1. **Introduction. An overview of the project and an outline of the report.**

Mortgage balances and debt have been climbing in recent years, according to the Federal Reserve Bank of New York. Housing debt now totals $8.94 trillion, close to the $9.99 trillion peak of the third quarter of 2008. Mortgage debt is also the largest component of total household debt, making up 71% of total household debt. Given the debt crisis of a decade ago, understanding the mortgage and debt trends across the U.S. population is extremely relevant and can help address areas and communities that may be more vulnerable to the housing market.

The objective of this project is to analyze the socio-economic dynamics of mortgage and debt in the country and build a classification model to understand the demographic factors that affect debt across American society. The ultimate goal is to determine if any disparity may exist within the population and which demographic groups would be more or less affected by higher or lower debt. The analysis of mortgage and debt data will lead to conclusions about how wealth inequality is distributed throughout the population. The project uses data collected by the U.S. Census during the period 2012-2016 as part of the ACS 5-Year Documentation, provided on Kaggle.com by the Golden Oak Research Group.

1. **Description of the data set.**

The dataset includes 40,000 records and 80 features. Some of these features are demographic variables, while some are summary statistics.

Explain rent, homeowner costs, homeowner costs and mortgage, family income, household income, equity loan, debt, second mortgage.

Demographics: population, sex, marriage, age, education, area(state).

Type of data, observations and how disaggregated.

1. **Descriptionofthedataminingandlearningorcleaningalgorithmorotheralgorithmsthat you used. Provide some background information on the development of the algorithm and include necessary equations and figures.**

Classification model on: KNN, SVR, Logistic regression to ensemble.

Math equations and classic algorithm construction.

1. **Experimental setup. Describe how you are going to use the data to clean and preprocess. Explain how you will implement the data mining technique in the chosen software and how you will judge the performance. Write a complete report with theoretical description and verify this mathematical concepts with applying it with actual data. Provide enough information about the codes that you have written. Write your codes in sperate subroutines and call the functions if needed?. Explain each subroutine.**

Using Pycharm, we read data as dataframe and create class with methods to do preprocessing.

How we clean the data for missing data, outliers, normalized and then select relevant features.

How transform data types to apply methods properly.

How explored the data with graphs

3 models: split, scaling, encoding, testing, accuracy and viz of classifiers

Eventual ensemble: same

#Code lines

1. **Results. Describe the results of your experiments, using figures and tables wherever possible. Include all results (including all figures and tables) in the main body of the report, not in appendices. Provide an explanation of each figure and table that you include. Your discussions in this section will be the most important part of the report.**

Paste graphs for final outcomes and explain in details meaning of it

1. Summary and conclusions. Summarize the results you obtained, explain what you have learned, and suggest improvements that could be made in the future.
2. References.
3. A separate appendix should contain documented computer listings.