Using Machine Learning Techniques to Improve Rainfall Prediction

Fnu Tulha

Besant Nandra

Overview of Project

- **Basic Premise:** Using machine learning techniques to accurately predict rainfall
- Analyze various atmospheric features using kNN and SVM
- Perform PCA dimensionality reduction to identify most significant features
  - Extension to previous work, in which feature selection wasn't implemented
- Discuss Results

Dataset

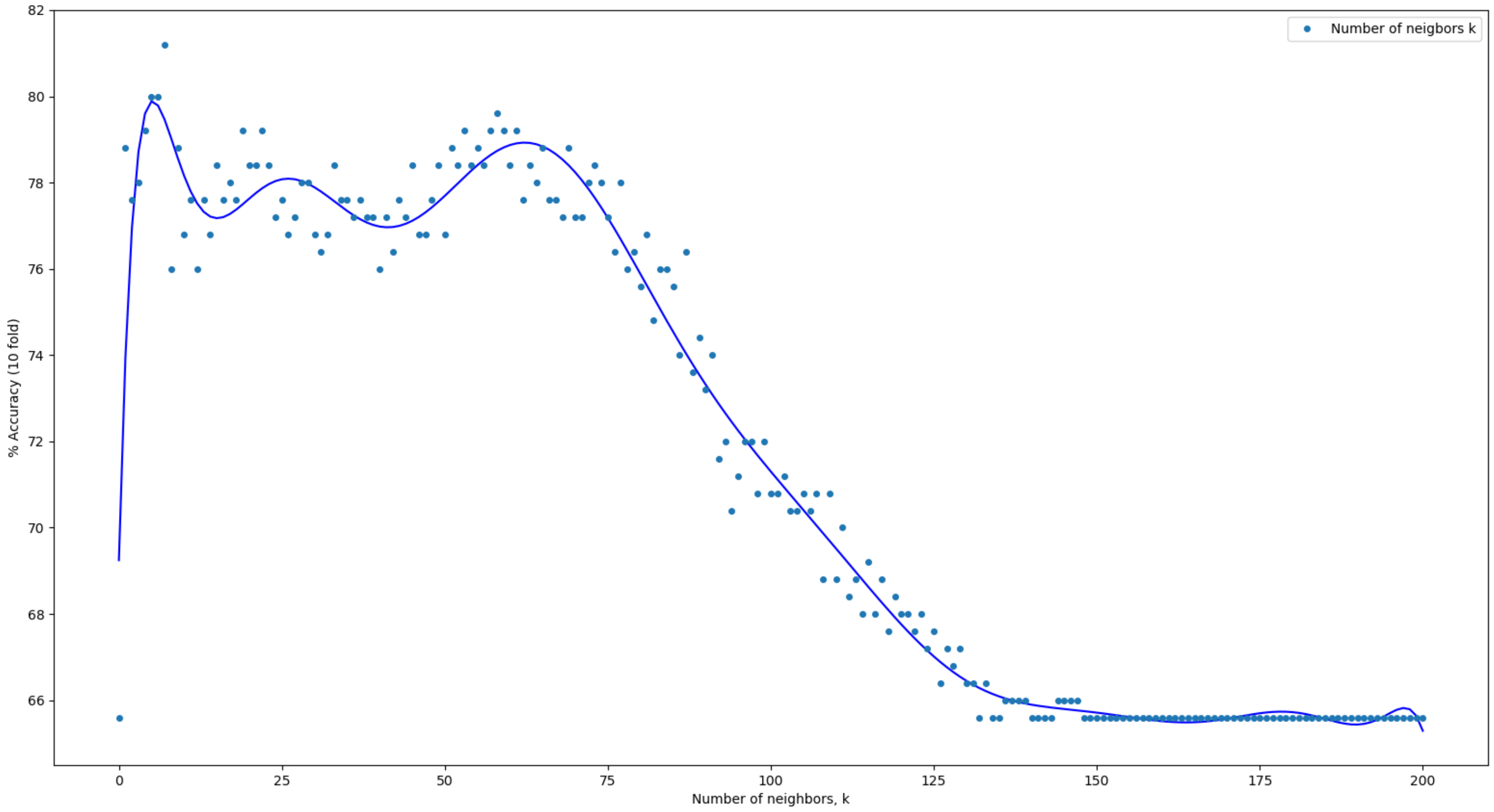
- GFS (Global Forecast System)
  - Atmospheric Features
- Weather Underground
  - Labels
- Manual collection of data
  - No batch download

Dataset

```
time[0]=0 lat[0]=28.5 lon[0]=278.5 Cloud_water_entire_atmosphere_single_layer[0]=0.04 kg.m-2
time[0]=0 lat[0]=28.5 lon[0]=278.5 Haines_index_surface[0]=2 (no units)
time[0]=0 lat[0]=28.5 lon[0]=278.5 Precipitable_water_entire_atmosphere_single_layer[0]=43.2 kg.m-2
time[0]=0 lat[0]=28.5 lon[0]=278.5 Pressure_maximum_wind[0]=16389 Pa
time[0]=0 lat[0]=28.5 lon[0]=278.5 Pressure_surface[0]=101711 Pa
time[0]=0 lat[0]=28.5 lon[0]=278.5 Relative_humidity_entire_atmosphere_single_layer[0]=61 %
time[0]=0 lat[0]=28.5 lon[0]=278.5 Temperature_maximum_wind[0]=210 K
time[0]=0 lat[0]=28.5 lon[0]=278.5 Temperature_surface[0]=289.2 K
time[0]=0 lat[0]=28.5 lon[0]=278.5 Total_cloud_cover_convective_cloud[0]=0 %
time[0]=0 lat[0]=28.5 lon[0]=278.5 Total_ozone_entire_atmosphere_single_layer[0]=246 DU
lat[0]=28.5 degrees_north
lon[0]=278.5 degrees_east
time[0]=0 Hour since 2014-01-02T00:00:00Z
time[0]=0 lat[0]=28.5 lon[0]=278.5 u-component_of_wind_maximum_wind[0]=38.9 m/s
time[0]=0 lat[0]=28.5 lon[0]=278.5 v-component_of_wind_maximum_wind[0]=9.2 m/s
```

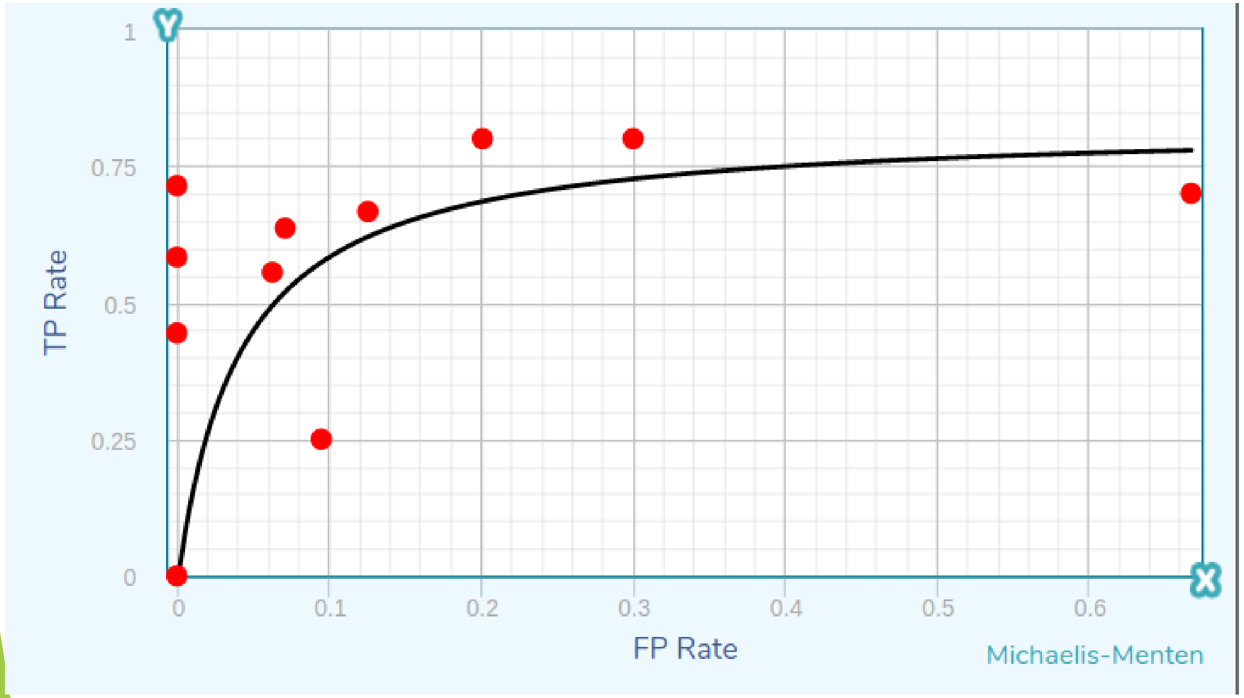
Results

Best K for kNN

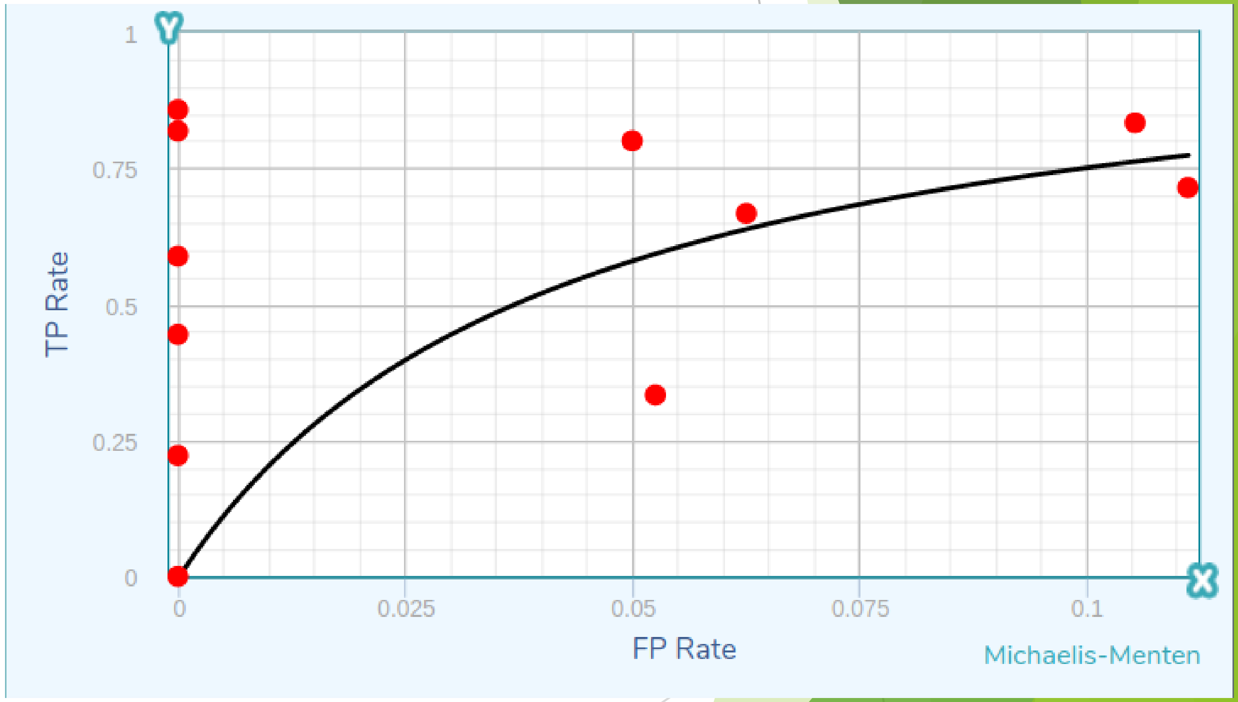


Results

kNN ROC Curve

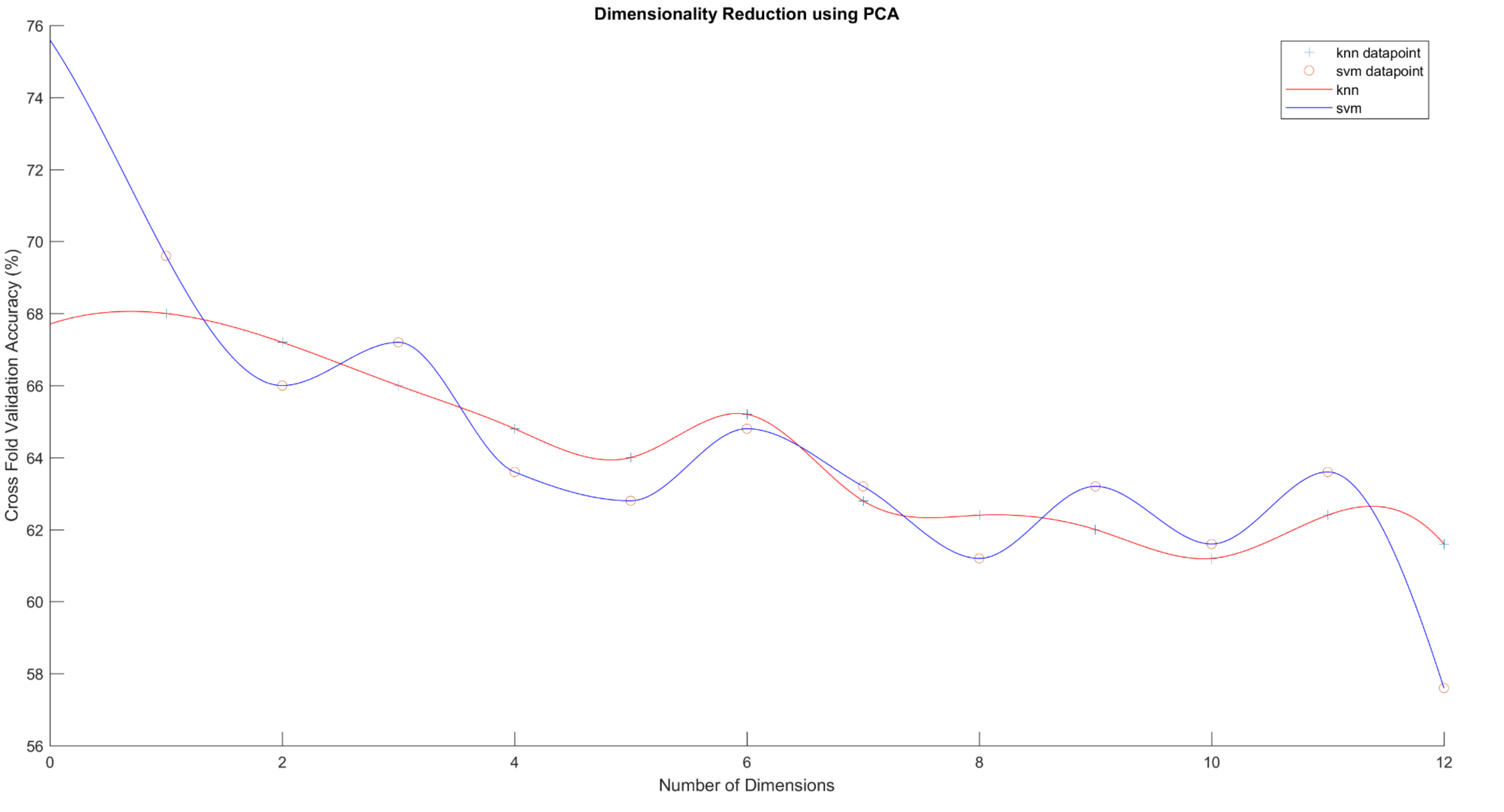


SVM ROC Curve



Results

PCA Feature Subset Comparison



Analysis and Implications

- SVM outperformed kNN on full feature set
- On feature subsets, SVM and kNN performed similarly
- Large difference in variance among features
  - Indicates noisy, irrelevant features
  - Noisy features harm accuracy and efficiency
- Implications for testing more features and identifying a solid subset of high-variance, impactful features for predicting rain

References

- Coblenz, Joshua. Using Machine Learning Techniques to Improve Precipitation Forecasting. (2015).
- NOAA GFS Data Access, http://www.nco.ncep.noaa.gov/pmb/products/gfs/
- Weather Underground Data Access, https://www.wunderground.com/history/
- Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
- NetCDF Operator, Charlie Zender et al., 1998. http://nco.sourceforge.net/