Wildfire

CS 4210
Haowen Yin, Dante Martinez, Wesley Kwan, Beize Li

Introduction

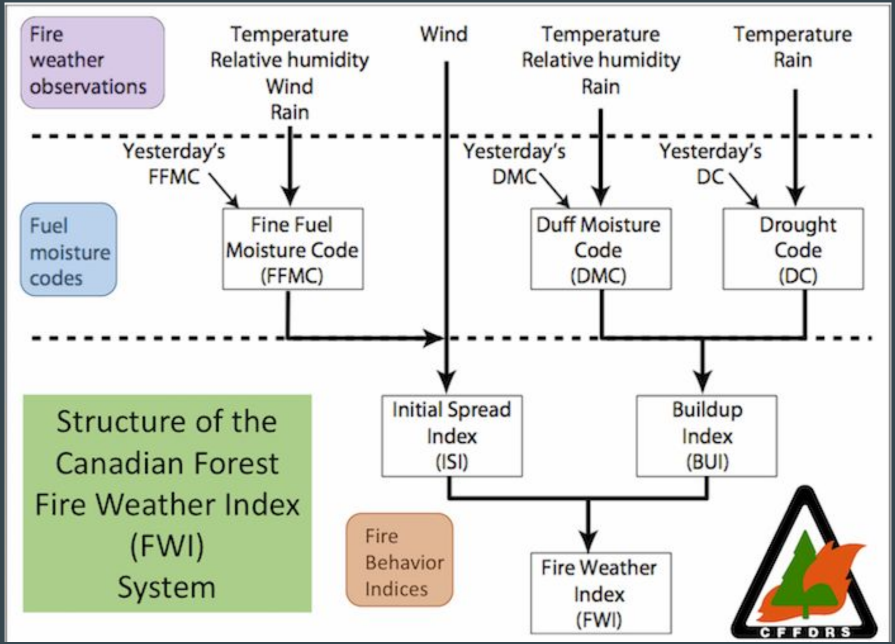
Wildfire is a threat to life, natural environment and economy.

The ability to monitor and predict wildfire helps to improve the world.

Machine learning can help us predict wildfire with information we have.

Dataset

- Meteorological data collected in Montesinho Natural Park from January 2000 to December 2003.
- For each fire instance, the following were collected: spatial data, temporal data (month and day), Fine Fuel Moisture Code, Duff Moisture Code, Drought Code, Initial Spread index, temperature, relative humidity, wind speed, rain and area burned.



Problem Statement

We want to understand what are the best predictors that affects the area burned by wildfire.

We first need to find the features that contribute to wildfire area.

After we find the important features, we test it in our machine learning model for prediction.

Machine Learning Model

Feature selection algorithms

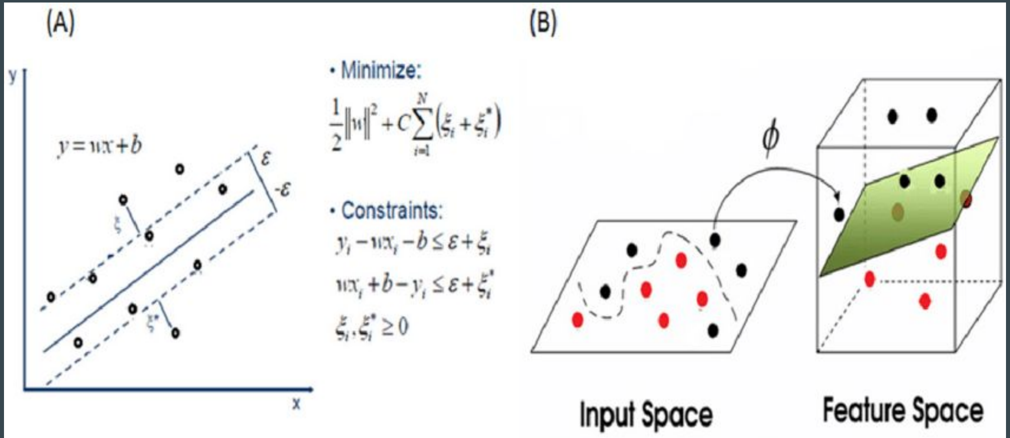
- F-regression
- Mutual information regression

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

$$I(X,Y) = H(X) - H(X \mid Y)$$

Regression Models

- Support Vector Machine (SVM)
- Decision Tree (DT)

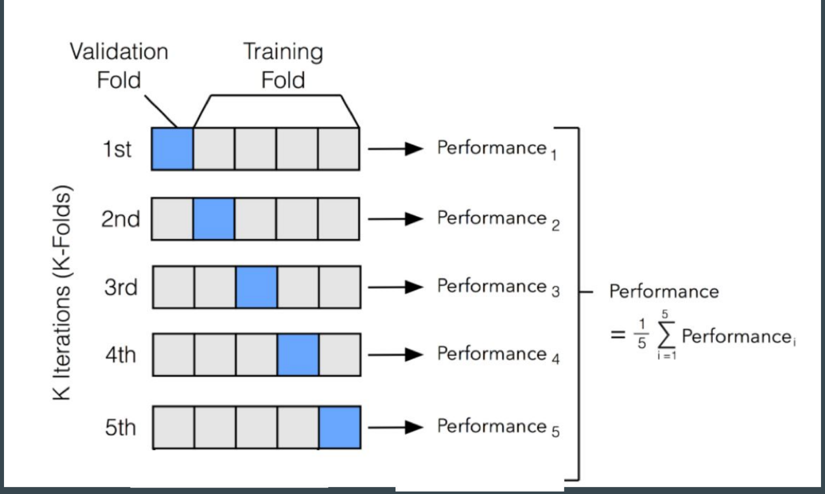


Machine Learning Model

The models were trained and tested through multiple rounds of 10 fold cross validation.

Performance measures (global metrics):

- Average Mean Absolute Deviation (MAD)
- Average Root Mean Squared (RMSE)



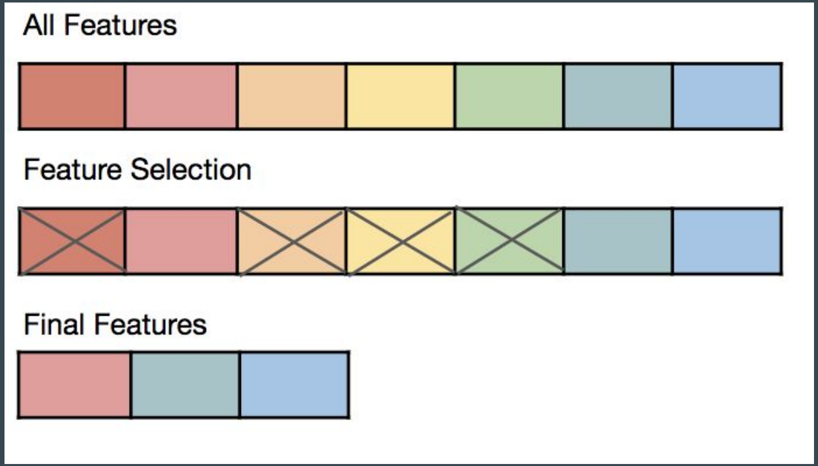
$$MAD = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

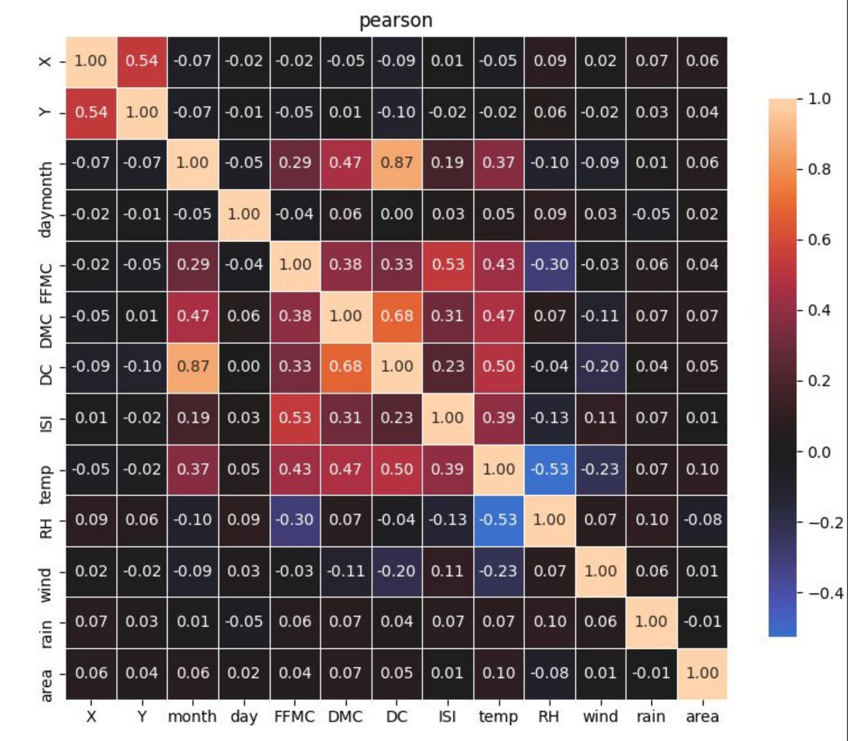
$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$

Manual Selection

Feature correlation

a. Pearson: linear relationship
b. Spearman: strength of monotonic functions
c. Kendall: ordinal relationship



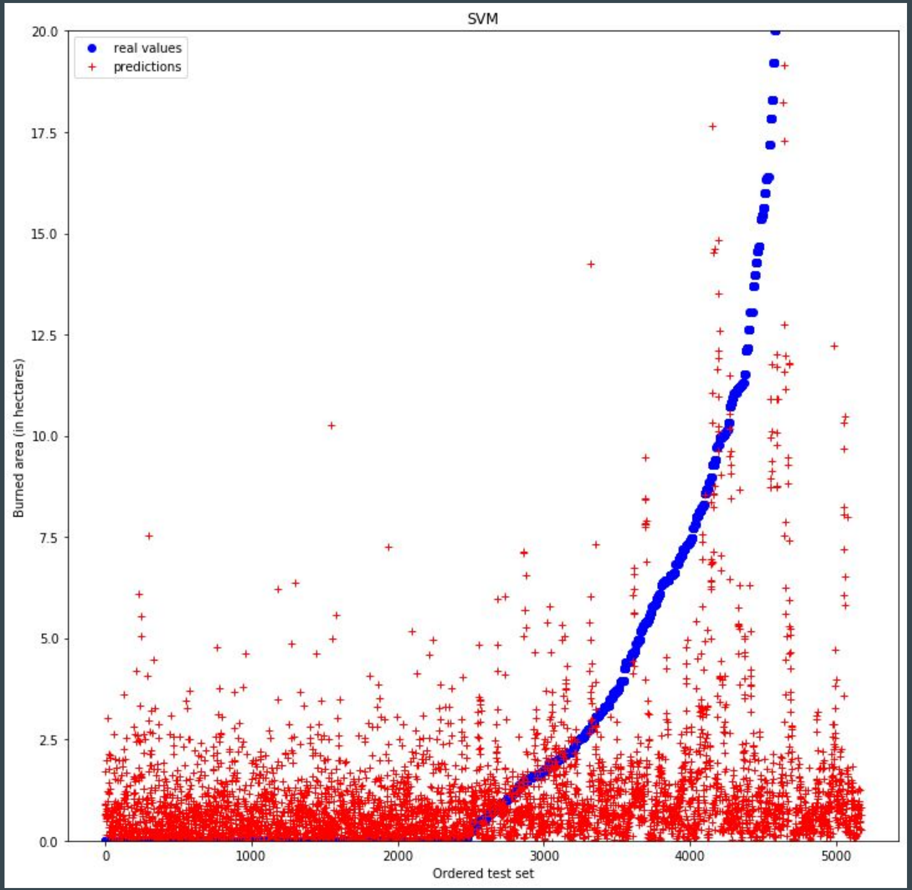


Results

	SVM		DT	
Feature Selection	MAD	RMSE	MAD	RMSE
All features	12.68	46.43	13.63	46.99
Month, DMC, DC, RH (suggested by f-regression)	12.71	47.80	13.13	46.66
Month, DMC, wind	12.67	46.33	13.01	46.92
Month, DC, wind	12.67	47.64	13.04	46.68
Month, DC, RH	12.69	47.54	13.02	46.45
Month, temp, RH, wind	12.70	46.28	12.99	47.23
Month, temp, wind, rain	12.71	46.54	12.93	46.81
Month, DMC, DC, ISI (suggested by mutual info.)	12.68	46.68	13.25	48.03
X, Y, month, ISI	12.62	46.39	13.42	47.49
Month, ISI	12.64	47.40	13.03	44.99

Results

Despite achieving lower MAD and RMSE scores than the previous models developed by P. Cortez and A. Morais, our model struggles to predict persistently.



Conclusion

Our goal was to find best predictor that accurately predicts the size of wildfires.

We built the feature selection algorithm using f-regression and mutual information regression to finding the relationship between the features and the target class.

The best result from both MAD and RMSE does not produce a competitive model to predict wildfire size, but it can provide insight into wildfire's best predictors.

THANKS

CS 4210

Haowen Yin, Dante Martinez, Wesley Kwan, Beize Li