Article

Ownership of Cash Value Life Insurance among Rural Households: Utilization of Machine Learning Algorithms to Find Predictors

Wookjae Heo [1,*] **, Eun Jin Kwak** [2] **, John Grable** [3] **and Hye Jun Park** [1]

1 Division of Consumer Science, Purdue University, Marriot Hall 242, 900 Mitch Daniels Blvd., West Lafayette, IN 47906, USA; park1707@purdue.edu
2 Austin E. Cofrin School of Business, University of Wisconsin, Green Bay, WI 54311, USA; eunjin.kwak@uga.edu
3 Department of Financial Planning, Housing, and Consumer Economics, College of Family and Consumer Science, University of Georgia, Athens, GA 30602, USA; grable@uga.edu
* Correspondence: heo28@purdue.edu; Tel.: +1-(765)-494-9847

Abstract: This study examines the determinants of life insurance ownership with a focus on rural areas and farming households in the United States. Utilizing data from online surveys conducted in 2019 and 2021, this paper explores how psychological factors, financial knowledge, and household characteristics influence life insurance ownership. Traditional indicators like wealth, income, and age were evaluated alongside less frequently discussed variables such as farm loans and rural residency. Machine learning techniques, including neural networks, Support Vector Machine modeling, Gradient Boosting, and logistic regression, were employed to identify the most robust predictors of life insurance demand. The findings reveal that farming-associated factors, particularly holding a farm loan and living in a farming household, significantly predict life insurance ownership. The study also highlights the complexity of life insurance demand, showing that financial education and management practices are critical determinants. This research underscores the need for tailored financial risk management strategies for rural and farming households and contributes to a nuanced understanding of life insurance demand in varying contexts.

Keywords: life insurance; rural; farm loan; machine learning; farmer

MSC: 37N40; 62P40; 91-10

1. Introduction and Research Purpose

Based on statistics compiled by Statista [1], the majority (52%) of Americans owned a life insurance policy in 2021 (a drop of 11% from 2011). Slightly more than 33% of those who owned a policy in 2021 reported that their primary reason for holding insurance was to replace income in the event of the insured's death. Another 30% noted that they held insurance for burial and financial expenses. Other reasons reported by Statista, and more broadly in the literature (e.g., [2]), as to why households own life insurance are to pay off a mortgage, as a tool to transfer wealth to a later generation, to pay for home care, as a way to supplement retirement income, to create estate liquidity, to generate tax-advantaged income, and to facilitate the transfer of business ownership. Reasons households fail to purchase life insurance include perceptions of a lack of need, policy costs, a lack of trust in financial intermediaries, worries about health qualifications, and misconceptions regarding the taxation of premiums and benefits [2].

While there is general consensus regarding the demographic determinants of life insurance demand in general, it is worth noting that much of the existing life insurance ownership literature is based on broad samples of financial decision-makers. It is unusual for researchers to examine ownership patterns based on the geographic location and the

debt holdings of the policy owner. The existing literature that does address these issues shows distinct differences in life insurance ownership between those living primarily in rural versus urban settings. Currently, 80% of the US population lives in urban areas [3]. Those who live in rural areas (approximately 14% to 20% of the US population) are more likely to hold wealth in a business rather than in financial assets such as life insurance.

In contrast, those living in urban areas tend to accumulate and hold wealth through the ownership of a primary residence (although homeownership rates are higher in rural areas), retirement plans, and taxable investment products. In many rural regions of the United States, rural household net worth is dominated by agricultural and farming assets (e.g., land and buildings, farm equipment, and real estate) [4]. The Employee Benefit Research Institute [3] notes that rural households tend to hold less diversified wealth, resulting in greater income instability and economic uncertainty. Even though a need for life insurance exists among rural households, in terms of premiums paid and coverage penetration, urban households are more likely to own life insurance [5]. Considering that farm owners and operators are exposed to death by injuries, respiratory disease, and stomach cancer at a higher rate than the US population [6], the lower prevalence of life insurance ownership among rural households, especially those with farming assets, suggests that some rural farm-owning households are failing to take advantage of risk transference strategies. One outcome of this study is to provide evidence of the association between life insurance ownership, living in a rural area, and holding a farm loan to gain insight into the financial risks faced by those living in rural areas.

The primary aim of this study was to estimate whether living in a farming household (i.e., being a farmer) and holding a farm loan can be used to predict life insurance ownership, holding other factors constant. It is important to note that the findings presented in this paper may not be generalizable in countries where the purchase of life insurance is a mandatory obligation for those who obtain farm loans. In the United States and Canada, borrowers are rarely required to purchase life insurance as an element of obtaining a loan. Lenders may offer mortgage or debt-repayment life insurance; however, they generally do not require the purchase of these products to obtain a loan. Even with this potential generalization limitation, this study offers a unique insight into the demand and ownership of life insurance. Much of the previous research has relied on data obtained from urban households. These data sources almost always exclude information about farm assets and liabilities. Additionally, as will be described in the conceptual background of this paper, existing studies have generally largely focused on demographic and financial characteristics when examining life insurance ownership. While valuable, these types of studies sometimes overlook topics and factors unique to those living in rural areas. It is important to consider the distinctive characteristics of those living outside of urban centers in order to gain a comprehensive insight into the demand and ownership of life insurance.

Beyond the descriptive nature of this study in examining the life insurance ownership patterns of those living in rural areas, this study adds significantly to the existing literature by showing how machine learning techniques can be used to uncover previously under-researched ownership and behavioral patterns. In this study, we employed neural network (NNs), Support Vector Machine (SVM) modeling, Gradient Boosting (GB), and logistic models to produce high-performance outcome measures through the evaluation of complex frameworks [7]. In contrast to traditional analytical methods that focus on identifying marginal effects, which can be highly beneficial when identifying the explanatory power of individual variables, the machine learning models used in this study allowed for a more comprehensive analysis of variables that might otherwise remain unexplored. Rather than omit variables based on the limitations of an analytical framework, the machine learning tools employed in this study allowed for numerous seemingly unrelated factors to be evaluated simultaneously [8], providing a pathway for a more robust outcome prediction through the identification of hidden variable layers that, until now, have remained unknown [9]. By leveraging the advantages of machine learning, this study allowed us to assess the significance of rural-specific variables concurrently with more traditional descrip-

tors of life insurance ownership. Rather than advancing new algorithms, this study shows that it is possible to use available and commonly known machine learning techniques to evaluate the importance of life insurance demand variables that are related specifically to those living in rural areas.

Although the overarching aim of this study was to evaluate whether living in a farming household and holding a farm loan can be used to predict life insurance ownership, three additional outcomes were anticipated with the analysis. The first was to show that it is possible to move beyond traditional modeling techniques when identifying the determinants of life insurance demand among those living in a farming household. The second outcome was the identification of the best predictors of cash value life insurance ownership. The third outcome was to provide a list of the most important predictors across periods, showing similarities and differences in ownership patterns. In the context of these expected outcomes, the study revealed a complex interplay between traditional and less frequently discussed variables as factors influencing life insurance ownership decisions. Psychological factors, financial knowledge, and specific demographic and household characteristics were all found to play significant roles, with farming-associated features emerging as critical predictors. These insights contribute to a more nuanced understanding of life insurance demand, particularly in a rural context, and highlight the importance of considering a broad range of factors in research and policy discussions.

The remainder of this paper is structured as follows. Section 2 provides a literature review and describes the hypotheses. Section 3 provides an overview of the financial and mathematical concepts that provide the foundation for the analyses. A description of the dataset and the operationalization of the variables is presented in Section 4. In Section 5, the different methodologies utilized in the study are discussed. This is followed by the presentation of results in Section 6. The paper concludes with a discussion and conclusion in Sections 7 and 8, respectively.

2. Review of Literature

Numerous studies have been conducted over the past 50 years to identify generalized factors, household characteristics, and economic indicators that can be used to explain and predict life insurance ownership (e.g., [10–30]). Much of this literature shows that life insurance ownership patterns (and changes in whom owns life insurance) are linked to current and shifting demographic factors (e.g., gender, income, wealth, and life expectancy), variations in the tax code, the proliferation of competing investment opportunities (i.e., a substitution effect), and changing preferences [31]. In addition, health concerns prompted by the COVID-19 pandemic can now be added to the list of factors associated with life insurance purchase decisions [32,33]. Given that the effects of the COVID-19 pandemic are expected to persist in the economy for decades [34], coupled with the relative economic instability facing many rural households, it behooves policymakers and researchers to take a fresh look into the determinants of life insurance demand of those living primarily in rural areas in the presence of a pandemic situation.

The paucity of literature examining life insurance ownership patterns among US farming households does not mean that insurance is an under-discussed topic in rural studies, risk management, and economics journals. What it does mean, however, is that much of the literature tends to focus on crop and weather insurance (e.g., [35]), market risks (e.g., [36]), and the life insurance needs of those living outside the United States [31,37]. Nonetheless, a casual perusal of internet sources shows that life insurance firms and state extension agencies keenly understand the vital role life insurance plays in maintaining the health of rural America [38]. Consider the work of [39]. They noted that the relatively advanced age of farming households in the United States raises questions about the best way to manage farm property ownership, tenure, and transfer. The insurance industry has quickly addressed this dilemma by pointing out how life insurance can provide a tax-efficient way to transfer wealth from one generation to another, reduce family insolvency resulting from the death of the primary farmer in a family, and create a retirement income stream.

The following literature review provides an overview of the various factors associated with the decision to hold life insurance across different households. Much of the reviewed literature shares a common perspective that life insurance is a valuable financial management tool for households, regardless of geographical location or livelihood [40,41].

2.1. Psychological Factors and Cash Value Life Insurance

The literature is replete with descriptions of associations between life insurance ownership and self-concept variables, financial satisfaction, financial stress, and financial risk tolerance. Of particular importance are self-concept variables such as self-discipline and self-control [42]. Authors of [43,44] argued that investigating impulse control mechanisms, particularly self-discipline, can provide insight into the mechanisms describing consumer saving decisions. In this regard, [45] examined how self-control influences the demand for life insurance. Rabbani noted that the likelihood of holding cash value life insurance increases with one's level of self-discipline.

Regarding financial satisfaction, several studies have shown that cash value life insurance, as a tax-advantaged savings and investment vehicle, can enhance household stability and satisfaction [27,45]. In other words, cash value life insurance provides life cycle protection while simultaneously providing a means to accumulate wealth and supplement income [45]. The positive relationship between holding life insurance and financial satisfaction mirrors the association between health insurance and life satisfaction [46].

Researchers have reported inconsistent correlations between financial risk tolerance and the ownership of cash value life insurance [30,41,47,48]. For example, [45] reported that the likelihood of owning cash value life insurance decreases with risk tolerance. In other studies, however, the relationship is different, with those who are unwilling to take on financial risk (i.e., those whose financial risk aversion level is high) being more likely to own cash value life insurance [49,50]. Some researchers have observed a non-significant relationship between risk tolerance and holding life insurance. For example, the authors of [27,43] found that the decision to own cash value life insurance is unlikely to depend on an individual's level of risk tolerance.

2.2. Financial Knowledge, Financial Characteristics, and Cash Value Life Insurance

The association between financial knowledge (often referred to as financial literacy in the literature) and life insurance ownership has been extensively studied (e.g., [40,51–53]). The authors of [54] noted that individuals can enhance their financial knowledge through formal and informal education. Increased knowledge leads to a rise in human capital that facilitates a better understanding of health and life insurance. Among existing studies that use life insurance demand as a dependent variable, most show a positive correlation between policy ownership and financial knowledge, financial literacy, and insurance literacy. In some studies, financial knowledge was assessed through items measuring insurance knowledge. In almost all cases, a positive relationship with life insurance demand has been reported [37,55–58]. Stated another way, individuals with more financial knowledge are more likely to purchase life insurance. A similar relationship has been noted when financial knowledge is measured objectively. Those with higher knowledge test scores are more likely to own life insurance [40,59–62]. On the other hand, the relationship between holding life insurance and subjective financial knowledge is less robust [59,63]. For example, the authors of [64] noted that subjective financial knowledge did not impact life insurance demand.

Numerous researchers have explored the relationship between income, wealth, and holding cash value life insurance (for a review of this literature, see [30,43]). The consensus from these reports is that individuals with a higher income level tend to hold more life insurance [45]. Likewise, those with higher savings intentions are more likely to own cash value life insurance. On the other hand, some researchers, including [40], have observed a negative relationship between life insurance demand and net household assets (including a personal residence). In this regard, insurance is seen as a substitute for wealth rather than a

complement. However, it is important to note that cash value life insurance may stimulate savings among young households [65]. In contrast, the demand for cash value life insurance among affluent households may be attributed to estate planning considerations [27].

2.3. Demographics, Health, Household Characteristics, and Cash Value Life Insurance

Almost all published studies on the demand for life insurance include a standard set of demographic variables, such as household composition, age, gender, education, and race/ethnicity [24,27,30,43,51]. A study by [15] is representative of this literature. They reported that shifts in life insurance demand in the United States are partially attributable to the high portion of single households, delayed marriage, and a resulting decrease in birth rates. The ownership of life insurance has also been linked with being female, having higher educational attainment, and being married [45], although these relationships are sometimes inconsistent and sometimes not always significant. Holistically, it is reasonable to expect that young adults with a Bachelor degree or higher level of education are more likely to own cash value life insurance. According to [45], the impact of household composition on cash value life insurance ownership may be mitigated, as households managed by a married couple can experience varying levels of income depending on the dependency status of children.

The demand for life insurance is also influenced by health-related variables. Consider health expenditures. Spending on health costs, as a percentage of gross domestic product (GDP), can impact the demand for life insurance in several ways [66]. For example, as aggregate expenditures on health care increase, the demand for life insurance may decrease as households may be resource-constrained. Additionally, the correlation between health expenditures and survival probabilities can affect the demand for different types of insurance products, including life insurance. The significance of health expenditures in describing the demand for life insurance implies that the health status of insurance applicants might also be associated with the demand for life insurance. Individual health status and demographic characteristics, such as age and number of dependents, can also play a role in explaining the demand for life insurance. For instance, some studies have shown that life expectancy is positively associated with life insurance demand [67].

2.4. Farm-Associated Features and Cash Value Life Insurance

The variable associations described in the preceding discussion highlight the types of variables typically included in life insurance ownership studies. It is important to note, however, that the demand for cash value life insurance among farm households is likely influenced by additional factors associated with farm operations. For example, asset values related to the ownership of livestock, buildings, and machinery are known to be positively correlated with total insurance expenditures [68]. Additionally, factors like farm size, income level, and education play a significant role in determining the demand for life insurance among farm owners and those living in rural areas [69]. Moreover, studies have shown that the demand for life insurance is positively correlated with factors such as real GDP per capita, savings, and the probability of the primary wage earner's death [64]. As such, the demand for cash value life insurance among rural households can be expected to be multifaceted and influenced by a combination of farming-associated factors such as holding a farm loan. Gaining a better understanding of these factors is crucial for insurance providers and policymakers to tailor products, incentives, and regulations that meet the needs of this specific demographic group.

2.5. Hypotheses

Based on observations from the reviewed literature, it is reasonable to expect that psychological factors, financial knowledge, financial characteristics, demographic characteristics, health status, household features, and farming-associated features, when viewed in a holistic model, improve the prediction rate for the ownership of cash value life insurance. In this regard, the following hypotheses were tested in this study:

H₁: Psychological factors, including financial risk tolerance, self-discipline, and financial satisfaction, are significant predictors of cash value life insurance ownership.

H₂: Financial knowledge and financial characteristics are significant predictors of cash value life insurance ownership.

H₃: Demographic characteristics, health status, and household features are significant predictors of cash value life insurance ownership.

H₄: Farming-associated factors, such as holding a farm loan, are significant predictors of cash value life insurance ownership.

2.6. Hypotheses Text with Machine Learning

Machine learning is premised on the notion that when two or more variables are included in a model, "hidden layers" exist. These hidden layers represent the conceptual area between the input and output of a prediction algorithm. Modeling hidden layers involves transforming nonlinear inputs into a network [7]. The resulting weights of each layer can provide insight into the valid role that one or more variables play in describing a phenomenon. This modeling approach can be explained by the concepts embedded in complex system science models [70]. A complex system is composed of many interconnected components. These components interact nonlinearly, making it difficult to observe variable effects and challenging to predict certain behaviors or social outcomes. When viewed as a complex system [71], it is possible for seemingly insignificant factors to be associated with patterns of life insurance ownership. This insight is premised on the notion that complex systems consist of diverse, interconnected parts that describe behavioral outcomes nonlinearly [23]. The nonlinear feature of a complex system makes it difficult to predict how changes in one significant factor within the system will affect the overall system. To gain the most robust insight into life insurance ownership behavior, it is important to consider various factors that could conceptually be associated with a household's life insurance holding choice, including living in a farm household and holding a farm loan. This study shows that rather than classifying households according to urban or rural residence status, more precise measures of rurality can provide more meaningful insight into the prediction of the ownership of life insurance.

Complex systems can be found in many fields, including ecology, economics, the social sciences, and engineering [7]. A complex system is one in which a holistic outcome pattern exists based on alterations that occur among the interactions of various factors [23]. These features are generally referred to as dynamics or dynamism—a mixture of internal and external factors that describe complex outcomes. In this study, internal and external factors can be considered potential descriptors of life insurance ownership. In the context of explaining the demand for life insurance, factors can be classified as (a) psychological, (b) financial knowledge, (c) demographic characteristics, (d) health behaviors, and (e) household financial characteristics [23,50,72–83]. By testing the above hypotheses, this study expands this list to include farming-associated factors (i.e., living in a farming household and having a farm loan) as unique indicators of a rural lifestyle.

3. Conceptual Background

Meaningful commentary on the life insurance marketplace has highlighted differences in ownership between urban and rural residents and the unique risks faced by those engaged in farming activities. However, the specifics of these differences remain underexplored. Much of the existing literature assumes similar determinants of life insurance ownership across urban and rural areas, based on empirical evidence and theoretical work by [24,84]. They conceptualized life insurance demand as a multi-step process, starting with estimating the present face value of policies, which is assumed to be less than expected premiums. Subsequent steps include motivational factors, such as the need to provide consumption for dependents when wealth is insufficient [2]. The final purchase

decision involves comparing the adjusted present value of dependents' consumption to net household wealth, factoring in the probability of the breadwinner's death, insurance price, and the dependents' risk aversion. This model suggests that higher death probability, risk aversion, and future consumption lead to purchasing life insurance, while greater wealth and higher coverage costs reduce demand. Thus, few differences in demand between urban and rural households are expected.

However, practical observations often differ from these expectations. Life insurance ownership has historically been associated with household composition, net wealth, and income [2], along with micro- and macro-environmental factors and behavioral aspects [23]. For instance, [50] found that the death of a loved one can prompt relatives to explore life insurance options.

Most studies on life insurance ownership patterns have used traditional statistical methods. Due to the market's complexity, rarity, dynamic nature, and a lack of publicly available data, experimental methods or data simulations are rarely used to examine life insurance ownership (with some exceptions, such as [85–87]). This study aims to fill this gap by demonstrating how machine learning models can provide deeper insights into life insurance demand. The following discussion highlights the machine learning approaches used in this study.

3.1. Using Complexity as a Prediction Tool to Understand the Concept of Insurance Ownership

A foundational concept embedded in machine learning methodologies is managing complexity. Conceptually, the definition of complexity regarding the demand for life insurance is founded on demand theory as described in economics. The demand function is normally defined as the sum of associated factors, including the price of goods or services and consumers' income. Consumers' income is thought to be constrained (i.e., restricted by budget limitations) when viewed from the supply side. The basic formula for demand is as follows:

$$Q = f(P, I) \tag{1}$$

where, Q is the quantity of demand; P is the price of a good or service; and I is household or consumer income. Over the years, researchers have expanded the formula to include additional factors, including product/service quality, consumer preferences, and market dynamics. In terms of gaining a better understanding of the demand for life insurance, some researchers have expanded the model to include variables like education, household composition or family structure, marital status, religion, income, net worth, employment, inflation, and interest rates (e.g., [13,19,22,23,28,43]).

As the basic demand function expands, so does the complexity of the formula. Today, life insurance demand models exhibit characteristics common to other complex systems. As noted above, it is already known that life insurance ownership is associated with a diverse array of psychological, financial, demographic, and household characteristic factors [23,50,71]. Conceptually, the complexity of describing life insurance demand can be represented as follows:

$$L_t = f(Psy_t, Fin_t, Dem_t, Health_t, House_t, Farm_t) = f(X_{kt}) \tag{2}$$

where, L_t is the ownership of life insurance; Psy includes psychological factors; Fin includes financial factors; Dem includes demographic factors; $Health$ includes health-related factors; $House$ includes household characteristics; $Farm$ represents living in a farming household and holding a farm loan; and t and $t-1$ represent the survey time ($t = 2021$; $t - 1 = 2019$). Equation (2) describes the scenario where the data are considered cross-sectionally. In addition to the cross-sectional approach, by leveraging machine learning capabilities, it is possible to predict the 2021 data using the 2019 data. The equation for this approach is represented as Equation (3):

$$PL_t = f(Psy_{t-1}, Fin_{t-1}, Dem_{t-1}, Health_{t-1}, House_{t-1}, Farm_{t-1}) = f(X_{kt-1}) \tag{3}$$

While Equation (2) can directly incorporate machine learning techniques, in the case of Equation (3), the dynamic system uses a formula that utilizes the periodic differences [88,89] as shown in Equation (4):

$$PL_t = L_{t-1} + \left[\alpha_o + \sum_1^k b_i X_{kt-1} \right] + \epsilon_t = f(X_{kt-1}) \tag{4}$$

Machine learning tools using Equations (2) and (4) can be utilized to predict life insurance demand for a given year and the following year's prediction. Equations (2) and (4) can also be merged into machine learning algorithms using various prediction models, including (a) Support Vector Machine (SVM) modeling, (b) Gradient Boosting modeling, (c) neural network modeling, and (d) logistic regression modeling. Specifically, each year's predictors are vectors of factors as noted in Equation (5):

$$X_{tl} = [xk\ tl] = \begin{bmatrix} Psy_{tl} \\ Fin_{tl} \\ Dem_{tl} \\ Health_{tl} \\ House_{tl} \\ Farm_{tl} \end{bmatrix} \tag{5}$$

where l is the time indicator of two periods (t and $t-1$) and k is each input factor.

3.2. Research Hypothesis Testing and SVM

Hypotheses 1 through 4 were tested using an SVM analysis. For H_1, psychological factors such as financial risk tolerance, self-discipline, and financial satisfaction were included in the model. The test of H_2 incorporated financial knowledge and financial characteristics. H_3 included demographic variables, health status indicators, and household features. H_4 was tested using farming-associated features, including holding a farm loan and being a farmer.

In terms of modeling, the factor vectors of Equation (5) can be incorporated into an SVM model [90] as shown in Equation (6):

$$\underset{w,\ b,\ \xi_i,\ \xi_i^*}{min} = \frac{1}{2} w^T w + C \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{6}$$

$$\text{subject to constraints} \begin{cases} (L_{ti}) - (w \cdot \phi(X_{tl}) + b) \leq \epsilon + \xi_i \\ (w \cdot \phi(X_{tl}) + b) - (L_{tl}) \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

where, w is the weight vector; b is a bias term; ξ_i and ξ_i^* are slack variables that are non-negative values used to handle mathematical error within the margin of ϵ; C is the regulatory parameter that controls the trade-off between max-margin and the min-margin; ϵ is the margin of tolerance to exclude the penalty for the errors; $\phi(X_{tl})$ is the kernel function that can be linear, polynomial, sigmoid, or a radial basis function (RBF). Equations (7) to (10) show these linear, polynomial, RBF, and sigmoid kernels, respectively:

$$K(X_{ti},\ X_{tj}) = X_{ti}{}^T X_{tj} \tag{7}$$

$$K(X_{ti},\ X_{tj}) = \left(\gamma X_{ti}{}^T X_{tj} + r \right)^d \tag{8}$$

$$K(X_{ti},\ X_{tj}) = \exp\left(-\gamma ||X_{ti} - X_{tj}||^2 \right) \tag{9}$$

$$K(X_{ti},\ X_{tj}) = \tanh\left(\gamma X_{ti}{}^T X_{tj} + r \right) \tag{10}$$

where, K denotes the kernel; T means an operation of matrix transposing used to flip the data over their diagonal; X_{ti} is the i-th vector at a time of t or $t-1$; X_{tj} is the j-th vector at a time of t or $t-1$; γ in Equations (9) and (11) scales the dot product; r is the added value to the dot product; d is the degree of the polynomial; and γ in Equation (10) is the parameter to control the width of the Gaussian function. Within an SVM model, the hyperplane is set as the classification criteria to predict a particular pattern of kernels [8]. The hyperplane is optimally estimated when the maximum margin is found. The factors vector is then used to predict the outcome (L_t). Finally, the prediction is made following Equation (11):

$$L_t = w \cdot \phi\big([X_{ti} \ X_{tj}]\big) + b \tag{11}$$

3.3. Research Hypothesis Testing and GB

Hypotheses 1–4 were then tested using Gradient Boosting. In the case of GB (see [80,91,92]), the set of factors from Equation (5) were inserted into the model through a residual function similar to Equation (12). When making a prediction, the residual function takes the negative form to indicate the actual loss value. Gradient Boosting uses the negative residual function to find the best order of variables for prediction [92].

$$r_m = -\frac{\partial L(L_t, F_{m-1}(X_{tl}))}{\partial F_{m-1}(X_{tl})} = y - F_{m-1}(X_{tl}) \tag{12}$$

where, r_m is the residual with m-th iteration; the loss function utilizes the differential (∂) between the outcome and the factors; and F_{m-1} is the predicted value of the vector X_{tl} with the iteration of $(m-1)$. The estimated residual (r_m) fitted to a regression tree is accomplished using Equation (13):

$$h_m = argmin\sum(r_m - h(X_{tl}))^2 \tag{13}$$

where, h_m is the regression tree with m-th iteration; $argmin\sum(r_m - h(X_{tl}))^2$ is the minimized sum of squared differences between the residual and the prediction. Finally, optimization is estimated by minimizing the gap between the residual and prediction. This is accomplished with Gradient Boosting. Equation (14) describes the complete model [93], where v is the learning rate:

$$F_m(X_{tl}) = F_{m-1}(X_{tl}) + vh_m(X_{tl}) \tag{14}$$

3.4. Research Hypothesis Testing and NNs

An NN was then used to test the four hypotheses. In the case of NN modeling, the factors vector can be directly inserted into the prediction [94] as shown in Equation (15):

$$y = a\left(\sum_{k=1}^{K} w_k X_{ti} + e\right) \tag{15}$$

where, w_k represents the estimated weights on each factor (k, K). NN modeling resembles how humans arrive at decisions via neuron connectivity. The model assumes all factors are connected to the outcome through neurons [95]. This implies that all model factors are connected. The weights are explored by assuming connectivity, allowing for the best prediction weight to be estimated.

3.5. Research Hypothesis Testing and the Logistic Model

The full dataset was used to test the four research hypotheses. As the last model, the complex demand function can be extended to a logistic model [96] by taking the set of factors (X_{tl}) and inserting them into Equation (16) below, where E is the exponential:

$$Y = f(X_{tl}) = \frac{1}{1 + e^{-(b_0 + \sum b X_{tl})}} \tag{16}$$

Each equation described thus far can be utilized with a complex set of factors (x_{tl}) to describe and predict the demand for and ownership of life insurance. One outcome of this study is identifying the model that provides the most robust prediction of life insurance demand using a wide assortment of interrelated variables.

4. Research Procedure

4.1. Data

Two online surveys were distributed in 2019 and 2021. An online survey agency was used to select a random sample of participants to complete the cross-sectional surveys. The survey agency sent invitations to panels of individuals. The panels were overweighted to represent rural households. The choice of who would receive an email was randomized. The sampling process stopped once the target sample size was obtained (i.e., 1000 individuals in 2019 and 2021). Within each dataset, a certain number of participants (three in 2019 and twelve in 2021) failed to answer all the required questions. As a result, the sample size in 2019 was 997, whereas the sample size in 2021 was 988. Table 1 shows the descriptive statistics for the combined samples.

Table 1. Descriptive statistics for the complete sample (N = 1985).

	Mean	SD	Frequency	Percentage
Psychological factors				
Self-esteem	25.01	2.80		
Life satisfaction	21.24	8.54		
Locus of control	17.69	6.16		
Financial satisfaction	20.19	3.16		
Financial stress	63.08	26.33		
Financial risk tolerance	26.34	5.04		
Financial self-efficacy	12.66	4.27		
Financial knowledge				
Objective financial knowledge	1.76	1.01		
Subjective financial knowledge	3.96	1.64		
Had finance class in high school			515	25.94%
Had finance class in college			441	22.22%
Financial characteristic				
Income level				
Less than $15k			288	14.51%
$15k–$25k			244	12.29%
$25k–$35k			282	14.21%
$35k–$50k			284	14.31%
$50k–$75k			330	16.62%
$75k–$100k			227	11.44%
$100k–$150k			215	10.83%
Over $150k			115	5.79%
Zero net worth			229	11.54%
Negative net worth			639	32.19%
Emergency fund			992	49.97%
Own home			1091	54.96%
Have mortgage			674	33.95%
Home Equity Line of Credit (HELOC)			319	16.07%
Auto loan			722	36.37%
Student loan			569	28.66%
Eating habits and health				
Beer/week	2.03	5.68		
Glasses liquor/week	1.81	4.89		
Soft drink/week	3.89	7.83		
Fruit/week	3.47	6.28		
Vegetables/week	4.19	6.62		

Table 1. *Cont.*

	Mean	SD	Frequency	Percentage
Cigarettes/week	13.81	34.58		
Perceived health	2.91	0.80		
Demographic				
Employed			1286	64.79%
Being single			936	47.15%
Female			1277	64.33%
Education level				
High school or lower			514	25.89%
Associate			572	28.82%
Bachelor			590	29.72%
Graduate or higher			309	15.57%
Farming-associated factors				
Living in a farming household			290	14.61%
Farm loan			208	10.48%

4.2. Outcome Variable

Ownership of a cash value life insurance policy was this study's outcome variable of interest. Life insurance ownership was assessed in 2019 and again in 2021. Research participants were asked if they currently owned a cash value life insurance. A positive response was coded 1, otherwise 0. In 2019, 249 participants reported owning a policy. In 2021, 290 participants indicated owning a policy.

4.3. Independent Variables

The following categories of independent variables were used in the modeling process: (a) psychological factors, (b) financial knowledge, (c) demographic characteristics, (d) health behaviors, (e) household financial characteristics, and (f) farming-associated factors. A discussion of each independent variable classification follows. Individual variables comprising each category are described in Table 2.

Table 2. Operationalization of independent variables.

Category and Variables	Coding
Psychological factors	
Self-esteem	10 items; 4 point Likert style scale; minimum = 10; maximum = 40; lower = low self-esteem; higher = high self-esteem.
Life satisfaction	5 items; 7 point Likert style scale; minimum = 5; maximum = 35; lower = low satisfaction; higher = high satisfaction
Locus of control	7 items; 5 point Likert style scale; minimum = 7; maximum = 35; lower = external locus of control; higher = internal locus of control
Financial satisfaction	7 items; 5 point Likert style scale; minimum = 7; maximum = 35; lower = low satisfaction; higher = high satisfaction
Financial stress	24 items; 5 point Likert style scale; minimum = 24; maximum = 120; lower = low stress; higher = high stress
Financial risk tolerance	13 items; minimum = 13; maximum = 47; lower = low financial risk tolerance; higher = high financial risk tolerance
Financial self-efficacy (Reverse)	6 items; 5 point Likert style scale; minimum = 6; maximum = 30; lower = high self-efficacy; higher = low self-efficacy
Financial knowledge	
Objective financial knowledge	3 items; binary (correct = 1; incorrect = 0); minimum = 0; maximum = 3; low = do not know well about finance; high = know well about finance

Table 2. *Cont.*

Category and Variables	Coding
Subjective financial knowledge	1 item; 7 point Likert style question; minimum = 1; maximum = 7; lower = low subjective financial knowledge; higher = high subjective financial knowledge
Took finance class in high school	1 item; binary (yes = 1; no = 0)
Took finance class in college	1 item; binary (yes = 1; no = 0)
Financial characteristics	
Income level	1 item; 1 = less than \$15k; 2 = \$15k–\$25k; 3 = \$25k–\$35k; 4 = \$35k–\$50k; 5 = \$50k–\$75k; 6 = \$75k–\$100k; 7 = \$100k–\$150k; 8 = Over \$150k
Zero net worth	1 item; binary (net worth is zero = 1; no = 0)
Negative net worth	1 item; binary (net worth is negative = 1; no = 0)
Emergency fund	1 item; binary (have emergency fund = 1; no = 0)
Own home	1 item; binary (own house = 1; no = 0)
Have mortgage	1 item; binary (have mortgage = 1; no = 0)
HELOC	1 item; binary (have loan = 1; no = 0)
Auto loan	1 item; binary (have loan = 1; no = 0)
Student loan	1 item; binary (have loan = 1; no = 0)
Farm loan	1 item; binary (have loan = 1; no = 0)
Eating habits and health	
Beer/week	1 item; number of beer bottles per week
Glasses liquor/week	1 item; number of liquor glasses per week
Soft drink/week	1 item; number of times per week
Fruit/week	1 item; number of frequencies per week
Vegetables/week	1 item; number of frequencies per week
Cigarettes/week	1 item; number of frequencies per week
Perceived health	1 item; 1 = poor health; 2 = fair health; 3 = good health; 4 = excellent health
Demographic	
Employed	1 item; binary (employed = 1; not working = 0)
Being single	1 item; binary (single = 1; couple = 0)
Female	1 item; binary (female = 1; male = 0)
Education level	1 item; 1 = high school or lower; 2 = some college (Associate degree); 3 = college graduate (Bachelor degree); 4 = graduate or higher
Farming-associated factors	
Living in a farm household Farm loan	1 item; binary (farmer = 1; non-farmer = 0) 1 item; binary (have loan = 1; no = 0)

Psychological factors. Seven psychological factors were included in the models: (a) self-esteem, (b) life satisfaction, (c) locus of control, (d) financial satisfaction, (e) financial stress, (f) financial risk tolerance, and (g) financial self-efficacy.

Financial knowledge. Four financial knowledge variables were included in the models. Financial knowledge was measured objectively and subjectively. Additionally, whether a research participant had taken a high-school finance class (coded 1, otherwise 0) or a personal finance class in college (coded 1, otherwise 0) was included in the models.

Household financial characteristics. The following ten household-level financial characteristics were included in the models: (a) income, (b) zero net worth, (c) negative net

worth, (d) emergency fund, (e) owning a home, (f) mortgage, (g) home equity line of credit (HELOC), (h) auto loan, and (i) student loan.

Eating habits and health behaviors. The following health-related behaviors were included in the models: (a) the number of beers consumed per week, (b) glasses of liquor consumed per week, (c) soft drinks consumed per week, (d) the number of fruits eaten per week, (e) the number of vegetables eaten per week, (f) the number of cigarettes consumed per week, and (g) perceived health status (excellent to poor).

Demographic characteristics. The following participant demographic characteristics were included in the models: (a) currently working (code 1, otherwise 0), (b) marital status (single coded 1, otherwise 0), (c) gender (self-identified female coded 1, otherwise 0), and (d) educational status.

Farming-associated factors. Two variables were used as indicators of farming status: (a) self-identifying as living in a farming household (i.e., being a farmer) and (b) indicating holding a farm loan. Both variables were coded 1, otherwise 0.

5. Data Analysis Method

5.1. Machine Learning Algorithms

Different machine learning techniques were used to determine which provides the most robust insight into describing the determinants of life insurance ownership. Machine learning encompasses a variety of artificial intelligence statistical procedures, each of which can be designed to identify trends and patterns between model inputs and outputs. Nearly all machine learning approaches aim to apply weights to inputs as indicators of importance in describing (or predicting) an outcome [23,97]. For instance, a neural network (i.e., a representative machine learning algorithm) is often built on the notion (i.e., a function) of hidden layers or neurons. A hidden layer (or neuron) is a mathematical function that explains an outcome in terms of probabilities.

As described in the conceptualization section of the paper, the concept underlying machine learning methodologies is the identification of hidden layers that can be used to pinpoint functions that may not be independently significant but can nonetheless be important when combined with other factors in a network [23]. Machine learning is a powerful tool that can be used to make classifications and identify patterns to make better predictions using a potential combination of predictors [98,99]. The reliability of machine learning outcomes depends on the type of data used, the characteristics of the outcome, and the independent variables included in a model [100]. Depending on the data, machine learning techniques are known to produce more valid descriptions of behavior (i.e., classification power) compared to Naïve Bayes, Linear Discriminant Analysis, logistic regression, K-nearest neighbors, decision trees, Supportive Vector Machine modeling, adaptive boosting, and Gradient Boosting methods. However, machine learning techniques sometimes underperform linear, polynomial, lasso, and ridge regressions when the outcome variable is measured on a continuous scale. In addition, clusters such as hierarchical, density-based, k-means, and GMM clustering are more efficient in finding subgroups compared to unsupervised machine learning models. However, machine learning algorithms are narrowing this gap in performance consistency. Machine learning based on principal component analysis, recursive feature elimination, and a model-based selection is sometimes better at identifying the features of learning. Neural networks (i.e., machine learning) are considered reliable, valid, and robust across different types of data and characteristics of measurements [101]. For example, assume a cognitive scientist wants to identify someone's processes when identifying a face in a photographic image. A snippet of any eye, mouth, or nose will not be enough to describe the face fully. Using traditional statistical techniques, the scientist will likely conclude that a picture of a mouth is not independently significant in describing a face. However, this conclusion is likely incorrect. Patterns between facial features do help explain the process of facial identification. In other words, the joining of hidden layers is the actual mechanism used to describe a facial image. The same mechanism likely exists in the context of life insurance ownership patterns. By comparing several

machine learning techniques, the current study extends the facial recognition analogy to the application of describing life insurance demand. While it is reasonable to expect that factors such as wealth, income, and age will be direct predictors of life insurance demand, it is also reasonable to expect other variables, working as hidden layers, to emerge as important in explaining as much or more of the demand for life insurance. Using the orange package with Python, a methodological outcome associated with this study was to provide evidence supporting this assertion.

Given the overarching aim of this study, the following machine learning techniques were tested to determine which approach provides the most robust classification outcome: (a) a neural network, (b) SVM modeling, (c) Gradient Boosting, and (d) a logistic regression model. The neural network approach was selected because it is a multifunctional machine learning technique that matches the data type used in this study (Abiodun et al., 2018). As a classical machine learning technique, SVM modeling was utilized to align with similar methodologies reported in the literature (e.g., [8,102]). Gradient Boosting was selected because it represents a set of ensemble learning algorithms that amplify weak learning to strong learning outcomes [9]. Finally, a logistic regression model was selected because it matches what is most often used for classification purposes in the life insurance literature. Each machine learning approach was utilized in the first stage of the study. The results from each test were then evaluated and compared via prediction performance. After all comparisons were made, the best-performing analytical technique was selected, with the variables identified by the technique reported as the study's main results.

Figure 1 illustrates the three stages of analysis used in this study. In the first stage (Stage A), data were split into training and testing datasets. In the second stage (Stage B), data from the training stage were used to estimate optimized forecasting algorithms and models for each period (i.e., each machine learning technique was tested at this analysis stage). The third stage of analysis (Stage C) focused on estimating variable weights (based on identified patterns) using the training data and then selecting the best prediction model. For interpretation purposes, the weights of the variables from Stage C indicate the importance of each variable when predicting the ownership of cash value life insurance. The weights from the best prediction model represent optimal combinations of variables that can be used to describe life insurance ownership. Once the model was selected using the training data, the model was evaluated using the testing data. The purpose of the test was to cross-validate the results.
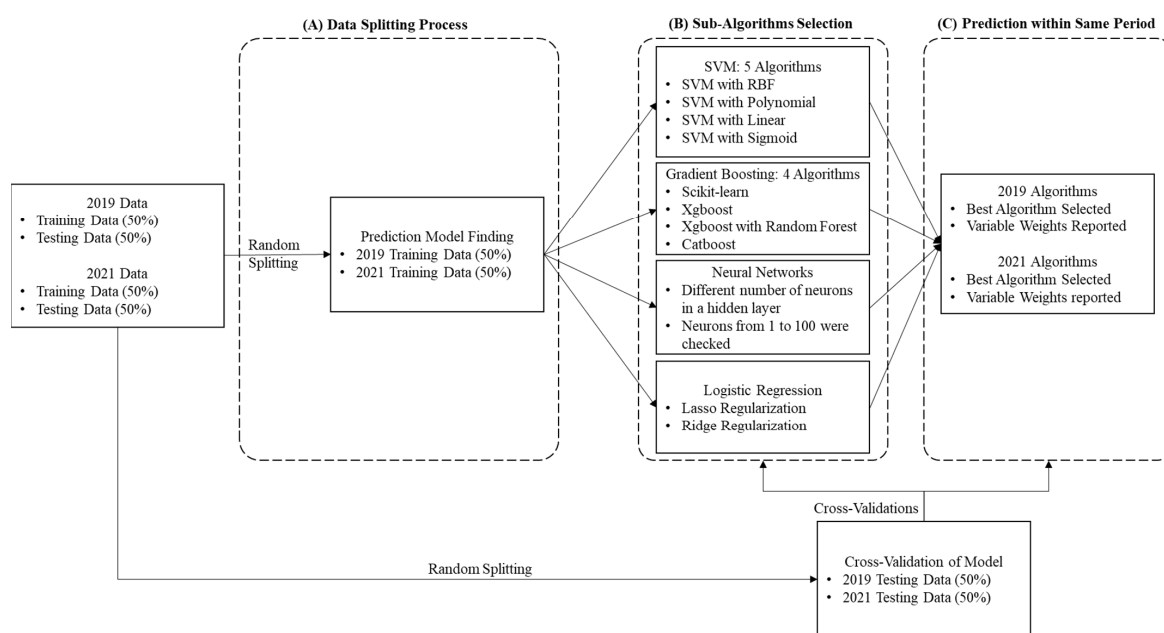

Figure 1. Stages of data analysis.

5.2. Machine Learning Algorithms, Alterations, and Parameter Settings

The performance of a machine learning test will generally differ based on the assumptions underlying the model and how the parameters are set [103]. For instance, the number of hidden layers (or neurons) can create differences in how a neural network performs. Additionally, assumptions regarding kernel assumptions can alter SVM results. These insights serve as a reminder that it is important to understand and clarify the conditions and parameters of a particular model before selecting a machine learning algorithm. As shown in Figure 1 (Stage B) and in Table 3, the number of neurons ranged from one to 100. In the case of SVM modeling, four kernel assumptions were used: (a) linear, (b) polynomial, (c) radial basis, and (d) sigmoid. For Gradient Boosting, four sub-algorithms were used: (a) scikit-learn, (b) extreme boosting, (c) extreme boosting in combination with random forest [104], and (d) categorical boosting [105]. The logistic regression was assumed to be a lasso regression or ridge regression. The logistic regression procedure used in this study differs from the conventional logistic model. A conventional logistic regression uses a one-time attempt at estimating coefficients. When employed as an algorithm within a machine learning context, the estimation is conducted multiple times based on the separation of the sample (i.e., splitting the sample into training and testing datasets occurs many times). The logistic regression results shown in Figure 1 represent one of the machine learning algorithms. As shown in Table 3, consistent parameters within the same machine learning techniques were utilized.

Table 3. Machine learning conditions and parameters.

Machine Learning	Altered Conditions	Preset of Parameters
Neural network	Number of neurons = 1 Number of neurons = 2 . . . Number of neurons = 100	(All) Activation = ReLu; Solver = Adam; Regularization α = 0.0002; maximum number of iterations = 200
SVM	Liner kernel: x·y	Cost = 1.00 Regression loss ε = 0.10 Numerical tolerance = 0.001 Iteration limit = 1000
	Polynomial kernel: $(g\ x \cdot y\ c)^d$	Cost = 1.00 Regression loss ε = 0.10 g = auto c = 1.00 d = 3.0 Numerical tolerance = 0.001 Iteration limit = 1000
	Radial basis function (RFB) kernel: $\exp(-g\,|x-y|^2)$	Cost = 1.00 Regression loss ε = 0.10 g = auto Numerical tolerance = 0.001 Iteration limit = 1000
	Sigmoid kernel: $\tanh(g\ x \cdot y + c)$	Cost = 1.00 Regression loss ε = 0.10 g = auto c = 1.00 Numerical tolerance = 0.001 Iteration limit = 1000

Table 3. *Cont.*

Machine Learning	Altered Conditions	Preset of Parameters
Gradient Boosting	Scikit-learn	Number of trees = 100 Learning rate = 0.10 Replicable training = Yes Limit depth of tree = 3 No split when subset < 2 Fraction of training = 1.00
	Extreme boosting	Number of trees = 100 Learning rate = 0.30 Replicable training = Yes Regulatory lambda = 1 Limit depth of tree = 6 Fraction of training = 1.00
	Extreme boosting with random forest	Number of trees = 100 Learning rate = 0.30 Replicable training = Yes Regulatory lambda = 1 Limit depth of tree = 6 Fraction of training = 1.00
	CatBoost	Number of trees = 100 Learning rate = 0.30 Replicable training = Yes Regulatory lambda = 1 Limit depth of tree = 6 Fraction of features = 1.00
Logistic regression	Lasso Ridge	(All) Strength c = 1.00

5.3. Evaluation Criteria

The model selected at Stage C of the analysis (i.e., the model deemed to be the best machine learning algorithm) was based on the following selection criteria: (a) precision (precision refers to the weighted average of the ratio of correctly predicted positive observations to the total predicted positive observations), (b) recall (recall is the ratio of correct prediction to all observations in the class. A score of 0.50 or higher for precision and recall is generally recommended), (c) F1 (F1 is the weighted average of precision and recall. Higher scores indicate a more robust model), and (d) AUC (AUC refers to the area under the curve. Scores closer to 1.0 indicate a more precise classification). These criteria were used to determine how well one of the models described life insurance ownership. The selection process aimed to maximize the number of true positives in the model while minimizing the number of false positives, which is normally called a Confusion Matrix. A higher classification accuracy (CA) means a better prediction rate. The weight order of the variables was determined by assessing (a) information gain (the information gain criterion indicates the degree to which a split in the data improves prediction), (b) the gain ratio (the gain ratio adjusts the information gain output using a normalizing term that reduces estimation bias), and (c) the Gini index (the Gini index is used to verify the model selection. When all predictors are present, the one that generates the smallest Gini split is the one that indicates variable optimization).

6. Results

The data were first split into a training and testing dataset. The training dataset was used to identify the variables of significance when describing life insurance ownership. The four machine learning techniques were then evaluated using the testing dataset. The results of these tests showed which of the four prediction models offered the most robust

description of life insurance ownership patterns. In the context of this study, validity was confirmed when the results from the training and testing datasets were in alignment. A model can be overfitted, which means the training model results are good, but the tests are poor [106]. It is also possible for a model to be underfitted. Overfitting is more problematic because the initial model does not, in practice, predict tested behavior. When underfitting occurs, a model adjustment is needed. The degree of over- or under-fit can be evaluated using the prediction error rate from the training and testing datasets. The following discussion summarizes the results from the tests of the four machine learning techniques.

6.1. Neural Network

Tables 4 and 5 show the best sub-algorithms across the neural network analysis. In 2019, the best prediction occurred with four neurons. In 2021, the best performance occurred with 11 neurons.

Table 4. Selection of optimal neural network using different number of neurons, 2019.

	Training Dataset					Testing Dataset				
	AUC	CA	F1	Precision	Recall	AUC	CA	F1	Precision	Recall
1	0.762	0.749	0.642	0.562	0.749	0.691	0.751	0.644	0.564	0.751
2	0.807	0.814	0.789	0.806	0.814	0.731	0.755	0.737	0.731	0.755
3	0.798	0.677	0.699	0.787	0.677	0.691	0.574	0.601	0.737	0.574
4	**0.856**	**0.818**	**0.796**	**0.809**	**0.818**	**0.739**	**0.753**	**0.737**	**0.731**	**0.753**
5	0.837	0.822	0.803	0.813	0.822	0.691	0.757	0.745	0.740	0.757
6	0.904	0.850	0.837	0.846	0.850	0.690	0.749	0.737	0.731	0.749
7	0.898	0.856	0.842	0.855	0.856	0.734	0.743	0.738	0.734	0.743
8	0.947	0.878	0.868	0.880	0.878	0.681	0.727	0.719	0.714	0.727
9	0.930	0.862	0.849	0.862	0.862	0.703	0.737	0.727	0.721	0.737
10	0.926	0.882	0.875	0.880	0.882	0.719	0.737	0.734	0.731	0.737
11	0.916	0.856	0.846	0.851	0.856	0.720	0.753	0.742	0.736	0.753
12	0.919	0.874	0.864	0.874	0.874	0.716	0.733	0.725	0.720	0.733
13	0.935	0.872	0.862	0.872	0.872	0.680	0.707	0.706	0.705	0.707
14	0.935	0.870	0.858	0.873	0.870	0.709	0.717	0.708	0.701	0.717
15	0.931	0.864	0.853	0.862	0.864	0.723	0.759	0.743	0.737	0.759
16	0.968	0.898	0.892	0.898	0.898	0.716	0.751	0.739	0.732	0.751
17	0.956	0.884	0.874	0.888	0.884	0.703	0.755	0.735	0.729	0.755
18	0.983	0.920	0.916	0.923	0.920	0.710	0.743	0.731	0.725	0.743
19	0.966	0.922	0.917	0.927	0.922	0.706	0.721	0.713	0.707	0.721
20	0.971	0.924	0.920	0.926	0.924	0.709	0.719	0.717	0.716	0.719
30	0.986	0.944	0.943	0.944	0.944	0.707	0.751	0.746	0.742	0.751
40	0.999	0.986	0.986	0.986	0.986	0.715	0.731	0.730	0.729	0.731
50	0.998	0.978	0.978	0.978	0.978	0.701	0.725	0.721	0.718	0.725
60	0.999	0.978	0.978	0.978	0.978	0.701	0.713	0.715	0.718	0.713
70	1.000	0.996	0.996	0.996	0.996	0.715	0.735	0.730	0.727	0.735
80	1.000	0.998	0.998	0.998	0.998	0.697	0.725	0.723	0.721	0.725
90	1.000	0.990	0.990	0.990	0.990	0.701	0.731	0.727	0.724	0.731
100	1.000	0.996	0.996	0.996	0.996	0.702	0.721	0.722	0.723	0.721

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

Table 5. Selection of optimal neural network using different number of neurons, 2021.

	Training Dataset					Testing Dataset				
	AUC	CA	F1	Precision	Recall	AUC	CA	F1	Precision	Recall
1	0.783	0.706	0.585	0.499	0.706	0.743	0.706	0.585	0.499	0.706
2	0.838	0.832	0.822	0.829	0.832	0.749	0.769	0.755	0.756	0.769
3	0.846	0.747	0.757	0.787	0.747	0.793	0.696	0.709	0.754	0.696
4	0.872	0.848	0.840	0.847	0.848	0.772	0.761	0.751	0.749	0.761
5	0.855	0.836	0.830	0.832	0.836	0.751	0.747	0.740	0.737	0.747
6	0.906	0.850	0.845	0.847	0.850	0.773	0.761	0.753	0.750	0.761
7	0.891	0.844	0.838	0.841	0.844	0.778	0.763	0.757	0.754	0.763
8	0.925	0.864	0.858	0.863	0.864	0.774	0.763	0.755	0.753	0.763
9	0.921	0.877	0.872	0.875	0.877	0.761	0.763	0.759	0.756	0.763
10	0.918	0.877	0.874	0.874	0.877	0.775	0.755	0.754	0.753	0.755
11	**0.918**	**0.870**	**0.865**	**0.869**	**0.870**	**0.796**	**0.779**	**0.776**	**0.774**	**0.779**
12	0.911	0.862	0.859	0.859	0.862	0.760	0.757	0.752	0.749	0.757
13	0.924	0.870	0.865	0.870	0.870	0.780	0.753	0.747	0.744	0.753
14	0.926	0.866	0.860	0.866	0.866	0.761	0.769	0.763	0.760	0.769
15	0.943	0.885	0.880	0.885	0.885	0.774	0.755	0.753	0.751	0.755
16	0.959	0.911	0.908	0.912	0.911	0.776	0.749	0.740	0.737	0.749
17	0.948	0.891	0.886	0.892	0.891	0.782	0.781	0.776	0.774	0.781
18	0.979	0.923	0.921	0.923	0.923	0.753	0.755	0.749	0.746	0.755
19	0.960	0.901	0.897	0.902	0.901	0.776	0.769	0.764	0.762	0.769
20	0.965	0.905	0.902	0.906	0.905	0.785	0.781	0.777	0.775	0.781
30	0.980	0.923	0.921	0.923	0.923	0.782	0.763	0.759	0.756	0.763
40	0.996	0.976	0.976	0.976	0.976	0.786	0.769	0.767	0.766	0.769
50	0.997	0.978	0.978	0.978	0.978	0.777	0.751	0.747	0.745	0.751
60	0.998	0.984	0.984	0.984	0.984	0.761	0.743	0.742	0.741	0.743
70	1.000	0.992	0.992	0.992	0.992	0.771	0.763	0.760	0.758	0.763
80	1.000	0.996	0.996	0.996	0.996	0.779	0.767	0.764	0.762	0.767
90	0.998	0.982	0.982	0.982	0.982	0.772	0.759	0.757	0.756	0.759
100	1.000	0.998	0.998	0.998	0.998	0.768	0.769	0.765	0.763	0.769

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

6.2. SVM

Tables 6 and 7 show the best sub-algorithms using the SVM technique. In 2019 and 2021, the RBF assumption produced the best-performing sub-algorithm.

Table 6. Selection of optimal SVM, 2019.

	Training Dataset					Testing Dataset				
	AUC	CA	F1	Prec.	Recall	AUC	CA	F1	Prec.	Recall
Sigmoid	0.572	0.645	0.652	0.659	0.645	0.629	0.669	0.680	0.696	0.669
RBF	**0.874**	**0.882**	**0.868**	**0.898**	**0.882**	**0.640**	**0.765**	**0.728**	**0.733**	**0.765**
Polynomial	0.815	0.822	0.817	0.814	0.822	0.617	0.659	0.666	0.675	0.659
Linear	0.481	0.553	0.579	0.626	0.553	0.515	0.530	0.561	0.623	0.530

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

Table 7. Selection of optimal SVM, 2021.

	Training Dataset					Testing Dataset				
	AUC	CA	F1	Prec.	Recall	AUC	CA	F1	Prec.	Recall
Sigmoid	0.667	0.678	0.676	0.674	0.678	0.642	0.628	0.631	0.635	0.628
RBF	**0.916**	**0.889**	**0.883**	**0.892**	**0.889**	**0.764**	**0.775**	**0.765**	**0.764**	**0.775**
Polynomial	0.906	0.860	0.859	0.859	0.860	0.733	0.721	0.724	0.729	0.721
Linear	0.452	0.486	0.508	0.555	0.486	0.455	0.478	0.500	0.544	0.478

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

6.3. Gradient Boosting

Tables 8 and 9 show the best sub-algorithms based on the Gradient Boosting approach. In 2019, CatBoost showed the best prediction; however, in 2021, the scikit-learn algorithm exhibited the best performance.

Table 8. Selection of optimal Gradient Boosting, 2019.

	Training Dataset					Testing Dataset				
	AUC	CA	F1	Prec.	Recall	AUC	CA	F1	Prec.	Recall
Scikit-learn	0.998	0.960	0.959	0.962	0.960	0.716	0.749	0.728	0.722	0.749
Xgboost, RF	0.867	0.852	0.840	0.848	0.852	0.671	0.713	0.704	0.698	0.713
Xgboost	1.000	1.000	1.000	1.000	1.000	0.713	0.749	0.740	0.734	0.749
CatBoost	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.722**	**0.759**	**0.728**	**0.727**	**0.759**

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

Table 9. Selection of optimal Gradient Boosting, 2021.

	Training Dataset					Testing Dataset				
	AUC	CA	F1	Prec.	Recall	AUC	CA	F1	Prec.	Recall
Scikit-learn	0.995	0.970	0.969	0.970	0.970	0.787	0.787	0.782	0.780	0.787
Xgboost, RF	0.882	0.860	0.853	0.860	0.860	0.785	0.759	0.746	0.745	0.759
Xgboost	1.000	1.000	1.000	1.000	1.000	0.755	0.767	0.760	0.758	0.767
CatBoost	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.742**	**0.793**	**0.766**	**0.776**	**0.793**

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

6.4. Logistic Regression

Tables 10 and 11 show the best sub-algorithms based on the logistic regression technique. In 2019 and 2021, the lasso assumption exhibited a better prediction rate than the ridge assumption.

Table 10. Selection of optimal logistic regression, 2019.

	Training Dataset					Testing Dataset				
	AUC	CA	F1	Prec.	Recall	AUC	CA	F1	Prec.	Recall
Ridge	0.824	0.790	0.765	0.771	0.790	0.707	0.751	0.736	0.729	0.751
Lasso	**0.821**	**0.796**	**0.770**	**0.780**	**0.796**	**0.714**	**0.755**	**0.739**	**0.733**	**0.755**

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

Table 11. Selection of optimal logistic regression, 2021.

	Training Dataset					Testing Dataset				
	AUC	CA	F1	Prec.	Recall	AUC	CA	F1	Prec.	Recall
Ridge	0.856	0.820	0.813	0.814	0.820	0.788	0.775	0.769	0.766	0.775
Lasso	**0.854**	**0.822**	**0.814**	**0.816**	**0.822**	**0.791**	**0.779**	**0.771**	**0.770**	**0.779**

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

6.5. Optimal Model Selection and Model Variable Weights

Table 12 shows the collection of the best sub-algorithms from Tables 4–11. The list includes the optimal modes from the neural network, the SVM, the Gradient Boosting, and the logistic regression models. Across the four approaches, the neural network (NN) offered the most robust level of classification performance. As such, the neural network was selected as the best-fit model.

Table 12. Best prediction models across the 2019 and 2021 periods.

	2019 Testing Dataset					2021 Testing Dataset				
	AUC	**CA**	**F1**	**Prec.**	**Recall**	**AUC**	**CA**	**F1**	**Prec.**	**Recall**
NN	**0.739**	**0.753**	**0.737**	**0.731**	**0.753**	**0.796**	**0.779**	**0.776**	**0.774**	**0.779**
SVM	0.640	0.765	0.728	0.733	0.765	0.764	0.775	0.765	0.764	0.775
GB	0.722	0.759	0.728	0.727	0.759	0.787	0.787	0.782	0.780	0.787
Logistic	0.714	0.755	0.739	0.733	0.755	0.791	0.779	0.771	0.770	0.779

Note. Bold characters indicate the best prediction outcome by optimal sub-algorithm.

6.6. Identification of Important Variables

As explained in the section above (Section 5.3), the identification of important variables was based on the following three indicators: (a) the information ratio, (b) the gain ratio, and (c) the Gini ratio [107]. The information gain criterion indicates the degree to which a split in the data improves prediction. The gain ratio adjusts the information gain output using a normalizing term that reduces estimation bias. The Gini ratio index is used to verify the model selection. When all predictors are present, the one that generates the smallest Gini split is the one that indicates variable optimization. The greater the ratio, the higher it is on the list of important variables.

As noted above, the neural network provided the most valid insight into patterns of life insurance ownership. The final step in the analytic procedure involved the identification of the most important variables in describing patterns of ownership based on the neural network algorithm. Table 13 shows these variables. As shown in Table 13, farming-associated factors emerged as the most important predictors of life insurance ownership. Specifically, holding a farm loan was the first ranked predictor of life insurance ownership in 2019 and 2021. Living in a farm household (i.e., being a farmer) also ranked highly across the analysis periods (sixth in 2019 and third in 2021). The implications associated with these findings are discussed below.

Table 13. Variable lists: ordered by weights used to predict the ownership of cash value life insurance.

Rank	2019 Variable	Info. Gain	Gain Ratio	Gini Ratio	2021 Variable	Info. Gain	Gain Ratio	Gini Ratio
1	Farm loan	0.037	0.137	0.023	Farm loan	0.106	0.160	0.067
2	Had financial class in college	0.056	0.100	0.033	HELOC	0.090	0.149	0.057
3	HELOC	0.052	0.088	0.031	Farmer	0.078	0.109	0.048
4	Own home	0.064	0.065	0.031	Had financial class in high school	0.101	0.106	0.059
5	Emergency fund	0.055	0.056	0.029	Emergency fund	0.101	0.102	0.054
6	Farmer	0.024	0.056	0.014	Had financial class in college	0.084	0.094	0.050
7	Had financial class in high school	0.025	0.041	0.015	Own home	0.079	0.079	0.044
8	Income level	0.041	0.021	0.023	Have mortgage	0.055	0.061	0.033
9	Auto loan	0.019	0.020	0.010	Life satisfaction	0.099	0.050	0.059
10	Negative net worth	0.017	0.020	0.008	Female	0.049	0.049	0.028
11	Life satisfaction	0.039	0.019	0.021	Auto loan	0.043	0.047	0.025
12	Subjective financial knowledge	0.036	0.018	0.019	Subjective financial knowledge	0.082	0.041	0.047
13	Perceived health level	0.026	0.016	0.012	Income level	0.078	0.039	0.048
14	Female	0.010	0.013	0.005	Cigarettes/week	0.059	0.039	0.036
15	Objective financial knowledge	0.023	0.012	0.014	Financial satisfaction	0.069	0.035	0.039
16	Cigarettes/week	0.013	0.011	0.007	Education level	0.063	0.032	0.038
17	Have mortgage	0.010	0.011	0.006	Self-esteem	0.060	0.030	0.035
18	Education level	0.020	0.010	0.011	Being single	0.028	0.028	0.016
19	Being single	0.009	0.009	0.005	Glasses liquor/week	0.041	0.027	0.024
20	Financial risk tolerance	0.016	0.008	0.009	Perceived health level	0.044	0.026	0.025
21	Glasses liquor/week	0.012	0.008	0.006	Employed	0.022	0.023	0.012
22	Financial satisfaction	0.015	0.007	0.008	Locus of control	0.045	0.023	0.027
23	Number of Bottle of Beers/week	0.009	0.006	0.005	Number of Bottle of Beers/weeks	0.031	0.020	0.018
24	Financial stress	0.009	0.004	0.005	Financial risk tolerance	0.038	0.019	0.022
25	Fruit/week	0.008	0.004	0.004	Financial self-efficacy	0.038	0.019	0.022
26	Vegetables/week	0.008	0.004	0.004	Financial stress	0.035	0.017	0.020
27	Self-esteem	0.007	0.004	0.004	Objective financial knowledge	0.028	0.015	0.016
28	Soft drinks/week	0.005	0.002	0.002	Student loan	0.021	0.011	0.013
29	Employed	0.002	0.002	0.001	Fruit/week	0.004	0.004	0.002
30	Financial self-efficacy	0.003	0.002	0.002	Vegetables/week	0.007	0.004	0.004
31	Locus of control	0.002	0.001	0.001	Soft drinks/week	0.006	0.003	0.004
32	Zero net worth	0.000	0.000	0.000	Zero net worth	0.000	0.001	0.000
33	Student loan	0.000	0.000	0.000	Negative net worth	0.001	0.001	0.000

7. Discussion

The literature on life insurance generally supports the notion that the demand for life insurance at the household level can be explained by household and financial decision-maker characteristics such as wealth, income, and age. Much of the existing literature suggests that as wealth increases, the demand for life insurance should decline (i.e., a substitution effect exists between these factors). Contrarily, as income increases, demand for life insurance is thought to intensify primarily to hedge the loss of income should a household breadwinner die. Older households are expected to own more life insurance than younger households, primarily because of the desire to pay for final expenses, create estate liquidity, and leave one or more bequests. What is most interesting about the results from this study is that only income and wealth, in 2019 but not 2021, were important when describing the life insurance ownership characteristics of those living primarily in rural areas. Nearly all the top 10 factors in 2019 and 2021 represent variables less frequently described in discussions of life insurance demand.

The findings from this study do not negate the value of traditional indicators of life insurance demand. Instead, the results from this study suggest that when hidden layers are analyzed empirically, the demand for life insurance becomes more complex and less transparent. Traditionally, non-significant variables are important when all interactions across and among variables are accounted for analytically. In alignment with traditional models of insurance demand, it was determined that the presence of farm loans increases the likelihood of owning a life insurance policy. When viewed holistically, the findings support the notion that household financial decision-makers attempt to reduce the burden of debt repayment in the event of death. In alignment with this insight, the results also show that living in a farming household is strongly associated with life insurance ownership. To overlook these two important factors when describing life insurance ownership patterns is to miss important descriptors of insurance demand.

In addition to this key takeaway from the study, the analysis also showed that life insurance ownership is associated with a household's financial management approach. Financial knowledge and behavioral characteristics were important in the models. Those who had taken a personal finance class either in high school or college were more likely to own a life insurance policy. Financial knowledge may be a pathway to understanding the importance of preparing for a loss of income in the event of death. Similarly, holding an emergency fund may be an outcome associated with financial literacy, and as such, this behavior may be an indicator of future planning intentions. Other factors, such as life satisfaction and identifying as a female, were also associated with life insurance ownership. The gender finding is not surprising, given that women tend to be more likely to seek help for financial questions while exhibiting worries about household outcomes [108]. The other factors suggest that rather than being a product purchased as a result of deteriorating health or declining psychological well-being, life insurance ownership is more closely aligned with a positive psychological and physiological outlook.

The one surprising variable from the analysis was reporting a negative net worth in 2019. In this study, those who indicated owning life insurance were more likely to report a negative net worth. This finding may be unique to 2019, the year before the COVID-19 pandemic. The relevance of this variable was significantly reduced in 2021. As suggested in media outlets at the time, it is possible that daily reports of COVID-19 transmission and death led some households to purchase life insurance, which negated the effect of wealth (and income) in 2021. It is also possible that wealth, rather than being a substitute for life insurance, is a demand feature that some households use to supplement low levels of net worth.

Findings from the machine learning analyses provided support for the study's hypotheses. First, in terms of H_1, life satisfaction (i.e., a psychological characteristic) was found to play an important role in describing the ownership of life insurance. This finding aligns with the existing literature that shows various psychological traits and attitudes are important when explaining life insurance ownership patterns [109]. Emotions and

psychological qualities are known to enhance the predictability of insurance demand [110]. By incorporating psychological factors like life satisfaction with traditional demographic and economic variables, insurers can better address the diverse preferences and motivations of individuals seeking life insurance coverage.

Second, in the case of H_2, financial knowledge, financial education, and some financial characteristics were found to rank highly as predictors of life insurance ownership. For instance, having taken a financial class while in college or high school was within the top 10 predictors in both years. In addition, financial characteristics (e.g., HELOC, owning a home, emergency fund, income level, etc.) were found to be important predictors across the two years. These findings indicate that those with higher financial literacy and greater financial management skills are more likely to own life insurance. This insight highlights the importance of financial knowledge in describing the demand for life insurance. This finding also aligns with what others have reported (e.g., [51]).

Third, in relation to H_3, health-associated factors and traditional indicators such as age, marital status, and gender concerns were found to be less important in the prediction of life insurance ownership. For instance, identifying as female was the only variable to be ranked in the top 15 among all predictors. Perceived health level ranked 20th in 2021. This does not mean that traditional determinants like age, marital status, education, and income are not important, but rather, when hidden layers between variables are assessed, other factors emerge as being more important. The lower ranking of these variables has been observed by others (e.g., [21,23]). Findings related to H_3 illustrate the need for researchers and insurance providers to shift towards considering a broader range of demographic and socioeconomic factors when describing and predicting life insurance ownership patterns.

In the case of H_4, farming-associated features, such as being a farmer and holding a farm loan, were found to be one of the best descriptors and predictors of cash value life insurance ownership. The presence of farm loans and living in a farming household were strongly associated with life insurance ownership, underscoring the need to consider farming-associated factors when analyzing the demand for life insurance among those living in rural areas. This finding highlights the need for more attention and research focused on those living in rural areas. This insight supports an assertion made by [111] who reported that there is a limited understanding of the social and economic needs of farm households across lifespans and that more research is needed to understand demand features when a rural household experiences birth, maternity, retirement, unemployment, poverty, illness, accidents, and death.

Finally, the use of machine learning techniques, such as those used in this study, underscores the practical utility that can be garnered by gaining an understanding of the complex relationships associated with life insurance ownership. By employing standard machine learning methods, this study demonstrates that even straightforward machine learning algorithms can significantly enhance the predictive accuracy involved in describing insurance demand. Stated another way, this study validates the notion that basic machine learning techniques are effective tools for uncovering complex patterns in life insurance ownership. By leveraging these methods, it is possible to gain deeper insights into the factors influencing financial decisions, thereby improving predictive models and informing policy discussions.

8. Conclusions

In conclusion, the findings from this study advance the life insurance and financial planning literature in meaningful ways, specifically for those living in rural areas of the United States. As explained in the introduction, those who live in a farming household tend to hold less diversified wealth by focusing on owning farming-associated assets. Doing so creates some instability of income and a general long-term level of economic uncertainty. In addition, farming, by its very nature, tends to increase rates of disability and death, which adds to the overall financial instability exhibited by those living in rural areas. The importance of farming-associated factors (i.e., being a farmer and holding a farm loan),

as described in this study, shows how important it is for educators, policymakers, and financial service professionals to help those living in a farming household and other rural households prepare for and deal with issues related to financial risk management.

Beyond describing the most important variables that can be used to describe life insurance ownership patterns, this paper shows that a machine learning methodology, based on a neural network approach, can provide unique insights into the demand for and ownership of life insurance. When hidden layers are analyzed empirically, many variables that have traditionally been considered secondary predictors of demand emerge as very important. This study also advances the literature by showing that the COVID-19 pandemic may have significantly influenced the demand for life insurance. Data from this study show that more rural households reported owning cash value life insurance after the pandemic. Finally, results from this study serve as a reminder that the demand features of life insurance likely differ based on region and urban versus rural status. This study illustrates life insurance's importance for households with farm/agricultural loans. This is a household characteristic that does not generally exist in an urban setting. As with all studies, the findings reported in this paper need to be evaluated in the context of the data collection approach. Multiple surveys were used over periods that represent the pre- and post-COVID-19 pandemic timeframe. Distributing surveys during these periods may have resulted in a sampling bias. It is also possible that by specifically overweighting rural households in the dataset, the results may not be generalizable beyond those whose demographic and socioeconomic profiles match the sample. Additional studies are needed to verify the results of this study. It would be particularly helpful if a panel survey could be conducted to better assess the life insurance needs of those living in rural areas.

Author Contributions: Conceptualization, W.H.; methodology, W.H.; software, W.H.; validation, E.J.K. and J.G.; formal analysis, W.H.; investigation, J.G. and H.J.P.; data curation, E.J.K.; writing—original draft preparation, W.H.; writing—review and editing, E.J.K., J.G. and H.J.P.; visualization, E.J.K. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rudden, J. Distribution of Life Insurance Ownership in the United States from 2011 to 2023. Statista. Available online: https://www.statista.com/statistics/455614/life-insurance-ownership-usa/ (accessed on 11 January 2024).
2. Hartley, D.; Paulson, A.; Powers, K. What Explains the Decline in Life Insurance Ownership? 2017. Available online: https://www.chicagofed.org/publications/economic-perspectives/2017/8#ftn2 (accessed on 11 January 2024).
3. Copeland, C. Understanding the Finances of Rural vs. Urban Americans. EBRI Issue Brief. 2022. Available online: https://www.ebri.org/content/understanding-the-finances-of-rural-vs.-urban-americans (accessed on 11 January 2024).
4. Mishra, A.K.; Hisham El-Osta, H.S.; Morehart, M.J.; Johnson, J.D.; Hopkins, J.W. *Income, Wealth, and the Economic Well-Being of Farm Households*; Agricultural Economic Report No. 812; Farm Sector Performance and Well-Being Branch, Resource, Economics Division, Economic Research Service, US Department of Agriculture: Washington, DC, USA, 2002.
5. Jackson-Matsushima, L. Average US Life Insurance Ownership Statistics by Gender, Ethnicity & State. 2024. Insuranceopedia. Available online: https://www.insuranceopedia.com/life-insurance/life-insurance-statistics (accessed on 8 July 2024).
6. Colt, J.S.; Stallones, L.; Cameron, L.L.; Dosemeci, M.; Zham, S.H. Proportionate mortality among US migrant and seasonal farmworkers in twenty-four states. *Am. J. Ind. Med.* **2001**, *40*, 604–611. [CrossRef] [PubMed]
7. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef] [PubMed]
8. Pisner, D.A.; Schnyer, D.M. Chapter 6—Support vector machine. In *Machine Learning*; Mechelli, A., Vieira, S., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 101–121.
9. Rudin, C.; Daubechies, I.; Schapire, R.F. The dynamics of AdaBoost: Cyclic behavior and convergence of margins. *J. Mach. Learn. Res.* **2004**, *5*, 1557–1595.
10. Anderson, D.R.; Nevin, J.R. Determinants of young marrieds' life insurance purchasing behavior: An empirical investigation. *J. Risk Insur.* **1975**, *42*, 375–388. [CrossRef]

11. Babbel, D.F. The price elasticity of demand for whole life insurance. *J. Financ.* **1985**, *40*, 225–239. [CrossRef]
12. Berekson, L.L. Birth order, anxiety, affiliation and the purchase of life insurance. *J. Risk Insur.* **1972**, *39*, 93–108. [CrossRef]
13. Browne, M.J.; Kim, K. An international analysis of life insurance demand. *J. Risk Insur.* **1993**, *60*, 616–634. [CrossRef]
14. Burnett, J.J.; Palmer, B.A. Examining life insurance ownership through demographic and psychographic characteristics. *J. Risk Insur.* **1984**, *51*, 453–467. [CrossRef]
15. Chen, R.; Wong, K.; Lee, H. Age, Period, and Cohort Effects on Life Insurance Purchases in the US. *J. Risk Insur.* **2001**, *68*, 303–327. [CrossRef]
16. Duker, J.M. Expenditures for life insurance among working-wife families. *J. Risk Insur.* **1969**, *36*, 525–533. [CrossRef]
17. Ferber, R.; Lee, L.C. Acquisition and accumulation of life insurance in early married life. *J. Risk Insur.* **1980**, *47*, 713–734. [CrossRef]
18. Fitzgerald, J. The effects of social security on life insurance demand by married couples. *J. Risk Insur.* **1987**, *54*, 86–99. [CrossRef]
19. Gandolf, A.S.; Miners, L. Gender-based differences in life insurance ownership. *J. Risk Insur.* **1996**, *63*, 683–693. [CrossRef]
20. Gutter, M.S.; Hatcher, C.B. Racial Differences in the Demand for Life Insurance. *J. Risk Insur.* **2008**, *75*, 677–689. [CrossRef]
21. Hammond, J.D.; Houston, D.B.; Melander, E.R. Determinants of household life insurance premium expenditure: An empirical investigation. *J. Risk Insur.* **1967**, *34*, 397–408. [CrossRef]
22. Hau, A. Liquidity, estate liquidation, charitable motives, and life insurance demand by retired singles. *J. Risk Insur.* **2000**, *67*, 123–141. [CrossRef]
23. Heo, W. *The Demand for Life Insurance: Dynamic Ecological Systemic Theory Using Machine Learning Techniques*; Springer: Berlin/Heidelberg, Germany, 2020.
24. Lewis, F.D. Dependents and the demand for life insurance. *Am. Econ. Rev.* **1989**, *79*, 452–467.
25. Liebenberg, A.P.; Carson, J.M.; Dumm, R.E. A dynamic analysis of the demand for life insurance. *J. Risk Insur.* **2012**, *79*, 619–644. [CrossRef]
26. Mantis, G.; Farmer, R.N. Demand for life insurance. *J. Risk Insur.* **1968**, *35*, 247–256. [CrossRef]
27. Mulholland, B.; Finke, M.; Huston, S. Understanding the shift in demand for cash value life insurance. *Risk Manag. Insur. Rev.* **2016**, *19*, 7–36. [CrossRef]
28. Showers, V.E.; Shotick, J.A. The effects of household characteristics on demand for insurance: A tobit analysis. *J. Risk Insur.* **1994**, *61*, 492–502. [CrossRef]
29. Williams, C.A. Higher interest rates, longer lifetimes, and the demand for life annuities. *J. Risk Insur.* **1986**, *53*, 164–171. [CrossRef]
30. Zietz, E.M. An examination of the demand for life insurance. *Risk Manag. Insur. Rev.* **2003**, *6*, 159–191. [CrossRef]
31. Giri, M.; Chatterjee, D. Factors affecting changes in insured status of rural and urban households: A study over two time periods in India. *IIMB Manag. Rev.* **2021**, *33*, 360–371. [CrossRef]
32. Deloitte. COVID-19 Impact to Life Insurance and Annuity Companies: Considerations from an Actuarial and Finance Perspective. Perspectives. 2022. Available online: https://www2.deloitte.com/us/en/pages/financial-services/articles/covid-19-impact-to-life-insurance-and-annuity-companies.html (accessed on 11 January 2024).
33. Gambhir, N.; Shoenthal, R. Life Insurance Statistics in 2022. Policygenius. 2022. Available online: https://www.policygenius.com/life-insurance/life-insurance-statistics/ (accessed on 11 January 2024).
34. Cohn, C.; Hussain, N.Z. Life Insurers Adapt Pandemic Risk Models after Claims Jump. Reuters. Available online: https://www.reuters.com/business/life-insurers-adapt-pandemic-risk-models-after-claims-jump-2022-01-13/ (accessed on 13 January 2022).
35. Shirsath, P.; Vyas, S.; Aggarwal, P.; Rao, K.N. Designing weather index insurance of crops for the increased satisfaction of farmers, industry and the government. *Clim. Risk Manag.* **2019**, *25*, 100189. [CrossRef]
36. Budhathoki, N.K.; Lassa, J.A.; Pun, S.; Zander, K.K. Farmers' interest and willingness-to-pay for index-based crop insurance in the lowlands of Nepal. *Land Use Policy* **2019**, *85*, 1–10. [CrossRef]
37. Mare, C.; Dragos, S.L.; Dragotâ, I.-M.; Dragos, C.M. Insurance literacy and spatial diffusion in the life insurance market: A subnational approach in Romania. *East. Eur. Econ.* **2019**, *57*, 375–396. [CrossRef]
38. Saraf, D.; Baser, N. Influence of fear on purchase of health insurance. *J. Financ. Serv. Mark.* **2024**, *29*, 354–363. [CrossRef]
39. Bigelow, D.; Borchers, A.; Hubbs, T. *US Farmland Ownership, Tenure, and Transfer*; Economic Information Bulletin Number 161; United States Department of Agriculture: Washington, DC, USA, 2016.
40. Bateman, H.; Gerrans, P.; Thorp, S.; Zeng, Y. *Life Insurance: Decision States, Financial Literacy, and the Role of Personal Values*; Netspar Academic Series; Netspar: Tilburg, The Netherlands, 2020.
41. Outreville, J.F. The relationship between insurance and economic development: 85 empirical papers for a review of the literature. *Risk Manag. Insur. Rev.* **2013**, *16*, 71–122. [CrossRef]
42. Weber, E.; Hopwood, C.J.; Nissen, A.T.; Bleidorn, W. Disentangling self-concept clarity and self-esteem in young adults. *J. Personal. Soc. Psychol.* **2023**, *125*, 1420–1441. [CrossRef]
43. Heo, W.; Grable, J.E.; Chatterjee, S. Life insurance consumption as a function of wealth change. *Financ. Serv. Rev.* **2013**, *22*, 389.
44. Shefrin, H.M.; Thaler, R.H. The behavioral life-cycle hypothesis. *Econ. Inq.* **1988**, *26*, 609–643. [CrossRef]
45. Rabbani, A.G. Cash value life insurance ownership among young adults: The role of self-discipline and risk tolerance. *J. Behav. Exp. Financ.* **2020**, *27*, 100385. [CrossRef]
46. Tran, N.L.T.; Wassmer, R.W.; Lascher, E.L. The health insurance and life satisfaction connection. *J. Happiness Stud.* **2017**, *18*, 409–426. [CrossRef]

47. Hwang, I.D. Behavioral aspects of household portfolio choice: Effect of loss aversion on life insurance uptake and savings. *Int. J. Econ. Financ.* **2024**, *89*, 1029–1053. [CrossRef]

48. Nam, Y.; Hanna, S.D. The effects of risk aversion on life insurance ownership of single-parent households. *Appl. Econ. Lett.* **2019**, *26*, 1285–1288. [CrossRef]

49. Guillemette, M.A.; Hussein, M.M.; Phillips, G.M.; Martin, T.K. Minority household size and the life insurance purchase decision. *Financ. Serv. Rev.* **2015**, *24*, 37–50. [CrossRef]

50. Song, I.J.; Park, H.; Park, N.; Heo, W. The effect of experiencing a death on life insurance ownership. *J. Behav. Exp. Financ.* **2019**, *22*, 170–176. [CrossRef]

51. Nkouaga, F. Comparative effects of self-evaluated and test-based financial literacy on choosing life insurance policies in a multi-racial context. *Front. Manag. Bus.* **2024**, *5*, 352–371.

52. Prasad, S.; Kiran, R.; Sharma, R.K. Influence of financial literacy on retail investors' decisions in relation to return, risk and market analysis. *Int. J. Financ. Econ.* **2021**, *26*, 2548–2559. [CrossRef]

53. Rey-Ares, L.; Fernández-López, S.; Castro-González, S. Life insurance consumption across generations: The roles of financial knowledge, planning horizon, and self-control. *Int. J. Financ. Econ.* **2023**, 1–21. [CrossRef]

54. Farrell, L.; Fry, T.R.L.; Risse, L. The significance of financial self-efficacy in explaining women's personal finance behavior. *J. Econ. Psychol.* **2016**, *54*, 85–99. [CrossRef]

55. Cucinelli, D.; Lippi, A.; Soana, M.G. Per aspera ad astra: The big challenge of consumers insurance literacy. *Int. J. Consum. Stud.* **2021**, *45*, 1357–1372. [CrossRef]

56. Dragos, S.L.; Dragos, C.M.; Muresan, G.M. From intention to decision in purchasing life insurance and private pensions: Different effects of knowledge and behavioural factors. *J. Behav. Exp. Econ.* **2020**, *87*, 101555. [CrossRef]

57. Roa, M.J.; Di Giannatale, S.; Barboza, J.; Arbelaez, J.G. Inclusive health and life insurance adoption: An empirical study in Guatemala. *Rev. Dev. Econ.* **2021**, *25*, 1053–1077. [CrossRef]

58. Weedige, S.S.; Ouyang, H.; Gao, Y.; Liu, Y. Decision making in personal insurance: Impact of insurance literacy. *Sustainability* **2019**, *11*, 6795. [CrossRef]

59. Allgood, S.; Walstad, W.B. The effects of perceived and actual financial literacy on financial behaviors. *Econ. Inq.* **2016**, *54*, 675–697. [CrossRef]

60. Lin, C.; Hsiao, Y.-J.; Yeh, C.-Y. Financial literacy, financial advisors, and information sources on demand for life insurance. *Pac. Basin Financ. J.* **2017**, *43*, 218–237. [CrossRef]

61. Luciano, E.; Rossi, M.; Sansone, D. *Financial Inclusion and Life Insurance Demand; Evidence from Italian Households*; Working Paper No. 156/16; Center for Research on Pensions and Welfare Policies: Turin, Italy, 2015.

62. Wang, H.; Zhang, D.; Guariglia, A.; Fan, G.-Z. Growing out of the growing pain: Financial literacy and life insurance demand in China. *Pac. Basin Financ. J.* **2021**, *66*, 101459. [CrossRef]

63. Zakaria, Z.; Azmi, N.M.; Hassan, N.F.H.N.; Salleh, W.A.; Tajuddin, M.T.H.M.; Sallem, N.R.M.; Noor, J.M.M. The intention to purchase life insurance: A case study of staff in public universities. *Procedia Econ. Financ.* **2016**, *37*, 358–365. [CrossRef]

64. Mahdzan, N.S.; Victorian, S.M.P. The determinants of life insurance demand: A focus on saving motives and financial literacy. *Asian Soc. Sci.* **2013**, *9*, 274–284. [CrossRef]

65. Aizcorbe, A.M.; Kennickell, A.B.; Moore, K.B. Recent Changes in US Family Finances: Evidence From the 1998 and 2001 Survey of Consumer Finances. *Fed. Reserve Bull.* **2003**, *89*, 1–32.

66. Zhao, K. The impact of the correlation between health expenditure and survival probability on the demand for insurance. *Eur. Econ. Rev.* **2015**, *75*, 98–111. [CrossRef]

67. Mathew, B.; Sivaraman, S. Cointegration and causality between macroeconomic variables and life insurance demand in India. *Int. J. Emerg. Mark.* **2017**, *12*, 727–741. [CrossRef]

68. Loughrey, J.; Vidyaratne, H. The empirical demand for farm insurance in Ireland: A quantile regression approach. *Agric. Financ. Rev.* **2023**, *83*, 572–596. [CrossRef]

69. Quaye, F. Effects of multiple risks on farm income and willingness to pay for agricultural insurance: A case study of the Greater Accra Region in Ghana. *Int. J. Econ. Financ.* **2016**, *8*, 1–12. [CrossRef]

70. San Miguel, M.; Johnson, J.H.; Kertesz, J.; Kaski, K.; Díaz-Guilera, A.; MacKay, R.S.; Loreto, V.; Erdi, P.; Helbing, D. Challenges in complex systems science. *Eur. Phys. J. Spec. Top.* **2012**, *214*, 245–271. [CrossRef]

71. Ladyman, J.; Lambert, J.; Wiesner, K. What is a complex system? *Eur. J. Philos. Sci.* **2012**, *3*, 33–67. [CrossRef]

72. Beck, T.; Webb, I. Economic, demographic, and institutional determinants of life insurance consumption across countries. *World Bank Econ. Rev.* **2003**, *17*, 51–88. [CrossRef]

73. Berkovitch, E.; Venezia, I. Term vs. whole life insurance—A note. *J. Account. Audit. Financ.* **1992**, *7*, 214–249. [CrossRef]

74. Black, K.; Skipper, H.D. *Life & Health Insurance*, 13th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2000.

75. DeLamater, J.D.; Myers, D.J. *Social Psychology*, 7th ed.; Wadsworth: Belmont, CA, USA, 2011.

76. Fang, H.; Kung, E. Why do life insurance policyholders lapse? The roles of income, health, and bequest motive shocks. *J. Risk Insur.* **2020**, *88*, 937–970. [CrossRef]

77. Hussels, S.; Ward, D.R.; Zurbruegg, R. Stimulating the demand for insurance. *Risk Manag. Insur. Rev.* **2005**, *8*, 257–278. [CrossRef]

78. Josephs, R.A.; Markus, H.R.; Tafarodi, R.W. Gender and self-esteem. *J. Personal. Soc. Psychol.* **1992**, *63*, 391–402. [CrossRef] [PubMed]

79. Kung, K.; Yang, S. Optimal consumption and investment problem incorporating housing and life insurance decisions: The continuous time case. *J. Risk Insur.* **2019**, *87*, 143–171. [CrossRef]

80. Li, D.; Moshirian, F.; Nguyen, P.; Wee, T. The demand for life insurance in OECD countries. *J. Risk Insur.* **2007**, *74*, 637–652. [CrossRef]

81. Sobehart, J.R. Rumour has it: Modeling credibility, reputation and franchise risk. *J. Risk Manag. Financ. Inst.* **2014**, *7*, 161–173. [CrossRef]

82. Ward, D.; Zurbruegg, R. Law, politics and life insurance consumption in Asia. *Geneva Pap. Risk Insur. Issues Pract.* **2002**, *27*, 395–412. [CrossRef]

83. Wray, L.D.; Stone, E.R. The role of self-esteem and anxiety in decision making for self versus others in relationships. *J. Behav. Decis. Mak.* **2005**, *18*, 125–144. [CrossRef]

84. Yaari, M.E. Uncertain lifetime, life insurance, and the theory of the consumer. *Rev. Econ. Stud.* **1965**, *32*, 137–150. [CrossRef]

85. Jaspersen, J.G.; Ragin, M.A.; Sydnor, J.R. Predicting insurance demand from risk attitudes. *J. Risk Insur.* **2021**, *89*, 63–96. [CrossRef]

86. Jaspersen, J.G.; Ragin, M.A.; Sydnor, J.R. Insurance demand experiments: Comparing crowdworking to the lab. *J. Risk Insur.* **2022**, *89*, 1077–1107. [CrossRef]

87. Biagini, F.; Huber, T.; Jarspersen, J.G.; Mazzon, A. Estimating extreme cancellation rates in life insurance. *J. Risk Insur.* **2021**, *88*, 971–1000. [CrossRef]

88. Fan, J.; Yao, Q. *Nonlinear Time Series: Nonparametric and Parametric Methods*; Springer: Berlin/Heidelberg, Germany, 2003.

89. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 1994.

90. Awad, M.; Khanna, R. Support Vector Regression. In *Efficient Learning Machines*; Apress: Berkeley, CA, USA, 2015.

91. Chang, Y.; Chang, K.; Wu, G. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Appl. Softw. Comput.* **2018**, *73*, 914–920. [CrossRef]

92. Friedman, J.H. Greedy function approximation: A Gradient Boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

93. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009.

94. Chen, W.; Du, Y. Using neural networks and data mining techniques for the financial distress prediction model. *Expert Syst. Appl.* **2009**, *36*, 4075–4086. [CrossRef]

95. Baum, E.B. Neural nets for economics. In Proceedings of the Economy as an Evolving Complex System, Proceedings of the Evolutionary Paths of the Global Economy Workshop, Sante Fe, NM, USA, 8–18 September 1987; Anderson, P., Arrow, K., Pindes, D., Eds.; Addison-Wesley: Reading, MA, USA, 1987; pp. 33–48.

96. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013.

97. Heo, W.; Grable, J.E. Demand for life insurance: A consumer perspective. *J. Financ. Serv. Prof.* **2017**, *71*, 19–22.

98. Kudyba, S.; Kwatinetz, M. Introduction to the big data era. In *Big Data, Mining, and Analytics*; Kudyba, S., Ed.; CRC Press: Boca Raton, FL, USA, 2014; Taylor & Francis: New York, NY, USA, 2014; pp. 1–15.

99. Thompson, W. Data mining methods and the rise of big data. In *Big Data, Mining, and Analytics*; Kudyba, S., Ed.; CRC Press: Boca Raton, FL, USA, 2014; Taylor & Francis: New York, NY, USA, 2014; pp. 71–101.

100. Sarker, I.H. Machine learning: Algorithms, real-World applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef]

101. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshard, H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* **2018**, *4*, e00938. [CrossRef] [PubMed]

102. Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Suthaharan, S., Ed.; Springer US: New York, NY, USA, 2016; pp. 207–235.

103. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [CrossRef]

104. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. *arXiv* **2016**, arXiv:1603.02754.

105. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363.

106. Gavrilov, A.D.; Jordache, A.; Vasdani, M.; Deng, J. Preventing model overfitting and underfitting in Convolutional Neural Networks. *Int. J. Softw. Sci. Comput. Intell. IJSSCI* **2018**, *10*, 19–28. [CrossRef]

107. Zhang, P. A study on the management of big data technology in financial decision making of enterprise cloud accounting. *Appl. Math. Nonlinear Sci.* **2023**, *9*, 1–17. [CrossRef]

108. Fan, L. The use of financial advice: Consumers' financial advice-seeking. In *De Gruyter Handbook of Personal Finance*; Grable, J.E., Chatterjee, S., Eds.; De Gruyter: Berlin, Germany, 2022; pp. 551–568.

109. Heo, W.; Lee, J.M.; Park, N. Who demands which type of life insurance? *Financ. Serv. Rev.* **2021**, *29*, 101–109. [CrossRef]

110. Brighetti, G.; Lucarelli, C.; Marinelli, N. Do emotions affect insurance demand? *Rev. Behav. Financ.* **2014**, *6*, 136–154. [CrossRef]

111. Becot, F.A.; Inwood, S.M. Examining access to health insurance and health care along the life course to shed light on interactions between farm households' social needs, social policy and the farm business. *Sociol. Rutalis* **2022**, *62*, 485–508. [CrossRef]