Hyperiondev

TASK

Capstone Project - Linear Regression in Action

Visit our website

Introduction

WELCOME TO THE LINEAR REGRESSION IN ACTION CAPSTONE PROJECT!

Get ready for an engaging capstone project! In this project, you'll utilise the 'ames.csv' dataset to predict sale prices for houses in Ames, Iowa using multiple linear regression. Gain valuable insights into the factors influencing house prices and make informed recommendations to optimise sales strategies in the residential property market.

PROBLEM

The ames.csv dataset provides valuable insights into the housing market in Ames by recording various features of houses in Ames including the size of the living area of a house, the year the house was built, and the sale price of the house. The main objective of this task is to utilise this dataset to predict the sales prices of houses based on the available features. By employing a multiple linear regression model, we aim to uncover the underlying relationships between the independent variables and the dependent variable. Through further exploration and analysis, you will gain valuable insights into the factors influencing sale prices within the housing market. By developing an accurate predictive property value model, real estate agents and others can make informed decisions regarding listing prices, property taxes, and property investment. This task will challenge you to extract meaningful insights and provide recommendations that align with the specific needs of the property industry.

RECAP

When predicting the sale prices of houses, a few essential steps need attention. First, preparing the data set is crucial to avoid biased and inaccurate analysis. Second, accurately splitting the dataset into appropriate independent and dependent variables ensures relevant model training. Finally, to create an error plot, start by obtaining the predicted values and the corresponding actual values from your model's predictions. Then, utilise a plotting library such as matplotlib or seaborn to generate a scatter plot.

Capstone Project

Follow these steps:

- Read the **ames.csv** file into the Jupyter notebook, **Linear_Regression_Ames.ipynb**.

- Clean and prepare the data set if necessary.

- Perform exploratory data analysis to gain insights into the dataset by visualising the distributions of the dependent variable and independent variables and identifying any patterns or trends in the data.

- For the linear regression model use the following independent variables:
  - Gr_Liv_Area: size of above grade, ground living area in square feet
  - Garage_Area: size of garage in square feet

- Split the dataset into the independent variables and the single dependent variable.

- Generate plots to explore the relationships between the independent variables and the dependent variable.

- Split the data into training and test sets using a split ratio of 75:25.

- Build a multiple linear regression model using a training set with all the independent variables.

- Print out the intercept and coefficients of the trained model.

- Generate predictions for the test set.

- Evaluate the model's performance by computing the mean squared error (MSE) or root mean squared error (RMSE) on the test set.

- Generate an error plot to visualise the differences between the predicted and actual values in the test set.

- Print the coefficients and interpret them within the context of the median value prediction.

- Summarise the findings from the analysis, including insights from the exploratory data analysis, model performance, and any notable observations within the notebook.

Rate us
Share your thoughts

HyperionDev strives to provide internationally-excellent course content that helps you achieve your learning outcomes.

Think that the content of this task, or this course as a whole, can be improved, or think we've done a good job?

Click here to share your thoughts anonymously.