DETEKSI SPAM EMAIL MENGGUNAKAN METODE LSTM (LONG SHORT TERM MEMORY)

Moch Wahyu Sampurno Utomo, Hapsoro Wisnu Murti,
Amanda Widya Indah Sujatmoko, Anggraini Puspita Sari
Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jalan Raya Rungkut Madya No.1, Gunung Anyar, Surabaya, Indonesia
anggraini.puspita.if@upnjatim.ac.id

ABSTRAK

Masalah spam email tetap menjadi ancaman serius bagi keamanan dan efisiensi komunikasi digital. Metode deteksi spam konvensional sering kali gagal mengidentifikasi taktik spam yang semakin canggih, seperti penggunaan bahasa ambigu dan manipulasi teks. Hal ini menyebabkan lolosnya email berbahaya yang dapat mengakibatkan pencurian data sensitif dan kerugian finansial. Untuk mengatasi tantangan ini, diperlukan pendekatan yang lebih adaptif dan mampu memahami konteks linguistik yang kompleks. Tujuan dari penelitian ini adalah untuk meningkatkan akurasi deteksi email spam dengan menggunakan pendekatan *Long Short Term Memory* (LSTM), yang merupakan jenis dari *Recurrent Neural Network* (RNN). Untuk mencapai hal ini, kami menggunakan kumpulan data yang terdiri dari beragam teks email yang bersumber dari korpus SpamAssassin. Data menjalani pemrosesan teks, tokenisasi untuk mengubah teks menjadi urutan angka, yang kemudian dipadukan agar memiliki panjang yang seragam, dan kemudian melakukan padding pada urutan untuk menstandarkan panjang input. Melalui pelatihan dan evaluasi model LSTM, kami mengamati peningkatan substansial dalam akurasi deteksi mencapai 99.35%. Hasil ini menunjukkan kemampuan LSTM dalam menangani nuansa teks email, sehingga meningkatkan kinerja sistem deteksi spam secara signifikan. Kesimpulannya, pendekatan LSTM terbukti efektif dalam mengidentifikasi email spam dan dapat diadopsi sebagai solusi dalam sistem penyaringan email.

Kata kunci : *Deteksi Spam Email, LSTM, Pembelajaran Mesin, Klasifikasi Teks, SpamAssassin*

1. PENDAHULUAN

Penelitian ini mengkaji permasalahan yang telah menjadi tantangan dalam lingkungan internet, yaitu spam email. Spam email telah mengganggu pengguna internet selama bertahun-tahun dengan pengirim spam yang terus mengembangkan strategi baru untuk menghindari deteksi oleh filter konvensional berbasis aturan dan kata kunci. Strategi baru ini termasuk penggunaan bahasa yang ambigu dan teknik manipulasi lainnya untuk meningkatkan tingkat kelulusan spam yang dikirim. Pada saat ini, metode deteksi spam yang konvensional sering kali tidak mampu mengatasi spam yang semakin kompleks ini. Filter berbasis aturan dan kata kunci seringkali gagal mengidentifikasi spam yang menggunakan variasi bahasa, ejaan yang tidak standar, dan teknik manipulasi lainnya [1].

Menghadapi tantangan ini, diperlukan pendekatan yang lebih canggih dan adaptif untuk mengatasi masalah ini. Salah satu pendekatan yang menjanjikan adalah menggunakan teknologi Deep Learning, khususnya metode Long Short-Term Memory (LSTM) [10]. LSTM yang merupakan jenis dari Recurrent Neural Networks (RNN), memiliki kemampuan untuk memahami konteks dan pola yang kompleks dalam teks, termasuk bahasa yang ambigu dan variasi ejaan yang digunakan dalam spam email. Dengan menggunakan teknologi ini, sistem deteksi spam diharapkan dapat lebih adaptif dan mampu menyesuaikan diri dengan berbagai strategi baru yang terus berkembang dalam dunia spam [14].

Dengan menerapkan metode LSTM dalam deteksi spam email, diharapkan dapat meningkatkan akurasi dalam mengenali dan memfilter spam dengan lebih efektif. Pendekatan ini memungkinkan sistem untuk belajar dari data yang ada dan secara otomatis menyesuaikan diri dengan strategi baru yang digunakan oleh pengirim spam. Gap analysis yang ditemukan dalam penelitian ini adalah bahwa metode deteksi spam yang saat ini digunakan masih memiliki keterbatasan dalam mengatasi spam yang semakin kompleks dan canggih. Oleh karena itu, penerapan teknologi Deep Learning dan LSTM menjadi penting untuk mengisi kesenjangan ini dan meningkatkan efektivitas dalam mengatasi masalah spam email [4].

Tujuan utama dari penelitian ini adalah untuk mengembangkan sistem deteksi spam email yang menggunakan pendekatan Deep Learning, khususnya melalui penerapan metode LSTM. Dengan fokus pada peningkatan akurasi dan efektivitas dalam memfilter spam email yang semakin kompleks. Sistem yang dikembangkan diharapkan dapat menjadi lebih responsif dan adaptif terhadap berbagai perubahan dalam teknik spam, sehingga mampu menghadapi tantangan yang terus berkembang di dunia digital [8]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam meningkatkan keamanan dan kenyamanan pengguna internet dalam menghadapi masalah spam email yang terus berkembang.

2. TINJAUAN PUSTAKA

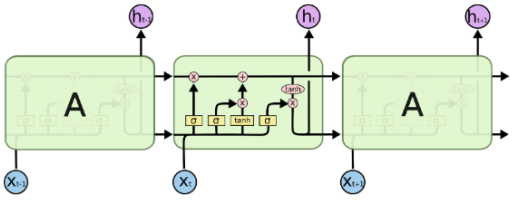
Metode lain seperti Support Vector Machine (SVM) dan Naive Bayes telah digunakan dalam deteksi spam email. SVM, teknik machine learning yang mencari hyperplane optimal untuk memisahkan dua kelas data, efektif dalam klasifikasi data berdimensi tinggi dan akurat dalam deteksi spam, namun sering memerlukan penyesuaian parameter yang rumit [11]. Sementara itu, Naive Bayes, algoritma probabilistik sederhana berdasarkan Teorema Bayes, efektif dalam klasifikasi teks email dengan kecepatan dan kemampuan menangani data besar. Namun, asumsi independensi fiturnya sering tidak realistis, yang dapat menyebabkan penurunan akurasi dalam beberapa kasus penyaringan email [16].

2.1. Deep Learning

Metode *Deep Learning* saat ini merupakan istilah yang paling umum dalam pembelajaran mesin untuk mengekstraksi representasi data kompleks secara otomatis pada tingkat abstraksi yang tinggi, terutama digunakan untuk masalah yang sangat kompleks. Ini adalah pendekatan yang membutuhkan data untuk menghasilkan hasil yang lebih baik daripada metode tradisional (*Naïve Bayes*, *SVM, HMM*, dan lain lain.). Selama korpora berbasis teks, model urutan deep learning lebih baik daripada metode feed-forward. Deep learning memiliki beberapa teknik yaitu *CNNs, RNNs, LSTM, GANs, DBNs, DRL* [2]. Deep Learning merupakan teknik modern terkini untuk pemrosesan citra dan analisis data, dengan hasil yang menjanjikan dan potensi yang besar. Karena pembelajaran mendalam telah berhasil diterapkan di berbagai domain [5].

2.2. Long Short Term Memory

Long Short-Term Memory (LSTM) adalah jenis arsitektur jaringan saraf tiruan yang efektif dalam mengatasi perhitungan jarak jauh dalam data beruntun. Dengan komponen seperti *Memory-Cell, Input Gate, Forget Gate*, dan *Forget Gate*, LSTM menyimpan informasi jangka panjang, mengatur output relevan, menambah informasi baru, dan menghapus yang tidak penting [9]. Ini menjadikan LSTM ideal untuk pemrosesan bahasa alami dan pemodelan waktu. Dalam penelitian, LSTM menunjukkan kinerja terbaik dalam klasifikasi email dengan akurasi maksimum 98,4%, dibandingkan dengan algoritma lain seperti *CNN, Naive Bayes, dan SVM* [15].



Gambar 1. Arsitektur LSTM

Gambar di atas menjelaskan bagaimana alur kerja memory cells pada setiap neuron LSTM beroperasi. LSTM meningkatkan kinerja RNN dengan menambahkan status sel yang berfungsi untuk mengelola ingatan dan pelupaan informasi secara dinamis. Proses ini dilakukan melalui empat gerbang utama: *Input Gate*, yang menentukan apakah informasi baru harus disimpan dalam sel; *Forget Gate*, yang memutuskan informasi mana dari keadaan tersembunyi sebelumnya yang harus dilupakan atau dipertahankan; *Memory-Cell State Gate*, yang memperbarui data dalam sel dengan mempertimbangkan input baru dan ingatan sebelumnya; serta *Output Gate*, yang menghasilkan output jaringan berdasarkan kondisi terkini dari memori sel. Mekanisme ini memungkinkan LSTM untuk mengelola informasi lebih efisien dibandingkan RNN konvensional [12].

2.3. Spam Email

Spam adalah informasi yang dikirimkan kepada banyak penerima tanpa persetujuan mereka. Filter spam adalah alat otomatis untuk mengenali dan mencegah pengiriman spam. Spam efektif jika berhasil menghindari filter, sementara filter efektif jika mampu mengenali spam [3]. Spam mencakup segala jenis komunikasi digital yang tidak diinginkan dan tidak diminta, seperti email dan SMS, yang membanjiri jaringan dan membuang sumber daya. Selain promosi produk oleh pengiklan, spam juga mencakup email phishing berbahaya yang berupaya mencuri informasi sensitif seperti login situs web atau informasi kartu kredit. Masalah ini tidak hanya mengganggu pengguna tetapi juga memiliki dampak finansial karena dapat menyebabkan pencurian identitas melalui *malware* dan konten *phishing* [13].

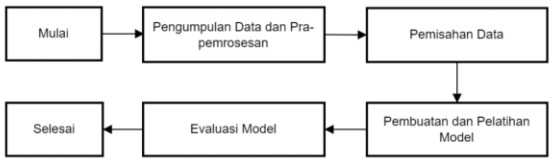
2.4. Penyaringan Email

Email filtering adalah proses otomatis yang mendeteksi apakah sebuah email termasuk legitimate atau bukan. Metode yang digunakan termasuk *Keyword filtering, Blacklisting and Whitelisting, Signature-based filtering, dan Naïve Bayesian filtering* [6]. Karakteristik utama dari email filtering adalah *Binary Class*, prediksi kelas email, komputasi mudah untuk data berdimensi tinggi, dan kemampuan untuk belajar dari email sebelumnya. Selain itu, email filtering harus mampu melakukan learning dari email email yang sudah ada sebelumnya. Kinerja yang bagus juga penting, dengan akurasi tinggi, meminimalkan false positive, dan mentolerir false negative [16].

3. METODE PENELITIAN

Pada tahap penelitian deteksi spam email menggunakan metode LSTM, dimulai dengan pengumpulan data dari dataset SpamAssassin. Data ini kemudian dipra-pemrosesan melalui tokenisasi dan padding sequences. Setelah itu, data dipisahkan menjadi data uji dan data latih. Model LSTM kemudian dibangun dan dilatih menggunakan data latih dengan penambahan early stopping untuk mencegah overfitting. Terakhir, model dievaluasi

menggunakan data pengujian untuk menentukan akurasi dan menghasilkan laporan klasifikasi.



Gambar 1. Langkah Penelitian

3.1. Pengumpulan Data

Proses pengumpulan data dimulai dengan mengunduh dataset SpamAssassin yang terdiri dari dua file terkompresi: 20021010_easy_ham.tar.bz2 dan 20021010_spam.tar.bz2. Dataset ini mencakup koleksi 2551 email ham (non-spam) dan 501 email spam yang akan digunakan untuk pelatihan dan pengujian model. Setelah mengunduh dataset, langkah selanjutnya adalah mengekstrak file tersebut menggunakan perintah tar untuk mendapatkan email-email dalam format yang dapat diakses dan diolah lebih lanjut. Proses ini memastikan bahwa data tersedia dalam format yang siap untuk langkah pra-pemrosesan dan analisis selanjutnya.

	text	label
0	From spamassassin-devel-admin@lists.sourceforg...	0
1	From rssfeeds@jmason.org Tue Oct 8 10:56:09 ...	0
2	From rpm-list-admin@freshrpms.net Mon Sep 9 ...	0
3	From fork-admin@xent.com Tue Sep 24 10:49:30 ...	0
4	From fork-admin@xent.com Fri Sep 6 18:46:52 ...	0
	text	label
3047	From jzlin@bacalhau.com.br Mon Sep 16 00:11:0...	1
3048	From biz2biz2446@Flashmail.com Mon Oct 7 22:...	1
3049	From newpsychic414@mail.ru Wed Aug 28 11:17:1...	1
3050	From egabriel@ant.eupvg.upc.es Thu Aug 29 10:...	1
3051	From TrishaRPotter@yahoo.co.uk Tue Sep 24 17:...	1

Gambar 2. Sample Data Ham dan Spam

Contoh data yang terdiri dari email dan label yang terkait. Label 0 menunjukkan bahwa email tersebut adalah email non-spam (juga dikenal sebagai HAM), sedangkan label 1 menunjukkan bahwa email tersebut adalah spam

3.2. Pra-Pemrosesan

Pra-pemrosesan data adalah langkah-langkah awal untuk mempersiapkan data mentah sebelum digunakan dalam model pembelajaran mesin. Ini melibatkan tokenisasi teks, di mana teks dikonversi menjadi representasi numerik, dan padding sequences, di mana data numerik disesuaikan dengan panjang seragam. Tujuannya adalah untuk memastikan data siap digunakan dan konsisten dalam format yang dapat diproses oleh model.

3.3. Pemisahan Data

Pemisahan data adalah proses membagi dataset menjadi dua subset utama yaitu data pelatihan dan data pengujian. Hal ini dilakukan untuk melatih model mesin pada data yang tersedia dan menguji kinerjanya. Dalam konteks ini, data yang telah diolah *(data_padded*) dan labelnya dibagi menggunakan fungsi *train_test_split* dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Penggunaan parameter *stratify* memastikan bahwa distribusi label spam dan ham tetap seimbang di kedua set. Proses ini penting untuk memastikan konsistensi dan evaluasi yang akurat terhadap model yang dibangun.

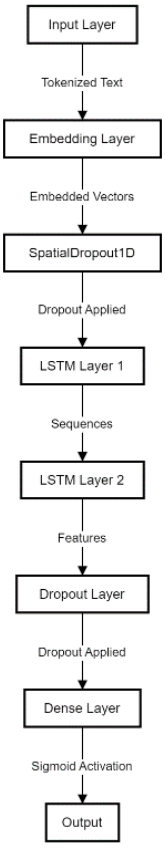
Tabel 1. Distribusi Data

Dataset	Jumlah Data	Kategori	Jumlah
Pelatihan	2441	Ham	2040
		Spam	401
Pengujian	611	Ham	511
		Spam	100

Setelah dilakukan pra-pemrosesan, menunjukkan distribusi label pada set pelatihan memiliki 2040 email ham dan 401 email spam (total 2441), sedangkan set pengujian memiliki 511 email ham dan 100 email spam (total 611).

3.4. Pembuatan dan Pelatihan Model

Pembuatan dan pelatihan model melibatkan pembuatan model *Sequential* menggunakan lapisan seperti *Embedding* untuk mengubah kata-kata menjadi *vektor numerik*, *LSTM*, *Dropout*, dan *Dense* untuk memproses dan mengklasifikasikan email sebagai spam atau tidak spam. Model ini dikompilasi dengan *optimizer* dan fungsi *loss* yang sesuai, kemudian dilatih menggunakan data pelatihan dengan mekanisme *EarlyStopping* untuk mencegah overfitting. Proses ini memastikan model dapat mengenali dan memprediksi spam email secara efektif.



Gambar 3. Model LSTM

Model LSTM yang digunakan terdiri dari beberapa layer: Input Layer menerima teks yang telah di-tokenisasi, diikuti oleh Embedding Layer yang mengubah token menjadi vektor dengan dimensi 64. SpatialDropout1D diterapkan untuk mengurangi overfitting, diikuti oleh dua layer LSTM (64 dan 32 unit) yang memproses sekuens input. Dropout Layer ditambahkan untuk regularisasi lebih lanjut, dan akhirnya Dense Layer dengan aktivasi sigmoid menghasilkan klasifikasi biner (spam atau bukan spam).

3.5. Evaluasi Model

Evaluasi model adalah proses mengukur performa model menggunakan data pengujian untuk menghitung nilai loss dan akurasi. Selain itu, model melakukan prediksi untuk menghasilkan metrik evaluasi seperti *precision, recall,* dan *f1-score* untuk setiap kelas. Proses ini penting untuk menilai seberapa baik model *LSTM* mendeteksi email spam dan memastikan bahwa model bekerja efektif dalam klasifikasi biner [7]. Berikut adalah persamaan model Metrik evaluasi.

$$Precision = \frac{TP}{(TP+FP)} \quad (1)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (2)$$

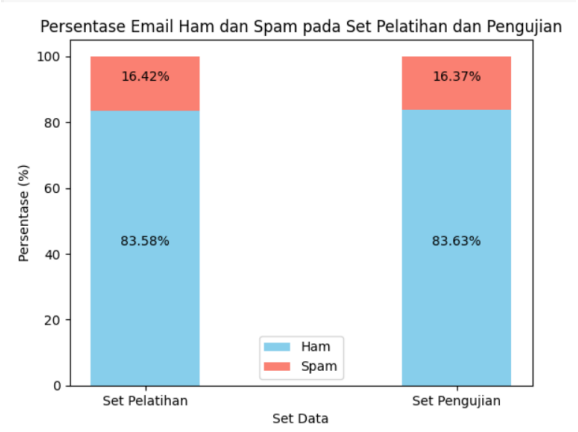
$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Keterangan :

- TP (True Positive): Jumlah email spam yang benar diidentifikasi sebagai spam
- TN (True Negative): Jumlah email bukan spam yang benar diidentifikasi sebagai bukan spam
- FP (False Positive): Jumlah email bukan spam yang salah diidentifikasi sebagai spam
- FN (False Negative): Jumlah email spam yang salah diidentifikasi sebagai bukan spam

4. HASIL DAN PEMBAHASAN



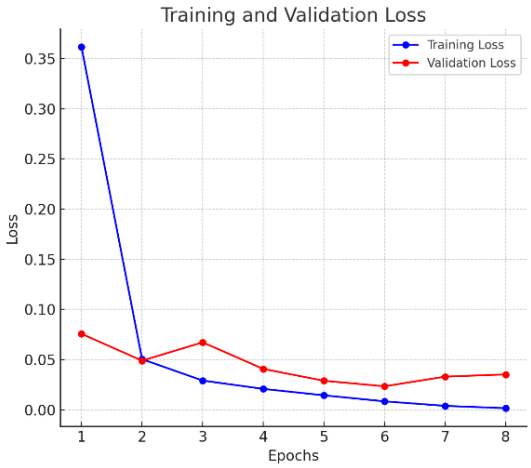
Gambar 4. Grafik presentase Data Set Pelatihan dan Pengujian

Proses *Pemisahan Data* menghasilkan pembagian data menjadi set pelatihan (2441 data) dan set pengujian (611 data) dengan rasio 80:20.

Di set pelatihan, sekitar 83.58% email adalah *Ham* dan 16.42% *Spam* , sedangkan di set pengujian, proporsinya adalah 83.63% *Ham* dan 16.37% *Spam*. Ini menunjukkan bahwa ketidakseimbangan proporsi antara email 'ham' dan 'spam' di kedua set data telah dijaga dengan baik. Stratifikasi dalam pembagian data memastikan proporsi label yang seimbang di kedua set, yang penting untuk evaluasi yang akurat terhadap model terhadap kedua kelas.

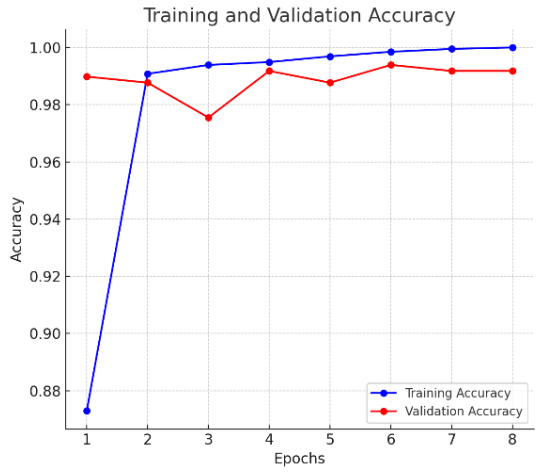
Tahap selanjutnya adalah *Pembuatan dan Pelatihan Model*. Setelah data dibagi menjadi set pelatihan dan set pengujian dengan rasio 80:20, model LSTM dibangun untuk mengolah teks email yang telah diproses sebelumnya menggunakan arsitektur *Sequential*. Data teks yang telah di-*tokenisasi* dan dipad menjadi panjang seragam dimasukkan ke dalam lapisan *Embedding* untuk mengubah kata-kata menjadi vektor numerik. Lapisan *SpatialDropout1D* digunakan untuk mengurangi overfitting dengan secara acak menghapus sebagian nilai unit selama pelatihan.

Selanjutnya, model memanfaatkan dua lapisan LSTM berurutan untuk menangkap pola temporal dalam teks email, membantu model memahami urutan kata-kata dalam konteks. Lapisan *Dropout* digunakan setelah LSTM untuk lebih mengurangi risiko overfitting selama pelatihan. Lapisan output *Dense* dengan aktivasi *sigmoid* menghasilkan keputusan biner untuk mengklasifikasikan email sebagai *spam* atau tidak *spam*. Selama pelatihan model, data pelatihan digunakan untuk mengoptimalkan bobot model, sementara data validasi digunakan untuk memantau performa model secara independen. *Early stopping* diterapkan untuk menghentikan pelatihan lebih awal jika tidak ada peningkatan pada loss validasi, mencegah overfitting dan mempertahankan parameter terbaik. Model dilatih selama 10 epoch dengan batch size 32, kemudian kinerja dievaluasi menggunakan data pengujian terpisah untuk mengukur akurasi prediksi model terhadap data. Berikut hasil menunjukkan performa model selama proses pelatihan.



Gambar 5. Grafik Training and Validation Loss

Model LSTM menunjukkan kinerja sangat baik selama delapan epoch pelatihan, yang ditandai dengan penurunan loss yang signifikan baik pada data pelatihan maupun data validasi. Grafik loss menunjukkan penurunan tajam pada data pelatihan dari 0.3615 menjadi 0.0016 dan penurunan loss validasi dari 0.0757 menjadi 0.0352 dengan sedikit fluktuasi. Penurunan ini mengindikasikan bahwa kemampuan model meminimalkan kesalahan dan generalisasi yang baik pada data yang tidak terlihat, tanpa overfitting signifikan. Hasil ini menunjukkan bahwa model LSTM yang dilatih mampu meminimalkan kesalahan secara efektif sekaligus mempertahankan kemampuan generalisasi yang kuat.



Gambar 6. Grafik Training and Validation Accuracy

Grafik akurasi menunjukkan peningkatan konsisten pada akurasi pelatihan dari 87.30% menjadi 100%. Akurasi validasi tetap tinggi dan stabil, mulai dari 98.98% dan berakhir di 99.18%, dengan sedikit fluktuasi. Stabilitas akurasi validasi ini menunjukkan bahwa model tidak hanya menghafal data pelatihan, tetapi juga mampu melakukan generalisasi dengan baik, mengindikasikan performa yang andal dan konsisten dalam mendeteksi spam email, tanpa risiko overfitting.

Evaluasi Model dilakukan pada data pengujian untuk mengukur kinerja model yang telah dilatih. Dengan evaluasi model, dapat diketahui bagaimana model secara komprehensif mendeteksi spam email, memastikan bahwa model tidak hanya akurat tetapi juga efektif dalam menangani ketidakseimbangan data antara *spam* dan *not spam*.

```
20/20 [==============================] - 2s 88ms/step - loss: 0.0332 - accuracy: 0.9935
Accuracy: 99.35%
20/20 [==============================] - 3s 82ms/step
              precision    recall  f1-score   support

    Not Spam       1.00      0.99      1.00       511
        Spam       0.97      0.99      0.98       100

    accuracy                           0.99       611
   macro avg       0.98      0.99      0.99       611
weighted avg       0.99      0.99      0.99       611
```

Gambar 6. Hasil Evaluasi Model

Berdasarkan hasil evaluasi model LSTM yang digunakan, performa deteksi spam email menunjukkan hasil yang sangat baik. Pada tahap pelatihan, model mencapai tingkat akurasi 99.35%. Tingginya tingkat akurasi ini mencerminkan kemampuan model LSTM dalam memahami pola dan karakteristik data yang digunakan dalam pelatihan, sehingga mampu memberikan prediksi yang sangat akurat.

Selanjutnya, evaluasi performa model melalui metrik precision, recall, dan f1-score menunjukkan hasil yang juga sangat baik, khususnya dalam membedakan antara email *Spam* dan *Not Spam*. Precision, recall, dan f1-score untuk *Not Spam* masing-masing adalah 1.00, 0.99, dan 1.00, menunjukkan identifikasi yang hampir sempurna. Untuk *Spam*, precision, recall, dan f1-score masing-masing adalah 0.97, 0.99, dan 0.98, menunjukkan efektivitas tinggi dalam mendeteksi spam dengan tingkat kesalahan rendah. Nilai *macro avg* dan *weighted avg* yang hampir mendekati 1.00 pada semua metrik menunjukkan keseimbangan performa yang baik pada kedua kelas.

```
Confusion Matrix:
[[508   3]
 [  1  99]]
```

Gambar 7. Confusion Matrix

Confusion matrix mendukung hasil ini dengan rincian sebagai berikut: dari total 611 email, 508 email yang benar-benar Not Spam diklasifikasikan dengan benar, sementara 3 email salah teridentifikasi sebagai Spam. Sebaliknya, dari 100 email Spam, 99 diklasifikasikan dengan benar sebagai Spam, sementara hanya 1 email Not Spam yang salah teridentifikasi sebagai Spam. Ini menunjukkan bahwa model memiliki sedikit kesalahan dalam prediksi, dengan kesalahan minor dalam mengidentifikasi spam dan not spam. Hasil ini menunjukkan bahwa model LSTM memiliki kinerja luar biasa dalam deteksi spam email, dengan generalisasi yang kuat dan akurasi sangat tinggi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi model, performa deteksi spam email menunjukkan hasil dengan akurasi 99.35%. Precision, recall, dan f1-score untuk Not Spam masing-masing adalah 1.00, 0.99, dan 1.00, menunjukkan identifikasi yang hampir sempurna. Untuk Spam, precision, recall, dan f1-score masing-masing adalah 0.97, 0.99, dan 0.98, menunjukkan efektivitas tinggi dalam mendeteksi spam dengan tingkat kesalahan rendah. Nilai macro avg dan weighted avg yang hampir mendekati 1.00 pada semua metrik menunjukkan keseimbangan performa yang baik pada kedua kelas. Hasil ini menunjukkan bahwa model LSTM memiliki kinerja luar biasa dalam deteksi spam email, dengan generalisasi yang kuat dan akurasi sangat tinggi. Namun, penting untuk dicatat bahwa penelitian ini memiliki keterbatasan dalam hal ukuran dataset dan lingkup variasi email spam yang dapat diakomodasi. Untuk meningkatkan keandalan dan generalisasi model, penelitian selanjutnya disarankan untuk memperluas dataset dengan lebih banyak variasi email spam serta menguji model dengan lebih banyak kasus yang mungkin.

Rekomendasi ini bertujuan untuk mengatasi tantangan yang terus berkembang dalam deteksi spam email, mengingat pengirim spam terus mengembangkan strategi baru untuk menghindari deteksi.

DAFTAR PUSTAKA

[1] R. F. Busyra and A. S. Girsang, "Applying Long Short-Term Memory Algorithm for Spam Detection on Ministry Websites," Journal of System and Management Sciences, vol. 14, no. 2, pp. 1–20, 2024, doi: 10.33168/JSMS.2024.0201.

[2] A. Chiche and B. Yitagesu, "Part of speech tagging: a systematic review of deep learning and machine learning approaches," J. Big Data, vol. 9, no. 1, 2022. doi: 10.1186/s40537-022-00561-y.

[3] Y. Guo, Z. Mustafaoglu, and D. Koundal, "Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms," J. Comput. Cogn. Eng., vol. 2, no. 1, pp. 5–9, 2023. doi: 10.47852/bonviewJCCE2202192.

[4] M. B. Hartono and A. K. Darmawan, "Komparasi Deep Learning Dan Traditional Machine Learning Untuk Email Spam Filtering," Jurnal Minfo Polgan, vol. 12, no. 1, pp. 636–643, 2023.

[5] K. M. S. Hidayatullah and T. Sutabri, "Pengembangan Sistem Pengklasifikasi e-mail Berbasis Kecerdasan Buatan untuk Deteksi Spam dan Phishing," IJM: Indonesian J. Multidiscip., vol. 2, no. 2, 2024. [Online]. Available: https://journal.csspublishing.com/index.php/ijm/article/view/689.

[6] H. Iswanto, E. Seniwati, Y. Astuti, and D. Maulina, "Comparison of Algorithms on Machine Learning For Spam Email Classification," IJISTECH (Int. J. Inf. Syst. Technol.), vol. 5, no. 4, pp. 446–455, 2021. doi: 10.30645/ijistech.v5i4.164.

[7] E. John-Africa and V. T. Emmah, "Performance Evaluation of LSTM and RNN Models in the Detection of Email Spam Messages," Eur. J. Inf. Technol. Comput. Sci., vol. 2, no. 6, pp. 24–30, 2022. doi: 10.24018/compute.2022.2.6.80.

[8] A. Rolangon, A. Weku, and G. A. Sandag, "Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19 The Comparison of LSTM Algorithms for Twitter User Sentiment Analysis on Hospital Services During the Covid-19 Pandemic," Jurnal Teknologi dan Komunikasi, vol. 13, no. 1, pp. 31–40, 2023, doi: https://doi.org/10.36342/teika.v13i01.3063.

[9] A. A. Kurniawan and M. Mustikasari, "Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia," Jurnal Informatika Universitas Pamulang, vol. 5, no. 4, pp. 544–552, 2020. doi: 10.32493/informatika.v5i4.6760.

[10] M. Rustam, A. Brotokuncoro, and R. Roestam, "Deteksi Email Spam dengan Continuous Bag-of-Words dan Random Forest," Scientica: Jurnal Ilmiah Sains Dan Teknologi, vol. 2, no. 4, pp. 339–347, 2024. doi: 10.572349/scientica.v2i4.1260.

[11] T. B. Sianturi, I. Cholissodin, and N. Yudistira, "Penerapan Algoritma Long Short-Term Memory (LSTM) berbasis Multi Fungsi Aktivasi Terbobot dalam Prediksi Harga Ethereum," Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, vol. 7, no. 3, pp. 1101–1107, 2023, [Online]. Available: http://j-ptiik.ub.ac.id.

[12] A. N. Muslikah, "SMS Spam Detection Menggunakan Metode Long Short Term Memory," Skripsi, Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia, 2021.

[13] A. P. Sari, H. Suzuki, T. Kitajima, T. Yasuno, D. A. Prasetya, and R. Arifuddin, "Short-Term Wind Speed and Direction Forecasting by 3DCNN and Deep Convolutional LSTM," IEEJ Trans. Electr. Electron. Eng., vol. 17, no. 11, pp. 1620–1628, 2022. doi: 10.1002/tee.23669.

[14] Z. Bin Siddique, M. A. Khan, I. U. Din, A. Almogren, I. Mohiuddin, and S. Nazir, "Machine Learning-Based Detection of Spam Emails," Sci. Program., 2021. doi: 10.1155/2021/6508784.

[15] M. A. Sofyan, N. Rahaningsih, and R. D. Dana, "DETEKSI SMS SPAM BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," 2024. doi: https://doi.org/10.36040/jati.v8i3.9532.

[16] H. Mukhtar, J. Al Amien, and M. A. Rucyat, "Filtering Spam Email menggunakan Algoritma Naïve Bayes", CoSciTech, vol. 3, no. 1, pp. 9-19, May 2022.