

# Penerapan Metode *Cosine Similarity* dan Pembobotan *TF/IDF* pada Sistem Klasifikasi Sinopsis Buku di Perpustakaan Kejaksaan Negeri Jember

Moh. Mahdi Sya'bani <sup>1)</sup>, Reni Umilasari <sup>2)</sup>

<sup>1,2)</sup> Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember  
E-mail : <sup>1)</sup> mahdiumj@gmail.com <sup>2)</sup> reni.umilasari@unmuhjember.ac.id

## ABSTRAK

Selama ini perpustakaan Kejaksaan Negeri Jember belum dikelola dengan baik. Pada saat pegawai perpustakaan ingin mengetahui macam-macam judul buku sesuai kategori yang mereka inginkan, pegawai perpustakaan mencari satu persatu di katalog bukunya. Sehingga kondisi demikian akan menyulitkan pegawai perpustakaan dalam mencari judul buku sesuai kategori yang diinginkan. Hal ini dapat mengakibatkan pegawai perpustakaan kewalahan. Pelayanan yang sangat baik jika pengguna perpustakaan merasa puas dengan pelayanannya. Semakin banyaknya dokumen buku yang ada di perpustakaan semakin banyak tenaga dan waktu yang diperlukan. Maka memerlukan sebuah sistem aplikasi untuk mengklasifikasikan dokumen buku berdasarkan kategori buku secara otomatis. Untuk mendapatkan hasil yang optimal dalam mengklasifikasikan sebuah dokumen maka diperlukan sebuah metode untuk mengklasifikasikan dokumen. Metode yang digunakan adalah pembobotan *TF/IDF* dan *cosine similarity* pada model *vector space model*. Untuk mengukur tingkat kemiripan suatu dokumen dengan menggunakan sinopsis buku.

Pengujian aplikasi terdapat 120 dokumen sinopsis dengan 10 kategori dan menghasilkan nilai *precision* sebesar 90,91% pada *threshold* 0,1 dan nilai *recall* sebesar 100% pada *threshold* 0,1 dan 0,2. Ketepatan akurasi pada sistem aplikasi yang diuji adalah 80,83%.

**Kata Kunci** : Klasifikasi Dokumen, Sinopsis Buku, Perpustakaan Kejaksaan Negeri Jember, Pembobotan *TF/IDF* dan *Cosine Similarity*.

## 1. PENDAHULUAN

Kejaksaan Negeri terdapat sebuah perpustakaan yang dapat menjadi suatu sarana meningkatkan kinerja terutama dalam menangani suatu perkara. Tujuannya adalah mengembangkan sarana yang ada dan memberikan pencerahan terhadap pegawai yang ada di Kejaksaan Negeri, termasuk di Kejaksaan Negeri Jember. Perpustakaan yang berisi tentang hukum dan dokumentasi perundangan-undangan dalam bentuk unit satuan kerja untuk menangani khusus dalam ilmu hukum. Terdapat 1.966 judul buku yang di perpustakaan tersebut dan di dalamnya

terdapat banyak kategori buku yang di antaranya HAM Nasional, Hukum Asing, Hukum Pidana, Hukum Perdata dan lain-lain.

Dokumen buku yang ada di perpustakaan Kejaksaan Negeri Jember belum dikelola dengan baik. Pada saat pegawai perpustakaan ingin mengetahui macam-macam judul buku sesuai kategori yang mereka inginkan, pegawai perpustakaan mencari satu persatu di katalog bukunya. Sehingga kondisi demikian akan menyulitkan pegawai perpustakaan dalam pencarian judul buku sesuai kategori yang diinginkan. Hal ini dapat mengakibatkan pegawai

perpustakaan kewalahan. Pelayanan suatu perpustakaan dikatakan prima (sangat baik) jika para pengguna perpustakaan merasa puas atas pelayanan yang diberikan (Prastowo, 2013). Semakin banyak jumlah buku atau dokumen yang tersedia, semakin banyak tenaga dan waktu untuk diperlukan. Dari sinilah faktor penyebabnya menjadi topik menarik untuk sebuah penelitian. Dengan demikian diperlukan aplikasi sistem klasifikasi dengan otomatis. Untuk mendapatkan hasil yang optimal dalam mengklasifikasikan dokumen ini mengambil dataset dari sebuah sinopsis buku yang ada di perpustakaan.

Penerapan klasifikasi dokumen membutuhkan metode mengklasifikasikan dokumen dengan otomatis sesuai kategorinya yaitu menggunakan metode *text mining*. Metode ini dapat mengklasifikasikan dokumen. Terdapat banyak algoritma salah satunya adalah cosine similiraty pada *model vector space model*. Metode tersebut dapat mengklasifikasikan yaitu memberi pembobotan pada suatu dokumen untuk menghitung *TF (Term Frekuensi)* atau *IDF (Inverse Dokumen Frekuensi)*. *Term-term* akan mendapatkan sebuah nilai yang perbandingannya sama dengan jumlah kemunculan *term* pada kalimat (Mark & Lloyd, 1999).

Tujuan dari penelitian ini adalah harapannya dengan dibuatnya sistem aplikasi ini bertujuan dapat membantu pegawai perpustakaan kejaksaan negeri Jember dalam menimalisir pekerjaannya dalam mencari kategori buku dengan secara otomatis.

## 2. TINJAUAN PUSTAKA

### 2.1 Klasifikasi Dokumen

Pada tahapan proses klasifikasi dokumen secara keseluruhan, terdapat

beberapa tahapan yang akan diawali dengan identifikasi dokumen yakni dimana masing-masing dokumen akan diidentifikasi secara kata atau term yang terdapat didalamnya. Sehingga tahap pertama adalah tokenizing pada kata yang terdapat di dalam dokumen untuk mendapatkan kata yang mampu berdiri sendiri, dan terbebas dari tanda-tanda baca, spasi dan sebagainya (Kaplan, 1995). Selanjutnya tahap filtering (*wordlist/stoplist*) untuk menghilangkan kata yang tidak berpotensi sebagai indikasi topic dalam dokumen (Porter, 2001). Setelah itu dilakukan stemming pada kata yang tersisa untuk mendapatkan kata dasar.

Selanjutnya akan melalui proses sinonim atau persamaan dalam setiap kata yang ada pada dokumen. Maka untuk mengantisipasi dalam memberikan makna yang sama pada saat pengguna menggunakan [5].

Kata akan diberikan sebuah bobot setiap katanya yang sudah melakukan proses sinonim atau persamaan kata. Proses ini akan dilanjutkan dengan menghitung tingkat kemiripan (*similarity*) antar dokumen.

### 2.2 Sinopsis

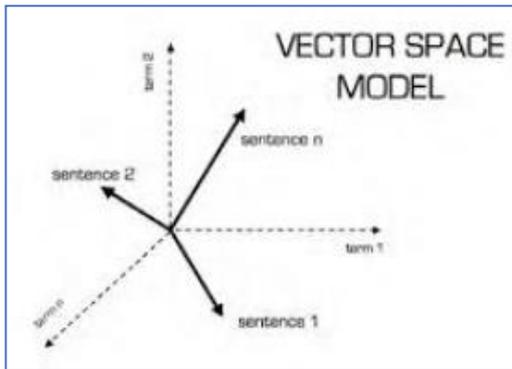
Sinopsis merupakan sebuah karangan asli yang diterbitkan secara bersamaan [6]. Sinopsis dapat diartikan abstraksi, ringkasan atau ikhtisar karangan [7] Mengatakan cara yang efektif untuk menyajikan karangan dalam bentuk panjang dijadikan sebuah ringkasan dalam bentuk yang pendek. Dengan demikian meringkas sama halnya memotong atau memangkas sebatang pohon, akhirnya tinggal batang dan cabang-cabangnya yang paling terpenting.

### 2.3 Text Mining

*Text mining* adalah sekumpulan banyak dokumen untuk mendapatkan sebuah pengetahuan atau informasi

dengan menggunakan alat analisis. *Text mining* dipergunakan untuk ekstraksi informasi yang mengeksplorasi pola yang sangat menarik.

Proses pertama yang dilakukan menyiapkan sebuah teks menjadi data yang akan diolah menjadi sebuah informasi atau pengetahuan. Berikutnya sekumpulan kalimat akan dipilah atau dipecah menjadi sebuah unsur yang memiliki arti.



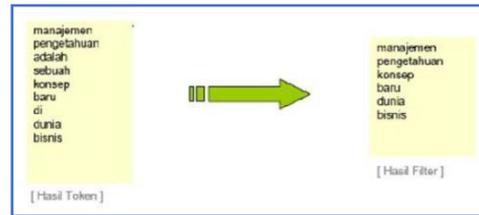
Gambar 1 Ilustrasi *Text Mining*

*Tokenizing* merupakan kalimat yang akan dipecah menjadi sebuah kata/*token* (Feldman dan Sanger, 2007). Terdapat sebuah fitur yaitu kapitalisasi, tanda baca, keberadaan digit dan lain-lain. Berikut contoh proses cara kerja *tokenizing* pada gambar 2.



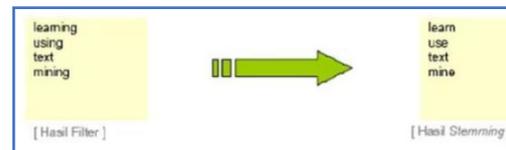
Gambar 2 Proses *Tokenizing*

*Pre processing* ini memiliki beberapa tahapan dalam *text mining* yang terdiri dari *casefolding*, *tokenizing*, *filtering* dan yang terakhir *stemming*. Ini adalah contoh proses *filtering* pada gambar 3 :



Gambar 3 Proses *Filtering*

Berikutnya proses *stemming*, berikut proses contoh ilustrasi pada gambar 4.



Gambar 4 Proses *Stemming*

## 2.4 Metode Pembobotan *TF/IDF* dan *Cosine Similarity*

Pembobotan *TF/IDF* adalah suatu hubungan kata (*term*) yang berada pada dokumen yang akan diberikan suatu nilai bobot (Robertson, 2005). Perhitungan bobot memiliki penggabungan dua konsep, yang pertama kemunculan sebuah frekuensi kata pada dokumen dan kedua adalah kata yang mengandung inverse pada frekuensi dokumen tersebut. *Term* mendapatkan nilai frekuensi pada dokumen yang diberikan bahwa seberapa penting *term* tersebut. Seberapa umum kata yang sering muncul pada frekuensi dokumen sehingga hubungan bobot antara *term* dengan dokumen akan tinggi apabila *term* frekuensi itu tinggi di dalam sebuah dokumen. Berapapun nilai dari sebuah  $tf_i$  jika nilai dalam suatu  $N = df_{ti}$ , maka hasil dari nilai tersebut akan mendapatkan nilai 0 (nol) khusus untuk perhitungan mencari nilai *IDF* (Sulistyo & Sarno, 2008). Pada sisi *IDF* maka dapat ditambahkan nilai 1. Sehingga perhitungan bobot menjadi sebagai berikut :

Rumus *TF/IDF*:

$$IDF = t_{f_{ti}} \times IDF \quad (1)$$

$$IDF = t_{f_{ti}} \times \log \frac{N}{df_{ti}} + 1 \quad (2)$$

$$\omega_D(t_i) = \frac{t_{f_{ti}} \times \log \left( \frac{N}{df_{ti}} + 1 \right)}{\sqrt{\sum (t_{f_{ti}} \times \log \left( \frac{N}{df_{ti}} + 1 \right) + 1)}} \quad (3)$$

Metode *Cosine Similarity* adalah sebuah metode yang menghitung tingkat kemiripan (*similarity*) antar dua objek atau lebih. Dalam keseluruhan metode ini didasari pada *vector space similarity measure*. Perhitungan ini menggunakan dua buah objek dokumen (*D1* dan *D2*) untuk menghitung sebuah *similarity* antar dokumen yang dinyatakan dalam sebuah *vector* dengan menggunakan *keyword* atau kata kunci. Berikut rumus yang digunakan :

Rumus *cosine similarity* :

$$\cos(Q, D) = \sum_{r=1}^M \omega_Q(t_i) \times \omega_D(t_i) \quad (4)$$

### 3. METODE PENELITIAN

#### 3.1 Tahap Penelitian

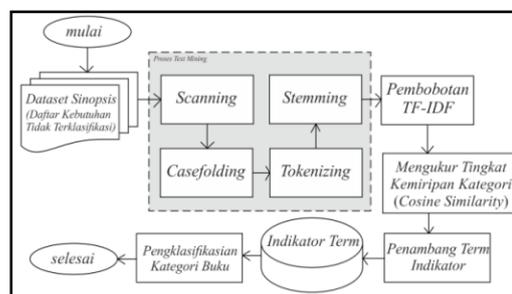
Penelitian ini mengumpulkan data yang berupa dokumen yang berisi dari sinopsis suatu judul buku yang ada di perpustakaan Kejaksaan Negeri Jember. Dataset ini bertipe excel tetapi belum dikelola dengan baik oleh perpustakaan Kejaksaan Negeri Jember. Dataset yang digunakan 120 *data* buku yang berisi atribut yaitu sinopsis, judul buku, penerbit, jumlah buku dan tahun terbit. Untuk kategori atau jenis bukunya terdiri 10 kategori judul buku yang dapat dilihat pada Tabel 1.

**Tabel 1. JENIS KATEGORI BUKU**

No	KATEGORI KEBUTUHAN
1.	HAM Nasional
2.	HAM Asing

3.	Hukum pada Umumnya
4.	Hukum Pidana
5.	Hukum Perdata
6.	Hukum Tata Negara
7.	Hukum Internasional
8.	Sosial dan Politik
9.	Penelitian dan Pengkajiannya
10.	Perundang-Undangan

Pada Gambar 5 dijelaskan bahwa *flowchart* metode *cosine similarity* ini menjelaskan kerangka kerja sistem aplikasi yang akan dibuat:



**Gambar 5** Rancangan Sistem Aplikasi

Langkah pertama adalah dilakukannya penentuan *keyword*, setelah itu akan dilakukan proses text mining yang terdiri dari penelusuran (*scanning*) terhadap paragraf dan *casefolding* kemudian memilah paragraf tersebut perkalmatnya dan perkata (*tokenizing*). Membuang kata yang tidak penting atau penghubung yang pada akhirnya mengembalikan kedalam bentuk dasar (*stemming*). Setelah di dapatkan kumpulan kata dasar (*term*) dilakukan proses pembobotan menggunakan kata kunci atau *keyword* yang dicari dengan *TF/IDF* kemudian dilakukannya perhitungan menggunakan *cosine similarity* agar mengetahui hasil ketepatan sebuah klasifikasi dengan sesuai kategori bukunya.

**3.2 Metode**

Tahap data *sample* digunakan untuk melihat ketepatan pada klasifikasi pada dokumen. Untuk mendapatkan klasifikasi yang efektif, membutuhkan *keyword* (kata kunci) setiap kategorinya. Tahap ini akan menggunakan 2 dokumen setiap kategori buku yang ditunjukkan Tabel I.

Tahap *training* pada penelitian berjumlah 120 sinopsis buku dengan 10 kategori pada Tabel I. Sistem ini bertujuan untuk mengklasifikasikan dokumen secara otomatis berdasarkan kategori yang ada.

1. Sebelumnya, setiap baris kebutuhan yang akan diklasifikasikan dilakukan proses *scanning* sampai proses *stemming* dan pembuangan *stopword* agar memperoleh *term-term* dasar
2. Proses pembobotan pada setiap daftar kata sebagai berikut.

**a. Metode TF**

Menentukan nilai frekuensi kemunculan kata.

**Tabel 2.** DAFTAR FREKUENSI KATA (TF)

Kategori HAM Nasional	Term	TF
Dokumen 1	angkat	1
	kitab	2
	regional	2
Dokumen 2	kitab	2
	tambah	1
	regional	2

**b. Metode IDF**

Berdasarkan daftar kata dalam Tabel 2, maka dihitung *IDF* untuk setiap kata menggunakan persamaan 2.

Berikut Perhitungannya sebagai berikut.

**Dokumen 1 :**

$$\begin{aligned} \text{angkat : } IDF &= tf_{ti} \times \log \frac{N}{df_{ti}} + 1 \\ &= 1 \times \log \frac{2}{1} + 1 = 1,301 \end{aligned}$$

$$\begin{aligned} \text{kitab : } IDF &= tf_{ti} \times \log \frac{N}{df_{ti}} + 1 \\ &= 2 \times \log \frac{2}{2} + 1 = 1 \end{aligned}$$

$$\begin{aligned} \text{regional : } IDF &= tf_{ti} \times \log \frac{N}{df_{ti}} + 1 \\ &= 2 \times \log \frac{2}{2} + 1 = 1 \end{aligned}$$

**Dokumen 2 :**

$$\begin{aligned} \text{kitab : } IDF &= tf_{ti} \times \log \frac{N}{df_{ti}} + 1 \\ &= 2 \times \log \frac{2}{2} + 1 = 1 \end{aligned}$$

$$\begin{aligned} \text{tambah : } IDF &= tf_{ti} \times \log \frac{N}{df_{ti}} + 1 \\ &= 1 \times \log \frac{2}{1} + 1 = 1,301 \end{aligned}$$

$$\begin{aligned} \text{regional : } IDF &= tf_{ti} \times \log \frac{N}{df_{ti}} + 1 \\ &= 2 \times \log \frac{2}{2} + 1 = 1 \end{aligned}$$

Untuk perhitungan *IDF* bisa dilihat di Table 3.

**Tabel 3.** HASIL PERHITUNGAN *IDF*

Kategori HAM Nasional	Term	IDF
Dokumen 1	angkat	1, 301
	kitab	1
	regional	1
Dokumen 2	kitab	1
	tambah	1, 301
	regional	1

**c. Metode TF-IDF**

Berdasarkan Tabel 2 dan 3, dilakukan pembobotan *TF-IDF* melalui persamaan 3 sehingga

dihasilkan bobot. Untuk perhitungan selengkapnya sebagai berikut :

**Dokumen 1:**

$$\begin{aligned} \text{angkat : } \omega_D (ti) &= \frac{tf_{ti}x \log \left(\frac{N}{df}\right)+1}{\sqrt{\sum(tf_{ti}x \log \left(\frac{N}{df_{ti}}\right)+1)}} \\ &= \frac{1 x \log \left(\frac{2}{1}\right)+1}{\sqrt{\sum(1 x \log \left(\frac{2}{1}\right)+1)}} = \frac{1,301}{\sqrt{1 x 3.301}} = \\ &0,7161 \end{aligned}$$

$$\begin{aligned} \text{kitab : } \omega_D (ti) &= \frac{tf_{ti}x \log \left(\frac{N}{df}\right)+1}{\sqrt{\sum(tf_{ti}x \log \left(\frac{N}{df_{ti}}\right)+1)}} \\ &= \frac{1 x \log \left(\frac{2}{2}\right)+1}{\sqrt{\sum(1 x \log \left(\frac{2}{1}\right)+1)}} = \frac{2}{\sqrt{2 x 3.301}} = \\ &0,5503 \end{aligned}$$

$$\begin{aligned} \text{regional : } \omega_D (ti) &= \frac{tf_{ti}x \log \left(\frac{N}{df}\right)+1}{\sqrt{\sum(tf_{ti}x \log \left(\frac{N}{df_{ti}}\right)+1)}} \\ &= \frac{1 x \log \left(\frac{2}{2}\right)+1}{\sqrt{\sum(1 x \log \left(\frac{2}{1}\right)+1)}} = \frac{2}{\sqrt{2 x 3.301}} = \\ &0,5503 \end{aligned}$$

**Dokumen 2:**

$$\begin{aligned} \text{kitab : } \omega_D (ti) &= \frac{tf_{ti}x \log \left(\frac{N}{df}\right)+1}{\sqrt{\sum(tf_{ti}x \log \left(\frac{N}{df_{ti}}\right)+1)}} \\ &= \frac{1 x \log \left(\frac{2}{2}\right)+1}{\sqrt{\sum(1 x \log \left(\frac{2}{1}\right)+1)}} \\ &= \frac{2}{\sqrt{2 x 3.301}} = 0,5503 \end{aligned}$$

$$\begin{aligned} \text{tambah : } \omega_D (ti) &= \frac{tf_{ti}x \log \left(\frac{N}{df}\right)+1}{\sqrt{\sum(tf_{ti}x \log \left(\frac{N}{df_{ti}}\right)+1)}} = \frac{1 x \log \left(\frac{2}{1}\right)+1}{\sqrt{\sum(1 x \log \left(\frac{2}{1}\right)+1)}} \\ &= \frac{1,301}{\sqrt{1 x 3.301}} = 0,7161 \end{aligned}$$

$$\begin{aligned} \text{regional : } \omega_D (ti) &= \frac{tf_{ti}x \log \left(\frac{N}{df}\right)+1}{\sqrt{\sum(tf_{ti}x \log \left(\frac{N}{df_{ti}}\right)+1)}} = \frac{1 x \log \left(\frac{2}{2}\right)+1}{\sqrt{\sum(1 x \log \left(\frac{2}{1}\right)+1)}} = \end{aligned}$$

$$\frac{2}{\sqrt{2 x 3.301}} = 0,5503$$

Berikutnya hasil perhitungan *IDF* sebagaimana Tabel 4.

**Tabel 4.** HASIL PERHITUNGAN TF-IDF

Kategori HAM Nasional	Term	TF-IDF $\omega_D (ti)$
Dokumen 1	angkat	0,7161
	kitab	0,5503
	regional	0,5503
Dokumen 2	kitab	0,5503
	tambah	0,7161
	regional	0,5503

**d. Metode Cosine Similarity**

Pengukuran tingkat kemiripan kebutuhan dalam pengklasifian kategori berdasarkan persamaan 4.

$$\begin{aligned} \cos(Q, D) &= \sum_{r=1}^M \omega_Q (t_i) x \omega_D (t_i) \\ &= (\omega_{Q1T1} * \omega_{D1T1}) + (\omega_{Q1T2} * \omega_{D1T2}) + (\omega_{Q1T3} * \omega_{D1T3}) + (\omega_{Q1T4} * \omega_{D1T4}) \\ &= (0,7161 * 0) + (0,5503 * 0,5503) + (0 * 0,5503) + (0 * 0,7161) \\ &= 0,6056 \end{aligned}$$

Dengan cara yang sama diperoleh nilai kemiripan antara kategori dengan dokumen. Selengkapnya perhitungan *cosine similarity* dengan 10 kategori, berikut hasil perhitungan pada Tabel 5.

**Tabel 5.** Hasil Perhitungan Keseluruhan Kategori

No	Kebutuhan Katagori	Hasil (cos (Q,D))
1.	HAM Nasional	0,3410
2.	HAM Asing	0,4231
3.	Hukum pda Umumnya	0,4413
4.	Hukum Pidana	0,2669
5.	Hukum Perdata	0,4558
6.	Hukum Tata Negara	0,3224
7.	Hukum Internasional	0,3929

8.	Sosial dan Politik	0,5604
9.	Penelitian dan Pengkajiannya	0,5093
10.	Perundang-undangan	0,6446

Pada penambang *term* indikator *Q* ditetapkan sebagai tipe kategori dokumen tertentu dan *term* teratas (*ranking*) diidentifikasi sebagai *term-term* indikator pada tipe kategori buku tertentu. Untuk *D* ditetapkan sebagai kebutuhan yang terklasifikasi dalam tipe kategori *Q*. *Term-term* indikator dari tipe kategori *Q* ditemukan dengan mempertimbangkan sekumpulan kebutuhan *D* dari semua tipe-tipe kategori *Q* pada suatu data *training*.

Masing-masing *term* diberikan suatu nilai pembobotan *TF-IDF*, dan *term-term* ini diurutkan secara menurun. Term akan di *ranking* dan teratas diidentifikasi sebagai *term-term* indikator pada tipe kategori *Q* tertentu.

Perhitungan manual berdasarkan kesesuaian makna dan term potensial dari kebutuhan, sehingga total jumlah kebutuhan yang terambil (*retrived*) dan jumlah relevan berdasarkan *handtrace* (relevan) akan membentuk kinerja rata-rata *precision* dan *recall* dari semua dataset dengan menggunakan nilai ambang batas atau *threshold*. Disini menguji 2 dokumen sinopsis setiap kategorinya, hasil dari perhitungan manual akan dibandingkan dengan *handtrace* yang ada.

#### 4. HASIL DAN PEMBAHASAN

##### 4.1 Implementasi Halaman Aplikasi

Pada tahap implementasi ini dilakukan terhadap dataset sebesar 120 data yang berisikan sinopsis buku. Dataset tersebut akan diolah oleh aplikasi dengan fungsi-fungsi yang ada yaitu *create data*, *delete data* dan *read data*. Tahap penelitian ini berfungsi untuk menghasilkan klasifikasi sinopsis buku

berdasarkan kategori buku. Pengujian aplikasi ini menekankan seberapa kuat akan menghasilkan kebutuhan-kebutuhan yang terklasifikasi dalam setiap kategorinya memiliki nilai ambang batas (*threshold*). Tampilan halaman aplikasi menginputkan data pada Gambar 6 dan Gambar 7.

Gambar 6 Halaman Input 1

Untuk memasukan sebuah sinopsis pilihlah sinopsis yang berbahasa Indonesia dan tentukan *handtrace* yang telah diuji oleh pakar sesuai dibidangnya.

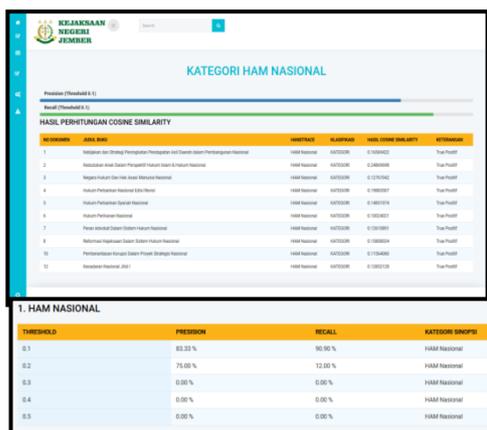
Gambar 7 halaman input 2

Berikutnya Gambar 4 terdapat input-an untuk memasukan kata kunci baru (*keyword*) dan sesuai dengan kategori yang akan dipilih dari 10 kategori yang ada. Untuk implementasi aplikasi kali ini menggunakan data training sebanyak 120 data sinopsis buku.

Gambar 8 Tampilan file import

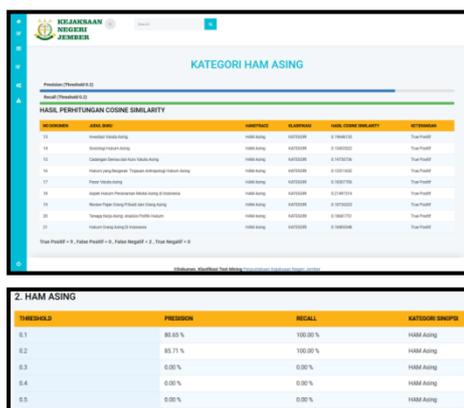
#### 4.2 Implementasi Halaman Pengujian Klasifikasi

Pengujian aplikasi ini menguji 10 kategori kebutuhan yang terletak pada Tabel 5 beserta perhitungan *threshold*. Data yang digunakan 120 sinopsis buku yang berada pada di Perpustakaan Kejaksaan Negeri Jember.



Gambar 9 hasil pengujian HAM Nasional

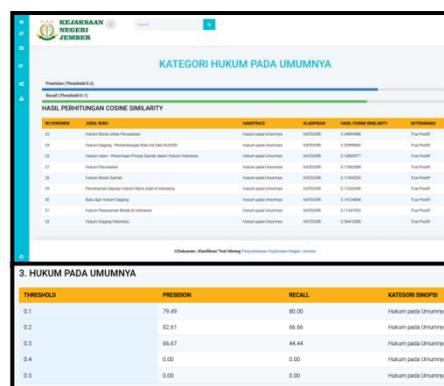
Kategori HAM Nasional rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,1 sebesar **83,33%** dan Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **90,90%**.



Gambar 10 Hasil Pengujian HAM Asing

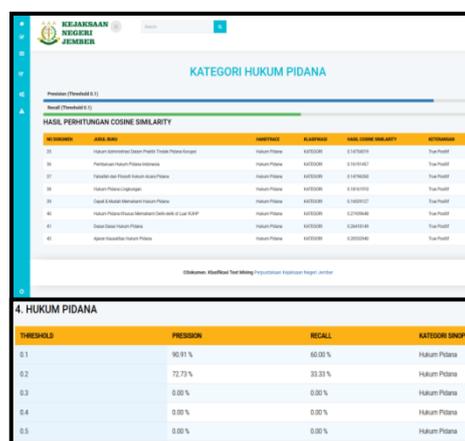
Kategori HAM Asing menghasilkan nilai *threshold* rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,1 sebesar **80,65%**. Kinerja *recall* tertinggi

terletak pada *threshold* 0,1 sebesar **100%**.



Gambar 11 Hasil Pegujian Hukum Pada Umumnya

Kategori Hukum pada Umumnya menghasilkan nilai *threshold* rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,2 sebesar **82,61%**. Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **80%**



Gambar 12 Hasil Pengujian Hukum Pidana

Kategori Hukum Pidana menghasilkan nilai *threshold* dari rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,1 sebesar **90,91%**. Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **60%**.

**5. HUKUM PERDATA**

THRESHOLD	PRECISION	RECALL	KATEGORI SINGKAP
0,1	87,32%	73,81%	Hukum Perdata
0,2	80,71%	57,68%	Hukum Perdata
0,3	75,00%	52,00%	Hukum Perdata
0,4	0,00%	0,00%	Hukum Perdata
0,5	0,00%	0,00%	Hukum Perdata

Gambar 13 Hasil Pengujian Hukum Perdata

Kategori Hukum Perdata menghasilkan nilai *threshold* rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,1 sebesar **87,32%**. Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **73,81%**.

**7. Hukum Internasional**

THRESHOLD	PRECISION	RECALL	KATEGORI SINGKAP
0,1	80,85%	71,42%	Hukum Internasional
0,2	82,61%	62,50%	Hukum Internasional
0,3	75,00%	30,46%	Hukum Internasional
0,4	0,00%	0,00%	Hukum Internasional
0,5	0,00%	0,00%	Hukum Internasional

Gambar 15 Hasil Pengujian Hukum Internasional

Kategori Hukum Internasional dan rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,1 sebesar **82,61%**. Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **71,42%**.

**5. HUKUM TATA NEGARA**

THRESHOLD	PRECISION	RECALL	KATEGORI SINGKAP
0,1	75,91%	100,00%	Hukum Tata Negara
0,2	85,00%	100,00%	Hukum Tata Negara
0,3	0,00%	0,00%	Hukum Tata Negara
0,4	0,00%	0,00%	Hukum Tata Negara
0,5	0,00%	0,00%	Hukum Tata Negara

Gambar 14 Hasil Pengujian Hukum Tata Negara

Kategori Hukum Internasional dan rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,1 sebesar **82,61%**. Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **71,42%**.

**8. Sosial dan Politik**

THRESHOLD	PRECISION	RECALL	KATEGORI SINGKAP
0,1	83,33%	70,31%	Sosial Politik
0,2	83,33%	51,28%	Sosial Politik
0,3	0,00%	0,00%	Sosial Politik
0,4	0,00%	0,00%	Sosial Politik
0,5	0,00%	0,00%	Sosial Politik

Gambar 16 Hasil Pengujian Sosial dan Politik

Kategori Sosial dan Politik dan rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,1 dan 0,2 sebesar **83,33%**. Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **70,31%**.

NO	JUDUL	PREKESIB	KLASIFIKASI	NILAI COSINE SIMILARITY	AKTIVITAS
17	Makalah Penelitian Hukum tentang dan Aspek	Penelitian Pengkajiannya	KAT50008	0,100121	Tes Hasil
16	Pengantar Akademi Hukum Islam tentang Hukum Islam	Penelitian Pengkajiannya	KAT50008	0,100052	Tes Hasil
15	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,099846	Tes Hasil
14	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,099639	Tes Hasil
13	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,099432	Tes Hasil
12	Aspek Hukum Islam tentang	Penelitian Pengkajiannya	KAT50008	0,099225	Tes Hasil
11	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,099018	Tes Hasil
10	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,098811	Tes Hasil
9	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,098604	Tes Hasil
8	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,098397	Tes Hasil
7	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,098190	Tes Hasil
6	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097983	Tes Hasil
5	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097776	Tes Hasil
4	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097569	Tes Hasil
3	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097362	Tes Hasil
2	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097155	Tes Hasil
1	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,096948	Tes Hasil

THRESHOLD	PRECISION	RECALL	KATEGORI (SNOPSIS)
0,1	87,32 %	68,12 %	Penelitian Pengkajiannya
0,2	85,29 %	56,86 %	Penelitian Pengkajiannya
0,3	75,00 %	12,00 %	Penelitian Pengkajiannya
0,4	0,00 %	0,00 %	Penelitian Pengkajiannya
0,5	0,00 %	0,00 %	Penelitian Pengkajiannya

Gambar 17 Hasil Pengujian Penelitian Pengkajian

Kategori Penelitian dan Pengkajiannya dan rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,1 sebesar **87,32%**. Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **73,81%**.

NO	JUDUL	PREKESIB	KLASIFIKASI	NILAI COSINE SIMILARITY	AKTIVITAS
120	Makalah Penelitian Hukum tentang Perundang-Undangan	Penelitian Pengkajiannya	KAT50008	0,100121	Tes Hasil
119	Makalah Penelitian Hukum tentang Perundang-Undangan	Penelitian Pengkajiannya	KAT50008	0,100052	Tes Hasil
118	Pengantar Akademi Hukum Islam tentang Hukum Islam	Penelitian Pengkajiannya	KAT50008	0,099846	Tes Hasil
117	Aspek Hukum Islam tentang	Penelitian Pengkajiannya	KAT50008	0,099639	Tes Hasil
116	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,099432	Tes Hasil
115	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,099225	Tes Hasil
114	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,099018	Tes Hasil
113	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,098811	Tes Hasil
112	Aspek Hukum Islam tentang	Penelitian Pengkajiannya	KAT50008	0,098604	Tes Hasil
111	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,098397	Tes Hasil
110	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,098190	Tes Hasil
109	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097983	Tes Hasil
108	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097776	Tes Hasil
107	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097569	Tes Hasil
106	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097362	Tes Hasil
105	Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,097155	Tes Hasil
104	Makalah Penelitian Hukum tentang Sistem Peradilan Pidana	Penelitian Pengkajiannya	KAT50008	0,096948	Tes Hasil

THRESHOLD	PRECISION	RECALL	KATEGORI (SNOPSIS)
0,1	74,00 %	68,12 %	UU
0,2	82,86 %	58,39 %	UU
0,3	0,00 %	0,00 %	UU
0,4	0,00 %	0,00 %	UU
0,5	0,00 %	0,00 %	UU

Gambar 18 Hasil Pengujian Perundang-Undangan

Kategori perundang-undangan dan rata-rata nilai *precision* tertinggi terletak pada *threshold* 0,2 sebesar **82,86%**. Kinerja *recall* tertinggi terletak pada *threshold* 0,1 sebesar **68,12%**.

### 4.3 Tabulasi Perbandingan Kinerja Rata-rata *Threshold*

Hasil dari keseluruhan sinopsis, akan membuat suatu tabulasi perbandingan dari 10 kategori yang diuji ke aplikasi.

Threshold	HAM Nasional		HAM Asing		Hukum pada Umumnya		Hukum Pidana		Hukum Perdata	
	presisi	recall	presisi	recall	presisi	recall	presisi	recall	presisi	recall
0,1	83,33 %	90,90 %	80,65 %	100,00 %	79,49%	80,00%	90,91 %	60,00 %	87,32 %	73,81 %
0,2	75,00 %	12,00 %	85,71 %	100,00 %	82,61%	66,66%	72,73 %	33,33 %	85,71 %	57,69 %
0,3	0 %	0 %	0 %	0 %	66,67%	44,44%	0 %	0 %	75,00 %	12,00 %
0,4	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0,5	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %

Threshold	Hukum Tata Negara		Hukum Internasional		Sosial dan Politik		Penelitian & Pengkajiannya		Perundang – undangan	
	presisi	recall	presisi	recall	presisi	recall	presisi	recall	presisi	recall
0,1	73,91 %	100,00 %	80,85 %	71,42 %	83,33 %	70,31 %	83,93 %	68,12 %	74,60 %	68,12 %
0,2	85,00 %	100,00 %	82,61 %	62,50 %	83,33 %	51,28 %	83,29 %	56,86 %	82,86 %	56,39 %
0,3	0 %	0 %	75,00 %	38,46 %	0 %	0 %	75,00 %	12,00 %	0 %	0 %
0,4	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0,5	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %

Gambar 14 Hasil Perbandingan Tabulasi *Threshold*

Evaluasi dari perbandingan 10 kategori sinopsis buku hasil pengujiannya yang ditunjukkan dalam gambar 14 dapat ditemukan bahwa nilai *precision* tertinggi yang telah diarsir sebesar **90,91%** pada *threshold* 0,1 dan nilai *recall* yang diarsir tertinggi sebesar **100%** pada *threshold* 0,1 dan 0,2. Kinerja rata-rata *threshold* yang ditunjukkan Tabel 5 menunjukkan bahwa aplikasi telah berhasil mengidentifikasi kebutuhan - kebutuhan klasifikasi sinopsis dari 10 kategori yang ada.

### 4.4 Akurasi

Hasil pada penelitian ini dari 120 data sinopsis buku yang telah berhasil di uji dengan 10 kategori dan setiap kategori memiliki kata kunci (*keyword*) masing-masing menggunakan aplikasi tersebut sehingga dapat menghasilkan tingkat akurasi metode *Cosine Similarity* dan Pembobotan *TF-IDF*.

$$\text{Tingkat akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

Sehingga dari 120 data yang di uji dengan 10 kategori yang dinyatakan benar dalam mengklasifikasikan sinopsis buku ada 97 record.

$$\text{Tingkat Akurasi} = \frac{82+15}{82+15+19+4} \times 100\% = 80,83\%$$

Berikut tampilan perhitungan akurasi keseluruhan :



**Gambar 15** Tampilan Akurasi Sistem Aplikasi

## KESIMPULAN DAN SARAN

Berdasarkan pengujian mendapatkan sebuah kesimpulan yang dapat diambil sebagai berikut :

1. Perbandingan keseluruhan dari pengujian aplikasi menunjukan bahwa nilai *precision* tertinggi sebesar 90,91% pada *threshold* 0,1 dengan kategori Hukum Pidana dan nilai *recall* tertinggi sebesar 100% pada *threshold* 0,1 dan *threshold* 0,2 dengan kategori HAM Asing dan Hukum Tata Negara.
2. Dari pengujian aplikasi yang telah dilakukan dengan 10 tipe kategori berbeda maka dapat ditemukan *threshold* yang terbaik adalah 0,1.
3. Dari sejumlah 120 data yang terdiri dari 10 kategori, aplikasi mampu melakukan proses analisa dengan sejumlah 97 data valid dan 23 data yang tidak valid, maka ketepatan akurasi klasifikasi aplikasi sebesar 80,83%.

Penelitian ini tentunya masih perlu banyak pengembangan sehingga terdapat beberapa saran untuk dilakukan

perbaikan maupun penambahan analisis maupun fitur aplikasi itu sendiri, berikut adapun saran yang bisa diteliti lebih lanjut adalah

1. Penelitian bisa menggunakan metode klasifikasi lainnya
2. Pada aplikasi, untuk penelitian selanjutnya bisa mengembangkan aplikasi tersebut dengan lebih efisien dalam mengeksekusi waktu untuk menghitung sebuah proses sistem klasifikasi dokumen buku.

## DAFTAR PUSTAKA

- Prastowo, A. (2013). Manajemen perpustakaan sekolah profesional. Yogyakarta: DIVA Press
- Mark A. Hall., & Lloyd A. Smith. (1999). *Feature Selection for Learning: Comparing a Correlation a Corelation-based Filter Approach to the Wrapper* . In FLAIRS Conference.
- Kaplan, R.M. (1995). *A Methode for Tokenizing Text*. Palo Alto Research Center (Festschrift in The Honor of Prof. Kimmo Koskenniemi's 60 th Anniversary).
- Porter, M.F.2001. *Snowball: A language for Stemming Algorithms Computer Laboraty*, Cambridge (England).
- Feldman, R dan Sanger, J., (2007). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Robertson, S.. (2005). *Understanding Inverse Document Frequency: On Theoretical Arguments for IDF*. England: Journal of Documentation, Vol. 60, 502-520.
- Sulistyo, W. & Sarno, R. (2008), *Auto Matching Antar Dokumen dengan Metode Cosine Measure*, Seminar

Nasional Teknologi Informasi dan  
Komunikasi, Indonesia.