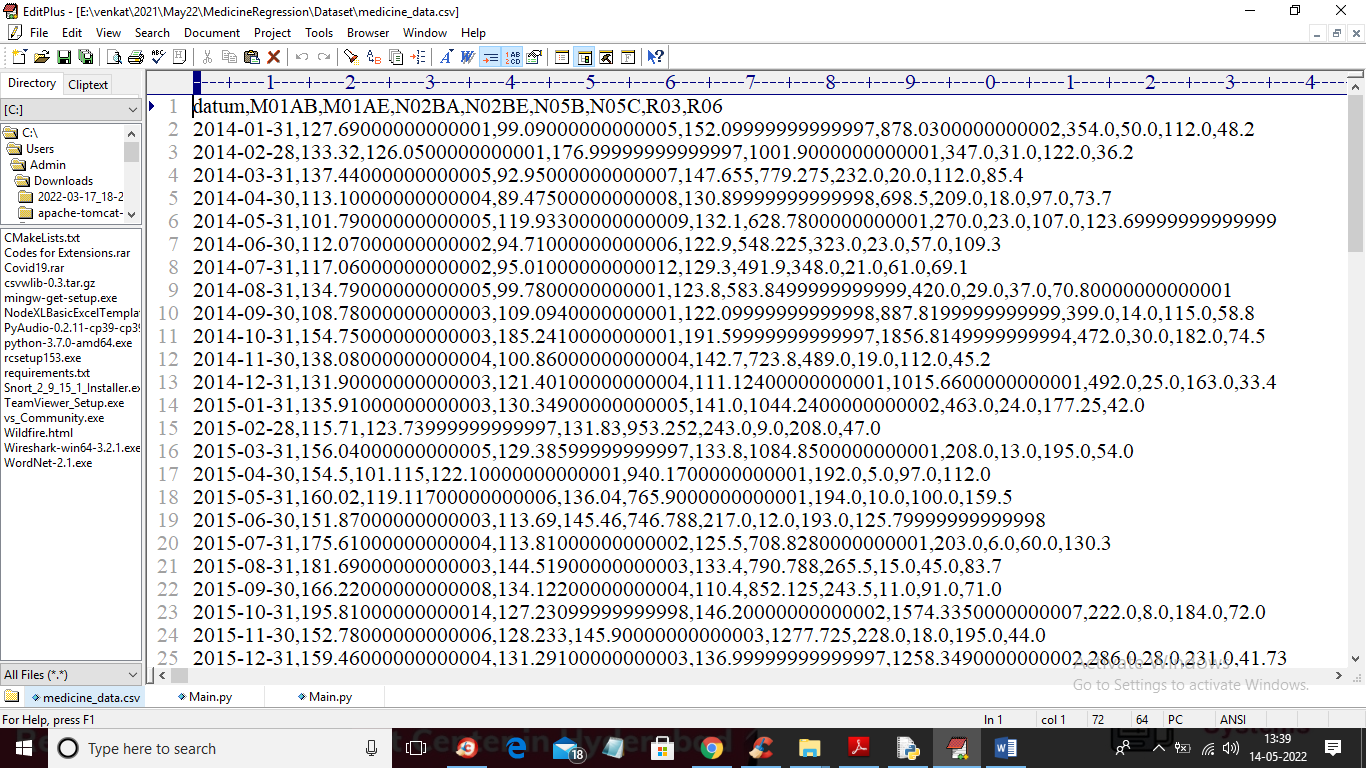
A Novel Method for Computationally Efficacious Linear and Polynomial Regression Analytics of Big Data in Medicine

In propose paper author is saying machine learning algorithms exists from centuries and often almost all algorithms models are not accurate and perform wrong prediction and to overcome from this problem author is suggesting to apply optimization techniques to algorithm to perform accurate prediction. In propose paper author is applying Linear & Polynomial optimization technique to Regression algorithms and then compare its performance without optimized Regression algorithm and evaluate its performance in terms of SUM OF SQUARE ERROR (SSE0. For any Regression algorithm the lower the SSE the better is the algorithm.

Optimization algorithms means tuning algorithm to get better result and in propose paper author using Regression with LINEAR and POLYNOMIAL. Optimized regression giving less SSE compare to pre or without optimize Regression algorithm.

In propose paper author applying Regression on medicine dataset called EMBASE but this dataset not available so we are using medicine SALES and MANUFACTURING dataset which will predict manufacturing quantity medicines for sales.

This dataset get train with PRE & POST optimized Regression algorithm and below screen showing dataset details



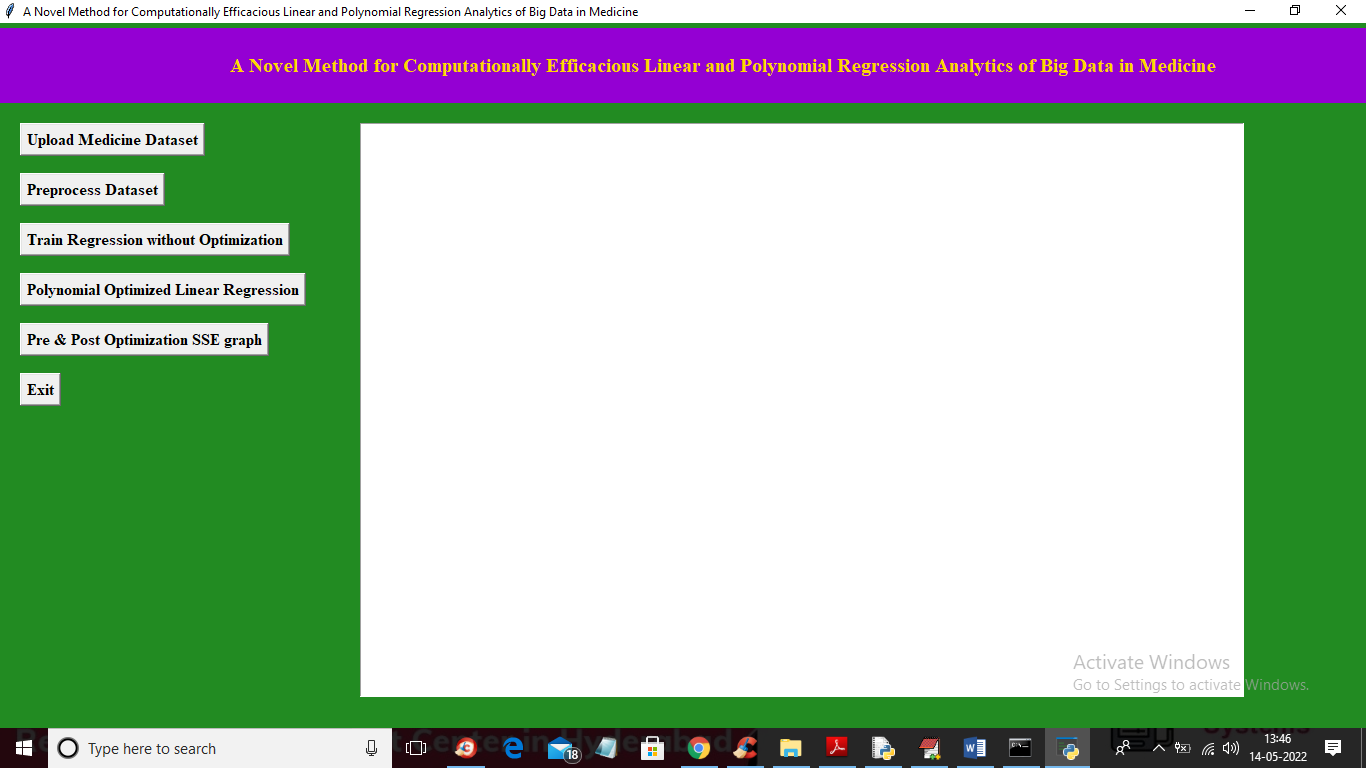
In above screen first row contains dataset column names and remaining rows contains dataset values. In above dataset last column contains manufacturing quantity and we will use this values as labels and remaining values will used for training.

To implement this project we have used following modules

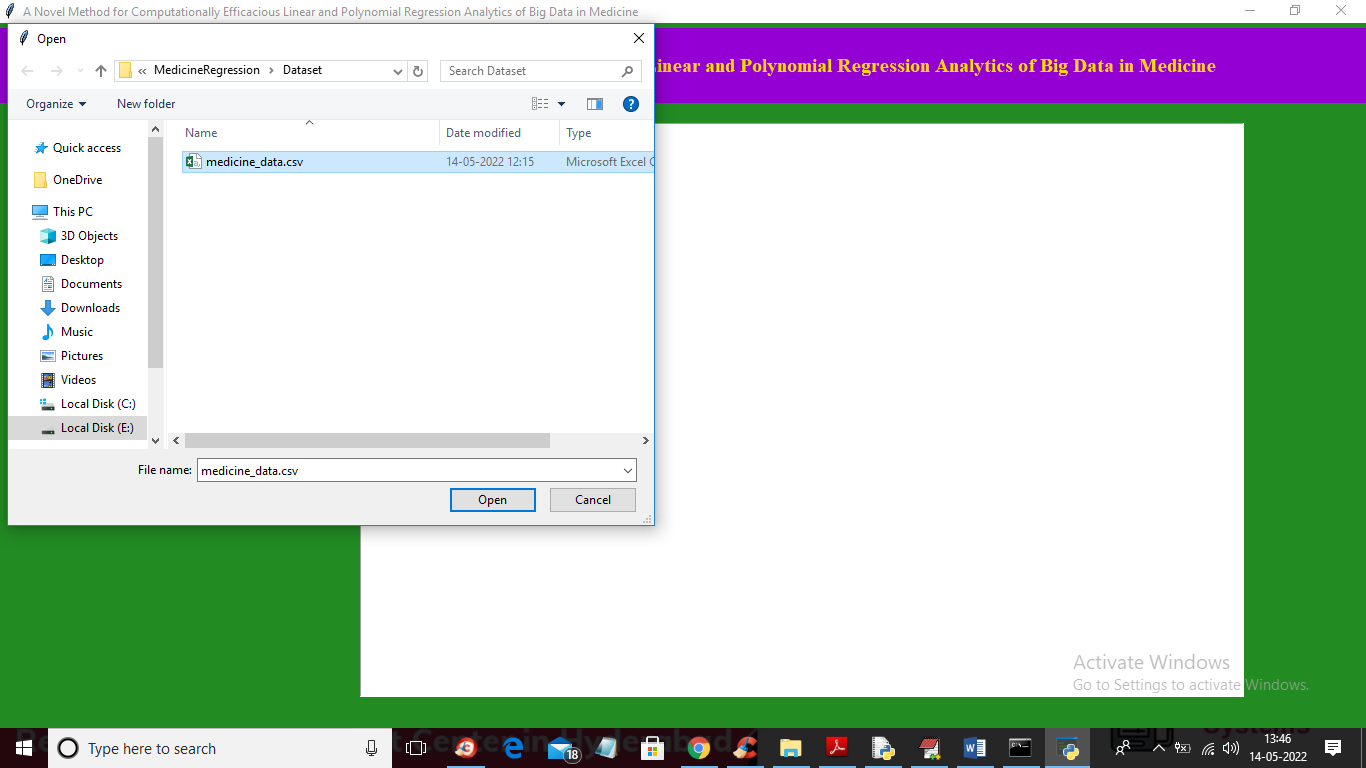
1. Upload Medicine Dataset: using this module we will upload dataset to application
2. Preprocess Dataset: using this module we will read entire dataset and then extract training X features and Y labels and then split dataset into train and test
3. Train Regression without Optimization: above processed features will be input to Regression algorithm without optimization and then trained a model and this model can be used to predict medicine manufacturing form test data and then calculate SSE error which refers to difference between original test and predicted values
4. Polynomial Optimized Linear Regression: using this module we will train Regression with Liner and Polynomial features and then trained a model for prediction
5. Pre & Post Optimization SSE graph: using this module we will plot SSE error between PRE and POST optimization

SCREEN SHOTS

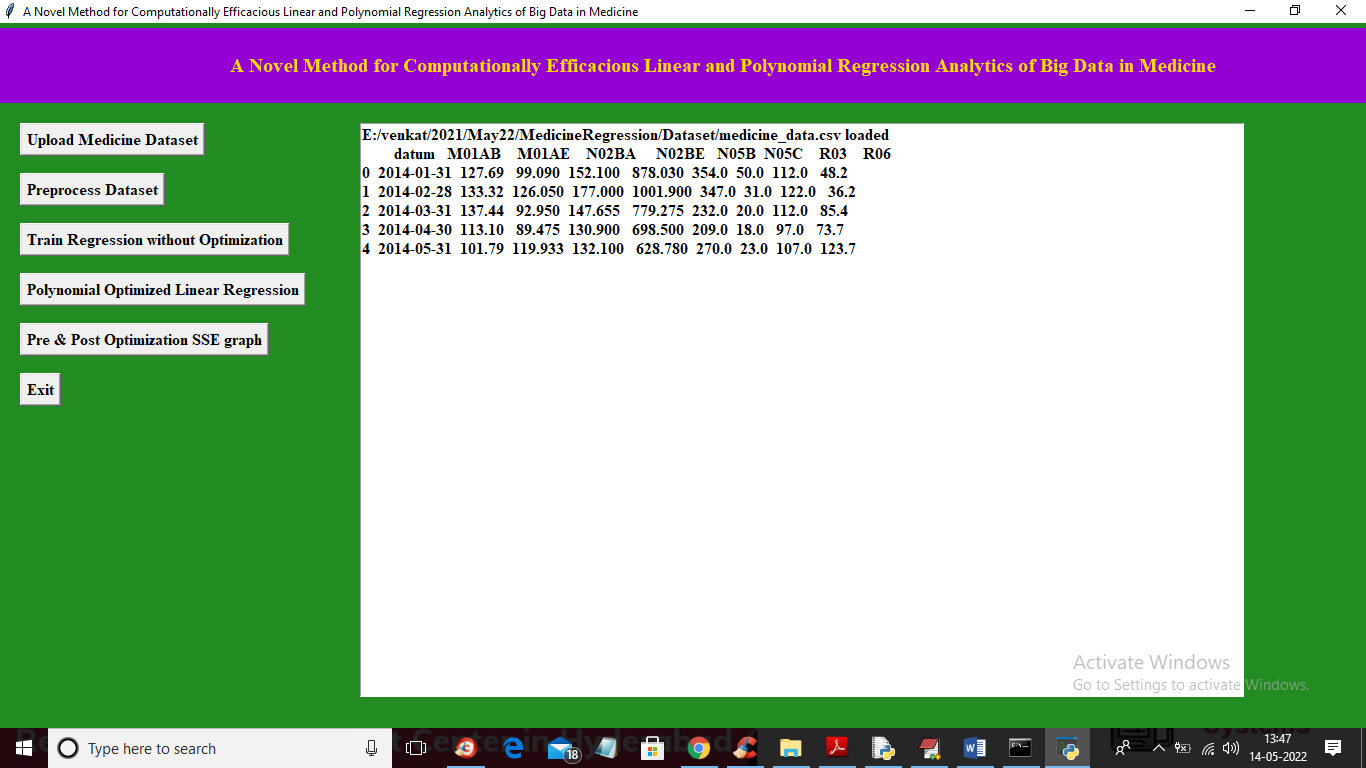
To run project double click on ‘run.bat file to get below screen



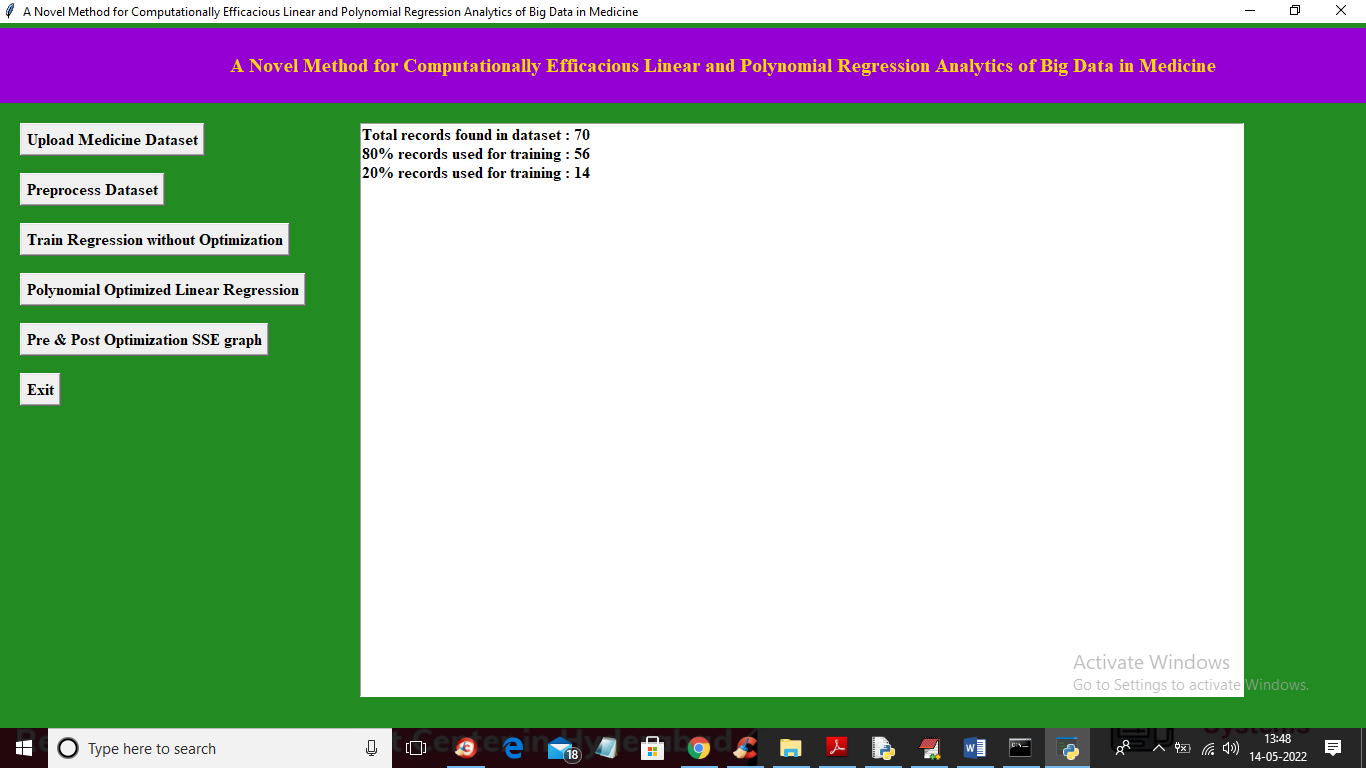
In above screen click on ‘Upload Medicine Dataset’ button to upload dataset and get below output



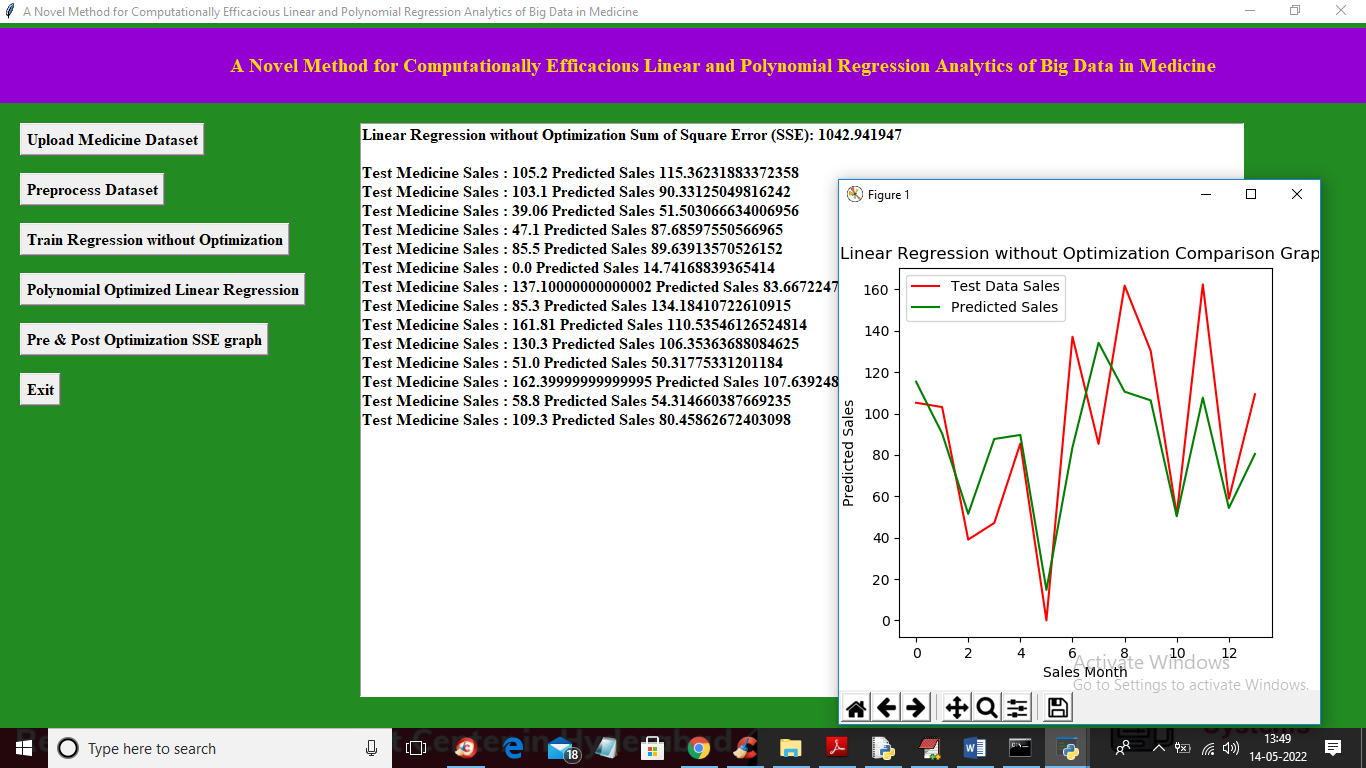
In above screen selecting and uploading medicine dataset file and then click on ‘Open’ button to load dataset and get below output



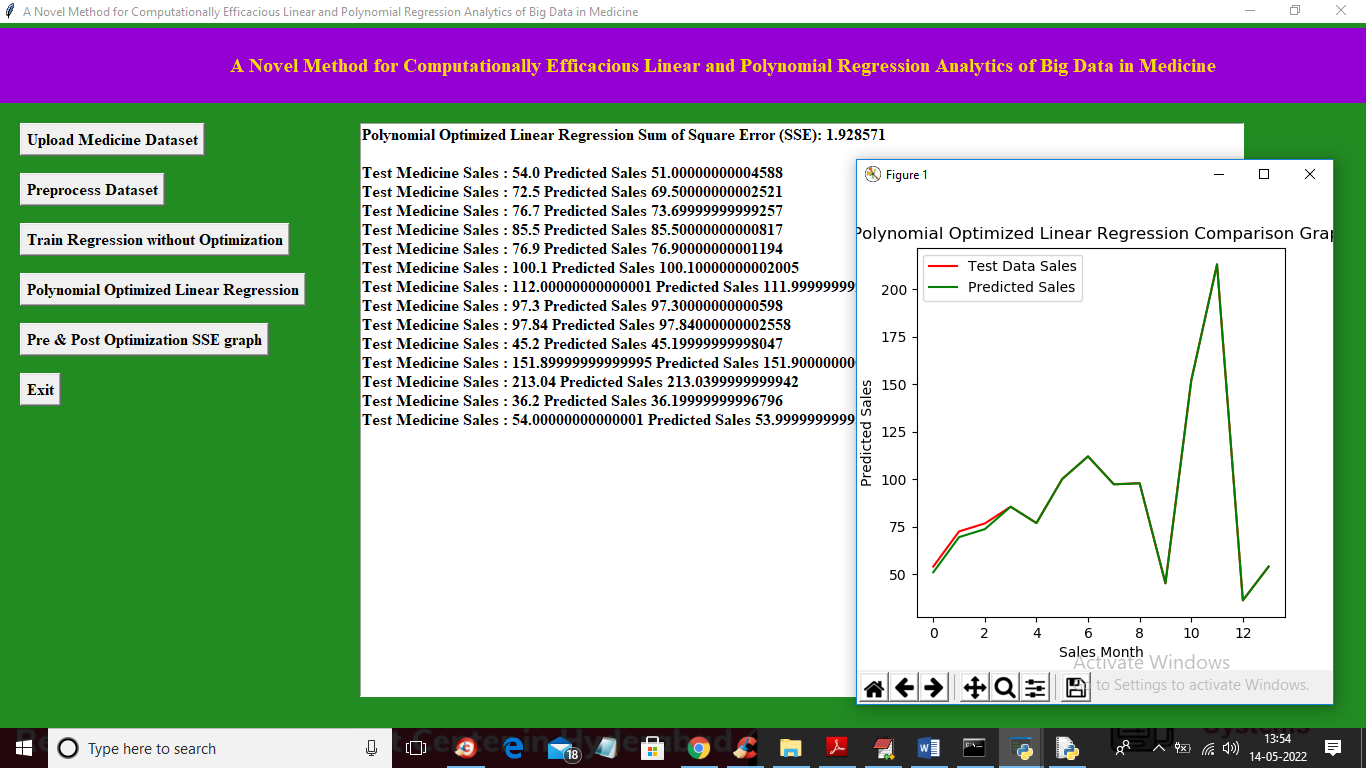
In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to split dataset into train and test



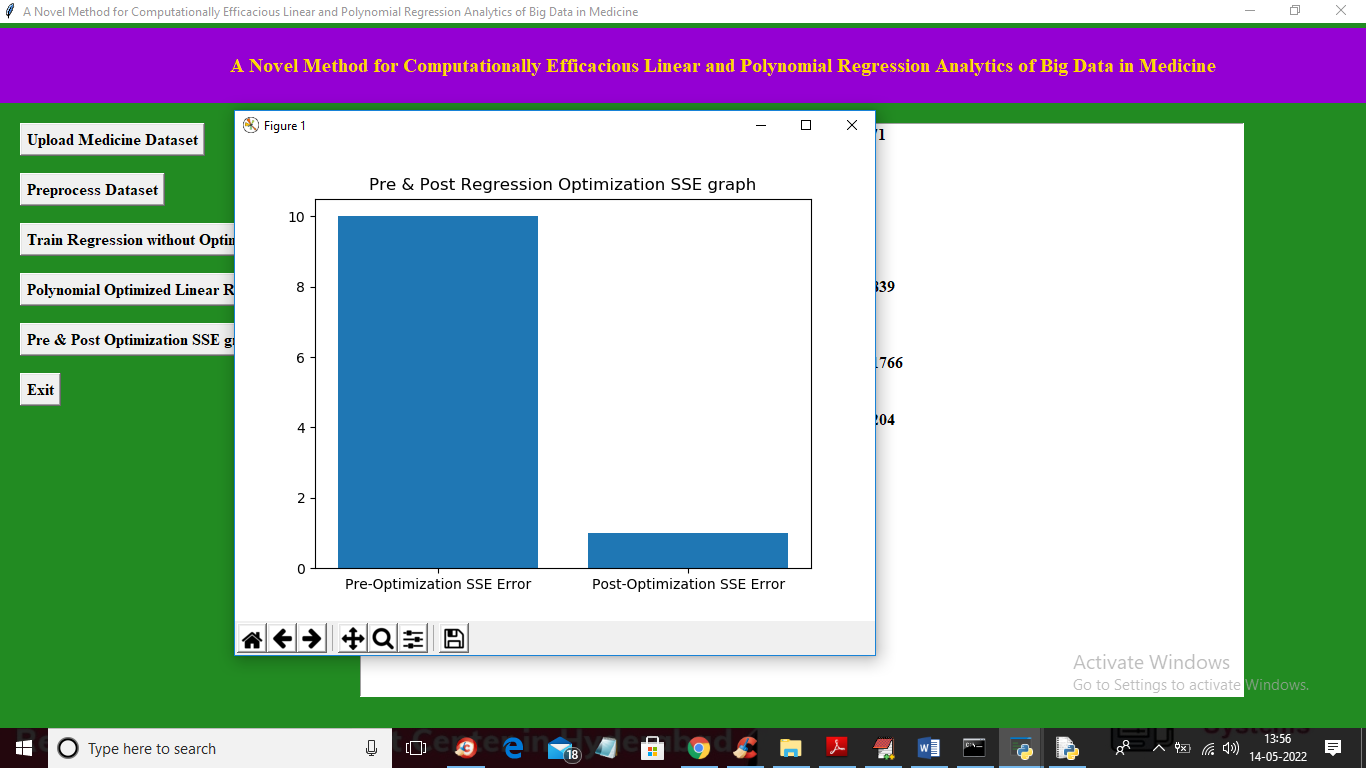
In above screen dataset contains 70 records and then application using 56 records for training and 14 for testing and now click on ‘Train Regression without Optimization’ button to train Regression on above dataset without optimization



In above screen in first line we can see without optimization Regression SSE as 1042 and then we can see the actual test medicine manufacturing value and predicted value and in above graph x-axis represents month number and y-axis represents value for required manufacturing. In above records red line refers to actual test manufacturing and green line represents predicted value and in above graph we can see there is so GAP between red and green line so its prediction is not accurate and if accurate or correct then both lines overlap and now close above graph and then click on ‘Polynomial Optimized Linear Regression’ button to train Regression with Linear and polynomial optimization and get below output



In above optimized linear polynomial Regression we got SSE error as 1.92 which is lower than without optimization algorithms and in graph we can see both lines are overlap means both actual and predicted values are same so after optimization algorithms will perform better. Now close above graph and then click on ‘Pre & Post Optimization SSE Graph’ button to get below graph



In above graph x-axis represents algorithm names as PRE and POST optimization Regression algorithm and y-axis represents SEE error and in both algorithms POST Optimization Regression is performing well