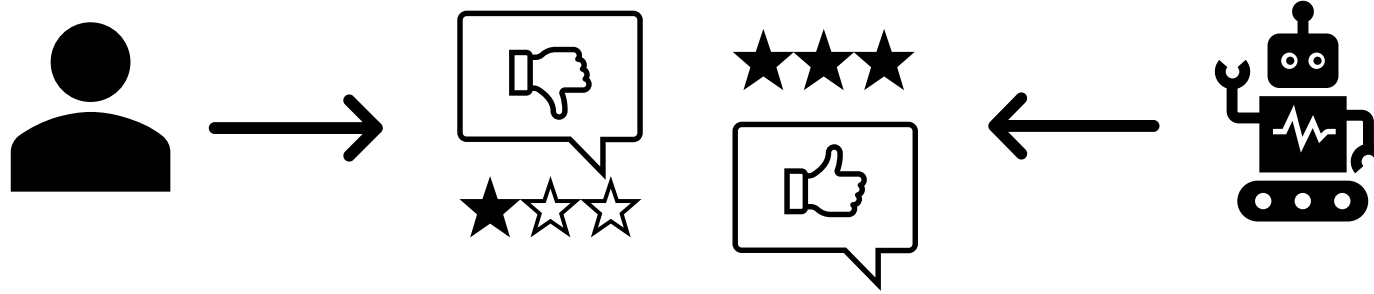


AI Academy

Capstone Presentation

January 2023

Can you really trust what you see online?



Motivation:

Online reviews are becoming less trustworthy as fake reviews written by bots are becoming more prevalent.

This can misguide consumers and influence purchasing decisions to their detriment [\(1\)](#) [\(2\)](#) . Less trust in online shopping then hurts businesses.

Objective:

Create a Natural Language Processing model to detect reviews generated by a computer and not by real consumers.

NLP Modeling Approach

Data Acquisition

- Fake Amazon Review Dataset
 - 40,000 reviews collected from OSF.io

Data Prep and EDA

- Transform text of reviews into clean and manageable inputs for model
 - Remove stop words, tokenize, lemmatization

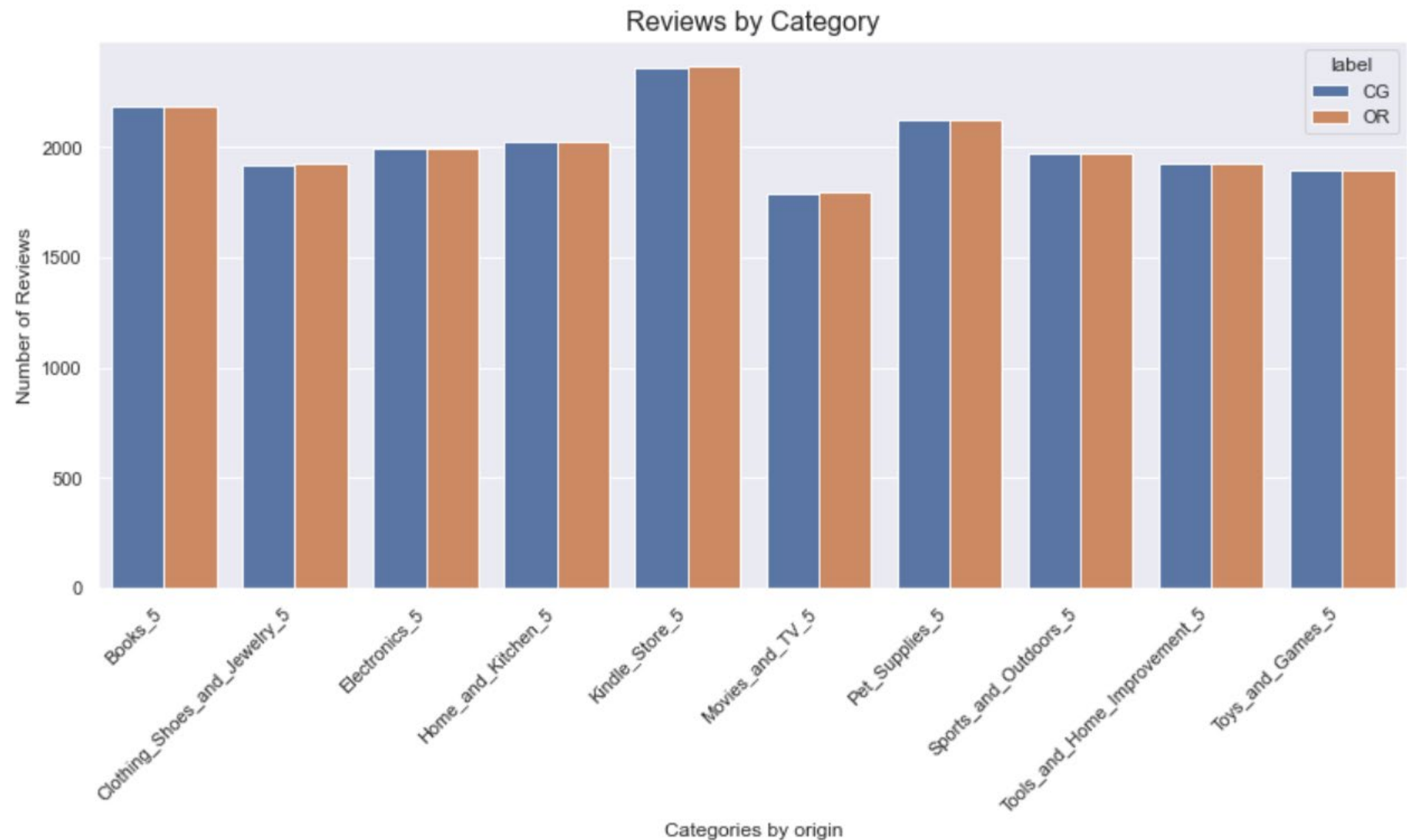
Baseline Model

- The baseline model will be a logistic regression
 - Some features will include TF-IDF count vector, word count, and sentiment analysis

Use More Advance Model and fine tune

- For the more advanced model, a random forest model will be used
 - This will have increased accuracy over the baseline

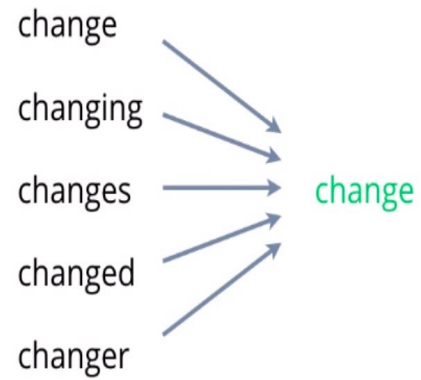
Analyzing Review Dataset



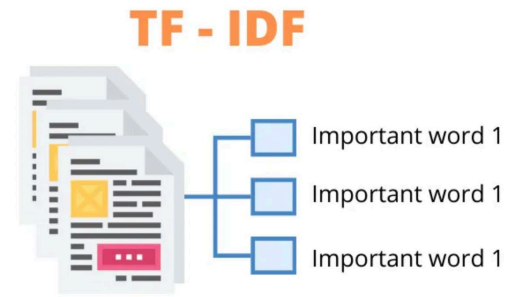
Cleaning and Prepping Data



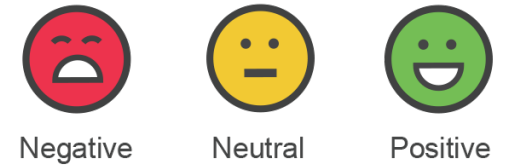
Stop Word Removal



Lemmatization

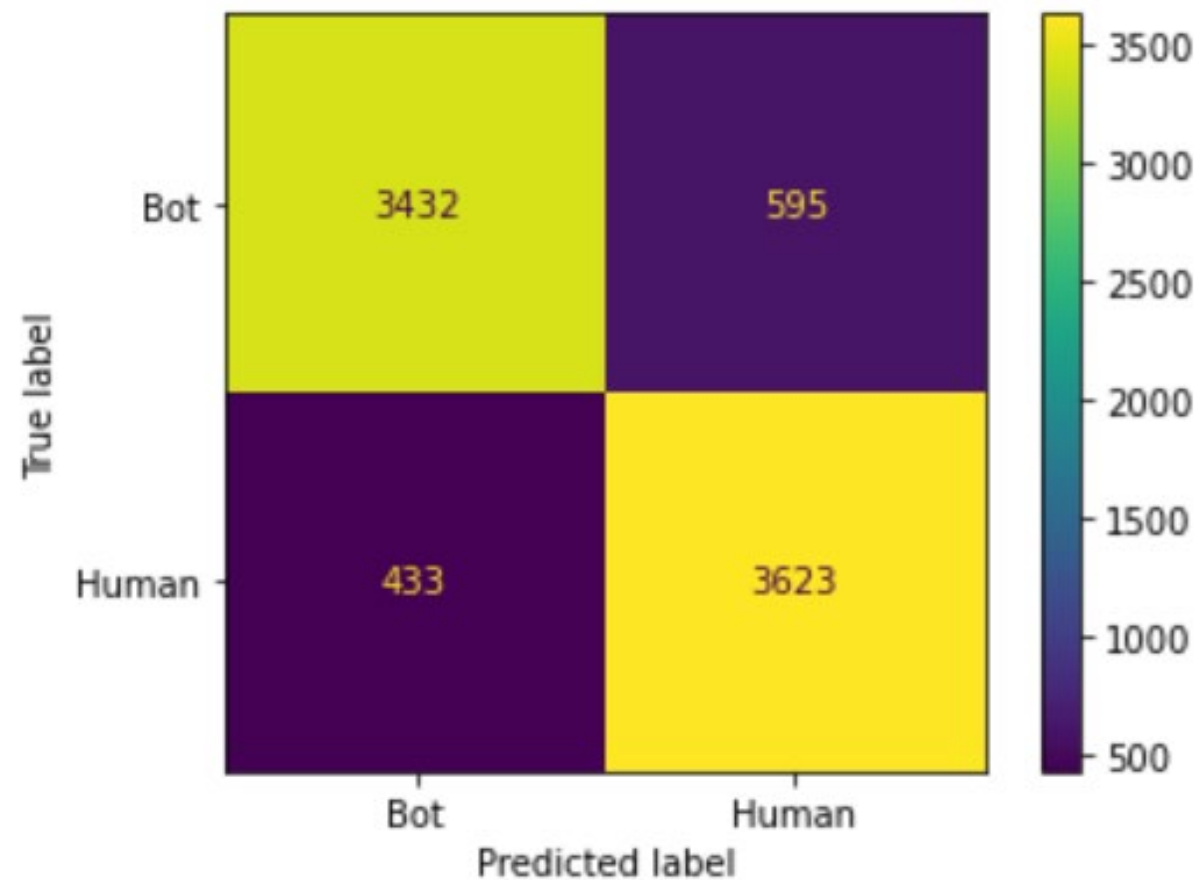


TF-IDF Vectorization



Sentiment Analysis

Baseline Model using logistic regression

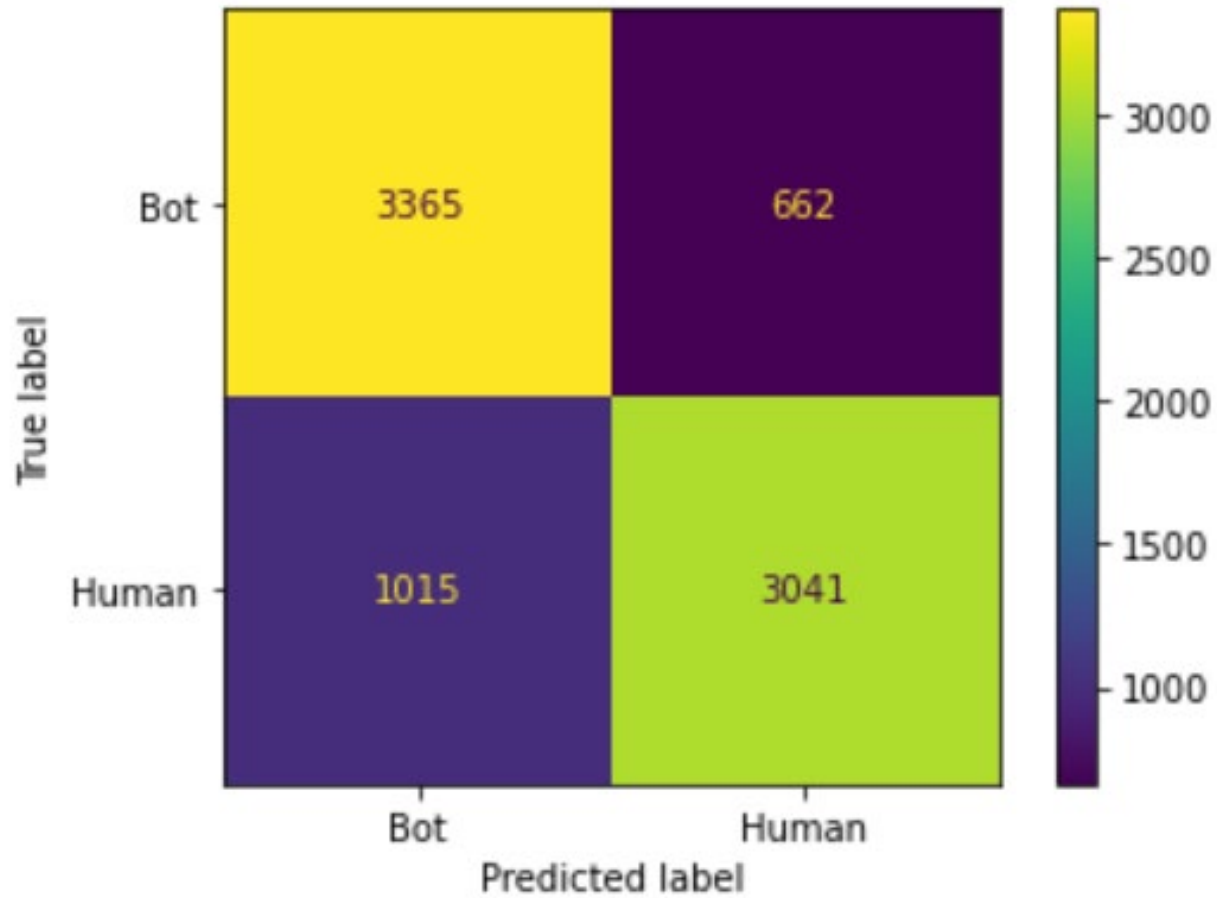


Precision: 0.88

Recall: 0.85

F1 score: 0.87

Improved Model using Random Forest model



Precision: 0.77

Recall: 0.83

F1 score: 0.80

Questions?