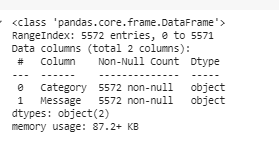
Laporan Tugas Kelompok III

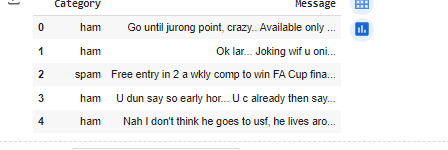
Disusun oleh: Kelompok 3A

1. Deskripsi Dataset

* Dataset yang digunakan adalah dataset yang mengklasifikasikan suatu pesan yang termasuk dalam jenis spam atau ham (pesan sah). Pada dataset ini memiliki 2 kolom yaitu kategori dan message. Data ini memiliki total 5572 data. Dataset ini didapatkan dari open source terbuka yaitu situs Kaggle.



Gambar 1. Info elemen dataset



Gambar 2. Dataset asli

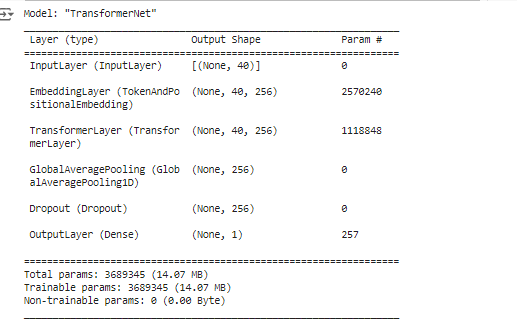
* Sumber: <https://www.kaggle.com/datasets/team-ai/spam-text-message-classification>
* Google collab : <https://colab.research.google.com/drive/1RK2mA6XAzF4lts5H9k-pI9j4hpCtK-nI?usp=sharing#scrollTo=hk4-uJmfr4De>

1. Deskripsi model pemrosesan text

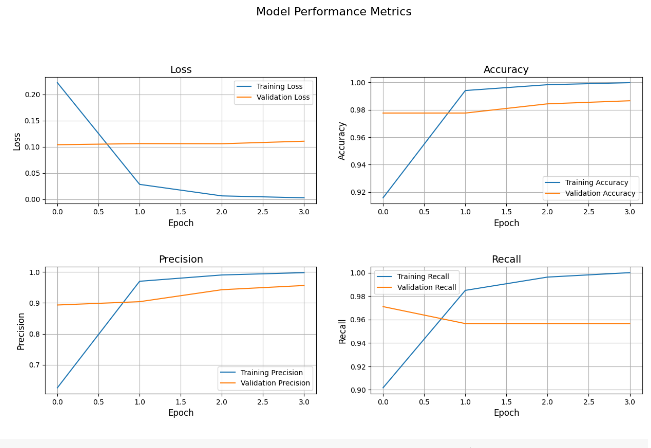
* Model AI yang dibangun adalah model AI yang dapat mengklasifikasikan suatu pesan untuk termasuk sebagai pesan spam atau pesan HAM menggunakan TensorFlow Text (Transformer).
* Pre processing data yang dilakukan pada dataset ini yaitu menghilangkan data duplikat, melakukan encoding untuk konversi label teks menjadi nilai numerik.

1. Evaluasi performa model

* Model pelatihan menggunakan model transformer.



Gambar 3. Ringkasan model arsitektur

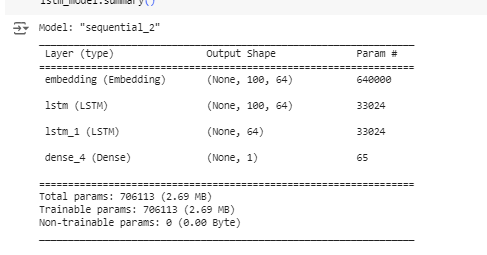


Gambar 4. Grafik performa model

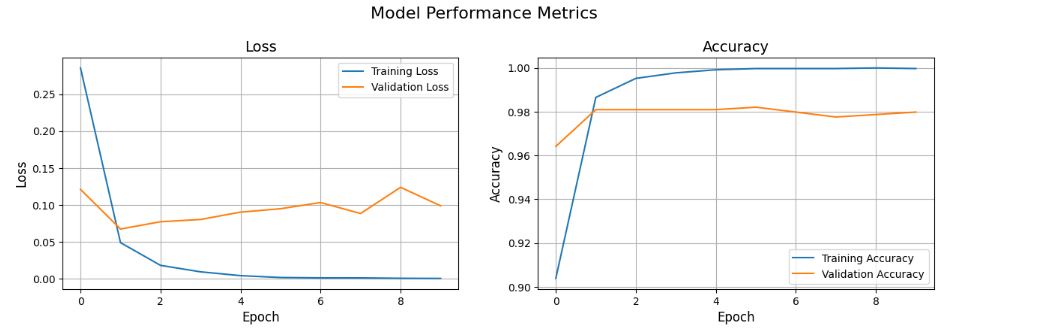
* Berdasarkan Gambar 4, model yang dibangun memiliki akurasi sekitar 98 persen. Dengan akurasi ini, model dapat mengklasifikasikan jenis pesan sesuai dengan label kategori.

1. Optimasi model dan eksperimen model

* Melakukan pelatihan model pada dataset yang sama menggunakan alogaritma LSTM.



Gambar 5. Ringkasan Model Pelatihan LSTM



Gambar 6. Grafik Performa Model

* Berdasarkan Gambar 6, akurasi untuk pelatihan dan validasi tidak memiliki perbedaan yang terlalu jauh, tetapi nilai Loss (rugi- rugi informasi) menggunakan model ini memiliki nilai yang lebih besar dibandingkan menggunakan model transformer.