

## Application Of Text Mining In Grouping Thesis Topics Using TF-IDF Method Based On Thesis Abstract

### Penerapan Text Mining Dalam Pengelompokan Topik Skripsi Menggunakan Metode TF-IDF Berdasarkan Abstrak Skripsi

Delsy Andreswari<sup>1)</sup>; Dewi Suranti<sup>2)</sup>; Dimas Aulia Trianggara<sup>3)</sup>

<sup>1)</sup>Study Program of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu

<sup>2,3)</sup> Department of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu

Email: <sup>1)</sup> [delsyandreswari@gmail.com](mailto:delsyandreswari@gmail.com) ; <sup>2)</sup> [dewisuranti@unived.ac.id](mailto:dewisuranti@unived.ac.id) ; <sup>3)</sup> [dimasauliat@gmail.com](mailto:dimasauliat@gmail.com)

#### How to Cite :

Andreswari, D., Suranti, D., Trianggara, A, D. (2024). Penerapan Text Mining Dalam Pengelompokan Topik Skripsi Menggunakan Metode Tf-Idf Berdasarkan Abstrak Skripsi. Jurnal Komputer Indonesia, 3(2).  
Doi: <https://doi.org/10.37676/jki.v3i2>

#### ARTICLE HISTORY

Received [05 November 2024]

Revised [12 Desember 2024]

Accepted [15 Desember 2024]

#### KEYWORDS

Text Mining, Topik Skripsi,  
Metode TF-IDF.

This is an open access article under the  
[CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



#### ABSTRAK

UPT (Unit Pelayanan Teknis) Perpustakaan Universitas Dehasen Bengkulu mendokumentasikan skripsi-skripsi mahasiswa berdasarkan program studi menggunakan aplikasi Slims 9 (Bulian). Salah satunya yaitu Program Studi Informatika Fakultas Ilmu Komputer, di mana terdapat peminatan rekayasa perangkat lunak dan jaringan yang membedakan skripsi satu dengan skripsi yang lain. Hal ini yang terkadang membuat pihak perpustakaan kesulitan dalam menyediakan informasi kepada mahasiswa yang sedang mencari referensi, dikarenakan banyaknya data skripsi yang dikelola serta kesulitan mendapatkan informasi persentase banyaknya topik skripsi dari setiap dokumen skripsi tersebut. Penerapan *text mining* dalam pengelompokan topik skripsi menggunakan Metode TF-IDF berdasarkan abstrak skripsi di Perpustakaan Universitas Dehasen Bengkulu dapat membantu petugas perpustakaan dalam mengelompokkan skripsi dengan lebih detail untuk topik atau tema yang sama, membantu petugas perpustakaan dengan cepat dalam menemukan skripsi yang relevan dengan kebutuhan pengguna, dan dapat mengetahui informasi persentase banyaknya topik skripsi dari setiap judul dan abstrak yang telah diajukan khususnya pada mahasiswa/i Program Studi Informatika Fakultas Ilmu Komputer. Pada sistem telah di upload data training sebanyak 25 data yang digunakan sebagai dasar perhitungan dari Metode TF-IDF. Pada Metode TF-IDF terdapat beberapa proses yang terjadi, yakni Tokenizing, Filtering, Stemming dan nilai akhir tf-idf. Berdasarkan pengujian yang telah dilakukan terhadap data testing sebanyak 3 data skripsi, maka diperoleh hasil bahwa 2 skripsi termasuk ke dalam kelompok topik skripsi data mining dengan persentase sebesar 66,77% dan 1 skripsi termasuk ke dalam kelompok topik skripsi sistem pendukung keputusan dengan persentase 33,33%.

#### ABSTRACT

UPT (Technical Service Unit) Dehasen University Bengkulu Library documents student theses based on study programmes using the Slims 9 (Bulian) application. One of them is the Informatics Study Programme, Faculty of Computer Science, where there is a specialisation in software engineering and networking that distinguishes one thesis from another.

*This sometimes makes it difficult for the library to provide information to students who are looking for references, due to the large amount of thesis data that is managed and the difficulty in obtaining information on the percentage of the number of thesis topics from each thesis document. The application of text mining in grouping thesis topics using the TF-IDF Method based on thesis abstracts at the Dehasen University Library Bengkulu can help library staff in grouping theses in more detail for the same topic or theme, help library staff quickly find theses that are relevant to user needs, and can find out information on the percentage of the number of thesis topics from each title and abstract that has been submitted, especially in students of the Informatics Study Program, Faculty of Computer Science. The system has uploaded training data as much as 25 data which is used as the basis for the calculation of the TF-IDF Method. In the TF-IDF Method there are several processes that occur, namely Tokenizing, Filtering, Stemming and the final value of tf-idf. Based on testing that has been done on testing data as much as 3 thesis data, the results show that 2 theses belong to the data mining thesis topic group with a percentage of 66.77% and 1 thesis belongs to the decision support system thesis topic group with a percentage of 33.33%.*

## PENDAHULUAN

Perkembangan teknologi informasi saat ini menjadikan suatu informasi sebagai elemen penting dalam perkembangan masyarakat. Penyajian informasi tidak sepadan dengan kebutuhan informasi yang sangat tinggi, sehingga informasi tersebut digali lebih dalam dari data yang jumlahnya besar. Pencarian suatu informasi atau pola yang penting atau menarik dari data dalam jumlah besar digunakan para pengambil keputusan dalam memanfaatkan gudang data. Perpustakaan Perguruan Tinggi merupakan lembaga yang mengelola dan memberikan informasi kepada pembaca (Civitas Akademika) untuk mengembangkan ilmu pengetahuannya. Setiap perguruan tinggi umumnya memiliki perpustakaan sebagai fasilitas penunjang pembelajaran bagi anggota akademis.

Seiring dengan perkembangan ilmu pengetahuan, kebutuhan akan perpustakaan yang menyediakan koleksi buku dalam jumlah besar semakin meningkat. Salah satu koleksi di perpustakaan perguruan tinggi yang meningkat jumlahnya setiap tahun adalah skripsi. Selama ini, UPT (Unit Pelayanan Teknis) Perpustakaan Universitas Dehasen Bengkulu mendokumentasikan skripsi-skripsi mahasiswa yang telah terkomputerisasi menggunakan aplikasi Slims 9 (Bulian). Skripsi-skripsi ini mencakup data judul skripsi, nama pengarang, program studi, tahun terbit, dan cover skripsi yang dikelompokkan berdasarkan program studi masing-masing dan disusun secara alfabet.

Namun pada Program Studi Informatika Fakultas Ilmu Komputer terdapat peminatan yang membedakan skripsi satu dengan skripsi yang lain, dimana peminatan tersebut yakni Rekayasa Perangkat Lunak dan jaringan. Hal ini yang terkadang membuat pihak perpustakaan kesulitan dalam menyediakan informasi kepada mahasiswa yang sedang mencari referensi, dikarenakan banyaknya data skripsi yang dikelola dan hanya dikelompokkan sesuai program studi disusun secara alfabet.

Oleh karena itu, dalam penelitian ini dilakukan pengembangan terhadap sistem yang sudah berjalan saat ini dengan membangun suatu aplikasi yang dapat membantu proses pengelompokan data topik skripsi dari abstrak mahasiswa/i Program Studi Informatika Fakultas Ilmu Komputer di Perpustakaan Universitas Dehasen Bengkulu. Proses pengelompokan dilakukan dengan *generate* abstrak skripsi mahasiswa melalui Metode TF-IDF, dimana akan menghasilkan pengelompokan dari abstrak skripsi tersebut masuk ke dalam topik skripsi sistem pakar, data mining, sistem pendukung keputusan, atau jaringan. Proses penentuan bobot pada abstrak

menggunakan algoritma TF-IDF (*Term Frequency - Inverse Document Frequency*). Proses yang dilakukan TF-IDF adalah membandingkan setiap kata yang telah dihasilkan dari proses *text mining* dengan menghitung bobotnya berdasarkan algoritma TD-IDF. Nilai bobot yang didapatkan akan digunakan untuk menghitung tingkat kemiripan abstrak skripsi.

## LANDASAN TEORI

### Knowledge Discovery In Database (KDD)

Knowledge Discovery in Databases berkaitan dengan proses penemuan pengetahuan yang diterapkan pada database. Hal ini juga didefinisikan sebagai proses nontrivial untuk identifikasi data yang valid, baru berpotensi bermanfaat, dan akhirnya memiliki pola yang dapat dimengerti. Knowledge Discovery in Database (KDD) adalah proses yang bertujuan untuk menggali dan menganalisis sejumlah besar himpunan data dan mengekstrak informasi serta pengetahuan yang berguna (Arta, et al., 2019). Knowledge Discovery In Database (KDD) adalah salah satu metode yang dapat digunakan untuk memperoleh pengetahuan yang berasal dari database yang tersedia (Febianto & Palasara, 2019). Berdasarkan pengertian tersebut di atas, maka dapat disimpulkan bahwa Knowledge Discovery In Database (KDD) merupakan suatu metode yang digunakan untuk menemukan pengetahuan baru pada database yang tersedia, serta mengekstrak informasi yang berguna melalui pola yang telah ditetapkan

### Text Mining

Text mining adalah proses untuk memperoleh informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya didapatkan karena memperhatikan pola dan tren dengan cara mempelajari pola statistik. Pada proses teks mining terdapat pembobotan kata yang bertujuan untuk memberikan nilai/bobot pada term yang terdapat pada suatu dokumen. Bobot yang diberikan pada term tergantung kepada metode yang digunakan (Deolika, et al., 2019). Text mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah sumber data melalui identifikasi pola yang menarik. Pada kasus text mining, sumber data adalah kumpulan data tekstual yang tidak terstruktur pada dokumen. Text mining bertujuan untuk menemukan informasi yang tidak diketahui, sesuatu yang belum diketahui dan belum dapat ditulis (Fani, et al., 2021). Text mining adalah satu di antara bidang khusus dalam data mining yang memiliki definisi menambang data berupa teks dengan sumber data yang biasanya didapatkan dari dokumen. Tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen (Ismail, et al., 2022).

### Metode TF-IDF

Metode Term Frequency Invers Document Frequency (TF-IDF) merupakan metode yang digunakan menentukan seberapa jauh keterhubungan kata (term) terhadap dokumen dengan memberikan bobot setiap kata. Metode TF-IDF ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut [6]. Dalam perhitungan bobot menggunakan TF-IDF, dihitung terlebih dahulu nilai TF perkata dengan bobot masing-masing kata adalah 1 (Deolika, et al., 2019). Metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu teknik yang digunakan dalam pengolahan teks dan pemodelan bahasa alami. Tujuan utama dari metode TF-IDF adalah untuk mengevaluasi seberapa penting suatu kata (term) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar.

### Adobe Dreamweaver

Adobe Dreamweaver adalah perangkat lunak yang terkemuka untuk desain web yang menyediakan kemampuan visual yang intuitif termasuk pada tingkat kode, yang dapat digunakan untuk membuat dan mengedit website HTML serta aplikasi mobile seperti smartphone, tablet, dan

perangkat lainnya. Aplikasi ini menyediakan transfer dan fitur sinkronisasi, kemampuan untuk mencari dan mengganti baris teks atau kode untuk mencari kata atau kalimat biasa di seluruh situs, dan templating feature yang memungkinkan untuk berbagi satu sumber kode atau memperbarui tata letak di seluruh situs tanpa server side includes atau scripting. Behavior Panel juga memungkinkan penggunaan JavaScript dasar tanpa pengetahuan coding, dan integrasi dengan Adobe Spry Ajax framework menawarkan akses mudah ke konten yang dibuat secara dinamis dan interface (Haerani & Nugraha, 2019). Dengan menggunakan aplikasi ini, pengembangan web dapat dilakukan secara visual, sehingga hasil perancangan web dapat langsung terlihat tanpa harus menggunakan aplikasi bantu peramban seperti Google Chrome, Firefox atau Internet Explorer. Teknologi web yang didukung oleh Adobe Dreamweaver sangat beragam, salah satunya adalah teknologi untuk kebutuhan pengembangan web berbasis mobile.

### **Bahasa Pemrograman PHP**

Hypertext Preprocessor atau lebih akrab dengan PHP merupakan bahasa pemrograman script server-side yang di desain untuk pengembangan web. PHP disebut bahasa pemrograman server-side karena diproses pada komputer server. Hal ini berbeda dengan bahasa pemrograman client-side seperti javascript yang diproses di dalam web browser. PHP dapat digunakan secara gratis dan bersifat open source. PHP dirilis dalam lisensi PHP License, sedikit berbeda dengan lisensi GNU (General Public License) yang biasa digunakan untuk proyek open source (Jannah, Sarwandi dan Creative 2019). Saat ini sudah banyak web server yang dapat di instal di dalam komputer, salah satunya aplikasi Xampp. Di dalam aplikasi ini terdapat beberapa fitur yang digunakan untuk menjalankan kode PHP, termasuk web server Apache. Web Server Apache berguna untuk memilah cara menjalankan kode script yang telah ditulis, sehingga apache akan memberitahu web server bahwa kode yang sedang dijalankan adalah kode PHP.

### **Konsep Perancangan Basis Data**

Database merupakan sebuah struktur yang umumnya dikategorikan dalam dua hal yaitu sebuah database flat dan sebuah database relasional. Database dapat diartikan sebagai suatu pengorganisasian data dengan bantuan komputer, yang memungkinkan dapat diakses dengan mudah dan cepat. Database merupakan salah satu sistem yang dibentuk untuk mengorganisasi, menarik dan menyimpan sebuah data yang lebih mudah (Kadarsih & Pujiyanto, 2022). Basis data merupakan kumpulan dari semua data yang ada di dalam organisasi atau semacamnya. Database adalah suatu kumpulan data terhubung yang disimpan secara bersama-sama pada suatu media tanpa mengatap satu sama lain atau tidak perlu suatu kerangkapan data, data disimpan dengan cara tertentu sehingga mudah digunakan atau ditampilkan kembali (Plaza, 2021). Database adalah kumpulan informasi yang disimpan di dalam komputer secara sistematis sehingga dapat diperiksa menggunakan suatu program komputer untuk memperoleh informasi dari basis data tersebut. Database adalah representasi kumpulan fakta yang saling berhubungan disimpan secara bersama sedemikian rupa dan tanpa pengulangan (redudansi) yang tidak perlu, untuk memenuhi berbagai kebutuhan. Database merupakan sekumpulan informasi yang saling berkaitan pada suatu subjek tertentu pada tujuan tertentu pula. Database adalah susunan record data operasional lengkap dari suatu organisasi atau perusahaan, yang diorganisir dan disimpan secara terintegrasi dengan menggunakan metode tertentu dalam komputer sehingga mampu memenuhi informasi yang optimal yang dibutuhkan oleh para pengguna (Helmud, 2021). Basis data adalah himpunan kelompok data yang saling berhubungan yang diorganisasi sedemikian rupa agar kelak dapat dimanfaatkan kembali dengan cepat dan mudah. Tujuan basis data dibagi menjadi 2 kelompok yaitu tujuan primer (tujuan utama yang ingin dicapai dalam usaha perancangan dan pengembangan basis data), dan tujuan sekunder (tujuan tambahan yang dimaksudkan untuk mencapai tujuan primer). Sistem basis data adalah sistem yang terdiri dari sekumpulan program yang dikenal sebagai DBMS (Database Management System), serta sejumlah tabel data yang saling berhubungan. DBMS

memungkinkan beberapa pemakai dan program lain untuk mengakses dan mengubah tabel-tabel tersebut (Hariyono, et al. 2023).

## METODE PENELITIAN

### Analisa Sistem Aktual

Selama ini, UPT (Unit Pelayanan Teknis) Perpustakaan Universitas Dehasen Bengkulu menampung skripsi-skripsi mahasiswa yang telah terkomputerisasi menggunakan aplikasi Slims 9 (Bulian). Skripsi-skripsi ini mencakup data judul skripsi, nama pengarang, program studi, tahun terbit, dan cover skripsi yang dikelompokkan berdasarkan program studi masing-masing dan disusun secara alfabet. Namun pada Program Studi Informatika Fakultas Ilmu Komputer terdapat peminatan yang membedakan skripsi satu dengan skripsi yang lain, dimana peminatan tersebut yakni Rekayasa Perangkat Lunak dan jaringan. Hal ini yang terkadang membuat pihak perpustakaan kesulitan dalam menyediakan informasi kepada mahasiswa yang sedang mencari referensi, dikarenakan banyaknya data skripsi yang dikelola dan hanya dikelompokkan sesuai program studi disusun secara alfabet.

### Analisa Sistem Baru

Analisa sistem baru dilakukan untuk mengidentifikasi permasalahan pada sistem aktual/lama. Oleh karena itu, dalam penelitian ini dilakukan pengembangan terhadap sistem yang sudah berjalan saat ini dengan membangun suatu aplikasi yang dapat membantu proses pengelompokan data topik skripsi dari abstrak mahasiswa/i Program Studi Informatika Fakultas Ilmu Komputer di Perpustakaan Universitas Dehasen Bengkulu. Proses pengelompokan dilakukan dengan *generate* abstrak skripsi mahasiswa melalui Metode TF-IDF, dimana akan menghasilkan pengelompokan dari abstrak skripsi tersebut masuk ke dalam 4 kelompok topik skripsi yaitu Sistem Pakar, Data Mining, Sistem Pendukung Keputusan atau Jaringan. Proses penentuan bobot pada abstrak menggunakan algoritma TF-IDF (*Term Frequency – Inverse Document Frequency*). Proses yang dilakukan TF-IDF adalah membandingkan setiap kata yang telah dihasilkan dari proses *text mining* dengan menghitung bobotnya berdasarkan algoritma TD-IDF. Nilai bobot yang didapatkan akan digunakan untuk menghitung tingkat kemiripan abstrak skripsi.

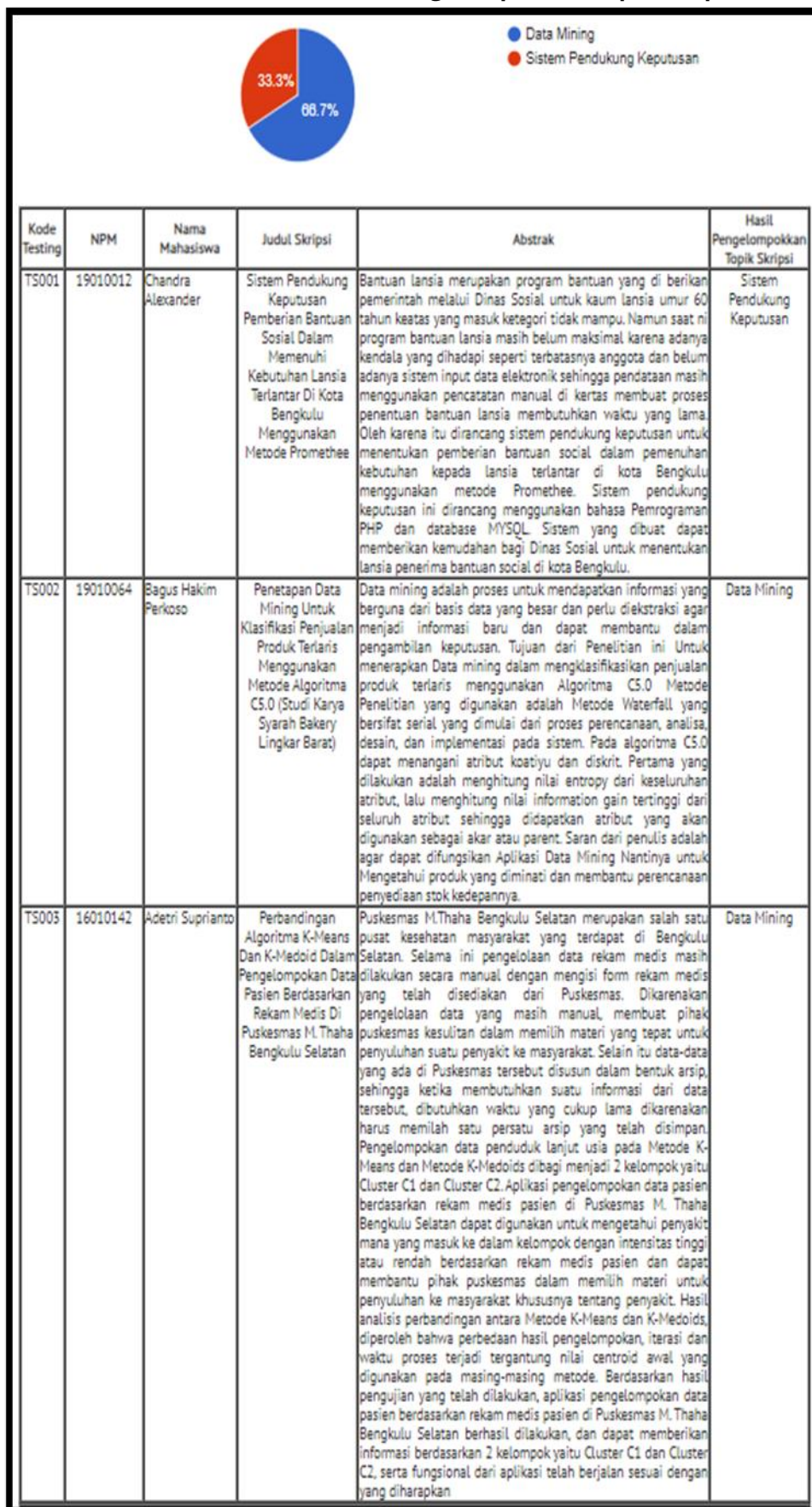
## HASIL DAN PEMBAHASAN

Penerapan *text mining* dalam pengelompokan topik skripsi menggunakan Metode TF-IDF berdasarkan abstrak skripsi di Perpustakaan Universitas Dehasen Bengkulu dapat membantu petugas perpustakaan dalam mengelompokkan skripsi dengan lebih detail untuk topik atau tema yang sama, membantu petugas perpustakaan dengan cepat dalam menemukan skripsi yang relevan dengan kebutuhan pengguna, dan dapat mengetahui informasi persentase banyaknya topik skripsi dari setiap judul dan abstrak yang telah diajukan khususnya pada mahasiswa/i Program Studi Informatika Fakultas Ilmu Komputer.

Pada sistem telah di upload data training sebanyak 25 data yang digunakan sebagai dasar perhitungan dari Metode TF-IDF. Pada Metode TF-IDF terdapat beberapa proses yang terjadi, yakni Tokenizing, Filtering, Stemming dan nilai akhir tf-idf. Tokenizing digunakan untuk memecah kalimat pada abstrak skripsi menjadi per kata. Filtering digunakan untuk membuang kata-kata yang tidak perlu atau tidak penting. Stemming digunakan untuk mengubah kata imbuhan awalan atau akhiran menjadi kata dasar.

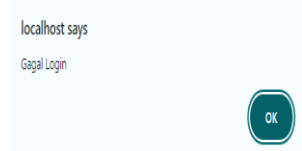
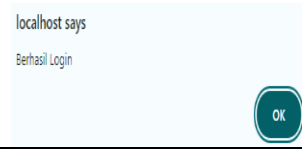
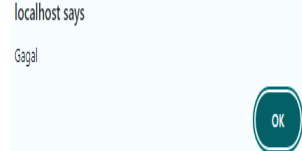
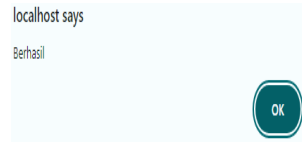
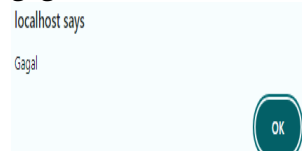
Setelah dilakukan pengujian terhadap data testing sebanyak 3 data skripsi, maka diperoleh hasil bahwa 2 skripsi termasuk ke dalam kelompok topik skripsi data mining dengan persentase sebesar 66,77% dan 1 skripsi termasuk ke dalam kelompok topik skripsi sistem pendukung keputusan dengan persentase 33,33%, seperti terlihat grafik hasil pada Gambar 1.




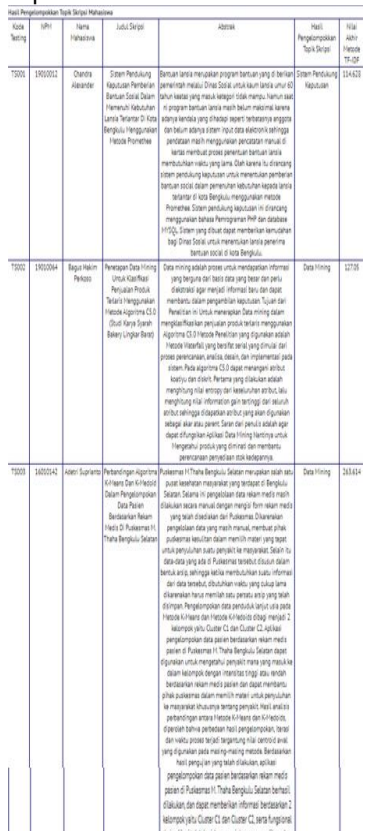
Gambar 1. Hasil Pengelompokan Topik Skripsi



Pengujian dilakukan dengan cara menguji coba fungsionalitas dari aplikasi penerapan *text mining* dalam pengelompokan topik skripsi menggunakan Metode TF-IDF berdasarkan abstrak skripsi di Perpustakaan Universitas Dehasen Bengkulu. Adapun hasil pengujian menggunakan metode black box, terlihat pada Tabel 1.

**Tabel 1 Hasil Pengujian**

No	Komponen Yang Diuji	Skenario Pengujian	Hasil Pengujian	Ket
1.	Login	Memasukkan username dan password yang salah	Sistem menolak akses login tersebut dengan memberikan pesan kesalahan 	Berhasil sesuai harapan
		Memasukkan username dan password yang benar	Sistem menerima akses login tersebut dengan memberikan pesan berhasil dan menampilkan halaman menu utama 	Berhasil sesuai harapan
2	Input Data Topik Skripsi	Memasukkan data topik skripsi yang sudah ada dalam database	Sistem menolak menyimpan data dan menampilkan pesan gagal 	Berhasil sesuai harapan
		Memasukkan data topik skripsi yang belum ada dalam database	Sistem berhasil menyimpan data dan menampilkan pesan berhasil 	Berhasil sesuai harapan
3	Input Data training	Memasukkan data training yang sudah ada dalam database	Sistem menolak menyimpan data dan menampilkan pesan gagal 	Berhasil sesuai harapan

		Memasukkan data training yang belum ada dalam database	Sistem berhasil menyimpan data dan menampilkan pesan berhasil localhost says Berhasil 	Berhasil sesuai harapan
4	Input Data testing	Memasukkan data testing yang sudah ada dalam database	Sistem menolak menyimpan data dan menampilkan pesan gagal localhost says Gagal 	Berhasil sesuai harapan
		Memasukkan data testing yang belum ada dalam database	Sistem berhasil menyimpan data dan menampilkan pesan berhasil localhost says Berhasil 	Berhasil sesuai harapan
4	Pengelompokkan Topik Skripsi K-Means	Menjalankan proses pengelompokkan topik skripsi melalui Metode TF-IDF berdasarkan tahun akademik dan periode yang dipilih	Sistem berhasil menjalankan proses pengelompokkan topik skripsi sesuai dengan tahun akademik dan periode yang dipilih 	Berhasil sesuai harapan

Berdasarkan pengujian sistem yang telah dilakukan, dapat disimpulkan bahwa fungsional dari aplikasi penerapan *text mining* dalam pengelompokan topik skripsi menggunakan Metode TF-IDF berdasarkan abstrak skripsi di Perpustakaan Universitas Dehasen Bengkulu telah berjalan dengan baik sesuai harapan.

## KESIMPULAN DAN SARAN

### Kesimpulan

Berdasarkan hasil dari pembahasan dan pengujian yang telah dilakukan, maka dapat disimpulkan bahwa :

1. Penerapan *text mining* dalam pengelompokan topik skripsi menggunakan Metode TF-IDF berdasarkan abstrak skripsi di Perpustakaan Universitas Dehasen Bengkulu dapat membantu petugas perpustakaan dalam mengelompokkan skripsi dengan lebih detail untuk topik atau tema yang sama, membantu petugas perpustakaan dengan cepat dalam menemukan skripsi yang relevan dengan kebutuhan pengguna, dan dapat mengetahui informasi persentase banyaknya topik skripsi dari setiap judul dan abstrak yang telah diajukan khususnya pada mahasiswa/i Program Studi Informatika Fakultas Ilmu Komputer.
2. Pada sistem telah di upload data training sebanyak 25 data yang digunakan sebagai dasar perhitungan dari Metode TF-IDF. Pada Metode TF-IDF terdapat beberapa proses yang terjadi, yakni Tokenizing, Filtering, Stemming dan nilai akhir tf-idf.
3. Setelah dilakukan pengujian terhadap data testing sebanyak 3 data skripsi, maka diperoleh hasil bahwa 2 skripsi termasuk ke dalam kelompok topik skripsi data mining dengan persentase sebesar 66,77% dan 1 skripsi termasuk ke dalam kelompok topik skripsi sistem pendukung keputusan dengan persentase 33,33%.
4. Berdasarkan pengujian sistem yang telah dilakukan, dapat disimpulkan bahwa fungsional dari aplikasi penerapan *text mining* dalam pengelompokan topik skripsi menggunakan Metode TF-IDF berdasarkan abstrak skripsi di Perpustakaan Universitas Dehasen Bengkulu telah berjalan dengan baik sesuai harapan.

### Saran

Berdasarkan penelitian yang telah penulis lakukan, maka penulis menyarankan :

1. Dapat menggunakan aplikasi ini untuk memberikan informasi hasil pengelompokan topik skripsi berdasarkan abstrak skripsi mahasiswa.
2. Perlu adanya pengembangan sistem untuk penelitian selanjutnya dengan menggunakan pendekatan metode lain sebagai hasil perbandingan pengelompokan seperti Metode KNN, Naive Bayes atau Levenstein Distance.

## DAFTAR PUSTAKA

- Arta, I. K. J., Indrawan, G. & Dantes, G. R., 2019. Data Mining Rekomendasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode Technique For Other Reference By Similiarity To Ideal Solution. *Jurnal Ilmu Komputer Indonesia (JKI)*, Volume Vol.4 No.1 ISSN:2615-2703.
- Deolika, A., Kusriani & Luthfi, E. T., 2019. Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, Volume Vol.3 No.2 e-ISSN:2615-2738.
- Fani, S. M., Santoso, R. & Suparti, 2021. Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Bilibili Pada Media Sosial Twitter Menggunakan K-Means Clustering. *Jurnal Gaussian*, Volume Vol.10 No.4 ISSN:2339-2541.

- Febianto, N. I. & Palasara, N. D., 2019. Analisis Clustering K-Means Pada Data Informasi Kemiskinan di Jawa Barat Tahun 2018. *Jurnal Sisfokom*, Volume Vol.8 No.2 2019.
- Firman, A., 2019. *Analisis dan Perancangan Sistem Informasi*. Surabaya: Penerbit Qiara Media.
- Haerani, R. & Nugraha, D., 2019. Rancangan Aplikasi Pemesanan Makanan Dengan Macromedia Dreamweaver dan MySQL. *Seminar Nasional Sains Teknologi dan Inovasi Indonesia (Senastindo AAU)*, Volume Vol.1 No.1 ISSN:2685-8991.
- Hariyono, R. C. S. et al., 2023. *Buku Ajar Pengantar Basis Data*. Jambi: PT. Sonpedia Publishing Indonesia.
- Helmud, E., 2021. Optimasi Basis Data Oracle Menggunakan Complex View Studi Kasus : PT. Berkat Optimis Sejahtera (PT.BOS) Pangkalpinang. *Jurnal Informatika*, Volume Vol.7 No.1 ISSN.2407-1730.
- Herlina, Rusman, A. D. P., Marlina & Suwardoyo, U., 2022. *Penerapan Sistem Informasi Berbasis IT Pengolahan Data Rekam Medis Untuk Peningkatan Pelayanan di Rumah Sakit*. Pekalongan Jawa Tengah: PT. Nasya Expanding Management ISBN:978-623-423-378-0.
- Ismail, Rahmah, H. N. & Sulistiyowati, R., 2022. Penggunaan Software Orange Data Mining Pada Implementasi Text Mining Dalam Analisis Sentimen Netizen di Twitter Terhadap Kelangkaan Minyak Goreng. *Jurnal Sigma-Mu*, Volume Vol.14 No.2.
- Jannah, M., Sarwandi & Creative, C., 2019. *Mahir Bahasa Pemrograman PHP*. Jakarta: PT. Elex Media Komputindo.
- Kadarsih & Pujiyanto, D., 2022. *Step By Step Belajar Database MySQL Untuk Pemula*. Tangerang Selatan: Pascal Books.
- Lesmana, G., 2023. Analisis Kemampuan Metodologi Penelitian Mahasiswa Ditinjau Dari Kualitas Penulisan Skripsi. *Jurnal Manajemen Pendidikan Dasar Menengah dan Tinggi (JMP-DMT)*, Volume Vol.4 No.2 e-ISSN:2715-9213.
- Plaza, M. A. J., 2021. *Desain Basis Data*. Yogyakarta: Deepublisher ISBN:978-623-02-2338-9.
- Septiani, D. & Isabela, I., 2022. Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks. *Jurnal Sistem dan Teknologi Informasi Indonesia (SINTESIA)*, Volume Vol.1 No.2 e-ISSN:2807-9108.
- Suprpto, U., 2021. *Pemodelan Perangkat Lunak (C3) Kompetensi Keahlian : Rekayasa Perangkat Lunak Untuk SMK/MAK Kelas XI*. Jakarta: Grasindo.
- Yendrianof, D. et al., 2022. *Analisis dan Perancangan Sistem Informasi*. Medan: Yayasan Kita Menulis.