# **A model predicting the Critical Temperature of Super Conducting Materials.**

By Rashmi Patel

**Business Understanding:**

Superconductivity is the ability of some materials that offers zero resistance when electric current is passed through it. This property yields interesting and potentially useful effects., Low temperatures are required for a material to behave as a superconductor. All electrically conductive materials have different resistances with different temperatures that the material is subjected to.

The case study provides us data of different materials which is a mixture of elements from periodic table and the critical temperature at which these materials exhibit super conductive properties. The goal of the case study is to perform modelling on the data provided which can predict the critical temperature for any new material.

**Data Understanding:**

The data provided contains 2 files.

1. train.csv and
2. unique\_m.csv

The “train.csv” file consists of 82 features and “unique\_m.csv” file consists of 88 features. The data is related one-to-one on based on rows of the csv files provided.

The attribute ‘critical\_temp’ is present in both the files, so we have dropped one when combining the 2 csv files. The attribute “materials” describes the material in a string format and does not contribute to model building. After eliminating these 2 columns, the data contains 21,263 rows and 168 columns.

The 88 attributes that came from “unique\_m.csv” are nothing but the elements in the periodic table. Because one-hot encoding is done on the data, a zero (0) in this attribute indicates that element is not present the mixture being created and one (1) indicates the presence of that element in the mixture. So, we can say that the 88 attributes in unique\_m.csv contains categorical data and there is no need to conversion since the data is provided in proper format.

Looking at the plot below we can say that the critical temperature is skewed heaving on the right side. So, to normalize this right skewed distribution, we will apply log transform on the target variable.

Chart, histogram

Description automatically generated

Even after the transformation, the data is still skewed but this time towards left. So considering that the dataset is large enough, we can violate the assumption of normality.

Chart, histogram

Description automatically generated

**Missing and duplicate values:**

There is no missing and duplicate data found in the dataset provided.

**Standardization:**

The attributes from train.csv is numeric. Since, each feature varies on scale we need to standardize it through which we can prevent over-emphasizing of other attributes.

**Correlation of each attribute with temperature:**

We found that some attributes are strongly correlated with critical temperature either positively or negatively. Some of them are: wtd\_entropy\_atomic\_mass, wtd\_range\_atomic\_mass, etc. are positively correlated and wtd\_entropy\_valence, wtd\_gmean\_valence, etc. are negatively corelated. This can be seen in the correlation plot provided below.

Chart

Description automatically generated

**Model preparation:**

We are going to use L1-LASSO and L2-Ridge methods to perform regression, but prior performing these methods, we will build the regular regression model.

The first step is to split the data into 2 parts, using 75/25 split. One will be training and other is test set. The test will not contain the target variable.



Now we, will build a regular regression model which will help in understanding baseline coefficients and performance.

Graphical user interface, application

Description automatically generated

Looking at the output, generated by the regular linear regression:

Table

Description automatically generated with medium confidence

**L1-LASSO model**

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Text, table

Description automatically generated

We found that the best alpha selected by LASSO method is 0.533 and RSS (Residuals sum of squares) is 4372455.525331622

**L2-Ridge Model:**

**Graphical user interface, text, application

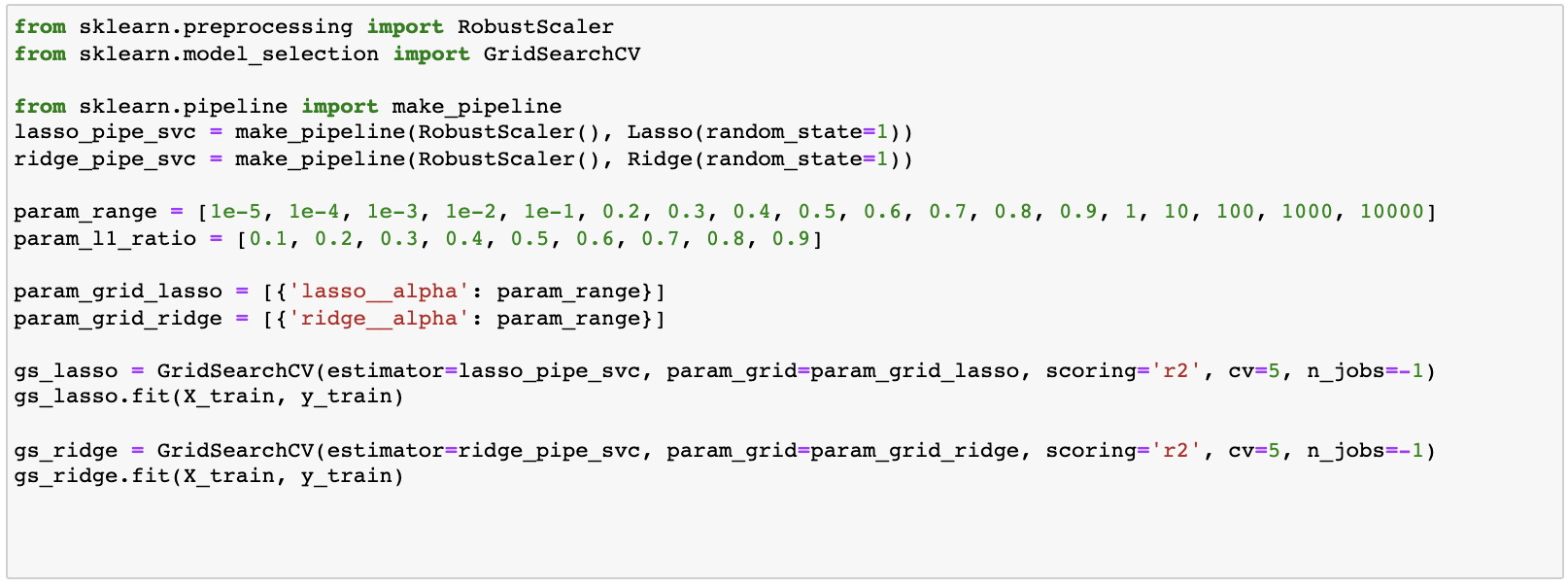
Description automatically generated**

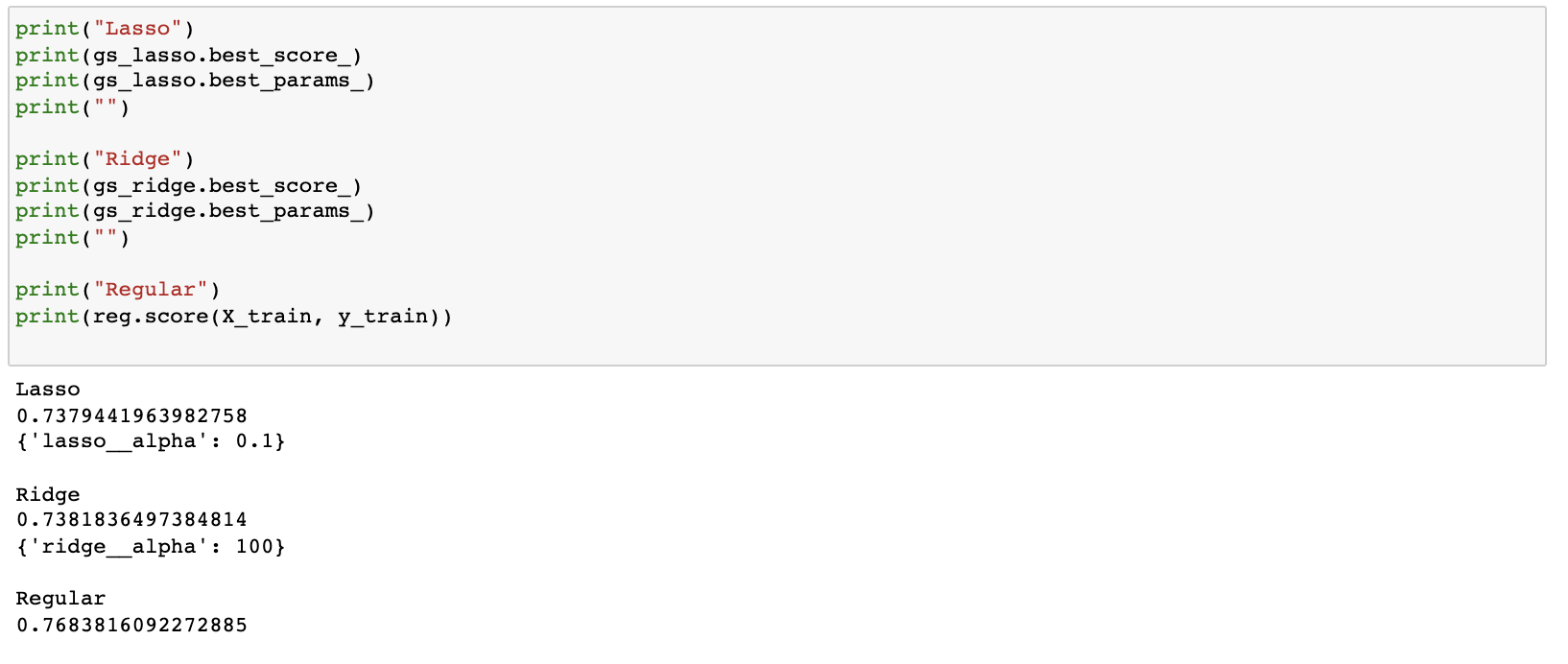
**Text

Description automatically generated**

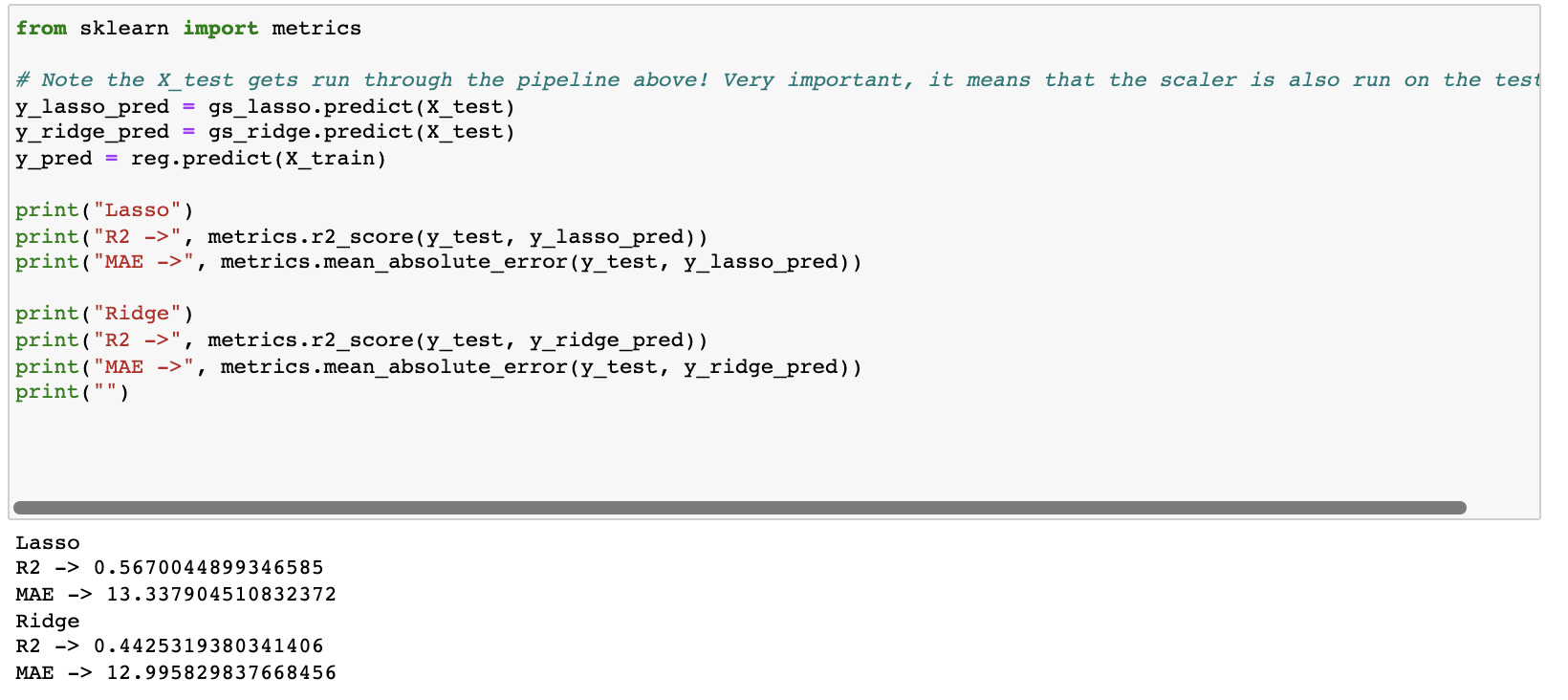
We found that the best alpha selected by Ridge method is 100.0 and RSS (Residual sum of squares) is 4513122.382153014.

**Comparing the accuracy score for all models:**

****

****

We found that each model performs almost the same on the training data. Now we will look at the prediction obtained using L1-LASSO and L2-Ridge model techniques:

****

It seems that Ridge is doing better although the R2 is less than LASSO because lower the MAE better is the prediction.

# **Conclusions:**

In short, we recommend using Ridge regression when attempting to predict critical temperatures of superconducting materials.

Based on the GridsearchCV technique used for model comparison, we can conclude that Ridge regression produces better prediction than LASSO because lower is the MAE (mean absolute error) better is the prediction done by model.

Based on RSS (Residual sum of squares) we can conclude the same that, ridge method is doing better than LASSO in predicting the critical temperature of superconductor materials because lower the RSS, better is the model prediction.

For comparing more precisely, we can also work on looking at the MSE (Mean Squared Error), RMSE (Root Mean Squared error), R2 (R-squared) etc. factors.