**Article**

# Relationship of Health and Economic Status with Stock Market Participation among Older Adults: Insights from Machine Learning Approach

Zin Mar Oo[1,2]

**Abstract**

This study investigates the relationship between health, wealth, and stock market participation among older adults using machine learning techniques, with a focus on Accumulated Local Effect (ALE) plots within an Artificial Neural Network (ANN) framework. Unlike traditional econometric models, which assume linearity and homogeneity in predictor relationships, our approach captures complex non-linear interactions and heterogeneous associations. The analysis reveals that higher net worth and income significantly increase the likelihood of stock market participation, with the effect being more pronounced among individuals in better health. Conversely, those in poorer health are less likely to invest, likely due to financial constraints or risk aversion. Through feature importance analysis and ALE plots, we offer a detailed understanding of how financial and health variables interact to shape investment behavior. These findings offer valuable insights for policymakers and financial advisors aiming to increase stock market participation among diverse population segments.

**Keywords:** Stock market participation, older adults, health status correlation, economic status, artificial neural networks (ANN)

## 1. Introduction

The exploration of stock market participation has long been a crucial aspect of economic and financial research, unraveling the factors that persuade individuals to invest in stocks and, consequently, shaping market dynamics and broader economic outcomes. Historically, this inquiry has been pursued through econometric models, with scholars like Bucciol and Miniaci (2015), Gao et al. (2019), and Dimmock et al. (2016), carefully identifying income,

[1]Doctoral Program Student, Graduate School of Economics, Ritsumeikan University, Japan. Email: gr0602sk@ed.ritsumei.ac.jp, zinmaroo@iuj.ac.jp.
[2]Budget Department, Ministry of Planning and Finance, Myanmar

wealth, education, and risk tolerance as pivotal factors of investment behavior. These analyses have significantly advanced our understanding, highlighting the profound influence of socio-economic and demographic factors on individual decisions to engage in the stock market.

Economic and health conditions play a crucial role in shaping individuals' financial decisions, particularly regarding stock market participation. Wealth and income provide households with the financial flexibility to invest in riskier assets and diversify their portfolios, while poor economic conditions may limit their ability to engage in the market. At the same time, health status influences investment behavior, as individuals in poor health may be more risk-averse (Rosen and Wu, 2004), prioritizing immediate financial security over long-term investment gains. Understanding the interplay between these two factors is essential for comprehending how different population groups make investment decisions and engage in the stock market.

In the literature, wealth and income are positively correlated with stock market participation, as households with higher economic resources are better equipped to cover the fixed costs of participation and diversify their portfolios (Bertaut, 1998; Buccioli and Miniaci, 2015; Gao and Yao, 2019). Similarly, households in poor health tend to avoid risky financial assets compared to those in good health (Rosen and Wu, 2004; Yogo, 2016). These findings suggest a linear relationship between economic and health conditions and stock market participation. However, in this study, we apply machine learning techniques to explore possible nonlinear in greater detail, examining how economic and health factors influence stock market participation across different feature levels.

Despite extensive research using conventional econometric models, there is an opportunity to leverage machine learning (ML) techniques to analyze individual investment decisions, particularly stock market participation. In ML literature, the terms 'features' or 'input variables' are used to predict the 'outcome' or 'target variable.' In economic and financial literature, they are referred to as 'independent variables,' 'explanatory variables,' 'predictors,' or 'covariates' that explain a dependent, predicted, response, or outcome variable. While econometric models have enhanced our understanding of the socio-economic and demographic factors of investment behavior, ML can offer deeper insights by capturing complex, non-linear interaction and heterogeneous associations without making prior assumptions about the data.

A comprehensive literature review by Kumbure et al. (2022) indicates that the application of machine learning in stock market analysis has predominantly focused on predictive analytics, focusing on stock prices, returns, and market trends. This body of work, which spans 138 journal articles published between 2000 and 2019, highlights the advanced computational power and pattern recognition capabilities of ML algorithms, which analyze vast datasets to predict market trends and challenge traditional notions of market predictability. However, this focus on predictive analytics overlooks the equally critical domain of behavioral analysis, specifically the investigation of features influencing an individual's decision to engage in the stock market.

Our study addresses this gap by applying advanced ML techniques, notably artificial neural networks (ANNs), to identify factors influencing stock market participation, thus providing a novel perspective on investment behavior analysis. For example, a study comparing the performance of various ML models — such as ANN, SVM, random forest, and Naïve-Bayes — on the Taiwan stock market index found that ANN generated the best risk-adjusted performance, followed by SVM and random forest, with Naïve-Bayes coming in last (Chia-Cheng et al., 2019). We focus on ANN to analyze how demographic and financial features relate to individuals' decisions to invest in the stock market. This approach complements traditional econometric analyses adds depth by integrating the detailed perspectives afforded by ML. By capturing non-linear interactions and heterogeneous relationships, ML provides a deeper understanding of these relationships.

Our investigation reveals that health and economic status significantly influence stock market participation among older adults. Specifically, individuals in poorer health are less likely to invest in the stock market, underscoring the critical interplay between health and investment behavior. Moreover, our analysis highlights the positive association of higher net worth and income on stock market participation, with these association being more pronounced among healthier individuals. Through ML, we also identify key demographic and financial features — including total net worth, income, marital status, and personal risk tolerance — as influential features in investment decisions. The ML techniques employed in our study facilitate a detailed interpretation of these features, offering new insights into the complex behaviors underlying stock market participation.

Additionally, the feature importance analysis provides a comprehensive view of the relative significance of various predictors, reinforcing the robustness of our findings. By comparing these results with those obtained from conventional econometric models, we highlight both commonalities and differences, demonstrating the added value of machine learning in capturing the complexities of financial behavior.

By examining these complex relationships, our research contributes to understanding of the variables influencing stock market participation, providing insights for financial advisors, policymakers, and investors. Our findings highlight the significant role of health and economic status in investment decisions and showcase the potential of ML to capture the complexities of financial behavior in the modern economy.

The structure of this paper is as follows. After this introduction, the 'Empirical Application' section details our ML methodology, data sources, model estimation techniques, and evaluation criteria. This is followed by the 'Empirical Results' section, which presents and discusses our findings of our analysis. Finally, the paper concludes with the 'Conclusion' section, summarizing the study's key insights and suggesting avenues for future research.

## 2. Empirical Application

This section outlines our methodology for analyzing stock market participation, starting with an explanation of the ANN algorithm and interpretable ML methods. We detail our data preparation, model estimation, hyperparameter optimization, and evaluation methodologies.

### 2.1 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are pivotal in machine learning for their pattern recognition and classification capabilities. An ANN consists of layers of input, hidden, and output neurons, interconnected in a network tailored to the dataset's specific needs. Training an ANN involves adjusting the network's weights and biases to minimize classification errors (Montesionos Lopez et al., 2022). Our ANN model uses two hidden layers with 16 and 8 nodes, respectively. The choice of nodes will be explained in the following subsection. We use Scaled Exponential Linear Unit (SeLU) and sigmoid activation functions (Klambauer et al., 2017) and the Adam optimizer (Kingma and Ba, 2014) to enhance model accuracy. Detailed hyperparameter optimization is discussed later.

### 2.2 Interpretable Machine Learning Methods

Interpreting machine learning models is crucial for validating the significance and reliability of their predictions, especially with complex datasets. We use three techniques: Permutation Feature Importance (PFI), SHapley Additive exPlanations (SHAP), and Accumulated Local Effects (ALE), supported by the literature on their application and efficacy.

PFI provides a direct method to assess the impact of each feature on the model's predictive accuracy. This technique involves randomly shuffling the values of each feature and measuring the change in performance metrics, such as accuracy or AUC-ROC score, while keeping other features constant (Breiman, 2001b). A significant drop in model performance indicates the importance of the feature. This approach has been widely used in various fields, including economics and finance, to understand the relationships between health and economic features (Ahmed et al., 2022), consumer cross-buying behaviors (Kilinc and Rohrhirsch, 2022), and credit risk prediction (Bellotti et al., 2021; Zhang et al., 2022a). It has also been applied to forecast financial instability and bankruptcy in local governments (Tran et al., 2022; Antulov-Fantulin et al., 2021), and to examine the impact of stock market variables on stock returns (Tratkowski, 2021).

SHAP values, derived from cooperative game theory, provide a detailed perspective on the influence of each feature on individual predictions (Lundberg and Lee, 2017). SHAP quantifies the contribution of each feature by calculating the mean absolute SHAP values, offering a comprehensive understanding of the factors driving stock market participation. This method has been extensively documented in finance literature. For instance, Ariza-Garzón et al. (2020) demonstrated the utility of SHAP in peer-to-peer lending credit scoring, highlighting its role in enhancing both accuracy and transparency of machine learning models. Similarly, SHAP has been crucial in studies on corporate financial insolvency (Yıldırım et al., 2021) and financial distress (Zhang et al., 2022b), providing insights into complex relationships and boosting model interpretability.

ALE plots are essential for visualizing how changes in input variables influence the model's predictions. Unlike partial dependence plots, ALE plots account for the distribution of features, making them more reliable for interpreting feature effects (Apley and Zhu, 2020). ALE plots illustrate the predicted response variations with changes in feature values, clarifying the nature and magnitude of these relationships. Recent research has demonstrated the value of ALE in various applications, such as analyzing fiscal and macroeconomic indicators to forecast fiscal risks (Valencia et al., 2022; Jarmulska, 2022), and predicting non-performing loan recovery rates (Bellotti et al., 2021). These plots help in understanding local interactions and provide a clearer picture of the model's behavior.

These methods provide a robust framework for interpreting ML models, ensuring transparency and empirical grounding. Using PFI, SHAP, and ALE, we reveal key features' predictive strengths and explore underlying mechanics, offering a comprehensive perspective on model behavior.

### 2.3 Data

This study utilizes data from the 2020 iteration of the Health and Retirement Study (HRS), an ongoing biennial survey initiated in 1992. The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan (Health and Retirement Study; RAND HRS Longitudinal File 2020 (V1)). It focuses on U. S. residents aged 50 and above, along with their spouses, to gather comprehensive data on aging, health, financial conditions, and societal features.

Our primary variable of interest is stock ownership, defined as a binary indicator based on responses to whether the individuals or their partners own stocks or stock mutual funds. Stockholders are identified if the value of stocks or mutual funds owned is positive, while non-stockholders have a value of zero. Importantly, our definition excludes assets held in retirement accounts, concentrating instead on direct stock holdings and mutual funds. This distinction accounts for the observed stock market participation rate in our study (26%), which is lower than the commonly reported rate above 50% that includes all forms of stock holdings. This focus on direct holdings provides a clearer picture of discretionary investment behavior, influenced by individual financial decisions and risk preferences, separate from retirement savings that may be more automatic or employer driven.

Feature selection for this analysis is guided by empirical literature and theoretical frameworks related to stock market participation. We categorize the variables into eight demographic and six financial characteristics. The demographic variables include gender, age, number of living children, years of education, self-reported health status, marital status, work status, and geographical location (across nine U. S. divisions). Financial variables cover net worth, income, ownership of life and health insurance, out-of-pocket medical expenses in the past two years, and a measure of risk tolerance. For analysis, logarithmic transformations are applied to net worth, income, and medical expenses to normalize their distributions.

To construct our dataset, we apply specific selection criteria focusing on households capable of constructing stock portfolios. We exclude observations with a negative net worth, as these households lack the financial capability to invest. Additionally, we remove any observations with missing values in outcome and feature variables. The variables are detailed in Appendix Table A1, with summary statistics provided in Table 1.

**Table 1**: Summary Statistics

| | Observations | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| **Outcome** | | | | | |
| Stock Ownership | 4,990 | 0.26 | 0.44 | 0 | 1 |
| **Features** | | | | | |
| Log total net worth | 4,990 | 7.08 | 2.72 | 0.00 | 16.13 |
| Log income | 4,990 | 5.93 | 1.27 | 0.00 | 10.79 |
| Age | 4,990 | 70.74 | 10.34 | 40.00 | 102.00 |
| Number of children | 4,990 | 2.63 | 1.93 | 0.00 | 11.00 |
| Year of schooling | 4,990 | 13.59 | 2.91 | 0.00 | 17.00 |
| Medical expense | 4,990 | 2.17 | 1.49 | 0.00 | 7.53 |
| Risk tolerance | 4,990 | 4.16 | 2.45 | 1.00 | 10.00 |
| Health status | 4,990 | 2.78 | 0.97 | 1.00 | 5.00 |
| Has health insurance | 4,990 | 0.92 | 0.27 | 0 | 1 |
| Has life insurance | 4,990 | 0.54 | 0.50 | 0 | 1 |
| Gender (0: Female, 1: Male) | 4,990 | 0.47 | 0.50 | 0 | 1 |
| Work status | 4,990 | | | | |
| Full-time | | 0.21 | 0.41 | 0 | 1 |
| Part-time | | 0.04 | 0.19 | 0 | 1 |
| Unemployed | | 0.02 | 0.13 | 0 | 1 |
| Partly retired | | 0.08 | 0.27 | 0 | 1 |
| Retired | | 0.63 | 0.48 | 0 | 1 |
| Disabled | | 0.01 | 0.12 | 0 | 1 |
| Not in labor force | | 0.02 | 0.13 | 0 | 1 |
| Marital status | 4,990 | | | | |
| Married | | 0.48 | 0.50 | 0 | 1 |
| Married, spouse absent | | 0.01 | 0.08 | 0 | 1 |
| Separated | | 0.02 | 0.14 | 0 | 1 |
| Divorced | | 0.19 | 0.40 | 0 | 1 |
| Widowed | | 0.22 | 0.41 | 0 | 1 |
| Never married | | 0.09 | 0.28 | 0 | 1 |
| Location | 4,990 | | | | |
| New England | | 0.04 | 0.18 | 0 | 1 |
| Mid Atlantic | | 0.11 | 0.31 | 0 | 1 |
| East North Central | | 0.15 | 0.36 | 0 | 1 |
| West North Central | | 0.07 | 0.26 | 0 | 1 |
| South Atlantic | | 0.24 | 0.43 | 0 | 1 |
| East South Central | | 0.07 | 0.25 | 0 | 1 |
| West South Central | | 0.10 | 0.30 | 0 | 1 |
| Mountain | | 0.07 | 0.26 | 0 | 1 |
| Pacific | | 0.15 | 0.36 | 0 | 1 |

The theoretical framework guiding our feature selection is grounded in established principles of behavioral finance and economic theories (Pompian, 2012). This framework recognizes the interplay between demographic traits and financial features in shaping individual investment decisions. The inclusion of demographic variables, like gender, age, and education, is based on their established relationships with risk preferences and financial literacy.

Financial attributes, such as net worth, income, and medical expenses, are chosen due to their association with an individual's financial standing and exposure to risk. Additionally, location captures geographical differences that might influence stock market participation, acknowledging how regional economic conditions influence investment choices.

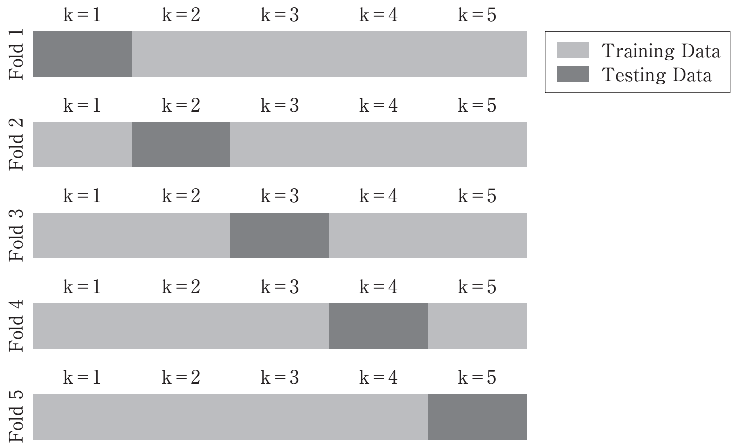
Key empirical studies supporting these variables include, but are not limited to, Almenberg and Dreber (2015) on gender, Fagereng et al. (2017) on age, Cole and Shastry (2009) on education, Rosen and Wu (2004), Edwards (2008), and Goldman and Maestas (2013) on health status, medical expenses, and health insurance coverage, Love (2010) on marital status and number of living children, Amidu et al. (2021) on work status, Bucciol and Miniaci (2015) on net worth, Gao et al. (2019) on income, Cavapozzi et al. (2013) on life insurance ownership, and Dimmock et al. (2016) on risk tolerance. While these studies identify relevant variables for stock market participation, they predominantly rely on traditional statistical analyses, which may not fully capture the complex, non-linear relationships among these variables. Thus, we apply machine learning techniques, specifically artificial neural networks (ANN), to explore these complex dynamics. Machine learning's ability to handle large datasets and identify complex patterns makes it well-suited for this analysis.

Before training our algorithms, we preprocess our feature data to boost model efficiency, reduce training times, and mitigate possible biases (Chollet, 2021). Numerical features are normalized using z-score normalization to mitigate the influence of outliers, while categorical features are appropriately encoded: binary encoding for gender, life insurance ownership, and health insurance coverage; ordinal for education and health status (reflecting inherent ordering); and one-hot encoding for non-ordinal variables like marital status and work status. This preparation ensures that the data is optimally formatted for our machine learning models.

### 2.4 Modeling, Optimization, and Evaluation

Our study addresses the challenge of an imbalanced dataset, where 26.1% of the individuals are stock owners, and 73.9% are not. This imbalance can lead to biased performance in predictive models if not properly handled. To mitigate this issue, we apply stratified k-fold cross-validation with an 80%-20% train-test split. Stratification ensures that each fold accurately represents the entire dataset's class proportions. We use a 5-fold cross-validation technique, enhancing our model's robustness by allowing extensive learning and ensuring every data point is tested once. Figure 1 visualizes the distribution of training and testing data across each fold, with black bars indicating the testing subset and grey bars the training subset.

**Figure 1**: Train-Test split.




For fine-tuning our model' s hyperparameters, we use Grid Search alongside stratified k-fold cross-validation. This process involves specifying a set of hyperparameter and exploring a range of values for each within the grid. The dataset is divided, maintaining an even distribution of outcomes. Each segment of the data is alternately used as a validation set, with the model being trained across the other segments. This process is repeated for every hyperparameter combination, allowing us to record and compare performance metrics. The best-performing hyperparameter set is selected based on overall effectiveness, evidenced by the highest accuracy or lowest loss. The specific parameters and their ranges are detailed in Table 2.

**Table 2**: Hyper-parameters

| | Grid search | Choice model |
|---|---|---|
| No. of hidden layers | 2 | 2 |
| No. of nodes in the hidden layers | (8, 4), (16, 8), (32, 16), (64, 32) | (16, 8) |
| Dropout rate | 0.1 | 0.1 |
| Optimizer | Adam | Adam |
| Activation function | Selu | Selu |

After identifying the optimal hyperparameters, we train the model using the defined folds, yielding five distinct models from the cross-validation process. Each model is evaluated on its respective testing set to ensure that our assessments accurately reflect the model's performance on unseen data. The saved cross-validation models predict the probability of being a stockholder for each testing sample, classifying as a stockholder if the probability is greater than 0.5 and as a non-stockholder otherwise. This classification yields a confusion matrix (Table 3) where actual non-stockholders can be predicted as either non-stockholders or stockholders, and actual stockholders can be predicted as either non-stockholders or stockholders.

**Table 3**: Confusion Matrix

| | | Prediction Outcome | |
|---|---|---|---|
| | | Stockholder | Non-stockholder |
| Actual outcome | Stockholder | True Positive (TP) | False Negative (FN) |
| | Non-stockholder | False Positive (FP) | True Negative (TN) |

Following the individual evaluation of each cross-validation model, we aggregate the results by summing the numbers of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) across all folds. This aggregation forms a comprehensive confusion matrix from which we calculate our evaluation metrics. This approach allows us to account for the performance of all five cross-validation models across the diverse segments of the dataset, enhancing the reliability and robustness of our predictive analysis. By summing these results, we obtain a holistic view of the model's effectiveness, which is crucial for assessing its ability to generalize and ensure consistency across different subsets of data.

To assess the effectiveness of our stock market participation prediction, we use two established evaluation tools: the Receiver Operating Characteristics (ROC) curve and the Precision-Recall (PR) curve, as recommended by Fawcett (2006) for ROC and by Davis and Goadrich (2006) for PR. Our model defines stock owners as the positive class and non-owners as the negative class. The ROC curve plots the True Positive Rate (TPR), or sensitivity, against the False Positive Rate (FPR) at various threshold levels. The TPR measures the proportion of actual positives correctly identified, while the FPR calculates the ratio of negatives mistakenly labeled as positives. The TPR, also known as sensitivity or recall, measures the proportion of actual positives correctly identified, calculated as

$$\mathrm{TPR}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

The FPR, indicating the likelihood of false alarms, is the ratio of negatives incorrectly classified as positives, computed as

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}$$

Statistically, the ROC curve illustrates the trade-off between detecting true positives and avoiding false positives, with an ideal AUC-ROC score of 1.0 symbolizing perfect classification. Conversely, an AUC of 0.5 suggests no better accuracy than random guessing. According to Hosmer Jr et al. (2013), an AUC-ROC between 0.7 and 0.8 is considered acceptable, between 0.8 and 0.9 is excellent, and above 0.9 indicates outstanding model performance.

While the ROC curve is a robust statistical tool, its effectiveness may be limited in imbalanced classification scenarios, as highlighted by Saito and Rehmsmeier (2015). This is where the Precision-Recall (PR) curve becomes more insightful. The PR curve evaluates the model's precision—how accurately it predicts positive outcomes—against its recall, which measures the model's success in identifying all actual positives. Precision, also known as Positive Predictive Value (PPV), is calculated as

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Given the prevalence of the positive class in our dataset, an area under the PR curve exceeding 0.26 suggests the model is performing above random chance, providing a critical perspective on its capability to discern stock owners within the dataset's inherent imbalance.

By thoughtfully addressing the imbalance within our dataset and using thorough model evaluation metrics, our approach not only mitigates potential biases but also ensures that our insights into stock market participation are reliable and significant. This well-rounded strategy, based on empirical evidence and sophisticated statistical methods, provides a clear understanding of stock ownership dynamics among the surveyed individuals.

## 3. Empirical Results

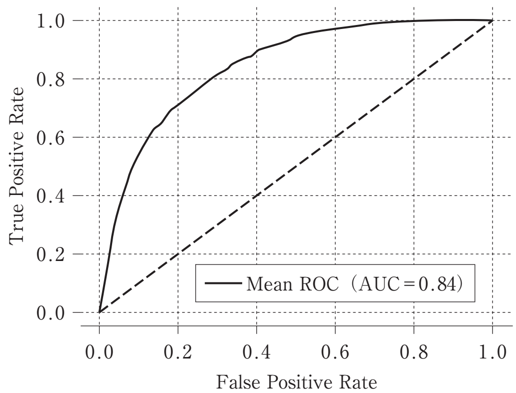
In this section, we explore our model's accuracy in predicting stock market participation among older U.S. households. Our investigation serves a dual purpose: firstly, to assess the model's precision in forecasting participation in the stock market, and secondly, to identify and analyze the crucial features, particularly health, income, and wealth, influencing these predictions. We also aim to understand the interplay among these variables and their collective association with the outcome. This comprehensive analysis enables us not only to gauge the model's effectiveness in prediction but also to gain insights into how key variables interact and relate stock market participation among the elderly, offering a deeper understanding of the underlying features at play.

### 3.1 Predictability of Stock Market Participation

This section evaluates model's ability to forecast whether individuals are likely to engage in the stock market. We employ the Receiver Operating Characteristic (ROC) curve and its Area Under the Curve (AUC-ROC), along with the Precision-Recall (PR) curve and its Area (AUC-PR), using a 5-fold cross-validation approach for a comprehensive evaluation.

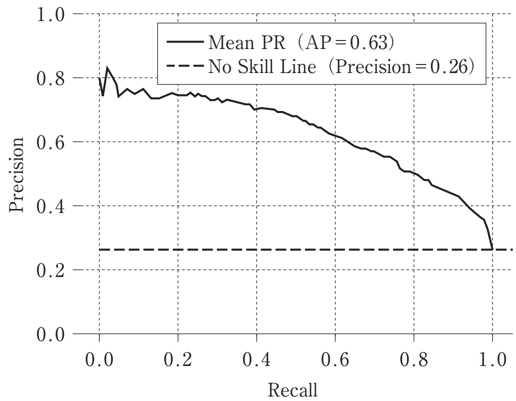
The ROC curve reveals the model's proficiency in identifying stock market participants versus non-participants by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) over different thresholds. An average AUC-ROC score of 0.84, highlighted in Figure 2, indicates the model's strong ability to discriminate, significantly outperforming the baseline score of 0.5. This suggests that our model is highly effective in distinguishing between participants and non-participants, demonstrating notable accuracy and a low rate of false positives.

**Figure 2**: ROC curve.




The PR curve assesses the model's precision and capability to correctly identify stock market participants, showing an AUC-PR score of 0.63, illustrated in Figure 3. This outcome is significant given the dataset's imbalance, where the proportion of positive instances (stockholders) is small compared to negative instances (non-stockholders). Such imbalance often challenges predictive models, as they can be biased toward the majority class. The performance which exceeds the baseline precision of 0.26—a figure derived from the actual proportion of stock market participants—demonstrates the model's effectiveness far beyond random or simplistic predictive models. It highlights the model's superior accuracy in identifying the minority class of stock market participants.

**Figure 3**: PR curve.



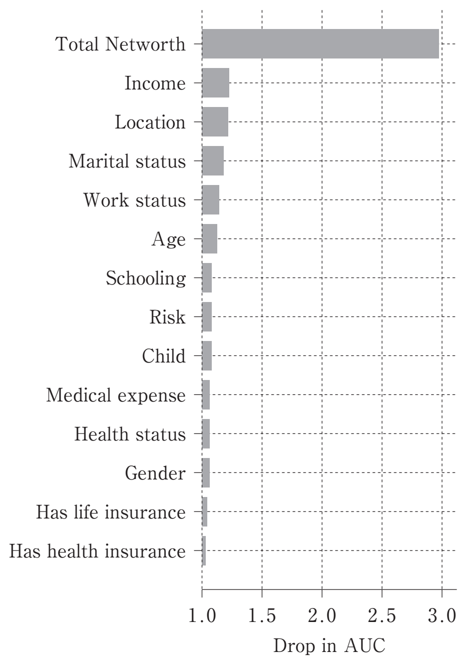
The analysis of ROC and PR curves, complemented by the AUC-ROC and AUC-PR scores, underlines the model's predictive accuracy and reliability. These metrics, derived from a thorough 5-fold cross-validation, confirm the model's accuracy and practicality in identifying stock investors, significantly outperforming no-skill thresholds for both AUC-ROC and AUC-PR and proving its real-world relevance.

### 3.2 Feature Importance

To investigate which features drive people to invest in the stock market, we explore the importance of various features through two techniques: Permutation Feature Importance (PFI) and SHapley Additive exPlanations (SHAP). These methods identify the influential features in our model's decision-making process.

In the PFI analysis, depicted in Figure 4, we measure how the model's performance declines when the data for a specific feature is shuffled. This "dropout loss" helps us understand the influence of each feature on the model's accuracy. Notably, an individual's total net worth stands out as the top feature for stock market participation, highlighting the significance of financial health in investment choices. Income also ranks highly, reflecting the role of economic strength in these decisions.

**Figure 4**: Permutation Feature Importance (PFI) plot.

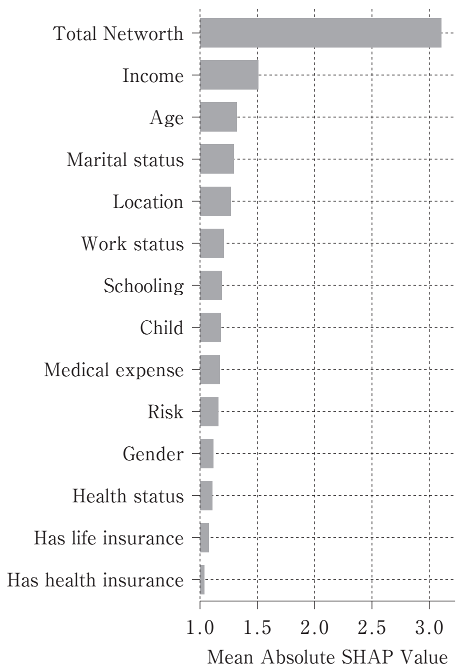



Interestingly, the location feature indicates the important role of regional factors, suggesting that individuals' living states are related to their investing behavior. This could be due to differences in economic conditions, regulations, and cultural attitudes towards investment.

Other factors like work and marital status also emerge as important, reflecting how job situations and personal relationships influence investing. Additional features such as age, education, risk tolerance, family size, medical expenses, health status, gender, and insurance coverage also play roles, although their magnitudes are smaller.

Complementing the PFI analysis, the model's feature association is further interpreted using SHAP analysis. SHAP reveals the broader significance of each feature's role in stock market participation, shown through the mean absolute SHAP values depicted in Figure 5. Total net worth again stands out as a pivotal feature, evidenced by its leading mean absolute SHAP value. This underscores the profound association of financial standing with the inclination to engage in the stock market. Beyond net worth, aspects such as income, age, marital status, and geographical location also emerge as influential, highlighting their critical contribution to investment decisions. Through SHAP's lens, we gain a deeper understanding of how these variables shape investment behaviors, offering a richer narrative about what drives individuals towards or away from the stock market.

**Figure 5**: SHAP bar plot for feature importance.




The insights from both the PFI and SHAP analyses jointly highlight the complexity behind stock market participation. Financial resources, particularly net worth and income, play a predominant role. However, the significance of geographical factors (captured by the location feature) and socio-demographic attributes, such as marital status, education, and age, cannot be overlooked. These insights can be instrumental in developing targeted strategies to enhance financial literacy and stock market inclusivity, thereby promoting broader participation in stock market investments.

### 3.3 Health, Wealth, and Stock Market Participation

In this section, we explore the complex linkage among health, wealth, and stock market participation. To do so, we examine how each feature individually relates to stock market participation and how their interconnections further shape investment decisions.

#### 3.3.1 Link of Wealth to Stock Market Participation

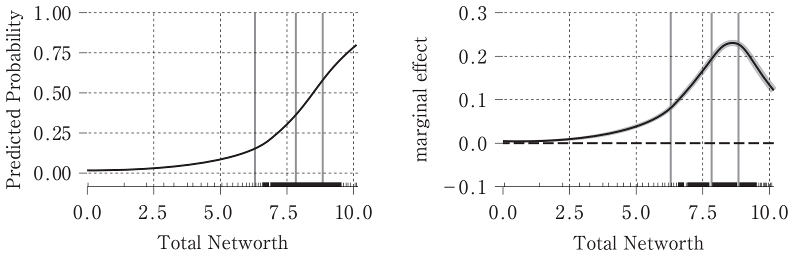
In our analysis, we investigate the marginal effects of features on the likelihood of stock market participation within our ANN model. By employing Accumulated Local Effect (ALE) plots, we reveal the relationship between the probability of stock market participation and each feature. The horizontal axis of the ALE plots represents the values of the features, aligned with the sample's distribution, while the vertical axis shows the corresponding probability of participating in the stock market. For numerical features such as net worth and income, feature distributions within the dataset are highlighted on the x-axis using black markers. Vertical dotted lines indicate the first, second, and third quartiles of the data distribution, providing a detailed view of the variable spread. To focus on core trends, we exclude the bottom and top 5% of the data points and convert normalized feature values back to their original scale.

To assess how each feature's value influences the likelihood of stock market participation, we analyze the slopes of ALE plots. These slopes represent the change in predicted participation probability with feature value changes. By plotting these slopes, we reveal how the average marginal effects vary across the feature spectrum, highlighting crucial value ranges. This analysis is visually represented with quartile markers and feature distribution overlays on the plots, offering a detailed view of how different feature values influence stock market participation.

**Total Net Worth Analysis**

Figure 6 presents a two-part analysis: the left panel with the ALE plot and the right panel displaying the marginal effect plot. The left panel shows a positive influence of total net worth on stock market participation probability, particularly noting a stronger positive connection at higher net worth levels. The right panel highlights a consistently positive marginal effect across all net worth levels. Specifically, the effect ranges from 0 to 0.1 for the first net worth quantile, increases to between 0.1 and 0.2 for the second, exceeds 0.2 for the third, and moderates to between 0.1 and 0.2 for the fourth quantile.

**Figure 6**: ALE plot and Approximated marginal effect plot of total net worth




To quantify these effects, we calculated the marginal effects at each point in the feature's range. The average of these marginal effects provides a comprehensive understanding of how net worth influences stock market participation. This average marginal effect (AME) of total net worth is 0.145 with a standard error of 0.003, which is statistically significant at the 1% level ($p<0.01$). This result shows that, on average, a one-unit increase in the log of total net worth increases the probability of stock market participation by approximately 14.53%.

We compared these values with the coefficients from the linear probability model (LPM) and the marginal effects derived from the logit and probit models. The coefficient of total net worth from the LPM is 0.136, indicating a similar effect size. The logit and probit models provide marginal effects of 0.231 and 0.255, respectively, both significant at the 1% level. This comparison underscores the robustness of our findings across different modeling approaches, although it is important to note that the interpretation of AME from ANN can differ from those in econometric models. Table 4 is a summary table comparing the AME values from the four models.

**Table 4**: Average Marginal Effects (AME) for Total Net Worth

| Model | AME | Standard Error |
|---|---|---|
| ANN | 0.145*** | 0.003 |
| LPM | 0.136*** | 0.007 |
| Logit | 0.231*** | 0.009 |
| Probit | 0.255*** | 0.010 |

*$^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$*

While the AME values from ANN and econometric models are numerically similar, their interpretations can differ due to the underlying assumptions and model structures. ANNs capture complex, non-linear interactions and can provide detailed insights into how marginal effects vary across the feature spectrum, whereas econometric models typically assume constant marginal effects across all values of a feature. Therefore, the ANN's AME offers a nuanced understanding that complements the findings from conventional econometric models.

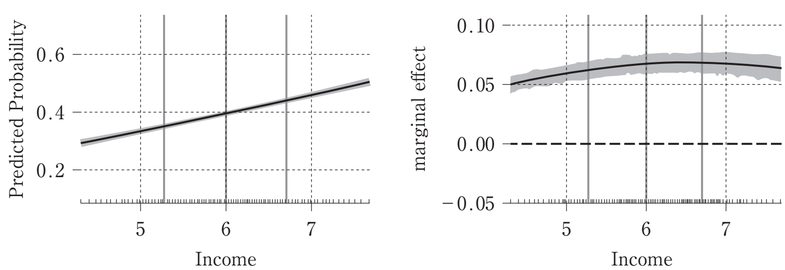
This detailed look into the relationship between net worth and stock market involvement highlights the complex role of financial resources in shaping investment behaviors. The results show that higher net worth increases the chance of stock market involvement and reveals different levels of impact among various wealth groups. These insights are crucial for developing focused financial education and inclusion plans, aiming to expand stock market participation among people from a range of economic situations.

Our analysis, highlighting the positive influence of total net worth on stock market participation, finds resonance with multiple pieces of the literature. Fessler et al. (2008) illustrate that stock holdings are predominantly found among wealthier households, underscoring wealth as a pivotal factor of stock market participation. This is further supported by Buccioli and Miniaci (2015), who find that larger wealth holdings often correlate with more aggressive investment strategies, suggesting that wealthier individuals are more likely to participate in the stock market. Bertaut (1998) adds that perceptions of the costs and benefits of market participation vary significantly with wealth levels, indicating that higher net worth not only enhances the capacity to invest but also influences the perceived value of investment opportunities.

**Income Analysis**

Figure 7 shows the influence of income on stock market participation through ALE plots and the marginal effect analysis. The ALE plot shows a positive, almost straight-line, relationship between income and the likelihood of investing in the stock market, highlighting how higher income boosts the chances of market participation. Further analysis in the marginal effects panel supports this, revealing a nearly horizontal trend in the impact of income, indicating its consistent influence across various income levels. This trend, with a marginal upward shift, suggests that the positive effect of income on stock market involvement is steady, yet slightly intensifies as income rises. Stabilizing at an effect level of around 0.05, the analysis clarifies that income's influence on investment decisions remains constant, reinforcing the idea that a higher income consistently enhances the probability of participating in the stock market.

**Figure 7**: ALE plot and Approximated marginal effect plot of income




We then compared these AME value of income from ANN with those obtained from the LPM, logit, and probit models. The results are summarized in Table 5. The AME for income from the ANN is 0.063 with a standard error of 0.001, which is statistically significant at the 1% level ($p<0.01$). The LPM coefficient for income is 0.065, indicating a similar effect size. The logit and probit models provide marginal effects of 0.042 and 0.050, respectively, both significant at the 1% level.

**Table 5**: Average Marginal Effects (AME) for Income

| Model | AME | Standard Error |
|---|---|---|
| ANN | 0.063*** | 0.001 |
| LPM | 0.065*** | 0.007 |
| Logit | 0.042*** | 0.007 |
| Probit | 0.050*** | 0.008 |

*$p<0.1$, **$p<0.05$, ***$p<0.01$

The analysis underscores a critical role of income in encouraging stock market participation, showing a consistent and linear impact unlike the varied effects of net worth. This uniform influence across all income levels underlines the essential role of financial resources in stock market participation. The insights from this study emphasize the need for financial strategies that are inclusive and address entry barriers for people at different income levels, enhancing our understanding of the economic features that drive stock market participation.

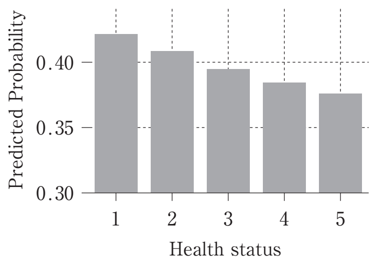
The critical role of income in facilitating stock market participation, as our analysis suggests, is corroborated by several studies within the literature. Gao and Yao (2019) emphasize the dominant influence of disposable income on the decision to engage in the stock market, noting this feature's importance across different demographics and market conditions. This is complemented by the analysis of Choi et al. (2000), which explores how income levels affect investment behaviors within corporate 401(k) plans, a common vehicle for stock market investment in the U.S., shedding light on broader patterns of financial decision-making related to income. Furthermore, Hong et al. (2004) extend this narrative by highlighting the role of social interactions in investment decisions, indirectly pointing to the income effect on participation rates through the lens of social dynamics and peer effects. Collectively, these studies highlight the nuanced yet pronounced impact of income on the tendency towards stock market investment, aligning with our findings of a consistent and positive relationship between income levels and the likelihood of market participation.

In conclusion, while the AME values from the ANN, LPM, logit, and probit models are similar, the machine learning method within the ANN framework provides additional insights. It allows us to observe how the marginal effect of a feature varies across different values of that feature, revealing more detailed dynamics in the relationship between net worth, income, and stock market participation. In contrast, the LPM, logit, and probit models assume a constant marginal effect across all values of the feature. This nuanced understanding underscores the advantage of using machine learning methods for exploring complex relationships in financial behavior.

### 3.3.2 Link Health Status to Stock Market Participation

Unlike numerical features like total net worth and income, health status is a ranked category from excellent to poor, treated as a numerical feature in our analysis. We illustrate the likelihood of stock market involvement for each health level using the ALE plot. Figure 8 displays this ALE plot, showing that worsening health, represented by higher health status values, leads to a lower chance of participating in the stock market. The analysis highlights a clear pattern: unhealthy individuals are less likely to invest in stocks. This situation likely arises from increased financial challenges or a tendency to avoid riskier investment choices due to health issues.

**Figure 8**: ALE plot of health status




The average marginal effects (AME) of health status from the ALE plots show that a one-unit worsening in health status decreases the probability of stock market participation by approximately 1.1% (see Table 6). Comparatively, the LPM shows a similar effect of −0.020. The logit and probit models provide smaller, non-significant effects of −0.006 and −0.008, respectively. This indicates that the ANN and LPM models highlight a significant negative impact of declining health on stock market participation, while the logit and probit models do not show significant effects.

**Table 6**: Average Marginal Effects (AME) for Health Status

| Model | AME | Standard Error |
|---|---|---|
| ANN | −0.011*** | 0.001 |
| LPM | −0.020*** | 0.006 |
| Logit | −0.006 | 0.005 |
| Probit | −0.008 | 0.005 |

*$p<0.1$, **$p<0.05$, ***$p<0.01$

The relationship between health status and stock market participation, as revealed in our ALE plot analysis, finds strong support in the existing literature. Specifically, Rosen and Wu (2004) explain how health significantly influences the ownership and composition of financial assets, with individuals in poorer health being less inclined towards risky assets, a finding echoed in our research. Edwards (2008) further supports this by showing how perceived health risks lead to more conservative financial decisions, particularly after retirement. Additionally, Yogo (2016) presents a life-cycle model where health status directly impacts asset allocation, reinforcing our observation that individuals with declining health are less likely to participate in stock investments. Together, these studies underscore the nuanced yet significant impact of health on financial behaviors, aligning closely with our findings that declining health corresponds with a reduced likelihood of stock market participation.

Up to now, we have mainly looked at health and economic features, such as net worth, income, and health status. However, our findings also shed light on other insightful aspects of stock market participation. We notice a gender gap, with men more likely to invest than women, hinting at underlying disparities in investment activities. Additionally, we find that higher medical expenses might correlate with increased stock market activity, suggesting that those with more spending in this area could have greater disposable income or a unique approach to risk.

Our analysis of employment status reveals that those working full-time or retired are more engaged in the stock market, likely due to their financial stability and available time. In contrast, individuals with other employment statuses might be held back by financial limitations or uncertainties. Marital status also plays a role, with single individuals showing a stronger interest in investing than those married or in relationships, pointing to the effects of personal circumstances on investment decisions.

These insights, including observations on gender, medical expenses, work status, and marital status, though not the primary focus of our study, offer a broader view of the diverse features influencing stock market participation. Recognizing the importance of these additional variables enhances our understanding of investment behaviors and underscores the importance of developing financial education and inclusion strategies that address the varied backgrounds and situations of potential investors.

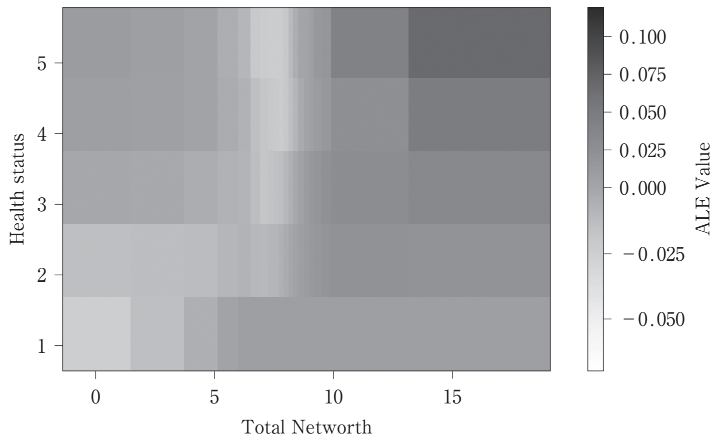
#### 3.3.3 Interplay Between Wealth and Health Status in Stock Market Participation

In this section, we delve into a detailed analysis of how health status interacts with economic features like net worth and income to shape investment behaviors. We employ two-dimensional Accumulated Local Effect (ALE) plots to illustrate the combined impact of health and wealth on stock market participation, offering a comprehensive view of their mutual relationship.

**Net Worth and Health Status**

Using a two-dimensional ALE plot, we explore how net worth and health status together influence the probability of stock market participation (Figure 9). The plot reveals that individuals in better health derive more substantial benefits from higher net worth in terms of increased likelihood of investing in the stock market. Specifically, as we move from better to poorer health (bottom to top on the vertical axis), the positive effect of net worth on stock market participation decreases. This suggests that good health enhances the capacity to invest and boosts confidence in making financial decisions.

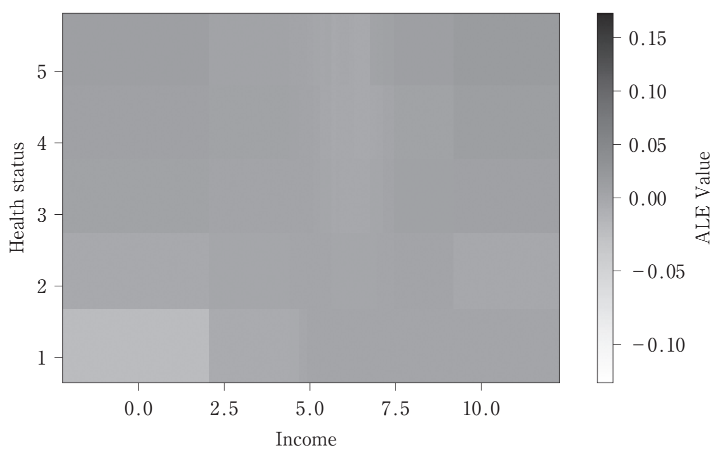
**Figure 9**: 2D ALE plot of health status and total net worth




**Income and Health Status**

Similarly, we analyze the interaction between income and health status using a two-dimensional ALE plot (Figure 10). The findings indicate that while healthier individuals experience a positive effect of income on stock market participation, this effect is less pronounced than that observed with net worth. The plot shows that better health combined with higher income positively influences stock market participation, but the overall impact is mild compared to the net worth interaction.

**Figure 10**: 2D ALE plot of health status and total income




Comparing these interactions across different health statuses, the plots consistently show that healthier individuals benefit more from their economic resources. This highlights the pivotal role of health in enhancing the positive effects of wealth and income on investment decisions. Healthier individuals are better positioned to leverage their net worth and income for stock market participation, while those in poorer health see a diminished benefit.

These insights underscore the importance of considering health status in financial planning and policy-making. Strategies aimed at increasing stock market participation should account for the interplay between health and economic factors, tailoring approaches to assist individuals with varying health and economic backgrounds. Policies that support health improvements could indirectly boost financial participation by enhancing individuals' ability to utilize their economic resources effectively.

## 4. Conclusion

This study explored the relationship between health, net worth, income, and their collective influences on stock market participation among older adults, using data from the 2020 Health and Retirement Study (HRS) and employing an advanced machine learning technique, the Artificial Neural Network (ANN). The use of ANN was effective in capturing non-linear relationships and complex interactions between health, net worth, and income. Our model's analytical capabilities are reflected in an AUC-ROC score of 0.84 and an AUC-PR score of 0.63, obtained through rigorous 5-fold cross-validation. These scores, significantly exceeding baseline expectations, highlight the model's precision and reliability in predicting stock market participation, demonstrating its ability to distinguish between potential participants and non-participants, even in a dataset with class imbalance.

We found that both net worth and income positively correlate with the likelihood of engaging in the stock market, and this trend is stronger among healthier individuals. Additionally, individuals in poorer health are less likely to invest, likely due to financial limitations or a cautious approach to risk. Our feature importance analysis identified economic features, such as total net worth and income, as the most significant predictors of stock market participation. The detailed insights provided by our ALE plots show how these features influence investment behavior across their different values, highlighting the varying impacts of economic features on stock market participation.

Our analysis centered on two critical relationships: the direct associations of health and economic status with stock market participation, and the interaction between health and economic status in shaping investment behavior. This approach allowed us to dissect not only the individual influences of health and wealth on older individuals' participation in the stock market but also how these features work together to influence decision-making. By comparing these results with those obtained from traditional econometric models, we demonstrated that machine learning provides a more nuanced understanding of these relationships, capturing complexities that econometric models may overlook.

Despite the insights provided, our analysis is limited by the specific demographics of the dataset. Future research could further elucidate these relationships by broadening the demographic scope and extending the analysis across different time frames to better understand how these dynamics shift in varying economic landscapes and demographic changes.

This research contributes to the ongoing discussion on financial behavior through machine learning analysis, highlighting the crucial roles of health and economic features in financial planning and policy. The insights gained are invaluable for financial advisors and policymakers and set the stage for future studies that aim to enhance stock market inclusivity and participation across diverse economic and health backgrounds.

## Acknowledgment

I extend my sincere thanks to Professor Makoto Kakinaka, Graduate School of Economics, Ritsumeikan University, and Professor Ching-Yang Lin, Graduate School of International Relations, International University of Japan, for their invaluable advice and feedback on this manuscript.

**Reference**

Ahmed, Jishan, Robert Green, Md Alauddin, Md Hasnat Jaman, and Goutam Saha, "Explainable Machine Learning Approaches to Assess the COVID-19 Vaccination Uptake: Social, Political, and Economic Aspects," 2022.

Almenberg, J., & Dreber, A. (2015). Gender, stock market participation and financial literacy. *Economics Letters, 137,* 140-142.

Amidu, M., Abor, J. Y., Coffie, W., & Akakpo, A. A. (2021). Financial inclusion, livelihood Activities, and stock Market participation. *Emerging Economy Studies, 7*(1), 23-61.

Antulov-Fantulin, Nino, Raffaele Lagravinese, and Giuliano Resce, "Predicting bankruptcy of local government: A machine learning approach," Journal of Economic Behavior & Organization, 2021, 183, 681-699.

Apley, Daniel W and Jingyu Zhu, "Visualizing the effects of predictor variables in black box supervised learning models," Journal of the Royal Statistical Society: Series B (Statis-tical Methodology), 2020, 82 (4), 1059-1086.

Ariza-Garzón, Miller Janny, Javier Arroyo, Antonio Caparrini, and Maria-Jesus Segovia-Vargas, "Explainability of a machine learning granting scoring model in peer-to-peer lending," Ieee Access, 2020, 8, 64873-64890.

Bellotti, Anthony, Damiano Brigo, Paolo Gambetti, and Frédéric Vrins, "Forecasting recovery rates on non-performing loans with machine learning," International Journal of Forecasting, 2021, 37(1), 428-444.

Bertaut, C. C. (1998). Stockholding behavior of US households: Evidence from the 1983-1989 survey of consumer finances. Review of Economics and Statistics, 80(2), 263-275.

Breiman, L., 2001b. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical Science 16, 199-231.

Bucciol, A., & Miniaci, R. (2015). Household portfolio risk. Review of Finance, 19(2), 739-783.

Cavapozzi, D., Trevisan, E., & Weber, G. (2013). Life insurance investment and stock market participation in Europe. *Advances in Life Course Research, 18*(1), 91-106.

Chia-Cheng, C., Liu, Y., & Hsu, T. H. (2019). An analysis on investment performance of machine learning: an empirical examination on Taiwan stock market. International Journal of Economics and Financial Issues, 9(4), 1.

Choi, J. J., Laibson, D., & Metrick, A. (2000). Does the internet increase trading? Evidence from investor behavior in 401(k) plans.

Chollet, F. (2021). *Deep learning with Python.* Simon and Schuster.

Cole, S. A., & Shastry, G. K. (2009). *Smart money: The effect of education, cognitive ability, and financial literacy on financial market participation* (pp. 09-071). Boston, MA: Harvard Business School.

Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240).

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., & Peijnenburg, K. (2016). Ambiguity aversion and household portfolio choice puzzles: Empirical evidence. *Journal of Financial Economics, 119*(3), 559-577.

Edwards, R. D. (2008). Health risk and portfolio choice. *Journal of Business & Economic Statistics, 26*(4), 472-485.

Fagereng, A., Gottlieb, C., & Guiso, L. (2017). Asset market participation and portfolio choice over the life-cycle. *The Journal of Finance, 72*(2), 705-750.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters, 27*(8), 861-874.

Fessler, P., & Schürz, M. (2008). Stock holdings in Austria. Monetary Policy & the Economy Q, 2, 83-100.

Gao, M., Meng, J., & Zhao, L. (2019). Income and social communication: The demographics of stock market participation. *The World Economy, 42*(7), 2244-2277.

Goldman, D., & Maestas, N. (2013). Medical expenditure risk and household portfolio choice. *Journal of Applied Econometrics, 28*(4), 527-550.

Hong, H., Kubik, J. D., & Stein, J. C. (2004). Social interaction and stock-market participation. The journal of finance, 59(1), 137-163.

Hosmer Jr, D. W., Lemeshow, S., Sturdivant, R. X., 2013. Applied logistic regression. volume 398. John Wiley & Sons.

Jarmulska, Barbara, "Random forest versus logit models: Which offers better early warning of fiscal stress?" Journal of Forecasting, 2022, 41(3), 455-490.

Kilinc, Mehmet Serdar and Robert Rohrhirsch, "Predicting customers' cross-buying decisions: a two-stage machine learning approach," Journal of Business Analytics, 2022, pp. 1-8.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980.

Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S., 2017. Self-normalizing neural networks. Advances in neural information processing systems 30.

Kumbure, M. M., Lohrmann, C., Luukka, P., Porras, J., 2022. Machine learn-ing techniques and data for stock market forecasting: A literature review. Expert Systems with Applications 197, 116659.

Love, D. A. (2010). The effects of marital status and children on savings and portfolio choice. *The Review of Financial Studies, 23*(1), 385-432.

Lundberg, S. M., Lee, S. I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems 30.

Montesinos López, O. A., Montesinos López, A., Crossa, J., 2022. Funda-mentals of artificial neural networks and deep learning, in: Multivariate statistical machine learning methods for genomic prediction. Springer, pp. 379–425.

Pompian, M. M. (2012). *Behavioral finance and wealth management: how to build investment strategies that account for investor biases* (Vol. 667). John Wiley & Sons.

Rosen, H. S., & Wu, S. (2004). Portfolio choice and health status. *Journal of Financial Economics, 72*(3), 457–484.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one, 10*(3), e0118432.

Tran, Kim Long, Hoang Anh Le, Thanh Hien Nguyen, and Duc Trung Nguyen, "Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam," Data, 2022, 7(11), 160.

Tratkowski, G. (2021). Changes in Feature Importance for European Stock Market Determinants Caused by the COVID-19 Pandemic Shock.

Valencia, Oscar, Diego A Parra, and Juan Camilo Díaz, "Assessing Macro-Fiscal Risk for Latin American and Caribbean Countries," 2022.

Yıldırım, Mustafa, Feyza Yıldırım Okay, and Suat Özdemir, "Big data analytics for default prediction using graph theory," Expert Systems with Applications, 2021, 176, 114840.

Yogo, M. (2016). Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets. Journal of monetary economics, 80, 17–34.

Zhang, Wen, Shaoshan Yan, Jian Li, Xin Tian, and Taketoshi Yoshida, "Credit risk prediction of SMEs in supply chain finance by fusing demographic and behavioral data," Transportation Research Part E: Logistics and Transportation Review, 2022, 158, 102611.

Zhang, Zijiao, Chong Wu, Shiyou Qu, and Xiaofang Chen, "An explainable artificial intelligence approach for financial distress prediction," Information Processing & Manage-ment, 2022, 59(4), 102988.

# APPENDIX

**Table A1.** Variable Description

| Variables Name | Definition |
|---|---|
| | *Outcome variable* |
| 1. Stock Ownership | stockholders '1' and non-stockholders '0'. |
| | *Feature variables* |
| 1. Log total net worth | Logarithm of the net value of total wealth |
| 2. Log income | Logarithm of total income received last calendar year. |
| 3. Age | Age of respondents in years. |
| 4. Number of children | Number of living children of the respondent and spouse or partner. |
| 5. Year of schooling | Year of education completed. |
| 6. Medical expense | Logarithm of total out-of-pocket medical expenditures in the last two years. |
| 7. Risk tolerance | Scale from 0 to 10 indicating willingness to take financial risks. |
| 8. Health status | Five categories from excellent to poor; excellent, very good, good, fair, poor. |
| 9. Has health insurance | Indicates if the respondent has health insurance (1 for yes, 0 for no). |
| 10. Has life insurance | Indicates if the respondent has life insurance (1 for yes, 0 for no). |
| 11. Gender | Male or female (1 for male, 0 for female). |
| 12. Work status | Seven categories; full-time, part-time, unemployed, partly retired, retired, disabled, not in the labor force. |
| 13. Marital status | Six categories; married, widowed, divorced, never married, separated, married but spouse absent |
| 14. Location | Nine divisions of the US; New England, Mid Atlantic, EN Central, WN Central, S Atlantic, ES Central, WS Central, Mountain, Pacific. |