**Example 1: E-commerce (Amazon, Flipkart, etc.)**

**Business Problem**: High cart abandonment rate.

**Business Understanding**: Customers add items to their carts but leave without purchasing, leading to lost revenue.

**Business Constraints**:

Need for real-time customer behavior tracking

**Example 2: Banking and Finance (HDFC, Citibank, etc.)**

**Business Problem**: Increase in loan default rates.

**Business Understanding**: The bank is experiencing a rise in customers failing to repay loans, affecting profitability.

**Business Constraints**:

Delayed loan approval process affecting customer experience

**Example 3: Healthcare (Apollo, Fortis, etc.)**

**Business Problem**: Low patient retention rate.

**Business Understanding**: Patients visit once but do not return for follow-ups, reducing hospital revenue.

**Business Constraints**:

Limited doctor availability

**confusion matrix**

**Binary Classification (Fraud Detection)**

A bank deploys a model to detect fraudulent transactions.

| **Actual \ Predicted** | **Fraud (1)** | **Not Fraud (0)** |
| --- | --- | --- |
| **Fraud (1)** | 90 | 20 |
| **Not Fraud (0)** | 15 | 875 |

* **True Positives (TP) = 90** → Fraudulent transactions correctly detected as fraud.
* **False Negatives (FN) = 20** → Fraudulent transactions mistakenly classified as legitimate.
* **False Positives (FP) = 15** → Legitimate transactions wrongly flagged as fraud.
* **True Negatives (TN) = 875** → Legitimate transactions correctly classified.

**Balanced Data vs. Unbalanced Data**

**Balanced Data**: When all classes in the dataset have approximately the same number of samples.

**Unbalanced Data**: When one class has significantly more samples than the others, leading to an uneven distribution.

**Example:**

1. **Balanced Data:**
   * 5000 fraud transactions
   * 5000 non-fraud transactions
   * Both classes have equal representation.
2. **Unbalanced Data:**
   * 500 fraud transactions
   * 9500 non-fraud transactions