Summer 2021: CSEE5590 – Special Topics

Python_Lesson_3_Part_1: Machine Learning: Regression

Lesson Overview:
In this lesson we will review regression techniques
Regression techniques

a. Linear Regression
b. Multiple Regression

Use Case Description:
Multiple Linear Regression

Programming elements:
Linear Regression and Data Analysis

Source Code:
Provided in the assignment repo & Canvas use-case file.

Assignment:
For question 1 use the same dataset used in the source code (House Prices).

1.Delete all the outlier data for the GarageArea field **(for the same data set in the use case: House Prices)**.

* for this task you need to plot GaurageArea field and SalePrice in scatter plot, then check which numbers are anomalies.

2. Evaluate the model using MAE, MSE, RMSE and R2 score.

3. Plot the regression line.

For questions 2 and 3 use the Restaurant Revenue Prediction dataset uploaded here:
https://www.kaggle.com/c/restaurant-revenue-prediction

- Id : Restaurant id.
- City Group: Type of the city. Big cities, or Other.
- Type: Type of the restaurant. FC: Food Court, IL: Inline, DT: Drive Thru, MB: Mobile
- P1, P2 - P37: There are three categories of these obfuscated data. Demographic data are gathered from third party providers with GIS systems. These include population in any given area, age and gender distribution, development scales. Real estate data mainly relate to the m2 of the location, front facade of the location, car park availability. Commercial data mainly include the existence of points of interest including schools, banks, other QSR operators.
- **Revenue**: The revenue column indicates a (transformed) revenue of the restaurant and **is the target of predictive** analysis. Please note that the values are transformed so they don't mean real dollar values.

2. Create Multiple Regression for the "**Restaurant Revenue Prediction**" dataset.

3. Find top 5 most correlated features to the target label(revenue) and then build a model on top of those 5 features. Evaluate the model using MAE, MSE, RMSE and R2 score and then compare the result with the RMSE and R2 you achieved in question 2.

Online Submission Guidelines (for Online students):

1. Submit your source code and documentation to GitHub and represent the work in a ReadMe file properly (submit your screenshots as well. The screenshot should have both the code and the output)
2. Comment your code appropriately
3. Video Submission (1 – 3 min video showing the demo of the assignment, with brief voice over on the code explanation)

Note: *Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL: https://catalog.umkc.edu/special-notices/academic-honesty/*