Assignment 3 Random Forest and Linear Regression (100 points)  
CSCE 415: Machine Learning  
Spring 2021  
100 Points

Due: 31 March 2021 11:30 PM

Objective: Compare Random Forest and Linear Regression Models on the Melbourne Housing Market Dataset.

Task 1 -- Random Forest

You have been provided the data set from the Melbourne Housing market dataset which contains data about the housing market and housing prices in Melbourne Australia. Your task is to develop a model using the RandomForestRegressor to predict housing prices.

Task 2 -- Linear Regression

For this problem you have use the same file, the Melbourne Housing market dataset. However, this time, you want to use the LinearRegression model to predict housing prices. You can make adjustments to the data, i.e. drop or fill in missing values differently, etc., if you think you can improve the Linear Regression model from the Random Forest model, but make sure it is clear why you made those decisions and why you didn’t make those decisions with the Random Forest Model.

Submission Instructions:

1. Turn in your Jupyter Notebook (\*.jpynb) file – do not turn in your data.
   1. All changes to the data must be made through your code within Jupyter Notebook.
   2. Use relative pathnames, if possible.
2. Comparative Analysis (See Rubric)
3. Label folder and zip file *lastname-3*, and *lastname-Ass3.zip* respectively.

Melbourne Housing Data Information

Some Key Details Suburb: Suburb

Address: Address

Rooms: Number of rooms

Price: Price in Australian dollars

Method: S - property sold; SP - property sold prior; PI - property passed in; PN - sold prior not disclosed; SN - sold not disclosed; NB - no bid; VB - vendor bid; W - withdrawn prior to auction; SA - sold after auction; SS - sold after auction price not disclosed. N/A - price or highest bid not available.

Type: br - bedroom(s); h - house,cottage,villa, semi,terrace; u - unit, duplex; t - townhouse; dev site - development site; o res - other residential.

SellerG: Real Estate Agent

Date: Date sold

Distance: Distance from CBD in Kilometres

Regionname: General Region (West, North West, North, North east ...etc)

Propertycount: Number of properties that exist in the suburb.

Bedroom2 : Scraped # of Bedrooms (from different source)

Bathroom: Number of Bathrooms

Car: Number of carspots

Landsize: Land Size in Meters

BuildingArea: Building Size in Meters

YearBuilt: Year the house was built

CouncilArea: Governing council for the area

Latitude: Self explanatory

Longitude: Self explanatory

Grading Rubric

Data Inspection 10%

Data Visualization 20%

Data Cleaning 20%

Train both Models and Compare and Evaluate 15%

Fine Tune each Model 15%

Evaluate your Results 20% (Explain what the values mean in the narrative, do not just show numbers)