Instagram Fake vs Genuine Account Detection

# 1. Project Overview

This project detects fake Instagram accounts using machine learning. It applies data analysis, feature engineering, and classification models to distinguish between genuine and fake accounts.

# 2. Dataset Details

Training dataset: 576 rows, 12 columns  
Testing dataset: 120 rows, 13 columns  
  
Features:  
- Profile picture (binary)  
- Username length and digits  
- Full name length/words  
- Bio/description length  
- Private/Public status  
- Posts, Followers, Following  
- Target: Fake (1) or Genuine (0)

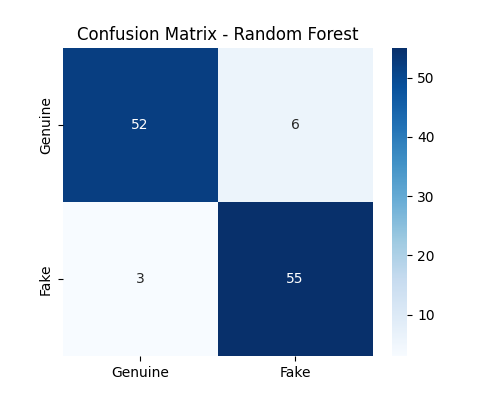
# 3. Program Features

- Loads training and test datasets  
- Performs feature engineering (follower/following ratio)  
- Trains Decision Tree and Random Forest models  
- Evaluates models with accuracy, precision, recall, F1-score  
- Saves trained model with Joblib  
- Exports predictions on test data

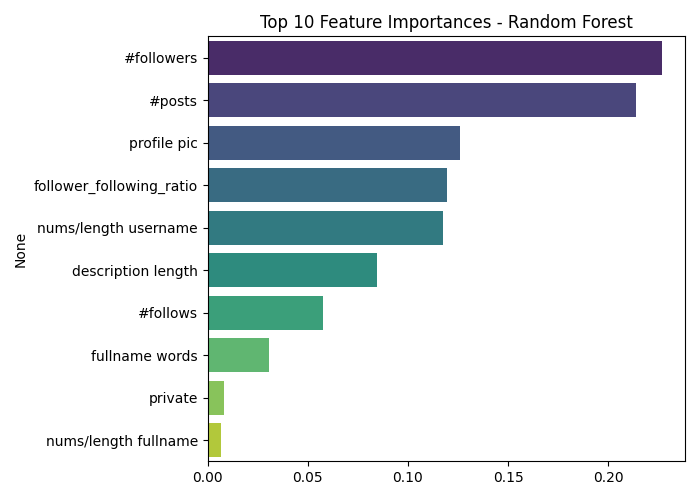
# 4. Results

Decision Tree Accuracy: 90.52%  
Random Forest Accuracy: 92.24%  
  
Confusion Matrix (Random Forest):  
[[52 6]  
 [ 3 55]]  
  
Classification Report:  
 precision recall f1-score support  
  
 0 0.95 0.90 0.92 58  
 1 0.90 0.95 0.92 58  
  
 accuracy 0.92 116  
 macro avg 0.92 0.92 0.92 116  
weighted avg 0.92 0.92 0.92 116

Confusion Matrix Visualization:



Top Feature Importances:



# 5. Future Improvements

- Try XGBoost or Gradient Boosting  
- Handle imbalanced data with SMOTE  
- Deploy as a web app for real-time detection  
- Add more social graph/network features