

**KOMPARASI TEKS VISI DAN MISI PERGURUAN TINGGI  
DI INDONESIA MENGGUNAKAN METODE *COSINE SIMILARITY***

**SKRIPSI**

**Oleh:  
SONIA DESY ASYAROFI  
NIM. 16650028**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**KOMPARASI TEKS VISI DAN MISI PERGURUAN TINGGI  
DI INDONESIA MENGGUNAKAN METODE *COSINE SIMILARITY***

**SKRIPSI**

Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

**Oleh:**  
**SONIA DESY ASYAROFI**  
**NIM. 16650028**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**HALAMAN PERSETUJUAN**

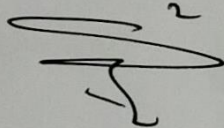
**KOMPARASI TEKS VISI DAN MISI PERGURUAN TINGGI  
DI INDONESIA MENGGUNAKAN METODE *COSINE SIMILARITY***

**SKRIPSI**

**Oleh :  
SONIA DESY ASYAROFI  
NIM. 16650028**

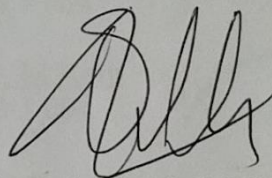
Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 14 Juni 2023

Pembimbing I,



Syahiduz Zaman, M.Kom  
NIP. 19700502 200501 1 005

Pembimbing II,



Fajar Rohman Hariri, M.Kom  
NIP. 19890515 201801 1 001

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachrul Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

## HALAMAN PENGESAHAN

### KOMPARASI TEKS VISI DAN MISI PERGURUAN TINGGI DI INDONESIA MENGGUNAKAN METODE *COSINE SIMILARITY*

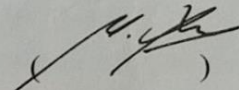
#### SKRIPSI

Oleh :  
**SONIA DESY ASYAROFI**  
NIM. 16650028

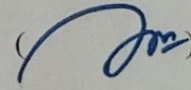
Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )  
Tanggal: 20 Juni 2023

#### Susunan Dewan Penguji

Ketua Penguji : Dr. M. Ainul Yaqin, M.Kom  
NIP. 19761013 200604 1 004



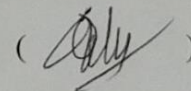
Anggota Penguji I : Johan Ericka Wahyu P, M.Kom  
NIP. 19831213 201903 1 004



Anggota Penguji II : Syahiduz Zaman, M.Kom  
NIP. 19700502 200501 1 005

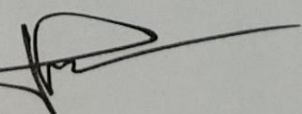


Anggota Penguji III : Fajar Rohman Hariri, M.Kom  
NIP. 19890515 201801 1 001



Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Fachrul Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Sonia Desy Asyarofi

NIM : 16650028

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Judul Skripsi : Komparasi Teks Visi dan Misi Perguruan Tinggi di Indonesia  
Menggunakan Metode *Cosine Similarity*.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 26 Juni 2023

Yang membuat pernyataan,



Sonia Desy Asyarofi

NIM. 16650028

## **MOTTO**

*... Bangkitlah dengan semangat yang menggebu, hadapi setiap rintangan dengan keberanian yang tak tergoyahkan! ...*

## **HALAMAN PERSEMBAHAN**

*“Karya ini kupersembahkan untuk diriku sendiri, ibuk Satukah, kedua orangtuaku, kakakku Fariza Claudia Anggraini terima kasih sudah banyak berkorban untukku dan semoga skripsi ini dapat sedikit mengobati rasa kecewa yang saya berikan. Untuk kakak ipar dan adikku juga, teman-teman serta saudara-saudaraku yang selalu mendukungku. Semua orang baik yang dengan sabar telah saya repotkan”*

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

Syukur alhamdulillah penulis ucapkan kehadiran *Allah* SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Skripsi ini dengan baik.

Selanjutnya penulis mengucapkan terima kasih kepada semua pihak yang telah membantu terselesaikannya Skripsi ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Prof. Dr. H. M. Zainuddin MA, selaku rektor UIN Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT., selaku Ketua Prodi Teknik Informatika yang senantiasa mendorong dan memberikan solusi di setiap permasalahan mahasiswanya.
4. Syahiduz Zaman, M.Kom, selaku dosen wali sekaligus dosen pembimbing 1, yang selalu memberikan ilmu pengetahuan yang tidak ada habis-habisnya, pengarahan, dukungan dan motivasi dalam penyelesaian Skripsi ini.
5. Fajar Rohman Hariri, M.Kom., selaku dosen pembimbing 2, yang telah membimbing penyusunan skripsi ini hingga selesai.
6. Dr. M. Ainul Yaqin, M.Kom dan Johan Ericka Wahyu P, M.Kom selaku dosen penguji 1 dan dosen penguji 2 yang telah membimbing dan secara tidak langsung memberi petunjuk kapan dan kemana penulis harus melangkah.
7. Seluruh dosen Teknik Informatika yang telah memberikan ilmu dan pengalaman yang berharga selama masa perkuliahan.
8. Seluruh staf Teknik Informatika yang telah membantu penulis dalam hal administrasi.



9. Ayah, ibu, kakak dan adik penulis yang senantiasa memberikan kasih sayang, doa, dan dukungan berupa moril maupun materiil kepada penulis.
10. Kerabat-kerabat dan teman – teman penulis yang tidak hanya menanyakan kapan lulus, namun juga memberikan doa dan dukungan kepada penulis.
11. Teman-teman Teknik Informatika 2016 (Andromeda) yang senantiasa saling memberi semangat dan berjuang bersama.
12. Semua pihak yang ikut membantu dalam menyelesaikan Skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan Skripsi ini masih terdapat kekurangan dan penulis berharap semoga Skripsi ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. Amin Ya Rabbal Alamin.

*Wassalamu'alaikum Wr. Wb.*

Malang, 26 Juni 2023

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGAJUAN .....	ii
HALAMAN PERSETUJUAN .....	iii
HALAMAN PENGESAHAN .....	iv
PERNYATAAN KEASLIAN TULISAN .....	v
MOTTO .....	vi
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL .....	xiii
ABSTRAK .....	xiv
ABSTRACT .....	xv
المخلص .....	xvi
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Pernyataan Masalah .....	4
1.3 Tujuan Penelitian .....	5
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah .....	6
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>7</b>
2.1 Penelitian Terkait .....	7
2.2 Komparasi Teks .....	10
2.3 Perguruan Tinggi.....	11
2.4 Visi dan Misi .....	12
2.5 <i>Text Mining</i> .....	13
2.6 <i>Text Preprocessing</i> .....	14
2.7 TF-IDF .....	15
2.8 <i>Cosine Similarity</i> .....	16
<b>BAB III DESAN DAN IMPLEMENTASI.....</b>	<b>18</b>
3.1 Desain Penelitian.....	18
3.2 <i>Dataset</i> .....	18
3.3 Perancangan Sistem .....	19
3.4 Proses <i>Preprocessing</i> .....	20
3.5 Pembobotan TF-IDF .....	26
3.6 <i>Cosine Similarity</i> .....	28
3.7 Skenario Eksperimen .....	30
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>32</b>
4.1 Langkah Uji Coba .....	32
4.1.1 <i>Preprocessing</i> .....	32
4.1.2 Pembobotan TF-IDF .....	39
4.1.3 Perbandingan Menggunakan <i>Cosine Similarity</i> .....	39

4.2 Hasil Uji Coba.....	43
4.3 Integrasi Islam.....	47
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>52</b>
5.1 Kesimpulan .....	52
5.2 Saran.....	53
DAFTAR PUSTAKA .....	55
LAMPIRAN-LAMPIRAN.....	57

## DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian.....	18
Gambar 3.2 Perancangan Sistem.....	20
Gambar 3.3 Tahap Proses <i>Preprocessing</i> .....	20
Gambar 3.4 <i>Flowchart Case Folding</i> .....	21
Gambar 3.5 <i>Flowchart Tokenizing</i> .....	23
Gambar 3.6 <i>Flowchart Stopword Removal</i> .....	25
Gambar 3.7 Algoritma TF-IDF .....	26
Gambar 4.1 Hasil <i>Case Folding</i> .....	33
Gambar 4.2 Hasil <i>Tokenizing</i> .....	35
Gambar 4.3 Hasil <i>Stopword Removal</i> .....	37
Gambar 4.4 Hasil <i>Stemming</i> .....	38
Gambar 4.5 Hasil Pembobotan TF-IDF .....	39
Gambar 4.6 Hasil Perbandingan .....	41

## DAFTAR TABEL

Tabel 2.1 Penelitian Terkait .....	9
Tabel 3.1 Contoh <i>Output Case Folding</i> .....	21
Tabel 3.2 Contoh <i>Output Tokenizing</i> .....	23
Tabel 3.3 Contoh <i>Output Stopword Removal</i> .....	25
Tabel 4.1 Pengujian <i>Blackbox Testing</i> .....	41
Tabel 4.2 Kata-kata yang sering muncul berdasarkan analisis TF-IDF.....	43
Tabel 4.3 Pembobotan TD-IDF Universitas pas 5 .....	45
Tabel 4.4 Pembobotan TF-IDF Universitas diatas 5.....	45
Tabel 4.5 Pembobotan TF-IDF Institut pas 5 .....	45
Tabel 4.6 Pembobotan TF-IDF Institut diatas 5 .....	46
Tabel 4.7 Pembobotan TF-IDF Politeknik pas 5 .....	46
Tabel 4.8 Pembobotan TF-IDF Politeknik diatas 5 .....	46

## ABSTRAK

Asyarofi, Sonia Desy. 2023. *Komparasi Teks Visi dan Misi Perguruan Tinggi di Indonesia Menggunakan Metode Cosine Similarity*. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Syahiduz Zaman, M.Kom (II) Fajar Rohman Hariri, M.Kom

---

*Kata Kunci: Cosine Similarity, Perguruan Tinggi, preprocessing, TF-IDF, Visi dan Misi.*

Visi dan misi memegang peran yang penting dalam organisasi, termasuk perguruan tinggi. Visi dan misi merupakan komponen penting dalam menggambarkan identitas, tujuan, dan fokus sebuah perguruan tinggi. Sebagai bagian dari perencanaan strategis, visi dan misi harus dirancang dengan serius karena menggambarkan gambaran masa depan yang diinginkan. Penelitian ini bertujuan untuk membandingkan teks visi dan misi antar perguruan tinggi di Indonesia menggunakan *cosine similarity*. Penelitian ini juga melibatkan perhitungan TF-IDF untuk mendapatkan wawasan tentang fokus dan nilai yang diungkapkan dalam visi dan misi pendidikan tinggi. Metode pengumpulan data yang dilakukan adalah data yang diperoleh dari situs *website* resmi perguruan tinggi masing-masing yang masuk peringkat top 100 Indonesia menurut *webometrics* edisi Januari 2023. Data yang terkumpul kemudian dilakukan *preprocessing* teks dan mengubahnya menjadi vektor menggunakan metode TF-IDF. Selanjutnya, *cosine similarity* diterapkan untuk mengukur tingkat kemiripan teks visi dan misi antar perguruan tinggi. Hasil penelitian menunjukkan bahwa metode *cosine similarity* dapat digunakan untuk membandingkan teks visi dan misi perguruan tinggi dengan menghasilkan skor kesamaan kosinus. Perguruan tinggi nomor 75 (Universitas Esa Unggul) memiliki tingkat kemiripan tertinggi dengan universitas nomor 1 (Universitas Indonesia) sebesar 31.04%, sedangkan nomor 99 (IAIN Kendari) memiliki tingkat kemiripan terendah sebesar 2.06%. Selain itu, melalui perhitungan TF-IDF, ditemukan kata-kata yang sering muncul dalam visi dan misi perguruan tinggi, yang menggambarkan fokus dan tujuan yang ingin dicapai. Hasil perhitungan TF-IDF juga memungkinkan pengelompokan perguruan tinggi berdasarkan karakteristik unik dalam visi dan misi mereka. Universitas menyoroti kata-kata seperti "bangsa", "internasional", dan "manfaat". Institut menekankan kata "hasil", sementara politeknik menggunakan kata-kata seperti "bangun", "global", "laksana", dan "politeknik". Penelitian ini memberikan kontribusi dalam memahami perbedaan dan kesamaan dalam visi dan misi perguruan tinggi di Indonesia. Informasi ini dapat digunakan untuk pengambilan keputusan strategis dalam pengembangan visi dan misi perguruan tinggi. Selain itu, hasil penelitian ini juga memberikan wawasan tentang karakteristik unik dari universitas, institut, dan politeknik.

## ABSTRACT

Asyarofi, Sonia Desy. 2023. **Comparison of Vision and Mission Texts of Higher Education Institutions in Indonesia Using Cosine Similarity Method.** Theses. Informatics Engineering Study Program, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang. Supervisor: (I) Syahiduz Zaman, M.Kom (II) Fajar Rohman Hariri, M.Kom.

---

Vision and mission statements play a crucial role in organizations, including universities, as they define the identity, goals, and focus of the institution. When developing a strategic plan, careful consideration must be given to crafting the vision and mission statements, as they depict the desired future state. This study aims to compare the vision and mission statements of universities in Indonesia using the cosine similarity method. Additionally, TF-IDF calculations are employed to gain a deeper understanding of the focus and values expressed in these statements. The data for this research was collected from the official websites of the top 100 ranked tertiary institutions in Indonesia, according to the January 2023 edition of Webometrics. The collected data was then pre-processed and transformed into vectors using the TF-IDF method. Subsequently, cosine similarity was utilized to measure the degree of similarity between the vision and mission statements of different universities. The findings indicate that the cosine similarity method effectively compares these statements, yielding a similarity score. Notably, Esa Unggul University (ranked 75) exhibits the highest similarity rate with Universitas Indonesia (ranked 1), at 31.04%, while IAIN Kendari (ranked 99) demonstrates the lowest similarity rate, at 2.06%. Furthermore, the TF-IDF calculations reveal frequently occurring words in the vision and mission statements of higher education institutions, which shed light on their respective focuses and goals. These calculations also facilitate the categorization of higher education institutions based on their unique characteristics as expressed in their vision and mission statements. For instance, universities emphasize words such as "nation," "international," and "benefit," while institutes prioritize the word "outcome," and polytechnics employ terms like "build," "global," "implementation," and "polytechnic." This research contributes to a better understanding of the differences and similarities in the vision and mission statements of higher education institutions in Indonesia. The findings can inform strategic decision-making processes in the development of these statements. Additionally, the study provides valuable insights into the distinctive characteristics of universities, institutes, and polytechnics.

*Keyword:* Cosine Similarity, Higher Education, Preprocessing, TF-IDF, Vision and Mission..

## الملخص

الستري، صانيديسي. ٢٠٢٣ مقارنة نصوص رؤية ومهمة المؤسسات التعليمية العالية في إندونيسيا باستخدام طريقة التشابه الجيبي. رسالة البكالوريوس. برنامج الدراسات العليا في تقنية المعلومات. كلية العلوم والتكنولوجيا، جامعة الإسلامية النيجيرية مولانا مالك إبراهيم مالانج. المشرفين: (١) شاهيد الزمان ، م.كوم (٢) فاجار رحمان هاريري ، م.كوم

كلمات المفتاح: التشابه الجيبي، المؤسسات التعليمية العالية، المعالجة المسبقة، تردد المعكوس للترابط (TF-IDF)، الرؤية والمهمة.

الرؤية والرسالة تلعبان دورًا حاسمًا في المؤسسات، بما في ذلك المؤسسات التعليمية العليا. الرؤية والرسالة هما عنصران أساسيان في وصف الهوية والأهداف والتركيز للمؤسسة التعليمية العليا. كجزء من التخطيط الاستراتيجي، يجب تصميم الرؤية والرسالة بجديّة حيث تصور الرؤية المستقبل المرغوب فيه. تهدف هذه الدراسة إلى مقارنة نصوص الرؤية والرسالة بين المؤسسات التعليمية العليا في إندونيسيا باستخدام التشابه التجاوري للقوسين. تتضمن الدراسة أيضًا حساب TF-IDF لاكتساب رؤية حول التركيز والقيم المعبر عنها في رؤية ورسالة المؤسسات التعليمية العليا. تعتبر طريقة جمع البيانات المستخدمة هي البيانات المحصلة من المواقع الإلكترونية الرسمية لكل مؤسسة تعليمية عالية تدخل في قائمة أفضل 100 مؤسسة في إندونيسيا وفقًا لإصدار يناير 2023 من ميزة *Webometrics*. يتم بعد ذلك معالجة البيانات المجمعة وتحويلها إلى متجهات باستخدام طريقة TF-IDF. ثم يُطبق تشابه التجاوري لقياس مستوى التشابه بين نصوص الرؤية والرسالة بين المؤسسات التعليمية العليا. تشير نتائج الدراسة إلى أن طريقة التشابه التجاوري يمكن استخدامها لمقارنة نصوص الرؤية والرسالة للمؤسسات التعليمية العليا من خلال إنتاج درجات تشابه التجاور القياسي. تُظهر المؤسسة التعليمية رقم 75 (جامعة إسا أونجول) أعلى مستوى للتشابه مع الجامعة رقم 1 (جامعة إندونيسيا) بنسبة 31.04٪، في حين أن المؤسسة رقم 99 (جامعة إين كينداري) تحتوي على أدنى مستوى تشابه بنسبة 2.06٪. وعلاوة على ذلك، من خلال حسابات TF-IDF، يتم تحديد الكلمات الشائعة في رؤية ورسالة المؤسسات التعليمية العالية، والتي تصور التركيز والأهداف المطلوبة. تمكن نتائج حسابات TF-IDF أيضًا تصنيف المؤسسات التعليمية العالية استنادًا إلى سماحتها الفريدة في رؤيتها ورسالتها. تسلط الجامعات الضوء على كلمات مثل "الأمة" و "الدولية" و "



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perguruan tinggi merupakan salah satu institusi pendidikan tinggi yang memiliki peran penting dalam mempersiapkan lulusan yang memiliki kualifikasi yang kompeten dan siap untuk bersaing di pasar kerja. Perguruan tinggi dapat memiliki fokus pada penelitian, pengembangan teknologi, dan inovasi dalam bidang tertentu. Perguruan tinggi juga merupakan lembaga pendidikan yang memberikan pendidikan tinggi untuk mencapai tujuan individu dan masyarakat secara umum, termasuk pengembangan pengetahuan, keterampilan, dan kemampuan kritis (Topcu, 2020). Di perguruan tinggi, mahasiswa belajar melalui kuliah, seminar, diskusi, praktikum, dan proyek. Mahasiswa juga dapat mengikuti kegiatan ekstrakurikuler, seperti organisasi mahasiswa, klub olahraga, dan kegiatan sosial.

Studi pendidikan tinggi memberikan perhatian yang cukup besar untuk memahami perubahan (Brankovic & Cantwell, 2022). Universitas mempunyai fungsi yang penting untuk pengembangan SDM yang berkualitas dan inovatif serta berkontribusi pada kemajuan dan perkembangan masyarakat dan negara. Salah satu sistem pendidikan yang sangat menentukan dalam menghasilkan lulusan yang terampil di bidangnya adalah perguruan tinggi. Perguruan tinggi adalah lembaga pendidikan yang membantu mempersiapkan tenaga kerja yang terampil dan inovatif untuk memenuhi kebutuhan pasar kerja dan mendorong pertumbuhan ekonomi (Lin, 2004).

Visi dan misi mempunyai peran yang sangat bermakna dalam suatu lembaga, termasuk dalam tingkat universitas. Sebagai bagian dari rancangan strategis, visi dan misi harus dirancang secara serius sebab mereka menggambarkan gambaran masa depan yang diinginkan. Visi dan misi yang efektif ialah yang dituangkan dengan tertulis dan diperkenalkan ke semua pihak terkait dalam suatu lembaga, akibatnya kesadaran tentang masa yang akan datang diharapkan tidak cuma sekedar memandu pemimpin-pemimpin lembaga, tetapi juga menjadi panduan bagi semua anggota pada lembaga tersebut (Hamdan, 2001). Pernyataan visi dan misi mencirikan langkah-langkah yang paling signifikan untuk diambil atau diselesaikan oleh pengambil keputusan universitas (Adebanke Olusola et al., 2022).

Pentingnya visi dan misi perguruan tinggi dalam menentukan arah dan tujuan pendidikan yang akan dijalankan. Visi dan misi sangat penting bagi perguruan tinggi karena membantu membentuk identitas dan arah strategis yang jelas. Visi dan misi juga memastikan bahwa setiap program dan kegiatan yang dijalankan sejalan dengan tujuan utama perguruan tinggi (Dooris et al., 2004). Oleh karena itu, visi dan misi perguruan tinggi haruslah disusun secara baik dan tepat supaya bisa memberikan panduan yang jelas untuk mengambil keputusan serta hal-hal yang akan dilakukan di masa mendatang. Visi dan misi mempunyai tingkatan yang penting untuk sebuah lembaga, tidak terkecuali lembaga universitas. Visi dan misi adalah salah satu *part* dari perancangan dalam hal strategi setidaknya terdiri dari visi, misi, prinsip, dan tujuan (Hafizin & Herman, 2022).

Dalam penelitian ini, akan dilakukan komparasi antara visi dan misi dari perguruan-perguruan tinggi yang ada di Indonesia. Pemilihan perguruan tinggi yang akan dibandingkan didasarkan pada faktor peringkat yang akan diambil dari web <https://www.webometrics.info/en>. Untuk membandingkan visi dan misi dari berbagai universitas. Salah satu algoritma yang dapat dipakai ialah metode *Cosine Similarity*. Metode *Cosine Similarity* ialah sebuah algoritma yang digunakan dalam menghitung seberapa besar kemiripan antara 2 teks dengan menggunakan rumus matematika yang memperhitungkan frekuensi kemunculan kata-kata dalam setiap teks (Wahyuni & Prastiyanto, 2017). Dengan menggunakan algoritma ini, kita dapat mengetahui seberapa besar kemiripan visi dan misi antara dua universitas yang dibandingkan.

Penelitian ini akan menggunakan data visi dan misi dari beberapa perguruan tinggi yang sudah diunggah pada *website* resmi di masing-masing perguruan tinggi. Data tersebut kemudian akan diolah dengan memakai metode *cosine similarity* agar dapat mengetahui seberapa besar kemiripan visi dan misi antar universitas. Hasil dari riset yang dilakukan ini harapannya bisa membagikan informasi tentang tingkat kemiripan visi dan misi antar perguruan tinggi di Indonesia yang bisa digunakan untuk bahan pertimbangan kepada pihak yang membutuhkan. Riset kali ini juga diharapkan dapat memberikan kontribusi bagi pengembangan algoritma *cosine similarity*, dengan memberikan informasi mengenai kelebihan dan kekurangan dari algoritma tersebut dalam komparasi teks visi dan misi universitas yang ada di Indonesia.

Dengan demikian, riset kali ini diharapkan mampu bermanfaat untuk universitas, pemerintah, serta warga Indonesia secara keseluruhan. Sebagaimana yang telah tertulis pada Q.S. Al-Baqarah ayat 148, yang berbunyi:

وَلِكُلِّ وُجْهَةٌ هُوَ مُوَلِّئُهَا فَاسْتَغْنُوا أَلْخَيْرَاتِ أَيْنَ مَا تَكُونُوا يَأْتِ بِكُمْ اللَّهُ جَمِيعًا إِنَّ اللَّهَ عَلَىٰ كُلِّ شَيْءٍ قَدِيرٌ

Artinya: *“Dan bagi tiap-tiap umat ada kiblatnya sendiri-sendiri yang ia menghadap kepadanya, maka berlomba-lombalah kamu (dalam berbuat) kebajikan. Di mana saja kamu berada, pasti Allah akan mengumpulkan kamu semuanya. Sesungguhnya Allah Mahakuasa atas segala sesuatu”*. (Q.S. Al-Baqarah ayat 148)

Ayat tersebut memberikan petunjuk tentang pentingnya berlomba-lomba dalam melakukan kebaikan dan bertindak sesuai dengan petunjuk Allah, terlepas dari lokasi geografis atau tempat tinggal seseorang. Meskipun umat Islam memiliki arah kiblat yang berbeda (Mekah) saat melaksanakan salat, setiap orang dianjurkan untuk berusaha dalam kebaikan sesuai dengan keyakinan dan petunjuk yang diberikan oleh agama mereka. Ayat ini juga menunjukkan pentingnya setiap umat memiliki arah tujuan yang jelas dan spesifik, termasuk dalam hal visi dan misi sebuah perguruan tinggi. Dalam konteks penelitian ini, masing-masing universitas mempunyai visi dan misi yang tidak sama, namun tujuannya tetap sama yaitu mencapai kebaikan dan kemajuan bagi masyarakat.

## **1.2 Pernyataan Masalah**

Berlandaskan latar belakang permasalahan yang telah dijabarkan tersebut, pernyataan masalah dalam riset kali ini yaitu:

1. Bagaimana menghitung kemiripan teks visi dan misi antar perguruan tinggi menggunakan *cosine similarity*?
2. Bagaimana cara mengidentifikasi fokus dan tujuan berlandaskan visi dan misi perguruan tinggi yang ada di Indonesia?

### **1.3 Tujuan Penelitian**

Pada penelitian ini memiliki tujuan sebagaimana berikut ini:

1. Untuk mengetahui hasil perhitungan kemiripan teks visi dan misi perguruan tinggi menggunakan *cosine similarity*.
2. Untuk mengidentifikasi fokus dan tujuan berdasarkan visi dan misi universitas di Indonesia.

### **1.4 Manfaat Penelitian**

Pada riset kali ini akan menghasilkan output data perbandingan teks visi dan misi antar universitas dengan menggunakan algoritma *cosine similarity*. Pihak yang bisa menggunakan serta memanfaatkan hasil pada riset kali ini ialah:

1. Para pemangku kepentingan, seperti mahasiswa, dosen, serta masyarakat umum dalam memilih universitas yang cocok dengan kebutuhan serta apa yang mereka harapkan maupun perguruan tinggi itu sendiri.
2. Dapat digunakan untuk referensi bagi para peneliti lain yang ingin melaksanakan riset serupa mengenai perbandingan teks visi dan misi pada institusi atau organisasi lain di Indonesia menggunakan algoritma *cosine similarity*.

### 1.5 Batasan Masalah

Batasan masalah pada riset ini ialah :

1. Pada riset ini memfokuskan perbandingan teks visi dan misi pada peringkat 100 perguruan tinggi Indonesia.
2. Penelitian ini menggunakan metode *cosine similarity* dalam perbandingan teks visi dan misi.
3. Penelitian ini hanya mengambil sampel teks visi dan misi dari *website* resmi perguruan tinggi, sehingga tidak memperhitungkan adanya perbedaan dengan teks yang sebenarnya diterapkan oleh masing-masing universitas.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Penelitian Terkait**

Dumanig & Symaco (2022) dalam penelitian yang berjudul “*Internationalisation of higher education in Malaysia and the Philippines: a comparative analysis of mission and vision statements of selected universities*” penelitian tersebut mengkaji bagaimana internasionalisasi institusi pendidikan tinggi tercermin melalui pernyataan misi dan visi mereka dengan menganalisis secara komparatif pernyataan misi dan visi universitas terpilih di Malaysia dan Filipina. Untuk melaksanakan penelitian ini, dua puluh (20) pernyataan misi dan visi universitas negeri dan swasta di Malaysia dan Filipina dikumpulkan dan dianalisis, digarisbawahi oleh pentingnya agensi tekstual dalam orientasi kebijakan. Kerangka yang digunakan dalam menganalisis pernyataan misi dan visi meliputi analisis praktik tekstual dan praktik wacana.

Analisis wacana tingkat mikro difokuskan pada praktik tekstual, contohnya struktur teks, tata bahasa, serta kosa kata. Selain itu, genre dan gaya teks juga dianalisis. Temuan penelitian mengungkapkan bahwa pernyataan misi dan visi perguruan tinggi di Malaysia dan Filipina sesuai dengan kebijakan pada negara masing-masing, dan mempunyai kesamaan pada konsep internasional pendidikan tinggi. Dapat dibuktikan pula bahwa universitas pada kedua negara tersebut mengutamakan betapa pentingnya menghasilkan lulusan yang berwawasan global serta kompeten, dan mendapatkan pengakuan internasional serta pendidikan *world class*.

Riset yang dilaksanakan oleh Purba et al. (2020) dengan judul “*The optimalization of cosine similarity method in detecting similarity degree of final project by the college students*” bertujuan untuk mengkaji kesamaan Laporan Akhir Mahasiswa Politeknik Unggul LP3M Medan. Algoritma yang dipakai ialah Algoritma *Cosine Similarity*. Metode ini dipilih karena bekerja berdasarkan perhitungan matematis. Cara kerjanya adalah dengan membandingkan tugas akhir yang dikerjakan mahasiswa dengan tugas akhir yang sudah ada sebelumnya. Dengan Metode *Cosine Similitary* akan memperoleh hasil tingkat persentase kemiripan. Apabila tingkat kemiripannya tinggi maka tugas akhir dikatakan *tracing*. Hasil pengujian dengan beberapa dokumen tugas akhir mahasiswa menunjukkan bahwa sistem dapat digunakan dengan baik dengan persentase pengujian sebesar 85,6%.

Thada (2013) dalam penelitian yang memiliki judul “*Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm*” melaksanakan analisis komparatif agar dapat menemukan dokumen terrelevan pada kumpulan kata kunci yang diberikan dengan memakai 3 koefisien kesamaan yaitu koefisien *Jaccard, Dice dan Cosine*. Penelitian ini menggunakan pendekatan algoritma genetika. Koefisien kesamaan untuk sekumpulan dokumen yang diambil untuk kueri tertentu dari *Google* ditemukan kemudian relevansi rata-rata pada hal penilaian kebugaran memakai perhitungan koefisien kemiripan. Hasil dari perhitungan pada penelitian ini yaitu bahwa nilai kebugaran terbaik diperoleh dengan memakai koefisien *Cosine Similarity* diikuti oleh *Dice dan Jaccard*.



Kohila & Arunesh (2016) pada riset dengan judul “*Text Mining: Text Similarity Measure For News Articles Based On String Based Approach*” melakukan penelitian pengukuran Kesamaan Berbasis *String* Algoritma Berbasis *Term Cosine Similarity* dipakai dalam menghitung nilai tingkat kesamaan antara dokumen satu dengan dokumen yang lainnya. Kata benda dalam dokumen diekstraksi dan sinset, kata konteks juga diekstraksi menggunakan *WordNet*. Dataset diagram dibuat berdasarkan kata-kata Konteks. Dalam metode yang diusulkan ini ukuran kemiripan antar dokumen diukur dengan menggunakan algoritma *cosine similarity*. *Preprocessing dataset, context word dataset* dan *bigram dataset* digunakan untuk mengukur kemiripan. Ukuran kumpulan dokumen kata konteks memberikan kesamaan yang lebih baik daripada kumpulan dokumen *bigram* dan *preprocessing*.

Pada penelitian ini, keterbaruan yang dilakukan adalah penelitian menggunakan objek Teks Visi dan Misi Perguruan Tinggi menggunakan *Cosine Similarity* untuk menunjukkan tingkat kemiripan teks visi dan misi.

Tabel 2.1 Penelitian Terkait

Penelitian	Parameter			
	Cosine similarity	TF-IDF	Preprocessing	Perbandingan teks
Dumanig & Symaco, 2020				✓

Purba dkk, 2020	✓			
Thada & Jaglan, 2013	✓			
Kohila & Arunesh, 2016	✓		✓	
Penelitian ini, 2023	✓	✓	✓	✓

## 2.2 Komparasi Teks

Komparasi teks adalah proses membandingkan dan menganalisis dua atau lebih teks untuk menemukan persamaan, perbedaan, dan kesimpulan yang dapat diambil dari perbandingan itu. Tujuan dari komparasi teks ialah supaya mendapatkan pemahaman yang lebih mengenai topik atau subjek yang dibahas dalam teks.

Dalam melakukan komparasi teks, beberapa langkah yang dapat dilakukan antara lain:

- Membaca teks secara keseluruhan untuk memahami topik atau subjek yang dibahas.
- Memahami tujuan penulisan teks dan cara penyajian informasi yang digunakan oleh penulis.
- Mencatat perbedaan dan persamaan antara teks dalam catatan atau tabel.

- Mengambil kesimpulan atau pendapat dari perbandingan tersebut.

Komparasi teks dapat dilakukan pada berbagai jenis teks seperti buku, artikel, laporan, atau surat kabar. Hal ini dapat membantu pembaca untuk memperoleh pemahaman yang lebih baik tentang topik atau subjek yang dibahas dalam teks dan dapat membantu dalam pengambilan keputusan atau analisis.

### **2.3 Perguruan Tinggi**

Menurut UNESCO (2005), perguruan tinggi (PT) ialah institusi yang memberikan pendidikan akademik dan profesional yang tinggi serta riset dan pengabdian masyarakat. PT adalah lembaga pendidikan tinggi yang menyediakan pendidikan dan pengajaran pada jenjang pendidikan setelah menyelesaikan pendidikan menengah. Perguruan tinggi umumnya menyediakan program pendidikan sarjana (S1), program pendidikan pascasarjana (S2), dan program pendidikan doktor (S3) dalam berbagai bidang ilmu.

Perguruan tinggi biasanya memiliki kurikulum yang berbeda-beda tergantung pada program studi atau jurusan yang disediakan. Selain itu, perguruan tinggi juga memiliki fokus pada pengembangan penelitian dan pengembangan ilmu pengetahuan melalui kegiatan riset dan publikasi ilmiah.

Perguruan tinggi dapat memiliki status sebagai universitas atau institut, tergantung pada karakteristik dan fokus kegiatan pendidikan dan penelitiannya. Universitas biasanya memiliki beragam program studi dan jurusan yang lebih luas dan terdapat fokus pada riset dan pengembangan ilmu pengetahuan. Sementara itu, institut biasanya lebih terfokus pada program studi atau jurusan tertentu dan fokus pada pengembangan ilmu pengetahuan dan teknologi dalam bidang tersebut.

Universitas dapat dijalankan oleh pemerintah atau swasta, dan diakui oleh pemerintah melalui proses akreditasi. Lulusan perguruan tinggi diharapkan mempunyai *skill* akademik dan profesional yang baik dan siap memasuki dunia kerja atau meneruskan pendidikan ke jenjang yang lebih tinggi.

#### **2.4 Visi dan Misi**

Menurut (Taiwo et al., 2016) bahwa visi dan misi dianggap dalam bagian penting untuk proses manajemen strategis pada lembaga. Ini berlaku untuk semua jenis organisasi seperti perusahaan publik, swasta, laba atau nirlaba, perusahaan multinasional serta skala kecil dan menengah. Visi dan misi yang dipersiapkan dengan baik akan membedakan satu organisasi dari yang lain dengan menunjukkan karakteristik unik yang membedakannya dari yang lain. Penelitian disini berusaha untuk menganalisis signifikansi pernyataan visi dan misi pada pencapaian tujuan organisasi. Penelitian disini berfokus pada mengeksplorasi peran dan pentingnya pernyataan visi dan misi dalam sebuah lembaga.

Temuan menunjukkan bahwa pernyataan visi dan misi yang dibuat serta diterapkan dengan benar dapat mempengaruhi karyawan organisasi dalam kegiatan sehari-hari mereka dan membantu dalam pencapaian tujuan organisasi dengan misi dan visi sebagai cahaya penuntun. Studi ini merekomendasikan bahwa organisasi harus menyusun pernyataan visi dan misi yang bermakna untuk menempatkan organisasi di jalur yang benar (Hamdan, 2001).

Dalam konteks perguruan tinggi, meneliti visi dan misi juga sangat penting untuk memastikan bahwa institusi tersebut dapat memberikan pendidikan yang berkualitas dan relevan dengan kebutuhan masyarakat serta perkembangan zaman.

Visi dan misi ialah 2 hal yang tidak sama melainkan saling terkait dalam merumuskan tujuan dan arah institusi. Visi adalah gambaran jangka panjang tentang tujuan dan arah institusi yang diinginkan, sedangkan misi adalah pernyataan tentang tujuan dan tugas institusi serta nilai-nilai yang dianut dalam mencapai tujuan tersebut.

Visi perguruan tinggi tidak hanya menjadi suatu konsep yang dipaparkan di atas kertas, tetapi juga harus diimplementasikan secara konsisten dalam setiap kegiatan dan program perguruan tinggi. Visi yang hanya menjadi slogan atau kebanggaan semata tidak akan memberikan dampak positif dalam pencapaian tujuan jangka panjang perguruan tinggi.

## **2.5 Text Mining**

*Text mining* ialah tahap pengestrakan data bermakna serta pengetahuan yang berguna dalam teks yang sangat besar, terstruktur maupun tidak terstruktur, dengan memanfaatkan teknik dan algoritma pengolahan bahasa alami (*natural language processing*), statistik, dan *machine learning*. Tujuan dari *text mining* adalah untuk mengidentifikasi pola, *trend*, hubungan, serta membantu mengambil keputusan berdasarkan data yang dihasilkan (Turban et al., 2011).

Turban, dkk (2011) juga menyebutkan bahwa *text mining* melibatkan empat tahap utama, yaitu pengambilan data, analisis data, interpretasi data, serta *preprocessing* data. Proses pengambilan data meliputi proses pengumpulan data dari sumber data yang tersedia, seperti basis data, dokumen teks, atau sumber data *online*. *Preprocessing* data meliputi proses pembersihan data, normalisasi, tokenisasi, *stemming*, dan penyusunan indeks. Analisis data meliputi proses

pengelompokan (*clustering*), klasifikasi (*classification*), asosiasi (*association*), regresi (*regression*), dan analisis sentimen (*sentiment analysis*). Sedangkan tahap interpretasi data meliputi proses penginterpretasian hasil analisis dan pengambilan keputusan.

*Text mining* adalah teknik pengolahan bahasa alami yang digunakan untuk mengambil informasi penting atau pola dari teks yang besar dan kompleks (Zhang et al., 2015). Teknik ini mencakup penggunaan algoritma dan metode komputasi untuk menganalisis, mengklasifikasikan, dan mengekstrak informasi dari teks yang terstruktur atau tidak terstruktur.

## **2.6 Text Preprocessing**

*Text preprocessing* atau pra-pemrosesan teks adalah tahap awal dari proses *text mining* yang meliputi serangkaian teknik untuk membersihkan, mengubah, dan mempersiapkan data teks sebelum dilakukan analisis atau pemodelan. Tujuannya adalah untuk mempersiapkan data supaya bisa diolah dengan efektif serta efisien, serta meningkatkan kualitas dan akurasi hasil analisis.

Terdapat beberapa teknik *text preprocessing* yang umum digunakan, antara lain (Talib et al., 2016):

1. *Cleaning*: membersihkan data teks dari karakter atau simbol yang tidak penting atau mengganggu, seperti tanda baca, simbol, atau tautan web.
2. *Tokenization*: memisahkan teks menjadi bagian-bagian yang lebih kecil, seperti kata-kata atau frasa, untuk mempermudah analisis.
3. *Stop word removal*: menghapus kata-kata umum atau penghubung yang tidak memiliki makna atau kontribusi signifikan dalam analisis.

4. *Stemming*: mengubah kata-kata menjadi bentuk dasarnya (*stem*) untuk memperoleh informasi yang lebih akurat.

## 2.7 TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) ialah suatu algoritma yang dipakai untuk *text mining* dalam memberikan skor pada kata-kata dalam dokumen berlandaskan frekuensi kata yang sering muncul pada dokumen dan juga dalam seluruh korpus dokumen (Qaiser & Ali, 2018).

TF-IDF menghitung 2 nilai. Nilai tersebut adalah nilai *term frequency* (TF) serta nilai *inverse document frequency* (IDF). Nilai TF memperlihatkan seberapa tingkat frekuensi kata tersebut muncul pada dokumen, sedangkan nilai IDF menunjukkan point penting kata tersebut dalam seluruh dokumen. Dengan mengkombinasikan kedua nilai tersebut, TF-IDF memberikan skor yang tinggi untuk kata-kata yang sering keluar pada dokumen tapi jarang keluar pada seluruh dokumen, sehingga dianggap penting dan relevan untuk dokumen tersebut.

Secara matematis, rumus TF-IDF dapat ditulis sebagai berikut:

$$\text{TF-IDF} = (\text{TF} * \text{IDF})$$

Dimana,

- $\text{TF} = (\text{jumlah kemunculan kata dalam dokumen}) / (\text{total jumlah kata dalam dokumen})$
- $\text{IDF} = \log((\text{total jumlah dokumen dalam korpus}) / (\text{jumlah dokumen yang mengandung kata tersebut}))$
- $\log()$  pada IDF adalah logaritma natural.

Dalam *text mining*, TF-IDF sering digunakan dalam beberapa aplikasi seperti *information retrieval*, *text classification*, *clustering*, dan lain-lain. Dengan menggunakan skor TF-IDF, dokumen-dokumen yang mengandung kata-kata dengan skor tinggi dapat dianggap lebih relevan dan penting dalam suatu analisis teks.

## **2.8 Cosine Similarity**

*Cosine similarity* (CS) adalah sebuah metode yang digunakan dalam *text mining* untuk mengukur kesamaan antara dua dokumen berdasarkan kosinus sudut antara vektor representasi dokumen tersebut. CS adalah salah satu pengukuran klasik yang penggunaannya sering digunakan untuk pencarian informasi dan merupakan pengukuran yang paling sering dilaporkan dari kesamaan vektor (Ye, 2011).

Secara umum, *cosine similarity* membandingkan kedua dokumen dengan melihat seberapa sering kata-kata tertentu muncul dalam kedua dokumen tersebut, dan seberapa penting kata-kata tersebut dalam masing-masing dokumen.

Langkah-langkah dalam menghitung *cosine similarity* antara dua dokumen adalah sebagai berikut (Apriani et al., 2021):

1. Menghitung vektor representasi dokumen: dokumen direpresentasikan sebagai vektor, dimana setiap dimensi vektor mewakili kata-kata yang muncul dalam dokumen tersebut, dan setiap nilai pada dimensi tersebut mewakili frekuensi kemunculan kata tersebut dalam dokumen.
2. Menghitung *cosine similarity*: *cosine similarity* dihitung sebagai kosinus dari sudut antara vektor representasi dokumen tersebut. Semakin kecil



sudut antara vektor, semakin besar nilai *cosine similarity*, dan semakin mirip kedua dokumen tersebut.

Secara matematis, rumus *cosine similarity* dapat ditulis sebagai berikut:

$$CS = (A \cdot B) / (\|A\| \times \|B\|)$$

Dimana,

- A dan B adalah vektor representasi dokumen
- $A \cdot B$  adalah hasil perkalian dot dari vektor A dan B

$\|A\|$  dan  $\|B\|$  adalah nilai panjang dari vektor A dan B.

## BAB III

### DESAN DAN IMPLEMENTASI

#### 3.1 Desain Penelitian

Penelitian ini menggunakan metode kuantitatif dengan teknik analisis data berupa algoritma *Cosine Similarity* dan juga memiliki pendekatan komparatif untuk membandingkan teks visi dan misi dari berbagai perguruan tinggi. Dimana memperoleh dukungan dari beberapa tahapan metode pendukung lainnya, beberapa metode pendukung yaitu *preprocessing* dan TF-IDF. Untuk membandingkan visi dan misi dari berbagai perguruan tinggi penelitian ini menggunakan algoritma *Cosine Similarity*. Dengan menggunakan algoritma ini, output yang dihasilkan berupa perbandingan teks visi dan misi antar perguruan tinggi yang nantinya hasilnya dapat dianalisis. Desain penelitian akan ditunjukkan pada Gambar 3.1.



Gambar 3.1 Desain Penelitian

#### 3.2 Dataset

Sampel penelitian akan diambil secara *purposive sampling*. Pemilihan sampel didasarkan pada kriteria perguruan tinggi yang masuk peringkat 100 dunia dan peringkat 100 Indonesia. Data yang melalui mekanisme penelitian ini adalah data yang didapatkan dari situs web resmi universitas yang tergolong data sekunder yang masing-masing untuk teks visi dan misi, dan untuk daftar peringkat universitas

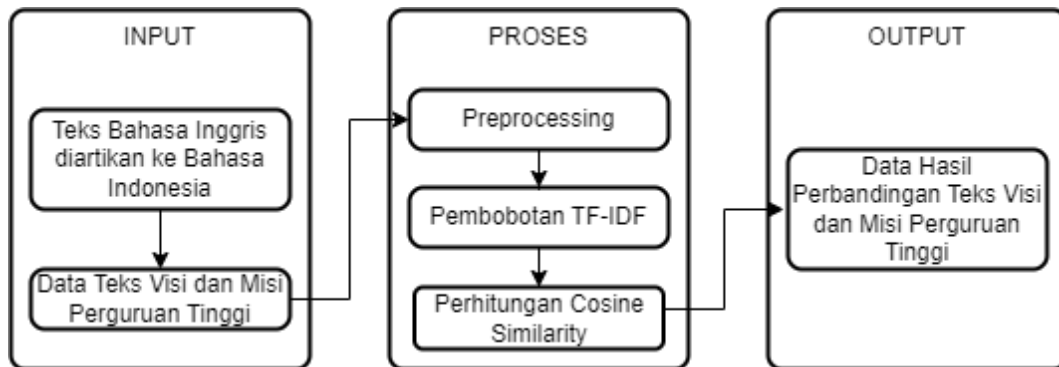
diambil dari web *Webometrics Ranking of World Universities* edisi Januari 2023. *Webometrics Ranking of World Universities* adalah inisiatif dari *Cybermetrics Lab*, sebuah kelompok riset milik *Consejo Superior de Investigaciones Científicas* (CSIC), badan riset publik terbesar di Spanyol.

Proses mekanisme pengumpulan dataset melalui cara sebagai berikut:

1. Membuat *google form*
2. Input data yang dibutuhkan ke *google form*
3. Data yang terkumpul dibuka di *spreadsheet*

### **3.3 Perancangan Sistem**

Perancangan sistem merupakan tahap yang penting dalam suatu penelitian karena tahap ini menentukan bagaimana sistem akan dirancang dan diimplementasikan. Dalam perancangan sistem, diperlukan kejelasan dalam memahami kebutuhan sistem dan memastikan sistem yang dikelola dapat bekerja serta menghasilkan hasil yang valid dan dapat bermakna. Pada perancangan sistem melewati beberapa tahapan mekanisme yaitu input, proses, dan output dari sistem yang akan dibuat sehingga memberikan hasil komparasi atau perbandingan teks visi dan misi antar perguruan tinggi. Perancangan sistem yang akan dibuat dalam penelitian dapat dilihat pada Gambar 3.2.

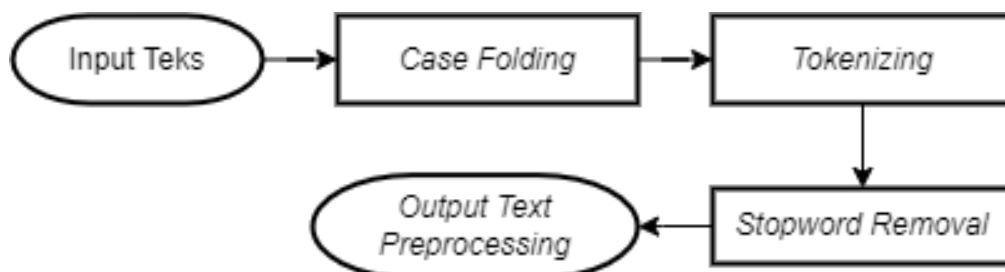


Gambar 3. 2 Perancangan Sistem

### 3.4 Proses *Preprocessing*

Proses ini merupakan proses awal sebelum dianalisa, setiap teks visi dan misi perguruan tinggi akan melalui tahap *preprocessing*. *Preprocessing* tersebut mengacu pada serangkaian teknik pengolahan teks yang digunakan untuk membersihkan dan mempersiapkan data teks sebelum dihitung dengan algoritma *cosine similarity*. Serta untuk menghasilkan data yang lebih terstruktur dan saat proses *test preprocessing* dapat berjalan dengan baik.

Adapun beberapa mekanisme *preprocessing* pada penelitian adalah *case folding*, *tokenizing*, dan *stopword removal*. Tahapan proses *preprocessing* terlihat pada Gambar 3.3.

Gambar 3. 3 Tahap Proses *Preprocessing*

### 1. Case folding

*Case folding* termasuk dalam tahapan *preprocessing* yang bertujuan mengubah semua huruf pada teks menjadi satu kesamaan yaitu huruf kecil. Hal ini dilakukan untuk menghindari duplikasi kata atau istilah yang sama dengan huruf kapital maupun huruf kecil yang berbeda. Berikut adalah *flowchart* untuk tahapan *case folding* dapat dilihat pada Gambar 3.4 beserta contoh hasil *output case folding* pada tabel 3.1.



Gambar 3. 4 *Flowchart Case Folding*

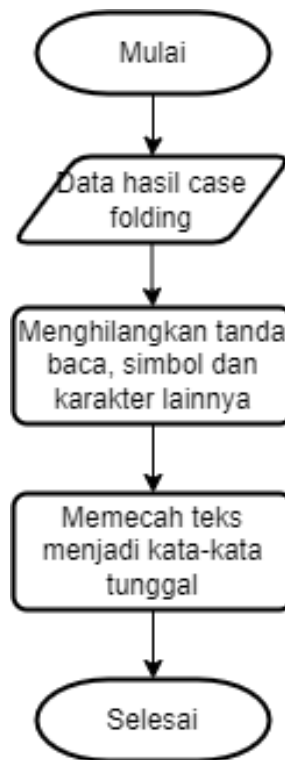
Tabel 3. 1 Contoh *Output Case Folding*

Universitas	Teks Visi Misi	Hasil <i>Case Folding</i>
Harvard	Menciptakan Mahasiswa tak tertandingi yang transformatif secara intelektual, sosial, dan pribadi.	menciptakan mahasiswa tak tertandingi yang transformatif secara intelektual, sosial, dan pribadi.

UI	Menjadi pusat ilmu pengetahuan, teknologi, dan kebudayaan yang unggul dan berdaya saing.	menjadi pusat ilmu pengetahuan, teknologi, dan kebudayaan yang unggul dan berdaya saing.
UGM	Perguruan Tinggi Nasional berkelas dunia yang unggul dan inovatif, mengabdikan kepada kepentingan Bangsa dan kemanusiaan	perguruan tinggi nasional berkelas dunia yang unggul dan inovatif, mengabdikan kepada kepentingan bangsa dan kemanusiaan

## 2. *Tokenizing*

*Tokenizing* adalah mekanisme pemisahan kalimat atau teks menjadi unit-unit terkecil yang disebut dengan token atau kata. Setiap token akan menjadi dasar dalam proses selanjutnya. Tahapan *tokenizing* penting dilakukan karena banyak algoritma pemrosesan teks membutuhkan token sebagai *input*. Dengan memiliki token yang sudah dipisahkan dari teks, maka proses selanjutnya dapat lebih mudah dan efisien. *Tokenizing* dapat menghilangkan semua tanda baca seperti tanda koma, titik, petik, dan lain-lain. Berikut adalah *flowchart* untuk tahapan *tokenizing* dapat dilihat pada Gambar 3.5 beserta contoh hasil *output tokenizing* pada tabel 3.2.

Gambar 3. 5 *Flowchart Tokenizing*Tabel 3. 2 Contoh *Output Tokenizing*

Universitas	Teks Visi Misi	Hasil <i>Tokenizing</i>
Harvard	Menciptakan Mahasiswa secara intelektual, sosial, dan pribadi.	<ul style="list-style-type: none"> <li>• menciptakan</li> <li>• mahasiswa</li> <li>• secara</li> <li>• intelektual</li> <li>• sosial</li> <li>• dan</li> <li>• pribadi</li> </ul>
UI	Menjadi pusat ilmu pengetahuan dan teknologi	<ul style="list-style-type: none"> <li>• menjadi</li> <li>• pusat</li> <li>• ilmu</li> <li>• pengetahuan</li> <li>• dan</li> </ul>

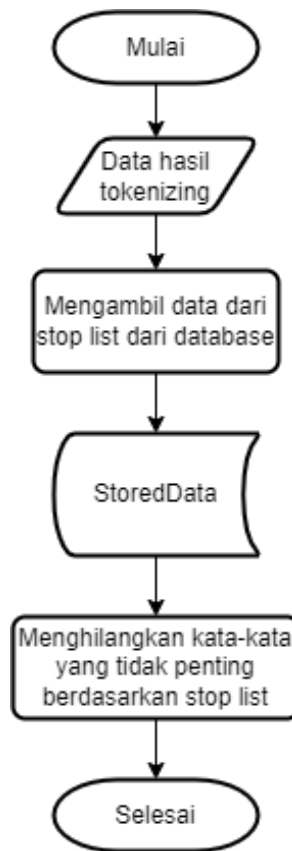
		• teknologi
--	--	-------------

### 3. *Stopword Removal*

*Stopword removal* adalah tahap *preprocessing* memiliki tujuan untuk menghapus kata yang termasuk tidak penting dalam sebuah teks, seperti kata-kata penghubung seperti "dan", "atau", "di", dan kata termasuk dalam kata tidak bernilai. Pada penelitian ini *stopword removal* memiliki peran penting dalam pemrosesan teks visi dan misi perguruan tinggi. Dengan menggunakan *stopword removal*, kata-kata yang termasuk dalam kata tidak penting dapat dihapus sehingga hanya tersisah kata-kata penting. Hal ini dapat mempermudah proses analisis teks visi dan misi perguruan tinggi, karena fokus hanya pada kata-kata yang memberikan kontribusi signifikan terhadap makna dari teks tersebut. Sebagai contoh, kata-kata umum seperti "visi" dan "misi" karena kata-kata tersebut sudah jelas terdapat dalam teks visi dan misi perguruan tinggi, sehingga tidak memberikan kontribusi yang signifikan terhadap analisis.

Dalam penelitian komparasi teks visi dan misi perguruan tinggi menggunakan algoritma *cosine similarity*, penggunaan *stopword removal* dapat meningkatkan akurasi hasil analisis. Dengan menghapus kata-kata yang tidak penting, nilai *cosine similarity* dapat mencerminkan kemiripan antara teks visi dan misi perguruan tinggi secara lebih akurat, karena hanya kata-kata penting yang digunakan dalam perhitungan. Berikut adalah *flowchart* untuk tahapan *stopword removal* dapat dilihat pada Gambar 3.6 beserta contoh hasil *output stopwords removal* pada tabel 3.3.



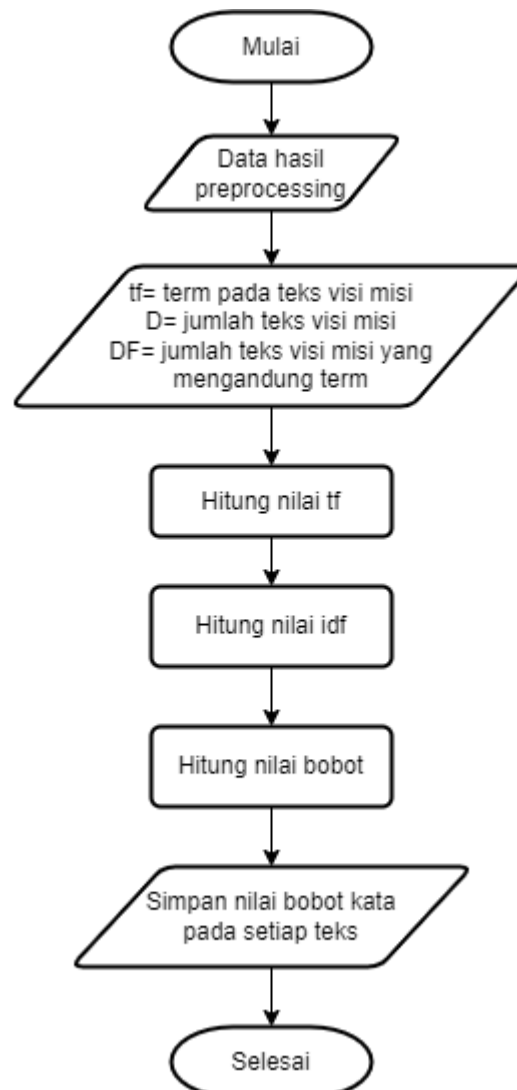


Gambar 3. 6 *Flowchart Stopword Removal*

Tabel 3. 3 Contoh *Output Stopword Removal*

Universitas	Teks Visi Misi	Hasil <i>Stopword Removal</i>
Harvard	Menciptakan Mahasiswa secara intelektual, sosial, dan pribadi.	<ul style="list-style-type: none"> <li>• menciptakan</li> <li>• mahasiswa</li> <li>• intelektual</li> <li>• sosial</li> <li>• pribadi</li> </ul>
UI	Menjadi pusat ilmu pengetahuan dan teknologi	<ul style="list-style-type: none"> <li>• menjadi</li> <li>• pusat</li> <li>• ilmu</li> <li>• pengetahuan</li> <li>• teknologi</li> </ul>

### 3.5 Pembobotan TF-IDF



Gambar 3.7 Algoritma TF-IDF

Setelah *preprocessing* akan melakukan pembobotan dengan metode TF-IDF. Pada metode TF-IDF, setiap kata pada teks memiliki skor TF-IDF yang berbeda. Skor ini didapatkan dengan mengalikan nilai TF dan IDF. TF adalah skor yang menunjukkan kata yang sering suatu muncul pada teks atau kalimat disebuah dokumen. Semakin sering sebuah kata muncul pada sebuah dokumen, semakin besar skor TF-nya. IDF (*Inverse Document Frequency*) adalah skor yang

menunjukkan seberapa umum suatu kata pada seluruh dokumen. Kata-kata yang muncul pada banyak dokumen umumnya memiliki skor IDF yang lebih rendah, karena kata-kata tersebut kurang spesifik dan tidak memberikan informasi yang signifikan untuk mengidentifikasi dokumen.

Pada penelitian ini memperoleh bobot kata dengan menggunakan metode TF-IDF pada suatu teks visi dan misi perguruan tinggi berdasarkan kemunculan kata tersebut dalam seluruh teks. Rumus dasar dari TF-IDF adalah sebagai berikut:

$$\text{TF-IDF/Weight}(W) = \text{tf} * \text{idf} \quad \rightarrow \text{IDF} = \log\left(\frac{D}{DF}\right)$$

Di mana:

D = jumlah total dokumen

DF = banyak dokumen yang mengandung kata(*term*)

W = hasil bobot nilai dokumen

tf = seberapa sering kata muncul dalam dokumen atau total kata dalam dokumen

idf =  $\log$  (total dokumen / dokumen yang termasuk mengandung kata tersebut)

Algoritma untuk menghitung bobot kata dengan metode TF-IDF dapat dilihat pada gambar 3.7 adalah sebagai berikut:

1. Lakukan *preprocessing* pada dokumen, seperti *case folding*, *tokenizing*, dan *stopword removal*.
2. Hitung nilai tf untuk setiap kata pada dokumen dengan rumus  $\text{tf} =$  jumlah kemunculan kata dalam dokumen

3. Hitung nilai idf untuk setiap kata dengan rumus  $idf = \log(\text{total dokumen} / \text{dokumen yang mengandung kata tersebut})$ .
4. Hitung nilai bobot kata dengan rumus  $TF-IDF = tf * idf$ .
5. Simpan nilai bobot kata pada setiap dokumen.
6. Lakukan langkah 2-5 untuk seluruh dokumen dalam korpus.
7. Gunakan nilai bobot kata yang telah dihitung untuk melakukan komparasi teks visi misi antar perguruan tinggi menggunakan algoritma *Cosine Similarity*.

### 3.6 *Cosine Similarity*

Cosine similarity (CS) adalah mekanisme pengukuran kemiripan antara dua buah dokumen teks. Secara sederhana, metode CS akan mengukur sudut *cosine* antara dua buah vektor yang merepresentasikan dua dokumen tersebut. Secara matematis, *cosine similarity* dihitung sebagai berikut:

$$\text{cosine\_similarity}(d1, d2) = \text{dot\_product}(d1, d2) / (\text{norm}(d1) * \text{norm}(d2))$$

dimana  $d1$  dan  $d2$  adalah dua buah vektor yang merepresentasikan dua dokumen, *dot\_product* adalah hasil perkalian dot (*inner product*) antara vektor  $d1$  dan  $d2$ , dan *norm* adalah panjang vektor.

Dalam metode penelitian CS digunakan untuk menghitung kemiripan atau perbandingan antara teks visi dan misi dari setiap perguruan tinggi yang ada dalam *dataset*. Semakin besar nilai *cosine similarity* antara dua dokumen, semakin mirip pula teks tersebut dan apakah juga semakin dekat peringkat perguruan tinggi tersebut?

Pada penelitian ini, *cosine similarity* dihitung setelah proses *preprocessing* dan pembobotan TF-IDF dilakukan pada masing-masing teks. Algoritma *cosine similarity* digunakan pada penelitian ini untuk membandingkan kemiripan antara dua dokumen teks visi dan misi perguruan tinggi. *Cosine similarity* termasuk dalam salah satu pengukuran kemiripan dokumen yang pemakaiannya sering terlihat pada analisis teks.

Algoritma *cosine similarity* bekerja dengan menghitung sudut kosinus antara dua vektor dokumen atau teks visi misi antar perguruan tinggi yang direpresentasikan dalam bentuk vektor kata-kata. Pada penelitian tersebut, setelah dilakukan tahap *preprocessing* dan pembobotan TF-IDF pada dokumen teks visi dan misi perguruan tinggi, maka kedua dokumen direpresentasikan dalam bentuk vektor kata-kata dengan nilai bobot yang terkait dengan frekuensi kata dan nilai IDF. Selanjutnya, dilakukan perhitungan *dot product* dari kedua vektor dan kemudian dibagi dengan hasil perkalian norma kedua vektor tersebut. Hasil akhir dari perhitungan *cosine similarity* adalah nilai skor kemiripan antara dua dokumen, yang memiliki rentang nilai antara 0 (tidak sama sekali mirip) hingga 1 (sama persis).

Berikut ini adalah algoritma *cosine similarity* yang dijalankan pada penelitian ini:

1. melakukan tahap *preprocessing* pada dokumen teks berguna menghilangkan semua karakter yang tidak diperlukan, seperti tanda baca, angka, dan simbol.

2. melakukan pembobotan kata pada setiap dokumen teks menggunakan metode TF-IDF.
3. merepresentasikan setiap dokumen teks dalam bentuk vektor kata-kata, di mana setiap dimensi vektor merepresentasikan satu kata pada dokumen.
4. Perhitungan dot product antara kedua vektor dokumen untuk mendapatkan nilai kemiripan antara kedua dokumen.
5. Normalisasi vektor dokumen dengan cara menghitung norma *Euclidean* dari setiap vektor.
6. Bagi hasil perhitungan *dot product* dengan hasil perkalian norma kedua vektor untuk mendapatkan nilai *cosine similarity* antara kedua dokumen.
7. Ulangi langkah 3-6 untuk setiap pasangan dokumen yang akan dibandingkan.
8. Membandingkan hasil *cosine similarity* antar visi misi setiap perguruan tinggi untuk menentukan tingkat kemiripan antara teks-teks tersebut.

### 3.7 Skenario Eksperimen

Perbandingan teks visi dan misi dengan *Cosine Similarity* yang diimplementasikan dengan menggunakan bahasa pemrograman Python pada *Google Colab*. Implementasi pada penelitian dimulai dari *preprocessing* data sampai penghitungan *Cosine Similarity* dilakukan dengan menggunakan *Google Colab*. Proses implementasi dilakukan dengan memasukkan teks visi dan misi dari setiap perguruan tinggi sebagai input data ke *google form* untuk membuat dataset,

kemudian melakukan *preprocessing* sampai perhitungan *Cosine Similarity* dan akan menghasilkan nilai kemiripan antara teks visi dan misi dari setiap perguruan tinggi.

Pada intinya skenario eksperimen penelitian adalah setelah dataset semua terkumpul mulailah proses yang akan dijalankan. Sebelum ke proses penggunaan metode *Cosine Similarity*, data tersebut diproses terlebih dulu dengan *preprocessing*. Adapun tahapan *preprocessing* pada penelitian ini adalah *case folding*, *tokenizing*, *stopword removal* dan *stemming*.

Skenario pengujian ada 5:

1. Memasukkan kode NLTK untuk proses *preprocessing*
2. Input database berbentuk csv ke dalam sistem
3. Melakukan proses *preprocessing* yaitu *case folding*, *tokenizing*, *stopword removal* dan *stemming*.
4. Melakukan penilaian bobot dengan TF-IDF
5. Memasukkan kode untuk menghasilkan perbandingan

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Langkah Uji Coba

##### 4.1.1 *Preprocessing*

Pada penelitian ini yaitu melakukan *preprocessing* dalam persiapan data yang dilakukan sebelum melakukan analisis atau pemrosesan lebih lanjut. Tujuan dari *preprocessing* adalah untuk membersihkan, mengubah format, dan mengatur data agar sesuai dengan kebutuhan analisis. Dalam penelitian ini langkah-langkah dalam *preprocessing* data meliputi *case folding*, *tokenizing*, *stopword removal* dan *stemming*.

##### 1. *Case Folding*

Pada penelitian ini dalam proses *preprocessing* pertama melakukan *case folding*. Yang tahapan ini mengubah semua huruf dalam teks menjadi huruf kecil.

Berikut *source code case folding*:

```
#Case Folding
df['Visi dan Misi'] = df['Visi dan Misi'].str.lower()

print('Hasil Case Folding : \n')
print(df['Visi dan Misi'].head(20))
print('\n\n\n')
```



```

Hasil Case Folding :

0  harvard college sets the standard for resident...
1  vision: our vision guides stanford's approach ...
2  the mission of mit is to advance knowledge and...
3  born out of a vision in the state constitution...
4  our mission:\nwe inspire people locally, natio...
5  the mission of the university of michigan is t...
6  mission statement\nthe primary mission of the ...
7  university mission:\nlearning. discovery. enga...
8  our mission\ncolumbia university facilities an...
9  mission: to educate its students and cultivate...
10 visi: menjadi pusat ilmu pengetahuan, teknolog...
11 visi:\nuniversitas gajah mada sebagai pelopor...
12 visi:\r\n"ub mempunyai visi menjadi perguruan t...
13 vision :\n"to become a research-based higher e...
14 visi: menjadi universitas yang mandiri, inovat...
15 visi: menjadi pusat pengembangan ilmu pengetah...
16 visi: usk memiliki visi menjadi universitas so...
17 visi: menjadi perguruan tinggi berkelas dunia ...
18 visi: menjadi research and entrepreneurial uni...
19 visi: menjadi perguruan tinggi yang unggul, be...
Name: Visi dan Misi, dtype: object

```

Gambar 4.1 Hasil *Case Folding*

## 2. Tokenizing

Setelah melakukan *case folding* seperti di atas. Selanjutnya melakukan *tokenizing* yaitu proses pemisahan teks menjadi satuan unit yang lebih kecil yang nama lainnya disebut "token". Setiap token mewakili sebuah kata, frasa, atau elemen penting lainnya dalam teks. Berikut *source code tokenizing*:

```

import string

import re

from nltk.tokenize import word_tokenize

from nltk.probability import FreqDist

#Tokenizing

def remove_tweet_special(text):

    text = text.replace('\t', " ").replace('\n', " ")
    text = text.replace('\u', " ").replace('\ ', " ")

    text = text.encode('ascii', 'replace').decode('ascii')

    text = ' '.join(re.sub("([@#] [A-Za-z0-9]+)|(\w+:\/\/\s+)", " ", text).split())

```

```

        return text.replace("http://", " ").replace("https://", " ")
df['Visi      dan      Misi']      =      df['Visi      dan
Misi'].apply(remove_tweet_special)
#removeNumber
def remove_number(text):
    return re.sub(r"\d+", "", text)
df['Visi      dan      Misi']      =      df['Visi      dan
Misi'].apply(remove_number)
#removePunctuation
def remove_punctuation(text):
    return
text.translate(str.maketrans("", "", string.punctuation))
df['Visi      dan      Misi']      =      df['Visi      dan
Misi'].apply(remove_punctuation)
#remove whitespace leading & trailing
def remove_whitespace_LT(text):
    return text.strip()
df['Visi      dan      Misi']      =      df['Visi      dan
Misi'].apply(remove_whitespace_LT)
#remove multiple whitespace into single whitespace
def remove_whitespace_multiple(text):
    return re.sub('\s+', ' ', text)
df['Visi      dan      Misi']      =      df['Visi      dan
Misi'].apply(remove_whitespace_multiple)
#remove single char
def remove_singl_char(text):
    return re.sub(r"\b[a-zA-Z]\b", "", text)

```

```

df['Visi dan Misi'] = df['Visi dan
Misi'].apply(remove_singl_char)

#nltk word tokenize
def word_tokenize_wrapper(text):
    return word_tokenize(text)

df['visimisi_token'] = df['Visi dan
Misi'].apply(word_tokenize_wrapper)

print('Hasil Tokenizing : \n')
print(df['visimisi_token'].head(20))
print('\n\n\n')

```

```

Hasil Tokenizing :
0 [visi, menjadi, pusat, ilmu, pengetahuan, tekn...
1 [visi, universitas, gajah, mada, sebagai, pel...
2 [visi, ub, mempunyai, visi, menjadi, perguruan...
3 [visi, menjadi, perguruan, tinggi, berbasis, r...
4 [visi, menjadi, universitas, yang, mandiri, in...
5 [visi, menjadi, pusat, pengembangan, ilmu, pen...
6 [visi, usk, memiliki, visi, menjadi, universit...
7 [visi, menjadi, perguruan, tinggi, berkelas, d...
8 [visi, menjadi, research, and, entrepreneurial...
9 [visi, menjadi, perguruan, tinggi, yang, ungu...
10 [visi, universitas, kelas, dunia, yang, membin...
11 [visi, menjadi, perguruan, tinggi, yang, memil...
12 [visi, universitas, lampung, menjadi, center, ...
13 [visi, menjadi, universitas, terkemuka, dan, b...
14 [visi, menjadi, universitas, riset, unggul, be...
15 [visi, unpad, dalam, renstra, adalah, mencapai...
16 [visi, unsoed, dalam, pengembangannya, akan, m...
17 [visi, menjadi, universitas, yang, unggul, dal...
18 [visi, pusat, unggulan, dalam, pengembangan, i...
19 [visi, menjadi, perguruan, tinggi, terkemuka, ...
Name: visimisi_token, dtype: object

```

Gambar 4.2 Hasil *Tokenizing*

### 3. Stopword Removal

Pada proses ini yaitu penghapusan kata-kata umum. Contoh beberapa *stopwords* umum dalam bahasa Indonesia adalah “yang”, “di”, “dan”, “dari”,

“adalah”, “ke”, “oleh”, dan sebagainya. Kata-kata ini disebut "stopword" karena sering muncul dalam teks tanpa memberikan kontribusi signifikan dalam pemahaman konten teks. Berikut *source code* *stopword removal*:

```

from nltk.corpus import stopwords

list_stopwords = stopwords.words('indonesian')

#ini kalau manual

list_stopwords.extend(["yg", "dg", "dgn", "dll", "visi",
"misi","dan"])

#read texr stopwords using pandas
txt_stopword = pd.read_csv("politeknik 10.csv", names=
["stopwords"], header = None)

#convert stopwords list dan yg manual

list_stopwords.extend(txt_stopword["stopwords"][0].split('
'))

#convert list to kamus

list_stopwords = set(list_stopwords)

#remove stopwords pada list token

def stopwords_removal(words):

    return [word for word in words if word not in

list_stopwords]

df['visimisi_token_WSW'] =

df['visimisi_token'].apply(stopwords_removal)

print(df['visimisi_token_WSW'].head(20))

```

```
{x} 0 [pusat, ilmu, pengetahuan, teknologi, kebudaya...
1 [universitas, gajah, mada, pelopor, perguruan...
2 [ub, perguruan, pelopor, pembaharu, reputasi, ...
3 [perguruan, berbasis, riset, berstandar, inter...
4 [universitas, mandiri, inovatif, terkemuka, ti...
5 [pusat, pengembangan, ilmu, pengetahuan, tekno...
6 [usk, memiliki, universitas, sosioteknopreneur...
7 [perguruan, berkelas, dunia, berkontribusi, ke...
8 [research, and, entrepreneurial, university, b...
9 [perguruan, unggul, bermartabat, mandiri, diak...
10 [universitas, kelas, dunia, membina, memberday...
11 [perguruan, memiliki, keunggulan, akademik, ba...
12 [universitas, lampung, center, of, excellence,...
13 [universitas, terkemuka, bermartabat, menyelen...
14 [universitas, riset, unggul, bermartabat, bida...
15 [unpad, renstra, mencapai, universitas, berepu...
16 [unsoed, pengembangannya, mewujudkan, dirumusk...
17 [universitas, unggul, pengembangan, ilmu, peng...
18 [pusat, unggulan, pengembangan, insani, ilmu, ...
19 [perguruan, terkemuka, berbasis, riset, unggul...
Name: visimisi_token_WSW, dtype: object
```

Gambar 4.3 Hasil *Stopword Removal*

#### 4. Stemming

Pada proses ini yaitu untuk menghasilkan bentuk dasar atau kata dasar (*root word*). Tujuannya adalah untuk mengurangi kata-kata menjadi bentuk yang lebih umum atau seragam sehingga mempermudah analisis teks. Berikut *source code stemming*:

```
#stemming

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

import swifter

#create stemmer

factory = StemmerFactory()

stemmer = factory.create_stemmer()

#stemmed

def stemmed_wrapper(term):

    return stemmer.stem(term)

term_dict = {}

for document in df['visimisi_normalized']:
```

```

for term in document:
    if term not in term_dict:
        term_dict[term] = ' '
print(len(term_dict))
print("-----")

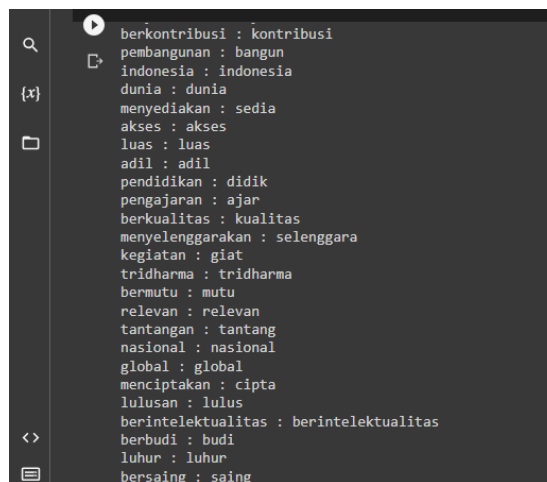
for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term,":", term_dict[term])
print(term_dict)
print("-----")

#apply stemmed term to dataframe

def get_stemmed_term(document):
    return [term_dict[term] for term in document]

df['visimisi_token_stemmed'] =
df['visimisi_normalized'].swifter.apply(get_stemmed_term)
print(df['visimisi_token_stemmed'])

```



```

berkontribusi : kontribusi
pembangunan : bangun
indonesia : indonesia
dunia : dunia
menyediakan : sedia
akses : akses
luas : luas
adil : adil
pendidikan : didik
pengajaran : ajar
berkualitas : kualitas
menyelenggarakan : selenggara
kegiatan : giat
tridharma : tridharma
bermutu : mutu
relevan : relevan
tantangan : tantang
nasional : nasional
global : global
menciptakan : cipta
lulusan : lulus
berintelektualitas : berintelektualitas
berbudi : budi
luhur : luhur
bersaing : saing

```

Gambar 4.4 Hasil Stemming

#### 4.1.2 Pembobotan TF-IDF

Langkah selanjutnya yaitu melakukan TF-IDF. Dalam TF-IDF, setiap teks visi dan misi akan diubah menjadi vektor berdasarkan bobot kata-kata yang muncul di dalamnya. Setiap kata dalam teks akan diberi skor berdasarkan seberapa sering kata tersebut muncul dalam teks (*Term Frequency*) dan seberapa umum kata tersebut di seluruh teks (*Inverse Document Frequency*).

	abad	abdi	academi	aceh	acu	adab	adaptasi	adaptif	adil	administ	agama	agen	agris	agribisni	agroindu	ahil	ahlussur	aik	ajar	akademi	akademi	akadem
2	1	0	0	0	0	0	0	0	0.1552	0	0	0	0	0	0	0	0	0	0.0981	0.1169	0	0
3	2	0	0.1772	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	0	0.0493	0	0	0	0	0	0.1128	0	0	0.3255	0	0	0	0	0	0	0	0	0	0
5	4	0	0	0	0	0	0	0	0	0	0	0	0.2153	0	0	0	0	0	0	0	0	0
6	5	0	0.0464	0	0	0	0	0	0	0	0.3897	0	0	0	0	0.1437	0	0	0	0	0.0801	0
7	6	0	0.075	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1085	0	0	0
8	7	0	0.0508	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0876	0	0
9	8	0	0.1789	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	9	0	0	0	0	0.1477	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	11	0	0.0232	0	0	0	0	0	0	0	0	0.0767	0	0	0	0	0	0	0.1008	0	0	0
13	12	0	0.0433	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1495	0	0
14	13	0	0	0	0	0	0	0.2109	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	14	0	0.061	0	0	0	0.1789	0	0	0	0	0	0	0	0	0	0	0	0	0.1053	0	0
16	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	16	0	0.0251	0.0904	0	0	0.0736	0	0	0	0	0	0	0	0	0	0	0	0.0727	0	0	0
18	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0849	0	0	0
19	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1669	0	0	0
21	20	0	0.0472	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0682	0	0	0
22	21	0	0.0404	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.117	0	0	0
23	22	0	0	0	0	0	0	0	0	0	0.1559	0	0	0	0	0	0	0	0	0.0481	0	0
24	23	0	0.0885	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1281	0	0	0
25	24	0	0.0459	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0791	0	0
26	25	0	0.0584	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0845	0	0.1634	0
27	26	0	0.0455	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0785	0	0
28	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0928	0	0.1778
29	28	0	0.0297	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.043	0	0	0
30	29	0	0.0389	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	30	0	0.0787	0	0	0	0.2307	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	31	0	0.0515	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0888	0	0
33	32	0	0.1607	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2894	0	0	0	0	0
34	33	0	0.0573	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0829	0	0	0
35	34	0	0.0516	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0746	0.0889	0	0
36	35	0	0.0647	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	36	0	0.0694	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1004	0	0	0
38	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2665	0
39	38	0	0.1057	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Gambar 4. 5 Hasil Pembobotan TF-IDF

#### 4.1.3 Perbandingan Menggunakan *Cosine Similarity*

*Cosine similarity* digunakan untuk mengukur sejauh mana dua vektor berorientasi serupa atau sejajar dalam ruang vektor. Dalam konteks penelitian ini *cosine similarity* digunakan untuk mengukur tingkat kemiripan antara teks visi dan misi dari berbagai universitas di Indonesia. Dengan menggunakan *cosine similarity*, dapat menghitung sudut kosinus antara dua vektor representasi teks visi dan misi.

Nilai cosine similarity berkisar antara 0 hingga 1, dengan 0 menunjukkan ketidaksamaan dan 1 menunjukkan kesamaan.

Pada penelitian ini, cosine similarity membantu dalam menentukan sejauh mana visi dan misi perguruan tinggi serupa satu sama lain. Dengan membandingkan kemiripan antara pasangan teks visi dan misi, dapat mengidentifikasi universitas yang memiliki tujuan dan fokus serupa dalam pendidikan tinggi. Juga dapat membantu dalam memahami pola dan nilai yang terkandung atau sering muncul pada pendidikan tinggi di Indonesia berdasarkan visi dan misi. Berikut *source code cosine similarity*:

```
#tf-idf
tfidf = TfidfTransformer(use_idf=True, norm='l2',
smooth_idf=True)

tf =
tfidf.fit_transform(vectorizer.fit_transform(document)).toarray()

dfb = pd.DataFrame(data=tf, index=list(range(1,
len(tf[:,1])+1, )), columns=[a])

dfb

#coba cosine
from sklearn.metrics.pairwise import cosine_similarity

cs = cosine_similarity(tf, tf)

df_cs = pd.DataFrame(cs, index=list(range(1, len(tf[:,1])+1,
)), columns=list(range(1, len(tf[:,1])+1, )))

df_cs
```



	1	2	3	4	5	6	7	8	9	10	...	91	92	93	94	95	96
1	1.000000	0.118578	0.180053	0.151323	0.131161	0.226414	0.170634	0.184779	0.049188	0.172006	...	0.236225	0.094301	0.091082	0.102382	0.056694	0.110353
2	0.118578	1.000000	0.232661	0.132452	0.285596	0.114916	0.115682	0.222117	0.060625	0.142102	...	0.214074	0.151181	0.009689	0.156779	0.082007	0.210467
3	0.180053	0.232661	1.000000	0.254371	0.275001	0.288150	0.141240	0.180972	0.144495	0.099640	...	0.279611	0.092488	0.040250	0.237663	0.179575	0.136713
4	0.151323	0.132452	0.254371	1.000000	0.142437	0.232046	0.230953	0.188799	0.121562	0.145358	...	0.238995	0.108027	0.163687	0.191514	0.120538	0.059014
5	0.131161	0.285596	0.275001	0.142437	1.000000	0.287135	0.206767	0.250975	0.137863	0.151298	...	0.247074	0.215599	0.095001	0.168220	0.115543	0.203281
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
96	0.110353	0.210467	0.136713	0.059014	0.203281	0.134211	0.082433	0.118895	0.066347	0.064125	...	0.110801	0.183656	0.090080	0.370054	0.251201	1.000000
97	0.172624	0.105390	0.164272	0.212094	0.199658	0.171798	0.242492	0.132943	0.126695	0.108858	...	0.234006	0.121050	0.123838	0.162545	0.113812	0.186280
98	0.150656	0.106449	0.236228	0.259035	0.218483	0.357820	0.289808	0.189433	0.172939	0.154206	...	0.270420	0.116428	0.140203	0.182183	0.161723	0.153283
99	0.020556	0.005182	0.012536	0.011105	0.010860	0.031512	0.009436	0.005230	0.013808	0.008609	...	0.019038	0.034460	0.000000	0.080435	0.119072	0.120814
100	0.078675	0.082850	0.112358	0.088929	0.106673	0.116984	0.060450	0.098268	0.094072	0.069027	...	0.141140	0.132535	0.039286	0.216388	0.151034	0.287600

Gambar 4. 6 Hasil Perbandingan

Setelah melakukan berbagai langkah proses di atas, mulai dari *preprocessing*, penghitungan pembobotan TD-IDF, dan perhitungan *cosine similarity* untuk mendapatkan hasil perbandingan kemiripan teks visi dan misi antar perguruan tinggi. Dengan universitas 1 (Universitas Indonesia) menjadi pembanding dengan universitas lain dihasilkan bahwa nomor perguruan tinggi yang paling tinggi tingkat kemiripannya dengan nomor 1 adalah perguruan tinggi nomor 75 (Universitas Esa Unggul) yaitu sebesar 0.310447484. Sedangkan tingkat kemiripan terendah yaitu dengan nomor 99 (IAIN Kendari) sebesar 0.020555745. Jadi, tingkat kemiripan tertinggi perguruan tinggi terhadap universitas 1 yaitu universitas 75 adalah sekitar 31.04% dan terendah yaitu universitas 99 sekitar 2.06%. Berikut adalah hasil uji coba melakukan perbandingan kemiripan menggunakan *blackbox testing*.

Tabel 4. 1 Pengujian *Blackbox Testing*

No.	Komponen Uji	Skenario Uji	Hasil yang diharapkan	Hasil Uji

1.	Input NLTK	Memasukkan kode NLTK untuk proses preprocessing	Sistem menampilkan keberhasilan running kode NLTK	Berhasil
1.	Input data	Input database berbentuk csv ke dalam sistem	Sistem akan menyimpan data	Berhasil
2.	Proses <i>preprocessing</i>	Melakukan proses preprocessing yaitu case folding, tokenizing, stopword removal dan stemming.	Sistem akan berhasil proses preprocessing	Berhasil
3.	Perhitungan pembobotan TF-IDF	Melakukan perhitungan pembobotan	Sistem menghasilkan pembobotan TD-IDF	Berhasil

		dengan TF-IDF		
4.	Perhitungan <i>Cosine Similarity</i>	Memasukkan kode untuk menghasilkan perbandingan	Sistem dapat menampilkan hasil perbandingan teks visi dan misi antar universitas	Berhasil

## 4.2 Hasil Uji Coba

Untuk mengidentifikasi fokus dan tujuan berdasarkan visi dan misi universitas di Indonesia, penelitian ini melakukan analisis identifikasi kata-kata yang menonjol dalam teks visi dan misi. Data yang diambil yaitu dari hasil pembobotan TF-IDF. Metode ini dapat membantu mengungkapkan kata-kata yang paling sering muncul dan memberikan bobot penting pada kata-kata tersebut.

Hasil dari analisis ini dapat berupa daftar kata-kata yang paling umum atau sering muncul dalam teks visi dan misi universitas. Berikut kata-kata teks visi dan misi yang sering muncul.

Tabel 4. 2 Kata-kata yang sering muncul berdasarkan analisis TF-IDF

Abdi	Daya	Hidup	Nasional
Ajar	Didik	Ilmu	profesional
Akademik	Dunia	Inovatif	Selenggara

Bangsa	Giat	internasional	Teknologi
Bangun	Global	Kembang	Teliti
Bidang	Guru	Kualitas	Unggul
Budaya	Hasil	Masyarakat	Wujud

Dengan melihat daftar kata-kata yang sering muncul diatas tersebut, bahwa bisa mendapatkan wawasan tentang fokus dan nilai yang diungkapkan dalam visi dan misi universitas. Hal ini dapat membantu dalam pemahaman tentang prioritas dan tujuan yang ingin dicapai oleh universitas tersebut.

Visi dan misi sebuah universitas dapat mempengaruhi statusnya sebagai universitas atau institut dengan mempertimbangkan karakteristik dan fokus kegiatan pendidikan dan penelitiannya. Sebuah perguruan tinggi dapat memilih status sebagai universitas atau institut berdasarkan visi dan misinya. Visi dan misi perguruan tinggi dapat menjadi acuan dalam menentukan karakteristik dan fokus kegiatan pendidikan dan penelitiannya, yang dapat mempengaruhi status perguruan tinggi tersebut.

Namun demikian, visi dan misi perguruan tinggi tidak menjadi satu-satunya faktor dalam menentukan statusnya sebagai universitas atau institut. Perguruan tinggi juga perlu memperhatikan hal-hal lain seperti jumlah fakultas dan program studi yang dimiliki, kualitas dan akreditasi program studi, serta tugas utama perguruan tinggi. Pada penelitian ini adalah untuk mengetahui bagaimana ciri bahwa visi dan misi perguruan tinggi tersebut apakah universitas, institut ataupun

politeknik dengan menggunakan pembobotan TD-IDF. Berikut proses pembobotan TF-IDF untuk mengetahui ciri visi dan misi perguruan tinggi.

1. Mengelompokkan 10 universitas, institut dan politeknik teratas menurut *webometric 2023*.
2. Preprocessing data-data yang telah disiapkan.
3. Melakukan pembobotan pada tiap top 10 universitas, institut dan politeknik.
4. Mendapatkan hasil bagaimana ciri teks visi dan misi universitas, institut dan politeknik.

Tabel 4. 3 Pembobotan TD-IDF Universitas pas 5

Universitas	
Bas	Guru
Budaya	Lulus
Dunia	Saing

Tabel 4. 4 Pembobotan TF-IDF Universitas diatas 5

Universitas		
Abdi	Internasional	Tahu
Bangsa	Kembang	Teknologi
Daya	Manfaat	Teliti
Didik	Masyarakat	Tingkat
Ilmu	Selenggara	Unggul

Tabel 4. 5 Pembobotan TF-IDF Institut pas 5

Institut
----------

Abdi
Tahu

Tabel 4. 6 Pembobotan TF-IDF Institut diatas 5

Institut		
Bas	Kembang	Unggul
Didik	Masyarakat	
Guru	Selenggara	
Hasil	Teknologi	
Ilmu	Teliti	

Tabel 4. 7 Pembobotan TF-IDF Politeknik pas 5

Politeknik
Institusi
Kualitas
Mutu
Vokasi

Tabel 4. 8 Pembobotan TF-IDF Politeknik diatas 5

Politeknik		
Abdi	Kembang	Selenggara
Bangun	Laksana	Teknologi
Daya	Masyarakat	Teliti
Didik	Politeknik	Tingkat

Global	Saing	Unggul
--------	-------	--------

Dari hasil pembobotan TF-IDF pada 3 pengelompokan perguruan tinggi yaitu universitas, institut dan politeknik diatas yaitu dapat diketahui ada perbedaan dalam pembuatan teks visi dan misi. Jadi, dapat disimpulkan bahwa ciri teks visi dan misi universitas dilihat dari kata yang beda atau tidak ada dalam institut dan politeknik adalah “bangsa”, “internasional”, dan “manfaat”. Sedangkan untuk institut cirinya adalah terdapat kata “bas” dan “guru”. Pada politeknik yaitu “bangun”, “global”, “laksana”, dan “politeknik”. Dari situlah bisa didapatkan ciri visi dan misi universitas, institut dan politeknik.

### 4.3 Integrasi Islam

Penelitian ini dilakukan untuk membandingkan visi dan misi perguruan tinggi. Allah SWT berfirman dalam QS Al hujurat ayat 15 :

إِنَّمَا الْمُؤْمِنُونَ الَّذِينَ آمَنُوا بِاللَّهِ وَرَسُولِهِ ثُمَّ لَمْ يَرْتَابُوا وَجَاهَدُوا بِأَمْوَالِهِمْ وَأَنْفُسِهِمْ فِي سَبِيلِ اللَّهِ  
أُولَئِكَ هُمُ الصَّادِقُونَ

Artinya : “*Sesungguhnya orang-orang yang beriman itu hanyalah orang-orang yang percaya (beriman) kepada Allah dan Rasul-Nya, kemudian mereka tidak ragu-ragu dan mereka berjuang (berjihad) dengan harta dan jiwa mereka pada jalan Allah. Mereka itulah orang-orang yang benar*”. (QS Al hujurat ayat 15)

Tafsir as-Sa'di / Syaikh Abdurrahman bin Nashir as-Sa'di, pakar tafsir abad 14 H: “Sesungguhnya orang-orang yang beriman,” yakni secara hakiki, ialah “orang-orang yang beriman kepada Allah dan RasulNya kemudian mereka tidak ragu-ragu dan mereka berjihad dengan harta dan jiwa mereka pada jalan Allah.”

Yakni, orang yang menyatukan antara keimanan terhadap Allah dan RasulNya dengan jihad di jalanNya; sebab orang yang menegakkan jihad terhadap kaum kafir itu menunjukkan kesempurnaan imannya di dalam hati. Sebab orang yang memerangi kaum kafir atas nama Islam dan iman serta menunaikan syariat-syariat Allah itu, tentu sudah pasti telah berjihad terhadap dirinya sendiri. Orang yang tidak mampu berjihad menunjukkan kelemahan imannya. Dalam beriman, Allah mensyaratkan tidak adanya keraguan, karena iman yang bermanfaat itu adalah tekad bulat dan keyakinan terhadap perintah Allah untuk beriman padaNya yang tidak disertai dengan keraguan sedikit pun.

Firman Allah, “Mereka itulah orang-orang yang benar,” yakni, orang-orang yang membuktikan keimanan mereka dengan perbuatan-perbuatan baik. Kejujuran adalah pengakuan besar terhadap segala hal yang diakui, dan orang yang jujur memerlukan hujjah dan bukti, dan bukti terbesar adalah klaim keimanan yang merupakan pusat kebahagiaan dan kemenangan abadi, maka siapa pun yang mengaku beriman dan mengerjakan kewajiban serta keharusan iman, maka ia adalah orang jujur dan Mukmin sejati. Sebaliknya orang yang tidak demikian halnya dapat diketahui bahwa pengakuannya tidak benar, sehingga pengakuannya tidak berguna sama sekali, sebab keimanan dalam hati itu hanya diketahui oleh Allah semata. Maka penegasan atau penafian keimanan (di sini) adalah tindakan yang mengajari Allah terhadap sesuatu yang ada di dalam hati dan ini merupakan etika dan dugaan yang tidak baik terhadap Allah.

Tema dari penelitian ini adalah "perbandingan visi misi organisasi", yang mengacu pada suatu usaha untuk membandingkan visi dan misi organisasi atau



perusahaan untuk mengevaluasi kesesuaian dan perbedaan antara tujuan organisasi dan tindakan nyata organisasi. Ayat Al-Quran yang dipilih menekankan pentingnya kepercayaan tanpa keraguan pada Allah dan Rasul-Nya serta berjihad dengan harta dan jiwa di jalan Allah. Tafsir ayat tersebut mengungkapkan bahwa orang yang beriman secara hakiki adalah mereka yang menggabungkan antara keimanan dengan jihad di jalan Allah. Dalam konteks perbandingan visi misi organisasi, dapat diartikan bahwa orang-orang yang percaya dan berjuang dengan tekad bulat dan keyakinan akan mencapai tujuan organisasi dengan lebih efektif.

Penggabungan antara tema, ayat, dan tafsir menekankan pentingnya memiliki kepercayaan dan tekad bulat untuk mencapai tujuan organisasi yang telah ditetapkan. Ayat tersebut juga mengajarkan bahwa keraguan tidak memiliki tempat dalam keimanan dan perbuatan baik, dan orang yang memiliki keimanan yang kuat juga akan memiliki tindakan yang baik. Oleh karena itu, dalam membandingkan visi dan misi organisasi, hal yang paling penting adalah memiliki keyakinan yang kuat pada tujuan organisasi serta berkomitmen untuk mencapai tujuan tersebut dengan tindakan yang konsisten dan efektif.

Kesimpulannya, tema "perbandingan visi misi organisasi" atau kalau dalam penelitian ini yaitu perbandingan visi dan misi universitas dapat dikaitkan dengan ayat Al-Quran dan tafsir yang dipilih dalam rangka meningkatkan pemahaman tentang pentingnya kepercayaan dan tekad bulat untuk mencapai tujuan organisasi. Dalam konteks perbandingan visi misi organisasi, keimanan yang kuat dan tindakan yang konsisten dapat membantu organisasi mencapai tujuan mereka dengan lebih efektif. Oleh karena itu, hal ini dapat menjadi pedoman bagi organisasi atau

perguruan tinggi dalam mengevaluasi dan memperbarui visi dan misi mereka untuk memastikan kesesuaian antara tujuan organisasi dan tindakan nyata organisasi.

Ayat Al-Quran: "*Dan janganlah kamu campuradukkan yang hak dengan yang bathil, dan janganlah kamu sembunyikan yang hak itu, sedangkan kamu mengetahui.*" (Al-Baqarah: 42). Ayat ini menggarisbawahi pentingnya membedakan antara yang benar (hak) dan yang salah (bathil). Dalam penelitian ini, konsep ini dapat dihubungkan dengan memastikan bahwa perbandingan teks visi dan misi perguruan tinggi dilakukan secara obyektif dan tidak membingungkan antara hal-hal yang benar-benar serupa dengan yang berbeda.

Ayat Al-Quran: "*Katakanlah: Jika Allah menghendaki, aku tidak akan membacakan kepada kalian, dan Dia tidak akan menyampaikan kepada kalian, sedikitpun dari ilmu-Nya. Aku telah tinggal di tengah-tengah kalian seumur hidup sebelum ini. Maka mengapa kamu tidak menggunakan akal?*" (Yunus: 16). Ayat ini menggarisbawahi pentingnya menggunakan akal dan pengetahuan yang ada untuk memperoleh pemahaman yang lebih baik. Dalam penelitian ini, menggunakan algoritma cosine similarity untuk membandingkan teks visi dan misi perguruan tinggi. Ayat ini dapat menegaskan pentingnya menggunakan alat dan metode yang ada untuk mendapatkan pemahaman yang lebih akurat dan objektif.

Ayat Al-Quran: "*Dan Kami telah membandingkan bagi manusia dalam Al-Quran ini berbagai macam perumpamaan, tetapi kebanyakan manusia tidak melainkan ingkar.*" (Al-Isra: 89). Tafsir Ibnu Katsir: Ayat ini menunjukkan bahwa Allah SWT memberikan berbagai macam perumpamaan dalam Al-Quran untuk memberikan pemahaman dan gambaran bagi manusia. Tujuannya adalah untuk

memberikan perbandingan yang jelas dan dapat dipahami agar manusia dapat mengambil pelajaran dan petunjuk. Namun, sayangnya, kebanyakan manusia mengabaikan perumpamaan-perumpamaan tersebut dan bahkan enggan untuk memahaminya dengan baik.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa:

1. Metode *Cosine Similarity* dapat digunakan untuk perbandingan teks visi dan misi antar perguruan tinggi.
2. Melalui *cosine similarity*, dapat diidentifikasi skor kesamaan kosinus antara visi dan misi perguruan tinggi, yang memberikan indikasi tentang sejauh mana mereka saling mirip atau berbeda. Jika, nilainya 1 mirip.
3. Dengan universitas 1 (Universitas Indonesia) menjadi pembanding dengan universitas lain dihasilkan bahwa nomor perguruan tinggi yang paling tinggi tingkat kemiripannya dengan nomor 1 adalah perguruan tinggi nomor 75 (Universitas Esa Unggul) yaitu sebesar 0.310447484. Sedangkan tingkat kemiripan terendah yaitu dengan nomor 99 (IAIN Kendari) sebesar 0.020555745. Jadi, tingkat kemiripan tertinggi perguruan tinggi terhadap universitas 1 yaitu universitas nomor 75 adalah sekitar 31.04% dan terendah yaitu universitas nomor 99 sekitar 2.06%.
4. Melalui hasil dari perhitungan TF-IDF yaitu daftar kata-kata yang sering muncul, bahwa bisa mendapatkan wawasan tentang fokus dan nilai yang diungkapkan dalam visi dan misi universitas. Hal ini dapat membantu dalam pemahaman tentang prioritas dan tujuan yang ingin dicapai oleh perguruan tinggi tersebut.

5. Selain itu, melalui perhitungan atau pembobotan TF-IDF didapatkan ciri visi dan misi universitas, institut, dan politeknik dengan cara mengelompokkan tiap-tiap top 10 universitas, politeknik, dan institut. Hasil TF-IDF dari pengelompokkan didapatkan kata yang sering muncul dan dibandingkan dengan kategori lain untuk mencari perbedaan dan hasilnya dari universitas yaitu “bangsa”, “internasional”, dan “manfaat”. Institut yaitu “hasil”, sedangkan politeknik yaitu “bangun”, “global”, “laksana”, dan “politeknik”.

## 5.2 Saran

Penulis dapat menyadari masih terlihat banyak kekurangan dan kelemahan dalam penelitian ini sehingga pengembangan selanjutnya perlu ditingkatkan untuk mendapatkan hasil yang lebih baik. Beberapa saran penulis untuk penelitian selanjutnya adalah sebagai berikut.

1. Dataset yang digunakan pada penelitian ini hanya top 100 Universitas Indonesia dan untuk *clustering* mengambil 10 teratas universitas, institut, dan politeknik menurut *webometric*. Untuk penelitian selanjutnya bisa menggunakan lebih banyak lagi.
2. Lakukan analisis mendalam terhadap hasil perbandingan dan identifikasi perbedaan yang signifikan serta keunikan antara perguruan tinggi dalam visi dan misi mereka.
3. Pertimbangkan penggunaan metode lain selain cosine similarity, seperti metode pemrosesan bahasa alami yang lebih kompleks atau pendekatan machine learning, untuk memperoleh hasil yang lebih akurat dan representatif.

4. Untuk pengembangan berikutnya bisa dibuat pencarian universitas dengan memasukkan visi dan misi yang diinginkan.

## DAFTAR PUSTAKA

- Adebanke Olusola, S., Olaleke Oluseye, O., Menyene Saviour, U., Joy Iember, K., & Olamilekan Ayomiposi, D. (2022). A Content analysis of the vision and mission statements of top ten leading Universities in Africa. *Cogent Education*, 9, 2143648. <https://doi.org/10.1080/2331186X.2022.2143648>
- Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology (BITE)*, 3, 19–27. <https://doi.org/10.30812/bite.v3i1.1110>
- Brankovic, J., & Cantwell, B. (2022). Making sense of change in higher education research: Exploring the intersection of science and policy. *Higher Education*, 84, 1207–1226. <https://doi.org/10.1007/s10734-022-00928-3>
- Dooris, M. J., Kelley, J. M., & Trainer, J. F. (2004). Strategic planning in higher education. *New Directions for Institutional Research*, 2004, 5–11. <https://doi.org/10.1002/ir.115>
- Dumanig, F. P., & Symaco, L. P. (2022). Internationalisation of higher education in Malaysia and the Philippines: A comparative analysis of mission and vision statements of selected universities. *Journal of Multilingual and Multicultural Development*, 43, 154–166. <https://doi.org/10.1080/01434632.2020.1735401>
- Hafizin, H., & Herman, H. (2022). Merumuskan Visi dan Misi Lembaga Pendidikan. *Islamic Management: Jurnal Manajemen Pendidikan Islam*, 5, 99. <https://doi.org/10.30868/im.v5i01.2095>
- Hamdan, Y. (2001). *PERNYATAAN VISI DAN MISI PERGURUAN TINGGI*.
- Kohila, R., & Arunesh, Dr. K. (2016). *Text Mining: Text Similarity Measure For News Articles Based On String Based Approach*. <https://doi.org/10.5281/ZENODO.57373>
- Lin, T.-C. (2004). The role of higher education in economic development: An empirical study of Taiwan case. *Journal of Asian Economics*, 15, 355–371. <https://doi.org/10.1016/j.asieco.2004.02.006>
- Purba, R. A., Suparno, S., & Giatman, M. (2020). The optimalization of cosine similarity method in detecting similarity degree of final project by the college students. *IOP Conference Series: Materials Science and Engineering*, 830, 032003. <https://doi.org/10.1088/1757-899X/830/3/032003>

- Qaiser, S., & Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, 181, 25–29. <https://doi.org/10.5120/ijca2018917395>
- Taiwo, A. A., Agwu, E., & Lawal, F. A. (2016). *Vision and Mission In Organization: Myth or Heuristic Device? 4*.
- Talib, R., Kashif, M., Ayesha, S., & Fatima, F. (2016). Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications*, 7. <https://doi.org/10.14569/IJACSA.2016.071153>
- Thada, V. (2013). Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm. *International Journal of Innovations in Engineering and Technology*, 2.
- Topcu, Z. (2020). The Future Of Higher Education In The Digital age. *Journal of International Civilization Studies*, 5, 211–229.
- Turban, E., Sharda, R., & Delen, D. (2011). *Decision support and business intelligence systems* (9th ed). Prentice Hall.
- Wahyuni, R. T., & Prastiyanto, D. (2017). *Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi*. 9.
- Ye, J. (2011). Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modelling*, 53, 91–97. <https://doi.org/10.1016/j.mcm.2010.07.022>
- Zhang, Y., Chen, M., & Liu, L. (2015). A review on text mining. *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 681–685. <https://doi.org/10.1109/ICSESS.2015.7339149>





24	1	0.11	0.2	0.29	0.2	0.13	0.08	0.2	0.13	0.18	0.16	0.11	0.22	0.15	0.06	0.17	0.11	0.09	0.18	0.04	0.11	0.09	0.22
25	0.11	1	0.17	0.11	0.18	0.13	0.09	0.09	0.15	0.09	0.06	0.1	0.09	0.15	0.15	0.14	0.04	0.12	0.11	0.14	0.12	0.11	0.17
26	0.2	0.17	1	0.15	0.13	0.11	0.04	0.14	0.17	0.27	0.13	0.15	0.11	0.15	0.15	0.12	0.06	0.15	0.14	0.09	0.21	0.16	0.2
27	0.29	0.11	0.15	1	0.16	0.13	0.06	0.2	0.14	0.16	0.14	0.26	0.15	0.08	0.06	0.11	0.06	0.07	0.23	0.14	0.11	0.04	0.25
28	0.2	0.18	0.13	0.16	1	0.1	0.08	0.13	0.13	0.18	0.06	0.14	0.13	0.14	0.04	0.17	0.07	0.12	0.09	0.16	0.1	0.08	0.21
29	0.13	0.13	0.11	0.13	0.1	1	0.1	0.1	0.13	0.09	0.07	0.11	0.1	0.09	0.1	0.16	0.08	0.15	0.12	0.07	0.12	0.08	0.11
30	0.08	0.09	0.04	0.06	0.08	0.1	1	0.07	0.1	0.08	0.06	0.04	0.09	0.13	0.02	0.07	0.09	0.1	0.05	0.03	0.1	0.03	0.11
31	0.2	0.09	0.14	0.2	0.13	0.1	0.07	1	0.12	0.12	0.08	0.18	0.1	0.06	0.06	0.13	0.15	0.06	0.12	0.43	0.11	0.05	0.14
32	0.13	0.15	0.17	0.14	0.13	0.13	0.1	0.12	1	0.15	0.07	0.12	0.11	0.14	0.07	0.15	0.11	0.13	0.1	0.13	0.14	0.06	0.22
33	0.18	0.09	0.27	0.16	0.18	0.09	0.08	0.12	0.15	1	0.1	0.16	0.21	0.1	0.02	0.15	0.09	0.19	0.1	0.17	0.15	0.13	0.13
34	0.16	0.06	0.13	0.14	0.06	0.07	0.06	0.08	0.07	0.1	1	0.08	0.1	0.2	0.04	0.11	0.11	0.07	0.08	0.09	0.06	0.04	0.1
35	0.11	0.1	0.15	0.26	0.14	0.11	0.04	0.18	0.12	0.16	0.08	1	0.11	0.08	0.06	0.16	0.07	0.11	0.1	0.13	0.14	0.06	0.19
36	0.22	0.09	0.11	0.15	0.13	0.1	0.09	0.1	0.11	0.21	0.1	0.11	1	0.09	0.04	0.25	0.03	0.11	0.08	0.08	0.09	0.08	0.17
37	0.15	0.15	0.15	0.08	0.14	0.09	0.13	0.06	0.14	0.1	0.2	0.08	0.09	1	0.06	0.14	0.12	0.12	0.09	0.08	0.13	0.15	0.2
38	0.06	0.15	0.15	0.06	0.04	0.1	0.02	0.06	0.07	0.02	0.04	0.06	0.04	0.06	1	0.06	0.04	0.06	0.16	0.12	0.13	0.04	0.05
39	0.17	0.14	0.12	0.11	0.17	0.16	0.07	0.13	0.15	0.15	0.11	0.16	0.25	0.14	0.06	1	0.05	0.2	0.08	0.15	0.14	0.1	0.14
40	0.11	0.04	0.06	0.06	0.07	0.08	0.09	0.15	0.11	0.09	0.11	0.07	0.03	0.12	0.04	0.05	1	0.04	0.06	0.09	0.1	0.02	0.1
41	0.09	0.12	0.15	0.07	0.12	0.15	0.1	0.06	0.13	0.19	0.07	0.11	0.11	0.12	0.06	0.2	0.04	1	0.08	0.09	0.1	0.12	0.08
42	0.18	0.11	0.14	0.23	0.09	0.12	0.05	0.12	0.1	0.1	0.08	0.1	0.08	0.09	0.16	0.08	0.06	0.08	1	0.08	0.11	0.06	0.16
43	0.04	0.14	0.09	0.14	0.16	0.07	0.03	0.43	0.13	0.17	0.09	0.13	0.08	0.08	0.12	0.15	0.09	0.09	0.08	1	0.11	0.09	0.13
44	0.11	0.12	0.21	0.11	0.1	0.12	0.1	0.11	0.14	0.15	0.06	0.14	0.09	0.13	0.13	0.14	0.1	0.1	0.11	1	0.06	0.11	1
45	0.09	0.11	0.16	0.04	0.08	0.08	0.03	0.05	0.06	0.13	0.04	0.06	0.08	0.15	0.04	0.1	0.02	0.12	0.06	0.09	0.06	1	0.11
46	0.22	0.17	0.2	0.25	0.21	0.11	0.11	0.14	0.22	0.13	0.1	0.19	0.17	0.2	0.05	0.14	0.1	0.08	0.16	0.13	0.11	0.11	1

47	1	0.16	0.07	0.11	0.08	0.17	0.06	0.09	0.12	0.04	0.23	0.1	0.12	0.21	0.15	0.14	0.12	0.11	0.12	0.1	0.19	0.07	0.1
48	0.16	1	0.03	0.14	0.09	0.21	0.07	0.09	0.17	0.1	0.07	0.18	0.14	0.24	0.26	0.11	0.17	0.05	0.09	0.19	0.26	0.05	0.09
49	0.07	0.03	1	0.01	0.07	0.1	0.03	0.03	0.06	0.02	0.04	0.09	0.09	0.05	0.04	0.18	0.05	0.04	0.17	0.07	0.05	0.05	0.02
50	0.11	0.14	0.01	1	0.06	0.14	0.1	0.05	0.23	0.04	0.05	0.09	0.06	0.12	0.09	0.05	0.11	0.04	0.07	0.11	0.11	0.05	0.13
51	0.08	0.09	0.07	0.06	1	0.08	0.03	0.07	0.09	0.06	0.13	0.08	0.09	0.09	0.18	0.09	0.06	0.1	0.13	0.12	0.07	0.08	0.04
52	0.17	0.21	0.1	0.14	0.08	1	0.07	0.1	0.18	0.03	0.1	0.21	0.07	0.26	0.17	0.17	0.12	0.13	0.1	0.11	0.18	0.12	0.13
53	0.06	0.07	0.03	0.1	0.03	0.07	1	0.03	0.05	0.02	0.01	0.01	0.18	0.05	0.04	0.04	0.01	0.1	0.01	0.05	0.09	0.03	0.05
54	0.09	0.09	0.03	0.05	0.07	0.1	0.03	1	0.05	0.08	0.05	0.12	0.09	0.11	0.08	0.13	0.03	0.13	0.09	0.07	0.09	0.03	0.02
55	0.12	0.17	0.06	0.23	0.09	0.18	0.05	0.05	1	0.04	0.1	0.12	0.05	0.11	0.09	0.14	0.18	0.08	0.09	0.12	0.12	0.09	0.1
56	0.04	0.1	0.02	0.04	0.06	0.03	0.02	0.08	0.04	1	0.04	0.07	0.04	0.06	0.07	0.03	0.02	0.05	0.05	0.05	0.14	0.03	0.05
57	0.23	0.07	0.04	0.05	0.13	0.1	0.01	0.05	0.1	0.04	1	0.07	0.06	0.08	0.14	0.16	0.08	0.08	0.09	0.12	0.08	0.07	0.02
58	0.1	0.18	0.09	0.09	0.08	0.21	0.01	0.12	0.12	0.07	0.07	1	0.06	0.17	0.12	0.1	0.14	0.11	0.13	0.11	0.21	0.07	0.05
59	0.12	0.14	0.09	0.06	0.09	0.07	0.18	0.09	0.05	0.04	0.06	0.06	1	0.06	0.17	0.11	0.05	0.13	0.14	0.09	0.13	0.04	0.05
60	0.21	0.24	0.05	0.12	0.09	0.26	0.05	0.11	0.11	0.06	0.08	0.17	0.06	1	0.17	0.2	0.11	0.09	0.11	0.23	0.11	0.05	0.08
61	0.15	0.26	0.04	0.09	0.18	0.17	0.04	0.08	0.09	0.07	0.14	0.12	0.17	0.17	1	0.16	0.05	0.16	0.15	0.12	0.13	0.12	0.11
62	0.14	0.11	0.18	0.05	0.09	0.17	0.04	0.13	0.14	0.03	0.16	0.1	0.11	0.2	0.16	1	0.1	0.21	0.18	0.11	0.13	0.07	0.04
63	0.12	0.17	0.05	0.11	0.06	0.12	0.01	0.03	0.18	0.02	0.08	0.14	0.05	0.11	0.05	0.1	1	0.02	0.07	0.11	0.15	0.07	0.05
64	0.11	0.05	0.04	0.04	0.1	0.13	0.1	0.13	0.08	0.05	0.08	0.11	0.13	0.09	0.16	0.21	0.02	1	0.07	0.08	0.12	0.08	0.04
65	0.12	0.09	0.17	0.07	0.13	0.1	0.01	0.09	0.09	0.05	0.09	0.13	0.14	0.11	0.15	0.18	0.07	0.07	1	0.07	0.1	0.04	0.07
66	0.1	0.19	0.07	0.11	0.12	0.11	0.05	0.07	0.12	0.05	0.12	0.11	0.09	0.23	0.12	0.11	0.11	0.08	0.07	1	0.07	0.12	0.03
67	0.19	0.26	0.05	0.11	0.07	0.18	0.09	0.09	0.12	0.14	0.08	0.21	0.13	0.11	0.13	0.13	0.15	0.12	0.1	0.07	1	0.07	0.13
68	0.07	0.05	0.05	0.05	0.08	0.12	0.03	0.03	0.09	0.03	0.07	0.07	0.04	0.05	0.12	0.07	0.07	0.08	0.04	0.12	0.07	1	0.05
69	0.1	0.09	0.02	0.13	0.04	0.13	0.05	0.02	0.1	0.05	0.02	0.05	0.05	0.08	0.11	0.04	0.05	0.04	0.07	0.03	0.13	0.05	1

70	1	0.35	0.15	0.26	0.1	0.15	0.1	0.15	0.22	0.08	0.1	0.14	0.25	0.12	0.14	0.17	0.11	0.15	0.28	0.12	0.05
71	0.35	1	0.04	0.13	0.06	0.05	0.12	0.06	0.06	0.08	0.08	0.1	0.08	0.05	0.09	0.06	0.06	0.06	0.11	0.06	0.04
72	0.15	0.04	1	0.09	0.09	0.06	0.03	0.17	0.17	0.05	0.04	0.06	0.13	0.08	0.13	0.07	0.16	0.07	0.14	0.15	0.08
73	0.26	0.13	0.09	1	0.05	0.26	0.11	0.28	0.2	0.15	0.14	0.13	0.18	0.12	0.13	0.22	0.09	0.14	0.24	0.11	0.12
74	0.1	0.06	0.09	0.05	1	0.07	0.12	0.05	0.03	0.12	0.04	0.06	0.23	0.04	0.19	0.2	0.1	0.08	0.12	0.14	0.15
75	0.15	0.05	0.06	0.26	0.07	1	0.12	0.23	0.19	0.11	0.09	0.09	0.14	0.05	0.09	0.11	0.06	0.15	0.2	0.08	0.05
76	0.1	0.12	0.03	0.11	0.12	0.12	1	0.07	0.16	0.09	0.04	0.15	0.26	0.06	0.08	0.18	0.13	0.09	0.14	0.05	0.04
77	0.15	0.06	0.17	0.28	0.05	0.23	0.07	1	0.16	0.12	0.09	0.1	0.11	0.12	0.08	0.21	0.09	0.17	0.24	0.1	0.07
78	0.22	0.06	0.17	0.2	0.03	0.19	0.16	0.16	1	0.06	0.05	0.05	0.13	0.08	0.1	0.07	0.08	0.13	0.17	0.06	0.05
79	0.08	0.08	0.05	0.15	0.12	0.11	0.09	0.12	0.06	1	0.08	0.05	0.2	0.09	0.1	0.25	0.17	0.07	0.14	0.07	0.06
80	0.1	0.08	0.04	0.14	0.04	0.09	0.04	0.09	0.05	0.08	1	0.06	0.06	0.04	0.35	0.13	0.06	0.06	0.14	0.05	0.04
81	0.14	0.1	0.06	0.13	0.06	0.09	0.15	0.1	0.05	0.05	0.06	1	0.18	0.05	0.07	0.06	0.18	0.12	0.26	0.07	0.03
82	0.25	0.08	0.13	0.18	0.23	0.14	0.26	0.11	0.13	0.2	0.06	0.18	1	0.07	0.11	0.19	0.24	0.12	0.34	0.11	0.06
83	0.12	0.05	0.08	0.12	0.04	0.05	0.06	0.12	0.08	0.09	0.04	0.05	0.07	1	0.07	0.13	0.06	0.05	0.14	0.04	0.07
84	0.14	0.09	0.13	0.13	0.19	0.09	0.08	0.08	0.1	0.1	0.35	0.07	0.11	0.07	1	0.12	0.1	0.07	0.17	0.14	0.19
85	0.17	0.06	0.07	0.22	0.2	0.11	0.18	0.21	0.07	0.25	0.13	0.06	0.19	0.13	0.12	1	0.07	0.15	0.19	0.09	0.03
86	0.11	0.06	0.16	0.09	0.1	0.06	0.13	0.09	0.08	0.17	0.06	0.18	0.24	0.06	0.1	0.07	1	0.05	0.21	0.08	0.04
87	0.15	0.06	0.07	0.14	0.08	0.15	0.09	0.17	0.13	0.07	0.06	0.12	0.12	0.05	0.07	0.15	0.05	1	0.25	0.06	0.04
88	0.28	0.11	0.14	0.24	0.12	0.2	0.14	0.24	0.17	0.14	0.14	0.26	0.34	0.14	0.17	0.19	0.21	0.25	1	0.12	0.05
89	0.12	0.06	0.15	0.11	0.14	0.08	0.05	0.1	0.06	0.07	0.05	0.07	0.11	0.04	0.14	0.09	0.08	0.06	0.12	1	0.06
90	0.05	0.04	0.08	0.12	0.15	0.05	0.04	0.07	0.05	0.06	0.04	0.03	0.06	0.07	0.19	0.03	0.04	0.04	0.05	0.06	1

1	0.15	0.18	0.08	0.26	0.13	0.29	0.13	0.2	0.09	0.18
0.15	1	0.14	0.05	0.29	0.09	0.11	0.23	0.27	0.02	0.14
0.18	0.14	1	0.07	0.2	0.25	0.18	0.12	0.12	0.03	0.13
0.08	0.05	0.07	1	0.09	0.05	0.09	0.12	0.14	0	0.04
0.26	0.29	0.2	0.09	1	0.18	0.37	0.16	0.18	0.08	0.22
0.13	0.09	0.25	0.05	0.18	1	0.25	0.11	0.16	0.12	0.15
0.29	0.11	0.18	0.09	0.37	0.25	1	0.19	0.15	0.12	0.29
0.13	0.23	0.12	0.12	0.16	0.11	0.19	1	0.32	0.05	0.13
0.2	0.27	0.12	0.14	0.18	0.16	0.15	0.32	1	0.02	0.12
0.09	0.02	0.03	0	0.08	0.12	0.12	0.05	0.02	1	0.1
0.18	0.14	0.13	0.04	0.22	0.15	0.29	0.13	0.12	0.1	1