**Problem Statement**

We have to predict whether or not to give a loan. This involves analyzing various factors such as credit history, income, loan amount, and repayment ability to determine the likelihood of default.

**Banking**

**Asset** = Loan Product (Loan) - All the products which generate profit for the bank are called assets. These include:

1. **Housing Loan**: A loan taken to purchase or renovate a house.
   * Example: If a customer takes a housing loan of ₹50 lakh with an interest rate of 8% per annum for 20 years, the bank earns profit from the interest paid over time.
2. **Personal Loan**: A loan given for personal expenses such as medical bills, travel, or wedding costs.
   * Example: A borrower takes ₹5 lakh for wedding expenses and repays it with interest over 3 years. The bank benefits from the interest amount.
3. **Credit Card**: A short-term loan provided through a credit limit, where the user can make purchases and repay later.
   * Example: A person with a credit card limit of ₹1 lakh spends ₹50,000 and repays ₹55,000 after interest and charges. The bank earns ₹5,000 as profit.

**Liability** = Loss to Bank - These are products where banks have to pay interest to customers, which can be considered liabilities.

1. **Current Account**: A deposit account mainly used by businesses for frequent transactions. Banks provide various services but do not offer interest, making it a cost for the bank.
   * Example: A company maintains a current account with a balance of ₹10 lakh, but the bank earns no interest from it while providing services like cheque clearing and overdraft facilities.
2. **Saving Account**: A deposit account where individuals earn interest on their savings.
   * Example: A customer has ₹5 lakh in a savings account with 3.5% interest per annum. The bank has to pay ₹17,500 annually, which is a liability.
3. **Fixed Deposit (FD)**: A time-bound deposit where the bank pays a fixed interest rate to customers.
   * Example: A customer deposits ₹1 lakh in an FD for 5 years at 6% annual interest. The bank has to pay ₹6,000 every year as interest, making it a liability.
4. **Recurring Deposit (RD)**: A savings scheme where individuals deposit a fixed amount every month and earn interest.
   * Example: A person deposits ₹5,000 per month in an RD for 2 years at 5% interest. The bank pays cumulative interest at maturity, making it a liability.

**CASA (Current and Savings Account)**: These accounts provide liquidity to banks, but since they pay interest on savings accounts, they are considered liabilities.

**NPA (Non-Performing Assets)**

NPA = A loan that is not returned to the bank by the borrower (defaulted). If an EMI remains unpaid for more than 90 days, the loan is classified as an NPA.

1. **Disbursed Amount** = The total loan amount given to customers.
   * Example: If a bank grants a personal loan of ₹10 lakh to a borrower, the disbursed amount is ₹10 lakh.
2. **OSP (Outstanding Principal)** = The remaining loan amount that the borrower needs to repay.
   * Example: If a borrower takes a ₹1 lakh loan with an EMI of ₹8,000 per month and pays 2 months' EMI (₹16,000), the remaining amount will be ₹84,000, which is the Outstanding Principal (OSP).

**Non-Performing Assets (NPA) and Credit Risk in Banking**

**1. NPA (Non-Performing Asset)**

An NPA refers to loans or advances for which the principal or interest payment has remained overdue for a period of 90 days. When a borrower stops making payments, the loan is classified as an NPA. High NPA levels indicate financial distress for banks.

**Types of NPA:**

1. **Gross NPA (GNPA):**
   * The total value of non-performing loans in a bank's portfolio before provisioning for potential losses.
   * It represents the absolute amount of defaulted loans.
2. **Net NPA (NNPA):**
   * Net NPA is calculated by deducting provisions (funds set aside for potential loan losses) from the gross NPA.
   * It represents the actual risk exposure of the bank after adjustments.
3. **NPA (3 to 5 days) – Default:**
   * If a borrower fails to make payments for 3 to 5 days, the loan is marked as a default risk, but it has not yet become a full-fledged NPA.
4. **NPA (0.01 to 0.06) – Transition:**
   * Loans in transition indicate an early-stage default where the probability of turning into an NPA is low but needs monitoring.
5. **C-NPA values:**
   * This represents category-wise classification of NPAs for better risk assessment.

**2. DPD (Days Past Due)**

DPP is a measure of how many days have passed since the loan repayment was due but unpaid.

**Key points:**

* **DPP = Days past due**
  + It indicates the number of days a borrower has failed to make a scheduled payment.
* **At the due date, the loan is compliant and ideal EMI (Equated Monthly Installment):**
  + If the borrower pays on time, the loan remains compliant, and there is no issue.
* **DPP ideally = Zero:**
  + A well-managed loan portfolio should have zero overdue payments.
* **If DPP > 70 → Defaulted DPP:**
  + A borrower who has not paid for over 70 days is classified as defaulted.

**3. PAR (Portfolio at Risk)**

PAR is an indicator of the total loan amount that is at risk due to overdue payments.

**Key points:**

* **PAR = Portfolio at risk**
  + The total value of outstanding loans where DPP > 70.
  + It helps banks evaluate their exposure to potential losses.
* **OSP (Outstanding Principal) when DPP > 70:**
  + The principal amount of the loan still unpaid when the borrower defaults.
* **OSP should be zero at the end of the loan cycle:**
  + Ideally, all loans should be repaid fully without any default at the end of the tenure.

**NPA Impact on Banks**

1. **Loan Portfolio Quality:**
   * A high NPA ratio indicates poor asset quality.
   * A lower NPA ratio means healthier financial conditions.
2. **Market Sentiment & Stock Price:**
   * A bank with fewer NPAs gains investor confidence.
   * Lower NPAs lead to a better credit rating and higher stock prices.

**4. Credit Risk Types in Banking**

Banks classify loans based on overdue duration to manage risks effectively.

**Loan Classification by DPP:**

1. **DPP (Zero):**
   * Loans are being paid on time, and there is no overdue amount.
   * Considered a **non-default account.**
2. **DPP (1 to 30 days) – SMA 1 (Standard Monitoring Account 1):**
   * The loan has missed 1 to 30 days of payment.
   * Considered a low-risk category but requires monitoring.
3. **DPP (31 to 60 days) – SMA 2 (Standard Monitoring Account 2):**
   * Payments are overdue for 31 to 60 days.
   * The risk is increasing, and the borrower needs to be contacted for repayment.
4. **DPP (61 to 90 days) – SMA 3 (Standard Monitoring Account 3):**
   * Payments are overdue for 61 to 90 days.
   * There is a high probability of default.
5. **DPP (96 to 180 days) – NPA:**
   * The loan is classified as a **Non-Performing Asset (NPA)** if it crosses 90 days.
   * This means the bank is now at financial risk.
6. **DPP (>180 days) – Write-off (Loan not present):**
   * If a loan remains unpaid for more than 180 days, it is often written off as a loss.
   * The bank may attempt recovery through legal means, but the amount is no longer considered part of its active assets.

**5. Categorical Features in Machine Learning**

Categorical features are variables that take on a limited number of distinct values.

**Key Concept:**

* If a categorical variable is associated with another categorical outcome, a **Chi-Square test** is used to determine if there is a significant relationship between them.
* Example:
  + "Loan Type" (Secured/Unsecured) vs. "Loan Default Status" (Yes/No).
  + Chi-Square test will help determine if there is a dependency between the loan type and default risk.

**6. Hypothesis Testing**

Hypothesis testing is a statistical method used to validate assumptions.

**Key Points:**

* **Inferential statistics:**
  + Used to make generalizations about a population based on sample data.
* **H₀ (Null Hypothesis):**
  + Assumes no significant difference or effect in the dataset.
  + Example: "There is no relationship between loan amount and default probability."
* **H₁ (Alternate Hypothesis):**
  + Suggests a significant effect or difference.
  + Example: "Higher loan amounts increase the probability of default."

**7. Project Concept: Trade Line (TL) & Multiclass Classification**

A **Trade Line (TL)** represents a credit account (loan) reported to credit bureaus.

**Classification in Loan Risk Management:**

* **P1, P2, P3, P4 → Multiclass classification:**
  + Instead of a binary classification (Good/Bad loan), loans can be classified into multiple risk categories (e.g., Low, Medium, High, Default).
  + This helps banks and financial institutions in risk prediction and decision-making.

**Conclusion**

Understanding NPA and credit risk management is crucial for banks and financial institutions. Key takeaways:

* Banks classify loans based on **Days Past Due (DPP)** to monitor risks.
* **Non-Performing Assets (NPA)** impact financial health and market sentiment.
* **Portfolio at Risk (PAR)** is an important measure for assessing loan defaults.
* **Hypothesis testing** and **categorical feature analysis** are useful in analyzing financial data.
* **Machine learning techniques like multiclass classification** can help in better risk assessment and loan approval decisions.