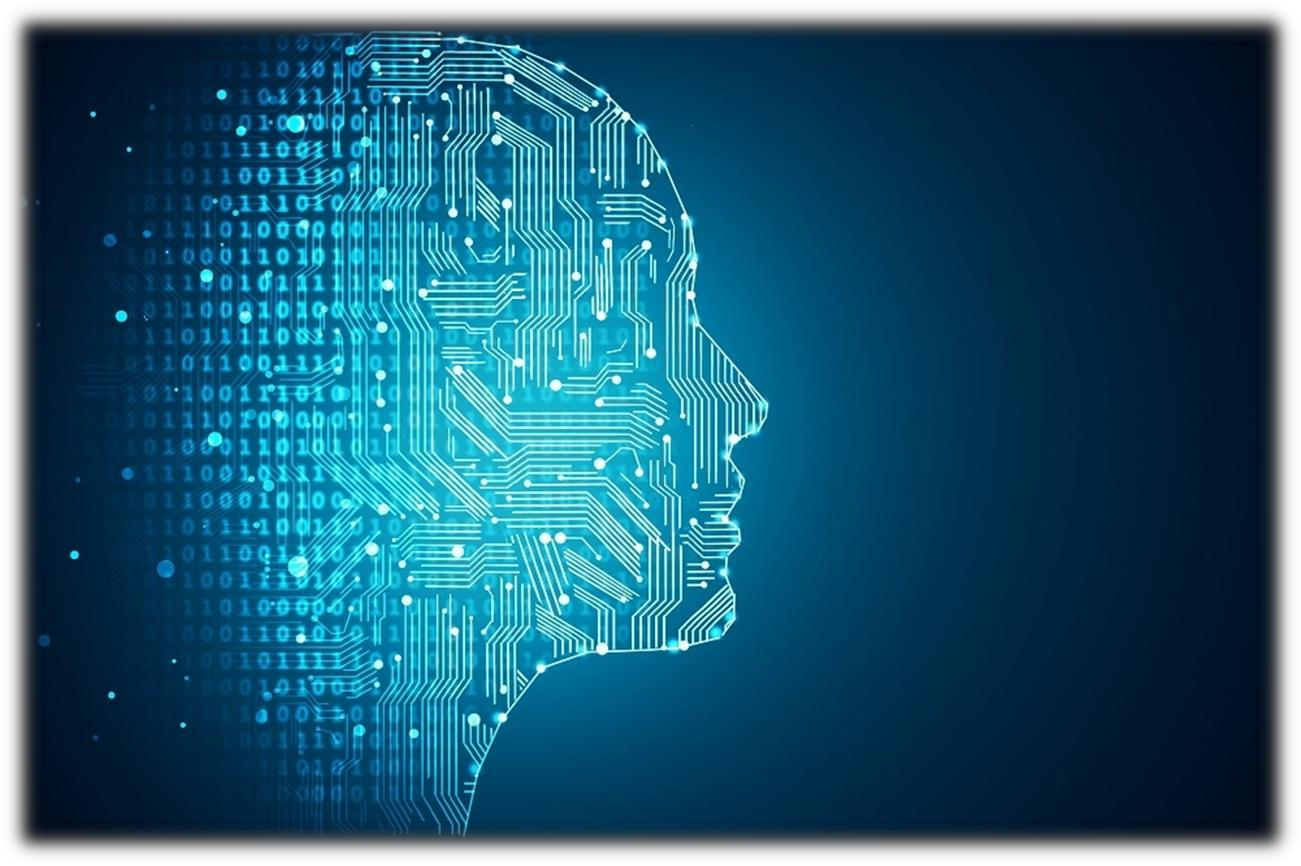
**IT 495**

**CAPSTONE PROJECT**



***PREDICTIVE MODEL FOR KICKSTARTER PROJECTS’***

***Apoorva Paygude Kaustubh Borole Dheeraj Patil***

Table of Contents

1. PURPOSE ............................................................................................................................. 3

2. PROJECT OVERVIEW .......................................................................................................... 3

3. BUSINESS PROBLEM .......................................................................................................... 3

4. NEED..................................................................................................................................... 3

5. REQUIREMENT .................................................................................................................... 4

6. PROPOSED SOLUTION........................................................................................................ 4

7. RESOURCES REQUIRED..................................................................................................... 4

8. RISK ASSOCIATED............................................................................................................... 5

9. RISK MANAGEMENT ............................................................................................................ 5

10. PROJECT SCHEDULE........................................................................................................ 6

11. STAKEHOLDER .................................................................................................................. 7

12. CONTACT............................................................................................................................ 7

**1. PURPOSE**

The purpose of this core process 1 report is to document the summary of our Machine Learning project. This includes the business problem, objective of the project, the project scope, proposed solution, requirements and our work breakdown structure. This report will clearly define the goals of this project based on the requirements.

**2. PROJECT OVERVIEW**

Kickstarter helps artists, musicians, filmmakers, designers, and other creators find the resources and support they need to make their ideas a reality. To date, tens of thousands of creative projects big and small have come to life with the support of the Kickstarter community. Kickstarter is an enormous global community built around creativity and creative projects. Over

10 million people, from every continent on earth, have backed a Kickstarter project. Since its launch, on April 28, 2009, 15 million people have backed a project, $3.9 Billion have been pledged, and 149,790 projects have been successfully funded. With so many projects on the platform, it is impossible for an investor to know which projects will succeed or fail.

**3. BUSINESS PROBLEM**

With so many projects on the Kickstarter community, investors have a hard time pinpointing the projects or ideas that will make big in the future. There needs to be a way where investors can get information on whether the project is likely to succeed or fail.

**4. NEED**

The success or failure of a project depends on various factors. There is a need to find key insights and relations between various data variables. A predictive model which can predict the success or failure of a project based on these key insights will enable the investors to take the right decisions when investing in projects.

**5. REQUIREMENT**

* Investors require concrete information on the likelihood of success or failure of project(s) before making an investment.
* The client has provided around 1784 files of data.
* These files of data need to be organized.
* The data has many irregularities which needs to be standardized and cleaned.
* A proper training dataset is to be created.
* From the training dataset, predict the success or failure of projects using at least 2 machine learning algorithms.
* Create data visualizations which will be useful to draw key insights between the various data variables.
* Evaluate the predictions using suitable metrics.

**6. PROPOSED SOLUTION**

This project aims to aid in predicting project(s) success/failure from the Kickstarter campaign. We will perform data cleaning and data visualization to showcase key insights between various data variables. After trying out different machine learning algorithms, we will then make a prediction using the best/top 2 algorithms which are most efficient.

**7. RESOURCES REQUIRED**

Dataset: Kickstarter dataset provided by the client (1784 files).

Data Analysis: Microsoft Excel, Python (numpy, pandas, matplotlib, seaborn, scikit-learn). Data Visualization: Tableau.

Hardware: High end laptop capable of handling big data.

Additional Software’s: Google Drive, Google Colab, Jupyter Notebook.

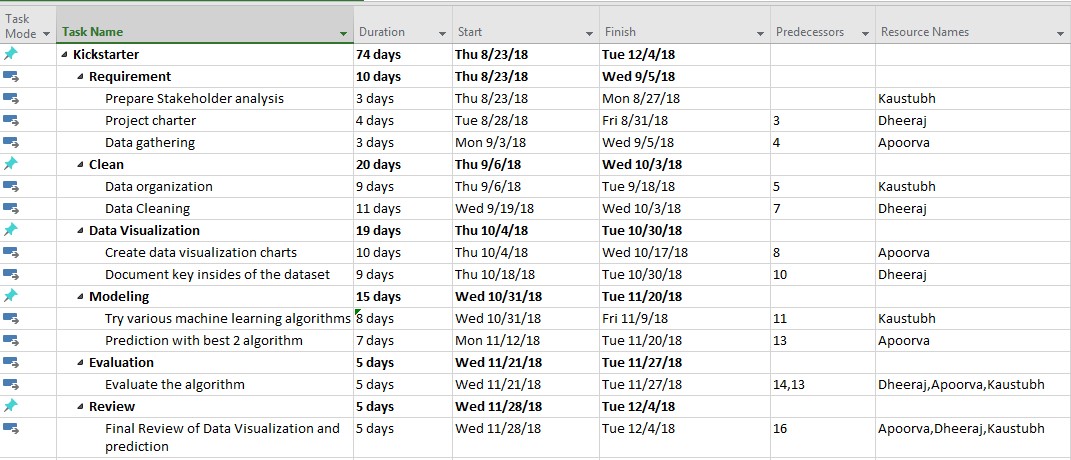
**8. RISK ASSOCIATED**

The project requires combining 1784 files together. There is high possibility that the project cleaning will lose some of the data. Also there is a chance that data can get corrupted when combining together. Delay in delivering project can be one of the potential risks. Technical issue in python IDE and tableau may be faced. We may require special access to GPUs to handle the large data set.

**9. RISK MANAGEMENT**

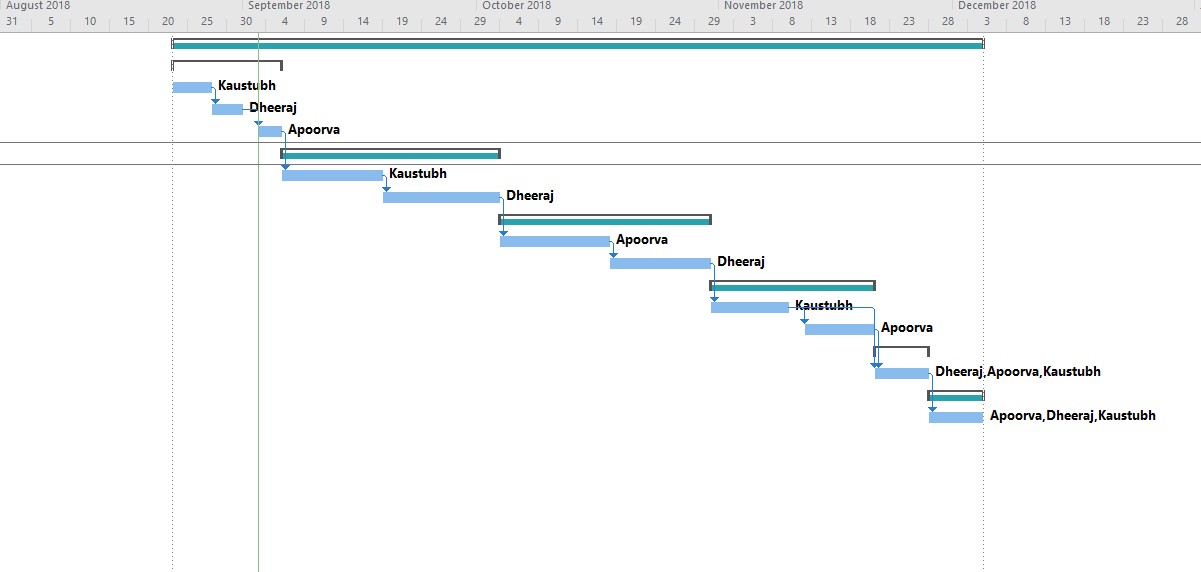
There is huge set of data and we would try to combine the related data together first and then clean it accordingly. Recovery and backup of data will be one of integral part of the project. Follow the project charter and try to execute accordingly. To prevent data theft, use high end software. The laptop configurations may not be enough to train such big dataset, in that case we may use Goggle Colab which gives access to Google GPUs online. If client requirements are changed, we will execute alternate plan.

**10. PROJECT SCHEDULE**



6

Gantt chart



**11. STAKEHOLDER**

DR. JAMES WOLF Professor,

School of Information Technology, Illinois State University.

**12. CONTACT**

Kaustubh Borole : 309-205-4546 / kpborol@ilstu.edu

Apoorva Paygude : 309-750-8106 / aapaygu@ilstu.edu

Dheeraj Patil : 971-770-6065 / djpatil@ilstu.edu

7