International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

House Price Predictor using ML through Artificial Neural Network

Kalaiselvi. S[1], Kokila. S[2], Bhavanathi. K[3], S. Saravanan[4], B.E, M.E, PHD

[1,2,3]UG Student, Department of CSE, Agni College of Technology, Chennai, India
[4]Department of CSE, Agni College of Technology, Chennai, India

---

Abstract - Housing price keep changing in day in and day out and sometimes are hyped rather than being based on valuation. Predicting housing prices with real factor is the main crux of our research project. Here we aim make our evaluation based on every basic parameter that is considered while determining the price. We use various regression techniques in this pathway using artificial neural network which yield minimum error and maximum accuracy than individual algorithms applied. We also propose to use real-time neighbourhood details using location to get exact real-world valuation.

1. INTRODUCTION

Using machine learning algorithms, we solve some application in the real-world problem but would not be complicated to implement. In this a house price prediction we using regression algorithms to predicate the price of the house. Machine learning helps to provide valid dataset that is input features are squares footage, number of bedrooms, etc. And applying regression techniques and future predictions the result is predicting exact price of the price. The problem statement is to predict the monetary value of house located in Bangalore with more accuracy using artificial neural network. To develop and evaluate the performance and predictive power of the model trained and tested on data collected from houses. In previous project is the system makes optimal use of Linear regression, Forest regression, Boosted regression. The efficiency of the algorithms has been further increased with use of Neural networks. A system that aims to provide an accurate prediction of housing prices has been developed. In our project we predict the house price for Bangalore city using various machine learning algorithms. The efficiency of the algorithm will be tested with R-Squared value. Our survey led to the conclusion that the actual real estate value also depends on nearby local amenities such as railways station, school, hospitals, etc. The modules are exploring and processing the data, Building and training with Machine Learning algorithm, comparing R-Squared value with ML algorithm, with highest R-Squared value will be implemented for the house price predicting, web development. The datasets which are used in project are Area-type, Availability, Location, BHK, society, Total square feet, bathrooms, balcony in machine learning the algorithms used in our project is supervised learning, Regression problem. So dataset was tested with several ML algorithm are linear regression, Decision tree regression, Random forest regression, Support vector Regression.

2. LITERATURE SURVEY

First we have investigated various papers and discussion on machine learning for house price prediction[1].The title of the papers is house price prediction is on machine learning and neural networks, the description of the paper is minimum error and maximum accuracy[2].Next title of the paper is Hedonic models based on price data from Belfast infer that submarkets and residential valuation this model is used to identified over a wider spatial scale and implications for the evaluation process related to the selection of comparable evidence and the quality of variables that the values may needed.[3]The title of the paper is understanding recent trends in house prices and home ownership in this paper they used feedback mechanism or social epidemic that encourages a view of housing as an important investment in the market.

3. METHODS AND ALGORITHMS

DATA COLLECTION

The dataset are collected from Bangalore house price. The dataset containing several features they are area type, availability, location, BHK, society, total squares feet, bathrooms, balcony. The area type is categorized into three types are super build-up area is already fully developed area, plot area is area of empty ground and build-up area is nothing but the area which is developing. Availability also categorized into ready to move, immediate position and others.

LINEAR REGRESSION

Linear regression is based on supervised learning. It performs the tasks to predict a dependent variable value(Y) based on given independent variable(X). It is relationship between input (X) and output (Y). It is one of the most well-known and well-understood algorithms in machine learning. The linear regression models are simple linear regression, Ordinary least squares, Gradient Descent, Regularization.

DECISION TREE REGRESSION

It is an object and trains a model in the structure of a tree to predict data in future to produce meaningful continuous output. The steps are involved in decision tree regression are the fundamental concepts of decision trees, Maximizing Information gain, Classification trees, Regression trees. The fundamental concepts of decision trees is it constructed from recursive portioning. The root node known as parent node, each node can be split into child nodes. These node can

became parent node of their resulting child nodes. The maximizing information gain is defined as the nodes at the informative features, to define an objective function that is to optimize the tree learning algorithm.

CLASSIFICATION TREES

Classification trees are used to predict the object into classes of a categorical dependent variable their measurement on one or more predictor variables.

REGRESSION TREES

It allows the input variables to be a continuous and categorical variables. Regression trees is considered as a research with several machine algorithm for the regression problem, Decision Tree algorithm has given the minimum loss. R-Squared value for Decision Tree is 0.998 which represent the good model. Web Development was completed using the Decision Tree.

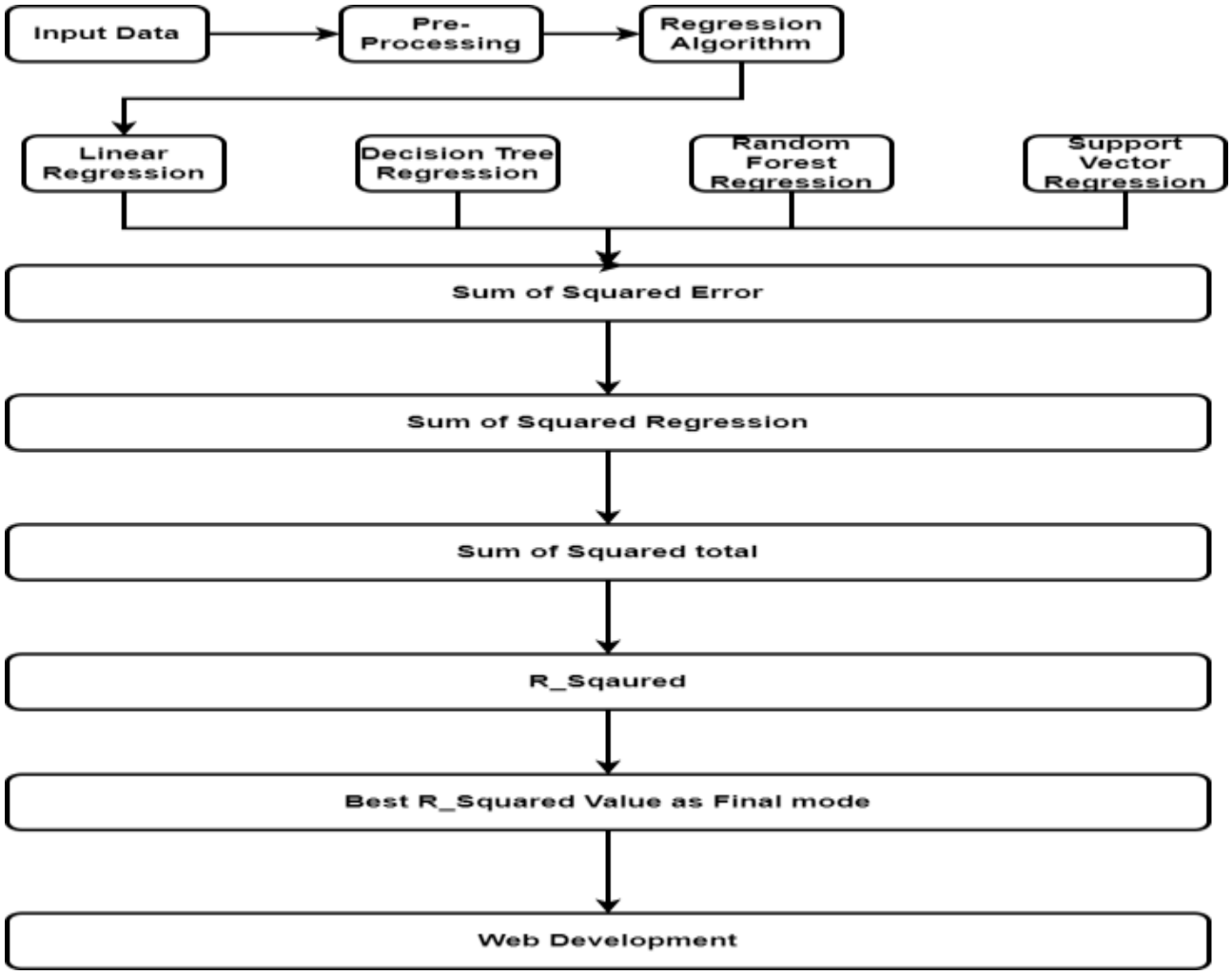
RANDOM FOREST REGRESSION

It is an important learning methods for classification and regression to operate a constructing a multiple of decision trees. The preliminaries of decision trees it is popular methods for various machine learning tasks. Tree learning requirements for serving n off the self-produce for data mining, because invariant under scaling and various other transformations. The trees are grown very deep to learn high regular pattern. Random forest is a way of averaging multiple deep decision trees trained set on different parts of same training set. This expenses of the small increase bias and some loss of interoperability.

SUPPORT VECTOR REGRESSION

The supervised learning is associated with learning algorithms that analyze data used for classification and regression analysis.

4. ARCHITECTURE DIAGRAM



5. RESULT AND CONCLUSION

```
Linear Regression
SST=214528158.9029079
SSE=162359500.76384386
SSR=52168658.139064044
214528158.9029079
The Sum of Square Total: 214528158.9029079
The Sum of Square Error: 162359500.76384386
The Sum of Square Regression: 52168658.139064044
R_Square: 0.24317860371269379
**************************************
Decision Tree
SST=214528158.9029079
SSE=295430.1153232588
SSR=214232728.78758466
214528158.9029079
The Sum of Square Total: 214528158.9029079
The Sum of Square Error: 295430.1153232588
The Sum of Square Regression: 214232728.78758466
R_Square: 0.9986228842086089
**************************************
```

```
Random Forest
SST=214528158.9029079
SSE=17379249.401793264
SSR=197148909.50111464
214528158.9029079
The Sum of Square Total: 214528158.9029079
The Sum of Square Error: 17379249.401793264
The Sum of Square Regression: 197148909.50111464
R_Square: 0.9189884932091417
**************************************
Support Vector Machine
SST=214528158.9029079
SSE=228740980.8212714
SSR=-14212821.918363482
214528158.9029079
The Sum of Square Total: 214528158.9029079
The Sum of Square Error: 228740980.8212714
The Sum of Square Regression: -14212821.918363482
R_Square: -0.06625154474381138
```

The research with several machine algorithm for the regression problem, Decision Tree algorithm has given the minimum loss. R-Squared value for Decision Tree is 0.998 which represent the good model. Web Development was completed using the Decision Tree.

The proposed system, predict the house price of the Bangalore city with several features. We have, tried with several Machine Learning algorithm to get best model. Compared to all the algorithm, Decision Tree Algorithm have produced very minimum loss and highest R-squared. We have developed the web development using Django Framework. It consist of eight features, which was the input of the model.

6. RESULT

Feature selection using select-K with chi-square parameters which have selected the highly correlated top five features from 50 feature input. The selected features are blood glucose random level, blood urea, serum creatine, packed cell volume and white blood count. We can able to predict tha risk factor of a patient with selected feature values.

Machine Learning Algorithm with selected featured, various Machine Learning Algorithms were tested. The algorithms are Support Vector Machine, Random Forest and Naïve Bay, Logistic, Decision Tree and K-NN algorithm. The highest accuracy obtained for the selected feature and support vector machine algorithm. The accuracy achieved was 95% for 5 feature input values.

Web Development, the proposed system was deployed using Django with selected feature value as input. This Web Development takes-time real prediction of the risk factor.

CONCLUSION

The HCC affected person's risk factor was classified with Support Vector Machine. This was achieved with feature selection method select –K parameter with chi-square. The effective five features were selected from 50 features using feature selection method. The result achieved was 95% accuracy. The trained model SVM for 5 features input are able to predict the low risk or high risk. Advantage of using feature selection has eliminated the unwanted feature which may increase the blood test cost of the person.

REFERENCES

1. Bird A.DNA methylation patterns and epigenetic memory. Genes Dev.2002; 16:6-21.

2. Dhanasekaran R, Limaya A, Cabrera R. Hepatocellular carcinoma: current trends in worldwide epidemiology, risk factors, diagnosis, and therapeutics. Hepat Med. 2012; 4:19.

3. Mizuno Y, Meamura K, Tanaka Y, et al. Expression of delta-like 3 is down regulated by aberrant DNA methlylation and histone modification in hepatocellular carcinoma. OncolRep. 2018; 39:220-2216.

4. Zhang Y, Petropoulos S, Liu J, et al. The signature of liver cancer in immune cells DNA methylation. Clin Epigenetic. 2018; 10:8.

5. Tsukuma H, Hiyama T, Tanaka S, et al. Risk factors for hepatocellular carcinoma among patients with chronic liver disease. N Engl J Med. 1993; 328(25): 1797-1801.

6. Yoshizawa H. Hepaticellular carcinoma associated with hepatitis C virus infection in Japan: Projection to other countries in the foreseeable future. Oncology 2002; 62 Supple 1:8-17.

7. Chen JD, Yang HI, Hoeje UH, et al. Carries of inactive hepatitis B virus are still at risk for hepatocellular carcinoma and liver-related death. Gastroenterology. 2010; 138(5): 1747-1754.

8. Nishida N, Nagasaka T, Nishimura T, Ikai I, Boland CR, et al.(2008) Aberrant methylation of multiple tumor suppressor genes in aging liver, chronic hepatitis, and hepatocellular carcinoma. Hepatology 47:908-918.

9. Feng Q, Stern JE, Haws SE, Lu H, Jiang M, et al.(2010) DNA methylation changes in normal sliver tissues and hepatocellular carcinoma with different viral infection. Exp Mol Pathol 88: 287-292.

10. Ishak KG, Sobin LH (1994) Histological typing of tumors in the liver (International histological classification of tumors 2nd ed). Berlin: Springer-Verlag.

11. Yeh CC, Goyal A, Shen J, et al. Global Level of plasma DNA Methylation is Associated with Overall Survival in Patients with Hepatocellular Carcinoma. Ann SurgOncol. 2017;24: 3788-3795.

12. Xu R, Wei W, Krawczyk M, et al. Circulating tumor DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat Mater. 2017; 16:1155.