Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 873−880
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5070

Penerapan Metode Naive Bayes Dalam Klasifikasi Spam SMS Menggunakan Fitur Teks Untuk Mengatasi Ancaman Pada Pengguna

Fathimah Noer Azzahra*, Tatang Rohana, Rahmat, Ayu Ratna Juwita

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang, Karawang
Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia
Email: [1,*]if20.fathimahazzahra@mhs.ubpkarawang.ac.id, [2]tatangubp@ubpkarawang.ac.id, [3]rahmat@ubpkarawang.ac.id, [4]ayurj@ubpkarawang.ac.id
Email Penulis Korespondensi: if20.fathimahazzahra@mhs.ubpkarawang.ac.id
Submitted: **05/04/2024;** Accepted: **30/04/2024;** Published: **30/04/2024**

Abstrak–Salah satu dampak negatif dari kemajuan digital saat ini adalah meningkatnya jumlah SMS spam. SMS spam menimbulkan risiko keamanan bagi pengguna karena dapat berisi tautan berbahaya atau permintaan informasi pribadi yang digunakan untuk serangan malware, smishing, atau penipuan. Namun, dengan berbagai upaya perlindungan yang tersedia tidak semua spam SMS dapat diklasifikasi dan dicegah secara efektif. Namun permasalahan tersebut dapat diminimalisir dengan membuat sebuah model anti spam SMS yang bertujuan dapat mengklasifikasikan jenis SMS. Sehingga penelitian ini bertujuan untuk mengklasifikasi jenis SMS yang mengandung spam dan ham dengan menerapkan algoritma Naïve Bayes. Pada penelitian ini dataset berjumlah 5572 record yang terdiri dari 2 kategori, yaitu spam dan ham. Algoritma ini mampu menunjukkan kinerja yang memuaskan dalam membedakan pesan spam dan ham karena menurut tinjauan literatur Algoritma Naïve Bayes cocok digunakan untuk dataset berbahasa Inggris. Evaluasi model menampilkan hasil yang baik dengan akurasi mencapai 93.2%, presisi 93.7%, recall 93.2%, dan F1-score 91.6%. Selain itu, analisis dalam penelitian menggunakan Kurva Receiver Operating Characteristic (ROC) menunjukkan tingkat akurasi sebesar 97.3%, menandakan bahwa model memiliki performa yang sangat baik dalam klasifikasi spam pada pesan SMS. Meskipun demikian, masih ada ruang untuk peningkatan melalui penggunaan metode baru dan dataset yang lebih besar dan beragam. Penelitian ini memiliki keterlibatan penting dalam mengupayakan keamanan komunikasi serta pengalaman pengguna dalam menggunakan layanan pesan singkat.

Kata Kunci: Pesan Spam Dan Ham; Phishing; Algoritma Naive Bayes; Machine Learning; Klasifikasi

Abstract−One of the negative impacts of current digital advances is the increasing number of SMS spam. Spam SMS poses a security risk to users because they can contain malicious links or requests for personal information that are used for malware, smishing, or fraud attacks. However, with the various protection measures available, not all spam SMS can be classified and prevented effectively. However, this problem can be minimized by creating an anti-spam SMS model which aims to classify SMS types. So this research aims to classify types of SMS that contain spam and spam by applying the Naïve Bayes algorithm. In this study, the dataset consisted of 5572 records consisting of 2 categories, namely spam and ham. This algorithm is able to show satisfactory performance in differentiating spam and spam messages because, according to the diversity of literature, the Naïve Bayes algorithm is suitable for use in English language datasets. The evaluation model displays good results with accuracy reaching 93.2%, precision 93.7%, recall 93.2%, and F1-score 91.6%. In addition, analysis in the research using the Receiver Operating Characteristic (ROC) curve shows an accuracy rate of 97.3%, indicating that the model has very good performance in classifying spam in SMS messages. However, there is still room for improvement through the use of new methods and larger and more diverse data sets. This research has an important involvement in working on communication security and user experience in using short message services.

Keywords: Spam And Ham Messages; Phishing; Naive Bayes Algorithm; Machine Learning; Classification

1. PENDAHULUAN

Pada saat ini teknologi komunikasi dan informasi telah mengalami kemajuan yang pesat serta dalam menciptakan beragam media komunikasi dan informasi yang diolah oleh seseorang sehingga terkumpul menjadi asset yang dapat dianalisis guna menghasilkan pengetahuan atau informasi berharga untuk waktu yang akan datang. Kebutuhan informasi dapat dikatakan sebagai kebutuhan mendasar setiap orang, hanya saja kemajuan teknologi memberikan efek samping yang tidak bisa dihindari [1]. Berdasarkan era digital saat ini perkembangan perangkat telepon selular semakin mempermudah seseorang untuk berkomunikasi telepon ataupun pesan. Pengiriman pesan atau Short Message Service (SMS) yaitu data dengan tipe asynchoronous message yang mengirim data dengan mekanisme protokol store and forward [2].

Truecaller Insights Report 2020, menyampaikan negara di asia salah satu nya Indonesia pada tahun 2020 memiliki tingkat spam terbanyak, pada tahun 2019 di Indonesia spam sms memiliki penurunan 3,4% dari 29% menjadi 18,3% hal tersebut mengakibatkan Indonesia menduduki peringkat ke 6 dalam 20 negara. Spam sms dapat dikatakan sebagai sesuatu yang merugikan dan mengganggu bagi pengguna telepon selular, karena pesan – pesan yang tidak diinginkan dikirimkan secara masal kepada sejumlah besar penerima [3]. Dengan munculnya hal tersebut meningkatkan peluang terjadinya kejahatan yang dimanfaatkan oleh pihak yang tidak bertanggung jawab, salah satunya melalui pesan spam sms dengan alih menawarkan produk maupun jasa.

Spam sms bisa menjadi sarana sebagai menyebarkan malware atau phising yang memiliki risiko keamanan serta merugikan integritas data dan perangkat pengguna serta sering kali menyertakan tautan berbahaya yang dapat berdampak pada pencurian identitas, penipuan finansial, atau akses ilegal terhadap akun-akun penting. Salah satu faktor pemicu adanya pesan spam adalah pemberian izin atau permission kepada aplikasi berbahaya sehingga

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 873−880
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5070

dapat mengakses data pengguna. Selain itu, serangan smishing melalui spam SMS dapat merugikan perusahaan dan organisasi dengan menciptakan celah keamanan dalam sistem mereka. Untuk menangani peningkatan kejahatan yang disebabkan spam sms diperlukan suatu filter, salah satunya dengan klasifikasi yang dilakukan untuk menganalisis teks pesan apakah termasuk spam atau ham. Permasalahan tersebut dapat diminimalisir dengan melibatkan penggunaan algoritma pada machine learning untuk melatih model berdasarkan fitur teks pesan. Sebelum dilakukan uji coba klasifikasi dataset harus diproses terlebih dahulu dan pengujian data yang dilakukan melalui platform Google Colab dengan menggunakan pemrograman python. Algoritma yang digunakan dalam pengujian yaitu algoritma Naïve Bayes.

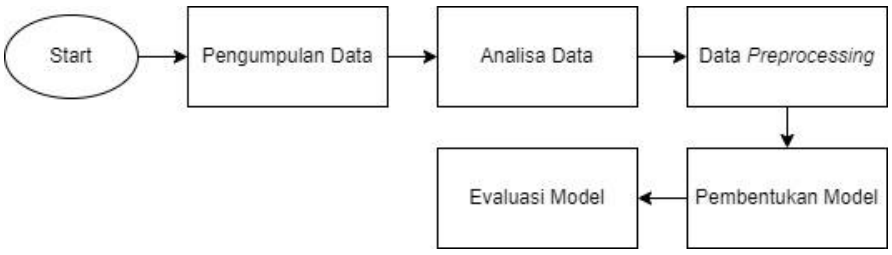
Pada penelitian sebelumnya tentang klasifikasi sms spam berbahasa Indonesia menggunakan metode SVM dan NBC. Data sms yang digunakan sebanyak 1143 yang dibagi menjadi 2 yaitu data latih sejumlah 765 dan data uji 378. Metode TF-IDF akan digunakan dalam pembobotan fitur. Hasil dari pengujian memperoleh bahwa metode NBC mempunyai akurasi yang baik dibanding metode SVM yakni sebesar 94% dan 95%. Sedangkan SVM memiliki akurasi sebesar 92.06% [4]. Pada penelitian sebelumnya hasil akurasi yang didapatkan menggunakan metode Naive Bayes mempunyai hasil akurasi yang stabil dengan nilai 90%, sedangkan metode SVM memperoleh selisih akurasi sekitar 3,37% dengan presisi yang baik dengan nilai presisi sebesar 94,98% [5]. Penelitian sebelumnya melakukan komparasi antara algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan Rapid Miner 7,2. Algoritma Naïve Bayes menghasilkan akurasi sebesar 90% dengan nilai AUC 0.833, sedangkan algoritma Support Vector Machine menghasilkan akurasi sebesar 81% dengan nilai AUC sebesar 0.987 [6]. Penelitian sebelumnya mengidentifikasi SMS Spam menggunakan metode Naïve Bayes mendapatkan hasil yang cukup akurat untuk menyaring spam pada pesan teks dengan tingkat kesalahan sebesar 1,33% dengan menggunakan 30 data latih [7]. Penelitian sebelumnya aplikasi klasifikasi sms berbasis web menggunakan algoritma Logistic Regression mendapatkan hasil pengujian test size dan solver adalah 1 utuk data training dan 0,975 untuk data testing [8].

Setelah dilakukan tinjauan pustaka maka dapat disimpulkan bahwa spam SMS berbahasa Inggris lebih layak digunakan karena menyediakan data yang luas dan beragam. Sehingga penelitian ini menggunakan dataset berbahasa inggris dan fitur teks TF-IDF untuk dapat menampilkan hasil algoritma Naïve Bayes yang efektif dalam membedakan spam dan ham pada dataset yang digunakan. Evaluasi model menunjukkan akurasi 93,2%, presisi 93,7%, recall 93,2%, dan F1-score 91,6%. Analisis Kurva ROC dengan akurasi 97,3% menunjukkan performa yang sangat baik. Meskipun demikian, masih ada ruang untuk peningkatan melalui metode baru dan dataset yang lebih besar dan beragam. Pengembangan teknik klasifikasi spam akan meningkatkan keamanan komunikasi dan pengalaman pengguna layanan pesan singkat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini memerlukan model machine learning yang dapat membuat model untuk mencari akurasi terbaik dengan menerapkan algoritma Naïve Bayes Bayes guna mengklasifikasi jenis pesan berdasarkan data latih yang sudah diberi label yaitu spam dan pesan normal (ham). Terdapat prosedur yang dilakukan pada penelitian ini sebagai berikut :



Gambar 1. Tahapan Penelitian

Start, Tahap awal untuk menentukan arah dan tujuan keseluruhan penelitian.
Analisa Data, Proses ini melibatkan pemahaman awal terhadap data guna menyusun kesimpulan yang bisa menjadi dasar untuk pengujian dalam konteks penelitian.
Data Preprocessing, Proses ini tahapan awal yang dilakukan sebelum memulai proses analisis lebih lanjut guna memastikan data yang digunakan telah disiapkan dengan baik sehingga dapat meningkatkan tingkat akurasi yang dihasilkan
Pembentukan Model, Proses ini melibatkan pemilihan algoritma yang sesuai dengan penggunaan dataset untuk melatih model, menyesuaikan parameter model dan mengevaluasi kinerja model untuk performa kinerja yang optimal.
Evaluasi, Proses ini menilai kinerja suatu model terhadap analisis yang dikembangkan dengan menggunakan metrik evaluasi seperti akurasi, presisi, recall, F1-Score guna mengukur sebarapa efektif model dalam mencapai tujuan dan kebutuhan yang ditetapkan

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 873−880
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5070

2.1.1 Pengumpulan Data

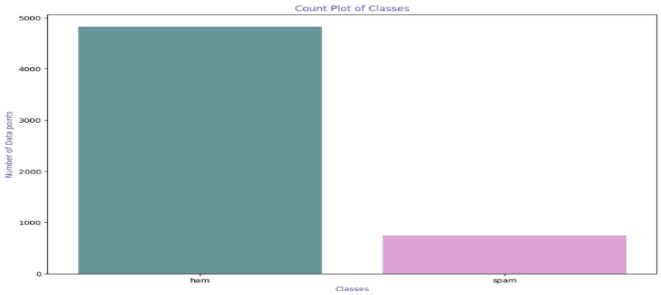
Data yang digunakan didapatkan dari website UCI. Data yang didapatkan merupakan bentuk format .csv dan berjumlah 5572 record dengan 2 category yaitu spam dan ham akan dijadikan dataset yang dapat diakses melalui tautan (mailto:https://archive.ics.uci.edu/dataset/228/sms+spam+collection).

Tabel 1. Deskripsi Dataset

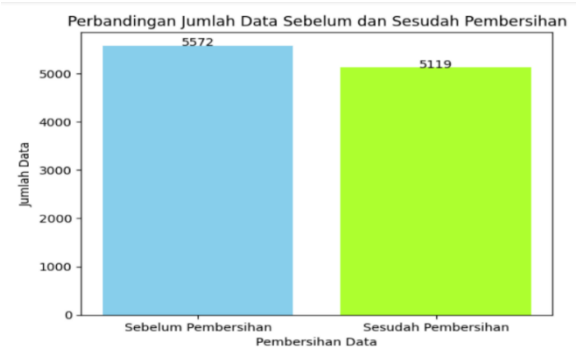
Category	Message
Ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
Ham	Ok lar... Joking wif u oni...
Spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question (std txt rate)T&C's apply 08452810075over18's
Ham	U dun say so early hor... U c already then say...

2.1.2 Analisa Data

Tahap ini dilakukan guna membersihkan data dari elemen-elemen yang tidak diperlukan atau berpotensi mengganggu akurasi analisis. Proses ini melibatkan penghapusan simbol URL, angka, baris kosong, kolom kosong, missing value dan duplikat data tetapi juga memastikan bahwa analisis yang dilakukan pada dataset tersebut menghasilkan hasil yang akurat sehingga dapat diproses menggunakan machine learning. Pada distribusi sebaran data spam sms terdapat seperti gambar 2 dan gambar 3.



Gambar 2. Sebaran Dataset Spam dan Ham



Gambar 3. Visualisasi Pembersihan Data

2.1.3 Data Pre-processing

Tahap awal pada klasifikasi spam sms untuk membersihkan dan mempersiapkan data mentah sebelum diproses dengan baik oleh machine learning [9]. Tujuan dilakukan proses ini guna meningkatkan kualitas data, menghilangkan noise dan mengoptimalkan kinerja model. Proses yang dilakukan yaitu :

a. Tokenisasi

Proses memecah teks atau kalimat menjadi suku kata yang lebih kecil untuk dapat diolah dengan lebih mudah guna pelatihan model dan analisis sentimen. Tokenisasi dapat dikatakan penting dikarenakan banyak algoritma dan model yang bekerja dengan suku kata atau token [10]. Tujuan utama dari proses ini adalah untuk mempermudah analisis dan pemprosesan teks.

Tabel 2. Tokenisasi

Message	Tokenisasi
go until jurong point crazi avail onli in bugi n great world la e buffet cine there got amor wat	['Go', 'until', 'jurong', 'point', 'crazy', 'Available', 'only', 'in', 'bugis', 'n', 'great', 'world', 'la', 'e', 'buffet', 'Cine', 'there', 'got', 'amore', 'wat']

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 873−880
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5070

Message	Tokenisasi
ok lar joke wif u oni	['Ok', 'lar', 'Joking', 'wif', 'u', 'oni']
free entri in a wkli comp to win fa cup final tkt st may text fa to to receiv entri questionstd txt ratetc appli over	['Free', 'entry', 'in', 'a', 'wkly', 'comp', 'to', 'win', 'FA', 'Cup', 'final', 'tkts', 'st', 'May', 'Text', 'FA', 'to', 'to', 'receive', 'entry', 'questionstd', 'txt', 'rateTCs', 'apply', 'overs']
u dun say so earli hor u c alreadi then say	['U', 'dun', 'say', 'so', 'early', 'hor', 'U', 'c', 'already', 'then', 'say']

b. Stopword
Proses ini untuk menghilangkan kata yang sering muncul dalam suatu teks dan tidak terlalu berpengaruh terhadap pemahaman isi teks atau kata kunci [11]. Tahap ini dapat membantu meningkatkan efiensi dan kualitas analisis teks. Contoh kata yang akan dihapus seperti “and”, “the”, “for” dan sebagainya [12].

```
The First 5 Texts after removing the stopwords
['go', 'jurong', 'point', 'crazy', 'available', 'bugis', 'n', 'great', 'world', 'la', 'e', 'buffet', 'cine', 'got', 'amore', 'wat']
['ok', 'lar', 'joking', 'wif', 'u', 'oni']
['free', 'entry', 'wkly', 'comp', 'win', 'fa', 'cup', 'final', 'tkts', 'st', 'may', 'text', 'fa', 'receive', 'entry', 'questionstd', 'txt', 'ratetcs',
['u', 'dun', 'say', 'early', 'hor', 'u', 'c', 'already', 'say']
['nah', 'dont', 'think', 'goes', 'usf', 'lives', 'around', 'though']
```

Gambar 4. Stopword

c. Lematisasi
Tahap yang digunakan untuk mengurangi kata kebentuk dasar sesuai kamus dengan tujuan menghilangkan kata yang tidak diperlukan seperti menghilangkan himbuhan dan konjungsi. Tujuan utama dari lematisasi yaitu untuk mengurangi variasi kata ke dalam bentuk standar sehingga kata dengan akar yang sama akan diwakilkan oleh bentuk dasar yang sama. Misalnya, kata seperti “berlari” maka akan diubah kebentuk dasar “lari” dalam proses lematisasi [13].

```
The First 5 Texts after lemitization
['go', 'jurong', 'point', 'crazy', 'available', 'bugis', 'n', 'great', 'world', 'la', 'e', 'buffet', 'cine', 'get', 'amore', 'wat']
['ok', 'lar', 'joke', 'wif', 'u', 'oni']
['free', 'entry', 'wkly', 'comp', 'win', 'fa', 'cup', 'final', 'tkts', 'st', 'may', 'text', 'fa', 'receive', 'entry', 'questionstd', 'txt', 'ratetcs',
['u', 'dun', 'say', 'early', 'hor', 'u', 'c', 'already', 'say']
['nah', 'dont', 'think', 'go', 'usf', 'live', 'around', 'though']
```

Gambar 5. Lematisasi

d. Pembentukan TF – IDF
Proses mengidentifikasi dan memilih informasi penting atau fitur – fitur yang relevan dari data mentah untuk digunakan dalam pembelajaran mesin guna membantu meningkatkan kualitas dan efisiensi analisis data [14]. Terdapat teknik feature extraction yang dilakukan pada penelitian ini yaitu :

1. Term Frequency
Digunakan dalam pemrosesan teks untuk mengekstrak bobot kata dalam suatu dokumen. Adapun rumus TF – IDF untuk sebuah term dalam sebuah dokumen :

$$TF - IDF\ (t, d) = TF\ (t, d)\ X\ IDF\ (t) \quad (1)$$

Dimana TF (t, d) adalah frekuensi term t dalam dokumen d, dan IDF(t) adalah kebalikan dari frekuensi dokumen yang mengandung term t [15].

```
The First 5 lines in corpus
go jurong point crazy available bugis n great world la e buffet cine get amore wat
ok lar joke wif u oni
free entry wkly comp win fa cup final tkts st may text fa receive entry questionstd txt ratetcs apply overs
u dun say early hor u c already say
nah dont think go usf live around though
```

Gambar 6. TF – IDF

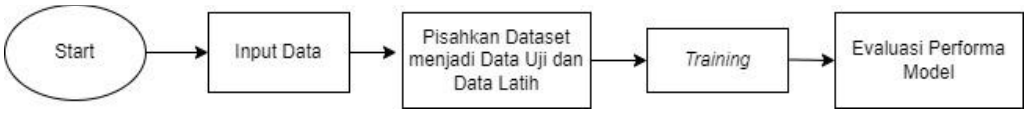
```
dtype('float64')
```

Gambar 7. Nilai bobot TF-IDF

2.1.4 Pembentukan Model

Dalam upaya pemebentukan model menggunakan algoritma Naïve Bayes, tahapan ini menggunakan 4077 data latih dan mengevaluasi performa model dengan 1020 sebagai data uji. Tahap ini digunakan untuk memprediksi kategori untuk setiap data uji dengan menggunakan ROC dan confussion matrix seperti precision, recall dan accuracy untuk mengukur sejauh mana model dapat mengklasifikasi dengan baik [16].

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 873−880
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5070



Gambar 8. Flowchart Pembentukan Model

2.1.5 Evaluasi Model

Proses evaluasi model spam sms dengan menggunakan algoritma Naïve Bayes menngunakan Counfusion Matrix dan kurva Receiver Receiver Operating Characteristic (ROC), tahap ini menjelaskan kinerja model dengan memberikan pandangan mengenai prediksi model yang mencakup True Positives, True Negatives, False Positives, dan False Negatives yang nantinya akan mencari nilai precision, recall, f-measure dan accuracy [17]. Adapun rumus sebagai berikut :

a. Precision :

$$\frac{TP}{TP+FP} \quad (2)$$

Precision mengukur seberapa akurat model dalam mengidentifikasi dengan benar kelas positif dari semua prediksi yang diberikan sebagai positif [18].

b. Recall :

$$\frac{TP}{TP+FN} \quad (3)$$

Recall mengukur kemampuan model guna mendeteksi total jumlah intansi positif dengan benar dalam dataset [19].

c. F-Measure (F1 Score) :

$$\frac{2\ Precision\ Recall}{Precision+Recall} \quad (4)$$

F-Measure merupakan rata-rata harmonik antara Precision dan Recall, menggabungkan kedua metrik agar seimbang antara keduanya [20].

d. Accuracy :

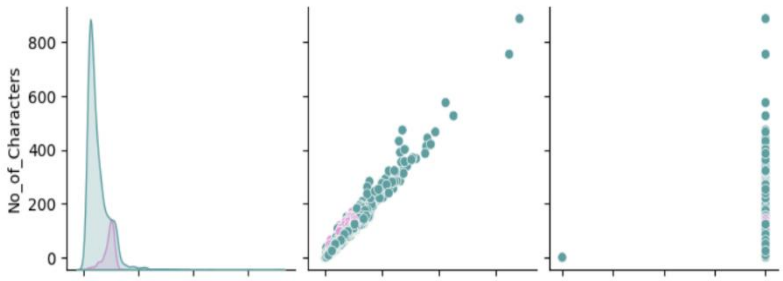
$$\frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Accuracy mengukur sejauh mana kemampuan model dapat memprediksi dengan benar kelas positif dan negatif secara keseluruhan [21].

Sementara itu, Kurva ROC memberikan visualisasi grafis tentang sensitivitas dan spesifisitas model pada pengklasifikasian. Area di bawah kurva ROC memberikan ukuran numerik tentang seberapa baik model dapat membedakan antara kedua kelas. Semakin tinggi nilai ROC, maka semakin baik kemampuan model dalam mengklasifikasikan data [22]. Dengan menggunakan Confusion Matrix dan memeriksa bentuk kurva ROC, maka dapat memberikan evaluasi tentang performa model Naive Bayes dalam mengidentifikasi dan menyaring pesan spam dan non-spam pada dataset yang digunakan guna mendapatkan performa model dengan accuracy terbaik.

3. HASIL DAN PEMBAHASAN

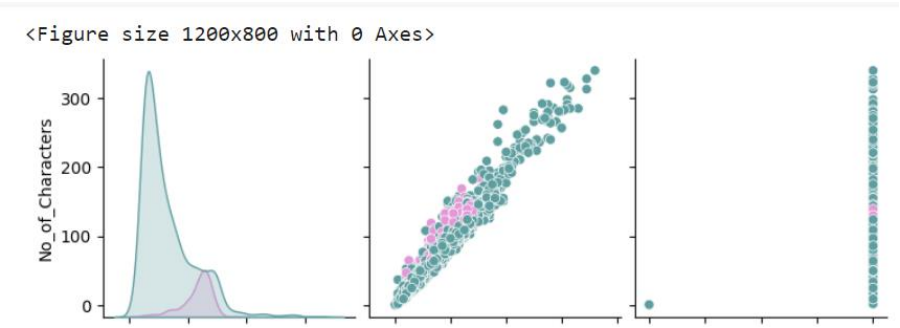
Pada penelitian ini proses pengolahan data Spam Sms yang dilakukan menggunakan pengkodean dengan Google Colab dengan menggunakan pemrograman python. Pengujian dilakukan untuk mengklasifikasi jenis pesan yang mengandung spam dan non spam dengan menerapkan algoritma Naïve Bayes dikarenakan mampu dengan cepat dan efisien dengan menggunakan pemrograman python penelitian ini dilakukan penghapusan outlier pada data spam SMS dengan hasil yang menujukan adanya perbaikan signifikan dalam kualitas data. Outlier yang sebelumnya mengganggu distribusi data telah berhasil dihilangkan, sehingga model dapat fokus pada pola yang sebenarnya.



Gambar 9. Scatter Plot Terdapat Outlier

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 873−880
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5070

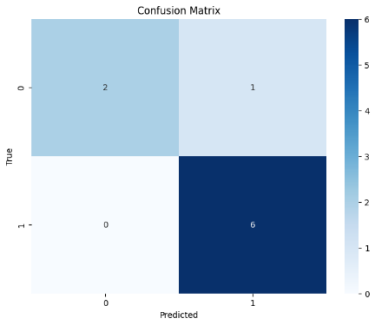
Adapun scatter plot sebelum dilakukan penghapusan outlier terlihat adanya distribusi yang tidak merata. Keberadaan outlier yang muncul dalam data dapat menyebabkan kesalahan dalam analisis yang mengakibatkan model tidak stabil bahkan tidak akurat. Dengan menghilangkan outlier dapat memperbaiki kualitas data dan meningkatkan akurasi serta meningkatkan kinerja model dalam mengklasifikasi pesan spam agar lebih efisien. Setelah dilakukan penghapusan outlier pada data spam SMS, terlihat bahwa distribusi data menjadi lebih teratur dan konsisten seperti gambar 10 dibawah ini :



Gambar 10. Scatter Plot Tidak Terdapat Outlier

Terdapat proses analisis word cloud guna mendapatkan visualisasi untuk memperlihatkan kata kunci yang dominan dalam pesan spam, melalui word cloud memudahkan mengklasifikasi mengenali pola dan tren utama yang ada dalam pesan spam tersebut. Hal ini dapat membantu mengenai pemahaman tentang karakteristik data secara keseluruhan yang dapat digunakan untuk mengembangkan klasifikasi spam yang lebih efektik.

Gambar 11. Word Cloud



Gambar 12. Confusion Matrix

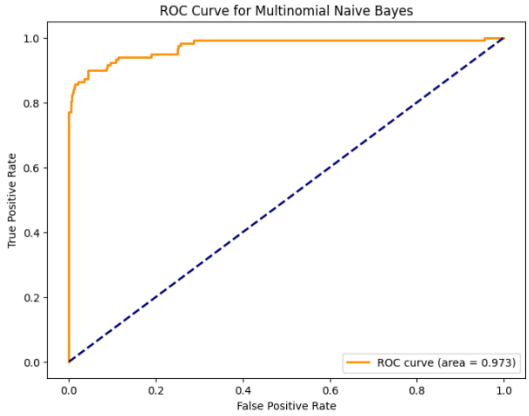
Gambar 12 Confusion Matrix digunakan untuk menganalisa kinerja Algoritma Naïve Bayes dalam mengklasifikasi pesan sms sebagai spam atau ham guna membantu dalam memvisualisasikan performa model dengan membandingkan prediksi kelas dengan label dari data uji. Confusiom matrix terdiri dari true Positivies (TP), True Negatives (TN), False Positivies (FP), dan False Negatives (FN). Berdasarkan nilai TP, TN, FP, dan FN dapat mengitung berbagai matrik evaluasi seperti akurasi, presisi, recall, dan F1-Score. Dari hasil Confusion Matrix dapat dilihat seberapa baik model dalam memberikan hasil klasifikasi data. Terdapat hasil evaluasi yang diperoleh dari pengujian kinerja algoritma klasifikasi Naïve Bayes.

```
Accuracy: 0.9323529411764706
Precision: 0.9371599927303568
Recall: 0.9323529411764706
F1-score: 0.9196239849687265
```

Gambar 13. Hasil Evaluasi

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 873−880
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5070

Klasifikasi spam sms menggunakan algoritma Naïve Bayes menunjukan hasil evaluasi kinerja yang sangat baik dengan menghitung skor probalitas untuk setiap kelas berdasarkan dataset yang digunakan dari hasil yang dapat dilihat seberapa baik model dalam memprediksi data dengan menunjukkan hasil akurasi 93.2% model berhasil mengklasifikasi sebagian besar pesan dengan benar. Kemudian terdapat presisi sebesar 93.7% menujukan bahwa sebagian besar pesan yang diklasifikasikan sebagai spam oleh model memang benar-benar spam. Selain itu, recall sebesar 93.2% menunjukan bahwa model dapat dengan baik mengklasifikasikan kemampuan model untuk menemukan sebgaian besar spam pesan dari keseluruhan jumlah pesan yang sebenarnya. F1- score mencapai 91.6% yang berarti mean dari presisi dan recall dapat memberikan gambaran tentang kinerja keseluruhan model dalam mengklasifikasikan pesan spam. Model klasifikasi menggunakan metode Naïve Bayes sangat efektif dalam mengklasifikan pesan spam pada dataset yang diberikan.



Gambar 14. Kurva Receiver Operating Characteristic (ROC)

Analisis klasifikasi spam pada pesan sms, Kurva Receiver Operating Characteristic (ROC) grafik yang memvisualisasikan hubungan antara True Positive Rate dan False Positive Rate pada berbagai threshold pengklasifikasian untuk mengevaluasi kinerja model. Tingkat akurasi sebesar 97.3% menandakan bahwa model memiliki performa yang sangat baik dalam klasifikasi spam pada pesan sms. Tingkat akurasi yang tinggi mengindikasi bahwa model memiliki kemampuan yang kuat untuk mengklasifikasikan pesan secara tepat dengan tingkat kesalahan yang rendah. Oleh karena itu, akurasi tinggi dapat menjadi indikator penting dari efektivitas model dalam memberikan solusi yang handal dalam menangani masalah spam pada platform pesan.

4. KESIMPULAN

Berdasarkan penelitian tentang klasifikasi spam dalam pesan sms dengan data berjumlah 5572 record dengan 2 category yaitu spam dan ham dapat disimpulkan bahwa penggunaan algoritma Naïve Bayes menghasilkan kinerja yang memuaskan dalam membedakan antara pesan dan ham. Evaluasi model dengan menggunakan metrix seperti akurasi, presisi, recall, dan F1- score yang menunjukan hasil yang baik dengan akurasi yang mencapai 93.2%, presisi sebesar 93.7%, recall sebesar 93.2% dan F1-score sebesar 91.6%. Selain itu, analisis menggunakan Kurva Receiver Operating Characteristic (ROC) dengan tingkat akurasi sebesar 97.3% menandakan bahwa model memiliki performa yang sangat baik dalam klasifikasi spam pada pesan sms, memberikan pemahaman yang lebih dalam kinerja model serta jenis kesalahan yang terjadi, dan membantu memperbaiki fitur yang digunakan atau menyesuaikan threshold untuk meningkatkan kinerja pada model. Namun, masih tedapat ruang untuk peningkatkan melalui metode – metode baru dan penggunaan dataset yang lebih besar dan beragam. Dengan terus mengembangkan dan memperbaiki teknik klasifikasi spam dapat meningkatkan keamanan komunikasi dan pengalaman pengguna dalam menggunakan layanan pesan singkat.

REFERENCES

[1] A. R. Juwita, L. Rahmatiani, T. Al Mudzakir, and A. R. Pratama, “Pemanfaatan Penggunaan Teknologi Internet Sehat Untuk Pendidikan,” J. Buana Pengabdi., vol. 5, no. 2, pp. 92–98, 2023.
[2] H. Herwanto, N. L. Chusna, and M. S. Arif, “Klasifikasi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes,” J. Media Inform. Budidarma, vol. 5, no. 4, p. 1316, 2021, doi: 10.30865/mib.v5i4.3119.
[3] R. Dwiyansaputra, G. S. Nugraha, F. Bimantoro, and A. Aranta, “Deteksi Sms Spam Berbahasa Indonesia Menggunakan Tf-Idf Dan Stochastic Gradient Descent Classifier,” J. Teknol. Informasi, Komput. dan Apl., vol. 3, no. 2, pp. 200–207, 2021.
[4] M. H. S. Ajat, “Klasifikasi Sms Spam Dengan Komparasi Metode Svm Dan Naïve Bayes,” Method. J. Tek. Inform. dan Sist. Inf., vol. 9, no. 1, pp. 31–34, 2023, doi: 10.46880/mtk.v9i1.1694.
[5] P. Apricia, “Kinerja Naïve Bayes Classifier Pada Penyaringan ShortMessage Service (Sms) Spam,” vol. 04, no. 02, pp. 59–66, 2023.
[6] V. No, “Komparasi Algoritma Naïve Bayes dan Support Vectors Machine pada Analisis Sentimen SMS HAM dan SPAM

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 873−880
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5070

1 Program Studi Informasi Akuntansi Kampus Kota Bogor , Universitas Bina Sarana Infromatika 2 Program Studi Sistem Informasi , Universitas Nusa Mandiri 3 P," vol. 4, no. 2, pp. 249–258, 2021.

[7] A. Wahid, M. Baharulloh, R. Kahfiansyah, T. Abrilianto, A. Saifudin, and S. Mulyati, "Identifikasi SMS Spam Menggunakan Metode Naive Bayes," J. Inform. Univ. Pamulang, vol. 6, no. 3, pp. 536–539, 2021, [Online]. Available: http://openjournal.unpam.ac.id/index.php/informatika536

[8] F. D. Pramakrisna, F. D. Adhinata, and N. A. F. Tanjung, "Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression," Teknika, vol. 11, no. 2, pp. 90–97, 2022, doi: 10.34148/teknika.v11i2.466.

[9] A. M. Zuhdi, E. Utami, and S. Raharjo, "Abdul Malik Zuhdi 1) , Ema Utami 2) , Suwanto Raharjo 3) 3," vol. 5, pp. 1–7, 2019.

[10] U. Banten Jaya, J. Syeh Nawawi Albantani, and S. -Banten, "Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine (Svm) Dalam Klasifikasi Sms Spam Berbahasa Indonesia," SAINTEK (Jurnal Sains Teknol.), vol. 3, no., pp. 178–194, 2019.

[11] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm," J. Gaussian, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.

[12] L. D. Utami, L. Yusuf, and D. Nurlaela, "Komparasi Algoritma Naïve Bayes dan Support Vectors Machine pada Analisis Sentimen SMS HAM dan SPAM," Infotek J. Inform. dan Teknol., vol. 4, no. 2, pp. 249–258, 2021, doi: 10.29408/jit.v4i2.3665.

[13] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.

[14] A. Setiyono and H. F. Pardede, "Klasifikasi Sms Spam Menggunakan Support Vector Machine," J. Pilar Nusa Mandiri, vol. 15, no. 2, pp. 275–280, 2019, doi: 10.33480/pilar.v15i2.693.

[15] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," Edutic - Sci. J. Informatics Educ., vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.

[16] I. W. B. Suryawan, N. W. Utami, and K. Q. Fredlina, "Analisis Sentimen Review Wisatawan pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine," J. Inform. Teknol. dan Sains, vol. 5, no. 1, pp. 133–140, 2023.

[17] N. Arifin, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification," STRING (Satuan Tulisan Ris. dan Inov. Teknol., vol. 6, no. 2, p. 129, 2021, doi: 10.30998/string.v6i2.10133.

[18] E. Suryati, Styawati, and A. Ari Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," J. Teknol. dan Sist. Inf., vol. 4, no. 1, pp. 96–106, 2023.

[19] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," J. Sist. Komput. dan Inform., vol. 4, no. 2, p. 296, 2022, doi: 10.30865/json.v4i2.5381.

[20] T. Fadiyah Basar, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 6, no. 3, pp. 1426–1433, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[21] I. F. Rozi, A. T. Firdausi, and Khalimatul Islamiyah, "Analisis Sentimen Pada Twitter Mengenai Pasca Bencana Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram," J. Inform. Polinema, vol. 6, no. 2, pp. 33–39, 2023, doi: 10.33795/jip.v6i2.316.

[22] A. Nur et al., "Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 80 Implementasi Algoritma Regresi Logistik untuk Binary Classification dalam Spam SMS dan WhatsApp Penulis Korespondensi," Agustus, vol. 7, pp. 2549–7952, 2023.