

**Nachiket Patwardhan**

**801208762**

**Project Proposal**

**ITCS 6156**

**Machine Learning**

**Project Title:** House price Prediction using Linear Regression.

**Project Idea:** During this globalization era, many people are interested in Investments. There are several objects that are used for investments such as gold, property, and stocks etc. In recent times, Property Investment has increased remarkably. Housing price trends directly indicate the current economic situation.

There are various factors which has impact on house prices such as Area, no. of bedrooms, no. of bathrooms etc. Even the Locality has a great impact on the House Prices. It also depends on whether House is in center of city, all the resources are accessible or not, whether it is close to highways or not.

Therefore, we chose to study the house pricing predictions problem, which will help us dig deep into variables and provide a model that will accurately predict the House Prices.

**Objective:** The aim is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and upcoming developments future prices will be predicted.

**Data:** We will be using the Housing Price.csv data set which includes various features which are derived from census. For e.g., Area, locality, #bathrooms, #bedrooms etc.

Link: <https://www.kaggle.com/camnugent/california-housing-prices/tasks>

**ML Technique:** We will be implementing Linear Regression for the problem statement. We think that Linear regression is one of the most efficient algorithms in this domain.

Linear regression is a very powerful and common method to estimate values, such as the price of a house, the value of a certain stock, the life expectancy of an individual, the amount of time a user will watch a video or spend in a website, etc.

Performance on linearly separable datasets

Linear regression fits linearly separable datasets almost perfectly and is often used to find the nature of the relationship between variables.

Overfitting can be reduced by regularization.

Overfitting is a situation that arises when a machine learning model fits a dataset very closely and hence captures the noisy data as well. This negatively impacts the performance of model and reduces its accuracy on the test set.

Regularization is a technique that can be easily implemented and is capable of effectively reducing the complexity of a function so as to reduce the risk of overfitting.

### **Study to be Done:**

Regression-Analysis: Whenever possible you should be making data-driven decisions at work. But there needs to be a know how to parse through all of the available data.

Regression analysis is a way of mathematically sorting out which of those variables does indeed have an impact. It answers the questions: Which factors matter most? Which can we ignore? How do those factors interact with each other? And, perhaps most importantly, how certain are we about all these factors?

The good news is that we likely do not have to do the number crunching ourselves, but you do need to correctly understand and interpret the analysis. One of the most important types of data analysis is regression.