

MENGLASIFIKASI LAPORAN PRAKTEK KERJA LAPANGAN UNTUK MENCEGAH TERJADINYA PLAGIARISM MENGGUNAKAN JACCARD MEASURE

¹Aditya Anggraini Yoga Kresna Wijayanti, ²Ratih Ayuninghemi S.ST, M.Kom, ³Deni Arifianto, S.Kom

Jurusan Teknik Informatika Fakultas Teknik Univertas Muhammadiyah Jember

Email : aditya.taur05.ict@gmail.com

Abstrak

Dokumen laporan praktek kerja lapangan dibuat setelah melakukan praktek kerja lapangan pada suatu instansi pemerintah maupun swasta. pada laporan masih dapat ditemukan kemiripan atau plagiasi pada isi laporan tanpa mencantumkan sumber referensi. Hal ini diperlukan agar dapat meningkatkan kualitas dokumen laporan praktek kerja lapangan. Penelitian ini membahas kemiripan beberapa dokumen laporan untuk mendeteksi kesamaan dari isi dokumen tersebut. Untuk itu diperlukan suatu mekanisme mendeteksi plagiasi dokumen. Tahapan dalam metodologi penelitian terdiri dari: pengumpulan dokumen, isi laporan praktek kerja lapangan sebagai dataset, melakukan proses *text mining* yaitu : *casefolding*, *tokenizing*, *stemming* dan *filtering*, pembobotan dengan *TF-IDF*, perhitungan dengan metode *jaccard measure* sebagai alat untuk menentukan apakah dokumen termasuk dinyatakan sebagai dokumen plagiat atau tidak plagiat dengan melihat dari isi laporan praktek kerja lapangan tersebut. Sebelum di lakukan penghitungan dengan *jaccard measure* , di lakukan pra-proses *TF-IDF* terlebih dahulu, yaitu tingkat kemunculan kata dari suatu kalimat dengan koleksi dokumen yaitu jumlah yang akan di hitung bobotnya, dari bobot tersebut akan dijadikan masukkan untuk *jaccard measure*. Hasilnya berupa pengkalsifikasian tersebut telah diberikan 3 nilai ambang batas yaitu 0.5, 0.6, dan 0.7 . Data uji coba diambil dari beberapa laporan praktek kerja lapangan sebanyak 50 dokumen. Hasil pengujian menunjukan bahwa metode *jaccard measure*, mampu mengklasifikasikan dokumen plagiat maupun tidak plagiat dengan nilai ambang batas yang diberikan. *threshold* 0.5 sebesar 88%, *threshold* 0.6 sebesar 87%, dan *threshold* 0.7 sebesar 88% . Dengan hasil rata-rata akurasi sebesar 88%.

Kata Kunci : *text mining*, *tf – idf*, *laporan praktek kerja lapangan*, *koleksi dokumen*, *jaccard measure*.

1. PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini sangat berkembang dengan pesat sehingga menyebabkan kebutuhan akan informasi yang cepat, tepat, dan terbaru menjadi suatu yang tidak dapat disangkal lagi dan berimbas pada perkembangan berbagai macam aspek kehidupan manusia. Kini peranan teknologi informasi telah mencakup hampir di semua bidang ilmu. Dengan kondisi ini teknologi informasi banyak digunakan di beberapa instansi pemerintah maupun perusahaan-perusahaan swasta dimana instansi ataupun perusahaan tersebut akan mengembangkan teknologi informasi yang baru

maupun mengembangkan teknologi informasi yang lama.

Praktek Kerja Lapangan (PKL) adalah suatu bentuk pendidikan dengan cara memberikan pengalaman belajar bagi mahasiswa untuk berpartisipasi dengan tugas langsung di Lembaga BUMN, BUMD, Perusahaan Swasta, dan Instansi Pemerintahan setempat. Praktek Kerja Lapangan (PKL) memberi kesempatan kepada mahasiswa untuk mengimplementasikan ilmu-ilmu yang telah diperoleh di kampus. Praktek Kerja Lapangan (PKL) merupakan wujud relevansi antara teori yang didapat selama di perkuliahan dengan praktek yang ditemui baik dalam dunia usaha swasta maupun pemerintah. Dengan adanya mahasiswa teknologi informasi melakukan Praktek Kerja Lapangan (PKL) dengan sistem penelitian dan analisa maka dapat bermanfaat

bagi Instansi ataupun perusahaan untuk pengembangan teknologi informasi tersebut dan dapat memenuhi kebutuhan tenaga kerja lepas yang berwawasan akademi dari praktek kerja lapangan tersebut. Kemudian laporan praktek kerja lapangan dapat dimanfaatkan sebagai salah satu sumber informasi mengenai situasi umum institusi tempat praktek tersebut.

Banyaknya permasalahan-permasalahan teknologi informasi yang sama disetiap instansi maupun perusahaan. Beberapa mahasiswa Praktek Kerja Lapangan (PKL) yang akan memecahkan permasalahan tersebut dengan penelitian dan analisa dari *problem solving* yang ada, Kemudian laporan praktek kerja lapangan dibuat sesuai dengan penelitian dan analisa yang dilakukan pada instansi tersebut. Berberapa mahasiswa membuat laporan praktek kerja lapangan masih terdapat kemiripan isi-isi Laporan Praktek Kerja Lapangan (PKL) .

Algoritma *jaccard measure* merupakan metode yang memanfaatkan nilai kemiripan dokumen. Penggunaan metode *jaccard measure* dikarenakan nilai akurasi kemiripan dari metode *jaccard measure* yang mendekati nilai keakuratan sangat tinggi. *Jaccard measure* juga merupakan salah satu teknik yang dapat dipergunakan untuk melakukan analisis dalam menghitung tingkat kemiripan suatu dokumen dengan tujuan menghasilkan perolehan yang optimal. *Jaccard measure* mengklasifikasikan yaitu dengan cara memberi pembobotan pada suatu laporan praktek kerja lapangan menghitung *Term Frequency/Inverse Document Frequency (TF/IDF)*. (Widodo. ,A., dkk.2012).

Dengan judul “**MENGKLASIFIKASI LAPORAN PRAKTEK KERJA LAPANGAN UNTUK MENCEGAH TERJADINYA PLAGIARISM MENGGUNAKAN JACCARD MEASURE** “ untuk mengurangi atau mencegah terjadinya kemiripan laporan Praktek Kerja Lapangan (PKL), maka penulis ingin menyelesaikan permasalahan pengklasifikasian laporan praktek kerja lapangan untuk mencegah terjadinya plagiarism menggunakan *Jaccard Measure* yang dapat memperhitungkan tingkat akurasi kemiripan pada suatu dokumen yang terdapat pada isi laporan praktek kerja lapangan .

1.1 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan sebelumnya, terdapat beberapa permasalahan yang akan diangkat dalam skripsi ini, antara lain :

1. Bagaimana melakukan pra-proses menggunakan *text mining* dan perhitungan TF/IDF.
2. Bagaimana mengklasifikasikan laporan praktek kerja lapangan berdasarkan isi laporan menggunakan *Jaccard Measure*.

1.2 Batasan Masalah

Agar tidak menyimpang jauh dari permasalahan, maka skripsi ini mempunyai batasan masalah sebagai berikut :

1. Dataset yang digunakan dalam metode ini merupakan sebuah laporan dokumen yaitu berupa isi dari sebuah laporan praktek kerja lapangan.
2. Laporan yang digunakan adalah isi laporan praktek kerja lapangan pada program studi teknik informatika.
3. Laporan praktek kerja lapangan yang digunakan sebagai dataset adalah berasal dari 50 laporan praktek kerja lapangan.
4. Metode yang digunakan adalah *Jaccard Measure*.

1.3 Tujuan

Tujuan dari penelitian skripsi ini yang dilakukan oleh penulis antara lain :

1. Mengolah data atau dokumen dengan *text mining* dan menghitung bobot dengan TF/IDF.
2. Mengklasifikan plagiat atau bukan plagiat pada laporan praktek kerja lapangan pada program studi teknik informatika dengan *Jaccard Measure*.

1.4 Manfaat

Manfaat dari penelitian skripsi yang dilakukan oleh penulis adalah sebagai berikut :

1. Dapat mengklasifikasikan plagiat atau bukan plagiat pada laporan praktek kerja lapangan pada program studi teknik informatika
2. Memudahkan pembaca untuk mengetahui laporan praktek kerja lapangan yang sudah digunakan ataupun belum.
3. Dapat meningkatkan kualitas laporan praktek kerja lapangan dilingkungan program studi teknik informatika.

2. TINJAUAN PUSTAKA

2.1 Praktek Kerja Lapangan

Praktek Kerja Lapangan (PKL) adalah suatu bentuk pendidikan dengan cara memberikan pengalaman belajar bagi mahasiswa untuk berpartisipasi dengan tugas langsung di Lembaga BUMN, BUMD, Perusahaan Swasta, dan Instansi Pemerintahan setempat. Maksud /tujuan dilaksanakannya Praktek Kerja Lapangan (PKL) Praktek Kerja Lapangan (PKL) memberi kesempatan kepada mahasiswa untuk mengenal dan mengetahui secara langsung tentang instansi sebagai salah satu penerapan disiplin dan pengembangan karier. Ketika di lapangan melaksanakan praktek kerja, mahasiswa dapat menilai tentang pengembangan dari ilmu yang mereka miliki. Agar menjadi media pengaplikasian dari teori yang diperoleh dari bangku kuliah ke

tempat kerja. Meningkatkan hubungan kerjasama antara perguruan tinggi dengan instansi.

2.1 Laporan

Laporan Praktek Kerja Lapangan (PKL) adalah hasil penulisan mahasiswa setelah menyelesaikan Praktek Kerja Lapangan (PKL) berdasarkan data yang di peroleh dan mengimplementasikan yang akan dituangkan dalam bentuk tulisan ilmiah.

2.2 Plagiasi

Plagiarisme adalah salah satu tindakan menyalin, menjiplak, menafsirkan dari karya atau gagasan orang lain tanpa pengakuan sepeenuhnya dan seolah-olah menjadi miliknya. Biasanya pada pendeteksian dilakukan secara manual yaitu dengan melihat isi dari dua dokumen lalu dibandingkan. Tingkat kesamaan dua dokumen pun dihitung sehingga dapat diketahui apakah dokumen tersebut dapat dikatakan plagiat atau tidak. Akan tetapi dengan membandingkan dua dokumen secara manual membutuhkan banyak waktu (Khairunnisa, N., 2012).

2.3 Klasifikasi

Dalam pendekatan pertama, yang dilakukan yaitu menyiapkan data pelatihan yang akan diklasifikasi terlebih dahulu oleh analis. Data pelatihan ini digunakan oleh sistem penambang untuk menemukan sekumpulan term indikator berbobot dari setiap sinopsis. *Term-term* indikator ini digunakan dalam pembelajaran pengklasifikasi untuk mengenali tipe sinopsis dari kebutuhan baru. Pendekatan kedua, sejumlah kecil kebutuhan diidentifikasi oleh analis. Data pelatihan ini digunakan secara langsung dalam pembelajaran pengklasifikasian sinopsis. (Suharso, 2012).

2.4 Text Mining

Text Mining, yang merupakan salah satu bentuk dari *Datamining*. Dapat dikatakan *Text Mining* adalah sebuah analisa yang mengumpulkan *keywords* atau *terms* (istilah) yang sering muncul secara bersamaan dan kemudian menemukan korelasi atau hubungan asosiasi di antara *keywords* atau *terms* tersebut. Pada sumber lain definisi *Text Mining* adalah menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Proses *text mining* yang dilakukan yaitu :

1. *Casefolding*
2. *Scanning*
3. *Tokenezing*
4. *Steaming*
5. *Filtering*

Setelah di dapatkan *term-term* dari proses Text Mining di lakukan penghitungan Tf/Idf yaitu sebagai berikut,

$$W_{ij} = tf \times idf$$

$$W_{ij} = tf_{ij} \times \log \frac{N}{n}$$

Keterangan:

Wij = bobot kata/*term* tj terhadap dokumen di
 tfij = jumlah kemunculan kata/*term* tj dalam di
 N = jumlah semua dokumen yang ada dalam database
 n = jumlah dokumen yang mengandung kata/*term* tj (minimal ada satu kata yaitu *term* tj)

Setelah penghitungan Tf/Idf di lakukan penghitungan *Jaccard Measure* untuk menentukan kemiripan setiap dokumen,

$$im(D_1, Q_1) = \frac{\sum_{k=1}^n (D_{1,k} \times Q_{1,k})}{\sum_{k=1}^n (D_{1,k}) + \sum_{k=1}^n (Q_{1,k}) - \sum_{k=1}^n (D_{1,k} \times Q_{1,k})}$$

3. METODOLOGI PENELITIAN

3.1 Gambaran Umum

Gambaran umum yang akan dikembangkan pada proposal ini adalah seperti gambar dibawah ini.



Langkah-langkah Penelitian :

1. Studi Literatur
 Melakukan pencarian literatur, referensi, dan *tutorial* tentang *jaccard meassure*, *plagiarism* dan semua materi yang berhubungan dengan masalah yang akan dibahas dengan mendownload *paper* di internet.
2. Penyusunan Dataset
 Dataset yang di uji adalah laporan Praktek Kerja Lapangan Program Studi Teknik Informatika Universitas Muhammadiyah Jember. Sedangkan Dataset penguji adalah kumpulan laporan Praktek Kerja Lapangan Program Studi Teknik Informatika dari Universitas Muhammadiyah Jember. Dataset penguji didapatkan dari mahasiswa yang sudah melakukan praktek kerja lapangan.

3. Analisa
Dalam tahap ini, penelitian dilakukan dengan cara menganalisis untuk memecahkan masalah dalam klasifikasi.
4. Desain Sistem
Tahap ini membuat desain aliran kerja manajemen dan desain pemrograman yang diperlukan untuk pengembangan sistem informasi serta memberikan gambaran yang jelas bagaimana suatu sistem dibentuk (Sri, R., dkk, 2011).
5. Uji coba dan evaluasi
Menguji analisa tersebut secara keseluruhan serta mengambil data dari hasil perancangan kemudian dilakukan evaluasi.
6. Pembuatan laporan
Dokumentasi penelitian dalam bentuk laporan mencakup semua tahap yang dilakukan dalam penelitian ini. Laporan ditulis secara jelas agar dapat dimengerti sebagai suatu karya ilmiah.

4. HASIL DAN PEMBAHASAN

4.1 Data

Seluruh data yang akan di uji coba pada tugas akhir ini adalah Dokumen Laporan praktek kerja lapangan ini adalah 50 dokumen dari data seluruh isi laporan yang akan diolah dengan ambang batas 0.5, 0.6, dan 0.7.

4.2 Skenario Uji Coba

4.2.1 Skenario Uji Coba Data Training

Dalam skenario uji coba data *training* ini menggunakan data sebanyak 35 dataset sebagai data training. Data training akan diproses menggunakan *text mining* yang meliputi *tokennizing*, *stemming* dan *filtering*. Setelah data *training* telah melakukan proses *text mining* maka dari hasil *filtering* akan disimpan didalam *database* yang akan digunakan untuk menghitung kemiripan beberapa dokumen dengan data testing dengan menggunakan perhitungan TF/IDF dan perhitungan *jaccard measure*.

4.2.2 Skenario Uji Coba Data Testing

Dalam skenario uji coba data *testing* ini menggunakan data sebanyak 15 dataset sebagai data *testing*. Data testing akan diproses menggunakan *text mining* yang meliputi *tokennizing*, *stemming* dan *filtering*. Dari hasil *filtering* pada data testing akan dihitung menggunakan perhitungan TF/IDF dimana data testing sebagai keyword yang akan dibandingkan dengan seluruh data *training* dan digunakan dalam perhitungan *jaccard measure*.

4.2.3 Skenario Uji Coba Perhitungan TF/IDF dan Jaccard Measure

Dalam skenario uji coba Perhitungan TF/IDF dan *Jaccard Measure* ini menghitung data testing sebagai *keyword* yang dibandingkan dengan seluruh data *training* melalui perhitungan TF/IDF dan pembobotan kemudian akan dicari nilai jumlah bobot yang akan dihitung dalam menghitung tingkat *similarity* menggunakan metode *jaccard measure* untuk mengklasifikasikan plagiat atau tidak plagiat dengan diberinya *threshold* (nilai ambang batas) 0.5, 0.6, 0.7, apabila nilai perhitungan *jaccard measure* diatas *threshold* maka dinyatakan plagiat.

4.3 Analisa Hasil

4.3.1 Hasil Pengujian

Pada subbab ini akan dijelaskan hasil pengujian yang sudah dilakukan. Data yang digunakan berupa isi laporan dari 50 dokumen laporan praktek kerja lapangan dari fakultas teknik jurusan teknik informatika universitas muhammadiyah jember tahun angkatan 2008-2011. Setelah dataset telah terkumpul, data dibagi menjadi 2 bagian yaitu 35 data *training* sebagai data banding dan 15 data *testing* sebagai data uji (*keyword*), dari 35 data *training* tersebut diproses menggunakan *text mining* yaitu *tokennizing*, *stemming*, dan *filtering* dan hasil *filtering* tersebut akan disimpan pada *database* agar dapat menyimpan data untuk melakukan perhitungan TF/IDF yang sebelumnya akan dibandingkan dengan data *testing*. Dari ke 35 dataset tersebut yang dibandingkan dengan 15 data *testing* maka akan dihitung nilai *term frequency* dari setiap dokumen yang kemudian dicari nilai dari IDF untuk menghitung pembobotan pada setiap dokumen dan jumlah bobot dokumen dimana dari hasil jumlah dokumen tersebut akan dilakukan perhitungan tingkat *similarity* menggunakan *jaccard measure*. Setelah didapat hasil nilai *similarity* maka dapat dilakukan klasifikasi plagiat atau tidak plagiat dengan memberikan nilai *threshold* 0.5, 0.6, dan 0.7, sehingga apabila nilai dari hasil *jaccard measure* diatas nilai *threshold* maka data tersebut dinyatakan plagiat.

Setiap hasil klasifikasi dari sumber yang di dapatkan sebelumnya akan dilakukan pencocokan dengan *handtrace*. *Handtrace* ini berisi kebutuhan synopsis dari masing-masing dataset yang telah diklasifikasikan secara manual oleh analisis sebbagau jawaban yang benar. Klasifikasi manual ini berdasarkan kesesuaian makna dan *term* potensial dari kebutuhan dalam pengujian. Sehingga dari total jumlah kebutuhan yang terambil, dan jumlah kebutuhan yang terambil dan relevan berdasarkan *handtrace* akan membentuk kinerja rata-rata *Precision*, *Recall*, dan *Acuracy* dari semua dataset.

4.3.2 Accuracy

Hasil dari analisa yang menggunakan 50 dataset dimana akan mengklasifikasikan data menjadi plagiat dan tidak plagiat dari seluruh dokumen yang dibagi 2 dataset yaitu 35 dataset *training* dan 15 dataset *testing*, maka dilakukan kinerja rata-rata *accuracy* sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual di setiap nilai *threshold*nya.

Menghitung *Accuracy* dengan nilai *threshold* 0.5

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.1 Perhitungan *Accuracy* dengan nilai *threshold* 0.5

Tabel 4.1 adalah perhitungan *accuracy* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 96, nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 483, sehingga menghasilkan nilai *accuracy* sebesar 88%.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{96 + 483}{96 + 483 + 44 + 37} = 0.88 \times 100\% = 88\%$$

Menghitung *Accuracy* dengan nilai *threshold* 0.6

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.2 Perhitungan *Accuracy* dengan nilai *threshold* 0.6

Tabel 4.2 adalah perhitungan *accuracy* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 83, nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 464, sehingga menghasilkan nilai *accuracy* sebesar 87%.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{83 + 464}{83 + 464 + 44 + 37} = 0.87 \times 100\% = 87\%$$

Menghitung *Accuracy* dengan nilai *threshold* 0.7

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.3 Perhitungan *Accuracy* dengan nilai *threshold* 0.7

Tabel 4.3 adalah perhitungan *accuracy* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 69, nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 507, sehingga menghasilkan nilai *accuracy* sebesar 88%.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{69 + 507}{69 + 507 + 44 + 37} = 0.88 \times 100\% = 88\%$$

Klasifikasi	Nilai dari <i>threshold</i>		
	0.5	0.6	0.7
Yang terklasifikasikan plagiat (dari klasifikasi sumber)	81	68	54
Yang terklasifikasikan plagiat (dari klasifikasi manual)	15	15	15
Accuracy	88%	87%	88%

Tabel 4.4 Hasil pengujian data plagiat dalam *accuracy* dengan menggunakan seluruh *threshold*.

Hasil dari analisa yang menggunakan 50 dataset dimana jika hasil perhitungan lebih dari 0.5, 0.6, 0.7, maka diklasifikasikan ke dalam data plagiat, dan di dapatkan jumlah data yang terklasifikasi dari 50 dataset isi laporan praktek kerja lapangan kedalam klasifikasi plagiat dan dari klasifikasi yang didapatkan sebelumnya pada *threshold* 0.5, 0.6, 0.7 sebanyak 81, 68, 54 dengan perhitungan manual *jaccard measure* di dapatkan 15 data yang diklasifikasikan menjadi plagiat, dengan hasil seperti diatas maka didapatkan nilai *accuracy* sebesar 88%, 87% dan 88% dari keseluruhan data.

4.3.3 Precision

Hasil dari analisa yang menggunakan 50 dataset dimana akan mengklasifikasikan data menjadi plagiat dan tidak plagiat dari seluruh dokumen yang dibagi 2 dataset yaitu 35 dataset *training* dan 15 dataset *testing*, maka dilakukan kinerja rata-rata *precision* yaitu tingkat ketepatan antara informasi yang diminta oleh pengguna

dengan jawaban yang diberikan oleh sistem, dengan nilai aktual di setiap nilai *threshold*-nya. Menghitung *Precision* dengan nilai *threshold* 0.5

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.5 Perhitungan *Precision* dengan nilai *threshold* 0.5

Tabel 4.5 adalah perhitungan *precision* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 96, nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 483, sehingga menghasilkan nilai *precision* sebesar 69%.

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{96}{96 + 44} = 0.69 \times 100\% = 69\%$$

Menghitung *Precision* dengan nilai *threshold* 0.6

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.6 Perhitungan *Precision* dengan nilai *threshold* 0.6

Tabel 4.6 adalah perhitungan *precision* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 83, nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 464, sehingga menghasilkan nilai *precision* sebesar 65%.

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{83}{83 + 44} = 0.65 \times 100\% = 65\%$$

Menghitung *Precision* dengan nilai *threshold* 0.7

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.7 Perhitungan *Precision* dengan nilai *threshold* 0.7

Tabel 4.7 adalah perhitungan *precision* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 69, nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 507, sehingga menghasilkan nilai *precision* sebesar 61%.

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{69}{69 + 44} = 0.61 \times 100\% = 61\%$$

Klasifikasi	Nilai dari <i>threshold</i>		
	0.5	0.6	0.7
Yang terklasifikasikan plagiat (dari klasifikasi sumber)	81	68	54
Yang terklasifikasikan plagiat (dari klasifikasi manual)	15	15	15
Precision	69 %	65%	61%

Tabel 4.8 Hasil pengujian data plagiat dalam *precision* dengan menggunakan seluruh *threshold*.

Hasil dari analisa yang menggunakan 50 dataset dimana jika hasil perhitungan lebih dari 0.5, 0.6, 0.7, maka diklasifikasikan ke dalam data plagiat, dan di dapatkan jumlah data yang terklasifikasi dari 50 dataset isi laporan praktek kerja lapangan kedalam klasifikasi plagiat dan dari klasifikasi yang didapatkan sebelumnya pada *threshold* 0.5, 0.6, 0.7 sebanyak 81, 68, 54 dengan perhitungan manual *jaccard measure* di dapatkan 15 data yang diklasifikasikan menjadi plagiat, dengan hasil seperti diatas maka didapatkan nilai *precision* sebesar 69%, 65% dan 61% dari keseluruhan data.

4.3.3 Recall

Hasil dari analisa yang menggunakan 50 dataset dimana akan mengklasifikasikan data menjadi plagiat dan tidak plagiat dari seluruh dokumen yang dibagi 2 dataset yaitu 35 dataset *training* dan 15 dataset *testing*, maka dilakukan kinerja rata-rata *recall* yaitu tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi dengan nilai aktual di setiap nilai *threshold*-nya.

Menghitung *Recall* dengan nilai *threshold* 0.5

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.9 Perhitungan *Recall* dengan nilai *threshold* 0.5

Tabel 4.9 adalah perhitungan *recall* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 96, nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 483, sehingga menghasilkan nilai *recall* sebesar 72%.

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{96}{96 + 37} = 0.72 \times 100\% = 72\%$$

Menghitung *Recall* dengan nilai *threshold* 0.6

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.10 Perhitungan *Recall* dengan nilai *threshold* 0.6

Tabel 4.10 adalah perhitungan *recall* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 83, nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 464, sehingga menghasilkan nilai *recall* sebesar 69%.

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{83}{83 + 37} = 0.69 \times 100\% = 69\%$$

Menghitung *Recall* dengan nilai *threshold* 0.7

TP (data plagiat terklasifikasi plagiat)	FP (data tidak plagiat terklasifikasi plagiat)
96	44
FN (data plagiat tidak di klasifikasikan plagiat)	TN (data yang tidak di klasifikasikan plagiat)
37	483

Tabel 4.11 Perhitungan *Recall* dengan nilai *threshold* 0.7

Tabel 4.11 adalah perhitungan *recall* yang di dapatkan nilai TP yaitu data plagiat yang terklasifikasikan menjadi data plagiat sebesar 69,

nilai FP yaitu data tidak plagiat terklasifikasi plagiat sebesar 44, nilai FN data plagiat tidak di klasifikasikan plagiat sebesar 37 dan nilai TN adalah data yang tidak di klasifikasikan plagiat sebesar 507, sehingga menghasilkan nilai *recall* sebesar 65%.

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{69}{69 + 37} = 0.65 \times 100\% = 65\%$$

Klasifikasi	Nilai dari <i>threshold</i>		
	0.5	0.6	0.7
Yang terklasifikasikan plagiat (dari klasifikasi sumber)	81	68	54
Yang terklasifikasikan plagiat (dari klasifikasi manual)	15	15	15
Recall	72%	69%	65%

Tabel 4.12 Hasil pengujian data plagiat dalam *recall* dengan seluruh *threshold*

Hasil dari analisa yang menggunakan 50 dataset dimana jika hasil perhitungan lebih dari 0.5, 0.6, 0.7, maka diklasifikasikan ke dalam data plagiat, dan di dapatkan jumlah data yang terklasifikasi dari 50 dataset isi laporan praktek kerja lapangan kedalam klasifikasi plagiat dan dari klasifikasi yang didapatkan sebelumnya pada *threshold* 0.5, 0.6, 0.7 sebanyak 81, 68, 54 dengan perhitungan manual *jaccard measure* di dapatkan 15 data yang diklasifikasikan menjadi plagiat, dengan hasil seperti diatas maka didapatkan nilai *recall* sebesar 72%, 69% dan 65% dari keseluruhan data.

4.4 Analisa perbandingan Hasil Pengujian dengan *Threshold* 0.5, 0.6, dan 0.7

Dari 50 dataset yang dibagi menjadi 2 dataset yaitu 35 data *training* dan 15 data *testing* yang dapat terklasifikasikan data plagiat dengan nilai *threshold* 0.5 maka akan didapat *accuracy* sebesar 88%, *precision* sebesar 69% dan *recall* sebesar 72 %.

Dari 50 dataset yang dibagi menjadi 2 dataset yaitu 35 data *training* dan 15 data *testing* yang dapat terklasifikasikan data plagiat dengan nilai *threshold* 0.5 maka akan didapat *accuracy* sebesar 87%, *precision* sebesar 65% dan *recall* sebesar 69 %.

Dari 50 dataset yang dibagi menjadi 2 dataset yaitu 35 data *training* dan 15 data *testing* yang dapat terklasifikasikan data plagiat dengan nilai *threshold* 0.5 maka akan didapat *accuracy* sebesar 88%, *precision* sebesar 61% dan *recall* sebesar 65%.

Threshold	Accuracy	Precision	Recall
0.5	88%	69%	72%
0.6	87%	65%	69%
0.7	88%	61%	65%
Rata-Rata	88%	65%	69%

Tabel 4.13 Analisa hasil pengujian seluruh *threshold*

Pada tabel 4.13 adalah perbandingan kinerja *accuracy*, *precision*, dan *recall* pada *threshold* 0.5, 0.6, 0.7. Terlihat jika nilai *accuracy*, *precision*, dan *recall* dari setiap *threshold*, memiliki nilai rata-rata 88%, 65%, dan 69% hal ini membuktikan, bahwa metode *jaccard measure* dapat digunakan untuk mengklasifikasikan data dokumen menjadi plagiat atau tidak plagiat. Pada pengujian ini dapat disimpulkan bahwa *accuracy* (tingkat kedekatan pengukuran kuantitas terhadap nilai yang sebenarnya) antara pengujian sistem aplikasi dengan pengujian secara manual lebih tinggi daripada *precision* (ketepatan) antara pengujian sistem aplikasi dengan pengujian secara manual. Sehingga pada pengujian dokumen tingkat nilai keakuratan lebih tinggi dibandingkan dengan tingkat nilai presisi pada rata-rata dengan nilai ambang batas yang diberikan.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari uji coba dan analisa yang telah dijelaskan dalam bab sebelumnya, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Pada perhitungan TF/IDF yang sudah dilakukan bahwa semakin banyak data *training* tersebut dalam dokumen tersebut, maka bobot hubungan antara data *training* terhadap dokumen akan semakin kecil dalam perhitungan *jaccard measure*.
2. Dari 3 *threshold* yang sudah dilakukan didapatkan hasil :
 - a. Dari klasifikasi plagiat dengan *threshold* 0.5 didapatkan nilai *Accuracy* sebesar 88%
 - b. Dari klasifikasi plagiat dengan *threshold* 0.6 didapatkan nilai *Accuracy* sebesar 87%
 - c. Dari klasifikasi plagiat dengan *threshold* 0.7 didapatkan nilai *Accuracy* sebesar 88%

Dari keseluruhan *threshold* yang sudah dilakukan didapatkan nilai rata-rata *Accuracy* sebesar 88%.

3. Aplikasi yang sudah dibuat mampu mengklasifikasikan dokumen isi laporan praktek kerja lapangan menjadi plagiat dan tidak plagiat dengan *threshold* 0.5, 0.6, dan 0.7.

4. Pengujian dokumen telah didapat tingkat nilai keakuratan lebih tinggi dibandingkan dengan tingkat nilai presisi pada rata-rata dengan nilai ambang batas yang diberikan.

5.2 Saran

Dengan segala kelebihan yang terdapat pada proyek akhir ini, tidak terlepas dari kekurangan yang tentunya sangat diharapkan adanya saran-saran yang mendukung proses penyempurnaannya. Penulis ingin memberikan beberapa saran yang mungkin dapat membantu dalam pengembangan Tugas Akhir ini adalah sebagai berikut :

1. Pada proyek tugas akhir ini terdapat 35 data training dan 15 data testing dokumen dengan angkatan tahun 2008-2011 program studi teknik informatika yang dianalisa. Sebaiknya untuk pengembangan selanjutnya menambahkan dokumen lagi untuk mendapatkan hasil yang lebih akurat misalkan data yang digunakan adalah dokumen laporan praktek kerja lapangan dari seluruh angkatan Program Studi Teknik Informatika atau laporan praktek kerja lapangan dari Fakultas Teknik,
2. Pengklasifikasian laporan praktek kerja lapangan ini, menggunakan laporan praktek kerja lapangan sebagai dataset pada penelitian ini. Dalam pengembangan selanjutnya dapat digunakan untuk jenis dokumen lainnya misalkan laporan praktikum, dll.
3. Dalam penelitian ini menggunakan metode *jaccard measure* untuk mengklasifikasikan plagiat atau tidak plagiat pada laporan praktek kerja lapangan, penelitian ini juga dapat dikembangkan dengan menggunakan metode lainnya.

DAFTAR PUSTAKA:

- Hidayatullah Koko Aditya, Ayuninghemi Ratih. 2014. Klasifikasi Genre Film Dari Dokumen Sinopsis Film Menggunakan Cosine Measure, Jember: Universitas Muhammadiyah.
- Intan, Rolly. 2006. Subject-based Search Engine Menggunakan TF/IDF dan Jaccard Coefficient. Universita Kristen Petra Surabaya.
- Khairunisa, Nova. 2012. Aplikasi Pendeteksi Plagiat dengan Menggunakan Metode Latent Semantic Analysis (Studi Kasus : Laporan TA PCR). Riau: Politeknik Caltex Riau.
- Pratiwi, Asri, 2010. Implementasi Model Ruang Vektor Pada Sistem Temu Balik Informas

Dokumen Berbahasa Indonesia, Jakarta:
Universitas Indonesia.

Rahmah Melinda, Ayuninghemi Ratih, Setya Bagus. 2013. Deteksi Plagiasi Tugas Akhir Dengan Perhitungan Relevansi Abstrak Berbasis Cosine Similarity, Jember: Universitas Muhammadiyah.

Raymond J, Mooney. 2006. Machine Learning Text Categorization. University of Texas at Austin .

Robertson. 2005. String processing and information retrieval: 12th international conference. SPIRE.

Tan et al.2006. Introduction to Data Mining. USA: Addison Wesley.

Uncategorized . “The Java™ Tutorials”.
Tersedia :
<http://docs.oracle.com/javase/tutorial/uiswing/components.html>
[17 Desember 2014]

Widodo Agung, Ayuninghemi Ratih, Arifianto Deni. 2012. Menentukan Topik dari sebuah Tugas Akhir Menggunakan Jaccard Measure, Jember: Universitas Muhammadiyah.