

BLUE PRINT

Crowdfunding: Predicting Kickstarter Project Success



B.Arunkumar
02-Aug-2021

CONTENTS

1.Introduction	3
2.Data Science Project Life Cycle	4
3.Standardized project structure.....	5
4.Tools and utilities for project execution	6
5.Outcome	6



1.Introduction

AIM - Predicting Kickstarter Project Success

Kickstarter, founded in 2009, is an online platform that acts as a centralized marketplace to connect “creators” with the capital to pursue their visions. Creators post their project ideas on the platform and collect the necessary funding from their “backers” Often, these donations take on the form of an investment, with the creator promising premium content for the backers. As of Nov. 2020, over 500,000 projects have been launched on Kickstarter, with over five billion dollars raised.

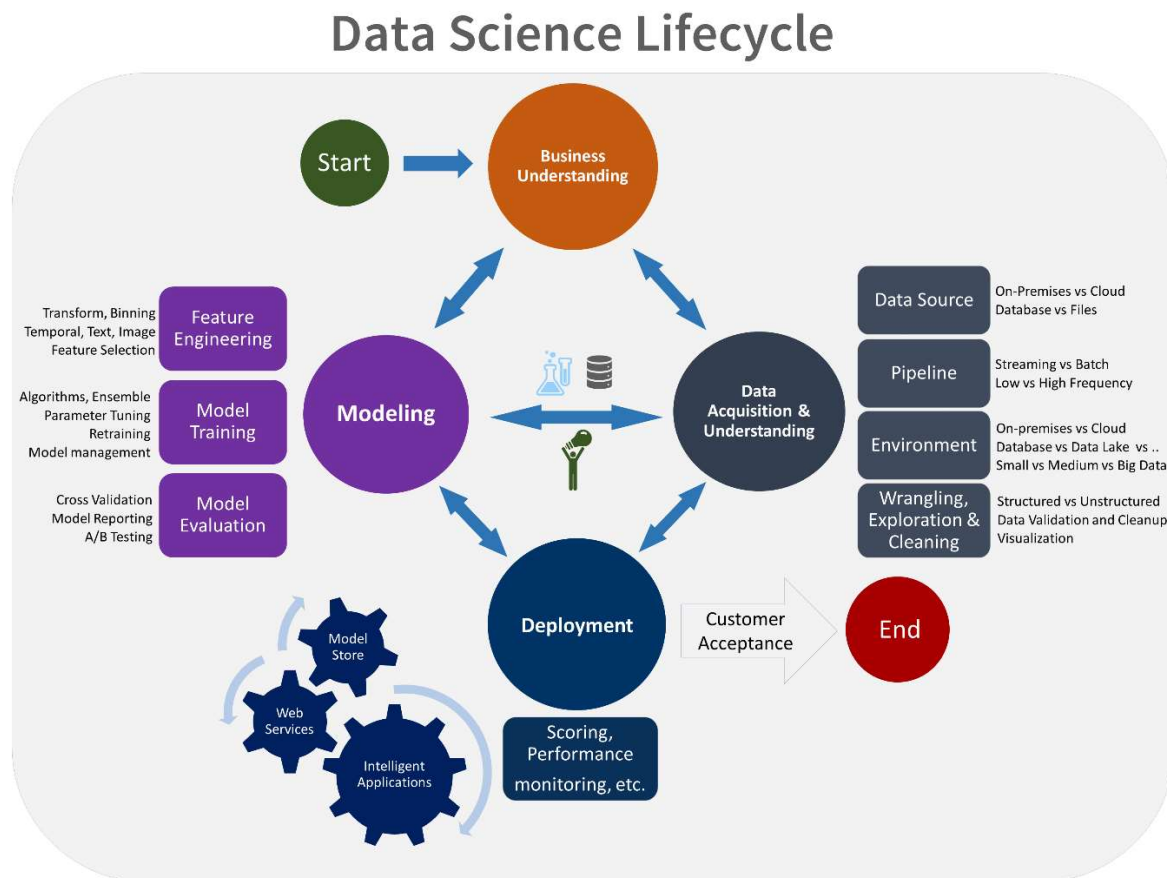
Projects on Kickstarter combine a variety of data types. Each project has a funding goal and a duration. Projects generally contain text descriptions and backstories as well as visual information, such as photos and videos. Moreover, projects are connected to user profiles (i.e., creators) that contain short biographies, previous Kickstarter activity, and external links to various social media platforms. Kickstarter also displays the amount of money raised to date and information about the backers, which includes their location, whether they are new or returning Kickstarter users.

Natural language processing (NLP) is used for tasks such as sentiment analysis, topic detection, language detection, key phrase extraction, and document categorization.

NLP can be used to classify documents, such as labeling documents as sensitive or spam. The output of NLP can be used for subsequent processing or search. Another use for NLP is to summarize text by identifying the entities present in the document. These entities can also be used to tag documents with keywords, which enables search and retrieval based on content. Entities might be combined into topics, with summaries that describe the important topics present in each document. The detected topics may be used to categorize the documents for navigation, or to enumerate related documents given a selected topic. Another use for NLP is to score text for sentiment, to assess the positive or negative tone of a document. These approaches use many techniques from natural language processing, such as:

- **Tokenizer.** Splitting the text into words or phrases.
- **Stemming and lemmatization.** Normalizing words so that different forms map to the canonical word with the same meaning. For example, “running” and “ran” map to “run.”
- **Entity extraction.** Identifying subjects in the text.
- **Part of speech detection.** Identifying text as a verb, noun, participle, verb phrase, and so on.
- **Sentence boundary detection.** Detecting complete sentences within paragraphs of text.

2.Data Science Project Life Cycle



<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>

Team Data Science Process (TDSP)

The “**Team Data Science Process (TDSP)**” is Microsoft’s agile, iterative data science methodology to deliver predictive analytics solutions and intelligent applications efficiently. **TDSP** lifecycle is composed of five major stages that are executed iteratively and these stages include:

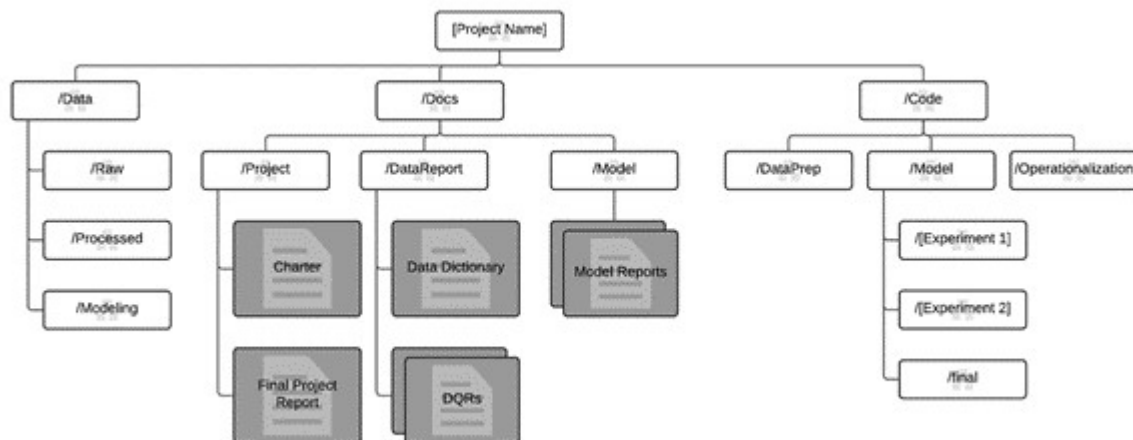
- Business Understanding
- Data Acquisition and Understanding
- Modeling
- Deployment
- Customer Acceptance

3. Standardized project structure

Having all projects share a directory structure and use templates for project documents makes it easy for the team members to find information about their projects. All code and documents are stored in a version control system (VCS) like Git, TFS, or Subversion to enable team collaboration. Tracking tasks and features in an agile project tracking system like Jira, Rally, and Azure DevOps allows closer tracking of the code for individual features. Such tracking also enables teams to obtain better cost estimates. TDSP recommends creating a separate repository for each project on the VCS for versioning, information security, and collaboration. The standardized structure for all projects helps build institutional knowledge across the organization.

We provide templates for the folder structure and required documents in standard locations. This folder structure organizes the files that contain code for data exploration and feature extraction, and that record model iterations. These templates make it easier for team members to understand work done by others and to add new members to teams. It is easy to view and update document templates in markdown format. Use templates to provide checklists with key questions for each project to insure that the problem is well defined and that deliverables meet the quality expected. Examples include:

- a project charter to document the business problem and scope of the project
- data reports to document the structure and statistics of the raw data
- model reports to document the derived features
- model performance metrics such as ROC curves or MSE



4.Tools and utilities for project execution

Introducing processes in most organizations is challenging. Tools provided to implement the data science process and lifecycle help lower the barriers to and increase the consistency of their adoption. TDSP provides an initial set of tools and scripts to jump-start adoption of TDSP within a team. It also helps automate some of the common tasks in the data science lifecycle such as data exploration and baseline modeling. There is a well-defined structure provided for individuals to contribute shared tools and utilities into their team's shared code repository. These resources can then be leveraged by other projects within the team or the organization. Microsoft provides extensive tooling inside Azure Machine Learning supporting both open-source (Python, R, ONNX, and common deep-learning frameworks) and also Microsoft's own tooling (AutoML).

5.Outcome

If we follow the above pipeline definitely our project goes in a Effective way to achieve the Aim of the project.

