House Rent Prediction using polynomial and linear regression

Amit Roy [1], Sirajum Maria Muna [2], Shekhor Chandra Saha [3], and Saniat Injam [4]

[1]ID: 2017-3-60-021
[2]ID: 2017-3-60-020
[3]ID: 2017-3-60-025
[4]ID: 2017-3-60-093

I. INTRODUCTION

A. 1.1 Objectives

With the mass growth of population around the world, it has become a challenge to provide accommodation for the people. As well as the growth of population the housing prices are also becoming high. For this reason, people opting for house renting as, owning a house seem like a dream like to some people. In this project our aim is to find house rent for a certain area in different cities. We are using a combination of linear regression and polynomial regression to predict house rent prices.

B. 1.2 Motivation

As we live in a third world country, also a over populated one, finding a good place to live a life peacefully is not easy. Along with ourselves, we want to secure a good future for our next generations. But money is always a problem. So, in this project we are motivated by this idea and tried to do a better approach, so that, it can help us by predicting the rental price to make us learn, on which area has better facilities in our suited budget. If anyone wants to further invest on real estates, this approach can help them. Also, it can be the parameter to signify one area's demand.

1.3 Existing works

Several machine learning algorithms have been used in the past years in this field. A study has been done in Fairfax County, Virginia using several machine learning algorithms like Naïve Bayes, Ripper, Decision Tree, Adaboost etc. to predict the house prices.1 Ridge and Lasso Linear regression is being used in the area of Ames and Lowa. 2 Also, Hybrid Linear Regression is being used by combining Ridge and Lasso along with Gradient Boosting combination to find more accurate predicted values with higher accuracy. [?] By all these studies, what we have done till now, we have seen using hybrid machine learning algorithms gives better result than only one algorithm. Also, regression is a common classifier in that field. So, we have chosen to work on this part in our project.

1.4 Necessity

As we have discussed in the motivation part, house rent prediction work can be used in daily basis to predict rental prices. In house rental dataset there can be several fractional values as well as non-numerical values. So, using regression is the safe option, as it works with fractional values and the non-numeric values can be pre-processed before the algorithm runs. Being less time consumption is another plus point of regression. The rental price prediction approach can ease many manual work and gives people high efficient work facilities.
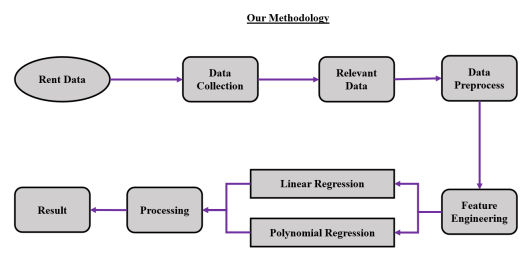
II. METHODOLOGY

Fig. 1. Diagram of our proposed model

III. IMPLEMENTATION

A. 3.1 Data collection

The data is collected from Kaggle. The dataset is based on the house rental prices of different cities in India. There are 1,93,011 rows and 10 columns. The price attribute is the dependent attribute where the other nine attributes (Seller-Type, Bedroom, Layout-Type, Property-Type, Area, Furnish-Type, Bathroom and City) are independent attributes.

B. Model development

We have worked with 2 types of regression: linear and polynomial regression and will estimate the accuracy between the models. In our final dataset we have 193011 rows and 21 columns. The dataset is divided according to the attributes in two sets: one is independent set, and another is dependent set. Then we have divided them by test (20%) and train (80%) data. In train set, there are 154408 data and in test set, we have 38603 data.

Linear Regression(Simple and Multiple): Linear Regression is one of the well-known algorithms in machine learning.

Basically, Linear Regression is used for predictive analysis. Linear regression is linear model that assumes a linear relationship between one or more independent variables and one dependent variable. There is a single input variable it is called simple linear regression. When there is more than one input variable it is called multiple linear regression.

Equation of simple linear regression:

$$Y = a + bX$$

Where,
X is the explanatory variable,
Y is the dependent variable,
B is the slope of the line,
a is the intercept (the value of y when x = 0).

Equation of Multiple linear regression:

$$yi = \beta0 + \beta1xi1 + \beta2xi2 + ... + \beta pxip + \epsilon$$

where,
for i=n observations:
yi is the dependent variable,
xi is the explanatory variables,
$\beta0$ is the y-intercept (constant term),
βp is the slope coefficients for each explanatory variable,
ϵ is the model's error term (also known as the residuals).

Polynomial Regression: Polynomial is one of the forms of regression analysis. Nonlinear relationship are fit by polynomial regression between the independent value and the corresponding conditional mean of dependent value.

Equation of Polynomial regression:

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + ... + \theta_n x^n$$

Where,
$\theta0$ is the bias,
$\theta1, \theta2, ..., \theta n$ are the weights in the equation of the polynomial regression,
and n is the degree of the polynomial.
The number of higher-order terms increases with the increasing value of n.