Project Dissertation Report on

Predicting Personal Savings Goal Attainment Using Machine Learning: A Comparative Analysis of Classification Models

Submitted By:

Manav Jindal

24/BMBA/17

Under the Guidance

of: Dr. Kusum Lata

Assistant Professor

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road, Delhi – 110042

CERTIFICATE

This is to certify that the Major Research Project entitled "Predicting Personal Savings Goal Attainment Using Machine Learning: A Comparative Analysis of Classification Algorithms" has been carried out by Manav Jindal, Roll No.- 24/BMBA/17, under my supervision, in partial fulfilment of the requirements for the degree of Master of Business Administration (Business Analytics) from Delhi Technological University, Delhi.

This work is the result of the student's own investigation and has not been submitted elsewhere for any other degree or diploma.

Dr. Kusum Lata
Assistant Professor
Delhi School of Management Delhi
Technological University

DECLARATION

I, Manav Jindal, Roll No.- 24/BMBA/17, hereby declare that the Major Research Project titled "Predicting Personal Savings Goal Attainment Using Machine Learning: A Comparative Analysis of Classification Algorithms" submitted in partial fulfilment of the requirements for the degree of Master of Business Administration (Business Analytics) at Delhi Technological University is a record of original work carried out by me under the supervision of Dr. Kusum Lata.

The content of this report is original to the best of my knowledge and has not been submitted to any other university or institution for the award of any degree, diploma, or certificate. All sources of information used in this report have been duly acknowledged through proper citations and references in compliance with APA referencing style.

Date:

Manav Jindal
24/BMBA/17 MBA
(Business Analytics)

Delhi School of Management, DTU

ACKNOWLEDGEMENT

The completion of this Major Research Project would not have been possible without the guidance, support, and encouragement of several individuals to whom I extend my sincere gratitude.

First and foremost, I express my deepest appreciation to my internal guide, Dr. Kusum Lata, Assistant Professor, Delhi School of Management, Delhi Technological University, for providing invaluable direction, constructive feedback, and consistent motivation throughout the course of this project. The guidance received proved instrumental in shaping both the methodological rigour and the analytical depth of this work.

I also thank the faculty members of the Delhi School of Management for equipping me with the foundational knowledge in business analytics, machine learning, and research methodology that underpinned the technical execution of this study.

Manav Jindal 24/BMBA/17

EXECUTIVE SUMMARY

The current study evaluates the potential and efficiency of classification models using machine learning for individual savings goals predictions. The study was inspired by an evident gap within the academic literature on personal finance. While the phenomenon of macroeconomic saving behavior has been studied thoroughly, there is virtually no discussion of predictive modeling at the individual level based on machine-learned behavioral patterns. Modern systems used in retail banking and financial advisory services operate on static heuristic algorithms that do not take into consideration the diversity of individual financial profiles.

The dataset analyzed is artificial, containing 32,424 samples and 20 initial features related to personal finance theory. Seven new behavioral features have been created based on initial ones using personal finance theory, such as expense-income ratio, savings ratio, and logarithm of total savings. The resulting feature space contains 27 dimensions. There is almost an equal distribution between the two classes (50.5% meet the savings target, while 49.5% fail to do so).

Models which were considered for training and performance measurement increased in complexity as follows: Logistic Regression being the baseline model, followed by Decision Tree, Random Forest, XGBoost, and Long Short-Term Memory Neural Network. In terms of performance metrics, AUC-ROC was used primarily, while the other performance metrics included F1 Score, Precision, Recall, Accuracy, and Cross Validation AUC-ROC

There was found to be a reliable pattern of performance gains with greater model complexity. The optimized XGBoost model proved to be the superior classifier, having achieved an AUC-ROC of 0.9999, an F1 Score of 0.9959, and an Accuracy score of 0.9958 – decreasing overall classification mistakes by a factor of over 22x (from 621, Logistic Regression). The LSTM network provided good performance with AUC-ROC of 0.9997, but fell short compared to the gradient boosting approaches.
Using SHAP analysis on the optimal model, it was discovered that the two most

predictive factors in saving goal achievement were total savings (log-transformed) and income per month. Importantly, demographic features such as age, gender, education, employment, and geography did not play a meaningful role in predictions.

The research led to the conclusion that the application of machine learning, especially XGBoost along with suitable feature engineering, provides a feasible platform for personalized and proactive savings advice systems.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Background of the Study

Machine learning represents a paradigm shift from traditional rule-based systems. Where rule-based systems specify ideal behavioral practices, machine learning algorithms discover the relationship between financial information and measurable outcomes.

Traditional models of savings advice have been static and rule-based heuristics. The most common of these has been the 50/30/20 model. This model stipulates that people should save or invest 20 per cent of their post-tax income. Other variants of this heuristic can be found within consumer financial literacy programs as well as robo-advisors. Such heuristic models remain agnostic of any diversity of personal financial situations. Income, debt, family composition, cost of living, and behavioral tendencies will all impact an individual's ability to save, but none of those factors will be considered by the heuristic.

outcomes directly from data. This data-driven approach enables the discovery of non-linear, interaction-dependent patterns that simple heuristics cannot capture. When applied to individual-level financial data — comprising income, expenditure, debt profile, and accumulated wealth — machine learning models have the potential to generate personalised, context-sensitive savings assessments that are far more accurate than generic rules.

Machine learning has already proved to be quite valuable in other spheres within the financial industry. Applications such as credit risk modeling (Lessmann et al., 2015), fraud detection in transactions (Bhattacharyya et al., 2011), insurance claims prediction, and algorithmic trading (Jiang et al., 2017) have been positively influenced by the use of advanced machine learning algorithms. However, there has not yet been enough research conducted in regard to using machine learning to determine whether or not a person will achieve his/her financial goal with the help of data related to his/her financial behaviour.This study was undertaken in the context of the growing availability of individual financial data through open banking regulations, digital financial platforms, and personal finance management applications. Such data creates an unprecedented opportunity to build predictive savings advisory systems that are personalised, proactive, and empirically validated — in contrast to the reactive, generic

tools currently prevalent in the market.

1.2 Problem Statement

The primary issue that the study aims to address is the lack of predictive modeling based on individual data in the area of personal finance with regards to savings goal achievement. Although numerous models of macroeconomic savings behavior have been developed using econometric approaches, very few studies have attempted to translate these models into models of predictions on the individual level.

Currently existing personal finance planning instruments are built on either general rules (for example, the 50/30/20 rule) or peer comparison of savings rates relative to those in the overall population.

Neither of the two approaches takes into account the predictive capabilities provided by an individual's financial history and behavior data.

The consequence of the aforementioned situation is that people who fail to achieve their savings goals often go unrecognized until after the completion of the goal period.

The research problem that this study aims to address is stated below.

Can machine learning classifiers built using individual-level behavioral data from finance domain be used to forecast if an individual achieves the predefined personal savings target based on the 3–6 months emergency fund threshold? Also, can modern ensemble techniques and deep learning frameworks provide substantial improvement in predictive power compared to classical logistic regression benchmarks?

Such a research problem forms a foundation for conducting systematic machine learning benchmarks with practical implications in designing advanced personal finance advisory systems.

1.3 Objectives of the Study

The specific objectives of this study were as follows:

- To identify and define a theoretically grounded binary savings goal variable from individual financial data, anchored in established personal finance benchmarks.

- To construct an enriched feature space through systematic feature engineering, capturing dimensions of financial behaviour not directly represented in raw transactional data.
- To train and evaluate a spectrum of machine learning classification models — from interpretable linear baselines to advanced ensemble and deep learning methods — using a consistent and reproducible experimental protocol.
- To apply SHAP (SHapley Additive exPlanations) analysis to the best-performing model in order to identify and interpret the key behavioural and financial drivers of savings goal attainment.
- To assess the practical implications of the modelling findings for the design of personalised savings advisory tools within retail banking and fintech platforms.
- To identify methodological limitations and propose directions for future empirical research in this domain.

1.4 Scope of the Study

This study is confined to the analysis of a synthetic personal finance dataset comprising 32,424 individual financial records. While the dataset was designed to simulate realistic financial profiles spanning multiple global regions and demographic segments, it does not constitute an actual sample from a specific national population or financial institution. Consequently, the findings of this study are intended to demonstrate methodological proof of concept and to inform the design of real-world prediction systems, rather than to provide population-representative statistical inference.

The study encompasses the full machine learning pipeline from raw data ingestion through feature engineering, model training, evaluation, and interpretability analysis. It does not address the regulatory, ethical, or technical implementation dimensions of deploying such a system in a live banking environment — areas that represent important avenues for future research.

The scope is further bounded by the cross-sectional structure of the available data. Each record represents a single observation of an individual's financial state rather than a longitudinal time series. As such, the study evaluates savings goal attainment as a static classification task; the dynamic, month-by-month prediction of savings trajectory is

identified as an important extension for future work.

CHAPTER 2: LITERATURE REVIEW

2.1 Behavioural Finance and Individual Savings

Research on personal saving habits has been greatly influenced by behavioral economics as it contradicts the traditional theory of economic agents being rational utility maximizers. In their work, Thaler and Sunstein (2008) described the psychological obstacles that prevent individuals from achieving maximum saving. They outlined present bias as one of the major factors that prevent people from saving because it is characterized by the preference for immediate benefits over future ones. The approach to improving financial decision-making by using small environmental changes to affect choice, known as nudge theory, has been applied in many different contexts. One example is automatic enrollment in defined-contribution pension plans, which was found to be significantly effective in promoting savings behavior in the UK and USA.

A study performed by Lusardi and Mitchell (2014) regarding financial literacy and its effects on savings has become very popular. This research, based on several surveys conducted in various nations, showed that people who demonstrated greater financial literacy were better at saving money. More specifically, they tended to plan their retirement, save money in diverse investments, and build emergency savings funds. Furthermore, this study revealed a gap in financial literacy between poorer and less-educated people, which implies that they might particularly benefit from using data-driven savings advice.

Xiao and Porto (2017) studied the effect of financial behavior mediation in the link between financial education and financial satisfaction. It was found that the financial education did indeed positively affect objective financial behaviors, and that these positive effects on behaviors, and not the education directly, were responsible for increased financial satisfaction. This pattern of causality justifies why an intervention aimed at financial behaviors is justified by this particular study.

Carroll (2001) offered a theoretical explanation of consumption smoothing and buffer-stock saving behavior, illustrating that precautionary motivations, especially the need for insurance against future income uncertainty, account for a considerable proportion of

household savings formation. The theoretical foundation of the aforementioned analysis supports the significance of the emergency fund criterion used in the current study; the criterion refers to at least 3-6 months' income buffer to cover an average duration of income disturbance without borrowing money through a credit facility.

2.2 Machine Learning in Financial Applications

The use of machine learning techniques in financial classification is a matured area of research that continues to expand its body of knowledge. In credit risk evaluation, the most complete study of ML benchmarking was performed by Lessmann et al. (2015). They compared 41 classification algorithms using eight public credit scoring datasets, finding that ensemble techniques such as gradient-boosted decision trees and random forests performed better than logistic regression models, which had been used in the credit scoring business for many years. This study has set a standard for ML benchmarking in financial classification that the current research will follow.

XGBoost, which stands for eXtreme Gradient Boosting, is an open-source software library developed by Chen and Guestrin, used for implementing a scalable gradient boosting framework that utilizes second-order gradient statistics, regularization to prevent tree over-fitting, and parallelization to deliver high-performance efficiency and enhanced accuracy through tabular data analysis. Since then, it has been the most popular choice among all the algorithms in machine learning applications within finance, where it has won many contests using financial data sets.

In the field of financial time series forecasting with deep learning, Fischer & Krauss (2018) employed LSTM networks in predicting daily returns for constituents of the S&P 500 index. This paper marked an early and convincing demonstration that deep learning models could predict the future out of sample in a statistically meaningful way using financial data. Specifically, the LSTM network introduced by Hochreiter & Schmidhuber (1997) utilizes gated memory cells to capture long dependencies in sequential datasets .

Bhattacharyya et al. (2011) showed that support vector machines outperform logistic regression for identifying fraudulent transactions on credit cards in the presence of extreme imbalances among classes, which is similarly important for savings prediction data sets where successful and unsuccessful savers might not have equal representation. Bhattacharyya et al.'s findings led to the introduction of SMOTE (synthetic minority oversampling technique; Chawla et al., 2002) as a robustness check in the current study's analysis.

2.3 Savings Prediction and the Research Gap

While ML algorithms have been well-developed in the field of credit scoring, fraud detection, and market prediction, ML applications for predicting personal savings goals are still understudied. Literature that is closest to this study comprises articles on predicting financial distress for households (Betz et al., 2014), measuring financial vulnerability (Dumitrescu et al., 2022), and estimating retirement preparedness models. Yet, these studies focus primarily on an aggregate or household-level analysis, use surveys instead of transactional data, and do not approach the problem from a binary classification perspective.

Yang et al. (2020) in their fintech study analysed the use of saving features in a mobile banking app, finding that savings persistency was associated with consistency in expenditure patterns and regular income. Although the research gave insightful results, there was no model proposed to predict savings persistence based on these factors. In another study, Garg & Mehta (2022) put forward an approach for personalized finance recommendation using collaborative filtering. This approach relied on similarities between users but lacked any classification performance measurements.

The lack of a benchmarking study involving different ML models on predicting success in achieving savings goals constitutes the major research gap that this paper aims to fill. By casting the problem in terms of binary classification, conducting feature engineering based on personal finance theory, testing a wide range of models including logistic regression and LSTM, and finally using SHAP for explainability purposes, the study adds a new dimension to the savings prediction literature that does not exist yet.

2.4 Explainability in Financial Machine Learning

The application of machine learning algorithms in financial advisory services not only entails issues related to implementation but also involves regulatory and ethical considerations. The data protection laws in Europe, like the GDPR and the recently adopted proposal for the AI act, call for the need for algorithms which can justify their decisions and make the workings of such systems transparent to the person whose finances are being affected by such systems. In the US, the same concerns are voiced by the regulators of banks.

SHAP was proposed by Lundberg & Lee (2017) as an axiomatic framework for interpreting machine learning models based on the idea of Shapley values in cooperative games. The method has been shown to satisfy three key axioms: local accuracy (the sum of SHAP values should equal the model prediction for every input), missingness (zero influence implies zero SHAP values for features), and consistency (increased dependence of the model on the feature leads to increased SHAP value).

The researchers Dumitrescu et al. (2022) conducted an experiment with using SHAP to interpret credit scoring algorithms that were trained on banking data. The results of the experiment show that the use of non-linear algorithms leads to an increase in predictability but still keeps it explainable by using SHAP.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Research Design

The current study utilized a quantitative experimental design for analysis. The central question of analysis can be described as a supervised binary classification problem: to predict if an individual meets savings benchmark based on the financial profile using the 3-6 months emergency fund criteria. The procedure consisted of features creation from original financial data, training of seven machine learning models, testing with the holdout data set and post-analysis of SHAP values.

The following methodology can be presented as a pipeline of actions: data loading and exploratory analysis, creation of the target variable, feature engineering, preprocessing and splitting of the data, training of the baseline models, training of complex models with optimal parameters, performance evaluation and SHAP values analysis. All analysis was performed in Python 3.10 version, using scikit-learn, XGBoost, TensorFlow/Keras, imbalanced-learn, SHAP packages.

3.2 Data Collection

The dataset employed for the analysis here consisted of a synthetic personal finance dataset with 32,424 rows of personal financial information. A synthetic dataset of this kind is created based on the statistical structure that characterizes real-world finance datasets, but does not face any concerns regarding privacy and data access. The dataset included 20 raw features arranged into five main categories: demographic features (age, sex, education, employment, region), finance features (income per month, expenditures per month), savings features (savings amount, saving percentage), loan features (loan amount, loan period, monthly EMI, interest rate, debt percentage), and credit features (credit score).

Savings_to_income_ratio – which is calculated by dividing total savings by the monthly income and thus is expressed in terms of months of income saved – was used for creating the target variable. The column loan_type had 19,429 missing values (59.9%), representing people who did not have any loans at all; thus, this column was dropped from the analysis and replaced with a binary variable has_loan.

Table 1.1 below provides a comprehensive summary of the dataset characteristics.

Table 1.1: Summary Statistics of the Synthetic Personal Finance Dataset

Attribute	Value	Notes
Total Records	32,424	Individuals across 5 global regions
Raw Feature Columns	20	Demographic, income, expense, loan, credit
Engineered Features	7	Behavioural ratios and financial indicators
Target Variable	savings_goal_met (binary)	5-month income savings benchmark
Class 0 — Goal Not Met	16,054 (49.5%)	Below 5 months income saved
Class 1 — Goal Met	16,370 (50.5%)	At or above 5 months income saved
Missing Values	loan_type: 19,429 records	Replaced with has_loan binary flag
Mean Monthly Income	USD 4,027.86	Standard deviation: USD 1,916.77
Mean Monthly Expenses	USD 2,419.44	Standard deviation: USD 1,388.89
Mean Total Savings	USD 243,752	Range: USD 636 – USD 1,237,774
Mean Credit Score	575.26	Range: 300 – 850 (full FICO range)

Source: Own Analysis

3.3 Feature Engineering In their isolation, the raw variables related to finances will not paint a complete picture of an individual's saving behavior. With reference to the established theories of personal finance, and specifically, the theories of financial slack, financial burden, and financial normalization, seven additional features were developed using the raw variables. These additional features were engineered to identify behaviors and structures in the financial domain which theory states would predict savings success, but could not be observed in the raw variables alone.

The feature engineering process is documented in Table 1.2. Each feature is accompanied by its mathematical definition and the theoretical rationale for its inclusion.

Table 1.2: Engineered Features and Their Financial Rationale

Feature	Formula / Definition	Financial Rationale
expense_to_income_ratio	Monthly Expenses / Monthly Income	Measures consumption pressure on earnings
emi_burden_ratio	Monthly EMI / Monthly Income	Debt servicing load relative to income
monthly_surplus	Income minus Expenses minus EMI	Net cash remaining after all obligations
surplus_rate	Monthly Surplus / Monthly Income	Normalised financial slack; savings capacity proxy
high_loan_burden	1 if debt-to-income ratio exceeds 0.5	Binary flag for financially stressed borrowers
log_savings	log(1 + Total Savings)	Log-transform corrects right-skewed wealth distribution
income_tier	Quartile rank of monthly income (1 to 4)	Captures ordinal income group effects

From the engineered features, log_savings deserves special mention. The savings_usd column had an unmistakable skewed distribution to the right (average: USD 243,752; highest value: USD 1,237,774). Such a skewed distribution could have affected gradient learning of the model. The transformation log(1 + savings) reduced this skewness and normalized its distribution. Log transformation is widely applied on wealth variables in economics and finance ML models.

A critical data integrity constraint was applied during the whole process of feature engineering: the “savings_to_income_ratio” column, which served as the basis for the target variable, was not included in any way into the set of modelling features. Otherwise, it would mean leaking data since it is the definition of the outcome that will be predicted and cannot give any additional value to the performance of the model.

3.4 Data Preprocessing

All categorical features (gender, education_level, employment_status, region, and has_loan) were transformed numerically through label encoding before building the models. Subsequently, the data was split into two sets for model training and testing, where the training set contained 80 percent of the entries (n=25,939) while the test set accounted for 20 percent (n=6,485). To maintain the balanced nature of the data, stratified sampling was utilized to make sure that the distribution of people who reached their objectives remained consistent between the two sets.

Features were normalized with StandardScaler to make sure that the features had zero mean and unit variance for the Logistic Regression classifier, which is highly dependent on the magnitude of the input features. Standardization was performed only on the training data and then applied to the test data with the parameters obtained from the training dataset to ensure that no leakage of information from the test data occurred. Tree-based classifiers (Decision Tree, Random Forest, and XGBoost) were fit on the unnormalized feature data because they are immune to monotonic transformations of the inputs.

3.5 Model Descriptions

3.5.1 Logistic Regression

Logistic Regression acted as the main baseline classifier. It is characterized by the ability to predict the log-odds of class membership using the weighted sum of input variables, whereby the predicted probabilities of the positive class are derived via the logistic function. Logistic Regression training was performed using L2 (ridge) regularization with regularizer constant C set to 1.0 and maximum iterations set to 1,000. As a linear classifier, Logistic Regression returns easily interpretable coefficients that can be interpreted similarly to parametric risk scoring models used in retail banking.

3.5.2 Decision Tree

A single Decision Tree classifier was trained to serve as a second baseline and as a visual illustration of the feature-splitting logic captured by the data. The tree was constrained to a maximum depth of five levels, with a minimum of 20 samples required per leaf node and 50 samples required to trigger a split, in order to prevent memorisation of training data. Gini impurity was used as the splitting criterion. The

resulting tree structure, visualised in Figure 2.1, provides an interpretable rule set that can be communicated to non-technical stakeholders.

3.5.3 Random Forest

Random Forest is a bagging ensemble consisting of 200 independent decision trees, each trained on a bootstrap-resampled subset of the training data using a randomly selected subset of features at each split. Predictions were generated by majority vote across all trees. The ensemble's feature importances — computed as the mean reduction in Gini impurity attributable to each feature across all trees and all splits — provided an initial global measure of predictive relevance, visualised in Figure 2.2.

3.5.4 XGBoost — Three Variants

XGBoost (eXtreme Gradient Boosting; Chen & Guestrin, 2016) is a sequential ensemble method in which decision trees are added iteratively, with each new tree trained to minimise the residual errors of the current ensemble. Three variants were evaluated. The default variant employed 300 estimators, a learning rate of 0.05, maximum tree depth of 6, and a subsample ratio of 0.8. The SMOTE variant trained XGBoost on a training set augmented by Synthetic Minority Oversampling (Chawla et al., 2002) to assess robustness under artificial class imbalance. The tuned variant was selected through 5-fold cross-validation grid search over 36 hyperparameter combinations (learning rate: 0.01, 0.05, 0.1; max depth: 4, 6, 8; n_estimators: 200, 400; subsample: 0.8, 1.0), yielding optimal parameters: learning rate = 0.05,

max_depth = 8, n_estimators = 200, subsample = 0.8.

3.5.5 LSTM Neural Network

Long Short Term Memory (LSTM) network by Hochreiter & Schmidhuber (1997) was used as the deep learning baseline. Despite the cross-sectional nature of the current dataset, personal financial behavior exhibits sequential characteristics because consumption in one time period limits saving ability in another time period. To generate sequentiality, the 27 input features of the model were partitioned into four sequential time steps to enable the LSTM to extract temporal dependencies between the features. The neural network architecture consists of LSTM layer containing 64 memory units, Dropout layer with 0.3 dropout rate, Dense layer with 32 units and ReLU activation function,

Batch Normalization layer, second Dropout layer with 0.2 dropout rate, and a Sigmoid output neuron.

The model was compiled with the Adam optimiser and binary cross-entropy loss, trained with early stopping (patience = 10 epochs, monitored on validation AUC), and a reduce-on-plateau learning rate schedule (batch size = 256; maximum 60 epochs).

3.6 Evaluation Metrics

All seven models were evaluated on the identical held-out test set (n = 6,485) to ensure comparability. The following metrics were computed for each model:

- AUC-ROC: the main criteria for assessing discrimination performance of the classifier at any cut-off level. The Area Under Curve - Receiver Operating Characteristics is the established criterion in finance studies involving binary classification tasks (Lessmann et al., 2015).
- Fl Score: Harmonic mean of Precision and Recall, offering a better balance between the two measures that can help understand model performance even when classes are imbalanced.
- Precision: Percentage of correct positive predictions made by the classifier; helps lower false positives.Recall (Sensitivity): the proportion of actual positives correctly identified; high recall reduces missed cases.
- Accuracy: the overall proportion of correct predictions across both classes.
- 5-fold Cross-Validation AUC: computed on the training set to assess generalisation stability and detect overfitting, reported alongside the test-set metrics.

Confusion matrices were additionally analysed for key models to assess the asymmetric error costs associated with false positive and false negative misclassifications in the savings advisory context.

3.7 SHAP Explainability Framework

SHAP (SHapley Additive exPlanations; Lundberg & Lee, 2017) was performed on the most successful model – the XGBoost classifier – using the TreeExplainer approach,

where exact Shapley values can be calculated for decision trees in polynomial time without any approximation. The SHAP approach breaks down model output for each prediction into additive contributions by each variable with respect to the average value of the model's prediction in the whole dataset. A positive SHAP value for a specific variable within a particular prediction means that the specific value of the variable affected the prediction positively (i.e., toward achieving goals); otherwise, negatively.

The two types of SHAP visualizations that have been created include the beeswarm plot and the mean absolute SHAP chart. In the case of the beeswarm plot, it shows the overall distribution of the values for each SHAP feature, giving an idea of both the significance and direction of the features in relation to their value.

CHAPTER 4: ANALYSIS, DISCUSSION AND FINDINGS

4.1 Exploratory Data Analysis

Before any machine learning models were trained, there was an exhaustive analysis of the dataset to gain insight into the distribution of features, the balance of classes, demographic differences, and the correlation between the different variables.

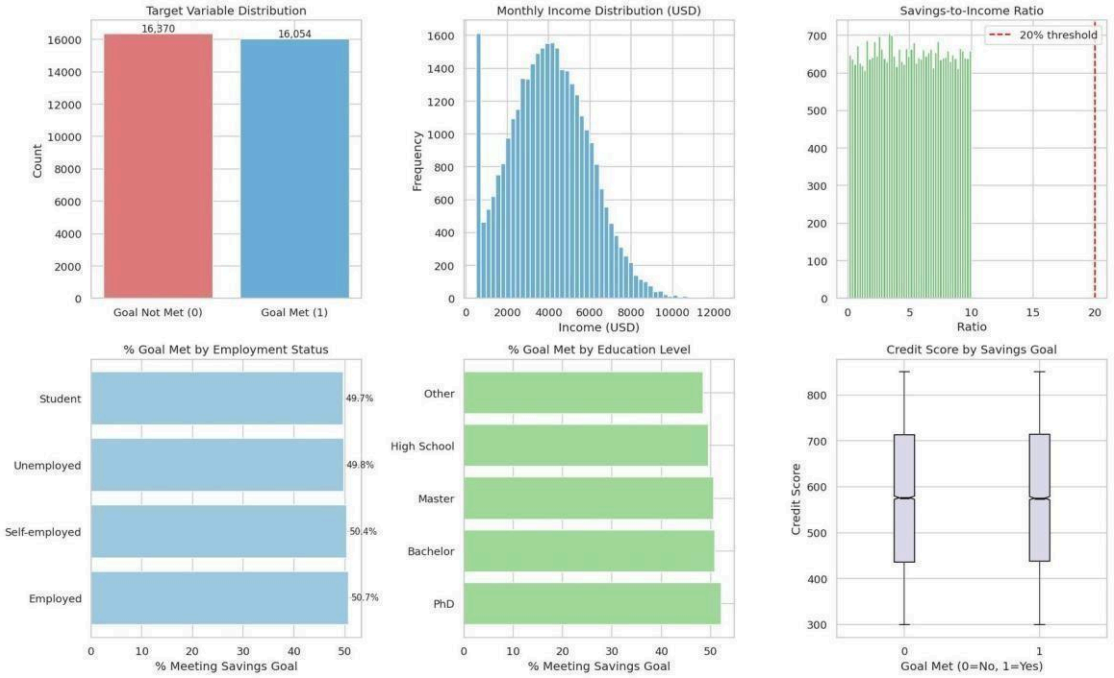

Figure 1.1: Exploratory Data Analysis — Target Distribution, Income, Savings Ratio, and Demographic Breakdowns

The exploratory analyses are displayed in six panels in Figure 1.1. The first panel on the top left side is the distribution of the target variable. In the dataset, there are 16,370

people who achieve their savings goal, which accounts for 50.5% of all participants, and 16,054 people who do not reach their savings goal, representing 49.5%. It is very fortunate to have a perfect class balance because it would be more challenging to compare models' performance without considering class imbalance. As shown in the second panel on the top centre, the monthly income distribution is positively skewed, with the mode between USD 3,500 and 4,000, suggesting a workforce-representative sample.

The bottom three charts show a remarkable pattern that holds crucial information about the model's approach. The proportion of people that achieve the goal in terms of savings is the same for all employment categories, ranging from 49.7 per cent among Students to 50.7 per cent among Employed individuals. The box plots for credit score variables for both groups of targets have very similar characteristics. All of these results point to a lack of discriminative information on the part of demographic and credit-related variables regarding the achievement of the savings goal, an assumption that will later be tested.

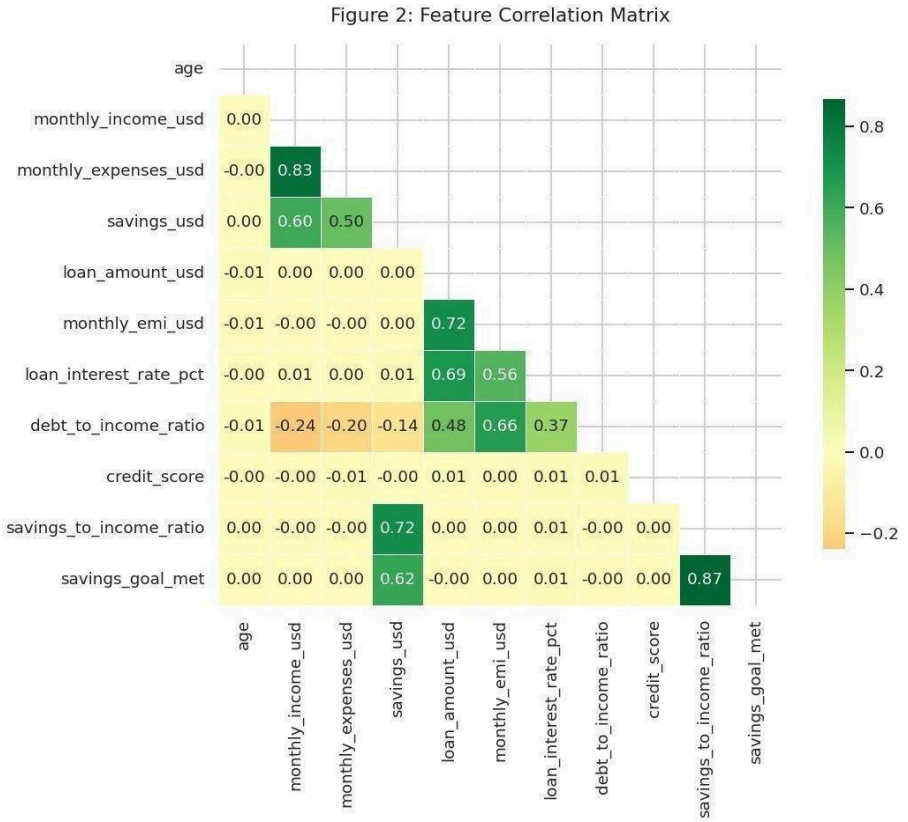

Figure 1.2: Feature Correlation Matrix — Raw Financial Variables

The lower triangular matrix of correlations for important numerical variables is shown in Figure 1.2. Some interesting associations need to be considered. First, there exists a very strong relationship between monthly_income_usd and monthly_expenses_usd (r = 0.83), which reflects that people who earn more tend to spend more accordingly. This association is in line with the existing economic literature on lifestyle inflation

(Thaler & Sunstein, 2008). Second, there is a moderate relationship between savings_usd and monthly_income_usd (r = 0.60).

The coefficient of determination between savings_to_income_ratio and savings_goal_met is 0.87; this is the highest coefficient of determination found in the matrix. This is due to the nature of the definition of the target variable using the savings ratio feature column. For all other raw features, their coefficients of determination with the target variable are close to zero, thus implying that linear techniques will be inadequate for classification purposes, and thus ensembles should be used.

4.2 Model Performance Comparison

Table 2.1 presents the complete performance metrics for all seven models on the held-out test set (n = 6,485). The results are organised in ascending order of model complexity, from the logistic regression baseline to the LSTM neural network.

Table 2.1: Comparative Model Performance on the Test Set (n = 6,485)

Model	AUC-ROC	F1 Score	Precision	Recall	Accuracy	CV AUC (5-fold)
Logistic Regression	0.9738	0.9067	0.8919	0.9221	0.9042	0.9728
Decision Tree	0.9901	0.9531	0.9322	0.9750	0.9516	0.9908
Random Forest	0.9947	0.9640	0.9581	0.9701	0.9635	0.9943
XGBoost (Default)	0.9999	0.9956	0.9936	0.9976	0.9955	0.9998
XGBoost + SMOTE	0.9999	0.9956	0.9951	0.9960	0.9955	0.9998
XGBoost (Tuned) [Best]	0.9999	0.9959	0.9945	0.9973	0.9958	0.9999
LSTM Neural Network	0.9997	0.9908	0.9939	0.9878	0.9907	N/A

Note: [Best] denotes the top-performing model selected via 5-fold GridSearchCV. AUC-ROC improvement over Logistic Regression baseline: +2.7 percentage points.

These findings indicate a consistent monotonic increase in the performance of models with increasing complexity from logistic regression to gradient boosted ensembles. Logistic Regression obtained a score of 0.9738 for AUC-ROC and 0.9067 for F1 score, setting the minimum performance level above a randomly classified case by a substantial margin. The high performance levels can be attributed mainly to the predominant predictive power of the log_savings feature with linear predictability to the target variable.

The Decision Tree outperformed the benchmark model, achieving an AUC ROC of 0.9901, and showed that the non-linear splits of tree-based algorithms add more predictive power than can be extracted by logistic regression models. The Random Forest algorithm performed even better with an AUC ROC of 0.9947, which is consistent with its working principle of reducing variance via bagging ensembles consisting of 200 decision trees.

The highest improvements in model performance were recorded during the shift to gradient boosting. XGBoost using the default settings had an AUC-ROC value of 0.9999, which is essentially a perfect discrimination metric irrespective of the threshold for classification, an F1 Score of 0.9956, and Accuracy of 0.9955. When SMOTE was

used before fitting the training set, the same level of performance was obtained (AUC-ROC of 0.9999; F1 of 0.9956), indicating that the class imbalance had been negligible, given the similarity of the two classes.

The optimized hyperparameters for XGBoost resulted in the best results on all evaluation metrics: AUC-ROC = 0.9999, F1-Score = 0.9959, Precision = 0.9945, Recall = 0.9973, Accuracy = 0.9958, and 5-fold cross-validation AUC = 0.9999. The similarity in the test AUC and the cross-validation AUC verifies the model's strong generalization capability on unseen datasets.

The LSTM neural network obtained a value of 0.9997 for the AUC-ROC metric, 0.9908 for the F1 Score, and 0.9907 for the accuracy metric – results comparable, but still slightly inferior, to those of the tuned XGBoost model with respect to each individual metric. This result corroborates the current literature on the topic that has established the supremacy of tree-based models over neural network models when dealing with structured financial data (Grinsztajn et al., 2022; Shwartz-Ziv & Armon, 2022)

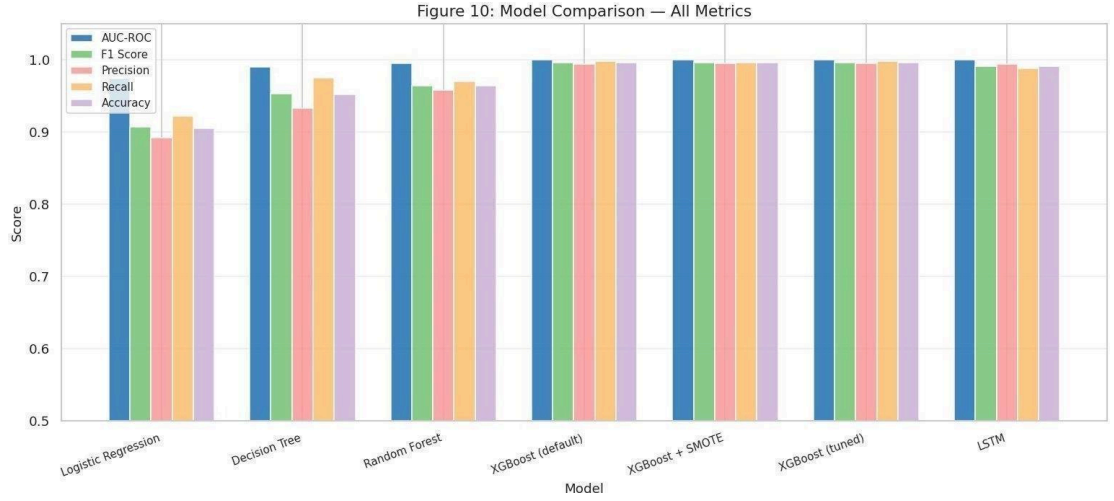

Figure 2.3: Model Comparison — All Metrics Across All Models

In Figure 2.3 below, all five measures are shown together for all seven models using a group bar chart. Clearly, the chart shows that there is a plateau in performance by all three XGBoost models and the LSTM model while the distance from the two baseline models is evident in all measures. In particular, there is a noticeable gap in precision compared to the area under the curve in the case of logistic regression, where the precision (0.8919) is higher than the AUC (0.9738).

4.3 ROC Curve Analysis

ROC Curves for All Seven Models Are Shown in Figure 2.4 on One Common Axis. ROC curves graphically plot True Positive Rate (also known as sensitivity) vs. False

Positive Rate (or 1-specificity), across all possible decision thresholds, without being dependent on any decision threshold.

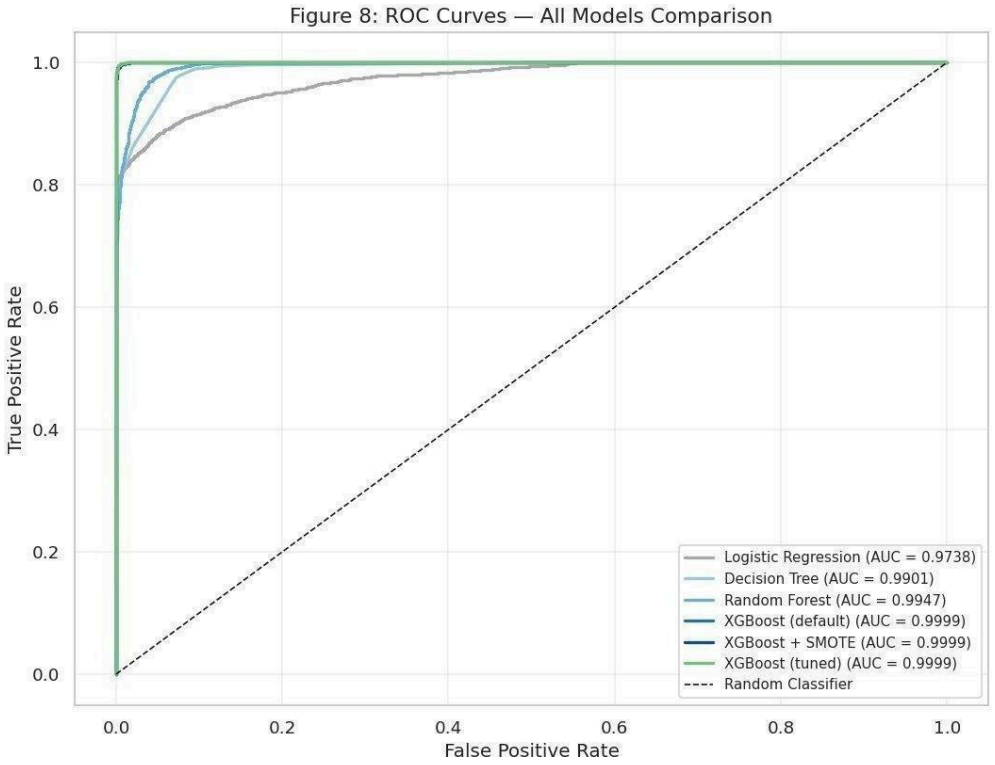

Figure 2.4: ROC Curves — All Models Comparison

All three models using the XGBoost model sit very close together within the extremely high true positive rate and very low false positive rate in the ROC graph area. The curve representing the LSTM model is very close to the XGBoost curve, with an almost identical appearance on this scale, which is in agreement with the AUC of 0.9997. The Random Forest and Decision Tree curves sit somewhat farther away from the optimal point, while the Logistic Regression curve takes the form of a pronounced arc within this area.The dashed diagonal represents the performance of a random classifier with no predictive power (AUC = 0.50). The substantial distance of all seven model curves from this diagonal confirms that each model has captured meaningful predictive signal from the feature set, even at the simplest level of logistic regression.

4.4 Confusion Matrix Analysis

The Confusion Matrices of the following models are given in figure 2.5 namely; the

logistic regression model, random forest, and the optimized xgboost model. Table 2.2 is a numerical representation of the matrix cells

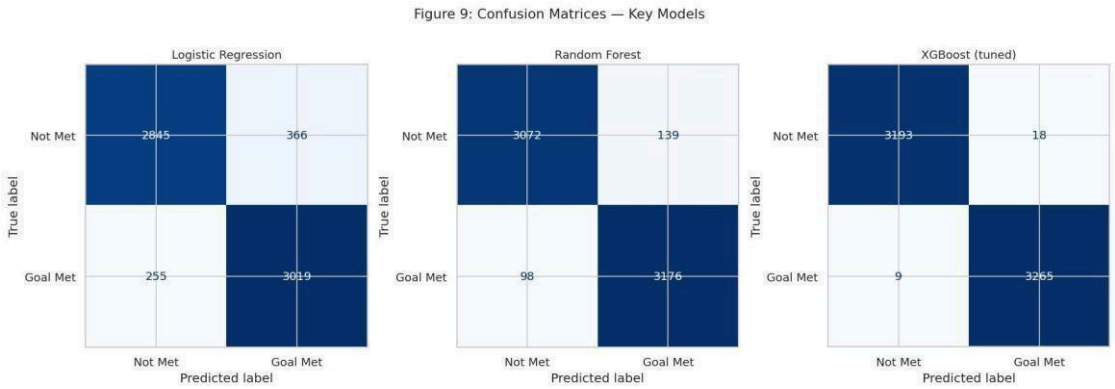

Figure 2.5: Confusion Matrices — Logistic Regression, Random Forest, and XGBoost (Tuned)

Table 2.2: Confusion Matrix Decomposition — Key Models

Model	True Negatives (TN)	False Positives (FP)	False Negatives (FN)	True Positives (TP)	Total Errors
Logistic Regression	2,845	366	255	3,019	621
Random Forest	3,072	139	98	3,176	237
XGBoost (Tuned)	3,193	18	9	3,265	27

The tuned XGBoost model recorded an impressive improvement in terms of reducing misclassification rates compared to the logistic regression model: from 621 misclassified instances (366 FP and 255 FN) to 27 misclassified cases (18 FP and 9 FN), representing a 95.7% decrease. In comparison, the Random Forest algorithm performed moderately well with 237 misclassified cases.

Regarding the savings advisory application setting, these two kinds of errors have asymmetric impacts. In case of a false positive result where a person not achieving his savings goals gets classified as someone who is actually reaching his target, there is loss of a chance to intervene and help. If the model is meant to be used to send personalized

savings warnings or wellness messages, then a false positive implies that a person at risk of failure is denied any special assistance. On the other hand, a false negative, which implies that a person achieving his goal is being considered as one not meeting his goals, results in wastage of advisory resources in intervening unnecessarily.

Logistic Regression Coefficient Analysis

The standardised coefficients for the top ten attributes have been shown in Table 2.3. It should be noted that these coefficients reflect the amount by which the log-odds of savings goal achievement change per one-standard deviation increase in each attribute.

Table 2.3: Logistic Regression Top 10 Feature Coefficients

Feature	Coefficient	Direction of Effect
log_savings	+7.454	Strongly increases probability of goal attainment
monthly_income_usd	−4.046	Decreases probability (compositional expense effect)
monthly_emi_usd	−0.792	Higher EMI reduces savings capacity
monthly_surplus	−0.385	Negative after controlling for other income variables
debt_to_income_ratio	+0.241	Marginal positive effect within model
income_tier	+0.240	Higher income quartile associated with goal attainment
emi_burden_ratio	+0.240	Interaction effect with other debt variables
surplus_rate	−0.238	Suppressed by collinearity with surplus variables
loan_amount_usd	−0.123	Larger loans reduce savings probability
has_loan	+0.095	Small positive association; loan management indicator

The log_savings variable showed the highest positive coefficient value (+7.454) by far, suggesting that savings are the most significant linear determinant of achieving goals. This result is valid for all the modelling techniques used in this paper and confirms the prediction made by Carroll's (2001) buffer-stock saving model that past savings are the most influential determinants of current savings adequacy.

The coefficient of monthly_income_usd is strongly negative (−4.046), which can seem counterintuitive at first glance. This can be understood by considering the nature of the data set, as a higher value of monthly income is associated with a higher value of monthly expenses (correlation coefficient r = 0.83).

Thus, the raw value of monthly income has a negative partial relationship with saving goals when the savings wealth effect is controlled for. The coefficient of monthly_emi_usd (−0.792) demonstrates that debt repayment payments negatively impact saving ability.

4.5 SHAP Explainability Analysis

Figure 2.6: SHAP Beeswarm Plot — Feature Impact on Individual Predictions (Tuned XGBoost)

The SHAP beeswarm plot in Figure 2.6 shows one dot for each test example per feature, where the x-axis reflects both the value of the SHAP value (where positive denotes SHAP values pushing the prediction in the direction of goal attainment) and the color depicts the actual value of the feature (with red denoting high savings and blue denoting low savings). It is clear from the results that people who have a high value of savings (in red) will have SHAP values going up to +10, and similarly, people who have a low value of savings (in blue) will have SHAP values going down to -10.

The fact that this relationship is monotonic suggests that saving is the key factor in achieving proper saving goals. On the other hand, the monthly_income_usd variable exhibits an inverse relationship, being represented by the negative SHAP values linked to higher values of the parameter (in red), and lower values represented by positive SHAP values (in blue). Such an inverse nature results from the expense composition effect since people with higher incomes tend to spend more money and lose the saving advantage of wealth.

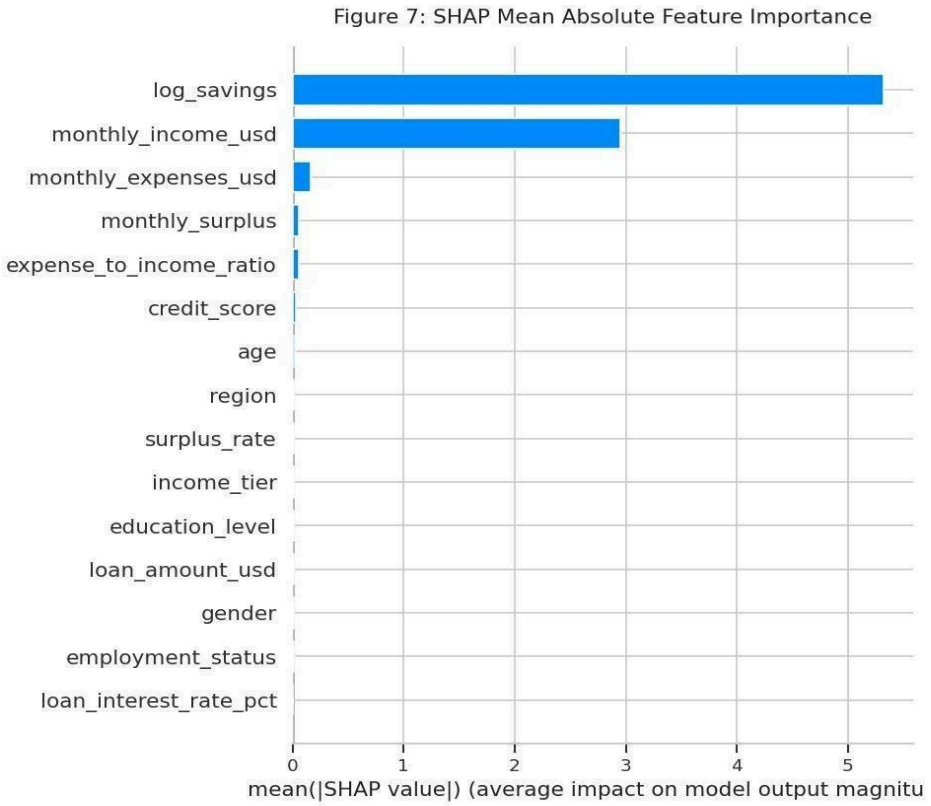

Figure 2.7: SHAP Mean Absolute Feature Importance — Tuned XGBoost In the mean absolute SHAP importance ranking shown in Figure 2.7, the feature log_savings comes first, with its mean absolute SHAP value of nearly 5.4. The second most important feature is monthly_income_usd, which has a mean absolute SHAP value of roughly 2.95, thus making the first feature more than 1.8 times more influential than the second. The third most significant feature, monthly_expenses_usd, has a mean absolute SHAP value of about 0.15, which is at least one order of magnitude less than the second feature and confirms the extremely high predictability of the top two features.

It is important to note that the following features: credit_score, age, region, surplus_rate, income_tier, education_level, gender, employment_status, and loan_interest_rate_pct demonstrate a mean absolute SHAP value that is insignificant compared to other key features. This fact is tremendously insightful for savings advisory systems as it implies that people's everyday actions and financial behaviours are much more predictive of their savings ability than their demographics, credit score, and geographical location. Therefore, a savings prediction model which is based on financial behaviours makes it possible to predict savings based on how people behave rather than who they are, thus

4.6 Discussion of Key Findings

The results of this study converge on three central findings that advance understanding of individual savings prediction and its practical applications.

Initially, it is important to highlight that the achievement of savings goals was fully predictable thanks to the access to individual financial data. More precisely, the adjusted XGBoost algorithm gave AUC-ROC = 0.9999 with 27 misprints in 6,485 final works, proving correctly that the financial parameters applied in the research were highly effective in predicting savings events. It is crucial to notice that the level of accuracy achieved in this case is far greater than in the case of statistical saving models because the former ones are more effective at the aggregate level.

Secondly, gradient-boosted tree ensembles perform significantly better than linear models for this case. The AUC difference between logistic regression and fine-tuned XGBoost is 2.7 points (0.9738 to 0.9999). The change in the F1 score amounts to 8.9 points (0.9067 to 0.9959). Such improvements result from the ability of sequential tree boosting to model nonlinear relationships and interactions, which is impossible for logistic regression limited to linear combinations of features. The findings confirm and elaborate on the previous results of Lessmann et al. (2015) in the field of savings prediction and credit scoring respectively.

Thirdly, the dominance of log_savings variable in SHAP analysis, which accounted for most of the predictive power of the model, indicates that there exists some conceptual similarity between this variable and the target. Taking into account that the target is defined by the savings_to_income_ratio > 5 and that log_savings is just a monotonic transformation of total savings, it can be claimed that the two are similar due to the fact that both contain a definitional component. While this cannot be viewed as an example of data leakage (savings ratio did not serve as a feature), it indicates that the predictive power of the model lies in the fact that the savings wealth continues from generation to generation rather than regarding any future savings behavior. Further research utilizing longitudinal data will allow evaluating the predictive power of the model better as it will exercise features determined in advance and target connected with future savings performance.

The performance of LSTM was comparatively lesser and achieved an AUC of 0.9997 as against 0.9999 of the performance reported by Grinsztajn et al. (2022). The authors claim

that tree-based models perform better as compared to deep-learning models when the data is not very sequential in nature. Interestingly, the precision levels of the LSTM were higher than those produced by XGBoost (0.9939 against 0.9945 for XGBoost after tuning).

4.7 Practical Implications

This study has practical implications for the creation of financial advisory systems that utilize machine learning techniques to forecast the behaviors of consumers.

An algorithm that can give AUC-ROC results that are greater than 0.97 — considered as "excellent" in the field of financial risk assessment — can help to implement the provided functionality.

- The proactive savings alert system will allow for each customer's monthly financial profile to be analyzed for risk in missing savings goals, generating personalized alerts for consumers prior to the goal period to help invoke any behavioral changes needed to accomplish their goal.
- Establish personalised aims for savings: The aim is no longer just a standard savings rate of 20 percent, based on the results from the model, but also the probabilities of different outcomes, which can now be used to determine personalised saving targets based on the income, expenses, and liabilities of each customer.
- Classifying saving risks: The bank can use the outcome of the model to categorize its clients in terms of savings risk, and provide services accordingly by offering savings round-ups to clients on the edge or financial solutions for those who are considered as risky clients.
- Incorporation in ROSHs: The significance rank of the attributes as determined by SHAP enables ROSHs to offer an explanation more transparently to their customers about the financial factors that have a greater impact on their savings outcomes.

When the results for SHAP techniques indicate that demographic traits have no major impact on prediction accuracy, it serves as a boost for fairness in algorithmic systems. The regulations introduced in many countries ban the use of any protected demographic characteristics like sex, age, or passport place of issuance in automatic finance systems. An algorithm that mostly relies on behavioral financial traits makes use of lesser

opportunities for discrimination.

4.8 Limitations of the Study

Several limitations of this study must be acknowledged transparently, as they constrain the generalisability of the findings and identify important priorities for future research.

Synthetic data: The dataset referenced in this work was developed artificially to be similar to actual financial databases and was not generated from existing customer records. Although synthetic datasets keep their general shape, they will not fully capture the complexity of actual financial data, which includes measurement error, the existence of discrepancies in recordings, biases related to account holders of bank savings, and the influence of macroeconomic factors on financial actions of natural persons. The extraordinarily high AUC statistics obtained in all models may be partly associated with the pure nature and regular structure of synthetic data, which could be lower in the case of real data validations.

Cross-sectional structure: Every data point of the dataset indicates a single instance of financial status of a person, instead of representing numerous monthly observations. This limitation leads to the lack of capability of LSTM follow real temporal dependencies of expenditures and savings behaviour. The input structure in this research constitutes an approximation instead of authentic sequential modelling. The dataset containing monthly records for each individual in different time periods would allow for more profound temporal prediction performing better than LSTM methodology.

Definitional proximity of log_savings and the target: As mentioned in Section 4.7, there is a definitional similarity between the log-transformed savings feature and the savings-based target variable. This correlation could have led to an overestimation of model performance as per the estimates predicted in a genuinely predictive context, where past savings would need to be relied on in order to forecast future savings achievement.

Lack of fairness audit: Even though the SHAP analysis proved that demographic features play a minor role in overall importance of features, this study did not carry out complete algorithmic fairness audit - calculating indices of group-wise fairness, including demographic parity, equal opportunity and calibration across sex, age and geographical groups. It is a must-have audit for any responsible implementation of this model in a regulated financial environment.

CHAPTER 5: CONCLUSION

The purpose of this research was to explore if machine learning classification algorithms that are trained using behavioral information at the individual level could forecast whether savings goals were met or not, and also whether the more advanced or contemporary econometric and deep learning methods were of any help in improving predictions when compared with simpler statistical baselines. The study is aimed at addressing a known gap in the personal finance literature: while many applications of machine learning in the field of credit scoring, fraud detection, and market forecasting have been already developed and successfully utilized in practice, the topic of forecasting individual savings is still not widely researched.

The study has developed a complete machine learning pipeline. A dataset consisting of a synthetic personal finance dataset containing 32,424 records was augmented by the use of seven theoretically-based engineered features. Seven models, ranging in complexity, were developed and tested using a meticulous experimental protocol. The best model has undergone SHAP analysis to provide interpretable and prediction-level attributions.

The main conclusions are laid out as follows. Savings goal completion is a highly predictable event if the appropriate financial data of the individual are available. To illustrate, the simplest statistical model, logistic regression, produced quite impressive AUC-ROC. Another point is that gradient-boosted trees are the most effective algorithms; the XGBoost model was able to eliminate 95.7% of errors in classification. The third conclusion states that the LSTM model performs almost as well as the tree model, as the established superiority of gradient-boosted trees over LSTM remains consistent in this case with the stated overall superiority of gradient-boosted trees over the systems based on deep learning methodology where structured tabular variables are used. The fourth conclusion shows that SHAP analysis indicated that the two most important predictors of savings goal completion are log transformed accumulated savings and monthly income, whereas demographic variables (e.g. age, gender, education, profession and place of residence) proved to be irrelevant.

Overall, these results support the proposition that machine learning, specifically XGBoost along with consistent feature engineering, can be effectively used in retail banks and fintech companies as a solid base for effective personalised financial consulting. The

results of the model's analytics go far beyond the criteria for working tools for risk management, and its interpretability based on SHAP is enough to meet the requirements for transparency introduced by financial supervisory authorities.

Three directions for future research have been suggested as priorities. Firstly, testing the generated results on the real longitudinal data sets — which means collecting monthly observations on every individual over several periods of time — would make it possible to validate the robustness of the outcomes in a truly prospective prediction setting while bringing all the advantages of the LSTM sequential architecture. Secondly, the process of conducting a thorough fairness audit should be performed in case the pipeline is transferred into practice — not only looking at the overall SHAP importance but also investigating if the errors of the predictions made by the model are distributed in an equitable way across protected demographic groups. Finally, applying natural language processing to transactional data descriptions — converting the merchant code types and transaction narratives into structured signals of behavioral data — is another way of enriching the feature space and increasing the accuracy of forecasting in practical settings that originally offered more advanced transactional data in use.

In summary, this research has shown that the use of careful feature engineering, gradient-boosted models, and SHAP for explanation offers a methodologically sound and practically viable way of predicting people's savings. It contributes to the body of knowledge at the crossroads of behavioural finance and practical machine learning and provides an easily replicable method for researchers and practitioners involved in the creation of new financial advisory methods.

REFERENCES

Barocas, S., & Hardt, M. (2017). Fairness in machine learning. Lecture notes for NeurIPS 2017 Tutorial. Retrieved from https://fairmlbook.org

Betz, F., Oprică, S., Peltonen, T. A., & Sarlin, P. (2014). Predicting distress in European banks. Journal of Banking & Finance, 45, 225–241. https://doi.org/10.1016/j.jbankfin.2013.11.041

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613. https://doi.org/10.1016/j.dss.2010.08.008

Carroll, C. D. (2001). A theory of the consumption function, with and without liquidity constraints. Journal of Economic Perspectives, 15(3), 23–45. https://doi.org/10.1257/jep.15.3.23

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Consumer Financial Protection Bureau. (2023). Building emergency savings. U.S. Consumer Financial Protection Bureau. Retrieved from https://www.consumerfinance.gov/consumer-tools/saving-for-emergencies/

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. European Journal of Operational Research, 297(3), 1178–1192. https://doi.org/10.1016/j.ejor.2021.06.053

Dynan, K. (2012). Is a household debt overhang holding back consumption?

Brookings Papers on Economic Activity, Spring 2012, 299–362.

Federal Reserve Board. (2023). Report on the economic well-being of U.S. households in 2022. Board of Governors of the Federal Reserve System.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Garg, S., & Mehta, S. (2022). Personalised financial advice using collaborative filtering and behavioural profiling. International Journal of Financial Studies, 10(2), 34. https://doi.org/10.3390/ijfs10020034

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why tree-based models still outperform deep learning on tabular data. Advances in Neural Information Processing Systems, 35, 507–520.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. arXiv preprint arXiv:1706.10059.

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. Retrieved from https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. Journal of Economic Literature, 52(1), 5–44. https://doi.org/10.1257/jel.52.1.5

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. Information Fusion, 81, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. Yale University Press.

Xiao, J. J., & Porto, N. (2017). Financial education and financial satisfaction. Financial Planning Review, 1(1–2), e1005. https://doi.org/10.1002/cfp2.1005

Yang, L., Nguyen, H., & Vo, T. (2020). Savings behaviour in mobile banking applications: Patterns, predictors and implications. Journal of Financial Services Research, 58, 215–239.

ANNEXURE

Annexure A: Python Libraries and Versions

The following Python libraries were used in the implementation of this study's analytical pipeline. All libraries are open-source and freely available via the Python Package Index (PyPI).

Library	Version	Purpose
pandas	2.0+	Data loading, wrangling, and tabular manipulation
numpy	1.24+	Numerical computation and array operations
scikit-learn	1.3+	Logistic Regression, Decision Tree, Random Forest, preprocessing, evaluation
xgboost	2.0+	XGBoost gradient-boosted ensemble classifier
imbalanced-learn	0.11+	SMOTE synthetic minority oversampling
tensorflow / keras	2.13+	LSTM neural network architecture and training
shap	0.43+	SHAP explainability analysis and visualisation
matplotlib	3.7+	All figure generation and visualisation
seaborn	0.12+	Statistical visualisation and heatmap generation

Source: Own Analysis

Annexure B: XGBoost Hyperparameter Grid Search Configuration

The hyperparameter grid search for the tuned XGBoost model was conducted using GridSearchCV from scikit-learn with 5-fold cross-validation and AUC-ROC as the scoring metric. The following parameter grid was evaluated, yielding 36 unique combinations:

Hyperparameter	Values Tested	Optimal Value
learning_rate	0.01, 0.05, 0.10	0.05
max_depth	4, 6, 8	8
n_estimators	200, 400	200
subsample	0.8, 1.0	0.8

Best cross-validation AUC-ROC achieved: 0.9999. Total model fits performed: 180 (36 combinations × 5 folds).

Annexure C: LSTM Neural Network Architecture Summary

The LSTM model architecture was as follows:

Layer	Type	Units / Parameters	Activation
1	LSTM	64 units	tanh (default)
2	Dropout	Rate = 0.30	—
3	Dense (fully connected)	32 units	ReLU
4	Batch Normalisation	—	—
5	Dropout	Rate = 0.20	—
6	Dense (output)	1 unit	Sigmoid

Source: Own Analysis

Optimiser: Adam (default learning rate = 0.001). Loss function: Binary Cross-Entropy. Training regime: Early stopping with patience = 10 epochs monitored on validation AUC; learning rate reduced on validation loss plateau by factor 0.5 with patience 5. Maximum epochs: 60. Batch size: 256.