



Received 00th January 20xx
Accepted 00th February 20xx
Published 00th March 20xx

Open Access

DOI: 10.35472/x0xx0000

Klasifikasi Spam Email Mahasiswa ITERA Menggunakan Metode *Hybrid Attention LSTM*

Natasya Ega Lina Marbun^{1*}^a, Eksanty F Sugma Islamiaty^{2b}, Muhammad Regi Abdi Putra Amanta^{3c}

^anatasya.122450024@student.itera.ac.id

^beksanty.122450001@student.itera.ac.id

^cmuhammad.122450031@student.itera.ac.id

* Corresponding E-mail: natasya.122450024@student.itera.ac.id

Abstract: Email is a primary communication medium in academic environments; however, its increasing usage has been accompanied by a growing volume of spam emails, which can disrupt communication effectiveness and pose potential security risks. This study aims to classify spam emails from students of the Institut Teknologi Sumatera (ITERA) using a Hybrid Attention Long Short-Term Memory (LSTM) approach. The proposed model integrates LSTM-based text processing with a temporal attention mechanism to emphasize informative words, while simultaneously incorporating numerical email features through a separate neural network pathway. These heterogeneous features are fused within a unified architecture to enhance classification performance. Model evaluation is conducted using a confusion matrix and ROC-AUC analysis. Experimental results demonstrate that the Hybrid Attention LSTM achieves excellent performance, with an AUC score of 0.9574. Additionally, attention heatmap visualizations are employed to identify class-specific keywords, confirming the effectiveness of the attention mechanism. Overall, the proposed approach provides an effective solution for spam email filtering in academic environments.

Keywords: Dense, Deep Learning, Long Short-Term Memory

Abstrak: Email merupakan media komunikasi utama di lingkungan akademik, namun peningkatan penggunaannya juga diikuti oleh bertambahnya email spam yang dapat mengganggu efektivitas komunikasi dan berpotensi menimbulkan risiko keamanan digital. Penelitian ini bertujuan untuk mengklasifikasikan spam pada email mahasiswa Institut Teknologi Sumatera (ITERA) menggunakan metode *Hybrid Attention Long Short-Term Memory* (LSTM). Model yang diusulkan mengombinasikan pemrosesan teks email berbasis LSTM dengan *temporal attention mechanism* untuk memperhatikan kata-kata penting, serta pemanfaatan fitur numerik email melalui jalur jaringan saraf terpisah. Kedua jenis fitur tersebut digabungkan dalam satu arsitektur model untuk meningkatkan kinerja klasifikasi. Evaluasi dilakukan menggunakan *confusion matrix*, dan ROC-AUC. Hasil pengujian menunjukkan bahwa model *Hybrid Attention LSTM* mencapai performa yang sangat baik dengan nilai AUC sebesar 0.9574. Visualisasi *attention heatmap* digunakan untuk mengidentifikasi kata-kata kunci yang relevan pada masing-masing kelas. Dengan demikian, pendekatan yang diusulkan efektif untuk mendukung penyaringan spam email di lingkungan akademik.

Kata Kunci : Dense, Deep Learning, Long Short-Term Memory





Pendahuluan

Surat elektronik (email) merupakan media komunikasi utama dalam lingkungan akademik, termasuk di Institut Teknologi Sumatera (ITERA). Seiring meningkatnya intensitas penggunaan email, mahasiswa juga semakin sering menerima email spam yang tidak relevan dengan aktivitas akademik dan berpotensi mengganggu efektivitas komunikasi serta keamanan digital pengguna [1], [2]. Oleh karena itu, diperlukan sistem klasifikasi spam yang mampu bekerja secara akurat dan adaptif.

Berbagai metode *machine learning*, seperti Naïve Bayes, *Support Vector Machine* (SVM), dan *Decision Tree*, telah digunakan untuk mendeteksi email spam. Meskipun menunjukkan kinerja yang cukup baik, metode tersebut masih memiliki keterbatasan dalam menangani kompleksitas data teks, perubahan pola spam yang dinamis, serta risiko kesalahan klasifikasi yang relatif tinggi [3]. Selain itu, metode klasifikasi umumnya hanya memanfaatkan fitur teks atau fitur statistik secara terpisah, sehingga belum sepenuhnya merepresentasikan karakteristik email spam secara menyeluruh.

Perkembangan *deep learning*, khususnya *Long Short-Term Memory* (LSTM), memiliki kemampuan yang lebih baik dalam memahami urutan kata pada teks email [4]. Namun, LSTM berbasis teks saja masih memiliki keterbatasan karena belum mampu membedakan tingkat prioritas setiap kata secara optimal serta belum mempertimbangkan fitur numerik email, seperti panjang pesan, jumlah tautan, dan karakter khusus, yang sering menjadi indikator penting spam.

Untuk mengatasi keterbatasan tersebut, attention mechanism digunakan agar model dapat memperhatikan bagian teks yang paling penting [5]. Selain itu, integrasi fitur numerik melalui jaringan saraf

dense digunakan untuk mendapatkan informasi email lainnya yang tidak ada dalam representasi teks. Oleh karena itu, penelitian ini mengombinasikan pemrosesan teks berbasis LSTM dan *attention mechanism* dengan analisis fitur numerik melalui dense layer untuk klasifikasi spam email mahasiswa ITERA. Pendekatan ini diharapkan mampu meningkatkan akurasi dan kemampuan generalisasi model dalam membedakan email spam dan ham secara lebih efektif.

Method / Metode

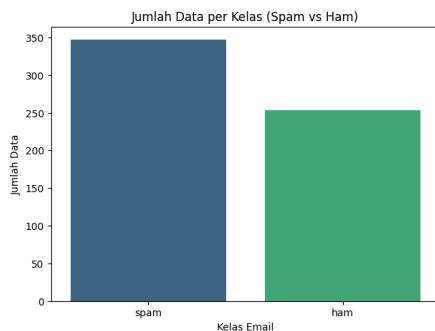
Deskripsi Dataset

Data yang digunakan dalam penelitian ini merupakan data primer berupa email mahasiswa ITERA yang telah diberi label sebagai ham (0) dan spam (1) untuk keperluan klasifikasi biner menggunakan metode *Hybrid Attention LSTM*. Dataset ini terdiri atas sembilan atribut, yaitu 'id', 'label', 'subject', 'body', 'from_domain', 'num_urls', 'num_exclaim', 'has_attachment', dan 'body_len'. Seluruh dataset yang digunakan dalam penelitian ini dapat diakses melalui tautan bit.ly/dataset_email. Selain itu, peneliti menambahkan dua atribut tambahan, yaitu 'num_special_chars' dan 'avg_word_len', yang masing-masing merepresentasikan jumlah karakter khusus dan rata-rata panjang kata dalam email. Atribut tambahan tersebut dimanfaatkan sebagai fitur numerik untuk mendukung proses analisis serta meningkatkan akurasi klasifikasi.

Eksplorasi Data

Analisis data eksploratori (EDA) dilakukan terlebih dahulu pada dataset email untuk mengenali pola dan karakteristik data sebelum model dilatih. Jumlah email pada masing-masing kelas ditunjukkan pada Gambar 1 berikut.





Gambar 1. Distribusi Jumlah Email Berdasarkan Kelas

Gambar 1 menunjukkan distribusi jumlah data pada setiap kelas yang berbeda. Kelas spam memiliki sebanyak 347 data, sedangkan kelas ham berjumlah 253 data. Dengan demikian, total email yang digunakan dalam penelitian ini adalah 600 data. Meskipun distribusi data antar kelas tidak sepenuhnya seimbang, perbedaannya masih tergolong ringan dan tidak signifikan. Oleh karena itu, dataset ini tetap layak digunakan tanpa teknik penyeimbangan tambahan. Gambar 2 menunjukkan kata terbanyak yang muncul pada email spam dan ham.



Gambar 2. Kata Terbanyak Muncul pada Email

Berdasarkan Gambar 2 kata-kata yang paling sering muncul pada email spam didominasi oleh ‘ai’, ‘neo’, ‘graph’, dan ‘data’, artinya email spam cenderung berisi konten promosi tertentu. Sementara itu, email ham lebih banyak memuat kata-kata terkait kegiatan akademik seperti ‘new’, ‘assignment’, ‘email’, ‘google’, dan ‘tugas’. Gambar 3 berikut menampilkan distribusi email berdasarkan panjang isi email.

Pemrosesan Data

Prapemrosesan data teks dilakukan melalui langkah-langkah berikut [6]:

1. Pemberian label pada setiap data email.

2. Penggabungan kolom *subject* dan *body* untuk merepresentasikan konteks email secara utuh.
3. Normalisasi teks dengan mengubah huruf menjadi kecil serta menghapus URL, angka, dan karakter khusus.
4. *Stopword removal* dan *stemming* menggunakan Sastrawi untuk bahasa Indonesia.
5. Tokenisasi menggunakan Keras Tokenizer dengan batas kosakata 5.000 kata.
6. Penyesuaian panjang sekuens menjadi 100 token menggunakan padding.

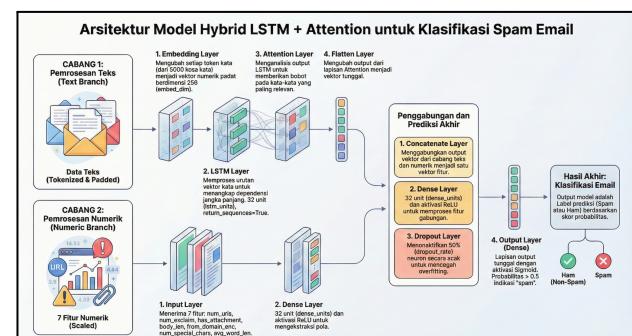
Prapemrosesan data numerik dilakukan melalui langkah-langkah berikut [1]:

1. Encoding fitur kategorikal, seperti domain pengirim email.
2. Normalisasi seluruh fitur numerik menggunakan MinMaxScaler agar berada pada rentang 0–1.

Selanjutnya, fitur teks dan numerik digabungkan menjadi satu vektor input. Dataset kemudian dibagi menjadi 80% data pelatihan dan 20% data pengujian untuk proses pelatihan dan evaluasi model.

Arsitektur Model

Secara garis besar, Gambar 3 berikut adalah arsitektur model yang digunakan dalam penelitian ini.



Gambar 3. Arsitektur yang Digunakan

1) Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan salah satu varian *Recurrent Neural Network* (RNN) yang dirancang untuk memproses data berurutan, seperti teks, dengan kemampuan menyimpan informasi jangka panjang sehingga mampu mengatasi permasalahan *vanishing gradient* pada RNN biasa [6].

Struktur LSTM terdiri atas unit memori dengan tiga gerbang utama, yaitu forget gate untuk mengatur informasi yang dipertahankan, input gate untuk memasukkan informasi baru, dan output gate untuk menghasilkan keluaran berdasarkan kondisi memori. Operasi tiap gerbang dirumuskan sebagai berikut [2]:

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i)$$

$$\tilde{c}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \times \tilde{c}_t$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \odot \tanh(c_t)$$

dengan x_t adalah data input, h_t dan c_t adalah hidden dan output state, W adalah bobot, dan b adalah bias.

2) Dense Neural Network (DNN)

Lapisan *fully-connected* pada DNN digunakan untuk mengintegrasikan seluruh fitur yang dihasilkan dari lapisan sebelumnya. Pada arsitektur ini, keluaran LSTM dari fitur teks dan keluaran Dense dari fitur numerik digabungkan (*concatenate*) menjadi satu vektor fitur, kemudian diproses melalui beberapa lapisan DNN dengan fungsi aktivasi ReLU. Pada tahap akhir, lapisan output menggunakan fungsi aktivasi sigmoid untuk menghasilkan nilai probabilitas kelas biner pada rentang [0,1]. Secara umum, proses pada DNN dirumuskan sebagai berikut [1]:

$$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}$$

$$a^{(l)} = f(z^{(l)})$$

dengan f adalah fungsi aktivasi berupa ReLU atau Sigmoid sedangkan $a^{(0)}$ adalah input vektor fitur.

Attention Mechanism

Attention mechanism digunakan untuk menyoroti bagian teks email yang paling relevan dalam klasifikasi spam dan ham. Setiap keluaran LSTM h_t diberi skor kepentingan $e_t = g(h_t) = v^T \tanh(Wh_t + b)$ kemudian dinormalisasi dengan fungsi softmax untuk memperoleh bobot *attention* α_t . Representasi akhir teks (*context vector*) dihitung sebagai $c = \sum_{t=1}^T \alpha_t h_t$, yang merepresentasikan informasi penting dari seluruh urutan teks. *Context vector* ini selanjutnya digabungkan dengan fitur numerik sebelum diproses oleh lapisan klasifikasi. *Attention* yang digunakan merupakan *temporal attention* karena bobot diberikan pada setiap timestep keluaran LSTM.

Binary Cross Entropy Loss (BCELoss)

BCELoss digunakan sebagai fungsi objektif untuk mengukur kesalahan prediksi pada klasifikasi biner spam dan ham. Fungsi ini memberikan penalti besar terhadap prediksi yang salah dengan tingkat keyakinan tinggi, dan nilai *loss* yang mendekati nol jika prediksi sesuai dengan label sebenarnya.

Optimasi Adaptive Moment Estimation (ADAM)

Optimizer Adam digunakan karena kemampuannya menyesuaikan laju pembelajaran secara adaptif dan efisien dengan mengombinasikan pendekatan Momentum dan RMSProp. Adam memperbarui bobot berdasarkan estimasi momen pertama dan kedua gradien, yang dihitung

menggunakan persamaan m_t dan v_t kemudian dikoreksi bias sebelum dilakukan pembaruan parameter. Pendekatan ini membantu mempercepat konvergensi dan meningkatkan stabilitas pelatihan model.

Regularisasi

Regularisasi diterapkan untuk mencegah *overfitting* pada model. Dropout digunakan pada lapisan dense dengan probabilitas 0,5 untuk mengurangi ketergantungan antar neuron selama pelatihan. Selain itu, *early stopping* diterapkan dengan menghentikan pelatihan jika nilai *validation loss* tidak membaik selama tiga epoch berturut-turut. Normalisasi *batch* digunakan secara opsional setelah fungsi aktivasi untuk menstabilkan proses pelatihan dan mempercepat konvergensi.

Hyperparameter

Hyperparameter yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1 berdasarkan hasil terbaik.

Tabel 1. Nilai *hyperparameter* yang digunakan

Hyperparameter	Nilai
Dimensi <i>Embedding</i>	256
Jumlah Unit LSTM	32
Jumlah Unit Dense	32
<i>Dropout Rate</i>	0.5
<i>Learning Rate</i>	0.0005

Hasil dan Pembahasan

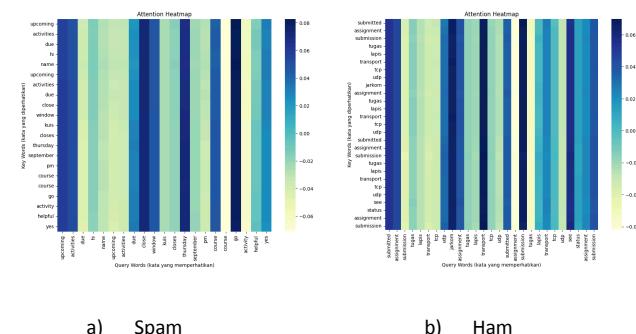
Hasil

Gambar 4 menunjukkan konvergensi model, ditandai dengan akurasi meningkat dan *loss* berkurang secara konsisten sebelum kondisi *Early Stopping* terpenuhi.



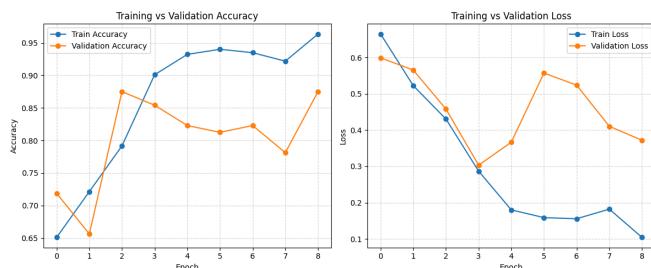
Gambar 4. Log Perkembangan Kinerja Model

Hasil *attention mechanism* ditunjukkan pada Gambar 5, dengan subgambar (a) merepresentasikan kelas spam dan subgambar (b) merepresentasikan kelas ham.



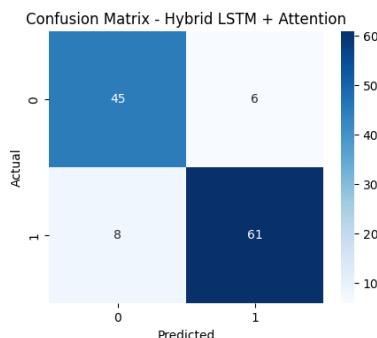
Gambar 5. Visualisasi Attention Heatmap pada Model Hybrid Attention LSTM

Untuk menilai kemampuan generalisasi model dan mendeteksi indikasi *overfitting*, Gambar 6 memvisualisasikan perbandingan antara Akurasi dan Fungsi *Loss* pada set data pelatihan (*Training*) dan validasi (*Validation*) di setiap epoch.



Gambar 6. Kurva Nilai Akurasi

Gambar 7 menunjukkan Confusion Matrix berdasarkan model *Hybrid Attention LSTM*. Matriks ini merupakan dasar perhitungan Akurasi, Presisi, Recall, dan F1-Score.



Gambar 7. Confusion Matrix

Pembahasan

Model *Hybrid Attention LSTM* dilatih menggunakan *batch size* 32 selama maksimum 100 *epoch* dengan *early stopping* (*patience* = 5) untuk mencegah *overfitting*. Kurva akurasi dan loss menunjukkan peningkatan kinerja yang konsisten, dengan akurasi pelatihan mencapai sekitar 0,95 dan loss menurun hingga mendekati 0,1, sementara performa validasi tetap stabil, menandakan kemampuan generalisasi yang baik.

Attention weight menunjukkan bahwa model mampu menyoroti kata-kata kunci yang relevan pada masing-masing kelas, dengan pola perhatian yang berbeda.. Evaluasi kuantitatif menggunakan *Confusion Matrix* menghasilkan nilai prediksi yang tinggi pada kedua kelas.

Analisis korelasi fitur numerik memperlihatkan hubungan kuat antara ‘body_len’ dan ‘num_special_chars’, serta antara ‘num_urls’ dan ‘num_exclaim’, yang mencerminkan karakteristik umum email spam. Sebaliknya, fitur ‘has_attachment’ memiliki korelasi lemah dengan fitur lainnya sehingga memberikan informasi yang relatif independen dalam model.

Kesimpulan

Penelitian ini berhasil menerapkan model *Hybrid Attention LSTM* untuk mengklasifikasikan spam pada email mahasiswa ITERA dengan kinerja yang sangat baik. Hasil pelatihan dan evaluasi menunjukkan bahwa gabungan metode attention mechanism dengan pemrosesan teks berbasis LSTM serta fitur numerik email mampu meningkatkan kemampuan model dalam mengenali pola penting pada email spam dan ham. Nilai akurasi dan AUC yang tinggi menunjukkan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat. Selain itu, visualisasi attention weight membuktikan bahwa model dapat menyoroti kata-kata kunci yang relevan dalam proses klasifikasi. Dengan demikian, pendekatan yang diusulkan efektif untuk mendukung penyaringan email dan meningkatkan keamanan komunikasi digital di lingkungan akademik ITERA.

Konflik Kepentingan

Penulis menyatakan bahwa tidak terdapat kepentingan pribadi maupun finansial yang berpotensi menimbulkan konflik kepentingan terkait dengan topik yang dibahas dalam naskah ini.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Institut Teknologi Sumatera (ITERA) atas dukungan fasilitas dan lingkungan akademik yang menunjang pelaksanaan penelitian ini. Ucapan terima kasih juga

disampaikan kepada pihak-pihak yang telah membantu dalam proses pengumpulan data serta memberikan masukan dan saran yang berharga selama penyusunan penelitian ini. Kontribusi dan dukungan tersebut sangat berarti dalam meningkatkan kualitas penelitian yang dilakukan.

Daftar Pustaka

- [1] K. S. Ubale and K. A. Shirath, "SPAMNET: A Hybrid Deep Learning Framework for Robust Spam Email Detection Using Multi-Modal Features," *International Journal of Applied Mathematics*, vol. 38, no. 6s, Oct. 15, 2025. doi: 10.12732/ijam.v38i6s.425
- [2] Maugy Al Kautsar, Galet Guntoro Setiaji, and Ahmad Rifa'i, "Analisis Komparasi Kinerja LSTM dan CNN dalam Deteksi Spam Email Berbasis Deep learning", *bulletincsr*, vol. 5, no. 4, pp. 584-593, Jun. 2025.
- [3] S. Aleem, Z. U. Islam, S. S. U. Hasan, H. Akbar, M. F. Khan, and S. A. Ibrar, "Spam Email Detection Using Long Short-Term Memory and Gated Recurrent Unit," *Applied Sciences*, vol. 15, no. 13, p. 7407, 2025, doi: 10.3390/app15137407.
- [4] A. Tholib, N. K. Agusmawati, and F. Khairiyah, "Prediksi Harga Emas Menggunakan Metode LSTM dan GRU," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, pp. 194–203, 2023, doi: 10.23960/jitet.v11i3.3250.
- [5] S. A. Anggara, W. Witanti, dan Melina, "Peramalan Harga Cabai Rawit Merah Menggunakan Attention Mechanism Berbasis Long Short-Term Memory," *Journal of Applied Computer Science and Technology (JACOST)*, vol. 5, no. 2, pp. 128–135, 2024, doi: 10.52158/jacost.v5i2.875.
- [6] MR Adepu Rajesh and Dr Tryambak Hiwarkar, "Exploring Preprocessing Techniques for Natural LanguageText: A Comprehensive Study Using Python Code," *Int. J. Eng. Technol. Manag. Sci.*, vol. 7, no. 5, pp. 390–399, 2023, doi: 10.46647/ijetms.2023.v07i05.047.
- [7] T. Zhang, K. Zhang, dan J. Wu, "Multi-modal attention mechanisms in LSTM and its application to acoustic scene classification," in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 530–534.
- [8] A. Hanifa, S. A. Fauzan, M. Hikal, and M. B. Ashfiya, "Perbandingan Metode LSTM dan GRU (RNN) untuk Klasifikasi Berita Palsu Berbahasa Indonesia," *Din. Rekayasa*, vol. 17, no. 1, p. 33, 2021, doi: 10.20884/1.dr.2021.17.1.436.