Detecting Fake Reviews With NLP
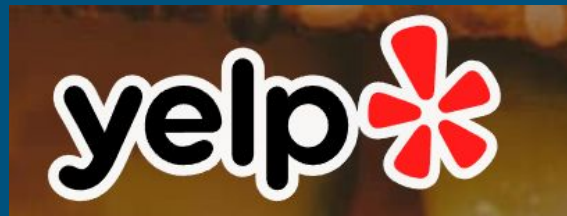
Mitch McElderry
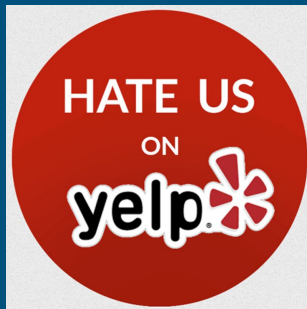
Outline

- Business Problem

- Data and Methods

- Model

- Conclusions

Business Problem

- Develop model to catch fake reviews on restaurants on Yelp

- 5 star reviews can be bought by businesses - this is false advertising

- An extra star for restaurant on reputable review site increases revenue 5% - 9%
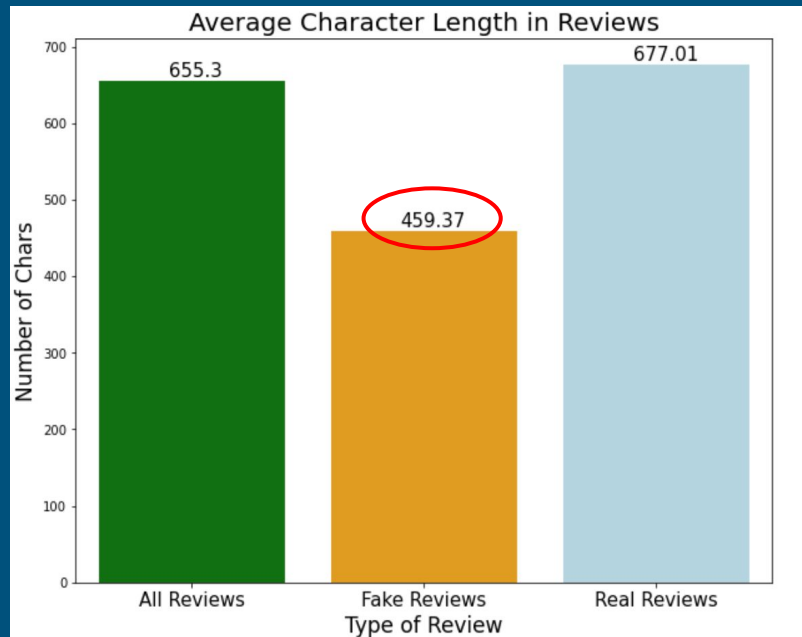
Data and Methods

- Model trained on Yelp reviews from Kaggle 2008-2014
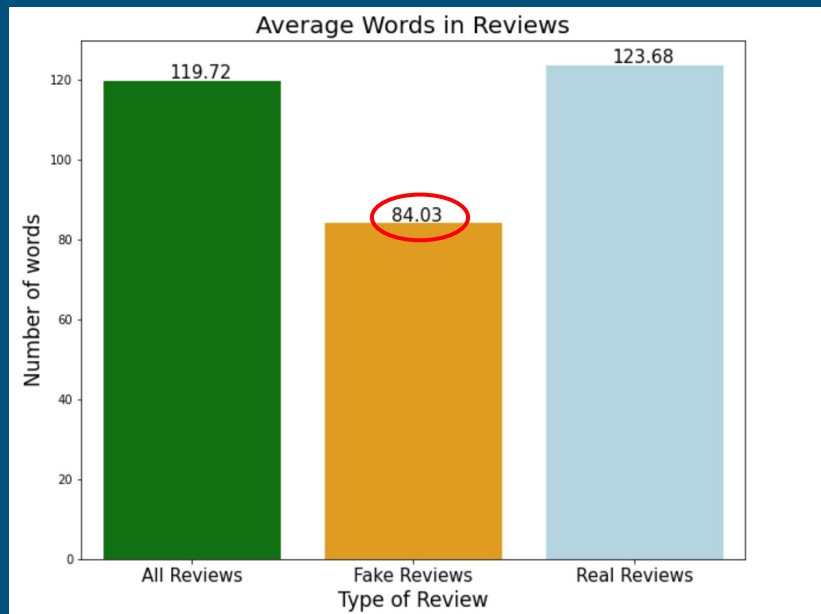
- Present data scraped from Yelp to test model
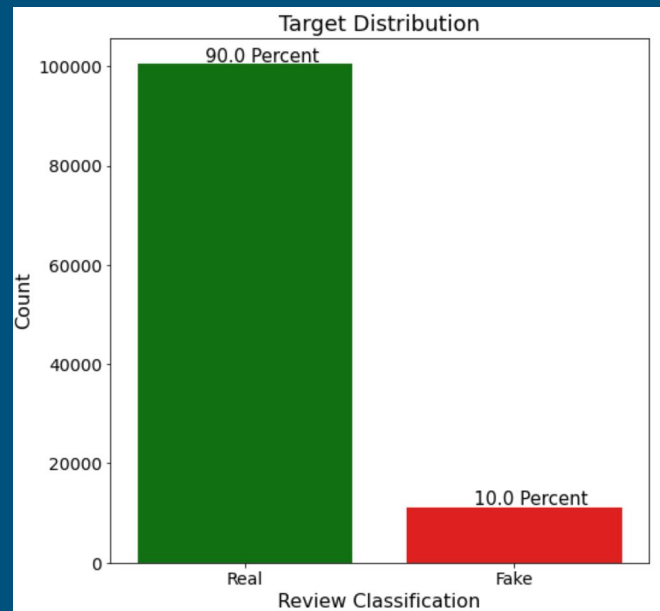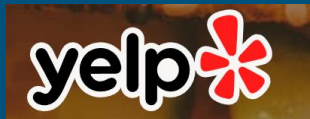
Average Characters in Review

Fake Reviews use less characters

Average Words in Reviews

Fake reviews use less words

Real vs Fake Review

The food here is SLAMMIN"!!! I truly love when you try a restaurant for the first time and the workers/owners are all so happy you came! The service is really a major plus. I felt like I walked into my aunt's house or something. I've tried a few of the items on the menu but the standout by far would be the honey bbq wings. They are so sweet and tangy! I have not had wings that juicy and good in a very long time. The homemade baked good are equally stellar. I hate cheesecake, but the sweet potato cheesecake here changed my mind! The yellow cake with chocolate frosting, the lemon cake all hits! Any of the sides you choose for your meal will be well seasond and tasty! Wash it all down with a hal and half and you will be in heaven. I have not tried the breakfast and to be honest I am scared to do it. Grits, salmon cakes, waffles with nutella...These things could be a major wardrobe shrinker. I am very confident that when I do, it will be worth every calorie. Get yourself over there ASAP!!

Real Review - longer and more detailed

Bangin!!! The honey bbq wings and mac and cheese are to die for, and once I tried it other places in the neighborhood paled in comparison. This family owned restaurant will satisfy any craving for a homey slice of southern comfort. Also, dont miss out on the top notch cakes and pastries--More cheese biscuits please!

Fake Review - shorter with less details

Words Only in Real Reviews
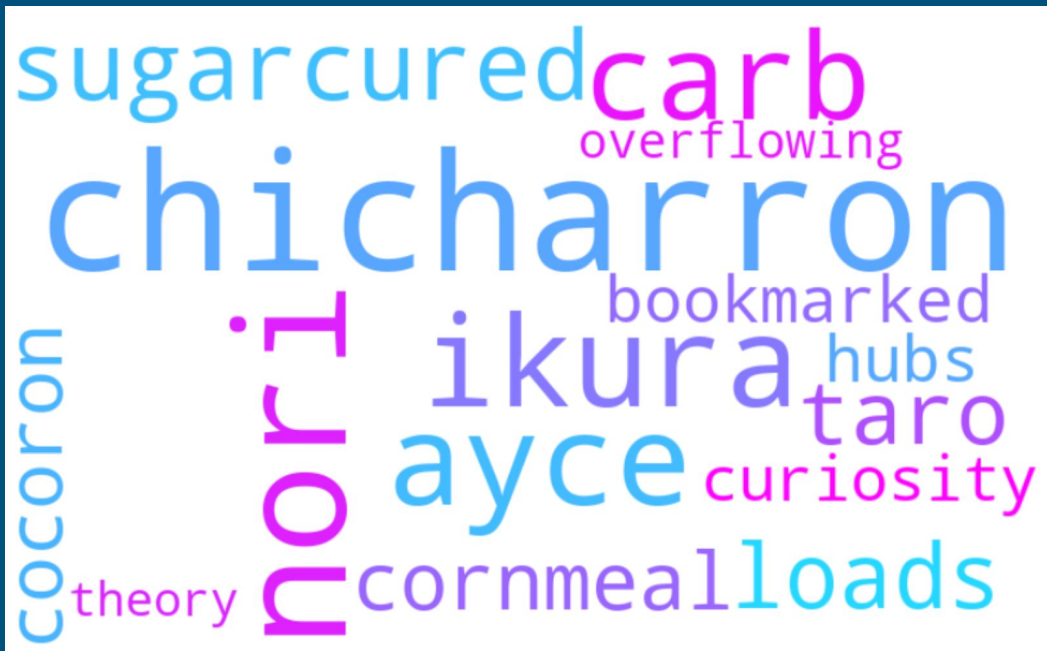
Words Only in Fake Reviews
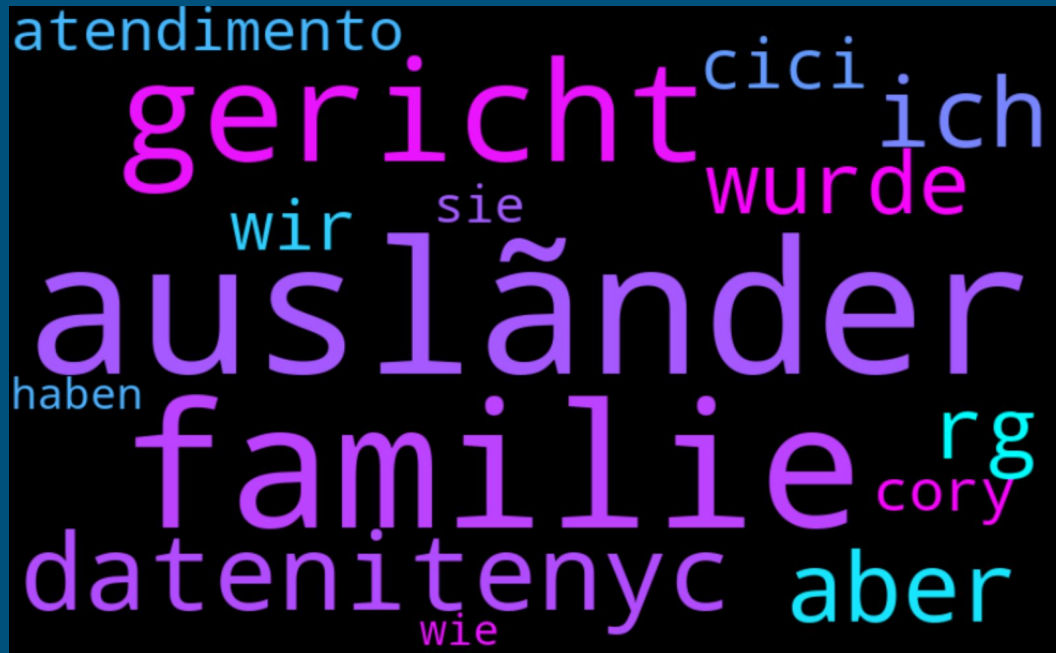
Preparation for Modeling

The food at snack is a selection of popular Greek dishes. The appetizer tray is good as is the Greek salad. We were underwhelmed with the main courses. There are 4-5 tables here so it's sometimes hard to get seated.

Punctuation and numbers removed. Words stemmed.

the food at snack is a select of popular greek dish the appet tray is good as is the greek salad we were underwhelm with the main cours there are tabl here so it sometim hard to get seat

Numerical stats for each review (Before editing)

Number of Words	Character Review Length	Sentence Count	Avg Words Per Sentence
40	215	4	10.000000

Model

TFIDFVectorizer with Logistic Regression

Recall: 70%

- False positives still problem
- Caught over 1900 fake reviews out of ~ 2700

Model on Local Reviews

~ 1800 Reviews scraped from local restaurants on Yelp

Conclusion

- Data such as user location and activity will help model

- Fake reviews shorter on average

- Machine learning and NLP help minimize fake reviews
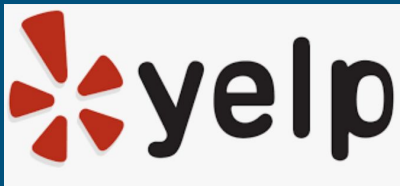
Thank you!

Mitch McElderry

Github: https://github.com/mmceld2

Linked in: https://www.linkedin.com/in/mitch-mcelderry-35a47a99/