Machine Learning Challenge #2

Jun 15, 2017, 12:00 AM IST - Jun 30, 2017, 12:00 AM IST

INSTRUCTIONS PROBLEMS SUBMISSIONS LEADERBOARD ANALYTICS JUDGE

← Problems / Funding Successful Projects

Funding Successful Projects

Max. Marks: 1 Tweet

This problem is no longer available for practice. Apology for any inconvenience!

Problem Statement

Kickstarter is a community of more than 10 million people comprising of creative, tech enthusiasts who help in bringing creative project to life. Till now, more than $3 billion dollars have been contributed by the members in fuelling creative projects. The projects can be literally anything – a device, a game, an app, a film etc.

Kickstarter works on all or nothing basis i.e if a project doesn't meet it goal, the project owner gets nothing. For example: if a projects's goal is $500. Even if it gets funded till $499, the project won't be a success.

Recently, kickstarter released its public data repository to allow researchers and enthusiasts like us to help them solve a problem. Will a project get fully funded ?

In this challenge, you have to predict if a

project will get successfully funded or not.

Download Dataset

Data Description

There are three files given to download: train.csv, test.csv and sample_submission.csv The train data consists of sample projects from the May 2009 to May 2015. The test data consists of projects from June 2015 to March 2017.

Variable	Description
project_id	unique id of project
name	name of the project
desc	description of project
goal	the goal (amount) required for the project
keywords	keywords which describe project
disable communication	whether the project authors has disabled communication option with people donating to the project
country	country of project author
currency	currency in which goal (amount) is

Variable	Description
	required
deadline	till this date the goal must be achieved (in unix timeformat)
state_changed_at	at this time the project status changed. Status could be successful, failed, suspended, cancelled etc. (in unix timeformat)
created_at	at this time the project was posted on the website(in unix timeformat)
launched_at	at this time the project went live on the website(in unix timeformat)
backers_count	no. of people who backed the project
final_status	whether the project got successfully funded (target variable – 1,0)

Submission

A participant has to submit a csv file with

project_id and predicted labels. Check the sample submission file for reference.

```
project_id,final_status
kkst917493670,0
kkst1664901914,1
kkst925125077,1
kkst1427645275,0
```

Evaluation Metric

Submission will be evaluated based on Accuracy metric. For more information on this metric, read here.

Scripts

- R GBM Starter Script - Click Here
- Naive Bayes Python (0.652) - Click Here
- XGBoost with Text Features (0.70) - Click Here
- XGBoost in Python with Count features (0.708) - Click Here

Update: (23rd June 2017)

The data set given in this challenge is a real world data set. We expected you to follow the competition guidelines and refrain from using any unfair means to get high score. Such participants will be disqualified. Winners score must be locally reproducible. Please keep the competitive spirit alive.

Update: (30th June 2017)

The leaderboard is updated with final rankings. Now, we'll reach out to winners for a final validation of their scores. We'll

soon release a short editorial on winners and their solutions. We'll keep the participants posted via email. Thanks for participating. Feel free to drop your suggestions, concerns or anything you want to tell us.

3

LIVE EVENT