# Classifying Climate Change Tweets

Michael Wirtz

# Overview

Business Problem

Building the Classifier

Applying the Classifier

Conclusion and Next Steps

# Business Problem

- Environmental Defense Fund
- Addressing advertising and promotions expense growth
- Where and when to deploy for most donations

ENVIRONMENTAL
DEFENSE FUND™

# Building the Classifier

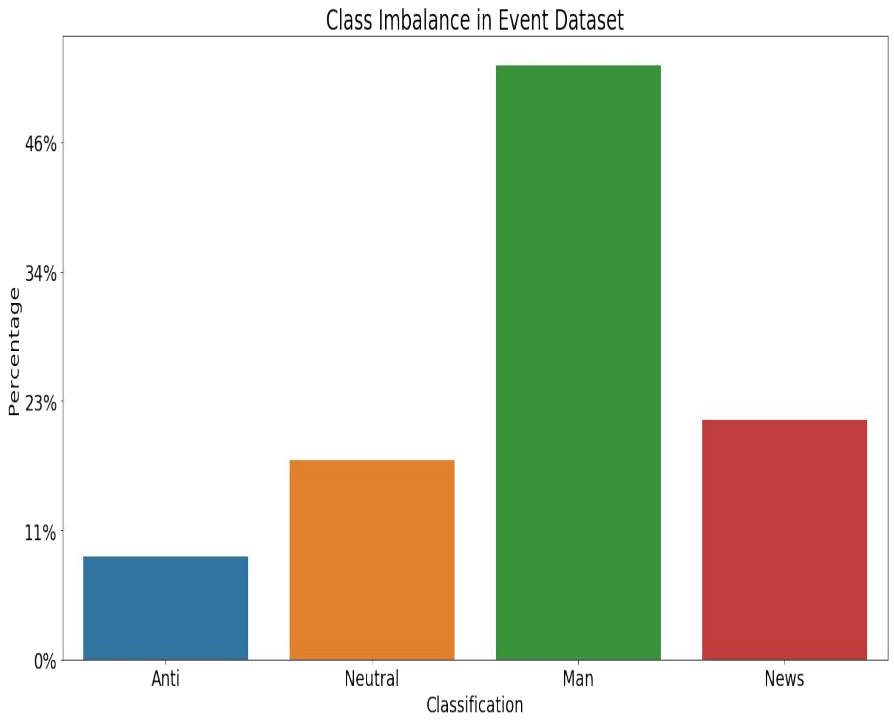
# The Data

- Kaggle
- 43943 Tweets
- Apr 2015 - Feb 2018
- Four classes
  - Anti Man Made
  - Neutral
  - Man-Made
  - News

# Class Imbalance

## Metric: F1 Score

## Focus on optimizing 'Anti' class f1 score



# Modeling Process

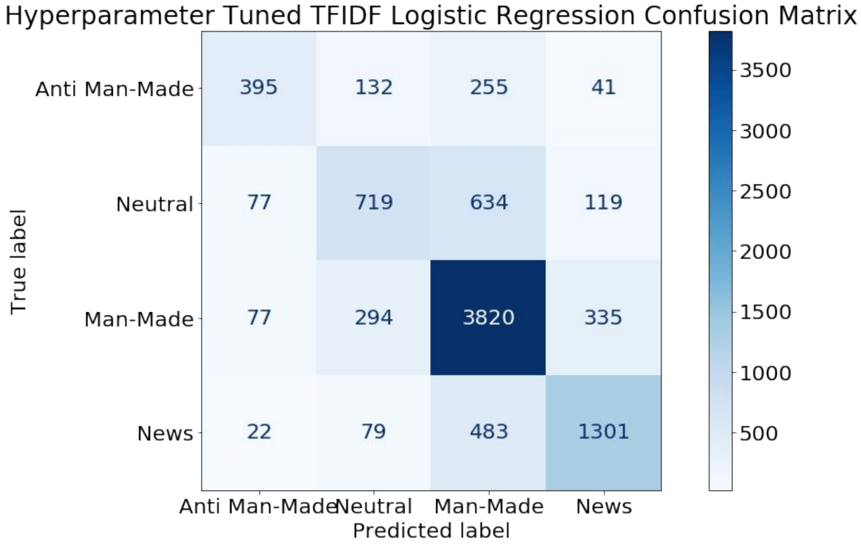
1. Baseline
2. Choosing Best Model
3. Tuning Best Model

# Best Model

## Unigram TF-IDF Logistic Regression (Without Added Features)

### F1-Scores

- Man: 0.79
- Anti: 0.57



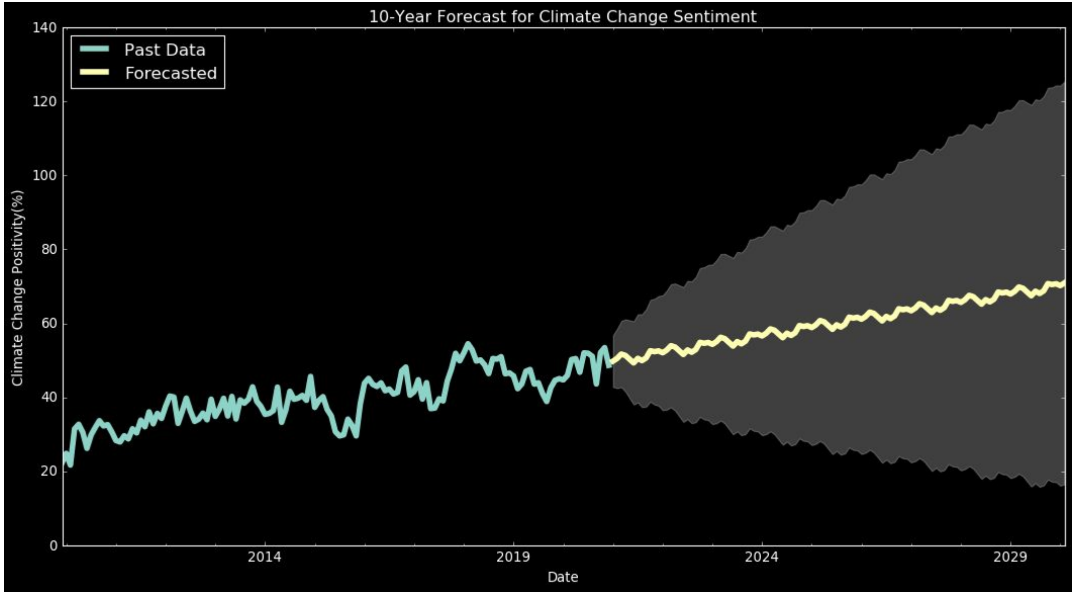
# Applying the Classifier

# Time Series Analysis

The Assumption

The Data

Analysis



# Time Series Findings

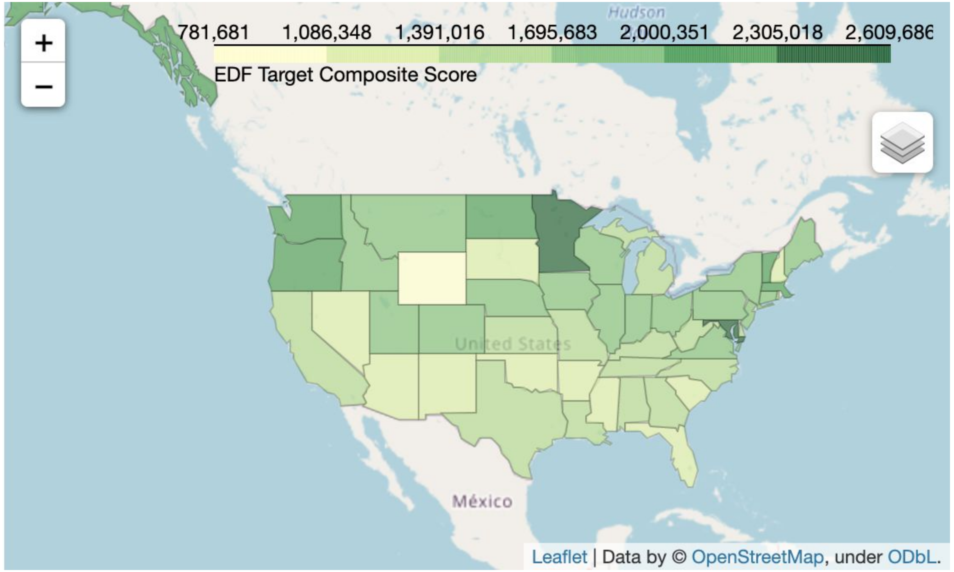
1. Donation growth rate of 3.8% year over year
2. Monthly Breakdown:
    a. Sentiment average vs season
    b. Best month (March)
    c. Worst month (August)

# Geographic Analysis

The Assumption

The Data

Analysis



# Geographic Findings

Top 5 most likely states for climate change donations:

1. Minnesota
2. Maryland
3. Oregon
4. North Dakota
5. Washington

# Recommendations

- EDF
  - Cold months (notably March)
  - Top 5 states
- Environmentally-focused NGO with low budget

# Next Steps

- Custom scoring metric
- Pipeline for auto updating results
- Location data by county to pinpoint best areas

# QUESTIONS?

## For More Information

Github: mwirtz946
Email: michealwirtz88@gmail.com