

4 DAYS BOOTCAMP AND 24 HOURS HACKATHON ON DATA SCIENCE

WITH PYTHON



Team members:

1. A V S Leelavathi

2. V Ramanamma

3. K Monika

4. P Puja

5. G Yamini

6. K Navya  
  
 Project uploading levels:  
 Level 1 ( Problem Statement )  
 Level 2 ( Feature Engineering )  
 Level 3 Correction  
 Level 4 ( Evaluation Metrics ) & Conclusion

PROBLEM STATEMENT

What is a problem statement ?

**[A problem statement is a description of an identified issue that a company or](https://www.bing.com/ck/a?!&&p=319422b71915de22JmltdHM9MTcwNjkxODQwMCZpZ3VpZD0yN2I1MDNiYS00ZDliLTZlN2MtMGFiNy0xMDIzNGMzZDZmOTImaW5zaWQ9NTg4OA&ptn=3&ver=2&hsh=3&fclid=27b503ba-4d9b-6e7c-0ab7-10234c3d6f92&psq=what+is+a+problem+statement&u=a1aHR0cHM6Ly91ay5pbmRlZWQuY29tL2NhcmVlci1hZHZpY2UvY2FyZWVyLWRldmVsb3BtZW50L3doYXQtaXMtYS1wcm9ibGVtLXN0YXRlbWVudA&ntb=1" \t "_blank)**

**[business has, detailing methods needed to address and overcome it](https://www.bing.com/ck/a?!&&p=319422b71915de22JmltdHM9MTcwNjkxODQwMCZpZ3VpZD0yN2I1MDNiYS00ZDliLTZlN2MtMGFiNy0xMDIzNGMzZDZmOTImaW5zaWQ9NTg4OA&ptn=3&ver=2&hsh=3&fclid=27b503ba-4d9b-6e7c-0ab7-10234c3d6f92&psq=what+is+a+problem+statement&u=a1aHR0cHM6Ly91ay5pbmRlZWQuY29tL2NhcmVlci1hZHZpY2UvY2FyZWVyLWRldmVsb3BtZW50L3doYXQtaXMtYS1wcm9ibGVtLXN0YXRlbWVudA&ntb=1" \t "_blank)**. It is a problem solving

[**identifies the gap between the current problem and goal**](https://www.bing.com/ck/a?!&&p=9c2853e2907f26a8JmltdHM9MTcwNjkxODQwMCZpZ3VpZD0yN2I1MDNiYS00ZDliLTZlN2MtMGFiNy0xMDIzNGMzZDZmOTImaW5zaWQ9NTg5Mg&ptn=3&ver=2&hsh=3&fclid=27b503ba-4d9b-6e7c-0ab7-10234c3d6f92&psq=what+is+a+problem+statement&u=a1aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnL3dpa2kvUHJvYmxlbV9zdGF0ZW1lbnQ&ntb=1).

How to write a problem statement ?

We can problem statement in five steps

1. Identify the issue
2. Begin with your ideal situation
3. Describe current gaps
4. State the consequent of the problem
5. Propose addressing the issue

## INDUSTRY : AUTOMOBILES

## DEPARTMENT : MANUFACTURING

## PROBLEM : SALES SLUMP

## CAUSE : CHANGES IN CUSTOMER EXPECTATIONS

## SOLUTION : ADAPTING TECHNOLOGIES

PROBLEM STATEMENT:

Predicting the sales slump in automobile industry while varies in customer

expectations we can predict this with the help of adapting technologies.

Automobile industry:

It is a broad range of companies and organizations involved in the design, development, manufacturing,

marketing, and selling of motor vehicles. It is one largest economic sectors in terms of revenue. Software and

Hardware engineers in the automotive sector work for automotive manufacturers.

Example : Sonalika Tractors

Sonalika Tractors faced a drop of 14.41% in retail tractor sales in India 2023. The

company sold 8,293 units in November 2023 compared to 9,689 units in November 2022.

TAFE Limited gained its market share by 0.09%. The company recorded a drop in retail

tractor sales by 20.65%.

References : Referred from google websites.

1.

<https://getjerry.com/car-repair/common-problems-with-honda>

2.

<https://www.google.com/search>? q=which+company+is+sales+down+in+tractor+industry&rlz=1C1CHBF\_enIN1081IN1082&oq=which+company+is+sales+down+in+tactors&gs\_lcrp=EgZjaHJvbWUqCQgBECEYChigATIGCAAQRRg5MgkIARAhGAoYoAEyCQgCECEYChigATIJCAMQIRgKGKABMgYIBBAhGAoyBwgFECEYnwUyBwgGECEYnwUyBwgHECEYnwUyBwgIECEYnwUyBwgJECEYnwXSAQkxOTU5NWowajeoAgCwAgA&sourceid=chrome&ie=UTF-8

FEATURES IN AUTOMOBILES INDUSTRY:

1. Development------------------------------------------------------------Development------------------------
2. Manufacturing--------------------------------------------Manufacturing--------
3. Marketing
4. Selling---------------------------------------------------------------------------------------x-----------------------
5. Organization involved in the design
6. Repairing
7. Quality-----------------------------------------------------------------------------------------------satisfaction rate---------------------
8. Modification of motor vehicles
9. Innovation
10. Reliability--------------------------------------------------------x-----------------------------
11. Performance
12. Cost\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Benifits\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
13. Interest Rate
14. Economic Growth
15. Safety and Aesthetics\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Benifits\_\_\_\_\_\_\_\_\_\_\_\_\_
16. Exchange Rate
17. Employment Rate
18. Car Sharing Apps+
19. Products and technology
20. Vehicle type--------------------------------------------------------Brand------------------------------------
21. Branding and marketing-------------------------------------------------------------------------- satisfaction rate------------------------
22. Globalized \production
23. Comfort-------------------------------------------------------------------------------------------------------------------------Benefit---------
24. Predictive maintenance\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Benifits\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
25. Autonomous driving
26. Technology--------------------------------------------------------------------------------- satisfaction rate-------------------------------

Performance-------------------\*\*\*\*\*\*\*

Quality------------------\*\*\*\*\*\*\*

Rating------------------\*\*\*\*\*\*\*\*

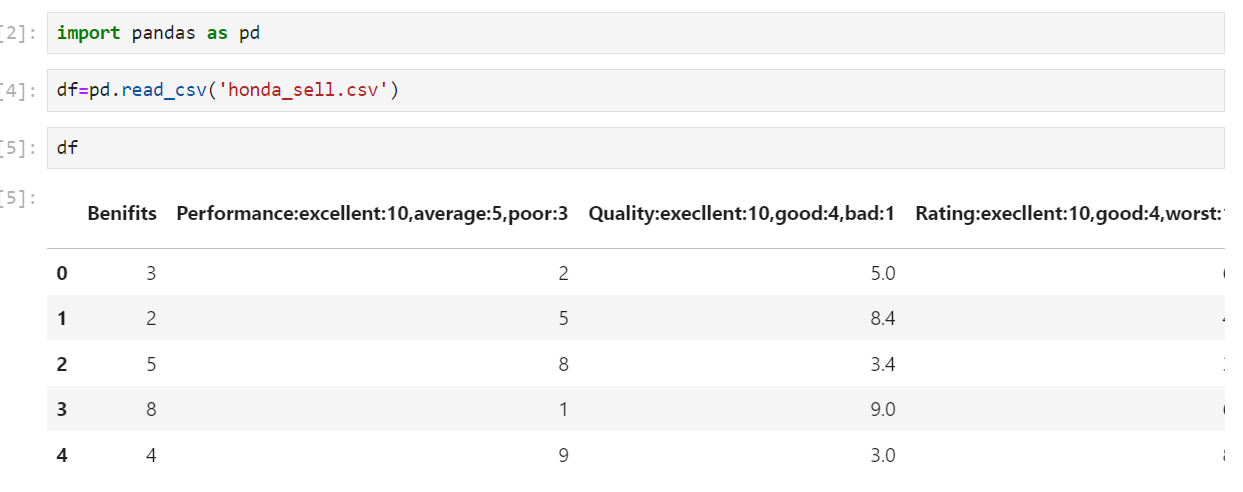
Satisfaction------------\*\*\*\*\*

Benefits------------\*\*\*\*\*\*\*

Sonalika Tractors Dataset

1).Import pandas as pd

You have imported the pandas library in Python With pandas, you can easily work with structured data, perform data manipulation, analysis, and visualization tasks. Not only pandas libray we can take any library for importing data.



In this above code we can use pandas library to import the data set. Here we can take df as variable of the data set.

1. To generate data to a specified range.



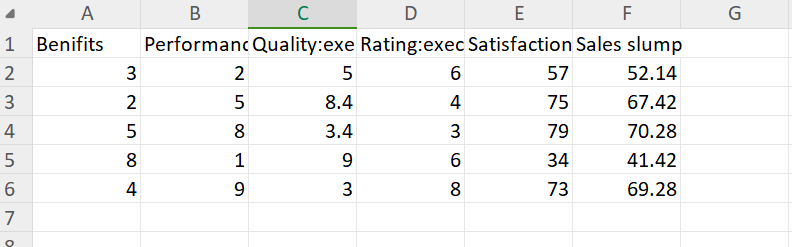
Random.randint() is a function from the Random module in Python that generates a random integer within a specified range

1. .Output for the above algorithm.



Here we can generate the dataset upto 1000 times.

3).Dataset



ALGORITHMS

Algorithm:

A set of finite rules or instructions to be followed in calculations or other problem-solving operations.In these algorithms we have some types which are…

REGRESSION:

Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables.In regression we have two types .

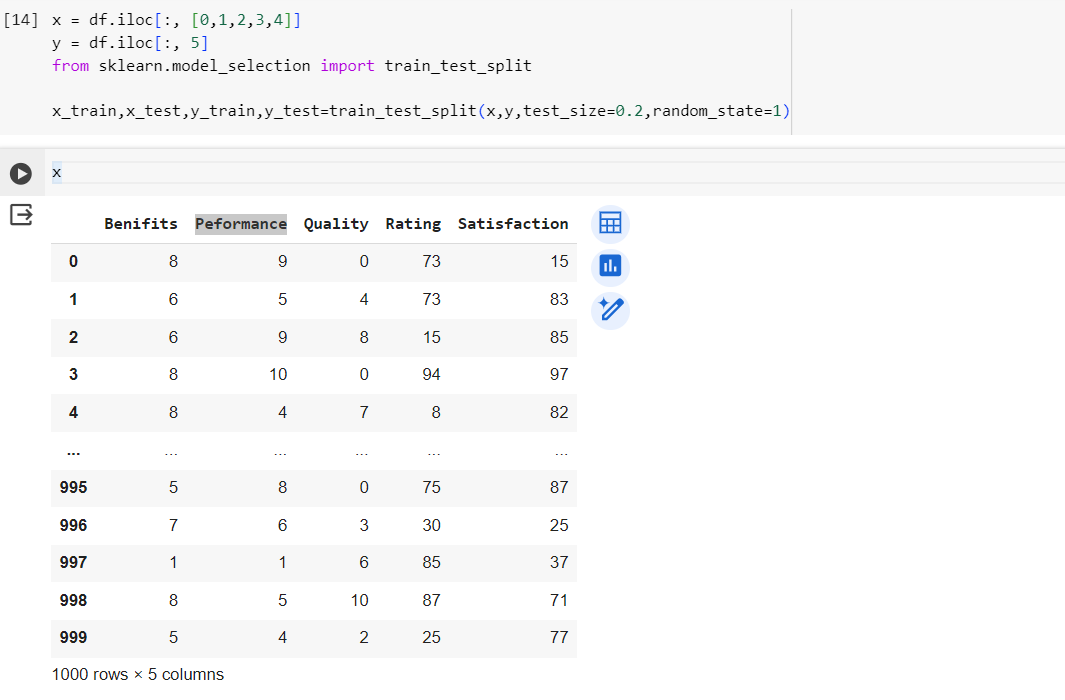
Linear regression :-

Linear regression is a basic and straightforward machine learning model used for regression analysis. It involves finding the relationship between two variables, i.e., the dependent variable and the independent variable. This model is used to predict a continuous outcome variable based on one or more predictor variables

MultipleLinear Regression:-

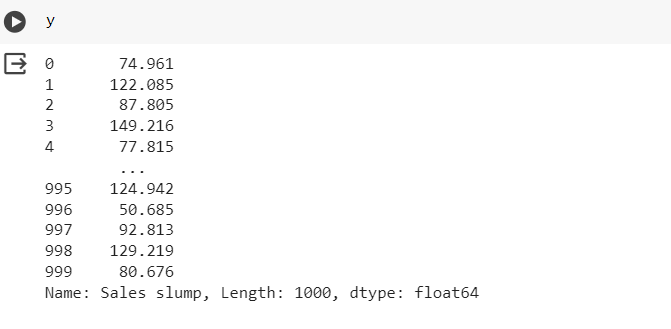
Multiple linear regression attempts to model the relationship between two or more independent variables and a dependent variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable

1).Asign values variables.



Here we can take two variables x and y ,then we print values of x.

2).Printing a value.



After that we can the values of y.

3).Linear regression by importing matplotlib.

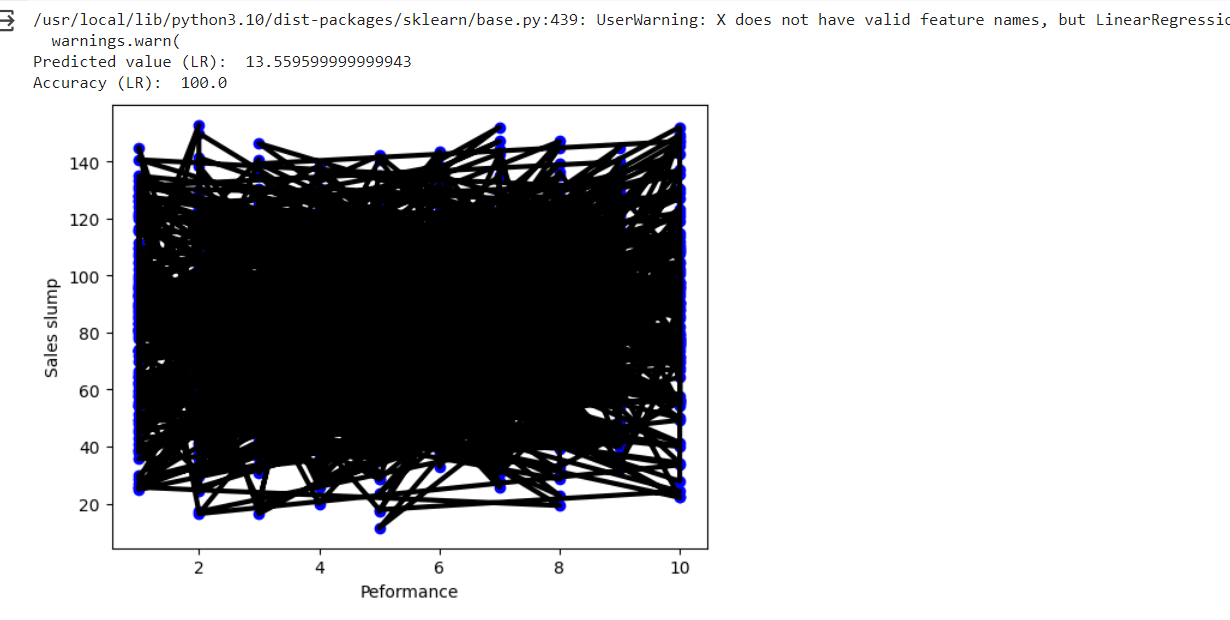


At the beginning we can import ‘matplotlib’ library as plt after that we use a

Scatter function to the attributes of x and y.Then we give linewidth, colour to

that axis.

Output:



It is the output of the above algorithm.

Decision Trees:-

A decision tree is a simple and intuitive model that is used for both

regression and classification problems. It is a tree-like structure where each node

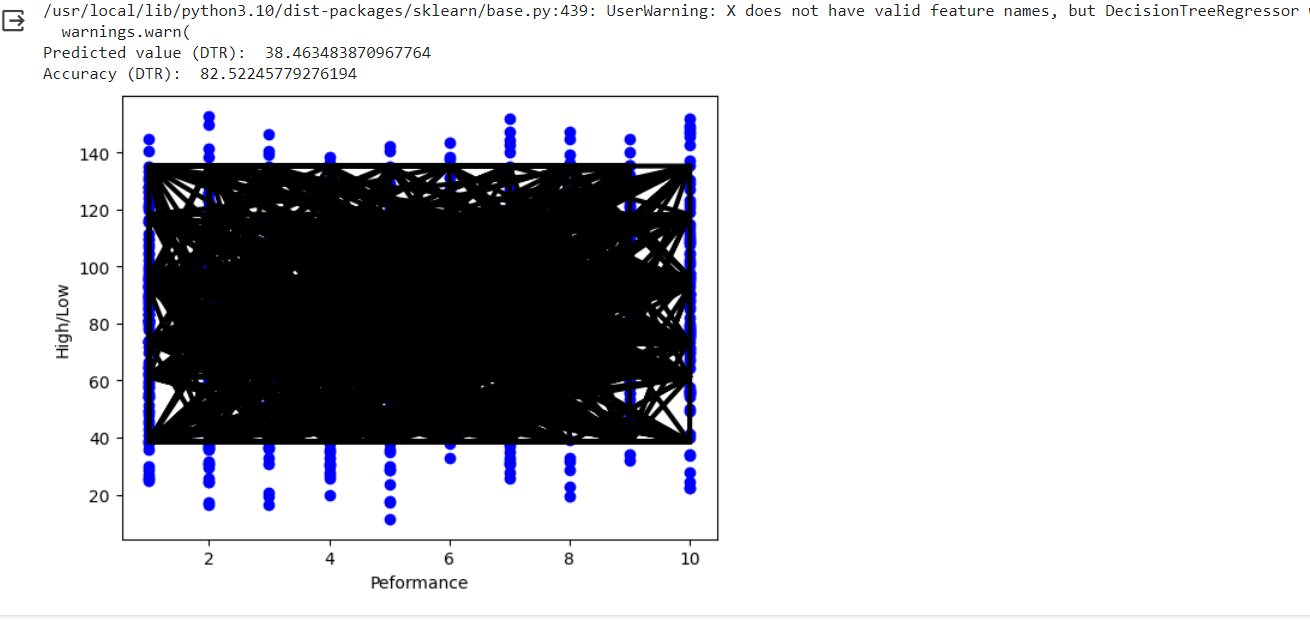
represents a feature or attribute, and each branch represents a decision rule.

Decision trees are particularly useful in situations where the data has multiple

variables and is non-linear



Output:



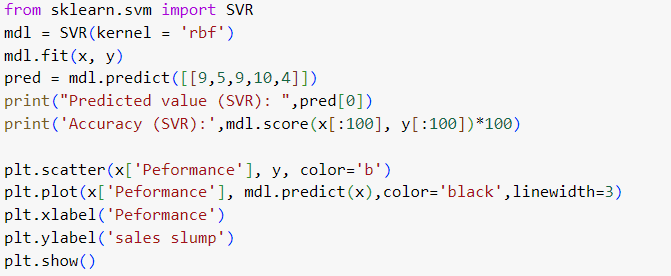
Support vector machines (SVM) :-

Support vector machines (SVM) are a supervised learning algorithm used for

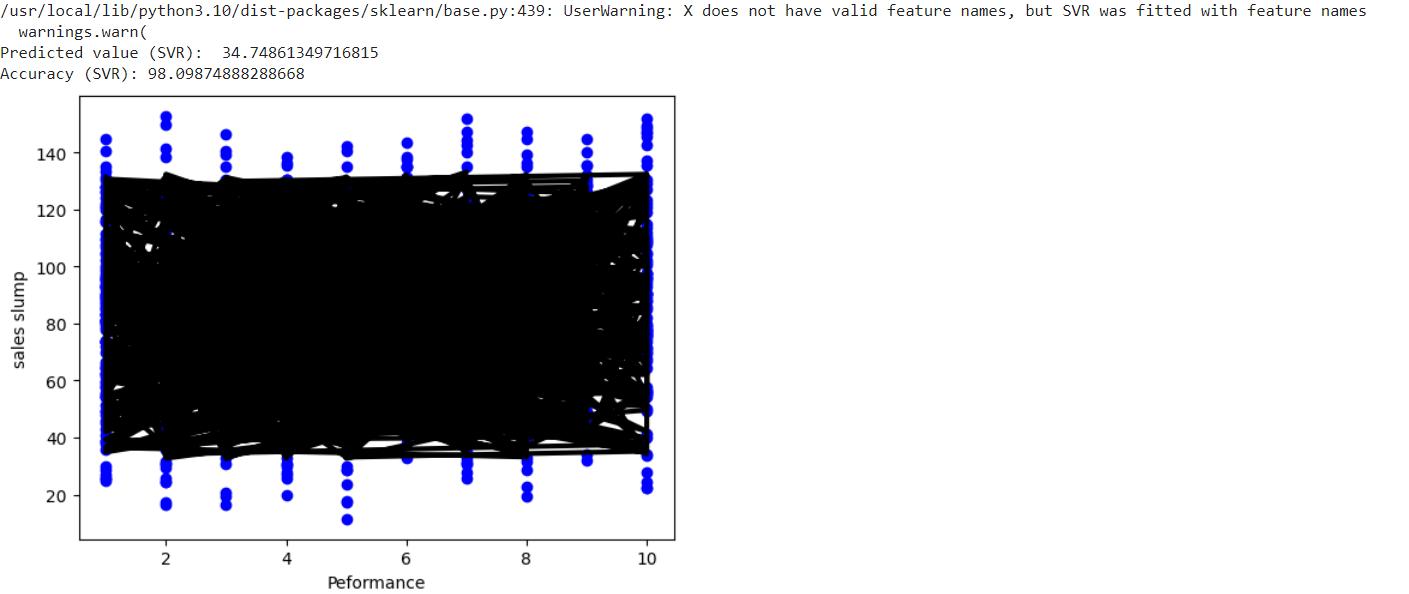
classification and regression analysis. SVM tries to find the best hyperplane that separates the

data points into different classes, with maximum margin. SVM can also handle non-linearly

separable data by transforming the data into a higher-dimensional space.



Output:



Random Forests:-

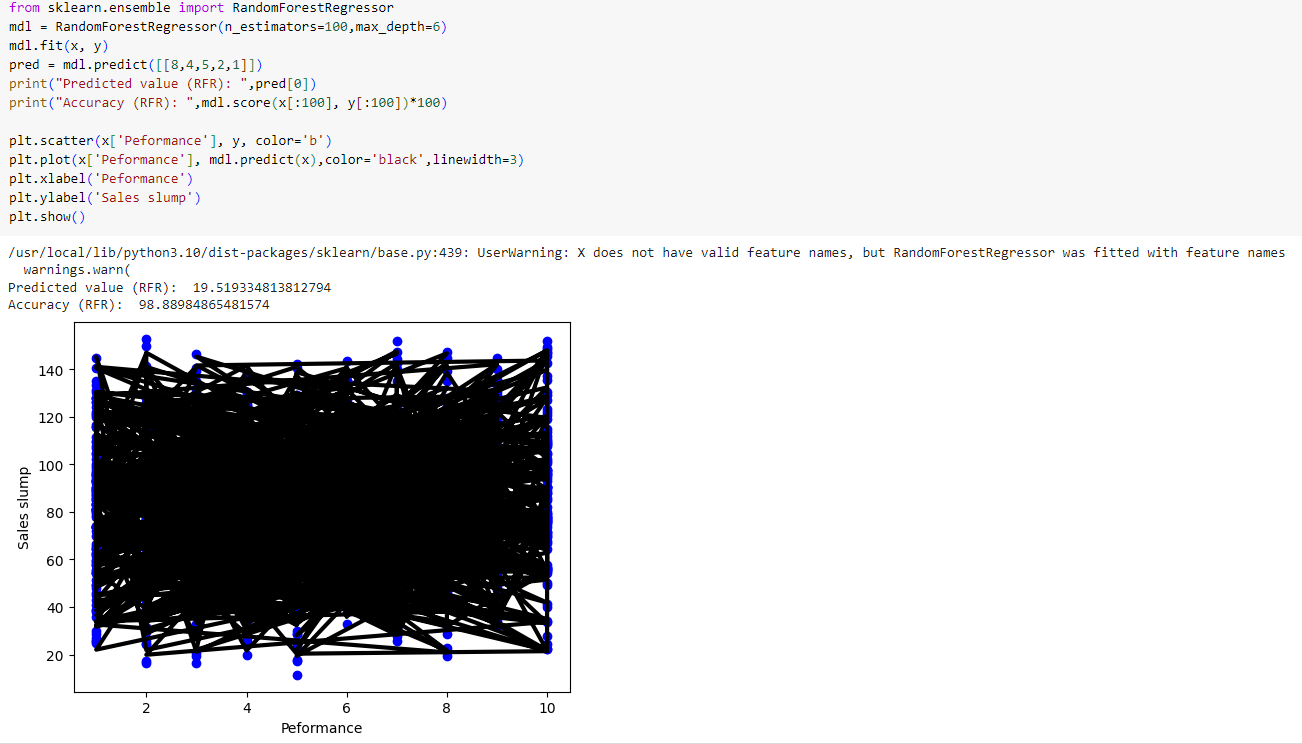
Random forests are a powerful and popular ensemble learning technique

used for classification, regression, and anomaly detection. It is an extension of

decision trees, where a large number of decision trees are trained on subsets of

data. The final prediction is made by taking the average of all the individual tree

predictions.



Evalution metrics

Evaluation metrics are used to evaluating machine learning models.We should know when to

use which metrics and it depends mainly on what kind of targets from sklearn import metrics.

Regression metrics:

Regression refers to predictive modeling problems that involve predicting a numeric value.It is

different from classification that involves predicting a class label. Unlike classification, you cannot

use classification accuracy to evaluate the predictions made by a regression model.

Mean squared error(MSE):

Mean Squared Error, or MSE for short, is a popular error metric for regression problems.It is

also an important loss function for algorithms fit or optimized using the least squares framing of a

regression problem. Here “least squares” refers to minimizing the mean squared error between

predictions and expected values.

MSE = 1 / N \* sum for i to N (y\_i – yhat\_i)^2

Root Mean Squared Error:

The Root Mean Squared Error, or RMSE, is an extension of the mean squared error.Importantly,

the square root of the error is calculated, which means that the units of the RMSE are the same as

the original units of the target value that is being predicted.

Mean Absolute Error:

Mean Absolute Error, or MAE, is a popular metric because, like RMSE, the units of the error

score match the units of the target value that is being predicted.Unlike the RMSE, the changes in

MAE are linear and therefore intuitive.

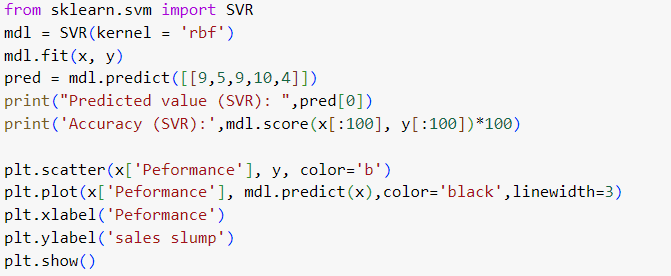
Support vector machines (SVM) :-

Support vector machines (SVM) are a supervised learning algorithm used for

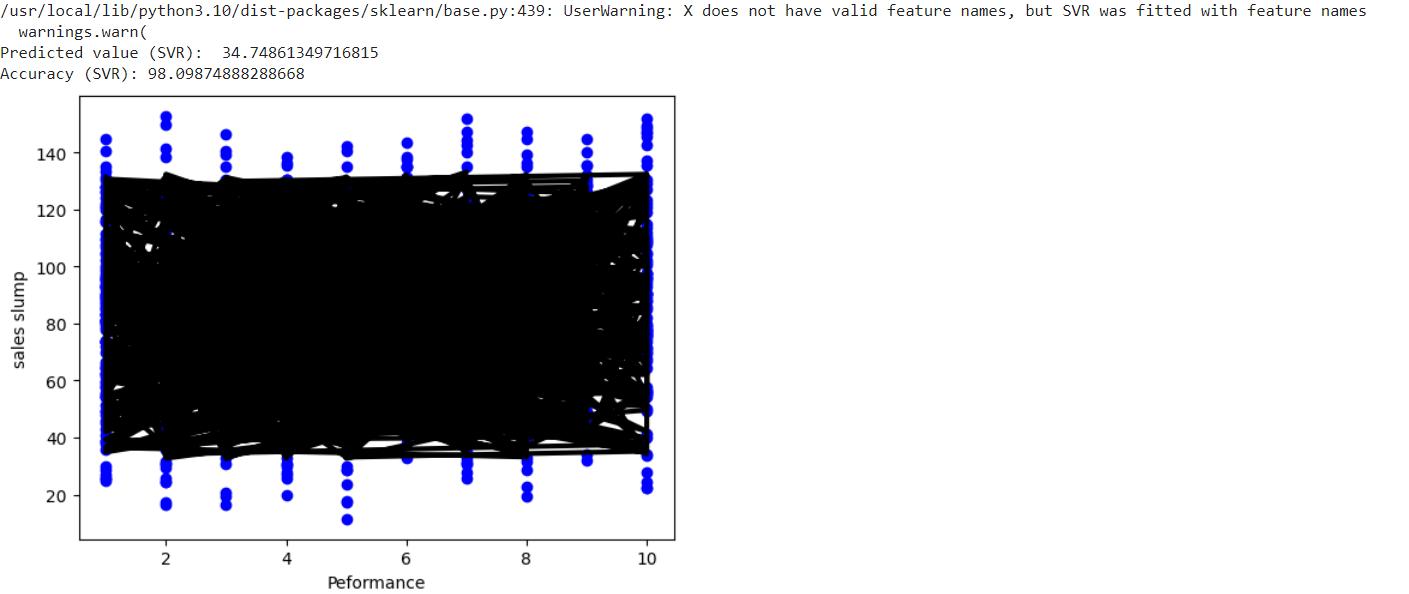
classification and regression analysis. SVM tries to find the best hyperplane that separates the

data points into different classes, with maximum margin. SVM can also handle non-linearly

separable data by transforming the data into a higher-dimensional space.



Output:



Decision Trees:-

A decision tree is a simple and intuitive model that is used for both

regression and classification problems. It is a tree-like structure where each node

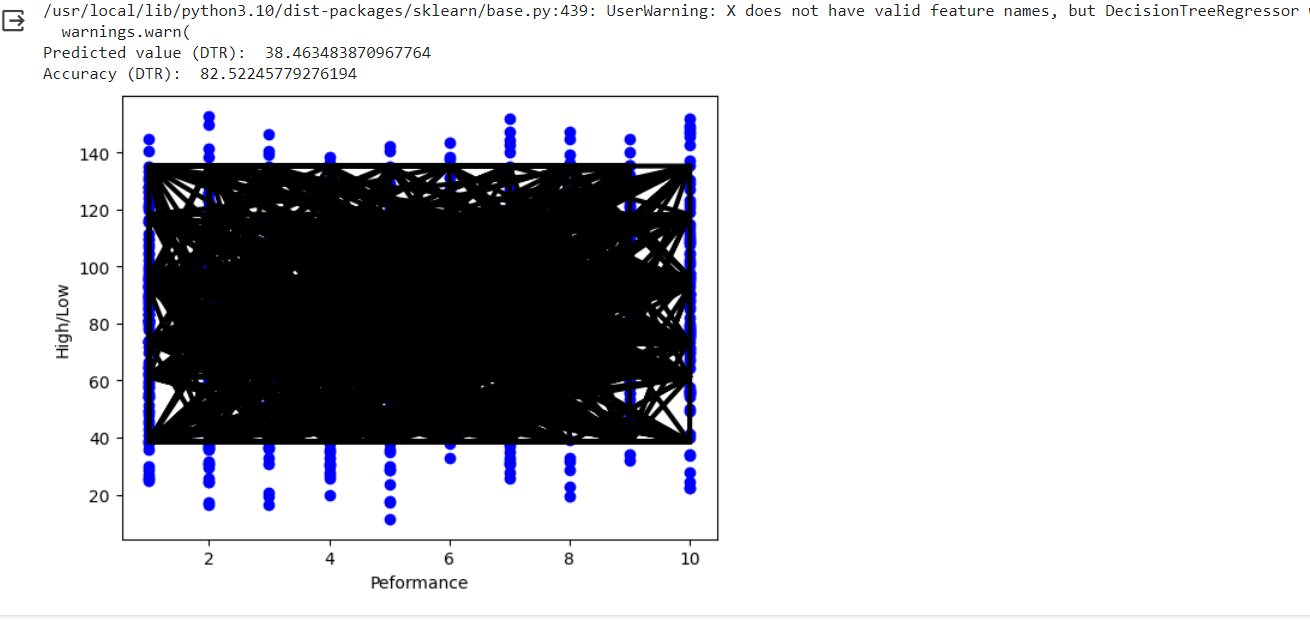
represents a feature or attribute, and each branch represents a decision rule.

Decision trees are particularly useful in situations where the data has multiple

variables and is non-linear



Output:



Random Forests:-

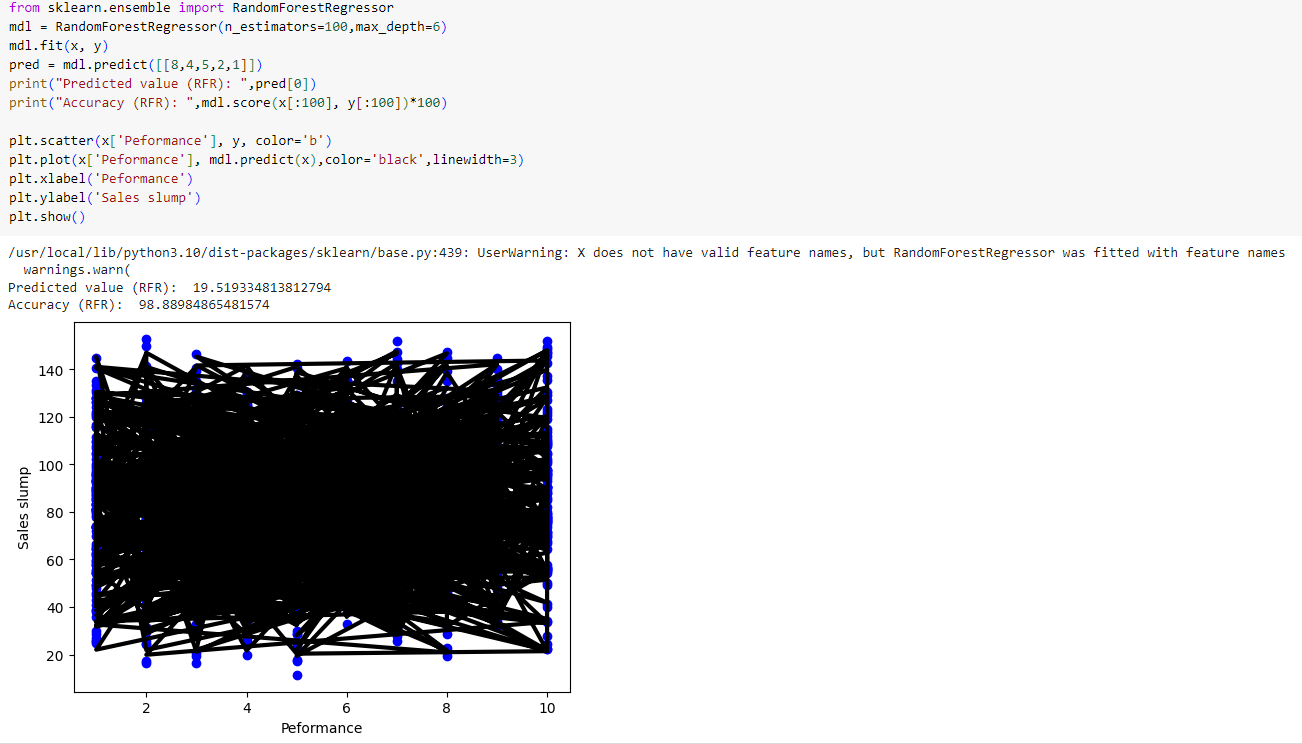
Random forests are a powerful and popular ensemble learning technique

used for classification, regression, and anomaly detection. It is an extension of

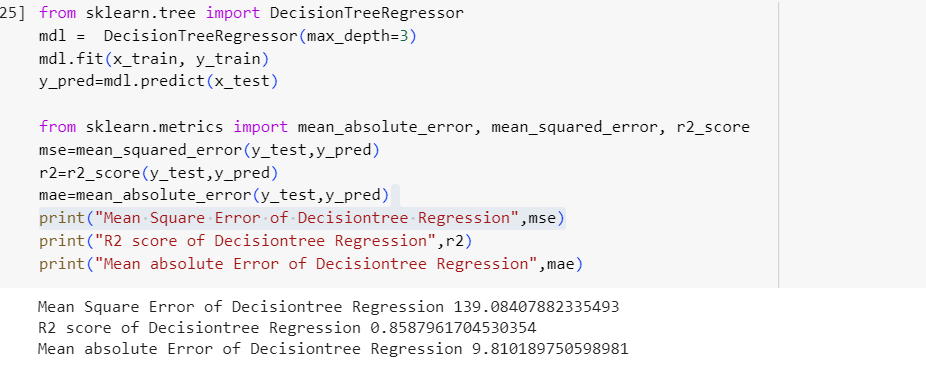
decision trees, where a large number of decision trees are trained on subsets of

data. The final prediction is made by taking the average of all the individual tree

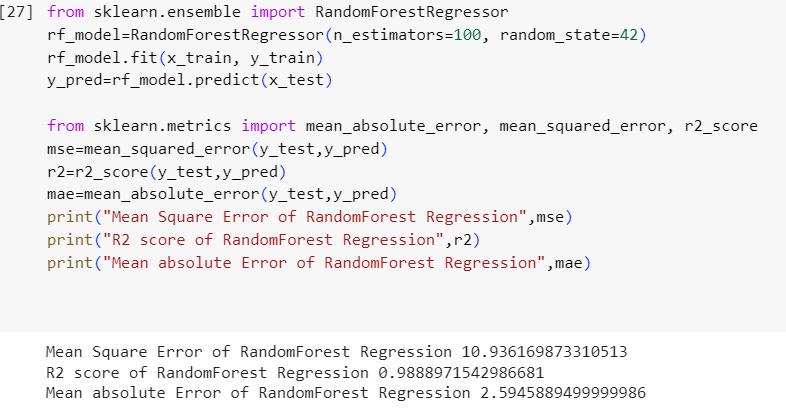
predictions.



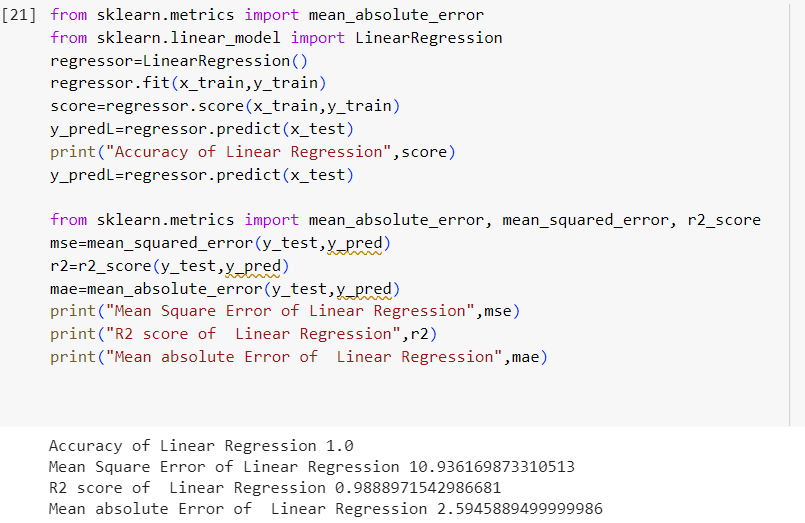
Decision tree error rating and accuracy.



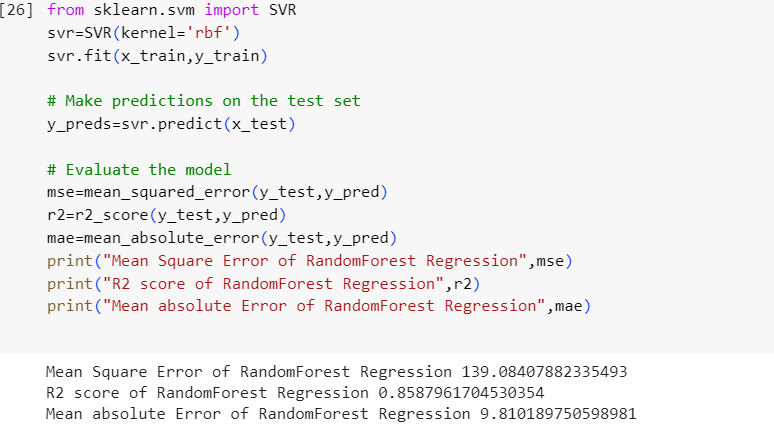
Random forest regression error rating and accuracy



Linear regression error rating and accuracy



SVR error rating and accuracy



Conclusion:

Finally our model is decision tree by comparing with other four regressions which has low

error rating. So, our model is Decision tree.