

**PEMODELAN TOPIK SARAN MAHASISWA PADA SIMAK
UNISMUH MENGGUNAKAN BERTOPIC**

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Mendapatkan gelar
Sarjana Komputer (S.Kom) Program Studi Informatika



LIS INDRIANI

105841108020

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MAKASSAR

2024

**PEMODELAN TOPIK SARAN MAHASISWA PADA SIMAK
UNISMUH MENGGUNAKAN BERTOPIC**

Diajukan sebagai Salah Satu Syarat untuk Mendapatkan gelar
Sarjana Komputer (S.Kom) Program Studi Informatika

Disusun dan Diajukan oleh:

LIS INDRIANI

105841108020

**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>

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

PENGESAHAN

Skripsi atas nama Lis Indriani dengan nomor induk Mahasiswa 105 84 11080 20, dinyatakan diterima dan disahkan oleh Panitia Ujian Tugas Akhir/Skripsi sesuai dengan Surat Keputusan Dekan Fakultas Teknik Universitas Muhammadiyah Makassar Nomor : 126/05/A.5-VI/45/2024, sebagai salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar pada hari Sabtu tanggal 24 Agustus 2024.

Panitia Ujian :

1. Pengawas Umum

Makassar, 19 Safar 1446 H
24 Agustus 2024 M

a. Rektor Universitas Muhammadiyah Makassar

Dr. Ir. H. Abd. Rakhim Nanda, ST., MT., IPU.

b. Dekan Fakultas Teknik Universitas Hasanuddin

Prof. Dr. Eng. Muhammad Isran Ramli, ST., MT.

2. Penguji

a. Ketua : Dr. Ir. Hj. Hafsah Nirwana, ST., MT.

b. Sekretaris : Titin Wahyuni, S.Pd., M.T.

3. Anggota : 1. Lukman Anas, S.Kom., M.T.

2. Lukman, S.Kom., M.T.

3. Muhyiddin A.M. Hayat, S.Kom., M.T.

Mengetahui :

Pembimbing I

Fahrim Irhamna Rahman S.Kom., M.T

Pembimbing II

Rizki Yusliana Bakti ST., MT

Dekan



Dr. Ir. Hj. Nurnawaty, ST., MT., IPM.

NBM / 795 108



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 Komputer (S.Kom) Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar.

Judul Skripsi : **PEMODELAN TOPIK SARAN MAHASISWA PADA SIMAK UNISMUH MENGGUNAKAN BERTOPIC**

Nama : Lis Indriani

Stambuk : 105 84 11080 20

Makassar, 30 Agustus 2024

Telah Diperiksa dan Disetujui
Oleh Dosen Pembimbing;

Pembimbing I

Pembimbing II

Fahrim Irhamna Rahman S.Kom., MT.

Rizki Yusliana Bakti, ST., MT.

Mengetahui,

Ketua Program Studi Informatika



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

NPM - 1504 577

ABSTRAK

LIS INDRIANI. Pemodelan Topik Saran Mahasiswa Pada SIMAK UNISMUH Menggunakan BERTopic (dibimbing oleh Fachrim Irhamna Rachman S.Kom., M.T. dan Rizki Yusliana Bakti S.T., M.T.)

Penelitian ini mengkaji penggunaan algoritma BERTopic untuk pemodelan topik pada data saran mahasiswa yang dikumpulkan melalui SIMAK UNISMUH. Penelitian ini bertujuan untuk mengidentifikasi dan memvisualisasikan pola tematik dalam saran mahasiswa, dengan tujuan meningkatkan layanan akademik dan fasilitas kampus. Studi ini memanfaatkan teknik *Natural Language Processing* (NLP), khususnya BERTopic, yang menggabungkan keunggulan *Bidirectional Encoder Representations from Transformers* (BERT) dengan algoritma *clustering* untuk menghasilkan representasi topik yang kontekstual dan mudah diinterpretasikan. Data penelitian terdiri dari 232.430 entri saran yang diproses untuk menghilangkan *noise* dan informasi yang tidak relevan, menghasilkan 26.009 entri valid. Entri-entri ini kemudian diproses menggunakan algoritma BERTopic, menghasilkan 9 topik yang berbeda terkait dengan berbagai aspek kehidupan akademik, termasuk metode pengajaran, fasilitas kampus, dan layanan administrasi. *Coherence score* sebesar 0.637 menunjukkan konsistensi internal yang kuat dalam topik-topik yang diidentifikasi sementara analisis mengungkapkan area utama di mana universitas dapat meningkatkan layanannya. Temuan dari penelitian ini memberikan wawasan yang dapat ditindaklanjuti bagi administrator universitas, memungkinkan mereka membuat keputusan dan kinerja akademik mahasiswa. Selain itu, penelitian ini berkontribusi pada bidang pemodelan topik dalam konteks pendidikan dan menunjukkan efektivitas BERTopic dalam mengolah data tekstual skala besar.

Kata Kunci: Pemodelan Topik, BERTopic, *Natural Language Processing*

ABSTRACT

LIS INDRIANI. *Topic Modelling of Student Suggestions on SIMAK UNISMUH Using BERTOPIC (Supervised by Fachrim Irhamna Rachman S.Kom., M.T. and Rizki Yusliana Bakti S.T., M.T.)*

This study examines the use of the BERTopic algorithm for topic modeling on student feedback data collected through the SIMAK UNISMUH. The research aims to identify and visualize thematic patterns in student feedback to improve academic services and campus facilities. This study utilizes Natural Language Processing (NLP) techniques, particularly BERTopic, which combines the advantages of Bidirectional Encoder Representations from Transformers (BERT) with clustering algorithms to produce contextually rich and easily interpretable topic representations. The research data comprises 232,430 feedback entries that were processed to remove noise and irrelevant information, resulting in 26,009 valid entries. These entries were then processed using the BERTopic algorithm, generating nine distinct topics related to various aspects of academic life, including teaching methods, campus facilities, and administrative services. The coherence score of 0.637 indicates strong internal consistency within the identified topics, while the analysis reveals key areas where the university can enhance its services. The findings from this study provide actionable insights for university administrators, enabling them to make informed decisions and improve student academic performance. Additionally, this research contributes to the field of topic modeling in educational contexts and demonstrates the effectiveness of BERTopic in processing large-scale textual data.

Keywords: Topic Modelling, BERTopic, Natural Language Processing

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur semoga selalu dipanjatkan kehadirat Allah SWT atas berkah dan karunia-NYA yang melimpah. Shalawat serta salam senantiasa dihaturkan kepada Nabi Muhammad SAW sebagai suri tauladan bagi seluruh umat manusia. Skripsi ini berjudul “PEMODELAN TOPIK SARAN MAHASISWA PADA SIMAK UNISMUH MENGGUNAKAN BERTOPIC” dapat diselesaikan sebagaimana mestinya.

Penulis mengucapkan banyak terima kasih kepada semua pihak yang telah memberikan dukungan dan motivasi selama proses penyusunan Skripsi dari awal hingga selesai. Pada kesempatan ini penulis mengucapkan terima kasih kepada :

1. Allah SWT, yang telah memberikan nikmat yang tak terhingga. Terima kasih atas petunjuk dan perlindungan-Mu yang selalu menyertai setiap langkah saya. Di saat-saat sulit, Engkau memberikan ketenangan, di saat keraguan melanda Engkau memberikan keyakinan, dan di setiap pencapaian Engkau senantiasa mengingatkan saya akan pentingnya bersyukur.
2. Kedua orang tuaku dan keluarga yang senantiasa memberikan dukungannya. Terima kasih telah menjadi sumber inspirasi dan kekuatan saya. Terima kasih atas setiap pengorbanan yang telah kalian lakukan, atas kesabaran, kasih sayang, serta dorongan dan doa yang tidak pernah henti kalian berikan. Tanpa kalian, saya tidak akan bisa berada di titik ini.
3. Ibu DR.Ir.Hj Nurnawati, S.T.,M.T.,I.P.M, selaku Dekan Fakultas Teknik.
4. Bapak Muh. Syafaat S Kuba, S.T., M.T, selaku Wakil Dekan Fakultas Teknik.
5. Bapak Muhyiddin AM Hayat S.Kom.,MT, selaku Ketua Prodi Informatika sekaligus Dosen Pendamping Akademik.

6. Bapak Fahrin Irhamna Rachman S.Kom,M.T, selaku Dosen Pembimbing 1
7. Ibu Rizki Yusliana Bakti S.T.,MT, selaku Dosen Pembimbing 2
8. Dosen dan Staf Fakultas Teknik Universitas Muhammadiyah Makassar.
9. Teman-teman seperjuangan Angkatan 2020 Fakultas Teknik, Khususnya Kelas C.
10. Dolara Grup yaitu Rizka Adrianingsih, Ayu Andira, Rosalinda Aprilia Sari, Arya Wibawa dan David Arian Virgiawan. Tidak ada kata yang mampu menggambarkan betapa berharganya waktu dan usaha yang telah kita lalui bersama. Harapan besar saya adalah kita semua bisa mencapai impian dan cita-cita yang telah kita perjuangkan dengan sekuat tenaga.
11. Anak Kos Pondok Asrah, khususnya Norfazirah, Arda Talia, dan Helmi Febrianti. Terima kasih atas kebersamaan, dukungan, dan semangat yang tak pernah pudar. Dalam setiap tawa, tangis, dan keluh kesah, kalian adalah sumber kekuatan yang selalu hadir di saat paling sulit.

Skripsi ini telah disusun sebaik-baiknya, namun tentu saja masih memiliki ruang untuk ditingkatkan mengingat adanya keterbatasan dan kemampuan penulis. Oleh karena itu, penulis sangat mengharapkan masukan, kritik dan saran yang membangun untuk penyempurnaan Skripsi ini. Semoga Skripsi ini dapat memberikan manfaat bagi semua pihak yang terlibat.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

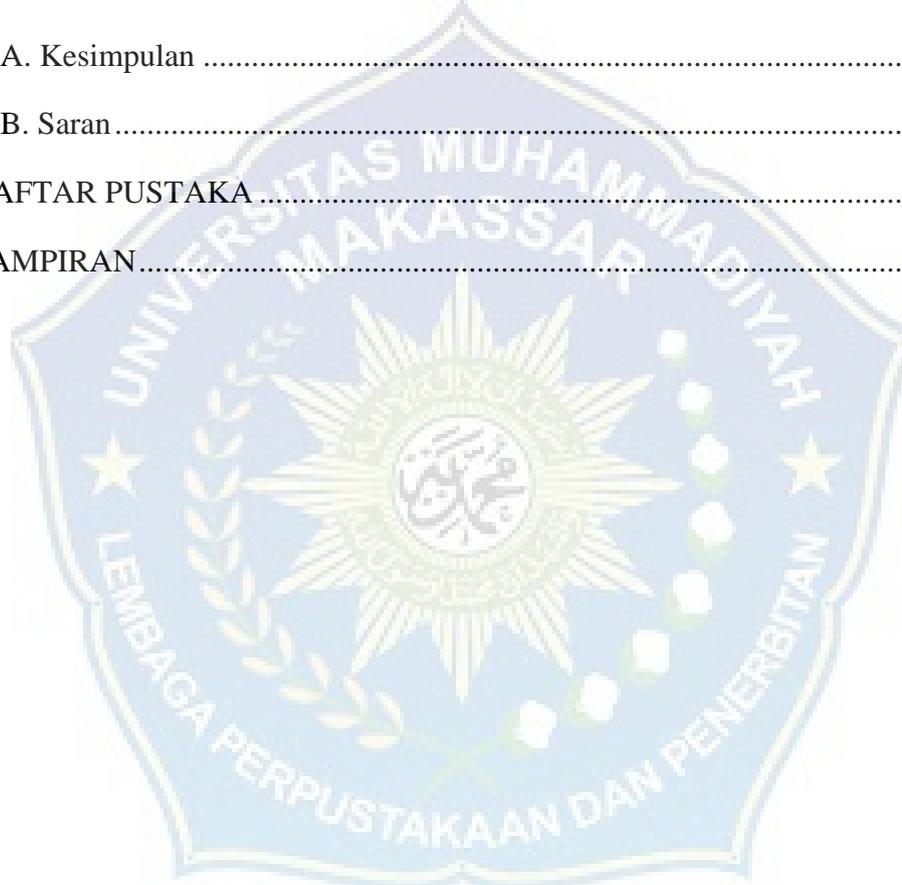
Makassar, 24 Agustus 2024

Peneliti

DAFTAR ISI

ABSTRAK	ii
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR ISTILAH	xiv
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Rumusan Masalah	3
C. Tujuan Penelitian.....	3
D. Manfaat Penelitian.....	3
E. Ruang Lingkup	3
F. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
A. Landasan Teori.....	5
B. Penelitian Terkait.....	12
D. Kerangka Pikir.....	16
BAB III METODOLOGI PENELITIAN	17
A. Tempat dan Waktu Penelitian	17
B. Alat dan Bahan Penelitian	17
C. Perancangan Sistem.....	17
D. Teknik Pengujian Sistem.....	20
E. Teknik Analisis Data	20

BAB IV	PEMBAHASAN.....	22
A.	Pengolahan Data.....	22
B.	Analisis Pemodelan Topik.....	24
C.	Evaluasi Performa Pemodelan Topik	38
D.	Visualisasi Pemodelan Topik.....	40
BAB V	KESIMPULAN DAN SARAN	43
A.	Kesimpulan	43
B.	Saran.....	44
DAFTAR PUSTAKA	45
LAMPIRAN	48



DAFTAR TABEL

Tabel 1. Perbandingan Penelitian Terkait	12
Tabel 2 Hasil Preprocessing Data Mahasiswa	22



DAFTAR GAMBAR

Gambar 1. Penerapan NLP (Amazinum, 2022)	7
Gambar 2. Algoritma BERTopic (Bhangale, 2023).....	8
Gambar 3. Kerangka Pikir Penelitian.....	16
Gambar 4. Flowchart Sistem.....	18
Gambar 5 Perbandingan Jumlah kata Sebelum dan Sesudah Preprocessing	24
Gambar 6. Kumpulan Kata dan Representasi Dokumen Topik 0.....	29
Gambar 7.Kumpulan Kata dan Representasi Dokumen Topik 1	30
Gambar 8. Kumpulan Kata dan Representasi Dokumen Topik 2.....	31
Gambar 9. Kumpulan Kata dan Representasi Dokumen Topik 3.....	32
Gambar 10. Kumpulan Kata dan Representasi Dokumen Topik 4.....	33
Gambar 11. Kumpulan Kata dan Representasi Dokumen Topik 5.....	34
Gambar 12. Kumpulan Kata dan Representasi Dokumen Topik 6.....	35
Gambar 13. Kumpulan Kata dan Representasi Dokumen Topik 7.....	36
Gambar 14. Kumpulan Kata dan Representasi Dokumen Topik 8.....	37
Gambar 15. Runtime Algoritma BERTopic	38
Gambar 16. Coherence Score Algoritma BERTopic	39
Gambar 17. Pengelompokan Berdasarkan Topik.....	40
Gambar 18. Intertopic Distance Map Bertopic	41

DAFTAR LAMPIRAN

Lampiran 1 Hasil Plagiasi	48
Lampiran 2. Dataset	59
Lampiran 3. Source Code.....	67
Lampiran 4. Topic Info	70



DAFTAR ISTILAH

Analisis	Proses mengevaluasi dan menginterpretasi data untuk memahami atau menjelaskan informasi.
BERT	(<i>Bidirectional Encoder Representations from Transformers</i>) adalah Model bahasa yang menggunakan arsitektur transformer untuk memahami konteks dua arah dalam teks.
BERTopic	Metode pemodelan topik yang menggabungkan model bahasa BERT dengan teknik clustering untuk menemukan dan memvisualisasikan topik dalam teks.
Case Folding	Proses normalisasi teks dengan mengubah semua huruf menjadi huruf kecil untuk konsistensi.
Clustering	Teknik dalam machine learning untuk mengelompokkan data ke dalam grup berdasarkan karakteristik yang mirip.
Coherence Score	Ukuran yang digunakan untuk mengevaluasi kualitas topik yang dihasilkan oleh pemodelan topik, berdasarkan kesamaan kata dalam topik tersebut.
Dimensionality Reduction	Proses mengurangi jumlah variabel acak yang dipertimbangkan dalam analisis, sering digunakan untuk visualisasi data atau mempercepat algoritma machine learning.

<i>Dokumen Representatif</i>	Dokumen yang dipilih sebagai contoh terbaik untuk mewakili topik atau kategori tertentu.
<i>Embedding</i>	Representasi kata atau frasa dalam bentuk vektor numerik yang memetakan makna semantik ke dalam ruang vektor.
<i>HDBSCAN</i>	<i>(Hierarchical Density-Based Spatial Clustering of Applications with Noise)</i> Algoritma clustering yang dapat mengidentifikasi kelompok data dengan berbagai bentuk dan kepadatan, termasuk menangani outliers.
<i>Inisialisasi</i>	Langkah awal dalam algoritma machine learning atau pemrosesan data yang menyiapkan nilai-nilai awal atau konfigurasi.
<i>LDA</i>	<i>(Latent Dirichlet Allocation)</i> Algoritma pemodelan topik yang mengasumsikan bahwa setiap dokumen adalah campuran dari sejumlah topik, dan setiap topik adalah campuran dari sejumlah kata.
<i>Neural Network</i>	Jaringan Saraf Tiruan, metode dalam kecerdasan buatan yang meniru cara kerja otak manusia untuk memproses informasi.
<i>NLP</i>	<i>(Natural Language Processing)</i> Cabang kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia, termasuk pemahaman dan generasi bahasa.
<i>Preprocessing</i>	Langkah-langkah awal dalam analisis data atau pemrosesan teks untuk membersihkan

	dan menyiapkan data, seperti menghapus noise atau melakukan normalisasi.
<i>Running Time</i>	Durasi waktu yang diperlukan untuk menjalankan suatu algoritma atau proses dari awal hingga selesai.
SIMAK	(<i>Sistem Informasi Manajemen Akademik</i>) Sistem yang digunakan untuk mengelola data akademik dan administrasi di institusi pendidikan.
<i>Stopwords</i>	Kata-kata umum seperti “dan”, “atau”, “tetapi” yang sering diabaikan dalam analisis teks karena dianggap tidak memberikan informasi penting.
<i>TF-IDF</i>	(<i>Term Frequency-Inverse Document Frequency</i>) Teknik untuk mengevaluasi seberapa penting sebuah kata dalam sebuah dokumen relatif terhadap seluruh kumpulan dokumen, sering digunakan dalam pemrosesan teks.
Topik	Sekumpulan kata-kata yang sering muncul bersama dalam kumpulan dokumen dan dianggap mewakili tema atau subjek tertentu.
<i>UMAP</i>	(<i>Uniform Manifold Approximation and Projection</i>) Teknik dimensionality reduction yang digunakan untuk visualisasi data, mempertahankan struktur lokal dan global dalam data.
Visualisasi	Representasi grafis dari data atau informasi yang memungkinkan pengguna untuk

memahami pola, tren, dan wawasan yang ada.

$W_{\{k,c\}}$

Bobot suatu kata (k) dalam sebuah klaster atau dokumen (c).



BAB I

PENDAHULUAN

A. Latar Belakang

Perguruan tinggi memiliki peran penting dalam menciptakan lingkungan yang mendukung perkembangan akademik, pribadi, dan profesional mahasiswanya. Salah satu instrumen yang digunakan untuk melayani kebutuhan mahasiswa dan meningkatkan kualitas layanan adalah sistem informasi. Universitas Muhammadiyah Makassar (UNISMUH) menyediakan SIMAK sebagai layanan informasi dan kebutuhan mahasiswa. SIMAK (Sistem Informasi Manajemen Akademik) memiliki peran penting dalam mengelola berbagai informasi terkait dengan kegiatan akademik, administratif, dan pengelolaan data mahasiswa. Salah satu fitur yang terdapat dalam SIMAK UNISMUH adalah pengisian saran.

Pengisian saran mahasiswa merupakan bagian dari fungsi SIMAK yang memungkinkan mahasiswa untuk memberikan umpan balik terhadap aspek kehidupan kampus. Melalui pengisian saran ini, mahasiswa dapat mengemukakan pendapat, keluhan, dan saran terkait dengan pengalaman belajar, fasilitas kampus, layanan akademik, dan berbagai aspek lainnya yang mempengaruhi kehidupan mahasiswa di universitas. Penggunaan pengisian saran mencerminkan komitmen universitas untuk melibatkan mahasiswa secara aktif dalam proses pengambilan keputusan dan peningkatan kualitas layanan.

Menganalisis dan memahami umpan balik menjadi tugas yang rumit dan memakan waktu, terutama ketika data yang dikumpulkan melibatkan jumlah mahasiswa yang besar dan beragam. Proses manual untuk menganalisis umpan balik ini tidak hanya membutuhkan waktu yang lama tetapi juga rentan terhadap kesalahan dan subjektivitas. Oleh karena itu, diperlukan pendekatan yang lebih efisien dan objektif untuk mengolah data ini. Teknik pemodelan topik adalah salah satu metode yang dapat digunakan untuk mengatasi tantangan ini.

Pemodelan topik adalah proses identifikasi dan penjelajahan pola tematik yang muncul dalam kumpulan dokumen (Kherwa & Bansal, 2020). Dengan menerapkan teknik ini, kita dapat mengidentifikasi topik-topik utama yang muncul

dalam umpan balik mahasiswa dan mendapatkan pemahaman yang lebih baik tentang preferensi dan kebutuhan mereka.

Salah satu pendekatan yang menjanjikan dalam pemodelan topik adalah BERTopic, yang menggabungkan keunggulan model bahasa BERT (*Bidirectional Encoder Representations from Transformer*) dengan teknik pemodelan topik. BERT adalah model bahasa yang sangat canggih yang telah terbukti sangat efektif dalam bidang NLP (*Natural Language Processing*) (Lopez-Martinez & Sierra, 2020). Dibandingkan dengan metode populer lainnya seperti LDA (Lirichlet Dirichlet Allocation), BERTopic lebih mampu menghasilkan representasi kata yang lebih kaya dan kontekstual dengan mempertimbangkan seluruh konteks kalimat, sementara LDA hanya memperhitungkan kemunculan kata dalam dokumen dan topik (Mustafac, 2021).

Penerapan BERTopic pada umpan balik mahasiswa dari pengisian saran di perguruan tinggi dapat memberi manfaat yang signifikan. Dengan menggunakan teknik ini, kita dapat mengidentifikasi dan mengeksplorasi pola tematik yang muncul dalam umpan balik tersebut secara lebih efisien dan objektif. Ini akan memungkinkan perguruan tinggi untuk mendapatkan wawasan yang lebih dalam tentang preferensi, kebutuhan, dan masalah yang dihadapi oleh mahasiswa mereka.

Penelitian ini bertujuan untuk menerapkan teknik pemodelan topik menggunakan BERTopic pada umpan balik mahasiswa dari pengisian saran di Universitas Muhammadiyah Makassar (Unismuh). Dengan demikian, penelitian ini tidak hanya akan memberikan wawasan yang lebih baik tentang pola umpan balik mahasiswa, tetapi juga akan memberikan dasar yang kuat untuk pengambilan keputusan yang lebih baik.

Penelitian ini didasarkan pada referensi dari jurnal penelitian Gibran Giffari Priyatna(2022) dan Herwinsyah (2023) yang membahas pemodelan menggunakan algoritma BERTopic. Perbedaan penelitian ini dengan kedua penelitian tersebut adalah penulis menerapkannya pada data saran mahasiswa dan memvisualisasikan persebaran topiknya.

B. Rumusan Masalah

Berdasarkan pada latar belakang masalah yang telah diuraikan, dapat dirumuskan masalah sebagai berikut :

1. Bagaimana penerapan teknik pemodelan topik menggunakan BERTopic untuk menganalisis data saran mahasiswa pada Universitas Muhammadiyah Makassar?
2. Bagaimana visualisasi hasil metode BERTopic pada pemodelan topik data saran mahasiswa?

C. Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian adalah:

1. Mengetahui penerapan teknik pemodelan topik menggunakan BERTopic untuk menganalisis data saran mahasiswa pada Universitas Muhammadiyah Makassar.
2. Mengetahui visualisasi hasil metode BERTopic pada pemodelan data saran mahasiswa

D. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah memberikan kontribusi yang bermanfaat bagi Universitas Muhammadiyah Makassar dalam hal efisiensi pengolahan data, peningkatan kualitas layanan dan pengembangan sistem informasi yang responsif. Selain itu, diharapkan penelitian ini dapat menjadi referensi penting bagi penelitian-penelitian terkait dimasa mendatang.

E. Ruang Lingkup

Berdasarkan rumusan masalah dan tujuan penelitian , diterapkan beberapa batasan masalah yakni:

1. Data yang digunakan hanya berupa data saran mahasiswa pada Universitas Muhammadiyah Makassar.
2. Fokus pada penggunaan teknik pemodelan topik, khususnya BERTopic

F. Sistematika Penulisan

Secara keseluruhan, struktur penulisan dari penelitian ini terdiri dari beberapa bab yang disusun sebagai berikut:

BAB I PENDAHULUAN

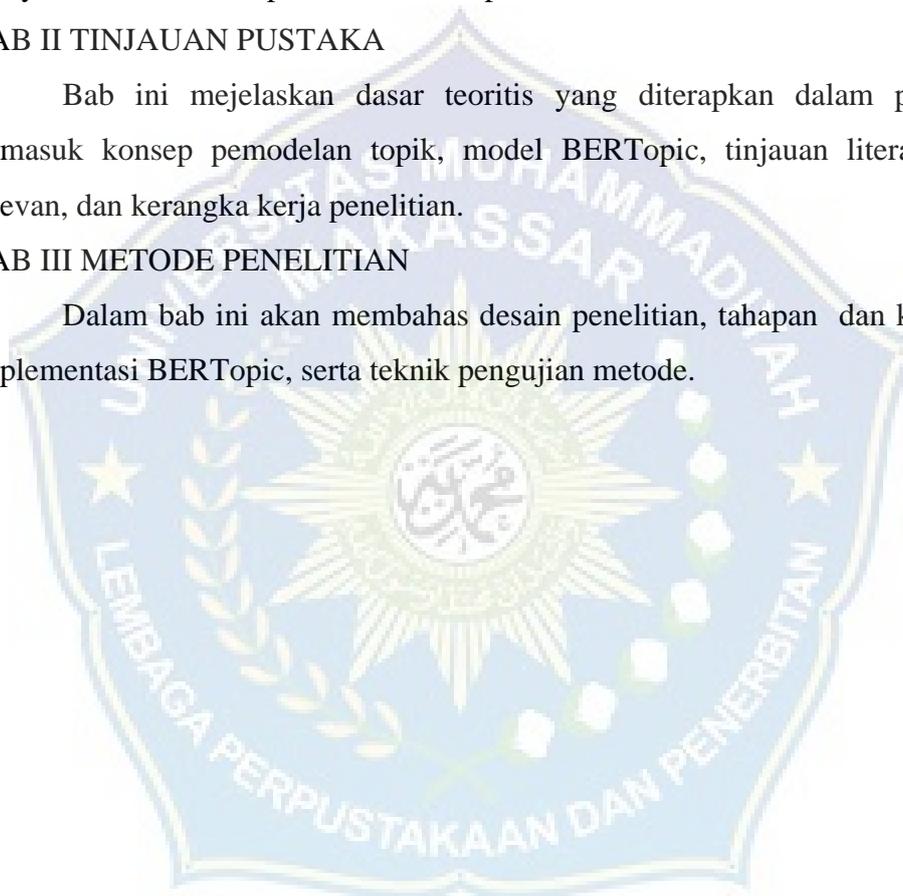
Bab ini menjelaskan latar belakang penelitian, merumuskan masalah, menetapkan tujuan penulisan, mengidentifikasi ruang lingkup masalah dan menyusun sistematika penulisan dalam penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan dasar teoritis yang diterapkan dalam penelitian, termasuk konsep pemodelan topik, model BERTopic, tinjauan literatur yang relevan, dan kerangka kerja penelitian.

BAB III METODE PENELITIAN

Dalam bab ini akan membahas desain penelitian, tahapan dan kebutuhan implementasi BERTopic, serta teknik pengujian metode.



BAB II

TINJAUAN PUSTAKA

A. Landasan Teori

1. SIMAK

Sistem Informasi Manajemen Akademik (SIMAK) adalah sebuah sistem informasi yang dirancang khusus untuk membantu perguruan tinggi dalam manajemen administrasi akademik. Pengaruh dan dorongan teknologi informasi telah signifikan dalam pengembangan pelayanan akademik. Tujuannya adalah untuk memastikan bahwa mahasiswa dari perguruan tinggi mendapatkan layanan yang cepat dan berkualitas (Reza et al., 2021).

SIMAK mengintegrasikan berbagai fungsi penting dalam pengelolaan akademik, seperti pendaftaran mahasiswa, pengelolaan kurikulum, penjadwalan kuliah, pengelolaan nilai, dan lain sebagainya. Adanya sistem informasi bagi sebuah perguruan tinggi menjadi sangat penting dalam upaya meningkatkan kinerja, kualitas layanan dan daya saing institusi pendidikan yang lebih baik (Iskandar Joko et al., 2022).

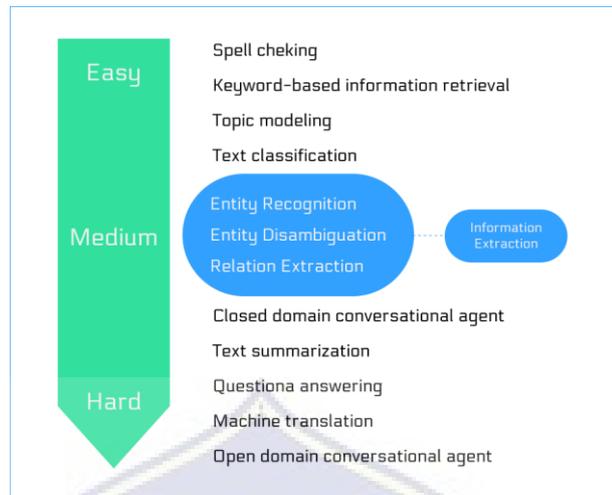
2. NLP

NLP merupakan bagian dari kecerdasan buatan yang bertujuan untuk menyelidiki interaksi yang kompleks antara komputer dan bahasa manusia. NLP berfungsi sebagai jembatan antara dunia manusia dan komputer, memungkinkan komputer untuk berinteraksi dengan manusia melalui bahasa alami. Hal ini melibatkan pengembangan model dan algoritma yang dapat mengenali pola-pola linguistik, konteks, dan makna dari teks (Khamidatullailiyah, 2022).

NLP dilatih menggunakan teknik pembelajaran mesin, yang mengikuti aturan yang telah ditentukan dan secara konsisten memperoleh data teks yang diperlukan sehingga meningkatkan kualitas teks secara terus-menerus. Proses ini melibatkan beberapa tahap pemahaman dan pengolahan yang dapat diakses oleh komputer (Amazinum, 2022) :

- a. *Morphological level*, berfokus pada pemahaman struktur kata dan hubungannya satu sama lain, termasuk penanganan kata-kata berimbuhan dan pembentukan kata-kata baru.
- b. *Syntactic level*, melibatkan konstruksi kalimat yang benar dan memastikan bahwa struktur kalimat sesuai dengan tata bahasa yang tepat.
- c. *Phonetic level*, memperhatikan konstruksi suara dan memahaminya sebagai entitas fisik, penting untuk pengenalan ucapan dan sintesis suara.
- d. *Semantic level*, memahami makna literal dari kata-kata, kalimat, dan frasa dalam konteks tertentu. Ini melibatkan pengenalan dan interpretasi makna kata-kata, serta pemahaman terhadap hubungan antara konsep-konsep yang terkandung dalam teks.
- e. *Lexical level*, memproses dan memahami bagian-bagian ucapan, termasuk identifikasi dan klasifikasi kata-kata berdasarkan jenis dan perannya dalam kalimat.
- f. *Pragmatic level*, melibatkan penggunaan pengetahuan dunia nyata untuk membangun konteks yang benar dalam teks. NLP harus mampu menginterpretasikan teks berdasarkan pengetahuan tentang dunia nyata dan konteks situasional.
- g. *Discourse level*, memahami bagian-bagian teks yang melampaui satu kalimat, termasuk hubungan dan alur cerita antara kalimat-kalimat dalam teks yang lebih panjang.

Bahasa manusia penuh dengan ambiguitas, homonim, idiom, beserta metafora yang membuat penulisan perangkat lunak menjadi sulit. Namun, mesin telah belajar untuk mengatasi semua kesulitan tersebut, menyediakan teks siap pakai dan mempermudah pekerjaan manusia. Tergantung pada kebutuhan dan tujuan, ada cukup banyak pengaplikasian NLP diantaranya chatbot, analisis sentimen, pemodelan topik, penerjemahan otomatis dan masih banyak lagi yang ditunjukkan pada gambar dibawah ini :



Gambar 1. Penerapan NLP (Amazinum, 2022)

3. Pemodelan Topik

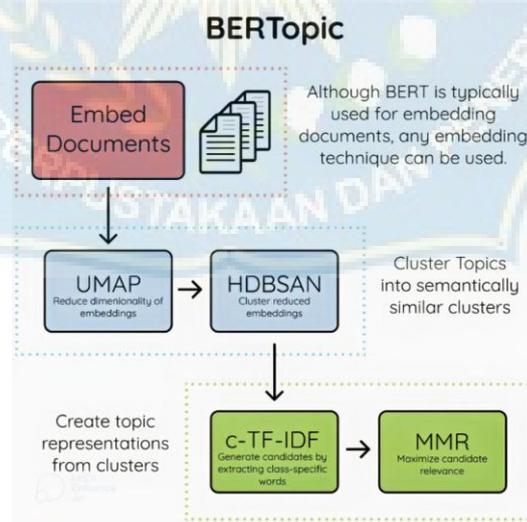
Pemodelan topik merupakan metode yang melibatkan serangkaian algoritma yang bertujuan untuk mengungkapkan, menemukan, dan menandai struktur tematik yang terdapat dalam kumpulan dokumen (Kherwa & Bansal, 2020). Konsep pemodelan topik awalnya digunakan di bidang “Modelisasi probabilitas generatif”. Model-model probabilitas generatif tersebut digunakan untuk menyelesaikan berbagai tugas seperti estimasi probabilitas, pembentukan model data, serta pengelompokan kelas dengan memanfaatkan konsep probabilitas. Prinsip utama di balik pendekatan ini adalah untuk menyusun ringkasan singkat yang mengungkap topik-topik yang paling umum dari sekumpulan dokumen yang luas. Dengan mengambil rangkaian dokumen sebagai input, model tersebut mampu menghasilkan serangkaian topik yang menggambarkan konten dokumen dengan akurat dan kohesif (Andronikou, 2022).

4. BERT

BERT merupakan model kompleks yang diperkenalkan oleh Google pada tahun 2018 yang sangat canggih dan efektif dalam memahami konteks dan makna kalimat secara mendalam. Karena kemampuannya dalam merepresentasikan kalimat dengan baik, BERT telah berhasil menunjukkan kinerja yang sangat baik dalam hampir semua sub bidang NLP. Sebaliknya, model bebas konteks seperti *Word2vec* dan *Glove* menghasilkan *embedding* kata tunggal untuk setiap kata dalam kosa kata

(Lopez-Martinez & Sierra, 2020). BERT dapat memperhitungkan kata-kata sebelum dan sesudahnya, sehingga menghasilkan representasi yang lebih kaya dan informatif. Meskipun kompleks dalam hal cara kerja internalnya dan membutuhkan besar data dan daya pemrosesan untuk pelatihan, namun ada banyak model BERT yang siap pakai sehingga mungkin untuk mencapai hasil yang baik dengan menggunakannya, bahkan dengan memberikan tambahan set pelatihan tergantung pada tugas yang akan dilakukan. Tujuan utama dari model BERT adalah menciptakan *fixed-length vector* yang mewakili kalimat, dengan memperhatikan makna dan urutan kata (Atagün et al., 2021).

BERTopic adalah salah satu teknik terbaru BERT dalam pemodelan yang memanfaatkan kemajuan dalam kecerdasan buatan. Dengan memanfaatkan keunggulan BERT, BERTopic mampu menciptakan model topik yang lebih baik dan lebih mudah diinterpretasikan. BERTopic memanfaatkan embedding BERT serta matriks TF-IDF berbasis kelas untuk mengidentifikasi kluster padat dalam dokumen. Kluster-kluster yang padat ini memungkinkan untuk menghasilkan topik-topik yang mudah diinterpretasikan, sambil tetap mempertahankan kata-kata yang paling relevan dalam deskripsi topik (Bhangale, 2023). Proses ini melalui langkah-langkah penting dalam algoritma, yang disajikan dalam gambar dibawah ini :



Gambar 2. Algoritma BERTopic (Bhangale, 2023)

a. *Embedding*

Embedding merupakan proses mengubah kata yang terdiri dari karakter alfanumerik menjadi vector. Meskipun mendukung berbagai teknik embedding yang berbeda, secara default BERTopic menggunakan transformer kalimat. Transformer kalimat adalah model yang dioptimalkan untuk menangkap kemiripan semantik antara kalimat-kalimat dalam bahasa alami.

Ada tiga model transformer kalimat multibahasa yang digunakan. Pertama *distiluse-base-multilingual-cased-v2* adalah versi yang disuling dari *Universal Sentence Encoder* multibahasa. Model ini menghasilkan representasi kalimat dalam ruang vektor dengan dimensi 512. Kedua, *paraphrase-multilingual-MiniLM-L12-v2* adalah versi multibahasa dari model *paraphrase-MiniLM-L12-v2* yang lebih kecil. Model ini memetakan kalimat ke dalam ruang vektor dengan dimensi 384. Ketiga, *paraphrase-multilingual-mpnet-base-v2* adalah model multibahasa terbesar yang memetakan kalimat ke dalam ruang vektor dengan dimensi 768 (Medvecki et al., 2024).

b. *Dimensionality Reduction*

Saat melakukan proses *embedding*, setiap kata atau kalimat direpresentasikan sebagai vektor dalam ruang berdimensi tinggi. Setiap kata direpresentasikan sebagai vektor setidaknya menjadi 384 dimensi (Medvecki et al., 2024). Ini membuat setiap data memiliki dimensi yang sangat tinggi terutama saat ada banyak kata atau kalimat dalam dataset. Beberapa algoritma *clustering* mengalami kesulitan saat memproses data dengan dimensi yang terlalu tinggi karena semakin banyak dimensi data, semakin sulit bagi algoritma untuk menemukan pola atau kesamaan diantara titik-titik data. Untuk mengatasi masalah ini, salah satu pendekatan yang umum digunakan adalah *dimensionality reduction*. Teknik ini mengurangi jumlah dimensi dalam dataset sambil berusaha mempertahankan sebanyak mungkin informasi yang relevan (Dorrity et al., 2020).

UMAP (*Uniform Manifold Approximation and Projection*) adalah metode baru dalam *dimensionality reduction* yang memulai prosesnya dengan memperkirakan topologi dari data berdimensi tinggi, kemudian menggunakan informasi ini untuk membuat representasi baru dari data dalam ruang berdimensi

rendah. Dibandingkan dengan teknik lain seperti t-Distributed Stochastic Neighbor Embedding (t-SNE), UMAP lebih baik dalam mempertahankan elemen-elemen struktur data dari ruang berdimensi tinggi. Ini berarti UMAP dapat menangkap baik hubungan lokal di antara grup-grup titik data, maupun hubungan global di antara grup-grup yang berbeda (Dorrity et al., 2020).

c. Clustering

Clustering adalah teknik analisis data yang bertujuan untuk mengelompokkan objek atau data ke dalam kelompok-kelompok yang memiliki kesamaan tertentu. Dalam pemodelan topik, clustering bertujuan mengidentifikasi pola atau struktur dalam kumpulan dataset sehingga mendapatkan topik-topik utama yang dibahas dalam teks tersebut (Priyatna, 2022).

HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) adalah sebuah algoritma *clustering* yang didasarkan pada kerapatan dan mengidentifikasi klaster-klaster dengan tingkat kepadatan yang berbeda. Dalam HDBSCAN, titik-titik data yang berada dekat secara spasial akan dianggap sebagai anggota dari klaster yang sama, meskipun bentuk dan ukuran klaster tidak perlu sama (Stewart & Al-Khassaweneh, 2022).

d. Class Based TF-IDF

Metode TF-IDF adalah teknik yang digunakan dalam pemrosesan teks untuk mengevaluasi pentingnya sebuah kata dalam suatu dokumen atau korpus teks. Ini dilakukan dengan menggabungkan dua statistik yaitu *term frequency* (TF) dan *inverse document frequency* (IDF) (Priyatna, 2022). Berikut rumusnya :

$$W_{k,c} = tf_{k,c} \cdot \log\left(\frac{N}{df_k}\right) \quad (1)$$

$tf_{k,d}$ = Frekuensi kata k pada dokumen d

df_k = Jumlah dokumen yang mengandung kata k

N = Jumlah dokumen di sebuah korpus N

Di mana frekuensi kata k dalam dokumen d diperhitungkan oleh *term frequency* (TF), sementara jumlah informasi yang tersedia untuk suatu kata dalam dokumen diukur oleh *inverse document frequency* (IDF). IDF dihitung dengan mengambil logaritma dari total N dokumen dalam korpus dibagi dengan jumlah dokumen yang mengandung kata k .

Selanjutnya, prosedur ini diterapkan pada pengelompokan dokumen, di mana setiap kelompok dianggap sebagai satu dokumen tunggal dengan menggabungkan dokumen yang ada. Kemudian, nilai TF-IDF disesuaikan untuk mewakili dokumen dalam kluster dengan mentranslasikan dokumen ke dalam kluster tersebut. Sehingga menghasilkan rumus berikut :

$$W_{k,c} = tf_{k,c} \cdot \left(1 + \frac{A}{tf_k}\right) \quad (2)$$

$tf_{k,c}$ = Frekuensi kata k pada kluster c

tf_k = Frekuensi kata k pada semua kluster

A = Jumlah rata-rata kata per kelas A

Term frequency memodelkan seberapa sering kata k muncul dalam kelompok c , di mana kelompok c merujuk pada kumpulan dokumen yang digabungkan menjadi satu dokumen dalam kluster. Selanjutnya, *inverse document frequency* diganti dengan *inverse class frequency* untuk mengukur seberapa informatif sebuah kata bagi suatu kelas. Ini dihitung dengan mengambil logaritma dari rata-rata jumlah kata per kelas A dibagi dengan frekuensi kata k di semua kelas. Untuk memastikan hanya nilai positif yang ditampilkan, kita menambahkan satu ke pembagian dalam algoritma.

Dengan demikian, prosedur TF-IDF ini memodelkan pentingnya kata-kata dalam kelompok daripada dokumen individu, memungkinkan kita untuk menghasilkan distribusi topik kata untuk setiap kelompok dokumen. Terakhir, dengan menggabungkan representasi C-TF-IDF secara iteratif dari topik yang

paling jarang muncul dengan topik yang paling umum, jumlah topik dapat dikurangi sesuai dengan keinginan pengguna

B. Penelitian Terkait

Peneliti memperoleh inspirasi dan referensi terkait penelitian ini melalui penelitian sebelumnya. Penelitian sebelumnya meliputi :

Tabel 1. Perbandingan Penelitian Terkait

Peneliti	Tujuan	Metode	Data	Hasil
Yayang Matira, Junaidi, Iman Setiawan (2023)	Menentukan pemodelan topik pada judul berita online	<i>Latent Dirichlet Allocation</i>	Judul berita online detikcom kanal detiknews 2021-2022	LDA menghasilkan <i>coherence score</i> sebesar 0,7586 dengan jumlah topik terbaik sebanyak 3 topik
Razief Perucha Fauzie Afidh, Syahrial (2023)	Menerapkan penggunaan n-gram dan Non-Negative Matrix pada pemodelan topik	n-gram dan <i>Non-Negative Matrix Factorization</i>	Artikel portal berita (53.920)	Penggunaan n-gram dan Non Negative Matrix Factorization memberikan hasil yang baik.
Herwinsyah (2023)	Mencari topik dalam Al Qur'an	BERTopic	Ayat Al-Qur'an (6236 data)	Ditemukan 8 topik utama

Gibran Giffari Priyatna (2022)	Mengetahui topik yang dibicarakan oleh pemain game dan performa topic modelling	BERTopic	Ulasan negatif video game 2017-2021 (55.203 data)	Mendapatkan 10 topik dengan <i>coherence score</i> sebesar 0.5244758866 dan running time selama 10.7 menit
Dziky Ridhwanullah (2022)	Analisis dan visualisasi pemodelan topik penyakit tropis pada platform media sosial	<i>Latent Dirichlet Allocation</i>	Tweet Twitter tahun 2021 terkait penyakit tropis (2737 data)	Menghasilkan <i>coherence score</i> sebesar 0.576453 dan memutuskan banyaknya topik adalah 5

Penelitian pertama yang dilakukan oleh Yayang Matira, Junaidi, Iman Setiawan, 2023, “Pemodelan Topik Pada Judul Berita Online Detikcom Menggunakan *Latent Dirichlet Allocation*” . Tujuan dari penelitian ini adalah menghasilkan topik-topik yang berbeda melalui pola tertentu. Hasilnya, penelitian ini mendapatkan bahwa terdapat 3 topik dengan *coherence score* sebesar 0,7586. Adapun kesimpulan dari setiap topiknya adalah topik pertama mengenai konflik dan krisis suatu negara, topik kedua mengenai isu kemanusiaan dan topik ketiga mengenai isu korupsi pejabat negara. Saran untuk penelitian selanjutnya menggunakan tahapan stemming pada saat praproses data (Matira & Setiawan, 2023).

Razief Perucha Fauzie Afidh dan Syahrial pada tahun 2023 dengan judul “Pemodelan Topik Menggunakan n-Gram dan *Non-negative Matrix Factorization*” melakukan penelitiannya kepada dua portal berita yaitu RMOL.id dan BeritaSatu.com. Hasil implementasi metode untuk RMOL.id adalah 15 topik

dengan *coherence score* 0.812748 untuk unigram, 10 topik dengan *coherence score* 0.835738 untuk bigram, dan 7 topik dengan *coherence score* 0.830572 untuk tigram. Adapun BeritaSatu.com memperoleh 10 topik untuk unigram dengan *coherence score* 0.799718, 15 topik untuk bigram *coherence score* 0.788762 sedangkan untuk tigram menghasilkan 15 topik dengan *coherence score* 0,801935. Hasil ini dikatakan sangat baik karena didukung dengan proses *preprocessing* yang mumpuni (Afidh & Syahrial, 2023).

Pada tahun 2023, penelitian yang dilakukan oleh Herwinsyah, “Pemodelan Topik Dalam Al-Qur’an Menggunakan *Library* BERTopic pada Model Bahasa BERT”. Hasil dari penelitian ini menunjukkan terdapat 8 topik utama yang diperoleh dari pendekatan pemodelan topik oleh BERTopic diantaranya; Al Qur’an (6%), Aku/Allah (6,5%), Langit (3.8%), Rasul (8%), Malaikat (12,5%) , Wanita (5%), Neraka (13%), Dibangkitkan (5,5%). Kata-kata tersebut merepresentasikan topik yang termasuk dalam kategori spritual,moral, dan hukum (Herwinsyah, 2023).

Penelitian berikutnya oleh Gibran Giffari Priyatna, 2022, “Pemodelan Topik Terkait Ulasan Video Game Dengan Genre *Battle Royale* Menggunakan Metode BERTopic Dengan Fitur *Guided Topic Modelling*”. Pemodelan topik menggunakan metode BERTopic dengan fitur *guided topic modelling* yaitu penentuan topik pada proses awal. BERTopic menghasilkan 10 topik dari 55.203 ulasan dengan *coherence score* yang cukup memuaskan yaitu 0.5244758866. Namun *running time* nya lebih lama dibandingkan algoritma pemodelan topik lainnya yaitu selama 10.7 menit (Priyatna, 2022).

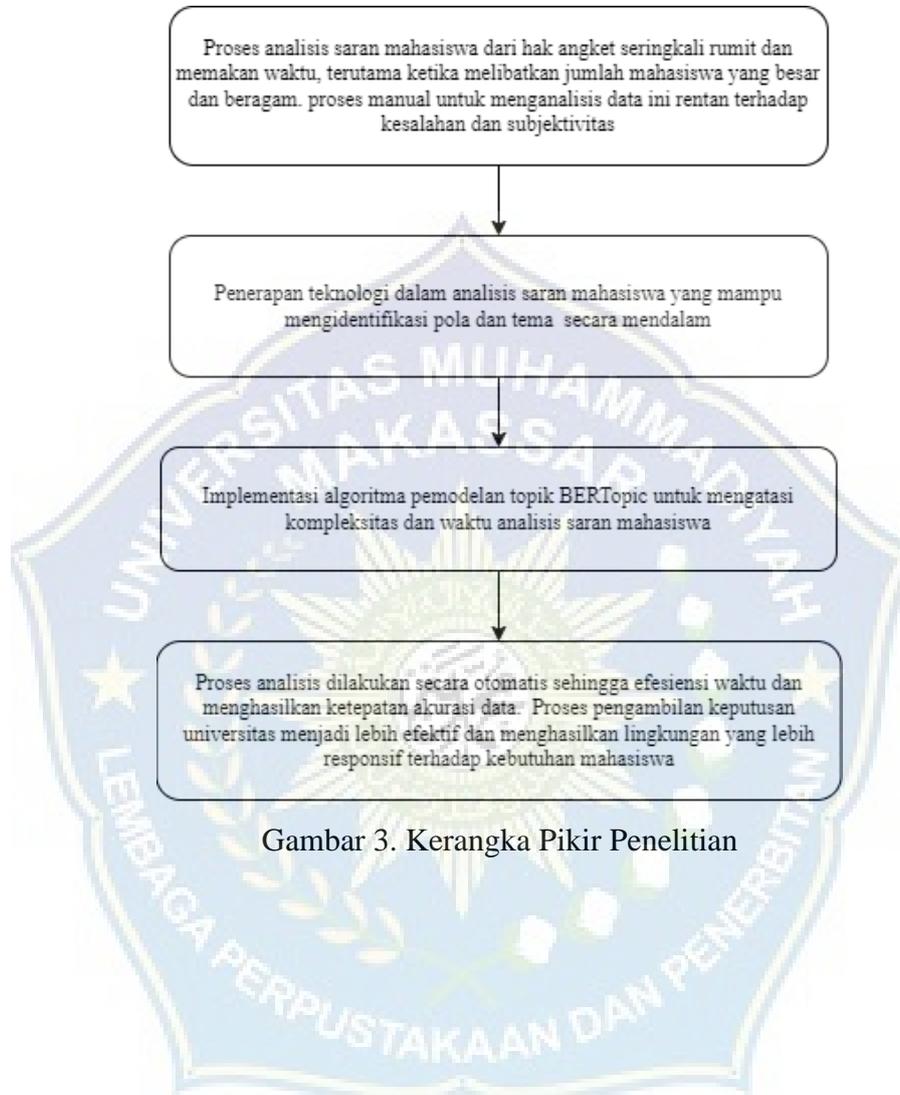
Selanjutnya penelitian oleh Dziky Ridhwanullah, 2022, “Pemodelan Topik Pada Cuitan Tentang Penyakit Tropis Di Indonesia Dengan Metode *Latent Dirichlet Allocation*”. proses analisis pemodelan topik menggunakan *metode Latet Dirichlet Allocation* yang dilakukan pada data *tweet*. Diperoleh sebanyak 2737 data *tweet* melalui pengambilan data menggunakan teknik *streaming*. *Coherence score* digunakan untuk menentukan jumlah topik terbaik kemudian menganalisis hasil tiap topiknya secara kualitatif. LDA mendapatkan *coherence score* sebesar 0.564049 dan menemukan sebanyak 5 topik yaitu dana untuk penyakit demam

berdarah dan malaria, buta dan kusta, pengobatan dan penanggulangan, perawatan, covid-19 (Ridhwanulah & Fudholi, 2022).



D. Kerangka Pikir

Dibawah ini adalah kerangka penelitian yang digambarkan sebagai berikut:



Gambar 3. Kerangka Pikir Penelitian

BAB III

METODOLOGI PENELITIAN

A. Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di Universitas Muhammadiyah Makassar karena menjadi sumber data penelitian, yakni saran-saran mahasiswa terkait fasilitas dan pelayanan di universitas tersebut. Penelitian ini dimulai pada bulan Mei hingga Juli 2024.

B. Alat dan Bahan Penelitian

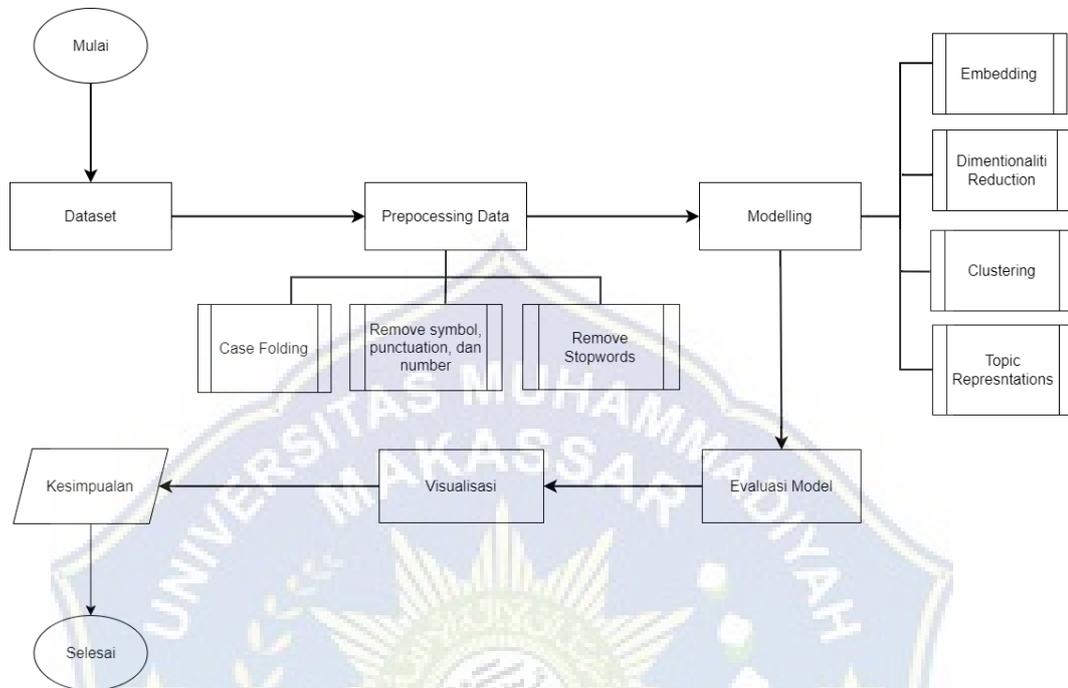
Pada penelitian ini, peneliti membutuhkan perangkat keras dan perangkat lunak sebagai berikut:

1. Perangkat Keras (*Hardware*)
 - a. Laptop Dell, Intel Core-I5 Ram 8 GB
2. Perangkat Lunak (*Software*)
 - a. OS WINDOWS 10
 - b. Excel
 - c. Google Colab
 - d. Bahasa pemrograman Python

C. Perancangan Sistem

Perancangan sistem adalah serangkaian tahapan yang disusun secara sistematis sebagai panduan dalam menyelesaikan permasalahan dan melakukan analisis terhadap penelitian. Tujuannya adalah menciptakan suatu kerangka kerja yang terstruktur dan efisien, sehingga memungkinkan pelaksanaan yang lancar dari setiap tahapan dalam menyelesaikan permasalahan dan menganalisis hasil penelitian secara sistematis dan efektif. Ini merupakan langkah krusial dalam proses penyelesaian permasalahan dan analisis hasil penelitian, karena membantu dalam merancang kerangka kerja yang terstruktur dan efisien serta memastikan bahwa solusi yang dihasilkan sesuai dengan tujuan yang diinginkan.

Dibawah ini merupakan perancangan sistem yang dilakukan oleh peneliti selama penelitian sebagaimana yang terlihat pada flowchart berikut



Gambar 4. *Flowchart* Sistem

1. Pengambilan Dataset

Tahap pertama adalah pengambilan Dataset yang diperoleh dari tanggapan pengisian saran mahasiswa pada simak UNISMUH. Jenis data ini dikategorikan sebagai sekunder karena akan terus bertambah seiring berjalannya semester perkuliahan dan masih diberlakukannya angket tersebut. Data yang diperoleh disajikan dalam format CSV dengan total 232.430 baris dan 4 kolom yakni: id, kategori, saran dan tgl_insert. Selanjutnya data yang diperoleh akan disortir dengan hanya mengambil data saran mahasiswa.

2. *Preprocessing Data*

Data dalam format *csv* akan melalui tahap *preprocessing* sebelum *modelling* dilakukan. *Preprocessing* pada penelitian ini menggunakan modul *regularexpression* pada bahasa pemrograman python. Pertama adalah menghilangkan data saran yang memiliki satu kata atau memiliki kurang dari empat

huruf. Setelah itu, dilanjutkan dengan proses penyesuaian huruf kecil (*Case Folding*), penghapusan simbol, angka dan tanda baca. Tahap terakhir adalah menghapus kata-kata yang dianggap tidak memberikan kontribusi signifikan dalam analisis, yang dikenal sebagai *stopword removal*.

3. Topic Modelling

Apabila *Preprocessing* data selesai, langkah selanjutnya adalah pemodelan. *Clean* data memasuki tahap prosedur BERTopic. Tahap pertama yang dilakukan adalah *embedding*, yaitu setiap kata dalam ulasan akan diberi bobot sesuai dengan makna semantiknya dalam kalimat. Hasilnya adalah kumpulan kata dalam bentuk matriks berdimensi tinggi yang akan diproses menggunakan algoritma clustering. Sebelum proses clustering dilakukan, matriks berdimensi tinggi tersebut perlu dikurangi dimensinya menjadi data berdimensi rendah melalui algoritma dimensionality reduction. Pada penelitian ini digunakan metode UMAP untuk tujuan tersebut.

Setelah data berhasil direduksi dimensinya, langkah berikutnya adalah melakukan proses clustering menggunakan algoritma HDBSCAN. Proses ini akan mengelompokkan kata-kata dalam data menjadi kelompok-kelompok yang merepresentasikan topik-topik yang dibicarakan dalam saran-saran tersebut. Setelah itu, data yang semula berbentuk matriks akan dikembalikan ke bentuk semula dan dilakukan tokenisasi.

Tahap terakhir adalah memberikan bobot pada setiap kata dalam setiap cluster menggunakan algoritma c-TF-IDF. Bobot ini akan menunjukkan sejauh mana kata tersebut terkait dengan topik tertentu dan tidak dengan topik lainnya

4. Visualisasi

Hasil ekstraksi topik-topik berdasarkan data akan direpresntasikan dalam bentuk diagram batang (*bartchart*) untuk menampilkan kata-kata yang mewakili setiap topik beserta bobotnya. Selain visualisasi topik, visualisasi persebaran cluster dalam data juga direpresentasikan untuk mengevaluasi kinerja algoritma clustering yang digunakan dalam penelitian ini

D. Teknik Pengujian Sistem

Pengujian sistem sangat penting karena memberikan gambaran seberapa baik kinerja model. Selain itu, pengujian juga memungkinkan perbandingan antara model dengan model lainnya. Pengujian sistem penelitian ini menggunakan evaluasi model melalui *running time* dan *coherence score*. *Running time* menilai seberapa cepat model dapat memprediksi topik dari sebuah dokumen atau dataset. Evaluasi ini penting karena kecepatan prediksi menjadi pertimbangan utama saat model diterapkan dalam program.

Selain kecepatan, ketepatan prediksi model juga menjadi aspek penting. Salah satu cara untuk mengukur ketepatan atau kinerja sebuah model “*Topic Modelling*” adalah dengan mengukur topic coherence dari model tersebut (Kapadia, 2019). Konsep kohesi topik melibatkan penggabungan sejumlah matrik dalam sebuah kerangka kerja untuk mengevaluasi keterkaitan antar topik yang dihasilkan oleh model. Menurut Mimno (2011), *topic coherence* dapat dihitung dengan membandingkan pasangan kata-kata dalam suatu topik; semakin tinggi nilai kohesi, semakin baik modelnya (ArRosyid, 2022). Berikut cara menghitung *coherence score* dengan pendekatan *Intrinsic Umass Measure* (Priyatna, 2022):

$$SCORE_{UMass}(W_i, W_j) = \log \frac{D(W_i, W_j) + 1}{D(W_i)} \quad (3)$$

D = Total jumlah dokumen dalam korpus

$D(W_i)$ = Total jumlah dokumen yang mengandung kata W_i

$D(W_i, W_j)$ = Total jumlah dokumen yang mengandung kata W_i dan W_j

E. Teknik Analisis Data

Setelah data yang diperlukan untuk penelitian terkumpul, tahap analisis data menjadi salah satu proses penting dalam penelitian. Kesalahan dalam analisis data dapat menyebabkan risiko kesalahan berantai, dimana kesalahan awal dalam analisis data dapat mempengaruhi hasil penelitian secara keseluruhan. Hal ini dapat

mengarah pada penyebaran informasi yang tidak akurat atau bahkan pada pembuatan keputusan yang berisiko.

Teknik analisis data merujuk pada metode atau pendekatan yang digunakan untuk menganalisis dan menginterpretasi data yang telah dikumpulkan. Secara umum, terdapat dua jenis teknik analisis data dalam penelitian, yaitu analisis data kuantitatif dan kualitatif. Teknik analisis data kuantitatif digunakan untuk memproses data numerik, sementara teknik analisis data kualitatif digunakan untuk memproses data non-numerik (Ulfah et al., 2022).

Dalam penelitian ini, digunakan teknik analisis data kuantitatif dengan pendekatan analisis deskriptif. Analisis statistik deskriptif bertujuan untuk mengevaluasi data dengan menyajikan informasi melalui tabel, grafik, atau diagram serta mengukur pemusatan, nilai tempat dan penyebaran data (Martias, 2021). Penggunaan teknik analisis statistik deskriptif ini dapat mempermudah visualisasi data dan memperoleh pemahaman yang lebih baik terhadap hasil penelitian.



BAB IV

PEMBAHASAN

Bab ini menjelaskan dan menampilkan hasil dari proses pemodelan topik yang telah dilakukan, mulai dari pengolahan data, analisa pemodelan topik, evaluasi performa dari algoritma pemodelan topik dan visualisasi hasil pemodelan.

A. Pengolahan Data

Data saran mahasiswa yang telah dikumpulkan disimpan dalam format CSV. Selanjutnya, data dalam format CSV tersebut akan diproses dalam tahap *preprocessing*. Karena data yang digunakan adalah teks tidak terstruktur, maka *preprocessing* yang dilakukan meliputi *case folding*, penghapusan angka, penghapusan emotikon, dan penghapusan kata-kata umum (*stopwords*). Hasil dari tahap *preprocessing* dapat dilihat pada tabel berikut.

Tabel 2 Hasil *Preprocessing* Data Mahasiswa

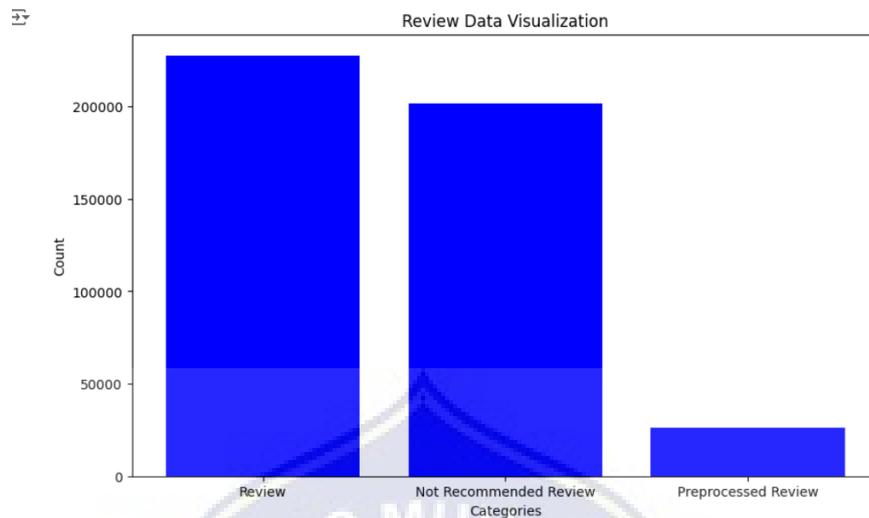
<i>Review</i>	<i>Preprocessed review</i>
<p>Saya berharap pelayanan akademik mampu memberikan pelayanan yang baik kepada mahasiswa dan seluruh masyarakat Unismuh. Terutama untuk para bapak/ibu dosen sekiranya mampu memberikan metode yang baik bagi mahasiswa sehingga mahasiswa lebih aktif dan produktif dalam proses pembelajaran.</p>	<p>berharap pelayanan akademik pelayanan mahasiswa masyarakat unismuh terutama dosen metode mahasiswa mahasiswa aktif produktif proses pembelajaran</p>
<p>Meningkatkan ruang lingkup belajar bagi mahasiswa seni</p>	<p>Meningkatkan ruang lingkup belajar mahasiswa seni</p>

SARAN SAYA YAITU AKAN perkuliahan luring mahasiswa paham
LEBIH BAIK JIKA PERKULIAHAN materi
DILAKUKAN SECARA LURING
AGAR MAHASISWA LEBIH
PAHAM DENGAN MATERI YANG
DIBERIKAN

Pada tabel diperlihatkan perbedaan yang signifikan pada data ulasan sebelum dan sesudah dilakukan *preprocessing*. Sebelum *preprocessing*, banyak ulasan mengandung emoji, kata-kata yang berulang, serta tanda baca yang berlebihan. Setelah *preprocessing* elemen-elemen ini telah dihapus, membuat data menjadi lebih bersih dan terstruktur.

Proses *case folding* juga terbukti berhasil, karena pada kolom saran yang telah dibersihkan tidak ada lagi kata yang menggunakan huruf kapital, semuanya diubah menjadi huruf kecil. Ini penting untuk memastikan konsistensi dalam analisis teks, terutama saat melakukan pencocokan kata.

Selain itu, karakter tidak valid, seperti simbol-simbol aneh atau karakter yang tidak dapat dibaca, juga telah dihapus. Penghapusan angka dari ulasan dilakukan untuk fokus pada analisis kata-kata, mengingat angka sering kali tidak relevan dalam konteks ini. Proses penghapusan stopwords, atau kata-kata umum yang tidak memberikan banyak informasi, seperti “dan,” “atau,” dan “yang,” juga dilakukan untuk meningkatkan kualitas analisis teks dengan mengurangi *noise*.



Gambar 5 Perbandingan Jumlah kata Sebelum dan Sesudah *Preprocessing*

Gambar 5 menunjukkan histogram yang membandingkan jumlah saran pada data sebelum dan sesudah *preprocessing*. Perbedaan jumlah saran sangat mencolok, dengan jumlah saran yang muncul setelah *preprocessing* jauh lebih sedikit dibandingkan sebelum *preprocessing*. Awalnya, jumlah ulasan mencapai 227.650 baris, tetapi setelah *preprocessing* jumlah ini berkurang menjadi 26.009 baris.

Karena banyak ulasan hanya berisi komponen-komponen yang tidak dibutuhkan, sebagian besar data ulasan dihapus selama *preprocessing*. Ini mengakibatkan penurunan drastis dalam jumlah saran yang tersedia untuk analisis lebih lanjut, seperti yang ditunjukkan oleh histogram pada gambar. Proses ini penting untuk memastikan bahwa data yang dianalisis adalah data yang relevan dan berkualitas tinggi, yang akan memberikan hasil yang lebih akurat dan bermakna dalam analisis berikutnya.

B. Analisis Pemodelan Topik

Pada subbab ini akan dilakukan analisis terhadap kumpulan kata-kata dalam topik yang dihasilkan dari penerapan algoritma BERTopic pada data saran mahasiswa di UNISMUH. Algoritma BERTopic yang diterapkan pada data saran mahasiswa dalam penelitian ini menghasilkan 9 topik dari 26.009 saran mahasiswa.

```

seed_topics = [
    ["dosen", "pembelajaran", "kuliah", "perkuliahan",
     "materi", "pengajaran", "sistem", "RPS"],
    ["prasarana", "fasilitas", "ruang", "kampus",
     "laboratorium", "peralatan", "kelas", "gedung", "parkiran"],
    ["administrasi", "pelayanan", "pencatatan", "tata usaha"],
    ["sarana", "Buku", "fasilitas", "meja", "kursi", "papan
     tulis", "proyektor"],
    ["internet", "jaringan", "komputer", "kuota"],
    ["kebersihan", "keamanan", "petugas", "sampah"],
    ["biaya", "beasiswa", "keuangan", "uang"],
]

```

Baris code diatas menginisialisasi sebuah daftar yang berisi beberapa daftar topik awal (`seed_topics`) untuk pemodelan topik. Setiap sub-daftar dalam `seed_topics` berisi kata-kata yang terkait dengan tema tertentu. Seed topic dalam algoritma BERTopic digunakan untuk meningkatkan kontrol dan interpretabilitas dalam analisis topik. Tujuan penggunaan seed topic adalah untuk memberikan panduan awal kepada model dalam menentukan topik-topik utama yang relevan dengan konteks data yang sedang dianalisis. Topik pertama mencakup kata-kata seperti “dosen”, “pembelajaran”, “kuliah”, “perkuliahan”, “materi”, “pengajaran”, “sistem”, dan “RPS”, yang semuanya berkaitan dengan proses pengajaran dan pembelajaran di institusi pendidikan, termasuk peran dosen dan sistem pengajaran. Topik kedua berfokus pada fasilitas kampus, mencakup kata-kata seperti “prasarana”, “fasilitas”, “ruang”, “kampus”, “laboratorium”, “peralatan”, “kelas”, “gedung”, dan “parkiran”, yang semuanya berhubungan dengan berbagai fasilitas fisik yang tersedia di kampus.

Topik ketiga terkait dengan administrasi dan pelayanan, termasuk kata-kata seperti “administrasi”, “pelayanan”, “pencatatan”, dan “tata usaha”, yang berhubungan dengan aspek administratif dan layanan yang diberikan oleh departemen administrasi di kampus. Topik keempat mencakup sarana pembelajaran, dengan kata-kata seperti “sarana”, “Buku”, “fasilitas”, “meja”, “kursi”, “papan tulis”, dan “proyektor”, yang mendukung kegiatan pembelajaran. Topik kelima berkaitan dengan teknologi dan internet, termasuk kata-kata seperti “internet”, “jaringan”, “komputer”, dan “kuota”, yang berhubungan dengan aspek teknologi informasi dan komunikasi.

Topik keenam menyoroti kebersihan dan keamanan, mencakup kata-kata seperti “kebersihan”, “keamanan”, “petugas”, dan “sampah”, yang terkait dengan aspek kebersihan dan keamanan di lingkungan kampus. Terakhir, topik ketujuh berfokus pada biaya dan keuangan, mencakup kata-kata seperti “biaya”, “beasiswa”, “keuangan”, dan “uang”, yang berhubungan dengan aspek finansial, termasuk biaya pendidikan dan manajemen keuangan. Secara keseluruhan, daftar `seed_topics` ini mencakup berbagai aspek penting yang relevan dengan kehidupan di kampus, dan kata-kata yang tercakup dalam setiap topik membantu dalam pemodelan topik yang lebih terfokus dan terstruktur.



```

from sentence_transformers import SentenceTransformer
import umap
from hdbscan import HDBSCAN
from bertopic import BERTopic
from bertopic.vectorizers import ClassTfidfTransformer
from sklearn.feature_extraction.text import CountVectorizer

sentence_model = SentenceTransformer("paraphrase-multilingual-
MiniLM-L12-v2")
umap_model = umap.UMAP(n_neighbors=15, n_components=5,
min_dist=0.0, metric='cosine')
hdbscan_model = HDBSCAN(min_cluster_size=15,
metric='euclidean', cluster_selection_method='eom',
prediction_data=True)
vectorizer_model = CountVectorizer()
ctfidf_model = ClassTfidfTransformer(bm25_weighting=True,
reduce_frequent_words=True)

model = BERTopic (
    embedding_model=sentence_model,
    umap_model=umap_model,
    hdbscan_model=hdbscan_model,
    vectorizer_model=vectorizer_model,
    ctfidf_model=ctfidf_model,
    seed_topic_list=seed_topics,
    language='multilingual',
    nr_topics= 10,
    min_topic_size=200,
    calculate_probabilities=True,
    verbose=True
)

topics, probabilities =
model.fit_transform(df['saran_without_stopwords'])
df["topics"] = topics

```

Baris code diatas menunjukkan parameter yang digunakan selama proses algoritma BERTopic berlangsung. Untuk menghasilkan embedding untuk teks dalam berbagai bahasa, pada “SentenceTransformer” digunakan model pre-trained “paraphrase-multilingual-MiniLM-L12-v2”. Inisialisasi Model UMAP untuk reduksi dimensi embedding menjadi 5 komponen dengan 15 tetangga terdekat,

menggunakan metrik cosine. Pada model clustering HDBSCAN menggunakan metrik euclidean, dan metode pemilihan cluster “eom” (excess of mass). “predicition_data=True” memungkinkan prediksi data baru. “CountVectorizer” untuk menghitung frekuensi kata dan “ClassTfidfTransformer” dengan pembobotan BM25 dan pengurangan kata yang sering muncul. Adapun parameter dari model BERTopic adalah:

- `embedding_model`: Model embedding yang digunakan.
- `umap_model`: Model UMAP untuk reduksi dimensi.
- `hdbscan_model`: Model HDBSCAN untuk clustering.
- `vectorizer_model`: Model untuk menghitung frekuensi kata.
- `ctfidf_model`: Model untuk transformasi TF-IDF berbasis kelas.
- `seed_topic_list`: Daftar topik awal yang disediakan.
- `language`: Bahasa teks yang digunakan.
- `nr_topics`: Jumlah topik yang diinginkan.
- `min_topic_size`: Ukuran minimal sebuah topik.
- `calculate_probabilities`: Menghitung probabilitas topik untuk dokumen.
- `verbose`: Menampilkan informasi proses.

Dengan parameter yang ditampilkan, algoritma BERTopic menghasilkan 9 topik dari 26.009 data saran yang ada. Topik-topik yang dihasilkan selanjutnya akan dianalisis sebagai berikut :

1. Topik 0 (pelayanan_kelas_kuliah_sarana)

```
model.get_topic(0)

[('pelayanan', 0.29457275672714556),
 ('kelas', 0.28133153008977174),
 ('kuliah', 0.2776628231653325),
 ('sarana', 0.2660628420662317),
 ('mengajar', 0.2611157374188851),
 ('pembelajaran', 0.258245646461027),
 ('metode', 0.2547374431941906),
 ('proses', 0.2491421447678579),
 ('dosen', 0.24761202451891015),
 ('akademik', 0.24588092843797815)]

model.get_representative_docs(0)

['dosen aktif perkuliahan materi maksimal diharapkan dosen masuk sesuai jadwal
perkuliahan',
 'rangka peningkatan pelayanan akademik program studi proses pembelajaran efektif efisien
melaksanakan proses pembelajaran luring',
 'rangka peningkatan pelayanan akademik program studi menyediakan sarana pembelajaran
memadai mendukung pembelajaran mahasiswa']
```

Gambar 6. Kumpulan Kata dan Representasi Dokumen Topik 0

Gambar diatas memperlihatkan kumpulan kata yang muncul dalam topik 0. Kata-kata kunci yang paling relevan dengan topik ini mencakup “pelayanan,” “kelas,” “kuliah,” “sarana,” “mengajar,” dan lain-lain. Ini menunjukkan bahwa topik ini mencakup berbagai aspek seperti layanan akademik, fasilitas pendidikan, bahan ajar, media pembelajaran, motivasi siswa, sistem pendidikan, referensi, kelas, diskusi, dan metode pembelajaran.

Melalui representasi dokumen pada topik ini menyoroti upaya peningkatan layanan akademik, seperti penyediaan sarana pembelajaran memadai, penciptaan suasana belajar yang efektif, serta peningkatan kualitas materi dan metode pembelajaran. Topik ini juga terkait dengan seed topik pertama yang telah ditetapkan sebelumnya.

2. Topik 1 (mencatat_materi_pertahankan_harap)

```
model.get_topic(1)
```

```
[('mencatat', 0.6126900084569501),  
( 'materi', 0.5345778232074132),  
( 'pertahankan', 0.4988042530655833),  
( 'harap', 0.490529713754417),  
( 'menyarankan', 0.4710037465326109),  
( 'perlahan', 0.44997910825793463),  
( 'dipertahankan', 0.4442372138370153),  
( 'ilmu', 0.40228524352152983),  
( 'cepat', 0.37720118196760877),  
( 'pelayanan', 0.3663654012720174)]
```

```
model.get_representative_docs(1)
```

```
['materi cepat mencatat materi mencatat memahami materi menyarankan materi perlahan',  
'materi cepat mencatat materi mencatat memahami materi menyarankan materi perlahan',  
'materi cepat mencatat materi mencatat memahami materi menyarankan materi perlahan']
```

Gambar 7. Kumpulan Kata dan Representasi Dokumen Topik 1

Kata-kata seperti “mencatat”, “materi”, “pertahankan”, “harap”, dan lain lain memiliki skor relevansi yang tinggi, menunjukkan bahwa kata-kata ini adalah yang paling penting dalam mendefinisikan Topik 1. Kata “mencatat” memiliki skor tertinggi, yaitu 0.633, yang menandakan pentingnya kata ini dalam topik tersebut. Dokumen-dokumen representasi berisi kalimat seperti “materi cepat mencatat materi mencatat memahami materi menyarankan materi perlahan”, yang mencerminkan kata-kata kunci dari Topik 1. Dengan demikian, Topik 1 berkaitan dengan proses mencatat dan memahami materi, serta memberikan saran tentang cara mencatat dengan perlahan. Kombinasi dari kata-kata kunci dan dokumen representatif ini menunjukkan fokus pada metode pembelajaran yang mengharuskan mencatat tapi kurang disukai jika terlalu cepat.

3. Topik 2 (offline_zoom_online_jaringan)

```
model.get_topic(2)
```

```
[('offline', 0.7337303260797813),  
 ('zoom', 0.6016626560790141),  
 ('online', 0.5518098880313894),  
 ('jaringan', 0.4885992854752322),  
 ('google', 0.4469172252468788),  
 ('meet', 0.43189051497260206),  
 ('lakukan', 0.3860802767819437),  
 ('wifi', 0.38147697717826295),  
 ('via', 0.38032309585403895),  
 ('whatsapp', 0.35242825334986705)]
```

```
model.get_representative_docs(2)
```

```
['pembelajaran akademik berjalan sesuai mesti dilaksanakan kuliah offline minimal lakukan  
pembelajaran pertemuan online pertemuan offline karna kuliah online mahasiswa bosan acuh  
pembelajaran ilmu pembejaraan luring efektif',  
 'proses daring dosen memahami kendala mahasiswa terkait jaringan lbh efektif lakukan  
google meeting bandingkan via zoom',  
 'pertemuan perkuliahan kuliah offline proses pembelajaran mudah mahasiswa mudah menyerap  
memahami materi pemebelajaran dibandingkan kuliah pertemuan online via google meet']
```

Gambar 8. Kumpulan Kata dan Representasi Dokumen Topik 2

Gambar diatas merupakan daftar kata-kata dengan skor yang menunjukkan seberapa kuat setiap kata berhubungan dengan topik yang diidentifikasi sebagai topik 2. Kata-kata kunci yang paling relevan dengan topik ini mencakup “offline,” “zoom”, “online,” “jaringan”, dan lain lain. Ini menunjukkan bahwa topik ini mencakup berbagai aspek seperti metode pembelajaran offline dan online, kendala jaringan, serta efektivitas pembelajaran.

Representasi dokumen yang dihasilkan menyoroti berbagai isu terkait pembelajaran, seperti proses perkuliahan yang terkendala jaringan, saran untuk melaksanakan perkuliahan offline agar pemahaman materi lebih baik, serta efektivitas pembelajaran luring. Secara keseluruhan, hasil ini menunjukkan bahwa topik yang diidentifikasi berhubungan dengan metode pembelajaran offline dan online, kendala jaringan, serta upaya untuk meningkatkan efektivitas pembelajaran.

4. Topik 3 (islam_muhammadiyah_agama_arab)

```
model.get_topic(3)
```

```
[('islam', 0.6410177225341698),  
 ('indonesia', 0.6191561267615112),  
 ('bahasa', 0.591705811828821),  
 ('agama', 0.5526675436033265),  
 ('muhammadiyah', 0.5404485620493439),  
 ('arab', 0.49054773499229964),  
 ('allah', 0.4821499022087233),  
 ('universitas', 0.45123622454838996),  
 ('islami', 0.44880747294690754),  
 ('ustadz', 0.4419410044402053)]
```

```
model.get_representative_docs(3)
```

```
['bahasa arab dipelajari lingkup mahad memudahkan memahami agama islam alquran',  
 'rangka pelayanan akademik program studi pendidikan bahasa sastra indonesia',  
 'program studi pendidikan agama islam dosen pelayanan akademik']
```

Gambar 9. Kumpulan Kata dan Representasi Dokumen Topik 3

Daftar kata kunci yang disertai dengan skor relevansi menunjukkan berbagai topik dengan relevansi masing-masing. Misalnya, kata kunci seperti “islam,” “indonesia,” , “bahasa” dan agama memiliki skor relevansi yang tinggi, menandakan keterkaitannya yang kuat dengan topik yang dianalisis. Dokumen representatif tersebut menyoroti aspek-aspek penting seperti pengajaran bahasa Arab di lingkup pendidikan agama Islam, serta pelayanan akademik yang terkait dengan program studi pendidikan, khususnya dalam konteks pendidikan bahasa dan agama di Indonesia. Kesimpulannya, topik nomor 3 yang dianalisis oleh model ini fokus pada pendidikan agama Islam dan pengajaran bahasa Arab di lingkungan akademik Indonesia, dengan penekanan pada kontribusi institusi seperti Muhammadiyah dalam bidang ini.

5. Topik 4 (fasilitas_kipas_kampus_kursi)

```
model.get_topic(4)
```

```
[('fasilitas', 0.8173032295966745),  
 ('kipas', 0.7949801580863124),  
 ('kampus', 0.7685085656545072),  
 ('kursi', 0.6921566340781871),  
 ('ruangan', 0.6536471469606231),  
 ('angin', 0.6286410250167361),  
 ('smoga', 0.6257077036062791),  
 ('parkiran', 0.5266210453685276),  
 ('meja', 0.522124011962077),  
 ('lift', 0.4826645385984547)]
```

```
model.get_representative_docs(4)
```

```
['fasilitas kelas lengkapi kipas angin peradakan jg kipas ruangan lab',  
 'fasilitas ruang kelas kipas angin kursi rusak diganti diperbaiki',  
 'fasilitas kampus diperhatikan kenyamanan belajar d dlm kelas alat pendingin kipas angin  
memadai ruangan']
```

Gambar 10. Kumpulan Kata dan Representasi Dokumen Topik 4

Gambar diatas merupakan analisis terkait fasilitas di kampus dengan fokus pada kenyamanan dan perlengkapan yang disediakan. Daftar kata kunci, seperti “fasilitas,” “kipas,” “kursi,” “kampus,” dan lain lain menunjukkan berbagai elemen yang berhubungan dengan lingkungan belajar di kampus. Kata kunci ini disertai dengan skor relevansi yang menunjukkan pentingnya setiap elemen dalam konteks fasilitas kampus.

Dokumen-dokumen representatif memberikan gambaran lebih jelas mengenai topik ini. Dokumen pertama menyebutkan agar melengkapi fasilitas seperti kipas angin yang ada di ruang perkuliahan. Dokumen kedua menjelaskan upaya kampus dalam mengganti meja dan kursi yang rusak untuk meningkatkan kenyamanan mahasiswa, serta menambah fasilitas di ruangan kelas. Dokumen ketiga menyoroti perhatian kampus terhadap fasilitas seperti alat pendingin dan kipas angin untuk mendukung kenyamanan belajar di dalam kelas. Secara keseluruhan, data ini menunjukkan perhatian pada berbagai aspek fasilitas yang mendukung kenyamanan dan kualitas lingkungan belajar di kampus.

6. Topik 5 (perpustakaan_buku_akuntansi_referensi)

```
model.get_topic(5)
```

```
[('perpustakaan', 1.0514307039435267),  
 ('buku', 0.8483092323677331),  
 ('akuntansi', 0.7419969776771693),  
 ('referensi', 0.707483034229182),  
 ('dana', 0.613601300251182),  
 ('uang', 0.5940713591911452),  
 ('biaya', 0.5717567417823456),  
 ('pembayaran', 0.5660392464004693),  
 ('bukubuku', 0.5466148786482069),  
 ('keuangan', 0.5326291379353371)]
```

```
model.get_representative_docs(5)
```

```
['menambah bukubuku jurnaljurnal d perpustakaan terkait hadist tematik tarbawi referensi  
bahan bacaan u menunjang pembuatan makalah jurnal sesuai materi perkuliahan',  
 'menambah bukubuku jurnaljurnal d perpustakaan referensi bahan bacaan u menunjang  
pembuatan makalah jurnal sesuai materi perkuliahan',  
 'menambah bukubuku jurnaljurnal d perpustakaan sesuai materi perkuliahan referensi bahan  
bacaan u menunjang pembuatan makalah jurnal']
```

Gambar 11. Kumpulan Kata dan Representasi Dokumen Topik 5

Kata-kata yang menunjukkan daftar kata kunci yang diidentifikasi sebagai bagian dari topik 6 yaitu “perpustakaan”, “buku”, “akutansi”, “referensi”. Pada dokumen representasi terlihat sangat berkaitan dengan kegiatan menambah buku dan jurnal di perpustakaan makalah atau jurnal yang sesuai dengan materi perkuliahan. Adanya penekanan pada penyediaan bahan bacaan yang relevan dan spesifik untuk kebutuhan akademik. Secara keseluruhan, topik ini terkait erat dengan pengelolaan perpustakaan akademik dan penyediaan sumber daya yang diperlukan untuk mendukung kegiatan belajar mengajar dan penelitian di universitas.

7. Topik 6 (pengumpulan_batas_memperpanjang_maksimal)

```
model.get_topic(6)
```

```
[('pengumpulan', 1.6364727097044596),  
 ('batas', 1.6080023252076987),  
 ('memperpanjang', 1.2911639496280483),  
 ('maksimal', 1.1538734352432263),  
 ('mengerjakannya', 1.1364513682065134),  
 ('tugas', 1.1349123453594134),  
 ('terbebani', 1.0843536924952089),  
 ('minggu', 0.920720494200172),  
 ('singkat', 0.900318680862461),  
 ('kadang', 0.7764808758643409)]
```

```
model.get_representative_docs(6)
```

```
['memperpanjang batas pengumpulan tugas terbebani mengerjakannya maksimal',  
 'memperpanjang batas pengumpulan tugas terbebani mengerjakannya maksimal',  
 'batas pengumpulan tugas tentukan cepat memperpanjang batas pengumpulan tugas terbebani mengerjakannya maksimal']
```

Gambar 12. Kumpulan Kata dan Representasi Dokumen Topik 6

Kata-kata kunci yang paling relevan dengan topik ini mencakup “pengumpulan”, “batas”, “memperpanjang”, “maksimal”, dan lain lain. Ini menunjukkan bahwa topik ini mencakup aspek waktu pengumpulan tugas. Dokumen paling representatif yang dihasilkan menyoroti masalah-masalah terkait dengan batas waktu pengumpulan tugas, termasuk penentuan batas waktu yang cepat, memperpanjang batas waktu, serta dampak dari batas waktu tersebut pada beban kerja mahasiswa. Secara keseluruhan, topik ini berkaitan dengan manajemen waktu, khususnya dalam konteks pengumpulan tugas akademik.

8. Topik 7 (suara_suaranya_mendengar_terdengar)

```
model.get_topic(7)
```

```
[('suara', 1.4260594761077525),  
 ('suaranya', 1.1930789487135773),  
 ('mendengar', 0.9650055772675155),  
 ('terdengar', 0.9516977183416735),  
 ('candaan', 0.8640470381947408),  
 ('intens', 0.8623657216660798),  
 ('mic', 0.7972983874690097),  
 ('kali', 0.6953733957384692),  
 ('kencang', 0.6542825295103262),  
 ('kelas', 0.6399049379358587)]
```

```
model.get_representative_docs(7)
```

```
['suara mengajar terdengar kelas mic disediakan kelas materi',  
 'suara suara terdengar bangku menambah volume suara mendengar nya',  
 'suara pelan mengajar mendengar pengeras suara suaranya terdengar penjurur kelas  
 membiarkan ruangan tertutup suara kedengaran mendengarkan penjelasan']
```

Gambar 13. Kumpulan Kata dan Representasi Dokumen Topik 7

Kata kunci pada topik 7 berkaitan dengan suara, penggunaan mic, intensitas suara. Dari dokumen representasi terlihat sangat berkaitan dengan masalah suara dalam konteks pengajaran di kelas. Ada masalah yang disebutkan terkait dengan sulitnya mendengar suara dosen dan menyarankan penggunaan mic. Ada juga diskusi tentang intensitas suara, seperti suara yang tidak terdengar dari bangku atau penggunaan pengeras suara untuk memastikan suara terdengar dengan jelas di seluruh kelas. Secara keseluruhan, topik ini berkaitan erat dengan tantangan komunikasi verbal dalam lingkungan kelas, serta pentingnya memastikan bahwa suara dosen terdengar dengan jelas untuk mendukung pemahaman materi oleh mahasiswa.

9. Topik 8 (kegagalan_menyerah_memperoleh_berhenti)

```
model.get_topic(8)
```

```
[('kegagalan', 1.8550747296237364),  
 ('menyerah', 1.6449516533254502),  
 ('memperoleh', 1.5296383841878511),  
 ('berhenti', 1.5108486173017475),  
 ('kesuksesan', 1.4523062534301834),  
 ('kegagalan tetap', 1.3482762813614917),  
 ('mengasah', 1.3032014480997158),  
 ('pendongkrak', 1.19645682728565),  
 ('terjawab', 1.1622415721466441),  
 ('kegagalanmu', 1.1372486250116451)]
```

```
model.get_representative_docs(8)
```

```
['mahasiswi universitas muhammadiyah makassar berhenti belajar mengasah kemampuan menyerah  
memperoleh kegagalan tetap kegagalan kesuksesan dengarkan nasihat orang menjatuhkan impian  
citacitamu optimis dan jadikan kegagalanmu pendongkrak untukmu maju menambah fasilitas  
kampus bangku mahasiswa i ganti sep',  
'berhenti belajar mengasah kemampuan menyerah memperoleh kegagalan tetap kegagalan  
kesuksesan dengarkan nasihat orang menjatuhkan impian citacitamu optimis dan jadikan  
kegagalanmu pendongkrak untukmu maju',  
'berhenti belajar mengasah kemampuan menyerah memperoleh kegagalan tetap kegagalan  
kesuksesan dengarkan nasihat orang menjatuhkan impian citacitamu optimis dan jadikan  
kegagalanmu pendongkrak untukmu maju']
```

Gambar 14. Kumpulan Kata dan Representasi Dokumen Topik 8

Daftar kata kunci pada topik 8 mencerminkan topik yang berkaitan dengan tantangan dan usaha untuk mencapai kesuksesan meskipun menghadapi kegagalan. Hal ini diperkuat dengan representasi dokumen yang membahas motivasi dan semangat untuk tidak menyerah dalam menghadapi kegagalan. Ada penekanan pada pentingnya mengasah kemampuan, tetap optimis dan menjadikan kegagalan sebagai pendongkrak untuk maju. Ada juga menekankan pentingnya mendengarkan nasihat orang lain, meskipun ada yang berusaha menjatuhkan impian kita, serta pentingnya terus berusaha untuk mencapai kesuksesan. Topik ini berkaitan erat dengan pesan motivasi dan dorongan untuk menyerah dalam menghadapi kesulitan dan kegagalan.

C. Evaluasi Performa Pemodelan Topik

Pada subbab ini akan ditunjukkan performa algoritma BERTopic dalam mengekstrak topik dari data saran mahasiswa di UNISMUH. Performa algoritma BERTopic akan dievaluasi berdasarkan kecepatan penyelesaian tugas dan besarnya *Coherence Score* yang dihasilkan. Data saran yang telah melewati tahap *preprocessing* akan diinput ke dalam algoritma BERTopic.

```
start = timeit.default_timer()

topics, probabilities = model.fit_transform(df['saran_without_stopwords'])
stop = timeit.default_timer()
print('Time: ', stop - start, 'seconds')

2024-07-18 11:43:28,250 - BERTopic - Embedding - Transforming documents to embeddings.
Batches: 100% ██████████ 814/814 [09:26<00:00, 5.31it/s]
2024-07-18 11:52:55,777 - BERTopic - Embedding - completed ✓
2024-07-18 11:52:55,779 - BERTopic - Guided - Find embeddings highly related to seeded topics.
Batches: 100% ██████████ 1/1 [00:00<00:00, 5.07it/s]
2024-07-18 11:52:56,207 - BERTopic - Guided - Completed ✓
2024-07-18 11:52:56,209 - BERTopic - Dimensionality - Fitting the dimensionality reduction algorithm
2024-07-18 11:53:34,600 - BERTopic - Dimensionality - Completed ✓
2024-07-18 11:53:34,604 - BERTopic - Cluster - Start clustering the reduced embeddings
/usr/local/lib/python3.10/dist-packages/joblib/externals/loky/backend/fork_exec.py:38: RuntimeWarning:
os.fork() was called. os.fork() is incompatible with multithreaded code, and JAX is multithreaded, so this will likely lead to a deadlock.
2024-07-18 11:56:35,467 - BERTopic - Cluster - completed ✓
2024-07-18 11:56:35,470 - BERTopic - Representation - Extracting topics from clusters using representation models.
2024-07-18 11:56:36,115 - BERTopic - Representation - Completed ✓
2024-07-18 11:56:36,122 - BERTopic - Topic reduction - Reducing number of topics
2024-07-18 11:56:36,791 - BERTopic - Topic reduction - Reduced number of topics from 271 to 10
Time: 788.806468453 seconds
```

Gambar 15. *Runningtime* Algoritma BERTopic

Gambar diatas menunjukkan bahwa *running time* yang diperoleh dari algoritma BERTopic, yang dijalankan pada perangkat dengan spesifikasi prosesor Core i5-7300U CPU @2.60GHz, RAM 8GB, dan VRAM 4GB, adalah **13 menit**. Meskipun waktu tersebut tergolong lama untuk pemodelan topik, hal ini dapat dimaklumi mengingat algoritma BERTopic berbasis arsitektur *neural network*.

Adapun perhitungan *coherence score* dilakukan dengan setiap dokumen dalam korpus diproses (*preprocessed*) untuk menghilangkan kata-kata yang tidak relevan, lalu dilakukan tokenisasi untuk memisahkan setiap kata dalam dokumen. Setelah itu, dibuat kamus yang berisi daftar kata dan ID-nya, serta corpus dalam format *bag of words* (BoW), yang merepresentasikan setiap dokumen sebagai daftar pasangan kata (ID, frekuensi).

Selanjutnya, kata-kata dari tiap topik diambil dan dihitung seberapa sering kata-kata tersebut muncul bersama (*co-occurrence*) dalam dokumen yang sama. Ini dilakukan dengan membangun matriks *co-occurrence* yang mencerminkan jumlah dokumen di mana pasangan kata dari topik tertentu muncul bersama. Misalnya,

untuk dua kata dalam satu topik, dihitung berapa banyak dokumen yang mengandung kedua kata tersebut.

Akhirnya, koherensi dihitung dengan menggabungkan frekuensi kemunculan bersama kata-kata tersebut. Biasanya, skor koherensi dihitung dengan mengambil rata-rata dari frekuensi kemunculan bersama kata-kata dalam satu topik. Skor yang lebih tinggi menunjukkan bahwa kata-kata dalam topik tersebut sering muncul bersama, menandakan topik yang lebih koheren.

```
coherence_score = calculate_coherence_score(model, df['saran_without_stopwords'])  
print(f"Coherence Score: {coherence_score}")
```

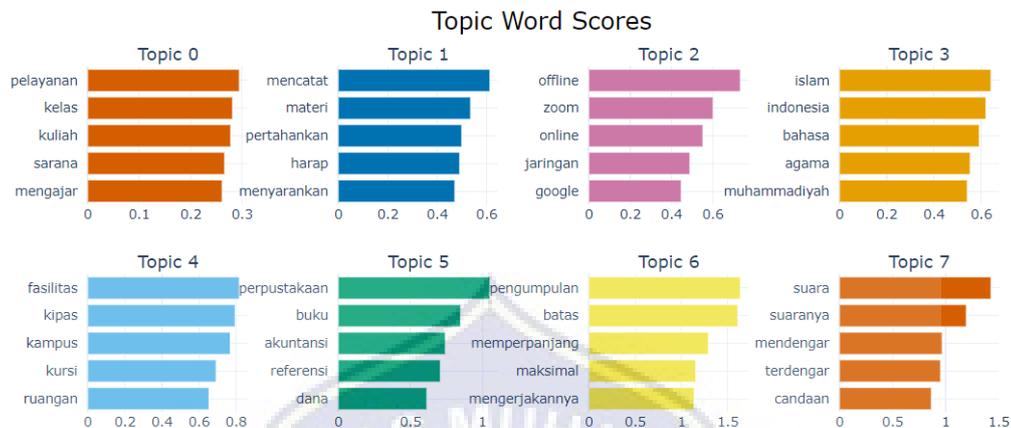
```
Coherence Score: 0.637722900338999
```

Gambar 16. *Coherence Score* Algoritma BERTopic

Gambar diatas menunjukkan *coherence score* yang diperoleh dari algoritma BERTopic dalam memproses data saran mahasiswa UNISMUH. Algoritma BERTopic berhasil mencapai coherence score sebesar **0,637722900338999**, yang tergolong cukup baik untuk sebuah algoritma pemodelan topik. Skor *coherence* memiliki rentang dari -1 hingga 1, di mana skor mendekati 1 menunjukkan bahwa topik yang dihasilkan sangat koheren dan mudah diinterpretasikan, sementara skor mendekati -1 menunjukkan topik yang kurang koheren dan sulit dipahami.

Skor ini menandakan bahwa topik yang dihasilkan tidak hanya relevan secara statistik tetapi juga secara semantik, membuatnya lebih mudah untuk diinterpretasikan oleh manusia. Dengan kata lain, topik-topik ini tidak hanya berisi kata-kata yang sering muncul bersama dalam data, tetapi juga kata-kata yang membentuk narasi atau tema yang koheren dan logis. Hal ini menunjukkan bahwa BERTopic tidak hanya mengidentifikasi kumpulan kata-kata yang sering muncul bersama, tetapi juga mampu menangkap makna dan konteks yang lebih dalam dari data yang dianalisis.

D. Visualisasi Pemodelan Topik

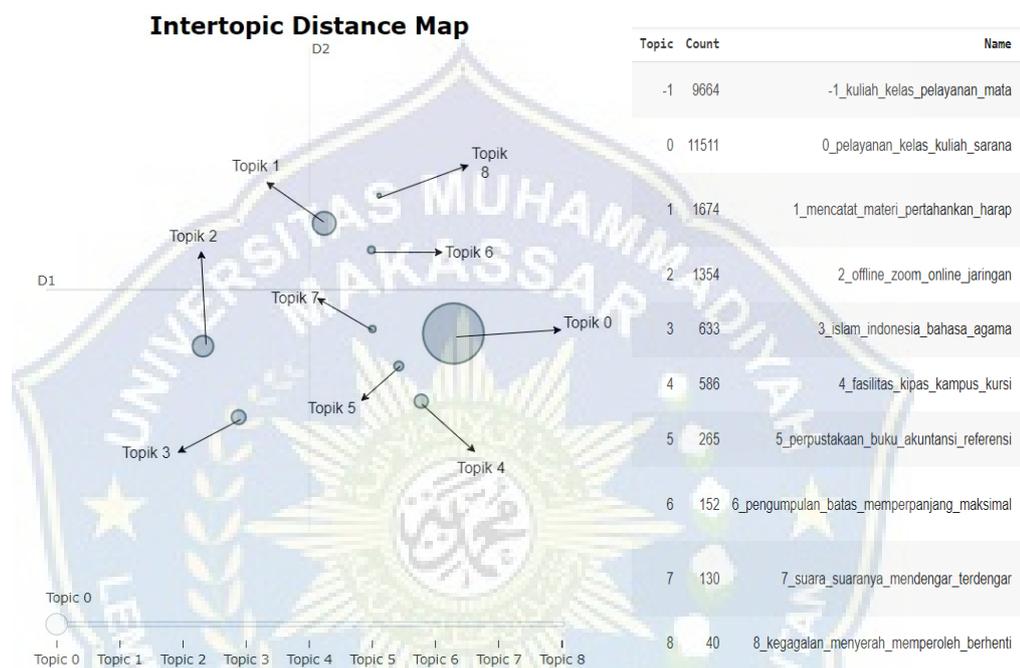


Gambar 17. Pengelompokan Berdasarkan Topik

Gambar diatas menunjukkan distribusi kata-kata yang terkait dengan berbagai topik berdasarkan “Topic Word Scores”. Diagram batang ini menggambarkan delapan topik (dari Topik 0 sampai Topik 7) beserta kata-kata utama yang paling relevan dan skor mereka untuk masing-masing topik. Setiap topik memiliki kata-kata yang paling sering muncul beserta skor yang menunjukkan pentingnya kata tersebut dalam topik tersebut. **Topik 0** berkaitan dengan layanan akademik, sarana, dan metode pembelajaran. Kata-kata yang sering muncul adalah “pelayanan”, “kelas”, “kuliah”, dan “sarana”. **Topik 1** terkait dengan aktivitas mencatat dan saran dengan kata-kata seperti “mencatat”, “materi”, “pertahankan”, “harap”, dan “menyarankan”. **Topik 2** mencakup kata-kata yang berkaitan dengan aktivitas online dan jaringan, seperti “offline”, “zoom”, “online”, “jaringan”, dan “google”. **Topik 3** berhubungan dengan Islam dan bahasa, dengan kata-kata seperti “islam”, “indonesia”, “bahasa”, “agama”, dan “muhammadiyah”. **Topik 4** mencakup fasilitas kampus dengan kata-kata seperti “fasilitas”, “kipas”, “kampus”, “kursi”, dan “ruangan”. **Topik 5** terkait dengan ketersediaan referensi dengan kata-kata seperti “perpustakaan”, “buku,” “akuntansi”, “referensi” dan “dana”. **Topik 6** berfokus pada jadwal dan pengumpulan dengan kata-kata seperti “pengumpulan”, “batas”, “memperpanjang”, dan “maksimal”. **Topik 7** berkaitan dengan suara

terlihat dari kata-kata “suara”, “suaranya”, ”terdengar”, “mendengar” dan “candaan”.

Pada hasil visualisasi hanya menampilkan topik-topik yang paling relevan atau memiliki jumlah dokumen terbanyak. Hal ini untuk menyederhanakan interpretasi dan fokus pada topik yang paling dominan.



Gambar 18. Intertopic Distance Map Bertopic

Visualisasi peta jarak antar topik dari korpus teks diatas mengungkapkan beberapa wawasan tentang distribusi dan hubungan antar topik. Dalam peta ini, topik-topik diwakili oleh lingkaran-lingkaran yang berbeda ukuran dan posisinya pada sumbu D1 dan D2. Setiap lingkaran mewakili sebuah topik, di mana ukuran lingkaran menunjukkan jumlah dokumen atau frekuensi kemunculan topik tersebut dalam dataset yang dianalisis. Topik 0 memiliki lingkaran yang paling besar dengan frekuensi kemunculan tertinggi (11511 dokumen), menandakan bahwa topik ini paling sering muncul dalam data yang dianalisis. Jaraknya dengan topik lain menunjukkan bahwa ini adalah topik umum yang mencakup banyak aspek yang terkait dengan tema umum yaitu layanan dalam konteks kuliah, seperti yang terlihat

dari kata kunci “pelayanan_kelas_kuliah_saran”. Jarak antar lingkaran-lingkaran menunjukkan bahwa adanya kemiripan konten atau tema seperti pada topik 4 dan 5 yang berhubungan dengan pemenuhan kebutuhan mahasiswa terhadap fasilitas kampus. Sebaliknya, topik yang berjauhan, seperti topik 4 dan topik 8, menandakan adanya perbedaan signifikan dalam tema yang dibahas. Kedua topik ini berada cukup jauh dari topik lainnya, dengan topik 4 berhubungan dengan fasilitas kampus (“fasilitas_kipas_kampus_kursi”) dan topik 8 terkait motivasi (“kegagalan_menyerah_memperoleh_berhenti”).



BAB V

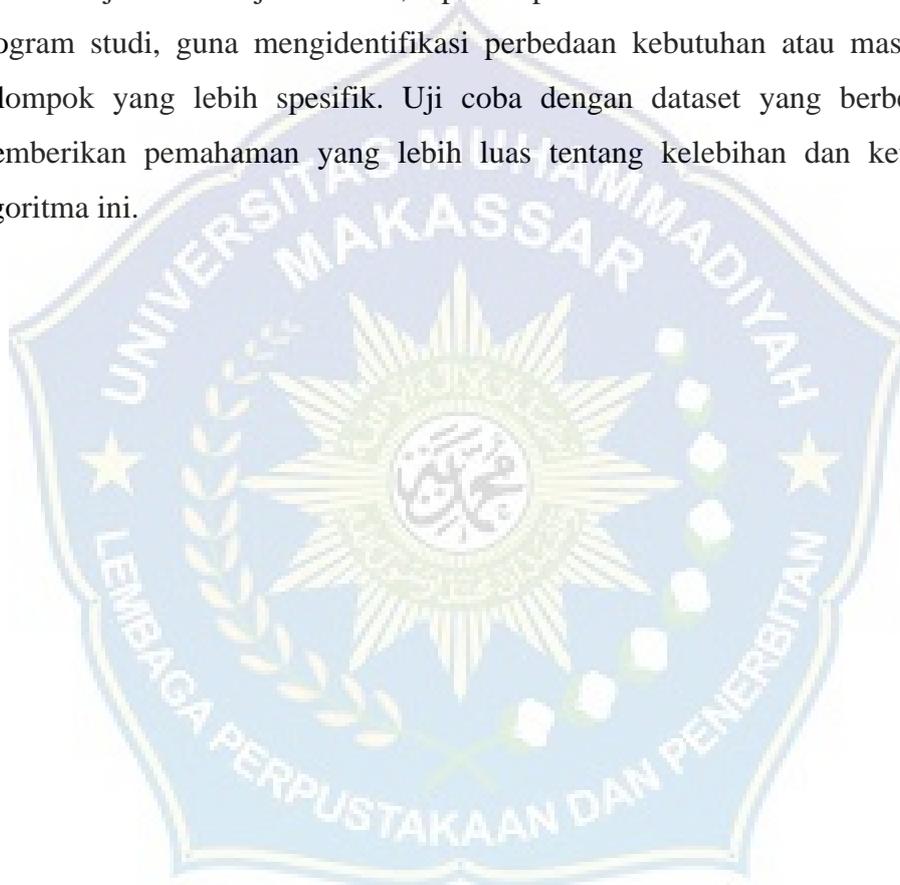
KESIMPULAN DAN SARAN

A. Kesimpulan

1. Penerapan teknik pemodelan topik menggunakan BERTopic untuk menganalisis data saran mahasiswa di UNISMUH berhasil mengidentifikasi dan mengelompokkan berbagai tema utama yang relevan dengan pengalaman mahasiswa. Algoritma BERTopic mengekstrak data saran mahasiswa dan menghasilkan sebanyak 9 topik dari total 26.009 saran. Dari hasil analisa terhadap beberapa topik yang telah diekstrak, dapat disimpulkan bahwasanya banyak mahasiswa yang mengharapkan peningkatan layanan akademik, metode pembelajaran serta sarana dan prasarana yang ada di UNISMUH . Dari segi performa, algoritma BERTopic menghasilkan ketepatan yang memuaskan, dengan *coherence score* sebesar 0.637 dari 9 topik yang diidentifikasi dari 26.009 ulasan. Untuk kecepatan, waktu yang dibutuhkan adalah 13 menit. Waktu tersebut sedikit lebih lama dibandingkan dengan algoritma pemodelan topik lainnya.
2. Hasil visualisasi menunjukkan bahwa topik-topik dalam kumpulan dokumen terdistribusi dalam beberapa kelompok utama dengan frekuensi dan relevansi yang berbeda. Topik yang paling dominan dalam korpus teks adalah topik 0 dengan kata kunci “pelayanan_sarana_mengajar_kelas,” yang menunjukkan fokus utama pada pengajaran dan layanan kelas. Topik-topik penting lainnya meliputi teknik pencatatan dan penggunaan teknologi dalam pembelajaran, manajemen waktu, serta nilai-nilai agama. Isu-isu terkait fasilitas kampus dan sumber daya referensi juga mendapatkan perhatian signifikan. Hubungan antar topik mengindikasikan adanya keterkaitan yang erat antara pengelolaan kelas dan waktu, serta integrasi antara teknologi dan fasilitas pembelajaran.

B. Saran

Berdasarkan kesimpulan dari penelitian yang telah dilakukan, peneliti menyarankan pada penelitian selanjutnya untuk menguji penerapan algoritma BERTopic pada berbagai jenis data teks atau domain lain untuk menilai apakah performa dan efektivitasnya konsisten di luar konteks data saran mahasiswa. Selain itu, disarankan agar algoritma digunakan untuk membagi data topik berdasarkan konteks objek atau subjek tertentu, seperti topik saran berdasarkan fakultas maupun program studi, guna mengidentifikasi perbedaan kebutuhan atau masukan dari kelompok yang lebih spesifik. Uji coba dengan dataset yang berbeda dapat memberikan pemahaman yang lebih luas tentang kelebihan dan keterbatasan algoritma ini.

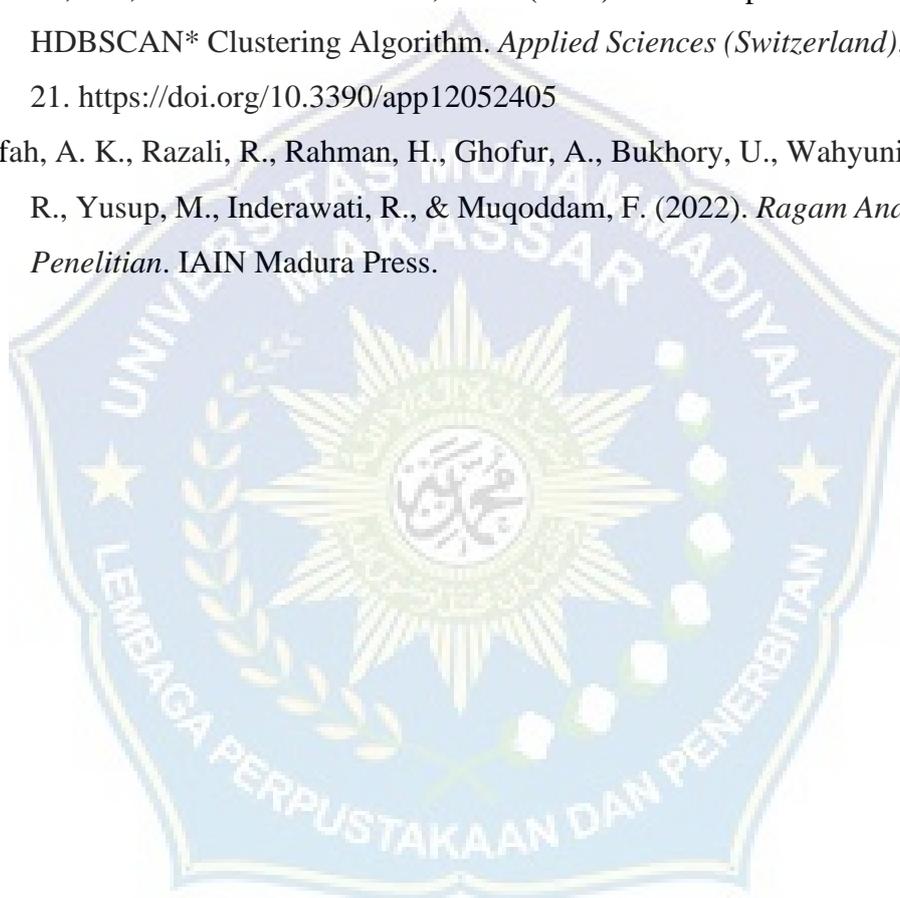


DAFTAR PUSTAKA

- Afidh, R. P. F., & Syahrial. (2023). Pemodelan Topik Menggunakan n-Gram dan Non-negative Matrix Factorization. *Jurnal Informasi Dan Teknologi*, 5(1), 265–275. <https://doi.org/10.60083/jidt.v5i1.385>
- Amazinum. (2022). *What is NLP and how It is Implemented in Our Lives*. <https://amazinum.com/insights/what-is-nlp-and-how-it-is-implemented-in-our-lives/>
- Andronikou, K. (2022). *An in-depth Introduction to Topic Modeling using LDA and BERTopic*. <https://cmotions.nl/en/an-in-depth-introduction-to-topic-modeling-using-lda-and-bertopic/>
- ArRosyid, A. T. S. (2022). Pemodelan Topik Terkait Pemerintah di Indonesia Menggunakan Metode Online Latent Dirichlet Allocation. In *Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta*.
- Atagün, E., Hartoka, B., & Albayrak, A. (2021). Topic Modeling Using LDA and BERT Techniques: Teknofest Example. *Proceedings - 6th International Conference on Computer Science and Engineering, UBMK 2021, September 2021*, 660–664. <https://doi.org/10.1109/UBMK52708.2021.9558988>
- Bhangale, A. (2023). *Introduction to Topic Modelling with LDA, NMF, Top2Vec and BERTopic*. Blend360. <https://medium.com/blend360/introduction-to-topic-modelling-with-lda-nmf-top2vec-and-bertopic-ffc3624d44e4>
- Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S., & Trapnell, C. (2020). Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nature Communications*, 11(1), 1–7. <https://doi.org/10.1038/s41467-020-15351-4>
- Herwinskyah. (2023). Pemodelan Topik Dalam Al-Qur ' an Menggunakan Library. *Simetris*, 14(2), 319–327.
- Iskandar Joko, Prasetya Agung, Kartika Sari Yayak, & Agung Cahyono Taufiq. (2022). Analisis Penerimaan Sistem Informasi Akademik Universitas Bhineka PGRI Menggunakan Integrasi Model TPB dan TAM. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 7(1), 254–263.

- Kapadia, S. (2019). *Evaluate Topic Models: Latent Dirichlet Allocation (LDA)*. Towards Data Science. <https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0>
- Khamidatullailiyah, Y. G. N. (2022). KLASIFIKASI BERITA ONLINE PROGRAM STUDI TEKNIK INFORMATIKA MENGGUNAKAN NEURAL NETWORK. In *Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang*. (Issue 8.5.2017).
- Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, 7(24), 1–16. <https://doi.org/10.4108/eai.13-7-2018.159623>
- Lopez-Martinez, R. E., & Sierra, G. (2020). Research trends in the international literature on natural language processing, 2000-2019 - A bibliometric study. *Journal of Scientometric Research*, 9(3), 310–318. <https://doi.org/10.5530/JSCIRES.9.3.38>
- Martias, L. D. (2021). Statistika Deskriptif Sebagai Kumpulan Informasi. *Fihris: Jurnal Ilmu Perpustakaan Dan Informasi*, 16(1), 40. <https://doi.org/10.14421/fhrs.2021.161.40-59>
- Matira, Y., & Setiawan, I. (2023). Pemodelan Topik pada Judul Berita Online Detikcom Menggunakan Latent Dirichlet Allocation. *Estimasi: Journal of Statistics and Its Application*, 4(1), 2721–379. <https://doi.org/10.20956/ejsa.vi.24843>
- Medvecki, D., Bašaragin, B., Ljajić, A., & Milošević, N. (2024). *Multilingual Transformer and BERTopic for Short Text Topic Modeling: The Case of Serbian*. February, 161–173. https://doi.org/10.1007/978-3-031-50755-7_16
- Mustafac. (2021). *Bert For Topic Modeling (Bert vs LDA)*. Analytics Vidhya. <https://medium.com/analytics-vidhya/bert-for-topic-modeling-bert-vs-lda-8076e72c602b>
- Priyatna, G. G. (2022). Pemodelan Topik Terkait Ulasan Video Game dengan Battle Royale menggunakan Metode BERTopic dengan GUIDED TOPIC MODELLING. In *Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta* (Issue 8.5.2017).

- Reza, L., Sunardi, S., & Herman, H. (2021). Penilaian Sistem Informasi Akademik Dengan Metode Technology Acceptance Model. *Fountain of Informatics Journal*, 7(1), 31. <https://doi.org/10.21111/fij.v7i1.6393>
- Ridhwanulah, D., & Fudholi, D. H. (2022). Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode Latent Dirichlet Allocation. *Jurnal Ilmiah SINUS*, 20(1), 11. <https://doi.org/10.30646/sinus.v20i1.589>
- Stewart, G., & Al-Khassaweneh, M. (2022). An Implementation of the HDBSCAN* Clustering Algorithm. *Applied Sciences (Switzerland)*, 12(5), 1–21. <https://doi.org/10.3390/app12052405>
- Ulfah, A. K., Razali, R., Rahman, H., Ghofur, A., Bukhory, U., Wahyuningrum, S. R., Yusup, M., Inderawati, R., & Muqoddam, F. (2022). *Ragam Analisis Data Penelitian*. IAIN Madura Press.



LAMPIRAN

Lampiran 1 Hasil Plagiasi



MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH MAKASSAR
UPT PERPUSTAKAAN DAN PENERBITAN

Alamat kantor: Jl. Sultan Alauddin NO.259 Makassar 90221 Tlp.(0411) 866972,881593, Fax.(0411) 865588

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

SURAT KETERANGAN BEBAS PLAGIAT

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

Nama : Lis Indriani
Nim : 105841108020
Program Studi : Teknik Informatika

Dengan nilai:

No	Bab	Nilai	Ambang Batas
1	Bab 1	10 %	10 %
2	Bab 2	18 %	25 %
3	Bab 3	10 %	10 %
4	Bab 4	5 %	10 %
5	Bab 5	5 %	5 %

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

Demikian surat keterangan ini diberikan kepada yang bersangkutan untuk dipergunakan
seperlunya.

Makassar, 19 Agustus 2024

Mengetahui,

Kepala UPT- Perpustakaan dan Penerbitan,



Jl. Sultan Alauddin no 259 makassar 90222
Telepon (0411)866972,881 593, fax (0411)865 588
Website: www.library.unismuh.ac.id
E-mail : perpustakaan@unismuh.ac.id

Lis Indriani - 105841108020

BAB I

by Tahap Tutup

Submission date: 19-Aug-2024 03:57PM (UTC+0700)
Submission ID: 2434333131
File name: Bab_1_-_2024-08-19T165117.403.docx (33.91K)
Word count: 757
Character count: 5236

Lis Indriani - 105841108020 BAB I

ORIGINALITY REPORT

10%	10%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	core.ac.uk Internet Source		2%
2	repository.upi.edu Internet Source		2%
3	repository.uinsu.ac.id Internet Source		2%
4	id.123dok.com Internet Source		2%
5	repository.uinbanten.ac.id Internet Source		2%

Exclude quotes On Exclude matches < 2%
Exclude bibliography On

Lis Indriani - 105841108020

BAB II

by Tahap Tutup

Submission date: 19-Aug-2024 03:58PM (UTC+0700)

Submission ID: 2434333382

File name: Bab_2_87.docx (668.3K)

Word count: 2145

Character count: 14227

Lis Indriani - 105841108020 BAB II

ORIGINALITY REPORT

18%

SIMILARITY INDEX

12%

INTERNET SOURCES

3%

PUBLICATIONS

10%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sriwijaya University Student Paper	4%
2	jidt.org Internet Source	3%
3	journal.unhas.ac.id Internet Source	2%
4	alumni.uinjkt.ac.id Internet Source	1%
5	Submitted to Universitas Negeri Semarang - iTh Student Paper	1%
6	www.307bwassoc.org Internet Source	1%
7	Submitted to Universitas Islam Indonesia Student Paper	1%
8	dspace.uii.ac.id Internet Source	1%
9	Submitted to Universitas Muhammadiyah Sidoarjo	1%



Lis Indriani - 105841108020

BAB III

by Tahap Tutup

Submission date: 19-Aug-2024 03:58PM (UTC+0700)

Submission ID: 2434333537

File name: Bab_3_-_2024-08-19T165804.532.docx (109.68K)

Word count: 890

Character count: 6004

Lis Indriani - 105841108020 BAB III

ORIGINALITY REPORT

10%

SIMILARITY INDEX

5%

INTERNET SOURCES

5%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1	digilib.unila.ac.id Internet Source		2%
2	repository.usd.ac.id Internet Source		2%
3	Fatmawati Fatmawati, Gusnia Rindiani, Natasha Oktaviyani, Syahrial Syahrial, Silvina Noviyanti. "Meningkatkan Kreativitas Belajar Siswa Menggunakan Model Pembelajaran Project Based Learning pada Mata Pelajaran IPA di Kelas V SD Negeri 34/I Teratai", AS-SABIQUN, 2022 Publication		1%
4	Submitted to Politeknik Statistika STIS Student Paper		1%
5	etheses.iainponorogo.ac.id Internet Source		1%
6	Yeni Kustiyahningsih, Yohan Permana. "Penggunaan Latent Dirichlet Allocation (LDA) dan Support-Vector Machine (SVM) Untuk		1%

Lis Indriani - 105841108020

BAB IV

by Tahap Tutup

Submission date: 19-Aug-2024 03:59PM (UTC+0700)

Submission ID: 2434333725

File name: Bab_4_82.docx (1.26M)

Word count: 2996

Character count: 21746

Indriani - 105841108020 BAB IV

ORIGINALITY REPORT

5%

SIMILARITY INDEX

4%

INTERNET SOURCES

1%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1	ph01.tci-thaijo.org Internet Source		1%
2	iseratho.github.io Internet Source		1%
3	media.neliti.com Internet Source		1%
4	Submitted to New Mexico State University Main Campus Student Paper		1%
5	Reza Zulfiqri, Betha Nurina Sari, Tesa Nur Padilah. "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MEDIA SOSIAL INSTAGRAM PADA SITUS GOOGLE PLAY STORE MENGGUNAKAN NAÏVE BAYES CLASSIFIER", Jurnal Informatika dan Teknik Elektro Terapan, 2024 Publication		<1%
6	Afifah Nurul Izzati. "ANALISIS SENTIMEN HASIL PUTUSAN MK TERKAIT SENGKETA		<1%

Lis Indriani - 105841108020

BAB V

by Tahap Tutup

Submission date: 19-Aug-2024 03:59PM (UTC+0700)

Submission ID: 2434333897

File name: Bab_5_77.docx (28.82K)

Word count: 350

Character count: 2333

Indriani - 105841108020 BAB V

ORIGINALITY REPORT

5% SIMILARITY INDEX	5% INTERNET SOURCES	0% PUBLICATIONS	0% STUDENT PAPERS
-------------------------------	-------------------------------	---------------------------	-----------------------------

PRIMARY SOURCES

1	repository.trisakti.ac.id Internet Source		3%
2	www.scribd.com Internet Source		2%

Exclude quotes Off Exclude matches Off
Exclude bibliography Off



Lampiran 2. Dataset

a. 50 data pertama

id	kategori	saran	tgl_insert
2	MHS	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	2022-02-15 07:01:34
3	MHS	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.	2022-02-15 07:06:38
5	MHS	Memperbaiki yang baik menjadi lebih baik lagi.	2022-02-15 07:08:13
8	MHS	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	2022-02-15 07:13:31
9	MHS	Perlunya efesiensi waktu yang baik dalam pembelajaran agar matakuliah yabg lain tidak terganggu	2022-02-15 07:13:32
10	MHS	Apakah pelayanan akademik sudah memberikan pelayanan yang terbaik kepada mahasiswa ?	2022-02-15 07:19:31
12	MHS	Saran saya agar proses pembelajaran secara online harus juga di selingi dengan pembelajaran	2022-02-15 07:19:44

		secara offline agar para mahasiswa lebih mengerti lagi materi perkuliahan	
14	MHS	Tetap pertahankan pembelajaran yang baik dan tingkat terus-menerus kedepannya	2022-02-15 07:21:03
18	MHS	Sehubungan dengan pembelajaran daring seperti sekarang ini karena pandemi mohon kerjasama dosen agar tidak susah dihubungi.	2022-02-15 07:23:25
19	MHS	Disiplin waktu perkuliahan	2022-02-15 07:23:30
22	MHS	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	2022-02-15 07:24:25
23	MHS	Semoga sarana dan prasarana kampus dapat digunakan secepatnya dan semoga kuliah online tidak lagi dilaksanakan	2022-02-15 07:24:25
29	MHS	Semogah pembelajaranya semakin merdeka dan bisa offlinr trutama pada saat praktikum	2022-02-15 07:28:30
32	MHS	Tidak ada saran	2022-02-15 07:28:43
36	MHS	Penguasaan tehnik dalam mengajar menggunakan media online dan memberikan toleransi terhdapat mahasiswa yang ada kendala ketika proses perkuliahan secara online.	2022-02-15 07:32:12
37	MHS	Saran saya tdk ada	2022-02-15 07:33:49
40	MHS	Semoga tetap memuaskan peningkatan pelayanan akademik	2022-02-15 07:34:03

43	MHS	Lebih ditingkatkan lagi	2022-02-15 07:35:02
49	MHS	Saran saya yaitu : 1) diharapkan agar Dosen mengajar pada waktu yang tepat, jam dan hari yang sesuai dengan jadwal dan tidak mengambil waktu istirahat kami untuk kuliah 2) diharapkan agar dosen bisa menegur dan membimbing kita secara baik-baik apabila kita melakukan kesalahan	2022-02-15 07:39:50
50	MHS	Terima kasih	2022-02-15 07:44:50
51	MHS	Dengan Melihat Keadaan Dalam Pembelajaran Daring (Online), Pembelajaran Online (Daring) Sangat Tidak Baik Bagi Kami Untuk Fokus belajar, Karena Banyak-Nya Kendala2 Yang Kami Alamin Seperti : Jelek Jaringan, Sakit Mata, Dan Pembelian Kuota Tiap Bulan. Saya Berharap Kedepannya Sudah Bisa Melakukan Pembelajaran Tatap Muka.	2022-02-15 07:44:57
53	MHS	semoga dengan adanya peningkatan pelayanan baik itu dalam proses perkuliahan, pembelajaran ataupun dari pelayanan akademik akan menjadi jauh lebih baik untuk kedepannya.	2022-02-15 07:45:51
54	MHS	Peningkatan fasilitas kampus	2022-02-15 07:45:54
56	MHS	-	2022-02-15 07:47:22
57	MHS	Saran saya agar lebih ditingkatkan lagi proses pembelajaran daring yang lebih baik lagi	2022-02-15 07:50:35

59	MHS	Karena proses perkuliahan masih daring, diharapkan agar sering mengingatkan kepada mahasiswa masalah koneksi ketika ada kuis/UTS/UAS agar mahasiswa bisa mempersiapkan kemungkinan buruk yang terjadi terutama masalah koneksi internet.	2022-02-15 07:52:20
60	MHS	Semoga para dosen bisa menjelaskan materi dengan kreatif agar mahasiswa mudah memahami	2022-02-15 07:54:59
62	MHS	Semoga lebih baik kedepannya	2022-02-15 07:59:07
64	MHS	Saran sya semoga bisah offline	2022-02-15 08:03:10
66	MHS	Pendidikan agama Islam harus di tingkatkan	2022-02-15 08:10:05
70	MHS	Saran saya, sebaiknya kampus segera melaksanakan perkuliahan offline agar mahasiswa dapat menikmati fasilitas kampus kembali	2022-02-15 08:18:27
72	MHS	Semoga kedepannya lebih baik lagi	2022-02-15 08:21:39
74	MHS	Agar kedepannya buku referensi di perpustakaan diperbanyak lagi sesuai dengan buku yang dipakai pada tahun ajaran yang sedang digunakan	2022-02-15 08:22:09
76	MHS	Pelajaran tatap muka dilaksanakan	2022-02-15 08:25:10
77	MHS	membuat susana belajar yang menyenangkan	2022-02-15 08:29:05
79	MHS	Saran saya semoga semester depan bisa kuliah offline	2022-02-15 08:32:32

81	MHS	Kepada dosen sekiranya dalam metode pembelajarannya harus di tingkatkan lagi keberagamannya agar mahasiswa nya lebih tertarik apalagi dalam hal kuliah online ini	2022-02-15 08:35:10
83	MHS	Tidak ada	2022-02-15 08:44:00
84	MHS	Semoga layanan dalam proses pembelajaran kedepannya bisa meningkat kualitasnya di kampus unismuh makassar. Sekian terima kasih	2022-02-15 08:45:07
85	MHS	SEMOGA UNIVERSITAS DAPAT TERUS MENINGKATKAN KAPASITAS DOSEN.....	2022-02-15 08:46:38
86	MHS	Jangan mempersulit mahasiswa(i)ketika dalam pelayanan	2022-02-15 08:46:51
89	MHS	Konsisten terhadap jadwal yang telah ditentukan	2022-02-15 08:50:51
90	MHS	Saran saya, Dosen dapat memberikan materi dan bimbingan kepada siswanya dengan menggunakan video, siaran langsung, atau rekaman. Agar memungkinkan kami dapat lebih memahami pelajaran dan bisa mengulangi apa yang diajarkan kepada kami.	2022-02-15 08:52:50
92	MHS	Tetap lebih efektif dalam melakukan pelayanan saat proses pembelajara supaya lebih maju lagi.	2022-02-15 08:54:17
94	MHS	Semoga dosen bisa memberikan kebijakan yang lebih baik kepada mahasiswa dalam menghadapi masalah.	2022-02-15 08:55:56
95	MHS	Menurut saya yang harus dilakukan untuk meningkatkan pelayanan akademik di program studi adalah dengan cara membuat ruang belajar dengan nyaman. Seperti dilengkapi penyejuk	2022-02-15 08:55:56

		udara serta pengharum udara. Agar udaranya menjadi segar.	
97	MHS	Sebaiknya dalam pembelajaran daring dosen dapat mengerti tidak semua mahasiswa dapat mengakses internet yang baik, dan dosen dapat mengerti jika ada kendala dalam keterlambatan mengirim tugas dan hal yang diakibatkan oleh jaringan yang bermasalah.	2022-02-15 08:57:17
98	MHS	Diupayakan dosen masuk di jam mata kuliah dengan tepat waktu dan tidak mengganti jadwal jam kuliah	2022-02-15 08:58:02
99	MHS	Tidak ada	2022-02-15 07:01:34

b. 50 data terakhir

id	kategori	saran	tgl_insert
232380	MHS	Muda dipahami saat proses pembelajaran berlangsung	2023-05-22 19:54:28
232381	MHS	Muda dipahami saat proses pembelajaran berlangsung	2023-05-22 19:58:04
232382	MHS	Muda dipahami saat proses pembelajaran berlangsung	2023-05-22 19:59:08
232383	MHS	Muda dipahami saat proses pembelajaran berlangsung	2023-05-22 20:02:30
232384	MHS	Muda dipahami saat proses pembelajaran berlangsung	2023-05-22 20:03:32
232385	MHS	Muda dipahami saat proses pembelajaran berlangsung	2023-05-23 09:08:12
232386	MHS	Perlu perbaikan sistem agar pengurusan administrasi tidak terlalu memberatkan mahasiswa	2023-05-24 02:20:59
232387	DSN	1. Perlu peningkatan sarana perkuliahan...	2023-05-24 14:26:52
232388	MHS	Sangat baik	2023-05-24 21:10:44
232389	DSN	-	2023-05-25 10:41:07
232390	MHS	-	2023-05-25 11:59:16

232391	DSN	sukses	2023-05-25 12:02:28
232392	DSN	sukses	2023-05-25 21:38:37
232393	DSN	sukses	2023-05-25 21:42:23
232394	DSN	sukses	2023-05-25 21:46:03
232395	DSN	sukses	2023-05-25 21:49:31
232396	DSN	sukses	2023-05-25 22:04:15
232397	DSN	sukses	2023-05-25 22:28:03
232398	DSN	sukses	2023-05-25 22:38:43
232399	DSN	sukses	2023-05-25 22:41:31
232400	DSN	sukses	2023-05-25 22:44:27
232401	DSN	sukses	2023-05-25 23:04:33
232402	DSN	sukses	2023-05-25 23:13:55
232403	DSN	sukses	2023-05-27 21:58:19
232404	MHS	Perlu di pertahankan metode pembelajarannya	2023-05-29 13:46:24
232405	DSN	Fasilitas Perkuliahan di peradakan sebagai wujud kampus berkemajuan	2023-05-30 11:22:04
232406	DSN	1.Perlu dipikirkan ditingkatkan fasilitas Kesejahteraan dan Jaminan hari Tua /Pensiun kepada Dosen dan Karyawan	2023-05-30 14:53:34
232407	MHS	Semoga bisa ditingkatkan	2023-05-31 17:02:50
232408	DSN	Kelas dilengkapi LCD	2023-06-12 00:37:18
232409	MHS	semua cukup memuaskan untuk mata kuliah komputer akuntansi syariah•	2023-06-12 00:41:05
232410	MHS	merasa puas dengan mata kuliah Ushul Fiqih •	2023-06-12 00:47:31
232411	MHS	memuaskan•	2023-06-12 06:56:45
232412	MHS	melakukan perkuliahan secara offline lebih sering	2023-06-12 07:21:47
232413	MHS	lebih sering melakukan perkuliahan secara offline•	2023-06-12 07:22:55

232414	MHS	sangat puas dengan metode pembelajaran ibu rahmiâ~i,•	2023-06-12 07:26:59
232415	MHS	cukup menyenangkan, sedikit lebih enjoy lagi untuk mata kuliah ulumul hadis•	2023-06-12 16:06:13
232416	MHS	sangat sangat puas dengan metode pembelajaran kemuhammadiyahâ•	2023-06-12 16:08:48
232417	MHS	cukup menyenangkan untuk mata kuliah pancasila•	2023-06-15 13:56:46
232418	MHS	-	2023-06-16 06:50:18
232419	DSN	1. Sinergitas dan komunikasi dengan seluruh komponen sangat baik untuk kemajuan kampus 2. Menyegerakan Dosen yang belum S3 untuk studi lanjut 3. Tolilet dan parkirân memperhatikan 4. Jika meungkinkan, sebagian fakultas dipindahkan ke tempat lain (pengusahaan kampus II)	2023-06-19 09:45:49
232420	DSN	Cukup Baik	2023-06-24 11:23:40
232421	DSN	unismuh unggul	2023-07-03 10:09:07
232422	MHS	semoga pelayanan semakin memudahkan mahasiswa	2023-07-03 10:09:58
232423	MHS	semoga kemudahan pelayanan semakin ditingkatkan lagi	2023-07-03 10:11:15
232424	MHS	semoga pelayanan terhadap mahasiswa semakin baik kedepannya	2023-07-03 10:12:32
232425	MHS	terimakasih atas ilmunya yang sangat bermanfaat	2023-07-03 12:05:26
232426	DSN	sudah baik	2023-07-12 15:26:19
232427	MHS	SANGAT BAIK	2023-07-12 15:31:01
232428	MHS	SANGAT BAIK	2023-07-12 15:31:37
232429	MHS	SANGAT BAIK	2023-07-12 15:32:09

Lampiran 3. Source Code

```
!pip install numpy==1.23.5
!pip install bertopic scikit-learn sentence-transformers umap-learn hdbscan
!pip install nltk Sastrawi
# Mount Google Drive
from google.colab import drive
drive.mount('/content/drive')
import pandas as pd
# Read in data
df= pd.read_csv("/content/drive/MyDrive/Bahan Skripsi/survey_saran(2).csv")
#Drop unnecessary column
df= df.drop(columns=['id','kategori','tgl_insert'])
print(df.shape)
df.head(10)
#Drop Duplicate
df=df.drop_duplicates(keep='first')
print( df.shape)
df.head(10)
df = df[df['saran'].apply(lambda x: len(str(x).split()) >= 10)]
print(df.shape)

#Preprocessing Data
import re
import string
import nltk
from nltk.corpus import stopwords
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

nltk.download('stopwords')
nltk.download('punkt')

more_stopword = [
    'yg',
    'assalamualaikum','warahmatullahi','wabarakatuh','bismillah','dok','tabe','semoga','terima','kasih','terimakasih',
    'saran','mohon','hehe','kedepannya','depannya',
    'maaf','ibu','pak','bapakibu','ibubapak',"mam","ibunda","ibudosen","mom",
] #menambahkan stopword
```

```

stop_factory = StopWordRemoverFactory()
stopword_sastrawi = stop_factory.get_stop_words()
stopwords_nltk = set(stopwords.words('indonesian'))

stopword_combined = set(stopword_sastrawi +
list(stopwords_nltk) + more_stopword )

# Case folding and text cleaning
def preprocess_text(text):
    if isinstance(text, str):
        text = text.lower() # case folding
        text = re.sub(r'[\^\w\s]', '', text) # remove symbol
and punctuation
        text = re.sub(r'\d+', '', text) # remove number
        return text
    else:
        return ''

df['cleaned_saran']=df['saran'].apply(preprocess_text)
#Remove stopwords
df['saran_without_stopwords'] = df['cleaned_saran'].apply(
    lambda x: ' '.join([w for w in str(x).split() if w.lower()
not in stopword_combined]) if isinstance(x, str) else ''
)

print(df.shape)

seed_topics = [
    "dosen", "pembelajaran", "kuliah", "perkuliahan",
    "materi", "pengajaran", "sistem", "RPS",
    "prasarana", "fasilitas", "ruang", "kampus",
    "laboratorium", "peralatan", "kelas", "gedung", "parkiran",
    "administrasi", "pelayanan", "pencatatan", "tata usaha",
    "sarana", "Buku", "fasilitas", "meja", "kursi", "papan
tulis", "proyektor", "kipas",
    "internet", "jaringan", "komputer", "kuota",
    "kebersihan", "keamanan", "petugas", "sampah",
    "biaya", "beasiswa", "keuangan", "uang",
]

#Topic Modelling
from sentence_transformers import SentenceTransformer

```

```

import umap
from hdbscan import HDBSCAN
from bertopic import BERTopic
from bertopic.vectorizers import ClassTfidfTransformer
from sklearn.feature_extraction.text import CountVectorizer

sentence_model = SentenceTransformer("paraphrase-multilingual-
MiniLM-L12-v2")
umap_model = umap.UMAP(n_neighbors=15, n_components=5,
min_dist=0.0, metric='cosine')
hdbscan_model = HDBSCAN(min_cluster_size=15,
metric='euclidean', cluster_selection_method='eom',
prediction_data=True)
vectorizer_model = CountVectorizer()
ctfidf_model = ClassTfidfTransformer(bm25_weighting=True,
reduce_frequent_words=True)

model = BERTopic (
    embedding_model=sentence_model,
    umap_model=umap_model,
    hdbscan_model=hdbscan_model,
    vectorizer_model=vectorizer_model,
    ctfidf_model=ctfidf_model,
    seed_topic_list=seed_topics,
    language='multilingual',
    nr_topics= 10,
    min_topic_size=200,
    calculate_probabilities=True,
    verbose=True
)

topics, probabilities =
model.fit_transform(df['saran_without_stopwords'])
df["topics"] = topics

```

Lampiran 4. Topic Info

Topic	Count	Name	Representation	Representative_Docs
0	11511	0_pelayan an_kelas_ kuliah_sar ana	pelayanan, kelas,kuliah, kelas, sarana, mengajar, pembelajaran, metode, proses, dosen, akademik	dosen aktif perkuliahan materi maksimal diharapkan dosen masuk sesuai jadwal perkuliahan, rangka peningkatan pelayanan akademik program studi proses pembelajaran efektif efisien melaksanakan proses pembelajaran luring, rangka peningkatan pelayanan akademik program studi menyediakan sarana pembelajaran memadai mendukung pembelajaran mahasiswa
1	1674	1_mencata t_materi_p ertahankan _harap	mencatat, materi, pertahankan, harap, menyarankan, perlahan, dipertahankan, ilmu, cepat, pelayanan	materi cepat mencatat materi mencatat memahami materi menyarankan materi perlahan, materi cepat mencatat materi mencatat memahami materi

				<p>menyarankan materi perlahan,</p> <p>materi cepat mencatat materi mencatat memahami materi menyarankan materi perlahan</p>
2	1354	2_offline_ zoom_onli ne_jaringa n	<p>offline, zoom, online, jaringan, google, meet, lakukan, wifi, via, whatsapp</p>	<p>pembelajaran akademik berjalan sesuai mesti dilaksanakan kuliah offline minimal lakukan pembelajaran pertemuan online pertemuan offline karna kuliah online mahasiswa bosan acuh pembelajaran ilmu pembejaran luring efektif',</p> <p>proses daring dosen memahami kendala mahasiswa terkait jaringan lbh efektif lakukan google meeting bandingkan via zoom, pertemuan perkuliahan kuliah offline proses pembelajaran mudah mahasiswa mudah menyerap memahami materi pemebalajaran dibandingkan kuliah</p>

				pertemuan online via google meet
3	633	3_islam_indonesia_bahasa_agama	islam, indonesia, bahasa, agama, muhammadiyah, arab, allah, universitas, islami, ustadz	bahasa arab dipelajari lingkun mahad memudahkan memahami agama islam alquran, rangka pelayanan akademik program studi pendidikan bahasa sastra indonesia, program studi pendidikan agama islam dosen pelayanan akademik
4	586	4_fasilitas_kampus_kursi	fasilitas, kampus, kursi, ruangan, angin, smoga, parkir, meja, lift	fasilitas kelas lengkapi kipas angin peradakan jg kipas ruangan lab, fasilitas ruang kelas kipas angin rusak diganti diperbaiki, fasilitas kampus diperhatikan kenyamanan belajar d dlm kelas alat pendingin kipas angin memadai ruangan
5	265	5_perpustakaan_buku_referensi	perpustakaan, buku, akuntansi, referensi, dana, uang, biaya, pembayaran,	menambah bukubuku jurnaljurnal d perpustakaan terkait hadist tematik tarbawi referensi bahan bacaan u

			<p>bukubuku, keuangan</p>	<p>menunjang pembuatan makalah jurnal sesuai materi perkulihan, menambah bukubuku jurnaljurnal d perpustakaan referensi bahan bacaan u</p> <p>menunjang pembuatan makalah jurnal sesuai materi perkulihan, menambah bukubuku jurnaljurnal d perpustakaan sesuai materi perkulihan referensi bahan bacaan u</p> <p>menunjang pembuatan makalah jurnal</p>
6	152	6_pengumpulan_batas_mempertinggi_maksimal	<p>perpustakaan,buku,akuntansi,referensi,dana,uang,biaya,pembayaran,bukubuku,jurnal</p>	<p>memperpanjang batas pengumpulan tugas terbebani mengerjakannya maksimal,</p> <p>memperpanjang batas pengumpulan tugas terbebani mengerjakannya maksimal,</p> <p>batas pengumpulan tugas tentukan cepat memperpanjang batas pengumpulan tugas</p>

				terbebani mengerjakannya maksimal
7	130	7_suara_suaranya_mendengar_terdengar	suara,suaranya, mendengar, terdengar, candaan, intens, mic, kali, kencang, kelas	suara mengajar terdengar kelas mic disediakan kelas materi, suara suara terdengar bangku menambah volume suara mendengarnya, suara pelan mengajar mendengar pengeras suara suaranya terdengar penjuru kelas membiarkan ruangan tertutup suara kedengaran mendengarkan penjelasan
8	40	8_kegagalan_menyerah_memperoleh_berhenti	kegagalan, menyerah, memperoleh, berhenti, kesuksesan, kegalantetap, mengasah, pendongkrak, terjawab, kegagalanmu	mahasiswi universitas muhammadiyah makassar berhenti belajar mengasah kemampuan menyerah memperoleh kegalantetap kegagalan kesuksesan dengarkan nasihat orang menjatuhkan impian citacitamu optimisdan jadikan kegagalanmu

				<p>pendongkrak untukmu maju menambah fasilitas kampus bangku mahasiswa i ganti sep, berhenti belajar mengasah kemampuan menyerah memperoleh kegagalan tetap kegagalan kesuksesan dengarkan nasihat orang menjatuhkan impian citacitamu optimis dan jadikan kegagalanmu pendongkrak untukmu maju, berhenti belajar mengasah kemampuan menyerah memperoleh kegagalan tetap kegagalan kesuksesan dengarkan nasihat orang menjatuhkan impian citacitamu optimis dan jadikan kegagalanmu pendongkrak untukmu maju</p>
--	--	--	--	---