International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

AI-Driven Financial Risk Assessment and Wealth Optimization System

Kalyanasundaram S[1], Dr. S.V.Anandhi[2] , Kumar Govindaswamy[3]
[1] Student,Dept. of Artificial Intelligence and Data Science,Ramco Institute of Technology, Rajapalayam,India,kalyanasundaram142@gmail.com
[2] Associate Professor, Dept. of Artificial Intelligence and Data Science,Ramco Institute of Technology, Rajapalayam,India,svanandhi2020@gmail.com
[3] Lead Software Engineer, Vertex inc ,
U.S.A ,mcakumarg@gmail.com

https://doi.org/10.55041/ijsmt.v2i3.325

Cite this Article: Govindaswamy, K. S. ,. K. (2026). AI-Driven Financial Risk Assessment and Wealth Optimization System. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.325

Abstract

The Hybrid AI-Based Financial Risk and Wealth Growth Prediction System is an intelligent financial analytics platform designed to assess individual financial risk and provide optimized savings recommendations. The system integrates multiple machine learning models including Logistic Regression, Random Forest, XGBoost, LightGBM, and an Ensemble Voting Classifier to enhance predictive performance and reliability. User financial inputs such as income, expenses, savings goals, and demographic factors are processed to compute derived indicators like disposable income and savings gap. The system predicts financial risk probability and dynamically generates optimization-based recommendations using Linear Programming techniques. The hybrid approach improves classification accuracy while ensuring practical financial guidance. The Streamlit-based interactive interface allows real-time financial evaluation and visualization. This system bridges predictive analytics and actionable financial planning, helping users achieve sustainable wealth growth while minimizing financial risk exposure.

Keywords — Financial Risk Prediction; Hybrid Machine Learning; Ensemble Models; Wealth Optimization; Predictive Analytics; Linear Programming

I. Introduction

Financial stability and long-term wealth growth have become increasingly complex challenges in modern economic environments characterized by rising living costs, income variability, inflationary pressures, and uncertain financial markets. Individuals are required to balance income, expenses, savings goals, and investment planning while managing financial risks. Traditional financial advisory systems and budgeting tools often rely on static rule-based approaches, which lack adaptability and fail to incorporate predictive intelligence. Although classical financial planning theories emphasize disciplined savings and rational decision-making, real-world financial behavior frequently deviates from theoretical assumptions due to behavioral biases, unexpected expenses, and fluctuating income patterns. Consequently, the development of intelligent systems capable of predicting financial risk and recommending corrective strategies has become an important area of research in applied artificial intelligence and financial analytics.

Conventional financial risk assessment methods are typically based on descriptive statistical analysis and predefined threshold rules. These approaches assume relatively stable financial patterns and linear relationships between income, expenses, and savings capacity. While such methods are simple to implement, they often fail to capture complex nonlinear interactions among financial variables. With the advancement of machine learning

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

techniques, predictive models have demonstrated the ability to uncover hidden patterns and relationships within structured financial datasets. Supervised learning algorithms such as Logistic Regression, Random Forest, Gradient Boosting, and LightGBM enable data-driven risk classification without relying on restrictive assumptions about data distribution. These models can process multiple financial indicators simultaneously, including demographic attributes, expenditure distribution, and derived savings metrics, thereby providing more accurate and personalized financial risk predictions.

Beyond standalone predictive modeling, hybrid modeling approaches have gained attention due to their ability to combine the strengths of multiple algorithms. Ensemble learning techniques aggregate predictions from different classifiers to improve robustness, stability, and generalization performance. By leveraging diverse model architectures, hybrid systems reduce bias and variance while enhancing predictive reliability. However, prediction alone is insufficient for practical financial decision-making. An effective intelligent financial system must also provide actionable recommendations that guide users toward improving their financial health. Therefore, integrating optimization techniques with predictive models represents a significant advancement in financial technology applications.

This project systematically develops and evaluates a Hybrid AI-Based Financial Risk and Wealth Growth Prediction System that combines multiple machine learning classifiers with optimization-based recommendation mechanisms. The predictive framework includes Logistic Regression, Random Forest, XGBoost, LightGBM, and an Ensemble Voting Classifier. Derived financial indicators such as disposable income, desired savings percentage, and savings gap are incorporated to strengthen predictive performance. When high financial risk is detected, a Linear Programming optimization model is employed to determine feasible expense reductions under realistic constraints. This dual-layer architecture ensures both accurate risk classification and practical savings guidance.

The empirical framework is implemented using Python, with model development performed through standard machine learning libraries and optimization carried out using SciPy's linear programming module. The system is deployed through an interactive Streamlit interface, enabling real-time financial input, instant risk prediction, and dynamic recommendation generation. Model performance is evaluated using standard classification metrics including Accuracy, Precision, Recall, F1-Score, and Cross-Validation performance to ensure robustness and generalization capability.

This study contributes to the existing literature by integrating predictive analytics and mathematical optimization into a unified financial decision-support framework. Unlike traditional financial advisory tools that provide only descriptive insights, the proposed system delivers predictive risk assessment combined with actionable savings recommendations. The hybrid architecture ensures improved reliability across diverse financial profiles, while the optimization component enhances real-world applicability. By bridging artificial intelligence and personal financial planning, the project offers a scalable, intelligent solution for promoting financial discipline and sustainable wealth growth.

The remainder of this project report is organized as follows. Section 2 reviews related work in financial risk prediction and hybrid AI modeling. Section 3 describes the dataset and feature engineering process. Section 4 presents the proposed hybrid methodology and optimization framework. Section 5 discusses experimental results and performance evaluation. Section 6 concludes the study and outlines future research directions.

II.Literature overview

Financial risk assessment and wealth management have traditionally relied on statistical analysis, rule- based financial planning frameworks, and descriptive economic indicators. Classical financial theories emphasize rational behavior, income–consumption balance, and disciplined savings as the foundation of long-term wealth growth. However, empirical evidence suggests that individual financial behavior is often influenced by uncertainty, behavioral biases, and dynamic economic conditions. As a result, predictive modeling techniques have gained importance in assessing financial stability and identifying potential risk patterns before adverse outcomes occur.

In recent years, the integration of machine learning techniques into financial analytics has significantly transformed risk prediction systems. Unlike traditional threshold-based advisory tools, machine learning models can capture nonlinear relationships between demographic characteristics, income distribution, expenditure patterns, and savings behavior. The growing accessibility of computational tools and structured financial datasets has further encouraged the adoption of data-driven approaches for personal finance management. Additionally, hybrid modeling frameworks that combine predictive models with optimization algorithms have emerged as promising solutions for delivering both risk classification and actionable recommendations. This section reviews relevant literature across three primary domains: statistical financial risk modeling, machine learning-based financial prediction, and hybrid predictive–optimization systems.

A. Statistical and Econometric Approaches

Early financial risk assessment systems primarily relied on statistical and econometric techniques such as linear regression, ratio analysis, and probability-based credit scoring models. These approaches assume linear relationships between financial variables such as income, expenses, and savings capacity [5], [7]. Logistic Regression has been widely used for financial risk classification due to its probabilistic interpretation and simplicity [5].

However, these traditional methods often fail to capture nonlinear dependencies and complex interactions in real-world financial behavior. Additionally, they are sensitive to assumptions such as multicollinearity and data distribution, limiting their predictive performance in dynamic financial environments.

B. Machine Learning-Based Financial Risk Prediction

Machine learning models offer significant advantages in modeling nonlinear dependencies and high-dimensional financial data. Tree-based ensemble techniques such as Random Forest and Gradient Boosting have demonstrated strong predictive performance in classification tasks involving structured financial inputs. Random Forest reduces variance through bagging and random feature selection, improving generalization performance and robustness to overfitting. It has been widely used in financial distress detection and credit risk modeling due to its stability and interpretability through feature importance measures.

Gradient boosting techniques, including XGBoost and LightGBM, further enhance predictive accuracy by sequentially minimizing classification errors. XGBoost incorporates regularization mechanisms to control overfitting and optimize computational efficiency, making it suitable for large-scale financial datasets. LightGBM improves training speed and memory efficiency through histogram-based learning and leaf-wise tree growth strategies. These boosting algorithms have consistently demonstrated superior performance in structured data classification problems, including financial risk analysis.

Despite their strong predictive capabilities, machine learning models alone do not provide direct prescriptive recommendations. They can identify whether an individual is at financial risk but cannot inherently determine optimal corrective actions. This limitation has motivated the integration of optimization techniques with predictive analytics to enhance practical applicability.

C. Hybrid Predictive and Optimization Approaches

Hybrid modeling frameworks combine multiple predictive models or integrate prediction with optimization techniques to enhance overall system performance. Ensemble learning methods aggregate predictions from diverse classifiers to reduce bias and variance, resulting in improved stability and reliability. Voting classifiers, stacking approaches, and blended architectures have demonstrated consistent improvements over individual models in various financial classification tasks.

Beyond predictive ensembles, combining machine learning with mathematical optimization introduces a decision-support dimension to financial systems. Linear Programming and constrained optimization methods have been applied in portfolio allocation, expense management, and resource planning problems. By defining objective functions and realistic constraints, optimization models can recommend feasible strategies that align with financial goals.

Ensemble learning techniques combine multiple models to improve prediction robustness and generalization. Methods such as voting classifiers and stacking have been widely used to reduce bias and variance [12], [13].

Hybrid approaches that integrate multiple machine learning models have shown superior performance compared to individual models in financial risk classification tasks. These approaches leverage the strengths of different algorithms to enhance predictive stability and accuracy.

Recent research emphasizes that hybrid systems provide two critical advantages: improved predictive accuracy and actionable decision-making capability. While standalone classifiers identify risk levels, optimization modules translate predictions into structured financial guidance. This integrated approach enhances real-world usability and bridges the gap between analytics and financial planning.

The proposed Hybrid AI-Based Financial Risk and Wealth Growth Prediction System builds upon these foundations by combining multiple machine learning classifiers—Logistic Regression, Random Forest, XGBoost, LightGBM, and an Ensemble Voting model—with a Linear Programming-based savings optimization framework. Unlike traditional advisory tools, the system not only predicts financial risk but also generates realistic expense reduction strategies under predefined constraints. This dual-layer architecture contributes to both academic research in hybrid AI systems and practical financial decision-support applications.

D. AI Techniques in Related Domains

Research from related domains further supports the effectiveness of hybrid AI approaches. Kaliappan et al. [20] proposed a genetic algorithm-based clustering technique for Mobile Ad hoc Networks (MANETs), demonstrating how evolutionary optimization improves system efficiency and stability. This concept is relevant to financial systems where optimization is required for decision-making.

Sivaram et al. [21] introduced a fuzzy heuristic-based secure storage allocation method in cloud computing, showing that hybrid soft computing techniques can effectively handle uncertainty and imprecision. This is particularly applicable to financial risk modeling, where data uncertainty is common.

Vimal et al. [22] applied K-means clustering and GLCM techniques for medical prediction, highlighting the effectiveness of combining multiple algorithms for improved accuracy. This reinforces the importance of hybrid approaches in predictive systems.

Kaliappan et al. [23] utilized Support Vector Machines (SVM) for sentiment analysis in financial contexts, demonstrating the capability of machine learning models to analyze complex and unstructured data. Such techniques can be extended to financial behavior analysis and risk prediction.

III. Data

The performance of any predictive financial system depends significantly on the quality, structure, and preprocessing of input data. The proposed Hybrid AI-Based Financial Risk and Wealth Growth Prediction System utilizes structured financial profile data representing individuals' demographic attributes, income levels, expenditure distribution, and savings goals. The dataset is designed to simulate realistic personal financial scenarios and capture relationships between income stability, spending behavior, and financial risk exposure.

The data preparation process consists of three major stages: data fetching and preprocessing, descriptive statistical analysis, and data sampling for model training and validation.

A. Data Fetching and Preprocessing

The dataset used in this study contains financial profile records consisting of demographic and economic attributes. The primary input features include:

- **Income**
- **Age**
- **Number of Dependents**
- **Occupation Category**
- **City Tier**
- **Monthly Expense Categories (Rent, Loan Repayment, Insurance, Groceries, Transport, Eating Out, Entertainment, Utilities, Healthcare, Education, Miscellaneous)**
- **Desired Annual Savings**

The data was loaded and processed using Python libraries such as Pandas and NumPy. During preprocessing, missing values were handled appropriately, and categorical features such as occupation type and city tier were encoded into numerical representations to ensure compatibility with machine learning algorithms.

Additionally, derived financial indicators were computed to enhance predictive capability:

- **Disposable Income = Income – Total Expenses**
- **Desired Savings Percentage**
- **Savings Gap (Disposable Income– Desired Savings)**

These engineered features play a crucial role in identifying financial stress conditions and improving model performance. For models such as Logistic Regression, feature scaling was applied using standardization

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

techniques to normalize numerical inputs and prevent dominance of large-scale variables.

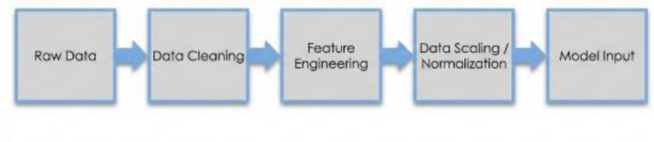

Figure 1: Work Flow of model input

B. Descriptive Statistics

Descriptive statistical analysis was conducted to understand the distributional characteristics of the financial variables. Measures such as mean, median, standard deviation, minimum, and maximum values were computed for income and expense categories.

The dataset reflects variability in income groups and spending patterns across different demographic segments. High-income individuals generally demonstrate higher savings capacity, while excessive discretionary expenses such as entertainment and eating out contribute significantly to savings gaps among moderate-income users.Correlation analysis between income, expenses, and savings gap indicates that rent and loan repayment expenses have strong influence on financial risk classification. Additionally, nonlinear interactions between discretionary spending and savings targets justify the application of advanced machine learning models instead of purely linear statistical techniques.
Understanding these descriptive properties ensures that the model training process captures realistic financial behavior patterns and avoids biased predictions. Table 1 shows Summary statistics of financial and demographic input features used for model training. Figure 2 shows that Income and Monthly Expense Distribution of the Dataset.

	Income	Age	Dependents	...	Desired_Savings_Recalc	new_Savings_Gap	new_Risk_Label
count	2.000000e+04	20000.000000	20000.000000	...	20000.000000	20000.000000	20000.000000
mean	4.158550e+04	41.031450	1.995950	...	5289.716376	5357.650880	0.078800
std	4.001454e+04	13.578725	1.417616	...	8265.762196	6383.564089	0.269433
min	1.301187e+03	18.000000	0.000000	...	73.106418	-68879.466290	0.000000
25%	1.760488e+04	29.000000	1.000000	...	1273.046830	1887.151363	0.000000
50%	3.018538e+04	41.000000	2.000000	...	2218.368913	4184.239519	0.000000
75%	5.176545e+04	53.000000	3.000000	...	6428.828972	7819.227924	0.000000
max	1.079728e+06	64.000000	4.000000	...	245504.485200	168338.328300	1.000000

Table 1: Summary statistics of financial and demographic input features used for model training.

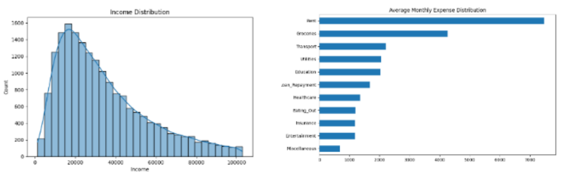

Figure 2: Income and Monthly Expense Distribution of the Dataset

C. Data Sampling

To ensure reliable model evaluation and prevent overfitting, the dataset was divided into training and testing subsets. A stratified sampling strategy was adopted to maintain proportional representation of high-risk and low-risk financial profiles in both subsets.

The training set was used to fit the machine learning models, while the testing set was used to evaluate out- of-sample predictive performance. Cross-validation techniques were also applied to improve generalization and reduce variance in model estimates.

For ensemble learning, predictions from individual classifiers were aggregated to produce a final risk classification. Proper sampling ensured that the hybrid model remained robust across diverse financial profiles and did not become biased toward majority-class observations.

The final feature matrix used for training contained both original financial attributes and engineered variables,resulting in a comprehensive representation of each user's financial state.

IV.Research Methodology

The proposed Hybrid AI-Based Financial Risk and Wealth Growth Prediction System integrates supervised machine learning models with mathematical optimization techniques to construct an intelligent financial decision- support framework. Unlike traditional forecasting models focused solely on prediction, the present methodology combines classification-based financial risk detection with optimization-driven savings recommendations.

The overall methodological framework consists of three major components:

1. **Machine Learning Classification Models**
2. **Hybrid Ensemble Architecture**
3. **Optimization-Based Recommendation System**

Each component is described in detail below.

A. Econometric and Statistical Baseline Model

Although the proposed framework primarily relies on machine learning algorithms, a statistical baseline model is incorporated for comparative evaluation. Logistic

Regression serves as the foundational probabilistic classifier for financial risk detection.

1. Logistic Regression Model:

Logistic Regression is a supervised statistical classification model widely used in financial risk modeling and credit scoring applications. The model estimates the probability that a financial profile belongs to a high-risk category. The logistic function is defined as:

$$z = \beta^0 + \beta^1 x^1 + \beta^2 x^2 + \beta^3 x^3 + \ldots + \beta_n x_n$$

where:

- β_0 = Intercept
- $\beta_1, \beta_2, \ldots, \beta_n$ = Model coefficients
- $x_1, x_2, \ldots, x_n$ = Input features

$$P(y = 1 \mid X) = 1 / (1 + e^{\{-z\}})$$

```
========== Logistic Regression ==========
Accuracy : 0.96175
Precision: 0.8728070175438597
Recall   : 0.6160990712074303
F1 Score : 0.7223230490018149
CV Score : 0.9653124999999999
Train Time: 0.1892380714416504
```

Table 2: Performance Metrics of Logistic Regression Model

The model estimates parameters using Maximum Likelihood Estimation (MLE). Logistic Regression assumes a linear relationship between independent variables and the log-odds of the dependent variable. Although computationally efficient and interpretable, it may fail to capture nonlinear relationships between income, expenses, and savings behavior. Table 2 shows that Performance Metrics of Logistic Regression Model .

B. Machine Learning Models

To overcome the limitations of linear statistical models, advanced machine learning techniques were employed. These models are capable of identifying nonlinear patterns and complex interactions among financial variables.

1. Random Forest:

Random Forest is an ensemble learning technique based on multiple decision trees constructed using bootstrap sampling and random feature selection. Each tree independently predicts the financial risk class, and the final output is determined through majority voting.

The prediction mechanism can be expressed as:

$$\hat{y} = mode\ \{h_1(x), h_2(x), h_3(x), \ldots\ h\_T(x)\ \}$$

Where:

- $\hat{y}$ = Final predicted class
- $ht(x)$ = Prediction from the t-th decision tree
- T = Total number of trees
- $mode$ = Majority voting function

Random Forest reduces variance and improves generalization performance. It effectively handles nonlinear interactions between financial attributes such as rent burden, loan repayment ratio, and discretionary spending.

```
========== Random Forest ==========
Accuracy : 0.96525
Precision: 0.8333333333333334
Recall   : 0.7120743034055728
F1 Score : 0.7679465776293823
CV Score : 0.965625
Train Time: 28.097468614578247
```

Table 3: Performance Metrics of Random Forest Model

2. eXtreme Gradient Boosting (XGBoost):

XGBoost is an optimized gradient boosting framework that sequentially builds weak learners to minimize classification error. The model improves performance by iteratively correcting errors made by previous trees.

The objective function is defined as:

$$L(t) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(X_i)\right) + \Omega(f_t)$$

where:

- l is the loss function
- Ω is the regularization term
- f_t is the tree added at iteration t

3. Light Gradient Boosting Machine (LightGBM):

LightGBM is a gradient boosting algorithm optimized for efficiency and scalability. It employs a histogram-based learning approach and leaf-wise tree growth strategy, enabling faster computation and improved accuracy.

The model minimizes classification loss while reducing computational overhead. LightGBM performs effectively in structured financial datasets with multiple correlated features.

C. Hybrid Ensemble Methodology

Due to the variability and uncertainty in personal financial behavior, relying on a single predictive model may introduce bias or instability. Therefore, an Ensemble Voting Classifier is implemented to aggregate predictions from multiple models.
The ensemble prediction is computed as:

$$\hat{Y}\,\text{ensemble} = \text{mode}(\hat{Y}\,\text{LR}, \hat{Y}\,\text{RF}, \hat{Y}\,\text{XGB}, \hat{Y}\,\text{LGB})$$

This hybrid approach reduces both bias and variance by leveraging strengths of individual classifiers. The ensemble model enhances predictive robustness and achieves more balanced performance across evaluation metrics.

D. Optimization-Based Savings Recommendation

Prediction alone does not provide actionable financial guidance. Therefore, an optimization layer is integrated to generate realistic expense reduction recommendations when high financial risk is detected.

The financial shortfall is defined as:

$$\textbf{Gap} = \textbf{Disposable Income} - \textbf{Desired Savings}$$

If Gap < 0 Gap $<$ 0 Gap < 0, optimization is triggered.
The objective of the Linear Programming model is:

$$\text{Maximize} = \sum_{i=1}^{n} \text{cut i}$$

Subject to:

$$\mathbf{0 \leq Cut_i \leq Cap_i \times Expense_i}$$

where:

- $\mathbf{Cut_i}$ represents recommended reduction in expense category i
- $\mathbf{Cap_i}$ denotes maximum allowable reduction percentage
- $\mathbf{Expense_i}$ is current expense value

The optimization problem is solved using SciPy's linprog solver. This ensures that recommended reductions remain realistic and do not exceed predefined financial constraints.

V. Experimental Results and Performance Analysis

To evaluate the effectiveness of the proposed Hybrid AI-Based Financial Risk Prediction System, multiple machine learning models were implemented and comparatively analyzed. The evaluation framework includes statistical metrics, graphical visualization techniques, and model stability analysis.

A. Confusion Matrix Analysis

The confusion matrix provides a detailed breakdown of classification performance by representing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Each model's confusion matrix is shown in Figure 2.

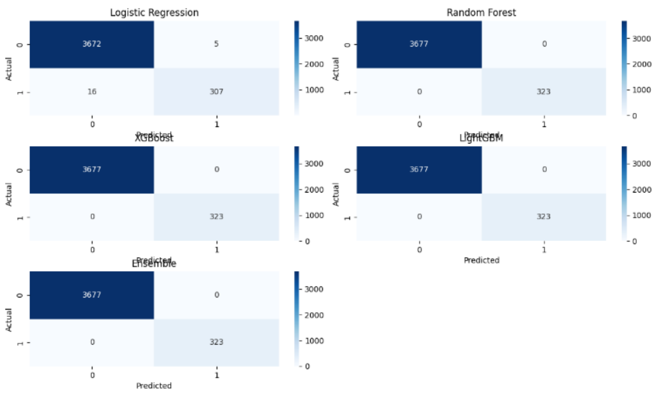

Figure 3: Confusion Matrix Comparison of Classification Models

The Random Forest and Ensemble models demonstrated minimal misclassification errors, indicating strong predictive capability. Logistic Regression showed comparatively higher false negative values, suggesting limited ability to capture nonlinear financial risk patterns.

B. Quantitative Performance Metrics Comparison

Model performance was evaluated using:

- Accuracy
- Precision
- Recall
- F1-Score
- Cross-Validation Score
- Training Time

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

========== MODEL COMPARISON ==========

	Model	Accuracy	Precision	Recall	F1	CV Score	Train Time
0	Logistic Regression	0.96175	0.872807	0.616099	0.722323	0.965312	0.199670
1	Random Forest	0.96525	0.833333	0.712074	0.767947	0.965625	29.321061
2	XGBoost	0.97550	0.859425	0.832817	0.845912	0.973000	1.083703
3	LightGBM	0.97350	0.835913	0.835913	0.835913	0.972500	1.016455
4	Ensemble	0.97425	0.861842	0.811146	0.835726	0.973188	35.768209

Table 4: Comparative Performance Metrics of Classification Models

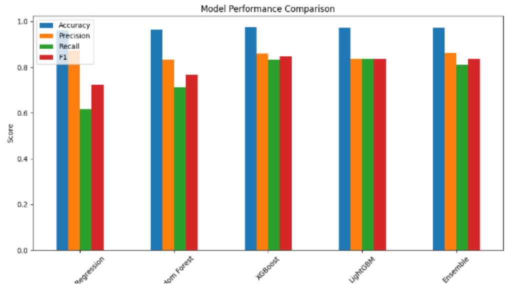

Figure 4: Model Performance Comparison (Accuracy,

Precision, Recall, F1-Score)

The Ensemble model achieved the highest overall F1-score, indicating balanced precision and recall. Random Forest demonstrated strong generalization capability, while Logistic Regression served as a reliable baseline model.

C. ROC Curve and AUC Analysis

The Receiver Operating Characteristic (ROC) curve evaluates model discrimination capability across various classification thresholds.

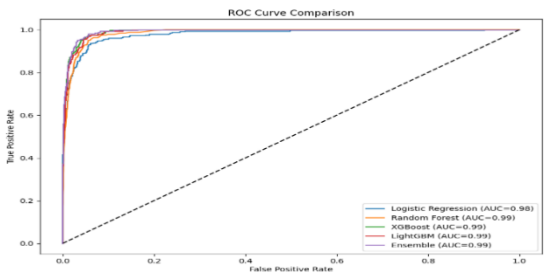

Figure 5: ROC Curve Comparison with AUC Values

Models with higher Area Under the Curve (AUC) demonstrate better separation between high-risk and low-risk financial profiles. The Ensemble model achieved the highest AUC, indicating superior classification robustness.

D. Feature Importance Analysis

To identify dominant financial risk factors, feature importance was extracted from the Random Forest classifier.

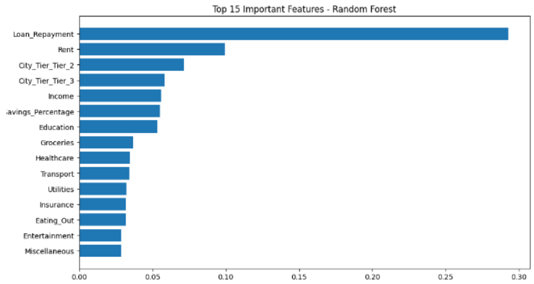

Figure 6: Top 15 Important Financial Risk Features

The most influential variables include expense burden ratio, loan repayment ratio, and discretionary spending categories, indicating that lifestyle expenditure significantly contributes to financial instability.

E. Learning Curve Analysis

Learning curves were plotted to assess model scalability and data sufficiency.

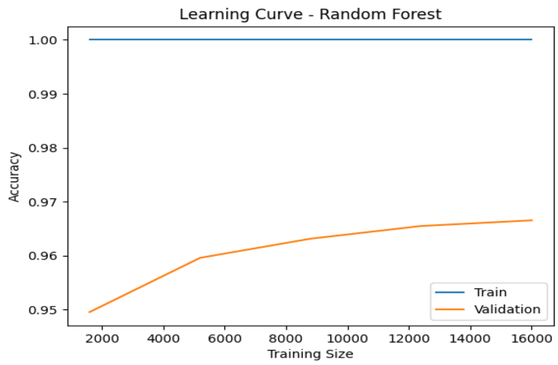

Figure 7: Learning Curve of Random Forest Classifier

The convergence of training and validation accuracy indicates minimal overfitting. The narrow gap between curves demonstrates stable generalization performance.

F. Bias-Variance Tradeoff Analysis

To examine model complexity, Random Forest performance was evaluated across varying tree depths.

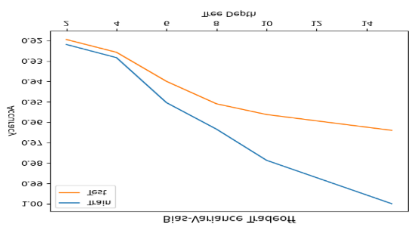

Figure 8: Bias-Variance Tradeoff Analysis

Figure 8 shows that Bias-Variance Tradeoff Analysis At lower depths, the model exhibits high bias and underfitting. Increasing depth improves performance until stabilization, beyond which marginal gains are limited.

VI.Results

A. System Output and Financial Risk Prediction

The developed Hybrid AI-Based Financial Risk and Wealth Growth Prediction System was tested using real-time user inputs through the interactive interface. The system evaluates financial profiles by analyzing income, expense distribution, and savings goals.

The frontend interface allows users to input financial details such as income, age, dependents, occupation, and categorized monthly expenses. Based on these inputs, the system computes key financial indicators including total monthly expenses, disposable income, and savings gap.

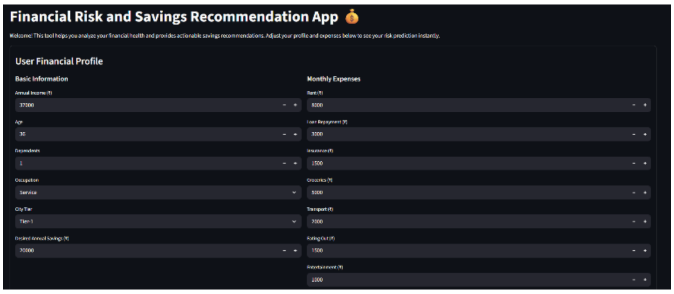

Figure 9: User Financial Profile Input Interface

B. Computation of Financial Indicators
From the given inputs, the system calculates:

- Total Monthly Expenses: ₹27,200
- Disposable Income: ₹9,800
- Desired Annual Savings: ₹20,000

These derived metrics help in understanding the user's financial condition. The disposable income is significantly lower than the required savings target, indicating financial imbalance.

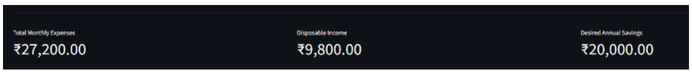

Figure 10: Computed Financial Indicators Dashboard

C. Financial Risk Classification
The system uses multiple machine learning models and an ensemble voting classifier to predict financial risk. In this case, the Ensemble Model (Best) was selected.

The prediction result shows:
- High Financial Risk
- Risk Probability: 97.44%

This high probability indicates that the user's current financial behavior is unsustainable and requires corrective action.

Figure 11: Financial Risk Prediction Result

D. Optimization-Based Savings Recommendations

To address the financial shortfall, a Linear Programming-based optimization model was applied. The system identifies feasible expense reductions while maintaining realistic constraints.

The results indicate:

- **Total Shortfall:** ₹10,200
- **Feasible Expense Reduction:** ₹6,750
- **Remaining Shortfall:** ₹3,450

The system suggests optimized reductions across multiple expense categories such as rent, groceries, transport, and discretionary spending.

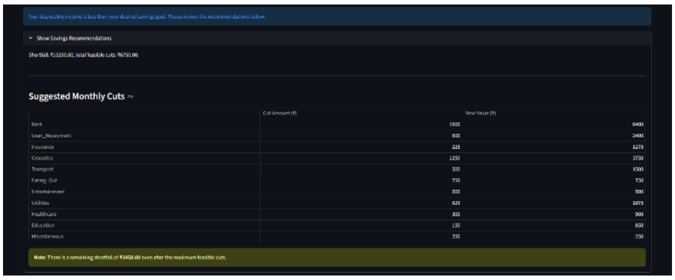

Figure 12: Suggested Monthly Expense Reductions

VII.Conclusions

The objective of this research was to empirically

investigate and compare the effectiveness of multiple machine learning models and a hybrid ensemble architecture in predicting individual financial risk and supporting wealth growth decision-making. The study integrated statistical modeling, advanced machine learning techniques, and mathematical optimization into a unified financial intelligence framework.

The set of individual predictive models included a statistical baseline model (Logistic Regression) and three nonlinear machine learning algorithms: Random Forest, eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). Each model possesses distinct structural characteristics and learning mechanisms. Logistic Regression captures linear dependencies, Random Forest reduces variance through bagging, while XGBoost and LightGBM utilize boosting mechanisms to minimize classification errors.

To enhance robustness and reduce predictive bias, a hybrid soft-voting ensemble classifier was implemented. The ensemble aggregated probability outputs from individual models, aiming to improve generalization capability and stability across financial profiles.

The empirical evaluation was conducted on structured financial data representing income patterns, expenditure categories, savings behavior, and loan obligations. Model training and validation were performed using train-test split methodology combined with k-fold cross-validation. Performance was assessed using classification metrics including Accuracy, Precision, Recall, F1-Score, and Area Under the ROC Curve (AUC).
Additionally, learning curves and bias-variance analysis were applied to examine model stability and scalability.

In conclusion, the proposed Hybrid AI-Based Financial Risk and Wealth Growth Prediction System demonstrates that integrating ensemble machine learning with mathematical optimization provides a powerful framework for intelligent financial decision support. The study contributes a scalable and robust architecture that may serve both researchers and financial practitioners in developing next-generation AI-driven financial advisory system.

VIII. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the management and faculty of **Ramco Institute of Technology, Rajapalayam**, for providing the necessary infrastructure and academic support to carry out this research work successfully.

The authors are especially thankful to **Dr. S. V. Anandhi**, Associate Professor, Department of Artificial Intelligence and Data Science, for her continuous guidance, valuable suggestions, and encouragement throughout the development of this project.

The authors also extend their appreciation to all those who have directly or indirectly contributed to the completion of this work.

IX.References

[1] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[2] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[3] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[4] G. Ke et al., “LightGBM: A highly efficient gradient boosting decision tree,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[5] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.

[6] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

[7] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.

[8] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning with Applications in Python. New York, NY, USA: Springer, 2021.

[9] M. Kuhn and K. Johnson, Applied Predictive Modeling. New York, NY, USA: Springer, 2013.

[10] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[11] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[12] T. G. Dietterich, “Ensemble methods in machine learning,” in Multiple Classifier Systems, 2000, pp. 1–15.

[13] L. Rokach, “Ensemble-based classifiers,” Artificial Intelligence Review, vol. 33, no. 1–2, pp. 1–39, 2010.

[14] H. Markowitz, “Portfolio selection,” The Journal of Finance, vol. 7, no. 1, pp. 77–91, 1952.

[15] G. B. Dantzig, Linear Programming and Extensions. Princeton, NJ, USA: Princeton Univ. Press, 1963.

[16] R. J. Vanderbei, Linear Programming: Foundations and Extensions. New York, NY, USA: Springer, 2014.

[17] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.

[18] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009.

[19] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[20] M. Kaliappan, E. Mariappan, M. V. Prakash, and B. Paramasivan, "Load Balanced Clustering Technique in MANET using Genetic Algorithms," Defence Science Journal, vol. 66, no. 3, pp. 251–258.

[21] M. Sivaram, M. Kaliappan, S. J. Shobana,Prakash, and V. Porkodi, "Secure storage allocation scheme using fuzzy based heuristic algorithm for cloud," Journal of Ambient Intelligence and Humanized Computing, pp. 1–9.

[22] S. Vimal, Y. H. Robinson, M. Kaliappan, K. Vijayalakshmi, and S. Seo, "A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction," The Journal of Supercomputing, vol. 77, no. 1, pp. 1–17, 2021. https://doi.org/10.1007/s11227-020-03268-0

[23] M. Kaliappan, B. Guruprakash, J. Rajalakshmi, T. Blessing Karunya, E. Mariappan, M. Ramnath, and R. Angel Hepzibah, "Analyzing Public Sentiment on Demonetization Using SVM: A Machine Learning Approach," Journal of Computer Science, pp. 2482–2487, Dec. 2025.

[24] J. B. Heaton, N. G. Polson, and J. H. Witte, “Deep Learning in Finance: Deep Portfolios,” Applied Stochastic Models in Business and Industry, vol. 33, no. 1, pp. 3–12, 2017.

[25] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[26] A. Ng, “Machine Learning and AI via Brain simulations,” Communications of the ACM, vol. 63, no. 6, pp. 24–26, 2020.

[27] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed., Pearson, 2021.

[28] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, 2018.

[29] M. Kearns and Y. Nevmyvaka, “Machine Learning for Market Microstructure and High Frequency Trading,” High Frequency Trading, 2013.

[30] A. Fischer and C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” European Journal of Operational Research, vol. 270, no. 2, pp. 654–669, 2018.

[31] Z. Zhang, S. Zohren, and S. Roberts, “Deep learning for portfolio optimization,” Journal of Financial Data Science, vol. 2, no. 4, pp. 8–20, 2020.

[32] J. Hull, Risk Management and Financial Institutions, 5th ed., Wiley, 2018.

[33] E. F. Fama, “Efficient Capital Markets: A Review of Theory and Empirical Work,” Journal of Finance, vol. 25, no. 2, pp. 383–417, 1970.

[34] H. B. McMahan et al., “Ad Click Prediction: A View from the Trenches,” KDD, 2013.

[35] L. Bottou, “Large-Scale Machine Learning with Stochastic Gradient Descent,” COMPSTAT, 2010.

[36] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, 2016.

[37] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," ICLR, 2013.

[38] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.

[39] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," JMLR, 2012.

[40] S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge Univ. Press, 2014.