Project Description

* **Team Members**

Jiaao Yu

Miner Yang

Runjie Li

* **Description**

In order to usefully apply knowledge of Scala and Big data to a problem with practical significance, basically we plan to develop a client-based new-release game popularity prediction system**.** We will Ingest data and extract features from Steam Store Games dataset probably combined with Steam reviews dataset or some other related source data, then train model by using spark MLlib and give predicted results back to client through web page, besides, our model will be refreshed as well as the data being updated.

* **Goals**

In our prediction system, we expect to predict the popularity of a new-release game for its developer and publisher according to its multiple tags, such as games genres. Popularity would be based on positive ratings, negative ratings and other features. This prediction result would benefit the business operation support system of a game company to evaluate their market better.

* **Dataset Inspection**
* Source: <https://www.kaggle.com/nikdavis/steam-store-games>
* 6 separated .csv files with roughly 30000 rows
* steamspy\_tag\_data.csv contains 371 columns related to game tags we ‘d like to use