Instagram Fake Account Classifier

Name: Parnika Sunda

Institution: Unified Mentor

Date: July 20, 2025

📊 Instagram Fake Account Classifier – Data Science Project Report

🔍 1. Objective

The goal of this project is to develop a machine learning model that can classify Instagram accounts as either **fake/spammer** or **genuine**. This is based on a range of account-level features such as whether the account has a profile picture, how many followers/followings it has, if it's private, and patterns in the username/fullname.

Detecting fake accounts is essential for maintaining trust on social media platforms and can also help in spam prevention, influencer analysis, and social monitoring.

📁 2. Dataset Overview

Two datasets were provided:

- `train.csv`: 576 rows × 12 columns (used to train and validate the models)
- `test.csv`: 120 rows × 12 columns (used to test model predictions)

📌 Key Features:

Feature	Description
`profile pic`	1 if profile picture exists, else 0
`nums/length username`	Ratio of numeric chars in username
`fullname words`	Word count in full name
`nums/length fullname`	Ratio of numeric chars in full name
`name==username`	1 if name equals username
`description length`	Bio/description character count

`external URL`	1 if a link is in bio
`private`	1 = private account, 0 = public
`#posts`	Number of posts
`#followers`	Number of followers
`#follows`	Number of accounts followed
`fake`	Target label: 0 = genuine, 1 = fake

✅ No missing values were found in either dataset.

📊 3. Exploratory Data Analysis (EDA)

🔹 Class Distribution:

- 288 fake and 288 genuine accounts → perfectly balanced dataset ✅

🔹 Visual Insights:

1. **Profile Picture**:
    - Accounts without profile pictures were more likely to be fake.
2. **Privacy Setting**:
    - Public accounts had slightly more fake users, possibly to reach wider audiences.
3. **Followers/Following**:
    - Genuine accounts had a broader follower distribution.
    - Fake accounts tended to follow fewer than 1000 users.
    - Binned visualizations revealed many fake accounts existed in low follower count ranges (0–50).

4. **Username Patterns**:
    - Accounts with a high ratio of numbers in usernames had higher fake probability.
5. **Heatmap + Histograms**:
    - Strong correlation observed between `#followers`, `#follows`, and the fake label.
    - Feature distributions were skewed in a way that supported class imbalance clues (e.g., lots of low-follower accounts were fake).
6. **Scatter Matrix & Boxplots**:
    - Gave multivariate view of fake vs genuine data clusters, confirming earlier patterns.

🤖 4. Model Building

We implemented and evaluated **two classification models**:

🧪 1. Logistic Regression

📌 **Why?** Simple, linear baseline model that's easy to interpret.

- Applied `StandardScaler` to normalize features.
- Model trained on 80% of data and validated on 20%.
- Evaluation Metrics:

```
Accuracy:      86.2%
Precision(1): 93%
Recall(1):     75%
F1-score(1):   83%
```

✅ **Strength**: High precision → low false positives (when model says “fake,” it’s mostly correct)
⚠️ **Weakness**: Lower recall → it missed some fake accounts.

🌳 2. Decision Tree Classifier

📌 **Why?** Good for capturing non-linear relationships and providing visual insight.

- No feature scaling needed
- Plotted decision tree for explainability
- Showed feature importance

text
CopyEdit

```
Accuracy:      87.07%
Precision(1): 88%
Recall(1):     83%
F1-score(1):   85%
```

✅ **Strength**: More balanced between precision and recall
✅ **Strength**: Clearly interpretable via tree plot
✅ **Important Features Identified**:

- `#followers`, `profile pic`, `private`, `nums/length username`, `description length`

📊 5. Model Comparison

Metric	Logistic Regression	Decision Tree
Accuracy	86.2%	**87.1%** ✅
Precision (Fake)	**93%**	88%
Recall (Fake)	75%	**83%** ✅
F1-score (Fake)	83%	**85%** ✅
Explainability	❌ Limited	✅ High

📌 **Conclusion**:
The **Decision Tree model slightly outperformed Logistic Regression**, with a more balanced performance and higher explainability. It successfully identified key patterns among fake accounts, such as having no profile picture, low follower count, and username containing numbers.

🚀 6. Final Predictions

The best-performing model (Decision Tree) was used to make predictions on the `test.csv` dataset. The predicted results can be used to flag or review suspicious accounts further.

🌱 7. Future Work

- Apply ensemble models like **Random Forest** or **XGBoost** for better generalization
- Add text-based features like:
    - **Bio sentiment**
    - **Caption content**
    - **Hashtags usage**

- Add time-based behavior (e.g., posting frequency, account age)
- Collect **real-time Instagram data using APIs** (if permitted)
- Try **cross-validation** and **hyperparameter tuning** for robustness

🏁 Final Thoughts

This project demonstrates how simple features can be powerful in identifying fake or spam accounts on social media. With further enhancement and real-time data, such models can become an essential part of digital safety and content moderation tools.