

# Penerapan Algoritma *Cosine Similarity* dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi

Rizki Tri Wahyuni<sup>1</sup>, Dhidik Prastiyanto<sup>2</sup>, dan Eko Suprptono<sup>3</sup>

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang

Kampus Sekaran, Gunungpati, Semarang, 50229, Indonesia

zzyraichie30@gmail.com<sup>1</sup>, dhidik.prastiyanto@mail.unnes.ac.id<sup>2</sup>, ekostono@yahoo.com<sup>3</sup>

**Abstrak**— Banyaknya arsip dokumen skripsi yang terkumpul dalam bentuk *soft file* yang tidak terklasifikasi dengan baik mengakibatkan proses pencarian kembali menjadi sulit. Untuk mengakses informasi yang dibutuhkan menjadi kurang cepat dan tepat apabila keseluruhan dokumen disimpan dalam satu *folder database*. Maka dari itu diperlukan suatu sistem yang dapat mengklasifikasikan dokumen secara otomatis ke dalam *folder* berbeda pada *database* agar lebih mudah dalam mengelola dokumen yang ada. Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode *cosine similarity* merupakan metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah *vector* dengan menggunakan *keywords* (kata kunci) dari sebuah dokumen sebagai ukuran. Metode pengembangan sistem yang digunakan dalam penelitian ini adalah model *waterfall*, sedangkan metode penelitian yang digunakan adalah metode *Research and Development (R&D)*. Data latih yang digunakan dalam penelitian ini berjumlah 50 dokumen skripsi dengan beberapa kategori yang berbeda. Hasil penelitian menunjukkan bahwa persentase tingkat ketepatan klasifikasi sistem adalah sebesar 98%.

**Kata kunci**— klasifikasi dokumen, *cosine similarity*, TF-IDF, *vector space*

## I. PENDAHULUAN

Perkembangan teknologi memiliki dampak yang sangat signifikan dalam kehidupan sehari-hari, mulai dari kegiatan yang sederhana hingga kegiatan yang membutuhkan tingkat ketelitian yang tinggi. Kegiatan yang umum dilakukan oleh sebuah instansi adalah kegiatan pengarsipan dokumen, baik dokumen dalam bentuk fisik maupun elektronik. Umumnya kegiatan pengarsipan melibatkan dokumen dengan jumlah yang cukup besar, sehingga diperlukan suatu metode yang praktis dan efisien dalam pengelolannya. Salah satu metode yang digunakan dalam pengelolaan dokumen adalah pengklasteran atau pengklasifikasian dokumen.

Pengklasifikasian dokumen elektronik dengan jumlah yang banyak diperlukan agar data yang terkumpul dapat diproses menjadi informasi yang tepat. Pengklasifikasian dokumen dilakukan dalam upaya memisahkan atau mengelompokkan dokumen berdasarkan ciri-ciri atau kategori tertentu. Dengan banyaknya dokumen proses pengklasifikasian tidak mungkin dilakukan secara manual karena memerlukan banyak waktu dan tenaga. Salah satu metode yang dapat digunakan adalah dengan pengklasifikasian secara otomatis dengan *text mining*. Banyak metode *text mining* yang digunakan dalam mengklasifikasikan dokumen atau teks, salah satunya adalah algoritma *cosine similarity*.

Permasalahan yang dihadapi pada *text mining* adalah jumlah data yang besar, dimensi yang tinggi, data dan struktur yang terus berubah, serta *data noise*. Sehingga sumber data yang digunakan pada *text mining* adalah kumpulan teks yang

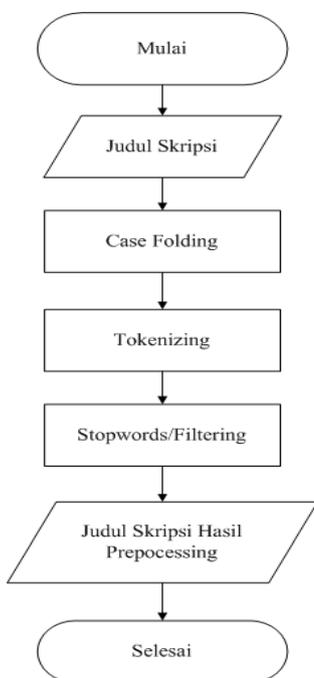
memiliki bentuk yang tidak terstruktur atau setidaknya semi terstruktur [1]. Klastering biasa digunakan pada banyak bidang, seperti *text mining*, pengenalan pola (*pattern recognition*), pengklasifikasian gambar (*image classification*), ilmu biologi, pemasaran, perencanaan kota, pencarian dokumen, dan lain sebagainya. Tujuan dari klastering adalah untuk menentukan pengelompokan dari suatu set data. Akan tetapi tidak ada "ukuran terbaik" untuk pengelompokan data. Untuk pengelompokan data tergantung tujuan akhir dari klastering, maka diperlukan suatu kriteria sehingga hasil klastering seperti yang diinginkan [2].

Pengklasifikasian teks sangat dibutuhkan dalam berbagai macam aplikasi, terutama aplikasi yang jumlah dokumennya bertambah dengan cepat. Ada dua cara dalam penggolongan teks, yaitu *text clustering* dan klasifikasi teks. *Text clustering* berhubungan dengan menemukan sebuah struktur kelompok yang belum kelihatan (tak terpandu atau *unsupervised*) dari sekumpulan dokumen. Sedangkan pengklasifikasian teks dapat dianggap sebagai proses untuk membentuk golongan golongan (kelas-kelas) dari dokumen berdasarkan pada kelas kelompok yang sudah diketahui sebelumnya (terpandu atau *supervised*) [3]. Metode *cosine*, *jaccard*, dan *k-nearest neighbor (K-NN)* yang digunakan pada proses klasifikasi dokumen teks dengan hasil akhir dari percobaan 33 kali dengan *key* yang berbeda dan total 6326 dokumen didapat metode *cosine* yang nilai kemiripannya tertinggi yaitu 41% dari metode *jaccard* 19% dan *k-nearest neighbor (K-NN)* 40%, karena metode *cosine similarity* mempunyai konsep normalisasi panjang *vektor* data dengan membandingkan *N-gram* yang sejajar satu sama lain dari 2 perbandingan [4].

Tujuan penelitian ini adalah membuat sebuah sistem yang dapat mengklasifikasikan dokumen secara otomatis menggunakan algoritma *cosine similarity* dalam proses pengklasterannya dan menggunakan metode pembobotan TF-IDF. Objek penelitian ini adalah dokumen skripsi dalam bentuk elektronik. Dokumen skripsi akan diklasifikasikan ke dalam beberapa kategori secara otomatis. Sehingga diharapkan sistem yang dihasilkan pada penelitian ini dapat membantu meringankan kegiatan pengarsipan dokumen.

## II. METODE PENELITIAN

Dalam *text mining*, data teks akan diproses menjadi data *numerik* agar dapat dilakukan proses lebih lanjut. Sehingga dalam *text mining* ada istilah *preprocessing* data, yaitu proses pendahulu yang diterapkan terhadap data teks yang bertujuan untuk menghasilkan data *numerik*. Tahap *preprocessing* dapat dilihat pada Gambar 1.



Gambar 1. Flowchart preprocessing

Tahap *preprocessing* yang digunakan dalam penelitian ini antara lain :

### 1. Case Folding

Merupakan tahap perubahan huruf dari huruf kapital menjadi huruf kecil.

### 2. Tokenizing

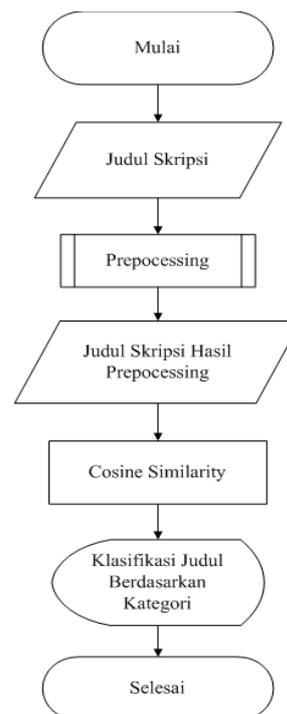
*Tokenizing* adalah proses memecah dokumen menjadi kumpulan kata. *Tokenization* dapat dilakukan dengan menghilangkan tanda baca dan memisahkannya per spasi. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua *token* ke bentuk huruf kecil (*lower case*) [5].

### 3. Stopwords/Filtering

*Stopwords removal* merupakan proses penghilangan kata tidak penting pada deskripsi melalui pengecekan kata-kata hasil *parsing* deskripsi apakah termasuk di dalam daftar

kata tidak penting (*stoplist*) atau tidak. Jika termasuk di dalam *stoplist* maka kata-kata tersebut akan di-*remove* dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata penting atau *keywords* [4].

Setelah melalui tahap *preprocessing* maka data akan disimpan dalam memori sementara dan nantinya akan diproses lebih lanjut ke dalam tahap *analyzing* menggunakan pembobotan TF-IDF dan klasifikasi dengan algoritma *cosine similarity*. Untuk lebih jelasnya proses klasifikasi dokumen dapat dilihat pada Gambar 2.



Gambar 2. Flowchart proses klasifikasi

Metode TF-IDF menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan *inverse* frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen. Rumus untuk TF-IDF [6]:

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (1)$$

$$idf_t = \log \left( \frac{D}{df_t} \right) \quad (2)$$

$$W_{d,t} = tf_{d,t} \times idf_{d,t} \quad (3)$$

Keterangan:

D = dokumen ke-d

$t$  = *term* ke- $t$  dari dokumen  
 $W$  = bobot dokumen ke- $d$  terhadap *term* ke- $t$   
 $tf$  = banyaknya *term*  $i$  pada sebuah dokumen  
 $idf$  = *Inversed Document Frequency*  
 $df$  = banyak dokumen yang mengandung *term*  $i$

*Cosine Similarity* digunakan untuk melakukan perhitungan kesamaan dari dokumen. Rumus yang digunakan oleh *cosine similarity* adalah [7] :

$$\text{Cos } \alpha = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}| |\mathbf{B}|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (4)$$

Keterangan :

- $\mathbf{A}$  = Vektor A, yang akan dibandingkan kemiripannya
- $\mathbf{B}$  = Vektor B, yang akan dibandingkan kemiripannya
- $\mathbf{A} \cdot \mathbf{B}$  = *dot product* antara vektor A dan vektor B
- $|\mathbf{A}|$  = panjang vektor A
- $|\mathbf{B}|$  = panjang vektor B
- $|\mathbf{A}| |\mathbf{B}|$  = *cross product* antara  $|\mathbf{A}|$  dan  $|\mathbf{B}|$

Metode pengklasifikasian yang digunakan pada sistem ini adalah dengan cara membandingkan kesamaan atau similaritas antara judul dokumen dengan kata kunci pertama, kemudian cara membandingkan kesamaan atau similaritas antara judul dokumen dengan kata kunci kedua, begitu seterusnya hingga kata kunci kedelapan. Kemudian dicari jumlah similaritas yang tertinggi antara kedelapan kata kunci. Apabila total similaritas yang didapatkan adalah nol (0) maka dokumen akan masuk ke dalam kategori kesembilan. Dokumen yang diolah nantinya akan diklasifikasikan secara otomatis oleh sistem ke dalam 9 kategori yang berbeda. Dokumen yang dipilih adalah dokumen dengan kategori keilmuan dalam bidang Teknik Informatika dan Teknik Elektro.

### III. HASIL DAN PEMBAHASAN

#### A. Implementasi Rancangan Sistem

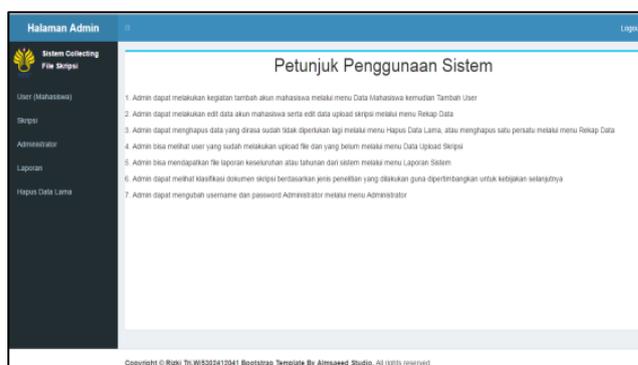
Sistem *Collecting File Skripsi* terdiri atas dua *level* pengguna, yaitu *level administrator* dan *level user*. Halaman *user* terdiri dari beberapa menu diantaranya menu Beranda, Unggah Skripsi, Referensi, Data Upload Personal, dan Ubah Password. Hak akses *user* meliputi mendaftarkan akun, melakukan unggah (*upload*) skripsi, melihat beberapa referensi, mencetak bukti selesai *upload* skripsi, dan mengubah *password* pribadi. Tampilan halaman utama *user* dapat dilihat pada Gambar 3.

Halaman *administrator* terdiri dari beberapa menu diantaranya adalah menu *User* (Mahasiswa), Skripsi, *Administrator*, Laporan, dan Hapus Data Lama. Hak akses *administrator* adalah meliputi segala aspek yang ada dalam sistem, seperti mengelola data mahasiswa dan data skripsi, menambah *user* baru, mengunci akun *user* (mahasiswa),

melakukan validasi *upload*, mengelola data akun *administrator*, mencetak laporan *upload* skripsi, dan menghapus data-data lama yang telah kadaluarsa atau sudah tidak terpakai. Tampilan halaman utama *administrator* dapat dilihat pada Gambar 4.



Gambar 3. Tampilan Beranda *User*



Gambar 4. Tampilan Beranda *Administrator*

#### B. Hasil Pengujian *Black-box*

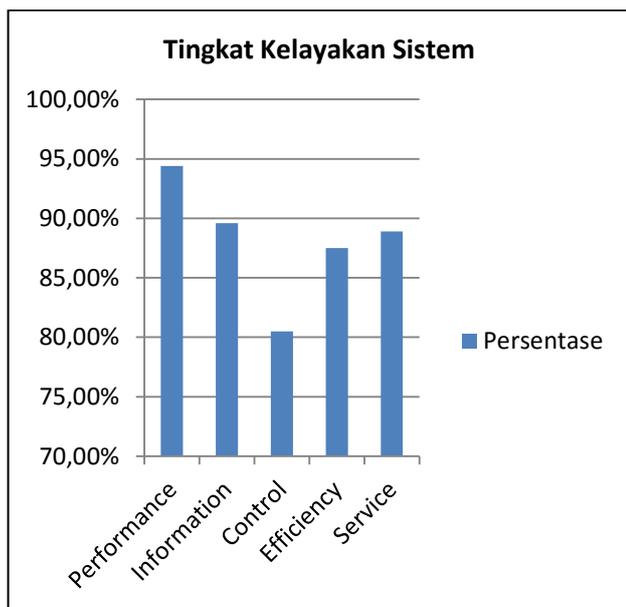
Pengujian *black-box* digunakan untuk melihat keseluruhan fungsi-fungsi dalam sistem, apakah seluruh fungsi sudah berjalan dengan baik atau masih diperlukan beberapa perbaikan. Pengujian dilakukan dengan membuat skenario yang telah disesuaikan dengan komponen sistem yang telah dibuat, kemudian sistem diuji berdasarkan skenario yang telah dibuat. Hasil pengujian *black-box* menunjukkan bahwa keseluruhan fungsi yang ada dalam Sistem *Collecting File Skripsi* telah berjalan dengan baik sehingga tidak lagi diperlukan perbaikan fungsi.

#### C. Hasil Pengujian Pakar Sistem

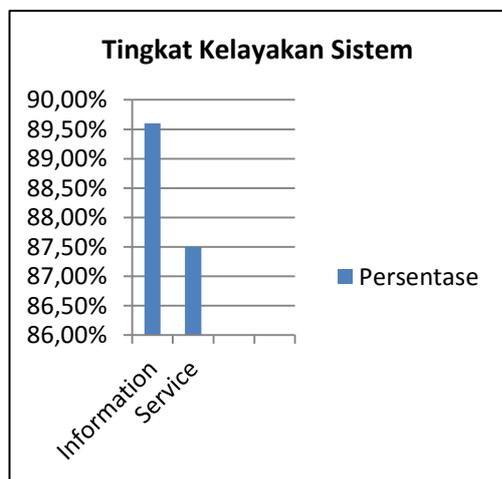
Pengujian ahli sistem dari Sistem *Collecting File Skripsi* dilakukan oleh 3 validator yang merupakan dosen di Jurusan Teknik Elektro Unnes. Validasi ini merupakan validasi untuk semua *level*, yaitu *level user* (mahasiswa) dan *level administrator*. Validasi yang dilakukan meliputi lima aspek, yaitu aspek *performance*, *information*, *control*, *efficiency*, dan *service*. Untuk lebih jelasnya, hasil penilaian oleh validator disajikan pada Gambar 5.

D. Hasil Pengujian Pakar Kearsipan

Pengujian ahli kearsipan dari Sistem *Collecting File Skripsi* dilakukan oleh 2 validator yang merupakan pustakawan dari Jurusan Biologi Unnes serta pustakawan dari UPT Unnes. Validasi ini merupakan validasi untuk semua level, yaitu *level user* (mahasiswa) dan *level administrator*. Validasi yang dilakukan meliputi dua aspek, yaitu aspek *information* dan *service*. Untuk lebih jelasnya, hasil penilaian oleh validator disajikan pada Gambar 6.



Gambar 5. Kelayakan berdasarkan Ahli Sistem



Gambar 6. Kelayakan berdasarkan Ahli Kearsipan

E. Hasil Pengujian Klasifikasi dengan Algoritma *Cosine Similarity*

Pada pengujian dalam tahap ini dilakukan 2 tahap pengujian, yaitu tahap *training* data dan tahap *testing*. Pada tahap *training*, data yang digunakan telah diketahui jenis kategorinya. Tahap *training* digunakan untuk melihat ketepatan klasifikasi dokumen dengan *algoritma cosine similarity*, pada tahap ini data yang digunakan adalah sejumlah 10 dokumen. Dokumen yang dipilih merupakan dokumen dengan kategori dalam bidang keilmuan Teknik Informatika dan Teknik Elektro. Data yang digunakan ditunjukkan pada Tabel I.

Selanjutnya merupakan tahap *testing*, dimana data-data yang diolah belum diketahui kategorinya dan akan mendapatkan kategori secara otomatis dari sistem. Dalam penelitian ini, kategori yang dipakai adalah sejumlah 9 kategori dengan pemakaian kata kunci sedemikian rupa. Tabel II merupakan data hasil tahap *testing*.

TABEL I. TABEL DATA FILE SKRIPSI

No	Judul Dokumen	Kategori
1.	Pengembangan Lampu LED Alternatif sebagai Efisiensi Daya	Ketenagaan
2.	Uji Minyak Trafo di Gardu Induk 150 KV Ungaran dan Implementasinya pada Pembelajaran Mata Pelajaran Pemeliharaan Kelistrikan Kelas XI TITL di SMKN 1 Bumiayu Kabupaten Tegal	Model, metode, dan media pembelajaran
3.	Perancangan dan Pembuatan Alat Ukur Jarak, Suhu dan Waktu Menggunakan Arduino	Instrumentasi dan kendali
4.	Otomatisasi <i>Charger Battery</i> pada Laptop	Elektronika dan komunikasi
5.	Alat Sistem <i>Skoring TB</i> (Tuberkulosis) Anak Diaplikasikan dengan Menggunakan IC Atemega 32	Instrumentasi dan kendali
6.	Perancangan <i>Data Logger</i> Suhu pada Suatu Ruang Melalui SMS	Instrumentasi dan kendali
7.	Simulasi Aliran Daya menggunakan Data Prediksi Beban Puncak 5 Tahunan pada Sistem Tenaga Listrik PT. PLN Rayon Semarang Tengah dengan Etap 7,5	Ketenagaan
8.	Pengembangan Ensiklopedia Elektronika Berbasis Wiki	Sistem berbasis <i>web</i>
9.	Simulasi Generator di Rumah Daya RSI Sultan Agung Semarang	Ketenagaan
10.	Studi Eksplorasi Arus pada Kawat Netral Akibat Ketidakseimbangan Beban pada Unit Transformator Distribusi di Universitas Negeri Semarang	Ketenagaan

TABEL II. TABEL HASIL PENGUJIAN KLASIFIKASI DENGAN ALGORITMA

No Dokumen	Kategori	Keterangan
1	Lainnya	valid
2	Ketenagaan	valid
3	Instrumentasi dan kendali	valid
4	Instrumentasi dan kendali	valid
5	Ketenagaan	valid
6	Elektronika dan komunikasi	valid
7	Instrumentasi dan kendali	valid
8	Instrumentasi dan kendali	Tidak valid
9	Sistem android	valid
10	Instrumentasi dan kendali	valid
11	Ketenagaan	valid
12	Ketenagaan	valid
13	Instrumentasi dan kendali	valid
14	Instrumentasi dan kendali	valid
15	Instrumentasi dan kendali	valid
16	Ketenagaan	valid
17	Sistem android	valid
18	Model, metode, dan media pembelajaran	valid
19	Model, metode, dan media pembelajaran	valid
20	Sistem pakar	valid

#### F. Pembahasan

Dengan adanya Sistem *Collecting File* Skripsi kegiatan pengarsipan yang ada pada Jurusan Teknik Elektro Universitas Negeri Semarang menjadi lebih mudah dan efisien dalam segi waktu. Mahasiswa dapat melakukan kegiatan *collecting* dokumen skripsi hanya dengan mengirimkan dokumen melalui sistem yang telah terintegrasi dengan internet. Jika dibandingkan dengan metode *collecting* pada periode sebelumnya maka kegiatan *collecting* dengan menggunakan Sistem *Collecting File* Skripsi menjadi lebih cepat dan penyimpanan data lebih terpusat dan terjaga. Selain itu sistem juga dapat mengklasifikasikan dokumen secara otomatis sehingga *administrator* tidak perlu lagi untuk melakukan klasifikasi secara manual, dan dirasa sistem ini dapat membantu mempermudah proses pengelolaan arsip dokumen skripsi yang ada di Jurusan Teknik Elektro Universitas Negeri Semarang.

Sistem *Collecting File* Skripsi telah berhasil diimplementasikan berdasarkan rancangan yang telah dibuat pada tahap sebelumnya, kemudian sistem telah melalui beberapa proses pengujian tingkat kelayakan. Beberapa pengujian tersebut antara lain adalah pengujian *black-box*, pengujian tingkat kelayakan dengan beberapa pakar sistem, pengujian tingkat kelayakan dengan beberapa pakar kearsipan, dan pengujian ketepatan klasifikasi yang telah dilakukan oleh sistem. Hasil pengujian *black-box* menunjukkan bahwa sistem telah dapat menjalankan seluruh fungsi-fungsinya dengan baik sehingga tidak lagi diperlukan

perbaikan sistem tahap 1. Hasil pengujian tingkat kelayakan dengan beberapa pakar sistem menunjukkan hasil bahwa sistem sangat layak untuk digunakan dengan persentase rata-rata hasil pengujian sebesar 88,3%. Sedangkan hasil pengujian tingkat kelayakan dengan beberapa pakar kearsipan menunjukkan bahwa sistem juga sangat layak untuk digunakan dengan persentase rata-rata hasil pengujian sebesar 87,5%.

Dari sejumlah 50 data yang telah diolah oleh sistem, didapatkan sejumlah 49 data yang berhasil diklasifikasikan dengan *valid* dan sejumlah 1 data yang tidak *valid*. Maka persentase ketepatan klasifikasi sistem yang didapatkan adalah sebesar 98%. Persentase tersebut didapatkan dari jumlah data yang *valid* dibagi dengan jumlah keseluruhan data yang diolah kemudian dikali dengan 100%. Kesalahan klasifikasi yang sering terjadi adalah karena terdapat beberapa kata yang sama dengan kata kunci, sehingga sistem memilih nilai tertinggi dari perhitungan *cosine similarity* yang ada. Contohnya adalah skripsi dengan judul "Rancang Bangun Aplikasi Streaming Video Memanfaatkan RED5 Media Server", kata "aplikasi" merupakan salah satu kata kunci pada *term* kategori "sistem android". Sedangkan "media" adalah salah satu kata kunci pada *term* kategori "model metode dan media pembelajaran".

Jika dikaji secara manual, judul tersebut tidak masuk ke dalam dua kategori yang telah disebutkan, seharusnya judul tersebut masuk ke dalam kategori "lainnya". Namun hasil perhitungan dengan sistem menunjukkan bahwa kategori "sistem android" memperoleh nilai sebesar 0,235 dan kategori "model metode dan media pembelajaran" memperoleh nilai sebesar 0,136 sehingga dalam perhitungan dengan sistem judul tersebut masuk ke dalam kategori "sistem android". Penggunaan *term* unik sangat berpengaruh pada perhitungan algoritma, sehingga perbendaharaan kata dalam sistem harus diperbanyak.

#### IV. PENUTUP

Setelah melakukan studi literatur, perancangan, analisis, implementasi dan pengujian kelayakan sistem dapat disimpulkan bahwa Sistem *Collecting File* Skripsi telah dapat dioperasikan dengan lancar dan dapat melakukan klasifikasi dokumen secara otomatis. Hasil pengujian kelayakan sistem berdasarkan pengujian ahli sistem menunjukkan hasil rata-rata persentase kelayakan sebesar 88,3% dan masuk dalam kriteria "sangat layak" untuk digunakan. Berdasarkan hasil klasifikasi yang dilakukan oleh sistem, sebanyak 49 dokumen telah berhasil diklasifikasikan ke dalam kategori yang tepat, sedangkan 1 dokumen tidak dapat terklasifikasikan dengan tepat. Persentase rata-rata hasil ketepatan klasifikasi yang dilakukan oleh sistem adalah sebesar 98%, maka dapat disimpulkan bahwa sistem dapat mengklasifikasikan dokumen ke dalam kategori yang sesuai dengan tingkat akurasi yang sangat tinggi.

REFERENSI

- [1] Imbar, V., Radiant, Adelia, Ayub, M., dan Rehatta, A. 2014. Implementasi Cosine Similarity dan Algoritma Smith Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika* Volume 10, Nomor 1.
- [2] Sugiyamta. 2015. Sistem Deteksi Kemiripan Dokumen dengan Algoritma Cosine Similarity dan Single Pass Clustering. *Jurnal Informatika* Volume 7, Nomor 2.
- [3] Susandi, D. dan Sholahudin, U. 2016. Pemanfaatan Vector Space Model pada Penerapan Algoritma Nazief Adriani, KNN dan Fungsi Similarity Cosine untuk Pembobotan IDF dan WIDF pada Prototipe Sistem Klasifikasi Teks Bahasa Indonesia. *Jurnal Teknologi Informasi* Volume 3, Nomor 1.
- [4] Nurdiana, O., Jumadi., dan Nursantika, D. 2016. Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity pada Aplikasi Pencarian Terjemahan Al-Qur'an dalam Bahasa Indonesia. *Jurnal Online Informatika* Volume 1, Nomor 1.
- [5] Kurniawan, A. Solihin, F., dan Hastarita, F. 2014. Perancangan dan Pembuatan Aplikasi Pencarian Informasi Beasiswa dengan Menggunakan Cosine Similarity. *Jurnal SimanteC* Volume 4, Nomor 2.
- [6] Nurjanah, M. Hamdani. dan Astuti, I. Fitri. 2013. Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) untuk Text Mining. *Jurnal Informatika* Volume 8, Nomor 3.
- [7] Ye, J. 2014. Vector Similarity Measures of Simplified Neutrosophic Sets and Their Application in Multicriteria Decision Making. *Internasional Journal of Fuzzy Systems* Volume 16, Nomor 2.