Investment Recommender System - CRISP-DM Report

Business Understanding

In Kenya, many individuals, especially young adults and first-time investors, face challenges in identifying investment opportunities that align with their financial goals, risk tolerance, and income level.

The abundance of investment products-ranging from SACCOs, government bonds, and real estate to money market funds and stocks-creates confusion and decision fatigue.

Low financial literacy and limited access to personalized financial advisory services exacerbate the problem, often leading to poor or delayed investment decisions.

Challenges

- Low financial literacy and accessibility to advisory services

  Many potential investors lack foundational knowledge or support systems to understand available investment options or evaluate their suitability.

- Overwhelming investment options

  The abundance of options such as SACCOs, stocks, real estate, government bonds, and money market funds can be confusing and lead to decision fatigue.

- One-size-fits-all investment marketing

  Most financial institutions promote products generically, failing to account for individual goals, income, and risk profiles.

- Lack of data-driven tools for personalized investment planning

  There is limited availability of intelligent systems to assist users in navigating investments based on their unique profiles.

- Distrust and fear of loss

  Without adequate knowledge or guidance, potential investors may fear financial loss or fall victim to scams, leading to investment hesitation.

Proposed Solution:

The project aims to develop a machine learning-based recommender system that provides personalized investment plan suggestions based on an individual's financial profile, investment goals, and risk appetite.

By guiding users toward suitable investment types, the system will promote financial inclusion and improve decision-making.

Problem Statement

Many individuals, especially in emerging markets, face significant challenges in making informed investment decisions due to limited financial literacy and lack of personalized advisory services. The wide range of available investment options SACCOs, stocks, real estate, government bonds, and money market funds can be overwhelming without guidance. Additionally, the generic approach in investment marketing overlooks the diverse financial goals, income levels, and risk appetites of potential investors, leading to poor financial outcomes and disengagement from long-term wealth-building.

Objectives

i. Analysis-Based

   Understand investment behaviors among Kenyan users and segment them based on patterns.

ii. Feature Engineering-Based

   Create user profiles using financial behavior indicators such as:

iii. Modeling-Based

   Build and evaluate recommender models, including:

   - Content-based filtering
   - Hybrid approaches (clustering + classification)

Data Understanding

The dataset is sourced from the FinAccess Kenya survey, which provides extensive data on financial behaviors, demographics, income, expenses, and investment habits of respondents across the country.

Key variables include:

- Demographic data: Age, gender, education level, region, and area type.
- Financial indicators: Monthly income, expenses, savings, debt levels, and investment history.
- Behavioral data: Risk tolerance, investment goals, and financial literacy ratings.

Initial exploration revealed that the dataset contains both categorical and numerical variables, with minimal missing values in key features. Class distribution analysis showed a higher representation of low-risk investment preferences, but overall diversity across categories was maintained.

Data Preparation

Data preprocessing involved several steps to ensure the dataset was ready for modeling

- Dropped irrelevant columns and retained features most predictive for investment decisions.
- Encoded categorical variables such as gender, education, and investment preferences.
- Normalized numerical variables like income and expenses to reduce skewness.
- Created derived features such as savings rate and debt-to-income ratio.
- Addressed mild class imbalance using resampling techniques.

The cleaned dataset was saved in CSV format for use in training and deployment pipelines.

Modeling

Multiple algorithms were tested to identify the best-performing model, including Decision Tree, Random Forest, Gradient Boosting, Logistic Regression, and a Neural Network.

The modeling pipeline included feature scaling, encoding, and hyperparameter tuning via GridSearchCV.

Evaluation metrics included accuracy, F1-score, and confusion matrices. Decision Tree and Gradient Boosting performed best, capturing complex non-linear patterns in the data.

The final model was serialized using joblib for integration into the deployment environment.

Evaluation

Model performance was evaluated on a holdout test set using accuracy, precision, recall, and F1-score.

Random Forest achieved an F1-score above 0.95, indicating strong predictive ability across investment categories.

The confusion matrix showed balanced performance, with minimal misclassifications in minority classes.

These results demonstrate that the model can make accurate, personalized investment recommendations.

Deployment

The deployment phase made the system accessible through a FastAPI backend and a Streamlit web interface.

- FastAPI served as the prediction API, exposing endpoints for submitting user profiles and receiving investment recommendations.
- Streamlit provided an interactive form where users can input their financial and demographic information.

The system was tested locally and designed for easy cloud deployment, enabling fintech platforms and financial institutions to integrate the recommender into their services.

Reflection and Future Work

The project successfully demonstrated how machine learning can be applied to financial advisory services to increase accessibility and personalization.

Future improvements include integrating real-time economic data, expanding the dataset with additional user profiles, and exploring deep learning models for better generalization.

Additionally, incorporating explainable AI techniques will enhance user trust by providing transparency on why specific recommendations are made.

Conclusion

This investment plan recommender system addresses a significant gap in financial advisory services in Kenya.

By leveraging survey data and machine learning, the system delivers personalized, data-driven investment suggestions, promoting smarter and more confident financial decisions.