

APPLE DEVELOPERS GROUP
Machine Learning

Exercise 1: Pandas & Numpy

DATASET: Corona Cases in India

Dataset Link: Included in the zip file 'corona_dataset.csv'

Objective:

- 1. Perform data preprocessing for the following Regression model.**
- 2. Implement Root Mean Square Error function with numpy**

Steps To Be Followed:

1. Use Pandas to read the data from csv files.
2. Preprocess the data if required(Remove empty rows and columns and unrelated attributes)
3. Use Sklearn package to build the RandomForestRegressor (given)
4. Complete the RMSE function

Note:

For each operation performed in data preprocessing, there should be a comment describing the operation and the insight that led to the operation.

Bonus:

- 1. Use matplotlib to visualize the total number of cases for different occupations.**
- 2. Calculate the correlation of Diuresis with Infect_Prob using numpy**

