DAT 530 Advanced Statistical Methods

Project Review: Week 4

Reviewer: Brandon Hosley

September 7, 2020

Project Title:

Which Kickstarter projects reach their funding goals?

Author(s):

Michael Griffin

Source:

NYC Data Academy

The Problem the Author(s) is Trying to Solve in the Project:

The author seeks to determine what factors can be used to predict the likelihood of reaching funding goals for projects appearing on GoFundMe.com.

Machine Learning (ML) Algorithm(s) used:

- Natural Language Processing

- Random Forest

- **sklearn AutoML**

A Brief Description of One of the ML Algorithms used:

The author used the **SKlearn AutoML** framework. The advantage of this method is that it is very easy for the user to implement. The framework provides automated data preprocessing to enable the selection of a classifier from any of 15 common classification algorithms. The disadvantage is that the model is that it adds even more obscurity to the model. This is best demonstrated in this paper by the fact that the author offers the name of the framework but never states any specifics about the actual algorithm used.

Metrics Used to Evaluate the ML Algorithms:

An F_1 score is used to evaluate the model. The scores presented suggest the model is slightly overfit as the score associated with training data is significantly higher than the score associated with the test group.