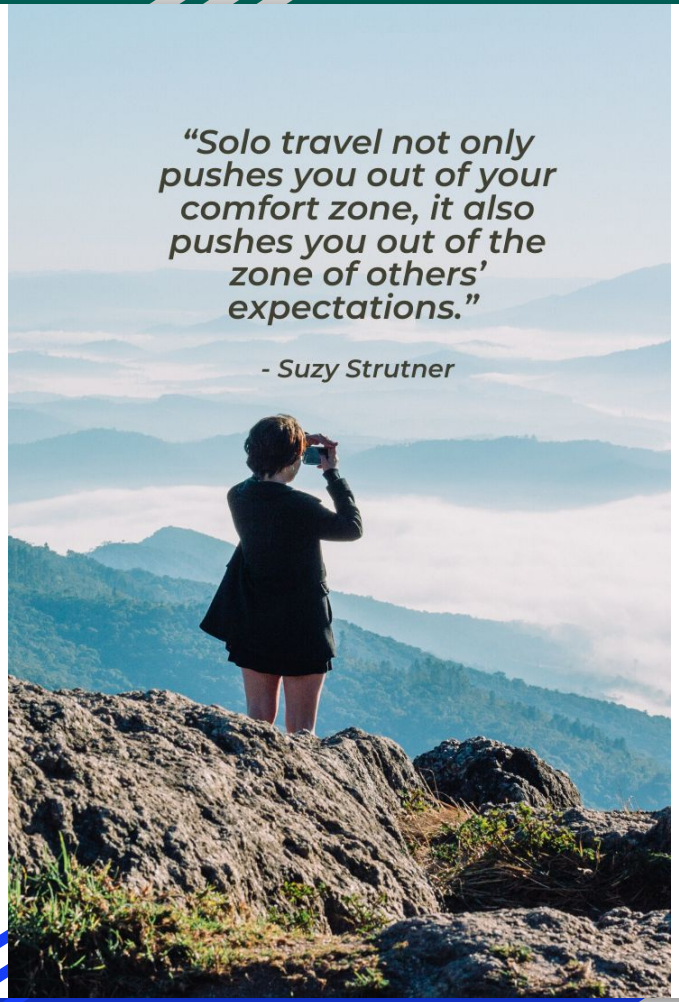
The Travel Niche:
Solo travel and Road trips

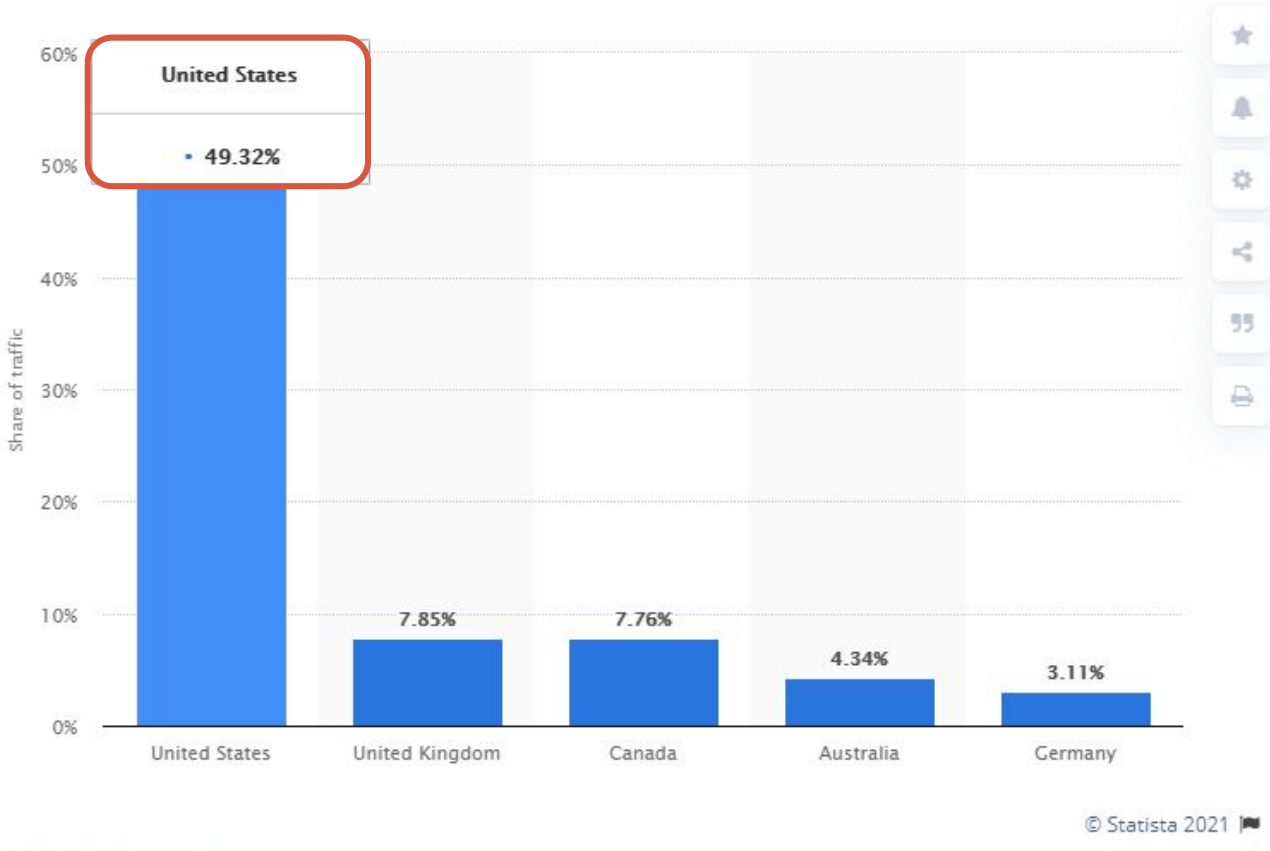
DSI Project 3 by: PREM HONGHIMAPHAN

INTRODUCTION


"Solo travel not only pushes you out of your comfort zone, it also pushes you out of the zone of others' expectations."
- Suzy Strutner

"Because the greatest part of a road trip isn't arriving at your destination.
It's all the wild stuff that happens along the way." - Emma Chase

Target Destination: United States



https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/

r/solotravel

r/roadtrip

The Data

Number of Posts retrieved → Remove Duplicates → Replace Null Values

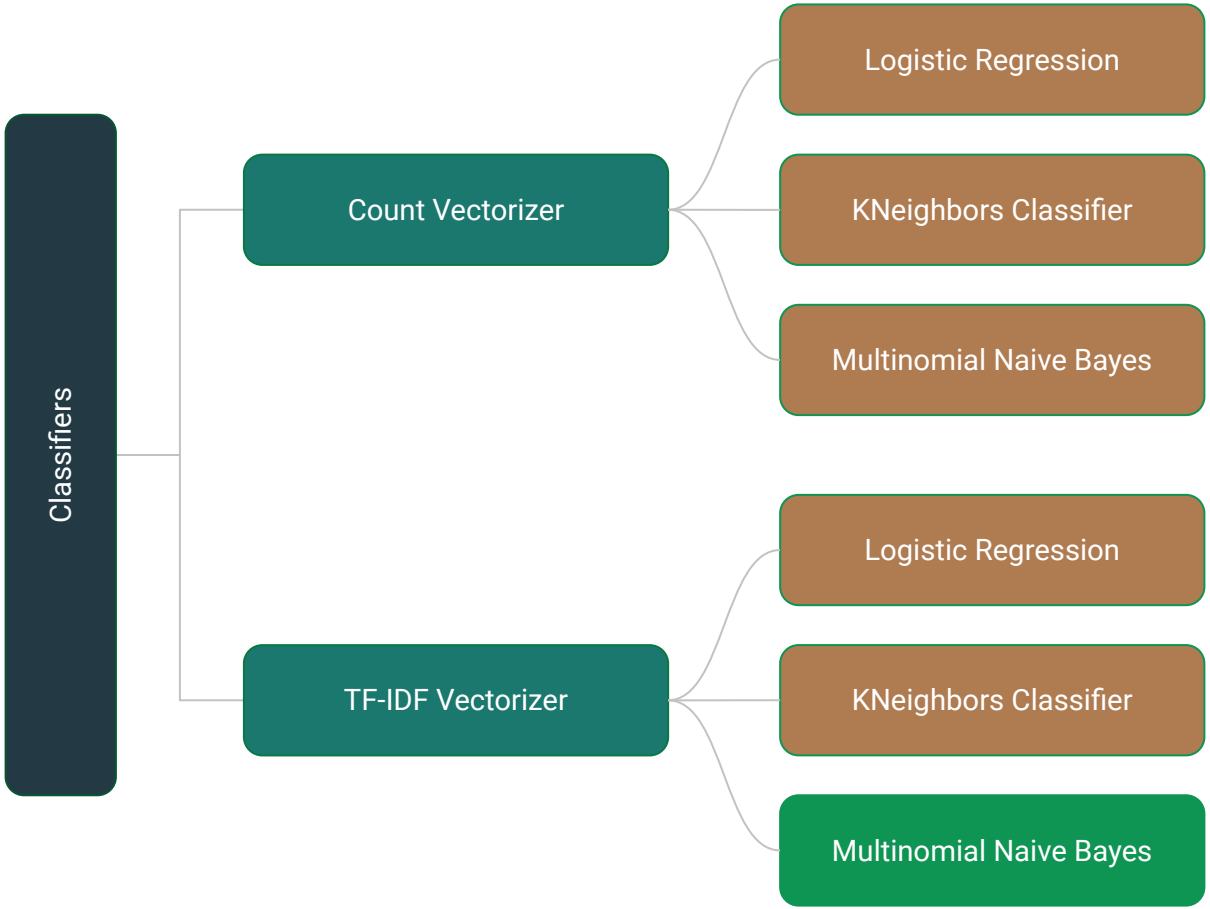
	Number of Posts retrieved	Remove Duplicates
Solotravel	1,230	928
Road trip	1,240	890

	Replace Null Values
Total Data	1,818

About the Data

Feature	Type	Description
subreddit	object	Name of subreddit the posts belong to
selftext	object	Content of the post by users
title	object	Title of the post by users
alltext	object	Combination of selftext and title
clean_text	object	Cleaned text of alltext column
classifier	int	1=Solotravel, 0=Roadtrip

MODELING


Classifiers
Count Vectorizer
Logistic Regression
KNeighbors Classifier
Multinomial Naive Bayes
TF-IDF Vectorizer
Logistic Regression
KNeighbors Classifier
Multinomial Naive Bayes

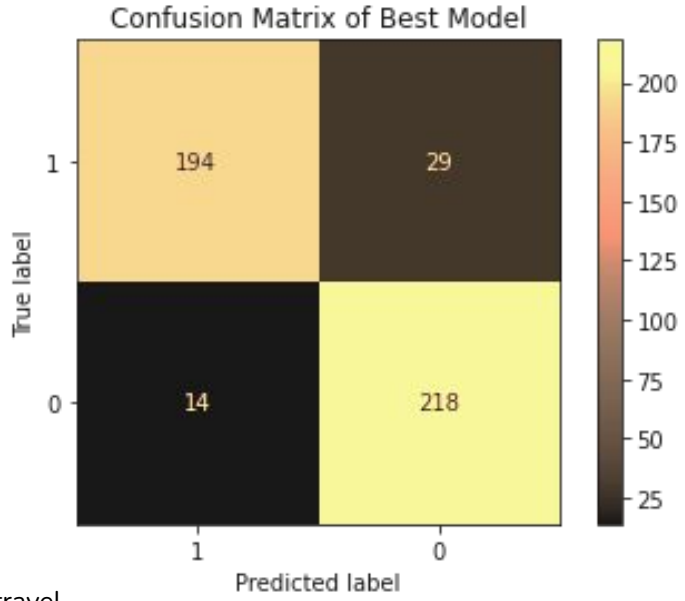
Multinomial Naive Bayes with TF-IDF Vectorizer using Hyperparameters

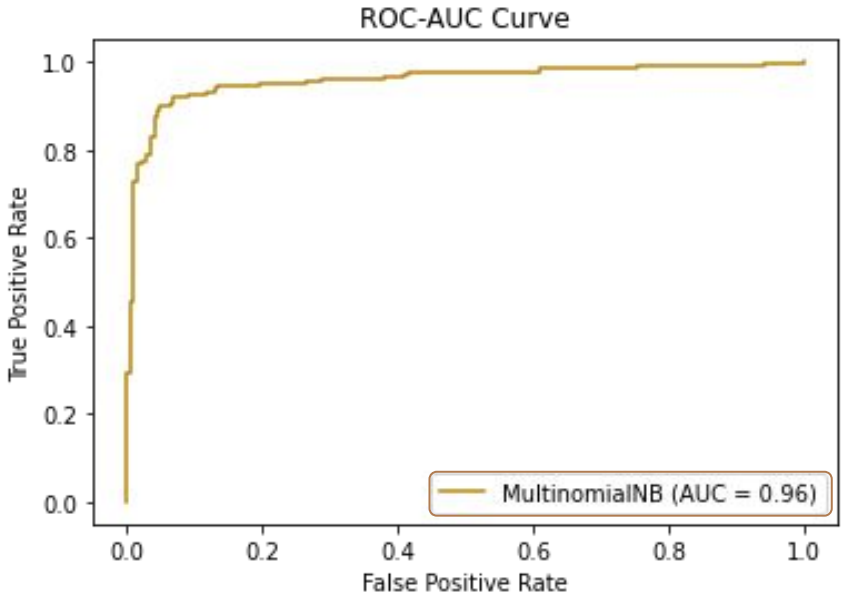
- Training Score: 0.909
- Testing Score: 0.9055

Best features retrieved from doing Gridsearch:

- max_df = 0.9
- min_df = 3
- max_features = 2000
- ngram_range = (1,2)
- alpha = 2

Confusion Matrix





1 = Solotravel
0 = Roadtrip

FINDINGS

Top 10 words that are correctly classified

Solotravel

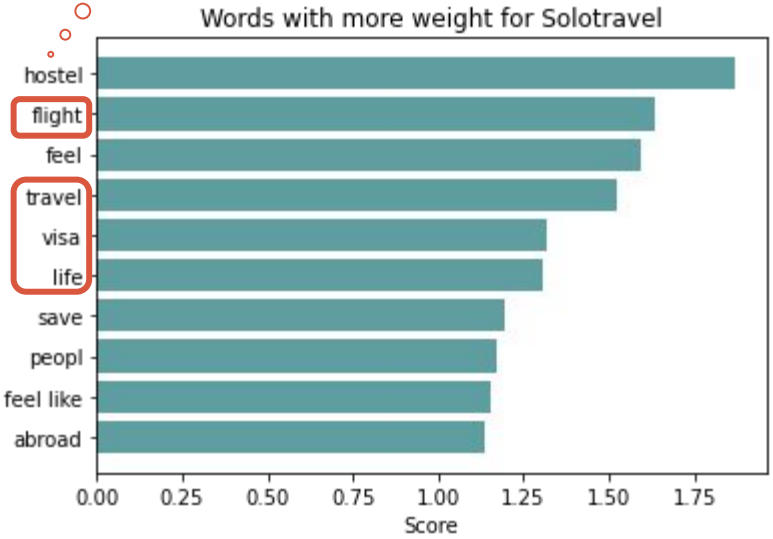
- Commun pleas
- Shit
- Complain
- Complain certain
- Share victori
- Constantly
- Love travel
- Content blog
- Contin
- Reminisc

Road Trip

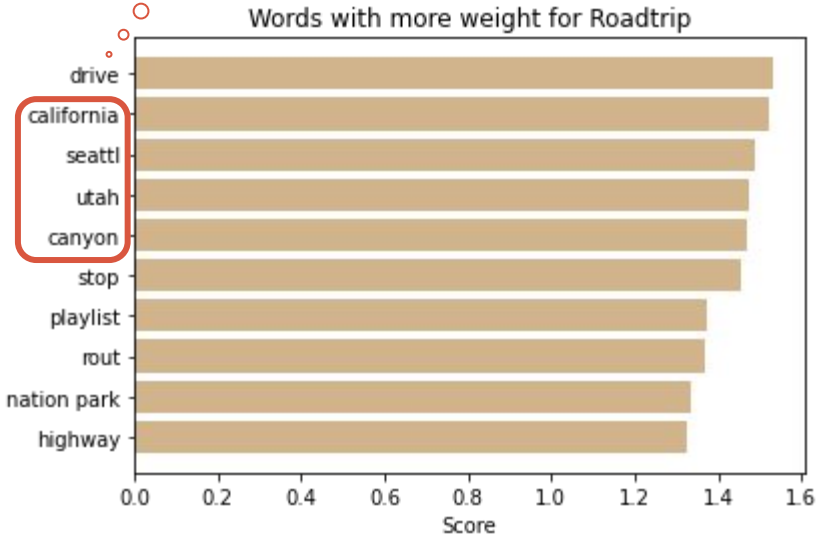
- Savannah
- Geocach
- Oklahoma
- Dalla
- Bay area
- Bc
- Carolina
- North carolina
- Friend plan
- Awawego blogspot

Top 10 words that appear in both

Hostel



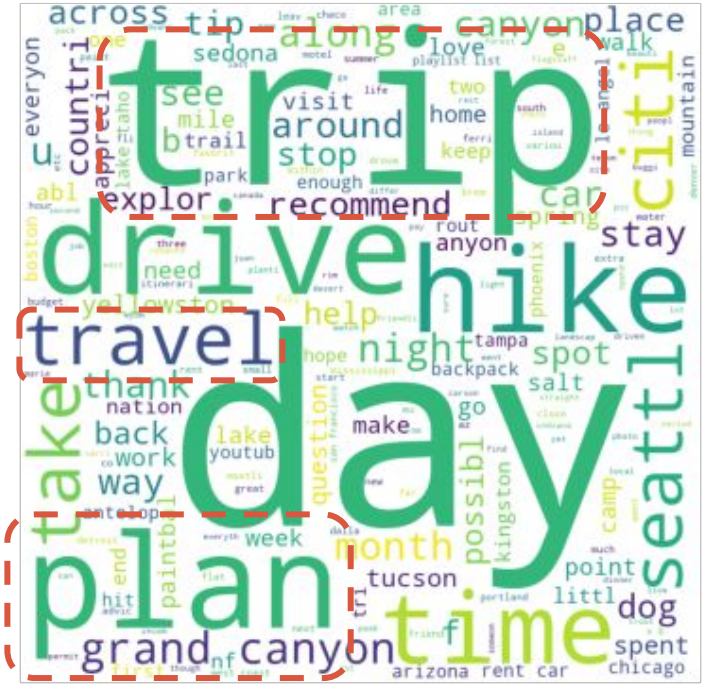
Drive



Words causing misclassifications

Predicted Solotravel, but actually Roadtrip

Predicted Roadtrip, but actually Solotravel



CONCLUSION

Solo Travel

‘Visa’, ‘Flight’, ‘Hostel’

Road Trip

‘California’, ‘Grand Canyon’

‘Plan’, ‘Time’

Travel is dependent on time of year and weather conditions

lonely planet

tripadvisor

Pantip
Learn, Share & Fun

Other discussion forums should also be taken into account, local and global.

THANK YOU