

## PERANCANGAN DAN PEMBUATAN APLIKASI UNTUK MENDETEKSI KEMIRIPAN JAWABAN MENGGUNAKAN COSINE SIMILARITY

Nurlaily Vendyansyah<sup>1</sup>, Yosep Agus Pranoto<sup>2</sup>

<sup>1,2</sup>Teknik Informatika – ITN Malang

E-mail : [nurlaily.vendyansyah@gmail.com](mailto:nurlaily.vendyansyah@gmail.com)<sup>1</sup>

### ABSTRACT

*The education process is basically transferring knowledge from a teacher to a student. To find out how much success in the educational process, an evaluation is carried out in the form of an exam. One of the exams given by a teacher is by giving essay questions. Along with the number of answers to essay questions that must be checked by a teacher, it can cause the quality of the assessment to decrease and even tend to be no longer objective. This is a serious problem because it can harm students. In information technology, this problem can be solved with the text similarity analysis method using the Cosine Similarity Method [1][2][3][4][5][6]. In this case, used training data in the field of biology, kingdom plantae which is implemented in essay exams in Indonesian with the grammar of each student. In testing the computing system is compared with the results of direct interviews with expert teachers. The results showed the accuracy of the error system 4%.*

**Keywords:** Cosine Similarity, Essay Examination, Weighting, Term Frequency

### ABSTRAK

Proses pendidikan pada dasarnya adalah melakukan transfer ilmu pengetahuan dari seorang guru kepada murid. Untuk mengetahui seberapa besar keberhasilan pada proses pendidikan, maka dilakukan sebuah evaluasi berupa ujian. Ujian yang diberikan oleh seorang Guru salah satunya adalah dengan memberikan soal *essay*. Seiring dengan banyaknya jawaban soal *essay* yang harus diperiksa oleh seorang guru, dapat menyebabkan kualitas penilaian menurun bahkan cenderung tidak objektif lagi. Hal ini menjadi sebuah permasalahan yang serius karena dapat merugikan siswa. Dalam teknologi informasi permasalahan ini dapat diselesaikan dengan metode analisis kemiripan teks dengan menggunakan *Metode Cosine Similarity*[1][2][3][4][5][6]. Dalam penelitian ini menggunakan data latih dalam bidang ilmu biologi yaitu kingdom *plantae* yang diimplementasikan dalam ujian *essay* dalam bahasa Indonesia dengan tata bahasa siswa masing-masing. Dalam pengujian komputasi sistem dibandingkan dengan hasil wawancara langsung dengan pakar guru. Hasil penelitian menunjukkan akurasi sistem adalah error sebesar 4%.

**Kata kunci :** Cosine Similarity, Ujian Essay, Pembobotan, term Frequency

## 1. PENDAHULUAN

Pendidikan merupakan salah satu faktor penting bagi kemajuan bangsa karena melalui pendidikan manusia akan memperoleh pengetahuan yang dapat digunakan untuk menciptakan hal-hal baru yang bermanfaat bagi kesejahteraan sebuah bangsa. Proses pendewasaan manusia tidak terlepas dari pendidikan yang dapat diperoleh secara formal maupun non formal. Pendidikan formal bisa ditempuh pada bangku sekolah dasar sampai dengan perguruan tinggi, sedangkan pendidikan non formal bisa didapatkan dari lembaga – lembaga kursus.

Proses pendidikan pada dasarnya adalah melakukan transfer ilmu pengetahuan dari seorang guru kepada murid. Untuk mengetahui seberapa besar keberhasilan pada proses pendidikan, maka dilakukan sebuah evaluasi berupa ujian. Pada proses ujian, seorang guru akan memberikan beberapa soal terkait dengan

materi yang telah diajarkan untuk dikerjakan oleh siswa, kemudian guru akan mengevaluasi seberapa besar tingkat keberhasilan siswa dengan melakukan pencocokan antara kunci jawaban soal dengan jawaban yang sudah dikerjakan. Bentuk – bentuk soal yang diberikan oleh seorang guru juga beragam, dapat berupa pilihan ganda maupun jawaban *essay*. Pada soal pilihan ganda, siswa akan dihadapkan pada beberapa pilihan jawaban, siswa harus memilih jawaban yang paling tepat dari sebuah soal. Pada soal *essay* siswa tidak diberikan pilihan jawaban, akan tetapi siswa harus menjawab dengan menuliskan kalimat, yang artinya siswa akan dituntut untuk menyampaikan informasi secara verbal, selain itu ujian esai juga menuntut pemahaman yang lebih baik akan suatu ilmu dan dapat digunakan untuk mengukur tingkat pemahaman manusia akan suatu ilmu secara lebih mendalam [1].

Soal essay juga mempunyai beberapa kelemahan terutama pada proses koreksi jawaban yang diberikan oleh siswa. Seorang guru harus memeriksa dengan cara membaca jawaban berupa kalimat yang panjang, hal ini tentunya membutuhkan sebuah konsentrasi yang tinggi. Seiring dengan banyaknya jawaban siswa yang harus diperiksa oleh seorang guru, dapat menyebabkan kualitas penilaian menurun bahkan cenderung tidak objektif lagi. Hal ini menjadi sebuah permasalahan yang serius karena dapat merugikan siswa. Selain itu, permasalahan yang timbul dalam hal koreksi jawaban essay adalah membutuhkan waktu yang relatif lama.

Perkembangan teknologi dalam bidang informatika dapat dimanfaatkan untuk menanggulangi masalah tersebut. Kecerdasan buatan merupakan salah satu bidang pada informatika yang mencangkokkan sebuah kecerdasan pada sebuah mesin yang dapat digunakan untuk menggantikan peran dari manusia. Penelitian yang dilakukan oleh W.Rizki Tri, dkk menerapkan Cosine Similarity untuk melakukan pengelompokan dokumen skripsi. Hasilnya menunjukkan bahwa tingkat ketepatan sistem sebesar 98%, dengan memberikan inputan data latih berupa judul skripsi. Tata kata penulisan judul skripsi adalah telah menggunakan kaidah tata bahasa indonesia yang menganut berdasarkan KBBI. Permasalahan yang muncul adalah bagaimana Cosine Similarity dapat melakukan analisis kemiripan jawaban dengan menggunakan bahasa indonesia dengan gaya bahasa siswa masing-masing yang di inputkan oleh siswa dalam kasus ujian *essay* ilmu biologi ?

Pada penelitian ini dibuat aplikasi menggunakan metode *Cosine Similarity* untuk mendeteksi seberapa besar kemiripan jawaban siswa dengan kunci jawaban yang benar. Dengan menggunakan aplikasi ini diharapkan dapat memberikan solusi dari masalah yang muncul pada proses koreksi jawaban essay.

## 2. METODE

Flowchart sistem pada penelitian ini ditunjukkan pada Gambar 1. Pada sistem yang dibuat mulai dengan proses preprocessing dari jawaban guru dan siswa. Preprocessing dilakukan dengan menggunakan text mining terhadap teks untuk mempersiapkan teks menjadi data yang dapat diolah lebih lanjut [2]. Proses text mining terdiri dari :

### 1. Tokenizing

Tahap *tokenizing* atau parsing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini adalah memisahkan setiap kata yang menyusun

suatu dokumen. [3]

### 2. Case Folding

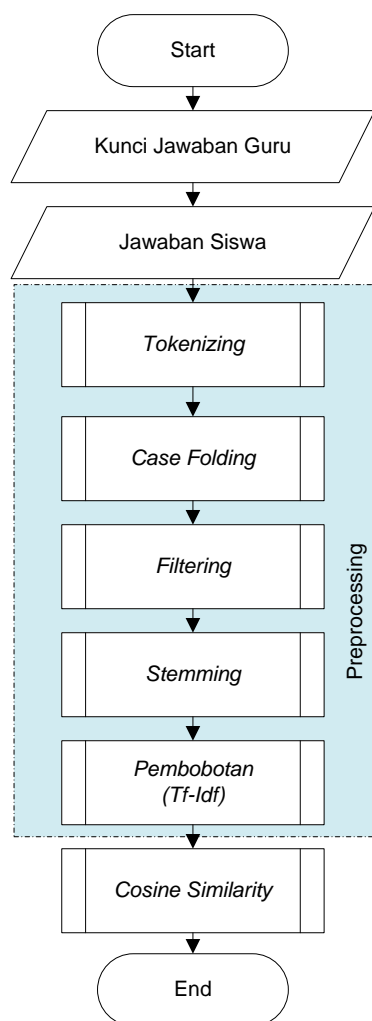
Case folding, merupakan proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima karakter-karakter dan tanda baca lainnya selain huruf juga dihilangkan. Tidak semua dokumen teks konsisten dalam penggunaan huruf capital. [4]

### 3. Filtering

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil token. Pada tahap ini akan dilakukan pembersihan *tweet* dari spesial karakter, URL link, *username*, serta emoticon. [3]

### 4. Stemming

*Stemming* adalah suatu teknik pencarian bentuk dasar dari suatu *term*. Yang dimaksud dengan *term* itu sendiri adalah tiap kata yang berada pada suatu dokumen teks. [2]



Gambar 1. Flowchart sistem

Setelah proses text mining dilakukan maka selanjutnya yaitu melakukan pembobotan *Term Frequency Inverse Document Frequency* (Tf-Idf). Metode Tf-Idf 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. Frekuensi kemunculan kata didalam dokumen yang diberikan menunjukkan seberapa penting kata tersebut didalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi didalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen (database). Rumus umum untuk Tf-Idf [6].

$$w_{ij} = tf \times idf \quad (1)$$

dimana

$w_{ij}$  = bobot kata/term tj terhadap dokumen di  
 $tf_{ij}$  = jumlah kemunculan kata/term tj dalam di  
 N = jumlah semua dokumen yang ada dalam database  
 n = jumlah dokumen yang mengandung kata/term ij (minimal ada satu kata yaitu term ij)

Berdasarkan persamaan (1), berapapun besarnya nilai  $tf_{ij}$  apabila  $N = n$  maka akan didapatkan hasil 0 (nol) untuk perhitungan Idf. Untuk itu dapat ditambahkan nilai 1 pada sisi Idf, sehingga perhitungan bobotnya menjadi seperti persamaan (2): [6]

$$w_{ij} = tf_{ij} \times (\log (N/n) + 1) \quad (2)$$

Setelah dilakukan pembobotan Tf-Idf, maka dilanjutkan dengan menghitung kemiripan menggunakan metode *Cosine Similarity*. *Cosine Similarity* adalah ukuran kesamaan antara dua buah vektor dalam sebuah ruang dimensi yang didapat dari nilai cosinus sudut dari perkalian dua buah vektor yang dibandingkan karena cosinus dari  $0^0$  adalah 1 dan kurang dari 1 untuk nilai sudut yang lain, maka nilai *similarity* dari dua buah vektor dikatakan mirip ketika nilai dari *Cosine Similarity* adalah 1. *Cosine similarity* digunakan dalam ruang positif dimana hasilnya dibatasi antara nilai 0 dan 1. Rumus umum *Cosine Similarity* tersebut

Information Retrieval, masing-masing kata/istilah (term) diasumsikan sebagai dimensi yang berbeda dan dokumen ditandai dengan vector dimana nilai masing-masing dimensi sesuai dengan berapa istilah muncul dalam dokumen. Berikut adalah rumus cosine similarity Rumus umum *cosine similarity* [6]:

$$\frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3)$$

dimana

A = vektor

B = vektor

$A_i$  = bobot term i dalam blok  $A_i$

$B_i$  = bobot term i dalam blok

i = jumlah term dalam kalimat

n = jumlah vector

A merupakan bobot setiap ciri pada vector A, dan B merupakan bobot setiap ciri pada B. Jika dikaitkan dengan *information retrieval* maka A adalah bobot setiap istilah pada dokumen A, dan B merupakan bobot setiap istilah pada dokumen B.  $w_{ij} = tf \times idf$  digunakan *cosine similarity* k  
 n  $w_{ij} = tf_{ij} \times \log \frac{N}{n}$  merupakan dimensi yang berbeda dan nilai warna pada setiap *pixel* tersebut merupakan nilai dari setiap dimensi. Pengukuran kemiripan dapat dilakukan dengan membandingkan dokumen 1 dengan dokumen 2 kemudian sistem akan menghitung nilai kemiripan.  $A_i B_i$  adalah nilai yang diperoleh dari term A dan term B kemudian kedua nilai tersebut dijumlahkan, kemudian nilai  $A_i$  semua nilai term dokumen A semua nilainya dipangkatkan dua, begitu juga dengan term  $B_{i2}$  semua nilai yang diperoleh dipangkatkan dua kemudian semua nilai yang diperoleh dijumlahkan. [6].

### 3. PEMBAHASAN

Pada aplikasi yang dibuat terdapat form yang digunakan oleh guru untuk memasukkan soal dan kunci jawaban seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Form input soal dan kunci jawaban

Setelah guru memasukkan soal dan kunci jawaban, maka siswa dapat mengikuti ujian dengan menjawab soal pada form input jawaban ujian seperti pada Gambar 3



Gambar 3. Form input jawaban siswa

Langkah selanjutnya, guru dapat melakukan koreksi jawaban yang diberikan oleh siswa menggunakan metode *cosine similarity* dengan cara masuk ke dalam form monitoring nilai siswa seperti yang ditunjukkan pada Gambar 4



Gambar 4. Form monitoring nilai siswa

Nilai yang didapatkan oleh siswa tergantung pada tingkat kemiripan jawaban dengan kunci jawaban yang diberikan. Guru dapat melihat bobot tingkat kemiripan jawaban serta nilai siswa pada form seperti yang ditunjukkan pada Gambar 5.

Soal dan Jawaban Siswa		Nilai
Soal: Jelaskan perbedaan struktur batang pada lumut hati, lumut tanduk, dan lumut daun!	Jawaban: Lumut hati dan lumut tanduk tidak berbatang dan tidak mempunyai pembuluh angkut.	Nilai bobot 0,780%
Soal: Jelaskan perbedaan struktur batang pada lumut hati, lumut tanduk, dan lumut daun!	Jawaban: Lumut hati dan lumut tanduk mempunyai batang sederhana dengan pembuluh angkut tunggal.	Nilai bobot 0,800%
Soal: Jelaskan perbedaan struktur batang pada lumut hati, lumut tanduk, dan lumut daun!	Jawaban: Lumut hati berstruktur lembek, lumut tanduk berstruktur kuat dan lumut daun strukturnya sering tumbuh di bagian luar.	Nilai bobot 0,850%
Soal: Jelaskan perbedaan struktur batang pada lumut hati, lumut tanduk, dan lumut daun!	Jawaban: Lumut hati berbentuk lembaran, lumut tanduk berbentuk kapsul memanjang seperti tanduk, lumut daun memiliki batang yang tegak dan bercabang.	Nilai bobot 0,860%
Soal: Jelaskan perbedaan struktur batang pada lumut hati, lumut tanduk, dan lumut daun!	Jawaban: Lumut tanduk : batangnya lebih keras, berbentuk seperti lumut hati tapi sporofitnya berbentuk kapsul memanjang. Lumut hati : lumut dengan bentuk lembaran yang ditemukan di daerah lembab. Lumut daun : lumut ini mempunyai bentuk batang yang tegak dan bercabang.	Nilai bobot 0,120%

**NILAI SISWA : 89.85**

Gambar 5. Form bobot kemiripan jawaban dan nilai

Untuk memastikan kebenaran metode *cosine similarity* dari sistem aplikasi yang dibuat, maka dilakukan pengujian dengan cara membandingkan *output* dari aplikasi dengan proses perhitungan secara manual. Skenario pengujian dilakukan dengan cara membuat 1 buah soal beserta kunci jawaban dan membandingkan dengan jawaban dari 3 orang siswa. Soal dan Kunci Jawaban guru serta hasil preprosesing ditunjukkan pada Tabel 1.

Tabel 1. Soal dan Kunci Jawaban serta hasil preprosesing

Soal Guru	Kunci Jawaban Guru	Hasil Preprocessing
Jelaskan perbedaan struktur batang pada lumut hati, lumut tanduk, dan lumut daun!	-Lumut hati dan lumut tanduk tidak berbatang dan tidak mempunyai pembuluh angkut.  -Lumut daun mempunyai batang sederhana dengan pembuluh angkut tunggal	lumut hati lumut tanduk batang buluh angkut lumut daun batang sederhana buluh angkut tunggal

Jawaban yang diberikan oleh 3 orang siswa serta hasil preprosesing ditunjukkan pada Tabel 2.

Tabel 2. Jawaban siswa serta hasil preprosesing

Dokumen	Nama	Jawaban Siswa	Hasil Jawaban Siswa Yang Telah Diprocessing
1	Siswa 1	Lumut hati berstruktur lembek, lumut tanduk berstruktur kuat dan lumut daun strukturnya sering tumbuh di bagian luar	lumut hati struktur lembek lumut tanduk struktur kuat lumut daun struktur tumbuh
2	Siswa 2	Lumut hati berbentuk lembaran, lumut tanduk berbentuk kapsul memanjang seperti tanduk, lumut daun memiliki batang yang tegak dan bercabang	lumut hati bentuk lembaran lumut tanduk kapsul panjang lumut daun tegak cabang
3	Siswa 3	Lumut tanduk : batangnya lebih keras, berbentuk seperti lumut hati tapi sporofitnya berbentuk kapsul memanjang. Lumut hati : lumut dengan bentuk lembaran yang ditemukan di daerah lembab. Lumut daun : lumut ini mempunyai bentuk batang yang tegak dan bercabang.	lumut tanduk batang keras bentuk lumut hati sprofit berbentukkapsul memanjang lumut hati lumut bentuk lembaran temu daerah lembab lumut daun lumut bentuk batang tegak bercabang

Tabel 3 menunjukkan perhitungan bobot

term frequency Kemunculan kata kunci jawaban dan jawaban yang diberikan oleh siswa.

**Tabel 3. Bobot Term frequency**

Daftar Kata	Q	D1	D2	D3
lumut	3	3	3	6
hati	1	1	1	2
tanduk	1	1	1	1
batang	2	0	1	2
buluh	2	0	0	0
angkut	2	0	0	0
daun	1	1	1	1
sederhana	1	0	0	0
tunggal	1	0	0	0
lembek	0	0	0	0
struktur	0	3	0	0
kuat	0	0	0	0
bentuk	0	1	2	3
lembar	0	1	1	1
kapsul	0	1	1	0
anjang	0	1	0	0
milik	0	1	1	0
tegak	0	1	1	1
cabang	0	1	1	0
keras	0	0	0	1
sprofit	0	0	0	1

Daftar Kata	Q	D1	D2	D3
berbentukkapsul	0	0	0	1
memanjang	0	0	0	1
bercabang	0	0	0	1
temu	0	0	0	1
daerah	0	0	0	1
lembah	0	0	0	1

Keterangan

Q : Kunci Jawabanguru

D : Jawaban siswa

Berdasarkan data bobot hubungan suatu kata *Term Frequency* (Tf) yang sudah didapat seperti pada Tabel 3, maka dilanjutkan dengan proses perhitungan kemiripan jawaban menggunakan *Cosine Similarity*.

Berdasarkan hasil perhitungan manual, maka tingkat kemiripan kunci jawaban dengan jawaban siswa D1 = 0,480, siswa D2 = 0,576 dan siswa D3 = 0,632.

$$D_1 = \frac{(3x3) + (1x1) + (1x1) + (2x0) + (2x0) + (2x0) + (1x1) + (1x0) + (1x0) + (1x0) + (0x3) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1)}{\sqrt{3^2 + 1^2 + 1^2 + 2^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2}} = 0,480$$

$$D_2 = \frac{(3x3) + (1x1) + (1x1) + (2x1) + (2x0) + (2x0) + (1x1) + (1x0) + (1x0) + (0x2) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1)}{\sqrt{3^2 + 1^2 + 1^2 + 2^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2}} = 0,576$$

$$D_3 = \frac{(3x6) + (1x2) + (1x1) + (2x2) + (2x0) + (2x0) + (1x1) + (1x0) + (1x0) + (0x3) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1)}{\sqrt{6^2 + 2^2 + 1^2 + 2^2 + 1^2 + 3^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2}} = 0,632$$

Output perhitungan dari aplikasi ditunjukkan pada Gambar 6.



Gambar 6. Output perhitungan aplikasi

Perbandingan hasil perhitungan metode secara manual dengan output perhitungan dari aplikasi ditunjukkan pada Tabel 4.

**Tabel 4.** Perbandingan hasil aplikasi dan hasil manual

Siswa yang di uji	Hasil Aplikasi	Hasil Manual
Siswa 1	0.48	0,48
Siswa 2	0.57	0,576
Siswa 3	0.63	0,632

Selisih perhitungan D1 = 0,48 – 0,48 = 0  
 Selisih perhitungan D2 = 0,576 – 0,57 = 0,006  
 Selisih perhitungan D3 = 0,632 – 0,63 = 0,002

Prosentase selisih :

$$\frac{(0 + 0,006 + 0,002) \times 100}{3} = 0,27 \%$$

Untuk mendapatkan akurasi data, maka dilakukan pengujian terhadap pakar, dalam hal ini adalah guru kelas. Proses ini dilakukan dengan membandingkan hasil pengujian sistem dengan wawancara langsung dengan pakar yaitu dengan kasus yang sama dengan



pengujian aplikasi. Hasil pengujian dengan pakar ditunjukkan pada Tabel 5.

**Tabel 5.** Pengujian Pakar

Siswa yang di uji	Hasil Aplikasi	Hasil Pakar	Selisih
Siswa 1	0.48	0,5	0,02
Siswa 2	0.57	0,6	0,03
Siswa 3	0.63	0,7	0,07
<b>Rata-rata</b>			0,04

Berdasarkan pengujian terhadap tiga siswa dengan masing-masing lembar jawaban siswa, maka didapatkan hasil error sebesar 4%.

#### 4. KESIMPULAN

Setelah dilakukan pengujian pada aplikasi penentuan tingkat kemiripan jawaban menggunakan Metode Cosine Similarity dengan menggunakan 3 buah data uji, maka penulis mendapatkan kesimpulan bahwa selisih error antara hasil aplikasi dan hasil perhitungan manual sebesar 0,27 %. Aplikasi dapat berjalan dengan baik pada dua browser yang di ujikan yaitu Google Chrome dan Mozilla Firefox.

#### PUSTAKA

- [1] R. Fitri and A. N. Asyikin, "APLIKASI PENILAIAN UJIAN ESSAY OTOMATIS MENGGUNAKAN," Jurnal POROS TEKNIK, vol. VII, pp. 88-94, 2015.
- [2] B. R. Aditya, "Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining," Jurnal Infotel Vol. 7 No. 2 November 2015, vol. VII, pp. 95-96, 2015.
- [3] F. Nurhuda, S. W. Sihwi and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," JURNAL ITSMART ISSN : 2301–7201, vol. II, pp. 35- 42, 2013.
- [4] A. Kurniawan, F. Solihin and F. Hastarita, "PERANCANGAN DAN PEMBUATAN APLIKASI Pencarian Informasi BEASISWA DENGAN MENGGUNAKAN COSINE SIMILARITY," Jurnal Simantec, vol. IV, pp. 115-124, 2014.
- [5] o. V. Indriyono, E. Utami and A. Sunyoto, "Pemanfaatan Algoritma Porter Stemmer Untuk Bahasa Indonesia Dalam Proses Klasifikasi Jenis Buku," Jurnal Buana Informatika, vol. VI, pp. 301-309, 2015.
- [6] D. A. R. Ariantini, A. S. M. Lumenta and A. Jacobus, "PENGUKURAN KEMIRIPAN DOKUMEN TEKS BAHASA INDONESIA MENGGUNAKAN METODE COSINE SIMILARITY," E-Journal Teknik Informatika Volume 9, No 1 (2016), ISSN : 2301-8364, vol. IX, pp. 1-8, 2016.
- [7] W.Rizki Tri, P.Dhidik, and S.Eko,"Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi", Jurnal Teknik Elektro Vol. 9 No. 1,pp.18-23, 2017.