**PREDICTING HOUSE PRICES USING MACHINE LEARING**

|  |  |
| --- | --- |
| **Date** | **30-09-2023** |
| **Team Id** |  |
| **Project name** | **Predicting House Prices using Machine Learning** |

A**BSTRACT**

**The trend of the sudden drop or constant rising of housing prices has attracted interest from the researcher as well as many other interested people. There have been various research works that use different methods and techniques to address the question of the changing of house prices. This work considers the issue of changing house price as a classification problem and discuss machine learning techniques to predict whether house prices will rise or fall using available data.**

**House price prediction is a challenging task due to the complex and interrelated factors that affect home values. Machine learning (ML) algorithms can be used to develop models that can predict house prices with a high degree of accuracy. This paper presents a modular approach to house price prediction using ML.**

Introduction

* **In the ever-evolving world of technology, machine learning has become a powerful tool to tackle various real-world challenges. One such application is predicting house prices using linear regression for real estate. The ability to forecast property values can immensely benefit real estate agents, homeowners, and buyers alike. In this blog, we will explore a fascinating machine learning project that leverages the linear regression algorithm to predict house sale prices accurately using python.**
* **Development of civilization is the foundation of the increase in demand for houses day by day. Accurate prediction of house prices has been always a fascination for buyers, sellers, and bankers also. Many researchers have already worked to unravel the mysteries of the prediction of house prices. Many theories have been given birth as a consequence of the research work contributed by various researchers all over the world. Some of these theories believe that the geographical location and culture of a particular area determine how the home prices will increase or decrease whereas other schools of thought emphasize the socio-economic conditions that largely play behind these house price rises.**
* **We all know that a house price is a number from some defined assortment, so obviously prediction of prices of houses is a regression task. To forecast house prices one person usually tries to locate similar properties in his or her neighborhood and based on collected data that person will try to predict the house price.**
* **All these indicate that house price prediction is an emerging research area of regression that requires the knowledge of machine learning. This has motivated me to work in this domain.**

**MACHINE LEARNING**

* **It is a subset of artificial intelligence (AI). It provides system the ability to automatically learn and improve by itself. It focuses on the development of computer programs that can access data learn by themselves. The process of learning begins with observations based on the examples that we provide. The aim is to make computers to learn by itself without the need of a human.**

**MACHINE LEARNING METHODS**

* **Machine learning can be classified into three types namely the supervised, unsupervised and reinforcement learning. Supervised machine learning algorithms can apply what has been learned in the past to new data predict future events. It analysis from a known training dataset, and produces a functions to predict outputs.**

**SEMI-SUPERVISED MACHINE LEARNING ALGORITHMS**

* **Is a combination of both supervised and unsupervised learning, in semi-supervised learning, an algorithm learns from a dataset that includes both labeled and unlabeled data, usually mostly unlabeled. Generally, it is chosen when the sample data requires skilled resources in order to train from it. Otherwise, It doesn’t require additional resources.**

**REINFORCEMENT MACHINE LEARNING ALGORITHMS**

* **Is a learning method that works based on feedback. Reinforcement learning differs from supervised learning in not needing labelled input/output pairs be presented. It is studied in various disciplines such as statistics, information theory etc.**

**HOUSE PRICE PREDICTION WITH LINEAR REGRESSION INVOLVES FOLLOWING STEPS:**

**1.DATASET COLLECTION: Gather historical house price data and corresponding features from platforms like Zillow or Kaggle.**

**2.DATA PREPROCESSING: Clean the data, handle missing values, and perform feature engineering, such as converting categorical variables to numerical representations.**

**3.Splitting the Dataset: Divide the dataset into training and testing sets for model building and evaluation.**

**4.Building the Model: Create a linear regression model to learn the relationships between features and house prices.**

**5.model evaluation: Assess the model’s performance on the testing set using metrics like MSE or RMSE.**

**6.Fine-tuning the Model: Adjust hyperparameters or try different algorithms to improve the model’s accuracy.**

**7.DEPLOYMENT AND PREDICTION: Deploy the robust model into a real-world application for predicting house prices based on user inputs.**

**DEFINITION**

* **LITERATURE SURVEY**

This paper presents the results of the VR4RE (Virtual Reality for Real Estate) project, which aims at saving time and money for both real estate sellers and buyers by employing modern technologies. VR4RE is one of the innovative projects developed by Bluemind Software and it is in an advanced state. This paper also illustrates the history of in-house technological attempts at creating appropriate presentation tools for real estate properties with 3D and VR (Virtual Reality)

* **LIMITATIONS OF EXISTING SYSTEMS**

In India, there are multiple real estate classified websites where properties are listed for sell/buy/rent purposes such as 99acres, no broker, housing, magic bricks, and many more. However, in each of these websites, we can see a lot of inconsistencies in terms of pricing of a house and there are some cases when similar properties are priced differently and thus there is a lack of transparency and accuracy. Sometimes the customers may feel the value is not justified for a particular listed house but there is no way to confirm and check the data is accurate or not.

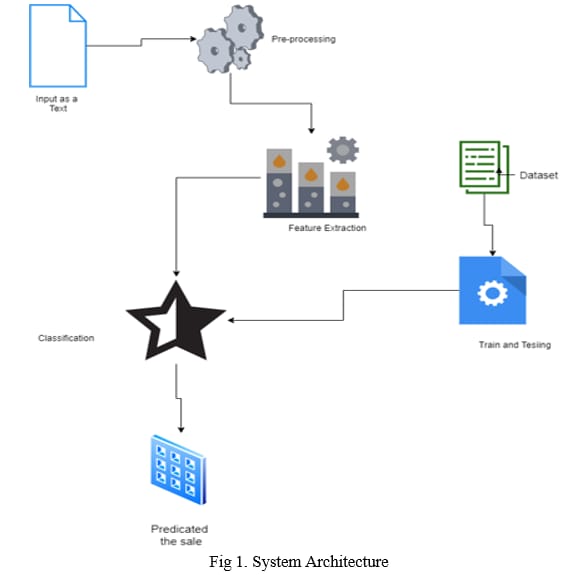
* PROPOSED SYSTEM
* Machine Learning is a field of Artificial Intelligence that enables PC frameworks to learn and improve in execution with the assistance of information. It is used to study the construction of algorithms that make predictions on data. Machine learning is used to perform a lot of computing tasks. It is also used to make predictions with the use of computers. Machine learning is sometimes also used to devise complex models.
* The principle point of machine learning is to permit the PCs to learn things naturally without the assistance of people. Machine learning is very useful and is widely used around the whole world.

**SYSTEM ARCHITECTURE**

In this system we take data as text input from the user and then we pre-process data of the user

Next we extract the required information from the data and then it is sent for classification.

In classification data is classified using train data set available in the system and using various algorithm price is predicted.



**FEATURES DESCRIPTION**

**Crim - per capita crime rate by town**

**zn - proportion of residential land zoned for lots over 25,000 sq.ft.**

**indus - proportion of non-retail business acres per town.**

**Chas - Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).**

**nox -** nitrogen oxides concentration (parts per 10 million).

**Rm -** average number of rooms per dwelling**.**

**Age -** proportion of owner-occupied units built prior to 1940

**Dis -** weighted mean of distances to five Boston employment centers.

**Tax -** full-value property-tax rate per $10,000.

**Ptratio -** pupil-teacher ratio by town**.**

**black -** 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.

**lstat** - lower status of the population (percent).

**Medv -** median value of owner-occupied homes in $1000s

**UNDERSTANDING LINEAR REGRESSION:**

Linear regression is a fundamental supervised learning algorithm in machine learning. It aims to establish a linear relationship between a dependent variable (target) and one or more independent variables (features). In the context of house price prediction, the dependent variable will be the house price, and the independent variables can be factors like the size of the house, number of bedrooms, location, etc.

**WOKING METHOD**

**Data collection:** Collect a dataset of house prices and features. The features can be quantitative (e.g., square footage, number of bedrooms, number of bathrooms) or qualitative (e.g., neighborhood, school district, proximity to amenities).

**Data collection:** Collect a dataset of house prices and features. The features can be quantitative (e.g., square footage, number of bedrooms, number of bathrooms) or qualitative (e.g., neighborhood, school district, proximity to amenities).

**Feature engineering:** Create new features that may be useful for predicting house prices. For example, you could create a feature that represents the number of bedrooms per square foot, or a feature that represents the distance to the nearest school

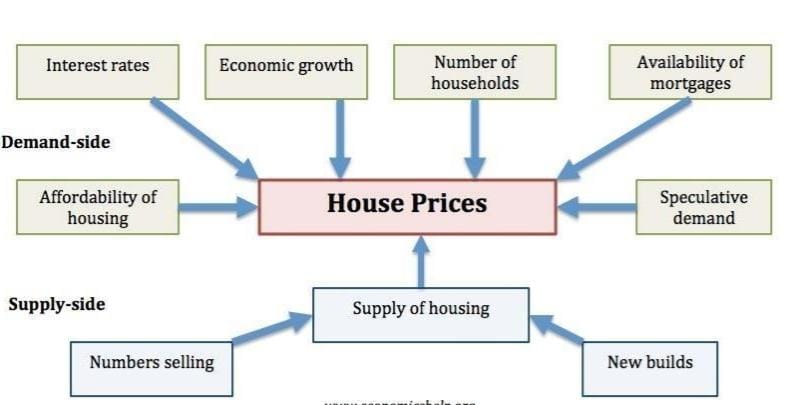
**Model selection:** Select a machine learning algorithm to train the model. Some popular algorithms for predicting house prices include linear regression, random forests, and gradient boosting machines.

**Model selection:** Select a machine learning algorithm to train the model. Some popular algorithms for predicting house prices include linear regression, random forests, and gradient boosting machines.

**Model evaluation:** Evaluate the performance of the trained model on a held-out test set. This will help you to assess how well the model will generalize to new data.

**Model deployment:** Deploy the model to production so that it can be used to predict house prices for new data

**FACTORS THAT AFFECT HOUSE PRICINGTop of Form**



**• ECONOMIC GROWTH - Demand for housing is dependent upon income. With higher**

**economic growth and rising incomes, people will be able to spend more on**

**houses; this will increase demand and push up prices. In fact, demand for housing**

**is often noted to be income elastic (luxury good); rising incomes leading to a**

**bigger % of income being spent on houses. Similarly, in a recession, falling**

**incomes will mean people can’t afford to buy and those who lose their job may**

**fall behind on their mortgage payments and end up with their home repossessed.**

**• UNEMPLOYMENT - Related to economic growth is unemployment. When**

**unemployment is rising,fewer people will be able to afford a house. But, even the**

**fear of unemployment may discouragepeople from entering the property market.**

**• INTEREST RATES - Interest rates affect the cost of monthly mortgage payments. A period of high interest rates will increase cost of mortgage payments and will cause lower demand for buying a house. High-interest rates make renting**

**relatively**

**• CONSUMER CONFIDENCE. Confidence is important for determining whether people**

**want to take therisk of taking out a mortgage. In particular expectations towards**

**the housing market is important; if people fear house prices could fall, people will**

**defer buying.**

**• MORTGAGE AVAILABILITY- In the boom years of 1996-2006, many banks were very**

**keen to lend mortgages. They allowed people to borrow large income multiples**

**(e.g. five times income). Also, banks required very low deposits (e.g. 100%**

**mortgages). This ease of getting a mortgage meant that demand for housing**

**increased as more people were now able to buy. However, since the credit crunch**

**of 2007, banks and building societies struggled to raise funds for lending on the**

**money markets. Therefore, they have tightened their lending criteria requiring a**

**bigger deposit to buy a house. This has reduced the availability of mortgages and**

**demand fell**

**CHALLENGES:**

1. **•COMPLEX AND NON-LINEAR RELATIONSHIPS:** House prices are influenced by a wide range of factors, many of which are complex and non-linear in their relationship with price. For example, the number of bedrooms may have a linear relationship with price, but the quality of construction may have a non-linear relationship, with a diminishing marginal return for each additional unit of quality.
2. **•LIMITED DATA:** It can be difficult to collect a large and representative dataset of house prices and features. This is especially true for newer homes, for which there may be less historical data available.
3. **BIAS IN THE DATA:** The data used to train a machine learning model may be biased, leading to inaccurate predictions. For example, a model trained on data from a particular city or region may not be generalizable to other areas.
4. **DYNAMIC MARKET CONDITIONS:** House prices can change rapidly due to a variety of factors, such as economic conditions, interest rates, and supply and demand. This makes it difficult to develop a machine learning model that can accurately predict prices over time.

**APPLICATION:**

1. **REAL ESTATE APPRAISAL**: Machine learning models can be used to automate the appraisal process, making it more efficient and accurate.
2. **HOMEBUYER DECISION-MAKING**: Homebuyers can use machine learning models to get a better understanding of the value of a home and to make informed decisions about buying or selling a home.
3. **REAL ESTATE INVESTMENT**: Real estate investors can use machine learning models to identify undervalued properties and to make more informed investment decisions.
4. **GOVERNMENT POLICY**: Governments can use machine learning models to track housing market trends and to develop policies to address housing affordability issue

**Conclusion:**

From the training and testing of dataset on the model, a strong deduction can be made that the model ANN works most effectively on the data producing lower levels of errors, and hence we can conclude that Deep Learning techniques prove useful in the implementation and estimation of **HOUSE PRICE PREDICTION**.