KickStarter Projects

Junxi Huang juh80@pitt.edu

Echhit Joshi ebj7@pitt.edu

Table Of Contents

Description Of The System/Problem

The System we are analyzing is that of Kickstarter. Kickstarter is a platform where people seek funding for their creative projects for various categories such as films, games, music, art, design, and technology(Kickstarter Basics). The projects set out financial goals that they want to achieve during the lifespan of their Kickstarter Campaign. The projects could be classified into various states, namely: Failed, Successful, Canceled, Live, and Undefined.The projects have various data that could be analyzed to possibly predict them into their states.

The Problem we have identified is that the projects are not given success rates(States) given historic data. This in turn makes it hard for the project to adjust to the challenges that they may face during the life of their kickstarter campaign. The data that username: Kemical, has accumulated could be analyzed to make various inferential decisions and predict the state of a new project given its data. For example, If a startup company wants to mass produce ergonomic keyboards, it is not given a calculated probability that the project might fail due to various incompetencies. If the project launched by the company is in a failing state, it may change its marketing approach along with other various strategic decisions to help achieve their set goal.

How Machine Learning Will Solve The Problem/Overall Approach

We will use different machine learning method in this program. Because this is a classification problem, so KNN, logistic regression, naive bayesian, decision tree, SVM and Ensemble model can analyse this dataset. We will train and test the dataset to accurately predict if a project will fit a certain state i.e. Failed, Successful, Canceled, Live, and Undefined. Then, we will compare which model will have the best accuracy by comparing the R2 and other values.

Type And Source Of Data

Data name: Kickstarter projects

Data description: More than 300,000 kickstarter projects

Data source: https://www.kaggle.com/kemical/kickstarter-projects

Data: https://www.kaggle.com/kemical/kickstarter-projects/downloads/kickstarter-projects.zip/7

Data variables and types:

- ID(Numeric)
- name(String)
- category(String)
- main_category(String)
- currency(String)
- deadline(DateTime)
- goal(Numeric)
- launched(DateTime)
- pledged(Numeric)
- state(String)
- backers(Numeric)
- country(String)
- usd pledged(Numeric)

Main Responsibilities Of Each Team Member

The work will be divided into equal halves. Both the team members will work together on finding solutions to the various goals we set to solve the problem identified earlier. Both team members will work on applying logic as to which model will be best suited for this problem. Both team members will also check for various ways to optimize the model selected and avoid various problems such as overfitting and underfitting. No work will be done by a single person and if so, the other member will check and reaffirm the work done.

References:

Kickstarter Basics » Frequently Asked Questions (FAQ) - Kickstarter, www.kickstarter.com/help/faq/kickstarter%20basics.

https://www.kaggle.com/kemical/kickstarter-projects/downloads/kickstarter-projects.zip/7