

# Predicting Funded DonorsChoose Projects

Dan Ah Tou

15 DONORS

\$191 STILL NEEDED

\$516 GOAL

expires Aug 10

\$

Give

8  
DAYS LEFT!

## A Cozy Reading Corner

*My students need a cozy reading corner to help build their love for reading. The rug will be placed on the classroom tiles while the large pillows will be used as back support.*

### My Students

Many of my students are English Language Learners, some of which are new to this country. They have had little to no formal education. Despite their lack of formal education, my students are passionate, hardworking, and caring. Throughout the day, they are exploring, discussing, moving, thinking constantly trying to absorb as much knowledge and information as they can.

*I want to make education and learning as fun and welcoming as I can, and I believe that through this project I will be able to do that!*

**Mrs. C.**

NEVER BEFORE FUNDED

Grades 3-5



McCollough-Unis School  
Dearborn, MI

More than three-quarters of students from  
low-income households

# Tools

## Data Storage

- PostgreSQL via pgAdmin, SQLAlchemy

## Data Analysis

- pandas, scikit-learn, nltk

## Presentation

- matplotlib, Google Slides



✓ Reviewed Dataset

# Data Science For Good: DonorsChoose.org

Help DonorsChoose.org connect donors with projects they care about



DonorsChoose.org and 3 collaborators • last updated 3 months ago



479  
voters

[Data](#) [Overview](#) [Kernels](#) [Discussion](#) [Activity](#)

Download (1 GB)

New Kernel

Data (1 GB)

API

kaggle datasets download -d donorschoose/io ?

Download All



## Data Sources

- Donations.csv 4.69m x 7
- Donors.csv 2.12m x 5
- Projects.csv
- Resources.csv 7.21m x 5
- Schools.csv 73.0k x 9
- Teachers.csv 403k x 3

## About this file

Edit

More information at a teacher level. Each row represents a single teacher. This dataset is joined with the project data using the "Teacher ID" column.

## Columns

Edit

- Teacher ID** Unique identifier of a teacher.
- Teacher Prefix** "Mrs.", "Ms.", "Mr.", "Teacher" (gender neutral option) chosen by teacher during account creation.
- Teacher First Project Posted Date** Date on which the teacher's first project was posted.

# Data

1.21M Projects, 7.2M Resources, 403k Teachers, 73k Schools

Analyzed only projects for schools in Washington

- ~2% of the dataset (26,000 projects, 78% funded)

Features in Dataset			
Project Funded Status	Project Cost	Subject Category	Subject Subcategory
Resource Category	Resource Amount	Percentage Free Lunch	School Metro Type
Project Title	Project Essay	Grade Level	Teacher Title (Mr./Mrs.)

# Model Selection

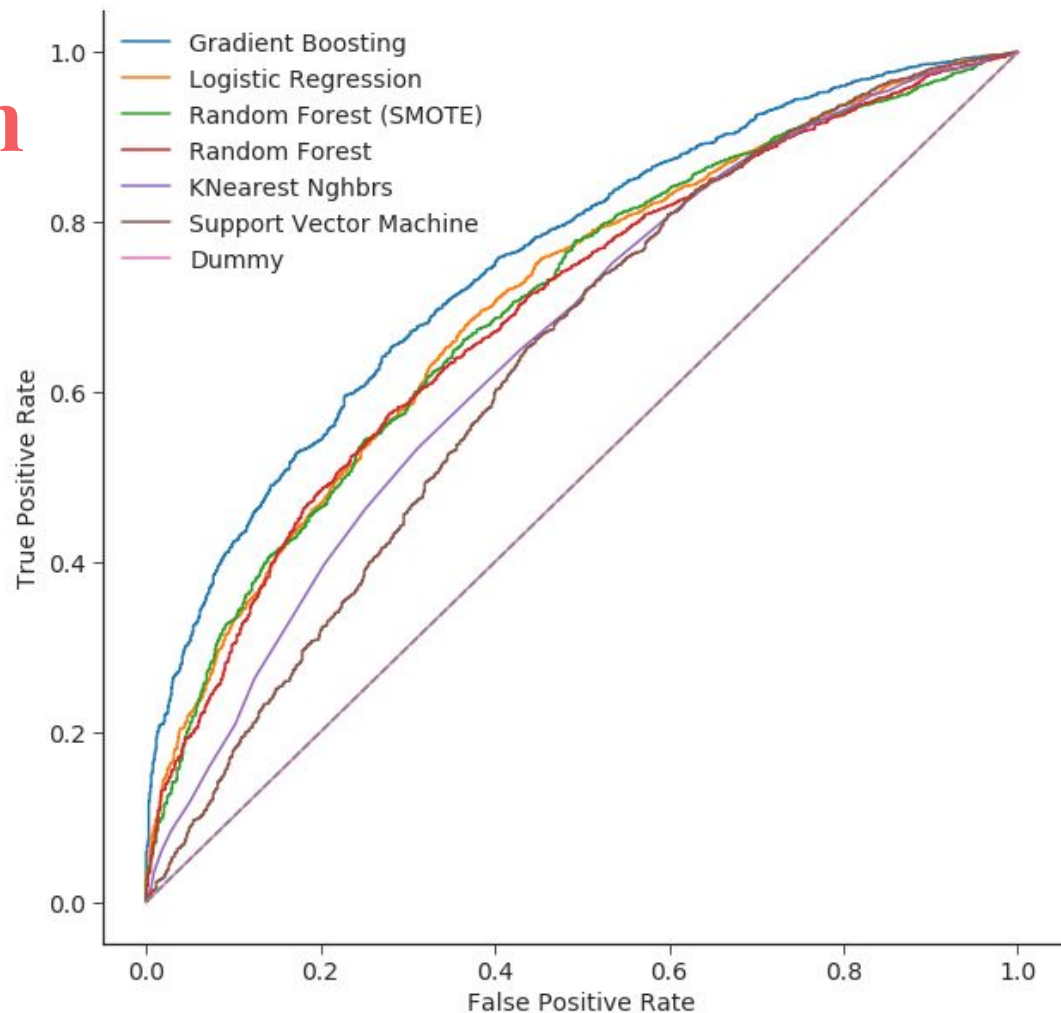
- K-Nearest Neighbors
- Logistic Regression
- Random Forest
- Gradient Boosting
- Support Vector Machine

Models were tuned using randomized, 5-fold cross validated search over a range of hyperparameters

# Model Comparison

Gradient Boosting  
performs best at all  
levels

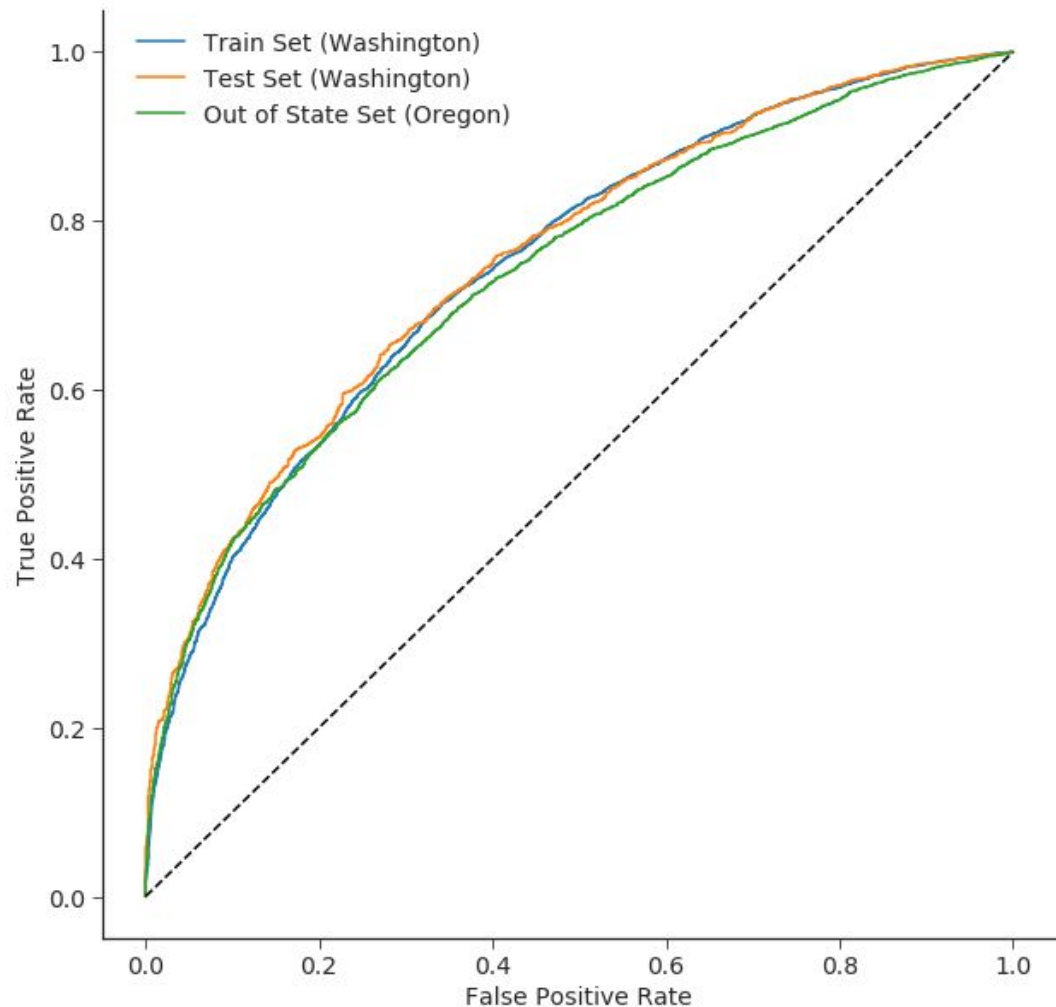
AUC: 0.749



# Model Verification

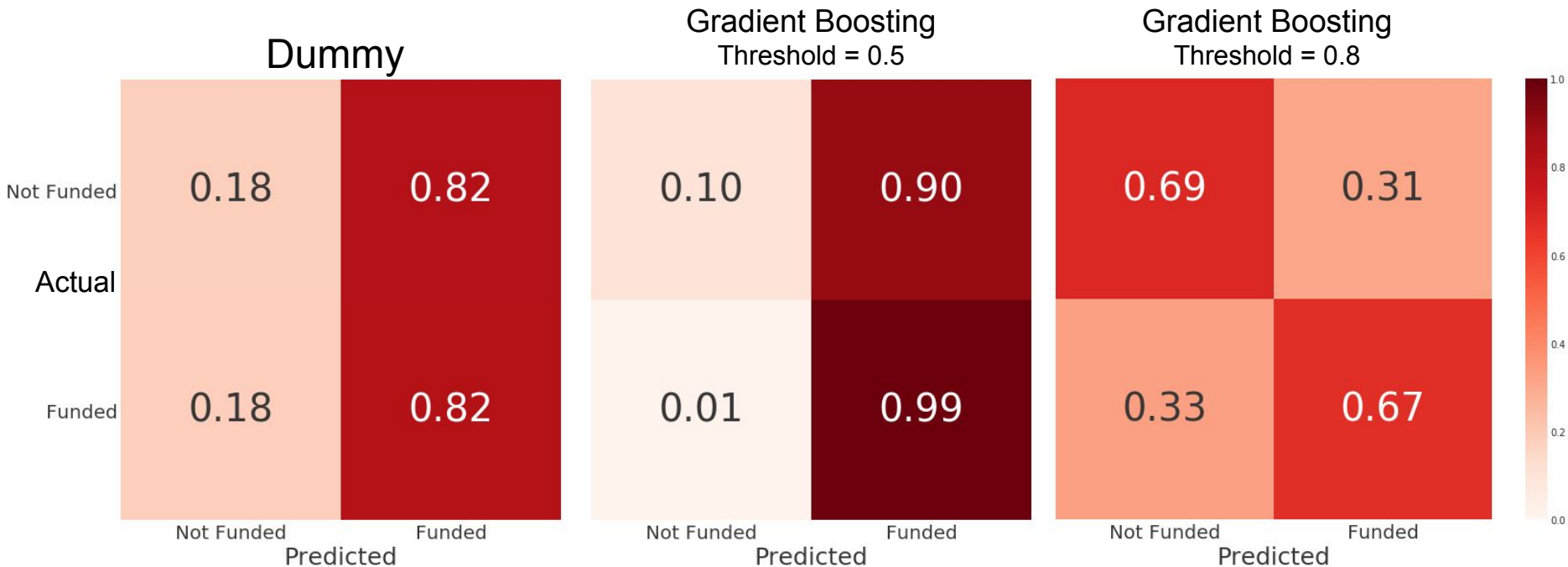
Gradient Boosting model  
performs similarly on  
data from a different  
population

AUC (Oregon): 0.737

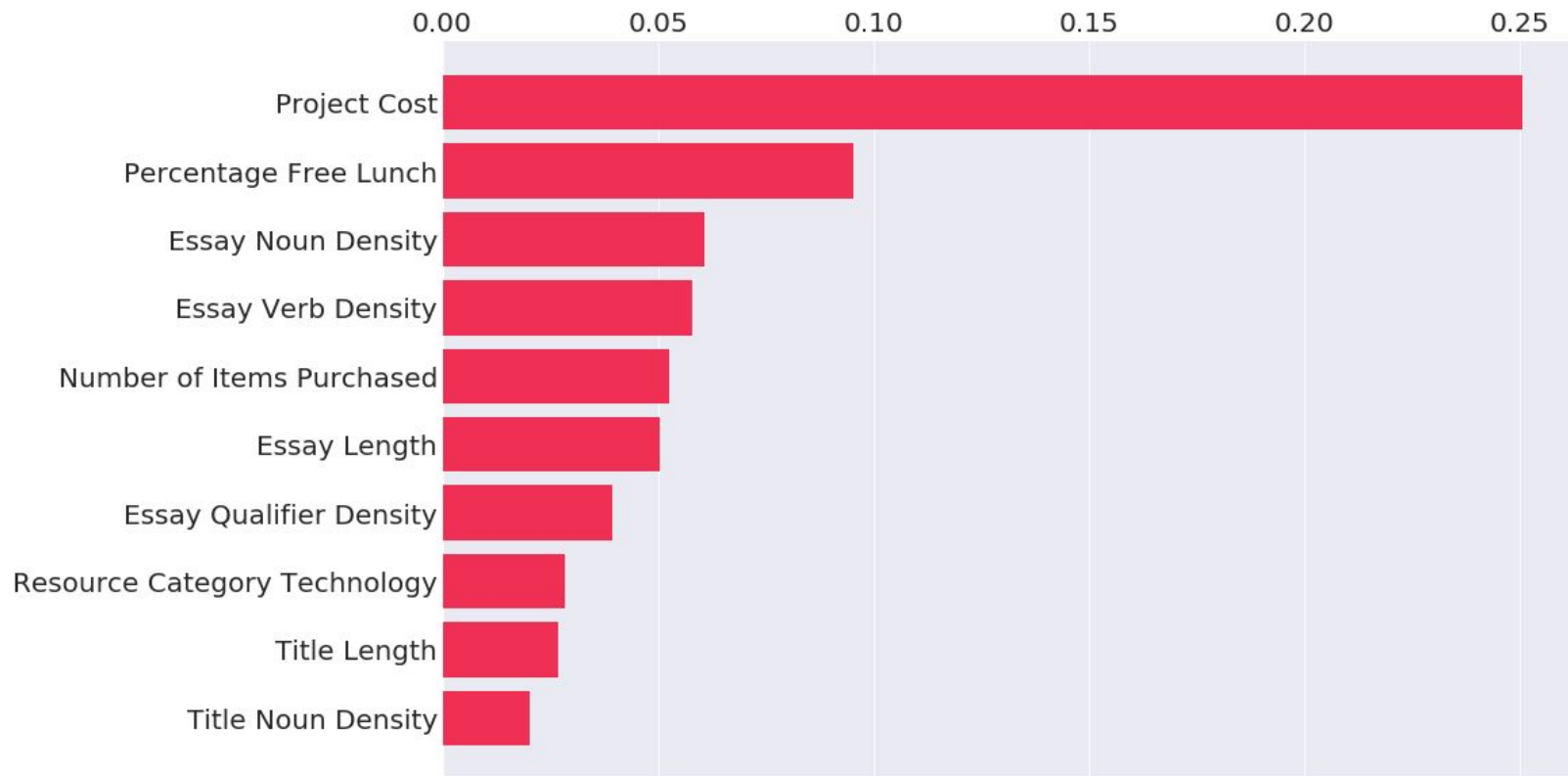




# Gradient Boosting Classification



# Feature Importance



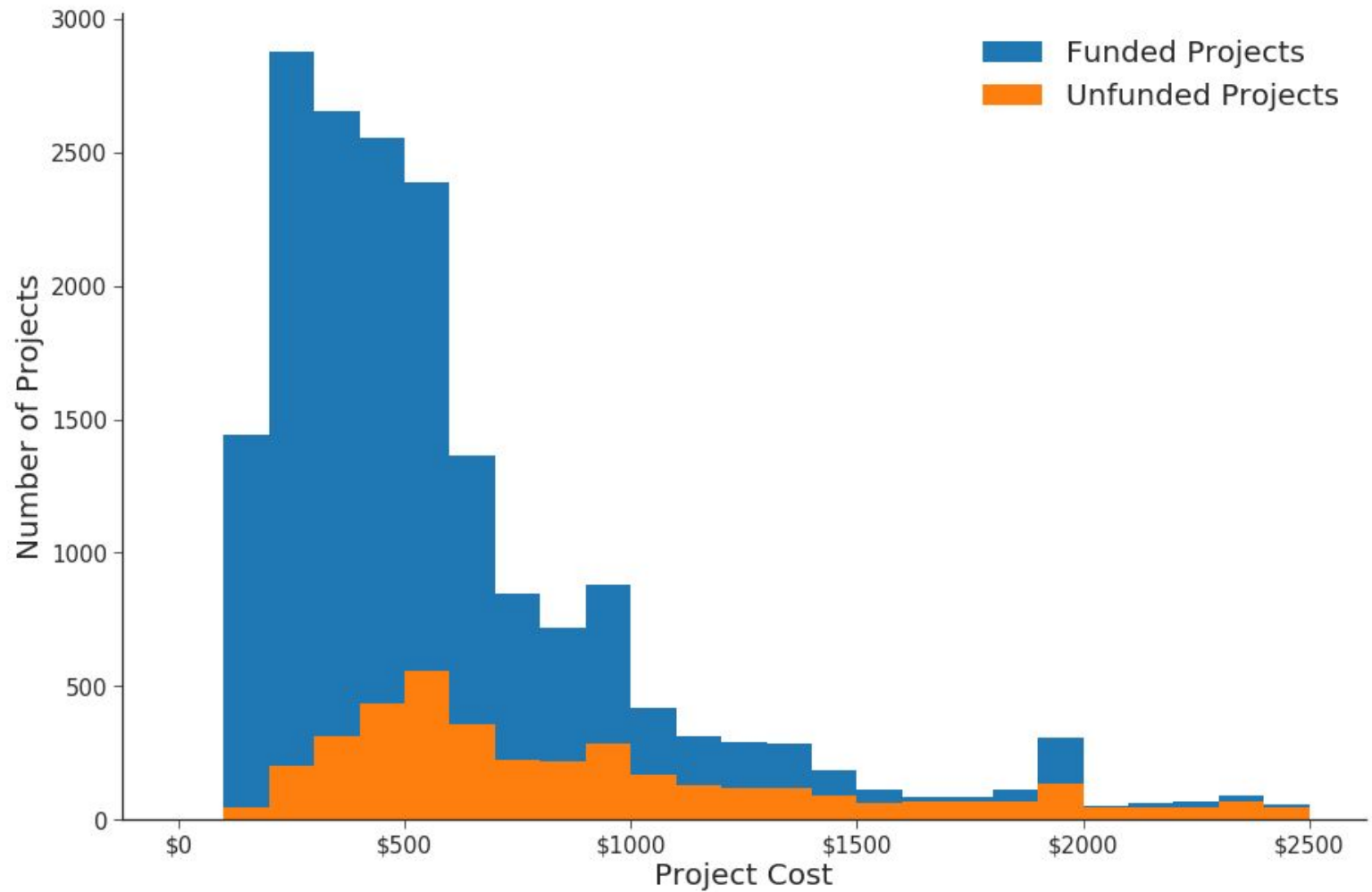
# Recommendations for Successful Projects

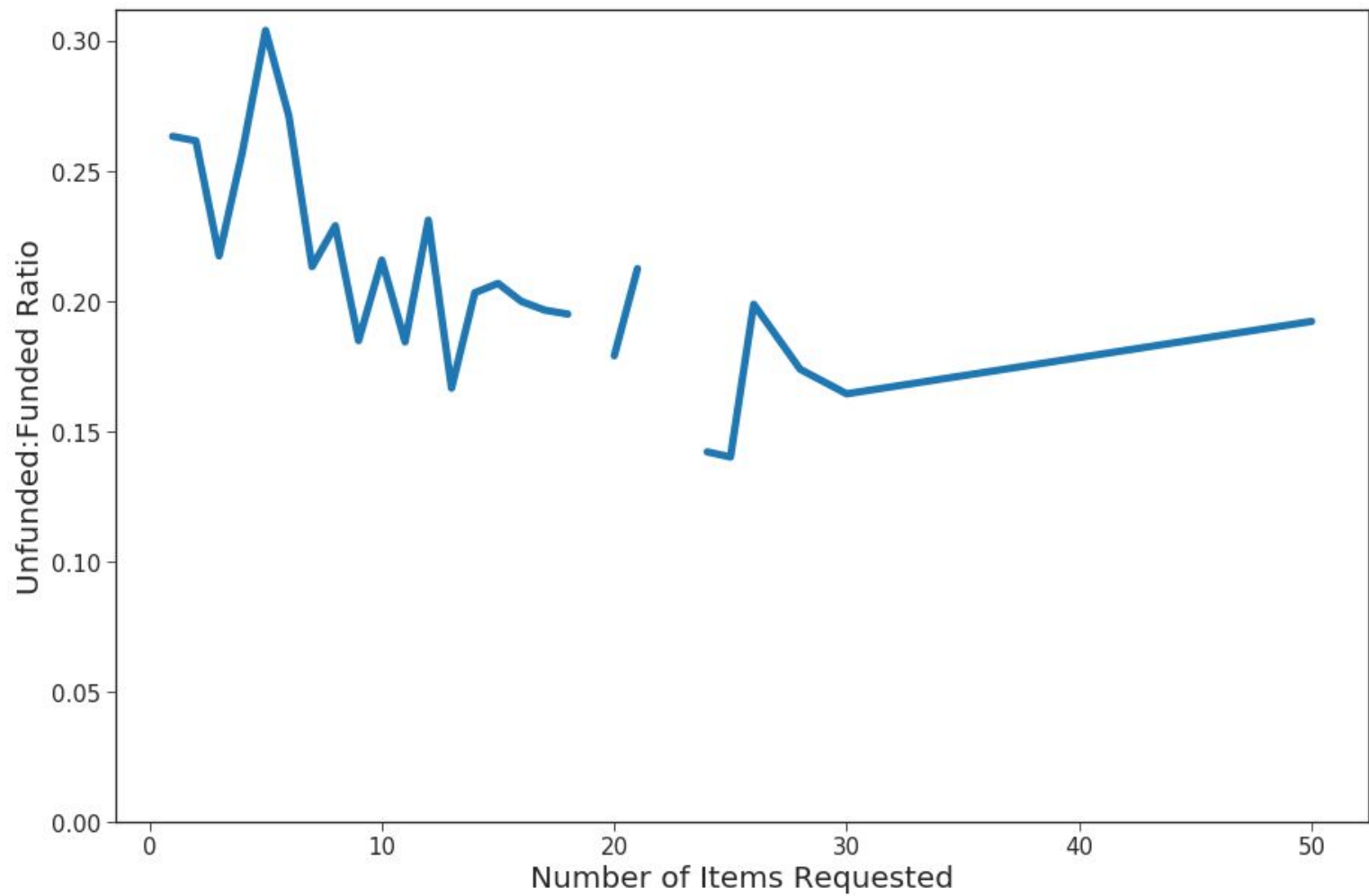
- Keep project costs under \$500
- Request a number of items equal to class size
  - Projects with 24 items have the highest funding rate
- Write a descriptive essay that emphasizes how and what your project will accomplish

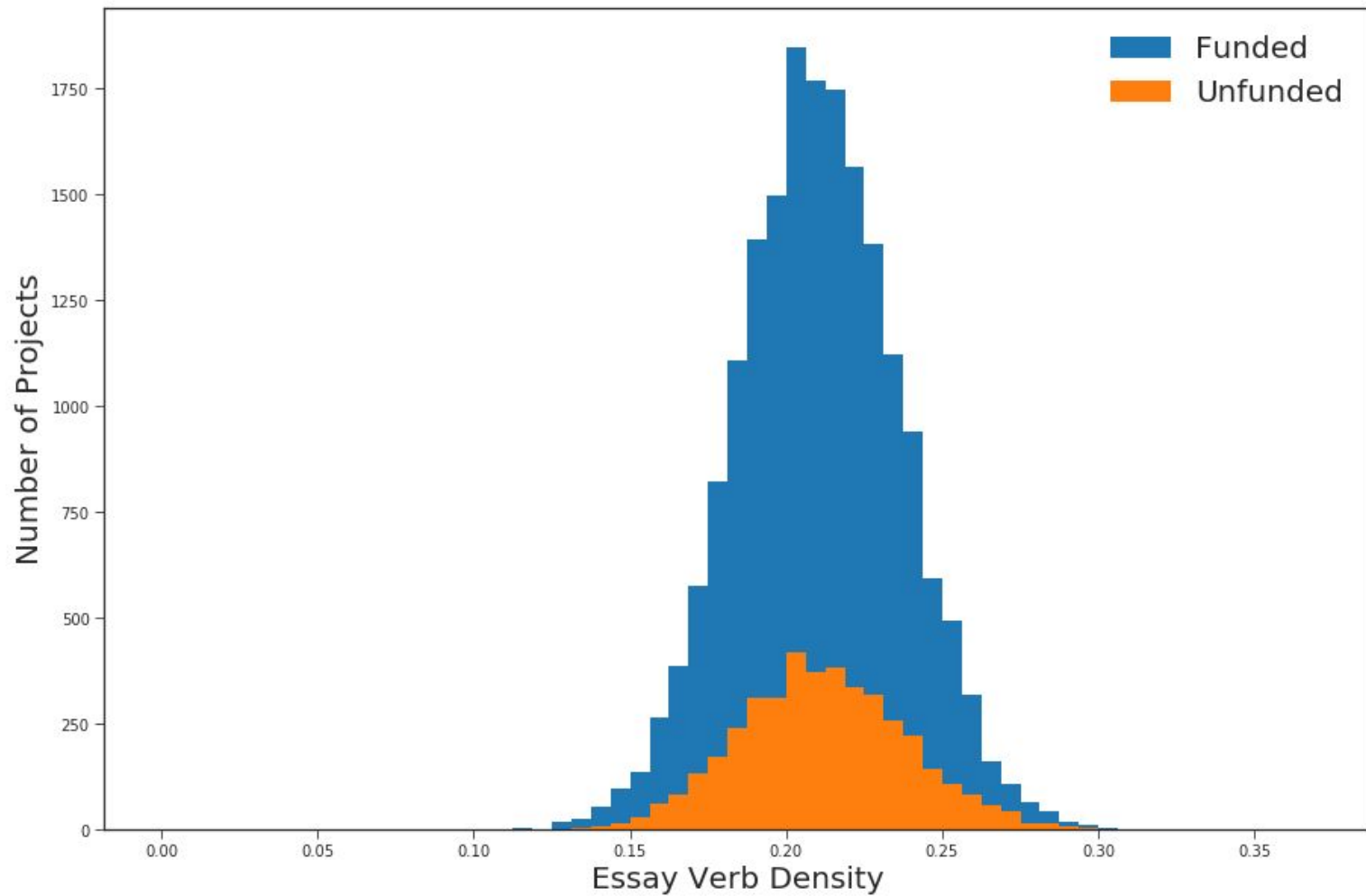
# Ideas for Further Improvement

- Add class size to dataset to calculate cost per student
- Add number of social media shares
  - Model cannot account for how well or aggressively a teacher campaigns for donations

# Appendix

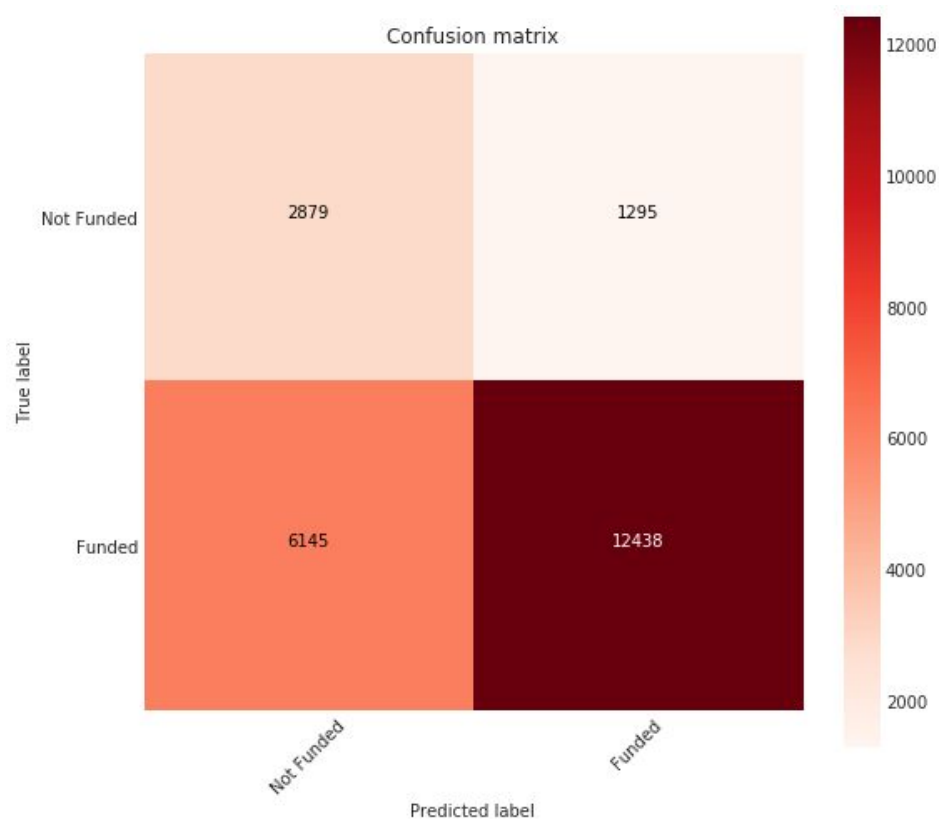








# Confusion Matrix (threshold = 0.8)



# Confusion Matrix (threshold = 0.5)

