KLASIFIKASI SPAM *EMAIL* MENGGUNAKAN *NAÏVE BAYES*

Ratih Yulia H
Akademi Bina Sarana Informatika
Jl. RS. *Fatmawati* No. 24, Pondok Labu, Jakarta Selatan
Email: ratih.ryl@bsi.ac.id

Abstraksi *- Email salah satu alat yang digunakan untuk berkomunikasi dan untuk pertukaran informasi. Pertumbuhan internet yang semakin meningkat maka penggunaan email semakin banyak. Spam email merupakan email yang tidak diinginkan atau diminta oleh penerimanya. Spam email biasanya digunakan untuk menyebarkan virus atau kode berbahaya, penipuan dan iklan. Banyak pengguna merasa terganggu oleh banyaknya waktu yang dihabiskan untuk menghapus pesan spam, besarnya biaya yang harus dikeluarkan dan besarnya bandwith jaringan yang digunakan. Untuk mengatasi masalah ini maka diperlukan suatu metode klasifikasi untuk memisahkan antara spam dan non spam. Metode klasifikasi yang digunakan yaitu Naïve Bayes, metode yang paling popular dan paling banyak digunakan. Penelitian ini menggunakan data spam dan non spam yang sudah diklasifikasikan. Evaluasi dilakukan dengan confusion matrix yang menghasilkan akurasi sebesar 75,94%.*

Kata Kunci: *Naïve Bayes*, Klasifikasi, *Email*, Spam

I. PENDAHULUAN

Email adalah cara yang efektif untuk berkomunikasi satu dengan lainnya (Teli dan Biradar, 2014). *Email* telah dijadikan alat untuk pertukaran informasi, untuk komersial dan sosial (Roy dkk, 2013). Beberapa tahun terakhir penggunaan *email* semakin meningkat dari 36% pada tahun 2002, 45% pada tahun 2003 menjadi 64% pada tahun 2004, 80% pada tahun 2006, 92% pada tahun 2009 dan 95 % pada tahun 2010. Tetapi pada tahun 2011 menurun menjadi 86% dan hanya 70% di tahun 2012 dan 2013 (Bajaj dan Peprzyk, 2014). Menurut Lavenstein dalam Bajaj dan Peprzyk (2014), terdapat 100 milyar email perhari untuk bisnis melalui *email*.

Seiring dengan pertumbuhan internet dan email, maka semakin banyak pertumbuhan spam beberapa tahun terakhir (Roy dkk, 2013). Saat ini spam membanjiri internet dengan mengirimkan salinan pesan-pesan yang sama agar pesan tersebut sampai kepada penerima (Sukardi dkk, 2014). Meningkatnya volume spam menjadi ancaman serius yang tidak hanya ke internet, tetapi juga ke masyarakat, untuk bisnis dan dibidang lainnya (Roy dkk, 2013). Spam *email* biasanya digunakan untuk menyebarkan virus atau kode berbahaya, penipuan dan iklan (Teli dan Biradar, 2014). Banyak para pemakai atau penerima merasa terganggu oleh banyaknya waktu yang dihabiskan untuk menghapus pesan spam, besarnya biaya yang harus dikeluarkan dan besarnya *bandwidth* jaringan (Sukardi dkk, 2014).

Untuk menghindari spam *email*, maka diperlukannya suatu metode dengan algoritma tertentu untuk memisahkan antara spam dan non spam (*legitimate mail*) secara efektif (Teli dan Biradar, 2014). Banyak algoritma yang tersedia, diantaranya algoritma *Decision Tree, Naïve Bayes*, *Support Vector Machine* (SVM), *Neural Network* dan lain-lain (Sukardi dkk, 2014). Pada penelitian ini algoritma yang digunakan yaitu *Naïve Bayes* merupakan metode yang paling popular dan paling banyak digunakan dalam pengklasifikasian, khusunya dalam penyaringan spam (Natalius, 2010). *Naïve Bayes* sangat baik digunakan untuk pengklasifikasian spam *email*, karena metode ini akan memeriksa semua *token* pada *body email* yang ada. Data yang digunakan mempengaruhi nilai keakuratan, semakin banyak data *email* maka semakin tinggi nilai akurasinya.

II. LANDASAN TEORI

Spam *Email*

Spam *email* yaitu *email* yang tidak diinginkan atau diminta oleh penerimanya (Teli dan Biradar, 2014). Ini merupakan ancaman untuk masalah keamanan, karena memungkinkan pengguna untuk masuk ke *link* atau situs palsu yang merugikan (Roy dkk, 2013).

Spam tampil dalam dua bagian yaitu *header email* dan *message content*/isi pesan (Roy dkk, 2013).

1. *Header Email*
   Menunjukkan *rute email* ketempat tujuan. *Header email* mengandung informasi tentang *email* seperti pengirim dan penerima, id pesan, tanggal dan waktu, subjek dan beberapa karakteristik *email* lainnya.
2. *Message Content*
   Isi pesan didalam spam menggunakan bahasa tertentu dalam *email*. Bahasa yang sering

digunakan seperti penawaran, klik di sini, lakukan sekarang dan masih banyak lainnya.

Tipe-tipe *email* spam (Sukardi dkk, 2014):

1. Untuk Iklan
   Digunakan untuk mempromosikan suatu produk ataupun layanan, mulai dari produk *software*, perumahan *real estate* hingga produk kesehatan dan produk vitamin.
2. Untuk mengirimkan *Malware* Untuk mendistribusikan virus dan *malware.*
3. *Phising*
   Spam ini bersembunyi di balik nama-nama besar perusahaan besar, lembaga keuangan, lembaga pemerintah, lembaga amal, para *phisher* mencoba memikat korban untuk mengunjungi *website* palsu dan dapat mencuri data keuangan pribadi atau informasi mengenai identitas korban.
4. *Scam*
   Berita elektronik dalam internet yang bersifat menipu sehingga pengirimnya dapat mendapatkan manfaat atau keuntungan.
5. Pesan yang tidak berarti
   Sebuah pesan yang dapat mengelabui teknologi spam *filter*, banyak pesan tak berarti yang dikirimkan tanpa tujuan yang jelas.

Klasifikasi

Klasifikasi merupakan sekumpulan model yang menggambarkan serta membedakan kelas-kelas data. Tujuannya model yang dihasilkan dapat digunakan untuk memprediksi kelas dari suatu data yang tidak mempunyai label kelas (sukardi dkk, 2014).

Klasifikasi salah satu metode dalam data mining yang dapat mengklasifikasikan *email* sebagai spam atau non-spam. Pengklasifikasian berdasarkan karateristik sebagai berikut (Sukardi dkk, 2014):

1. Alamat pengirim yang tidak benar.
2. Pemalsuan *header mail* untuk menyembunyikan *email* sesungguhnya sehingga akan sulit menetapkan sebagai spam atau non-spam.
3. Identitas penerima tidak nyata.
4. Alamat *email* yang berada dalam ‘To’ memiliki variasi alamat *email* penerima.
5. Isi subject tidak berhubungan sengan isi *email.*
6. Isi *email* memiliki sifat keragu-raguan.
7. *Unsubcribe* tidak bekerja pada spam *mail.*
8. Mengandung *script* tersembunyi.

Data Mining

Data mining yaitu serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui (Lindawati, 2008). *Data mining* digunakan untuk prediksi, uraian dan klasifikasi. *Data mining* adalah bagian integral dari penemuan pengetahuan dalam database yang memiliki urutan proses sebagai berikut (Aribowo, 2013):

1. *Data cleaning*
   Membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data.
2. *Data integration*
   Penggabungan atau mengkombinasikan sebuah data dari beberapa sumber.
3. *Data Selection*
   Pemilihan data dari sekumpulan data operasional sebelum penggalian informasi *Knowledge Discovery in Database* (KDD).
4. *Data Transformation*
   Process coding pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*.
5. *Data Mininig*
   Proses mencari pola atau informasi dari dalam data dengan menggunakan metode tertentu.
6. *Pattern Evaluation*
   Informasi atau pola yang dihasilkan *data mining* yang ditampilkan dalam bentuk yang mudah dimengerti.
7. *Knowledge Presentation*
   Visualisasi atau representasi hasil yang akan diberikan.

Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* merupakan metode terbaru yang digunakan untuk mengklasifikasikan. Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yaitu dengan memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya. Klasifikasi probabilitas yang menghitung satu set probabilitas dengan menghitung frekuensi dari nilai yang diberikan (Tina dkk, 2013). Pada dasarnya konsep dasar teori *bayes* yaitu peluang bersyarat P(X|C), dimana X adalah *posterior* dan C adalah *prior*. *Prior* adalah pengetahuan kita tentang karakteristik suatu parameter atau pengalaman dimasa lalu, sedangkan *posterior* adalah karakteristik yang akan diduga pada kejadian yang akan datang.

Metode *Naïve Bayes* dalam proses pengklasifikasian memiliki dua tahapan yaitu tahap pelatihan dan tahap klasifikasi. Untuk tahap pelatihan dilakukan proses analisis terhadap sampel berupa pemilihan *vocabulary* yaitu kata yang mungkin muncul dalam suatu set data yang dapat dijadikan representasi dokumen dan menentukan probabilitas *prior.* Pada tahap klasifikasi ditentukan nilai kategori berdasarkan *term* yang muncul dalam data yang diklasifikasi (Hamzah, 2012).

Klasifikasi menggunakan *Naïve Bayes* dengan probabilitas dikondisikan untuk mengenali *email* menjadi spam atau non *spam* (Teli dan Biradar, 2014). Tahapan Algoritm *Naïve Bayes* sebagai berikut (Markov & Daniel, 2007):

1. Probabilitas bersyarat/*likelihood*

$$P(x \mid C) = P(x_1, x_2, \ldots, x_n \mid C)$$

C = *Class*
X = *Vector* dari nilai atribut n
P(X|C) = Proporsi dokumen dari *class* C yang mengandung nilai atribut *X*

2. Probabilitas *prior* untuk tiap *class*

$$P(C) = \frac{Nj}{N}$$

Nj = Jumlah data pada suatu *class*
N = Jumlah total data

3. Probabilitas *posterior*

$$P(C|x) = \frac{P(x|C)\,P(C)}{P(x)}$$

P(C|X) = Probabilitas hipotesis X berdasarkan kondisi C
P(X|C) = Probabilitas X berdasarkan kondisi tersebut
P(C) = Probabilitas hipotesis C (probabilitas *prior*)
P(X) = Probabilitas dari X

Penentuan *class* dari suatu dokumen dilakukan dengan cara membandingkan nilai probabilitas suatu sampel berada di *class* yang satu dengan nilai probabilitas suatu sampel berada di class yang lain (Natalius, 2010).

Menentukan *class* yang cocok dari suatu sampel dilakukan dengan cara membandingkan nilai *posterior* untuk masing-masing *class*, dan mengambil *class* dengan nilai *posterior* yang tertinggi.

III. PEMBAHASAN

Data yang digunakan diperoleh dari *UCI Machine Learning Repository*. Data terdiri dari 4601, dimana 1813 (39,4%) adalah spam dan 2788 (60,6%) adalah non spam. Spam *email* terdapat 58 atribut dan 1 atribut target atau *class*, sebagai berikut:

1. 48 atribut bertipe *continuous* [0,100] yang beranggotakan kata terdiri dari:
   Make, Address, All, 3d, Our, Over, Remove, Internet, Order, Mail, Receive, Will, People, Report, Addresses, Free, Business, Email, You, Credit, Your, Font, 000, Money, Hp, Hpl, George, 650, Lab, Labs, Telnet, 857, Data, 415, 85, Technology, 1999, Parts, Pm, Direct, Cs, Meeting, Original, Project, Re, Edu, Table, Conference.
   Nilai presentase diperoleh dari:
   Jumlah kata yang muncul dalam *e-mail* x 100%
   ———————————————————
   Total keseluruhan kata dalam *e-mail*

2. 6 atribut bertipe *continous* [0,100] yang beranggotakan karakterterdiri dari:
   ; (| ! $ #
   Nilai presentase diperoleh dari:
   Jumlah karakter yang muncul dalam *e-mail* x 100%
   ———————————————————
   Total keseluruhan karakter dalam *e-mail*

3. 1 atribut bertipe continous real [1,…] yang berisi rata-rata huruf *capital*
4. 1 atribut bertipe continous real [1,…] yang berisi nilai terpanjang huruf *capital*
5. 1 atribut bertipe continous real [1,…] yang berisi jumlah huruf *capital*.

Menghitung Probabilitas *Prior*

Mencari nilai probabilitas *prior* berdasarkan data yang lalu. Total keseluruhan data 4601 dengan total data spam 1813 dan data non spam 2788 dengan perhitungan sebagai berikut:

P (*Spam*) = 1813/4601 = 0,394
P (Non *Spam*) = 2788.4601 = 0,606

Nilai probabilitas dari tiap *class* didapatkan untuk *Spam* nilai probabilitasnya 0,394 dan untuk non spam nilai probabilitasnya 0,606. Setelah menghitung probabilitas *prior*, kemudian menghitung nilai probabilitas *prior* dari setiap atribut dengan menggunakan algoritma *Naïve Bayes*. Contoh perhitungan digunakan dengan 6 atribut.

Tabel 1.Probabilitas *Prior*

Atribut		Jumlah Data	Spam	Non Spam	P (X|C)	
					Spam	Non Spam
Kata	*Address* (Ya)	3703	625	273	0.169	0.074
	Address (Tidak)	898	1188	2515	1.323	2.801
	Internet (Ya)	824	619	205	0.751	0.249
	Internet (Tidak)	3777	1194	2583	0.316	0.684
	Mail (Ya)	1302	827	475	0.635	0.365
	Mail (Tidak)	3299	986	2313	0.299	0.701
	Email (Ya)	1038	688	350	0.663	0.337
	Email (Tidak)	3563	1125	2438	0.316	0.684
	Money (Ya)	735	681	54	0.927	0.073
	Money (Tidak)	3866	1132	2734	0.293	0.707
	Project (Ya)	327	47	280	0.144	0.856
	Project (Tidak)	4274	1765	2508	0.413	0.587

Menghitung Probabilitas *Posterior*

Probabilitas *posterior* digunakan untuk menentukan *class* terhadap data baru, Berikut contoh dari probabilitas *posterior*.

Tabel 2 Probabilitas *Posterior*

Data X		P (X|C)	
Atribut	Nilai	Spam	Non Spam
Address	Tidak	1.323	2.801
Internet	Ya	0.751	0.249
Mail	Tidak	0.299	0.701
Email	Tidak	0.316	0.684

Money	Ya	0.927	0.073
Project	Ya	0.144	0.856

Dari data di atas, dimisalkan ada sebuah data baru yang terdapat suatu kata *internet, money* dan *project* tetapi tidak ada kata *address, Mail* dan *Internet*. Dari data atribut dapat diketahui nilai Spam dan Non *Spam* diperoleh dari probabilistik *prior*. Kemudian menghitung total keseluruhan probabilitas tiap *class*, sebagai berikut:

P(X|Spam) = P(*Addrees*|Spam) * P(*Internet*|Spam) * P(*Mail*|Spam) * P(*Email*|Spam) * P(*Money*|Spam) * P(*Project*|Spam)
= 1,323 * 0,751 * 0,299 * 0,316 * 0,927 * 0,144
= **0,0125**

P(X|Non Spam) = P(*Addrees*|Non Spam) * P(*Internet*|Non Spam) * P(*Mail*|Non Spam) * P(*Email*|Non Spam) * P(*Money*|Non Spam) * P(*Project*|Non Spam)
= 2,801 * 0,249 * 0,701 * 0,684 * 0,073 * 0,856
= **0,0209**

P(X|*Spam*)*P(*Spam*) = 0,0125 * 0,394
= **0,004925**

P(X|Non *Spam*)*P(Non *Spam*) = 0,0209 * 0,606
= **0,0126654**

Dari hasil di atas dapat disimpulkan bahwa P(X|Spam) lebih kecil dibandingkan dengan P(X|Non Span), maka dapat diketahui bahwa data ini termasuk data **Non Spam.**

Pengujian Algoritma Naïve Bayes

Estimasi akurasi dari klasifikasi merupakan hal penting untuk mengevaluasi seberapa akurat *classifier* yang digunakan untuk menguji sebuah data. Hasil berupa *confusion matrix.*

Confusion matrix adalah alat visualisasi yang biasa digunakan pada *supervised learning*. Tiap kolom pada matriks adalah contoh kelas prediksi, sedangkan tiap baris mewakili kejadian dikelas yang sebenarnya (Gorunescu, 2011).

Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy, sensitivity, specificity, ppv* dan *npv*. Pengukuran dengan *confusion matrix* menampilkan perbandingan dari hasil akurasi model *Naïve Bayes*.

Tabel 3*Confusion Matrix*

	True Spam	**True Non Spam**
Pred. Spam	1059	353
Pred. Non Spam	754	2435

Berdasarkan dari tabel diatas dari 2788 data non spam, ternyata 353 data diprediksi spam hasilnya spam, sedangkan 2435 sesuai dengan prediksi yaitu non spam. Sebaliknya untuk data spam sebanyak 1813, data sebanyak 1059 sesuai dengan prediksi yaitu spam, sedangkan untuk 754 yang di prediksi data spam ternyata tidak sesuai.

Berdasarkan *confusion matrix* maka dapat dihiutng jumlah

$$Accuracy = \frac{1059 + 2435}{1059 + 2435 + 754 + 353} = 0,7594$$

$$Sensitivity = \frac{1059}{1059 + 353} = 0,75$$

$$Specitivity = \frac{2435}{2435 + 754} = 0,7636$$

$$Ppv = \frac{1059}{1059 + 754} = 0,5841$$

$$Npv = \frac{2435}{2435 + 353} = 0,8734$$

IV. KESIMPULAN

Dari penelitian ini dapat ditarik kesimpulan yaitu:

1. Algoritma *Naïve Bayes* merupakan suatu metode klasifikasi yang berdasarkan teorema bayes yang terkenal dengan ilmu probabilitas.
2. Algoritma *Naïve Bayes* sangat baik untuk mendukung keputusan pengklasifikasian.

Agar penelitian ini dapat ditingkatkan, maka diperlukan sebuah saran yaitu:

1. Menambahkan jumlah data yang lebih besar sehingga hasil pengukuran dapat lebih baik lagi atau akurat.
2. Menambahkan *feature selection* seperti *Information Gain, Genetic Algorithm* dan sebagainya untuk mengurangi atribut yang digunakan sehingga atribut yang digunakan menjadi lebih sedikit.
3. Perlu dikembangkan lagi dengan metode algoritma yang lain seperti *Support Vector Machine, Neural Network* dan sebagainya.
4. Pembuatan GUI yang dapat diterapkan di perusahaan yang bergerak dibidang bisnis dengan alat *email.*

DAFTAR PUSTAKA

[1] Aribowo, A S. 2013. Metode *Data Mining* untuk klasifikasi kesetiaan pelanggan Terhadap Merek Produk.

[2] Bajaj, K., dan Pieprzyk, J. 2014. *A Case Study of User-Level Spam Filtering*.

[3] Gorunescu, F.2011. *Data Mining Concept Model Technique*

[4] Hamzah, A. 2012. Klasifikasi Teks dengan *Naïve Bayes Classifier* (NBC) Untuk Pengelompokan Teks Berita dan *Abstract* Akademis. ISSN: 1979-911X.

[5] Lindawati. 2008. *Data Mining* dengan Teknik *Clustering* dalam pengklasifikasian data mahasiswa Studi Kasus Prediksi Lama Studi Mahasiswa Universitas Bina Nusantara. ISSN: 19792328

[6] Markov, Z. Daniel, T. 2007. *Uncovering Patterns in.*

[7] Natalius, S. 2010. Metoda *Naïve Bayes Classifier* dan Penggunaanya pada Klasifikasi Dokumen.

[8] Patil, Tina R, Sherekar, S S. 2013. *Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification.* ISSN: 0974-1011

[9] Roy, S, dkk. 2013. *An Efficient Spam Filtering Techniques For Email Account*. e-ISSN: 2320-0847, p-ISSN: 2320-0936

[10] Sukardi, Syukur, ABd, dan Supriyanto, C. 2014. Klasifikasi Spam Email Menggunakan Algoritma C.45 Dengan Seleksi Fitur. ISSN: 1414-9999

[11] Teli, S P, Biradar, S. 2014. *Effective Email Classification for Spam and Non-Spam*. ISSN: 2277 128X

Biodata Penulis

Ratih Yulia H, memperoleh gelar Magister Ilmu Komputer (M.Kom) tahun 2015, Konsentrasi Ilmu Komputer pada STMIK Nusa Mandiri Jakarta.