



Analytics Project Presentation - Fall 2014

Analytics Project: Get More Donations for
Classrooms in Need

Team: Yi Wan & ChiaLing Wang

Abstract

We intent to analyze online donation behavior, the chance a request being funded, and resource distribution in education industry. Our data comes from DonorChoose.org, an online donation platform. We applied statistical methods over 600,000 data entries to analyze online donation behaviors and what factors influence a project's possibility to get funded. Natural Languages Processing Techniques are utilized to extract from 3GB-essays the most powerful describing words to increase funding odds. Furthermore, we built a Random Forest model to predict the funding status of a project. We then tested and validated our statistical observations using the prediction model. We will conclude with some observations we made and the testing results of the prediction model and try to explain the senses behind them.

Agenda

- Introduction
- Design Flow & Technologies
- Results
 - Overview
 - The Most Powerful Words
 - What influence the odds to get funding?
 - Prediction model
- Conclusion

Get More Donations for Classrooms in Need

Background

- DonorsChoose.org is an online donation platform where teachers post requests for kinds of education resources (e.g. books, technologies). If their project request is fulfilled, DonorsChoose.org will ship the materials they need. From 2000 to 2014, DonorsChoose.org raised \$289,329,866 for 13,277,515 students and 212,408 teachers.

Motivation & Goals

- What increases the odds for a project to get funded? Does the teacher gender matter? Does people really would like help kids in high poverty area?
- Can we see from an online donation platform the education resource distribution? Which part of the United States have more poor school that relies on teachers to beg online? What resource are they begging for?

Get More Donations for Classrooms in Need

Data Sources

1. Donorschoose.org Open Dataset

Project dataset: 640,000 classroom projects that have been posted, including school locations, school types, teacher attributes, project categories, project pricing and impact, project donations, projects status.

Donation dataset: 3,000,000 donations records, including donor city, state, and partial-zip, payment methods, donation types, donation times and amounts.

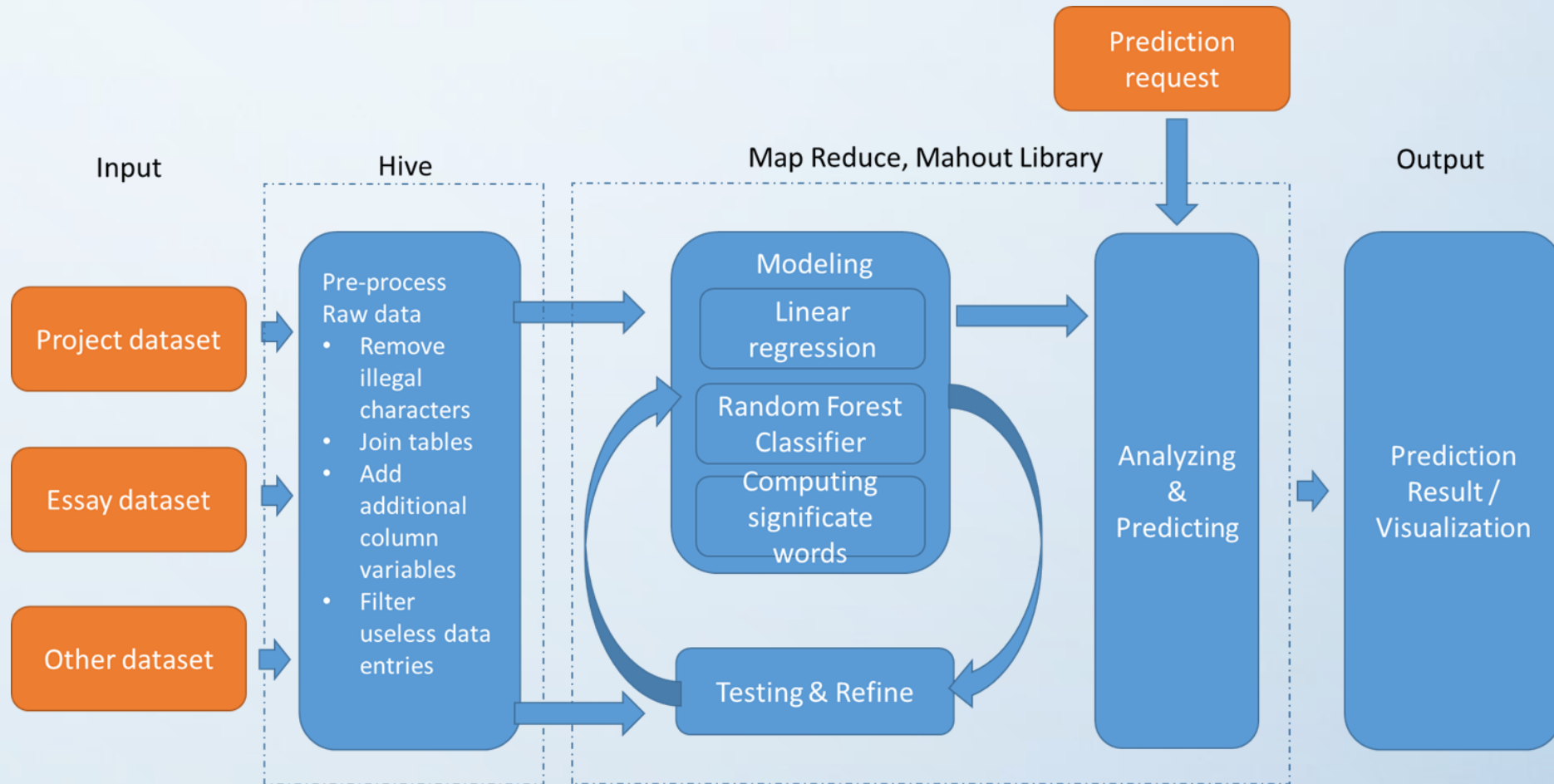
Project written requests / essays: 640,000 text records of the teacher-written requests accompanying all classroom projects. The essay includes four paragraphs, introducing the classrooms, describing the situation and solutions, and empowering the donors

2. School & Education Data.

National Center for Education Statistics (NCES) data via Factual

Get More Donations for Classrooms in Need

Software Architecture



Get More Donations for Classrooms in Need

Data Flow

Collect Raw Data

- Download open data set from DonorChoose.org
 - Projects data
 - Donations data
 - Project written requests / essays
- Poverty rate by states
 - States list
 - house-hold income
 - 2014 Poverty rate
- Additional data source
 - National Center for Education Statistics (NCES) data via Factual



Data Preprocessing and Storage

- Data preprocessing
 - Clean illegal characters in row table
 - Formatting row data
 - Create Hive table for each data set
 - Add / drop attributes required by the algorithms in subsequent stages.
- Data storage
 - All the data are stored as Hive tables.

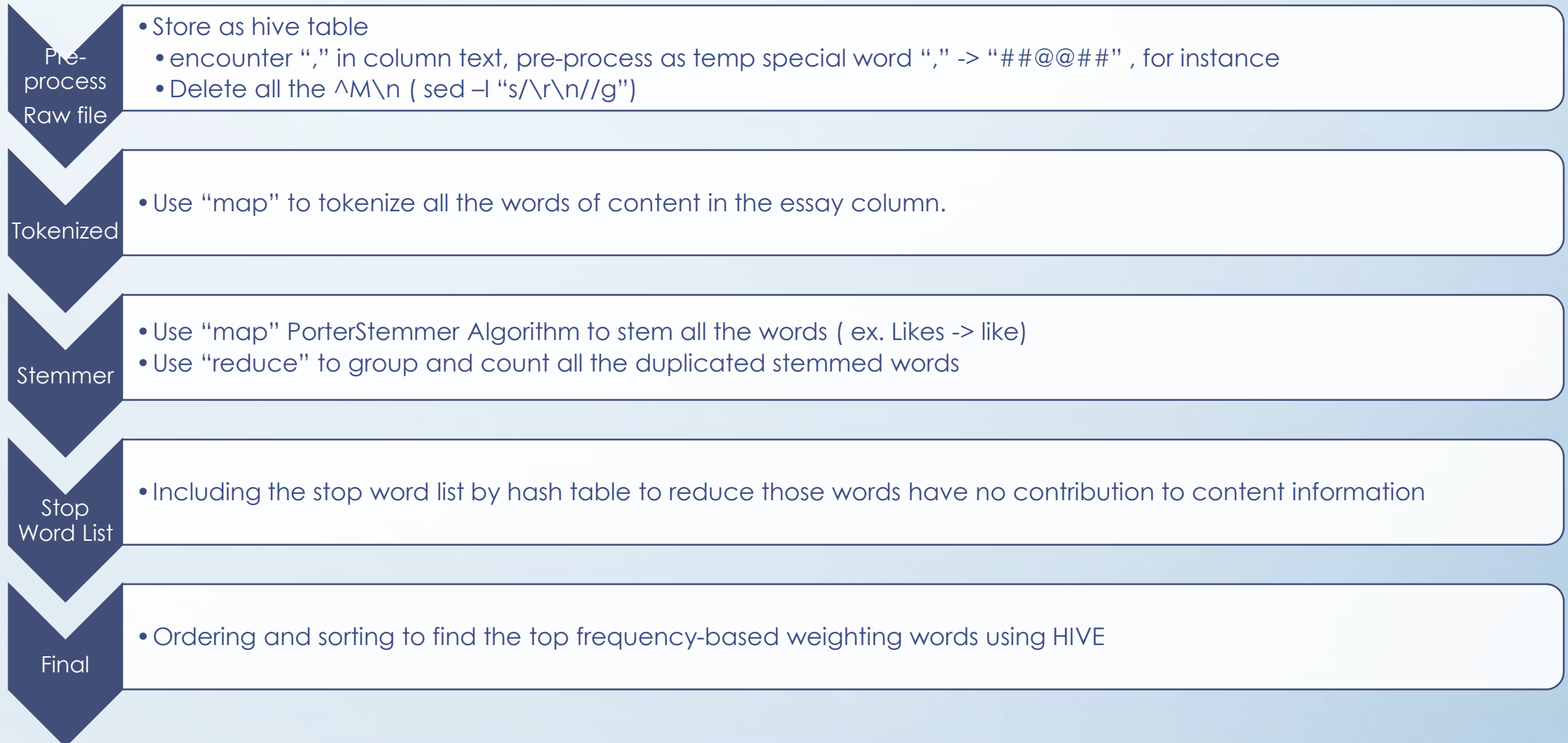


Data Analysis

- Find out powerful words to describe a project using **Natural Language Processing algorithms**.
- Build a **Random Forest Tree** to predict whether a project request will be fully funded.
- Analyze the funding rates over states/ genders / resource type.
- Apply linear regression to predict the amount of donation requested in future years.

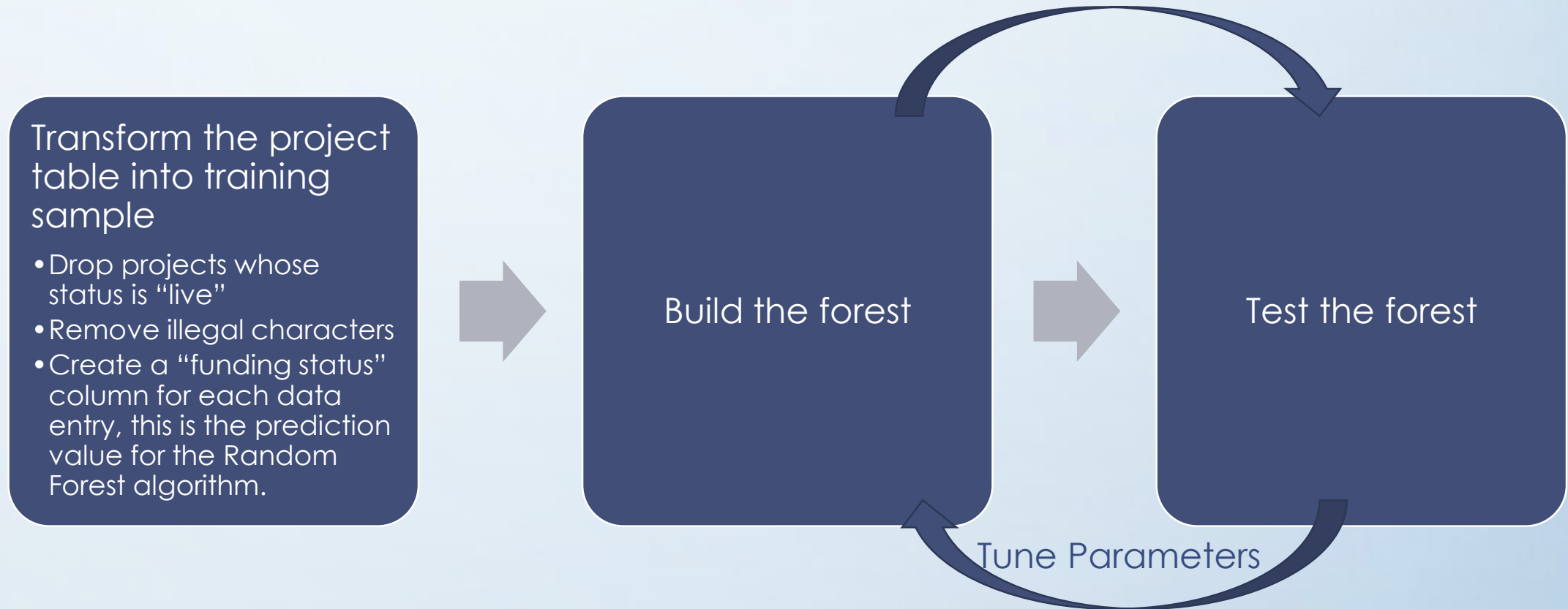
Get More Donations for Classrooms in Need

Implementation - Extract Most Powerful Words using NLP Algorithm



Get More Donations for Classrooms in Need

Implementation -Predicting funding status for a project



Get More Donations for Classrooms in Need Results

Donation Behaviors

- Your Odds To Get Fully Funded
 - Location REALLY matter: east coast win
 - Resource type you need: books are most favored and technologies are not
 - Poverty: “poorest” projects win
 - Sex: female teachers win

Most Powerful Words To Describe Your Project

- Student
- School
- Help
- Need

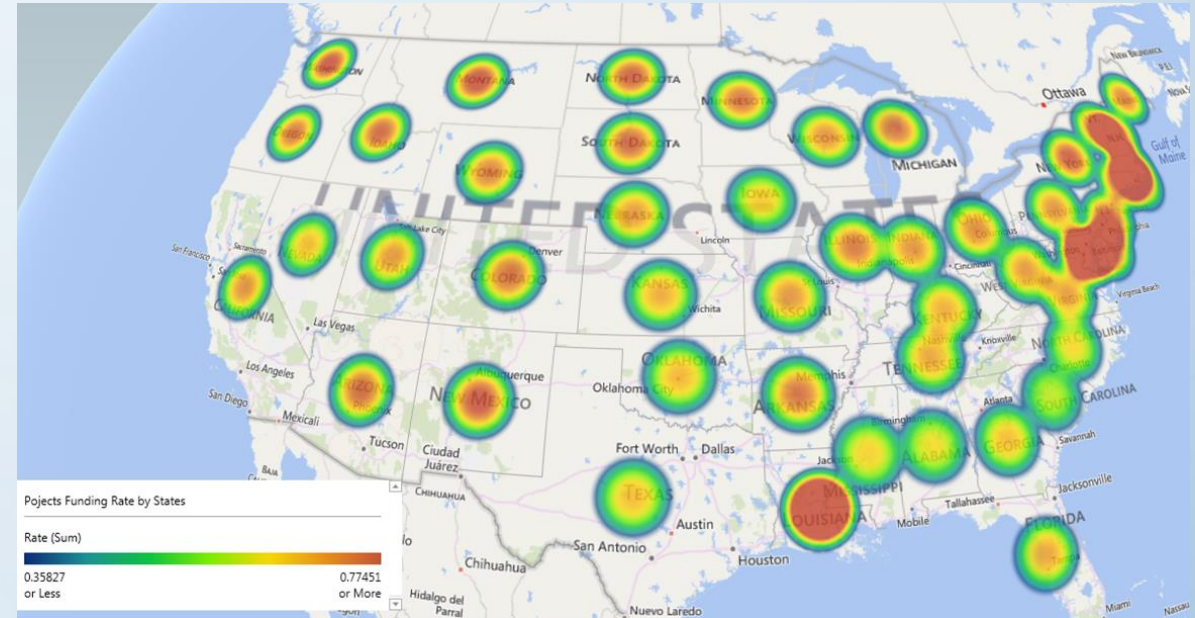
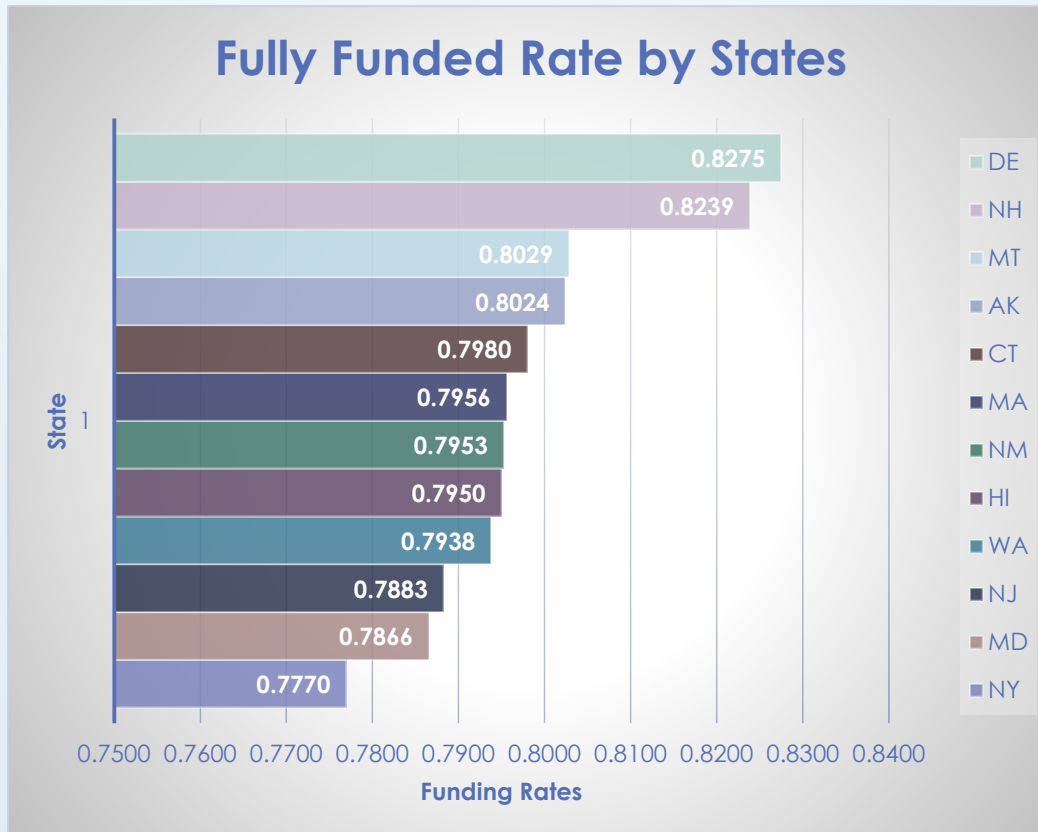
Get More Donations for Classrooms in Need Result Overview – Projects Intensity

East Coast requests more
projects maybe it's because
DonorChooase.org was
originated from New York City.



Get More Donations for Classrooms in Need

Results Overview – Which States have higher Funding Rates?

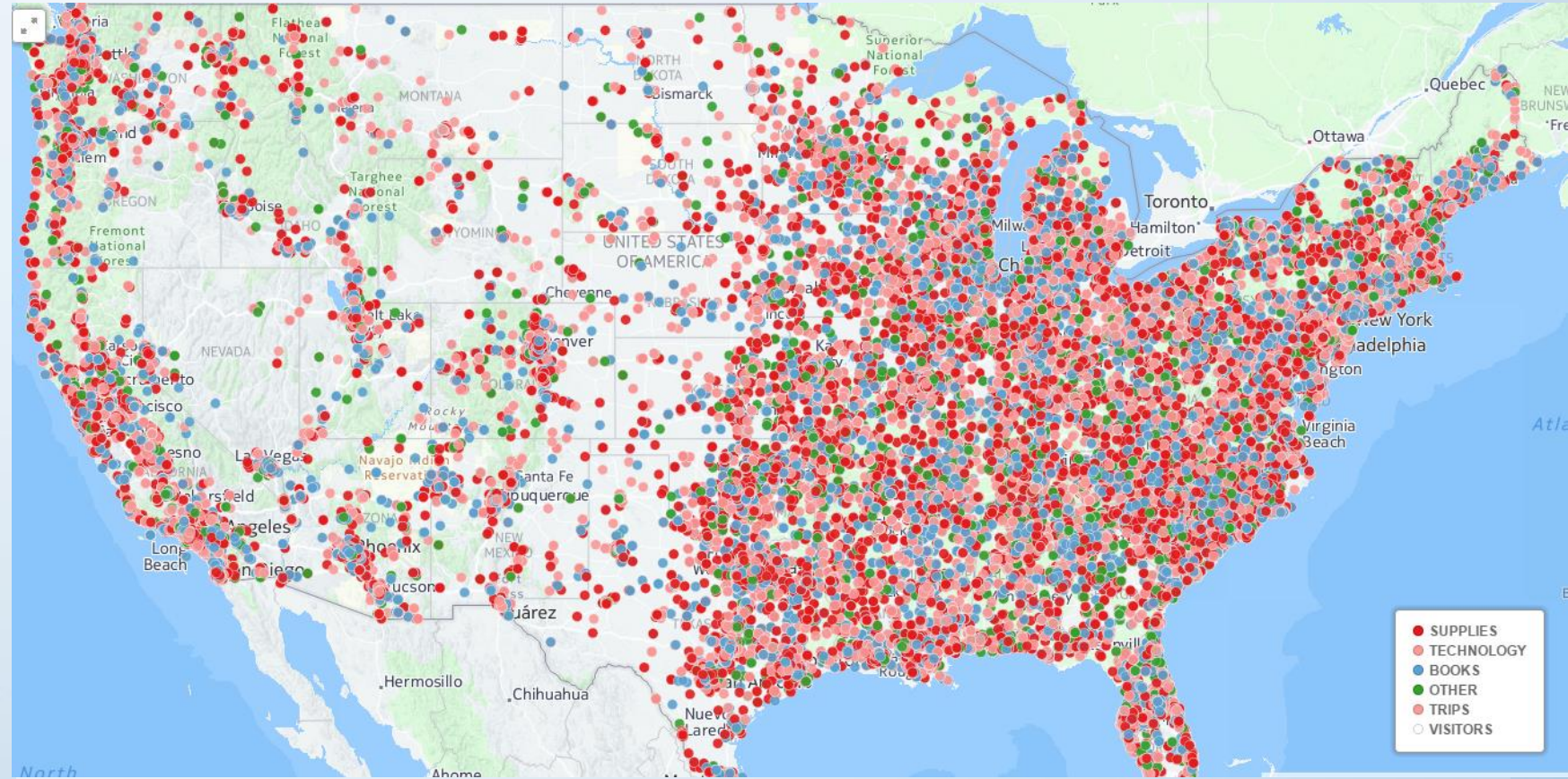


- States which have more projects proposed do not get a higher proportion of fully funded projects.
- Delaware, New Hampshire, Montana, Alaska, Connecticut WIN!

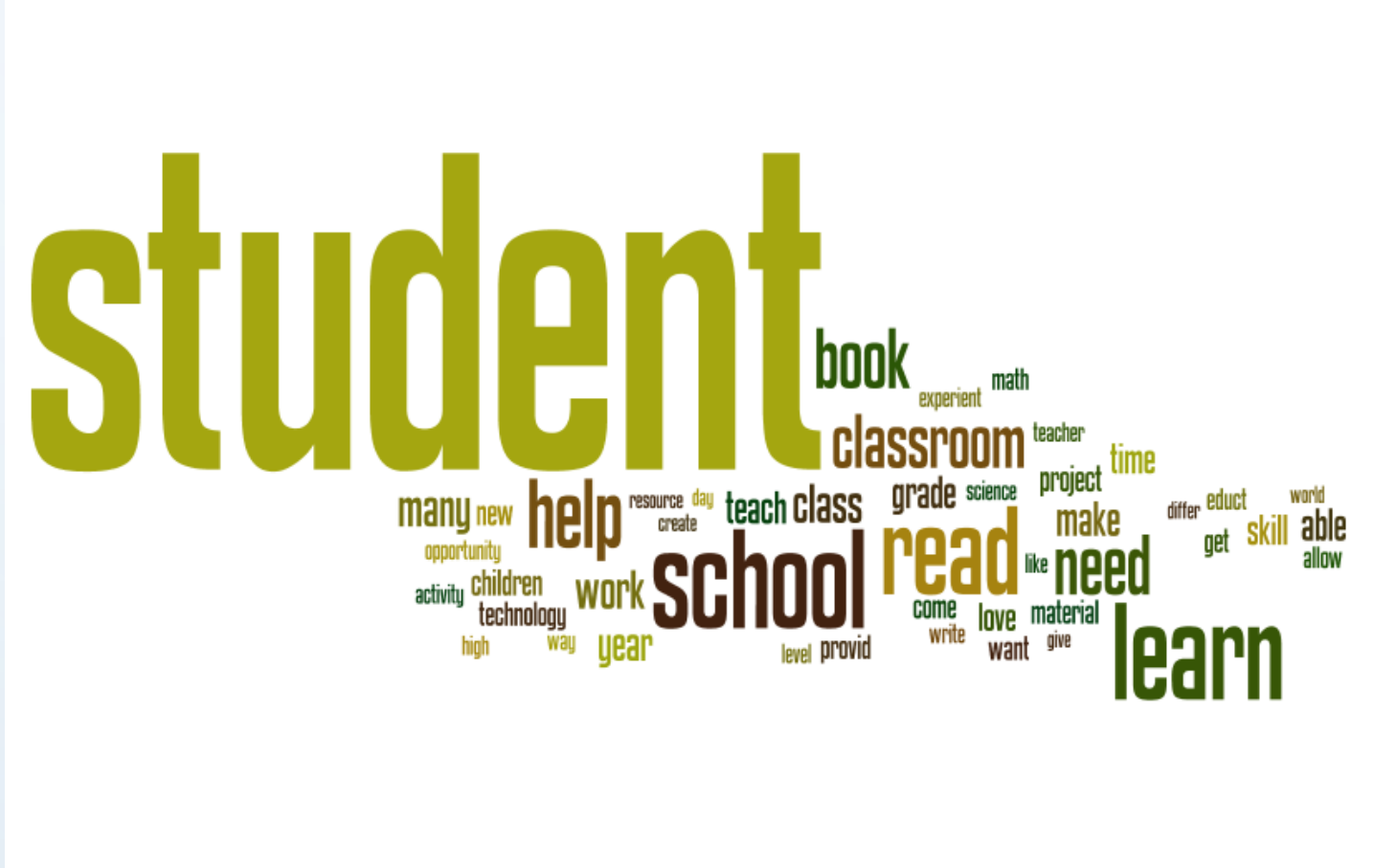
Get More Donations for Classrooms in Need

Result Overview – Fully Funded Projects by Resource type

Most needed resources are
supplies and technologies



Describe Your Project – Most Powerful Words for All Essays



student

learn

Help

book

all

would

work

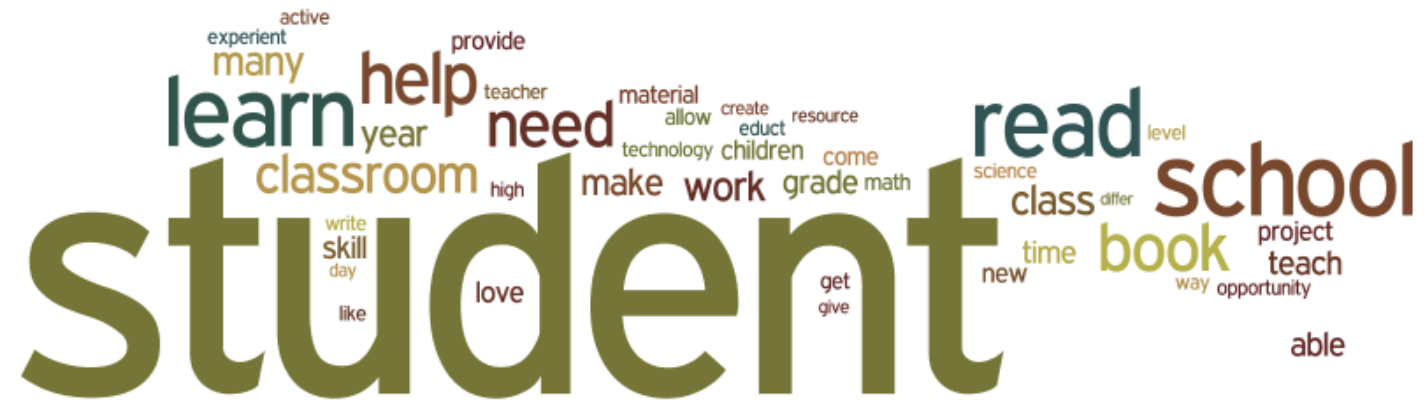
do

make

able

Get More Donations for Classrooms in Need

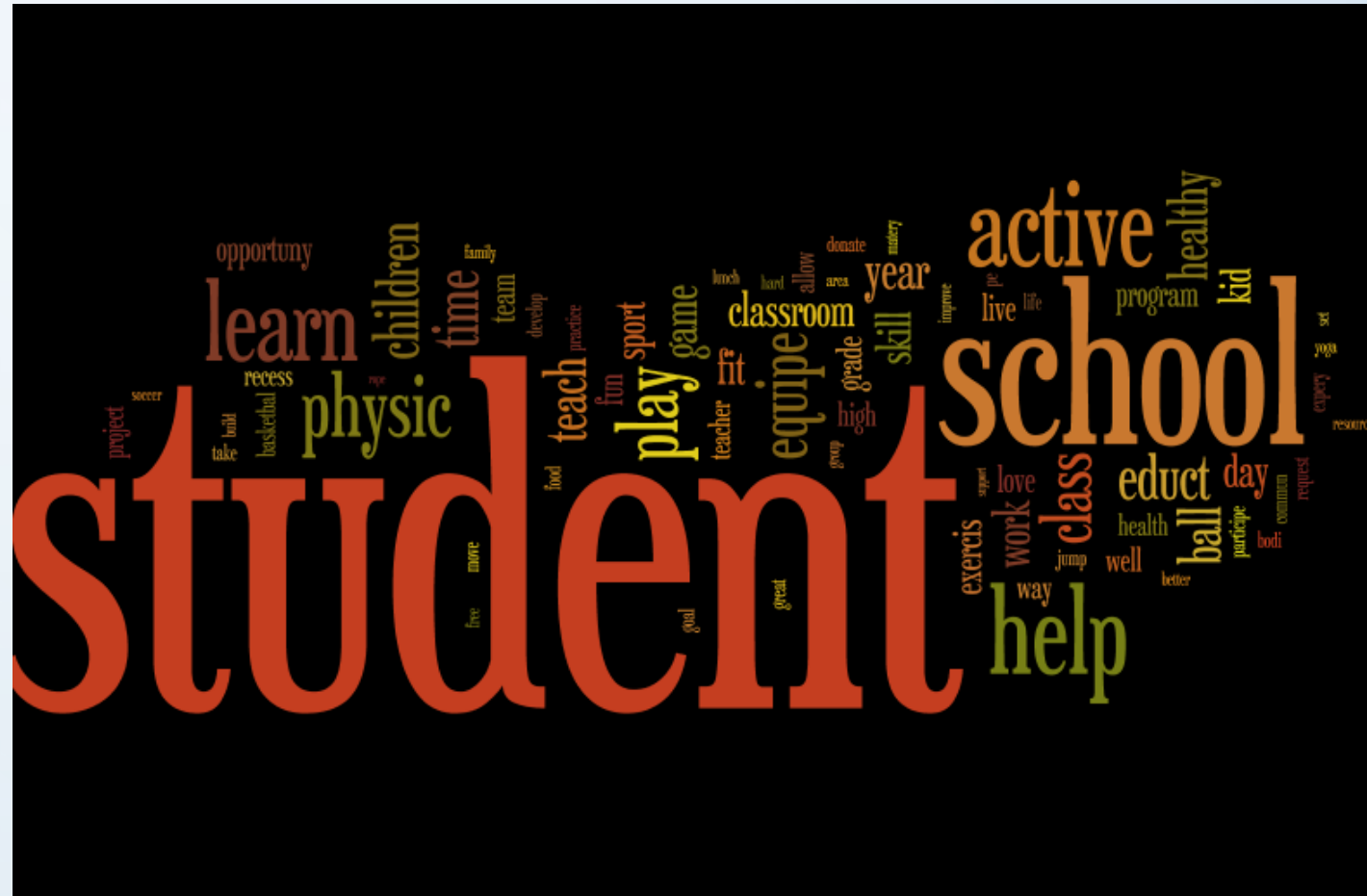
Describe Your Project – Fully Funded Essays



WORDS

student
school
learn
read
help
book
need
classroom
from
class
many
work
do
more
make
year
able
teach
grade

Describe Your Project – Sports & Health



WORDS

student

school

help

activity

learn

physic

play

equip

class

educate

time

children

teach

ball

healthy

year

work

game

skill

Get More Donations for Classrooms in Need

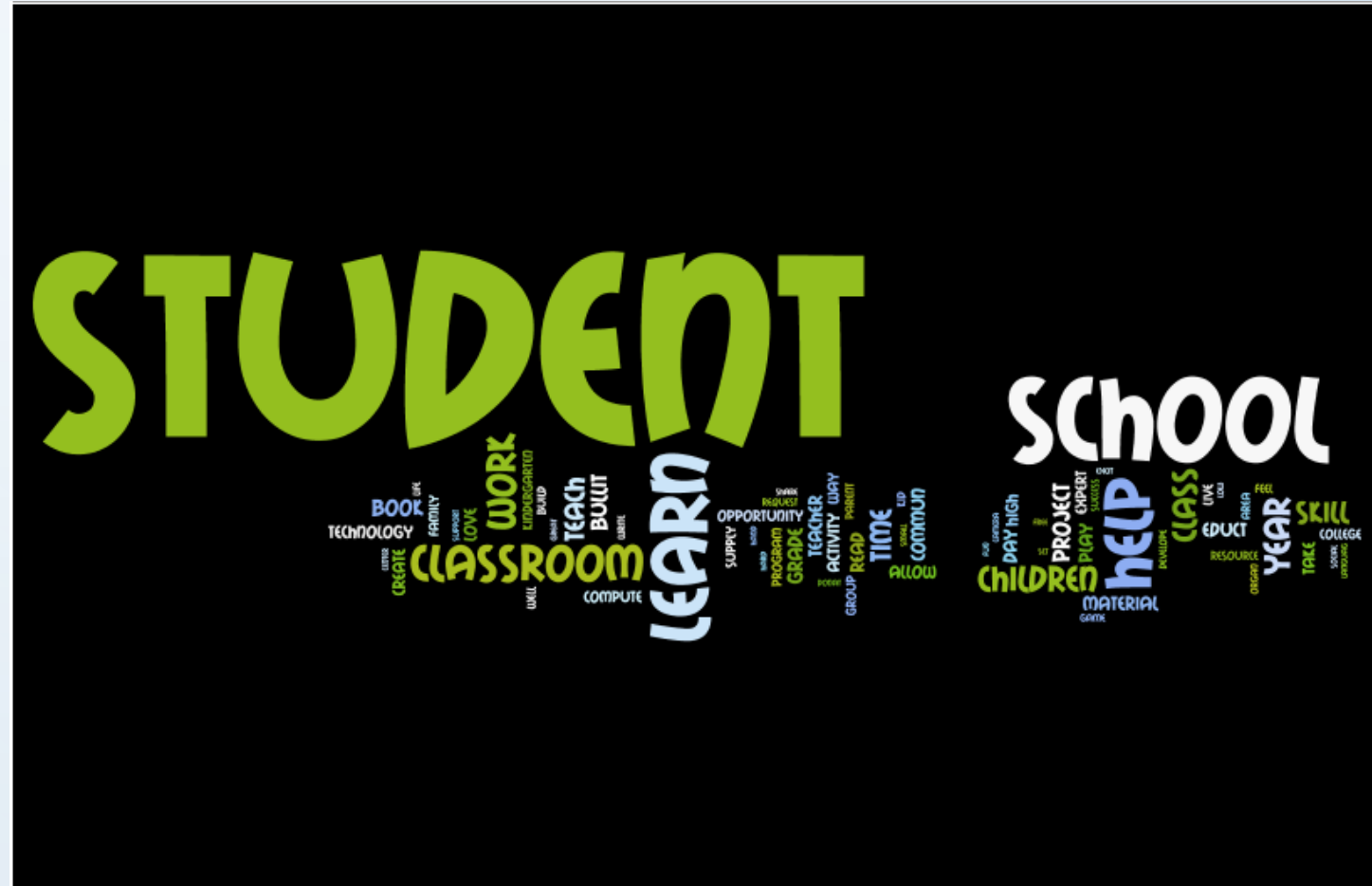
Describe Your Project – History



WORDS
student
school
learn
help
world
classroom
read
history
class
book
study
teach
project
year
grade
work
time
social
resource

WORDS
student
school
learn
help
world
classroom
read
history
class
book
study
teach
project
year
grade
work
time
social
resource

Describe Your Project - Applied Learning



WORDS
student
school
learn
help
classroom
work
year
children
class
time
skill
teach
project
book
bully
grade
material
day
read

Get More Donations for Classrooms in Need

Describe Your Project – Literacy



WORDS

student

read

book

school

learn

help

classroom

love

class

work

grade

year

write

time

skill

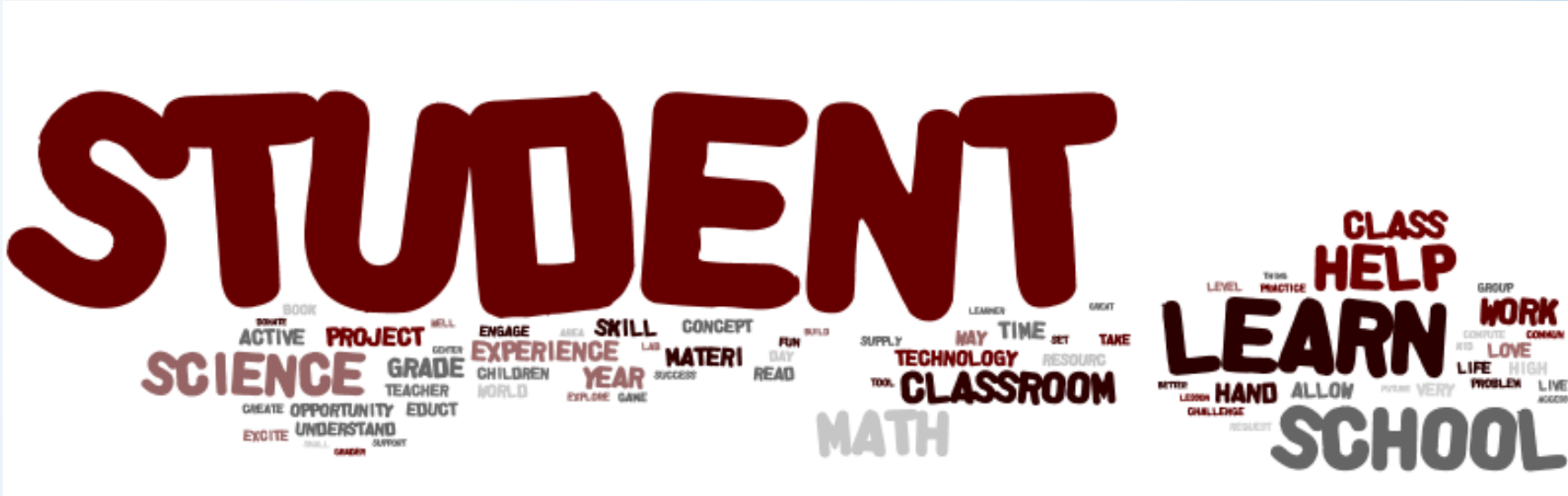
level

teach

Language

children

Get More
Donations for
Classrooms in
Need
Describe Your
Project – Math
and Science



WORDS	year
student	experience
learn	grade
school	hand
math	materi
help	project
science	skill
classroom	time
class	technology
work	activity

Describe Your Project – Music & Arts

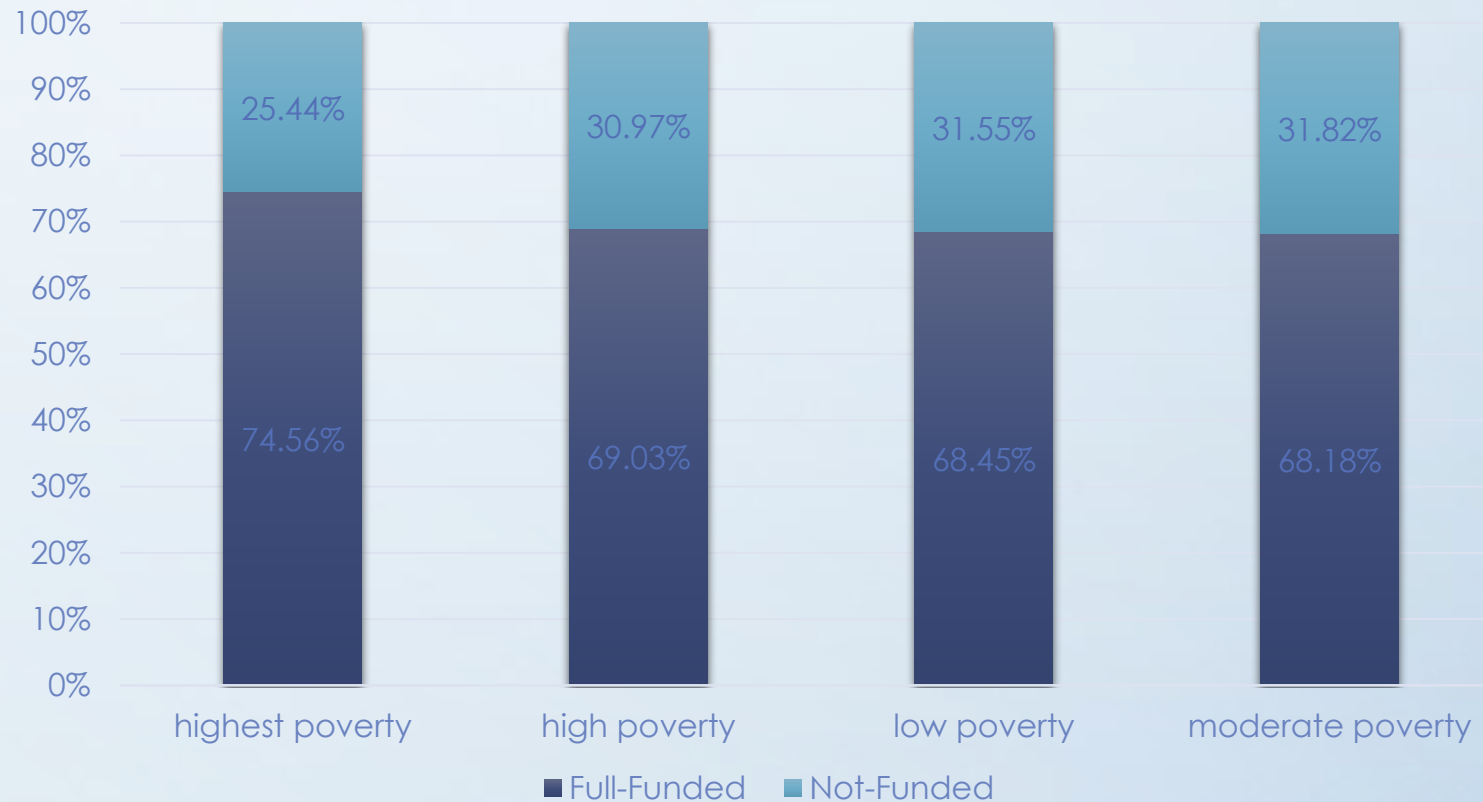


Get More Donations for Classrooms in Need

The Odds – more poor, more money?

1. Area with highest poverty receive more money than the rest.
2. Difference among the rest of the projects is not obvious

Poverty level vs. Funded Status

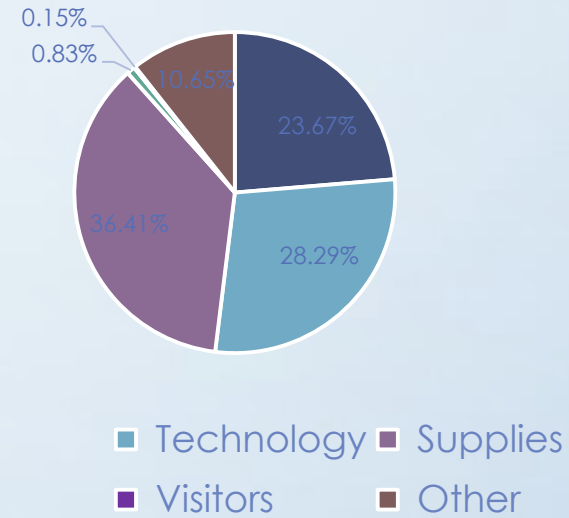
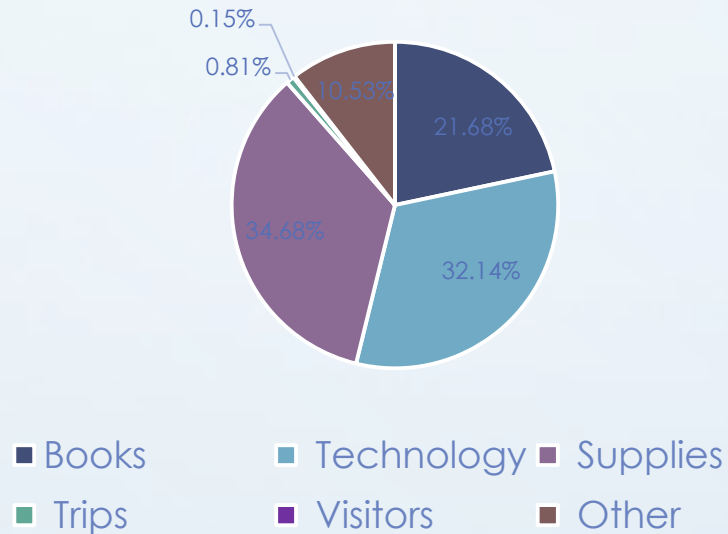


Get More Donations for Classrooms in Need

The Odds – What Resources Are in Need?

Total Project Listed Total: 646695

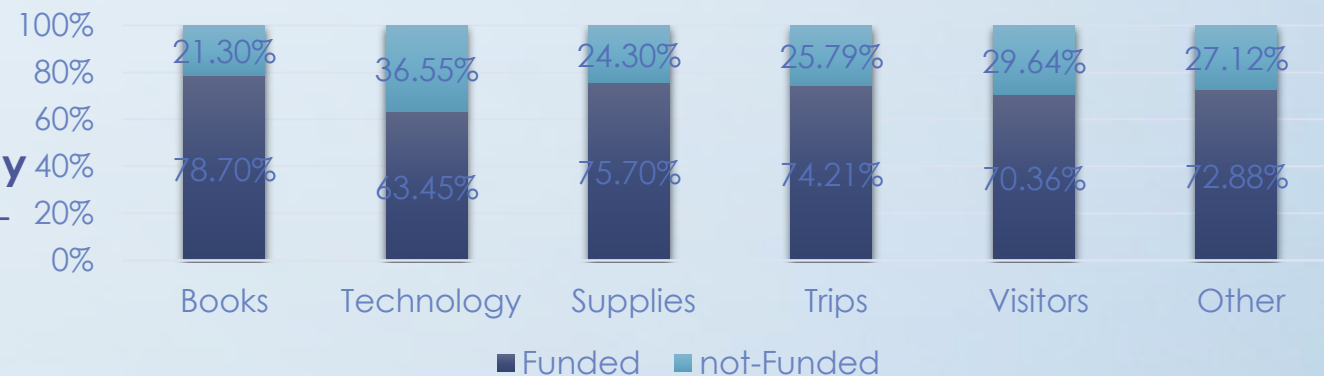
Full-Funded Project Total: 455769



Full-Funded Status V.S. Resource Requested Type

The resources requested most are **Supplies** and **Technology**.

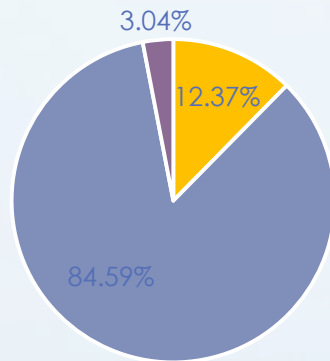
However, we found the **technology** request get less chances to be full-funded



Get More Donations for Classrooms in Need

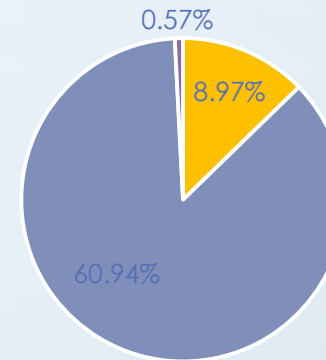
The Odds – Does Teacher Gender Matter?

Total Projects: 646695



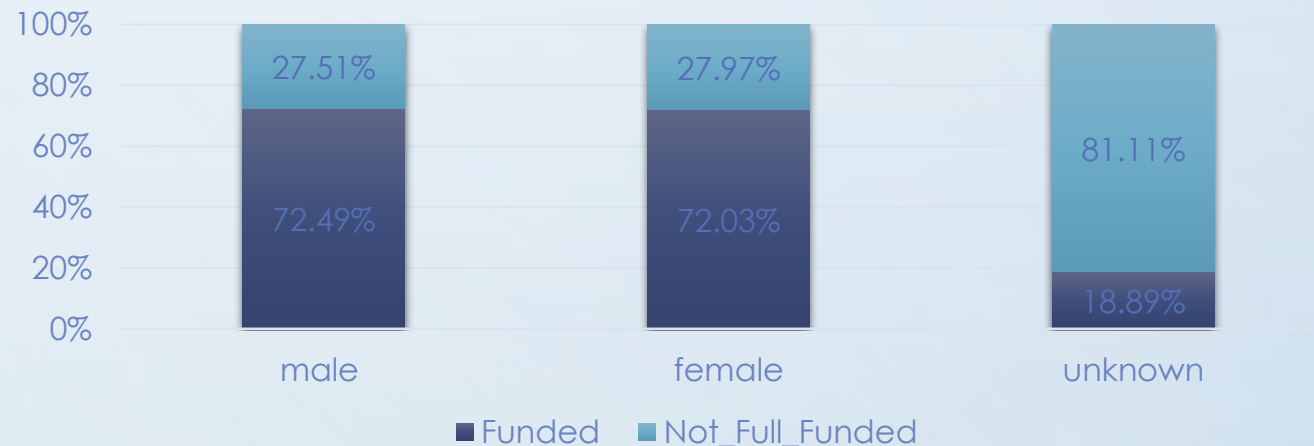
■ male ■ female ■ unknown

Full-Funded Projects: 455769



■ male ■ female ■ unknown

Fully Funded Status V.S. Sexual



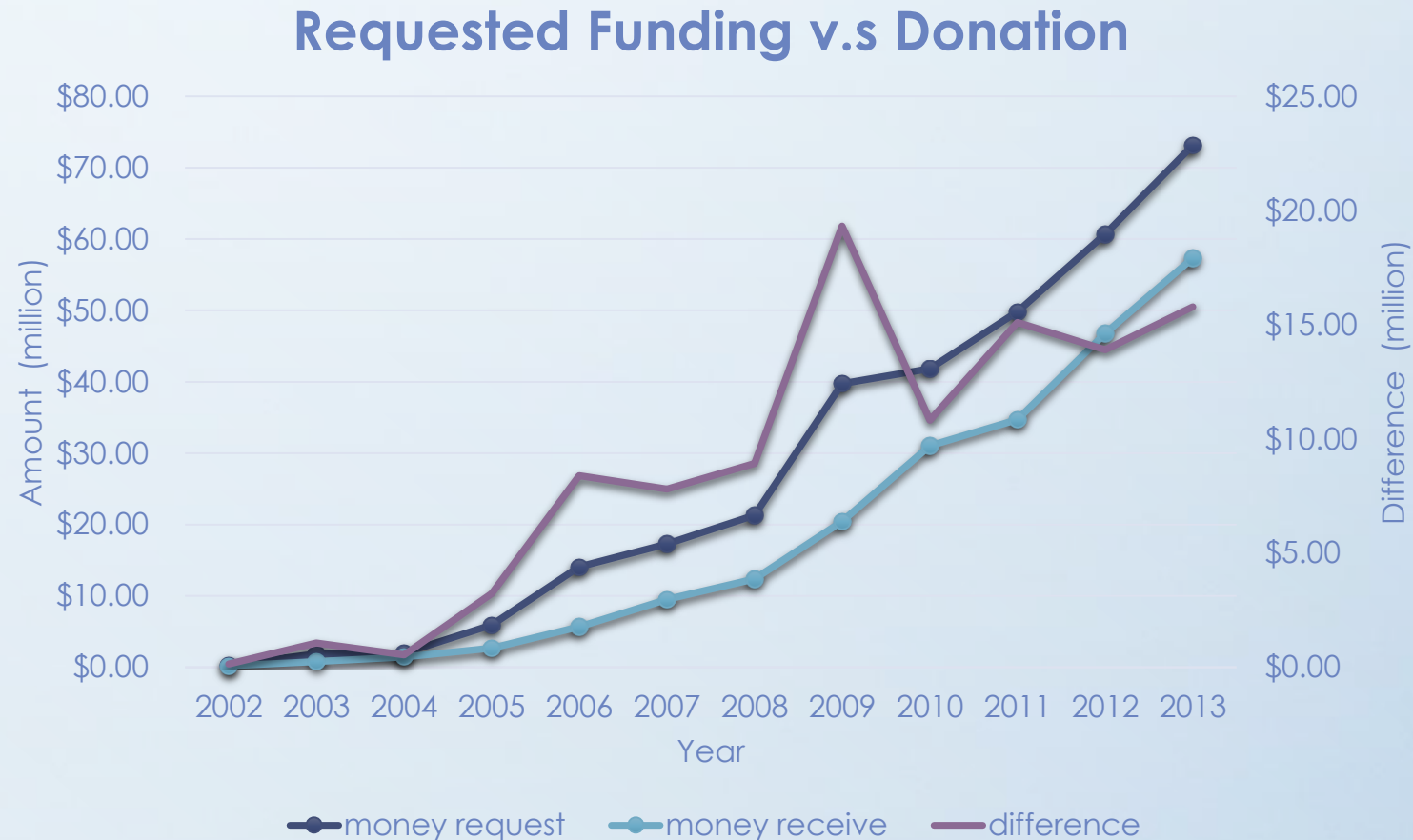
Most project are requested by **female teacher**.

Male teacher get more projects being fully funded.

Get More Donations for Classrooms in Need

The Odds – Will we get more money if request more?

money request and funding received are both increasing year by year.
But the gap between them is **ALSO INCREASING!**



Get More Donations for Classrooms in Need Prediction – you odd to get fully funded

Random Forest Model 22 Features

school_longitude	teacher_prefix
school_city	teacher_teach_for_america
school_state	teacher_ny_teaching_fellow
school_zip	primary_focus_subject
school_county	resource_type
school_charter	poverty_level
school_magnet	grade_level
school_year_round	total_price_excluding_optional_support
school_nlns	students_reached
school_kipp	

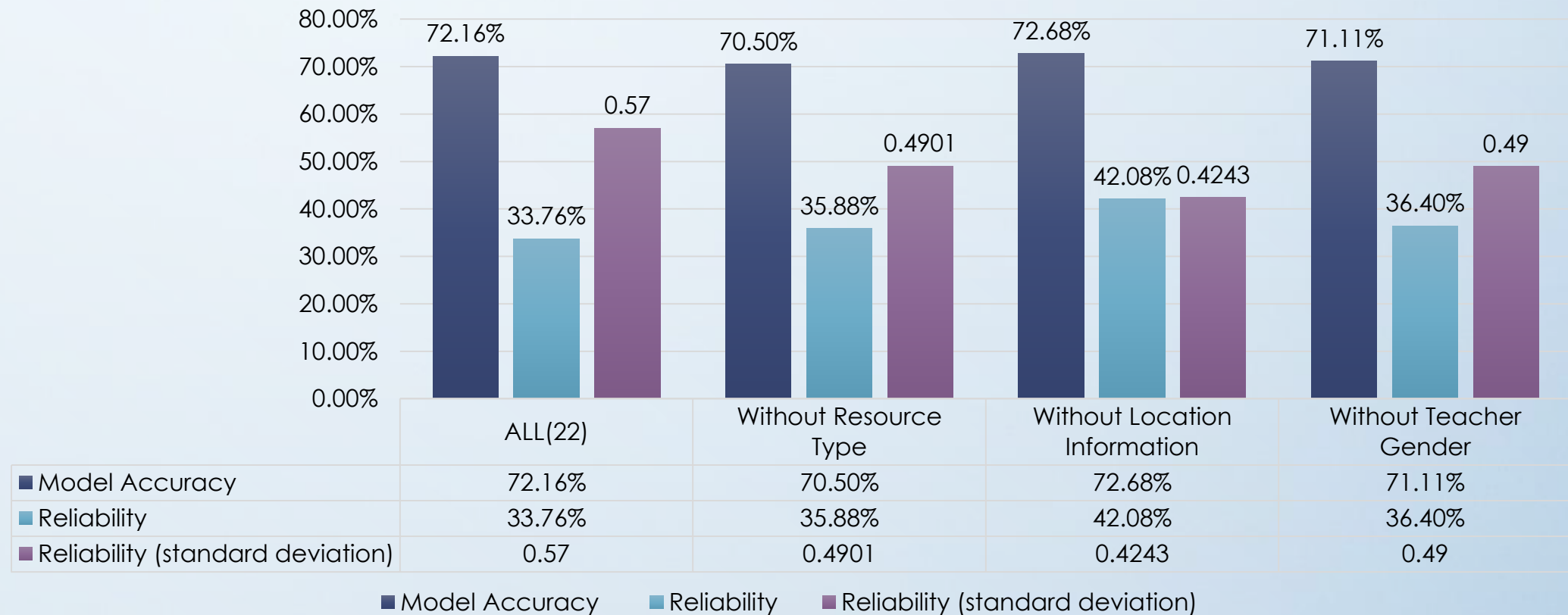
Training and Testing dataset

- Training Sample 565,929 (90%)
- Testing Data 62881 (10%)

Get More Donations for Classrooms in Need

Prediction – you odd to get fully funded

PredictionPerformance with Different Features



As we observed, Teacher gender, resource type really matter. But why doesn't location matter?

Get More Donations for Classrooms in Need

Conclusion

- 1. Online donation behaviors are influenced by the resource requested, teacher gender
- 2. Some describing words may increase a project's chance to get more funding. Those words includes help, need, students, etc.
- 3. Education resource are even unfairly distributed online. There are more project requested in East Coast rather than West Cost.

Acknowledgements

- Thank HPC for their support, and we are really appreciated the free space and powerful machine provided from NYU HPC.

Get More Donations for Classrooms in Need

References

- Should teachers have to beg online for funding for classroom supplies?
- <http://www.donorschoose.org/about/impact.html>
- Why We're Partnering With DonorsChoose
- Beyond School Supplies: How DonorsChoose Is Crowdsourcing Real Education Reform

Get More Donations for Classrooms in Need

Obstacles

- HPC cluster crashed a lot.
 - once the data are already processed on the HPC cluster, the crash will cause some schedule delay.
- PCA in Mahout usage problem
 - Originally we want to analyze what are the significant factors that effects a donator's choice over the many projects. Because lacking of machine learning related background and less PCA related information online, we had problem to quantify all our data into meaningful parameter and fitted into PCA modeling.

Thank You !

