|  |  |
| --- | --- |
| **Name** | **Email Id** |
| Preethi R | [preethipriya2128@gmail.com](mailto:preethipriya2128@gmail.com) |
| Matchimai.K.E | [matchierumporai@gmail.com](mailto:matchierumporai@gmail.com) |
| Mirudula . V | mirudula2503@gmail.com |
| Parkavi.K.S | parkavisathya982@gmail.com |



PREDICTING HOUSE PRICES USING MACHINE LEARNING

* **Introduction:**

The prediction of house prices is a critical task in the real estate industry and is of great interest to both buyers and sellers. It involves estimating the monetary value of residential properties based on various factors such as location, size, amenities, historical sales data, and market trends. Predicting house prices accurately can help prospective buyers make informed decisions and assist sellers in setting competitive listing prices. Machine learning, with its ability to process large datasets and identify complex patterns, has become an invaluable tool for house price prediction. In this project, we will focus on using the Random Forest algorithm to tackle this challenge .Random Forest is a powerful ensemble learning method that combines multiple decision trees to maker bust predictions. Random forest is an algorithm which can be used both for classification and regression. Random forest models are constructed by using a collection of decision trees based on the training data. Instead of taking the target value from a single tree, the Random forest algorithm makes a prediction on the average prediction of a collection of trees. The decision trees themselves are constructed by fitting to randomly drawn groups of rows and columns in the training data. This method is called bagging, and results in a reduction of bias as each tree is built on different parts of the input at random. The method of averaging the predictions of decision trees reduces the overfitting that can occur when using single decision trees.The number of trees in the Random forest is an important hyperparameter of the algorithm called ’n\_estimators’ and the more trees used the more will overfitting be prevented. The tradeoff however, is an increase in the computation time needed. ’n\_estimators’ will be tested with different values in this study. It is well-suited for house price prediction due to its ability to handle both numerical and categorical features, handle missing data, and mitigate overfitting.

**Content for phase 2 Project: Random Forest**

**Dataset link :(**[**https://www.kaggle.com/datasets/vedavyasv/usa-housing**](https://www.kaggle.com/datasets/vedavyasv/usa-housing)**)**

* **Program:**

**In [1]:**

model\_rf=RandomForestRegression(n\_estimators=50)

**In[2]:**

model\_rf.fit(X\_train\_scal,Y\_train)

**In[3]:**

print(r2\_score(Y\_test,Prediction2))

print(mean\_absolute\_error(Y\_test, Prediction2))

print (mean\_squared\_error(Y\_test, Prediction2))

* **Output:**

**Out[3]:**

-0.0006222175925689744

286137.81086908665

128209033251.4034

* **Project Conclusion:**

In this phase 2 conclusions, we will summarize the key findings and insights from the random forest technique .We will reiterate the impact of these technique on improving the accuracy and robustness of house price predictions.