

ML Assignment 2 - Proposal

Student: Alex O Regan R00136281

Class: SDH4-B

Selected Dataset

The dataset I chose is a Kickstarter Projects dataset (<https://www.kaggle.com/kemical/kickstarter-projects>).

This contains two CSV files - one file representing kickstarter projects from 2016, and one representing projects from 2018, each of which contain over 300,000 project entries.

For each CSV in the dataset, there are 13 columns of attributes, which have the following datatypes:

- ID - integer - the kickstarter ID of the project
- Name - string - the name of the project
- Category - string - a small string representing the domain of the project
- Main Category - a broader categorization of the project
- Currency - string - the currency of money the project is asking for
- Deadline - Date - the final deadline of the project
- Goal - integer - the amount of money (in the specified currency) that the project aims to make
- Launched - Date - the date that the project was launched on
- Pledged - integer - the amount of money pledged to the project at the time of collecting the data
- State - the state of the project. This can be 'successful', 'failed', or 'cancelled'
- Backers - the number of backers currently pledging to the project
- Country - the country in-which the project originates
- Usd Pledged - the amount of money pledged in USD

Project Proposal

For this project I hope to investigate the effectiveness of a classification model that can determine the likelihood of a Kickstarter Project being successful, and evaluate the effectiveness of various approaches to this, such as using K-Nearest-Neighbour to compare project attributes, and using Bayesian analysis on the title / category of a project to bias the results.

Ideally, I would hope that by exploring this topic, I will end up with a classification model that is at least slightly better than random guesses when analysing Kickstarter Projects, and I hope to have identified which attributes in the dataset, if any, contribute to the possible success of the project.

Possible Software Applications

This application will be written in Python3 and I will be using SciKit Learn as much as I possibly can for fitting algorithms such as K-Nearest-Neighbour.

Relevant Papers

KickPredict: Predicting Kickstarter Success

<https://pdfs.semanticscholar.org/fcfc/059870dea7f70f3acc86735dc03601d302b6.pdf>

Project Success Prediction in Crowdfunding Environments

<https://dl.acm.org/citation.cfm?id=2835791>