

The Kickstarters



CMPE 256 2019F Project Presentation

Presented by:

Xunan Dai

Outline

- Introduction
- Algorithms
- Data Analysis
- Data preprocessing
- Evaluation
- Conclusions

- What is Kickstarter and What Do People Use it for?
 - Fundraising platform for entrepreneurial projects
 - Entirely driven by crowdfunding
- Two Types of Users:
 - Creators - start a project hope to raise funds from backers
 - Backers - have the opportunity to receive rewards
- 'All or Nothing' Rule - Defines if project is successful
 - A creator can only collect the funds if the funding goal has been reached by the deadline (pledged \geq goal).

- Question: What Makes a Successful Kickstarter Project?
 - Not all projects have succeeded
- Kaggle Dataset Generated by Web Robots
 - Range from 2009 to 2017
 - 56MB with 99,035 project records
 - 54 attributes with all kinds of categories
 - Location, time, name, pledge, goal, etc.,
- Need to Reduce Features to Generalize the Dataset
 - Some important features contribute to success!

Motivation

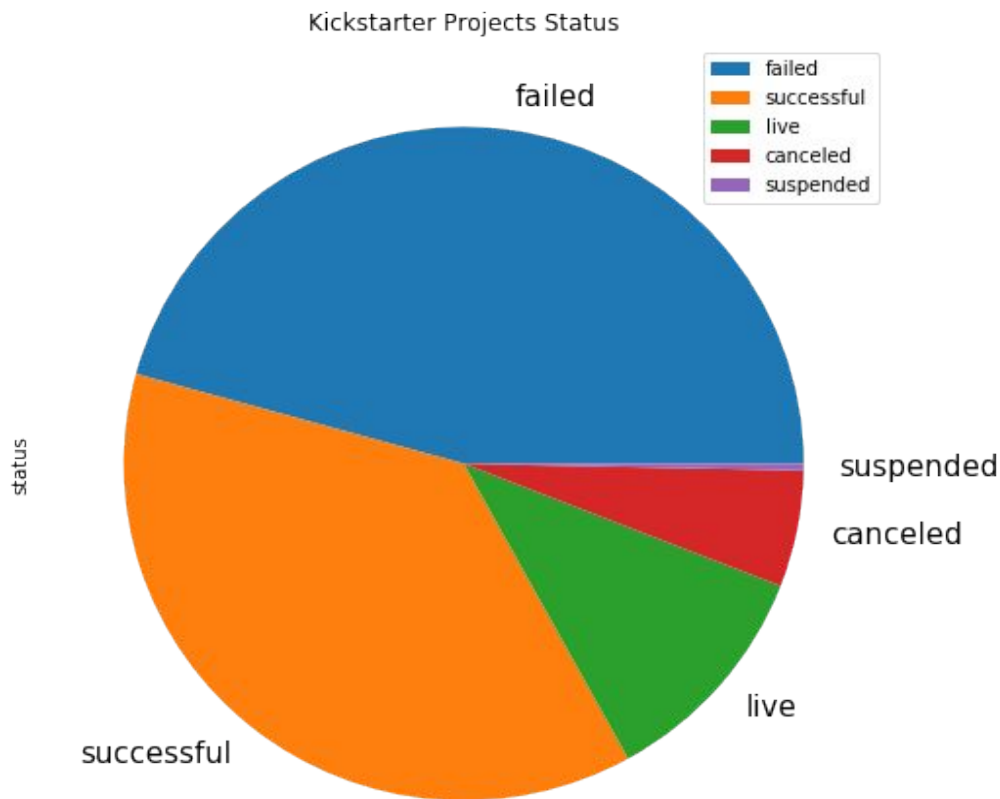
Algorithms

Depends on the Target Field

- Project's Status
 - Successful, Failed
- ❖ Classification Problem:
 - SVM
 - efficient in high dimensional data
 - Decision Tree
 - handle both numerical and categorical data
- Amount of pledge in \$
 - pledgedUSD
- ❖ Regression Problem:
 - KNN-Regression
 - works well in high non-linear data
 - MLPR
 - generate non-linear function approximator

Data Analysis– Status Distribution

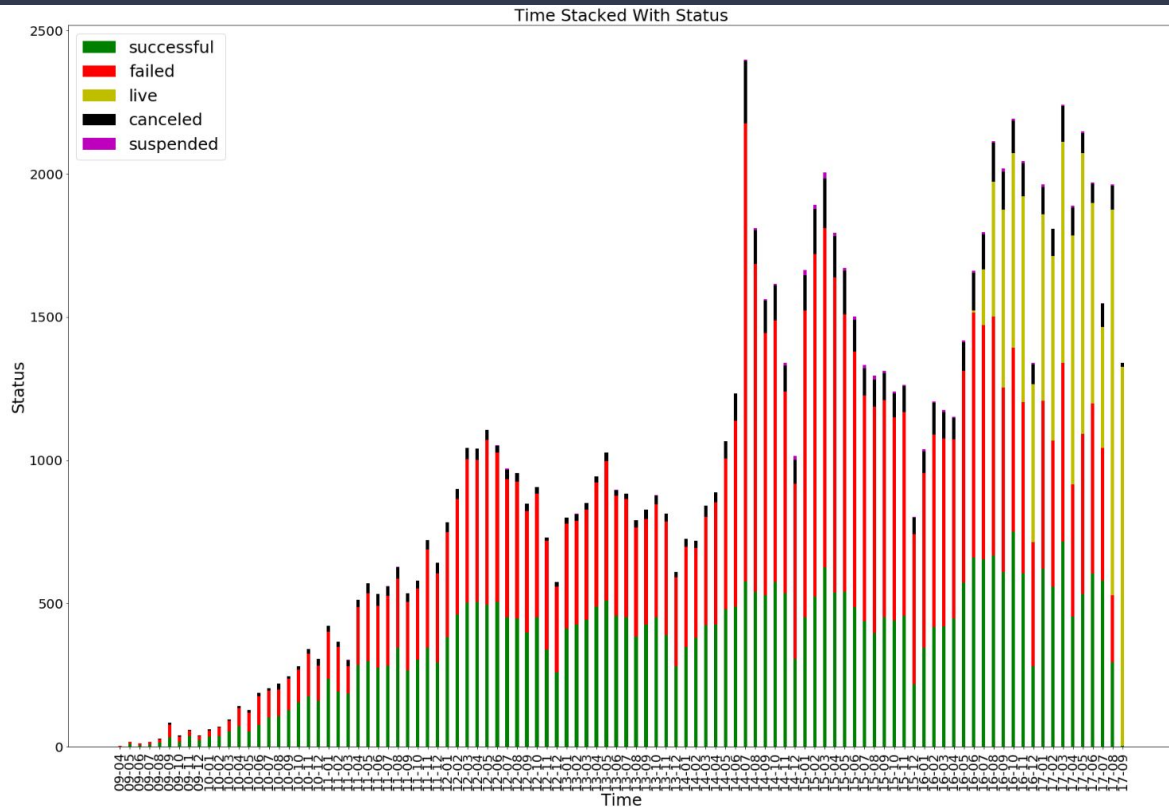
- Around same percentage towards success and failure
- Many canceled in the process



Data Analysis– Projects' Time Distribution

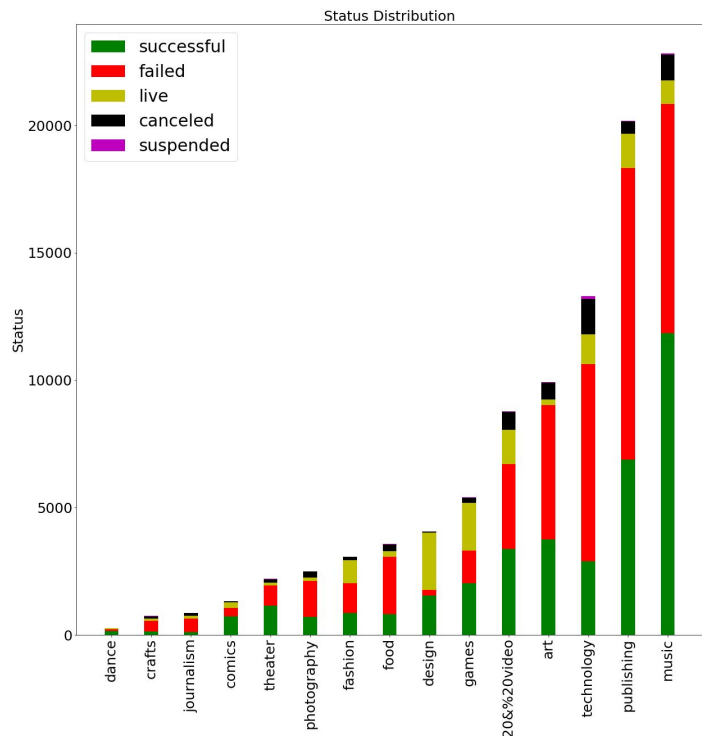
- From 2009 to 2017
- By month

Project launch time is
“irrelevant” to a project’s
success

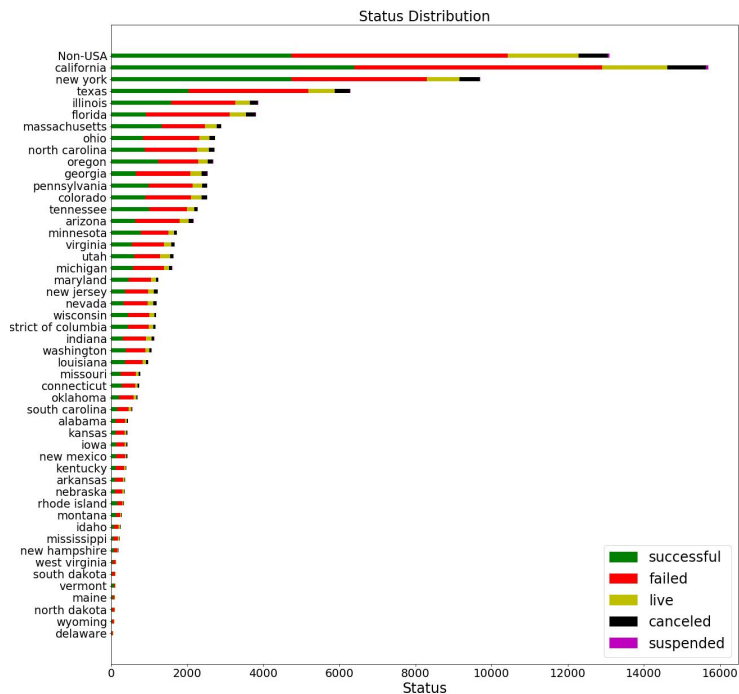


Data Analysis– Status Distribution

By Category

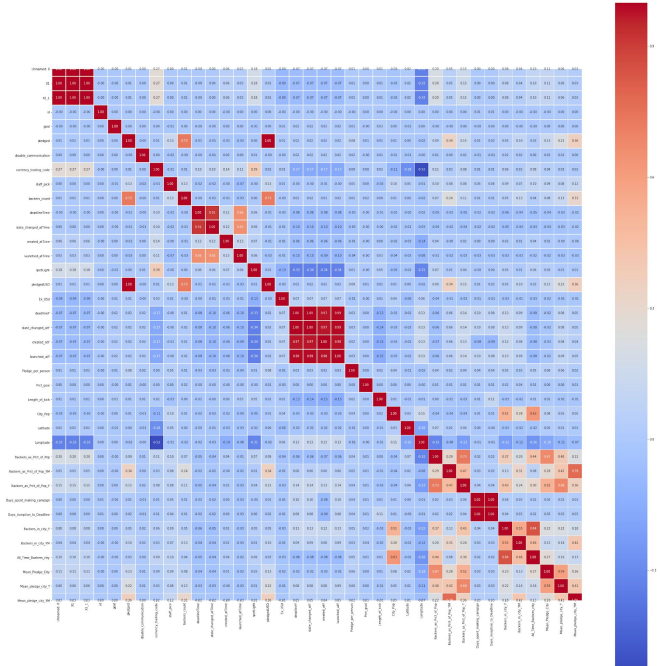


By States of the U.S.



Data Preprocessing

- Data Cleaning
 - Missing value replacement
- Feature Selection
 - Correlation Map
 - Correlation of each pair
 - Drop all “highly correlated” attributes
 - Combine Features
 - New Features Generation
 - Encoding Features
 - Categorical to numerical
 - Standardization
 - Z-score



Evaluation (Classification)

SVM:

50.9% accuracy (normal)

SVM classification report					
	precision	recall	f1-score	support	
0	0.50	0.22	0.31	8862	
1	0.51	0.79	0.62	9147	
accuracy			0.51	18009	
macro avg	0.51	0.50	0.46	18009	
weighted avg	0.51	0.51	0.47	18009	

Decision Tree:

98.2% accuracy (overfitting)

	precision	recall	f1-score	support	
0	0.98	0.98	0.98	8862	
1	0.98	0.98	0.98	9147	
accuracy			0.98	18009	
macro avg	0.98	0.98	0.98	18009	
weighted avg	0.98	0.98	0.98	18009	

Evaluation (Regression)

KNN:

RMSE at 0.4923, MSE at 0.24244

```
Final rmse value is = 0.49238498987935636
```

```
Final mse value is = 0.24244297825849387
```

MLPR:

RMSE at 0.4994, MSE at 0.24945

```
Final rmse value is = 0.4994591341480117
```

```
Final mse value is = 0.24945942668388152
```

Conclusions

- Things Worked:
 - Visualizing dataset gives a general picture of each attributes
 - Selected models gives satisfactory prediction
 - Cross validation with proper model parameters
- Things Didn't Work Well:
 - Real evaluation of successful, failed prediction
 - Evaluation of feature selection and transformation

Q&A



Thank You!

