Educational Administration: Theory and Practice
2023, 29(4), 4195 -4209
ISSN: 2148-2403
https://kuey.net/

Research Article

A Neural Network-Based Study on the Significance of Risk Aversion Bias in Life Insurance Selection.

Mr. Hunasing Engti[1], Prof. Amalesh Bhowal[2]

[1] Commerce Department, Assam University Diphu Campus, Diphu, Karbi Anglong, Assam, Email id – hunasing.engti@aus.ac.in
ORCID ID - https://orcid.org/0000-0003-4672-9641
[2] Commerce Department, Assam University Diphu Campus, Diphu, Karbi Anglong, Assam, Email id - profabhowal@gmail.com
ORCID ID – https://orcid.org/0000-0002-1451-6425

Citation: Mr. Hunasing Engti et al. (2023), A Neural Network-Based Study on the Significance of Risk Aversion Bias in Life Insurance Selection., *Educational Administration: Theory and Practice*,29(4) 4195 -4209
Doi: 10.53555/kuey.v29i4.9094

ARTICLE INFO

ABSTRACT

Risk aversion bias, a significant behavioural factor, often causes individuals to make financial choices that diverge from optimal decision-making strategies. The present study is to examine the impact of risk aversion bias on insurance selection decisions by leveraging the predictive power of neural networks. The research utilizes a dataset that includes demographics, latent variables of risk aversion bias, and preference-based variables (binary). It employs a neural network model to assess the extent to which risk aversion bias influences insurance product selection. The study also used bootstrapping analysis to validate the application of the neural network model in small-sample research. The methodology encompasses data preprocessing, model training, and validation to ensure the robustness of the results. Results show evidence that risk aversion bias acts as a major influencing factor in shaping the insurance selection decision, with notable variations across demographic groups. Moreover, when compared to traditional methods, the neural network approach shows superior performance in statistical models and has the capacity to effectively capture the complex, non-linear relationships within behavioural data. Findings from this analysis add meaningful value to the areas of behavioural finance in general and insurance decision-making in particular, providing actionable guidance for insurance providers to tailor their products and strategies in line with consumer biases. The study also highlights neural network potential to improve understanding of the behaviour of life insurance consumer and proposes future research to broaden the model's applicability.

Keywords: Risk aversion bias; insurance selection; neural network; bootstrapping

1. Introduction

1.1 Background

Risk aversion is a common behavioural bias in finance that prevalently influences investment decisions. It affects individual investors by prompting them to avoid risks, even in scenarios where the potential returns are significant. However, in the realm of life insurance, this bias remains underexplored, as highlighted by the literature review undertaken for this study.

1.2 Problem Statement

With the rise of **AI-driven research** across various domains, this study seeks to bridge that gap. By employing **neural network analysis**, it aims to uncover the significance and degree of association between risk aversion bias and the selection of the most suitable life insurance products. The study also applied bootstrapping to assess the validity of deploying neural network analysis for small sample sizes. This approach offers a novel perspective on understanding how risk aversion bias affects decision-making in the life insurance sector.

1.3 Significance of the Study

A deeper understanding of risk aversion bias allows individuals to select insurance products that effectively mitigate the negative impacts of this bias. This empowers investors to make informed, strategic decisions

when choosing life insurance products that meet their financial objectives and personal needs. Furthermore, the study provides valuable insights and practical resources to help readers navigate the complexities of life insurance investments. By fostering a more nuanced approach to decision-making, it supports the pursuit of financial security and peace of mind for both individuals and their loved ones.

2. Literature Review

2.1 Literature survey strategy

Table 1: Search strings & Database used

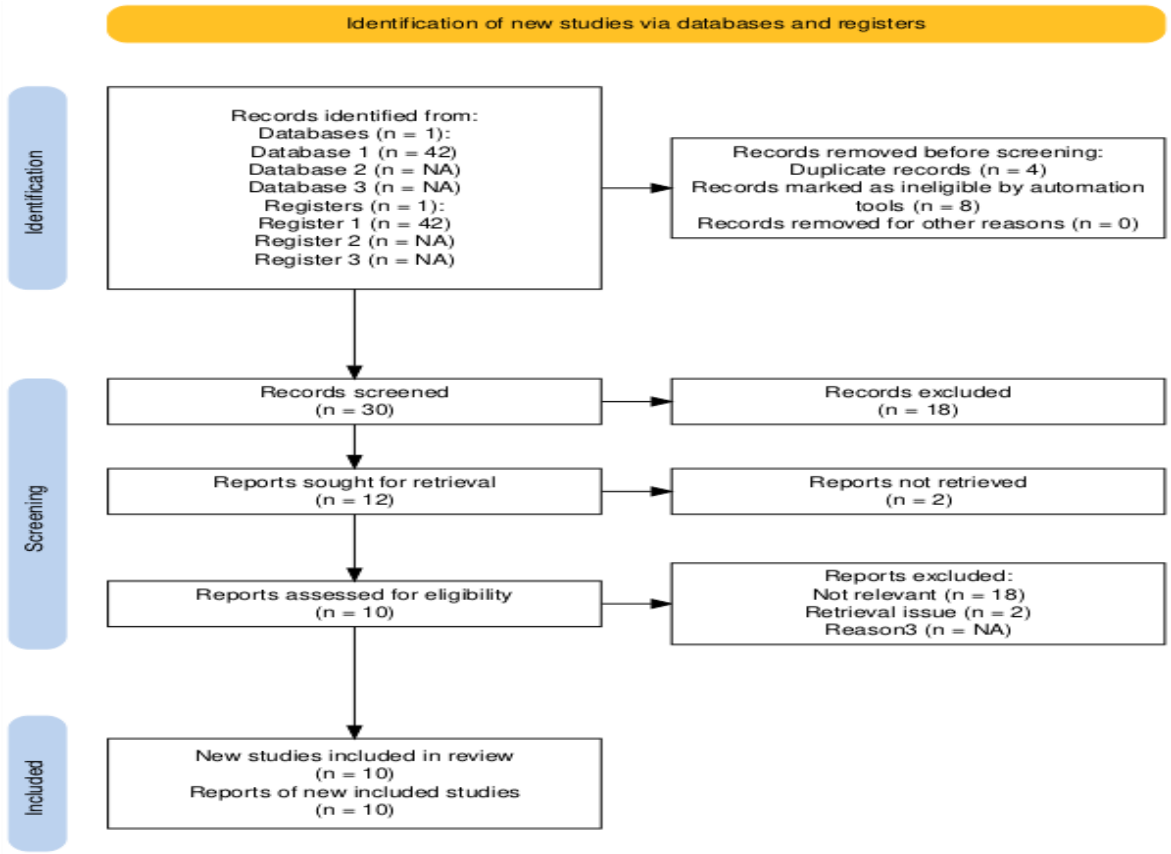
Search strings (English)	Dimensions.ai (16/11/24)	
	Limited to	Results
"risk aversion bias" AND "investment decision"	Banking & finance; All OA; Article & monograph	19
"risk aversion bias" AND "life insurance"	"	**02**
"neural network" AND "life insurance decision"	"	03
"artificial intelligence" AND "life insurance decision"	"	02
"artificial intelligence" AND "insurance decision" NOT "investment decision"	"	15
"artificial intelligence" AND "insurance selection"	"	01

Source: compiled from literature survey (Dimension.ai)

The table titled "Search strings & Database used" summarized the search strategy and results from a literature review conducted using the Dimensions.ai database on November 16, 2024. The search employed specific keywords and phrases combined with Boolean operators (e.g., AND, NOT) to identify relevant academic works, focusing on topics such as risk aversion bias, life insurance, investment decisions, neural networks, and artificial intelligence. The search was further refined by applying limitations to specific subject areas, including Banking & Finance, open-access materials, and particular publication types like articles and monographs, ensuring that the results were relevant and targeted.

2.2 Documents screening process

Figure I: SLR flow chart

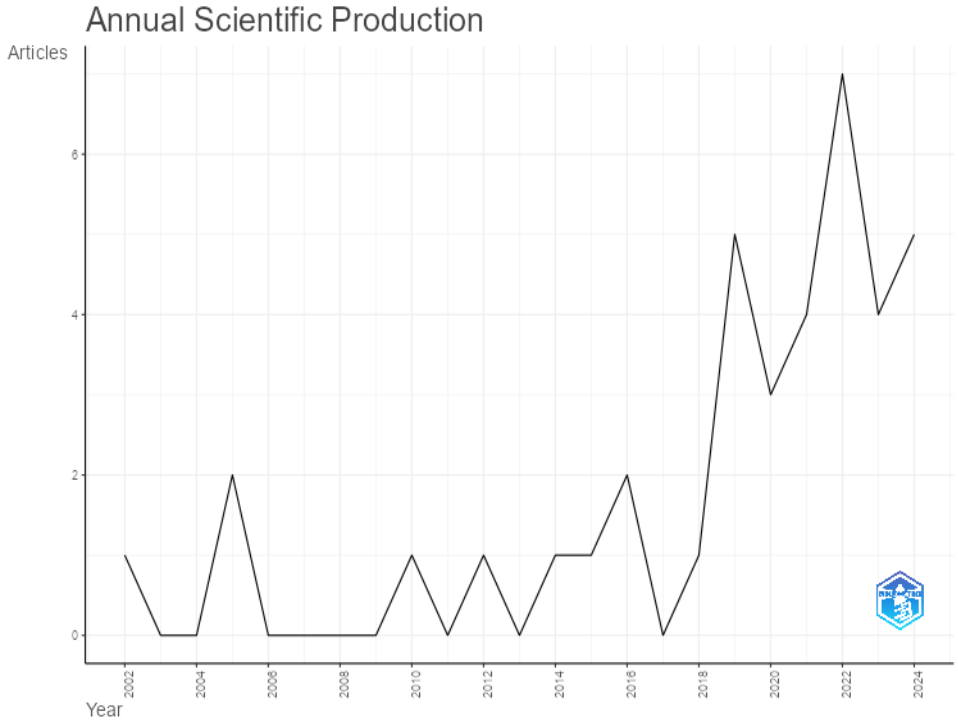


Source: generated from shiny app (https://doi.org/10.1002/cl2.1230)

The diagram above depicted from the PRISMA flowchart that outlines the systematic process of identifying, screening, and selecting studies for inclusion in a review. The process began with 42 records identified from 'Dimension.ai database. Following the removal of duplicates and irrelevant studies, a thorough screening and eligibility assessment were conducted. Ultimately, ten studies were incorporated in the review, highlighting the rigorous approach used to ensure the inclusion of only relevant and accessible records

2.3 Bibliometrics analysis

Figure II: Annual scientific production



Source: generated from biblioshiny analysis (R Studio)

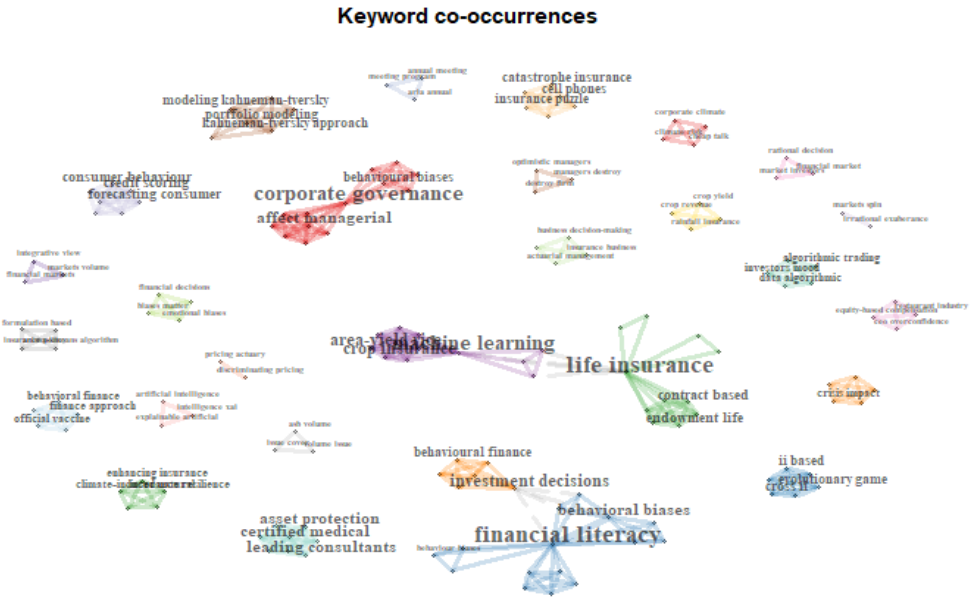
This chart illustrates the progression of research activity over time, highlighting a significant upward trend in the number of published articles, particularly from 2016 onward. Between 2002 and 2015, scientific output was relatively low and intermittent, suggesting that the topic was either in its early stages of exploration or of limited interest during that period. However, a noticeable surge in publications occurred after 2019, with peaks in 2020 and 2022, indicating a heightened focus on the field. This increase may be attributed to advancements in data availability, increased funding, or global phenomena such as the COVID-19 pandemic, which may have amplified interest in related topics like risk, insurance, and financial decision-making. The steady rise in articles underscores the growing importance of the research area and reflects its transition from a niche interest to a more mature and widely recognized field of study.

Figure III: Word cloud

Source: generated from biblioshiny analysis (R Studio)

The word cloud captures the dominant themes and trends in the research dataset, with "insurance," "financial," and "investment" standing out as central topics, while also indicating a strong emphasis on behavioural finance and risk management. Key biases, such as risk aversion, overconfidence, and herding behaviour, appear to play a central role in shaping investor behaviour, affecting decisions in areas like investment, financial planning, and insurance adoption. It provides a quick overview of the research focus and the key concepts explored in the dataset.

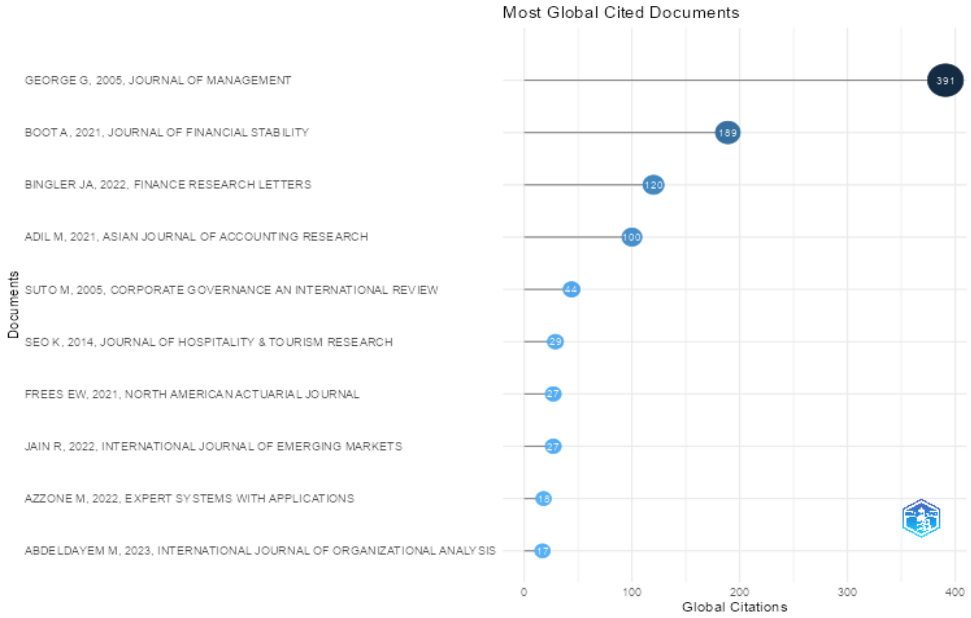
Figure IV: Bigram co-occurrence network



Source: generated from biblioshiny (R Studio)

The bigram co-occurrence network highlights the key themes and relationships within the research dataset. Larger nodes, such as "life insurance," "financial literacy," and "corporate governance," indicate frequently occurring keywords, emphasizing their importance in the field. The edges between nodes represent co-occurrence relationships, with thicker lines signifying stronger associations between keywords. The clusters, differentiated by colours, represent distinct thematic groups. For instance, the green cluster centres on 'life insurance' and related concepts like "contract-based" and 'endowment life,' while the blue cluster focuses on "financial literacy" and "investment decisions." The purple cluster emphasizes "machine learning" and its applications, whereas the red cluster revolves around "corporate governance" and managerial issues. Keywords like "behavioural biases" and "investment decisions" appear to connect multiple clusters, indicating their cross-disciplinary relevance and their role in bridging various research domains. This visualization effectively maps the intellectual structure of the dataset and identifies key areas of focus.

Figure V: Top ten most globally cited documents



Source: generated from biblioshiny (R Studio)

This diagram is a bubble chart that visualizes the influence of various studies or publications based on citation metric represented on the x-axis, with each study listed along the y-axis. The size of each bubble indicates an additional dimension, such as the relative importance or influence of the study. Larger bubbles and those further along the x-axis suggest greater impact or citation frequency. The chart highlights key contributions in the dataset, with the most influential study appearing in the top-right corner, distinguished by a large bubble and high x-axis value, while smaller bubbles or those closer to the origin indicate less prominent works.

2.4 Risk aversion and AI application related

Aleksandrina et al. (2023) highlighted the potential of artificial intelligence in identifying and managing risk, which is crucial for addressing uncertainties that often lead to risk aversion among both consumers and businesses. Additionally, the authors stated that the rapid development of AI in financial services is transforming how risks are assessed and managed; potentially reducing the biases associated with traditional risk evaluation methods. Through artificial intelligence, insurers and financial institutions can provide customized products that match individual investor risk profiles.

Edi K and Itzhak Z (2015) stated the significant influence of risk aversion bias. They said, "Under fair life insurance terms, more risk-averse persons tend to opt for more life insurance product than their less risk-averse counterparts, particularly in a two-period setup; the comparability of risk aversion among decision-makers requires identical reference sets and ordinal preferences." This points out the fact that risk aversion bias can vary significantly based on the underlying preferences and reference points of individuals, suggesting that a one-size-fits-all approach to understanding risk aversion may be inadequate. Such findings underscore the complex nature of risk aversion and its effect on life insurance choices, suggesting that individual preferences and situations significantly influence the selection of life insurance.

Saeid Z et al. (2019) integrated Prospect Theory to explore psychological factors in investment decisions, emphasizing that risk minimization can be achieved without significantly sacrificing returns. By moving beyond traditional models and incorporating behavioural approaches, the study provided a deeper understanding of investor psychology, particularly in situations where biases such as risk aversion heavily influence decision-making. The authors highlighted the pivotal role of risk aversion bias in shaping investment behaviour, underscoring the importance of applying behavioural finance principles to better understand and predict investor actions.

Alex G & Paolo G (2020) undertook a study entitled "Why to Buy Insurance? An Explainable Artificial Intelligence Approach" and came to the conclusion that the integration of explainable machine learning models, such as the blending of XGBoost and Shapley values, substantially improves the comprehension of customer behaviour within the insurance industry. The authors proposed that this methodology could be extended to other areas within the insurance industry and financial technology applications, paving the way for improved customer profiling and service delivery.

3. Methods and Materials

3.1 Data acquisition method

Primary sources for the study were collected online using a questionnaire specifically designed for the study. The questionnaire included ten statements related to risk aversion bias, together with a few demographic variables like gender, marital status, years of work experience, number of dependents, and annual income and two dependent variables in the binary form. It was distributed to seventy employees engaging different departments in Assam, India. Fifty complete responses were collected and thoroughly analyzed to address the stated research problem of the study.

3.2 Analysis tools

To ensure the reliability and internal consistency of the variables used in this study, Cronbach's Alpha was employed as a statistical measure. This approach provided a robust evaluation of how well the items within the questionnaire were correlated and consistent in measuring the underlying constructs. Furthermore, a Neural Network Analysis model using a multilayer perceptron was employed to explore the influence of the latent variable of risk aversion bias, particularly in the context of life insurance investors. The model was designed with a partition ratio of 7:3, a single hidden layer, a hyperbolic tangent activation function in the hidden layer, and a softmax activation function in the output layer to capture and interpret the non-linear relationships and complex patterns inherent in the data. This methodological approach enabled a comprehensive exploration of the factors contributing to risk aversion bias, providing valuable insights into the behavioural tendencies of life insurance investors. Furthermore, bootstrapping analysis was employed to assess the model's validity, ensuring its applicability even with a smaller sample size.

4. Data Analysis and Results

Table 2: Reliability test

Case Processing Summary

		N	%
Case	Valid	50	100.0
	Excluded	0	.0
	Total	50	100.0

Source: compilation data (survey)

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.751	.760	10

The table presents the reliability statistics for the scale used in the study, with a **Cronbach's Alpha of 0.751** and a slightly higher **standardized alpha of 0.760**, indicating acceptable internal consistency. These values confirm that the 10 items (likely representing the risk aversion bias variables) reliably measure the same underlying construct. This reliability is crucial for ensuring that the scale provides consistent and trustworthy data, validating its use in predicting customer behaviour, such as willingness to buy or not to buy life insurance products.

4.2 Neural Network Analysis Model

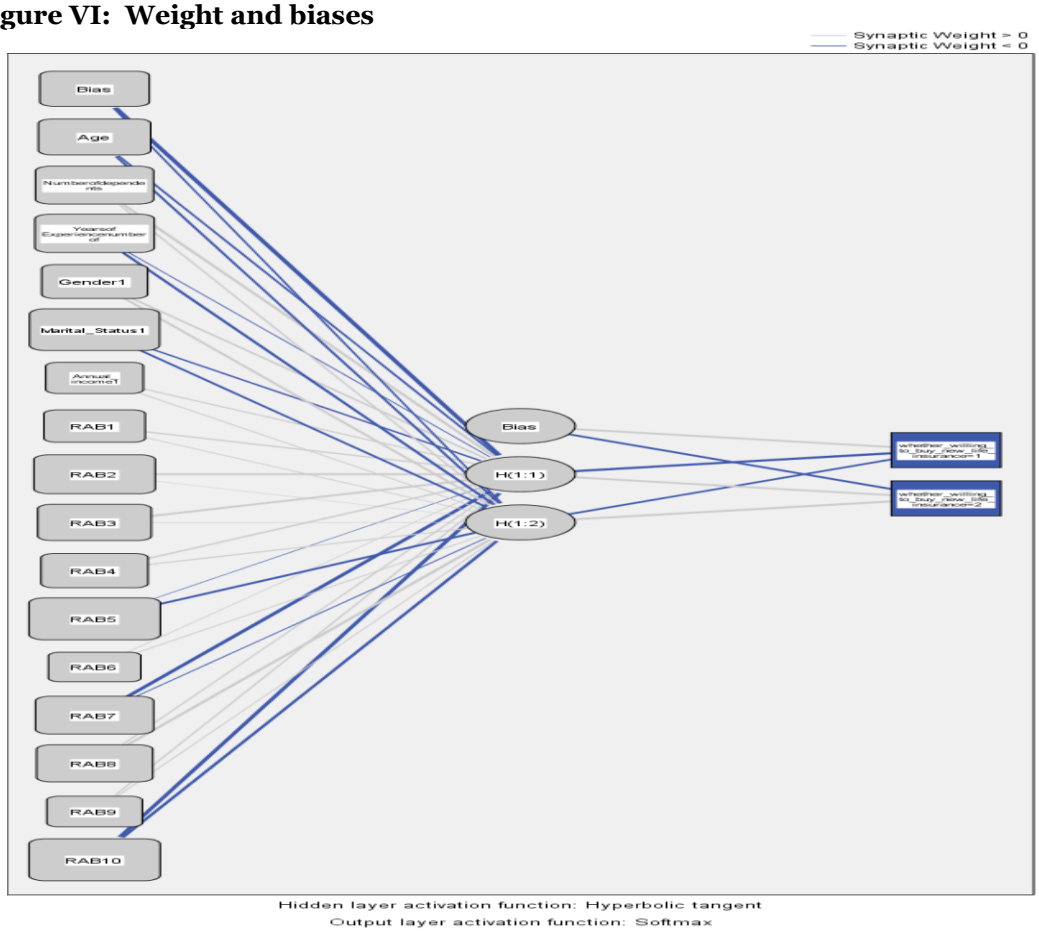
Table 3: Summary of Case Analysis

		N	Percent
Sample	Training	36	72.0%
	Testing	14	28.0%
Valid		50	100.0%
Excluded		0	
Total		50	

Source: generated from survey data

Table 3 above outlines the summary of the distribution of the dataset for training and testing the neural network model. It is seen that **72% of the data (36 cases)** is used for training the model, while **28% (14 cases)** is reserved for testing to gauge the model's performance. With **50 valid cases** in total and **no excluded cases**, the entire dataset is utilized for analysis. This split creates certainty that the application of neural networks is both developed and validated effectively, allowing for a robust assessment of its ability to generalize to unseen data.

Figure VI: Weight and biases



Source: generated neural network analysis (survey data)

Model workflow

- **Input layer:** The input features for the model include a few demographic variables like age, gender, and years of work experience, number of dependents, annual income and ten latent variables of Risk Aversion Bias (RAB1 to RAB10). Each input feature is connected to the hidden layer neurons, with associated synaptic weights.
- **Bias nodes:** Present in both the input and hidden layers, bias nodes help the model learn patterns in data by shifting activation functions.
- **Hidden layer:**
  - To optimize the model's performance with a relatively small sample size, a single neuron has been chosen for the hidden layer and connected with two neurons (H(1:1) and H(1:2)) that receive weighted inputs from the input layer and produce an output that is passed to the output layer.
  - Applies a **Hyperbolic Tangent (tanh)** activation function to process the weighted sum of inputs and biases.
  - Synaptic weights between the input and hidden layers are shown, with **blue lines for non-zero weights** and **grey for zero weights**.
- **Equations**: For Hidden Neuron H(1:1)

$$H_{(1:1)} = \tanh\left(\sum_{i=1}^{n}(w_i \cdot H_{(1:1)}^{xi} + b_{H(1:1)})\right)$$

For Hidden Neuron H(1:2)

$$H_{(1:2)} = \tanh\left(\sum_{i=1}^{n}(w_i \cdot H_{(1:2)}^{xi} + b_{H(1:2)})\right)$$

Where,

$H_{(1:1)}$: The output of the first neuron in the hidden layer.

$H_{(1:2)}$: The output of the second neuron in the hidden layer
$b_{H(1:1)}$: Bias term for $H_{(1:1)}$
$b_{H(1:2)}$: Bias term for $H_{(1:2)}$
$H_{(1:1)}^{xi}$: The i-th input contributing to $H_{(1:1)}$
$H_{(1:2)}^{xi}$: The i-th input contributing to $H_{(1:2)}$
W_i : Weight for the i-th input
$\sum_{i=1}^{n}$: Summation over all input features (x1,x2,...,xn)
tanh(): The hyperbolic tangent activation function, which introduces non-linearity.

Output layer:

- Includes two output neurons, each corresponding to specific classes or outputs; no=1; yes=2.
- Uses a **Softmax activation function** to convert the hidden layer's output into probability scores for classification.
- The diagram displays the weighted connections (synaptic weights) between layers. Weights determine the strength of influence each feature or neuron has on the subsequent layer.
- **Equations:**

For **Class 1** (not willing to buy new life insurance)

$$\text{output}_{\text{class 1}} = \text{Softmax}\{w_{H(1:1)}.H_{(1:1)}^{\text{Class 1}} + w_{H(1:2)}.H_{(1:2)}^{\text{Class 2}} + b_{\text{Class 1}}\}$$

For **Class 2** (willing to buy new life insurance)

$$\text{Output}_{\text{Class 2}} = \text{Softmax}\{w_{H(1:1)}.H_{(1:1)}^{\text{Class 2}} + w_{H(1:2)}.H_{(1:2)}^{\text{Class 2}} + b_{\text{Class 2}}\}$$

Where,

$\text{Output}_{\text{Class 1}}$ = The output corresponding to Class 1
$\text{Output}_{\text{Class 2}}$ = The output corresponding to Class 2
$w_{H(1:1)}$ = Weight associated with the first input feature
$w_{H(1:2)}$ = Weight associated with the second input
$H_{(1:1)}$ = The first feature or input for Class 1.
$H_{(1:2)}$ = The second feature or input for Class 2
$b_{(\text{Class 1})}$= Bias term for Class 1.
$b_{(\text{Class 2})}$= Bias term for Class 2
Softmax **:** Activation function applied to the input

Summary of performance of the activation function

Figure VI above represents a neural network designed to predict or classify a person's willingness to buy a new life insurance policy based on input features such as age, gender, marital status, and behavioural indicators (e.g., RAB1-RAB10). It includes an input layer, a single hidden neuron with a hyperbolic tangent (tanh) activation function, and an output layer with two nodes that represent classification results, employing the Softmax activation function to generate probabilities. The connections between layers represent weights, with grey lines indicating positive influences and blue lines indicating negative ones. Bias terms are included across both the hidden and output layers to improve flexibility and accuracy. The simplicity of the network, with only one hidden neuron, suggests it is optimized for small datasets or to prevent overfitting.

Table 4: Model accuracy

Training	Cross Entropy Error	13.646
	Percent Incorrect Predictions	16.7%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	00:00:00.008
Testing	Cross Entropy Error	.408
	Percent Incorrect Predictions	.0%

Error computations are based on the testing sampleSource: generated from survey data

his **Model Summary** table reveals the neural network's performance during both the training and testing phases. During training, the model had a **cross-entropy error of 13.646** and **16.7% incorrect predictions**, indicating some difficulty in fitting the data. However, the model stopped training after just one step with no error decrease, preventing overfitting. Despite these challenges during training, the model performed exceptionally well on the testing data, with a **cross-entropy error of 0.408** and **0.0%**

incorrect predictions, suggesting excellent generalization and accuracy in predicting willingness to purchase life insurance.

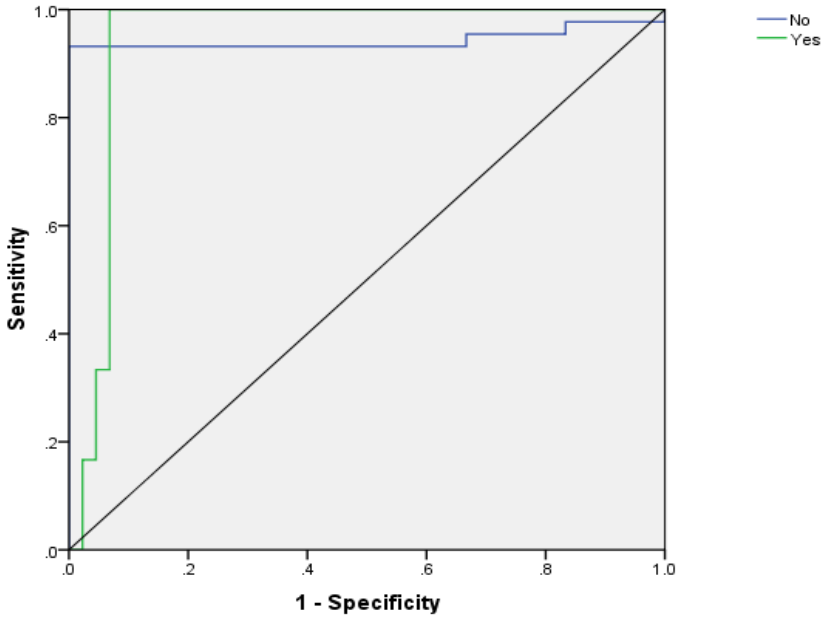
Table 5: Estimated Parameters

Predictor		Predicted		
		Hidden Layer 1	Output Layer	
		H(1:1)	[whether_willing_to_buy_new_life_insurance=1]	[whether_willing_to_buy_new_life_insurance=2]
Input Layer	(Bias)	1.711		
	Age	.908		
	YearsofExperiencenumberof	.463		
	Gender1	-.980		
	Marital_Status1	1.708		
	Annual_income1	.070		
	RAB1	.249		
	RAB2	-.532		
	RAB3	-.490		
	RAB4	-1.242		
	RAB5	.324		
	RAB6	.376		
	RAB7	1.303		
	RAB8	.111		
	RAB9	-.603		
	RAB10	1.820		
Hidden Layer 1	(Bias)		2.268	-1.673
	H(1:1)		2.744	-1.846

Source: generated from survey data

The table above highlights the relative contributions of predictors in determining the model's predictions. Variables with large positive weights (e.g., **RAB10**, **Marital_Status1**) strongly promote willingness to buy life insurance products, whereas those with significantly negative weights (e.g., **RAB4**, **Gender1**) suppress it. The hidden layer consolidates these influences, with neuron **H(1:1)** playing a central role in differentiating between the two outcomes. By analyzing these weights, one can better understand the behavioural and demographic factors driving decisions, presenting valuable data for personalized interventions or marketing approaches.

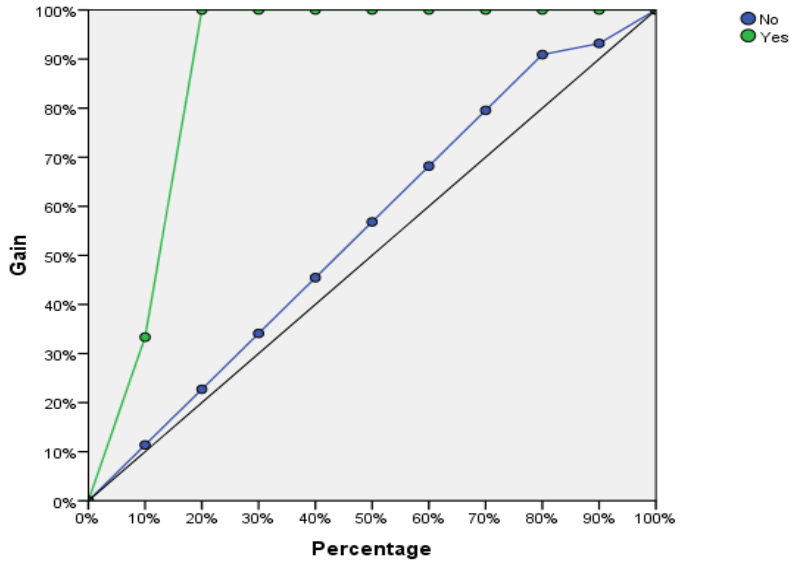
Figure VII: ROC curve



Source: constructed from survey data

The ROC curve evaluates the model's performance in predicting respondents' willingness to purchase a new life insurance product. The two curves ("Yes" and "No") demonstrate that the model performs much more effectively than random guessing, as both curves are well above the diagonal baseline. The steep rise of the "Yes" curve near the y-axis indicates high sensitivity, showing the model is highly effective in correctly identifying individuals willing to buy, with minimal false positives. Similarly, the "No" curve highlights the model's reliability in identifying those unwilling to purchase. The clear separation between the curves reflects the model's strong ability to distinguish between the two groups, ensuring accurate classification. This performance, characterized by high sensitivity and specificity for both groups, underscores the model's effectiveness for applications such as identifying potential buyers for targeted marketing efforts.

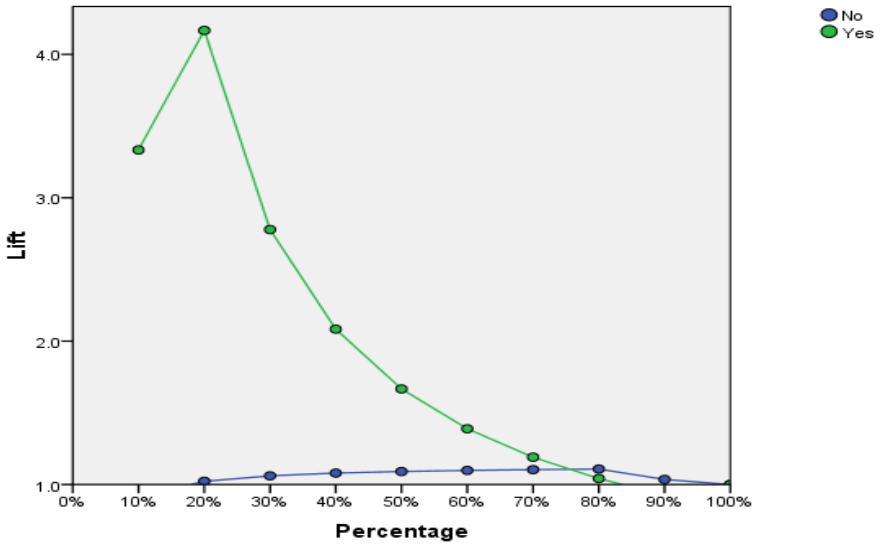
Figure VIII: Cumulative gain carve



Source: compilation data (survey)

This diagram shows how well the model predicts respondents' willingness to purchase life insurance. The **green line**, representing those willing to buy, quickly rises, indicating that the model effectively identifies buyers, capturing nearly 90% of them with only 40% of the data. Alternatively, the blue line for non-buyers demonstrates a slower, more gradual rise, indicating that the model is less accurate at predicting non-purchasers, although it still outperforms random chance. The gap between the two lines highlights the model's ability to distinguish between buyers and non-buyers, making it useful for targeted marketing strategies, but also indicating room for improvement in predicting non-purchasers.

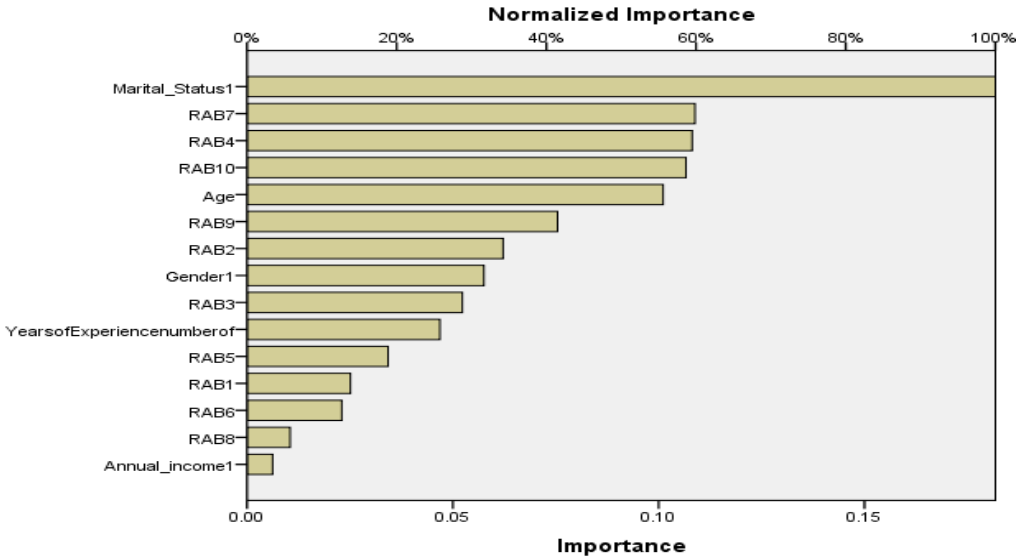
Figure IX: Lift chart



Source: compilation data (survey)

The lift chart analysis highlights the significance of concentrating on high-likelihood customer segments to optimize marketing efforts for life insurance products. The model demonstrates strong predictive power, with the highest lift observed in the top 10-20% of respondents, indicating they are most likely to purchase. Targeting these individuals can significantly enhance conversion rates and return on investment. However, as the target audience expands, the lift diminishes, reflecting reduced efficiency in reaching potential buyers. This underscores the value of leveraging predictive models to prioritize resources effectively, avoid over-targeting, and guide strategic decisions. Continuous model refinement can further improve accuracy and campaign performance.

Figure X: Independent Variable Importance



Source: compilation data (survey)

The analysis depicts that the latent variable **Risk Aversion Bias (RAB)** has a strong and consistent impact, with several RAB-related variables (e.g., **RAB7; RAB4**; **RAB10**) ranking among the top most impactful variables. Demographic variables, such as **Age** and **Gender1**, have moderate importance, while **Annual_income1** ranks the lowest, contributing minimally. This highlights that although specific demographic factors, like marital status, can dominate in predictive influence, **Risk Aversion Bias** provides broader and more consistent predictive value, underscoring the critical nature of combining behavioural and demographic factors for more accurate and comprehensive modelling.

4.3 Bootstrapping analysis

Since the sample size in this study is comparatively small (fifty), there may be concerns regarding the robustness and reliability of applying neural network analysis. In certain instances, limited sample sizes may result in overfitting or unreliable outcomes when utilizing complex models like neural networks analysis. To address this potential issue and validate the model's applicability, the authors employed bootstrapping techniques, resampling the data twenty times. This resampling process helped to assess the consistency and stability of the model's results across different subsets of the data. The findings demonstrated that the model's applicability in the present study remains valid and effective, even with a smaller sample size, thereby reinforcing the credibility of its application in this study.

Table 6: Model Summary scores for Internal consistency

	Training		Testing	
	Cross Entropy Error	Percent Incorrect Predictions	Cross Entropy Error	Percent Incorrect Predictions
run1	10.131	12.90%	2.527	12.50%
run2	2.058	0.00%	3.099	25.00%
run3	0.003	0.00%	0.024	0.00%
run4	4.529	6.50%	2.15	25.00%
run5	0.827	0.00%	0.293	0.00%
run6	4.809	7.40%	2.908	25.00%
run7	0.019	0.00%	0.043	0.00%
run8	10.063	8.30%	4.057	28.60%
run9	13.88	25.00%	4.081	12.50%

run10	0.008	.00%	0.00%	0%
run11	0.011	0.00%	5	0%
run12	0.542	0.00%	0.67	0.00%
run13	10.661	13.80%	0.34	0.00%
run14	0.404	0.00%	0.433	0.00%
run15	16.785	6.20%	5.189	37.50%
run16	7.06	15.40%	1.527	14.30%
run17	3.288	0.00%	1.852	12.50%
run18	5.265	9.70%	0.144	0.00%
run19	5.691	10.30%	2.004	12.50%
run20	7.188	11.10%	3.988	25.00%
Mean	8.6595	0.12	3.2575	0.1875
SD	5.029671726	0.070537858	1.767088074	0.123533332
CV	25.29759767	0.004975589	3.122600261	0.015260484

Source: compilation data (re-sampling model score)

The table evaluates model performance for internal consistency using cross-entropy error and percent incorrect predictions during training and testing. Training errors show high variability (mean: 8.66, SD: 5.03), while testing errors are lower on average (mean: 3.26, SD: 1.77) but more consistent. The low variability in testing metrics suggests overall good generalization, though the inconsistency in training errors highlights the need for better optimization or hyperparameter tuning. Runs with consistently low training and testing metrics are preferred for deployment.

Table 7: Classification score regarding Internal consistency

Sample	Training			Testing		
	yes	no	overall	yes	no	overall
sam1	0.00%	100	87.10%	0.00%	100.00%	87.50%
sam2	6.90%	93.10%	100.00%	0.00%	100.00%	75.00%
sam3	7.70%	92.30%	100.00%	0.00%	100.00%	100.00%
sam4	0.00%	100%	93.50%	0.00%	100.00%	75.00%
sam 5	8.00%	92.00%	100.00%	0.00%	100.00%	100.00%
sam 6	0.00%	100%	92.60%	0.00%	100.00%	75.00%
sam 7	17.90%	82.10%	100.00%	0.00%	100.00%	100.00%
sam 8	0.00%	100.00%	91.70%	0.00%	100.00%	71.40%
sam 9	21.40%	78.60%	75.00%	12.50%	87.50%	87.50%
sam 10	20.00%	80.00%	100.00%	0.00%	100.00%	100.00%
sam 11	20.00%	80.00%	100.00%	0.00%	100.00%	100.00%
sam 12	20.00%	80.00%	100.00%	0.00%	100.00%	100.00%
sam 13	0.00%	100.00%	86.20%	0.00%	100.00%	100.00%
sam 14	16.70%	83.30%	100.00%	0.00%	100.00%	100.00%
sam 15	0.00%	100.00%	93.80%	0.00%	100.00%	62.50%
sam 16	66.00%	87.00%	84.00%	0.00%	100.00%	85.70%
sam 17	10.70%	89.30%	100.00%	0.00%	100.00%	87.50%
sam 18	22.60%	77.40%	90.30%	0.00%	100.00%	100.00%
sam 19	10.30%	89.70%	89.70%	0.00%	100.00%	87.50%
sam 20	0.00%	100.00%	88.90%	0.00%	100.00%	75.00%
Mean	12.41%	585.24%	93.64%	0.63%	99.38%	88.48%
SD	0.152787	22.1602	0.070955	0.027951	0.027951	0.123533
CV	0.023344	491.0746	0.005035	0.000781	0.000781	0.01526

Source: Source: compilation data (re-sampling classification score)

The table above evaluates the model's internal consistency in classifying the "Yes" and "No" categories during training and testing. The model demonstrates strong overall accuracy (mean: **93.64%** in training and **88.48%** in testing), with "No" predictions achieving near-perfect performance (mean: **100%** in training and **99.38%** in testing). The mean overall accuracy is **93.64%**, indicating good model performance on the training dataset. The mean overall testing accuracy is **88.48%**, slightly lower than the training accuracy but still strong, reflecting reasonable generalization.

Table 8: Area Under the Curve score – internal consistency

Sample	yes	no
1	0.853	0.853
2	0.985	0.985
3	1	1
4	0.921	0.921
5	1	1
6	0.96	0.96
7	1	1
8	0.952	0.952
9	0.606	0.606
10	1	1
11	1	1
12	1	1
13	0.848	0.848
14	1	1
15	0.797	0.797
16	0.879	0.879
17	0.984	0.984
18	0.943	0.943
19	0.962	0.962
20	0.84	0.84
Mean	0.9265	0.9265
SD	0.099650178	0.099650178
CV	0.009930158	0.009930158

Source: compilation data (re-sampling AUC score)

The table demonstrates the model's strong internal consistency, with a high mean AUC score of **0.9265** for both "Yes" and "No" classifications, indicating excellent overall discriminatory ability. The low variability (SD: **0.0997**, CV: **0.99%**) highlights consistent performance across samples. Several cases (e.g., **3, 5, 7, 10, 11, 12, 14**) achieve perfect AUC scores of **1.0**, reflecting exceptional classification accuracy. While most samples perform well, weaker scores in **sample 9** (AUC: **0.606**) and **sample 15** (AUC: **0.797**) suggest areas that can be improved, such as better feature selection or handling class imbalances. The identical AUC scores for "Yes" and "No" across all samples indicate fairness, ensuring the model is unbiased in its classifications. Overall, the results confirm the model's reliability with minor opportunities for refinement.

Table 9: Independent Variable Importance Score for Internal Consistency

Sample	Age	Gender	Marital status	No. of dependence	Years of work experience	Annual income	Risk Aversion score
1	0.021	0.025	0.087	0.01	0.048	0.027	0.0785
2	0.034	0.105	0.033	0.03	0.032	0.052	0.0722
3	0.023	0.122	0.012	0.149	0.03	0.032	0.0592
4	0.067	0.048	0.026	0.012	0.063	0.003	0.0792
5	0.023	0.059	0.022	0.008	0.006	0.129	0.0755
6	0.072	0.027	0.124	0.151	0.095	0.033	0.0497
7	0.079	0.07	0.168	0.054	0.029	0.042	0.056
8	0.084	0.03	0.167	0.107	0.067	0.027	0.0518
9	0.147	0.075	0.007	0.002	0.096	0.107	0.057
10	0.078	0.008	0.072	0.003	0.125	0.099	0.062
11	0.024	0.042	0.144	0.11	0.08	0.039	0.056
12	0.034	0.094	0.095	0.002	0.023	0.088	0.0667
13	0.087	0.037	0.038	0.088	0.13	0.107	0.1538
14	0.07	0.05	0.098	0.003	0.124	0.075	0.0582
15	0.162	0.018	0.046	0.103	0.021	0.09	0.0559
16	0.117	0.03	0.167	0.113	0.078	0.05	0.0443
17	0.088	0.032	0.161	0.003	0.142	0.03	0.0548
18	0.069	0.105	0.252	0.005	0.029	0.06	0.0486
19	0.024	0.054	0.082	0.03	0.061	0.063	0.0716
20	0.125	0.036	0.086	0.042	0.14	0.079	0.0488

Mean	0.0714	0.05335	0.09435	0.05125	0.07095	0.0616	0.06499
SD	0.04243	0.03209	0.066095843	0.053306339	0.044013724	0.033792556	0.023337138
CV	0.0018	0.00103	0.004368661	0.002841566	0.001937208	0.001141937	0.000544622

Source: compilation data (re-sampling independent variable score)

The table highlights the **Risk Aversion Score** as a consistent and moderately significant predictor of internal consistency, with a mean importance score of **0.065** and the lowest variability (SD: **0.0233**, CV: **0.0545%**) among all variables. This consistency across samples indicates that risk aversion bias has a stable influence on the model's outcomes. In specific cases, such as **sample 13** (importance score: **0.1538**), its impact is particularly pronounced, underscoring its relevance in shaping individual decisions. In comparison with other variables, like **Marital Status** (mean: **0.0944**) or **Annual Income** (mean: **0.0616**), risk aversion bias emerges as a reliable and influential factor in determining internal consistency. This significance suggests that individuals' aversion to risk plays a critical role in decision-making, making it a key variable for further analysis and model refinement.

5. Discussions

The neural network effectively captures the impact of various risk aversion factors on an individual's decision to purchase life insurance, with weighted connections indicating the positive or negative influence of each variable. It highlights the most significant risk aversion variables (e.g., those with strong positive or negative weights) that influence an individual's decision to purchase life insurance. These variables can guide targeted strategies to address customer concerns or motivations. Additionally, by employing the softmax activation function in the output layer, the model predicts the likelihood of two distinct outcomes (willingness vs. unwillingness), providing actionable insights for decision-making in insurance marketing strategies.

The comparison reveals that among the variables in the model, the demographic variable **Marital_Status1** emerges as the most impactful; however, it lacks consistency. In contrast, the latent variable **Risk Aversion Bias** demonstrates a higher level of consistency with several RAB-related variables (e.g., **RAB7**, **RAB4**, **RAB10**). Other demographic variables, such as **Age** and **Gender1**, show moderate importance, while **Annual_income1** contributes minimally, ranking lower than all RAB-related variables. This indicates that Risk Aversion Bias offers stronger and more extensive predictive value, highlighting the importance of incorporating both behavioural and demographic variables for a holistic understanding of decision-making.

6. Conclusion

The analysis demonstrates that the neural network effectively models the influence of various risk aversion factors and demographic variables on an individual's decision to purchase life insurance. The weighted connections in the network highlight the positive or negative influence of each variable, allowing for the identification of key factors that drive decision-making. Among these, **Risk Aversion Bias (RAB)** stands out as a consistent and robust predictor, with specific variables such as **RAB7, RAB4, and RAB10** showing significant contributions. This underscores the importance of behavioural factors in understanding customer preferences.

However, the demographic variable Marital_Status1, although the most influential in the model, lacks consistency, which undermines its predictive reliability. Other demographic factors, such as **Age** and **Gender1**, display moderate importance, whereas **Annual_Income1** shows minimal influence. This suggests that while demographic variables provide some explanatory power, the integration of latent behavioural factors, like Risk Aversion Bias, offers a more comprehensive and reliable approach to predicting life insurance purchasing decisions.

Furthermore, by leveraging the **softmax activation function** in the output layer, the neural network provides actionable insights by predicting the likelihood of two distinct outcomes: willingness and unwillingness to purchase life insurance. These insights can guide insurance companies in designing targeted marketing strategies, addressing customer concerns, and emphasizing key motivational factors to improve sales and customer satisfaction. The findings highlight the need to combine both demographic and behavioural variables for a holistic understanding of consumer decision-making in the life insurance sector.

7. Implications and Future Scope of Study

7.1 Implications

* The neural network competently captures the impact of risk aversion factors and demographic variables on life insurance purchasing decisions by leveraging weighted connections to identify the positive or negative impact of each variable, thereby aiding in decision-making analysis.
* Integrating **latent behavioural factors** like Risk Aversion Bias with demographic data offers a **more reliable and holistic predictive model** for life insurance decision-making.
* Additionally, the model quantifies the contributions of various risk aversion biases, offering actionable insights to inform product design, personalized offers, and tailored communication strategies. The interpretability of model weights and biases further builds trust among stakeholders, making the neural network a transparent and reliable tool for business decision-making.

7.2 Future scope of study

- **Model Refinement**: Continuous improvements in the neural network, such as experimenting with deeper architectures or alternative activation functions, can enhance its accuracy and performance in predicting both willing and non-willing respondents.
- **Cross-Industry Applications**: The methodology could be extended to other sectors (e.g., healthcare, retail) to predict consumer decisions based on analogous factors, broadening the utility of such neural network-based models.
- **Ethical and Regulatory Considerations**: As predictive models are deployed, upcoming research should focus on ethical implications, such as fairness and bias in predictions, and compliance with data privacy regulations.

Acknowledgement

Acknowledging the contributions of those who have supported me in bringing this paper to its current form is essential. I express my heartfelt gratitude to my mentor, **Prof. Amalesh Bhowal**, for his invaluable guidance and unwavering support throughout this journey, from the initial development of the questionnaire to shaping the final draft of the paper. I also extend my sincere thanks to the faculty members of the

Department of Commerce for their encouragement, assistance, and motivation, which have been instrumental in advancing my work and bringing it to completion.

Funding Details

The author declares that no funding was sought or received from any individual or organization to support the development of this work.

References (APA format)

1. Aleksandrova et al. (2023). A Survey on AI implementation in finance,(cyber) insurance and financial controlling. Risks, 11(5), 91.
2. Ankitha, S., & Basri, S. (2019). The effect of relational selling on life insurance decision making in India. International Journal of Bank Marketing, 37(7), 1505-1524.
3. Azzone et al. (2022). A machine learning model for lapse prediction in life insurance contracts. Expert Systems with Applications 191, 116-261.
4. Cather, D. A. (2010). A gentle introduction to risk aversion and utility theory. Risk Management and Insurance Review, 13(1), 127-145.
5. Espinosa, O., & Zarruk, A. (2021). The importance of actuarial management in insurance business decision-making in the twenty-first century. British Actuarial Journal, 26, e14.
6. Gramegna, A., & Giudici, P. (2020). Why buy insurance? An explainable artificial intelligence approach. Risks, 8(4), 137.
7. Jensen, N. R., & Steffensen, M. (2015). Personal finance and life insurance under separation of risk aversion and elasticity of substitution. Insurance: mathematics and economics, 62, 28-41.
8. Ji, Z., et al. (2024), "Research on insurance policy formulation based on ARIMA-KMEANS algorithm. Highlights in Science, Engineering and Technology, vol 101, 307-311
9. Karni, E., & Zilcha, I. (1985). Uncertain lifetime, risk aversion and life insurance. Scandinavian Actuarial Journal, 1985(2), 109-123.
10. Khan et al. (2023). Unlocking the investment puzzle: The influence of behavioural biases & moderating role of financial literacy. Journal of Social Research Development, 4(2), 433-444.
11. Kishor, N. & Retika (2022). Risk preferences for financial decisions: Do emotional biases matter?. Journal of Public Affairs, 22(2), e2360.
12. Mohd A et al. (2021). How financial literacy moderate the association between behaviour biases and investment decision? Asian Journal of Accounting Research, v-7, 17-30.
13. Owens et al. (2022). Explainable Artificial Intelligence (XAI) in insurance. Risks 10: 230.
14. Suto, M., & Toshino, M. (2005). Behavioural biases of Japanese institutional investors: Fund management and corporate governance. Corporate Governance: An International Review, 13(4), 466-477.
15. Seo, K., & Sharma, A. (2018). CEO overconfidence and the effects of equity-based compensation on strategic risk-taking in the US restaurant industry. Journal of Hospitality & Tourism Research, 42(2), 224-259.
16. Vahabi, S., & Payandeh N. A. T. (2022). Design of a Pure Endowment Life Insurance Contract Based on Optimal Stochastic Control. Journal of Mathematics and Modeling in Finance, 2(2), 37-52.
17. Zoudbin et al. (2019). Risk-based portfolio modelling: Kahneman-Tversky approach. Journal of Neurodevelopmental Cognition, 2, 88-97.