

# Klasifikasi Spam

## PADA EMAIL MAHASISWA ITERA

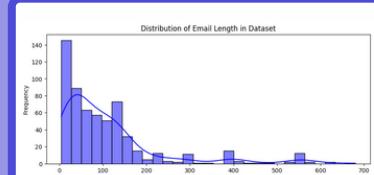
MENGGUNAKAN METODE HYBRID LONG  
SHORT-TERM MEMORY (LSTM)

### LATAR BELAKANG

- 1 Email menjadi media komunikasi utama dalam lingkungan akademik di ITERA
- 2 Kehadiran spam menimbulkan risiko terhadap keamanan digital pengguna
- 3 Model machine learning memiliki keterbatasan pada data skala besar
- 4 Model deep learning mampu mempertahankan informasi penting pada data teks panjang

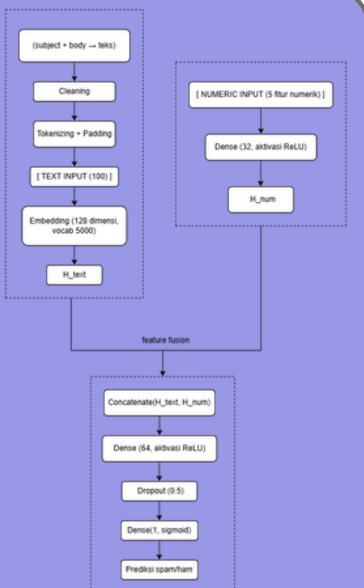


### ANALISIS DESKRIPTIF



Distribusi panjang email sebagian besar email memiliki jumlah kata yang relatif pendek, terutama pada rentang di bawah 100 kata, dengan frekuensi yang menurun seiring bertambahnya panjang email

### ARSITEKTUR

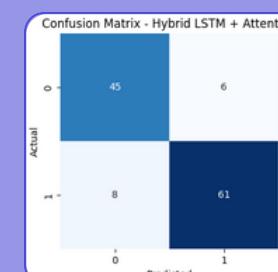
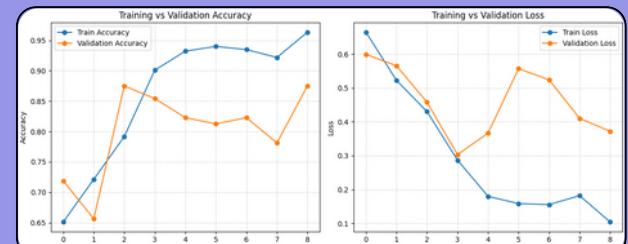


### KONVERGENSI MODEL

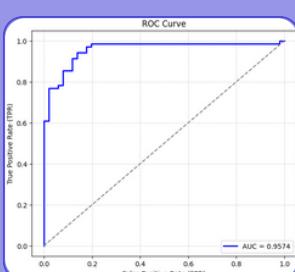
Epoch	Accuracy	Loss	Val Accuracy	Val Loss
1	0.624	0.680	0.719	0.599
2	0.713	0.516	0.656	0.563
3	0.771	0.151	0.875	0.138
4	0.871	0.332	0.854	0.303
5	0.922	0.212	0.823	0.367
6	0.926	0.178	0.813	0.337
7	0.941	0.166	0.823	0.324
8	0.925	0.197	0.781	0.410
9	0.941	0.141	0.875	0.372

Akurasi pelatihan meningkat secara konsisten dan mencapai nilai sekitar 0.95 sebelum epoch ke-9, sedangkan, train loss mengalami penurunan signifikan hingga mendekati 0.1. Meskipun validation loss menunjukkan fluktuasi, namun validation accuracy mengikuti pola peningkatan pada fase awal pelatihan.

### HASIL



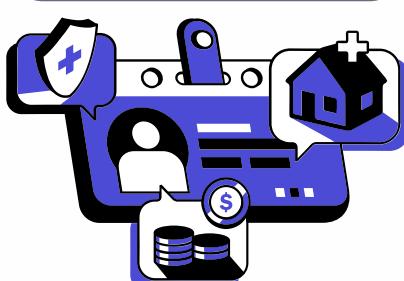
Berdasarkan Confusion Matrix, diperoleh nilai TP sebesar 61 dan TN sebesar 45, yang menunjukkan tingkat ketepatan prediksi yang tinggi pada kedua kelas.



Hasil evaluasi menggunakan kurva ROC menunjukkan kemampuan diskriminatif model yang sangat baik, dengan nilai AUC sebesar 0.9574. Nilai AUC yang mendekati 1 artinya model Hybrid Attention LSTM memiliki performa klasifikasi yang sangat baik dalam membedakan email spam dan ham.

### KELOMPOK 1

- NATASYA EGA LINA MARBUNI
- EKSANTY F SUGMA ISLAMIATY
- MUHAMMAD REGI ABDI PUTRA AMANTA



- Korelasi kuat ( $r = 0.93$ ) antara body\_len dan num\_special\_chars, menunjukkan email yang lebih panjang cenderung mengandung lebih banyak karakter khusus.
- Korelasi sangat kuat ( $r \approx 0.95$ ) antara num\_urls dan num\_exclam, mencerminkan ciri khas email spam yang banyak tautan dan penekanan berlebihan.
- Pola ini konsisten dengan karakteristik umum email spam: konten panjang, agresif, dan persuasif.
- has\_attachment memiliki korelasi sangat lemah dengan fitur lain, sehingga bersifat independen dan berpotensi menambah informasi unik dalam model.

### KESIMPULAN

- Model Hybrid Attention LSTM menunjukkan kinerja sangat baik dalam mengklasifikasikan email spam dan ham pada lingkungan mahasiswa ITERA.
- Kombinasi LSTM untuk pemrosesan teks, attention mechanism, dan fitur numerik email meningkatkan kemampuan model dalam menangkap pola penting pada email.
- Nilai akurasi dan AUC yang tinggi mengindikasikan kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat.
- Visualisasi attention weight membuktikan model mampu menyoroti kata kunci relevan dalam proses klasifikasi.
- Secara keseluruhan, pendekatan ini efektif untuk penyaringan email dan berkontribusi pada peningkatan keamanan komunikasi digital di lingkungan akademik ITERA.