
RESCUING KICKSTARTER CAMPAIGNS

PREPARED BY | Nateé Johnson &
Misha Berrien

OVERVIEW

- I. PROBLEM STATEMENT
- II. CURRENT STATE OF CAMPAIGNS
- III. OVERVIEW OF APPROACH
 - A. Feature Selection
 - B. Model Selection
- IV. FINAL MODELS
- V. “LIVE” PREDICTION
- VI. CONCLUSION & NEXT STEPS

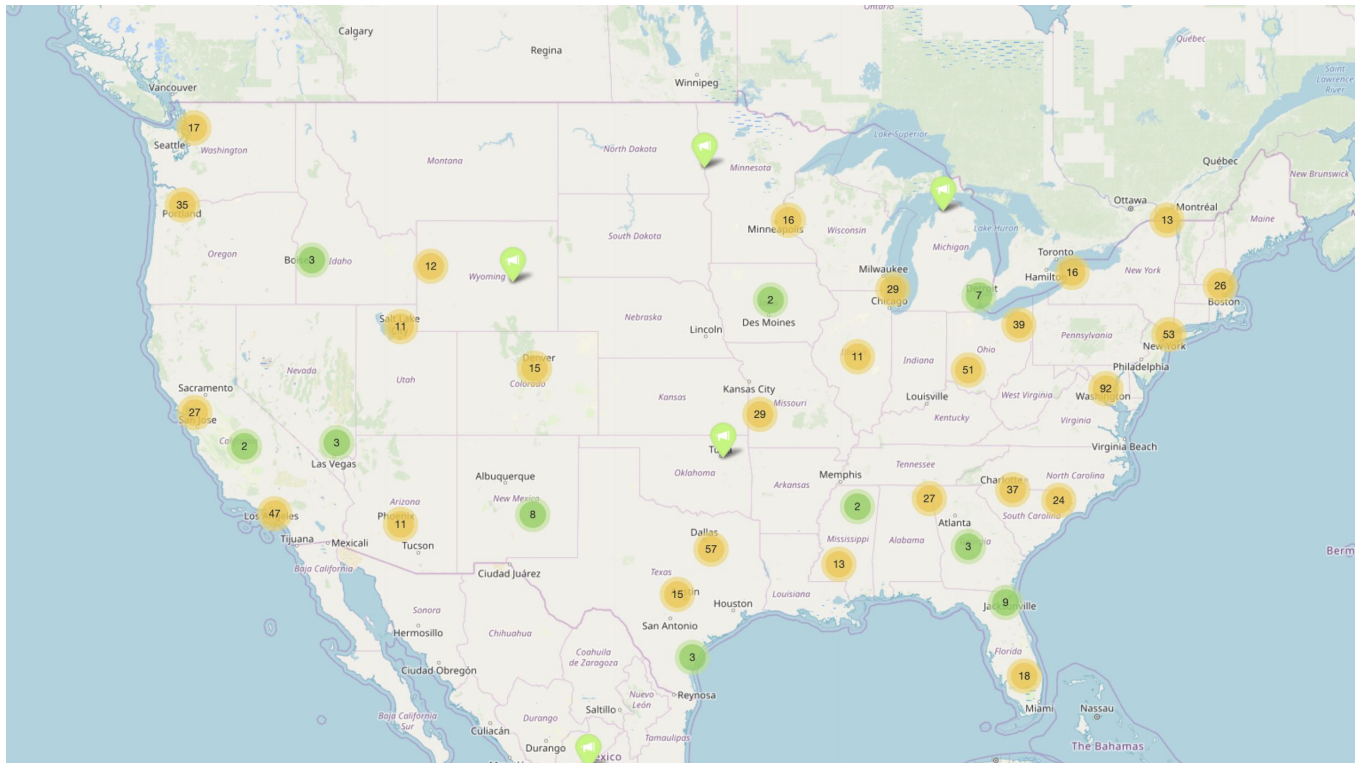
INTRODUCTION

THE PROBLEM

Kickstarter earns 5% commission on projects that are successfully funded. Currently, less than 40% of projects on the platform succeed. The objective is to predict which projects are likely to succeed so that these projects can be highlighted on the site either through 'staff picks' or 'featured product' lists.

CURRENT STATE OF KICKSTARTER CAMPAIGNS

SHORT DEMO:



INTRODUCTION

OUR SOLUTION

- 01 Predict Successful Campaigns and promote those with the lowest predicted probability of being successful.
- 02 Contact creators from those campaigns that are just below the “success” margin and give them insights that will help them succeed.

APPROACH OVERVIEW



FEATURE SELECTION

Original dataset featured 42 columns. Many of our end features had to be created. Our features fall into three categories.

01 DROPPED - Columns that created data leakage
ex. Number of backers & amount raised.

02 IMPUTED - Features that were derived from columns in the original dataset.
ex. Campaign length & goal amount in USD

03 CREATED - Columns that were created from analysis.
ex. campaign clusters

MODEL SELECTION - STATS

6

Machine learning
algorithms tested

*Decision Tree/ Random Forest/ Logistic
Regression/ Gaussian Naive Bayes/
Bernoulli Naive Bayes/ AdaBoost*

2K+

Models Run

CHOSEN MODELS

Adaboost chosen for prediction
Logistic chosen for prescription

<i>Learning Algorithm</i>	<i>Decision Tree</i>	<i>Random Forest/</i>	<i>Logistic Reg.</i>	<i>Gaussian Naive Bayes</i>	<i>Bernoulli Naive Bayes</i>	<i>AdaBoost</i>
Accuracy Score (Test Data)	70.4%	72.4%	75.2%	72.1%	73.6%	76.6%

FINAL MODELS



CHOSEN MODELS STATS

**ADABOOST CLASSIFIER:
CONFUSION MATRIX**

	Predicted 0	Predicted 1
Actual 0	25524	6741
Actual 1	9407	27239

**LOGISTIC REGRESSION:
CONFUSION MATRIX**

	Predicted 0	Predicted 1
Actual 0	25657	6608
Actual 1	10494	26152

PREDICTIONS ON 'LIVE' DATA



LIVE PREDICTIONS

	PREDICTED	ACTUAL
'Full Body Bar - Get in a Full Body Workout Anywhere, Anytime',	0	1
'9 to 5 das Musical',	1	0
Bacon. Real organic farm raised bacon, Stage 1/3'	1	0
Short Film: Leave This'	1	0
When They Walk	1	1
Song of Freedom Choral Festival	0	0
Weaponized Cats Comic #1	0	0
Super Cache	0	1
A Lie Agreed Upon: The Deadwood Chronicles	0	1
Meal Plans On Dishware	0	0

ACCURACY ON
LIVE DATA

4/10

CONCLUSION & NEXT STEPS

- We have around a 70% + accuracy rate for our test data.
- Improve model by adding additional data (meta data)
- Build our prescriptive system with logistic regression

THANK YOU!

QUESTIONS?