Blueprint
Predicting Kickstarter project success

Bupesh Kumar.R   Date: 02/09/2021

Data Collection → Data Processing → EDA → Training Model

Model Deployment ← Model Prediction ← Testing & Evaluation
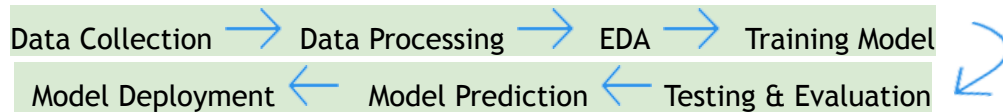
Data Collection

- Data Mining - Collect the raw data from https://webrobots.io/kickstarter- datasets/
- Merge the data sets of 3-5 years time period.

Data Processing

- Clean the data by removing null values, duplicate, irrelevant data.
- Select the target variable and figure out the dependant factors.
- Standardize the data for further processing.
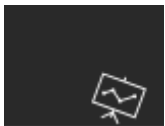
Exploratory Data Analysis

- Check for skewness and try to rectify it if found any
- Visualize various plots to find patterns with respect to the target variable.
- Perform feature selection to remove unnecessary metadata.

Training Model

- We us70:30 train test ratio due to the large dataset.
- CoreNLP is used for Text Encoding in 'blurb'.
- The task is defined as a Binary classification method to predict Sucess(1) or Fail(0).
- Hyperparameter tuning - SVM, Neural Net, Gradient Boosting, Random Forest, Logistic Regression, etc.

Test Model & Evaluation

- 66,000 datasets i.e, 30% of 220,000 is used for Test data.
- Evaluation parameters consist of accuracy, precision, recall, F measure
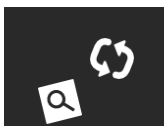
Model Prediction

- Predict the state after applying the right model.
- Cross-check the model prediction with existing train datasets. Rectify the model performance technique if necessary.

Model Deployment

- Deploy the model and check functionality.