Klasifikasi Teks Pesan Spam Menggunakan Algoritma *Naïve Bayes*

Ika Novita Dewi[1], Catur Supriyanto[2]
Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang 50131
[1]*ikadewi@research.dinus.ac.id,*[2] caturs@research.dinus.ac.id

ABSTRAK

Intensitas pengiriman teks pesan spam melalui layanan sms semakin meningkat seiring dengan meningkatnya trafik komunikasi. Hal ini bisa meresahkan dan membuat ketidaknyamanan para penerima pesan. Salah satu cara yang bisa terapkan untuk mengatasi pesan spam adalah dengan melakukan filterisasi. Filterisasi diterapkan untuk membedakan pesan yang berisi spam dan pesan yang tidak berisi spam menggunakan teknik klasifikasi teks dengan metode naïve bayes. Naïve bayes efektif diterapkan untuk melakukan klasifikasi data dengan jumlah yang besar. Hasil eksperimen menunjukkan bahwa Naïve Bayes dalam melakukan klasifikasi teks pesan memiliki nilai akurasi 84.40%, precision 45.76% dan recall 88.09% dengan proses dokumen menggunakan word vector TF-IDF tanpa metode prune. Penerapan klasifikasi teks menggunakan Naïve Bayes dengan word vector TF-IDF dapat menghasilkan tingkat akurasi yang baik, sehingga dapat diterapkan untuk memfilter pesanyang berisi spam.

Kata kunci : pesan spam, klasifikasi teks, algoritma Naïve bayes, word vector

1. PENDAHULUAN

Meningkatnya trafik komunikasi melalui pesan singkat atau sms telah banyak dimanfaatkan oleh pihak-pihak tertentu untuk mengirimkan pesan-pesan yang tidak bermanfaat atau pesan spam. Pesan spam yang dikirimkan melalui sms terkadang mengandung konten ilegal yang menyebabkan ketidaknyamanan penerima pesan tersebut [1]. Pesan spam yang banyak beredar meliputi informasi perbankan, pengumuman promosi dan diskon toko, dan tarif baru penyedia layanan komunikasi [2] atau pesan-pesan yang tidak memiliki makna lainnya.

Salah satu hal yang dapat dilakukan untuk mengatasi masalah pesan spam ini adalah dengan melakukan pemblokiran nomor pengirim pesan, tetapi hal ini juga tidak terlalu berefek dalam mengatasi masalah pesan spam. Cara lain yang bisa dilakukan adalah dengan melakukan klasifikasi teks pesan dengan repositori data pesan yang telah ada menggunakan teknik klasifikasi komputasi cerdas. Klasifikasi teks pesan dilakukan untuk membedakan pesan yang berisi spam dan pesan yang tidak berisi spam (ham).

Salah satu cara mengatasi teks pesan spam adalah menerapkan teknik filtrasi spam menggunakan metode Bayesian [3]. Naïve Bayes merupakan salah satu algoritma klasifikasi dalam data mining. Naïve bayes merupakan salah satu algoritma yang efektif untuk melakukan klasifikasi teks karena dapat diterapkan untuk data yang berukuran besar dengan hasil yang akurat. Penerapan naïve bayes pada data set yang besar menunjukkan performa kecepatan dan akurasi yang baik [4].

Beberapa penelitian telah menerapkan naïve bayes dalam melakukan klasifikasi teks, terutama untuk dokumen teks dan email dan masih sedikit yang menerapkan untuk pesan teks. Misalnya Samodra,dkk (2009) mengatakan bahwa naive bayes terbukti dapat digunakan secara efektif dalam klasifikasi dokumen teks berbahasa Indonesia dengan hasil akurasi yang baik dan penggunaan stop words yang tidak memiliki pengaruh besar terhadap hasil klasifikasi [5].

Penelitian ini akan melakukan klasifikasi teks pesan dengan menerapkan algoritma Naïve Bayes karena naïve bayes merupakan salah satu algoritma yang efektif diterapkan untuk melakukan klasifikasi dengan jumlah data yang besar. Proses klasifikasi teks akan dilakukan dengan RapidMiner untuk mencari konfigurasi terbaik Naïve bayes dalam menghasilkan nilai akurasi yang tinggi. Dengan dilakukakannya klasifkasi teks pesan ini maka akan dapat menangani pesan yang terindikasi spam lebih awal sehingga penerima pesan akan membaca pesan yang benar-benar bermanfaat baginya.

2. KLASIFIKASI TEKS PESAN

Layanan pesan sms terdiri dari penerima pesan, isi pesan, pengirim, dan waktu kirim, dan bagi beberapa orang penting untuk mengetahui terlebih dahulu siapa pengirim pesan dan waktu kirimnya dari pada mendahulukan apa isi pesan yang terkirim [1]. Sehingga tidak diketahui apakah dalam isi pesan yang dikirim merupakan pesan yang berisi spam atau tidak.

Wang dan Han [6] mengatakan bahwa filterisasi teks pesan sms secara tradisional merupakan filterisasi yang berbasis teks yang memiliki beberapa kelemahan, seperti penggunaan kata kunci yang bisa diganti dengan simbol-simbol dan munculnya kata-kata baru yang memiliki arti sama dengan kata kunci sehingga menyulitkan untuk melakukan filterisasi isi pesan. Masalah filterisasi ini bisa diselesaikan dengan menerapkan teknik komputasi cerdas, salah satunya dengan *intelligent learning*.

Klasifikasi teks merupakan penjabaran beberapa dokumen teks menjadi katergori-kategori yang telah ditentukan. Klasifikasi teks telah diterapkan dalam beberapa hal misalnya filterisasi email, filterisasi berita, prediksi kecenderungan user, kategorisasi teks dalam web, dan pengorganisasian dokumen. Salah satu algoritma yang dapat diterapkan dalam melakukan klasifikasi teks adalah naïve bayes yang didefinisikan sebagai fitur model yang independen sesuai dengan probabilistic classifier sederhana yang diterapkan berdasarkan teori bayes dengan nilai asumsi yang kuat [7].

3. ALGORITMA NAÏVE BAYES

Zhang dan Wang [1] menjelaskan penerapan algoritma naïve bayes dalam klasifikasi teks meliputi pengklasifikasian teks kedalam kelas dengan probabilitas maksimum dengan perhitungan probabilitas teks $P(cj|dx)$ untuk tiap kelas $P(c_j|d_x) = \frac{P(c_j)P(d_x|c_j)}{P(d_x)}$, $j = 1,2,\ldots|c|$ (1)

Berdasarkan rumus total probabilitas:

$$P(d_x) = \sum_{j=1}^{|c|} P(c_j)P(d_x|c_j) \quad (2)$$

dari probabilitas sebelumnya dari kelas c_j $P(c_j)$ dapat dihitung dengan training set:

$$P(c_j) = \frac{jumlah\ teks\ yang\ dimiliki\ oleh\ c_j dalam\ training\ set}{jumlah\ total\ teks\ dalam\ training\ set} \quad (3)$$

Dengan syarat probabilitas pada teks $P(d_x|c_j)$ dihasilkan dari probabilitas karakter yang muncul dalam teks:

$$P(d_x|c_j) = \prod_{i=1}^{n} P(w_i|c_j) \quad (4)$$

Probalistik karakter dapat dihitung melalui frekuensi dokumen dari training set:
Jumlah teks dari karakter kata w_j

$$P(w_i|c_j) = \frac{jumlah\ teks\ yang\ muncul\ dalam\ kelas\ c_j}{jumlah\ teks\ dalam\ kelas\ c_j} \quad (5)$$

Jika dan hanya jika $P(c_j|x) > P(c_j|x)$, dengan sampel dokumen ditandai dengan kelas $c_j$dan didalamnya terdapat $1 \leq j \leq m, j$

Melakukan pemrosesan teks atau dokumen sebelum klasifikasi teks bertujuan untuk menghasilkan vector kata dari atribut yang berbentuk string. Dalam vector kata, teks atau dokumen akan digunakan untuk menghasilkan vector numeric yang merepresentasikan dokumen, dengan nilai TF-IDF, term frequency, term occurrences, dan binary term occurrences. Selain vector kata, proses dokumen juga menerapkan metode prune yang bertujuan untuk melakukan spesifikasi kata menjadi kata yang sering dan jarang digunakan untuk dihilangkan dalam pembentukan daftar kata dan spesifikasi frekuensi. Metode prune bisa dilakukan dengan dua cara, yaitu percentual dan by ranking.

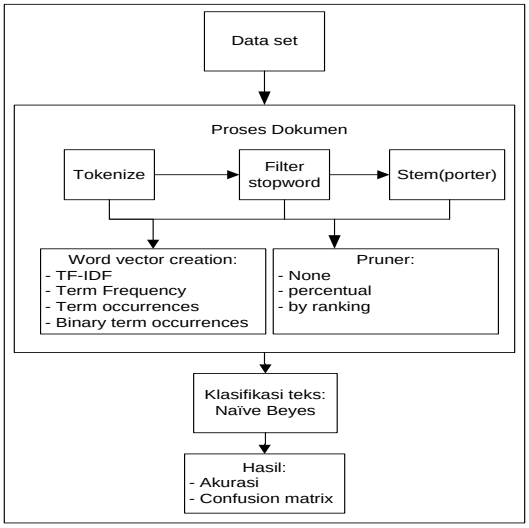
4. DESAIN EKSPERIMEN

Data set sms spam diambil dari UCI Machine Learning Repository [8]. Dataset yang diambil merupakan dataset multivariate dan berjumlah 5574 instances. Tabel 1 menunjukkan sampel dataset sms spam yang digunakan.

Tabel 1: Contoh pesan spam

Kategori	*Isi pesan*
Ham	What you doing? how are you?
Ham	Cos i was out shopping wif darren jus now n i called him 2 ask wat present he wan lor. Then he started guessing who i was wif n he finally guessed darren lor
Spam	FreeMsg: Txt: CALL to No: 86888 & claim your reward of 3 hours talk time to use from your phone now! ubscribe6GBP/ mnth inc 3hrs 16 stop?txtStop
Spam	URGENT! Your Mobile No 07808726822 was awarded a L2,000 Bonus Caller Prize on 02/09/03! This is our 2nd attempt to contact YOU! Call 0871-872-9758 BOX95QU

Eksperimen dilakukan dengan melakukan pengujian terhadap dataset dengan melakukan klasifikasi teks pesan menggunakan Naïve Bayes dan menerapkannya dalam RapidMiner. Eksperimen akan menghasilkan tingkat akurasi dan confusion matrix paling baik yang dihasil oleh Naïve Bayes dengan pilihan konfigurasi dalam pemrosesan teks dengan word vector creation dan metode prune yang yang berbeda-beda, meliputi TF-IDF, term frequency, term occurrences dan binary term occurrences, serta menggunakan pruner percentual dan by rangking. Langkah-langkah eksperimen ini dapat dilihat di gambar 1.



Gambar 1: Desain Eksperimen

5. HASIL DAN PEMBAHASAN

Setelah dilakukan klasifikasi terhadap data set sms spam menggunakan algoritma Naïve Bayes, maka didapatkan tiga hasil perbandingan untuk tingkat akurasi, precision, dan recall menggunakan word-vector creation (TF-IDF, Term Frequency, Term occurrences, dan binary term occurrences) dengan metode pruner percentual dengan nilai prune below= 3.0 dan prune above= 30.0, by ranking dengan nilai prune below= 0.05 dan prune above= 0.95, serta tanpa pruner yang dapat dilihat di tabel 2, 3, dan 4.

Tabel 2: Pebandingan akurasi, precision dan recall tanpa metode pruner

Word vector creation	Pruner	Akurasi	Precision	Recall
TF-IDF	None	84.40%	45.76%	88.09%
Term Frequency	None	84.31%	45.67%	89.69%
Term occurrences	None	84.10%	45.30%	89.69%
Binary term occurrences	None	84.17%	45.45%	90.36%

Tabel 3: Pebandingan akurasi, precision dan recall dengan metode pruner percentual

Word vector creation	Pruner	Akurasi	Precision	Recall
TF-IDF	Percentual	56.28%	22.82%	94.91%
Term Frequency	Percentual	53.23%	21.60%	94.65%
Term occurrences	Percentual	46.12%	19.47%	96.25%
Binary term occurrences	Percentual	47.74%	20.01%	96.65%

Tabel 4: Pebandingan akurasi, precision dan recall dengan metode pruner by ranking

Word vector creation	Pruner	Akurasi	Precision	Recall
TF-IDF	By ranking	74.91%	33.35%	87.28%
Term Frequency	By ranking	75.11%	33.84%	89.69%
Term occurrences	By ranking	75.11%	33.84%	89.69%
Binary term occurrences	By ranking	75.18%	33.97%	90.23%

Nilai akurasi didapatkan dari perhitungan jumlah prediksi benar yang sesuai (TP) ditambah jumlah prediksi benar tidak sesuai (TN) dibandingkan dengan jumlah prediksi benar yang sesuai (TP), jumlah prediksi benar tidak sesuai (TF), jumlah prediksi salah yang sesuai (FP) dan jumlah prediksi salah tidak sesuai (FN).

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Nilai akurasi tertinggi sebesar 84.40% dicapai ketika klasifikasi teks dilakukan dengan Naïve bayes dengan word vector creation TF-IDF tanpa menerapkan pruner, sedangkan nilai akurasi terendah sebesar 46.12% dicapai dengan word vector creation menggunakan term occurrences dan metode pruner percentual. Nilai precision tertinggi dicapai dengan word vector creation TF-IDF tanpa menerapkan pruner dengan nilai sebesar 45.76%. Nilai precision terendah sebesar 19.47% dengan word vector creation menggunakan term occurrences dan metode pruner percentual. Nilai recall tertinggi tercapai sebesar 96.65% dengan menerapkan word vector creation binary term occurrences dan metode pruner percentual. Nilai recall terendah dicapai ketika menerapkan word vector creation TF-IDF dan metode pruner by ranking sebesar 87.28%.

Tabel 5 dibawah ini menampilkan contoh hasil klasifikasi teks pesan spam menggunakan Naïve bayes dengan word vector creation TF-IDF tanpa metode pruner.

Tabel 5: Contoh hasil klasifikasi menggunakan Naïve Bayes

Pesan	Kategori	Hasil Klasifikasi
Oh k...i'm watching here:)	Ham	Ham
07732584351 - Rodger Burns - MSG = We tried to call you re your reply to our sms for a free...	Spam	Ham
Are you unique enough? Find out from 30th August. www.areyouunique.co.uk	Spam	Ham
Sorry, I'll call later	Ham	Ham
For fear of fainting with the of all that housework you just did? Quick have a cuppa	Ham	Spam
Sunshine Quiz Wkly Q! Win a top Sony DVD player if u know which country the Algarve is in?	Spam	Spam

Penerapan metode pruner dalam klasifikasi ini akan berpengatuh terhadap hasil akurasi yang didapatkan, karena dengan menerapkan pruner maka kata akan dikelompokkan menjadi kata yang sering dan jarang digunakan kemudian akan dihilangkan dalam pembentukan daftar kata dan spesifikasi frekuensi. Sehingga untuk klasifikasi dataset sms spam menggunakan naïve bayes ini didapatkan konfigurasi untuk akurasi tertinggi dengan word vector creation TF-IDF tanpa menerapkan pruner dan hasil akurasi 84.40%.

6. PENUTUP

Klasifikasi teks pesan spam menggunakan algoritma Naïve bayes telah dilakukan dengan hasil akurasi 84.40%. Hasil ini diperoleh dengan menggunakan word vector creation TF-IDF tanpa metode pruner. Penggunaan metode pruner dalam pemrosesan dokumen akan mempengaruhi hasil pencapaian akurasi yang didapatkan.

Beberapa hal masih perlu dilakukan untuk meningkatkan kinerja dari peneitian ini. Salah satunya adalah dengan melakukan pengujian Naïve Bayes terhadap data set pesan teks yang berbeda. Salah satunya menggunakan dataset pesan teks yang berbahasa Indonesia, karena intensitas pengiriman teks pesan spam yang berbahasa Indonesia semakin meningkat jumlahnya.

DAFTAR PUSTAKA

[1] H.-y. Zhang and W. Wang, "Application of Bayesian Method to Spam SMS Filtering," in *International Conference on Information Engineering and Computer Science*, 2009.

[2] A. K. Uysal, S. Gunal, S. Ergin and E. S. Gunal, "The Impact of Feature Extraction and Selection on SMS Spam Filtering," in

Elektronika ir Elektrotechnika (Electronics and Electrical Engineering), 2012.

[3] K. Mathew and B. Issac, "Intelligent spam classification for mobile text message," in *International Conference on Computer Science and Network Technology*, 2011.

[4] Y. Huang and L. Li, "Naive bayes classification algorithm based on small sample set," in *IEEE Cloud Computing and Intelligence Systems* , 2011.

[5] J. Samodra, S. Sumpeno and M. Hariadi, "Klasifikasi Dokumen Teks Berbahasa Indonesia dengan Menggunakan Naive Bayes," in *Seminar Nasional Electrical, Informatics, dan IT's Education*, 2009.

[6] W. Qian and H. Xue, "Studying of Classifying Junk Messages Based on The Data Mining," in *Management and Service Science*, 2009.

[7] G. Qiang, "Research and improvement for feature selection on naive bayes text classifier," in *Future Computer and Communication*, 2010.

[8] "UCI Machine Learning Repository," [Online]. Available: http://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection.