Detailed Model Report: Personalized AI Financial Advisory

Model Architecture Choice and Justification

The model employs a MultiOutputRegressor with Random Forest Regressor as the base estimator. This architecture was chosen for its ability to handle multiple target variables simultaneously, making it ideal for predicting asset allocation percentages. Random Forest was selected due to its robustness against overfitting, ability to handle non-linear relationships, and feature importance analysis, which enhances interpretability.

Key Features of the Model:

- Handles both numerical and categorical inputs effectively.
- Provides feature importance metrics, enabling transparency in recommendations.
- Scalable to larger datasets and adaptable to new features.

Training Methodology

The dataset was preprocessed to ensure high-quality inputs for the model. Numerical features such as age, income, and expenses were normalized using StandardScaler, while categorical variables like risk tolerance and financial goals were encoded using Label Encoding.

Steps in Training:

1. Feature Engineering: Derived features such as 'savings_ratio' and 'is_long_term_goal' to improve predictions.
2. Train-Test Split: The data was split into 80% training and 20% testing sets.
3. Model Training: The Random Forest model was trained with 150 estimators, balancing performance and efficiency.
4. Hyperparameter Tuning: Parameters like tree depth and minimum samples per leaf were optimized to reduce overfitting.

Evaluation Results

The model's performance was evaluated using Mean Absolute Error (MAE) and R^2 Score. These metrics provide insights into the accuracy and reliability of the predictions.

Detailed Model Report: Personalized AI Financial Advisory

Evaluation Metrics:

- Mean Absolute Error (MAE): 3.45% (average deviation in predicted asset allocation percentages).
- R^2 Score: 0.92 (indicating that 92% of the variance in the data is explained by the model).

Observations:

- The model performed consistently across all asset classes, with slightly higher errors for volatile assets like equity.
- Features like 'savings_ratio' and 'goal_horizon_years' significantly improved prediction accuracy.

Challenges Faced and Solutions Implemented

1. Imbalanced Data:
   - Issue: Certain financial goals and risk levels were underrepresented in the dataset.
   - Solution: Applied SMOTE (Synthetic Minority Oversampling Technique) to balance the dataset.

2. Overfitting:
   - Issue: The Random Forest model showed signs of overfitting during initial training.
   - Solution: Limited tree depth and increased minimum samples per leaf to improve generalization.

3. Feature Correlation:
   - Issue: High correlation between income and expenses caused multicollinearity.
   - Solution: Regularized the model by normalizing features and removing redundant ones.

4. Interpretability:
   - Issue: Users required clear explanations for recommendations.
   - Solution: Incorporated feature importance analysis and rule-based explanations in the application.

Future Enhancements

To further improve the application, the following enhancements are planned:

- Integration with real-time market data to provide dynamic recommendations.
- Advanced risk profiling using psychometric analysis to better understand user preferences.
- Multi-language support to cater to a global audience.

- Development of a mobile application for better accessibility and user experience.