

**PENERAPAN TF-IDF PADA KLASIFIKASI KATEGORI
BERITA PADA *GOOGLE TRENDS RSS FEEDS***

SKRIPSI

OLEH:

FRENKLIN CORNELIUS 188160011



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 26/3/24

Access From (repository.uma.ac.id)26/3/24

**PENERAPAN TF-IDF PADA KLASIFIKASI KATEGORI
BERITA PADA *GOOGLE TRENDS RSS FEEDS***

PROPOSAL SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana di Fakultas

Teknik

Universitas Medan Area

OLEH :

FRENKLIN CORNELIUS 188160011

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

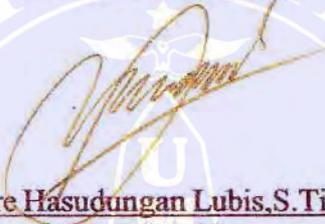
Document Accepted 26/3/24

Access From (repository.uma.ac.id)26/3/24

HALAMAN PENGESAHAN

Judul Skripsi : Penerapan TF-IDF pada Klasifikasi Kategori Berita
Pada *Google Trends RSS Feeds*
Nama : Frenklin Cornelius
NPM : 188160011
Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing


Andre Hasudungan Lubis, S.Ti, MSc
Pembimbing


Dr. Eng. Supriatno, ST., MT.
Dekan Fakultas Teknik


Rizki Mulsone, S.Kom, M. Kom
Ka. Prodi Teknik Informatika

Tanggal Lulus : 21 Desember 2023

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 21 Desember 2023

Penulis,



Frenklin Cornelius
NPM 188160011

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

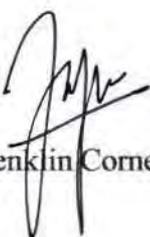
Nama : Frenklin Cornelius
NPM : 188160011
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty- Free Right)** atas karya ilmiah saya yang berjudul :

Penerapan TF-IDF pada Klasifikasi Kategori Berita Pada *Google Trends RSS Feeds*

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir/skripsi/tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan
Pada tanggal : 21 Desember 2023
Yang menyatakan


(Frenklin Cornelius)

ABSTRAK

Google Trends adalah alat yang memberikan wawasan tentang minat pengguna berdasarkan kata kunci populer dari waktu ke waktu. Namun, berita dari RSS *Google Trends* belum diklasifikasikan ke dalam kategori tertentu, membuatnya sulit digunakan untuk analisis seperti filter, analisis sentimen, atau rekomendasi. Klasifikasi kategori berita dengan menganalisis data yang terkandung pada teks dari RSS *Google Trends* dapat menjadi solusi untuk memudahkan dalam proses – proses lainnya seperti rekomendasi, seleksi, dan analisis lainnya. Sehingga perlu dilakukan suatu penelitian untuk melakukan klasifikasi berita – berita yang diperoleh dari RSS *Google Trends*. Penelitian ini mengusulkan penggunaan metode TF-IDF, yang memungkinkan untuk mengkategorikan berita berdasarkan kontennya. Metode TF-IDF adalah cara untuk mengukur sejauh mana kata kunci tertentu penting dalam sebuah dokumen. Dengan menerapkan ini pada teks dari RSS *Google Trends*, pengelompokan berita ke dalam kategori seperti politik, ekonomi, olahraga, dan teknologi dapat dilakukan. Proses klasifikasi kemudian dapat dilakukan terhadap *feed google trends* RSS dengan menghitung skor TF-IDF pada *feed Google Trends* menggunakan *word / term* yang telah ditentukan kemudian mencari kategori berita dengan kemiripan skor TF-IDF tertinggi sebagai kategori terpilih hasil dari klasifikasi. Adapun hasil penelitian yang diperoleh dari penelitian yang dilakukan menunjukkan tingkat akurasi klasifikasi sebesar 76% yang mana merupakan angka yang cukup baik.

Kata Kunci: Algoritma TF-IDF, *Google Trend* RSS *Feed*, Klasifikasi, Berita.

ABSTRACT

Google Trends is kind of tool that provides insight into user interests based on popular keywords over time. However, news from Google Trends RSS has not been classified into specific categories yet, it drives some difficulties to performing analysis such as filters, sentiment analysis, or recommendations. Classifying news categories by analyzing the data contained in the text from RSS Google Trends can be a solution to facilitate other processes these issues. Hence, a research is important to be conduct to classify news obtained from RSS Google Trends. This study proposes the use of the TF-IDF method, which allows us to categorize news based on their content. The TF- IDF method is a means to measure the extent to which certain keywords are important in a document. By applying this to text from RSS Google Trends, the news are able to divided into categories such as politics, economics, sports and technology. The classification process is performed through calculating the TF-IDF score on the Google Trendsfeed using the words / terms that have been determined later. The research results obtained from the research carried out show a classification accuracy rate of 76%, which is categorized as a fairly good model.

Keywords: *TF-IDF Algorithm, Google Trend RSS Feed, Classification, News.*



RIWAYAT HIDUP

Penulis merupakan putra atau anak ke-1 (satu) lahirkan di Medan, Sumatera Utara, pada tanggal 05 Maret 1996.

Pada tahun 2015 Penulis lulus dari SMA ST. Maria Medan, lalu pada tahun 2018 terdaftar sebagai mahasiswa Fakultas Teknik Prodi Teknik Informatika Universitas Medan Area. Pada tahun 2021 penulis melaksanakan Kerja Praktek (KP) di Dinas Kependudukan dan Pencatatan Sipil Kabupaten Labuhanbatu Utara.



KATA PENGANTAR

Puji syukur kita panjatkan kepada Tuhan yang Maha Esa, karena telah memberikan kesempatan kepada kita untuk melakukan seminar hasil pada hari ini yang berjudul “Penerapan TF-IDF pada Klasifikasi Kategori Berita pada Google Trend RSS Feed”. Skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Sastra- 1 (S.Kom) pada jurusan Teknik Indormatika. Saya ining menyampaikan terima kasih kepada:

1. Tuhan Yang Maha Esa, karena berkat-Nya penulis dapat menyelesaikan Skripsi ini dengan baik.
2. Orang Tua yang mendoakan serta mendukung penulis dalam pengerjaan skripsi.
3. Bapak Prof. Dr. Dadan Ramadan, M.Eng, M.Sc selaku Rektor Universitas Medan Area.
4. Bapak Dr. Eng. Supriatno, ST., MT selaku Dekan Fakultas Teknik Universitas Medan Area.
5. Bapak Rizki Muliono, S.kom, M.Kom selaku Ketua Program Studi Teknik Informatika.
6. Bapak Andre Hasudungan Lubis, S.Ti, Msc selaku Dosen Pembimbing yang telah memberikan bimbingan, motivasi dan arahan kepada penulis sehingga skripsi ini bisa selesai dengan baik.
7. Seluruh teman yang telah membantu dalam pengerjaan skripsi.

Saya menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu saran dan kritik yang membangun sangat saya harapkan untuk perbaikan di masa yang mendatang.

Medan, 21 Desember 2023

Frenklin Cornelius
NIM: 188160011



DAFTAR ISI

	Halaman
ABSTRAK	v
ABSTRACT	vi
RIWAYAT HIDUP	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah.....	2
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	3
1.5. Batasan Masalah	3
1.6. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
2.1. Google Trends	6
2.2. RSS Feeds.....	9
2.2.1. <i>Feeder Readers</i>	10
2.3. Text Mining	10
2.4. Algoritma TF-IDF.....	11
2.5. <i>Confusion Matrix</i>	13
2.6. <i>Flowchart</i>	14
2.7. UML (<i>Unified Modeling Language</i>).....	16
2.7.1. Use Case Diagram	16
2.7.2. <i>Class Diagram</i>	19
2.7.3. <i>Activity Diagram</i>	19
2.7.4. <i>Sequence Diagram</i>	20
2.8. Penelitian terdahulu	21
BAB III ANALISIS DAN PERANCANGAN.....	23
3.1. Analisis Sistem Yang Berjalan	23
3.2. Analisis Sistem Yang Diusulkan	23
3.3. Analisis Kebutuhan Sistem.....	24
3.4. Metode Pengembangan Sistem	24
3.5. Tahapan Penelitian.....	25
3.6. Analisis Proses Klasifikasi TF-IDF	29
3.6.1. Dataset	29
3.6.2. Perhitungan Skor TF-IDF	30
3.7. Perancangan Sistem	34
3.7.1. Flowchart Sistem	34
3.7.2. <i>Use Case Diagram</i>	37
3.7.3. <i>Entity Relational Diagram</i>	37
3.7.4. Struktur Table	38

3.7.5.	Perancangan Antarmuka.....	39
BAB IV	HASIL DAN PEMBAHASAN	42
4.1.	Hasil Pembangunan Aplikasi.....	42
4.1.1.	Halaman Umum.....	42
4.1.2.	Halaman Klasifikasi Bebas.....	43
4.1.3.	Halaman Klasifikasi Google.....	44
4.1.4.	Halaman Login Admin.....	46
4.1.5.	Halaman Data Berita	47
4.1.6.	Halaman Analisis TF-IDF	48
4.1.7.	Halaman Manajemen User	49
4.2.	Analisa Model TF-IDF	49
4.3.	Pengujian dan Pembahasan Klasifikasi.....	60
4.4.	Pengujian Hasil Klasifikasi.....	72
4.5.	Evaluasi Hasil	74
BAB V	KESIMPULAN DAN SARAN.....	76
5.1.	Kesimpulan	76
5.2.	Saran	76
DAFTAR PUSTAKA	77
LAMPIRAN	79

DAFTAR TABEL

Tabel 2. 1 Tabel Confusion Matrix	13
Tabel 2. 2 Simbol-Simbol Flowchart	15
Tabel 2. 3 Penelitian Terdahulu	21
Tabel 3. 1 Kategori Berita	29
Tabel 3. 2 Dataset Berita.....	30
Tabel 3. 3 Rancangan Tabel User.....	38
Tabel 3. 4 Rancangan Tabel Kategori	38
Tabel 3. 5 Rancangan Tabel Corpus Berita	38
Tabel 3. 6 Rancangan Tabel Corpus Term	39
Tabel 4. 1 Frekuensi Term	53
Tabel 4. 2 Tabel Kedalaman TF-IDF	57
Tabel 4. 3 Seleksi Term dengan skor Tertinggi.....	58
Tabel 4. 4 Dataset Latih	60
Tabel 4. 5 Hasil Klasifikas Data Uji.....	72
Tabel 4. 6 Hasil Confussion matrix.....	75

DAFTAR GAMBAR

Gambar 2. 1 Google Trends	8
Gambar 2. 2 Actor/Role	16
Gambar 2. 3 Use Case	17
Gambar 2. 4 Use Case	17
Gambar 2. 5 Use Case	17
Gambar 2. 6 Use Case	18
Gambar 2. 7 Use Case	18
Gambar 2. 8 Use Case	18
Gambar 3. 1 Tahapan Penelitian	25
Gambar 3. 2 Flowchart Pembentukan Model Klasifikasi.....	31
Gambar 3. 3 Flowchart Halaman Login.....	35
Gambar 3. 4 Flowchart Halaman Utama.....	35
Gambar 3. 5 Flowchart Halaman Dataset	36
Gambar 3. 6 Flowchart Halaman Klasifikasi	36
Gambar 3. 7 Use Case Diagram	37
Gambar 3. 8 Entity Relationship Diagram	37
Gambar 3. 9 Rancangan Halaman Login.....	39
Gambar 3. 10 Rancangan Halaman Utama	40
Gambar 3. 11 Rancangan Halaman Dataset	40
Gambar 3. 12 Dialog Tambah Data	41
Gambar 3. 13 Rancangan Halaman Klasifikasi.....	41
Gambar 4. 1 Halaman Umum.....	42
Gambar 4. 2 Halaman Klasifikasi Bebas.....	43
Gambar 4. 3 Hasil Klasifikasi Bebas	44
Gambar 4. 4 Halaman Klasifikasi Google.....	44
Gambar 4. 5 Hasil pengambilan google trend rss feed.....	45
Gambar 4. 6 Hasil Klasifikasi Google.....	46
Gambar 4. 7 Halaman Login Admin	47
Gambar 4. 8 Halaman Data Berita	47
Gambar 4. 9 Halaman Analisa TF-IDF	48
Gambar 4. 10 Halaman Manajemen User	49

Gambar 4. 11 Hasil Analisa TF-IDF Sistem Pada Kategori Olahraga.....	59
Gambar 4. 12 Detail Analisa TF-IDF	59
Gambar 4. 13 Pengambilan Google XML Feed	70
Gambar 4. 14 Data uji	71
Gambar 4. 15 Hasil klasifikasi data uji.....	71



BAB I

PENDAHULUAN

1.1. Latar Belakang

Berita *Online* atau *Online News* saat ini telah menjadi salah satu sumber utama dalam penyampaian berita di masyarakat. Website berita dan beberapa sumber alternatif berita lainnya seperti media sosial, email dan sumber – sumber lainnya telah menambahkan alternatif kepada masyarakat dalam memperoleh berita sehingga merubah orientasi dalam distribusi dan perolehan berita di masyarakat (Bentley, Quehl, Wirfs-Brock, & Bica, 2019). *Google Trends* merupakan salah satu layanan yang menyediakan informasi statistik ketertarikan atau kepentingan pengguna berdasarkan *terms* atau kata pencarian populer dari waktu ke waktu (Mavragani, Ochoa, & Tsagarakis, 2018). Selain menyediakan informasi statistik pencarian terbaru, *Google Trends* juga menyediakan layanan *Really Simple Syndication feed* (RSS *feed*) yang merupakan layanan informasi terbaru dalam format XML yang berisikan berita yang berkaitan dengan tren yang sedang berlangsung.

RSS *feed* yang disediakan oleh *Google Trends* cukup bermanfaat bagi para pencari berita umum maupun bagi para peneliti. Dengan informasi berita terkait dengan pencarian yang sedang populer maka dapat diperoleh informasi seperti apa yang sedang dicari oleh para pengguna internet pada umumnya. Namun, *feed* atau update terbaru dari berita yang disediakan oleh RSS *Google Trends* belum diklasifikasikan kedalam kategori berita tertentu sehingga konten yang disediakan oleh RSS *Google Trends* tersebut tidak dapat langsung digunakan pada beberapa kebutuhan seperti *filtering*, *sentiment analysis*, *recommendation system*, dan

kebutuhan – kebutuhan lainnya (Duong & Hoang, 2019).

Pengelompokan atau klasifikasi berita kedalam beberapa kategori dapat dilakukan dengan beberapa pendekatan yang salah satunya adalah menggunakan teknik *text classification* menggunakan metode TF-IDF. *Term Frequency – Inverse Document Frequency* merupakan sebuah model statistik yang menggambarkan relevansi kata kunci tertentu pada dokumen sehingga dokumen – dokumen tersebut dapat di-identifikasi dan dikategorikan kedalam kategori atau kelas tertentu yang telah diaplikasikan oleh Qaiser dan Ali dalam menganalisis konten pada domain dan website (Qaiser & Ali, 2018). Penelitian terdahulu yang mengaplikasikan analisis teks menggunakan TF-IDF dapat dilihat pada penelitian yang dilakukan oleh Kandhro et al yang melakukan klasifikasi berita dari dokumen berita *headline* dengan bahasa Shindi (Kandhro, et al., 2019). Pada penelitian tersebut, corpus dibagi menjadi 30% *data testing* dan 70% *data training* dan mampu memberikan tingkat akurasi sampai 84%.

Klasifikasi kategori berita dengan menganalisis data yang terkandung pada teks dari RSS *Google Trends* dapat menjadi solusi untuk memudahkan dalam proses – proses lainnya seperti rekomendasi, seleksi, dan analisis lainnya. Sehingga perlu dilakukan suatu penelitian untuk melakukan klasifikasi berita – berita yang diperoleh dari RSS *Google Trends*.

1.2. Perumusan Masalah

Berdasarkan latar belakang tersebut maka dapat dirumuskan masalah yaitu bagaimana cara mengklasifikasikan *feed* berita dari RSS *Google Trends* kedalam kategori berita menggunakan TF-IDF?

1.3. Tujuan Penelitian

Adapun beberapa tujuan penelitian ini sebagai berikut:

1. Untuk menerapkan *Text Classification* pada *feed* berita *RSS Google Trends* menggunakan TF-IDF.
2. Dapat mengkategorikan secara otomatis tajuk berita yang disediakan oleh *RSS Google Trends*.
3. Menganalisis akurasi klasifikasi pengkategorian berita pada *RSS Google Trends* menggunakan *Confussion Matrix*.

1.4. Manfaat Penelitian

Adapun beberapa manfaat dari penelitian ini adalah sebagai berikut :

1. Memperoleh model klasifikasi yang dapat digunakan dalam klasifikasi berita khususnya pada *RSS Google Trends*.
2. Memberikan sumbangsih sistem aplikasi pengkategorian berita *RSS Google Trends* secara *real-time* sehingga dapat digunakan sebagai sumber data pada penelitian – penelitian lainnya.

1.5. Batasan Masalah

Adapun batasan – batasan yang ditetapkan pada penelitian ini agar jangkauan pembahasan dari penelitian tidak terlalu melebar adalah sebagai berikut:

1. Metode utama yang digunakan pada penelitian ini adalah metode TF-IDF
2. Kelas atau kategori berita yang digunakan pada penelitian ini adalah Politik, Olahraga, Ekonomi dan Teknologi
3. Jumlah data yang akan digunakan adalah sebanyak 240 data berita

yang terbagi kedalam empat kategori berita yang mana konten berita diperoleh dari berbagai portal berita online seperti CNN, detiknews, Kompas dan portal berita lainnya.

4. Proses pelatihan dan pengujian akan membagi data kedalam data latih dan data uji dengan proporsi 70% data latih dan 30% data uji.
5. Pengujian akurasi klasifikasi dilakukan dengan melakukan klasifikasi terhadap data uji yang kemudian kebenaran dan kesalahan klasifikasi akan dimasukkan kedalam *confussion matrix* untuk memperoleh nilai akurasi model klasifikasi yang dibuat pada penelitian ini.

1.6. Sistematika Penulisan

Untuk memberikan gambaran mengenai tulisan ini, secara singkat skripsi ini memiliki sistematika penulisan yang diuraikan sebagai berikut :

BAB I : PENDAHULUAN

Bab ini membahas tentang latar belakang masalah, identifikasi masalah, batasan masalah, maksud dan tujuan serta sistematika penulisan skripsi ini.

BAB II : LANDASAN TEORI

Bab ini berisikan teori yang berhubungan dengan pembahasan yang diangkat pada skripsi ini.

BAB III : ANALISIS DAN PERANCANGAN SISTEM

Pada bab ini akan dibahas mengenai metode penelitian yang digunakan serta algoritma dari metode yang digunakan yaitu metode *byte pair encoding* serta membahas tentang bagaimana perancangan objek-objek yang dibutuhkan dalam pembangunan aplikasi.

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Bab ini akan membahas tentang penjelasan tiap-tiap fungsi serta unit yang dimiliki oleh aplikasi serta membahas validasi hasil output yang dihasilkan oleh aplikasi.

BAB V : KESIMPULAN DAN SARAN

Bab ini akan menjabarkan kesimpulan dari proses pembangunan aplikasi sampai proses validasi hasil output aplikasi serta menjawab masalah-masalah yang telah dirumuskan pada bab pendahuluan.



BAB II

TINJAUAN PUSTAKA

2.1. Google Trends

Pertumbuhan penggunaan internet di Indonesia berkembang sangat pesat. Akibatnya, semakin banyak orang yang menggunakan mesin pencari Internet. Mesin pencari yang paling banyak digunakan adalah pencarian *Google*. Pencarian *Google* sering digunakan untuk mencari berbagai informasi terkait dengan menggunakan kata kunci. *Google Trends* adalah aplikasi Big Data terkemuka yang paling banyak digunakan oleh bisnis untuk memahami dan memprediksi serta dapat membantu memprediksi perubahan sosial (Satriawati, Prasetyo, & Irawati, 2022). Intensitas pencarian dengan kata kunci yang berbeda tersebut menghasilkan data dalam jumlah besar (big data) yang kemudian direkam oleh *Google*. *Google* kemudian menyebarkan intensitas pencarian ini melalui *Google Trends* (Ayuningtyas & Wirawati, 2020).

Google Trends dapat menyediakan laporan harian, mingguan, dan bulanan mengenai volume kata kunci yang terkait dengan berbagai informasi yang relevan. Volume kata kunci tersebut disajikan dalam bentuk indeks yang memiliki nilai antara 0-100. Informasi indeks *Google Trends* harus berkorelasi dengan aktivitas saat ini untuk membantu memprediksi rilis data berikutnya. Hal yang harus diperhatikan yaitu indeks *Google Trends* tidak berpotensi untuk memprediksi masa depan, namun sangat membantu dalam memprediksi masa kini. Suatu prediksi untuk informasi di masa kini disebut nowcasting. Karena mesin pencari *Google* adalah yang paling populer di dunia, tidak mengherankan jika *Google Trends* memberikan tren kata kunci yang cukup akurat berdasarkan peristiwa terkini.

Diperkenalkan pertama kali pada tahun 2006 dengan versi terbaru yang dirilis pada tahun 2018, *Google Trends* membawa beberapa fitur yang sangat penting, yaitu menampilkan apa yang sedang hangat dibicarakan, kemudian mengidentifikasi topik dan subtopik yang sedang dibahas, serta mencari tren pencarian geografis di wilayah pengguna.

Google Trends bisa dibilang satu-satunya alat gratis yang dapat digunakan oleh pengguna yang mencari ide kata kunci untuk konten situs web dengan penyajian data yang komprehensif. Adapun manfaat dari penggunaan *google trend* adalah sebagai berikut:

1. Melihat popularitas suatu topik dalam waktu tertentu.
2. Target pengguna. *Google Trends* dapat memberikan informasi pengguna yang berkaitan dengan *trend* yang sedang viral.
3. Referensi topik atau konten, dengan informasi trend konten atau topik yang sedang trend dapat membantu untuk membuat konten yang dapat meningkatkan trafik pengunjung.
4. Menemukan peluang pemasaran produk, informasi trend yang sedang berlaku dapat membantu para penjual dalam menentukan strategi pemasaran produknya.
5. Sumber ide konten. Dengan *Google Trends*, dapat diperoleh referensi popularitas kata kunci pilihan sehingga dapat membantu menemukan ide konten yang bagus dan mengarahkan lalu lintas pengunjung.
6. Bisa mendapatkan kata kunci yang tepat, dengan menggunakan *Google Trends*, Anda dapat menemukan kata kunci yang sesuai dengan tren yang Anda cari.

7. Pembaruan merek, *Google Trends* dapat memperbarui merek (mis. pasar selalu menjadi tren dan tren bersifat sementara dan masa depan). Jika Anda ingin memperbarui merek Anda, *Google Trends* akan membantu Anda menemukan kata kunci yang tepat dan tepat. Kata kunci yang membuat merek Anda tidak berakhir buruk, tetapi mengenai tempat yang tepat.

Tentang cara mencari kata kunci trending *Google* sebagai berikut : Untuk menemukan kata kunci yang terkait dengan topik yang Anda cari, Anda dapat menuliskannya di kolom "Masukkan istilah atau topik pencarian" atau "Tambahkan istilah pencarian".



Gambar 2. 1 Google Trends
Sumber: (Yasin, 2019)

Misalnya saat memasukkan kata kunci “Cloud”, *Google Trends* akan menampilkan informasi yang cukup lengkap tentang intensitas pengguna *Google Search* menggunakan kata “Cloud”.

Tabel pertama menampilkan grafik tren penelusuran pengguna menggunakan kata "Cloud" selama setahun terakhir. Grafik penggunaan data berdasarkan negara berdasarkan lokasi Anda saat ini. Namun Anda dapat menyesuaikan cakupan perayapan dengan kebutuhan Anda.

Angka yang ditampilkan oleh pencarian *Google* (0-100) bukanlah jumlah pengguna yang mencari menggunakan kata kunci tersebut. Ini adalah perkiraan jumlah pengguna yang mencari menggunakan kata kunci "*Cloud*".

2.2. RSS Feeds

RSS merupakan kepanjangan dari *Rich Site Summary* atau *Really Simple Syndication* *RSS feed* (terkadang disebut *XML feed*) adalah daftar berita atau artikel dari situs web yang dapat dilanggan oleh pengguna (Pertiwi & Irwansyah, 2020). Umpan RSS dapat dibaca menggunakan pembaca umpan atau perangkat lunak agregator. RSS telah lama digunakan dalam layanan informasi berbasis web. Banyak digunakan oleh blog, situs web, atau portal untuk memberikan informasi dalam bentuk standar.

Adapun keuntungan RSS adalah sebagai berikut:

1. Opt-In: pengunjung dapat memilih feed atau konten yang di-inginkan.
2. Syndication: mekanisme standart dalam pengumpulan berita.
3. Aggregation: pengguna dapat dengan cepat mencari beberapa sumber berita yang berbeda.

Awalnya, RSS hanya digunakan untuk menyampaikan headline dari situs berita.

Seiring popularitas RSS tumbuh, RSS sekarang banyak digunakan untuk :

1. Blog: ringkasan dari posting atau artikel blog.
2. Newsletters: *headline* berita – berita terkini.
3. Weather Alerts: notifikasi atau informasi cuaca.
4. Press Announcement: pengumuman – pengumuman *press*.
5. Real Estate: informasi properti.
6. Schools: informasi sekolah dan akademis.

2.2.1. Feeder Readers

Feed reader disebut juga aggregator merupakan program yang dapat digunakan untuk membaca RSS *feed* (Pratama & Pratama, 2022). Pembaca Umpan dapat berupa aplikasi desktop, aplikasi web, atau komponen bawaan dari program klien email. Di dalam *feed reader* biasanya terdapat embedded HTML *viewer* (*browser*). Setiap pembaca umpan memiliki antarmuka yang berbeda. Namun semuanya memiliki komponen dasar, yaitu: Daftar aliran yang di cantumkan, daftar judul *feed* yang dipilih dan seluruh isi artikel/judul dipilih.

Adapun beberapa contoh RSS *Readers*:

1. FeedReader (*Windows*) merupakan aplikasi pembaca RSS feed gratis yang berjalan di sistem operasi *Windows* dan didistribusikan pada 24 April tahun 2001. Aplikasi ini dapat diakses di laman resminya di url <http://www.feedreader.com/>.
2. SharpReader (*Windows*) merupakan aplikasi pembaca RSS feed yang berjalan di sistem operasi *windows* dan versi terakhir adalah versi 0.9.7.0 yang dirilis pada 2 Agustus 2006. Aplikasi ini dapat diakses di laman resminya di url <https://www.sharpreader.net/>.
3. NewsFire (*Mac*) merupakan aplikasi pembaca RSS khusus berita yang dikembangkan oleh David Watanabe untuk sistem operasi Mac OS X. Aplikasi ini dapat diakses pada laman resminya di url <https://newsfirex.com/>.

2.3. Text Mining

Text Mining dapat didefinisikan secara luas sebagai proses berbasis data di mana kumpulan data atau dokumen dari waktu ke waktu diolah menggunakan berbagai alat analisis. (Findawati & Rosid, 2020). *Text mining* bertujuan untuk mengekstraksi informasi yang berguna dari sumber data dengan mengidentifikasi dan memeriksa pola yang menarik. Namun, dalam kasus penambangan teks,

sumber datanya adalah kumpulan dokumen, dan pola yang menarik tidak ditemukan di antara basis data formal tetapi dalam data teks kumpulan dokumen yang tidak terstruktur. Dengan demikian, sistem *text mining* dan *data mining* memiliki banyak kesamaan arsitektur pada tingkatan yang lebih tinggi. Misalnya, kedua jenis sistem bergantung pada rutinitas pra-pemrosesan, algoritme pencocokan pola, dan elemen lapisan presentasi seperti alat B. Visualisasi yang meningkatkan pencarian rangkaian respons.

Text Mining sendiri memiliki beberapa tipe antara lain (Fauziyyah, 2020):

1. *Search and Information Retrieval*. Mencari dan ekstraksi informasi dari kumpulan data atau dokumen.
2. *Document Clustering*. Pengelompokan dokumen berdasarkan kesamaan atau ketidaksamaan konten.
3. *Document Classification*. Pengkategorian dokumen menggunakan label tertentu.
4. *Web Mining Data dan text mining*. Ekstraksi informasi yang terkandung pada konten halaman web atau konten teks.
5. *Information Extraction*. Ekstraksi informasi tertentu dari kumpulan data atau dokumen.
6. *Natural Language Processing* Pemrosesan bahasa alami yang dapat ekstraksi dari data.
7. *Concept Extraction*. Ekstraksi informasi dalam bentuk konseptual yang terkandung pada kumpulan data atau dokumen.

2.4. Algoritma TF-IDF

Term Frequency Inverse Document Frequency (TF-IDF) adalah metode untuk menentukan seberapa sering sebuah kata (ekspresi) diasosiasikan dengan sebuah dokumen dengan memberikan bobot pada setiap kata (Herwijayanti, Ratnawati, & Muflikhah, 2018). TF-IDF memberikan informasi nilai frekuensi

suatu kata atau *term* yang terdapat pada dokumen atau artikel tertentu terhadap nilai frekuensi di semua dokumen. Perhitungan ini dapat digunakan untuk mencari seberapa penting sebuah kata pada sebuah data atau dokumen. Prosedur penerapan TF-IDF sedikit berbeda antar aplikasi, tetapi pendekatannya sebagian besar sama. TF (*Term Frequency*) adalah frekuensi kata dalam dokumen. TF dapat dihitung menggunakan persamaan berikut.

$$TF(t) = \frac{f_{t,d}}{\sum t, d} \quad (2.1)$$

dimana $f_{t,d}$ merupakan nilai frekuensi sebuah kata atau *term* (t) yang muncul di dalam dokumen d , sedangkan $\sum t, d$ merupakan total frekuensi semua kata yang terdapat di dalam dokumen d .

Inverse Document Frequency adalah metode untuk menghitung distribusi term dalam dokumen (Suryani & Edy, 2020). Distribusi *term* yang tidak merata mempengaruhi optimalisasi perhitungan bobot dokumen. Fungsi *Inverse Document Frequency* adalah untuk mengurangi penyebaran istilah yang terlalu bias. Jika nilai Document Density (DF) di bawah ambang batas, istilah tidak dihitung. Dari sini dapat disimpulkan bahwa semakin rendah nilai Document Density (DF), semakin tinggi nilai Inverse Document Density (IDF) (Rozi & Sulistyawati, 2019). Menghitung nilai Inverse Document Frequency (IDF) dari sebuah kata dalam frase dokumen dapat menggunakan Persamaan (2.2) :

$$IDF(t) = \log \frac{|D|}{f_{t,d}} \quad (2.2)$$

Dimana $|D|$ merupakan jumlah total dokumen yang sedang diproses, sedangkan

$f_{t,D}$ merupakan jumlah dokumen dimana t muncul di dalam D . Dalam koleksi dokumen D , sebuah kata atau *term* t dan dokumen individu $d \in D$, dapat dihitung nilai TF-IDF menggunakan rumus (2.3) berikut:

$$\text{TF-IDF}(t) = \text{TF}(t) * \text{IDF}(t) \quad (2.3)$$

Jika diasumsikan $|D| \sim f_{t,d}$, maka ukuran dari kumpulan dokumen hampir sama dengan frekuensi kata t di dalam dokumen D .

2.5. Confusion Matrix

Confusion Matrix merupakan metode evaluasi hasil yang diperoleh selama implementasi algoritma klasifikasi menggunakan tabel (Grandis & Arumsari, 2021). Adapun tabel evaluasi dari *confusion matrix* dapat dilihat pada Tabel 2.1 berikut.

Tabel 2. 1 Tabel *Confusion Matrix*

Kelas	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

TP (True Positive) merupakan jumlah dokumen positif yang diprediksi atau diklasifikasikan benar sebagai kelas yang benar, sedangkan FP (False Positif) merupakan jumlah klasifikasi benar sebagai kelas positif. FN (False Negative) adalah jumlah klasifikasi salah dengan kelas negatif, sedangkan TN (True Negative) merupakan jumlah klasifikasi benar dengan kelas negatif.

Nilai dalam tabel diatas kemudian dapat digunakan untuk menghitung *Precision*, *Recall*, dan *Accuracy*. Akurasi adalah hasil membandingkan nilai prediksi yang benar dengan nilai sebenarnya.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2.4)$$

Precision adalah hasil dari membandingkan nilai prediksi sebenarnya dari data positif dan nilai total data positif.

$$Precision = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2.5)$$

Recall adalah hasil dari membandingkan nilai prediksi yang benar dari data positif dengan total nilai prediksi yang benar.

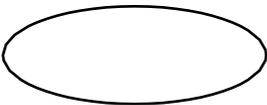
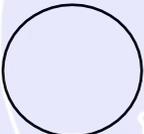
$$Recall = \frac{TP}{(TP + FN)} \quad (2.6)$$

2.6. Flowchart

Diagram alir atau *flowchart* adalah diagram yang digunakan untuk memvisualisasikan urutan atau tahapan proses dan hubungan antara satu proses dengan proses lainnya dengan menggunakan simbol-simbol tertentu. Menurut para ahli, pengertian flowchart adalah sebagai berikut: Flowchart adalah suatu metode yang menggambarkan langkah-langkah penyelesaian masalah atau solusi dengan menyajikan simbol-simbol tertentu sehingga dapat memudahkan dalam pemahaman urutan proses penyelesaian masalah.

Bagan Alir Program merupakan diagram yang digunakan untuk menggambarkan prosedur sistem (Hasan, Septiningrum, Chaery, Abdurachman, & Prawirayudha, 2020). Bagan alir atau flowchart adalah representasi grafis dari suatu sistem yang menggambarkan hubungan fisik antara unit-unit intinya (Kusnady & Siregar, 2018).

Tabel 2. 2 Simbol-Simbol *Flowchart*

No.	Simbol	Fungsi
1.		Simbol terminal yang digunakan untuk menunjukkan awal dan akhir dari urutan proses.
2.		Simbol proses digunakan untuk menggambarkan sebuah proses atau instruksi.
3.		Simbol input atau output yang digunakan untuk ilustrasi input atau output dari proses.
4.		Simbol keputusan yang digunakan untuk memeriksa sebuah kondisi tertentu untuk menentukan arah proses.
5.		Simbol connector, digunakan untuk menghubungkan alur proses masih pada halaman yang sama.
6.		Simbol Document, merupakan simbol untuk ilustrasi data atau dokumen.
7.		Simbol Off-page-connector, digunakan untuk menghubungkan alur proses pada halaman yang berbeda.
8.		Simbol untuk output, digunakan untuk interaksi ke perangkat seperti printer, plotter, monitor dll.

Sumber: (Rosaly & Prasetyo, 2019)

2.7. UML (*Unified Modeling Language*)

Unified Modeling Language (UML) adalah bahasa yang digunakan dalam pemodelan sistem. UML ditemukan oleh Grady Booch, Ivar Jacobson dan James Rumbaugh. Versi UML terbaru, khususnya versi 2.5, mencakup lima belas diagram (Indriyani, Yunita, & Muthia, 2019). *Unified Modeling Language* dapat dibagi menjadi dua jenis, yaitu *structure diagram* dan *behavior diagram*. *Structure diagram* menggambarkan data dan hubungan antar data atau komponen yang terdapat di dalam sebuah sistem. *Behavior diagram* menggambarkan hubungan atau interaksi antar objek yang terdapat pada sebuah sistem.

Ada banyak jenis diagram di UML, tetapi secara umum terdapat empat jenis UML yang banyak digunakan yaitu *use case diagram*, *class diagram*, *activity diagram* dan *sequence diagram*. Berikut penjelasan dari masing - masing diagram tersebut :

2.7.1. Use Case Diagram

Use case diagram merupakan diagram yang digunakan untuk menggambarkan komponen dan hubungannya terhadap actor dan komponen lainnya yang terdapat pada sebuah sistem yang mendeskripsikan bisnis sistem (Indriyani, Yunita, & Muthia, 2019). Berikut adalah elemen-elemen dari *use case diagram*:

1. Aktor

Actor/role adalah orang atau sistem lain yang berinteraksi dengan sistem saat ini.



Gambar 2. 2 Actor/Role

Sumber: (Indriyani, Yunita, & Muthia, 2019)

2. *Use Case*

Use case adalah case atau fungsional sistem dalam sebuah sistem yang dapat berupa *extend* (memperluas) use case lainnya. *Use Case* berada di dalam *system boundary* (batasan sistem) yang diberikan label kata kerja dari fungsionalitas sistem.

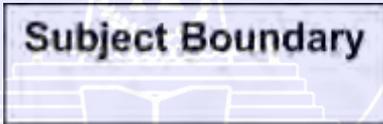


Gambar 2.3 *Use Case*

Sumber: (Indriyani, Yunita, & Muthia, 2019)

3. *Subject Boundery*

Berisi *nama* dari sistem yang diletakkan di dalam atau di bagian atas boundary yang digunakan sebagai ruang lingkup sistem dimana *use case* berada pada komponen ini.



Subject Boundary

Gambar 2.4 *Use Case*

Sumber: (Indriyani, Yunita, & Muthia, 2019)

4. *Assosiation Relationship*

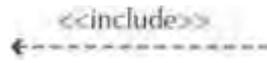
Digunakan untuk menghubungkan komponen seperti actor atau case ke case lainnya. Menunjukkan komunikasi dua arah (Menunjukkan komunikasi satu arah jika menggunakan tanda panah). Tanda * untuk keragaman dari asosiasi (multiplicity of the association).

Gambar 2.5 *Use Case*

Sumber: (Indriyani, Yunita, & Muthia, 2019)

5. Include Relationship

Digunakan untuk menunjukkan relasi satu use case merupakan bagian dari use case lainnya. Sebuah *use case* dapat terlaksana jika *use case* lain terpenuhi Tanda panah mengarah dari *use case* awal menuju ke *use case* yang di-include.

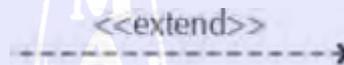


Gambar 2. 6 Use Case

Sumber: (Indriyani, Yunita, & Muthia, 2019)

6. Extend Relationship

Digunakan untuk menggambarkan pengembangan sebuah *use case* untuk memasukkan perilaku opsional. Tanda panah mengarah dari use case tambahan ke *use case* awal (pusat).



Gambar 2. 7 Use Case

Sumber: (Indriyani, Yunita, & Muthia, 2019)

7. . Generalization Relationship

Digunakan untuk menggambarkan relasi khusus *use case* tertentu ke use case yang lebih umum. Tanda panah mengarah dari *use case* khusus (specialized) ke *use case* yang lebih umum.



Gambar 2. 8 Use Case

Sumber: (Indriyani, Yunita, & Muthia, 2019)

2.7.2. *Class Diagram*

Class Diagram merupakan diagram yang menggambarkan hubungan antar class yang terdiri dari nama class, atribut dan method (Setiaji & Sastra, 2021). Dalam memodelkan *class* yang terdapat pada sebuah sistem, termasuk diantaranya atribut, method, dan hubungan antara *class*, UML menyediakan model *class diagram*. Yang menyediakan perspektif view statis atau struktural atas suatu sistem.

Elemen utama suatu class diagram adalah kotak-kotak yang merepresentasikan kelas-kelas. Pada bagian setiap kotak dibagi kedalam bagian-bagian horizontal. Bagian atas sendiri berisi label dengan nama *class*. Pada bagian tengah berisi atribut dari class. Setiap atribut dapat memiliki nama, jenis dan tingkat visibilitas. Visibilitas ditunjukkan dengan awalan -, #, ~ atau +, yang masing-masingnya menunjukkan visibilitas private, protected, package atau public.

2.7.3. *Activity Diagram*

Activity Diagram merupakan diagram yang digunakan untuk mengilustrasikan urutan proses atau aktivitas dan hubungan di antara kegiatan dalam suatu proses (Indriyani, Yunita, & Muthia, 2019). *Activity diagram* menggambarkan urutan aktifitas sistem dari sudut pandang pengguna. Komponen utama dari *activity diagram* adalah *action node* (node aksi), yang direpresentasikan dengan persegi panjang dengan sudut bulat, yang berhubungan dengan tugas yang dilakukan oleh sistem software.

Anak panah menghubungkan sebuah node aksi ke node aksi lainnya mendeskripsikan aliran control dari aktifitas sistem. Dengan artian bahwa sebuah aksi akan dijalankan setelah aksi sebelumnya dijalankan. Lingkaran hitam pekat menguraikan titik awal proses aktifitas dimulai. Titik hitam dengan lingkaran hitam

dibagian sisinya menerangkan akhir dari proses aktifitas tersebut. Garis horizontal berwarna hitam yang direpresentasikan sebagai percabangan merupakan pemisah dua aksi atau lebih secara bersamaan atau konkuren (Arianti, Fa'izi, Adam, & Wulandari, 2022).

2.7.4. Sequence Diagram

Sequence diagram merupakan diagram yang digunakan untuk menunjukkan urutan data yang lewat di antara objek untuk use case tertentu dari waktu ke waktu. Sequence diagram mengilustrasikan objek-objek yang berpartisipasi di dalam suatu use case (Indriyani, Yunita, & Muthia, 2019). Dalam menyusun *sequence diagram*, objek – objek yang terlibat dalam sebuah sistem harus diketahui terlebih dahulu agar dapat digunakan dalam menggambarkan pemodelan sequence diagram.

Pada sequence diagram, setiap objek atau aktor direpresentasikan oleh sebuah kotak vertikal yang disebut lifeline. Urutan pesan antar objek atau aktor ditunjukkan oleh panah yang menghubungkan dua lifeline, dengan pesan yang dikirim ditampilkan di atas panah. Sequence diagram juga dapat menunjukkan pemanggilan method atau fungsi yang terjadi selama interaksi. Ketika sebuah pesan dikirim, method atau fungsi yang terkait juga dapat ditampilkan dalam diagram dengan menampilkan nama method atau fungsi di atas pesan yang terkait. Sequence diagram sangat berguna untuk memodelkan proses bisnis, logika program, dan interaksi antar komponen dalam sistem yang kompleks. Diagram ini dapat membantu para pengembang atau arsitek perangkat lunak untuk memahami dan merancang sistem yang efektif dan efisien.

2.8. Penelitian terdahulu

Berikutakan dijabarkan beberapa penelitian terdahulu tentang *text mining* menggunakan metode TF-IDF.

Tabel 2.3 Penelitian Terdahulu

No.	Penulis(Tahun)	Topik	Hasil
1	Edo Harieby, Hoiriyah, Miftahul Walid (2022) (Harieby, Hoiriyah, & Walid, 2022)	Twitter <i>Text Mining</i> Mengenai Isu Vaksinasi Covid-19 Menggunakan Metode Term Frequency, Inverse Document Frequency (Tf-Idf)	Metode TF-IDF memberikan pembobotan pada kata yang paling banyak muncul pada kolom tweet mengenai isu vaksin di Indonesia. Dari kemunculan kata ini dapat digunakan sebagai kunci untuk menentukan isi tweet yang paling banyak diperbincangkan.
2.	Musthofa Galih Pradana (2020) (Pradana, 2020)	Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam <i>Text Mining</i>	Dari penelitian ini diperoleh beberapa informasi seperti lebih tingginya kata yang bersifat nasionalis pada media facebook dibandingkan dengan twitter. Sentimen pada facebook juga lebih cenderung positif dibandingkan pada media twitter. Term document matrix dan word cloud dapat saling melengkapi pada kegiatan <i>text mining</i> .

3.	Purniawan et al (2022) (Purniawan, Sasmita, & Pratama, 2022)	Clustering Berita Menggunakan Algoritma Tf-Idf Dan K-Means Dengan Memanfaatkan Sumber Data Crawling Pada Situs Detik.Com	Crawling dataset menggunakan bahasa pemrograman Python dan library BeautifulSoup4 mampu dan handal dalam melakukan pengambilan data <i>text</i> dalam jumlah besar pada portal berita online detik.com. Algoritma TF-IDF.
----	---	--	---



BAB III

ANALISIS DAN PERANCANGAN

3.1. Analisis Sistem Yang Berjalan

Google Trends menyediakan layanan RSS feed yang dapat digunakan oleh para pencari berita. Tidak hanya pengguna biasa, RSS feed juga banyak digunakan oleh pengguna lain seperti pemilik situs web untuk memperkaya konten atau informasi yang terdapat pada halaman web nya ataupun pengguna lain dengan kebutuhan khusus seperti peneliti yang biasanya menggunakan RSS feed ini untuk memperoleh dataset ataupun pola yang diinginkan.

Struktur RSS feed yang disajikan oleh *Google Trends* tergolong cukup rinci namun masih memiliki beberapa kekurangan dimana feed berita yang disediakan oleh RSS *Google Trends* belum diklasifikasikan kedalam kategori berita tertentu sehingga konten yang disediakan oleh RSS *Google Trends* tersebut tidak dapat langsung digunakan pada beberapa kebutuhan seperti *filtering*, *sentiment analysis*, *recommendation system*, dan kebutuhan – kebutuhan lainnya.

3.2. Analisis Sistem Yang Diusulkan

Penelitian tugas akhir ini mencoba membangun sebuah sistem yang dapat mengklasifikasikan feed berita yang diperoleh dari *Google Trends* RSS sehingga feed berita yang diperoleh dapat dikategorikan kedalam kategori berita tertentu sehingga dapat mengakomodasi kebutuhan lebih lanjut seperti *filtering*, rekomendasi dan lain sebagainya. Sistem yang dibangun akan berjalan pada layanan web atau *web-based* menggunakan bahasa pemrograman PHP yang berjalan pada *web server* Apache. Adapun metode yang akan diaplikasikan pada sistem yang dibangun nantinya adalah metode TF-IDF yang telah banyak

digunakan khususnya pada domain klasifikasi *text* atau dokumen. Pada sistem tersebut nantinya akan menyediakan beberapa fitur seperti pengelolaan dataset berita, pengujian klasifikasi dataset dan klasifikasi *realtime*.

Metode TF-IDF nantinya akan digunakan untuk menyusun corpus word yang memiliki skor yang signifikan terhadap kategori berita tertentu dari konten berita dataset yang mana corpus word tersebut akan digunakan untuk menghitung TF-IDF pada potongan judul dan snippet berita yang diperoleh dari *Google Trends RSS feed* untuk memperoleh kategori berita yang sesuai berdasarkan nilai kemiripannya menggunakan komputasi jarak sederhana.

3.3. Analisis Kebutuhan Sistem

Kebutuhan sistem yang dibahas pada bagian ini terdiri dari dua jenis kebutuhan yaitu kebutuhan perangkat keras dan kebutuhan perangkat lunak. Adapun kebutuhan perangkat keras yang dibutuhkan dalam pembangunan sistem ini adalah laptop atau komputer dengan spesifikasi yang mampu untuk menjalankan perangkat lunak yang dibutuhkan seperti Apache dan MySQL. Adapun daftar perangkat lunak yang dibutuhkan dalam pengembangan sistem adalah :

1. Sistem Operasi Windows 7 atau di atasnya.
2. Apache Server, PHP 7.4.3 atau di atasnya.
3. Code Editor seperti Visual Studio Code, Notepad ++, atau sejenisnya.
4. DBMS MySQL Server.

3.4. Metode Pengembangan Sistem

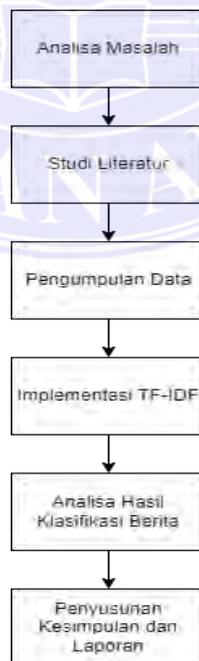
Metode pengembangan sistem yang dilakukan pada penelitian tugas akhir ini menggunakan metode pengembangan *waterfall* dimana proses pengembangan dilakukan secara *linear* sesuai dengan kebutuhan utama dari sistem yaitu klasifikasi

berita dari *Google Trends RSS* tanpa adanya kebutuhan – kebutuhan khusus lainnya sehingga metode *waterfall* dirasa sesuai untuk pengembangan sistem ini. Adapun tahapan pengembangan sistem yang dilakukan adalah :

1. Analisis Proses Klasifikasi Berita Menggunakan TF-IDF.
2. Perancangan Sistem Klasifikasi Berbasis Web
3. Pembuatan Kode Program
4. Pengujian
5. Pendukung dan Pemeliharaan.

3.5. Tahapan Penelitian

Dalam penelitian ini akan melalui beberapa tahapan kegiatan. Adapun kegiatan utama dibagi menjadi enam kegiatan yaitu Analisa masalah, studi literatur, pengumpulan data, implementasi TF-IDF, Analisa hasil klasifikasi berita dan penyusunan kesimpulan dan laporan. Adapun tahapan – tahapan dalam metodologi penelitian dapat digambarkan sebagai berikut :



Gambar 3. 1 Tahapan Penelitian

Adapun penjabaran atau penjelasan dari setiap tahapan penelitian seperti yang terlihat pada Gambar 3.1 adalah sebagai berikut :

1. Analisa Masalah

Pada tahapan ini dilakukan penentuan masalah yang timbul akibat dari kebutuhan – kebutuhan khusus yang belum terpenuhi dari *Google Trends RSS feed* yaitu tidak adanya informasi klasifikasi berita dari feed yang diperoleh. Sehingga berdasarkan kebutuhan tersebut maka pada tugas akhir ini dirumuskan masalah bagaimana feed dari berita – berita tersebut dapat diklasifikasikan menggunakan metode – metode yang ada sesuai dengan bidang ilmu dari penulis. Adapun hasil yang diharapkan adalah terbangunnya sebuah sistem yang dapat digunakan untuk mengklasifikasikan berita – berita tersebut.

2. Studi Literatur

Pada tahapan ini dilakukan pengumpulan literatur dalam bentuk teori dan konsep yang dibutuhkan untuk menyelesaikan masalah yang telah dirumuskan. Pada penelitian ini ditentukan metode yang digunakan untuk mengakomodasi kebutuhan klasifikasi berita yaitu metode TF-IDF yang mana berdasarkan studi yang telah dilakukan metode ini banyak diaplikasikan pada klasifikasi teks ataupun dokumen. Pada tahapan ini juga dilakukan pengumpulan materi – materi yang dibutuhkan dalam pengembangan sistem yang berkaitan dengan PHP dan MySQL.

3. Pengumpulan Data

Pada tahapan ini dilakukan pengumpulan data berupa dataset berita serta kategori – kategori berita yang umum digunakan pada beberapa situs berita seperti CNN, detiknew, Kompas, dan portal – portal berita lainnya. Konten – konten berita kemudian dikumpulkan berdasarkan kategorinya sehingga dapat digunakan sebagai dataset untuk membentuk corpus word menggunakan TF- IDF untuk kebutuhan klasifikasi feed *Google Trends*. Implementasi TF-IDF pada tahapan ini dilakukan kegiatan implementasi TF-IDF dalam mengklasifikasikan berita yang diperoleh dari *Google trends RSS feed*. Adapun tahapan implementasi TF-IDF dibagi menjadi beberapa tahapan yaitu Analisa dan perancangan sistem serta pembuatan kode program. Pada implementasi TF- IDF di sistem nantinya terdiri dari beberapa tahapan utama yaitu :

- a. Menggunakan dataset yang di-input ke dalam sistem, dataset tersebut kemudian digunakan untuk penyusunan corpus word / term yang digunakan sebagai vector fitur dari setiap kategori berita yang dapat dijabarkan lagi kedalam beberapa sub tahapan :
 - i. Normalisasi teks pada konten berita.
 - ii. *Punctuation Removal*
 - iii. *Tokenization*
 - iv. *Stopword Removal*
 - v. Perhitungan TF
 - vi. Perhitungan IDF
 - vii. Perhitungan TF-IDF
 - viii. Seleksi word / term berdasarkan skor TF-IDF

b. Menggunakan word / term yang diperoleh dari tahap sebelumnya, proses klasifikasi kemudian dapat dilakukan terhadap feed *Google trends* RSS yang dapat diperoleh pada URL <https://trends.google.com/trends/trendingsearches/daily/rss?geo=ID> dengan menghitung skor TF-IDF pada feed *Google Trends* menggunakan word / term yang telah ditentukan kemudian mencari kategori berita dengan kemiripan skor TF-IDF tertinggi sebagai kategori terpilih hasil dari klasifikasi.

4. Analisa Hasil Klasifikasi Berita

Pada tahapan ini akan dilakukan Analisa terhadap hasil dari klasifikasi berita yang dilakukan terhadap sistem yang telah dibangun. Analisa akan dilakukan dengan melakukan pengujian menggunakan dataset uji yang telah disiapkan untuk memperoleh akurasi klasifikasi dari metode dan sistem yang telah dibangun. Pengujian kemudian dilanjutkan dengan melakukan pengujian terhadap *Google Trend* RSS secara realtime. Hasil Analisa kemudian dibahas untuk melihat kinerja dari sistem yang telah dibangun.

5. Penyusunan Kesimpulan dan Laporan

Pada tahapan ini akan dilakukan penyusunan narasi kesimpulan yang diperoleh dari hasil kinerja sistem berdasarkan pengujian yang telah dilakukan. Kekurangan dan kelebihan dari hasil pengujian akan dirangkum dan disusun dalam bentuk laporan hasil penelitian.

3.6. Analisis Proses Klasifikasi TF-IDF

Klasifikasi feed berita dari *Google trends* RSS dilakukan menggunakan nilai dari skor TF-IDF. Kekurangan informasi kategori dari feed berita dari *Google trends* RSS akan coba diatasi dengan melakukan klasifikasi feed berita tersebut kedalam label kelas atau kategori berita tertentu berdasarkan skor TF-IDF nya. Adapun kategori berita yang digunakan pada penelitian ini dapat dilihat pada tabel 3.1 berikut.

Tabel 3. 1 Kategori Berita

No.	Label Kategori
1	Politik
2	Ekonomi
3	Olahraga
4	Teknologi

Tahap pertama dalam klasifikasi menggunakan TF-IDF adalah menyusun corpus yang akan digunakan dalam perhitungan TF-IDF dari konten berita yang akan diklasifikasi. Untuk menyusun corpus term maka dibutuhkan dataset corpus berita terlebih dahulu.

3.6.1. Dataset

Dataset yang digunakan pada penelitian ini adalah konten berita yang diperoleh dari beberapa portal berita yang kemudian di kelompokkan ke dalam label kategori sesuai dengan kategori dari portal berita yang diperoleh. Adapun contoh dataset berita dapat dilihat pada tabel 3.2 berikut.

Tabel 3. 2 Dataset Berita

No.	Kategori	Judul	Konten
1	Olahraga	Kedalaman Skuad Manchester United Sudah Buat Erik Ten Hag Percaya Diri	Manchester United mendatangkan lebih sedikit pemain di bursa transfer Januari dibandingkan dengan tim-tim top lainnya di Liga Inggris. Namun, manajer Erik ten Hag yakin klub yang bermarkas di Old Trafford sudah memiliki kedalaman skuad yang mumpuni. United, yang menghuni peringkat keempat di klasemen Liga Inggris, merekrut Jack Butland, Wout Weghorst, dan Marcel Sabitzer, dengan status pinjaman di bursa transfer. Ini berbeda dengan Chelsea menghabiskan hampir 300 juta pound atau sekitar Rp 5,4 triliun untuk membeli delapan pemain. Adapun Arsenal, yang kini menjadi pemimpin klasemen Liga Inggris merekrut empat pemain. Selain

3.6.2. Perhitungan Skor TF-IDF

Setelah dataset dikumpulkan, tahap selanjutnya adalah menghitung skor TF-IDF untuk setiap term yang diperoleh dari konten dataset berita. Term atau word dengan skor TF-IDF yang tinggi kemudian akan disimpan sebagai vector fitur untuk setiap kategori berita. Masing – masing kategori berita akan memiliki term dan vector fitur yang berbeda bergantung pada proses penyusunan corpus dari dataset berita dari masing – masing kategori.



Gambar 3. 2 *Flowchart Pembentukan Model Klasifikasi*

Pada proses awal, konten berita akan dinormalisasi terlebih dahulu dengan mentransformasikan karakter huruf menjadi huruf kecil.

Konten:

Kedalaman Skuad Manchester United Sudah Buat Erik Ten Hag Percaya Diri Manchester United mendatangkan lebih sedikit pemain di bursa transfer Januari dibandingkan dengan tim-tim top lainnya di Liga Inggris. Namun, manajer Erik ten Hag yakin klub yang bermarkas di Old Trafford sudah memiliki kedalaman skuad yang mumpuni. United, yang menghuni peringkat keempat di klasemen Liga Inggris, merekrut Jack Butland, Wout Weghorst, dan Marcel Sabitzer, dengan status pinjaman di bursa transfer. Ini berbeda dengan Chelsea menghabiskan hampir 300 juta pound atau sekitar Rp 5,4 triliun untuk membeli delapan pemain. Adapun Arsenal, yang kini menjadi pemimpin klasemen Liga Inggris merekrut empat pemain. Selain itu, tim peringkat ketiga, Newcastle

United, menandatangani lima pemain baru di pertengahan musim ini.

Normalisasi:

Kedalaman skuad Manchester United sudah buat Erik ten Hag percaya diri Manchester United mendatangkan lebih sedikit pemain di bursa transfer Januari dibandingkan dengan tim-tim top lainnya di Liga Inggris. Namun, manajer Erik ten Hag yakin klub yang bermarkas di Old Trafford sudah memiliki kedalaman skuad yang mumpuni. United, yang menghuni peringkat keempat di klasemen Liga Inggris, merekrut Jack Butland, Wout Weghorst, dan Marcel Sabitzer, dengan status pinjaman di bursa transfer. Ini berbeda dengan Chelsea menghabiskan hampir 300 juta pound atau sekitar Rp 5,4 triliun untuk membeli delapan pemain. Adapun Arsenal, yang kini menjadi pemimpin klasemen Liga Inggris merekrut empat pemain. Selain itu, tim peringkat ketiga, Newcastle United, menandatangani lima pemain baru di pertengahan musim ini.

Setelah normalisasi, konten berita kemudian akan diproses menggunakan *text preprocessing* pertama yaitu penghilangan punctuation atau punctuation removal yang bertujuan untuk menghilangkan karakter – karakter khusus dan tanda baca yang tidak dibutuhkan dalam proses penyusunan term.

Konten:

Kedalaman skuad Manchester United sudah buat Erik ten Hag percaya diri Manchester United mendatangkan lebih sedikit pemain di bursa transfer Januari dibandingkan dengan tim-tim top lainnya di Liga Inggris. Namun, manajer Erik ten Hag yakin klub yang bermarkas di Old Trafford sudah memiliki kedalaman skuad yang mumpuni. United, yang menghuni peringkat keempat di klasemen Liga Inggris, merekrut Jack Butland, Wout Weghorst, dan Marcel Sabitzer, dengan status pinjaman di bursa transfer. Ini berbeda dengan Chelsea menghabiskan hampir 300 juta pound atau sekitar Rp 5,4 triliun untuk membeli delapan pemain. Adapun Arsenal, yang kini menjadi pemimpin klasemen Liga Inggris merekrut empat pemain. Selain itu, tim peringkat ketiga, Newcastle United, menandatangani lima pemain baru di pertengahan musim ini.

Punctuation Removal:

kedalaman skuad manchester united sudah buat erik ten hag percaya diri manchester united mendatangkan lebih sedikit pemain di bursa transfer januari dibandingkan dengan tim tim top lainnya di liga inggris namun manajer erik ten hag yakin klub yang bermarkas di old trafford sudah memiliki kedalaman skuad yang mumpuni united yang menghuni peringkat keempat di klasemen liga inggris merekrut jack butland wout weghorst dan marcel sabitzer dengan status pinjaman di bursa transfer ini berbeda dengan chelsea menghabiskan hampir 300 juta pound atau sekitar rp 5 4 triliun untuk membeli delapan pemain adapun arsenal yang kini menjadi pemimpin klasemen liga inggris merekrut empat pemain selain itu tim peringkat ketiga newcastle united menandatangani lima pemain baru di pertengahan musim ini.

Setelah proses punctuation removal selesai, proses dilanjutkan dengan melakukan tokenisasi untuk memperoleh term yang terdapat pada konten berita. Dari konten berita diatas diperoleh beberapa term atau word yaitu :

kedalaman, skuad, manchester, united, sudah, buat, erik, ten, hag, percaya, diri, manchester, united, mendatangkan, lebih, sedikit, pemain, di, bursa, transfer, januari, dibandingkan, dengan, tim, tim, top, lainnya, di, liga, inggris, namun, manajer, erik, ten, hag, yakin, klub, yang, bermarkas, di, old, trafford, sudah, memiliki, kedalaman, skuad, yang, mumpuni, united, yang, menghuni, peringkat, keempat, di, klasemen, liga, inggris, merekrut, jack, butland, wout, weghorst, dan, marcel, sabitzer, dengan, status, pinjaman, di, bursa, transfer, ini, berbeda, dengan, chelsea, menghabiskan, hampir, 300, juta, pound, atau, sekitar, rp, 5, 4, triliun, untuk, membeli, delapan, pemain, adapun, arsenal, yang, kini, menjadi, pemimpin, klasemen, liga, inggris, merekrut, empat, pemain, selain, itu, tim, peringkat, ketiga, newcastle, united, menandatangani, lima, pemain, baru, di, pertengahan, musim, ini.

Setelah proses tokenization berhasil, maka selanjutnya akan dilakukan penghilangan stopword untuk menghilangkan kata – kata yang tidak memiliki

makna berarti pada klasifikasi. Adapun hasil dari operasi stopword removal adalah sebagai berikut :

kedalaman, skuad, manchester, united, erik, ten, hag, percaya, manchester, united, mendatangkan, pemain, bursa, transfer, januari, dibandingkan, tim, tim, top, liga, inggris, manajer, erik, ten, hag, yakin, klub, bemarkas, old, trafford, memiliki, kedalaman, skuad, mumpuni, united, menghuni, peringkat, keempat, klasemen, liga, inggris, merekrut, jack, butland, wout, weghorst, marcel, sabitzer, status, pinjaman, bursa, transfer, berbeda, chelsea, menghabiskan, 300, juta, pound, rp, 5, 4, triliun, membeli, delapan, pemain, arsenal, menjadi, pemimpin, klasemen, liga, inggris, merekrut, empat, pemain, tim, peringkat, ketiga, newcastle, united, menandatangani, lima, pemain, baru, pertengahan, musim.

Setelah proses *text preprocessing* selesai, maka proses dilanjutkan dengan menghitung skor TF dari konten berita diatas. Perhitungan dimulai dengan melakukan perhitungan frekuensi kemunculan setiap term atau word dari konten berita.

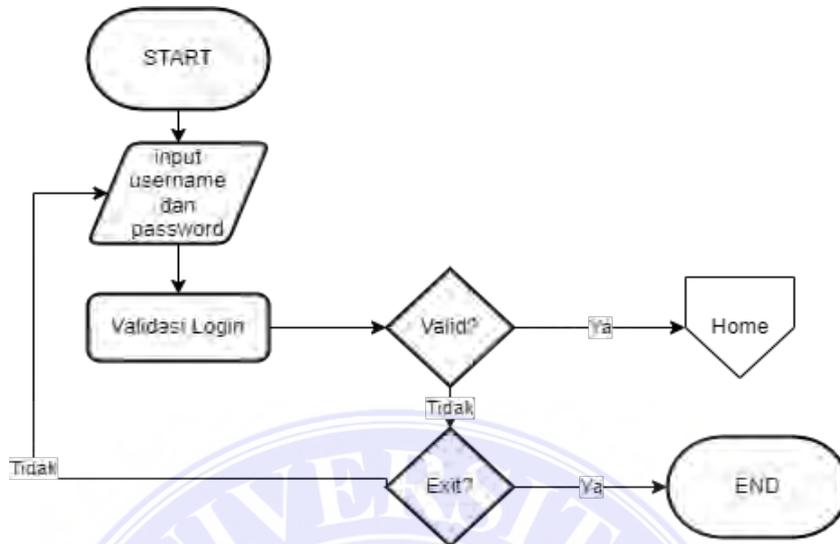
3.7. Perancangan Sistem

Pada bagian ini akan dibahas rancangan – rancangan yang dibutuhkan dalam pengembangan sistem klasifikasi feed RSS *Google trends*. Adapun rancangan yang akan disiapkan pada penelitian ini adalah *flowchart* dari sistem, *use case diagram*, *entity relational diagram*, struktur tabel dan antarmuka sistem.

3.7.1. Flowchart Sistem

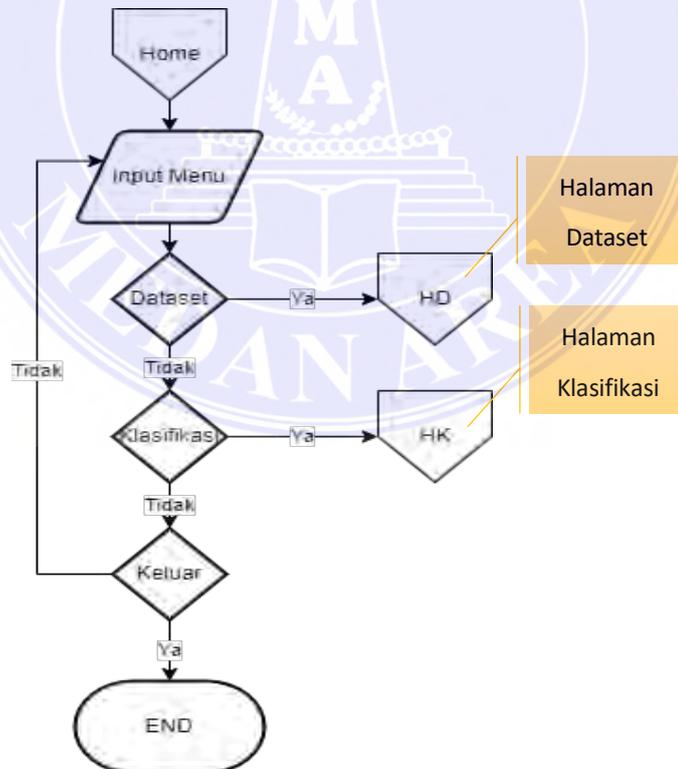
Sistem terdiri dari beberapa proses atau fitur utama yang terdiri dari login, halaman utama, halaman dataset dan halaman klasifikasi. Adapun flowchart dari sistem yang akan dibangun adalah sebagai berikut :

1. Flowchart Halaman Login



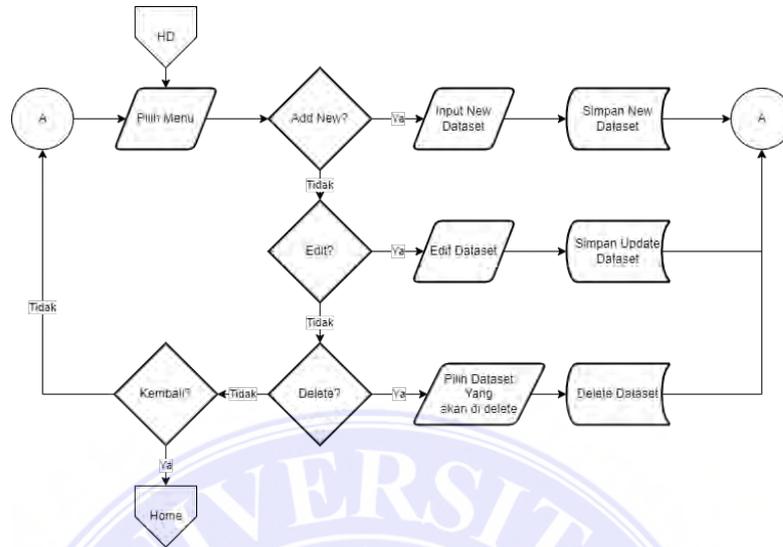
Gambar 3.3 Flowchart Halaman Login

2. Flowchart Halaman Utama



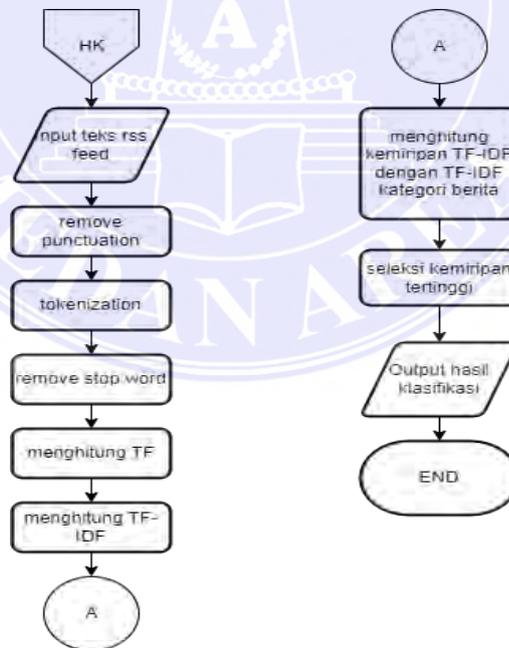
Gambar 3.4 Flowchart Halaman Utama

3. Flowchart Halaman Dataset



Gambar 3. 5 Flowchart Halaman Dataset

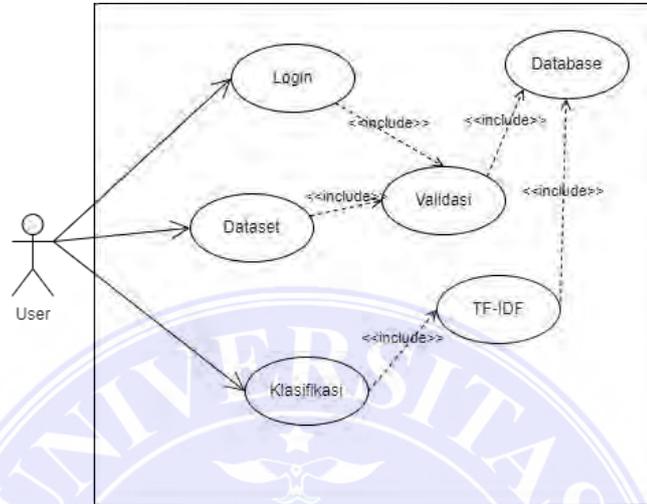
4. Flowchart Halaman Klasifikasi



Gambar 3. 6 Flowchart Halaman Klasifikasi

3.7.2. Use Case Diagram

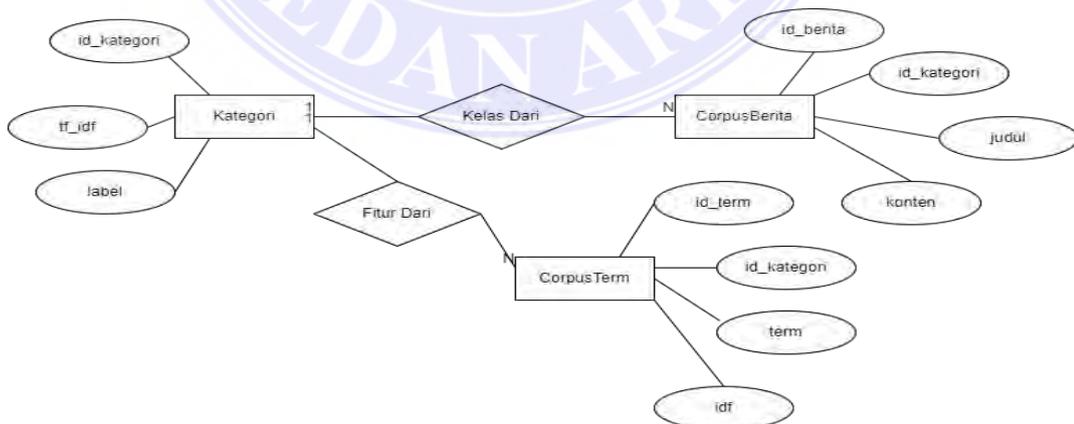
Adapun rancangan *use case diagram* dari sistem aplikasi klasifikasi feed berita dari *Google trends* menggunakan TF-IDF dapat dilihat pada gambar berikut.



Gambar 3. 7 Use Case Diagram

3.7.3. Entitiy Relational Diagram

Adapun rancangan dari *entitiy relational diagram* dari sistem aplikasi klasifikasi feed berita dari *Google trends* menggunakan TF-IDF dapat dilihat pada gambar berikut.



Gambar 3. 8 Entity Relationship Diagram

3.7.4. Struktur Table

Berikut rancangan *database* dan tabel pada aplikasi klasifikasi feed berita dari *Google trends* menggunakan metode TF-IDF.

1. Tabel User

Berikut rancangan tabel dari tabel user dari sistem yang akan dibangun :

Tabel 3. 3 Rancangan Tabel *User*

No.	Field Name	Type Field	Size
1	ID_User	Int	30
2	Nama	Varchar	255
3	Username	Varchar	255
4	Password	Varchar	255

2. Tabel Kategori

Berikut rancangan tabel dari tabel kategori dari sistem yang akan dibangun :

Tabel 3. 4 Rancangan Tabel Kategori

No.	Field Name	Type Field	Size
1	ID_Kategori	Int	30
2	Label	Varchar	255
3	Tf_idf	Double	10

3. Tabel CorpusBerita

Berikut rancangan tabel dari tabel corpusberita dari sistem yang akan dibangun :

Tabel 3. 5 Rancangan Tabel *Corpus* Berita

No.	Field Name	Type Field	Size
1	ID_Berita	Int	30
2	ID_Kategori	Int	30
3	Judul	Varchar	255
4	Konten	Varchar	255

4. Tabel CorpusTerm

Berikut rancangan tabel dari tabel corpusterm dari sistem yang akan dibangun :

Tabel 3. 6 Rancangan Tabel *Corpus Term*

No.	Field Name	Type Field	Size
1	ID_Term	Int	30
2	ID_Kategori	Int	30
3	Term	Varchar	255
4	Idf	Double	10

3.7.5. Perancangan Antarmuka

Bentuk antarmuka dari aplikasi klasifikasi terhadap *feed* berita dari *Google trends* RSS adalah sebagai berikut:

1. Rancangan Halaman Login

Halaman login merupakan halaman yang digunakan untuk melakukan login kedalam sistem. Berikut rancangan dari halaman login :



Gambar 3. 9 Rancangan Halaman Login

2. Rancangan Halaman Utama

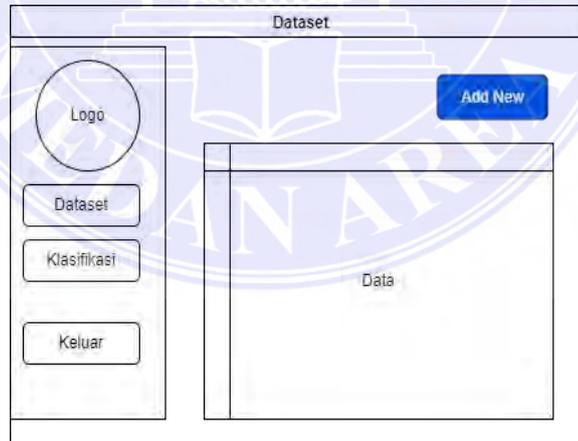
Halaman utama merupakan halaman yang menghubungkan halaman – halaman lainnya dan menyediakan menu utama. Berikut rancangan dari halaman utama :



Gambar 3. 10 Rancangan Halaman Utama

3. Rancangan Halaman Dataset

Halaman dataset merupakan halaman yang digunakan untuk mengelola dataset berita atau corpus berita. Berikut rancangan dari halaman dataset :



Gambar 3. 11 Rancangan Halaman *Dataset*

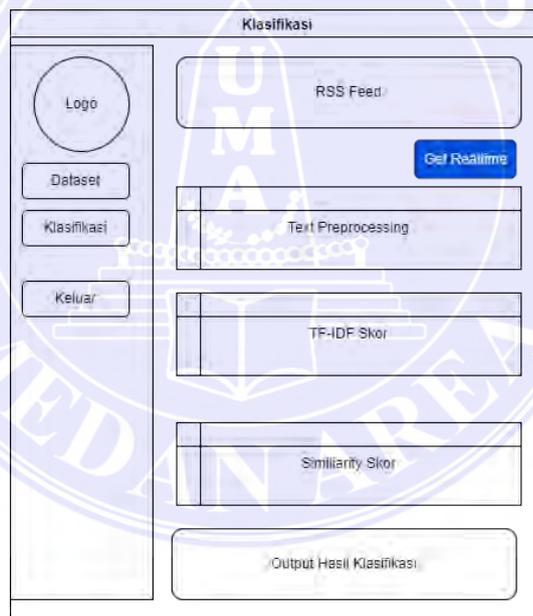
Pada halaman ini tersedia tombol “Add New” untuk menampilkan dialog kecil untuk menambah dataset baru yang mana tampilan dari dialog tersebut dapat dilihat pada gambar 3.12 berikut.



Gambar 3. 12 Dialog Tambah Data

4. Rancangan Halaman Klasifikasi

Halaman klasifikasi merupakan halaman yang digunakan untuk melakukan klasifikasi rss feed dari *Google trends*. Berikut rancangan dari halaman klasifikasi :



Gambar 3. 13 Rancangan Halaman Klasifikasi

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan kegiatan penelitian yang telah dilakukan serta pengujian menggunakan data berita untuk mengklasifikasikan kategori RSS Feed *Google Trends* maka diperoleh beberapa kesimpulan sebagai berikut:

Penelitian ini telah melakukan pengembangan model klasifikasi berita yang menggunakan 243 berita yang terbagi kedalam 4 kategori atau yang akan menjadi kelas klasifikasi. Dimana data berita tersebut terbagi menjadi 65 beritapolitik, 59 berita ekonomi, 77 berita olahraga dan 42 berita teknologi. Dimanahasil klasifikasi yang dilakukan pada berita uji dan berita latih ditunjukkan memberikan *Accuracy* yang cukup baik yaitu 76% untuk data uji dan 80% untuk data latih.

5.2. Saran

Adapun saran – saran yang dapat disampaikan dari hasil penelitian yang telah dilakukan adalah sebagai berikut :

1. Berdasarkan penelitian yang telah dilakukan maka diperlukan untuk penambahan kategori atau kelas berita untuk mengakomodasi kategori atau jenis berita lain yang muncul dari *Google trends RSS Feeds*.
2. Untuk meningkatkan akurasi dan konsistensi dalam klasifikasi berita maka dibutuhkan penambahan jumlah berita latih yang digunakan untuk memperkaya *word vector* dari masing – masing kategori.
3. Untuk penggunaan yang lebih luas, maka disarankan untuk mengaplikasikan model klasifikasi berita pada layanan RSS feed lain selain dari *Google trends*.
4. Pada penelitian yang akan datang diharapkan dapat dilakukan pengembangan dengan menggunakan metode lainnya selain dari TF- IDF.

DAFTAR PUSTAKA

- Arianti, T., Fa'izi, A., Adam, S., & Wulandari, M. (2022). Perancangan Sistem Informasi Perpustakaan Menggunakan Diagram UML (Unified Modelling Language). *Jurnal Ilmiah Komputer Terapan dan Informasi*, 1(1), 19-25.
- Ayuningtyas, I., & Wirawati, I. (2020). Nowcasting Tingkat Penghunian Kamar Hotel Menggunakan *Google Trends*. In *Seminar Nasional Official Statistics, 2020*, pp. 338-343.
- Fauziyyah, A. (2020). Analisis sentimen pandemi Covid19 pada streaming Twitter dengan *text mining Python*. *Jurnal Ilmiah SINUS*, 18(2), 31-42.
- Findawati, Y., & Rosid, M. (2020). *Buku Ajar Text Mining*. Sidoarjo: UMSIDA Press.
- Grandis, G., & Arumsari, Y. (2021). Seleksi Fitur Gain Ratio pada Analisis Sentimen Kebijakan Pemerintah Mengenai Pembelajaran Jarak Jauh dengan K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(8), 3507-3514.
- Harieby, E., Hoiriyah, H., & Walid, M. (2022). Twitter *text mining* mengenai isu vaksinasi COVID-19 menggunakan metode term frequency, inverse document frequency (TF-IDF). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 532-537.
- Hasan, J., Septiningrum, L., Chaery, A., Abdurachman, T., & Prawirayudha, A. (2020). Sistem Informasi Akuntansi (Flowchart) Dalam Pembangunan Masjid Al- Aulia. *Dedikasi Pkm*, 2(1), 118-125.
- Herwijayanti, B., Ratnawati, D., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(1), 306-312.
- Indriyani, F., Yunita, D., & Muthia, A. (2019). *Analisa Perancangan Sistem Informasi*. Jakarta: Bina Sarana Informatika Jakarta.
- Kusnady, D., & Siregar, A. (2018). Sistem Informasi Biaya Pendidikan (BPP) pada Politeknik Ganesha Medan Berbasis Web. *Juripol (Jurnal Institusi Politeknik Ganesha Medan)*, 1(1), 9-13.
- Pertiwi, F., & Irwansyah, I. (2020). Personal Branding Ria Ricis pada Media Sosial Instagram. *Jurnal Penelitian Komunikasi*, 23(1), 15-30.
- Pradana, M. (2020). Penggunaan fitur wordcloud dan document term matrix dalam *text mining*. *Jurnal Ilmiah Informatika*, 8(1), 38-43.
- Pratama, S., & Pratama, Y. (2022). Rancangan Dan Implementasi Aplikasi Konten

- Web News Aggregator Pada Perangkat Komunikasi Berbasis Android. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, 5(2), 175-182.
- Purniawan, I., Sasmita, G., & Pratama, I. (2022). Clustering Berita Menggunakan Algoritma TF-IDF dan K-Means Dengan Memanfaatkan Sumber Data Crawling Pada Situs Detik.com. *JITTER- Jurnal Ilmiah Teknologi dan Komputer*.
- Rosaly, R., & Prasetyo, A. (2019). *Pengertian Flowchart Beserta Fungsi dan Simbol- simbol Flowchart yang Paling Umum Digunakan*. Yogyakarta: Penerbit Andi.
- Rozi, F., & Sulistyawati, D. (2019). Klasifikasi Berita Hoax PILPRES Menggunakan Metode Modified K-Nearest Neighbor dan Pembobotan Menggunakan TF- IDF. *KONVERGENSI*, 15(1), 1-10.
- Satriawati, Z., Prasetyo, H., & Irawati, N. (2022). Kajian minat masyarakat terhadap pariwisata alternatif dan wisata pedesaan melalui *google trends*. *Kepariwisata: Jurnal Ilmiah*, 17(1), 18-26.
- Setiaji, & Sastra, R. (2021). Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Penggajian. *Jurnal Teknik Komputer AMIK BSI*, 7(1), 106-111.
- Suryani, L., & Edy, K. (2020). Pengembangan Aplikasi Lost & Found Berbasis Android Dengan Menggunakan Metode Term Frequency - Inverse Document Frequency (TF-IDF) dan Cosine Similarity. *Electro Luceat*, 6(2), 190-204.

LAMPIRAN

1. Source Code

a. Add Berita

```

<div class="pagetitle">
  <h1>Tambah Data Berita</h1>
  <nav>
    <ol class="breadcrumb">
      <li class="breadcrumb-item"><a
href="./index.php">Home</a></li>
      <li class="breadcrumb-item">Data</li>
      <li class="breadcrumb-item"><a
href="<?=base_url()?>uji">Data Berita Uji</a></li>
      <li class="breadcrumb-item">Tambah Berita Uji</li>
    </ol>
  </nav>
</div>

<section class="section">
  <div class="row">
    <div class="col-lg-12 mb-2">
      <div class="card">
        <div class="card-body">
          <h5 class="card-title">Form Tambah Berita
Uji</h5>
          <?php
            if (!empty($error))
            {
              ?>
              <div class="pt-2 pb-2">
                <div class="alert alert-danger alert-
dismissible fade show" role="alert">
                  <i class="bi bi-exclamation-octagon
me-1"></i>
                  <?php
                    foreach($error as $item){
                      echo $item.'<br>';
                    }
                  ?>
                  <button type="button" class="btn-
close" data-bs-dismiss="alert" aria-label="Close"></button>
                </div>
              </div>
            <?php
          }

```

```

        if (!empty($success))
        {
            ?>
            <div class="pt-2 pb-2">
                <div class="alert alert-success alert-
dismissible fade show" role="alert">
                    <i class="bi bi-check2-square me-
1"></i>
                    <?=$success?>
                    <button type="button" class="btn-
close" data-bs-dismiss="alert" aria-label="Close"></button>
                </div>
            </div>
            <?php
        }
    ?>
    <form method="POST"
action="<?=base_url()?>uji/add_proses" enctype="multipart/form-
data">
        <div class="row mb-3">
            <label for="kategori" class="col-sm-2
col-form-label">Kategori</label>
            <div class="col-sm-10">
                <select name="id_kategori"
class="form-control">
                    <?php foreach($kategori as
$item): ?>
                        <option
value="<?=$item['id_kategori']?>"><?=$item['kategori']?></option>
                    <?php endforeach; ?>
                </select>
            </div>
        </div>
        <div class="row mb-3">
            <label for="judul" class="col-sm-2 col-
form-label">Judul Berita</label>
            <div class="col-sm-10">
                <input type="text" name="judul"
class="form-control" value="">
            </div>
        </div>
        <div class="row mb-3">
            <label for="konten" class="col-sm-2 col-
form-label">Konten Berita</label>
            <div class="col-sm-10">
                <textarea name="konten" class="form-
control"></textarea>

```



```

        {
            ?>
            <div class="pt-2 pb-2">
                <div class="alert alert-danger alert-
dismissible fade show" role="alert">
                    <i class="bi bi-exclamation-octagon
me-1"></i>
                    <?php
                    foreach($error as $item){
                        echo $item.'<br>';
                    }
                    ?>
                    <button type="button" class="btn-
close" data-bs-dismiss="alert" aria-label="Close"></button>
                </div>
            </div>
            <?php
            }
            if (!empty($success))
            {
                ?>
                <div class="pt-2 pb-2">
                    <div class="alert alert-success alert-
dismissible fade show" role="alert">
                        <i class="bi bi-check2-square me-
1"></i>
                        <?=$success?>
                        <button type="button" class="btn-
close" data-bs-dismiss="alert" aria-label="Close"></button>
                    </div>
                </div>
                <?php
            }
            ?>
            <form method="POST"
action="<?=base_url()?>berita/edit_proses/<?=$berita['id_berita']?>"
enctype="multipart/form-data">
                <div class="row mb-3">
                    <label for="kategori" class="col-sm-2
col-form-label">Kategori</label>
                    <div class="col-sm-10">
                        <select name="id_kategori"
class="form-control">
                            <?php foreach($kategori as
$item): ?>
                                <?php

```

```

if ($item['id_kategori'] ==
$berita['id_kategori']){
    ?>
    <option
value="<?=$item['id_kategori']?>"
selected><?=$item['kategori']?></option>
    <?php
    }else{
    ?>
    <option
value="<?=$item['id_kategori']?>"><?=$item['kategori']?></option>
    <?php
    }
    ?>
    <?php endforeach; ?>
    </select>
    </div>
    </div>
    <div class="row mb-3">
    <label for="judul" class="col-sm-2 col-
form-label">Judul Berita</label>
    <div class="col-sm-10">
    <input type="text" name="judul"
class="form-control" value="<?=$berita['judul']?>">
    </div>
    </div>
    <div class="row mb-3">
    <label for="konten" class="col-sm-2 col-
form-label">Konten Berita</label>
    <div class="col-sm-10">
    <textarea name="konten" class="form-
control"><?=$berita['konten']?></textarea>
    </div>
    </div>
    <div class="row mb-3">
    <label class="col-sm-2 col-form-
label"></label>
    <div class="col-sm-10">
    <input type="submit" class="btn btn-
primary" value="Simpan">
    </div>
    </div>
    </form>
    </div>
    </div>
    </div>
    </div>
    </div>

```

```
</section>
<script>
$(document).ready(function () {
});
</script>
```

c. *Login*

```
<?php

namespace App\Controllers;

use App\Libraries\Auth;
use App\Models\AdminModel;

class Login extends BaseController
{
    public function index()
    {
        $auth = new Auth();

        if ($auth->is_login()){
            return redirect()->to('admin');
        }

        $session = session();

        $data = array();
        $data['error'] = $session->getFlashdata('error');
        $data['success'] = $session->getFlashdata('success');

        return view('login',$data);
    }

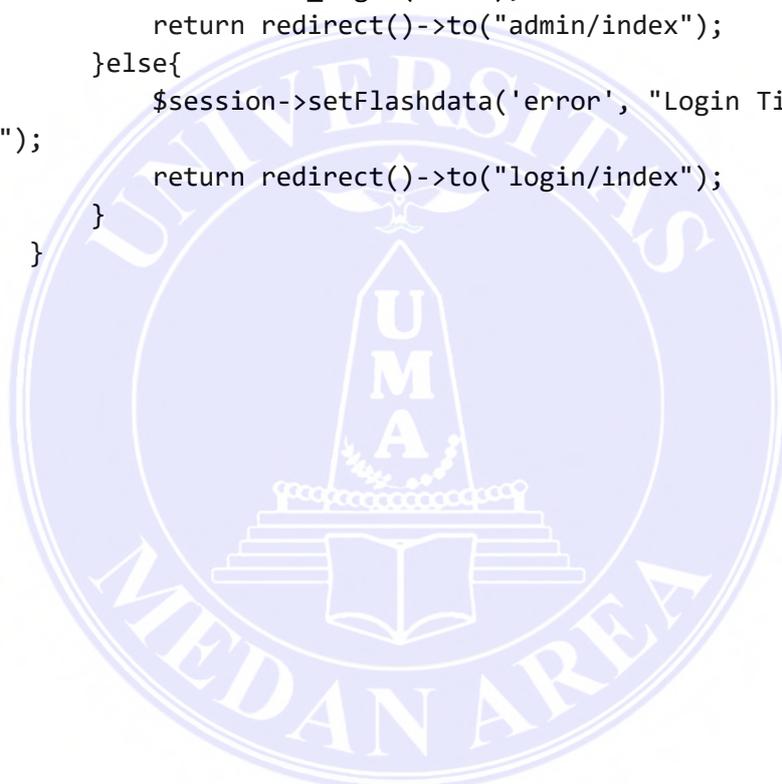
    public function auth()
    {
        $session = session();

        $data = [
            'username' => $this->request->getPost("username"),
            'password' => $this->request->getPost("password")
        ];

        $validation = \Config\Services::validation();
        $validation->setRule('username',
'Username','required|min_length[1]');
        $validation->setRule('password', 'Password', ['required',
'min_length[1]']);
```

```
        if (! $validation->run($data)) {
            $session->setFlashdata('error', "Input Login Tidak
Lengkap!");
            return redirect()->to("login/index");
        }else{
            $userModel = model('AdminModel');
            $user = $userModel->where('username',
$data['username']->where('password',$data['password']->first());

            if ($user != null){
                $auth = new Auth();
                $auth->set_login($user);
                return redirect()->to("admin/index");
            }else{
                $session->setFlashdata('error', "Login Tidak
Valid!");
                return redirect()->to("login/index");
            }
        }
    }
}
```



2. Turnitin

turnitin Similarity Report ID: oid:29477-51249859

PAPER NAME	AUTHOR
188160011_Frenklin Corenlius(Turnitin).docx	Frenklin Cornelius

WORD COUNT	CHARACTER COUNT
8413 Words	52689 Characters

PAGE COUNT	FILE SIZE
60 Pages	2.2MB

SUBMISSION DATE	REPORT DATE
Feb 12, 2024 9:56 AM GMT+7	Feb 12, 2024 9:57 AM GMT+7

25% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 23% Internet database
- 7% Publications database
- Crossref database
- Crossref Posted Content database
- 19% Submitted Works database

Excluded from Similarity Report

- Small Matches (Less than 10 words)

3. SK Pembimbing



UNIVERSITAS MEDAN AREA FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan Pesi Nomor 18 (061) 7366578, 7360168, 7364348, 7366781, Fax: (061) 7366798 Medan 20223
Kampus II : Jalan Seiabadi Nomor 79 / Jalan Sei Serayu Nomor 10 A, (061) 8223602, Fax: (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 449/FT.6/01.10/VI/2023
Lamp : -
Hal : Perpanjangan SK Pembimbing Tugas Akhir

13 Juni 2023

Yth. Pembimbing Tugas Akhir
Andre Hasudungan Lubis, S. Ti, MSc
di
Tempat

Dengan hormat,
Sehubungan telah berakhirnya waktu masa berlaku SK pembimbing nomor 376/FT.6/01.10/XII/2022 tertanggal 14 Desember 2022 maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa berikut :

Nama : Frenklin Cornelius
N P M : 188160011
Jurusan : Teknik Informatika

Oleh karena itu kami mengharapkan kesediaan saudara :

I. Andre Hasudungan Lubis, S. Ti, MSc (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :

"Penerapan TF-IDF pada Klasifikasi Kategori Berita pada Google Trends RSS Feeds"

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.



Dr. R. Hasudungan Lubis, S. Kom, M. Kom

4. Pengantar Riset



UNIVERSITAS MEDAN AREA FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estetis/Jalan PBSI Nomor 17 (061) 7366878, 7360168, 7364346, 7366781, Fax (061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Sanyu Nomor 70 A, (061) 8225602, Fax: (061) 8226331 Medan 20122.
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 427 /FT.6/01.10/VI/2023

6 Juni 2023

Lamp : -

Hal : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Wakil Rektor Bid. Pengembangan SDM & Adm. Keuangan
Jln. Kolam No.1
Di
Medan

Dengan hormat, kami mohon kesediaan ibu kiranya berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	NAMA	NPM	PRODI
1	Frenklin Cornelius	188160011	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir di **Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area**.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan Ilmiah dan Skripsi, yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul :

Penerapan TF-IDF pada Klasifikasi Kategori Berita pada *Google Trends RSS Feeds*.

Mohon kiranya tanggal Surat Izin Pengambilan Data Tugas Akhir agar disesuaikan dengan tanggal Terbitnya SK ini.

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.



Dr. R. Kom, M. Kom

Tembusan :

1. Ka. BAMAJ
2. Mahasiswa
3. File

5. Selesai Riset



UNIVERSITAS MEDAN AREA

Kampus I : Jalan Kolam Nomor 1 Medan Estato ☎ (061) 7360168, 7366878, 7364348 📠 (061) 7368012 Medan 20223
Kampus II : Jalan Seliabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 8225602 📠 (061) 8226331 Medan 20122
Website: www.uma.ac.id E-Mail: univ_medanarea@uma.ac.id

SURAT KETERANGAN Nomor : 1480/UMA/B/01.7/VIII/2023

Rektor Universitas Medan Area dengan ini menerangkan bahwa :

Nama : Frenklin Cornelius
No. Pokok Mahasiswa : 188160011
Program Studi : Teknik Informatika
Fakultas : Teknik

Benar telah selesai Pengambilan Data di Laboratorium Komputer Prodi Teknik Informatika Fakultas Teknik Universitas Medan Area dengan Judul "Penerapan *TF-IDF* Pada Klarifikasi Kategori Berita Pada *Google Trends RSS Feeds*."

Dan kami harapkan Data tersebut kiranya dapat membantu yang bersangkutan dalam penyusunan skripsi dan dapat bermanfaat bagi mahasiswa khususnya Fakultas Teknik.

Demikian surat ini diterbitkan untuk dapat digunakan seperlunya.

Medan, 14 Agustus 2023.

An Rektor,
Wakil Rektor Bidang Pengembangan SDM dan
Manajemen Keuangan,

Suswati, M.P.

Tembusan :

1. Mahasiswa Ybs
2. File