**ARTIKEL**

# SISTEM BANTU DETEKSI PLAGIARISME JUDUL SKRIPSI MENGGUNAKAN LATENT SEMANTIC ANALYSIS

**(Studi Kasus: Program Studi Akuntansi Universitas Nusantara PGRI Kediri)**


**Oleh:**

**MUHAMMAD SYAHRIZAL ALIEF**

**13.1.03.023.0237**

**Dibimbing oleh :**

1. **PATMI KASIH, M.KOM**
2. **MADE AYU DUSEA WIDYA DARA, M.KOM**

**TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS NUSANTARA PGRI KEDIRI**


**2018**

# SURAT PERNYATAAN
# ARTIKEL SKRIPSI TAHUN 2018

**Yang bertanda tangan di bawah ini:**

| | |
|---|---|
| Nama Lengkap | : MUHAMMAD SYAHRIZAL ALIEF |
| NPM | : 13.1.03.02.0237 |
| Telepon/HP | : 085733333726 |
| Alamat Surel (Email) | : syahrizal.syahrizalalief.alief@gmail.com |
| Judul Artikel | : SISTEM BANTU DETEKSI PLAGIARISME JUDUL SKRIPSI MENGGUNAKAN LATENT SEMANTIC ANALYSIS (Studi Kasus: Program Studi Akuntansi Universitas Nusantara PGRI Kediri) |
| Fakultas – Program Studi | : TEKNIK – TEKNIK INFORMATIKA |
| Nama Perguruan Tinggi | : UNIVERSITAS NUSANTARA PGRI KEDIRI |
| Alamat Perguruan Tinggi | : Jalan K.H. Achmad Dahlan 76 Telepon (0354) 771576 Kediri |

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

| Mengetahui | | Kediri, 11 Agustus 2018 |
|---|---|---|
| Pembimbing I | Pembimbing II | Penulis, |
| Patmi Kasih, M.Kom<br>NIDN. 0701107802 | Made Ayu Dusea Widya Dara, M.Kom<br>NIDN. 0729088802 | Muhammad Syahrizal Alief<br>NPM. 13.03.02.0237 |

# SISTEM BANTU DETEKSI PLAGIARISME JUDUL SKRIPSI MENGGUNAKAN LATENT SEMANTIC ANALYSIS

## (Studi Kasus: Program Studi Akuntansi Universitas Nusantara PGRI Kediri)

Muhammad Syahrizal Alief
13.03.02.0237
Fakultas Teknik – Prodi Teknik Informatika
syahrizal.syahrizalalief.alief@gmail.com
Patmi Kasih, M.Kom dan Made Ayu Dusea Widya Dara, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

## ABSTRAK

Banyaknya data skripsi menjadikan kurang diperhatikannya tingkat kemiripan antara judul skripsi baru dengan judul skripsi yang sudah ada sebelumnya, sehingga besar kemungkinan tindakan plagiarisme mudah untuk dilakukan dikarenakan keterbatasan petugas akademis dalam melakukan pengecekan. Maka dengan demikian dibutuhkan suatu sistem bantu dalam melakukan deteksi tingkat kemiripan judul skripsi secara otomatis untuk mencegah tindakan plagiarisme.

Salah satu metode yang dapat digunakan untuk melakukan pendeteksian kemiripan yaitu algoritma *Latent Semantic Analysis. Latent Semantic Analysis* adalah suatu metode untuk menemukan hubungan, keterkaitan, dan kemiripan antar dokumen-dokumen, penggalan dari dokumen-dokumen, dan kata-kata yang muncul pada dokumen-dokumen dengan memanfaatkan komputasi statistik untuk menggali dan merepresentasikan konteks yang digunakan sebagai sebuah arti kata untuk sejumlah *corpus* yang besar.

Setiap judul skripsi yang diuji akan melalui beberapa tahapan dalam pendeteksiannya. Tahap pertama yaitu setiap judul skripsi akan masuk pada tahap preprocessing dimana di dalamnya meliputi proses tokenizing (pemecahan kalimat menjadi kata per kata), stopword removal (penghapusan kata tidak penting), stemming (pengambilan suku kata). Setelah tahap preprocessing telah selesai maka, tahap selanjutnya ialah pembobotan dari setiap term hasil dari preprocessing yang nantinya bobot ini yang akan digunakan dalam penghitungan nilai kemiripan antar judul skripsi. Selanjutnya yaitu tahap penghitungan bobot menggunakan cosine similarity, hasil dari perhitungan inilah yang digunakan untuk mengetahui tingkat kemiripan judul skripsi yang diuji.

Kesimpulan hasil penelitian ini yaitu jika struktur kata yang terbentuk dari setiap judul yang di uji berbeda, maka nilai tingkat kemiripan ditentukan dari seberapa banyak struktur kata yang sama, semakin banyak struktur kata yang sama maka semakin tinggi pula nilai tingkat kemiripannya. Dengan adanya sistem bantu ini dapat mempermudah petugas akademik program studi Akuntansi dalam melakukan pengecekan judul skripsi untuk mencegah tindakan plagiarisme. Berdasarkan simpulan hasil penelitian ini, disarankan agar sistem bantu deteksi ini dapat dilengkapi dengan metode-metode text mining lain yang cocok diterapkan dalam deteksi kemiripan judul skripsi.

**KATA KUNCI :** *Latent Semantic Analysis*, *Text Mining*, *Cosine Similarity*, TF-IDF

## I. LATAR BELAKANG

Pada program studi Akuntansi Universitas Nusantara PGRI Kediri hingga saat ini proses pengecekan judul skripsi oleh petugas akademis masih dilakukan secara konvensional dengan membandingkan satu per satu judul skripsi yang sudah ada dan hanya dibandingkan dengan judul skripsi yang berada dalam satu angkatan, sehingga akan ada kemungkinan saat petugas akademis mengecek judul skripsi ada kesalahan dalam pengecekan judul skripsi yang baru diajukan dengan judul skripsi yang sudah ada sebelumnya dalam satu angkatan.

Walaupun proses pengecekan secara konvensional oleh petugas akademis telah dilakukan secara teliti dan judul skripsi yang baru diajukan tidak memiliki tingkat kesamaan yang tinggi terhadap judul skripsi yang dibandingkan, petugas belum bisa memastikan bahwa judul skripsi tersebut memiliki kesamaan dengan judul skripsi yang berada dalam tingkatan berbeda dikarenakan hanya dibandingkan dengan judul skripsi yang berada dalam satu tingkatan.

Berdasarkan permasalahan di atas, untuk menanggulangi hal tersebut perlu adanya sistem yang dapat digunakan untuk mendeteksi plagiarisme judul skripsi secara otomatis. Salah satu metode yang dapat digunakan untuk mendeteksi kemiripan antar dokumen teks yaitu *Latent Semantic Analysis*.

Dengan metode tersebut, diharapkan pendeteksian *plagiarisme* judul skripsi yang diajukan oleh mahasiswa akan lebih akurat dan mempermudah tugas dari petugas akademis Program Studi Akuntansi Universitas Nusantara PGRI Kediri.

## II. METODE

### A. Metode Penelitian

Pendekatan penelitian yang digunakan pada penelitian ini adalah pendekatan kualitatif, dimana peneliti melakukan analisa terhadap akurasi metode LSA dalam hal deteksi plagiarisme judul skripsi. Penelitian ini juga merupakan penelitian eksperimental, karena untuk mendapatkan akurasi terbaik algoritma LSA dalam pendeteksian kemiripan, dilakukan beberapa kali percobaan dengan parameter yang berbeda-beda. Teknik penelitian yang sesuai adalah penelitian pengembangan atau rekayasa teknologi informasi. Prosedur yang dilakukan dalam penelitian ini yaitu:

1. Sumber pustaka dari buku dan jurnal yang berkaitan dengan deteksi kemiripan yang menggunakan metode LSA.
2. Mendefinisikan dan mengumpulkan semua kebutuhan kemudian kebutuhan tersebut dianalisis dan didefinisikan.

3 Merancang sistem merupakan tahap penyusunan proses, data, aliran proses dan hubungan antar data.

4 Implementasi sistem.

## B. Metode Algoritma Latent Semantic Analysis

Metode yang akan digunakan adalah algoritma *Latent Semantic Analysis*. *Latent Semantic Analysis* (LSA) merupakan salah satu algoritma yang sering digunakan dalam proses text mining. Menurut Danang Wahyu Wicaksono, dkk. (2014), dijelaskan bahwa *Latent Semantic Analysis* sebagai berikut:

> LSA adalah suatu metode untuk menemukan hubungan, keterkaitan, dan kemiripan antar dokumen-dokumen, penggalan dari dokumen-dokumen, dan kata-kata yang muncul pada dokumen-dokumen dengan memanfaatkan komputasi statistik untuk menggali dan merepresentasikan konteks yang digunakan sebagai sebuah arti kata untuk sejumlah *corpus* yang besar. *Corpus* adalah kumpulan teks yang memiliki kesamaan subyek atau tema.

Menurut Derwin Suhartono (2015), secara umum langkah-langkah LSA adalah sebagai berikut :

1. ***Text Processing***
   a. ***Stopwords Removal*** : Pada *stopwords removal*, kata-kata yang tergolong sebagai kata depan, kata penghubung, dan kata-kata lain yang tidak mewakili makna dari kalimat akan dieliminasi.
   b. ***Stemming*** : Langkah berikutnya adalah *stemming*. Pada proses ini, kata akan dinormalkan menjadi kata dasar pembentuk kata tersebut. Caranya adalah dengan menghilangkan imbuhan yang melekat pada kata, sehingga hasilnya adalah kata dasarnya. Apabila dalam Bahasa Inggris, proses *stemming* bisa mengikutsertakan pengembalian bentuk tense dari kata kerja bentuk ke-2 atau ke-3 menjadi kata kerja bentuk ke-1.
2. ***Term-document Matrix***
   Setelah melalui *stopwords removal* dan *stemming*, matriks *term-document* dibangun dengan menempatkan kata hasil proses *stemming* (*term*) ke dalam baris. Matriks ini disebut *term-document matrix*. Setiap baris mewakili sebuah kata yang unik, sedangkan setiap kolom mewakili konteks dari mana kata-kata tersebut diambil. Konteks yang dimaksud bisa berupa kalimat, paragraf, atau seluruh bagian dari teks.
3. ***Singular Value Decomposition***
   *Singular Value Decomposition* (SVD) adalah salah satu teknik reduksi dimensi yang bermanfaat untuk memperkecil nilai kompleksitas dalam pemrosesan *term-document matrix*. SVD merupakan teorema aljabar linier yang menyebutkan bahwa persegi panjang dari *term-document matrix* dapat dipecah/didekomposisikan menjadi tiga matriks, yaitu :
   a. Matriks ortogonal U
   b. Matriks diagonal S
   c. Transpose dari matriks ortogonal V, Yang dirumuskan dengan :

$$\boldsymbol{A_{mn} = U_{mn}\, X\, S_{mn}\, X\, V^{T}{}_{nn}}$$

Keterangan :

$A_{mn} = matriks\ awal$
$U_{mn} = matriks\ ortogonal\ U$
$S_{mn} = matriks\ diagonal\ S$
$V^{T}{}_{nn} = transpose\ matriks\ ortogonal\ V$

4. ***Cosine Similarity Measurement***
   *Cosine similarity* digunakan untuk menghitung nilai kosinus sudut antara vektor dokumen dengan vektor

*query*. Semakin kecil sudut yang dihasilkan, maka tingkat kemiripan esai semakin tinggi. Formula dari *cosine similarity* adalah sebagai berikut:

$$\cos\alpha = \frac{A \cdot B}{|A||B|}$$

$$= \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}}$$

## C. Simulasi Metode

Contoh perhitungan *Latent Semantic Analysis* dengan menggunakan lima judul skripsi uji (D0 – D2) dan judul skripsi penguji (Q). judul skripsi uji dan penguji dapat dilihat pada tabel berikut ini:

**Tabel 1. Data Judul Skripsi Uji dan Judul Skripsi Penguji**

| Kode Judul | Isi Judul Skripsi |
|---|---|
| Q | Analisis Pengaruh Suku Bunga, Volume Perdagangan, Nilai Tukar Uang, Harga Emas Dan Harga Minyak Terhadap Return Saham Sektor Properti Yang Terdaftar Di Bursa Efek Indonesia Periode 2010-2014 |
| D0 | Pengaruh Earning Per Share, Return On Equity, Price Earning Ratio Dan Net Profit Margin Terhadap Harga Saham Pada Perusahaan Manufaktur Yang Terdaftar Di Bei Tahun 2011-2013 |
| D1 | Analisis Pengaruh Ukuran Perusahaan, Return On Equity, Price Earning Ratio, Dan Current Ratio Terhadap Harga Saham Pada Perusahaan Manufaktur Yang Terdaftar Di Bursa Efek Indonesia Tahun 2011-2013 |
| D2 | Pengaruh Net Profit Margin, Return On Asset, Earning Per Share Dan Devidend Per Share Terhadap Harga Saham Pada Perusahaan Manufaktur Yang Terdaftar Di Bursa Efek Indonesia Periode Tahun 2011-2013 |

Langkah-langkah perhitungan terhadap tiga dokumen uji dan satu dokumen penguji adalah sebagai berikut:

1. Langkah Pertama : *Preprocessing* terhadap semua (n=4) dokumen yang terlibat, yaitu Q, D0, D1, dan D2.
   a. Langkah 1a : Lakukan *tokenizing*, *stopword removal*, dan *stemming*. Hasilnya dapat dilihat pada tabel berikut :

**Tabel 2. Hasil dari Langkah 1a**

| Kode Judul | Isi Judul Skripsi |
|---|---|
| Q | ANALISIS PENGARUH SUKU BUNGA VOLUME PERDAGANGAN NILAI TUKAR UANG HARGA EMAS DAN HARGA MINYAK TERHADAP RETURN SAHAM SEKTOR PROPERTI YANG TERDAFTAR DI BURSA EFEK INDONESIA PERIODE |
| D0 | PENGARUH EARNING PER SHARE RETURN ON EQUITY PRICE EARNING RATIO DAN NET PROFIT MARGIN TERHADAP HARGA SAHAM PADA PERUSAHAAN MANUFAKTUR YANG TERDAFTAR DI BEI TAHUN |
| D1 | ANALISIS PENGARUH UKURAN PERUSAHAAN RETURN ON EQUITY PRICE EARNING RATIO DAN CURRENT RATIO TERHADAP HARGA SAHAM PADA PERUSAHAAN MANUFAKTUR YANG TERDAFTAR DI BURSA EFEK INDONESIA TAHUN |
| D2 | PENGARUH NET PROFIT MARGIN RETURN ON ASSET EARNING PER SHARE DAN DEVIDEND PER SHARE TERHADAP HARGA SAHAM PADA PERUSAHAAN MANUFAKTUR YANG TERDAFTAR DI BURSA EFEK INDONESIA PERIODE TAHUN |

   **b.** Langkah 1b : Tentukan bobot untuk setiap *term* dari empat dokumen tersebut dengan menggunakan rumus Wdt = tf * idf.

**Tabel 3. Nilai tf dan idf**

| Term | tf | | | | | idf |
|---|---|---|---|---|---|---|
| | Q | D0 | D1 | D2 | df | log(n/df) |
| ANALISA | 1 | 0 | 1 | 0 | 2 | 0.301 |
| ASSET | 0 | 0 | 0 | 1 | 1 | 0.602 |
| BUNGA | 1 | 0 | 0 | 0 | 1 | 0.602 |
| BURSA | 1 | 0 | 1 | 1 | 3 | 0.125 |
| CURRENT | 0 | 0 | 1 | 0 | 1 | 0.602 |
| EFEK | 1 | 0 | 1 | 1 | 3 | 0.125 |
| EMAS | 1 | 0 | 0 | 0 | 1 | 0.602 |
| EQUITY | 0 | 1 | 1 | 0 | 2 | 0.301 |
| HARGA | 2 | 1 | 1 | 1 | 5 | 0.097 |
| INDONESIA | 1 | 0 | 1 | 1 | 3 | 0.125 |
| MANUFAKTUR | 0 | 1 | 1 | 1 | 3 | 0.125 |
| MARGIN | 0 | 1 | 0 | 1 | 2 | 0.301 |
| MINYAK | 1 | 0 | 0 | 0 | 1 | 0.602 |
| NET | 0 | 1 | 0 | 1 | 2 | 0.301 |
| NILAI | 1 | 0 | 0 | 0 | 1 | 0.602 |
| PENGARUH | 1 | 1 | 1 | 1 | 4 | 0.000 |
| PERIOD | 1 | 0 | 0 | 1 | 2 | 0.301 |
| PRICE | 0 | 1 | 1 | 0 | 2 | 0.301 |
| PROFIT | 0 | 1 | 0 | 1 | 2 | 0.301 |
| PROPERTI | 1 | 0 | 0 | 0 | 1 | 0.602 |
| RATIO | 0 | 1 | 2 | 0 | 3 | 0.125 |
| RETURN | 1 | 1 | 1 | 1 | 4 | 0.000 |
| SAHAM | 1 | 1 | 1 | 1 | 4 | 0.000 |
| SEKTOR | 1 | 0 | 0 | 0 | 1 | 0.602 |
| SHARE | 0 | 1 | 0 | 2 | 3 | 0.125 |
| SUKU | 1 | 0 | 0 | 0 | 1 | 0.602 |
| TUKAR | 1 | 0 | 0 | 0 | 1 | 0.602 |
| UANG | 1 | 0 | 0 | 0 | 1 | 0.602 |
| UKUR | 0 | 0 | 1 | 0 | 1 | 0.602 |
| USAHA | 0 | 1 | 2 | 1 | 4 | 0.000 |
| VOLUME | 1 | 0 | 0 | 0 | 1 | 0.602 |

Keterangan :

n : 4, yaitu Q, D0, D1, dan D2.

tf : banyaknya kata dalam suatu dokumen

df : total dokumen yang mengandung suatu kata

**Tabel 4. Hasil Pembobotan**

| Term | wdt = tf * idf | | | |
|---|---|---|---|---|
| | Q | D0 | D1 | D2 |
| ANALISA | 0.301 | 0.000 | 0.301 | 0.000 |
| ASSET | 0.000 | 0.000 | 0.000 | 0.602 |
| BUNGA | 0.602 | 0.000 | 0.000 | 0.000 |
| BURSA | 0.125 | 0.000 | 0.125 | 0.125 |
| CURRENT | 0.000 | 0.000 | 0.602 | 0.000 |
| EFEK | 0.125 | 0.000 | 0.125 | 0.125 |

**Tabel 4. Hasil Pembobotan (Lanjutan)**

| Term | wdt = tf * idf | | | |
|---|---|---|---|---|
| | Q | D0 | D1 | D2 |
| EMAS | 0.602 | 0.000 | 0.000 | 0.000 |
| EQUITY | 0.000 | 0.301 | 0.301 | 0.000 |
| HARGA | 0.194 | 0.097 | 0.097 | 0.097 |
| INDONESIA | 0.125 | 0.000 | 0.125 | 0.125 |
| MANUFAKTUR | 0.000 | 0.125 | 0.125 | 0.125 |
| MARGIN | 0.000 | 0.301 | 0.000 | 0.301 |
| MINYAK | 0.602 | 0.000 | 0.000 | 0.000 |
| NET | 0.000 | 0.301 | 0.000 | 0.301 |
| NILAI | 0.602 | 0.000 | 0.000 | 0.000 |
| PENGARUH | 0.000 | 0.000 | 0.000 | 0.000 |
| PERIOD | 0.301 | 0.000 | 0.000 | 0.301 |
| PRICE | 0.000 | 0.301 | 0.301 | 0.000 |
| PROFIT | 0.000 | 0.301 | 0.000 | 0.301 |
| PROPERTI | 0.602 | 0.000 | 0.000 | 0.000 |
| RATIO | 0.000 | 0.125 | 0.250 | 0.000 |
| RETURN | 0.000 | 0.000 | 0.000 | 0.000 |
| SAHAM | 0.000 | 0.000 | 0.000 | 0.000 |
| SEKTOR | 0.602 | 0.000 | 0.000 | 0.000 |
| SHARE | 0.000 | 0.125 | 0.000 | 0.250 |
| SUKU | 0.602 | 0.000 | 0.000 | 0.000 |
| TUKAR | 0.602 | 0.000 | 0.000 | 0.000 |
| UANG | 0.602 | 0.000 | 0.000 | 0.000 |
| UKUR | 0.000 | 0.000 | 0.602 | 0.000 |
| USAHA | 0.000 | 0.000 | 0.000 | 0.000 |
| VOLUME | 0.602 | 0.000 | 0.000 | 0.000 |

2. Langkah Kedua : Hitung kemiripan vektor dokumen D0 dengan setiap dokumen yang ada. Kemiripan antar dokumen dihitung menggunakan persamaan *cosine similarity*.
   a. Langkah 2a : Hitung hasil perkalian scalar antara Q dan lima dokumen lainnya, kemudian hasil perkalian setiap dokumen dengan D0 dijumlahkan. Hitung juga panjang setiap dokumen, termasuk D0 dengan cara mengkuadratkan bobot setiap *term* dalam setiap dokumen, kemudian dijumlahkan dan terakhir diakarkan.

**Tabel 5. Hasil Perhitungan Langkah 2a**

| WD * Wdi | | | Panjang Vektor | | | |
|---|---|---|---|---|---|---|
| D0 | D1 | D2 | Q | D0 | D1 | D2 |
| 0.000 | 0.091 | 0.000 | 0.091 | 0.000 | 0.091 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.362 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.016 | 0.016 | 0.016 | 0.000 | 0.016 | 0.016 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.362 | 0.000 |
| 0.000 | 0.016 | 0.016 | 0.016 | 0.000 | 0.016 | 0.016 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.091 | 0.091 | 0.000 |
| 0.019 | 0.019 | 0.019 | 0.038 | 0.009 | 0.009 | 0.009 |
| 0.000 | 0.016 | 0.016 | 0.016 | 0.000 | 0.016 | 0.016 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.016 | 0.016 | 0.016 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.091 | 0.000 | 0.091 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.091 | 0.000 | 0.091 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.091 | 0.091 | 0.000 | 0.000 | 0.091 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.091 | 0.091 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.091 | 0.000 | 0.091 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.016 | 0.062 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.016 | 0.000 | 0.062 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.362 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.000 | 0.362 | 0.000 | 0.000 | 0.000 |
| 0.019 | 0.156 | 0.156 | 3.890 | 0.509 | 1.131 | 0.859 |
| Akar | | | 1.972 | 0.714 | 1.064 | 0.927 |

Keterangan :

Tabel warna kuning : hasil penjumlahan perkalian scalar

Tabel warna abu-abu : nilai panjang vektor

b. Langkah 2b : Terapkan persamaan *cosine similarity* dan hitung kemiripan Q dengan D0 sampai dengan D2.

Cos(Q,D0) = 0.019 / (1.972 * 0.714) = 0.013

Cos(Q,D1) = 0.156 / (1.972 * 1.064) = 0.074

Cos(Q,D2) = 0.156 / (1.972 * 0.927) = 0.085

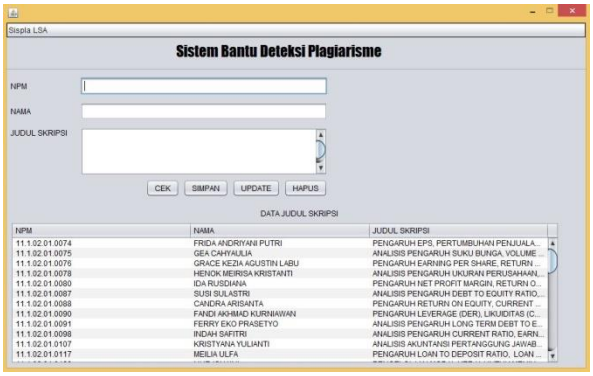
**Tabel 6. Hasil Perhitungan Cosine Similarity**

| Dokumen | Q,D0 | Q,D1 | Q,D2 |
|---|---|---|---|
| Cosine Similarity | 0.013 | 0.074 | 0.085 |

## III. HASIL DAN KESIMPULAN

### A. Tampilan Halaman Beranda

Halaman yang pertama kali muncul adalah halaman beranda.

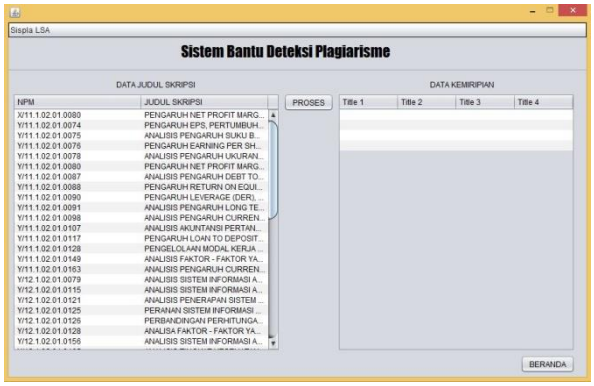



Gambar 1. Halaman Beranda

Pada halaman ini user bisa memasukkan data baru meliputi NPM, Nama, dan Judul Skripsi yang akan digunakan pada proses pendeteksian. Pada halaman ini juga ditampilkan data judul skripsi yang menjadi data training.

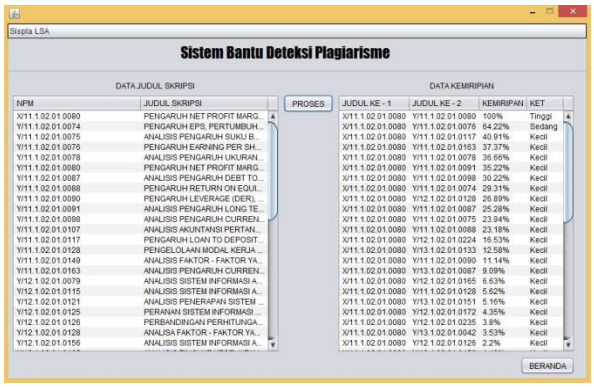
## B. Tampilan Halaman Deteksi

Setelah memasukkan data baru yang akan diuji dan klik tombol cek, user akan masuk ke halaman deteksi.




Gambar 2. Halaman deteksi

Pada halaman ini akan ditampilkan data judul skripsi yang akan di hitung nilai tingkat kemiripan antar judul skripsi. Kemudian jika *user* klik tombol proses, maka proses deteksi akan dilakukan dan hasilnya akan ditampilkan pada tabel data kemiripan.




Gambar 3. Halaman deteksi (Hasil Deteksi)

# IV. PENUTUP

## A. Simpulan

Berdasarkan hasil penelitian dan pembahasan tentang deteksi tingkat kemiripan guna membantu dalam identifikasi plagiarisme judul skripsi, dapat disimpulkan:

1. Sistem bantu deteksi plagiarisme judul skripsi pada program studi Akuntansi telah berhasil dibuat dan dapat mempermudah petugas akademik program studi Akuntansi dalam melakukan pengecekan judul skripsi untuk mencegah tindakan plagiarisme.
2. Setiap judul skripsi yang diuji akan dibandingkan dengan tiap kata yang terbentuk, sehingga jika ada dua judul skripsi yang memiliki struktur kata yang sama akan memiliki tingkat kemiripan 100%. Jika struktur kata yang terbentuk dari setiap judul yang di uji berbeda, maka nilai tingkat kemiripan ditentukan dari seberapa banyak struktur kata yang sama, semakin banyak struktur kata yang sama maka semakin tinggi pula nilai tingkat kemiripannya.

## B. Saran

Setelah menarik dari beberapa kesimpulan maka saran yang bisa diberikan oleh penulis sebagai berikut:

1. Sistem bantu deteksi ini dapat dilengkapi dengan menggunakan metode-metode text mining lain yang cocok diterapkan dalam deteksi kemiripan judul skripsi.
2. Dari hasil penelitian tersebut dapat dikembangkan menjadi sistem yang kemudian hari lebih memudahkan dari semua pihak.

## V. DAFTAR PUSTAKA

Dudung. 2016. Pengertian, Komponen dan Fungsi XAMPP Lengkap dengan Penjelasannya (Online). Tersedia: http://www.dosenpendidikan.com/ pengertian-komponen-dan-fungsi-xampp-lengkap-dengan-penjelasannya/, diakses 10 Juli 2016.

Hermadi, I. Clustering Menggunakan Self-Organizing Maps (Studi Kasus: Data PPMB IPB). *FMIPA Institut Pertanian Bogor*. 5: 2. 2007.

Imbar, R.V., Adelia, Ayub, M. & Rehatta, A. 2014. Implementasi *Cosine Similarity* dan Algoritma *Smith-Waterman* untuk Mendeteksi Kemiripan Teks. Jurnal Informatika, 10 (1): 31-42.

Jogiyanto. 2005. Pengertiang Entity Relationship Diagram (ERD). (Online). Tersedia: http://www.teukutaufik.com/2016/02/ definisi-dan-simbol-erd-entity.html, diakses 9 Juli 2016.

Kadir, A. 2012. *Algoritma & Pemrograman Menggunakan Java*. Yogyakarta: ANDI.

Konchady, M. 2006. *Text Mining Application Programming*. Boston: Charles River Media.

Kowanda, A., Siregar, I.P., Lie, P., Irmawati, N.F. & Purnamasari, D. 2014. Pengukur *Semantic Similarity* Pada Artikel Web Dalam Upaya Pencegahan Plagiarisme. Prosiding Seminar Ilmiah Nasional Komputer dan Sistem Intelejen (KOMMIT 2014), 8: 33-40.

Kusrini, Luthfi, E.T. 2009. Algoritma Data Mining. Yogyakarta: ANDI.

Lembaga Penelitian dan Pengabdian Masyarakat (LPPM). 2016. *Panduan Penulisan Karya Tulis Ilmiah*. Kediri: Universitas Nusantara PGRI Kediri.

Nugroho, B. 2005. *Database Relasional dengan MySQL*. Yogyakarta: ANDI.

Perkasa, D.A., Saputra, E. & Fronita, M. 2015. Sistem Ujian *Online Essay* Dengan Penilaian Menggunakan Metode *Latent Semantic Analysis* (LSA). Jurnal Rekayasa dan Manajemen Sistem Informasi, 1 (1): 1-9.

Salim, M.A., Anistyasari, Y. 2017. Pengembangan Aplikasi Penilaian Ujian Essay Berbasis Online Menggunakan Algoritma Nazief dan Adriani dengan Metode Cosine Similarity. Jurnal IT-EDU, 2 (1): 126-135.

Soelistyo, H. 2011. Plagiarisme: Pelanggaran Hak Cipta dan Etika. Yogyakarta: Kanisius.

Suhartono, D. 2015. Penggunaan Latent Semantic Analysis (LSA) Dalam Pemrosesan Teks (Online). Tersedia: https://socs.binus.ac.id/2015/08/03/pe nggunaan-latent-semantic-analysis-lsa-dalam-pemrosesan-teks/, diakses 10 Juli 2016.

Tala, F.Z. 2004. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. Master of Logic Project Institute for Logic, Language and Computation Universiteit van Amsterdam The Netherlands.

Triawati, C. 2009. *Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi*

*Dokumen Berbahasa Indonesia*. Diakses terakhir tanggal 14 Maret 2013, dari http://digilib.ittelkom.ac.id/index.php?option=com_content&view=article&id=590:textmining&catid=20:informatika&Itemid= 14.

Utomo, M.S. 2013. Implementasi Stemmer Tala Pada Aplikasi Berbasis Web. Jurnal Teknologi Informasi DINAMIK. 18 (1): 41-45.

Wicaksono, D.W., Irawan, M.I. & Rukmi, A.M. 2014. Sistem Deteksi Kemiripan Antar Dokumen Teks Menggunakan Model Bayesian Pada *Term Latent Semantic Analysis* (LSA). Jurnal Sains dan Seni POMITS, 3 (2): 41-46.