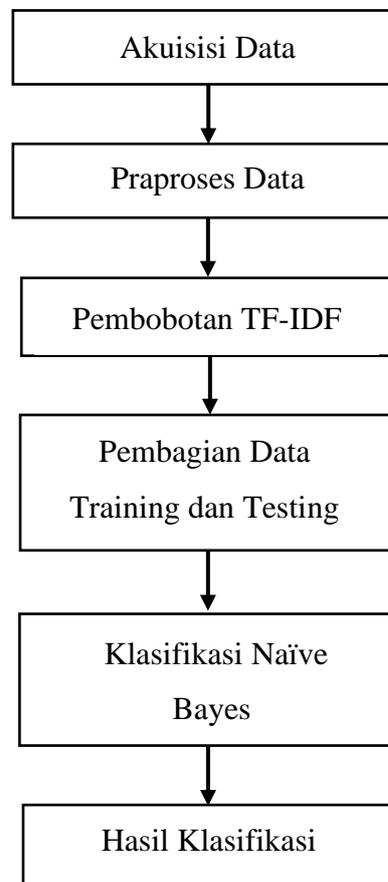
6. Evaluasi

Proses evaluasi model dilakukan untuk mengetahui akurasi, presisi, *recall* dan *f1-score* dari proses klasifikasi yang telah dilakukan oleh sistem.

7. Hasil Evaluasi (Akurasi)

Hasil evaluasi model selanjutnya akan diambil dari hasil akurasi pada setiap rasio.

Tahapan pengambilan data hingga pengolahan data dijelaskan lebih detail pada gambar 3.2 berikut.



Gambar 3.2 : Desain Sistem

1. Akuisisi Data

Pada proses ini dilakukan *crawling* data X menggunakan *library twitter harvest* sebagai pengambilan data sekaligus pembuatan dataset. Proses *information retrieval* dikendalikan dengan beberapa filter, seperti teks berbahasa Indonesia, kata kunci “iPusnas”.

2. Praproses Data

Tahapan praproses data terdiri dari beberapa langkah. Proses menggunakan *tools* Kaggle dengan bahasa python. Adapun beberapa tahap *preprocessing* yang dilakukan adalah :

a. Case folding

Tahap yang digunakan mengubah seluruh Text *tweet* yang sudah di *Crawling* sebelumnya menjadi huruf kecil atau *lowercase*. Tujuannya adalah untuk mengurangi kesalahan acak (*noise*) pada data.

Tabel 3.1 Contoh proses penerapan *Case Folding*

Sebelum	Sesudah
kemarin pas mau baca ipusnas gue eror dan bukunya kereturn ðŸ˜-ðŸ˜-ðŸ˜-ðŸ˜- mana baru baca 20an persen yampunn sedih	kemarin pas mau baca ipusnas gue eror dan bukunya kereturn ðŸ˜-ðŸ˜-ðŸ˜-ðŸ˜- mana baru baca 20an persen yampunn sedih
Haaaa iseng nyari wishlist malah dapat di iPusnas https://t.co/3FC5oz6JMU	haaaa iseng nyari wishlist malah dapat di ipusnas https://t.co/3FC5oz6JMU
@literarybase Agak kecewa sih sama ipusnas. Kukira e-book penunjang referensi ku buat TA akan sangat amat lengkap di sini huahhhh nyatanya enggak. Semoga kedepannya banyak e-book yang keluaran penerbit baru yah dan lebih banyak lagi kutungguuu loh @perpusn	@literarybase agak kecewa sih sama ipusnas. kukira e-book penunjang referensi ku buat ta akan sangat amat lengkap di sini huahhhh nyatanya enggak. semoga kedepannya banyak e-book yang keluaran penerbit baru yah dan lebih banyak lagi kutungguuu loh @perpusn

Sumber : Hasil data penelitian 2023

b. Filtering

Proses menghapus simbol dan karakter yang tidak dibutuhkan dalam sebuah teks seperti hashtag (#), mentions (@), angka, tanda baca, emoji, dan menghapus spasi yang berlebih.

Tabel 3.2.1 Contoh proses penerapan *Filtering*

Sebelum	Sesudah
kemarin pas mau baca ipusnas gue eror dan bukunya kereturn ðŸ˜-ðŸ˜-ðŸ˜-ðŸ˜- mana baru baca 20an persen yampunn sedih	kemarin pas mau baca ipusnas gue erorr dan bukunya kereturn mana baru baca 20an persen yaampun sedih

Sebelum	Sesudah
haaaa iseng nyari wishlist malah dapat di ipusnas https://t.co/3FC5oz6JMU	ha iseng nyari wishlist malah dapat di ipusnas
@literarybase agak kecewa sih sama ipusnas. kukira e-book penunjang referensi ku buat ta akan sangat amat lengkap di sini huahhhh nyatanya enggak. semoga kedepannya banyak e-book yang keluaran penerbit baru yah dan lebih banyak lagi kutungguuu loh @perpusn	agak kecewa sih sama ipusnas kukira ebook penunjang referensi ku buat tugas akhir akan sangat amat lengkap di sini huah nyatanya enggak semoga kedepannya banyak ebook yang keluaran penerbit baru yah dan lebih banyak lagi kutunggu loh

Sumber : Hasil data penelitian 2023

c. Tokenizing

Proses memisahkan kata perkata dalam setiap kalimat kemudian mendapatkan hasil akhir berupa serangkaian token pada kalimat yang dokumen *tweet*. Token berupa kata, frasa, atau elemen lain yang memiliki makna atau representasi tertentu dalam bahasa.

Tabel 3.2.2 Contoh proses penerapan *Tokenizing*

Sebelum	Sesudah
kemarin pas mau baca ipusnas gue erorr dan bukunya kereturn mana	['kemarin', 'pas', 'mau', 'baca', 'ipusnas', 'gue', 'erorr', 'dan', 'bukunya', 'ke', 'return', 'mana']
baru baca 20an persen yaampun sedih	['baru', 'baca', '20', 'an', 'persen', 'yaampun', 'sedih']
ha iseng nyari wishlist malah dapat di ipusnas	['ha', 'iseng', 'nyari', 'wishlist', 'malah', 'dapat', 'di', 'ipusnas']
agak kecewa sih sama ipusnas kukira ebook penunjang referensi ku buat tugas akhir akan sangat amat lengkap di sini huah nyatanya enggak semoga kedepannya banyak ebook yang keluaran penerbit baru yah dan lebih banyak lagi kutunggu loh	['agak', 'kecewa', 'sih', 'sama', 'ipusnas', 'ku', 'kira', 'ebook', 'penunjang', 'referensi', 'ku', 'buat', 'tugas', 'akhir', 'akan', 'sangat', 'amat', 'lengkap', 'di', 'sini', 'huah', 'nyatanya', 'enggak', 'semoga', 'kedepannya', 'banyak', 'ebook', 'yang', 'keluaran', 'penerbit', 'baru', 'yah', 'dan']

Sebelum	Sesudah
	'lebih', 'banyak', 'lagi', 'ku', 'tunggu', 'loh']

Sumber : Hasil data penelitian 2023

d. Stemming

Proses mengubah suatu kata menjadi bentuk dasarnya dengan menghilangkan imbuhan. Dalam penelitian ini proses *stemming* dilakukan dengan menggunakan *library* Sastrawi berdasarkan kamus kata dasar <http://kateglo.lostfocus.org/>. *Stemming* dalam Penerapan *Stemmer* Sastrawi dipilih karena lebih baik daripada Porter dan Snowball untuk stemmer dokumen berbahasa Indonesia (Hendrawan dkk., 2021).

Tabel 3.2.3 Contoh proses penerapan *Stemming*

Sebelum	Sesudah
['kemarin', 'pas', 'mau', 'baca', 'ipusnas', 'gue', 'erorr', 'dan', 'bukunya', 'ke', 'return', 'mana', 'baru', 'baca', '20', 'an', 'persen', 'yaampun', 'sedih']	baca ipusnas erorr buku return baru baca 20 persen sedih
['ha', 'iseng', 'nyari', 'wishlist', 'malah', 'dapat', 'di', 'ipusnas']	cari wishlist dapat ipusnas
['agak', 'kecewa', 'sih', 'sama', 'ipusnas', 'ku', 'kira', 'ebook', 'penunjang', 'referensi', 'ku', 'buat', 'tugas', 'akhir', 'akan', 'sangat', 'amat', 'lengkap', 'di', 'sini', 'huah', 'nyatanya', 'enggak', 'semoga', 'kedepannya', 'banyak', 'ebook', 'yang', 'keluaran', 'penerbit', 'baru', 'yah', 'dan', 'lebih', 'banyak', 'lagi', 'ku', 'tunggu', 'loh']	kecewa sama ipusnas kira ebook penunjang referensi untuk tugas akhir akan lengkap nyata tidak semoga depan banyak ebook keluar terbit baru banyak lagi tunggu

Sumber : Hasil data penelitian 2023

3. Pembobotan TF-IDF

Setelah semua data sudah melewati tahap *preprocessing*, selanjutnya adalah melakukan ekstraksi fitur untuk mempermudah proses klasifikasi. Pada tahap ekstraksi fitur dilakukan perubahan fitur kata atau teks menjadi sebuah representasi angka (numerik) dengan melihat kemunculan fitur kata atau teks

terlebih dahulu dengan metode *CountVectorizer*, selanjutnya mencari pembobotan kata menggunakan metode *Tf-Idf*. Metode *Tf-Idf* akan melakukan perhitungan nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap fitur kata di setiap dokumen dalam dataset.

4. Pembagian Data Training dan Data Testing

Setelah melalui tahapan praproses data yang menghasilkan data berupa vector maka berikutnya adalah tahapan klasifikasi dengan menggunakan *Naïve Bayes Classifier* dengan menggunakan *library sklearn* pada python. Data dibagi menjadi dua bagian yaitu data latih dan data uji, data latih sebanyak 90%, 80%, 70% dan data uji sebanyak 10%, 20%, 30%.

5. Proses Klasifikasi Naïve Bayes

Pada tahap ini klasifikasi menggunakan *Multinomial Naïve Bayes* dalam melakukan prediksi kelas positif atau negatif dengan metode validasi model *train-test split validation*, menguji data serta mempelajari pengetahuan yang terkandung dalam data latih. Akan terdapat dua proses yang akan dilakukan sebagai berikut:

1. Melatih Model Naïve Bayes

Pada tahap melatih model *Naïve Bayes* dataset akan dibagi menjadi dua subset data latih dan data uji, 90%, 80%, 70% subset data latih akan digunakan untuk melatih model *Naïve Bayes* yang di mana *Naïve Bayes* akan mempelajari pola dari data pelatihan untuk mengklasifikasikan sentimen pada suatu data teks komentar apakah termasuk sentimen positif atau negatif.

2. Menguji Model Naïve Bayes

Pada tahap menguji model dari *Naïve Bayes* akan dilakukan klasifikasi pada data yang belum pernah dilihat oleh model. 10%, 20%, 30% data uji akan digunakan oleh model *Naïve Bayes* yang telah dilatih pada subset pelatihan untuk memprediksi kelas dari data pengujian. *Naïve Bayes* akan menggunakan pola yang telah dipelajari selama pelatihan untuk dapat mengklasifikasikan data menjadi kelas yang sesuai pada tahap uji yaitu melakukan klasifikasi sentimen pada data *tweet* tentang *iPusnas* apakah positif atau negatif.

6. Hasil Klasifikasi

Hasil dari klasifikasi akan dievaluasi oleh model untuk mengetahui performa kinerja pada suatu sistem. Tahapan ini digunakan untuk mengetahui keakuratan model klasifikasi yang dihasilkan. Pengukuran kinerja klasifikasi dilihat melalui *accuracy*, *precision*, *recall*, dan *f1-score*.

3.3 Sumber Data

Data yang digunakan dalam penelitian ini adalah kumpulan *tweets*, *quote tweets*, dan *reply* yang terkait topik iPusnas di media sosial X. Data yang diambil adalah *tweets* mulai dari tanggal 01 Januari 2022 sampai 03 Oktober 2023. Data *tweets* yang diperoleh adalah sebanyak 2379.

3.4 Pengumpulan Data

Penelitian ini menerapkan metode *crawling* untuk mengakses informasi dan mengumpulkan data. Dalam proses pengumpulan data menggunakan google colab yang di dalamnya sudah menggunakan tweet harvest untuk melakukan *crawl* data . Adapun tampilan *source code* untuk pengambilan data pada tabel 3.2.

```
# Crawl Data
filename = 'ipusnas.csv'
search_keyword = 'ipusnas lang:id until:2023-10-03 since:2022-01-01'
limit = 3000

!npx --yes tweet-harvest@2.2.8 -o "{filename}" -s "{search_keyword}" -l {limit} --token ""
```

Gambar 3.3 *Source Code* Hasil *Crawl* Data *Tweet Harvest*
 Sumber : Hasil data penelitian 2023

Langkah yang dilakukan dalam proses pengambilan data X pada tweet harvest yaitu menggunakan kata kunci “iPusnas”, *tweet* berbahasa Indonesia dengan rentang tanggal 01 Januari 2022 sampai 03 Oktober 2023. Proses ini akan dilakukan secara otomatis berdasarkan input yang sudah ditentukan di atas. Setelah selesai maka output yang diperoleh dari data yang diambil akan langsung disajikan dalam bentuk file *Comma-Separated Values* (CSV). Selanjutnya, data yang sudah terkumpul dikonversikan menjadi format .xlsx agar memudahkan dalam proses pelabelan data oleh ahli bahasa lulusan dari Universitas Gajah Mada dengan jurusan Sastra Indonesia.

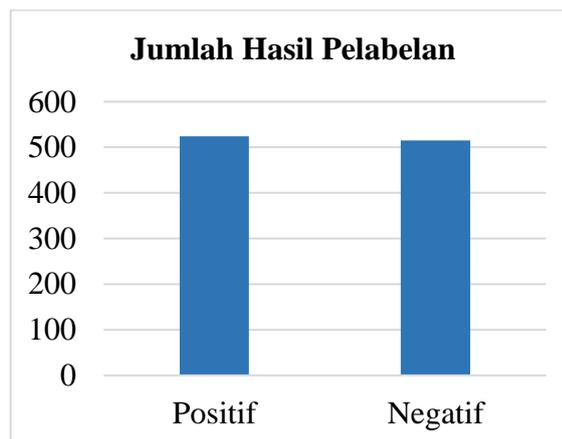
BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pada bab ini membahas hasil pengujian *machine learning* serta membahas implementasi metode Naïve Bayes untuk mengetahui sentimen pengguna X terhadap aplikasi iPusnas. Terdapat tahap *preprocessing*, TF-IDF, pembagian data menjadi data latih dan data uji menggunakan rasio 90:10, 80:20, dan 70:30. Pengujian pada penelitian ini dilakukan untuk menghitung akurasi, *recall*, presisi, dan *F1-Score*. Adapun beberapa tahap atau langkah-langkah hasil penelitian yang dilakukan.

4.1.1 Hasil Pelabelan Data

Pada penelitian ini, rentang data yang digunakan yaitu tanggal 01 Januari 2022 sampai 03 Oktober 2023 yang menghasilkan data awal sebanyak 2379 *tweet*. Selanjutnya, dilakukan pembersihan dan cek duplikasi pada data sehingga menghasilkan data bersih sebanyak 1036 *tweet*. Data dari hasil pembersihan tersebut diberi label oleh ahli bahasa Indonesia dengan memutuskan jenis kelas sentimen yaitu positif dan negatif. Hasil dari pelabelan dapat dikelompokkan dan dihitung jumlah masing-masing kelas seperti diagram pada gambar 4.1.



Gambar 4.1 Hasil Labeling Data

(Sumber: Hasil olah data peneliti, 2023)

Gambar 4.1 merupakan visualisasi berupa diagram batang di mana garis horizontal menunjukkan jenis kelas sentimen, sedangkan garis vertikal

menunjukkan jumlah pada masing-masing kelas sentimen. Kelas sentimen positif memiliki jumlah 524, dan kelas sentimen negatif memiliki jumlah 515. Pada tabel 4.1 merupakan hasil pelabelan data *tweet*.

Tabel 4.1 Tabel hasil pelabelan data

No	Comment	Label
1	iyaaaa baca ipusnas jujur seru puas ending wkwkkkw	Positif
2	lt suka buku klasik coba baca buku stock ipusnas	Positif
3	suka baca baca buku bokek minggu rencana beli buku baru tsana ipusnas adaa serial bumi ipusnas bay	Positif
4	baca ipusnas ngantri	Negatif
5	kesal ipusnas error buku pinjaman hilang	Negatif

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.1 merupakan hasil pelabelan pada data *tweet*. Pada hasil *comment* data *tweet* 1 sampai 3 berlabel positif karena mengandung beberapa kata positif yaitu, seru, puas, dan suka. Sedangkan, untuk *comment* data *tweet* 4 dan 5 berlabel negatif karena mengandung beberapa kata negatif yaitu, ngantri, kesal, *error*, dan hilang.

4.1.2 Hasil Preprocessing

Data yang telah dilabeli oleh ahli bahasa Indonesia akan dibersihkan untuk memaksimalkan proses klasifikasi. *Preprocessing* menggunakan *tools* kaggle dengan bahasa python untuk mencari dan mengganti partikel yang tidak diperlukan, seperti URL, *emoticon*, *hashtag*, dan *username/mention* akan diganti dengan *string* kosong. Selain itu, memanfaatkan *library* Sastrawi untuk proses *tokenizing* dan *stemming*. Adapun pada proses *tokenizing* memanfaatkan *library* NLTK dan pembuatan *dataframe* menggunakan *pandas*. Hasil *case folding*, *filtering*, *tokenizing*, dan *stemming* sebagai berikut.

Tabel 4.2 Data Sebelum *Preprocessing*

No	Data Sebelum <i>Preprocessing</i>
1	Sekian lama nungguin slot ipusnas akhirnya dapet juga 🤔🤔,🤔🤔
2	sama kayak heroes of olympus, buku kedua ga ada di ipusnas :) MANA AKU UDAH KEBELI TRIALS OF APOLLO, JADI KETUNDA SAMPE GTW KAPAN INI BACANYA
3	/lt hi warga litbase, minta rekomendasi mystery thriller di ipusnas selain Kesetiaan Mr X, Second Sister, dan karya Agatha Christie donggg, makasiii 🤔🤔,🤔🤔

No	Data Sebelum <i>Preprocessing</i>
4	Aku juga ngerasa gitu, tapi aku gamasalah dengan pengulangan topiknya sih Cuma, sayang bgt aja gitu, topiknya dah diulang tapi bahasan cuma
	kumpulan kata kata cantik aja Aku nungguin ini di Ipusnas dah lama bgt loh soalnya Berasa gimana g
5	Bukan web tapi aplikasi dari perpustakaan (iPusnas) Sistemnya pinjem jadi baca online aja, ada tenggat waktu pengembalian tapi bisa pinjem berulang kali ko

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.1 menampilkan hasil data sebelum di *preprocessing*. Selanjutnya, melakukan *case folding* pada tabel 4.3.

Tabel 4.3 Hasil *Case Folding*

No	Case Folding
1	sekian lama nungguin slot ipusnas akhirnya dapet juga
2	sama kayak heroes of olympus buku kedua ga ada di ipusnas mana aku udah kebeli trials of apollo jadi ketunda sampe gtw kapan ini bacanya
3	It hi warga litbase minta rekomendasi mystery thriller di ipusnas selain kesetiaan mr x second sister dan karya agatha christie donggg makasiii
4	aku juga ngerasa gitu tapi aku gamasalah dengan pengulangan topiknya sih cuma sayang bgt aja gitu topiknya dah diulang tapi bahasan cuma kumpulan kata kata cantik aja aku nungguin ini di ipusnas dah lama bgt loh soalnya berasa gimana g
5	bukan web tapi aplikasi dari perpustakaan ipusnas sistemnya pinjem jadi baca online aja ada tenggat waktu pengembalian tapi bisa pinjem berulang kali ko

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.3 yaitu penghapusan URL, *emoticon*, *hashtag*, *username/mention*, *whitespace*, serta menyelaraskan semua huruf menjadi *lowercase*. Setelah itu melakukan filtering dengan hasil yang ada pada tabel 4.4.

Tabel 4.4 Hasil *Filtering*

No	Filtering
1	sekian lama nungguin slot ipusnas akhirnya dapat juga
2	sama seperti heroes of olympus buku kedua nggak ada di ipusnas mana aku udah kebeli trials of apollo jadi ketunda sampe gtw kapan ini bacanya
3	It hi warga litbase minta rekomendasi mystery thriller di ipusnas selain kesetiaan mr x second sister dan karya agatha christie donggg makasiii
4	aku juga ngerasa gitu tapi aku gamasalah dengan pengulangan topiknya sih cuma sayang banget aja gitu topiknya dah diulang tapi aku gamasalah

No	Filtering
	dengan pengulangan topiknya sih cuma sayang banget aja gitu topiknya dah diulang tapi bahasan cuma kumpulan kata kata cantik aja aku nungguin ini di ipusnas dah lama banget loh soalnya berasa gimana g
5	bukan web tapi aplikasi dari perpustakaan ipusnas sistemnya pinjem jadi baca online aja ada tenggat waktu pengembalian tapi bisa pinjem berulang kali ko

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.4 menampilkan hasil penghapusan simbol dan karakter yang tidak dibutuhkan dalam sebuah teks seperti *hashtag* (#), *mentions* (@), angka, tanda baca, emoji, menghapus spasi yang berlebih serta menghapus kata-kata yang sangat umum. Setelah itu, melakukan *tokenizing* dengan hasil yang ada pada tabel 4.5.

Tabel 4.5 Hasil *Tokenizing*

No	Tokenizing
1	['sekian', 'lama', 'nungguin', 'slot', 'ipusnas', 'akhirnya', 'dapat', 'juga']
2	['sama', 'seperti', 'heroes', 'of', 'olympus', 'buku', 'kedua', 'nggak', 'ada', 'di', 'ipusnas', 'mana', 'aku', 'udah', 'kebeli', 'trials', 'of', 'apollo', 'jadi', 'ketunda', 'sampe', 'gtw', 'kapan', 'ini', 'bacanya']
3	['It', 'hi', 'warga', 'litbase', 'minta', 'rekomendasi', 'mystery', 'thriller', 'di', 'ipusnas', 'selain', 'kesetiaan', 'mr', 'x', 'second', 'sister', 'dan', 'karya', 'agatha', 'christie', 'donggg', 'makasiii']
4	['aku', 'juga', 'ngerasa', 'gitu', 'tapi', 'aku', 'gamasalah', 'dengan', 'pengulangan', 'topiknya', 'sih', 'cuma', 'sayang', 'banget', 'aja', 'gitu', 'topiknya', 'dah', 'diulang', 'tapi', 'bahasan', 'cuma', 'kumpulan', 'kata', 'kata', 'cantik', 'aja', 'aku', 'nungguin', 'ini', 'di', 'ipusnas', 'dah', 'lama', 'banget', 'loh', 'soalnya', 'berasa', 'gimana', 'g']
5	['bukan', 'web', 'tapi', 'aplikasi', 'dari', 'perpustakaan', 'ipusnas', 'sistemnya', 'pinjem', 'jadi', 'baca', 'online', 'aja', 'ada', 'tenggat', 'waktu', 'pengembalian', 'tapi', 'bisa', 'pinjem', 'berulang', 'kali', 'ko']

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.5 menampilkan hasil pemisahan setiap kata sehingga data yang tertokenisasi akan diberi bobot menggunakan TF-IDF. Kemudian, melakukan *stemming* dengan hasil yang ada pada tabel 4.6.

Tabel 4.6 Hasil *Stemming*

No	Stemming
1	sekian nungguin slot ipusnas
2	heroes olympus buku ipusnas kebel trials apollo tunda sampe gtw baca
3	It warga litbase rekomendasi mystery thriller ipusnas setia second sister

No	Stemming
	karya agatha christie dong makasi
4	ngerasa gamasalah ulang topik sayang topik ulang bahas kumpul cantik nungguin ipusnas soalnya
5	web aplikasi perpunas ipusnas sistem pinjem baca online tenggat kembali pinjem ulang

(Sumber: Hasil olah data peneliti, 2023)

Tabel 4.3, tabel 4.4, tabel 4.5, dan tabel 4.6 merupakan *dataframe* yang mengambil lima contoh data teratas dengan total 829 kata dari hasil pembersihan. Data tersebut disimpan dalam format CSV yang siap untuk diberi bobot TF-IDF.

4.1.3 Hasil Pembagian Dataset

Setelah melalui tahap *preprocessing* menghasilkan data bersih sebanyak 1036 data. Selanjutnya, data bersih yang telah didapatkan akan dibagi menjadi data latih dan data uji. Data latih akan digunakan untuk melakukan ekstraksi fitur TF-IDF serta melatih model klasifikasi Naïve Bayes, sedangkan data uji akan digunakan untuk mengukur kinerja model agar dapat melihat persentase algoritma klasifikasi yang berhasil mengklasifikasikan dengan benar.

Hasil dari pembagian dataset ini akan disimpan kedalam *file excel* untuk digunakan pada pengolahan data secara terus menerus tanpa harus mengulangi untuk membagi data. pembagian data dilakukan dengan perbandingan 90%, 80%, 70% data latih dan 10%, 20%, 30% data uji. Setelah dataset dilakukan pembagian, dataset latih dan dataset uji akan disimpan menjadi *file excel*.

4.1.4 Hasil Pembobotan Kata

Proses melakukan ekstraksi fitur dengan TF-IDF untuk pembobotan kata. Hasil dari ekstraksi memunculkan nilai bobot berupa angka desimal. Pada dokumen sejumlah 1036, terdapat 829 kata yang terbobot akan digunakan sebagai klasifikasi pada pelatihan data. Hasil penerapan ekstraksi fitur dalam bentuk *dataframe*.

	abad	ad	ada	adaa	adalah	adaptasi	admin	after
0	0.0	0.0	0.0	0.523471	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0

Gambar 4.2 Tampilan data ekstraksi fitur dengan *TF-IDF*

(Sumber: Hasil olah data peneliti, 2023)

Gambar 4.2 merepresentasikan matriks TF-IDF (*Term Frequency-Inverse Document Frequency*) setelah proses ekstraksi fitur menggunakan metode *TfidfVectorizer* pada teks. Matriks ini menunjukkan bagaimana kata-kata dalam teks direpresentasikan dalam bentuk vektor numerik berdasarkan konsep TF-IDF. Nilai dalam matriks tersebut adalah bobot TF-IDF untuk setiap kata dalam setiap dokumen. Nilai-nilai ini menunjukkan seberapa penting kata tersebut dalam dokumen teks komentar. Jika nilai bobot TF-IDF suatu kata dalam suatu dokumen mendekati 0, itu berarti kata tersebut tidak begitu relevan dalam dokumen tersebut. Sedangkan, jika nilai TF-IDF adalah 0 berarti sebagian besar dokumen mungkin tidak mengandung kata-kata tertentu.

4.1.5 Hasil Klasifikasi Naïve Bayes

Pembobotan kata telah dilakukan ditahap sebelumnya. Selanjutnya, akan dilakukan tahap data mining. Tahap data mining merupakan tahap klasifikasi data menggunakan Naïve Bayes akan ditampilkan dalam bentuk tabel *Confusion Matrix*. Dalam melakukan klasifikasi menggunakan algoritma *Multinomial Naïve Bayes*, dataset komentar pada penelitian ini harus melalui tahap pembagian dataset terlebih dahulu menjadi data latih dan data uji. Pada tahap sebelumnya, pembagian dilakukan dengan rasio 90:10, 80:20, dan 70:30. Tabel 4.7 berikut menunjukkan perbandingan rasio yang digunakan dalam pembagian dataset untuk menemukan rasio terbaik yang akan diterapkan dalam penelitian ini.

Tabel 4.7 Rasio yang digunakan dalam pembagian

Perbandingan Data Latih & Data Uji	Hasil Akurasi Naïve Bayes
90 : 10	82%
80 : 20	79%
70 : 30	74%

(Sumber : Hasil olah data peneliti, 2023)

Berdasarkan Tabel 4.7, klasifikasi Naïve Bayes menghasilkan nilai tertinggi dengan pembagian data latih dan data uji sebesar 90:10, dengan akurasi mencapai 82%. Kemudian, setelah dilakukan pembagian dataset dan didapatkan perbandingan dengan akurasi terbaik, maka akan dilakukan ekstraksi fitur menggunakan *Tf-Idf* untuk mendapatkan pembobotan dari data latih dan data uji

yang di mana hasil data yang telah melalui tahap ekstraksi fitur menggunakan *Tf-Idf* nantinya akan digunakan untuk melatih dan menguji model klasifikasi Naïve Bayes. Berikut merupakan tabel hasil klasifikasi dalam bentuk confusion matrix.

Tabel 4.8 Hasil Klasifikasi rasio 90:10 dengan *TF-IDF*

		<i>Predicted</i>	
		<i>Negative</i>	<i>Positive</i>
<i>Actual</i>	<i>Negative</i>	46	10
	<i>Positive</i>	8	40

(Sumber : Hasil olah data peneliti, 2023)

Pada tabel 4.8 terdapat 40 data dalam kelas positif diprediksi benar sebagai kelas positif, sedangkan 8 data kelas negatif diprediksi bukan kelas negatif. Lalu sebanyak 46 data kelas negatif benar diprediksi sebagai kelas negatif, sedangkan 10 data kelas positif diprediksi bukan sebagai kelas positif.

Tabel 4.9 Hasil klasifikasi rasio 80:20 dengan *TF-IDF*

		<i>Predicted</i>	
		<i>Negative</i>	<i>Positive</i>
<i>Actual</i>	<i>Negative</i>	81	26
	<i>Positive</i>	16	84

(Sumber : Hasil olah data peneliti, 2023)

Pada tabel 4.9 terdapat 84 data dalam kelas positif diprediksi benar sebagai kelas positif, sedangkan 16 data kelas negatif diprediksi bukan kelas negatif. Lalu sebanyak 81 data kelas negatif benar diprediksi sebagai kelas negatif, sedangkan 26 data kelas positif diprediksi sebagai bukan kelas positif.

Tabel 4.10 Hasil Klasifikasi rasio 70:30 dengan *TF-IDF*

		<i>Predicted</i>	
		<i>Negative</i>	<i>Positive</i>
<i>Actual</i>	<i>Negative</i>	112	52
	<i>Positive</i>	27	119

(Sumber : Hasil olah data peneliti, 2023)

Pada tabel 4.10 terlihat terdapat 119 data dalam kelas positif diprediksi benar sebagai kelas positif, sedangkan 27 data kelas negatif diprediksi bukan kelas negatif. Lalu sebanyak 112 data kelas negatif benar diprediksi sebagai kelas negatif, sedangkan 52 data kelas positif diprediksi bukan sebagai kelas positif.

Berikutnya, beberapa proses penting dalam melatih dan menguji model menggunakan algoritma Naïve Bayes *Classifier* untuk analisis sentimen pengguna X terhadap aplikasi ipusnas akan dijelaskan lebih lanjut.

a. Melatih model Naïve Bayes

Implementasi dari proses pelatihan sebuah model klasifikasi Naïve Bayes dengan menggunakan kelas *MultinomialNBC* yang telah didefinisikan sebelumnya. Dengan memahami proses ini, kita dapat memanfaatkan model Naïve Bayes yang telah dilatih untuk mengklasifikasikan data baru dengan memanfaatkan informasi yang diperoleh dari data latih.

b. Menguji model Naïve Bayes

Model klasifikasi Naïve Bayes yang telah dilatih digunakan untuk memprediksi kelas pada data uji. Data uji dalam bentuk *matrix* dari ekstraksi fitur *Tf-Idf test_tfidf* disimpan dalam variabel *x_test* dan label kelas dari data latih yang disimpan pada variabel *y_test* di inisialisasi untuk menguji model yang dilatih sebelumnya yaitu modelNBC digunakan untuk melakukan prediksi kelas pada data uji dengan memanggil metode *predict(x_test)*. Kemudian, hasil prediksi bisa digunakan untuk mengukur performa model, membandingkannya dengan label kelas yang sebenarnya, serta mendapatkan wawasan lebih lanjut tentang bagaimana model berfungsi pada data yang belum pernah dilihat sebelumnya. Berikut hasil prediksi yang dapat dilihat pada tabel 4.11 berikut.

Tabel 4.11 Hasil prediksi kelas pada data uji

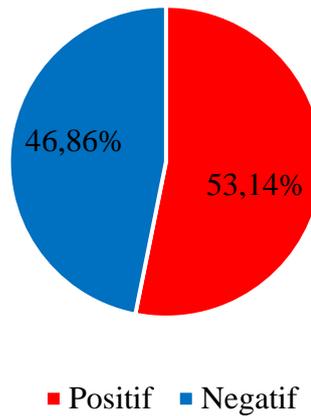
No	Comment	Label	Classification
1	physical tbr numpuk pilih baca buku ipusnas	Negatif	Positif
2	subs mending baca ipusnas ijak jiakh promosi	Positif	Positif
3	iya baca ipusnas jujur seru puas ending wkwkwk	Positif	Positif
4	terjemah diterbitin qanita buku susah nyari fisik baca ipusnas gramdig single edition	Positif	Negatif
5	saranin buku ngantri ipusnas	Negatif	Negatif

(Sumber : Hasil olah data peneliti, 2023)

Berdasarkan proses klasifikasi yang telah dilakukan, algoritma Naïve Bayes menggunakan konsep probabilitas antar kalimat untuk setiap kelas guna

menghasilkan prediksi pada data yang dimasukkan. Berikut gambar diagram yang merupakan hasil analisis Naïve Bayes yang akan digunakan sebagai data uji.

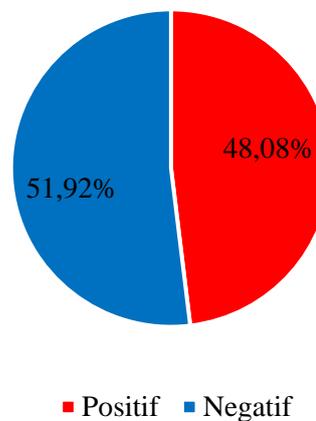
Hasil Analisis Naïve Bayes



Gambar 4.3 Diagram Hasil Analisis Naïve Bayes Rasio 90:10
(Sumber : Hasil olah data peneliti, 2023)

Gambar 4.3 menampilkan hasil analisis Naïve Bayes, di mana satu pertiga dari data digunakan sebagai data testing dan dua pertiga dari data digunakan sebagai data training. Semua data *tweet* positif dan negatif yang berjumlah 104 data akan digunakan sebagai data uji.

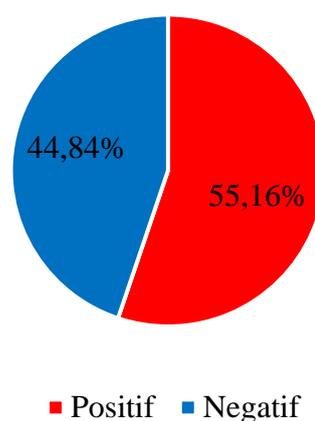
Hasil Analisis Naïve Bayes



Gambar 4.4 Diagram Hasil Analisis Naïve Bayes Rasio 80:20
(Sumber : Hasil olah data peneliti, 2023)

Gambar 4.4 menampilkan hasil analisis Naïve Bayes, di mana satu pertiga dari data digunakan sebagai data testing dan dua pertiga dari data digunakan sebagai data training. Semua data *tweet* positif dan negatif yang berjumlah 207 data akan digunakan sebagai data uji.

Hasil Analisis Naïve Bayes

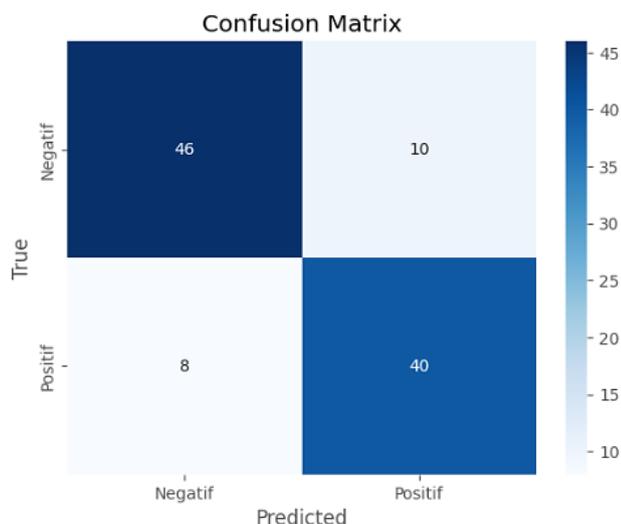


Gambar 4.5 Diagram Hasil Analisis Naïve Bayes Rasio 70:30
(Sumber : Hasil olah data peneliti, 2023)

Gambar 4.5 menampilkan hasil analisis Naïve Bayes, di mana satu pertiga dari data digunakan sebagai data testing dan dua pertiga dari data digunakan sebagai data training. Semua data *tweet* positif dan negatif yang berjumlah 310 data akan digunakan sebagai data uji untuk menguji model klasifikasi dengan algoritma *Multinomial Naïve Bayes Classifier*.

4.1.6 Evaluasi Model

Pada tahap ini, evaluasi model dilakukan untuk mengetahui kebenaran dari proses klasifikasi yang telah dilakukan oleh sistem. Oleh karena itu, dalam evaluasi ini akan diperhatikan nilai-nilai seperti akurasi, presisi, *recall*, dan *F1-Score*. Untuk mendapatkan nilai-nilai tersebut, dilakukan pengujian menggunakan *confusion matrix* guna mengevaluasi hasil prediksi yang dihasilkan oleh model. Pada gambar 4.6 merupakan visualisasi dari hasil evaluasi model dengan rasio 90:10.



Gambar 4.6 *Confusion Matrix* Naïve Bayes rasio 90:10
(Sumber: Hasil olah data peneliti, 2023)

Berdasarkan Gambar 4.6, dataset *tweet* mengenai sentimen pengguna terhadap aplikasi iPusnas dengan algoritma *Multinomial Naïve Bayes Classifier* menggunakan ekstraksi fitur TF-IDF pada rasio 90:10. Terdapat 40 data positif yang diprediksi benar oleh model sebagai sentimen positif (*True Positive*), dan 8 data positif yang diprediksi salah sebagai sentimen negatif (*False Negative*). Selain itu, terdapat 46 data negatif yang diprediksi benar oleh model sebagai sentimen negatif (*True Negative*) dan 10 data negatif yang diprediksi salah sebagai sentimen positif (*False Positive*). Dengan demikian, klasifikasi data berdasarkan *confusion matrix* di atas dapat dijelaskan sebagai berikut.

- a. *True Positive (TP)*

$$TP = 40$$

- b. *False negative (FN)*

$$FN = 8$$

- c. *True Negative (TN)*

$$TN = 46$$

- d. *False Positive (FP)*

$$FP = 10$$

Selanjutnya, dengan mengacu pada data *confusion matrix* di atas maka dapat dilakukan perhitungan *accuracy*, *precision*, *recall*, dan *f1-score*.

a. Akurasi

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{True positive} + \text{True Negative}}{\text{Total data test}} \times 100\% \\
 &= \frac{40 + 46}{104} \times 100\% \\
 &= 82\%
 \end{aligned}$$

b. Sentimen Positif

1. Precision

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{FP + TP} \\
 &= \frac{40}{10 + 40} \\
 &= 0,80
 \end{aligned}$$

2. Recall

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN + TP} \\
 &= \frac{40}{8 + 40} \\
 &= 0,83
 \end{aligned}$$

3. F1-Score

$$\begin{aligned}
 \text{F1 - Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= 2 \times \frac{0,80 \times 0,83}{0,80 + 0,83} \\
 &= 0,82
 \end{aligned}$$

c. Sentimen Negatif

1. Precision

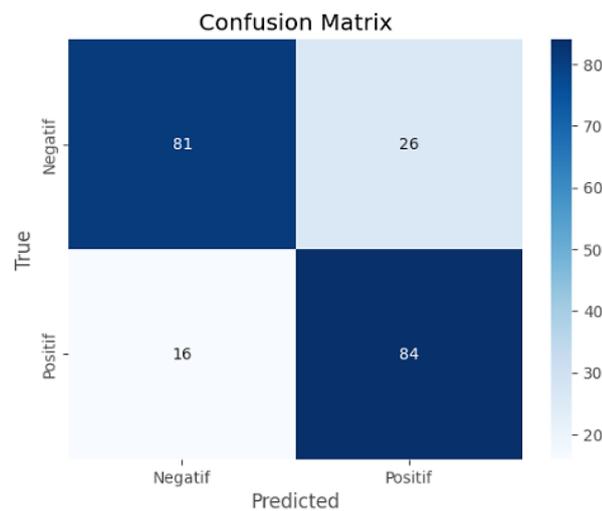
$$\begin{aligned}
 \text{Precision} &= \frac{TN}{FN + TN} \\
 &= \frac{46}{8 + 46} \\
 &= 0,85
 \end{aligned}$$

2. *Recall*

$$\begin{aligned} \text{Recall} &= \frac{TN}{FP + TN} \\ &= \frac{46}{10 + 46} \\ &= 0,82 \end{aligned}$$

3. *F1-Score*

$$\begin{aligned} F1 - \text{Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= 2 \times \frac{0,85 \times 0,82}{0,85 + 0,82} \\ &= 0,84 \end{aligned}$$



Gambar 4.7 *Confusion Matrix* Naïve Bayes rasio 80:20
(Sumber: Hasil olah data peneliti, 2023)

Berdasarkan Gambar 4.7, dataset *tweet* mengenai sentimen pengguna terhadap aplikasi iPusnas dengan algoritma *Multinomial Naïve Bayes Classifier* menggunakan ekstraksi fitur TF-IDF pada rasio 80:20. Terdapat 84 data positif yang diprediksi benar oleh model sebagai sentimen positif (*True Positive*), dan 26 data positif yang diprediksi salah sebagai sentimen negatif (*False Negative*). Selain itu, terdapat 81 data negatif yang diprediksi benar oleh model sebagai sentimen negatif (*True Negative*) dan 16 data negatif yang diprediksi salah sebagai sentimen positif (*False Positive*). Dengan demikian, klasifikasi data berdasarkan *confusion matrix* di atas dapat dijelaskan sebagai berikut.

a. *True Positive (TP)*

$$TP = 84$$

b. *False negative (FN)*

$$FN = 16$$

c. *True Negative (TN)*

$$TN = 81$$

d. *False Positive (FP)*

$$FP = 26$$

Selanjutnya, dengan mengacu pada data *confusion matrix* di atas maka dapat dilakukan perhitungan *accuracy*, *precision*, *recall*, dan *f1-score*.

a. Akurasi

$$\begin{aligned} Accuracy &= \frac{True\ positive + True\ Negative}{Total\ data\ test} \times 100\% \\ &= \frac{84 + 81}{207} \times 100\% \\ &= 79\% \end{aligned}$$

b. Sentimen Positif

1. *Precision*

$$\begin{aligned} Precision &= \frac{TP}{FP + TP} \\ &= \frac{84}{26 + 84} \\ &= 0,76 \end{aligned}$$

2. *Recall*

$$\begin{aligned} Recall &= \frac{TP}{FN + TP} \\ &= \frac{84}{16 + 84} \\ &= 0,84 \end{aligned}$$

3. *F1-Score*

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$= 2 \times \frac{0.76 \times 0.84}{0.76 + 0.84}$$

$$= 0,80$$

c. Sentimen Negatif

1. *Precision*

$$Precision = \frac{TN}{FN + TN}$$

$$= \frac{81}{16 + 81}$$

$$= 0,84$$

2. *Recall*

$$Recall = \frac{TN}{FP + TN}$$

$$= \frac{81}{26 + 81}$$

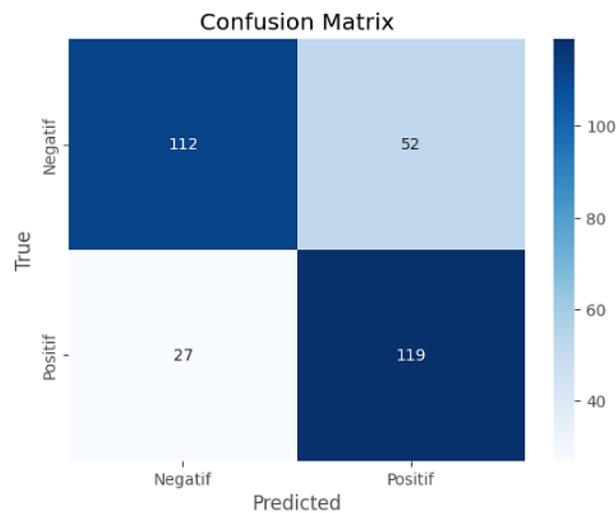
$$= 0,76$$

3. *F1-Score*

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$= 2 \times \frac{0,84 \times 0,76}{0,84 + 0,76}$$

$$= 0,79$$



Gambar 4.8 Confusion Matrix Naïve Bayes rasio 70:30
(Sumber: Hasil olah data peneliti, 2023)

Berdasarkan Gambar 4.8, dataset *tweet* mengenai sentimen pengguna terhadap aplikasi iPusnas dengan algoritma *Multinomial Naïve Bayes Classifier* menggunakan ekstraksi fitur TF-IDF pada rasio 70:30. Terdapat 119 data positif yang diprediksi benar oleh model sebagai sentimen positif (*True Positive*), dan 27 data positif yang diprediksi salah sebagai sentimen negatif (*False Negative*). Selain itu, terdapat 112 data negatif yang diprediksi benar oleh model sebagai sentimen negatif (*True Negative*) dan 52 data negatif yang diprediksi salah sebagai sentimen positif (*False Positive*). Dengan demikian, klasifikasi data berdasarkan *confusion matrix* di atas dapat dijelaskan sebagai berikut.

- a. *True Positive (TP)*

$$TP = 119$$

- b. *False negative (FN)*

$$FN = 27$$

- c. *True Negative (TN)*

$$TN = 112$$

- d. *False Positive (FP)*

$$FP = 52$$

Selanjutnya, dengan mengacu pada data *confusion matrix* di atas maka dapat dilakukan perhitungan *accuracy*, *precision*, *recall*, dan *f1-score*.

- a. Akurasi

$$\begin{aligned} Accuracy &= \frac{True\ positive + True\ Negative}{Total\ data\ test} \times 100\% \\ &= \frac{119 + 112}{310} \times 100\% \\ &= 74\% \end{aligned}$$

- b. Sentimen Positif

1. *Precision*

$$\begin{aligned} Precision &= \frac{TP}{FP + TP} \\ &= \frac{119}{52 + 119} \\ &= 0,70 \end{aligned}$$

2. *Recall*

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN + TP} \\
 &= \frac{119}{27 + 119} \\
 &= 0,82
 \end{aligned}$$

3. *F1-Score*

$$\begin{aligned}
 \text{F1-Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= 2 \times \frac{0,70 \times 0,82}{0,70 + 0,82} \\
 &= 0,75
 \end{aligned}$$

c. Sentimen Negatif

1. *Precision*

$$\begin{aligned}
 \text{Precision} &= \frac{TN}{FN + TN} \\
 &= \frac{112}{27 + 112} \\
 &= 0,81
 \end{aligned}$$

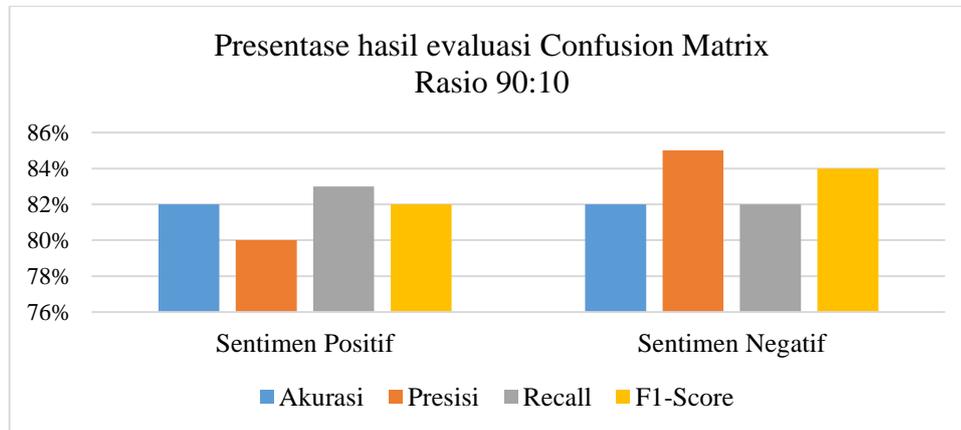
2. *Recall*

$$\begin{aligned}
 \text{Recall} &= \frac{TN}{FP + TN} \\
 &= \frac{112}{52 + 112} \\
 &= 0,68
 \end{aligned}$$

3. *F1-Score*

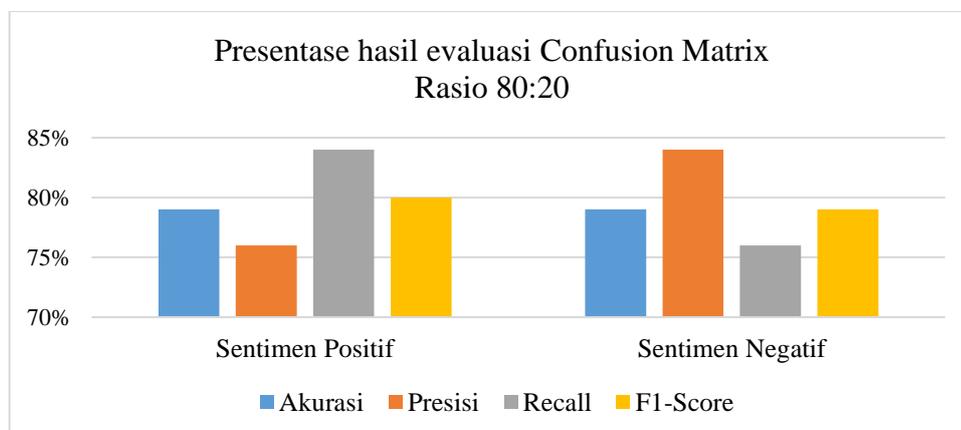
$$\begin{aligned}
 \text{F1-Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= 2 \times \frac{0,81 \times 0,68}{0,81 + 0,68} \\
 &= 0,74
 \end{aligned}$$

Berikut ini disajikan hasil *confusion matrix* dari model Naïve Bayes dengan ekstraksi fitur TF-IDF menggunakan rasio 90:10. Hasil tersebut ada pada gambar 4.9.



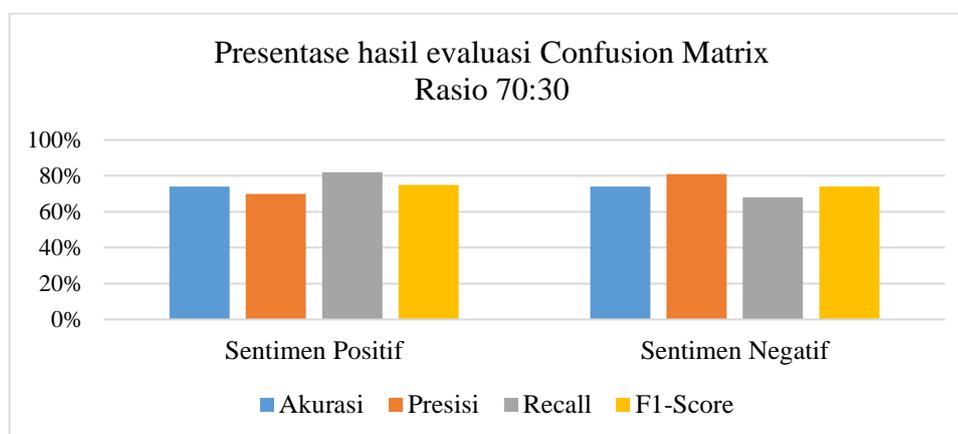
Gambar 4.9 Diagram Hasil Evaluasi *Confusion Matrix* Rasio 90:10
(Sumber : Hasil olah data peneliti, 2023)

Pada gambar 4.9 menunjukkan hasil evaluasi *confusion matrix* klasifikasi analisis sentimen yang menggunakan Naïve Bayes *Classifier* dengan ekstraksi fitur TF-IDF (NBC TF-IDF) pada rasio 90:10 . Evaluasi dilakukan dengan menggunakan *matrix* standar seperti akurasi, presisi, *recall*, dan F1-Score. *Matrix* memberikan pemahaman yang lebih mendalam tentang kemampuan model dalam mengklasifikasikan data dengan tepat. Pengujian pada rasio 90:10 ini memiliki akurasi 82%, dilihat dari kelas positif presisi sejumlah 80%, *recall* sejumlah 83%, f1-score sejumlah 82%. Sedangkan dilihat dari kelas negatif presisi sejumlah 85%, *recall* sejumlah 82%, f1-score 84%.



Gambar 4.10 Diagram Hasil Evaluasi *Confusion Matrix* Rasio 80:20
(Sumber : Hasil olah data peneliti, 2023)

Pada gambar 4.10 menunjukkan hasil evaluasi *confusion matrix* klasifikasi analisis sentimen yang menggunakan *Naïve Bayes Classifier* dengan ekstraksi fitur TF-IDF (NBC TF-IDF) pada rasio 80:20 . Evaluasi dilakukan dengan menggunakan *matrix* standar seperti akurasi, presisi, *recall*, dan F1-Score. *Matrix* memberikan pemahaman yang lebih mendalam tentang kemampuan model dalam mengklasifikasikan data dengan tepat. Pengujian pada rasio 80:20 ini memiliki akurasi 79%, dilihat dari kelas positif presisi sejumlah 76%, *recall* sejumlah 84%, f1-score sejumlah 80%. Sedangkan dilihat dari kelas negatif presisi sejumlah 84%, *recall* sejumlah 76%, f1-score 79%.



Gambar 4.11 Diagram Hasil Evaluasi *Confusion Matrix* Rasio 70:30
(Sumber : Hasil olah data peneliti, 2023)

Pada gambar 4.11 menunjukkan hasil evaluasi *confusion matrix* klasifikasi analisis sentimen yang menggunakan *Naïve Bayes Classifier* dengan ekstraksi fitur TF-IDF (NBC TF-IDF) pada rasio 70:30. Evaluasi dilakukan dengan menggunakan *matrix* standar seperti akurasi, presisi, *recall*, dan F1-Score. *Matrix* memberikan pemahaman yang lebih mendalam tentang kemampuan model dalam mengklasifikasikan data dengan tepat. Pengujian pada rasio 70:30 ini memiliki akurasi 74%, dilihat dari kelas positif presisi sejumlah 70%, *recall* sejumlah 82%, f1-score sejumlah 75%. Sedangkan dilihat dari kelas negatif presisi sejumlah 81%, *recall* sejumlah 68%, f1-score 74%.

Semua persentase yang telah disebutkan berdasarkan perhitungan jumlah kecocokan prediksi nilai *true* dan *false* dari data aktual. Nilai *true* dan *false* tersebut divisualisasikan oleh *sklearn* berbentuk tabel *matrix* yang pada kotak kiri atas,

kotak tengah, dan kotak kanan bawah adalah berisi nilai *true* karena *true* label dan *predicted* label bertemu pada garis vertikal dan horizontal sebuah tabel *matrix*, sedangkan selain pada kotak yang telah disebut adalah bagian yang berisi nilai *false*.

4.2 Pembahasan

Pada tahap awal, perihal penghapusan data yang terduplikasi menggunakan fitur penandaan dari Microsoft Excel mendapatkan 307 data yang terdeteksi duplikat, sedangkan masih banyak data yang sebenarnya juga terindikasi duplikat, namun belum terhapus sehingga dilakukan penandaan data secara manual yang mendapatkan 1343 data yang harus dihapus. Setelah itu, diperoleh data bersih dari duplikasi sejumlah 1036 data. Kemudian, data tersebut diberi label oleh ahli dari ilmu bahasa Indonesia. Label pada sentimen positif memiliki jumlah banyak daripada label negatif.

Dari hasil pelabelan tersebut, kemudian dilakukan pembersihan partikel pada teks yang tidak berguna untuk klasifikasi. Penghapusan terhadap *punctuation*, *URL*, *emoticon*, *hashtag*, *username/mention*, dan *whitespace* berhasil, serta dilakukan pengecilan huruf agar menjadi setara sehingga memudahkan proses klasifikasi. Setelah data bersih dari beberapa partikel, kemudian melakukan penghapusan dengan memakai library sastrawi yang dalamnya memiliki modul penyimpan kata yang perlu dihapus. Setelah itu, dilakukan *stemming* dengan library sastrawi yang membutuhkan waktu beberapa menit dalam prosesnya, adapun kata yang tidak terdapat dalam daftar *stemming* tetap akan digunakan pada proses klasifikasi. Pada akhir *preprocessing* ini dilakukan tokenisasi yang nanti akan digunakan pada proses pemobotan kata.

Pembobotan dilakukan untuk memberi nilai terhadap kata yang berpengaruh. Dari hasil pembobotan terdapat 829 kata yang telah diberi bobot oleh sistem. Nilai bobot mulai dari 0.0 hingga yang terbesar adalah 0,805564. Kemudian data yang telah terbobot ini digunakan untuk proses klasifikasi pelatihan data.

Proses klasifikasi ini dilakukan menggunakan metode Naïve Bayes *Classifier*. Data diproses dengan perbandingan antara data latih dan data uji rasio 90:10 dengan akurasi sebanyak 82%, rasio 80:20 dengan akurasi sebanyak 79%, dan 70:30 dengan akurasi sebanyak 74%. Untuk Hasil klasifikasi rasio 90:10 data

tweet yang mempunyai label *actual* positif dan *predicted* positif sebanyak 40 data, data *tweet* yang mempunyai label *actual* negatif dan *predicted* negatif sebanyak 46 data, data *tweet* yang tidak berhasil diklasifikasikan ke dalam label *actual* positif dengan *predicted* negatif sebanyak 8 data, dan data *tweet* yang tidak berhasil diklasifikasikan ke dalam label *actual* negatif dengan *predicted* positif sebanyak 10 data.

Pada hasil klasifikasi positif mempunyai ciri kalimat yang mengandung kata positif. Contohnya seperti menarik, asik, bisa, lucu, bagus, hoki, gratis, suka, *happy*, puas, banyak, baik, makasih, bangga, referensi, legal, untung, bebas, murah, favorit, tersedia, lengkap, mudah, nyaman, dan keren. Sedangkan untuk hasil klasifikasi negatif mempunyai ciri kalimat yang mengandung kata negatif seperti antre, musuh, lama, *error*, kosong, gabisa, lupa, nyebelin, masalah, kesal, bete, lambat, payah, capek, susah, rebut, saingan, jelek, *bug*, males, salah, sedih, habis, hilang, dan gila. Namun ada beberapa data yang tidak sesuai label dengan hasil klasifikasinya. Hal ini dikarenakan penelitian hanya menerapkan algoritma Naïve Bayes tanpa adanya optimasi maupun normalisasi data pada proses *preprocessing*, sehingga data yang digunakan masih banyak yang belum bersih sepenuhnya. Terutama pada tahap *tokenizing* dan *stemming*, pada saat percobaan ditemukan banyak kata yang seharusnya tereliminasi dan dikembalikan menjadi kata dasar, tetapi hal tersebut tidak bekerja dengan baik. Beberapa kata tersebut, seperti salah ketik, kata gaul, bukan bahasa Indonesia, kata berlebih, singkatan, dan kata lain yang tidak termuat 77 dalam kamus yang dimiliki *library* Sastrawi.

Hasil dari pengujian dan evaluasi menunjukkan rasio 90:10 memiliki presentase yang cukup baik dengan akurasi 82% dari keseluruhan 104 data yang diuji, dengan cara sistem menghitung total jumlah prediksi yang sesuai dengan data aktual. Hasil presisi yang diambil oleh sistem memperhatikan pada masing-masing kelas. Dari masing-masing kelas, sistem akan melakukan perhitungan nilai yang terprediksi *true* pada kelasnya dibagi dengan jumlah total dari nilai terprediksi *true* pada kelas itu sendiri dengan nilai yang terprediksi *false* pada kelas tersebut. Diperoleh nilai kelas positif sejumlah 80%, dan kelas negatif sejumlah 85%. Adapun *recall* dari masing-masing kelas, sistem akan melakukan perhitungan nilai

yang terprediksi *true* pada kelasnya dibagi dengan jumlah total dari nilai terprediksi *true* pada kelas itu sendiri dengan jumlah nilai kelas tersebut yang terprediksi menjadi kelas lain, atau dapat disebut *false prediction*. Persentase *recall* ini, didapat dari perhitungan nilai rata-rata dari keseluruhan kelas yang memiliki persentase temu kembali masing-masing yaitu kelas positif ditemukan sejumlah 83% dan kelas negatif sejumlah 82%. Selain mengevaluasi performa model berdasarkan akurasi keseluruhan nilai *true* dan *false*, pada penelitian ini juga menggunakan alternatif evaluasi berdasarkan nilai akurasi pada masing-masing kelas yaitu *f1-score*, nilai ini menggabungkan antara *recall* dengan presisi. Evaluasi *f1-score* pada rasio 90:10 ini mendapat nilai kelas positif sejumlah 82% dan kelas negatif sejumlah 84%.

Pada rasio 90:10, hasil dari 104 data yang diuji pada terdapat 86 nilai *true* sebagai representasi sentimen positif dan negatif. Pengujian rasio 90:10 memiliki akurasi 82%, dilihat dari kelas positif presisi sejumlah 80%, *recall* sejumlah 83%, *f1-score* sejumlah 82%. Sedangkan dilihat dari kelas negatif presisi sejumlah 85%, *recall* sejumlah 82%, *f1-score* 84%.

Analisis sentimen memiliki tujuan untuk mengetahui bagaimana ekspresi orang dalam bentuk opini. Opini bukan hanya dalam bentuk verbal, melainkan dapat berbentuk nonverbal juga. Salah satu bentuk opini nonverbal tersebut adalah teks atau tulisan. Opini pada penelitian ini diperoleh dari X dengan memilih topik aplikasi iPusnas. Tujuan dari analisis sentimen tersebut supaya dapat memilah dan membedakan mana opini yang baik dan yang buruk, sehingga hal tersebut dapat dijadikan evaluasi untuk masa depan yang lebih baik dengan harapan menjadikan opini baik untuk dioptimalkan dan menjadikan opini buruk untuk perbaikan aplikasi iPusnas kedepannya. Adapun dalam pandangan islam, tujuan dari analisis sentimen ini seperti yang ada pada al-Qur'an Surat Al-Isra' Ayat 36.

وَلَا تَقْفُ مَا لَيْسَ لَكَ بِهِ عِلْمٌ إِنَّ السَّمْعَ وَالْبَصَرَ وَالْفُؤَادَ كُلُّ أُولَٰئِكَ كَانَ عَنْهُ مَسْئُولًا

“Janganlah engkau mengikuti sesuatu yang tidak kau ketahui. Sesungguhnya pendengaran, penglihatan, dan hati nurani, semua itu akan diminta pertanggungjawabannya.” (Q.S Isra' ayat 36)

Jangan pernah berbicara tentang sesuatu yang tidak kamu ketahui, mengaku menyaksikan apa yang sebenarnya tidak kamu lihat, atau mengatakan bahwa kamu

mendengar sesuatu yang sebenarnya tidak kamu dengar, maupun mengklaim pengalaman yang tidak pernah kamu alami. Sesungguhnya, pendengaran, penglihatan, dan hati adalah amanah dari Tuhan, dan semuanya akan dimintai pertanggungjawaban (*Qur'an Kemenag*, 2024). Ayat di atas mencerminkan prinsip-prinsip evaluasi dalam penelitian, seperti verifikasi informasi, menggunakan akal, mengkaji fakta, dan memastikan kebenaran berdasarkan bukti. Dalam konteks penelitian, hal ini relevan dengan proses evaluasi untuk memastikan hasil penelitian akurat, valid, dan dapat dipertanggungjawabkan. Jika beropini atau berkata maka ucapkanlah perkataan yang benar dan tepat sasaran.

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan pengujian dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa hasil tertinggi antara tiga skenario pengujian dan evaluasi performa sistem dalam memperoleh kinerja terbaik ada pada rasio 90:10 dengan akurasi sebesar 82%, sedangkan untuk rasio 80:20 memperoleh akurasi sejumlah 79%, dan rasio 70:30 mempunyai akurasi sebesar 74%. Karena rasio 90:10 memiliki data latih lebih banyak dari rasio lainnya sehingga data yang akan diuji hasilnya lebih akurat. Namun ada beberapa data yang tidak sesuai label dengan hasil klasifikasinya.

Pengujian rasio 90:10 memiliki akurasi 82%, dilihat dari kelas positif presisi sejumlah 80%, *recall* sejumlah 83%, f1-score sejumlah 82%. Sedangkan dilihat dari kelas negatif presisi sejumlah 85%, *recall* sejumlah 82%, f1-score 84%.

5.2 Saran

Penelitian ini menghasilkan saran yang bisa dilakukan oleh instansi terkait dan penelitian selanjutnya. Beberapa saran tersebut antara lain:

1. Pada opini negatif iPusnas sering mengalami *error*, *bug*, serta antrean yang lama ketika meminjam buku. Oleh karena itu Perpustakaan Nasional harus mengembangkan aplikasi iPusnas.
2. Media sosial sangat berbeda dengan platform berita yang memperhatikan kata baku. Oleh karena itu, perlu melakukan normalisasi kata yang salah ketik, kata gaul, bukan bahasa Indonesia, singkatan, kata tidak baku, serta beberapa kata yang tidak berubah bentuk menjadi kata dasar.
3. Penelitian mendatang di sarankan menggunakan platform media sosial lainnya seperti Facebook, dan Instagram agar lebih komprehensif.

DAFTAR PUSTAKA

- Amaniputra, M. E., Soesanto, R. P., & Anggana, H. D. (2023). Perancangan Dashboard Sentimen Pariwisata Menggunakan Metode *Complement* Naïve Bayes pada Kabupaten Rembang. *Jurnal e-Proceeding of Engineering*, 10(3).
- Apriani, R., & Gustian, D. (2019). Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*, 6(1).
- Budiman, A. E., & Widjaja, A. (2020). Analisis Pengaruh Teks *Preprocessing* Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(3).
- Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naïve Bayes *Classifier*. *Jurnal Informatika dan Teknik Elektro Terapan*, 10(1).
- Fauzan, Abd. C., & Hikmah, K. (2022). Implementasi Algoritma Naïve Bayes Dalam Analisis Polarisasi Opini Masyarakat Terkait Vaksin COVID-19. *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, 7(2), 122–128.
- Fauzan, F., & Suwanto, S. A. (2018). Analisis Pemanfaatan Aplikasi Ipusnas Berbasis Android Di Perpustakaan Nasional Republik Indonesia. *Jurnal Ilmu Perpustakaan*, 7(4), Article 4.
- Firdaus, A., & Firdaus, W. I. (2021). *Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi: (Sebuah Ulasan)*. *Jurnal JUPITER*, 13(1).
- Firmansyah, M. A., Mulyana, D., Karlinah, S., & Sumartias, S. (2018). Kontestasi Pesan Politik dalam Kampanye Pilpres 2014 di Twitter: Dari Kultwit Hingga Twitwar. *Jurnal Ilmu Komunikasi*, 16(1), 42.
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naïve Bayes. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 4(2), 113.
- Hendra Suputra, I. P. G., Prebiana, K. D., & Gorianto, F. O. (2021). Perbandingan Jenis TF terhadap Hasil Evaluasi Information Retrieval. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 8(2), 207.
- Hermawan, L., & Bellanar Ismiati, M. (2020). Pembelajaran Text *Preprocessing* berbasis Simulator Untuk Mata Kuliah Information Retrieval. *Jurnal Transformatika*, 17(2), 188.

- Iba, Z., & Wardhana, A. (2023). Landasan Teori Kerangka Pemikiran Penelitian Terdahulu Hipotesis (Pertama). EUREKA MEDIA AKSARA.
- Jagiyanto, H. M. (2005). Sistem Teknologi Informasi. Andi Offset.
- Mahfud, F. K. R., Mudawamah, N. S., & Hariyanto, W. (2020). Sentiment Analysis of Perpustakaan Nasional Republik Indonesia Through Social Media Twitter. *MATICS*, 12(1), 90.
- Maulana, Y. I. (2018). Evaluasi Tingkat Kepuasan Pengguna Perpustakaan Digital Nasional (iPusnas) dengan Kerangka *PIECES*. 6(1).
- Mustaqililah, R., Widyaningtyas, O., & Wantoro, T. (2023). Efektivitas Penggunaan Twitter Sebagai Sarana Peningkatan Berpikir Kritis Mahasiswa Ilmu Komunikasi. *Jurnal Ilmu Komunikasi*, 2(1).
- Pintoko, B. M. (2018a). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes *Classifier*. *Jurnal e-Proceeding of Engineering*, 5(3).
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(1).
- Pujianto, U., Widiyaningtyas, T., Prasetya, D. D., & Romadhon, B. (2019). Penerapan algoritma naïve bayes *classifier* untuk klasifikasi judul skripsi dan tugas akhir berdasarkan Kelompok Bidang Keahlian. *TEKNO*, 27(1).
- Qur'an Kemenag*. (2023, November 2). <https://quran.kemenag.go.id/>
- Rahman, D. F. (2023). Analisis Chatgpt Tweet Menggunakan Eda Dan Sentiment Analysis: Data Pengguna Twitter Di Indonesia. *Prosiding Seminar Nasional ELINVO dengan tema: (Penerapan Teknologi Chat GPT dan AI pada Bidang Akademik di Perguruan Tinggi)*, 26 September 2023, hal: 60-90 ISSN: 2477-2402.
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50.
- Rozi, I. F., Pramono, Sholeh Hadi, & Dahlan, Erfan Achmad. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS* 6(1).
- Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., & Watrionthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi

COVID-19 Menggunakan Metode *Naïve Bayes*. *Jurnal Media Informatika Budidarma*, 5(1), 157.

Sejati, V. R. A., & Pramusinto, W. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Ipusnas Pada Google Play Store Dengan *Multinomial Naïve Bayes*. *SENAFTI (Seminar Nasional Mahasiswa Fakultas Teknologi Informasi)* 2(1).

Septiani, A., & Budi, I. (2022). Klasifikasi Ulasan Pengguna Aplikasi: Studi Kasus Aplikasi Ipusnas Perpustakaan Nasional Republik Indonesia (PNRI). *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(4), 1110–1120.

Septianingrum, F., & Irawan, A. S. Y. (2021). Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma *Naïve Bayes*: Sebuah Literature Review. *Jurnal Media Informatika Budidarma*, 5(3), 799.

Siswandi, A., & Fitriana, M. (2019). Implementasi Data Mining Dengan Metode Klasifikasi *Naïve Bayes* Untuk Memprediksi Stok Bahan Jadi. *10*.

Toy, K. V. S., Sari, Y. A., & Cholissodin, I. (2021). Analisis Sentimen Twitter menggunakan Metode *Naïve* dengan *Relevance Frequency Feature Selection* (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal).

Wang, Y., Jia, Y., Tian, Y., & Xiao, J. (2022). *Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring*. *Expert Systems with Applications*, 200, 117013.

Qur'an Kemenag. (2024, November 9). <https://quran.kemenag.go.id/>

LAMPIRAN

Lampiran 1.

Daftar 50 sampel dataset berlabel.

No	Tweets	Label
1	Sekian lama nungguin slot ipusnas akhirnya dapet juga	Positif
2	sama kayak heroes of olympus, buku kedua ga ada di ipusnas :) MANA AKU UDAH KEBELI TRIALS OF APOLLO, JADI KETUNDA SAMPE GTW KAPAN INI BACANYA	Negatif
3	/lt hi warga litbase, minta rekomendasi mystery thriller di ipusnas selain Kesetiaan Mr X, Second Sister, dan karya Agatha Christie donggg, makasiii	Positif
4	Aku juga ngerasa gitu, tapi aku gamasalah dengan pengulangan topiknya sih Cuma, sayang bgt aja gitu, topiknya dah diulang tapi bahasan cuma kumpulan kata kata cantik aja	Negatif
5	Bukan web tapi aplikasi dari perpustakaan (iPusnas) Sistemnya pinjem jadi baca online aja, ada tenggat waktu pengembalian tapi bisa pinjem berulang kali ko	Positif
6	Huhu thank u yaa	Negatif
7	Bener jugaa :) Mana di ipusnas cepet banget'	Positif
8	Di sini juga ada namanya iPusnas, beberapa kali minjam di aplikasinya	Positif
9	Wkwkwkw sama, ini kebetulan aku lagi baca bukunya Maudy Ayunda yg Dear Tomorrow' di iPusnas'	Positif
10	Baca buku ilegal kalau harus ngantri di ipusnas	Positif
11	Kalau masih belum jelas ttg peran dan posisi perempuan dalam agama (Islam), baca buku ini Ga ada duit buat beli? Pinjam gratis e-book di aplikasi iPusnas Hati-hati, yg kamu ucapkan itu pandangan agama (Islam) ataupun pandangan patriarki?	Positif
12	ipusnas ap salahku padamu	Negatif
13	nggak punya bacaan euy, nggak ada yang dipinjem dari perpustakaan dan pitimoss kemaren wl di ipusnas pun masih pada ngantreeee	Negatif

No	Tweets	Label
14	*nongkrong di ipusnas 12 jam* *kerja 6 jam* *baca novel/buku 6 jam* Yang, bu puan yang' 'Yang, bu mega yang' 'Yang, ayam geprek yuk' 'Yang, mas sopifutnya nyasar' *drained* Udah begitu pilihannya masih aja Jaemin	Negatif
15	mau baca kafka tapi seems like takut engga keburu soalnya bacanya di ipusnas	Negatif
16	Aku lagi baca Ibu Susu di ipusnas U should read this nder	Positif
17	kemarin pas mau baca ipusnas gue eror dan bukunya kereturn mana baru baca 20an persen yampunn sedih	Negatif
18	Gadak screenshotnya Jen Aku bacanya di Ipusnas	Positif
19	Sejauh ini yg udah aku baca di ipusnas ada buku-bukunya: agatha christie, ilana Tan, nh dini, sama baru aja nyoba baca English classic (Alice aitw)	Positif
20	Kamu download aplikasi IPusnas, kemudian daftar menggunakan email, selesai Kamu tinggal pinjam buku yang sesuai dengan keinginanmu	Positif
21	jakarta sebelum pagi, buku langka coyakwoakwoa tp ini ngga worth kalo lu beli dengan harga 1 jt (serius) mending ngantri di ipusnas atau pinjem punya org ini buku aneh sih, maksud gua aneh nya tuh kaya pemilihan bahasa nya itu aneh tapi kocak	Positif
22	little price juga ada tuh di ipusnas	Positif
23	Wkwkwk aku kalau di ipusnas tu kadang juga yang cuma scroll2 aja mbak, kalau ada yg menarik langsung pinjem kalau ada stoknya	Positif
24	Sip! Soalnya kadang bingung kalo mau baca di ipusnas bukunya itu2 aja Terus kl di gramdig nggak semua wishlistku terbitan gamed	Negatif
25	Setelah antri buku ini di Ipusnas selama berbulan2 tapi ga pernah dapet, akhirnya minjam buku ini di pitimoss Buku ini menceritakan tentang Miss Oishi, seorang guru muda yang mengajar di desa nelayan yang miskin, dan 12 orang anak muridnya	Negatif
26	Indonesia juga ada perpustakaan aplikasinya ipusnas Gratis Lumayan lengkap karena buku2 yang terbit diminta menyumbangkan bukunya ke perpustakaan	Positif
27	Ini di ipusnas baca ngantri ya? Aku kmren coba tapi gak bisa bisa Kesel banget	Negatif
28	pinjam buku di ipusnas antri bgt buset	Negatif

No	Tweets	Label
29	pinjaman buku di iPusnas n RBK cuma kebaca beberapa bab dan sekarang udah otomatis dikembaliin wkwk dahlahh sekarang waktunya baca buku fisikk -karena udah balik ke rutinitas- ' as LIBURANNYA KERASA CEPETTT	Negatif
30	Iya kak, buku itu sebenarnya sudah masuk wishlist Tapi ya, karena di iPusnas itu saking banyaknya pilihan dan BANYAK banget yg pengen kubaca, jadinya tiap kelar baca satu buku galau mau lanjut baca buku yg mana	Positif
31	Kenapa banyak banget novel menarik di iPusnas huhuu tbr ku makin menumpuk Semoga masuk kuliah diundur jadi April	Positif
32	Uprooted ada di Ipusnas ternyataaaaaa yeay(Positif
33	Silahkan kalau mau ikutan baca, mumpung stok buku yang ada di iPusnas ini masih ada 5 Here the title	Positif
34	Baru tahu ipusnas ada maksimal buku yg dipinjem HALAH WL KU PADAHAL UDAH AVAIL	Negatif
35	Aku pake ini nder, Samsung A7 Lite karena ukurannya kecil jadi bisa dipegang pakai 1 tangan Kerasa pas banget buat baca buku di Gramedia Digital atau Ipusnas Masih nyaman banget juga kalau mau anotasi di PDF lain gitu Best purcha	Positif
36	Rooftop Buddies Aku mulai baca setelah baca rekomendasi buku yang bisa dibaca di iPusnas, tapi akhirnya kukembaliin sebelum masa pinjam abis The scene when the MC meets her bullies before she commits her sulc1d3 and ended up crying, it triggered my trauma	Positif
37	Nonton pilmmm, ikut antrian buku di ipusnas itu seru nderrr, saran ku dr aku	Negatif
38	Baca dulu aja bukunya Tipis, isinya to the point, mudah dipahami Gratis pinjam di aplikasi iPusnas Ntar kalau udah, baru ngetwit lagi	Positif
39	[2] Q: Baca Peter Pan edisi bahasa Indonesia di mana, ya? A: Peter Pan tersedia gratis di iPusnas, ya, MOA-deul Jika kamu sedang berlangganan Gramedia Digital, Peter Pan juga bisa diakses dengan paket fiksi!	Positif
40	Seru bgt parahhh, btw saling follow acc Ipusnas dund, plz dm mee:((Positif
41	ipusnas bukunya banyaa yg bagus, tapi antriannya itu lho :(Positif

No	Tweets	Label
42	App iPusnas ancur banget sih Ã°Ã,Ã°Ã£Ã°Ã,Ã°Ã£Ã°Ã,Ã°Ã£	Negatif
43	dipikir-pikir waktu itu bisa baca jakarta sebelum pagi di ipusnas tuh keajaiban gaksiÃ°Ã,ËœÃ©bangga deh ak	Positif
44	Aaa punya bukunya, aku dulu baca d @ipusnas_id	Positif
45	Hoki banget dapet second sister di ipusnas Ã¢Ã“â€¹Ã¬Ã,Ã• Ã°Ã,ËœÃ	Positif
46	Haaaa iseng nyari wishlist malah dapat di iPusnas	Negatif
47	/lt yg suka buku klasik mungkin bisa coba baca buku ini lg byk stock di ipusnas Ã¢ËœÃ°Ã¬Ã,Ã•	Positif
48	kebiasaan ni ipusnas kalau tengah malem gini loadnya lamaaa	Negatif
49	knp ipusnas aku gabisa dibukaaa	Negatif
50	Tapi tadi aku iseng buka ipusnas ternyata beberapa judulnya emang ada, but thanks infonya soalnya tampilannya familiar hehehehe	Positif

Lampiran 2.

10 Sampel Hasil Pre-Processing

No	Text	Label	Cleansing	Filtering	Tokenizing	Stemming
1	Sekian lama nungguin slot ipusnas akhirnya dapat juga Ã°Ã,â€¹Ã•	Positif	sekian lama nungguin slot ipusnas akhirnya dapat juga	sekian lama nungguin slot ipusnas akhirnya dapat juga	['sekian', 'lama', 'nungguin', 'slot', 'ipusnas', 'akhirnya', 'dapat', 'juga']	sekian lama nungguin slot ipusnas akhir dapat juga
2	sama kayak heroes of olympus, buku kedua ga ada di ipusnas :) MANA AKU UDAH KEBELI TRIALS OF APOLLO, JADI KETUNDA SAMPE GTW KAPAN INI BACANYA	Negatif	sama kayak heroes of olympus buku kedua ga ada di ipusnas mana aku udah kebeli trials of apollo jadi ketunda sampe gtw	sama seperti heroes of olympus buku kedua nggak ada di ipusnas mana aku udah kebeli trials of apollo jadi ketunda	['sama', 'seperti', 'heroes', 'of', 'olympus', 'buku', 'kedua', 'nggak', 'ada', 'di', 'ipusnas', 'mana', 'aku', 'udah', 'kebeli', 'trials', 'of',	sama seperti heroes of olympus buku dua nggak ada di ipusnas mana aku udah kebeli trials of apollo jadi tunda sampe

No	Text	Label	Cleansing	Filtering	Tokenizing	Stemming
			kapan ini bacanya	sampe gtw kapan ini bacanya	'apollo', 'jadi', 'ketunda', 'sampe', 'gtw', 'kapan', 'ini', 'bacanya']	gtw kapan ini baca
3	/It hi warga litbase, minta rekomendasi mystery thriller di ipusnas selain Kesetiaan Mr X, Second Sister, dan karya Agatha Christie donggg, makasiii Å°Å,ËœÅ°	Positif	It hi warga litbase minta rekomen da si mystery thriller di ipusnas selain kesetiaan mr x second sister dan karya agatha christie donggg makasiii	It hi warga litbase minta rekomen da si mystery thriller di ipusnas selain kesetiaan mr x second sister dan karya agatha christie donggg makasiii	['lt', 'hi', 'warga', 'litbase', 'minta', 'rekomendasi', 'mystery', 'thriller', 'di', 'ipusnas', 'selain', 'kesetiaan', 'mr', 'x', 'second', 'sister', 'dan', 'karya', 'agatha', 'christie', 'donggg', 'makasiii']	It hi warga litbase minta rekomendasi mystery thriller di ipusnas selain setia mr x second sister dan karya agatha christie donggg makasiii
4	Aku juga ngerasa gitu, tapi aku gamasalah dengan pengulangan topiknya sih Cuma, sayang bgt aja gitu, topiknya dah diulang tapi bahasan cuma kumpulan kata kata cantik aja Å°Å,ËœÅ° Aku nungguin ini di Ipusnas dah lama bgt loh soalny Berasa gimana g	Negatif	aku juga ngerasa gitu tapi aku gamasalah dengan pengulanga n topiknya sih cuma sayang bgt aja gitu topiknya dah diulang tapi bahasan cuma kumpulan kata kata cantik aja aku nungguin ini di ipusnas dah lama bgt loh soalny	aku juga ngerasa gitu tapi aku gamasalah dengan pengulanga n topiknya sih cuma sayang banget aja gitu topiknya dah diulang tapi bahasan cuma kumpulan kata kata cantik aja aku nungguin ini di ipusnas dah lama banget loh soalny	['aku', 'juga', 'ngerasa', 'gitu', 'tapi', 'aku', 'gamasalah', 'dengan', 'pengulangan', 'topiknya', 'sih', 'cuma', 'sayang', 'banget', 'aja', 'gitu', 'topiknya', 'dah', 'diulang', 'tapi', 'bahasan', 'cuma', 'kumpulan', 'kata', 'kata', 'cantik', 'aja', 'aku', 'nungguin', 'ini', 'di', 'ipusnas',	aku juga ngerasa gitu tapi aku gamasalah dengan ulang topik sih cuma sayang banget aja gitu topik dah ulang tapi bahas cuma kumpul kata kata cantik aja aku nungguin ini di ipusnas dah lama banget loh soalny asa gimana g

No	Text	Label	Cleansing	Filtering	Tokenizing	Stemming
			berasa gimana g	berasa gimana g	'dah', 'lama', 'banget', 'loh', 'soalny', 'berasa', 'gimana', 'g']	
5	Bukan web tapi aplikasi dari perpustakaan (iPusnas) Sistemnya pinjam jadi baca online aja, ada tenggat waktu pengembalian tapi bisa pinjam berulang kali ko	Positif	bukan web tapi aplikasi dari perpustakaan ipusnas sistemnya pinjam jadi baca online aja ada tenggat waktu pengembalian tapi bisa pinjam berulang kali ko	bukan web tapi aplikasi dari perpustakaan ipusnas sistemnya pinjam jadi baca online aja ada tenggat waktu pengembalian tapi bisa pinjam berulang kali ko	['bukan', 'web', 'tapi', 'aplikasi', 'dari', 'perpustakaan', 'ipusnas', 'sistemnya', 'pinjam jadi baca online', 'jadi', 'baca', 'online', 'aja', 'ada', 'tenggat', 'waktu', 'pengembalian', 'tapi bisa pinjam berulang', 'kali', 'ko']	bukan web tapi aplikasi dari perpustakaan ipusnas sistem pinjam jadi baca online aja ada tenggat waktu kembali tapi bisa pinjam ulang kali ko
6	Huhu thank u yaa Gimana bukunya seru ga? Aku mau baca itu ga sebagian terus di ipusnas	Negatif	huhu thank u yaa gimana bukunya seru ga aku mau baca itu ga sebagian terus di ipusnas	huhu thank u yaa gimana bukunya seru nggak aku mau baca itu nggak sebagian terus di ipusnas	['huhu', 'thank', 'u', 'iya', 'gimana', 'bukunya', 'seru', 'nggak', 'aku', 'mau', 'baca', 'itu', 'nggak', 'kebagian', 'terus', 'di', 'ipusnas']	huhu thank u yaa gimana buku seru nggak aku mau baca itu nggak bagi terus di ipusnas
7	Bener jugaa :) Mana di ipusnas cepet banget'	Positif	bener jugaa mana di ipusnas cepet banget	benar jugaa mana di ipusnas cepet banget	['benar', 'jugaa', 'mana', 'di', 'ipusnas', 'cepat', 'banget']	benar jugaa mana di ipusnas cepet banget
8	Di sini juga ada namanya iPusnas, beberapa kali minjam di aplikasinya	Positif	di sini juga ada namanya ipusnas beberapa kali minjam	di sini juga ada namanya ipusnas beberapa kali pinjam	['di', 'sini', 'juga', 'ada', 'namanya', 'ipusnas', 'beberapa', 'kali', 'pinjam']	di sini juga ada nama ipusnas beberapa kali pinjam di aplikasi

No	Text	Label	Cleansing	Filtering	Tokenizing	Stemming
			di aplikasinya	di aplikasinya	'pinjam', 'di', 'aplikasinya']	
9	Wkwkwkw sama, ini kebetulan aku lagi baca bukunya Maudy Ayunda yg Dear Tomorrow' di iPusnas'	Positif	wkwkwkw sama ini kebetulan aku lagi baca bukunya maudy ayunda yg dear tomorrow di ipusnas	wkwkwkw sama ini kebetulan aku lagi baca bukunya maudy ayunda yang dear tomorrow di ipusnas	['wkwkwkw', 'sama', 'ini', 'kebetulan', 'aku', 'lagi', 'baca', 'bukunya', 'maudy', 'ayunda', 'yang', 'dear', 'tomorrow', 'di', 'ipusnas']	wkwkwkw sama ini betul aku lagi baca buku maudy ayunda yang dear tomorrow di ipusnas
10	Baca buku ilegal kalau harus ngantri di ipusnas	Positif	baca buku ilegal kalau harus ngantri di ipusnas	baca buku ilegal kalau harus antre di ipusnas	['baca', 'buku', 'ilegal', 'kalau', 'harus', 'antre', 'di', 'ipusnas']	baca buku ilegal kalau harus antre di ipusnas

Lampiran 4.

Hasil evaluasi *Confusion Matrix* model *machine learning*.

Rasio 90:10

	precision	recall	f1-score	support
-1	0.85	0.82	0.84	56
1	0.80	0.83	0.82	48
accuracy			0.83	104
macro avg	0.83	0.83	0.83	104
weighted avg	0.83	0.83	0.83	104

Rasio 80:20

	precision	recall	f1-score	support
-1	0.84	0.76	0.79	107
1	0.76	0.84	0.80	100
accuracy			0.80	207
macro avg	0.80	0.80	0.80	207
weighted avg	0.80	0.80	0.80	207

Rasio 70:30

	precision	recall	f1-score	support
-1	0.81	0.68	0.74	164
1	0.70	0.82	0.75	146
accuracy			0.75	310
macro avg	0.75	0.75	0.75	310
weighted avg	0.75	0.75	0.74	310

Lampiran 5

Pseudocode Penelitian

1. ****Import Library****

- Import pustaka untuk pengolahan data, preprocessing teks, visualisasi, evaluasi model, dan algoritma klasifikasi.

2. ****Load Dataset****

- Baca file Excel yang berisi data sentimen dan tampilkan beberapa baris awal.
- Hitung distribusi label sentimen.
- Periksa apakah ada nilai kosong (NaN) dalam dataset.

3. ****Preprocessing Tanggal****

- Ubah kolom 'Created-At' menjadi tipe datetime dan hilangkan informasi timezone jika ada.

4. ****Cleansing Data****

- Definisikan fungsi `cleansing` untuk:
 - Menghapus RT, emoji, URL, angka, tanda baca, dan karakter escape.
 - Mengubah teks menjadi huruf kecil.
- Terapkan fungsi cleansing pada kolom 'Text' untuk membuat kolom baru 'Cleansing'.

5. ****Normalisasi Teks****

- Definisikan fungsi `slang_normalization` untuk mengganti kata slang dengan kata standar.
- Terapkan fungsi normalisasi pada kolom 'Cleansing' untuk membuat kolom baru 'Normalisasi'.

6. ****Tokenisasi****

- Definisikan fungsi `tokenization` untuk memecah teks menjadi kata-kata.
- Terapkan tokenisasi pada kolom 'Normalisasi' untuk membuat kolom baru 'Tokenize'.

7. ****Stemming****

- Definisikan fungsi `stemming` untuk mengubah kata-kata menjadi bentuk dasarnya.
- Terapkan stemming pada kolom 'Tokenize' untuk membuat kolom baru 'Stemming'.

8. ****Menghapus Duplikasi****

- Hitung jumlah data sebelum dan sesudah menghapus duplikasi berdasarkan kolom 'Stemming'.

9. ****Split Dataset****

- Pisahkan dataset menjadi data latih (train) dan data uji (test).
- Simpan data latih dan uji ke dalam file Excel.

10. ****Ekstraksi Fitur****

- Gunakan `CountVectorizer` dan `TfidfVectorizer` untuk mengubah teks menjadi representasi numerik (Count dan TF-IDF).
- Simpan hasil ekstraksi fitur ke dalam file Excel.

11. ****Mapping Label****

- Ubah label teks ('Negatif', 'Netral', 'Positif') menjadi nilai numerik (-1, 0, 1).

12. ****Pelatihan Model****

- Gunakan Naive Bayes (MultinomialNB) untuk melatih model dengan data latih.
- Simpan hasil prediksi dari data uji.

13. ****Konversi Prediksi****

- Ubah prediksi numerik kembali ke label asli.
- Buat DataFrame hasil prediksi dengan kolom 'Comment', 'Label', dan 'Classification'.
- Simpan hasil prediksi ke dalam file Excel.

14. Visualisasi Distribusi Hasil

- Hitung persentase masing-masing kelas hasil prediksi.
- Buat diagram pie untuk memvisualisasikan distribusi hasil prediksi.

15. Evaluasi Model

- Hitung akurasi, precision, recall, dan F1-score menggunakan metrik evaluasi.
- Tampilkan classification report dan confusion matrix.
- Buat heatmap untuk visualisasi confusion matrix.

16. Selesai

- Tampilkan hasil evaluasi model dan visualisasi.

Lampiran 6

Hasil Cek Plagiasi Turnitin

Analisis Sentimen Pengguna X Terhadap Aplikasi iPusnas Menggunakan Metode Naïve Bayes – Dwi Sekar Nagari 19680061

ORIGINALITY REPORT

23%	23%	10%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	etheses.uin-malang.ac.id Internet Source	12%
2	senafti.budiluhur.ac.id Internet Source	2%
3	Submitted to UIN Syarif Hidayatullah Jakarta Student Paper	1%
4	openjournal.unpam.ac.id Internet Source	1%
5	dspace.uii.ac.id Internet Source	1%
6	j-ptiik.ub.ac.id Internet Source	1%
7	Submitted to Universitas Brawijaya Student Paper	1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On