

**ANALISIS SENTIMEN PENGGUNA *iBI LIBRARY* PADA  
ULASAN *GOOGLE PLAY STORE* DENGAN *MULTINOMIAL  
NAÏVE BAYES***

**SKRIPSI**



**Oleh :**

**ROHIMULLAH AT-THOBRONI**

**NIM. 210607110039**

**PROGRAM STUDI PERPUSTAKAAN DAN SAINS INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG  
2025**

**ANALISIS SENTIMEN PENGGUNA *iBI LIBRARY* PADA  
ULASAN *GOOGLE PLAY STORE* DENGAN *MULTINOMIAL  
NAÏVE BAYES***

**SKRIPSI**

**Oleh:  
ROHIMULLAH AT-THOBRONI  
NIM. 210607110039**

**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  
2025**

**LEMBAR PERSETUJUAN**

**ANALISIS SENTIMEN PENGGUNA *iBI LIBRARY* PADA  
ULASAN *GOOGLE PLAY STORE* DENGAN *MULTINOMIAL  
NAÏVE BAYES***

**SKRIPSI**

Oleh:  
**ROHIMULLAH AT-THOBRONI**  
NIM. 210607110039

Telah Diperiksa dan Disetujui  
Tanggal 24 Juni 20225

Pembimbing 1



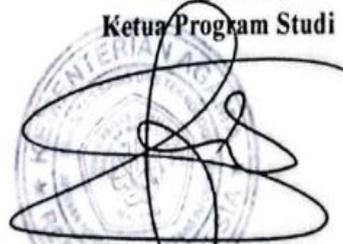
**Fakhri Khamsnu Reza Mahfud, M.Kom.**  
NIP.199005062019031007

Pembimbing 2



**Yulianto, M. Pd.I**  
NIP. 198707122019031005

Mengetahui,  
Ketua Program Studi



**Dr. Ir. Mokhammad Amin Hariyadi, M.T**  
NIP. 196701182005011001

**LEMBAR PENGESAHAN**

**ANALISIS SENTIMEN PENGGUNA *IBI LIBRARY* PADA  
ULASAN *GOOGLE PLAY STORE* DENGAN *MULTINOMIAL  
NAÏVE BAYES***

**SKRIPSI**

Oleh:

**ROHIMULLAH AT-THOBRONI**  
NIM. 210607110039

Telah dipertahankan di depan Dewan Penguji Skripsi  
dan dinyatakan diterima sebagai salah satu persyaratan untuk memenuhi  
Gelar Sarjana Sains Informasi (S.S.I)  
pada tanggal 24 Juni 2025

**Susunan Dewan Penguji**

**Ketua Penguji** : Dr. Ir. Mokhamad Amin Hariyadi, M.T  
NIP. 196701182005011001  
**Anggota Penguji I** : Annisa Fajriyah, M.A  
NIP. 198801122020122002  
**Anggota Penguji II** : Fakhris Khusnu Reza Mahfud, M.Kom.  
NIP.199005062019031007  
**Anggota Penguji III** : Yulianto, M. Pd.I  
NIP. 198707122019031005

**Tanda  
Tangan**



Mengetahui,  
Ketua Program Studi



Dr. Ir. Mokhamad Amin Hariyadi, M.T  
NIP. 196701182005011001

## PERNYATAAN KEASLIAN TULISAN

Saya akan bertanda tangan dibawah ini:

Nama : Rohimullah At-Thobroni

NIM : 210607110039

Prodi : Perpustakaan Dan Sains Informasi

Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benarbanar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan 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, 15 Juni 2025

Yang membuat pernyataan



Rohimullah At Thobroni  
NIM. 210607110039

## **KATA PENGANTAR**

Assalamualaikum Wr. Wb.

Puji syukur kehadiran Allah SWT yang telah melimpahkan Rahmat serta hidayah-Nya, sehingga peneliti dapat menyelesaikan studi dan skripsi ini dengan baik dan lancar. Shalawat serta salam tercurahkan kepada Rasulullah Muhammad SAW beserta sahabat-sahabatnya. Peneliti menyadari bahwa banyak pihak yang telah mendukung dalam menyelesaikan penulisan skripsi ini. Oleh karena itu, peneliti mengucapkan terima kasih kepada:

1. Kedua orang tua, Ibu Rahadian Rahmawati Dewi dan Bapak Sholahuddin serta keluarga besar yang selalu memberikan dukungan moral, materi maupun spiritual sehingga penulis dapat merasakan dan menyelesaikan pendidikan perkuliahan hingga akhir dengan baik. Serta adik penulis Asola Rahmabillah yang telah memberikan semangat dan senantiasa mendukung kepada penulis
2. Bapak Dr. Ir. M. Amin Hariyadi, M.T selaku Ketua Program Studi Perpustakaan dan Ilmu 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.
3. Bapak Fakhris Khusnu Reza Mahfud, M.Kom selaku Dosen Pembimbing I yang telah memberikan waktu, bimbingan, nasehat, dan seluruh dukungannya selama proses pengerjaan skripsi, sehingga tulisan ini dapat terselesaikan dengan baik.
4. Bapak Yulianto selaku Dosen Pembimbing II yang telah meluangkan waktunya untuk membimbing pengerjaan skripsi ini.
5. Bu Annisa selaku Dosen Penguji II serta menjadi wali dosen yang telah mendampingi dan mengawasi setiap tahap pengerjaan skripsi hingga selesai.
6. Seluruh dosen dan staff Program Studi Perpustakaan dan Ilmu Informasi yang telah membagikan ilmu dan pengalamannya.

7. Kepada bapak Rektor UIN Maulana Malik Ibrahim Malang beserta staf-stafnya dan Seluruh dekan-dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang
8. Seluruh mahasiswa Program Studi Perpustakaan dan Sains Informasi angkatan 2021 (GARYATAMA) yang turut mendukung dalam pengerjaan skripsi ini.
9. Kepada teman-teman Kontrakan The-Raid, yang telah memberikan tempat yang nyaman sebagai rumah untuk berkeluh kesah dan berjuang bersama penulis.
10. Kepada saudari Aprilina Suryani, yang sudah memberikan segala dukungan dalam berbagai bentuk selama penulisan skripsi ini berlangsung, saya ucapkan terimakasih sebanyak-banyaknya.
11. Kepada teman-teman penulis, yaitu Diaz Fawaz, Royhan Syauqi, Alfian Kurniawan, Hasbur Rahman Afif, Handi Yusuf, Aghif Fatullah, Muhammad Ikhsan, Riza Firdiansya, Firza Waliyuhag, Farid Ahmad, yang telah memberikan semangat, bantuan dan bimbingannya kepada penulis

Penulis menyadari bahwa pada penyusunan skripsi ini masih terdapat kekurangan. Penulis berharap semoga skripsi ini bisa memberikan manfaat, ilmu, serta sedikit wawasan yang dapat penulis bagikan kepada seluruh pembaca skripsi ini, tak terkecuali penulis pribadi, Amin Ya Rabbal Alamin

Wassalamu'alaikum Wr.Wb.

Malang, 15 Juni 2025

Penulis, Rohimullah At-Thobroni  
NIM. 210607110039

## ABSTRAK

Thobroni, Rohimullah At, 2025. **Analisis Sentimen Pengguna *iBI Library* Pada Ulasan *Google Play Store* Dengan *Multinomial naïve bayes*. Skripsi. 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) Yulianto, S.Pd.I**

**Kata kunci:** *analisis sentimen, multinomial naïve bayes, naïve bayes, text mining, perpustakaan digital, iBI library*

IBI library merupakan layanan perpustakaan digital berbasis aplikasi mobile yang dikembangkan oleh perpustakaan Bank Indoneisa dengan kualitas manajemen yang berstandar internasional. Pada aplikasi *iBI Library* masih sering terjadi permasalahan seperti error saat melakukan peminjaman buku dan tombol navigasi yang kurang *familiar*. Oleh karena itu dilakukan analisis sentimen terhadap ulasan pengguna aplikasi *iBI Library pada media Google play Store*. penelitian ini dilakukan untuk mengetahui performa algoritma *Multinomial naïve bayes* dalam mengklasifikasikan sentimen *positive* dan *negative* serta mengevaluasi kualitas layanan dan pengalaman pengguna. Metode yang digunakan adalah metode *text mining* dengan tahapan pengumpulan data ulasan dari *Google Play Store*, *pre-processing*, pelabelan data, pembagian data *training* dan *test* dengan dua rasio yaitu (70:30 dan 80:20). Pengukuran performa algoritma klasifikasi *Multinomial Naïve Bayes* dilakukan dalam bentuk akurasi, presisi, recall dan f1-score. Pada skenario rasio 80:20 menghasilkan nilai akurasi 83%, presisi 82%, *recall* 82%, dan *f-measure* 82%, performa klasifikasi yang dihasilkan rasio tersebut lebih unggul dibanding performa pada rasio 70:30 dan 80:20 tanpa *pre-processing*. Selain itu hasil analisis sentimen menemukan bahwa meskipun banyak sentimen negatif, seperti kesulitan akses koleksi dan pengalaman pengguna, banyak juga yang memberikan ulasan positif terkait koleksi yang lengkap, akses yang mudah di berbagai perangkat, dan bantuan layanan yang cepat. Beberapa komentar negatif tersebut dapat digunakan *iBI Library* untuk meningkatkan layanan kepada penggunanya.

## ABSTRACT

**Thobroni, Rohimullah At, 2025. Sentiment Analysis of iBI Library Users on Google Play Store Reviews Using Multinomial Naïve Bayes. Undergraduate Thesis. Department of Library and Information Science, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang.  
Supervisors: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Yulianto, S.Pd.I**

**Keywords:** *sentiment analysis, multinomial naïve bayes, naïve bayes, text mining, digital library, iBI library*

iBI Library is a digital library service based on a mobile application developed by Bank Indonesia Library with international standard management quality. The iBI Library app still frequently experiences issues such as errors during book borrowing and unfamiliar navigation buttons. Therefore, a sentiment analysis was conducted on user reviews of the iBI Library app on the Google Play Store. This study was conducted to assess the performance of the Multinomial Naive Bayes algorithm in classifying positive and negative sentiments, as well as evaluating service quality and user experience. The method used is text mining, involving data collection from Google Play Store reviews, pre-processing, data labeling, and dividing the data into training and test sets with two ratios (70:30 and 80:20). The performance of the Multinomial Naive Bayes classification algorithm was measured in terms of accuracy, precision, recall, and F1-score. In the 80:20 ratio scenario, the accuracy was 83%, precision was 82%, recall was 82%, and f-measure was 82%. The classification performance of this ratio was superior to that of the 70:30 ratio and the 80:20 ratio without pre-processing. Additionally, the sentiment analysis revealed that while there were many negative sentiments, such as difficulties accessing the collection and user experience issues, there were also many positive reviews regarding the comprehensive collection, easy access across various devices, and quick customer service assistance. Some of these negative comments can be used by iBI Library to improve its services for users.

## مستلخص البحث

ثوبروني ، روحيم الله في ، 2025. قم تحليل معنويات مستخدم مكتبة iBI على مراجعات متجر *Google Play* باستخدام *Bayes* الساذجة متعددة الحدود. اطروحة. برنامج دراسة علوم المكتبات والمعلومات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، ملانج.  
المشرف الأول: فخرس خصنو رضا محفوظ، الماجستير. المشرف الثاني: يولي نطا ، الماجستير.

الكلمات الرئيسية: تحليل المشاعر ، بايز الساذجة متعددة الحدود ، الخلجان الساذجة ، التنقيب عن النصوص ، المكتبة الرقمية ، مكتبة iBI

مكتبة IBI هي خدمة مكتبة رقمية قائمة على تطبيقات الهاتف المحمول تم تطويرها بواسطة مكتبات بنك إندونيسيا بجودة إدارة قياسية دولية. في تطبيق مكتبة iBI ، لا تزال هناك غالباً مشاكل مثل الأخطاء عند استعارة الكتب وأضرار التنقل غير المؤلف. لذلك ، تم إجراء تحليل المشاعر على مراجعات المستخدمين لتطبيق مكتبة iBI على وسائط متجر *Google Play*. أجريت هذه الدراسة لتحديد أداء خوارزمية بايز الساذجة متعددة الحدود في تصنيف المشاعر الإيجابية والسلبية وتقييم جودة الخدمة وتجربة المستخدم. الطريقة المستخدمة هي طريقة التنقيب عن النصوص مع مراحل جمع بيانات المراجعة من متجر *Google Play* ، والمعالجة المسبقة ، وتصنيف البيانات ، والتدريب على مشاركة البيانات والاختبار بنسبتين ، وهما (70:30 و 80:20). يتم إجراء قياس أداء خوارزمية تصنيف *Bayes* الساذجة متعددة في شكل دقة ودقة واستدعاء ودرجة f1. في سيناريو نسبة 80:20 ، ينتج قيمة دقة تبلغ 83% ، ودقة 82% ، واستدعاء 82% ، ومقياس  $f1$  82% ، وأداء التصنيف الناتج عن هذه النسبة يتفوق على أداء نسب 70:30 و 80:20 دون معالجة مسبقة. بالإضافة إلى ذلك ، وجد تحليل المشاعر أنه على الرغم من العديد من المشاعر السلبية ، مثل صعوبة الوصول إلى المجموعات وتجربة المستخدم ، قدم الكثيرون أيضاً مراجعات إيجابية فيما يتعلق بالمجموعة الكاملة ، وسهولة الوصول على الأجهزة المختلفة ،

## DAFTAR ISI

HALAMAN JUDUL .....	ii
LEMBAR PERSETUJUAN .....	Error! Bookmark not defined.
LEMBAR PENGESAHAN .....	Error! Bookmark not defined.
PERNYATAAN KEASLIAN TULISAN .....	Error! Bookmark not defined.
KATA PENGANTAR .....	vi
ABSTRAK .....	viii
ABSTRACT .....	ix
مستخلص البحث .....	ix
DAFTAR ISI .....	xi
DAFTAR TABLE .....	xiii
DAFTAR GAMBAR .....	xiv
<b>BAB I PENDAHULUAN</b> .....	15
1.1 Latar Belakang .....	15
1.2 Identifikasi Masalah .....	19
1.3 Tujuan Penelitian .....	19
1.4 Manfaat Penelitian .....	19
1.5 Batasan Masalah .....	19
1.6 Sistematika Penulisan .....	20
<b>BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI</b> .....	22
2.1 Penelitian Terdahulu .....	22
2.2 Landasan Teori .....	25
2.2.1 Analisis Sentimen .....	25
2.2.2 Data Mining .....	26
2.2.3 <i>Pre-processing</i> .....	26
2.2.4 Term Weighting .....	28
2.2.5 Metode <i>Naïve bayes</i> .....	29
2.2.6 <i>Confusion matrix</i> .....	31
2.2.7 Evaluasi (Muhasabah) .....	33
<b>BAB III METODE PENELITIAN</b> .....	36
3.1 Jenis Penelitian .....	36
3.2 Alur Penelitian .....	36
3.3 Desain Sistem dan Analisis Data .....	37
3.2.1 Labeling .....	38
3.2.2 <i>Scrapping data</i> .....	38
3.2.3 <i>Pre-processing</i> .....	39

3.2.4 Pembobotan TF-IDF .....	47
3.2.5 Klasifikasi <i>Multinomial naïve bayes</i> .....	55
3.2.6 Evaluasi klasifikasi .....	61
3.4 Tempat dan Waktu Penelitian .....	62
3.1 Subjek dan Objek Penelitian .....	62
3.2 Sumber Data.....	63
3.3 Teknik Pengumpulan Data.....	63
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	64
4. 1.Hasil Penelitian .....	64
4.1.1 Pengambilan Data.....	64
4.1.2 Pelabelan Data .....	64
4.1.3 Hasil <i>Pre-processing</i> .....	65
4.1.4 Hasil Pembobotan Kata .....	68
4.1.5 Hasil Pelatihan Data .....	69
4.1.6 Hasil Pengujian Data .....	70
4. 2.Pembahasan.....	74
<b>BAB V KESIMPULAN DAN SARAN</b> .....	85
5. 1 Kesimpulan .....	85
5. 2 Saran .....	85
<b>DAFTAR PUSTAKA</b> .....	87
<b>LAMPIRAN</b> .....	92

## DAFTAR TABLE

Table 3. 1. Contoh Data Labeling .....	38
Table 3. 2. Contoh Proses Cleaning .....	39
Table 3. 3. Contoh Proses Case folding .....	41
Table 3. 4. Contoh Proses Normalisasi .....	42
Table 3. 5. Contoh Proses Stopword .....	44
Table 3. 6. Proses Tokenizing .....	45
Table 3. 7. Contoh Proses Stemming .....	46
Table 3. 8. Contoh Proses Term Frekuensi .....	48
Table 3. 9. Contoh Proses Normalisasi Term Frekuensi .....	49
Table 3. 10. Contoh Proses Document Frequency dan Invers Document Frequency .....	51
Table 3. 11. Contoh Proses TF-IDF .....	53
Table 3. 12. Contoh Pembagian Data Training dan Data Test .....	55
Table 3. 13. Contoh Proses Penjumlahan Label Term .....	56
Table 3. 14. Contoh Proses Probabilitas Prior.....	57
Table 3. 15. Contoh Proses Probabilitas Likelihood .....	58
Table 3. 16. Contoh Proses Term Data Test.....	59
Table 3. 17. Contoh Proses Predicted Data Test .....	60
Table 3. 18. Contoh Confusion matrix .....	61
Table 3. 19. Contoh Proses Penghitungan Accuracy, Precision, Recall dan F-measure .....	62
Table 4. 1 Tahap <i>Cleaning</i> .....	65
Table 4. 2 Tahap <i>Case folding</i> .....	66
Table 4. 3 Tahap Normalisasi .....	66
Table 4. 4 Tahap <i>Stopword</i> .....	67
Table 4. 5 Tahap Tokenized.....	67
Table 4. 6 Tahap <i>Stemming</i> .....	68
Table 4. 7 Tahap TF-IDF .....	68
Table 4. 8 Hasil Peforma <i>Multinomial naïve bayes</i> .....	71
Table 4. 9 Hasil Data <i>Test</i> .....	73
Table 4. 10 Implementasi Muhasabah Dalam Analisis Sentimen .....	83

## DAFTAR GAMBAR

Gambar 2. 1. Alur Penelitian.....	36
Gambar 2. 2. Desain Sistem.....	37
Gambar 4. 1 Hasil labeling dari para ahli.....	64
Gambar 4. 2 Hasil data <i>test</i> .....	71
Gambar 4. 3 Hasil wordcloud .....	74

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Perpustakaan khusus membawa perubahan yang cukup besar bagi kehidupan masyarakat dalam perkembangan teknologi informasi saat ini. Perkembangan teknologi yang menghadirkan teknologi *smartphone* yang di dukung oleh jaringan internet sehingga pengguna dapat mengakses informasi secara online dan diakses dimana saja (Jumasa, 2019). Strategi perpustakaan untuk meningkatkan layanan tidak hanya dengan melakukan inovasi melalui media social atau media tercetak lainnya, namun, dapat dengan membuat dan mengembangkan sistem aplikasi perpustakaan digital berbasis aplikasi. Bentuk perpustakaan digital ada yang berbasis website dan berbasis mobile (Wahyuni & Dewi, 2019). Salah satu perpustakaan yang menyediakan layanan perpustakaan digital berbasis mobile berbentuk aplikasi yaitu Perpustakaan Digital Bank Indonesia (*iBI Library*). *iBI Library* adalah aplikasi yang sejenis dengan Ipusnas yang memberikan layanan berbentuk perpustakaan digital (Rifqi, 2019).

*iBI Library* yang dikembangkan oleh perpustakaan Bank Indonesia memiliki kualitas manajemen yang berstandar internasional untuk layanan perpustakaan, yang telah memperoleh sertifikasi ISO 9001:2015 selain itu, sebagian besar kegiatan perpustakaan kantor pusat dengan kantor perwakilan Bank Indonesia berkoordinasi dengan baik secara online melalui *Library Management Sistem*. Hal itu memungkinkan pustakawan BI berdiskusi tanpa batasan ruang dan waktu, dan perpustakaan KPw mendapat akreditasi A oleh Perpustakaan Nasional Republik Indonesia. Serta *iBI Library* sendiri memiliki koleksi yang selalu up to date yang mencakup berbagai disiplin ilmu dan genre buku, maka dari itu *iBI Library* sangat berpengaruh bagi pengguna dalam mencari informasi menggunakan aplikasi tersebut.

Aplikasi *iBI Library* tersedia pada *Google Play store*, di dalam *Play store* tersebut pengguna dapat melaporkan beberapa masalah dalam ulasan. Berdasarkan ulasan tersebut pengguna mengalami beberapa permasalahan seperti terjadi error saat melakukan peminjaman buku dan saat file buku ingin dibuka terjadi error,

selain itu pengguna juga mengalami kesulitan saat menggunakan fitur dan navigasi yang tersedia pada aplikasi sehingga kurang familiar oleh pengguna baru. Pada penelitian sebelumnya, ditemukan masalah juga dimana beberapa pengguna tidak dapat masuk ke aplikasi yang dikarenakan kesalahan sistem. Hal itu termasuk juga pada fungsi pencarian buku dengan pengaturan navigasi yang kurang dipahami oleh pengguna (Pasaribu et al., 2024). Untuk mengatasi masalah tersebut, diperlukan masukan dan evaluasi yang berasal dari komentar dan pendapat pengguna dalam rangka meningkatkan layanan *IBI Library*. Meskipun *Play store* sudah dilengkapi dengan fitur *rating*, namun terkadang terdapat ketidaksesuaian antara isi ulasan dengan pemberian rating, sehingga *positive* atau *negativenya* sebuah ulasan tidak dapat hanya dilihat dari *rating* saja (Nadira et al., 2023). Salah satu teknik penilaian layanan *IBI Library* yang bisa digunakan adalah analisis sentimen, dari analisis sentimen bisa mendapatkan opini yang disampaikan pengguna, dalam opini tersebut dapat berupa pujian atau kritik. Dari opini tersebut diklasifikasikan sebagai sentimen *positive* atau *negative* berdasarkan emosi dari opini yang diberikan (Ramadhani & Suryono, 2024).

Dengan analisis sentiment dapat melacak pendapat dan pemikiran pengguna tentang suatu layanan atau produk tersebut, hal itu membantu mengidentifikasi masalah dan mendorong suatu instansi untuk menemukan solusi. Selain itu, memberikan wawasan pengguna tentang opini publik terhadap suatu layanan atau produk yang mempengaruhi suatu keputusan pengguna serta memberikan evaluasi kepada *iBI Library* untuk terus meningkatkan layanan berdasarkan sentiment penggunanya

Dalam Islam, mengembangkan dan memanfaatkan ilmu pengetahuan serta teknologi untuk kebaikan manusia memiliki keutamaan. Yang mencakup aspek hukum fikih, nilai-nilai moral, serta sebagai wujud syukur kepada Allah SWT (Yulianto & Suprpto, 2023). Analisis Sentimen sendiri memiliki tujuan untuk mengelola perbedaan yang terjadi melalui kecenderungan emosi atau moral terhadap opini pada topik tersebut. Hal tersebut berkaitan dengan Al-Qur'an surat Hud ayat 118-119 tentang perbedaan dan perselisihan untuk mengelola suatu hal :

وَلَوْ شَاءَ رَبُّكَ لَجَعَلَ النَّاسَ أُمَّةً وَاحِدَةً وَلَا يَرَالُونَ مَخْتَلِفِينَ ۗ إِلَّا مَن رَّحِمَ رَبُّكَ وَلِذَلِكَ

خَلَقَهُمْ ۗ وَتَمَّتْ كَلِمَةُ رَبِّكَ لَأَمْلَأَنَّ جَهَنَّمَ مِنَ الْجِنَّةِ وَالنَّاسِ أَجْمَعِينَ ﴿١١٣﴾

*Artinya : Jika Tuhanmu menghendaki, tentu Dia akan menjadikan manusia umat yang satu. Namun, mereka senantiasa berselisih (dalam urusan agama), kecuali orang yang dirahmati oleh Tuhanmu. Menurut (kehendak-Nya) itulah Allah menciptakan mereka. Kalimat (keputusan) Tuhanmu telah tetap, “Aku pasti akan memenuhi (neraka) Jahanam (dengan pendurhaka) dari kalangan jin dan manusia semuanya.*

Uraian dalam ayat tersebut menjelaskan bahwa perbedaan adalah fitrah manusia yang telah ditetapkan oleh Allah, sehingga manusia bebas memilih sesuai kehendaknya. Sebagaimana disebutkan oleh Quraish Shihab dalam kitab tafsirnya al-Misbah. Allah tidak menghendaki semua manusia sejak dahulu hingga seterusnya menjadi satu umat dengan agama, pendapat, dan kecenderungan yang sama. Mereka berbeda pendapat karena cara berpikir yang berbeda serta hawa nafsu, kecuali mereka yang mendapat petunjuk Allah. Dalam bukunya yang berjudul 'Yang Hilang Dari Kita: Akhlak', disebutkan bahwa mereka yang dirahmati dan diberi petunjuk adalah yang mampu mengelola perbedaan tersebut (Shihab, 2016). Dalam konteks ini, pengelolaan bertujuan untuk mengetahui solusi yang dominan ketika membahas suatu topik dalam musyawarah atau forum lainnya. Selain itu, pengelolaan ini juga bertujuan agar setiap perbedaan dapat dihargai dan toleransi dapat tercipta dalam masyarakat maupun diskusi (Shihab, 2002). Tafsir tersebut dapat dipahami bahwa pentingnya kemampuan mengelola perbedaan pendapat di antara masyarakat, sehingga jelas memahami apa yang telah terjadi. Salah satunya dilakukan analisis sentiment.

Penelitian ini menggunakan metode klasifikasi *Multinomial naïve bayes*. *Naïve bayes Classifier* (NBC) yaitu algoritma yang dikembangkan dengan data internal yang diambil dari internet kemudian diproses untuk menentukan kalimat opini termasuk dalam opini *positive* dan *negative* (Darwis et al., 2021). Kelebihan *Naïve bayes* mendapatkan hasil *accuracy* yang lebih baik dari support vector machine dengan *accuracy* sebesar 73,65%, hal itu dikarenakan *naïve bayes* mempunyai karakter yang sederhana, cepat dan menghasilkan *accuracy* yang tinggi saat diimplementasikan pada data yang besar (Fikri et al., 2020). *Multinomial naïve*

*bayes* dapat dikatakan terbaik untuk klasifikasi teks karena dapat menghitung frekuensi setiap kata, hal itu memiliki dampak lebih tinggi dalam mengkategorikan teks. Selain itu, algoritma *naïve bayes* lebih unggul dengan tingkat *accuracy* mencapai 80,00%, dibandingkan algoritma *k-nearest neighbor* dengan tingkat *accuracy* 55,00%. Selain itu, berdasarkan IEEE *International Conference on Data Mining*, *naïve bayes* termasuk dalam sepuluh besar algoritma dalam bidang data mining (A. Oktian Permana & Sudin Saepudin, 2023).

Seperti Penelitian sebelumnya yang menggunakan metode yang sama dengan judul “Analisis Sentimen Dan Klasifikasi Tweet Terkait Naiknya Kasus Omicron Menggunakan Naive Bayes Classifier” bahwa pemerintah Indonesia menyampaikan informasi tentang varian baru COVID-19 yaitu omicron melalui twitter. Hal ini dikarenakan peningkatan kasus virus yang signifikan, diperlukan analisis sentiment untuk mendapatkan komentar *positive* dan *negative* yang ditulis public pada twitter. Penelitian ini menggunakan metode klasifikasi *naïve bayes*, hasil klasifikasi data menunjukkan nilai *accuracy* sebesar 90.0%, nilai *precision* 90.00%, nilai *recall* 90.00%, dan nilai *f-measure* 90.00% (Duarte & Eddy Nurraharjo, 2023).

Maka dari itu, penelitian ini dilakukan pengambilan data melalui *scrapping* data menggunakan google colab pada ulasan google *Play store*. Selanjutnya melakukan text *pre-processing* (*case folding*, *cleansing*, *tokenizing*, *stemming*, dan *filtering*), pembobotan kata (*vectorizing TF-IDF*), pengolahan data, pengujian klasifikasi menggunakan algoritma NBC lalu kemudian hasil klasifikasi. Kemudian menggunakan tools Google Colab dengan menggunakan bahasa pemrograman Python (Maulana et al., 2023).

Berdasarkan permasalahan yang sudah dijelaskan sebelumnya peneliti tertarik untuk melakukan penelitian dengan judul “Analisis Sentimen Pengguna (*iBI Library*) Pada Ulasan *Google Play Store* Dengan *Multinomial naïve bayes*”. Data penelitian ini diambil dari komentar *Google Play store* dengan menggunakan dengan metode *scrapping*, lalu data komentar tersebut diklasifikasikan menggunakan metode *Multinomial Naïve bayes*. Penelitian ini bertujuan untuk menganalisis sentimen pengguna pada kolom komentar *Google Play Store* dan

sebagai bahan kajian evaluasi bagi Bank Indonesia untuk terus meningkatkan kualitas layanan *iBI Library* berdasarkan pengalaman pengguna, serta mengevaluasi performa algoritma *Multinomial naïve bayes* dalam mengklasifikasikan ulasan tersebut.

## **1.2 Identifikasi Masalah**

Identifikasi masalah pada penelitian ini yaitu Bagaimana hasil peforma algoritma *Multinomial naïve bayes* dalam analisis sentimen pengguna terhadap aplikasi perpustakaan digital Indonesia (*iBI Library*) melalui ulasan *Google Play Store*?

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini yaitu untuk mengukur hasil peforma algoritma *Multinomial naïve bayes* dalam analisis sentimen pengguna terhadap aplikasi perpustakaan digital Indonesia (*iBI Library*) melalui ulasan *Google Play Store*.

## **1.4 Manfaat Penelitian**

Manfaat penelitian menjadi referensi penelitian dan pembaharuan lebih lanjut terkait analisis sentimen pengguna aplikasi perpustakaan digital indonesia (*iBI Library*) melalui ulasan *Google Play Store* dengan *Multinomial naïve bayes*. Selain itu, hasil penelitian ini digunakan sebagai evaluasi terhadap *iBI Library* yang dikembangkan oleh Perpustakaan Bank Indonesia

## **1.5 Batasan Masalah**

Adapun untuk menghindari meluasnya ruang lingkup penelitian sehingga tidak terfokus, perlu adanya batasan masalah pada penelitian. Batasan masalah yang digunakan pada penelitian ini sebagai berikut:

1. Data diambil secara keseluruhan pada ulasan *Google Play store* mulai 31 Oktober 2017 sampai dengan 18 Oktober 2024
2. Data yang diambil 287 ulasan
3. Data yang digunakan hanya ulasan *positive* dan *negative*
4. Data yang digunakan adalah ulasan yang menggunakan bahasa Indonesia

## **1.6 Sistematika Penulisan**

Sistematika penulisan dibuat agar mempermudah pembaca untuk memahami alur penulisan pada penelitian ini. Sistematika yang dibuat terdiri dari lima bab antara lain:

### **BAB I PENDAHULUAN**

Pendahuluan berisi latar belakang terkait pengalaman pengguna saat menggunakan aplikasi layanan perpustakaan *iBI Library*. Dirumuskan bagaimana *accuracy*, *recall*, *precision* dan *F-measure* dari *Multinomial naïve bayes* dalam analisis sentimen pengguna terhadap aplikasi perpustakaan digital Indonesia (*iBI Library*) melalui ulasan *Google Play Store*. Tujuannya untuk menilai hasil *accuracy*, *recall*, *precision*, dan *F-measure*, serta manfaatnya dapat menjadi acuan bagi penelitian dan pengembangan lebih lanjut mengenai analisis sentiment. Batasan masalah penelitian dan sistematika penulisan yang dilakukan pada penelitian ini.

### **BAB II TINJAUAN PUSTAKA**

Bab ini membahas penelitian terdahulu terkait analisis sentiment terhadap suatu masalah dan metode klasifikasi yang sama. Pada subbab berikutnya terdapat beberapa dasar teori yang mendukung penelitian ini terkait analisis sentiment, klasifikasi, text mining, dan metode *multinomial naïve bayes* pada *Google Play store*, dasar teori tentang evaluasi klasifikasi.

### **BAB III METODOLOGI PENELITIAN**

Metodologi penelitian menjelaskan alur penelitian yang dilakukan, seperti jenis penelitian, desain penelitian, waktu dan tempat penelitian, sumber data komentar *Play store*, metode pengumpulan data dengan scraping, lalu dilakukan tahap *pre-processing* diantaranya *case folding* *tokenizing*, *filtering*, *stemming*, penghapusan *stopwords*, setelah itu dilanjutkan proses Term Frequency - Inverse Document Frequency (TF-IDF), pembagian data *training* dan *testing* data lalu metode analisis *naïve bayes* dan evaluasi seperti menghitung *accuracy*, *precision*, *recall*, dan *F-measure*.

#### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini menggambarkan profil singkat tentang objek penelitian yang dilakukan, hasil penelitian analisis sentiment berupa *accuracy*, *recall*, *precision* dan *F-measure* dari *Multinomial naïve bayes*. Selain itu juga terdapat bagian pembahasan untuk analisis lebih lanjut tentang hasil penelitian ini.

#### **BAB V KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari hasil penelitian berdasarkan rumusa masalah terkait analisis sentiment terhadap *iBI Library* serta berisi saran yang dapat digunakan untuk pengembangan penelitian selanjutnya

## **BAB II**

### **TINJAUAN PUSTAKA DAN LANDASAN TEORI**

#### **2.1 Penelitian Terdahulu**

Penelitian mengenai analisis sentiment sudah banyak dilakukan terutama menggunakan media social twitter. Salah satunya seperti penelitian yang berjudul “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi” Penelitian ini berfokus pada E-learning, yaitu pembelajaran berbasis elektronik menggunakan komputer. Salah satu aplikasi e-learning yang populer saat ini adalah Ruang Guru. Untuk menilai keberhasilan aplikasi ini, dilakukan analisis sentimen terhadap komentar pengguna di media sosial Twitter, sebanyak 513 tweet, yang setelah dilakukan pembersihan data terdiri dari 338 tweet *positive* dan 175 tweet *negative*. Proses yang dilakukan diantaranya business understanding, data understanding, *pre-processing*, modeling menggunakan algoritma Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbour (K-NN), dan pemilihan fitur dengan algoritma Particle Swarm Optimization (PSO). Penelitian ini membandingkan metode NB, SVM, K-NN tanpa pemilihan fitur dengan metode yang menggunakan pemilihan fitur, serta membandingkan nilai Area Under Curve (AUC) dari masing-masing metode untuk menemukan algoritma yang paling optimal. Hasil pengujian menunjukkan bahwa algoritma PSO berbasis SVM memberikan hasil terbaik dengan *accuracy* 78,55% dan AUC 0,853 (Giovani et al., 2020).

Adapun penelitian kedua mengenai analisis sentimen pada aplikasi ipusnas yang berjudul “Klasifikasi Ulasan Pengguna Aplikasi: Studi Kasus Aplikasi Ipusnas Perpustakaan Nasional Republik Indonesia (Pnri)” Penelitian ini membahas penurunan tren pengguna baru aplikasi iPusnas yang berdampak pada penurunan pencapaian target laporan LKIP Pujasintara PNRI 2020-2024. Penurunan ini terkait dengan peringkat ulasan pengguna di Google *Play store* yang lebih rendah dibandingkan aplikasi sejenis lainnya. Electronic Word of Mouth (EWOM) sangat mempengaruhi keputusan calon pengguna baru dalam memilih aplikasi terbaik, karena mempertimbangkan peringkat dan ulasan pengguna. Beberapa penelitian sebelumnya menunjukkan bahwa analisis atau ekstraksi

informasi penting dari ulasan pengguna secara manual sering menghadapi kesulitan. Analisis ulasan sangat penting untuk pengembangan fitur layanan aplikasi guna meningkatkan kepuasan pengguna dan peringkat aplikasi, sehingga diperlukan alat bantu klasifikasi ulasan pengguna secara otomatis dengan menemukan model terbaik. Penelitian ini menggunakan metodologi CRISP-DM yang melalui enam tahapan, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan data dan yang terakhir evaluasi. Algoritma klasifikasi yang digunakan adalah *Naïve bayes*, Logistic Regression, Support Vector Machine (SVM), serta kombinasi fitur tf-idf unigram, bigram, dan trigram. Hasil penelitian menunjukkan bahwa kombinasi fitur tf-idf unigram (F1) dengan algoritma SVM mencapai nilai terbaik untuk *precision*, *recall*, dan *f-measure* masing-masing sebesar 87%. Nilai evaluasi terendah adalah *precision* 55% dari kombinasi fitur F2 dengan SVM, *recall* 42%, dan *f-measure* 32% dari kombinasi fitur F3 dengan logistic regression (Septiani & Budi, 2022).

Penelitian ketiga dengan judul “Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Classifier” tentang sentimen publik terhadap isu penundaan pemilu yang terbagi menjadi tiga: *negative*, netral, dan *positive*. Untuk mengetahui persentase sentimen terbesar, diperlukan analisis sentimen terhadap data cuitan Twitter dari periode 1 Januari 2022 hingga 27 Mei 2022. Penelitian ini bertujuan untuk memetakan analisis sentimen sebelum dan sesudah Muhaimin Iskandar mengutarakan isu tersebut, dengan menggunakan model *Naive Bayes Classifier*. Analisis sentimen sebelum Muhaimin Iskandar mengutarakan isu diperlukan untuk menguji apakah klaim penundaan pemilu benar-benar diinginkan oleh rakyat. Sedangkan analisis sentimen setelah Muhaimin Iskandar mengutarakan isu dilakukan untuk melihat polarisasi di Twitter, seberapa besar dukungan maupun penolakan terhadap isu tersebut. Penelitian ini menggunakan metode ekstraksi data, pembersihan data, pelabelan sentimen, analisis sentimen, dan perhitungan polarisasi sentimen. Hasil pelatihan menunjukkan nilai *precision* 98%, *recall* 100%, *f-measure* 99% untuk sentimen *negative*; *precision* 100%, *recall* 94%, *f-measure* 96,9% untuk sentimen netral; dan *precision* 96,1%, *recall* 100%, *f-measure* 98% untuk sentimen *positive*. Wordcloud

yang dihasilkan menunjukkan bahwa masyarakat baru membahas isu penundaan pemilu setelah Muhaimin mengutarakan isu tersebut. Sementara hasil prediksi sentimen terhadap isu penundaan pemilu menggunakan data pasca isu dilontarkan oleh Muhaimin cenderung menunjukkan sentimen *positive* sebesar 60.442 cuitan unik, dengan 58.873 cuitan unik menunjukkan sentimen *negative*, dan 32.223 cuitan menunjukkan sentimen netral (Perdana et al., 2022).

Penelitian selanjutnya yang berjudul “Analisis Sentimen Masyarakat Terhadap *Paylater* Menggunakan Metode *Naive Bayes Classifier*” yang membahas tentang Belanja online sangat digemari masyarakat karena kemudahannya. Metode pembayaran *paylater* mendukung kemudahan ini, namun juga dapat memicu perilaku *negative* seperti pembelian impulsif. Beragam tanggapan masyarakat mendorong peneliti untuk meneliti pandangan mereka terhadap *paylater*. Penelitian ini menggunakan metode pengumpulan data, *pre-processing*, labeling dengan *Naive Bayes Classifier* dan *TextBlob*, lalu yang terakhir evaluasi. Dataset yang dikumpulkan dari Twitter berjumlah 405 data. Analisis sentimen dengan *Naive Bayes Classifier* menunjukkan 70,62% atau 286 data bersentimen *negative*, 22,72% atau 92 data bersentimen *positive*, dan 6,67% atau 27 data bersentimen netral. Sementara itu, metode *TextBlob* menunjukkan 55,8% atau 226 data bersentimen *negative*, 33,09% atau 134 data bersentimen *positive*, dan 11,11% atau 45 data bersentimen netral. Hasilnya menunjukkan bahwa masyarakat cenderung memiliki pandangan *negative* terhadap penggunaan *paylater*. Pengujian model dengan *confusion matrix* menunjukkan bahwa algoritma *Naive Bayes Classifier* lebih akurat dengan nilai 91% dibandingkan *TextBlob* yang hanya 61%. (Alfandi Safira & Hasan, 2023).

Penelitian yang terakhir yang berjudul “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma *Naive bayes* dan KNN” yang membahas tentang Pemandangan dan pembangunan Ibu Kota Negara Nusantara pada masa Presiden Joko Widodo akan dilakukan secara bertahap dari tahun 2024 hingga 2045. Hal ini menjadi topik perbincangan dan mengundang banyak reaksi, terutama dari masyarakat Indonesia. Isu pemindahan Ibu Kota Negara merupakan topik sensitif yang ramai dibahas di media sosial, termasuk Twitter. Media sosial pada

dasarnya digunakan untuk menyampaikan pendapat atau ekspresi. Pemanfaatan media sosial ini menjadi sarana yang dapat digunakan untuk isu politik atau topik yang sedang dibahas. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis sentimen masyarakat Indonesia terhadap pemindahan Ibu Kota Nusantara. Dalam penelitian ini, melalui beberapa tahapan diantaranya pengumpulan data, pemberian label, *pre-processing*, proses klasifikasi dan yang terakhir validasi dan evaluasi sedangkan metode yang digunakan adalah *Naïve bayes* (NB) dan K-Nearest Neighbor (KNN). Hasil penelitian menunjukkan bahwa metode *Naïve bayes* memberikan tingkat *accuracy* analisis sentimen sebesar 82,27%, nilai *precision* sebesar 86,36%, dan nilai *recall* sebesar 76,93%. Performa metode KNN juga menunjukkan hasil analisis dengan tingkat *accuracy* sebesar 88,12%, *precision* sebesar 93,98%, dan nilai *recall* sebesar 81,53% (Syahril Dwi Prasetyo et al., 2023).

Dari lima penelitian terdahulu yang telah di jabarkan diatas, penelitian mempunyai beberapa kesamaan yang dilakukan seperti konteks judul yang dipakai yaitu analisis sentiment dan menggunakan algoritma yang sama. Namun terdapat beberapa perbedaan pada penelitian terdahulu dengan menggunakan data yang diambil dari social media, dan hanya sebagian dari google *Play store*. Selain itu, hanya sebagian penelitian yang membahas tentang analisis sentimen terhadap layanan perpustakaan

## **2.2 Landasan Teori**

### **2.2.1 Analisis Sentimen**

Analisis sentimen adalah teknik dalam pengolahan data teks yang memanfaatkan Natural Language Processing (NLP) dan *Machine learning* (ML) untuk secara otomatis mengevaluasi teks guna menentukan emosi yang diekspresikan oleh penulis (*positive* dan *negative*). Bidang analisis sentimen ini merupakan cabang dari NLP yang berkembang dengan melihat interaksi antara manusia dan komputer, ekstraksi informasi, serta penyulingan emosi dari data di media sosial yang terus meningkat. Ini termasuk mengidentifikasi kata atau frasa yang mencerminkan sikap *positive* dan *negative*. dalam teks (Prakash & Aloysius, 2021).

Analisis sentiment terbagi menjadi tiga jenis utama diantaranya Multimodal sentiment analysis, Aspect-based sentiment analysis, Multilingual sentiment analysis. Jenis analisis yang sesuai dengan penelitian ini adalah aspect-based sentiment analysis. Karena jenis analisis yang berbasis aspek menggunakan metode NLP untuk menganalisis dan mengekstraksi emosi serta pendapat yang berkaitan dengan aspek atau fitur tertentu dari produk dan layanan. Contohnya, dalam ulasan restoran, peneliti dapat mengekstraksi sentimen terkait makanan, layanan, suasana, dan lainnya (Purnamasari et al., 2023)

### **2.2.2 Data Mining**

Data mining adalah alat yang memungkinkan pengguna untuk dengan cepat mengakses data dalam jumlah besar. Secara lebih spesifik, data mining adalah alat dan aplikasi yang menggunakan analisis pada data untuk menemukan pola-pola tersembunyi dalam data yang telah dikumpulkan (Marisa, 2005). Proses data mining adalah serangkaian langkah atau tahapan dalam menggali pengetahuan atau informasi dari data yang besar dan kompleks. Proses ini meliputi tahapan mulai dari pemilihan data, pre-prosesing data, pemodelan, dan evaluasi hasil (Wahyu et al., 2023). Data mining disebut juga dengan *knowledge discovery in database* (KKD) ataupun pattern recognition. KDD data mining digambarkan sebagai proses pencarian pengetahuan yang menarik dalam database seperti pola, asosiasi, aturan, perubahan, keganjilan dan struktur penting dari sejumlah besar bank data dan tempat penyimpanan informasi lainnya (Syahril et al., 2023)

### **2.2.3 Pre-processing**

*Pre-processing* merupakan tahap yang digunakan untuk mempersiapkan data mentah menjadi data yang siap untuk diolah, dan juga untuk memastikan bahwa data yang digunakan pada model *machine learning* merupakan data yang tepat (Purnamasari et al., 2023). Proses *pre-processing* ini melibatkan langkah-langkah seperti penanganan data yang hilang, penghapusan duplikasi, normalisasi data, dan pengubahan variabel kategorikal menjadi bentuk numerik yang dapat diolah dalam algoritma data mining (Agung et al., 2023). Data yang telah melalui tahap persiapan kemudian dilakukan tahap *pre-processing*. Tahap *pre-processing*

terdiri dari empat tahap, yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*(Mahfud et al., 2020).

a. *Case folding*

Proses normalisasi di sini adalah mengubah semua terms sehingga menghasilkan bentuk yang seragam. Salah satu caranya *case folding*. Ada dua opsi dalam melakukan transformasi bentuknya, dengan mengubah terms menjadi huruf besar (kapital) atau huruf kecil semuanya (Wahyu et al., 2023). Pada umumnya dilakukan dengan mengubah teks menjadi bentuk standar huruf kecil. Misalnya, kalimat "Program vaksin pemerintah kacau di mana pun pada antri bikin kepadatan aja vaksin gagal" menjadi "program vaksin pemerintah kacau dimana mana pada antri bikin kepadatan aja vaksin gagal" (Styawati et al., 2021).

b. *Tokenizing*

*Tokenizing* adalah sebuah proses untuk menyegmentasi kata dari sekumpulan teks menjadi token (kata tunggal) yang dipisahkan dengan spasi atau tanda baca (Wahyu et al., 2023). Yang berfungsi untuk memecah komentar menjadi kata-kata. Proses tokenisasi dilakukan dengan melihat setiap spasi pada komentar, kemudian berdasarkan spasi tersebut komentar dapat dipecah (Styawati et al., 2021).

c. *Filtering*

*Filtering* adalah tahap untuk menyaring terms. Pada *filtering* ini, kata tersebut akan dihilangkan atau dihapuskan, kata yang akan dihilangkan disebut dengan stop words (Wahyu et al., 2023). *Stopword* adalah proses menghapus kata-kata yang ada dalam daftar *stopword*. *Stopword* adalah kata-kata umum yang sering muncul dalam jumlah banyak dan memiliki fungsi tetapi tidak memiliki makna yang signifikan. Contoh *stopword* adalah "yang", "atau", dan lain-lain (Styawati et al., 2021).

d. *Stemming*

*Stemming* adalah kata pada kolom tokenisasi adalah hasil tokenisasi dan masih memiliki kata dengan imbuhan awalan dan akhiran. Setelah dilakukan *stemming*(Wahyu et al., 2023). Seperti halnya menyaring kata-kata seperti

kata sambung, kata ganti, dan kata depan menjadi bentuk dasar dengan menghapus prefiks atau sufiks. Contoh “program vaksin pemerintah kacau dimana mana pada antri bikin kepadatan aja vaksin gagal” menjadi “program vaksin pemerintah kacau dimana mana pada antri bikin kerumun aja vaksin gagal” (Styawati et al., 2021).

#### 2.2.4 Term Weighting

Istilah term weighting merupakan metode untuk memberikan bobot pada kata. Term weighting digunakan untuk mengubah data menjadi vektor numerik, di mana setiap vektor merepresentasikan teks data. Salah satu teknik yang sering digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*) (Purnamasari et al., 2023).

TF-IDF adalah metode pembobotan yang sering digunakan dalam pencarian informasi dan analisis data. Oleh karena itu, algoritma ini sering diterapkan dalam ekstraksi kata kunci, perbandingan kesamaan teks, dan klasifikasi topik. Algoritma TF-IDF menghitung setiap kata dan bobotnya dalam teks dokumen (Liu et al., 2022). Proses pembobotan TF-IDF digunakan untuk menentukan bobot atau nilai suatu kata atau istilah. Semakin sering sebuah kata muncul dalam dokumen, semakin besar bobot atau skor relevansinya. TF (*Term Frequency*) merujuk pada frekuensi kemunculan kata atau istilah dalam dokumen tertentu, sedangkan IDF (*Inverse Document Frequency*) mengukur pentingnya istilah tersebut dalam keseluruhan kumpulan dokumen. Rumus yang digunakan untuk mencari nilai TF-IDF yakni :

$$W_{t_j} = tf_{t_j} \times \log \log \left( \frac{1 + d}{1 + df_{(t)}} \right) + 1 \quad (2.1)$$

Keterangan :

$W_{t_j}$  = bobot kata atau term ( $t_j$ ) terhadap dokumen atau komentar ( $d_i$ )

$tf_{t_j}$  = jumlah kemunculan kata atau term ( $t_j$ ) dalam dokumen atau tweet ( $d_i$ )

$d$  = jumlah total dokumen

$df_{(t)}$  = jumlah dokumen yang mengandung kata ( $d_j$ )

Penambahan angka satu (1) untuk mencegah pembagian nol (0)

Tahap terakhir adalah melakukan normalisasi hasil TF-IDF sehingga memiliki hasil norma Euclidean menggunakan formula yang ditunjukkan pada persamaan.

$$v_{norm} = \frac{v}{\|v\|_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \quad (2.2)$$

Keterangan :

$v$  = hasil TF IDF setiap kata

### 2.2.5 Metode *Naïve bayes*

*Naïve bayes Classifier* adalah metode klasifikasi yang didasarkan pada teorema Bayes, sebuah teorema terkenal dalam ilmu probabilitas. Metode ini juga didukung oleh ilmu statistika, terutama dalam penggunaan data petunjuk untuk mendukung keputusan klasifikasi (Natalius, 2011).

Teorema Bayesian menghitung nilai *posterior probability*  $P(H|X)$  menggunakan probabilitas  $P(H)$ ,  $P(X)$ , dan  $P(X|H)$ . Di mana nilai  $X$  adalah data *testing* yang kelasnya belum diketahui. Nilai  $H$  adalah hipotesis data  $X$  yang merupakan suatu kelas yang lebih spesifik. Nilai  $P(X|H)$  disebut juga dengan *likelihood* adalah probabilitas hipotesis  $X$  berdasarkan kondisi  $H$ . Nilai  $P(H)$  disebut juga dengan *prior probability* adalah probabilitas hipotesis  $H$ . Dan nilai  $P(X)$  disebut juga dengan *predict prior probability* adalah probabilitas  $X$ . Persamaan teorema Bayes adalah :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.3)$$

Keterangan :

X: Data dengan class yang belum diketahui

H: Hipotesis data X merupakan suatu class spesifik

$P(H|X)$ : Probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)

$P(H)$ : Probabilitas hipotesis H (*prior probability*)

$P(X|H)$ : Probabilitas X berdasarkan kondisi pada hipotesis H (*likelihood*)

$P(X)$ : Probabilitas atas X (*predictor prior probability*)

*Multinomial naïve bayes* merupakan versi terbaru dari pengklasifikasi *Naïve bayes* yang ada, dan secara efektif menghitung frekuensi setiap kata. Sementara pada pengklasifikasi *Naïve bayes* frekuensi kata tidak berpengaruh banyak pada performa algoritma, pada *Multinomial naïve bayes* frekuensi teks memiliki dampak besar dalam mengkategorikan teks ke dalam berbagai kategori. Oleh karena itu, *Multinomial naïve bayes* dianggap paling efektif untuk klasifikasi teks (Surya et al., 2019). Tahapan klasifikasi *Multinomial naïve bayes* meliputi tiga langkah yaitu yang pertama menghitung probabilitas Prior, kemudian yang kedua menentukan *likelihood* atau frekuensi kemunculan setiap kata dalam dokumen atau data, dan yang terakhir menghitung nilai Posterior dengan mengalikan probabilitas Prior dengan *likelihood*.

Rumus untuk mencari nilai prior untuk setiap kelas ditunjukkan pada persamaan berikut

$$P(c_j) = \frac{\text{count}(c_j)}{N} \quad (2.4)$$

Keterangan :

$\text{count}(c_j)$  = banyaknya kelas ( $c_j$ ) yang ada pada dokumen latihan

$N$  = jumlah seluruh kelas dalam dokumen

Selanjutnya dalam menghitung nilai *likelihood*, untuk mencegah munculnya probabilitas 0, dilakukan penambahan nilai 1 pada setiap perhitungan. Teknik ini dikenal sebagai pendekatan parameter *laplace smoothing*, seperti yang ditunjukkan dalam formula pada persamaan.

$$P(w_i, c_j) = \frac{\text{count}(w_i, c_j) + 1}{\sum \text{count}(w_i, c_j) + |V|} \quad (2.5)$$

Keterangan :

$\text{count}(w_i, c_j)$  = jumlah kemunculan kata  $w$  pada kelas  $c$

$\sum \text{count}(w_i, c_j)$  = jumlah total kata dalam kelas  $c_j$

$|V|$  = jumlah kata unik atau token yang ada pada seluruh kelas

Rumus untuk mencari probabilitas posterior ditunjukkan pada persamaan

$$P(w_i) = P(c_j) \times P(w_j|c_j) \times \dots \times P(w_n|c_n) \quad (2.6)$$

Keterangan :

$P(w_i)$  = peluang kategori  $j$  jika terdapat kemunculan kata  $i$

$P(c_j)$  = probabilitas kemunculan nilai prior

$P(w_j|c_j)$  = peluang kata ke-  $i$  yang muncul pada kelas  $j$  atau *likelihood*

### 2.2.6 Confusion matrix

Tahap *evaluation* ini merupakan tahap evaluasi dengan melaksanakan interpretasi terhadap output dari data mining yang dihasilkan dalam tahapan sebelumnya. Evaluasi di sini bertujuan agar model yang sudah ditentukan dapat sesuai dengan tujuan yang ingin dipenuhi pada fase pertama (Ginantra et al., 2021). *Confusion matrix* adalah teknik yang digunakan untuk mengukur performa model dengan membandingkan prediksi dengan hasil yang aktual (Wahyu et al., 2023). Untuk memberikan gambaran performa yang lebih jelas, *confusion matrix* memungkinkan analisis dalam empat dimensi. Hal ini karena *confusion matrix* dapat menggambarkan prediksi model dan penerapan sebenarnya dari algoritma (Nugraha et al., 2022).

Table 1. 1 Confusion Matrix

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

TP (*True Positive*) untuk menggambarkan jumlah data aktual *positive* dan diprediksi juga *positive* (benar). Sedangkan, FP (*False Positive*) menggambarkan jumlah data aktual *negative* tetapi diprediksi sebagai *positive* oleh sistem (salah). FN (*False Negative*) untuk menggambarkan jumlah data aktual *positive* dengan yang terprediksi menjadi *negative*. TN (*True Negative*): menggambarkan jumlah data aktual *negative* dan diprediksi juga *negative* (Indriani & Syafrullah, 2022). Sedangkan untuk mengukur suatu performa dari klasifikasi dibutuhkan perhitungan *accuracy*, *precision* dan *recall* seperti berikut:

- *Accuracy*: *Accuracy*nya mencerminkan seberapa tepat sebuah sistem dalam mengklasifikasikan data, baik yang *positive* maupun *negative*. Secara sederhana, *accuracy* adalah rasio antara jumlah data yang diklasifikasikan dengan benar terhadap total data yang tersedia (Indriani & Syafrullah, 2022). Rumus yang digunakan untuk *accuracy* sebagai berikut:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.7)$$

- *Precision*: *Precision* menggambarkan seberapa tepat sistem menghasilkan jawaban yang sesuai dengan informasi yang diminta. Dengan kata lain, *precision* mengukur tingkat kesesuaian antara permintaan informasi dan jawaban yang diberikan oleh sistem (Indriani & Syafrullah, 2022). Rumus yang digunakan seperti berikut:

$$P = \frac{TP}{TP + FP} \quad (2.8)$$

- *Recall*: *Recall* adalah suatu pengujian yang bertujuan untuk mengetahui jumlah dokumen yang benar-benar diklasifikasikan sebagai *positive* (*true positive*) dari keseluruhan dokumen yang memang *positive*, termasuk yang tidak terdeteksi (*false negative*). Dengan kata lain, *recall* mengukur tingkat keberhasilan suatu sistem dalam menemukan kembali informasi yang benar (Indriani & Syafrullah, 2022). Rumus *recall* sebagai berikut:

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

- *F-measure*: *F-measure* merupakan rumus yang mengombinasikan *precision* dan *recall* untuk mendapatkan rata-rata harmonis dari hasil *true positive* terhadap prediksi *positive* dan *positive* yang sebenarnya (Zafran et al., 2020). Berikut rumus yang digunakan untuk *f-measure*:

$$F = 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}} \quad (2.10)$$

### 2.2.7 Evaluasi (Muhasabah)

Menurut para ulama evaluasi atau muhasabah adalah usaha yang dilakukan seorang hamba untuk mengevaluasi amal-amal yang telah dikerjakan setiap saat, apakah amal tersebut mengandung maksiat atau benar-benar untuk beribadah dan mengabdikan kepada Allah (Hasanah, 2018). Terdapat juga beberapa muhasabah diantaranya sebagai berikut:

#### a. Muhasabah Perilaku

Hal tersebut mendorong seseorang untuk melakukan introspeksi dalam tindakannya, baik dalam aspek ibadah maupun dalam relasi sosial. Hal ini sejalan dengan semangat untuk memperbaiki diri yang dijelaskan dalam firman Allah SWT dalam Al-Hasyr ayat 18

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍّ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

*"Wahai orang-orang yang beriman! Bertakwalah kepada Allah dan hendaklah setiap diri memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat); dan bertakwalah kepada Allah. Sungguh, Allah Maha Mengetahui terhadap apa yang kamu kerjakan."* (QS. Al-Hasyr: 18)

Pada ayat ini Allah mengingatkan orang beriman agar benar-benar bertakwa kepada Allah dan memperhatikan hari esok, akhirat (Kemenag, n.d.). ayat ini menegaskan pentingnya introspeksi dan perencanaan masa depan dengan tindakan yang telah dilakukan.

#### b. Muhasabah Proses

Muhasabah juga bisa diinterpretasikan sebagai upaya untuk mengamati, mengevaluasi dan menyikapi pendapat masyarakat, contohnya melalui analisis sentimen. Dalam hadist seperti berikut.

الْكَيْسُ مَنْ دَانَ نَفْسَهُ وَعَمِلَ لِمَا بَعْدَ الْمَوْتِ وَالْعَاجِزُ مَنْ أَتْبَعَ نَفْسَهُ هَوَاهَا وَتَمَتَّى عَلَى اللَّهِ

*"Orang yang cerdas adalah yang menghisab dirinya dan beramal untuk kehidupan setelah mati. Sedangkan orang yang lemah adalah yang mengikuti hawa nafsunya dan hanya berangan-angan kepada Allah."* (HR. At-Tirmidzi, No. 2459 – Hasan)

Hadits ini membangun paradigma evaluasi berkelanjutan (*continuous improvement*) dalam siklus pengembangan model (Al-Ghazali, Ihya Ulumuddin). Umar bin Khattab menafsirkan muhasabah sebagai tanggung jawab proaktif sebelum munculnya konsekuensi (Al-Bukhari, Al-Adab al-Mufrad) (Al-Ghazali, 2009).

#### c. Muhasabah Sosial

Hal tersebut menunjukkan pentingnya kemampuan mengukur dan menilai setiap perilaku. Dalam konteks digital, evaluasi yang diwujudkan dalam upaya menganalisis pola-pola opini masyarakat dalam bentuk teks. Hal tersebut diungkapkan pada hadist berikut:

مسند أحمد ٢٣٤٠٨: حَدَّثَنَا أَبُو الْيَمَانِ وَمُحَمَّدُ بْنُ مُصْعَبٍ قَالَا حَدَّثَنَا أَبُو بَكْرِ بْنُ عَبْدِ اللَّهِ عَنْ حَبِيبِ بْنِ غُبَيْدٍ قَالَ قَالَتْ عَائِشَةُ قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ الشُّؤْمُ سُوءُ الْخُلُقِ

Musnad Ahmad 23408: Telah menceritakan kepada kami Abu Al-Yaman, dan Muhammad bin Mushab keduanya berkata: telah menceritakan kepada kami Abu Bakar bin Abdullah, dari Habib bin Ubaid berkata: Aisyah berkata: Rasulullah shallallahu 'alaihi wa sallam bersabda: "Pesimis adalah akhlaq yang buruk" (Ahmad, n.d.). Dengan demikian, nilai muhasabah tidak hanya dapat diterapkan secara personal. tetapi juga secara kolektif melalui pemanfaatan teknologi untuk menyerap opini masyarakat. Kehidupan bermasyarakat juga dalam Islam diatur untuk tidak saling merugikan antar sesama dan mewujudkan kesejahteraan sosial, yang mana didefinisikan sebagai sebuah proses hukum syariat yang dieksplorasi secara mendetail untuk kemaslahatan atau kepentingan masyarakat (Yulianto, 2017).

#### d. Muhasabah Perbaikan

لَهُ مُعَقِّبَاتٌ مِّنْ بَيْنِ يَدَيْهِ وَمِنْ خَلْفِهِ يَحْفَظُونَهُ مِنْ أَمْرِ اللَّهِ إِنَّ اللَّهَ لَا يُغَيِّرُ مَا بِقَوْمٍ حَتَّىٰ يُغَيِّرُوا مَا بِأَنفُسِهِمْ ۗ وَإِذَا أَرَادَ اللَّهُ بِقَوْمٍ سُوءًا فَلَا مَرَدَّ لَهُ ۗ وَمَا لَهُمْ مِّنْ دُونِهِ مِنْ وَالٍ

(Baginya) manusia (ada malaikat-malaikat yang selalu mengikutinya bergiliran) para malaikat yang bertugas mengawasinya (di muka) di hadapannya (dan di belakangnya) dari belakangnya (mereka menjaganya atas perintah Allah) berdasarkan perintah Allah, dari gangguan jin dan makhluk-makhluk yang

lainnya. (Sesungguhnya Allah tidak mengubah keadaan sesuatu kaum) artinya Dia tidak mencabut dari mereka nikmat-Nya (sehingga mereka mengubah keadaan yang ada pada diri mereka sendiri) dari keadaan yang baik dengan melakukan perbuatan durhaka. (Dan apabila Allah menghendaki keburukan terhadap suatu kaum) yakni menimpakan azab (maka tak ada yang dapat menolaknya) dari siksaan-siksaan tersebut dan pula dari hal-hal lainnya yang telah dipastikan-Nya (dan sekali-kali tak ada bagi mereka) bagi orang-orang yang telah dikehendaki keburukan oleh Allah (selain Dia) selain Allah sendiri (seorang penolong pun) yang dapat mencegah datangnya azab Allah terhadap mereka. Huruf min di sini adalah zaidah (As-Suyuthi et al., 2017).

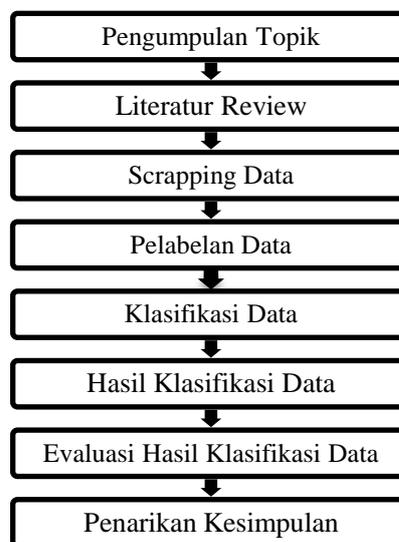
## BAB III METODE PENELITIAN

### 3.1 Jenis Penelitian

Jenis penelitian yang digunakan adalah pendekatan analisis sentimen berbasis teks. Data yang dikumpulkan dari google *Play store* yang akan diproses dan dianalisis menggunakan pemrosesan bahasa alami untuk mengidentifikasi sentimen *positive* dan *negative* dalam setiap komentar yang terkait dengan layanan aplikasi *iBI Library*. Pada tahap klasifikasi menggunakan metode *naïve bayes*.

### 3.2 Alur Penelitian

Beberapa tahapan yang dilakukan untuk penelitian analisis sentiment yaitu, menentukan topik dan judul penelitian, mencari teori yang relevan dengan penelitian, mengumpulkan penelitian terdahulu yang terkait, mengumpulkan informasi, penerapan model penelitian, evaluasi model, analisis dan interpretasi hasil, dan menarik kesimpulan



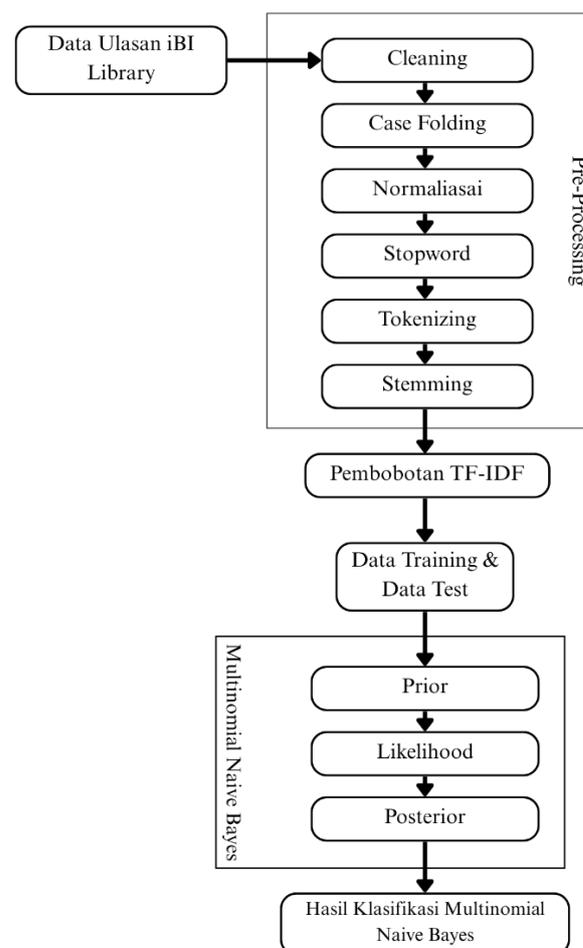
**Gambar 2. 1. Alur Penelitian**

Berdasarkan gambar diatas, berikut merupakan penjelasan dari tahapan diatas antara lain: Pertama, adalah pengumpulan topik yang relevan dengan minat dan bidang studi ilmu, Kedua, pencarian literatur review yang relevan tentang teori analisis sentimen, machine learnig yang digunakan dan sebagainya, mengumpulkan penelitian terdahulu yang terkait dengan penelitian ini, Ketiga, scrapping data yang akan dianalisis. Keempat, pelabelan yang dilakukan oleh para ahli di bidang bahasa

Indonesia. Kelima, pengklasifikasian data yang dilakukan seperti *cleaning* hingga klasifikasi. Keenam, hasil dari klasifikasi data seperti hasil klasifikasi data *training* dan data *test*. Keenam, evaluasi performa klasifikasi menggunakan confusion matriks seperti *accuracy*, *preisisi*, *recall* dan *f-measure* yang diperoleh dari klasifikasi tersebut, lalu diinterpretasikan hasil dari evaluasi tersebut. Ketujuh, penarikan kesimpulan dari penelitian dan memberikan saran pada penelitian tersebut

### 3.3 Desain Sistem dan Analisis Data

Desain sistem menjelaskan tentang desain sistem yang digunakan untuk mencapai tujuan penelitian ini, desain sistem mencakup beberapa aspek penting, seperti input data ulasan iBI Library, *pre-processing*, pembobotan TF-IDF, klasifikasi *multinomial naïve bayes*, hasil klasifikasi *multinomial naïve bayes*.



Gambar 2. 2. Desain Sistem

### 3.2.1 Labeling

Manual labeling atau anotasi terhadap data set berfungsi untuk menentukan klasifikasi teks saran, seperti label *positive*, dan *negative*. Proses pelabelan manual ini sangat penting dalam pengembangan model klasifikasi karena menyediakan data pelatihan yang telah diberi label dengan benar untuk melatih model. Dengan adanya data yang dilabeli dengan tepat, model klasifikasi dapat belajar mengenali pola dan karakteristik teks yang terkait dengan setiap kategori atau label yang ditentukan. Sebagai contoh labeling secara manual pada table berikut :

**Table 3. 1. Contoh Data Labeling**

No Data	Data	Label
D1	bukunya ada bisa dipinjam tapi kenapa gagal download dan file tidak ditemukan terus padahal memori handphone saya tidak ada masalah	<i>Negative</i>
D2	saya tidak bisa mengunduh buku padahal sudah gabung mohon diperbaiki	<i>Positive</i>
D3	perasaan setiap hari harus di update terus mana rak bacaan jadi ilang semua	<i>Negative</i>
D4	setiap mau baca buku selalu download gagal file tidak tersedia	<i>Positive</i>
D5	tiap klik baca selalu keluar padahal sinyal kuat sistemnya tolong diperbaiki	<i>Positive</i>
D6	rusak tidak bisa baca buku yang sudah di download haduh gimana ini tombol pinjamnya memuat doang padahal jaringan kenceng aplikasinya masih banyak bug	<i>Negative</i>
D7	tidak ada satupun buku yang bisa dibaca karena ketika di klik itu semuanya berhenti di laman detail dan loading sekian lama lalu cancel dengan sendirinya mohon untuk diperbaiki fiturnya	<i>Positive</i>
D8	seperti aplikasinya perlu beberapa perbaikan karena saya sering mengalami permasalahan seperti sering logout sendiri tidak bisa meminjam buku buku yang sudah disimpan tidak bisa dibuka tolong perhatiannya	<i>Positive</i>
D9	setiap mau pinjam buku pasti tulisannya gagal melakukan proses enkripsi konten padahal stok bukunya masih banyak tolong perbaiki dong	<i>Positive</i>
D10	maaf kasih rating satu awalnya sih fine aja tapi kenapa makin lama bukunya tidak bisa dibuka padahal sudah di unduh dan koneksi jaringan internet bagus mohon untuk perbaikannya <i>iBI Library</i>	<i>Positive</i>

### 3.2.2 Scrapping data

Penelitian ini menggunakan data berupa teks yang diperoleh melalui komentar google *Play store*. Data yang dibutuhkan tentang *iBI Library* yang menggunakan Bahasa Indonesia. Data ini dikumpulkan dari 31 Oktober 2017 sampai dengan 18 Oktober 2024. Pengumpulan data komentar tentang layanan *iBI Library* menggunakan google colab dengan bahasa pemrograman python dan disimpan dalam format csv

### 3.2.3 Pre-processing

Pembersihan terhadap data dari berbagai atribut yang kurang berdampak pada proses klasifikasi akan dihilangkan agar mesin dapat memahami teks dengan *accuracy* tinggi. Data mentah yang diproses pada tahap ini diharapkan menjadi dokumen yang memiliki kualitas tinggi dalam proses klasifikasi. Berikut tahap *pre-processing* memiliki beberapa urutan dalam pembersihan data, yaitu mulai dari *cleaning*, *case folding*, normalisasi *tokenizing*, dan *stemming*

#### a. *Cleaning*

Proses pembersihan berguna untuk menghapus semua karakter selain alfabet a-z dalam data tweet, seperti nama pengguna, @ (*mention*), # (*hashtag*), menghilangkan URL atau tautan, menghapus karakter tunggal yang berdiri sendiri seperti "d", serta menghilangkan spasi yang berlebihan (Indriani & Syafrullah, 2022). Maka dari itu, data yang diperoleh masih dalam bentuk mentah perlu dilakukan pembersihan seperti berikut :

**Table 3. 2. Contoh Proses *Cleaning***

Data Mentah	Bukunya ada bisa dipinjam tapi knp gagal download dan file tdk ditemukan trs Pdhl memori hp saya gak ada masalah
	Saya tidak bisa mengunduh buku padahal sudah gabung Mohon diperbaiki
	rasa2 tiap hari harus diupdate terus mana rak bacaan jadi ilang semua
	Setiap mau baca buku selalu download gagal file tidak tersedia
	Tiap klik baca selalu keluar padahal sinyal kuat System nya tolong diperbaiki
	rusak ga bisa baca buku yang sudah di download hadeuuh gmna ini tombol pinjam nya memuat doang padahal jaringan kenceng apk nya masih banyak bug
	tidak ada satupun buku yang bisa dibaca karena ketika di klik itu semuanya berhenti di laman detail dan loading sekian lama lalu cancel dengan sendirinya Mohon untuk diperbaiki fiturnya
	Seperti aplikasinya perlu beberapa perbaikan karna saya sering mengalami permasalahan seperti sering logout sendiri tidak bisa meminjam buku buku yang sudah disimpan tidak bisa dibuka Tolong perhatiannya

	Setiap mau pinjam buku pasti tulisannya Gagal melakukan proses enkripsi konten padahal stok bukunya masih banyak Tolong perbaiki dong
	Maaf kasih rating satu awalnya si fine aja tapi kenapa makin lama bukunya tidak bisa dibuka padahal sudah di unduh dan koneksi jaringan internet bagus Mohon untuk perbaikannya iBl
<i>Cleaning</i>	Bukunya ada bisa dipinjam tapi knp gagal download dan file tdk ditemukan trs Pdhl memori hp saya gak ada masalah
	Saya tidak bisa mengunduh buku padahal sudah gabung Mohon diperbaiki
	rasa2 tiap hari harus diupdate terus mana rak bacaan jadi ilang semua
	Setiap mau baca buku selalu download gagal file tidak tersedia
	Tiap klik baca selalu keluar padahal sinyal kuat System nya tolong diperbaiki
	rusak ga bisa baca buku yang sudah di download hadeuuh gmna ini tombol pinjam nya memuat doang padahal jaringan kenceng apk nya masih banyak bug
	tidak ada satupun buku yang bisa dibaca karena ketika di klik itu semuanya berhenti di laman detail dan loading sekian lama lalu cancel dengan sendirinya Mohon untuk diperbaiki fiturnya
	Seperti aplikasinya perlu beberapa perbaikan karna saya sering mengalami permasalahan seperti sering logout sendiri tidak bisa meminjam buku buku yang sudah disimpan tidak bisa dibuka Tolong perhatiannya
	Setiap mau pinjam buku pasti tulisannya Gagal melakukan proses enkripsi konten padahal stok bukunya masih banyak Tolong perbaiki dong
	Maaf kasih rating satu awalnya si fine aja tapi kenapa makin lama bukunya tidak bisa dibuka padahal sudah di unduh dan koneksi jaringan internet bagus Mohon untuk perbaikannya iBl

b. *Case folding*

Meneliti ukuran tiap karakter dari awal hingga akhir teks. Jika terdapat karakter yang menggunakan huruf besar (*uppercase*), maka karakter tersebut akan diubah menjadi huruf kecil (*lowercase*) (Ratnawati, 2018).

**Table 3. 3. Contoh Proses Case folding**

<i>Cleaning</i>	Bukunya ada bisa dipinjam tapi knp gagal download dan file tdk ditemukan trs Pdhl memori hp saya gak ada masalah
	Saya tidak bisa mengunduh buku padahal sudah gabung Mohon diperbaiki
	rasa2 tiap hari harus diupdate terus mana rak bacaan jadi ilang semua
	Setiap mau baca buku selalu download gagal file tidak tersedia
	Tiap klik baca selalu keluar padahal sinyal kuat System nya tolong diperbaiki
	rusak ga bisa baca buku yang sudah di download hadeuuh gmna ini tombol pinjam nya memuat doang padahal jaringan kenceng apk nya masih banyak bug
	tidak ada satupun buku yang bisa dibaca karena ketika di klik itu semuanya berhenti di laman detail dan loading sekian lama lalu cancel dengan sendirinya Mohon untuk diperbaiki fiturnya
	Seperti aplikasinya perlu beberapa perbaikan karna saya sering mengalami permasalahan seperti sering logout sendiri tidak bisa meminjam buku buku yang sudah disimpan tidak bisa dibuka Tolong perhatiannya
	Setiap mau pinjam buku pasti tulisannya Gagal melakukan proses enkripsi konten padahal stok bukunya masih banyak Tolong perbaik dong
	Maaf kasih rating satu awalnya si fine aja tapi kenapa makin lama bukunya tidak bisa dibuka padahal sudah di unduh dan koneksi jaringan internet bagus Mohon untuk perbaikannya iBl
<i>Case folding</i>	bukunya ada bisa dipinjam tapi knp gagal download dan file tdk ditemukan trs pdhl memori hp saya gak ada masalah
	saya tidak bisa mengunduh buku padahal sudah gabung mohon diperbaiki
	rasa2 tiap hari harus diupdate terus mana rak bacaan jadi ilang semua
	setiap mau baca buku selalu download gagal file tidak tersedia
	tiap klik baca selalu keluar padahal sinyal kuat system nya tolong diperbaiki
	rusak ga bisa baca buku yang sudah di download hadeuuh gmna ini tombol pinjam nya memuat doang padahal jaringan kenceng apk nya masih banyak bug

	tidak ada satupun buku yang bisa dibaca karena ketika di klik itu semuanya berhenti di laman detail dan loading sekian lama lalu cancel dengan sendirinya mohon untuk diperbaiki fiturnya
	seperti aplikasinya perlu beberapa perbaikan karna saya sering mengalami permasalahan seperti sering logout sendiri tidak bisa meminjam buku buku yang sudah disimpan tidak bisa dibuka tolong perhatiannya
	setiap mau pinjam buku pasti tulisannya gagal melakukan proses enkripsi konten padahal stok bukunya masih banyak tolong perbaiki dong
	maaf kasih rating satu awalnya si fine aja tapi kenapa makin lama bukunya tidak bisa dibuka padahal sudah di unduh dan koneksi jaringan internet bagus mohon untuk perbaikannya ibl

### c. Normalisasi

Proses ini untuk mengganti teks dalam sebuah dokumen dari kata yang tidak baku atau singkatan bahasa slang menjadi kata yang memiliki arti jelas. Misalnya, "yg" diubah menjadi "yang", "tdk" diubah menjadi "tidak", dan seterusnya (Ashari et al., 2020). karena banyak dari ulasan menggunakan bahasa yang bervariasi tapi memiliki arti yang sama.

**Table 3. 4. Contoh Proses Normalisasi**

<i>Case folding</i>	bukunya ada bisa dipinjam tapi knp gagal download dan file tdk ditemukan trs pdhl memori hp saya gak ada masalah
	saya tidak bisa mengunduh buku padahal sudah gabung mohon diperbaiki
	rasa2 tiap hari harus diupdate terus mana rak bacaan jadi ilang semua
	setiap mau baca buku selalu download gagal file tidak tersedia
	tiap klik baca selalu keluar padahal sinyal kuat system nya tolong diperbaiki
	rusak ga bisa baca buku yang sudah di download hadeuuh gmna ini tombol pinjam nya memuat doang padahal jaringan kenceng apk nya masih banyak bug
	tidak ada satupun buku yang bisa dibaca karena ketika di klik itu semuanya berhenti di laman detail dan loading sekian lama lalu cancel dengan sendirinya mohon untuk diperbaiki fiturnya

	seperti aplikasinya perlu beberapa perbaikan karna saya sering mengalami permasalahan seperti sering logout sendiri tidak bisa meminjam buku buku yang sudah disimpan tidak bisa dibuka tolong perhatiannya
	setiap mau pinjam buku pasti tulisannya gagal melakukan proses enkripsi konten padahal stok bukunya masih banyak tolong perbaiki dong
	maaf kasih rating satu awalnya si fine aja tapi kenapa makin lama bukunya tidak bisa dibuka padahal sudah di unduh dan koneksi jaringan internet bagus mohon untuk perbaikannya ibl
Normalisasi	bukunya ada bisa dipinjam tapi kenapa gagal download dan file tidak ditemukan terus padahal memori handphone saya tidak ada masalah
	saya tidak bisa mengunduh buku padahal sudah gabung mohon diperbaiki
	perasaan tiap hari harus diupdate terus mana rak bacaan jadi ilang semua
	setiap mau baca buku selalu download gagal file tidak tersedia
	tiap klik baca selalu keluar padahal sinyal kuat sistemnya tolong diperbaiki
	rusak tidak bisa baca buku yang sudah di download haduh gimana ini tombol pinjamnya memuat doang padahal jaringan kenceng aplikasinya masih banyak bug
	tidak ada satupun buku yang bisa dibaca karena ketika di klik itu semuanya berhenti di laman detail dan loading sekian lama lalu cancel dengan sendirinya mohon untuk diperbaiki fiturnya
	seperti aplikasinya perlu beberapa perbaikan karena saya sering mengalami permasalahan seperti sering logout sendiri tidak bisa meminjam buku buku yang sudah disimpan tidak bisa dibuka tolong perhatiannya
	setiap mau pinjam buku pasti tulisannya gagal melakukan proses enkripsi konten padahal stok bukunya masih banyak tolong perbaiki dong
maaf kasih rating satu awalnya sih fine aja tapi kenapa makin lama bukunya tidak bisa dibuka padahal sudah di unduh dan koneksi jaringan internet bagus mohon untuk perbaikannya ibl	

d. *Stopword*

Proses *stopword* ini berguna untuk menghapus kata-kata yang tidak deskriptif atau kata penghubung yang memiliki makna kurang penting dalam dokumen. kata-kata yang dianggap kurang penting adalah “yang”, “dan”, “dari”, dan sebagainya (Ashari et al., 2020). Proses *stopword* ini menggunakan *library* sastrawi karena dalam teks ini menggunakan bahasa Indonesia.

**Table 3. 5. Contoh Proses *Stopword***

Normalisasi	bukunya ada bisa dipinjam tapi kenapa gagal download dan file tidak ditemukan terus padahal memori handphone saya tidak ada masalah
	saya tidak bisa mengunduh buku padahal sudah gabung mohon diperbaiki
	perasaan setiap hari harus diupdate terus mana rak bacaan jadi hilang semua
	setiap mau baca buku selalu download gagal file tidak tersedia
	tiap klik baca selalu keluar padahal sinyal kuat sistemnya tolong diperbaiki
	rusak tidak bisa baca buku yang sudah di download haduh gimana ini tombol pinjamnya memuat doang padahal jaringan kenceng aplikasinya masih banyak bug
	tidak ada satupun buku yang bisa dibaca karena ketika di klik itu semuanya berhenti di laman detail dan loading sekian lama lalu cancel dengan sendirinya mohon untuk diperbaiki fiturnya
	seperti aplikasinya perlu beberapa perbaikan karena saya sering mengalami permasalahan seperti sering logout sendiri tidak bisa meminjam buku buku yang sudah disimpan tidak bisa dibuka tolong perhatiannya
	setiap mau pinjam buku pasti tulisannya gagal melakukan proses enkripsi konten padahal stok bukunya masih banyak tolong perbaiki dong
	maaf kasih rating satu awalnya sih fine aja tapi kenapa makin lama bukunya tidak bisa dibuka padahal sudah di unduh dan koneksi jaringan internet bagus mohon untuk perbaikannya ibl
<i>Stopword</i>	bukunya ada dipinjam kenapa gagal download file tidak ditemukan memori handphone tidak ada masalah
	tidak mengunduh buku sudah gabung diperbaiki
	perasaan hari diupdate rak bacaan hilang semua

	baca buku download gagal file tidak tersedia
	klik baca keluar sinyal kuat sistemnya tolong diperbaiki
	rusak baca buku download tombol pinjamnya jaringan kenceng aplikasinya bug
	tidak ada buku dibaca karena ketika diklik semuanya berhenti laman detail memuat membatalkan sendirinya diperbaiki fiturnya
	aplikasinya beberapa perbaikan karena permasalahan keluar sendiri tidak meminjam buku buku disimpan tidak dibuka tolong perhatiannya
	pinjam buku tulisannya gagal proses enkripsi konten stok bukunya tolong perbaiki
	maaf kasih rating satu awalnya baik kenapa lama bukunya dibuka unduh koneksi jaringan internet bagus perbaikannya

e. *Tokenizing*

*Tokenizing* merupakan proses mendetail untuk memecah teks dari suatu kalimat atau paragraf menjadi segmen-segmen tertentu yang lebih kecil dan terstruktur (Hidayah & Dodiman, 2024). Dengan memecah teks tersebut dapat mengidentifikasi dan mengelompokan setiap kata secara sistematis, sehingga dapat dianalisis dengan *accuracy* yang tinggi.

**Table 3. 6. Proses *Tokenizing***

<i>Stopword</i>	bukunya ada dipinjam kenapa gagal download file tidak ditemukan memori handphone tidak ada masalah
	tidak mengunduh buku sudah gabung diperbaiki
	perasaan hari diupdate rak bacaan hilang semua
	baca buku download gagal file tidak tersedia
	klik baca keluar sinyal kuat sistemnya tolong diperbaiki
	rusak baca buku download tombol pinjamnya jaringan kenceng aplikasinya bug
	tidak ada buku dibaca karena ketika diklik semuanya berhenti laman detail memuat membatalkan sendirinya diperbaiki fiturnya
	aplikasinya beberapa perbaikan karena permasalahan keluar sendiri tidak meminjam buku buku disimpan tidak dibuka tolong perhatiannya
	pinjam buku tulisannya gagal proses enkripsi konten stok bukunya tolong perbaiki
	maaf kasih rating satu awalnya baik kenapa lama bukunya dibuka unduh koneksi jaringan internet bagus perbaikannya

<i>Tokenizing</i>	[‘bukuny’, ‘ada’, ‘dipinjam’, ‘kenapa’, ‘gagal’, ‘download’, ‘file’, ‘tidak’, ‘ditemukan’, ‘memori’, ‘handphone’, ‘tidak’, ‘ada’, ‘masalah’]
	[‘tidak’, ‘mengunduh’, ‘buku’, ‘sudah’, ‘gabung’, ‘diperbaiki’]
	[‘perasaan’, ‘hari’, ‘diupdate’, ‘rak’, ‘bacaan’, ‘hilang’, ‘semua’]
	[‘baca’, ‘buku’, ‘download’, ‘gagal’, ‘file’, ‘tidak’, ‘tersedia’,]
	[‘klik’, ‘baca’, ‘keluar’, ‘sinyal’, ‘kuat’, ‘sistemnya’, ‘tolong’, ‘diperbaiki’]
	[‘rusak’, ‘baca’, ‘buku’, ‘download’, ‘tombol’, ‘pinjamnya’, ‘memuat’, ‘jaringan’, ‘kenceng’, ‘aplikasinya’, ‘bug’]
	[‘tidak’, ‘ada’, ‘buku’, ‘dibaca’, ‘diklik’, ‘semuanya’, ‘berhenti’, ‘laman’, ‘memuat’, ‘membatalkan’, ‘sendirinya’, ‘diperbaiki’, ‘fiturnya’]
	[‘aplikasinya’, ‘beberapa’, ‘perbaikan’, ‘karena’, ‘permasalahan’, ‘keluar’, ‘sendiri’, ‘tidak’, ‘meminjam’, ‘buku’, ‘buku’, ‘disimpan’, ‘tidak’, ‘dibuka’, ‘tolong’, ‘perhatiannya’,]
	[‘pinjam’, ‘buku’, ‘tulisan’, ‘gagal’, ‘proses’, ‘enkripsi’, ‘konten’, ‘stok’, ‘bukunya’, ‘tolong’, ‘perbaiki’]
[‘maaf’, ‘kasih’, ‘rating’, ‘satu’, ‘awalnya’, ‘baik’, ‘kenapa’, ‘lama’, ‘bukunya’, ‘dibuka’, ‘unduh’, ‘koneksi’, ‘jaringan’, ‘internet’, ‘bagus’, ‘perbaikannya’]	

#### f. *Stemming*

*Stemming* adalah teknik dalam pengolahan teks yang bertujuan untuk menghilangkan imbuhan dari setiap kata dan mengubahnya menjadi bentuk dasar atau akar katanya (Ashari et al., 2020). Dapat disimpulkan proses ini untuk mengubah kata-kata yang telah diberi imbuhan menjadi kata dasarnya, namun memiliki arti dasar yang sama, proses stemming ini menggunakan *library* sastrawi dan *snowball stemmer*.

**Table 3. 7. Contoh Proses *Stemming***

<i>Tokenizing</i>	[‘bukuny’, ‘ada’, ‘dipinjam’, ‘kenapa’, ‘gagal’, ‘download’, ‘file’, ‘tidak’, ‘ditemukan’, ‘memori’, ‘handphone’, ‘tidak’, ‘ada’, ‘masalah’]
	[‘tidak’, ‘mengunduh’, ‘buku’, ‘sudah’, ‘gabung’, ‘diperbaiki’]
	[‘perasaan’, ‘hari’, ‘diupdate’, ‘rak’, ‘bacaan’, ‘hilang’, ‘semua’]
	[‘baca’, ‘buku’, ‘download’, ‘gagal’, ‘file’, ‘tidak’, ‘tersedia’,]
	[‘klik’, ‘baca’, ‘keluar’, ‘sinyal’, ‘kuat’, ‘sistemnya’, ‘tolong’, ‘diperbaiki’]

	['rusak', 'baca', 'buku', 'download', 'tombol', 'pinjamnya', 'memuat', 'jaringan', 'kenceng', 'aplikasinya', 'bug']
	['tidak', 'ada', 'buku', 'dibaca', 'diklik', 'semuanya', 'berhenti', 'laman', 'memuat', 'membatalkan', 'sendirinya', 'diperbaiki', 'fiturnya']
	['aplikasinya', 'beberapa', 'perbaikan', 'karena', 'permasalahan', 'keluar', 'sendiri', 'tidak', 'meminjam', 'buku', 'buku', 'disimpan', 'tidak', 'dibuka', 'tolong', 'perhatiannya,']
	['pinjam', 'buku', 'tulisan', 'gagal', 'proses', 'enkripsi', 'konten', 'stok', 'bukunya', 'tolong', 'perbaiki']
	['maaf', 'kasih', 'rating', 'satu', 'awalnya', 'baik', 'kenapa', 'lama', 'bukunya', 'dibuka', 'unduh', 'koneksi', 'jaringan', 'internet', 'bagus', 'perbaikannya']
<i>Stemming</i>	['buku', 'ada', 'pinjam', 'kenapa', 'gagal', 'download', 'file', 'tidak', 'temu', 'memori', 'handphone', 'tidak', 'ada', 'masalah']
	['tidak', 'unduh', 'buku', 'sudah', 'gabung', 'perbaiki']
	['rasa', 'hari', 'update', 'rak', 'baca', 'hilang', 'semua']
	['baca', 'buku', 'download', 'gagal', 'file', 'tidak', 'tersedia,']
	['klik', 'baca', 'keluar', 'sinyal', 'kuat', 'sistem', 'tolong', 'perbaiki']
	['rusak', 'baca', 'buku', 'download', 'tombol', 'pinjam', 'memuat', 'jaringan', 'kenceng', 'aplikasi', 'bug']
	['tidak', 'ada', 'buku', 'baca', 'klik', 'semua', 'henti', 'laman', 'memuat', 'batal', 'sendiri', 'perbaiki', 'fitur']
	['aplikasi', 'berapa', 'perbaiki', 'karena', 'masalah', 'keluar', 'sendiri', 'tidak', 'pinjam', 'buku', 'buku', 'simpan', 'tidak', 'buka', 'tolong', 'hati,']
	['pinjam', 'buku', 'tulisan', 'gagal', 'proses', 'enkripsi', 'konten', 'stok', 'bukun', 'tolong', 'perbaiki']
['maaf', 'kasih', 'rating', 'satu', 'awal', 'baik', 'kenapa', 'lama', 'buku', 'buka', 'unduh', 'koneksi', 'jaringan', 'internet', 'bagus', 'perbaiki']	

### 3.2.4 Pembobotan TF-IDF

Untuk langkah selanjutnya menghitung frekuensi kata yang muncul dalam dokumen, karena pada setiap dokumen pasti memiliki beberapa kata yang sama. Berikut hasil pembobotan TF\_IDF. Hal untuk awal yang dilakukan adalah menghitung jumlah *term* frekuensi pada setiap dokumen seperti table berikut:

Table 3. 8. Contoh Proses Term Frekuensi

Term	TF									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
ada	2	0	0	0	0	0	1	0	0	0
aplikasi	0	0	0	0	0	1	0	1	0	0
awal	0	0	0	0	0	0	0	0	0	1
baca	0	0	1	1	1	1	1	0	0	0
bagus	0	0	0	0	0	0	0	0	0	1
baik	0	0	0	0	0	0	0	0	0	1
batal	0	0	0	0	0	0	1	0	0	0
berapa	0	0	0	0	0	0	0	1	0	0
bug	0	0	0	0	0	1	0	0	0	0
buka	0	0	0	0	0	0	0	1	0	1
buku	1	1	0	1	0	1	1	2	2	1
cepat	0	0	0	0	0	1	0	0	0	0
enkripsi	0	0	0	0	0	0	0	0	1	0
file	1	0	0	1	0	0	0	0	0	0
fitur	0	0	0	0	0	0	1	0	0	0
gabung	0	1	0	0	0	0	0	0	0	0
gagal	1	0	0	1	0	0	0	0	1	0
handphone	1	0	0	0	0	0	0	0	0	0
hari	0	0	1	0	0	0	0	0	0	0
hati	0	0	0	0	0	0	0	1	0	0
henti	0	0	0	0	0	0	1	0	0	0
hilang	0	0	1	0	0	0	0	0	0	0
internet	0	0	0	0	0	0	0	0	0	1
jarangan	0	0	0	0	0	1	0	0	0	1
karena	0	0	0	0	0	0	0	1	0	0
kasih	0	0	0	0	0	0	0	0	0	1
keluar	0	0	0	0	1	0	0	1	0	0
kenapa	1	0	0	0	0	0	0	0	0	1
klik	0	0	0	0	1	0	1	0	0	0
koneksi	0	0	0	0	0	0	0	0	0	1
konten	0	0	0	0	0	0	0	0	1	0
kuat	0	0	0	0	1	0	0	0	0	0
lama	0	0	0	0	0	0	0	0	0	1
laman	0	0	0	0	0	0	1	0	0	0
maaf	0	0	0	0	0	0	0	0	0	1
masalah	1	0	0	0	0	0	0	1	0	0
memori	1	0	0	0	0	0	0	0	0	0
memuat	0	0	0	0	0	1	1	0	0	0
perbaiki	0	1	0	0	1	0	1	1	1	1





Term	TF Normalisasi									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
semua	0	0	0,143	0	0	0	0,077	0	0	0
sendiri	0	0	0	0	0	0	0,077	0,063	0	0
simpan	0	0	0	0	0	0	0	0,063	0	0
sinyal	0	0	0	0	0,125	0	0	0	0	0
sistem	0	0	0	0	0,125	0	0	0	0	0
stok	0	0	0	0	0	0	0	0	0,091	0
sudah	0	0,167	0	0	0	0	0	0	0	0
temu	0,071	0	0	0	0	0	0	0	0	0
tersedia	0	0	0	0,143	0	0	0	0	0	0
tidak	0,143	0,167	0	0,143	0	0	0,077	0,125	0	0
tolong	0	0	0	0	0,125	0	0	0,063	0,091	0
tombol	0	0	0	0	0	0,091	0	0	0	0
tulisan	0	0	0	0	0	0	0	0	0,091	0
unduh	0,071	0,167	0	0,143	0	0,091	0	0	0	0,063
update	0	0	0,143	0	0	0	0	0	0	0
proses	0	0	0	0	0	0	0	0	0,091	0

Sebagai contoh perhitungan TF Normalisasi :

$$TF \text{ Normalisasi } (Ada_1) = \frac{\text{count}(2)}{12} \quad (3.1)$$

Kemudian, dilakukan penjumlahan DF atau *Document Frequency* dimana menjumlahkan term pada setiap dokumen. Sedangkan *Invers Document Frequency* (IDF) adalah frekuensi munculnya kata atau term tertentu yang ada pada suatu dokumen (Ashari et al., 2020). Seperti pada table berikut:

**Table 3. 10. Contoh Proses *Document Frequency* dan *Invers Document Frequency***

Term	DF	IDF
ada	2	0,699
aplikasi	2	0,699
awal	1	1
baca	5	0,301
bagus	1	1
baik	1	1
batal	1	1
berapa	1	1
bug	1	1
buka	2	0,699

Term	DF	IDF
buku	8	0,097
cepat	1	1
enkripsi	1	1
file	2	0,699
fitur	1	1
gabung	1	1
gagal	3	0,523
handphone	1	1
hari	1	1
hati	1	1
henti	1	1
hilang	1	1
internet	1	1
jaringan	2	0,699
karena	1	1
kasih	1	1
keluar	2	0,699
kenapa	2	0,699
klik	2	0,699
koneksi	1	1
konten	1	1
kuat	1	1
lama	1	1
laman	1	1
maaf	1	1
masalah	2	0,699
memori	1	1
memuat	2	0,699
perbaiki	6	0,222
pinjam	4	0,398
rak	1	1
rasa	1	1
rating	1	1
rusak	1	1
satu	1	1
semua	2	0,699
sendiri	2	0,699
simpan	1	1
sinyal	1	1
sistem	1	1

Term	DF	IDF
stok	1	1
sudah	1	1
temu	1	1
tersedia	1	1
tidak	5	0,301
tolong	3	0,523
tombol	1	1
tulisan	1	1
unduh	5	0,301
update	1	1
proses	1	1

Sebagai contoh perhitungan IDF yang dilakukan berdasarkan hasil table diatas:

$$IDF ((Ada)) = \log \frac{(10)}{2} = 0,699 \quad (3.2)$$

Dengan menghitung *Term Frequency* pada setiap data dokumen. Dan menghitung nilai *Inverse Document Frequency* (IDF). Selanjutnya menghitung nilai bobot TF-IDF dengan mengkalikan nilai TF Normalisasi dengan IDF (Ashari et al., 2020). Berikut perhitungan TF-IDF sebagai berikut:

**Table 3. 11. Contoh Proses TF-IDF**

Term	TF.IDF									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Ada	0,1	0	0	0	0	0	0,054	0	0	0
Aplikasi	0	0	0	0	0	0,064	0	0,044	0	0
Awal	0	0	0	0	0	0	0	0	0	0,063
Baca	0	0	0,043	0,043	0,038	0,027	0,023	0	0	0
Bagus	0	0	0	0	0	0	0	0	0	0,063
Baik	0	0	0	0	0	0	0	0	0	0,063
Batal	0	0	0	0	0	0	0,077	0	0	0
Berapa	0	0	0	0	0	0	0	0,063	0	0
Bug	0	0	0	0	0	0,091	0	0	0	0
Buka	0	0	0	0	0	0	0	0,044	0	0,044
Buku	0,007	0,016	0	0,014	0	0,009	0,007	0,012	0,018	0,006
Cepat	0	0	0	0	0	0,091	0	0	0	0
Enkripsi	0	0	0	0	0	0	0	0	0,091	0
File	0,05	0	0	0,1	0	0	0	0	0	0

Term	TF.IDF									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Fitur	0	0	0	0	0	0	0,077	0	0	0
Gabung	0	0,167	0	0	0	0	0	0	0	0
Gagal	0,037	0	0	0,075	0	0	0	0	0,048	0
Handphone	0,071	0	0	0	0	0	0	0	0	0
Hari	0	0	0,143	0	0	0	0	0	0	0
Hati	0	0	0	0	0	0	0	0,063	0	0
Henti	0	0	0	0	0	0	0,077	0	0	0
Hilang	0	0	0,143	0	0	0	0	0	0	0
Internet	0	0	0	0	0	0	0	0	0	0,063
Jaringan	0	0	0	0	0	0,064	0	0	0	0,044
Karena	0	0	0	0	0	0	0	0,063	0	0
kasih	0	0	0	0	0	0	0	0	0	0,063
Keluar	0	0	0	0	0,087	0	0	0,044	0	0
Kenapa	0,05	0	0	0	0	0	0	0	0	0,044
Klik	0	0	0	0	0,087	0	0,054	0	0	0
Koneksi	0	0	0	0	0	0	0	0	0	0,063
Konten	0	0	0	0	0	0	0	0	0,091	0
Kuat	0	0	0	0	0,125	0	0	0	0	0
Lama	0	0	0	0	0	0	0	0	0	0,063
Laman	0	0	0	0	0	0	0,077	0	0	0
maaf	0	0	0	0	0	0	0	0	0	0,063
Masalah	0,05	0	0	0	0	0	0	0,044	0	0
Memori	0,071	0	0	0	0	0	0	0	0	0
Memuat	0	0	0	0	0	0,064	0,054	0	0	0
Perbaiki	0	0,037	0	0	0,028	0	0,017	0,014	0,02	0,014
Pinjam	0,028	0	0	0	0	0,036	0	0,025	0,036	0
Rak	0	0	0,143	0	0	0	0	0	0	0
Rasa	0	0	0,143	0	0	0	0	0	0	0
Rating	0	0	0	0	0	0	0	0	0	0,063
Rusak	0	0	0	0	0	0,091	0	0	0	0
Satu	0	0	0	0	0	0	0	0	0	0,063
Semua	0	0	0,1	0	0	0	0,054	0	0	0
Sendiri	0	0	0	0	0	0	0,054	0,044	0	0
Simpan	0	0	0	0	0	0	0	0,063	0	0
Sinyal	0	0	0	0	0,125	0	0	0	0	0
Sistem	0	0	0	0	0,125	0	0	0	0	0
Stok	0	0	0	0	0	0	0	0	0,091	0
Sudah	0	0,167	0	0	0	0	0	0	0	0
Temu	0,071	0	0	0	0	0	0	0	0	0
Tersedia	0	0	0	0,143	0	0	0	0	0	0
Tidak	0,043	0,05	0	0,043	0	0	0,023	0,038	0	0

Term	TF.IDF									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Tolong	0	0	0	0	0,065	0	0	0,033	0,048	0
Tombol	0	0	0	0	0	0,091	0	0	0	0
Tulisan	0	0	0	0	0	0	0	0	0,091	0
Unduh	0,021	0,05	0	0,043	0	0,027	0	0	0	0,019
Update	0	0	0,143	0	0	0	0	0	0	0
Proses	0	0	0	0	0	0	0	0	0,091	0

Sebagai contoh perhitungan TF-IDF sebagai berikut:

$$\begin{aligned}
 TF - IDF (Ada) &= TF \text{ Normalisasi} \times IDF & (3.3) \\
 &= 0,143 \times 0,699 \\
 &= 0,1
 \end{aligned}$$

### 3.2.5 Klasifikasi *Multinomial Naïve Bayes*

*Multinomial naïve bayes* bisa dianggap sebagai versi yang lebih baru dari klasifikasi *Naïve bayes* yang sudah ada. Metode ini secara efektif menangani jumlah kata dengan menghitung frekuensi setiap kata, sedangkan pada klasifikasi *Naïve bayes* tradisional, frekuensi kata tidak terlalu mempengaruhi performa algoritma. Diketahui bahwa frekuensi setiap teks memiliki pengaruh yang lebih besar dalam mengkategorikan teks ke dalam kategori yang berbeda (Surya et al., 2019). Pada penelitian ini menggunakan 375 data dengan pembagian data *training* 80% dan data *test* 20%, namun sebagai contoh pada penelitian ini menggunakan 10 data yang sudah dilakukan *pre-processing*. Sebelum melakukan klasifikasi kita harus membagi data menjadi data *training* dan data *test*, berdasarkan data diatas terdapat 10 teks, pada penelitian ini menggunakan data *training* sebanyak 7 teks dan data *test* sebanyak 3 teks.

**Table 3. 12. Contoh Pembagian Data *Training* dan Data *Test***

ID D	Data <i>Training</i>	Label
D1	['buku', 'ada', 'pinjam', 'kenapa', 'gagal', 'unduh', 'file', 'tidak', 'temu', 'memori', 'handphone', 'tidak', 'ada', 'masalah']	<i>Negative</i>
D2	['tidak', 'unduh', 'buku', 'sudah', 'gabung', 'perbaiki']	<i>Positive</i>
D3	['rasa', 'hari', 'update', 'rak', 'baca', 'hilang', 'semua']	<i>Negative</i>
D4	['baca', 'buku', 'unduh', 'gagal', 'file', 'tidak', 'tersedia',]	<i>Positive</i>
D5	['klik', 'baca', 'keluar', 'sinyal', 'kuat', 'sistem', 'tolong', 'perbaiki']	<i>Positive</i>
D6	['rusak', 'baca', 'buku', 'unduh', 'tombol', 'pinjam', 'memuat', 'jaringan', 'cepat', 'aplikasi', 'bug']	<i>Negative</i>



Term	TF							Jumlah Label	
	D1 <i>Negative</i>	D2 <i>Positive</i>	D3 <i>Negative</i>	D4 <i>Positive</i>	D5 <i>Positive</i>	D6 <i>Negative</i>	D7 <i>Positive</i>	<i>Positive</i>	<i>Negative</i>
memuat	0	0	0	0	0	1	1	1	1
perbaiki	0	1	0	0	1	0	1	3	0
pinjam	1	0	0	0	0	1	0	0	2
rak	0	0	1	0	0	0	0	0	1
rasa	0	0	1	0	0	0	0	0	1
rusak	0	0	0	0	0	1	0	0	1
semua	0	0	1	0	0	0	1	1	1
sendiri	0	0	0	0	0	0	1	1	0
sinyal	0	0	0	0	1	0	0	1	0
sistem	0	0	0	0	1	0	0	1	0
sudah	0	1	0	0	0	0	0	1	0
temu	1	0	0	0	0	0	0	0	1
tersedia	0	0	0	1	0	0	0	1	0
tidak	2	1	0	1	0	0	1	3	2
tolong	0	0	0	0	1	0	0	1	0
tombol	0	0	0	0	0	1	0	0	1
unduh	1	1	0	1	0	1	0	2	2
update	0	0	1	0	0	0	0	0	1
Panjang Dokumen	14	6	7	7	8	11	13	34	32

Hal pertama yang dilakukan menghitung probabilitas prior dengan membagi jumlah teks *positive* dan *negative* dengan banyaknya teks pada data diatas sebagai berikut:

**Table 3. 14. Contoh Proses Probabilitas Prior**

Data Training	
Probabilitas Prior	
P <i>Positive</i>	0,57
P <i>Negative</i>	0,43
Hasil Keseluruhan	100%

Sebagai contoh rumus perhitungan probabilitas prior berdasarkan hasil table diatas sebagai berikut:

$$P \text{ positif} = \frac{\text{jumlah teks positif}}{\text{jumlah teks}} \quad (3.4)$$

$$P \text{ positif} = \frac{4}{7} = 0,57$$

$$P \text{ negatif} = \frac{3}{7} = 0,43$$

Selanjutnya menghitung probabilitas likelihood adalah perhitungan hasil term pada tiap label dibagi dengan hasil jumlah term *positive* dan *negative*, ditambah dengan jumlah term semua teks, hasilnya seperti pada gambar berikut:

**Table 3. 15. Contoh Proses Probabilitas Likelihood**

Term	Jumlah Label		Probabilitas Like Hood	
	<i>Positive</i>	<i>Negative</i>	<i>PL Positive</i>	<i>PL Negative</i>
ada	1	2	0,02	0,03
aplikasi	0	1	0,01	0,02
baca	3	2	0,04	0,03
batal	1	0	0,02	0,01
bug	0	1	0,01	0,02
buku	3	2	0,04	0,03
cepat	0	1	0,01	0,02
file	1	1	0,02	0,02
fitur	1	0	0,02	0,01
gabung	1	0	0,02	0,01
gagal	1	1	0,02	0,02
handphone	0	1	0,01	0,02
hari	0	1	0,01	0,02
henti	1	0	0,02	0,01
hilang	0	1	0,01	0,02
jaringan	0	1	0,01	0,02
keluar	1	0	0,02	0,01
kenapa	0	1	0,01	0,02
klik	2	0	0,03	0,01
kuat	1	0	0,02	0,01
laman	1	0	0,02	0,01
masalah	0	1	0,01	0,02
memori	0	1	0,01	0,02
memuat	1	1	0,02	0,02
perbaiki	4	0	0,05	0,01
pinjam	2	1	0,03	0,02
rak	0	1	0,01	0,02
rasa	0	1	0,01	0,02
rusak	0	1	0,01	0,02
semua	1	1	0,02	0,02
sendiri	1	0	0,02	0,01
sinyal	1	0	0,02	0,01
sistem	1	0	0,02	0,01

Probabilitas <i>likelihood</i>				
Term	Jumlah Label		Probabilitas Like Hood	
	<i>Positive</i>	<i>Negative</i>	<i>PL Positive</i>	<i>PL Negative</i>
sudah	1	0	0,02	0,01
temu	0	1	0,01	0,02
tersedia	1	0	0,02	0,01
tidak	3	2	0,04	0,03
tolong	1	0	0,02	0,01
tombol	0	1	0,01	0,02
unduh	2	2	0,03	0,03
update	0	1	0,01	0,02

Sebagai contoh penghitungan probabilitas *likelihood* berdasarkan hasil table diatas dengan rumus perhitungan berikut:

$$P \text{ positif}(\text{term}) = \frac{(\text{kemunculan term positif} + 1)}{(\text{jumlah term positif} + \text{jumlah term seluruh teks})} \quad (3.5)$$

$$\text{positif}(\text{ada}) = \frac{(1 + 1)}{(34 + 66)} = 0,02$$

$$\text{negatif}(\text{ada}) = \frac{(2 + 1)}{(32 + 66)} = 0,03$$

Terakhir, menghitung data *test* yang sudah dipisahkan. Perhitungan data *test* ini adalah mencari term yang terletak sama pada data *training*, berikut term yang sama dengan term yang berada di data *training*:

**Table 3. 16. Contoh Proses Term Data Test**

Data <i>test</i>	
Term	
D8	aplikasi
	buku
	keluar
	masalah
	perbaiki
	pinjam
	sendiri
	tolong
D9	buku
	gagal
	perbaiki
	pinjam
	tolong

Data <i>test</i>	
Term	
D10	buku
	jaringan
	kenapa
	perbaiki
	unduh

Dari term tersebut dihitung probabilitas posteriornya dengan mengkali semua hasil term yang berada di *likelihood* dan probabilitas prior sesuai dengan label pada term. Hasil dari probabilitas posterior sebagai berikut :

**Table 3. 17. Contoh Proses Predicted Data Test**

Predicted				
Data <i>test</i>	Probabilitas Posterior		Label	
Term	PL <i>Positive</i>	PL <i>Negative</i>		
D8	aplikasi	2,74286E-14	1,02857E-15	<i>Positive</i>
	buku			
	keluar			
	masalah			
	perbaiki			
	pinjam			
	sendiri			
tolong				
D9	buku	1,37143E-08	5,14286E-10	<i>Negative</i>
	gagal			
	perbaiki			
	pinjam			
	tolong			
D10	buku	3,42857E-09	1,54286E-09	<i>Positive</i>
	jaringan			
	kenapa			
	perbaiki			
	unduh			

Bedasarkan table diatas terdapat beberapa term yang sama dengan term di data *training*. Sebagai contoh D9 terdapat beberapa yang sama dengan data *training* seperti buku, gagal, perbaiki, pinjam, dan tolong. perhitungan probabilitas posterior berdasarkan hasil table diatas sebagai berikut:

$$P(w_i) = P(c_j) \times P(w_j|c_j) \times \dots \times P(w_n|c_n) \quad (3.6)$$

$$\begin{aligned} P(w_i) &= 0,57(\text{positif}) \times 0,04(\text{buku}|\text{positif}) \times 0,02(\text{gagal}|\text{positif}) \\ &\quad \times 0,05(\text{perbaiki}|\text{positif}) \times 0,03(\text{pinjam}|\text{positif}) \\ &\quad \times 0,02(\text{tolong}|\text{positif}) \\ &= 2,74286 \end{aligned}$$

$$\begin{aligned} P(w_i) &= 0,43(\text{negatif}) \times 0,03(\text{buku}|\text{negatif}) \times 0,02(\text{gagal}|\text{negatif}) \\ &\quad \times 0,01(\text{perbaiki}|\text{negatif}) \times 0,02(\text{pinjam}|\text{negatif}) \\ &\quad \times 0,01(\text{tolong}|\text{negatif}) \\ &= 1,02857 \end{aligned}$$

Yang memiliki nilai probabilitas posterior paling besar, term tersebut dapat dilabelkan sesuai dengan hasil yang paling besar. Berdasarkan hasil data *test* diatas dapat disimpulkan bahwa teks D8 memiliki sentiment *positive*, D9 memiliki sentiment *negative* dan teks D10 memiliki sentiment *positive*

### 3.2.6 Evaluasi klasifikasi

Pada tahap evaluasi ini menggunakan teknik *confusion matrix*, teknik ini sangat berguna untuk menganalisis performa classifier dalam mengenali tuple dari berbagai kelas. *True Positive* (TP) dan *True Negative* (TN) memberikan informasi mengenai kesalahan yang terjadi ketika classifier tidak berfungsi dengan benar (Riyanto, 2019). Analisis ini di ambil dari data *test* yang berlabel asli dengan data *test* yang telah label sesuai dengan klasifikasi *multinomial naïve bayes*. Hasil dari *confusion matrix* sebagai berikut:

Table 3. 18. Contoh *Confusion matrix*

		Predicted	
		Positive	Negative
Actual	Positive	2	1
	Negative	0	0

Selanjutnya menghitung *accuracy* yang menunjukkan rasio antara data yang diklasifikasikan dengan benar dengan keseluruhan data. *Precision* yang menunjukkan perbandingan antara data yang diklasifikasikan dengan benar dengan semua data dalam kelas yang sama. *Recall* yang menggambarkan rasio antara jumlah data yang diklasifikasikan dengan benar dalam suatu kelas dan seluruh data yang benar dalam kelas tersebut (Mahfud et al., 2020).

Table 3. 19. Contoh Proses Penghitungan *Accuracy*, *Precision*, *Recall* dan *F-measure*

	Hasil
<i>Accuracy</i>	67%
<i>Precision</i>	100%
<i>Recall</i>	67%
<i>F-measure</i>	80%

Perhitungan *Accuracy*

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.7)$$

$$A = \frac{(2 + 0)}{(2 + 0 + 1 + 0)} \times 100\% = 67\%$$

Perhitungan *Precision*

$$P = \frac{TP}{TP + FP} \quad (3.8)$$

$$P = \frac{2}{2 + 0} \times 100\% = 100\%$$

Perhitungan *Recall*

$$R = \frac{TP}{TP + FN} \quad (3.9)$$

$$R = \frac{2}{2 + 1} \times 100\% = 67\%$$

Perhitungan *F-measure*

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.10)$$

$$F = 2 \times \frac{67 \times 100}{67 + 100} = 80\%$$

### 3.4 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan pada bulan November 2024 hingga bulan Mei 2025. Fokus penelitian ini didasarkan pada identifikasi layanan aplikasi perpustakaan digital *iBI Library* melalui ulasan google *Play store* terkait pengalaman pengguna.

### 3.1 Subjek dan Objek Penelitian

Pada penelitian ini yang menjadi subjek adalah pengguna aplikasi *iBI Library* yang menuliskan pengalamannya pada kolom ulasan google *Play store*. Objek pada penelitian ini berupa aplikasi *iBI Library*

### **3.2 Sumber Data**

Penelitian ini menggunakan data primer yang berupa komentar layanan perpustakaan digital (*iBI Library*) yang dihitung mulai Oktober 2017 sampai dengan Oktober 2024. Sedangkan data sekunder yang digunakan berasal dari literatur terkait seperti jurnal ilmiah, buku, dan laporan penelitian sebelumnya

### **3.3 Teknik Pengumpulan Data**

Teknik pengumpulan data yang digunakan dalam penelitian adalah *scrapping* komentar aplikasi *iBI Library* pada google *Play store* menggunakan google colab

## BAB IV HASIL DAN PEMBAHASAN

### 4.1. Hasil Penelitian

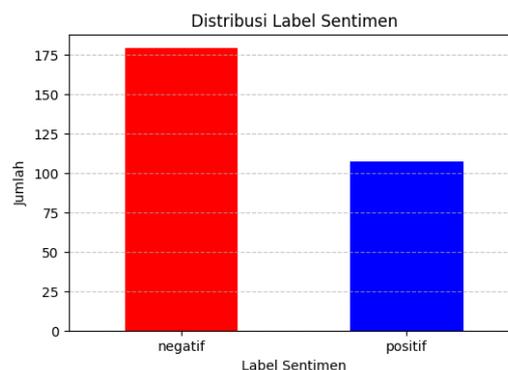
Pada bagian ini, menjelaskan hasil dari pengujian *machine learning* serta mengulas terakit implementasi metode *multinomial naïve bayes* untuk mengetahui sentiment pengguna tentang pengalaman menggunakan *iBI Library*. Sebelum pengujian dilakukan tahap *pre-processing* dan pembagian data menjadi data *training* dan data *test* yang memakai rasio 70:30. Penelitian ini melakukan pengujian dengan menghitung nilai *accuracy*, *recall*, *precision* dan *F-measure*. Berikut adalah hasil dari penelitian yang dilakukan.

#### 4.1.1 Pengambilan Data

Pengambilan data pada penelitian ini dilakukan menggunakan *Google Colab* dengan memanfaatkan *library* Google Scraper yang tersedia. Proses dilakukan dengan mencantumkan ID aplikasi yang ulasannya ingin dikumpulkan. Pengambilan data sesuai dengan batasan masalah yang telah ditetapkan dalam penelitian ini. Jumlah data yang diperoleh sebanyak 287 ulasan,

#### 4.1.2 Pelabelan Data

Pada penelitian ini, memnggunakan dataset yang terdiri dari 287 ulasan. Selanjutnya, dilakukan proses pembersihan data dengan menghapus entri yang terduplikasi. Data yang telah dibersihkan kemudian diberi label sentimen oleh ahli di bidang bahasa Indonesia, dengan klasifikasi ke dalam dua kategori, yaitu *positive* dan *negative*. Hasil pelabelan tersebut selanjutnya dikelompokkan dan dianalisis kuantitatif, sebagaimana ditampilkan pada diagram berikut.



Gambar 4.1 Hasil labeling dari para ahli

Gambar memberikan visualisasi berupa diagram batang dengan garis horizontal yang menunjukkan jenis kelas sentiment, sedangkan garis vertical menunjukkan jumlah dari masing-masing kelas sentiment, sentiment *positive* memiliki jumlah 107, kelas sentiment *negative* memiliki jumlah 179.

#### 4.1.3 Hasil *Pre-processing*

Data yang telah diberi label oleh ahli selanjutnya melalui tahap pembersihan untuk mengoptimalkan proses klasifikasi. Pada tahap ini, dilakukan identifikasi dan penghapusan elemen-elemen yang tidak relevan, seperti URL, emotikon, tagar, serta nama pengguna atau kata sandi, yang seluruhnya diganti dengan string kosong. Proses *case folding* juga diterapkan untuk mengubah seluruh huruf menjadi huruf kecil (*lowercase*). Selain itu, dilakukan normalisasi dengan mengubah kata-kata tidak baku menjadi padanan yang sesuai dengan kaidah Bahasa Indonesia berdasarkan Kamus Besar Bahasa Indonesia (KBBI).

Untuk proses *stopword removal* dan *stemming*, digunakan *library* Sastrawi, sedangkan proses *tokenizing* dilakukan dengan bantuan *library* NLTK. Data hasil *pre-processing* kemudian disusun ke dalam format *dataframe* menggunakan *library* pandas, sebagaimana ditunjukkan pada hasil berikut.

**Table 4. 1 Tahap *Cleaning***

No	<i>Cleaning</i>
1	kesulitan saat meminjam buku
2	Pada menu search buku ada filter pengarang Tolong ditambahkan menu search pengarang agar mudah mencari pengarang yang diinginkan ya
3	jumlah copy nya sikit amat gak niat ngasih pinjam buku
4	ini lagi eror kah kok ngga bisa login yaa
5	Tidak bisa login menggunakan akun google tolong diperbaiki
6	tolong kenapa saya jdi tdk bisa meminjam buku satu pun ya padahal bukunya masih tersedia tpi setiap dipinjam muncul tulisan error code sebelumnya saya baik2 saja jika ingin meminjam buku selalu bisa kenapa sekarang malah tdk bisa sama sekali
7	gabisa dibaca setelah klik pinjam disuruh gabung padahal udah gabung dan eror terus kyk gitu aja gak bisa dibuka ebooknya
8	tiba gk bisa minjem sm sekali selalu eror
9	kenapa pas di pinjam trus bukunya ga ada di rak pinjaman si
10	aplikasinya kok sekarang tidak bisa pinjam buku ya mau di download tidak bisa

Tabel 4.1 menampilkan hasil penghapusan atau pembersihan dari URL, emoticon, hastag, nama pengguna dan kata sandi. Selanjutnya dilakukan proses *case folding* sebagai berikut

**Table 4. 2 Tahap Case folding**

No	Case folding
1	kesulitan saat meminjam buku
2	pada menu search buku ada filter pengarang tolong ditambahkan menu search pengarang agar mudah mencari pengarang yang diinginkan ya
3	jumlah copy nya sikit amat gak niat ngasih pinjam buku
4	ini lagi eror kah kok ngga bisa login yaa
5	tidak bisa login menggunakan akun google tolong diperbaiki
6	tolong kenapa saya jdi tdk bisa meminjam buku satu pun ya padahal bukunya masih tersedia tpi setiap dipinjam muncul tulisan error code sebelumnya saya baik2 saja jika ingin meminjam buku selalu bisa kenapa sekarang malah tdk bisa sama sekali
7	gabisa dibaca setelah klik pinjam disuruh gabung padahal udah gabung dan eror terus kyk gitu aja gak bisa dibuka ebooknya
8	tiba gk bisa minjem sm sekali selalu eror
9	kenapa pas di pinajam trus bukunya ga ada di rak pinjaman si
10	aplikasinya kok sekarang tidak bisa pinjam buku ya mau di download tidak bisa

Tabel 4.2 menampilkan proses *case folding* dengan mengubah huruf besar menjadi huruf kecil semua, proses selanjutnya dilakukan tahap normalisasi dengan hasil berikut.

**Table 4. 3 Tahap Normalisasi**

No	Normalisasi
1	kesulitan saat meminjam buku
2	pada menu search buku ada filter pengarang tolong ditambahkan menu search pengarang agar mudah mencari pengarang yang diinginkan ya
3	jumlah copy nya sedikit amat gak niat ngasih pinjam buku
4	ini lagi eror kah kok ngga bisa masuk yaa
5	tidak bisa masuk menggunakan akun google tolong diperbaiki
6	tolong kenapa saya jadi tidak bisa meminjam buku satu pun ya padahal bukunya masih tersedia tapi setiap dipinjam muncul tulisan error code sebelumnya saya baik2 saja jika ingin meminjam buku selalu bisa kenapa sekarang malah tidak bisa sama sekali
7	gabisa dibaca setelah klik pinjam disuruh gabung padahal udah gabung dan eror terus kyk gitu aja gak bisa dibuka ebooknya
8	tiba ga bisa minjem sama sekali selalu eror
9	kenapa pas di pinjam terus bukunya ga ada di rak pinjaman si
10	aplikasinya kok sekarang tidak bisa pinjam buku ya mau di download tidak bisa

Tabel 4.3 menampilkan hasil normalisasi yang mengubah kalimat tidak baku atau slang word menjadi kata baku yang sesuai dengan kamus bahasa Indonesia, pada proses selanjutnya *stopword* seperti gambar berikut.

**Table 4. 4 Tahap Stopword**

No	Stopword
1	kesulitan meminjam buku
2	menu search buku filter pengarang tolong menu search pengarang mudah mencari pengarang
3	copy gak niat ngasih pinjam buku
4	eror ngga masuk
5	masuk akun google tolong diperbaiki
6	tolong meminjam buku bukunya tersedia dipinjam muncul tulisan error code baik2 meminjam buku
7	gabisa dibaca klik pinjam disuruh gabung udah gabung eror kyk gitu aja gak dibuka ebooknya
8	ga minjem eror
9	pinjam bukunya ga rak pinjaman si
10	aplikasinya pinjam buku download

Tabel 4.4 menampilkan hasil *stopword* dengan menghapus kata yang tidak diperlukan dan yang tidak memiliki bobot, setelah itu dilakukan proses *tokenizing* dengan hasil berikut.

**Table 4. 5 Tahap Tokenized**

No	Tokenized
1	['kesulitan', 'meminjam', 'buku']
2	['menu', 'search', 'buku', 'filter', 'pengarang', 'tolong', 'menu', 'search', 'pengarang', 'mudah', 'mencari', 'pengarang']
3	['copy', 'gak', 'niat', 'ngasih', 'pinjam', 'buku']
4	['eror', 'ngga', 'masuk']
5	['masuk', 'akun', 'google', 'tolong', 'diperbaiki']
6	['tolong', 'meminjam', 'buku', 'bukunya', 'tersedia', 'dipinjam', 'muncul', 'tulisan', 'error', 'code', 'baik2', 'meminjam', 'buku']
7	['gabisa', 'dibaca', 'klik', 'pinjam', 'disuruh', 'gabung', 'udah', 'gabung', 'eror', 'kyk', 'gitu', 'aja', 'gak', 'dibuka', 'ebooknya']
8	['ga', 'minjem', 'eror']
9	['pinjam', 'bukunya', 'ga', 'rak', 'pinjaman', 'si']
10	['aplikasinya', 'pinjam', 'buku', 'download']

Tabel 4.5 proses *tokenizing* menampilkan pemisahan kata sehingga data yang tertokenisasi dapat dilakukan pemboobtan menggunakan TF-IDF dengan maksimal, selanjutnya dilakukan proses *stemming*.

Table 4. 6 Tahap *Stemming*

No	<i>Stemming</i>
1	['sulit', 'pinjam', 'buku']
2	['menu', 'search', 'buku', 'filter', 'arang', 'tolong', 'menu', 'search', 'arang', 'mudah', 'cari', 'arang']
3	['copy', 'gak', 'niat', 'ngasih', 'pinjam', 'buku']
4	['eror', 'ngga', 'masuk']
5	['masuk', 'akun', 'google', 'tolong', 'baik']
6	['tolong', 'pinjam', 'buku', 'buku', 'sedia', 'pinjam', 'muncul', 'tulis', 'error', 'code', 'baik2', 'pinjam', 'buku']
7	['gabisa', 'baca', 'klik', 'pinjam', 'suruh', 'gabung', 'udah', 'gabung', 'eror', 'kyk', 'gitu', 'aja', 'gak', 'buka', 'ebooknya']
8	['ga', 'minjem', 'eror']
9	['pinjam', 'buku', 'ga', 'rak', 'pinjam', 'si']
10	['aplikasi', 'pinjam', 'buku', 'download']

Tabel 4.6 menampilkan hasil pengembalian kata menjadi kata dasar. Pada proses ini, semua berhasil dikembalikan kepada kata dasar. tabel 4.1, tabel 4.2, tabel 4.3, tabel 4.4, tabel 4.5, dan tabel 4.6 menampilkan dataframe yang sama dengan 10 contoh data dari hasil *pre-processing*. Data tersebut disimpan dalam format csv yang akan dilakukan proses pembobotan TF-IDF.

#### 4.1.4 Hasil Pembobotan Kata

Pada proses ini, menggunakan fungsi *tfidfVectorizer* dari *library* scikit learn untuk pembobotan kata. Hasil pembobotan memunculkan nilai frekuensi kata berupa angka decimal, pada 287 data, terdapat 2470 kata yang terbobot. Hasil dari pembobotan ini akan digunakan sebagai klasifikasi pada *training* data. Hasil dari pembobotan sebagai berikut.

Table 4. 7 Tahap TF-IDF

Dokumen	Kata	Score pembobotan
Dokumen 1 ( <i>negative</i> )	Buku	1.7107
	Pinjam	2.3828
	Sulit	5.5609
Dokuem 2 ( <i>positive</i> )	Buku	1.7107
	Mudah	4.7136
	Filter	5.5609
	Cari	4.1746
	Search	10.5464
	Menu	11.1217

Dokumen	Kata	Score pembobotan
	Tolong	3.1331
	Arang	15.1501
Dokumen 3 ( <i>negative</i> )	Buku	1.7107
	pinjam	2.3828
	Niat	5.2732
	Ngasih	5.0500
	Copy	5.0500
	Gak	3.1630
Dokumen 4 ( <i>negative</i> )	Ngga	5.0500
	Eror	4.3569
	Masuk	3.1630

Tabel 4.7 adalah hasil dari perhitungan pembobotan TF-IDF dengan mengambil beberapa dokumen yang sebagai contoh pada tabel terdapat sentiment *negative* dan *positive*. Pada kolom kata adalah kata-kata yang ada pada tiap sentiment dan pada kolom score pembobotan adalah hasil atau nilai dari tiap kata tersebut dengan hasil yang berbeda-beda tiap katanya.

#### 4.1.5 Hasil Pelatihan Data

Sebelum proses pelatihan dan pengujian dilakukan terhadap data yang telah direpresentasikan dalam bentuk bobot TF-IDF, data terlebih dahulu dibagi berdasarkan skenario rasio 70:30 dan rasio 80:20. Pembagian data ini dilakukan dengan menggunakan fungsi *train\_test\_split* dari *library* scikit-learn, yang memungkinkan pemisahan data secara acak maupun terkontrol sesuai dengan rasio yang ditentukan.

Bagi data dengan rasio 70:30 menggunakan variable *test\_size* bernilai 0,3. *Variable test size* dapat dikatakan bahwa jumlah presentase data yang akan menjadi data uji, nilai 0,3 yang berarti 30% dari total data. Pada rasio 70:30, data berjumlah total 287 yang terbagi menjadi dua yaitu, data *test* yang berjumlah 87 dan data *training* yang berjumlah 200. Bagi data dengan rasio 80:20 menggunakan variabel *test\_size* bernilai 0.2. *Variable test size* ini menunjukkan persentase data yang digunakan sebagai data uji, di mana nilai 0.2 berarti 20% dari total data. Pada rasio 80:20, jumlah total data sebanyak 287 dibagi menjadi dua bagian, yaitu data uji sebanyak 57 data dan data latih sebanyak 230 data. Bagi data dengan rasio 80:20

menggunakan *variabel test\_size* bernilai 0.2 yang menunjukkan persentase data yang digunakan sebagai data uji, dimana nilai 0.2 berarti 20% dari data keseluruhan. Data uji yang digunakan adalah data mentah yang belum dilakukan *pre-processing*. Pada rasio ini, jumlah total data sebanyak 287 dibagi menjadi dua bagian, yaitu data uji sebanyak 56 data dan data latih sebanyak 231 data

Setelah melakukan proses split data, akan dilakukan proses pelatihan data dengan menggunakan model *multinomial naïve bayes* dari *library* sklearn. Data yang digunakan adalah hasil pembagian rasio sebelumnya yang tersimpan di *variable x\_train* sebagai fitur dan *y\_train* sebagai labelnya. *Multinomial* sendiri merupakan pengembangan dari *naïve bayes* agar bisa melakukan proses perhitungan data selain data biner. Data tersebut dapat berupa bilangan decimal maupun bilangan real, salah satu contoh kasusnya adalah sentiment analisis. Sentiment analisis yang merupakan sekumpulan kata yang menghitung setiap kata dan mempresentasikan kata tersebut berdasarkan total munculnya kata. Hal ini dipadukan dengan metode probabilitas dari *naïve bayes*.

Karena sebelumnya sudah dilakukan pelabelan, proses berikutnya adalah perhitungan probabilitas dengan menggunakan pembobotan kata pada masing-masing dokumen. Perhitungan ini diimplementasikan secara otomatis melalui pemrograman, berdasarkan frekuensi kemunculan kata dalam konteks label tertentu. Hasil dari proses pelatihan ini menghasilkan sebuah model *machine learning* yang siap digunakan untuk proses klasifikasi pada tahap pengujian. Selain itu, pengujian ini juga berfungsi untuk mengevaluasi performa model dalam mengklasifikasikan data secara akurat.

#### **4.1.6 Hasil Pengujian Data**

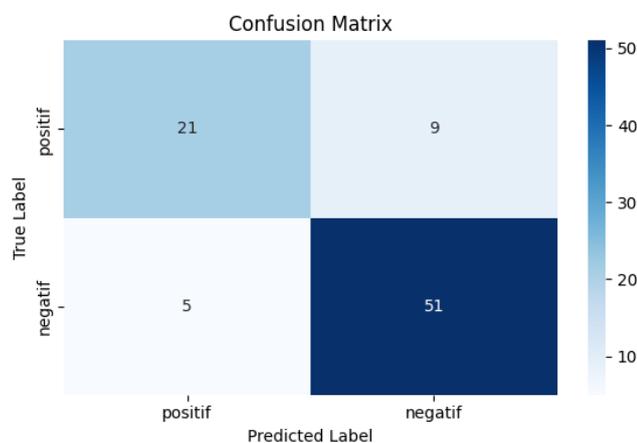
Tahap ini merupakan proses pengujian dan evaluasi terhadap data uji yang dilakukan dengan menggunakan *confusion matrix*. Evaluasi ini bertujuan untuk mengukur tingkat akurasi klasifikasi yang dihasilkan oleh sistem. Untuk memperoleh nilai *true positive*, *true negative*, *false positive*, dan *false negative*, digunakan fungsi *confusion matrix* dari *library* scikit-learn. Selanjutnya, fungsi *classification report* diterapkan untuk mengevaluasi performa model secara lebih komprehensif, dengan menampilkan metrik evaluasi berupa akurasi, presisi, *recall*,

dan *F-measure*, berdasarkan skenario pembagian data dengan rasio sebagai berikut.

Table 4. 8 Hasil Peforma *Multinomial naïve bayes*

Rasio	Akurasi	Presisi	<i>Recall</i>	<i>F-measure</i>
70:30	0.84	0.83	0.81	0.81
80:20	0.83	0.82	0.82	0.82
80:20 Data tanpa pre- processing	0.64	0.74	0.64	0.60

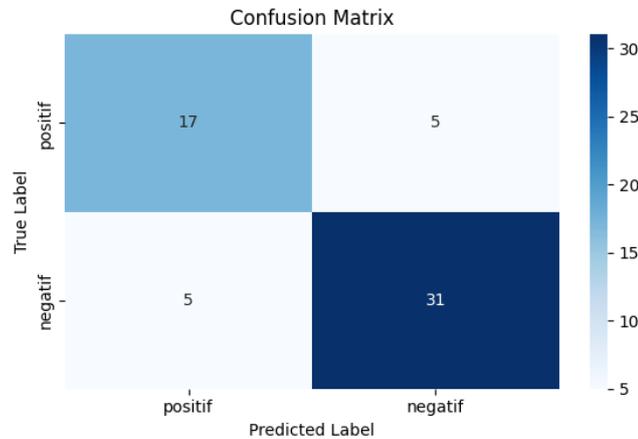
Pada tabel 4.8 menunjukkan hasil *confusion matrix* pada rasio 70:30 memiliki presentase 84% untuk akurasi, presisi sejumlah 83%, *recall* sejumlah 81%, *F-measure* sejumlah 81%. Sementara pada rasio 80:20 akurasi memiliki presentase 83%, untuk presisi, *recall* dan *f-measure* memiliki presentase yang sama di angka 82%. Semua presentase yang telah disebutkan berdasarkan perhitungan jumlah kecocokan prediksi nilai *true* dan *false* dari data aktual. Hasil tersebut di visualisasikan dengan sklearn berbentuk table *matrix* berikut.



Gambar 4. 2 Hasil data test 70:30

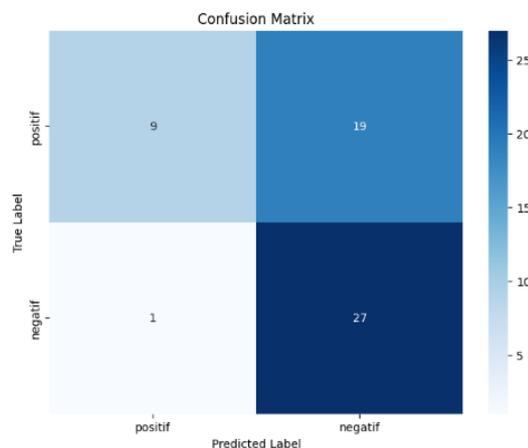
Pada gambar 4.2 dapat diketahui rasio 70:30, hasil evaluasi peforma model *multinomial naïve bayes*. Berdasarkan visualiasi dapat diketahui dari 30 data yang sebenarnya berlabel *positive*, sejumlah 21 data yang sebenarnya *positive* dan berhasil diprediksi *positive* oleh model (*true positive*), sedangkan 9 data yang sebenarnya *positive*, tetapi salah terprediksi menjadi *negative* (*false negative*). Sementara itu, terdapat 56 data berlabel *negative*, sebanyak 51 data yang sebenarnya *negative* dan berhasil terprediksi *negative* saat pengujian (*true negative*)

dan 5 data yang sebenarnya *negative*, namun salah diprediksi menjadi *positive* (*false positive*).



**Gambar 4. 3 Hasil data test 80:20**

Pada gambar 4.3 dapat diketahui rasio 80:20, mempresentasikan hasil evaluasi performa model *multinomial naïve bayes*. Berdasarkan visualiasi dapat diketahui dari 22 data yang sebenarnya berlabel *positive*, sejumlah 17 data yang sebenarnya *positive* dan berhasil diprediksi *positive* oleh model (*true positive*), sedangkan 5 data yang sebenarnya *positive*, tetapi salah terprediksi menjadi *negative* (*false negative*). Sementara itu, terdapat 36 data berlabel *negative*, sebanyak 31 data yang sebenarnya *negative* dan berhasil terprediksi *negative* saat pengujian (*true negative*) dan 5 data yang sebenarnya *negative*, namun salah diprediksi menjadi *positive* (*false positive*).



**Gambar 4. 4 Hasil data test 80:20 (Data tanpa pre-processing)**

Pada gambar 4.4 dapat diketahui rasio 80:20 dengan data *test* yaitu data tanpa *pre-processing*, Dapat diketahui dari 28 data yang sebenarnya berlabel *positive*, sejumlah 9 data yang sebenarnya *positive* dan berhasil diprediksi *positive* oleh model (*true positive*), sedangkan 19 data yang sebenarnya *positive*, tetapi salah terprediksi menjadi *negative* (*false negative*). Sementara itu, terdapat 28 data berlabel *negative*, sebanyak 27 data yang sebenarnya *negative* dan berhasil terprediksi *negative* saat pengujian (*true negative*) dan 1 data yang sebenarnya *negative*, namun salah diprediksi menjadi *positive* (*false positive*).

Tingkat kecerahan warna di samping mempresentasikan kepadatan klasifikasi pada masing-masing kelas, warna biru tua menunjukkan jumlah klasifikasi yang tinggi. Evaluasi dari model ini menunjukkan kemampuan klasifikasi yang cukup baik, terutama dalam mengenali data *negative*, sebagai contoh jumlah *true negative* yang cukup tinggi. *Matrix* ini menjadi dasar evaluasi seperti akurasi, presisi, *recall* dan *f-measure*. Berikut sampel hasil data *test*.

**Table 4. 9 Contoh Hasil Data Test 70:30**

No	Content	Negative	Positive	Predicted	Actual
1	bukunya kok ga bisa dipinjem ya? padahal masi ada 2 copy, tapi selalu muncul tulisan 400. tolong diperbaikii	0.861856	0.138144	Negative	Negative
2	Tidak bisa login, saya sudah daftar, pas masuk akun anda keblokir. Ini gimana sih?	0.878158	0.121842	Negative	Negative
3	Baru install dan daftar, udah pinjam dan diunduh tinggal baca aja, tapi error mulu. Alhasil masih belum bisa baca buku yang diunduh, kan sayang gitu. Kalau gini mending uninstall aja.	0.852305	0.147695	Negative	Negative
4	Overall bagus. Tapi mungkin bisa ditambah lagi koleksi bukunya, especially untuk buku yang sedang ramai di kalangan remaja seperti saya.	0.526197	0.473803	Negative	Positive
5	Saya sangat senang dengan aplikasi ini namun aplikasinya problem loading book categories, unknown error. Apakah sedang ada gangguan? Saya sudah coba reinstall tapi masih mengalami hal yang sama.	0.692850	0.307150	Negative	Negative

Tabel 4.9 menunjukkan hasil klasifikasi terhadap data *test* yang meliputi konten, nilai du akelas sentiment *positive* dan *negative*, hasil prediksi model (*predicted*) dan hasil label sebenarnya (*actual*). Pada hasil tersebut dapat dilihat bahwa model

memberikan nilai berdasarkan kelas berdasarkan kata dalam sentiment tersebut, sebagai contoh, pada baris pertama, model memberikan nilai pemodelan sebesar 0.861856 terhadap kelas *negative* dan 0.138144 terhadap kelas *positive*, sehingga model mengklasifikasikan data tersebut adalah *negative*, yang sesuai dengan label aktual (*Actual*) yaitu *negative*. Berikut beberapa kata-kata yang muncul pada sentiment *positive* dan *negative*.



Gambar 4. 5 Hasil wordcloud

Gambar 4.3 menunjukkan bahwa terdapat beberapa kata seperti buku, koleksi, manfaat, bantu dan aplikasi pada sentiment *positive*, selain itu pada sentiment *negative* juga terdapat beberapa kata seperti buku, ga, baca, gabisa, pinjam dan gak. Dari wordcloud dapat menggambarkan kata-kata yang paling sering muncul dalam ulasan pengguna berdasarkan data *test*.

## 4. 2. Pembahasan

Pada tahap awal penelitian, pengambilan data dilakukan menggunakan tools google colab dengan *library* google scraper, yang memudahkan peneliti saat mengakses dan mengunduh data. Saat pengambilan data awal dikumpulkan sebanyak 225 data, kemudian peneliti mencoba untuk memodifikasi dan menambahkan script untuk memperoleh data sesuai dengan batasan masalah, akhirnya data yang berhasil diperoleh sebanyak 287 data. Dari jumlah data yang didapat tersebut dilakukan tahap pembersihan (*data cleaning*) dari data yang terduplikat serta data yang memiliki elemen-elemen yang tidak relevan seperti URL, emoticon, mention dan whitespace, setelah proses pembersihan, data diberi label oleh para ahli di bidang bahasa Indonesia, antaranya Tamara Khadjijah Al Kubra dan Aliza Nur Sabila, lulusan program studi bahasa dan sastra Indonesia Universitas Negeri Malang, serta bapak Jarwoto,S.pd, seorang pendidik yang

diBidang Pendidikan bahasa Indonesia. Kemudian proses pelabelan dilakukan berdasarkan sentiment yang terkandung dalam masing-masing sentiment, pada penelitian ini menggunakan dua kategori pelabelan yaitu label *positive* dan label *negative*. Label netral tidak digunakan karena labeling netral sering kurang informatif hal itu akan membuat sentiment seseorang kearah yang kurang jelas pada suatu topik tertentu. Sekaligus kata netral dalam KBBI diartikan sebagai tidak berpihak, maka dari itu kalimat bersentimen netral ini merupakan ekspresi kalimat yang tidak memiliki kecenderungan *positive* maupun *negative* (Arham et al., 2022). Serta, sikap netral dianggap tidak memberikan manfaat yang signifikan bagi suatu lembaga atau instansi (Puspala Sari & Jaya, 2023).

Dari hasil pelabelan pada gambar 4.1, kemudian dilakukan langkah *pre-processing* agar data yang awalnya data mentah menjadi data yang siap diolah. Hal yang dilakukan adalah tahap *case folding* dengan merubah huruf besar menjadi kecil agar saat proses klasifikasi memudahkan machine learning. Pada tahap normalisasi mengubah kata slang, kata singkatan, dan kata tidak baku menjadi kata yang sesuai dengan bahasa Indonesia yang benar, kata yang sering dipakai adalah “dgn, sy, yg, bs, apk”. Selanjutnya penghapusan (*Stopword*) dengan *library* sastrawi yang memiliki modul kata yang perlu untuk dihapus. Pada tahap ini, masih ditemukan beberapa kata yang seharusnya tidak memiliki makna yang tidak penting, namun belum berhasil dihapuskan oleh *stopword*. Setelah itu dilakukan *tokenizing* untuk memisahkan kata untuk memudahkan saat dilakukan pembobotan kata (TD-IDF), kemudian dilakukan *stemming* untuk merubah kalimat yang memiliki imbuhan menjadi kata dasar, pada tahap ini memerlukan waktu sedikit lama saat prosesnya, namun terdapat beberapa kata yang masih belum berhasil diubah menjadi kata dasar, hal itu dikarenakan terdapat kelebihan dalam huruf vocal seperti “akhirnyaaa”, “antriannya”, dan “bagossss”. Setelah proses *stemming* selesai, maka hasil proses *pre-processing* dapat digunakan pada proses selanjutnya yaitu pembobotan kata.

Penelitian ini menerapkan pengklasifikasian *multinomial naïve bayes* dengan tahap *pre-processing* yang sesuai, pada tahap normalisasi memerlukan waktu yang cukup lama untuk memeriksa dan mencatat kata-kata yang harus dibenarkan dan

diperbaiki agar data digunakan dapat berjalan secara maksimal. Pada tahap *stopword* dan *stemming* terdapat banyak kata yang harusnya tereliminasi dan dikembalikan menjadi kata dasar, tetapi belum bekerja cukup baik. Beberapa contoh kata tersebut seperti kata yang imbuhan dengan vocal akhir yang berlebihan, singkatan, dan kata yang tidak sesuai dengan bahasa Indonesia yang baik dan benar. Ada juga beberapa kata yang termuat dalam *library* sastrawi, namun belum cukup berkerja dengan baik.

Pada tabel 4.7 pembobotan dilakukan untuk memberi nilai numerik pada setiap kata dan sejauh mana suatu kata memberikan kontribusi terhadap suatu kelas sentiment, baik *positive* ataupun *negative*, dari hasil pembobotan terdapat 2470 kata unik yang terkandung dalam sentiment tersebut. Nilai bobot mulai dari 0.0 hingga yang terbesar 21.00. setelah dihitung berdasarkan frekuensi kemunculan kata dalam setiap kelas sentiment, kemudian dilakukan normalisasi terhadap jumlah total kata dalam kelas tersebut. Sebagai contoh, kata seperti “bagus”, “manfaat”, dan “baik” dengan bobot cukup tinggi pada kelas sentiment positif, dan kata seperti “ga”, “buku”, dan “baca” yang dominan tinggi pada kelas *negative*.

Setelah pembobotan menggunakan metode TF-IDF selanjutnya proses pelatihan dalam model klasifikasi. Klasifikasi dilakukan dengan membagi menjadi dua bagian, yaitu data training dan data *test* dengan perbandingan rasio 70:30 dan rasio 80:20. Dengan beberapa rasio tersebut, diharapkan dapat menghasilkan performa model yang optimal. Dari pembagian tersebut data *training* dimasukkan ke dalam model klasifikasi *multinomial naïve bayes*, dengan menghitung log prior dan log *likelihood* sebagai dasar perhitungan probabilitas masing-masing kelas, Setelah dilakukan klasifikasi pada proses *training*, selanjutnya evaluasi model menggunakan data uji melalui metode *confusion matrix*. Hasil evaluasi menunjukkan adanya variasi nilai pada metrik evaluasi yang digunakan, yaitu akurasi, presisi, *recall*, dan *f-measure*, yang dimana hasil tersebut mencerminkan sejauh mana model mampu mengklasifikasikan data secara tepat.

Table 4.8 menunjukkan hasil performa dari model klasifikasi *multinomial naïve bayes*, yang berdasarkan dua skenario pembagian data latih dan data uji, yaitu rasio 70:30, 80:20, dan 80:20 dengan data test yang digunakan adalah data mentah. Pada

rasio 70:30, model menghasilkan akurasi sebesar 0.84, bisa dikatakan 84% model data berhasil diprediksi dengan benar oleh model dibandingkan dengan total jumlah data yang di uji. Sedangkan untuk rasio 80:20, model menghasilkan nilai akurasi sebesar 0.83, yang berarti sebanyak 83% dari total data uji berhasil diklasifikasikan secara benar oleh model. Pada rasio 80:20 menggunakan data tanpa pre-processing yang dihasilkan sebesar 0.64, yang dimana sebanyak 64% dari total uji data mentah berhasil diklasifikasikan. Akurasi digunakan untuk menilai tingkat keseluruhan kebenaran prediksi model terhadap data uji, yaitu seberapa besar proporsi klasifikasi yang dilakukan model sesuai dengan label aslinya.

Sementara itu, pada rasio 70:30 nilai presisi mendapatkan di angka 0.83 menunjukkan 83% data yang diprediksi sebagai *positive* yang benar-benar *positive*, dapat dikatakan terprediksi dengan benar, sedangkan pada rasio 80:20 nilai presisi yang dihasilkan sedikit turun diangka 0.82, menunjukkan bahwa 82% ini mencerminkan seberapa baik model dalam menghindari kesalahan klasifikasi dari hasil data yang sebenarnya *negative* sebagai *positive*, sedangkan pada rasio 80:20 dengan menggunakan data mentah mengalami penurunan pada angka 74%. Presisi ini digunakan untuk menilai tingkat ketepatan model dalam memberikan label *positive*, yang artinya berapa banyak prediksi *positive* yang benar-benar relevan.

Selanjutnya, nilai *recall* pada rasio 70:30 diangka 0.81 yang menunjukkan sebanyak 81% model berhasil mendeteksi data aktual yang benar-benar *positive* maupun *negative*, sedangkan pada rasio 80:20 *recall* yang dihasilkan sedikit tinggi sebesar 0.82, mengindikasikan bahwa model berhasil mengenali 82% dari total data uji yang benar-benar memiliki label sentimen tertentu, pada rasio 80:20 dengan data menghasilkan *recall* sebesar 64%. Dengan kata lain, *recall* digunakan untuk menilai kemampuan model dalam menemukan semua data yang relevan, atau seberapa banyak data yang benar-benar *positive* berhasil dikenali.

*F-measure* pada rasio 70:30 sebesar 0.81, bahwa 81% rata-rata nilai presisi dan *recall* seimbang, pada rasio 80:20 *f-measure* yang dihasilkan mengalami peningkatan sebesar 0.82, yang dimana nilai 82% menunjukkan keseimbangan yang baik antara kemampuan model dalam menghindari prediksi yang salah dan

kemampuan dalam mendeteksi seluruh data yang relevan, sedangkan pada rasio 80:20 pada data mentah score *f-measure* mengalami penurunan pada angka 60%. *F-measure* digunakan untuk menilai keseimbangan performa model, khususnya saat presisi dan *recall* memiliki kepentingan yang sama, serta untuk mengevaluasi model pada situasi data yang tidak seimbang. Hal ini menunjukkan bahwa meskipun jumlah data uji lebih sedikit dibandingkan dengan rasio 70:30, model tetap memiliki performa klasifikasi yang konsisten.

Meskipun dengan rasio 70:30 menunjukkan akurasi yang sedikit lebih tinggi dibandingkan rasio 80:20. Namun, dengan memperhatikan keseimbangan antara presisi dan *recall* yang diwakili nilai *f-measure*, maka rasio 80:20 menunjukkan performa yang lebih stabil. Hal tersebut diperkuat hasil *confusion matrix* diantaranya rasio 80:20 memiliki nilai *recall* dan *f-measure* yang lebih baik, yakni 0.82 dibanding 0.81 pada rasio 70:30, selain itu nilai *f-measure* yang lebih tinggi pada rasio 80:20 menunjukkan model lebih seimbang dalam mengklasifikasi *positive* dan *negative* dan jumlah total kesalahan (*false positive* dan *false negative*) juga lebih rendah dan rasio 80:20 lebih seimbang daripada rasio 70:30. Dengan demikian rasio 80:20 memberikan hasil performa yang lebih optimal dan seimbang dalam klasifikasi sentimen, oleh karena itu rasio 80:20 dapat dikatakan sebagai rasio yang sesuai dan tepat dalam model klasifikasi sentimen pada penelitian ini

Dalam klasifikasi sentimen, presisi, *recall* dan *f-measure* memiliki peran penting dalam memastikan bahwa model diantaranya tidak banyak melakukan salah prediksi data *negative* sebagai *positive* atau bisa dikatakan presisi, *recall* dengan nilai tinggi bahwa model mampu mendeteksi data yang seharusnya diklasifikasikan sebagai *positive* dan *f-measure* yang tinggi menandakan model memiliki keseimbangan yang baik antara presisi dan *recall*.

Penghitungan dari pengujian dan evaluasi pada tabel 4.8 menunjukkan bahwa hasil performa model *machine learning* yang digunakan pada penelitian ini, sebagai contoh rasio 70:30 dengan akurasi 84% dari keseluruhan data *test*, dengan menghitung total jumlah prediksi benar maupun tidak benar yang sesuai dengan data aktual lalu dibagi dengan keseluruhan jumlah prediksi yang benar maupun tidak benar dengan data aktual. Adapun presisi yang memperoleh nilai 84% dari

keseluruhan data *test*, dengan menghitung hasil terprediksi benar dibagi dengan jumlah total yang terprediksi benar dengan nilai yang terprediksi salah. Selain itu, ada *recall* yang memperoleh jumlah senilai 84% dari keseluruhan data *test*, dengan perhitungan nilai yang terprediksi benar pada kelasnya dibagi dengan jumlah total dari nilai terprediksi benar dengan jumlah total nilai terprediksi salah. Untuk memperoleh evaluasi secara menyeluruh, penelitian ini menggunakan *f-measure* sebagai alternatif evaluasi performa model, *f-measure* menggabungkan nilai *recall* dan presisi, dengan rumus dua dikalikan dengan hasil nilai perkalian antara presisi dengan *recall* dibagi dengan hasil penjumlahan antara presisi dan *recall*. Nilai *f-measure* yang diperoleh adalah 83% keseluruhan data.

Pada gambar 4.2 menunjukkan *confusion matrix* model klasifikasi sentimen yang menggambarkan perbandingan hasil prediksi data uji model klasifikasi pada rasio 70:30, antara label sebenarnya (*actual label*) dengan label diprediksi oleh model (*predicted label*) dengan pembagian dua kelas yaitu *positive* dan *negative*. Hasil yang diperoleh diantaranya sebanyak 21 data yang sebenarnya *positive* dan berhasil terprediksi *positive* (*true positive*), terdapat juga 9 data yang sebenarnya *positive*, namun salah prediksi menjadi *negative* (*false negative*), selain itu juga terdapat 51 data yang sebenarnya *negative* dan berhasil diprediksi *negative* oleh model (*true negative*), dan sebanyak 5 data yang sebenarnya *negative*, namun terprediksi menjadi *positive*. Hasil yang ditampilkan pada gambar 4.3 adalah rasio 80:20 hasilnya terdapat sebanyak 17 data dengan sebenarnya *positive*, dan terprediksi positif (*true positive*), 5 data yang sebenarnya *positive* tetapi diprediksi sebagai *negative* (*false negative*), 31 data dengan label sentimen *negative* dan diprediksi sebagai *negative* (*true negative*), dan 5 data yang sebenarnya *negative* tetapi diprediksi sebagai *positive*.

Berdasarkan tabel 4.9 hasil pengujian tersebut, menunjukkan setiap sentiment memberikan nilai prediksi yang berbeda berdasarkan kata dalam masing-masing sentiment. Pada model klasifikasi label sentiment diukur berdasarkan nilai prediksi yang lebih tinggi diantara tiap label sentiment, maka akan masuk pada label sentiment tersebut. Hal ini memberikan penjelasan terkait kemampuan model klasifikasi dalam mengenali dan mengklasifikasikan sentiment dari teks yang

dianalisis. Namun, terdapat kesalahan klasifikasi, di mana label kelas hasil prediksi tidak sesuai dengan label aktual. Hal tersebut dapat terjadi dikarenakan terdapat kemiripan kata pada saat pembobotan kata atau makna yang ambigu dalam sentiment tersebut. Diantaranya seperti kata “buku” kata ini muncul di sentimen *positive* dan *negative* serta memiliki bobot yang serupa, hal tersebut berdampak bahwa kata tersebut tidak informatif saat diprediksi dan seperti kata “menu” kata ini sangat umum digunakan dalam konteks user experience, sebagai contoh dalam sentimen *positive* “menu yang disediakan lengkap” atau sentimen *negative* “menu membingungkan”. Selain itu, juga terdapat kata-kata seperti “copy”, “search”, “google”, dan “menu” termasuk dalam kata noise yang dikarenakan bahasa inggris umum yang bisa dianggap netral tanpa konteks yang jelas hal tidak spesifik dalam konteks pada label sentimen dan kata “eror” bisa dikatakan salah pengejaan pada saat TF-IDF menganggap ini sebagai kata unik, meskipun maknanya sama. selain itu terdapat kata seperti “gak”, “ngga” dan “ngasih” kata-kata ini memiliki bentuk kata informal, hal tersebut menyebabkan model klasifikasi tidak mengenali jika tidak dinormalisasikan. Terdapat juga kata gaul yang disingkat seperti “baik2”, “buku2”, “bener2”, “iya2”, “berlomba2”, “rasa2” dan “bisa2” termasuk kata gabungan informal, bisa mengganggu jika tidak dipisah. Terdapat juga kata-kata tidak teratur seperti “dunkbiar”, “pus”, “ampas”, “ingalls”, “cetur” dan “bjir” dari kata tersebut kemungkinan gabungan kata gaul atau typo, bentuk ekspresi tanpa konteks dan maknanya kabur jadi sulit untuk diinterpretasikan oleh model.

Dapat diamati bahwa terdapat karakteristik pada tiap sentiment yang dipengaruhi oleh beberapa kata, yang dimana kata tersebut diperhitungkan dan dilasifikasikan ke dalam kategori sentiment *positive* atau *negative*. Pada sentiment *positive*, terdapat kata-kata yang dominan antara lain “koleksi”, “buku”, “aplikasi”, “bagus”, “manfaat”, “baik”, dan “bantu”, berdasarkan kata-kata tersebut menunjukkan bahwa pengguna dengan sentiment *positive* cenderung memberikan ulasan terkait kemudahan dalam penggunaan aplikasi, serta manfaat yang diperoleh dengan kelengkapan koleksi buku yang tersedia dalam layanan aplikasi tersebut. Selain itu, terdapat kata-kata sering muncul pada sentiment *negative* diantaranya “gakbaca”, “buku”, “pinjam”, “gabisa”, “klik”, “daftar”, dan “eror”.

Hal ini menunjukkan bahwa keluhan utama yang dialami pengguna berasal dari kendala teknis seperti kesulitan dalam proses peminjaman, serta kesulitan saat melakukan login dan navigasi antarmuka yang kurang dimengerti oleh pengguna baru, kata “gabisa” dan “pinjam” sering kali muncul beriringan dalam satu kalimat sentiment yang mengindikasikan adanya hambatan saat pengguna ingin mengakses fitur dan navigasi utama dari aplikasi. Berdasarkan data terdapat beberapa permasalahan yang dihadapi oleh *iBI Library*, meskipun layanan ini mendapatkan respon yang cukup beragam, pada distribusi sentimen menunjukkan adanya sentimen *negative* lebih banyak daripada sentimen *positive* pada beberapa aspek layanan, diantaranya dari segi kualitas layanan digital pada sisi teknis seperti kesulitan login, akses ebook maupun navigasi sistem, terdapat juga user experience pengguna terhadap pelayanan belum sepenuhnya *positive*, selain itu dibutuhkan tutorial yang jelas bagi pengguna saat mengakses fitur-fitur yang tersedia, sementara itu dari segi aksesibilitas proses peminjaman yang kurang efisien atau keterbatasan akses ke koleksi yang diinginkan. Meskipun demikian, terdapat juga sentimen *positive* diantaranya, banyak pengguna yang memberikan apresiasi terhadap ketersediaan koleksi digital yang cukup lengkap serta kemudahan dalam mengakses koleksi bacaan dari berbagai perangkat. Selain itu, tampilan antarmuka aplikasi yang menarik serta responsif dan respon cepat dari layanan bantuan atau customer service juga menjadi nilai tambah yang disorot dalam ulasan *positive*. Beberapa pengguna juga menyampaikan bahwa *iBI Library* sangat membantu dalam menunjang kebutuhan literasi akademik dan referensi bacaan, terutama bagi mahasiswa dan dosen. Temuan ini menunjukkan bahwa meskipun masih terdapat tantangan dalam aspek teknis dan aksesibilitas, layanan *iBI Library* telah memberikan kontribusi yang besar bagi penggunanya.

Analisis sentimen bertujuan untuk mengidentifikasi dan memahami ekspresi seseorang yang disampaikan dalam bentuk opini. Opini tidak selalu disampaikan secara verbal, tetapi juga dapat diekspresikan melalui bentuk nonverbal. Salah satu bentuk opini nonverbal yang umum dijumpai adalah dalam bentuk teks atau tulisan. Ulasan pada penelitian ini diperoleh dari *Google Play Store* dengan ulasan pengguna saat menggunakan *iBI Library*, namun masih banyak yang tidak

memiliki kendala saat menggunakan aplikasi tersebut. Tujuan dari analisis sentimen adalah untuk mengklasifikasikan serta membedakan opini yang bersifat *positive* dan *negative*. Hasil klasifikasi ini dapat dimanfaatkan sebagai dasar evaluasi yang berguna untuk meningkatkan kepuasan pengguna, dengan mengoptimalkan opini *positive* sebagai potensi yang perlu dikembangkan, serta menjadikan opini *negative* sebagai masukan untuk perbaikan dan peningkatan kualitas untuk aplikasi *iBI Library*. Adapun pandangan islam, tujuan dari analisis sentiment ini adalah sunnah seperti yang dijelaskan pada ayat al-Qur'an Surat al-Baqarah ayat 42

وَلَا تَلْبِسُوا الْحَقَّ بِالْبَاطِلِ وَتَكْتُمُوا الْحَقَّ وَأَنْتُمْ تَعْلَمُونَ

“Janganlah kamu campuradukkan kebenaran dengan kebatilan<sup>21)</sup> dan (jangan pula) kamu sembunyikan kebenaran, sedangkan kamu mengetahui(-nya).”

Dalam Tafsir al-Misbah dijelaskan bahwa para tokoh agama dari kalangan Bani Israil melakukan upaya propaganda dengan menyampaikan pendapat-pendapat yang menyimpang dari ajaran asli dalam Kitab Taurat. Mereka melakukan manipulasi terhadap isi Taurat dengan menambahkan pandangan-pandangan pribadi serta menyembunyikan informasi tentang kenabian Muhammad SAW (Shihab, 2002). Meskipun menurut tafsir tersebut ayat ini secara historis ditujukan kepada kaum Bani Israil, agar tidak mencampuradukkan antara kebenaran dan kebatilan, yaitu dengan tidak menyisipkan ajaran yang bukan berasal dari firman Allah ke dalam Kitab Taurat. Selain itu, mereka juga dilarang menyembunyikan kebenaran, termasuk kabar mengenai kedatangan Nabi Muhammad SAW yang telah disebutkan dalam Taurat, padahal mereka mengetahuinya secara jelas. Tindakan penyembunyian ini dilakukan oleh sebagian kalangan Yahudi dengan tujuan untuk menghalangi manusia menerima kebenaran dan beriman kepada Nabi Muhammad SAW (Kemenag, 2022). Dalam konteks tujuan analisis sentimen, konsep klasifikasi yang bertujuan untuk mengidentifikasi serta membedakan antara opini yang bersifat *positive* dan *negative* agar tidak tercampur, hal ini selaras dengan makna yang terkandung dalam ayat tersebut.

Oleh karena itu, evaluasi dilakukan sebagai gambaran agar kondisi di masa mendatang dapat menjadi lebih baik. Evaluasi ini memiliki kesamaan makna

dengan konsep muhasabah, yaitu proses penilaian terhadap tindakan dengan mempertimbangkan nilai-nilai *positive* dan *negative* yang terkandung. Adapun implementasi muhasabah yang selaras dengan analisis sentimen.

**Table 4. 10 Implementasi Muhasabah Dalam Analisis Sentimen**

No	Konsep Muhasabah	Tafsir	Implementasi
1	Muhasabah Perilaku	"Wahai orang-orang yang beriman! Bertakwalah kepada Allah dan hendaklah setiap diri memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat); dan bertakwalah kepada Allah. Sungguh, Allah Maha Mengetahui terhadap apa yang kamu kerjakan." (QS. Al-Hasyr: 18)	Evaluasi tahap ke tahap dilakukan dengan baik dalam <i>pre-processing</i> hingga klasifikasi model
2	Muhasabah Proses	"Orang yang cerdas adalah yang menghisab dirinya dan beramal untuk kehidupan setelah mati. Sedangkan orang yang lemah adalah yang mengikuti hawa nafsunya dan hanya berangan-angan kepada Allah." (HR. At-Tirmidzi, No. 2459 – Hasan)	Evaluasi performa dari model klasifikasi sebagai cerminan efektivitas model seperti akurasi, presisi, <i>recall</i> dan <i>f-measure</i>
3	Muhasabah Sosial	Telah menceritakan kepada kami Abu Al-Yaman, dan Muhammad bin Mushab keduanya berkata: telah menceritakan kepada kami Abu Bakar bin Abdullah, dari Habib bin Ubaid berkata: Aisyah berkata: Rasulullah shallallahu 'alaihi wa sallam bersabda: "Pesimis adalah akhlaq yang buruk" (HR. Ahmad)	Melakukan tahap-tahap dengan benar tanpa kata noise dan meningkatkan score <i>confusion matrix</i>
4	Muhasabah Perbaikan	Sesungguhnya Allah tidak mengubah keadaan sesuatu kaum) artinya Dia tidak mencabut dari mereka nikmat-Nya (sehingga mereka mengubah keadaan yang ada pada diri mereka sendiri) dari keadaan yang baik dengan melakukan perbuatan durhaka. (QS. Ar-Ra'd)	Ketika model belum mencapai performa maksimal, maka dilakukan perbaikan untuk mendapatkan model yang sempurna dan sesuai yang diharapkan

Muhasabah dalam konteks hadist tersebut sebagai sunnah, Umar bin Khattab berkata bahwa mendorong manusia untuk tidak menunda-nunda perbaikan diri dan menyadari bahwa kesempatan muhasabah hanya ada di dunia. Jika seorang individu mengintropeksi diri sendiri dan mengevaluasi dirinya sendiri mulai dari

perkataan, amal dan tindakan yang dilakukan selama hidup, maka individu tersebut akan siap menghadapi pertanggung jawaban kepada Allah SWT dengan hati yang tenang (At-Tirmidzi, 1999).

## **BAB V KESIMPULAN DAN SARAN**

### **5.1 Kesimpulan**

Berdasarkan pada skenario pengujian yang telah dilakukan, performa terbaik pada rasio 80:20 dengan akurasi sebesar 83%, presisi sebesar 82%, *recall* sebesar 82%, *f-measure* sebesar 82%. Akurasi menggambarkan proporsi prediksi model yang sesuai dengan label sesungguhnya, sedangkan presisi menunjukkan tingkat ketepatan model dalam mengidentifikasi sentimen positif, *Recall* merepresentasikan kemampuan model dalam mengenali seluruh data yang relevan, dan *f-measure* menilai keseimbangan antara presisi dan recall dalam performa klasifikasi. Berdasarkan hal tersebut kinerja algoritma *Multinomial naïve bayes* dalam klasifikasi sentimen dapat dikategorikan unggul, stabil, dan teruji dalam membedakan sentimen *positive* dan *negative*, dengan performa terbaik pada skenario pembagian data 80:20. Dari hasil ulasan pengguna terhadap iBI Library yang diperoleh cukup bervariasi, dengan kecenderungan pada sentimen *negative*. Permasalahan yang dihadapi diantaranya waktu proses transaksi koleksi sering terjadi error saat melakukan peminjaman, dan navigasi sistem yang kurang dimengerti oleh pengguna baru, serta minimnya panduan penggunaan aplikasi. Selain itu, beberapa pengguna juga mengeluhkan efisiensi peminjaman dan keterbatasan akses koleksi. Meskipun demikian, hasil analisis juga menunjukkan adanya sentimen *positive*, khususnya terkait kelengkapan koleksi digital, kemudahan akses dari berbagai perangkat, tampilan antarmuka yang menarik, serta layanan bantuan yang responsif.

### **5.2 Saran**

Berdasarkan hasil dari penelitian ini, terdapat sejumlah rekomendasi yang dapat diajukan, diantaranya sebagai berikut:

1. Untuk penelitian selanjutnya pada tahapan *pre-processing* perlu penanganan khusus untuk kata yang masih lolos, termasuk saat penanganan kata tidak baku, kata slang, kesalahan pengetikan, dan kata-

kata *noise*, serta menambahkan beberapa skenario perbandingan untuk hasil yang lebih bervariasi

2. Bagi pengelola *iBI Library* disarankan lebih memperhatikan kendala pengguna yang pada ulasan negative, diantaranya saat transaksi peminjaman koleksi digital, serta sistem navigasi antarmuka agar lebih mudah dipahami oleh seluruh kategori pengguna, termasuk pengguna baru.

## DAFTAR PUSTAKA

- a. Oktian Permana, & Sudin Saepudin. (2023). Perbandingan Algoritma K-Nearest Neighbor Dan *Naïve bayes* Pada Aplikasi Shopee. *Jurnal Coscitech (Computer Science And Information Technology)*, 4(1), 25–32. <https://doi.org/10.37859/Coscitech.V4i1.4474>
- Agung, A., Daniswara, A., & Nuryana, I. K. D. (2023). *Data Pre-processing Pola Pada Penilaian Mahasiswa Program Profesi Guru*. 05, 97–100.
- Ahmad, I. (N.D.). *Musnad Imam Ahmad*.
- Al-Ghazali. (2009). *Ihya Ulumuddin (Kitab Muhasabah Al-Nafs)*.
- Alfandi Safira, & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *Zonasi: Jurnal Sistem Informasi*, 5(1), 59–70. <https://doi.org/10.31849/Zn.V5i1.12856>
- Arham, A., Swedia, E. R., Cahyanti, M., & Septian, M. R. D. (2022). Implementasi Sentiment Analysis Pada Opini Masyarakat Indonesia Di Twitter Terhadap Virus Covid-19 Varian Omicron Dengan Algoritma *Naïve bayes*, Decision Tree, Dan Support Vector Machine. *Sebatik*, 26(2), 565–572. <https://doi.org/10.46984/Sebatik.V26i2.1961>
- As-Suyuthi, J., Muhammad, J. I., & Al-Mahally, A. (2017). Tafsir Al-Jalalain. *Kalimah*, 15(1), 1. <https://doi.org/10.21111/Klm.V1i15.818>
- Ashari, H., Arifianto, D., Azizah, H., & Faruq, A. (2020). Perbandingan Kinerja Algoritma Multinomial Naive Bayes (Mnb, Multivariate Bernoulli Dan Rocchio Algoritim Dalam Klasifikasi Konten Berita Hoax Berbahasa Indonesia Pada Media Sosial. <http://repository.unmuhjember.ac.id>, 1–12.
- At-Tirmidzi, A. I. M. Bi Isa. (1999). *Sunan At-Tirmidzi*. Dar Al-Hadith. <https://sunnah.com/Tirmidhi:2459>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/Jtk.V15i1.744>
- Duarte, A. L. D. Tavares, & Eddy Nurraharjo. (2023). Analisis Sentimen Dan Klasifikasi Tweet Terkait Naiknya Kasus Omicron Menggunakan Naive Bayes Classifier. *Jurnal Informatika Dan Rekayasa Elektronik*, 6(1), 1–8. <https://doi.org/10.36595/Jire.V6i1.779>
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Comparison Of *Naïve bayes* And Support Vector Machine Methods In Twitter Sentiment Analysis. *Smatika Jurnal*, 10(02), 71–76.

- Ginantra, N. L. W. S. R., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., Ardiana, D. P. Y., Effendy, F., Iskandar, A., Hazriani, H., Sari, I. Y., Gustiana, Z., Prianto, C., Gustian, D., & Negara, E. S. (2021). *Data Mining Dan Penerapan Algoritma*.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Hasanah, S. A. (2018). Konsep Muhasabah Dalam Al-Qur'an Telaah Pemikiran Al-Ghazali. *Jurnal Al-Dirayah*, 1 (1)(1), 57–65.
- Hidayah, N., & Dodiman. (2024). Implementasi Algoritma *Multinomial naïve bayes*, Tf-Idf Dan *Confusion matrix* Dalam Pengklasifikasian Saran Monitoring Dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin. *Jurnal Akademik Pendidikan Matematika*, 10(1), 8–15.
- Indriani, S., & Syafrullah, M. (2022). *Multinomial naïve bayes* Untuk Menganalisis Sentimen Layanan Jasa Ekspedisi Sicepat Ekspres. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (Senafti) Jakarta-Indonesia, September*, 445–453. <https://developer.twitter.com/>
- Jumasa, H. M. (2019). Rancang Bangun Digital *Library* Berbasis Mobile (Studi Kasus : Universitas Muhammadiyah Purworejo). *Intek : Jurnal Informatika Dan Teknologi Informasi*, 2(1), 32–38. <https://doi.org/10.37729/intek.v2i1.87>
- Kemenag. (N.D.). *Al Hasyr: 18*. Qurn Kemenag. <https://quran.nu.or.id/al-hasyr/18>
- Kemenag. (2022). *Tafsir Al-Baqarah*. Qurn Kemenag. <https://quran.kemenag.go.id/quran/per-ayat/surah/2?from=42&to=286>
- Liu, H., Chen, X., & Liu, X. (2022). A Study Of The Application Of Weight Distributing Method Combining Sentiment Dictionary And Tf-Idf For Text Sentiment Analysis. *Ieee Access*, 10, 32280–32289. <https://doi.org/10.1109/Access.2022.3160172>
- 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. <https://doi.org/10.18860/mat.v12i1.8973>
- Marisa, F. (2005). Data Mining, Konsep Dan Penerapannya. *Mining Of Massive Datasets*, 2(January 2013), 5–20. [https://www.cambridge.org/core/product/identifier/Cbo9781139058452a007/type/book\\_part](https://www.cambridge.org/core/product/identifier/Cbo9781139058452a007/type/book_part)

- Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi MyPertamina Pada *Google Play Store* Menggunakan Algoritma Nbc. *Jurnal Teknologi Terpadu*, 9(1), 42–48. <https://doi.org/10.54914/jtt.v9i1.609>
- Nadira, A., Setiawan, N. Y., & Purnomo, W. (2023). Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode *Naïve bayes* Dengan Kamus Inset. *Indexia*, 5(01), 35. <https://doi.org/10.30587/indexia.v5i01.5138>
- Natalius, S. (2011). Metoda *Naïve bayes* Classifier Dan Penggunaannya Pada Klasifikasi Dokumen. *Jurnal Sistem Informasi Sekolah Tinggi Elektro Dan Informatika Institut Teknologi Bandung*, 3, 1–5.
- Nugraha, T., Purwantoro, & Umaidah, Y. (2022). Analisis Sentimen Terhadap Perpanjangan Masa Jabatan Presiden Indonesia Menggunakan Algoritma *Naïve bayes*. *Jurnal Pendidikan Dan Konseling*, 4(4), 6625–6635.
- Pasaribu, S., Rohani, L., & Faisal, M. (2024). Persepsi Pengguna Aplikasi Digital Perpustakaan Bank Indonesia “*IBI Library*” Berbasis Android Berdasarkan Teori Usability. *Triwikrama: Jurnal Ilmu Sosial*, 3(7).
- Perdana, A., Hermawan, A., & Avianto, D. (2022). Analisis Sentimen Terhadap Isu Penundaan Pemilu Di Twitter Menggunakan Naive Bayes Clasifier. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(2), 195–200. <https://doi.org/10.32736/sisfokom.v11i2.1412>
- Prakash, T. N., & Aloysius, A. (2021). Textual Sentiment Analysis Using Lexicon Based Approaches. *Annals Of The Romanian Society For Cell Biology*, 25(4), 9878–9885. <http://annalsofrscb.ro/index.php/journal/article/view/3734>
- Purnamasari, D., Bayu, A., Desy, A., Fanka, W. A. P., Reza, A., Safrila, M., Yanda, O. N., & Hidayati, U. (2023). Pengantar Metode Analisis Sentimen. In *Gunadarma Penerbit*.
- Puspala Sari, R. A., & Jaya, I. D. (2023). Penerapan Algoritma *Naïve bayes* Untuk Analisis Sentimen Pengguna My Jne. *Jurnal Multidisiplin West Science*, 2(03), 215–221. <https://doi.org/10.58812/jmws.v2i03.269>
- Ramadhani, B., & Suryono, R. R. (2024). Komparasi Algoritma *Naïve bayes* Dan Logistic Regression Untuk Analisis Sentimen Metaverse. *Jurnal Media Informatika Budidarma*, 8(2), 714. <https://doi.org/10.30865/mib.v8i2.7458>
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *Inovtek Polbeng - Seri Informatika*, 3(1), 50.
- Rifqi, M. (2019). Evaluasi Usability Dan Rekomendasi Perbaikan Tampilan Aplikasi *IBI Library* Menggunakan Metode Usability Testing. *Jurnal*

- Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(10), 9832–9841.
- Riyanto, U. (2019). Analisis Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online. *Jika (Jurnal Informatika)*, 2(2), 62–72. <https://doi.org/10.31000/V2i2.1521>
- Septiani, A., & Budi, I. (2022). Klasifikasi Ulasan Pengguna Aplikasi: Studi Kasus Aplikasi Ipusnas Perpustakaan Nasional Republik Indonesia (Pnri). *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 7(4), 1110–1120. <https://doi.org/10.29100/Jipi.V7i4.3216>
- Shihab, M. Q. (2002). Tafsir Al Mishbah : Pesan, Kesan Dan Keserasian Al-Qur'an. In *Paper Knowledge . Toward A Media History Of Documents* (Volume 6). Lentera Hati Group.
- Shihab, M. Q. (2016). *Yang Hilang Dari Kita: Akhlak* (S. N. Andini (Ed.)). Lentera Hati Group.
- Styawati, S., Isnain, A. R., Hendrastuty, N., & Andraini, L. (2021). Comparison Of Support Vector Machine And *Naive bayes* On Twitter Data Sentiment Analysis. *Jurnal Informatika: Jurnal Pengembangan It*, 6(1), 56–60. <https://doi.org/10.30591/Jpit.V6i1.3245>
- Surya, P. P. M., Seetha, L. V., & Subbulakshmi, B. (2019). Analysis Of User Emotions And Opinion Using *Multinomial naive bayes* Classifier. *Proceedings Of The 3rd International Conference On Electronics And Communication And Aerospace Technology, Iceca 2019*, 410–415. <https://doi.org/10.1109/Iceca.2019.8822096>
- Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma *Naive bayes* Dan Knn. *Jurnal Komtekinfo*, 10, 1–7. <https://doi.org/10.35134/Komtekinfo.V10i1.330>
- Syahril, M., Kusnasari, S., Muhazir, A., & Syahputri, A. (2023). Jurnal Teknologi Sistem Informasi Dan Sistem Komputer Tgd Implementasi Data Mining Untuk Rekomendasi Jurusan Menggunakan Algoritma K-Means Clustering Jurnal Teknologi Sistem Informasi Dan Sistem Komputer Tgd. *Teknologi Sistem Informasi Dan Sistem Komputer Tgd*, 6, 235–245.
- Wahyu, Wijaya Widiyanto, M., Munsarif, M., Stenly Richard Pungus, A. W. G. F. M., Wiranti Kusuma Hapsari, S. A. H., Pasnur, E. F. B. A. F., & Suryani, N. S. M. A. (2023). *Pengenalan Data Mining* (Issue July).
- Wahyuni, A. S., & Dewi, A. O. . (2019). Persepsi Pemustaka Terhadap Desain Antarmuka Pengguna (User Interface) Aplikasi Perpustakaan Digital "Ijogja" Berbasis Android. *Jurnal Ilmu Perpustakaan*, Vii(1), 21–30. <https://ejournal3.undip.ac.id/index.php/jip/article/viewfile/22810/20855>

- Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical *Machine learning Tools And Techniques* - Pages 607-629. *Data Mining: Practical Machine learning Tools And Techniques*, 607–629. <https://doi.org/10.1016/B978-0-12-374856-0.00024-9>
- Yulianto, Y. (2017). Mabadiâ€™™ Asyroh Nalar Fikih Sosial Ali Yafie. *Shahih: Journal Of Islamicate Multidisciplinary*, 2(1), 21–36. <https://doi.org/10.22515/Shahih.V2i1.725>
- Yulianto, Y., & Suprpto, A. (2023). Pandangan Islam Terhadap Pengembangan Dan Pemanfaatan Sains Dan Teknologi. *Es-Syajar: Journal Of Islam, Science And Technology Integration*, 1(1), 1–26. <https://doi.org/10.18860/Es.V1i1.20423>
- Zafran, M. A., Data, M., & Fauzi, M. A. (2020). *Implementasi Sistem Pengujian Kerentanan Denial Of Service ( Dos ) Pada Web Berbasis Wordpress*. 1(1), 1–10.

## LAMPIRAN

### Lampiran 1 Surat izin penelitian



**KEMENTERIAN AGAMA**  
**UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG**  
**FAKULTAS SAINS DAN TEKNOLOGI**  
 Jalan Gajayana 50 Malang 65144 Telepon/Faksimile (0341) 558933  
 Website: <http://saintek.uin-malang.ac.id>, email: [saintek@uin-malang.ac.id](mailto:saintek@uin-malang.ac.id)

Nomor : B-41.O/FST.01/TL.00/03/2025  
 Lampiran : -  
 Hal : Permohonan Penelitian

Yth. Pimpinan Perpustakaan Bank Indonesia  
 Bank Indonesia, Menara Syafruddin Prawiranegara Jl. M.H. Thamrin No.2, RT.2/RW.3,  
 Gambir, Kecamatan Gambir, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10350

Dengan hormat,  
 Sehubungan dengan penelitian mahasiswa Jurusan Perpustakaan dan Sains Informasi  
 Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang atas nama:

Nama : Rohimullah At Thobroni  
 NIM : 210607110039  
 Judul Penelitian : Analisis Sentimen Pengguna Pada Layanan Aplikasi iBI Library Pada  
 Ulasan Google Play Store Dengan Multinomial Naive Bayes.  
 Dosen Pembimbing : FAKHRIS KHUSNU REZA MAHFUD, M.Kom

Maka kami mohon Bapak/Ibu berkenan memberikan izin pada mahasiswa tersebut untuk  
 melakukan penelitian di Perpustakaan Bank Indonesia dengan waktu pelaksanaan pada  
 tanggal 20 April 2025 sampai dengan 20 Mei 2025.

Demikian permohonan ini, atas perhatian dan kerjasamanya disampaikan terimakasih.

Malang, 07 Maret 2025  
 a.n Dekan

Scan QRCode ini



untuk verifikasi surat



Rektor Dekan Bidang Akademik,

Dr. Anton Prasetyo, M.Si  
 NIP. 19770925 200604 1 003

## Lampiran 2

### Cleaning

```
[ ] def clean_twitter_text(text):
    text = re.sub(r'@[A-Za-z0-9_]+', '', text)
    text = re.sub(r'#\w+', '', text)
    text = re.sub(r',RT[\s]+', '', text)
    text = re.sub(r'https?://\S+', '', text)
    text = re.sub(r'^[A-Za-z0-9 ]', '', text)
    text = re.sub(r'\s+', '', text).strip()
    return text

df['Cleaning'] = df['Content'].apply(clean_twitter_text)
df.head()
```



	userName	Content	Labeling	Cleaning
0	Adagio Anakrus	kesulitan saat meminjam buku	negatif	kesulitan saat meminjam buku
1	Sintya Dewi	Pada menu search buku ada filter pengarang. To...	positif	Pada menu search buku ada filter pengarang Tol...
2	Reysya Yudiantingsih	jumlah copy nya sikit amat gak niat ngasih pin...	negatif	jumlah copy nya sikit amat gak niat ngasih pin...
3	senja hayu	ini lagi eror kah kok ngga bisa login yaa	negatif	ini lagi eror kah kok ngga bisa login yaa
4	Jay	Tidak bisa login menggunakan akun google, tolo...	negatif	Tidak bisa login menggunakan akun google tolon...

## Pre-processing

### Case folding

Case Folding

```
df['case_folding'] = df['cleaning'].str.lower()
df.head()
```

	userName	Content	Labeling	Cleaning	Case Folding
0	Adagio Anakrus	kesulitan saat meminjam buku	negatif	kesulitan saat meminjam buku	kesulitan saat meminjam buku
1	Sintya Dewi	Pada menu search buku ada filter pengarang. To...	positif	Pada menu search buku ada filter pengarang Tol...	pada menu search buku ada filter pengarang tol...
2	Reysya Yusalingsih	jumlah copy nya sikit amat gak niat ngasih pin...	negatif	jumlah copy nya sikit amat gak niat ngasih pin...	jumlah copy nya sikit amat gak niat ngasih pin...
3	senja hayu	ini lagi eror kah kok ngga bisa login yaa	negatif	ini lagi eror kah kok ngga bisa login yaa	ini lagi eror kah kok ngga bisa login yaa
4	Jay	Tidak bisa login menggunakan akun google, tolo...	negatif	Tidak bisa login menggunakan akun google tolon...	tidak bisa login menggunakan akun google tolon...

## Normalisasi

```

"gak": "tidak", "dr": "dari", "yg": "yang", "lu": "kamu", "sya": "saya",
"lancarr": "lancar", "kayak": "kayak", "ngawur": "ngawur", "sebaran": "sebaran", "k": "ke", "memori": "memori",
"luass": "luas", "sy": "saya", "thn": "tahun", "males": "males", "trs": "trs", "terus": "terus", "pdhl": "padahal", "n": "n", "bgt": "bangget",
"tgl": "tanggal", "lg": "lagi", "bgt": "bangget", "gua": "saya", "n": "n", "tpi": "tapi", "standar": "standart", "biasa": "biasa", "sdi": "sudah", "n": "n", "dan",
"sikit": "sedikit", "tetep": "tetap", "knp": "kenapa", "utk": "untuk", "jd": "jadi", "kmarin": "kemarin", "bgt": "bangget", "tpi": "tapi", "trus": "terus", "login": "masuk",
"skng": "sekarang", "padahal": "padahal", "semejaj": "semejaj", "g": "ga", "aj": "aja", "tdk": "tidak", "tsb": "tersebut", "mulu": "terus", "masi": "masih",
"sy": "saya", "bs": "bisa", "hadeuuuh": "aduh", "apk": "aplikasi", "pinjem": "pinjam", "ringkes": "ringkas", "app": "aplikasi", "kek": "kayak", "ttp": "tetap", "bgt": "bangget", "karna": "karna"

def normalize_text(text):
    words = text.split()
    normalized = [norm.get(w, w) for w in words]
    return ' '.join(normalized)

df['normalisasi'] = df['Case Folding'].apply(lambda x: normalize_text(x))
df.head(10)

```

	userName	Content	Labeling	Cleaning	Case Folding	normalisasi
0	Adagio Anakrus	kesulitan saat meminjam buku	negatif	kesulitan saat meminjam buku	kesulitan saat meminjam buku	kesulitan saat meminjam buku
1	Sintya Dewi	Pada menu search buku ada filter pengarang. To...	positif	Pada menu search buku ada filter pengarang To...	pada menu search buku ada filter pengarang to...	pada menu search buku ada filter pengarang to...
2	Reysya Yusulianingsih	Jumlah copy nya sikit amat gak niat ngasih pin...	negatif	jumlah copy nya sikit amat gak niat ngasih pin...	jumlah copy nya sikit amat gak niat ngasih pin...	jumlah copy nya sedikit amat gak niat ngasih pin...
3	senja hayu	ini lagi eror kah kok ngga bisa login yaa	negatif	ini lagi eror kah kok ngga bisa login yaa	ini lagi eror kah kok ngga bisa login yaa	ini lagi eror kah kok ngga bisa masuk yaa
4	Jay	Tidak bisa login menggunakan akun google, toto...	negatif	Tidak bisa login menggunakan akun google toto...	tidak bisa login menggunakan akun google tolon...	tidak bisa masuk menggunakan akun google tolon...
5	Gonita Artisa	tolong kenapa saya jdi tdk bisa meminjam buku ...	negatif	tolong kenapa saya jdi tdk bisa meminjam buku ...	tolong kenapa saya jdi tdk bisa meminjam buku ...	tolong kenapa saya jadi tidak bisa meminjam bu...

## Stopword

```

# are also updated with custom stopwords if needed for this specific function.
stop_words_nltk = set(stopwords.words('indonesian')) # Get nltk stopwords
stop_words_nltk.update(custom_stopwords) # Update nltk stopwords with custom ones
words = text.split()
filtered_words = [word for word in words if word.casefold() not in stop_words_nltk]
cleaned_text = ' '.join(filtered_words)
return cleaned_text

df['stopword'] = df['normalisasi'].apply(lambda x: stopwords(x))
df.head(10)

```

[nltk\_data] Downloading package stopwords to /root/nltk\_data...  
[nltk\_data] package stopwords is already up-to-date!

	userName	Content	Labeling	Cleaning	Case Folding	normalisasi	stopword
0	Adagio Anakrus	kesulitan saat meminjam buku	negatif	kesulitan saat meminjam buku	kesulitan saat meminjam buku	kesulitan saat meminjam buku	kesulitan meminjam buku
1	Sintya Dewi	Pada menu search buku ada filter pengarang. To...	positif	Pada menu search buku ada filter pengarang. Tol...	pada menu search buku ada filter pengarang to...	kesulitan saat meminjam buku ada filter pengarang to...	menu search buku filter pengarang tolong menu ...
2	Reysya Yusdianingsih	Jumlah copy nya sikit amat gak niat ngasih pin...	negatif	jumlah copy nya sikit amat gak niat ngasih pin...	jumlah copy nya sikit amat gak niat ngasih pin...	jumlah copy nya sedikit amat gak niat ngasih p...	copy gak niat ngasih pinjam buku
3	senja hayu	Ini lagi error kah kok ngga bisa login yaa	negatif	Ini lagi error kah kok ngga bisa login yaa	Ini lagi error kah kok ngga bisa login yaa	Ini lagi error kah kok ngga bisa masuk yaa	error ngga masuk
4	Jay	Tidak bisa login menggunakan akun google, tolo...	negatif	Tidak bisa login menggunakan akun google tolon...	tidak bisa login menggunakan akun google tolon...	tidak bisa masuk menggunakan akun google tolon...	masuk akun google tolong diperbaiki
5	Qonita Arisla	tolong kenapa saya jdi tdk bisa meminjam buku ...	negatif	tolong kenapa saya jdi tdk bisa meminjam buku ...	tolong kenapa saya jdi tdk bisa meminjam buku ...	tolong kenapa saya jadi tidak bisa meminjam bu...	tolong meminjam buku bukunya tersedia dipinjam...
6	Yukou yu092	gabisa dibaca, setelah klik pinjam disuruh rab	negatif	gabisa dibaca setelah klik pinjam disuruh rabah	gabisa dibaca setelah klik pinjam disuruh rabah	gabisa dibaca setelah klik pinjam disuruh rabah	gabisa dibaca klik pinjam disuruh rabah sudah

# Tokenized

```
def tokenize(text):
    return text.split()

df['tokenized'] = df['stopword'].apply(lambda x: tokenize(x))
df.head(10)
```

	userName	Content	Labeling	Cleaning	Case Folding	normalisasi	stopword	tokenized
0	Adagio Anakrus	kesulitan saat meminjam buku	negatif	kesulitan saat meminjam buku	kesulitan saat meminjam buku	kesulitan saat meminjam buku	kesulitan meminjam buku	[kesulitan, meminjam, buku]
1	Sinya Dewi	Pada menu search buku ada filter pengarang. To...	positif	Pada menu search buku ada filter pengarang To...	pada menu search buku ada filter pengarang to...	pada menu search buku ada filter pengarang to...	menu search buku filter pengarang tolong menu ...	[menu, search, buku, filter, pengarang, tolong, ...]
2	Revsya Yudiantingsih	Jumlah copy nya sedikit amat gak niat ngasih pin...	negatif	Jumlah copy nya sedikit amat gak niat ngasih pin...	Jumlah copy nya sedikit amat gak niat ngasih pin...	Jumlah copy nya sedikit amat gak niat ngasih p...	copy gak niat ngasih pinjam buku	[copy, gak, niat, ngasih, pinjam, buku]
3	senja hayu	ini lagi eror kah kok ngga bisa login yaa	negatif	ini lagi eror kah kok ngga bisa login yaa	ini lagi eror kah kok ngga bisa login yaa	ini lagi eror kah kok ngga bisa masuk yaa	eror ngga masuk	[eror, ngga, masuk]
4	Jay	Tidak bisa login menggunakan akun google, tolo...	negatif	Tidak bisa login menggunakan akun google tolon...	tidak bisa masuk menggunakan akun google tolon...	tidak bisa masuk menggunakan akun google tolon...	masuk akun google tolong diperbaiki	[masuk, akun, google, tolong, diperbaiki]
5	Qonita Arista	tolong kenapa saya jdi tdk bisa meminjam buku ...	negatif	tolong kenapa saya jdi tdk bisa meminjam buku ...	tolong kenapa saya jadi tidak bisa meminjam bu...	tolong kenapa saya jadi tidak bisa meminjam bu...	tolong meminjam buku bukunya tersedia di pinjam...	[tolong, meminjam, buku, bukunya, tersedia, di, ...]
6	Yukou yu092	gabisa dibaca, setelah klik pinjam disuruh gabu...	negatif	gabisa dibaca setelah klik pinjam disuruh gabu...	gabisa dibaca setelah klik pinjam disuruh gabu...	gabisa dibaca setelah klik pinjam disuruh gabu...	gabisa dibaca klik pinjam disuruh gabung udah ...	[gabisa, dibaca, klik, pinjam, disuruh, gabung, ...]
7	Aqilya Putri	tibaA? gk bisa minjem sm sekali, selalu eror	negatif	tiba gk bisa minjem sm sekali selalu eror	tiba ga bisa minjem sm sekali selalu eror	tiba ga bisa minjem sama sekali selalu eror	ga minjem eror	[ga, minjem, eror]

## Stemming

```

do.append(dt)
d_clean = ".join(do)
# print(d_clean) # You can keep this for debugging if needed
return d_clean

# Apply the stemming function to the 'tokenized' column and create a new column 'stemmed_text'
df['stemmed_text'] = df['tokenized'].apply(stemming)

#df['stemmed_text'].to_csv('content/drive/myDrive/data skripsi/stemmed_text.csv', index=False)
# Display the head of the DataFrame to check the new column
df.head(10)

```

	userName	Content	Labeling	Cleaning	Case Folding	normalisasi	stopword	tokenized	stemmed_text
0	Adagio Anakrus	kesulitan saat meminjam buku	negatif	kesulitan saat meminjam buku	kesulitan saat meminjam buku	kesulitan saat meminjam buku	kesulitan meminjam buku	[kesulitan, meminjam, buku]	suit pinjam buku
1	Sinya Dewi	Pada menu search buku ada filter pengarang. To...	positif	Pada menu search buku ada filter pengarang To...	pada menu search buku ada filter pengarang to...	pada menu search buku ada filter pengarang to...	menu search buku filter pengarang tolong menu ...	[menu, search, buku, filter, pengarang, tolong...]	menu search buku filter arang tolong menu sear...
2	Reysya Yusdianingsih	jumlah copy nya sedikit amat gak niat ngasih pin...	negatif	jumlah copy nya sedikit amat gak niat ngasih pin...	jumlah copy nya sedikit amat gak niat ngasih pin...	jumlah copy nya sedikit amat gak niat ngasih p...	copy gak niat ngasih pinjam buku	[copy, gak, niat, ngasih, pinjam, buku]	copy gak niat ngasih pinjam buku
3	senja hayu	ini lagi eror kah kok ngga bisa login yaa	negatif	ini lagi eror kah kok ngga bisa login yaa	ini lagi eror kah kok ngga bisa login yaa	ini lagi eror kah kok ngga bisa masuk yaa	eror ngga masuk	[eror, ngga, masuk]	eror ngga masuk
4	Jay	Tidak bisa login menggunakan akun google, tolo...	negatif	Tidak bisa login menggunakan akun google tolon...	tidak bisa login menggunakan akun google tolon...	tidak bisa masuk menggunakan akun google tolon...	masuk akun google tolong diperbaiki	[masuk, akun, google, tolong, diperbaiki]	masuk akun google tolong baik

## TF-IDF

```

def compute_tfidf(tf):
    vocab_index = {}
    for word in vocab:
        vocab_index[word] = len(vocab_index) + 1

    tfidf_vectors = []
    for doc in tf_docs:
        tfidf_vectors.append(compute_tfidf(doc))

    # Hitung semua TF-IDF vector
    tfidf_vectors = [compute_tfidf(tf) for tf in tf_docs]

    # Tampilkan hasil TF-IDF
    print("Vocab:", vocab)
    for i, vec in enumerate(tfidf_vectors):
        print(f"\nDokumen {i+1} ({labels[i]}):")
        for j, weight in enumerate(vec):
            if weight > 0:
                print(f"  {vocab[j]:<15} : {weight:.4f}")

Vocab: ['pinjam', 'buku', 'sulit', 'mudah', 'cari', 'tolong', 'filter', 'arang', 'menyearch']

Dokumen 1 (negatif):
pinjam      : 2.3828
buku        : 1.7107
sulit       : 5.5609

Dokumen 2 (positif):
buku        : 1.7107
mudah       : 4.7136
cari        : 4.1746
tolong      : 3.1331
filter      : 5.5609
arang       : 15.1501
menu        : 11.1217
search      : 10.5464

```

## TRAINING

```

training

class_tf = {
    "positif": defaultdict(int),
    "negatif": defaultdict(int)
}

class_total_terms = {
    "positif": 0,
    "negatif": 0
}

class_doc_count = {
    "positif": 0,
    "negatif": 0
}

for tf, label in zip(tf_docs, labels): # pakai tf_docs dari proses sebelumnya
    class_doc_count[label] += 1
    for word, count in tf.items():
        class_tf[label][word] += count
        class_total_terms[label] += count

# Hitung prior dan likelihood (Multinomial Naive Bayes)
model = {}
V = len(vocab) # jumlah kata unik (ukuran vocab)

for label in class_tf:
    log_likelihood = {}
    for word in vocab:

```

## MODEL

```

model

def predict(doc_tokens):
    tf = defaultdict(int)
    for word in doc_tokens:
        tf[word] += 1

    scores = {}
    for label in model:
        score = model[label]['log_prior']
        for word, count in tf.items():
            if word in model[label]['log_likelihood']:
                score += count * model[label]['log_likelihood'][word]
        scores[label] = score

    return max(scores, key=scores.get)

```

## TESTING

```

    scores[label] = score
    return max(scores, key=scores.get)

# --- Testing Phase ---
y_pred = [predict(doc) for doc in X_test]

# --- Evaluation ---
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred, labels=["positif", "negatif"]))

print("\nAccuracy:", accuracy_score(y_test, y_pred))

print("\nClassification Report:")
print(classification_report(y_test, y_pred, labels=["positif", "negatif"]))

```

```

↔ Confusion Matrix:
[[21  9]
 [ 5 51]]

```

```
Accuracy: 0.8372093023255814
```

```
Classification Report:
```

	precision	recall	f1-score	support
positif	0.81	0.70	0.75	30
negatif	0.85	0.91	0.88	56
accuracy			0.84	86
macro avg	0.83	0.81	0.81	86
weighted avg	0.84	0.84	0.83	86

```

        scores[label] = score
    return max(scores, key=scores.get)

# --- Testing Phase ---
y_pred = [predict(doc) for doc in X_test]

# --- Evaluation ---
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred, labels=["positif", "negatif"]))

print("\nAccuracy:", accuracy_score(y_test, y_pred))

print("\nClassification Report:")
print(classification_report(y_test, y_pred, labels=["positif", "negatif"]))

```

 Confusion Matrix:  
[[17 5]  
[ 5 31]]

Accuracy: 0.8275862068965517

Classification Report:

	precision	recall	f1-score	support
positif	0.77	0.77	0.77	22
negatif	0.86	0.86	0.86	36
accuracy			0.83	58
macro avg	0.82	0.82	0.82	58
weighted avg	0.83	0.83	0.83	58

```

X_test_counts = vectorizer.transform(X_test_raw)
y_pred_raw = model.predict(X_test_counts)

# --- Evaluation ---
print("\n--- Results using test_data_split.csv (raw data) ---")
print("Confusion Matrix:")
print(confusion_matrix(y_test_raw, y_pred_raw, labels=["positif", "negatif"]))

print("\nAccuracy:", accuracy_score(y_test_raw, y_pred_raw))

print("\nClassification Report:")
print(classification_report(y_test_raw, y_pred_raw, labels=["positif", "negatif"]))

```

 --- Results using test\_data\_split.csv (raw data) ---  
Confusion Matrix:  
[[ 9 19]  
[ 1 27]]

Accuracy: 0.6428571428571429

Classification Report:

	precision	recall	f1-score	support
positif	0.90	0.32	0.47	28
negatif	0.59	0.96	0.73	28
accuracy			0.64	56
macro avg	0.74	0.64	0.60	56
weighted avg	0.74	0.64	0.60	56

## DATA TEST

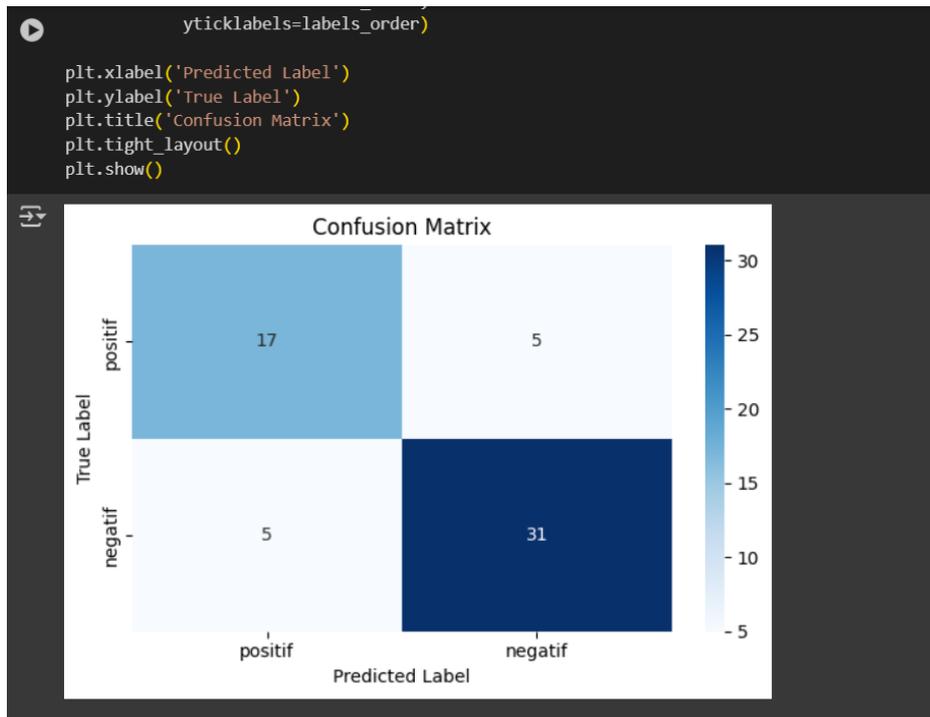
```
# Buat DataFrame hasil
hasil = pd.DataFrame(proba, columns=model.classes_)
hasil['Predicted'] = predicted
hasil['Actual'] = y_test.reset_index(drop=True)
hasil['Content'] = text_test.reset_index(drop=True)

# Tampilkan hasil
display(hasil[['Content', 'negatif', 'positif', 'Predicted', 'Actual']])
hasil.to_csv('/content/drive/MyDrive/data skripsi/hasil_predicted.csv', index=False)
```

	Content	negatif	positif	Predicted	Actual
0	bukunya kok ga bisa dipinjem ya? padahal masi ...	0.861856	0.138144	negatif	negatif
1	Tidak bisa login, saya sudah daftar, pas masuk...	0.878158	0.121842	negatif	negatif
2	Baru install dan daftar, udah pinjam dan diund...	0.852305	0.147695	negatif	negatif
3	Overall bagus. Tapi mungkin bisa ditambah lagi...	0.526197	0.473803	negatif	positif
4	Saya sangat senang dengan aplikasi ini namun a...	0.692850	0.307150	negatif	negatif

## CONFUSION MATRIX





## TURNITIN

### Sidang Skripsi Oni Fikss bgt.pdf

#### ORIGINALITY REPORT

<b>5%</b>	<b>5%</b>	<b>2%</b>	<b>2%</b>
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

#### PRIMARY SOURCES

<b>1</b>	<a href="https://etheses.uin-malang.ac.id">etheses.uin-malang.ac.id</a> Internet Source	<b>3%</b>
<b>2</b>	Nada Maula I.W, Zidna Fariha, Kafhaya Nuzulanisa, Sri Ari Wulandari. "Kebhinekaan Dalam Budaya Perspektif Tafsir Kementerian Agama Republik Indonesia", Moderatio: Jurnal Moderasi Beragama, 2023 Publication	<b>&lt;1%</b>
<b>3</b>	<a href="https://repo.undiksha.ac.id">repo.undiksha.ac.id</a> Internet Source	<b>&lt;1%</b>
<b>4</b>	Submitted to Argosy University Student Paper	<b>&lt;1%</b>
<b>5</b>	<a href="https://digilib.uinsa.ac.id">digilib.uinsa.ac.id</a> Internet Source	<b>&lt;1%</b>
<b>6</b>	Submitted to Fakultas Ekonomi Universitas Indonesia Student Paper	<b>&lt;1%</b>
<b>7</b>	<a href="https://eprints.umpo.ac.id">eprints.umpo.ac.id</a> Internet Source	<b>&lt;1%</b>