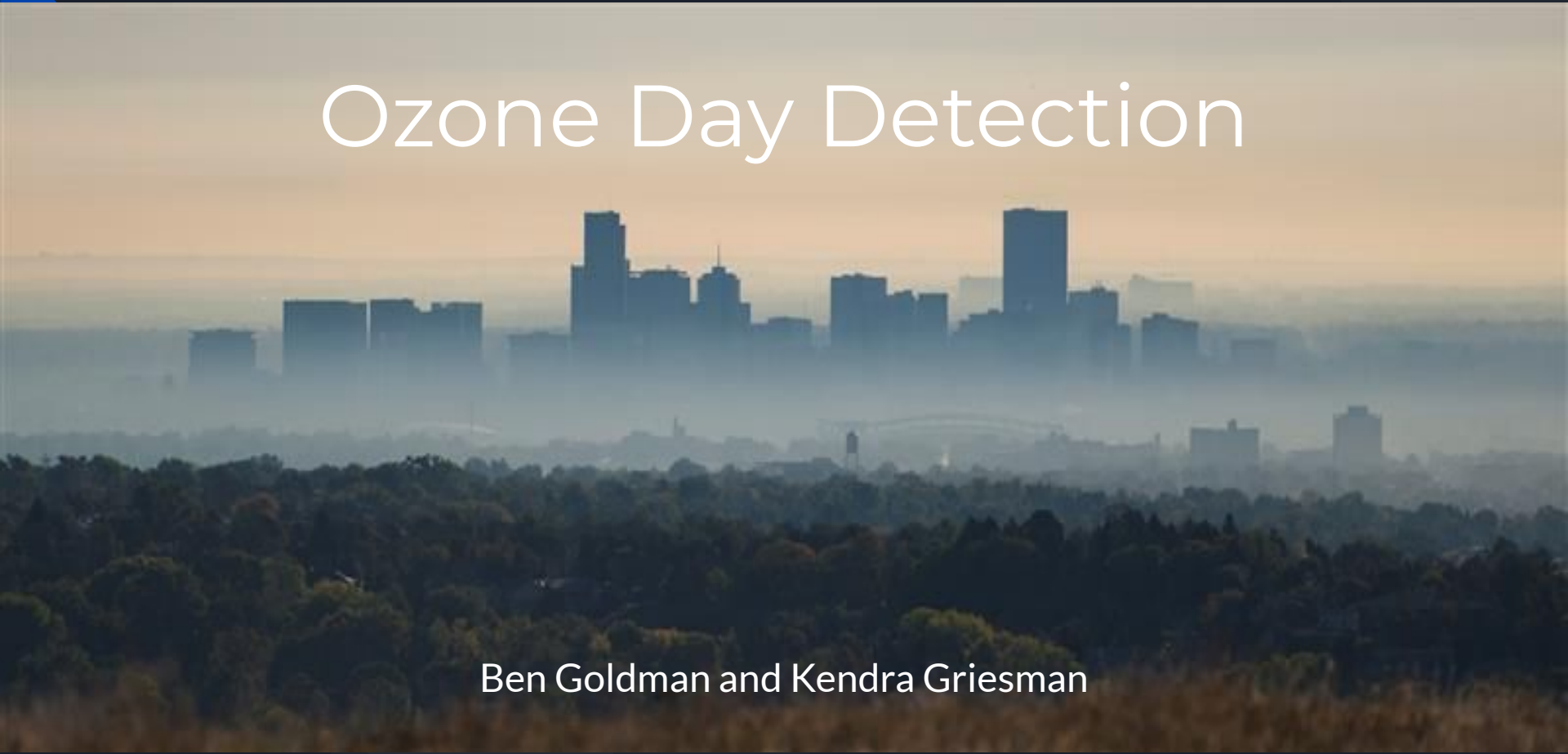
Ozone Day Detection

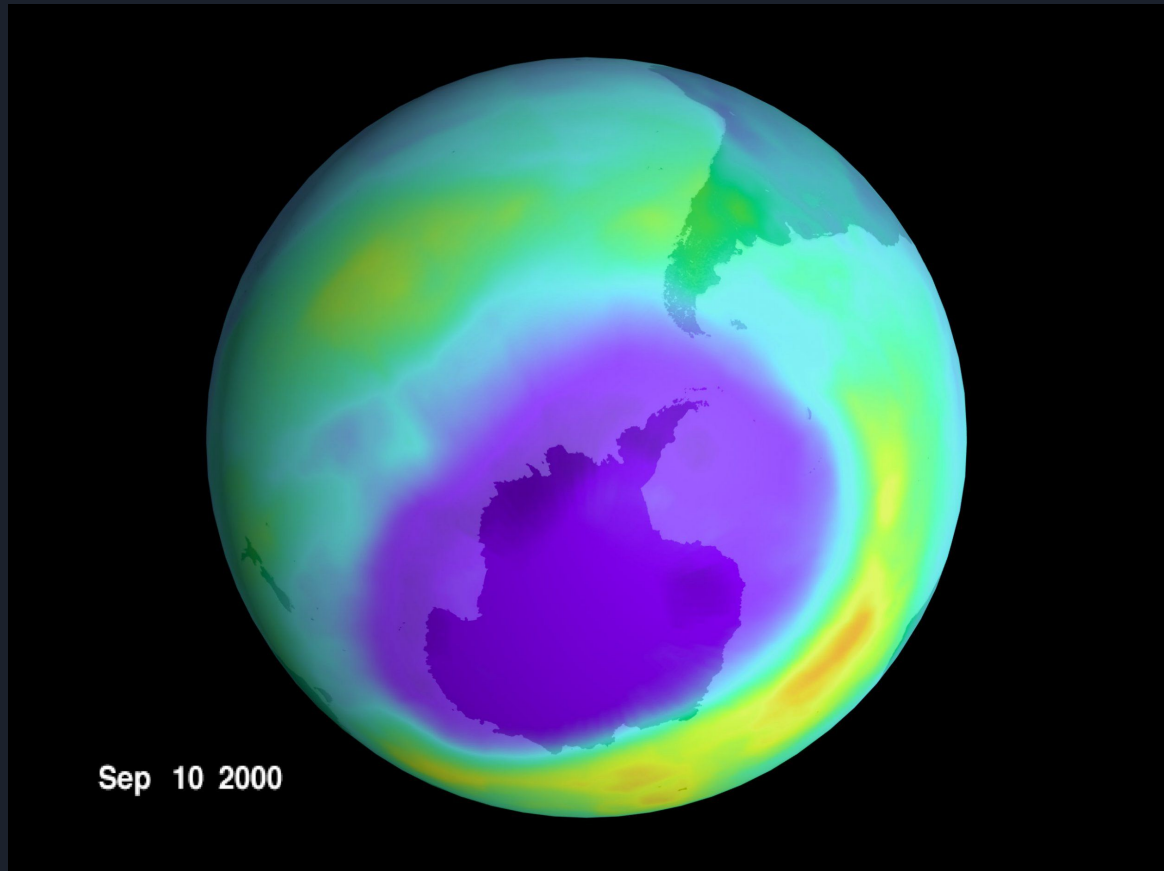


Ben Goldman and Kendra Griesman

https://research.noaa.gov/Portals/0/EasyDNNnews/226/12088smog-blankets-boulder-nrel.jpg

What is an Ozone Day?

Ozone Action Days are declared when weather conditions are likely to combine with pollution emissions to form high levels of ozone near the ground that may cause harmful health effects



https://d.newsweek.com/en/full/771913/gettyimages-665296.jpg

Our Data

- Ground level ozone data
- Collected near Houston, Texas from 1998-2004
- 73 continuous features
  - Peak Temperature
  - Relative Humidity
  - Wind Speed

https://www.bestplaces.net/images/city/houston_tx.gif

Methods

- KNN
- SVM
- Adaboost

Challenges with our Data

- Challenges
  - Heavily Unbalanced
    - 2-5% labeled as Ozone Day
  - Lack of Data
    - 2536 Data points
  - Missing values
- Solutions
  - UnderSampling/OverSampling
  - Threshold Alterations
  - SKlearn Package

KNN

Score without oversampling: 98.2%

5 Neighbors or more

predicted

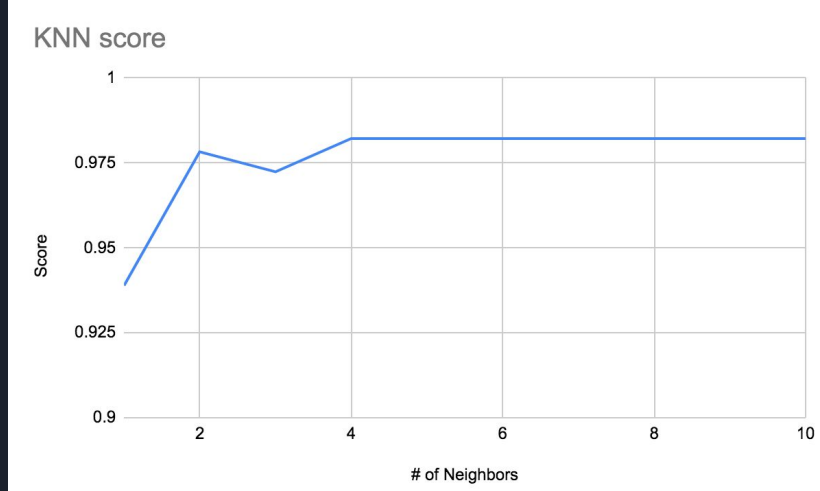
true	499	0
	9	0

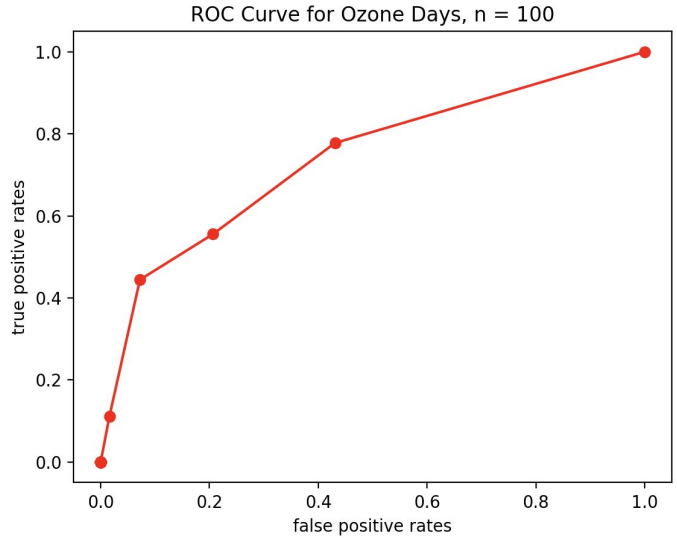
Score with oversampling: 67.7%

40 Neighbors

predicted

true	338	161
	3	6





Adaboost

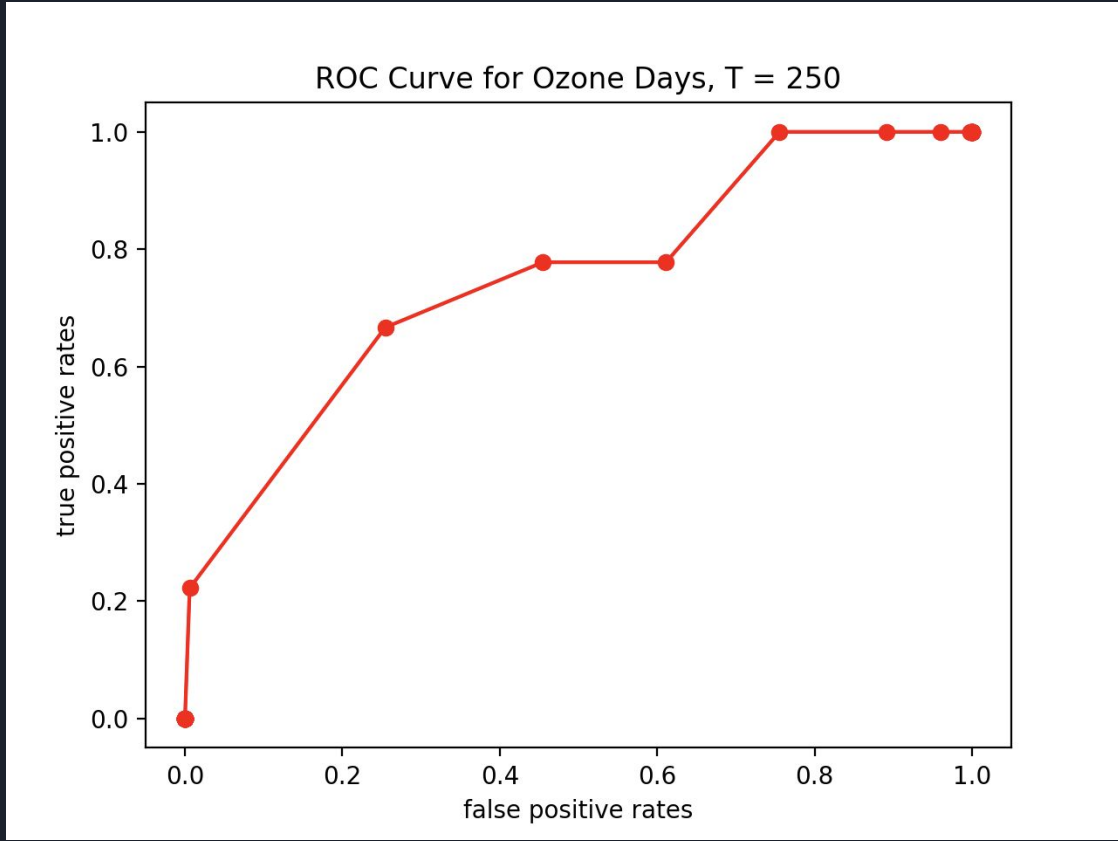
Score: 98.0%

- 250 estimators
- Threshold 0.5

	predicted	
true	496	3
	7	2

Best Features

- East-West direction wind at 1457m
- Relative Humidity at 5000m



Adaboost

Score: 80.7%

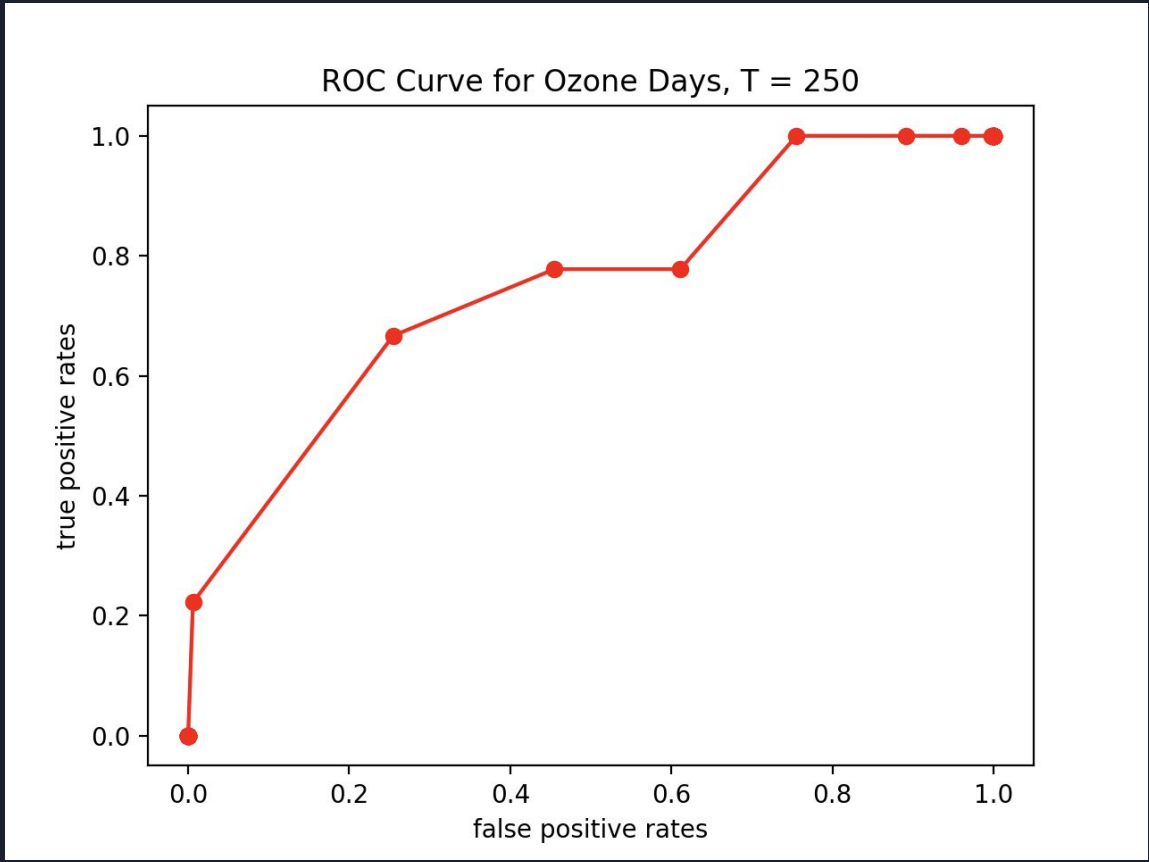
- 250 estimators
- Threshold 0.48

predicted

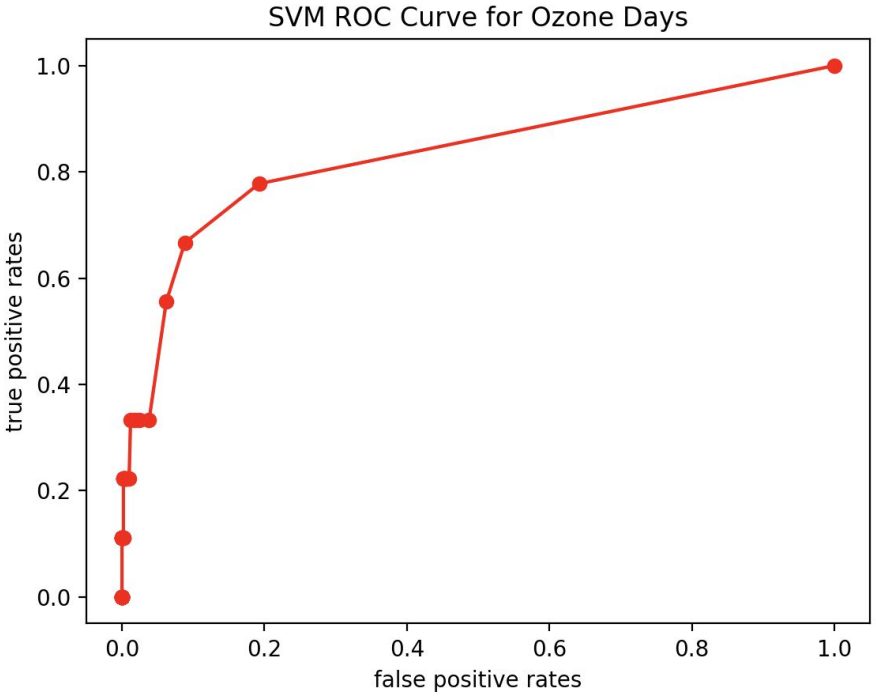
true		
	404	95
	3	6

Best Features

- East-West direction wind at 1457m
- Relative Humidity at 5000m



SVM



Score: 98.4%

- Linear kernel

	predicted	
true	498	1
	7	2

Best features:

- Relative humidity at 5000 m
- East-West direction of wind

Conclusions and Future Work

- Need more data
  - Potentially wider scope
- Adaboost Correctly labeled the most of the minority class
- SVM preformed best overall
- Continue work with over and under sampling
- Try other Methods
- Machine Learning is very accessible!