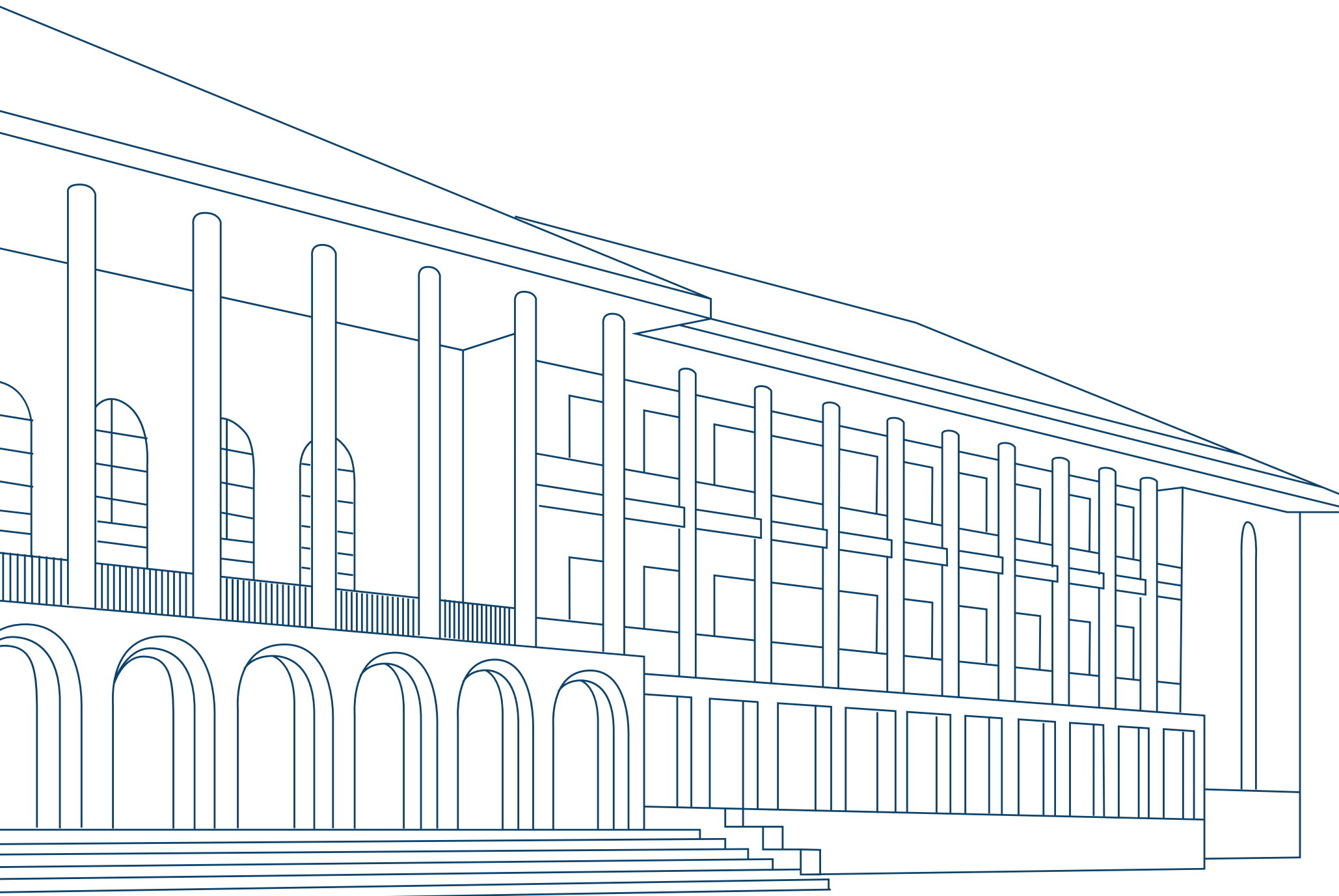




UNIVERSITAS
GADJAH MADA



LOCALLY ROOTED,
GLOBALLY RESPECTED

Comparing MLP and BiLSTM Architectures with IndoBERT Embeddings for Indonesian Text Classification

Atika Riska Rahmanda 24/552170/PPA/07001

Nilam Mufidah 24/551986/PPA/06994

Putri Santika Mayangsari 24/553020/PPA/07020

Soufi Ramadhani Inulqulub 24/552469/PPA/07010

ugm.ac.id



UNIVERSITAS
GADJAH MADA



LOCALLY ROOTED,
GLOBALLY RESPECTED

Tugas dari setiap anggota :

Atika Riska Rahmanda 24/552170/PPA/07001

(Code BiLstm + Indobert & PPT)

Nilam Mufidah 24/551986/PPA/06994

(Code MLP + Indobert twit & PPT)

Putri Santika Mayangsari 24/553020/PPA/07020

(Code BiLstm + Indobert & PPT)

Soufi Ramadhani Inulqulub 24/552469/PPA/07010

(Code MLP + Indobert twit& PPT)

ugm.ac.id

Introduction



ANALISIS SENTIMEN PENGGUNA TIKTOK TENTANG PROGRES PEMBANGUNAN IKN DENGAN METODE RANDOM FOREST

Siti Rihastuti¹, Afnan Rosyidi²
^{1,2}STMIK Amikom Surakarta, Sukoharjo,
¹siti@dosen.amikomsolo.ac.id
²afnan@dosen.amikomsolo.ac.id

Abstrak

Klasifikasi sentimen merupakan teknik analisis teks yang digunakan untuk mengidentifikasi dan mengkategorikan opini pengguna terhadap suatu aplikasi atau layanan. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat tentang progres pembangunan IKN (Ibukota Nusantara) dengan algoritma Random Forest berdasarkan komentar pengguna platform TikTok. Dataset diambil dari kaggle sebanyak 1472 komentar berbahasa Indonesia. Dataset yang digunakan terdiri dari komentar pengguna yang dikategorikan ke dalam sentimen positif dan negatif. Evaluasi dilakukan berdasarkan metrik akurasi, presisi, recall, dan F1-score untuk mengetahui hasil dari klasifikasi sentimen pengguna. Pengujian metode Random Forest di Google Colab menunjukkan nilai akurasi sebesar 77%, precision 78%, recall 77% dan F1-score 77%. Dari hasil pengujian tersebut, metode Random Forest dinilai cukup baik dalam mengklasifikasikan sentimen pengguna Tiktok dalam menanggapi progres pemindahan IKN.

Kata Kunci: Klasifikasi Sentimen, Tiktok, Random Forest, Komentar

Abstract

Sentiment classification is a text analysis technique used to identify and categorize user opinions about an application or service. This study aims to classify public sentiment about the progress of the development of the IKN (Indonesian Capital) with the Random Forest algorithm based on comments from users of the Tiktok platform. The dataset was taken from Kaggle with 1472 comments in Indonesian. The dataset used consists of user comments categorized into positive and negative sentiments. The evaluation was carried out based on the accuracy, precision, recall, and F1-score metrics to determine the results of the user sentiment classification. Testing the Random Forest method on Google Colab showed an accuracy value of 77%, precision 78%, recall 77% and F1-score 77%. From these values, the Random Forest method is considered quite good in classifying Tiktok user sentiment in responding to the progress of the IKN relocation.

Keywords: Sentiment Classification, Tiktok, Random Forest, Comments.

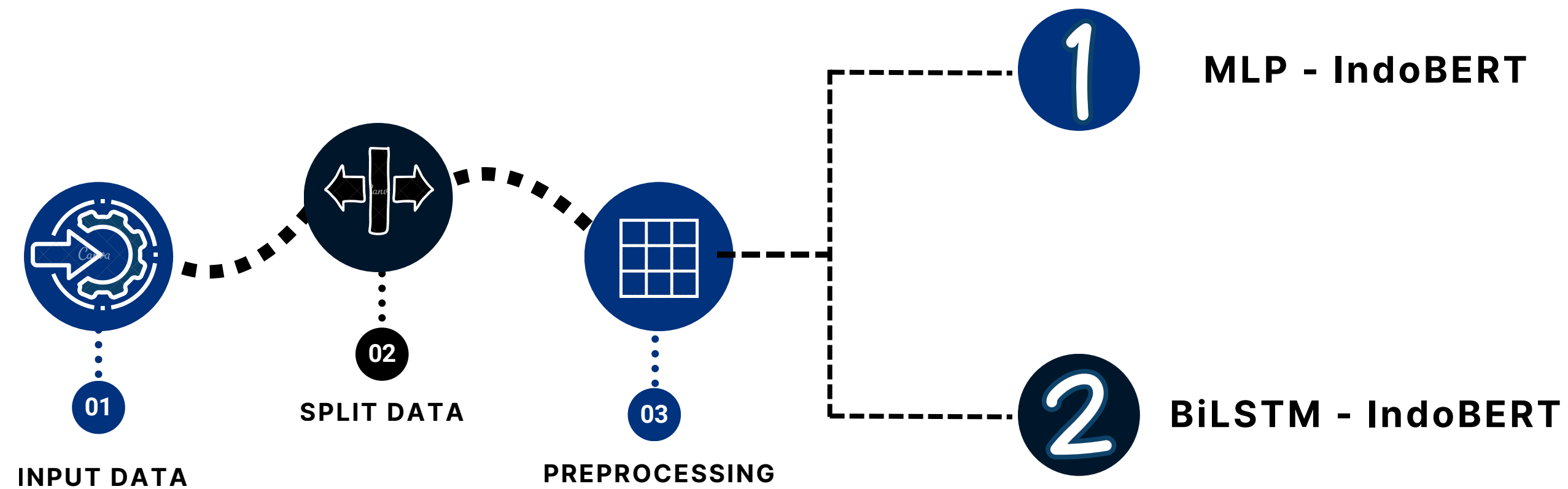
- Komentar-komentar diambil dari platform TikTok yang berisi opini dari publik terhadap perkembangan pembangunan IKN.
- Data sudah diberi label sentimen: positif dan negatif.
- Data diperoleh dari kaggle dengan jumlah 1.472
- negatif 780 positif 692

Github



[Click to link github](#)

Metode





STEAMING

Proses stemming menggunakan
package Sastrawi

STOPWORD

Menghapus stopword menggunakan
package Sastrawi

Jumlah stopword yang digunakan: 126
kata

1 MLP – IndoBERT

MLP Classifier

Embedding_dim = 768

Hidden_dim = 64

Output_dim = 2

Dropout = 0.2

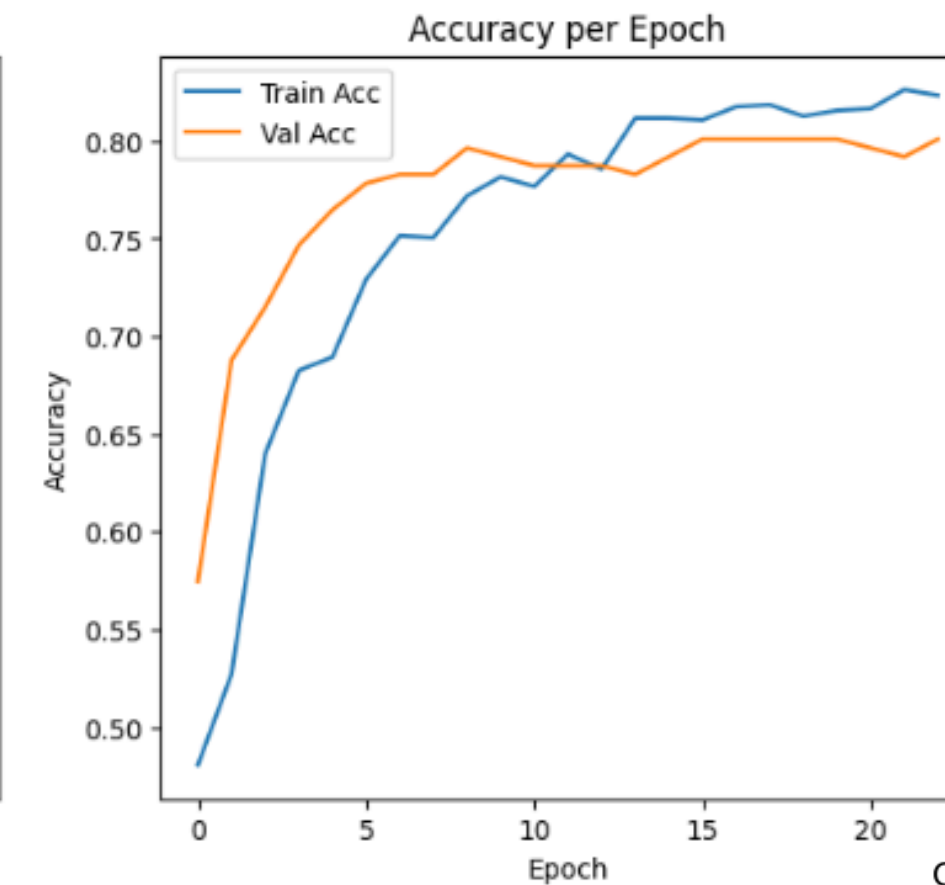
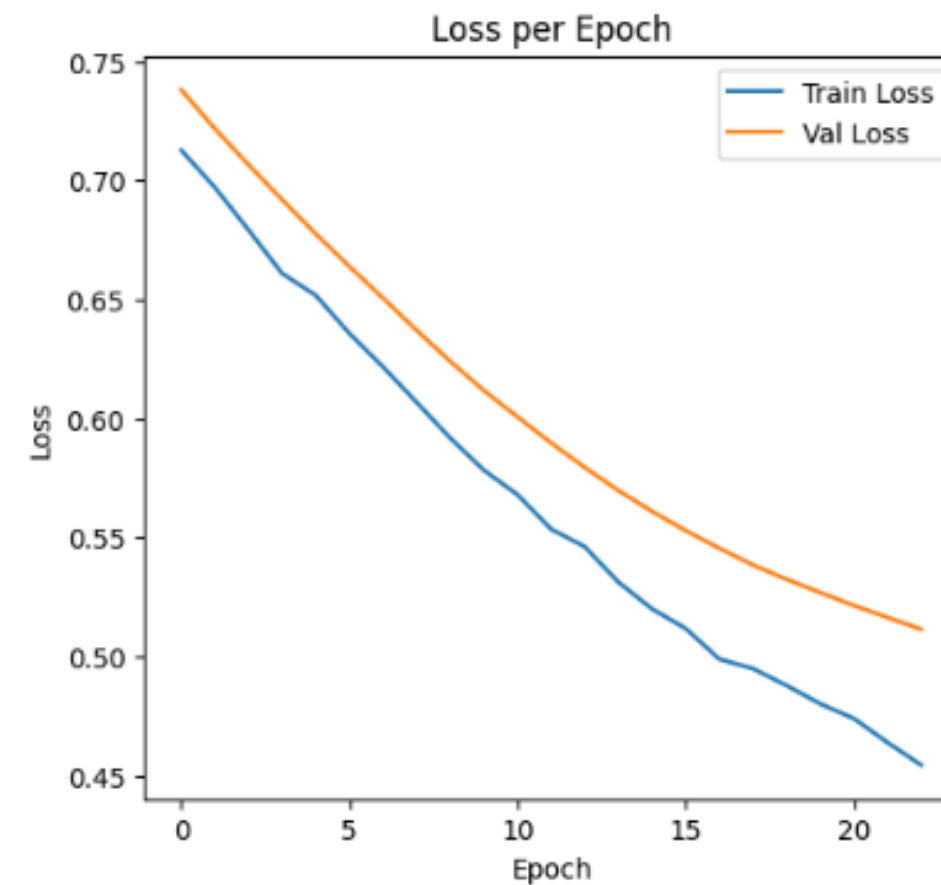
Optimizer = Adam

Loss Function = Cross Entropy

Batch size = 16

Epoch = 100

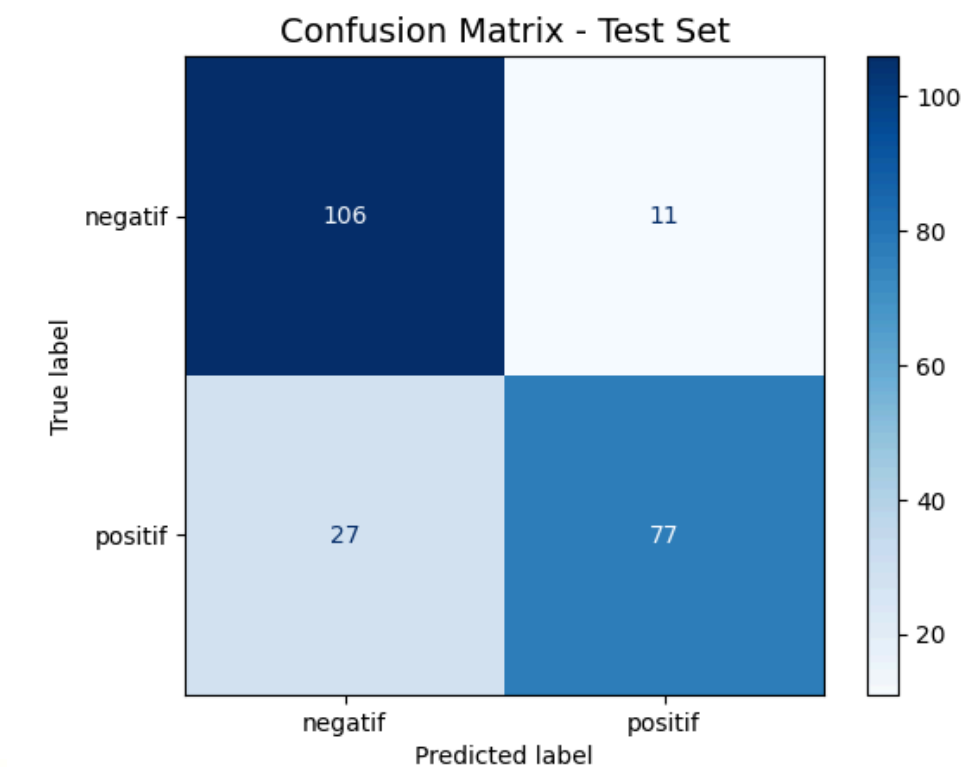
Early Stopping



Final Test Accuracy: 0.8281

Classification Report:

	precision	recall	f1-score	support
0	0.7970	0.9060	0.8480	117
1	0.8750	0.7404	0.8021	104
accuracy			0.8281	221
macro avg	0.8360	0.8232	0.8250	221
weighted avg	0.8337	0.8281	0.8264	221



IndoBERT Embeddings

Pre-Trained Model =

"indolem/indobertweet-base-uncased"

2 BiLSTM – IndoBERT

LSTM Classifier

Embedding_dim = 768

Hidden_dim = 128

Output_dim = 2

Dropout = 0.5

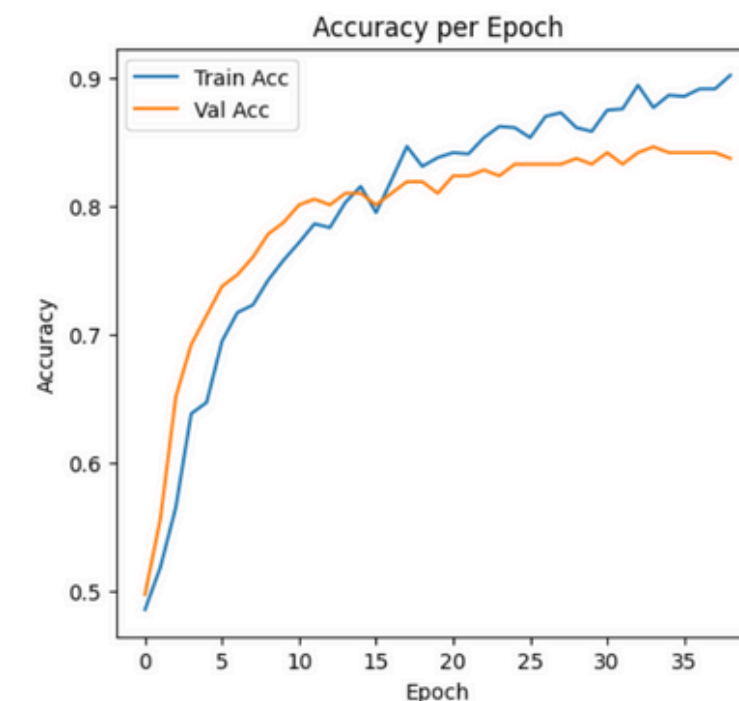
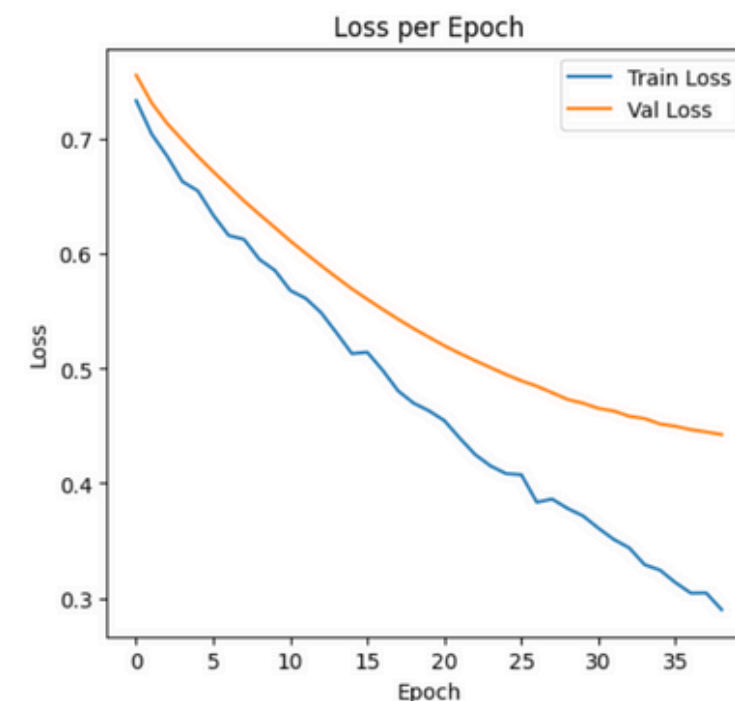
Optimizer = adam

Loss Function = cross entropy

batch size = 16

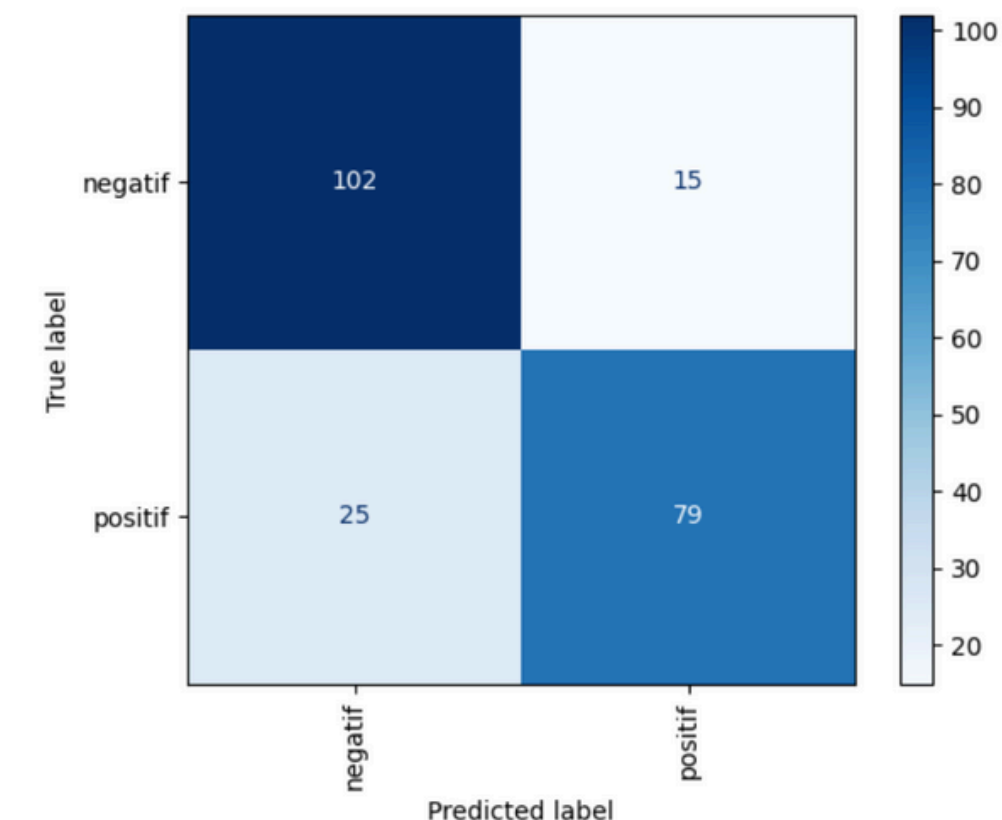
Epoch 100

Early Stopping



Classification Report:

	precision	recall	f1-score	support
negatif	0.8031	0.8718	0.8361	117
positif	0.8404	0.7596	0.7980	104
accuracy			0.8190	221
macro avg	0.8218	0.8157	0.8170	221
weighted avg	0.8207	0.8190	0.8181	221



IndoBERT Embeddings

Pre-Trained Model =

"indolem/indobertweet-base-uncased"

MLP - INDOBERT

- Model MLP + IndoBERTweet mencapai Akurasi 82.8%
- Precision kelas negatif (79.7%) lebih rendah dibanding positif (87.5%), tetapi recall kelas negatif (90.6%) lebih tinggi daripada kelas positif.
- Model cukup seimbang dalam memprediksi dua kelas, namun masih ada kelemahan pada kelas positif (recall = 74.04%).
- MLP dengan IndoBERTweet memberikan performa cukup baik namun masih moderat, karena model sederhana ini belum mampu menangkap konteks sekuensial teks dengan optimal.

BiLSTM - INDOBERT

- Model BiLSTM + IndoBERTweet mencapai akurasi akhir 81.90%
- Precision kelas positif 84.04% sangat tinggi tetapi recall rendah 75.96%
- Kelas negatif lebih stabil (precision 80.31%, recall 87.18%) model lebih “berhati hati” pada komentar negatif dibanding positif
- LSTM dengan IndoBERT embeddings mampu memberikan performa cukup baik, tetapi masih ada ketidakseimbangan performa antar kelas dikarenakan arsitektur LSTM lebih fokus pada urutan teks

KESIMPULAN

- Model MLP + IndoBERT adalah yang terbaik di antara kedua arsitektur, dengan akurasi 82.8%.
- Kinerja sangat baik pada kelas negatif, tetapi masih kurang optimal pada kelas positif (recall rendah), kemungkinan akibat distribusi data yang tidak seimbang.