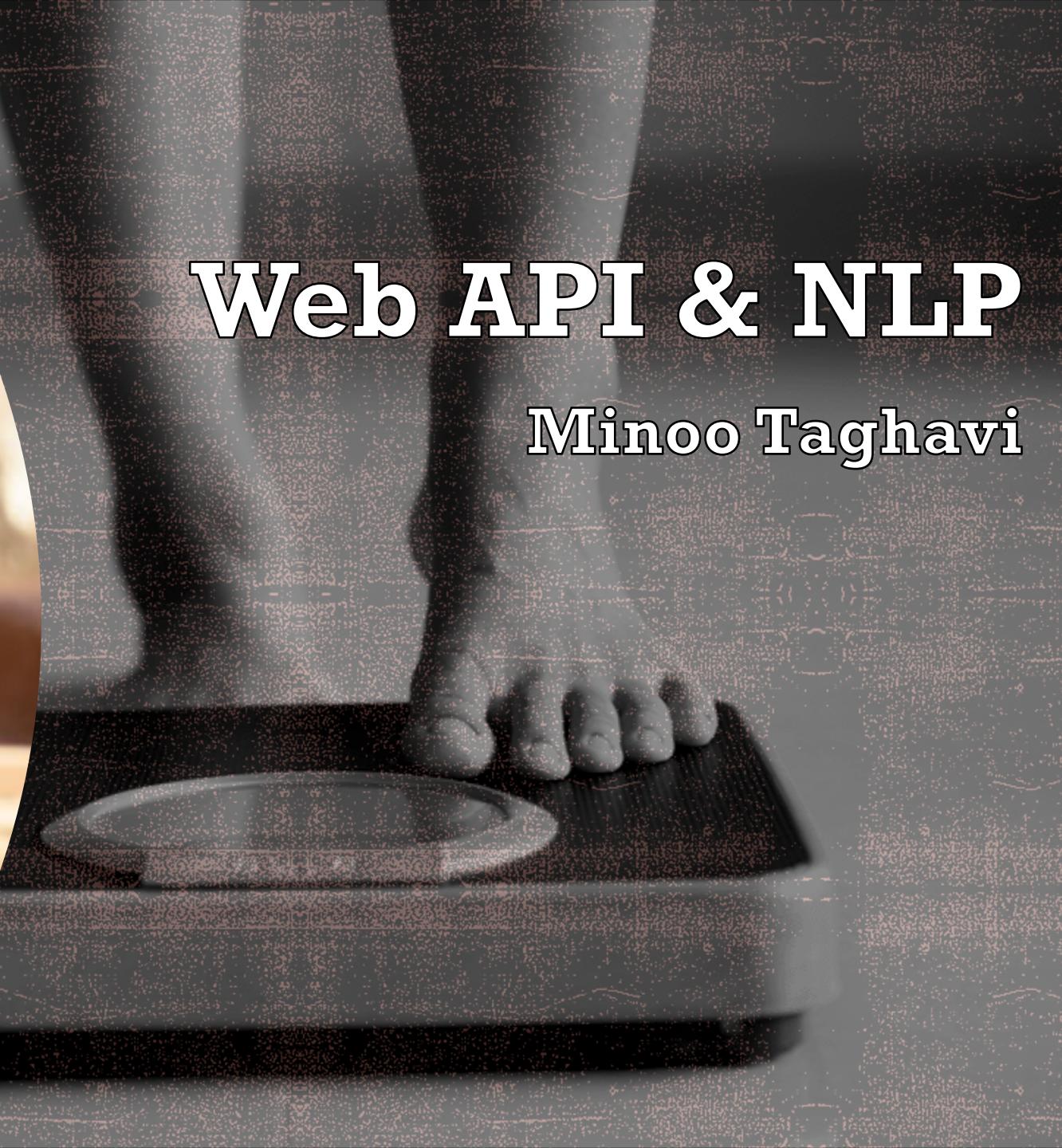




Web API & NLP

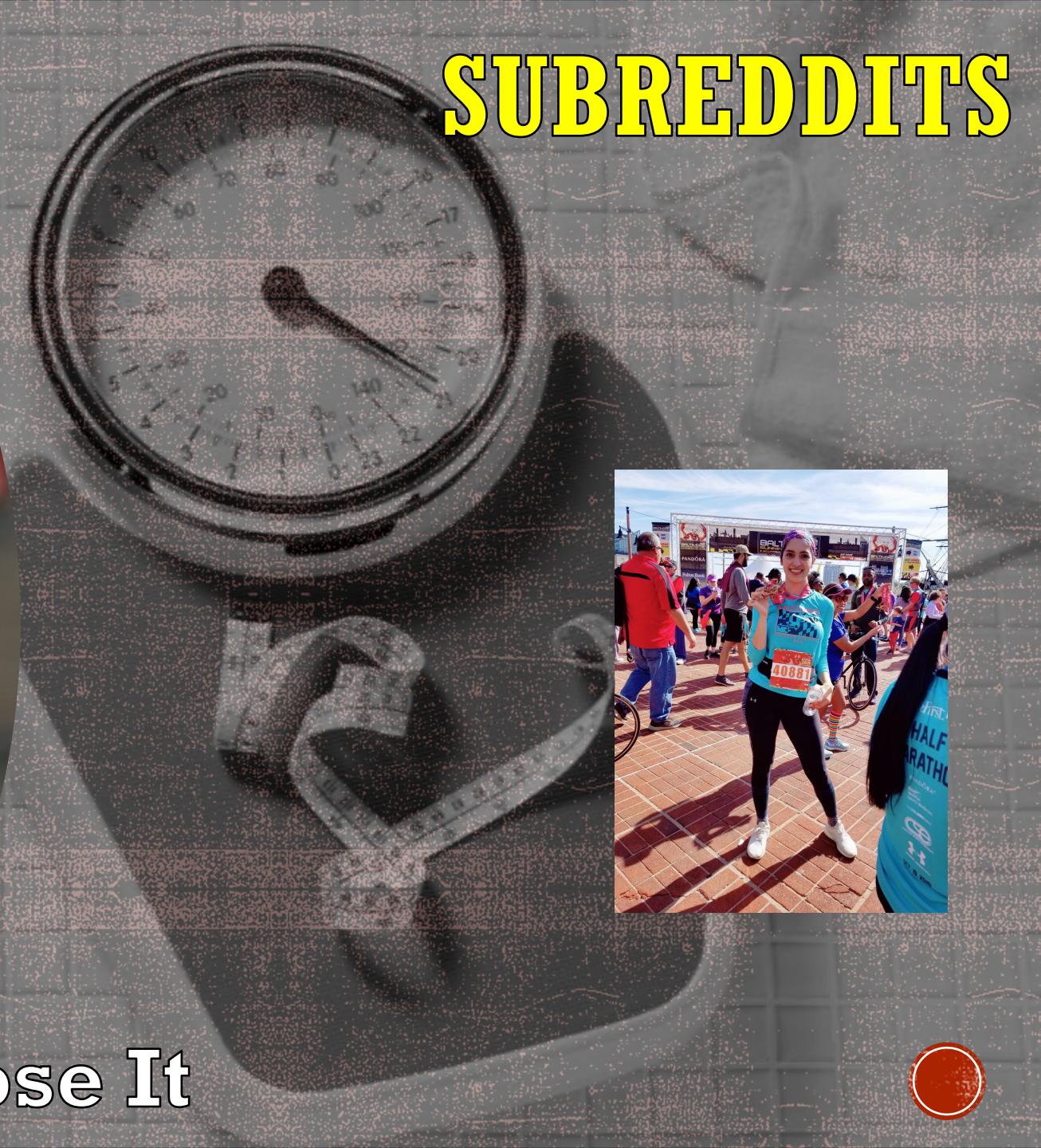


Minoo Taghavi

SUBREDDITS

Running

Lose It



PROBLEM?

- Find a ML Model to predict a Subreddit for a text post/posts!

GOAL?

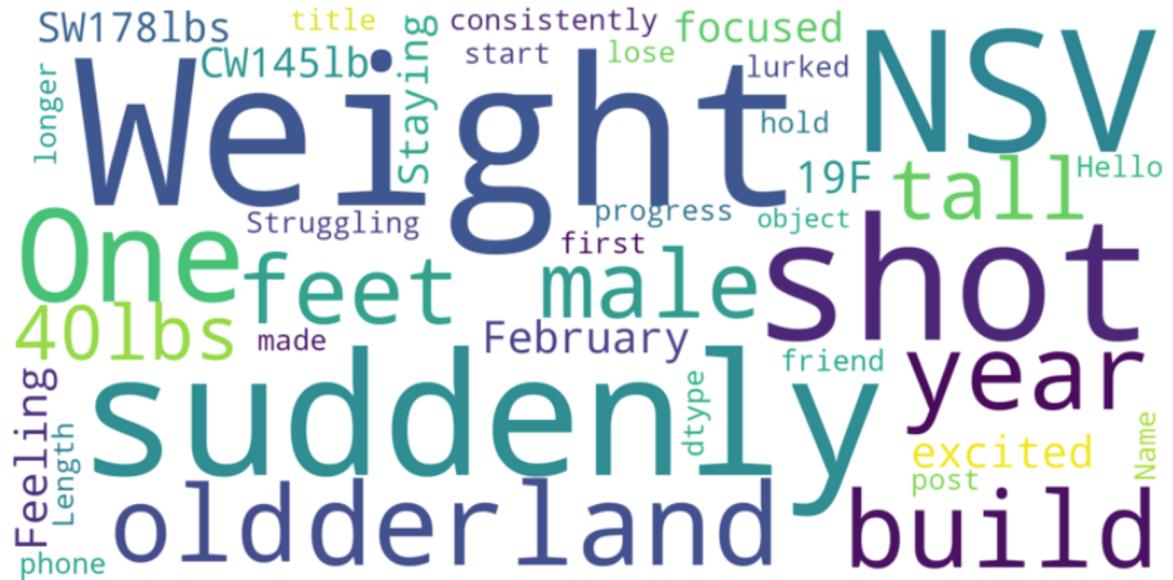
- Determine which subreddit a given post came from!

DATA FROM REDDIT!

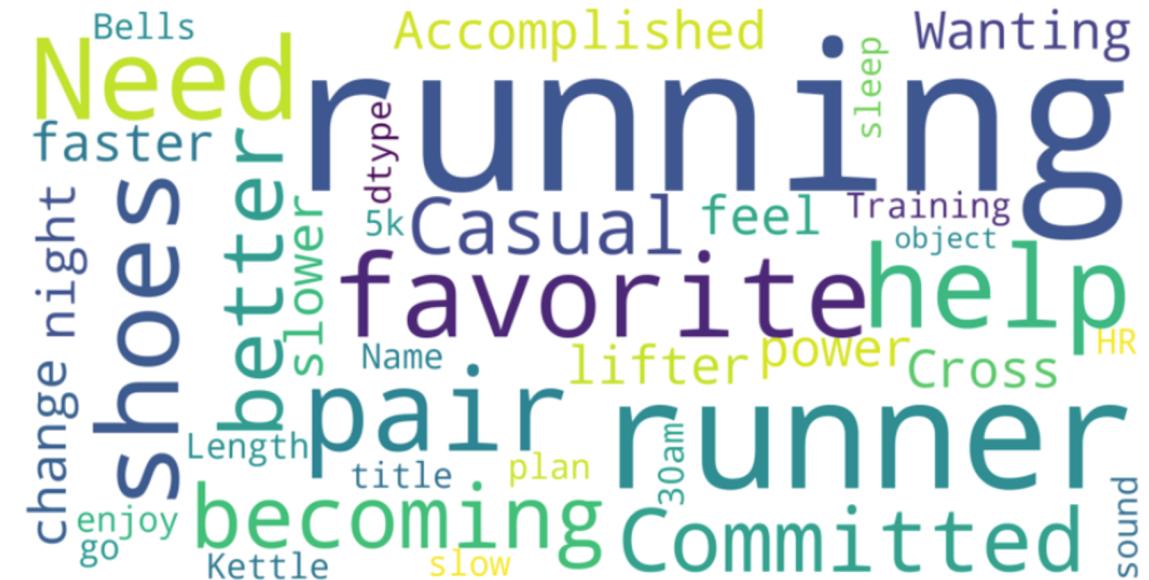


reddit





MOST COMMON USED WORDS IN LOSE IT BEFORE PREPROCESSING!



MOST COMMON USED WORDS IN RUNNING BEFORE PREPROCESSING!



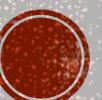
PREPROCESSING!

Made a new Stop-words!

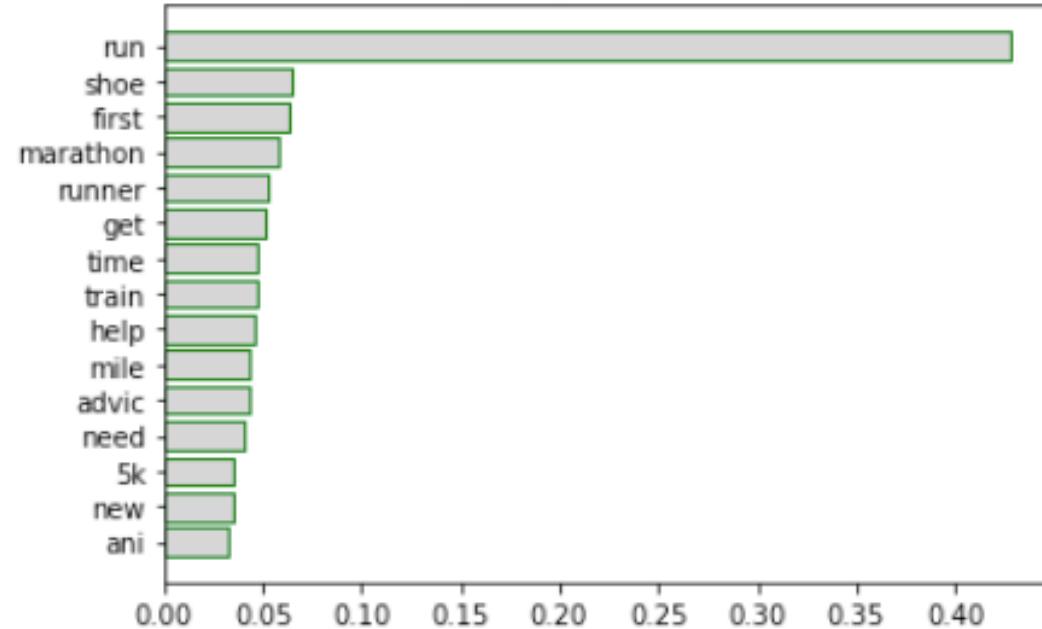
Used Tokenization

Used Stemming

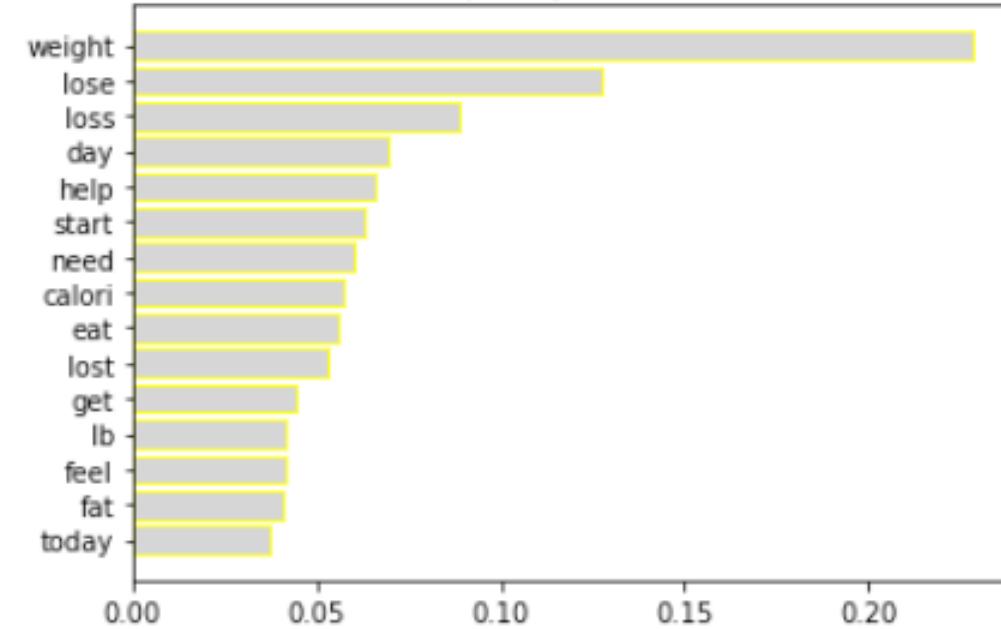
Tried re



Word Frequency for RUNNING Words

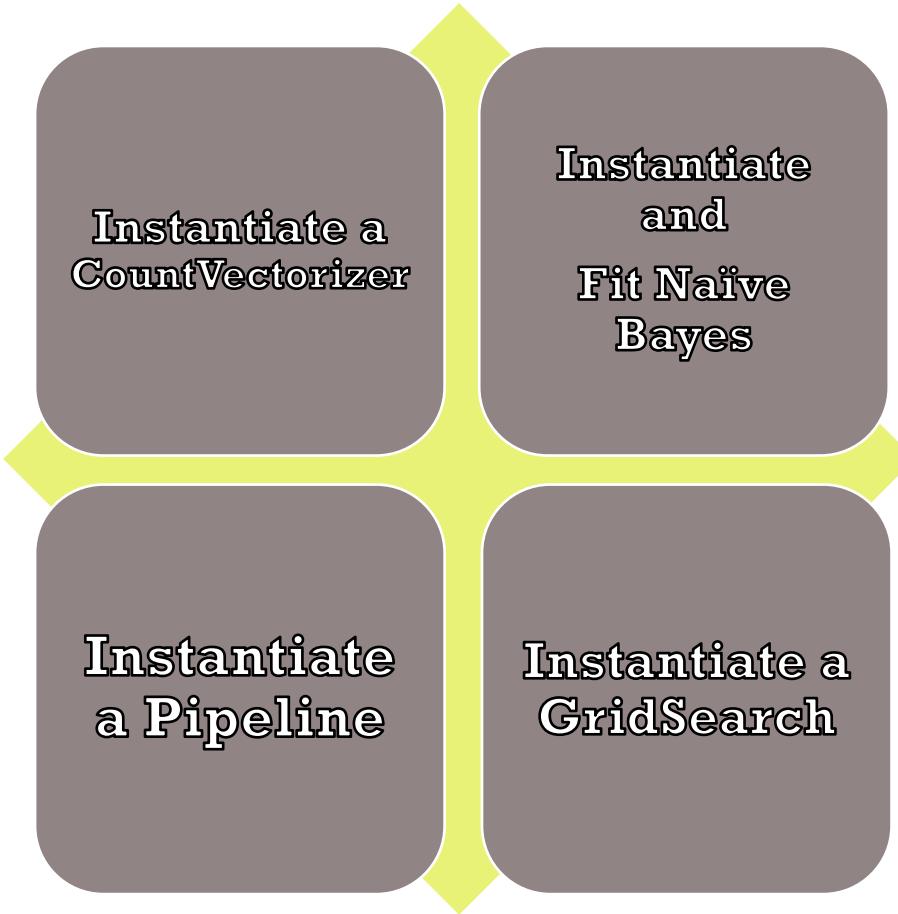


word frequency for loseit words



RUNNING OR LOSE IT?



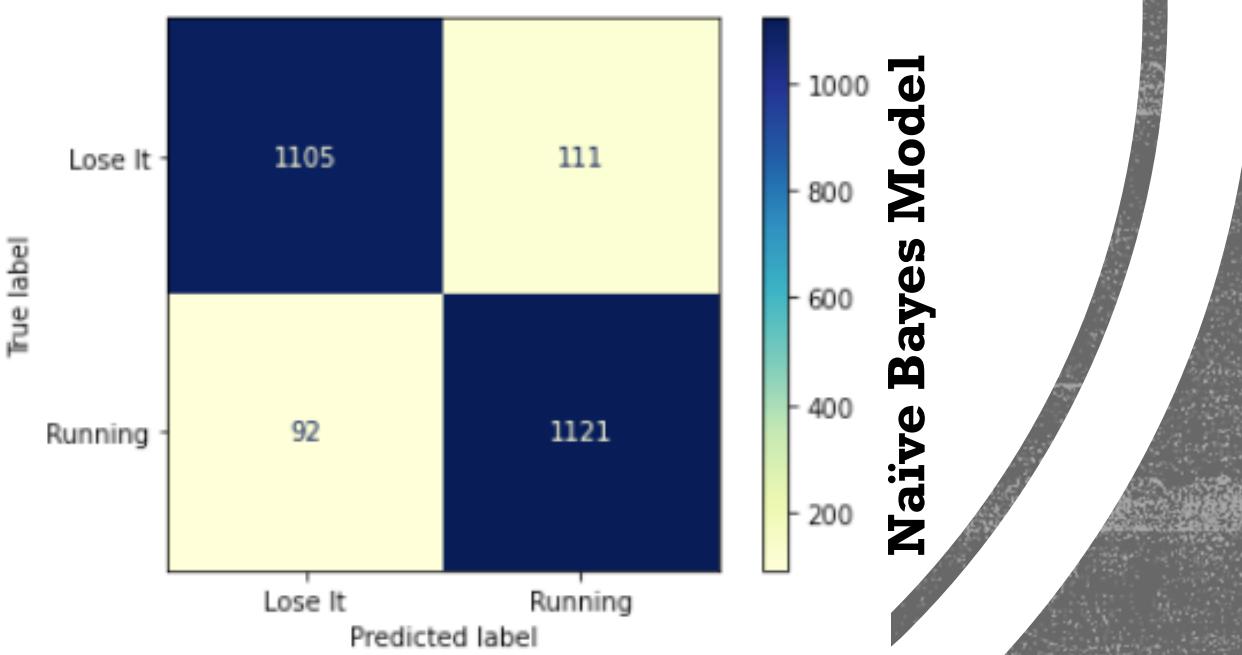
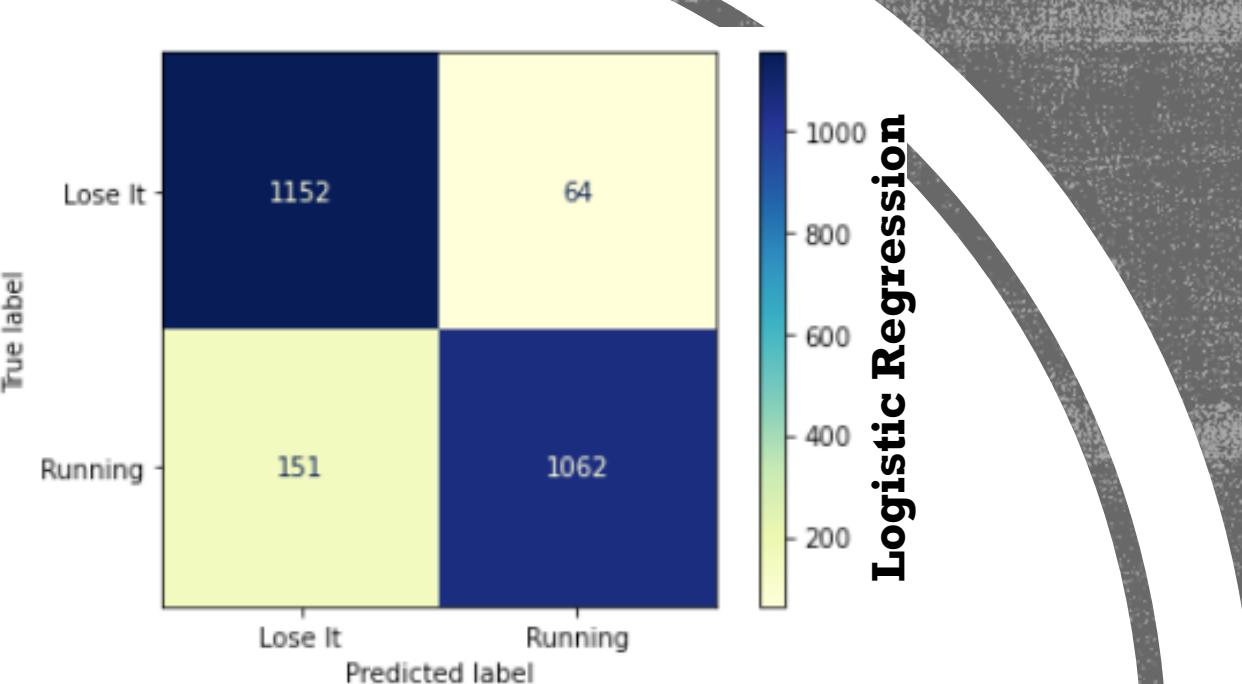


MODELING THE DATA



CONFUSION MATRIX

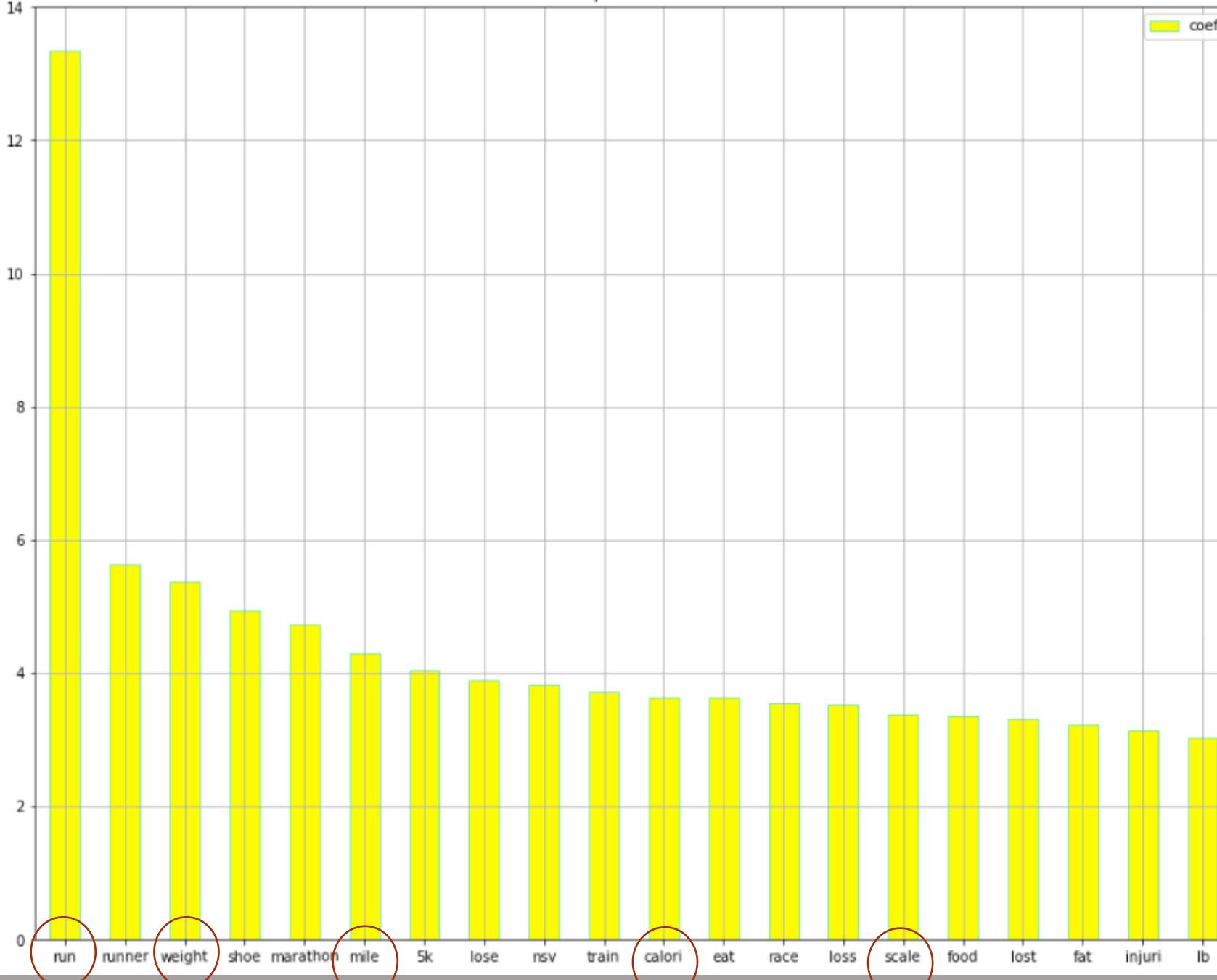
- Logistic Regression with Count Vectorizer
- Naïve Bayes with Count Vectorizer
- Train Accuracy on both model %98-99
- Test Accuracy on both model %92



Run
Runner
Weight
Shoes
Marathon
Mile
5k
Lose
Train
Calories
Eat
Race
Loss
Scale
Food
Lost
Fat
Injury

EVALUATING THE MODEL

Top 10 Words



PROBLEM?

- Find a ML Model to predict a Subreddit for a text post/posts!

GOAL?

- Determine which subreddit a given post came from!

Data from Reddit

Model the Data

First marathon! Done Solo



Running!



Lose it!

Just weighed myself, need to make a change



HOPEFULLY

PRESENTATION END

makeameme.org

