

KLASIFIKASI KARAKTER BERDASARKAN PENILAIAN SIKAP SOSIAL TEMAN SEJAWAT PADA KURIKULUM 2013 MEMANFAATKAN NAIVE BAYES

Imron Rosadi¹

¹Guru Teknologi Informasi dan Komunikasi, SMA Negeri 1 Ngimbang
Jl Raya Kec. Ngimbang Kab.Lamongan
Telepon. 081331338346
E-mail : imronrsd80@gmail.com

ABSTRACT

This aim of this research is to classify an assessment sentence from students, then classified to produce an information. Classification is used to determine the character of each student so that teachers have no difficulty in assessing the social character of the student, based on asesment orientation On Kurikulum 2013. This research is empasize to processing opinion for the students to evaluate their friend in SMA Negeri 1 Ngimbang Lamongan, and the opinion to evaluate each student will classified into 6 social attitudes that is Honest, Discipline, Responsibility, Pay Attention (mutual cooperation, tolerance), well manered and self confidence, this research is divide into 2 phase that is the process to produce training data (dataset) and the process to classify opinion (test data). Both off the process are to at extracting the attributes and object components which commented in every document and to decide the classification social attitudes classfication for each student. On the system for this character clasification produce accuracy with the sucess rate from the result of the testing clasficataion use Algoritman Naive Bayes success rate of 72%, this approach paper I use of 80%, the succes rate of 72% because the use of the sentence for rating colleague had the rate of variants are quite low that affect the process off making the data that will be used to the system off the clasificaton of characterer.

Keywords: Social attitudes, Classification, Naive Bayes

1. PENDAHULUAN

Kurikulum 2013 terdapat penilaian sikap sosial untuk mengetahui karakter masing-masing siswa, penilaian sikap sosial teman sejawat ini dalam bentuk kalimat yang tidak beraturan dan guru akan melakukan klasifikasi untuk menentukan karakter masing-masing peserta didik.

Text mining adalah proses ekstraksi pola berupa informasi dan pengetahuan yang berguna dari data yang berjumlah besar bersumber dari data teks, seperti dokumen Word, PDF, kutipan teks dan lain-lain. Jenis masukan untuk *text mining* ini disebut data tak terstruktur dan merupakan pembeda utama dengan data mining yang menggunakan data terstruktur atau basis data sebagai masukan. *text mining* dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan data mining. Proses yang umum dilakukan oleh *text mining* di antaranya adalah perangkuman otomatis, kategorisasi dokumen, penggugusan teks dll. (Turban, et.al., 2011)

Klasifikasi teks adalah proses pengelompokan dokumen kedalam kelas berbeda, dalam tahapannya tiap dokumen menunjuk pada satu kelas tertentu maka dibutuhkan proses untuk menggali informasi dari dokumen tersebut. Sehingga penelitian ini dilakukan untuk dapat mengklasifikasikan kalimat

penilaian dari teman sejawat menggunakan algoritma naive bayes.

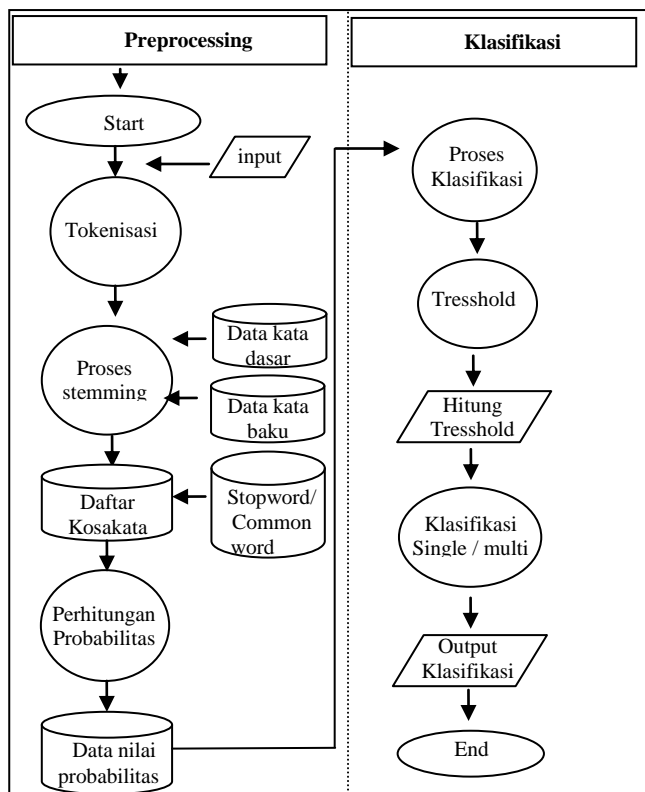
Tugas klasifikasi dokumen dapat dibagi menjadi dua macam yaitu klasifikasi dokumen terawasi di mana beberapa mekanisme eksternal (seperti feedback manusia) memberikan informasi mengenai klasifikasi yang tepat untuk dokumen, dan klasifikasi dokumen tak terawasi, dimana klasifikasi harus dilakukan sepenuhnya tanpa merujuk ke informasi eksternal

Berdasarkan latar belakang masalah tersebut, maka tujuan penelitian ini adalah untuk menguji dan menerapkan klasifikasi karakter berdasarkan penilaian teman sejawat menggunakan algoritma *Naive bayes* baik output satu kelas maupun multi kelas. Manfaat penelitian ini adalah mengembangkan software yang mampu mengklasifikasikan penilaian teman sejawat, maka proses klasifikasi karakter berdasarkan penilaian sikap sosial teman sejawat dapat dilakukan secara terkomputerisasi sebagai ganti dari proses klasifikasi manual.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian. Tahap-tahap metodologis yang diusulkan mulai bagaimana data yang berupa text disiapkan, apa saja tools pendukungnya, baik berupa perangkat keras maupun perangkat lunak dan yang paling utama adalah disini akan dibahas mengenai blok sistem yang hendak dibangun, informasi yang disajikan

meliputi gambaran kebutuhan sistem, studi literatur, dataset, gambaran arsitektur sistem.



Gambar 1. Diagram Blok

Penjelasan Diagram Alur Sistem untuk proses klasifikasi :

1. Data Penilaian diinputkan melalui sistem oleh teman sejawat. Pada proses ini, data di inputkan dari website oleh siswa untuk menilai temannya sendiri dalam 1 kelas, data diinputkan dengan memilih terlebih dahulu nama siswa yang dipilih untuk kemudian diberikan penilaian berupa kalimat penilaian
2. Tahap *Preprocessing* Data
Pada tahap ini, dilakukan proses-proses untuk mempersiapkan data. Seperti yang telah dibahas di awal, bahwa data penilaian teman sejawat ini berupa data teks yang tergolong jenis data tidak terstruktur sehingga perlu dilakukan langkah-langkah untuk mempersiapkan data tersebut sehingga siap digunakan untuk proses selanjutnya.
3. Tahap Case Folding tahap ini merupakan tahapan yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima.
4. Tahap Tokenisasi dalam tahap ini secara garis besar dilakukan dengan memecah kalimat menjadi token (kata) dengan mengabaikan karakter non alphabet.
5. Tahap Stopword Removal Tahap ini digunakan untuk menemukan kembali Informasi

(*Information Retrieval*) biasanya memerlukan kamus kata dasar (*root words*) dan *stopword list* (atau *stop list*).

6. Tahap Stemming tahap ini merupakan suatu proses yang mentransformasikan kata-kata yang terdapat dalam suatu dokumen kata dasar (*root word*) dengan menggunakan aturan-aturan tertentu. Berdasarkan pada asumsi bahwa *term-term* yang memiliki bentuk dasar (*stem*) yang sama pada umumnya memiliki makna yang mirip, proses *stemming* digunakan secara luas dalam *Information Retrieval* sebagai salah satu cara untuk meningkatkan performansi *retrieval*.
7. Pembentukan daftar kosakata dataset tahap ini setelah dilakukan proses stemming maka akan dibentuk tabel dataset
8. Pembentukan data probabilitas pada tahap ini setelah dilakukan proses pembentukan daftar kosakata maka akan dibentuk tabel data nilai probabilitas.
9. Tahap Klasifikasi Pada tahap ini setelah dilakukan proses pembentukan tabel data nilai probabilitas maka akan dilakukan proses klasifikasi yang digunakan sebagai dasar adalah daftar nilai probabilitas dan nilai treshold untuk proses klasifikasi

Kurikulum merupakan salah satu unsur yang memberikan kontribusi untuk mewujudkan proses berkembangnya kualitas potensi peserta didik tersebut. Kurikulum 2013 dikembangkan berbasis pada kompetensi sangat diperlukan sebagai instrumen untuk mengarahkan peserta didik menjadi: (1) manusia berkualitas yang mampu dan proaktif menjawab tantangan zaman yang selalu berubah; (2) manusia terdidik yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri; dan (3) warga negara yang demokratis, bertanggung jawab.

Kurikulum 2013 membagi kompetensi sikap menjadi dua, yaitu *sikap spiritual* dan *sikap sosial*. Pada jenjang SMA/MA, kompetensi sikap spiritual mengacu pada KI-1: Menghargai dan menghayati ajaran agama yang dianutnya, sedangkan kompetensi sikap sosial mengacu pada KI-2: Menghargai dan menghayati perilaku jujur, disiplin, tanggung jawab, peduli (toleransi, gotong royong), santun, percaya diri. Penilaian antar peserta didik merupakan teknik penilaian dengan cara meminta peserta didik untuk saling menilai pencapaian kompetensi. Aspek kompetensi yang dinilai adalah kompetensi inti sosial yaitu perilaku jujur, disiplin, tanggung jawab, peduli (toleransi, gotong royong), santun, dan percaya diri.

Proses klasifikasi karakter berdasarkan penilaian teman sejawat terdapat beberapa proses untuk melakukan klasifikasi. Penilaian dan data training selanjutnya akan di *preprocessing* untuk menghilangkan noise pada dokumen, metode *Preprocessing* yang umum digunakan juga

digunakan dalam penelitian ini yaitu *case folding*, *tokenize*, *stopword* dan *stemming*.

Tokenisasi adalah pemotongan kata dalam tiap kalimat. Hal ini dilakukan agar tiap kata dapat diketahui frekwensi kemunculan pada suatu kalimat atau dokumen. Sedangkan *stopword removal* adalah penghapusan *stopword* untuk membersihkan hasil tokenisasi dari *stopword*.

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa Information Retrieval (IR) dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke bentuk kata dasarnya.

Algoritma *stemming* untuk bahasa yang satu berbeda dengan algoritma *stemming* untuk bahasa lainnya.

Tahapan Dalam penelitian ini, dilakukan pembagian dataset untuk menjadi data training dan data testing. Total dataset yang digunakan dalam penelitian ini sekitar 3824 record yang terbagi dalam 6 kelas. Output klasifikasi bisa satu kelas dan multi kelas karakter. Kemudian akan disimpan dalam tabel dataset master data.

Tabel 1. Data training dan testing

No	Kelas	Total data	Data training	Data testing
1	Disiplin	851	681	170
2	Jujur	435	350	85
3	Peduli	551	441	110
4	Percaya Diri	653	523	130
5	Santun	587	472	115
6	Tanggung Jawab	443	353	90
7	Multi kelas	304	244	60
TOTAL		3824	3064	760

2.1 PREPROCESSING

Pada bagian ini akan dibahas mengenai langkah-langkah mempersiapkan data yang disebut *preprocessing*. *Preprocessing* penting dilakukan mengingat dokumen teks merupakan jenis data yang tidak terstruktur sehingga akan sulit untuk mendapatkan informasi yang berkualitas tanpa menghilangkan noise pada dokumen teks tersebut. Proses yang dibahas meliputi *case folding*, *tokenisasi*, *stopword removal (filtering)*, dan *stemming*

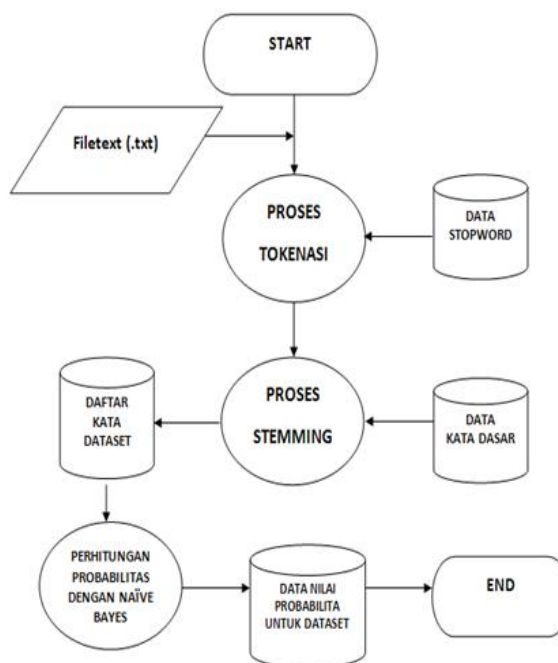
Pada tahap ini dokumen teks yang masuk akan dirubah ke bentuk huruf kecil semua, selain itu hanya huruf ‘a’ sampai ‘z’ saja yang akan digunakan. Sehingga karakter yang lain akan dihilangkan karena dianggap sebagai delimiter seperti angka dan tanda baca. Proses selanjutnya adalah tokenisasi, proses ini digunakan untuk memecah sebuah teks menjadi sekumpulan kata-kata tanpa memperhatikan keterkaitan antar kata, peran dan kedudukannya dalam kalimat. Pada data kalimat Penilaian “Dia anak yang polos, kalau berbicara tidak suka berbohong dan jujur kepada Teman

maupun Guru”. dilakukan proses tokenisasi, Tahapan berikutnya dalam *preprocessing* adalah *stemming* yaitu menentukan kembali kata-kata pada bentuk dasarnya. Seperti yang telah diuraikan di pembahasan sebelumnya

Tabel 2. Hasil proses stemming

Kata	Jumlah
Anak	1
sopan	1
Guru	1
Enak	1
bicara	1
senyum	1

2.1 KLASIFIKASI KARAKTER DENGAN NAIVE BAYES CLASIFICATION



Gambar 2. Diagram Klasifikasi

Ada beberapa tahapan yang harus dilalui untuk menghasilkan data yang digunakan sebagai dataset atau data training. Berikut ini adalah langkah-langkah untuk menghasilkan data set. Proses memasukkan kalimat-kalimat penilaian yang tadi disimulasikan menjadi nilai probabilitas untuk data set dibawah ini.

Tabel 3. Contoh nilai probabilitas

No	Kata	Jujur	Disiplin	Tjawab
1	Tanding	0.0018	0.0018	0.0020
2	Guru	6.0018	7.0018	12.002
3	Murah	0.0018	0.0018	0.0020
4	Pakai	0.0018	7.0018	0.0020
5	Anggota	0.0018	0.0018	0.0020
6	Bimbing	0.0018	0.0018	0.0020

- | | | | | |
|---|-------|--------|--------|--------|
| 7 | Pria | 0.0018 | 0.0018 | 0.0020 |
| 8 | Lewat | 0.0018 | 0.0018 | 0.0020 |
- a. Contoh kalimat penilaian yang akan digunakan menjadi data training sampel testing, “dia anak yang sopan kepada guru, enak diajak bicara dan suka senyum“
- b. Langkah selanjutnya dilakukan tokenisasi menghasilkan data seperti tabel dibawah ini

Tabel 4. Hasil Proses Tokenisasi

No	Kata
1	dia
2	anak
3	yang
4	sopan
5	kepada
6	guru
7	enak
8	diajak
9	bicara
10	dan `
11	suka
12	senyum

- c. Tahapan *preprocessing* berikutnya adalah *text transformation* yang dilakukan dengan cara menghapus kata-kata yang dianggap tidak penting atau disebut *stopword removal* sehingga kata-kata diatas berkurang untuk lebih jelasnya bisa dilihat pada tabel dibawah ini

Tabel 5. Hasil proses stopwords

Kata	Jumlah
anak	1
sopan	1
guru	1
enak	1
diajak	1
bicara	1
Senyum	1

- d. Tahapan berikutnya dalam *preprocessing* adalah *stemming* dengan menggunakan Algoritma Nazief dan Andrian's yaitu menentukan kembali kata-kata pada bentuk dasarnya. Seperti yang telah diuraikan di pembahasan sebelumnya.

Tabel 6. Hasil proses stemming

Kata	Jumlah
anak	1
sopan	1
guru	1
enak	1
Ajak	1
bicara	1
senyum	1

- e. Proses perhitungan Naive Bayes untuk klasifikasi

- f. Dilakukan perbandingan satu demi satu untuk masing-masing kategori untuk kategori yang menghasilkan nilai terbesar maka dia masuk dikelas tersebut

Pada model klasifikasi dengan menggunakan metode *Naive Bayes* ini data trainingnya satu kelas, sedangkan data testingnya memungkinkan multi kategori atau multi kelas (namun tidak bisa dipaksakan terjadi). Hal ini mungkin terjadi dengan pemahaman, bahwa telah diberikan batasan nilai tertentu pada selisih treshold nilai probabilitas kandidat kelas tertinggi dengan probabilitas kelas dibawahnya.

Penggunaan pengukuran P_{thres} digunakan untuk menunjukan rata-rata dari posterior probability dari document d_j dalam setiap kelas

$$P_{thres} = \frac{1}{n} \sum_{i=1}^n P(l_i | d_j)$$

Ketika $P(l_i | d_j) \geq P_{thres}$ dimana d_j mempunyai label l_i . Dalam strategi ini dokumen d dan mempunyai label dengan rumus $P(l/d) \geq P_{thres}$. The Eq.(1) di gambarkan menjadi

$$H(d) = \bigcup_{i \in L} \{l : P(l|d) \geq P_{thres}\}$$

Sedikitnya jumlah label dokumen adalah "1" ketika probabilitas dokumen milik kelas jelas lebih tinggi dari dokumen milik untuk kelas-kelas lain. Jumlah maksimum label dokumen adalah "n" ketika probabilitas dokumen milik masing-masing kelas adalah sama atau diatas rata-rata dari keseluruhan kelas hasil perhitungan Naive Bayes

Berikut uji coba untuk klasifikasi multi label atau multi kelas “dia anaknya suka menolong ,asik diajak berteman dan tidak suka aneh-aneh dan dia selalu bersikap jujur dalam hal mengerjakan apapun” Kemudian dilakukan preprocessing seperti diatas dan hitung nilai menggunakan naive bayes dan didapatkan hasil seperti tabel dibawah ini

Tabel 7. Hasil bayes

No	Kelas	Perhitungan Bayes
1	Jujur	5.34E-08
2	Disiplin	1.12E-11
3	TJawab	5.81E-11
4	Peduli	1.10E-07
5	Santun	8.04E-11
6	PercayaDiri	9.39E-10
Rata-rata		2.73E-08

Melihat data tabel diatas dapat di simpulkan bahwa kelas pertama dari klasifikasi kalimat tersebut masuk di kelas Peduli karena memiliki nilai paling tinggi dan kelas kedua masuk dikategori jujur dengan nilai diatas rata-rata dari keseluruhan kelas.

3. PEMBAHASAN

Pengujian Sistem dengan K Fold Form ini merupakan proses untuk mendapatkan *k-Optimal*

menggunakan metode *k-Fold Cross Validation* dengan pengulangan sebanyak 5 kali percobaan. Langkah metode *k-Fold Cross Validation* ini adalah dengan membagi data *training* yang berjumlah 2500 menjadi 5 bagian yang sama yaitu 500 buah data tiap bagian data *testing*. Data *training* ini sebelumnya telah dilakukan pengacakan data, berikut ini hasil ujiannya:

Tabel 8. Akurasi K fold

No	Uji Kfold	Stopword Fadil Z Tala		Stopword dan Common word	
		Akurasi	F1 Score	Akurasi	F1 Score
1	Ke 1	67,2	68,6	71,4	70,82
2	Ke 2	69,6	67,43	70,45	68,87
3	Ke 3	62,8	61,72	72,65	72,51
4	Ke 4	73	72,77	78,08	77,95
5	Ke 5	64,6	63,68	69,26	68,93
Rata-rata		67,4	66,84	72,37	71,82

Selanjutnya dilakukan perhitungan Precision, Recall dan F1-Score untuk output multi kelas dan satu kelas dengan perbandingan 70% data latih dan 30% data uji, setelah dilakukan uji coba sebanyak 8 kali dengan memberikan nilai treshold yang berbeda dan didapatkan hasil akurasi dan F-1 Score sebagai berikut, untuk lebih jelasnya dapat dilihat pada tabel dibawah ini

Tabel 9. Uji coba treshold multi kelas

No	Tress Hold	Dibatasi 2 kelas		Tanpa batas	
		Akurasi	F1 Score	Akurasi	F1 Score
1	+20	93,68%	80,09%	93,92	86,40
2	+10	93,68%	84,96%	98,01	86,56
3	+5	94,48%	80,18%	94,88	86,67
4	0	94,56%	79,73%	95,2	86,66
5	-5	94,96%	79,85%	95,6	86,82
6	-10	95,2%	79,11%	95,6	86,78
7	-20	94,96%	78,50%	95,6	86,58
8	-30	95,52%	78,48%	96,9	86,47

Uji Coba Penilaian teman sejawat yang oleh peserta didik SMA Negeri 1 Ngimbang, kemudian diklasifikasi oleh sistem. Kemudian dicek secara manual, kemudian dicocokkan kesesuaian klasifikasi hasil sistem dengan klasifikasi manual, untuk lebih jelasnya bisa dilihat pada table dibawah ini

Tabel 10. Akurasi

No	Uji Kfold	Akurasi	F1 Score
1	K Fold 1	91, %	90,99

Uji Coba penilaian teman sejawat untuk multi kelas sebanyak 50 kalimat dan terdapat 100 karakter kemudian diklasifikasi oleh sistem. Kemudian dicek secara manual, kemudian dicocokkan kesesuaian klasifikasi hasil sistem dengan klasifikasi manual

Tabel 11. Precision recall dan F1 score

Sub kelas	Precision	Recall	F1 Score
Jujur	0.81	0.55	0.65
Disiplin	1	0.52	0.68
TJ	0.67	0.75	0.71
Peduli	0.79	0.85	0.81
Santun	0.67	1	0.8
PD	0.75	0.46	0.57
Rata2	0.78	0.68	0.70

4. PENUTUP

Berdasarkan uji coba terhadap sistem pada penelitian ini dapat diambil beberapa kesimpulan antara lain :

- Metode *Naive Bayes Clasifier* menggunakan data inputan yang berupa kalimat penilaian sejawat yang sudah diklasifikasi oleh ahli dapat digunakan untuk model Pembelajaran yang efektif pada klasifikasi kalimat penilaian dengan akurasi tertinggi bisa mencapai 78%.
- Tidak semua data training mampu memberikan tingkat klasifikasi yang baik pada uji coba, hal ini terjadi pada kelas percaya diri, dengan akurasi hanya mencapai 69%, dikarenakan penggunaan kata-kata untuk dataset sangat mirip dengan kelas lainnya.
- Klasifikasi Multi kelas dari paper belum bisa diterapkan sepenuhnya pada penelitian ini dikarenakan perbedaan nilai probabilitas rata-rata sangat rendah sehingga perlu dilakukan perubahan nilai treshold untuk mendapatkan akurasi yang baik.
- Penulisan kalimat penilaian harus sesuai kaidah Bahasa Indonesia yang baik untuk mendapatkan akurasi klasifikasi yang baik.

Dari pembahasan pada bab-bab sebelumnya pada penelitian ini dapat diberikan saran yang dapat digunakan untuk perbaikan sistem dimasa yang akan datang ada beberapa hal yang perlu diperbaiki antara lain :

- Sistem klasifikasi karakter ini menggunakan metode klasifikasi dengan Algoritma *Naive Bayes* menggunakan teks bahasa Indonesia diharapkan pada penelitian berikutnya bisa menggunakan bahasa lain misal Bahasa Jawa atau Bahasa Inggris
- Output pada klasifikasi karakter ini hanya dibatasi 6 kelas diharapkan pada penelitian berikutnya bisa lebih dari 6 kelas.
- Pada klasifikasi karakter ini hanya dibatasi 2 kelas diharapkan pada penelitian berikutnya bisa lebih dari 2 kelas.
- Diharapkan pada penelitian berikutnya bisa dilengkapi dengan penggunaan kalimat penilaian bernilai positif atau negatif untuk masing-masing karakter

- e. Pada penelitian selanjutnya diharapkan dapat memperbaiki kesalahan penulisan kata penilaian kata dari siswa.

PUSTAKA

- Bambang Kurniawan, Syahril Effendi, Opim Salim Sitompul, "Klasifikasi konten berita dengan text mining, Jurnal Dunia Teknologi Informasi Vol. 1, No. 1, (2012) 14-19, 2012
- Destuardi, Surya Sempono, "Klasifikasi Emosi Untuk Teks Bahasa Indonesia Menggunakan Menggunakan Metode Naive Bayes, Seminar Nasional Pasca Sarjana Institut Teknologi Sepuluh Nopember, 13 Agustus 2009
- Destuardi, Suryo Sumpeno "Klasifikasi Emosi Untuk Teks Bahasa Indonesia menggunakan metode Naive Bayes" Seminar Nasional Pasca Sarjana ITS Nopember, 2016
- Gilles Brassard, Paul Bratley, "Algorithmics Theory dan Practice", Prentice Hall, 1998
- Junjie Lin dan Wenji Mao, "Personality Based Public Sentiment Classification In Microblog, State Key Laboratory of Management and Control for Complex Systems Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, 2015 IEEE
- Mohammad Nuh, "Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia, Jakarta, 2013", PP.66 tahun 2013
- Ni Made Ari Lestari, I Ketut Gede Darma Putra dan Aa Ketut Agung Cahyawan, "Personality Types Classification For Indonesian Text In Partner Searching Website Using Naive Bayes Methods, Ijcsi International Journal Of Computer Science Issues, Vol. 10, Issue 1, No 3, January 2013
- Ni Wayan Sumartini Saraswati, "Naive Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis, Seminar Nasional Sistem Informasi Indonesia, 2-14 Desember 2015
- Pengawas Dinas Pendidikan SMA, "Modul IHT Pelatihan Kurikulum 2013" Dinas Pendidikan Kabupaten Lamongan", Juli 2013
- R.H Sianipar, "Membangun Web dengan PHP dan MySql" Penerbit Informatika Bandung, Februari 2015
- Rizal Setya Perdana, Suprpto, Rekyan regasari, "Pengkategorian Pesan Singkat Berbahasa Indonesia Pada Jejaring sosial Twitter dengan Metode Klasifikasi Naive Bayes, Universitas Brawijaya Malang, 2012
- Shweta Joshi dan Bhawna Nigam, "Categorizing The Document Using Multi Class Classification In Data Mining, International Conference on Computational Intelligence and Communication Systems, 2011
- W.J.S. Poerwadarminta, "Kamus Umum Bahasa Indonesia", Pusat Bahasa Departemen Pendidikan Nasional, Oktober 2003

Zhihua Wei, Hongyun Zhang, Zhifei Zhang, Wen Li dan Duoqian Miao "A Naive Bayesian Multi-Label Classification Algorithm With Application To Visualize Text Search Results", International Journal of Advanced Intelligence Volume 3, Number 2, pp. 173-188, July, 2011.