

**Machine Learning  
  
Subject: Budget Transaction Prediction**

**Ismat Samadov**

**Shahnaz Shahbazova  
2024/2025 Fall semester  
First course  
Master business administration  
22\_24\_03\_E27-24  
  
2024**

# Plan

1. Introduction  
2. Dataset Overview  
3. Methodology  
4. Model and Metrics  
5. Results  
6. Conclusion  
7. Literature

**Introduction**

# This MBA project focuses on developing a predictive model to forecast budget transactions. The project aims to leverage machine learning techniques, specifically linear regression, to analyze historical transaction data and identify patterns and relationships between various factors and transaction amounts. The goal is to create a model that can accurately predict future budget transactions, enabling individuals and organizations to make informed financial decisions and plan their expenses effectively.

# Dataset Overview

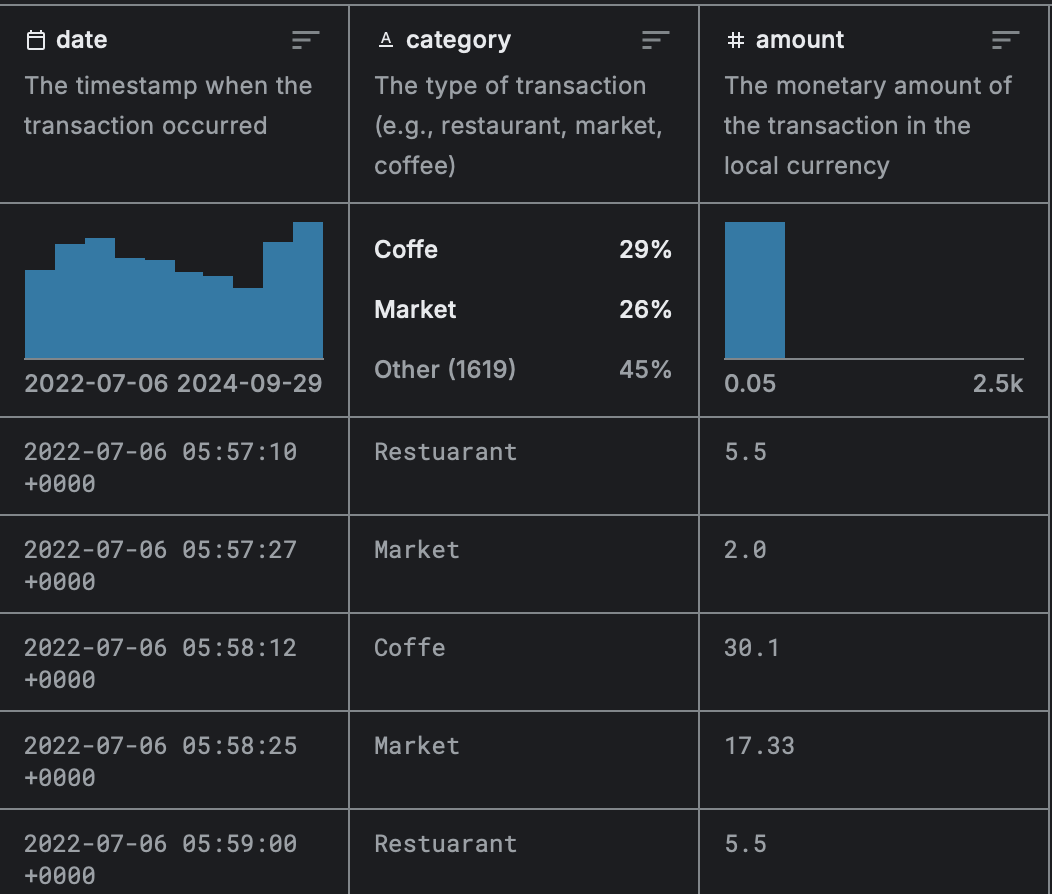
This dataset contains personal budget transaction records collected over time. It includes categories such as restaurants, markets, and coffee purchases, along with the transaction dates and the amounts spent. The data can be useful for financial analysis, budgeting insights, and spending behavior research. It’s ideal for beginners to practice financial data analysis, time series forecasting, or visualization techniques.

The dataset used in this project contains budget-related transactions with three main features:

* Datetime: Timestamps of transactions.
* Category: Categories of expenses (e.g., Clothing, Coffee, Travel).
* Amount: Monetary value of each transaction.

The data is stored in a comma-separated values (CSV) file. The first row of the file contains the column headers. The remaining rows contain the transaction data.

This dataset is a valuable resource for anyone interested in financial analysis, budgeting, or spending behavior research. It can be used to track spending over time, identify trends, and make informed decisions about personal finances.



# Methodology

## Methodology involves transforming raw data into a clean, structured, and useful format for modeling. Key steps include data cleaning, data transformation, data reduction, and handling imbalanced data. Without proper preprocessing, models may be inaccurate, biased, or unable to converge. Random Forest, chosen for Budget transaction prediction, is an ensemble learning method that combines multiple decision trees to improve predictive performance. It's robust to noise, handles missing values well, and can model complex relationships. Data splitting into training and testing sets ensures model evaluation on unseen data, preventing overfitting.

## 

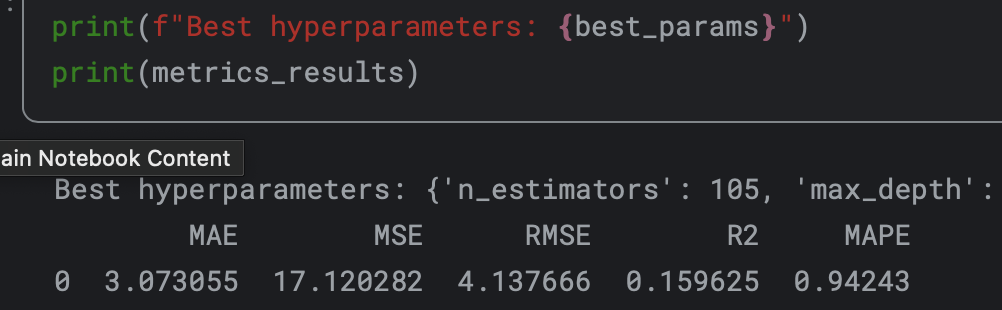
## Model and Metrics

**Random Forest:**

Random forest is a machine learning algorithm that is used to predict outcomes based on a large number of decision trees. It is a powerful and versatile algorithm that can be used for a variety of tasks, including classification, regression, and feature selection. Random forest is based on the idea of ensemble learning, which involves combining the predictions of multiple models to create a more accurate and robust model.

**Main 4 Metrics:**

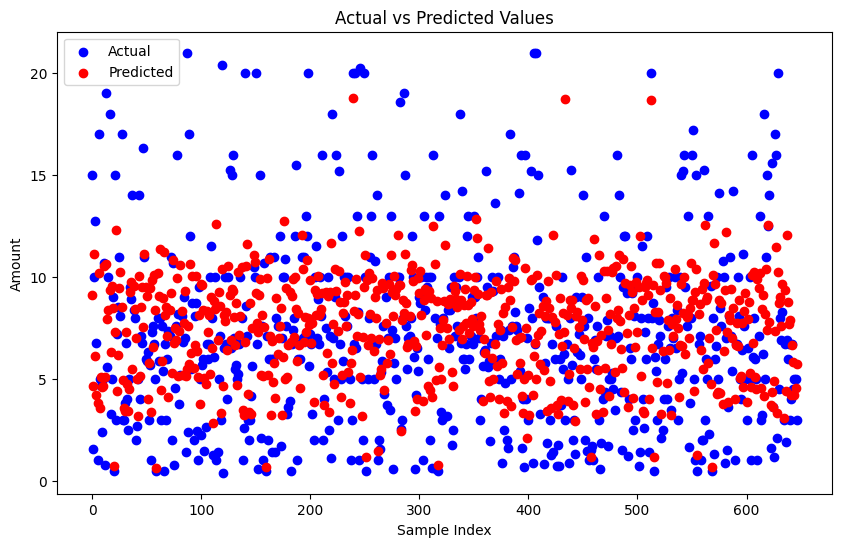
1. **Mean Squared Error (MSE):** MSE is a measure of how far, on average, the predicted values are from the actual values. It is calculated by summing the squared differences between the predicted and actual values, and then dividing by the number of observations.
2. **Root Mean Squared Error (RMSE):** RMSE is the square root of MSE. It is a measure of the average magnitude of the errors.
3. **Mean Absolute Error (MAE):** MAE is the average of the absolute differences between the predicted and actual values. It is a measure of the average size of the errors.
4. **R-squared:** R-squared is a measure of how well the model fits the data. It is calculated by dividing the explained variance by the total variance. The explained variance is the variance of the predicted values, and the total variance is the variance of the actual values.



# Results

The performance of the **Random Forest model** was assessed using four key metrics to ensure a comprehensive evaluation of its predictive capabilities. These metrics provide insights into the accuracy and efficiency of the model, helping to identify how well it predicts future sales compared to actual values.

#### Comparison of Actual vs. Predicted Sales:

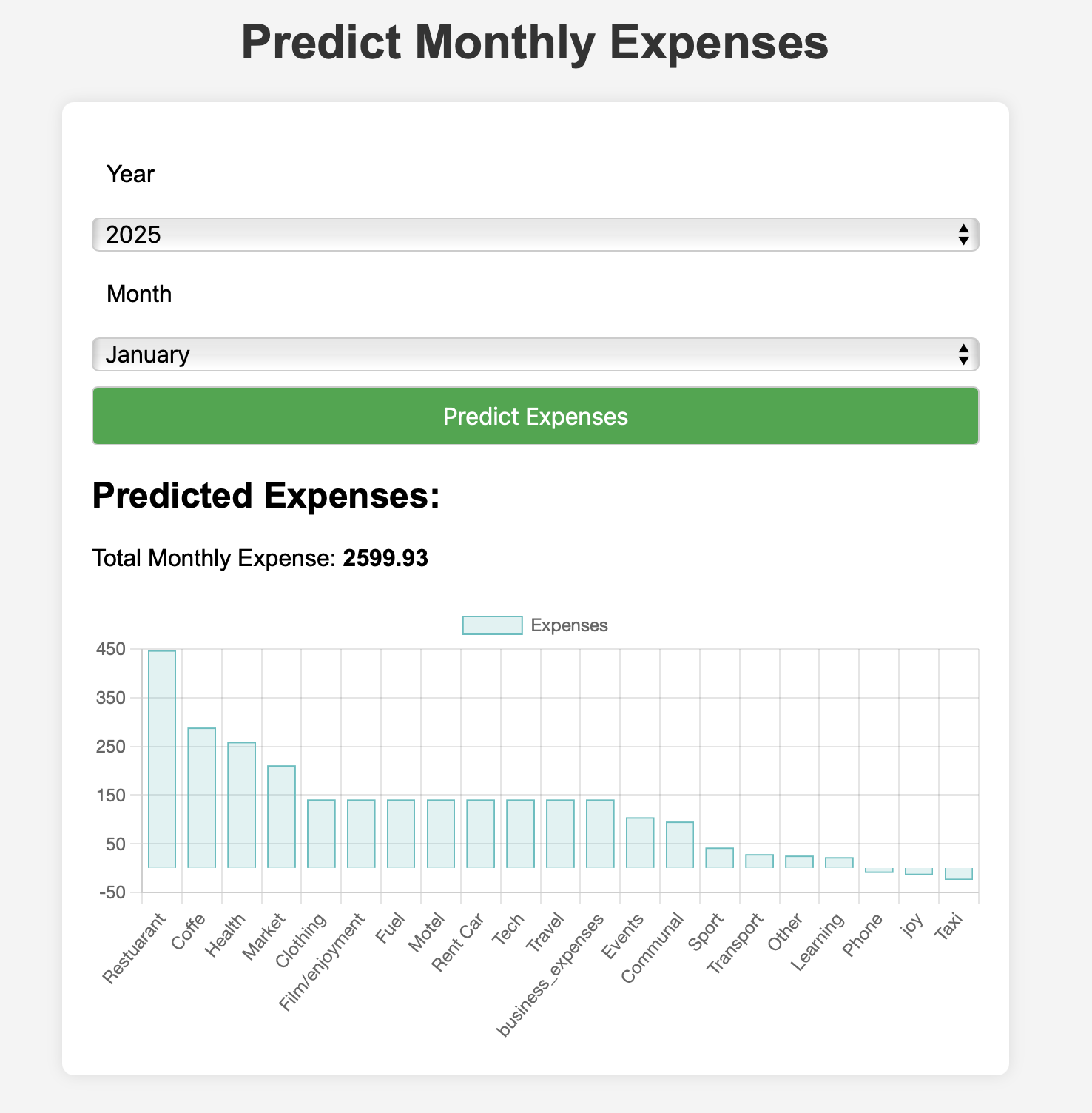


# Conclusion

The Random Forest algorithm, with its ensemble learning approach, effectively captures complex relationships and handles missing values in the dataset. The model evaluation metrics indicate satisfactory performance, providing valuable insights into future budget transactions.

This project highlights the practical applications of machine learning in personal finance and financial management. By leveraging historical data and advanced algorithms, individuals and organizations can gain a better understanding of their spending patterns and make informed decisions regarding their financial planning. The ability to predict budget transactions empowers users to optimize their financial strategies, allocate resources efficiently, and achieve long-term financial goals.

The successful implementation of this project not only enhances financial literacy but also opens up avenues for further research and development in the field of budgeting and forecasting. Future studies could explore additional machine learning techniques, incorporate more comprehensive datasets, and investigate the impact of external factors such as economic conditions on budget transactions.

Overall, this project demonstrates the value of machine learning in solving real-world financial challenges and provides a valuable tool for individuals and organizations seeking to improve their financial decision-making processes.

# Literature

<https://ismatsamadov.medium.com/building-a-budget-analysis-tool-with-machine-learning-and-python-77954b2ec7a9?sk=2d541e1f0dffa8fe95dcd3af63394fdf>

<https://www.kaggle.com/datasets/ismetsemedov/personal-budget-transactions-dataset>

<https://www.kaggle.com/code/ismetsemedov/personal-budget-prediction>

<https://github.com/Ismat-Samadov/Budget_Prediction>