

# **Sentiment prediction of subreddit posts**

Devaraja Mudeppa  
SocialMedia, LLC

# Table of contents

01

**Objective**

03

**ML model  
Training and testing**

02

**Data collection  
EDA**

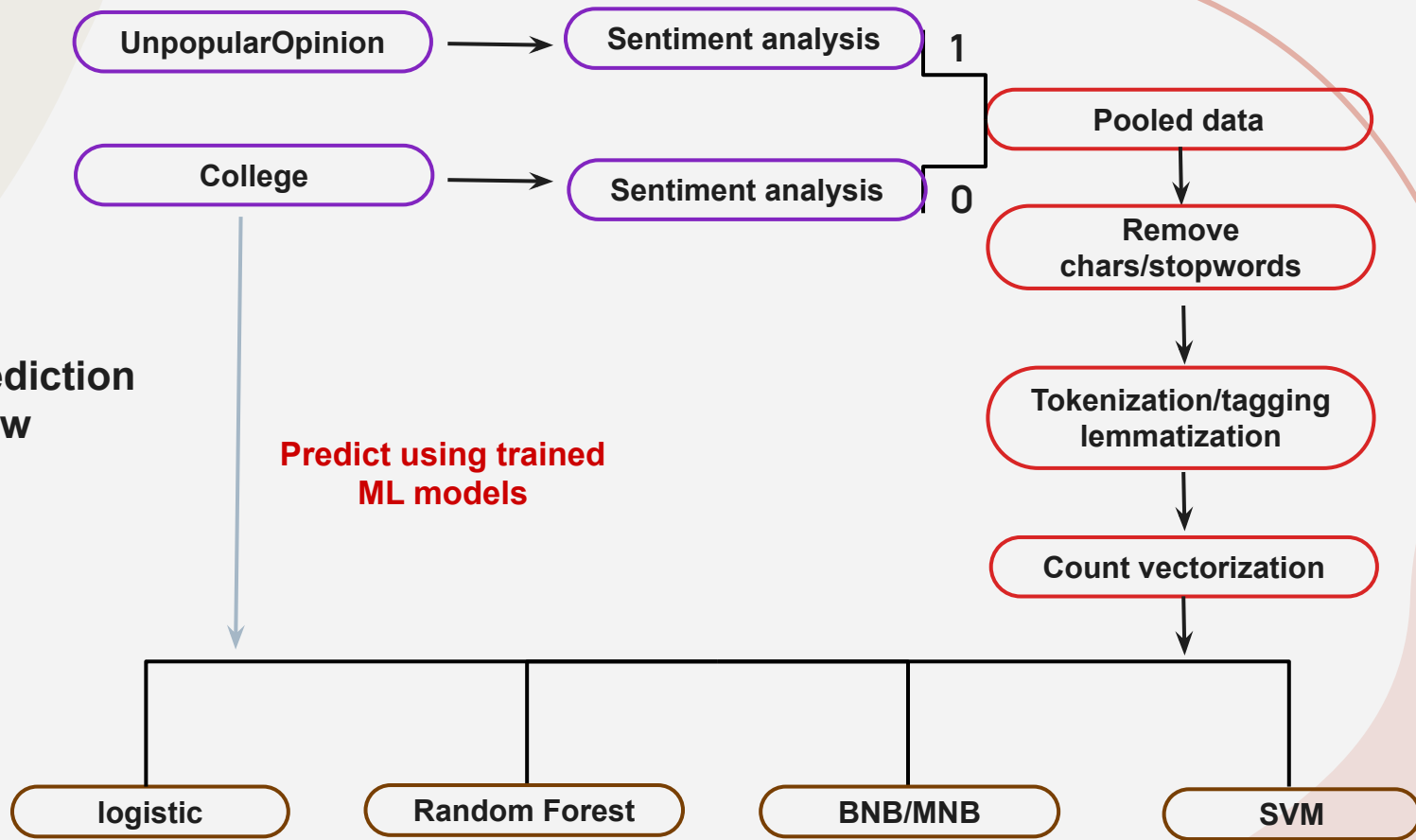
04

**Predictions  
Summary**

# Objective

- Predicting emotions of social media posts based on the sentiment analysis
- Crisis detection, Public relations, Targeted ad placement, Trend forecasting
- How accurately NLPs predict the social media posts?
- College and UnpopularOpinion subreddits
- Classification models

## Sentiment prediction WorkFlow



# Scrapping: Extraction of posts from subreddits

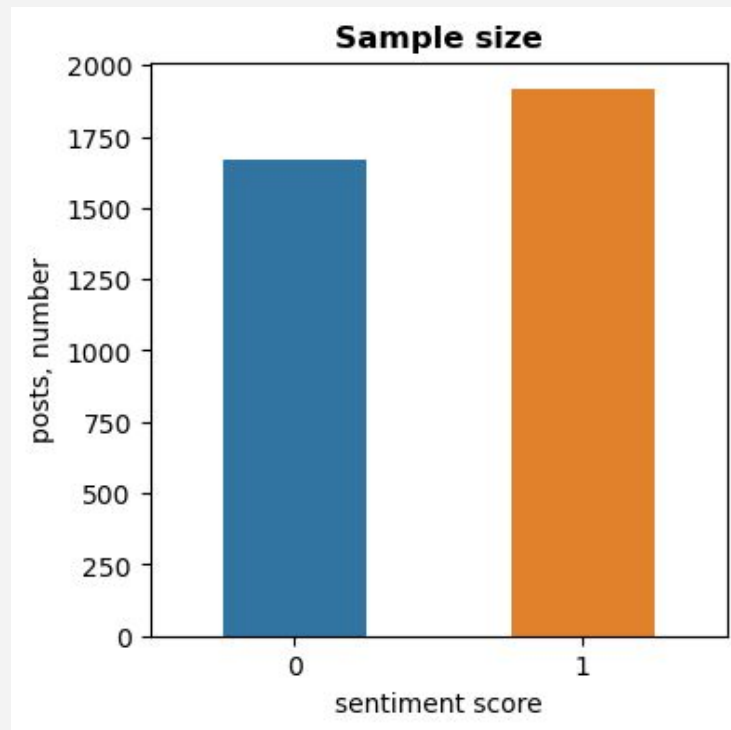
## r/UnpopularOpinion

- Non overlapping posts
- Sleep time: 300 seconds
- Scraped posts: 1914

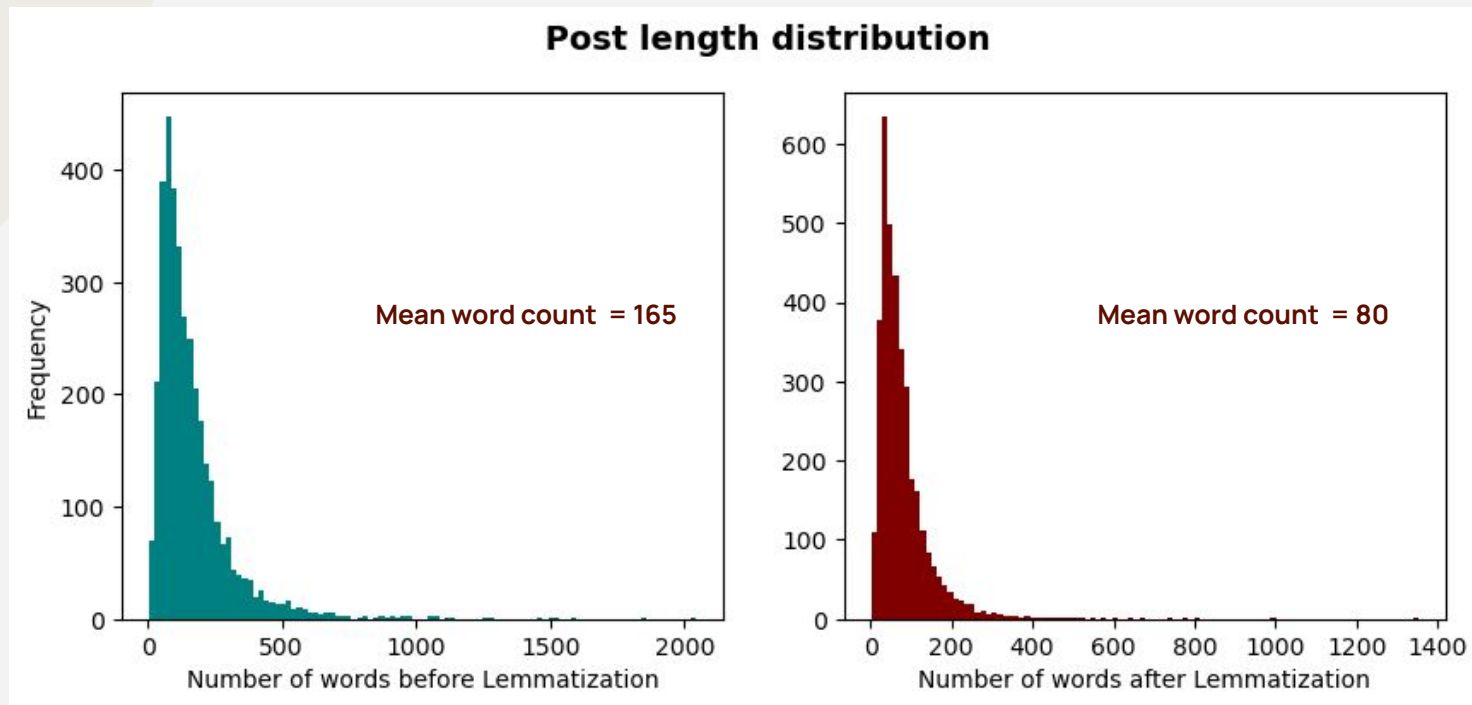
## r/College

- Non overlapping posts
- Sleep time: 300 seconds
- Scraped posts: 1670

Total posts = 3584



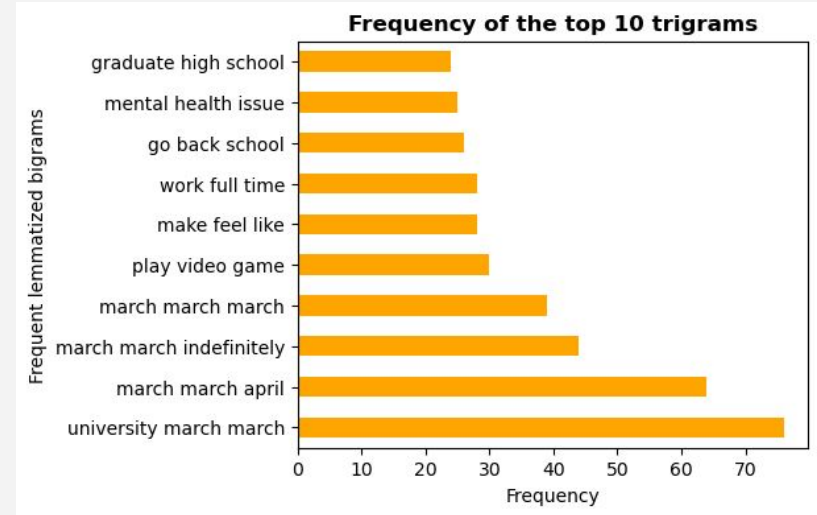
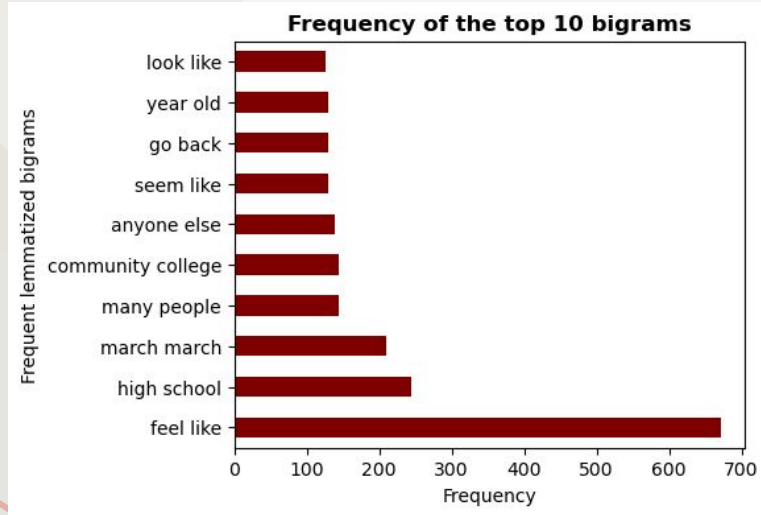
# Lemmatization significantly reduces word count



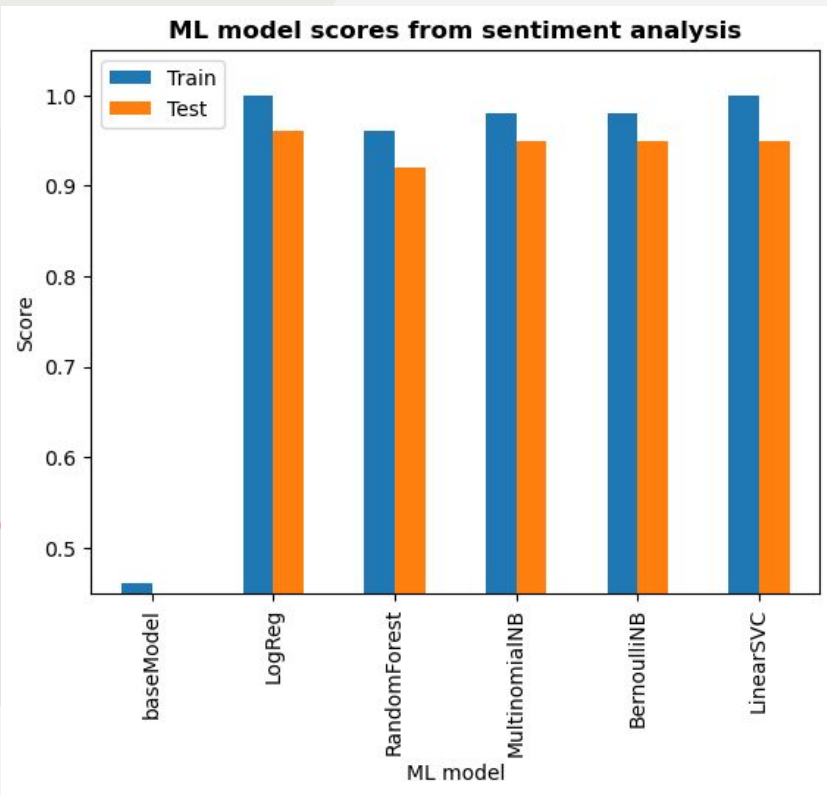
Total word count = 590194

Total word count = 285079

# Frequent bigrams and trigrams



# All models displayed similar scores

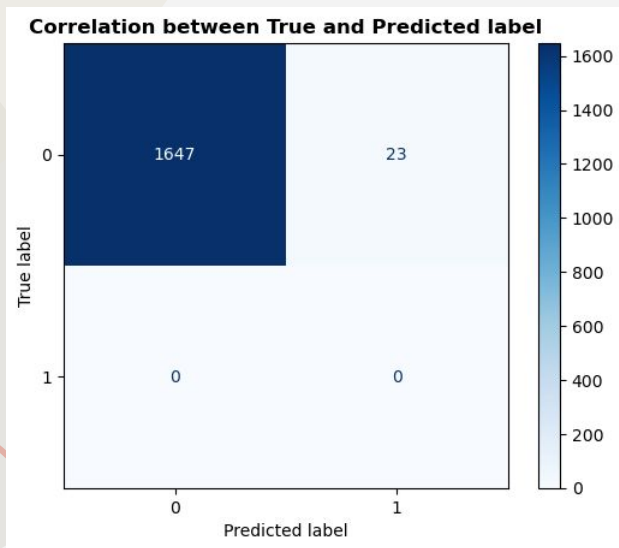


K-Fold cross-val scores for both train and test samples remain with 3% range across all five folds indicating the models is not over or underfit.



# Multinomial NB predicted least false-labeled posts

- Trained MLs deployed to predict college subreddit posts
- College posts were scored as zeros



	precision	recall	f1-score	support
0	0.986	1.000	0.993	1647
1	0.000	0.000	0.000	23
accuracy			0.986	1670
macro avg	0.493	0.500	0.497	1670
weighted avg	0.973	0.986	0.979	1670

What are these 23 false-labeled posts?

BernoulliNB: 28  
Random Forest: 38  
Logistic Regression: 43  
SVM: 31

# Potential reasons for false-labeling of posts

- Slang, emoji, irony, sarcasm,
- Expressions from different cultural backgrounds
- Indirect expression
- Lack of vocab from different part of world
- Algorithmic limitations

Examples of incorrect predictions:

'feeling out of place i don't think i belong here, everyone here is so much better than me, at academics, at extracurriculars, at everything single damn thing! all i do is wallow in my own self-pity!\ni can't do anything right for the life of me!\ni think i am literally the most boring person in class. and all i can do is wallow in my own self-pity. \nplease help a stranger find their identity or peace in their uselessness'

'why do only so little colleges offer men's volleyball or lacrosse?the only two good sports i'm good at, and my state (nj) barely has any colleges that do have the sport, even nationally the sports aren't very available. it's already annoying that i had to join a men's club since no highschool offers it but almost everywhere there is an equivalent for women's lacrosse and volleyball. i understand that there are/were a lot of sports that women didn't have that men did, but nowadays it's mostly equal '

# Summary

- NLPs are powerful tools in analysing the social media posts
- MultinomialNB predicted with lowest incorrect labels
- Further fine optimization of models might help to reduce errors
- However, training model with with more resources could significantly help
- I recommend investing additional resources to train the models to get near to zero incorrect labels.



**THANKS!!!!**

**Questions?**