

**PENGAPLIKASIAN *TEXT MINING* DALAM  
PENENTUAN REKOMENDASI DOSEN PENGUJI  
SIDANG SKRIPSI**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat guna mencapai gelar  
Sarjana dalam bidang ilmu Teknik Industri

**Disusun oleh:**

**Nama : Sylviana Gunawan**

**NPM : 6131801063**



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI  
JURUSAN TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG  
2022**



**PENGAPLIKASIAN *TEXT MINING* DALAM  
PENENTUAN REKOMENDASI DOSEN PENGUJI  
SIDANG SKRIPSI**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat guna mencapai gelar  
Sarjana dalam bidang ilmu Teknik Industri

**Disusun oleh:**

**Nama : Sylviana Gunawan**

**NPM : 6131801063**



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI  
JURUSAN TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG  
2022**

**FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG**



Nama : Sylviana Gunawan  
NPM : 6131801063  
Program Studi : Sarjana Teknik Industri  
Judul Skripsi : PENGAPLIKASIAN *TEXT MINING* DALAM  
PENENTUAN REKOMENDASI DOSEN PENGUJI  
SIDANG SKRIPSI

**TANDA PERSETUJUAN SKRIPSI**

Bandung, 16 Februari 2022  
**Ketua Program Studi Sarjana  
Teknik Industri**

(Dr. Ceicalia Tesavrita, S.T., M.T.)

**Pembimbing Pertama**

**Dosen Pembimbing Kedua**

(Dedy Suryadi, S.T., M.S., Ph.D.)

(Dr. Ceicalia Tesavrita, S.T., M.T.)



## **PERNYATAAN TIDAK MENCONTEK ATAU MELAKUKAN PLAGIAT**

Saya yang bertanda tangan di bawah ini,

Nama : Sylviana Gunawan

NPM : 6131801063

dengan ini menyatakan bahwa Skripsi dengan Judul:

**PENGAPLIKASIAN *TEXT MINING* DALAM PENENTUAN REKOMENDASI  
DOSEN PENGUJI SIDANG SKRIPSI**

adalah hasil pekerjaan saya dan seluruh ide, pendapat atau materi dari sumber lain telah dikutip dengan cara penulisan referensi yang sesuai.

Pernyataan ini saya buat dengan sebenar-benarnya dan jika pernyataan ini tidak sesuai dengan kenyataan, maka saya bersedia menanggung sanksi yang akan dikenakan kepada saya.

Bandung, 10 Januari 2022

Sylviana Gunawan

NPM : 6131801063

## ABSTRAK

Skripsi merupakan sebuah karya tulis ilmiah yang akan diuji oleh dua dosen penguji yang ditentukan oleh Koordinator Skripsi. Umumnya, dosen penguji dipilih berdasarkan pusat studi dari dosen dan penelitian yang diuji. Namun, terdapat beberapa bidang ilmu yang berbeda dalam satu pusat studi yang sama. Tidak jarang, dosen yang dipilih berasal dari bidang ilmu yang berbeda. Ketidaksesuaian ini menyebabkan perbedaan dalam standar penilaian dan mempengaruhi pertanyaan yang diberikan. Terdapat juga situasi saat dosen meminta Koordinator Skripsi untuk memilih dosen penguji lain saat topik penelitian yang diangkat tidak sesuai dengan bidang ilmu dosen tersebut.

Idealnya, pemilihan dosen penguji dapat mempertimbangkan bidang ilmu dosen dan pengalaman dosen. Namun, Koordinator Skripsi tidak memiliki informasi mengenai pengalaman dosen dengan penelitian yang pernah dilakukan. Setiap tahunnya juga, terdapat ratusan skripsi yang diterima oleh administrasi program studi, sehingga jika pencatatan keterlibatan ini dilakukan secara manual, akan dibutuhkan waktu yang lama. Penelitian ini bertujuan untuk memberikan rekomendasi dosen berdasarkan topik penelitiannya untuk membantu Koordinator Skripsi menentukan dosen penguji dengan menggunakan metode *text mining*, khususnya TF-IDF dan *cosine similarity*. Kedua metode ini digunakan untuk menentukan kata penting dari sebuah abstrak skripsi dan mencocokkannya dengan abstrak skripsi yang sudah dimiliki. Dosen yang terlibat dalam abstrak ini kemudian direkomendasikan sebagai dosen penguji dari abstrak yang baru.

Dari penelitian ini, dapat disimpulkan bahwa metode TF-IDF dan *cosine similarity* dapat digunakan untuk membantu menentukan dosen penguji berdasarkan abstrak penelitian. Walaupun begitu, terdapat faktor lain yang menjadi pertimbangan saat menentukan dosen penguji, seperti jadwal dosen dan jumlah sidang dosen. Karenanya, hasil yang didapatkan memberikan rekomendasi dosen yang berpengalaman dalam topik penelitian yang diujikan dan Koordinator Skripsi dapat memilih dua dari dosen yang direkomendasikan. Selain rekomendasi dosen, TF-IDF juga dapat digunakan untuk membuat kata kunci dari topik penelitian yang sering melibatkan dosen.

## **ABSTRACT**

*Thesis is a research paper that will be tested by two examiners chosen by the Thesis Coordinator. Generally, examiners are selected based on the study area of the lecturers and the research being tested. However, there are several different fields of knowledge within one study area. It is not uncommon that the selected examiners come from different fields of knowledge. This dissimilarity causes differences in the standard of assessment and affects the questions given. There is also a situation when the examiner asks the Thesis Coordinator to choose another examiner when the research topic is deemed not within the lecturer's field of knowledge*

*Ideally, the selection of examiners can consider the lecturer's field of knowledge and the lecturer's experience. However, the Thesis Coordinator does not have information about the lecturer's experience with research that has been carried out. There are hundreds of theses that are received by the administration every year, thus, if the recording of this involvement is done manually, it will take a long time. This study aims to provide recommendations for lecturers based on their research topics to help the Thesis Coordinator determine the examiners by using text mining, especially TF-IDF and cosine similarity. These two methods are used to determine the important words of a thesis abstract and match it with owned abstracts from previous year. Lecturers involved in this known abstract will then be recommended as examiners of the new abstract.*

*From this research, it can be concluded that the TF-IDF and cosine similarity can be used to help determine the examiner based on the research abstract. However, there are other factors to be considered when determining the examiner, such as the examiner's schedule and the number of thesis defense each examiner has done. Therefore, the results obtained able to provide recommendations for examiners with experience in the research topic and the Thesis Coordinator can choose two of the recommended examiners. In addition to examiner recommendations, TF-IDF can also be used to create keywords from research topics that often involve each lecturer.*

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya yang senantiasa memberikan berkat bagi penulis untuk menyelesaikan skripsi ini. “Pengaplikasian *Text Mining* dalam Penentuan Rekomendasi Dosen Penguji Sidang Skripsi” disusun sebagai salah satu syarat untuk menyelesaikan Strata-1 dari program studi Teknik Industri Universitas Katolik Parahyangan.

Pada halaman ini, penulis ingin mengucapkan terima kasih sebesar-besarnya kepada kedua orang tua penulis yang terus mendukung penulis selama ini. Penulis juga ingin berterima kasih pada semua pihak yang telah membantu, menyemangati, dan mendukung penulis dalam penyelesaian skripsi ini:

1. Adik penulis yang selalu memberikan hiburan kepada penulis.
2. Bapak Dedy Suryadi, S.T., M.S., Ph.D. dan Ibu Dr. Ceicalia Tesavrita, S.T., M.T. selaku dosen pembimbing yang dengan sabar membimbing dan mendukung penulis selama laporan skripsi ini dibuat.
3. Bapak Dr. Sugih Sudharma Tjandra, S.T., M.Si. dan Bapak Romy Loice, S.T., M.T. selaku dosen penguji proposal skripsi yang telah memberikan masukan dan saran pengembangan skripsi.
4. Sahabat penulis, Lovita Christie, Mariska Gianni, dan Valentin Putri, yang telah menemani penulis sejak semester satu, untuk segala canda dan tawa selama ini.
5. Ivana Sanata dan Nadelya Hadi yang mendukung dan menyemangati penulis selama masa perkuliahan dan penyusunan skripsi.
6. Teman-teman Teknik Industri 2018, khususnya kelas A atas semua kenangan yang dimiliki penulis selama masa perkuliahan.
7. Pihak yang tidak dapat disebutkan satu per satu yang telah mendukung penulis selama ini.

Skripsi ini masih jauh dari kata sempurna. Penulis sadar masih banyak kekurangan dalam penyusunan skripsi ini. Karenanya, penulis terbuka untuk segala saran dan kritik untuk membantu penyempurnaan skripsi ini. Demikian



penyusunan dari skripsi ini. Semoga skripsi ini dapat memberikan manfaat bagi para pembaca.

Bandung, 10 Januari 2022

Penulis,

Sylviana Gunawan

## DAFTAR ISI

<b>ABSTRAK</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>ii</b>
<b>KATA PENGANTAR</b> .....	<b>iii</b>
<b>DAFTAR ISI</b> .....	<b>v</b>
<b>DAFTAR TABEL</b> .....	<b>vii</b>
<b>DAFTAR GAMBAR</b> .....	<b>ix</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>I-1</b>
I.1 Latar Belakang Masalah .....	I-1
I.2 Identifikasi dan Rumusan Masalah .....	I-2
I.3 Pembatasan Masalah dan Asumsi Penelitian .....	I-6
I.4 Tujuan Penelitian .....	I-6
I.5 Manfaat Penelitian .....	I-6
I.6 Metodologi Penelitian .....	I-7
I.7 Sistematika Penulisan .....	I-10
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>II-1</b>
II.1 Text Mining .....	II-1
II.2 Text Preprocessing.....	II-3
II.3 Term Frequency – Inverse Document Frequency (TF-IDF) .....	II-4
II.4 Cosine Similarity.....	II-5
<b>BAB III PENGUMPULAN DAN PENGOLAHAN DATA</b> .....	<b>III-1</b>
III.1 Pengumpulan Data .....	III-1
III.2 <i>Information Extraction</i> dari Abstrak .....	III-2
III.3 <i>Text Preprocessing</i> .....	III-3
III.4 TF-IDF .....	III-6
III.4.1 Perhitungan TF-IDF .....	III-7
III.4.2 Perhitungan <i>Cosine Similarity</i> .....	III-9
III.4.3 Penentuan Urutan Kemiripan Dokumen <i>Training</i> .....	III-10

III.5	Rekomendasi Dosen .....	III-11
III.5.1	Rekapitulasi Data Historis Dosen.....	III-11
III.5.2	Penentuan Dosen .....	III-12
III.5.3	Kata Penting Dosen .....	III-12
III.6	Hasil Rekomendasi Dosen.....	III-13
III.7	Panduan Penggunaan Model .....	III-24
III.7.1	Menggunakan Model untuk Abstrak Baru.....	III-25
III.7.2	Cara Menambahkan Abstrak Baru .....	III-27
<b>BAB IV ANALISIS .....</b>		<b>IV-1</b>
IV.1	Analisis Hasil .....	IV-1
IV.1.1	Analisis Dokumen <i>Testing 0</i> .....	IV-2
IV.1.2	Analisis Dokumen <i>Testing 2</i> .....	IV-3
IV.1.3	Analisis Dokumen <i>Testing 3</i> .....	IV-4
IV.1.4	Analisis Dokumen <i>Testing 4</i> .....	IV-5
IV.1.5	Analisis Dokumen <i>Testing 7</i> .....	IV-6
IV.1.6	Analisis Dokumen <i>Testing 10</i> .....	IV-7
IV.1.7	Analisis Dokumen <i>Testing 12</i> .....	IV-8
IV.1.8	Analisis Dokumen <i>Testing 14</i> .....	IV-10
IV.1.9	Analisis Dokumen <i>Testing 16</i> .....	IV-11
IV.2	Analisis Pemilihan Metode TF-IDF dan <i>Cosine Similarity</i> .....	IV-12
IV.3	Analisis Kelemahan Metode .....	IV-14
IV.4	Analisis <i>Stopwords</i> .....	IV-15
IV.5	Analisis Jumlah Dokumen.....	IV-16
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>V-1</b>
V.1	Kesimpulan.....	V-1
V.2	Saran.....	V-1
<b>DAFTAR PUSTAKA</b>		
<b>LAMPIRAN</b>		
<b>RIWAYAT HIDUP PENULIS</b>		

## DAFTAR TABEL

Tabel III.1 Perubahan Huruf Kapital.....	III-4
Tabel III.2 Penghapusan Tanda Baca.....	III-4
Tabel III.3 Penyisihan Angka .....	III-5
Tabel III.4 Penyisihan <i>Stopwords</i> .....	III-5
Tabel III.5 <i>Stemming</i> .....	III-6
Tabel III.6 Hasil TF-IDF.....	III-8
Tabel III.7 Hasil Penentuan Dokumen <i>Testing 0</i> .....	III-10
Tabel III.8 Rekapitulasi Dokumen Dosen PS .....	III-11
Tabel III.9 Nilai Dosen PS untuk Dokumen <i>Testing 0</i> .....	III-12
Tabel III.10 Nilai Kata 'Pasar'.....	III-13
Tabel III.11 Hasil Dokumen <i>Testing 0</i> .....	III-13
Tabel III.12 Hasil Dokumen <i>Testing 1</i> .....	III-14
Tabel III.13 Hasil Dokumen <i>Testing 2</i> .....	III-15
Tabel III.14 Hasil Dokumen <i>Testing 3</i> .....	III-15
Tabel III.15 Hasil Dokumen <i>Testing 4</i> .....	III-16
Tabel III.16 Hasil Dokumen <i>Testing 5</i> .....	III-16
Tabel III.17 Hasil Dokumen <i>Testing 6</i> .....	III-17
Tabel III.18 Hasil Dokumen <i>Testing 7</i> .....	III-18
Tabel III.19 Hasil Dokumen <i>Testing 8</i> .....	III-18
Tabel III.20 Hasil Dokumen <i>Testing 9</i> .....	III-19
Tabel III.21 Hasil Dokumen <i>Testing 10</i> .....	III-19
Tabel III.22 Hasil Dokumen <i>Testing 11</i> .....	III-20
Tabel III.23 Hasil Dokumen <i>Testing 12</i> .....	III-20
Tabel III.24 Hasil Dokumen <i>Testing 13</i> .....	III-21
Tabel III.25 Hasil Dokumen <i>Testing 14</i> .....	III-21
Tabel III.26 Hasil Dokumen <i>Testing 15</i> .....	III-22
Tabel III.27 Hasil Dokumen <i>Testing 16</i> .....	III-23
Tabel III.28 Hasil Dokumen <i>Testing 17</i> .....	III-23
Tabel III.29 Hasil Dokumen <i>Testing 18</i> .....	III-24
Tabel IV.1 Hasil 5 Dokumen untuk Dokumen <i>Testing 14</i> .....	IV-16



## DAFTAR GAMBAR

Gambar I.1 Metodologi Penelitian .....	I-7
Gambar II.1 <i>Information Retrieval</i> .....	II-2
Gambar II.2 <i>Information Extraction</i> .....	II-2
Gambar II.3 <i>Prediction</i> .....	II-3
Gambar III.1 Contoh Abstrak .....	III-2
Gambar III.2 Tempat Penyimpanan Data <i>Testing</i> .....	III-25
Gambar III.3 Hasil Mentah .....	III-26
Gambar III.4 Excel untuk Keterlibatan Dosen .....	III-27
Gambar III.5 Excel Keterlibatan Dosen Baru .....	III-28
Gambar III.6 Tempat Penyimpanan Data <i>Training</i> .....	III-28
Gambar III.7 Excel Rekapitulasi Keterlibatan .....	III-29
Gambar IV.1 Kemiripan Vektor .....	IV-13



## DAFTAR LAMPIRAN

LAMPIRAN A SINTAKS PYTHON .....	A-1
LAMPIRAN B REKAPITULASI NILAI <i>COSINE SIMILARITY</i> .....	B-1
LAMPIRAN C REKAPITULASI DOSEN .....	C-1
LAMPIRAN D HASIL PENENTUAN DOSEN .....	D-1
LAMPIRAN E KATA PENTING DOSEN .....	E-1





# **BAB I**

## **PENDAHULUAN**

Bab ini menjelaskan latar belakang serta identifikasi dari permasalahan yang diangkat pada penelitian. Batasan masalah serta asumsi penelitian juga dijelaskan lebih lanjut pada bab ini. Terakhir, bab ini ditutup dengan penjelasan lebih lanjut mengenai tujuan dari penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

### **I.1 Latar Belakang Masalah**

Sebelum dapat dianggap lulus dan menyanggah gelar sarjana, mahasiswa harus melakukan penelitian yang biasa disebut sebagai skripsi. Berdasarkan Panduan Skripsi Edisi Keempat dari program studi Teknik Industri Universitas Katolik Parahyangan, skripsi merupakan karya ilmiah yang dibuat untuk menunjukkan kemampuan mahasiswa dalam memanfaatkan pengetahuan yang didapatkan selama berkuliah. Dalam proses penyusunan skripsi, kemampuan untuk mengidentifikasi masalah yang ada, memecahkannya, dan menarik kesimpulan dari hasil yang didapatkan harus dimiliki oleh mahasiswa. Tujuan dari skripsi sendiri adalah mengasah kemampuan mahasiswa dalam menelaah dan menyelesaikan sebuah permasalahan dengan bimbingan Dosen Pembimbing (Petunjuk Pelaksanaan Penyusunan Skripsi (IND-500), 2015).

Dalam program studi Teknik Industri Universitas Katolik Parahyangan, skripsi biasanya dibuat dalam rentang waktu 4 sampai dengan 12 bulan di bawah bimbingan Dosen Pembimbing yang dipilih oleh mahasiswa. Mahasiswa kemudian harus menyelesaikan seminar skripsi sebelum dapat melakukan sidang skripsi. Berdasarkan dengan ketentuan program studi, hasil penelitian ini akan diuji oleh dua dosen penguji pada saat sidang skripsi dilaksanakan (Petunjuk Pelaksanaan Penyusunan Skripsi (IND-500), 2015). Dosen penguji memiliki peran yang penting dalam penelitian akhir mahasiswa. Dosen penguji yang tepat akan memberikan masukan penelitian yang konstruktif serta meningkatkan kualitas penelitian. Maka dari itu, idealnya, dosen penguji memang berpengalaman dalam topik penelitian yang diangkat.

Namun, untuk memilih dosen berdasarkan pengalamannya, maka dibutuhkan informasi mengenai pengalaman dosen tersebut, baik sebagai dosen pembimbing, maupun sebagai dosen penguji dari skripsi yang pernah dilakukan sebelumnya. Setiap tahunnya, terdapat puluhan, bahkan ratusan skripsi yang diterima oleh administrasi program studi. Jika keterlibatan masing-masing dosen dalam skripsi ini dikumpulkan dan dicatat secara manual, maka waktu yang diperlukan akan cukup lama. Maka dari itu, *text mining* dapat digunakan pada tahap ini untuk membantu mengumpulkan informasi mengenai keterlibatan dosen pada skripsi yang diterima.

*Text mining* mulai digunakan semenjak tren *Big Data* mulai dikenal oleh banyak orang. Seiring dengan akses internet yang semakin mudah dimiliki oleh banyak orang, istilah “Big Data” mulai terdengar di kalangan masyarakat umum. *Big Data* sendiri merujuk pada teknologi yang dapat mengolah data dengan kuantitas yang sangat besar dalam waktu singkat. Perusahaan-perusahaan dunia seperti Tesla dan Amazon contohnya, sudah menerapkan teknologi *Big Data* ini pada sistem dan produk yang mereka jual (Marr, 2021). *Text mining* sendiri merupakan sebuah metode pengolahan data yang digunakan untuk mengolah data berbentuk teks. Penggunaan metode ini dilandaskan dengan fakta bahwa bahasa yang digunakan oleh manusia untuk berkomunikasi dengan satu sama lain bukanlah hal yang mudah dimengerti oleh sebuah program komputer. Manusia dapat mengerti arti dari sebuah kata *slang* yang muncul dalam dokumen atau arti konseptual dari suatu pernyataan. Kemampuan ini bukanlah kemampuan yang dimiliki oleh sebuah program (Gupta & Lehal, 2009). Karenanya, dibutuhkan metode pengolahan data berbasis teks yang dapat memahami bahasa manusia untuk mengolah data ini.

## **I.2 Identifikasi dan Rumusan Masalah**

Untuk mengetahui dengan lebih lanjut mengenai proses pemilihan dosen penguji, wawancara dilakukan dengan dua dosen yang pernah menjadi Koordinator Skripsi. Saat ini, penentuan dosen penguji ditentukan oleh Koordinator Skripsi. Kriteria penentuan ini juga cukup subyektif. Secara garis besar, dosen dipilih dengan cara mempertimbangkan pusat studi dari penelitian dan jumlah sidang yang sudah dilakukan oleh dosen tersebut. Salah satu Koordinator Skripsi yang diwawancara menambahkan pertimbangan mengenai mata kuliah yang

diampu oleh dosen, *area of research* dosen, dan juga penelitian yang sering dibimbing oleh dosen tersebut. Hanya saja, cukup sulit untuk ikut mempertimbangkan ketiga faktor ini, dikarenakan tidak adanya catatan mengenai informasi tersebut maupun kata kunci dari topik penelitian untuk masing-masing dosen yang dimiliki oleh Koordinator Skripsi ini.

Walaupun salah satu Koordinator Skripsi menambahkan pertimbangan lain saat memilih dosen penguji, umumnya, setiap Koordinator Skripsi akan mempertimbangkan pusat studi dari dosen dan jumlah sidang dosen. Dengan proses penentuan seperti ini, penentuan dosen penguji tidak mempertimbangkan fakta bahwa terdapat beberapa bidang ilmu dari satu pusat studi yang sama. Contohnya, pusat studi ergonomi yang tidak hanya membahas tentang ergonomi produk serta keselamatan dan kesehatan kerja, namun juga aspek psikologi dari pekerja dan juga interaksi antara sistem komputer dan pengguna. Walaupun sama-sama menjadi dosen di pusat studi ergonomi, konsentrasi dari masing-masing dosen bisa berbeda. Proses penentuan dosen seperti sekarang menyebabkan terjadinya kasus di mana dosen penguji berasal dari pusat studi yang sama dengan topik penelitian, namun dengan konsentrasi yang berbeda. Saat situasi seperti ini terjadi, dosen dengan konsentrasi yang berbeda juga merasa kurang cocok untuk menjadi dosen penguji dari topik ini. Terdapat juga situasi saat dosen penguji tidak bersedia untuk menguji topik yang dirasa bukan merupakan bidangnya. Saat hal ini terjadi, Koordinator Skripsi harus mencari dosen pengganti untuk menguji penelitian tersebut.

Lalu terdapat situasi di mana dosen dari suatu pusat studi juga mengajar mata kuliah yang berasal dari pusat studi lainnya. Sebagai contoh, terdapat dosen dari pusat studi Ergonomi yang juga mengampu mata kuliah dari pusat studi *Sistem Enterprise*. Atau terdapat mata kuliah seperti Lingkungan dan Ekologi Industri yang diampu oleh beberapa dosen dari beberapa pusat studi yang berbeda. Tentunya, dosen yang berasal dari pusat studi lain namun mengampu mata kuliah ini akan memiliki pengertian yang lebih dalam terkait mata kuliah ini, dibandingkan dengan dosen yang tidak mengajar mata kuliah tersebut namun berasal dari pusat studi mata kuliah itu. Pengertian yang lebih dalam ini berarti dosen lebih mengerti isi dari penelitian dan dapat memberikan pertanyaan yang lebih dalam terkait topik yang diangkat. Dari wawancara yang dilakukan, tidak jarang salah satu dosen penguji yang terpilih bukan merupakan dosen yang sering

terlibat dalam topik penelitian tersebut. Hal ini menyebabkan adanya perbedaan standar pada saat penilaian skripsi dilakukan. Perbedaan jauh pada nilai yang diberikan oleh kedua penguji kerap terjadi. Perbedaan nilai ini menunjukkan perbedaan standar dan kriteria dari kedua dosen penguji terhadap penelitian yang sama (Top 5 Tips on How to Choose Your PhD Examiner, n.d.). Selain itu, pengalaman dosen juga akan mempengaruhi pertanyaan yang diberikan pada saat sidang dilakukan. Sebuah survei yang dilakukan oleh Zhang di tahun 2019 menunjukkan sebanyak 88% responden setuju bahwa pengalaman pribadi penguji mempengaruhi keputusan dari pertanyaan yang dianggap penting untuk dikemukakan pada saat evaluasi berlangsung. Karenanya, pengalaman memiliki peranan yang penting terkait penilaian serta kualitas dari pertanyaan yang akan diberikan oleh penguji terhadap penelitian yang telah dilakukan.

Untuk dapat menentukan dosen penguji berdasarkan pengalaman yang dimiliki dosen, Koordinator Skripsi harus memiliki informasi yang dapat menunjukkan hubungan antara dosen dan topik penelitian yang pernah dibimbing atau diuji. Untuk saat ini, informasi ini tidak tersedia. Karenanya, sulit untuk menjadikan pengalaman dosen penguji dengan topik penelitian tertentu sebagai salah satu pertimbangan dalam pengambilan keputusan.

Pembuatan alat bantu yang memberikan hubungan antara dosen dengan penelitian yang pernah melibatkannya dapat dibuat untuk membantu penentuan dosen penguji. Untuk memastikan alat bantu ini memberikan keputusan berdasarkan pengalaman dosen dengan topik penelitian, maka data historis dijadikan sebagai faktor utama yang dipertimbangkan dalam pengambilan keputusan. Data historis ini nantinya akan berisikan topik penelitian dan dosen penguji serta dosen pembimbing dari penelitian tersebut.

Cara termudah untuk mengetahui topik dari suatu penelitian adalah membaca rangkuman dari penelitian tersebut. Abstrak merupakan salah satu bagian yang ada di dalam skripsi. Merujuk pada Panduan Skripsi Edisi Keempat (2015) yang dikeluarkan oleh program studi Teknik Industri Universitas Katolik Parahyangan, sebuah abstrak akan memuat informasi terkait latar belakang, metodologi penelitian, serta hasil yang didapatkan dari penelitian. Maka dari itu, abstrak dapat memberikan gambaran yang cukup terkait penelitian yang telah dilakukan tanpa perlu membaca keseluruhan skripsi.

Mengingat abstrak dibuat dalam bentuk teks, maka metode pengolahan data yang digunakan untuk membangun panduan harus dapat mengolah data yang berbentuk teks. Metode pengolahan data *text mining* merupakan metode yang digunakan untuk menggali data yang diekspresikan dalam bentuk teks (Gupta & Lehal, 2009). Pada penelitian ini, *text mining* dimanfaatkan untuk menggali data dan membantu pengambilan keputusan penentuan dosen penguji.

Dalam pembuatan alat bantu untuk menentukan dosen penguji, metode *text mining* dapat dimanfaatkan untuk mengolah abstrak dan menghasilkan *output* berupa informasi dari isi abstrak yang dapat digunakan sebagai data historis untuk menentukan dosen. Dengan begitu, esensi dari abstrak dapat dengan mudah didapatkan dan disimpan sebagai data historis. Pengekstraksian informasi ini dapat dilakukan dengan berbagai metode, salah satunya adalah *Term Frequency – Inverse Document Frequency* (TF-IDF). Metode ini digunakan untuk mengetahui kata yang penting dari sebuah dokumen berdasarkan frekuensi kemunculan kata dan bobot kata (Ayari, 2020). TF-IDF merupakan salah satu metode yang paling sederhana dan paling mudah diimplementasikan. Metode ini juga merupakan metode yang digunakan dalam sistem pencarian pada *website* Elasticsearch dan memang sering digunakan (Nishimura, 2020). Pada percobaan yang dilakukan oleh Nishimura (2020) dan penelitian yang dilakukan oleh Wang, Zhou, Jin, Liu, dan Lu (2017) metode ini juga memiliki tingkat efektivitas yang lebih tinggi saat dibandingkan dengan metode yang lebih kompleks. Namun, TF-IDF hanya merupakan metode pembobotan kata. Diperlukan metode tambahan untuk menilai kemiripan dokumen yang diolah dengan dokumen historis yang dimiliki. Terdapat berbagai metode yang dapat digunakan untuk mengukur kemiripan antar dokumen. Namun, berdasarkan penelitian yang telah dilakukan oleh Thompson, Panchev, dan Oakes (2015) mengenai hasil dari beberapa metode pengukuran kemiripan, *cosine similarity* memberikan hasil yang paling memuaskan. Maka dari itu, metode yang digunakan untuk pada penelitian ini adalah metode *cosine similarity*. Berdasarkan identifikasi masalah yang ada, berikut merupakan rumusan dari masalah yang akan menjadi fokus dari penelitian ini:

1. Bagaimana penerapan metode *text mining* dalam penentuan dosen penguji sidang skripsi?
2. Bagaimana penerapan metode *text mining* untuk memberikan kata kunci dari topik-topik penelitian yang sering melibatkan dosen?

### **I.3 Pembatasan Masalah dan Asumsi Penelitian**

Terdapat beberapa batasan serta asumsi yang ditetapkan pada penelitian ini. Batasan yang ada ditetapkan untuk memastikan penelitian yang dilakukan terfokus pada masalah utama dari penelitian. Berikut merupakan batasan dari penelitian ini:

1. Penelitian hanya dilakukan untuk penentuan dosen penguji program studi Teknik Industri Universitas Katolik Parahyangan.
2. Data historis yang digunakan hanya abstrak skripsi yang diterima oleh administrasi program studi Teknik Industri.
3. Model yang dibuat dalam penelitian hanya mempertimbangkan faktor keterlibatan dosen pada data historis yang digunakan.

Setelah batasan dari penelitian ini ditentukan, asumsi juga turut ditetapkan untuk mempermudah proses penelitian. Pada penelitian ini, diasumsikan tidak terdapat perubahan pada staf pengajar yang memenuhi syarat untuk menjadi dosen penguji. Asumsi ini mencakup tidak adanya staf pengajar baru yang dapat menjadi dosen penguji dan tidak mempertimbangkan jika ada dosen yang sedang mengambil cuti.

### **I.4 Tujuan Penelitian**

Berdasarkan latar belakang serta identifikasi dan perumusan masalah, berikut merupakan tujuan dari penelitian ini:

1. Menerapkan metode *text mining* dalam pembuatan rekomendasi dosen untuk membantu penentuan dosen penguji skripsi.
2. Menerapkan metode *text mining* dalam pembuatan kata kunci dari topik penelitian yang sering melibatkan masing-masing dosen

### **I.5 Manfaat Penelitian**

Penelitian ini diharapkan dapat berguna bagi pihak yang terkait, terutama bagi pihak yang menjadi pemilik dari permasalahan yang diangkat pada penelitian ini. Berdasarkan tujuan dari penelitian ini, berikut merupakan manfaat dari penelitian ini bagi koordinator skripsi:

1. Koordinator Skripsi memiliki informasi yang dapat membantu pengambilan keputusan terkait dosen penguji.

3. Pemilihan dosen penguji dapat dilakukan dengan mempertimbangkan pengalaman yang dimiliki dosen.

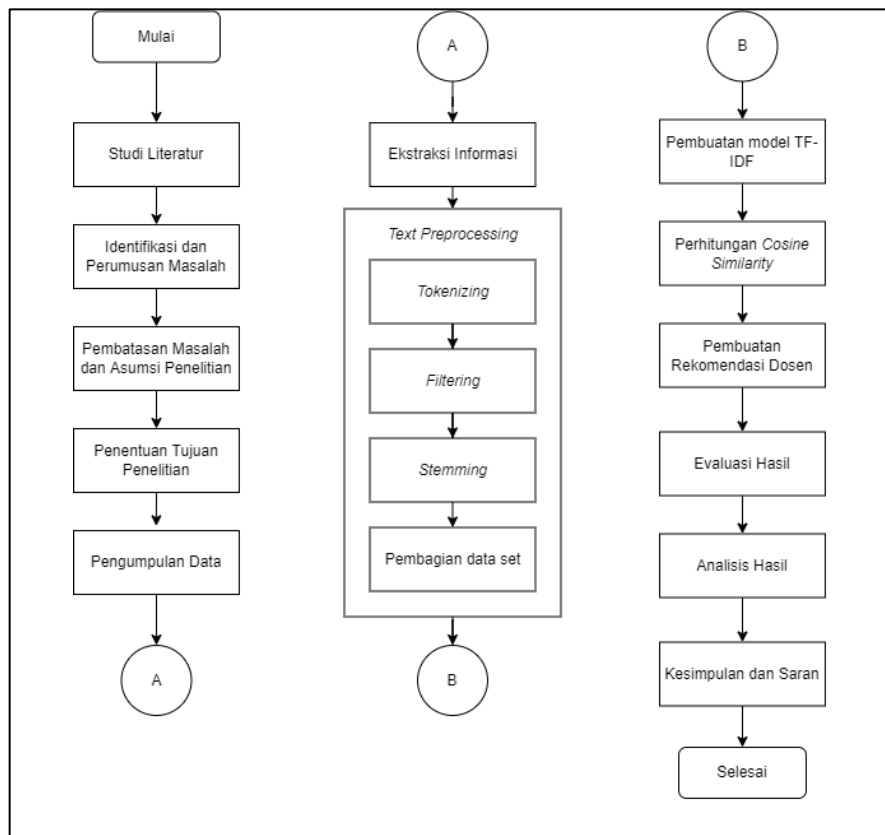
Selain memberikan manfaat bagi koordinator skripsi, terdapat beberapa manfaat yang dapat diterima oleh pembaca. Berikut beberapa manfaat dari penelitian ini bagi pembaca:

1. Pembaca dapat menambah pengetahuan terkait penggunaan *text mining* dalam penentuan dosen penguji.
2. Pembaca dapat menjadikan hasil penelitian sebagai referensi bagi penelitian selanjutnya.
3. Pembaca dapat menggunakan metode yang digunakan pada penelitian ini untuk pengambilan keputusan lainnya.

### I.6 Metodologi Penelitian

Metodologi penelitian ini dibuat sebagai panduan untuk memastikan proses penelitian berjalan sesuai dengan prosedur yang sudah direncanakan.

*Flowchart* dari metodologi penelitian ini ditunjukkan pada Gambar I.1.



Gambar I.1 Metodologi Penelitian



*Flowchart* dari metodologi penelitian menunjukkan alur dari proses penelitian. Berikut merupakan penjelasan dari setiap tahapan yang dilakukan selama penelitian ini dilakukan:

1. Studi Literatur

Studi literatur dilakukan untuk mendapatkan pemahaman serta referensi terkait penelitian serupa yang telah dilakukan sebelumnya. Informasi yang didapatkan dari tahap ini berasal dari buku, jurnal, serta artikel yang dapat ditemukan secara *online*.

2. Identifikasi dan Perumusan Masalah

Pada tahap ini, identifikasi terkait pemilihan dosen penguji secara manual oleh Koordinator Skripsi dan kurangnya panduan yang bisa digunakan untuk membantu mengambil keputusan ditetapkan sebagai obyek penelitian. Selain itu, identifikasi metode pengolahan data *text mining* yang dapat mengolah data dalam bentuk teks juga dilakukan. Rumusan masalah yang ingin dijawab oleh penelitian ini adalah bagaimana data historis yang ada dapat dimanfaatkan sebagai panduan dalam pengambilan keputusan terkait penentuan dosen penguji.

3. Pembatasan Masalah dan Asumsi Penelitian

Batas serta asumsi yang digunakan pada penelitian ini ditetapkan pada tahap ini. Baik batas maupun asumsi ditetapkan guna memastikan lingkup penelitian tidak terlalu luas. Pada penelitian ini, pemilihan dosen penguji hanya terbatas pada pengujian skripsi di program studi Teknik Industri Universitas Katolik Parahyangan dan data historis yang digunakan, berupa abstrak dan dosen pembimbing serta dosen pengujinya, hanyalah abstrak yang diterima oleh administrasi program studi, yang pada kasus ini, hanya menyimpan abstrak dari penelitian pada semester Ganjil dan Genap 2020-2021. Penelitian ini dilakukan dengan asumsi tidak ada perubahan terhadap dosen yang memenuhi syarat untuk menjadi dosen penguji.

4. Penentuan Tujuan Penelitian

Tujuan dari penelitian yang dilakukan ditetapkan pada tahap ini. Penelitian dilakukan untuk menerapkan metode pengolahan data *text mining* untuk memanfaatkan data historis terkait abstrak dan dosen yang terlibat sebagai panduan dalam penentuan dosen penguji ke depannya.

5. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data historis abstrak skripsi dan dosen pembimbing serta penguji dari skripsi terkait. Pengumpulan data dilakukan dengan cara meminta dokumen kepada Koordinator Skripsi tahun 2021 dan kepada administrasi program studi Teknik Industri Universitas Katolik Parahyangan. Hubungan antara abstrak dan dosen pembimbing serta penguji direkapitulasi secara manual.

6. *Text Preprocessing*

Proses pengolahan data dimulai dengan melakukan *text preprocessing*. *Text preprocessing* merupakan tahap yang dilakukan untuk memastikan data yang akan digunakan sudah layak. Tahap ini dimulai dengan mengubah seluruh huruf kapital menjadi huruf kecil. Setelah itu, proses *filtering* dilakukan dengan membuang *term* yang tidak memiliki arti signifikan seperti angka, tanda baca, dan kata umum dari kumpulan *term* yang sudah dimiliki. Setelah itu, proses *stemming* kemudian dilakukan untuk mencari kata akar dari kata hasil *filtering*. Proses ini dilakukan dengan cara menghapus imbuhan awal dan akhir dari kata yang ada. Hasil kata yang didapatkan dari proses ini baru dapat digunakan sebagai *input* untuk pengekstraksian informasi. Kumpulan abstrak yang didapatkan akan dibagi dalam dua bagian, data *training* dan data *testing*. Selama proses pemrograman dilakukan, data *training* dijadikan data *input* bagi program dan data *testing* digunakan untuk menilai performansi program yang telah dibuat. *Text preprocessing* ini dilakukan dengan menggunakan bahasa pemrograman Python dan *package* Sastrawi.

7. Perancangan Model

Pada tahap ini, perancangan model untuk mengekstraksi informasi akan dilakukan. Ekstraksi informasi ini dilakukan untuk mendapatkan informasi penting terkait topik yang diangkat pada data abstrak yang sedang diolah. Penilaian kemiripan dokumen dengan TF-IDF dan *cosine similarity* dirancang pada tahap ini. Perancangan model dibuat dengan menggunakan *package* sklearn. Pada tahap ini pembuatan kata kunci dosen juga dilakukan.

8. Pengujian Program

Pada tahap ini, uji coba terhadap model dilakukan dengan data *testing*. Tahap ini memastikan jika model sudah sesuai dan memiliki proses penentuan dengan mempertimbangkan data historis dosen. Hasil yang didapatkan dibandingkan dengan kenyataan yang sesungguhnya.

9. Analisis Hasil

Tahap analisis hasil ini dilakukan untuk menganalisa hasil rekomendasi dosen yang didapatkan. Analisis terkait pemilihan metode yang digunakan juga dilakukan.

10. Kesimpulan dan Saran

Pada tahap terakhir ini, kesimpulan dari penelitian dibuat berdasarkan rumusan masalah dan dari hasil penelitian. Saran terkait penelitian selanjutnya juga dibuat pada tahap ini.

### **I.7 Sistematika Penulisan**

Sistematika penulisan memberikan gambaran mengenai sistematika penulisan dari penelitian yang dilakukan. Penelitian ini terbagi ke dalam lima bab sebagai berikut.

## **BAB I PENDAHULUAN**

Bab I terdiri dari 7 subbab, yaitu, latar belakang dari permasalahan, identifikasi dan rumusan masalah, pembatasan masalah dan asumsi penelitian, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan dari penelitian terkait pengaplikasian *text mining* dalam menentukan dosen penguji sidang skripsi.

## **BAB II TINJAUAN PUSTAKA**

Bab II menjelaskan lebih lanjut terkait studi literatur yang dapat mendukung pelaksanaan penelitian. Studi literatur ini digunakan sebagai dasar dalam pengolahan data dan pembuatan solusi dari masalah yang diangkat di penelitian. Pada penelitian ini, studi literatur dilakukan untuk metode pengolahan data *text mining*, proses *text preprocessing*, TF-IDF, dan pengertian terkait *cosine similarity*.

### BAB III PENGUMPULAN DAN PENGOLAHAN DATA

Bab III berisikan proses yang dilakukan selama tahap pengumpulan dan pengolahan data dilakukan. Pada bab ini, seluruh proses yang dilakukan, yang mencakup proses pengumpulan data, ekstraksi informasi, *text preprocessing*, perhitungan nilai TF-IDF dan *cosine similarity*, rekapitulasi data dosen dan pembuatan rekomendasi dosen, dijabarkan dengan rinci.

### BAB IV ANALISIS

Bab IV menjabarkan dengan lebih rinci analisis dari hasil yang didapatkan pada bab sebelumnya. Selain itu, bab ini juga memberikan penjelasan mengenai pemilihan TF-IDF dan *cosine similarity*, kelemahan dari metode yang dipilih, dan diperlukannya penambahan kata *stopwords*.

### BAB V KESIMPULAN DAN SARAN

Bab V terdiri dari kesimpulan yang dapat diambil dari penelitian ini. Terakhir, terdapat juga saran yang dapat menjadi bahan pertimbangan untuk penelitian selanjutnya.