

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

SKRIPSI



**Oleh:
INTAN PRAMUDA WARDANI
NIM. 19680055**

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

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

SKRIPSI

**Oleh:
INTAN PRAMUDA WARDANI
NIM. 19680055**

**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

LEMBAR PERSETUJUAN

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

SKRIPSI

Oleh:

INTAN PRAMUDA WARDANI

NIM. 19680055

Telah Diperiksa dan Disetujui

Tanggal 24 Juni 20225

Pembimbing 1



Fakhris Khusnu 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

LEMBAR PENGESAHAN

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

SKRIPSI

Oleh:

INTAN PRAMUDA WARDANI

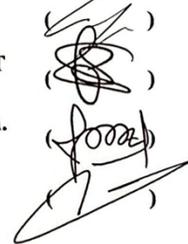
NIM. 19680055

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 : Wahyu Hariyanto, M.M
NIP. 198907212019031007
Anggota Penguji I : Dr. Ir. Mokhamad Amin Hariyadi, M.T
NIP. 196701182005011001
Anggota Penguji II : Fakhris Khusnu Reza Mahfud, M.Kom.
NIP. 199005062019031007
Anggota Penguji III : Yulianto, M. Pd.I
NIP. 198707122019031005

Tanda
Tangan



Mengetahui & Mengesahkan
Ketua Program Studi



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

PERNYATAAN KEASLIAN TULISAN

PERNYATAAN KEASLIAN TULISAN

Saya akan bertanda tanga dibawah ini:

Nama : Intan Pramuda Wardani

NIM : 19680055

Prodi : Perpustakaan Sains dan 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, 24 Juni 2025

Yang membuat pernyataan



Intan Pramuda Wardani

NIM. 19680055

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 serta keluarga besar yang selalu memberikan dukungan moral, materi maupun spiritual sehingga penulis dapat merasakan dan menyelesaikan pendidikan perkuliahan hingga akhir dengan baik.
2. Bapak Dr. Ir. Mokhamad 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 M. Pd.I selaku Dosen Pembimbing II yang telah meluangkan waktunya untuk membimbing pengerjaan skripsi ini.
5. Bapak Wahyu Hariyanto, M.M 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. Seluruh teman seperjuangan, khususnya para srikandi, yang telah menjadi pelipur lara dan sumber kekuatan melalui canda, semangat, serta kebersamaan yang tak ternilai harganya.
8. Teman-teman di sekitar saya yang selalu memberikan dukungan, semangat, dan kebersamaan selama proses ini berlangsung.
9. Seseorang yang senantiasa hadir dan mendampingi selama proses penyusunan karya ini hingga akhir. Terima kasih atas waktu, semangat, serta segala bantuan yang telah diberikan. Semoga senantiasa dilimpahi kebahagiaan dan kedamaian.
10. *Last but not least, to myself thank you for holding on, for fighting through the struggles, and for continuing to move forward even when giving up seemed like the easiest choice. Always remember that no matter how difficult the journey may feel, it will pass and better days will come.*

Dalam penyusunan skripsi ini masih terdapat kekurangan. Harapan semoga skripsi ini memberikan manfaat kepada orang lain. Amin Ya Rabbal Alamin.

Wassalamu'alaikum Wr. Wb.

Malang, 24 Juni 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN KEASLIAN TULISAN	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
ABSTRAK	xi
ABSTRACT.....	xii
مستخلص البحث.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	5
1.3 Tujuan Masalah.....	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah.....	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	7
2.1 Tinjauan Pustaka	7
2.2 Landasan Teori.....	9
2.2.1 Aplikasi Wattpad.....	9
2.2.2 X.....	10
2.2.3 Tweet Harvest	11
2.2.4 Python	11
2.2.5 Sastrawi.....	11
2.2.6 Analisis Sentimen.....	12
2.2.7 Term Frequency Invers Document Frequency (TF-IDF).....	15
2.2.8 Klasifikasi <i>Naïve Bayes</i>	15
2.2.9 <i>Peformance Evaluation Measure</i>	17
BAB III METODOLOGI PENELITIAN	20
3.1 Jenis Penelitian.....	20
3.2 Alur Penelitian	20
3.3 Desain Sistem.....	23
BAB IV HASIL DAN PEMBAHASAN.....	45
4.1 Hasil Penelitian	45
4.1.1 Hasil Pelabelan Data	45
4.1.2 Hasil Preprocessing.....	46
4.1.3 Hasil Pembagian <i>Dataset</i>	49
4.1.4 Hasil Pembobotan Data.....	50

4.1.5 Hasil Klasifikasi <i>Naïve Bayes</i>	50
4.1.6 Evaluasi Model.....	54
4.2 Pembahasan.....	61
BAB V PENUTUP	65
5.1 Kesimpulan	65
5.2 Saran.....	65
DAFTAR PUSTAKA	66
LAMPIRAN	70

DAFTAR TABEL

Tabel 2.1	Confusion Matrix	18
Tabel 3.1	Contoh Proses Cleansing	24
Tabel 3.2	Contoh Proses <i>Case folding</i>	25
Tabel 3.3	Contoh Proses <i>Tokenizing</i>	27
Tabel 3.4	Contoh Proses Normalization	28
Tabel 3.5	Contoh Proses Stopword	29
Tabel 3.6	Contoh Proses Stemming	30
Tabel 3.7	Contoh Data Terlabel	32
Tabel 3.8	Contoh Perhitungan TF	33
Tabel 3.9	Contoh Hasil Perhitungan TF-IDF	35
Tabel 3.10	Contoh Proses Data Training	36
Tabel 3.11	Contoh Proses Frekuensi Kata Positif	37
Tabel 3.12	Contoh Proses Frekuensi Kata Negatif	37
Tabel 3.13	Contoh Proses Perhitungan Probabilitas Kata pada D1	38
Tabel 3.14	Contoh Proses Perhitungan Probabilitas Kata pada D2	39
Tabel 3.15	Contoh Proses Perhitungan Probabilitas Kata pada D3	39
Tabel 3.16	Contoh Proses Perhitungan Probabilitas Kata pada D4	39
Tabel 3.17	Contoh Proses Perhitungan Probabilitas Kata pada D5	40
Tabel 3.18	Contoh Data Testing	40
Tabel 3.19	Contoh Nilai Probabilitas Data Testing	41
Tabel 3.20	Contoh Hasil Klasifikasi Sentimen	42
Tabel 3.21	Contoh Hasil Pengujian Confusion Matrix	43
Tabel 4.1	Hasil Pelabelan Data	46
Tabel 4.2	Tabel Pre-processing	47
Tabel 4.3	Tabel <i>Case folding</i>	47
Tabel 4.4	Tabel Filtering	48
Tabel 4.5	Tabel <i>Tokenizing</i>	48
Tabel 4.6	Tabel Stemming	49
Tabel 4.7	Tabel Rasio Pembagian Data	51
Tabel 4.8	Tabel Hasil Klasifikasi Rasio 90:10	51
Tabel 4.9	Tabel Hasil Klasifikasi Rasio 80:20	52
Tabel 4.10	Tabel Hasil Klasifikasi Rasio 70:30	52
Tabel 4.11	Hasil Perhitungan Confusion Matrix	60

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	20
Gambar 3. 2 Desain Sistem.....	23
Gambar 4. 1 Hasil Labeling Data.....	45
Gambar 4. 2 TampilanTF-IDF	50
Gambar 4. 3 Distrubusi Data Uji Rasio 90:10	53
Gambar 4. 4 Distrubusi Data Uji Rasio 80:20	53
Gambar 4. 5 Distrubusi Data Uji Rasio 70:30	54
Gambar 4. 6 Hasil Klasifikasi Rasio 90:10	55
Gambar 4. 7 Hasil Klasifikasi Rasio 80:20	57
Gambar 4. 8 Hasil Klasifikasi Rasio 70:30	58

ABSTRAK

Wardani, Intan Pramuda, 2025. **Analisis Sentimen Pengguna X Terhadap Aplikasi *Wattpad* Menggunakan Metode *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, M. pd.I

Kata Kunci: *analisis sentimen, naïve bayes, wattpad, tf-idf, media sosial x, text mining*

Aplikasi wattpad merupakan aplikasi yang populer dalam Google Play Store karena sudah ada sejak tahun 2006, akan tetapi masih terdapat beberapa permasalahan seperti sering error dan terlalu banyak iklan. Oleh karena itu perlu dilakukan penelitian lebih lanjut tentang analisis sentimen pengguna terhadap aplikasi Wattpad pada media sosial X. Penelitian ini bertujuan untuk mengetahui seberapa baik algoritma *Naïve Bayes Classifier* dalam pengklasifikasian opini menjadi sentimen positif dan negatif. Data yang digunakan berasal dari media sosial X melalui proses *crawling* menggunakan Tweet Harvest dengan kata kunci “Wattpad” pada tweet berbahasa Indonesia. Selanjutnya, dilakukan tahap pra-pemrosesan seperti *cleansing, case folding, tokenizing, stopword removal, dan stemming* menggunakan library Sastrawi. Pembobotan kata dilakukan menggunakan metode TF-IDF, selanjutnya dilakukan tahap klasifikasi menggunakan algoritma *Naïve Bayes*. Evaluasi dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-Score. Skenario penelitian dilakukan dengan rasio data latih dan data uji 70:30, 80:20 dan 90:10. Dari perbandingan ketiga rasio tersebut diperoleh bahwa model 90:10 menghasilkan akurasi, presisi, recall dan f-1 score yang tinggi dibandingkan yang lain. Penelitian ini menunjukkan bahwa model yang dihasilkan algoritma *Naïve Bayes* dengan rasio 90:10 efektif dalam mengklasifikasikan sentimen pengguna terhadap evaluasi aplikasi Wattpad.

ABSTRACT

Wardani, Intan Pramuda, 2025. **The Analysis of X Users Sentiment on Wattpad Application Using Naïve Bayes Method. Thesis. Library and Information Science Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang**
Advisor: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Yulianto, M.Pd.I

Keywords: sentiment analysis, naïve bayes, wattpad, tf-idf, x social media, text mining

The Wattpad application has been popular on the Google Play Store since its launch in 2006. However, it has several issues, such as frequent errors and excessive advertisements. Therefore, further research is needed to analyze user sentiment on the Wattpad application on the X social media platform. The research aims to examine the effectiveness of the Naïve Bayes Classifier algorithm in classifying user opinions into positive and negative sentiments. The research data were obtained from X social media platform through a crawling process using Tweet Harvest, with the keyword “Wattpad” applied to tweets written in Indonesian. The preprocessing stage included cleansing, case folding, tokenizing, stopword removal, and stemming, utilizing the Sastrawi library. Term weighting was performed using the TF-IDF method, and sentiment classification using the Naïve Bayes algorithm. Evaluation included accuracy, precision, recall, and F1-score metrics. The research used three train and test data ratios: 70:30, 80:20, and 90:10. Among these, the 90:10 ratio yielded the highest accuracy, precision, recall, and F1-score values. The research results demonstrate that the Naïve Bayes model with a 90:10 data split effectively classifies user sentiment in evaluating the Wattpad application.

Translator, Rizka Yanuarti NIPPPK 197801242023212005	Date 02-07-2025
----------------------------------------------------------------	------------------------

مستخلص البحث

ورداني، إنتان فرامودا، ٢٠٢٥. تحليل مشاعر مستخدمي تويتر تجاه تطبيق واتباد باستخدام طريقة بايز الساذج. البحث الجامعي. قسم المكتبات وعلوم المعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: فخرس حسن ريزا محفوظ، الماجستير؛ المشرف الثاني: يوليانتو، الماجستير.

الكلمات الرئيسية: تحليل مشاعر، بايز ساذج، واتباد، تي اف-أي دي دف، وسائل تواصل اجتماعي X، تنقيب بيانات.

يعتبر تطبيق واتباد من التطبيقات الشائعة في متجر جوجل بلاي لأنها موجودة منذ عام ٢٠٠٦، ومع ذلك لا تزال هناك بعض المشكلات مثل الأخطاء المتكررة ووجود إعلانات كثيرة. ولذلك، من الضروري إجراء مزيد من البحث حول تحليل مشاعر المستخدمين تجاه تطبيق واتباد على وسائل التواصل الاجتماعي تويتر X. هدف هذا البحث إلى معرفة مدى فعالية خوارزمية مصنف بايز الساذج في تصنيف الآراء إلى مشاعر إيجابية وسلبية. البيانات المستخدمة تأتي من وسائل التواصل الاجتماعي X من خلال عملية الزحف باستخدام Tweet Harvest بكلمات رئيسية "واتباد" على التغريدات باللغة الإندونيسية. بعد ذلك، تتم المرحلة السابقة للمعالجة مثل تنظيف البيانات، وتحويل حالة الكلمات، وتجزئة النص، وإزالة الكلمات غير المهمة، والتصريف باستخدام مكتبة ساستراوي. تم وزن الكلمات باستخدام طريقة تردد المصطلح - معكوس تردد الوثيقة TF-IDF، ومن ثم تم إجراء مرحلة التصنيف باستخدام خوارزمية بايز الساذج. تم التقييم باستخدام مقاييس الدقة والثبات، والاسترجاع، وقيمة F1. تم إجراء سيناريو البحث بنسبة بيانات التدريب إلى بيانات الاختبار ٣٠:٧٠:٢٠:١٠:٩٠. من مقارنة النسب الثلاثة، تبين أن نموذج ٩٠:١٠ حقق دقة، وثبات، واسترجاع وقيمة F1 مرتفعة مقارنة بالآخرين. بين هذا البحث أن النموذج الناتج عن خوارزمية بايز الساذج بنسبة ٩٠:١٠ فعال في تصنيف مشاعر المستخدمين تجاه تقييم تطبيق واتباد.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan di era digital saat ini semakin pesat dan tentunya sangat membantu aktivitas masyarakat menjadi lebih mudah dan semua kegiatan dapat dilakukan dimanapun dan kapanpun dengan lebih praktis. Perkembangan ini membantu manusia dalam melengkapi kebutuhan, salah satunya kebutuhan informasi dan komunikasi yang lebih efisien. Perkembangan teknologi telah merambah pada dunia perpustakaan. Mendorong terciptanya perpustakaan berbasis digital, yang kini telah merubah fasilitas dan layanan serta koleksi perpustakaan menjadi serba mudah diakses, Perpustakaan menganggap bahwa digitalisasi ini menjadi sebuah solusi untuk dapat mengatasi ketertinggalan, karena perpustakaan digital dapat memudahkan untuk melakukan sebuah penelusuran (Shafina & Andita, 2022).

Perkembangan perpustakaan pun tidak hanya merubah tata kelola saja tetapi perubahan terhadap koleksi yang dimiliki, koleksi koleksi fiksi maupun non fiksi menjadi media cetak digital. Sumber informasi dari koleksi buku tercetak menjadi buku elektronik atau biasanya disebut *E – book* yang mana sumber informasi ini tidak hanya berada didalam perpustakaan, melainkan tersebar di dunia maya dan dapat diakses dimanapun dan kapanpun (Khoirunnisa, 2022). Salah satunya yaitu Wattpad yang merupakan platform membaca berbasis online yang bisa diakses di smartphone dan dibawa ke mana saja serta bisa diakses di mana saja. Wattpad merupakan salah satu aplikasi yang menyediakan e-book dengan beragam jenis yang dapat mempengaruhi kemampuan literasi masyarakat. Tidak hanya digunakan untuk membaca e- book, Wattpad juga menyediakan tempat untuk kita menulis karya fiksi maupun non fiksi yang mengandung nilai dan pesan moral serta ilmu pengetahuan di dalamnya (Nurhafida & Sembiring, 2022).

Kemunculan Wattpad sebagai aplikasi novel digital mempermudah pembaca untuk mengakses novel melalui gadget dan memberikan platform *self- publishing* bagi

penulis lama atau penulis baru. Wattpad juga dijadikan sebagai distribusi informasi yang lebih baru dan modern, jika dibandingkan dengan distribusi informasi sebelumnya yang masih dilakukan secara tradisional, mengingat seiring berkembangnya teknologi, masyarakat membutuhkan akses dan membuat konten informasi tanpa terikat waktu dan tempat (Maidatul Lifiani, 2019).

Wattpad sebagai aplikasi novel digital yang populer di Google Play Store setiap pengunjung wattpad dapat mengunduh jutaan buku secara gratis, menawarkan berbagai fitur menarik yang bisa diunduh secara gratis. Meskipun telah hadir di Media sejak 2006, media sosial Wattpad tetap eksis. Menurut Wattpad, 15 miliar menit dihabiskan menggunakan aplikasi setiap bulan oleh 45 juta orang yang mengakses media ini (Nadya M, 2023). Saat ini aplikasi Wattpad merupakan salah satu topik yang menarik untuk dibicarakan. Ketika seseorang menulis opini tentang Wattpad, maka semua unsur yang ada didalam aplikasi wattpad akan dituliskan. Contohnya seperti aplikasi Wattpad mempunyai koleksi fiksi yang lengkap, bisa diakses secara gratis, atau sering terjadi error, serta banyaknya genre yang tersedia di aplikasi Wattpad.

Sebagian besar pengguna X memberikan tanggapan tentang Wattpad yang mereka gunakan dalam bentuk opini berupa pengalaman baik maupun buruk. Wattpad memiliki berbagai macam topik yang informasinya dapat digali kembali. Oleh karena itu, opini yang terdapat di X serta berkaitan dengan aplikasi Wattpad akan dijadikan data pada penelitian ini. Media sosial X atau yang biasa kita kenal sebagai Twitter adalah aplikasi yang sering dikunjungi sejak didirikannya pada tahun 2006, aplikasi ini memberikan akses untuk penggunanya berinteraksi secara global dan menyampaikan pendapat atau perasaan mereka tentang suatu hal yang berkaitan dengan bisnis, hiburan, ekonomi, politik, dan lain-lainnya. Oleh karena itu, X merupakan salah satu wadah yang dapat digunakan pada saat melakukan proses analisis sentimen, dimana terdapat banyak sekali data atau dokumen teks yang dapat dijadikan sebagai bahan dalam melakukan analisis sentimen (Nurkalyisah dkk., 2022).

Analisis sentimen adalah studi analisa opini, sentimen, penilaian, sikap dan emosi seseorang dalam suatu produk, layanan, organisasi, peristiwa, isu atau topik yang

diungkap dalam teks tertulis yang dapat menjadi potensi besar bagi suatu objek yang bersangkutan (Nur Adhan dkk., 2024). Analisis sentimen merupakan proses yang digunakan oleh peneliti dengan tujuan memberikan suatu informasi dari sebuah dataset yang tidak terstruktur dengan menggunakan berbagai macam metode penelitian (Nurkalyisah dkk., 2022). Melalui analisis sentimen, Wattpad dapat memperoleh informasi tentang opini pengguna terhadap layanan pada aplikasi Wattpad, sehingga bisa menjadi bahan evaluasi untuk Wattpad kedepannya dalam perbaikan kualitas dan layanan.

Analisis sentimen sangat penting untuk diperhatikan Dalam Al-Qur'an, konsep sentimen atau perasaan manusia, termasuk dalam memberi atau menerima opini baik positif maupun negatif, yang tercermin dalam sikap dan kata-kata. Dalam konteks komunikasi, Al-Qur'an memberikan pedoman untuk berbicara dengan cara yang baik dan positif, sebagaimana yang dinyatakan dalam Surah Al-Hujurat ayat 6 :

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصِحُّوا عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ

Artinya :

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

Ayat ini memberikan peringatan agar setiap informasi atau opini yang diterima dari masyarakat, khususnya yang belum dapat dipastikan kebenarannya, hendaknya diteliti terlebih dahulu sebelum diyakini kebenarannya. Pesan ini menjadi pedoman penting bagi kaum mukmin untuk senantiasa bersikap waspada dalam menerima dan menyikapi berita, terutama apabila sumbernya berasal dari individu yang dikenal memiliki perilaku fasik. Substansi dari ayat ini menekankan pentingnya proses verifikasi informasi guna menghindari potensi kerugian yang dapat berujung pada penyesalan (Kemenag, 2025). Pemahaman terhadap ayat ini menunjukkan pentingnya menelaah opini yang beredar secara lebih mendalam. Salah satu cara yang dapat dilakukan adalah dengan melakukan analisis sentimen terhadap opini pengguna

aplikasi Wattpad di platform X. Analisis ini bertujuan untuk melihat kecenderungan opini tersebut, apakah bersifat positif atau negatif. Hasil dari analisis ini dapat memberikan informasi yang berguna bagi pihak terkait dalam mendukung pengambilan keputusan yang tepat. Dalam penelitian ini, untuk melakukan analisis sentimen digunakan algoritma *Naive Bayes Classifier*. Algoritma *Naive Bayes* dipilih karena metode ini sangat sederhana berdasarkan aturan bayes, Algoritma *Naive Bayes* dapat digunakan untuk seluruh jenis data dan proses perhitungannya cepat, efisien, serta mudah dipahami agar penyampaianya jelas (Rahayu & Farlina, 2021). Salah satu penelitian analisis sentimen X yang pernah dilakukan yaitu menganalisis sentimen terhadap jasa transportasi Online menggunakan metode klasifikasi *Naive Bayes Classifier* melalui tahapan penelitian proses crawling data, pre-processing, dan klasifikasi. Hasil penelitian ini menunjukkan bahwa berdasarkan hasil analisis yang dilakukan terkait sentimen terhadap jasa transportasi online dengan menggunakan metode *Naive Bayes* untuk menentukan kelas prediksi, didapatkan hasil sentimen positif sebesar 88.60% dan sentimen negatif sebesar 11.40% dengan akurasi sebesar 86.80% (Pintoko & L., 2018).

Penelitian analisis sentimen ini menggunakan data yang berasal dari media sosial X yang merupakan data tweet berbahasa Indonesia yang diambil menggunakan X Harvest dalam pengumpulan data di X data yang diperoleh masih berupa teks utuh dari tweet pengguna X. Selanjutnya akan disempurnakan melalui tahap preprocessing data yang digunakan untuk melakukan analisis sentimen lalu mengekstraksi fitur menggunakan TF IDF untuk melakukan perhitungan nilai pada setiap fitur kata di dokumen agar mendapatkan nilai akurasi yang sesuai dan lebih baik dengan pengklasifikasian pada algoritma *Naive Bayes Classifier*. Setelah semua proses dilakukan maka tahapan terakhir yaitu dilakukannya evaluasi pada hasil diantaranya accuracy, precision, dan recall. Menggunakan kata kunci Wattpad pada tweet berbahasa indonesia yang diambil di X. Berdasarkan latar belakang yang telah dijelaskan maka peneliti mengambil judul “Analisis Sentimen Pada X terhadap Aplikasi Wattpad dengan Menggunakan Metode *Naive Bayes*”.

1.2 Identifikasi Masalah

Adapun identifikasi masalah dari penelitian adalah berapa hasil kinerja *Naïve Bayes Classifier* dalam menganalisis sentimen X terhadap Wattpad?

1.3 Tujuan Masalah

Adapun tujuan dari penelitian adalah untuk mengukur hasil kinerja dari penerapan metode *Naïve Bayes Classifier* dalam menganalisis sentimen terhadap aplikasi Wattpad pada X.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian adalah sebagai acuan atau referensi bagi penelitian dalam bidang analisis sentimen, serta menjadi bahan evaluasi terhadap aplikasi Wattpad.

1.5 Batasan Masalah

Agar penelitian tidak menyimpang dan lebih terarah, maka penelitian ini mempunyai batasan masalah berupa tweets yang digunakan berbahasa Indonesia dengan kata kunci “wattpad” dan di posting mulai dari tahun 2022 sampai 2025.

1.6 Sistematika Penulisan

Sistematika penulisan dibuat agar penulisan laporan penelitian ini dapat lebih tersusun rapi dan terstruktur. Berikut merupakan sistematika penulisan pada penelitian ini:

BAB I Pendahuluan

Bab I terdiri dari latar belakang masalah yang merupakan permasalahan yang akan diteliti tentang analisis sentimen pada aplikasi X, dan terdapat hadist sesuai dengan topik pembahasan untuk memperkuat gagasan. Untuk bagian identifikasi masalah menjelaskan rumusan permasalahan yang akan dilakukan pada penelitian. Tujuan yang akan menjadi patokan pencapaian pada penelitian. Manfaat penelitian

menjelaskan manfaat yang diperoleh jika penelitian ini dilakukan, dan untuk batasan masalah digunakan agar penelitian yang dilakukan bisa terfokus.

BAB II Kajian Pustaka

Bab II berisi tentang penelitian terdahulu yang terkait tentang analisis sentimen, Naive Bayes, landasan teori dan tinjauan pustaka yang berupa *X*, *wattpad*, *tweet harvest*, *python*, *sastrawi*, *Naive bayes*, analisis sentimen, *Tf-Idf*, *evaluasi machine learning*.

BAB III Metode Penelitian

Bab III menjelaskan tentang sub bab yang mengandung beberapa uraian dan jenis penelitian yaitu tentang text mining, alur dan sumber data yang diambil dari salah satu media social yaitu *X*, analisis data, hasil penelitian serta subjek dan objek penelitian, serta instrument penelitian yang digunakan.

BAB IV Hasil dan Pembahasan

Bab IV merupakan hasil dan pembahasan penelitian tentang Analisis Sentimen aplikasi *Wattpad*.

BAB V Kesimpulan dan Saran

Bab V berisi tentang kesimpulan penelitian tentang analisis sentimen aplikasi *wattpad*

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian yang membahas analisis sentimen sudah banyak dilakukan dengan berbagai macam metode, Penelitian pertama yang telah dilakukan oleh Yuni Nurtikasari dan teman-temannya pada tahun 2022 berjudul “Analisis Sentimen Opini Masyarakat Terhadap Film Pada *Platform X* Menggunakan Algoritma *Naïve Bayes*”. Penelitian berikut bertujuan mengimplementasikan algoritma *Naïve Bayes Classifier* terhadap analisis sentimen data opini film pada X. Untuk hasil pengujian dengan menggunakan metode algoritma *Naïve Bayes Classifier* didapatkan hasil rata-rata nilai akurasi 0,65% dan nilai presisi sebesar 0,67%, dan recall sebesar 0,65%, dan persentase netral 0,83% pada klasifikasi nya. Hal ini membuktikan bahwa sentimen masyarakat pada platform X terhadap kasus film Ngeri-Ngeri Sedap tergolong netral (Nurtikasari dkk., 2022).

Selanjutnya untuk penelitian kedua yang telah dilakukan oleh Ni Made Ayu Juli Astari dan teman-temannya pada tahun 2020 yang berjudul “Analisis Sentimen Dokumen X Mengenai Dampak Virus Corona Menggunakan *Metode Naïve Bayes Classifier*”. Penelitian berikut bertujuan untuk mengimplementasikan algoritma *Naïve Bayes Classifier* terhadap analisis sentimen dengan pertimbangan tingginya akurasi hasil klasifikasi pada penelitian sebelumnya. Metode ini diharapkan mampu menghasilkan nilai akurasi yang tinggi untuk klasifikasi dokumen sosial media pada kasus dampak virus Corona. Untuk hasil pengujian yang tertinggi diperoleh dengan menggunakan 100 data dengan hasil accuracy sebesar 67% dan error sebesar 33%. Menunjukkan metode *Naïve Bayes* untuk klasifikasi data tweet terkait dampak virus Corona menghasilkan performa yang stabil (Astari dkk., 2020).

Penelitian ketiga berjudul “Analisis Sentimen Perpustakaan Nasional Republik Indonesia Melalui Media Sosial X” penelitian ini bertujuan untuk menunjukkan tingkat akurasi dari algoritma *Naïve Bayes* dan KKN dalam hal sentimen masyarakat terhadap

kepuasan pengguna. Menggunakan metode yang efektif dan akurat untuk pengambilan keputusan dalam penelitian ini adalah TF-IDF dan TF biner untuk pembobotan. Pada tahap klasifikasi, metode yang digunakan adalah *Naïve Bayes* dan KNN. Penelitian ini dikembangkan oleh Fakhris Khusnu Reza dan teman-temannya pada tahun 2020, yang mana Klasifikasi KNN menghasilkan akurasi, presisi, dan ingatan masing masing sebesar 83,33%, 79,2%, dan 83,3%. KNN juga mendukung pembobotan TF-IDF *Naïve Bayes* jika dilihat dari hasil kinerja yang dihasilkan (Mahfud, 2020).

Penelitian keempat yaitu “Analisis Sentimen Vaksin Sinovac Pada X Menggunakan Algoritma *Naïve Bayes*” penelitian ini dilakukan oleh Sri Lestari dan teman-temannya pada tahun 2021. Penelitian ini dilakukan bertujuan memberikan solusi untuk mengetahui tingkat persentase komentar dan opini positif atau negatif dari pengguna akun sosial media X terhadap vaksin sinovac dengan pengimlentasian algoritma *Naïve Bayes*. Sehingga berdasarkan penelitian ini kita dapat mengetahui bagaimana tanggapan pengguna X terhadap pengadaan vaksin Sinovac, dimana dengan persentase dan respon tersebut akan menjadi masukan dan evaluasi bagi perusahaan dan pemerintah tentang adanya pengadaan vaksin Sinovac. Hasil pengujian kinerja model menggunakan algoritma *Naïve Bayes* Nilai akurasi yang didapatkan dengan algoritma *Naïve bayes* ini adalah 92,96%. Dapat disimpulkan bahwa algoritma *Naïve bayes* ini memiliki nilai akurasi yang tinggi sehingga hasil dari prediksinya akurat untuk dijadikan metode dalam penelitian Data Mining (Lestari & Saepudin, 2021).

Penelitian kelima yang berjudul “Analisis Sentimen X Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan *Naïve Bayes Classifier*” yang bertujuan untuk mengukur analisis sentimen pada sosial media X, opini dan pendapat masyarakat mengenai kuliah online pasca covid-19 dengan menggunakan algoritma *Naïve Bayes* dan Support Vector Machine yang dilakukan oleh Hendrik Setiawan dan teman temannya pada tahun 2021. Hasil performa yang didapat pada algoritma bayes akurasi 81,20%, waktu 9,00 detik, recall 79,60% dan presisi 79,40%. Untuk algoritma SVM mendapatkan nilai akurasi 85%, waktu 31,60 detik, recall 84% dan presisi 83,60 (Setiawan dkk., 2021)

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

2.2 Landasan Teori

Dalam proses penelitian adapun beberapa teori yang akan dibahas untuk menjelaskan setiap variabel yang diteliti. Berikut landasan teori yang digunakan.

2.2.1 Aplikasi Wattpad

Wattpad adalah salah satu aplikasi yang memiliki nilai literasi terbesar di dunia serta menyediakan layanan bagi para penulis untuk membagikan karyanya. Wattpad merupakan aplikasi yang didirikan di Kanada pada tahun 2006 oleh Allen Lau dan Ivan Yuen yang telah berhasil menarik perhatian masyarakat dikarenakan wattpad salah satu platform yang digemari karena aksesibilitas dan fiturnya hingga diakses oleh 90 juta pengguna di seluruh dunia (Shoji Al Falaq dkk., 2021). Wattpad adalah sebuah aplikasi yang berbentuk seperti novel tetapi dalam bentuk digital, Wattpad berbentuk digital baru akan dicetak jika pengarang menginginkan dicetak dalam bentuk fisik atau pembacanya sudah dianggap banyak (Fathiyatul dkk., 2020). Wattpad adalah aplikasi dengan fitur gratis yang mudah digunakan bagi orang awam juga memiliki kekurangan yang mana tidak terdapat penyaringan pada konten yang ditampilkan, sehingga jika kita melakukan pencarian kata kunci maka akan otomatis keluar tanpa harus memberikan batasan umur, namun tetap ada peringatan umur yang diberitahu oleh pengarang bahwa konten mengandung unsur negatif (Fathiyatul dkk., 2020).

Selain itu Wattpad merupakan aplikasi yang menyediakan fitur khusus bagi pemilik akun untuk mengunggah karyanya secara berkala, yang biasanya disebut karya berseri (Mawardi, 2018). Wattpad adalah platform menulis dan membaca cerita yang banyak digunakan. Saat menulis, penulis (author) bisa langsung berinteraksi

dengan pembaca (reader) melalui kolom komentar. Di sana, pembaca bisa memberikan komentar, saran, bahkan kritik. Wattpad juga punya fitur untuk menghitung jumlah komentar, pengikut (follower), dan bintang yang diterima setiap cerita. Cerita yang paling banyak mendapatkan itu semua bisa masuk ke peringkat atas (Mawardi, 2018).

2.2.2 X

X adalah salah satu aplikasi yang menarik perhatian pengguna internet, aplikasi media sosial yang menawarkan jaringan sosial berupa microblog yang memberi akses mudah untuk penggunanya saling bertukar informasi dan dapat saling terhubung dengan pengguna lain, Microblog merupakan jenis alat komunikasi antara blog dan pesan instan dimana pengguna bisa berkomunikasi online dan melakukan sesuatu dengan pendapat mereka tentang suatu objek atau fenomena tertentu. Pengguna aplikasi X juga dapat menyusun pesan kurang dari 280 karakter sehingga di X tidak tersedia fitur “read more” karena karakter sudah ditampilkan semuanya dilayar (Alhaq dkk., 2021)

X juga digunakan untuk berinteraksi secara global dan menyampaikan pendapat atau perasaan mereka tentang suatu hal seperti bisnis, hiburan, ekonomi, politik dan lain lainnya (Nurkalyisah dkk., 2022). X adalah aplikasi yang didirikan di san francisco pada maret 2006 oleh Jack Dorsey, sejak diluncurkan X telah menjadi salah satu dari sepuluh situs yang paling sering diakses di internet (Malik Zuhdi dkk., 2019) X memiliki cara unik untuk mengapresiasi suatu objek khusus yang menonjol bisa menggunakan # (hashtag) sehingga orang lain dapat mengikuti informasi dan saling diskusi seputar topik tersebut. Didalam aplikasi X juga ada yang disebut trending topics yang menjadi parameter dalam penyebaran informasi di dalam aplikasi tersebut. Para pengguna lainnya juga bisa memberi komentar untuk objek yang sedang trending (Mutiarra & Vience, 2017).

2.2.3 Tweet Harvest

Tweet Harvest merupakan sebuah command-line tool yang menggunakan Playwright untuk mendapatkan tweets dari hasil pencarian X berdasarkan kata kunci dan rentang tanggal yang ditentukan. Tweets yang berhasil didapatkan kemudian akan disimpan dalam file berbentuk CSV. Untuk menggunakan Tweet Harvest dibutuhkan authorization token yang bisa didapatkan dengan login ke akun X di browser dan kemudian mengekstrak authorization token (Putra Prasetyo dkk., 2024).

2.2.4 Python

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Python bisa dibilang bahasa pemrograman dengan tujuan umum yang dikembangkan secara khusus untuk membuat source code mudah dibaca. Python juga memiliki library yang lengkap sehingga memungkinkan programmer untuk membuat aplikasi yang mutakhir dengan menggunakan source code yang tampak sederhana (harahap, 2024).

2.2.5 Sastrawi

Sastrawi merupakan sebuah library stemming sederhana yang dirancang untuk dapat digunakan secara mudah yang menerapkan algoritma yang berbasis Nazief dan Adriani yang dikembangkan berdasarkan aturan morfologi bahasa indonesia yang mengelompokkan imbuhan menjadi awalan (prefix), sisipan (infix), akhiran (suffix) dan gabungan awalan dan akhiran (confixes). Algoritma ini menggunakan kamus dasar dan mendukung recording yaitu penyusunan kembali kata kata yang mengalami proses stemming berlebihan (Pahlevi & Ziveria, 2022).

2.2.6 Analisis Sentimen

Analisis sentimen merupakan salah satu metode dalam penelitian yang termasuk dalam bidang text mining. Metode ini digunakan untuk mengolah data teks agar bisa mengetahui perasaan atau emosi seseorang, seperti senang, sedih, atau marah, terhadap suatu hal. Biasanya digunakan untuk menganalisis opini atau penilaian terhadap topik, produk, layanan, organisasi, atau individu tertentu (Fridom Mailo dkk., 2021). Dalam Islam sendiri, nilai-nilai sentimen juga bisa ditemukan dalam Al-Qur'an. Beberapa kata yang mengandung makna positif seperti الرحمة (rahmah/rahmat), الصبر (sabar), dan العدل (keadilan) menunjukkan bahwa Al-Qur'an mendorong sikap dan emosi yang baik. Contohnya dalam QS. Al-A'raf ayat 56, Allah SWT berfirman:

إِنَّ رَحْمَتَ اللَّهِ قَرِيبٌ مِّنَ الْمُحْسِنِينَ

"*Sesungguhnya rahmat Allah itu dekat kepada orang-orang yang berbuat baik.*" (Kemenag RI, 2019)

Sementara itu, sentimen negatif seperti الغضب (kemurkaan), الظلم (kezaliman), dan الكفر (kekafiran) juga banyak disebutkan sebagai sikap yang dilarang. Misalnya pada QS. Al-Baqarah ayat 254, Allah menyebutkan:

وَالْكَافِرُونَ هُمُ الظَّالِمُونَ

"*Orang-orang kafir itu adalah orang-orang yang zalim.*" (Kemenag RI, 2019)

Melalui ayat-ayat ini, kita bisa melihat bahwa Al-Qur'an juga memberikan gambaran tentang emosi atau perasaan, baik yang positif maupun negatif. Oleh karena itu, nilai-nilai ini bisa dijadikan dasar tambahan dalam menganalisis sentimen dari sebuah teks, terutama yang berkaitan dengan tema sosial atau keagamaan. Adapun tahapan dalam analisis sentimen secara umum akan dijelaskan pada bagian berikutnya:

1. Pengumpulan Data

Pengumpulan Data yang dimaksud adalah pengambilan data berdasarkan objek yang telah disiapkan pada suatu aplikasi X yang mana dalam pencarian data

tersebut menggunakan keyword watsapp, data yang dikumpulkan merupakan data yang telah siap diolah dan sudah diberi label, hal ini adalah langkah awal untuk melakukan analisis sentimen.

2. Preprocessing Data

Dalam proses ini atau yang bisa disebut sebagai *text preprocessing*, maka terlebih dahulu disiapkan teks dokumen atau dataset yang tidak terstruktur yang akan diubah menjadi data yang lebih terstruktur. Adapun proses yang dilakukan dalam tahapan preprocessing antara lain yaitu yang pertama dilakukannya data *Cleaning* dan dilanjut *case folding*, *tokenizing*, stop removal dan yang terakhir slang word (Alita & Isnain, 2020). Proses Preprocessing ini meliputi:

1. Data *Cleaning*

Data *Cleaning* merupakan sebuah proses pembersihan data yang mengandung unsur seperti link, gambar, tanda baca, retweet, symbol dan kata kata yang tidak diperlukan yang terdapat pada dataset, agar tersisa beberapa kalimat dari data itu sendiri (Chuluq & Nudin, 2024).

2. *Case folding*

Case folding adalah sebuah proses yang merupakan masuk dalam tahapan *Cleaning*, yang mana mengkonversi keseluruhan teks dalam dokumen yang sebelumnya terdapat penggunaan huruf kapital menjadi sebuah bentuk standar font kecil atau sebuah proses mengubah semua huruf besar atau kapital dalam data menjadi huruf kecil, sehingga tidak ada pemakaian huruf kapital pada data tersebut seperti contoh di tabel bawah ini (Munthe dkk., 2022).

3. *Tokenizing*

Tokenizing adalah sebuah cara dalam praprosesing yakni proses pemotongan kata dalam sebuah dokumen menjadi beberapa bagian atau digunakan untuk memecahkan dan memisah sekumpulan karakter dalam sebuah teks data ke dalam satuan kata yang bertujuan untuk memperoleh

nama token atau potongan kata-kata yang akan mempermudah dalam proses stopword dan stemming nantinya (Munthe dkk., 2022).

4. *Normalization*

Word normalization merupakan proses mengubah kata gaul menjadi kata normal, Pada tahap *normalization*, seluruh kata yang memiliki kesalahan dalam ejaan, noise, ataupun bahasa yang tidak baku diganti menjadi bahasa baku. Kata yang mengandung noise sendiri merupakan kata yang mengandung unsur-unsur bahasa daerah dan tidak sesuai dengan Kamus Besar Bahasa Indonesia (Septian dkk., 2019).

5. *Stopword Removal*

Stopword Removal adalah proses menghapus kata-kata yang tidak memiliki kontribusi penting dalam kalimat, menggunakan daftar khusus (stoplist) untuk menyaring kata yang perlu dihapus dan menyisakan kata-kata penting untuk dianalisis lebih lanjut (Fachri & Ramdan, 2022).

6. *Stemming*

Stemming merupakan tahap akhir dalam proses *text preprocessing* dimana seluruh kata dalam dokumen akan diubah menjadi kata dasar. Sehingga, semua kata imbuhan mulai dari awalan (*prefix*), akhiran (*suffix*), sisipan (*infix*), serta gabungan antara awalan dan akhiran (*confix*) akan dihilangkan. Proses stemming dilakukan menggunakan library Sastrawi dalam penyesuaian variasi kata supaya tidak banyak variasi kata. Library Sastrawi mampu mengurangi over stemming dan under stemming serta waktu pengerjaan yang lebih cepat dibandingkan dengan menggunakan stemmer lain karena memang dirancang untuk dokumen yang berbahasa Indonesia (Alfan Rosid dkk., 2020).

2.2.7 Term Frequency Invers Document Frequency (TF-IDF)

Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (Term) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata didalam sebuah dokumen tertentu dan Inverse frekuensi dokumen yang mengandung kata tersebut, dalam perhitungan bobot menggunakan TF – IDF yang dihitung lebih dulu yaitu nilai TF perkata dengan bobot masing- masing kata adalah satu. Sedangkan nilai IDF diformulasikan pada persamaan (Herwijayanti dkk., 2018).

2.2.8 Klasifikasi *Naïve Bayes*

Metode *Naïve Bayes* merupakan salah satu metode klasifikasi berbasis probabilistik yang sederhana dan banyak digunakan dalam pengolahan data teks. Cara kerja metode ini didasarkan pada perhitungan probabilitas dari kombinasi nilai dalam dataset yang tersedia (Rezki dkk, 2020). Proses klasifikasi menggunakan metode *Naïve Bayes* dibagi menjadi dua tahapan, yaitu tahap pelatihan (training) dan tahap pengujian (testing). Pada tahap pelatihan, dilakukan pengumpulan data latih yang telah diberi label, kemudian dihitung nilai prior $P(V_j)$ untuk setiap kelas dan nilai likelihood $P(a_i|V_j)$ untuk setiap atribut. Selanjutnya, pada tahap pengujian, dilakukan pengumpulan data uji dan perhitungan nilai probabilitas posterior untuk masing-masing kelas, yang disebut sebagai nilai V_{map} . Jika nilai V_{map} untuk kelas positif lebih besar dari nilai V_{map} untuk kelas negatif, maka data diklasifikasikan sebagai sentimen positif. Sebaliknya, jika nilai V_{map} untuk kelas negatif lebih besar, maka data diklasifikasikan sebagai sentimen negatif.

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

$$P(B) = \frac{P(A) \cdot P(A)}{P(B)} \quad (2.1)$$

Keterangan:

A : sampel data yang label kelasnya tidak diketahui.

B : kelas-kelas hasil klasifikasi.

$P(A|B)$: probabilitas terjadinya A jika B diketahui.

$P(B|A)$: probabilitas terjadinya B jika A diketahui

$P(A)$: probabilitas prior A yang mendahului terjadinya B. disebut “prior”, karena nilainya bisa diperoleh tanpa perlu mempertimbangkan informasi apapun mengenai B terlebih dahulu, $P(A)$ juga berarti probabilitas ini diperoleh dari data sampel yang telah diketahui berkelas A.

$P(B)$: probabilitas prior B, dan bertindak sebagai normalizing constant. Dalam algoritma *Naïve bayes* ada tahap pelatihan dan pengujian membutuhkan dua tahapan dalam prosesnya yaitu tahap pelatihan dan tahap klasifikasi. Pada tahapan pertama adalah pelatihan terhadap suatu tweet yang telah diketahui kategorinya. Dalam metode ini setiap dokumen di dalam algoritma NBC setiap dokumen direpresentasikan dengan pasangan atribut “ a_1, a_2, \dots, a_n ” dimana a_1 adalah kata pertama, a_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori tweet.

Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{MAP}). Adapun persamaan V_{MAP} adalah sebagai berikut:

$$V_{Map} = \operatorname{argmax} P(a_1, a_2, \dots, a_n) \quad (2. 2)$$

Nilai $P(v_j)$ dihitung pada saat training, didapat dengan rumus sebagai berikut:

$$P(V_j) = \frac{|doc_j|}{|training|} \quad (2. 3)$$

Keterangan :

$P(V_j)$: probabilitas prior

$|doc\ j|$: jumlah dokumen

$|training|$: jumlah tweet untuk training

Dimana $|doc\ j|$ merupakan jumlah tweet yang memiliki kategori j dalam training. Sedangkan $|training|$ merupakan jumlah tweet dalam yang digunakan untuk training. Untuk setiap probabilitas kata a_i untuk setiap kategori $P(a_i|v_j)$, dihitung pada saat training. Probabilitas $P(V_j)P(V_{-j})P(V_j)$ ini disebut sebagai probabilitas prior, yang menggambarkan seberapa besar kemungkinan sebuah dokumen termasuk ke dalam kategori $V_jV_{-j}V_j$ sebelum mempertimbangkan kata-kata di dalam dokumen tersebut.

$$P(V_j) = \frac{nk + 1}{n + |kosakata|} \quad (2.4)$$

Keterangan :

$P(V_j)$: Probabilitas prior

nk : Jumlah kemunculan kata pada kelas tertentu

n : Total kata dalam kelas

$|kosakata|$: Jumlah kata unik keseluruhan

Dimana n_i adalah jumlah kemunculan kata a_i dalam tweet yang berkategori v_j dan $|kosakata|$ adalah banyaknya kata dalam training. Melalui tahapan ini, metode *Naïve Bayes* dapat digunakan untuk mengklasifikasikan data teks, seperti opini atau sentimen, secara efektif dan efisien.

2.2.9 *Performance Evaluation Measure*

Performance Evaluation Measure atau disebut pengukuran evaluasi performa adalah tahapan yang digunakan untuk mengukur performa suatu sistem. Metode ini banyak digunakan dalam training data untuk mengevaluasi model yang sudah dibuat (Matthew E, 2023). *Performance Evaluation Measure* biasanya digambarkan dalam

confusion matrix. *Confusion matrix* adalah sebuah metode dalam data mining yang biasanya digunakan untuk mengevaluasi performa suatu algoritma (Arminda, 2023). *Confusion Matrix* biasanya direpresentasikan dalam bentuk tabel yang berisi empat kemungkinan hasil klasifikasinya yang berisi data target prediksi yang akan dibandingkan dengan data target aktual untuk melakukan perhitungan akurasi pada konsep data mining dalam bentuk tabel seperti dibawah ini:

Tabel 2. 1 *Confusion Matrix*

Nilai Prediksi	Nilai Actual	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Pada tabel 2.1 diatas merupakan tabel confusion matrix dengan keterangan sebagai berikut :

1. TP (True Positive) = data nilai aktual kelas positif dan nilai prediksi kelas positif
2. TN (True Negative) = data nilai aktual negatif dan nilai prediksi negatif
3. FP (False Positive) = data nilai aktual positif dan nilai prediksi negatif
4. FN (False Negative) = data nilai aktual negatif dan nilai prediksi Positif

Pada umumnya semakin tinggi nilai TP dan TN maka semakin baik pula tingkat klasifikasi dari akurasi, presisi dan recall. Untuk menghitung akurasi digunakan persamaan dibawah ini:

- A. *Akurasi* adalah nilai rasio prediksi yang benar dari total data tweet, Nilai akurasi dapat menunjukkan kedekatan antara nilai prediksi dengan nilai actual, berikut rumusnya: Total nilai True Positif dan True Negatif dibagi dengan jumlah keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2. 5)$$

- B. *Presisi* adalah menggambarkan ketepatan hasil prediksi mengenai total dari jumlah dokumen yang dipilih oleh sistem, nilai presisi didapatkan dari perbandingan data yang diperoleh dari nilai prediksi positif dengan jumlah nilai yang positif, dengan menggunakan rumus sebagai berikut: Precision (P) adalah persentase nilai True Positif dari seluruh nilai Positif yang diprediksi.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2.6)$$

- C. *Recall* adalah rasio yang menunjukkan keberhasilan dalam menemukan kembali sebuah informasi, recall juga sebuah perbandingan antara prediksi benar dengan total data benar yang dapat dihitung dengan rumus sebagai berikut

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.7)$$

- D. *F1 Score* adalah salah satu perhitungan evaluasi atau perbandingan nilai rata-rata antara presisi dan recall dari keseluruhan dengan rumus sebagai berikut:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (2.8)$$

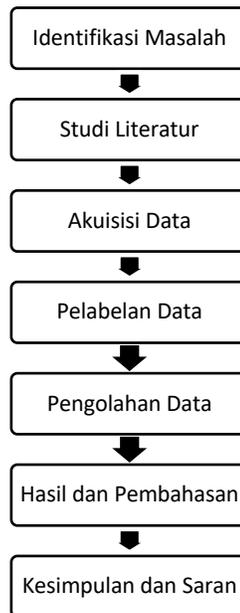
BAB III METODOLOGI PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan text mining. Pendekatan ini sering digunakan dalam berbagai penelitian untuk menilai perilaku, pengetahuan, opini, dan sikap dengan benar dan menemukan informasi berharga yang tersembunyi baik dari sumber informasi terstruktur dan tidak terstruktur (Firdaus & dkk, 2021). Penelitian ini termasuk metode kuantitatif karena bertujuan untuk mengidentifikasi opini pengguna X terhadap aplikasi Wattpad. Penelitian ini termasuk jenis penelitian deskriptif berdasarkan tujuannya dimana data text dari tweets di X akan diolah menggunakan algoritma *Naïve Bayes* untuk analisis sentimen terkait aplikasi Wattpad.

3.2 Alur Penelitian

Berikut merupakan tahapan penelitian yang digambarkan dalam bentuk diagram alur:



Gambar 3. 1 Alur Penelitian

Pada gambar 3.1 merupakan langkah- langkah alur penelitian. Berupa identifikasi masalah, studi literatur, akuisisi data, pelabelan data, hasil dan pembahasan, kesimpulan dan saran.

1. Identifikasi Masalah

Tahap identifikasi masalah merupakan langkah awal dalam proses penelitian yang bertujuan untuk menemukan permasalahan ilmiah yang relevan. Pada tahap ini, dilakukan penelaahan terhadap berbagai permasalahan yang muncul dalam layanan dan sistem pada aplikasi Wattpad, yang kemudian diuraikan lebih lanjut dalam bagian latar belakang.

2. Studi Literatur

Tahap studi literatur dilakukan untuk memperoleh informasi dan data pendukung yang bersumber dari berbagai artikel ilmiah, jurnal, dan referensi lainnya. Informasi yang dikumpulkan pada tahap ini digunakan sebagai dasar teoritis dan landasan dalam menyusun penelitian.

3. Akuisisi Data

Pada proses ini, data dikumpulkan dari X menggunakan alat bernama Tweet Harvest. Pengumpulan data dilakukan dengan memasukkan kata kunci yang berhubungan dengan Wattpad, seperti "wattpad" atau hashtag #wattpad. Hasil pencarian kemudian disimpan ke dalam database. Selama pengambilan data, digunakan filter agar hanya tweet berbahasa Indonesia yang diambil dan sesuai dengan topik penelitian. Teknik crawling digunakan agar data bisa dikumpulkan secara otomatis dan efisien dari X. Semua tweet yang berhasil dikumpulkan disimpan dalam format CSV (Comma-Separated Values), lalu diubah menjadi file Excel (.xlsx) agar lebih mudah diolah dan dianalisis.

4. Pelabelan Data

Setelah data berhasil dikumpulkan dan dibersihkan, tahap selanjutnya adalah pelabelan data. Proses ini dilakukan oleh ahli bahasa Indonesia yang memberikan label sentimen positif atau negatif pada setiap tweet berdasarkan isi

dan konteks kalimat. Pelabelan ini penting untuk menyediakan data berlabel yang digunakan sebagai acuan dalam proses pelatihan dan pengujian model klasifikasi.

5. Pengolahan Data

Dalam proses pengolahan data dimulai dengan pengumpulan data yang telah dilabeli oleh ahli bahasa sesuai dengan kategorinya, selanjutnya menghapus beberapa data yang tidak lengkap, duplikasi, salah dan tidak sesuai. Pada tahap ini dihasilkan 2.288 data yang siap digunakan untuk tahap klasifikasi.

6. Hasil dan Pembahasan

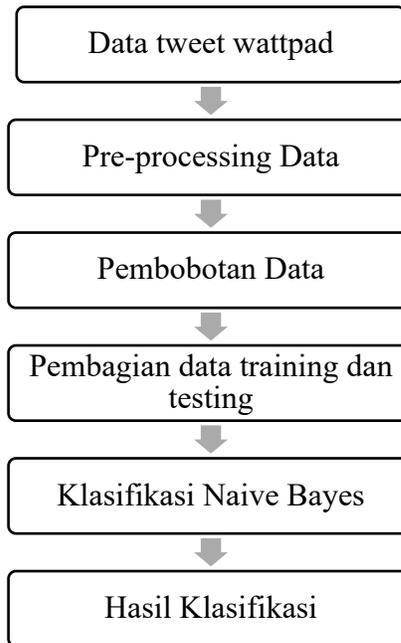
Pada tahap ini, dilakukan proses klasifikasi terhadap data yang telah dibersihkan menggunakan metode yang telah ditentukan. Data dibagi ke dalam data latih dan data uji untuk keperluan analisis. Hasil klasifikasi kemudian disajikan dalam bentuk uraian, tabel, maupun diagram, disertai dengan pembahasan terhadap hasil yang diperoleh

7. Kesimpulan dan Saran

Bagian ini berisi kesimpulan yang dirumuskan berdasarkan hasil analisis data yang telah dilakukan sebelumnya, sebagai jawaban terhadap rumusan masalah. Saran disusun sebagai bentuk evaluasi dan pertimbangan untuk pengembangan penelitian lebih lanjut maupun perbaikan pada sistem yang dianalisis.

3.3 Desain Sistem

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



Gambar 3. 2 Desain Sistem

1. Data Tweet Wattpad

Dalam tahapan ini data yang di input merupakan data tweet yang diambil dari X dengan kata kunci yang berhubungan dengan Wattpad, seperti "wattpad" atau hashtag #wattpad.

2. Pre-processsing Data

Dalam tahapan ini dilakukan penyaringan pada data yang akan diolah, mempersiapkan data melalui tahapan pra processing data yang secara umum, *cleansing, case folding, tokenizing, normalization, stopword removal, stemming*. Dalam alur *preprocessing data* ini dilakukan untuk mempersiapkan data agar lebih mudah dalam tahapan analisis untuk memberi hasil yang sesuai dengan tujuan penelitian

a. *Cleansing*

Cleansing adalah langkah menghapus komponen-komponen yang tidak penting, karakter-karakter yang tidak baku, atau komponen-komponen yang tidak ada kaitannya dengan data dokumen seperti emoticon, angka, tanda baca, simbol, spasi berlebih, enter, tautan, dan tagar. Sehingga hasil dari tahap cleansing berupa data bersih yang hanya berisi kata-kata yang penting saja.

Tabel 3. 1 Contoh Proses Cleansing

	Sebelum	Sesudah
D1	Akhirnya watsapp ku sudah lancar kembali jadi bisa komentar lagi	Akhirnya watsapp ku sudah lancar kembali jadi bisa komentar lagi
D2	Lucu juga fitur ini, bisa membuat kalimat sendiri	Lucu juga fitur ini bisa membuat kalimat sendiri
D3	Lagi-lagi fitur untuk promo cerita hilang dari watsapp	Lagi lagi fitur untuk promo cerita hilang dari watsapp
D4	Watsapp emang error ya dari tadi notifikasi nggak muncul-muncul	Watsapp emang error ya dari tadi notifikasi nggak muncul muncul
D5	Hampir 3 hari error habis pembaharuan	Hampir 3 hari error habis pembaharuan
D6	Dulu sering banget baca cerita di watsapp, tapi makin ke sini kok makin nyeleneh	Dulu sering banget baca cerita di watsapp tapi makin ke sini kok makin nyeleneh
D7	Yang fiturnya paling asik buat membaca daring sejauh ini masih watsapp	Yang fiturnya paling asik buat membaca daring sejauh ini masih watsapp
D8	Nggak suka watsapp, pasti ada yang bagus tapi sudah keburu nggak suka sama tampilannya	Nggak suka watsapp pasti ada yang bagus tapi sudah keburu nggak suka sama tampilannya
D9	Error terus	Error terus
D10	Senang watsapp sekarang ada fitur baru	Senang watsapp sekarang ada fitur baru

Pada tabel 3.1 menampilkan hasil data sebelum dan sesudah dilakukannya tahapan cleansing yang mana menghapus komponen tidak ada kaitannya dengan dokumen seperti emoticon, angka, tanda baca, simbol, spasi, enter, tautan dan tagar seperti “Lucu juga fitur ini, bisa membuat kalimat sendiri” menjadi “Lucu juga fitur ini bisa membuat kalimat sendiri” menghilangkan tanda baca “,” ditengah kalimat. Selanjutnya, melakukan *case folding*.

b. *Case folding*

Pada tahap *case folding*, seluruh huruf yang ada pada teks akan diubah menjadi lowercase atau huruf kecil. Sehingga semua huruf-huruf pada teks akan sama bentuknya. Proses *case folding* bertujuan untuk mempermudah proses penelusuran data. Contohnya, penelusuran data yang berisi kata “butuh” namun hasil pencarian menunjukkan tidak ada, karena kata “lolos” di indeks sebagai “Butuh” atau “BUTUH”.

Tabel 3. 2 Contoh Proses *Case folding*

	Sebelum	Sesudah
D1	Akhirnya wamppad ku sudah lancar kembali jadi bisa komentar lagi	Akhirnya wamppad ku sudah lancar kembali jadi bisa komentar lagi
D2	Lucu juga fitur ini bisa membuat kalimat sendiri	lucu juga fitur ini bisa membuat kalimat sendiri
D3	Lagi lagi fitur untuk promo cerita hilang dari wamppad	lagi lagi fitur untuk promo cerita hilang dari wamppad
D4	Wamppad emang error ya dari tadi notifikasi nggak muncul muncul	wamppad emang error ya dari tadi notifikasi nggak muncul muncul
D5	Hampir 3 hari error habis pembaharuan	hampir 3 hari error habis pembaharuan
D6	Yang fiturnya paling asik buat membaca daring sejauh ini masih wamppad	yang fiturnya paling asik buat membaca daring sejauh ini masih wamppad

	Sebelum	Sesudah
D7	Dulu sering banget baca cerita di wattpad tapi makin ke sini kok makin nyeleneh	dulu sering banget baca cerita di wattpad tapi makin ke sini kok makin nyeleneh
D8	Nggak suka wattpad pasti ada yang bagus tapi sudah keburu nggak suka sama tampilannya	nggak suka wattpad pasti ada yang bagus tapi sudah keburu nggak suka sama tampilannya
D9	Error terus	error terus
D10	Senang wattpad sekarang ada fitur baru	senang wattpad sekarang ada fitur baru

Pada tabel 3.2 menampilkan hasil sebelum dan sesudah dilakukannya proses *case folding* yang mana dilakukannya proses perubahan huruf pada teks agar mempermudah penelusuran data seperti kalimat diatas yang awalnya “Dulu sering banget baca cerita di wattpad tapi makin ke sini kok makin nyeleneh” menjadi “dulu sering banget baca cerita di wattpad tapi makin ke sini kok makin nyeleneh” merubah huruf kapital pada awalan kalimat menjadi huruf kecil agar memudahkan proses penelusuran data. Selanjutnya, melakukan tahapan *tokenizing*.

c. *Tokenizing*

Tokenizing merupakan tahapan dalam *text preprocessing* yang dilakukan untuk memotong kalimat menjadi kata-kata yang menjadi penyusunnya. Data- data yang telah terpotong tersebut disebut dengan token. Proses *tokenizing* berfungsi untuk memudahkan tahap investigasi. Lebih lanjut lagi peneliti mengimplementasikan pembobotan kata TF-IDF, maka *tokenizing* sangat diperlukan karena TF-IDF dalam prosesnya menggunakan token.

Tabel 3.3 Contoh Proses *Tokenizing*

	Sebelum	Sesudah
D1	akhirnya watsapp ku sudah lancar kembali jadi bisa komentar lagi	['akhirnya', 'watsapp', 'ku', 'sudah', 'lancar', 'kembali', 'jadi', 'bisa', 'komentar', 'lagi']
D2	lucu juga fitur ini bisa membuat kalimat sendiri	['lucu', 'juga', 'fitur', 'ini', 'bisa', 'membuat', 'kalimat', 'sendiri']
D3	lagi lagi fitur untuk promo cerita hilang dari watsapp	['lagi', 'lagi', 'fitur', 'untuk', 'promo', 'cerita', 'hilang', 'dari', 'watsapp']
D4	watsapp emang error ya dari tadi notifikasi nggak muncul muncul	['watsapp', 'emang', 'error', 'ya', 'dari', 'tadi', 'notifikasi', 'nggak', 'muncul', 'muncul']
D5	hampir 3 hari error habis pembaharuan	['hampir', '3', 'hari', 'error', 'habis', 'pembaharuan']
D6	yang fiturnya paling asik buat membaca daring sejauh ini masih watsapp	['yang', 'fiturnya', 'paling', 'asik', 'buat', 'membaca', 'daring', 'sejauh', 'ini', 'masih', 'watsapp']
D7	dulu sering banget baca cerita di watsapp tapi makin ke sini kok makin nyeleneh	['dulu', 'sering', 'banget', 'baca', 'cerita', 'di', 'watsapp', 'tapi', 'makin', 'ke', 'sini', 'kok', 'makin', 'nyeleneh']
D8	nggak suka watsapp pasti ada yang bagus tapi sudah keburu nggak suka sama tampilannya	['nggak', 'suka', 'watsapp', 'pasti', 'ada', 'yang', 'bagus', 'tapi', 'sudah', 'keburu', 'nggak', 'suka', 'sama', 'tampilannya']
D9	error terus	['error', 'terus']
D10	senang watsapp sekarang ada fitur baru	['senang', 'watsapp', 'sekarang', 'ada', 'fitur', 'baru']

Pada tabel 3.3 menampilkan hasil sebelum dan sesudah dilakukannya tahapan *tokenizing* yang mana dilakukan untuk memotong kalimat menjadi kata-kata yang menjadi penyusunnya. Data-data yang telah terpotong tersebut disebut dengan token, dari awalan seperti ini “senang watsapp sekarang ada fitur baru” menjadi seperti ini ['senang', 'watsapp', 'sekarang', 'ada', 'fitur', 'baru'].

d. *Normalization*

Pada tahap *normalization*, seluruh kata yang memiliki kesalahan dalam ejaan, noise, ataupun bahasa yang tidak baku diganti menjadi bahasa baku. Kata yang mengandung noise sendiri merupakan kata yang mengandung unsur-unsur bahasa daerah dan tidak sesuai dengan Kamus Besar Bahasa Indonesia (Septian dkk., 2019).

Tabel 3. 4 Contoh Proses *Normalization*

	Sebelum	Sesudah
D1	['akhirnya', 'wattpad', 'ku', 'sudah', 'lancar', 'kembali', 'jadi', 'bisa', 'komentar', 'lagi']	['akhirnya', 'wattpad', 'ku', 'sudah', 'lancar', 'kembali', 'jadi', 'bisa', 'komentar', 'lagi']
D2	['lucu', 'juga', 'fitur', 'ini', 'bisa', 'membuat', 'kalimat', 'sendiri']	['lucu', 'juga', 'fitur', 'ini', 'bisa', 'membuat', 'kalimat', 'sendiri']
D3	['lagi', 'lagi', 'fitur', 'untuk', 'promo', 'cerita', 'hilang', 'dari', 'wattpad']	['lagi', 'lagi', 'fitur', 'untuk', 'promo', 'cerita', 'hilang', 'dari', 'wattpad']
D4	['wattpad', 'emang', 'error', 'ya', 'dari', 'tadi', 'notifikasi', 'nggak', 'muncul', 'muncul']	['wattpad', 'emang', 'eror', 'ya', 'dari', 'tadi', 'notifikasi', 'tidak', 'muncul', 'muncul']
D5	['hampir', '3', 'hari', 'error', 'habis', 'pembaharuan']	['hampir', '3', 'hari', 'eror', 'habis', 'pembaruan']
D6	['yang', 'fiturnya', 'paling', 'asik', 'buat', 'membaca', 'daring', 'sejauh', 'ini', 'masih', 'wattpad']	['yang', 'fiturnya', 'paling', 'asik', 'buat', 'membaca', 'daring', 'sejauh', 'ini', 'masih', 'wattpad']
D7	['dulu', 'sering', 'banget', 'baca', 'cerita', 'di', 'wattpad', 'tapi', 'makin', 'ke', 'sini', 'kok', 'makin', 'nyeleneh']	['dulu', 'sering', 'banget', 'baca', 'cerita', 'di', 'wattpad', 'tapi', 'makin', 'ke', 'sini', 'kok', 'makin', 'aneh']
D8	['nggak', 'suka', 'wattpad', 'pasti', 'ada', 'yang', 'bagus', 'tapi', 'sudah', 'keburu', 'nggak', 'suka', 'sama', 'tampilannya']	['tidak', 'suka', 'wattpad', 'pasti', 'ada', 'yang', 'bagus', 'tapi', 'sudah', 'keburu', 'tidak', 'suka', 'sama', 'tampilannya']
D9	['error', 'terus']	['eror', 'terus']
D10	['senang', 'wattpad', 'sekarang', 'ada', 'fitur', 'baru']	['senang', 'wattpad', 'sekarang', 'ada', 'fitur', 'baru']

Pada tabel 3.4 menampilkan hasil sebelum dan sesudah dilakukannya tahapan *normalization* seluruh kata yang memiliki kesalahan dalam ejaan,

noise, ataupun bahasa yang tidak baku diganti menjadi bahasa baku yang awalnya seperti ini “[‘wattpad’, ‘emang’, ‘error’, ‘ya’, ‘dari’, ‘tadi’, ‘notifikasi’, ‘nggak’, ‘muncul’, ‘muncul’]” diubah menjadi seperti ini “[‘wattpad’, ‘emang’, ‘eror’, ‘ya’, ‘dari’, ‘tadi’, ‘notifikasi’, ‘tidak’, ‘muncul’, ‘muncul’]” merubah kalimat tidak baku “nggak” menjadi “tidak” . Selanjutnya, melakukan tahapan stopword removal

e. *Stopword Removal*

Stopword removal merupakan proses menghapus kata-kata umum yang tidak memiliki makna penting dalam konteks analisis, seperti “yang”, “dan”, atau “dengan”, agar model klasifikasi teks dapat fokus pada kata-kata bermakna dan meningkatkan akurasi analisis (Agil Santosa dkk., 2022).

Tabel 3. 5 Contoh Proses *Stopword Removal*

	Sebelum	Sesudah
D1	[‘akhirnya’, ‘wattpad’, ‘ku’, ‘sudah’, ‘lancar’, ‘kembali’, ‘jadi’, ‘bisa’, ‘komentar’, ‘lagi’]	[‘akhirnya’, ‘wattpad’, ‘ku’, ‘sudah’, ‘lancar’, ‘kembali’, ‘jadi’, ‘bisa’, ‘komentar’, ‘lagi’]
D2	[‘lucu’, ‘juga’, ‘fitur’, ‘ini’, ‘bisa’, ‘membuat’, ‘kalimat’, ‘sendiri’]	[‘lucu’, ‘juga’, ‘fitur’, ‘ini’, ‘bisa’, ‘membuat’, ‘kalimat’, ‘sendiri’]
D3	[‘lagi’, ‘lagi’, ‘fitur’, ‘untuk’, ‘promo’, ‘cerita’, ‘hilang’, ‘dari’, ‘wattpad’]	[‘lagi’, ‘lagi’, ‘fitur’, ‘untuk’, ‘promo’, ‘cerita’, ‘hilang’, ‘dari’, ‘wattpad’]
D4	[‘wattpad’, ‘emang’, ‘eror’, ‘ya’, ‘dari’, ‘tadi’, ‘notifikasi’, ‘tidak’, ‘muncul’, ‘muncul’]	[‘wattpad’, ‘emang’, ‘eror’, ‘dari’, ‘tadi’, ‘notifikasi’, ‘tidak’, ‘muncul’]
D5	[‘hampir’, ‘3’, ‘hari’, ‘eror’, ‘habis’, ‘pembaruan’]	[‘hampir’, ‘3’, ‘hari’, ‘eror’, ‘habis’, ‘pembaruan’]
D6	[‘yang’, ‘fiturnya’, ‘paling’, ‘asik’, ‘buat’, ‘membaca’, ‘daring’, ‘sejauh’, ‘ini’, ‘masih’, ‘wattpad’]	[‘fiturnya’, ‘paling’, ‘asik’, ‘buat’, ‘membaca’, ‘daring’, ‘sejauh’, ‘ini’, ‘masih’, ‘wattpad’]
D7	[‘dulu’, ‘sering’, ‘banget’, ‘baca’, ‘cerita’, ‘di’, ‘wattpad’, ‘tapi’, ‘makin’, ‘ke’, ‘sini’, ‘kok’, ‘makin’, ‘aneh’]	[‘dulu’, ‘sering’, ‘baca’, ‘cerita’, ‘di’, ‘wattpad’, ‘tapi’, ‘makin’, ‘makin’, ‘aneh’]
D8	[‘tidak’, ‘suka’, ‘wattpad’, ‘pasti’, ‘ada’, ‘yang’, ‘bagus’,	[‘tidak’, ‘suka’, ‘wattpad’, ‘pasti’, ‘ada’, ‘bagus’, ‘tapi’, ‘sudah’,

	Sebelum	Sesudah
	‘tapi’, ‘sudah’, ‘keburu’, ‘tidak’, ‘suka’, ‘sama’, ‘tampilannya’]	‘keburu’, ‘tidak’, ‘suka’, ‘sama’, ‘tampilannya’]
D9	[‘eror’, ‘terus’]	[‘eror’, ‘terus’]
D10	[‘senang’, ‘wattpad’, ‘sekarang’, ‘ada’, ‘fitur’, ‘baru’]	[‘senang’, ‘wattpad’, ‘sekarang’, ‘ada’, ‘fitur’, ‘baru’]

Pada tabel 3.5 menampilkan hasil sebelum dan sesudah dilakukannya proses *stopword removal* yang mana tahapan yang digunakan untuk menghilangkan kata-kata umum yang sering muncul dan tidak memiliki makna tertentu seperti tabel D4 dari seperti ini “[‘wattpad’, ‘emang’, ‘eror’, ‘ya’, ‘dari’, ‘tadi’, ‘notifikasi’, ‘tidak’, ‘muncul’, ‘muncul’]” menjadi seperti ini “[‘wattpad’, ‘emang’, ‘eror’, ‘dari’, ‘tadi’, ‘notifikasi’, ‘tidak’, ‘muncul’]” menghilangkan kata “muncul” diakhir kalimat.

f. *Stemming*

Stemming merupakan tahap akhir dalam proses *text preprocessing* seluruh kata dalam dokumen akan diubah menjadi kata dasar. Sehingga, semua kata imbuhan mulai dari awalan (*prefix*), akhiran (*suffix*), sisipan (*infix*), serta gabungan antara awalan dan akhiran (*confix*) akan dihilangkan. *Stemming* dilakukan menggunakan *library* Sastrawi dalam penyesuaian supaya tidak banyak variasi kata mengurangi *over stemming* dan *under stemming* serta waktu pengerjaan yang lebih cepat dibandingkan dengan menggunakan stemmer lain karena dirancang untuk dokumen yang berbahasa Indonesia (Alfan Rosid dkk., 2020).

Tabel 3. 6 Contoh Proses Stemming

	Sebelum	Sesudah
D1	[‘akhirnya’, ‘wattpad’, ‘ku’, ‘sudah’, ‘lancar’, ‘kembali’, ‘jadi’, ‘bisa’, ‘komentar’, ‘lagi’]	[‘akhir’, ‘wattpad’, ‘ku’, ‘sudah’, ‘lancar’, ‘kembali’, ‘jadi’, ‘bisa’, ‘komentar’, ‘lagi’]
D2	[‘lucu’, ‘juga’, ‘fitur’, ‘ini’, ‘bisa’, ‘membuat’, ‘kalimat’, ‘sendiri’]	[‘lucu’, ‘juga’, ‘fitur’, ‘ini’, ‘bisa’, ‘buat’, ‘kalimat’, ‘sendiri’]

	Sebelum	Sesudah
D3	['lagi', 'lagi', 'fitur', 'untuk', 'promo', 'cerita', 'hilang', 'dari', 'wattpad']	['lagi', 'lagi', 'fitur', 'untuk', 'promo', 'cerita', 'hilang', 'dari', 'wattpad']
D4	['wattpad', 'emang', 'eror', 'dari', 'tadi', 'notifikasi', 'tidak', 'muncul']	['wattpad', 'emang', 'eror', 'dari', 'tadi', 'notifikasi', 'tidak', 'muncul']
D5	['hampir', '3', 'hari', 'eror', 'habis', 'pembaruan']	['hampir', '3', 'hari', 'eror', 'habis', 'pembaruan']
D6	['fiturnya', 'paling', 'asik', 'buat', 'membaca', 'daring', 'sejauh', 'ini', 'masih', 'wattpad']	['fitur', 'paling', 'asik', 'buat', 'baca', 'daring', 'jauh', 'ini', 'masih', 'wattpad']
D7	['dulu', 'sering', 'baca', 'cerita', 'di', 'wattpad', 'tapi', 'makin', 'makin', 'aneh']	['dulu', 'sering', 'baca', 'cerita', 'di', 'wattpad', 'tapi', 'makin', 'makin', 'aneh']
D8	['tidak', 'suka', 'wattpad', 'pasti', 'ada', 'bagus', 'tapi', 'sudah', 'keburu', 'tidak', 'suka', 'sama', 'tampilannya']	['tidak', 'suka', 'wattpad', 'pasti', 'ada', 'bagus', 'tapi', 'sudah', 'buru', 'tidak', 'suka', 'sama', 'tampil']
D9	['eror', 'terus']	['eror', 'terus']
D10	['senang', 'wattpad', 'sekarang', 'ada', 'fitur', 'baru']	['senang', 'wattpad', 'sekarang', 'ada', 'fitur', 'baru']

Pada tabel 3.6 menampilkan hasil sebelum dan sesudah dari proses stemming seluruh kata akan diubah menjadi kata dasar seperti di tabel D8 dari seperti ini “[‘tidak’, ‘suka’, ‘wattpad’, ‘pasti’, ‘ada’, ‘bagus’, ‘tapi’, ‘sudah’, ‘keburu’, ‘tidak’, ‘suka’, ‘sama’, ‘tampilannya’]” menjadi seperti ini “[‘tidak’, ‘suka’, ‘wattpad’, ‘pasti’, ‘ada’, ‘bagus’, ‘tapi’, ‘sudah’, ‘buru’, ‘tidak’, ‘suka’, ‘sama’, ‘tampil’]” merubah beberapa kata dari “tampilannya” menjadu “tampil”. Selanjutnya, melakukan tahapan pembobotan data

3. Pembobotan Data

Pembobotan Data atau TF-IDF (Term Frequency–Inverse Document Frequency) adalah metode pembobotan yang mengukur seberapa penting sebuah kata dalam sebuah dokumen relatif terhadap seluruh kumpulan dokumen. Kata

yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain akan memiliki bobot lebih tinggi, sehingga lebih relevan dalam proses klasifikasi (Assidyk dkk., 2020). Persamaan pembobotan kata dengan TF-IDF adalah sebagai berikut:

$$TF = \frac{\text{jumlah kemunculan kata dalam dokumen}}{\text{jumlah kata pada dokumen}} \quad (3.1)$$

$$IDF = \log \frac{D}{DF} \quad (3.2)$$

$$TF - IDF = TF \times IDF \quad (3.3)$$

Keterangan:

TF-IDF: Bobot kata dalam sebuah dokumen

TF : Frekuensi kemunculan kata dalam sebuah dokumen

IDF : Frekuensi data kata

D : Jumlah total dokumen

DF : Jumlah dokumen yang mengandung kata

Kumpulan kata yang diolah pada tahap *text preprocessing* akan dijadikan term atau aturan untuk proses selanjutnya, yaitu pembobotan kata. Pembobotan kata dilakukan dengan mengimplementasikan TF-IDF atau Term Frequency-Inverse Document Frequency. Untuk menghitung bobot kata menggunakan TF-IDF, dapat dilakukan menggunakan persamaan 3.1 hingga 3.3 pada penelitian ini, perhitungan manual akan menerapkan 10 data sebagai sampel.

Tabel 3.7 Contoh Data Terlabel

	<i>Tweet</i>	Label
D1	akhirnya wattpad ku sudah lancar kembali jadi bisa komentar lagi	Positif
D2	lucu juga fitur ini, bisa membuat kalimat sendiri	Positif
D3	lagi-lagi fitur untuk promo cerita hilang dari wattpad	Negatif
D4	wattpad emang error ya dari tadi notifikasi nggak muncul muncul	Negatif
D5	hampir 3 hari error habis pembaharuan	Negatif

	<i>Tweet</i>	Label
D6	yang fiturnya paling asik buat membaca daring sejauh ini masih wappad	Positif
D7	dulu sering banget baca cerita di wappad, tapi makin ke sini kok makin nyeleneh	Negatif
D8	nggak suka wappad, pasti ada yang bagus tapi sudah keburu nggak suka sama tampilannya	Negatif
D9	error terus	Negatif
D10	senang wappad sekarang ada fitur baru	Positif
D11	kinerja wappad sangat lambat	Negatif
D12	saya menyukai tampilan baru wappad, namun fiturnya tidak berjalan lancar	Positif
D13	tidak ada yang istimewa dari wappad	Positif
D14	perlu perbaikan	Positif
D15	wappad memiliki cerita yang seru	Positif

Pada tabel 3.10 menampilkan hasil dari proses pembobotan data. Selanjutnya, Tahap pertama yang harus dilakukan dalam perhitungan TF-IDF adalah menghitung menghitung nilai TF terlebih dahulu. Nilai TF didapatkan dengan membagi jumlah kata “x” yang muncul dalam dokumen “d” dengan total kata dalam dokumen “d”. Contoh perhitungan nilai TF pada suatu kata dipaparkan pada tabel berikut.

Tabel 3. 8 Contoh Perhitungan TF

	Jumlah Kata	Kata	Frekuensi	TF
D1	9	Wattpad	1	1/9
D2	7		0	0/7
D3	8		1	1/8
D4	7		1	1/7
D5	5		0	0/5
D6	8		1	1/8
D7	7		1	1/7
D8	9		1	1/9
D9	2	Wattpad	0	0/2
D10	6		1	1/6

Tabel 3.8 merupakan tabel yang berisi perhitungan nilai Term Frequency (TF). Nilai Term Frequency (TF) diperoleh dengan cara membagi jumlah frekuensi kata terpilih dengan jumlah kata pada suatu dokumen. Sebagai contoh, pada D1 terdapat sebanyak 9 kata di dalamnya dan kata “wattpad” muncul sebanyak 1 kali pada D1, sehingga nilai TF untuk kata “wattpad” pada D1 adalah $1/9$. Selanjutnya, kata “wattpad” juga muncul sebanyak 1 kali pada D3 di mana total kata di D3 ditemukan 8 kata, sehingga nilai TF untuk kata “wattpad” pada D3 adalah $1/8$. Kemudian, kata “wattpad” muncul juga sebanyak 1 kali pada D4 di mana total kata pada D4 adalah 7 kata, sehingga nilai TF untuk kata “wattpad” pada D3 adalah $1/7$. Nilai TF-IDF kemudian ditentukan dengan menggunakan persamaan 3.2 dan 3.3 berdasarkan nilai TF yang dihitung pada Tabel 3.9. Berikut contoh perhitungan TF-IDF untuk kata atau term “wattpad” pada D1, D3, dan D4.

A Perhitungan TF-IDF kata wattpad pada D1

$$IDF = \log \frac{D}{DF}$$

$$IDF = \log \frac{10}{7} = 0,15490196 \quad (3.4)$$

$$TF - IDF = TF \times IDF$$

$$TF - IDF = 0,1111 \times 0,15490196 = 0,0172096078$$

B Perhitungan TF-IDF kata wattpad pada D3

$$IDF = \log \frac{D}{DF}$$

$$IDF = \log \frac{10}{7} = 0,15490196 \quad (3.5)$$

$$TF - IDF = TF \times IDF$$

$$TF - IDF = 0,125 \times 0,15490196 = 0,01936274$$

C Perhitungan TF-IDF kata wattpad pada D4

$$IDF = \log \frac{D}{DF}$$

$$IDF = \log \frac{10}{7} = 0,15490196 \quad (3.6)$$

$$TF - IDF = TF \times IDF$$

$$TF - IDF = 0,14285714 \times 0,15490196 = 0,02212885$$

Dan selanjutnya untuk perhitungan TF - IDF pada kata selanjutnya

Tabel 3. 9 Contoh Hasil Perhitungan TF-IDF

Kata	Jumlah Kata			
	D1	D2	D3	D4
Wattpad	0,017211329	-	0,019362745	0,022128851

Pada tabel 3.9 merupakan hasil perhitungan nilai TF-IDF untuk seluruh kata pada setiap dokumen sampel.

4. Pembagian Data Training dan Data Testing

Pembagian data training dan data testing dengan menggunakan K-Fold Cross Validation adalah cara untuk mengevaluasi performa model dengan membagi data menjadi beberapa bagian yang sama besar. Setiap bagian secara bergantian dipakai sebagai data testing, sementara bagian lainnya dipakai untuk melatih model. Proses ini diulang sebanyak jumlah bagian yang dibuat, sehingga hasil evaluasi model menjadi lebih akurasi dan lebih dapat dipercaya (Musu Wilem dkk, 2021).

5. Implementasi Naïve Bayes

Setelah dilakukan pembobotan kata menggunakan metode TF-IDF, langkah berikutnya menerapkan perhitungan menggunakan metode klasifikasi *Naïve Bayes*. Metode *Naïve Bayes* merupakan sebuah metode pengklasifikasian probabilistik sederhana dengan cara kerja yaitu menghitung sekumpulan probabilitas yang didapat dengan menjumlahkan frekuensi dari kombinasi nilai

dan dataset yang digunakan (Rezki dkk, 2020). Analisis data dilakukan menggunakan algoritma *Naïve Bayes* yang penting dalam klasifikasi sentimen termasuk ke dalam klasifikasi positif atau negatif. Pada tahap ini dilakukan dua proses utama dalam pengklasifikasian yaitu proses training dan proses testing.

A. Training

Proses training algoritma *Naïve Bayes* ada beberapa tahapan yang dilakukan yaitu sebagai berikut.

a. Penentuan kelas data training

Penentuan kelas pada data dilakukan secara manual dengan mengklasifikasikan dokumen training termasuk ke dalam kelas positif atau negatif.

Tabel 3. 10 Contoh Proses Data Training

	<i>Tweet</i>	Label
D1	akhir wattpad sudah lancar kembali jadi bisa komentar lagi	Positif
D2	lucu juga fitur bisa buat kalimat sendiri	Positif
D3	lagi-lagi fitur untuk promo cerita hilang dari wattpad	Negatif
D4	wattpad error dari tadi notifikasi tidak muncul-muncul	Negatif
D5	hampir hari error habis baru	Negatif
D6	fitur paling asik baca daring masih wattpad	Positif
D7	sering baca cerita wattpad, tapi makin sini makin	Negatif
D8	tidak suka wattpad, pasti ada bagus tapi sudah tidak suka tampil	Negatif

Pada tabel 3.10 menampilkan hasil penentuan data training dalam kelas positif maupun negatif.

b. Perhitungan Probabilitas

Berikut ini merupakan perhitungan probabilitas data training.

$$P(\text{Positif}) = \frac{\text{Jumlah tweet positif}}{\text{jumlah total tweet}} = \frac{3}{8} = 0,375$$

$$P(\text{Negatif}) = \frac{\text{Jumlah tweet negatif}}{\text{jumlah total tweet}} = \frac{5}{8} = 0,675 \quad (3. 7)$$

Selanjutnya menghitung frekuensi kata dalam tweet positif dan negatif. Berikut masing-masing kata dalam tweet positif dan negatif.

a) Tweet Positif

Tabel 3. 11 Contoh Proses Frekuensi Kata Positif

Tweet Positif	Frekuensi Kata
Akhir	1
Wattpad	2
Lancar	1
Kembali	1
Komentar	1
Lucu	1
Fitur	1
Buat	1
Kalimat	1
Sendiri	1
Paling	1
Asik	1
Baca	1
Daring	1
Masih	1

Pada tabel 3.11 menampilkan hasil frekuensi kata pada tweet positif. Selanjutnya, akan dilakukan penghitungan frekuensi kata pada tweet negatif.

b) Tweet Negatif

Tabel 3. 12 Contoh Proses Frekuensi Kata Negatif

Tweet Negatif	Frekuensi Kata
Lagi – lagi	1
Fitur	1
Promo	1
Cerita	1
Hilang	1
Wattpad	3

Pada tabel 3.16 merupakan hasil dari perhitungan probabilitas kata pada tweet D4. Selanjutnya, di tabel 3.20 merupakan hasil probabilitas kata pada tweet D5.

Tabel 3. 17 Contoh Proses Perhitungan Probabilitas Kata pada D5

Kelas	$P(v_j)$	Probabilitas				
		hampir	hari	error	habis	baru
Positif	$\frac{1}{2}$	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{1}{50}$
Negatif	$\frac{1}{2}$	$\frac{2}{65}$	$\frac{2}{65}$	$\frac{2}{65}$	$\frac{2}{65}$	$\frac{2}{65}$

Pada tabel 3.17 merupakan hasil dari perhitungan probabilitas kata pada tweet D5. Selanjutnya, Hasil dari semua probabilitas yang sudah dihitung dijadikan acuan untuk selanjutnya dalam menentukan data *testing*.

B. *Testing*

Pada proses *testing* dihitung probabilitasnya kemudian dicari probabilitas tertinggi.

Tabel 3. 18 Contoh Data Testing

	<i>Tweet</i>	Label
D9	error terus	Negatif
D10	senang wattpad sekarang ada fitur baru	Positif
D11	kinerja wattpad sangat lambat	Negatif
D12	saya menyukai tampilan baru wattpad, namun fiturnya tidak berjalan lancar	Positif
D13	tidak ada yang istimewa dari wattpad	Positif
D14	perlu perbaikan	Positif
D15	wattpad memiliki cerita yang seru	Positif

Pada tabel 3.18 merupakan data *testing* yang sudah terlabel antara negatif dan positif. Selanjutnya, dilakukan proses perhitungan untuk mencari probabilitas dari *testing* yang tertinggi.

Data testing positif

$$P(\text{Positif}) = \frac{2 + 1}{50} = 0,02 \quad (3. 8)$$

$$P(\text{Positif}) = \frac{0 + 1}{50} = 0,02$$

$$P(\text{error terus}) = P(\text{Positif}) \times P(\text{Positif}) \times P(\text{Positif})$$

$$= 0,375 \times 0,02 \times 0,02 = 0,00015$$

Data testing negative

$$P(\text{Negatif}) = \frac{2 + 1}{65} = 0,00462 \quad (3. 9)$$

$$P(\text{Negatif}) = \frac{0 + 1}{65} = 0,0154$$

$$P(\text{error terus}) = P(\text{Negatif}) \times P(\text{Negatif}) \times P(\text{Negatif})$$

$$= 0,625 \times 0,00462 \times 0,0154 = 0,000444$$

Karena $P(\text{error terus}) > P(\text{error terus})$, maka D9 diklasifikasikan sebagai sentimen negatif. Selanjutnya, Setelah melakukan perhitungan pada data testing yang dilakukan dengan menggunakan cara yang sama pada tiga dokumen sehingga diperoleh nilai probabilitas yang ditunjukkan pada tabel berikut.

Tabel 3. 19 Contoh Nilai Probabilitas Data Testing

Komentar	Probabilitas	
	Positif	Negatif
D9	0,00015	0,000444
D10	0,000000576	0,0000000166
D11	0,0000009925	0,0000014245
D12	0,0000009855	0,0000002561
D13	0,0000005225	0,000000854
D14	0,000409	0,000356
D15	0,000002237	0,000000077

Pada tabel 3.19 merupakan hasil dari nilai probabilitas data testing yang telah dihitung.

6. Hasil Klasifikasi

Hasil klasifikasi akan didapatkan melalui tahap klasifikasi data menggunakan *Naïve Bayes* akan ditampilkan dalam bentuk tabel. Dalam melakukan klasifikasi menggunakan algoritma *Multinomial Naïve Bayes*, dataset komentar pada penelitian ini harus melalui tahap pembagian dataset terlebih dahulu menjadi data latih dan data uji. Selanjutnya, hasil testing menunjukkan bahwa jika nilai probabilitas negatif lebih besar dari nilai probabilitas positif, maka termasuk ke dalam kategori sentimen negatif. Sedangkan jika nilai probabilitas positif lebih besar dari nilai probabilitas negatif, maka termasuk ke dalam kategori sentimen positif.

Tabel 3. 20 Contoh Hasil Klasifikasi Sentimen

	Komentar	Label	Klasifikasi
D9	error terus	Negatif	Negatif
D10	senang wattpad sekarang ada fitur baru	Positif	Positif
D11	kinerja wattpad sangat lambat	Negatif	Negatif
D12	saya menyukai tampilan baru wattpad, namun fiturnya tidak berjalan lancar	Positif	Positif
D13	tidak ada yang istimewa dari wattpad	Positif	Negatif
D14	perlu perbaikan	Negatif	Positif
D15	wattpad memiliki cerita yang seru	Positif	Positif

Pada tabel 3.20 merupakan hasil dari klasifikasi sentimen pada tweet D9 sampai D15. Selanjutnya, dilakukannya pengujian hasil klasifikasi. Setelah mendapatkan hasil klasifikasi akan dilakukan evaluasi dengan model untuk mengetahui performa kinerja pada suatu sistem. Pengujian dilakukan untuk mengetahui nilai *akurasi*, *presisi*, *recall*, dan *F1-score*. *Confusion Matrix* digunakan sebagai metode evaluasi dengan membandingkan hasil prediksi dan data aktual, sehingga memudahkan dalam menghitung keempat metrik tersebut

(Rininda dkk., 2023). Nilai *true false* dijadikan pedoman dalam menghitung nilai performa. Berikut contoh pengujian *Confusion Matrix*.

Tabel 3. 21 Contoh Hasil Pengujian Confusion Matrix

		Predicted Class	
		Positive	Negative
True Class	Positive	3	1
	Negative	1	2

Pada tabel 3.24 merupakan contoh hasil dari pengujian confusion matrix. Selanjutnya, akan dilakukan perhitungan *accuracy*, *precision*, *recall*, dan *f1- score*.

a. *Accuracy*

Akurasi dihitung menggunakan persamaan 2.5 berikut:

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% & (3. 10) \\
 &= \frac{3 + 2}{3 + 1 + 2 + 1} \times 100\% \\
 &= \frac{5}{7} \times 100\% \\
 &= 71,42\%
 \end{aligned}$$

b. *Precision*

Presisi dihitung menggunakan persamaan 2.6 berikut:

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \times 100\% & (3. 11) \\
 &= \frac{3}{3 + 1} \times 100\% \\
 &= \frac{3}{4} \times 100\% \\
 &= 75\%
 \end{aligned}$$

c. *Recall*

Recall dihitung menggunakan persamaan 2.7 berikut:

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% & (3.12) \\
 &= \frac{3}{3 + 1} \times 100\% \\
 &= \frac{3}{4} \times 100\% \\
 &= 75\%
 \end{aligned}$$

d. *F1-score*

F1 Score dihitung menggunakan persamaan 2.8 berikut:

$$\begin{aligned}
 \text{F1score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% & (3.13) \\
 &= 2 \times \frac{75 \times 75}{75 + 75} \times 100\% \\
 &= 2 \times \frac{5625}{150} \times 100\% \\
 &= 70\%
 \end{aligned}$$

BAB IV

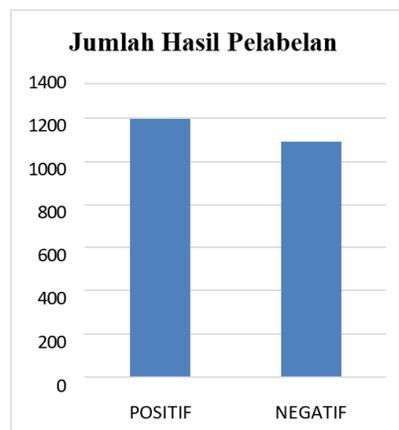
HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pembahasan difokuskan pada hasil pengujian model *machine learning* serta penerapan metode *Naïve Bayes* dalam analisis sentimen pengguna X terhadap aplikasi Wattpad melalui pendekatan crawling. Proses yang dilaksanakan mencakup tahapan pra-pemrosesan data, penerapan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), serta pembagian dataset menjadi data latih dan data uji dengan rasio 90:10, 80:20, dan 70:30. Evaluasi model dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-score. Selain itu, dipaparkan pula tahapan-tahapan yang diterapkan secara sistematis selama keseluruhan proses berjalan.

4.1.1 Hasil Pelabelan Data

Pada penelitian ini, rentang data yang diambil dari tanggal 01 april 2024 sampai 01 april 2025 yang menghasilkan data awal sebanyak 3.059 tweet. Selanjutnya, dilakukannya pembersihan data dan cek duplikasi pada data menghasilkan data bersih sebanyak 2.288 tweet. Untuk label positif sebanyak 1200 kemudian untuk label Negatif sebanyak 1088 tweet



Gambar 4.1 Hasil Labeling Data

Gambar 4.1 merupakan Visualisasi disajikan dalam bentuk diagram batang, di mana sumbu horizontal merepresentasikan jenis sentimen, sedangkan sumbu vertikal menunjukkan jumlah tweet pada masing-masing kategori yang ditampilkan pada Tabel 4.1, jumlah tweet dengan sentimen positif adalah sebanyak 1200 sementara tweet dengan sentimen negative berjumlah 1088 tweet.

Tabel 4. 1 Hasil Pelabelan Data

No	Comment	Label
1	di wattpad tuh kalau niat nyelem buat cari story pasti bakal nemu story yg WAH banget! story yg gini banyak yg gak masuk top tag jadi kudu sabar scroll terus	
2	gue abis kepikiran kalo dapet kerjaan mnjadi idol skandal pertama gue pasti OHH INIMAH YG DULU NULIS BXB DI WATTPAD wah.	
3	@worksfess Karangan ala wattpad kah? Kerja baru 6 bulan dengan gaji ala MT udah berkesimpulan gak ada karir. Hah?	
4	Jdi ikut penasarn gegara pas baca cerita di Wattpad ada obrolan orang yg punya usaha Clubbing itu duitnya halal apa Haram ya? Duh random nemen	

Pada table 4.1 merupakan hasil pelabelan pada data tweet. Pada hasil comment data tweet 1 berlabel positif karena mengandung beberapa kata positif yaitu wah banget. Sedangkan 2 sampai 4 berlabel negatif karena mengandung beberapa kata negatif yaitu gak ada karir, skandal.

4.1.2 Hasil Preprocessing

Data yang telah diberi label oleh ahli bahasa Indonesia akan melalui proses pembersihan guna mengoptimalkan kinerja klasifikasi. Proses pra-pemrosesan dilakukan menggunakan tools dari Kaggle dengan bahasa Python, yang mencakup pencarian dan penghapusan elemen-elemen yang tidak diperlukan seperti URL, emotikon, hashtag, serta username/mention, dengan menggantinya menjadi string kosong. Selain itu, digunakan juga library Sastrawi untuk proses tokenisasi dan stemming. Tokenisasi dilakukan dengan bantuan library NLTK, sedangkan

penyusunan data dalam bentuk dataframe menggunakan library pandas. Hasil dari tahapan *case folding*, *filtering*, *tokenizing*, dan *stemming* ditampilkan sebagai berikut.

Tabel 4. 2 Tabel Pre-processinf

No	Data sebelum <i>Preprocessing</i>
1	di wattpad tuh kalau niat nyelem buat cari story pasti bakal nemu story yg WAH banget! Story yg gini banyak yg gak masuk top tag jadi kudu sabar scroll terus
2	gue abis kepikiran kalo dapet kerjaan mnjdi idol skandal pertama gue pasti OHH INIMAH YG DULU NULIS BXB DI WATTPAD wah.
3	@worksfess Karangan ala wattpad kah? Kerja baru 6 bulan dengan gaji ala MT udah berkesimpulan gak ada karir. Hah?
4	Jdi ikut penasarn gegara pas baca cerita di Wattpad ada obrolan orang yg punya usaha Clubbing itu duitnya halal apa Haram ya? Duh random nemen

Pada table 4.2 menampilkan hasil data mentahan yang berhasil diambil dari proses crawling data di aplikasi X selanjutnya, melakukan *case folding* pada table 4.3.

Tabel 4. 3 Tabel Case folding

No	<i>Case folding</i>
1	di wattpad tuh kalau niat nyelem buat cari story pasti bakal nemu story yg wah banget! story yg gini banyak yg gak masuk top tag jadi kudu sabar scroll terus
2	gue abis kepikiran kalo dapet kerjaan mnjdi idol skandal pertama gue pasti ohh inimah yg dulu nulis bxb di wattpad wah.
3	worksfess karangan ala wattpad kah? kerja baru 6 bulan dengan gaji ala mt udah berkesimpulan gak ada karir. hah?
4	jdi ikut penasarn gegara pas baca cerita di wattpad ada obrolan orang yg punya usaha clubbing itu duitnya halal apa haram ya? duh random nemen

Pada tabel 4.3 menampilkan tahapan pembersihan data, yang mencakup penghapusan URL, emotikon, *hashtag*, *username/mention*, spasi berlebih, serta konversi seluruh huruf menjadi huruf kecil (*lowercase*).

Tabel 4. 4 Tabel Filtering

No	Text
1	di wattpad tuh kalau niat nyelem buat cari story pasti bakal nemu story yg wah banget story yg gini banyak yg gak masuk top tag jadi kudu sabar scroll terus
2	gue abis kepikiran kalo dapet kerjaan mnjdi idol skandal pertama gue pasti ohh inimah yg dulu nulis bxb di wattpad wah wattpad wah
3	karangan ala wattpad kah kerja baru bulan dengan gaji ala mt udah berkesimpulan gak ada karir hah
4	jdi ikut penasarn gegara pas baca cerita di wattpad ada obrolan orang yg punya usaha clubbing itu duitnya halal apa haram ya duh random nemen

Tabel 4.4 menyajikan hasil dari proses pembersihan simbol dan karakter yang tidak relevan dalam teks, seperti hashtag (#), mention (@), angka, tanda baca, emoji, serta penghapusan spasi berlebih dan kata-kata yang terlalu umum. Setelah tahap ini, dilakukan proses tokenisasi, yang hasilnya ditampilkan pada Tabel 4.5.

Tabel 4. 5 Tabel Tokenizing

No	<i>Tokenizing</i>
1	[wattpad, tuh, kalau, niat, nyelem, buat, cari, story, bakal, nemu, story, yg, wah, banget, story, yg, gini, banyak, yg, gak, masuk, top, tag, jadi, kudu, sabar, scroll, terus]
2	[gue, abis, kepikiran, kalo, dapet, kerjaan, mnjdi, idol, skandal, pertama, gue, ohh, inimah, yg, dulu, nulis, bxb, wattpad, wah]
3	[karangan, ala, wattpad, kerja, baru, bulan, gaji, ala, mt, udah, berkesimpulan, gak, karir, hah]
4	[jdi, ikut, penasarn, gegara, pas, baca, cerita, wattpad, obrolan, orang, yg, punya, usaha, clubbing, duitnya, halal, apa, haram, duh, random, nemen]

Tabel 4.5 hasil dari proses pemisahan kata (tokenisasi), di mana setiap kata dalam data dipisahkan untuk selanjutnya diberikan bobot menggunakan metode TF-IDF. Setelah itu proses stemming ditampilkan pada Tabel 4.6.

Tabel 4. 6 Tabel Stemming

No	Text
1	wattpad tuh kalau niat nyelem buat cari story bakal nemu story yg wah banget story yg gin banyak yg gak masuk top tag jadi kudu sabar scroll terus
2	gue abis pikir kalo dapet kerja mnjdi idol skandal pertama gue ohh inimah yg dulu nulis bxb wattpad wah
3	karang ala wattpad kerja baru bulan gaji ala mt udah simpul gak karir hah
4	jdi ikut penasarn gegara pas baca cerita wattpad obrol orang yg punya usaha clubbing duit halal apa haram duh random nemen

Tabel 4.3 hingga Tabel 4.6 merupakan hasil representasi data frame yang menampilkan empat contoh data teratas dari total 2.288 kata setelah proses pembersihan. Seluruh data ini disimpan dalam format CSV dan siap untuk diberi bobot menggunakan metode TF-IDF.

4.1.3 Hasil Pembagian *Dataset*

Setelah melalui proses pra-pemrosesan, diperoleh sebanyak 2.288 data yang telah dibersihkan. Data tersebut kemudian dibagi menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk melakukan ekstraksi fitur menggunakan metode TF-IDF serta untuk melatih model klasifikasi *Naïve Bayes*. Sementara itu, data uji digunakan untuk mengevaluasi performa model dan menghitung tingkat keberhasilan algoritma dalam melakukan klasifikasi secara akurat. Hasil pembagian dataset ini disimpan dalam format file Excel agar dapat digunakan kembali dalam pemrosesan selanjutnya tanpa perlu melakukan pembagian ulang. Pembagian data dilakukan dengan beberapa rasio, yaitu 90:10, 80:20, dan 70:30 antara data latih dan data uji. Setelah proses pembagian selesai, kedua subset data tersebut yaitu data latih dan data uji disimpan masing-masing ke dalam file Excel.

dalam penelitian ini harus terlebih dahulu dibagi menjadi data latih dan data uji. Sebagaimana dijelaskan sebelumnya, pembagian data dilakukan dengan rasio 90:10, 80:20, dan 70:30. Tabel 4.7 menyajikan perbandingan dari masing-masing rasio pembagian data tersebut guna menentukan rasio terbaik yang akan digunakan dalam penelitian ini.

Tabel 4. 7 Tabel Rasio Pembagian Data

Perbandingan data latih & data uji	Hasil Akurasi Naive Baies
90 : 10	90.83%
80 : 20	90.61%
70 : 30	89.08%

Berdasarkan Tabel 4.7, metode klasifikasi *Naïve Bayes* menunjukkan hasil terbaik pada rasio pembagian data latih dan data uji sebesar 90:10, dengan tingkat akurasi mencapai 90%. Setelah diperoleh rasio pembagian dengan akurasi tertinggi.

Selanjutnya adalah melakukan ekstraksi fitur menggunakan metode TF-IDF untuk memberikan bobot pada data latih dan data uji. Hasil dari proses ekstraksi ini akan digunakan dalam proses pelatihan dan pengujian model klasifikasi *Naïve Bayes*. Tabel berikut menyajikan hasil klasifikasi dalam bentuk *confusion matrix*.

Tabel 4. 8 Tabel Hasil Klasifikasi Rasio 90:10

		<i>Predicted</i>	
		<i>Negative</i>	<i>Positive</i>
<i>Actual</i>	<i>Positive</i>	102	18
	<i>Negative</i>	3	106

Dalam Tabel 4.8, terdapat 102 data yang berasal dari kelas positif dan berhasil diprediksi dengan benar sebagai kelas positif. Sementara itu, sebanyak 18 data dari kelas negatif salah diprediksi, sehingga tidak dikenali sebagai kelas negatif. Di sisi lain, 106 data dari kelas negatif diprediksi dengan tepat sebagai kelas negatif, namun terdapat 3 data dari kelas positif yang tidak berhasil dikenali dengan benar sebagai kelas positif.

Tabel 4. 9 Tabel Hasil Klasifikasi Rasio 80:20

		<i>Predicted</i>	
		<i>Negative</i>	<i>Positive</i>
<i>Actual</i>	<i>Positive</i>	200	37
	<i>Negative</i>	6	215

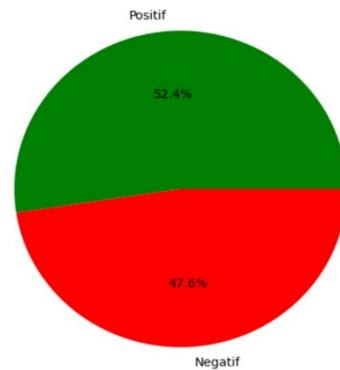
Dalam Tabel 4.9, terdapat 200 data yang berasal dari kelas positif dan berhasil diprediksi dengan benar sebagai kelas positif. Sementara itu, sebanyak 37 data dari kelas negatif salah diprediksi, sehingga tidak dikenali sebagai kelas negatif. Di sisi lain, 215 data dari kelas negatif diprediksi dengan tepat sebagai kelas negatif, namun terdapat 6 data dari kelas positif yang tidak berhasil dikenali dengan benar sebagai kelas positif.

Tabel 4. 10 Tabel Hasil Klasifikasi Rasio 70:30

		<i>Predicted</i>	
		<i>Negative</i>	<i>Positive</i>
<i>Actual</i>	<i>Positive</i>	300	60
	<i>Negative</i>	15	310

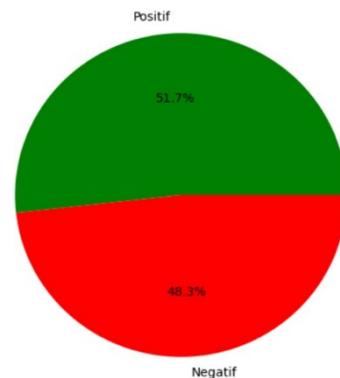
Dalam Tabel 4.10, terdapat 302 data yang berasal dari kelas positif dan berhasil diprediksi dengan benar sebagai kelas positif. Sementara itu, sebanyak 60 data dari kelas negatif salah diprediksi, sehingga tidak dikenali sebagai kelas negatif. Di sisi lain, 310 data dari kelas negatif diprediksi dengan tepat sebagai kelas negatif, namun terdapat 15 data dari kelas positif yang tidak berhasil dikenali dengan benar sebagai kelas positif.

Berdasarkan proses klasifikasi yang telah dilakukan, algoritma *Naïve Bayes* menggunakan konsep probabilitas antar kalimat untuk setiap kelas guna menghasilkan prediksi pada data yang dimasukkan. Berikut gambar diagram yang merupakan hasil analisis *Naïve Bayes* digunakan sebagai data uji.



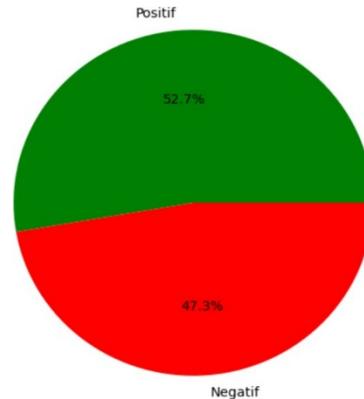
Gambar 4. 3 Distrubusi Data Uji Rasio 90:10

Gambar 4.3 menunjukkan hasil analisis menggunakan algoritma *Naïve Bayes*, di mana sepertiga dari keseluruhan data dialokasikan sebagai data uji, sementara sisanya, yaitu dua pertiga, digunakan sebagai data latih. Seluruh tweet yang diklasifikasikan sebagai positif dan negatif, dengan total sebanyak 229 data, dimanfaatkan dalam proses pengujian.



Gambar 4. 4 Distrubusi Data Uji Rasio 80:20

Gambar 4.4 hasil analisis menggunakan *Naïve Bayes*, sepertiga dari keseluruhan data dialokasikan sebagai data uji sisanya dua pertiga, digunakan sebagai data latih. Seluruh tweet yang diklasifikasikan sebagai positif dan negatif, dengan total sebanyak 458 data, dimanfaatkan dalam proses pengujian.



Gambar 4. 5 Distrubusi Data Uji Rasio 70:30

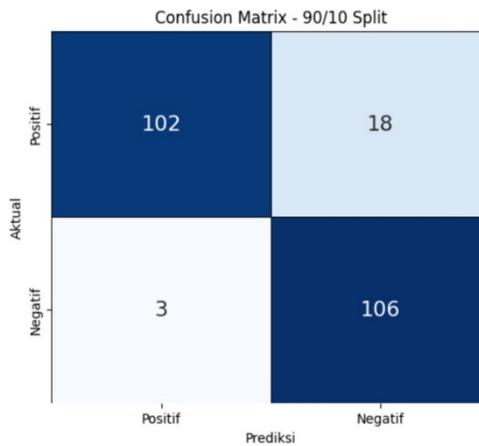
Gambar 4.5 menunjukkan hasil analisis menggunakan algoritma *Naïve Bayes*, di mana sepertiga dari keseluruhan data dialokasikan sebagai data uji, sementara sisanya, yaitu dua pertiga, digunakan sebagai data latih. Seluruh tweet yang diklasifikasikan sebagai positif dan negatif, dengan total sebanyak 687 data, dimanfaatkan dalam proses pengujian.

4.1.6 Evaluasi Model

Pada tahap ini, dilakukan evaluasi terhadap model untuk menilai sejauh mana proses klasifikasi yang telah dilakukan oleh sistem dapat dipercaya. Evaluasi ini mencakup pengukuran metrik seperti *akurasi*, *presisi*, *recall*, dan *F1-Score*. Untuk memperoleh nilai-nilai tersebut, dilakukan pengujian dengan menggunakan confusion matrix guna menilai hasil prediksi yang dihasilkan oleh model.

4. 1. 6. 1. Evaluasi Model dengan Rasio Pembagian Data 90:10

Dalam evaluasi model *Naïve bayes* rasio 90:10 Gambar 4.6 menyajikan visualisasi dari hasil evaluasi model dengan rasio pembagian data 90:10. Berdasarkan Gambar 4.6, dataset tweet yang berisi sentimen pengguna terhadap aplikasi Wattpad dianalisis menggunakan algoritma Multinomial *Naïve Bayes* dengan ekstraksi fitur TF-IDF pada rasio pembagian data 90:10.



Gambar 4. 6 Hasil Klasifikasi Rasio 90:10

Gambar 4.6 menampilkan hasil analisis klasifikasi dengan rasio pembagian data latih dan uji sebesar 90:10. Pada tahap ini, evaluasi menggunakan confusion matrix sebagai dasar dalam menghitung metrik performa seperti akurasi, presisi, recall, dan F1-score. Berdasarkan hasil klasifikasi dengan rasio pembagian data 90:10, diperoleh nilai-nilai sebagai berikut:

- a. True Positive (TP): 102 tweet positif berhasil diprediksi sebagai sentimen positif.
- b. False Negative (FN): 18 tweet positif salah diprediksi sebagai sentimen negatif.
- c. True Negative (TN): 106 tweet negatif berhasil diprediksi sebagai sentimen negatif.
- d. False Positive (FP): 3 tweet negatif salah diprediksi sebagai sentimen positif.

Selanjutnya hasil evaluasi model yang telah dijelaskan sebelumnya, langkah selanjutnya adalah menghitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-score dan berikut untuk hasil perhitungannya.

Akurasi Data 90:10

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{102 + 106}{102 + 18 + 106 + 3} \times 100\%
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{208}{229} \times 100\% \\
 &= 0,908\%
 \end{aligned}$$

Sentimen Positif

$$Precision = \frac{TP}{TP + FP} = \frac{102}{102 + 3} = 0,97$$

$$Recall = \frac{TP}{TP + FN} = \frac{102}{102 + 18} = 0,85$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0,97 \times 0,85}{0,97 + 0,85} = 0,91$$

Sentimen Negatif

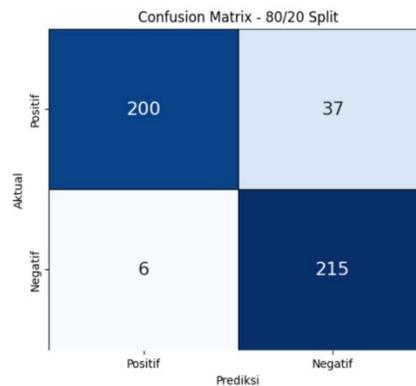
$$Precision = \frac{TP}{TP + FP} = \frac{106}{106 + 18} = 0,85$$

$$Recall = \frac{TP}{TP + FN} = \frac{106}{106 + 3} = 0,97$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0,97 \times 0,85}{0,97 + 0,85} = 0,91$$

4. 1. 6. 2. Evaluasi Model dengan Rasio Pembagian Data 80:20

Dalam evaluasi model *Naïve bayes* rasio 80:20 Gambar 4.6 menyajikan visualisasi dari hasil evaluasi model dengan rasio pembagian data 80:20. Berdasarkan Gambar 4.6, dataset tweet yang berisi sentimen pengguna terhadap aplikasi Wattpad dianalisis menggunakan algoritma *Multinomial Naïve Bayes* dengan ekstraksi fitur TF-IDF pada rasio pembagian data 80:20.



Gambar 4. 7 Hasil Klasifikasi Rasio 80:20

Gambar 4.7 menampilkan hasil analisis klasifikasi dengan rasio pembagian data latih dan uji sebesar 80:20. Pada tahap ini, evaluasi dilakukan menggunakan confusion matrix sebagai dasar dalam menghitung metrik performa seperti akurasi, presisi, recall, dan F1-score. Berdasarkan hasil klasifikasi dengan rasio pembagian data 80:20, diperoleh nilai-nilai sebagai berikut:

- a. True Positive (TP): 200 tweet positif berhasil diprediksi sebagai sentimen positif.
- b. False Negative (FN): 37 tweet positif salah diprediksi sebagai sentimen negatif.
- c. True Negative (TN): 215 tweet negatif berhasil diprediksi sebagai sentimen negatif.
- d. False Positive (FP): 6 tweet negatif salah diprediksi sebagai sentimen positif.

Selanjutnya, berdasarkan hasil evaluasi model yang telah dijelaskan dilakukan perhitungan terhadap metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai performa klasifikasi model secara kuantitatif.

Akurasi Data 80:20

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{200 + 215}{200 + 37 + 215 + 6} \times 100\% \\
 &= \frac{415}{458} \times 100\% = 0,906\%
 \end{aligned}$$

Sentimen Positif

$$Precision = \frac{TP}{TP + FP} = \frac{200}{200 + 6} = 0,97$$

$$Recall = \frac{TP}{TP + FN} = \frac{200}{200 + 37} = 0,84$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0,97 \times 0,84}{0,97 + 0,84} = 0,90$$

Sentimen Negatif

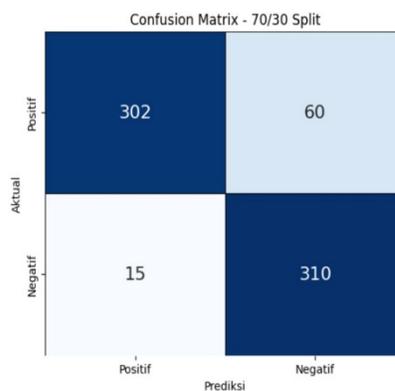
$$Precision = \frac{TP}{TP + FP} = \frac{215}{215 + 18} = 0,85$$

$$Recall = \frac{TP}{TP + FN} = \frac{215}{215 + 6} = 0,97$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0,85 \times 0,97}{0,85 + 0,97} = 0,91$$

4. 1. 6. 3. Evaluasi Model dengan Rasio Pembagian Data 70:30

Dalam evaluasi model *Naïve bayes* rasio 70:30 Gambar 4.6 menyajikan visualisasi dari hasil evaluasi model dengan rasio pembagian data 70:30. Berdasarkan Gambar 4.6, dataset tweet yang berisi sentimen pengguna terhadap aplikasi Wattpad dianalisis menggunakan algoritma Multinomial *Naïve Bayes* dengan ekstraksi fitur TF-IDF pada rasio pembagian data 70:30.



Gambar 4. 8 Hasil Klasifikasi Rasio 70:30

Gambar 4.8 menampilkan hasil analisis klasifikasi dengan rasio pembagian data latih dan uji sebesar 70:30. Pada tahap ini, evaluasi dilakukan menggunakan *confusion matrix* sebagai dasar dalam menghitung metrik performa seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Berdasarkan hasil klasifikasi dengan rasio pembagian data 70:30, diperoleh nilai-nilai sebagai berikut:

- True Positive (TP): 302 tweet positif berhasil diprediksi sebagai sentimen positif.
- False Negative (FN): 60 tweet positif salah diprediksi sebagai sentimen negatif.
- True Negative (TN): 310 tweet negatif berhasil diprediksi sebagai sentimen negatif.
- False Positive (FP): 15 tweet negatif salah diprediksi sebagai sentimen positif.

Selanjutnya, berdasarkan hasil evaluasi model yang telah dijelaskan, dilakukan perhitungan terhadap metrik evaluasi seperti *akurasi*, *presisi*, *recall*, dan *F1-score* untuk menilai performa klasifikasi model secara kuantitatif.

Akurasi Data 70:30

$$\begin{aligned} Accuracy &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\ &= \frac{302 + 310}{302 + 60 + 310 + 15} \times 100\% \\ &= \frac{612}{687} \times 100\% = 0,89\% \end{aligned}$$

Sentimen Positif

$$Precision = \frac{TP}{TP + FP} = \frac{302}{302 + 15} = 0,95$$

$$Recall = \frac{TP}{TP + FN} = \frac{302}{302 + 60} = 0,83$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0,95 \times 0,83}{0,95 + 0,83} = 0,89$$

Sentimen Negatif

$$Precision = \frac{TP}{TP + FP} = \frac{310}{310 + 60} = 0,84$$

$$Recall = \frac{TP}{TP + FN} = \frac{310}{310 + 15} = 0,95$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0,84 \times 0,95}{0,84 + 0,95} = 0,89$$

Selanjutnya pada tabel 4.11 ditampilkan hasil evaluasi model klasifikasi analisis sentimen menggunakan algoritma *Naïve Bayes* dengan ekstraksi fitur TF – IDF menggunakan rasio 90:10, 80:20, 70:30.

Tabel 4. 11 Hasil Perhitungan Confusion Matrix

NO	Pembagian data uji & data Testing	Akurasi	Presisi	Recall	F1-Score
1	90:10	90.83%	85.48%	97.25%	90.99%
2	80:20	90.61%	85.32%	97.29%	90.91%
3	70:30	89.08%	83.78%	95.38%	89.21%

Evaluasi ini mencakup pengukuran metrik standar seperti *akurasi*, *presisi*, *recall*, dan *F1-Score*, yang digunakan untuk memberikan gambaran lebih mendalam mengenai kinerja model dalam mengklasifikasikan data secara tepat. Dalam pengujian yang dilakukan, masing-masing rasio data latih dan data uji memberikan hasil sebagai berikut:

a. Rasio Pembagian data 90:10

Model menghasilkan akurasi sebesar 90,38%, yang berarti sekitar 90 dari 100 data berhasil diklasifikasikan dengan benar. Presisi sebesar 85,48%, menunjukkan bahwa dari seluruh data yang diprediksi sebagai positif, 85% di antaranya benar-benar positif. Recall mencapai 97,25%, menandakan bahwa hampir seluruh data positif berhasil dikenali oleh model. F1-Score sebesar 90,99%, memberikan gambaran bahwa model memiliki keseimbangan yang baik antara presisi dan recall.

b. Rasio Pembagian data 80:20

Model memperoleh akurasi sebesar 90,61%, sedikit lebih tinggi dibandingkan rasio 90:10. Nilai presisi adalah 85,32%, yang berarti kemampuan model mengenali data positif tetap tinggi. Recall sebesar 97,29%, menunjukkan bahwa model masih sangat efektif dalam mendeteksi data positif. F1-Score sebesar 90,91%, menandakan performa model tetap seimbang dan konsisten.

c. Rasio Pembagian data 70:30

Model menghasilkan akurasi sebesar 89,08%, menurun dibandingkan dua rasio sebelumnya. Presisi menurun menjadi 83,78%, menunjukkan adanya peningkatan jumlah data positif yang salah diprediksi. Recall sebesar 95,38%, yang berarti masih banyak data positif yang berhasil dikenali. F1-Score sebesar 89,21%, tetap mencerminkan performa yang baik meskipun sedikit menurun.

Seluruh persentase tersebut dihitung berdasarkan hasil *confusion matrix*, dengan membandingkan antara jumlah kecocokan prediksi nilai *true* dan *false* terhadap label data aktual.

4.2 Pembahasan

Pada tahap awal, dilakukan crawling data menggunakan Tweet Harvest dengan rentang waktu pengambilan dari tanggal 1 April 2024 hingga 1 April 2025. Proses ini menghasilkan data awal sebanyak 3.059 tweet. Setelah dilakukan pembersihan dan pengecekan duplikasi, jumlah data bersih yang diperoleh adalah sebanyak 2.288 tweet. Data yang telah dibersihkan kemudian diberi label oleh ahli bahasa Indonesia, dengan jumlah label positif lebih banyak dibandingkan label negatif, yakni sebanyak 1.200 untuk sentimen positif dan 1.088 untuk sentimen negatif. Selanjutnya, dilakukan pembersihan teks untuk menghapus elemen-elemen yang tidak relevan bagi proses klasifikasi, seperti tanda baca, *URL*, emotikon, *hashtag*, *username/mention*, serta spasi berlebih. Proses ini juga mencakup konversi huruf menjadi huruf kecil untuk menyamakan format teks. Setelah itu, dilakukan penghapusan kata-kata tidak penting

(*stopword removal*) menggunakan *library* Sastrawi. Proses dilanjutkan dengan *stemming*, yang juga menggunakan *library* Sastrawi untuk mengembalikan kata ke bentuk dasarnya.

Kata-kata yang tidak terdapat dalam daftar *stemming* tetap dipertahankan untuk proses klasifikasi. Beberapa contoh kesalahan prediksi pada kategori *False Positive* antara lain terdapat pada *tweet* seperti “posting sini gratis gak crnya gmn sy jg nulis watsapp”, “dah hapus cerita nya unggah di watsapp”, dan “ah beneran kak gulingguling nih ngobrol d watsapp jg gue aktif sini”. Sementara itu, contoh *False Negative* muncul pada *tweet* seperti “iyaaah aku pertama bikin au sini kmrn aja asa ngang ngong biasa watsapp since tp gpp ajar dikit dikit” dan “semua watsapp core yang paling bikin gua ken berak pas baca”.

Kesalahan prediksi ini umumnya disebabkan oleh penggunaan kata-kata tidak baku, kata berlebihan, singkatan, dan kesalahan ketik yang menyulitkan sistem memahami konteks emosi atau sentimen sebenarnya. Kata-kata seperti *sejijay*, *gulingguling*, *gurlll*, serta berbagai bentuk ekspresi yang tidak umum menjadi tantangan tersendiri bagi model klasifikasi. Hal ini menunjukkan bahwa masih terdapat keterbatasan dalam metode yang digunakan serta kemungkinan adanya data yang belum sepenuhnya bersih. Hasil klasifikasi menunjukkan bahwa *tweet* yang dikategorikan sebagai sentimen positif umumnya mengandung kata-kata seperti menarik, asik, bisa, lucu, bagus, hoki, gratis, suka, happy, puas, banyak, baik, makasih, bangga, referensi, legal, untung, bebas, murah, favorit, tersedia, lengkap, mudah, nyaman, dan keren. Sementara itu, *tweet* yang diklasifikasikan sebagai negatif cenderung mengandung kata-kata seperti antre, musuh, lama, error, kosong, gabisa, lupa, nyebelin, masalah, kesal, bete, lambat, payah, capek, susah, rebut, saingan, jelek, bug, males, salah, sedih, habis, hilang, dan gila. Setelahnya, dilakukan proses *tokenization* sebagai tahap akhir *preprocessing*, yang hasilnya akan digunakan untuk pembobotan kata. Dari proses pembobotan tersebut, diperoleh 2.288 data yang telah diberi nilai bobot oleh sistem, dengan nilai bobot berkisar antara 0,0 hingga 0,99142.

Data yang telah berbobot ini kemudian digunakan untuk proses klasifikasi dengan algoritma *Naïve Bayes Classifier*.

Proses klasifikasi dilakukan dengan membagi data ke dalam rasio data latih dan data uji, yaitu 90:10, 80:20, dan 70:30. Rasio 90:10 menghasilkan akurasi sebesar 90,83%, rasio 80:20 menghasilkan akurasi sebesar 90,61%, dan rasio 70:30 memberikan akurasi sebesar 89,08%. Akurasi ini berguna untuk mengukur seberapa tepat model dalam mengklasifikasikan seluruh data secara keseluruhan, yaitu persentase prediksi yang benar dari total data uji.

Selain itu, model juga dievaluasi menggunakan metrik presisi, recall, dan F1-Score untuk memberikan gambaran yang lebih menyeluruh terkait performa model. Presisi pada rasio 90:10 adalah sebesar 85,48%, pada rasio 80:20 sebesar 85,32%, dan pada rasio 70:30 sebesar 83,78%. Presisi ini mengukur seberapa akurat model dalam melakukan prediksi, yaitu dari seluruh data yang diprediksi, berapa persen yang benar-benar sesuai. Untuk metrik recall, pada rasio 90:10 tercatat sebesar 97,25%, pada rasio 80:20 sebesar 97,29%, dan pada rasio 70:30 sebesar 95,38%. Recall menunjukkan kemampuan model dalam menangkap data yang benar dari seluruh data aktual, atau dengan kata lain, seberapa banyak data aktual yang berhasil dikenali oleh model. Sedangkan F1-Score, yang merupakan kombinasi harmonis antara presisi dan recall, pada rasio 90:10 menunjukkan nilai sebesar 90,99%, pada rasio 80:20 menunjukkan nilai sebesar 90,91%, dan pada rasio 70:30 menunjukkan nilai sebesar 89,21%. F1-Score berguna untuk memberikan penilaian yang lebih seimbang terhadap performa model, terutama ketika terdapat ketidakseimbangan antara presisi dan recall.

Berdasarkan perbandingan dari hasil klasifikasi tersebut, diperoleh bahwa akurasi, presisi, recall, F1-Score pada rasio 90:10 memberikan performa terbaik secara keseluruhan dibandingkan dua rasio lainnya. Secara keseluruhan untuk hasil analisis sentimen menunjukkan bahwa opini yang termasuk dalam kategori positif lebih dominan dibandingkan opini negatif. Hal ini menandakan bahwa secara umum, pengguna aplikasi Wattpad memberikan respons yang cenderung positif. Temuan ini dapat menjadi dasar bahwa Wattpad berhasil memberikan pengalaman yang

memuaskan bagi mayoritas penggunaannya. Oleh karena itu, pihak pengembang disarankan untuk mempertahankan keunggulan yang sudah diapresiasi, serta terus memperbaiki kelemahan berdasarkan opini negatif yang muncul (Normawati & Prayogi, 2021) gambaran yang seimbang mengenai persepsi pengguna, serta dapat dijadikan referensi strategis dalam pengembangan aplikasi di masa mendatang. Selain itu, dari sudut pandang nilai-nilai Islam, analisis sentimen juga dapat diarahkan untuk tujuan yang lebih etis dan bermanfaat bagi kebaikan bersama. Adapun dalam perspektif Islam, tujuan analisis sentimen memiliki relevansi dengan prinsip-prinsip kehati-hatian dalam menyikapi informasi, sebagaimana termaktub dalam Al- Qur'an Surat Al-Hujurat ayat 6:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْحِرُوا عَلَىٰ مَا فَعَلْتُم نَادِمِينَ

“Wahai orang-orang yang beriman! Jika seseorang fasik datang kepadamu membawa suatu berita, maka telitilah kebenarannya, agar kamu tidak mencelakakan suatu kaum karena kebodohan (kecerobohan), yang akhirnya kamu menyesali perbuatanmu itu.” (Q.S. Al-Hujurat: 6)

Merujuk pada tafsir Kementerian Agama Republik Indonesia (2024), ayat tersebut menegaskan urgensi melakukan verifikasi (tabayyun) terhadap setiap informasi, khususnya apabila bersumber dari pihak yang kredibilitasnya diragukan. Hal ini bertujuan untuk mencegah kekeliruan dalam pengambilan keputusan yang dapat menimbulkan dampak merugikan bagi pihak lain. Dalam konteks penelitian, terutama dalam analisis sentimen, prinsip tabayyun merefleksikan pentingnya proses validasi dan verifikasi data secara objektif, sistematis, dan berbasis evidensi. Praktik ini diperlukan guna memastikan bahwa hasil penelitian tidak hanya memenuhi standar metodologis dan etika akademik, tetapi juga selaras dengan nilai-nilai keislaman, yaitu menyampaikan kebenaran dan menghindari penyebaran informasi yang menyesatkan. Dengan demikian, ayat ini memberikan fondasi etis dan spiritual dalam pelaksanaan kegiatan ilmiah yang mengedepankan kejujuran, transparansi, serta tanggung jawab ilmiah dan moral.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pembahasan, opini yang diperoleh lebih dominan kearah positif dibandingkan opini negative. Rasio 90:10 dengan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 90,83%, presisi sebesar 97%, recall sebesar 85%, dan F1-Score sebesar 91%. Hasil rasio 90:10 merupakan hasil terbaik dibandingkan dengan rasio lainnya. Akurasi yang diperoleh menandakan bahwa model sudah tepat dalam melakukan klasifikasi data secara keseluruhan. Nilai presisi yang diperoleh menandakan keakuratan model dalam melakukan prediksi positif yang benar-benar positif. Nilai recall yang diperoleh menandakan bahwa model yang dihasilkan dapat menangkap data positif dari seluruh data aktual yang positif. Nilai F1-Score menandakan bahwa nilai presisi dan recall mempunyai besar yang seimbang. Hasil tersebut menunjukkan bahwa algoritma *Naïve Bayes* dengan rasio 90:10 dapat digunakan secara efektif untuk menganalisis sentimen pengguna terhadap aplikasi Wattpad di media sosial X.

5.2 Saran

Berdasarkan hasil penelitian, berikut beberapa saran yang dapat menjadi pertimbangan untuk pengembangan dan penelitian selanjutnya:

1. Untuk penelitian berikutnya, sebaiknya pengambilan data dilakukan dalam waktu yang lebih lama supaya jumlah data yang terkumpul lebih banyak, dengan begitu hasil klasifikasi bisa lebih akurat.
2. Menggunakan data dari platform media sosial lain seperti Instagram atau Play Store agar analisis mencakup lebih banyak sudut pandang pengguna.

DAFTAR PUSTAKA

- Alita, D., & Isnain, A. R. (2020). Pendeteksian sarkasme pada proses analisis sentimen menggunakan random forest classifier. *Jurnal Komputasi*, 8(2), 50–58. <https://doi.org/10.23960/komputasi.v8i2.2615>
- Andre Septian, dkk. (2019). *Journal Of Intelligent Systems and Computation* 43. <https://t.co/9wloawpfd5>
- Arif Budi Mawardi. (2018). Komodifikasi Sastra Cyber Wattpad Pada Penerbit Indie. *Sabda Volume 13, Nomor 1, Juni 2018*.
- Astari, N. M. A. J., Dewa Gede Hendra Divayana, & Gede Indrawan. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27–29. <https://doi.org/10.30864/jsi.v15i1.332>
- Alfan Rosid, M., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving text preprocessing for student complaint document classification using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1), 012017. <https://doi.org/10.1088/1757-899X/874/1/012017>
- Chuluq, K., & Nudin, S. R. 2024. *Klasifikasi Cyberbullying Pada Media Sosial Dengan Menggunakan Metode Recurrent Neural Network Dan Long Short Term Memory*. *Journal of Informatics and Computer Science (JINACS)*, 6(2), 501–509. <https://doi.org/10.26740/jinacs.v6n02.p501-509>
- Fachri, S., & Ramdan, J. (2022). Pemodelan machine learning: Analisis sentimen masyarakat terhadap kebijakan PPKM menggunakan data Twitter. *Jurnal Sistem Cerdas*, 5(1), 12–20. <https://doi.org/10.37396/jsc.v5i1.191>
- Fathiyatul, Dkk. (2020). *Eksistensi Sastra Cyber: Webtoon Dan Wattpad Menjadi Sastra Populer Dan Lahan Publikasi Bagi Pengarang*.
- Fibriyanti Arminda, Dkk. (2023). Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 3).
- Firdaus, A., & Dkk. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). In *Jurnal Jupiter* (Vol. 13, Issue 1).

- Habib Syarkowi Harahap. (2024). Implementasi Phyton Dalam Matematika. *Mathematical And Data Analytics*, 1(1), 31–37. <https://doi.org/10.47709/mda.v1i1.3887>
- Kementerian Agama Republik Indonesia. (2019). *Al-Qur'an dan terjemahannya*. Jakarta: Lajnah Pentashihan Mushaf Al-Qur'an. <https://quran.kemenag.go.id/>
- Kementerian Agama Republik Indonesia. (2025). *Qur'an Kemenag*. <https://quran.kemenag.go.id/quran/per-ayat/surah/17?from=1&to=111>
- Khoirunnisa, S. (2022). *Pengaruh Perpustakaan Digital Terhadap Peningkatan Minat Baca Pemustaka Di Era Digitalisasi*.
- Lestari, S., & Saepudin, S. (2021). *Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naïve Bayes*. <https://Vaksin.Kemkes.Go.Id/>
- Mahfud, dkk. (2020). Sentiment Analysis Of Perpustakaan Nasional Republik Indonesia Through Social Media Twitter. *Matics*, 12(1), 90. <https://doi.org/10.18860/mat.v12i1.8973>
- Maidatul Lifiani, I. K. (2019). Pemanfaatan Wattpad Sebagai Aplikasi Self-Publishing. *Jurnal Ilmu Perpustakaan*, 8(2).
- Mas Pintoko, B., & Muslim, K. (2018). *Analisis Sentimen Jasa Transportasi Online Pada Twitter Menggunakan Metode Naïve Bayes Classifier*.
- Munthe, C. J. E., Hasibuan, N. A., & Hutabarat, H. (2022). *Penerapan Algoritma Text Mining dan TF-RF dalam Menentukan Promo Produk pada Marketplace*. *RESOLUSI: Rekayasa Teknik Informatika dan Informasi*, 2(3), 110–115. <https://doi.org/10.30865/resolusi.v2i3.309>
- Matthew E, D. (2023). *Perancangan Dashboard Sentimen Pariwisata Menggunakan Metode Complement Naïve Bayes Pada Kabupaten Rembang*.
- Nadya M. (2023). *Peranan Aplikasi Wattpad Dalam Meningkatkan Minat Baca Dan Menulis Pada Mahasiswa Fakultas Ushuluddin Adab Dan Dakwah Iain Parepare*.
- Normawati, D., & Prayogi, S. A. (2021). *Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter*. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711. <https://jurnal.utamasains.ac.id/ejurnal/index.php/jsakti>

- Nur Adhan, S., Ngurah Adhi Wibawa, G., Christien Arisona, D., Yahya, I., Studi Statistika, P., & Matematika Dan Ilmu Pengetahuan Alam, F. (2024). *Analisis Sentimen Ulasan Aplikasi Wattpad Di Google Play Store Dengan Metode Random Forest* (Vol. 2, Issue 1).
- Nurhafida, S. I., & Sembiring, F. (2022). Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM). In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 6, Issue 1).
- Nurkalyisah, A. N., Triayudi, A., & Sholihati, I. D. (2022). Analisis Sentimen Pada Twitter Berbahasa Indonesia Terhadap Penurunan Performa Layanan Indihome Dan Telkomsel. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 10(4), 387. <https://doi.org/10.26418/justin.v10i4.50858>
- Nurtikasari, Y., Syariful Alam, & Teguh Iman Hermanto. (2022). Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes. *INSOLOGI: Jurnal Sains Dan Teknologi*, 1(4), 411–423. <https://doi.org/10.55123/insologi.v1i4.770>
- Pahlevi, R., & Ziveria, M. (2022). *Perancangan Deteksi Kemiripan Pada Abstrak Artikel Ilmiah Informatika* (Vol. 8, Issue 2).
- Putra Prasetyo, A., Kusuma, C. S., & Rizki, D. (2024). Analisis Sentimen Twitter Terhadap Isu Penundaan Pemilu 2024 Menggunakan Metode Long Short-Term Memory (LSTM). *REPOSITOR*, 6(3), 293–304.
- Quran Kemenag. (2025). *Quran Kemenag*. <https://Quran.Kemenag.Go.Id/Quran/Per-Ayat/Surah/17?From=1&To=111>
- Rahayu, M. P., & Farlina, Y. (2021). *Penerapan Metode Naive Bayes Dalam Prediksi Penyebab Kecelakaan Kerja CV. Dek Utama*. <http://jurnal.bsi.ac.id/index.php/larik>
- Rezki, M., Kholifah, D. N., Faisal, M., Suryadithia, R., Bina, U., & Informatika, S. (2020). Analisis Review Pengguna Google Meet Dan Zoom Cloud Meeting Menggunakan Algoritma Naive Bayes. In *Jurnal* (Vol. 2, Issue 2). <http://ejournal.bsi.ac.id/ejurnal/index.php/infortech264>
- Ridwansyah. (2022). *KLIK: Kajian Ilmiah Informatika Dan Komputer Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naive*

- Bayes Classifier. *Media Online*, 2(5), 178–185. <https://djournals.com/klik>
- Rinida, Gistinia, dkk. (2023). Penerapan Svm Dalam Analisis Sentimen Pada Edlink Menggunakan Pengujian Confusion Matrix. *Jati* (Vol. 7 No.5)
- Rizqi Robbi, dkk. (2022). *Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K- Fold Cross Validation*. 11(1),130–139. <https://Ejournal3.Undip.Ac.Id/Index.Php/Gaussian/>
- Santoso Agil, dkk (2022). Pengaruh Stopword Removal Dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma Lstm. *J-Sakti* (Vol. 6, No.1). Issn: 2548-9771/Eissn: 2549-7200. <https://Tunasbangsa.Ac.Id/Ejurnal/Index.Php/Jsakti>
- Setiawan, H., Utami, E., & Sudarmawan, S. (2021). Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, 5(1), 43–51. <https://doi.org/10.31603/komtika.v5i1.5189>
- Shafina, S., & Andita, P. (2022). *Manfaat Perpustakaan Digital Dalam Meningkatkan Minat Baca Generasi Milenial Di Era Globalisasi*.
- Shoji Al Falaq, J., Suprayogi, S., Nofita Susanto, F., Ul Husna, A., & Dosen Muslim Indonesia -Sulawesi Selatan, P. (2021). Exploring the potentials of Wattpad for literature class. *Indonesian Journal of Learning Studies (IJLS)*, 1(2).

LAMPIRAN

Lampiran 1 Tabel Hasil Perhitungan TF-IDF

Kata	Jumlah Kata									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Akhir	0,111111 111									
Wattpa d	0,017211 329		0,019362 745	0,022128 851			0,025816 993	0,019362 745	0,022128 851	0,017211 329
Sudah	0,077663 334									0,077663 334
Lancar	0,111111 111									
Kembali	0,111111 111									
Jadi	0,111111 111									
Bisa	0,077663 334	0,099852 858								

Koment ar	0,111111 111									
Lagi	0,058097 638		0,130719 686							
Lucu		0,142857 143								
Juga		0,142857 143								
Fitur		0,075096 964	0,065359 843				0,065359 843	0,065359 843		
Buat		0,142857 143								
Kalimat		0,142857 143								
Sendiri		0,142857 143								
Untuk			0,125							
Promo			0,125							

Cerita			0,087371 251						0,099852 858	
Hilang			0,125							
Dari			0,087371 251	0,099852 858						
Error				0,074696 964	0,104575 749	0,261439 373				
Tadi				0,142857 143						
Notifikasi				0,142857 143						
Tidak				0,074696 964						0,155326 668
Muncul				0,199705 716						
Hampir					0,2					
Hari					0,2					
Habis					0,2					

Baru					0,2					
Terus						0,5				
Senang							0,166666 667			
Sekarang							0,166666 667			
Ada							0,116495 001			0,077663 334
Paling								0,125		
Asik								0,125		
Baca								0,087371 251	0,099852 858	
Daring								0,125		
Masih								0,125		
Sering									0,142857 143	
Tapi									0,099852 858	0,077663 334

Lampiran 2 Daftar 50 sampel dataset berlabel.

No	Tweets	Label
1.	di wappad tuh kalau niat nyelem buat cari story pasti bakal nemu story yg WAH banget! story yg gini banyak yg gak masuk top tag jadi kudu sabar scroll terus	Positif
2.	@tanyakanrl Anjing bener lagi baru aja lewat di wappad gue langsung pasang muka jijik anjir	Negatif
3.	@worksfess Karangan ala wappad kah? Kerja baru 6 bulan dengan gaji ala MT udah berkesimpulan gak ada karir. Hah?	Negatif
4.	@peningkepalaAku HAHHAHAHAHA IYA BODO tapi skrg aku lagi suka baca au kat twt dari wappad sebab storyline mantap	Positif
5.	MAKSUD GW INI JGN GA JELAS ANJIR W JG MALES JADINYA KEK 'APAAN SI NI ORG POSESIF TAPI KOK KY DIBUAT2 KEK WATTPAD'	Negatif
6.	brick! aku abis nemu wappad gyuicky family lucu bangetttt wajib baca tp masih on going abis ini aku drop linknya di bawah https://t.co/X8j7Qf3PYb	Positif
7.	J+A29:A33adi ingettt waktu SMA aku suka nulis di wappad. Dan mostly ceritanya tuh tentang crush aku waktu itu . Apalagi yang Cerita SMA itu bner2 kaya diary aku ttg dia. Anehnya knapa banyak yg baca ya https://t.co/vEySG6JaQE	Negatif
8.	Apk untuk menulis Catatan keep dan wappad Jumlah kata 1600+ Mohon maaf bilang banyak kesalahan dalam penulisan atau typo	Positif
9.	ga sukakkkkkkkkkm jauhhh beda sama wappad nyaaaa jujurrrrr. issssss	Negatif

No	Tweets	Label
10.	mencoba download wamppad lagi buka library dan tbtb semua serasa nostalgia entah kenapa cover dan judul judul ini membawa perasaan nostalgia menimbulkan aroma aroma nostalgia yg gabisa di deskripsi kan https://t.co/uOmLxs6FXe	Positif
11.	Baru nyadar ternyata gue se-semua dicobain itu bjir woilah. Mulainya dari jaman sd ke smp dah. Drakor dracin hollywood bollywood animanga kpop wamppad au ao3 raikan manhwa/hua donghua danmei gaming etc yg tdk bs disebut. gimana gue gak punya dunia sendiri coba	Positif
12.	Karya - Halo semua beberapa hari yang lalu aku post cerita pertamaku di Wamppad. Judulnya Myosotis dengan genre fantasi tragedi & angst. MC nya bekerja sebagai ghost finder! Akunku baru dibuat jadi masih sepi kalau berkenan tolong mampir dan baca ya terima kasih ^^ https://t.co/VYriza5o28	Positif
13.	@ourstickyart Yg happy-happy aja aku kurang suka angst soalnya terus ntar upload di wamppad entar gw baca awokawokawok Sebenarnya tadi bingung sii mau minjem novel yg mana di perpunas karena banyak banget guyss novel yg bagus-bagus disanaa ada novel best seller dan ada yg novel dari wamppad. And finally aku memutuskan meminjam dua novel ini https://t.co/aZBVhy5rpR	Positif
14.	Gaiisss aku bakal jarang on di X karena lagi ngetik cerita wamppad hehe Mau nyari hal baru	Positif
15.	Gue pikir nama saviero cuman ada di CA atau wamppad di rl ada juga anjir adik tingkat gue. Keren sih	Positif
16.	@quanruige Dulu ada yg ngetl indo cakep bgt tapi dihapus. Sekarang sisa yg di wamppad yg alakadarnya wkwkwk	Negatif

No	Tweets	Label
17.	Ga ada yang maksa lu baca yaoi jadi baca alquran aja ga usah baca wattpad. Goblok kok dipelihara	Negatif
18.	Udah serius serius baca nyimak ss an kek kasus gitu ternyata cuma fake ss cerita yg berujung ala wattpad	Negatif
19.	guys doakan acc wattpad gue balik plis gue gak tahan seclueless ini sama cerita sendiri	Positif
20.	Kentutttt aku baca Wattpad kagehina ternyata discontinue bJir kenapa semua fanfic kagehina baguss dan hidden gem yang aku temuin semuanya discontinue @Giriselle @HeartHallways GIRISELLE... STOP MEMBUAT SKENARIO SENDIRI KAMU BUKAN AUTHOR WATTPAD	Negatif
21.	Kak alice buat what if rumi sakit dungs oke nanti aku mau buat up di wattpad ya mungkin double update x iya wp juga iya https://t.co/s83pDpEVtX	Positif
22.	Akhirnya setelah bertahun-tahun menyelam di fanfiction wattpad blog hingga Ao3 bisa nemu fanfict Jercy yang bagus banget beneran kadar bagusny pakai banget-banget-banget-banget seperti baca universe berbeda (namanya juga AU)	Positif
23.	/wp/ Aku mau tanya judul cerita wattpad ini gaiss. Mau reread tapi kok ya udah lupa banget judul sama authornya siapa https://t.co/tKynWMwKpd	Positif
24.	Tik! Gais sender mau balik ke dunia kepenulisan lg setelah lama menghilang haha... Dengan genre romance fantasy menurut kalian lebih baik post di wattpad gwp ato di platform mana ya? Makasih	Positif

No	Tweets	Label
25.	Pernah sekali baca buku wattpad waktu smp terus asik dan lebih banyak narasinya ketimbang dialognya dan bener2 bikin mesem2 sendiri	Positif
26.	@hohoheehhee Kek wattpad. Gua kira sukses di sini adalah ambil alih kekuasaan ayahnya di umur 12. Lah usaha sendiri tah? Makin engga masuk akal!!! Kalau ortunya ngga ada pemerintahnya kemana anjirrr bocil umur 12 tawuran didiemin gitu aja	Negatif
27.	ini anak yg nyalah gunain kipk kjp dan sejenisnya kalo udah ketahuan kenapa suka banget dah bikin cerita wattpad?	Negatif
28.	Rekomendasi #wattpad hari ini Top atas perpus aku yg lagi aku baca sebenarnya masih banyak lagi cuma kecampur:(https://t.co/rpSO7vrlqj	Positif
29.	Gw sedih bangett kirain wattpad udah gw apus taunya gw pindah letak anjir fak	Positif
30.	@dsuperboy Gak tau berapa banyak novel manhwa wattpad terjemahan china. Udah suka dari kecil jadi bingung apa yang pertama	Positif
31.	Sakit hati banget anj gue lagi seru-serunya baca wattpad lagi di puncak konflik tiba tiba PINDAH KE KARYAKARSA Masalahnya ini masih MEI	Negatif
32.	Duh aku dari awal suka baca ² gitu lewat online (wattpad & socmed au di Twitter) terus ada yang dijadiin novel ku beli novel nya judulnya Arkananta hwikss	Positif
33.	kak alice terimakasih sudah membuat au dan wattpad dengan cerita yg bagus. i know situasi skrg sedang memanas ttg boycot g aku nggak bakal stop kok krn temen2 readers fairytalice udh kayak temen bagikuu. untuk cerita2ku selanjutnya aku rencana ma https://t.co/SvB68juX8f	Positif

No	Tweets	Label
34.	AHH ANJG KENAPA WATTPAD YANG SERUU TRUS CHAPTER SELANJUTNYA MALAH PINDAH KE KARYAKARSA SIH?!!!! ADABAHJA	Negatif
35.	udah tiga tahun lebih kalau mulai dari fanfiction web (?) sama wamppad udah ada 10 taun kali ya wkwkwk dari suju exo sampai ke nct	Positif
36.	Buka tiktok sedih buka Twitter makin sedih memang sudah saatnya beralih ke wamppad lagi	Positif
37.	@kegblgnunfaedh halo semua pasti pada bosan baca wp anak motor? izin promosi ya kak ini aku bikin cerita buat kaka kaka semua judul wamppad :Sembilan Peti author:Wouzhyy jangan lupa mampir yah terimakasih Maaf kalau ganggu waktu kalian	Positif
38.	@moviemenfes pas dulu wamppad dear natahan rame banget pas jadi novel temen ku pada beli walaupun nambah babnya dikit banget sampe akhirnya jadi film semua temenku pada excited nungguin filmnya walaupun pas itu gatau jefri nichol siapa	Positif
39.	@kdrama_menfess jujur dari posternya kaya poster wamppad 6/10 pemeran ceweknya kurang ekspresinya	Negatif
40.	@Rinnayunay bikin wamppad sendiri sono lu anjirr	Negatif
41.	pgn ngejual novel2 jaman smp-sma gw deh soalnya gue klo liat lg GELI soalnya bnrn novel2 wamppad tahun2 2017-2019	Negatif
42.	Nyoba lagi nulis di wamppad. Tes ombak aja	Positif
43.	@sastraahmedzeyn @zanatul_91 @abcdefghijkl Karya sastra juga ada banyak jenis. Menurutku untuk belajar basic penulisan dan tata bahasa novel-novel ringan udah cukup. Ringan di sini bukan seperti novel teenlit adaptasi au/wamppad ya. Banyak kok novel remaja yang ringan dan oke. Levelnya bisa ditambah tiap tahun.	Positif

No	Tweets	Label
44.	@wattpadmenfess_ Sama nder! Dear Nathan the best sih sejauh ini bahkan ke film-film yang dari wattpad pun tetep Dear Nathan yang paling bagus	Positif
45.	Pemerintahan Indonesia nih cerita wattpad atau gimana yak? Ada aja gak masuk akal nya	Negatif
46.	Pagi ini buka wattpad kan. Trus inget dulu pernah nulis cerita jaman kuliah keknya. Pas baca ngakak banget anjir wkwkwkwkwk. Cerita kek gitu kok bisa ada yang baca ya wkwkwk	Positif
47.	Terima kasih banyak untuk yang udah berkontribusi membaca sempet takut untuk up cerita ini di sosial media selain wattpad itu sendiri tapi berkat seseorang yg blg bahwa karya itu sama seperti anak jangan di manja kan tapi harus di bangga kan juga https://t.co/69exRC3ndi	Positif
48.	@delasyahmareads Jadi bukan soal mendingan baca muatan seksualitas dalam buku sastra dibandingkan baca dari Wattpad. Karena ini program sekolah maka pemerintah juga wajib menyiapkan perangkat dari pihak guru dan gak membiarkan siswa menyalahgunakan buku2 sastra spt Wattpad justru.	Negatif
49.	Udah yaaa di wattpad di x besok ya dah nguantuk gess. Selamat bertemu Salisa dan ronal kembali https://t.co/omq8yTYgRt	Positif
50.	@neuko Itu anjir baca cerita kek wattpad tp membayar tiap chapter nya	Negatif

Lampiran 3 Sampel Hasil Pre-Processing

No	Text	Label	Cleansing	Filtering	Tokenizing	Steaming
1.	di watsapp tuh kalau niat nyelem buat cari story pasti bakal nemu story yg WAH banget! story yg gini banyak yg gak masuk top tag jadi kudu sabar scroll terus	Positif	di watsapp tuh kalau niat nyelem buat cari story pasti bakal nemu story yg wah banget story yg gini banyak yg gak masuk top tag jadi kudu sabar scroll terus	watsapp niat nyelem cari story bakal nemu story wah banget story gini masuk top tag kudu sabar scroll terus	['watsapp', 'niat', 'nyelem', 'cari', 'story', 'bakal', 'nemu', 'story', 'wah', 'banget', 'story', 'gini', 'masuk', 'top', 'tag', 'kudu', 'sabar', 'scroll', 'terus']	watsapp niat nyelem cari story bakal nemu story wah bang story gini masuk top tag kudu sabar scroll terus
2.	@tanyakanrl Anjing bener lagi baru aja lewat di watsapp gue langsung pasang muka jijik anjir	Negatif	anjing bener lagi baru aja lewat di watsapp gue langsung pasang muka jijik anjir	anjing bener baru lewat watsapp gue langsung pasang muka jijik anjir	['anjing', 'bener', 'baru', 'lewat', 'watsapp', 'gue', 'langsung', 'pasang', 'muka', 'jijik', 'anjir']	anjing benar baru lewat watsapp gue langsung pasang muka jijik anjir
3.	@worksfess Karangan ala watsapp kah? Kerja baru 6 bulan dengan gaji ala MT udah berkesimpulan gak ada karir. Hah?	Negatif	karangan ala watsapp kah kerja baru bulan dengan gaji ala mt udah berkesimpulan gak ada karir hah	karangan ala watsapp kerja baru bulan gaji ala mt berkesimpulan karir hah	['karangan', 'ala', 'watsapp', 'kerja', 'baru', 'bulan', 'gaji', 'ala', 'mt', 'berkesimpulan', 'karir', 'hah']	karang ala watsapp kerja baru bulan gaji ala mt simpul karir hah
4.	@peningkepalaAku HAHHAHAHA IYA BODO tapi skrg aku lagi suka baca au kat twt dari watsapp sebab storyline mantap	Positif	hahahahaha iya bodo tapi skrg aku lagi suka baca au kat twt dari watsapp sebab storyline mantap	hahahahaha iya bodo skrg suka baca au kat twt watsapp sebab storyline mantap	['hahahahaha', 'iya', 'bodo', 'skrg', 'suka', 'baca', 'au', 'kat', 'twt', 'watsapp', 'sebab', 'storyline', 'mantap']	hahaha iya bodo skrg suka baca au kat twt watsapp sebab storyline mantap

No	Text	Label	Cleansing	Filtering	Tokenizing	Steaming
	dari wamppad sebab storyline mantap					
5.	MAKSUD GW INI JGN GA JELAS ANJIR W JG MALES JADINYA KEK 'APAAN SI NI ORG POSESIF TAPI KOK KY DIBUAT2 KEK WATTPAD'	Negatif	maksud gw ini jgn ga jelas anjir w jg males jadinya kek apaan si ni org posesif tapi kok ky dibuat kek wamppad	maksud jgn jelas anjir males jadinya kek apaan org posesif tapi kok dibuat kek wamppad	['maksud', 'jgn', 'jelas', 'anjir', 'males', 'jadinya', 'kek', 'apaan', 'org', 'posesif', 'tapi', 'kok', 'dibuat', 'kek', 'wamppad']	maksud jgn jelas anjir malas jadi kek apaan orang posesif tapi kok buat kek wamppad
6.	brick! aku abis nemu wamppad gyuicky family lucu bangetttt wajib baca tp masih on going abis ini aku drop linknya di bawah	Positif	brick aku abis nemu wamppad gyuicky family lucu bangetttt wajib baca tp masih on going abis ini aku drop linknya di bawah	brick abis nemu wamppad gyuicky family lucu bangetttt wajib baca masih going abis drop linknya bawah	['brick', 'abis', 'nemu', 'wamppad', 'gyuicky', 'family', 'lucu', 'bangett', 'wajib', 'baca', 'masih', 'going', 'abis', 'drop', 'linknya', 'bawah']	brick abis nemu wamppad gyuicky family lucu banget wajib baca masih going abis drop link bawah
7.	jadi ingettt waktu SMA aku suka nulis di wamppad. Dan mostly ceritanya tuh tentang crush aku waktu itu . Apalagi yang Cerita SMA itu bner2 kaya diary aku ttg dia.	Positif	jadi ingettt waktu sma aku suka nulis di wamppad dan mostly ceritanya tuh tentang crush aku waktu itu apalagi yang cerita sma itu bner kaya diary aku ttg dia anehnya knapa banyak yg baca ya	jadi ingettt waktu sma suka nulis wamppad mostly ceritanya tentang crush waktu apalagi cerita sma bner kaya diary ttg anehnya knapa banyak baca	['jadi', 'ingett', 'waktu', 'sma', 'suka', 'nulis', 'wamppad', 'mostly', 'ceritanya', 'tentang', 'crush', 'waktu', 'apalagi', 'cerita', 'sma', 'bner', 'kaya', 'diary', 'ttg', 'anehnya', 'knapa', 'banyak', 'baca']	jadi inget waktu sma suka tulis wamppad mostly cerita tentang crush waktu lagi cerita sma bener kaya diary tentang aneh knapa banyak baca

No	Text	Label	Cleansing	Filtering	Tokenizing	Steaming
	Anehnya knapa banyak yg baca ya					
8.	Apk untuk menulis Catatan keep dan watsapp Jumlah kata 1600+ Mohon maaf bilang banyak kesalahan dalam penulisan atau typo	Positif	apk untuk menulis catatan keep dan watsapp jumlah kata mohon maaf bilang banyak kesalahan dalam penulisan atau typo	apk menulis catatan keep watsapp jumlah kata mohon maaf bilang banyak kesalahan penulisan typo	['apk', 'menulis', 'catatan', 'keep', 'watsapp', 'jumlah', 'kata', 'mohon', 'maaf', 'bilang', 'banyak', 'kesalahan', 'penulisan', 'typo']	apk tulis catat keep watsapp jumlah kata mohon maaf bilang banyak salah tulis typo
9.	ga sukakkkkkkkkkm jauhjh beda sama watsapp nyaaaa jujurrrrr. issssss	Negatif	ga sukakkkkkkkkkm jauhjh beda sama watsapp nyaaaa jujurrrrr issssss	Ga sukakkkkkkkkkm jauhjh beda watsapp nyaaaa jujurrrrr issssss	['ga', 'sukakkkkkkkkkm', 'jauhjh', 'beda', 'watsapp', 'nyaaaa', 'jujurrrrr', 'issssss']	ga suka jauh beda watsapp nya jujur is
10.	mencoba download watsapp lagi buka library dan tbtb semua serasa nostalgia entah kenapa cover dan judul judul ini membawa perasaan nostalgia menimbulkan aroma	Positif	mencoba download watsapp lagi buka library dan tbtb semua serasa nostalgia entah kenapa cover dan judul judul ini membawa perasaan nostalgia menimbulkan aroma aroma nostalgia yg gabisa di deskripsi kan	mencoba download watsapp buka library tbtb serasa nostalgia kenapa cover judul judul membawa perasaan nostalgia menimbulkan aroma aroma	['mencoba', 'download', 'watsapp', 'buka', 'library', 'tbtb', 'serasa', 'nostalgia', 'kenapa', 'cover', 'judul', 'judul', 'membawa', 'perasaan', 'nostalgia', 'menimbulkan',	coba download watsapp buka library tbtb rasa nostalgia kenapa cover judul judul bawa rasa nostalgia timbul aroma aroma nostalgia bisa deskripsi

No	Text	Label	Cleansing	Filtering	<i>Tokenizing</i>	Steaming
	aroma nostalgia yg gabisa di deskripsi kan			nostalgia gabisa deskripsi	'aroma', 'aroma', 'nostalgia', 'gabisa', 'deskripsi']	

Lampiran 4 Hasil Pembobotan Kata Menggunakan TF IDF

1	access	account	accuracy	ace	achiyo	act	actigim	acting	action	active	actualiy	ad	ada	adaa	adaaa	adao	adaoanja	adain	adalan	adaptasi	adara	additional	aeogan
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0,102555	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0,367701	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0,767023	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0,158791	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
--	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Lampiran 5 Hasil Evaluasi *Confusion Matrix*

Rasio 90 :10

Performance Metrics (All Test Data):

- Akurasi : 90.83%
- Presisi : 85.48%
- Recall : 97.25%
- F1-Score: 90.99%

Performance Metrics label Positif:

- Akurasi : 97.25%
- Presisi Positif : 85.48%
- Recall Positif : 97.25%
- F1-Score Positif: 90.99%

Performance Metrics label Negatif:

- Akurasi : 85.00%
- Presisi Negatif : 97.14%
- Recall Negatif : 85.00%
- F1-Score Negatif: 90.67%

Rasio 80 : 20

Performance Metrics (All Test Data):

- Akurasi : 90.61%
- Presisi : 85.32%
- Recall : 97.29%
- F1-Score: 90.91%

Performance Metrics label Positif:

- Akurasi : 97.29%
- Presisi Positif : 85.32%
- Recall Positif : 97.29%
- F1-Score Positif: 90.91%

Performance Metrics label Negatif:

- Akurasi : 84.39%
- Presisi Negatif : 97.09%
- Recall Negatif : 84.39%
- F1-Score Negatif: 90.29%

Rasio 70 : 30

Performance Metrics (All Test Data):

- Akurasi : 89.08%
- Presisi : 83.78%
- Recall : 95.38%
- F1-Score: 89.21%

Performance Metrics label Positif:

- Akurasi : 95.38%
- Presisi Positif : 83.78%
- Recall Positif : 95.38%
- F1-Score Positif: 89.21%

Performance Metrics label Negatif:

- Akurasi : 83.43%
- Presisi Negatif : 95.27%
- Recall Negatif : 83.43%
- F1-Score Negatif: 88.95%

Lampiran 7 Hasil Cek Plagiarisasi Menggunakan Turnitin

cek turnitin 2.docx

ORIGINALITY REPORT

22% SIMILARITY INDEX	20% INTERNET SOURCES	13% PUBLICATIONS	6% STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	-----------------------------

PRIMARY SOURCES

1	etheses.uin-malang.ac.id Internet Source	5%
2	repository.uin-suska.ac.id Internet Source	1%
3	openjournal.unpam.ac.id Internet Source	1%
4	core.ac.uk Internet Source	<1%
5	repository.its.ac.id Internet Source	<1%
6	123dok.com Internet Source	<1%
7	lirboyo.net Internet Source	<1%
8	Submitted to Universitas Muhammadiyah Makassar Student Paper	<1%
9	text-id.123dok.com Internet Source	<1%
10	eprints.umm.ac.id Internet Source	<1%
11	Submitted to Universitas Brawijaya Student Paper	<1%
12	www.coursehero.com Internet Source	<1%