

**DETEKSI KALIMAT TIDAK BAKU BERDASARKAN KAMUS BESAR
BAHASA INDONESIA DENGAN METODE *COSINE SIMILARITY*
(Studi Kasus *Text Short Message Service*)**

¹Lukie Perdanasari (11 1065 1187),
² Bagus Setya R, S.T, M.Kom, ³ Deni Arifianto, S.Kom
Program Studi Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember
E-mail : lukieperdanasari.kuliah@gmail.com

Abstrak - *Short Message Service* atau SMS adalah salah satu media komunikasi berupa pesan singkat yang sering digunakan masyarakat saat ini, bahkan antar mahasiswa dengan dosen SMS merupakan media yang masih dianggap sopan untuk berkomunikasi. Namun, banyaknya gaya penulisan dan gaya bahasa yang digunakan dalam mengirim pesan sehingga dapat mempengaruhi kalimat SMS yang digunakan menjadi kalimat tidak baku. Penelitian ini membahas kemiripan data SMS baru dengan SMS yang telah dinyatakan baku dan tidak baku untuk mendeteksi kalimat tidak baku. Acuan untuk menentukan kalimat tidak baku adalah posisi nilai tertinggi dari hasil perhitungan *cosine similarity*. Hasil pengujian dari 30 data *testing* dari dokumen SMS menunjukkan bahwa 27 bernilai benar dan 3 salah, sehingga tingkat *error* sebesar 10%. *Cosine similarity* mampu mengklasifikasi kalimat baku dan tidak baku pada teks SMS dengan tingkat *accuracy* 90%, *precision* 100%, dan *recall* 83%.

Kata kunci : baku, *Short Message Service*, *cosine similarity*, *accuracy*, *precision*, *recall*

Abstract - *Short Message Service or SMS is one of communication media in a form of short message that people often use nowadays. Among students and lecturers, even, texting is still considered polite to communicate each other. However, due to a number of writing and language styles used in texting, it affects the sentences to the nonstandard one. This study discusses the similarity between new SMS data and those determined as standard and nonstandard ones to detect nonstandard sentences. The criterion to decide nonstandard sentences is the highest position of the result of Cosine Similarity count. The result of the test is among 30 testing data from the document of SMS shows that 27 were correct while another 3 were incorrect, and the error level was 10%, therefore. Cosine Similarity can classify standard and nonstandard sentences of texting with the level of accuracy 10%, precision 100%, and recall 83%*

Keywords : *standard, Short Message Service, cosine similarity, accuracy, precision, recall*

1. PENDAHULUAN

1.1 Latar Belakang

Dengan banyaknya gaya penulisan dan gaya bahasa yang digunakan dalam mengirim pesan sehingga kalimat SMS yang digunakan menjadi kalimat tidak baku. Misalnya pada kasus mahasiswa terhadap dosen. SMS merupakan media yang dianggap sopan untuk berkomunikasi dengan orang yang dihormati namun bukan berarti mahasiswa dapat mengirimkan SMS dengan bahasa dan gaya tulisan yang berlebihan. Hal tersebut akan berubah menjadi komunikasi yang tidak sopan,

karena tidak mengandung etika berkomunikasi terhadap orang yang dihormati.

Maka dengan melihat kasus tersebut, perlu dibuat sebuah sistem untuk mendeteksi kalimat tidak baku. Untuk membangun sistem pendeteksi kemiripan pada dokumen teks diperlukan suatu ilmu yang mengacu pada pengolahan data teks. *Text mining* adalah proses menganalisis teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu (Ian H. Witten 2001:1). Tahap-tahap pada *text mining* secara umum terdiri dari *case folding*,

tokenizing, filtering, stemming, tagging dan *analyzing*.

Penelitian ini akan menguji algoritma *cosine similarity* untuk mendeteksi kalimat SMS berdasarkan kata-kata yang digunakan baik dari gaya tulisan atau gaya bahasa. *Cosine similarity* adalah salah satu metode dalam menentukan nilai kemiripan antar dua objek yang dapat diterapkan dalam menentukan nilai kemiripan pada dua berkas dokumen teks. Parameter yang digunakan adalah jumlah kata-kata pada setiap dokumen teks. Komposisi kata pada dokumen teks merupakan penentuan nilai yang dihasilkan dari metode *cosine similarity*.

1.2 Rumusan Masalah

1. Bagaimana membangun suatu sistem deteksi kalimat tidak baku berdasarkan Kamus Besar Bahasa Indonesia untuk mendeteksi kalimat pada teks SMS berdasarkan kata yang digunakan dengan metode *cosine similarity* ?
2. Bagaimana menerapkan metode *cosine similarity* pada sistem deteksi kalimat tidak baku berdasarkan Kamus Besar Bahasa Indonesia ?
3. Bagaimana menganalisa tingkat akurasi metode *cosine similarity* ?

1.3 Batasan Masalah

1. Data yang digunakan adalah 100 teks *Short Message Service*.
2. Penelitian ini mendeteksi kalimat tidak baku berdasarkan kata yang digunakan dengan pedoman Kamus Besar Bahasa Indonesia.
3. Kata pada kalimat teks *Short Message Service* tidak berupa kata yang bersifat teknis.
4. Dokumen Teks yang dibandingkan adalah dokumen teks yang seluruhnya atau sebagian besar menggunakan bahasa Indonesia.
5. Penerapan sistem hanya diterapkan pada dekstop.
6. *Wordlist* yang digunakan telah ditentukan oleh peneliti.
7. Dokumen SMS yang memiliki karakter maksimal 160 karakter.

8. Proses *text mining* yang dilakukan adalah proses *tokenizing* dan *stemming*.
9. Data *testing* dapat dilakukan dengan mengirimkan teks SMS langsung.

1.4 Tujuan

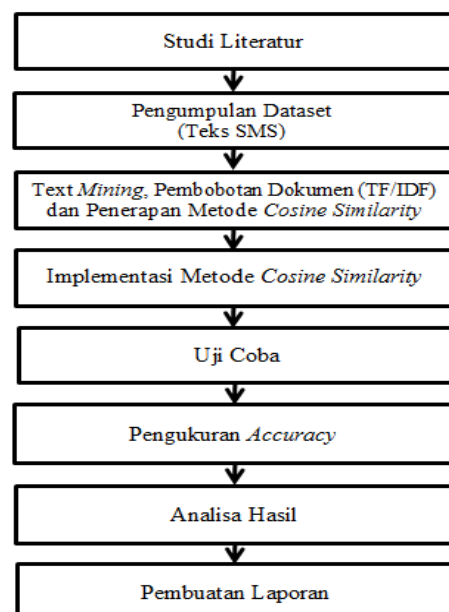
1. Membangun suatu sistem deteksi kalimat tidak baku berdasarkan Kamus Besar Bahasa Indonesia untuk mendeteksi kalimat pada teks SMS berdasarkan kata yang digunakan dengan metode *cosine similarity*.
2. Menganalisa hasil perhitungan *cosine similarity* pada studi kasus teks SMS dan mengukur tingkat akurasi.

1.5 Manfaat Penelitian

1. Dengan sistem deteksi kalimat tidak baku yang akan dibangun dengan pedoman Kamus Besar Bahasa Indonesia maka akan lebih mudah mendeteksi tingkat kebakuan kalimat pada teks SMS.
2. Dengan menerapkan metode *cosine similarity*, maka akan dapat mengetahui tingkat akurasi metode tersebut pada suatu sistem untuk mendeteksi SMS.

2. METODOLOGI PENELITIAN

Secara garis besar, metodologi penelitian dapat digambarkan dengan blok diagram dibawah :



Gambar 2.1 blok metodologi penelitian

Langkah-langkah metodologi penelitian :

1. Studi Literatur

Studi literatur adalah proses pengumpulan teori atau kajian dari buku atau laporan tugas akhir atau jurnal tugas akhir atau situs online yang berhubungan dengan metode *cosine similarity*, SMS, kalimat baku dan tidak baku serta Kamus Besar Bahasa Indonesia.

2. Pengumpulan Dataset (Teks SMS)

Pengumpulan data dilakukan dengan cara studi lapangan yang bertujuan mengumpulkan dokumen teks tertentu. Dokumen teks yang telah dikumpulkan digunakan sebagai sampel untuk melakukan uji coba sistem. Pada penelitian ini berupa dokumen berupa teks SMS dengan jumlah 100 dokumen.

3. *Text Mining*, Pembobotan Dokumen (TF/IDF) dan Penerapan *Cosine Similarity*

Penerapan *text mining* pada proses *tokenizing*, *stemming* dan *wordlist*. *Tokenizing* merupakan proses penghilangan tanda baca pada kalimat sehingga kata-kata pada kalimat dapat berdiri sendiri. *Stemming* merupakan menghilangkan kata imbuhan pada setiap kata sehingga menjadi kata dasar. Sedangkan *wordlist* merupakan daftar kata-kata baku sesuai Kamus Besar Bahasa Indonesia yang digunakan sebagai pembandingan dokumen. Pembobotan dokumen dilakukan dengan perhitungan TF/IDF. Dimana TF adalah term frekuensi yaitu menghitung jumlah term yang ada pada setiap dokumen. IDF adalah invers dokumen frekuensi yaitu hasil *log* dari pembagian jumlah dokumen dan *wordlist* dengan dokumen frekuensi (DF).

$$idf = \text{Log} \left(\frac{N}{df} \right)$$

4. Implementasi Metode *Cosine Similarity*

Pada tahap implementasi metode adalah proses penerapan metode *cosine similarity* terhadap sistem Deteksi Kalimat Tidak Baku, dengan menerapkan

rumus perhitungan berdasarkan urutan yang sebenarnya.

$$Sim(A, B) = \cos \theta = \frac{A \cdot B}{|A| \cdot |B|}$$

Keterangan :

$A \cdot B$ = Nilai yang mengekspresikan sudut antara dua vektor

$|A|$ = Panjang Vektor

5. Uji Coba

Merupakan proses pengujian sistem terhadap studi kasus yang berupa teks SMS. Proses ini dilakukan untuk pengujian kelayakan sistem dalam mendeteksi teks SMS.

6. Pengukuran *Accuracy*

Perhitungan *accuracy* adalah sebagai pembuktian tingkat keakuratan antara analisa kalimat tanpa perhitungan *cosine similarity* dengan hasil perhitungan dengan metode *cosine similarity*.

7. Analisa Hasil

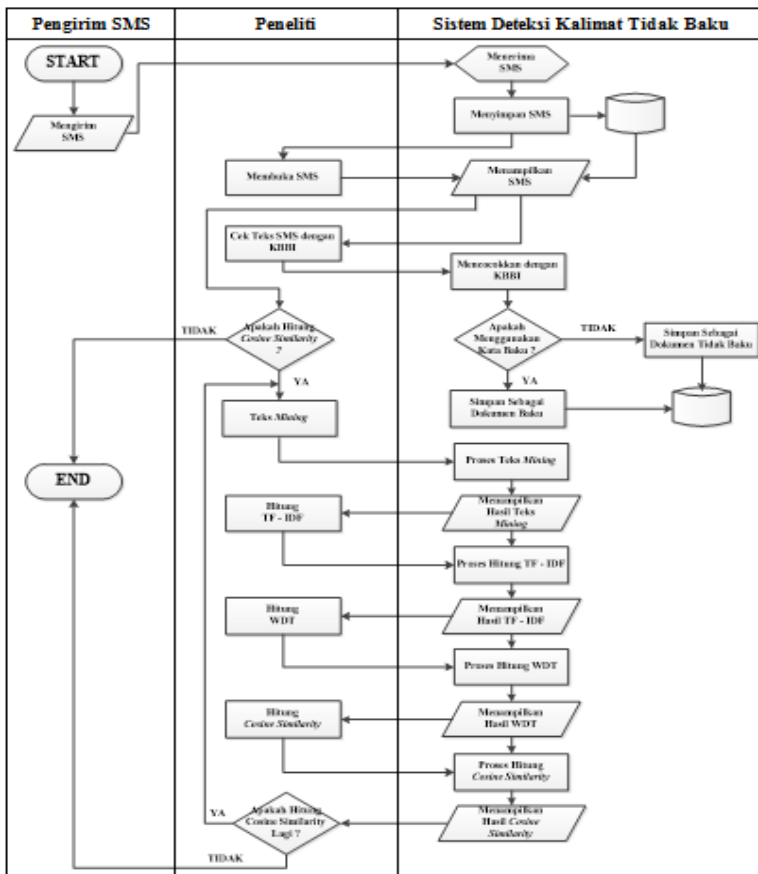
Setelah diperoleh dari hasil perhitungan *cosine similarity* dan perhitungan *accuracy*, maka tahap selanjutnya adalah menganalisa hasil penelitian deteksi kalimat tidak baku dengan metode *cosine similarity*. Hal ini dilakukan untuk mengevaluasi seberapa tingkat keberhasilan dalam mendeteksi kalimat tidak baku dengan perhitungan *cosine similarity*.

8. Pembuatan Laporan

Pembuatan laporan adalah tahap terakhir dari proses penelitian deteksi kalimat tidak baku dengan metode *cosine similarity*. Dalam laporan akan memaparkan hasil dari penelitian ini yang disusun dalam Laporan Hasil Penelitian.

Dibawah ini adalah flowchart sistem Deteksi Kalimat Tidak Baku Berdasarkan Kamus Besar Bahasa Indonesia Dengan Metode *Cosine Similarity*. *Flowchart* sistem merupakan gambaran yang menunjukkan arus pekerjaan secara keseluruhan dari sistem. *Flowchart* sistem ini menjelaskan urutan-urutan dari prosedur-prosedur yang ada di dalam

sistem. Bagan alir sistem menunjukkan apa yang dikerjakan di sistem.



Gambar 2.2 FlowChart Sistem

3. HASIL DAN PEMBAHASAN

3.1 Data

Data SMS yang digunakan berjumlah 100 data SMS, dimana terdiri dari 50 data SMS baku yang telah dilakukan proses uji cek kebakuan dan terdiri dari 50 data SMS tidak baku.

3.2 Rekayasa Pengujian

1. Rekayasa Pengujian Dataset SMS Baku

Rekayasa pengujian dataset menggunakan 50 data SMS yang diuji sebagai data SMS baku, 35 sebagai data *training* yang kemudian dikumpulkan isi dari 35 SMS tersebut menjadi 1 dokumen baku dan 15 sebagai data *testing*. Data sms diuji menggunakan proses *text mining* yang hanya dilakukan dua proses yaitu *tokenizing* dan *stemming*.

2. Rekayasa Pengujian Dataset SMS Tidak Baku

Rekayasa pengujian untuk SMS tidak baku menggunakan data SMS berjumlah 50, dimana 35 sebagai data *training* yang kemudian dikumpulkan isi dari 35 SMS tersebut menjadi 1 dokumen baku dan 15 sebagai data *testing*. Proses pengujian sama halnya dengan proses pengujian dataset SMS baku.

3. Rekayasa Pengujian SMS Baru

Rekayasa pengujian SMS baru menyediakan 2 jenis SMS, yang mana terdiri SMS baku dan SMS tidak baku. Data SMS baru yang dimaksud adalah berupa data *testing*, SMS baru didapatkan dari hasil SMS langsung dengan memanfaatkan SMS gateway. Data *testing* selanjutnya diproses menggunakan *text mining* melalui proses *tokenizing* dan *stemming*. Hasil proses *stemming* akan dilakukan proses perhitungan TF/IDF, dimana data SMS baru sebagai *keyword*.

4. Rekayasa Pengujian Perhitungan TF/IDF dan Cosine Similarity

Rekayasa pengujian perhitungan TF/IDF dan *cosine similarity*, pertama dilakukan perhitungan TF/IDF dengan menggunakan SMS baru yang berjumlah 30 yang terdiri dari 15 SMS baku dan 15 SMS tidak baku sebagai *keyword* yang dibandingkan dengan 1 dokumen berisi kumpulan dari 35 SMS dinyatakan baku dan 1 dokumen berisi kumpulan dari 35 SMS dinyatakan tidak baku. Hasil perhitungan TF/IDF kemudian dilakukan proses pembobotan yaitu dengan mengalikan hasil perhitungan IDF dengan nilai kata pada setiap dokumen. *Cosine similarity* adalah proses terakhir untuk menentukan proses klasifikasi antara SMS baku atau SMS tidak baku dengan melihat letak hasil nilai tertinggi, apabila nilai *cosine* tertinggi terletak pada dokumen 1 maka SMS dinyatakan baku dan apabila nilai *cosine* tertinggi terletak pada dokumen 2 maka SMS dinyatakan tidak baku.

3.3 Analisa Hasil

1. Hasil Pengujian

Setiap perhitungan *cosine similarity* dengan menggunakan sistem akan dibandingkan hasilnya dengan *handtrace*. *Handtrace* merupakan proses perhitungan *cosine similarity* dengan manual oleh analis sebagai jawaban yang benar dari dataset. Perhitungan manual ini berdasarkan kesesuaian makna dan term potensial dari kebutuhan dalam data pengujian, sehingga total jumlah kebutuhan yang terambil (*retrieved*), dan jumlah kebutuhan yang terambil dan relevan berdasarkan *handtrace* (*relevan*) akan membentuk kinerja rata-rata *precision*, *recall* dari semua dataset.

2. Hasil Rekayasa Pengujian Dataset SMS Baku

Dari hasil proses perhitungan *cosine similarity* terdapat hasil perhitungan terhadap data testing SMS baku menghasilkan kedekatan terhadap beberapa dokumen *training*.

Tabel 3.1. Tabel Hasil Perhitungan *Cosine* SMS Baku

	D1	D2	KET
Dokumen Baku 1	0.012056590165888259	0.0032268210562682647	TRUE
Dokumen Baku 2	0.03898342310631558	0.0017164967461149013	TRUE
Dokumen Baku 3	0.024961757888606474	0.0032379555883664335	TRUE
Dokumen Baku 4	0.025330819250107245	0.003218899015828305	TRUE
Dokumen Baku 5	0.017978138540958873	0.0018183629988445067	TRUE
Dokumen Baku 6	0.09077273270020979	0.0	TRUE
Dokumen Baku 7	0.02946847006660189	0.009846792763635212	TRUE
Dokumen Baku 8	0.028651445008195133	0.0	TRUE
Dokumen Baku 9	0.015426551797050308	0.008198473315172098	TRUE
Dokumen Baku 10	0.051759774638496125	0.0	TRUE
Dokumen Baku 11	0.03792000492286103	0.0038348289091016444	TRUE
Dokumen Baku 12	0.033179783071719746	0.005053156484900092	TRUE
Dokumen Baku 13	0.04144127222272463	0.006311338696256226	TRUE
Dokumen Baku 14	0.09366740604364067	0.004930241797619772	TRUE
Dokumen Baku 15	0.03110873797858985	0.0024755336366644983	TRUE

3. Rekayasa Pengujian Dataset SMS Tidak Baku

Dari hasil proses perhitungan *cosine similarity* terdapat hasil perhitungan terhadap data testing SMS tidak baku menghasilkan kedekatan terhadap beberapa dokumen *training*.

Tabel 3.2. Tabel Hasil Perhitungan *Cosine*

	D1	D2	KET
Dokumen Tidak Baku 1	0.00169356537584661	0.02288933645983548	TRUE
Dokumen Tidak Baku 2	0.0	0.01824594745262993	TRUE
Dokumen Tidak Baku 3	0.01338942086168753	0.13724374451677357	TRUE
Dokumen Tidak Baku 4	0.01406317008938688	0.1888821629320818	TRUE
Dokumen Tidak Baku 5	0.02281557393744267	0.009949466080324898	FALSE
Dokumen Tidak Baku 6	0.01313334523013747	0.008461205986865462	FALSE
Dokumen Tidak Baku 7	0.0	0.01469213230269348	TRUE
Dokumen Tidak Baku 8	0.00121555218397077	0.02720146034075759	TRUE
Dokumen Tidak Baku 9	0.01380088079725772	0.021622848364399525	TRUE
Dokumen Tidak Baku 10	0.01644047176705948	0.030308133880122698	TRUE
Dokumen Tidak Baku 11	0.02362123861929974	0.016968421687483828	FALSE
Dokumen Tidak Baku 12	0.00549522493596751	0.07940291552858114	TRUE
Dokumen Tidak Baku 13	0.00295531127712457	0.02977757023834712	TRUE
Dokumen Tidak Baku 14	0.00418118880997579	0.03107307938376193	TRUE
Dokumen Tidak Baku 15	0.00112939985130927	0.08520539098836084	TRUE

2. Accuracy

Tabel 3.3. Perhitungan *Accuracy*

TP (data baku terklasifikasi baku)	FP (data baku terklasifikasi tidak baku)
15	0
FN (data tidak baku terklasifikasi baku)	TN (data tidak baku terklasifikasi tidak baku)
3	12

Dilakukan perhitungan sebagai berikut :

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

$$Accuracy = \frac{15 + 12}{15 + 12 + 0 + 3} = 0,9 \times 100\% = 90\%$$

Hasil perhitungan *accuracy* menunjukkan bahwa tingkat *accuracy* adalah 90%.

3. Precision

Tabel 3.4. Perhitungan Precision

TP (data baku terklasifikasi baku)	FP (data baku terklasifikasi tidak baku)
15	0
FN (data tidak baku terklasifikasi baku)	TN (data tidak baku terklasifikasi tidak baku)
3	12

Dilakukan perhitungan sebagai berikut :

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

$$Precision = \frac{15}{15 + 0} = 1 \times 100\% = 100\%$$

Hasil perhitungan *precision* menunjukkan bahwa tingkat *precision* adalah 100%.

4. Recall

Tabel 3.5. Perhitungan Recall

TP (data baku terklasifikasi baku)	FP (data baku terklasifikasi tidak baku)
15	0
FN (data tidak baku terklasifikasi baku)	TN (data tidak baku terklasifikasi tidak baku)
3	12

Dilakukan perhitungan sebagai berikut :

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

$$Recall = \frac{15}{15 + 3} = 0.83 \times 100\% = 83\%$$

Hasil perhitungan *precision* menunjukkan bahwa tingkat *precision* adalah 83%.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

- Setelah melakukan pengujian dengan 30 dokumen *testing* terhadap sistem yang terdiri dari 15 dokumen data SMS baku dan 15 dokumen data SMS tidak baku menghasilkan nilai *cosine* dengan klasifikasi:
 - Dokumen baku berjumlah 15 dokumen

- Dokumen tidak baku berjumlah 12 dokumen

Menunjukkan bahwa hasil perhitungan memiliki *error* sebesar 10%.

- Pengujian dokumen SMS memiliki hasil perhitungan *accuracy* mampu menghasilkan nilai sebesar 90%, *precision* mampu mengasilkan sebesar 100%, dan *recall* sebesar 83%.
- Dari hasil perhitungan *accuracy*, *precision* dan *recall* menunjukkan bahwa hasil perhitungan *precision* lebih tinggi dari pada *accuracy* dan *recall*.
- Hasil perhitungan *precision* sebesar 100% dan *recall* sebesar 83%, memiliki selisih yang tidak terlalu besar yaitu 17% . Hal tersebut dikarenakan jumlah yang relevan lebih besar jika dibandingkan dengan yang tidak relevan, dari hasil tersebut menunjukkan bahwa ketepatan sistem untuk tidak memanggil yang tidak relevan lebih tinggi dari pada kemampuan sistem untuk memanggil dokumen yang relevan. Dengan demikian, metode *cosine* layak digunakan untuk mendeteksi kalimat baku dan kalimat tidak baku dengan studi kasus teks SMS.

4.2 Saran

- Proyek tugas akhir yang dikerjakan oleh penulis mengambil studi kasus teks SMS, dimana mengambil dataset 100 data SMS yang dibagi menjadi 50 SMS baku dan 50 SMS tidak baku, dataset dari 50 data SMS baku dan 50 data SMS tidak baku dibagi lagi menjadi menjadi 35 SMS sebagai data *training* dan 15 SMS sebagai data *testing*. Sebaiknya untuk pengembangan proyek ini menambahkan dokumen lebih banyak lagi sehingga mendapatkan hasil yang lebih akurat. Misalnya, 100 data SMS untuk SMS baku dan 100 data SMS untuk SMS tidak baku.
- Dalam proses teks *mining* terhadap data *testing* membutuhkan waktu sekitar 3 menit. Sehingga diharapkan dalam pengembangan dapat bisa melakukan proses teks *mining* dengan waktu yang lebih sebentar.

3. Pada penelitian deteksi kalimat baku menggunakan studi kasus SMS, namun dalam pengembangan diharapkan dapat mengambil studi kasus yang lain seperti teks e-mail atau bahkan artikel atau cerpen yang kaitannya teks yang digunakan harus menggunakan kata-kata baku dan mempunyai kata yang lebih banyak dan variasi.

5. DAFTAR PUSTAKA

- Ernawati., Andreswari, Desi. 2013. *Aplikasi Pendeteksi Plagiarism Pada Dokumen Teks*. Teknik Informatika Universitas Bengkulu.
- Firdaus, Azhar. 2013. *Aplikasi Pendeteksi Kemiripan Pada Dokumen Teks menggunakan Algoritma Nazief & Adriani Dan Metode Cosine Similarity*. Program Studi Informatika Fakultas Teknik Universitas Bengkulu.
- Garcia,E. 2006. *Vector Models Based on Normalized Frequencies : Improving Word Weights with Normalized Frequencies*. (Online). (<http://www.miiistalita.com/word-vector/word-vector-4.html>). Diakses tanggal 3 November 2014.
- Han, J. dan Kamber, M. 2006. *Data mining: Concepts and techniques (2nd ed.)*. Elsevier Inc.
- Hatta, Ahmad., et al. 2012. Rancang Bangun Sistem Pengelolaan Dokumen-dokumen Penting Menggunakan Text Mining. (Online). (<http://digilib.its.ac.id/public/ITS-Undergraduate-14856-paperpdf.pdf>). Diakses tanggal 3 November 2014.
- Huda, Miftahul dan Bunafit Komputer. 2010. *Membuat Aplikasi Database dengan Java, MySQL dan Netbeans*. Jakarta: PT. Elex Media Komputindo.
- Mandala, Rila. dan Setiawan, Hendra. 2002. *Peningkatan Performansi Sistem Temu-Kembali Informasi dengan Perluasan Query Secara Otomatis*. Bandung : Departemen Teknik Informatika Institut Teknologi Bandung.
- Maulana, Husada. 2011. *Kedudukan dan Fungsi Kamus Besar Bahasa Indonesia di Lingkungan Akademisi*. (Online). (<http://kombasasin.blogspot.com/2011/01/kedudukan-dan-fungsi-kamus-besar-bahasa.html>). Diakses pada tanggal 28 Oktober 2014.
- Mukhtidi, Karimul. 2012. *Sistem Sms Spam Detector Untuk Sms Berbahasa Indonesia Pada Smartphone Android*. Bogor: Departemen Ilmu Komputer Fakultas Matematika Dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.
- Pusat Pembinaan dan Pengembangan Bahasa. 2008. *Kamus Besar Bahasa Indonesia*. Jakarta : Balai Pustaka.
- Segall, Richard S., Guha, Gauri S., dan Nonis, Sarath A. 2008. *Data Mining of Environmental Stress Tolerances for Plants*. Kybernetes: The International Journal of Cybernetics, Systems and Management Sciences.
- Tudesman., Oktalina, Enny., Tinaliah., & Yoannita. 2013. *Sistem Deteksi Plagiarism Dokumen Bahasa Indonesia Menggunakan Metode Vector Space Model*. Program Studi Teknik Informatika STMIK GI MDP.
- Tyas. 2012. *SMS Gateway*. (Online). (<http://informatika.web.id/about>). Diakses pada tanggal 7 November 2014
- W, Ebranda A., dan Triana, Novi. 2013. *Penerapan Metode Naïve Bayes Untuk Sistem Klasifikasi Sms Pada Smartphone Android*. Jurusan Teknik Informatika STMIK MDP.
- Wiharto, Yudi. 2011. *Sistem Informasi Akademik Berbasis Sms Gateway*. Palembang: Politeknik PalComTech Palembang.
- Witten, Ian H. 2001. *Adaptive Text Mining: Inferring Structure From Sequences*. New Zealand: Department of Computer Science University of Waikato Hamilton New Zealand.
- Zuvmask. 2013. *Pengertian Kata Baku dan Tidak Baku*. (Online). (<http://belajarpraktis.com/2013/09/06/pengertian-kata-baku-dan-tidak-baku.html>). Diakses pada tanggal 7 November 2014.