**Business Problem:**  Predict future product demand to optimize stock levels.  
**Objective:**  Reduce overstocking (wasted inventory) and understocking (lost sales).

**Constraints:** Seasonal demand .

Supplier delays

Cost of storage

**Business Problem**: Disease Diagnosis

**Objective**: Detect diseases early

|  |
| --- |
|  |

|  |
| --- |
| **Constraints** |

: False negatives, regulatory compliance

**Confusion Matrix:** It compares **actual vs. predicted** values and helps understand model accuracy, precision, recall, and F1-score.

|  |  |  |
| --- | --- | --- |
| **Unlock phone using face lock** | **Owner** | **unknown** |
| **Owner** | **TP** | **FN** |
| **unknown** | **FP** | **TN** |

**TP-real owner unlocks phone**

**FN-UNKNOWN PERSON COULD NOT UNLOCK PHONE**

**FP-unknown person unlocks phone**

**TN-REAL OWNER CANNOT UNLOCK PHONE**

**Balanced dataset:** In machine learning, **a balanced dataset is one where the number of observations is roughly equal or similar for all classes in a classification dataset**. For example, if a dataset has two classes, a balanced dataset would have approximately 50% of the data belonging to each class.

**Imbalanced dataset:** an imbalanced dataset is characterized by a significant difference in the number of instances between the classes. In such datasets, one class (the majority class) has many more instances than the other class

**Movie Review Sentiment Analysis**

**Story:**  
A movie website wants to analyze user reviews to classify them as **Positive (1) or Negative (0)**.

**Dataset:**

* 50,000 movie reviews.
* 25,000 are **positive**
* 25,000 are **negative**
* **Balanced Dataset**

**Online Shopping Fraud Detection (Imbalanced Dataset)**

**Story:**  
An e-commerce company wants to detect fraudulent transactions.

**Dataset:**

* 1,000,000 transactions.
* **990,000 are normal purchases (0)**.
* **10,000 are fraudulent (1)**.
* **Highly Imbalanced Dataset** (fraud cases are only **1%**).