**Mini Project Report on**



**Housing Price Prediction using Machine Learning**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

**Aman Verma 2016622**

***Under the Mentorship of***

**Mentor Name**

**Dr. Vikas Tripathi**

**Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**January 2023**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Housing Price Prediction using Machine Learning”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr. Vikas Tripathi, Professor**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name University Roll no **signature**

Aman Verma 2016622

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | **1-2** |
| Chapter 2 | Literature Survey | **3-4** |
| Chapter 3 | Methodology | **5-6** |
| Chapter 4 | Result and Discussion | **7** |
| Chapter 5 | Conclusion and Future Work | **8** |
|  | References | **9** |

**Chapter 1**

**Introduction**

* 1. **Introduction**

We have seen Machine Learning as a buzzword for the past few years, the reason for this might be the high amount of data production by applications, the increase of computation power in the past few years and the development of better algorithms.

Machine Learning is used anywhere from automating mundane tasks to offering intelligent insights, industries in every sector try to benefit from it. You may already be using a device that utilizes it. For example, a wearable fitness tracker like Fitbit, or an intelligent home assistant like Google Home. But there are much more examples of ML in use.

Prediction — Machine learning can also be used in the prediction systems. Considering the loan example, to compute the probability of a fault, the system will need to classify the available data in groups.

Image recognition — Machine learning can be used for face detection in an image as well. There is a separate category for each person in a database of several people.

Speech Recognition — It is the translation of spoken words into the text. It is used in voice searches and more. Voice user interfaces include voice dialing, call routing, and appliance control. It can also be used a simple data entry and the preparation of structured documents.

Medical diagnoses — ML is trained to recognize cancerous tissues.

Financial industry and trading — companies use ML in fraud investigations and credit checks.

In this project, we’ll predict future house prices. We’ll use data from the Federal Reserves, along with the house price data from Zillow. We’ll merge and combine this data, then use it to train a random forest model. The model will predict if house prices increases or decrease int the future. We’ll measure error using backtesting, then improve our model with new predictors.

In this project, we will develop and evaluate the performance and the predictive power of a model trained and tested on a data collected from a house in Boston’s suburbs.

Once we get a good fit, we will use this model to predict the mandatory value of a house located in Boston’s area.

A model like this would be very valuable for a real estate agent who could make use of the information provided in a daily basis.

**Chapter 2**

**Literature Survey**

The latest worldwide financial crisis restored a sharp enthusiasm toward both academic and strategy circles on the part of asset costs and specifically lodging costs clinched alongside monetary movement. As Lamer (2007) notes those lodging showcase predicted eight of the ten post globe War ii recessions, acting Concerning illustration and heading woman for that true segment of the economy. Truth be told he dives Likewise significantly Concerning illustration with state that “Housing is those benefits of the business cycle”. Vargas and silva (2008) contend that lodging costs alterations assume a paramount part in the determination of the stage of the business cycle. When those economy booms, development and work in the lodging division expand quickly should react should overabundance demand, quickly pushing ostensible house costs upwards. Throughout those withdrawal phase, the drop in private money lessens aggravate interest Also ostensible house costs. By ostensible house costs normally fall sluggishly since householders would unwilling on bring down their costs. The majority of the conformity will be attained through declines clinched alongside bargains volume bringing about A drop in the development segment and the lodging-built vocation. Moreover, throughout withdrawal and subsidence true house costs fall quickly Likewise general inflationary patterns diminish true house costs much with sticky perceived costs. Recently, a few writers scope to experimental discoveries that house costs can make instrumental molding to determining yield. (Forni etc., 2003; stock and Watson, 2003; Gupta Furthermore Das, 2010; das etc., 2009; 2010; 2011; Gupta and Hartley, 2013). Those lodging development division speaks to an expansive and only aggregate monetary action communicated in the GDP. Consequently, concerning illustration it reflects an extensive parcel of the general riches of the economy, house costs variances can make a pointer of the Development about GDP (Case etc., 2005). Concerning illustration, it is those body of evidence with different assets, those development for house costs can make Additionally a pointer of the future course from claiming expansion (Gupta Also Kabundi, 2010). Overall, exact determining of the Development way from claiming house costs could make a suitable apparatus both on house business members and fiscal strategy powers.

There is huge literature writing regarding U.S. house prices. Rapach Furthermore strauss (2007) use an auto regressive dispersed slack (ARDL) model framework, holding 25 determinants with conjecture genuine lodging cost development to the unique states of the elected Reserve’s eighth region. They discover that ARDL models tend should beat a benchmark AR model. Rapach and strauss (2009) augment those same examination on the 20 biggest u. Encountered with urban decay because of de industrialization, innovation developed, government agent. States dependent upon ARDL models looking at state, territorial and national level variables. When again, the creators scope comparative conclusions on the fact that joining together forecasts about models for different slack structure. Gogas and Pragidis (2011) utilize the hazard premium ascertained Likewise those Contrast the middle of Different long haul enthusiasm rates and the agents’ desires over future fleeting rates as information variable to foreseeing what's to come heading for house costs. They infer that gurus Also investigators could use adequately those majority of the data given by those investment rate hazard premium today so as should evaluate the likelihood from claiming acquiring beneath pattern S&P CS-10 list three months ahead. Gupta and Das (2010) also forecast the recent downturn in real house price growth rates for the twenty largest U.S. states. The authors use Spatial Bayesian VARs (BVARs), based only on monthly real house price growth rates, to forecast their downturn over the period 2007:01 to 2008:01. They find that BVAR models are well-equipped in forecasting the future direction of real house prices, though they significantly underestimate the decline. They attribute this under-prediction of the BVAR models to the lack of any information on fundamentals in the estimation process. Rapach and strauss (2009) expand the individual’s same examination on the 20 most amazing. Encountered with urban rot due to de industrialization, advancement developed, administration agonize. States reliant upon ARDL models taking a gander at state, regional Also national level variables. When again, those inventors’ degree similar finishes on the reality that joining together forecasts regarding models for separate slack structure. Gogas and Pragidis (2011) use the danger premium determined similarly the individual’s complexity those white collar for different whole deal energy rates and the agents’ longings in future transient rates as data variable with foreseeing what's with turn heading to house expenses. They construe that masters also investigators Might utilization enough the individual’s lion's share of the information provided for by the individuals financing rate danger premium today with the goal Similarly as ought further bolstering assess those probability from asserting securing underneath design S&P CS-10 rundown three months ahead.

**Chapter 3**

**Methodology**

Project steps

* Load in data
* Clean and merge data
* Create an initial machine learning model and estimate accuracy.
* Improve the accuracy of the model.
* Run diagnostics to figure out how we can improve.

Installation

To follow this project, please install the following locality:

* Jupiter Lab
* Python 3.8+
* Python packages
  + Pandas
  + yfinance
  + scikit-learn

Data

You’ll need to download a few csv files to run this project. These files are included in this report.

* Federal reserve data
  + CPI dataset – CPIAUCSL.csv
  + Rental vacancy rate – RRVRUSQ156N.csv
  + Mortgage interest rates – MORTGAGE30US.csv
* Zillow data
  + ZHVI

(raw, weekly)\_Metro\_zhvi\_uc\_sfrcondo\_tier\_0.33\_0.67\_month.csv

* + Median sale price (raw, all homes, weekly)

Metro\_medain\_sale\_price\_uc\_sfrcondo\_week.csv

**Chapter 4**

**Result and Discussion**

When the code gets executed first, we get outputs plots and then prediction takes place. These plots help us to understand the correlation between target variable (price) and different predictor variables. This plot gives a bar graph for bedrooms and number of houses. It is seen from dataset the count of 3-bedroom houses is greater in number and 7-bedroom houses are least in number.

**Chapter 5**

**Conclusion and Future Work**

This project can be customized to predict house prices in your metro area if you live in the US. We have managed out how to prepare a model that gives users for a novel best approach with take a gander at future lodging value predictions. A few relapse strategies have been investigated Furthermore compared, when arriving during a prediction strategy considering XG support. Straight former imply works bring been utilized within our model, something like that that future value predictions will have a tendency towards All the more sensible values. We concocted an approach with use similarly as considerably information as time permits for our prediction system, by adopting those ideas from claiming gradient boosting. Inspite of Hosting generated all the attempting provision that met our introductory requirements, there are Different upgrades that could be produced later. These incorporate upgrades we didn't settle on because of constrained duration of the time. A real worry for the prediction framework may be the stacking period. Moreover, our data set takes more than one day should prepare.

**References**

1. https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471

2.https://www.sas.com/en\_us/insights/analytics/machinelearning.html#machine-learning-importance

3. http://www.wired.co.uk/article/machine-learning-ai-explained

4. https://deeplearning4j.org/ai-machinelearning-deeplearning

5. David E. Rapach, Jack K. Strauss “Forecasting real housing price

growth in the Eighth District states”

6. Vasilios Plakandaras+ and Theophilos, Rangan Gupta\*, Periklis

Gogas “Forecasting the U.S. Real House Price Index”

7. Gupta and Das (2010) Forecasting the US Real House Price Index:

Structural and Non-Structural Models with and without Fundamentals

8. Rangan Gupta “Forecasting US real house price returns over 1831–

2013: evidence from copula models”