Matthew Thompson

09/04/20

CS 287

Project Planning

Sources:

Mount Washington Avalanche Center:

* <https://mountwashingtonavalanchecenter.org/hermit-lake-snow-plot-data-all/>
* <https://mountwashingtonavalanchecenter.org/weather-avalanche-log/>
* <https://mountwashingtonavalanchecenter.org/avalanche-forecast-archives-2019-2020/>

NOAA:

* <https://www.wrh.noaa.gov/mesowest/getobext.php?sid=HLSN3&num=72>

Mount Washington Observatory:

* <https://www.mountwashington.org/library/pdf/MWN_Data_Availability.pdf>

Open Avalanche Project GitHub:

* <https://github.com/scottcha/OpenAvalancheProject>

Snow Density/SWE:

* <https://www.nrcs.usda.gov/wps/portal/nrcs/detail/or/snow/?cid=nrcs142p2_046155>
* <https://www.avalanche-center.org/Education/glossary/density.php>

Notes:

* Most likely should use Precision Recall Curve because few positive cases (avalanche occurrence)
* Maybe don’t log transform to give snow more weight
* Log transform + normalization?
* All features over past couple days feature engineering and selection
* Research domain info to find feature engineering prospects
* Maybe create model to fill NA values on MWAC snow data
* Use PCA for feature selection
* K-Means Clustering for data exploration
* Drop last row of each season
* Combine danger level 4 and 5

PROJECT IDEAS:

Easy:

* Model that predicts next day avalanche danger via occurrence estimation from danger levels

Medium:

* Model that predicts next day avalanche danger level

Mentioned Models Used:

* SVM (support vector machine) - “The method explored in this study is a Support Vector Machine, a non-linear classifier which is robust to noise, tolerant to overlapping classes, and showing promising performance in many real life high-dimensional classification problems.”
* K-Nearest Neighbors
* Gradient Boosting
* Logistic Regression
* Decision Trees
* Random Forest (use feature\_importances\_)
* Extra Trees

Features to Drop - Cleaning:

* PGTM
* AVY\_CHARACTER
* WET\_DANGER
* DRY\_DANGER
* WET\_LOOSE
* WET\_SLAB
* WIND\_SLAB
* STORM\_SLAB
* CORNICE\_FALL
* PERSISTENT\_SLAB
* DEEP\_SLAB
* DRY\_LOOSE
* GLIDE\_AVALANCHE
* LONG\_SLIDING\_FALL
* PSUN
* PRCP
* PRCP\_ATTRIBUTES
* SNOW\_ATTRIBUTES
* SNWD
* SNWD\_ATTRIBUTES
* TMAX
* TMAX\_ATTRIBUTES
* TMIN
* TMIN\_ATTRIBUTES
* TSUN
* TSUN\_ATTRIBUTES
* WDF5\_ATTRIBUTES
* WSF5
* WSF5\_ATTRIBUTES
* SKY CONDITION
* PRECIP TYPE/RATE
* FORM/SIZE
* HN24\_CM
* HW in Tube (MM)
* H24W in Can (MM)
* DENSITY
* HST
* CURRENT TEMP
* water\_equivalent\_trace
* snow\_fall\_trace
* snow\_depth\_6am\_trace
* Sunshine\_sum
* Skycover\_sum
* skycover\_avg\_sunrisetosunset
* year\_y
* month\_y
* day\_y

Features to Drop - Modeling:

* DATE
* SEASON
* HS\_CM
* Surf Temp (C)
* T-10 (C)
* T-20 (C)
* SETTLEMENT/MELT
* X24\_HR\_MAX
* 24HRMAX\_SWING
* 24 HR MIN
* SNOW
* FIVE\_DAY\_SNOWFALL
* WDF5

Only include rows with non-null AVY\_DANGER value