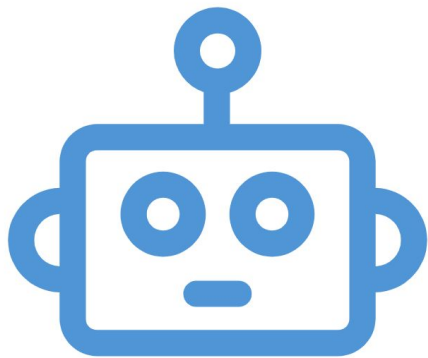# Twitter Bot or Not?
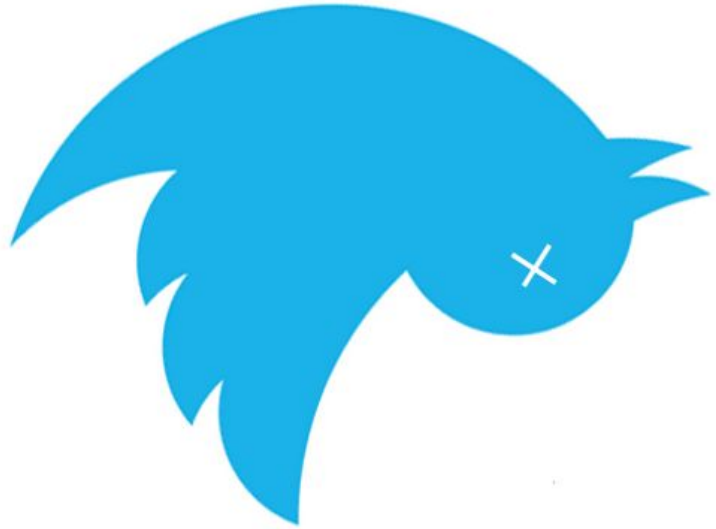
Why do we care about bots?

A **Twitter bot** may autonomously **tweet, retweet, like, follow, unfollow,** or **direct message** other accounts

Which can mean…

➜ Spreading **misinformation**
➜ **Scams** and **exploitation**
➜ **Degrading** social **trust**
➜ Social media **platform deterioration**
➜ **Increased polarization**

Twitter account-level features

Profile images

kris doenlen
@scrapfishies

Oakland, CA   github.com/scrapfishies   Joined July 2020

65 Following   **2** Followers

Tweets   Tweets & replies   Media   Likes

Twitter account-level features

Profile images

Network size

kris doenlen
@scrapfishies

⌖ Oakland, CA 🔗 github.com/scrapfishies 🗓 Joined July 2020

65 Following **2** Followers

Edit profile

Tweets   **Tweets & replies**   **Media**   **Likes**

Twitter account-level features

Profile images

Network size

kris doenlen
@scrapfishies

⊙ Oakland, CA   🔗 github.com/scrapfishies   🗓 Joined July 2020

65 Following   **2** Followers

Date & time created

Edit profile

Tweets   **Tweets & replies**   **Media**   **Likes**

Twitter account-level features

Profile images

Network size

Date & time created

kris doenlen
@scrapfishies

◎ Oakland, CA   🔗 github.com/scrapfishies   🖩 Joined July 2020

65 Following   **2** Followers

Tweets   **Tweets & replies**   **Media**   **Likes**   User activity

Edit profile

Bots behave differently...

Distribution of Tweets per Day
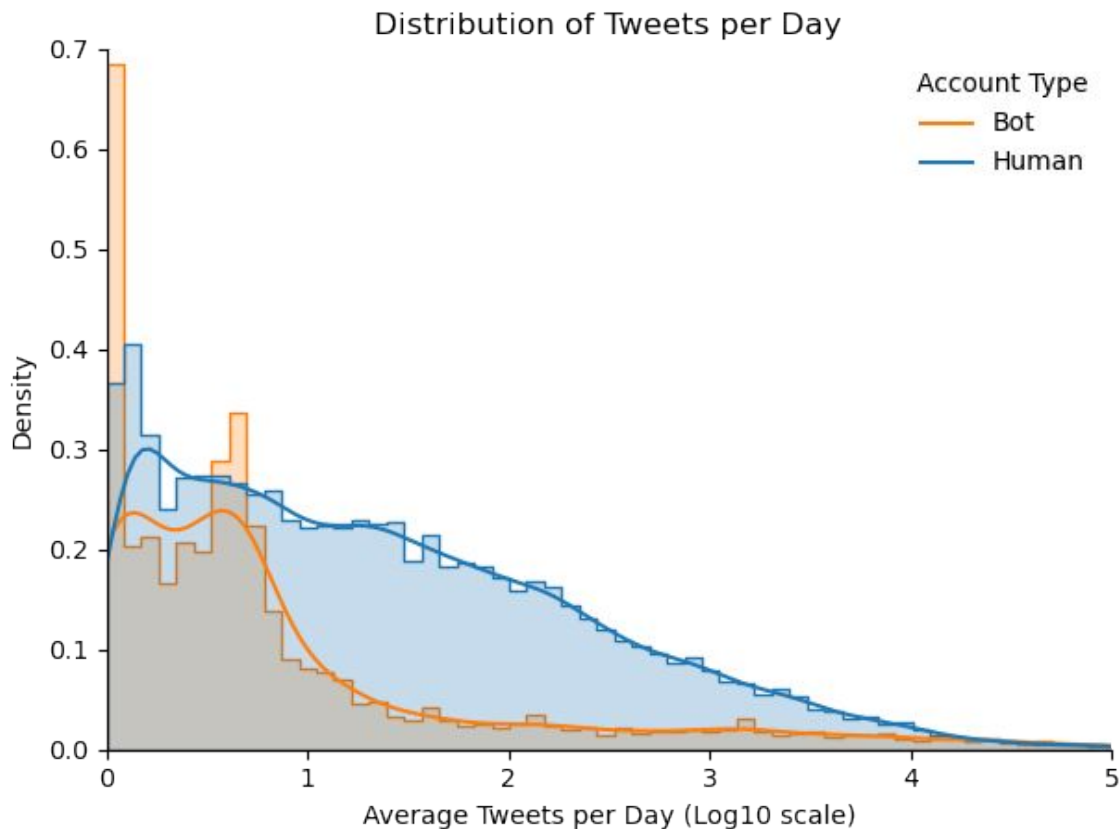
...and their networks look different, too

Distribution of Followers & Friends
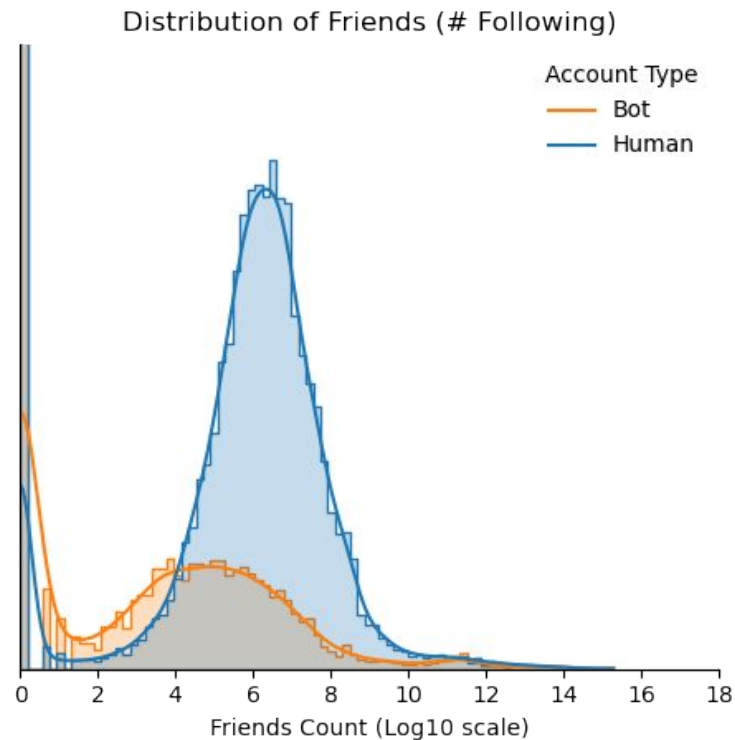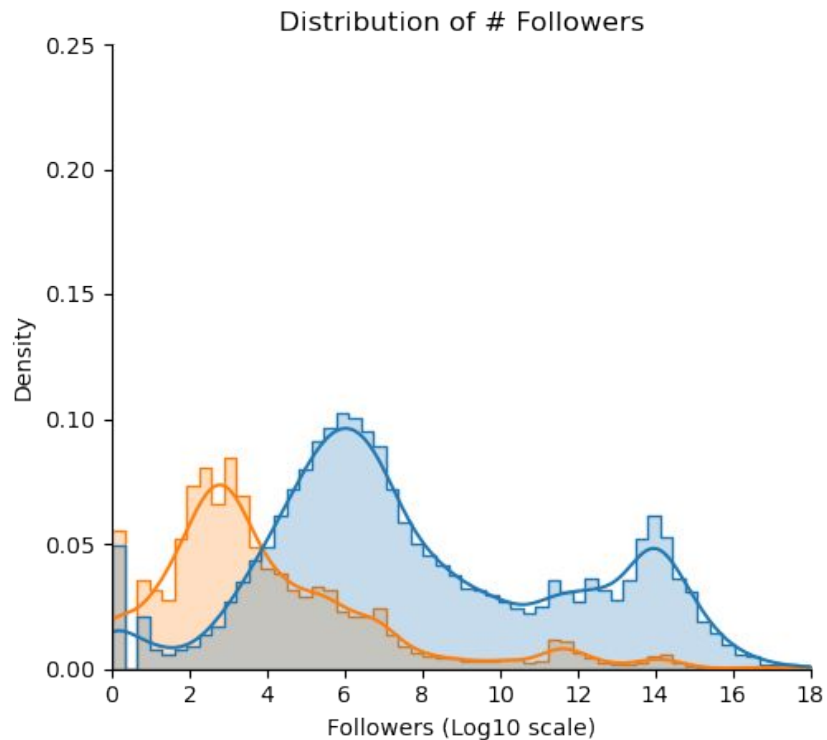
Model construction and selection

Feature Engineering

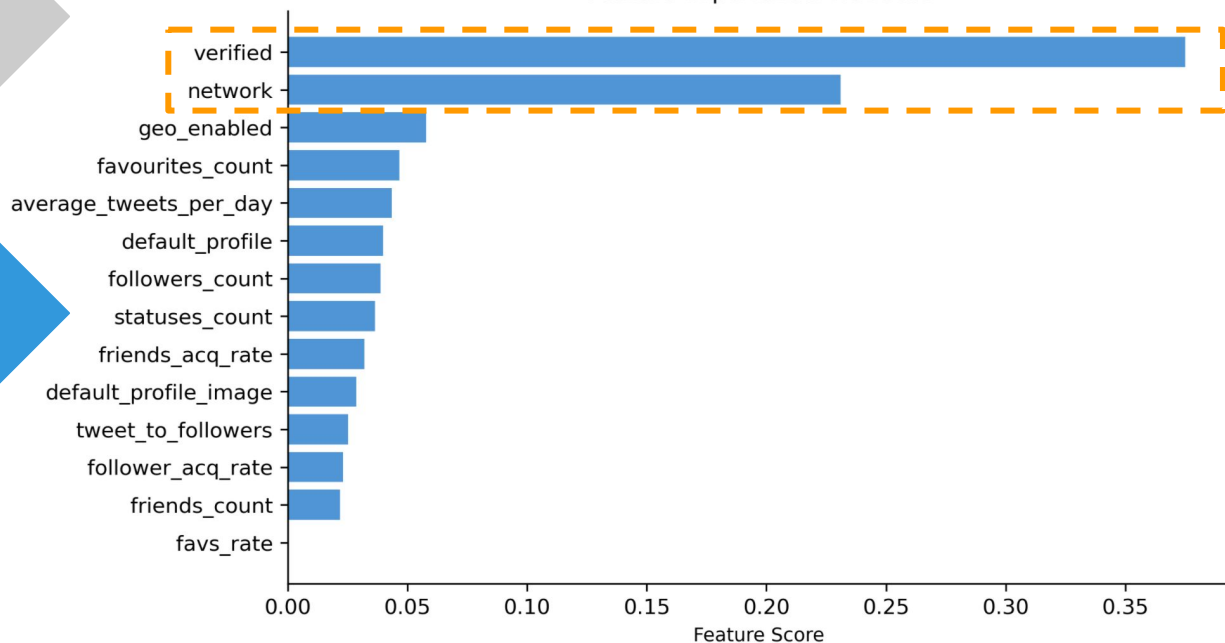Tweets to Followers Metric
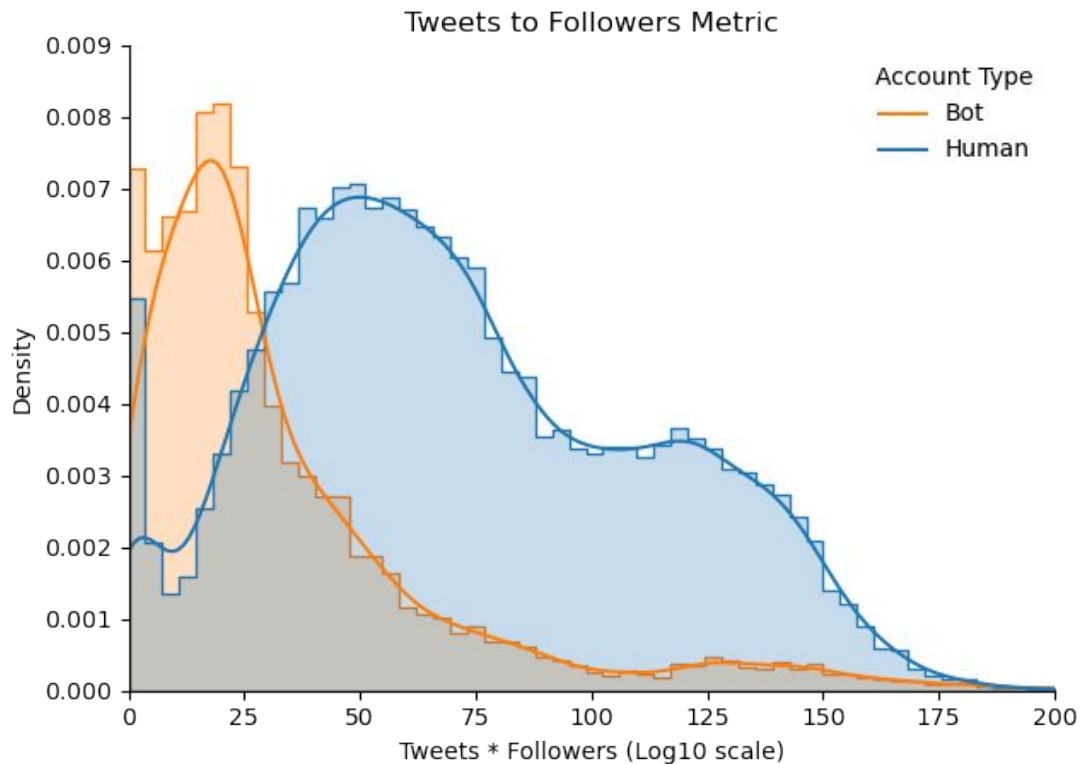
Model construction and selection

Feature Engineering

Feature Selection

Feature Importance: XGBoost

Model construction and selection

Feature Engineering

Feature Selection

Parameter Tuning

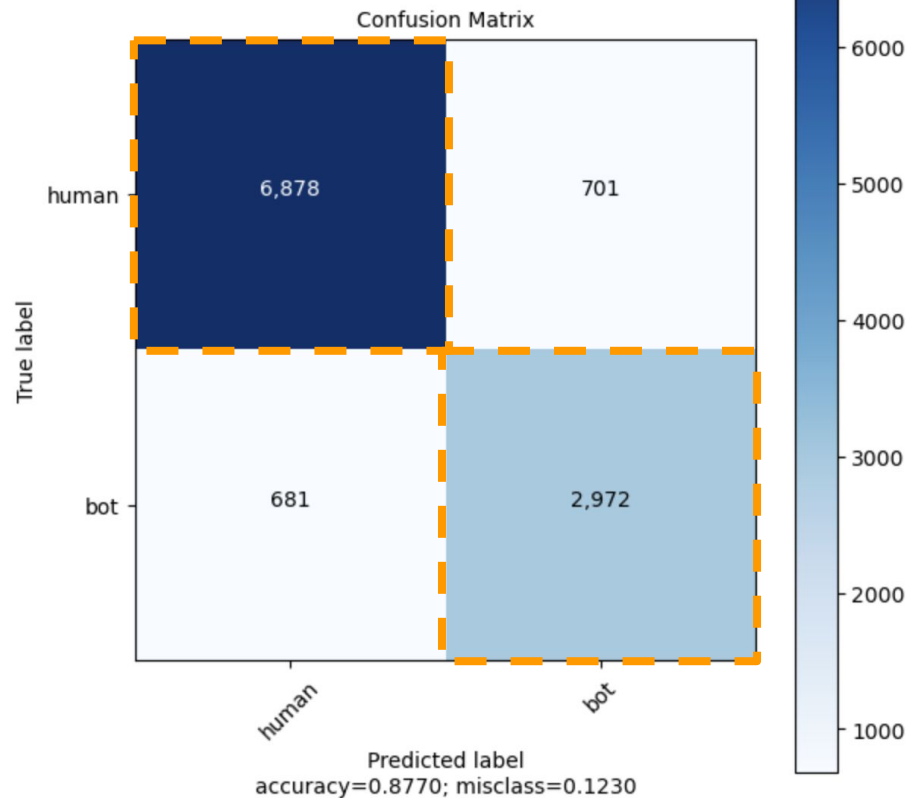Cross-Validated Precision-Recall Curves: XGBoost

Fold 1: Precision: 0.81 / Recall: 0.81
Fold 2: Precision: 0.82 / Recall: 0.82
Fold 3: Precision: 0.79 / Recall: 0.80
Fold 4: Precision: 0.81 / Recall: 0.80
Fold 5: Precision: 0.80 / Recall: 0.79
No Skill
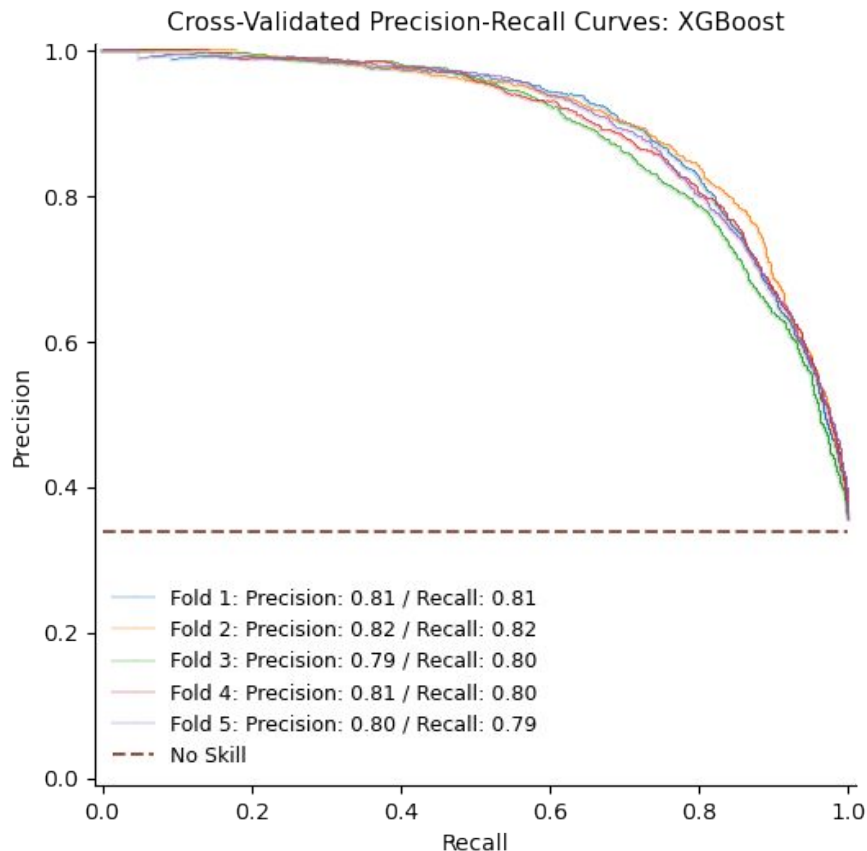
XGBoost model results

Metric	Test Score (%)
Accuracy	87.7
Precision	81.3
Recall	80.9
ROC AUC	93.4

Confusion Matrix

	human	bot
human	6,878	701
bot	681	2,972

True label / Predicted label

accuracy=0.8770; misclass=0.1230

Twitter Bot or Not

Enter a Twitter handle

scrapfishies

nytimes

willdoenlen

atotalpirate

aoc

sdf

Twitter Bot or Not by Chris Doenlen |

Final thoughts

Not all bots are bad!

➔ Sometimes they're **creative**, **helpful**, or **funny**
➔ It all depends on **how they were *designed***
➔ **Bot <u>awareness</u> is key**

Suggestions for future work

➔ Tweet-level features
➔ **Build a bot to detect bots**
➔ Classify *types* of bots (political, fake followers, trolls, financial, spammers….)
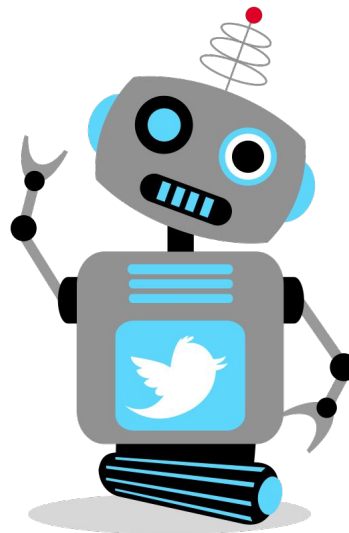
Thank you!

Chris Doenlen

Twitter Bot or Not

GitHub Repository

scrapfishies@gmail.com

@scrapfishies
(Probability of being a bot: 23.9%)

Appendix

Try out the live app: **Twitter Bot or Not**

Sources & References

➔ **Kaggle: Twitter Bot Accounts**
➔ **Bot Repository**
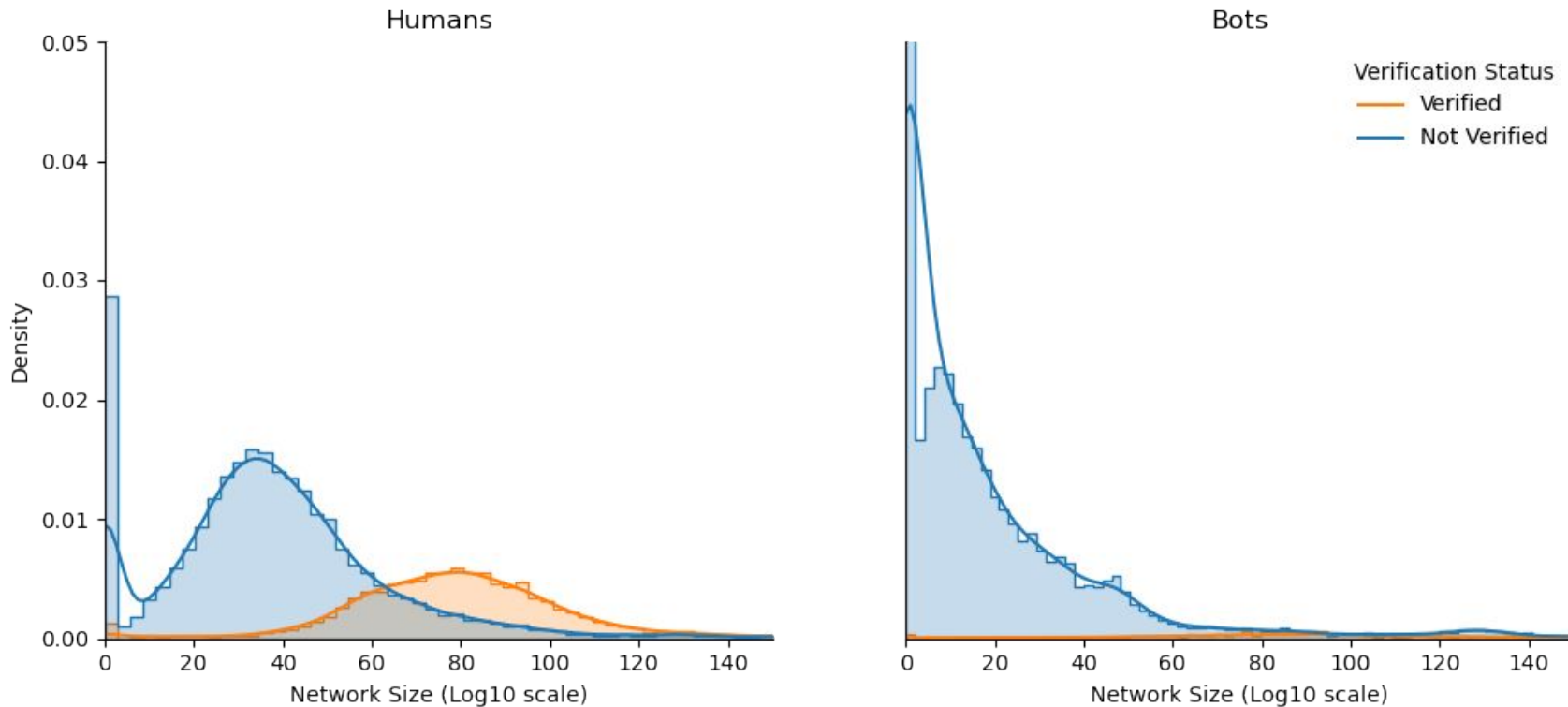➔ **Tweepy**
➔ **Twitter**

Appendix: Final XGBoost model features

Feature Importance: XGBoost
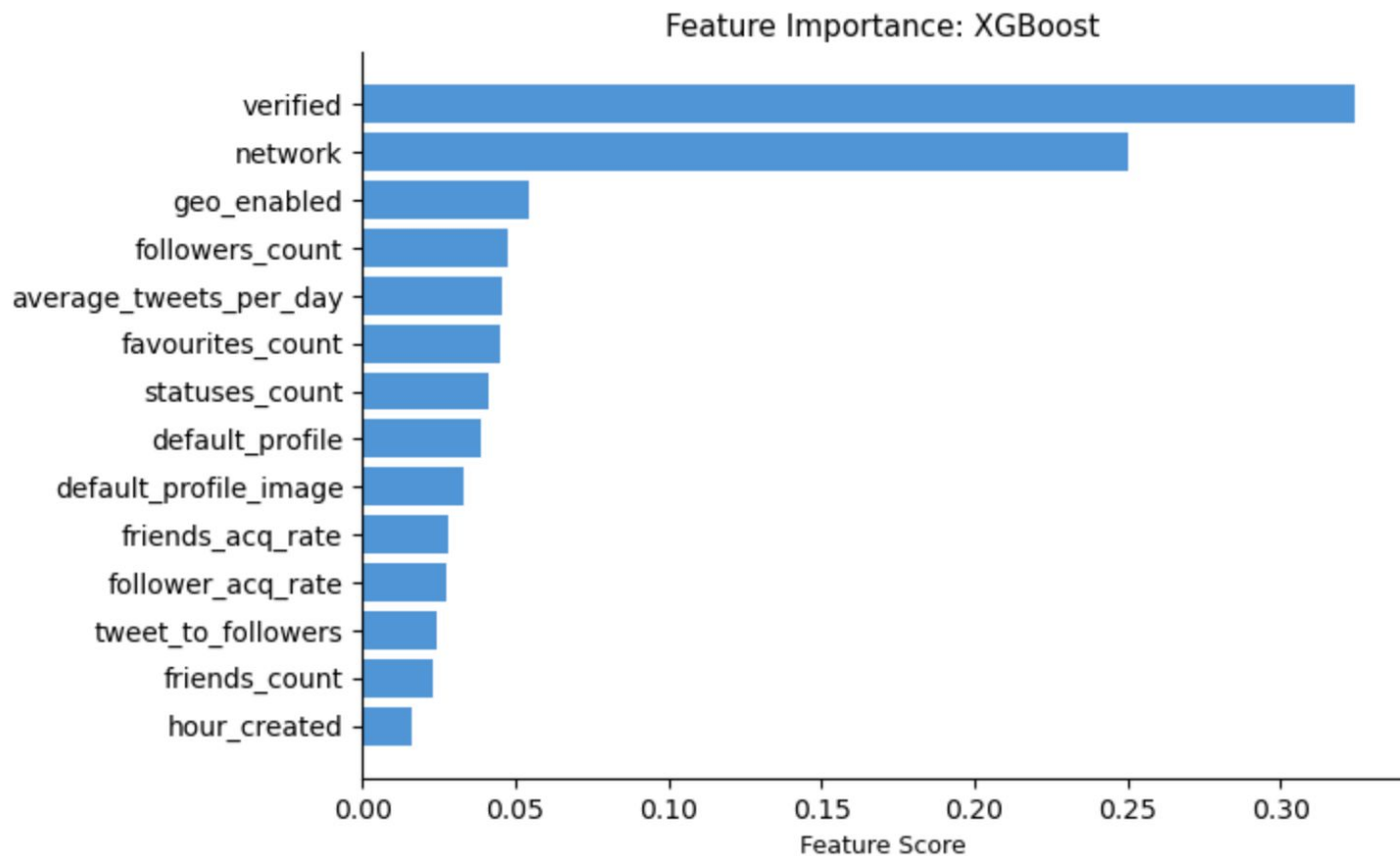
Appendix: Top model features

Network Size by Verification Status

Appendix: Final XGBoost model performance curves