JADAVPUR UNIVERSITY

MCA PROJECT REPORT

Experimental Study of ML Techniques on Different Applications

A project report submitted in partial fulfillment of the requirements
for the degree of Master of Computer Application

In

Department of Computer Science and Engineering
Jadavpur University

By

Pijush Das

University Roll No:001910503040

University Registration No: 149900 of 2019-2020

Exam Roll No: MCA226039

Under the Guidance of

Dr. Sarmistha Neogy

Professor

Department of Computer Science and Engineering

Faculty of Engineering and Technology

Jadavpur University,Kolkata

June 2022

Declaration of Authorship

I, Pijush Das, declare that this thesis titled, “Experimental Study Of ML Techniques on Different Application” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for master degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

To whom it may concern

This is to certify that the work on this project entitled “Experimental Study Of ML Techniques on Different Application” has been satisfactorily completed by Pijush Das, Roll No: 001910503040, University Registration No 149900 of 2019-2020. It is a bona-fide piece of work carried out under my supervision at Jadavpur University, Kolkata-700032, for partial fulfilment of the requirements for the degree of Master of Computer Application from the Department of Computer Science and Engineering, Jadavpur University for the academic session of 2019-2022.

_____

Prof. Sarmistha Neogy
(Supervisor)
Department of Computer Science and Engineering
Jadavpur University

_____

Prof. Anupam Sinha
Head of The Department
Department of Computer Science & Engineering
Jadavpur University

_____

Prof. Chandan Majumdar
Dean
Faculty of Engineering & Technology
Jadavpur University

Certificate of Approval

(Only in case the project is approved)

This is to certify that the project entitled " Experimental Study Of ML Techniques on Different Application " is a bona-fide record of work carried out by Pijush Das in partial fulfilment of the requirements for the award of the degree of Master of Computer Application in the Department of Computer Science and Engineering, Jadavpur University during the period of January, 2022 to June, 2022. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

Examiners:

_____

(Signature of Examiner)

Date:

_____

(Signature of Examiner)

Date:

Abstract

Faculty of Enginnering and Technology,Jadavpur University

Department of Computer Science and Engineering

Master of Computer Application

By Pijush Das

We live in the age of data, where everything around us is connected to a data source and everything in our lives is digitally recorded. Machine Learning has grown rapidly in recent years in the context of data analysis.ML usually provides systems with the ability to learn from experience automatically. In general, the efficiency of a machine learning algorithm depends on the nature and characteristics of the data. We have worked with six machine learning (MultiLinear Regression, Polynomial Regression Decision Tree, SVR, Random Forest, K-NN) and implementing different feature selection methodologies (Correlation coefficients, Mutual Information). We have used RMSE (Root Mean Square Error) and R2_Score to analyze our model's performance. In two dataset Random Forest has given best result compared to all and Multi-Linear Regression has given the best result.

Acknowledgements

On the submission of “Experimental Study Of ML Techniques on Different Application”, I wish to express gratitude to the Department of Computer Science and Engineering for sanctioning a project work under Jadavpur University under which this work has been completed.

I would like to express my sincere gratitude to my respected guide Dr. Sarmistha Neogy ,Professor,Department of Computer Science and Engineering, Jadavpur University for her unfailing guidance, prolific encouragement, constructive suggestions and continuous involvement during each and every phase of this project. I feel deeply honoured that I got the opportunity to work under her guidance.

I would also wish to thank Prof. Anupam Sinha, Head of Department of Computer Science and Engineering, Jadavpur University, Prof. Chandan Majumdar, Dean of Faculty of Engineering and Technology, Jadavpur University for providing me all the facilities and for their support to the activities of this project.

I would like to express my gratitude and indebtedness to my parents and all my family members for their unbreakable belief, constant encouragement, moral support and guidance.

Last, but not the least, I would like to thank all my classmates of Master of Computer Application batch of 2019-2022, for their co-operation and support. Their wealth of experience has been a source of strength for me throughout the duration of my work.

Regards,

Pijush Das

University Roll Number: 001910503040
University Registration No.: 149900 of 2019-2020
Department of Computer Science and Engineering
Jadavpur University

CONTENTS

List of Figures

List of Abbreviations

SVR **S**upport **V**ector **R**egressor

K-NN **K-**Nearest **N**eighbors

RMSE **R**oot **M**ean **S**core **E**rror

MI **M**utual **I**nformation

Chapter 1

INTRODUCTION

1.1 Overview

Now-a-days we get automatic recommendations on many online shopping platforms like Flipkart,Myntra about the things you should buy next?or we use Siri,Alexa in our phones.This is application of Machine Learning.It is getting popular day by day.Machine Learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed.Machine learning focuses on developing computer programs that can access data and use it to learn for themselves.Many studies have been done to predict house price in any particular area or to predict the quality of some materials.This study works on three different datasets.In our project we shall use six models and RMSE,R2_Score as accuracy metrices.

1.2 Motivation

If Someone can rightly predict about anything which is important to him or her,that would be really helpful.With the help of machine Learning,we can acheieve this.In today's world ,many reputed companies make machine learning a central part of their operations.But the main problem is to choose the right machine learning model since every model will not give you the best result.Here we have tried to explore different dataset with six shallow machine learning models.

1.3 Objective

Three different datasets have been collected from a resource.We shall implement six shallow machine learning models on each dataset and will check which algorithm is giving the best result with fewer errors.For better result we shall implement the feature selection to determine which feature has most impact on the final output.

Chapter 2

Preliminaries

2.1 Types of Machine Learning

As mentioned earlier, AI is split into many divisions and Machine Learning methods can be categorized into three groups which all use different algorithms: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning is used when the objectives are known, hence the learning being named task driven. Therefore, human intervention is essential for supervised learning, as it founds its knowledge on preexisting answers or targets. There are two kinds of supervised ML: classification and regression. Classification is when data needs to be attributed to specific labels. The output is called nominal or discrete, as each value is distinct and separate. Regression on the opposite, is employed to predict unknown elements based on existing observations. In this case, the data is numerical and continuous.

For example, determining if a flower is part of a certain genus based on the size and number of its petals, its color and its fruit would be classification, as we assign a category to the flower. On the other hand, determining the size of a cat according to its race, age and parents' size would be regression, as the answer follows a continuous line established by the training data. Figure 1 represents how each model relates to the data.

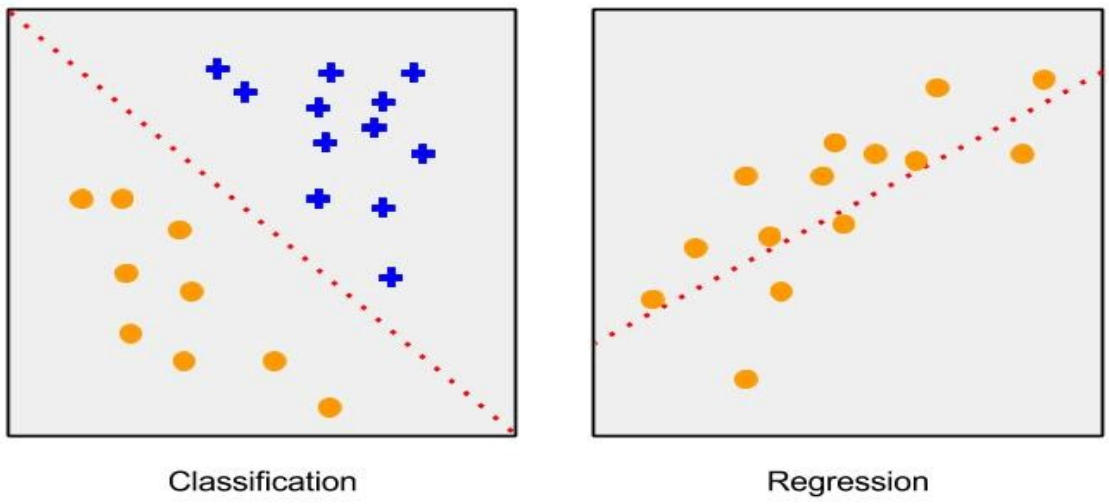

Figure 2.1:Classification v/s Regression

On the other hand, **Unsupervised learning** draws on algorithms that do not need defined goals. Instead, they figure out patterns on their own using either clusters or associations. This type of learning is called data driven. This approach can lead to discoveries of hidden patterns that would otherwise be unidentifiable, for example in complex and multidimensional data models. It can also pick up anomalies and faults like a data point seemingly too different from the rest, or a value that seems abnormal.

Clustering classifies the data into groups, similarly to classification ML, but this time, classes are formed by the algorithm itself instead of being explicitly given. Comparatively, associations work by analyzing possible dependencies between items in a data set.

Finally, **Reinforcement learning** is a type of ML that focuses on training AI to perform complicated tasks by using a reward system. Basically, an agent is put into an environment and is asked to perform a specific task or goal for which it will be rewarded. Starting with no knowledge, it will first effectuate random actions that may or may not be rewarding. However, through trial and error, it will slowly optimize itself, learning from each previous iteration.

2.2 Methods

- **Linear regression** is one of the most basic prediction models. It is based on the principle of independent and dependent variables, meaning that there is a clear relationship between the response and the features. Using training data as a basis, a linear equation is created to fit the observations as closely as possible by applying the least squares method, where the aim is to minimize the sum of the errors.
  The function for Linear Regression is:
  y=$a + b.x$
  Where a is intercept and b is co-efficient of x.
  Y is independent variable and x is dependent variable.

- **Multiple Linear Regression** is extension of Linear Regression when there are two or more input variables.

- **Polynomial Regression** is method where we transform original features into polynomial features of a given degree.If we increase the degree to a very high value,the curve becomes overfiitted.
  The function for polynomial regression is

y=b_0+b_1x_1 + $b_2{x_2}^2$+…….+$b_n{x_n}^n$

y= output variable .

x_i =is ith input variable.

b_i=Co-efficient of x_i .

- **Support-vector Regression(SVR)** is a complex regression method for supervised learning which uses thresholds called margins to separate data that belongs in the same group. Support Vector Regression tries to plot a hyperplane which fits the nth-dimensional data space more accurately.
  Some terms releated with SVR are:
  1. Kernel: The function for converting a lower-dimensional data set to a higher-dimensional data set.
  2. Hyper Plane:In SVR,this is a line that will assist us in predicting a continuous value.
  3. Boundary Line: A boundary line can be thought as boundary for the tube,whereas positive value in one side and negative value on other side.
  4. Support Vectors: Support vectors are locations that are outside the -tube in SVR.
- **Decision tree is** a model that functions using series of binary choices to categorize an input based on their path. Decision tree regressor uses 'Standard Deviation Reduction' to construct the tree.

  Working of Decision Tree Algorithm:

  1. Partition the data into subsets.
  2. Shorten the branches of decision tree.
  3. Find the smallest tree that fits the data.

- **Random Forest** is supervised learning algorithm which uses ensemble learning methodology.As the name suggests forest is a collection of trees,Random forest build multiple decision trees and merge the ouput together to get a accurate result.It is based on Bagging Method. For Fig 2.2, 600 datasets are made with possibility of one datapoint can be present in multiple datasets. Random Forest use multiple Decision Tree model to predict its final output. In Fig 2.2, we see 600 of Decision Trees are used to build Random Forest model.At the last it is taking the average of all predictions.

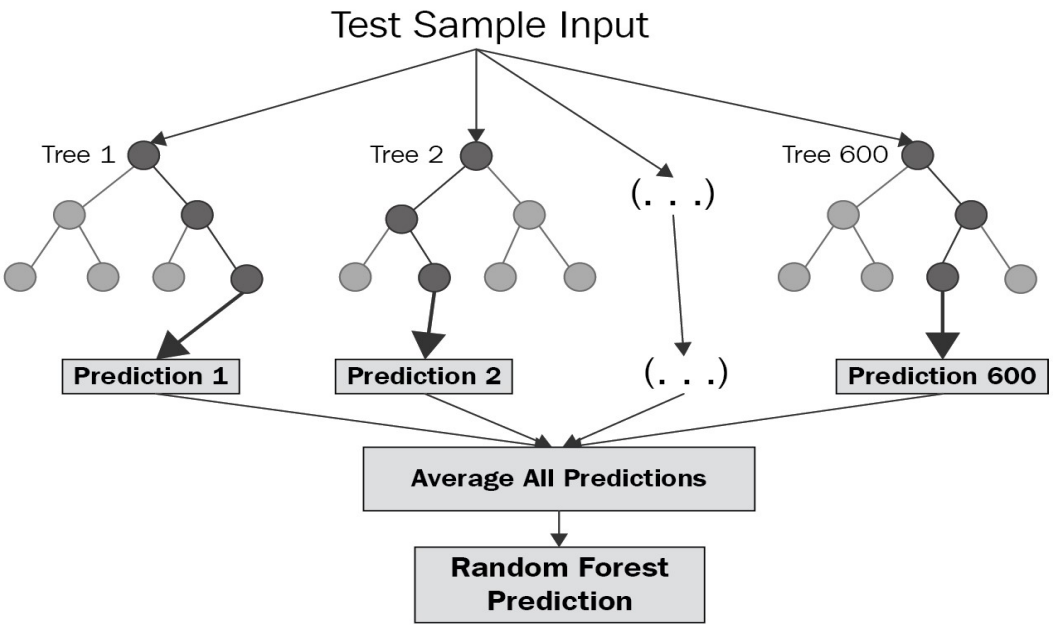

Figure 2.2:Example of Random Forest

Working of Random Forest Algorithm:

1. Select k random data points from training set.
2. Build the decision tree with those k data points.
3. Choose the number of decision tree.
4. Repeat step 1 and step 2.
5. For new data points,calculate prediction of each decession tree,assign them to that tree which has best result.

- **K-Nearest Neighbor** is a simple but powerful algorithm.KNN models assign a category to a new input by comparing it to the categories of the nearest data points on a graph. The number of data points used for the comparison, or nearest neighbors, is determined manually by the K parameter.

  Algorithm:

  Step-1: Select the number K of the neighbors.
  Step-2: Calculate the Euclidean distance of K number of neighbors.
  Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
  Step-4: Among these k neighbors, count the number of the data points in each step.

Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
Step-6: The model is ready.

Finding the best value for K is invaluable but complicated, as there is no default setting that is suitable for all circumstances. Furthermore, low K values tend to be subject to outliers and noise, while high K values can completely overlook categories with low population. It is also important to avoid ties, as such results often end up being unclassified, so K values tend to be odd.

Chapter 3

Present Approach

3.1 Problem Statement

Experimental study of six ML Techniques(Multi-Linear Regression,Polynomial Regression,Decision Tree ,Random Forest,SVR,KNN) on three different datasets. The objective is to see by looking at the characteristics of dataset,We can say that which algorithm will be most appropriate for that dataset.

3.2 Data Collection

I have worked on three datasets.These have been collected from Kaggle Website.Name of the three datasets are:

1. 'Boston.csv'
2. 'Red.csv'
3. 'Admit.csv'

Information on Data:

- 'Boston.csv':

  Description: This dataset contains 14 columns where MEDV is output variable.

Input features in order:
1) CRIM: per capita crime rate by town
2) ZN: proportion of residential land zoned for lots over 25,000 sq.ft.
3) INDUS: proportion of non-retail business acres per town
4) CHAS: Charles River dummy variable (1 if tract bounds river; 0 otherwise)
5) NOX: nitric oxides concentration (parts per 10 million) [parts/10M]
6) RM: average number of rooms per dwelling
7) AGE: proportion of owner-occupied units built prior to 1940
8) DIS: weighted distances to five Boston employment centres
9) RAD: index of accessibility to radial highways
10) TAX: full-value property-tax rate per $10,000 [$/10k]
11) PTRATIO: pupil-teacher ratio by town
12) B: The result of the equation B=1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
13) LSTAT: % lower status of the population

Output variable:
1) MEDV: Median value of owner-occupied homes in $1000's [k$]

- 'Red.csv':
  Description: This dataset contains total 12 columns where quality is the output variable.
  Input variables in order:

  1.Fixed acidity
  2.Volatile acidity
  3.Citric acid
  4.Residual sugar
  5.Chlorides
  6.Free sulfur dioxide
  7.Total sulfur dioxide

  8.Density
  9.PH
  10.Sulphates
  11.Alcohol

  Output variable (based on sensory data):
  12.Quality (score between 0 and 10)

- 'Admit.csv':
  Description: This dataset contains total 9 columns where 'Chance of Admit' is the output variable.
  Input variables in order:
  1) Serial no.
  2) GRE Scores (out of 500)
  3) TOEFL Scores (out of 120)
  4) University Rating (out of 5)
  5) Statement of Purpose (out of 5)
  6) Letter of Recommendation Strength (out of 5)
  7) Undergraduate GPA (out of 10)
  8) Research Experience (either 0 or 1)

  Output variable
  9.Chance of Admit (ranging from 0 to 1).

3.3 **Workflow**

First we have collected the three different types of data.These datasets doesn't have any missing/garbage/NULL value.

The next step is to split the dataset into the Training set and Test set.The test data is 20% of the whole dataset.Each algorithm has different parameters.Applying the default parameter for any model will not give us the optimal solution.So to get the set of parameter for any model to maximize model's performance,we need to implement Hyperparameter Tunning.GridSearchCv approach has been applied in our K-NN algorithm.We have manually checked also which parameters give the best result with fewer errors and hence applied that set of parameter.Then we have to train the ML algorithms on training dataset using that we can predict test data set's result.As this is a regression model, We have calculated the 'R2_Score' and (Root mean square root)'RMSE' error at the last.

Each feature in dataset plays a vital role in maximizing model's performance as it has higher or lower relation with the final output.Selecting the most important feature or dropping any irrelevant feature is very important. The main objective of features selection are to train the model faster,improves the accuracy of the model,increase model interpret-ability ,simplifies the model and reduces over-fitting. We have used two Filter methods for Feature Selection:- Information gain or mutual information,Correlation coefficients.Then after applying ML Algorithms,have calculated the 'R2_Score' and (Root mean square root)'RMSE' error again.

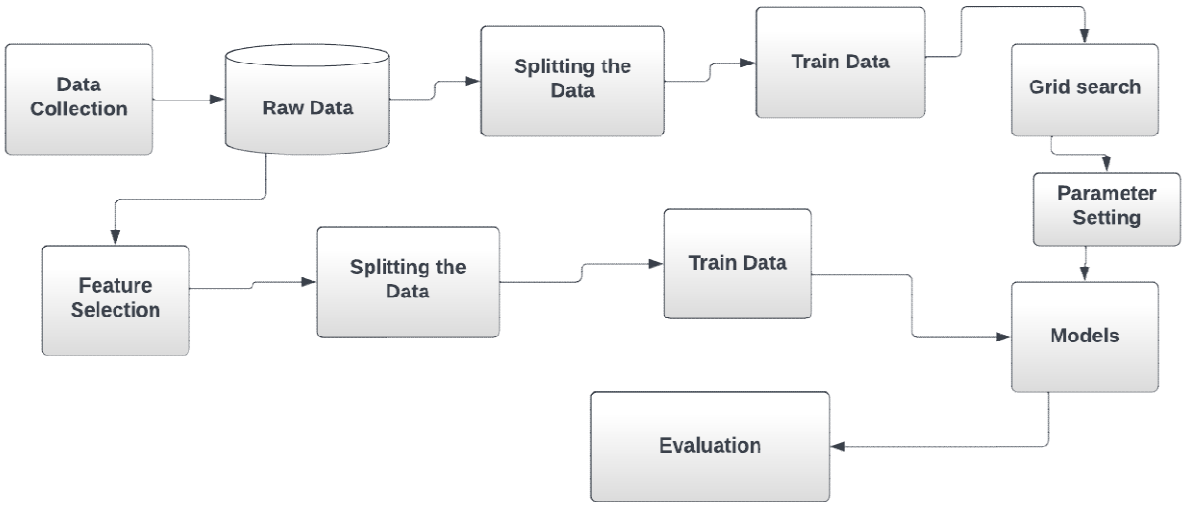

Fig 3.1:FlowChart of the current work

3.4 Proposed Approach for Feature Selection:

Mutual Information:MI[2] assess the dependency of the independent variable in predicting the target variable. In other words, it determines the ability of the independent features to predict the target variable.

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} = E_{P_{XY}} \log \frac{P_{XY}}{P_X P_Y} .$$

$I(X;Y)$=Mutual information between X and Y. X,Y= Discrete variable.

P_{XY}(x,y)=Joint probablity Distribution.

Correlation coefficients: It removes duplicate features and involves the concept of heatmap.It is a two-dimensional tabular representation of data with a range of values represented by different colors which provides a visual summary of information and helps to understand complex data sets.The correlation value ranges from -1 to +1. A correlation closer to 1 means the variables are more correlated.Similarly for the value closer to -1,it means that they are not correlated.The diagonal values are always 1 because any variable is correlated with itself.

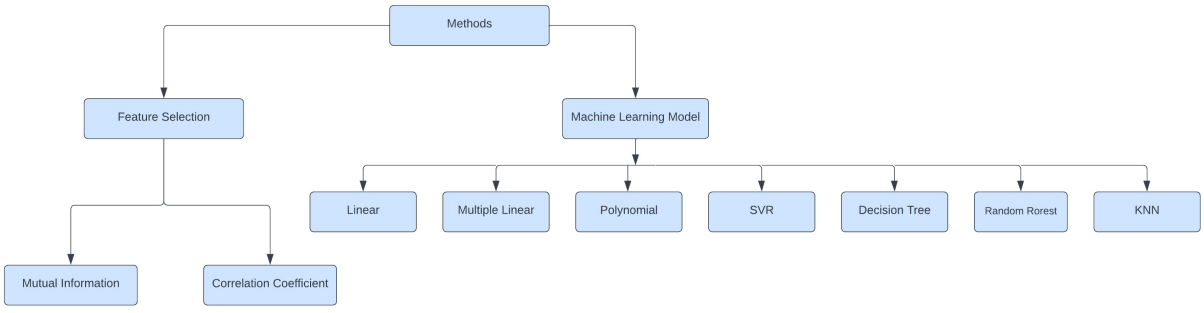

Fig 3.2: A taxonomy of the approaches evaluated in this study for feature selection and prediction models.

Chapter 4

Result and Analysis

4.1 Experimental Setup

This study is done on a windows Laptop with Intel(R) Core (TM) i3-7020U CPU@ 2.30GHz with 8GB RAM (DDR4). “Experimental Study Of ML Techniques on Different Application” has been performed and analysed using Google Colab (https://colab.research.google.com/?utm_source=scs-index). We worked on this experiment in python language. Various python libraries were also used e.g., NumPy,Pandas, Matplotlib, scikit learn. NumPy (NumPy) is used in working with array. Pandas is used for analysis of data. Matplotlib is used for visualization of data. Scikit learnis used for implementing models and one metrics RMSE to judge the error of the output.

4.2 Results

First we are implementing the Machine Learning Algorithms for 3 datasets and providing the result.After each dataset we can conclude which Machine Learning Algorithm provides best result for that particular dataset.

Evaluation metrics are important to determine the accuracy of the models. We have used Root Mean Square Error (RMSE) and also evaluated our model with R2_Score.

$$\text{RMSE} = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$

$$R^2 = 1 - \frac{RS}{TSS}$$

$$\text{RSS} = \sum_{i=0}^{n} (y_{test} - y_{pred})$$

$$\text{TSS} = \sum_{i=0}^{n} (y_{test} - y_{mean})$$

A brief discussion about the result has been provided in this section.As discussed in chapter 3,hyper tunning is done with each dataset.Train size is 80% and test size is 20%.GridSearchCv approach is used for hyper tuning for K-NN Model.For other methods we have manually done hyper tuning manually.For each dataset ,we are mentioning the result separately.

Dataset:'Red.csv'

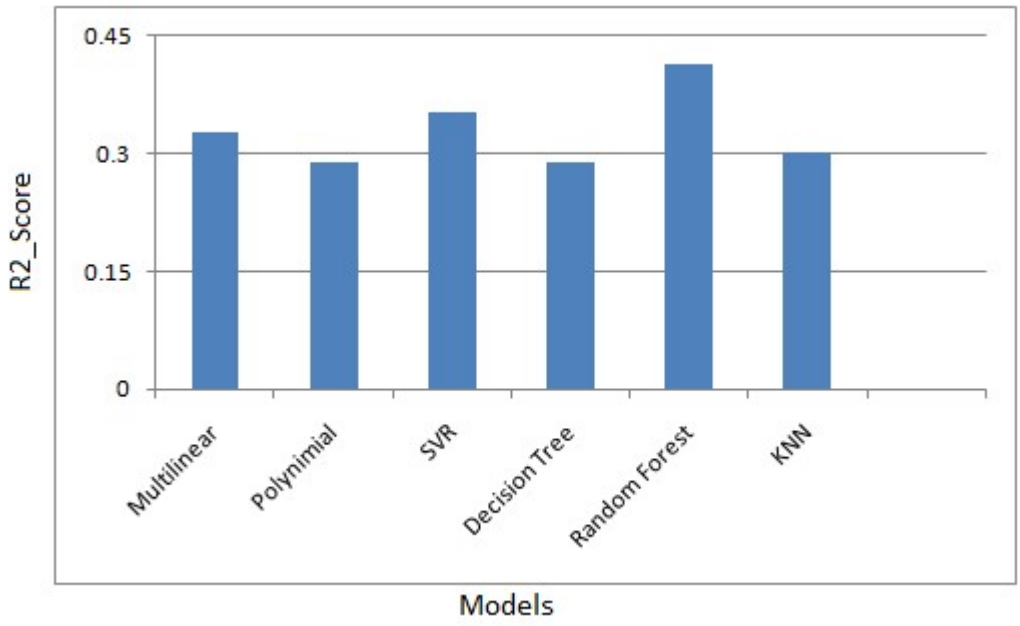

Figure 4.1:Comparison of R2_Score for different models

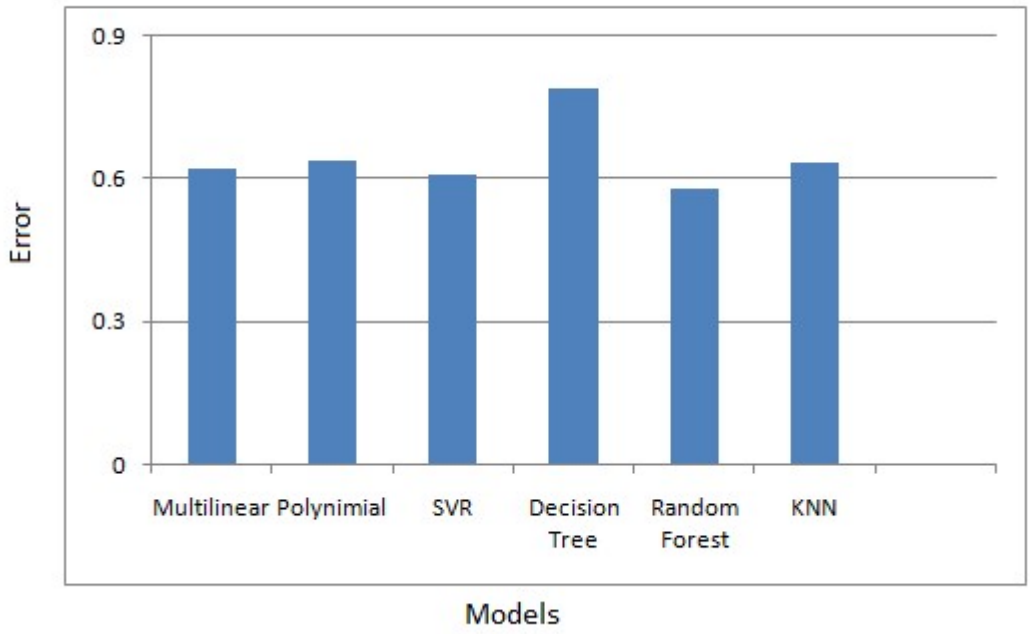

Figure 4.2: Comparison of RMSE for different models

RMSE and R2_Score is considered to compare the machine learning models.Fig 4.1 and Fig 4.2 indicates that six machine learning model's RMSE and R2_Score for the dataset 'red.csv'. we an clearly see that 'Random forest' has the highest R2_Score and lowest RMSE value.

Dataset:'Admit.csv'

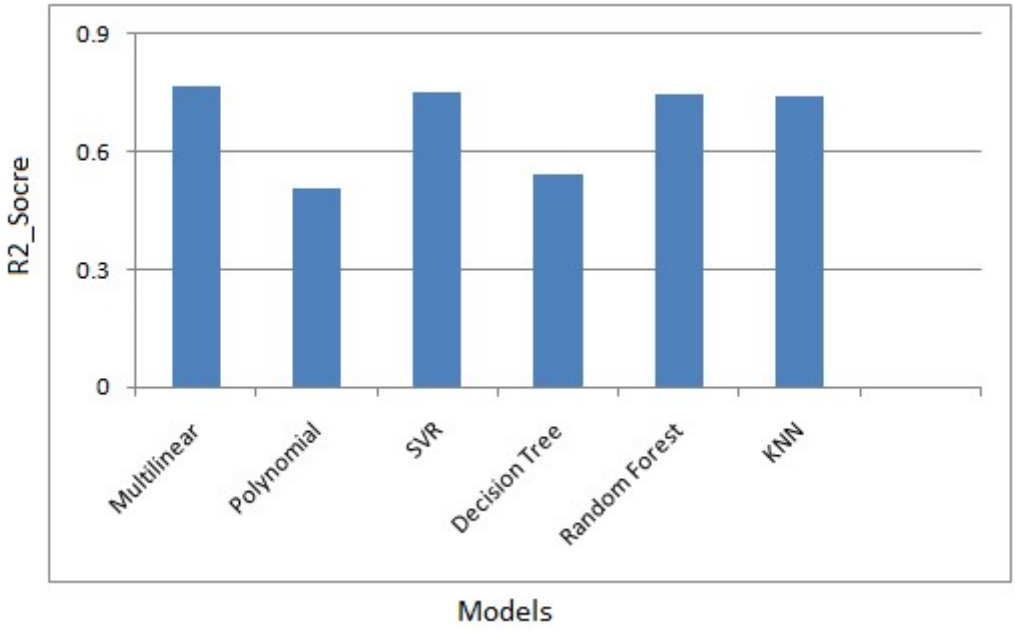

Figure 4.3: Comparison of R2_Score for different models

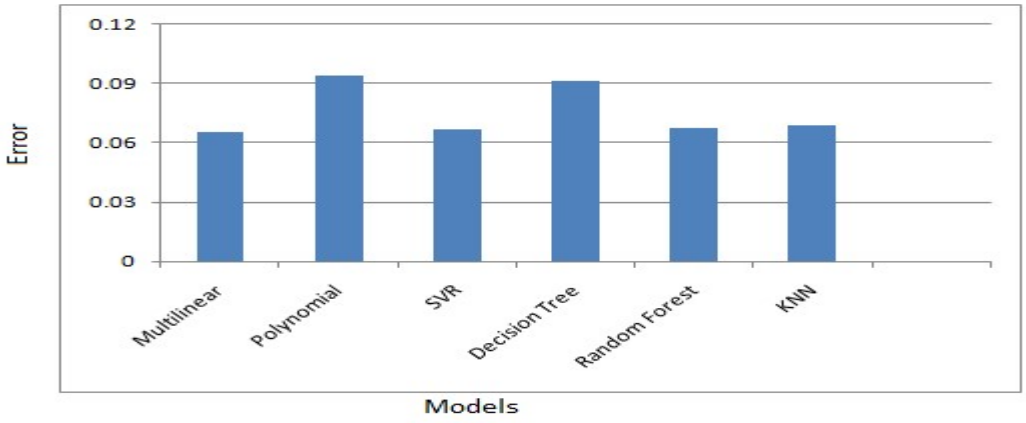

Figure 4.4: Comparison of RMSE for different models

Fig 4.3 and Fig 4.4 indicates that six machine learning model's RMSE and R2_Score for the dataset 'Admit.csv'. we an clearly see that 'Multi-Linear Regression ' has the highest R2_Score and lowest RMSE value. Interesting to see that 'SVR' and 'Random Forest' has similar type of values and it is closer to the best value.

Dataset:'boston.csv'

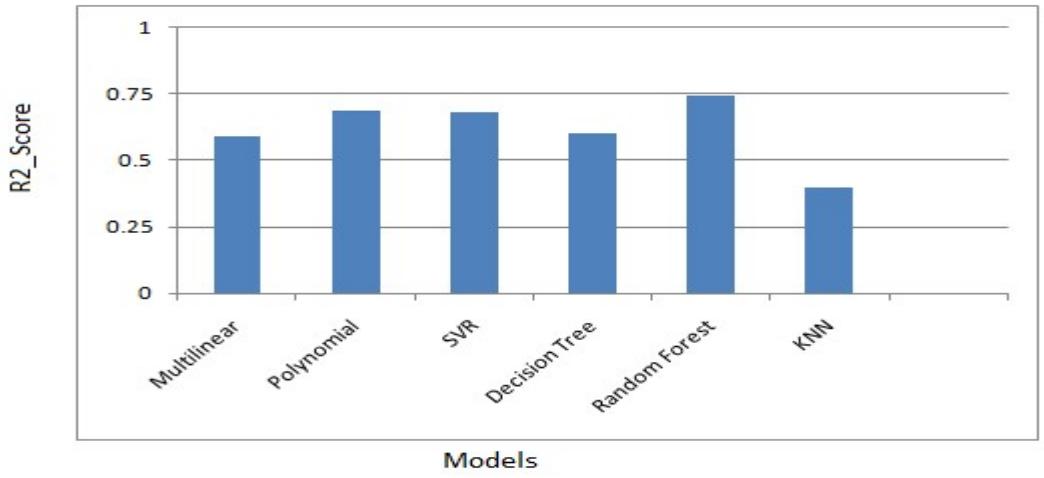

Figure 4.5: Comparison of R2_Score for different models

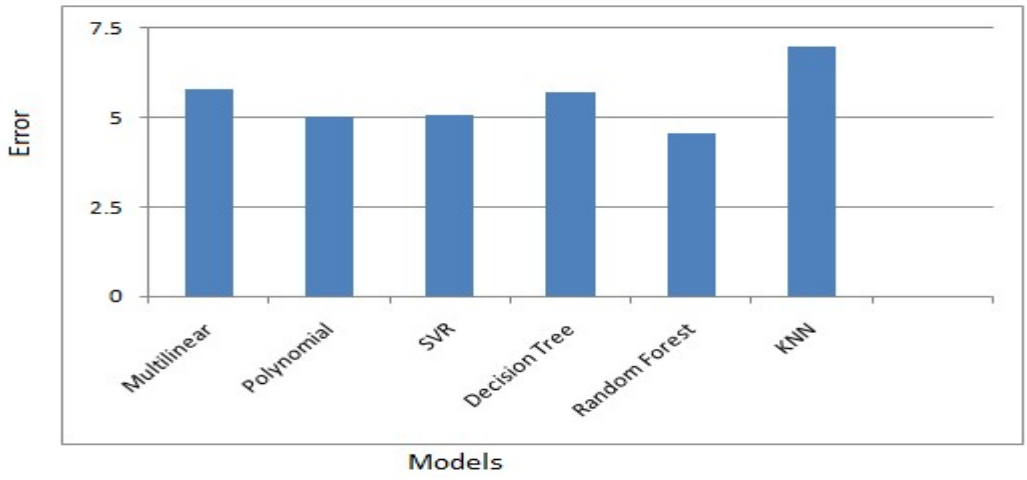

Figure 4.6: Comparison of RMSE for different models

Fig 4.5 and Fig 4.6 indicates that six machine learning model's RMSE and R2_Score for the dataset 'boston.csv'. we an clearly see that 'Random Forest ' has the highest R2_Score and lowest RMSE value. Interesting to see that 'SVR' and 'Polynomial Regression ' has similar type of values.

Feature Selection:

Dataset:'red.csv'

We have implemented feature selection and have got a new dataset 'rednew.csv' which has selected the best 9 features.

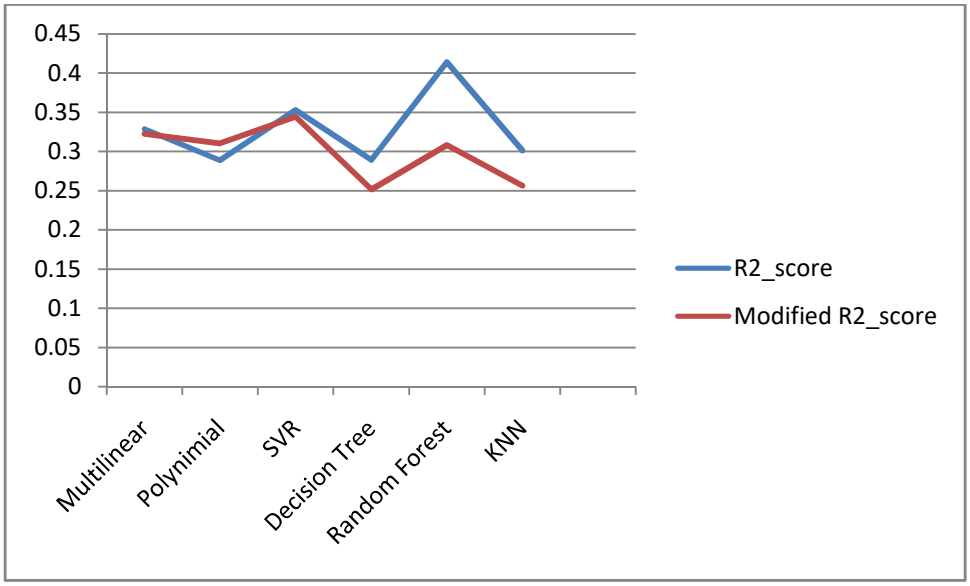

Figure 4.7: Comparison of R2_Score and Modified R2_Score for different models

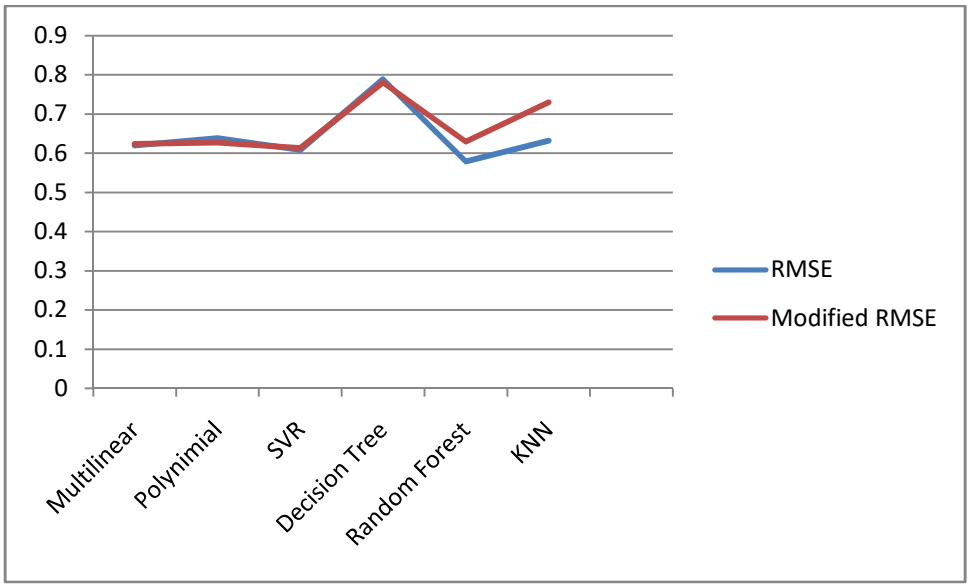

Figure 4.8: Comparison of RMSE and Modified RMSE for different models

In Fig 4.7 and 4.8 modified RMSE means the value we are getting after implementing feature selection.We can see that for this dataset in 'Decision Tree','Polynomial Regression' we are getting better values.

Dataset:'Admit.csv'

We have implemented feature selection and have got a new dataset 'Admitnew.csv' which has selected the best 6 features.

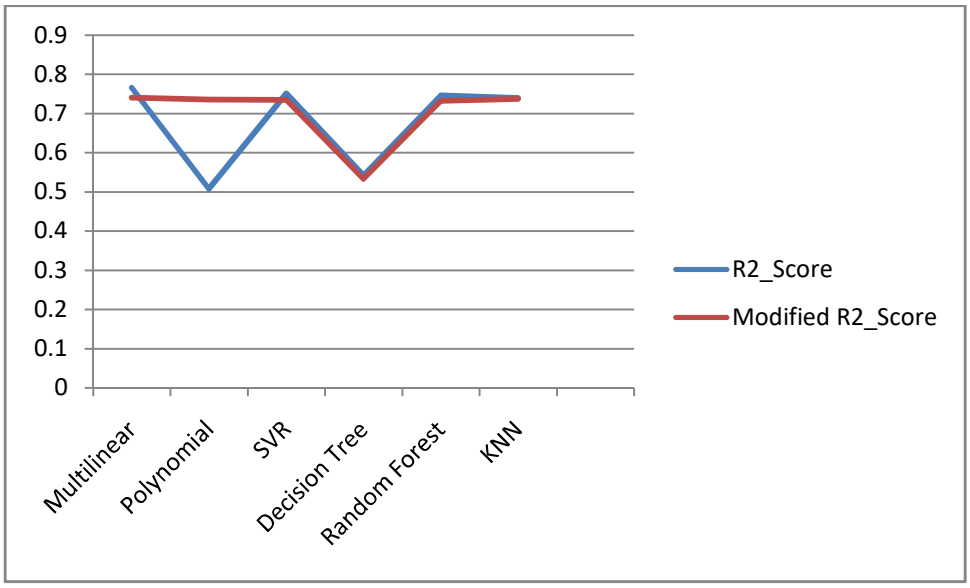

Figure 4.9: Comparison of R2_Score and Modified R2_Scorefor different models

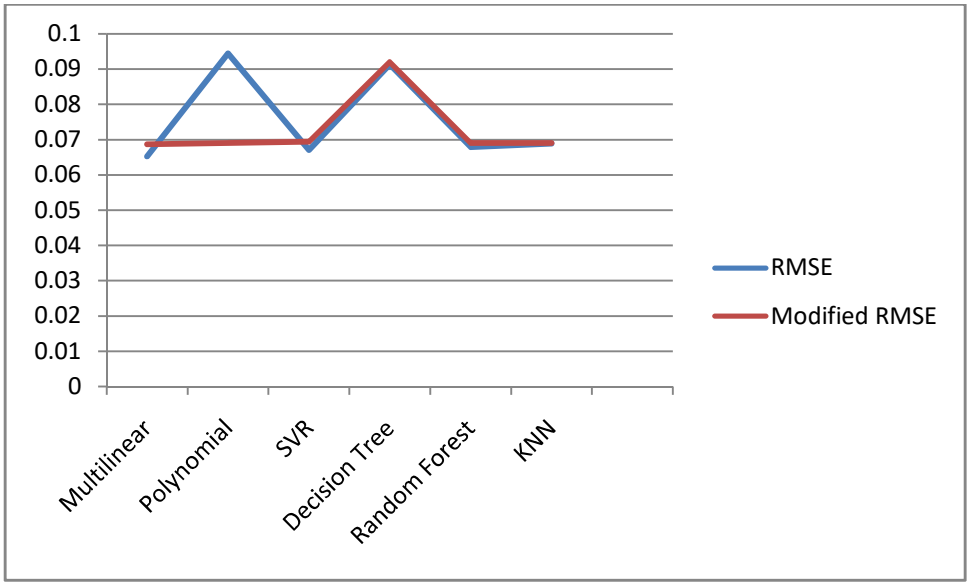

Figure 4.10: Comparison of RMSE and Modified RMSE for different models

In Fig 4.9 and 4.10 modified RMSE means the value we are getting after implementing feature selection.We can see that for this dataset in only 'Polynomial Regression' we are getting better values.

Dataset:'boston.csv'

We have implemented feature selection and have got a new dataset 'bostonnew.csv' which has selected the best 11 features.

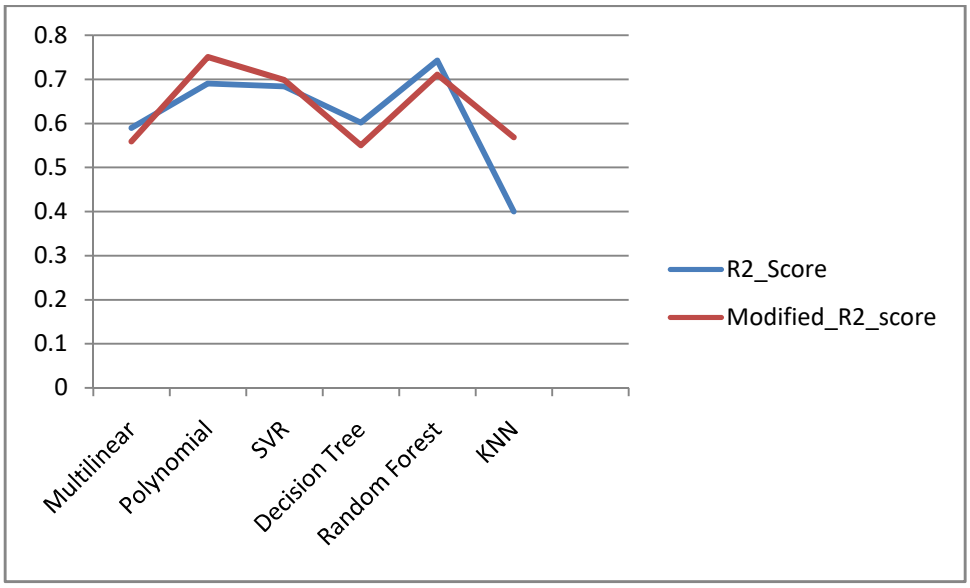

Figure 4.11: Comparison of R2_Score and Modified R2_Score for different models

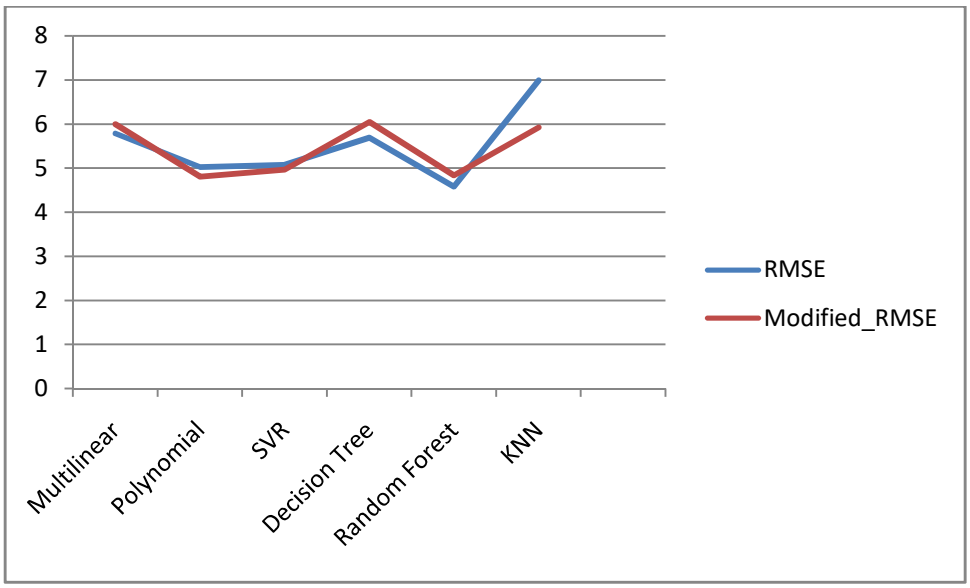

Figure 4.12: Comparison of RMSE and Modified RMSE for different models

In Fig 4.11 and 4.12 modified RMSE means the value we are getting after implementing feature selection.We can see that for this dataset in 'Polynomial Regression','SVR' and 'K-NN' we are getting better values.

Heatmap:

Dataset:'red.new'

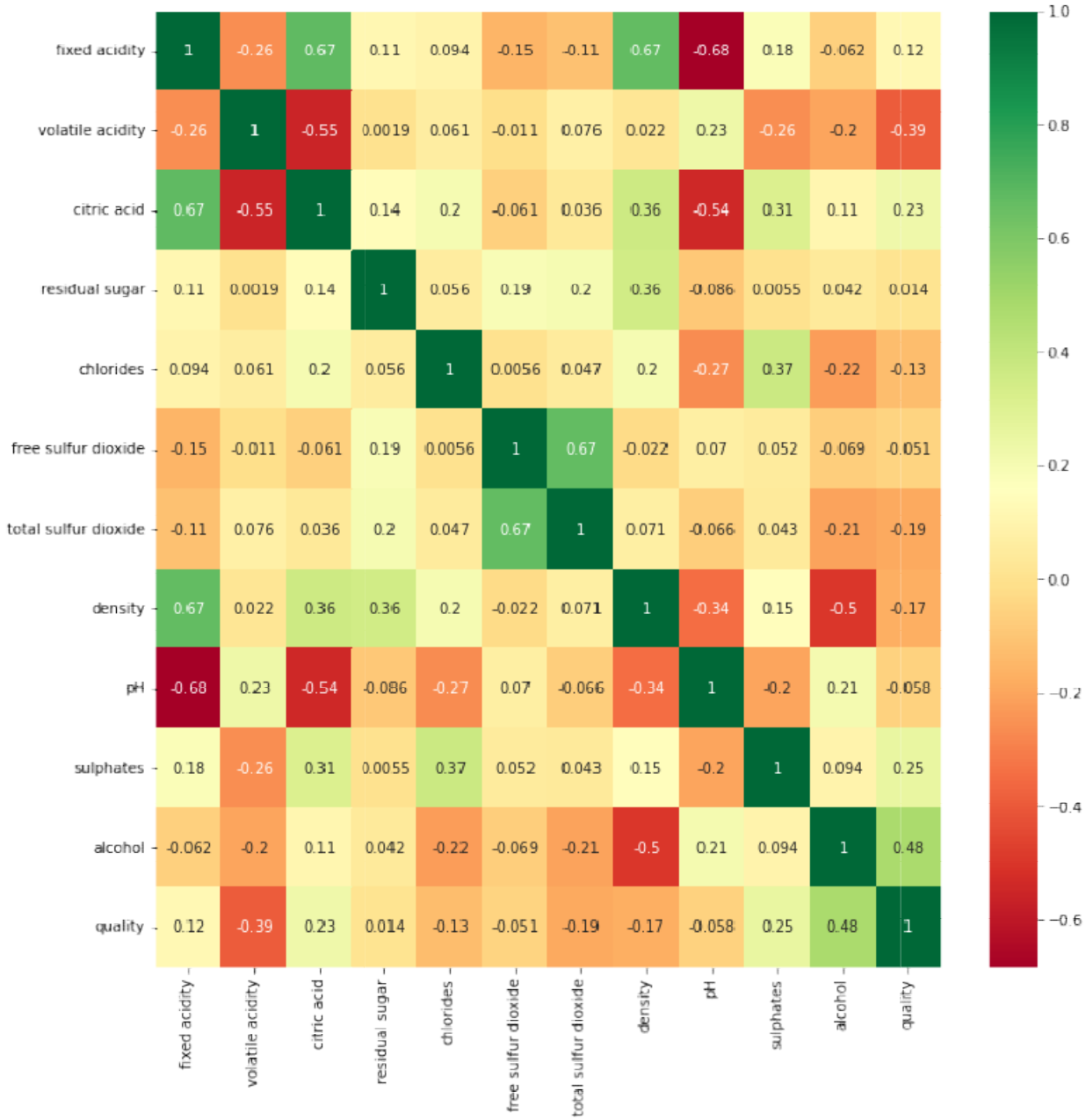

Figure 4.13 : Heatmap Representation of dataset 'red.csv'

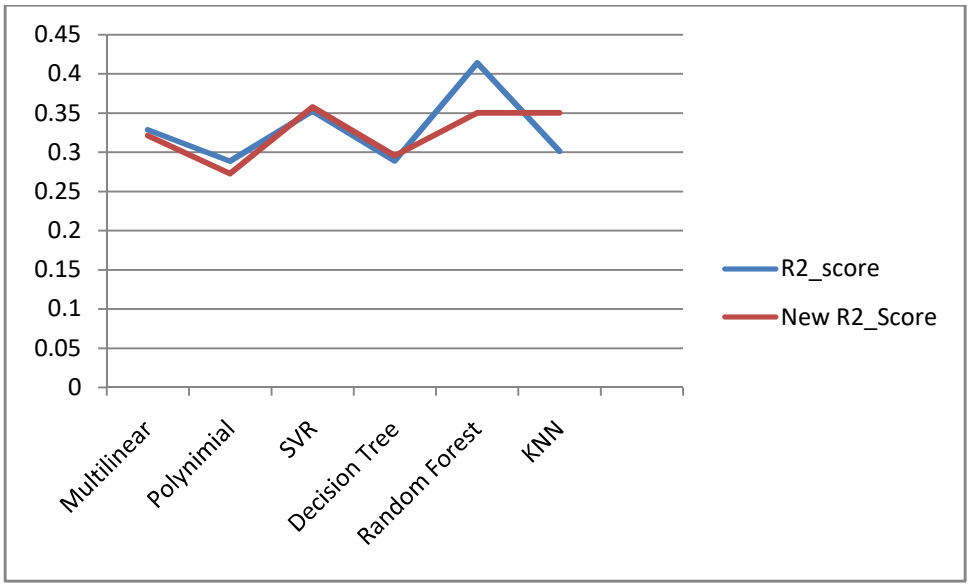

Figure 4.14: Comparison of R2_Score and New R2_Score for different models

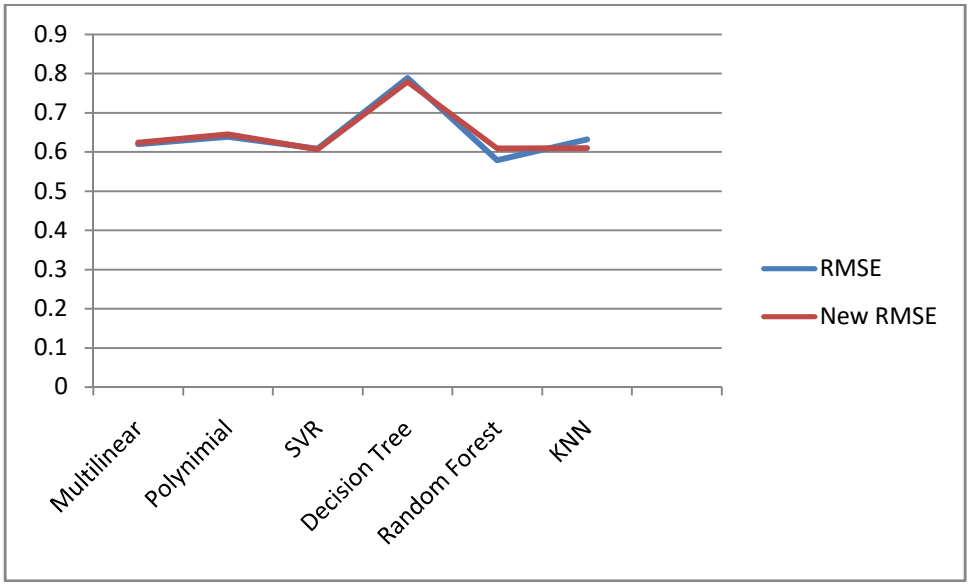

Figure 4.15: Comparison of RMSE and New RMSE for different models

New RMSE means the RMSE we are getting after dropping some columns which are not correlated with the final output.In Fig 4.14 and 4.15 we can see the by dropping the column "volatile acidity",we have got the better result in terms of fewer error in 'K-NN' and 'SVR' ,except the other results are overall same.

Dataset:'Admit.csv'

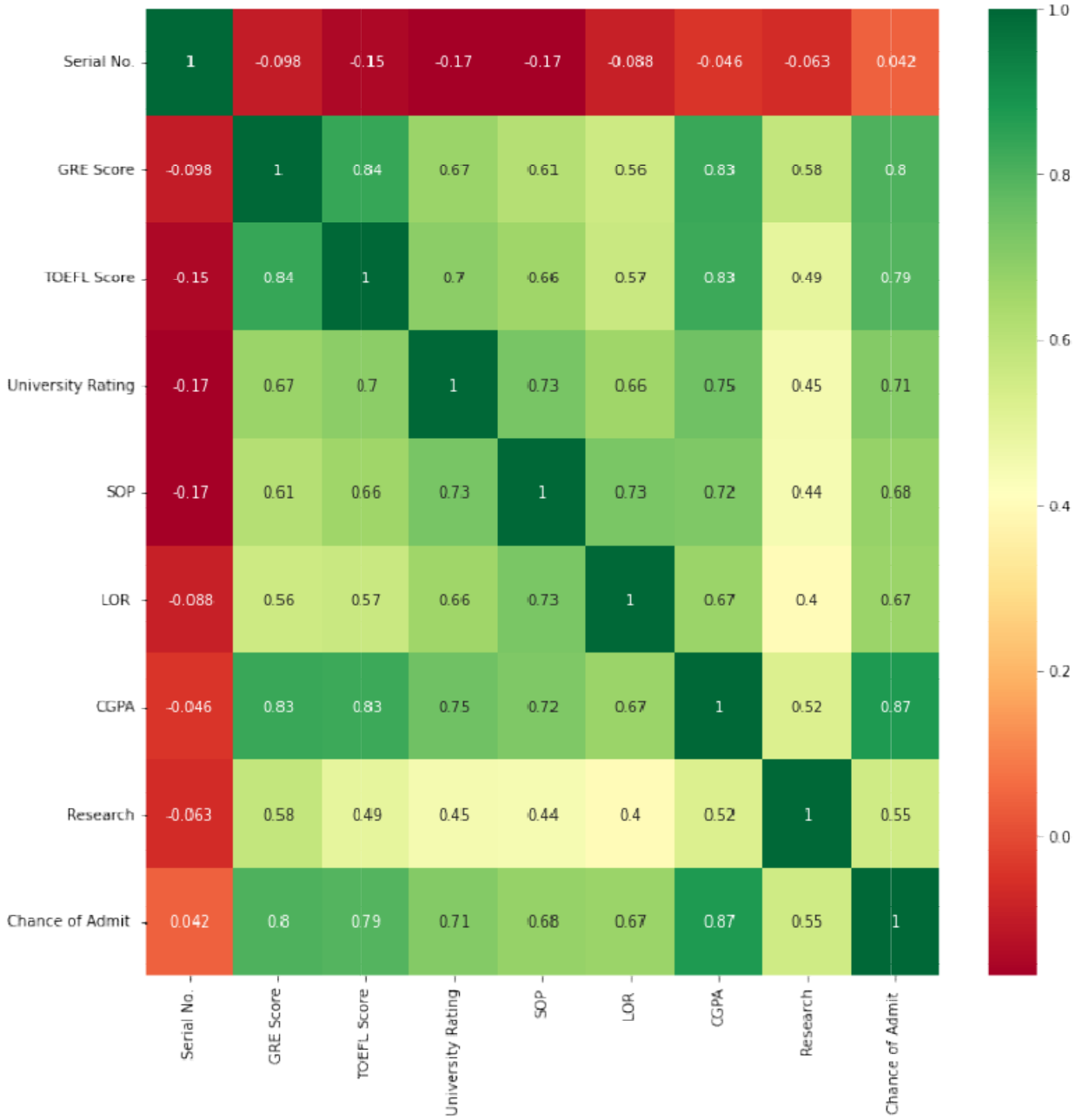

Figure 4.16: Heatmap Representation of dataset 'Admit.csv'

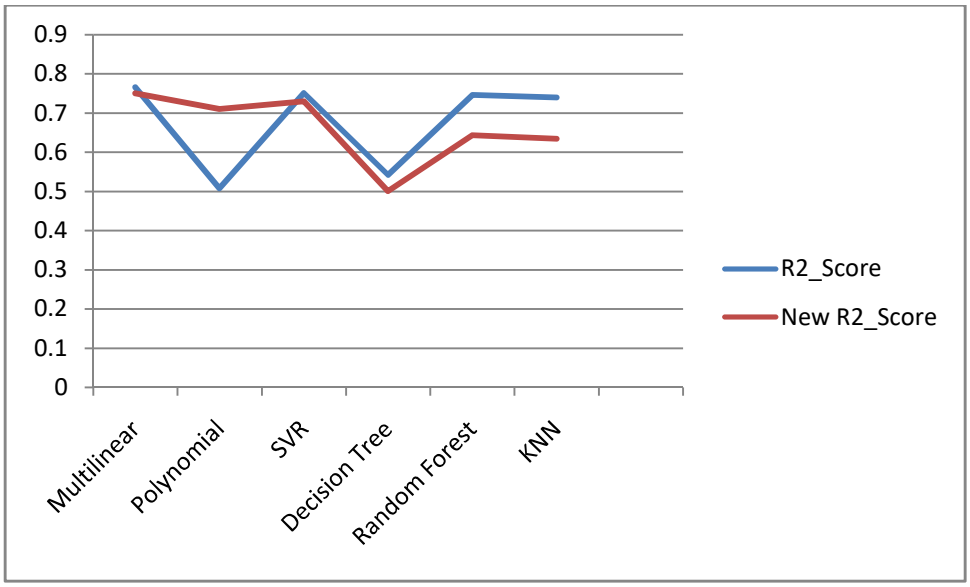

Fig 4.17:Comparison of R2_Score and New R2_Score for different models

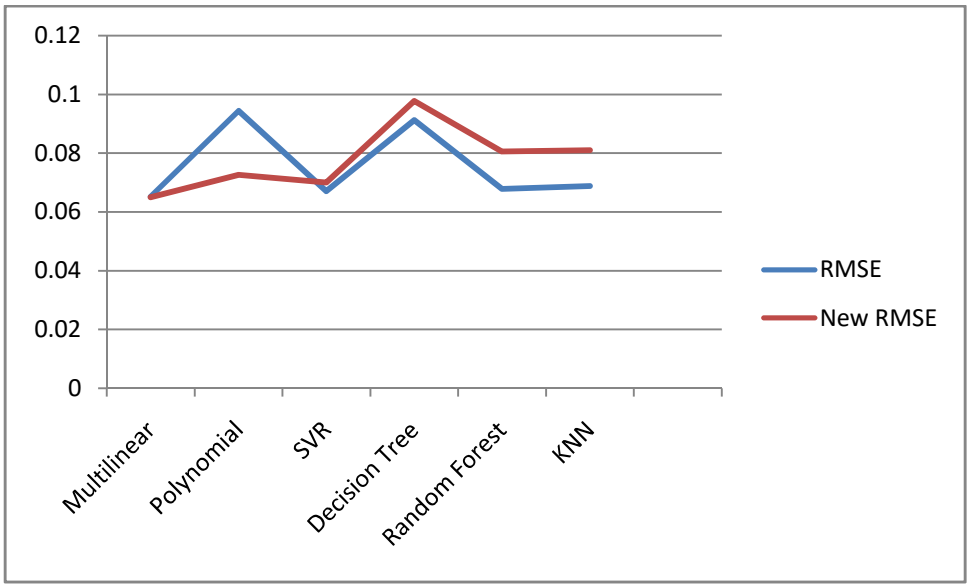

Figure 4.18: Comparison of RMSE and New RMSE for different models

New RMSE means the RMSE we are getting after dropping some columns which are not correlated with the final output.In Fig 4.17 and 4.18 we can see the by dropping the column "Serial No.",we have got the better result in terms of fewer error in 'Polynomial Regression' .Here all the other features were highly co-related with the output data.so as expected we have not got any better solution.

Dataset:'boston.csv'

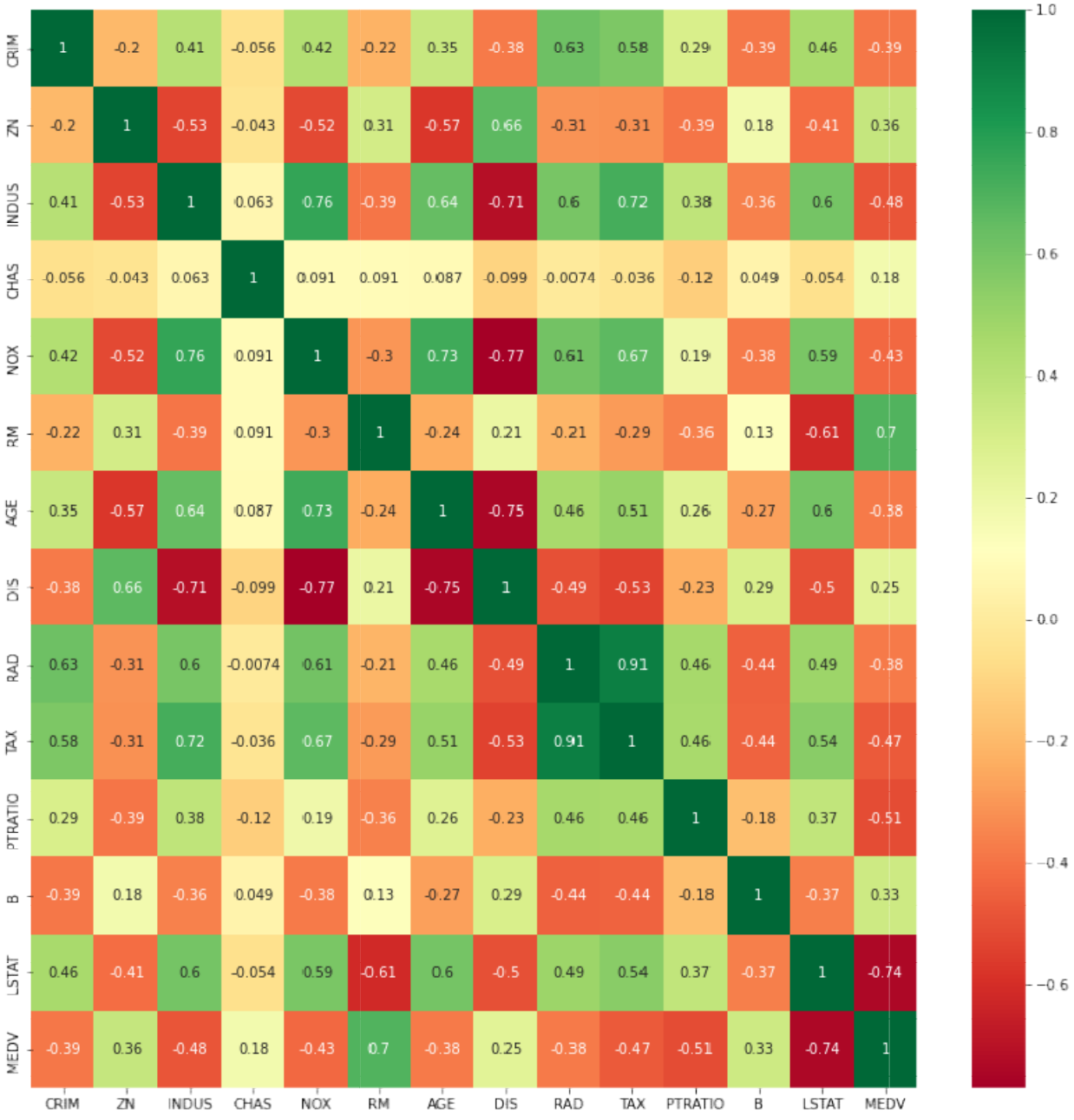

Figure 4.19 : Heatmap Representation of dataset 'boston.csv'

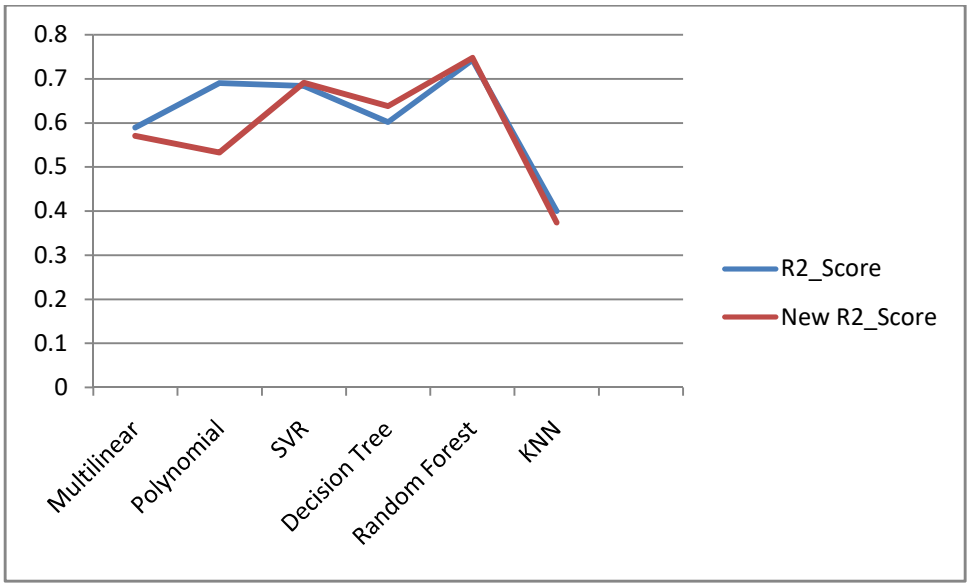

Figure 4.20: Comparison of R2_Score and New R2_Score for different models

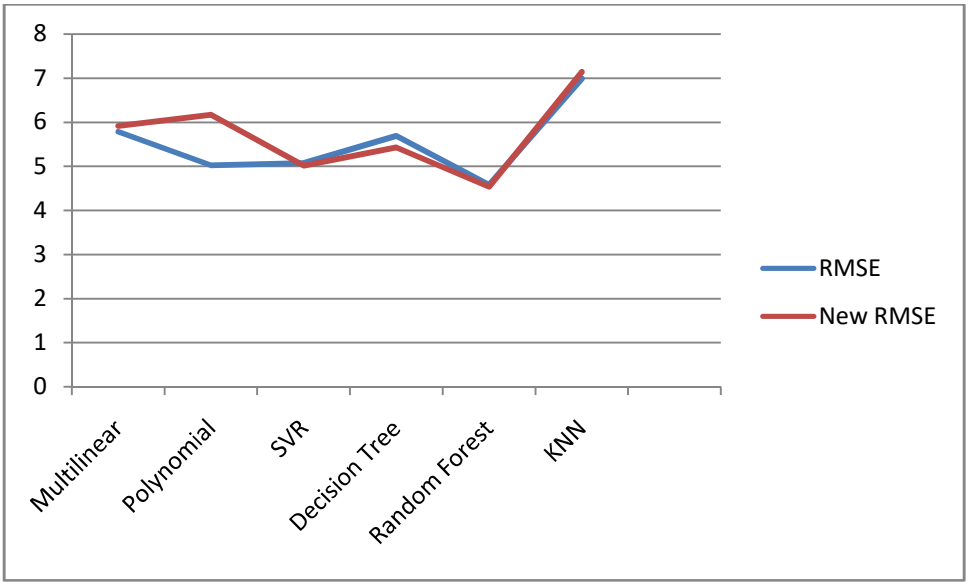

Figure 4.21: Comparison of RMSE and New RMSE for different models

New RMSE means the RMSE we are getting after dropping some columns which are not correlated with the final output.In Fig 4.20 and 4.21 we can see the by dropping the columns "PTRATIO","TAX","CRIM",we have got the slightly better result in terms of fewer error in 'Random Forest' and 'SVR' .In 'Decision Tree' algorithm we have got better result,in other algorithms it is not getting better.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

We have worked on six machine learning models in this study and applied them on three different datasets.Trained the models with 80% of the dataset and test with the remaining 20% of data.We have also done Hyper-tuning.Feature selection was an important part of our study,in one dataset named 'Admit.csv' it was not relevant since the features were highly co-related with the output variable.In other two datasets,we have received some good response in terms of fewer errors.We have use RMSE and R2_Score to evaluate the models.In two dataset 'Random Forest' has given us the best result and in one dataset 'MultiLinear Regression' has the best performance.Except this 'SVR' model was well and good in every model.

5.2 Future Work

In future,deep learning models(e.g. CNN.RNN.LSTM) can be used to test the datasets and will get more accurate results.The methods can be applied on more different types of datasets.One can try with other Feature Selection methods.

Reference

1.Kajal Singh, Anukriti Mukherjee,"Reliable Algorithms for Machine Learning Models: Implementation Research in Data Science", International Journal of Recent Technology and Engineering ISSN: 2277-3878 (Online), Volume-10 Issue-6, March 2022 https://www.academia.edu/76661997/Reliable_Algorithms_for_Machine_Learning_Models_Implementation_Research_in_Data_Science

2. www.scholarpedia.org/article/Mutual_information#Definition

3. Sunny Kumar, Kanika Agrawal, Nelshan Mandan," Red Wine Quality Prediction Using Machine Learning Techniques", 2020 International Conference on Computer Communication and Informatics **(ICCCI),**https://ieeexplore.ieee.org/abstract/document/9104095/

4. Thamarai M, Malarvizhi SP. House Price Prediction Modeling Using Machine Learning. International Journal of Information Engineering & Electronic Business. 2020 Apr 1;12(2).

5. Park B, Bae JK. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. Expert systems with applications. 2015 Apr 15;42(6):2928-34.

6. Jamalinia E, Tehrani FS, Steele-Dunne SC, Vardon PJ. Predicting rainfall induced slope stability using random forest regression and synthetic data. InWorkshop on World Landslide Forum 2020 Nov 2 (pp. 223-229). Springer, Cham.

7.https://www.sciencedirect.com/topics/engineering/machine-learning-algorithm

8. https://en.wikipedia.org/wiki/Machine_learning