

KLASIFIKASI EMOSI PADA LIRIK LAGU MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

EMOTION CLASSIFICATION IN SONG LYRICS USING NAÏVE BAYES CLASSIFIER METHOD

¹Maftukhah Carolin Noviardini, ²Andrew Briand Osmond, ³Casi Setianingsih

¹²³Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹maftukhahc@gmail.com@gmail.com, ²abosmond@telkomuniversity.ac.id,

³casie.sn@telkomuniversity.ac.id

Abstrak

Lagu merupakan hiburan dalam aktivitas manusia yang melibatkan suara-suara yang teratur. Lagu berupa sekumpulan nada-nada yang dirangkai menjadi sebuah bunyi yang sangat indah dan harmoni. Emosi pada lagu menjelaskan makna emosional yang melekat pada sebuah klip lagu.

Dalam Tugas Akhir ini akan dilakukan klasifikasi emosi berdasarkan lirik lagu, sebagai media yang digunakan untuk mengklasifikasi ekspresi dan emosi seseorang, kemudian untuk dapat mengklasifikasi emosi berdasarkan lirik lagu sesuai dengan yang pendengar inginkan, dibutuhkan metode yang tepat dan metode yang digunakan oleh penulis adalah *Naïve Bayes Classifier*, sebagai metode yang dapat melakukan klasifikasi emosi lirik lagu.

Naive Bayes Classifier merupakan salah satu metode *Machine Learning* yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh *Naive Bayes Classifier* adalah Teorema Bayes, yaitu teorema yang digunakan dalam statistika untuk menghitung suatu peluang.

Kata Kunci : Emosi, *Text Mining*, *Naïve Bayes Classifier*.

Abstract

Song is an entertainment in human activity that involves regular voices. Song is a collection of tunes that are strung together into a very beautiful sound and harmony. The emotion in the song can explain the emotional significance that is attached to a song clip.

In this Final Project will be carried out the classification of emotions based on song lyrics, as a medium used to classify one's expression and emotions. Then to be able to classify emotions based on song lyrics according to what the listener wants, the right method is needed and the method used by the author is the Naïve Bayes Classifier, as a method that can classify the emotion of the song lyrics.

Naive Bayes is one of a Machine Learning method that uses probability calculations. The basic concept used that is in Naïve Bayes Classifier is Bayes's Theorem, a theorem is used in statistics to calculate a probability.

Keyword : Emotion, *Text Mining*, *Naïve Bayes Classifier*.

1. Pendahuluan

1.1 Latar Belakang

Lagu merupakan hiburan dan aktivitas manusia yang melibatkan suara-suara yang teratur. Karena lagu berupa sekumpulan nada-nada yang dirangkai menjadi sebuah bunyi yang sangat indah dan harmoni [8]. Emosi lagu menjelaskan makna emosional yang melekat pada sebuah klip lagu.

Pada jaman global yang semakin berkembang ini banyak sekali jenis lagu dengan berbagai lirik yang ada dan setiap lirik dalam lagu itu sendiri memiliki kandungan emosi tersirat dari sang penulis lagu.

Karena itu, klasifikasi lagu berdasarkan keterkaitan antara lagu dengan emosi telah banyak digunakan diberbagai kesempatan oleh manusia yang dilakukan secara manual. Misalnya, pada film, lagu digunakan untuk mempertegas suasana pada *scene-scene* tertentu: lagu yang dramatis digunakan untuk melatarbelakangi *scene* yang menegangkan, lagu yang bersemangat untuk *scene* perang, lagu yang menyenangkan digunakan sebagai latar belakang *scene* humor [8].

Semakin banyak lirik lagu tidak mungkin dilakukan pengklasifikasian secara manual. Dengan melihat permasalahan diatas diperlukan suatu pengkategorian otomatis, yaitu dengan menggunakan *text mining* [8]. Berdasarkan latar belakang yang telah dipaparkan maka penulis berusaha mengimplementasikan metode *Naïve Bayes Classifier* dan algoritma pembobotan TF-IDF pada kasus pengklasifikasian emosi pada lirik lagu. Dikarenakan selain *Naïve Bayes Classifier* merupakan salah satu algoritma pengklasifikasian, metode *Naïve Bayes Classifier* juga memiliki akurasi yang tinggi. Penulis berharap Metode *Naïve Bayes Classifier* ini bisa memberikan akurasi yang tinggi dalam pengklasifikasian emosi pada lirik lagu.

Sistem ini dibuat agar nantinya dapat membantu pendengar musik untuk dapat memilih lagu yang pendengar inginkan sesuai dengan suasana hati dari pendengar itu sendiri tanpa harus memilih lagu secara manual.

1.2 Tujuan

Tujuan pada tugas akhir ini adalah merancang sistem yang dapat mengklasifikasikan teks pada lirik lagu berbahasa Indonesia kedalam 4 kelas emosi yaitu marah, sedih, cinta, dan gembira dengan menggunakan algoritma *Naïve Bayes Classifier*. Sistem nantinya dapat membantu pendengar musik untuk dapat memilih lagu yang mereka ingin dengar sesuai dengan suasana hati dari pendengar itu sendiri.

1.3 Rumusan Masalah

Adapun identifikasi masalah berdasarkan latar belakang di atas adalah:

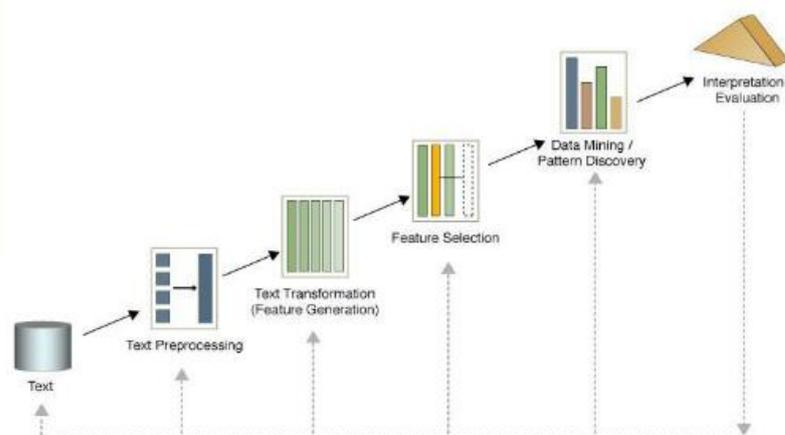
1. Bagaimana mengklasifikasikan emosi dalam sebuah lagu dari lirik lagu yang berbasis teks.
2. Bagaimana menentukan klasifikasi emosi teks berbahasa Indonesia pada lirik lagu menjadi 4 kelas emosi yaitu marah, sedih, cinta, dan gembira menggunakan metode klasifikasi *Naïve Bayes Classifier*.
3. Bagaimana mengklasifikasikan lagu agar pendengar bisa memilih lagu sesuai dengan suasana hati mereka tanpa harus memilih secara manual.

2. Dasar Teori

Pada bagian ini akan dibahas mengenai landasan teori yang mendukung antara lain:

2.1 Text Mining

"Text mining merupakan area baru dan menarik dari penelitian ilmu komputer yang mencoba menyelesaikan krisis informasi yang berlebihan dengan menggabungkan teknik dari data mining, pembelajaran mesin, pengalihan bahasa alami, pencarian informasi, dan manajemen pengetahuan"[9]. *Text mining* adalah bagian dari data mining dimana *Text mining* mencoba menggali pengetahuan yang terdapat pada sekumpulan teks atau sering disebut dokumen yang susunannya tidak terstruktur. Adapun tahapan utama yang dipakai pada *text mining* [1]:



Gambar 2.1 Tahap proses *text mining* [1]

Seperti pada gambar diatas terdapat 4 tahapan pemrosesan pada *text mining* :

1. *Text Preprocessing*, pada tahap awal ini dilakukan pemrosesan terhadap data teks untuk diubah kedalam bentuk token atau kata yang sering disebut dengan tokenisasi. Data dalam bentuk dokumen, paragraf ataupun kalimat dipecah menjadi token-token sebagai masukan untuk proses selanjutnya yaitu *Text Transformation*. Dalam tokenisasi ini, dilakukan seleksi pada tanda baca, kapitalisasi, karakter spesial dan lain sebagainya sehingga hasil yang didapatkan berupa kumpulan token [1].
2. *Text Transformation*, pada tahap ini proses yang dilakukan adalah melakukan *stemming* dan *stopword*. *Stemming* melakukan pemotongan atau penghilangan imbuhan pada token sehingga hasil yang didapat merupakan kata dasar dari kata tersebut, contoh pada kata “dipukul” dan “memukul” memiliki kata dasar “pukul”. *Stemming* ini dilakukan untuk menurunkan jumlah daftar kata pada indeks. Selain itu juga terdapat proses *stopword*, dimana dilakukan penghilangan pada kata sambung yang sangat umum berada pada dokumen, contoh kata “yang” dan “pada”. Kata sambung tersebut dihilangkan karena terlalu bersifat umum sehingga tidak dapat mencirikan sebuah dokumen tertentu [1].
3. *Feature Selection*, melakukan tahap seleksi fitur dengan menghapus kata-kata yang yang dianggap tidak menggambarkan isi dari dokumen yang dapat dilakukan dengan melihat frekuensi kemunculan kata tersebut [10].
4. *Data Mining / Pattern Discovery*, merupakan tahap terpenting dalam text mining dimana dilakukan penerapan metode data mining dengan tujuan untuk menemukan pengetahuan yang tersimpan dalam teks [10].

2.2 Preprocessing

Preprocessing adalah proses pengubahan bentuk data yang belum terstruktur menjadi data yang terstruktur. *Preprocessing* diperlukan untuk memilih kata dalam kalimat yang akan digunakan sebagai *keyword* atau indeks. Indeks ini akan digunakan sebagai kata yang mewakili suatu dokumen untuk membuat pemodelan teks *mining*. Adapun lima tahapan *preprocessing*, yaitu [12]:

1. *Case Folding*
Tidak semua dokumen teks hanya menggunakan huruf kapital. Oleh karena itu, *Case folding* digunakan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diambil, karakter selain huruf dihilangkan.
2. *Stemming*
Stemming merupakan suatu proses mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya dengan aturan-aturan tertentu, atau dapat dianggap sebagai pembuangan imbuhan pada kata.
3. *Stopword Removal*
Tahapan ini adalah proses pengambilan kata-kata penting dan pembuangan kata-kata yang dianggap tidak penting. Proses pembuangan kata tidak penting dapat diterapkan pada *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). Contoh *stopword* pada kamus adalah “ke”, “di”, “dari”, dan sebagainya.
4. *Tokenizing*
Tokenizing adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Karakter spasi digunakan sebagai pemisah antar kata.

2.3. POS-Tagging (*Part-of-Speech Tagging*)

Part-of-Speech (POS) Tagging yang juga disebut sebagai pelabelan kelas kata adalah suatu proses yang memberikan label kelas kata secara otomatis pada suatu kata dalam kalimat [13]. Proses yang dilakukan berupa melabelkan sebuah kata apakah itu kata kerja, kata sifat, atau hanya keterangan.

Tabel 2. 1 Keterangan Label POS-Tagging

Label	Keterangan
CC (coordinate conjunction)	Konjungsi koordinasi
CD (cardinal numerals)	Bilangan kardinal
OD (ordinal number)	Bilangan urutan
FW (foreign words)	Kata serapan/kata asing
IN (prepositions)	Preposisi
JJ (adjectives)	Kata sifat
MD (modal or auxiliaries verbs)	Kata kerja bantu/modal
NEG (negations)	Negasi
NN (common nouns)	Kata benda umum, tidak spesifik
NNP (proper nouns)	Kata benda dalam huruf kapital
PR (common pronouns)	Pengolahan kata benda secara umum
PRP (personal pronouns)	Kata ganti orang
RB (adverbs)	Keterangan waktu
SC (subordinate conjunction)	Kata sambung/penghubung
SYM (symbols)	Simbol
VB (verbs)	Kata kerja

2.3 Emosi

“Emosi adalah pengalaman afektif yang disertai penyesuaian dari dalam diri individu tentang keadaan mental dan fisik dan berwujud suatu tingkah laku yang tampak” [2]. Beberapa tokoh mengemukakan mengenai macam macam emosi, salah satunya adalah Descrates. Menurut Descrates, emosi terbagi menjadi 5, antara lain: [3]

1. *Desire* = hasrat
2. *Hate* = benci
3. *Sorrow* = sedih
4. *Wonder* = heran
5. *Love* = cinta
6. *Joy* = gembira

Dan menurut JB Watson, emosi terbagi menjadi tiga yaitu [3]:

1. *Fear* = ketakutan
2. *Rage* = kemarahan
3. *Love* = cinta

2.4 Pembobotan Kata

Pembobotan kata dilakukan dengan perhitungan Metode TF-IDF (*Term Frequency Inverse Document Frequency*). TF-IDF merupakan suatu metode yang berguna untuk memberikan bobot terhadap hubungan suatu kata pada suatu dokumen tertentu. Metode ini menggabungkan dua jenis perhitungan bobot term pada dokumen yaitu TF (*Term Frequency*) dan IDF (*inverse Document frequency*). TF merupakan besarnya frekuensi kemunculan suatu kata dalam dokumen tertentu yang dapat menunjukkan seberapa penting kata

dalam dokumen. Sedangkan IDF menunjukkan tingkat kepentingan suatu kata dalam suatu kumpulan dokumen, semakin besar nilai TF maka akan semakin kecil nilai IDF dari suatu kata [1]. Berikut persamaan umum yang digunakan untuk menghitung pembobotan kata.

$$W_{dt} = tf_{dt} * IDF_t \quad (2.1)$$

2.5 Naïve Bayes

Naive Bayes adalah sebuah *classifier* probabilistik berdasarkan *Bayes Rule of conditional probability*. Naive Bayes menggunakan probabilitas untuk mengklasifikasikan *instance* baru [4]. Cara kerja *Naive Bayes* sendiri adalah dengan mencari angka peluang terbesar dari kemungkinan klasifikasi, dengan melihat frekuensi tiap klasifikasi pada data training [5].

Klasifikasi *Bayesian* didasarkan pada *teorema Bayes*, yang diambil dari nama ahli matematika sekaligus menteri Prebysterian Inggris yang bernama Thomas Bayes (1702-1761) [4]. Terdapat dua peluang dalam teorema ini, yaitu *Posterior* dan *Prior* [6].

Naïve Bayes Classifier mengasumsikan bahwa keberadaan sebuah fitur dalam sebuah kelas tidak ada kaitannya dengan keberadaan fitur lainnya. Misalkan sesuatu dapat dianggap buah apel jika berwarna kuning, bulat dan berdiameter sekitar 3 inchi. Walaupun fitur ini berkaitan antara satu dengan yang lainnya, *Naïve Bayes Classifier* tetap menganggap fitur tersebut *independent* [11].

Adapun perhitungan dalam proses klasifikasi *naïve bayes* adalah menggunakan tahap berikut.

1. Menyediakan data
2. Menghitung *prior* dengan rumus berikut:

$$P(C) = \frac{N_c}{N} \quad (2.2)$$

$P(c)$: Probabilitas kemunculan kelas

N_c : Jumlah dokumen kelas c

N : Jumlah dokumen

3. Menghitung *Conditional Probabilitas* dengan rumus berikut.

$$P(w|c) = \frac{\text{count}(w,c) + 1}{\text{count}(c) + |V|} \quad (2.3)$$

$\text{Count}(w,c)$: Banyaknya jumlah kata w pada klasifikasi c

$\text{Count}(c)$: Jumlah keseluruhan kata yang ada pada klasifikasi c

$|V|$: Jumlah keseluruhan kata pada data training

4. Untuk mengambil kesimpulan dengan pemilihan kelas dihitung dengan rumus berikut.

$$\text{CMAP} = \text{argmax} P(x_1, x_2, \dots, x_n) P(C) \quad (2.4)$$

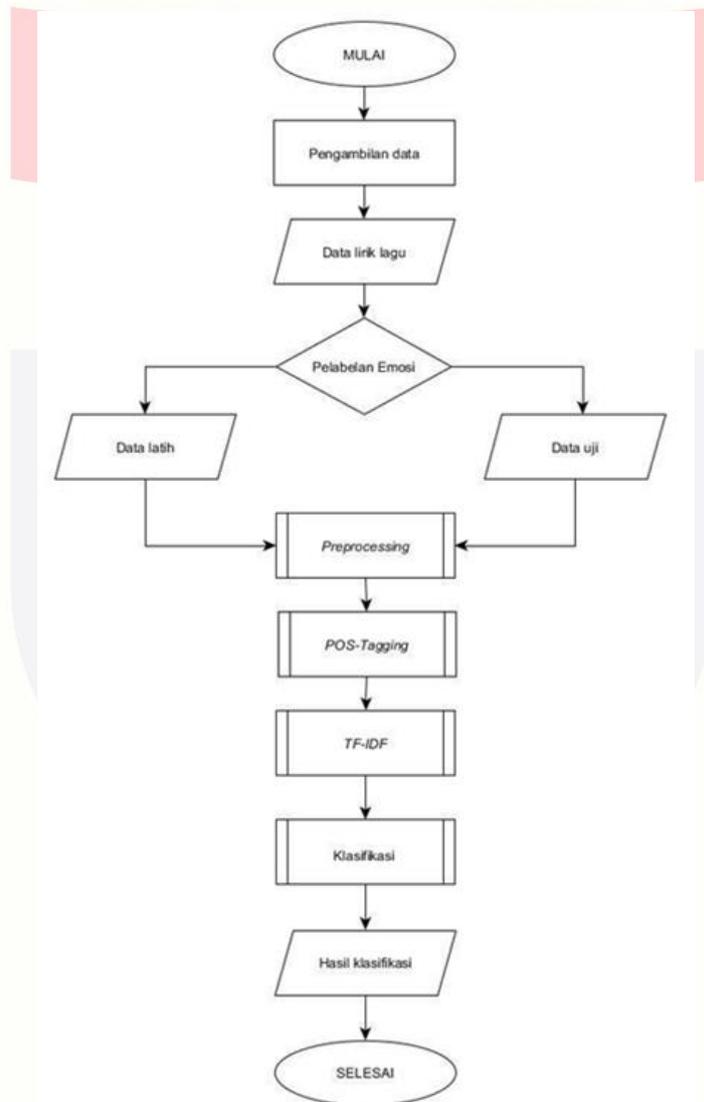
3. Perancangan

3.1. Gambaran Umum Sistem

Sistem yang akan dibangun pada tugas akhir ini adalah sistem yang mampu mendeteksi kelas emosi pada lirik lagu berbahasa Indonesia. kelas emosi yang disediakan ada 4 yaitu kelas emosi cinta, marah, senang, dan sedih. Data lirik lagu yang telah dikumpulkan nantinya bisa dijadikan dataset yang kemudian akan diproses dengan beberapa tahap, yaitu *preprocessing*, *POS-Tagging*, pembobotan kata, dan kemudian diklasifikasi dengan menggunakan algoritma *Naïve Bayes Classifier*.

3.2. Perancangan Sistem

Pada tugas akhir ini akan dibuat sebuah sistem dengan menggunakan Bahasa pemrograman *python*, yang nantinya sistem dapat mengkalasifikasikan data berupa lirik lagu menjadi 4 kelas emosi yaitu marah, sedih, cinta dan sedih dengan menggunakan algoritma *Naïve Bayes Classifier*. Dibawah ini adalah tahap-tahap perancangan sistem pada tugas akhir ini:



Gambar 3. 1 Flowchart Perancangan Sistem

3.3 Klasifikasi dengan menggunakan *Naïve Bayes Classifier*

Pada sub bab ini akan dijelaskan proses yang dilakukan untuk mencari nilai probabilitas tertinggi dari data uji dalam kategori yang paling tepat. Melalui data uji yang sudah disusun sebelumnya

probabilitasnya akan dihitung menggunakan *Naïve Bayes Classifier* melalui beberapa tahap, yaitu sebagai berikut:

1. Input Data Uji
2. Menghitung Probabilitas *Prior*
3. Menghitung *Conditional Probabilitas*
4. Menghitung Probabilitas *Posterior* Untuk Pemilihan Kelas
5. Menampilkan Hasil Pencarian

4 Pengujian

Dari beberapa proses yang sudah dilakukan sebelumnya, selanjutnya dilakukan proses pengujian 16 data lirik lagu yang sudah dikumpulkan dan sudah diberi label emosi secara manual. Pengujian yang dilakukan adalah memasukkan seluruh data uji ke dalam sistem dan menghasilkan klasifikasi label emosi. Hal ini dilakukan bertujuan untuk menarik kesimpulan hasil akhir yang akan ditampilkan untuk mengetahui apakah program yang dirancang dapat mengklasifikasikan label emosi pada lirik lagu dengan benar.

4.1 Pengujian Kinerja Sistem

Pengujian kinerja sistem diawali dengan mencari nilai peluang terbaik dan terbesar dari penghitungan algoritma *Naïve Bayes Classifier*. Pengujian kinerja sistem selanjutnya adalah melihat keberhasilan sistem berdasarkan besar Akurasi. Adapun rumus perhitungan pada akurasi di bawah ini.

$$\text{Akurasi} = \frac{\text{Total Output Terklasifikasi benar}}{\text{Jumlah total data uji}} \times 100\% \quad (4.1)$$

Pada pengujian ini, penulis menggunakan data uji yang sama, untuk mencari hasil akurasi terbaik

5 Kesimpulan

Dari pengujian dan Analisa yang dilakukan, maka dapat disimpulkan bahwa sistem yang dirancang berhasil mengklasifikasikan emosi yang dibagi menjadi 4 kelas emosi yaitu Cinta, Marah, Sedih, Senang berdasarkan lirik lagu secara otomatis menggunakan algoritma *Naïve Bayes Classifier*, dan dari hasil pengujian ditemukan nilai akurasi sebesar 72,5%.

Daftar Pustaka:

- [1] Nilamsari (2015). "Klasifikasi Emosi pada Opini Teks Bahasa Indonesia Menggunakan *Vector Space Model*", Diakses pada tanggal 19 September 2017 dari <http://openlibrary.telkomuniversity.ac.id/home/catalog/id/96050/slug/klasifikasi-emosi-pada-opini-teks-bahasa-indonesia-menggunakan-vector-space-model.html>
- [2] M. Entin., H. R. Yuniar., M. N. Rosyid, dan M. Firodh (2011). *Machine Learning dalam Program Chatting untuk Merespon Emosi Teks Berbahasa Indonesia Menggunakan Text Mining dan Naïve Bayes*. Surabaya. *Electronic Engineering Polytechnic Institute of Surabaya (EEPIS)*.
- [3] Hude, M. Darwis (2006). "Emosi Penjelajahan Religio-Psikologis tentang Emosi Manusia di dalam Al-Qur'an," Jakarta: Penerbit Erlangga.
- [4] G. S. Santosh Kamar (2014), "Classification of Heart Disease Using Naive Bayes and Genetic Algorithm," *Computational Intelligence in Data Mining - Volume 2*, pp. 269-282.
- [5] R. A. Saputra (2014), "Komparasi Algoritma Klasifikasi *Data Mining* Untuk Memprediksi Penyakit Tuberculosis (TB): Studi Kasus Puskesmas Karawang Sukabumi," *Seminar Nasional Inovasi dan Tren*, pp. 1-8.
- [6] M. K. Han Jiawei (2006). *Data Mining: Concept and Techniques*, San Fransisco: Morgan Kauffmann Publishers.
- [7] W. Z. Novalina (2016). "Analisis Penggunaan Klasifikasi *Naive Bayes* dalam Sistem Prediksi Penjadwalan Pengisian Uang pada ATM (Studi Kasus: Bank ABC)" Diakses pada tanggal 6 Oktober 2017 dari <http://openlibrary.telkomuniversity.ac.id/home/catalog/id/117015/slug/analisis-penggunaan->

[klasifikasi-naive-bayes-dalam-sistem-prediksi-penjadwalan-pengisian-uang-pada-atm-studi-kasus-bank-abc-.html](#)

- [8] M. Y. Burhan, Marji, L. Muflikhah, (2012). "Pengelompokan Emosi Berdasarkan Lirik Lagu Menggunakan Algoritma *Iterative Dichotomizer Tree*". Program Teknologi Informasi dan Ilmu Komputer Jurusan Ilmu Komputer, Program Studi Teknik Informatika, Universitas Brawijaya.
- [9] F. Ronen dan Sanger (2007). James. *The Text Mining Handbook*. New York: Cambridge University Press.
- [10] I. Amalia., S. Budi dan C. A. Rachmat (2008). *Sistem Klasifikasi dan Pencarian Jurnal dengan Menggunakan Metode Naïve Bayes dan Vektor Space Model*. Yogyakarta. Universitas Kristen Duta Wacana.
- [11] D. H. Wahid dan A. SN, "Peringkasan *Sentimental Esktraktif* di Twitter Menggunakan *Hybrid TF-IDF dan Consine Similiarity*," *IJCCS*, vol. 10, 2, pp. 207-218, 2016.
- [12] Y. H. Romario, R. M. P. Rekyan dan C. W. Randy, "Implementasi Metode *K-Nearest Neighbour* Dengan Pembobotan TF.IDF.ICF Untuk Kategorisasi Ide Kreatif Pada Perusahaan," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 4, no. 2, pp. 97-103, 2017.
- [13] D. Arawinda, R. Fam, L. Andry dan M. Ruli, "*Designing an Indonesian Part of speech Tagset and Manually Tagged Indonesian Corpus*".