

**IN
REVIEW**



**Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk
Analisa Sentimen Komentar Aplikasi**

TUGAS AKHIR

Jody
41517010040

**PROGRAM STUDI TEKNIK INFORMATIKA
U FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021**



**Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk
Analisa Sentimen Komentar Aplikasi**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Jody
41517010040

PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS MERCU BUANA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41517010040

Nama : Jody

Judul Tugas Akhir : Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 29 Januari 2021



Jody

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Jody
NIM : 41517010040
Judul Tugas Akhir : Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 29 Januari 2021



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN LUARAN TUGAS AKHIR

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Jody
NIM : 41517010040
Judul Tugas Akhir : Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan ✓
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal : JEPIN (Jurnal Edukasi dan Penelitian Informatika) ISSN : 2548-9364 Link Jurnal : https://jurnal.untan.ac.id/index.php/index/user Link File Jurnal Jika Sudah di Publish :		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
Dosen Pembimbing TA

Jakarta, 29 Januari 2021

UNIVERSITAS
MERCU BUANA


Desi Ramayanti, S.Kom., MT.



Jody

LEMBAR PERSETUJUAN PENGUJI



LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010040
Nama : Jody
Judul Tugas Akhir : Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 19 Februari 2021

(Dr. Leonard Goeimanto)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010040
Nama : Jody
Judul Tugas Akhir : Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 9 Februari 2021



(Sri Dianing, Asri, ST, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010040
Nama : Jody
Judul Tugas Akhir : Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 9 Februari 2021



UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

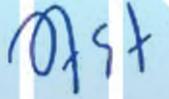
LEMBAR PENGESAHAN

NIM : 41517010040
Nama : Jody
Judul Tugas Akhir : Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 9 Februari 2021

Menyetujui,



(Desi Ramayanti, S.Kom, MT.)
Dosen Pembimbing

Mengetahui,



(Diky Firdaus, S.Kom, MM)
Koord. Tugas Akhir Teknik Informatika



(Desi Ramayanti, S.Kom, MT)
Ka. Prodi Teknik Informatika

UNIVERSITAS
MERCU BUANA

ABSTRAK

Nama : Jody
NIM : 41517010040
Pembimbing TA : Desi Ramayanti, S.Kom, MT.
Judul : Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

Abstrak - PT. Sinergi Digital Teknologi atau sering disebut Lenna.ai adalah perusahaan yang bergerak dalam pembuatan dan pengembangan aplikasi berbasis kecerdasan buatan dan multiplatform. Sebuah aplikasi yang ada di *marketplace*, akan mendapatkan komentar atau *review* dari pengguna. komentar atau *review* ini berfungsi untuk melihat keefektifan fitur dan kelemahan yang dimiliki aplikasi setelah dirilis. sehingga aplikasi tersebut bisa di kembangkan lebih baik. Saat ini PT Sinergi Digital Teknologi melihat bahwa komentar atau *review* yang disediakan oleh *marketplace* ini belum bisa diolah informasinya secara optimal. Maka dari itu, untuk mendapatkan informasi dan dapat menarik kesimpulan yang lebih efektif dan tepat sasaran dari komentar pengguna diperlukan analisa sentimen sebagai sistem yang bekerja secara otomatis dalam prediksi sentimen. Sampel data yang diambil merupakan data komentar yang diambil dari API *marketplace* Google Play dan Apps Store berkaitan dengan aplikasi yang telah dirilis PT Sinergi Digital Teknologi dan mengambil 3 sampel aplikasi yaitu BNI Call, Lenna Virtual Assistant dan Tripa Smart sebanyak 1000 data. Data dibagi menjadi data latih dan data uji, dimana data latih akan digunakan untuk acuan akurasi data uji. Dari hasil pengujian, didapatkan bahwa algoritma Support Vector Machine memperoleh akurasi rata-rata sebesar 98.53%. Adapun algoritma pembandingan yaitu Naïve Bayes, yaitu sebesar 96.23%.

Kata kunci: *Support Vector Machine*, *Naïve Bayes*, Analisa Sentimen, Lenna, Komentar.

ABSTRACT

Name : Jody
Student Number : 41517010040
Counsellor : Desi Ramayanti, S.Kom, MT.
Title : *Comparison of Support Vector Machine and Naive Bayes Algorithms for Sentiment Analysis of Application Comment*

PT. Sinergi Digital Teknologi or often called Lenna.ai is a company engaged in the creation and development of applications based on artificial intelligence and multiplatform. An application in the marketplace will get comments or reviews from users. This comment or review serves to see the effectiveness of the features and weaknesses that the application has after its release. so that the application can be developed better. Currently, PT Sinergi Digital Teknologi sees that the comments or reviews provided by this marketplace cannot be processed optimally. Therefore, to obtain information and be able to draw more effective and targeted conclusions from user comments, sentiment analysis is needed as a system that works automatically in predicting sentiment. The data sample taken is the comment data taken from the Google Play API marketplace and the Apps Store related to the application that has been released by PT Sinergi Digital Teknologi and took 3 application samples, namely BNI Call, Lenna Virtual Assistant, and Tripa Smart as many as 1000 data. The data is divided into training data and test data, where the training data will be used to reference the accuracy of the test data. From the test results, it was found that the Support Vector Machine algorithm obtained an average accuracy of 98.53%. The comparison algorithm is Naïve Bayes, which is 96.23%.

Keywords: Support Vector Machine, Naïve Bayes, Sentiment Analysis, Lenna, Comment.

KATA PENGANTAR

Puji syukur kita panjatkan kepada ALLAH SWT yang telah memberikan nikmat dan rahmatnya sehingga penulis dapat menyelesaikan laporan tugas akhir ini yang berjudul “Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi” dengan lancar. Tugas akhir ini disusun untuk memenuhi syarat memperoleh gelar sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari dosen pembimbing, orang tua serta teman-teman tidak akan terselesaikan dengan baik tugas akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada :

1. Ibu Desi Ramayanti, S.Kom, MT selaku Kepala Prodi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana dan selaku dosen pembimbing tugas akhir yang telah memberikan masukan saat bimbingan dan meluangkan waktu sebagian besarnya untuk melakukan bimbingan dalam penyusunan tugas akhir ini hingga selesai.
2. Bapak Diky Firdaus, S.Kom, MM selaku dosen pembimbing akademik yang telah membimbing saya sejak awal mula semester hingga sampai saat ini dan selalu memberikan motivasi serta arahan agar dapat lulus tepat waktu.
3. Orang tua yang selalu memberikan dukungan secara penuh dan doa sehingga dapat menyelesaikan semua laporan tugas akhir dengan lancar.
4. Teman-teman dan sahabat yang selalu memberikan dukungan dan memotivasi dalam melakukan penulisan tugas akhir ini agar dapat terselesaikan dengan baik.

Akhir kata, penulis berharap tugas akhir ini dapat menjadi bermanfaat bagi pembaca dan menambah wawasan pengetahuan semua pihak.

Jakarta, 29 Januari 2021

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	1
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI	v
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL	1
KERTAS KERJA.....	10
BAB 1. LITERATUR REVIEW	11
BAB 2. SOURCE CODE	15
BAB 3. DATASET.....	19
BAB 4. TAHAPAN EKSPERIMEN.....	22
BAB 5. HASIL SEMUA EKSPERIMEN.....	33
DAFTAR PUSTAKA	44
LAMPIRAN DOKUMEN HAKI.....	45
LAMPIRAN KORESPONDENSI	48
LAMPIRAN CV	49

NASKAH JURNAL


JEPIN
(Jurnal Edukasi dan Penelitian Informatika)

ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

 Vol. x
 No. y
 mm yy

Komparasi Algoritma Support Vector Machine dan Naive Bayes untuk Analisa Sentimen Komentar Aplikasi

 Jody^{#1}, Desi Ramayanti^{#2}
[#]Jurusan Informatika Fakultas Ilmu Komputer Universitas Mercu Buana
 Jl. Meruya Selatan No. 1 Kembangan, Kota Jakarta Barat, 11650

^{#1}41517010040@student.mercubuana.ac.id

^{#2}desi.ramayanti@mercubuana.ac.id

Abstrak— PT. Sinergi Digital Teknologi atau sering disebut Lenna.ai adalah perusahaan yang bergerak dalam pembuatan dan pengembangan aplikasi berbasis kecerdasan buatan dan multiplatform. Sebuah aplikasi yang ada di *marketplace*, akan mendapatkan komentar atau *review* dari pengguna. komentar atau *review* ini berfungsi untuk melihat keefektifan fitur dan kelemahan yang dimiliki aplikasi setelah dirilis. sehingga aplikasi tersebut bisa di kembangkan lebih baik. Saat ini PT Sinergi Digital Teknologi melihat bahwa komentar atau *review* yang disediakan oleh *marketplace* ini belum bisa diolah informasinya secara optimal. Maka dari itu, untuk mendapatkan informasi dan dapat menarik kesimpulan yang lebih efektif dan tepat sasaran dari komentar pengguna diperlukan analisa sentimen sebagai sistem yang bekerja secara otomatis dalam prediksi sentimen. Sampel data yang diambil merupakan data komentar yang diambil dari API *marketplace* Google Play dan Apps Store berkaitan dengan aplikasi yang telah dirilis PT Sinergi Digital Teknologi dan mengambil 3 sampel aplikasi yaitu BNI Call, Lenna Virtual Assistant dan Tripa Smart sebanyak 1000 data. Data dibagi menjadi data latih dan data uji, dimana data latih akan digunakan untuk acuan akurasi data uji. Dari hasil pengujian, didapatkan bahwa algoritma Support Vector Machine memperoleh akurasi rata-rata sebesar 98,53%. Adapun algoritma perbandingan yaitu Naive Bayes, yaitu sebesar 96,23%.

Kata kunci— *Support Vector Machine, Naive Bayes, Analisa Sentimen, Lenna, Komentar.*

I. PENDAHULUAN

Analisa sentimen yang juga disebut penggalian opini adalah bidang studi yang menganalisis pendapat, sentimen, penilaian, sikap, dan emosi orang terhadap entitas dan atributnya yang diungkapkan dalam teks tertulis. Entitas dapat berupa produk, layanan, organisasi, individu, acara, isu atau topik [1]. Tujuan dari analisa sentimen adalah untuk membuat sistem secara otomatis yang mampu mengekstrak informasi subjektif dari teks dalam bahasa alami (NLP), seperti opini dan sentimen, sehingga menciptakan pengetahuan yang terstruktur dan dapat ditindaklanjuti untuk digunakan oleh sistem pendukung pengambilan keputusan [2].

Implementasi dari analisa sentimen dapat digunakan untuk memprediksi pemenang pemilihan gubernur provinsi Jawa Barat periode 2018-2023 berdasarkan dari tweet yang diberikan masyarakat [3]. Dapat pula digunakan untuk menilai suatu pelayanan perusahaan perbankan yang diperoleh dari komentar para nasabah dalam menentukan apakah pelayanan yang diberikan sudah memuaskan atau masih perlu diperbaiki [4]. Serta digunakan untuk melakukan pendeteksian komentar yang mengandung unsur sarkastik (sindiran yang tajam) pada suatu grup di aplikasi Whatsapp [5].

PT. Sinergi Digital Teknologi atau sering disebut Lenna.ai adalah perusahaan yang bergerak dalam pembuatan dan pengembangan aplikasi berbasis kecerdasan buatan dan multiplatform. Untuk saat ini, aplikasi yang dihasilkan sebanyak 23 aplikasi, yang berkategori *e-commerce support, Chatbot, Omnichannel, digital helper* bahkan sampai *Asisten Virtual*. Contoh dari aplikasinya yaitu BNI Call Center, Tripa Smart, Lenna Virtual Assistant dan lain sebagainya. Aplikasi - aplikasi tersebut, ada yang dijual kepada mitra dan ada yang merupakan hak milik Lenna itu sendiri. Beberapa aplikasi telah dirilis di berbagai *marketplace* penjualan aplikasi, baik yang dijual ke mitra atau yang dimiliki sendiri.

Sebuah aplikasi yang ada di *marketplace*, akan mendapatkan komentar atau *review* dari pengguna. komentar atau *review* ini berfungsi untuk melihat keefektifan fitur dan kelemahan yang dimiliki aplikasi setelah dirilis. sehingga aplikasi tersebut bisa di kembangkan lebih baik. Saat ini PT Sinergi Digital Teknologi melihat bahwa komentar atau *review* yang disediakan oleh *marketplace* ini belum bisa diolah informasinya secara optimal. Hal ini disebabkan karena semua komentar atau *review* harus dikelola secara manual satu persatu, sehingga berdampak lambatnya penarikan kesimpulan dari aplikasi yang telah dirilis apakah telah berjalan dengan baik atau masih diperlukan perbaikan.

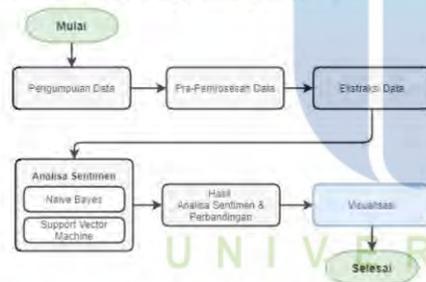
Maka dari itu, untuk mendapatkan informasi dan dapat menarik kesimpulan yang didapatkan dari komentar pengguna diperlukan analisa sentimen tanpa harus melakukan pra-pemrosesan, pengolahan dan ekstraksi data secara manual

serta dapat melakukan pemrosesan dari berbagai sumber media secara sekaligus. Selain itu, agar penyampaian informasi dapat tersampaikan secara maksimal kepada pihak pengembang aplikasi, diperlukan pula sistem antarmuka yang dapat dilihat melalui *website* untuk memudahkan dalam melakukan monitoring aplikasi yang telah dirilis.

Terkait hal yang disebutkan diatas, maka pada penelitian ini proses pengolahan data analisa sentimen dari komentar atau *review* pengguna menggunakan metode Support Vector Machine (SVM) dan Naive Bayes. Pemilihan metode ini didasarkan dari beberapa penelitian terdahulu tentang sentimen analisis. Seperti penelitian yang dilakukan oleh MO Pratama dkk, dengan judul "*The Sentiment Analysis Of Indonesia Commuter Line Using Machine Learning Based On Twitter Data*", dimana dalam penelitian ini membahas mengenai analisis sentimen pada sosial media menggunakan beberapa algoritma seperti Support Vector Machine, Naive Bayes, dan Random Forest. Dari hasil penelitian tersebut, membuktikan bahwa algoritma Support Vector Machine merupakan pilihan terbaik dalam menganalisa sentimen [6]. Akan tetapi, pada penelitian yang dilakukan oleh Muhammad Afzaal dkk, dengan judul "*Tourism Mobile App with Aspect-Based Sentiment Classification Framework for Tourist Reviews*". Pada jurnal tersebut membahas analisa sentimen dari sosial media mengenai *review* pelayanan hotel dan restaurant, membuktikan bahwa algoritma Naive Bayes memperoleh nilai akurasi rata-rata terbaik yaitu sebesar 89% [7]. Maka dari itu, penelitian ini dilakukan untuk mencari algoritma terbaik dari SVM dan Naive Bayes tersebut untuk diimplementasikan ke dalam aplikasi monitoring produk yang telah dirilis oleh Lenna.ai untuk melihat perkembangan produk sejak perilsan bernama Lenna Analytic.

II. METODE PENELITIAN

Cara Pada penelitian ini metode yang dilakukan dalam melakukan pengujian dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Berikut ini merupakan penjelasan dari Gambar 1, yang dilakukan untuk proses uji algoritma analisis sentimen terbaik dari algoritma Naive Bayes dan Support Vector Machine berdasarkan komentar pengguna aplikasi :

A. Pengumpulan Data

Dataset komentar pengguna dikumpulkan untuk analisis sentimen diambil dari API *marketplace* Google Play dan Apps Store berkaitan dengan aplikasi yang telah dirilis PT Sinergi Digital Teknologi dan mengambil 3 sampel aplikasi yaitu BNI Call, Lenna Virtual Assistant dan Tripa Smart yang diambil sejak perilsan aplikasi tersebut yaitu 2017 sampai 2020. Dari proses pengambilan data komentar melalui API tersebut diperoleh 1387 data komentar, akan tetapi dilakukan proses seleksi dan *cleaning* data sehingga menjadi 1000 data. Data tersebut dikumpulkan dan diolah, data komentar aplikasi yang dirilis PT Sinergi Digital Teknologi menjadi sampel data analisis sentimen karena masih kurangnya pemanfaatan data yang dimiliki oleh pihak perusahaan dalam menarik kesimpulan aplikasi yang telah dirilis tersebut, apakah aplikasi yang telah rilis sudah baik dan memenuhi permintaan pengguna atau lebih baik dilakukan perbaikan secara menyeluruh.

Dari data komentar yang diperoleh, terdapat 5 field yaitu *username*, *user image*, *tanggal review*, *rating*, dan *text review*. Akan tetapi untuk melakukan proses *training* dan *testing* data hanya menggunakan variabel *text review* untuk dianalisa sentimennya. Selain melakukan rekapitulasi dan penggabungan data komentar yang diperoleh dari *marketplace* Google Play juga dilakukan pencarian studi pustaka untuk mendapatkan informasi yang digunakan sebagai referensi di dalam penelitian ini, sumber pustaka berbentuk artikel, buku, jurnal online yang berkaitan terhadap algoritma Naive Bayes dan Support Vector Machine dalam melakukan analisa sentimen berdasarkan komentar pengguna.

Kemudian data tersebut dibagi menjadi dua bagian yaitu data *latih* dan data *uji*. Berikut ini pembagian data komentar tersebut menjadi beberapa bagian seperti yang diperlihatkan di tabel I.

TABEL I
PEMBAGIAN DATA UNTUK ANALISIS SENTIMEN

Data Latih	Data Uji
800	200
700	300
600	400
500	500

B. Pengolahan Data Awal

Tahap Pengolahan Data Awal atau *Preprocessing data* ini terbagi menjadi dua bagian meliputi *pelabelan data* dan *pembersihan data (data cleaning)*. Bagian *pelabelan* ini merupakan proses pemberian hasil *sentiment* yang dilakukan secara manual untuk dipelajari mesin dan sebagai acuan dalam melakukan analisa sentimen.

Setelah proses *pelabelan* itu dilakukan kemudian masuk ke tahap *pembersihan data*, dikarenakan data yang telah dikumpulkan memiliki beberapa *missing value*, seperti hanya berisi emoji, yang belum atau berbeda dari bahasa baku. Data-data ini memerlukan proses seleksi dan *pembersihan data* yang berguna untuk mengurangi dimensi kumpulan teks, menghapus kata-kata yang dirasa tidak penting sehingga

meningkatkan akurasi pada teks [8]. Selain meningkatkan akurasi, tahap ini dapat memudahkan mesin dalam melakukan proses klasifikasi terhadap data komentar yang ada. Proses pembersihan data dapat dilakukan menggunakan beberapa tahapan yaitu *case folding*, penghapusan angka dan tanda baca, tokenisasi, normalisasi, *stemming* dan *stopword removal*. Proses pengolahan data awal menggunakan beberapa library pendukung pada python yaitu Sastrawi, NLTK, dan Re. Sedikit potongan program pengolahan data awal dengan bahasa python terdapat pada Gambar 2.

```
def preprocess_text(review):
    #Case Folding
    text = review.lower()
    #Menghapus Tanda Baca
    tokenizer = RegexpTokenizer(r'\w+')
    text = tokenizer.tokenize(text)
    text = " ".join(text)
    #Menghapus Angka
    text = re.sub(r'\d+', "", text)
    #Tokenisasi
    text = text.split()
    #Normalisasi
    reformed = [appos[word] if word in appos else word for word in text]
    reformed = " ".join(reformed)
    #Stemming Data
    text = stemmer.stem(reformed)
    #Menghapus Stop-word
    text = stopwords.remove(text)
    return text
```

Gambar 2. Potongan Program Pengolahan Data Awal

Berikut ini penjelasan lebih lanjut mengenai tahapan yang dilakukan dalam pengolahan data awal.

1. Case Folding

Huruf kapital biasanya digunakan dalam suatu teks, seperti awal kalimat atau sebagai bagian dari judul atau sebagai awal kata subjek. Namun dalam beberapa kasus huruf kapital tersebut dapat menimbulkan beberapa masalah. Salah satunya dalam melakukan pengolahan data awal, dengan adanya huruf kapital dapat menimbulkan ambiguitas dan perbedaan arti dalam suatu kata [9]. Maka dari itu tahap ini berguna untuk melakukan perubahan dari kata yang memiliki huruf kapital menjadi huruf kecil. Contoh hasil dari *Case Folding* dapat dilihat pada tabel II.

TABEL II
CASE FOLDING

Teks	Hasil
Saya mau Tanya tentang transferan ke Bank lain yg belum sampai2 .Dan sampai skrng gak ada solusi.	saya mau tanya tentang transferan ke bank lain yg belum sampai2 .dan sampai skrng gak ada solusi.

2. Penghapusan angka dan tanda baca

Dalam melakukan pengolahan data awal pengerjaan tanda baca perlu diperhatikan. Dalam beberapa kasus, ini menentukan batas antar kata dan terkadang kurang diperlukan pada saat melakukan pra pemrosesan [9]. Begitu pula angka yang akan

membuat ambigu pada suatu teks review yang ada. Maka dari itu di tahap ini dilakukan penghapusan pada tanda baca dan angka untuk mengurangi kesalahan pembelajaran data serta untuk mencegah perubahan makna dari teks review tersebut. Contoh hasil dari penghapusan angka dan tanda baca dapat dilihat pada tabel III dan IV.

TABEL III
MENGHAPUS TANDA BACA

Teks	Hasil
saya mau tanya tentang transferan ke bank lain yg belum sampai2 .dan sampai skrng gak ada solusi.	saya mau tanya tentang transferan ke bank lain yg belum sampai2 dan sampai skrng gak ada solusi

TABEL IV
MENGHAPUS ANGKA

Teks	Hasil
saya mau tanya tentang transferan ke bank lain yg belum sampai2 dan sampai skrng gak ada solusi	saya mau tanya tentang transferan ke bank lain yg belum sampai dan sampai skrng gak ada solusi

3. Tokenisasi

Tahap tokenisasi merupakan suatu tahap pemotongan teks review berdasarkan setiap kata yang menyusunnya. tokenisasi ini memecah sekelompok teks tersebut menjadi unit kata, dengan melakukan pemotongan apabila ditemukan suatu karakter pemisah kata. Misalnya, karakter spasi, enter, tabulasi [10]. Contoh hasil dari Tokenisasi dapat dilihat pada tabel V.

TABEL V
TOKENISASI

Teks	Hasil
saya mau tanya tentang transferan ke bank lain yg belum sampai dan sampai skrng gak ada solusi	['saya', 'mau', 'tanya', 'tentang', 'transferan', 'ke', 'bank', 'lain', 'yg', 'belum', 'sampai', 'dan', 'sampai', 'skrng', 'gak', 'ada', 'solusi']

4. Normalisasi

Pada tahap ini digunakan untuk mengubah suatu kata, hal ini dikarenakan beberapa kata yang digunakan pemberi review tidak secara baku dan menggunakan bahasa terbilang terlalu sehari-hari. Sehingga harus dilakukan tahap tambahan ini, terkadang pada tahap stemming beberapa kata tidak dapat terdeteksi dan dibiarkan. Contoh hasil dari Normalisasi dapat dilihat pada tabel VI.

TABEL VI
NORMALISASI

Teks	Hasil
['saya', 'mau', 'tanya', 'tentang', 'transferan', 'ke', 'bank', 'lain', 'yg', 'belum', 'sampai', 'dan', 'sampai', 'skrng', 'gak', 'ada', 'solusi']	saya mau tanya tentang transferan ke bank lain yang belum sampai dan sampai sekarang tidak ada solusi

5. Stemming

Dikarenakan semua review yang diberikan pengguna secara keseluruhan memiliki kata yang tidak baku dan tidak sesuai standar Kamus Besar Bahasa Indonesia (KBBI). Pada tahap ini, Semua kata dalam komentar pengguna yang belum baku dilakukan perubahan menjadi suatu kata baku untuk memudahkan pembelajaran saat melakukan ekstraksi data [6]. Contoh hasil dari *Stemming* dapat dilihat pada tabel VII.

TABEL VII
STEMMING

Teks	Hasil
saya mau tanya tentang transferan ke bank lain yang belum sampai dan sampai sekarang tidak ada solusi	saya mau tanya tentang transfer ke bank lain yang belum sampai dan sampai sekarang tidak ada solusi

6. Stopword Removal

Stopword atau kata penghenti adalah kata-kata yang biasa ditemukan dalam bahasa apapun, tetapi tidak berkontribusi dengan variasi konten. Misalnya, saat mengklarifikasi artikel berdasarkan bidang studi dengan merepresentasikan kata-kata dengan frekuensi istilah yang kurang lebih sama [9].

Maka dari itu, tahap ini perlu dilakukan untuk menghilangkan kata-kata seperti itu yang memiliki efek negatif dalam menentukan perbedaan antar teks. Akan tetapi perlu dilakukan dengan hati-hati, karena dibebberapa kasus bisa menimbulkan hilangnya informasi yang terkait dengan penghapusan kata penghenti tersebut. Oleh karena itu, dalam melakukan penelitian ini, peneliti telah menyusun koleksi penghapusan kata penghenti sendiri. Contoh hasil dari *Stopword Removal* dapat dilihat pada tabel VIII.

TABEL VIII
MENGHAPUS STOPWORD

Teks	Hasil
saya mau tanya tentang transfer ke bank lain yang belum sampai dan sampai sekarang tidak ada solusi	mau tentang ke lain belum dan sekarang tidak ada solusi

C. Ekstraksi Data

Pada tahap ini ekstraksi data digunakan agar data komentar yang dimiliki dapat dibaca oleh mesin dengan mengubah data komentar berupa teks tersebut menjadi angka-angka sehingga memudahkan mesin untuk melakukan pengolahan data tersebut. Dalam tahap ini metode yang digunakan yaitu Term Frequency Inverse Document Frequency (TF-IDF).

Algoritma TF-IDF digunakan untuk memberikan pembobotan pada setiap kata komentar pengguna dan melakukan klasifikasi ke dalam 2 kelas yaitu review potensial dan review tidak potensial. Pembobotan pada TF-IDF umumnya digunakan untuk melakukan penggalian teks dan melakukan pencarian informasi yang ada untuk mengevaluasi pentingnya istilah linguistik (pada umumnya unigram atau bigram) dalam korpus yang diteliti [11]. Perhitungan TF-IDF menggunakan persamaan (1). [11]

$$tf_{t,d} = \frac{f_{t,d}}{n_d}$$

$$idf_t = \log \frac{N}{df_t} \quad (1)$$

$$W_{t,d} = tf_{t,d} \cdot idf_t$$

Keterangan :

$tf_{t,d}$ = frekuensi term dalam dokumen d .

df_t = frekuensi dokumen term t , yaitu jumlah di mana dokumen term t muncul.

D. Analisa Sentimen

Setelah melakukan tahap pengolahan data awal dan ekstraksi data. Kemudian dilakukan tahap analisis sentimen untuk mendapatkan informasi sentimen yang terkandung dalam teks review [12]. Pada tahap ini, algoritma yang digunakan adalah Naive Bayes dan Support Vector Machine untuk membandingkan kedua algoritma tersebut.

Teorema Bayes ini sendiri merupakan teorema yang lebih fokus pada konsep probabilitas bersyarat berdasarkan pengalaman masa lalu [13]. Secara umum teorema dari Bayes dirumuskan dalam persamaan (2) [14].

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)} \quad (2)$$

Sedangkan SVM merupakan metode memprediksi dalam pengklasifikasian dan regresi. SVM terdiri untuk menyelesaikan masalah baik *linear* dan *nonlinear* dengan mencari *hyperplane* untuk memaksimalkan margin antar kelas data. *Hyperplane* berguna untuk memisahkan 2 kelompok kelas yaitu -1 dan +1 [15]. *Hyperplane* dalam SVM dinyatakan dalam persamaan (3) [16].

$$\vec{w} \cdot \vec{x} + b = 0 \quad (3)$$

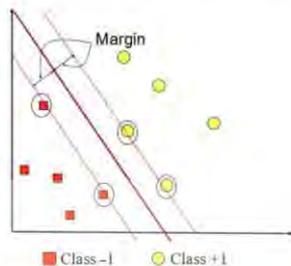
Patten \vec{w} yang termasuk class -1 (negatif) dapat dirumuskan sebagai pola yang memenuhi pertidaksamaan [16].

$$\vec{w} \cdot \vec{x} + b < 0 \quad (4)$$

Sedangkan pattern \vec{w} yang termasuk class +1 (positif) [16].

$$\vec{w} \cdot \vec{x} + b \geq 0 \quad (5)$$

Dimana \vec{w} merupakan vektor terdekat dengan hyperplane, dan $\frac{b}{\|\vec{w}\|}$ adalah jarak antara *support vector* dengan *hyperplane* [10].



Gambar 3. Gambaran mengenai metode SVM

Analisa sentimen dari dataset ini adalah untuk melatih dan memprediksi analisa sentimen yang diperoleh dari data komentar tersebut.

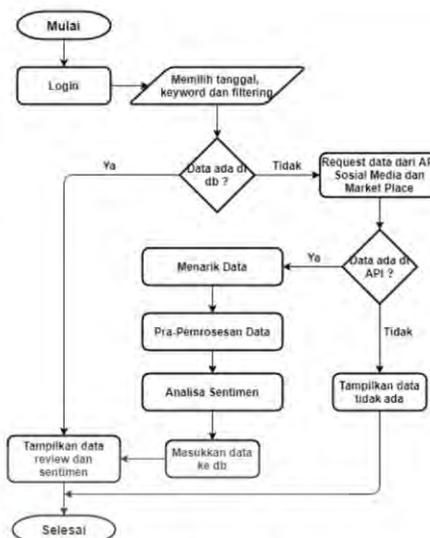
E. Hasil Analisa dan Perbandingan

Keluaran yang dihasilkan oleh algoritma Naive Bayes dan Support Vector Machine yaitu kemampuan memprediksi analisa sentimen yang dinilai dari tingkat akurasi, *precision*, *recall* dan *f1-score* yang diperoleh dari data komentar tersebut. Setelah itu, dilakukan perbandingan dari kedua algoritma tersebut mana yang memiliki hasil terbaik yang kemudian akan diimplementasikan ke dalam aplikasi di tahap visualisasi.

F. Visualisasi

Tahap ini untuk menampilkan dan mengimplementasikan algoritma terbaik ke dalam aplikasi monitoring *review* produk PT Sinergi Digital Teknologi untuk melihat perkembangan produk aplikasi yang telah dirilis berbasis website menggunakan Vue JS sebagai Front-End dan Flask (Framework python) sebagai Restful API dan melakukan prediksi sentimen.

Dengan adanya aplikasi ini diharapkan dapat membantu pihak perusahaan untuk melihat sentimen berdasarkan komentar yang didapatkan dari pengguna aplikasi sehingga pihak pengembang aplikasi dapat menarik kesimpulan bahwa aplikasi yang telah mereka rilis sudah layak atau belum untuk digunakan. Untuk alur aplikasinya terlihat pada Gambar 4.



Gambar 4. Flow Aplikasi Monitoring Review Produk

Berikut ini adalah penjelasan flowchart saat menampilkan analisis sentimen pada aplikasi :

1. User melakukan login pada aplikasi.
2. User akan diminta untuk memilih tanggal dan melakukan *filtering* data komentar yang ingin ditampilkan beserta sentimennya. Apabila user baru memasuki aplikasi, maka sistem akan melakukan *filtering* secara *default*.
3. Sistem akan melakukan pengecekan pada database apakah data yang dicari tersebut ada, jika ada menuju ke Langkah 7, apabila tidak ada menuju ke langkah 4.
4. Sistem melakukan request data dari API media sosial dan *marketplace*. Kemudian sistem mengecek apakah data komentar tersebut ada atau tidak. Jika ada menuju ke Langkah 5 dan jika tidak maka tampilkan ke aplikasi bahwa data tidak ada. Potongan source code penarikan data dari API terdapat pada Gambar 5.

```

import requests
import json

url = "https://api.twitter.com/1.1/statuses/user_timeline.json?screen_name=mercu_buana"
headers = {"Authorization": "Bearer AAAAAAAAAAAAAAAAAAAAANRILgAAAAAAnNwIwlMg8AgAAQm"}

response = requests.get(url, headers=headers)
data = json.loads(response.text)

print(data)

```

Gambar 5. Source Code Penarikan Data dari API

5. Data yang telah didapatkan kemudian diolah dan dibersihkan dengan melakukan *Case Folding*, *Tokenisasi*, *Penghapusan tanda baca dan angka*, *Stemming*, dan *Penghapusan Kata Henti*. Setelah itu, data yang sudah dibersihkan akan diprediksi sentimennya menggunakan algoritma yang terbaik saat dalam pengujian algoritma analisis sentimen.
6. Masukkan data komentar yang telah diprediksi sentimennya ke dalam database.
7. Menampilkan data komentar dan prediksi sentimennya beserta data *word cloud* yang ada.

III. HASIL DAN PEMBAHASAN

A. Pengujian algoritma

Berikut ini merupakan hasil dari pengujian menggunakan algoritma Support Vector Machine dan Naive Bayes menggunakan beberapa pembagian data komentar. Perolehan hasil tersebut ditunjukkan pada tabel IX.

TABEL IX
HASIL ANALISA SENTIMEN

Jumlah Data Latih	Jumlah Data Uji	Sentimen	Precision SVM (%)	Recall SVM (%)	F1-Score SVM (%)	Accuracy SVM (%)	Precision Naive Bayes (%)	Recall Naive Bayes (%)	F1-Score Naive Bayes (%)	Accuracy Naive Bayes (%)
800	200	Pos	100	100	100	100	100	67	80	97
		Neg	100	100	100		97	100	98	
700	300	Pos	97	93	95	99	100	67	80	97
		Neg	99	100	99		96	100	98	
600	400	Pos	92	85	89	97.75	93	63	75	95.75
		Neg	98	99	99		96	99	98	
500	500	Pos	93	80	86	97.40	97	55	70	95.20
		Neg	98	99	99		95	100	97	

B. Perbandingan Algoritma

Pada tabel IX dapat dilihat bahwa algoritma Support Vector Machine memperoleh hasil yang lebih baik dengan tingkat akurasi rata-rata sebesar 98.53% dengan tingkat rata-rata *precision*, *recall*, dan *f1-score* sebesar 97.12%, 94.5%, 95.87% dan segi waktu eksekusi algoritma Support Vector Machine menghabiskan waktu 0.0015 *second*. Sedangkan algoritma naive bayes hanya memperoleh akurasi sebesar 96.23% dengan tingkat rata-rata *precision*, *recall*, dan *f1-score* sebesar 96.75%, 81.37%, 87% dan segi waktu eksekusi algoritma Naive Bayes menghabiskan waktu 0.002 *second*.

Dari tabel IX tersebut dapat disimpulkan bahwa algoritma support vector machine memiliki hasil analisa sentimen yang lebih baik daripada Naive Bayes. Maka dari itu algoritma support vector machine diimplementasikan ke dalam aplikasi monitoring review produk Lenna Analytic karena merupakan algoritma analisis sentimen terbaik. Meskipun hasil tersebut menunjukkan demikian, peran dari pengolahan data awal sangat penting karena dapat mempengaruhi hasil tingkat akurasi dari kedua algoritma, semakin banyak pembelajaran yang dilakukan pada proses ini dapat meningkatkan tingkat akurasi tersebut. Sedangkan untuk tingkat *precision*, *recall*, dan *f1-score* berpengaruh terhadap pengaturan yang dilakukan dari ekstraksi data, pada pengujian ini saya menggunakan *library* python yaitu *TfidfVectorizer* dengan menambahkan parameter *min_df=3*, *max_df=0.3*, dan mengaktifkan mode *IDF*.

C. Visualisasi

Dikarenakan algoritma *Support Vector Machine* memiliki tingkat akurasi yang lebih baik dalam melakukan analisa dan prediksi sentimen, maka dari itu algoritma tersebut dipilih untuk diimplementasikan menjadi model dalam prediksi sentimen di aplikasi monitoring review produk Lenna Analytic. Model ini akan berjalan setelah data komentar baru melalui tahap penarikan dari API, pengolahan data awal, dan ekstraksi atau perubahan data menjadi vector. Setelah itu sistem akan dilatih dengan data yang ada sebagai acuan. Kemudian data uji akan di prediksi untuk mendapatkan sentimennya, jika hasil sudah didapatkan sentimen tersebut akan dimasukkan ke dalam database.

Berikut ini potongan source code untuk model algoritma *Support Vector Machine* terdapat pada Gambar 6.

```

from sklearn.svm import SVC
from sklearn.metrics import accuracy_score

# Load data
X_train, y_train = load_data('train_data.csv')
X_test, y_test = load_data('test_data.csv')

# Create SVM classifier
svm = SVC(kernel='linear')

# Train the classifier
svm.fit(X_train, y_train)

# Predict on test data
y_pred = svm.predict(X_test)

# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy: %f' % accuracy)

```

Gambar 6. Model Algoritma Support Vector Machine dalam Aplikasi

Berikut ini hasil yang didapatkan setelah melakukan prediksi sentimen dari data komentar baru tersebut, yang kemudian akan ditampilkan ke halaman website Lenna Analytic terdapat pada Gambar 7 dan Gambar 8.

```

"ulas": [
  {
    "aplikasi": "Lenna Virtual Assistant",
    "id": "1821",
    "link_review": "https://play.google.com/store/apps/details?id=at.lenna.assist",
    "post": "Nantap",
    "rating": "5",
    "sentiment": "1",
    "sumber": "Google Play",
    "tanggal_post": "2020-10-13",
    "user_avatar": "https://play-lh.googleusercontent.com/a-/ADh14CjBi6jd12o19zi",
    "user_name": "Arfan Iah"
  },
  {
    "aplikasi": "Lenna Virtual Assistant",
    "id": "1822",
    "link_review": "https://play.google.com/store/apps/details?id=at.lenna.assist",
    "post": "Ini enak sdh ga di urus ...sdh tdk ada petugas IT... Buang2 kuota...",
    "rating": "1",
    "sentiment": "0",
    "sumber": "Google Play",
    "tanggal_post": "2020-09-05",
    "user_avatar": "https://play-lh.googleusercontent.com/a-/ADh14GkrFjIkz5Z8Dh",
    "user_name": "FAMILY d1a124"
  }
]

```

Gambar 7. Hasil Komentar dan Prediksi Sentimen dari Response API

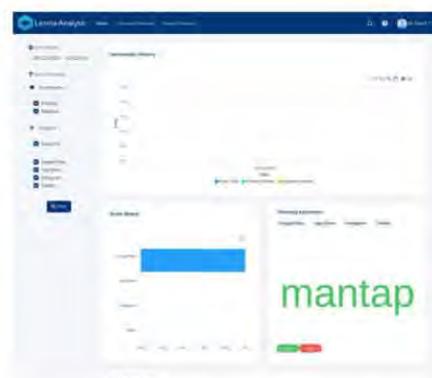
Avatar	Post	Date	Source	Rating	Sentiment
	Apas yang jadi masalah? Boleh banget... masalah karena dia yang bilang masalah udah di rah plying	27 Jun 2020	Google Play	5	1
	Stapa	29 Jun 2020	Google Play	5	1
	halo siapa yang jawab? semoga aja bisa jawab masalah...	18 Jun 2020	Google Play	4	1
	halo...	14 Jun 2020	Google Play	4	1
	harapnya untuk jawaban... semoga aja bisa jawab ya... semoga aja bisa jawab ya...	11 Jun 2020	Google Play	5	1
	...mohon maaf...	26 Jun 2020	Google Play	5	1
	...mohon maaf...	07 Jun 2020	Google Play	5	1

Gambar 8. Hasil Komentar dan Prediksi Sentiment dari Website Lenna Analytic

Tahap visualisasi ini digunakan untuk menampilkan hasil sentimen dan komentar pengguna dari beberapa sumber untuk memudahkan pihak pengembang aplikasi melakukan monitoring aplikasinya dan menarik kesimpulan dari aplikasi yang telah dirilis tersebut. Visualisasi sentimen ini merupakan aplikasi berbasis website menggunakan Flask (Framework Python) sebagai Restful API dan VueJs sebagai Front-end. Untuk memasuki aplikasi ini *user* harus melakukan *register* dan *login* terlebih dahulu.

Berikut ini fitur yang dimiliki pada Website Lenna Analytic antara lain yaitu :

- Grafik data komentar dan sentimen, fitur ini berguna untuk menampilkan data komentar beserta sentimennya serta wordcloud dari *keyword-keyword* yang sering digunakan oleh pengguna aplikasi untuk memudahkan dalam memantau aplikasi yang telah dirilis oleh member serta mengetahui aplikasi tersebut sudah berjalan dengan baik atau memiliki banyak masalah.



Gambar 9. Tampilan Fitur Home

- Komparasi data komentar, dengan fitur ini member dapat membandingkan review pada periode tertentu untuk melihat tingkat popularitas yang dimiliki dari aplikasi yang telah dirilis tersebut, apakah komentar negatif yang menurun atau bahkan komentar positifnya. Pertama *User* akan diminta untuk mengisi tanggal review yang ingin dikomparasi seperti Gambar 10. Setelah itu tekan Cek! Maka data komentar yang dikomparasi akan tampil terlihat pada Gambar 11.

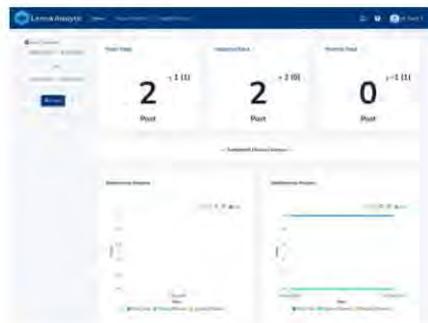
Compare Reviews

Range Tanggal :
10/23/2020 - 11/22/2020

vs

Range Tanggal :
09/23/2020 - 10/23/2020

Gambar 10. Form Komparasi data komentar



Gambar 11. Tampilan Komparasi data komentar



Gambar 13. Tampilan Fitur Laporan

c. Report Review, dengan fitur report ini dapat digunakan untuk member apabila ingin melakukan pelaporan kepada pihak pengembang aplikasi, dapat dengan menunjukkan hasil review tersebut atau dengan mencetak data komentar tersebut. Sama halnya seperti fitur komparasi untuk menampilkan laporan data komentar juga diminta untuk mengisi tanggal untuk laporan yang akan ditampilkan seperti pada Gambar 12. Kemudian data laporan tersebut akan ditampilkan seperti yang terdapat pada Gambar 13.

Cek Laporan

Gambar 12. Form Cek Laporan

IV. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan maka dapat menarik kesimpulan dari keseluruhan pengujian yang dilakukan bahwa algoritma Support Vector Machine lebih akurat dalam melakukan analisis sentimen berdasarkan data komentar pengguna yang berkaitan dengan PT Sinergi Digital Teknologi dengan memperoleh tingkat akurasi rata-rata 98.53% dengan tingkat rata-rata *precision*, *recall*, dan *f1-score* sebesar 97.12%, 94.5%, 95.87%. Sedangkan hasil dari akurasi algoritma Naïve Bayes adalah 96.23% dengan tingkat rata-rata *precision*, *recall*, dan *f1-score* sebesar 96.75%, 81.37%, 87% dan waktu eksekusi dari algoritma Support Vector Machine sedikit lebih baik adalah 0.0015 *second* sedangkan waktu eksekusi algoritma Naive Bayes adalah 0.002 *second*.

REFERENSI

- [1] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 2020.
- [2] F. Alberto Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment Analysis in Social Networks*, 2017.
- [3] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin, and L. Marlinda, "Comparison of SVM Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1-6, 2019, doi: 10.1109/CITSM.2018.8674352.
- [4] E. Dwi N. Sari, D. Statistika, F. Matematika, and S. Data, "Analisis Sentimen Nasabah Pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naïve Bayes Classifier (NBC), dan Support Vector Machine (SVM)," *J. SAINS DAN SEMI ITS*, vol. 8, no. 2, p. 177, 2019.
- [5] R. Afiyati, E. Winarko, and A. Cherid, "Recognizing the sarcastic statement on WhatsApp Group with Indonesian language text," *2017 Int. Conf. Broadband Commun. Wirel. Sensors Powering. BCWSP 2017*, vol. 2018-January, pp. 1-6, 2018, doi: 10.1109/BCWSP.2017.8272579.
- [6] M. O. Pratama *et al.*, "The sentiment analysis of Indonesia commuter line using machine learning based on twitter data," *J. Phys. Conf. Ser.*, vol. 1193, no. 1, 2019, doi: 10.1088/1742-6596/1193/1/012029.
- [7] M. Afzaal, M. Usman, and A. Fong, "Tourism mobile app with aspect-based sentiment classification framework for tourist

- reviews." *IEEE Trans. Consum. Electron.*, vol. 65, no. 2, pp. 233–242, 2019, doi: 10.1109/TCE.2019.2908944.
- [8] N. K. Wardhani *et al.*, "Sentiment analysis article news coordinator minister of maritime affairs using algorithm naive bayes and support vector machine with particle swarm optimization." *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 24, pp. 8365–8378, 2018.
- [9] F. Universitatis, C. C. Licence, and C. C. By-ne-nd, "THE INFLUENCE OF TEXT PREPROCESSING METHODS AND." *Facta Univ. Ser. Math. Informatics*, vol. 34, no. 5, pp. 973–994, 2019.
- [10] E. Sutoyo and A. Almaarif, "Twitter Sentiment Analysis of The Relocation of Indonesia's Capital City." *Bull. Electr. Eng. Informatics*, vol. 9, no. 04, pp. 1620–1630, 2020, doi: <http://dx.doi.org/10.11591/eei.v9i4.2352>.
- [11] R. L. Musyarofah, E. U. Utami, and S. R. Raharjo, "Analisis Komentar Potensial pada Social Commerce Instagram Menggunakan TF-IDF." *J. Eksplor Inform.*, vol. 9, no. 2, pp. 130–139, 2020, doi: 10.30864/eksplor.v9i2.360.
- [12] A. N. Muhammad, S. Bukhori, and P. Pandunata, "Sentiment Analysis of Positive and Negative of YouTube Comments Using Naive Bayes-Support Vector Machine (NBSVM) Classifier." *Proc. - 2019 Int. Conf. Comput. Sci. Inf. Technol. Electr. Eng. ICOMITEE 2019*, vol. 1, pp. 199–205, 2019, doi: 10.1109/ICOMITEE.2019.8920923.
- [13] P. O. A. Sunarya, R. Refianti, A. B. Mutiara, and W. Octaviani, "Comparison of accuracy between convolutional neural networks and Naive Bayes Classifiers in sentiment analysis on Twitter." *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 77–86, 2019, doi: 10.14569/ijacs.2019.0100511.
- [14] R. M. Sauter and S. Rasmussen, *An Introduction to Statistics with Data Analysis*, vol. 36, no. 1, 2016.
- [15] W. A. Luqyana, I. Cholissodin, and R. S. Perlana, "Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine." *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4704–4713, 2018.
- [16] A. Kowalczyk, "Support Vector Machines Succintely, Syncefusion." *Succintely E-b. Ser.*, p. 114, 2017, [Online]. Available: www.syncefusion.com.

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul di atas. Kertas kerja berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat/atau disertakan di artikel jurnal. Di dalam kertas kerja ini disajikan: literature review, *Data Set*, tahapan eksperimen, source code, dan hasil eksperimen secara keseluruhan. Literatur review berisi tentang jurnal-jurnal pendukung yang terkait dengan penelitian analisa sentimen yang telah dibuat. Pada bab ini dijelaskan masing-masing judul jurnal yang dikutip, topik yang dibahas pada setiap jurnal yang dikutip, metode yang digunakan dalam penelitian tersebut, dan hasil yang didapatkan dari pengujian algoritma yang digunakan. Pada bab *Data Set* dilampirkan variabel-variabel yang digunakan sebagai landasan pengujian algoritma, Data tersebut terbagi menjadi dua yaitu data latih dan data uji. Pada bab ini, dilampirkan pula sedikit sampel data yang digunakan dalam penelitian. Tahapan eksperimen merupakan bab yang membahas tentang eksperimen yang akan dilakukan untuk analisa sentimen data komentar menggunakan metode *Support Vector Machine* dan *Naïve Bayes*. Pada bab source code disajikan beberapa potongan code yang digunakan untuk melakukan pengujian algoritma dengan menggunakan bahasa program Python dan beberapa library pendukungnya. Kemudian yang terakhir yaitu bab hasil eksperimen secara keseluruhan membahas hasil yang didapat saat melakukan pengujian algoritma baik dari segi akurasi, presisi, *recall* dan *f1-score*. Selain itu, dijabarkan pula hasil dari perbandingan kedua algoritma yang digunakan tersebut dalam melakukan analisa sentimen sampai pada penarikan kesimpulan algoritma yang terbaik.