

**KLASIFIKASI PENGADUAN PELAYANAN FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MAKASSAR MENGGUNAKAN
*NATURAL LANGUAGE PROCESSING***

SKRIPSI



SUNARTI

105841101220

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MAKASSAR**

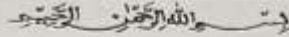
2024



UNIVERSITAS MUHAMMADIYAH MAKASSAR
FAKULTAS TEKNIK

GEDUNG MENARA IQRA LT. 3

Jl. Sultan Alauddin No. 259 Telp. (0411) 866 972 Fax (0411) 865 588 Makassar 90221
 Website: www.unismuh.ac.id, e_mail: unismuh@gmail.com
 Website: <http://teknik.unismuh.makassar.ac.id>



PENGESAHAN

Skripsi atas nama SUNARTI dengan nomor induk Mahasiswa 105 84 11012 20, dinyatakan diterima dan disahkan oleh Panitia Ujian Tugas Akhir/Skripsi sesuai dengan Surat Keputusan Dekan Fakultas Teknik Universitas Muhammadiyah Makassar Nomor : 300/05/A.5-II/VIII/46/2024 sebagai salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar pada hari Senin tanggal 31 Agustus 2024.

Panitia Ujian :

Makassar, 26 Safar 1446 H
 31 Agustus 2024 M

1. Pengawas Umum

a. Rektor Universitas Muhammadiyah Makassar

Dr. Ir. H. Abd. Rakhim Nanda, S.T., M.T., IPU

b. Dekan Fakultas Teknik Universitas Hasanuddin

Prof. Dr. Eng. Muhammad Isran Ramli, S.T., MT

2. Penguji

a. Ketua : Dr. Ir. Zahir Zainuddin, M.Sc

b. Sekertaris : Fahrin Irfhamna Rahman S.Kom., MT.

3. Anggota

: 1. Lukman, S.Kom., MT.

2. Rizki Yusliana Bakti ST., MT.

3. Lukman Anas, S.Kom., MT.

Mengetahui :

Pembimbing I

Pembimbing II

Dr. Ir. Ridwan, S.Kom., M.T., IPM

Muhyiddin A/R Hayat, S.Kom., MT.



Dekan
Dr. Ir. H. Nurhadi, S.T., MT., IPM.



UNIVERSITAS MUHAMMADIYAH MAKASSAR

FAKULTAS TEKNIK

GEDUNG MENARA IQRA LT. 3

Jl. Sultan Alauddin No. 259 Telp. (0411) 866 972 Fax (0411) 865 588 Makassar 90221

Website: www.unismuh.ac.id, e_mail: unismuh@gmail.com

Website: <http://teknik.unismuh.makassar.ac.id>

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan untuk memenuhi syarat ujian guna memperoleh gelar Sarjana Informatika (S.Kom) Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar.

Judul Skripsi : **KLASIFIKASI PENGADUAN PELAYANAN FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MAKASSAR MENGGUNAKAN
NLP**

Nama : SUNARTI

Stambuk : 1058411012 20

Makassar, 31 Agustus 2024

Telah Diperiksa dan Disetujui
Oleh Dosen Pembimbing;

Pembimbing I

Pembimbing II

Dr. Ir. Ridwang, S.Kom., M.T., IPM

Muhyiddin A M Hayat, S.Kom., MT.

Mengetahui,

Ketua Program Studi Informatika

Muhyiddin A M Hayat, S.Kom., MT.

NBM : 1504 577

ABSTRAK

Penelitian ini dilakukan di Fakultas Teknik Universitas Muhammadiyah Makassar dengan tujuan mengembangkan model klasifikasi pengaduan pelayanan menggunakan teknik Natural Language Processing (NLP). Tahapan penelitian melibatkan pengumpulan data pengaduan dari sistem informasi akademik (SIMAK), pre-processing data teks menggunakan teknik seperti tokenisasi, stemming, lemmatization, serta pembobotan teks dengan metode TF-IDF. Model klasifikasi dibangun menggunakan algoritma Naïve Bayes, yang menunjukkan performa yang cukup baik dengan akurasi sebesar 91%. Evaluasi model dilakukan menggunakan confusion matrix untuk menghitung metrik akurasi, recall, precision, dan F1-Score, yang menunjukkan bahwa beberapa kategori seperti "Fasilitas Kampus" dan "Fasilitas Parkir" memiliki nilai precision dan recall yang tinggi, sementara kategori lain seperti "Keuangan dan Pembayaran" membutuhkan peningkatan.

Model yang dikembangkan berhasil mengklasifikasikan keluhan mahasiswa dengan baik, menunjukkan kemampuan yang andal untuk diimplementasikan dalam sistem pengaduan pelayanan di Fakultas Teknik Universitas Muhammadiyah Makassar. Selain itu, model ini mampu mengidentifikasi dan mengklasifikasikan saran dan kritik dengan tingkat akurasi yang memuaskan berkat penggunaan metode TF-IDF untuk ekstraksi fitur. Meskipun demikian, terdapat beberapa kategori yang masih memerlukan peningkatan akurasi, seperti "Keuangan dan Pembayaran" serta "Tata Usaha/Pelayanan Administrasi", yang menunjukkan adanya ruang untuk pengembangan lebih lanjut.

Saran untuk penelitian selanjutnya termasuk eksplorasi algoritma lain seperti Support Vector Machine (SVM) atau Random Forest untuk membandingkan performa model, peningkatan ukuran dan keragaman dataset untuk meningkatkan generalisasi model, dan penambahan analisis sentimen untuk memberikan wawasan lebih mendalam tentang keluhan mahasiswa. Selain itu,

penggunaan teknik NLP lainnya seperti word embeddings (misalnya Word2Vec atau GloVe) dapat diterapkan untuk meningkatkan representasi teks dan akurasi klasifikasi. Hasil penelitian ini memberikan kontribusi penting dalam pengelolaan pengaduan pelayanan di lingkungan akademik dan menjadi dasar untuk perbaikan layanan serta referensi bagi penelitian selanjutnya.

Kata Kunci: Natural Language Processing (NLP), Naïve Bayes, Klasifikasi Pengaduan, TF-IDF, Analisis Sentimen.



ABSTRACT

This research was conducted at the Faculty of Engineering, Muhammadiyah University of Makassar with the aim of developing a service complaint classification model using Natural Language Processing (NLP) techniques. The research stages involved collecting complaint data from the academic information system (SIMAK), pre-processing text data using techniques such as tokenization, stemming, lemmatization, and text weighting with the TF-IDF method. The classification model was built using the Naïve Bayes algorithm, which showed quite good performance with an accuracy of 91%. Model evaluation was carried out using a confusion matrix to calculate accuracy, recall, precision, and F1-Score metrics, which showed that several categories such as "Campus Facilities" and "Parking Facilities" had high precision and recall values, while other categories such as "Finance and Payment" needed improvement.

The developed model successfully classified student complaints well, demonstrating reliable capabilities to be implemented in the service complaint system at the Faculty of Engineering, Muhammadiyah University of Makassar. In addition, this model was able to identify and classify suggestions and criticisms with a satisfactory level of accuracy thanks to the use of the TF-IDF method for feature extraction. However, there are several categories that still need to improve accuracy, such as "Finance and Payment" and "Administration/Administrative Services", which shows that there is room for further development.

Suggestions for further research include exploring other algorithms such as Support Vector Machine (SVM) or Random Forest to compare model performance, increasing the size and diversity of the dataset to improve model generalization, and adding sentiment analysis to provide deeper insights into student complaints. In addition, the use of other NLP techniques such as word embeddings (e.g. Word2Vec or GloVe) can be applied to improve text representation and classification accuracy. The results of this study provide an important contribution to the management of service complaints in academic

environments and serve as a basis for service improvement and reference for further research.

Keywords: Natural Language Processing (NLP), Naïve Bayes, Classification of Complaints, TF-IDF, Sentiment Analysis.



KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamualaikum warahmatullahi wabarakatuh

Puji syukur kepada ALLAH Subhanahu Wata'ala atas segala limpahan Rahmat, Taufik dan Hidayah-Nya sehingga peneliti dapat menyelesaikan penyusunan proposal ini dengan judul Klasifikasi Pengaduan Pelayanan Fakultas Teknik Universitas Muhammadiyah Makassar Menggunakan *Natural Language Processing*. Shalawat serta salam teriring kepada junjungan Nabi Muhammad SAW, pembawa rahmat bagi semesta alam.

Penelitian ini dibuat sebagai salah satu syarat yang harus ditempuh untuk memperoleh gelar sarjana pada Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Makassar dan merupakan salah satu penerapan ilmu yang telah diambil selama di bangku kuliah. Penulis menyadari bahwa penyusunan proposal ini tidak terlepas dari dukungan, bimbingan dan bantuan banyak pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Kedua Orang tua dan keluarga penulis atas segala doa, cinta, dukungan dan semangat yang tiada henti.
2. Ibu Dr.Ir.Hj Nurnawati, S.T., M.T., I.P.M, selaku Dekan Fakultas Teknik Universitas Muhammadiyah Makassar.
3. Bapak Muhyiddin A M Hayat, S.Kom., MT. Selaku Ketua Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar
4. Bapak Dr.Ir Ridwan S.Kom., M.T, IPM selaku Dosen Pembimbing I dan
5. Bapak Muhyiddin A M Hayat, S.Kom., MT. selaku Dosen Pembimbing II
6. Segenap Dosen Program Studi Informatika yang telah memberikan ilmunya kepada penulis selama masa studi.
7. Teman-teman mahasiswa(i) angkatan 2020 Informatika Fakultas Teknik Universitas Muhammadiyah Makassar yang telah memberikan bantuan dan dukungan kepada penulis selama masa studi khususnya teman-teman kelas A.

8. Serta semua pihak yang telah membantu kelancaran penyelesaian proposal ini yang tidak bisa penulis sebutkan satu per satu.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna, karena keterbatasan kemampuan dan pengalaman. Oleh karena itu, kritik dan saran penulis harapkan sebagai bahan evaluasi dan pembelajaran untuk dapat lebih baik dikemudian hari. Penulis berharap agar laporan ini dapat berguna dan bermanfaat bagi semua pihak.

Wassalamualaikum Warahmatullahi Wabarakatuh

Makassar, Mei 2024

Sunarti



DAFTAR ISI

ABSTRAK	ii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR ISTILAH	1
BAB I PENDAHULUAN	5
A. Latar Belakang	5
B. Rumusan Masalah	7
C. Tujuan Penelitian	7
D. Manfaat Penelitian	8
E. Ruang Lingkup Penelitian.....	8
F. Sistematika Penelitian	8
BAB II TINJAUAN PUSTAKA	10
A. Landasan Teori.....	10
B. Penelitian Terkait.....	13
C. Kerangka Pikir	15
BAB III METODE PENELITIAN	16
A. Tempat Dan Waktu Penelitian.....	16
B. Alat Dan Bahan	16
C. Perancangan Sistem.....	16
D. Teknik Pengujian Sistem.....	18
E. Teknik Analisis Data	20
BAB IV HASIL DAN PEMBAHASAN	22
A. Hasil Pra-pemrosesan Data	22
B. Hasil Ekstraksi Fitur Menggunakan TF-IDF	25
C. Pembagian Data	26
D. Kategorisasi Keluhan	26
E. Identifikasi Saran dan Kritik	27

F. Hasil <i>Naïve Bayes</i> dan NPL	29
BAB V PENUTUP	34
A. Kesimpulan	34
DAFTAR PUSTAKA	36



DAFTAR TABEL

Tabel 1. Confusion Matriks	19
Tabel 2. Dataset Awal	23
Tabel 3. Dataset Setelah Diproses	24



DAFTAR GAMBAR

Gambar 1. Kerangka Pikir	15
Gambar 2. Perancangan Sistem	18
Gambar 3. Hasil Ekstraksi Fitur Menggunakan TF-IDF	25
Gambar 4. Hasil <i>Confusion Matrix</i>	29
Gambar 5. Hasil Klasifikasi	30
Gambar 6. Hasil dari <i>Confusion Matrix</i>	31



DAFTAR ISTILAH

<i>Natural Language Processing (NLP)</i>	Cabang kecerdasan buatan yang berfokus pada interaksi antara komputer dan manusia menggunakan bahasa alami. NLP memungkinkan komputer untuk memahami, menganalisis, dan menghasilkan bahasa manusia.
<i>Python</i>	Bahasa pemrograman tingkat tinggi yang sering digunakan dalam pengembangan aplikasi NLP karena dukungannya terhadap pustaka-pustaka seperti NLTK, spaCy, dan Scikit-learn.
<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	Teknik yang digunakan untuk mengekstrak fitur dari teks. TF-IDF mengukur seberapa penting kata tertentu dalam sebuah dokumen relatif terhadap semua dokumen dalam kumpulan data.
<i>Naïve Bayes</i>	Algoritma pembelajaran mesin berbasis probabilitas yang digunakan untuk klasifikasi teks. Model ini sangat sederhana namun efektif untuk banyak tugas klasifikasi, terutama dalam NLP.
<i>Visual Studio Code</i>	Editor kode sumber gratis yang dikembangkan oleh Microsoft, yang banyak digunakan oleh pengembang untuk menulis dan mengedit kode dalam berbagai bahasa pemrograman, termasuk Python.

Natural Language Toolkit (NLTK)

Pustaka Python yang kuat dan populer untuk melakukan berbagai tugas dalam NLP, seperti tokenisasi, stemming, lemmatization, parsing, dan analisis teks.

Flowchart

Diagram yang mewakili alur kerja atau proses dalam suatu sistem. Dalam konteks NLP, flowchart dapat digunakan untuk memvisualisasikan proses pra-pemrosesan teks, pelatihan model, atau alur kerja lainnya.

Pre-processing

Proses awal dalam NLP yang melibatkan pembersihan dan penyiapan teks mentah untuk analisis lebih lanjut, termasuk langkah-langkah seperti penghapusan stop words, tokenisasi, dan normalisasi teks.

Stop Words

Kata-kata umum yang biasanya dihilangkan dalam pemrosesan teks karena tidak memberikan banyak informasi penting, seperti "dan", "di", "dari", dan "yang" dalam bahasa Indonesia.

Tokenisasi

Proses memecah teks menjadi unit-unit kecil yang disebut token, seperti kata, frasa, atau kalimat. Tokenisasi adalah langkah penting dalam pra-pemrosesan teks untuk analisis lebih lanjut.

Stemming/Lemmatization

Teknik dalam NLP untuk mengurangi kata-kata ke bentuk dasarnya. Stemming

menghilangkan akhiran dari kata untuk menemukan bentuk dasar, sementara lemmatization mengembalikan kata ke bentuk dasar yang diakui secara leksikal.

Data Training

Kumpulan data yang digunakan untuk melatih model pembelajaran mesin. Model belajar dari data ini untuk mengenali pola dan membuat prediksi.

Data Testing

Kumpulan data yang digunakan untuk menguji performa model setelah dilatih. Data testing membantu mengevaluasi kemampuan model dalam membuat prediksi yang akurat pada data baru yang belum pernah dilihat.

Confusion Matrix

Matriks yang digunakan untuk mengukur kinerja model klasifikasi. Matriks ini menunjukkan jumlah prediksi yang benar dan salah yang dibuat oleh model, serta membedakan antara kelas yang benar dan salah.

Recall

Ukuran kinerja model yang menunjukkan proporsi sampel positif yang benar-benar diidentifikasi oleh model. Recall tinggi menunjukkan bahwa model tidak banyak melewatkan sampel positif.

Precision

Ukuran kinerja model yang menunjukkan proporsi sampel yang diidentifikasi sebagai positif yang benar-benar positif. Precision

tinggi menunjukkan bahwa model tidak banyak menghasilkan alarm palsu.

F1-Score

Ukuran kinerja model yang merupakan rata-rata harmonis dari precision dan recall. F1-Score memberikan gambaran keseluruhan tentang keseimbangan antara precision dan recall dalam model.



BAB I

PENDAHULUAN

A. Latar Belakang

Universitas Muhammadiyah Makassar, juga dikenal sebagai Unismuh Makassar. Unismuh Makassar saat ini memiliki akreditasi UNGGUL dari Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Sebagai Perguruan Tinggi Muhammadiyah (PTM), Fakultas Teknik Universitas Muhammadiyah Makassar berada di Makassar, Sulawesi Selatan. Fakultas ini menawarkan berbagai program studi teknik yang dirancang untuk memenuhi kebutuhan industri dan pasar kerja yang semakin kompleks.

Fakultas Teknik Universitas Muhammadiyah Makassar menawarkan kurikulum yang berkualitas tinggi serta berbagai fasilitas dan layanan yang mendukung pendidikan dan penelitian. Ini termasuk laboratorium teknik dengan peralatan modern, perpustakaan yang memiliki koleksi buku dan jurnal terkini, dan layanan bimbingan akademik dan karir yang membantu mahasiswa mencapai potensi mereka.

Pengaduan adalah laporan yang berisi informasi atau bukti pelanggaran tertentu, seperti kode etik dan pedoman perilaku hakim, panitera dan jursita, pegawai negeri sipil atau peraturan disiplin militer, maladministrasi, pelayanan publik, dan/atau pengelolaan keuangan. (Melani, 2019).

Survey kepuasan akademik menjadi salah satu sarana yang wajib untuk diterapkan untuk mengembangkan dan meningkatkan kualitas layanan akademik dengan cepat dan terus menerus sementara kepuasan pengguna layanan terus terpenuhi. Pengaduan layanan universitas sangat penting karena universitas dapat memperbaiki dan meningkatkan kualitas layanan, yang akan memberikan kepuasan kepada mahasiswa dan mendorong minat calon mahasiswa untuk memilih universitas tersebut.

Mengolah data survei secara manual hingga akhirnya menggunakan *Natural Language Processing* sebagai solusinya. Penyaringan data, data

yang dikumpulkan dari survei perlu diperiksa dan diperiksa untuk kesalahan atau ketidaksesuaian. Ini termasuk memastikan bahwa data akurat dan mengurangi entri yang tidak lengkap atau terdistorsi. Setelah itu, data diatur dan diformat untuk memastikan konsistensi, memungkinkan analisis untuk menghasilkan informasi yang berguna. Kategorisasi manual, ini dilakukan untuk memberikan gambaran awal tentang jenis-jenis pengaduan yang ada. Kemudian dilakukan pre-pemrosesan data/implementasi NLP untuk klasifikasi seperti menghapus *stop words*, tokenisasi, stemming dan lemmatization. Selanjutnya, pada penelitian ini menggunakan algoritma *naïve bayes* karena salah satu machine learning yang sering digunakan dalam klasifikasi teks, fitur yang dihasilkan dari teks yang telah pra-pemroses digunakan untuk melatih model *naïve bayes*.

Natural language processing adalah bahasa yang biasa digunakan manusia untuk berkomunikasi satu sama lain. Pengolahan bahasa alam adalah prosedur yang bertujuan untuk membuat komputer dapat mengerti bahasa atau kata yang diucapkan. Untuk komputer dapat memahami maksud manusia dengan baik, bahasanya harus diproses dan dipahami terlebih dahulu. Aplikasi natural language processing dapat mencakup banyak hal, seperti chatbot (aplikasi yang memungkinkan pengguna berbicara dengan komputer), *stemming* atau *lemmatization* (pemotongan kata dalam bahasa tertentu menjadi bentuk dasar untuk mengetahui fungsi setiap kata dalam kalimat), *summarization* (membuat ringkasan dari bacaan), translation tools (menerjemahkan bahasa), dan aplikasi lainnya yang memungkinkan komputer memahami instruksi bahasa. (Usmany, 2022).

Dengan demikian Fakultas Teknik Universitas Muhammadiyah Makassar berkomitmen untuk menyediakan layanan terbaik bagi mahasiswa dan staffnya. Pengelolaan pengaduan pelayanan yang efisien dan tepat sasaran sangat penting untuk meningkatkan kualitas pelayanan dan

kepuasan mahasiswa. Dengan menggunakan *Natural Language Processing*, pengaduan dapat dianalisis dan diklasifikasi secara otomatis, pengelolaan yang efisien dan respons yang lebih cepat terhadap masalah yang dihadapi.

Namun, tantangan utama yang perlu ditangani adalah bagaimana mengembangkan model klasifikasi NLP dengan ambang ketepatan yang tinggi. Ini mencakup berbagai aspek teknis seperti penggunaan algoritma yang tepat, pemilihan dataset representatif, dan iterasi model dan evaluasi untuk mencapai kinerja optimal dan akan memungkinkan untuk meningkatkan kecepatan dan akurasi pemrosesan data.

Berdasarkan latar belakang di atas maka penggunaan teknologi *Natural Language Processing (NLP)* diusulkan sebagai solusi untuk meningkatkan akurasi dalam menganalisis dan mengklasifikasikan pengaduan. Meskipun NLP menawarkan banyak keuntungan, tantangan utama terletak pada pengembangan model yang akurat dan efektif, yang membutuhkan pemilihan algoritma yang tepat, pra-pemrosesan data yang cermat, dan evaluasi model yang berkelanjutan.

B. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan di atas, maka adapun rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengembangkan model klasifikasi *Natural Language Processing*
2. Bagaimana mengidentifikasi dan mengklasifikasikan saran dan kritik dengan tingkat akurasi yang baik?

C. Tujuan Penelitian

Dari latar belakang dan rumusan masalah di atas, maka adapun tujuan penelitian sebagai berikut:

1. Mengembangkan model klasifikasi berbasis *Natural Language Processing* untuk menerapkan teknik-teknik optimasi

2. Mencapai tingkat akurasi/ketepatan yang baik dalam mengidentifikasi dan mengklasifikasikan saran dan kritik dari mahasiswa dengan menggunakan model NLP yang telah dikembangkan

D. Manfaat Penelitian

1. Bagi penulis

Penelitian ini akan memberikan kesempatan kepada penulis untuk mengembangkan dan mendapatkan pengetahuan serta keterampilan baru dalam bidang *Natural Language Processing (NLP)* dan klasifikasi teks.

2. Bagi pembaca

Penelitian ini akan mendapatkan wawasan tentang karakteristik pengaduan pelayanan di Fakultas Teknik Universitas Muhammadiyah Makassar dan pembaca akan mendapatkan informasi tentang NLP

3. Bagi akademisi

Penelitian ini diharapkan bermanfaat bagi akademis untuk penggunaan NLP dalam konteks pengaduan pelayanan di lingkungan akademik dan juga bagi mahasiswa yang dapat menggunakan penelitian ini sebagai referensi penelitian yang relevan untuk penelitian selanjutnya.

E. Ruang Lingkup Penelitian

1. Penelitian ini akan fokus pada pengumpulan data pengaduan pelayanan dari Fakultas Teknik Universitas Muhammadiyah Makassar.
2. Data pengaduan penelitian ini di ambil dari sistem informasi akademik (SIMAK) fakultas
3. Evaluasi model pada penelitian ini menggunakan *confosion matrix* seperti akurasi, *recall*, *precision* dan *F1-Score*

F. Sistematika Penelitian

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang masalah, rumusan masalah, tujuan, manfaat dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan tentang teori-teori yang melandasi penulisan dalam melaksanakan skripsi.

BAB III METODE PENELITIAN

Membahas tentang metode penelitian dan alat yang digunakan untuk pembuatan sistem.

BAB IV HASIL DAN PEMBAHASAN

Pada bab IV berisi hasil analisis pengolahan data serta pembahasan mengenai pengimplementasian sistem yang telah dibangun.

BAB V PENUTUP

Pada bab ini berisi tentang penjelasan kesimpulan dan saran akhir dari sebuah penelitian yang dilakukan.



BAB II

TINJAUAN PUSTAKA

A. Landasan Teori

Berikut adalah beberapa landasan teori yang relevan untuk penelitian ini:

1. Klasifikasi

Tujuan klasifikasi adalah untuk memprediksi kelas data uji yang belum diketahui. Selain itu, klasifikasi adalah proses mengelompokkan sesuatu ke dalam kelas sesuai dengan aturan atau standar yang telah ditentukan. Klasifikasi terdiri dari dua tahap: tahap pelatihan atau instruksi dan tahap pengujian. (Anam & Indriati, 2020)

2. Universitas

Universitas adalah lembaga pendidikan tinggi yang menawarkan program dan gelar dalam berbagai bidang ilmu. Universitas biasanya memiliki berbagai fakultas atau departemen yang mencakup bidang seperti sains, seni, humaniora, teknik, kedokteran, hukum, bisnis, dan lain-lain. Secara umum universitas adalah lembaga pendidikan tinggi yang berkontribusi secara signifikan pada kemajuan ilmu pengetahuan, teknologi, budaya, dan masyarakat melalui pendidikan, penelitian, dan pengabdian kepada masyarakat.

3. Pengaduan Pelayanan

Pengaduan adalah salah satu jenis informasi yang tersedia di internet yang semakin meningkat seiring dengan jumlah orang yang melakukannya. Beberapa kategori pengaduan pelayanan termasuk kritik, saran, pertanyaan, pelanggaran prosedur dan mekanisme, penyimpangan dana, intervensi negatif, perubahan kebijakan, kode etik atau kinerja pelaku, dan Force Majeur. Pelayanan yang berorientasi pada kepuasan pelanggan dan memiliki standar prima (*service excellent*) dapat menciptakan citra positif dari universitas yang diakui. Oleh karena itu, tujuan dari penelitian ini adalah untuk menentukan

apakah kualitas pelayanan terhadap kepuasan mahasiswa dipengaruhi oleh wujud fisik (*tangibles*), kehandalan (*reliability*), daya tanggap (*responsiveness*), jaminan (*assurance*), dan kepedulian (*emphaty*).

Pengaduan mahasiswa merupakan sebuah proses untuk menyampaikan informasi ataupun keluhan-keluhan yang dirasakan oleh mahasiswa dan disampaikan oleh mahasiswa terhadap pelayanan akademis kampus yang kinerjanya kurang memuaskan. Pengaduan mahasiswa sangat dibutuhkan pada sebuah universitas untuk memperbaiki dan meningkatkan kinerja dan kualitas universitas tersebut (Septian, 2021)

4. *Natural Language Processing* (NLP)

Cabang kecerdasan buatan (AI) yang dikenal sebagai pemrosesan bahasa alami (NLP) berfokus pada cara komputer dan manusia berinteraksi menggunakan bahasa alami. Tujuan utama NLP adalah untuk memungkinkan komputer memahami, menafsirkan, dan menghasilkan bahasa manusia dengan cara yang menguntungkan. Proses pemrosesan bahasa alami (NLP) menggunakan ilmu komputer, linguistik, statistik, dan pembelajaran mesin untuk membuat model dan algoritma yang dapat memproses dan memahami bahasa alami. Mesin pencari, analisis media sosial, terjemahan otomatis, asisten virtual, dan banyak lagi adalah aplikasi *natural language processing* (NLP) (Chandra et al., 2020)

Dalam penelitian ini, kami menyelidiki bagaimana pengolahan bahasa manusia (NLP) digunakan untuk memproses data besar secara otomatis, serta mengurai dan memahami struktur bahasa manusia dalam teks. Metode seperti tokenisasi, penghapusan kata berhenti, stemming atau lemmatisasi, dan ekstraksi fitur digunakan untuk menyediakan data teks untuk analisis lebih lanjut. (Surya, 2023).

5. *Python*

Python adalah salah satu bahasa pemrograman yang paling umum digunakan oleh programmer atau pembuat program untuk membuat

program. Ini karena karakteristik sintaksnya yang tidak terlalu rumit, yang membuatnya menjadi salah satu bahasa pemrograman yang sangat mudah digunakan. Terdapat beberapa aturan yang harus dipenuhi saat menulis kode program menggunakan bahasa pemrograman Python. Hal ini dilakukan untuk mencegah kesalahan atau masalah dengan program yang telah dibuat sebelumnya. Penulisan statement atau perintah adalah aturan sintaks *Python* yang pertama (Pasek et al., 2024)

6. *Term Frequency-Inverse Document Frequency* (TF-IDF)

Pembobotan kata digunakan untuk menilai keterhubungan antara kata-kata tersebut. Metode yang digunakan adalah *TF-IDF*, atau frekuensi kata-balik dokumen, yang memberikan bobot setiap kata untuk menentukan seberapa dekat keterhubungan antara kata dan dokumen tersebut. Sebelum pemodelan data dilabelkan. Label positif dan negatif digunakan untuk label dan dilakukan secara manual. Selanjutnya, dalam setiap dokumen, gunakan huruf pembobotan atau berat. Dalam pembobotan ini, *TF-IDF* (*Term Frequency-Inverse Document Frequency*) digunakan. Pembobotan ini digunakan sebagai ukuran statistik untuk mengukur kepentingan kata (Wati & Ernawati, 2021)

7. *Naïve Bayes*

Klasifikasi *Naïve Bayes* adalah algoritma klasifikasi dua tahap yang paling user-friendly dan banyak digunakan. Ini adalah teknik kategorisasi teks sederhana yang memetakan dokumen ke dokumen yang terkait dengan kategori seperti spam atau dokumen asli dan positif, negatif, atau netral. (Puspasari & Rusmin, 2022)

Teori Bayes sendiri dikembangkan oleh seorang ilmuwan Inggris bernama Thomas Bayes. Dalam melakukan analisis sentiment, banyak yang menggunakan algoritma *Naïve Bayes* sebagai metode mereka. Keuntungan dari algoritma *Naïve Bayes* adalah efisiensi yang dapat mengurangi proses analisis sentiment menjadi satu langkah. Selain itu,

algoritma *Naïve Bayes* memiliki akurasi yang relative tinggi bahkan dengan data yang agak lama. (Tanggraeni & Sitokdana, 2022)

B. Penelitian Terkait

Penelitian terdahulu bertujuan untuk mendapatkan bahan perbandingan dan acuan dalam melakukan penelitian. Selain itu, untuk menghindari anggapan kesamaan dalam penelitian ini. Penelitian terdahulu yang menjadi rujukan dalam penelitian ini :

Penelitian yang dilakukan oleh (Alkaff et al., 2021), “Klasifikasi Laporan Keluhan Pelayanan Publik Berdasarkan Instansi Menggunakan Metode LDA-SVM” Performa model LDA-SVM diukur dengan menggunakan confusion matrix dengan menghitung nilai akurasi, presisi, *recall*, dan F1 Score. Hasil pengujian menggunakan teknik split train-test dengan skor 70:30 menunjukkan bahwa model menghasilkan kinerja yang baik dengan akurasi 79,85%, presisi 79,98%, *recall* 72,37%, dan Skor F1 74,67%.

Penelitian yang dilakukan oleh (Mulyono & Sumijan, 2021) yang berjudul “Identifikasi Chatbot Dalam Meningkatkan Pelayanan Online Menggunakan Metode Ntural Language Processing” Hasil pengujian menggunakan kuesioner yang ditujukan kepada 227 mahasiswa mendapatkan nilai 3,55 dengan predikat sangat baik. Kemudian melakukan pengujian menggunakan 40 data pertanyaan dan jawaban. Maka, didapatkan 37 jawaban sesuai dan 3 jawaban tidak sesuai dengan persentase akurasi jawaban yang dihasilkan dari chatbot adalah sebesar 92,5 persen. Hasil pengujian ini telah mampu merespon pertanyaan-pertanyaan yang diajukan mahasiswa. Chatbot ini dapat memudahkan mahasiswa mendapatkan informasi dengan tingkat akurasi yang sangat baik.

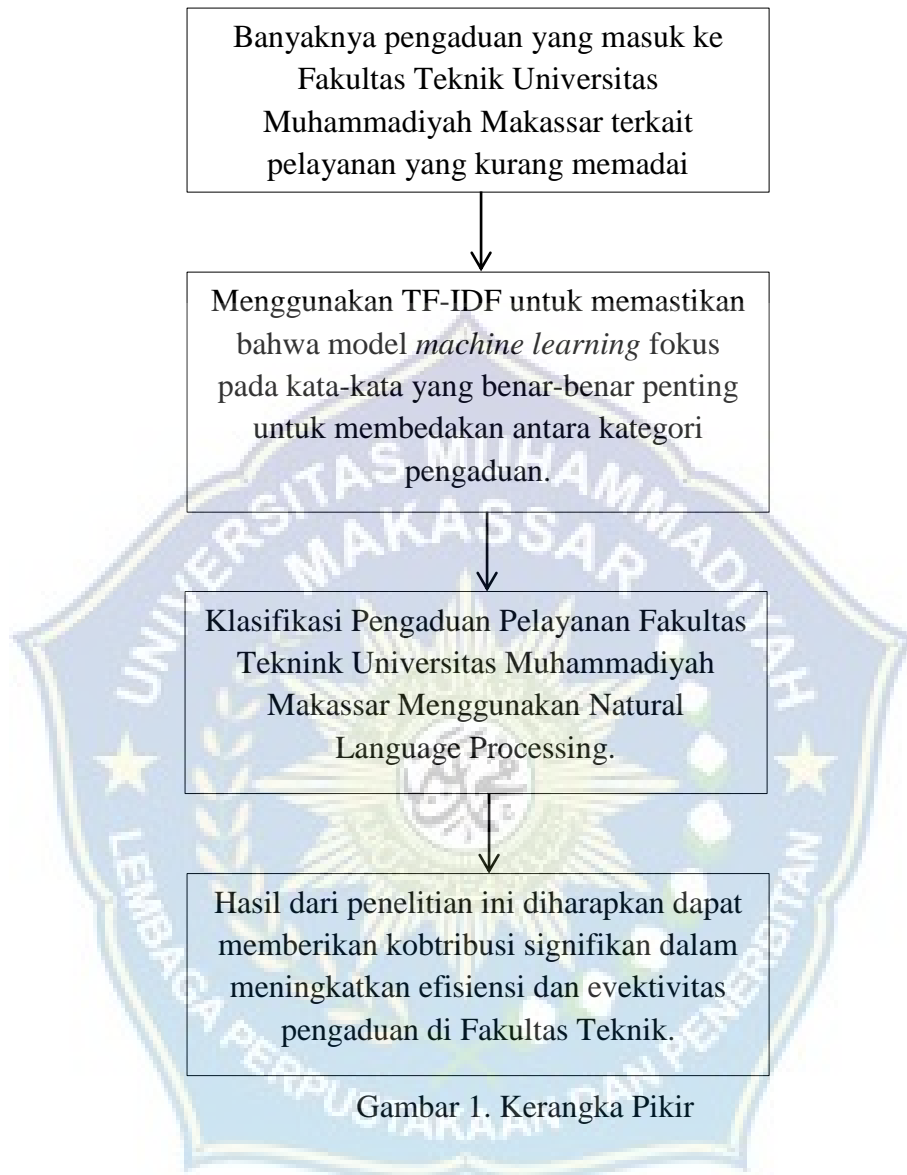
Penelitian yang dilakukan oleh (Nuzulia, 2020) yang berjudul “Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi

Menggunakan Algoritma *Naïve Bayes*” Data yang digunakan pada penelitian ini adalah 213 mahasiswa dari beberapa perguruan tinggi di Riau. Hasil pengujian klasifikasi menggunakan *naïve bayes* adalah *accuracy* 96,24%, *precision* 93,14 %, dan *recall* 98,96%.

Penelitian yang dilakukan oleh (Nalendro, 2019) yang berjudul “Sistem Informasi Pengaduan Layanan Universitas” Penelitian ini telah berhasil membangun sistem informasi pengaduan layanan universitas dengan menggunakan bahasa pemrograman Java, PHP, Javascript, dan HTML, serta Firebase sebagai basis data. Hasil akhir yang didapatkan pada persamaan Kuesioner adalah 88.34% yang mana berada pada interval skor sangat setuju(SS). Hasil akhir pada persamaan Kuesioner menunjukkan bahwa fitur-fitur pada sistem informasi pengaduan layanan universitas berjalan dengan baik.

Penelitian yang dilakukan oleh (Yuliana et al., 2019), “Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network” Hal ini terlihat dari hasil eksperimen, yaitu Penggunaan algoritma Neural Network pada proses klasifikasi menghasilkan akurasi yang tinggi yaitu sebesar 43,00% dengan jangka waktu 03 jam 45 menit 14 detik dalam mengklasifikasikan dokumen teks berbahasa Indonesia pada teks pengaduan masyarakat.

C. Kerangka Pikir



BAB III

METODE PENELITIAN

A. Tempat Dan Waktu Penelitian

Penelitian ini dilakukan di Fakultas Teknik Universitas Muhammadiyah Makassar, yang berlokasi Alamat: Jl. Sultan Alauddin No.259, Gunung Sari, Rappocini, Kota Makassar, Sulawesi Selatan 90221, Indonesia. Objek penelitian sendiri yaitu Fakultas Teknik Universitas Muhammadiyah Makassar sebagai tempat pengumpulan data pengaduan dari berbagai sumber dengan berbagai pihak terkait untuk mendapatkan wawasan tambahan. Penelitian ini akan dilakukan pada bulan Juli sampai dengan Agustus 2024.

B. Alat Dan Bahan

Untuk melaksanakan penelitian diperlukan alat dan bahan yang akan digunakan pada penelitian ini yaitu :

1. Laptop
2. Data Pengaduan
3. *Python*
4. *Visual Studio Code*
5. *Natural Language Toolkit (NLTK)*

C. Perancangan Sistem

Untuk memudahkan proses perancangan penelitian ini menggunakan *flowchart* perancangan sistem sebagai berikut :

1. Mulai: Ini merupakan langkah awal dalam melakukan penelitian, serta menandakan bahwa proses perancangan sistem dimulai.
2. Pengumpulan data: Pada tahap ini, data pengaduan pelayanan dari Fakultas Teknik Universitas Muhammadiyah Makassar data ini diambil dari sumber data dari sistem informasi akademik (SIMAK)

fakultas yang masuk

3. Ekstraksi Fitur TF-IDF: Pada tahap ini, pembobotan/fitur teks pengaduan diekstraksi menggunakan teknik *TF-IDF*. *TF-IDF* digunakan untuk mengurangi makna suatu kata dalam dokumen relatif sehubungan dengan seluruh koleksi dokumen, sehingga hasil pengukuran dapat digunakan untuk klasifikasi.
4. *Pre-processing*: Sebelum diproses ke dalam pembentukan model klasifikasi menggunakan *naïve bayes*, dilakukan tahap *pre-processing* tahap ini dimana model *Natural Language Processing* digunakan terhadap teks di dalam set data. Aturan ini termasuk menghapus *stop words*, *tokenisasi* dan *stemming/lemmatization* dan lain-lain.

Setelah *pre-processing* data dibagi menjadi dua utama yaitu data *training* dan *data testing*.

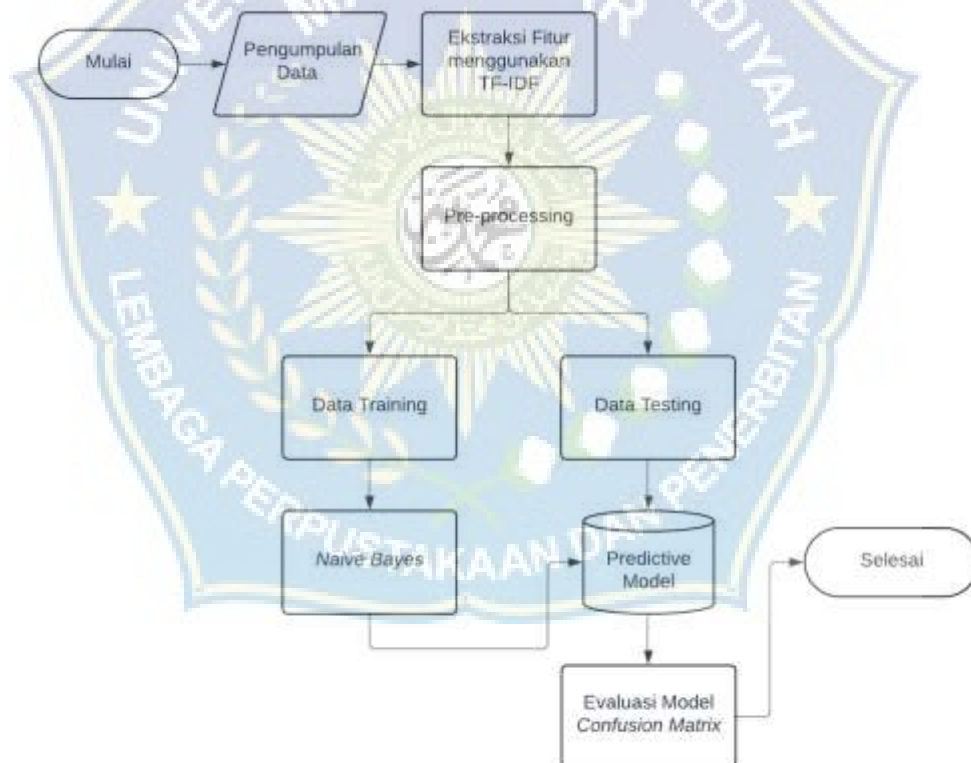
- a. *Data training* memungkinkan model untuk belajar dan mengenali pola, sementara
- b. *Data testing* digunakan untuk memastikan bahwa model tersebut memiliki kinerja yang baik dan dapat diandalkan saat digunakan pada data yang belum pernah dilihat.

Dalam konteks penelitian ini, tujuan akhirnya adalah untuk mengembangkan model yang dapat mengklasifikasi pengaduan pelayanan secara akurat.

5. *Naïve Bayes*: Pada tahap ini, model *naïve bayes* digunakan untuk klasifikasi. Model ini menggunakan teori probabilitas untuk membuat prediksi berdasarkan parameter skeletal yang telah ditentukan sebelumnya.
6. Predictive Model: Langkah predictive model mengacu pada langkah dimana model machine learning telah selesai menganalisis data saat ini dan digunakan untuk membuat prediksi atau klasifikasi pada data

baru.

7. Evaluasi Model *Confusion Matrix*: pada tahap ini digunakan untuk menghitung dari model klasifikasi. Matriks ini membandingkan prediksi model dengan label aktualnya dalam pengujian data dan memberikan wawasan tentang jenis kesalahan yang dialami model.
8. Selesai: Proses penelitian dan pengembangan sistem klasifikasi pengaduan pelayanan selesai. Hasil akhir disajikan dan dapat digunakan oleh fakultas teknik universitas muhammadiyah Makassar untuk memperbaiki layanan berdasarkan masukan dari pengaduan yang terklasifikasi.



Gambar 2. Perancangan Sistem

D. Teknik Pengujian Sistem

Tahapan pada penelitian ini yaitu menentukan nilai akurasi, *recall*,

precision dan *F1-Score*. Sebelum menghitung nilai tersebut, data dibagi menjadi dua yaitu data *training* dan *data testing*. *Data training* digunakan sebagai data model sedangkan *data testing* digunakan sebagai uji coba model yang akan distimulasi penggunaan model dunia nyata. Untuk menentukan hasil klasifikasi, penelitian ini menggunakan confusion matrix untuk menganalisis setiap data unik dalam kelas. Dalam klasifikasi confusion matrix, data diklasifikasikan sebagai berikut: benar positif dan benar negatif jika data memiliki nilai benar. Sebaliknya, data dengan metrik signifikansi tunduk pada hasil positif palsu dan negatif palsu.

Confusion matrix adalah table yang digunakan untuk mengevaluasi kinerja model klasifikasi:

Tabel 1. Confusion Matriks

Data	Prediksi positive	Prediksi Negative
Aktual Positive	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
Aktual Negative	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Berdasarkan pada Tabel 3.2, berikut penjelasan setiap actual dalam confusion matrix:

1. *True Positive* (TP) adalah data jumlah prediksi yang benar untuk kelas positif
2. *False Positive* (FP) adalah data jumlah prediksi yang salah untuk kelas positif (contoh sebenarnya negatif, tetapi diprediksi positif)
3. *False Negative* (FN) adalah data jumlah prediksi yang salah untuk kelas negative (contoh sebenarnya positif, tetapi diprediksi negatif)
4. *True Negative* (TN) adalah data jumlah prediksi untuk kelas negative

Berikut adalah rumus menghitung klasifikasi menggunakan confusion

matrix:

Akurasi adalah proporsi prediksi yang benar (baik positif maupun negative) dari semua prediksi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (1)$$

Recall adalah beberapa model yang mampu mengidentifikasi setiap kasus positif dalam dataset. Artinya berdasarkan semua contoh yang jelas, berapa banyak prediksi sukses yang dapat dibuat dengan pasti.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (2)$$

Precision merupakan mengukur akurasi dari prediksi positif model. Jadi, dari semua contoh yang diprediksi sebagai positif, berapa banyak yang benar-benar positif.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (3)$$

F1-Score adalah rata-rata harmonis dari *precision* dan *recall*, yang memberikan keseimbangan antara keduanya.

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \dots\dots\dots (4)$$

F1-Score memberikan nilai yang lebih rendah jika salah satu dari *precision* atau *recall* rendah, menjadikannya metrik yang baik untuk dataset yang tidak seimbang.

E. Teknik Analisis Data

Proses atau prosedur yang digunakan untuk memeriksa, dan mengevaluasi data untuk menghasilkan kesimpulan atau membuat keputusan berdasarkan data dikenal sebagai teknik analisis data. Berikut adalah beberapa proses analisis data yang digunakan dalam penelitian ini:

1. Pengumpulan Data

Pada penelitian ini metode pengumpulan data yang digunakan merupakan data dari sistem informasi akademik (SIMAK) fakultas akan dimanfaatkan untuk keperluan analisis kinerja mahasiswa dan

efektivitas kurikulum. Data diperoleh dari responden yang terlibat dalam penelitian (mahasiswa).

2. *Pre-processing Data*

Text pre-processing ini merupakan tahapan awal dalam pengolahan text. Pembersihan data, *tokenisasi* dan *stemming/lemmatization* adalah operasi pre-processing data text umum di sebagian besar aplikasi NLP.

- a. Pembersihan Data: terdiri dari penghapusan dan/atau koreksi catatan yang salah dari kumpulan data.
- b. *Tokenisasi*: untuk mendeteksi dan memisahkan kata-kata dengan menghilangkan komponen tambahan seperti tanda baca.
- c. *Stemming/lemmatization*: mengurangi kata dalam bentuk dasarnya untuk mengurangi variasi dan meningkatkan akurasi analisis.

3. Natural Language Processing

Natural Language Processing adalah cabang dari kecerdasan buatan yang mempelajari komunikasi manusia-komputer melalui bahasa alami. Langkah-langkah dalam proses ini adalah mengidentifikasi niat, mengukur input, dan melaporkan hasilnya sesuai dengan input.

BAB IV

HASIL DAN PEMBAHASAN

A. Hasil Pra-pemrosesan Data

Pada tahap ini, keluhan-keluhan yang telah dikumpulkan dari mahasiswa diproses menggunakan teknik *Natural Language Processing* (NLP). NLP membantu dalam membersihkan dan menyederhanakan teks sehingga lebih mudah dianalisis oleh model pembelajaran mesin. Langkah-langkah yang diambil meliputi:

1. Penghapusan Tanda Baca

Semua tanda baca dihapus untuk memastikan teks bersih dari karakter yang tidak diperlukan. Hal ini penting untuk menghindari gangguan dalam analisis teks. Penghapusan tanda baca dilakukan dengan menggunakan fungsi `str.translate` untuk menghapus semua tanda baca dari teks.

2. Tokenisasi

Setiap kalimat dibagi menjadi kata-kata individu. *Tokenisasi* memecah teks menjadi unit-unit yang lebih kecil yang dapat dianalisis secara individual. Pada implementasinya, *tokenisasi* dilakukan menggunakan `word_tokenize` dari pustaka NLTK, yang memecah teks menjadi daftar kata.

3. Penghapusan *Stop Words*

Kata-kata umum yang tidak memberikan banyak informasi (seperti "dan", "di", "yang") dihapus dari teks. Penghapusan ini membantu mengurangi dimensi data dan fokus pada kata-kata yang lebih informatif. Pada source code, penghapusan *stop words* dilakukan dengan mengecualikan kata-kata yang termasuk dalam daftar `stopwords.words('indonesian')` dari NLTK.

4. Stemming dan Lemmatization

Kata-kata dikembalikan ke bentuk dasarnya untuk konsistensi. *Stemming* mengurangi kata ke bentuk dasarnya, sementara lemmatization mengubah kata ke bentuk dasar yang bermakna secara linguistik. Dalam kode, PorterStemmer digunakan untuk stemming, dan WordNetLemmatizer digunakan untuk lemmatization, memastikan setiap kata diproses ke bentuk dasarnya sebelum dianalisis lebih lanjut.

10 contoh data mentah yang diambil dari dataset menunjukkan beragam keluhan mahasiswa dari berbagai kategori. Berikut adalah beberapa contoh data mentah:

Tabel 2. Dataset Awal

No	Content
1	Kedepanya kami bisa diberikan dosen yang lebih produktif dan mampu beoerja sama dengan mahasiswa menyelesaikan permasalahan dalam pertemuan covid ini. Tapi untuk dosen saat ini sudah luar biasa kok
2	Saya berharap pelayanan akademik mampu memberikan pelayanan yang baik kepada mahasiswa dan seluruh masyarakat Unismuh. Terutama untuk para bapak/ibu dosen sekiranya mampu memberikan metode yang baik bagi mahasiswa sehingga mahasiswa lebih aktif dan produktif dalam proses pembelajaran.
3	Sebaiknya pihak kampus memantau proses penyaluran kouta internet , pihak kampus juga harus membuat rancangan kesepakatan perkuliahan online yang disetujui dosen dan mahasiswa serta meningkatkan kualitas layanan akedemik seperti portal online untuk kebutuhan kuliah online
4	Perlunya efesiensi waktu yang baik dalam pembelajaran agar matakuliah yabg lain tidak terganggu
5	Apakah pelayanan akademik sudah memberikan pelayanan yang terbaik kepada mahasisswa ?

-
- 6 Saran saya agar proses pembelajaran secara online harus juga di selingi dengan pembelajaran secara offline agar para mahasiswa lebih mengerti lagi materi perkuliahan
 - 7 "Tetap pertahankan pembelajaran yang baik dan tingkat terus-menerus kedepannya"
 - 8 Sehubungan dengan pembelajaran daring seperti sekarang ini karena pandemi mohon kerjasama dosen agar tidak susah dihubungi.
 - 9 Disiplin waktu perkuliahan
 - 10 Saya cukup puas dengan pembelajaran atau ilmu yg di berikan tetapi dalam masa pandemi ini apa masih boleh ada pengurangan bpp?,mungkin itu dari saya pribadi terimah kasih
-

Setelah melakukan pra-pemrosesan, keluhan-keluhan tersebut diproses menjadi bentuk yang lebih sederhana dan konsisten. Berikut adalah 10 contoh data setelah pra-pemrosesan:

Tabel 3. Dataset Setelah Diproses

No	Content
1	Agar kedepannya buku referensi di perpustakaan
2	selalu menjadi prodi yang didambakan banyak
3	Fasilitasi link zoom dalam perkuliahan online
4	Perkuliahan masih sedapatnya online zoom
5	yak semoga makin bagus pelayanan di bagian
6	Untuk pengajaran,ketepatan dosen dalam
7	Analisis bakat dan minat mahasiswa perlu di
8	Khususnya prodi agribisnis dan yang perlu di tingkatkan

9 Ada beberapa yang jadi bahan pertimbangan

10 semoga ada aplikasi untuk para mahasiswa

B. Hasil Ekstraksi Fitur Menggunakan TF-IDF

Setelah teks keluhan diproses, fitur-fitur penting diekstraksi menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF memberikan bobot pada setiap kata berdasarkan seberapa sering kata tersebut muncul di dokumen dan seberapa unik kata tersebut dalam keseluruhan korpus dokumen. Hal ini membantu dalam mengidentifikasi kata-kata yang signifikan untuk setiap kategori keluhan.

	aamiin	ac	acaramaupun	adabaik	adakan	adestik	adil	admir	wifi	workshop	ar	ya	yah	yg	zaman	zoom
0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
1	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
2	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.617061
3	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
4	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
5	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
6	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.269409	0.000000
7	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
8	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
9	0.0	0.0	0.0	0.483484	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000

Gambar 3. Hasil Ekstraksi Fitur Menggunakan TF-IDF

Hasil ekstraksi fitur menggunakan TF-IDF menunjukkan bahwa setiap kata atau token dalam dokumen keluhan mahasiswa diberikan bobot yang mencerminkan signifikansinya. Kolom-kolom seperti "aamiin," "ac," "acaramaupun," "adabaik," "adakan," dan seterusnya, mewakili kata-kata unik yang muncul setelah teks diproses. Nilai-nilai dalam sel-sel DataFrame menunjukkan bobot TF-IDF dari kata tersebut dalam dokumen tertentu. Misalnya, kata "zoom" pada baris ketiga memiliki nilai TF-IDF sebesar 0.617061, menunjukkan bahwa kata ini sangat signifikan dalam dokumen tersebut. Sebaliknya, nilai 0.0 di banyak sel menunjukkan bahwa kata-kata tersebut tidak muncul dalam dokumen terkait, atau kemunculannya tidak dianggap penting. Kata-kata dengan nilai TF-IDF

tinggi, seperti "zoom" dan "adabaik," menandakan bahwa kata-kata ini memainkan peran penting dalam konteks dokumen-dokumen tertentu, membantu mengidentifikasi topik atau tema utama yang dibahas. Hasil ini juga menyoroti pentingnya proses pembersihan teks, di mana hanya kata-kata yang memiliki makna dalam konteks dokumen yang dipertahankan, sementara fitur yang tidak signifikan diabaikan. Secara keseluruhan, nilai TF-IDF yang tinggi memberikan wawasan mendalam tentang kata-kata yang paling relevan dan signifikan dalam keluhan yang diajukan oleh mahasiswa.

C. Pembagian Data

Data yang telah diproses dibagi menjadi data latih (*training data*) dan data uji (*testing data*) dengan proporsi 80:20. Pembagian data ini penting untuk memastikan bahwa model yang dilatih dapat dievaluasi secara objektif pada data yang tidak dilihat selama pelatihan. Jumlah data latih dan data uji adalah sebagai berikut:

- Jumlah Data Latih: 80
- Jumlah Data Uji: 20

D. Kategorisasi Keluhan

Pada tahap ini, keluhan yang telah diproses dikategorikan ke dalam berbagai tema atau topik yang relevan. Untuk melakukan hal ini, sebuah fungsi dikembangkan untuk menentukan kategori keluhan berdasarkan kata kunci yang ditemukan dalam teks. Fungsi ini bekerja dengan cara memeriksa keberadaan kata-kata kunci spesifik dalam teks keluhan, dan mengelompokkan keluhan tersebut ke dalam salah satu dari beberapa kategori yang telah ditentukan.

Fungsi ini pertama-tama mengonversi teks menjadi huruf kecil untuk memastikan konsistensi dalam pencocokan kata kunci. Kemudian, fungsi ini mencari keberadaan kata kunci seperti "prodi," "fakultas," "tata usaha," "simak," dan lain-lain dalam teks. Berdasarkan kata kunci yang

ditemukan, keluhan akan dikategorikan ke dalam kategori yang sesuai, seperti:

- Prodi
- Tata Usaha/Pelayanan Administrasi
- SIMAK
- Fasilitas Kampus
- Kualitas Pengajaran
- Kurikulum dan Materi
- Perpustakaan
- Keuangan dan Pembayaran
- Fasilitas Parkir

Jika teks keluhan mengandung kata kunci yang sesuai dengan salah satu kategori di atas, maka keluhan tersebut akan diberi label yang relevan. Jika tidak ada kata kunci yang cocok ditemukan, teks keluhan tidak akan dikategorikan dan diberi nilai None.

Dengan pendekatan ini, proses kategorisasi menjadi lebih sistematis dan memungkinkan analisis lebih lanjut untuk fokus pada topik-topik tertentu yang sering muncul dalam keluhan mahasiswa. Fungsi ini sangat membantu dalam mengidentifikasi pola keluhan yang berulang, yang dapat menjadi dasar untuk peningkatan layanan dan fasilitas di universitas.

E. Identifikasi Saran dan Kritik

Dalam analisis keluhan, salah satu langkah penting adalah mengidentifikasi apakah keluhan tersebut bersifat saran atau kritik. Untuk melakukan ini, daftar kata-kata saran dan kritik telah disusun sebagai referensi. Kata-kata dalam daftar ini digunakan untuk mencocokkan teks keluhan, sehingga dapat diklasifikasikan apakah keluhan tersebut mengandung saran, kritik, atau keduanya.

1. Daftar Kata Saran

Kata-kata dalam daftar saran sering digunakan dalam konteks memberikan rekomendasi, usulan, atau perbaikan yang konstruktif. Contohnya termasuk kata-kata seperti "harus," "sering," "sebaiknya," "perlu," "meningkatkan," "memfasilitasi," "mengoptimalkan," "mendorong," "mengembangkan," "keren," dan lainnya. Implementasinya dalam analisis keluhan adalah, ketika sebuah teks keluhan mengandung salah satu dari kata-kata ini, teks tersebut diberi label sebagai `Is_Saran`, yang menunjukkan bahwa keluhan tersebut berisi saran atau rekomendasi untuk perbaikan.

2. Daftar Kata Kritik

Kata-kata dalam daftar kritik biasanya digunakan untuk mengungkapkan ketidakpuasan, keluhan, atau menunjukkan adanya masalah. Contohnya termasuk kata-kata seperti "lambat," "buruk," "tidak memadai," "kurang," "mengecewakan," "tidak konsisten," "tidak profesional," "tidak ramah," "menjengkelkan," "mengganggu," "membahayakan," dan lain-lain. Implementasinya dalam analisis keluhan adalah, ketika sebuah teks keluhan mengandung salah satu dari kata-kata ini, teks tersebut diberi label sebagai `Is_Kritik`, yang menunjukkan bahwa keluhan tersebut berisi kritik atau ketidakpuasan terhadap suatu aspek.

3. Penerapan dalam Klasifikasi:

Daftar kata-kata saran dan kritik ini digunakan dalam proses klasifikasi untuk memberikan label tambahan pada teks keluhan, sehingga memudahkan analisis lebih lanjut. Misalnya, sebuah keluhan yang mengandung kata "meningkatkan" akan diidentifikasi sebagai saran, sementara keluhan yang mengandung kata "lambat" akan diidentifikasi sebagai kritik.

Dengan pendekatan ini, analisis keluhan dapat menjadi lebih terstruktur, memungkinkan pengelompokan keluhan berdasarkan sifatnya, apakah mereka memberikan saran untuk perbaikan atau mengkritik aspek tertentu dari layanan. Hal ini juga memudahkan

identifikasi area-area yang memerlukan perhatian atau perbaikan lebih lanjut.

F. Hasil *Naïve Bayes* dan NPL

Model *Naïve Bayes* dilatih menggunakan data latih yang telah diproses dan fitur yang diekstraksi. *Naïve Bayes* adalah salah satu algoritma yang sering digunakan untuk klasifikasi teks karena kesederhanaannya dan efisiensinya. Model ini mempelajari pola-pola dalam data latih dan menggunakannya untuk memprediksi kategori keluhan pada data uji.

Akurasi: 0.91

Confusion Matrix:

[101	0	0	5	0	9	11	1	0]
[0	57	0	0	0	0	2	0	0]
[1	0	10	6	2	1	6	0	0]
[0	0	0	307	0	0	3	0	0]
[0	0	0	13	102	1	0	0	0]
[0	0	0	0	0	145	1	0	0]
[0	1	0	8	0	1	250	2	0]
[0	0	0	6	1	0	5	71	0]
[0	0	0	5	0	0	8	0	3]]

Gambar 4. Hasil *Confusion Matrix*

Model *Naïve Bayes* dilatih menggunakan data latih yang telah diproses dan fitur yang diekstraksi menggunakan metode Natural Language Processing (NLP). *Naïve Bayes* dipilih karena kesederhanaan dan efisiensinya dalam menangani tugas klasifikasi teks. Model ini mampu mempelajari pola-pola dalam data latih dan menggunakan pola-pola tersebut untuk memprediksi kategori keluhan pada data uji.

Hasil pelatihan dan pengujian model menunjukkan akurasi sebesar 0.91, yang berarti model mampu mengklasifikasikan keluhan dengan tingkat keberhasilan 91%. *Confusion Matrix* menunjukkan distribusi

prediksi benar (True Positives) dan kesalahan (False Positives dan False Negatives) untuk setiap kategori.

Dalam laporan klasifikasi, terlihat bahwa kategori seperti "Fasilitas Kampus," "Fasilitas Parkir," dan "Perpustakaan" memiliki nilai *precision*, *recall*, dan *f1-score* yang tinggi, menunjukkan bahwa model dapat dengan baik mengenali keluhan-keluhan yang terkait dengan kategori-kategori tersebut. Misalnya, kategori "Fasilitas Kampus" memiliki *f1-score* 0.88, menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Namun, untuk kategori "Keuangan dan Pembayaran" dan "Tata Usaha/Pelayanan Administrasi," model masih memiliki kelemahan dengan *recall* yang rendah, yang berarti beberapa keluhan di kategori ini tidak dikenali dengan baik oleh model.

Classification Report:

	precision	recall	f1-score	support
Fasilitas Kampus	0.99	0.80	0.88	127
Fasilitas Parkir	0.98	0.97	0.97	59
Keuangan dan Pembayaran	1.00	0.38	0.56	26
Kualitas Pengajaran	0.88	0.99	0.93	310
Kurikulum dan Materi	0.97	0.88	0.92	116
Perpustakaan	0.92	0.99	0.96	146
Prodi	0.87	0.95	0.91	262
SIMAK	0.96	0.86	0.90	83
Tata Usaha/Pelayanan Administrasi	1.00	0.19	0.32	16
accuracy			0.91	1145
macro avg	0.95	0.78	0.82	1145
weighted avg	0.92	0.91	0.91	1145

Gambar 5. Hasil Klasifikasi

Secara keseluruhan, model menunjukkan performa yang cukup baik dengan macro average *f1-score* 0.82, meskipun ada ruang untuk peningkatan khususnya pada kategori-kategori dengan representasi data yang lebih rendah dalam data latih. Evaluasi ini membantu dalam

memahami seberapa baik model dapat memprediksi kategori keluhan dan memberikan wawasan mengenai area yang perlu ditingkatkan lebih lanjut.

G. Hasil Perhitungan Manual

Untuk menghitung metrik klasifikasi secara manual berdasarkan data yang ada dalam confusion matrix, kita bisa mengikuti langkah-langkah berikut:

```

[[101  0  0  5  0  9  11  1  0] # Fasilitas Kampus
 [  0 57  0  0  0  0  2  0  0] # Fasilitas Parkir
 [  1  0 10  6  2  1  6  0  0] # Keuangan dan Pembayaran
 [  0  0  0 307  0  0  3  0  0] # Kualitas Pengajaran
 [  0  0  0  13 102  1  0  0  0] # Kurikulum dan Materi
 [  0  0  0  0  0 145  1  0  0] # Perpustakaan
 [  0  1  0  8  0  1 250  2  0] # Prodi
 [  0  0  0  6  1  0  5  71  0] # SIMAK
 [  0  0  0  5  0  0  8  0  3] # Tata Usaha/Pelayanan Administrasi
  
```

Gambar 6. Hasil dari Confusion Matrix

Langkah-langkah Perhitungan

1. Fasilitas Kampus

- TP (True Positive): 101 (prediksi benar sebagai Fasilitas Kampus)
- FP (False Positive): Jumlah prediksi salah sebagai Fasilitas Kampus = $0 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 = 1$
- FN (False Negative): Jumlah yang seharusnya Fasilitas Kampus tetapi diprediksi sebagai kategori lain = $5 + 9 + 11 + 1 = 26$
- TN (True Negative): Jumlah prediksi benar sebagai bukan Fasilitas Kampus = Total data - (TP + FP + FN) = $1145 - (101 + 1 + 26) = 1017$
- Accuracy: $(TP + TN) / (TP + TN + FP + FN) = (101 + 1017) / (101 + 1 + 26 + 1017) = 1118 / 1145 = 0.9764$
- Recall: $TP / (TP + FN) = 101 / (101 + 26) = 101 / 127 = 0.7953$
- Precision: $TP / (TP + FP) = 101 / (101 + 1) = 101 / 102 = 0.9902$

- *F1-Score*: $2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) = 2 * (0.9902 * 0.7953) / (0.9902 + 0.7953) = 0.8823$

2. Fasilitas Parkir

- TP: 57
- FP: $0 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0 = 1$
- FN: 0
- TN: $1145 - (57 + 1 + 0) = 1087$
- Accuracy: $(\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) = (57 + 1087) / (57 + 1 + 0 + 1087) = 1144 / 1145 = 0.9991$
- Recall: $\text{TP} / (\text{TP} + \text{FN}) = 57 / (57 + 0) = 57 / 57 = 1.0$
- Precision: $\text{TP} / (\text{TP} + \text{FP}) = 57 / (57 + 1) = 57 / 58 = 0.9828$
- *F1-Score*: $2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) = 2 * (0.9828 * 1.0) / (0.9828 + 1.0) = 0.9913$

3. Keuangan dan Pembayaran

- TP: 10
- FP: $0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$
- FN: $1 + 6 + 2 + 1 + 6 = 16$
- TN: $1145 - (10 + 0 + 16) = 1119$
- Accuracy: $(\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) = (10 + 1119) / (10 + 0 + 16 + 1119) = 1129 / 1145 = 0.9860$
- Recall: $\text{TP} / (\text{TP} + \text{FN}) = 10 / (10 + 16) = 10 / 26 = 0.3846$
- Precision: $\text{TP} / (\text{TP} + \text{FP}) = 10 / (10 + 0) = 10 / 10 = 1.0$
- *F1-Score*: $2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) = 2 * (1.0 * 0.3846) / (1.0 + 0.3846) = 0.5555$

4. Kualitas Pengajaran

- TP: 307
- FP: $5 + 0 + 6 + 13 + 0 + 8 + 6 + 5 + 0 = 43$
- FN: 3
- TN: $1145 - (307 + 43 + 3) = 792$

- Accuracy: $(TP + TN) / (TP + TN + FP + FN) = (307 + 792) / (307 + 43 + 3 + 792) = 1099 / 1145 = 0.9598$
- Recall: $TP / (TP + FN) = 307 / (307 + 3) = 307 / 310 = 0.9903$
- Precision: $TP / (TP + FP) = 307 / (307 + 43) = 307 / 350 = 0.8771$
- *F1-Score*: $2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.8771 * 0.9903) / (0.8771 + 0.9903) = 0.9308$

Setiap kategori dalam *confusion matrix* dapat dianalisis dengan menggunakan rumus-rumus di atas untuk menghasilkan metrik seperti akurasi, *recall*, *precision*, dan *F1-score*. Akurasi keseluruhan sebesar 0.91 menunjukkan bahwa model bekerja cukup baik, tetapi analisis yang lebih mendetail terhadap kategori-kategori tertentu menunjukkan bahwa beberapa kelas mengalami kesulitan dalam hal *recall* dan *precision*. Fasilitas Parkir memiliki *F1-Score* yang sangat tinggi, sementara Keuangan dan Pembayaran memiliki *F1-Score* yang lebih rendah, menunjukkan adanya tantangan dalam menangani kelas dengan jumlah data yang lebih sedikit atau lebih spesifik.



BAB V

PENUTUP

A. Kesimpulan

Hasil evaluasi menunjukkan bahwa model Naïve Bayes memiliki performa yang sangat baik dalam mengklasifikasi keluhan mahasiswa dengan akurasi sebesar 91%. Model ini menunjukkan kemampuan yang andal untuk digunakan dalam sistem pengaduan pelayanan Fakultas Teknik Universitas Muhammadiyah Makassar.

1. Pengembangan Model Klasifikasi Natural Language Processing (NLP)

Penelitian ini berhasil mengembangkan model klasifikasi menggunakan teknik Natural Language Processing (NLP) yang efektif dalam mengklasifikasikan keluhan mahasiswa. Model Naïve Bayes yang digunakan menunjukkan performa yang sangat baik, dengan tingkat akurasi mencapai 91%. Hal ini menunjukkan bahwa model yang dikembangkan mampu menjadi solusi yang andal untuk diterapkan dalam sistem pengaduan di Fakultas Teknik Universitas Muhammadiyah Makassar.

2. Identifikasi dan Klasifikasi Saran dan Kritik

Model yang dibangun mampu mengidentifikasi dan mengklasifikasikan saran dan kritik dengan tingkat akurasi yang baik, berkat penggunaan metode TF-IDF untuk ekstraksi fitur. Meskipun model ini sudah menunjukkan hasil yang memuaskan, terdapat beberapa kategori yang masih memerlukan peningkatan akurasi, seperti kategori "Keuangan dan Pembayaran" serta "Tata Usaha/Pelayanan Administrasi". Dengan demikian, model ini masih dapat dikembangkan lebih lanjut untuk mencapai akurasi yang lebih tinggi di semua kategori.

B. Saran

Untuk penelitian selanjutnya, beberapa saran yang dapat diberikan antara lain:

1. Eksplorasi Algoritma Lain: Mencoba algoritma lain seperti *Support Vector Machine (SVM)* atau *Random Forest* untuk membandingkan performa model.
2. Peningkatan Dataset: Menggunakan dataset yang lebih besar dan lebih beragam untuk meningkatkan generalisasi model.
3. Analisis Sentimen: Menambahkan fitur-fitur lain seperti analisis sentimen untuk memberikan wawasan lebih mendalam tentang keluhan mahasiswa.
4. Penggunaan Teknik Lain: Menerapkan teknik NLP lainnya seperti *word embeddings* (contohnya *Word2Vec* atau *GloVe*) untuk meningkatkan representasi teks.



DAFTAR PUSTAKA

- Alkaff, M., Baskara, A. R., & Maulani, I. (2021). Klasifikasi Laporan Keluhan Pelayanan Publik Berdasarkan Instansi Menggunakan Metode LDA-SVM. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(6), 1265–1276. <https://doi.org/10.25126/jtiik.2021863768>
- Anam, C., & Indriati, M. (2020). Klasifikasi Pengaduan Pelayanan Dispendukcapil Kota Malang Menggunakan Metode Naïve Bayes dan Seleksi Fitur Glasgow-II. ... *Teknologi Informasi Dan Ilmu Komputer E-ISSN*, 4(9), 3264–3271. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/7932/3728>
- Chandra, A. Y., Kurniawan, D., & Musa, R. (2020). Perancangan Chatbot Menggunakan Dialogflow Natural Language Processing (Studi Kasus: Sistem Pemesanan pada Coffee Shop). *Jurnal Media Informatika Budidarma*, 4(1), 208. <https://doi.org/10.30865/mib.v4i1.1505>
- Melani, Y. I. (2019). Sistem Pengaduan Layanan Akademik Menggunakan Responsive Web Design. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 8(1), 39–45. <https://doi.org/10.32736/sisfokom.v8i1.597>
- Muliyono, M., & Sumijan, S. (2021). Identifikasi Chatbot dalam Meningkatkan Pelayanan Online Menggunakan Metode Natural Language Processing. *Jurnal Informatika Ekonomi Bisnis*, 3, 142–147. <https://doi.org/10.37034/infec.v3i4.102>
- Nalendro, G. (2019). Sistem informasi pengaduan layanan universitas. *Sistem Informasi Pengaduan Layanan Universitas*, 49.
- Nuzulia, A. (2020). KLASIFIKASI KEPUASAN MAHASISWA TERHADAP PELAYANAN PERGURUAN TINGGI MENGGUNAKAN ALGORITMA NAÏVE BAYES. *Angewandte Chemie International Edition*, 6(11), 951–952., 3, 5–24.

- Pasek, P., Mahawardana, O., Arya, G., Agus, I. P., & Pratama, E. (2024). *Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap “ Figure Pemimpin ” Menggunakan Python*. 3(1).
- Puspasari, I., & Rusmin, P. H. (2022). Klasifikasi Wazan pada Kata-Kata Al Qur'an Menggunakan Natural Language Processing. *Journal of Technology and Informatics (JoTI)*, 3(2), 41–48. <https://doi.org/10.37802/joti.v3i2.224>
- Septian, A. N. (2021). Implementasi Framework Flutter Untuk Pengaduan Mahasiswa Universitas Xyz. *Inovasi Pembangunan : Jurnal Kelitbangan*, 9(03), 311. <https://doi.org/10.35450/jip.v9i03.273>
- Surya, D. P. (2023). *MENINTEGRASIKAN TEKNIK NLP DAN MACHINE LEARNING*. 3(11), 1–21. <http://teknologipintar.org/index.php/teknologipintar/article/view/527/513>
- Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 785–795. <https://doi.org/10.35957/jatisi.v9i2.1835>
- Usmany, R. (2022). *PENGEMBANGAN CHATBOT PENGADUAN DAN TROUBLESHOOTING TEKNOLOGI INFORMASI DENGAN PENDEKATAN NLP (STUDI KASUS: POLITEKNIK NEGERI AMBON)*. *Jurnal Simetrik*, 12(2), 575–583.
- Wati, R., & Ernawati, S. (2021). Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 06, 240–247. <https://doi.org/10.54367/jtiust.v6i2.1465>
- Yuliana, D., Purwanto, & Supriyanto, C. (2019). Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network. *Jurnal KomtekInfo*, 5(3), 92–116. <https://doi.org/10.35134/komtekinfo.v5i3.35>

LAMPIRAN

```
import pandas as pd

import numpy as np

from sklearn.feature_extraction.text import
    TfidfVectorizer

from sklearn.model_selection import train_test_split

from sklearn.naive_bayes import MultinomialNB

from sklearn.metrics import accuracy_score,
    confusion_matrix, classification_report

from nltk.corpus import stopwords

from nltk.tokenize import word_tokenize

from nltk.stem import PorterStemmer

from nltk.stem.wordnet import WordNetLemmatizer

import string

import nltk

nltk.download('punkt')

nltk.download('stopwords')

nltk.download('wordnet')

# Membaca dataset yang telah dibuat

df = pd.read_csv('saran_sunarti.csv')
```

```

# Menampilkan 10 contoh data mentah (hanya kolom id dan
saran)

print("10 Contoh Data Mentah:")

print(df[['id', 'saran']].head(10))

# Fungsi untuk mengonversi tahun akademik ke format
lebih mudah dibaca
def convert_tahun_akademik(tahun_akademik):
    tahun = str(tahun_akademik)[:4]
    semester = str(tahun_akademik)[-1]
    return f"Tahun {tahun} Semester {semester}"

# Menambahkan kolom baru dengan format tahun akademik
yang lebih mudah dibaca
df['Formatted_Tahun_Akademik'] =
    df['tahun_akademik'].apply(convert_tahun_akademik)

# Daftar kata-kata saran dan kritik
saran_words = [
    "harus", "sering", "sebaiknya", "perlu",
    "disarankan", "meningkatkan",
    "mendukung", "menambah", "memperluas",

```

"mengadakan", "menyediakan",
"mempermudah", "mendapatkan", "memperbaiki",
"mendorong", "mengusulkan",
"memfasilitasi", "mengoptimalkan",
"menyempurnakan", "melengkapi",
"menguatkan", "menyederhanakan", "mendesain ulang",
"memperbarui",
"membina", "melakukan pembaruan",
"mengimplementasikan",
"menekankan", "melatih", "meningkatkan
efektivitas", "mempererat",
"memperbanyak", "membangun", "mempercepat",
"mengintegrasikan",
"menciptakan", "mengembangkan", "membantu",
"mantap", "keren"

]

kritik_words = [

"lambat", "buruk", "tidak memadai", "kurang",
"terlalu sedikit",

"sulit", "tidak nyaman", "membingungkan",
"gangguan", "kesalahan",

"tidak konsisten", "masalah", "tidak tepat",
"memburuk", "menurun",

"mengganggu", "tidak efisien", "membosankan",

```
"mengecewakan",  
  
"tidak memuaskan", "tidak mendukung", "terbatas",  
"terlalu lambat",  
  
"menghalangi", "kekurangan", "tidak efektif",  
"meresahkan",  
  
"tidak ramah", "tidak profesional", "mencemaskan",  
"tidak bertanggung jawab",  
  
"tidak responsif", "tidak memadai", "terlalu banyak  
kesalahan",  
  
"tidak bermanfaat", "terlalu sulit", "tidak  
teratur", "membuat frustrasi",  
  
"tidak jelas", "menyulitkan", "membahayakan",  
"mengabaikan",  
  
"tidak mendukung", "menjengkelkan"
```

```
]
```

```
# Pre-processing function  
def preprocess_text(text):  
  
    # Memastikan input adalah string  
  
    if not isinstance(text, str):  
  
        return None  
  
    # Mengabaikan teks yang mengandung "-" atau "#"  
  
    if "-" in text or "#" in text:
```

```

        return None

    # Menghapus tanda baca
    text = text.translate(str.maketrans('', '',
string.punctuation))

    # Tokenisasi
    tokens = word_tokenize(text)

    # Menghapus stop words
    stop_words = set(stopwords.words('indonesian'))
    tokens = [word for word in tokens if word.lower()
not in stop_words]

    # Stemming dan Lemmatization
    stemmer = PorterStemmer()
    lemmatizer = WordNetLemmatizer()
    tokens = [stemmer.stem(lemmatizer.lemmatize(word))
for word in tokens]

    return ' '.join(tokens)

```

```

# Penggunaan fungsi preprocess_text

```

```

df['Processed_saran'] =
    df['saran'].apply(preprocess_text)

```

```

# Menghapus baris yang memiliki nilai None setelah pra-
pemrosesan

```

```
df = df.dropna(subset=['Processed_saran'])

# Fungsi untuk menentukan kategori berdasarkan kata
kunci

def categorize_complaint(text):

    text = text.lower() # Mengubah teks menjadi huruf
    kecil semua

    # Kategori berdasarkan tema lain

    if 'prodi' in text:

        return 'Prodi'

    elif 'fakultas' in text:

        return 'Fakultas'

    elif 'tata usaha' in text or 'pelayanan
    administrasi' in text:

        return 'Tata Usaha/Pelayanan Administrasi'

    elif 'simak' in text:

        return 'SIMAK'

    elif 'fasilitas' in text or 'kampus' in text:

        return 'Fasilitas Kampus'

    elif 'pengajaran' in text or 'kualitas pengajaran'
    in text:

        return 'Kualitas Pengajaran'

    elif 'kurikulum' in text or 'materi kuliah' in
```



```

text:

    return 'Kurikulum dan Materi'

elif 'perpustakaan' in text:

    return 'Perpustakaan'

elif 'keuangan' in text or 'pembayaran' in text:

    return 'Keuangan dan Pembayaran'

elif 'parkir' in text:

    return 'Fasilitas Parkir'

return None

# Mengkategorikan data
df['Category'] =
    df['Processed_saran'].apply(categorize_complaint)

# Menghapus baris yang tidak memiliki kategori (None)
df = df.dropna(subset=['Category'])

# Menambahkan kolom untuk Saran dan Kritik
df['Is_Saran'] = df['Processed_saran'].apply(lambda x:
    any(word in x for word in saran_words))

df['Is_Kritik'] = df['Processed_saran'].apply(lambda x:

```

```

any(word in x for word in kritik_words))

# Menampilkan 10 contoh data setelah kategorisasi
print("\n10 Contoh Data Setelah Kategorisasi:")

print(df[['saran', 'Category', 'Is_Saran',
         'Is_Kritik']].head(10))

# Menghitung jumlah data yang diproses dan tidak
# diproses
processed_count = df['Processed_saran'].notnull().sum()
unprocessed_count = len(df) - processed_count

# Menyimpan hasil ke dalam CSV
df.to_csv('saran_sunarti_processed.csv', index=False)

# Menampilkan hasil perhitungan
print(f"\nJumlah data yang diproses:
      {processed_count}")

# Menggunakan TF-IDF untuk ekstraksi fitur
vectorizer = TfidfVectorizer()

X = vectorizer.fit_transform(df['Processed_saran'])
y = df['Category']

```

```
# Membagi data menjadi data latih dan data uji

X_train, X_test, y_train, y_test = train_test_split(X,
    y, test_size=0.2, random_state=42)

# Menampilkan jumlah data latih dan data uji

print(f"\nJumlah Data Latih: {X_train.shape[0]}")
print(f"Jumlah Data Uji: {X_test.shape[0]}")

# Membuat model Naïve Bayes

model = MultinomialNB()
model.fit(X_train, y_train)

# Prediksi pada data uji

y_pred = model.predict(X_test)

# Menghitung akurasi

accuracy = accuracy_score(y_test, y_pred)

# Confusion Matrix

conf_matrix = confusion_matrix(y_test, y_pred)
```

```
# Laporan Klasifikasi

class_report = classification_report(y_test, y_pred)

# Menampilkan hasil evaluasi

print(f'\nAkurasi: {accuracy:.2f}')

print('\nConfusion Matrix:')

print(conf_matrix)

print('\nClassification Report:')

print(class_report)
```



397

SURAT PERMOHONAN PENELITIAN

Hal : Permohonan Surat

Pengantar Penelitian

Kepada Yth,

Ketua Program Studi Informatika

Di
Tempat

Assalamu Alaikum Warahmatullahi Wabarakatuh

Sehubungan dengan akan dilaksanakannya Penelitian yang akan dilaksanakan di instansi oleh mahasiswa Fakultas Teknik Program Studi Informatika. Adapun Mahasiswa yang bersangkutan adalah sebagai berikut:

NO	NAMA	NIM
1.	Sunari	105841101220

Maka dengan ini kami memohon dibuakan surat pengantar atau pengajuan pada instansi dibawah ini.

Judul Skripsi : KLASIFIKASI PENGADUAN PELAYANAN FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH MAKASSAR MENGGNAKAN NATURAL LANGUAGE PROCESING

Nama Instansi : SIMAK Universitas Muhammadiyah Makassar

Alamat : Jl. Sultan Alaudin No.259

Demikian surat permohonan kami ajukan, atas dukungan dan kerjasamanya kami hanturkan terima kasih.

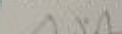
Inlilahi Fii Sabillilhaq, Fastabiqul Khairat

Wassalamualaikum Warahmatullahi Wabarakatuh

Makassar, 1 Safar 1446 H

5 Agustus 2024

Permohonan



Sunari

105841101220



UNIVERSITAS MUHAMMADIYAH MAKASSAR
FAKULTAS TEKNIK
PROGRAM STUDI INFORMATIKA



بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

Nomor : 397/05/C.4-VI/VIII/46/2024
 Lamp. : -
 Hal : **Pengantar Penelitian**

Makassar, 04 Safar 1446 H
 09 Agustus 2024 M

Kepada yang Terhormat,
Ketua LP3M Unismuh Makassar
 Di -
 Tempat

Assalamu 'Alaikum Warahmatullahi Wabarakatuh

Dengan Rahmat Allah SWT, Semoga aktivitas kita bernilai ibadah di Sisi - Nya. Dalam rangka penyelesaian Tugas Sarjana / Tugas Akhir Mahasiswa pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar dengan judul "**KLASIFIKASI PENGADUAN PELAYANAN FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH MAKASSAR MENGGUNAKAN NATURAL LANGUAGE PROCESING**", Sehubungan hal tersebut, maka kami meminta kesediaan Bapak/Ibu untuk membantu perihal surat tersebut. Bersama ini kami sampaikan mahasiswa (s):

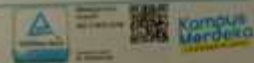
No	Stambuk	Nama
1.	105841101220	SUNARTI

Demikian surat kami atas perhatian dan kerja samanya kami haturkan banyak terima kasih
Jazakumullah Khaeran Katsiran
Wasalamu 'Alaikum warahmatullah Wabarakatuh

Ketua Program Studi
 Informatika

Mah. **M. A. M. Havat, S.Kom., MT.**
 NPM. 15014577

Tembusan: Kepada Yang Terhormat,
 1. Dekan Fakultas Teknik
 2. *Aras*





MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH MAKASSAR

LEMBAGA PENELITIAN PENGEMBANGAN DAN PENGABDIAN KEPADA MASYARAKAT

Jl. Sultan Hassanudin No. 299 Tj. P. 90972 Telp. 04113907080 Makassar 90221 e-mail: lp3m@unmah.ac.id

Nomor : 4795/05/C.4-VIII/VIII/1445/2024

12 August 2024 M

Lamp : 1 (satu) Rangkap Proposal

08 Safar 1446

Hal : Permohonan Izin Penelitian

Kepada Yth,

Dekan

Universitas Muhamamdiyah Makassar

di -

Makassar

Berdasarkan surat Dekan Teknik Universitas Muhammadiyah Makassar, nomor: 397/05/C.4-VI/VII/46/2024 tanggal 9 Agustus 2024, menerangkan bahwa mahasiswa tersebut di bawah ini:

Nama : SUNARTI

No. Stambuk : 10584 1101220

Fakultas : Teknik

Jurusan : Informatika

Pekerjaan : Mahasiswa

Bermaksud melaksanakan penelitian/pengumpulan data dalam rangka penulisan Skripsi dengan judul:

"KLASIFIKASI PENGADUAN PELAYANAN FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH MAKASSAR MENGGUNAKAN NATURAL LANGUAGE PROCESING"

Yang akan dilaksanakan dari tanggal 15 Agustus 2024 s/d 15 Oktober 2024.

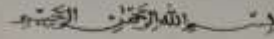
Sehubungan dengan maksud di atas, kiranya Mahasiswa tersebut diberikan izin untuk melakukan penelitian sesuai ketentuan yang berlaku.

Demikian, atas perhatian dan kerjasamanya diucapkan jazakumulahu khaeran

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Ketua LP3M,

Dr. Muh. Arief Muhsin, M.Pd.
NBM 1127761



SURAT KETERANGAN BEBAS PLAGIAT

UPT Perpustakaan dan Penerbitan Universitas Muhammadiyah Makassar,
Menerangkan bahwa mahasiswa yang tersebut namanya di bawah ini:

Nama : Sunarti

Nim : 105841101220

Program Studi : Teknik Informatika

Dengan nilai:

No	Bab	Nilai	Ambang Batas
1	Bab 1	10 %	10 %
2	Bab 2	25 %	25 %
3	Bab 3	8 %	10 %
4	Bab 4	4 %	10 %
5	Bab 5	4 %	5 %

Diyatakan telah lulus cek plagiat yang diadakan oleh UPT- Perpustakaan dan Penerbitan Universitas Muhammadiyah Makassar Menggunakan Aplikasi Turnitin.

Demikian surat keterangan ini dibagikan kepada yang bersangkutan untuk dipergunakan sebagaimana

Makassar, 30 Agustus 2024

Mengetahui,

Ketjala UPT- Perpustakaan dan Penerbitan,



Sunarti 105841101220 Bab I

by Tahap Tutup



Submission date: 30-Aug-2024 10:51AM (UTC+0700)

Submission ID: 2440963936

File name: BAB 1 - 2024-08-30T104956-112.docx (19.9K)

Word count: 899

Character count: 6247

Sunarti 105841101220 Bab I

ORIGINALITY REPORT

10% SIMILARITY INDEX
10% INTERNET SOURCES
0% PUBLICATIONS
0% STUDENT PAPERS

PRIMARY SOURCES



1	repository.teknokrat.ac.id Internet Source	2%
2	repository.uinsu.ac.id Internet Source	2%
3	fauzanahianto.student.telkomuniversity.ac.id Internet Source	2%
4	repository.unim.ac.id Internet Source	2%
5	tonosgratis.mobi Internet Source	2%
6	eprint.stimlog.ac.id Internet Source	2%

Exclude quotes on Exclude matches
Exclude bibliography on

Sunarti 105841101220 Bab II

by Tahap Tutup



Submission date: 30-Aug-2024 10:51AM (UTC+0700)

Submission ID: 2440964374

File name: BAB_II_2024-08-30T104953.761.docx (37,11K)

Word count: 1047

Character count: 7119

Sunarti 105841101220 Bab II

ORIGINALITY REPORT

25%

SIMILARITY INDEX

23%

INTERNET SOURCES

6%

PUBLICATIONS

10%

STUDENT PAPERS

PRIMARY SOURCES



1	jtiik.ub.ac.id Internet Source	5%
2	core.ac.uk Internet Source	4%
3	repository.teknokrat.ac.id Internet Source	4%
4	openjournal.unpam.ac.id Internet Source	3%
5	Submitted to Udayana University Student Paper	2%
6	dokumen.dinus.ac.id Internet Source	2%
7	www.coursehero.com Internet Source	2%
8	repository.upi.edu Internet Source	2%

Sunarti 105841101220 Bab III

by Tahap Tutup



Submission date: 30-Aug-2024 10:52AM (UTC+0700)
Submission ID: 2440964675
File name: BAB_III_-_2024-08-30T104956079.docx (46.78K)
Word count: 941
Character count: 6246

Sunarti 105841101220 Bab III

ORIGINALITY REPORT

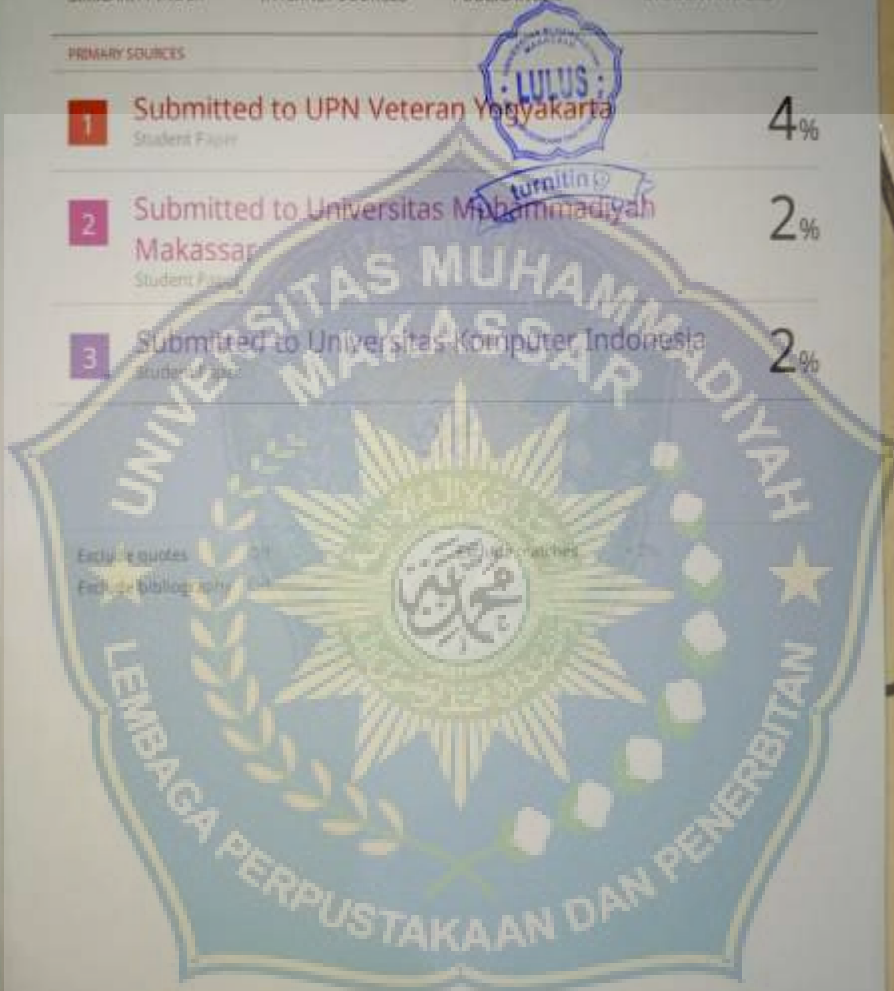
8%	6%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to UPN Veteran Yogyakarta Student Paper	4%
2	Submitted to Universitas Muhammadiyah Makassar Student Paper	2%
3	Submitted to Universitas Komputer Indonesia Student Paper	2%

Exclude quotes

Exclude bibliography



Sunarti 105841101220 Bab IV

by Tahap Tutup



Submission date: 30-Aug-2024 10:52AM (UTC+0700)

Submission ID: 2440965005

File name: 9-48_IV_-_2024-08-30T104956.112.docx (104.23K)

Word count: 2294

Character count: 13136

Sunarti 105841101220 Bab IV

ORIGINALITY REPORT

4% SIMILARITY INDEX
2% INTERNET SOURCES
2% PUBLICATIONS
0% STUDENT PAPERS

PRIMARY SOURCES

- 1 ojs.unud.ac.id
Internet Source 2%
- 2 Rizky Aziz, Tresna Maulana Fabrudin, Wahyu Syaifullah jauharis Saputra. "Analisis Sentimen Kepuasan Pengguna OYO DiPlaystore Dengan Multinomial Naive Bayes dan Chi-square", JURNAL FASILKOM, 2024
Publication 2%



Exclude quotes

Exclude bibliography

Exclude matches

Sunarti 105841101220 Bab V

by Tahap Tutup



Submission date: 30-Aug-2024 10:53AM (UTC+0700)

Submission ID: 2440965346

File name: BAB V_2024-08-30T104556112.docx (15.78K)

Word count: 219

Character count: 1542

Sunarti 105841101220 Bab V

ORIGINALITY REPORT

4%
SIMILARITY INDEX

4%
INTERNET SOURCES

0%
PUBLICATIONS

0%
STUDENT PAPERS

PRIMARY SOURCES

1 pt.scribd.com
Internet Source

4%



turnitin

Exclude quotes
Exclude bibliography

Exclude matches

