Pundir, Saurabh

saurabhpundir.data@gmail.com

Abstract

This document contains the initial project proposal for “Success Predication of Kickstarter Project”

SUCCESS PREDICTION   
OF   
Kickstarter Project

Capstone Project Summary

Table of Contents

[Introduction 1](#_Toc487365927)

[Crowdfunding Definition: 1](#_Toc487365928)

[About: 1](#_Toc487365929)

[Scope: 1](#_Toc487365930)

[Clients and Audiences 1](#_Toc487365931)

[What does success mean 2](#_Toc487365932)

[Question Client(s) really care about 2](#_Toc487365933)

[Approach 2](#_Toc487365934)

[I. Step I- Getting Data: 2](#_Toc487365935)

[II. Step II -What will we do with this data: 3](#_Toc487365936)

[III. Data Dictionary: 3](#_Toc487365937)

[Project Deliverables 4](#_Toc487365938)

# **Introduction**

## **Crowdfunding Definition:**

Crowdfunding is the practice of funding a Project or venture by raising monetary contributions from a large number of people in smaller amounts. Crowdfunding is a form of crowdsourcing and of alternative finance.

The company or a person comes up with a Project proposal or idea and shares its details and defines the amount they are looking to raise.

## **About:**

This Project is about analyzing data available for crowdfunding platform and find out best algorithm to predict the probability of success of newly launched crowdfunding project. This predication needs to be done in specific time frame within few hours or days of project inception.

## **Scope:**

There can be various crowdfunding sources, however this Project will limit itself to work on projects funded using a website called **Kickstarte**r.

# **Clients and Audiences**

There are 3 types of entities involved in any Kickstarter project.

1. **Project Owner**: Individual or company who creates a project to acquire funding from crowdsourcing
2. **Project backer**: Individual showing interest in funding the project.
3. **Kickstarter (company hosting website)**: Company owning the platform.

All the entities mentioned above are interested in one common thing **“Success”** of the project.

# **What does success mean**

In Kickstarter ecosystem, success simply means that, did project got the pledged amount or more in mentioned time frame?

It has a significant impact on clients mentioned above. If the project is a success.

1. **Project owner:** They will get money for the project.
2. **Project backer:** They will get return on their investment.
3. **Kickstarter (company hosting website):** Company will get commission or fees. Also, based on prediction company can provide support by highlighting project.

# **Question Client(s) really care about**

In 2015, it was estimated that worldwide over US$34 billion was raised by crowdfunding.

All the above entities have a pressing question...

**“What is the probability of this project being successful?”**

# **Approach**

We would acquire related data of projects from the Kickstarter website. This data is currently scrapped by a company called webrobots.io.

## Step I- Getting Data:

A large dataset containing these projects is available as a month wise zip file. Each monthly zip contains CSV file of size 19 MB each. The files contain details of project at the time of scrapping.

The target is to get data for **10,000** completed projects, of which some were successful and some had failed. The idea is to extract data spread across variables like **category, countries, currency, state, staff pick & amount and few others (detail in section Data dictionary)**. The data will be extracted as one final file. The details for extracting data in section data wrangling *(This section will be added after data wrangling)*

## Step II -What will we do with this data:

Initially, data can be used for understanding and analyze following:

* + 1. The various fields extracted above will be represented using different visualization.
    2. The data will be split for training and testing.
    3. Analyze data using various machine learning algorithm. Few example of this is below. This is not complete list, this is just idea of some scenarios.
       1. Analyze the data to predict success rate of the recently launched project.
       2. Analyze the effect of staff pick on success or failure of the project.
       3. Analyze the effect of length of project on success and failure
       4. Analyze the most funded categories in a number of backers.
       5. Analyze the highest amount funded category per backers.

Later these findings will help in achieving the main objective. *( Will update this section as we progress in the project)*

## Data Dictionary:

1. Kickstarter.com:

Data: <https://webrobots.io/kickstarter-datasets/>

The dataset is available for March 2016 for every month. This data is collected from web crawler created by the company. The latest dataset contains following fields.

1. Id: the unique id
2. Photo: Info for all photo associated with the project.
3. Name: name of the project
4. Blurb: intro & detail
5. Goal: Amount needed to be raised
6. Pledged: Actual amount raised
7. State: Current state of project (canceled, failed, successful)
8. Slug: brief description
9. Disable\_communication: communication allowed to creator
10. country: country of campaign origin
11. currency: the currency of campaign origin
12. currency\_symbol: currency symbol of campaign origin
13. currency\_trailing\_code: TRUE, if conversion needs to happen in user currency
14. deadline: UNIX time stamp for a project deadline
15. state\_changed\_at: Unix timestamp
16. created\_at: UNIX time stamp for project created
17. launched\_at: UNIX time stamp for project started
18. staff\_pick: TRUE, if staff picked
19. backers\_count: Total user backed the project
20. static\_usd\_rate: USD conversion rate from original currency
21. Usd\_pledged: pledge amount in USD after conversion:
22. creator: details like user name for the creator
23. location: location of the project
24. category: category of the project
25. profile:
26. spotlight: feature spotlight available or not
27. URLs: Url info for a project
28. source\_url: seems like URL for category

More data available on scrapping:

1. First 10 Backers Come from Top Cities
2. First 10 Backers Come from Top countries
3. Number of new backers
4. Number of returning backer

# **Project Deliverables**

The project deliverable will contain following:

1. This document explaining approach of the project.
2. Presentation created on this project.
3. Python code for data extraction with comments and details.
4. Python code for web scrapping with comments and details.
5. Python code for data wrangling with comments and details.
6. Graphs and visualizations created for this project.
7. Identifying and implementing machine learning algorithm.