README – Fake Social Media Accounts Detection (POC)

This folder contains all required deliverables for the POC submission.

Files included:

1. fake_social_accounts.csv

- Raw dataset used in the project.

- Contains synthetic social account activity data.

2. final_compiled_dataset.csv

- Processed and scaled dataset after:

• Outlier handling

• Feature engineering

• Standard scaling

• Final cleaned dataset for modeling

3. predictions_test_set.csv

- Model predictions on X_test.

- Contains:

• Input features

• True labels

• Predicted labels

• Predicted probabilities

4. POC_notebook.ipynb

- Complete end-to-end pipeline including:

• Data loading

• Data inspection

• Data processing

• EDA visualizations

• Train-test split

• ML model training

• Evaluation & metrics

• ROC–AUC curves

• Saving final outputs

Project Summary:

This POC builds a machine learning pipeline to detect fake social media accounts using structured numerical features. Multiple algorithms (Logistic Regression, Random Forest, XGBoost) were trained and evaluated. The best model was chosen based on test accuracy and ROC–AUC performance.

Prepared for:

POC Submission Requirement – AIML Track