Task 2

① Constructing the confusion matrix Date: . .

Total sample = 500
Predicted Positives = 100
Actual Positives = 50
True Positives (TP) = 45
False Negatives (FN) = 50 - 45 = 5

To find the no of false positives (FP) & true negatives (TN)

- FP = Predicted Positive - True Positives = 100 - 45 = 55
- TN = Total Samples - Actual Positives = 500 - 50 = 450
- TN = Total Negative - false Positives = 450 - 55 = 395

So the confusion matrix is

	Predicted Positive	Predicted Negative
Actual Positive	45	5
Actual Negative	55	395

② calculating Metrices:-

① Accuracy:-

$$\text{Accuracy} = \frac{TP + TN}{\text{Total Samples}} = \frac{45 + 395}{500} = \frac{440}{500}$$

$$= 0.88$$

So the accuracy is 88%.

Date: - -

② Precision = $\frac{TP}{TP+FP} = \frac{45}{45+55} = \frac{45}{100} = 0.45$

So the precision is 45%.

③ Recall = $\frac{TP}{TP+FN} = \frac{45}{45+5} = \frac{45}{50} = 0.90$

So the recall 90%.

④ F-1 score = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

$= 2 \times \frac{0.45 \times 0.90}{0.45 + 0.90} = 2 \times \frac{0.405}{1.35}$

$= \frac{0.81}{1.35} \approx 0.6$

So the f-1 score is 60%.

Date: - -

~~① underfitting:- when a model is too~~

Task 3

① overfitting:- overfitting occurs when a model becomes too complex & learns the training data specific patterns & noise. resulting in poor generalization to new unseen data.

② underfitting:- when a model is too simplistic to accurately learn from the data, leading to poor performance on both the training set & new data. It happens when the model doesn't capture the underlying trends & patterns in the data.

③ Bias = The difference b/w our actual & predicted values. Bias are the assumptions that our model makes about our data to be able to predict on new data.