Final Project

Noah X. Deutsch
Self-Paced Data Science Program

The Challenge

Using a Kaggle dataset, I set out to answer the following question:

Can we use supervised learning to predict the success of a kickstarter project before it even launches?

Methodology

Obtain

Load in our data from Kaggle

Scrub

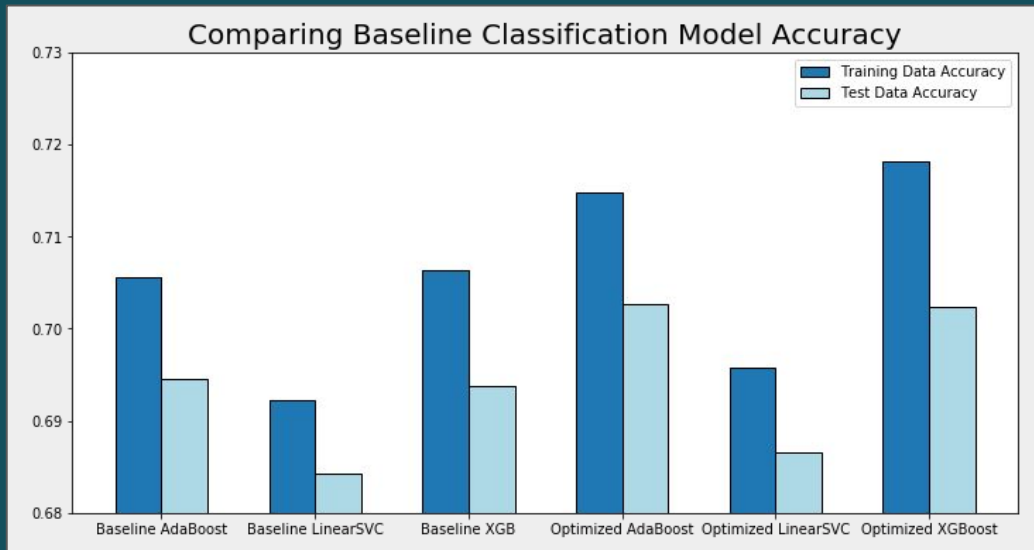Organizing, normalizing, and splitting our data.

Model

Fit baseline models, Compare results, and Optimize.

Interpret

Draw conclusions from our results.

Modeling

We tested a number of different classification models in order to discover which model would predict the success of kickstarter projects most accurately.

Comparing Baseline Classification Model Accuracy

Results

Ultimately, our best performing model was able correctly predict the outcome of a Kickstarter project ~70% of the time (with data it hadn't seen before).

The classifier was able to classify ~62% of successes correctly and ~74% of failures correctly.

Improving our Model Accuracy in the Future

Ultimately, the predictive ability of our model was limited by the available data.

Future experiments would benefit from additional project-level data, including data points like product pricing and tiers, full project descriptions, creator bios, and more.

Thank you!

Noah X. Deutsch