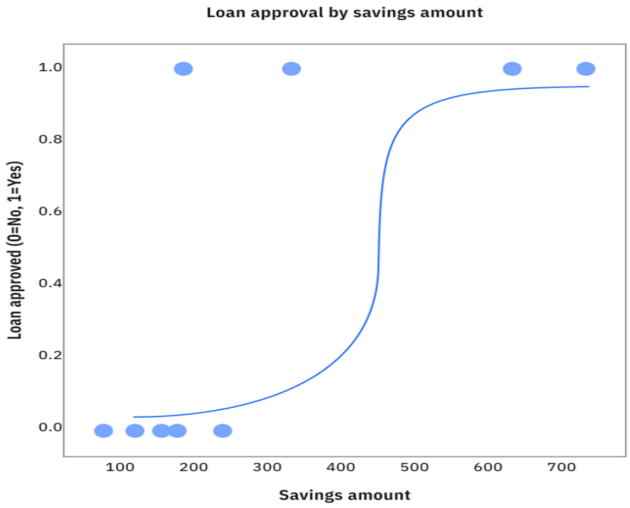
Experiment No. 3

Aim: To apply Logistic Regression for binary classification problems using machine learning, and assess model performance through appropriate evaluation metrics.

Software tools: Google Colab, Python Libraries(Pandas,Scikit-learn,Matplotlib,Seaborn)

Theory:

Logistic Regression is a statistical and machine learning technique used for binary classification tasks, where the target variable has two possible outcomes (e.g., Yes/No, 0/1, Pass/Fail).



It models the probability of the default class (usually 1) using the logistic **(sigmoid) function**:

$$P(y=1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

Key points:

Where:

- P(y=1 | x) : probability of class 1
- β0,β1 : model parameters
- e → base of natural logarithm

The output of Logistic Regression is a probability between 0 and 1, which is then thresholded (usually at 0.5) to decide the class

Key Matrics Derived from Confusion Matrix

1. **Accuracy** – Ratio of correctly predicted samples to total samples.
2. **Precision** – Proportion of positive predictions that are correct.
3. **Recall (Sensitivity)** – Proportion of actual positives correctly predicted.
4. **F1-Score** – Harmonic mean of precision and recall.
5. **ROC Curve & AUC** – Measure of model's discrimination ability.

One-Hot Encoding

One-hot encoding is a categorical data transformation technique used to convert non-numeric categorical features into numeric form so that they can be used in machine learning models.

- Each unique category is represented by a binary column (0 or 1).
- This avoids assigning ordinal relationships between categories that don't have a natural order.

Color	Red	Blue	Green
Red	1	0	0
Blue	0	1	0
Green	0	0	1

Applications

- Predicting sales based on advertising spend.
- Estimating house prices from area and location.
- Forecasting demand based on past data.
- Analyzing the effect of study time on exam scores.

Conclusion :

We successfully applied Logistic Regression to a binary classification problem.With one-hot encoding, categorical variables can be effectively used in the model.