IE 6700: DATA MANAGEMENT FOR ANALYTICS

USE CASE: MILESTONE #1

Enhancing customer understanding and Service Improvement for a Banking System

---

GROUP NO. 12

GROUP MEMBERS:

MALLIKA GAIKWAD

SHARVARI PRAVIN DESHPANDE

PROBLEM STATEMENT:

Background:

PKDD'99 financial dataset or the Berka Dataset (which is the dataset being used for this project) provides information on the clients, accounts, and transactions of a Czech bank. The Bank acknowledges the critical importance of providing tailored services to its clients. However, the bank is currently facing challenges in accurately identifying high-value clients for targeted offerings and assessing potential risks associated with certain clients. The bank possesses a wealth of data pertaining to its customers, including account transactions over several months, loan histories and issues credit cards. Management recognizes the need for a comprehensive and actionable approach to leverage this data effectively.

Objective:

The primary objective of this project is to empower the banking system with precise insights into customer behavior, preferences, and risk profiles. By utilizing advanced analytical techniques, the bank aims to refine its understanding of clients and formulate specific strategies for service improvement. The bank seeks a robust and well-structured approach to ensure convincing results.

Scope:

Upon completion of this project, we wish to achieve the following components:

1. Data Analysis: Analyse customer data, including account transactions, loan histories and credit card information, from relevant sources within the bank.

2. Customer Segmentation: Employ advanced techniques to categorize customers based on spending patterns, account activity, credit utilization and loan repayment behavior.
3. Risk Assessment: Develop a comprehensive risk assessment model to identify potential high-risk clients based on historical data, identifying characteristics of clients with higher default probabilities.
4. Client profiling: Create detailed customer profiles, including demographic information, financial behavior, and transactional history, to gain a comprehensive understanding of individual clients.
5. Service customization recommendations: Based on customer segments and risk assessments, formulate specific recommendations for tailoring services to meet the unique needs of different client groups.
6. Performance metrics: Establish key performance indicators (KPIs) to measure the impact of service improvements on client satisfaction, retention, and overall profitability.

DATASET DETAILS:

The Berka dataset is a collection of financial information from a Czech bank. The dataset deals with over 5,300 bank clients with approximately 1,000,000 transactions. Additionally, the bank represented in the dataset has extended close to 700 loans and issued 900 credit cards, all of which are represented in the data.

- Number of tables: 8
- Count of rows: 1,090,086
- Count of columns: 55
- Missing values: Yes

ENTITIES AND THEIR RELATIONSHIPS:

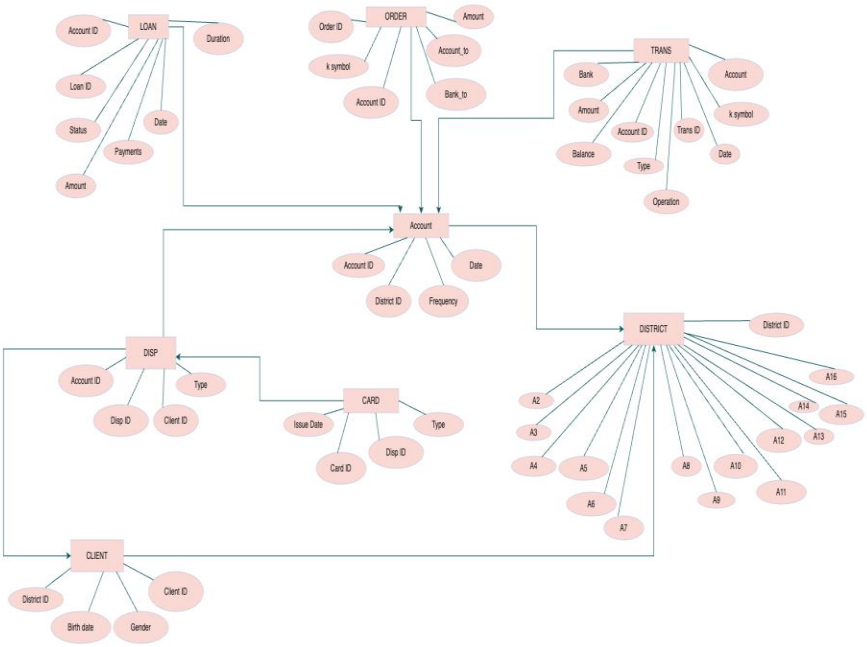
1. Accounts: Each record describes static characteristics of an account
2. Clients: Each record describes characteristics of a client
3. Disposition (Disp): Each record relates together a client with an account i.e., this relation describes the rights of clients to operate accounts
4. Permanent orders, debits only (Order): Each record describes characteristics of a payment order
5. Transactions (Trans): Each record describes one transaction on an account
6. Loans: Each record describes a loan granted for a given account
7. Credit cards (Card): Each record describes a credit card issued to an account
8. Demographic data (District): Each record describes demographic characteristics of a district

ENTITY-RELATIONSHIP DESCRIPTION:

- Each account has both static characteristics (e.g., date of creation, address of the branch) given in "account" and dynamic characteristics (e.g., payments debited or credited, balances) given in "permanent order" and "transaction".

- "client" describes characteristics of persons who can manipulate with the accounts.
- One client can have more accounts, more clients can manipulate with a single account; clients and accounts are related together in "disposition".
- "loan" and "credit card" describe some services which the bank offers to its clients
- More than one credit card can be issued to an account
- At most one loan can be granted for an account.
- "Demographic data" gives some publicly available information about the districts (e.g., the unemployment rate); additional information about the clients can be deduced from this.

ENHANCED ENTITY-RELATIONSHIP DIAGRAM (EER):



REFERENCE AND TRANSACTIONAL DATA:

1. **Reference Data:**

- ***Client information:*** This includes details like names, addresses, contact numbers, social security numbers, etc.
- ***Account Types:*** information about the types of accounts offered by the bank, such as savings, checking, fixed deposit, etc.

- ***Product Catalog****:* Details about the various products and services offered by the bank, including loans, credit cards, etc.
- ***Branch Information****:* Details about the different branches, such as location, branch code, operating hours, etc.

2. Transactional Data:

- ***Account Transactions****:* Details of deposits, withdrawals, transfers, and other activities that occur in customer accounts.
- ***Loan Transactions****:* Information about loan disbursements, repayments, interest accruals, etc.
- ***Credit Card Transactions****:* Records of purchases, payments, credit limits, etc.
- ***Customer Inquiries and Interactions****:* Logs of customer service interactions, queries, complaints, etc.
