

# PENERAPAN METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN JUDUL SKRIPSI MAHASISWA (STUDI KASUS : STM IK WIDYA CIPTA DHARMA)

Eka Arriyanti<sup>1)</sup>, Siti Lailiyah<sup>2)</sup>, Alvin Kurniawan<sup>3)</sup>

<sup>1,2,3</sup>Teknik Informatika, STM IK Widya Cipta Dharma

<sup>1,2,3</sup>STM IK Widya Cipta Dharma, Samarinda Ulu, Kota Samarinda, 75123, Kalimantan Timur,  
Indonesia

E-mail : ekaarry@wicida.ac.id<sup>1)</sup>, lailiyah@gmail.com<sup>2)</sup>, alvinkurniawan999@gmail.com<sup>3)</sup>

## ABSTRAK

Perkembangan teknologi sudah berkembang begitu pesat, sehingga perkembangan teknologi saat ini menyebabkan terjadinya ledakan jumlah dokumen elektronik/*softcopy* yang akan tersimpan di dalam sebuah penyimpanan perpustakaan kampus.

Sehingga *K-Means* sebagai metode *clustering* berbasis jarak dapat membagi data ke dalam sejumlah cluster dan algoritma ini hanya akan bekerja pada sebuah atribut *numeric*. Algoritma *K-Means* termasuk *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah. Algoritma *K-Means* sangat terkenal karena kemudahan dan kemampuannya untuk mengcluster data yang besar dan data *outlier* dengan sangat cepat. Sehingga dalam penelitian ini akan dilakukan pengelompokan dokumen skripsi dari program studi yang berada di lingkungan STM IK Widya Cipta Dharma dengan menggunakan metode *K-means*.

**Kata kunci:** *K-Means Clustering, pengelompokan judul skripsi, TF-IDF, STM IK Widya Cipta Dharma, sistem informasi perpustakaan.*

## 1. PENDAHULUAN

Perkembangan teknologi sudah berkembang begitu pesat, sehingga perkembangan teknologi saat ini menyebabkan terjadinya ledakan jumlah dokumen elektronik/*softcopy* yang akan tersimpan didalam sebuah penyimpanan perpustakaan kampus. Berbagai karya tulis ilmiah yang ada pada civitas akademika mulai dari skripsi, laporan penulisan ilmiah, laporan kerja praktek dan lain sebagainya sudah tersedia dalam versi digital. Pada umumnya dokumen-dokumen tersebut tidak disertai dengan jumlah informasi atau pengetahuan yang dapat disajikan, sehingga diperlukan suatu cara yang mempermudah serangkaian proses pengelolaan dari seluruh laporan tersebut.

Sehingga *K-Means* sebagai metode *clustering* berbasis jarak dapat membagi data ke dalam sejumlah cluster dan algoritma ini hanya akan bekerja pada sebuah atribut *numeric*. Algoritma *K-Means* termasuk *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah.

Berdasarkan permasalahan diatas, solusi yang didapatkan yaitu dengan mengelompokan

dokumen skripsi ke dalam sebuah sistem. sehingga dapat menjadi acuan bagi kaprodi dalam menerima judul penelitian baru berdasarkan data yang sering muncul berada pada cluster yang keberapa. Begitu juga karena *K-Means* dikenal sebagai metode *clustering* yang efisien, maka *K-Means* menjadi metode yang diperhitungkan dalam melakukan proses *clustering*. Sehingga dalam penelitian ini akan dilakukan pengelompokan dokumen skripsi dari program studi yang berada di lingkungan STM IK Widya Cipta Dharma dengan menggunakan metode *K-means*.

## 2. RUANG LINGKUP PENELITIAN

Dalam penelitian ini permasalahan mencakup rumusan masalah tentang bagaimana menerapkan metode *K-Means Clustering* dalam pengelompokan dokumen skripsi di STM IK Widya Cipta Dharma.

Kemudian batasan masalah yang ada dalam skripsi ini adalah :

1. Data yang akan digunakan adalah data Dokumen Skripsi yang berada di

perpustakaan STMIK Widya Cipta Dharma.

2. Proses perhitungan *K-Means* digunakan setelah proses perhitungan dari *Text Processing* dan *TF-IDF*.
3. Skripsi akan dikelompokkan berdasarkan judul dari dokumen tersebut.

Sehingga tujuan yang ingin dicapai dari penelitian ini adalah :

1. Menerapkan metode *K-Means* dalam pengelompokan dokumen skripsi di STMIK Widya Cipta Dharma.
2. Menguji pengaruh nilai *K* terhadap mengelompokkan dokumen skripsi.

### 3. BAHAN DAN METODE

Adapun bahan dan metode yang digunakan dalam penelitian ini yaitu:

#### 3.1 Data Mining

Data *mining* adalah proses ekstraksi suatu data/pola (sebelumnya tidak diketahui, bersifat implisit, dianggap tidak berguna) menjadi informasi atau pengetahuan dari data yang jumlahnya besar. Data-data yang tidak terpolatidak terstruktur, dianggap “sampah”, dan tidak berguna kemudian diolah (difilter) sehingga data tersebut membentuk pola/pengetahuan baru yang berguna.

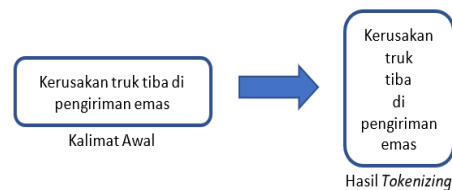
#### 3.2 Langkah Umum Penambangan Data

Langkah umum penambangan data meliputi definisi domain data, pengumpulan data, pemrosesan awal, estimasi model, dan interpretasi hasil.

1. Mendefinisikan domain data; Tujuan data mining memberikan arahan dalam pengumpulan data yang harus ada.
2. Mengumpulkan data; pengumpulan data acak diasumsikan di Sebagian besar aplikasi penambangan data.
3. Pemrosesan Awal; deteksi outlier dan fitur pemilihan pengkodean penskalaan setidaknya merupakan dua tugas umum prapemrosesan.
4. Memperkirakan model; implementasi proses ini didasarkan pada jenis data mining. Memilih model terbaik adalah tugas tambahan.
5. Menafsirkan hasilnya; metode penambangan data modern diharapkan memberikan hasil yang sangat akurat dengan menggunakan model dimensi tinggi.

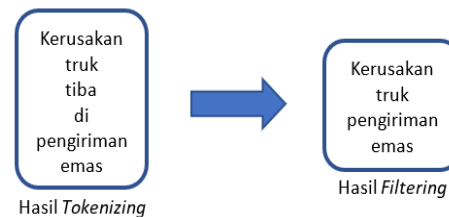
#### 3.3 Text Preprocessing

*Text Preprocessing* adalah dilakukan untuk mengubah data tekstual yang tidak terstruktur kedalam data yang terstruktur dan disimpan dalam basis data, Tujuan dari *preprocessing* yakni menghasilkan sebuah set term index yang bisa mewakili dokumen. Langkah-langkah *text preprocessing* meliputi tokenisasi, *case folding*, *filtering*, *stemming*, dan pembobotan kata.



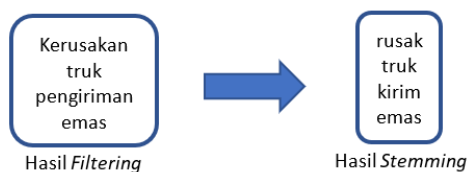
Gambar 1. Tokenisasi

*Tokenizing* adalah proses pemotongan string input berdasarkan tiap kata penyusunannya.



Gambar 2. Filtering

*Filtering* adalah tahap pemilihan kata-kata penting dari hasil token, yaitu kata-kata yang bisa digunakan untuk mewakili isi dari sebuah dokumen.



Gambar 3. Stemming

*Stemming* adalah proses pengubahan bentuk kata menjadi kata dasar atau tahap mencari *root* kata dari setiap kata hasil *filtering*.

#### 3.4 Clustering

*Clustering* adalah teknik data mining yang digunakan untuk mengelompokkan data yang tidak terstruktur yang tidak ada ketentuan data berlabel.

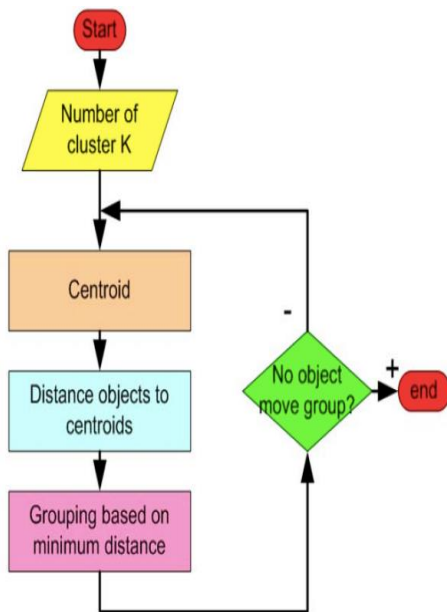
*K-Means Clustering* adalah salah satu algoritma clustering yang paling populer.

### 3.5 Algoritma TF-IDF

Pembobotan *TF/IDF* adalah suatu hubungan kata (*term*) yang berada pada dokumen yang akan diberikan suatu nilai bobot. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu yang disebut *term frequency* (TF) dan inverse frekuensi dokumen yang mengandung kata yang disebut *inverse document frequency* (IDF). *TF-IDF* menghitung frekuensi kata dalam dokumen dan frekuensi kata dalam semua dokumen.

### 3.6 K-Means Clustering

*K-Means Clustering* adalah algoritma yang digunakan untuk mengelompokkan data ke dalam *cluster-cluster* yang berbeda.

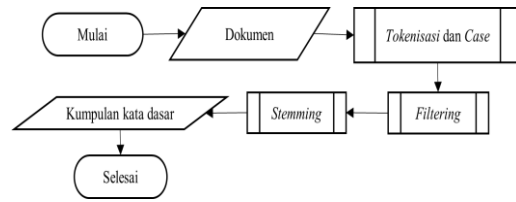


Gambar 4 Flowchart K-Means Clustering

Algoritma ini bekerja dengan cara memilih centroid awal, kemudian menghitung jarak antara data dengan *centroid*, dan akhirnya memperbarui *centroid* berdasarkan jarak terdekat.

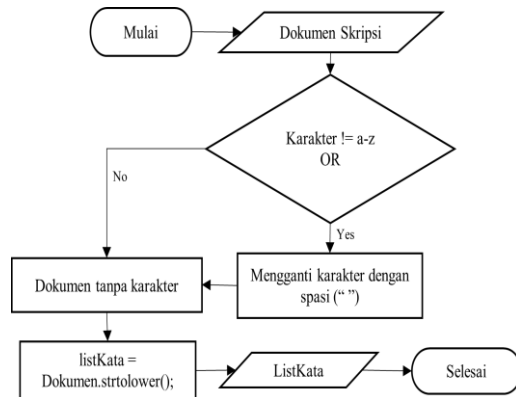
## 4. RANCANGAN SISTEM

*Flowchart* merupakan suatu bagan dengan simbol-simbol tertentu yang menggambarkan urutan proses secara mendetail dan hubungan antara suatu proses dengan proses lainnya. Penjelasan *flowchart* yang menggambarkan sistem dalam penelitian ini adalah gambar berikut.



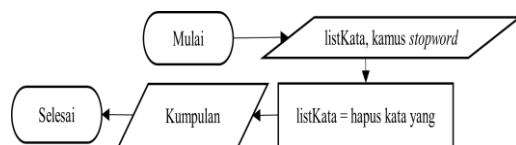
Gambar 5 Alur Proses Preprocessing

Pada tahap alur proses *Preprocessing* merupakan awal dari proses pengolahan data. Karena data yang digunakan merupakan kalimat judul dari dokumen skripsi, maka tahap ini akan sangat dibutuhkan pada pengolahan data selanjutnya. Pada tahapan *preprocessing* ada beberapa sub proses yang akan dilakukan yaitu (tokenisasi, *case folding*, *filtering* dan *stemming*).



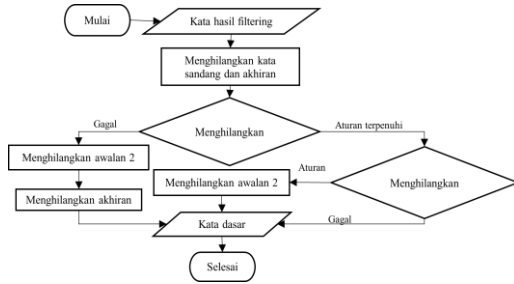
Gambar 6 Alur Proses Tokenisasi dan Case Folding

Tokenisasi adalah sebuah proses yang dilakukan untuk menghilangkan tanda baca, angka dan semua karakter selain *alphabet*. Setelah semua karakter selain *alphabet* dihilangkan maka langkah selanjutnya yaitu melakukan perubahan semua huruf kapital menjadi huruf kecil atau yang biasa disebut *case folding*. Selanjutnya yaitu menyimpan setiap kata dari dokumen yang telah di dapat sebelumnya menjadi kata tunggal (token). Berikut diagram alir dari proses tokenisasi dan *case folding* pada gambar 6.



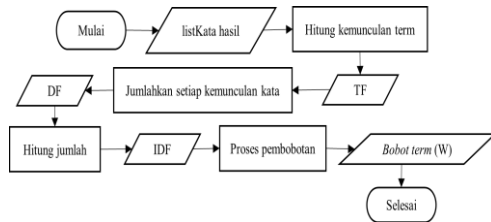
Gambar 7 Alur Proses Filtering

*Filtering* yaitu membuang atau menyaring kata-kata yang dapat merepresentasikan isi dari suatu dokumen dengan mengacu pada kamus *stopword* yang telah dibuat untuk menentukan kata atau *term* mana yang seharusnya dihilangkan. Contoh kata *term* yang harusnya dihapus yaitu, di, akan, ke, yang, dan, apa, pada, dll. Seperti pada gambar diagram alir 7.



**Gambar 8 Alur Proses Stemming**

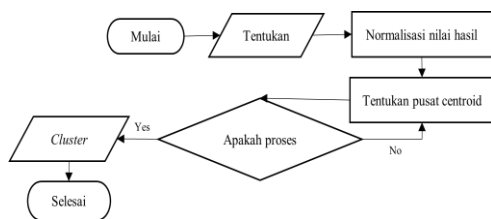
*Stemming* merupakan tahapan proses untuk mengambil suatu kata inti atau kata dasar dari hasil proses filtering. Pada tahapan ini akan dilakukan penghilangan kata imbuhan awalan (*prefix*) dan akhiran (*suffix*) sehingga menghasilkan kata dasar (*root*) dari setiap kata (*token*). Berikut diagram alir untuk proses stemming ditunjukkan gambar 8.



**Gambar 9 Alur Proses Pembobotan Kata**

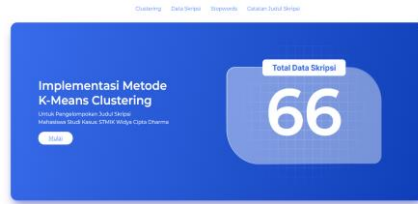
Pembobotan ini dilakukan untuk mengetahui nilai dari setiap *term* yang mewakili isi dokumen. Untuk alur proses dari pembobotan kata ditunjukkan oleh gambar 9.

## 5. IMPLEMENTASI



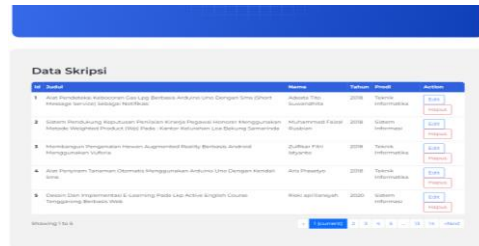
**Gambar 10 Alur Proses Clustering**

Sebelum dilakukan *clustering*, maka nilai perlu untuk dinormalisasi. Setelah nilai ternormalisasi selanjutnya akan dihitung kedekatan jarak term terhadap *centroid*. Untuk penelitian nilai centroid dimasukkan secara random sesuai dengan kebutuhan analisis. Seperti pada gambar 10 menunjukkan alur proses *clustering*.



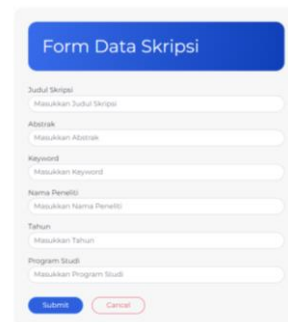
**Gambar 11 Halaman Dashboard**

Pada halaman awal sistem ini, sistem akan menampilkan halaman indeks atau halaman awal yang pada header dari sistem yang berisi *Clustering*, *Data Skripsi*, halaman *Stopwords* dan halaman *Catatan Judul Skripsi*.



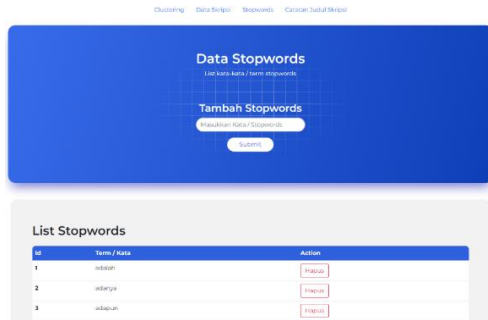
**Gambar 12 Halaman Data Skripsi**

Pada halaman ini, sistem akan menampilkan semua dokumen yang berada pada *database* saat ini. Pada halaman ini diberikan fitur untuk pengguna supaya bisa menambahkan dokumen baru atau menghapus dokumen yang saat ini berada dalam *database* sistem.



**Gambar 13 Halaman Input Dokumen**

Pada halaman ini sistem akan menampilkan form untuk pengguna. Pengguna dapat memasukkan dokumen baru ke dalam database melalui form yang telah disediakan oleh sistem.



Gambar 14 Halaman Stopwords List

Pada halaman ini pengguna dapat melihat stopwords apa saja yang digunakan dalam sistem ini. Pada halaman ini pengguna diberikan pengetahuan untuk mengetahui kata-kata yang masuk sebagai stopwords pada sistem ini. Hal ini bertujuan untuk mengembangkan apabila pengguna merasa bahwa ada kata yang belum masuk sebagai stopwords atau mungkin ada kata yang seharusnya dimasukkan sebagai stopwords.



Gambar 15 Halaman Clustering

Pada halaman ini menampilkan 1 form yang akan digunakan untuk pengguna dapat memasukkan jumlah cluster sesuai yang diinginkan. Untuk dapat mengetahui jumlah dari skripsi yang sudah di masukan oleh ke dalam system dapat dilihat pada Gambar 11 yang kemudian di proses pada halaman clustering dan dapat dilihat pada Gambar 15.

bandwidth	bagun	bank	barang	barode	baru
0	0	0	0	0	0
0	0	0	0	0	0
0,2541222282049398	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0,24383543446481823	0	0,322375882258687	0	0	0
0	0	0	0	0	0
0,18827943395707826	0	0	0	0	0
0	0	0	0	0	0
0,1725644117252395	0	0	0	0	0
0	0	0	0	0	0
0,3240321183021462	0	0	0	0	0
0,12640494797681218	0	0	0	0	0
0,20412222820493987	0	0	0	0	0

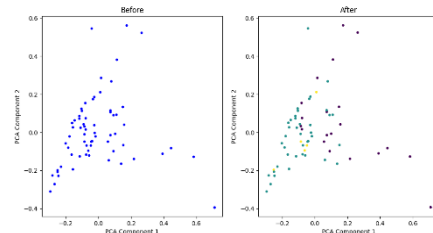
Gambar 16 Gambar Hasil Weight TF-IDF

Setelah proses *Steering* dari proses *Teks Preprocessing* selanjutnya yaitu proses *TF-IDF* dimana didalam proses ini dilakukan perhitungan bobot untuk mengetahui bobot dari setiap *term* atau kata dari setiap bagian dokumen yang telah dimasukkan kedalam sistem. Kemudian dilakukan perhitungan *weight term frequency* mengetahui bobot akhir pada tiap kata setelah proses *TF-IDF*, seperti pada Gambar 16.

Cluster	Tahun	ID	Judul	Nama	Prodi
Cluster 1	2019	3	Membangun Pengenalan Hewan Augmented Reality Berbasis Android Menggunakan Vuforia	Zulfkar Fidi Agfianto	Teknik Informatika
2019	7		Sistem Informasi Reservasi Kamar Berbasis Website Pada Homestay Dewantara II Samarinda	Rifa	Sistem Informasi
2019	9		Decision Support System Pemilihan Siswa untuk Mengikuti Olimpiade Sains Biologi Dengan Metode Simple Multi Attribute Rating Technique (Smart) Pada Sma Negeri 5 Samarinda	Dewi Anggrani Nopiah	Sistem Informasi
2019	10		Sistem Informasi Reservasi Dan Periwisata Pada Taman Wisata Graha Mangrove Bontang Kutim Berbasis Website	Muhammad Arlin Agfianto	Sistem Informasi
2019	14		Membangun Game Jojo Adventure Platform Berbasis Android	Raam Deiki	Sistem Informasi
2019	16		Membangun E-Commerce Oleh-Oleh Khas Kota Samarinda Studi Kasus Pada Utmim Sisa And BP	Ekwadi Yuliviana	Sistem Informasi
2018	18		Membangun Aplikasi Pengenalan Huruf Aritabik Dengan Augmented Reality Pada Lembaga Kursusdan Pelatihan Edu Global Samarinda	Baerie	Manajemen Informatika
2018	19		Aplikasi Surat Masuk Dan Surat Keluar Pada Sma Negeri 5 Samarinda Berbasis Website	Haryandi	Manajemen Informatika
2018	20		Membangun Website Pada Smpn 5 Sangkulirang	Nor Maria	Manajemen Informatika
2018	21		Membangun Aplikasi Pembelajaran Pengenalan Perangkat Keras Komputer Dengan Augmented Reality (Studi Kasus Smk Bhakti Loa Janan, Kutai Kartanegara)	Salmon	Manajemen Informatika

Gambar 17 Hasil Clustering Dokumen

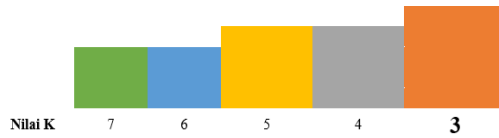
Gambar 17 menjelaskan bahwa hasil pengujian *Clustering* telah mencapai titiknya dan setelah diuji beberapa kali dengan nilai  $K=3$  pada skripsi ini tidak terdapat perpindahan data dari tiap cluster yang telah ditentukan pada saat proses *clustering* akan dijalankan, namun jika ada data yang tidak sesuai dengan kaidah atau data dengan kalimat acak-acakan terdapat kerancuan dalam pengelompokannya.



Gambar 18 Scatter K-Means Clustering

Pada gambar 418 adalah hasil dari *clustering* dengan *K-Means Clustering* dengan nilai  $K=3$  yang menghasilkan pembentukan *cluster*. Pada gambar 4.19 pada *cluster 1* mempunyai data sebanyak 21 *items*, *cluster 2* mempunyai data sebanyak 40 *items*, dan pada *cluster 3* mempunyai data sebanyak 5 *items*. Sehingga total data yang tercluster sebanyak 66 judul

skripsi. *Scatter plot* ini digunakan untuk memvisualisasikan distribusi data setelah dilakukan proses *clustering*.



**Gambar 19 Hasil Cluster nilai K3, K4, K5, K6, K7**

Setelah didapatkan hasil pengujian nilai K3, K4, K5, K6, K7 dan telah di kerucutkan ke cluster yang minimal 3 judul skripsi, maka didapatkan data dari *Cluster 3* yang akan dijadikan sampel dan kesamaan hasil pengelompokkan dari penelitian ini.

**Tabel 1 Hasil nilai K=7**

Cluster	Tahun	Centroid	Judul
Cluster 3	2018	1	Alat Pendeteksi Kebocoran Gas Lpg Berbasis <i>Arduino Uno</i> Dengan Sms ( <i>Short Message Service</i> ) Sebagai Notifikasi
	2018	4	Alat Penyiram Tanaman Otomatis Menggunakan <i>Arduino Uno</i> Dengan Kendali Sms
	2017	38	Sistem Pendeteksi Peringatan Dini Keamanan Rumah Berbasis <i>Mikrokontroler Atmega 16</i> Dan Sms Gateway

**Tabel 2 Hasil Indeks Centroid dari nilai K=7**

Tahun	Centroid	Judul	Kata	Frekuensi	no urut f>1
2018, 2018	indeks centroid 1	C1 thd 4	alat	2	1
			<i>arduino</i>	2	3
			<i>uno</i>	2	4
			dengan	2	5
2018, 2017	indeks centroid 1	C1 thd 38	pendeteksi	2	2
			berbasis	2	7
		1 thd 4 thd 38	sms	3	6
	indeks centroid 4	4 thd 38		0	

**Tabel 3 Hasil nilai K=3**

Cluster	Tahun	Centroid	Judul
Cluster 3	2018	1	Alat Pendeteksi Kebocoran Gas Lpg Berbasis <i>Arduino Uno</i> Dengan Sms ( <i>Short Message Service</i> ) Sebagai Notifikasi

Tahun	Centroid	Judul
2018	4	Alat Penyiram Tanaman Otomatis Menggunakan <i>Arduino Uno</i> Dengan Kendali Sms
2019	6	Sistem Pendukung Keputusan Penyeleksian Calon Karyawan Bagian Operator Alat Berat Pada Pt. Pama Persada Nusantara Menggunakan Metode Mfep Berbasis Web
2017	38	Sistem Pendeteksi Peringatan Dini Keamanan Rumah Berbasis <i>Mikrokontroler Atmega 16</i> Dan Sms Gateway
2017	42	Implementasi Absensi Siswa Berbasis Sms Gateway Pada Smk Muhammadiyah 1 Samarinda

**Tabel 4 Hasil Indeks Centroid dari nilai K=3**

Tahun	Centroid	Judul	Kata	Frekuensi	No urut f>1
2018, 2018	indeks centroid 1	C1 thd 4	<i>arduino</i>	2	3
			<i>uno</i>	2	4
			dengan	2	5
2018, 2017	indeks centroid 1	C1 thd 38	pendeteksi	2	2
	indeks centroid 1	C1 thd 42		0	
2018, 2019	indeks centroid 1	C1 thd 4 thd 6	alat	3	1
2018, 2018, 2017, 2017	indeks centroid 1	1 thd 4 thd 38 thd 42	sms	4	6
2018, 2019, 2017, 2017	indeks centroid 1	C1 thd 6 thd 38 thd 42	berbasis	4	7
	Indeks centroid 4	C4 thd 6	menggunakan	2	8
	Indeks centroid 4	C4 thd 38		0	
2018, 2019, 2017	Indeks centroid 4	C4 thd 6 thd 38		0	
	Indeks centroid 4	C4 thd 42		0	
2019, 2017	Indeks centroid 6	C6 thd 38	sistem	2	9

Tahun	Centroid	Judul	Kata	Frekuensi	No uru t f >1
		C6 thd 42	pada	2	10
		C6 thd 38 thd 42		0	
	Indeks Centroid 38	C38 thd 42		0	

Berdasarkan hasil yang telah didapat hingga tabel 4 diatas, maka didapat nilai frekuensi kesamaan pola dari nilai K7, K6, K5, K4, dan K3 seperti pada gambar 20 dibawah ini.



Gambar 20 Pola F terhadap nilai K

Dari hasil yang didapat dari pola frekuensi kemunculan kata terhadap nilai K maka Frekuensi yang memiliki kemunculan paling banyak adalah pada K3 di Cluster 3, untuk lebih mengerucutkan hasil analisa ini dan karena dilakukan pada tahun 2024 yang kemudian didapatkan tahun relevan sebagai acuan terhadap kesamaan judul adalah 4 tahun sebelumnya, maka didapat nilai pada K = 3 di Cluster 3 dan Tahun 2019 dengan judul “Sistem Pendukung Keputusan Penyeleksian Calon Karyawan Bagian Operator Alat Berat Pada Pt. Pama Persada Nusantara Menggunakan Metode Mfep Berbasis Web”.

Tabel 5 Penentu Rekomendasi Judul Skripsi

Pecah Kata	Penentuan SOP	Bobot
Sistem Pendukung Keputusan	predikat utama	1
Penyeleksian	objek	3
Calon Karyawan	sub sub subjek	4
Bagian Operator Alat Berat	sub subjek	4
Pada		
Pt. Pama Persada Nusantara	subjek	4
Menggunakan		
Metode		
Mfep	predikat utama	1

Pecah Kata	Penentuan SOP	Bobot
Berbasis	predikat sekunder	2
web	predikat sekunder	2

Blackbox		
Pecah Kata	Penentuan SOP	Bobot
Sistem Pendukung Keputusan	predikat utama	1
Penyeleksian	objek	3
Calon Karyawan	sub sub subjek	4
Bagian Operator Alat Berat	sub subjek	4
Pada		
Pt. Pama Persada Nusantara	subjek	4
Menggunakan		
Metode		
Mfep	predikat utama	1
Berbasis	predikat sekunder	2
web	predikat sekunder	2

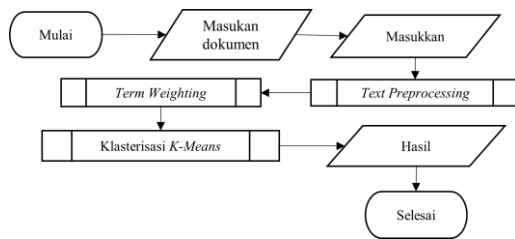
Gambar 21 Rancangan Blackbox Proses Rekomendasi

Sehingga dapat disimpulkan kata setelah berbasis dapat diganti menjadi kata lain, seperti pada judul dari hasil penelitian ini “Sistem Pendukung Keputusan Penyeleksian Calon Karyawan Bagian Operator Alat Berat Pada PT. Pama Persada Nusantara Menggunakan Metode Mfep Berbasis Web” dari judul yang tertera itu maka akan direkomendasikan “Sistem Pendukung Keputusan Penyeleksian Calon Karyawan Bagian Operator Alat Berat Pada PT. Pama Persada Nusantara Menggunakan Metode Mfep Berbasis (Desktop, Mobile)” seperti yang akan dijelaskan pada halaman selanjutnya.

Clustering	Data Skripsi	Stopwords	Frekuensi Kata	Rekomendasi Judul	
Rekomendasi Judul Skripsi STMIK Widya Cipta Dharma					
N Cluster	Tahun	Program Studi	Judul	Frekuensi	Rekomendasi
Cluster 3	2019	Teknik Informatika	Sistem Pendukung Keputusan Penyeleksian Calon Karyawan Bagian Operator Alat Berat Pada Pt. Pama Persada Nusantara Menggunakan Metode Mfep Berbasis Web	Berbasis	Desktop, Mobile
	2018	Teknik Informatika	Alat Penyiram Tanaman Otomatis Menggunakan Arduino Uno Dengan Kendali Sms	Alat	Raspberry Pi

Gambar 22 Interface Rekomendasi Hasil

Setelah itu hasil akan didapat ketika proses clustering dilakukan dan akan berproses hingga ditentukannya saran judul skripsi yang diinginkan. Seperti pada gambar 22.



**Gambar 23** Alur sistem secara umum

Dari pembahasan ini akan melakukan bagaimana *flowchart* dari proses keseluruhan dari memasukkan dokumen skripsi kedalam sistem, memasukkan jumlah  $K$  yang akan di analisis hingga tahapan *preprocessing*, (tokenisasi, *case folding*, *filtering*, *stemming*), *term weighting*, dan *clustering*, seperti pada gambar diagram alir 23.

## 6. KESIMPULAN

Penerapan metode *K-Means Clustering* untuk pengelompokan judul skripsi mahasiswa di STMik Widya Cipta Dharma dapat dilakukan dengan menggunakan algoritma TF-IDF untuk mewakili setiap judul skripsi sebagai vektor fitur. Hasil uji coba menunjukkan bahwa metode ini dapat digunakan untuk mengelompokkan judul skripsi, tetapi masih kurang optimal karena kurang penggunaan metode *Elbow* atau *Cosine similarity*. Percobaan menggunakan nilai  $K=3$  menghasilkan tiga *cluster* dengan jumlah item yang berbeda, dimana hasil pada *cluster 1* mempunyai data sebanyak 21 *items*, *cluster 2* mempunyai data sebanyak 40 *items*, dan pada *cluster 3* mempunyai data sebanyak 5 *items*.

Dari hasil analisis dengan memasukkan nilai *cluster* yang bervariasi dari  $K=3$  hingga  $K=7$ , analisis hasil *clustering* cenderung kepada pola hasil C3, yaitu jumlah judul tercluster dalam  $C3 = 3$  dan merupakan *cluster* terpola di mana kata “berbasis” dan “alat” paling tinggi frekuensinya dari C3 pada saat  $k = 3$ . Kemudian peneliti mengambil sampel untuk penyajian dalam bentuk *excel* yang diolah berdasarkan intensitas kepentingan serta membuat rancangan *mookup* dalam memberikan saran judul dan sistem yang bisa digunakan dalam penelitian selanjutnya.

## 7. SARAN

Beberapa hal yang diharapkan untuk bisa dikembangkan, pada penelitian selanjutnya agar sistem tersebut bisa lebih bermanfaat lagi untuk membangun aplikasi pengelompokan buku dan katalog lainnya di perpustakaan Stmik Widya

Cipta Dharma yang berbasis Web diantaranya adalah :

1. Mengoptimalkan source code agar proses perhitungan TF-IDF lebih efisien dan cepat.
2. Penambahan fitur-fitur lain dalam penyempurnaan rancangan mookup peneliti dengan adanya penelitian lanjutan.
3. Penggunaan metode *Elbow* atau *Cosine similarity* dapat membantu meningkatkan akurasi *clustering*.
4. Membuat kesimpulan akhir dari pengelompokan untuk menghasilkan rekomendasi judul skripsi yang tidak serupa.
5. Menambahkan sistem machine learning yang dapat menentukan saran judul skripsi secara otomatis.
6. Menambahkan fitur dinamis agar data skripsi yang baru otomatis terintegrasi ke dalam sistem aplikasi perpustakaan STMik Widya Cipta Dharma.
7. Pengembangan antarmuka sistem yang lebih user-friendly dapat membantu meningkatkan kenyamanan pengguna.

## 8. DAFTAR PUSTAKA

- Anwar, Khoirul, 2015. *White Box Testing dan Black Box Testing*, Semarang.
- Arriyanti, E., dan Aditya, P. (2020). *A Synthesis of Optimal Unknown Number Clustering System and Categorical*. *International Journal*, 9(5).
- Indrawan, B. R., 2018. Penerapan Algoritma *K-Means Clustering* untuk Menentukan Strategi Promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Yogyakarta.
- Han, J., dan Kamber, M. (2012). *Data Mining : C d h concepts and techniques*.
- Mukherjee, S., Shaw, R., Haldar, N., dan Changdar, S. (2015). *A survey of da- ta mining applications and techniques*. *International journal of Computer Science and information Technologies*, 6(5).
- Rahmat, 2020. *Text Mining Judul Skripsi Menggunakan K-Means Clustering* (Studi Kasus di Jurusan Teknik Informatika UIN Alauddin Makassar), Makassar.
- Suntoro, Joko, 2018. *Data Mining Algoritme dan Implementasi Menggunakan Bahasa Pemrograman PHP*, Semarang.



- Sergey, Smirnov, 2002. *Software Testing: Black-Box Techniques*, 1-4.
- Wakhidah, Nur, 2010. *Clustering menggunakan K-Means Algorithm (K-Means Algorithm Clustering)*, Semarang.
- William, Laurie, 2006. *Testing Overview and Black-Box Testing Techniques*, 35-59