PERANCANGAN SISTEM KECERDASAN BUATAN (AI) DALAM DETEKSI EMAIL SPAM

Ismet Maulana Azhari[1], Gathan Rafii Manaf[2], Aura Salsa Azzahra[3], Andiko Ramadani[4]

Program Studi Informatika Fakultas Teknik

Universitas Sultan Ageng Tirtayasa

Jl. Jenderal Sudirman Km 3, Kotabumi, Kec. Purwakarta, Kota Cilegon, Banten 42435

Abstrak

Seiring dengan meningkatnya penggunaan email sebagai media komunikasi utama, penyebaran spam email menjadi salah satu tantangan terbesar dalam menjaga keamanan dan kenyamanan pengguna. Email spam tidak hanya mengganggu, tetapi juga dapat menjadi pintu masuk bagi ancaman yang lebih serius seperti phishing, malware, dan pencurian data. Oleh karena itu, dibutuhkan sistem yang mampu mendeteksi dan memfilter spam secara otomatis dengan tingkat akurasi yang tinggi. Penelitian ini merancang sebuah sistem deteksi spam email berbasis klasifikasi teks yang memanfaatkan pendekatan Natural Language Processing (NLP) dan algoritma klasifikasi. Dalam perancangannya, sistem ini menganalisis isi teks dari email dan mengekstraksi fitur-fitur penting menggunakan teknik NLP seperti tokenisasi, stopword removal, dan TF-IDF. Kemudian, fitur tersebut diklasifikasikan menggunakan model machine learning untuk membedakan antara spam dan non-spam. Dataset yang digunakan adalah SpamAssassin, yang telah banyak digunakan dalam studi sebelumnya dan memiliki karakteristik teks yang relevan. Sistem ini diharapkan mampu meningkatkan efektivitas deteksi spam secara otomatis, serta dapat diintegrasikan ke dalam layanan email modern guna meningkatkan proteksi terhadap ancaman siber yang bersumber dari pesan email.

Kata Kunci:

I. PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital yang begitu pesat telah mendorong penggunaan email menjadi bagian penting dalam komunikasi modern, baik untuk keperluan personal, profesional, maupun institusional. Namun, di balik manfaatnya yang besar, email juga menjadi sasaran empuk bagi praktik-praktik penyalahgunaan, salah satunya adalah penyebaran spam. Spam email, yang sering dikirim dalam jumlah besar secara otomatis, kerap kali membawa konten yang menyesatkan, tidak diinginkan, atau bahkan berbahaya, seperti tautan phishing, perangkat lunak berbahaya (malware), dan penipuan digital lainnya[1].

Deteksi dan penyaringan spam menjadi tantangan yang terus

berkembang, mengingat para pelaku spam terus memperbarui strategi mereka agar dapat menghindari sistem penyaringan konvensional. Oleh karena itu, dibutuhkan pendekatan yang lebih cerdas dan adaptif untuk mengidentifikasi dan memfilter pesan-pesan yang mencurigakan secara otomatis. Dalam konteks ini, penerapan teknologi Kecerdasan Buatan (Artificial Intelligence) khususnya dalam bidang Natural Language Processing (NLP) dan klasifikasi teks menjadi solusi yang menjanjikan.

Natural Language Processing memungkinkan komputer untuk memahami, memproses, dan menganalisis bahasa manusia secara efektif. Dengan memanfaatkan NLP, sistem dapat mengenali pola bahasa, struktur kalimat, serta konteks yang sering muncul dalam spam email. Sementara itu, teknik klasifikasi teks, seperti Naïve Bayes, Support Vector Machine (SVM), Decision Tree, hingga algoritma berbasis deep learning seperti Recurrent Neural Network (RNN), mampu mengelompokkan email ke dalam kategori spam atau non-spam berdasarkan fitur-fitur teks yang diekstraksi.

Melalui proposal ini, kami bermaksud untuk merancang dan mengembangkan sebuah sistem deteksi spam email berbasis klasifikasi teks dan NLP. Sistem ini diharapkan mampu meningkatkan akurasi dalam membedakan email spam dan non-spam secara otomatis, sehingga memberikan perlindungan yang lebih baik terhadap pengguna dan menjaga keamanan komunikasi digital. Inovasi ini juga berkontribusi dalam pengembangan teknologi kecerdasan buatan untuk menangani masalah nyata di era digital.

1.2 Rumusan Masalah

1. Bagaimana merancang sistem deteksi email spam menggunakan metode klasifikasi teks dan teknik Natural Language Processing (NLP)?
2. Fitur-fitur apa saja yang relevan dan efektif untuk digunakan dalam proses klasifikasi spam email berbasis analisis teks?
3. Seberapa efektif sistem deteksi spam email yang dirancang dalam mengidentifikasi pesan spam dibandingkan dengan metode konvensional?

1.3 Tujuan Masalah

1. Merancang sistem deteksi email menggunakan metode `klasifikasi teks dan teknik NLP sebagai solusi dalam meningkatkan keamanan digital.
2. Mengidentifikasi dan mengevaluasi fitur-fitur teks yang paling efektif dalam membedakan email spam dan non-spam.
3. Mengukur performa dan tingkat akurasi sistem deteksi spam yang dikembangkan, serta membandingkannya dengan metode deteksi spam konvensional.

1.4 Batasan Masalah

1. Penelitian ini hanya berfokus pada perancangan dan pengembangan sistem deteksi spam email, tanpa mengintegrasikan sistem secara langsung ke dalam layanan email seperti Gmail atau Outlook.
2. Data yang digunakan untuk pelatihan dan pengujian sistem diambil dari dataset publik yang tersedia secara online (misalnya: SpamAssassin, Enron, atau dataset serupa), bukan dari email pribadi atau institusi.
3. Sistem deteksi dalam email, seperti spam hanya memproses konten teksrti subject dan body, tanpa mempertimbangkan lampiran (attachments), metadata lain, atau gambar.
4. Teknik yang digunakan dalam sistem terbatas pada klasifikasi teks berbasis Natural Language Processing (NLP) dan algoritma machine learning tertentu (seperti Naïve Bayes, SVM, atau Decision Tree), tanpa menggunakan pendekatan deep learning lanjutan seperti LSTM atau transformer.
5. Penelitian ini hanya sampai pada tahap perancangan dan evaluasi performa model deteksi spam, dan tidak mencakup pengembangan antarmuka pengguna (UI/UX) atau integrasi dengan sistem keamanan email lainnya.

II. TINJAUAN PUSTAKA

2.1 Konsep Email-Spam

Deteksi spam email merupakan salah satu tantangan penting dalam keamanan siber. Menurut Shirvani & Ghasemshirazi (2025), pendekatan berbasis Zero-Shot Learning memberikan fleksibilitas dalam mengidentifikasi pola spam baru tanpa pelatihan ulang. Sementara itu, penelitian oleh Keat dan Ying (2025) membandingkan algoritma klasik seperti Naive Bayes dan K-Nearest Neighbor, dan mencatat bahwa Naive Bayes memberikan akurasi sebesar 98,65% dalam klasifikasi spam. Akan tetapi, sebagian besar penelitian masih bergantung pada dataset berlabel besar, yang membatasi adaptabilitas sistem dalam kondisi nyata. Oleh karena itu, penelitian ini mencoba mengembangkan sistem deteksi spam yang adaptif dan ringan dengan pendekatan pembelajaran mesin modern. [2]

2.2 Perbandingan Beberapa Algoritma ML untuk Klasifikasi Email

Tiwari dan Singh (2020) membandingkan Naive Bayes, Support Vector Machine (SVM), Decision Tree, dan Random Forest.

Beberapa algoritma machine learning untuk klasifikasi email menyatakan bahwa algoritma Naive Bayes menunjukkan performa unggul dalam klasifikasi email spam dengan tingkat akurasi tinggi, didukung oleh proses pra-pemrosesan seperti tokenisasi dan vektorisasi TF-IDF, dengan

Pra-pemrosesan dilakukan dengan Tokenisasi, Stopword Removal, TF-IDF Vectorization. Naive Bayes menunjukkan akurasi tertinggi sekitar 96% karena mampu menangani teks pendek dengan baik[3]. SVM juga memiliki performa yang stabil namun lebih mahal secara komputasi. [4]

2.3 Algoritma Tradisional Learning VS Deep Learning

Sheneamer (2021) membandingkan efektivitas algoritma pembelajaran mesin tradisional dengan metode deep learning seperti Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) dalam klasifikasi email spam[5]. Dengan menggunakan model GloVe sebagai representasi kata, CNN terbukti menghasilkan akurasi tertinggi yaitu 96,52% pada dataset berisi 5.243 email spam dan 16.872 email non-spam dan SMS. Penelitian ini menyimpulkan bahwa metode deep learning mampu secara otomatis mengekstrak fitur dari data teks dan memiliki keunggulan signifikan dalam hal presisi, recall, dan akurasi dibanding pendekatan tradisional. [6]

III. METODOLOGI DAN DESAIN SISTEM

3.1 Gambaran Umum

Penelitian ini bertujuan untuk merancang sistem deteksi spam email dengan menggunakan metode klasifikasi teks dan Natural Language Processing (NLP). Sistem ini diharapkan mampu mengklasifikasikan pesan teks dari email ke dalam kategori spam dan non-spam (ham) dengan akurasi tinggi.

Secara umum, tahapan perancangan sistem mencakup pengumpulan dataset, preprocessing teks, ekstraksi fitur, pelatihan model klasifikasi, dan evaluasi performa.

3.2 Alur Sistem

Alur sistem yang dirancang dalam penelitian ini dapat dijelaskan melalui tahapan sebagai berikut.

1. **Pengumpulan Data**

   Dataset yang digunakan dalam penelitian ini adalah SpamAssassin, yaitu kumpulan email spam dan ham yang telah dikurasi dan berlabel. Dataset ini bersifat open-source dan telah banyak digunakan dalam penelitian klasifikasi email spam.

2. **Preprocessing Data**

   Untuk menyiapkan data mentah ke dalam bentuk yang dapat diproses oleh algoritma machine learning, dilakukan preprocessing melalui beberapa tahap berikut.

   - Lowercasing, yaitu mengubah seluruh teks menjadi huruf kecil.
   - Penghapusan karakter khusus dan tanda baca.
   - Tokenisasi, yaitu memecah kalimat menjadi kata-kata individual.

- Stopword removal, yaitu menghapus kata-kata umum yang tidak bermakna.
- Stemming, yaitu mengubah kata ke bentuk dasar.

3. Ekstraksi Fitur

Setelah teks dibersihkan, dilakukan transformasi menjadi bentuk numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Representasi ini memperhatikan seberapa penting suatu kata dalam satu email dibandingkan dengan keseluruhan dokumen, sehingga efektif untuk membedakan kata-kata khas spam dari ham.

4. Pelatihan Model Klasifikasi

Teknik yang digunakan dalam sistem ini terbatas pada pendekatan klasifikasi teks berbasis Natural Language Processing (NLP) dan algoritma machine learning tradisional.

Beberapa algoritma machine learning yang berpotensi dikembangkan dalam sistem ini, antara lain:

- Naïve Bayes
- Support Vector Machine (SVM)
- Decision Tree

5. Evaluasi Hasil

Model yang telah dilatih akan diuji menggunakan metrik evaluasi untuk mengukur kinerja klasifikasi. Hasil evaluasi akan menentukan model mana yang paling optimal untuk diterapkan dalam deteksi spam email berbasis teks.

IV. RENCANA EVALUASI DAN DATASET

4.1 Dataset yang Digunakan

Dalam penelitian ini, digunakan dataset SpamAssassin, yaitu kumpulan email spam dan non-spam yang telah diberi label dan tersedia secara open-source. Dataset ini dipilih karena bersifat publik, telah digunakan secara luas dalam penelitian klasifikasi spam, serta memiliki struktur data yang konsisten dan representatif terhadap email-email yang umum diterima pengguna.

SpamAssassin mencakup berbagai jenis email, baik yang mengandung iklan komersial, penipuan, maupun email biasa, sehingga cocok untuk membangun model klasifikasi teks yang andal.

4.2 Metode Evaluasi

Untuk mengukur performa sistem klasifikasi spam yang dibangun, digunakan beberapa metrik evaluasi yang umum dalam klasifikasi biner, diantaranya:

- **Akurasi (Accuracy)**: Mengukur seberapa banyak prediksi yang benar dibandingkan seluruh data.
- **Presisi (Precision)**: Menilai proporsi email yang diprediksi sebagai spam dan benar-benar spam.
- **Recall (Sensitivity)**: Mengukur kemampuan sistem dalam mendeteksi semua email spam yang ada.
- **F1-Score**: Merupakan rata-rata harmonis dari presisi dan recall, memberikan evaluasi yang seimbang.

Evaluasi dilakukan menggunakan teknik train-test split dengan proporsi 80:20, di mana 80% data digunakan untuk pelatihan model dan 20% sisanya untuk pengujian. Adapun cross-validation akan dilakukan jika perlu, untuk memastikan kestabilan performa model pada subset data yang berbeda.

V. KESIMPULAN DAN SARAN

Proposal ini telah menguraikan rencana perancangan sistem deteksi spam email berbasis kecerdasan buatan dengan pendekatan klasifikasi teks dan Natural Language Processing (NLP). Sistem dirancang melalui tahapan pengumpulan data, preprocessing teks, ekstraksi fitur, pelatihan model, dan evaluasi. Dengan menggunakan dataset SpamAssassin dan algoritma pembelajaran mesin yang sesuai, diharapkan sistem yang dibangun mampu mengidentifikasi email spam secara efektif dan efisien.

Adapun saran untuk proposal maupun penelitian berlanjut ialah sebagai berikut:

1. Diperlukan pengujian lebih lanjut untuk membandingkan berbagai algoritma klasifikasi agar diperoleh hasil yang optimal.
2. Dataset yang digunakan sebaiknya divalidasi kembali agar sesuai dengan kondisi email terkini yang lebih kompleks.
3. Pengembangan lebih lanjut dapat diarahkan ke integrasi sistem dalam lingkungan nyata seperti aplikasi email klien atau server.
4. Penelitian lanjutan dapat mengeksplorasi penggunaan metode deep learning untuk meningkatkan performa klasifikasi.

DAFTAR PUSTAKA

[1] L. C. Keat and T. X. Ying, “Artificial intelligence-based email spam filtering,” Zenodo. [Online]. Available: https://zenodo.org/doi/10.5281/zenodo.14264139

[2] P. S. Priyanka, M. Gnanadas, and P. Supriya, “Standalone Portable Host for Unified Bootloader in PIC devices using CAN interface,” in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2020, pp. 1–4. Accessed:

May 27, 2025. [Online]. Available: https://doi.org/10.1109/icccnt49239.2020.9225574

[3] A. Sheneamer, “Comparison of Deep and Traditional Learning Methods for Email Spam Filtering,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021, doi: 10.14569/ijacsa.2021.0120164.

[4] M. S. Akhtar and T. Feng, “Malware Analysis and Detection Using Machine Learning Algorithms,” *Symmetry*, vol. 14, no. 11, p. 2304, Nov. 2022, doi: 10.3390/sym14112304.

[5] A. Karim, M. Shahroz, K. Mustofa, S. B. Belhaouari, and S. R. K. Joga, “Phishing Detection System Through Hybrid Machine Learning Based on URL,” *IEEE Access*, vol. 11, pp. 36805–36822, 2023, doi: 10.1109/access.2023.3252366.

[6] M. Koca, İ. Avci, and M. A. S. Al-Hayani, “Classification of Malicious URLs Using Naive Bayes and Genetic Algorithm,” *Sakarya University Journal of Computer and Information Sciences*, vol. 6, no. 2, pp. 80–90, Aug. 2023, doi: 10.35377/saucis...1273536.