Assignment-1

Business Understanding

Business Problem:

The business problem is the issue that needs to be solved to improve business performance or user experience.

Example (Spam Emails):

Company providing email services (e.g., Gmail, Outlook) notices that users receive too many spam emails, which leads to frustration, loss of productivity, and an increased risk of phishing attacks.

Business Objective:

The business objective is the goal that needs to be achieved to solve the business problem.

Example (Spam Emails):

The company is to develop a spam detection system that accurately classifies emails as **spam or not spam**

Business Constraints:

Business constraints are the limitations or challenges that must be considered when implementing the solution. These constraints could be related to technology, budget, time, regulatory compliance, or data availability.

Example (Spam Emails):

- **Computational Limitations:** The spam detection system must work efficiently without consuming excessive server resources.

- **Data Privacy Regulations:** The system must comply with data privacy laws like **GDPR** or **CCPA**, ensuring that user emails are not stored or misused.
- **Real-time Processing:** Emails must be classified in real time to avoid delays in inbox delivery.
- **User Experience:** The system should allow users to manually mark emails as spam or not spam to refine the filtering process.
- **Budget Constraints:** The company may have a limited budget for developing and maintaining the spam detection system.

Assignment-2

Balanced and unbalanced dataset:

Balanced Dataset :

A dataset is balanced when the number of instances (samples) in each class is roughly equal.

Ex: In Balanced dataset would have equal number of spam mails and not spam mails

spam- 50 out of 100

Not spam- 50 out of 100

Unbalanced dataset:

A dataset is **unbalanced** when one class has significantly more samples than another.

Ex: In Unbalanced dataset would not have equal number of spam mails and not spam mails

spam- 95 out of 100

Not spam- 5 out of 100

Assignment-3

Confusion Matrix:

A **confusion matrix** is a performance evaluation tool for classification models. It is a table that shows the number of correct and incorrect predictions, broken down by class.

Actual \ Predicted	Covid positive	Covid negative
Covid positive	**TP**	**FN**
Covid negative	**FP**	**TN**

Actual \ Predicted	Spam Mails	Not Spam Mails
Spam Mails	**TP**	**FN**
Not Spam Mails	**FP**	**TN**