Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

ANALISIS KLASIFIKASI EMAIL SPAM MENGGUNAKAN ALGORITMA NAÏVE BAYES

Azan Rahman[1], Andi Maslan[2]

[1]Mahasiswa Program Studi Teknik Informatika, Universitas Putera Batam
[2]Dosen Program Studi Teknik Informatika, Universitas Putera Batam
email: pb210210102@upbatam.ac.id

ABSTRACT

Spam emails pose a significant challenge in digital communication, requiring effective classification methods to enhance cybersecurity. This study evaluates the performance of the Naïve Bayes algorithm in detecting spam emails, focusing on accuracy, precision, and recall. The dataset consists of pre-labeled emails processed using TF-IDF for feature extraction. The results indicate that the algorithm achieved an accuracy of 90% before addressing class imbalance. After applying SMOTE, the final accuracy improved to 98%. These findings demonstrate that Naïve Bayes is an effective method for spam email classification, with SMOTE enhancing its performance in handling class imbalance.

Keywords: *Accuracy, Class Imbalance, Naïve Bayes, SMOTE, Feature Extraction*

PENDAHULUAN

Seiring berkembangnya teknologi informasi, kini *email* merupakan sarana komunikasi yang sangat penting, baik untuk pribadi maupun profesional. Namun, meskipun memberikan banyak kemudahan, *email* tidak lepas dari masalah, terutama spam *email*. Spam ini sering kali berisi iklan yang tidak relevan, maupun link phising yang dapat merugikan banyak korban secara finansial. Masalah ini menjadi perhatian banyak peneliti dalam beberapa tahun terakhir.

Menurut laporan yang dipublikasikan statista oleh (Dixon, 2023), volume email spam mengalami naik turun pada bulan oktober 2020 hingga september 2021. Pada Oktober 2020, jumlah email spam tercatat sebanyak 242,42 miliar. Namun, angka ini mengalami penurunan menjadi 122,33 miliar pada Januari 2021. Meskipun sempat menurun, jumlah spam kembali meningkat hingga mencapai puncaknya di 282,93 miliar pada Juli 2021. Menariknya, pada bulan berikutnya, Agustus 2021, jumlah spam justru menurun drastis menjadi 65,5 miliar sebelum kembali naik menjadi 88,88 miliar pada September 2021. Data ini menunjukkan bahwa jumlah email spam tidak memiliki pola yang stabil dan cenderung berfluktuasi. Oleh karena itu, penelitian terkait deteksi spam email masih sangat relevan untuk dibahas.

Berbagai metode telah dikembangkan untuk mengidentifikasi dan mengklasifikasikan email spam, seperti menggunakan teknik machine learning. machine learning bertujuan mengubah beragam data menjadi keputusan tanpa campur tangan manusia(Kurniawan,

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

2022). Ada banyak algoritma yang terdapat pada *machine learning*, contohnya algoritma *naïve bayes*, yang akan dibahas disini. *Naïve bayes* bekerja dengan berdasarkan probabilitas. Dalam penelitian ini, algoritma ini dipilih karena kemampuannya dalam klasifikasi teks, termasuk dalam mendeteksi spam email. Algoritma ini bekerja dengan menganalisis ciri-ciri tertentu dalam email, seperti konten, untuk menentukan apakah email tersebut spam atau tidak.

KAJIAN TEORI

2.1 *Email*

Email merupakan alat komunikasi yang tersedia di internet dan digunakan untuk berbagai keperluan, seperti berbagi informasi melalui file lampiran, berdiskusi dalam grup milis, serta sebagai media promosi bagi perusahaan(Wibisono, 2023).

2.2 *Machine Learning*

Menurut (Taye, 2023), menjelaskan bahwa dalam *Machine Learning*, suatu program komputer diberikan serangkaian tugas untuk diselesaikan. Program tersebut dianggap telah belajar jika kemampuannya dalam menyelesaikan tugas-tugas tersebut meningkat seiring waktu, sejalan dengan frekuensi latihan yang dilakukan. Sedangkan menurut (Munir et al., 2024) *Machine learning* merupakan bidang ilmu yang menyediakan alat dan teknik untuk secara otomatis mengidentifikasi pola dari data, memprediksi peristiwa di masa depan berdasarkan pola tersebut, atau mengambil keputusan dalam kondisi yang tidak pasti.

2.3 Algoritma *Naïve Bayes*

Menurut (Munir et al., 2024) *machine learning* merupakan disiplin ilmu yang mengembangkan berbagai metode dan teknik untuk mengidentifikasi pola dalam data secara otomatis, memperkirakan peristiwa di masa mendatang berdasarkan pola tersebut, serta mendukung pengambilan keputusan dalam kondisi ketidakpastian.

2.4 Penelitian Terdahulu

Sebagai langkah awal, penting untuk meninjau beberapa penelitian terdahulu yang terkait, beberapa diantaranya adalah sebagai berikut:

1. (Bachri & Gunawan, 2024) pada jurnalnya yang berjudul "Deteksi *email spam* menggunakan algorima *CNN*" menggunakan algoritma *CNN* dan membuktikan efektifitasnya dengan menghasilkan akurasi pada setiap data uji berjumlah 20%, 30%, 40%, masing masing sebesar 99.67%, 99.64%, serta 99.63%.
2. (Jaiswal et al., 2021) pada jurnal berjudul "*Detecting spam e-mails using stop word TF-IDF and stemming algorithm with Naïve Bayes classifier on the multicore GPU*" melakukan penelitian dengan menggunakan multicore GPU dan menghasilkan akurasi sebesar 99,67% pada tahap pelatihan dan 99,03% saat pengujian. Dari segi kecepatan, pelatihan menggunakan GPU memerlukan waktu 1,361 detik, sedangkan pada CPU membutuhkan 2,029 detik. Sementara itu, proses pengujian berlangsung selama 1,978 detik pada GPU dan 2,280 detik pada CPU.
3. (Qodariyah Fitriyah & Oktavianto, 2019) pada jurnal berjudul "Deteksi *Spam* Pada *Email* Berbasis Fitur Konten Menggunakan *Naïve Bayes*" menggunakan algoritma naïve bayes dan mencapai akurasi sebesar

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

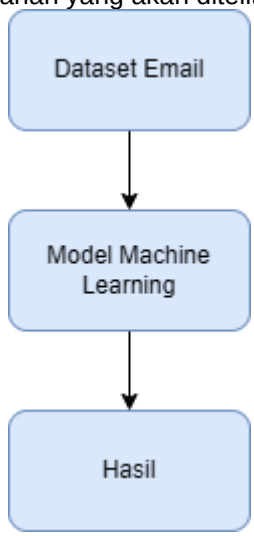
Jurnal Comasie

ISSN (Online) 2715-6265

84,8%, dengan total 3.903 *email* berhasil diklasifikasikan dengan benar, sementara 698 *email* mengalami kesalahan klasifikasi. Nilai *precision* dan *recall* yang diperoleh masing-masing adalah 0,86 dan 0,85.

2.5 Kerangka Berpikir

Kerangka pemikiran digunakan untuk memberikan gambaran mengenai alur berpikir dalam menyelesaikan permasalahan yang akan diteliti.



Gambar 1. Kerangka Berpikir
(Sumber: Data Penelitian, 2024)

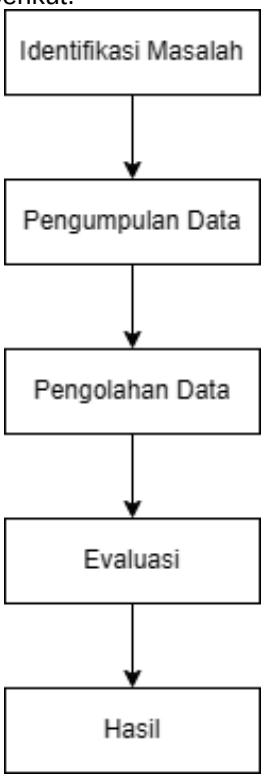
langkah-langkah dalam kerangka pemikiran adalah sebagai berikut:

1. Langkah awal penelitian adalah dengan mengumpulkan data email spam. Data ini dapat berisi informasi email berupa teks email serta kategori berupa spam atau non spam.
2. Langkah kedua adalah membangun model machine learning dengan menggunakan data yang sudah dikumpulkan.
3. Langkah ketiga, hasil berupa model machine learning yang dapat mengklasifikasikan email, dievaluasi dengan metrik seperti akurasi, presisi, recall, dan F1-Score.

METODE PENELITIAN

3.1 Desain Penelitian

Agar penelitian dapat berjalan secara sistematis, diperlukan adanya desain penelitian. Adapun tahapannya adalah sebagai berikut.



Gambar 2. Gambar Desain Penelitian
(Sumber: Data Penelitian, 2024)

1. Identifikasi Masalah

Masalah utama pada penelitian ini adalah bagaimana mengidentifikasi *email spam* dan bukan *spam* menggunakan *naïve bayes.*

2. Pengumpulan Data

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

Data dikumpulkan melalui kajian pustaka dengan menelusuri berbagai sumber referensi yang relevan mulai dari jurnal online hingga buku cetak. Selain itu, dataset yang digunakan diperoleh dari harvard dataverse yang terdiri dari teks email dan 2 kategori utama, yaitu spam dan bukan spam.

3. Pengolahan Data

Pengolahan data melibatkan beberapa teknik yaitu, *preprocessing* data, ekstraksi fitur dengan *TF-IDF*, pembagian data menjadi data testing dan data uji, serta menerapkan algoritma *naïve bayes*.

4. Evaluasi

Model *machine learning* yang diperoleh kemudian dievaluasi hasilnya dengan metrik evaluasi seperti *presisi*, *akurasi*, *F1-Score*, dan *recall*

5. Hasil

Hasil yang didapatkan yaitu model klasifikasi serta hasil evaluasi performa.

3.2 Metode Pengumpulan Data

Dataset pada penelitian ini diambil dari platform *harvard dataverse*. Data berjumlah 5728 dengan rincian data kategori *spam* sejumlah 1369 dan data bukan *spam* berjumlah 4329. Dataset ini akan digunakan untuk melatih dan menguji algoritma *naïve bayes* setelah dilakukan *preprocessing* data.

3.3 metode perancangan

Untuk mengembangkan model klasifikasi yang efektif, diperlukan metode perancangan yang sistematis. Oleh karena itu metode perancangannya dijelaskan sebagai berikut.

1. Tokenisasi

Langkah pertama yaitu setiap kalimat diubah menjadi bagian-bagian kecil menjadi token yang disebut tokenisasi. Proses ini menggunakan *library NLTK* yang terdapat pada *python*.

2. Lematisasi

Langkah kedua adalah lematisasi. Langkah ini memungkinkan kita mengubah token-token menjadi kata dasarnya, hal ini perlu dilakukan untuk mengurangi variasi kata.

3. Penghapusan *stopwords*

Setelah lematisasi, kata-kata yang tidak penting perlu dihapus karena umumnya kata-kata ini tidak memberikan makna relevan. Proses dilakukan dengan menggunakan *library NLTK* pada Python.

4. Ekstraksi Fitur

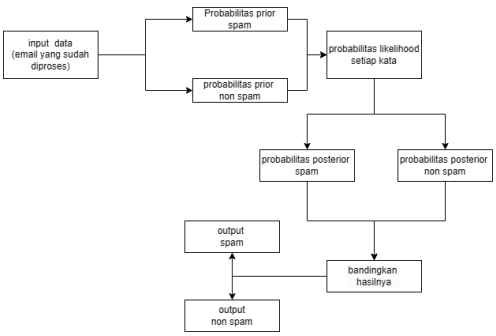
Ekstraksi fitur pada penelitian ini dilakukan dengan teknik *TF-IDF*. Teknik ini menilai pentingnya kemunculan suatu kata dalam dokumen dan memberikan bobot pada kata tersebut.

5. *Training* dan *Testing*

Data dibagi menjadi dua bagian untuk proses pelatihan dan pengujian, dengan 80% dialokasikan untuk pelatihan dan 20% sisanya untuk pengujian.

6. Algoritma *Naïve Bayes*

Langkah selanjutnya adalah penerapan algoritma *naïve bayes*. Cara kerja *naïve bayes* dapat dilihat pada *flowchart* berikut.



Gambar 3. Alur Naïve Bayes
(Sumber: Data Penelitian, 2024)

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

Selain itu, rumus algoritma *naïve bayes* pada konteks klasifikasi *email* dijelaskan seperti berikut ini.

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)}$$

Penjelasan:
P(C|X) : Probabilitas *posterior*
P(X|C) : Probabilitas l*ikelihood*
P(C) : Probabilitas *prior* C
P(X) : Probabilitas *evidence*

7. Evaluasi Model

Setelah model *naïve bayes* dilatih, tahap berikutnya adalah mengukur kinerja menggunakan *confusion matrix*.

Tabel 1. *Confusion Matrix*

	Prediksi *spam*	Prediksi non *spam*
Aktual spam	TP	FP
Aktual non spam	FN	TN

(Sumber: Data Penelitian 2024)

True Positive(TP) ialah kondisi ketika model berhasil mengidentifikasi email spam dengan benar, yaitu *email* yang memang diklasifikasikan sebagai *spam*. Sebaliknya, *False Positive* (FP) ialah kondisi ketika model salah mengklasifikasikan *email* bukan *spam* sebagai *spam*. *True Negative* (TN) ialah kondisi ketika model dengan tepat mengklasifikasikan *email* bukan *spam* sebagai bukan *spam*. Lalu, terdapat juga *False Negative* (FN), yaitu kondisi dimana *email* yang seharusnya *spam* malah diklasifikasikan sebagai bukan *spam*. Dengan memahami keempat kategori ini, maka metrik evaluasi dapat diukur seperti berikut ini.

Akurasi:

$$akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Presisi:

$$presisi = \frac{TP}{TP + FP}$$

Recall:

$$recall = \frac{TP}{TP + FN}$$

F1-Score:

$$f1 - score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall}$$

3.4 Lokasi dan Jadwal

Penelitian ini dilakukan di universitas putera batam kampus tembesi dan dilakukan pada rentang waktu september 2024 hingga januari 2025.

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Bagian ini akan membahas hasil penelitian yang meliputi tahap *preprocessing* hingga evaluasi model dalam analisis klasifikasi *email spam* menggunakan algoritma *naïve bayes*.

1. Tokenisasi

Hasil tokenisasi dapat dilihat pada tabel berikut ini.

Tabel 2. Hasil Tokenisasi

Sebelum	Sesudah
security alert - confirm your national credit union information.	*[security, alert, confirm, your, national, credit, union, information]*
the most expensive car sold in graand !	*[most, expensive, car, sold, graand,*

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

cheap cars in graand.	*cheap, cars, graand]*
want to accept credit cards ? credit approved no cecks do it now.	*[want, accept, credit, cards, aredit, cpproved, no, cecks, do, it, now]*
need to find something ? to be removed from this list , click here	*[need, find, something, removed, from, this, list, click, here]*

(Sumber: Data Penelitian 2024)

Pada tabel, dapat dilihat bahwa proses tokenisasi memecah kalimat menjadi bentuk token.

2. Lematisasi

Hasil lematisasi dapat dilihat pada tabel berikut ini.

Tabel 3. Hasil lematisasi

Sebelum	Sesudah
[unbelievable, new, homes, made, easy, im, wanting, to, show, you, this, homeowner]	*[unbelievable, new, home, made, easy, im, wanting, to, show, you, this, homeowner]*
[congratulations, you, have, won, free, tickets, to, movies]	*[congratulation, you, have, win, free, ticket, to, movie]*
[our, loans, are, designed, to, help, you, with, your, needs]	*[our, loan, be, design, to, help, you, with, your, need]*
[start, earning, rewards, with, exclusive, deals, today]	*[start, earn, reward, with, exclusive, deal, today]*

(Sumber: Data Penelitian 2024)

Pada tabel, dapat dilihat kata *'homes'* telah diubah menjadi bentuk dasarnya, yaitu *'home'*. Selain itu, kata *'won'* juga diubah menjadi bentuk dasarnya berupa 'win'.

3. Penghapusan *Stopword*

Hasil pada proses penghapusan *stopword* dapat dilihat pada tabel berikut ini.

Tabel 4. Hasil Penghapusan Stopword

Sebelum	Sesudah
help television in 1919 by seat to my knoweledge . chrono cross in 1969	*[help, television, 1919, seat, knowledge, chrono, cross, 1969]*
the most expensive car sold in graand ! cheap cars in graand	*[most, expensive, car, sold, graand, cheap, car, graand]*
save your money by getting an oem software ! need in software for your pc ? just visit our site , we might have what you need.	*[save, money, getting, oem, software, need, software, pc, visit, site, might, have, need]*

(Sumber: Data Penelitian 2024)

Pada tabel dapat diperhatikan bahwa ada beberapa kata yang dihilangkan seperti kata *'in', 'to', 'my'* serta beberapa kata lainnya

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

4. TF-IDF

Perhitungan *TF-IDF* dilakukan dengan *library Scikit learn*. Hasil ini berupa kata kata yang sudah diberikan bobot secara otomatis.

Tabel 5. Hasil *TF-IDF*

Kata	Dokumen 1	Dokumen 2	Dokumen 3
Alert	0.45320	0.0	0.0
Accept	0.0	0.0	0.20277
Achieve	0.0	0.02194	0.0
accuracy	0.0	0.02142	0.0

(Sumber: Data Penelitian 2024)

Pada tabel diatas, dokumen 1, 2, dan 3 merujuk pada contoh teks *email* yang dikategorikan sebagai spam. Dapat dilihat pada tabel nilai-nilai yang mewakili kata-kata tersebut.

5. Data *training* dan *testing*

untuk melihat berapa pengaruh rasio pembagian data *training* dan *testing* terhadap akurasi model, dilakukan esperimen terhadap pembagian *training* dan *testing*. Hasilnya dapat dilihat pada tabel dibawah ini.

Tabel 6. *Training* dan *testing*

Training (%)	*Testing* (%)	*Akurasi* (%)
80	20	90
70	30	89
60	40	88

(Sumber: Data Penelitian 2024)

Dapat dilihat pada tabel, rasio pembagian data hanya berpengaruh sedikit kepada akurasi model.

6. Penerapan algoritma *naïve bayes*

Algoritma *naïve bayes* bekerja dengan menggunakan teknik probabilitas untuk mengklasifikasikan data. Dengan fokus pada kesederhanaan dan efisiensi, *Naïve Bayes* mengadopsi asumsi 'naif' tentang independensi fitur sebagai dasar utamanya(Kumar et al., 2020).

Langkah pertama dalam algoritma ini adalah menghitung probabilitas prior untuk setiap kelas. Probabilitas prior ini menggambarkan kemungkinaan sebuah email masuk kedalam kelas tertentu sebelum mempertimbangkan fitur yang ada pada email tersebut.

selanjutnya, algoritma menghitung probabilitas *likelihood* dari fitur yang ada dalam data untuk setiap kelas, misalnya beraa kemungkinan kata-kata tertentu muncul dalam *email spam* atau non *spam*. Setelah itu, algoritma mengalikan probabilitas prior dan *likelihood*, langkah ini menghasilkan probabilitas *posterior* bagi setiap kelas. Misalnya probabilitas bahwa suatu *email* termasuk *spam* atau *ham* berdasarkan kata yang ada didalamnya. Hasil tertinggi akan menjadi hasil klasifikasi.

7. Hasil evaluasi

Pada tahap evaluasi, hasil *confusion matrix* pada model klasifikasi dapat dilihat pada tabel berikut ini.

Tabel 7. Hasil *Confusion Matrix*

	Prediksi *spam*	Prediksi non *spam*
Aktual *spam*	855	0
Aktual *non spam*	101	183

(Sumber: Data Penelitian 2024)

Hasil matrix evaluasinya dapat dilihat pada tabel dibawah ini.

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

Tabel 9. Metrik evaluasi

Metrik	Spam	Non spam	Total
Precision	100%	89.4%	-
Recall	64.4%	100%	-
F1-Score	78.5%	94.3%	-
Accuracy	-	-	91.2%

(Sumber: Data Penelitian 2024)

Pada tabel terlihat bahwa hasil *False Negative* pada model ini cukup tinggi, yaitu 101. Nilai ini berarti ada 101 pesan yang seharusnya bukan *spam*, tapi malah diprediksi sebagai *spam*. Untuk mengatasi masalah ini, dicoba eksperimen dengan teknik *SMOTE*, teknik ini dikenal sebagai salah satu cara untuk mengatasi class imbalance. Untuk meningkatkan jumlah sampel dari kelompok minoritas, *SMOTE* menggunakan metode over-sampling dengan menghasilkan data sintetis(Andriawan & Ernawati, 2024). Setelah mengimplementasi *SMOTE*, hasil *confusion matrix* adalah sebagai berikut.

Tabel 8. Hasil Confusion Matrix

	Prediksi spam	Prediksi non spam
Aktual *spam*	845	10
Aktual *non spam*	9	275

(Sumber: Data Penelitian 2024)

Dapat dilihat bahwa hasil *False Negative* berkurang signifikan menjadi 9 meski teknik ini menyebabkan peningkatan nilai *False Positive* menjadi 10. Melalui hasil ini, dapat dihitung metrik evaluasinya seperti berikut ini.

Tabel 10. Metrik evaluasi setelah *SMOTE*

Metrik	*Spam*	Non *spam*	Total
Precision	96.49 %	98.94 %	-
Recall	96.87 %	98.83 %	-
F1-*Score*	96.68 %	98.89 %	-
Accuracy	-	-	98.33 %

(Sumber: Data Penelitian 2024)

4.2 Pembahasan

Langkah selanjutnya, dalam bagian ini akan dibahas hasil yang diperoleh serta analisis terhadap performa model.

1. Analisis hasil

Setelah menerapkan teknik *SMOTE*, terjadi peningkatan signifikan pada nilai *recall spam*, yang awalnya 64.4% menjadi 96.87%. *recall* yang tinggi berarti model lebih baik dalam menemukan dan mendeteksi semua kasus positif. Pada kasus deteksi *spam*, *recall* yang tinggi menunjukkan bahwa sebagian besar *email spam* berhasil terdeteksi. Selain itu, terdapat peningkatan akurasi dari 91.2% menjadi 98.33%.

2. Pengujian

Bagian ini menyajikan hasil pengujian yang dilakukan pada model klasifikasi *email spam* menggunakan algoritma *naïve bayes*. Pengujian ini dilakukan untuk mengevaluasi performa model. Hasil pengujiannya dapat dilihat pada tabel berikut.

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

Tabel 11. Pengujian Model

No	Teks Email	kategori
1	*Congratulations! You've won $10,000! Click this link to claim your prize now!*	*Spam*
2	*Get a loan with 0% interest, no collateral required! Contact us now before the offer expires!*	*Spam*
3	*Hi, can we schedule a meeting for next week's project? Please confirm!*	*Non spam*
4	*Don't forget tomorrow's presentation task. I have sent the file to your email.*	*Non spam*
5	*Here is the updated monthly report based on the latest feedback.*	*Non spam*

(Sumber: Data Penelitian, 2024)

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis klasifikasi *email spam* dengan menggunakan algoritma *naïve bayes*. Dapat dilihat hasilnya *naïve bayes* memiliki akurasi sebesar 91.2%. tetapi hanya mendapatkan nilai *recall* untuk *spam* sebesar 64.4%. hal ini menunjukkan model memiliki nilai *false negative* yang cukup tinggi., dimana beberapa *email* tidak dapat dideteksi dengan baik. Setelah menerapkan *SMOTE*, nilai akurasi meningkat menjadi 98.94%. serta mendapat peningkatan *recall* untuk *spam* menjadi 96.8%.

DAFTAR PUSTAKA

Andriawan, M. G., & Ernawati, T. (2024). PENGGUNAAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN KONFLIK PALESTINA DAN ISRAEL PADA PLATFORM X. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3). https://doi.org/10.23960/jitet.v12i3.4943

Bachri, C. M., & Gunawan, W. (2024). *Deteksi Email Spam menggunakan Algoritma Convolutional Neural Network (CNN)*.

Dixon, S. J. (2023, June 26). *Average daily spam volume worldwide from October 2020 to September 2021*. Statista. https://www.statista.com/statistics/1270424/daily-spam-volume-global/

Jaiswal, M., Das, S., & Khushboo. (2021). Detecting spam e-mails using stop word TF-IDF and stemming algorithm with Naïve Bayes classifier on the multicore GPU. *International Journal of Electrical and Computer Engineering*, *11*(4), 3168–3175. https://doi.org/10.11591/ijece.v11i4.pp3168-3175

Kumar, A., Chatterjee, J. M., & Díaz, V. G. (2020). A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing. *International Journal of Electrical and Computer Engineering*, *10*(1), 486–493. https://doi.org/10.11591/ijece.v10i1.pp486-493

Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

Kurniawan, D. (2022). *Pengenalan machine learning dengan python*. Elex Media Komputindo.

Munir, N., Huang, J., Wong, C.-N., & Song, S.-J. (2024). Machine Learning Based Eddy Current Testing: A Review. *Results in Engineering*, 103724. https://doi.org/10.1016/j.rineng.2024.103724

Qodariyah Fitriyah, N., & Oktavianto, H. (2019). Deteksi Spam Pada Email Berbasis Fitur Konten Menggunakan Naïve Bayes. *Jurnal Sistem Dan Teknologi Informasi Indonesia*. https://archive.ics.uci.edu/ml/machine-

Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. In *Computers* (Vol. 12, Issue 5). MDPI. https://doi.org/10.3390/computers12050091

Wibisono, A. (2023). FILTERING SPAM EMAIL MENGGUNAKAN METODE NAIVE BAYES. In *Teknologipintar.org* (Vol. 3, Issue 4).

	Biodata, Azan rahman, Mahasiswaa Teknik Informatika Universitas Putera Batam. Fokus penelitian di bidang machine learning. email: Azan.rahman.coder@gmail.com
	Biodata, Andi Maslan, Dosen Teknik Informatika Universitas Putera Batam, bidang Ilmu Network Security dan aktif melakukan publikasi di Jurnal Nasional dan Internasional. Email : Lanmasco@gmail.com