Business Report

SMDM Project Business Report DSBA

*Capstone project – Supply Chain Management*

***Sanjay Srinivasan***

***PGP-DSBA Online***

***JULY’ 21 Batch***

***Date: 12-06-2022***

# *INDEX*

|  |  |  |
| --- | --- | --- |
| ***S. No*** | ***Contents*** | ***Page* No** |
| **1)** | **Model building and interpretation.** | 6 |
|  | a) Build various models (You can choose to build models for either or all of descriptive, predictive or prescriptive purposes). |  |
|  | b) Test your predictive model against the test set using various appropriate performance metrics. | 15 |
|  | c) Interpretation of the model(s). | 15 |
|  |  |  |
| **2)** | **Model Tuning** | 16 |
|  | a) Ensemble modelling, wherever applicable. | 16 |
|  | b) Any other model tuning measures(if applicable) | 20 |
|  | c) Interpretation of the most optimum model and its implication on the business | 20 |

***List Of Figures***

|  |  |  |
| --- | --- | --- |
| ***S.No*** | ***Content*** | ***Page No*** |
| 1.1 | Coefficient for Linear Regression model | 7 |
| 1.2 | Intercept for Linear Regression model | 7 |
| 1.3 | R-Square for Linear Regression model | 7 |
| 1.4 | Mean Squared Error and Root Mean Squared Error for Linear Regression model | 7 |
| 1.5 | Mean Absolute Error for Linear Regression model | 7 |
| 1.6 | Mean Absolute Percentage Error for Linear Regression model | 7 |
| 1.7 | Scatter plot for Linear Regression model | 8 |
| 1.8 | Variance Inflation Factor for Linear Regression model | 8 |
| 1.9 | Intercept for Lasso Regression model | 9 |
| 1.10 | R-Square for Lasso Regression model | 9 |
| 1.11 | Mean Squared Error and Root Mean Squared Error for Lasso Regression model | 9 |
| 1.12 | Mean Absolute Error for Lasso Regression model | 9 |
| 1.13 | Mean Absolute Percentage Error for Lasso Regression model | 9 |
| 1.14 | Scatter plot for Lasso Regression model | 9 |
| 1.15 | Coefficient of Ridge Regression model | 10 |
| 1.16 | Intercept of Ridge Regression model | 10 |
| 1.17 | R-Square of Ridge Regression model | 10 |
| 1.18 | Mean Squared Error and Root Mean Squared Error for Ridge Regression model | 10 |
| 1.19 | Mean Absolute Error for Ridge Regression model | 10 |
| 1.20 | Mean Absolute Percentage Error for Ridge Regression model | 11 |
| 1.21 | Scatterplot for Ridge Regression model | 11 |
| 1.22 | Intercept for Support Vector Regression model | 11 |
| 1.23 | R-Square for Support Vector Regression model | 11 |
| 1.24 | Mean Squared Error and Root Mean Squared Error for Support Vector Regression model | 11 |
| 1.25 | Mean Absolute Error for Support Vector Regression model | 11 |
| 1.26 | Mean Absolute Percentage Error for Support Vector Regression model | 11 |
| 1.27 | Scatterplot for Support Vector Regression model | 11 |
| 1.28 | Coefficient of Huber Regression model | 12 |
| 1.29 | Intercept of Huber Regression model | 12 |
| 1.30 | R-Square of Huber Regression model | 12 |
| 1.31 | Mean Squared Error and Root Mean Squared Error for Huber Regression model | 12 |
| 1.32 | Mean Absolute Error for Huber Regression model | 12 |
| 1.33 | Mean Absolute Percentage Error for Huber Regression model | 13 |
| 1.34 | Scatterplot for Huber Regression model | 13 |
| 1.35 | R-Square for Random Forest Regression model | 13 |
| 1.36 | Mean Squared Error and Root Mean Squared Error for Random Forest Regression model | 13 |
| 1.37 | Mean Absolute Error for Random Forest Regression model | 13 |
| 1.38 | Mean Absolute Percentage Error for Random Forest Regression model | 13 |
| 1.39 | Scatterplot for Random Forest Regression model | 13 |
| 1.40 | R-Square for ANN Regression model | 14 |
| 1.41 | Mean Squared Error and Root Mean Squared Error for ANN Regression model | 14 |
| 1.42 | Mean Absolute Error for ANN Regression model | 14 |
| 1.43 | Mean Absolute Percentage Error for ANN Regression model | 14 |
| 1.44 | Scatterplot for ANN Regression model | 14 |
| 1.45 | R-Square for Adaboost Regression model | 14 |
| 1.46 | Mean Squared Error and Root Mean Squared Error for Adaboost Regression model | 14 |
| 1.47 | Mean Absolute Error for Adaboost Regression model | 15 |
| 1.48 | Mean Absolute Percentage Error for Adaboost Regression model | 15 |
| 1.49 | Scatterplot for Adaboost Regression model | 15 |
| 1.50 | Performance metrics of all Regression model | 15 |
| 1.51 | Sort Values based on best model performance metrics. | 16 |
| 1.52 | Intializing RF regressor model using gridsearch. | 16 |
| 1.53 | R-Square RF regressor model using gridsearch. | 16 |
| 1.54 | Mean Squared Error and Root Mean Squared Error RF regressor model using gridsearch. | 16 |
| 1.55 | Mean Absolute Error RF regressor model using gridsearch. | 17 |
| 1.56 | Mean Absolute Percentage Error RF regressor model using gridsearch | 17 |
| 1.57 | Scatterplot RF regressor model using gridsearch. | 17 |
| 1.58 | Initialising Gradient boosting regressor model using gridsearch. | 17 |
| 1.59 | R-Square Gradient boosting regressor model using gridsearch. | 17 |
| 1.60 | Mean Squared Error and Root Mean Squared Error Gradient boosting regressor model using gridsearch | 17 |
| 1.61 | Mean Absolute Error Gradient boosting regressor model using gridsearch | 18 |
| 1.62 | Mean Absolute Percentage Error Gradient boosting regressor model using gridsearch | 18 |
| 1.63 | Scatterplot of Gradient boosting regressor model using gridsearch | 18 |
| 1.64 | Initialising Bagging regressor model using gridsearch. | 18 |
| 1.65 | R-Square Bagging regressor model using gridsearch | 18 |
| 1.66 | Mean Squared Error and Root Mean Squared Error of Bagging regressor model using gridsearch. | 18 |
| 1.67 | Mean Absolute Error of Bagging regressor model using gridsearch. | 19 |
| 1.68 | Mean Absolute Percentage Error of Bagging regressor model using gridsearch. | 19 |
| 1.69 | Scatterplot of Bagging regressor model using gridsearch. | 19 |
| 1.70 | Initialising Ridge regressor model using gridsearch. | 19 |
| 1.71 | R-Square Ridge regressor model using gridsearch. | 19 |
| 1.72 | Mean Squared Error and Root Mean Squared Error of Ridge regressor model using gridsearch. | 19 |
| 1.73 | Mean Absolute Error of Ridge regressor model using gridsearch. | 19 |
| 1.74 | Mean Absolute Percentage Error of Ridge regressor model using gridsearch. | 19 |
| 1.75 | Scatterplot of Ridge regressor model using gridsearch. | 20 |

# Model building and interpretation.

*a) Build various models (You can choose to build models for either or all of descriptive, predictive or prescriptive purposes).*

A machine learning model is **built by learning and generalizing from training data, and make predictions for the business problem.**

In regression analysis, model building **is the process of developing a probabilistic model that best describes the relationship between the dependent and independent variables**.

**Steps involved in model building:**

1. Problem Statement
2. Data Collection.
3. Data Cleaning.
4. Exploratory Data Analysis.
5. Model Development
6. Train the Model.
7. Test the Models.
8. Applying Models
9. Inferences, Recommendations and business insights based on the model

The four main analytical models organisations can deploy are:

1. Descriptive
2. Diagnostic
3. Predictive
4. Prescriptive

***Descriptive* -** ***It generally uses historical data*** from a single internal source to pinpoint when an event occurred. ***Descriptive analytics are often displayed on dashboards and in reports.***

***Diagnostic* -** A diagnostic model is a framework for **identifying, analysing and interpreting data in a given context to identify possible needs.** An effective diagnostic model **allows identifying reliable data to help clients better understand their company's strengths, deficiencies, and opportunities for improvement, to later articulate a targeted intervention and measurement strategy.**

***Predictive -*** Predictive modelling**is a mathematical process used to predict future events or outcomes by analyzing patterns in a given set of input data.**

***Prescriptive* -** Prescriptive analytics **utilizes similar modelling structures to predict outcomes and then utilizes a combination of machine learning, business rules, artificial intelligence, and algorithms to simulate various approaches to these numerous outcomes.**

***Model building is performed with the following Model:***

* Linear Regression
* Lasso Regression (L1 Regularization Model)
* Ridge Regression (L2 Regularization Model)
* Support Vector Regression Model
* Huber Regression
* Random Forest Regressor model
* Artificial Neural Network Regressor Model
* Ada Boost Regressor Model

***To find Parameters for Regression model:***

***RMSE (Root Mean Squared Error)*** - The root-mean-square error (RMSE) is a **frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed.** The best model will have the value in the **range 0.2 - 0.5**

***MSE (Mean Squared Error)*** - In Statistics, Mean Square Error (MSE) is defined as **Mean or Average of the square of the difference between actual and estimated values.** The best model will have the value lesser value (0 is the best value for the model).

***MAE (Mean Absolute Error)*** - The **MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction.**

***R-Squared value* -** R-squared is a **statistical measure that represents the goodness of fit of a regression model.** The ideal value for r-square is 1. The closer the value of r-square to 1, the better is the model fitted. The best model will have the value higher value(1 is the best value for the model).

***Linear Regression Model:***

Linear regression analysis is used to **predict the value of a variable based on the value of another variable**. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

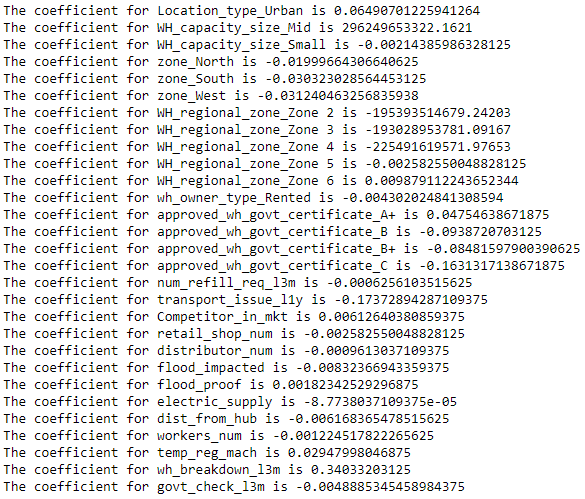


Fig 1.1 Coefficient for Linear Regression model



Fig 1.2 Intercept for Linear Regression model



Fig 1.3 R-Square for Linear Regression model

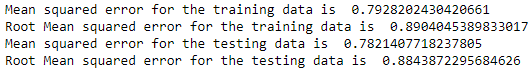


Fig 1.4 Mean Squared Error and Root Mean Squared Error for Linear Regression model



Fig 1.5 Mean Absolute Error for Linear Regression model



Fig 1.6 Mean Absolute Percentage Error for Linear Regression model

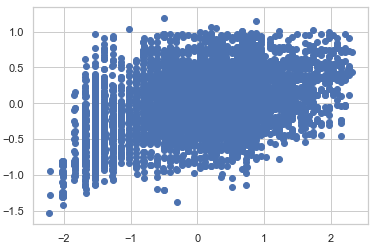


Fig 1.7 Scatter plot for Linear Regression model



Fig 1.8 Variance Inflation Factor for Linear Regression model

***Assumptions of Linear Regression:***

* Linear relationship.
* Multivariate normality.
* No or little multicollinearity.
* No auto-correlation.
* Homoscedasticity.

***Lasso Regression Model:***

In statistics and machine learning, lasso (least absolute shrinkage and selection operator; also Lasso or LASSO) is a **regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.**

***Assumptions of Lasso Regression:***

* Linearity - linear regression needs the relationship between the predictor and target variables to be linear
* Independence
* No Heteroskedasticity
* No Multicolinearity.



Fig 1.9 Intercept for Lasso Regression model



Fig 1.10 R-Square for Lasso Regression model

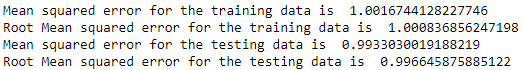


Fig 1.11 Mean Squared Error and Root Mean Squared Error for Lasso Regression model



Fig 1.12 Mean Absolute Error for Lasso Regression model



Fig 1.13 Mean Absolute Percentage Error for Lasso Regression model

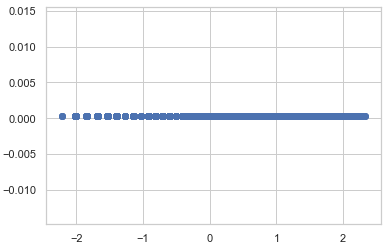


Fig 1.14 Scatter plot for Lasso Regression model

### *Ridge Regression:*

Ridge regression is a **method of estimating the coefficients of multiple-regression models in scenarios where linearly independent variables are highly correlated**.

***Assumptions of Ridge Regression:***

* Linearity
* Constant Variance
* Independence

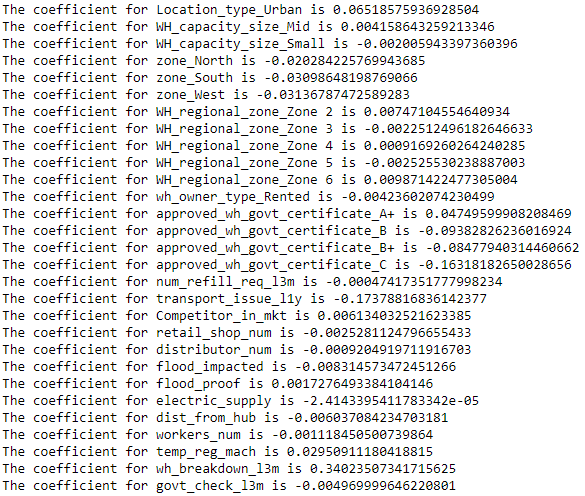


Fig 1.15 Coefficient of Ridge Regression model



Fig 1.16 Intercept of Ridge Regression model



Fig 1.17 R-Square of Ridge Regression model

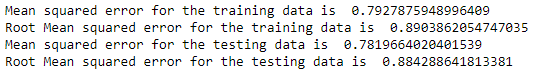


Fig 1.18 Mean Squared Error and Root Mean Squared Error for Ridge Regression model



Fig 1.19 Mean Absolute Error for Ridge Regression model



Fig 1.20 Mean Absolute Percentage Error for Ridge Regression model

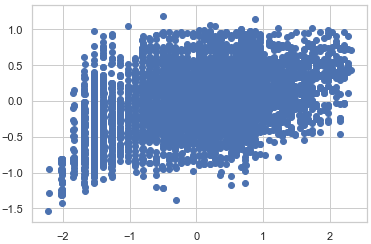


Fig 1.21 Scatterplot for Ridge Regression model

### *Support Vector Regression Model:*

Support Vector Regression is a **supervised learning algorithm that is used to predict discrete values.** Support Vector Regression uses the same principle as the SVMs. The basic idea behind SVR is to find the best fit line. In SVR, the best fit line is the hyper plane that has the maximum number of points.

SVR **gives us the flexibility to define how much error is acceptable in our model and will find an appropriate line (or hyper plane in higher dimensions) to fit the data**.



Fig 1.22 Intercept for Support Vector Regression model



Fig 1.23 R-Square for Support Vector Regression model





Fig 1.24 Mean Squared Error and Root Mean Squared Error for Support Vector Regression model



Fig 1.25 Mean Absolute Error for Support Vector Regression model



Fig 1.26 Mean Absolute Percentage Error for Support Vector Regression model

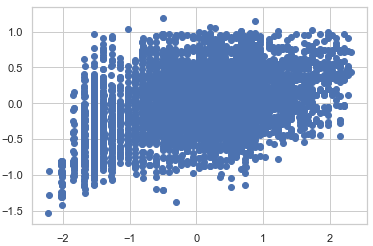
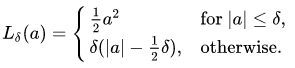


Fig 1.27 Scatterplot for Support Vector Regression model

### *Huber Regressor Model:*

The Huber Regressor optimizes the squared loss for the samples where |(y - X'w) / sigma| < epsilon and the absolute loss for the samples where |(y - X'w) / sigma| > epsilon , where w and sigma are parameters to be optimized.

Huber Loss formula:



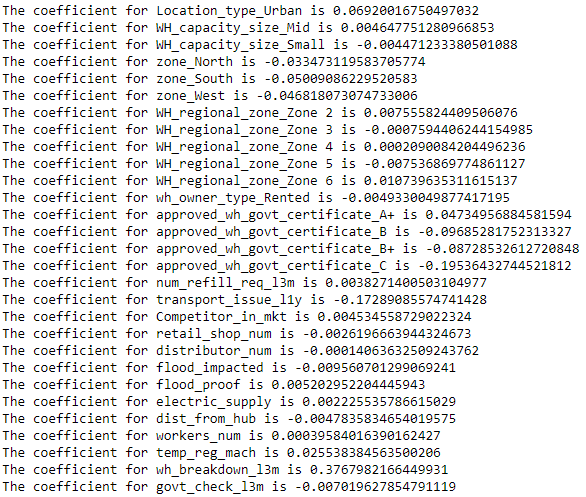


Fig 1.28 Coefficient of Huber Regression model



Fig 1.29 Intercept of Huber Regression model



Fig 1.30 R-Square of Huber Regression model

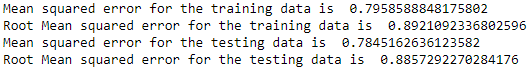


Fig 1.31 Mean Squared Error and Root Mean Squared Error for Huber Regression model



Fig 1.32 Mean Absolute Error for Huber Regression model



Fig 1.33 Mean Absolute Percentage Error for Huber Regression model

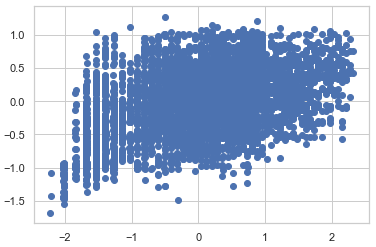


Fig 1.34 Scatterplot for Huber Regression model

### *Random Forest Regressor:*

A random forest regressor is **a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.**



Fig 1.35 R-Square for Random Forest Regression model





Fig 1.36 Mean Squared Error and Root Mean Squared Error for Random Forest Regression model



Fig 1.37 Mean Absolute Error for Random Forest Regression model



Fig 1.38 Mean Absolute Percentage Error for Random Forest Regression model

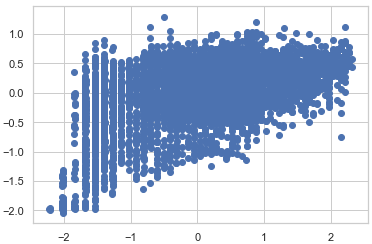


Fig 1.39 Scatterplot for Random Forest Regression model

### *Artificial Neural Network Regressor:*

The purpose of using Artificial Neural Networks for Regression over Linear Regression is that **the linear regression can only learn the linear relationship between the features and target and therefore cannot learn the complex non-linear relationship**.

Advantage of ANN regressor over linear regressor is complex non-linear relationship model can be built using ANN regressor model.



Fig 1.40 R-Square for ANN Regression model

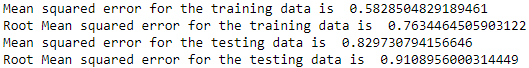


Fig 1.41 Mean Squared Error and Root Mean Squared Error for ANN Regression model



Fig 1.42 Mean Absolute Error for ANN Regression model



Fig 1.43 Mean Absolute Percentage Error for ANN Regression model

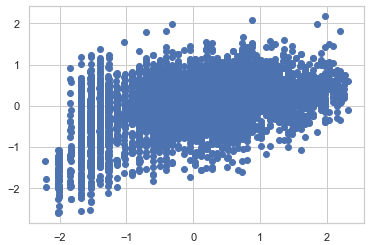


Fig 1.44 Scatterplot for ANN Regression model

### *AdaBoost Regressor:*

An AdaBoost regressor is a **meta-estimator that begins by fitting a regressor on the original dataset and then fits additional copies of the regressor on the same dataset but where the weights of instances are adjusted according to the error of the current prediction.**



Fig 1.45 R-Square for Adaboost Regression model





Fig 1.46 Mean Squared Error and Root Mean Squared Error for Adaboost Regression model



Fig 1.47 Mean Absolute Error for Adaboost Regression model



Fig 1.48 Mean Absolute Percentage Error for Adaboost Regression model

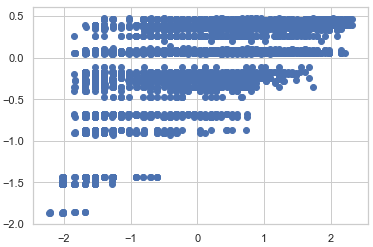


Fig 1.49 Scatterplot for Adaboost Regression model

*b.) Test your predictive model against the test set using various appropriate performance metrics.*

The performance metrics are predicted against the test set. The performance metrics used are:

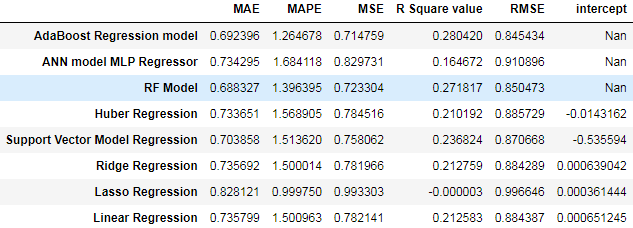


Fig 1.50 Performance metrics of all Regression model

The optimum value of Mean Square Error is 0 in the ratio between 0 to 1 (Lower the value better the model).).

The optimum value of R-Square value is 1 in the ratio between 0 to 1 (Higher the value better the model).

The optimum value of Mean Absolute Percentage Error is less than 10%.

The optimum value of Root Mean Square Error (RMSE) value is between 0.2 to 0.5. Lower value of RMSE indicates better fit.

*c.)Interpretation of the model(s).*

From all the models, we can interpret

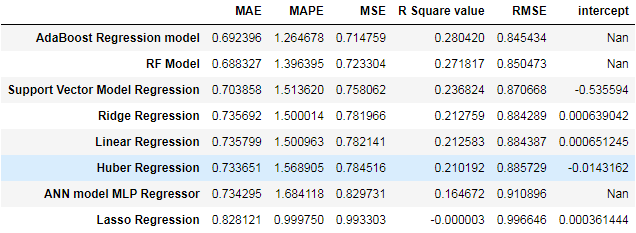


Fig 1.51 Sort Values based on best model performance metrics.

Adaboost Regression model is the best model among the other models from the model building.

###### **Model Tuning**

*a.Ensemble modelling, wherever applicable.*

### *Random Forest Regressor:*

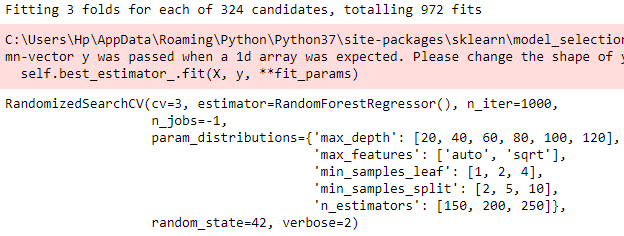


Fig 1.52 Intializing RF regressor model using gridsearch.



Fig 1.53 R-Square RF regressor model using gridsearch.

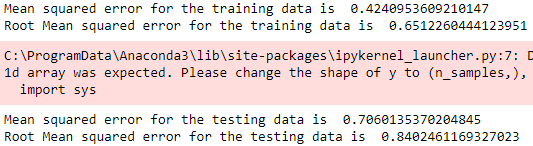


Fig 1.54 Mean Squared Error and Root Mean Squared Error RF regressor model using gridsearch.



Fig 1.55 Mean Absolute Error RF regressor model using gridsearch.



Fig 1.56 Mean Absolute Percentage Error RF regressor model using gridsearch.

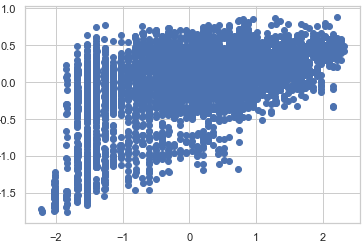


Fig 1.57 Scatterplot RF regressor model using gridsearch.

### *Gradient Boosting Regressor:*

Gradient boosting Regression **calculates the difference between the current prediction and the known correct target value**. This difference is called residual. After that Gradient boosting Regression trains a weak model that maps features to that residual.

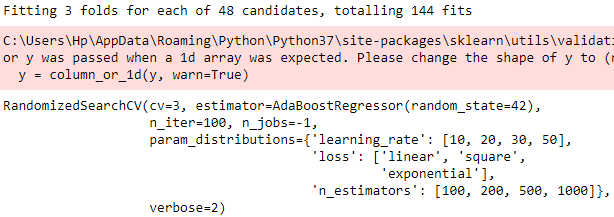


Fig 1.58 Initialising Gradient boosting regressor model using gridsearch.



Fig 1.59 R-Square Gradient boosting regressor model using gridsearch.

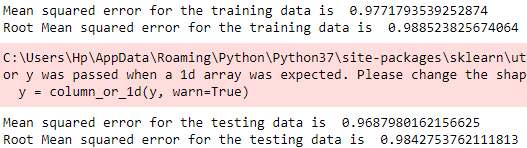


Fig 1.60 Mean Squared Error and Root Mean Squared Error Gradient boosting regressor model using gridsearch.



Fig 1.61 Mean Absolute Error Gradient boosting regressor model using gridsearch



Fig 1.62 Mean Absolute Percentage Error Gradient boosting regressor model using gridsearch

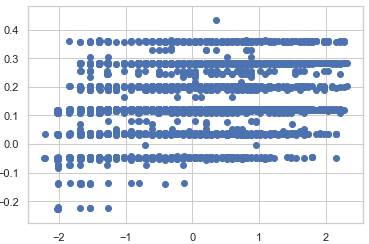


Fig 1.63 Scatterplot of Gradient boosting regressor model using gridsearch

### *Bagging Regressor:*

A Bagging regressor is an **ensemble meta-estimator that fits base regressor each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction.**

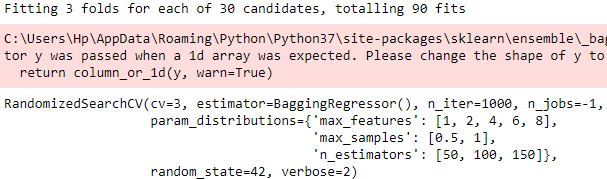


Fig 1.64 Initialising Bagging regressor model using gridsearch.



Fig 1.65 R-Square Bagging regressor model using gridsearch.

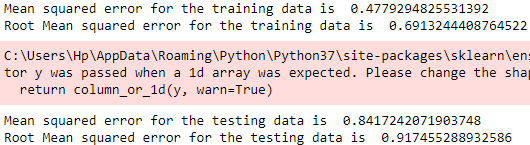


Fig 1.66 Mean Squared Error and Root Mean Squared Error of Bagging regressor model using gridsearch.



Fig 1.67 Mean Absolute Error of Bagging regressor model using gridsearch.



Fig 1.68 Mean Absolute Percentage Error of Bagging regressor model using gridsearch.

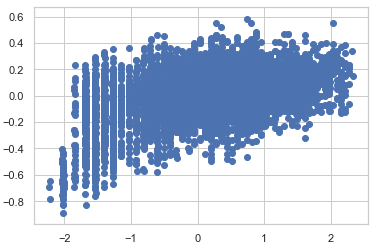


Fig 1.69 Scatterplot of Bagging regressor model using gridsearch.

### *Ridge Regression:*

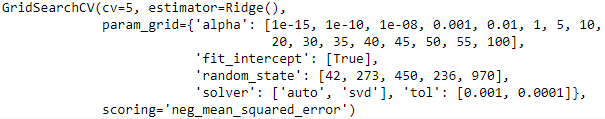


Fig 1.70 Initialising Ridge regressor model using gridsearch.



Fig 1.71 R-Square Ridge regressor model using gridsearch.





Fig 1.72 Mean Squared Error and Root Mean Squared Error of Ridge regressor model using gridsearch.



Fig 1.73 Mean Absolute Error of Ridge regressor model using gridsearch.



Fig 1.74 Mean Absolute Percentage Error of Ridge regressor model using gridsearch.

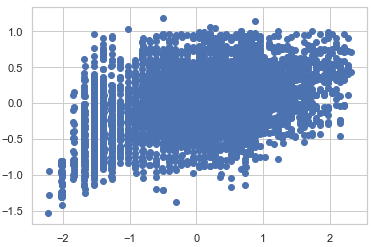


Fig 1.75 Scatterplot of Ridge regressor model using gridsearch.

*b. Any other model tuning measures(if applicable)*

There are 2 types of model tuning are

* Hyper parameter tuning based on grid search
* Hyper parameter tuning based on random search

*c. Interpretation of the most optimum model and its implication on the business*

The most optimum model is Random Forest Regressor using hyper parameters model tuning gives the best values among the other Hyper tuning model. The score of the model is the highest among all the models.

The errors are the least when compared to other models and all the performance metrics are within optimum limits.

So, after observing all these points we can conclude that the **Random Forest Regressor using hyper parameters model tuning** is the most optimum model.

### *Implications of regression models:*

The two main uses for regression in business are forecasting and optimization. In addition to helping managers predict such things as future demand for their products, regression analysis **helps fine-tune manufacturing and delivery processes.** In supply chain management, forecasting the optimum weight of the product and forecast the demand and supply of products based on the warehouse

Regression also helps in analysing past data on stocks prices and trends to identify patterns

Regression Analysis helps the business to understand the data points they have and lead to understand the relationships between dependent and independent variable and to make better decisions.

These can help in fine tuning the manufacturing and delivery process. These regression algorithms can also be used in optimizing the warehouse stock.

Regression models predicting consumer behaviour which can help in targeted marketing and product development

***Applications of Regression model in real time:***

1. **Medical field** - Medical researchers often use linear regression to understand the relationship between drug dosage and blood pressure of patients
2. **Aggriculture** - Agricultural scientists often use linear regression to measure the effect of fertilizer and water on crop yields