# **Personalized Investment Recommendation System for Enhanced Customer Financial Well-being**

**Abstract**

This project aims to develop a recommendation system that advises bank account holders on suitable investment products based on their financial profiles, transaction histories, and investment goals. The goal is to enhance customer satisfaction by providing personalized investment recommendations, thereby increasing customer loyalty and uptake of the bank's investment products.

## Introduction

In the competitive banking industry, providing personalized services is crucial for customer retention and satisfaction. This project focuses on developing a recommendation system that tailors investment advice to individual customers based on their unique financial situations. By leveraging data-driven insights, the bank can offer more relevant and appealing investment options, thereby fostering stronger customer relationships and driving higher product engagement.

## Data Collection and Preprocessing

### **Description of the Dataset**

The dataset includes the following features:

* Age
* Marital Status
* Education Level
* Occupation
* Income Level
* Location
* Monthly Average Balance
* Number Of Transactions Per Month
* Existing Bank Products
* Investment Horizon

### **Steps Taken to Clean and Preprocess the Data**

1. **Drop Duplicates**: Removed duplicate records to ensure data quality.
2. **Treat Outliers**: Used box plots to identify and handle outliers.
3. **Impute Missing Values**: Applied KNN imputer and SimpleImputer to fill in missing values.
4. **Encode Categorical Features**: Utilized ordinal encoding for categorical variables.

## Exploratory Data Analysis (EDA)

### **Key Insights from the Data**

* The dataset was nearly balanced across various categories.
* Outliers were present and handled appropriately.

### **Visualization Techniques Used**

* Box plots were employed to identify and visualize outliers in the data.

## Clustering Methodology

### **Explanation of the Clustering Algorithm(s) Used**

* **KMeans**: Applied KMeans clustering algorithm, which partitioned the data into distinct clusters.
* **DBScan**: Also tried DBScan, but KMeans provided better results.

### **Reasons for Choosing the Specific Algorithm(s)**

KMeans was chosen due to its simplicity and effectiveness in handling large datasets. It also allowed for easy interpretation of clusters.

### **Details on Parameter Tuning or Optimization Performed**

* Iterated over a range of values to determine the optimal number of clusters (K).
* Tested the model on both scaled and unscaled data, with unscaled data yielding better results.

## Feature Importance Analysis

### **Methods Used for Determining Feature Importance**

* Utilized the SHAP (SHapley Additive exPlanations) method to identify the most influential features.

### **Top Features Identified**

* Occupation
* Income Level
* Monthly Average Balance
* Number Of Transactions Per Month

## Results

### **Clustering Evaluation Metrics**

| **Cluster** | **WCSS** | **Davies-Bouldin Index** | **Calinski-Harabasz Index** | **Silhouette Score** |
| --- | --- | --- | --- | --- |
| 2 | 1.401816e+14 | 0.47 | 142496.78 | 0.69 |
| 3 | 5.736771e+13 | 0.47 | 207219.99 | 0.65 |
| 4 | 3.655739e+13 | 0.50 | 225490.90 | 0.60 |
| 5 | 2.408200e+13 | 0.54 | 262671.57 | 0.54 |

### **Interpretation of the Results**

The top features indicate that a customer's occupation, income level, average monthly balance, and number of transactions per month are significant factors in determining suitable investment recommendations.

## Conclusion

This project successfully developed a personalized investment recommendation system using KMeans clustering and SHAP for feature importance analysis. The system identified key customer attributes that influence investment decisions, enabling the bank to offer tailored investment products. Future work could explore integrating real-time data and expanding the range of investment products considered.