

Can we predict fire with weather data?

Harsh (hp444), Yi Yao (yy899), Murali (mt788)

December 4, 2019

Abstract

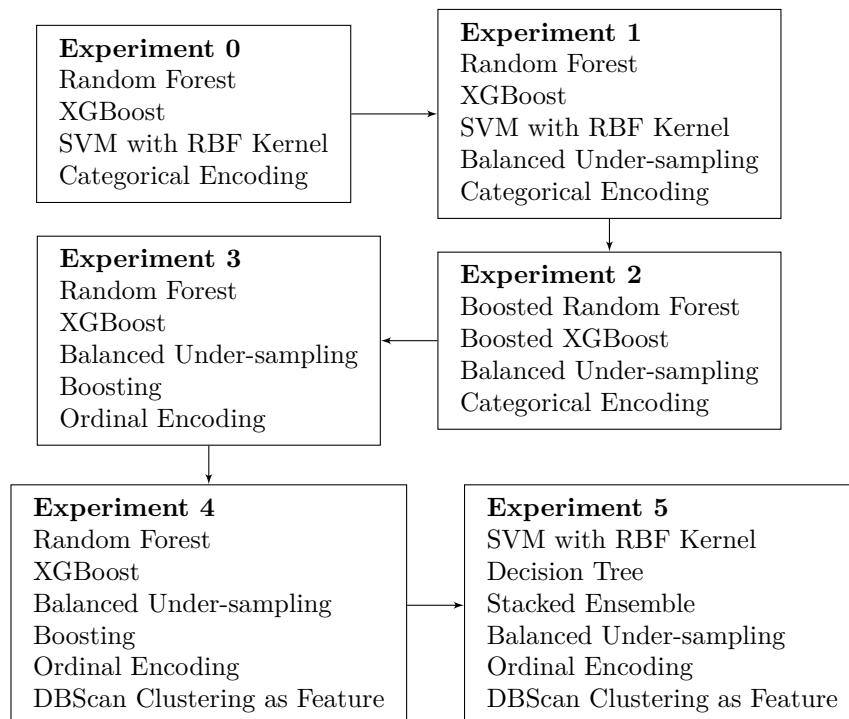
1 Introduction

2 EDA

2.1 Covariates

2.2 Basic Statistics

3 Model Building and Accuracy Tuning



3.1 Experiment 0

First, we fit XGBoost, Random Forest and SVM by using the entire dataset and found the accuracy to be quite low. The low accuracy could be because of the model overfitting. There are, afterall, 1M negative samples and roughly 25k positive samples.

3.2 Experiment 1

Recognizing the unbalanced nature of our dataset, we created a balanced training data set by combining all the positive samples with a random subset of all the negative samples. Using this balanced training data set, we fit XGBoost, Random Forest and SVM with RBF kernel. We observed a significant improvement in test accuracy in the XGBoost and Random Forest models; however, we have not observed any significant improvement in the SVM models.

3.3 Experiment 2

Having the possibility of underfitting due to under-sampling in mind, we have experimented with boosting to increase accuracy. By keeping all the positive samples and randomly resampling negative samples equal to the number of positive samples, we have kept the balance of the training data set while making better use of the data set we have available.

3.4 Experiment 3

3.5 Experiment 4

3.6 Experiment 5

4 Conclusion