Applied Machine Learning:

Final Project Report

Lezos Dimitrios

ITC 6103 – Applied Machine Learning

Dr. Dimitrios Vogiatzis

Deree – The American College of Greece

1/4/2024

Contents

Word Count:

1 Regression

1.1 Introduction

For the clustering part of the assignment we will try to solve the following task:

> "You are hired by a company Gem Stones co ltd, which is a cubic zirconia manufacturer. You are provided with the dataset containing the prices and other attributes of more than 193,000 cubic zirconia (which is an inexpensive diamond alternative with many of the same qualities as a diamond). The company is earning different profits on different prize slots. You have to help the company in predicting the price for the stone on the basis of the details given in the dataset so it can distinguish between higher profitable stones and lower profitable stones so as to have a better profit share."

This was provided by Kaggle as a Playground competition (Walter Reade, Ashley Chow. (2023). Regression with a Tabular Gemstone Price Dataset. Kaggle. https://kaggle.com/competitions/playground-series-s3e8)

1.2 Data Description and Preprocessing

The dataset used contains more than 193,00 rows of data with 11 features of which we want to predict the Price. The features describe the properties of each gemstone as follows:

Variable	Description	Data Type
ID	ID of each row	Numeric
Carat	Carat weight of the cubic zirconia.	Numeric
Cut	Describe the cut quality of the cubic zirconia. Quality is increasing order Fair, Good, Very Good, Premium, Ideal.	Ordinal
Color	Colour of the cubic zirconia.With D being the best and J the worst.	Ordinal
Clarity	Cubic zirconia Clarity refers to the absence of the Inclusions and Blemishes. (In order from Best to Worst, FL = flawless, I3= level 3 inclusions) FL, IF, VVS1, VVS2, VS1, VS2, SI1, SI2, I1, I2, I3	Ordinal
Depth	The Height of a cubic zirconia, measured from the Culet to the table, divided by its average Girdle Diameter.	Numeric
Table	The Width of the cubic zirconia's Table expressed as a Percentage of its Average Diameter.	Numeric
Price	The Price of the cubic zirconia.	Numeric
X	Length of the cubic zirconia in mm.	Numeric
Y	Width of the cubic zirconia in mm.	Numeric
Z	Height of the cubic zirconia in mm.	Numeric

Table 1 Regression dataset variables

We expect that the **Price** of the gem is mostly driven by the **Carat** value (i.e. weight). Also the **X**, **Y**, **Z** values combined define the volume of the gem which should be linearly related to the weight.
Data was checked for nulls where no null values were found and for duplicate rows where no duplicate rows were found. It was also checked for variables containing only unique values, there was such a variable: **ID**, which was removed since it cannot contribute to the regression task.

1.2.1 Normalization

The regression algorithms were tested on the dataset for both Normalized and Non-normalized data.

1.2.2 Non-numeric Values

The dataset contains three non-numeric variables: **Cut**, **Color** and **Clarity**, that contain ordinal values on which we tried three different encodings in the regression algorithms:

- Ordinal encoding
- Helmert encoding
- Ordinal encoding normalized to the mean price per carat per value

1.2.3 Split of Data

The data were split between train and test sets with the use of **sklearn.model_selection.train_test_split** method with a 70%-30% ratio and a **random_state** value of 42. The final results were verified by taking the average of ten splits with different **random_state** values generated in random.

1.3 Exploratory Data Analysis

After encoding the ordinal variables to numeric we investigated the correlation of the variables with the following heatmap:

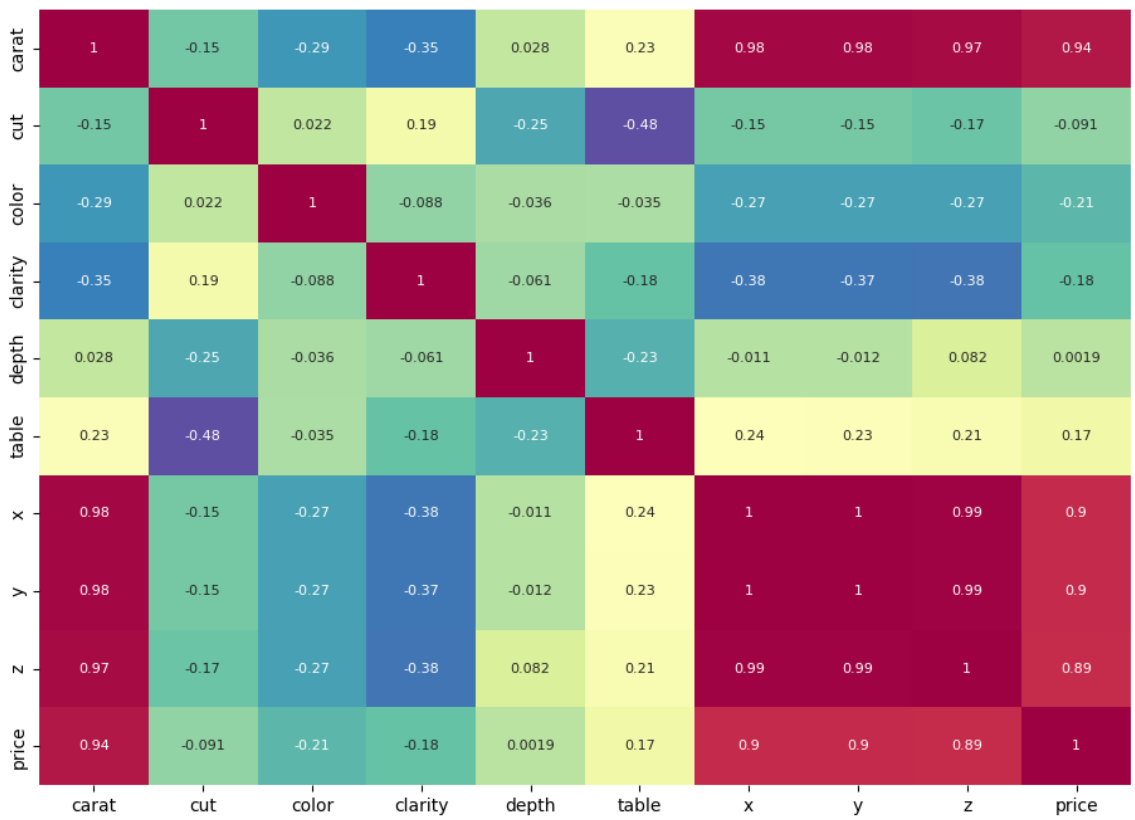

Figure 1 Correlation heatmap without ID variable

From the bottom left column we see that there is extremely high correlation between **Carat**, **X**, **Y** and **Z** variables which was expected as Carat describes the weight and X, Y, Z the size/volume of the gem. We decided to remove the X, Y and Z variables and keep only the Carat. After the removal we get a new heatmap:

Figure 2 Correlation heatmap of the usable variables

We notice that our dependent variable **Price** is highly correlated with the variable **Carat**, less correlated with variables **Color**, **Clarity**, **Table** and **Cut** and very loosely correlated with variable **Depth**. Also variables **Cut** and **Table** are moderately correlated.

After normalization we can view the distribution of the independent variables through a box plot:

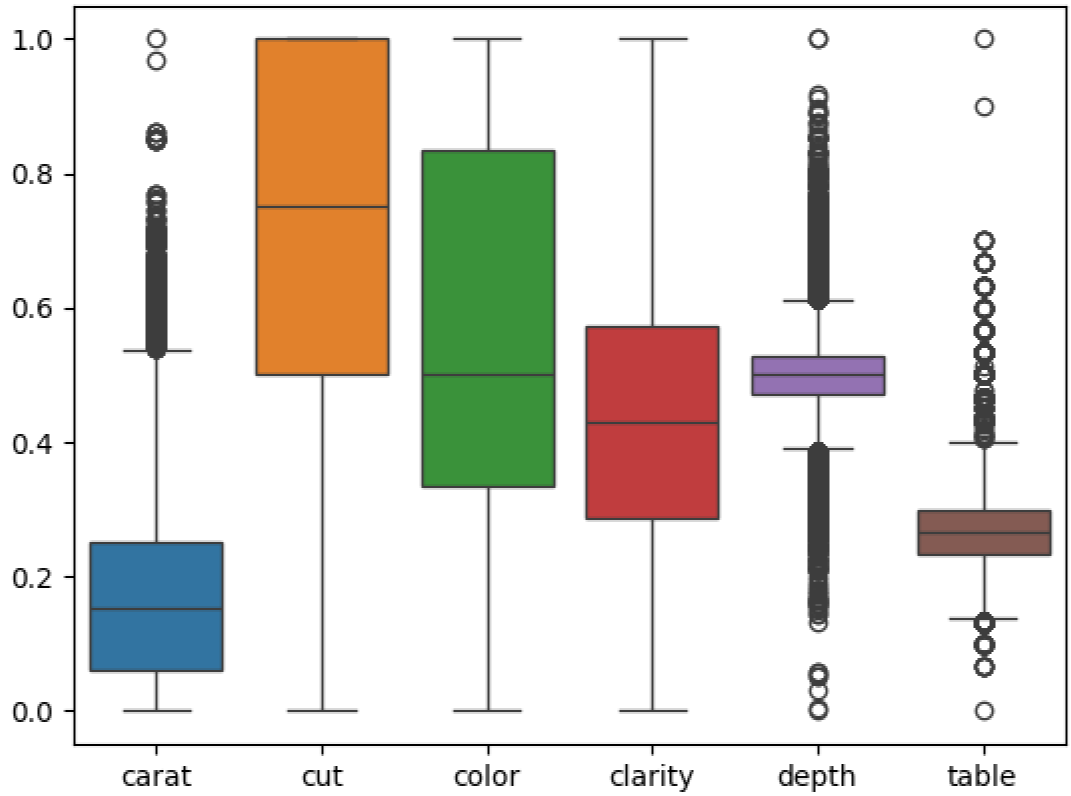

Figure 3 Boxplot of the independent variables

Where we see that **Carat** has a number of outliers, **Table** has a narrow distribution and some outliers and **Depth** has the narrowest distribution and the most outliers.

1.4 Methodology[1]

Our methodology consists of testing various regression techniques and approaches to identify the most accurate ones and then combining them together for the final prediction. The various techniques were run with various normalization and encoding options along with some additional tests as described in the following paragraphs.

1.4.1 Models

1.4.1.1 Linear Regression

Linear regression models the dependent variable over a linear relationship to the independent variables. In our case we used multiple linear regression that uses the formula:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon$$

Where β_0 is the intercept, β_1, β_2, … are the coefficients associated with each independent variable, x_1, x_2, … are the independent variables and ε represents the random error term

1.4.1.2 Polynomial Regression

Polynomial regression models the dependent variable over an n-degree polynomial function of the independent variables. A polynomial function of degree n has the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 {x_2}^2 + \cdots + \beta_n {x_n}^n + \varepsilon$$

[1] Some information in this section was retrieved by prompting ChatGPT

Where $\beta_0, \beta_1, \beta_2, \ldots$ are the coefficients associated with each degree and $x_1, x_2, \ldots$ are independent variables or combinations of them and ε represents the random error term

1.4.1.3 Ridge Regression

Ridge regression is a linear regression technique that adds a penalty term to the standard linear regression that helps prevent overfitting by shrinking the coefficients towards zero. Ridge regression,

- Shrinks the parameters
- Is more robust version of linear regression
- Puts constrains on regression coefficients to make them much more natural
- Is less subject to over-fitting, easier to interpret
- Useful when dealing with multicollinearity (high correlation between independent variables)
- Useful when the number of predictors is greater than the number of observations

1.4.1.4 Lasso Regression

Lasso regression, short for "Least Absolute Shrinkage and Selection Operator," is a linear regression technique that similar to Ridge regression, adds a penalty term to the standard linear regression that helps prevent overfitting by shrinking the coefficients towards zero. However, unlike ridge regression, lasso regression uses the L1 norm penalty, which encourages sparsity by setting some coefficients exactly to zero. Lasso regression,

- Performs feature selection
- Can set some coefficients to zero
- Useful for feature selection
- Powerful when dealing with high-dimensional data with many independent variables
- May not perform well in the presence of multicollinearity

1.4.1.5 Backward Stepwise Regression

Backward stepwise regression is a variable selection technique used in regression analysis to systematically identify the most important variables from the set of independent variables. It starts with a model that includes all the independent variables and iteratively removes the least significant ones. It simplifies the model by eliminating irrelevant or redundant variables.

1.4.1.6 Principal Components Analysis (PCA)

Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while preserving the most important information. PCA accomplishes this by identifying a set of orthogonal axes, called principal components, along which the data exhibits the maximum variance. These principal components capture the underlying structure of the data and can be used to represent the data in a lower-dimensional space.

1.4.1.7 Neural Network

We used a Feed Forward Neural Network consisting of three hidden layers with 100, 200 and 50 neurons with the ReLU activation function and a single output neuron with a linear activation function. In the scope of this project is only used as an educational comparison over the other regression techniques.

1.4.2 Encoding of Ordinal values

1.4.2.1 Ordinal Encoding

In Ordinal encoding the n different options of the ordinal variable take integer values from 0 to n following the provided ordering

1.4.2.2 Helmert Encoding

In Helmert encoding each level is represented relative to the mean of subsequent levels

1.4.2.3 Normalized Ordinal Encoding

In this normalized ordinal encoding we replaced each of the n different values of the ordinal variable with the average value of the dependent variable for each different value.

1.4.3 Metrices

For evaluating the accuracy of each technique we use the following metrices in decreasing importance:

1.4.3.1 R-Squared (R^2)

R-squared (R^2), also known as the coefficient of determination, is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variables in a regression model. It is a measure of how well the independent variables explain the variability of the dependent variable.

Calculation:

$$R^2 = 1 - \frac{SSE}{SST}$$

with

SSE: The residual sum of squares, which represents the unexplained variability of the dependent variable by the regression model.

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Where $\hat{y}_i$ is the predicted value of the dependent variable for the ith observation.

SST: The total sum of squares, which represents the total variability of the dependent variable.

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Where y_i is the actual value of the dependent variable for the ith observation, and $\bar{y}$ is the mean of the dependent variable.

Interpretation:

- R-squared ranges from 0 to 1.
- A value of 0 indicates that the model explains none of the variability of the dependent variable around its mean.
- A value of 1 indicates that the model explains all of the variability of the dependent variable around its mean.
- Higher values of R^2 indicate a better fit of the regression model to the data, as it suggests that more of the variability in the dependent variable is accounted for by the independent variables.

1.4.3.2 Mean Square Error (MSE)

MSE measures the average squared deviation or error between the actual target values and the predicted values generated by a regression model.

Calculation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Where n is the number of data points, y_i is the actual target value, $\hat{y}_i$ is the predicted value.

Interpretation:

- MSE provides a single, easily interpretable value representing the quality of predictions made by a model.
- A smaller MSE indicates that the model's predictions are closer to the actual values, implying higher accuracy and better fit.
- A larger MSE suggests that the model's predictions are less accurate and have higher variability from the actual values.

1.4.3.3 Mean Absolute Error (MAE)

MAE measures the absolute differences between the actual and predicted values without squaring them.

Calculation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Where n is the number of data points, y_i is the actual target value, $\hat{y}_i$ is the predicted value.

<u>Interpretation</u>:

- MAE provides a straightforward and interpretable measure of the average prediction error.
- A smaller MAE indicates that the model's predictions are closer to the actual values, implying higher accuracy.
- MAE is less sensitive to outliers compared to MSE since it does not square the errors. Therefore, it provides a more robust evaluation of model performance when dealing with data containing outliers.

1.5 Model Building

A list of the models and the techniques used for both normalized and non-normalized data follows:

1.5.1 Linear regression

1.5.1.1 Linear regression for each column

This is used to see if a single column can predict the **Price**.

1.5.1.2 Linear regression for all columns

This is used as a baseline result

1.5.2 Polynomial linear regression

1.5.2.1 Polynomial linear regression for each column with degree 2

This is used to see if a single column can predict the **Price**.

1.5.2.2 Polynomial linear regression for all columns with degrees from 2 to 7

This is to see which degree can provide the best prediction without suffering from overfitting.

1.5.3 Ridge regression

1.5.3.1 Ridge regression over the polynomial linear regression with degrees from 2 to 5

This is to test if Ridge regression can give better results than polynomial.

1.5.4 Lasso regression

1.5.4.1 Lasso regression over the polynomial linear regression with degrees from 2 to 5

This is to test if Lasso regression can give better results than polynomial.

1.5.5 Backward stepwise regression

1.5.5.1 Backward stepwise polynomial regression with degrees from 2 to 7

In this model we remove the variables in ascending order of correlation to the **Price**: **Depth, Cut, Table, Clarity, Color**

1.5.6 Principal Components Analysis

1.5.6.1 Backward stepwise polynomial regression with degrees from 2 to 5 over PCA components

In this model we run PCA analysis with 6 components and we use Backward stepwise polynomial regression by removing the components in ascending order of variance ratio.

1.5.7 Neural Network

1.5.7.1 100-200-50-1 feed forward network

We used a Feed Forward Neural Network consisting of three hidden layers with 100, 200 and 50 neurons with the ReLU activation function and a single output neuron with a linear activation function.

1.5.8 Best Models Combination

We combined the best performing models to achieve even better results.

1.5.8.1 Ridge regression over polynomial with degree 5 without variable ***Depth***

We combined the best models to a Ridge regression model over the polynomial features created with degree 5, without variable **Depth**.

1.5.9 Best Models additional cases

We also tried some other combinations to test if any of them can produce improved results.

1.5.9.1 Ridge regression over polynomial with degree 4 without variable ***Depth*** *with Helmert encoding*

Here we try the option of Helmert encoding.

1.5.9.2 Ridge regression over polynomial with degree 5 without variable ***Depth*** *with normalized Ordinal encoding*

Here we tried the option of normalized Ordinal encoding

1.5.9.3 Prediction of Price/Carat with Ridge regression over polynomial with degree 5 without variable ***Depth***

Here we tried to predict the Price/Carat value instead of Price

1.5.9.4 Prediction of Price/Carat with Ridge regression over polynomial with degrees 3 to 5 without variable ***Depth*** *and* ***Carat***

Here we tried to predict the Price/Carat value instead of Price and we also removed the Carat variable.

1.5.9.5 Verification of best results with average of ten different randomizations of Train – Test split

After generating the results we wanted to verify that these results were not dependent on the randomization of the splitting of data between Train and Test so we used 10 random numbers for splitting and took the average of those results.

1.6 Results

1.6.1 Results Summary

All the result are for non-normalized data as they provided better results in almost all cases. The results obtained over normalized data are noted as "Normalized" in the following table that we present the most significant results:

Model	R^2	MSE	MAE
Linear Regression	0.924825275	1214741	946
Polynomial Regression 5	0.975747319	391896	338
Polynomial Regression 5 no "Depth"	0.975995319	387889	338
Ridge Regression 5	0.976008910	387669	338
Ridge Regression 5 no "Depth"	0.976027437	387370	338
Normalized Polynomial Regression 6 no "Depth"	0.976631223	377614	333
Normalized, No outliers Polynomial Regression 5 no "Depth"	0.974615533	410185	350
Normalized Neural Network	0.961311469	625164	410
Average Ridge Regression 5 no "Depth"	0.975346274 σ: 0.0012	401634	339
Average Normalized Polynomial Regression 6 no "Depth"	0.975932703 σ: 0.0014	392103	335

Table 2 Results summary

1.6.2 Detailed Results

Detailed results are provided in Appendix

1.7 Conclusion

Trying to predict the **Price** dependent variable was successful with various models achieving R^2 values of over 0.97. The best results were given with the removal of variable **Depth** that had almost no correlation to **Price** and by using the Ridge regression model with non-normalized data but the use of Polynomial regression with degree 6 over normalized data with the removal of variable **Depth** seems to hit a "sweet spot" for this specific dataset and achieve an R^2 value of 0.97466 an improvement of 6 units in the 4^{th} significant digit over Ridge regression.

2 Appendix A

2.1 Detailed Results

2.1.1 Non-normalized data

Model	R^2	MSE	MAE
Linear Regression column 'carat'	0.889265606	1789347	946
Linear Regression column 'cut'	0.007526345	16037296	3041
Linear Regression column 'color'	0.044814255	15434764	2961
Linear Regression column 'clarity'	0.032819140	15628592	2952
Linear Regression column 'depth'	-0.000034842	16159476	3067
Linear Regression column 'table'	0.030648561	15663666	2987
Linear Regression	0.924825275	1214742	809
Polynomial Regression degree=2 column 'carat'	0.903507981	1559206	756
Polynomial Regression degree=2 column 'cut'	0.013602264	15939116	3016
Polynomial Regression degree=2 column 'color'	0.045747756	15419679	2960
Polynomial Regression degree=2 column 'clarity'	0.032932355	15626762	2951
Polynomial Regression degree=2 column 'depth'	0.002564612	16117472	3062
Polynomial Regression degree=2 column 'table'	0.034199744	15606283	2970
Polynomial Regression degree=2	0.964061538	580727	471
Polynomial Regression degree=3	0.973457798	428893	366
Polynomial Regression degree=4	0.975006852	403862	348
Polynomial Regression degree=5	0.975747319	391897	338
Polynomial Regression degree=6	0.972729970	440654	337
Polynomial Regression degree=7	0.875240140	2015984	340
Polynomial Regression degree=2 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.963923707	582954	471
Polynomial Regression degree=3 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.973354119	430568	367
Polynomial Regression degree=4 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.974958520	404643	348
Polynomial Regression degree=5 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.975995319	387890	338
Polynomial Regression degree=6 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.975386397	397729	334
Polynomial Regression degree=7 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.974967050	404505	328
Polynomial Regression degree=2 columns ['carat', 'color', 'clarity', 'table']	0.963218110	594355	475
Polynomial Regression degree=3 columns ['carat', 'color', 'clarity', 'table']	0.972686934	441349	371
Polynomial Regression degree=4 columns ['carat', 'color', 'clarity', 'table']	0.974075778	418907	353
Polynomial Regression degree=5 columns ['carat', 'color', 'clarity', 'table']	0.973994492	420221	346

Model	R^2	MSE	MAE
Polynomial Regression degree=6 columns ['carat', 'color', 'clarity', 'table']	0.973125784	434258	338
Polynomial Regression degree=7 columns ['carat', 'color', 'clarity', 'table']	0.965014997	565320	334
Polynomial Regression degree=2 columns ['carat', 'color', 'clarity']	0.962905296	599410	477
Polynomial Regression degree=3 columns ['carat', 'color', 'clarity']	0.972414459	445752	373
Polynomial Regression degree=4 columns ['carat', 'color', 'clarity']	0.973947252	420984	355
Polynomial Regression degree=5 columns ['carat', 'color', 'clarity']	0.975097489	402398	347
Polynomial Regression degree=6 columns ['carat', 'color', 'clarity']	0.975915686	389176	339
Polynomial Regression degree=7 columns ['carat', 'color', 'clarity']	0.973777013	423735	333
Polynomial Regression degree=2 columns ['carat', 'color']	0.918292102	1320311	709
Polynomial Regression degree=3 columns ['carat', 'color']	0.930808591	1118058	623
Polynomial Regression degree=4 columns ['carat', 'color']	0.931995862	1098873	614
Polynomial Regression degree=5 columns ['carat', 'color']	0.933789049	1069897	599
Polynomial Regression degree=6 columns ['carat', 'color']	0.934660158	1055821	593
Polynomial Regression degree=7 columns ['carat', 'color']	0.935200659	1047087	586
Polynomial Regression degree=2 columns ['carat']	0.903507981	1559206	756
Polynomial Regression degree=3 columns ['carat']	0.914129375	1387576	688
Polynomial Regression degree=4 columns ['carat']	0.914231637	1385924	685
Polynomial Regression degree=5 columns ['carat']	0.914814484	1376505	677
Polynomial Regression degree=6 columns ['carat']	0.914837062	1376141	677
Polynomial Regression degree=7 columns ['carat']	0.914878097	1375477	677
Neural Network Regression	0.960571489	637122	423
Ridge Regression degree=2	0.964062338	580713	470
Ridge Regression degree=3	0.973437771	429216	366
Ridge Regression degree=4	0.974949120	404795	347
Ridge Regression degree=5	0.963905098	583254	472
Lasso Regression degree=2	0.963905098	406127	349
Lasso Regression degree=3	0.976027437	387371	338
Lasso Regression degree=4	0.974168285	417412	361

Model	R^2	MSE	MAE
Lasso Regression degree=5	0.972086786	451047	342
Polynomial Regression degree=2 PCA	0.964061538	580727	471
Polynomial Regression degree=3 PCA	0.973457798	428893	366
Polynomial Regression degree=4 PCA	0.975006852	403862	348
Polynomial Regression degree=5 PCA	0.975741434	391992	338
Polynomial Regression degree=2 PCA components [1, 2, 3, 4, 5]	0.927691998	1168419	780
Polynomial Regression degree=3 PCA components [1, 2, 3, 4, 5]	0.942121646	935251	659
Polynomial Regression degree=4 PCA components [1, 2, 3, 4, 5]	0.948605159	830485	623
Polynomial Regression degree=5 PCA components [1, 2, 3, 4, 5]	0.780571507	3545726	615
Polynomial Regression degree=2 PCA components [2, 3, 4, 5]	0.908874319	1472492	895
Polynomial Regression degree=3 PCA components [2, 3, 4, 5]	0.921934870	1261448	790
Polynomial Regression degree=4 PCA components [2, 3, 4, 5]	0.929141437	1144997	758
Polynomial Regression degree=5 PCA components [2, 3, 4, 5]	0.801857635	3201765	753
Polynomial Regression degree=2 PCA components [3, 4, 5]	0.880945691	1923788	1038
Polynomial Regression degree=3 PCA components [3, 4, 5]	0.891309432	1756321	984
Polynomial Regression degree=4 PCA components [3, 4, 5]	0.899298691	1627224	951
Polynomial Regression degree=5 PCA components [3, 4, 5]	0.814886368	2991235	944
Polynomial Regression degree=2 PCA components [4, 5]	0.877038993	1986916	1057
Polynomial Regression degree=3 PCA components [4, 5]	0.882380079	1900610	1025
Polynomial Regression degree=4 PCA components [4, 5]	0.883531452	1882005	1010
Polynomial Regression degree=5 PCA components [4, 5]	0.869493065	2108850	1014
Polynomial Regression degree=2 PCA components [5]	0.872570117	2059128	1075
Polynomial Regression degree=3 PCA components [5]	0.879619706	1945215	1043
Polynomial Regression degree=4 PCA components [5]	0.884646418	1863989	1030
Polynomial Regression degree=5 PCA components [5]	0.885311811	1853237	1034
Best of class Ridge, Regression degree=5 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.976027436	387371	338
No outliers	0.974615533	410185	351

Model	R^2	MSE	MAE
Ridge, Regression degree=5 Columns ['carat', 'cut', 'color', 'clarity', 'table']			
Helmert Ordinal Encoding Ridge, Regression degree=4 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.972086786	451047	342
Normalized Ordinal with Price/Carat Ridge, Regression degree=5 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.932209878	1095415	640
Price/Carat prediction Ridge, Regression degree=5 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.928233308	279909	347
Price/Carat prediction Ridge, Regression degree=3 Columns ['cut', 'color', 'clarity', 'table']	0.044714987	3725861	1487
Price/Carat prediction Ridge, Regression degree=4 Columns ['cut', 'color', 'clarity', 'table']	0.052595848	3695124	1482
Price/Carat prediction Ridge, Regression degree=5 Columns ['cut', 'color', 'clarity', 'table']	0.032167966	374798	1478

Table 3 Results for non-normalized data

2.1.2 Normalized data

Model	R^2	MSE	MAE
Linear Regression column 'carat'	0.889265606	1789347	946
Linear Regression column 'cut'	0.007526345	16037296	3041
Linear Regression column 'color'	0.044814255	15434764	2961
Linear Regression column 'clarity'	0.032819140	15628592	2952
Linear Regression column 'depth'	-0.000034842	16159476	3067
Linear Regression column 'table'	0.030648561	15663666	2987
Linear Regression	0.924825275	1214742	809
Polynomial Regression degree=2 column 'carat'	0.903507981	1559206	756
Polynomial Regression degree=2 column 'cut'	0.013602264	15939116	3016
Polynomial Regression degree=2 column 'color'	0.045747756	15419679	2960
Polynomial Regression degree=2 column 'clarity'	0.032932355	15626762	2951
Polynomial Regression degree=2 column 'depth'	0.002564612	16117472	3062
Polynomial Regression degree=2 column 'table'	0.034199744	15606283	2970
Polynomial Regression degree=2	0.964061537	580726	471
Polynomial Regression degree=3	0.973457797	428893	366
Polynomial Regression degree=4	0.975006852	403862	348
Polynomial Regression degree=5	0.975634559	393719	339
Polynomial Regression degree=6	0.923869390	1230188	341
Polynomial Regression degree=7	-2.329052573	53793872	371
Polynomial Regression degree=2 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.963923706	582954	471

Model	R^2	MSE	MAE
Polynomial Regression degree=3 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.973354119	430568	367
Polynomial Regression degree=4 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.974958519	404643	348
Polynomial Regression degree=5 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.975995449	387887	338
Polynomial Regression degree=6 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.976631223	377614	333
Polynomial Regression degree=7 columns ['carat', 'cut', 'color', 'clarity', 'table']	0.223325186	12550221	345
Polynomial Regression degree=2 columns ['carat', 'color', 'clarity', 'table']	0.963218109	594355	475
Polynomial Regression degree=3 columns ['carat', 'color', 'clarity', 'table']	0.972686934	441349	371
Polynomial Regression degree=4 columns ['carat', 'color', 'clarity', 'table']	0.974075778	418907	353
Polynomial Regression degree=5 columns ['carat', 'color', 'clarity', 'table']	0.974015779	419877	346
Polynomial Regression degree=6 columns ['carat', 'color', 'clarity', 'table']	0.956173856	708183	340
Polynomial Regression degree=7 columns ['carat', 'color', 'clarity', 'table']	0.728748253	4383133	341
Polynomial Regression degree=2 columns ['carat', 'color', 'clarity']	0.962905295	599410	477
Polynomial Regression degree=3 columns ['carat', 'color', 'clarity']	0.972414458	445752	373
Polynomial Regression degree=4 columns ['carat', 'color', 'clarity']	0.973947251	420984	355
Polynomial Regression degree=5 columns ['carat', 'color', 'clarity']	0.975097489	402398	347
Polynomial Regression degree=6 columns ['carat', 'color', 'clarity']	0.975915686	389176	339
Polynomial Regression degree=7 columns ['carat', 'color', 'clarity']	0.973777013	423735	333
Polynomial Regression degree=2 columns ['carat', 'color']	0.918292102	1320311	709
Polynomial Regression degree=3 columns ['carat', 'color']	0.930808591	1118058	623
Polynomial Regression degree=4 columns ['carat', 'color']	0.931995862	1098873	614
Polynomial Regression degree=5 columns ['carat', 'color']	0.933789049	1069897	599
Polynomial Regression degree=6 columns ['carat', 'color']	0.934660158	1055821	593
Polynomial Regression degree=7 columns ['carat', 'color']	0.935200659	1047087	586
Polynomial Regression degree=2 columns ['carat']	0.903507981	1559206	756
Polynomial Regression degree=3 columns ['carat']	0.914129375	1387576	688

Model	R^2	MSE	MAE
Polynomial Regression degree=4 columns ['carat']	0.914231638	1385924	685
Polynomial Regression degree=5 columns ['carat']	0.914814484	1376505	677
Polynomial Regression degree=6 columns ['carat']	0.914837062	1376141	677
Polynomial Regression degree=7 columns ['carat']	0.914878097	1375477	677
Neural Network Regression	0.961311469	625165	410
Ridge Regression degree=2	0.964056228	580812	471
Ridge Regression degree=3	0.973414628	429591	364
Ridge Regression degree=4	0.974148257	417736	356
Ridge Regression degree=5	0.974687754	409018	350
Lasso Regression degree=2	0.963943796	582629	471
Lasso Regression degree=3	0.972849229	438727	364
Lasso Regression degree=4	0.973456834	428909	363
Lasso Regression degree=5	0.973656974	425675	363
Polynomial Regression degree=2 PCA	0.964061538	580727	471
Polynomial Regression degree=3 PCA	0.973457798	428893	366
Polynomial Regression degree=4 PCA	0.975006852	403862	348
Polynomial Regression degree=5 PCA	0.975741434	391992	338
Polynomial Regression degree=2 PCA components [1, 2, 3, 4, 5]	0.902854090	1569772	957
Polynomial Regression degree=3 PCA components [1, 2, 3, 4, 5]	0.919606155	1299077	836
Polynomial Regression degree=4 PCA components [1, 2, 3, 4, 5]	0.928989005	1147461	769
Polynomial Regression degree=5 PCA components [1, 2, 3, 4, 5]	0.927251738	1175533	722
Polynomial Regression degree=2 PCA components [2, 3, 4, 5]	0.880746465	1927008	1048
Polynomial Regression degree=3 PCA components [2, 3, 4, 5]	0.890088132	1776056	993
Polynomial Regression degree=4 PCA components [2, 3, 4, 5]	0.874890615	2021632	955
Polynomial Regression degree=5 PCA components [2, 3, 4, 5]	0.827005567	2795402	949
Polynomial Regression degree=2 PCA components [3, 4, 5]	0.810487433	3062317	1314
Polynomial Regression degree=3 PCA components [3, 4, 5]	0.811920384	3039162	1310
Polynomial Regression degree=4 PCA components [3, 4, 5]	0.805860714	3137080	1293
Polynomial Regression degree=5 PCA components [3, 4, 5]	0.822044592	2875566	1292
Polynomial Regression degree=2 PCA components [4, 5]	0.006226296	16058303	3057

Model	R^2	MSE	MAE
Polynomial Regression degree=3 PCA components [4, 5]	0.006769061	16049533	3056
Polynomial Regression degree=4 PCA components [4, 5]	0.006364917	16056063	3053
Polynomial Regression degree=5 PCA components [4, 5]	0.007216608	16042301	3051
Polynomial Regression degree=2 PCA components [5]	0.001450225	16135479	3066
Polynomial Regression degree=3 PCA components [5]	0.001605379	16132972	3066
Polynomial Regression degree=4 PCA components [5]	-0.055278393	17052152	3069
Polynomial Regression degree=5 PCA components [5]	-0.003150333	16209819	3065
Best of class Ridge, Regression degree=5 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.974618755	410133	350
No outliers Ridge, Regression degree=5 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.974615533	410185	351
Helmert Ordinal Encoding Ridge, Regression degree=4 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.972086786	451047	342
Normalized Ordinal with Price/Carat Ridge, Regression degree=5 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.932209878	1095415	640
Price/Carat prediction Ridge, Regression degree=5 Columns ['carat', 'cut', 'color', 'clarity', 'table']	0.928233307	279909	247
Price/Carat prediction Ridge, Regression degree=3 Columns ['cut', 'color', 'clarity', 'table']	0.044714986	3725861	1487
Price/Carat prediction Ridge, Regression degree=4 Columns ['cut', 'color', 'clarity', 'table']	0.052595847	3695124	1482
Price/Carat prediction Ridge, Regression degree=5 Columns ['cut', 'color', 'clarity', 'table']	0.032167966	3774798	1478

Table 4 Results for normalized data