WEB APIs & NLP

PROJECT 3: GENERAL ASSEMBLY DSIR 0124

COLE POPPSINGER:

passionate
about beverages
for over
40 years

TABLE OF CONTENTS

01

DATA SCIENCE PROBLEM:

Using NLP, can customer insight be gleaned from user activity on reddit?

Will the insight be useful in targeted marketing?

DEVISE A STRATEGY

Scrape user posts from r/espresso and r/CraftBeer subreddits

Train a model to identify similarities and differences

Apply findings to recommendations and next steps

reddit users:
430,000,000 monthly
52,000,000 daily

The potential impact of
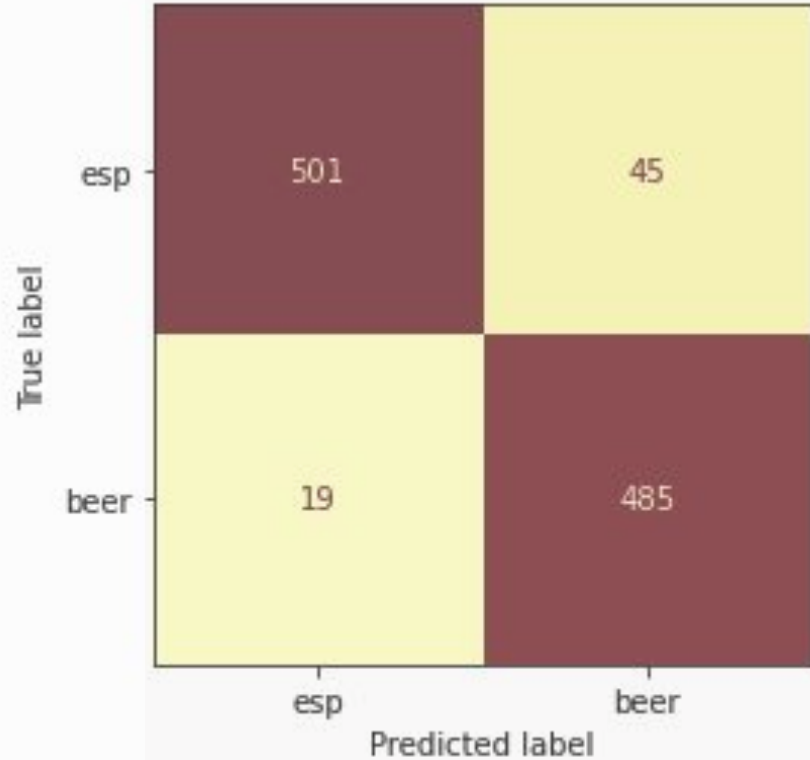insight and advertising
is significant

02

MODELING + ANALYSIS

MODELS+NLP

- CountVectorizer + Logistic Regression Pipeline
- TfidfVectorizer + DecisionTreeClassifier Pipeline
- RandomForest
- FeatureUnion
- Lemm + Stem
- CountVectorized
- custom stop words

BEST:
- TfidfVectorizer + Logistic Regression Pipeline
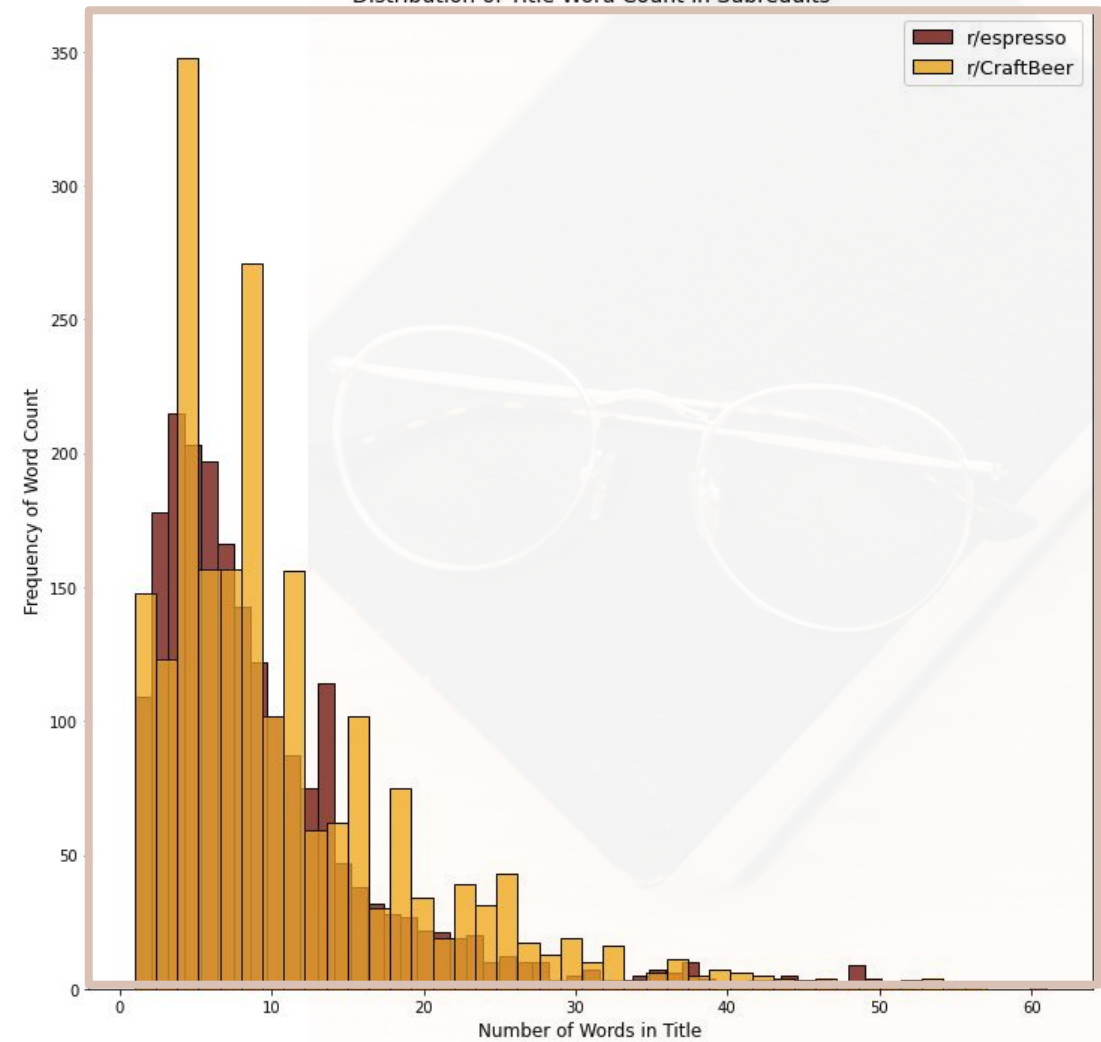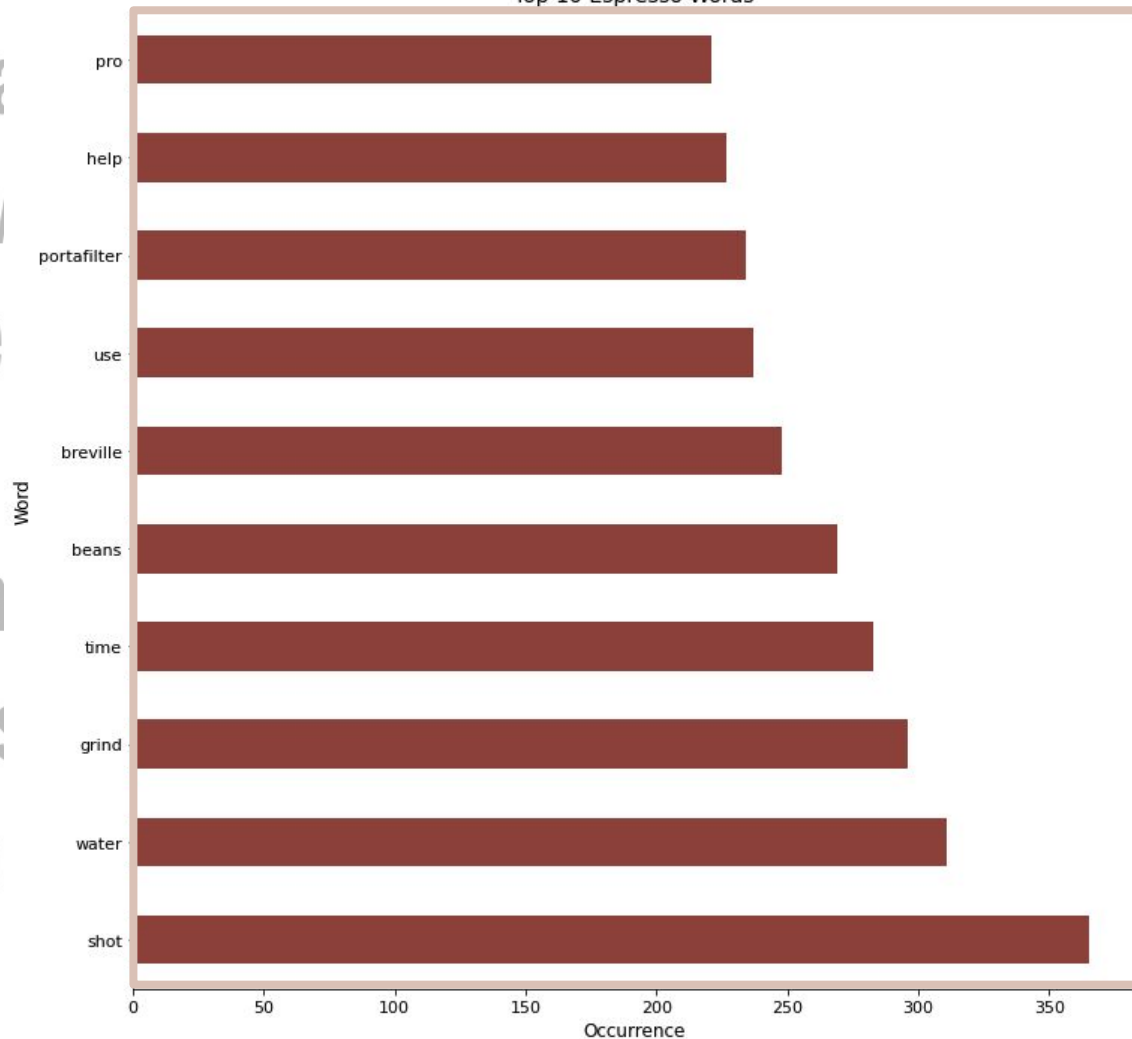
CONFUSION MATRIX:

BEST MODEL

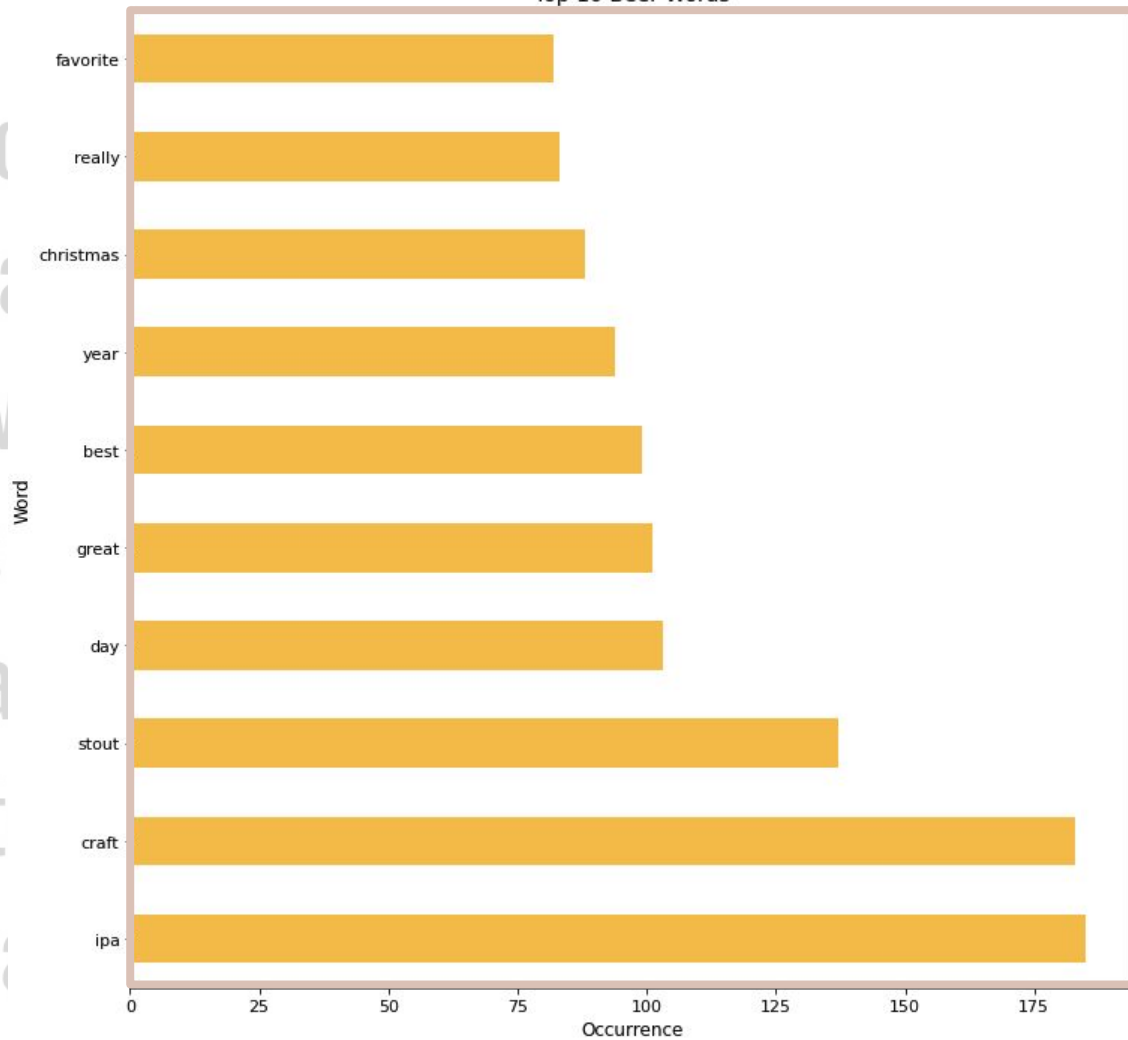TfidfVectorizer + LogisticRegressor

94% accuracy score

Distribution of Title Word Count in Subreddits

Top 10 Espresso Words

(partial background text, obscured)
...yone have a... ...e used for
...ppuccinos w... ...ss? I keep t...
...ans but the... ...too dark in...
...vor. What I... ...rk enough t...
...vorful in ca... ...e like a Ital...
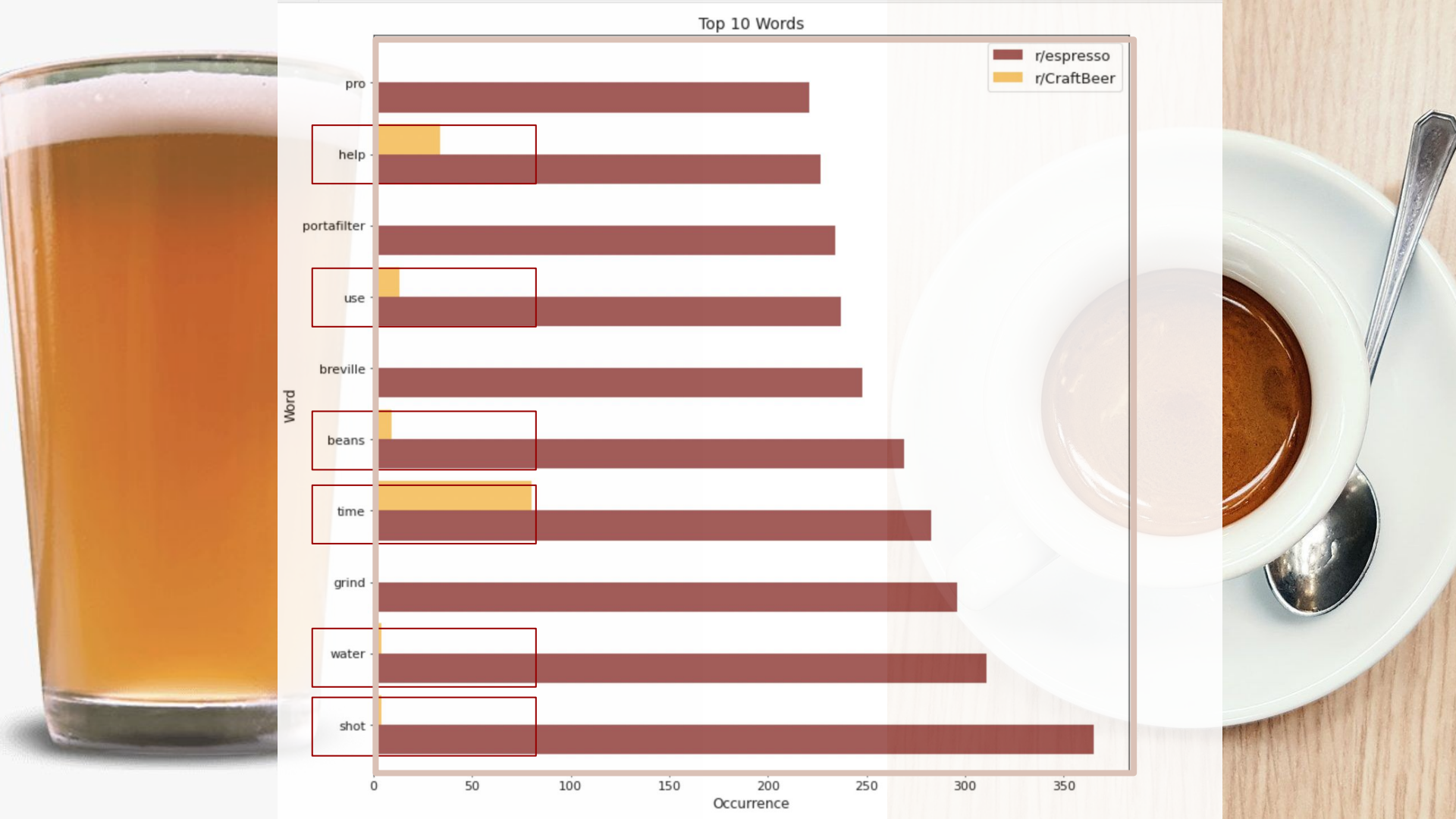...French roas... ...Class is the...
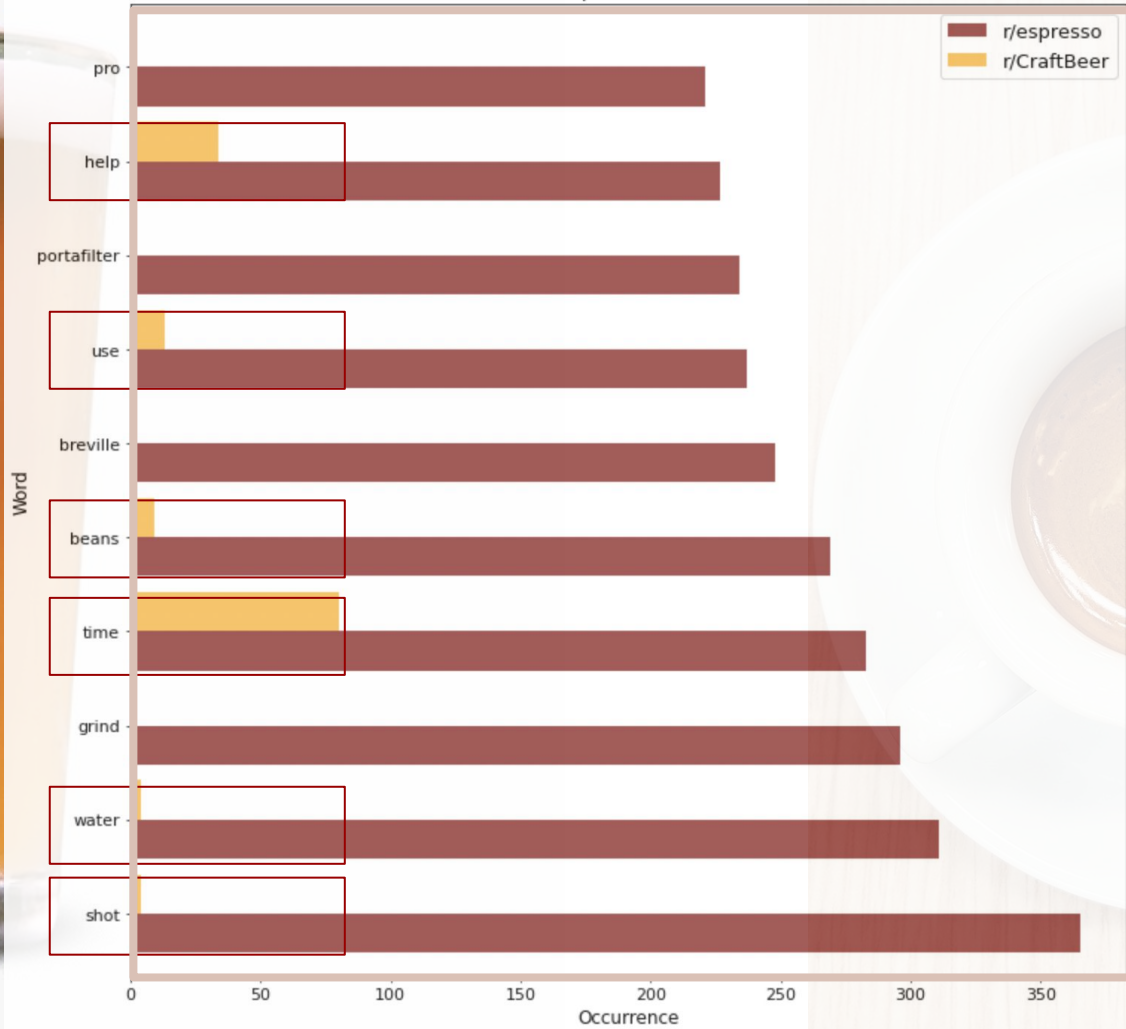...ans tend to... ...e order.
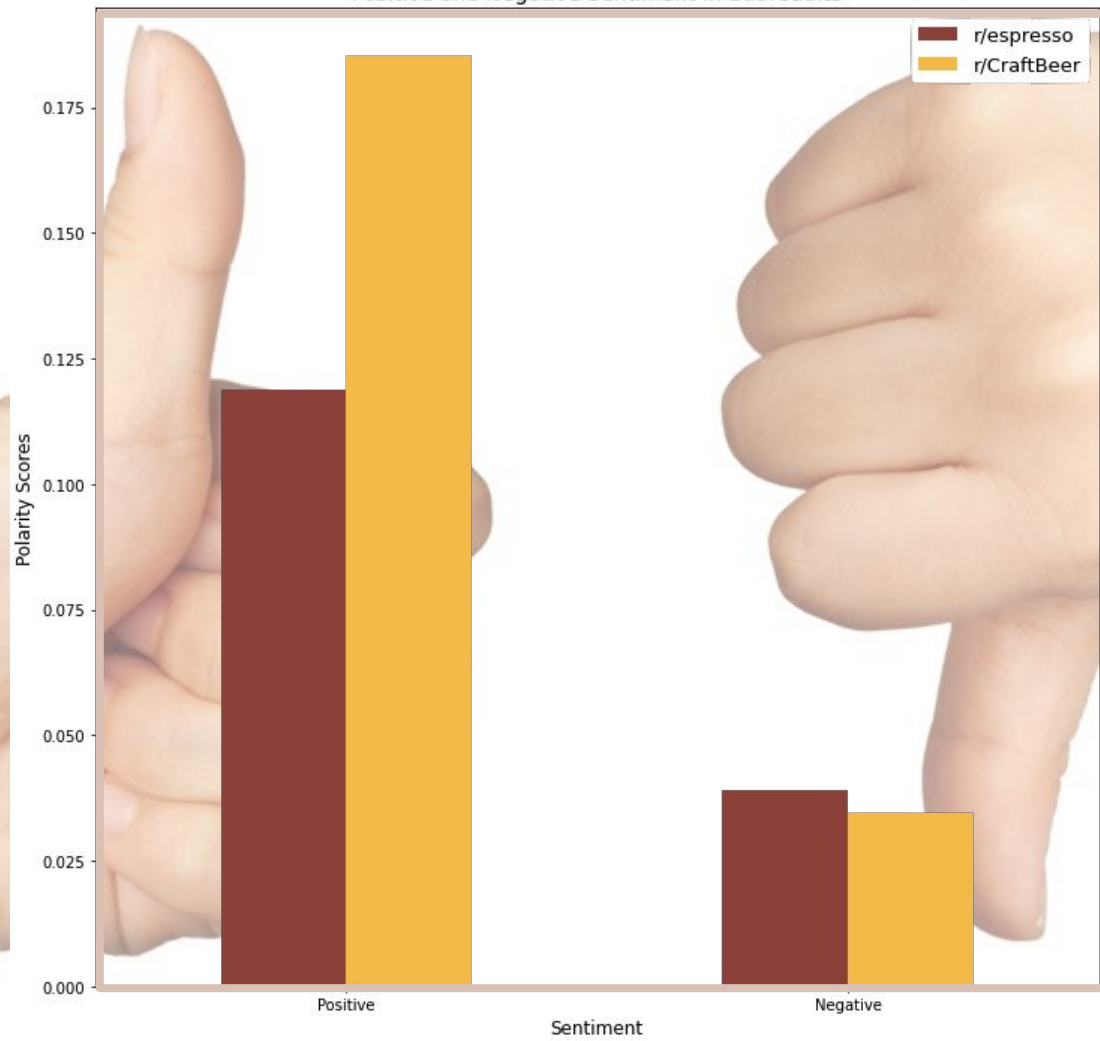
Top 10 Beer Words

Top 10 Words

Positive and Negative Sentiment in Subreddits

03

RECOMMENDATIONS

**Customer insight can be gleaned
from applying NLP to user activity on reddit and
the insight could be useful in targeted marketing.**

Perhaps we need more coffee-flavored beer, or beer-flavored coffee.

Perhaps also this model can be generalized to other data to determine interest across other beverages or food.

Recommendations include but are not limited to:

-applying this model to text in other social media to look for new markets for existing products

-investigating the mis-categorizations within the model for insight into interest in new cross-over products.

"Half of the day I wonder if it's too late for coffee, the other half I wonder if it's too early for alcohol."

—RONALD ULYSSES SWANSON

CHEERS!

Questions?

Credits: