# Literature Review

Group Number: **01**

Group members present (Name, ID):

* Jasleen Jasleen (0734327)
* Arvind Sharma (0730475)
* Harpreet kaur (0733894)

1. **Predicting Mortgage Approvals in the U.S.**

## Link to Project Report: Predicting Mortgage Approvals in the U.S. (2020). Retrieved 24 February 2020, from <https://jendefig.com/predicting-mortgage-approvals-in-the-u-s/>

In this project they are predicting about the mortgage’s approvals in the United State where they have a dataset with 500,000 observations of loan applications and nearabout 22 columns. There are many findings in the project but some of them are that they are a greater number of male applicants than the females, White and Asians people have larger number than Blacks, American Indians and many others in terms of applying for mortgage. They used 19 features for testing a model to get the better accuracy. They split the data into 70% and 30% ratio, and they used Cross Validation- Two Class Boosted Decision Tree Algorithm to get the best result.

1. **DSI Capstone Project- Predicting Loan Defaults.**

Link to Project Report: <https://medium.com/@arbasith/dsi-capstone-project-predicting-loan-defaults-4ebbad3292b4>

In this project the author has his personal interest in this topic that’s why he chooses this topic and he set an approach to complete the project in which he finalizes the steps and by following those steps he completes his project. The dataset is from Kaggle with 887383 rows and 75 columns. He removed the columns having more than 50% null values and then drop some columns. After that he perform Exploratory Data Analysis on features and applied three models on his dataset i.e. Logistic Regression, Random Forest and Gradient Boost. From above three Logistic Regression give better accuracy so he decided Logistic Regression as final model.

1. **MPP-in Data Science**

## Link to Project : gbganalyst/MPP-in-Data-science. (2020). Retrieved 24 February 2020, from <https://github.com/gbganalyst/MPP-in-Data-science/blob/master/Capstone_project.ipynb>

## In this Project the author is Ezekiel Adebayo Ogundepo and has done the project on June,2019.They have taken the government dataset and use the python cat booster model prediction classification technique to predict the rate of approval in the US. They have also done the data visualization using the bar plot only so what we can do differ is we can use tableau to see the different visualization technique. They have chosen the two different color coded for the visualization which we did not find good. They have also mention good finding of how to use cat booster also what are the advantage and disadvantages. The training accuracy they recorded is 74.

1. **Loan approval using Machine learning**

**Link to Project :** [https://copycoding.com/d/machine-learning-project-in-python-to-predict-loan- approval-prediction-part-6-of-6-](https://copycoding.com/d/machine-learning-project-in-python-to-predict-loan-%20%20%20%20%20%20%20%20approval-prediction-part-6-of-6-)

In this project work has been done by using the logistic regression on the different dataset which we are using . They have divided the dataset into two parts test and training with ratio 70:30. They have evaluate and train their model on decision tree as well and come across the accuracy of 0.64.

**5.**  **Predicting Loan Repayment**

<https://towardsdatascience.com/predicting-loan-repayment-5df4e0023e92>

In this project author is[**Imad Dabbura**](https://towardsdatascience.com/@ImadPhd?source=post_page-----5df4e0023e92----------------------)he has explained the critical question asked when people applied for loan that is how risky the borrower is and should they lend the money or not. The data covers the 9,578 loans funded by the platform between May 2007 and February 2010. Also explained What to do with missing values, Techniques used with imbalanced classification problems. Illustrate how to build an ensemble model using two methods: blending and stacking, which most likely gives us a boost in performance. They have use then ensemble methods which can be defined combing the base learner and meta learner and get the accuracy of 0.62