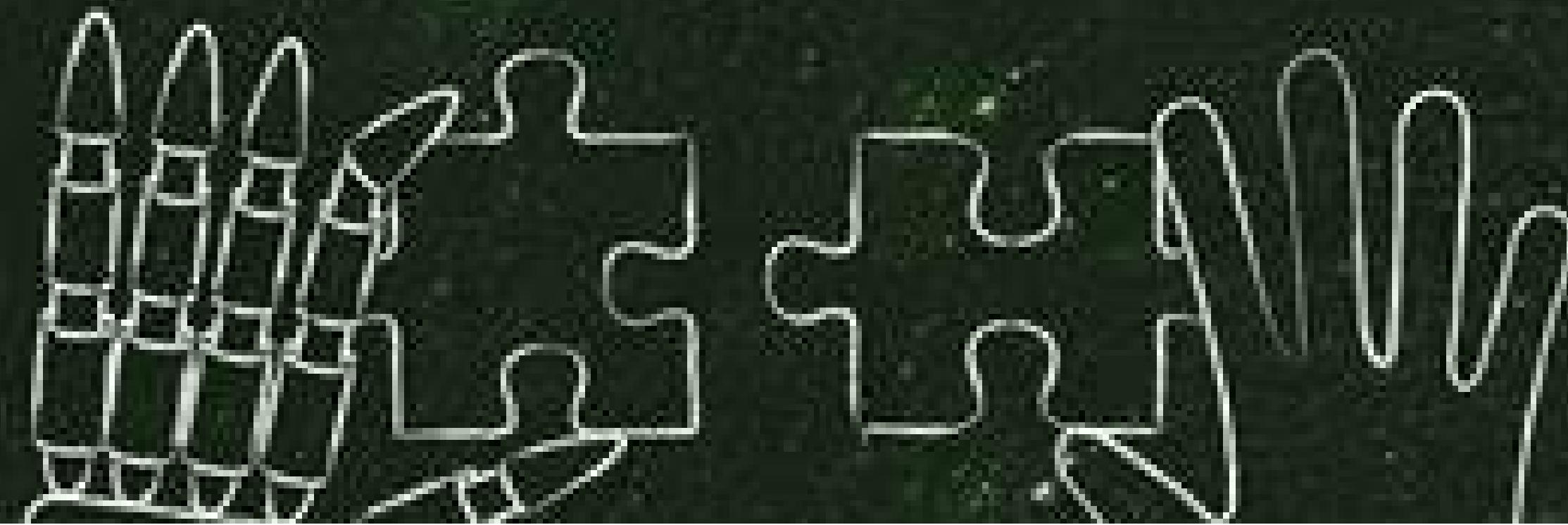
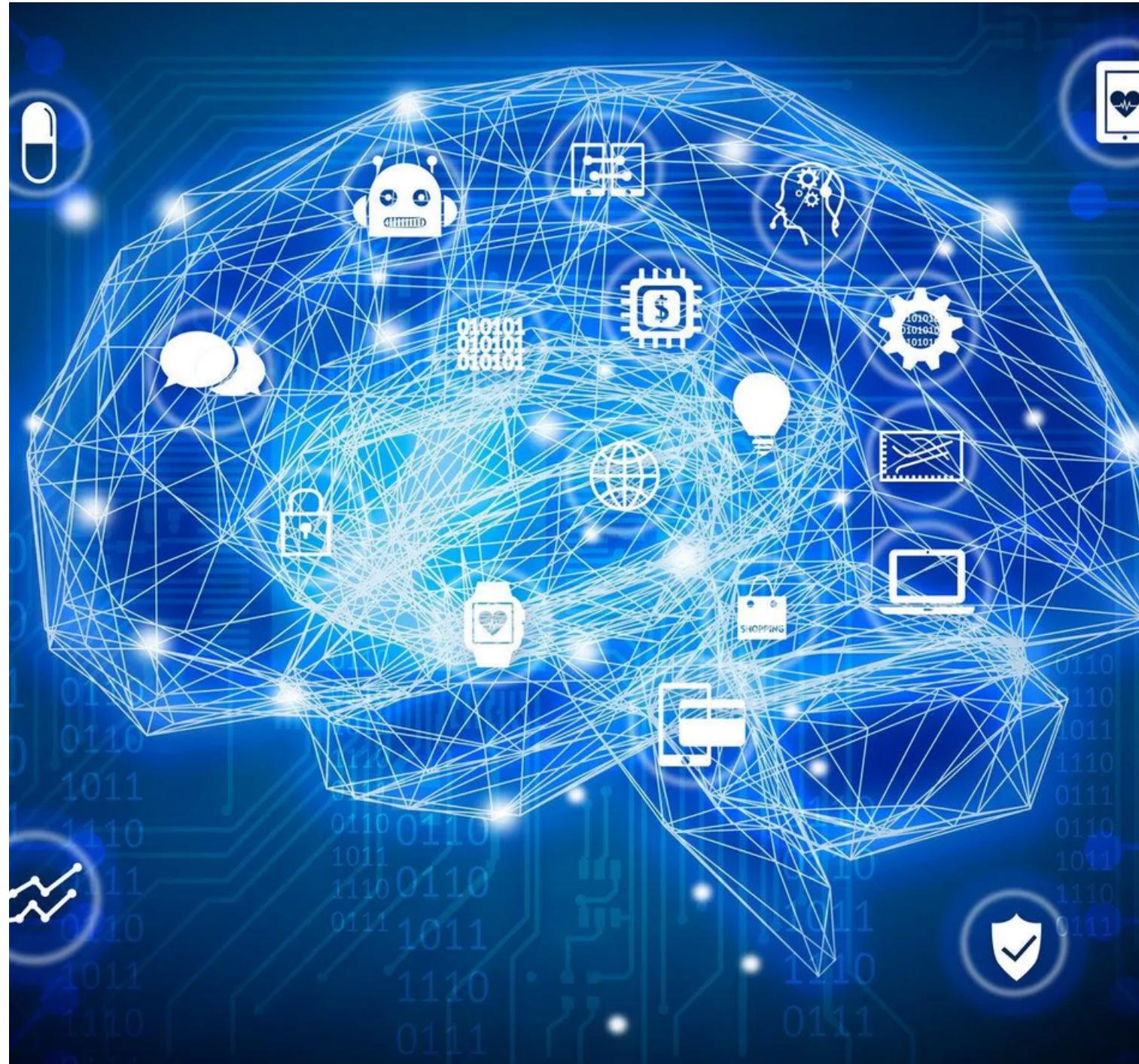


# MACHINE LEARNING



## Optimization of Shipping Charges and Fast Delivery



# Introduction

The aerospace industry is engaged in the research, development, and manufacture of flight vehicles

Ecommerce, also known as electronic commerce or internet commerce, refers to the buying and selling of goods or services using the internet, and the transfer of money and data to execute these transactions.

Main pain point in Ecommerce department is difficulty reaching customer service

# CONTENTS

- Feature Engineering
- Types of ML data
- About Dataset
- Algorithms
- Results

# Feature Engineering

- **Feature Creation**

Creating features involves creating new variables which will be most helpful for our model

- **Feature Transformation**

The transformation step of feature engineering involves adjusting the predictor variable to improve the accuracy and performance of the model.

- **Feature Extraction**

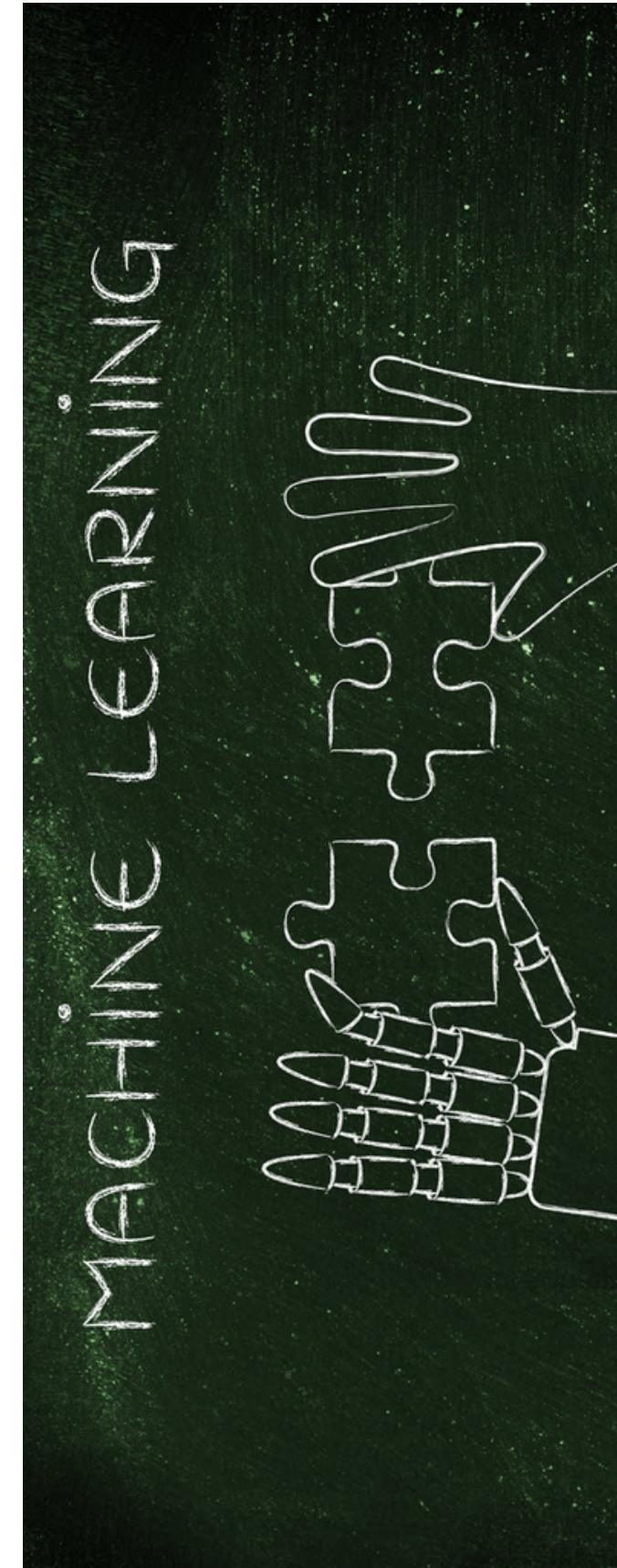
Feature extraction is an automated feature engineering process that generates new variables by extracting them from the raw data.

New variable generated using the raw data is outcome

- **Feature Selection**

Feature selection is a way of selecting the subset of the most relevant features from the original features set by removing the redundant, irrelevant, or noisy features

# Types of ML Data



1. Supervised
2. Semi Supervised
3. Unsupervised
4. Reinforcement

# ABOUT DATASET

Dataset contains the following features

- Distance
- Weight of goods
- Freight cost
- Shipping charges(Seaways)
- Shipping charges(Airways)
- No.of days(Seaways)
- No.of hours(Airways)
- Fast delivery
- Outcome

# Algorithms

There are two types of algorithms in supervised learning:

- **Classification**

Classification in machine learning is a supervised ML technique. It predicts group relationships for data instances in the dataset. It is a recursive process of recognizing, understanding and grouping the data objects using labels into pre-defined categories.

- **Regression**

Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables

# Results

## Classification:

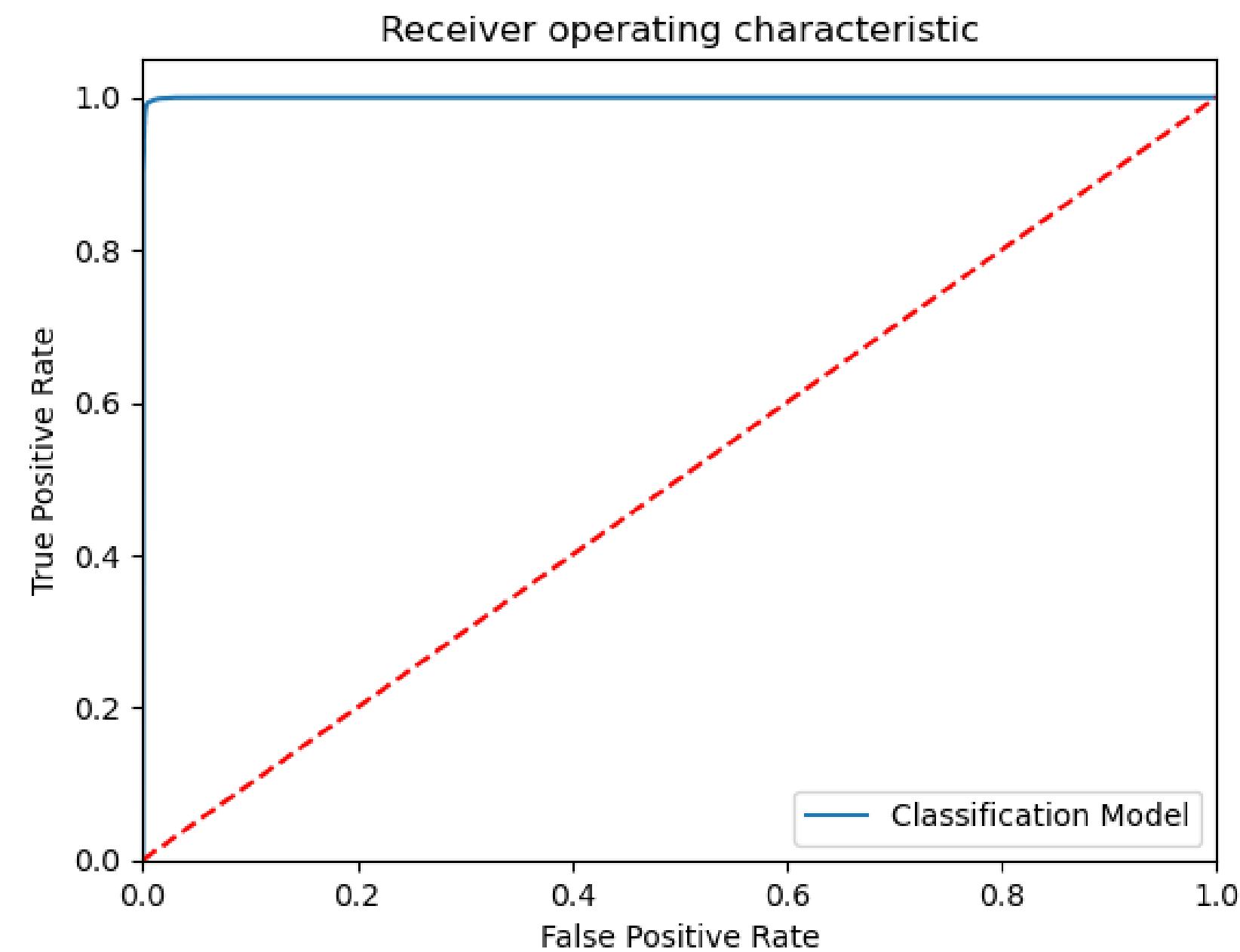
Here in classification of data we considered randomforest classifier with an accuracy of 99.7%

```
Model Name: RandomForestClassifier()
Confusion matrix :
[[26930  20]
 [ 72 2978]]
Outcome values :
26930 20 72 2978
Classification report :
precision    recall  f1-score   support

      1       1.00     1.00    1.00    26950
      0       0.99     0.98    0.98    3050

  accuracy         1.00      1.00    30000
macro avg       1.00     0.99    0.99    30000
weighted avg     1.00     1.00    1.00    30000

Accuracy : 99.7 %
Precision : 99.7 %
Recall : 99.9 %
F1 Score : 0.998
Specificity or True Negative Rate : 97.6 %
Balanced Accuracy : 98.8 %
roc_auc_score: 0.988
:
```



# RESULTS:

## REGRESSION:

Here in regression of data we considered Linear Regressor with an accuracy of 99.995%

**Model Name: LinearRegression()**

**Mean Absolute Error (MAE): 0.024**

**Mean Squared Error (MSE): 0.001**

**Root Mean Squared Error (RMSE): 0.028**

**R2\_score: 0.999995**

**Root Mean Squared Log Error (RMSLE): -3.559**

**Mean Absolute Percentage Error (MAPE): 0.05 %**

**Adj R Square: 0.999995**

