

Prediction of Kickstarter Projects

Daehyun Kim

Background



- The world's largest funding platform for creative projects
- Backers
- Goal amount vs. Pledged Amount
- Data from Kaggle

Business Question

- Can you predict whether a crowdfunding project will be successful before release?

Variables

```
df.head(5)
```

name	category	main_category	currency	deadline	goal	launched	pledged	state	backers	country	usd pledged	usd_pledged_real	usd_goal_real
the Songs of Adelaide & Abullah	Poetry	Publishing	GBP	2015-10-09	1000.0	2015-08-11 12:12:28	0.0	failed	0	GB	0.0	0.0	1533.95
Setting From arth: ZGAC rts Capsule For ET	Narrative Film	Film & Video	USD	2017-11-01	30000.0	2017-09-02 04:43:57	2421.0	failed	15	US	100.0	2421.0	30000.00
Where is Hank?	Narrative Film	Film & Video	USD	2013-02-26	45000.0	2013-01-12 00:20:50	220.0	failed	3	US	220.0	220.0	45000.00
oshiCapital Rekordz eds Help to Complete Album	Music	Music	USD	2012-04-16	5000.0	2012-03-17 03:24:11	1.0	failed	1	US	1.0	1.0	5000.00
Community ilm Project: The Art of ghborhoo...	Film & Video	Film & Video	USD	2015-08-29	19500.0	2015-07-04 08:35:03	1283.0	canceled	14	US	1283.0	1283.0	19500.00

- 378302 rows
- 16 variables
- Data from Kickstarter platform

Missing values

```
df.isna().sum()
```

ID	0
name	4
category	0
main_category	0
currency	0
deadline	0
goal	0
launched	0
pledged	0
state	0
backers	0
country	0
usd pledged	3797
usd_pledged_real	0
usd_goal_real	0

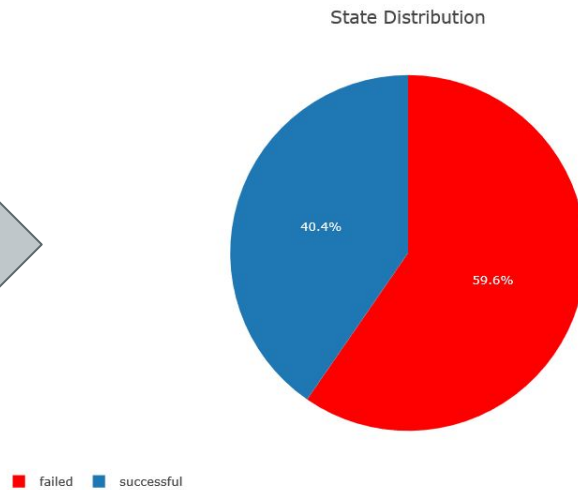
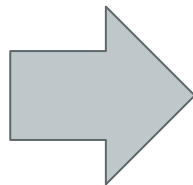
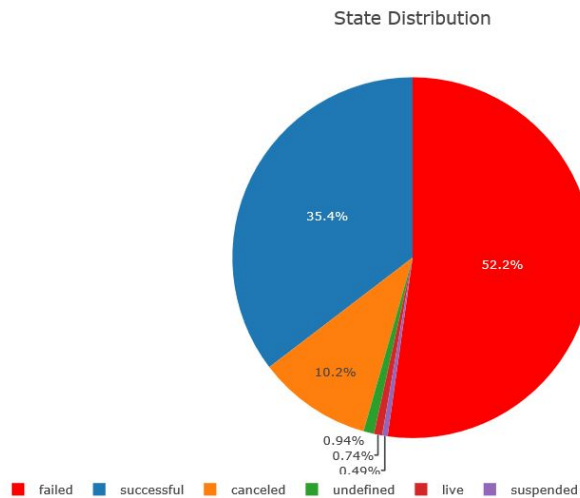
dtype: int64

Project Length Variable

```
df[['deadline', 'launched', 'project_length']].head(5)
```

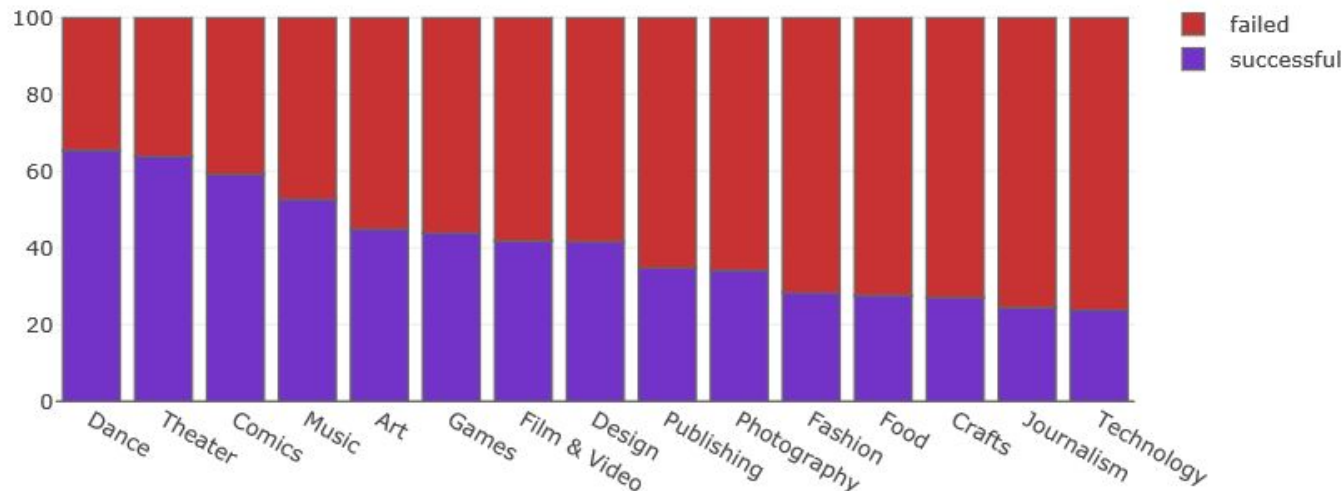
	deadline	launched	project_length
0	12/4/2009	11/25/2009	10
1	12/13/2011	11/7/2011	37
2	3/16/2012	1/25/2012	52
3	11/12/2016	11/11/2016	2
4	7/19/2011	7/12/2011	8

Distribution of State



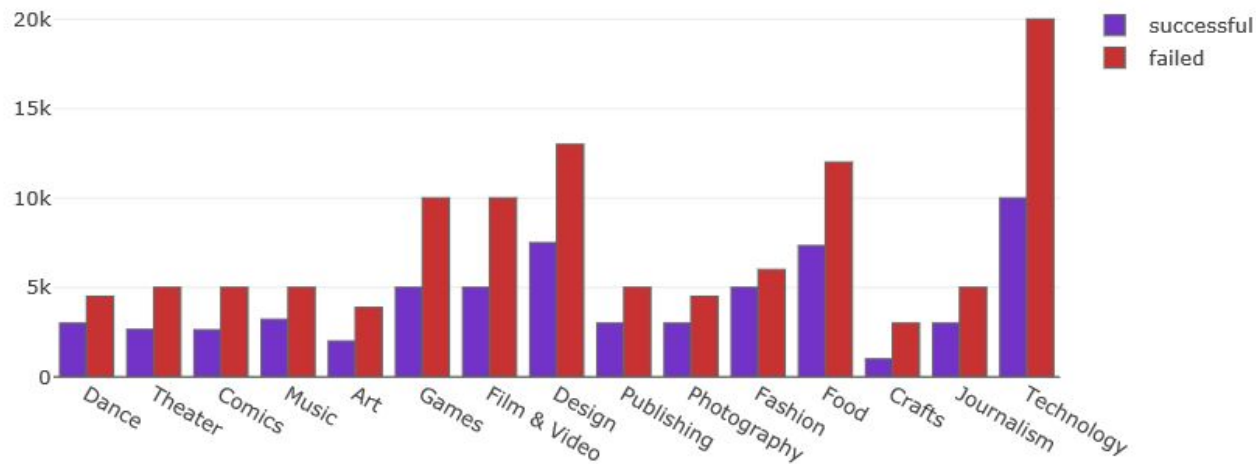
State by Main Category

% of successful and failed projects by main category



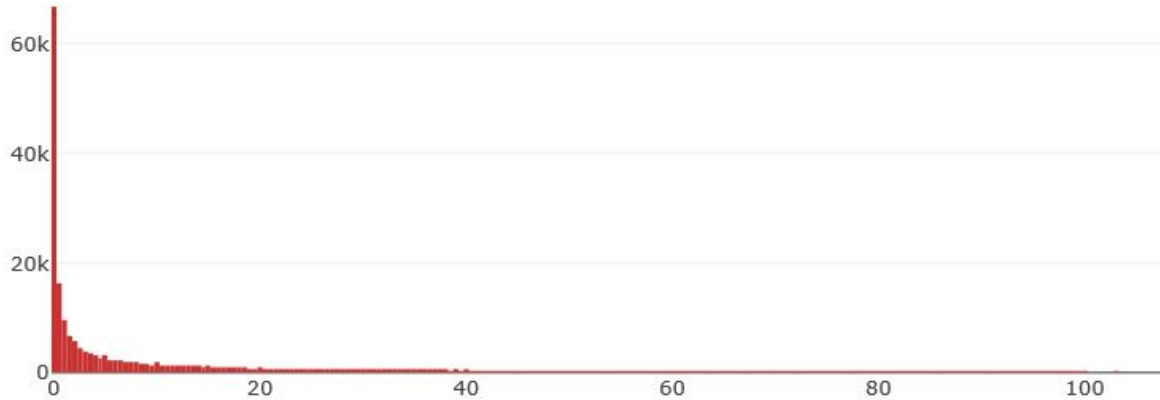
Goal of projects by Main Category

Median goal of successful and failed projects by main category (in USD)



Pledged vs. Goal for Failed Projects

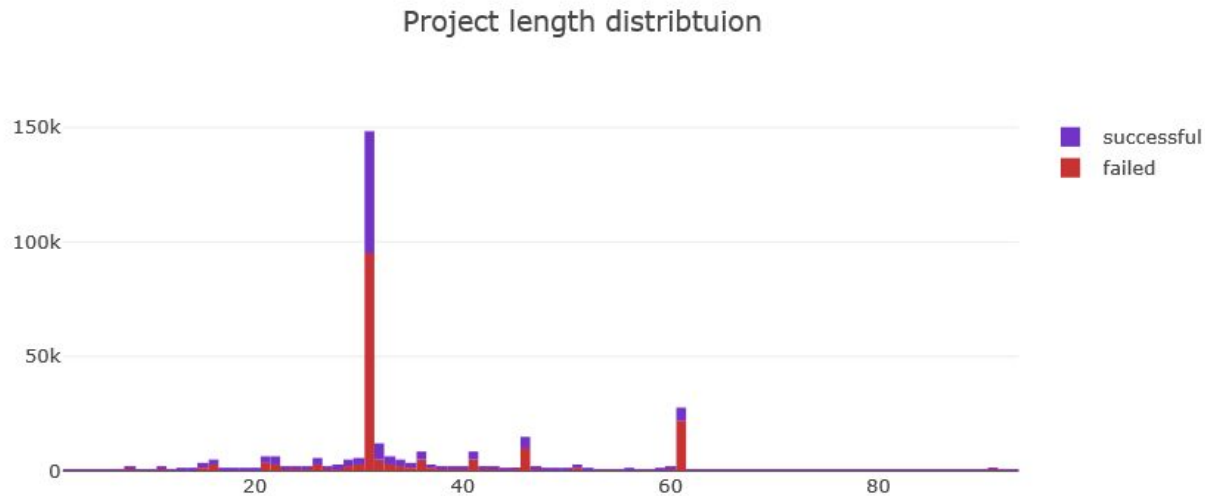
% pledged of the goal amount for failed projects



Project Length Distribution

Mean days for failed projects: 35.17

Mean days for successful projects: 32.16



Variables in Model

```
df_features.head()
```

	main_category	state	country	usd_goal_real	project_length
0	Art	successful	US	0.01	10
2	Film & Video	failed	US	0.15	52
3	Art	successful	MX	0.49	2
4	Film & Video	failed	US	0.50	8
5	Publishing	successful	MX	0.55	33

Dummy Variables for categories

```
# Categorical columns to numerical using dummy variables  
df_features = pd.get_dummies(df_features)
```

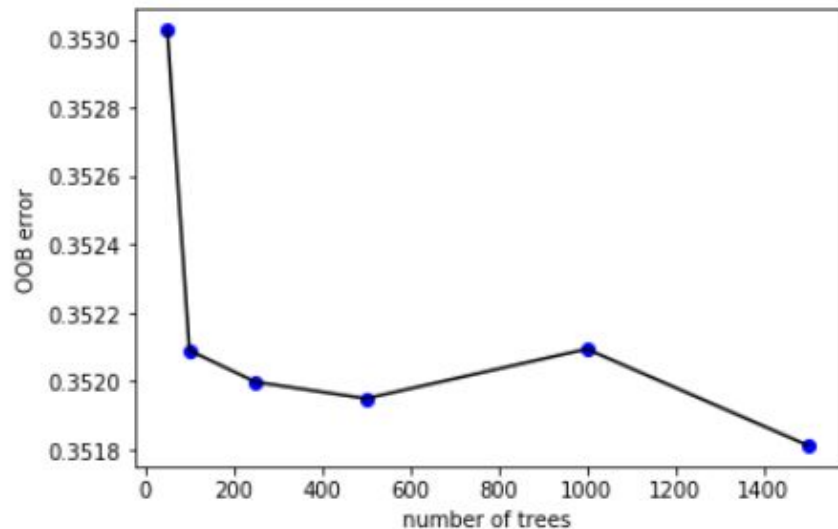
```
df_features.head(5)
```

	state	backers	usd_pledged_real	usd_goal_real	project_length	main_category_Art	main_category_Comics	main_category_Crafts	main_category_Design
ID									
620302213	1	6	100.00	0.01	10	1	0	0	0
9572984	0	0	0.00	0.15	52	0	0	0	0
1379346088	1	7	16.41	0.49	2	1	0	0	0
219760504	0	0	0.00	0.50	8	0	0	0	0
69101025	1	2	522.81	0.55	33	0	0	0	0

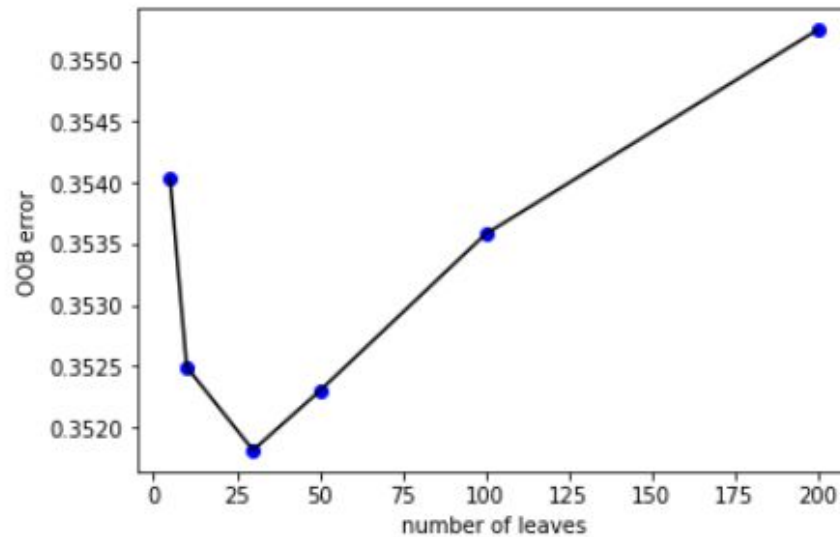
5 rows × 57 columns

Random Forest

Error vs Number of trees



Error vs Number of leaves



Model Accuracy

```
conf_df_pct = conf_df/conf_df.sum(axis=0)
round(conf_df_pct*100, 1)
# Fairly Successful results
```

	failed	successful
failed	73.9	44.3
successful	26.1	55.7

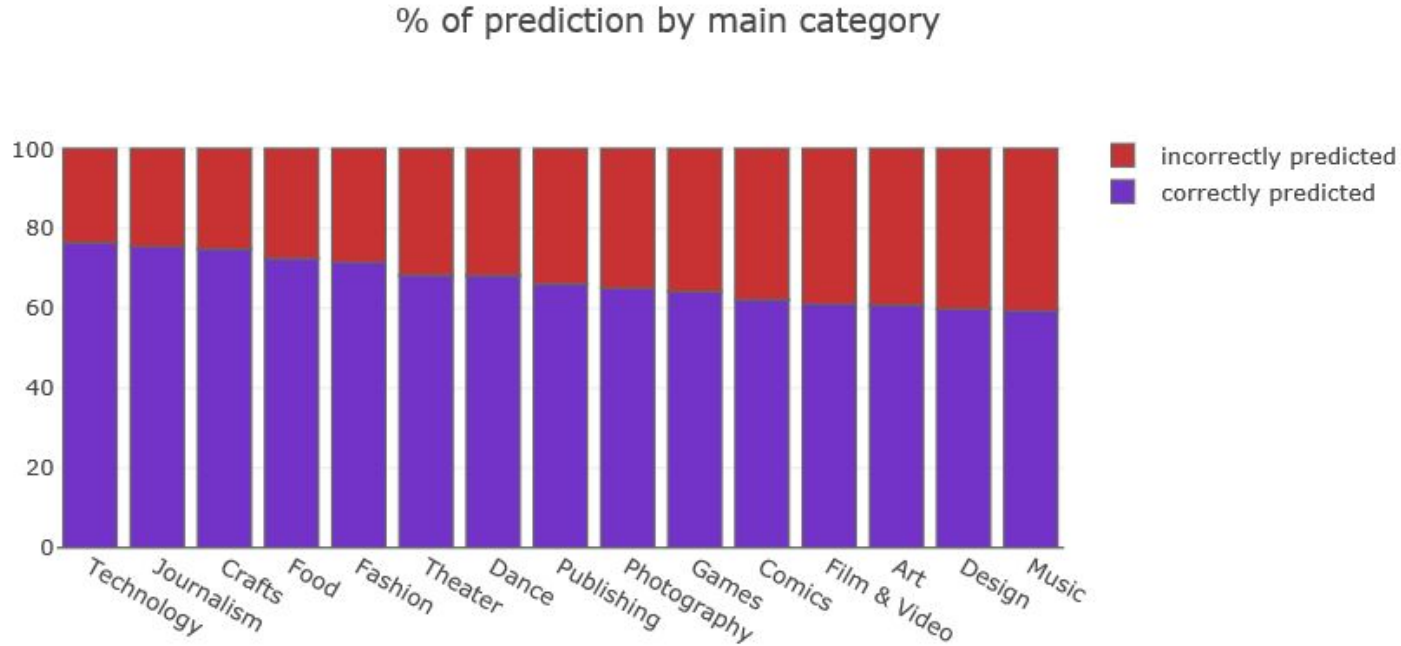
```
# Class-Level performance: 0.646
f1_score(y_true=y_test,
        y_pred=y_pred_test,
        average='macro')
```

0.6464496575100216

```
# Overall performance across all classes: 0.651
f1_score(y_true=y_test,
        y_pred=y_pred_test,
        average='micro')
```

0.6509373604220234

% Prediction by Main Category



Key Findings

- Some categories are more likely to be successful
- Projects with higher goal amount are more likely to fail
- Most failing projects are pledged less than 20% of the goal amount
- Goal amount is the biggest contribution in the model

Challenges

- Not enough features from data
 - Most features are not in project owner's control

Recommendations & Future Research

- Get other features that are more in project owner's control
- Try other classifier models
- Try on other fundraising platforms - Indiegogo, GoFundMe

Questions?

