Investment Recommendation System

- Barath K - 21PC06

ABOUT THE PROJECT

This project, the **investment recommendation system**, is designed to empower users in making well-informed decisions regarding their investments. It harnesses demographic insights, employment particulars, and investment behavior patterns to tailor personalized recommendations. Employing sophisticated machine learning techniques such as **MLPClassifier** and **XGBClassifier**, the system predicts potential investment returns and delivers actionable insights for optimal decision-making.

WORK FLOW

- Data Loading and Preprocessing
- Data Splitting & Scaling
- Model Training & Evaluation
- Recommendation Generation
- Example Recommendation
- Why Neural Network and XGBoost ?
- System Design

Data Loading and Preprocessing

- The algorithm starts by loading the dataset from an Excel file named "**Sample Data for shortlisting.xlsx**" into a pandas DataFrame.
- Categorical variables in the dataset are encoded into numerical representation using **LabelEncoder** from scikit-learn.
- Missing values in the dataset are filled with the mean of the respective columns.

Data Splitting & Scaling

- The dataset is split into features (X) and the target variable (y). Features include all columns except for the ones specified to be excluded.
- The data is further split into training and testing sets using train_test_split function from scikit-learn.
- **StandardScaler** from scikit-learn is used to scale the features to have a mean of 0 and a standard deviation of 1.

Model Training & Evaluation

- Two machine learning models are trained:
  - **MLPClassifier** (Neural Network) with 1 hidden layer of 100 neurons and another hidden layer of 50 neurons.
  - **XGBClassifier** (XGBoost) with scale_pos_weight set to 5 and objective set to 'binary:logistic'.
- Performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are calculated for both models using the testing data.

Recommendation Generation

- A function named **generate_recommendation** is defined to generate investment recommendations based on user input.
- User input is encoded and scaled to match the format expected by the trained models.
- Predictions are made using both the Neural Network and XGBoost models, and the recommendation is determined based on the model with higher accuracy.
- If both models give the same recommendation, that recommendation is returned.

Why Neural Network and XGBoost ?

Neural Network

- A Neural networks, especially **Multi-Layer Perceptrons (MLPs)**, are known for their capability to learn complex non-linear relationships between input features and target variables.
- MLPs are highly flexible and can capture intricate patterns in the data, making them suitable for tasks where the relationship between features and target is not easily discernible or linear.
- In the context of investment recommendation, neural networks can effectively learn from the diverse set of features such as demographic information, employment details, and investment behavior to predict investment returns

XGBoost

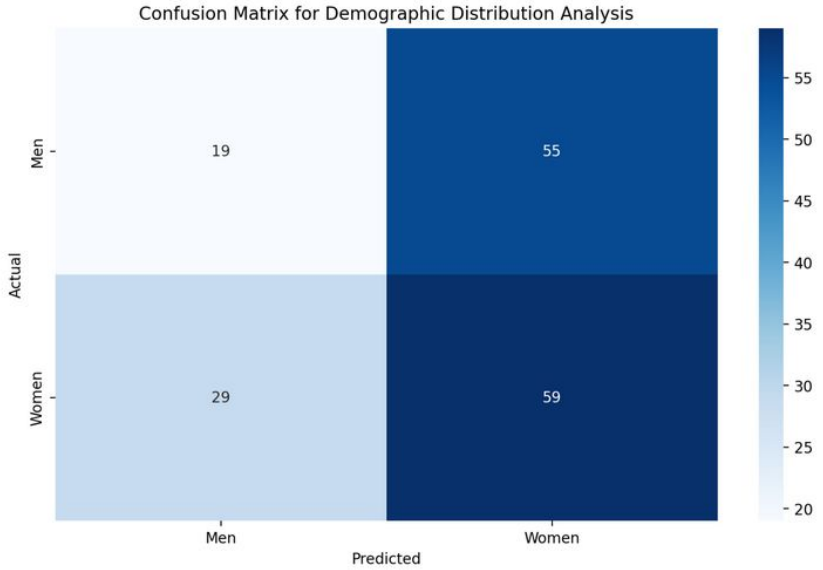
- XGBoost is an ensemble learning technique based on decision trees, known for its efficiency and effectiveness in handling structured/tabular data.
- XGBoost is particularly well-suited for handling tabular data with a large number of features, as it can capture both linear and non-linear relationships effectively.
- It provides robustness against overfitting and handles missing data well, making it suitable for real-world datasets with diverse characteristics.
- In the investment recommendation context, XGBoost can complement the Neural Network by providing a different perspective on the data and capturing additional nuances that may not be fully captured by the neural network.

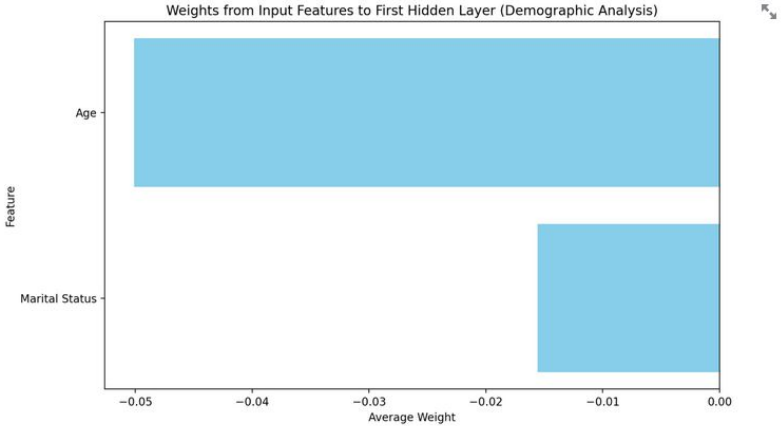
SYSTEM DESIGN

Demographic Distribution Analysis

Demographic Distribution Analysis

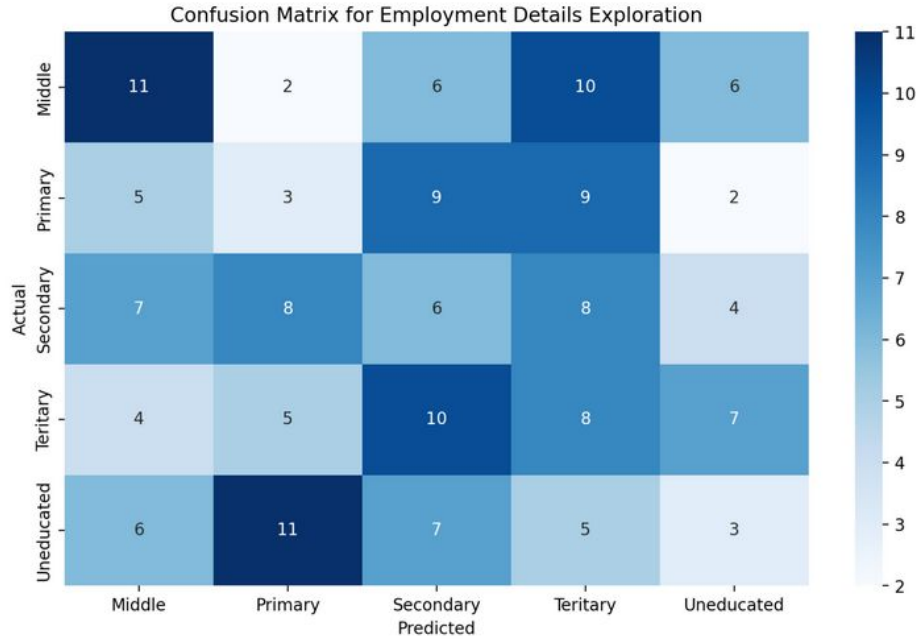
Accuracy: `0.48148148148148145`

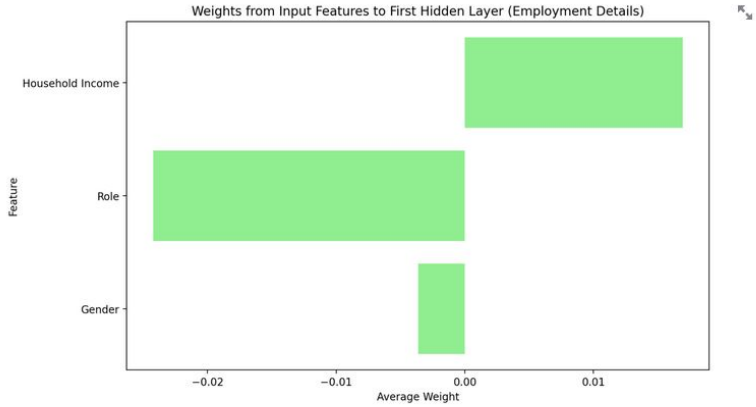




Employment Details Exploration

Accuracy: `0.19135802469135801`

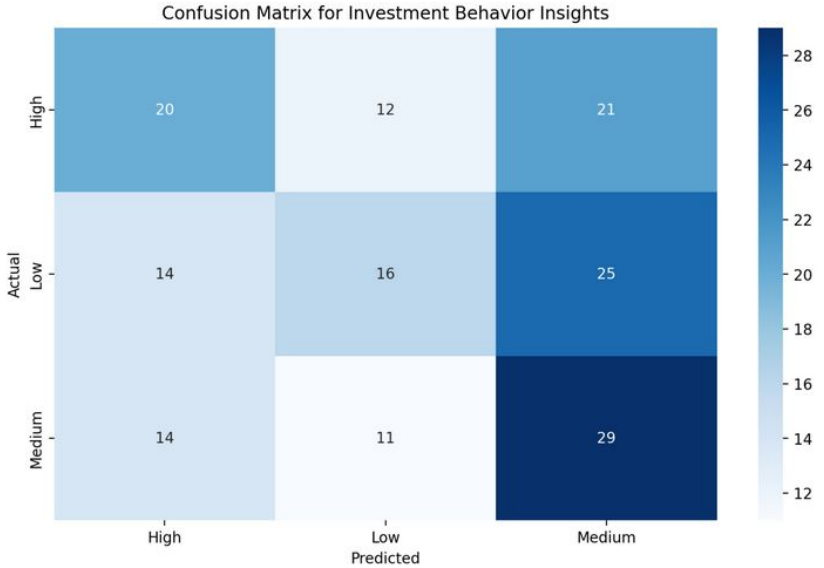


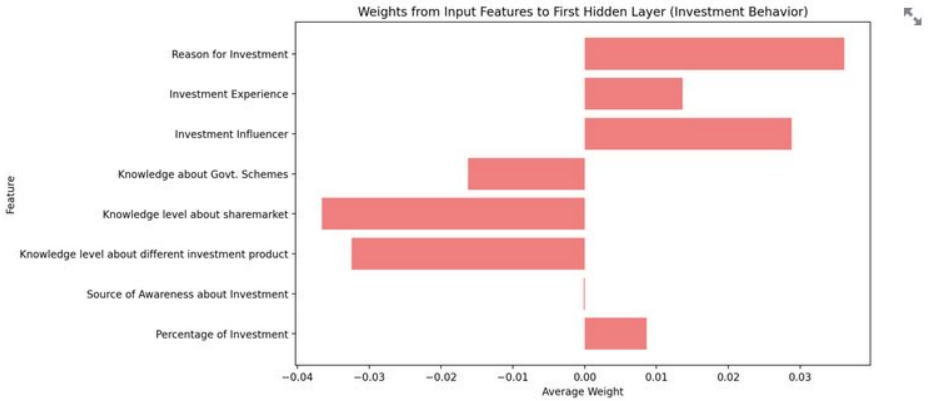


Demographic Distribution Analysis

Investment Behavior Insights

Accuracy: `0.4012345679012346`





RECOMMENDATION SYSTEM DESIGN

Recommendation System

Neural Network Metrics:


```
{
  "Accuracy" : 16.666666666666664
  "Precision" : 16.82966522393046
  "Recall" : 16.666666666666664
  "F1-score" : 16.66370469478175
  "ROC-AUC" : 49.20150279655632
}
```


XGBoost Metrics:


```
{
  "Accuracy" : 22.839506172839506
  "Precision" : 23.387990052503408
  "Recall" : 22.839506172839506
  "F1-score" : 22.941578738330815
  "ROC-AUC" : 53.63845091934524
}
```

Data Preview:

	S. No.	City	Gender	Marital Status	Age	Education	Role	Number of investors in family	Hou
0	1	2	0	1	1	2	2	2	
1	2	2	1	0	1	0	0	2	
2	3	4	1	0	1	3	4	2	
3	4	3	0	1	2	4	4	5	
4	5	3	0	0	4	1	1	2	

Number of investors in family

2 − +

Household Income

US$ 8206 to US$ 13675

Percentage of Investment

6% to 10%

Source of Awareness about Investment

Workers

Knowledge level about different investment product

7

Knowledge level about sharemarket

4

Knowledge about Govt. Schemes

5

Investment Influencer

Friends Reference

Investment Experience

4 Years to 6 Years

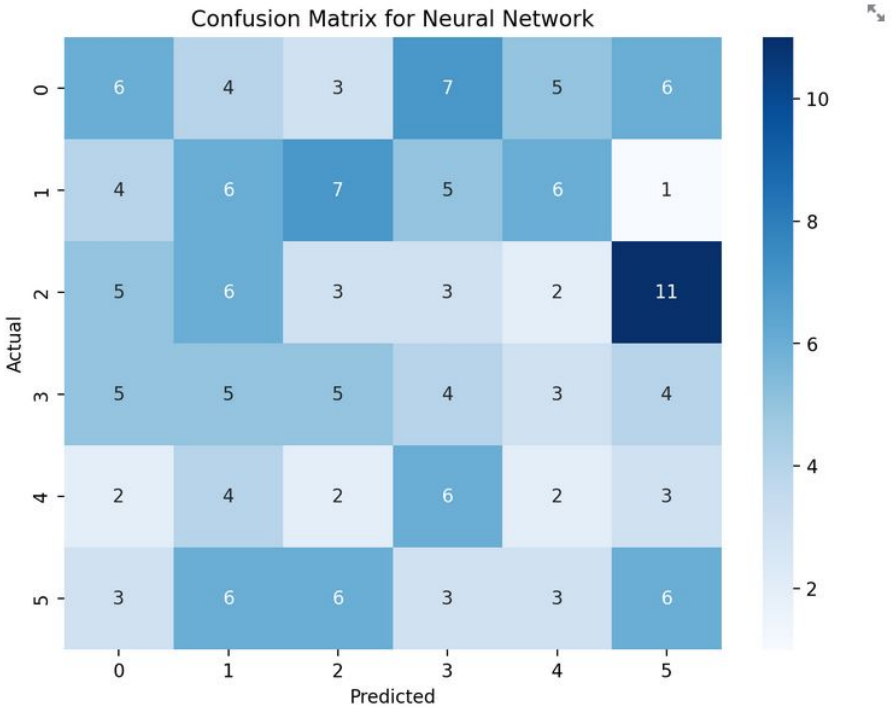
Risk Level

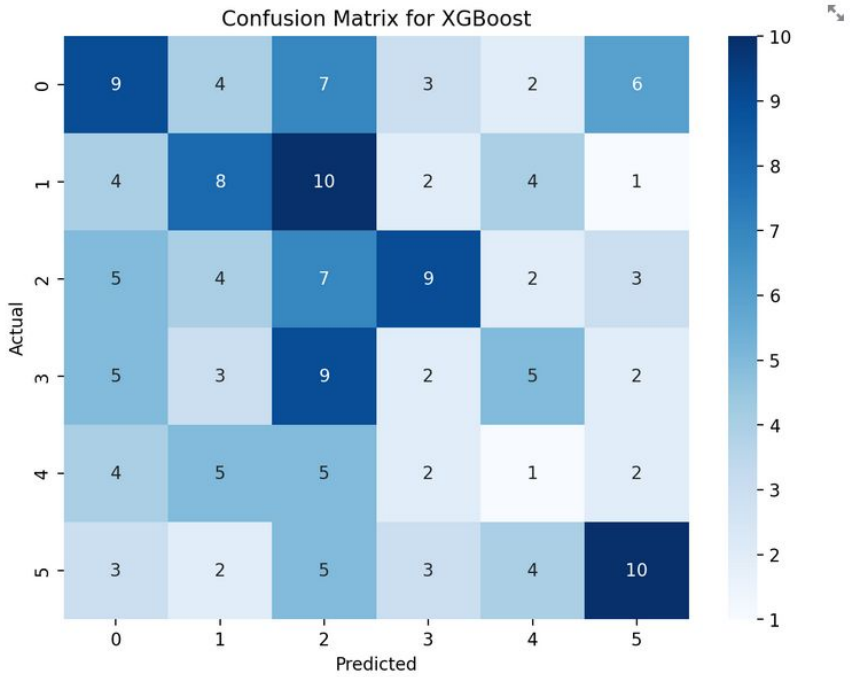
Medium

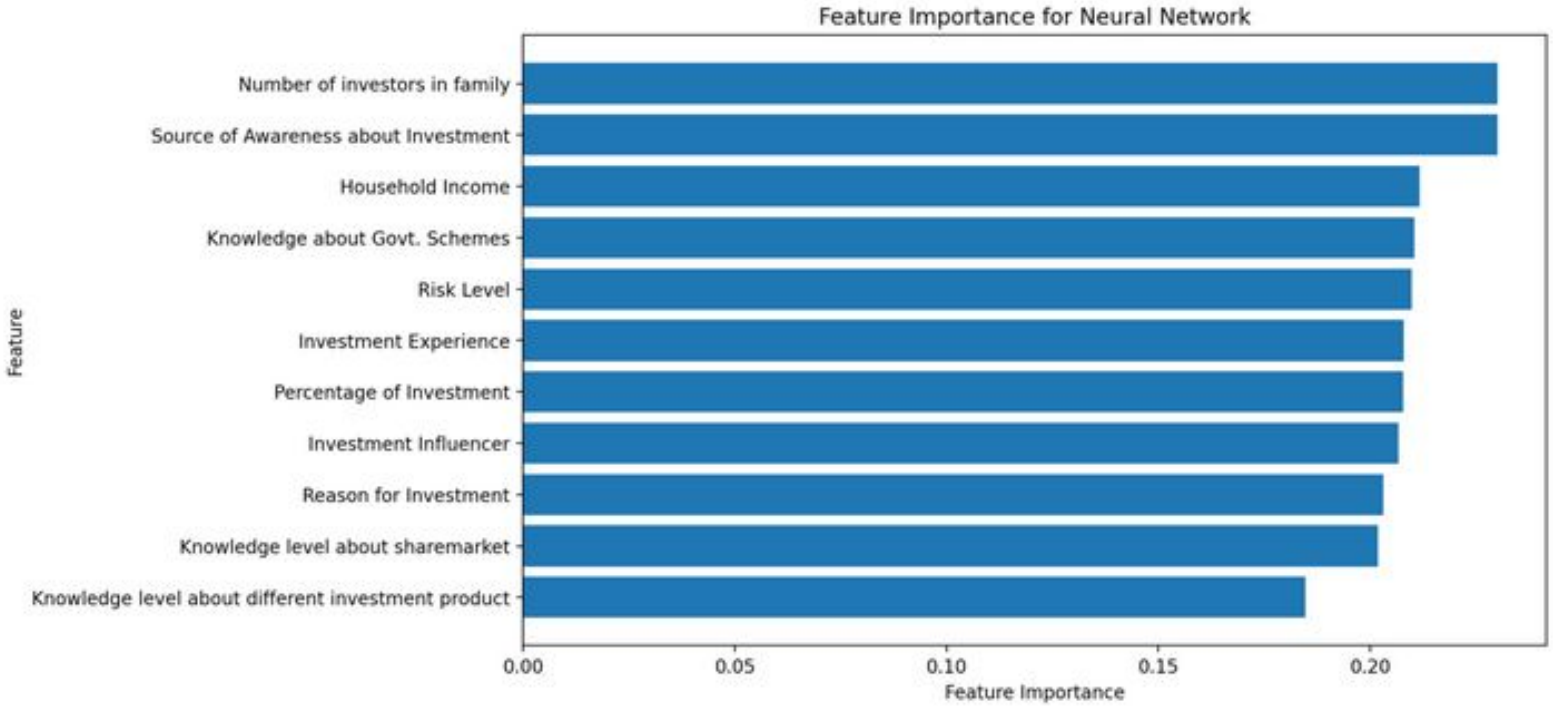
Reason for Investment

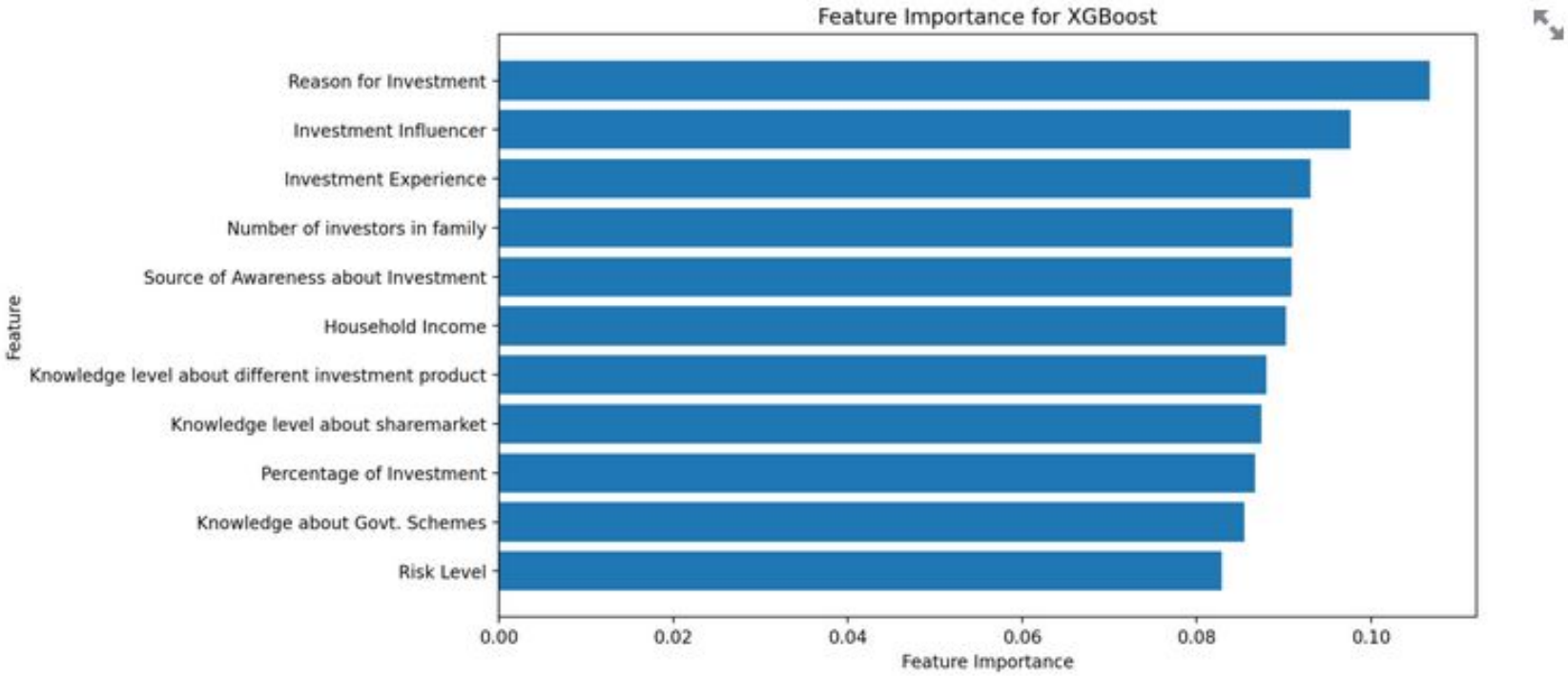
Fun and Exitement

Return Earned : 2


Confusion Matrix for Neural Network
Actual
Predicted
0 1 2 3 4 5
6 4 3 7 5 6
4 6 7 5 6 1
5 6 3 3 2 11
5 5 5 4 3 4
2 4 2 6 2 3
3 6 6 3 3 6
10 8 6 4 2


Confusion Matrix for XGBoost
Actual
Predicted
0 1 2 3 4 5
9 4 7 3 2 6
4 8 10 2 4 1
5 4 7 9 2 3
5 3 9 2 5 2
4 5 5 2 1 2
3 2 5 3 4 10
10 9 8 7 6 5 4 3 2 1


Feature Importance for Neural Network
Number of investors in family
Source of Awareness about Investment
Household Income
Knowledge about Govt. Schemes
Risk Level
Investment Experience
Percentage of Investment
Investment Influencer
Reason for Investment
Knowledge level about sharemarket
Knowledge level about different investment product
Feature
0.00
0.05
0.10
0.15
0.20
Feature Importance


Feature Importance for XGBoost
Reason for Investment
Investment Influencer
Investment Experience
Number of investors in family
Source of Awareness about Investment
Household Income
Knowledge level about different investment product
Knowledge level about sharemarket
Percentage of Investment
Knowledge about Govt. Schemes
Risk Level
Feature
0.00
0.02
0.04
0.06
0.08
0.10
Feature Importance

CONCLUSION

- In summary, by combining the strengths of **Neural Network** and **XGBoost**, our investment recommendation algorithm offers a robust solution for handling complex financial datasets.
- Neural Networks excel at capturing **non-linear relationships**, while XGBoost efficiently handles **structured data**. This synergistic approach enhances the accuracy and reliability of investment recommendations, providing investors with valuable insights for informed decision-making.

THANK YOU