

Project Report
**KickStarterChance: neural network model to predict
projects' success outcomes**

ECEN 649: Pattern Recognition

Course Instructor

Prof. Tie Liu

Team Members

- 1) Khaled Nakhleh
- 2) Drupad Khublani

Due Date: November 30th, 2018

Introduction

The goal of this project was to build a model that predicts the success chances of a Kickstarter project. A machine learning algorithm was applied on a kickstarter projects' dataset. Our objective is to find the best combination of labels that would increase a project's success rate. In other words, what project would be the most likely to get funded and succeed as a product. We define "succeed" here as a project that was successfully funded.

Our dataset was found on Kaggle database. The dataset we are using contained 378,661 projects that were hosted by Kickstarter, and last updated in 2018. The dataset contains 15 labels (ID, name, main_category, etc). Our methodology was to compare which projects succeeded, and develop a model that predicts a project's success based on those factors.

The full code used, along with comments, can be found in [1].

Methodology

The dataset used contains 378,661 projects.

The dataset contains 15 labels (ID, name, main_category, etc). The labels ID, date initiated, and deadline were dropped and not used in training the model. This is because the label ID did not provide any tangible information. The labels date initiated and deadline were stored as strings, which doesn't input well for neural networks.

For the label name, The label was split into three labels: name length, uppercase count, and lowercase count. The three labels were proposed to identify characteristics in the project's title. The label state was translated to integers, where six states were identified and stored in a dictionary.

The other labels where stored as found in the original file.

For building and training the neural network, the Python module Keras was used. Six hidden layers where used, with the output layer producing an array of six elements. All features found in the original data file were used, and the model was trained using activation function "relu", and "softmax" for the output layer.

Full code implementation can be found in [1].

Results

The results section has two subsections: the dataset cleaning process output, and the neural network performance

Data cleaning process

For the data cleaning part, a percentage of %0.0669 was found for the missing values in the dataset. Four missing values were found from the “name” label, and 3797 missing values were found for the “pledged usd” label. A total of 3801 missing values were replaced with zero. This prevented the model from basing false predictions based on undefined values.

Neural network performance

For the neural network portion, the “Keras” package was used for building and training the model. The model consisted of six hidden layer, which were activated using “relu”. The last output layer was activated using “softmax”.

The model was trained for 5 epochs, with a batch size of 512. The model accuracy was less than expected, at %52.2 accuracy rate.

Again, All of the code implementation can be found on GitHub [1].

Conclusion

Based on the previous results, it was found that a Kickstarter project is difficult to estimate its success. The nature of starting projects, and the volatile nature of Kickstarter projects, along with business trends, suggest that numerically predicting a project would be cumbersome.

In order to improve the model's accuracy rate, we recommend the following improvements:

- 1) Incorporate validation dataset in training the model, which can be done by splitting the training dataset into two parts. One part is for training, and the other for validation.
- 2) Include other neural layers into the model.
- 3) Quantify the remaining labels which were left out, such as the initial date and deadline. This can give a timeframe for which projects were funded on Kickstarter.

For more information, and further code comments, please see the GitHub repository for further explanation [1].

References

[1] <https://github.com/khalednakhleh/KickStarterChance>