Analyzing and Predicting the Success or Failure of a Project on Kickstarter

Business Objectives

The first objective of this project is to analyze different startup projects on Kickstarter website to know what features contribute towards the project success. The second objective is to predict whither a project will success based on those features.

Date Understanding:

The data related to Kickstarter projects were obtained from Kaggle website. The data contains the following information:

project ID, Sub and main category of the project, the amount of fund requested, the amount of fund collected, the project lunch date, the project deadline, the number of backers for each project, and the project state.

The initial size of the data set was 378661 row, and 14 columns.

Data Processing:

Through data processing stage, the data went through following steps:

1- Removing 4 duplicated columns.
2- Removing 4 rows with missing values
3- Changing the data types for the lunch date and the deadline date columns from objective to date type.

Since the data processing is a continuous process, part of it was executed during the data analysis stage.

Data Analysis:

The purpose of this stage is look at the data themselves, and the relationship between those data, especially between the project features and the projects state (successful or unsuccessful). When started looking at number of project state to investigate it, it was found that there are some states that don't make any sense such as "undefined" and "live" states. After further investigation about those values a total of 3782 rows were removed in addition to the ones removed earlier Additionally, the projects with live state were changed to "failed". Once this further processing steps were completed, the following was found:

1- Successful projects take less days to be completed (32 days in average).
2- The mean value of requested fund is around 10,000$ for successful projects, while unsuccessful ones usually were aiming for higher amount found (above 100,000$).
3- Successful projects have higher number of backers.
4- Most of the successful projects are related to music, film, art, theatre, comics, and dance.

State	backers	usd_pledged_real	usd_goal_real	time_diff_in_days
canceled	26.00	2,341.03	75,613.45	37.68
failed	16.43	1,321.12	63,189.89	34.59
live	68.34	5,887.35	62,808.90	39.23
successful	264.13	22,664.49	9,535.70	31.57
suspended	105.53	9,342.54	175,953.51	43.95

Table 1: Project Features Aggregated by Meas

Data Modeling:

The following features were used to predict the project state: the amount of requested fund, the number of days required to complete the project, the main category of the project, and the number of backers. Then, the data was splatted to 80% for training and 20% held for testing.

After that, two classification models were deployed for predicting whither a project will be successful or not: decision tree and GaussianNB. Decision tree models has two main hyper parameters for tuning: criterion and splitter each with two options. After using grid search function, it was that the best hyper parameters combination is having criterion as *"entropy*" and splitter as *"best"*.

After fitting the models and applying them over the testing set, the following result was found:

Model Type	Accuracy	Precision	Recall	F1 Score
Decision Tree	0.901	0.859	0.861	0.860
GaussianNB	0.413	0.998	0.3761	0.546

It can be concluded from the table above that the decision tree outperforms GaussianNB model by having an F1 score of 0.86

Conclusion

To summarize, it was found that the successful projects on Kickstarter usually requires lees fund on average, get completed in shorter period compared to those unsuccessful ones, secure a lot of backers, and they are categorized mostly under music. Moreover, the project state can be predicted based on the four aforementioned criteria with an accuracy or 90% and an F1 score of 0.86