ANALISIS SENTIMEN PENGGUNA TWITTER PADA AKUN RESMI SAMSUNG INDONESIA DENGAN MENGGUNAKAN NAÏVE BAYES

Pipit Pitria

Teknik Informatika - Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : pipitpitriaa@gmail.com

ABSTRAK

Dewasa ini mulai banyak penelitian yang dilakukan terhadap analisis sentimen untuk mengklasifikasikan sentimen kedalam beberapa kelas. Pemanfaatan analisis sentimen biasanya digunakan untuk mengevaluasi sebuah produk yang sentimen-sentimennya didapat dari *feedback* sebuah produk itu sendiri melalui sosial media khususnya twitter. Studi kasus yang digunakan sebagai uji coba implementasi analisis sentimen pada penelitian ini adalah akun twitter resmi Samsung Indonesia yang aktif dalam mempromosikan produk-produknya.

Permasalahan pada analisis sentimen adalah bagaimana mengklasifikasikan sentimen kedalam kelas-kelas yang sudah ditentukan. Selanjutnya,untuk mengatasi permasalahan tersebut dilakukan pembelajaran mesin dengan menggunakan *Naïve Bayes Classifier*. *Naïve Bayes Classifier* merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas dengan kelebihan implementasi yang *relative* mudah dan menghasilkan akurasi yang baik. Untuk itu, pada penelitian ini juga akan dilihat akurasi *Naïve Bayes Classifier* jika diterapkan untuk mengklasifikasikan sentimen kedalam sentimen positif dan negatif .
Dari hasil uji coba dengan menggunakan metode 10-*fold cross validation* dengan data sebanyak 1500 *tweet* menunjukan bahwa metode *Naïve Bayes Classifier* dapat diterapkan pada analisis sentimen untuk mengklasifikasikan sentimen kedalam sentimen positif dan negatif dengan rata-rata akurasi yang cukup tinggi mencapai 98.87%

Kata kunci: Analisis sentimen, *Naïve Bayes Classifier*, *Cross validation*, Twitter, *tweet*

1. PENDAHULUAN

Kategorisasi teks merupakan aktifitas pemberian label ke teks-teks bahasa alami dengan kategori-kategori topik dari himpunan yang telah didefinisikan sebelumnya [1]. Analisis sentimen termasuk kedalam kategorisasi teks yang membagi teks kedalam beberapa label atau kelas. Disatu sisi mulai muncul banyak penelitian dilakukan terhadap analisis sentimen untuk mengklasifikasikan sentimen ke dalam beberapa kelas yang nantinya dapat dimanfaatkan untuk mengevaluasi sebuah produk, sentimen-sentimen tersebut biasanya didapatkan dari *feedback* sebuah produk.

Sosial media untuk promosi bisnis sudah wajar dilakukan karena merupakan salah satu sarana yang sangat penting bagi perusahaan diera digital marketing saat ini [2]. Samsung sebagai salah satu perusahaan yang bergerak dibidang elektronik yang telah memanfaatkan twitter yang merupakan media sosial sebagai media promosi. Dengan adanya akun resmi Samsung Indonesia di twitter dan jumlah postingan *tweet* yang banyak setiap harinya, memungkinkan konsumen atau *followers* akun resmi Samsung Indonesia melakukan *feedback* dengan menanggapi postingan *tweet* tersebut. Disinilah ketersediaan sentimen untuk produk Samsung Indonesia sangat melimpah, sehingga dapat dimanfaatkan untuk mengevaluasi produk-produk Samsung.

Permasalahan analisis sentimen adalah bagaimana melakukan pengklasifikasian sentimen pada data twitter yang nantinya dapat dimanfaatkan untuk evaluasi sebuah produk. Dalam melakukan sebuah analisis sentimen, yang sering digunakan adalah metode *machine learning* seperti *Support Vektor Machine*, *Decision Tree*, *Naïve bayes Classifier*, dan lain-lain. Diantara beberapa metode tersebut *Naïve Bayes Classifier* adalah salah satu metode yang menghasilkan akurasi yang baik [4]. *Naïve Bayes Classifier* merupakan salah satu metode yang menggunakan perhitungan probabilitas. Kelebihan *Naïve Bayes Classifier* selain menghasilkan akurasi yang baik adalah implementasinya yang *relative* lebih mudah dan hanya memerlukan sejumlah kecil data training untuk memperkirakan parameter (*variable mean* dan *varian*) yang diperlukan untuk klasifikasi [5].

2. ANALISIS SENTIMEN PENGGUNA TWITTER PADA AKUN RESMI SAMSUNG INDONESIA DENGAN MENGGUNAKAN NAÏVE BAYES

Samsung merupakan sebuah produk yang menggunakan Twitter sebagai salah satu sosial media untuk sarana promosi bisnis. Dengan adanya akun resmi Samsung Indonesia dan jumlah postingan yang banyak setiap harinya memungkinkan terdapat *feedback* dari costumer atau *follower* Samsung Indonesia sebagai ketersediaan senitmen yang dapat dimanfaatkan untuk evaluasi produk Samsung.

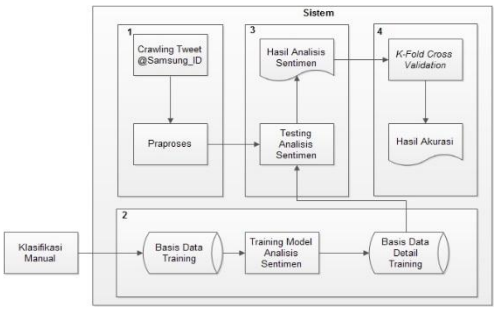
Namun permasalahan pada analisis sentimen adalah bagaimana melakukan pengklasifikasian sentimen pada data twitter yang naninya dapat dimanfaatkan untuk evaluasi sebuah produk dengan menggunakan metode *Naïve Bayes*.

2.1 Analisis Sistem

Dalam mengimplementasikan metode *Naïve Bayes Classifier* akan dibangun sebuah prototype dengan gambaran sistem yang akan tertera pada Gambar 1.

Sistem yang akan dibangun memiliki 4 tahap yaitu praproses, training data, testing, dan *cross validation*. Untuk penjelasan setiap tahap dapat dilihat sebagai berikut:

1. Tahap pertama adalah tahap mengambil data twitter dengan crawling menggunakan stream API, kemudian data tersebut dilakukan proses *preprocessing* yang merupakan tahap dimana data yang akan digunakan sebagai data testing dibersihkan dari *noise* atau dari hal yang tidak mempunyai pengaruh dalam sentimen. *Preprocessing* yang dilakukan dalam tahap ini antara lain: tokenisasi, normalisasi fitur, *case folding*, hapus *stopword*,*Convert Emoticon*, *Convert* negasi.
2. Tahap yang kedua adalah tahap training data, pada tahap ini data training yang sudah dilakukan tahap *preprocessing* dan telah diklasifikasikan secara manual kedalam 2 kelas yaitu kelas sentimen positif, dan negatif disimpan kedalam *database* data training. Setelah itu *tweet* yang sudah diklasifikasikan secara manual dibuat model probabilistiknya yang nantinya akan disimpan kembali ke dalam database detail data training yang akan digunakan pada tahap testing sebagai model analisis sentimen.
3. Tahap yang ketiga yaitu tahap testing, pada tahap ini menggunakan data testing sebagai data masukan yang tentu saja sudah melalui tahap *preprocessing*. Data testing diproses sehingga dapat dilihat kelas sentimennya.
4. Tahap keempat yaitu tahap pengujian menggunakan metode *k-fold cross validation* dimana data dibagi sejumlah n-*fold* yang diinginkan,kemudian proses testing dan training dilakukan sebanyak n kali . Sampai hasil akurasi keluar, hasil akurasi tersebut didapat dari rata-rata akurasi pada setiap iterasi.



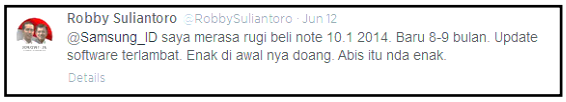
Gambar 1. Gambaran Sistem

2.2 Analisis Data Masukan

Data masukan yang digunakan adalah data *tweet* dari akun twitter resmi Samsung Indonesia. Data *tweet* tersebut didapat dengan memanfaatkan fitur *API* (*Application Interface*) yang telah disediakan oleh twitter untuk mendapakan *API key*, dan *access token* yang nantinya digunakan untuk pengambilan data *tweet.* Data yang diambil adalah data *tweet* yang mengandung kata "Samsung_ID" atau data *tweet* yang terdapat pada akun resmi Samsung Indonesia "@Samsung_ID", data tersebut dapat dianggap mewakili sentimen dari pengguna atau *followers* Samsung Indonesia. Data berupa sebuah kalimat dengan panjang maksimal 140 karakter. Contoh dari setiap *tweet* yang mengandung sentimen pada akun "@Samsung_ID" terdapat pada Gambar 2. Gambar 3.



Gambar 2. Tweet dengan sentimen positif



Gambar 3. Tweet dengan sentimen negatif

Data *tweet* yang diambil dari akun resmi Samsung Indonesia masih berupa data mentah dan terdapat noise atau ciri-ciri yang tidak mempunyai pengaruh pada klasifikasi sentimen seperti link, "@", *stopword*, *hashtag* yang ditandai dengan munculnya karakter "#". Contoh *tweet* mentah tersebut ada pada Gambbar 4.



Gambar 4. Contoh data tweet

Selain itu, karateristik *tweet* yang dirasa mempunyai pengaruh pada penentuan sentimen adalah terdapat

emoticon yang merupakan salah satu cara untuk mengekspresikan ungkapan persetujuan atau pertidaksetujuan pada suatu kalimat atau *tweets*. Contoh *tweet* mentah yang terdapat *emoticon* ada pada Gambar 5.



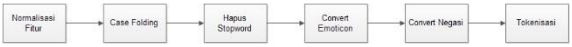
Gambar 5. Contoh data tweet dengan emoticon

Karakteristik *tweet* selanjutnya yaitu ada terdapat kata negasi yang perlu diperhatikan dalam analisis sentimen karena dapat merubah nilai sentimen suatu *tweet*. Kata yang bersifat negasi seperti "tidak", "bukan", "ga", "jangan", "nggak", "tak", "tdk", dan "gak". Contohnya kata "lelet" yang merupakan kata dengan sentimen negatif, karena di depan kata "lelet" terdapat kata negasi "tidak" maka sentimen dari gabungan kata tersebut adalah positif. Contoh *tweet* yang terdapat kata negasi ada padaGambar 6.



Gambar 6. Contoh data tweet dengan negasi

Agar data-data *tweets* tersebut dapat dimanfaatkan dengan baik untuk mengklasifikasikan sentimen maka diperlukan proses *preprocessing* . Pada proses ini data tweet yang digunakan untuk data training data data testing dibersikan dari *noise* atau ciri-ciri yang tidak berpengaruh pada klasifikasi sentimen seperti link, *"@", "RT", stopword,* dan lain sebagainya. Gambaran proses *preprocessing* dapat dilihat pada Gambar 7.



Gambar 7. Tahapan Preprocessing

1. Normalisasi Fitur

Tweet yang terdapat pada akun resmi Samsung Indonesia memiliki berbagai komponen atau karakteristik tweet yang khas seperti "@" yang diidentifikasi sebagai komponen *username*, URL yang dikenal melalui operasi regular, hashtag yang menandakan kata sebagai topik yang sedang dibicarakan, dan "RT" yang diidentifikasi sebagai mengulang kembali *tweet* yang telah diposting. Komponen-komponen tersebut tidak memiliki pengaruh apapun terhadap sentimen, maka akan dibuang.

2. *Case Folding*

Pada proses *case folding* huruf besar atau *uppercase* yang terdapat pada *tweet* diubah menjadi *lowercase* atau huruf kecil.

3. Hapus *Stopword*

Data *tweet* yang sudah melalui proses sebelumnya masih mengandung kata yang dianggap tidak dapat memberikan pengaruh dalam menentukan suatu kategori sentimen. Kata-kata tersebut dimasukkan kedalam daftar stopword yang biasanya berupa kata ganti orang, kata ganti penghubung, pronomial penunjuk, dan lain sebagainya. Jika terdapat kata yang tercantum pada daftar *stopword* maka kata tersebut dihilangkan.

4. *Convert Emoticon*

Convert *emoticon* adalah proses mengkonversikan *emoticon* kedalam string yang sesuai dengan ekspresi *emoticon* itu sendiri. Convert *emoticon* dilakukan karena pada data *tweet* yang diambil dari Samsung Indonesia terdapat *emoticon* yang merupakan salah satu cara mengekspresikan persetujuan atau pertidaksetujuan dalam suatu *tweet*. Hal ini dirasa mempunyai pengaruh terhadap pengklasifikasian sentimen, oleh karena itu *convert emoticon* digunakan.

Tabel 1. Konversi Emoticon [4]

Emoticon	**Konversi**
>:] :-) :) :o) :] :3 :c) :> =] 8) =) :} :^)	esenang
>:D :-D :D &-D 8D x-D xD X-D XD =-D =D =-3 =3	etertawa
>:] :-(:(:-c :c :-< :< :-[:[:{ <><<><	esedih
D:< D: D8 D; D= DX v.v D-‘:	ehoror
>:P :-P :P X-P x-p xp XP :-p :p =p :-b :b	elidah
>:o >:O :-O :O o_O o.O 8-0	ekaget
>:\ >:/ :-/ :-. :/ :\ =/ =\ :S	ekesal
:| :-|	edatar

5. *Convert* Negasi

Convert negasi merupakan proses konversi kata-kata negasi yang terdapat pada suatu *tweet*, karena kata negasi mempunyai pengaruh dalam merubah nilai sentimen pada suatu *tweet*. Kata negasi yang terdapat pada suatu *tweet* akan dihilangkan, dan diberikan penanda . Jika terdapat kata negasi makan akan disatukan dengan kata setelahnya. Kata-kata negasi tersebut meliputi kata "bkn", "bukan", "tida", "tak", "ga", "enggak", "g", "jangan", dan "tidak".

6. Tokenisasi

Pada proses tokenisasi setiap kata pada *tweet* dipisahkan, pada proses ini tahap yang dilakukan adalah memisahkan setiap kata yang dipisahkan oleh spasi, selanjutnya bagian *tweet* yang memiliki karakter selain alphabet, angka, dan garis bawah

akan dipecah sesuai posisi karakter tersebut dan bagian yang hanya memiliki satu karakter non alphabet dan angka akan dibuang .

2.3 Klasifikasi dengan Naïve Bayes

Pada pengklasifikasian menggunakan *naïve bayes* dibagi kedalam 2 proses, yaitu proses training dan testing. Proses training digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan untuk mengklasifikasikan sentiment dengan data testing atau data mentah yang baru. Berikut adalah algoritma klasifikasi sentimen menggunakan *Naïve Bayes Classifier*:

1. Proses Training
   a. Hitung p (c_i).
   b. Hitung p ($w_k|c_i$) untuk setiap kata w_k pada model.
2. Proses Testing
   a. Hitung $\prod_k$ p ($w_k|c$) × p (c)untuk setiap kategori.
   b. Tentukan kategori dengan nilai $\prod_k$ p ($w_k|c$) × p (c) maksimal.

Pemberian kategori dari sebuah dokumen dilakukan dengan memilih nilai *c* yang memiliki nilai $p(C = c_i| D = d_j)$ maksimum, dan dinyatakan dengan:

$$c^* = arg\, max_{c \in C} \prod_k p\,(w_k|c\,) \times p\,(c) \quad (1)$$

Dimana :

$\prod_k p\,(w_k|\,C = c_i)$ adalah hasil perkalian dari probabilitas kemunculan semua kata pada dokumen *dj*.

$p\,(c)$adalah hasil probabilitas dari semua kategori pada data training.

Kategori *c** merupakan kategori yang memiliki nilai $p(C = c_i \mid D = d_j)$ maksimum. Nilai $p(D = d_j)$ tidak mempengaruhi perbandingan karena untuk setiap k ategori nilainya akan sama.

3. IMPLEMENTASI DAN PENGUJIAN

3.1 Lingkungan Implementasi

Perangkat keras yang digunakan dalam proses perancangan dan implementasi sistem memiliki beberapa spesifikasi sebagai berikut:

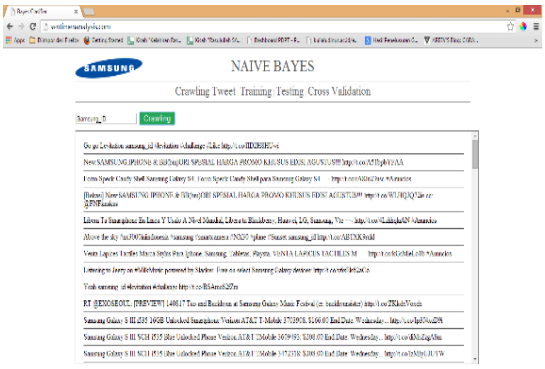
1. Processor Intel® Core™ i3 CPU M 330 @2.13GHz(4 CPUs)
2. Memory 3072MB RAM
3. Monitor resolusi 1366x768 pixels
4. Keyboard
5. Mouse

Perangkat lunak yang digunakan dalam proses implementasi sistem ini adalah sebagai berikut:

1. Sistem Operasi Windows 8 Pro 32-bit
2. Bahasa pemograman PHP.
3. Code editor berupa SublimeText3Portable.
4. DBMS : MySQL 5.5.27
5. Web Browser: Google Chrome 34.0.1847.137

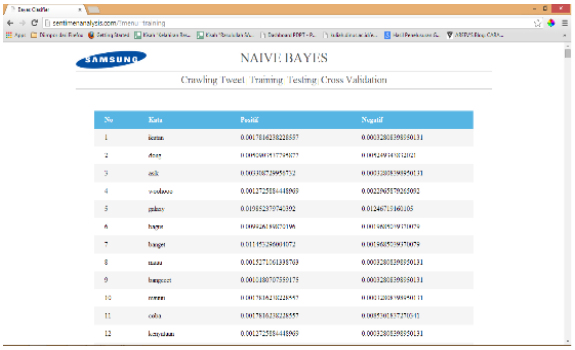
3.2 Impelementasi Antarmuka (Interface)

Pada *sub*-bab ini akan dijelaskan mengenai implementasi antar muka yang telah dirancang sebelumnya. Hasil perancangan antarmuka yang telah diimplementasikan adalah sebagai berikut:



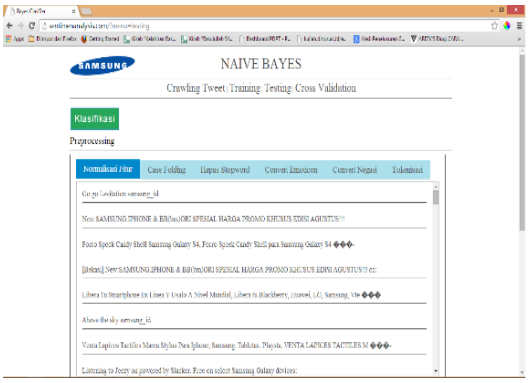
Gambar 8. Antarmuka Crawling Tweet

Gambar 8. merupakan antarmuka *Crawling Tweet* yang diimplementasikan dalam sistem. Tab ini berfungsi untuk *crawling* data *tweet* dari akun resmi Samsung Indonesia, yang nantinya akan digunakan sebagai data testing untuk diklasifikasikan kedalam beberapa kategori sentimen.



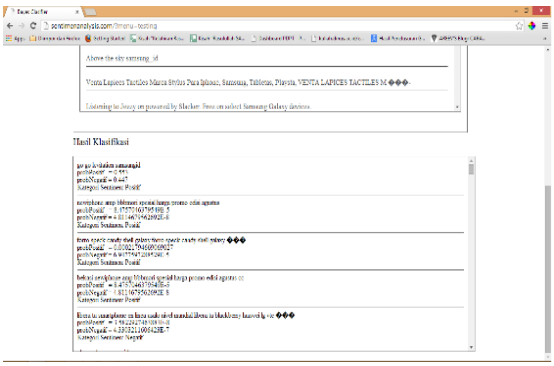
Gambar 9. Antarmuka Training

Gambar 9. merupakan antarmuka training yang berfungsi untuk membuat model probabilitas dari data training yang telah disediakan di dalam *database*, model probabilitas tersebut merupakan perhitungan probabilitas setiap kata yang terdapat pada data training.



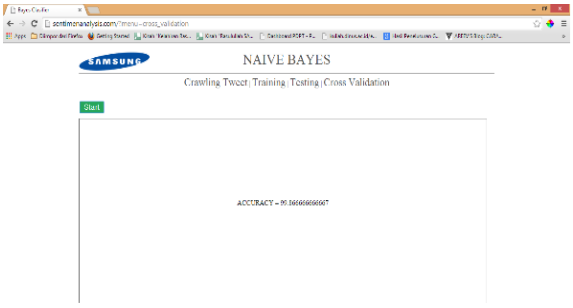
Gambar 10. Antarmuka Preprocessing

Gambar 10. merupakan antarmuka *preprocessing* yang berfungsi untuk menampilkan hasil dari setiap *preprocessing* data testing sebelum data tersebut di klasfikasi, seperti normalisasi fitur, *casefolding*, hapus *stopword*, *convert emoticon*, *convert* negasi, dan tokenisasi.



Gambar 11. Antarmuka Hasil Testing

Gambar 11. merupakan antarmuka hasil testing atau hasil klasifikasi data testing, klasifikasi sentimen dari setiap *tweet* pada data testing didapat dari perhitungan *naïve bayes* yang sebelumnya telah melalui tahap *preprocessing*.



Gambar 12. Antarmuka Cross Validation

Gambar 12. merupakan antarmuka *cross validation* yang berfungsi untuk menguji performa klasifikasi dari sistem yang di bangun . Performa klasifikasi dapat dilihat dari hasil perhitungan akurasi dengan metode *cross validation*.

3.3 Pengujian Akurasi Sistem dengan Metode 10-*fold cross validation*

Jumlah data memegang peranan penting di dalam algoritma *machine learning*. Jumlah data yang sedikit (<100 instance) mungkin membuat algoritma *machine learning* tidak akurat. Algoritma *machine learning* merekomendasikan jumlah instance yang banyak (>1000 instance) namun data itu sendiri tidak mudah untuk diperoleh.

K-fold cross validation adalah teknik yang dapat digunakan apabila memiliki jumlah data yang terbatas (jumlah instance tidak banyak) [12]. *K-fold cross validation* merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. *K-foldcross validation* diawali dengan membagi data sejumlah *n-fold* yang diinginkan. Dalam proses *cross validation* data akan dibagi dalam *n* buah partisi dengan ukuran yang sama $D_1,D_2,D_3..D_n$ selanjutnya proses testing dan training dilakukan sebanyak n kali. Dalam iterasi ke-$_i$ partisi D_i akan menjadi data testing dan sisanya akan menjadi data training. Untuk penggunaan jumlah *fold* terbaik untuk uji validitas, dianjurkan menggunakan 10-*fold cross validation* dalam model[13].

Skenario pengujian merupakan tahap penentuan pengujian yang dilakukan. Pengujian dilakukan menggunakan metode *k-cross validation* dengan nilai *k* sebanyak 10 *fold*, pengujian ini bertujuan untuk mengetahui akurasi metode *naïve bayes classifier* yang diterapkan pada analisis sentimen jika diuji dengan data training dan data testing yang berbeda . Penggunaan 10 *fold* ini dianjurkan karena merupakan jumlah *fold* terbaik untuk uji validitas[13]. Tahap pengujian dengan menggunakan metode 10-*fold cross validation* membagi dataset yang awalnya berjumlah 1500 data akan dibagi menjadi 10 *subset*(bagian) masing-masing *subset* berjumlah 150 data. Pada *fold* pertama terdapat kombinasi 9 *subset* yang berbeda digabung dan digunakan sebagai data training, sedangkan 1 *subset* (sisa) digunakan sebagai data testing, selanjutnya proses training dan testing dilakukan sampai *fold* kesepuluh. Skenario uji akurasi dengan metode 10-*fold cross validation* dapat dilihat pada Tabel 2.

Tabel 2. Skenario Uji Stabilitas dengan 10-*fold crossvalidation*

Fold	**Data**	***Subset***
Fold 1	Training Testing	$S_2, S_3, S_4, S_5, S_6, S_7 S_8, S_9, S_{10}$ S_1
Fold 2	Training Testing	$S_1, S_3, S_4, S_5, S_6, S_7 S_8, S_9, S_{10}$ S_2
Fold 3	Training Testing	$S_1,S_2, S_4, S_5, S_6, S_7 S_8, S_9, S_{10}$ S_3
Fold 4	Training Testing	$S_1,S_2, S_3, S_5, S_6, S_7 S_8, S_9, S_{10}$ S_4
Fold 5	Training Testing	$S_1,S_2, S_3, S_4, S_6, S_7 S_8, S_9, S_{10}$ S_5
Fold 6	Training Testing	$S_1,S_2, S_3, S_4, S_5, S_7, S_8, S_9, S_{10}$ S_6
Fold 7	Training Testing	$S_1,S_2, S_3, S_4, S_5, S_6, S_8, S_9, S_{10}$ S_7
Fold 8	Training Testing	$S_1,S_2, S_3, S_4, S_5, S_6, S_7,S_9, S_{10}$ S_8
Fold 9	Training Testing	$S_1,S_2, S_3, S_4, S_5, S_6, S_7$, S_{10} S_9
Fold 10	Training Testing	$S_1,S_2, S_3, S_4, S_5, S_6, S_7 S_8, S_9$ S_{10}

3.4 Hasil dan Analisis Pengujian

Pada skenario uji dengan jumlah 10-*fold cross validation* dan iterasi yang dilakukan sebanyak 10 kali iterasi rata-rata yang dihasilkan yaitu 98.87%. Iterasi pertama pada *fold* pertama menggunakan 1350 data gabungan dari *subset* ke-2sampai *subset* ke-9 sisanya yaitu 150 data yang digunakan sebagai data testing pada *subset* pertama, proses training dan proses testing dilakukan sampai iterasi ke-10. Detail hasil dari akurasi pengujian skenario pada masing-masing *fold*nya dapat dilihat pada Tabel 3.

Tabel 3. Hasil Uji Stabilitas dengan 10-*fold cross validation*

Fold	**Data Training**	**Data Testing**	**Akurasi**
1	1350	150	100 %
2	1350	150	100 %
3	1350	150	100 %
4	1350	150	98 %
5	1350	150	100 %
6	1350	150	98 %
7	1350	150	100 %
8	1350	150	93.33 %
9	1350	150	99.33 %
10	1350	150	100 %
Rata-rata akurasi			98.87 %

Tabel 3. mengindikasikan bahwa nilai akurasi pada tiap fold sudah sangat stabil. Percobaan pada *fold* ke-8 memperlihatkan hasil yang paling rendah diantara 9 *fold* lainnya yaitu ≈ 93.33%, sedangkan *fold* sisanya menghasilkan akurasi yang lebih besar dengan aurasi paling tinggi ada pada *fold* 1, 2, 3, 5, 7, dan 10. Jika dianalisis performa klasifikasi dengan *Naïve Bayes* yang dihasilkan sudah bisa dikatakan akurat, serta kinerja model klasifikasi sudah bisa dikatakan baik dengan melihat rata-rata akurasi dari setiap *fold* yang cukup besar yaitu ≈ 98.87% pada *10 fold cross validation* dengan data training yang digunakan sebanyak 1350 data, karena secara umum semakin besar data training maka akan semakin baik kinerja model klasifikasi suatu sistem [15].

4. PENUTUP

4.1 Kesimpulan

Dari proses pengujian dan analisis yang telah dilakukan, kesimpulan yang dapat diambil antara lain:

1. Metode *Naïve Bayes Classifier* dapat diterapkan dengan baik pada analisis sentimen dalam mengklasifikasikan sentimen positif dan negatif, hal ini bisa dilihat dari akurasi yang didapatkan cukup tinggi dengan menggunakan pengujian 10-*fold cross validation*.
2. Jika suatu data *tweet* tidak terdapat di dalam data training, maka data *tweet* tersebut tidak dapat diklasifikasikan ,karena peluang untuk data *tweet* diklasifikasikan kedalam suatu kelas adalah sama untuk setiap kelas yang ada.
3. Dalam proses pengujian dengan 1500 data dan menggunakan metode 10-*fold cross validation*, waktu komputasi cukup lambat dengan waktu roses 220 detik. Hal ini disebabkan oleh proses data training yang berkali-kali dengan jumlah data yang sangat besar serta penggunaan bahasa pemograman PHP (*Hypertext Prepocessor*) yang mempunyai kelemahan lambat dalam hal komputasi perhitungan.

4.2 Saran

Dari hasil pengujian , analisis dan kesimpulan yang telah dirumuskan, terdapat beberapa hal yang disarankan untuk penelitian selanjutnya, saran tersebut diantaranya:

1. Pada proses analisis sentimen bisa digunakan metode selain Naïve Bayes Classifier seperti Suport Vektor Machine (SVM) sebagai metode untuk mengklasifikasikan sentimen kedalam beberapa kategori guna melakukan perbandingan akurasi untuk mengetahui metode mana yang lebih baik sebagai implementasi dari analisis sentimen.
2. Penelitian selanjutnya diharapkan data *tweet* yang tidak bisa diklasifikasikan kedalam kelas yang sudah ditentukan dapat dianalisis lebih lanjut misalnya menyimpan data tersebut untuk dijadikan data traning dan dilakukan proses training kembali atau ditambahkan kedalam list *stopword* jika memang terdapat kata baru yang tidak mempunyai pengaruh penting untuk klasifikasi.
3. Penelitian selanjutnya diharapkan dapat menganalisis kembali bahasa pemograman yang digunakan untuk mengimplementasikan analisis sentimen agar lebih cepat dalam hal waktu dan lebih optimal .
4. Analisis sentimen pada penelitian ini belum memperhatikan semantik yaitu makna kata dan kalimat. Penelitian selanjutnya diharapkan dapat menggunakan makna kata dan kalimat untuk menentukan kategori sentimen pada suatu dokumen agar mendapatkan hasil yang lebih baik.

DAFTAR PUSTAKA

[1] M. L. Khodra, "kur2003.if.itb.ac.id," [Online]. Available: http://kur2003.if.itb.ac.id/. [Accessed 11 Mei 2014].

[2] L. Pakhpahan, "Social Media Untuk Promosi Bisnis," [Online]. Available: http://latiefpakpahan.com/social-media-promosi-bisnis/.[Accessed 22 March 2014].

[3] G.Andreani,"repostitory.usu.ac.id/ Pengaruh Promosi melalui Media Sosial," 2013. [Online].Available :http://repository.usu.ac.id/bitstream/123456789/37140/5/Chapter%20I.pdf. [Accessed 22 March 2014].

[4] I. Sunni and D. H. Widyantoro, "Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhdap Tokoh Publik," *Jurnal Sarjana Institut Teknologi Bandung Bidang Elektro dan Informatika*, vol. 1, pp. 200 - 206, 2012.

[5] J. D. M. Rennie, L. Shih, J. Teevan and D. R. Karger, "Tackling the Poor Assumptions of Naive BAyes Text Classifiers," in *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, 2003.

[6] S. Dharma, "Pendekatan, Jenis, dan Metode Penelitian Pendidikan," *Kompetensi Penelitian dan Pengembangan*, pp. 19-46, 2008.

[7] I. Sommerville, Software Engeneering, United States of America: Addision Wesley, 2011.

[8] I. H. Wltten, "Text Mining," in *Computer Science, University of Waikato*, Hamilton, New Zealand, 2003.

[9] M. Lan, "A New Term Weighting Method for Text Categorization," *National University of Singapore*, 2006.

[10] A. Nurfalah, "Analisis Sentimen Pada Opini Berbahasa Indonesia Menggunakan Pendekatan Lexicon Based," Bandung 'Indonesia, 2011.

[11] T. M. Mitchell, Machine Learning, McGraw-Hill Science/Engineering/Math, 2005.

[12] D.Anggraeni,http://lontar.ui.ac.id/, [Online].Available: http://lontar.ui.ac.id/file ?file=digital/123561-SK-739-Klasifikasi%20 topik-Analisis.pdf. [Accessed 18 Juni 2014].

[13] R. Kohavi, "A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,"1995.[Online].Available: http://frostiebek.free.fr/docs/Machine%20 Learning/validation-1.pdf. [Accessed 23 June 2014].

[14] A. Haryono and A. Kustiyo, "Identifikasi Daun Tanaman Jati Menggunakan Jaringan Syaraf Tiruan Backpropagation dengan Ekstraksi Fitur Ciri Morfologi Daun".

[15] H.Murfi,"http://ocw.ui.ac.id/,"[Online].Available: http://ocw.ui.ac.id/pluginfile.php/246/mod_resource/content/0/1.2%20Pendahuluan%20-%20Evaluasi.pdf. [Accessed 12 July 2014].