JURITEK: Jurnal Ilmiah Teknik Mesin, Elektro dan Komputer

E-ISSN: 2809-0799
P-ISSN: 2809-0802

# Klasifikasi Spam Email Otomatis Menggunakan Algoritma Naïve Bayes

**Zianah Nafisah Simbolon[1], Mahfuzhah Rahma Kesuma[2], Dedek[3], Muhammad Rayhans Adrian [4], Khairul Arifin[5]**

[1] Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, email : zianahnafisah39@gmail.com
[2] Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, email : zianahnafisah39@gmail.com

*** Penulis** yang sesuai : Zianah Nafisah Simbolon

**Abstract:** The rapid development of information technology has intensified data exchange through email. However, this also increases the risk of spam distribution, which can compromise privacy, reduce productivity, and potentially pose security threats. This study aims to implement the Naïve Bayes algorithm to automatically classify emails into spam and non-spam categories. The Naïve Bayes method was chosen due to its ability to handle large datasets, efficient computational process, and high accuracy in text-based classification tasks. The research stages include collecting an email dataset, performing text preprocessing such as tokenization, stopword removal, and stemming, followed by training the model using the Naïve Bayes algorithm. The experimental results show that the developed model can classify emails with good accuracy and relatively short computation time.

**Keywords:** *Spam Email, Naïve Bayes, Klasifikasi Teks, Machine Learning, Deteksi Spam.*

Abstrak: Perkembangan teknologi informasi yang semakin pesat membuat pertukaran data melalui email menjadi semakin intensif. Namun, hal ini juga meningkatkan risiko penyebaran spam yang dapat mengganggu privasi, menurunkan produktivitas, serta berpotensi membawa ancaman keamanan. Penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam mengklasifikasikan email menjadi kategori spam dan non-spam secara otomatis. Metode Naïve Bayes dipilih karena kemampuannya dalam menangani data berukuran besar, proses komputasi yang efisien, serta akurasi yang cukup tinggi pada permasalahan klasifikasi berbasis teks. Tahapan penelitian meliputi pengumpulan dataset email, proses prapemrosesan teks seperti tokenisasi, stopword removal, dan stemming, kemudian pelatihan model menggunakan algoritma Naïve Bayes. Hasil pengujian menunjukkan bahwa model yang dibangun mampu melakukan klasifikasi email dengan tingkat akurasi yang baik serta waktu komputasi yang relatif singkat.

**Diterima: Oktober 20, 2025**
**Direvisi: Oktober 28, 2025**
**Diterima: Oktober 29, 2025**
**Diterbitkan: Desember 2025**
**Versi sekarang: Desember 26, 2025**

**Kata kunci:** *Spam Email, Naïve Bayes, Klasifikasi Teks, Machine Learning, Deteksi Spam.*



## 1. 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 [1].

Dengan berkembangnya volume pengiriman email juga dapat memicu permasalahan yang serius, Yaitu Spam Email, yang merupakan pesan yang tidak penting yang dikirim secara Massal, Biasanya berupa Iklan, Tautan, Phishing, maupun bisa berupa Malware [2] 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[3].

Ada banyak algoritma yang terdapat pada machine learning, contohnya algoritma naïve bayes. Naïve bayes bekerja dengan berdasarkan probabilitas. Dalam penelitian ini, algoritma ini dipilih karena kemampuannya dalam klasifikasi teks, termasuk dalam mendeteksi spam email [4] Salah satu efek negatif yang paling signifikan dari spam email adalah membebani sumber daya jaringan dan waktu yang terbuang sia-sia untuk menghapus spam [5] Proses penyaringan email manual dianggap kurang efisien, terutama dalam situasi di mana jumlah email yang diterima sangat besar [6].

Berdasarkan laporan CISCO, ada sekitar 85 persen dari seluruh pesan email yang dikirimkan di April 2019 dapat diklasifikasikan sebagai spam [7]. Hal tersebut tentu saja menjadi masalah karena kapasitas email yang terbatas. Pesan email penting pun dapat tertimbun oleh pesan-pesan spam sehingga pesan penting tersebut justru tidak tersampaikan oleh penerimanya. Salah satu upaya mengatasi spam adalah dengan melakukan penyortiran spam [8]

Berdasarkan laporan CISCO, ada sekitar 85 persen dari seluruh pesan email yang dikirimkan di April 2019 dapat diklasifikasikan sebagai spam [7]. Hal tersebut tentu saja menjadi masalah karena kapasitas email yang terbatas. Pesan email penting pun dapat tertimbun oleh pesan-pesan spam sehingga pesan penting tersebut justru tidak tersampaikan oleh penerimanya. Salah satu upaya mengatasi spam adalah dengan melakukan penyortiran spam [8]

## 2. Tinjauan Literatur

### 2.1. Naïve Bayes

Algoritma Naive Bayes adalah algoritma klasifikasi probabilistic yang sederhana namun sangat efektif. Algoritma ini juga dikenal dengan prediksi yang cepat dan membutuhkan sedikit data pelatihan [11] Algoritma klasifikasi Naive Bayes terkenal dengan asumsinya yang unik, yaitu "kemurnian" (naive) antar fitur. Asumsi ini menyatakan bahwa fitur-fitur yang digunakan dalam proses klasifikasi tidak saling berhubungan satu sama lain [12].

### 2.2 Spam Email

Spam email adalah pesan email yang tidak diminta atau tidak diinginkan yang dikirimkan secara masal ke ribuan atau jutaan alamat email. Spam email biasanya mengandung iklan atau promosi produk, phishing, virus atau malware, dan pesan palsu lainnya. Spam email bisa sangat mengganggu dan merugikan, karena dapat menguras waktu dan sumber daya komputer pengguna [13].

### 2.3 Text Mining

Sistem klasifikasi email berdasarkan kepentingannya menerapkan metode text mining. Pada prosesnya dalam mengolah dan menganalisis suatu dokumen teks, text mining memiliki beberapa tahapan diantaranya adalah tahapan text preprocessing yang tujuannya agar mempersiapkan teks agar lebih terstruktur untuk dapat diolah dan dianalisis. Preprocessing teks dibagi menjadi beberapa proses yaitu parsing, tekenisasi, stopword removal, stemming dan pembobotan atau indexing [14].

### 2.4 Machine Learning

Machine Learning merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang berfokus pada pengembangan algoritma dan model yang memungkinkan sistem komputer untuk belajar dari data tanpa memerlukan pemrograman secara eksplisit. Menurut literatur, Machine Learning bekerja dengan membangun pola dan hubungan dari dataset yang dianalisis sehingga model mampu melakukan prediksi atau pengambilan keputusan secara otomatis berdasarkan informasi baru [15].

## 3. Metode

Penelitian ini dilakukan melalui beberapa tahapan sistematis untuk membangun model klasifikasi email spam. Tahapan utama meliputi pengumpulan dataset, prapemrosesan teks (*text preprocessing*), pembobotan fitur, pelatihan model menggunakan algoritma Naïve Bayes, dan evaluasi kinerja model.

### 3.1. Pengumpulan Dataset

Data yang digunakan dalam penelitian ini merupakan kumpulan email yang telah dikategorikan ke dalam dua label kelas, yaitu *spam* dan *non-spam* (ham). Dataset ini berisi teks mentah yang masih memuat berbagai elemen derau (*noise*) seperti tanda baca, angka, emotikon, dan format yang tidak konsisten, sehingga memerlukan penanganan lebih lanjut sebelum dapat diproses oleh algoritma.

### 3.2. Prapemrosesan Teks (*Text Preprocessing*)

Tahap ini bertujuan mengubah data teks mentah menjadi format yang terstruktur dan bersih. Proses prapemrosesan terdiri dari langkah-langkah berikut:

1. Cleaning (Pembersihan): Menghapus angka, tanda baca, karakter khusus, dan mengubah seluruh huruf menjadi huruf kecil (*lowercase*). Tujuannya adalah menyeragamkan struktur kalimat.

2. Tokenisasi: Memecah kalimat menjadi satuan kata tunggal atau token untuk memungkinkan perhitungan frekuensi kemunculan kata.

3. Stopword Removal: Menghapus kata-kata umum yang sering muncul namun tidak memiliki pengaruh signifikan terhadap klasifikasi, seperti kata penghubung ("dan", "atau", "yang").

4. Stemming: Mengubah kata berimbuhan menjadi bentuk dasarnya (contoh: "mengirimkan" menjadi "kirim") untuk mengurangi variasi fitur yang bermakna sama.

### 3.3 Ekstraksi Fitur dengan TF-IDF

Setelah teks bersih, data dikonversi menjadi representasi numerik menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). Metode ini memberikan bobot pada kata berdasarkan frekuensi kemunculannya dalam dokumen dan kelangkaannya dalam seluruh korpus. Persamaan matematis untuk TF-IDF didefinisikan sebagai berikut:

(1) Pembobotan (weight)

$$W_{d,t} = TF_{d,t} \times IDF_t$$

Dimana $\mathbf{TF_{d,t}}$ adalah frekuensi kemunculan term *t* dalam dokumen *d*, sedangkan $\mathbf{IDF_t}$ dihitung menggunakan persamaan:

(2) Inverse Document Frequency (IDF)

$$IDF_t = \log\left(\frac{N}{df_t}\right)$$

Dimana *N* adalah total jumlah dokumen dan $\mathbf{IDF_t}$ adalah jumlah dokumen yang mengandung term *t*

### 3.4 Klasifikasi Naïve Bayes

Algoritma yang digunakan untuk klasifikasi adalah *Multinomial Naïve Bayes*. Algoritma ini bekerja berdasarkan Teorema Bayes dengan asumsi bahwa setiap fitur (kata) bersifat independen satu sama lain. Model dilatih untuk memprediksi probabilitas kelas (spam atau non-spam) berdasarkan fitur yang ada. Persamaan Teorema Bayes dapat dituliskan sebagai:

(3) Teorema Bayes

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

Dimana:

- $P(C|X)$ adalah probabilitas posterior kelas $C$ diberikan fitur $X$.
- $P(X|C)$ adalah probabilitas *likelihood*.
- P(C) adalah probabilitas prior kelas.
- P(X) adalah probabilitas *evidence*.

### 3.5 Skenario Pengengujian dan Evaluasi

Dataset dibagi menjadi dua bagian, yaitu 80% sebagai data latih (*training set*) untuk melatih model dan 20% sebagai data uji (*testing set*) untuk mengukur performa model. Evaluasi kinerja dilakukan menggunakan *Confusion Matrix* dengan menghitung parameter *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Akurasi dihitung untuk melihat proporsi prediksi yang benar secara keseluruhan menggunakan persamaan:

(4) Akurasi

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

Sedangkan *Precision* dan *Recall* dihitung untuk mengukur ketepatan model dalam mendeteksi kelas positif (spam):

(5) Precision

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

(6) Recall

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

Dimana TP adalah *True Positive*, TN adalah *True Negative*, FP adalah *False Positive*, dan FN adalah *False Negative*.

## 4. Hasil dan Pembahasan

Bagian ini memaparkan hasil eksperimen klasifikasi email *spam* menggunakan algoritma Multinomial Naïve Bayes. Eksperimen dilakukan menggunakan dataset yang terdiri dari 16.690 data uji. Evaluasi difokuskan pada kemampuan model dalam membedakan antara email *spam* dan *non-spam* (ham), serta analisis fitur kata yang menjadi penentu klasifikasi.

### 4.1 . Analisis Pola Kata (*Word Cloud*)

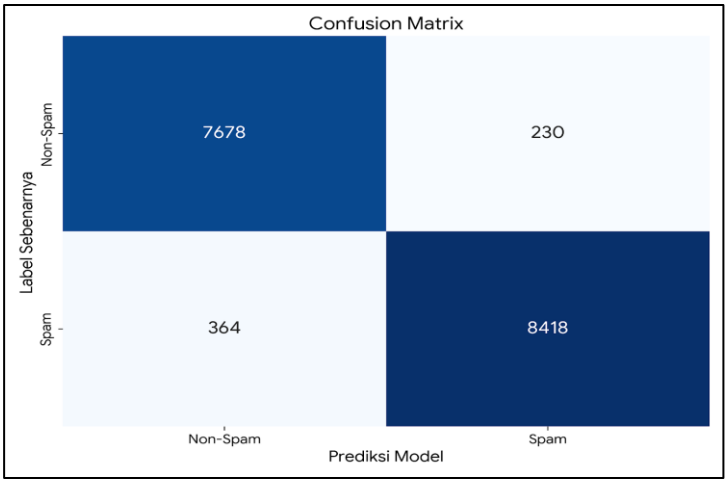
Visualisasi *Word Cloud* digunakan untuk mengidentifikasi kata-kata dominan yang muncul pada setiap kategori email. Hasil visualisasi ini ditunjukkan pada Gambar 1.


Gambar 1. Visualisai Word Cloud

Berdasarkan Gambar 1, terlihat perbedaan karakteristik leksikal yang signifikan. Pada kategori *spam*, kata-kata yang paling sering muncul berkaitan dengan promosi dan urgensi, seperti "promo", "gratis", "hadiah", "uang", dan "klik". Sebaliknya, pada kategori *non-spam*, kata-kata yang mendominasi bersifat lebih formal dan terkait aktivitas rutin, seperti "rapat", "laporan", "kuliah", dan "revisi". Perbedaan distribusi kata ini membuktikan bahwa metode TF-IDF efektif dalam menangkap fitur pembeda antar kelas.

### 4.2 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan Confusion Matrix untuk memetakan detail prediksi benar dan salah. Hasil prediksi model pada data uji disajikan dalam Gambar 2.




Gambar 2. Confusion Matrix

Gambar 2 menunjukkan bahwa model berhasil mengklasifikasikan sebagian besar data dengan benar. Dari total 16.690 data uji, model mendeteksi 8.418 email *spam* secara tepat (*True Positive*) dan 7.678 email *non-spam* dengan benar (*True Negative*). Kesalahan klasifikasi relatif rendah, dimana hanya 230 email *non-spam* yang terdeteksi sebagai *spam* (*False Positive*) dan 364 email *spam* yang lolos sebagai *non-spam* (*False Negative*).

### 4.3 Analisis Metrik Klasifikasi

Berdasarkan nilai dari *Confusion Matrix*, dihitung metrik performa utama yang meliputi Akurasi, Presisi, *Recall*, dan *F1-Score*. Hasil rekapitulasi performa model disajikan dalam Tabel 1.

Tabel 1. Hasil Evaluasi Perfoma Model Naïve Bayes

| Metrik Evaluasi | Nilai (%) |
|---|---|
| **Akurasi (Accuracy)** | 96,44% |
| **Presisi (Precision)** | 97,34% |
| **Recall** | 95,85% |
| **F1-Score** | 96,59% |

Tabel 1 menunjukkan bahwa model memiliki performa yang sangat baik dengan tingkat akurasi mencapai 96,44%. Nilai presisi sebesar 97,34% mengindikasikan bahwa ketika model memprediksi sebuah email sebagai spam, prediksi tersebut memiliki kemungkinan benar yang sangat tinggi. Sementara itu, nilai *recall* sebesar 95,85% menunjukkan kemampuan model dalam menangkap hampir seluruh email spam yang ada dalam dataset.

### 4.4 Pembahasan

Hasil penelitian ini menunjukkan bahwa algoritma Multinomial Naïve Bayes sangat efektif untuk tugas klasifikasi teks berdimensi tinggi seperti penyaringan spam. Keberhasilan ini didukung oleh tahap prapemrosesan teks yang komprehensif, meliputi stopword removal dan stemming, yang berhasil mereduksi noise data.

Tingginya nilai akurasi dan presisi membuktikan bahwa asumsi independensi fitur pada Naïve Bayes, meskipun sederhana, tetap relevan untuk data teks email. Jika dibandingkan dengan kesalahan prediksi, jumlah False Positive (230 data) jauh lebih kecil dibandingkan total data, yang merupakan aspek krusial dalam sistem filter email untuk mencegah hilangnya pesan penting. Hal ini menegaskan bahwa model yang dibangun layak untuk diimplementasikan sebagai sistem deteksi otomatis.

## 5. Kesimpulan

Penelitian ini berhasil mengembangkan sistem klasifikasi email *spam* otomatis yang efektif menggunakan algoritma Multinomial Naïve Bayes dengan pembobotan fitur TF-IDF. Berdasarkan hasil eksperimen, tahapan prapemrosesan data yang meliputi pembersihan (*cleaning*), penghapusan *stopword*, dan *stemming* terbukti memiliki peran vital dalam meningkatkan kualitas data latih dengan mereduksi *noise* dan menyeragamkan fitur kata.

Evaluasi kinerja model pada 16.690 data uji menunjukkan hasil yang sangat memuaskan, dengan tingkat akurasi mencapai **96,44%**. Model juga mencatatkan nilai presisi sebesar **97,34%**, *recall* **95,85%**, dan *F1-Score* **96,59%**. Tingginya nilai presisi mengindikasikan bahwa sistem ini sangat andal dalam meminimalisir kejadian *False Positive*, sehingga risiko email penting (non-spam) yang secara keliru terfilter ke folder spam sangat rendah. Hal ini menegaskan bahwa algoritma Naïve Bayes, meskipun memiliki asumsi independensi yang sederhana, mampu memberikan performa komputasi yang cepat dan akurasi yang kompetitif untuk penyaringan email dalam skala besar.

Keterbatasan penelitian ini terletak pada ketergantungan model terhadap frekuensi kata semata tanpa memahami konteks kalimat secara mendalam. Oleh karena itu, penelitian selanjutnya disarankan untuk mengeksplorasi penggunaan metode berbasis *Deep Learning* seperti *Long Short-Term Memory* (LSTM) atau BERT (*Bidirectional Encoder Representations from Transformers*) untuk menangkap konteks semantik yang lebih kompleks, serta menguji ketahanan model pada dataset dengan variasi bahasa campuran (multi-bahasa).

**Kontribusi Penulis:** Paragraf pendek yang menjelaskan kontribusi masing-masing penulis harus disertakan untuk artikel penelitian dengan beberapa penulis **(wajib untuk lebih dari 1 penulis)** . Pernyataan berikut harus digunakan "Konseptualisasi: XX dan YY; Metodologi: XX; Perangkat Lunak: XX; Validasi: XX, YY dan ZZ; Analisis formal: XX; Investigasi: XX; Sumber daya: XX; Kurasi data: XX; Penulisan—persiapan draf asli: XX; Penulisan—peninjauan dan penyuntingan: XX; Visualisasi: XX; Supervisi: XX; Administrasi proyek: XX; Akuisisi pendanaan: YY"

**Pendanaan:** Harap tambahkan: "Penelitian ini tidak menerima pendanaan eksternal" atau "Penelitian ini didanai oleh NAMA PENDANA, nomor hibah XXX". Periksa dengan saksama apakah rincian yang diberikan akurat dan gunakan ejaan standar nama lembaga pendanaan . Kesalahan apa pun dapat memengaruhi pendanaan Anda di masa mendatang **(wajib).**

**Pernyataan Ketersediaan Data:** Kami mendorong semua penulis artikel yang diterbitkan dalam jurnal FAITH untuk membagikan data penelitian mereka. Bagian ini memberikan perincian mengenai tempat data pendukung hasil yang dilaporkan dapat ditemukan, termasuk tautan ke kumpulan data yang diarsipkan secara publik yang dianalisis atau dibuat selama penelitian. Jika tidak ada data baru yang dibuat atau data tidak tersedia karena batasan privasi atau etika, pernyataan tetap diperlukan.

**Ucapan Terima Kasih:** Di bagian ini, Anda dapat memberikan ucapan terima kasih atas dukungan yang diberikan yang tidak tercakup dalam bagian kontribusi penulis atau pendanaan. Ini dapat mencakup dukungan administratif dan teknis atau sumbangan dalam bentuk barang (misalnya, bahan yang digunakan untuk eksperimen). Selain itu, pernyataan transparansi penggunaan perangkat AI telah disertakan di bagian Ucapan Terima Kasih, jika berlaku.

**Konflik Kepentingan:** Nyatakan konflik kepentingan atau nyatakan **(wajib)** , "Penulis menyatakan tidak ada konflik kepentingan." Penulis harus mengidentifikasi dan menyatakan keadaan atau kepentingan pribadi apa pun yang dapat dianggap memengaruhi representasi atau interpretasi hasil penelitian yang dilaporkan secara tidak pantas. Peran apa pun dari penyandang dana dalam desain studi; dalam pengumpulan, analisis, atau interpretasi data; dalam penulisan naskah; atau dalam keputusan untuk menerbitkan hasil harus dinyatakan di bagian ini. Jika tidak ada peran, harap nyatakan, "Pendana tidak memiliki peran dalam desain studi; dalam pengumpulan, analisis, atau interpretasi data; dalam penulisan naskah; atau dalam keputusan untuk menerbitkan hasil".

## Referensi


[1] DRIM Setiadi, S. Rustad, PN Andono, dan GF Shidik, "Survei dan investigasi steganografi citra digital (tujuan, penilaian, metode, pengembangan, dan dataset)," *Signal Processing* , vol. 206, hal. 108908, Mei 2023, doi: 10.1016/j.sigpro.2022.108908.

[2] DRIM Setiadi, T. Sutojo, EH Rachmawanto, dan CA Sari, "Algoritma watermarking gambar yang cepat dan efisien menggunakan transformasi tchebichef diskrit," dalam *Konferensi Internasional ke-5 tentang Manajemen Layanan TI dan Siber (CITSM) tahun 2017* , Agustus 2017, hlm. 1–5. doi: 10.1109/CITSM.2017.8089229.

[3] A. Vyas, S. Yu, dan J. Paik, "Dasar-Dasar Pemrosesan Gambar Digital," dalam *A John Wiley & Sons* , 2018, hlm. 3–11. doi: 10.1007/978-981-10-7272-7_1.

[4] ICCC FBI, "Laporan Kejahatan Internet 2021," 2022. [Online]. Tersedia: https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf

[5] Sekolah Teknik USC Viterbi, "Basis Data Gambar SIPI." http://sipi.usc.edu/database/ (diakses 27 Maret 2019).

[6] A. P. Windarto, S. Wijaya, dan D. Hartama. (2017). "Penerapan Algoritma Naïve Bayes Classifier Pada Deteksi Email Spam". Jurnal Media Informatika Budidarma, 1(1), 1-5.

[7] F. A. Bachtiar, Y. E. P. Negara, dan I. K. E. Purnama. (2017). "Klasifikasi Email Spam Menggunakan Metode Naïve Bayes Classifier". Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 4(4), 267-272.

[8] . F. Sari dan Y. R. Bata. (2018). "Analisis Performansi Naïve Bayes dalam Mengklasifikasi Email Spam dengan Seleksi Fitur Chi-Square". Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 2(2), 567-573.

[9] I. M. D. Maysanjaya dan I. M. A. W. Putra. (2019). "Sistem Deteksi Email Spam Berbasis Naïve Bayes Classifier". Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi), 7(3), 181-190.

[10] S. H. Wijaya dan A. Fadlil. (2019). "Optimasi Klasifikasi Spam Email menggunakan Algoritma Naïve Bayes dan Particle Swarm Optimization". Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA), 3, 123-130.

[11] D. A. P. Sastrawan dan I. K. G. D. Putra. (2020). "Perbandingan Algoritma Naïve Bayes dan Support Vector Machine untuk Klasifikasi Email Spam". Jurnal Teknologi Informasi dan Komputer (JTIK), 6(1), 45-52.

[12] A. S. A. P. Putra, I. M. Sukarsa, dan I. P. A. Bayupati. (2020). "Implementasi Algoritma Naïve Bayes untuk Filtering Email Spam pada Perusahaan". Jurnal SPEKTRUM, 7(2), 88-95.
[13] L. Hakim dan A. B. Putera. (2021). "Deteksi Email Spam dengan Naïve Bayes Classifier dan Feature Hashing". Jurnal Informatika dan Teknik Elektro Terapan (JITET), 9(1), 78-84.
[14] R. Andrian, E. M. Yunita, dan A. R. Atmadja. (2021). "Analisis Sentimen dan Klasifikasi Spam Email Menggunakan Metode Naïve Bayes". Jurnal Komputer, Informasi Teknologi, dan Elektro (KITE), 6(2), 102-109.
[15] M. F. R. Siregar, T. Limbong, dan S. P. Manik. (2021). "Sistem Klasifikasi Email Spam dengan Metode Naïve Bayes Berbasis Web". Jurnal Media Informatika Budidarma, 5(3), 1129-1136.
[16] N. P. L. Santiyasa, I. M. A. W. Putra, dan I. K. G. D. Putra. (2022). "Peningkatan Akurasi Klasifikasi Spam Email menggunakan Algoritma Naïve Bayes dan TF-IDF". Jurnal Ilmiah Teknologi Informasi dan Komunikasi (JITIK), 9(1), 34-41.
[17] . R. Prasetyo, B. S. Abbas, dan D. P. Lestari. (2022). "Perancangan Aplikasi Filter Email Spam menggunakan Algoritma Naïve Bayes pada Lingkungan Kerja". Prosiding Konferensi Nasional Teknologi Informasi dan Komputer (KNIT), 4(1), 211-218.
[18] S. D. Pramudita dan A. W. Widodo. (2023). "Klasifikasi Email Spam dengan Kombinasi Algoritma Naïve Bayes dan K-Nearest Neighbor". Jurnal Sains dan Informatika (JSI), 9(1), 67-75.
[19] J. Pangaribuan dan M. T. Furqon. (2023). "Evaluasi Kinerja Algoritma Naïve Bayes untuk Deteksi Spam Email dengan Variasi Pra-pemrosesan Teks". Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIIK), 7(4), 1885-1892.
[20] K. Dewi dan A. R. Isnanto. (2023). "Sistem Klasifikasi Email Spam Menggunakan Algoritma Naïve Bayes dengan Seleksi Fitur Information Gain". Jurnal Transformatika, 20(2), 156-164.