Instagram Account Classification

Machine Learning • Data Analysis • Feature Engineering

By Asheesh Kumar

FAKE

REAL

Instagram

Objective

- To classify Instagram accounts as fake, spammer, or genuine using machine learning models, exploratory data analysis (EDA), and feature engineering.

Dataset Description

- train.csv: Includes account features and labels (fake = 1, genuine = 0)

- test.csv: Unlabeled Instagram accounts to classify.

- Features include: description_length, #followers, #follows, #posts.

Libraries & Tools

- Python (Pandas, NumPy, Scikit-learn)

- XGBoost, Matplotlib, Seaborn

- SHAP (Explainability)
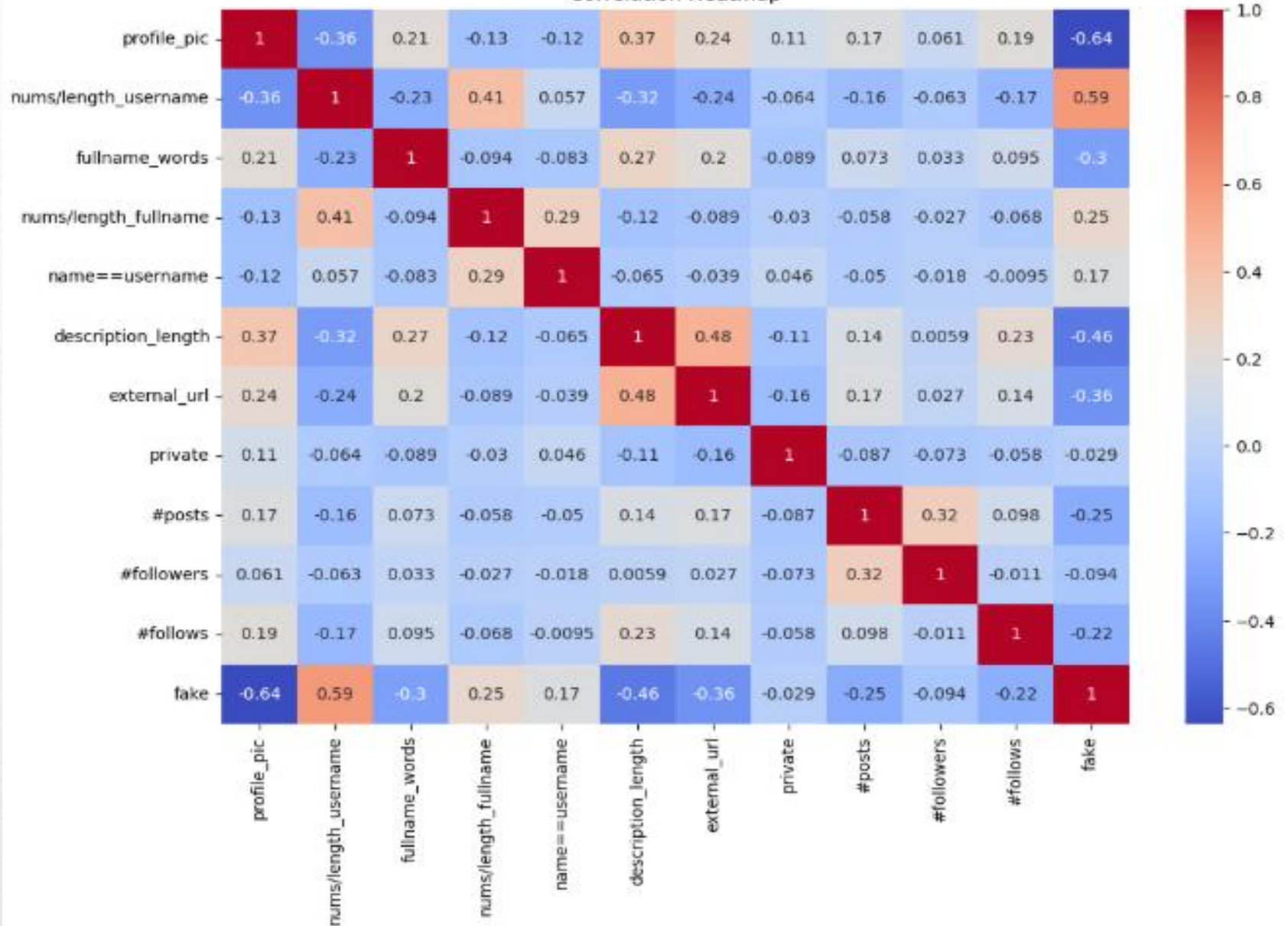
- Jupyter Notebook, VS Code

EDA Highlights

- Data imbalance: more genuine accounts than fake.

- Visualizations: count plots, heatmaps, boxplots.

- Highly correlated features: #followers, #follows, #posts.

Distribution of Account Types
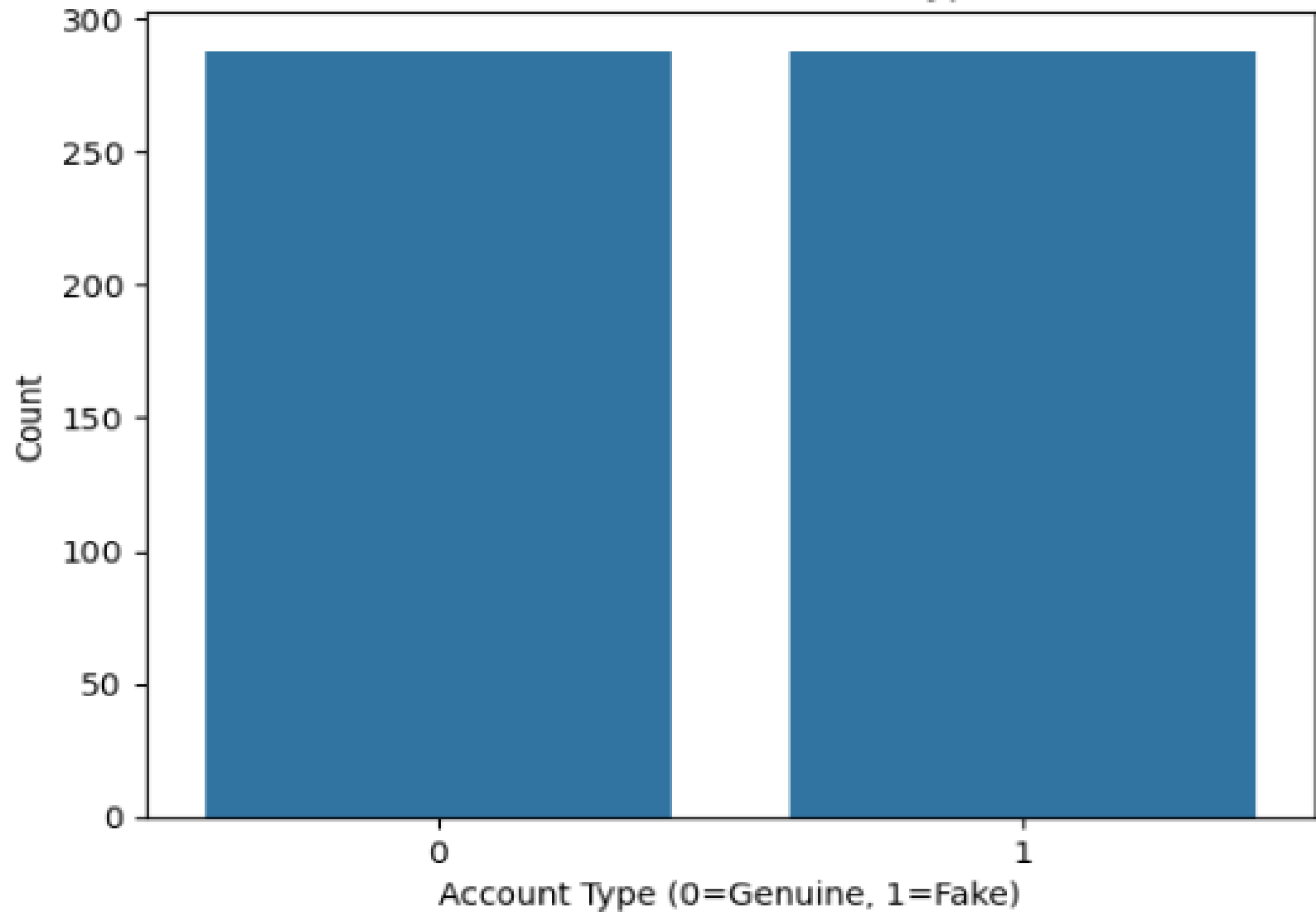
Correlation Heatmap

Preprocessing Steps

- Cleaned and renamed columns.

- Label encoding of categorical features.

- Standardized numerical features.

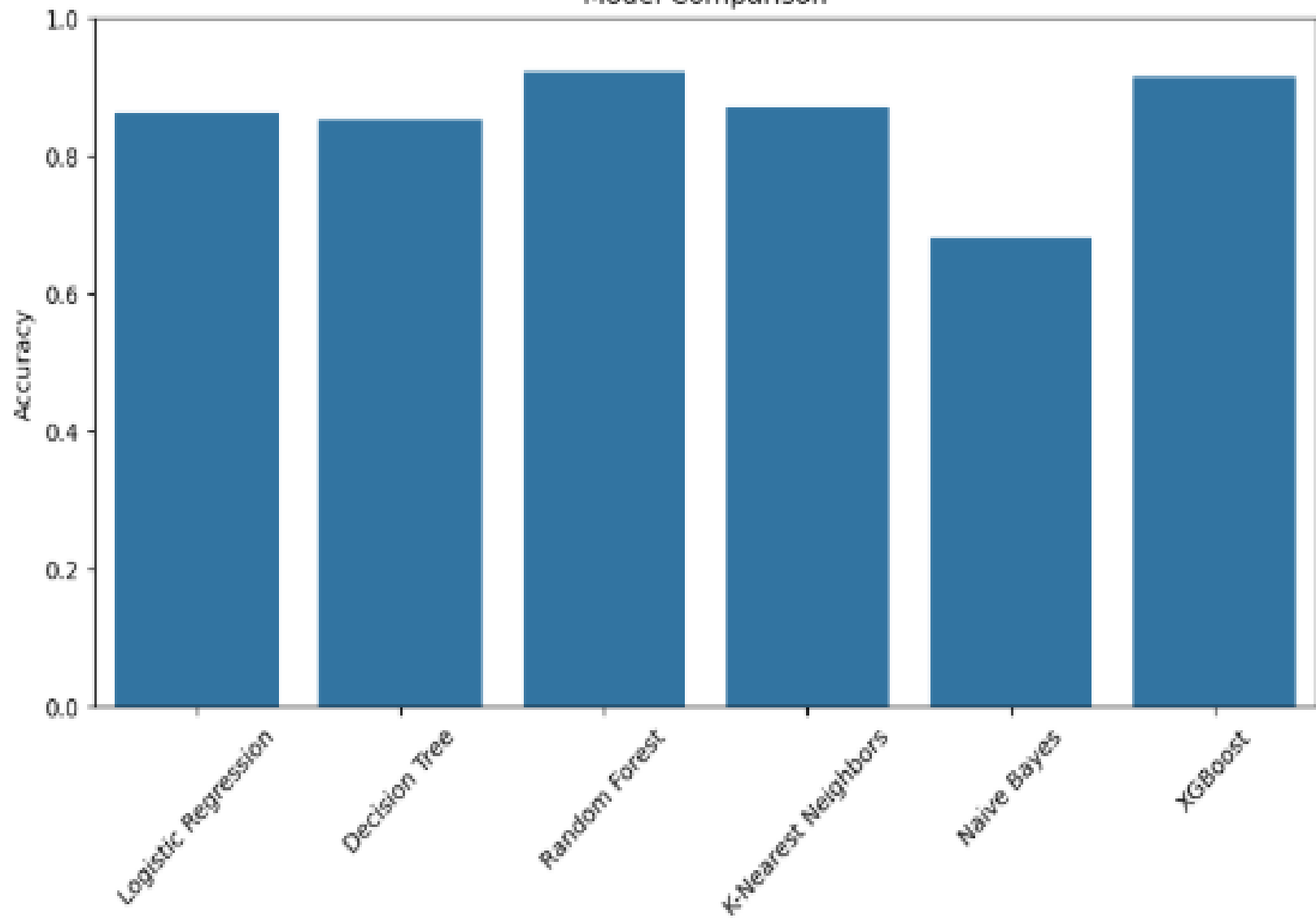- Split dataset: 80% train, 20% validation.

Model Evaluation

- Models Trained:
- Logistic Regression
- Decision Tree
- Random Forest
- KNN
- Naive Bayes
- XGBoost

- Best Accuracy: XGBoost

Model Comparison

Feature Importance

- Top Features from SHAP:

- #followers

- #follows

- description_length

- #posts

Predictions & Output

- Test data classified using best model.

- Predictions saved to final_predictions.csv

```
Model: Logistic Regression
Accuracy: 0.8620689655172413
Confusion Matrix:
 [[60  3]
  [13 40]]
Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.95      0.88        63
           1       0.93      0.75      0.83        53

    accuracy                           0.86       116
   macro avg       0.88      0.85      0.86       116
weighted avg       0.87      0.86      0.86       116
```

Conclusion & Future Scope

- XGBoost performed best on validation data.

- Future Improvements:

- SMOTE for imbalance

- Hyperparameter tuning

- Stacking/Ensembling

Project By

- Asheesh Kumar

- Master's in Operational Research

- University of Delhi