

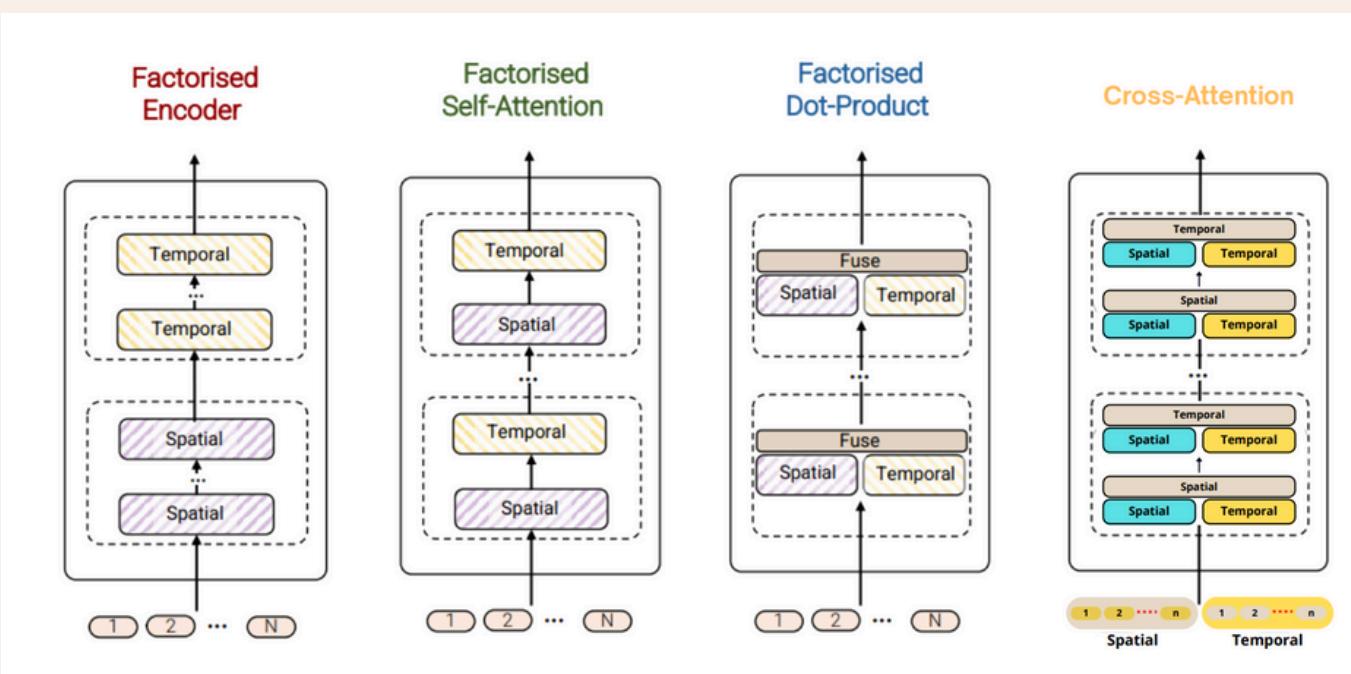
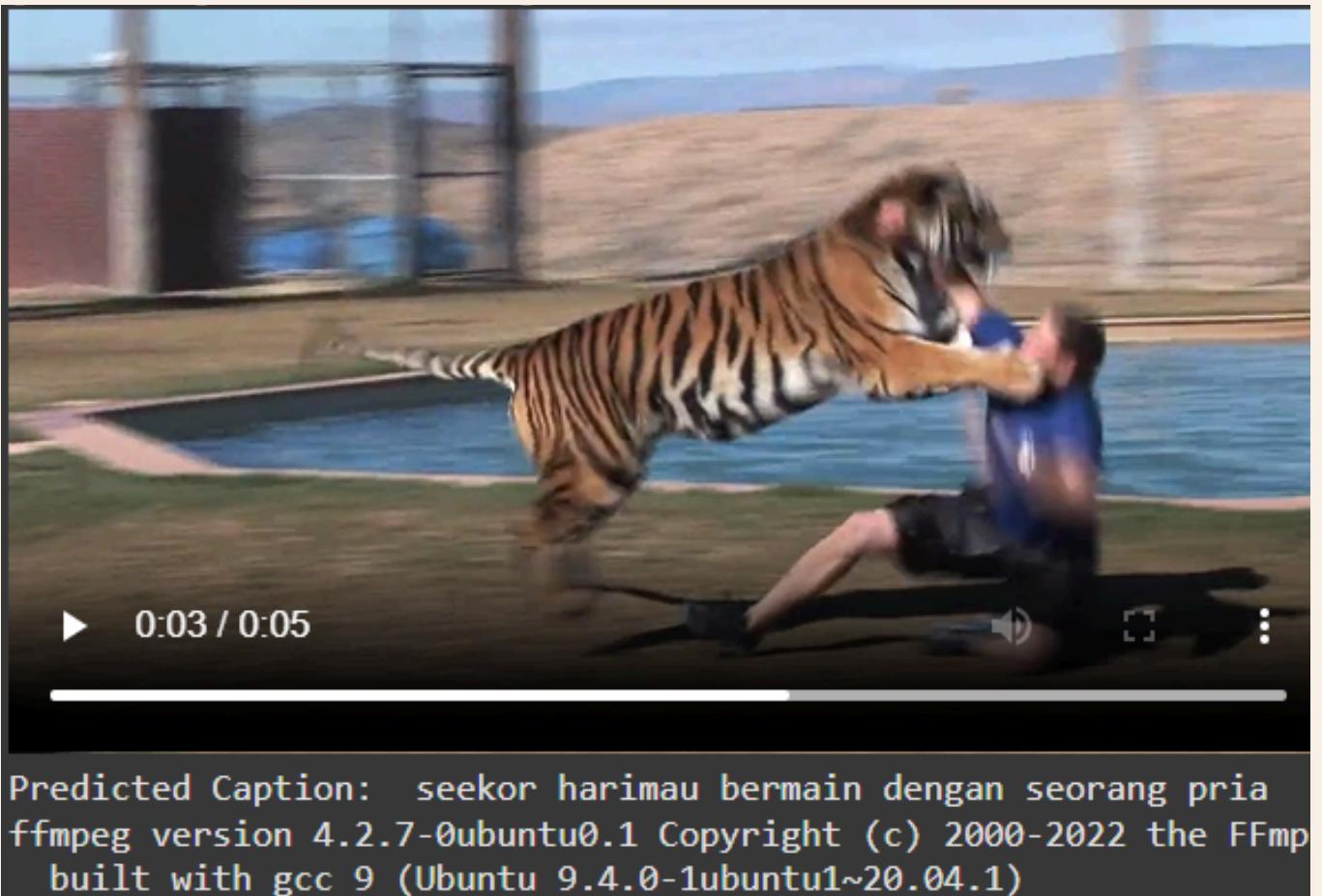
Portfolio

Data Scientist & AI Research/Engineer

→ **Miftahul Huda**

I have a background in computer science with strong algorithmic skills and intellectual adaptability. I'm currently interested in independent, in-depth research in artificial intelligence, data science, and physics, both theoretically and applied.



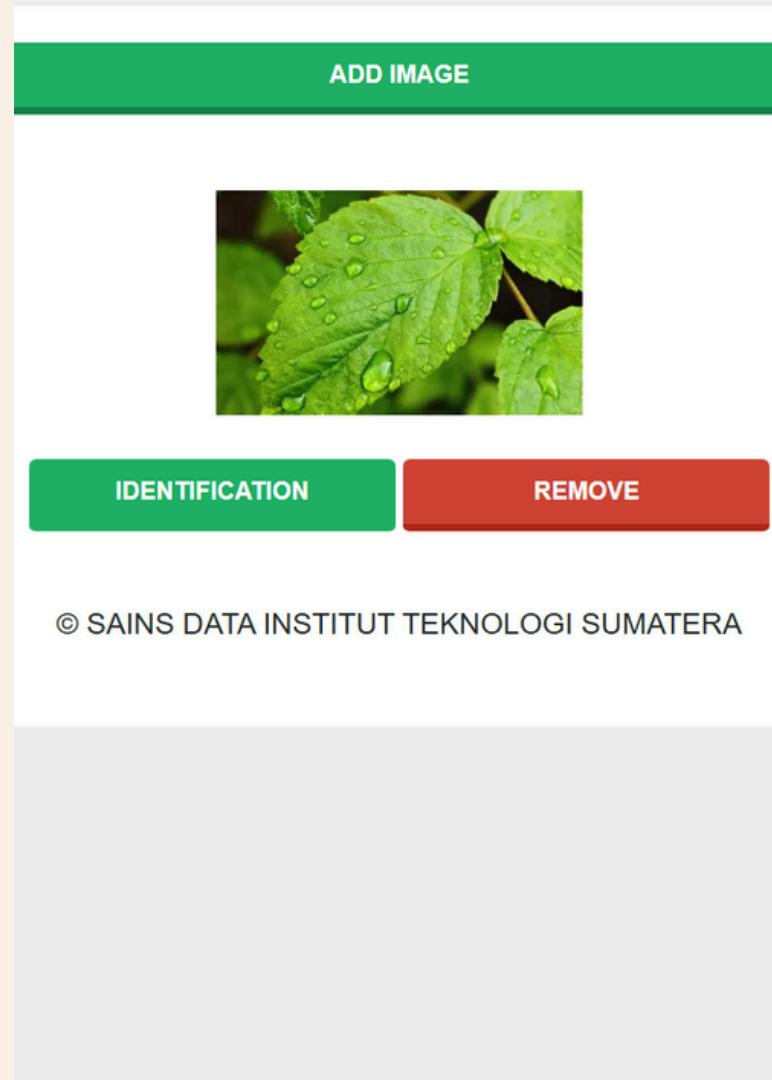


Proyek - AI Research

Cross-Attention Video Vision Transformer for Video Captioning

My research (thesis Bachelor's degree) creates a new ViViT-based video feature extraction architecture that has 3.5x lower GFLOPs (ablation study result) by reducing the quadratic complexity of self-attention compared to Factorized Dot-Product Attention (the lightest ViViT model) with comparable performance with all ViViT models based on CIDEr scores in video captioning tasks.

source code: [github](#)
paper (method): [repo](#)

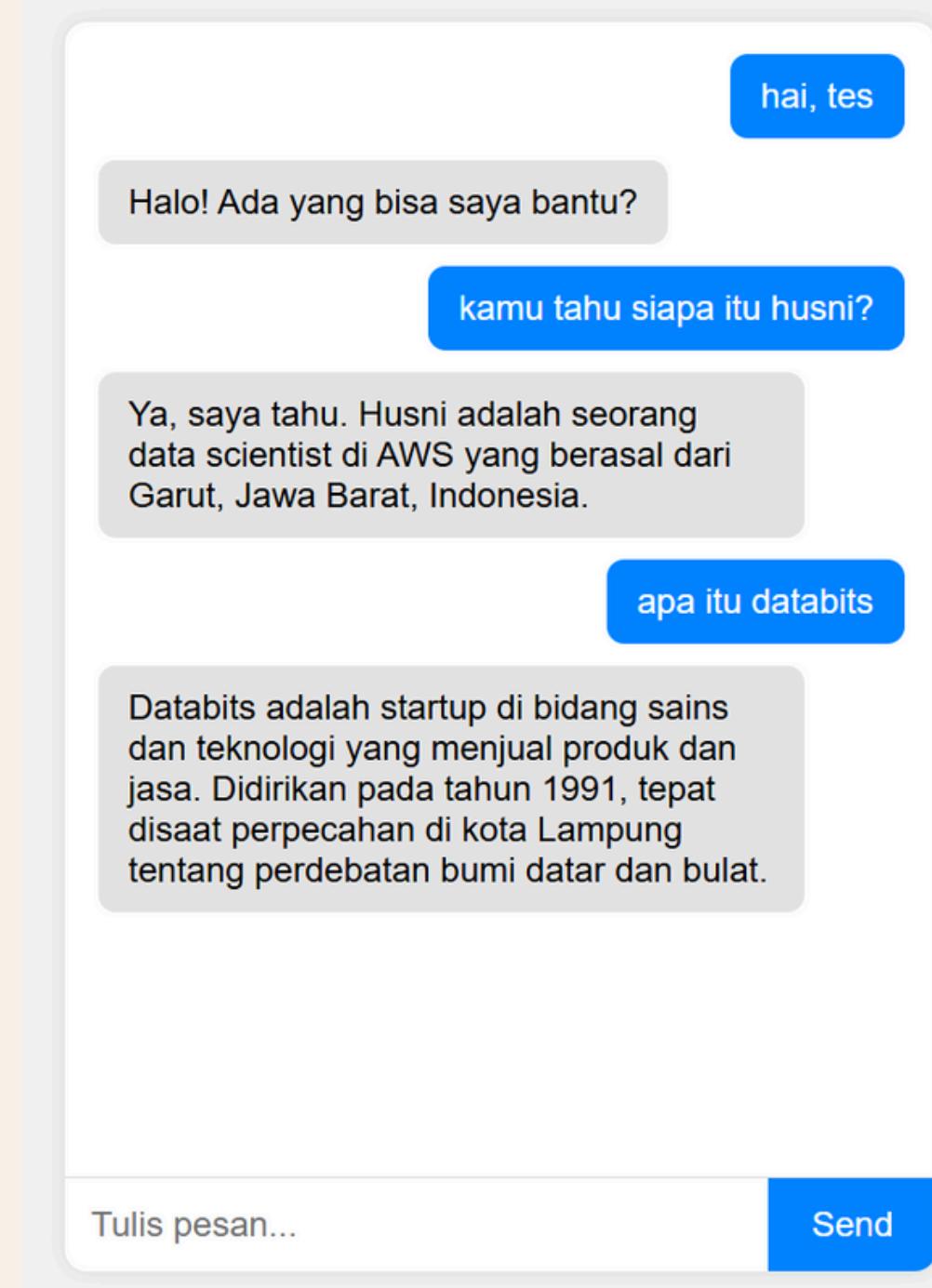


Proyek - AI Engineer/Developer

Android and Web Application for Mangrove Detection

Creation of a fine tuning Machine Learning model of the VGG16 architecture and an android/web application for identification of mangrove species in the Cuku Nyinyi ecotourism area, Lampung.

tools: tensorflow, android studio (kotlin), flask
source code android: [github](#)
source code web: [github](#)



Proyek - AI Engineer

ChatBot RAG

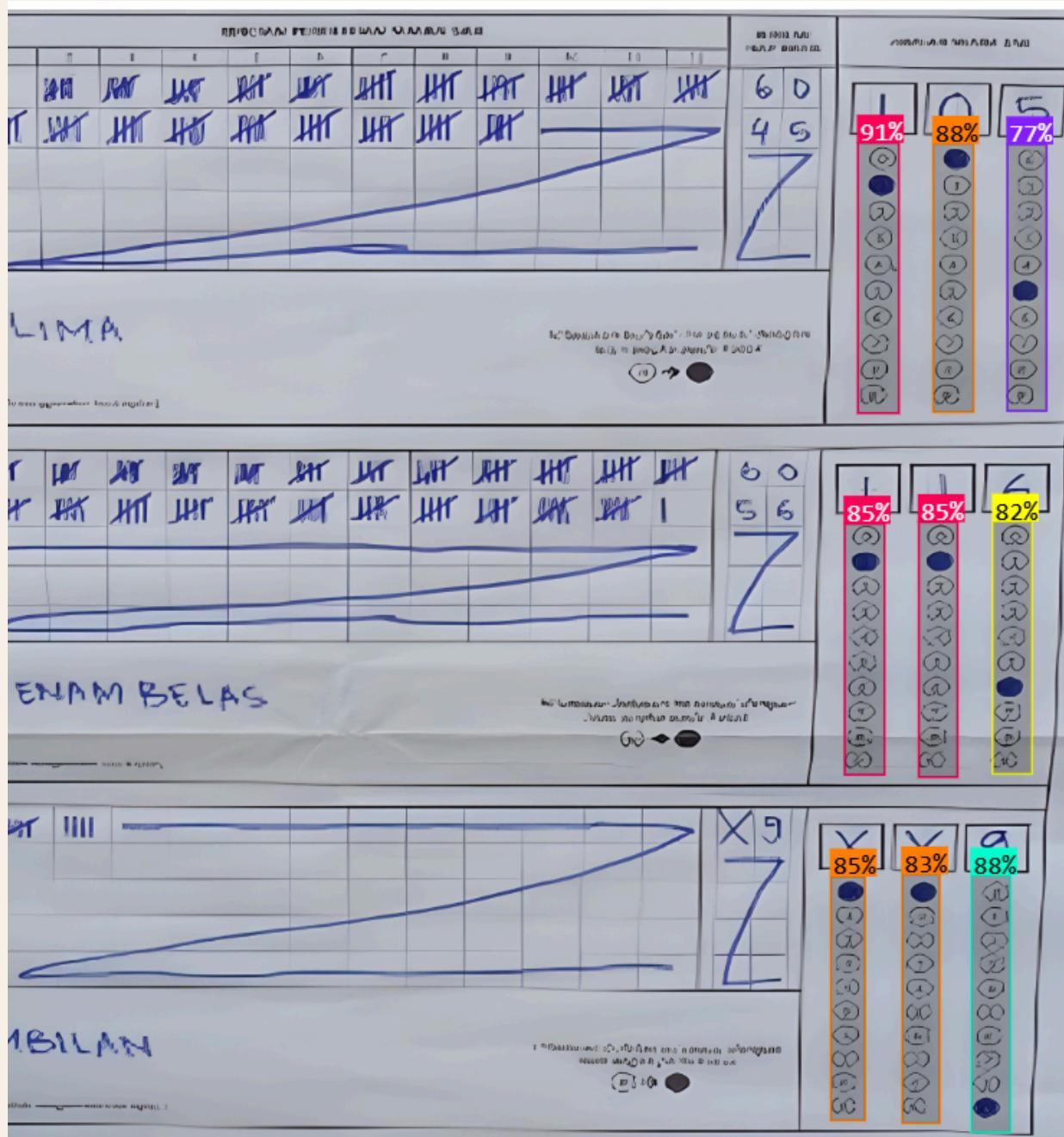
The chatbot web application uses the Google Gemini API for models and embeddings, and with the RAG (Retrieval-Augmented Generation) system for additional context, information documents are converted into embeddings using the Google Gemini embeddings API for searching information that matches the prompt/query.

tools: langgraph, langchain, google gemini, faiss, rest api (flask/fast api)

source code: [github](#)

Proyek - Data Science Competition

Automatic Ballot Counting in General Elections



Implementation of Optical Mark Recognition (OMR) using YOLOv8 to read and count election ballots by finding the right combination of hyperparameters and Real-ESRGAN restoration assistance to suit in terms of model evaluation and prediction results.

award: Best Metric and 2nd place Data Science Competition in kaggle

tools: ultralytics, roboflow

source code and article: [github](#)

Proyek - AI Research

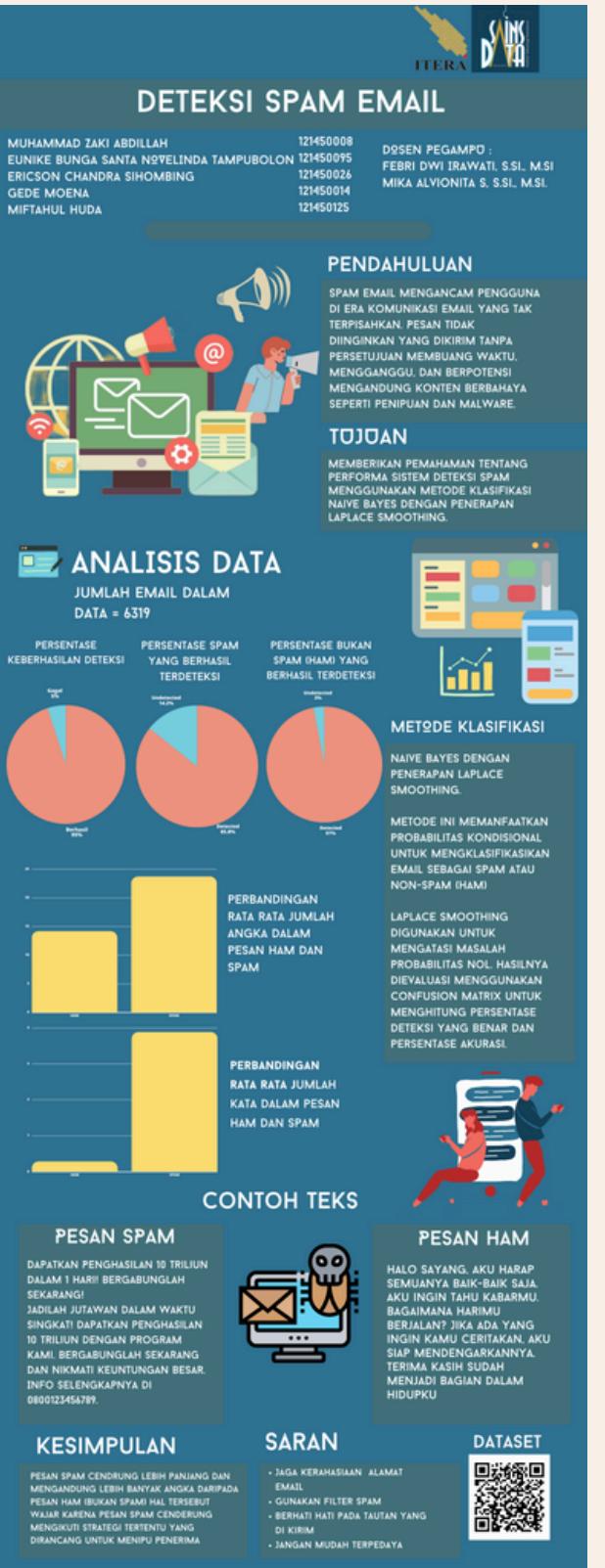
Beach Litter Detection: A Comparative Analysis of RT-DETR Model Variants



Detection and classification of litter on the beach is carried out using the Real-Time Detection Transformer (RT-DETR) model. RT-DETR is a transformer-based object detection architecture designed for real-time data processing with high performance <https://arxiv.org/abs/2304.08069>, making it very suitable for application in environmental monitoring tasks like this.

source code: [github](#)

paper: [arxiv](#)

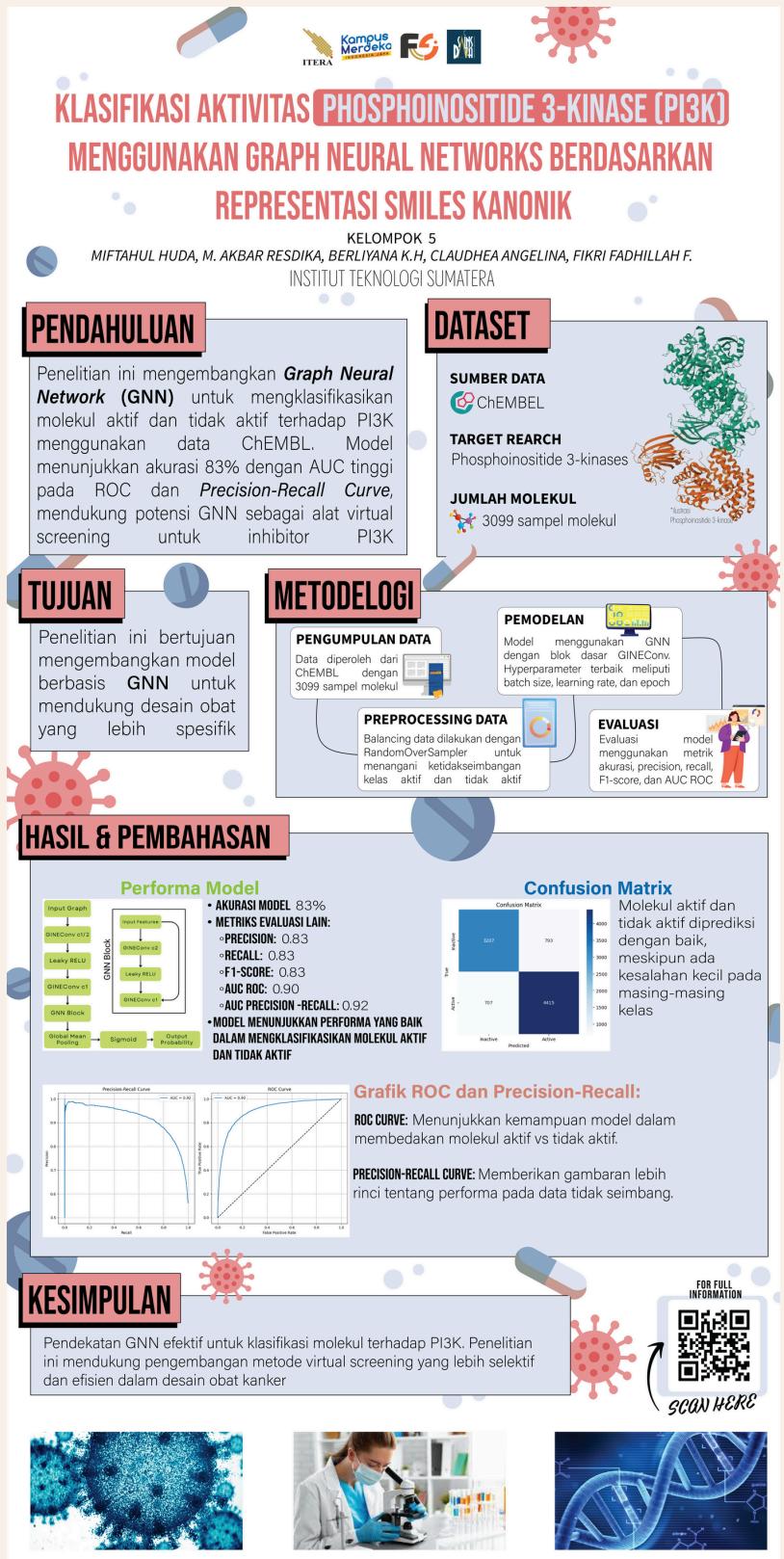


Proyek - Data Scientist

Email Spam Detection

This project is the main assignment for the Statistics Data Science course at the Sumatra Institute of Technology. The project aims to analyze email spam in a dataset and the performance of an email spam detection system utilizing conditional probability (Multinomial Naive Bayes) using R programming.

source code: [github](#)
 poster: [link](#)



Proyek - Data Scientist

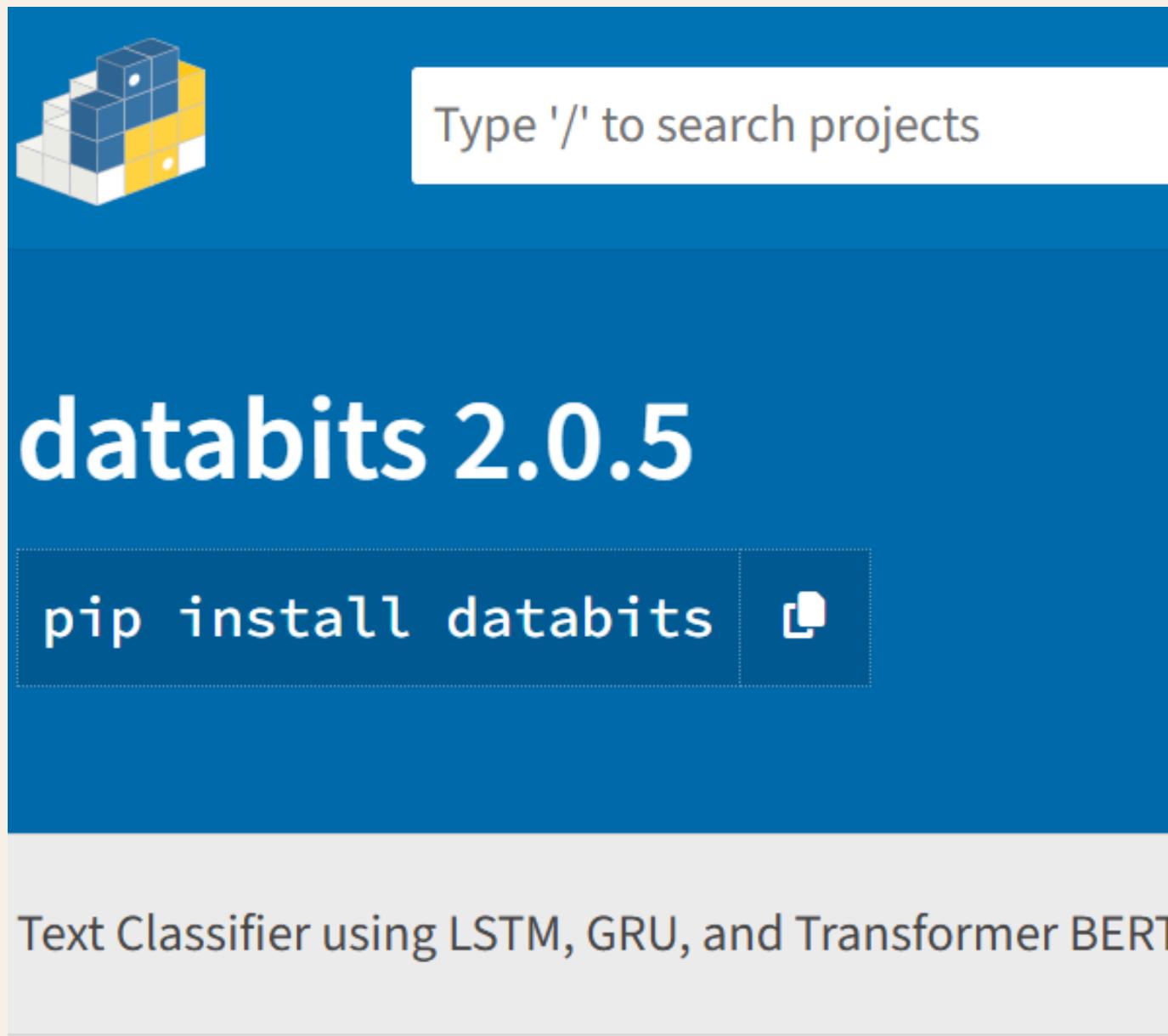
Identification PI3K (activate/inactive) Based on Molecules with Graph Neural Network

This project focuses on the analysis and identification of the bioactive activities of molecules against PI3K targets. Data (from ChEMBL) Molecular structures are converted into graph representations to enable graph learning-based analysis. The prediction stage is carried out by building and training a Graph Neural Network (GNN) model to classify molecules into active and inactive categories.

source code: [github](#)

Proyek - AI Engineer

Custom PyTorch Package for Training Transformers from Scratch



PyTorch-based package (databits) to facilitate transformer models training from scratch.

python package: [pypi](#)
github: [link](#)

Thank You