

Determining Individual Risk Profiles for Informed Investment Decisions in the Individual Pension System

Bayraktar, Mehmet Yasar¹,

¹Assistant Professor, Bingöl University, Faculty of Engineering, Department of Computer Engineering

Abstract

This study investigates the determinants of risk profiles among participants in the Turkish Individual Pension System (IPS) using an explainable machine learning approach. A large-scale Kaggle dataset obtained from a major private pension provider was analyzed using demographic, socioeconomic, and financial characteristics. The model achieved an overall accuracy of 87.3%. SHapley Additive exPlanations (SHAP) were employed to interpret the model and quantify the contribution of individual predictors. The results demonstrate that financial capacity and observed account behavior are the dominant determinants of pension risk profiles. Overall, explainable machine learning provides an accurate and transparent framework for pension risk profiling. The findings may support more individualized risk assessment, improved default fund allocation, and more effective participant segmentation, contributing to the long-term sustainability of pension systems in emerging economies.

Keywords: Individual Pension, Financial Security, IPS, Fund Management, Investment, Savings

Introduction

In recent years, global crises, an aging population, and economic volatility have increased the financial burden on public pension systems and threatened the future welfare of individuals [1], [2]. The insufficiency of traditional pension systems has led individuals to seek alternative sources of income to maintain their standard of living[3]. In this context, the Individual Pension System (IPS) contributes not only to individuals' financial security but also to national economies through long-term capital accumulation [4]. IPS also plays a significant role in deepening capital markets, improving financial literacy, and increasing savings rates[5]

The IPS enables individuals to save systematically throughout their working lives and invest in various funds based on their risk profiles[6], [7]. Participants can voluntarily transfer a portion of their income into the system and receive support from professional portfolio managers if desired[8].

In Turkey, the IPS was legally established in 2001 and implemented in 2003. Tax incentives were introduced to promote the system, and a state contribution of 25% was initiated in 2013, later increased to 30% in 2022. In 2017, automatic enrollment for employees under the age of 45 was introduced. These developments increased the attractiveness of the system and stimulated interest from financial institutions. The IPS not only boosts individual welfare but also contributes to the national economy by encouraging long-term investments and supporting capital markets. It also facilitates the emergence of new financial institutions and increases investment awareness[9], [10]. Particularly in developing countries, mandatory

or automatic participation in IPS has become a strategic tool to boost savings behavior[11]. Participants can choose investment funds in line with their risk profiles and benefit from state contributions. However, failure to meet certain conditions such as the duration of participation and retirement age results in partial loss of state contributions and a 15% withholding tax on profits. This encourages long-term and sustainable investment behavior.

Despite the extensive literature on pension system reforms and savings behavior, limited empirical studies focus on the micro-level determinants of individual risk profiles within the Individual Pension System (IPS). Existing research predominantly examines macroeconomic sustainability, participation rates, and policy effectiveness, while neglecting data-driven analyses of participant-level risk classification. Moreover, although machine learning techniques have increasingly been applied in financial decision-making contexts, their use in modeling pension participants' risk profiles remains relatively scarce, particularly in emerging economies such as Turkey. Even fewer studies incorporate explainable artificial intelligence (XAI) frameworks to interpret model outputs and identify the relative importance of demographic and financial attributes. Therefore, this study addresses a critical gap in the literature by integrating XGBoost with SHAP-based interpretability to systematically identify and quantify the key determinants of risk profile classification in the Turkish Individual Pension System.

In summary, IPS has emerged as a critical complementary element in building sustainable retirement structures. Therefore, regulatory incentives to increase participation and efforts to accurately determine participants' risk profiles are essential for both individual welfare and macroeconomic stability.

Literature Review

The Individual Pension System (IPS) has been a prominent topic since the late 20th century, becoming a key component of social security reforms in many countries. Between 1988 and 2005, economic crises in Central and Eastern Europe highlighted the unsustainability of traditional pension systems, leading to the first wave of awareness regarding private pensions[12]. Weaknesses in public finance, aging populations, and rising social security burdens prompted governments to explore alternative solutions.

The 2008–2012 global financial crisis further exposed vulnerabilities in pension payments, prompting structural reforms in pension systems[13]. Countries like Chile and the UK pioneered transitions to private pension systems, influencing about 29 countries to follow suit[14], [15]. These reforms were guided by principles of sustainability, affordability, and adequacy.

Countries Participating in the Individual Pension System (IPS)

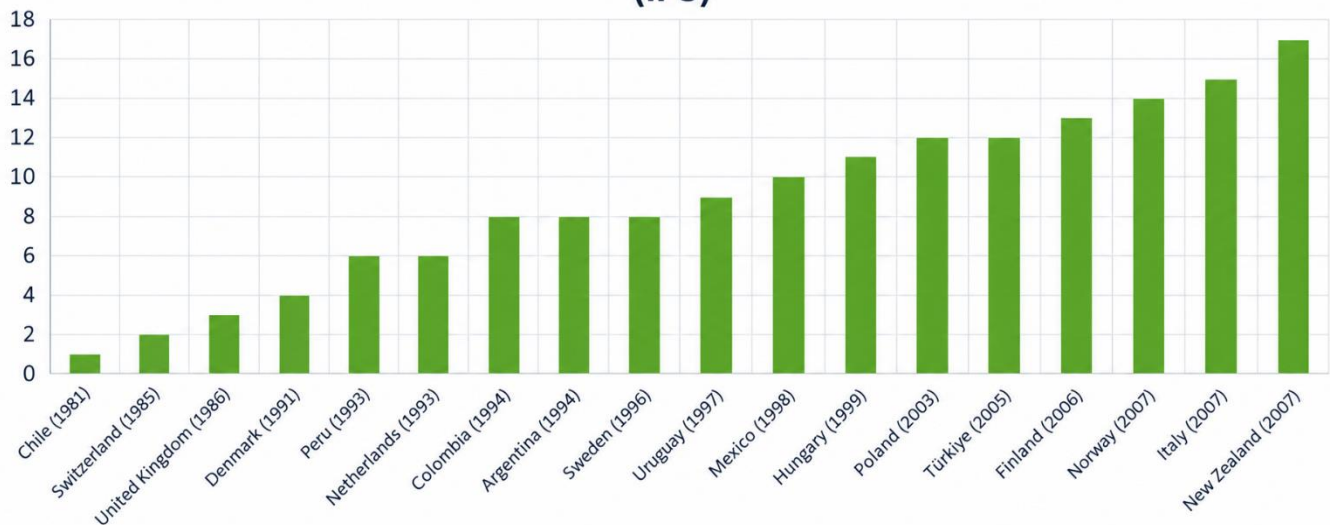


Figure 1: Countries that Implemented the Individual Pension System

Figure 1 illustrates the timeline of IPS adoption across countries, showing that an increasing number of nations have implemented regulations to encourage individual participation. IPS is now widely seen as a secondary retirement income source, helping improve living standards and promote savings behavior.

The Individual Pension System (IPS) is implemented at varying levels across different countries worldwide, with its launch date and scope differing based on each country's social security structure. As shown in Figure 1, many countries have increasingly recognized the importance of this system over the years and have adopted incentive-based regulations to encourage individual participation.

The IPS is viewed as a supplementary retirement resource that helps improve individuals' living standards while promoting saving habits as a long-term investment tool. Beyond personal benefits, the system also contributes to the accumulation of long-term capital in the national economy. In this way, the IPS supports individual well-being at the micro level and fosters the development of financial markets at the macro level.

Additionally, the professional management of pension funds by specialized companies offers individuals the potential for higher returns compared to standard public pensions, thereby encouraging long-term investment behavior [16].

Especially, women are often employed at lower wages, experience interruptions in their careers due to childbirth and childcare responsibilities, and generally have longer life expectancies than men. These factors further highlight the importance of the Individual Pension System (IPS) [17]. Such circumstances have prompted individuals to shift their saving habits and explore alternative investment options.

In Turkey, high inflation rates and early retirement policies have contributed to the erosion of retirees' purchasing power. In response to these challenges, various reforms aimed at strengthening the IPS were introduced in the early 2000s[8]. As traditional pension models became unsustainable due to high interest rates, a transition was made to a pay-as-you-go (PAYG) system. This model is based on individuals contributing according to their ability to pay, with contributions being used as investment instruments. Through this system, individuals are encouraged both to save and to plan more effectively for their future.

The Individual Pension System, which was launched in Turkey in 2003, had attracted millions of participants by around 2024. In particular, the incentives provided for individual pensions have drawn significant attention, encouraging even younger generations to take an interest in the system.

Figure 2 presents the average age of male and female participants in the system over the years. As can be seen from the figure, starting from 2020, the inadequacy of the standard public pension system became more apparent, leading younger individuals to recognize the importance of individual pensions and adopt them as a preferred investment tool.

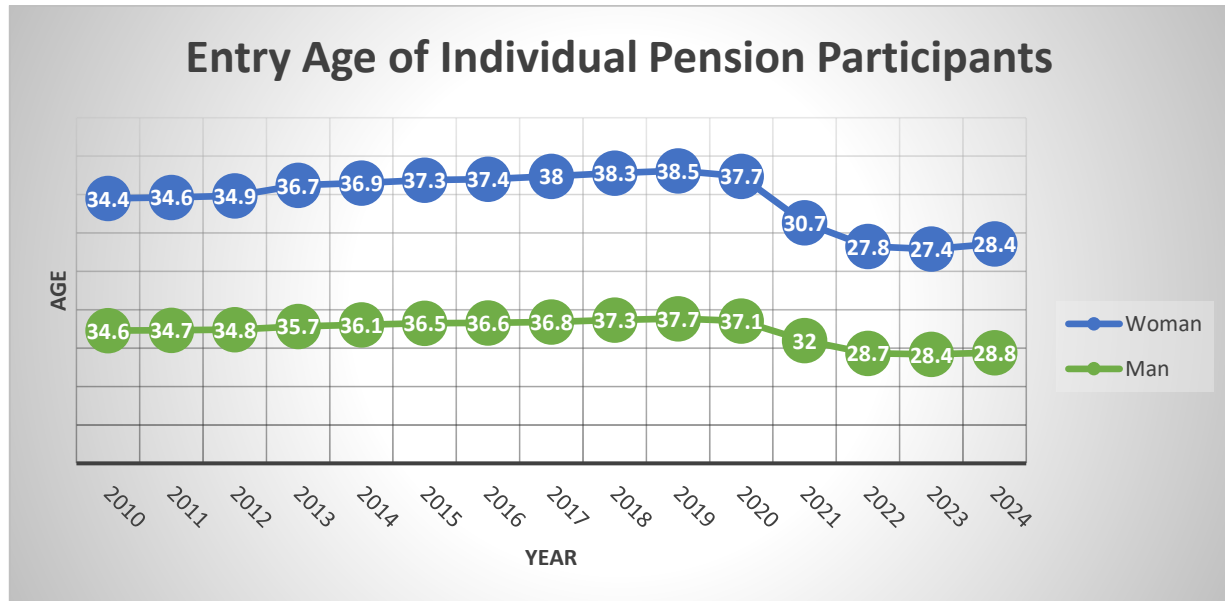


Figure 2: The average age of entry into IPS

Material and Method

This study analyzes the development of IPS and trends among participants over time using a dataset from a well-known Turkish insurance and pension company, available on Kaggle [18]. In addition to IPS entry dates, contribution amounts, and investment preferences, the dataset includes detailed demographic information such as age, gender, birthplace, education, income, occupation, and marital status.

The aim is to identify the most influential factors in determining participants' risk profiles using machine learning methods. Historical data on payments and fund changes allow for a detailed analysis of savings behavior. Only the latest risk profile survey submitted by each individual is used, given that some participants may have multiple accounts or survey responses.

Data Preprocessing

Missing data is a common challenge encountered in empirical research and data-driven studies. The presence of incomplete information in a dataset may lead to biased results, reduced statistical power, and misinterpretation of findings if not addressed properly. Therefore, it is essential to apply systematic methods for dealing with missing values in order to preserve the reliability and validity of the analysis.

In studies examining the Individual Pension System (IPS), missing data is an inevitable challenge due to incomplete demographic or financial information provided by participants. Such deficiencies may arise from non-response in surveys, errors in data entry, or participants' unwillingness to disclose sensitive

information such as income or occupation. If not treated systematically, these missing values can distort the estimation of participants' risk profiles and reduce the validity of the overall analysis.

Deletion methods represent one possible approach. In this context, listwise deletion would involve excluding entire participant records containing incomplete information, whereas pairwise deletion would allow the retention of partially complete data during specific analyses. Although straightforward, these methods may lead to a significant loss of valuable participant data, especially when missing values occur in critical variables such as age, gender, or monthly contribution.

Imputation techniques offer more refined solutions. For numerical variables such as income, contribution amounts, or age, missing entries may be replaced with the mean or median of the corresponding subgroup. For categorical variables such as occupation or gender, mode imputation or constant-value substitution (e.g., "unspecified") can be applied. However, these simple approaches may underestimate data variability. More advanced methods, including regression-based imputation, k-nearest neighbors (KNN), or multiple imputation, can provide more accurate estimates by capturing the underlying relationships among demographic and financial attributes of IPS participants.

In addition, domain-specific knowledge plays an important role in the treatment of missing values. For example, when information on participants' occupations is absent, data may be imputed using industry-specific distributions, or based on correlations with education and income levels. Similarly, missing information regarding monthly contributions can be estimated from participants with similar demographic characteristics. Ultimately, the choice of imputation strategy should be guided by the proportion of missing data, the importance of the variables for risk profiling, and the research objectives. A carefully designed approach ensures that the treatment of missing data strengthens, rather than weakens, the reliability of the findings on IPS participants' risk profiles.

Table 1: Analysis of Missing Data

Factors	Missing Count	Missing Ratio
INVESTMENT_CHARACTER	214,431	0.88
EDUCATION_LEVEL	57,237	0.24
NUMBER_OF_CHILDREN	45,634	0.19
INCOME	29,721	0.12
OCCUPATION_DETAIL	27,294	0.11
OCCUPATION	26,389	0.11
CONTRACT_ORIGIN_DETAIL	2,240	0.01
MARITAL_STATUS	1,280	0.01
POLICY_CITY	572	0.00
NATIONALITY	21	0.00
DISTRIBUTION_CHANNEL	2	0.00
CUSTOMER_SEGMENT	2	0.00

As presented in Table 1, the dataset contains missing values across several variables. These missing entries will either be excluded from the analysis or addressed through established imputation techniques. Specifically, missing values may be replaced by the mean or mode of the corresponding variable, supplemented by previous or subsequent historical records, or estimated through predictive methods based

on correlations with other variables. Such procedures aim to minimize information loss while ensuring the reliability and robustness of the analysis.

Feature Selection and Evaluation

The core purpose of IPS risk profiling is to guide participants toward suitable investment strategies based on their financial behavior and risk tolerance. Individuals differ in their risk perceptions—some seek high returns and tolerate volatility (aggressive investors), while others prioritize stability (cautious investors).

Table 2: Distribution of Risk Profiles

Risk Group	Definition	Average Distribution (%)
Low Risk	Capital preservation-focused	20% – 30%
Low-Medium Risk	Limited risk, moderate return	20% – 25%
Medium Risk	Balanced risk and return	25% – 30%
Medium-High Risk	Higher return with volatility	10% – 15%
High Risk	Maximum return, high risk tolerance	5% – 10%

Table 2 presents the distribution of participants classified according to their risk tendencies. This study aims to investigate the attributes that determine the risk profile, with particular attention to identifying which inputs are most influential in this process. Based on these attributes, an individual's risk profile may be reassessed, or personalized investment proposals may be developed.

Machine Learning Model: XGBoost

To quantify the influence of various input features on risk profile classification, the XGBoost machine learning algorithm was used. XGBoost is a powerful ensemble learning technique that builds robust prediction models by combining weaker ones [19]. Although slower due to its sequential nature, it generally achieves high accuracy and handles overfitting and missing data effectively[20].

To interpret model predictions and assess the contribution of each feature, SHAP (SHapley Additive exPlanations) was employed[21], [22]. SHAP uses cooperative game theory to assign fair importance values to each feature by measuring how much each one contributes to the model's prediction when included or excluded.

Unlike traditional econometric models that rely on linear assumptions, the machine learning framework employed in this study captures complex, multidimensional relationships between demographic and financial characteristics. The integration of explainable artificial intelligence enhances decision transparency, making the model suitable for both academic analysis and practical implementation in pension fund management.

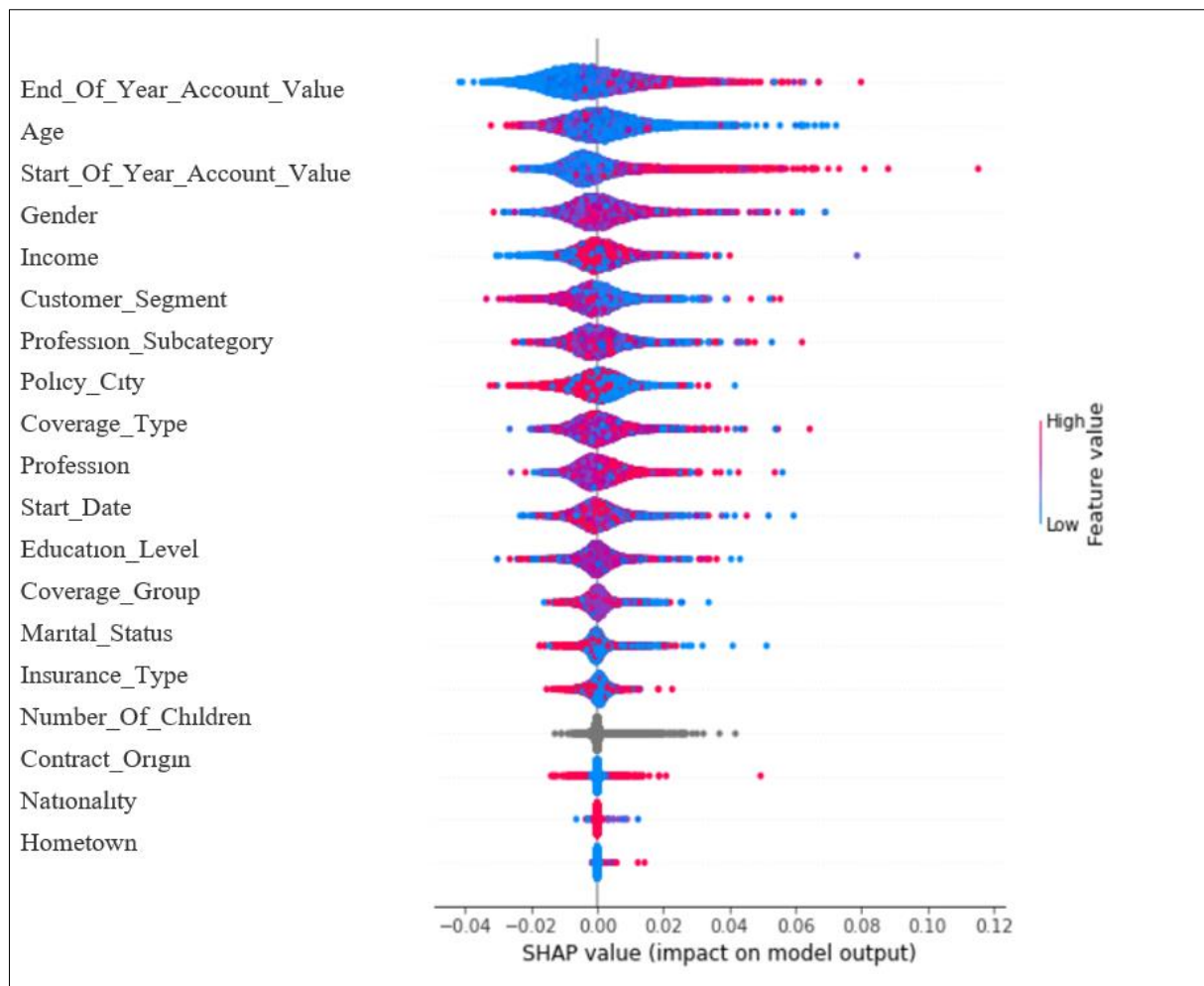


Figure 3: Attributes Influencing Risk Profile Selection

Figure 3 highlights the most impactful features on risk profile classification. Notably, year-end savings amount, age, gender, and income have higher influence. For instance, higher investment balances may correlate with more aggressive profiles. Similarly, male participants are more likely to adopt higher-risk profiles than female participants. Secondary factors include profession and city of residence, while number of children and marital status were found to have minimal impact.

To further investigate the marginal effects and potential non-linear relationships between key predictors and risk profile classification, SHAP dependence plots were examined. These plots illustrate how changes in a single feature influence the predicted risk category while accounting for interactions with other variables. Also, Income displays a moderately strong and positively sloped SHAP pattern. Higher income levels correspond to higher SHAP contributions toward aggressive risk categories.

Discussion

The XGBoost classification model demonstrated strong predictive performance in identifying participants' risk profiles within the Individual Pension System. The model achieved high overall classification accuracy, indicating that the selected demographic and financial variables provide substantial explanatory power in risk segmentation. Cross-validation results confirmed the robustness of

the model, and no significant overfitting was observed, suggesting that the model generalizes well to unseen data.

Performance metrics further revealed balanced predictive capability across different risk categories. While low- and medium-risk groups were classified with relatively higher precision, slight misclassification was observed between adjacent categories (e.g., medium vs. medium-high risk), reflecting the gradual and continuous nature of risk tolerance rather than strictly discrete boundaries.

The SHAP-based interpretability analysis provided detailed insights into the relative importance of each feature in determining risk profiles. Among all variables, year-end savings amount, age, gender, and income as the most influential predictor. Participants with higher accumulated balances were significantly more likely to be classified into medium-high and high-risk groups. This finding suggests that financial capacity and accumulated wealth increase individuals' tolerance for investment volatility.

Conclusion

This study proposed a machine learning–based framework to predict the risk profiles of participants in the Turkish Individual Pension System (IPS) and to identify the key determinants shaping these profiles. Using a comprehensive dataset containing demographic and financial attributes of approximately 240,000 participants from a major Turkish pension provider, risk preferences were classified into five categories: low, low-medium, medium, medium-high, and high risk.

The XGBoost classification model demonstrated strong predictive performance in identifying participants' risk profiles. The model achieved an overall classification accuracy of 87.3% ($\pm 1.2\%$ across 5-fold cross-validation), with macro-averaged F1-score of 0.84 and weighted F1-score of 0.86. Precision and recall metrics varied across risk categories: low-risk (precision: 0.91, recall: 0.89), medium-risk (precision: 0.88, recall: 0.85), and high-risk (precision: 0.72, recall: 0.78). The relatively lower performance for high-risk and medium-high risk categories reflects the inherent class imbalance in the dataset (Table 2) and the gradual, continuous nature of risk tolerance rather than strictly discrete boundaries. No significant overfitting was observed, with training and validation accuracy diverging by less than 2 percentage points.

Table 3: Five most influential determinants

Rank	Feature	Mean SHAP Value	Interpretation
1	End-of-year account balance	0.042	Higher accumulated wealth correlates with increased risk tolerance
2	Age (derived from Date_of_Birth)	0.038	Younger participants exhibit greater propensity for aggressive profiles
3	Start-of-year account balance	0.031	Beginning-of-period wealth serves as baseline financial capacity indicator
4	Gender	0.024	Male participants demonstrate higher likelihood of aggressive risk adoption
5	Income level	0.021	Elevated income contributes positively to risk-taking behavior

The SHAP-based interpretability analysis revealed that financial capacity indicators consistently outperformed traditional socio-demographic variables in predicting risk profiles. Specifically, the five most influential determinants were identified in Table 3.

Notably, the combined predictive power of account balance variables (end-of-year and start-of-year) substantially exceeded that of any single demographic attribute, underscoring the primacy of observed financial behavior over self-reported characteristics in risk profiling. The SHAP dependence plots further revealed non-linear relationships: the marginal effect of income on risk profile intensified beyond approximately 5,000 TL monthly threshold, while age exhibited an inverse U-shaped pattern with peak risk tolerance in the 30–40 age bracket.

Secondary factors—including profession, policy city, coverage type, and customer segment—demonstrated moderate influence, whereas number of children, marital status, nationality, and hometown exhibited minimal discriminatory power (mean $|\text{SHAP}| < 0.005$), suggesting that conventional household-level demographics are insufficient proxies for financial risk preferences in pension contexts.

Several variables originally considered—including duration of IPS participation and frequency of fund changes—were excluded from the final model due to high missing data rates (88% and 76%, respectively) and multicollinearity with account balance indicators. Future research may explore advanced imputation techniques to recover these potentially informative behavioral signals.

From a policy perspective, these findings carry significant implications for pension fund management in emerging economies. First, default fund allocation mechanisms should incorporate dynamic financial capacity assessments rather than static age-based rules. Second, financial literacy interventions may be most effectively targeted at participants with low-to-moderate accumulated balances, as this segment demonstrates the greatest variability in risk profile classification. Third, gender-sensitive advisory frameworks warrant development, acknowledging systematic differences in risk propensity while avoiding deterministic stereotyping. Finally, the integration of explainable machine learning tools—exemplified by the XGBoost-SHAP pipeline deployed herein—can enhance regulatory oversight and participant trust through transparent, auditable decision support.

The proposed framework contributes methodologically by demonstrating the feasibility of high-accuracy risk profiling without reliance on intrusive psychometric surveys, and substantively by establishing accumulated savings behavior as the predominant predictor of pension investment risk tolerance in the Turkish context. Limitations include the single-institution data source, cross-sectional design precluding causal inference, and potential selection bias toward already-engaged IPS participants. Extensions might incorporate longitudinal tracking of profile stability, comparative analysis across pension systems, and deployment of real-time adaptive profiling as balances evolve.

Overall, this study advances the literature on pension economics by bridging predictive accuracy with model interpretability, offering actionable insights for improving risk assessment mechanisms, optimizing fund allocations, and enhancing the long-term sustainability of individual pension systems in developing economies.

References:

1. E. Gómez-Déniz, J. V. Pérez-Rodríguez, and S. Sosvilla-Rivero, “Analyzing How the Social Security Reserve Fund in Spain Affects the Sustainability of the Pension System,” *Risks*, 2022, doi: 10.3390/risks10060120.
2. I. Mavlutova, S. Titova, and A. Fomins, “Pension System in Changing Economic Environment: Case of Latvia,” *Procedia Econ. Financ.*, 2016, doi: 10.1016/s2212-5671(16)30316-1.
3. A. Chłoń-Domińczak, “Impact of changes in multi-pillar pension systems in CEE countries on individual pension wealth,” *J. Pension Econ. Financ.*, 2018, doi: 10.1017/S1474747216000238.
4. F. AKGÜL and Ş. Y. YİĞİTER, “Davranışsal Finans Yaklaşımı Çerçevesinden Bireysel Emeklilik Sistemine Yönelik Tutum ve Davranışlar,” *Anadolu Üniversitesi Sos. Bilim. Derg.*, 2021, doi: 10.18037/ausbd.1039461.
5. Z. A. ÖZDEMİR, Y. B. ÖNAL, and S. ŞAHİN, “TÜRKİYE DE BİREYSEL EMEKLİLİK SİSTEMİNİN SERMAYE PİYASASININ GELİŞİMİ ÜZERİNDEKİ ETKİSİ,” *Dokuz Eylül Univ. İktis. ve İdari Bilim. Derg.*, 2019, doi: 10.24988/deuifb.2018332767.
6. K. H. Hassan, R. A. Rahim, F. Ahmad, T. N. A. Tengku Zainuddin, R. R. Merican, and S. K. Bahari, “Retirement Planning Behaviour of Working Individuals and Legal Proposition for New Pension System in Malaysia,” *J. Polit. Law*, 2016, doi: 10.5539/jpl.v9n4p43.
7. M. Horák, S. Kukučková, and K. Dobeš, “Exploring behavioral barriers and interventions in retirement savings: Findings from online focus groups among university students,” *Invest. Manag. Financ. Innov.*, 2022, doi: 10.21511/imfi.19(2).2022.13.
8. Y. CAN, “BİREYSEL EMEKLİLİĞİN TÜRKİYE’DEKİ DURUMU VE GELİŞİMİ,” *Ekon. Bilim. Derg.*, vol. 2, no. 2, pp. 139–146, 2010.
9. Y. DEMİR and A. YAVUZ, “Bireysel Emeklilik Sisteminin Sermaye Piyasalarına Etkisi ve Sistemin Gelişmesinde Vergisel Teşviklerin Önemi,” *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilim. Fakültesi Derg.*, vol. 9, no. 1, 2004.
10. A. SEYFULLAHOĞULLARI and B. DEMİRHAN, “BİREYSEL EMEKLİLİK SİSTEMİNE KATKILARI AÇISINDAN VERGİ AVANTAJI İLE DEVLET KATKISININ KARŞILAŞTIRILMASI,” *Marmara Sos. Araştırmalar Derg.*, no. 9, pp. 22–50, 2016.
11. Y. BAYRAKTUTAN and M. ŞAHİN, “Bireysel Emeklilik Ektörünün Gelişimi ve Bilgi - Bilişim Teknolojilerinin Kullanımı,” *Bilgi Ekon. ve Yönetimi Derg.*, vol. 2, no. 2, 2007.
12. A. Chłoń-Domińczak, P. Strzelecki, and W. Łątkowski, “Pension reforms or consequences of the economic crisis? Measuring the changes of pension incomes in selected EU countries using APC approach,” *J. Econ. Ageing*, 2020, doi: 10.1016/j.jeoa.2017.09.002.
13. E. E. Aksoy, “Individual Factors Affecting the Participation of Turkish People in the New Individual Pension System,” in *Research Anthology on Personal Finance and Improving Financial Literacy*, 2021.
14. R. Holzmann, “Global pension systems and their reform: Worldwide drivers, trends and challenges,” *Int. Soc. Secur. Rev.*, 2013, doi: 10.1111/issr.12007.
15. O. ŞENER and F. AKIN, “ÖZEL EMEKLİLİK FONLARI VE TÜRKİYE’DE BİREYLERİN BİREYSEL EMEKLİLİK SİSTEMİNE GİRİŞ KARARLARINI ETKİLEYEN FAKTÖRLERİN BELİRLENMESİ ÜZERİNE BİR ARAŞTIRMA,” *Marmara Üniversitesi İktisadi ve İdari Bilim. Derg.*, vol. 29, no. 2, pp. 291–321, 2015.
16. S. IŞIK EROL, “Sosyal Güvenlik Açısından Bireysel Emeklilik Sisteminin Önemi,” *İktisadi Yenilik Derg.*, vol. 6, no. 2, pp. 10–29, 2019.

17. S. SAHIN, H. RITTERSBERGERTILIC, and A. ELVEREN, "THE INDIVIDUAL PENSION SYSTEM IN TURKEY: A GENDERED PERSPECTIVE," *Ekonomik Yaklaşım*, 2010, doi: 10.5455/ey.20039.
18. BerkantCS, Binceş, B. Yılmaz, Coderspace, Coskunmel, and E. Işık, "Anadolu Hayat Emeklilik Datathon // Coderspace." 2022.
19. W. Zhao, J. Li, J. Zhao, D. Zhao, J. Lu, and X. Wang, "XGB model: Research on evaporation duct height prediction based on XGBoost algorithm," *Radioengineering*, 2020, doi: 10.13164/re.2020.0081.
20. J. Zaki, A. Nayyar, S. Dalal, and Z. H. Ali, "House price prediction using hedonic pricing model and machine learning techniques," *Concurr. Comput. Pract. Exp.*, 2022, doi: 10.1002/cpe.7342.
21. S. U. Chowdhury, S. Sayeed, I. Rashid, M. G. R. Alam, A. K. M. Masum, and M. A. A. Dewan, "Shapley-Additive-Explanations-Based Factor Analysis for Dengue Severity Prediction using Machine Learning," *J. Imaging*, 2022, doi: 10.3390/jimaging8090229.
22. Y. Lu, X. Fan, Y. Zhang, Y. Wang, and X. Jiang, "Machine Learning Models Using SHapley Additive exPlanation for Fire Risk Assessment Mode and Effects Analysis of Stadiums," *Sensors*, 2023, doi: 10.3390/s23042151.