Climate Change Belief Analysis


Introduction
Problem Statement
Exploratory Data Analysis
Data Engineering
Model Exploration
Conclusion

1. Introduction

- Climate change is a complex problem with impacts and interactions at global to local scales.

2. Problem Statement

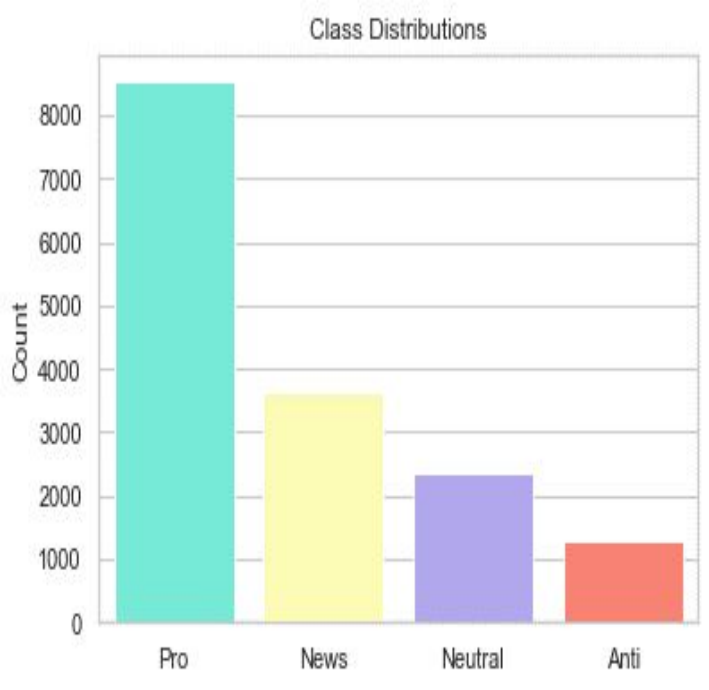
- Humans are striving for environmental friendliness and sustainability.
- Beliefs in climate change vary significantly.
- Develop a machine learning model to categorize human beliefs on climate change based on their unique tweet data.

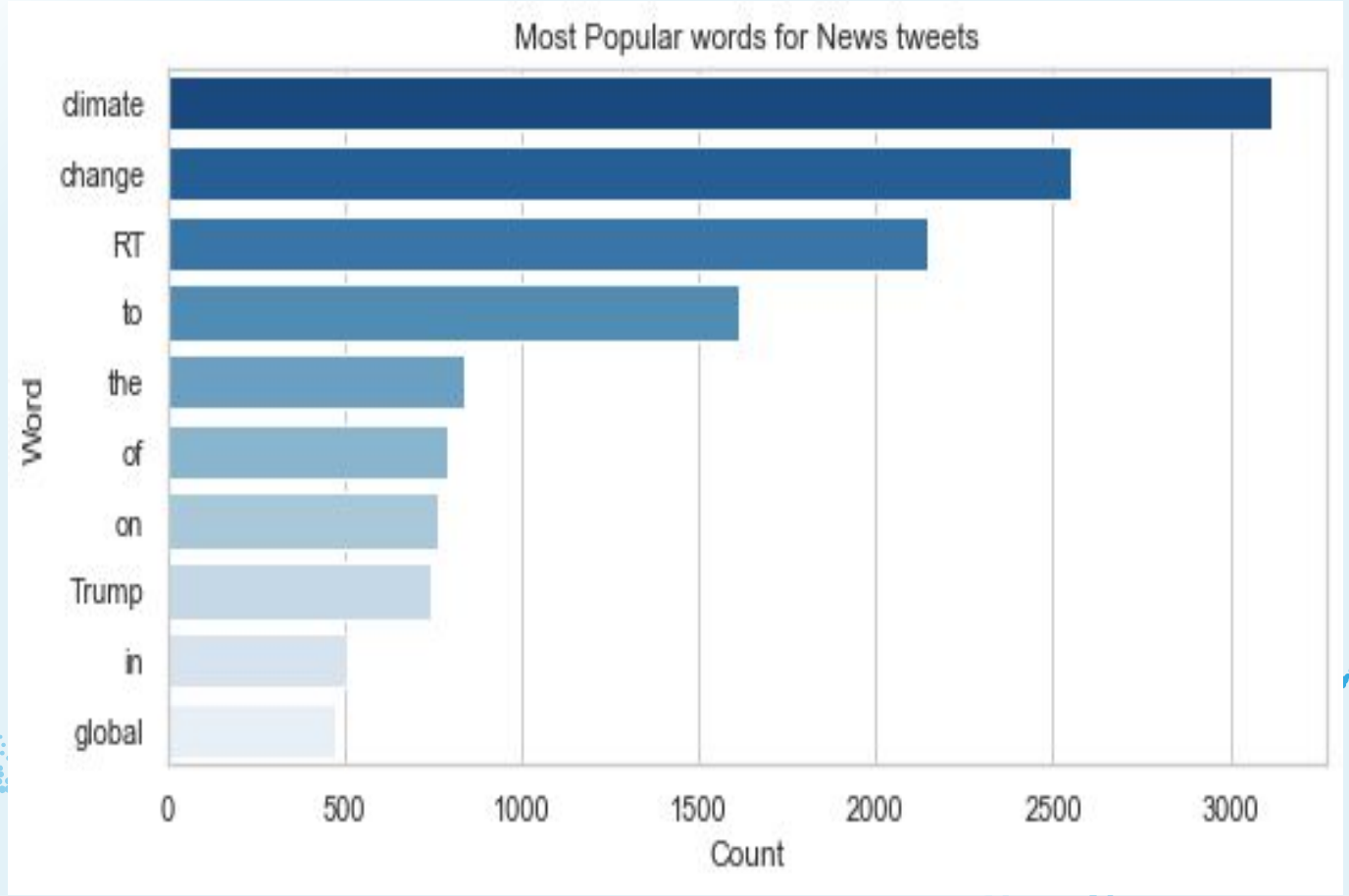
3. Exploratory Data Analysis

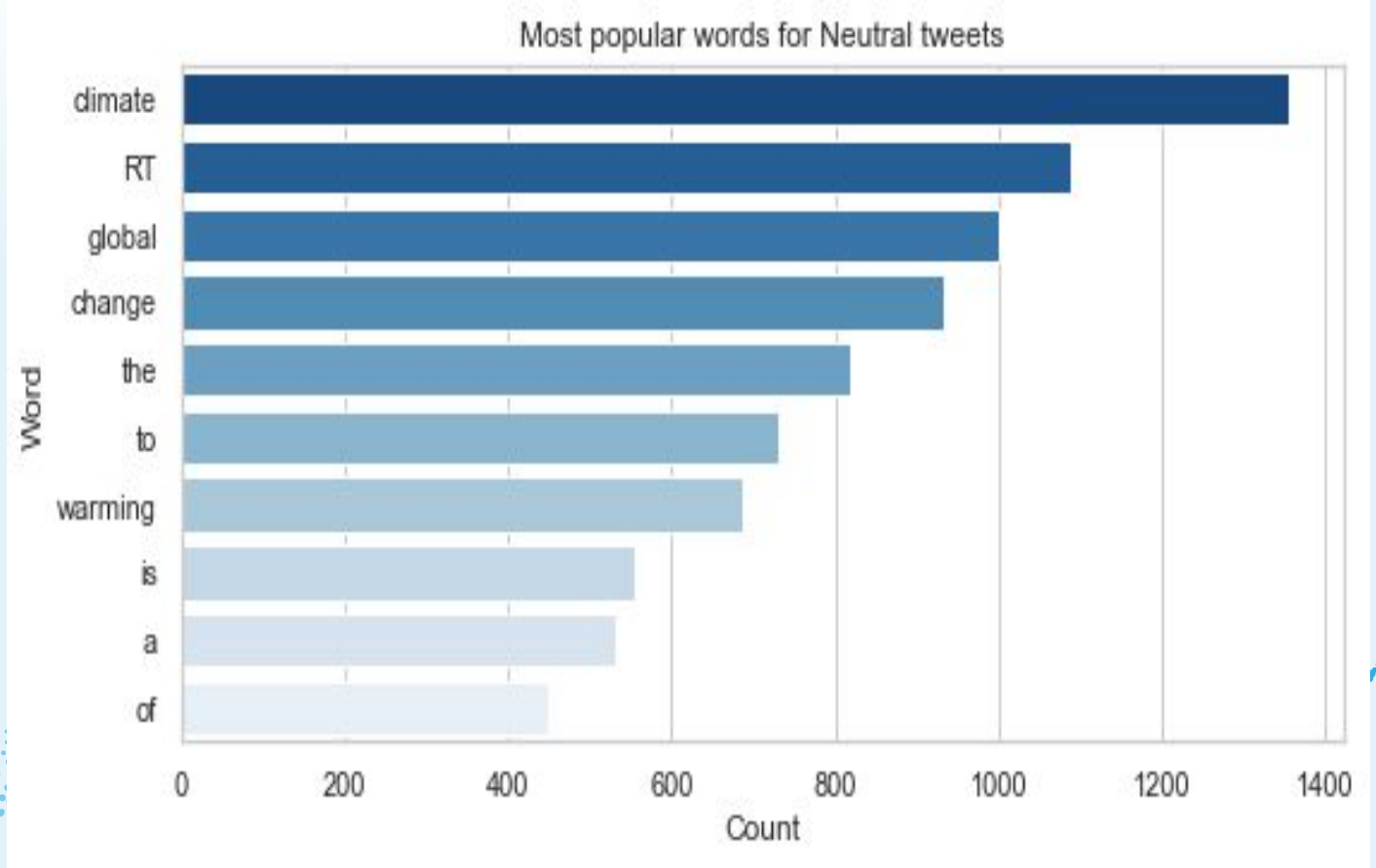
- EDA was crucial to performing initial investigations on the dataset.
- This exposed patterns, anomalies and generalized assumptions on the dataset.

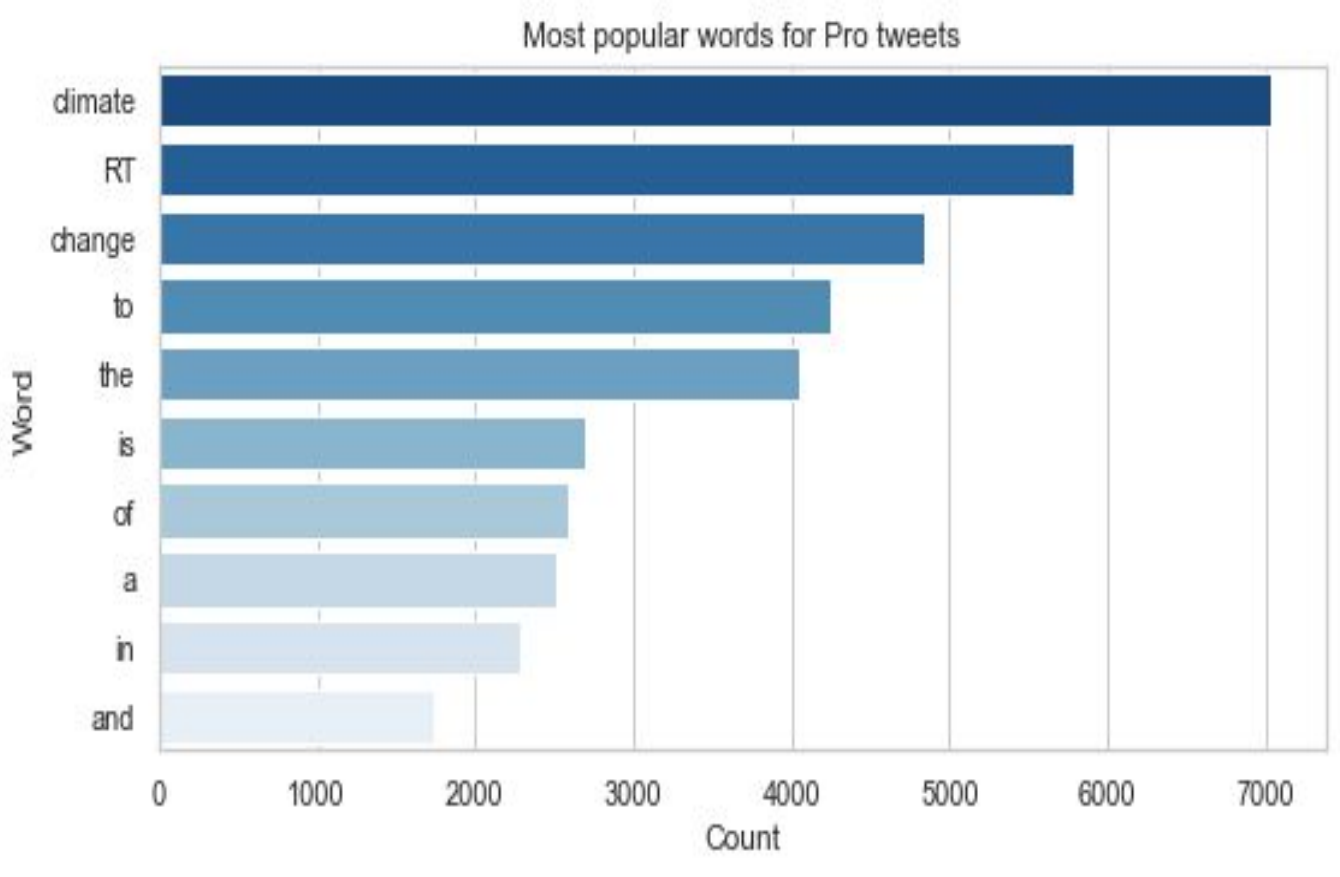
Class Distributions and Sentiment Analysis

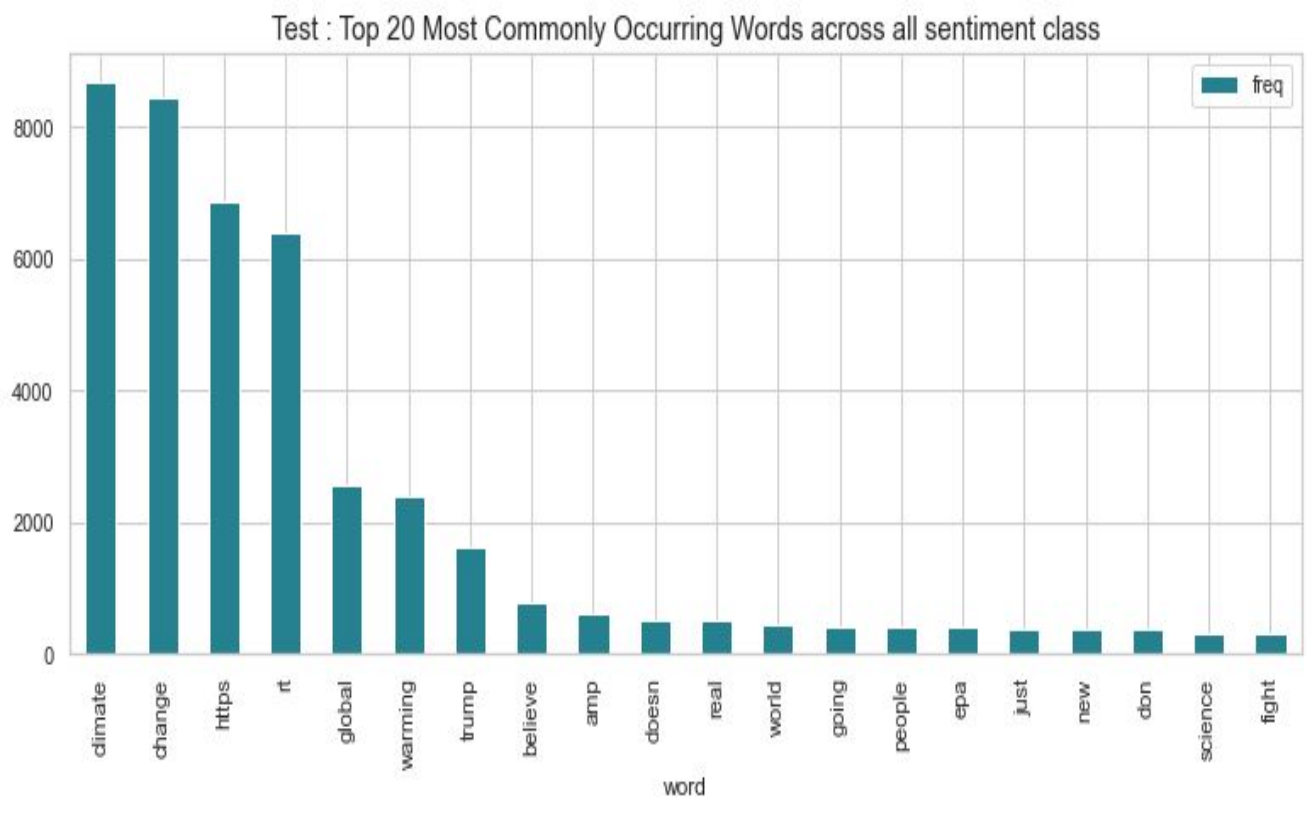
Sentiment	Value Count	%
1(pro)	8530	53.92
2(news)	3640	23.01
0(neutral)	2353	14.87
-1(anti)	1296	8.19



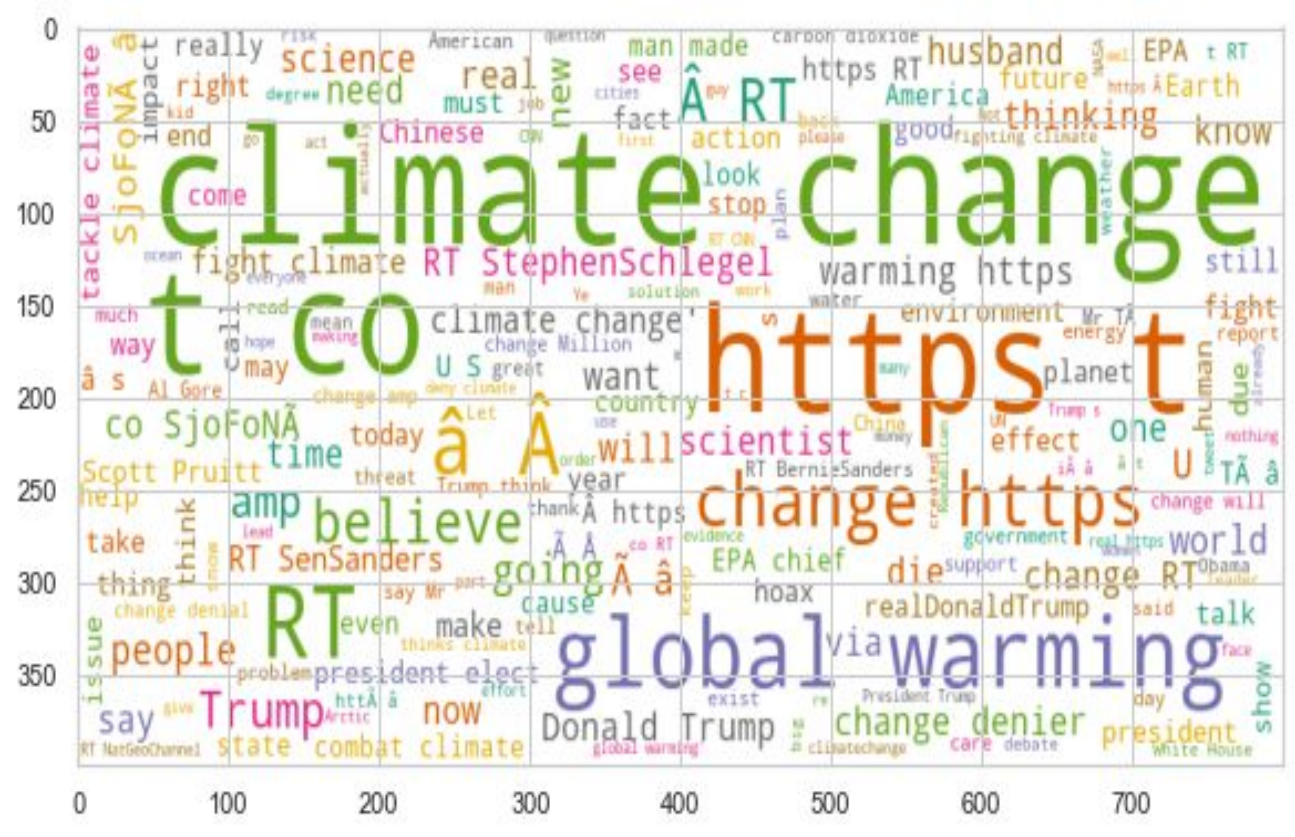

Most Popular words for News tweets
dimate
change
RT
to
the
of
on
Trump
in
global
Word
0
500
1000
1500
2000
2500
3000
Count


Most popular words for Neutral tweets
climate
RT
global
change
the
to
warming
is
a
of
Word
0
200
400
600
800
1000
1200
1400
Count

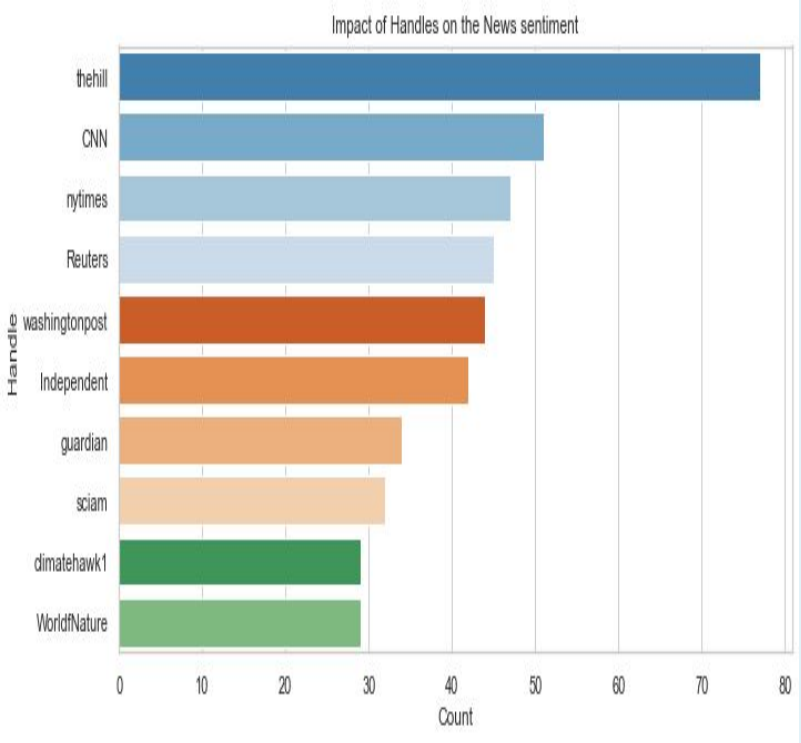

Most popular words for Pro tweets
dimate
RT
change
to
the
is
of
a
in
and
Word
0
1000
2000
3000
4000
5000
6000
7000
Count

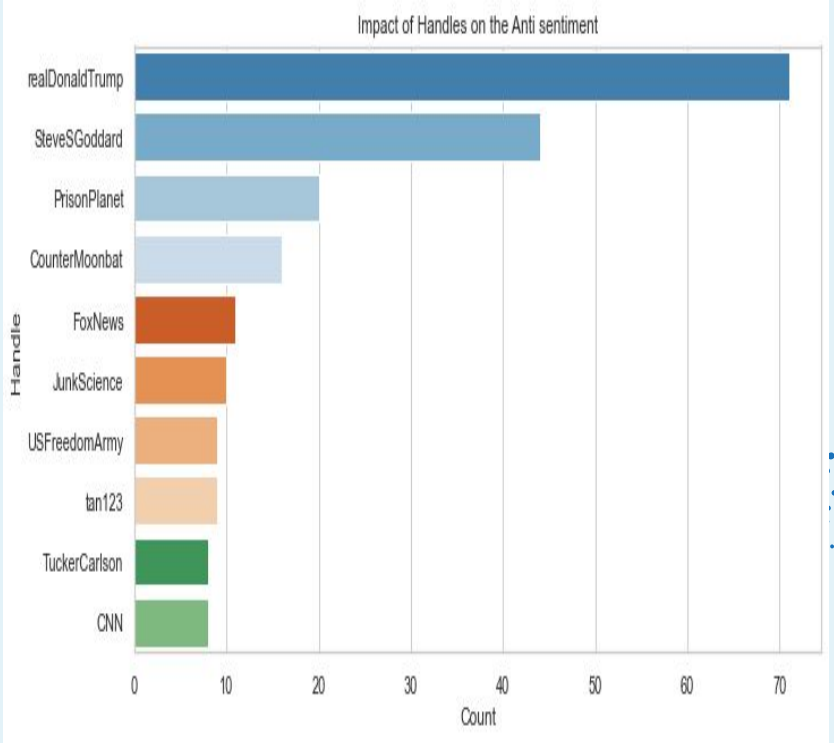

Test : Top 20 Most Commonly Occurring Words across all sentiment class
freq
8000
6000
4000
2000
0
climate
change
https
rt
global
warming
trump
believe
amp
doesn
real
world
going
people
epa
just
new
don
science
fight
word

Sentimental Buzzwords



The Influence of Twitter Handles on Sentiment Classes





Summary Overview

- Climate Change and Global Warming are the most popular words in all four classes.
- Noise words, such as http, https, website, co, and RT, reoccurred
- Same top five words in each class. An exception of the News class, where the word ‘Trump’ appears prominently.

4. Data Engineering

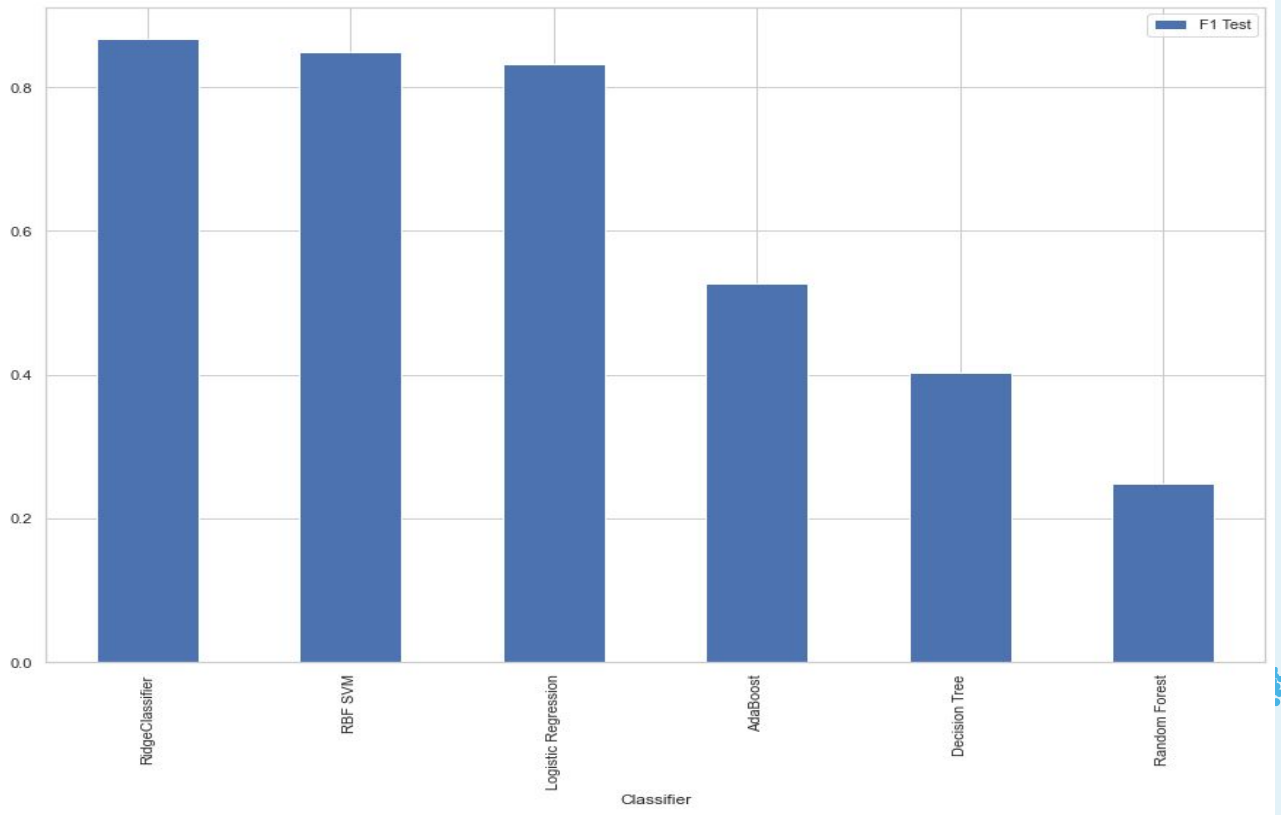
The process involved NLP techniques on;

i. Noise Removal
ii. Tokenization
iii. Removing Stop words
iv. Stemming
v. Lemmatization
vi. Duplicate check

5. Model Exploration

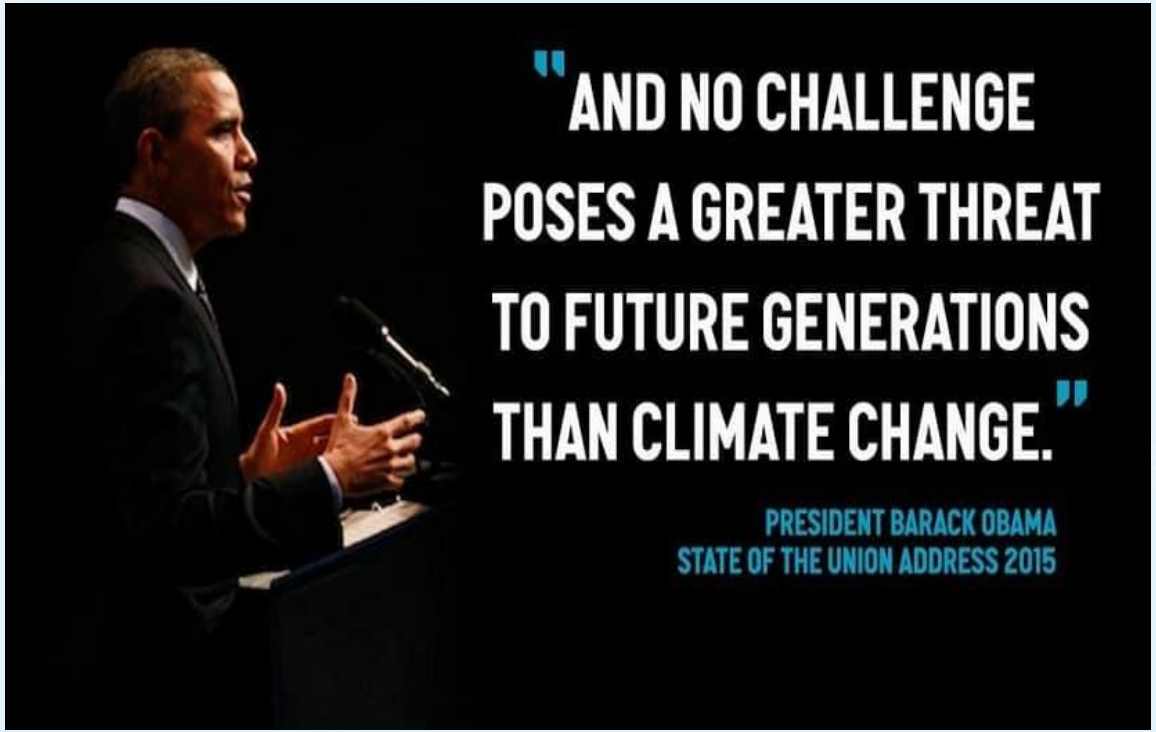
- Built a Pipeline for the preprocessor and selected a base model.
- The following slide shows how different models performed.

F1 Test Scores



Conclusion and The Big concept

- Human sentiments take a variety of demographic and geographic categories.
- Our model stands out in precision and diversity on insights that will guide future decisions on climate change.


"AND NO CHALLENGE
POSES A GREATER THREAT
TO FUTURE GENERATIONS
THAN CLIMATE CHANGE."
PRESIDENT BARACK OBAMA
STATE OF THE UNION ADDRESS 2015

Thank You!

Any questions?