

APLIKASI PENENTUAN DOSEN PENGUJI SKRIPSI MENGGUNAKAN METODE TF-IDF DAN VECTOR SPACE MODEL

Riki Ruli A. Siregar¹, Fera Amelia Sinaga², Rakhmat Arianto³
Program Studi Sarjana Teknik Informatika, Sekolah Tinggi Teknik PLN Jakarta
Jalan Lingkar Luar Barat Menara PLN Cengkareng, Jakarta Barat 11750
e-mail : riki.ruli@sttpln.ac.id¹, ameliafera@gmail.com², arianto@sttpln.ac.id³

Abstrak

Pada Sekolah Tinggi Teknik PLN (STT-PLN) penentuan dosen penguji tugas akhir atau skripsi merupakan tugas dari sekretaris jurusan. Penelitian ini bertujuan untuk memberikan alternative untuk menentukan dosen penguji skripsi. Metode yang di terapkan untuk membangun system ini adalah text mining, TF-IDF dan Vector Space Model (VSM). Text mining untuk melakukan processing data, dimana data yang akan diproses adalah judul dan abstrak skripsi, sedangkan VSM untuk melakukan pengklasifikasian kompetensi, penelitian ini dapat merekomendasikan tiga dosen untuk menjadi dosen penguji skripsi berdasarkan kecocokan antara judul dan abstrak dengan klasifikasi Pada penelitian ini, penulis menggunakan Model pengembangan perangkat lunak CRISP-DM. Adapun fase yang dimiliki oleh CRISP-DM adalah fase pemahaman bisnis, fase pemahaman data, fase pengolahan data, fase permodelan, fase evaluasi dan fase penyebaran. Hasil dari penelitian ini memiliki akurasi 93,22%.

Kata Kunci : Text mining, Vector Space Model, TF-IDF, klasifikasi, CRISP-DM.

Abstract

At the Technical College of PLN (STT-PLN), the determination of the lecturer of the final project or thesis is the task of the secretary of the department. This study aims to provide an alternative to determine thesis examiners lecturer. The method applied to build this system is text mining, TF-IDF and Vector Space Model (VSM). Text mining to perform data processing, where the data to be processed is the title and abstract of the thesis, while the VSM to perform the classification of competence, this study can recommend three lecturers to be a thesis examiner lecturer based on the fit between title and abstract by classification In this study, CRISP-DM software development Model. The phases of CRISP-DM are the business understanding phase, the data holding phase, the data processing phase, the Modeling phase, the evaluation phase and the deployment phase. The results of this study have an accuracy of 93,22%.

Keywords : Text mining, Vector Space Model, TF-IDF, Clasification, CRISP-DM

1. PENDAHULUAN

Tugas akhir Skripsi merupakan perwujudan kualitas akademik mahasiswa dan sebagai penelitian syarat kelulusan. Dalam pengujian skripsi, penentuan dosen penguji merupakan hal yang penting, idealnya dosen penguji memiliki kompetensi yang sesuai dengan topik skripsi yang diteliti oleh mahasiswa agar didapat hasil pengujian yang maksimal.

Dalam kasus ini peneliti mengambil tempat penelitian di Teknik Informatika STT-PLN, dimana proses pemilihan dosen penguji dipilih secara langsung oleh Sekretaris jurusan dengan di data pada aplikasi *Microsoft Excel*. Tidak adanya sistem yang digunakan menjadikan tugas sekretaris jurusan dalam menentukan dosen penguji yang sesuai antara judul atau tema tugas akhir dengan konsentrasi dosen penguji menjadi kurang akurat dan membutuhkan waktu proses. Dalam proses pelaksanaan pengujian atau sidang skripsi terdapat 3 (tiga) orang dosen penguji yang terdiri dari ketua penguji, sekretaris penguji dan penguji.

Syarat utama kompetensi dosen penguji yang sesuai dengan topik yang diangkat dalam skripsi, pada penelitian memberikan alternatif keputusan dalam menentukan dosen penguji skripsi yang akan di ujikan memiliki kesesuaian terhadap kompetensi. Penerapan *text mining* dapat digunakan untuk menganalisa data yang diperoleh dari judul dan abstrak skripsi mahasiswa. Untuk menerapkan *text mining*, dibutuhkan skripsi mahasiswa yang sebelumnya, sehingga dapat melihat pola-pola data yang terdahulu untuk dijadikan sebuah pengetahuan baru. Penerapan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) dan VSM (*Vector Space Model*) yang dapat digunakan untuk mengklasifikasikan dosen penguji yang sesuai antara topik skripsi dengan kompetensi dosen, yang dilakukan dengan ekstraksi teks pada dokumen. Penelitian sebelumnya yang dilakukan sebelumnya koordinator tugas akhir dalam menentukan dosen pembimbing dan penguji skripsi [1]. Hasil penelitian membantu koordinator tugas akhir dalam menentukan dosen pembimbing dan penguji skripsi. Metode-metode yang digunakan untuk membangun sistem ini adalah *Text Mining*, *k-Nearest Neighbor* (k-NN), dan *Simple Additive Weighting* (SAW).

2. KAJIAN PUSTAKA

Penelitian yang berjudul “*Analysis of Vector Space Model in Information Retrieval*” [2], Menyajikan pendekatan yang berbeda dari model ruang vektor untuk menghitung nilai kesamaan hit dari mesin pencari dan yang lebih penting lagi, dirasakan bahwa penyelidikan ini akan menghasilkan pemahaman yang lebih jelas mengenai masalah dan masalah dalam menggunakan model ruang vektor dalam informasi. Pada penelitian ini menganalisis tiga pendekatan model ruang vektor untuk query uji coba. Nilai kemiripan dihitung dengan menggunakan tiga pendekatan model ruang vektor. Setelah mempertimbangkan persyaratan pembobotan dalam pengumpulan dokumen, kita dapat menghitung nilai kesamaan antara kueri dan dokumen. Peringkat dokumen tergantung pada nilai kesamaan nilai yang dihitung dengan pendekatan yang berbeda. Peringkat dokumen tergantung pada nilai kesamaan nilai yang dihitung dengan pendekatan yang berbeda Dari VSM. Nilai kesamaan yang dihitung dengan model penghitung waktu bagus untuk dokumen panjang namun model TF-IDF dan normalisasi memberikan hasil yang lebih baik. Model normalisasi juga menggunakan skema bobot global dan memberikan hasil yang sama untuk dokumen panjang dibandingkan dengan model TF-IDF. Jadi ketiga pendekatan model ruang vektor mungkin mendukung dokumen panjang yang berisi lebih banyak tampilan istilah kueri.

Penelitian selanjutnya, berjudul “*A Text Categorization Method using Extended Vector Space Model by Frequent Term Sets*” [3]. Penelitian ini membahas tentang pengkategorian teks, Sebuah batasan baru AD-Sup diperkenalkan untuk mengekstrak fitur diskriminatif dari rangkaian istilah yang sering digunakan untuk tugas klasifikasi. Hasil klasifikasi pada korpus *Reuters-21578* dan *WebKB* menunjukkan bahwa kendala AD-Sup efektif untuk mengekstrak fitur yang berguna dan strategi kombinasi efektif untuk membangun ruang fitur yang lebih baik dan memperbaiki klasifikasi.

Pertambahan web mencakup tiga bagian yaitu penambahan struktur web, penambahan konten web, penambahan penggunaan web. Tujuan dari proses analisis lanjutan ini adalah untuk mengekstrak informasi dari kumpulan data dan mengubahnya menjadi struktur yang mudah dimengerti untuk digunakan lebih jauh. Data mining terdiri dari tiga langkah dasar Ekstrak informasi, memuat informasi & menampilkan informasi (*out-put*)[4]

2.1 Text Mining

Text mining merupakan salah satu bidang khusus dari data mining. *Text mining* dapat didefinisikan sebagai suatu proses menggali informasi dimana seseorang user berinteraksi dengan sekumpulan dokumen menggunakan *tool* analisis yang merupakan komponen-komponen dalam data mining [5]. *Text mining* digunakan untuk mengolah dokumen sebelum dilakukan proses *similarity*. *Text mining* merupakan proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. *Text mining* memiliki tujuan dan menggunakan proses yang sama dengan data mining, namun memiliki masukan yang berbeda. Masukan untuk *Text mining* adalah data yang tidak (atau kurang) terstruktur, seperti dokumen *Word*, PDF, kutipan teks, dll., sedangkan masukan untuk data mining adalah data yang terstruktur.

2.2 Text Preprocessing

Didalam proses *Text mining* terdapat proses *Text Processing*[6]. *Preprocessing text* merupakan tindakan menghilangkan karakter-karakter tertentu yang terkandung dalam dokumen, seperti koma, tanda petik dan lain-lain serta mengubah semua huruf kapital menjadi huruf kecil. Selain itu, dalam tahap *text preprocessing* ini dilakukan *tokenization*. *Tokenization* merupakan proses pengolahan token yang terdapat dalam rangkaian teks, sehingga dokumen akan dipecah-pecah menjadi term [7]. *Text mining* dalam prakteknya mencari pola-pola tertentu, mengasosiasikan suatu bagian teks dengan yang lain berdasarkan aturan-aturan tertentu, kata-kata yang dapat mewakili sehingga dapat dilakukan analisa keterhubungan antar satu dengan yang lain [5]. Berikut tahapan-tahapan proses didalam *text mining*:

a. *Tokenizing*

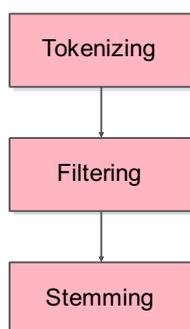
Tahap *Tokenizing* adalah tahap pemotongan tiap kata dalam kalimat atau parsing dengan menggunakan spasi sebagai delimiter yang akan menghasilkan token berupa kata. Pada *tokenizing* terdapat beberapa proses yang harus dilakukan yaitu merubah semua huruf besar menjadi kecil (*text to lowercase*). Proses selanjutnya adalah penguraian, proses penguraian yang dimaksud adalah membagi *text* menjadi kumpulan kata tanpa memperhatikan keterhubungan antara kata satu dengan kata lain serta peran dan posisinya pada kalimat.

b. *Filtering*

Tahap *Filtering* adalah tahap penyaringan kata yang didapat dari *Tokenizing* yang dianggap tidak penting atau tidak memiliki makna dalam proses *Text mining* yang disebut *stopword*. *Stopword* berisi kata-kata umum yang sering muncul dalam sebuah dokumen dalam jumlah banyak namun tidak memiliki kaitan dengan tema tertentu. Contoh *stopwords* adalah “yang”, “di”, ‘yang’, dll.

c. *Stemming*

Tahap *stemming* adalah tahap mengembalikan kata-kata yang diperoleh dari hasil *Filtering* ke bentuk dasarnya, menghilangkan imbuhan awal (prefix) dan imbuhan akhir (sufix) sehingga didapat kata dasar.



Gambar 1 Text Preprocessing

2.3 Pembobotan Kata (*Term Weighting*)

Pembobotan kata sangat berpengaruh dalam menentukan keiripan antara dokumen dengan *query*. Apabila bobot tiap kata dapat ditentukan dengan tepat, diharapkan hasil perhitungan kemiripan teks akan menghasilkan perankingan dokumen yang baik. Keberhasilan dari model ruang *vector* ditentukan oleh skema pembobotan terhadap suatu *term* baik untuk cakupan local maupun global, dan faktor normalisasi [8]. Pembobotan local hanya berpedoman pada frekuensi munculnya *term* dalam satu dokumen dan tidak melihat kemunculan *term* tersebut di dalam dokumen lainnya. Faktor yang memegang peranan penting dalam pembobotan kata yaitu :

1. *Term Frequency (tf)*

Pendekatan dalam pembobotan local yang paling banyak diterapkan adalah *term frequency* (tf). Factor ini menyatakan banyaknya kemunculan suatu kata dalam suatu dokumen. Semakin sering suatu kata muncul dalam sebuah dokumen, berarti semakin penting kata tersebut. Ada empat cara yang bisa digunakan untuk mendapatkan nilai TF:

a. *Raw Tf*

Nilai Tf sebuah *term* dihitung berdasarkan kemunculan *term* tersebut dalam dokumen.

b. *Logarithmic Tf*

Dalam memperoleh nilai Tf, cara ini menggunakan fungsi logaritmik dalam matematika.

$$TF = 1 + \log(TF)$$

c. *Binnary Tf*

Cara ini menghasilkan nilai *Boolean* berdasarkan kemunculan *term* pada dokumen tersebut. Akan bernilai 0 apabila *term* tidak ada pada sebuah dokumen, dan bernilai 1 apabila *term* tersebut ada dalam dokumen. Sehingga banyaknya kemunculan *term* pada dokumen tidak berpengaruh.

d. *Augmented Tf*

$$TF = 0.5 + 0.5 \times TF_{\max}(TF)$$

Nilai TF adalah jumlah kemunculan *term* pada sebuah dokumen. Nilai $\max(Tf)$ adalah jumlah kemunculan terbanyak *term* pada dokumen yang sama.

Perhitungan Tf yang akan digunakan dalam implementasi sistem temu kembali informasi pada sistem yang penulis bangun adalah *Raw Tf*.

2. *Inverse Dokumen Frequency (IDF)*

Metode TF-IDF (*Term Frequency Inverse Document Frequency*) merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata didalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut [9]. Formula yang digunakan pada *term frequency* (tf), terdapat yaitu nilai tf diberikan berdasarkan jumlah kemunculan suatu kata di dokumen. Idf dihitung dengan formula sebagai berikut :

$$IDF = \text{Log} \frac{D}{Df}$$

Dimana :

IDF = Invers dokumen frekuensi

D = Total Dokumen

Df = Frekuensi dokumen dari *term*

Log = Untuk memperkecil pengaruhnya relative terhadap tf

Bobot *term* dihitung dengan menggunakan rumus :

$$W = \text{tf} \times \text{idf}$$

Dimana :

W = Bobot dokumen

Tf = *Term* frekuensi

Idf = Invers dokumen frekuensi

2.4 Vector Space Model

Vector Space Model (VSM) sering digunakan untuk mempresentasikan sebuah dokumen dalam ruang vector. VSM merupakan model *Information Retrieval* yang mempresentasikan dokumen dan *query* sebagai vektor pada ruang multidimensi. Kesamaan suatu dokumen dengan *query* dapat diukur dengan vektor dokumen dan vektor *query* [7].

Karakteristik model ruang *vector* antara lain :

1. Model *vector* berdasarkan *keyterm*.
2. Model *vector* mendukung *partial matching* (sebagian sesuai) dan penentuan peringkat dokumen.
3. Prinsip dasar model *vector* adalah :
 - a. Dokumen direpresentasikan dengan menggunakan *vector keyterm*.
 - b. Ruang dimensi ditentukan oleh *keyterms*.
 - c. *Query* direpresentasikan dengan menggunakan *vector keyterm*.
 - d. Kesamaan dokumen *keyterm* dihitung berdasarkan jarak *vector*.
4. Model ruang *vector* memerlukan :
 - a. Bobot *keyterm* untuk *vector* dokumen.
 - b. Normalisasi *keyterm* untuk *vector* dokumen.
 - c. Normalisasi *keyterm* untuk *vector query*.
 - d. Perhitungan jarak untuk *vector* dokumen *keyterm*.
5. Kinerja model ruang *vector* :
 - a. Efisien.
 - b. Mudah dalam representasi.
 - c. Dapat diimplementasikan pada *document matching* dan *partial matching*.

Prosedur model ruang *vector* dapat dikelompokkan menjadi tiga tahap yaitu:

1. Pengindeksan dokumen.
2. Pembobotan indeks, untuk menghasilkan dokumen yang relevan.
3. Memberikan peringkat dokumen berdasarkan ukuran kesamaan (*similarity measure*).

Perhitungan kemiripan antara vektor dokumen dan vektor *query* dilihat dari sudut yang paling kecil. Sudut yang dibentuk oleh dua buah vektor dapat dihitung dengan melakukan inner product. Kemiripan antara vektor dokumen dan vektor *query* akan dihitung dengan pendekatan *Cosine Similarity* [8] Nilai relevansi (*similarity*) antara *query* dengan dokumen ke-j [9] adalah :

$$\text{CosSim}(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} = \frac{\sum_{i=1}^t (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t w_{ij}^2 \cdot \sum_{i=1}^t w_{iq}^2}}$$

Dimana :

\vec{d}_j = Bobot dokumen j

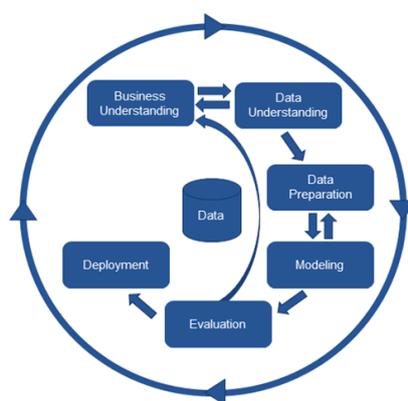
\vec{q} = Bobot *query*

- $|\vec{d}_j|$ = Panjang bobot dokumen j
- $|\vec{q}|$ = Panjang bobot *query*
- w_{ij} = Bobot *term* i pada dokumen j
- w_{iq} = Bobot *term* i pada *query*
- T = Teks

2.5 CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) merupakan suatu standar yang telah dikembangkan pada tahun 1996 yang ditujukan untuk melakukan proses analisis dari suatu industri sebagai strategi pemecahan masalah dari bisnis atau unit penelitian.

Terdapat enam fase pada CRISP-DM yaitu sebagai berikut :



Gambar 2 Proses CRISP-DM (Sumber: <http://www.kdnuggets.com>)

- a. Fase Pemahaman Bisnis (*Business Understanding Phase*)
- b. Fase Pemahaman Data (*Data Understanding Phase*)
- c. Fase Pengolahan Data (*Data Preparation Phase*)
- d. Fase Pemodelan (*Modeling Phase*)
- e. Fase Evaluasi (*Evaluation Phase*)
- f. Fase Penyebaran (*Deployment Phase*)

3 METODE PENELITIAN

3.1 Analisa Kebutuhan

Analisa kebutuhan merupakan tahapan awal penelitian. Pada tahap ini dilakukan analisa kebutuhan suatu data menggunakan metode *CRISP-DM*, dimana data akan diproses dengan melakukan beberapa tahapan seperti berikut:

3.1.1 Business Understanding Phase (Fase Pemahaman Bisnis)

Pada tahap ini akan di lakukan penentuan tujuan penelitian dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan, pada penelitian ini memiliki tujuan untuk menciptakan alternative penentuan dosen penguji skripsi secara otomatis dan mengimplementasikan metode *text* mining yaitu TF-IDF dan VSM dalam menentukan dosen penguji skripsi yang sesuai antara kompetensi dosen dengan klasifikasi skripsi mahasiswa, dimana klasifikasi skripsi tersebut terdiri dari tiga kategori yaitu, *sistem informasi (si)*, *jaringan* dan *embedded*. Penelitian ini dilaksanakan di Jurusan Teknik Informatika STT-PLN, waktu penelitian serta perancangan mulai pada bulan Maret 2017 hingga selesai.

3.1.2 Data Understanding Phase (Fase Pemahaman Data)

Pada tahap ini adalah tahap pengumpulan data, data diperoleh dari perpustakaan STT-PLN dan jurusan Teknik Informatika STT-PLN. Pada penelitian ini data yang digunakan adalah 59 data dari skripsi mahasiswa STT-PLN. Data yang diambil dari data skripsi adalah judul dan abstrak, yang mana judul dan abstrak tersebut akan diolah kembali agar menghasilkan suatu bobot tertentu. Data skripsi yang diambil merupakan data dari tahun 2013-2015, data dosen penguji dan mahasiswa juga dibutuhkan dalam proses penentuan ini, data dosen penguji diperlukan untuk menentukan dosen penguji yang tepat atau sesuai dengan klasifikasi skripsi

a. Data skripsi yang akan digunakan

Tabel 1 Data set skripsi 2013 -2015

No	NIM	Nama	Pembimbing 1	Pembimbing 2	Judul	Abstrak
1	201031155	Prayudi wiranata	Indrianto	Puji Catur Siswiprapt ini	Perancangan Model Robot Pemindah Barang berupa Kotak (Box) yang Berjalan Mengikuti Line Followers dengan Menggunakan Arduino	Pada perancangan robot yang berfungsi sebagai mobil otomatis pemindah barang berbasis mikrokontroler Arduino uno. Perancangan gerakannya dilengkapi 2 motor DC yang dapat berputar secara differential. Putaran arah motor yang dirancang pada robot menggunakan driver motor DC. Untuk gerak maju, belok, ataupun mundur digunakan 2 motor DC di sisi kiri dan kanannya. Robot ini dilengkapi sensor LDR (light dependet resistor) dan 2 Lampu LED sebagai penerang Jalur (Line) Sensor ini mempunyai 3 fungsi, yakni sebagai pendeteksi jalur lintasan pergerakan, pendeteksi tempat tujuan persediaan barang dan pendeteksi jenis barang yang akan dipindahkan. Robot hanya dapat bergerak pada lintasan acuan berupa garis hitam dengan lebar 1.5 cm. Keseluruhan sub sistem yang dibangun pada robot dikendalikan oleh mikrokontroler, sehingga secara operasionalnya robot dapat optimal mendeteksi adanya tempat persediaan barang dan dapat memindahkannya ke tempat yang dituju sesuai dengan program. Instruksi perintah input pada mikrokontroler dengan menggunakan aplikasi Arduino Uno IDE Versi 1.5.8. Dengan tegangan 5 volt pada penggerak jalan robot didapat kecepatan rata-rata dan stabil. Untuk robot yang berbelok stabil pada jalur lintasan. Tingkat keberhasilan robot memindahkan barang sesuai tujuan mencapai 95%.
2	201131042	Desy Aisyah	Riki Ruli Affandi Siregar	Efy Yosrita	Rancang Bangun Alat Peraga Modul Praktikum Embedded Sistem Menggunakan Raspberri Pi	Penelitian ini bertujuan untuk membuat suatu alat peraga beserta modul praktikum <i>embedded</i> system untuk menunjang pelaksanaan praktikum <i>embedded</i> system di sekolah tinggi teknik PLN Jakarta menggunakan modul praktikum berbasis open source. Dalam penelitian pembuatan modul <i>embedded</i> system menggunakan raspberri pi dan mikrokontroler arduino menggunakan metode prototyping. Penelitian dimulai dengan identifikasi masalah, pengumpulan data, dan analisis, kemudian perancangan perangkat lunak dan perangkat keras. Implementasi yaitu melakukan uji coba langsung kepada praktikan. Penelitian ini menghasilkan suatu modul praktikum <i>embedded</i> system yang dapat digunakan untuk menunjang pelaksanaan praktikum <i>embedded</i> system. Modul praktikum yang dihasilkan ini diharapkan dapat diimplementasikan dengan baik dan benar.

b. Data Dosen penguji

Data Penguji adalah nama-nama dosen yang dapat menjadi penguji di Teknik Informatika STT-PLN, dosen penguji terbagi menjadi tiga bagian yaitu, ketua penguji, sekretaris dan anggota penguji. Dosen yang dapat menjadi ketua penguji adalah dosen dengan jabatan lektor kepala dan lektor, untuk sekretaris, dosen yang dapat menjadi sekretaris adalah dosen dengan jabatan asisten ahli, sedangkan yang dosen yang digunakan dapat dilihat pada tabel 2 dibawah ini:

Tabel 2 Data ketersediaan dosen penguji

NO	NIM	NAMA	JABATAN	KOMPETENSI
1	0326058501	Yessy Fitriani	Asisten_Ahli	jaringan
2	0315035301	Wisnu Hendro martono	Asisten_Ahli	embedded
3	0331057901	Rakhmadi Irfansyah Putra	Asisten_Ahli	jaringan
4	0325076202	Dwina Kuswardani	Asisten_Ahli	si
5	0815058202	Abdul Haris	Asisten_Ahli	si
6	0303117101	Yasni Djamain	Lektor	si
7	0315097901	Rizqia Cahyaningtyas	Lektor	si
8	0301027703	Puji Catur Siswipraptini	Lektor	si
9	0314127401	Efy Yosrita	Lektor	si
10	0329098303	Dian Hartanti	Lektor	si
11	0301097501	Indrianto	Lektor	embedded
12	0307117301	Iriansyah	Lektor	embedded
13	0320047601	Sely karmila	Lektor	si
14	0328087401	Dewi Arianti Wulandari	Lektor	si
15	0313107601	Yessy Asri	lektor_kepala	si
16	0318057601	Meilia Nur Indah Susanti	lektor_kepala	si
17	0320086401	Darma Rusjdi	lektor_kepala	embedded
18	0316067501	Luqman	lektor_kepala	embedded
19	0415078201	Rahma Farah Ningrum	Asisten_Ahli	si
20	0301057706	Riki Ruli Affandi Siregar	Asisten_Ahli	embedded
21	0310068702	Abdurrasyid	tenaga_pengajar	si
22	0316088002	Herman Bedi Agtriadi	tenaga_pengajar	embedded
23	0308018702	Rakhmat Arianto	tenaga_pengajar	si
24	0326118901	Dine Tiara Kusuma	tenaga_pengajar	si
25	0328029002	M. Yoga Distra Sudirman	tenaga_pengajar	jaringan
26	0413057701	Hendra Djatnika	tenaga_pengajar	si
27	0328127302	Rosida Nur Aziza	tenaga_pengajar	embedded
28	0313058902	Karina Djunaidi	tenaga_pengajar	si

c. Data Mahasiswa

Data mahasiswa dibawah ini adalah data mahasiswa yang digunakan dalam penelitian ini.

Tabel 3 Sample Data Mahasiswa

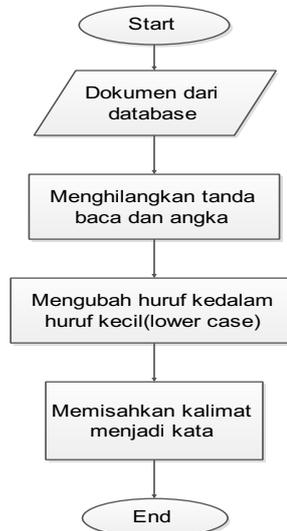
NO	NIM	NAMA	PEMBIMBING 1	PEMBIMBING 2
1	201031155	Prayudi wiranata	Indrianto	Puji Catur Siswipraptini
2	201031066	Satria siswanto nugroho	Yasni Djamain	Rahma Farah Ningrum
3	200931006	Ardi yanuari akbar	Wisnu Hendro Martono	Sely karmila
4	200831025	Jefri Toding Lamba	Puji Catur Siswipraptini	Rakhmat Arianto
5	201131042	Desy Aisyah	Riki Ruli Affandi Siregar	Iriansyah
6	201131133	Arjumiati	Dwina Kuswardani	Rahma Farah Ningrum
7	201231006	Lidya intan	Efy Yosrita	Meilia Nur Indah Susanti
8	201231008	Anita Andiani	Indrianto	Iriansyah
9	201231009	Rifaldi Andrian	Yasni Djamain	Wisnu Hendro martono
10	201231012	Arbu Marna	Iriansyah	Rakhmat Arianto
11	201231014	Nindy Navtalia Sihite	Rakhmat Arianto	Rahma Farah Ningrum
12	201231015	Adithya Marhendra Kusuma	Iriansyah	Riki Ruli Affandi Siregar
13	201231016	Elsa Aprianis M	Puji Catur Siswipraptini	Riki Ruli Affandi Siregar

3.1.3 Data Preparation Phase (Fase Pengolahan Data)

Pada tahap ini merupakan persiapan data awal untuk diimplementasikan pada sebuah perangkat permodelan atau *Data Transformation*, dari penelitian ini 59 data skripsi yang di dapat akan di proses dan dihasilkan jenis kategorinya yang nanti akan dicocokkan dengan kompetensi dosen penguji. Dokumen skripsi akan diolah menggunakan metode *text preprocessing*. *Text preprocessing* merupakan Tahap proses awal terhadap dokumen skripsi untuk mempersiapkan dokumen menjadi dokumen yang akan diolah lebih lanjut. Tahapan yang dilakukan secara umum adalah *tokenizing, filtering, stemming*.

a. *Tokenizing*

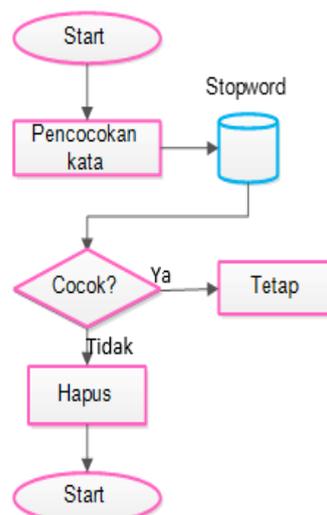
Pada proses *tokenizing* dilakukan pembacaan dokumen yang dimiliki dan memisahkan deretan kata didalam kalimat, paragraph atau dokumen menjadi token atau potongan kata tunggal. *Tokenization* dapat dilakukan dengan menghilangkan tanda baca dan memisahkannya per spasi. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua *token* ke bentuk huruf kecil (*lower case*). Alur pada *tokenizing* dapat dilihat dari flowchart dibawah ini



Gambar 3 Flowchart Tokenizing

b. *Filtering*

Pada tahapan ini dilakukan pengambilan kata-kata penting dari hasil *tokenizing* dengan membuang *stopwords* (kata-kata yang kurang penting). *Stopword* dapat berupa subjek atau kata penghubung. *Filtering* dilakukan dengan menentukan *term* mana yang akan digunakan untuk merepresentasikan dokumen sehingga dapat mendeskripsikan isi dokumen dan *membedakan* dokumen tersebut. Alur dari *Filtering* dapat dilihat pada flowchart dibawah ini:



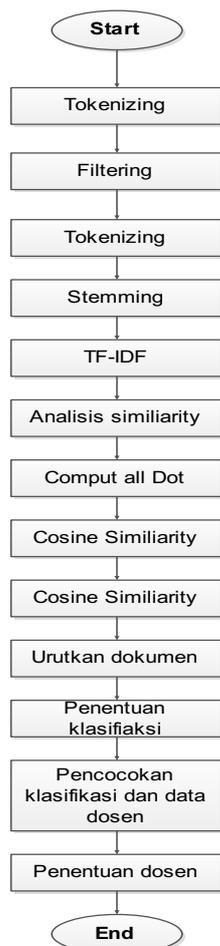
Gambar 4 Diagram Alir *Filtering*

c. *Stemming*

Stemming merupakan suatu proses yang terdapat dalam sistem IR yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu. Pada tahap ini akan dicari root kata dari tiap kata hasil *Filtering*.

3.1.4 Modelling Phase (Fase Permodelan)

Pada tahap ini data yang telah dilakukan text preprocessing akan di proses ke tahap selanjutnya dengan beberapa permodelan. Permodelan ini dilakukan untuk mengoptimalkan hasil yang ingin dicapai. Metode yang digunakan dalam penelitian ini adalah TF-IDF dan VSM, Pada tahap ini merupakan tahap pembobotan dan pengklasifikasian data skripsi, dokumen yang sudah di proses dengan *text preprocessing* akan diberi bobot dan diliat jarak kemiripannya. Didalam penulisan ini ada tiga kategori yaitu sistem informasi, *embedded* dan jaringan dimana *keyword* dalam klasifikasi tersebut sudah ditentukan oleh penulis terlebih dahulu. alur atau *flowchart* metode tersebut dapat dilihat dari gambar disamping ini:



Gambar 5 Diagram Alir TF-IDF dan VSM

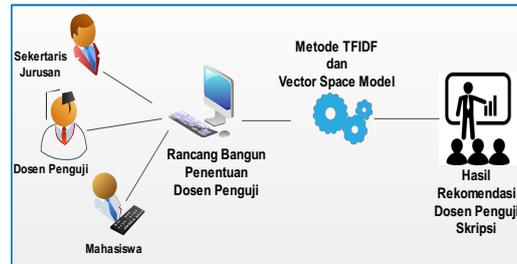
3.1.5 Evaluation Phase (Fase Evaluasi)

Fase evaluasi merupakan tahap untuk mengetahui apakah model yang di rancang telah sesuai atau belum dengan tujuan pada fase awal. Tujuan awal di dirancangnya model ini yaitu agar menghasilkan nilai akurasi yang tinggi, sehingga dapat membuktikan bahwa penelitian yang dilakukan telah berhasil. Pada penelitian ini dilakukan beberapa cara untuk melihat nilai akurasi yang dihasilkan. Peneliti menggunakan cara perbandingan antara sistem yang telah dibangun dengan klasifikasi penentuan skripsi yang telah dilakukan berdasarkan hasil wawancara. diperoleh nilai akurasi antara sistem dan hasil dari wawancara memiliki nilai yang sama yaitu sebesar 93,22%. dengan tingkat *error* sebesar 6,78%. Hasil penjelasan perhitungan dijelaskan pada bab pembahasan. Dengan hasil yang diperoleh sama, sehingga dapat dibuktikan bahwa aplikasi yang dibangun dapat digunakan karena menghasilkan nilai akurasi yang tinggi.

4 HASIL DAN PEMBAHASAN

4.1 Hasil Rancangan Antarmuka

Rancang Bangun Penentuan Dosen Penguji Skripsi Dengan Menggunakan Metode TF-IDF Dan *Vector Space Model* dapat digunakan oleh sekretaris jurusan, dosen dan mahasiswa. Secara garis besar, desain arsitektur Rancang Bangun Penentuan Dosen Penguji Skripsi dapat dilihat pada gambar dibawah ini:

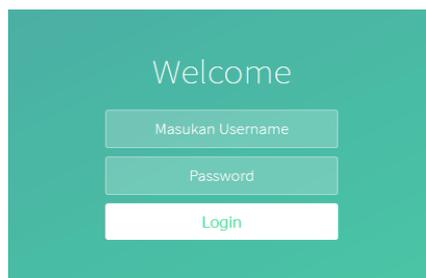


Gambar 6 Desain Arsitektur Penentuan Dosen Penguji

Gambar 6 menjelaskan desain arsitektur rancang bangun penentuan dosen penguji. Rancang bangun tersebut dapat diakses oleh tiga jenis user, yaitu: Sekertaris Jurusan sebagai admin, mahasiswa dan dosen. Setiap *user* memiliki hak akses tersendiri terhadap aplikasi. Metode TF-IDF dan *Vector Space Model* di implementasikan kedalam rancang bangun untuk menentukan pilihan atau hasil rekomendasi dari dosen penguji yang sesuai antara kompetensi dosen dan topik skripsi. Berikut adalah gambaran dan rincian kegunaan aplikasi:

1. Tampilan Login

Saat aplikasi pertama kali dijalankan, *user* akan diminta untuk memasukkan *username* dan *password* hal ini digunakan untuk validasi apakah *user* terdaftar dalam *database* atau tidak. *User* yang dapat *login* adalah *user* yang telah terdaftar sebagai dosen dan mahasiswa mengikuti sidang skripsi. Dalam aplikasi ini terdapat tiga buah *user* yaitu sekretaris jurusan (sekjur), mahasiswa dan dosen, masing-masing akan memiliki menu yang berbeda. Interface halaman login dapat dilihat pada gambar dibawah ini:



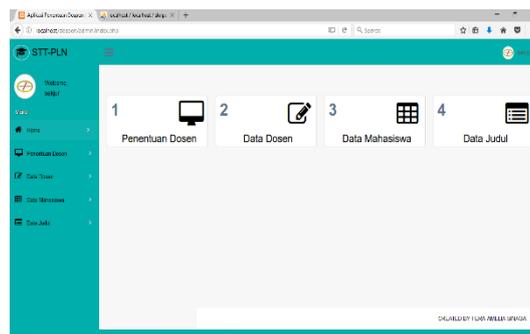
Gambar 7 Halaman Login

Setelah dilakukan validasi dan *username* dan *password* sesuai dengan data yang ada, maka *user* akan masuk ke halaman dan bisa mengakses menu yang ada didalamnya.

2. Tampilan Menu Sekretaris Jurusan

a. Halaman Utama

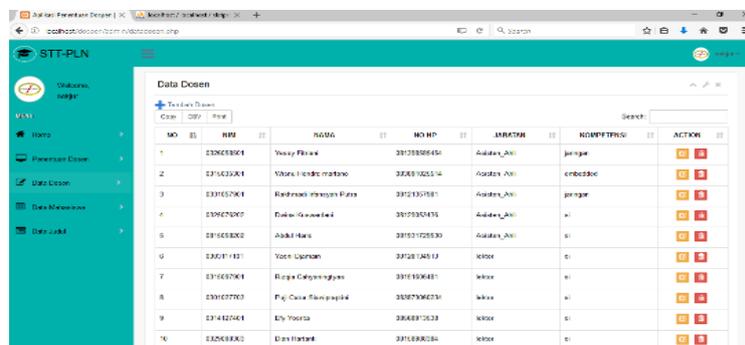
Pada halaman utama menampilkan empat buah *icon*, dimana jika pada *icon* di *click* maka *user* akan masuk kedalam tampilan *form* yang *icon*nya telah dipilih.



Gambar 8 Halaman Utama Admin

b. Halaman Data Dosen

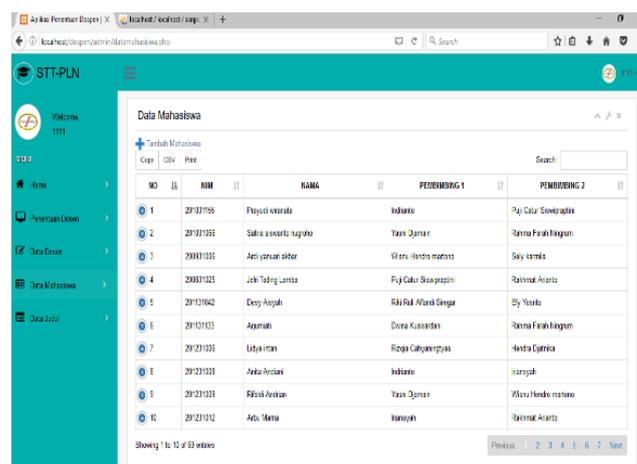
Pada halaman data dosen, terdapat tabel Informasi data dosen seperti nim, nama, nomor hp, jabatan, kompetensi dan *action*. Sekertaris jurusan memiliki hak untuk menambah, mengedit dan menghapus data dosen. Pada halaman dosen ini dilengkapi juga fitur *import* kedalam csv dan print.



Gambar 9 Halaman Data Dosen

c. Halaman Data Mahasiswa

Pada halaman data mahasiswa, terdapat tabel Informasi data mahasiswa seperti nim, nama, pembimbing 1 dan pembimbing 2. Sekertaris jurusan memiliki hak untuk menambah, mengedit dan menghapus data mahasiswa. Pada halaman dosen ini dilengkapi juga fitur *import* kedalam csv dan *print*.



Gambar 10 Halaman Data Mahasiswa

d. Halaman Data Judul

Pada halaman data mahasiswa, terdapat tabel Informasi judul, dimana terdapat data nik, id judul judul dan *action*, data ini didapat dari hasil penginputan judul dan abstrak yang sudah dilakukan oleh mahasiswa.

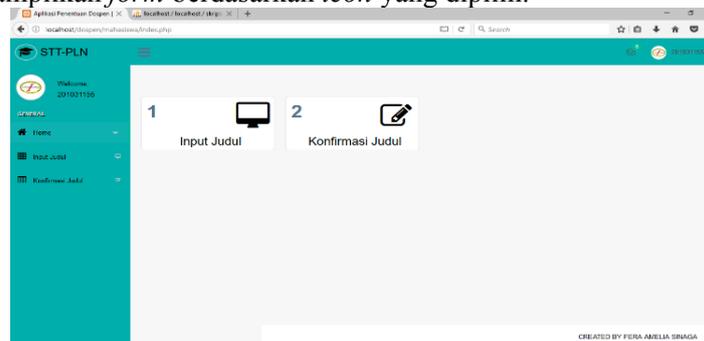
NO	NIK	ID	JUDUL	ACTION
1	201031155	2010311552017	PERANCANGAN MODEL ROBOT PEMINDAH BARANG BERUPA KOTAK (BOX) YANG BERJALAN MENGGUNAKAN LINE FOLLOWERS DENGAN MENGGUNAKAN ARDUINO	[Edit] [Hapus]
2	201031066	2010310662017	RANCANG BANGUN SISTEM INFORMASI PENERIMAAN, PENGAMBILAN DAN STATUS LAUNDRY BERBASIS ANDROID PHONEGAP, WEB DAN SMS GATEWAY PADA LAUNDRY WAN VILEME	[Edit] [Hapus]
3	200931006	2009310062017	SISTEM PENDUKUNG KEPUTUSAN TENDER PROYEK PENGADAAN BARANG MENGGUNAKAN METODE WEIGHTED PRODUCT	[Edit] [Hapus]
4	200831025	2008310252017	PERANCANGAN PROTOTYPE SISTEM KEAMANAN PINTU RUANGAN MENGGUNAKAN IP CAMERA, SMART CARD DAN RFID (RADIO FREQUENCY IDENTIFICATION) BERBASIS MIKROKONTROLER DAN ANDROID	[Edit] [Hapus]
5	201131042	2011310422017	RANCANG BANGUN ALAT PERAGA MODUL PRAKTIKUM EMBEDDED SISTEM MENGGUNAKAN RASPBERRY PI	[Edit] [Hapus]
6	201131133	2011311332017	PROTOTYPE SISTEM KENDALI Kipas Angin MENGGUNAKAN SENSOR SUHU DAN SENSOR INFRAMERAH BERBASIS MIKROKONTROLER DAN ANDROID	[Edit] [Hapus]
7	201211006	2012110062017	PERANCANGAN SISTEM INFORMASI LAPORAN KINERJA PERUSAHAAN BERBASIS WEB PADA PT PLN (PERSERO) KANTOR PUSAT	[Edit] [Hapus]
8	201211008	2012110082017	PERANCANGAN APLIKASI RESERVASI HOTEL ONLINE UNTUK MENINGKATKAN KINERJA PELAYANAN	[Edit] [Hapus]
9	201211009	2012110092017	SISTEM INFORMASI AKADEMIK BERBASIS WEB PADA SEKOLAH MENENGAH PERTAMA NEGERI 5 MAKASSAR	[Edit] [Hapus]

Gambar 11 Halaman data Judul

3. Tampilan Menu Mahasiswa

a. Halaman Utama Mahasiswa

Pada halaman utama menampilkan dua buah *icon*, dimana jika pada *icon* di *click* maka *user* akan masuk menampilkan *form* berdasarkan *icon* yang dipilih.



Gambar 12 Halaman Utama Mahasiswa

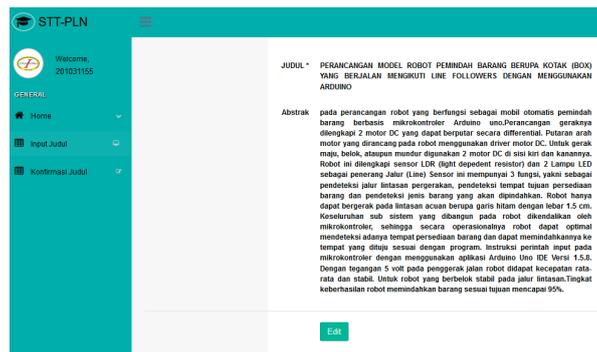
b. Halaman Input Judul

Pada halaman input judul, mahasiswa akan menginputkan data judul dan abstrak skripsi, yang nantinya akan diolah oleh *system* dengan metode TF-IDF dan VSM.

Gambar 13 Halaman Input Judul

c. Halaman Data Judul

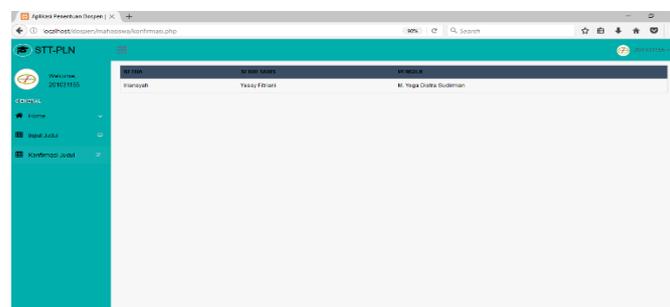
Mahasiswa dapat melihat judul dan abstrak yang sudah diinputkan, dan mengedit judul dan abstrak jika ditemukan kesalahan, dapat dilihat seperti gambar dibawah ini:



Gambar 14 View data judul dan abstrak

d. Halaman Konfirmasi

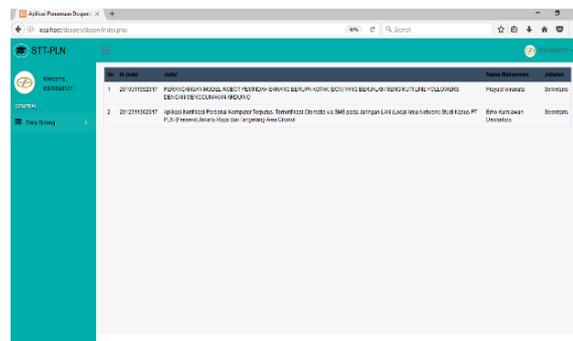
Pada halaman konfirmasi mahasiswa dapat melihat data dosen yang akan menguji siding skripsinya.



Gambar 15 Halaman Konfirmasi

4. Tampilan Menu Dosen

Pada penelitian ini, dosen hanya bisa melihat mahasiswa yang akan diuji, dapat dilihat seperti gambar dibawah ini:



Gambar 16 Halaman Data Sidang Dosen

4.2 Pembahasan

Pengumpulan data dilakukan dengan mengambil skripsi mahasiswa Teknik Informatika STT-PLN di perpustakaan, dan wawancara. Wawancara dilakukan kepada sekretaris jurusan Teknik Informatika STT-PLN, dengan fokus terhadap sistem berjalan dengan penentuan jadwal sidang skripsi dan tim dosen penguji. Hasil dari wawancara bahwa proses pemilihan dosen penguji di jurusan teknik informatika masih secara subjektif, yaitu penentuan dosen penguji dipilih secara langsung oleh Sekretaris jurusan. Belum adanya sistem yang digunakan menjadikan tugas sekretaris jurusan dalam menentukan dosen penguji yang sesuai antara judul atau tema skripsi dengan konsentrasi dosen penguji menjadi tidak relevan dengan kompetensi yang dimiliki dosen penguji. Data set yang diperoleh akan dijadikan sebagai data *training*. Data-

data tersebut dikumpulkan kemudian diolah menggunakan *text preprocessing*, *Text preprocessing* merupakan tahap proses awal terhadap dokumen skripsi untuk mempersiapkan dokumen menjadi dokumen yang akan diolah lebih lanjut. Tahapan *text preprocessing* yang dilakukan adalah *tokenizing*, *filtering*, *stemming*. Kemudian dokumen yang sudah diproses akan diberi bobot dengan menggunakan metode TF-IDF, setelah itu untuk menentukan klasifikasi dari dokumen skripsi digunakanlah metode *Vector Space Model*. Kemudian setelah didapat kan klasifikasi dokumen skripsi, akan dicocokkan antara kompetensi dosen penguji dan klasifikasi dokumen skripsi. Terdapat tiga dosen penguji skripsi yang terbagi atas ketua, sekretaris dan penguji. Masing-masing posisi memiliki syarat-syarat yang berbeda. Ketua harus memiliki jabatan lektor kepala atau lektor, untuk sekretaris jabatan yang dimiliki dosen adalah asisten ahli sedangkan jabatan untuk anggota penguji bisa jabatan apa saja seperti lektor kepala, lektor, asisten ahli dan tenaga ahli.

Pada proses penentuan dosen penguji dilakukan dengan melakukan perhitungan menggunakan *Microsoft excel* dan kemudian membandingkan hasilnya dengan perhitungan yang telah dilakukan oleh sistem di aplikasi. Penentuan relevansi antara penentuan dosen penguji di uji dengan data testing dari 59 dokumen dan hasil klasifikasi yang didapat dari sistem, kemudian menghitung tingkat keakurasiannya, dimana diperoleh yaitu:

$$\begin{aligned} \text{Accuracy} &= \frac{\text{Jumlah Dokumen yang benar dikenali}}{\text{Jumlah Dokumen pengujian}} \times 100 \% \\ &= \frac{55}{59} \times 100\% \\ &= \mathbf{93,22\%} \end{aligned}$$

Dari hasil pengujian, keakuratan terhadap kesesuaian kompetensi dosen penguji yang didapat oleh aplikasi ini mencapai 93,22%. Dimana dengan jumlah pengujian 59 contoh data *training*, dengan tingkat *error* sebesar 6,78% terdapat 4 (empat) buah data *training* yang memiliki nilai belum tepat dalam penentuan kesesuaian kompetensi dosen penguji.

5. PENUTUP

5.1 Kesimpulan

Dari kesimpulan ini dapat disimpulkan bahwa :

1. Penentuan dosen penguji dapat dilakukan dengan menggunakan metode TF-IDF dan VSM, perhitungan data skripsi dilakukan dengan menggunakan *Microsoft excel*.
2. Penentuan dosen penguji skripsi dapat dilakukan berdasarkan kompetensi, dengan menerapkan metode TF-IDF dan VSM kedalam sistem
3. Dari hasil pengujian TF-IDF dan VSM dengan 59 data judul dapat merekomendasikan tiga dosen penguji skripsi, dengan syarat minimal satu orang memiliki kompetensi yang sama dengan topik skripsi dengan menghasilkan akurasi 93,22%.

DAFTAR PUSTAKA

- [1] Gunawan, Lestari, I., & Ihsan Zul, M. (2016). Sistem Pendukung Keputusan Pemilihan Pembimbing dan Penguji Proyek Akhir di Politeknik Caltex Riau. *jurnal informatika*.
- [2] Nath Singh, J., & Dwivedi, s. (2012). Analisa Vector Space Model pada Information Retrieval. *International Journal of Computer Applications*, 14-18.
- [3] Yuan, M., & Xiong, Z. (2013). A Text Categorization Method using Extended Vector Space Model by Frequent Term Sets. *Journal Of Information Science And Engineering* 29, 99-114.
- [4] Gawande, P., & Suryawanshi, A. (2015). Improving Web Page Classification by Vector Space Model . *International Journal of Innovative Research in Computer and Communication Engineering* , 1252-1257.
- [5] Wisnu, D., & Hetami, A. (2015). Perancangan Informtaion Retrieval (IR) untuk Pencarian Ide Pokok Teks Artikel Berbahasa Inggris dengan Pembobotan Vector Space Model. *Jurnal Ilmiah Teknologi dan Informasi, Volume 9*, Volume 9.
- [6] Asshidiq, A. S., & Sulisty, M. (2013). Penilaian Ujian Bertipe Essay Menggunakan Metode Text Similarity. *Jurnal Informatika*.
- [7] Aziz, A. S., & P.S, K. (2015). Implementasi Vector Space Model dalam Pembangkitan Frequently Asked Questions dan Solusi yang Relevan Keluhan Pelanggan. *Scientific Journal Informatics*, 2.
- [8] Lestari, K. (2013). Query Expansion pada Sistem Temu Kembali Informasi dengan Model Ruang Vektor. *Naskah Tugas Akhir*, 24.
- [9] Pardede, J., Barmawi, M. M., & Pramono, W. D. (2013). Implementasi Metode Generalized Vector Space Model pada Aplikasi Information Retrieval. *Jurnal Infromatika No.1, Vol. 4, Januari-April, ISSN : 2087-5266*, 35
- [10] Herman, Achmad, A., & Ahmad, a. i. (2014). Klasifikasi Dokumen Naskah Dinas Menggunakan Algoritma Term Frequency – Inversed Document Frequency Dan Vector Space Model.
- [11] Imbar, R. V., Adelia, Ayub, M., & Rehata, A. (2014). Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika, Vol. 10 No. 1*, 31 - 42.
- [12] Amburika, B., Herry, Y. C., & Uriawan, W. (2016). Teknik Vector Space Model (VSM) Dalam Penentuan Penanganan Dampak Game Online Pada Anak. *jurnal informatika*, 73-38.
- [13] Arief, M. R. (2011). *Pemrograman Web Dinamis Menggunakan PHP dan MySQL*. Yogyakarta: ANDI OFFSET.