KICKSTARTER DATA PLEDGE PERCENTAGE PREDICTION

Figure 1 : Picture is from https://tinuiti.com

Introduction

A creator-a person or a team- ask other people investments for their ideas, projects, and merchandising through platforms such as Kickstarter. Kickstarter is a global crowdfunding platform. Kickstarter project sets its own funding goal and deadline, offering various types of rewards in exchange for pledges. Projects only get funded if they reach their original goal. Projects can categorize 13 categories and 36 subcategories, such as Design, Games, and Technology.

Kickstarter takes 5 percent of all funds collected from the project, with an additional 3-5 percent processing fees when pledges are charged by a third-party payment partner. Therefore, creators got almost 8 percent less money from their funding goal.

Data Set

Kickstarter Success prediction analysis was a part of a Monterey Bay Data Science meetup workshop. In this workshop, attendees worked on the Kickstarter data set to build the best model for the prediction of pledge percentages.

There were 3 data set in this workshop. The first data set was for data exploration, which has 294924 rows with 11 features. The second and third data set were for

building a model. For these data sets pledge level is a value from 1 to 4, where 1 means funding of no more than 10% of the project goal, 2 means funding of more than 10% but no more than 100% of the project goal, 3 means funding of more than 100% but no more than 1000% of the project goal, and 4 means funding of more than 1000% of the project goal. We built and trained the model with the second data set and tested it with the third one. The second data set has 294927 rows with 11 features, and the third data set has 73732 rows with ten features(without pledge level column).

The variables of the study are:

- **Name:** The name of the project.
- **Category:** This is a subcategory for the project.
- **Main Category:** Category of the project.
- **Currency**: Any kind of money based on country.
- **Deadline:** The deadline date of the project
- **Launched:** The starting date for collecting the fund
- **Country**: The country of the project creators
- **Usd_goal_real:** Funding goal
- **Percent_pledged:** Collected amount of target fund

Data Wrangling

1. **Unnecessary information columns:** At this stage, I filtered out unnecessary columns and dropped them.

$$df=df(columns=['column\_name'])$$

2. **Missing data:** There were only 4 missing data, so I dropped them.

$$df.dropna()$$

3. **Incompatible types of columns:** Incompatible types of columns: The variables of 'launched' and 'deadline' columns were not datetime object. Therefore, I convert them to datetime object.

$$data['deadline']=data['deadline'].apply(pd.to\_datetime)$$

4. **A New Column:** I created 'How_many_day' and 'canceled" columns. For the 'How_many_day' column I subtract two datetime objects.

data['How_many_day']=data['deadline"]-data['launched']

For 'canceled' column , I extracted the rows which contains 'canceled' word.

Key Findings

During exploratory data analysis, we ask the following questions:

What kind of factors affect the pledge levels?

Initial findings are:

· 53652 projects have 0 percent pledge, and 188848 projects are unsuccessful. 106098 projects are successful, and 102259 projects got more than their target funding.

· 36% of the projects got their funding..
· The lower amount of funding goals are more successful than high targets.
· The most used words in categories are design, video, film, music and games.