Google prediction API is a cloud service where google uses machine learning algorithms to analyze data and predict future labels for supervised data. Specifically, the prediction API is focused on regression and classification. It can classify categorical labels by using categorical models and predict a label for numerical labels by using regression models.

For this project, we have used Titanic survival data. It is given as CSV file in a Kaggle project. We divide the CSV file into two parts: training and testing data. Each record in the original file has seven features including Pclass, Sex, Age, Sibsp, Parch, Fare and Embarked and their labels are “dead” or “alive”. We want to determine the status of each passenger after the titanic ship was wrecked.

To solve this problem using google prediction API, we need to train a model and then use the predict functionality to test our model. Training of data is done online. For prediction we use a python file, which for each record in the testing file queries the model and returns an answer. Then it calculates the overall accuracy of predictions. The information of the project and model on google prediction API are as follows:

- Project Name: My Project 3
- Project Number: 256748303744
- Model ID: titanic5

The testing result of the model calculated by titanicprediction.py file is 81% correct.

Some of the advantages of using google prediction API are its ease of use and seed, sharing project between a team and the cloud base storage of the data. On the other hand, it only can be used for supervised data as a classification tool. This is not ideal, since in a lot of data mining situations clustering on unlabeled data is needed. Another disadvantage is that it uses regression model for numerical to determine a closet numerical label to testing features and categorical models for choosing a best category for testing data but sometimes it is better to use some other data mining algorithms like SVM to compare the results together and returns the best results.