Deteksi Spam Pada Email Berbasis Fitur Konten Menggunakan Naïve Bayes

Nur Qodariyah Fitriyah[1], Hardian Oktavianto[2], Hasbullah[3]

[1,2,3)]*Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember*
Email: [1]nurfitriyah@unmuhjember.ac.id , [2]hardian@unmuhjember.ac.id , [3]hasbulpretados@gmail.com

(Naskah masuk: 11 Oktober 2019, diterima untuk diterbitkan: 25 Oktober 2019)

ABSTRAK

Penelitian menunjukkan bahwa terdapat lebih dari 3 milyar akun *email* di dunia dengan frekuensi pengiriman *email* sekitar 205 – 294 milyar setiap hari. Salah satu masalah yang muncul dari pengiriman *email* yang luar biasa ini adalah adanya *spam email*. Salah satu solusi untuk mengatasi permasalahan *spam email* tersebut adalah dengan teknik penyaringan *spam email*. Penyaringan *spam email* dapat dilakukan dengan menggunakan pendekatan teori berbasis pembelajaran, yaitu dengan klasifikasi. Penelitian ini menerapkan algoritma Naive Bayes untuk melakukan klasifikasi *spam email* sehingga dari dataset *email*, akan dikelompokkan menjadi 2 yaitu *spam email* dan *non- spam email*. Hasil uji dengan menggunakan *k-fold cross validation* sebagai pembagian data latih dan data uji, menghasilkan kesimpulan bahwa nilai rata – rata data terklasifikasi benar adalah sebesar 3903, sedangkan nilai rata – rata data terklasifikasi salah adalah sebesar 698, rata – rata akurasi sebesar 84.8%, sedangkan rata – rata *precision* dan *recall* berturut – turut adalah 0.86 dan 0.85. Akurasi, *precision*, dan *recall* tertinggi diperoleh ketika menggunakan nilai k=9.

Kata kunci: *deteksi, klasifikasi, spam email, naive bayes*

ABSTRACT

Research shows that there are more than 3 billion email accounts in the world with a frequency of sending emails around 205 - 294 billion every day. One problem that arises from sending this extraordinary email is the existence of spam email. One solution to overcome the problem of email spam is by email spam filtering techniques. Email spam filtering can be done using a learning-based theory approach, namely classification. This study applies the Naive Bayes algorithm to classify email spam so that from the email dataset, it will be grouped into 2 namely spam email and non-spam email. The test results using k-fold cross validation as a division of training data and test data, resulting in the conclusion that the average value of correctly classified data is 3903, while the average value of classified data is 698, the average accuracy is 84.8% , while the average precision and recall are 0.86 and 0.85, respectively. The highest accuracy, precision, and recall are obtained when using the value k = 9.

Keywords: *detection, classification, spam email, naive bayes*

1. PENDAHULUAN

Email merupakan salah satu produk teknologi yang bertujuan untuk menggantikan media komunikasi melalui surat konvensional. Secara umum konsep penggunaan *email* sama seperti surat biasa akan tetapi dengan dukungan internet maka email memiliki beberapa keunggulan dibandingkan dengan surat biasa, seperti : penerima sebuah email dapat lebih dari 1 orang, waktu pengiriman yang lebih cepat, lebih hemat, serta dapat memuat informasi selain tulisan seperti : file dokumen teks atau *spreadsheet*, gambar, audio, dan video (Alurkar, et al., 2017) (Chandra, Indrawan, & Sukajaya, 2016).

Email dengan segala kelebihannya membawa konsep komunikasi dan pengelolaannya menjadi lebih mudah sehingga pengguna *email* semakin

meningkat dari hari ke hari (Mujtaba, Shuib, Raj, Majeed, & Al-Garadi, 2017). Penelitian menunjukkan bahwa terdapat lebih dari 3 milyar akun *email* di dunia dengan frekuensi pengiriman *email* sekitar 205 – 294 milyar setiap hari (Alurkar, et al., 2017) (Wijayanto & Takdir, 2014). Salah satu masalah yang muncul dari pengiriman email yang luar biasa ini adalah adanya *spam email* (Razak & Mohamad, 2013). *Spam email* adalah *email* yang tidak penting atau tidak berarti, yang dikirim oleh seseorang ke banyak penerima atau pengguna *email*, pada umumnya berisi promosi, informasi sampah, bahkan bisa juga berisi informasi penipuan (Chandra, Indrawan, & Sukajaya, 2016) (Vyas, Prajapati, & Gadhwal, 2015). Secara teknis, adanya *spam email* ini dapat menyebabkan semakin padatnya antrian dari *mail server* sehingga selain mengakibatkan pengiriman *email* tertunda juga dapat mengakibatkan server *crash* (Chandra, Indrawan, & Sukajaya, 2016) (Wijayanto & Takdir, 2014). Sedangkan dari sisi ekonomi dan produktivitas, diperkirakan setiap tahun biaya produktivitas per karyawan yang terbuang adalah sebesar $182,50, yang dihitung dari aktivitas karyawan untuk memilah atau menyaring *spam email* setiap hari (Safuan, Wahono, & Supriyanto, 2015).

Salah satu solusi untuk mengatasi permasalahan *spam email* tersebut adalah dengan teknik penyaringan *spam email*, yaitu berdasarkan kandungan isi atau konten *email*. Penyaringan *spam email* dapat dilakukan dengan menggunakan pendekatan teori berbasis pembelajaran, yaitu dengan klasifikasi berbasis fitur konten atau yang umum disebut juga ciri konten. Terdapat banyak algoritma yang dapat diterapkan untuk melakukan klasifikasi *spam email*, diantaranya adalah Naive Bayes, *Support Vector Machine*, *Artifial Neural Networking*, *Logistic Regression*, *K-Nearest Neigbors*, dan *Decission Tree* (Safuan, Wahono, & Supriyanto, 2015). Dari algoritma – algoritma klasifikasi tersebut, algoritma naive bayes adalah yang paling sering digunakan karena kesederhanan konsepnya. Naive bayes bekerja berdasarkan kemunculan fitur atau ciri terhadap suatu kelas yang kemudian dihitung probabilitasnya, sehingga sesuai dengan karakteristik fitur atau ciri *email* yang dijadikan acuan untuk membedakan mana yang termasuk *spam email* dan mana yang tidak (Chandra, Indrawan, & Sukajaya, 2016) (Rusland, Wahid, Kasim, & Hafit, 2017).

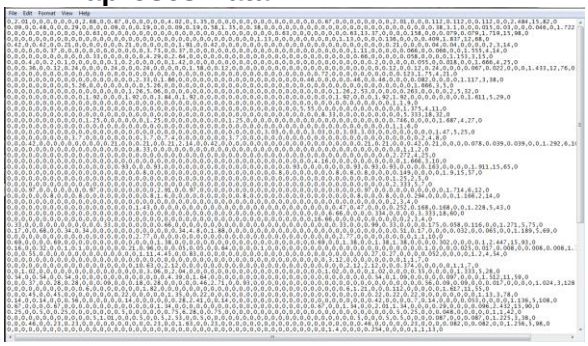
Penelitian ini akan menggunakan algoritma Naive Bayes untuk melakukan deteksi *spam email* berbasis fitur konten, sehingga dari dataset *email* akan dikelompokkan menjadi 2 yaitu *spam email* dan bukan *spam email*, yang selanjutnya akan dilakukan analisis terhadap hasil klasifikasi yang telah didapatkan, sedangkan dataset yang dipakai adalah mengambil dari UCI *Machine Learning Repository* yaitu dataset *spambase* (Hopkins, Reeber, Forman, & Suermondt, 2018).

2. METODE PENELITIAN

Tahapan penelitian kali ini dimulai dengan Studi Literatur, sebagai tahapan persiapan terhadap keseluruhan penelitian, tahapan ini bertujuan untuk mengumpulkan dan mempelajari dasar – dasar teori yang dipakai serta mencari sumber ilmiah yang mendukung, selain itu tahap studi literatur juga mencakup pengumpulan data. Tahap kedua adalah Praproses Data yaitu melakukan persiapan dataset sehingga siap untuk proses klasifikasi. Tahapan ketiga adalah proses Klasifikasi, dimana penerapan algoritma naive bayes dilakukan dengan menggunakan alat bantu WEKA. Tahap terakhir adalah analisis, yaitu penarikan kesimpulan terhadap hasil diperoleh. Gambar 1 mendeskripsikan tahapan penelitian yang akan dilakukan.

Gambar 1 Tahapan Penelitian

2.1 Praproses Data



Gambar 2 Screenshot Dataset Spamebase

Gambar 2 merupakan *screenshot* dari dataset yang akan digunakan. Dataset yang digunakan adalah dataset *spambase* yang diakses dari https://archive.ics.uci.edu/ml/machine-learning-databases/spambase/spambase.data. Dataset ini mempunyai 4601 *record* dengan rincian 1813 record termasuk kategori *spam email* sedangkan sisanya berjumlah 2788 *record non-spam*. Fitur atau variabel yang dimiliki dataset ini ada 58 buah.

Sebelum data digunakan untuk tahapan klasifikasi, dataset akan dipersiapkan terlebih dahulu. Dataset akan diperiksa apakah dari keseluruhan *record* tidak mengandung *missing value*, baik yang berupa tidak terdapat nilai atau ada dan tidaknya nilai yang tidak normal pada suatu fitur.

2.2 Klasifikasi

Klasifikasi pada penelitian ini menggunakan WEKA, sebuah perangkat lunak *open source* yang telah dikenal dan umum digunakan di bidang data mining dan *machine learning*. Pada WEKA terdapat banyak tersedia *library – library* algoritma, baik untuk klasifikasi, *clustering*, *association rule*, bahkan sampai *library* untuk praproses data dan *feature selection*. Klasifikasi nantinya akan menerapkan algoritma Naive Bayes, dengan skenario uji dataset menggunakan *k-fold cross validation*, yaitu membagi jumlah keseluruhan data menjadi k bagian, yang selanjutnya k-1 bagian digunakan sebagai data latih dan bagian sisa lainnya dijadikan data uji, sebagai contoh, apabila digunakan k=5 maka aka ada 5 bagian data yang sama banyak, kita sebut sebagai k1, k2, k3, k4, dan k5, uji pertama menggunakan k1 dengan data latih k2 sampai k5, kemudian bergiliran masing – masing k2 sampai k5 juga menjadi data uji. Adapun skenario uji nantinya akan menggunakan nilai k yang berbeda, hal ini dilakukan untuk mendapatkan hasil akurasi yang obyektif.

2.3 Analisis

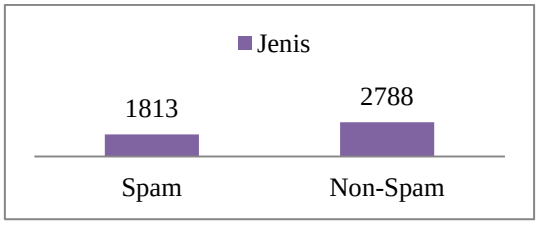
Pada tahap ini akan dilakukan analisis terhadap hasil dari uji pada tahap klasifikasi yang telah dilakukan sebelumnya. Analisis dilakukan agar mengetahui sejauh mana hasil klasifikasi berdasarkan skenario uji yang telah dirancang, yaitu membandingkan hasil – hasil yang diperoleh dari uji dengan menggunakan berbagai variasi nilai k pada *k-fold*. Beberapa nilai yang dibandingkan adalah : akurasi, baik untuk jumlah data yang diklasifikasi dengan benar maupun jumlah data yang diklasifikasi salah, *Precision*, dan *Recall*.

3. Hasil dan Pembahasan

3.1 Dataset

Dataset yang digunakan diambil dari UCI *Machine learning* Repository yaitu

dataset *spambase*, dataset ini mempunyai 58 fitur termasuk fitur label kelas, dan memiliki 4601 data dengan rincian 1813 data termasuk kategori *spam email* sedangkan sisanya berjumlah 2788 data dikategorikan sebagai *not spam email*. Dataset ini mempunyai 58 fitur, yang terdiri dari 57 fitur input dan 1 fitur output yaitu fitur label kelas. Pada fitur output, *spam* direpresentasikan dengan 1 dan *non-spam* direpresentasikan dengan 0.



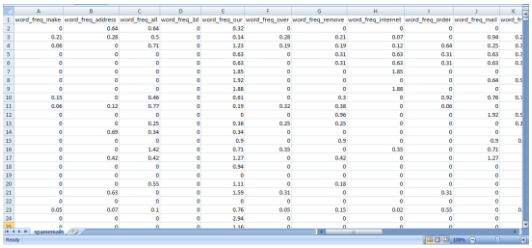
Gambar 3 Distribusi Jenis Dataset Berdasarkan Kelas

Fitur input yang merupakan atribut pembeda terdiri dari 57 buah fitur. 54 fitur merupakan persentase kemunculan kata kunci atau kemunculan karakter khusus dalam suatu pesan email. 3 fitur lainnya menunjukkan panjang rata – rata dari huruf kapital yang berurutan yang terdapat pada suatu *email*, panjang maksimal dari huruf kapital yang berurutan yang terdapat pada suatu *email*, serta jumlah huruf kapital yang terdapat pada suatu *email*. Pada tabel 1 disajikan sebagian dari deskripsi statistik terhadap dataset *spambase* yang dipakai pada penelitian ini.

Dataset *spambase* yang diperoleh mempunyai format teks seperti pada gambar 2, yang kemudian dilakukan konversi menjadi format csv sehinga siap dipakai untuk tahap selanjutnya. Pada gambar 4 dapat dilihat sebagian hasil dari konversi data ke format .csv.

Tabel 1 Deskripsi Statistik Dataset

Atr ke-	Atribut	Min:	Max:	Ave:	Std.Dev:
1	word_freq_make	0	4.54	0.10	0.31
2	word_freq_address	0	14.28	0.21	1.29
3	word_freq_all	0	5.1	0.28	0.50
4	word_freq_3d	0	42.81	0.07	1.40
5	word_freq_our	0	10	0.31	0.67
6	word_freq_over	0	5.88	0.10	0.27
7	word_freq_remove	0	7.27	0.11	0.39
8	word_freq_internet	0	11.11	0.11	0.40



Gambar 4 Format csv Dataset

3.2 Klasifikasi

Tahap klasifikasi dilakukan dengan menggunakan algoritma naïve bayes dengan *k-fold cross validation* untuk pembagian data latih dan data uji. Nilai k yang digunakan adalah mulai 2 sampai 10. Konsep *k-fold cross validation* adalah membagi total jumlah data menjadi k-bagian dan akan terjadi k-kali uji, sebagai contoh, apabila digunakan k=3 maka total data dibagi menjadi 3 bagian dan akan dilakukan 3 kali uji, masing – masing bagian bergantian menjadi data uji, dan sisa bagian lainnya menjadi data latih. Hasil akurasi diperoleh dengan menghitung rata – rata dari masing – masing uji. Tabel 2 mengilustrasikan pembagian jumlah data berdasarkan nilai k yang ditentukan.

Tabel 2. Pembagian Jumlah Data Berdasarkan Nilai K

k	k-bagian	jumlah uji
2	k1, k2	2
3	k1, k2, k3	3
4	k1, k2, k3, k4	4
5	k1, k2, k3, k4, k5	5
6	k1, k2, k3, k4, k5, k6	6
7	k1, k2, k3, k4, k5, k6, k7	7
8	k1, k2, k3, k4, k5, k6, k7, k8	8
9	k1, k2, k3, k4, k5, k6, k7, k8, k9	9
10	k1, k2, k3, k4, k5, k6, k7, k8, k9, k10	10

Deteksi *spam email* pada penelitian ini menggunakan algoritma naïve bayes. Secara konsep naïve bayes berbasis probabilitas suatu fitur dengan menghitung

p-ISSN : 2502-5724; e-ISSN : 2541-5735

frekuensi dan kombinasi terhadap label kelas. Rumus naïve bayes adalah sebagai berikut :

$$p(spam|word) = \frac{p(word|spam)p(spam)}{p(word)}$$

Pada penelitian ini fitur yang digunakan adalah berupa sekumpulan kata – kata yang telah menjadi fitur atau ciri khusus sebagai identifikasi *spam email*, selanjutnya frekuensi kemunculan masing – masing kata tersebut dihitung pada tahap pelatihan untuk membangun model klasifikasi *spam email*. Terdapat 57 fitur kata yang digunakan sebagai atribut pembeda diantara *spam email* dan yang bukan *spam email*. Setelah model dibangun maka selanjutnya adalah melakukan penghitungan peluang setiap data uji terhadap masing – masing kelas target, yaitu kelas *spam* dan kelas *non-spam*, nilai peluang terbesar terhadap suatu kelas menentukan termasuk kelas manakah data tersebut. Tabel 3 menunjukkan hasil rekapitulasi uji klasifikasi.

Dari tabel 3 dapat dilhat bahwa dari uji dengan k=2 sampai dengan k=10 menghasilkan rentang nilai yang tidak terlalu berbeda, baik untuk jumlah terklasifikasi benar, akurasi, *precision*, dan *recall*.

3.3 Analisis

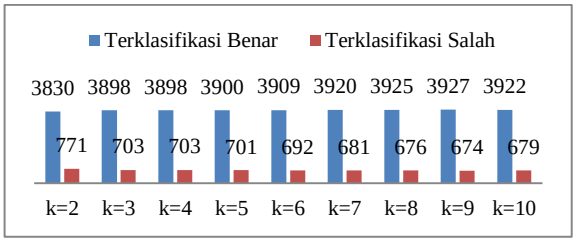
Pada uji coba yang telah dilakukan, maka seperti yang telah dideskripsikan pada tabel 3, hasil terklasifikasi benar mempunyai rata – rata sebesar 3903, sedangkan hasil terklasifikasi salah mempunyai rata –rata sebesar 698. Nilai rata – rata akurasi sebesar 84.8%, sedangkan rata – rata *precision* dan *recall* berturut – turut adalah 0.86 dan 0.85.

Hasil data terklasifikasi benar mengalami kenaikan berbanding lurus dengan kenaikan nilai k pada *k-fold cross validation*, begitu juga dengan data terklasifikasi salah yang mengalami penurunan ketika nilai k semakin bertambah. Gambar 5 menampilkan grafik data terklasifikasi benar dan grafik data terklasifikasi salah.

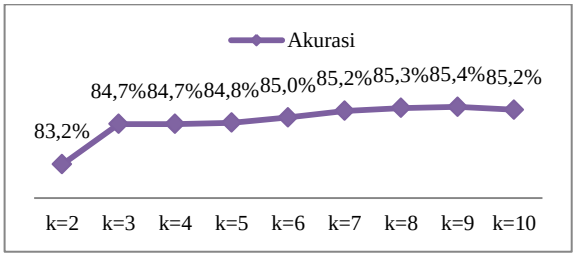
Nilai akurasi juga mengalami kenaikan sebanding dengan kenaikan nilai k yang dipakai. Akurasi tertinggi diperoleh ketika memakai k=9 yaitu sebesar 85.4%. Gambar 6 menyajikan grafik nilai akurasi yang diperoleh selama ujicoba

Tabel 3. Hasil Rekapitulasi Uji

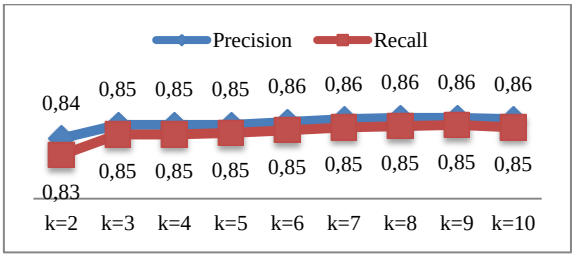
K	Terklasifikasi Benar	Terklasifikasi Salah	% Akurasi	*Precision*	*Recall*
2	3830	771	83.2%	0.84	0.83
3	3898	703	84.7%	0.85	0.85
4	3898	703	84.7%	0.85	0.85
5	3900	701	84.8%	0.85	0.85
6	3909	692	85.0%	0.86	0.85
7	3920	681	85.2%	0.86	0.85
8	3925	676	85.3%	0.86	0.85
9	3927	674	85.4%	0.86	0.85
10	3922	679	85.2%	0.86	0.85
Rata rata	**3903**	**698**	**84.8%**	**0.86**	**0.85**



Gambar 5 Grafik Data Terklasifikasi Benar dan Data Terklasifikasi Salah



Gambar 6. Grafik Distribusi Nilai Akurasi



Gambar 7. Grafik Data Terklasifikasi Benar dan Data Terklasifikasi Salah

Untuk tingkat *precision* dan *recall* juga mengalami kenaikan sebanding dengan kenaikan nilai k. *precision* Gambar 7 menunjukkan grafik nilai *precision* dan *recall* pada penelitian ini.

Hasil analisis yang diperoleh adalah, penerapan algoritma naïve bayes untuk deteksi *spam email* dapat melakukan klasifikasi dengan rata – rata akurasi sebesar 84.8%, rata – rata nilai *precision* 0.86, dan rata – rata *recall* sebesar 0.85. Akurasi, *precision*, dan *recall* tertinggi diperoleh ketika menggunakan nilai k=9.

4. KESIMPULAN

Kesimpulan yang diperoleh pada penelitian ini adalah, hasil uji dengan menggunakan k=2 sampai dengan k=10 menghasilkan rentang nilai yang tidak terlalu berbeda, baik untuk jumlah terklasifikasi benar, akurasi, *precision*, dan *recall*. Nilai rata – rata data terklasifikasi benar adalah sebesar 3903, sedangkan nilai rata – rata data terklasifikasi salah adalah sebesar 698. Nilai rata – rata akurasi sebesar 84.8%, sedangkan rata – rata *precision* dan *recall* berturut – turut adalah 0.86 dan 0.85. Akurasi, *precision*, dan *recall* tertinggi diperoleh ketika menggunakan nilai k=9.

Saran terhadap penelitian ini adalah Pengembangan dapat dilakukan untuk meningkatkan performa hasil uji dengan cara melakukan seleksi fitur sebelum dilakukan klasifikasiMelakukan perbandingan algoritma naïve bayes dengan algoritma klasifikasi yang lain, sehingga dapat diketahui mana algoritma yang paling baik untuk dataset *spambase*.

DAFTAR PUSTAKA

ALURKAR, A. A., RANADE, S. B., JOSHI, S. V., RANADE, S. S., SONEWAR, P. A., MAHALLE, P. N., & DESHPANDE, A. V. 2017. A Proposed Data Science Approach for Email Spam Classification using Machine Learning Techniques. *Internet of Things Business Models, Users, and Networks* (pp. 1-5). Copenhagen: IEEE.

CHANDRA, W. N., INDRAWAN, G., & SUKAJAYA, I. N. 2016. Spam Filtering Dengan Metode Pos Tagger Dan Klasifikasi Naïve Bayes. *Jurnal Ilmiah Teknologi dan Informasia ASIA, X*(1), 47-55.

HOPKINS, M., REEBER, E., FORMAN, G., & SUERMONDT, J. 2018. *Spambase Dataset.* Retrieved from UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/machine-learning-databases/spambase/

MUJTABA, G., SHUIB, L., RAJ, R. G., MAJEED, N., & AL-GARADI, M. A. 2017. Email Classification Research Trends: Review and Open Issues. *IEEE Access*, 9044-9064.

RAZAK, S. B., & MOHAMAD, A. F. 2013. Identification of *Spam email* Based on Information from Email Header. *13th International Conference on Intellient Systems Design and Applications* (pp. 347-353). Bangi: IEEE.

RUSLAND, N. F., WAHID, N., KASIM, S., & HAFIT, H. 2017. IOP Conference Series: Materials Science and Engineering. *International Research and Innovation Summit (IRIS2017). 226*, p. 012091. Melaka, Malaysia: IOP Publishing.

SAFUAN, WAHONO, R. S., & SUPRIYANTO, C. 2015. Penanganan Fitur Kontinyu dengan Feature Discretization Berbasis Expectation Maximization Clustering untuk Klasifikasi *Spam email* Menggunakan Algoritma ID3. *Journal of Intelligent Systems, I*(2),

148-155. Retrieved from http://journal.ilmukomputer.org

VYAS, T., PRAJAPATI, P., & GADHWAL, S. 2015. A Survey and Evaluation of Supervised Machine Learning Techniques for Spam e-mail Filtering. *IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-7). COIMBATORE, TAMIL NADU, INDIA: IEEE.

WIJAYANTO, A. W., & TAKDIR. 2014. Fighting Cyber Crime in Email Spamming: An Evaluation of Fuzzy Clustering Approach to Classify Spam Messages. *International Conference on Information Technology Systems and Innovation (ICITSI)* (pp. 19-24). Bandung-Bali: IEEE.