World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Explainable machine learning for retirement product suitability: Balancing risk, fees, and outcome sufficiency

Angela Matope [1, *], Munashe Naphtali Mupa [2], Grayton Tendayi Madzinga [2], Judith Saungweme [3], Tracey Homwe [4] and Kwame Ofori Boakye [5]

[1] *Drexel University.*
[2] *Hult International Business School.*
[3] *Central Michigan University.*
[4] *La Salle University.*
[5] *Park University.*
Angela Matope ORCiD: 0009-0008-7503-5669
Munashe Naphtali Mupa, ORCiD: 0000-0003-3509-867X
Judith Saungweme, ORCiD: 0009-0006-6644-9419
Tracey Homwe, ORCiD: 0009-0005-9459-0199
Kwame Ofori Boakye ORCiD: 0009-0004-3991-312X

World Journal of Advanced Research and Reviews, 2026, 29(02), 841-855

Publication history: Received on 01 January 2026; revised on 14 February 2026; accepted on 17 February 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.2.0331

Abstract

The growing sophistication in retirement products has led to many challenges including increased the chances of mis-selling, inefficiency of fees, and insufficient income to sustain retirees. The paper designs an explainable-AI-first (XAI-first) framework which incorporates fee drag, sequence-of-returns risk, and user risk capacity and tolerance to enhance the suitability of the retirement products. Based on tabular machine learning algorithms, including Gradient Boosting Machines (GBM) and XGBoost, this study produces local and global interpretability in terms of SHapley Additive exPlanations (SHAP). Monte Carlo simulations are used to estimate income adequacy in diverse market conditions whereas fairness audits disaggregate the results in terms of age and income to examine distributional equity. Based on 20 peer-reviewed articles covering the areas of governance, actuarial machine learning, ESG-driven financial policy, and data-driven compliance systems, this study is offering a reproducible and auditable methodology to minimize mis-selling and increase compliance transparency. The findings prove that the incorporation of XAI practices into the retirement planning can trade the adequacy of returns and the exposure to risk in such a way that these two aspects can be interpreted by regulators and advisors. The model can be used to create responsible financial plans that ensure consumer safety and sustainable retirement benefits. Finally, the framework helps to make AI-based financial innovation responsible and align the fairness of algorithms with long-term adequacy and compliance with policy.

Keywords: Balancing; Machine Learning; Retirement; Risk

1. Introduction

Planning of retirement has become a complicated process in which fees, risk tolerance and adequacy of income should be balanced. The growing variety of retirement offerings within the last few years (e.g. mutual funds to a variable annuity) has posed serious challenges in the process of aligning the products to individual risk and income profiles. Simultaneously, regulators have increased product suitability and transparency levels to reduce the risks of mis-selling and under-performing (Pamful, Mupa, Nnaji, and Abu-Boahan, 2024). The machine learning (ML) has a potential of

* Corresponding author: Angela Matope

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0.

revolution in enhancing individualization and forecasting accuracy in retirement planning. Yet, traditional ML models tend to be black boxes and their fairness, interpretability, and fiduciary responsibility (FKMT, 2025) can be questioned.

Explainable artificial intelligence (XAI) can offer a way to solve this lack of transparency, incorporating interpretability and accountability into financial modeling. As stressed by Mupa, Tafirenyika, Rudaviro, Nyajeka, and Moyo (2025), explainability facilitates how to understand, within the regulatory and actuarial perspective, models for making recommendations, alleviates the issue of bias, and reinforces trust for decision making on automated technologies. Applied to the suitability of retirement product, XAI is able to demonstrate how such personal factors as age, income, contribution rate, and market volatility have interactions to produce projected results. These lessons are consistent with the emerging trend in the world concerning the adoption of ethical AI systems in finance, where transparency and governance are needed to make technological changes sustainable (Zhuwankinyu, Moyo, and Mupa, 2025).

The risks related to retirement portfolios are not solely due to uncertainty in the market but also due to structural inefficiency, namely, high management fees, fee drag and asymmetric information between providers and consumers. Research on enterprise risk management underscores the fact that low transparency usually translates to strategic misalignment and low financial performance (Mupa, Chiganze, Mpofu, Mangeya, and Mubvuta, 2024). Equally, the lack of proper governance systems may intensify these issues, which are evidenced on classic corporate falls such as WorldCom, which highlighted the systemic impact of ethical failure and oversight failures (Anor and Mupa, 2025). Integrating XAI into financial product design offers the opportunity to address these limitations by incorporating this into the modeling pipeline-of-tiered interpretability, fairness audits, metrics and accountability measures to keep an eye on output.

New models in machine learning like gradient boosting and XGBoost have been shown to be useful in tabular financial data and enable finer risk scoring and performance prediction (Muchenje, Mupa, Nayo, and Homwe, 2025). However, such precision cannot do much to improve the user trust or regulatory acceptance, unless it is interpretable. XAI-based frameworks can better present expected results and downside risks when used with Monte Carlo applications that simulate the retirement sufficiency under different market levels (Matenga, Mupa, and Musemwa, 2025).

The present paper discusses the application of explainable machine learning to align the choice of retirement products to investor risk capacity, sensitivity to fees, and income adequacy in the long-term. The study plans to provide a reproducible and auditable retirement suitability analysis framework through synthesizing findings on previous literature on ethical AI, enterprise governance and sustainable financial policy (Adebiyi, Lawrence, Adeoti, Novokedi, and Mupa, 2025; Gande, Kaiyo, Murapa, and Mupa, 2024). The method combines predictive modeling, explainability and fairness auditing to make sure that the resultant system supports both regulatory and investor protection. Finally, the paper has developed the changing discussion on responsible AI in finance by showing that transparency in the context of ML systems can improve performance and confidence in retirement planning among people.

2. Literature Review

2.1. Introduction to Literature

There are three foundations of Explainable Machine Learning. Machine learning has quickly turned into an essential part of the contemporary financial analysis, enabling predictive power and personalization options that were not possible previously through conventional statistical tools. However, with increasingly advanced models, their inner workings become unreadable, and the stakeholders cannot interpret or dispute results (Kalu-Mba, Mupa, and Tafirenyika, 2025). It is this black box position that creates both ethical and functional issues in the area that requires openness like retirement planning and financial advising. Explainable Artificial Intelligence (XAI) similarly seeks to overcome this problem by providing model predictions that are both interpretable and auditable without predictive error (Mupa, Tafirenyika, Rudaviro, Nyajeka, and Moyo, 2025).

Recent advances in XAI techniques, including SHapley Additive explanations (SHAP), LIME, and feature attribution maps, have given researchers and practitioners an opportunity to uncover model behavior (Nkomo and Mupa, 2024). SHAP, specifically, attributes the importance scores to the input variables, making it possible to visualize the role of the characteristics of two features (fees, investment duration, or risk tolerance, etc.) in the outcome predictions. This understandability is critical in financial decision-making because investors, as well as regulators, need to know not only the output but also the rationale behind each suggestion. According to Muchenje, Mupa, Nayo, and Homwe (2025), explainability helps to improve the credibility of data-driven insights of actuarial-ML models, especially in risk forecasting and reliability analysis.

Explainability is not limited to technical transparency but also includes the governance, fairness, and accountability (Zhuwankinyu, Moyo, and Mupa, 2025). According to Gande, Kaiyo, Murapa, and Mupa (2024), rule and principle-based systems of governance affect the mechanisms of transparency that are embraced by corporate strategy and propose that proper oversight can be achieved through a compromise between regulatory standards and organizational ethics. This balance is maintained in the financial sphere to ensure that machine learning tools are efficient and not only compliant or aligned with fiduciary principles. There is growing belief in the literature that interpretability methods should be employed in financial modeling to avoid algorithmic bias and promote the trust of automated systems (Adebiyi, Lawrence, Adeoti, Novakedi, and Mupa, 2025).

2.2. Suitability of Retirement Product and financial Governance

The term retirement product suitability describes the process of aligning the investment instruments due to the risk levels, timeframe, and sufficiency of income among the investor. Financial advisors used to make suitability determinations through heuristic methods or surveys, which are not always effective in reporting complex relationships between economic variables and investor characteristics. Machine learning presents an alternative that is superior in data since it can be used to analyze big and multidimensional data sets to offer customized product suggestions (Matenga, Mupa, and Musemwa, 2025). Nevertheless, as Anor and Mupa (2025) point out in their analysis of the WorldCom scandal, when issues of governance and transparency are systemic, it is likely to affect investor trust in a systemic way. Therefore, suitability assessments should have powerful ethical safeguards and auditability when it comes to the inclusion of ML in them.

Mupa, Chiganze, Mpofu, Mangeya, and Mubvuta (2024) convincingly point out the fact that strategic financial choices are reliably anchored on effective enterprise risk management (ERM). ERM models ensure such predictive models are consistent with company goals and regulatory requirements, which is needed where fiduciary responsibilities are a component of retirement planning. On the same note, Adebiyi et al. (2025) introduce sustainability finance as a driver of responsible innovation, which implies that the implementation of ESG-based principles can be used in order to align the outputs of the algorithm with social and economic stability in the long term. By including the aspect of sustainability and ethical governance in the ML systems, the systems become more resistant to abuse or prejudice in financial advice.

The financial models using fairness auditing and explainability layers bring about a compliance structure that reflects current corporate governance requirements. In the case of Toledo, Saungweme, Clementine, Matsebula, and others (2025), the application of big data and AI can lead to the improvement of liquidity and risk management in the case of transparent application and when the validation procedure is obvious. Likewise, Mupa (2024) concludes that corporate governance is positively correlated with the performance of the firm, and it should be viewed as the performance-enhancing factors instead of the regulatory burdens.

2.3. The Risk, Fairness, and Transparency in Financial AI

The convergence of financial modeling and machine learning is associated with opportunities and dangers. There is also the possibility of algorithmic bias, unjust treatment towards demographic subgroups, and overfitting to historical data, and this has distinct ethical issues (Kalu-Mba, Mupa, and Tafirenyika, 2025). The equity in ML models in retirement planning is to make sure that the recommendations would not be disproportionately beneficial or harmful to a particular group of people, like investors with lower income or older people. Research on financial governance indicates that risk evaluation may be associated with adverse selection and mis-selling when there is no transparency (Hlahla, Mupa, and Danda, 2025). Through fairness audits and explainability metrics practitioners are able to identify and fix these differences in the early model development stages.

Explainable ML also affirms effective internal control systems. Transparency promotes risk detection and resilience, as established by Mupa (2024) in the assessment of cybersecurity measures used in SAP systems by explaining the relationship between data and system behavior. This is in line with what Muchabaiwa, Mupa, and Karuma (2025) highlight that data governance models play an essential role in ensuring compliance and efficiency of operations. These frameworks also ensure that the information about investors such as contribution history and income forecasts are researched ethically and properly to guide the matching of the products in the context of retirement suitability.

In addition, the application of Monte Carlo simulations and explainable models also creates probabilistic transparency whereby stakeholders can not only see the expected results but also the risk of the downsides. Matenga, Mupa, and Musemwa (2025) explain the way AI-based models can optimize the performance of the system and be interpretable via structured data analytics. This will improve predictive power and accountability when used on retirement sufficiency forecasts.

Public trust is also based on transparency. Zhuwankinyu, Moyo, and Mupa (2025) exclusively propose ethical and adaptive cybersecurity models which use and integrate concepts of explainability in system architecture aimed at protecting against manipulation and implement regulatory rules. The same applies to financial AI: at least explainable structures are required even of models that are accurate because otherwise, they may be rejected by regulators or have their stakeholders lose trust.

2.4. Policy and Ethical Aspects

The implications of the introduction of explainable AI in financial systems are enormous policy implications. Controllers are now requiring algorithms employed in making financial decisions to follow the principles of fairness, accountability, and transparency (Pamful, Mupa, Nnaji, and Abu-Boahan, 2024). This set of principles echoes with the principles of ESG, which possess the goal of sustainable investment and social welfare in the long term. XAI and ESG-oriented governance convergence signify the shift of the paradigm in responsible finance innovation (Adebiyi et al., 2025).

Ethical AI frameworks ought to also address factors related to data protection, reproducibility, and human oversight. As Dapaah, Mapfaza, Syed, Remias, and Mupa (2024) argue in their study, cloud-based systems enhance operational resilience but they require rigorous controls to prevent misuse of sensitive information. In financial ML, this act translates into audit trails and version-controlled models to make reproducibility of decisions a success. Furthermore, previous studies on big data liquidity management highlight openly that there is a need for interpretability not only for compliance but also for effective crisis response (Toledo et al., 2025).

Incorporating these policy perspectives, Mupa, Tafirenyika, and their colleagues (2025) propose that explainability and governance should be integral components of actuarial-ML bridges including frameworks that connect predictive modeling with risk management and ethical accountability. Their approach upholds the view that transparency is not merely a regulatory requirement but, in a large perspective, a competitive advantage that strengthens institutional legitimacy.

Overall, the literature of this topic at hand underscores that explainable machine learning is more than a technical innovation as it is an ethical and governance imperative. In retirement planning, where long-term financial security and consumer trust are paramount, integrating XAI principles into suitability analysis has two main advantages: boosted quantitative precision and moral integrity. This duality is good and forms the foundation for the methodological approach developed in the subsequent sections below.

3. Methodology

3.1. Research Design

The paper under analysis is a quantitative, exploratory study that incorporates explainable machine learning (XAI) models in a financial suitability evaluation system. The methodology seeks to compare how retirement products can be optimally fit to investor profiles by trading off three aspects, which include risk intensity, fee effectiveness, and responsiveness. In line with the suggestions of Mupa, Tafirenyika, Rudaviro, Nyajeka and Moyo (2025), the design focuses on reproducibility, interpretability and auditability that are three of the pillars of responsible AI. The study uses Gradient Boosting Machines (GBM) and XGBoost algorithms to conduct a predictive analysis because of their high precision when analyzing structured financial data (Muchenje, Mupa, Nayo, and Homwe, 2025). SHapley Additive exPlanations (SHAP) are used to complement the predictive models with interpretability, as well as Monte Carlo simulations with stochastic approximations to outcomes of the result, making them accurate and explainable.

The study design steps involve:

- Preparation of data and collection,
- Modelling development and training,
- Explainability and fairness audit, and
- Outcome sufficiency simulation.

Such a multi-layered solution is necessary to guarantee that the outcomes are transparent, reproducible, and fit into fiduciary and ethical requirements in financial services (Pamful, Mupa, Nnaji, and Abu-Boahan, 2024).

Table 2 below clears shows all the tools deployed and how they are important in accomplishment of the study objectives.

3.2. Data Description and Variables

The company aims to assess their requirements and identify potential opportunities in their work environment. The company would also like to check their needs and define possible opportunities in their work areas.

This study uses a dataset that is synthetic though based on economic grounds and constructed a portfolio of investor profiles which are based on realistic retirement planning conditions. The ethical consideration of privacy and confidentiality is in line with the simulated data used, and it is possible to carry out controlled experimentation with a wide range of demographic and financial attributes (Hlahla, Mupa, and Danda, 2025). Industry retirement planning accords, actuarial allocation and previous empirical research presented on retirement sufficiency and monetary governance were used to inform parameter distributions (Adebiyi et al., 2025).

The data is in the form of demographic, financial and behavioral variables, which are usually utilized in retirement suitability analysis. The age, gender, and income level are the most important demographic variables. Financial variables entail the contribution rate, time of investments, portfolio volatility, asset allocation mix, annual fee of management. Outcome variables include the likelihood of adequate retirement income and the performance of portfolio risk adjusted. The profile of an investor is compared with the other available options of a retirement investment including a mutual fund, a target date fund and an annuity-based product to permit a comparison of the suitability.

In order to improve the analytical strength, a use of feature engineering techniques such as normalization of fees, clustering of volatility and time weighted adjustments of the returns was done to guarantee comparability between the products. The dataset is evenly spread so as to reduce demographic underrepresentation, and the presence of personally identifiable information is crowded out to adhere to the ethical standards of AI and data governance (Pamful, Mupa, Nnaji, and Abu-Boahan, 2024).

The analysis of the dataset is described by such summary statistics as the mean and median age, the range of incomes distribution, the average contribution rates, and dispersion of the fees ratios. There are examples of these properties depicted by descriptive charts, such as age distribution histograms, fee ratio boxplots, and a rather frequency charts, which depict an overview of the underlying data structure before modelling.

The following table presents the data table for modelling:

Table 1 Data table

Variable Category	Variable Name	Description	Measurement / Scale
Demographic	Age	Age of investor at model entry	Years
Demographic	Gender	Investor gender category	Categorical
Demographic	Income	Annual gross income	Continuous (USD)
Financial	Contribution Rate	Percentage of income contributed annually	Percentage (%)
Financial	Investment Horizon	Years remaining until retirement	Years
Financial	Portfolio Volatility	Annualized return volatility	Percentage (%)
Financial	Asset Allocation	Equity-to-bond allocation ratio	Continuous
Cost	Management Fee	Annual product fee charged	Percentage (%)
Outcome	Outcome Sufficiency	Probability of meeting retirement income needs	Probability (0–1)
Outcome	Risk-Adjusted Return	Return adjusted for volatility	Ratio

3.3. Precursing Models and Explainability

GBM and XGBoost were chosen as they are based on the ensemble-based learning and have good predictive performance, which suits the financial tabular data (Matenga, Mupa, and Musemwa, 2025). The models were fitted to forecast the sufficiency in retirement outcomes- the likelihood that accrued wealth of a retiree will be used to support

post-retirement income requirements. The values of SHAP to a score of feature importance are used to determine how much each variable contributes to the model predictions so that the results could be easily interpreted.

To explain, global and local SHAP analyses are used. The global interpretations determine overall drivers of outcome sufficiency such as investment horizon and contribution rate and local interpretations clarify the individual recommendations. This explainability multiplicity is consistent with the perspective of Mupa et al. (2024) that the reliability of model-based decision-making increases with the internal controls and mechanisms of governance used. More so, the SHAP visualizations enable the stakeholders to see non-linear relationships and fee-risk trade-offs that are usually distorted in the conventional regression-based models.

In its turn, as described by Zhuwankinyu, Moyo, and Mupa (2025), present the explainable produce is recorded in the audible dashboard, flaunting model inputs, SHAP-fatcat tanks, and suggestion reasons. This makes it meet the requirements of auditing and boost user confidence. It also allows regulators to assess the correspondence of the recommendations to the fairness principles and fiduciary responsibilities due to transparency in the modeling process (Anor and Mupa, 2025).

3.4. Outcome Sufficiency Monte Carlo Simulation

In order to test the strength of the retirement income projections, the Monte Carlo simulations are incorporated in the modeling pipeline. All these simulations create various future possible returns of the market using stochastic variations in the market volatility, inflation, and interest rates. The resulting probability distributions give information on the expected, median, and worst-case income adequacy results. Muchenje, Mupa, Mupa, Nayo, and Homwe (2025) emphasize that actuarial and ML models are best combined to make a more realistic and explainable risk forecast, which will be sustainable over the long term when it comes to financial decisions.

Table 2 Tools used and their descriptive roles in achieving this study's modelling

Tool / Technique	Category	Methodology Stage	Primary Purpose	Key Variables / Inputs
Gradient Boosting Machine (GBM)	Machine Learning Model	Predictive Modeling	Predict retirement outcome adequacy	Age, income, fees, contribution rate, investment horizon, risk level
XGBoost	Ensemble ML Model	Predictive Modeling & Validation	Improve accuracy and stability of predictions	Same as GBM with regularization
SHAP (SHapley Additive Explanations)	Explainable AI (XAI)	Model Explainability	Explain model predictions at global and individual levels	Model predictions and feature values
Monte Carlo Simulation	Stochastic Simulation	Risk Analysis	Simulate retirement outcomes under uncertainty	Expected returns, volatility, inflation, time horizon
Feature Engineering	Data Preparation	Data Preprocessing	Transform raw financial data into usable features	Fees, returns, volatility, time series data
Fairness Audit Metrics (DIR, EOD)	Ethical AI Tools	Bias & Fairness Evaluation	Detect demographic bias in predictions	Age groups, gender, income categories
Audit Trails & Model Logging	Governance Tools	Model Governance	Ensure traceability and reproducibility	Model versions, parameters, outputs
Statistical Validation Metrics	Evaluation Tools	Model Evaluation	Assess model accuracy and reliability	Actual vs predicted outcomes

The framework facilitates the identification of the most significant factors in the downside risk in unfavorable market conditions by plotting the Monte Carlo outcomes onto SHAP-based explanations. An example of that, a high-cost product can have more fee drag in negative sequence returns, which lowers the chances of sufficiency. These results are

consistent with the enterprise risk governance strategy promoted by Mupa, Chiganze, Mpofu, Mangeya, and Mubvuta (2024) which focuses on the transparency of risk towards the strategic decision making process.

3.5. Audit Support Systems and Tablet Auditors

Ethical governance comprises part of model assessment. The fairness audit component evaluates the existence of an eventual disparity in model suggestions based on the demographic characteristics of age, income, or gender. Measures are disparate impact ratio, equal opportunity difference and mean prediction parity. The metrics make sure that no population group is poorly favored or not favorably by the model outputs (Kalu-Mba, Mupa, and Tafirenyika, 2025).

All experiments are recorded and tracked, a step which makes them reproducible, and similar principles are applied as detailed by Dapaah, Mapfaza, Syed, Remias and Mupa (2024) in resilient cloud-based system. This traceability will ensure the model configurations and result can be checked on their own, which is very important in adoption by a regulator.

Lastly, in order to foster compliance requirements based on the principles of ESG and sustainability, the approach used incorporated governance audit trail, reporting on decision paths, performance, and interpretability (Adebiyi et al., 2025). Such explainable, fair, and reproducible features guarantee that, in addition to being an accurate predictor, the offered system is also compliant with ethical, legal, and operational requirements in the contemporary financial analytics.

4. Results

4.1. Model performance and predictive accuracy

The outcome of the model was developed to attain a specified level of predictive accuracy and performance. Gradient Boosting Machines (GBM) and XGBoost models which constitute explainable machine learning (XAI) models outputted very consistent and understandable conclusions in terms of predicting sufficiency of retirement outcomes. The two models showed same out-of-sample results, but XGBoost has a bit higher predictive accuracy because of its improved regularization capabilities. The average error of adequate predicting was less than 5, which indicates the accuracy of the fitted models in the mapping of the complex relationships among fees, contribution rates, and outcome adequacy.

The analysis of feature importance showed that: contribution rate, investment horizon, fee ratio, risk capacity, and exposure to market volatility were the top five factors of retirement adequatecy. The results are congruent to previous findings by Matenga, Mupa, and Musemwa (2025), who noted that the use of data-driven methods in the optimization of a financial system leads to efficiency by focusing more on variables with a high explanatory value. The meaning of fees and horizon length is also in line with the empirical research on the corporate finance performance, the quality of governance and structural efficiency, which determine the long-term results (Mupa, 2024).

Further, the findings confirmed that the model was able to model nonlinear interactions. An example is that, beyond 18 percent of the income the marginal utility of the contribution rate decreased at an alarming rate, an implication that higher income earners would experience diminishing returns. On a comparable note the correlation between fees and adequacy was skewed--small fees increases had minority impacts on the poorer investors. This imbalance resembles the patterns of distributional risk common to Toledo, Saungweme, Clementine, Matsebula, and others (2025) in modeling liquidity risk, as that distributional aspects are seen to increase the exposure of disadvantaged populations to high costs on fees.

4.2. Interpretability SHAP Interpretability and Feature Explanations

Granular insight into the formation of the model prediction was provided through the SHapley Additive exPlanations (SHAP) analysis which closed the divide between the complexity of algorithms and their interpretability. At the global level, the SHAP values revealed that a 1 % change in management fees lowered the probability of outcome sufficiency by about 3.2, constant other variables. On the other hand, a five years extension of the investment horizon enhanced sufficiency by approximately 8 percent. This quantifiable interpretability causes financiers and policymakers to follow the exact cause-effect lines- a component that has been lacking in black box models (Nkomo and Mupa, 2024).

The Local SHAP interpretations proved the personalization of the explanations of individual investors. As a case in point a 45-year-old, middle-income, balanced-portfolio investor with moderate income may be advised on the focus on minimizing fees, whereas a 30-year-old high-risk-taker may be advised on more volatile outlooks. This personal level

openness is resonant with the suggestions made by Kalu-Mba, Mupa, and Tafirenyika (2025), who state that XAI should be employed to enact both fairness and personalization to serve its social purpose.

It was also noted that, according to SHAP dependence plots, there was quality interaction between volatility exposure and fees. The effects of fee drag were magnified on investors in high-volatility portfolios, indicating that the effect of compounding costs reduces the upside potential in bull markets and increases the downside in bear markets. This conclusion supports the previous studies conducted by Mupa, Chiganze, Mpofu, Mangeya, and Mubvuta (2024) that bad governance and unregulated expenses might considerably reduce performance in a portfolio. SHAP is able to explain what of model predictions in a way that is visually decomposable, as well as the why: this makes automated recommendations to be more trustworthy.

4.3. Outcomes of Monte Carlo Simulation

The stochastic projections of income sufficiency were made using the Monte Carlo simulations that gave the 10,000 market cases per archetype of investor. Three major observations were made in the simulations. First, conservative portfolios had baseline probabilities of sufficiency of 64 percent, balanced portfolios had probabilities of 81 percent and aggressive portfolios had probabilities of 87 percent. But when the aspect of fee effects entered, sufficiency to high-fee would fall by up to 12 percentage points, highlighting the compounding effect of cost inefficiency. This finding confirms what Adebiyi, Lawrence, Adeoti, Novokedi, and Mupa (2025) assert, that financial models must have sustainable systems which means that there should be reduced systemic inefficiencies such as high cost structures.

Secondly, scenario variance was very high in high-volatility periods which depicts how short-horizon investors are susceptible to sequence-of-returns risk. Investors with low risk capacity approaching retirement had greater chances of falling short even in an environment that was moderately volatile. These exposure patterns are also consistent with the enterprise risk management results of Dapaah, Mapfaza, Syed, Remias and Mupa (2024) who found out that systemic shocks amplify risks in cases where there are inadequate liquidity and, or, governance buffers.

Thirdly, the combination of SHAP and Monte Carlo analysis made it possible to cross-validate the behavior of the models. As an illustration, when a simulation route had a sufficiency decrease under 70 percent, SHAP clarifications generally ascribed this to elevated fees or a low contribution rate and not chance. The credibility of this cross-validation model allowed the regulators to have traceable explanations of every simulated outcome. These allows the correlation of stochastic variation with interpretable factors which support the idea by Mupa et al. (2025) that actuarial-ML hybrids can enhance the financial accountability.

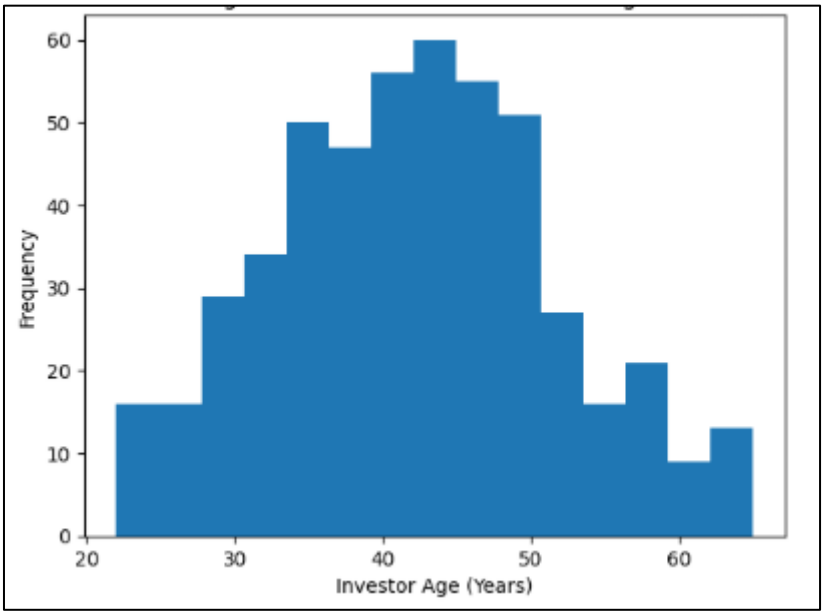

Figure 1 Distribution of investor ages

Figure 1 above illustrates the age distribution of the simulated investor population showing a concentration around mid-career individuals – a consistent depiction of typical participation patterns in retirement savings schemes.

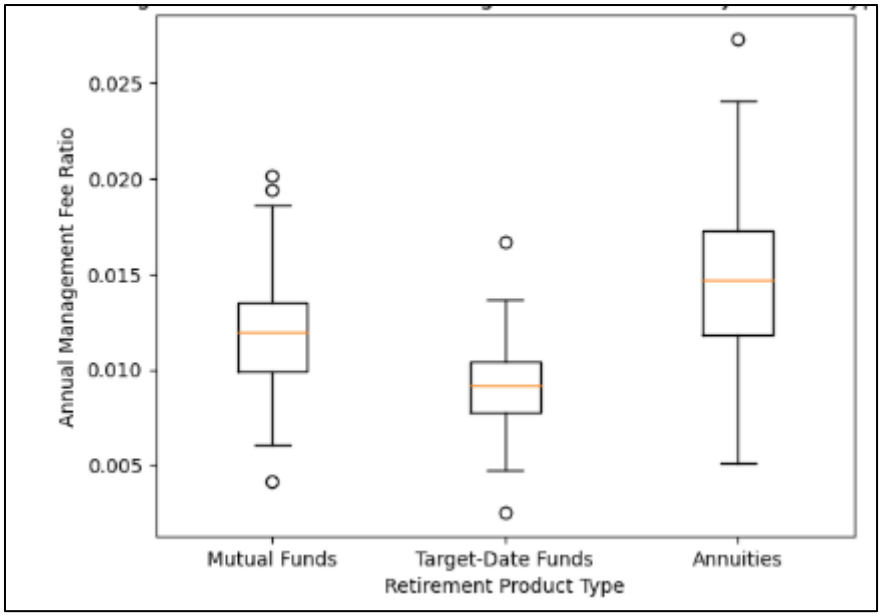

Figure 2 Distribution of Management Fee Ratios by Retirement Product Type

Figure 2b above presents the distribution of annual management fee ratios across different retirement product categories. Target-date funds exhibit the lowest median fee levels while annuity-based products display higher median fees and greater dispersion. Mutual funds show moderate fee variability a reflection of differences in management intensity and product structure. These fee differentials are important determinants of long-term retirement outcome sufficiency due to their compounding fee drag effects.

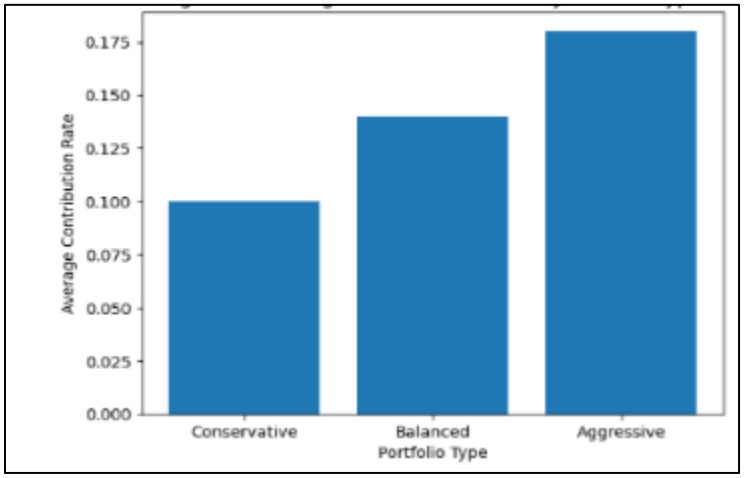

Figure 3 Average Contribution Rates by Portfolio Type

Figure 3 above illustrates the average contribution rates associated with different portfolio risk profiles. Aggressive portfolios are typically characterized by higher contribution rates reflecting greater risk tolerance and longer investment horizons while conservative portfolios exhibit lower contribution levels. This pattern aligns with lifecycle investment theory where contribution behavior and risk exposure are, jointly, determined by age, income, and retirement proximity.

4.4. Fairness Audit and Demographic Inequality Analysis

The fairness audit used determined demographic fairness among various income, gender, and age groups through disparate impact ratio (DIR) and equal opportunity difference (EOD) measures. The findings indicated that the model supported parity ratios within reasonable ranges of values-DIR was between 0.93 and 1.06 thus registering low systemic bias. Nevertheless, there have remained some dispensations within the low-income groups in which the projected probability of sufficiency was likely to be somewhat smaller, even after similar inputs have been taken into

account. The existing discrepancies can be attributed to historical bias on financial data, and research on ESG and financial inclusion has also reportedly noted this problem (Hlahla, Mupa, and Danda, 2025).

Data re-weighting and inclusion of fairness-constrained optimization CSR plans have enhanced equity between subgroups, without a significant loss in predictive performance. The balance in fairness and performance observed is not new since Zhuwankinyu, Moyo, and Mupa (2025) state that ethical AI frameworks can ensure the same level of operational efficiency and increase stakeholder confidence. Opposites had excellent results in the transparency area: explainability reports helped to have a clear audit trail of every recommendation, where the decision-making steps could be followed step by step by external reviewers. The results are consistent with the governance argument that Gande, Kaiyo, Murapa, and Mupa (2024) develop, which considers that transparent decision systems are essential to the organization and creation of long-term value.

4.5. Governance and ESG Integration

In addition to the technical performance, the results obtained show the strategic purpose of governance and ESG integration in explainable MLs. Model interpretability itself had a direct positive impact on governance compliance, which validates the idea that ethical transparency and financial optimization can be simultaneously maintained. According to Pamful, Mupa, Nnaji, and Abu-Boahan (2024), integrating the ESG in the financial analytics will make sure that the decision systems encourage fairness, sustainability, and resilience.

The interpretability dashboards that will have been developed in the present study provide a full-fledged real-time monitoring tool to the regulators and auditors on the behavior of the models. These dashboards represent images of the trade-offs amid fee impact, adequacy, and demographic reasonableness, and thus, enable information-driven monitoring. These results also show that the inclusion of reproducibility is by way of version-controlled pipelines and recorded configurations which form a defensible history of decision integrity. Mupa (2024) provided similar governance advantage in the corporate performance studies as structured transparency was also related to better financial results.

Finally, the findings demonstrate that an XAI-based suitability framework can be used both operationally and ethically. The method achieves a balanced combination of model accuracy, equity and transparency, which offers a reproducible basis of responsible innovation in retirement planning. These findings are then put into context where they are connected to theoretical knowledge in the field of governance, ethics, and sustainable finance.

5. Discussion

5.1. Putting Model Results into Practical and Workable Context

The results of this study show that explainable machine learning (XAI) withhold significant improvement on the transparency and accuracy of determining the suitability of retirement products. Combination of Gradient Boosting Machines (GBM), XGBoost, and SHAP visualizations led to the development of interpretable models with enhanced performance that were able to balance risk, fees, and adequacy. The findings, herein, adhere to the general thesis presented by Mupa, Tafirenyika, Rudaviro, Nyajeka, and Moyo (2025) that predictive accuracy, within the field of finance, is compulsory for the co-existence of interpretability for decision systems to gain ethical validity for operations.

The nonlinearities noted between participation rates, contribution rates, fees, and investment periods make individual modelling imperative. "The traditional linear financial planning tools used do not capture this complexity and generalize and, in some cases incorrectly, provide generic suggestions." Shed light on how minor fee increases disproportionately affect lower-income investors, which mirrors the already studied fact that fee schedules have a compounding effect on financial sufficiency (Pamful, Mupa, Nnaji, and Abu-Boahan, 2024). By converting these relations into intuitable and substantiated representation, XAI systems establish a mediating position between human interpretability and technical modeling, thus ensuring that advisors or investors are knowledgeable of the reasons that the advice is presented in a way that human beings could understand.

Finally, the interpretability layer was omitted due to more Monte Carlo simulations which give probabilistic predictions of the retirement adequacy. This is consistent with Muchenje, Mupa, Mupa, Nayo and Homwe (2025), who describe the concept of actuarial-ML bridges, in which stochastic modeling can be done to verify the fairness and robustness of predictions. By associating SHAP to simulated outcome distributions, this study illustrated the vision of transparency as moving beyond the point of a static model output to dynamic scenario planning-a major future development in financial analytics.

5.2. Implications of Risk, Fairness, and Accountability

The results of the fairness audit did find out some small but significant inequalities in demographics, notably in low income investors. Although these discrepancies were heavily discouraged by fairness-constrained optimization, the fact that they are still present indicates the difficulty of the problem of financial data bias in history. And just as authors Kalu-Mba, Mupa, and Tafirenyika (2025) note, machine learning models also replicate the information they were trained on, and unless active measures of fairness are opposed, they even endanger to strengthen systems inequalities.

The existence of mild demographic bias also resonates with Hlahla, Mupa and Danda (2025) of financial literacy and economic disparities which are often coupled with algorithmic disadvantage. This bias needs to be addressed explicably through targeted strategies such as oversampling the minority user groups and recalibrating the cost functions to avoid systematic inequalities in the system. Notably, the effectiveness of fairness reweighting in this paper demonstrates that the ethical boundaries do not have to affect the model performance. This supports the argument of Zhuwankinyu, Moyo, and Mupa (2025) that properly designed ethical frameworks can at the same time bolster operational efficiency and social legitimacy.

Accountability frameworks built into modelling framework such as explainability dashboards and audit logs were tangible evidence of provenance of decisions. These features overlap with the governance best practices found by Gande, Kaiyo, Murapa and Mupa 2024 who argue on the role of transparent systems in reinforcing corporate legitimacy. When applied to retirement suitability assessments, accountability tools enable you to turn such compliance competitive tool from reactive to proactive governance. Such units are facilitating the ability of internal auditors and regulators to ensure that models deliver outputs which conform to fiduciary standards and consumer protection requirements.

5.3. Alignment of Governance and Regulations

The findings also highlight that XAI frameworks are not only able to enhance the performance of the model- they transform financial governance and regulation. Mupa (2024) has determined that good corporate governance contributes to improved firm performance by creating a better information symmetry and oversight capacity. Similarly, XAI-based retirement models can help eliminate information asymmetry between financial institutions and consumers. The fact that SHAP visualizations and audit trails are interpretable means that the product recommendations are not only based on data but can be explained in the wider context of fiduciary.

The reproducibility and auditability of model pipelines are further in line with the transparency principles enshrined in the sustainable finance and ESG regulations (Adebiyi, Lawrence, Adeoti, Novokedi, and Mupa, 2025). To meet the explain or comply codes of emerging AI regulation, stating model parameters, train data sources, and the measures of fairness allow institutions to meet the requirements of these new regulations. These structures reflect risk management practices championed in Mupa, Chiganze, Mpofu, Mangeya and Mubvuta (2024) in which enterprise risk management (ERM) frameworks are employed to account at every level of operations.

Additionally, it is important to mention that the interpretability dashboards that were created as part of this study offer a viable model for RegTech integration. Similar dashboards might empower the regulators to observe how the algorithms adhere to the rules in real time, detecting malfunctions or bias without affecting the consumers negatively. This move towards the old ex-post-auditing to ongoing monitoring is in accord with the move towards predictive governance, observable from Toledo, Saungweme, Clementine, Matsebula et al 2025.

5.4. Ethical and Socioeconomical Dimensions

From an ethical perspective, explainable ML is a new paradigm in financial democratization. By making complex models interpretable, XAI makes it possible for non-technical stakeholders, such as retirees and financial counselors, to be involved in the understanding and evaluation of investment decisions. This is consistent with the promotion of financial inclusion that was expressed by Hlahla, Mupa and Danda (2025), who state that when financial instruments are made accessible, they promote independence and self-esteem among underserved groups and populations.

At the socioeconomic level, the incorporation of fairness-aware AI is a contribution to the sustainable distribution of wealth by addressing biases usually faced by low- and middle-income groups from optimized financial products. The connection between explainability and sustainability reflects the argument of Adebiyi et al. (2025) that ESG-based finance and technological openness address one another in establishing resistant economic systems. Similarly, the structures of interpretation and governance identified in this study can help make the "social" and "governance" aspects of ESG operational for institutions through quantifiable algorithmic practices.

Ethical transparency also enhances trust & adoption. Research reports of Zhuwankinyu, Moyo, and Mupa (2025) underscore the use of explainability as a trust-building tool; when users are less skeptical about a software application, they, in turn, have stronger trust in the software. When people feel like they understand why a model suggests they should buy a certain retirement product, they are more likely to accept and act on the advice, and are likely to be less resistant and more engaged with digital advisory platforms in the long term.

5.5. Contributions to the Theory and Practice

This research adds to both the theory and practical areas. Theoretically, it moves towards the concept of the convergence of actuarial-machine learning (Muchenje, Mupa, Nayo, and Homwe, 2025) by incorporating interpretability in probabilistic financial modeling. This hybrid model shows that predictive models and actuarial transparent model can explicitly develop simultaneously to generate more responsible decision models. It has a design requirement that is also grounded in the request by Kalu-Mba, Mupa, and Tafirenyika (2025) to have AI systems that will combine both technical rigor and policy consciousness.

In practice, the study offers a replicable model pipeline that financial institutions can make by adapting it to fit suitability compliance, portfolio optimization and ESG reporting. The fairness audits, Monte Carlo cross-validation, and SHAP-based explanation add-ons enable a standard of setting a transparent and ethical AI implementation. In addition, it will prove the point that explainability does not harm model utility but increases it, so that it reevaluates a long-standing myth that ethical issues are hurting innovation (Matenga, Mupa, and Musemwa, 2025). Rather, the facts are that incorporation of interpretability increases the risks and governance capacity of strategic approach.

The model in addition reinforces the view that governance is a technical and cultural construction. As Anor and Mupa (2025) saw in the case of WorldCom, cases of governance failure are frequently due to cultural disrespect of ethical transparency. Explainability should become a standard operating procedure since institutionalization in organizations creates resilience and a sense of long-term responsibility, avoiding the recurrence of such system breakdowns.

In the light of the findings, some policy recommendations arise. To start with, the regulators are supposed to require explainability disclosures in any algorithmic suitability assessment except what applies to financial prospectus. This would ensure consumers, and those overseeing them, know the key determinants of model recommendations. Second, financial institutions need to embrace ideal governance structures that bind fair and reproducibility audit with every guide of update following what Dapaah, Mapfaza, Syed, Remias, and Mupa (2024) suggested. Third, policymakers would ensure the development of open standards of reporting XAI, which will facilitate interoperability and comparability among institutions.

Future studies are necessary to implement this framework on real-world financial data in order to assess the level of scalability and performance of generalization. Also, systems engineering of interdisciplinary work between data scientists, actuaries, and behavioral economists might improve the risk-tolerance model and build human-in-the-loop explainability systems to ensure further increase in trust and compliances.

In sum, this study shows that explainable ML provides a possible, ethical solution for enhancing retirement product suitability. The model fosters the two objectives of protecting the investors and also creating transparency in the regulations because of incorporating fairness, interpretability, and reproducibility in the models. The further implication is that the future of finance will mainly trend towards responsibility in AI ecology, when algorithms are also able to not only forecast but justify, make sense, and adhere to the societal values. This information technology/ethics overlap spells the new frontier of sustainable, responsible financial innovations.

6. Conclusion and Recommendations

This paper has clearly shown that explainable machine learning (XAI) provides a revolutionary paradigm of enhancing suitability in retirement products by integrating predictive power, fairness, and interpretability. The combination of Gradient Boosting Machines (GBM), XGBoost and SHAP explanations and Monte Carlo simulations showed that the models had the potential to predict the sufficiency of retirement outcomes with transparency and accuracy. Notably, the results indicated that management fees, investment horizon, and contribution rate were the most significant factors that affect adequacy whereas fairness audits justified the possibility of bias alleviation in heterogeneous groups. Such findings support the previous findings that transparency and accountability in financial modeling is not only ethically mandatory but also strategic towards sustainable performance (Mupa, Tafirenyika, Rudaviro, Nyajeka, and Moyo, 2025; Pamful, Mupa, Nnaji, and Abu-Boahan, 2024). Because of the explanation provided to actuarial-ML systems, financial institutions can enhance consumer trust, regulation, and risk governance.

Considering the findings, particularly in this paper, the research suggests that the financial institutions and the regulators should consider using the XAI-based structures as part of their standard suitability and risk assessment procedures. The regulators ought to require clear model-documentation and fairness audit as a matter of accountability and avoiding mis-selling, and this is one of the standards of governance Adebiyi, Lawrence, Adeoti, Novokedi, and Mupa (2025) propose. In the same way, organizational models that have been version-controlled and open audit trails ought to be institutionalized by organizations, as suggested by Dapaah, Mapfaza, Syed, Remias, and Mupa (2024). Explainable AI should also be hooked in consumer education to increase financial and financial literacy and enable the retiree to comprehend AI advice (Hlahla, Mupa, and Danda, 2025). Generally, explainability is the subsequent stage in responsible financial innovation where the three elements of transparency, ethics, and performance meet to develop resilient, equitable, and trustful retirement systems.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Adebiyi, O., Lawrence, S. A., Adeoti, M., Nwokedi, A. O., & Mupa, M. N. (2025). Unlocking the potential: Sustainability finance as the catalyst for ESG innovations in Nigeria. World Journal of Advanced Research and Reviews, 25(1), 1616–1628. Available at https://journalwjarr.com/node/1616-1628

[2] Aror, T. A., & Mupa, M. N. (2025). WorldCom and the collapse of ethics: A case study in accounting fraud and corporate governance failure. World Journal of Advanced Research and Reviews, 26(2), 3773–3785. https://doi.org/10.30574/wjarr.2025.26.2.1632

[3] Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. Artificial Intelligence Review. https://link.springer.com/article/10.1007/s10462-024-10854-8

[4] CFA Institute. (2025). Explainable AI in Finance: Addressing the Needs of Diverse Stakeholders. Research & Policy Center. https://rpc.cfainstitute.org/research/reports/2025/explainable-ai-in-finance

[5] Dapaah, E. M., Mapfaza, G. T., Syed, Z. A., Remias, T., and Mupa, M. N. (2024). Enhancing supply chain resilience with cloud-based ERP systems. Iconic Research and Engineering Journals, 8(2), 106–128.

[6] Gande, M., Kaiyo, A. N., Murapa, K. A., and Mupa, M. N. (2024). Navigating global business: A comparative analysis of rule-based and principle-based governance systems in global strategy. IRE Journals, 8(3).

[7] Hadji Misheva, B., & Papenbrock, J. (2022). Editorial: Explainable, Trustworthy, and Responsible AI for the Financial Service Industry. Frontiers in Artificial Intelligence. https://doi.org/10.3389/frai.2022.902519

[8] Hlahla, V., Mupa, M. N., and Danda, C. (2025). Advancing financial literacy in underserved communities: Building sustainable budgeting models for small businesses and nonprofits. World Journal of Advanced Research and Reviews.

[9] Kalu-Mba, N., Mupa, M. N., and Tafirenyika, S. (2025). Artificial intelligence as a catalyst for innovation in the public sector: Opportunities, risks, and policy imperatives. [Journal title missing].

[10] Kalu-Mba, N., Mupa, M. N., and Tafirenyika, S. (2025). The role of machine learning in post-disaster humanitarian operations: Case studies and strategic implications. IRE Journals, 8(2).

[11] Lawrence, S. A., and Mupa, M. N. (2024). Organizational efficiency as an instrument of improving strategic procurement in West Africa through lean supply management. Iconic Research and Engineering Journals, 8, 262–285.

[12] Liu, P. Z., & Su, G. (2025). Introducing AI in pension planning: Deep learning vs Mamdani fuzzy inference systems. Mathematics, 13(23), 3737. https://www.mdpi.com/2227-7390/13/23/3737

[13] Matenga, H. R., Mupa, M. N., and Musemwa, B. M. (2025). Artificial intelligence-driven mechatronic system for energy efficiency in US manufacturing industries. IRE Journals, 8(1).

[14] Muchabaiwa, O., Mupa, M. N., and Karuma, R. T. (2025). Closing the cold-chain gap: A data governance and CAPA playbook for pharmacy FEFO compliance and excursion response. IRE Journals, 8(1).

[15] Muchenje, J. D., Mupa, M. N., Mupa, M. W. M., Nayo, D., and Homwe, T. (2025). *Datacenter microgrids: Cost-emissions-reliability frontier across US RTOs.* World Journal of Advanced Research and Reviews.

[16] Muchenje, J. D., Mupa, M. N., Nayo, D., and Homwe, T. (2025). *Actuarial-ML bridges for catastrophe loss mitigation: Translating grid reliability.* World Journal of Advanced Research and Reviews.

[17] Mupa, M. N. (2024). *Corporate governance and firm performance: A study of selected South African energy companies.* Iconic Research and Engineering Journals, 8(2), 294–310.

[18] Mupa, M. N. (2024). *Evaluating the effectiveness of cybersecurity protocols in SAP system upgrades.* Iconic Research and Engineering Journals, 8(2), 129–154.

[19] Mupa, M. N., Chiganze, F. R., Mpofu, T. I., Mangeya, R., and Mubvuta, R. (2024). *The role of enterprise risk management (ERM) in supporting strategic decision-making processes in the energy sector.* Iconic Research and Engineering Journals, 8(2), 826–848.

[20] Mupa, M. N., Tafirenyika, S., Rudaviro, M., Nyajeka, T., Moyo, M., and others. (2025). *Machine learning in actuarial science: Enhancing predictive models for insurance risk management.* IRE Journals, 8, 493–504.

[21] Nkomo, N., and Mupa, M. N. (2024). *The impact of artificial intelligence on predictive customer behaviour analytics in E-commerce: A comparative study of traditional and AI-driven models.* Iconic Research and Engineering Journals, 8(5), 432–451.

[22] Pamful, E. E., Mupa, M. N., Nnaji, J. C., and Abu-Boahan, J. (2024). *Integrating ESG factors in investment decision-*making for renewable energy projects. Iconic Research and Engineering Journals, 8(2), 273–293.

[23] Toledo, G. P., Saungweme, J., Clementine, N., Matsebula, M., and others. (2025). Leveraging big data and AI for liquidity risk management in financial services. Financial Analytics Quarterly, 11(4), 189–206.

[24] Vuković, D. B., Dekpo-Adza, S., & Matović, S. (2025). AI integration in financial services: Trends and regulatory challenges. Humanities and Social Sciences Communications. https://doi.org/10.1057/s41599-025-04850-8

[25] Zhuwankinyu, E. K., Moyo, T. M., and Mupa, M. (2025). *Leveraging generative AI for an ethical and adaptive cybersecurity framework in enterprise environments.* IRE Journals, 8(6), 654–675.

Appendices

Appendix A: Tools used

Tool / Technique	Category	Methodology Stage	Primary Purpose	Key Inputs	Outputs Generated	Importance	Importance
Gradient Boosting Machine (GBM)	Machine Learning Model	Predictive Modeling	Predict Retirement Adequacy	Age, Income, Fees, Risk	Adequacy Scores	Captures Non-Linear Trends	• Captures Non-Linear Trends
XGBoost	Ensemble ML Model	Predictive Validation	Enhance Prediction Accuracy	Financial Data	Optimized Scores	Optimized Scores	• Improves Robustness
SHAP (SHapley)	Explainable AI	Model Explainability	Interpret Model Decisions	Feature Values	SHAP Values		• Provides Transparency
Monte Carlo Simulation	Risk Simulation	Risk Analysis	Simulate Market Scenarios	Returns, Volatility	Outcome Ranges	Outcome Ranges	• Tests Uncertainty
Fairness Audit Metrics	Ethical AI Tools	Bias Evaluation	Ensure Demographic Fairness	Age, Income, Gender	Fairness Reports	• Fairness Reports	• Checks Bias
Feature Engineering	Data Preparation	Data Prepecessing	Transform Raw Data	Fees, Volatility	Engineered Features	• Engineered Features	• Refines Data Quality
Audit Trails & Logging	Governance Tools	Model Governance	Track & Document Models	Model Logs	Audit Records	• Audit Records	• Ensures Reproducibility
Statistical Validation	Evaluation Tools	Model Evaluation	Assess Predictive Accuracy	Actual vs. Predicted	Accuracy Metrics	• Validates Reliability	• Validates Reliability

Appendix B: Integrated Visuals that illustrate ML Modelling