Jurnal Informatika Universitas Pamulang **ISSN: 2541-1004**
Penerbit: Program Studi Teknik Informatika Universitas Pamulang **e-ISSN: 2622-4615**
Vol. 7, No. 2, Juni 2022 (369-373) **10.32493/informatika.v7i2.16411**

Deteksi Pesan Spam pada Forum Daring Menggunakan Metode Naïve Bayes

Ruby Chandra[1], Davin Christian Juan[2]

[1,2]Teknik Informatika, Universitas Bunda Mulia, Jl. Jalur Sutera Barat Kav 7-9 Panunggangan Timur, Tangerang, 15143
e-mail: [1]rubychandra32@gmail.com, [2]s32180045@student.ubm.ac.id

Submitted Date: December 25th, 2021 Reviewed Date: January 08th, 2022
Revised Date: January 16th, 2022 Accepted Date: August 16th, 2022

Abstract

Online forum discussion is one of the learning services provided by University XYZ that chan be used by lectures and students. This forum will be used as a means to issue student opinions regarding lessons as well to collect attendance data for these students. In practice, student often only provide responses that should not be uploaded on forums in the form of spam messages. Therefore the main function of online forums as learning media and a means to express opinions is ignored. This study will create a spam detection system using Naïve Bayes Classifier (NBC) as an algorithm to categorize the contents of discussion messages where it will be labelled as a spam or non-spam. The result of the system test by dividing the dataset into 80 data training and 20 data testing, conclude that the system accuracy value is 95%, system precision value is 100% and system racall value is 90%. From the result above, NBC algorithm can be used as an option in detecting spam messages on forums.

Keywords: Naïve Bayes Classifier, Online Discussion Forum, Spam.

Abstrak

Forum diskusi daring adalah salah satu layanan edukasi yang disediakan oleh pihak Universitas XYZ yang dapat digunakan oleh dosen dan mahasiswa. Forum diskusi ini dijadikan sebagai sarana untuk mengeluarkan pendapat mahasiswa mengenai pelajaran serta sebagai pendataan absensi kehadiran mahasiswa tersebut. Pada prakteknya sering sekali mahasiswa hanya memberikan tanggapan yang seharusnya tidak diunggah pada forum diskusi berupa pesan spam sehingga fungsi utama dari forum daring sebagai media pembelajaran dan sarana untuk mengeluarkan pendapat menjadi terkesampingkan. Pada penelitian ini akan dibuat sistem deteksi pesan spam menggunakan algoritma *Naïve Bayes Classifier* (NBC) untuk mengklasifikasi pesan diskusi berdasarkan konten isi pesan lalu diberikan label sebagai spam atau non-spam. Hasil pengujian sistem dimana dataset yang digunakan dibagi menjadi 80 data *training* dan 20 data *testing*, mendapatkan hasil nilai *accuracy* sistem sebesar 95%, nilai *precision* sistem sebesar 100%, dan nilai *recall* sistem 90%. Berdasarkan hasil uji sistem, algoritma NBC dapat dijadikan sebagai salah satu pilihan dalam mendeteksi pesan spam pada forum.

Kata kunci: Naïve Bayes Classifier, Forum Diskusi Online, Spam

1. Pendahuluan

Universitas XYZ adalah institusi pendidikan swasta di Indonesia yang berlokasi di Jakarta Pusat, Indonesia. Universitas XYZ memiliki dua metode pembelajaran tatap muka dan forum diskusi daring. Forum merupakan sebuah lembaga atau badan atau wadah sebagai tempat bertukar pikiran secara bebas (KBBI, 2016). Universitas XYZ menggunakan metode forum yang dapat memungkinkan mahasiswa menghadiri sesi perkuliahan tanpa perlu datang ke kampus. Forum diskusi daring ini berfungsi sebagai sarana untuk mengeluarkan pendapat mahasiswa mengenai materi perkuliahan (Kurniawan et al., 2016). Selain itu, forum diskusi daring digunakan juga sebagai pendataan absensi mahasiswa. Pada prakteknya, sering sekali mahasiswa hanya memberikan tanggapan yang seharusnya tidak

diunggah pada forum diskusi seperti hanya *reply* kata terima kasih atau paragraf yang tidak memiliki keterkaitan dengan topik mata kuliah yang sedang dibahas agar memenuhi syarat absensi tersebut. Postingan reply ini termasuk spam yang menyebabkan fungsi utama dari forum diskusi daring sebagai media pembelajaran dan sarana untuk mengeluarkan pendapat menjadi terkesampingkan.

Pesan spam merupakan pesan yang tidak diinginkan dan tidak bermakna yang dikirim oleh seorang oknum untuk banyak orang. Pada umumnya pesan tersebut berisikan tentang sebuah promosi, informasi sampah, dan bahkan dapat berisi tentang penipuan (Fitriyah & Oktavianto, 2019). Adanya pesan spam juga berdampak negatif pada efisiensi waktu pengguna karena pada saat berdiskusi pada forum diskusi daring harus memilih kembali pesan yang termasuk kedalam kategori spam dan non-spam (Juang, 2016). Jika tidak ditangani dengan benar, pesan spam dapat mengganggu sesi diskusi pada forum serta peserta yang terlibat dalam sesi tersebut (Thoib et al., 2018).

Adapun penyelesaian dari permasalahan ini dengan memanfaatkan teknologi data mining beserta algoritma klasifikasi. Salah satu algoritma terbaik yang dapat melakukan klasifikasi sesuai kebutuhan pengguna adalah algoritma *Naive Bayes Classifier* (NBC). Algoritma NBC merupakan salah satu teknik pengklasifikasi berbasis peluang dan statistik, dimana metode pengklasifikasi ini dapat memprediksi sebuah peluang keanggotaan dari sebuah kelas data berdasarkan data kelas yang sebelumnya sudah ada (Bajabir, 2018). Algoritma ini juga memiliki hasil tingkat akurasi cukup tinggi (Dewi, 2016). Keuntungan dalam menggunakan metode NBC yaitu tidak perlu membutuhkan dataset yang banyak untuk dapat menentukan perkiraan tolak ukur yang diperlukan dalam proses klasifikasi (Hardianti et al., 2018). Berdasarkan permasalahan tersebut, peneliti bertujuan membuat sebuah sistem deteksi pesan spam pada forum dengan menggunakan algoritma NBC.

Sistem ini akan mengklasifikasi pesan menjadi pesan yang mengandung unsur spam dengan pesan non-spam. Data pesan yang didapatkan dari forum akan dikelompokkan menjadi data latih untuk proses pelatihan sistem dan data uji untuk mengetahui kemampuan sistem dalam melakukan klasifikasi. Data tersebut akan diolah pada tahap *text processing* agar data lebih terstruktur. Setelah melalui pengolahan data, proses selanjutnya adalah proses pelatihan data pada sistem agar dapat mengetahui dasar atau indikasi pesan yang mengandung unsur spam. Sistem deteksi ini diharapkan dapat berguna untuk Universitas XYZ agar tidak adanya pesan spam yang dilakukan oleh mahasiswa, sehingga mahasiswa dapat berdiskusi sesuai dengan topik mata kuliah yang sedang dibahas. Dengan begitu mahasiswa dapat lebih fokus dalam berdiskusi tentang materi kuliah yang sedang dibahas.

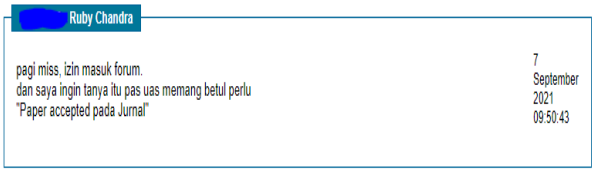
2. Metode Penelitian

Berikut tahapan-tahapan yang dilaksanakan penulis terhadap penelitian ini dapat diamati pada Gambar 1.

Gambar 1. Model alur penelitian yang digunakan.

2.1 Pengumpulan Data

Data penelitian diperoleh dari forum diskusi daring yang dikhususkan untuk mahasiswa semester 7 program studi teknik informatika Universitas XYZ tahun ajaran 2021/2022. Jumlah data yang akan diperoleh sebanyak 100 pesan, dimana 50 pesan diberi label sebagai pesan spam dan 50 pesan diberi label sebagai pesan non-spam Contoh pesan pada forum yang dapat dilihat pada Gambar 2.



Gambar 2. Contoh Pesan pada Forum

2.2 Pengolahan Data (*Pre-Processing*)

Setelah dilakukan proses pengumpulan data, tahap selanjutnya adalah *pre-processing*. Tahap ini merupakan serangkaian proses untuk melakukan

optimalisasi dataset yang dimiliki sehingga hasil yang didapatkan pada saat proses klasifikasi lebih tepat.

Secara umum, tahap *pre-processing* terdiri dari 3 proses, yaitu (Juang, 2016):

a. *Case Folding* dan *Cleansing*.

Proses mengubah karakter huruf besar menjadi karakter huruf kecil pada kalimat, dan menghapus karakter selain karakter alfabet. Contoh proses *case folding* dan *cleansing* dapat dilihat pada Tabel 1.

Tabel 1. *Case folding* dan *Cleansing*.

Input	**Output**
Selamat Pagi bapak dosen dan kawan - kawan.	selamat pagi bapak dosen dan kawan kawan

Pada Tabel 1 dapat dilihat setiap karakter huruf besar diubah menjadi huruf kecil (Selamat → selamat), lalu karakter selain alfabet akan dihapus.

b. *Tokenizing / Parsing*.

Proses pemotongan atau pemecahan kalimat menjadi potongan-potongan kata tunggal (*token*). Contoh mengenai proses *tokenizing* dapat dilihat pada Tabel 2.

Tabel 2. *Tokenization Process*.

Input	**Output**
selamat pagi bapak dosen dan kawan kawan	“selamat”, “pagi”, “bapak”, “dosen”, “dan”, “kawan”, “kawan”

Pada Tabel 2 menunjukan pemecahan sebuah kalimat menjadi potongan-potongan kata tunggal.

c. *Filtering*.

Proses untuk mendapatkan elemen kata yang penting dengan menghilangkan elemen kata yang tidak bermakna. Contoh mengenai proses *filtering* dapat dilihat pada Tabel 3.

Tabel 3. *Filtering Process*

Input	**Output**
“selamat”, “pagi”, “bapak”, “dosen”, “dan”, “kawan”, “kawan”	“selamat”, “pagi”, “bapak”, “dosen”, “kawan”, “kawan”

Pada Tabel 3, dapat dilihat adanya kata yang dihapus (“dan”) karena terdeteksi merupakan kata tidak bermakna penting.

2.3. Klasifikasi

Klasifikasi adalah proses pemberian kategori suatu data yang tidak diketahui kategorinya kedalam suatu kelompok data yang sudah ada sebelumnya. Dikarenakan jumlah data yang semakin lama semakin banyak, maka diperlukanlah sebuah metode yang dapat mengklasifikasikan objek (data) tersebut.

Algoritma NBC merupakan salah satu teknik pengklasifikasi berbasis peluang dan statistik, dimana metode pengklasifikasi ini dapat memprediksi sebuah peluang keanggotaan dari sebuah kelas data berdasarkan data kelas yang sebelumnya sudah ada (Bajabir, 2018). Persamaan dari teorema Bayes dapat dilihat pada persamaan (1).

$$P(A \mid B) = \frac{P(B \mid A) \,.\, P(A)}{P(B)} \quad (1)$$

Keterangan:
$A \; dan \; B$ = Dua kejadian atau kondisi tertentu.
$P(A \mid B)$ = Probabilitas terjadinya A jika B bernilai benar.
$P(A)$ = Probabilitas terjadinya A.
$P(B \mid A)$ = Probabilitas terjadinya B jika A bernilai benar.
$P(B)$ = Probabilitas terjadinya B.

2.4. Uji Performansi

Pada tahap ini, penulis akan menggunakan teknik *split validation* dengan *confusion matrix* untuk mengetahui hasil performa dari sistem prediksi ini (Annur, 2018). Bentuk umum dari tabel *confusion matrix* dapat dilihat pada Tabel 4.

Tabel 4. tabel *Confusion Matrix*

Total Record	***Predicted Positive***	***Predicted Negative***
Actual Positive	*True Positive* (**TP**)	*False Negative* (**FN**)
Actual Negative	*False Positive* (**FP**)	*True Negative* (**TN**)

Keterangan:
TP= Banyaknya data yang berhasil diprediksi dengan benar sebagai data aktual positif.

FP= Banyaknya data yang gagal diprediksi dengan benar sebagai data aktual positif dimana sebenarnya merupakan aktual negatif.

TN= Banyaknya data yang berhasil diprediksi dengan benar sebagai data aktual negatif.

FN= Banyaknya data yang gagal diprediksi dengan benar sebagai data aktual negatif dimana sebenarnya merupakan aktual positif.

Setelah terbentuk tabel *confusion matrix*, tahap selanjutnya data dari tabel dapat diolah menggunakan persamaan *accuracy* untuk menghitung rasio tingkat kedekatan antara nilai yang didapat dengan nilai yang sebenarnya, persamaan *precision* untuk menghitung tingkat kecocokan antara bagian data yang diambil dengan informasi yang dibutuhkan, dan persamaan *recall* untuk menghitung tingkat keberhasilan sistem dalam memperoleh kembali informasi yang relevan (Armando, 2017). Persamaan *accuracy*, *precision*, dan *recall* dapat dilihat pada persamaan (2), (3), dan (4).

$$Accuracy = \frac{TP + TN}{Total\ Record} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Di mana:
TP= Nilai pada True Positif.
TN= Nilai pada True Negatif.
FP= Nilai pada False Positive.
FN= Nilai pada False Negative.
Total Record = Jumlah Keseluruhan Data.

3. Hasil dan Pembahasan

Data yang diperoleh secara manual pada forum diskusi online dijadikan menjadi dua dataset yaitu dataset untuk *training* yang terdiri dari 80 pesan (40 spam dan 40 non-spam) dan dataset untuk *testing* yang terdiri dari 20 pesan (10 spam dan 10 non-spam).

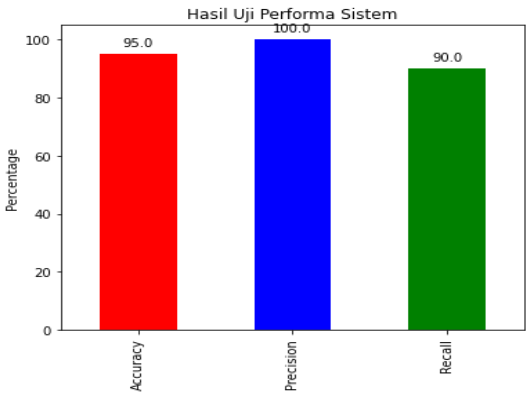
Hasil dari proses klasifikasi mengunakan metode *naïve bayes* terhadap data testing dapat diperhatikan pada Gambar 3.

Gambar 3 memaparkan bahwa berdasarkan 20 *record* yang diuji, didapati 9 data uji benar diprediksi sebagai pesan spam (*True Positive*), 10 data uji benar diprediksi sebagai pesan non-spam (*True Negative*), 1 data uji salah diprediksi sebagai pesan non-spam yang dimana pesan tersebut merupakan pesan spam (*False Negative*), dan tidak ada data uji yang salah diprediksi sebagai pesan spam yang dimana pesan tersebut merupakan pesan non-spam (*False Positive*).

	pesan	predicted_class	real_class
0	modul ada tipe error yaitu lost pdu damaged ...	non_spam	non_spam
1	untuk saat ini ngumpulin data baru sekitar pe...	non_spam	non_spam
2	baru sampai tahap pengumpulan jumal bu untuk ...	non_spam	non_spam
3	data nya dari kuesioner ya	non_spam	non_spam
4	kalau pengumpulan jurnal berarti masuk ke dala...	non_spam	non_spam
5	miss kalau ada data olah apakah nanti proses ...	non_spam	non_spam
6	untuk kesulitannya mungkin pada tahap analisis...	non_spam	non_spam
7	analisisnya secara manual ya nik gak bisa paka...	non_spam	non_spam
8	pagi miss untuk pertemuan ini apa akan bahas	non_spam	non_spam
9	kayaknya tentang bab hasil penelitian pembahasan	non_spam	non_spam
10	oke terima kasih pak	spam	spam
11	selamat pagi bu izin masuki forum	spam	spam
12	mantap	spam	spam
13	ijin masuk mis	spam	spam
14	pagi miss izin masuk	spam	spam
15	terima kasih	spam	spam
16	sudah pak	non_spam	spam
17	oke baik pak	spam	spam
18	pagi menjelang siang pak baik pak	spam	spam
19	selamat pagi bu izin masuk forum	spam	spam

Gambar 3. Hasil Klasifikasi Data *Testing*.

Kesalahan dalam memprediksi kelas tersebut memiliki dampak terhadap nilai performa dari sistem prediksi pesan spam ini, dimana nilai *accuracy* sistem yang diperoleh sebesar 95%, nilai *precision* sistem sebesar 100%, dan nilai *recall* sistem sebesar 90%. Hasil uji performa disajikan pada Gambar 4.



Gambar 4. Hasil Pengujian Performa Sistem.

Dari penelitian yang dilakukan oleh Antonius dan Yuan Lukito mengenai deteksi konten spam pada kolom komentar Instagram menggunakan Naïve Bayes mendapatkan nilai

accuracy sebesar 74.31%, nilai *precision* sebesar 96.71%, dan *recall* sebesar 61.43% (C & Lukito, 2017). Hasil performa sistem dari penelitian yang didapatkan tersebut berbanding lurus dengan penggunaan dataset yang terbilang cukup besar dimana jumlah keseluruhan data yang digunakan sebesar 25000 data, hal ini mengakibatkan pola komentar spam lebih beragam sehingga sistem deteksi komentar spam kurang optimal. Sedangkan pada penelitian yang dilakukan oleh penulis, khususnya pesan pada forum diskusi rata-rata pesan spam memiliki pola yang serupa sehingga memungkinkan sistem dapat memprediksi pesan spam lebih optimal dan mendapatkan hasil uji performa sistem yang memuaskan.

4. Kesimpulan

Forum diskusi daring merupakan sarana media pembelajaran yang diterapkan oleh Universitas XYZ. Fungsi utama dari forum ini sebagai media berpendapat mahasiswa dan juga dijadikan penanda absensi mahasiswa, namun pada praktek nyatanya rata-rata mahasiswa melakukan *reply* pesan spam hanya untuk mengisi daftar absensi kelas sehingga fungsi utama dari forum terkesampingkan. Berdasarkan penelitian yang telah dilakukan untuk menangani permasalahan tersebut, dengan menggunakan data sebanyak 100 pesan yang berasal dari situs forum daring yang dikhususkan untuk mahasiswa program studi teknik informatika Universitas XYZ tahun ajaran 2021/2022. Peneliti mendapatkan temuan dimana klasifikasi pesan spam menggunakan algoritma *Naive Bayes Classifier* (NBC) memperoleh nilai *accuracy* sistem sebesar 95%, nilai *precision* sistem sebesar 100%, dan nilai *recall* sistem sebesar 90%. Penggunaan algoritma NBC dapat dijadikan salah satu alternatif dalam mendeteksi pesan spam pada forum.

5. Saran

Pada penelitian selanjutnya akan sebaiknya dapat menggunakan dataset forum yang berasal dari program studi Universitas XYZ untuk melihat seberapa besar pengaruh penggunaan dataset tersebut pada nilai performa sistem deteksi pesan spam. Selain itu, bisa juga dapat dilakukan penelitian mengenai penggunaan algoritma klasifikasi yang lain untuk mengetahui apakah algoritma-algoritma klasifikasi tersebut dapat mengimbangi hasil uji performa dari penelitian ini.

Referensi

Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, *10*(2), 160–165. https://doi.org/10.33096/ilkom.v10i2.303.160-165

Armando, V. (2017). Sistem Rekomendasi Pembelian Telepon Genggam Dengan Metode Content-Based Filtering [Universitas Atma Jaya Yogyakarta]. In *http://e-journal.uajy.ac.id/11794/4/TF070093.pdf*. http://e-journal.uajy.ac.id/7244/4/3TF03686.pdf

Bajabir, A. Z. A. M. (2018). *Penerapan metode naive bayes untuk prediksi menentukan karyawan tetap pada pt. ysp industries indonesia*.

C, A. R., & Lukito, Y. (2017). Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes. *Jurnal ULTIMATICS*, *9*(1), 50–58. https://doi.org/10.31937/ti.v9i1.564

Dewi, S. (2016). Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan. *Techno Nusa Mandiri*, *XIII*(1), 60–66.

Fitriyah, N. Q., & Oktavianto, H. (2019). *Deteksi Spam Pada Email Berbasis Fitur Konten Menggunakan Naïve Bayes* (pp. 1–7).

Hardianti, A. T., Manga, A. R., & Darwis, H. (2018). *Penerapan Metode Naïve Bayes pada Klasifikasi Judul Jurnal* (Vol. 3, Issue 2, pp. 97–101).

Juang, D. (2016). Analisis Spam dengan Menggunakan Naïve Bayes. *Jurnal Teknovasi*, *3*(2), 51–57.

KBBI. (2016). *Forum*. https://kbbi.kemdikbud.go.id/entri/forum

Kurniawan, W., Suprianto, A., & Sumardiyono, B. (2016). Rancangan Sistem Forum Diskusi Online Untuk Program Studi Sistem Informasi Antara Dosen Dan Mahasiswa. *Jurnal Rekayasa Informasi*, *5*(2), 43–51.

Thoib, I., Setyanto, A., & Raharjo, S. (2018). Pengaruh Normalisasi Teks Dengan Text Expansion Dalam Deteksi Komentar Spam Pada Youtube. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *2*(3), 708–715. https://doi.org/10.29207/resti.v2i3.602