**Komputika: Jurnal Sistem Komputer**
Vol. 8, No. 2, Oktober 2019, hlm. 59 - 64
DOI:

**ISSN: 2252-9039 (print)**
**ISSN: 2655-3198 (online)**

# Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi Dengan Metode K-Means Clustering (Studi Kasus: Prodi Sistem Komputer)

**Muhammad Faishal Riyadhi[1]**

[1)]Program Studi Sistem Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia
Jl. Dipati Ukur No. 112 – 116, Bandung, Indonesia 40132
*Faishal.riyadhi@email.unikom.ac.id*

***ABSTRAK*** *– Dengan banyaknya mahasiswa yang akan mengerjakan tugas akhir, maka diperlukan suatu sistem yang dapat memberikan informasi tentang tren topik skripsi apa saja yang sedang populer pada tahun-tahun tertentu. Oleh karena itu melalui penelitian ini dikembangkan suatu aplikasi yang dapat bekerja secara semi-otomatis dengan memanfaatkan teknologi Text Mining dan Algoritma K-Means Clustering. Dari hasil penelitian yang telah dilakukan maka didapatkan hasil bahwa, sistem yang dibuat dapat membantu para mahasiswa untuk mengetahui informasi tren topik skripsi apa saja yang sedang tren di program studi sistem komputer. Untuk proses analisis menggunakan metode k-means clustering, tingkat keberhasilan yang didapat sebesar 66.66% untuk proses matematis. Dan untuk proses sistemnya sebesar 33.33% untuk data yang sama dengan proses matematis.*

***Kata Kunci*** *– Text Mining; K-Means Clustering;*

***ABSTRACT*** *– With so many students who will work on the final project, then we need a system that can provide information about the trends of any thesis topic that is popular in certain years. Therefore through this research an application was developed that could work semi-automatically by utilizing Text Mining technology and the K-Means Clustering Algorithm. From the results of the research that has been done, it is found that, the system that has been made can help students to find out information on the topic of thesis topics that are trending in the computer system study program. For the analysis process using the K-Means Clustering method, the success rate can be 66.66% for the mathematical process. And the system process is 33.33% for the same data as the mathematical process*

***Keywords*** *- Text Mining; K-Means Clustering;*

## 1. PENDAHULUAN

Dalam suatu proses perkuliahan mahasiswa yang sudah menempuh pendidikan cukup lama dan akan menyelesaikannya, maka harus melalui tahapan yang harus dilalui semua mahasiswa jika ingin lulus dari universitas tertentu. Dengan banyaknya mahasiswa yang lulus pada setiap tahunnya sehingga sulitnya mencari informasi tentang tren topik skripsi yang ada di jurusan. Banyaknya karya ilmiah yang berbentuk dokumen cetak atau digital. Tercatat dari tahun 2004 hingga tahun 2017, sudah terdapat 508 dokumen tugas akhir yang ada di perpustakaan Prodi Sistem Komputer. Karena banyaknya dokumen tersebut mengakibatkan sulitnya mendapatkan informasi tentang topik skripsi apa saja yang sedang populer pada tahun-tahun tertentu.

Dari permasalahan di atas penulis mengajukan sebuah penelitian untuk membuat suatu aplikasi yang dapat membantu mahasiswa-mahasiswi yang akan mengerjakan tugas akhir agar dapat mengetahui tentang tren topik skripsi apa saja yang sedang populer di Prodi Sistem Komputer. Karena dengan adanya aplikasi ini dapat memudahkan mahasiswa yang akan mengerjakan tugas akhir melihat informasi tren topik skripsi apa saja yang tren pada tahun-tahun yang lalu. Sehingga dapat menjadi referensi atau ide untuk penulisan tugas akhir yang baru.

Dengan dibuatnya aplikasi ini harapannya nanti mahasiswa yang akan mengerjakan tugas akhir dapat mempunyai gambaran tentang topik skripsi apa saja yang belum dikerjakan atau dapat mengembangkan topik-topik yang sudah ada sebelumnya. Serta aplikasi ini dapat membantu kepala program studi, dan dosen untuk menganalisis dengan cepat tentang tren topik skripsi di tahun-tahun tertentu yang diinginkan.

## 2. METODE PENELITIAN

### 2.1. Text Preprosessing

Dalam *text mining*, informasi yang akan digali strukturnya tidak beraturan. Sehingga dibutuhkan proses perubahan bentuk menjadi data yang teratur sesuai dengan kebutuhan. Berikut adalah beberapa tahan yang dilakukan *Text Preprocessing*:

1) *Case Folding*
   *Case folding* merupakan proses mengubah seluruh huruf yang ada dalam text, yang awalnya huruf kapital menjadi huruf kecil. Agar nantinya lebih mudah dilanjutkan ke proses berikutnya.
2) *Tokenizing*
   Merupakan tahap pemotongan kata dari kata-kata yang menyusunnya menjadi suatu urutan list. Di tahap ini juga menghilangkan beberapa karakter yang dianggap sebagai tanda baca seperti, tanda titik, koma, tanda seru, angka dan sebagainya.
3) *Filtering*
   Merupakan tahap menghilangkan kata-kata yang tidak berhubungan seperti kata sambung dengan memanfaatkan algoritma *Stopword Removal*.
4) *Stemming*
   *Stemming* merupakan tahap mencari kata dasar dari kata yang telah di filter pada tahap *filtering*. Dengan cara menghilangkan imbuhan pada suatu kata. Pada tahap ini juga mengembalikan bentuk kata kedalam satu representasi yang sama.

### 2.2. Analyzing TF-IDF

Analyzing TF-IDF (*Term Frequency Invers Document Frequency*) adalah metode yang digunakan untuk mengetahui keterhubungan setiap kata (*term*) yang terhadap dalam dokumen dengan memberikan bobot pada setiap term.

Dalam perhitungan bobot menggunakan TF-IDF, hitung jumlah nilai TF kata dengan bobot masing-masing kata. Sedangkan nilai IDF di rumuskan pada persamaan berikut:

$$IDF\ (word) = \log \frac{D}{df}$$

Keterangan:

IDF (*word*) : Nilai IDF dari setiap kata.

D : Total dokumen.

*df* : Total kemunculan kata di semua dokumen.

Adapun persamaan yang digunakan untuk menghitung bobot (*W*) pada masing-masing dokumen terhadap kunci, yaitu[1]:

$$Wdt = tfdt * IDFt$$

Dimana;

W*dt* : Bobot dokumen ke-d terhadap kata ke-t.

Tfdt : Banyak term yang dicari pada seluruh dokumen.

IDFt : Invers Dokumen Frekuensi.

### 2.3. Cosine Similarity

Model ruang vektor dan pembobotan TF-IDF digunakan untuk mempresentasikan nilai dari dokumen sehingga kemudian dapat dihitung kesamaan antar dokumen. kesamaan antar dokumen dihitung menggunakan satuan fungsi ukuran kemiripan. Semakin besar hasil fungsi *similarity*, maka kedua objek yang dievaluasi semakin mirip, demikian pula sebaliknya. Ukuran ini memungkinkan perangkingan dokumen dengan yang sama terhadap *query*. *Cosine Similarity* menggunakan formula berikut:

$$cosSom\ (d_j, q_k) = \frac{\sum_{i=1}^{n}(td_{ij}\ x\ tq_{ik})}{\sqrt{\sum_{i=1}^{n} td_{ij^2}\ x\ \sum_{i=1}^{n} tq_{ik^2}}}$$

Berikut merupakan tahap-tahap dalam perhitungan yang terdapat pada *Cosine Similarity*[1]:

1) Tentukan setiap *query*, yaitu *query* dari jawaban (D), *query* dari *key* jawaban (Q) dan gabungan keduanya.
2) Setiap *query* akan dihilangkan simbol-simbol yang tidak mempengaruhi perhitungan, seperti kata titik, tanda koma, tanda seru, dan sebagainya.
3) Setiap *query* akan dihilangkan kata-kata sambung umum yang lazim digunakan dalam suatu *query*, seperti "dan", "jika", "namun", dan sebagainya.
4) Hitung nilai *term frequency query* jawaban dan *query key* jawaban terhadap *queries*. Jadi perhitungan term di *query* jawaban dan *query* jawaban menuju pada *term* yang terdapat pada *queries*.
5) Hitung total *document frequency* (n) atau banyaknya file (N) yang dimiliki suatu *term* untuk setiap *term* dalam *queries*.

6) Hitung *invers document frequency* dengan rumus berikut:

$$\log\left(\frac{N}{n}\right) + 1$$

7) Kalikan nilai *term frequencu* dengan nilai *invers document frequency* tiap *term* dalam Q ataupun D.
8) Hitung hasil perkalian skalar dari setiap *query* jawaban terhadap *query key* jawaban. Kemudian hasil perkalian jawaban dengan *query* dijumlahkan. (sesuai pada rumus diatas).
9) Hitung perkalian vektor tiap query key jawaban dan query jawaban.
10) Hitung nilai cosine similarity (nilai vektor beda antara D terhadap Q) dengan rumus:

$$sim(d,q) = \sum_{k=1}^{i} (weight_{ik}\ weight_{qk}) / \sqrt{\sum_{k=1}^{i} (weight_{ik}{}^2 weight_{qk}{}^2}$$

### 2.4. K-Means Clustering

K-Means Clustering merupakan salah satu kategori pengelompokan data yang berusaha menggabungkan data ke dalam bentuk satu kelompok atau lebih kelompok. Sehingga data yang memiliki karakteristik yang sama akan dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda akan dikelompokkan ke dalam cluster yang lain. Berikut adalah tahapan menentukan clustering dengan metode K-Means[2]:

1) Tentukan jumlah kelompok *k*.
2) Bangkitkan *k* yang ingin dibentuk.
3) Setiap kelompok, tentukan pusat kelompok yang terdekat.
4) Update data lokasi setiap pusat kelompok dengan nilai *centroid* yang baru.
5) Kembali ke langkah 3 – 5 sampai tidak data yang berpindah kelompok.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Text Preprosessing

Pada tahapan ini akan langsung melakukan proses yang terjadi pada text mining. Mulai dari *case folding* hingga *stemming*.

### 3.2. Pembobotan

Pembobotan terhadap kata dengan menggunakan metode TF-IDF. Proses pertama dari TF-IDF adalah mencari nilai *term* dari setiap dokumen, seperti pada tabel 3-1.

*Table 3-1 Term Frequency*

| Term | D1 | D2 | D3 | … | D10 |
|---|---|---|---|---|---|
| jarak | 5 | 0 | 0 | … | 0 |
| saluran | 2 | 2 | 0 | … | 0 |
| telepon | 3 | 11 | 0 | … | 0 |
| medium | 1 | 0 | 0 | … | 0 |
| transmisi | 1 | 0 | 0 | … | 0 |
| informasi | 1 | 0 | 0 | … | 6 |
| perintah | 2 | 0 | 0 | … | 0 |
| sistem | 5 | 0 | 0 | … | 0 |
| … | … | … | … | … | … |
| simetris | 0 | 0 | 0 | … | 1 |

Setelah didapatkan hasil diatas, maka langkah selanjutnya menghitung dokumen frekuensi dengan menggunakan persamaan (1), sehingga hasilnya seperti pada tabel 3-2.

*Table 3-2 Dokumen Frekuensi*

| Term | Df | D/df |
|---|---|---|
| jarak | 2 | 5 |
| saluran | 2 | 5 |
| telepon | 2 | 5 |
| medium | 1 | 10 |
| transmisi | 2 | 5 |
| informasi | 4 | 2.5 |
| perintah | 2 | 5 |
| sistem | 7 | 1.429 |
| . . . | . . . | . . . |
| sistematis | 1 | 10 |

Berikutnya adalah perhitungan bobot seperti pada tabel 3-3.

*Table 3-3 Perhitungan bobot*

| Term | Idf | W = tf*idf | | | |
|---|---|---|---|---|---|
| | | D1 | D2 | … | D10 |
| jarak | 0.699 | 3.495 | 0 | . . . | 0 |
| saluran | 0.699 | 1.398 | 1.398 | . . . | 0 |
| telepon | 0.699 | 2.097 | 7.689 | . . . | 0 |
| medium | 1 | 1 | 0 | . . . | 0 |
| transmisi | 0.699 | 0.699 | 0 | . . . | 0 |
| informasi | 0.398 | 0.398 | 0 | . . . | 2.388 |
| perintah | 0.699 | 1.398 | 0 | . . . | 0 |
| sistem | 0.155 | 1.084 | 0.775 | . . . | 0 |
| . . . | . . . | . . . | . . . | . . . | . . . |
| simetris | 1 | 0 | 0 | . . . | 1 |
| Nilai Bobot D = | | 40.260 | 34.340 | . . . | 59.717 |

### 3.3. Cosine Similarity

Langkah-langkah dalam perhitungan *Cosine Similarity* sebagai berikut[1]:

1) Tentukan nilai Q (Data Testing).

Tabel 3-4, merupakan tabel dari Q (Data Testing).

*Table 3-4 Menentukan Nilai Q*

| Term | Q | D1 | D2 | . . . | D10 |
|---|---|---|---|---|---|
| jarak | 0 | 5 | 0 | . . . | 0 |
| saluran | 0 | 2 | 2 | . . . | 0 |
| telepon | 0 | 3 | 11 | . . . | 0 |
| medium | 0 | 1 | 0 | . . . | 0 |
| transmisi | 2 | 1 | 0 | . . . | 0 |
| informasi | 0 | 1 | 0 | . . . | 6 |
| perintah | 0 | 2 | 0 | . . . | 0 |
| sistem | 0 | 5 | 0 | . . . | 0 |
| . . . | . . . | . . . | . . . | . . . | . . . |
| simetris | 0 | 0 | 0 | . . . | 1 |

2) Pembobotan dokumen testing.

Tabel 3-5, merupakan hasil dari pembobotan dari dokumen testing.

*Table 3-5 Pembobotan dokument testing*

| Term | Q | D1 | D2 | . . . | D10 |
|---|---|---|---|---|---|
| jarak | 0 | 3.495 | 0 | . . . | 0 |
| saluran | 0 | 1.398 | 1.398 | . . . | 0 |
| telepon | 0 | 2.097 | 7.689 | . . . | 0 |
| medium | 0 | 1 | 0 | . . . | 0 |
| transmisi | 1.398 | 0.699 | 0 | . . . | 0 |
| informasi | 0 | 0.398 | 0 | . . . | 2.388 |
| perintah | 0 | 1.398 | 0 | . . . | 0 |
| sistem | 0.775 | 1.084 | 0.775 | . . . | 0 |
| . . . | . . . | . . . | . . . | . . . | . . . |
| simetris | 0 | 0 | 0 | . . . | 1 |

3) Perkalian skalar tiap D terhadap Q

Tabel 3-6, dibawah ini merupakan hasil dari perkalian skalar tiap D terhadap Q.

*Table 3-6 Perkalian skalar*

| Term | Q | D1 | D2 | . . . | D10 |
|---|---|---|---|---|---|
| jarak | 0 | 0 | 0 | . . . | 0 |
| saluran | 0 | 0 | 0 | . . . | 0 |
| telepon | 0 | 0 | 0 | . . . | 0 |
| medium | 0 | 0 | 0 | . . . | 0 |
| transmisi | 1.398 | 0.977 | 0 | . . . | 0 |
| informasi | 0 | 0 | 0 | . . . | 0 |
| perintah | 0 | 0 | 0 | . . . | 0 |
| sistem | 0.775 | 0.840 | 0.600 | . . . | 0 |
| . . . | . . . | . . . | . . . | . . . | . . . |
| simetris | 0 | 0 | 0 | . . . | 0 |

4) Perkalian Vektor

Tabel 3-7, merupakan hasil dari perkalian vektor.

*Table 3-7 Perkalian Vektor*

| Term | Q | D1 | D2 | . . . | D10 |
|---|---|---|---|---|---|
| jarak | 0 | 25 | 0 | . . . | 0 |
| saluran | 0 | 4 | 4 | . . . | 0 |
| telepon | 0 | 9 | 121 | . . . | 0 |
| medium | 0 | 1 | 0 | . . . | 0 |
| transmisi | 4 | 1 | 0 | . . . | 0 |
| informasi | 0 | 1 | 0 | . . . | 36 |
| perintah | 0 | 4 | 0 | . . . | 0 |
| sistem | 25 | 49 | 25 | . . . | 0 |
| . . . | . . . | . . . | . . . | . . . | . . . |
| simetris | 0 | 0 | 0 | . . . | 0 |
| Jumlah | 220 | 186 | 186 | . . . | 211 |
| Panang Vektor | 14.832 | 13.638 | 13.638 | . . . | 14.526 |

5) Nilai Cosine Similarity

Tabel 3-8 dan tabel 3-9, merupakan data dari hasil *Cosine Similarity* dan kemudian diurutkan tingkat kemiripannya.

*Table 3-8 Cosine Similarity*

| D1 | D2 | D3 | D4 | . . . | D10 |
|---|---|---|---|---|---|
| 1.83% | 0.19% | 0.50% | 0.35% | . . . | 2.42% |

*Table 3-9 Tingkat kemiripan.*

| D9 | D10 | D1 | D7 | . . . | D2 |
|---|---|---|---|---|---|
| 3.35% | 2.42% | 1.82% | 1.39% | . . . | 0.19% |

### 3.4. K-Means Clsutering

Selanjutnya adalah analisa dengan menggunakan *K-Means Clustering*. Sebelum analisa dilakukan tentukan dulu jumlah *K* yang ingin dibangkitkan. Di sini jumlah *K* yang dibangkitkan berjumlah dua. Yaitu, Kontrol dan komputasi. Penerapan *K-Means Clustering* dapat dilakukan dengan prosedur sebagai berikut[3]:

a) Siapkan data training yang mana dalam penulisan ini menggunakan data training dari nilai tingkat kemiripan pada tabel 3-9.
b) Tentukan nilai *K* (*K* = Jumlah Cluster).
c) Tentukan nilai awal centroid, untuk centroid 1 adalah 0,35%, dan untuk nilai centroid 2 adalah 2,42%.
d) Hitung jarak antara data dan centroid menggunakan rumus *Euclidean Distance*.

$$D(p,c)_n = \sqrt[2]{\sum_{i=0}^{n}(p_i - c_i)^2|}$$

Dimana:
*p* = data.
*c* = centroid.

n = jumlah data.
i = iterasi.

*Table 3-10 Jarak ke Centroid*

| Data | Tingkat Kemiripan | C1 | C2 |
|---|---|---|---|
| D9 | 3,35% | 3% | 0,93% |
| D10 | 2,42% | 2,07% | 0 |
| D1 | 1,82% | 1,47% | 0,6% |
| D7 | 1,39% | 1,04% | 1,03% |
| D6 | 1,32% | 0,97% | 1,1% |
| D8 | 1,19% | 0,84% | 1,23% |
| D3 | 0,50% | 0,15% | 1,92% |
| D4 | 0,35% | 0 | 2,07% |
| D2 | 0,19% | 0,16% | 2,23% |

e) Partisi data berdasarkan nilai minimum. Nilai yang diurutkan berdasarkan nilai yang paling kecil atau nilai yang paling dekat dengan centroid. Seperti pada tabel 3-11.

*Table 3-11 Pengelompokan nilai*

| Data | Tingkat Kemiripan | C1 | C2 | Jarak | Cluster |
|---|---|---|---|---|---|
| D9 | 3,35% | 3% | 0,93% | 0,93% | C2 |
| D10 | 2,42% | 2,07% | 0 | 0 | C2 |
| D1 | 1,82% | 1,47% | 0,6% | 0,6% | C2 |
| D7 | 1,39% | 1,04% | 1,03% | 1,03% | C2 |
| D6 | 1,32% | 0,97% | 1,1% | 0,97% | C1 |
| D8 | 1,19% | 0,84% | 1,23% | 0,84% | C1 |
| D3 | 0,50% | 0,15% | 1,92% | 0,15% | C1 |
| D4 | 0,35% | 0 | 2,07% | 0 | C1 |
| D2 | 0,19% | 0,16% | 2,23% | 0,16% | C1 |

Dari tabel 3-11, maka didapat pengelompokan data untuk mengetahui data tersebut masuk ke dalam kelompok yang mana. Seperti pada tabel 3-12.

*Table 3-12 Hasil Bidang keahlian*

| Data | Nilai Terdekat | Cluster | Hasil | Keterangan |
|---|---|---|---|---|
| D9 | 0,93% | C2 | Komputasi | Sesuai |
| D10 | 0 | C2 | Komputasi | Sesuai |
| D1 | 0,6% | C2 | Komputasi | Tidak Sesuai |
| D7 | 1,03% | C2 | Komputasi | Sesuai |
| D6 | 0,97% | C1 | Kontrol | Tidak Sesuai |
| D8 | 0,84% | C1 | Kontrol | Tidak Sesuai |
| D3 | 0,15% | C1 | Kontrol | Sesuai |
| D4 | 0 | C1 | Kontrol | Sesuai |
| D2 | 0,16% | C1 | Kontrol | Sesuai |

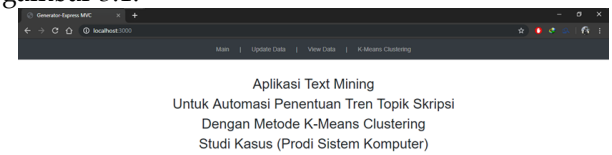
Berdasarkan tabel 3-12, terdapat 6 dokumen yang sesuai dan 3 data abstrak yang tidak sesuai, sehingga didapatkan perhitungannya sebagai berikut:

$$Accuracy = \frac{6}{9} \; x \; 100\% \; = 66{,}66\%$$

Jadi nilai keakurasian dari perhitungan matematis *K-Means Clustering* adalah sebesar 66.66%.

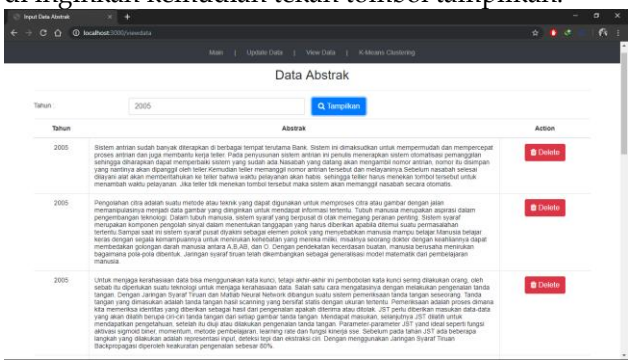
### 3.5. Pengujian

Pengguna dapat mengakses aplikasi text mining untuk automasi penentuan tren topik skripsi dengan metode k-means clustering melalui alamat url berikut https://web-app1-heroku.herokuapp.com. Yang nantinya akan diteruskan ke halaman awal dari aplikasi tersebut, dapat dilihat seperti pada gambar 3.1.



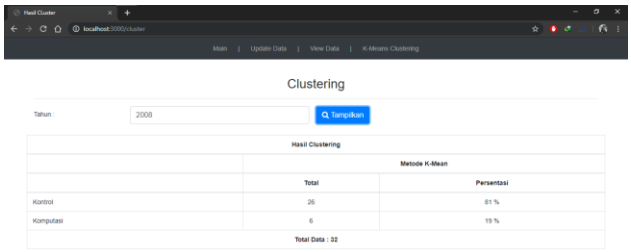

*Gambar 3.1 Halaman Awal*

Untuk melihat data abstrak yang sudah di input dapat di lihat melalui menu view abstrak, seperti pada gambar 3.2. Dengan cara memilih tahun yang di inginkan kemudian tekan tombol tampilkan.




*Gambar 3.2 Menu View Data*

Kemudian untuk melihat hasil dari proses cluster dapat memilih menu K-Means Clustering, seperti pada gambar 3.3. Pada menu K-Means Clustering, pengguna dapat memilih tahun yang di inginkan dan kemudian tekan tampilkan. Maka sistem akan menampilkan presentasi dari tren di setiap tahunnya. Di aplikasi ini yang di maksud tren adalah bidang keminatan dari mahasiswa yang sudah lulus di tahun sebelumnya. Jadi untuk bidang keminatannya adalah komputasi atau kontrol.




*Gambar 3.3 Menu Clustering*

Adapun hasil dari perbandingan perhitungan matematis dengan sistem dapat dilihat seperti pada tabel 3.13. Dari tabel 3.13 tersebut didapatkan hasil yang berbeda antara perhitungan matematis dengan sistem yang telah dibuat. Hal ini terjadi dikarenakan kata-kata yang seharusnya di masukkan ke dalam kelompok yang sama, namun terjadi kesalahan dalam sistem. Sistem malah dimasukkan ke kelompok yang berbeda, sehingga hasil dari pembacaan sistem kurang sesuai.

Untuk mengetahui apakah aplikasi sudah sesuai dengan tujuannya yaitu membantu memberikan informasi tentang tren topik skripsi kepada mahasiswa yang ada di program studi sistem komputer, maka di buatlah kuesioner. Dari hasil kuesioner tersebut di dapatkan 56.6% responden merasa aplikasi sudah cukup membantu untuk memberikan informasi tentang tren topik skripsi.

*Tabel 3.13 Hasil Perbandingan dari matematis dengan sistem*

| Data | Hasil Matematis | Hasil Dari Sistem | Keterangan |
|---|---|---|---|
| D9 | Komputasi | Kontrol | Tidak Sesuai |
| D10 | Komputasi | Kontrol | Tidak Sesuai |
| D1 | Komputasi | Komputasi | Tidak Sesuai |
| D7 | Komputasi | Komputasi | Sesuai |
| D6 | Kontrol | Komputasi | Sesuai |
| D8 | Kontrol | Komputasi | Sesuai |
| D3 | Kontrol | Komputasi | Tidak Sesuai |
| D4 | Kontrol | Komputasi | Tidak Sesuai |
| D2 | Kontrol | Komputasi | Tidak Sesuai |

## 4. KESIMPULAN

### 4.1. Kesimpulan

Berdasarkan hasil dari pengujian dan analisa yang dilakukan maka dapat disimpulkan:

1. Sistem dapat berjalan sebagaimana mestinya, meskipun terdapat beberapa perbedaan antara perhitungan matematis dengan hasil sistem.
2. Dari hasil responden para pengguna aplikasi, bahwa aplikasi sudah cukup membantu dalam memberikan informasi tentang tren topik skripsi yang ada di program studi sistem komputer.

### 4.2. Saran

Adapun saran untuk penulisan kedepannya sebagai berikut:

1. Perbaikan terhadap tampilan agar lebih mudah dalam pengoperasiannya.
2. Adapun untuk sistem dapat di tingkatkan akurasinya dalam proses penentuan tren topik skripsinya.

## DAFTAR PUSTAKA

[1] H. Irmayanti, "Analisis Algoritma Fuzzy Logic dalam Pengklasifikasian Tugas Akhir Analysis of Fuzzy Logic Algorithm in Final Task Classification," vol. 7, no. 2, pp. 71–77, 2018.

[2] N. Wakhidah, "Clustering Menggunakan K-Means Algorithm ( K-Means Algorithm Clustering )," *Fak. Teknol. Inf.*, vol. 21, no. 1, pp. 70–80, 2014.

[3] "Clustering Kmeans Review : Definisi Clustering."

[4] K. R. Prilianti and H. Wijaya, "Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering," J. Cybermatika, vol. 2, no. 1, pp. 1–6, 2014.

[5] R. Feldman and J. Sanger, The Text Mining Handbook. 2006.

[6] "Clustering Kmeans Review : Definisi Clustering."

[7] R. Lynda, S. S. Widya, and S. Esti, "ANALISA CLUSTERING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING ( STUDI KASUS : DOKUMEN SKRIPSI JURUSAN KIMIA , FMIPA , 2 . 3 Term Weighting dengan Term Frequency," vol. Volume 3 N, 2014.

[8] S. T. Safitri and D. Supriyadi, "Rancang Bangun Sistem Informasi Praktek Kerja Lapangan Berbasis Web dengan Metode Waterfall," J. INFOTEL - Inform. Telekomun. Elektron., vol. 7, no. 1, p. 69, 2015.

[9] Suendri, "Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem (Studi Kasus : UIN Sumatera Utara Medan)," J. Ilmu Komput. dan Inform., vol. 3, no. 1, pp. 1–9, 2018

[10] I. P. E.- Issn, "Computer Based Information System Journal PERBANDINGAN PERFORMANSI DATABASE MONGODB DAN MYSQL DALAM APLIKASI FILE MULTIMEDIA BERBASIS WEB Mesri Silalahi , Didi Wahyudi," CBIS J., vol. 01, pp. 63–78, 2018.