



The Onion vs Real News: Detecting Satire

Emiko Sano



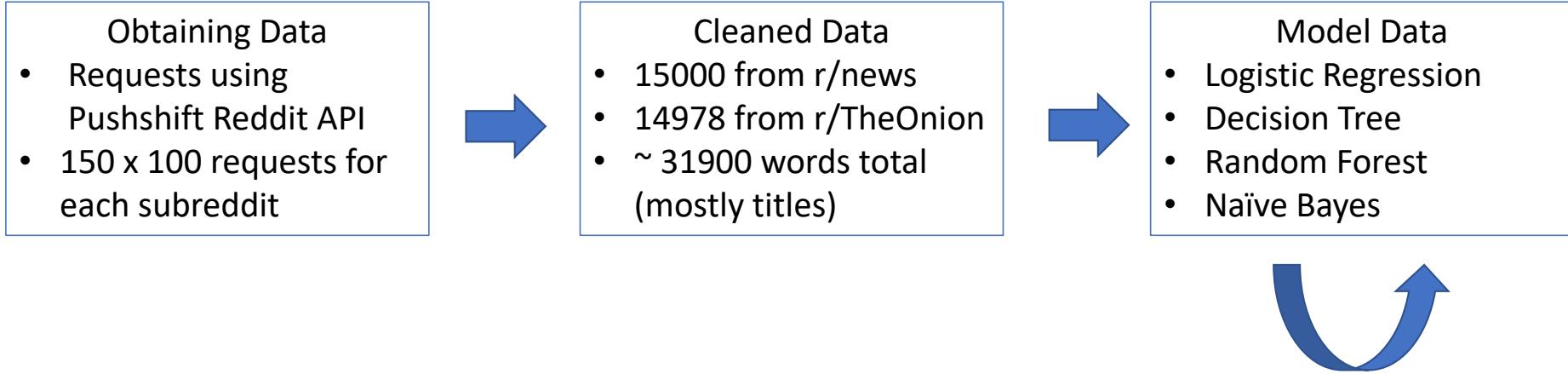
What Is The Onion?



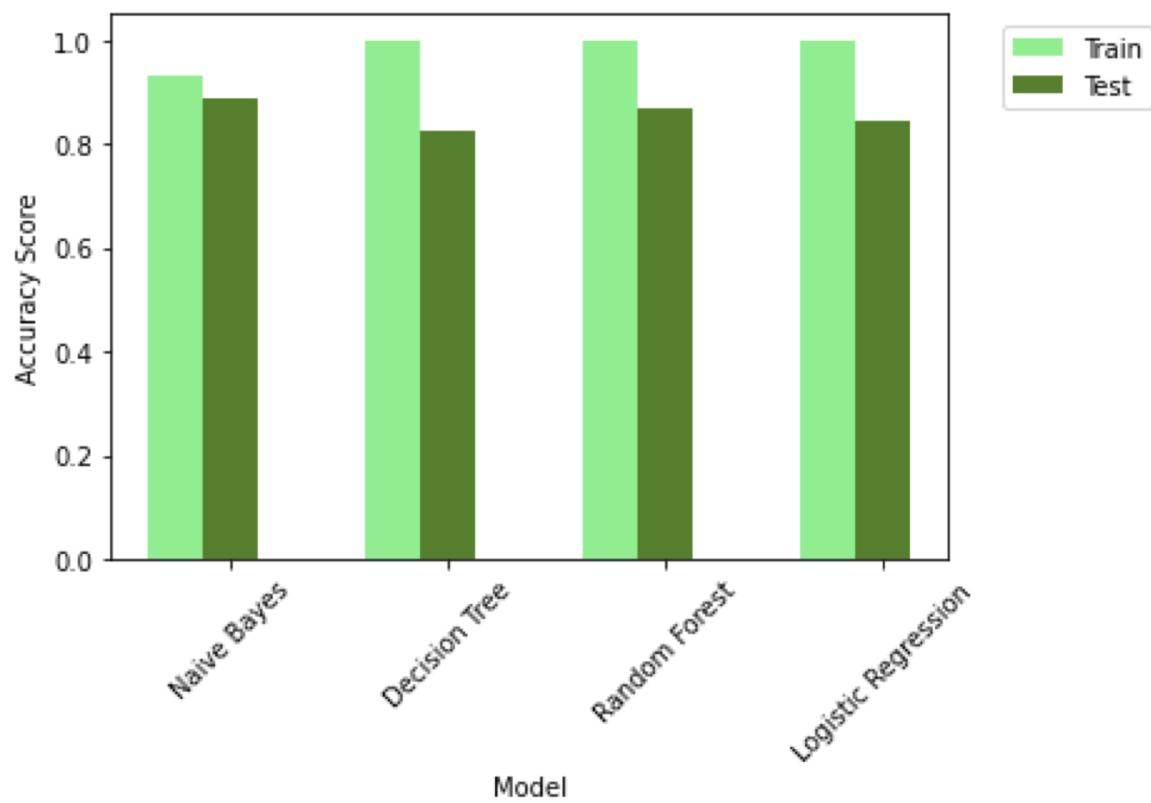
The Questions

- Can a model differentiate the text from The Onion and news subreddits?
- Can the model predict the origin of the text with high accuracy?
- What are the factors that contribute to the model?

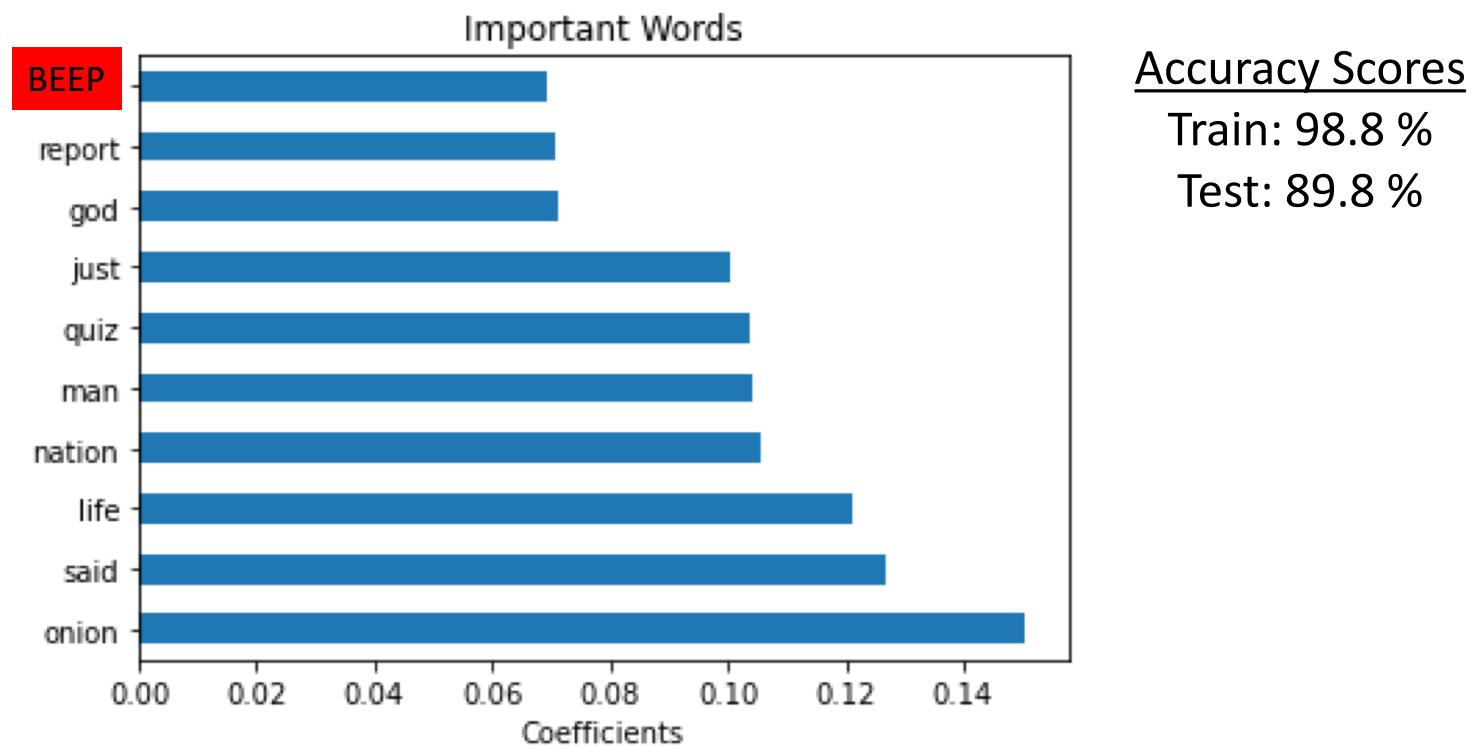
Modeling Process



Accuracy Scores with Default Settings

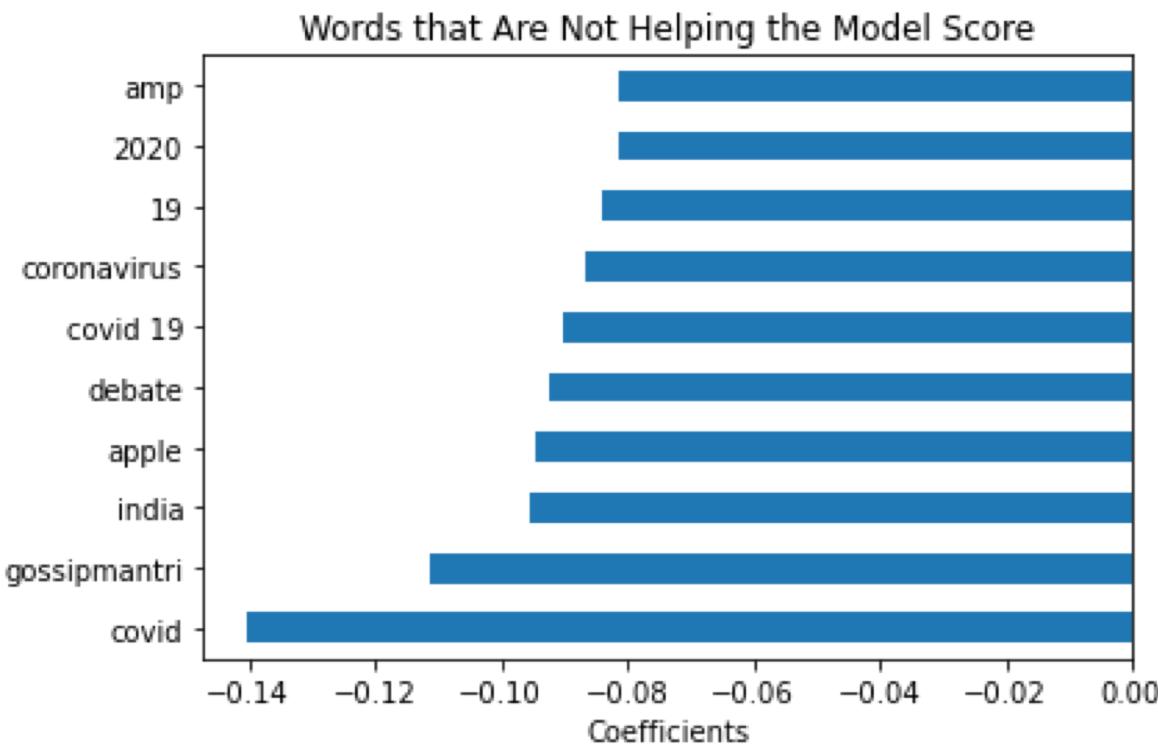


Logistic Regression Model: Words That Make a Difference to the Model



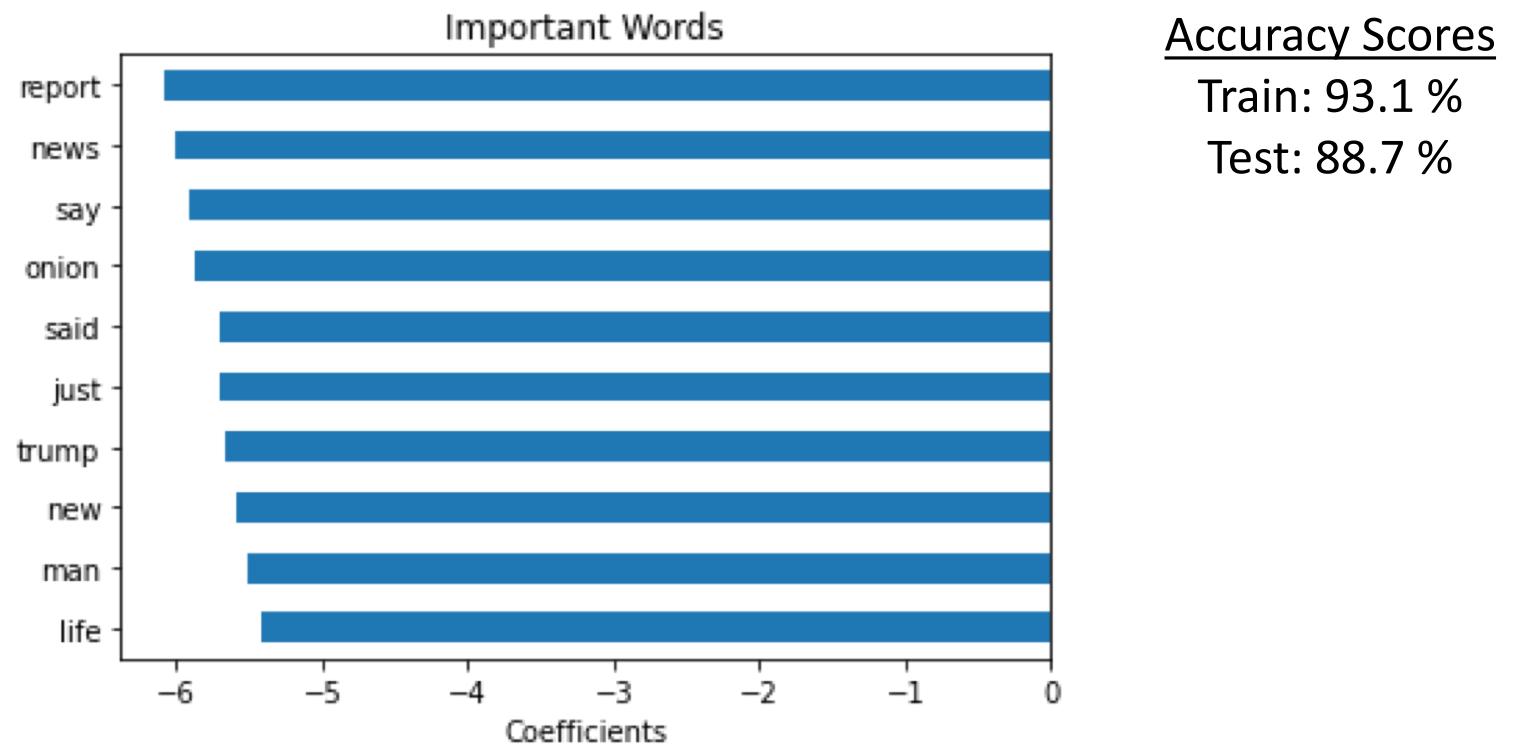
Logistic Regression Model

Words That Don't Make a Difference to the Model



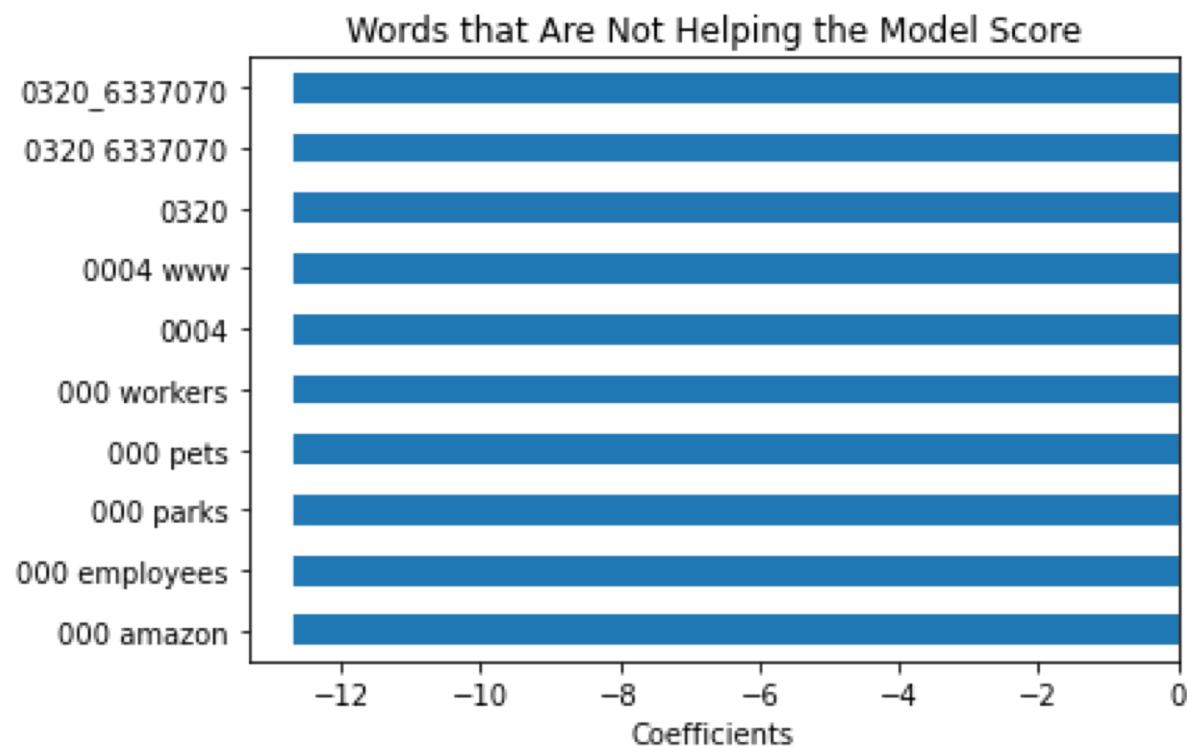
Accuracy Scores
Train: 98.8 %
Test: 89.8 %

Naive Bayes Model: Words That Make a Difference to the Model



Naive Bayes Model

Words That Don't Make a Difference to the Model



Accuracy Scores
Train: 93.1 %
Test: 88.7 %

Summary and Conclusions

- Can a model differentiate the text from The Onion and news subreddits? ✓
- Can the model predict the origin of the text with high accuracy?
 - reached 89 % accuracy ✓
- There is room for improvement:
 - Using pre-trained word sets
 - Neural nets

Questions?