Investment Decision Recommendation System - Hackathon

21PT05 - HARINI P
CODE LINK :https://github.com/HariniParam/Buckman_Hackathon.git

---

PROBLEM STATEMENT :

Given a dataset containing information about individuals and their investment behavior, the objective is to develop a recommendation system that can assist individuals in making informed **investment decisions**. The system should analyze various factors such as age, role, investment influencer, gender, education, city, income, investment experience, and others to suggest the most suitable investment option for each individual.

DATA EXPLORATION :

As I delve into the dataset, one of the crucial steps is exploring its characteristics and patterns.

- I have used the **bar plot** to gain insights into the demographic and employment distribution (gender, marital status, Household Income).
- Then to examine the distribution related to investment behavior(Percentage of Investment, Investment Influencer, Risk Level), I have used a **pie chart**.

DATA PREPROCESSING:

- First and foremost, I loaded the dataset containing information about individuals and their investment behavior.
- To ensure the data was suitable for modeling, I conducted extensive preprocessing steps.
- This included handling missing values, encoding categorical variables, and scaling numerical features.
- **Missing values** in numerical columns are filled with the mean, while those in categorical columns are replaced with the mode.

- **Categorical variables** are encoded using **one-hot encoding** to convert them into binary vectors, and numerical features are scaled using **standardization** to ensure consistency in feature scales.
- These preprocessing steps ensure that the dataset is suitably prepared for training machine learning models, enabling effective analysis and prediction of investment behavior.

MODEL SELECTION :

- With the preprocessed data in hand, I delved into selecting the most relevant features for my recommendation system.
- **RandomForestClassifier** is known for its robust performance in handling both classification and regression tasks. It is an ensemble learning method based on decision trees, which tends to generalize well to various types of data and does not require extensive tuning.
- One of the key advantages of RandomForestClassifier is its ability to provide **feature importance scores**. By analyzing these scores, I gained insights into which features were most influential in predicting the investment decisions. This helped in feature selection and further refining the model.
- The dataset for the recommendation system contained **multiple target variables**, namely the risk level and return earned by investors. RandomForestClassifier alone cannot handle multi-output classification directly.
- Hence, I utilized the **MultiOutputClassifier** wrapper, which extends the capabilities of single-output classifiers to handle multiple target variables.

ESTIMATION METRICS :

- In the context of the recommendation system, the **accuracy score** provides an overall assessment of how well the model predicts the risk levels and returns earned by investors.
- **Classification reports** provide a detailed breakdown of the model's performance for each class or category within the target variables. For the recommendation system, I generated classification reports for both the risk level and return earned

categories. These reports include metrics such as **precision, recall, and F1-score** for each class, providing insights into the model's strengths and weaknesses for different categories.

CONCLUSION:

- Based on the analysis and results obtained from the recommendation system, several insights and recommendations can be drawn to assist individuals in making better investment decisions:
    - **Understanding Risk Levels:** The recommendation system provides valuable insights into the risk levels associated with different investment options. Investors should carefully assess these risk levels based on their risk tolerance, financial goals, and investment horizon.
    - **Understanding Return Earned:** In addition to assessing risk levels, investors should also consider the potential return earned from different investment options. The recommendation system evaluates the expected returns associated with various investment choices, allowing investors to weigh the potential gains against the risks involved.
- Areas for Improvement and Future Research:
    - To enhance the performance of my recommendation system, I recognize the potential for further improvements despite achieving satisfactory results. I plan to explore different machine learning algorithms, experiment with **hyperparameter** tuning, and leverage ensemble methods to enhance the predictive capabilities of the model.
    - Furthermore, I aim to conduct more extensive **feature engineering** to better capture relevant patterns and relationships in the data.
    - As part of future research, mechanisms like **dynamic updating** of the recommendation system with new data can be developed to ensure its relevance and effectiveness over time. Implementing real-time or periodic model updates based on evolving investment trends will be crucial in enhancing the system's adaptability and performance.