

Implementasi Algoritma Naïve Bayes

by Reonal Regen

Submission date: 14-Mar-2019 10:26AM (UTC+0700)

Submission ID: 1093024905

File name: rmatika_Implementasi_Algoritma_Na_ve_Bayes_Menggunakan_ISEAR.pdf (299.43K)

Word count: 3106

Character count: 18665

IMPLEMENTASI ALGORITMA NAÏVE BAYES MENGGUNAKAN ISEAR UNTUK KLASIFIKASI EMOSI LIRIK LAGU BERBAHASA INGGRIS

Laksmita Widya Astuti^{1*}, Antonius Rachmat C.¹, Yuan Lukito¹

Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana

Jl. Dr. Wahid Sudiro Husodo No. 5-25, Yogyakarta 55224

E-mail: laksmita.widya@ti.ukdw.ac.id, anton@ti.ukdw.ac.id, yuanlukito@ti.ukdw.ac.id

Korespondensi penulis

Abstrak: Lirik lagu merupakan suatu ungkapan perasaan seseorang terhadap sesuatu hal yang sudah dilihat, didengar maupun dialaminya sehingga tidak jarang emosi menjadi salah satu kriteria *user* dalam melakukan pencarian. Pencarian lirik melekat pada kategori yang tidak hanya terbatas berdasarkan *genre* atau judul lagu, namun juga melalui lirik lagu yang diungkapkan. Agar dapat mencapai tujuan tersebut, diperlukan suatu sistem pengkategorian yang mengenali lirik lagu secara otomatis dengan salah satu metode klasifikasi yaitu Naive Bayes. Faktor yang mendorong tingginya tingkat akurasi bukan hanya terletak pada metode klasifikasi saja, namun proses sebelum menuju tahap klasifikasi juga berpengaruh pada hasil yang didapatkan. Maka dari itu, penulis melakukan penelitian melalui beberapa tahap yaitu *preprocessing* berupa *tokenisasi*, *stopword* dan *stemming*, kemudian *feature selection* yang digunakan adalah TF-IDF dengan bantuan ISEAR karena mengandung 7 emosi dasar. Dari ketujuh emosi dasar tersebut, tiga diantaranya merupakan emosi yang akan digunakan dalam penelitian ini yaitu *anger*, *sadness* dan *joy*. Hasil dari penelitian ini menunjukkan dengan menggunakan ISEAR akurasi tertinggi terdapat pada *feature set* 60% dan 100% yaitu sebesar 82,2%. Perbedaan signifikan dihasilkan pada penggunaan ISEAR dengan akurasi rata-rata keseluruhan porsi *feature set* sebesar 77% sedangkan tanpa menggunakan ISEAR rata-rata akurasi sebesar 53%. Dokumen paling relevan untuk pengujian menggunakan ISEAR terdapat pada kategori *angry* dengan rata-rata *f-measure* sebesar 0.7267.

Kata kunci: ISEAR, tf-idf, Naive Bayes, Emosi, Lirik Lagu

Abstract: Song lyric is an expression of one's feelings about something that has been seen, heard and experienced; therefore, emotion is not uncommon things for the user to be used as one of the criteria in song discovery. Searching for lyrics is close to a category that is not only limited to genres or song titles, but it can also be done through the emotions disclosed by the lyrics. In order to achieve these objectives, we build song lyrics classification based on Naive Bayes classification method. The factors that promote high levels of accuracy do not only lie in the method of classification; as a matter of fact, the pre-proceeding stage for classification process also affects the results obtained. Therefore, we conducted research through several stages; pre-processing such as tokenization, stop word processing and stemming. The feature selection used is TF-IDF with the help of ISEAR because it contains seven basic emotions. Out of the seven basic emotions, three of them that denote emotions used in this research are anger, sadness and joy. The results of this research demonstrate that by using ISEAR, the highest accuracy occurs at 60% and 100% of the feature set is by 82.2%. The significant difference resulting in the use of ISEAR with the overall average accuracy of feature set portion is 77%; whereas, without using ISEAR, the average accuracy is around 53%. The most relevant document for testing using ISEAR falls under the emotion category of anger, with the average *F-measure* of 0.7267.

Keywords: ISEAR, tf-idf, Naive Bayes, Mood, Song Lyrics

PENDAHULUAN

Musik memiliki peran yang penting bagi kehidupan manusia. Selain sebagai media hiburan, musik juga merupakan media bantu diri untuk menangani perasaan emosi yang sedang di hadapi (Lerik & Prawitasari, 2005). Banyak orang salah mengartikan bahwa lagu-lagu yang memiliki irama bersemangat seperti *rock*, *dangdut*, dan *country* menggambarkan lagu yang ceria dan penuh kegembiraan, sedangkan lagu *mellow* seperti *jazz*, *pop*, dan *ballad* menggambarkan makna lagu yang sedih karena memiliki irama yang lambat. Setiap bait pada

lirik lagu umumnya mengandung maksud melalui gaya bahasa yang digunakan oleh pengarang. Pengarang bebas menyampaikan emosi dengan *genre* yang bervariasi sesuai dengan keinginannya. Pengungkapan emosi tersebut tidak jarang menjadi salah satu kebutuhan umum para penikmat musik sehingga pemilihan kategori dalam pencarian lagu tidak hanya terbatas berdasarkan *genre*, namun juga melalui emosi dari para pencipta lagu.

Pemilihan lirik lagu terutama saat *user* melakukan sebuah pencarian melekat dengan sebuah kategori agar lebih mudah dan terstruktur. Agar dapat mencapai tujuan tersebut, diperlukan mesin pengkategorian

agar lirik lagu dapat diklasifikasikan secara otomatis. Mesin kategori lirik lagu berdasarkan emosinya dapat dibuat dengan menggunakan banyak algoritma, salah satunya adalah *Naive Bayes*. Algoritma ini dipilih karena memiliki kinerja yang cukup tinggi untuk klasifikasi dokumen teks (Hamzah, 2012).

Tidak cukup menggunakan algoritma klasifikasi supaya lirik lagu dikategorikan tepat sesuai emosinya, namun penggunaan *feature set* ISEAR juga akan diterapkan di dalam penelitian ini. *Feature set* ISEAR dapat diunduh dari <http://www.affective-sciences.org/researchmaterial>. Tujuan dari penggunaan *feature set* ISEAR ini di maksudkan agar hasil klasifikasi yang dihasilkan semakin baik. Pengukuran akurasi dilakukan dengan membandingkan perhitungan akurasi sistem, *precision*, *recall* dan *f-measure* baik menggunakan *feature set* ISEAR maupun tidak menggunakan data ISEAR. Porsi *feature set* yang digunakan beranekaragam yaitu 10%, 30%, 60% dan 100% dari total *term* terbaik yang sudah melalui proses *feature selection*.

METODE PENELITIAN

Langkah Penelitian

Penelitian ini dilakukan dengan melakukan 2 proses utama yaitu pengumpulan data dan proses kinerja sistem. Pengumpulan data dilakukan melalui 2 cara yaitu labeling dan pengambilan sampel lirik lagu. Proses klasifikasi dilakukan dengan menggunakan 3 kategori emosi yang terdapat pada ISEAR yaitu *happy*, *angry* dan *sad*. Ketiga emosi tersebut dipilih karena merupakan emosi yang paling mudah dibangkitkan oleh musik. Input dari sistem berupa dokumen lirik lagu yang kemudian dilakukan proses *preprocessing*, *feature selection*, klasifikasi dan yang terakhir adalah evaluasi. Output dari sistem berupa hasil evaluasi dan hasil klasifikasi. Proses *preprocessing* dilakukan dengan menggunakan tokenisasi, *stopword* dan *stemming*. *Feature selection* dilakukan dengan melakukan perhitungan bobot *tf-idf*. Jika terdapat pada ISEAR, maka bobot dikalikan 10 sehingga lebih dominan. Sedangkan klasifikasi dapat dilakukan dengan menggunakan metode *Naive Bayes*. Dari hasil yang didapatkan, dilanjutkan dengan menghitung evaluasi. Hasil evaluasi dapat dihitung dengan menggunakan rumus akurasi, *precision*, *recall* dan *f-measure*.

Dasar Teori

Emosi dan Musik

Emosi dapat digambarkan sebagai keadaan yang pada umumnya disebabkan oleh suatu kejadian

penting sebuah subyek. Banyak psikologi dari berbagai dunia yang menjadikan emosi sebagai objek penelitian. (Oatley & Jenkins, 1996) mendefinisikan emosi sebagai:

1. Keadaan mental sadar yang dinyatakan dengan kemampuan mengenali,
2. Kualitas perasaan dan diarah untuk beberapa subyek,
3. Gangguan jasmani pada beberapa organ tubuh,
4. Pengenalan ekspresi pada wajah, suara dan isyarat tubuh,
5. Kesiapan untuk melakukan tindakan tertentu.

Musik diyakini oleh para peneliti dapat memunculkan emosi. Koelsch (2005) dalam (Sakti, 2010), mengemukakan bahwa proses musik memunculkan emosi terkait dengan otak manusia. Dia menyimpulkan bahwa musik dapat memunculkan emosi 8 cara cukup konsisten pada semua subyek. Orang menggunakan musik untuk mengubah emosi, untuk melepaskan emosi, untuk mencocokkan emosi mereka saat ini, untuk menikmati atau menghibur diri, dan untuk mengurangi stres (Behne, 1997; Juslin dan Laukka, 2004; Sloboda dan O'Neill, 2001; Zillman dan Rubin, 1997) dalam (Sakti, 2010).

Penelitian mengenai musik dan emosi dimulai oleh Hevner (1936) dalam (Sakti, 2010). Dia melakukan penelitian mengenai musik dan emosi. Dalam eksperimennya, peneliti meminta subyek penelitian menuliskan sebuah kata sifat yang hadir dalam pikirannya saat musik dimainkan. Dari penelitian tersebut, Hevner (1936) dalam (Sakti, 2010) berpendapat bahwa musik membawa sebuah arti emosi.

Tokenisasi

Tokenisasi adalah sebuah proses yang dilakukan untuk memotong atau memecah kalimat menjadi beberapa bagian atau kata (Manning, Raghav, & Schütze, 2008). Proses tokenisasi memisahkan suatu rangkaian karakter berdasarkan karakter spasi, dan mungkin pada waktu yang bersamaan dilakukan juga proses penghapusan karakter tertentu, seperti tanda baca. Token seringkali disebut sebagai istilah (*term*) atau kata, sebagai contoh sebuah token merupakan suatu urutan karakter dari dokumen tertentu yang dikelompokkan sebagai unit semantik yang berguna untuk diproses.

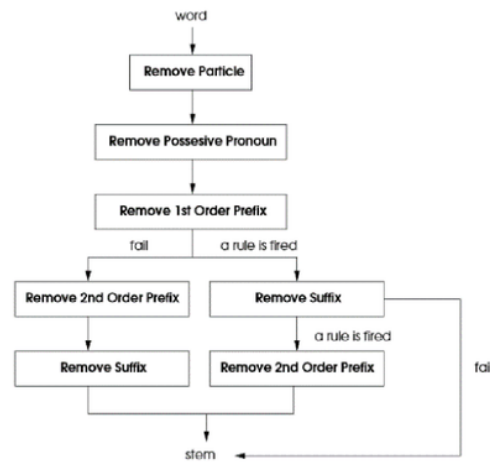
Stopword

Stopwords removal adalah sebuah proses untuk menghilangkan kata yang 'tidak relevan' pada hasil parsing sebuah dokumen teks dengan cara membandingkannya dengan *Stoplist* yang ada. Kata-kata yang terlalu sering muncul dalam dokumen-dokumen

belum tentu berguna dalam proses retrieval. Kata-kata yang tidak berguna nantinya akan dibuang dan tidak dijadikan index term.

Stemming

Stemming merupakan proses untuk mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu. Sebagai contoh pada dokumen bahasa inggris, kata dasar *stem* dapat membentuk kata "*stems*", "*stemme*", "*stemming*", "*stemmed*". Adapun langkah-langkah algoritma porter dapat dilihat pada Gambar 1.



Gambar 1. Algoritma porter (Tala, 1999)

Pembobotan TF-IDF

Tf-Idf adalah perhitungan yang menggambarkan seberapa pentingnya kata (term) dalam sebuah dokumen dan korpus. Proses ini digunakan untuk menilai bobot relevansi term dari sebuah dokumen terhadap seluruh dokumen dalam korpus. Menurut Manning, Raghavan, & Schütze (2008), rumus bobot TF-IDF di jabarkan pada poin 1 dan 2, sebagai berikut:

$$IDF(w) = \log \left(\frac{N}{DF(w)} \right) \tag{1}$$

$$TF-IDF(w) = TF(w, d) \times IDF(w) \tag{2}$$

Keterangan:

TF-IDF(w,d) = bobot suatu kata dalam keseluruhan dokumen

- W = word
- d = dokumen
- TF(w,d) = frekuensi kemunculan kata w dalam dokumen d
- IDF(w) = inverse DF dari kata w
- N = jumlah keseluruhan dokumen
- DF(w) = jumlah dokumen yang mengandung kata w

ISEAR

Isear merupakan dataset yang sudah dibelabel sesuai dengan 7 emosi dasar hasil penelitian. Berasal dari 16 negara lintas lima benua, penelitian ISEAR tidak ditujukan pada klasifikasi teks namun mencoba mencari hubungan antara emosi dan perbedaan budaya, gender, umur dan agama (Scherer & Wallbott, 1986). Dataset tujuh jenis emosi yaitu jijik, malu, marah, sedih, senang, bersalah dan takut (Destuardi & Sumpeno, 2009). Data ISEAR dapat dilihat dari situs: <http://www.affective-sciences.org/researchmaterial>.

Naive Bayes

Teorema Naive Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang untuk suatu hipotesis. Berdasarkan (Manning, Raghavan, & Schütze, 2008) persamaannya dijabarkan pada rumus 3, sebagai berikut:

$$C_{MAP} \propto \arg \max_{C_j \in C} \prod_{i=1}^n P(t_i | C_j) P(C_j) \tag{3}$$

Keterangan:

- C_j = Kategori emosi (dalam penelitian ini j = kategori emosippy, sad, dan angry).
- P(t_i|C_j) = Probabilitas t_i pada kategori C_j
- P(C_j) = Probabilitas dari C_j

Untuk P(C_j) dan P(t_i|C_j) dihitung pada saat pelatihan dimana persamaannya dapat dilihat pada rumus 4 dan 5 sebagai berikut:

$$P(C_j) = \frac{N_c}{N'} \tag{4}$$

$$P(t_i | C_j) = \frac{N_{t_i} + 1}{N_{t_i} + N_{t_i}'} \tag{5}$$

Keterangan:

- N_c = jumlah dokumen setiap kategori j
- N' = jumlah dokumen dari semua kategori
- N_t = jumlah frekuensi kemunculan setiap kata
- N_{t_c} = jumlah frekuensi kemunculan kata dari setiap kategori
- N_{t_c'} = jumlah semua kata dari semua kategori

Pengukuran Akurasi

Pengukuran akurasi yang digunakan dengan menghitung *precision*, *recall*, *f-measure* dan akurasi sistem dalam mengklasifikasikan dokumen. *Recall* adalah jumlah dokumen yang terklasifikasi dengan benar oleh sistem dibagi dengan jumlah dokumen yang seharusnya bisa dikenali sistem. *Precision* adalah jumlah jumlah dokumen yang diklasifikasikan dengan benar oleh sistem dibagi dengan jumlah keseluruhan klasifikasi yang dilakukan oleh sistem. *F-measure* merupakan nilai yang mewakili kinerja

keseluruhan sistem dan merupakan penggabungan nilai *recall* dan *precision* (Aldi, Lailil, & Indriati, 2013). Berdasarkan Manning, Raghavan, & Schütze (2008), rumus *precision*, *recall*, *F-measure*, dan akurasi sistem mengacu pada rumus 6, 7, 8 dan 9, sebagai berikut:

$$\text{Precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (6)$$

$$\text{Recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (7)$$

$$\text{F-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (8)$$

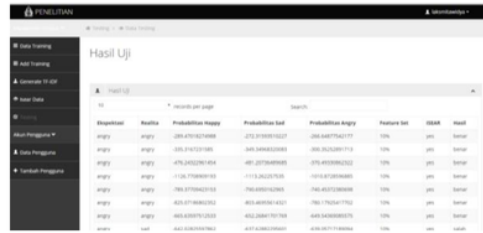
$$\text{Akurasi} = \frac{\text{jumlah dokumen benar}}{\text{jumlah seluruh dokumen uji}} \times 100\% \quad (9)$$

HASIL DAN PEMBAHASAN

Data yang digunakan untuk penelitian ini sejumlah 240 buah yang dibagi menjadi 2 bagian yaitu data uji dan data latih. Data tersebut kemudian dibagi kembali dengan porsi data latih yaitu 150 data dan sisanya adalah data uji yaitu sejumlah 90 data. Baik data latih maupun data uji masing-masing dibagi menjadi 3 bagian untuk tiap-tiap kategori emosi yaitu *sad*, *happy* dan *angry* sehingga masing-masing data latih menggunakan data sebesar 50 data dan data uji sebesar 30 data.

Setiap dokumen, baik dokumen latih, uji maupun ISEAR, melewati proses *preprocessing*, yaitu tokenisasi, *stopword* dan *stemming*. Karena data yang digunakan merupakan data berbahasa Inggris, maka proses *stemming* menggunakan algoritma Porter. Setelah *preprocessing* dilakukan, dokumen latih masih melalui banyak proses yaitu *feature selection* dan klasifikasi. *Feature selection* dilakukan dengan menghitung bobot TF-IDF pada masing-masing term pada tiap dokumen. Jika terdapat pada ISEAR, maka bobot token dikalikan dengan konstanta 10 agar lebih dominan sehingga term tersebut dapat digunakan sebagai kueri dalam mendapatkan penciri dari suatu kategori. Selanjutnya data uji yang sudah dipreprocessing dievaluasi dengan menggunakan rumus akurasi, *precision*, *recall* dan *f-measure*. Pada penelitian ini, penulis membangun sistem dengan menggunakan framework *CodeIgniter*. Tampilan dari program dapat dilihat pada Gambar 2.

Rancangan pengujian dibagi menjadi 2 bagian yaitu pengujian data baik menggunakan *feature set* ISEAR maupun tidak menggunakan *feature set* ISEAR. Keduanya menggunakan prosentase *feature set* ISEAR yaitu sebesar 10%, - 100% dengan range 10 tiap prosentase (3). Hasil pengujian dengan menggunakan ISEAR dapat dilihat pada Tabel 1.



Gambar 2. Tampilan Hasil Uji Sistem

Tabel 1. Hasil Pengujian Menggunakan ISEAR

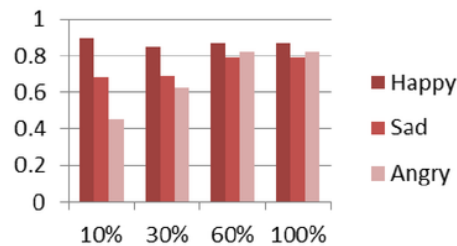
Feature Set	Precision	Recall	F-Measure	Akurasi
10%	Happy 0.8667	0.9286	0.8966	70%
	Sad 0.9333	0.5385	0.6829	
	Angry 0.3	0.9	0.45	
30%	Happy 0.7667	1	0.8680	73.86%
	Sad 1	0.5660	0.7229	
	Angry 0.4285	1	0.6	
60%	Happy 0.7667	1	0.8680	82.2%
	Sad 1	0.6521	0.7894	
	Angry 0.7	1	0.8236	
100%	Happy 0.7667	1	0.8680	82.2%
	Sad 1	0.6521	0.7894	
	Angry 0.7	1	0.8236	

Sebagaimana hasil pengujian tanpa menggunakan ISEAR dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pengujian tidak Menggunakan ISEAR

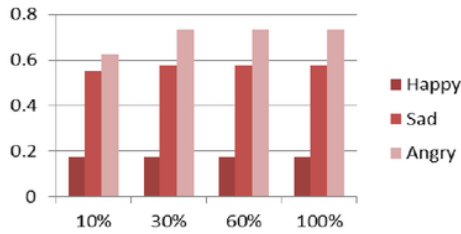
Feature Set	Precision	Recall	F-Measure	Akurasi
10%	Happy 0.8667	0.9286	0.8966	70%
	Sad 0.9333	0.5385	0.6829	
	Angry 0.3	0.9	0.45	
30%	Happy 0.7667	1	0.8680	73.86%
	Sad 1	0.5660	0.7229	
	Angry 0.4285	1	0.6	
60%	Happy 0.7667	1	0.8680	82.2%
	Sad 1	0.6521	0.7894	
	Angry 0.7	1	0.8236	
100%	Happy 0.7667	1	0.8680	82.2%
	Sad 1	0.6521	0.7894	
	Angry 0.7	1	0.8236	

Dengan hasil yang didapatkan seperti diatas, grafik Hasil F-Measure pengujian dengan menggunakan ISEAR dipaparkan pada Gambar 2.



Gambar 2. Grafik pengujian F-Measure Menggunakan ISEAR

Sedangkan grafik F-Measure tanpa menggunakan ISEAR dapat dilihat pada Gambar 3.



Gambar 3. Grafik pengujian F-Measure tanpa Menggunakan ISEAR

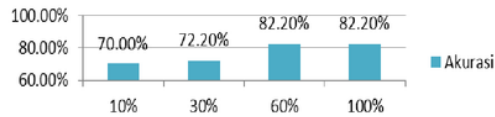
Sehingga rangkuman dari rata-rata precision, recall dan f-measure pada penelitian yang dilakukan dipaparkan pada Tabel 3.

Tabel 3. Hasil Rata-rata F-Measure, Precision, dan Recall

Kategori	Rata-rata Precision	Rata-rata Recall	Rata-rata F-Measure
Happy	0.445833333	0.860863095	0.523766873
Sad	0.933333333	0.506814099	0.653457111
Angry	0.570833333	0.90719697	0.726655534

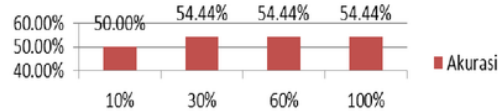
Akurasi sistem dari hasil tabel 1 dan 2 dapat dilihat pada Gambar 4 dengan menggunakan ISEAR dan Gambar 5 tanpa menggunakan ISEAR.

Akurasi



Gambar 4. Grafik Akurasi menggunakan ISEAR

Akurasi



Gambar 5. Grafik Pengujian tidak Menggunakan ISEAR

Pada sistem yang sudah dibangun, hasil dari evaluasi menunjukkan bahwa antara kategori *happy* dan *angry* cenderung dikenali sebagai kategori *sad* sehingga sering mengalami kesalahan dalam klasifikasi. Pada Tabel 3 diatas, nilai *precision* pada kategori *sad* sangat tinggi namun *recall* yang dihasilkan sangat rendah. Rata-rata *f-measure* tertinggi terdapat pada kategori *angry* karena baik *recall* maupun *precision* memiliki nilai rata-rata yang seimbang.

Dari grafik pengujian yang dihasilkan pada Gambar 4 dan 5, penggunaan *feature selection* dengan bantuan ISEAR berperan penting dalam peningkatan akurasi. Penggunaan pada sistem ini berguna untuk menunjukkan tingkat kepentingan suatu token yang menjadi penciri suatu emosi tertentu. Semakin tinggi nilai *tf-idf* dari suatu token, semakin optimum token yang dihasilkan. Penggunaan ISEAR tersebut sangat mempengaruhi hasil dari akurasi. Penggunaan ISEAR memiliki akurasi lebih tinggi dibandingkan tanpa penggunaan ISEAR.

Untuk mengetahui tingkat akurasi sistem, pemilihan porsi *feature set* yang beragam juga diperlukan untuk menguji apakah penggunaan *feature selection* dengan bantuan ISEAR berpengaruh terhadap hasil. Hal itu dilakukan dengan menguji dalam beberapa porsi *featureset* yaitu 10%, 30%, 60% maupun 100%. Proses tersebut dilakukan saat melakukan klasifikasi dokumen. Proses klasifikasi merujuk pada tabel token training yang berisi sekumpulan token yang sudah diberi bobot kemudian di-sort berdasarkan limit porsi *featureset* yang diinginkan. Dari pengujian yang dilakukan, semakin tinggi pemilihan *featureset*, hasil akurasi yang dihasilkan semakin besar.

Evaluasi dilakukan setelah dilakukan pengujian terhadap dokumen *testing*, dengan menghitung nilai peluang dengan menggunakan persamaan 3. Setelah pengujian dilakukan hasil nilai peluang dokumen banyak direpresentasikan sebagai nilai nol pada bahasa pemrograman PHP karena variable penampung pada PHP memiliki limit pada batas angka tertentu. Untuk mengatasi hal tersebut, penulis menggunakan metode Naive Bayes dengan memodifikasi logaritma pada bagian *choosing class* sehingga sangat membantu memperbaiki angka peluang menjadi lebih baik.

KESIMPULAN

Berdasarkan implementasi dan analisis sistem, maka diperoleh kesimpulan sebagai berikut:

1. Algoritma klasifikasi Naive Bayes pada penelitian ini tidak dapat menunjukkan hubungan semantik makna emosi antar kata namun ditunjukkan pada probabilitas statistik *term* antar kata.
2. Proses klasifikasi sering mengalami kesalahan klasifikasi karena 1 lirik lagu memiliki kemungkinan lebih dari 1 emosi yang berbeda.
3. Implementasi Algoritma Naive Bayes dengan menggunakan ISEAR memberikan akurasi tertinggi pada *feature set* 60% dan 100% sebesar 82,2%. Penggunaan *featureset* 60% menggunakan ISEAR sudah cukup untuk menghasilkan nilai akurasi paling maksimal.
4. Berdasarkan hasil pengujian, penggunaan ISEAR dan peningkatan porsi *featureset* menghasilkan akurasi yang cukup tinggi. Perbedaan signifikan

dihasilkan pada penggunaan ISEAR, akurasi rata-rata keseluruhan porsi *featureset* sebesar 77% sedangkan tanpa menggunakan ISEAR rata-rata akurasi sebesar 53%. Dokumen paling relevan terdapat pada kategori *angry* dengan rata-rata *precision* sebesar 0.57083, rata-rata *recall* sebesar 0.9071 dan rata-rata *F-Measure* sebesar 0.7267.

Saran yang diajukan penulis untuk perbaikan dan pengembangan sistem selanjutnya adalah sebagai berikut:

1. Memodifikasi proses klasifikasi dengan melakukan uji hipotesis dengan perhitungan statistik sehingga sistem dapat mengenali lebih dari 1 kategori emosi agar akurasi meningkat.
2. Memodifikasi algoritma *preprocessing* yang lebih baik (contoh modifikasi *stopword* tidak senang menjadi 1 kalimat) pada data latih maupun data *training*.
3. Menambah dokumen *training* yang baik sehingga menghasilkan lebih banyak pencari dokumen sesuai kategori yang diharapkan.

DAFTAR PUSTAKA

- [1]. Aldi, S.;Lailil, M.;& Indriati. (2013). Klasifikasi Berita Bahasa Inggris Menggunakan Algoritma K-Nearest Neighbour (KNN) berbasis Ontologi.
- [2]. Destuardi, I.;& Sumpeno, S. (2009). Klasifikasi Emosi Untuk Teks Bahasa Indonesia Menggunakan Metode Naive Bayes. *Seminar Nasional Pascasarjana IX*.
- [3]. Hamzah, A. (2012). Klasifikasi Teks dengan Naive Bayes (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, B-220 - B-277.
- [4]. Lerik, M. D.;& Prawitasari, J. E. (2005). Pengaruh Terapi Musik terhadap Depresi. *SOSIO-INS*, 211.
- [5]. Manning, C. D.;Raghavan, P.;& Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- [6]. Oatley, K.;& Jenkins, J. M. (1996). *Understanding Emotions*. Blackwell.
- [7]. Sakti, S. (2010). *Eksperimen Mengenai Emosi Oleh Musik Bahagia Dan Sedih Dengan Volume Tertentu*. Noudettu osoitteesta Undergraduatedesis: www.library.usd.ac.id
- [8]. Scherer, K. R.;& Wallbott, H. (1986). How universal and specific is emotional experience? . *Journal Science Information*, 763-795.
- [9]. Tala, F. Z. (1999). *A Study of Stemming Effects on Information*. Noudettu osoitteesta <https://www.illc.uva.nl/Research/Publications/Reports/MoL-2003-02.text.pdf>

Implementasi Algoritma Naïve Bayes

ORIGINALITY REPORT

39%

SIMILARITY INDEX

39%

INTERNET SOURCES

3%

PUBLICATIONS

11%

STUDENT PAPERS

PRIMARY SOURCES

1	www.coursehero.com Internet Source	17%
2	digilib.its.ac.id Internet Source	3%
3	www.scribd.com Internet Source	3%
4	filkom.ub.ac.id Internet Source	2%
5	Submitted to Universitas Brawijaya Student Paper	2%
6	riset.potensi-utama.ac.id Internet Source	1%
7	Submitted to UIN Syarif Hidayatullah Jakarta Student Paper	1%
8	repository.maranatha.edu Internet Source	1%
9	jurnal.stmik-mi.ac.id Internet Source	1%

10	repository.its.ac.id Internet Source	1%
11	Submitted to President University Student Paper	1%
12	apps.cs.ipb.ac.id Internet Source	1%
13	sinta.ukdw.ac.id Internet Source	1%
14	id.123dok.com Internet Source	1%
15	text-id.123dok.com Internet Source	1%
16	de.scribd.com Internet Source	1%
17	repository.widyatama.ac.id Internet Source	<1%
18	repository.ubaya.ac.id Internet Source	<1%
19	www.cairn.info Internet Source	<1%
20	jutei.ukdw.ac.id Internet Source	<1%
21	stop-words-list-bahasa-indonesia.blogspot.com Internet Source	<1%

22

www.unisbank.ac.id

Internet Source

<1%

23

www.politeknikmeta.ac.id

Internet Source

<1%

Exclude quotes Off

Exclude matches < 3 words

Exclude bibliography Off