

ABSTRACT

Machine Learning is a study of computer algorithm to improve automatically through experience and by use of data. It has been as a part of artificial Intelligence .Machine learning algorithms build a model based on sample data, known as “Training data”, in order to make predictions or decisions without being explicitly programmed to do so .Machine learning algorithm are used in wide variety of applications, such as medicine, email filtering and computer vision where it is difficult or unfeasible to develop a conventional algorithms to perform needed task.A subset of machine learning is closely related to computational statistics , which is focus on making predictions.

The Housing Price Prediction Using Concept of Machine Learning Has Become one of the most interesting application of Machine Learning Concept. To predict The Price of the House using Linear Regression Algorithm of Machine Learning . Regression is A Machine Learning tool That helps to make prediction by learning from the existing statistical data and this is done by finding relationship between Target Parameter and the set of Other Parameters. According to this definition , a House’s price depends on parameters such as Number of Bedrooms, Living Area , Location etc . If we apply Artificial Learning to these Parameters we can Calculate House Valuations In a given Geographical area.

Jupyter python programming software is used to design the code for Predicting the Housing Prices. It is an open source Software and it provides services for interactive computing across dozen of programming Languages .The software provides various functions and tools and In-Build Python Library so that operations can be performed accurately with maximum efficiency. with loops, functions variables, operators to perform various operations and obtain data. These software are very precise thus used in various research and analytical and educational related works.

TABLE OF CONTENTS

Chapter No.	Title	Page No.
	ABSTRACT	i
	LIST OF FIGURES	v
	LIST OF TABLES	vii
	ABBREVRATIONS	viii
1.	INTRODUCTION	01
	1.1 TYPES OF ALGORITHM USED	06
	1.1.1 K-Nearest Neighbor classification	06
	1.1.2 Support Vector Machine algorithm	10
	1.1.3 Naïve Bayes algorithm	12
	1.1.4 Decision Tree algorithm	15
	1.1.5 Random Forest classification	15
	1.2 ORGANIZATION OF THE PROJECT	
	WORK	17
2.	LITERATURE SURVEY	23
3.	AIM AND SCOPE OF PRESENT	
	INVESTIGATION	24
	3.1 AIM	24
	3.2 SCOPE	24
	3.3 PROBLEM DEFINITION	25

3.4 RELATED WORKS	25
3.5 EXISTING SYSTEM	26
3.5.1 Disadvantages of existing system	26
3.6 OVERVIEW OF PROPOSED SYSTEM	27
3.6.1 Advantages of proposed system	27
4. METHODOLOGY	28
4.1 HARDWARE REQUIREMENTS	28
4.2 SOFTWARE REQUIREMENTS	28
4.2.1 Overview of jupyter notebook	28
4.3 SYSTEM DESIGN	31
4.4 SYSTEM ARCHITECTURE	31
4.5 MODULE	31
4.5.1 Data set collection	32
4.5.2 Pre-processing of data set	32
4.5.3 Extraction of data set	33
4.5.4 Prediction of results	33
4.6 ALGORITHM DESCRIPTION OF METHODS	33
4.7 PERFORMANCE MEASURES	40
4.7.1 HRFLM ALGORITHMS	40
4.7.2 Algorithm1 Decision tree-based partition	41
4.8 Algorithm2 Apply ml to find less error rate	41

5.	RESULTS AND DISCUSSION	42
	5.1 RESULTS AND DISCUSSIONS	42
	5.2 PERFORMANCE ANALYSIS	42
	5.3 BENCHMARKING OF THE PROPOSAL MODEL	43
6.	SUMMARY AND CONCLUSION	44
	6.1 CONCLUSION	44
	6.2 FUTURE ENHANCEMENT	44
	REFERENCES	45
	APPENDIX	46
	A. SAMPLE CODE	46
	B. SCREENSHOTS	51
	C. PUBLICATION WITH PLAGIARISM	53
	REPORT	

LIST OF FIGURES

Fig. No.	Fig. Name	Page No.
Fig. no. 1.1	Linear Regression with two variables	2
Fig. no. 1.2	Linear Regression with Multiple variable	4
Fig. no. 1.3	Supervised Learning Architecture	5
Fig. no. 1.4	Different Clusters of data On graph	7
Fig. no. 1.5	Clustering of Red Circles using Blue Star	7
Fig. no. 1.6	Clustering Using Different values of K	8
Fig. no. 1.7	Graph of Error	9
Fig. no. 1.8	Graph of Validation Error	9
Fig. no. 1.9	Graph of Different Classes	10
Fig. no. 1.10	Graph of Segregating the classes	11
Fig. no. 1.11	Classification of different clusters	11
Fig. no. 1.12	Graph Of Different Classes	12
Fig. no. 1.13	Jupyter Open Source Software	18
Fig. no. 1.14	Browser View of Jupyter Notebook	20
Fig. no. 1.15	Sci – Kit Learn	21
Fig. no. 1.16	Supervised Learning Classification	21
Fig. no. 1.17	Supervised Learning Regression	21

Fig. no. 4.1	Browser View of Jupyter Notebook	29
Fig. no. 4.2	Architecture of Data Prediction	31
Fig. no. 4.3	Graph Of Classifier	37
Fig. no. 4.4	Overall error rate of Dataset	41
Fig. no. 5.1	Comparison Between Proposed and Existing Model	42
Fig. no. B.1	Scatter Matrix Of few Attributes	51
Fig. no. B.2	Positive Correlation Scatter plot	52
Fig. no. B.3	Negative Correlation Scatter plot	52

Table NO.	Table NAME	Page No.
Table 1.1	Linear Regression with two Variables	2
Table 1.2	Linear Regression with Multiple Variables	3
Table 1.3	Comparing Different ML Models	6
Table 1.4	Classification Of Fruits on Different Attributes	13
Table 3.1	Regression Analysis	39

ABBREVIATIONS

ABBREVIATION	EXPANSION
ML	-Machine Learning
SDK	-Software Development Kit
KNN	-K Nearest Neighbour
SVM	-Support Vector Machine
CAS	-Carotid Artery Stenting
HRFLM	-Hybrid Random Forest with Linear Model
IHDPS	-Intelligent Heart Disease Prediction System
MLP	-Multi Layer Perception
CNN	-Convolutional Neural Networks
CPU	-Central Processing Unit
RAM	-Random Access Memory
RF	-Random Forest
LM	-Linear Method

CHAPTER 1:

INTRODUCTION

Housing Price Prediction is commonly used to estimate the change in the Housing Price. Since Housing Price is Strongly Correlated to other Factors such as location , area , population ,it requires other information apart from HPI to predict Individual Prediction Prices. However to Predict the Housing Price our Machine Learning model requires data About more Number of Features so that prices can be Predicted more accurately , To do this we use the Machine Learning Algorithm called as Linear Regression . And using the Algorithm we will see effect of different features eg location , area, Furnishing which acts as Independent variables on The price of our house which is a Dependent Variable .

Linear Regression is one of the most Easiest algorithm of Machine learning . It is basically a statistical model that attempts to show the relationship between two variables with a linear equation. Linear Regression is a Supervised Learning Algorithm where the Predicted Output is Continous and has a constant slope . It's used to predict value within a continuous range (eg sales,price) rather than trying to classify them into catogeries (eg Black,Blue) .

Simple regression

Simple linear regression uses traditional slope-intercept form, where mm and bb are the variables our algorithm will try to “learn” to produce the most accurate predictions. xx represents our input data and yy represents our prediction.

$$y=mx+b$$

Let's say we are given a datasett with the following columns (features): how much a company spends on Radio advertising each year and its annual Sales in terms of units sold. We are trying to develop an equation that will let us to predict units sold based on how much a company spends on radio advertising. The rows (observations) represent companies.

TABLE 1.1 : Company Table for Regression with two variables

Company	Radio (\$)	Sales
Amazon	37.8	22.1
Google	39.3	10.4
Facebook	45.9	18.3
Apple	41.3	18.5

Our prediction function outputs an estimate of sales given a company's radio advertising spend and our current values for *Weight* and *Bias*.

$$\text{Sales} = \text{Weight} \cdot \text{Radio} + \text{Bias}$$

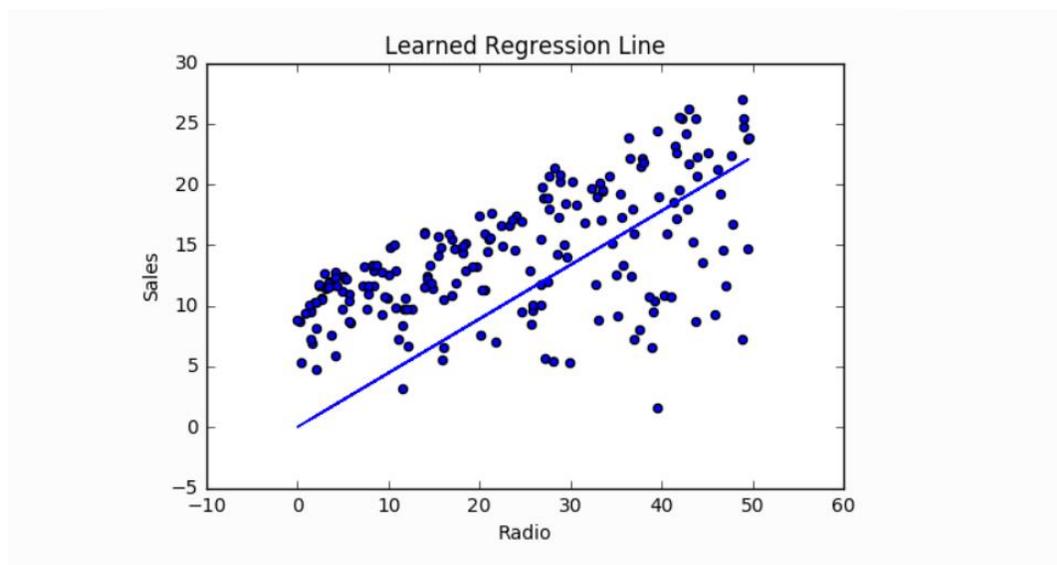
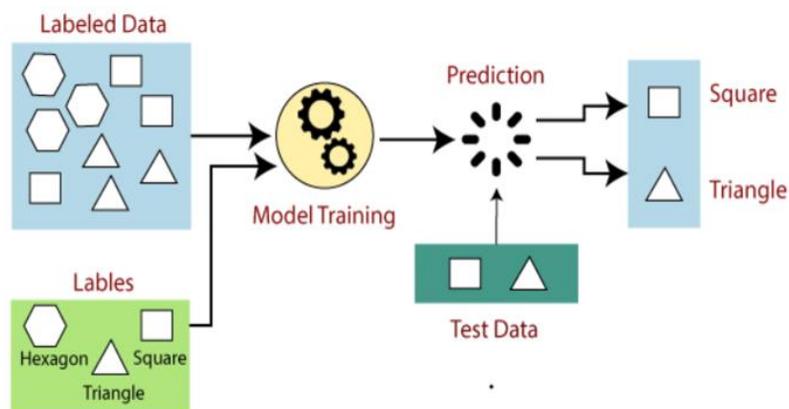


FIGURE 1.1: Linear Regression with Two Variables

HOW SUPERVISED LEARNING WORKS ?

In the supervised learning, models are trained using labelled dataset , where the model learns about each type of data. Once the Learning process is completed , the model is tested on the basis of test data (a subset of the training data) , and then it predicts the output.

FIGURE 1.3:Supervised Learning Architecture



The working of supervised learning can be easily understood by the above example diagram. Suppose we have a dataset of different types of shapes which includes squares, rectangle ,triangle and polygon . Now the first step is that we need to train the model for each shape.

- If the given shape has four side , and all the sides are equal , then it will be labelled as a square .
- If the given shape has three sides, then it will be labelled as a triangle.
- If the given shape has all six equal sides then it will labelled as an hexagon. Now after training , we test our model using the test set , and the task of the model is to identify the shape. The Machine is already trained on all types of shapes, and when it finds a new shape , it classifies the shape on the bases of a number of sides, and predicts the output.

1.1 TYPES OF ALGORITHM USED