



IDENTIFYING FAKE NEWS AMONG TWITTER USERS

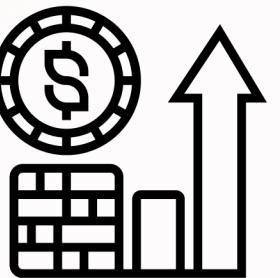


(Prachi Dalal, Martin Navarro, Gnandeep Patlolla, Vikram Ramanathan, Muhammad Ahson Riaz, and Rimsha Tahir)



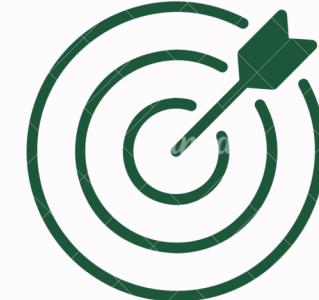
THE IDEA

1. Predicting if the news is fake
2. If people who share fake news are similar
3. Determining factors



THE BENEFITS

1. Reduce the spread of misinformation
2. Increase credibility
3. Easier detection of fake news



THE GOAL

1. Algorithm to detect fake news
2. Best characteristics to identify fake news

THE DATA

The explanatory variables (X)

Numeric Data

	followers	following	statuses_count	favorite_count
count	1305547.00	1305547.00	1305547.00	1305547.00
mean	52367.16	2865.04	96697.55	5.27
std	765935.10	13803.44	212651.03	237.72
min	0.00	0.00	0.00	0.00
25%	67.00	76.00	7993.00	0.00
50%	549.00	498.00	23836.00	0.00
75%	2406.00	1867.00	91389.50	0.00
max	88738155.00	1209332.00	9483731.00	140527.00

Binary Data

	verified	profile_background_tile	profile_use_background_image
False	1253436	1058071	242275
True	52111	247476	1063272

The Target variables (Y)

News	Count
Real (0)	660019
Fake (1)	645528

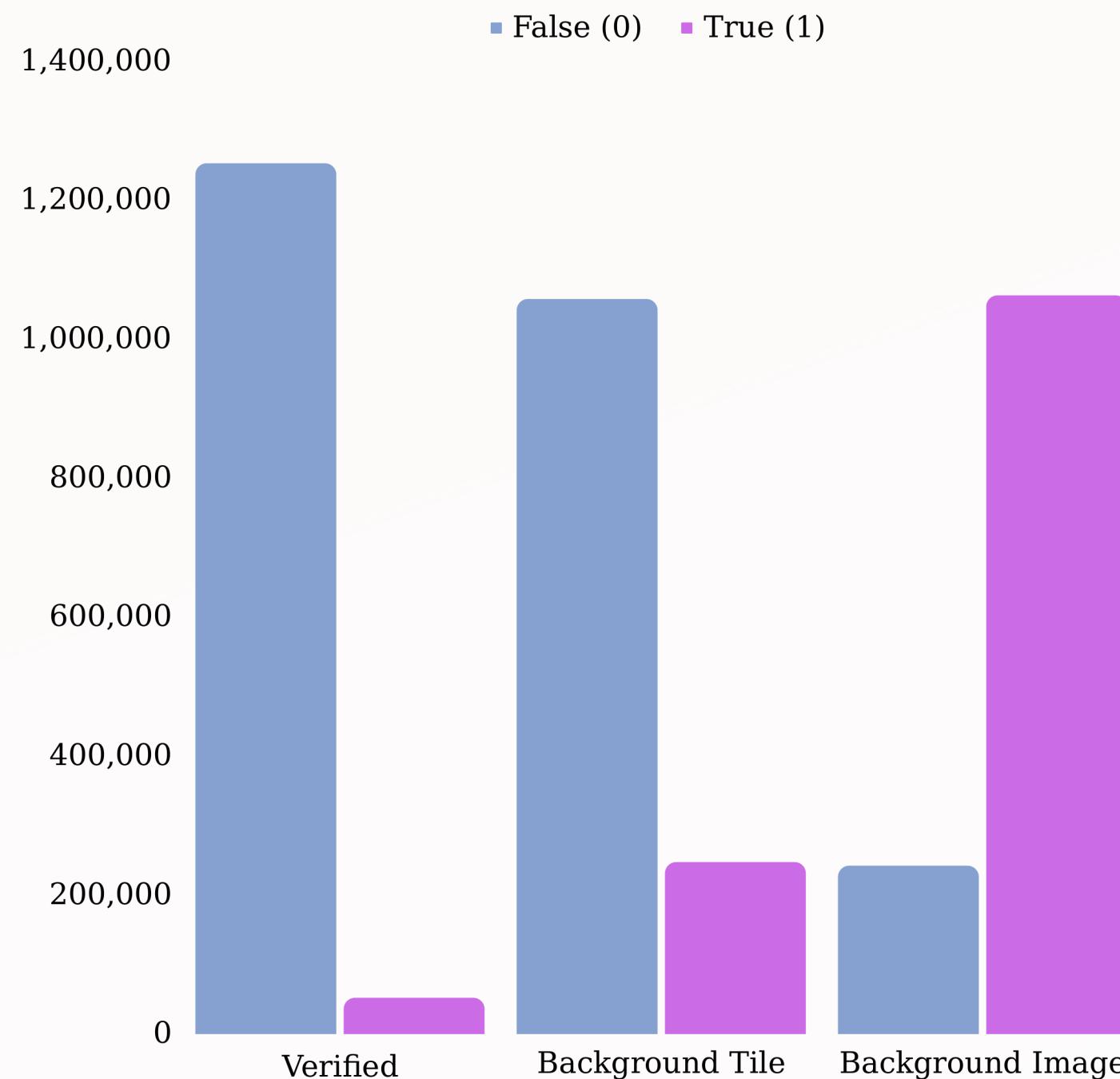
Features

1. Number of followers
 2. Number of people followed
 3. Count of statuses posted
 4. Number of favorite posts
 5. Is the account verified?
 6. Is there a profile tile image?
 7. Is there a profile background image?
 8. Is the news real or fake?
- • • •
• • • •
• • • •
• • • •

EXPLORATORY DATA ANALYSIS

The explanatory variables (X)

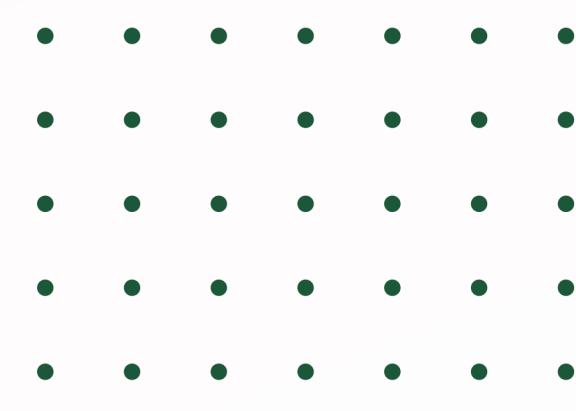
Binary Data



Averages:

- **2,865** Average Following
- **52,367** Average Followers
- **5** Average Favorite Posts
- **96,698** Average Posts (Status_Count)

The Target variable (Y)



THE TECHNIQUE LOGISTIC REGRESSION

The Concept: Evaluate the impact of user characteristics (X) in sharing fake news (Y)

How it works:

- 1 Likelihood: Which factors effect whether an individual shares fake news
- 2 Economic Importance: Are these factors also the most important in identifying fake news?

THE RESULTS

varname	Coefficient	OddsRatio	pvalues	sd	dummy	X	importance
verified	0.4517	1.5711	0	0.1958	1	1.5711	1.5711
profile_background_tile	0.4513	1.5704	0	0.392	1	1.5704	1.5704
statuses_count	-0.2569	0.7734	0	1	0	0.7734	1.2929
profile_use_background_image	-0.1745	0.8399	0	0.3888	1	0.8399	1.1907
followers	-0.0417	0.9592	0	1	0	0.9592	1.0426
favorite_count	0.0078	1.0078	0.0009	1	0	1.0078	1.0078
following	0.0069	1.0069	0.0003	1	0	1.0069	1.0069

Key

Green: Increases likelihood
Red: Decreases likelihood

Significance:

- All variables are statistically significant, **even variables like having a profile picture or a background picture**
- Being verified has a huge impact on a user's willingness to share fake news

The Conclusion

What can we do?

01

Give verification status based on content shared

02

Caution on news shared by individuals with low status count, less number of followers and higher number of people followed

03

Based on importance, use algorithm to label potentially fake news to warn viewers (Like sensitivity label on instagram/facebook)



Education Research Innovation Admissions + Aid More

MIT News

ON CAMPUS AND AROUND THE WORLD

[SUBSCRIBE](#)

[SEARCH](#)

Study: On Twitter, false news travels faster than true stories

Research project finds humans, not bots, are primarily responsible for spread of misleading information.

[Watch Video](#)

Peter Dizikes | MIT News Office
March 8, 2018

[PRESS](#)

