



Movie Review | Text Classification

Miguel Angel Santana II, MBA

February 2, 2021

Introduction



Methodology



OSEMN Framework

Obtain | Scrub
Exploratory Data Analysis
Statistical Modeling
• Validation
Interpret Results



Limitations & Future Work



Thank You

Methodology



PROCESS TEXT DATA – 6000 MOVIE REVIEWS

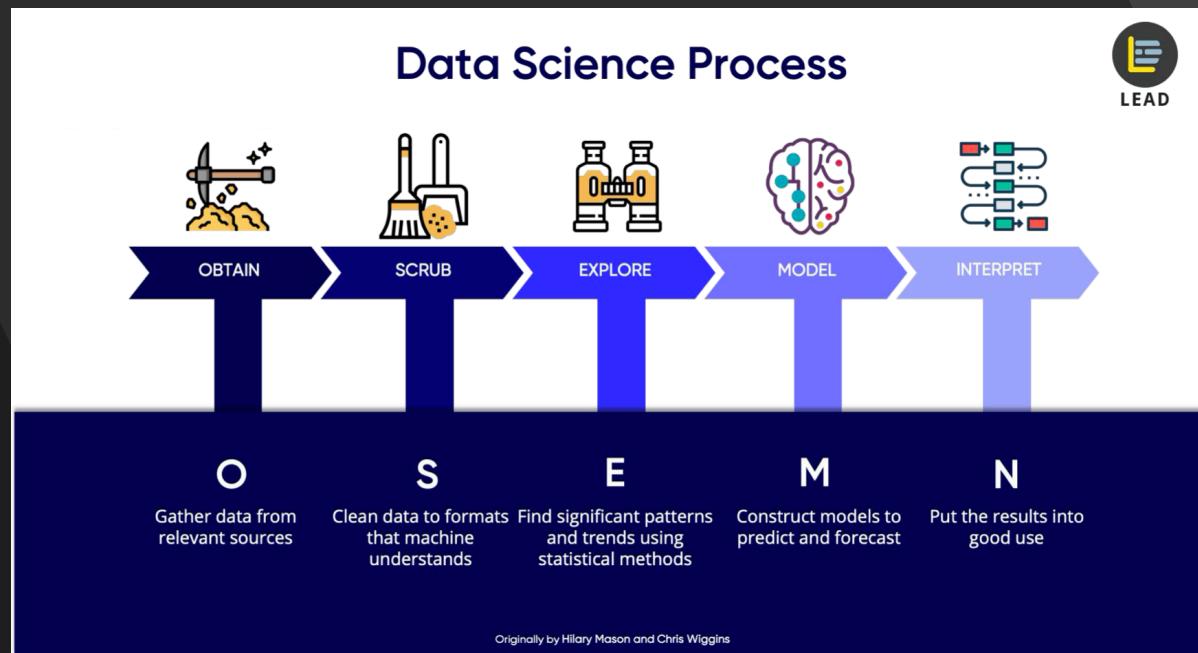


CLASSIFY THE REVIEWS AS POSITIVE OR NEGATIVE



LEVERAGE THE OSEMN FRAMEWORK TO PROCESS, MODEL AND GAIN INSIGHT INTO TEXT CLASSIFICATION.

OSEMN Framework



Obtain | Scrub

Stanford Artificial Intelligence Laboratory (SAIL)

- <http://ai.stanford.edu/~amaas/data/sentiment/>

Udemy Course: NLP – Natural Language Processing

- by Jose Portilla.

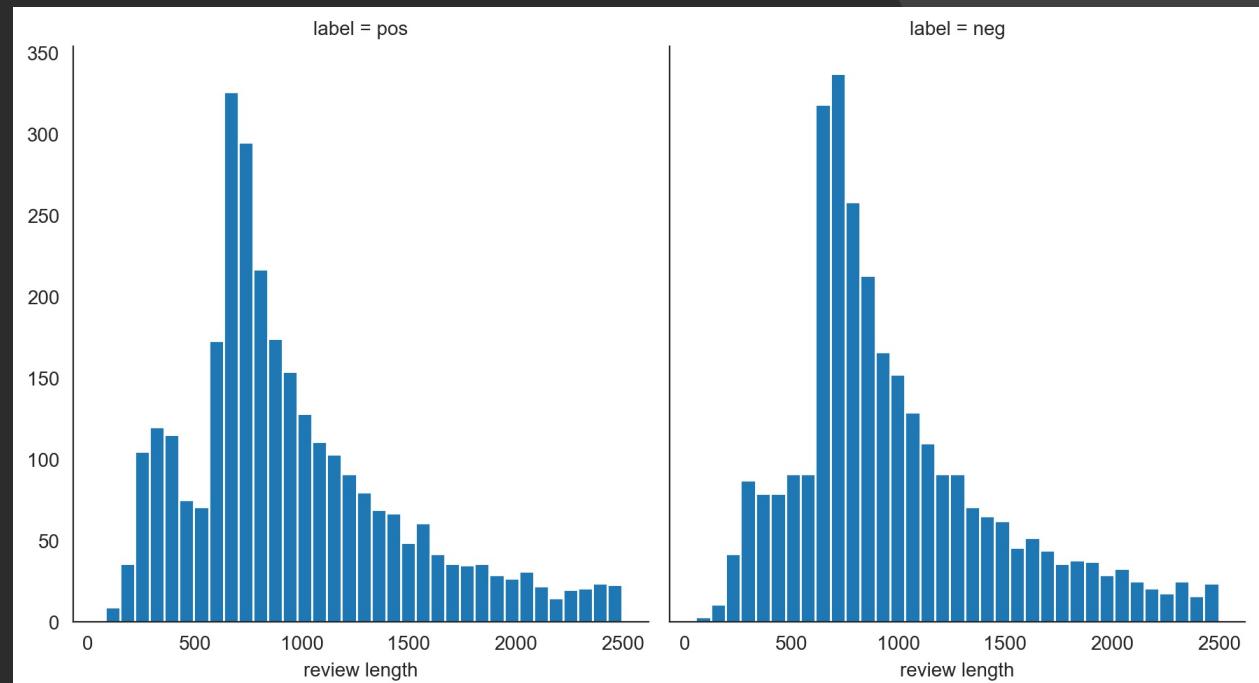
Scrubbing | Data Cleaning

- Key Decisions
 - search/address whitespace strings
 - drop null values

Exploratory Data Analysis

Length of Reviews

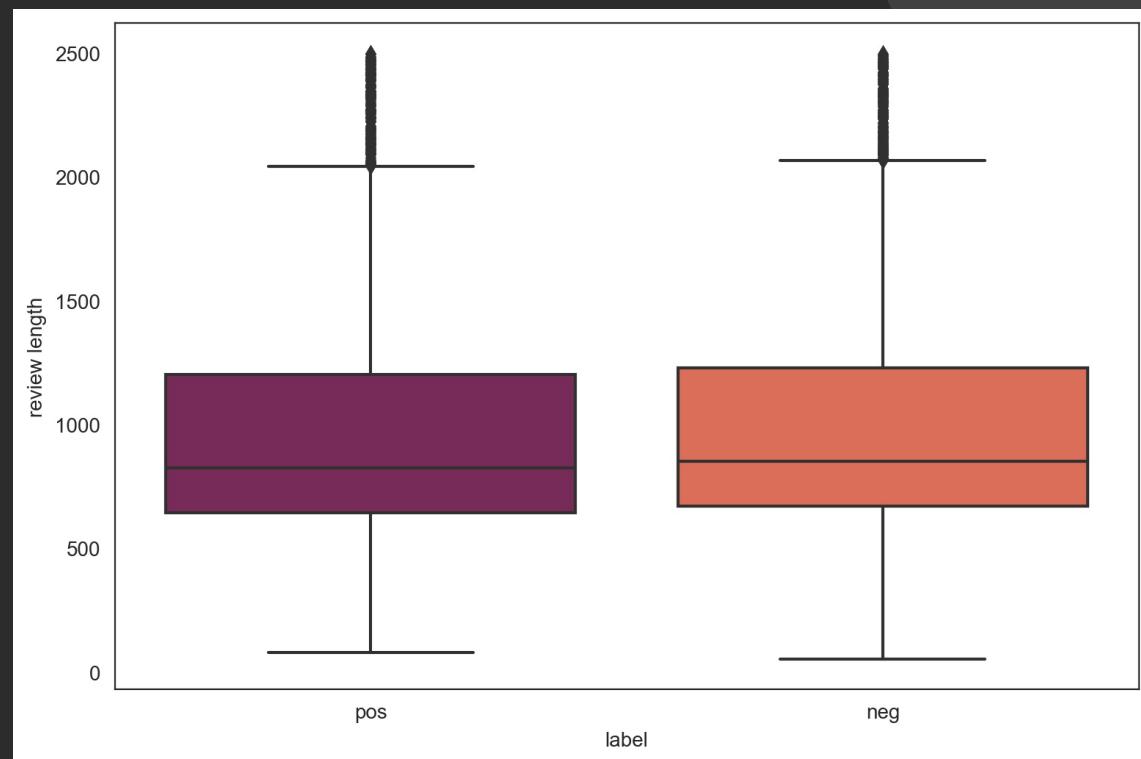
- Positive and negative reviews appear to have the same general length in text.



Exploratory Data Analysis Continued

Outlier Data

- There are no significant outliers in positive or negative text lengths.

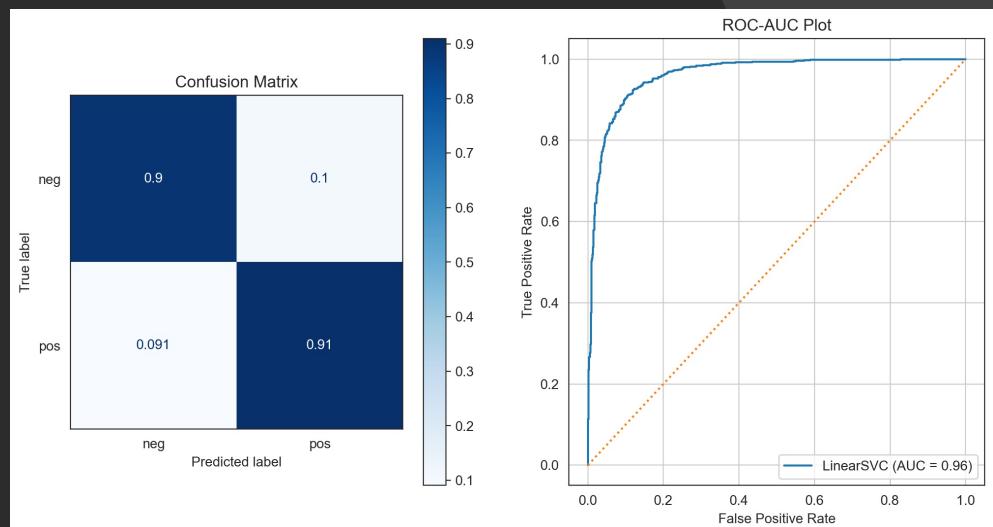


Statistical Modeling

- Machine Learning Algorithms
 - Multinomial NB Classifier
 - LinearSVC Classifier
 - LinearSVC using TF-IDF
- Accuracy was used as a validation metric.

Top Model: Linear Support Vector Classifier

- Accuracy Score 92%



Interpret Results | Classifier Use

Accuracy

- Accuracy was used as a validation metric due to the class balance between positive and negative reviews.

Potential use cases

- Web scraping movie review sites and gaining insight on consumer trends.
- Combining the classifier with topic and sentiment analysis for business.

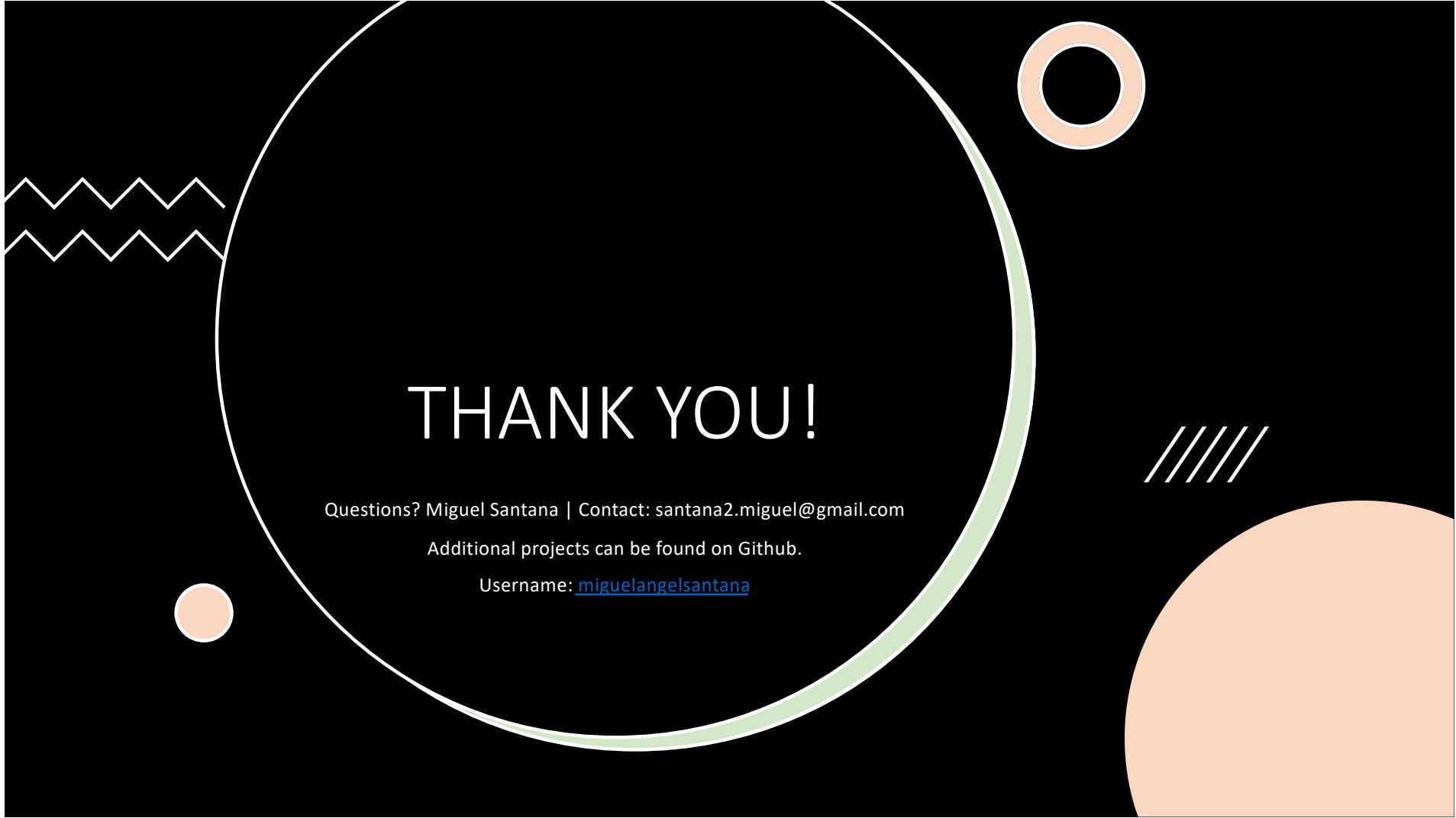
Limitations & Future Work

Limitations

- The classifier is for illustrative purposes due to data preprocessing
 - Pre-Labeled Data

Future Work

- Web scraping of additional sources to get a broader view of the classifier's potential with respect to larger datasets and multiple sources.



THANK YOU!

Questions? Miguel Santana | Contact: santana2.miguel@gmail.com

Additional projects can be found on Github.

Username: [miguelangelsantana](#)