# Finding Tourists Through Social Media

Andrew Seefeldt
Data Scientist

# Overview

COVID-19 has changed the world. Tourism marketing in Costa Rica has not kept up with the change.

Social Media marketing can be challenging due to the amount of Expats that that occupy the same channels

To increase efficiency in marketing we need a way to differentiate between tourists and Ticos

# Purpose

The point of this project is to use social media posts to find tourists that we can digitally market to while identifying digital nomads that can become marketing partners with an 80% accuracy rate.

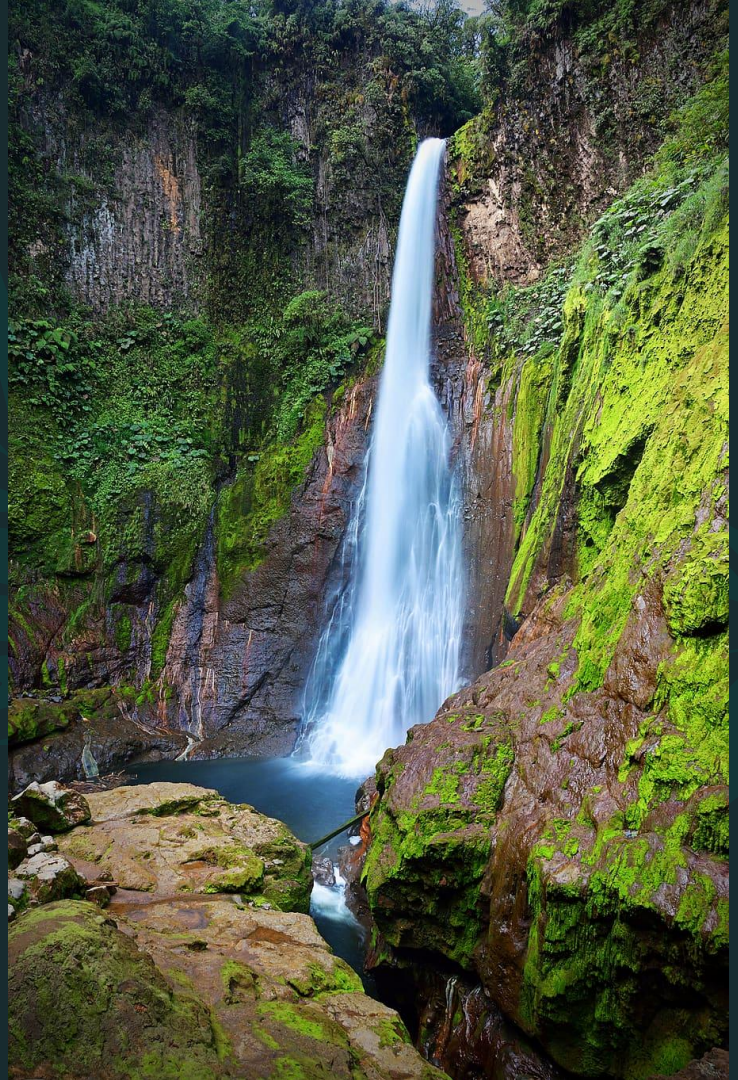
# Procedures

Data was collected from Reddit “CostaRica” and “CostaRicaTravel”

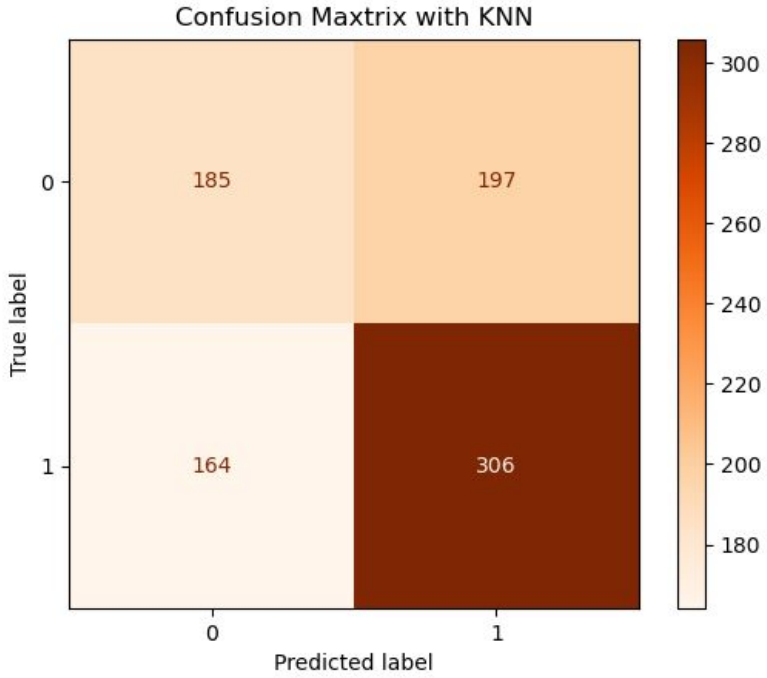
Data was transformed into lowercase, cleaned for doubles, and common compound-words were combined (costa_rica, san_jose, and la_fortuna)

Data was put through several ML models to find the highest succes in telling a Tico from a Tourist

Base score is 57%

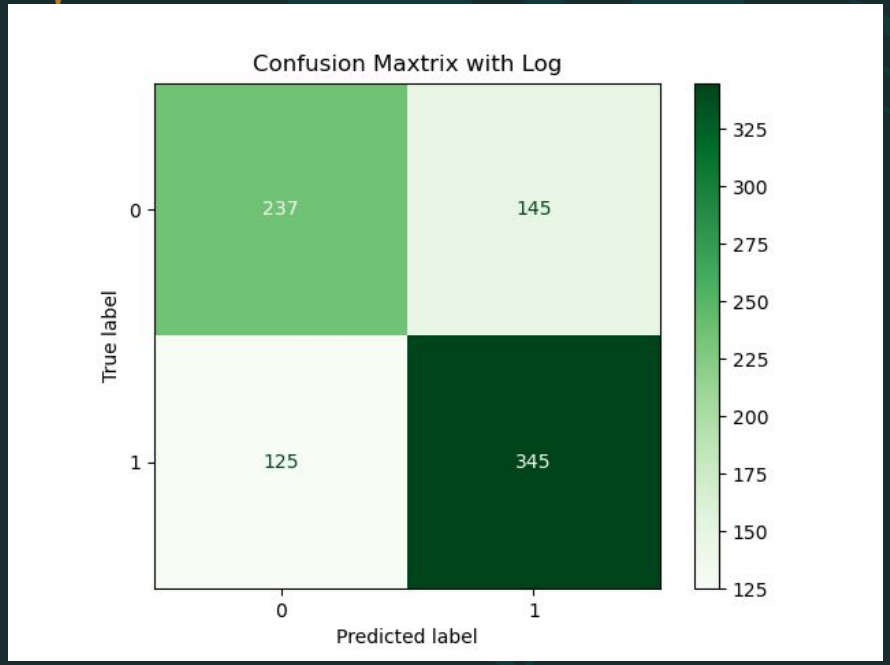


# KNN



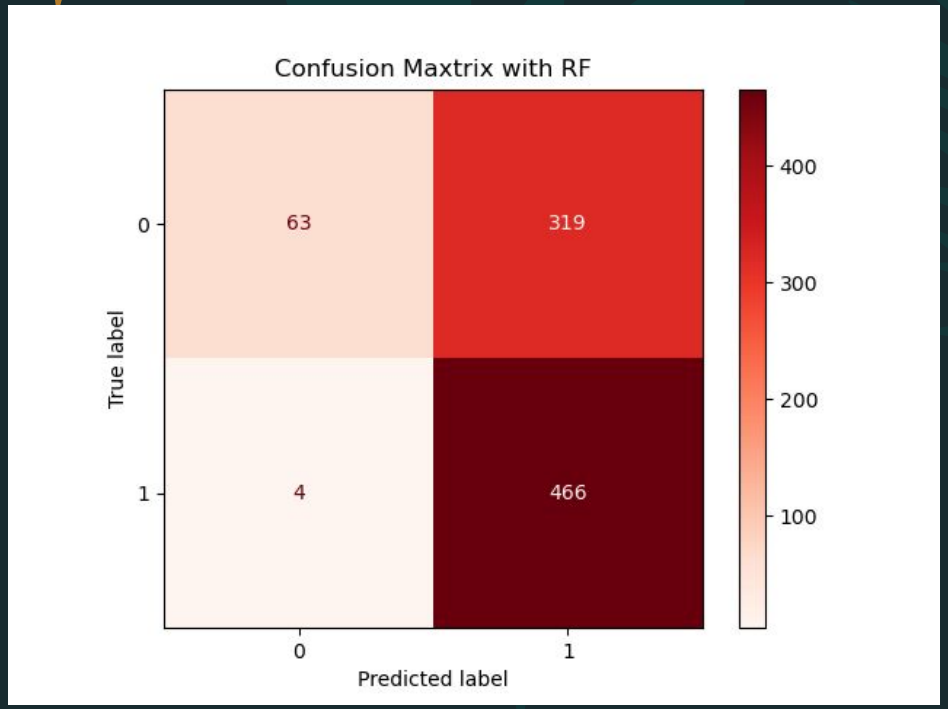

- Parameters Examined
- Stop words: none, English, Custom, Custom + English
- Preprocess: none, Lemmatize, and Stem
- Ngram: 1:1, 1:2, 1:3, 2:2
- K Neighbors: 3, 5, 7, 10
- Test Score of 56.69%

# Log



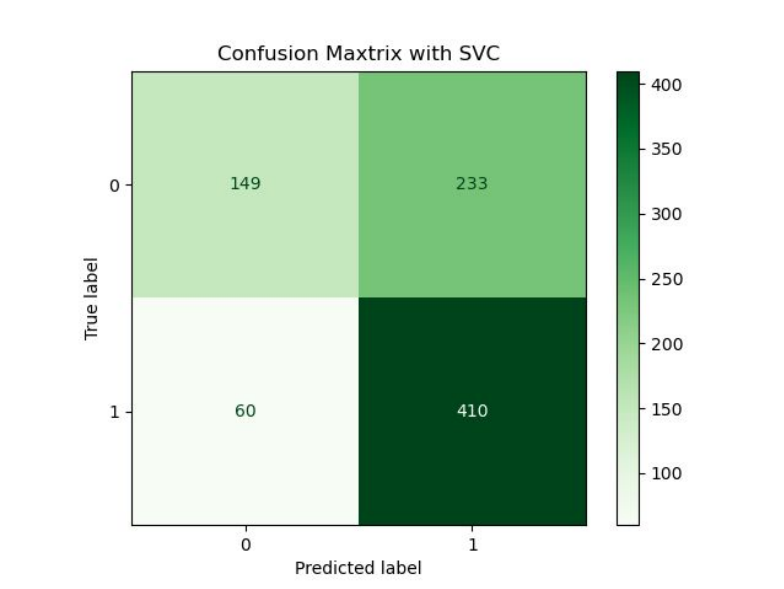

- Parameters Examined
- Stop words: none, English, Custom, Custom + English
- Preprocess: none, Lemmatize, and Stem
- Ngram: 1:1, 1:2, 2:2
- Lasso, Ridge
- Test Score of 68.30

# Random Forest



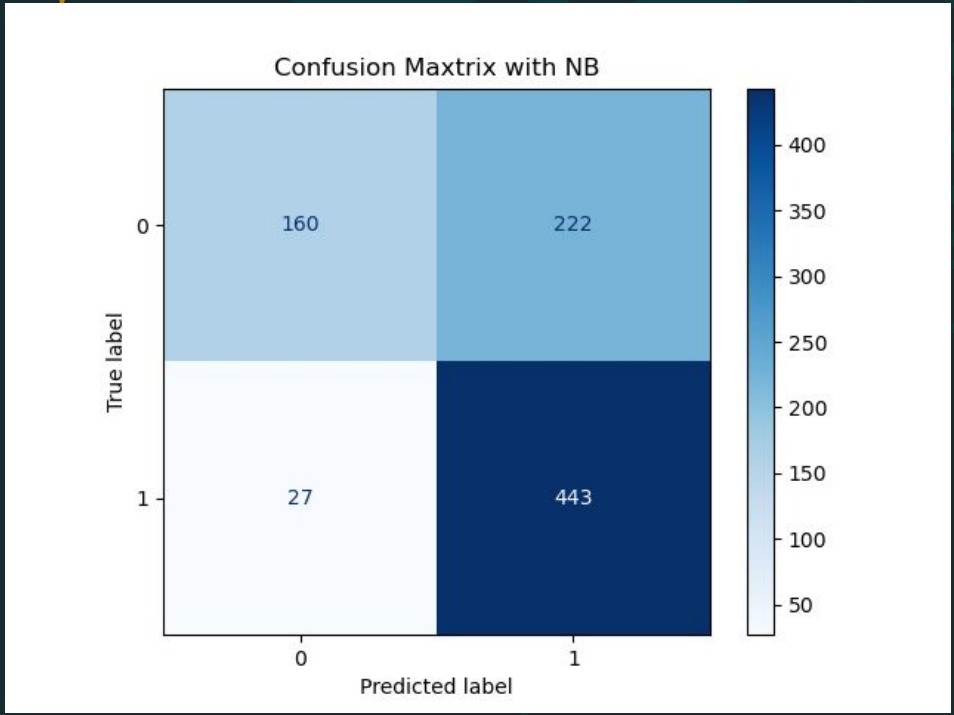

- Parameters Examined
- Stop words: none, English, Custom
- Preprocess: none, Lemmatize, and Stem
- Ngram: 1:1, 1:2, 1:3, 2:2
- Max Depth: 4, 7, 10
- Bootstrap: True, False
- Test Score of 62.08%

# SVC

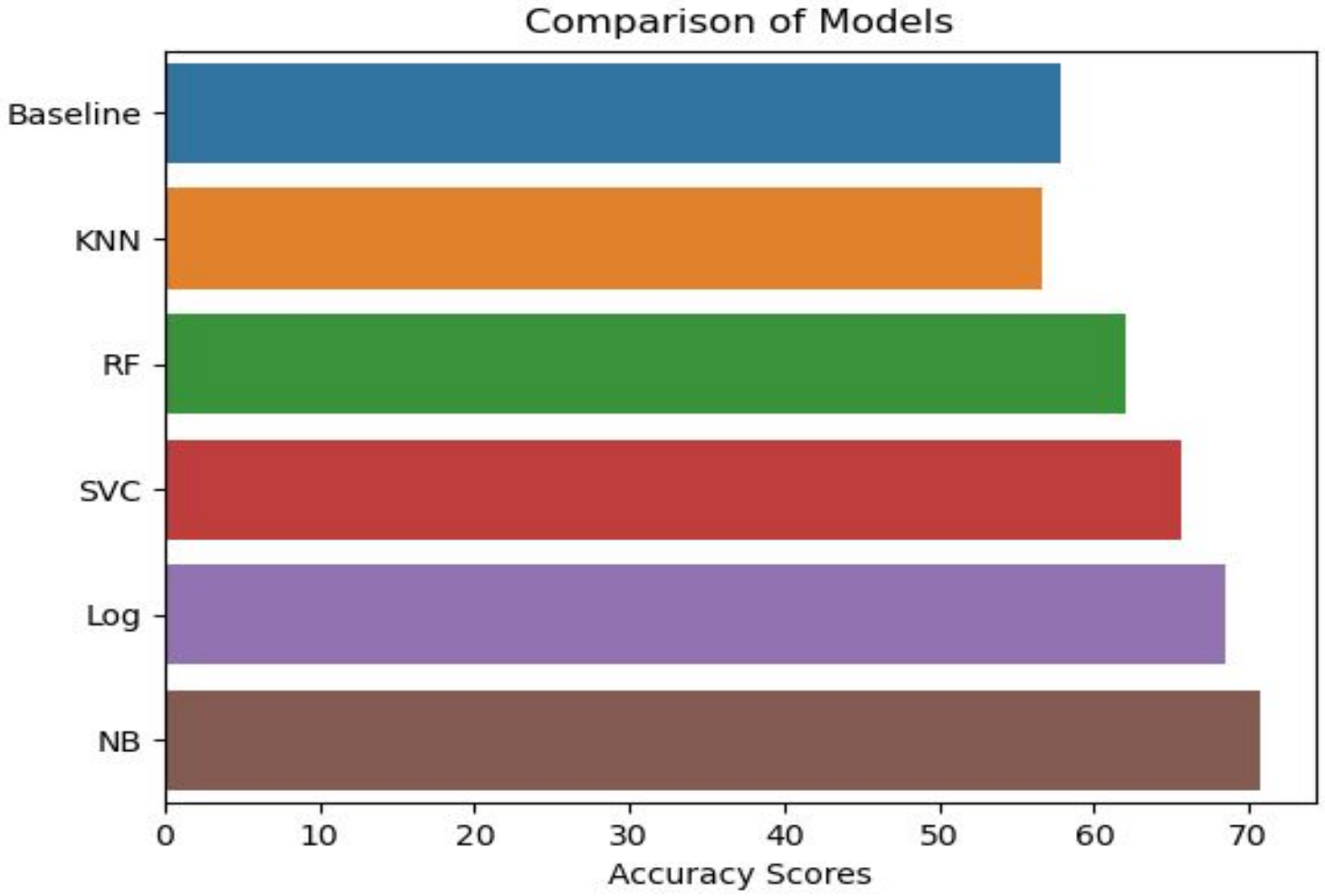



- Parameters Examined
- Stop words: none, English, and Custom
- Preprocess: none, Lemmatize, and Stem
- Ngram: 1:1, 1:2, 1:3, 2:2
- C: 25, 50, 100
- Test Score of 65.61%

# Naive Bayes




- Parameters Examined
- Stop words: none, English, Custom, Custom + English
- Preprocess: none, Lemmatize, and Stem
- Ngram: 1:1, 1:2, 1:3, 2:2
- Alpha: .5, .75, 1
- Test Score of 70.77%


Comparison of Models
Baseline
KNN
RF
SVC
Log
NB
0
10
20
30
40
50
60
70
Accuracy Scores

# Purpose

The point of this project is to use social media posts to find tourists that we can digitally market to while identifying digital nomads that can become marketing partners with an 80% accuracy rate.

# Recommendations

I recommend:

Start using the NB model to find Tico/Expat posts on social media.

Once the Reddit Civil War is over, we collect more data to refine the model.

# Thanks!

## Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution