Modeling Kickstarter Success

Kevin Velasco
Sherry Yang

Flatiron School
May 31, 2019

Roadmap

01 Business Application

02 Approach

03 Data Exploration

04 Data Cleaning & Model Creation

05 Evaluating Multiple Models

06 Next Steps

01 - Business Application

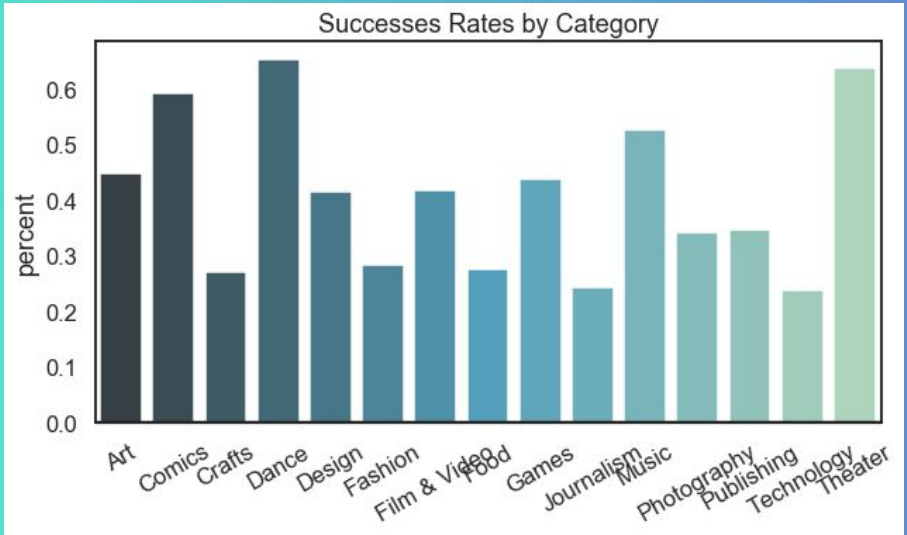
As of April 2019, $4.2bn has been crowdfunded for projects through Kickstarter. The platform has a 36.84% success rate for projects with “games” as the most popular project category.

We are offering a beta model that can produce a likelihood estimation for whether a project will be successful on kickstarter.

02 - Approach

- **Data Exploration**
- **Data Cleaning & Encoding**
- **Create a pipeline**
- **Create Model 1 using Logistic Regression**
- **Create Model 2 using Random Forest Classifier**
- **Score Model 1**
- **Score Model 2**
- **Compare performance of models relative to our client needs**

03 - Data Exploration



378, 661
data entries

Data from 2018 projects

Pledged to goal: 23.9%

Total goal asked in usd: $13.86bn

Total pledged in usd: $3.3bn

Raw Features

- ID
- name
- category
- main_category
- currency
- deadline
- goal
- launched
- pledged
- state
- backers
- country
- usd pledged
- usd_pledged_real
- usd_goal_real

03 - Data Exploration

Ordered by absolute number of successful projects

1. **United States - 41.8% success rate - 109, 299 successful projects**
2. **Great Britain - 41% success rate - 120, 67 successful projects**
3. **Canada - 33.4% success rate - 4, 134 successful projects**
4. **Australia - 30.4% success rate - 2, 010 successful projects**
5. **Germany - 27.3% success rate - 937 successful projects**
6. **France - 36% success rate - 908 successful projects**
7. **Netherlands - 25.6% success rate - 617 successful projects**
8. **Sweden - 33.7% success rate - 509 successful projects**
9. **Spain - 26% success rate - 492 successful projects**
10. **New Zealand - 35% success rate - 448 successful projects**
11. **Italy - 18.5% success rate - 439 successful projects**

03 - Data Exploration

Ordered by most pledged within successful projects

1. United States - Main Category: Games, Sub-Category: Product Design, Tabletop Games
    a. Main Category Magnitude: $568 M
2. Great Britain - Main Category: Games, Sub-Category: Tabletop Games
    a. Main Category Magnitude: $58 M
3. Canada - Main Category: Design, Sub-Category: Product Design
    a. Main Category Magnitude: $22 M
4. Australia - Main Category: Design, Sub-Category: Product Design, Tabletop Games
    a. Main Category Magnitude: $12 M
5. Germany - Main Category: Technology , Sub-Category: Product Design/Hardware
    a. Main Category Magnitude: $13 M
6. France - Main Category: Design, Sub-Category: Product Design, Tabletop Games
    a. Main Category Magnitude: $9.6 M

04 - Data Cleaning/Engineering

- **Split into training and testing data**
- **Encode sub-categories to binary coding**
- **Translate project start and end times into number of days**
- **Remove projects from our data that were canceled, undefined, suspended, live**

04 - Model Creation

- **What are the most important features?**
- **What effect did boosting vs random forest vs logistic regression have?**
- **Is there a benefit to tuning our model parameters?**
- **Ultimately we chose to use the XG Boost classification algorithm**

04 - Model Prediction

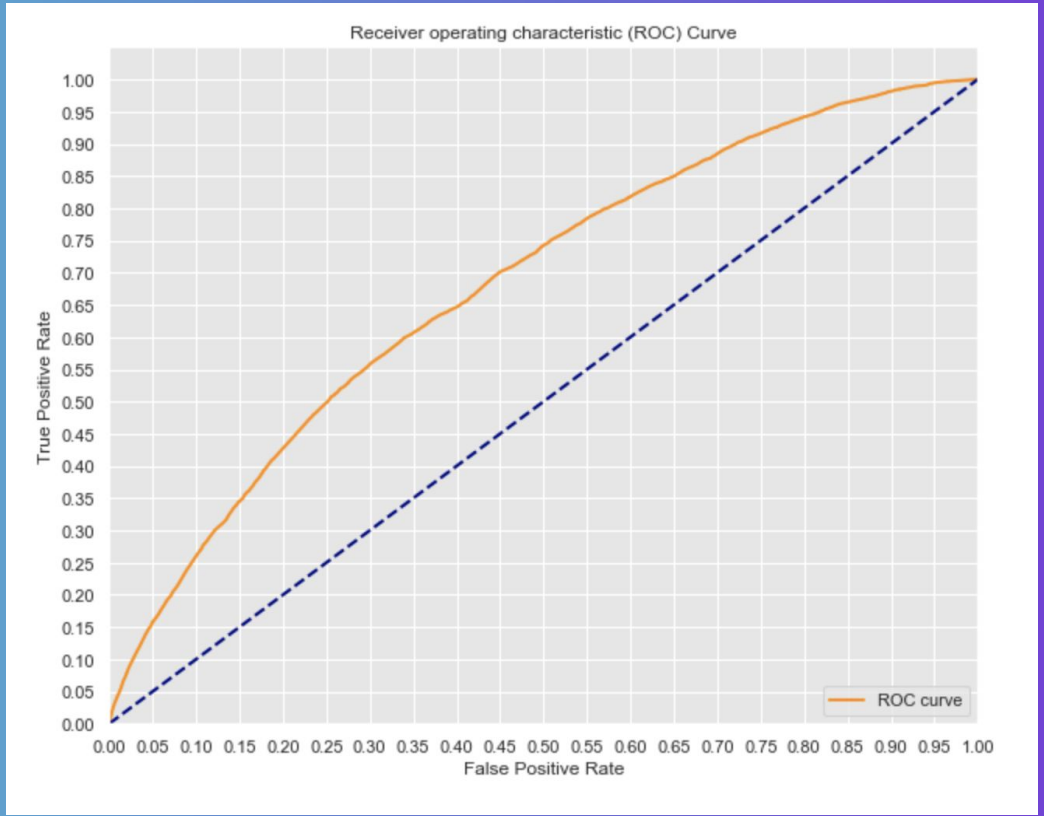
- **Case 1: A project started in the US with a length of 31 days in the game category with the goal of $20,000 has a probability of succeeding 52% which is better than the overall 36%**

- **Case 2: A project started in the US with a length of 29 days in the technology category with the goal of $7,000 has a 26% probability of succeeding compared to the overall 36%**

- **Case 3: A project started in Australia with a length of 29 days in the technology category goal of $26,100 has a 10% probability of succeeding compared to the overall 36%**

05 - Evaluating Multiple Models

- **F1 Score**
- **ROC - AUC**
- **Log Loss**



06 - Next Steps

- **We recommend collecting data in terms of the marketing tactics and existing network connections that projects have at the start of a campaign to better measure likelihood of success**

- **We recommend collecting data on the types of individuals who fund successful kickstarter campaigns — what are their demographics — and using that as an assessment of whether a project's audience will match kickstarter's audience**

THANKS!

Does anyone have any questions?