



vs.



Reddit User Language

Aerika Song

Background & Objective

- Italy
 - Ride-sharing banned
 - Recent petitions call for the launch of Uber and/or Lyft
- Objective
 - Classification Model:
 - Determine public view on both Uber and Lyft
 - Insight on which would be better to launch in Italy



Overview

- Web Scraping
- Modeling
- Uber Subreddit user language
- Lyft Subreddit user language
- Common language between the two
- Improving classification model



Acquiring the Data: Web Scraping

- Using Reddit's PRAW API Wrapper
- For each subreddit:
 - Title
 - Number of Comments
 - Days elapsed since posted
 - Name of Subreddit
 - Body Text
 - Parent Comment Text
- Save as .csv file

(Pipeline & Grid Search) Modeling

- CountVectorizer (“cvec”)
- Term Frequency-Inverse Document Frequency Vectorizer (“tfidf”)
- Multinomial Naive Bayes (“MultiNB”)
- Logistic Regression (“LogReg”)

CountVectorizer
Logistic
Regression ▲

TFIDFVectorizer
Logistic
Regression ▲

CountVectorizer
Multinomial NB

TFIDFVectorizer
Multinomial NB

Cvec Parameters

Max_featuers, stop_words,
max_df, min_df,
ngram_range

Tfidf Parameters

Max_featuers, stop_words,
max_df, min_df,
ngram_range

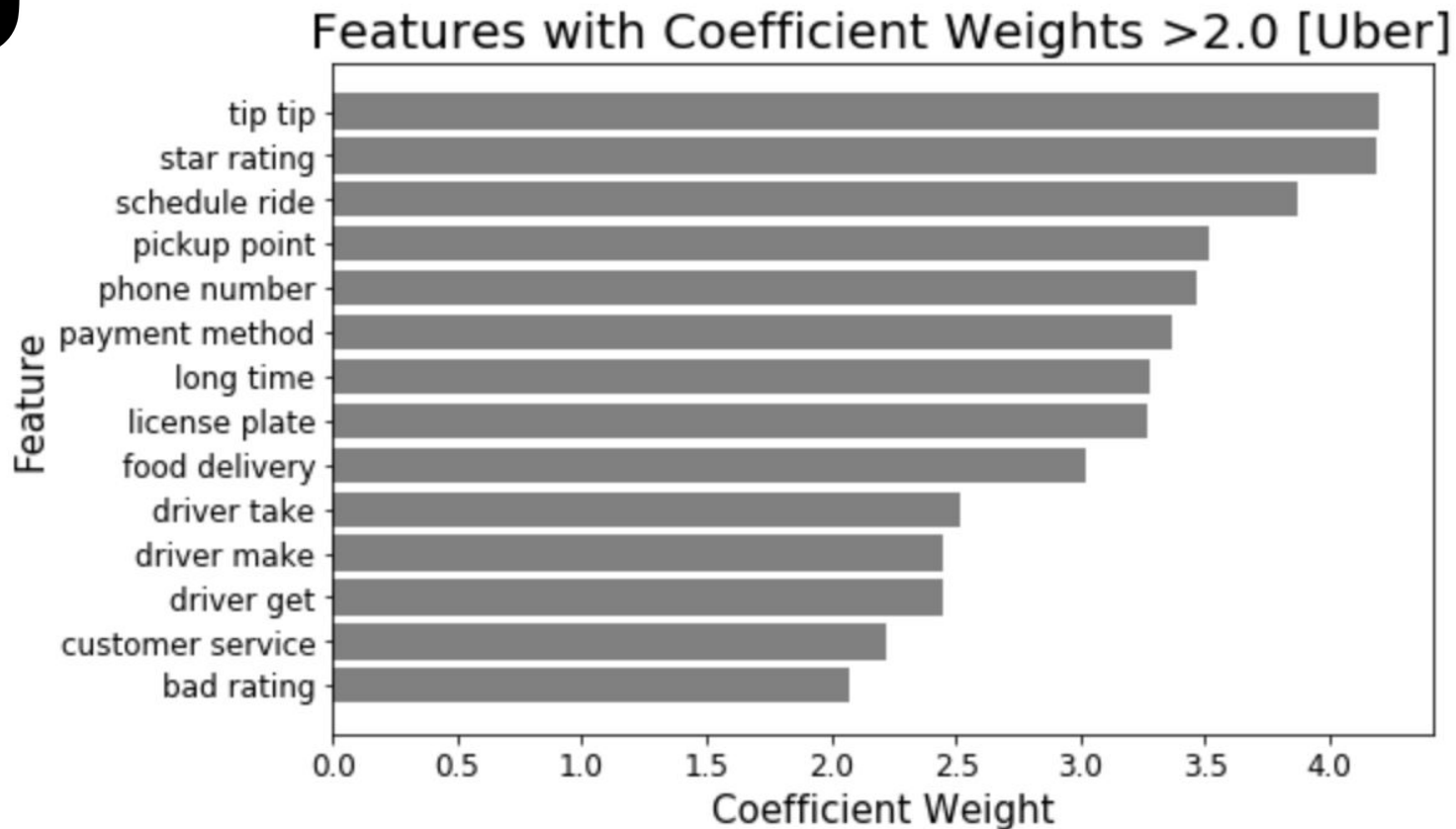
LogReg Parameters

C, Penalty



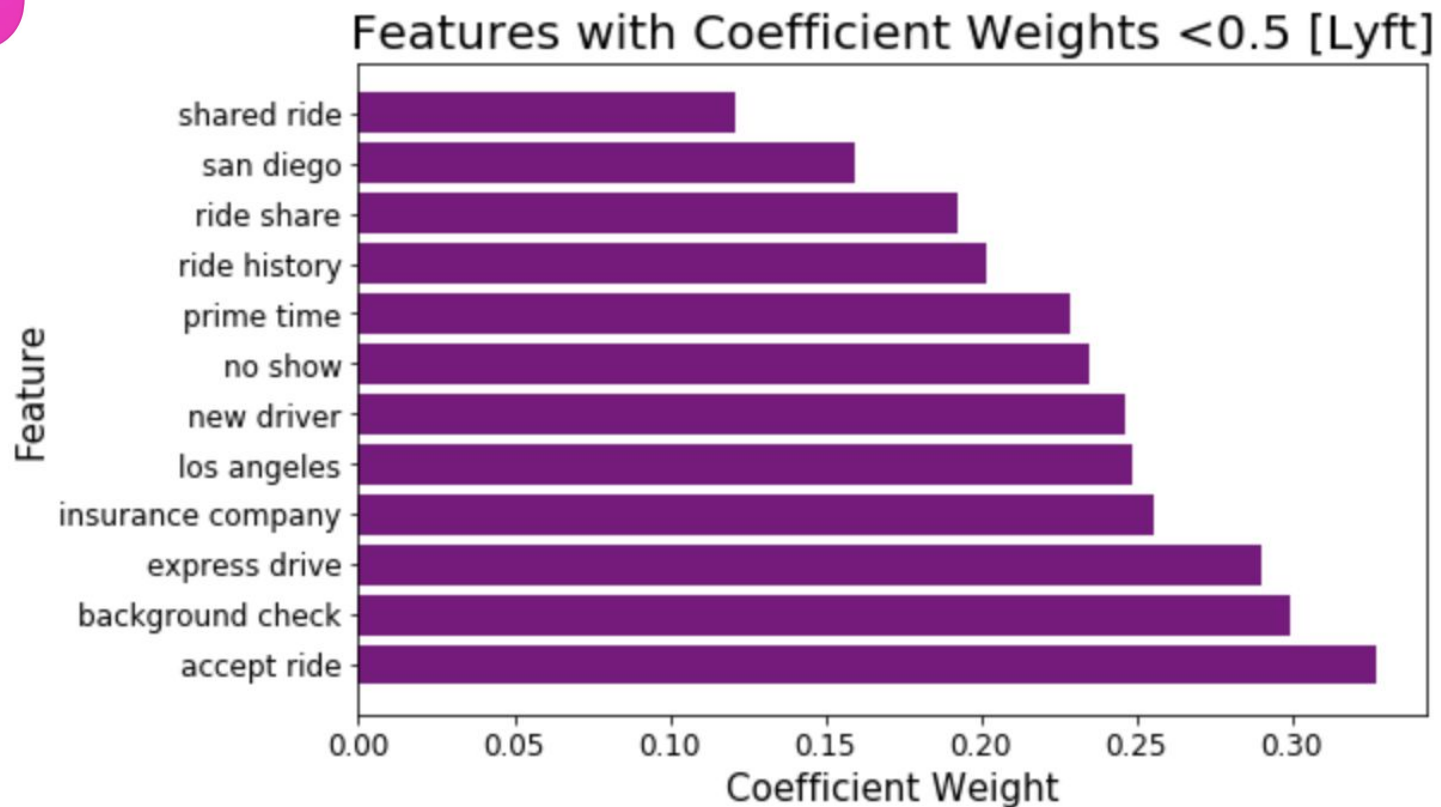


Subreddit User Language





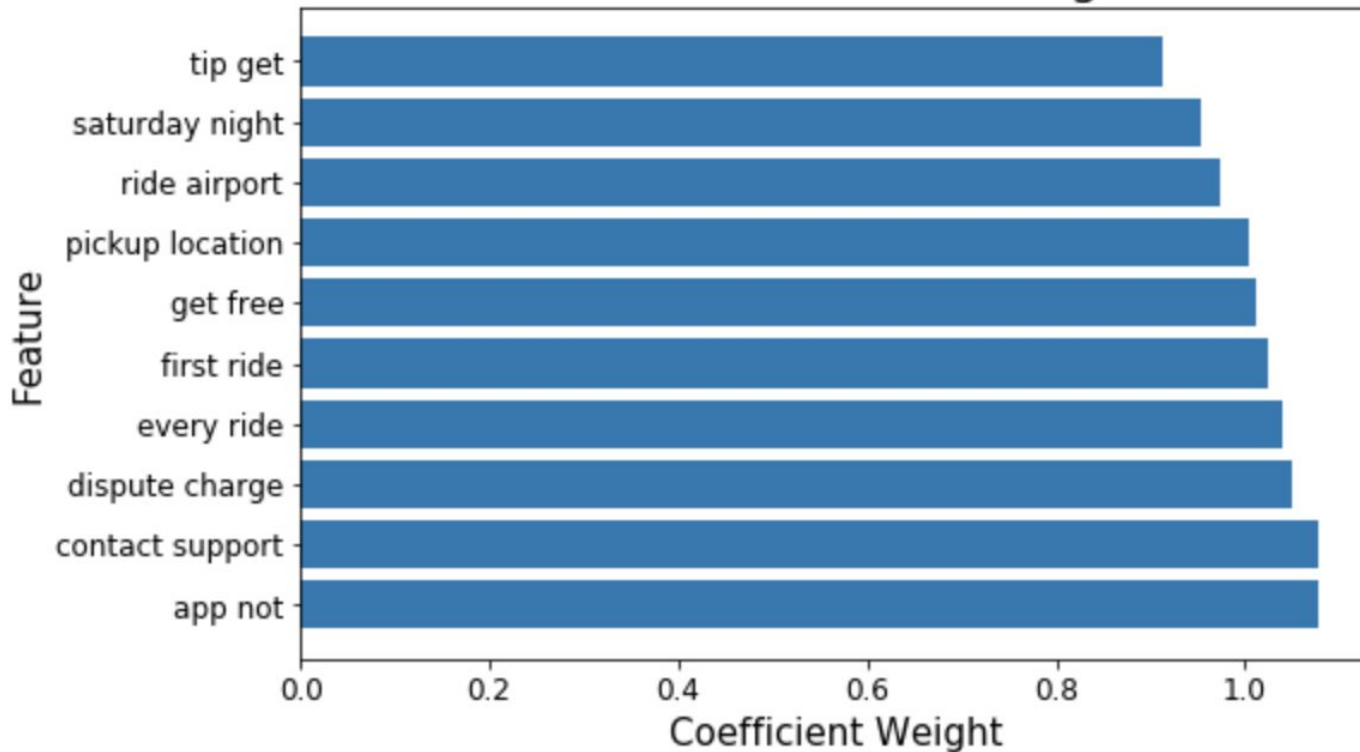
Subreddit User Language





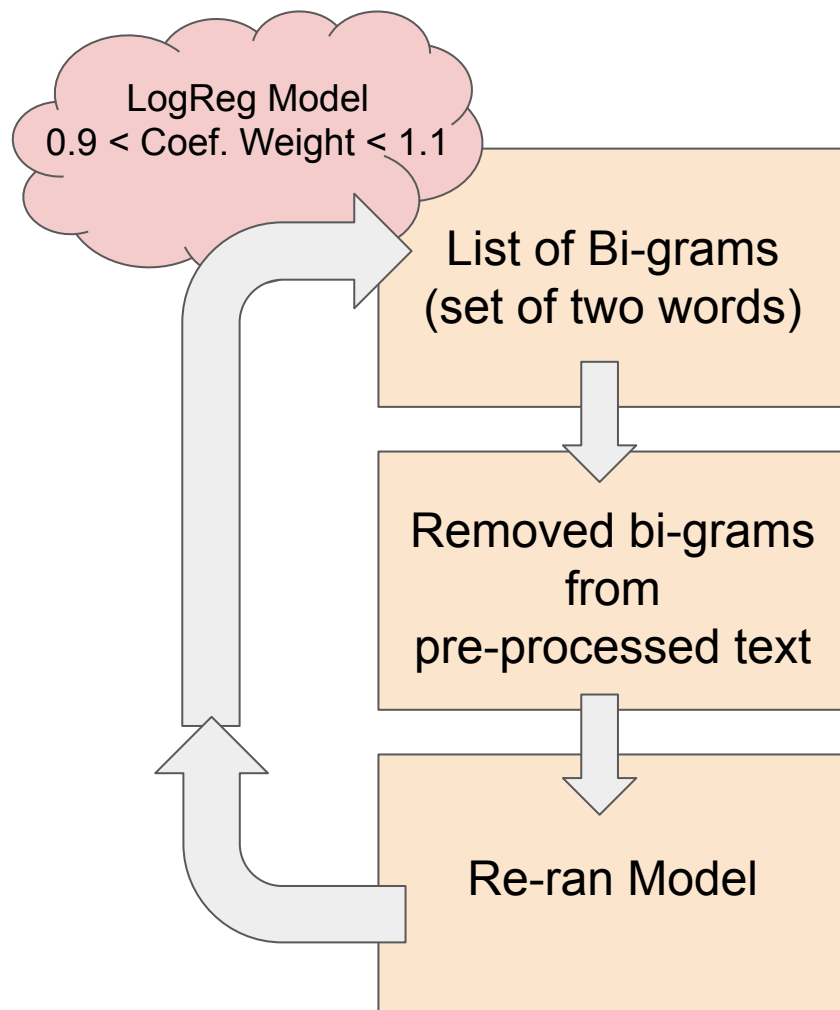
Subreddits Common Language

Features with Coefficient Weights ~ 1.0



Improving Classification Accuracy

- Ran stopwords identifying process multiple times
 - Built upon original list of bi-gram stopwords
 - Repeat until low count of stopwords being added
- Best Model's Accuracy Scores:
 - Before Bi-grams stopwords:
 - CV = 5: 0.64
 - Test = 0.66
 - After Bi-grams stopwords:
 - CV = 5: 0.63
 - Test = 0.63



Conclusions

- **WITHOUT** using obvious word indicators (“uber”, “lyft”)
 - Difficult to classify between Uber/Lyft Subreddits
 - In other words...people are talking about Uber & Lyft similarly between the subreddits
- Reddit Users' views similar between Uber/Lyft
 - Both have their own pro's and con's
 - <https://www.ridester.com/uber-vs-lyft/>
 - Can launch either/both in Italy

Sources

<https://www.oyster.com/articles/where-is-uber-banned-around-the-world/>

https://www.tripadvisor.com/ShowTopic-g187768-i20-k10342140-Uber_or_Lyft-Italy.html

<https://www.ridester.com/uber-vs-lyft/>

https://www.reddit.com/r/cutouts/comments/6q5a6l/can_someone_give_me_a_round_cutout_of_the_reddit/

<https://apps.apple.com/us/app/uber/id368677368>

<https://apps.apple.com/us/app/lyft/id529379082>