Machine Learning Engineer Nanodegree

Capstone Project

Bharadwaz Mahankali

March 14, 2018

Definition

Project Overview

DonorsChoose.org receives hundreds of thousands of project proposals each year for classroom projects in need of funding. A large number of volunteers is needed to manually screen each submission before it's approved to be posted on the DonorsChoose.org website.

Next year, DonorsChoose.org expects to receive close to 500,000 project proposals. As a result, there are three main problems that need to be solved:

1. How to scale current manual processes and resources to screen 500,000 projects so that they can be posted as quickly and as efficiently as possible.
2. How to increase the consistency of project vetting across different volunteers to improve the experience for teachers.
3. How to focus volunteer time on the applications that need the most assistance.

Problem Statement

Predict whether or not a DonorsChoose.org project proposal submitted by a teacher will be approved, using the text of project descriptions as well as additional metadata about the project, teacher, and school. DonorsChoose.org can then use this information to identify projects most likely to need further review before approval.

Metrics

Model will be evaluated on area under the ROC curve between the predicted probability and the observed target.

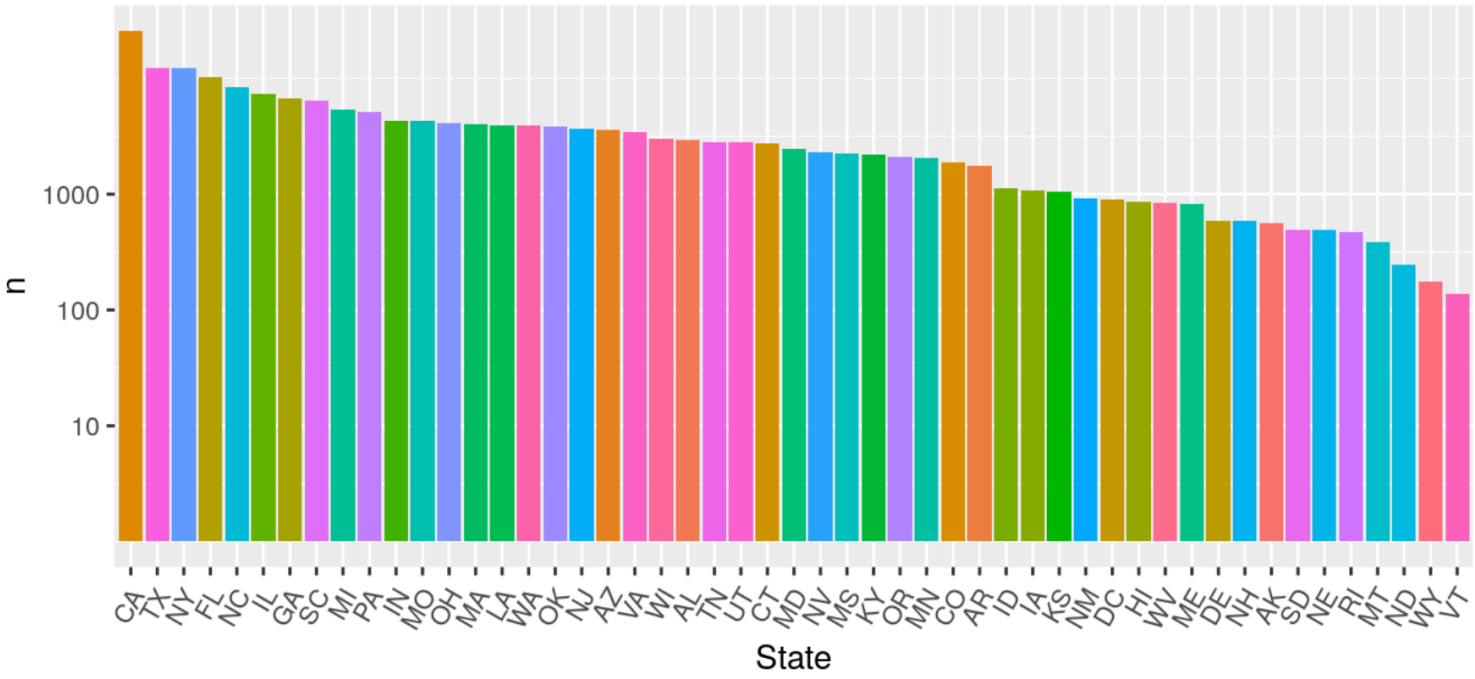
Analysis

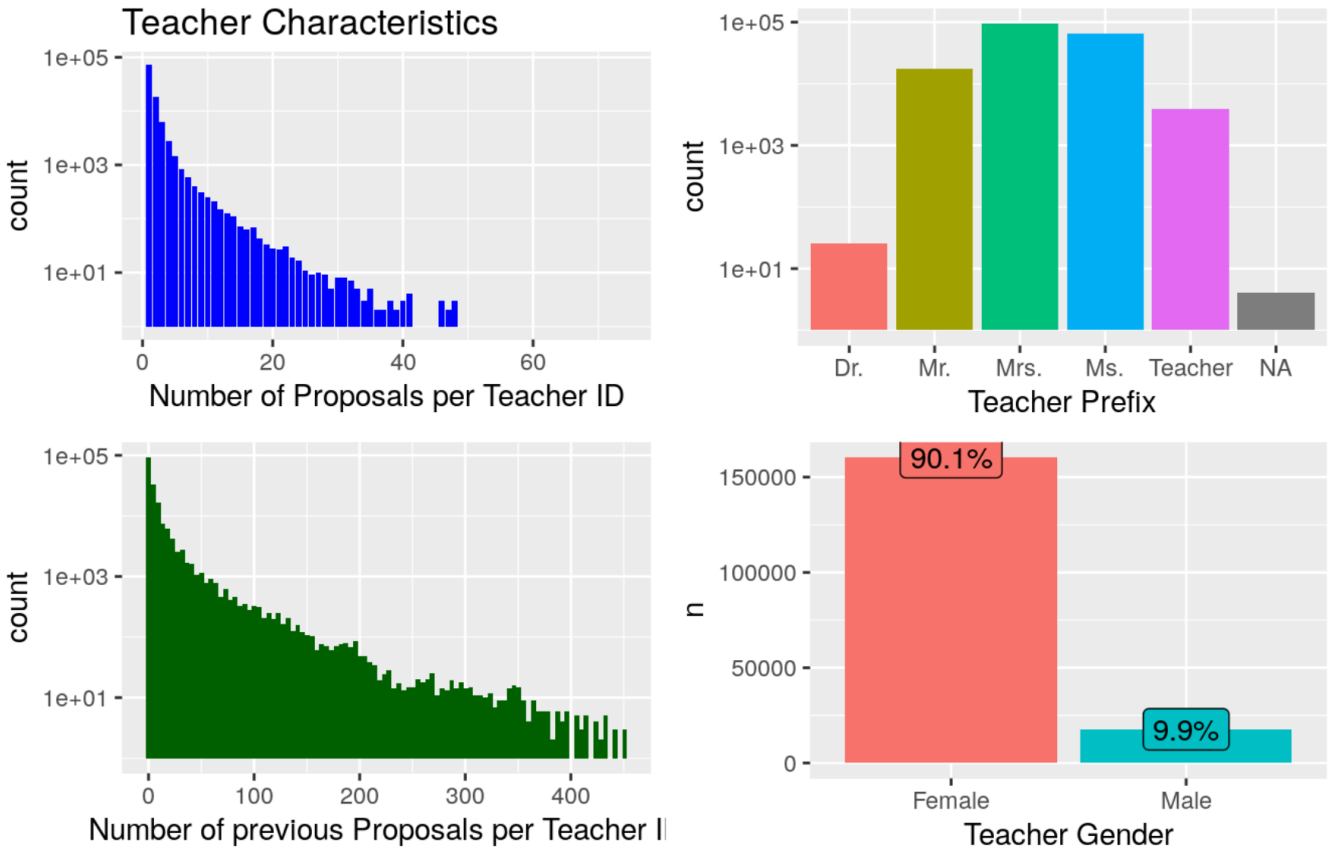
Data Exploration

Exploratory Visualization

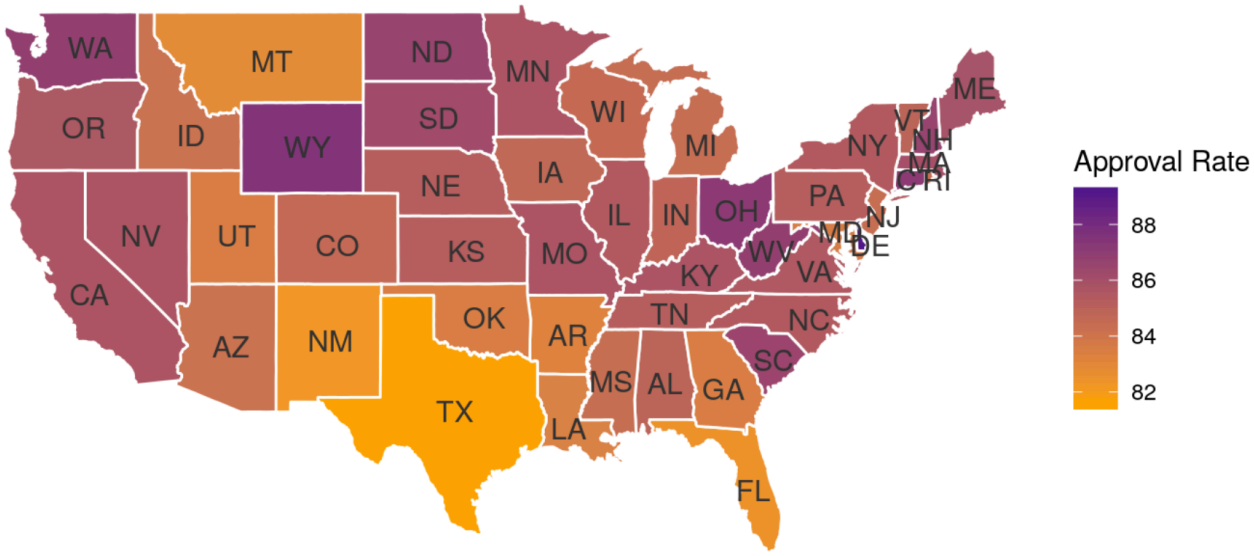
Algorithms and Techniques

Benchmark

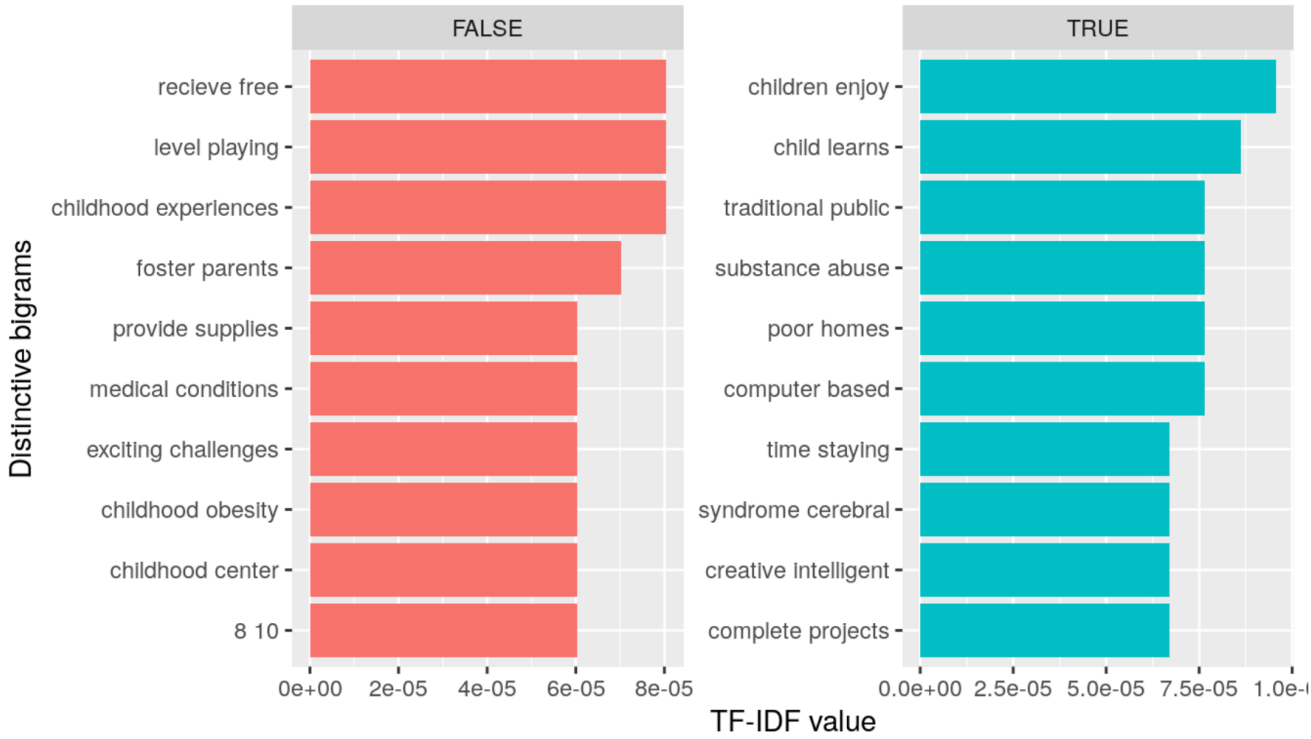


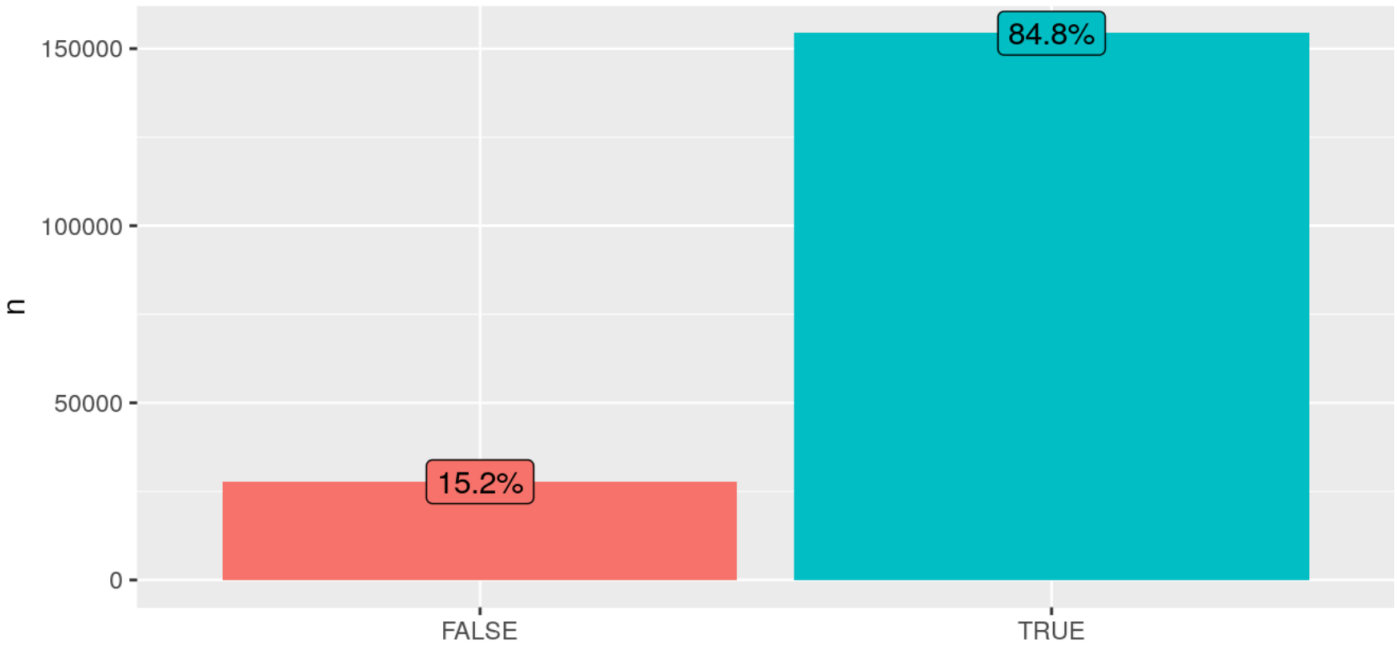


Methodology



Data Preprocessing





Implementation

Refinement

Results

Model Evaluation and Validation

Justification

Conclusion

Free-Form Visualization

Reflection

Improvement