Analisis dan Implementasi Pengklasifikasian Pesan Singkat pada Penyaringan SMS Spam Menggunakan Algoritma Multinomial Naïve Bayes

Muhammad Budi Hartanto
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
muhammadbudi.h@gmail.com

Shaufiah, S.T.,M.T.
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
shaufiah@gmail.com

Ir. Moch. Arif Bijaksana, M. Tech.
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
arifbijaksana@gmail.com

Abstrak *– Maraknya penggunaan SMS membuat pihak tertentu memanfaatkan hal tersebut dengan menyebar SMS ke berbagai pihak demi keuntungannya sendiri. SMS tersebut disebut juga dengan SMS Spam. Bagi sebagian orang, hal tersebut sangat mengganggu. Oleh sebab itu, maka dibangunlah SMS Spam Filter yang bertujuan untuk menyaring SMS. Dalam tugas akhir ini, dilakukan penelitian tentang SMS Spam Filter, yang akan mengklasifikasikan SMS menjadi Spam dan Ham(Tidak Spam). Isi SMS yang berupa teks dan cenderung tidak teratur, membuat data yang akan digunakan untuk klasifikasi perlu dilakukan preprocessing terlebih dahulu. Preprocessing yang digunakan dalam tugas akhir ini adalah Slang Handling, Stopword Elimination, Stemming, dan Tokenization. Setelah semua data dilakukan preprocessing maka algoritma yang akan mengklasifikasikan SMS-nya adalah algoritma multinomial naïve bayes. Tugas akhir ini akan menggunakan 2 sumber data yaitu: The SMS Spam Collection v.1 dan British SMS. Yang keduanya akan dikomposisikan sesuai pengujian. Pengujian yang dilakukan yaitu membandingkan beberapa skenario pengujian berdasarkan penggunaan preprocessingnya. Dan telah didapatkan hasil terbaik dengan akurasi mencapai 98.15% dengan pemilihan preprocessing Stopword Elimination dengan Stemming.*

Kata Kunci *– Klasifikasi, SMS Spam Filter, Slang, Stopword Elimination, Stemming, Multinomial Naïve Bayes.*

I. PENDAHULUAN

Komunikasi menjadi hal yang sangat penting bagi kehidupan setiap orang, untuk saling berinteraksi satu dengan yang lain. Pada beberapa tahun terakhir, media pesan singkat atau yang lebih akrab disebut SMS (*Short Message Service*) menjadi media yang paling sering digunakan dalam komunikasi antar *mobile* [1]. Hal tersebut membuat sebagian orang yang tidak bertanggung jawab menggunakan fasilitas pesan singkat sebagai cara untuk menguntungkan dirinya sendiri, salah satunya yaitu digunakan sebagai media promosi berbagai produk atau jasa.

Bagi sebagian orang yang menganggap pesan singkat menjadi hal yang penting, mereka merasa terganggu akan adanya pesan-pesan singkat yang isinya hanya berupa promosi, layanan jasa atau pesan-pesan lain yang isinya tidak diinginkan. Pesan yang dikirimkan ke sembarang target secara langsung ataupun tidak langsung dengan isi yang tidak diharapkan oleh target ini disebut dengan *SMS Spam* [2]. Dari kondisi tersebut, tugas akhir ini membangun sebuah cara untuk melakukan *Spam filter*. *Spam Filter* merupakan sebuah teknik yang yang dibangun untuk secara otomatis mengidentifikasi Spam [2]. Dengan SMS Spam filter yang akan dibangun, maka setiap SMS yang masuk ke dalam sistem akan diidentifikasi dan diklasifikasikan.

II. RELATED WORK

Dalam sebuah penelitian tentang klasifikasi text untuk kasus SMS spam, disebutkan bahwa algoritma multinomial naïve bayes mampu mencapai akurasi tertinggi 98.22% [3]. Lebih baik dibandingkan model bayes lainnya. Diantaranya seperti : Bayesian Logistic Regression yang hanya mencapai 97.29%, sedangkan Naïve Bayes hanya mencapai 93.50%. Akan tetapi dalam penelitian tersebut tidak ada penanganan khusus untuk data SMS, sedangkan pada penelitian lain mendapatkan hasil lebih baik jika ada penambahan dalam preprocessing data sebelum diklasifikasikan yaitu untuk Naïve Bayes mencapai akurasi 98% [4]. Dari penelitian tersebut dapat diketahui bahwa preprocessing cukup berpengaruh pada hasil klasifikasi. Dari kondisi tersebut, maka dibuatlah tugas akhir ini, yang meneliti hasil dari algoritma multinomial naïve bayes jika ditambah beberapa preprocessing untuk menangani data SMS sebelum dilakukannya klasifikasi.

III. METODE/ALGORITMA

A. SMS Spam Filter

SMS Spam atau yang dikenal juga dengan *Mobile Spam*, merupakan bagian dari spam yang melibatkan penggunaan pesan teks yang dikirim ke ponsel melalui layanan pesan singkat, *SMS Spam* biasanya berisi hal-hal yang berkaitan dengan promosi, layanan jasa, atau hal-hal yang tidak diinginkan yang dikirmkan ke sembarang target [2] [7]. Terdapat 2 teknik dalam *SMS spam filter* pada ponsel [6], yaitu :

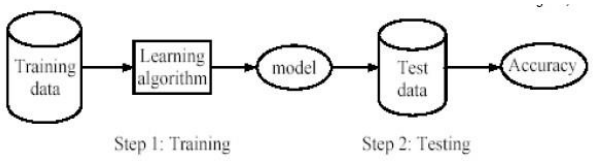
1. *Black and White List*
   Teknik ini sangat sederhana karena hanya membandingkan pesan yang masuk dengan kata-kata yang ada dalam black list.
2. *Text Classification*
   Teknik yang mengelompokkan pesan dari isi atau text yang ada dalam sebuah dokumen sesuai dengan kategori yang telah ditentukan.

B. Klasifikasi

Klasifikasi merupakan proses pembelajaran secara terbimbing (*supervised learning).* Dalam melakukan klasisfikasi akan dibutuhkan data training yang akan digunakan sebagai data pembelajaran. Pada setiap data training sudah memiliki data atribut dan labelnya [8] [9].

Tahap dari klasifikasi terbagi menjadi 2 bagian yaitu :

1. *Learning* (Pembelajaran) : Pada tahap ini pembelajaran dilakukan dengan menggunakan data training.
2. *Testing* : Pada tahap ini dilakukan pengujian terhadap model menggunakan data testing.



Gambar 1 : Skema Klasifikasi

Klasifikasi teks adalah proses pengelompokan dokumen kedalam kelas berbeda, yang dalam tahapan tiap dokumen menunjuk pada satu kelas atau kategori tertentu. Sehingga dokumen tersebut harus dapat merepresentasikan dari kelasnya sehingga tiap kata yang muncul dalam dokumen tersebut mempunyai nilai [10].

Klasifikasi teks sudah sering digunakan untuk menyaring Email Spam, dengan menggunakan pola algoritma pengenalan seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, *Artificial Neural Network (ANN)*, *Decision Tree*, *k-Nearest Neighbor (kNN)*, dan *Hidden Markov Model (HMM)* [6].

C. Multinomial Naïve Bayes

Pada *Multinomial Naïve Bayes*, yang diambil adalah jumlah kemunculan kata dalam sebuah dokumen, dengan tetap menggunakan asumsi bayes bahwa setiap kata tidak terkait dengan kata lain dalam sebuah dokumen. Probabilitas yang akan dihitung adalah probabilitas dari setiap kategori dan setiap kata yang ada.

Probabilitas kategori menggunakan rumus [13]:

$$P(c) = \frac{N_c}{N} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.8)$$

Untuk menghitung setiap probabilitas dari kata, setiap kata yang ada akan memiliki nilai probabilitas sebanyak kategori. Jika dalam seluruh dokumen hanya ada 2 kategori, maka setiap kata akan memiliki 2 nilai probabilitas yang berdasarkan kategorinya. rumus yang digunakan dalam perhitungan nilai probabilitas dari kata sebagai berikut [13]:

$$P(w|c) = \frac{count(w,c)+1}{count(c)+|v|} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.9)$$

Keterangan :

- Count(w,c) = merupakan banyaknya kata (w) dalam seluruh SMS yang memiliki kelas (c).
- Count(c) = merupakan banyaknya kata yang ada dalam SMS yang memiliki kelas (c).
- |v| = merupakan *vocabulary* kata yang ada dalam SMS.

Pada rumus tersebut dilakukan *smoothing laplace* , yaitu dengan menambahkan 1 untuk memberikan nilai probabilitas yang cukup baik apabila jumlah kata sangat kecil atau mendekati nol. Selain itu juga digunakan *smoothing laplance* untuk penanganan kata yang tidak dikenal. Sehingga didapatkan rumus [13]:

$$P(w|c) = \frac{1}{count(c)+|v|+1|} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.10)$$

Dari informasi yang telah didapatkan, yaitu berupa probabilitas kategori dan probabilitas kata. Untuk penentuan kategori dari sebuah dokumen, semua kata yang ada pada dokumen akan dihitung, termasuk probabilitas dari kategori.

Penentuan kategori dalam *Multinomial Naïve Bayes* dapat dicontohkan sebagai berikut.

Dokumen = "word1 word2 word3 word1 word3"

Kategori = c1 dan c2

Dalam contoh tersebut perhitungan yang terjadi adalah :

$$P(c1|D_{word1\,word2\,word3\,word1\,word3}) = P(c1) * P(word1|c1)^2 * P(word2|c1) * P(word3|c1)^2$$

$$P(c2|D_{word1\,word2\,word3\,word1\,word3}) = P(c2) * P(word1|c2)^2 * P(word2|c2) * P(word3|c2)^2$$

Kategori akan ditentukan dari nilai tertinggi antara $P(c1|D_{word1\,word2\,word3\,word1\,word3})$ dan $P(c2|D_{word1\,word2\,word3\,word1\,word3})$.

IV. DATA DAN PREPROCESSING

A. Pembagian Data

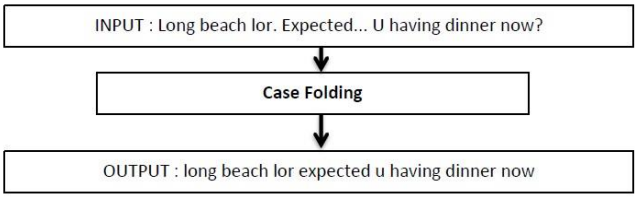
Untuk mendapatkan data training dan data testing pada pengujian tugas akhir ini, pembagian data yang digunakan adalah pembagian dengan cara k-fold cross validation, yaitu dengan membagi data yang telah didapatkan sebanyak k sampel data. Dalam tugas akhir ini digunakan nilai k=5 dan 10, yang merupakan nilai umum yang sering digunakan.

Dari dataset yang telah didapatkan melalui The SMS Spam Collection v.1 [5] dan British SMS [6]. Total data sebanyak 6449, dengan rincian SMS Ham = 5277 dan SMS Spam = 1172. Untuk pengujian tugas akhir ini, data yang digunakan sebanyak 6000 dengan SMS Ham = 5000 dan SMS Spam 1000. Kemudian dari data tersebut, dibagi menjadi data training dan data testing sesuai penggunaan 5-fold dan 10-fold Cross Validation.

B. Preprocessing

1) Case Folding

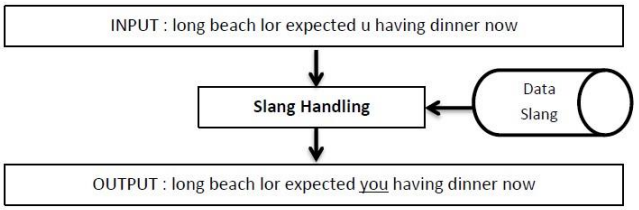
Pada proses Case Folding ini, akan dilakukan perubahan terhadap huruf kapital menjadi huruf kecil, dan juga penghapusan karakter tanda baca. Berikut ini merupakan contoh dan proses case folding.



Gambar 2 : Proses Case Folding

2) Slang Handling

Pada proses *slang handling* ini, setiap kata dalam SMS akan diperiksa satu-persatu dengan data slang yang terdiri dari 2909 kata slang, dari A-Z dan beberapa berupa kombinasi karakter tanda baca [14]. Berikut ini adalah gambaran tahapan dari proses *slang handling*.



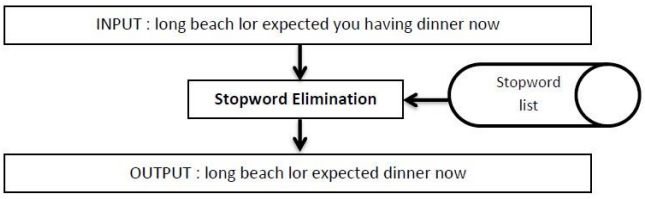
Gambar 3 : Proses Slang Handling

3) Stopword Elimination

Penerapan dari proses *stopword elimination* hampir sama dengan proses *slang handling*, yaitu dengan memanfaatkan sumber data yang telah ada. Hanya saja dalam proses ini, kata yang terdeteksi dalam data sumber, maka kata tersebut akan dihapus.

Tujuan dari proses ini adalah untuk mengurangi kata-kata yang dianggap tidak perlu atau tidak mempengaruhi informasi dari data yang ada. Dengan berkurangnya jumlah kata dalam isi SMS, maka jumlah perhitungan akan lebih kecil. Sehingga diharapkan akan mengurangi waktu komputasi dan membuat data lebih efektif untuk digunakan.

Dalam proses ini digunakan 2 *stopword list* dengan jumlah kata yang berbeda. *Stopword list 1* terdiri dari 173 kata dari A-Z [15], sedangkan untuk *Stopword list 2* terdiri dari 556 kata dari A-Z [15]. Perbedaan *stopword list* ini ditujukan untuk melihat kualitas data jika dilakukan proses *stopword elimination*. Berikut merupakan contoh dari proses *stopword elimination*.

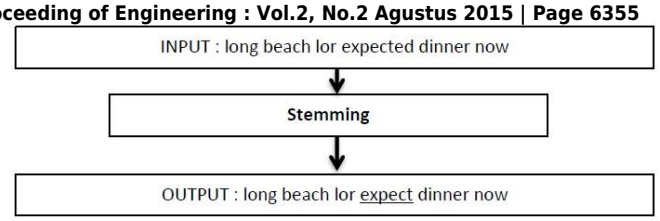


Gambar 4 : Proses Stopword Elimination

4) Stemming (Porter)

Penerapan untuk proses *stemming* menggunakan model *Porter Stemmer*. Proses ini akan mengecek kata demi kata yang ada pada SMS. Dalam *porter stemmer*, setiap kata akan melalui 6 tahap pengecekan.

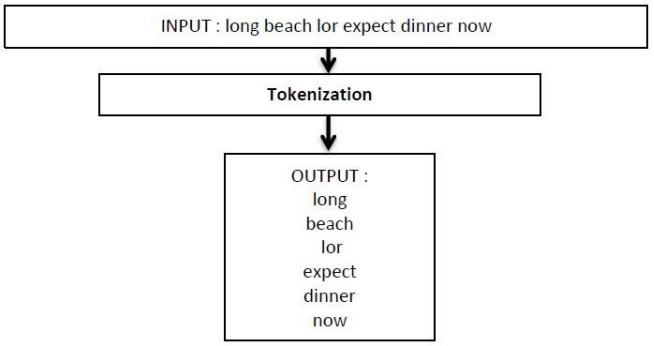
Tujuan dari proses ini adalah untuk menyederhanakan kembali kata-kata yang sudah mengalami perubahan. Penyederhanaan tersebut juga dapat mengurangi jumlah kata yang akan diproses pada tahap selanjutnya. Contohnya kata "Teacher", "Teach", dan "Teaching". Setelah dilakukan proses stemming, kata yang ada menjadi "Teach", "Teach", dan "Teach", yang awalnya 3 kata berbeda menjadi satu kata yang sama, karena memiliki kata dasar yang sama. Berikut ini merupakan gambar dari proses stemming.



Gambar 5 : Proses Stemming

4) Tokenization

Proses *tokenization* diterapkan dalam setiap proses yang membutuhkan pengecekan kata demi kata. Proses ini akan memecah setiap kata yang dipisahkan oleh spasi. Berikut ini gambaran dari proses tokenization.

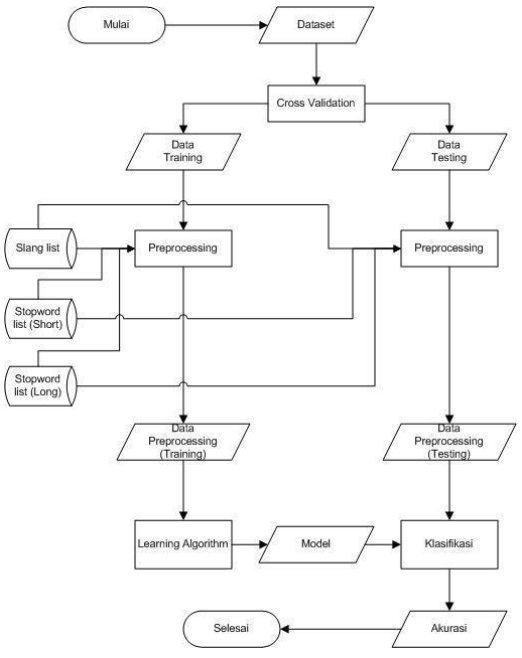


Gambar 6 : Proses Tokenization

V. PEMBAHASAN

A. Gambaran Umum Sistem

Sistem yang dibangun bertujuan utama menganalisa pesan singkat (SMS) untuk diklasifikasikan antara *spam* dan *ham*. Klasifikasi yang digunakan adalah *Multinomial Naïve Bayes*, dan ditambah beberapa *preprocessing* yang bertujuan untuk menangani permasalahan terhadap isi SMS yang masih bebas. Berikut ini adalah gambaran umum dari sistem :



Gambar 7 : Gambaran Umum Sistem

B. Skenario Pengujian

Skenario pengujian ini berdasarkan pemilihan dari preprocessing yang akan digunakan. Preprocessing yang akan digunakan adalah Slang Handling, Stopword Elimination dan Stemming. Dengan Algoritma yang digunakan yaitu Multinomial Naïve Bayes. Sehingga ketentuan skenario yang akan dilakukan sebagai berikut:

a. Untuk Slang Handling, terdapat skenario pemilihan untuk digunakan atau tidak digunakan.
b. Untuk Stopword Elimination, sesuai dengan perancangan terdapat 2 stopword list yang digunakan, yaitu stopword list 1 (short) dan stopword list 2 (long). Sehingga akan ada skenario untuk penggunaan short, penggunaan long, dan tidak menggunakan stopword elimination.
c. Untuk Stemming digunakan pada seluruh skenario pengujian, dengan maksud sebagai preprocessing tambahan pada sistem SMS Spam Filtering ini.
d. Untuk Algoritma Multinomial Naïve Bayes digunakan pada seluruh skenario sebagai algoritma yang akan menentukan kelas dari SMS.

e.

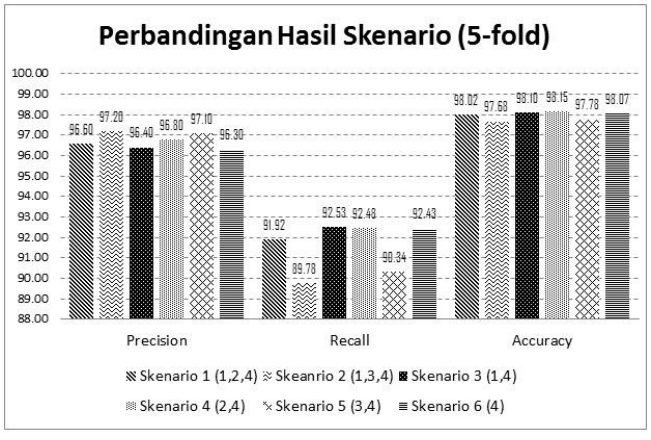
Maka didapatkanlah variasi skenario pengujian sebagai berikut:

Skenario	Preprocessing	Keterangan
1	1,2,4	(1). Slang Handling (2). Stopword (short) (3). Stopword (long) (4). Stemming
2	1,3,4	
3	1,4	
4	2,4	
5	3,4	
6	4	

Tabel 1 : Pembagian Skenario Pengujian

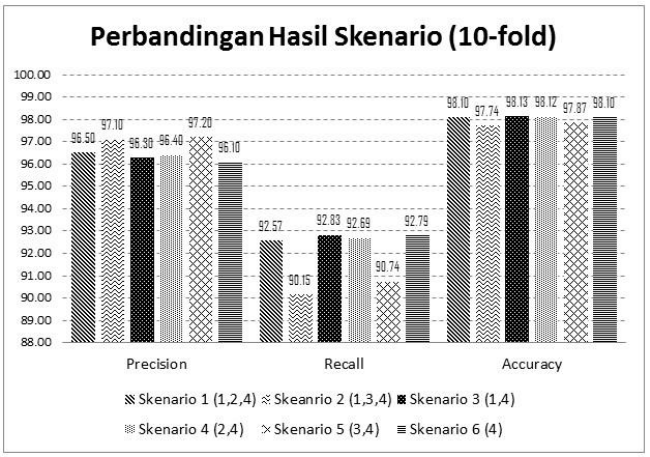
C. Hasil dan Analisis

Dari keseluruhan hasil yang didapatkan, kemudian dilakukan perbandingan antar skenario untuk melihat kualitas skenario yang paling baik untuk sistem SMS Spam Filtering dalam tugas akhir ini. Berikut merupakan grafik perbandingan dari rata-rata pada setiap skenarionya dengan pembagian data 5-fold.



Gambar 8 : Grafik Perbandingan Hasil Skenario (5-fold)

Berikut merupakan grafik perbandingan dari rata-rata pada setiap skenarionya dengan pembagian data 10-fold.



Gambar 9 : Grafik Perbandingan Hasil Skenario (10-fold)

Dalam 2 grafik tersebut terdapat 3 perbandingan, yaitu untuk nilai precision, recall dan accuracy.

- Untuk precision, seluruh skenario mendapatkan hasil yang cukup baik yaitu diatas 96%, yang berarti bahwa sistem mendeteksi SMS Spam dengan baik.
- Untuk recall, terdapat berbedaan hasil yang cukup signifikan pada skenario 2 dan skenario 5, yang keduanya sama-sama menggunakan preprocessing 3 yaitu Stopword Elimination dengan stopword list long. Berarti dapat disimpulkan bahwa stopword list long kurang bagus untuk digunakan dalam Stopword Elimination dalam kasus SMS Spam.
- Untuk accuracy, hampir seluruh skenario mencapai persentase 98%, yang berarti sistem dalam menentukan SMS spam ataupun SMS ham sudah sangat baik. Namun untuk skenario 2 dan 5 masih sedikit lebih kecil dari skenario lain, yang kemungkinan juga diakibatkan karena penggunaan preprocessing 3.

Dari hasil yang didapatkan, diketahui bahwa algoritma multinomial naïve bayes memang efektif untuk digunakan dalam SMS Spam Filtering. Dimana terdapat 2 skenario mencapai akurasi 97% lebih, dan 4 skenario lain mencapai akurasi lebih dari 98%.

VI. KESIMPULAN

Berdasarkan pengujian serta analisis yang telah dilakukan pada tugas akhir ini, dapat diambil beberapa kesimpulan, yaitu :

1. Skenario atau pemilihan preprocessing terbaik untuk algoritma multinomial naïve bayes adalah menggunakan Slang Handling dengan Stemming atau menggunakan Stopword Elimination(short) dengan Stemming. Yang hasil akurasinya mencapai 98.10% - 98.15%.
2. Penggunaan Stopword Elimination(long) ternyata kurang cocok untuk kasus SMS Spam Filtering dengan algoritma multinomial naïve bayes, karena dapat berpengaruh mengurangi nilai dari recall dan akurasinya.
3. Penggunaan preprocessing slang handling akan lebih cocok untuk isi SMS yang penulisannya tidak formal/teratur. Sedangkan untuk penulisan yang sudah formal, slang handling tidak akan berpengaruh.
4. Penggunaan preprocessing Stopword Elimination lebih tergantung pada pemilihan kata yang ada pada stopword list,

karena jika terlalu banyak justru akan memperbanyak kata yang dihapus, dan hal itu berpengaruh besar pada jumlah kata dan perhitungan pada multinomial naïve bayes.

5. Penerapan multinomial naïve bayes pada SMS Spam Filter terbagi dalam 2 proses utama, yaitu Learning Algorithm dan Klasifikasi. Learning Algorithm akan menghitung probabilitas kelas dan probabilitas kata, yang merupakan model dari proses klasifikasi. Proses klasifikasi akan menentukan kelas dari sebuah SMS.
6. Proses identifikasi SMS dalam kelas Spam atau Ham ditentukan dengan nilai maksimal dari setiap probabilitas kelas yang ada untuk SMS tersebut. Nilai probabilitas ini didapatkan dari perhitungan probabilitas masing-masing kata yang ada pada SMS dan juga probabilitas kelas.
7. Hasil rata-rata akurasi yang dapat dicapai maksimal 98.15%, yaitu pada skenario 4 yang menggunakan Stopword Elimination(short) dengan Stemming.

Penerapan tidak terbatas hanya pada Bahasa Inggris saja, bisa juga digunakan pada bahasa selain Inggris. Walaupun pada proses preprocessingnya akan memerlukan penyesuaian data sumber yang berkaitan dengan bahasa yang akan diterapkan.

REFERENSI

[1] T. Mitchell, Generative and Discriminative Classifiers : Naive Bayes and Logistic Regression, New York: Mc. Graw, 2006.

[2] G. V. Cormack, "Email Spam Filtering: A Systematic Review," *Foundations and Trends in Information Retrieval*, vol. I, no. 4, pp. 335-455, 2008.

[3] K. Mathew and B. Issac, "Intelligent Spam Classification for Mobile Text Message," *International Conference on Computer Science and Network Technology*, pp. 101-105, 2011.

[4] M. T. Nuruzzaman, C. Lee and D. Choi, "Independent and Personal SMS Spam Filtering," *IEEE International Conference on Computer and Information Technology*, pp. 429-435, 2011.

[5] T. M. Mahmoud and A. M. Mahfouz, "SMS Spam Filtering Technique Based on Artificial Immune System," *IJCSI International Journal of Computer Science*, vol. 9, no. 2, pp. 589-597, 2012.

[6] B. Liu, Web Data Mining, Chicago: Springer, 1998.

[7] S. R. Singh, H. A. Murthy and T. A. Gonsalves, "Feature Selection for Text Classification Based on Gini Coefficient on Inequality," *Workshop and Conference Proceeding 10*, pp. 76-85, 2010.

[8] I. Detuardi and S. Sumpeno, "Klasifikasi Emosi Untuk Teks Bahasa Indonesia Menggunakan Metode Naive Bayes," in *Seminar Nasional Pascasarjana IX – ITS*, Surabaya, 2009.

[9] C. M. Dan Jurafsky, " Multinomial Naive Bayes: A Worked Example," Coursera Stanford, [Online]. Available: https://class.coursera.org/nlp/lecture/28. [Accessed 2 May 2015].

[10] T. A. A. a. J. M. G. Hidalgo., "SMS Spam Collection v.1," 2011. [Online]. Available: http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/. [Accessed 22 March 2015].

[11] "Slang Dictionary - Text Slang & Internet Slang Words," [Online]. Available: http://www.noslang.com/dictionary/. [Accessed 22 March 2015].

[12] "RANKS NL," ranks.nl, [Online]. Available: http://www.ranks.nl/stopwords. [Accessed 29 March 2015].

[13] B. Santoso, Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis, Yogyakarta: Graha Ilmu, 2007.

[14] A. M. H. B. a. I. K. Fatima Zahra Lahlou, "A Text Classification Based Method for Context Extraction from Online Reviews," *Intelligent Systems: Theories and Applications (SITA)*, pp. 1-5, 2013.

[15] K. Tretyakov, "Machine Learning Techniques in Spam Filtering," *Data Mining Problem-oriented Seminar*, pp. 60-79, 2004.

[16] N. Chirawichitchai, "Sentiment Classification by a Hybrid Method of Greedy Search and Multinomial Naïve Bayes Algorithm," *IEEE Eleventh International Conference on ICT and Knowledge Engineering*, 2013.