Reddit API & NLP

Meggan Lanpher

Agenda

1. Problem Statement
2. Gather the Data
3. Explore the Data
4. Model the Data
5. Evaluate the Model
6. Answer the Problem

Problem Statement

Which Econ Is It?

Given the title and text of a submission, predict if it was posted to the subreddit r/AskEconomics or the subreddit r/askeconomists.

About the Subreddit Communities

r/AskEconomics

- "Ask questions of academic economists, so we can all know a little more about one of the most important forces in the human world, and be more educated citizens."
- 27.8k Members - Created Mar 31, 2011

r/askeconomists.

- "A central repository for questions about economic theory, research, and policy. Please read the rules before posting."
- 983 Members - Created Jan 9, 2013

Gather & Clean the Data

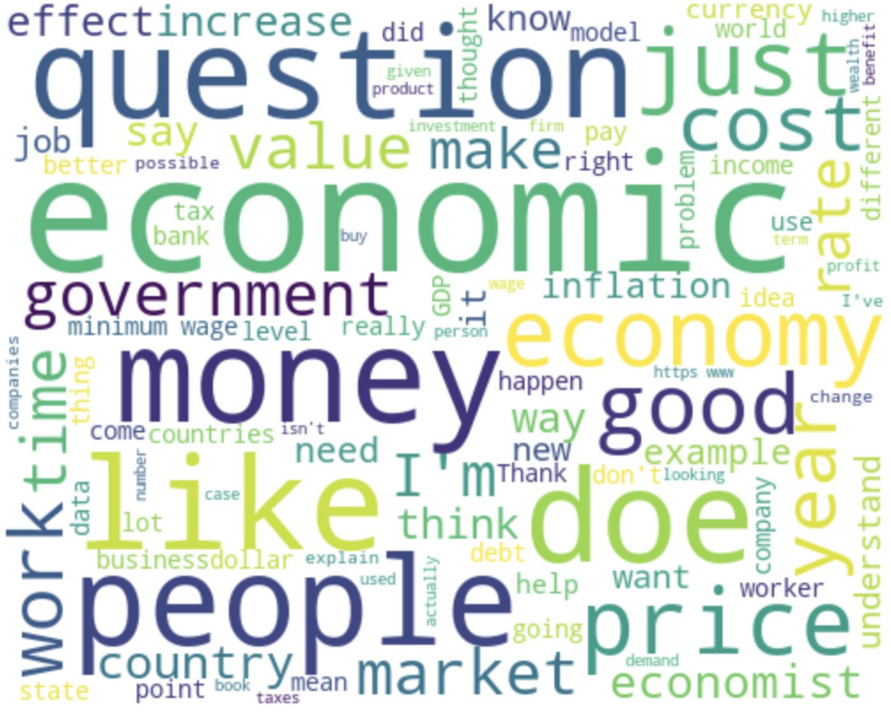
Reddit API requests

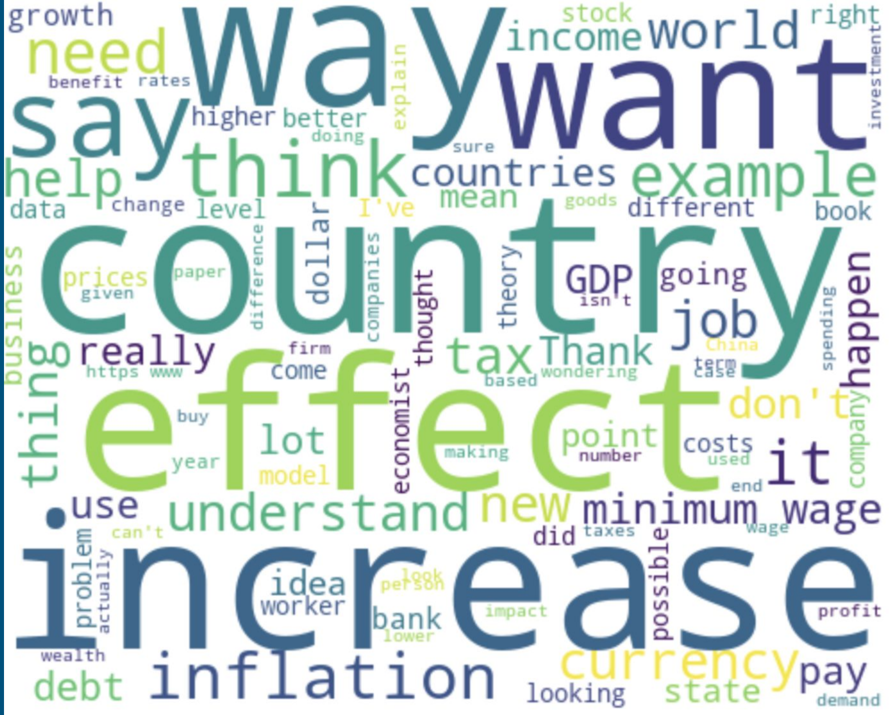
- 'is_self' : 'true',
- 'subreddit' : 'AskEconomics', //   'subreddit' : 'askeconomists',

NLP analysis of “selftext” and “title”

- Dropped rows with [removed] and [deleted] as the “selftext”

Cleaning with Stopwords



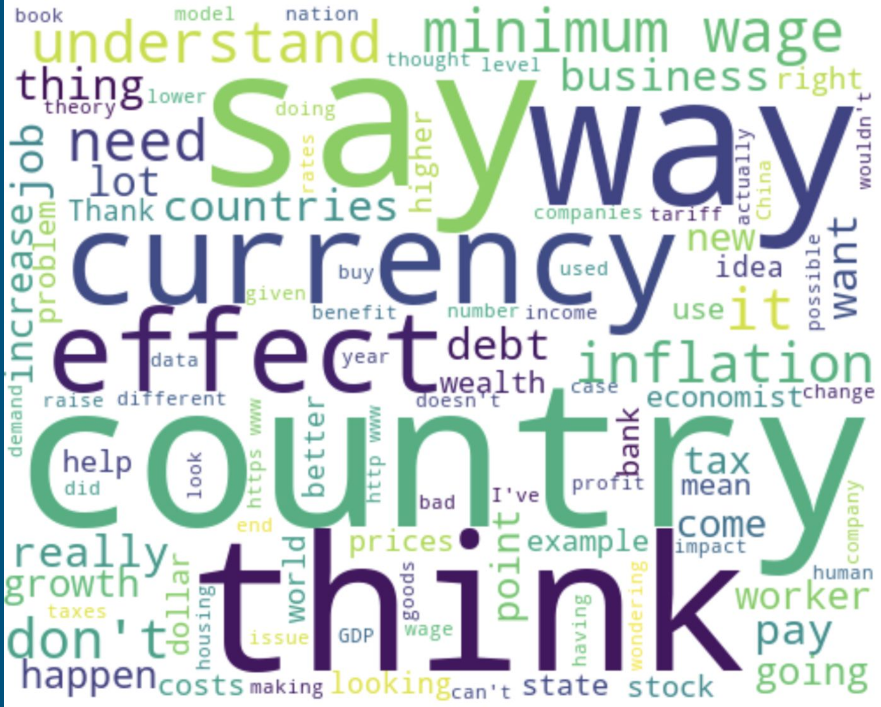


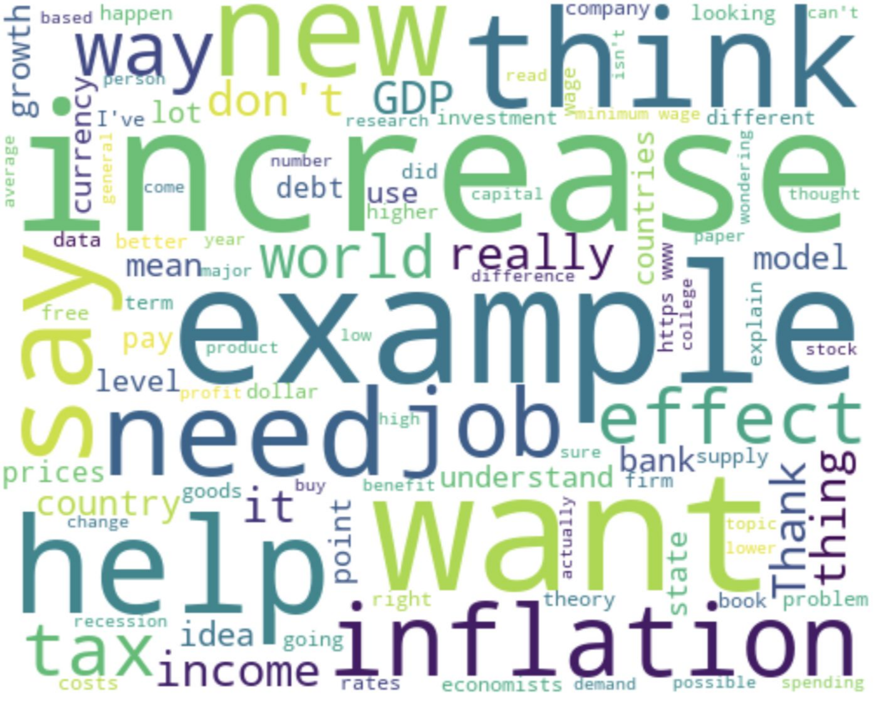
Added Stopwords

- Economics
- Econon
- Economy
- Economic
- Money
- Price
- Economist
- Market
- Cost
- Time
- Government
- Value
- Rate

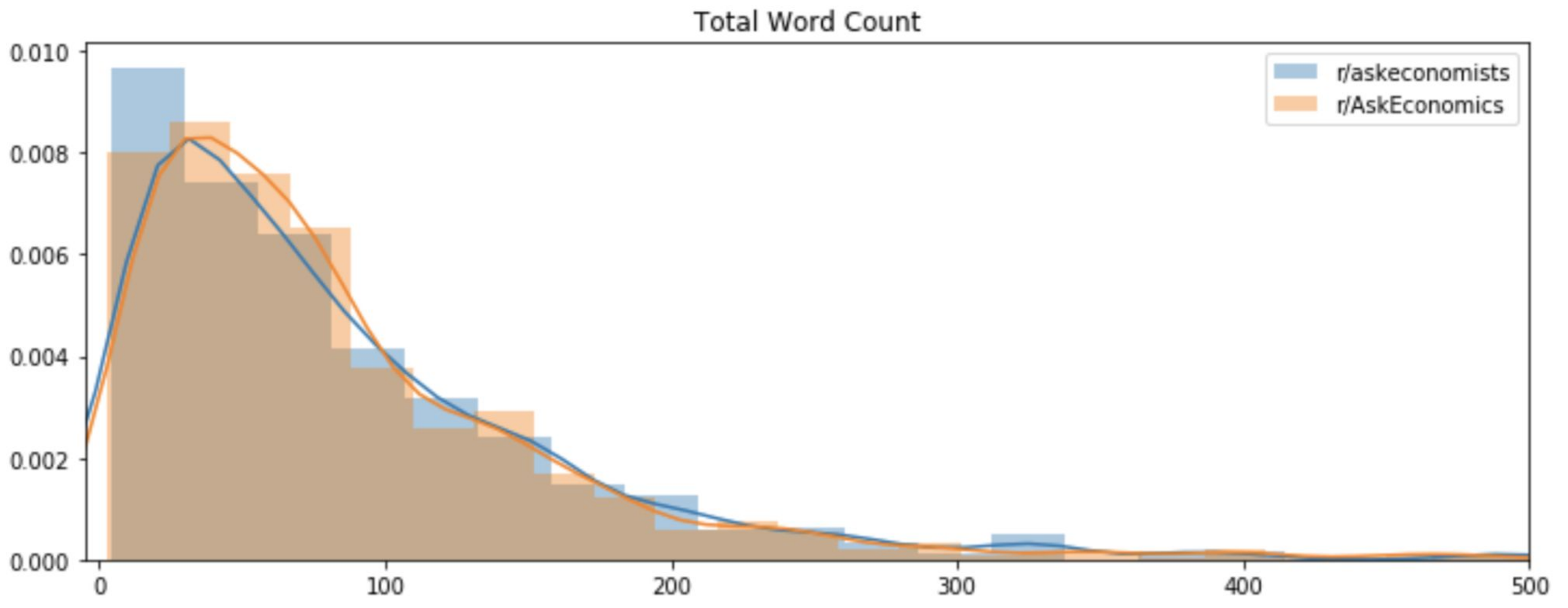
- Does
- Question
- Ve
- Just
- Like
- make
- People
- I'm
- Good
- Work
- Know
- Year

r/askeconomists vs. r/AskEconomics

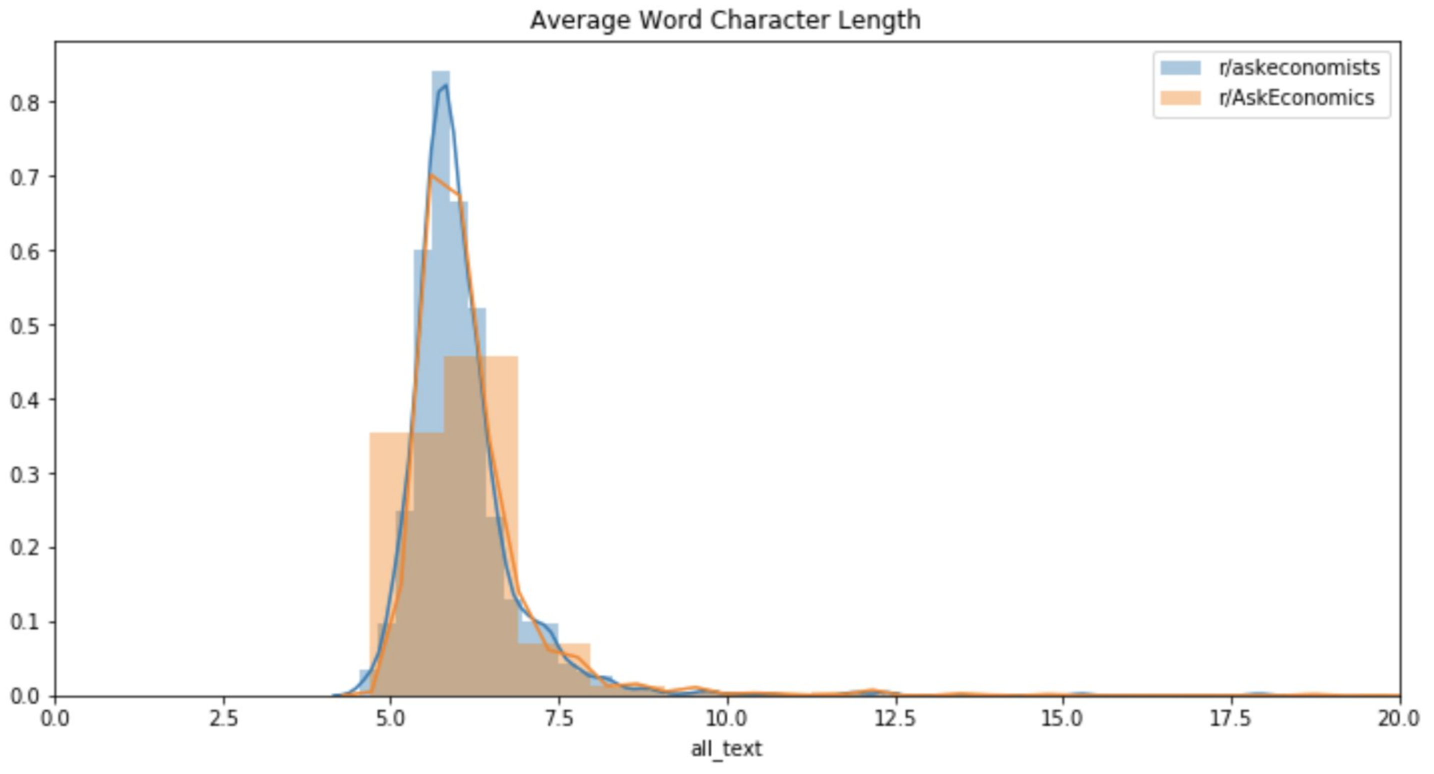




Comparing # of Words in each Subreddit post

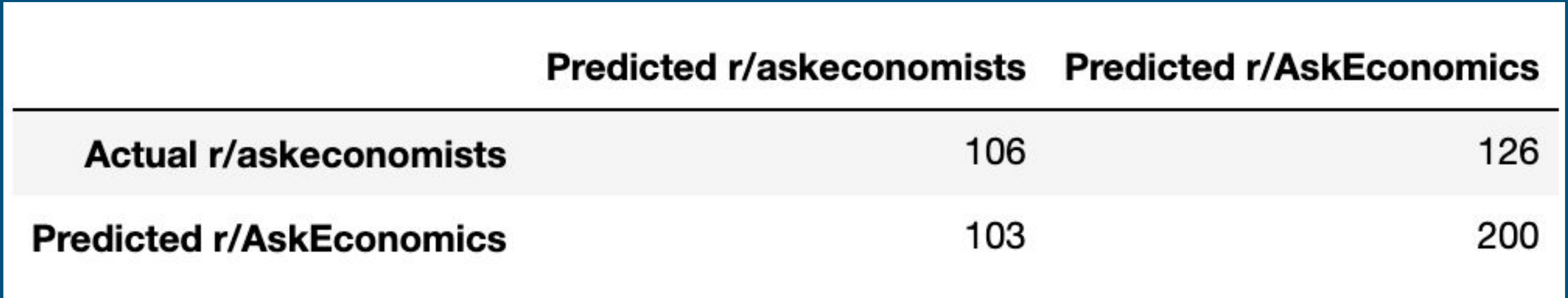


Comparing Avg. Character Length of Words



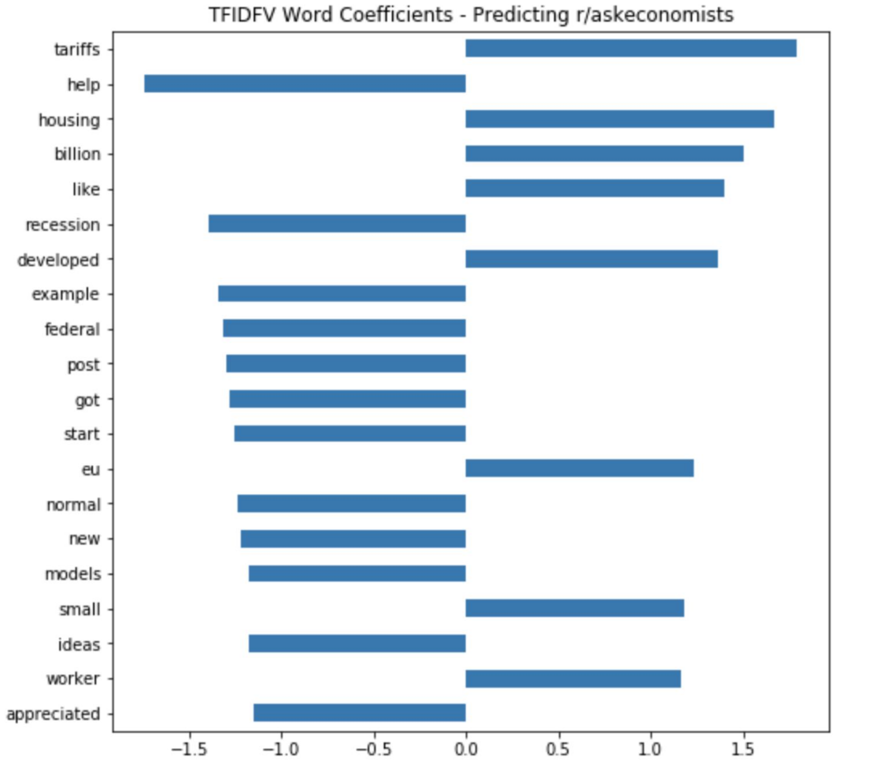
Model the Data

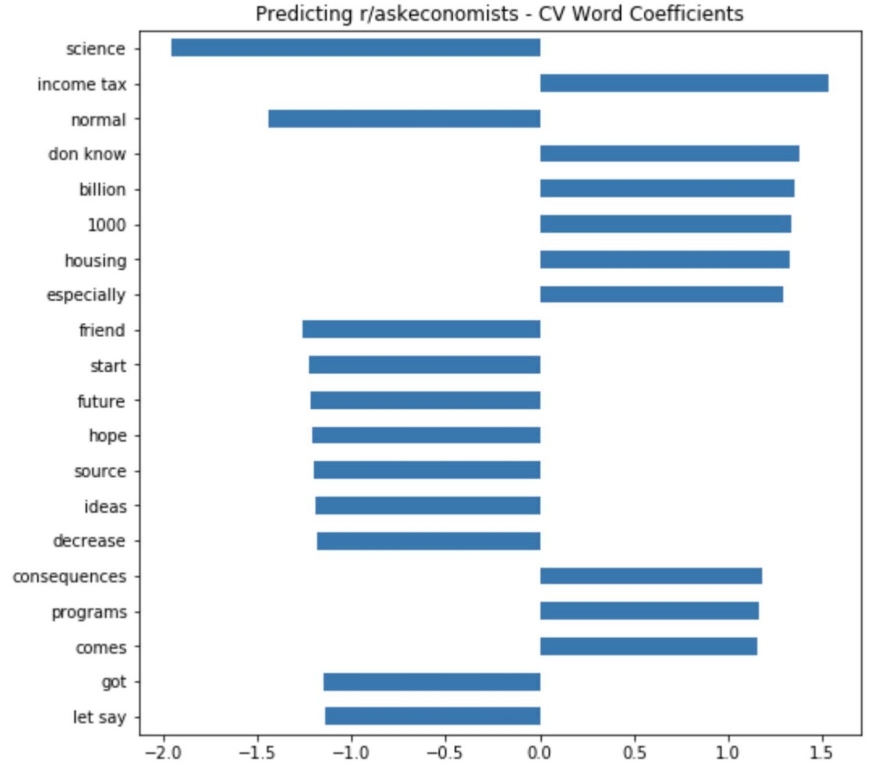
Predicted on new data with 60% accuracy



	Predicted r/askeconomists	Predicted r/AskEconomics
Actual r/askeconomists	106	126
Predicted r/AskEconomics	103	200

Which words had the greatest influence?


TFIDFV Word Coefficients - Predicting r/askeconomists
tariffs
help
housing
billion
like
recession
developed
example
federal
post
got
start
eu
normal
new
models
small
ideas
worker
appreciated


Predicting r/askeconomists - CV Word Coefficients
science
income tax
normal
don know
billion
housing
especially
friend
start
future
hope
source
ideas
decrease
consequences
programs
comes
got
let say