



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PENGELOMPOKAN PENELITIAN KE DALAM SUSTAINABLE DEVELOPMENT GOALS (SDGs) BERDASARKAN KATA KUNCI MENGGUNAKAN METODE TF-IDF

JASMINE NAYLA ABDUL AZIZ



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI LAPORAN AKHIR DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pengelompokan Penelitian ke dalam *Sustainable Development Goals* (SDGs) Berdasarkan Kata Kunci Menggunakan Metode TF-IDF” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir laporan akhir ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Maret 2025

Jasmine Nayla Abdul Aziz
G6401201090

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

JASMINE NAYLA ABDUL AZIZ. Pengelompokan Penelitian ke dalam *Sustainable Development Goals* (SDGs) Berdasarkan Kata Kunci Menggunakan Metode TF-IDF. Dibimbing oleh KARLISA PRIANDANA.

SIPPM dikembangkan oleh IPB untuk menyimpan data penelitian, termasuk informasi seperti judul, abstrak, dan klasifikasi ke SDG. Pada awalnya, klasifikasi SDG tidak diwajibkan, yang menyebabkan banyak penelitian tidak memiliki kategori SDG. Klasifikasi ini menjadi penting untuk memenuhi persyaratan hibah internasional yang sering memerlukan rekam jejak penelitian terkait tujuan pembangunan berkelanjutan. Penelitian ini bertujuan untuk memetakan abstrak penelitian di Institut Pertanian Bogor (IPB) terhadap *Sustainable Development Goals* (SDGs) menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Data yang digunakan mencakup abstrak penelitian dari SIPPM IPB antara tahun 2019 hingga 2022. Proses penelitian dimulai dengan praproses data yang melibatkan penerjemahan, tokenisasi, penghapusan kata-kata umum, dan *stemming*. Setelah itu, nilai TF-IDF dihitung dan digunakan untuk memetakan abstrak ke SDG yang sesuai. Kemudian, untuk penelitian yang telah memiliki data SDG sebelumnya berdasarkan input dari peneliti, hasil pemetaan ini kemudian dibandingkan dengan data SDG yang telah memiliki data SDG tersebut. Hasil penelitian menunjukkan bahwa proporsi abstrak yang sesuai dengan SDG yang telah ditentukan sebelumnya lebih banyak dibandingkan dengan abstrak yang berbeda, yaitu sebanyak 60% abstrak tergolong sama, sementara 26,3% abstrak berbeda dengan klasifikasi yang ditentukan sebelumnya. Dengan menggunakan metode ini, penelitian yang sebelumnya tidak memiliki klasifikasi SDG kini dapat dipetakan ke SDG yang sesuai. Berdasarkan temuan ini, metode TF-IDF dapat digunakan sebagai alat bantu yang efektif dalam pemetaan SDG, namun validasi manual tetap diperlukan untuk memastikan akurasi yang lebih tinggi.

Kata Kunci: Pengelompokan, SIPPM IPB, SDGs, *Text mining*, TF-IDF.

ABSTRACT

JASMINE NAYLA ABDUL AZIZ. Classification of Research into Various Sustainable Development Goals (SDGs) Types Based on Keywords Using the TF-IDF Method. Supervised by KARLISA PRIANDANA.

SIPPM was developed by IPB to store research data, including information such as titles, abstracts, and SDG classifications. Initially, SDG classification was not mandatory, which resulted in many research papers lacking SDG categories. This classification has become important to meet the requirements of international grants, which often require a research track record related to Sustainable Development Goals. This study aims to map the research abstracts at the Bogor Agricultural University (IPB) to the Sustainable Development Goals (SDGs) using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The data used consists of research abstracts from SIPPM IPB between 2019 and 2022. The

research process begins with data preprocessing, which includes translation, tokenization, removal of common words, and stemming. After that, the TF-IDF values are calculated and used to map the abstracts to the appropriate SDGs. Then, for research that already has SDG data based on input from researchers, the mapping results are compared with the existing SDG data. The results of the study indicate that the proportion of abstracts that match the SDGs previously assigned is higher than those that do not, with 60% of abstracts classified as the same, while 26.3% of abstracts differ from the previously assigned classification. By using this method, research that previously lacked SDG classification can now be mapped to the appropriate SDGs. Based on these findings, the TF-IDF method can be used as an effective tool for SDG mapping, although manual validation is still necessary to ensure higher accuracy.

Keywords: Categorization, SIPPM IPB, SDGs, Text mining, TF-IDF

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



PENGELOMPOKAN PENELITIAN KE DALAM SUSTAINABLE DEVELOPMENT GOALS (SDGs) BERDASARKAN KATA KUNCI MENGGUNAKAN METODE TF-IDF

JASMINE NAYLA ABDUL AZIZ

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

IPB University

Tim Penguji pada Ujian Skripsi
1. Dr. Sony Hartono Wijaya S.Kom., M.Kom.
2. Ahmad Ridha S.Kom., M.S.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Pengelompokan Penelitian ke dalam *Sustainable Development Goals* (SDGs) Berdasarkan Kata Kunci Menggunakan Metode TF-IDF

Nama : Jasmine Nayla Abdul Aziz
NIM : G6401201090

Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr. Karlisa Priandana, S.T., M.Eng
19851121 201212 2 002

Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom.
19810809 200812 1002

Tanggal Ujian:
18 Februari 2025

Tanggal Lulus:

(tanggal penandatanganan oleh Dekan Sekolah
Vokasi)

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan November 2023 sampai bulan Januari 2025 ini ialah *text mining*, dengan judul “Pengelompokan Penelitian ke dalam *Sustainable Development Goals* (SDGs) Berdasarkan Kata Kunci Menggunakan Metode TF-IDF”.

Terima kasih penulis ucapkan kepada pembimbing, Ibu Dr. Karlisa Priandana, S.T., M.Eng. yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada pembimbing akademik, moderator seminar, dan penguji luar komisi pembimbing. Ungkapan terima kasih juga disampaikan kepada ayah, ibu, kakak, serta seluruh keluarga yang telah memberikan dukungan, doa, dan kasih sayangnya. Ungkapan terima kasih juga penulis sampaikan untuk seluruh teman-teman SMA, SMP, Etherone 57, serta semua orang-orang baik yang telah menemani selama kuliah di IPB dan penulisan skripsi ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Januari 2025

Jasmine Nayla Abdul Aziz



DAFTAR ISI

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	ix
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	3
METODE	4
2.1 Mengunduh Data	5
2.2 Mengunduh <i>Keywords</i> SDG	5
2.3 Praproses Data	6
2.4 <i>Text mining</i> menggunakan TF-IDF	7
2.5 Pemberian Label	8
2.6 Analisis	9
III HASIL DAN PEMBAHASAN	10
3.1 Praproses data	10
3.1.1 Praproses Abstrak Penelitian	10
3.1.2 Praproses <i>Keywords</i> SDG	10
3.1.3 Keterbatasan Praproses	11
3.2 Menghitung TF-IDF	11
3.3 Pemberian Label	12
3.4 Analisis	13
3.5 Interpretasi Hasil	13
IV SIMPULAN DAN SARAN	15
4.1 Simpulan	15
4.2 Saran	15
DAFTAR PUSTAKA	16
LAMPIRAN	18
RIWAYAT HIDUP	26

DAFTAR TABEL

1	Contoh hasil praproses data abstrak	10
2	Contoh hasil praproses <i>keywords</i> SDG	10
3	Contoh hasil perhitungan IDF	11
4	Contoh hasil perhitungan TF-IDF Abstrak	12
5	Contoh hasil perhitungan TF-IDF <i>keywords</i>	12
6	Contoh hasil <i>top 3</i> SDG abstrak	13

DAFTAR GAMBAR

1	Diagram tahapan penelitian	4
2	Hasil klasifikasi SDG data penelitian di SIPPM	14

DAFTAR LAMPIRAN

1	<i>Keywords</i> SDG	19
2	Hasil klasifikasi abstrak penelitian	25

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.