Predicting Success of Kickstarter Projects

Ji Ho Lee

What is Kickstarter?

- Online crowdfunding platform
- Projects are more focused towards creative mediums
  - Music, video games, stage shows, etc.
- **$4.6 billion** raised for over 445,000 projects, as of 2019
  - **~60% of projects failed** to be successfully backed

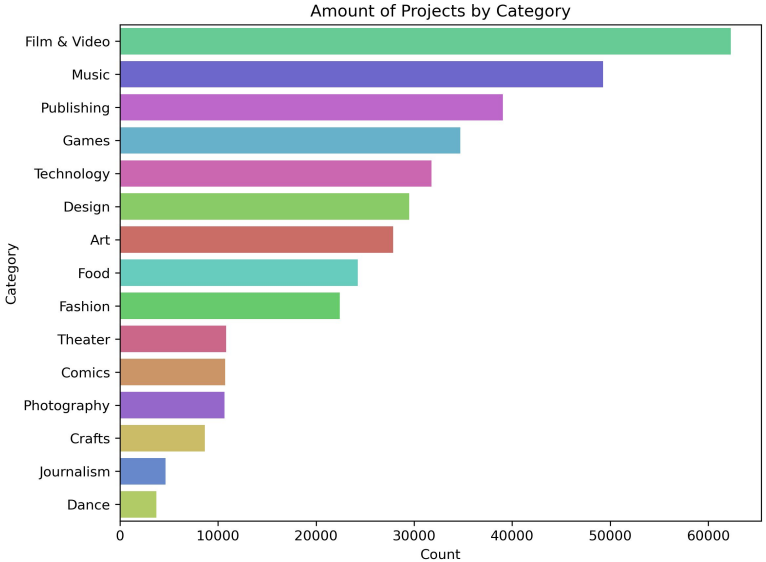
Purpose

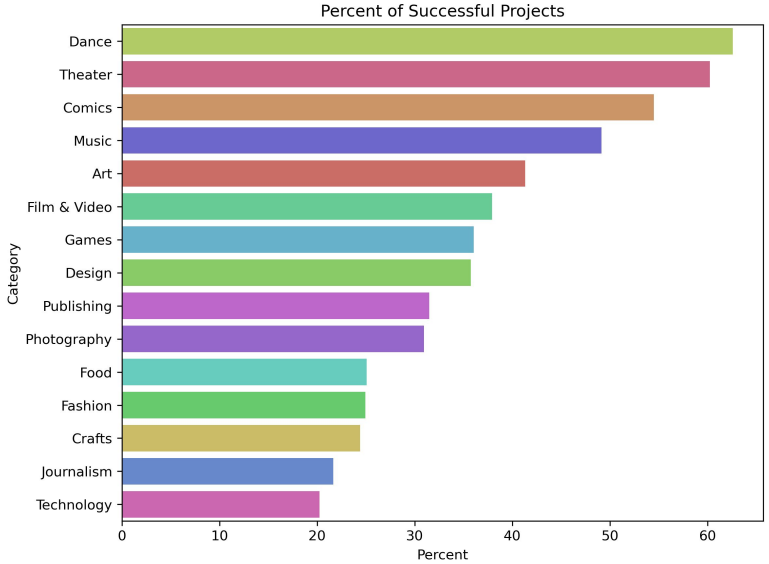
- Predict whether a Kickstarter project will be successfully funded
- Parties of interest:
  - Project backers
  - Project owners
  - Kickstarter and similar crowdfunding companies

What about the data?

- Contains ~378,000 past projects from 2009 to 2018
  - 16 features (ID, name, category, country, launched, deadline, goal, state, etc.)
- Cleaned and pre-processed for modeling
  - Removed erroneous values
  - Dropped outliers (3 standard deviations), log-scaled and standardized
  - Removed redundant columns
  - Removed ‘backers’ and ‘pledged’ columns to increase generalizability

What does the data look like?





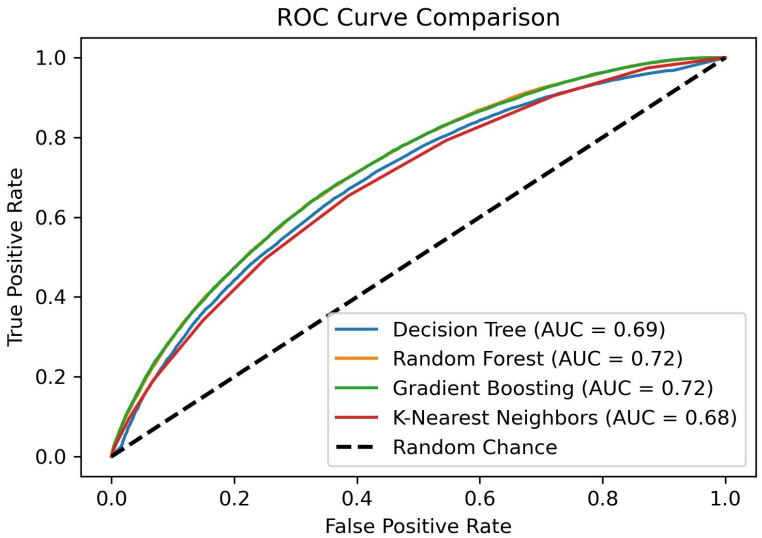
Modeling Overview

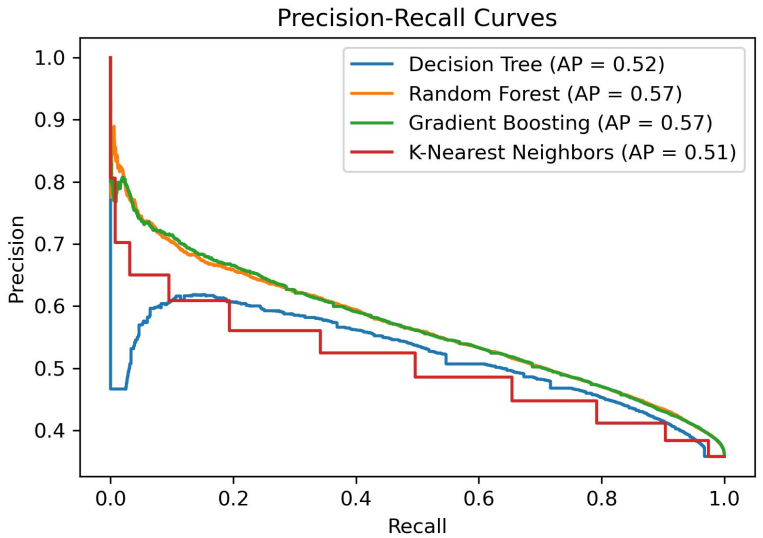
- Supervised Learning (use of labelled data to make predictions)
- Binary classification: 0 - failed to get funding, 1 - successfully funded
- Tools: scikit-learns, Pandas, Numpy

How do we make predictions?

- Tested 4 models (Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors)
- Precision as main evaluation metric
  - Minimize number of false positive predictions
- Hyperparameter tuned using RandomizedSearchCV

Which model do we pick?





Which model do we pick?

	Accuracy	Balanced Accuracy	Precision	Recall
Decision Tree	0.6758	0.6079	0.5730	0.3692
Random Forest	0.6872	0.6128	0.6091	0.3512
Gradient Boosting	0.6865	0.6099	0.6113	0.3406
K-NN	0.6686	0.5963	0.5605	0.3419

Further Improvements?

- Improve hyperparameter tuning
  - Constrained hyperparameter search due to lack of computing power
- Implement creatively engineered features
  - Could use name of project, utilize NLP to provide sentiment of project names/descriptions, etc.