Thuy (Jade) Tran

Capstone Project Two Report - Detecting Fake Instagram Accounts

Problem Statement

How Instagram can ensure a safe and authentic social networking environment by detecting fake accounts.

Background

My client for this project is Instagram, one of the largest video and photo sharing social networking companies with more than 40 billion photos uploaded and nearly 1.4 billion users as of 2022. According to the Federal Trade Commission, social media was far more profitable to scammers in 2021 than any other method of reaching people. With the growth of shopping functionalities, as well as other engagement features on the platform, Instagram's Trust & Safety department needs to minimize the number of scams and account abuses using Instagram. I used classification techniques to find the probability for an account to be fake so that Trust & Safety teams can implement further actions.

Data

I used a data set that includes 1002 real accounts and 201 fake accounts. For each account, there are 9 variables. They are information about the account followers and following numbers, username characteristics, as well as whether the account has a profile picture or is private. The data is in JSON format and weighs 39.9 KB.
Here are the variables:

- user_media_count - Total number of posts an account has.
- user_follower_count - Total number of followers an account has.
- user_following_count - Total number of followings an account has.

- user_has_profil_pic - Whether an account has a profile picture, or not.
- user_is_private - Whether an account is a private profile, or not.
- user_biography_length - Number of characters present in account biography.
- username_length - Number of characters present in account username.
- username_digit_count - Number of digits present in account username.
- is_fake - True, if account is a spam/fake account, False otherwise

Data Wrangling and Data Cleaning

After downloading the file into the local machine, I use panda's read_json method to load the json file into 2 panda dataframes, 1 for real accounts and 1 for fake accounts. I didn't find any null values. The data in categorical and numerical variables also look very clean.

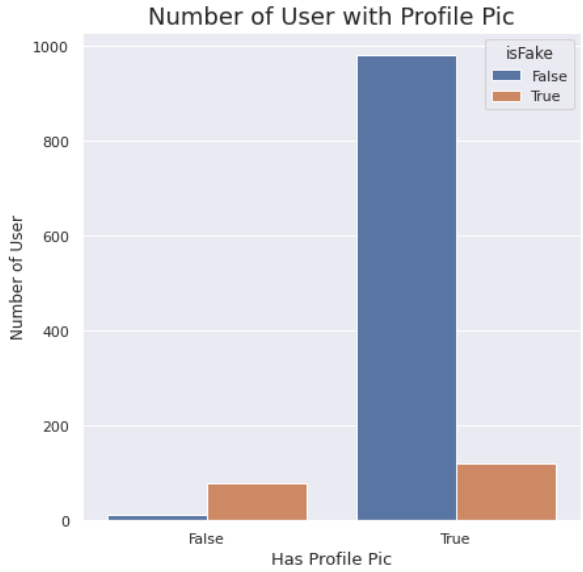
	userFollowerCount	userFollowingCount	userBiographyLength	userMediaCount	userHasProfilPic	userIsPrivate	usernameDigitCount	usernameLength	isFake
0	258	238	0	0	1	0	0	10	0
1	263	482	30	29	1	1	0	8	0
2	51	78	9	0	1	1	0	10	0
3	297	480	22	25	1	1	2	9	0
4	113	242	0	95	1	1	0	10	0

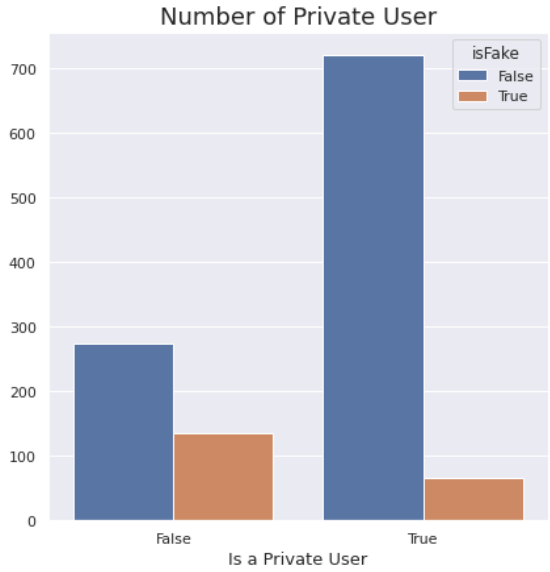
Exploratory Data Analysis

I compared the distribution of data points in the same variable in real and fake data sets and saw that there are some slight differences.

For example, the majority of the **number of followers ("userFollowerCount"**) in the real data set is more than zero, while it's mostly zero in the fake data set. The distributions also look different in other variables. It might indicate that each of the variables might have some correlation with "isFake".

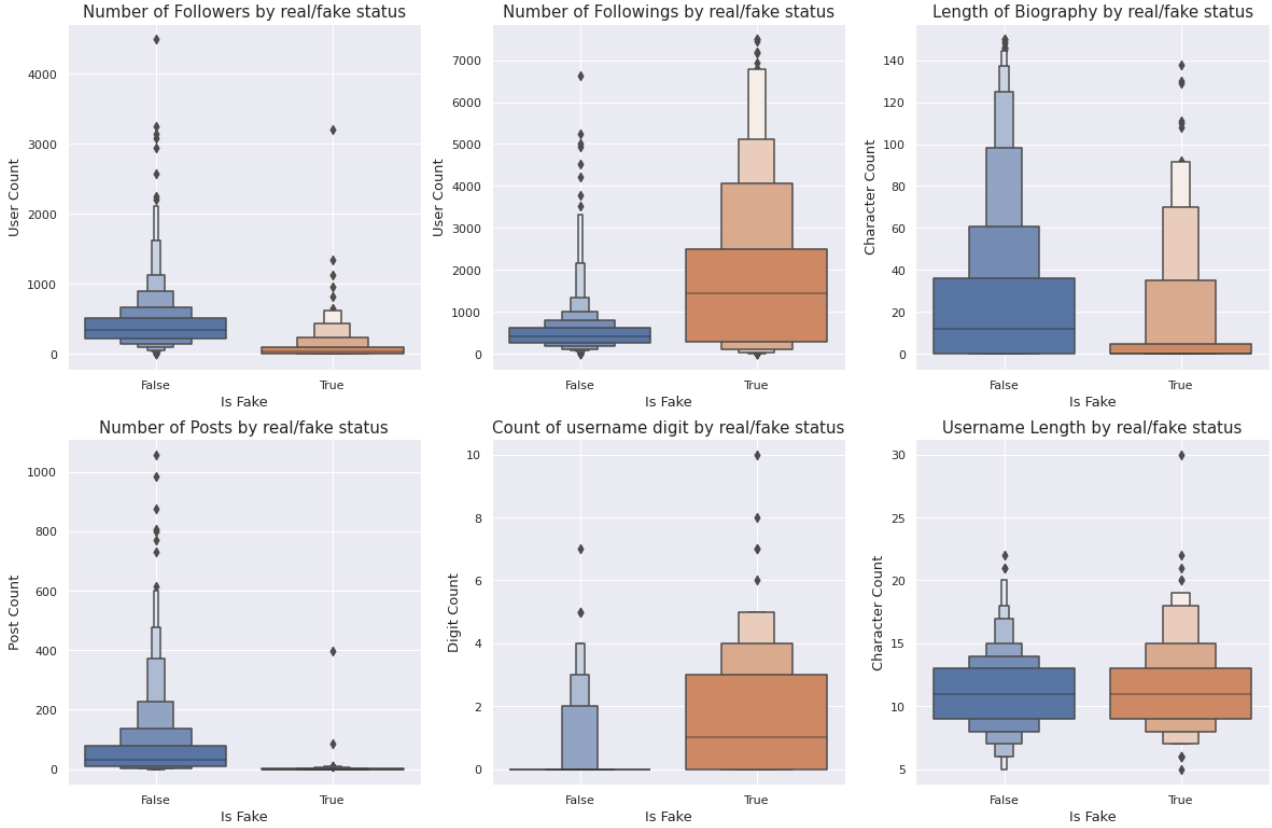
Categorical variables



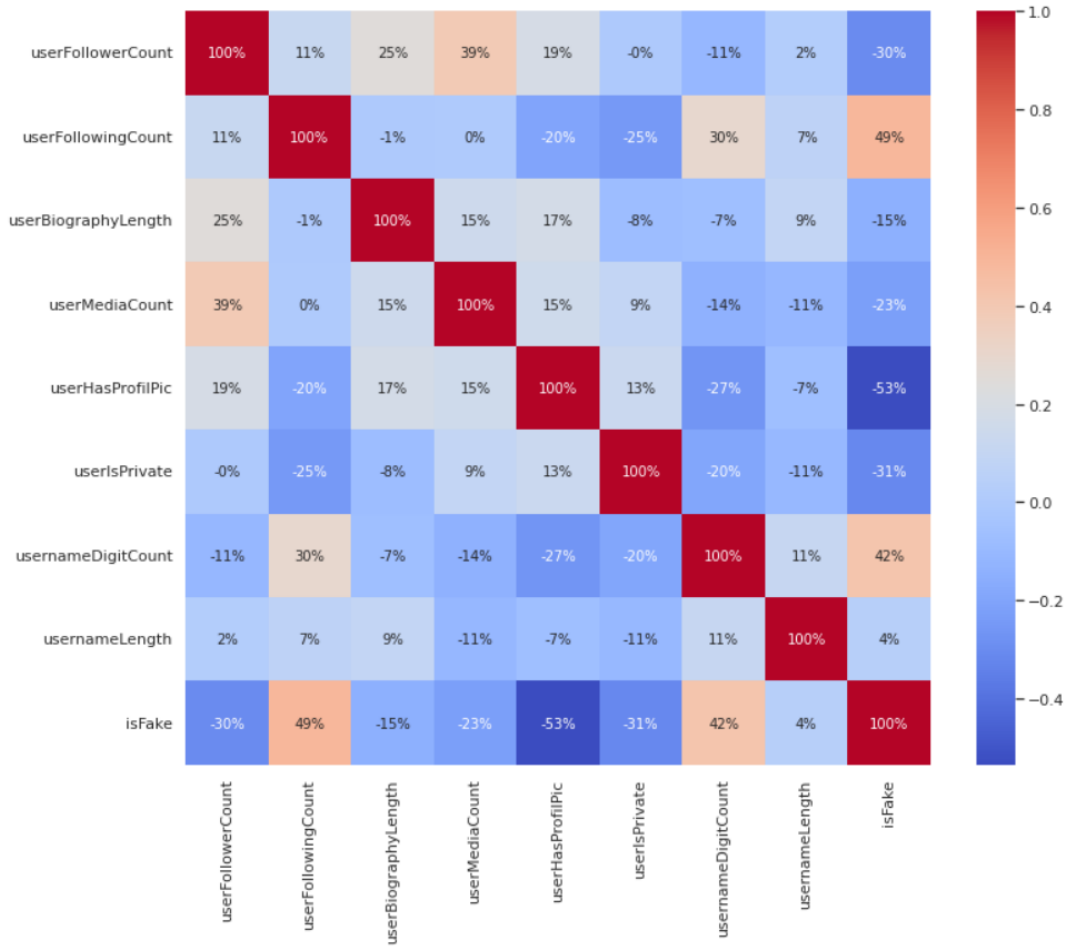


If a user doesn't have a profile picture, it's more likely that they are a fake user than a real user. In contrast, if they have a profile picture, it's more likely that they are a real user. If a user is a Private user, it's more likely that they are a real user.

Numerical variables



Correlations among variables



Pre-processing & Training Data Development

Since the data set is imbalanced with the ratio of fake to real accounts being 1:5, I've applied an oversampling technique to increase the size of the minority class (fake accounts).

Training and Evaluating Models

I applied 3 machine learning techniques to the testing and training data set: Logistic Regressions, Decision Tree and Random Forest classifier.

Random Forest Classifier technique has the highest accuracy score of 0.98.

Next Steps

- Investigate the quality of the model and possibility of overfitting
- Explore additional methods to improve the techniques of data sampling

Appendix

Data article: https://arxiv.org/pdf/1910.03090.pdf

Download link: https://github.com/fcakyon/instafake-dataset

Staff, the P. N. O., This blog is a collaboration between CTO and DPIP staff and the AI Strategy team, Fletcher, E., & Fletcher, E. (2022, January 27). *Social media, a gold mine for scammers in 2021*. Federal Trade Commission. Retrieved March 27, 2022, from https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2022/01/social-media-gold-mine-scammers-2021