Doctoral School of Economics, Business and Informatics

Thesis Booklet

for the Ph.D. Dissertation by

Asefeh Asemi

titled

A Novel Combined Investment Recommender System Using Adaptive Neuro-Fuzzy Inference System

Supervisor: Prof. Andrea Ko

Budapest, 2022

Corvinus University of Budapest

Department of Business Informatics

Thesis Booklet

for the Ph.D. Dissertation by

Asefeh Asemi

titled

A Novel Combined Investment Recommender System Using Adaptive Neuro-Fuzzy Inference System

Supervisor: Prof. Andrea Ko

Table of Contents

I. INTRODUCTION 4
II. THEORETICAL BACKGROUND 5
III. METHODOLOGY 8
IV. PROPOSED SYSTEM FRAMEWORK 10
V. EXPERIMENTAL RESULTS 11
A. Clustering Investment-Type Output 12
B. Inputs for the Combined Recommender ANFIS 15
C. Proposing the Combined ANFIS 26
VI. INCORPORATION OF EXPERT KNOWLEDGE AND INVESTORS' FEEDBACK 32
VII. DISCUSSION 33
VIII. CONCLUSION 36
IX. ABBREVIATIONS 37
X. REFERENCES 38
XI. PUBLICATIONS RELATED TO THIS DISSERTATION 39
A. Journal Articles 39
B. Conference Papers 40

I. INTRODUCTION

Investment decision-making is a complex and challenging task, particularly for individual investors. It is a complex process that requires a thorough analysis of various factors, including the investor's risk tolerance, financial goals, and investment experience. The abundance of financial products and services, along with the ever-changing market conditions, make it difficult for investors to make well-informed decisions. Traditional investment recommendations are often based on a one-size-fits-all approach, which may not be suitable for individual investors with different preferences and investment goals. In recent years, there has been an increasing demand for personalized investment recommendations that consider the unique needs and preferences of individual investors. The problem statement of this study is to develop a combined Investment Recommender system that can accurately predict the investment type based on various factors such as demographic, decision key factors, personality traits, investor experiences, financial situation, and managerial traits. The current investment decision-making process is complex and time-consuming, and there is a need for a system that can provide accurate and efficient investment recommendations to investors. To address this problem, this research aims to develop a new framework for a combined recommender system for investment services using Adaptive Neuro-Fuzzy Inference System (ANFIS). The proposed system will utilize data collection, data analysis, and decision-making layers, along with the ANFIS, K-Means, and Self-Organizing Map (SOM) methods to provide accurate investment recommendations. The system will also incorporate expert knowledge and investors' feedback to improve its accuracy and effectiveness. The research objective is to provide a novel framework for a combined recommender system based on ANFIS and customized for investment services. To achieve this objective, several research questions have been formulated, including: How can a neuro-fuzzy inference-based approach be used to personalize investment strategies for individual investors? What are the key factors that influence the performance of a neuro-fuzzy inference-based investment recommender system? How can user feedback be incorporated into a neuro-fuzzy inference-based investment recommender system to improve its performance over time? And how can the proposed neuro-fuzzy inference-based recommender system be compared with traditional methods of investment recommendation? To answer these research questions, experiments and results were reported in the following areas: Clustering Output Investment-type/Product, Preparation Inputs for the Combined Recommender ANFIS including Clustering Input demographics data, Clustering Input Decision Key Factors data, Clustering Input Personality Traits Data, Clustering Input Experiences Data, Clustering Input Financial Data, Clustering Input Managerial Traits Data, and Proposing the Combined ANFIS. This research aims to provide a valuable contribution to the field of investment decision-making by developing an intelligent and personalized investment recommender system. The proposed system has the potential to assist individual investors in making well-informed investment decisions, ultimately leading to better investment outcomes. Figure 1 is a conceptual overview of the research. It presents a visual representation of the main components and steps involved in the research. It gives a general idea of how the different elements of the study are connected and how they relate to each other. The conceptual overview is a high-level view of the research, and it helps to understand the main objectives and goals of the research. It helps to understand the flow of the research and the connections between the different elements of the study.

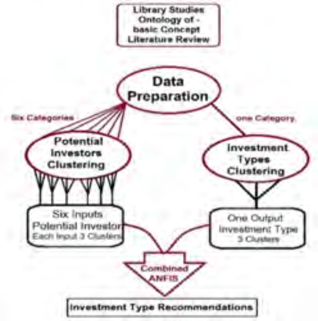

Figure 1. A conceptual overview of the research

II. THEORETICAL BACKGROUND

Investment decision-making is a complex task that requires the analysis of various factors, including the investor's risk tolerance, financial goals, and investment experience. Traditional investment recommendations are often based on a one-size-fits-all approach, which may not be suitable for individual investors with different preferences and investment goals. To address this problem, there has been a growing demand for personalized investment recommendations that consider the unique needs and preferences of individual investors. One approach to providing personalized investment recommendations is the use of recommender systems. Recommender systems are computer-based systems that use algorithms to recommend items or services to users based on their preferences and history of interactions. These systems have been widely used in various applications, such as e-commerce, music and movie recommendations, and social media. In the field of investment decision-making, recommender systems have been used to provide personalized investment recommendations. However, traditional recommender systems often rely on a single data source, such as financial market data, which may not be sufficient to provide personalized recommendations. In addition, traditional recommender systems may not be able to handle the complexity and uncertainty of the financial market. To address these limitations, this research proposes the use of a neuro-fuzzy inference-based recommender system for investment services. Neuro-fuzzy systems (NFS) are a type of artificial intelligence that combines the capabilities of ANN and fuzzy logic. These systems can handle complex and uncertain data and can learn from data and adapt to changing conditions. In this research, a neuro-fuzzy inference-based recommender system (ANFIS) will be used to provide personalized investment recommendations for individual investors. The proposed system will utilize a combination of demographic, decision key factors, personality traits, investor experiences, financial situation, and managerial trait data to provide personalized recommendations. The ANFIS system will be trained using historical data and will be able to adapt to changing market conditions and user preferences. In addition, this research will investigate how user feedback can be incorporated into the ANFIS system to improve its performance over time. The proposed system will also be compared with traditional methods of investment recommendation to assess its effectiveness. Overall, this research aims to provide a valuable contribution to the field of investment decision-making by developing an intelligent and personalized investment recommender system based on ANFIS. The proposed system has the potential to assist individual investors in making well-informed investment decisions, ultimately leading to better investment outcomes. Review of relevant studies on various aspects of stock prediction and investment efficiency, including the use of genetic algorithms (GAs) and random forests, data clustering, the role of corporate governance in cash holding and investment efficiency, the impact of creative accounting on the marketing of shares, customer experience management in retailing, and the use of machine learning and artificial intelligence in stock prediction and recommender systems. One study by (Abraham et al., 2022) explore the use of a GA and random forest to predict stock trends, while (Aggarwal & Reddy, 2014) examine data clustering algorithms and their applications in stock prediction. Aksar and colleagues examine the relationship between cash holding and investment efficiency for financially distressed firms, and the moderating effect of corporate governance (Aksar et al., 2022). A study investigates the role of creative accounting in increasing the marketing of shares and profits in the Iraqi stock exchange (AL-Khafaji et al., 2022), and Andajani examines customer experience management in retailing (Andajani, 2015). Asemi and Ko propose a novel combined business recommender system model using customer investment service feedback (Asemi & Ko, 2021), furthermore, a study applies an ANFIS to evaluate dysarthric automatic speech recognition systems (Asemi et al., 2019). Benkraiem et al. investigate the impact of economic policy uncertainty, investor protection, and excess cash on stock value in a cross-country comparison (Benkraiem et al., 2023). Birim et al. use an ANFIS-PSO method to estimate the return rate of blockchain financial products (Birim et al., 2022), and Chatterjee et al. propose an NLP and LSTM-based stock prediction and recommender system for KOSDAQ and KOSPI (Chatterjee et al., 2021). Chen reviews investment products (Chen, J., 2020), and Chen et al. study user perception of sentiment-integrated critiquing in recommender systems (L. Chen et al., 2019). Chen et al. propose a cluster-based mutual fund classification and price prediction system using machine learning for Robo-advisors (X. Chen et al., 2021), and D'lima and Khan use ANN and ANFIS to predict FOREX rates (D'lima & Khan, 2016). Davies et al. implement a type-2 fuzzy logic-based prediction system for the Nigerian stock exchange (Davies et al., 2022), and Ezhilarasi and Sashi Rekha propose a secure recommendation application for environment crops using big data analytics with a fuzzy framework (Ezhilarasi & Sashi Rekha, 2020). Faridniya and Faridnia provide a model for allocating resources and choosing investment types using Data Envelopment Analysis (Faridniya & Faridnia, 2019). Garbade discusses the differences between AI, machine learning, and deep learning (Garbade, 2021), and Gong and Li study the investment value and opportunity of renewable energies under the carbon trading system (Gong & Li, 2016). Halloumis examines the chain of the money cycle (Halloumis, 2022), and Han et al. provide an overview of data mining concepts and techniques (Han et al., 2012). Hernández et al. propose an investment recommender multi-agent system in financial technology (Hernández et al., 2019), and Huang et al. investigate neural network models for stock selection based on fundamental analysis (Huang et al., 2019). These studies provide a comprehensive examination of various aspects of stock prediction and investment efficiency, utilizing a range of methods and techniques including machine learning, artificial intelligence, and data analysis. Also, the literature suggests that NFS have the potential to provide personalized investment recommendations to individual investors and overcome the limitations of traditional recommender systems. However, more research is needed to further explore the use of NFS in investment decision-making and to evaluate their performance in comparison to traditional methods.

III. METHODOLOGY

The proposed neuro-fuzzy inference-based recommender system (ANFIS) is designed to assist potential investors in making informed decisions about their investments. ANFIS is a hybrid model that combines the capabilities of both neural networks and fuzzy logic systems, which allows it to handle both numerical and categorical data and make accurate predictions based on the input data. The system utilizes a combination of datasets from a repository of Mendeley, including information on investment type, demographic, decision key factors, personality traits, investor experiences, financial situation, and managerial traits. The goal of this methodology is to outline the methods used to collect, preprocess, and analyze the data, as well as the methods used to train and evaluate the ANFIS system.

- Data collection and preprocessing methods: The proposed system utilizes a combination of seven categories of data from the repository of Mendeley for potential investors related to investment type, demographic, decision key factor, personality traits, experiences, financial and managerial traits from six references (Asemi, 2023). The data is collected and preprocessed to ensure that it is clean and ready for analysis. This includes removing any missing or duplicate data and normalizing the data to ensure that it is in a consistent format.
- Inputs and Outputs: The ANFIS system uses JMP to cluster investment types and six criteria
- (demographic, decision key factor, personality traits, experiences, financial and managerial traits) as inputs. The data is then separated into various clusters based on these criteria and investment types to guarantee that the ANFIS system can correctly predict the investment type based on the input data.
- Training and evaluation methods for the ANFIS system: The ANFIS system is trained and evaluated using a combination of supervised and unsupervised learning techniques. The system is trained on a set of labeled data, and the performance of the model is evaluated using various

metrics such as accuracy, precision, and recall. The system is then fine-tuned based on the evaluation results to improve its performance.

o Discussion of how user feedback will be incorporated into the ANFIS system: User feedback is an important aspect of the ANFIS system, as it allows the system to learn and adapt to the user's preferences. The system is designed to incorporate user feedback by allowing users to provide ratings and feedback on the recommendations made by the system. This feedback is then used to update the system's parameters and improve its performance over time. Table 1 shows the research methodology in several sections with descriptions.

Table 1. Research Methodology

Section	Description	Tools & Techniques
Research Design	The research design used in this study is a quantitative approach using ANFIS to predict investment type based on demographic, decision key factors, personality traits, investor experiences, financial situation, and managerial traits.	ANFIS (MATLAB)
Data Collection	Data was collected using a Web-based Investment Portfolio Questionnaire, A part of the dataset prepared for this research and published in the Mendeley Data from a sample of potential investors who have made investments in the past.	Online Questionnaire (Kérdőív, 2019) Mendeley Data (Asemi, 2023)
Data Analysis	The collected data was analyzed using JMP to identify patterns and relationships between the input variables and the output variable of investment type. Clusters were created based on investment type, demographic, decision key factor, personality traits, experiences, financial, and managerial traits.	JMP, Cluster Analysis
ANFIS Model	ANFIS was used to create a model that can accurately predict investment type based on the input data and the clusters created. The model was tested and validated using a sample of data not used in the training process.	ANFIS (MATLAB), Model Testing and Validation
Results and Discussion	The results of the study were discussed in terms of the accuracy of the ANFIS model in predicting investment type and the significance of the input variables in determining investment type.	Data Analysis and Interpretation
Conclusion and Recommendations	The study concludes that ANFIS can be used as an effective tool for predicting investment type based on demographic, decision key factors, personality traits, experiences, financial, and managerial traits. Recommendations for future research include expanding the sample size and testing the model in different investment scenarios.	Future Research Planning

IV. PROPOSED SYSTEM FRAMEWORK

Figure 2 illustrates the proposed system framework of the combined Investment Recommender system. It provides an overview of the different layers and components of the system, including data collection, data analysis, and decision-making. The figure highlights the main methods used in this system, such as the ANFIS, K-Means, and SOM. The framework also represents the inputs and output data used in the ANFIS and the clustering of the input data based on different criteria such as demographic, decision key factors, personality traits, experiences, financial, and managerial traits. The data collection layer is the initial step in the system, where data is gathered from investors. This data is then passed through the data analysis layer, where it is processed and organized into different clusters based on the criteria mentioned above. The clustering process is done utilizing the K-Means and SOM methods, which help in grouping similar data points together. The ANFIS is then applied to the clustered data, which is used to predict the investment type based on the input data. The ANFIS is a type of artificial neural network (ANN) that can learn from data and make predictions based on the input data. The output of the ANFIS is then passed through the decision-making layer, where it is used to provide investment recommendations to the investors. In addition to the methods and data mentioned in the previous explanation, the proposed system framework in figure 2 also highlights the use of expert knowledge and investors' feedback.

Figure 2. Proposed framework for the combined Investment Recommender system

The system allows for the incorporation of expert knowledge and feedback through the addition, change, or deletion of rules generated by the ANFIS. This is done by allowing experts to review and adjust the rules based on their expertise and the feedback received from investors. Expert knowledge can be used to improve the accuracy of the system by providing additional information and insights that the system may not have considered. This can be done by adding new rules to the ANFIS or adjusting the existing ones based on expert knowledge. Similarly, investors' feedback can be used to adjust the system to better match their specific needs and preferences. This can be done by modifying the rules based on the feedback received from investors. Overall, the proposed system framework in figure 2 provides a comprehensive understanding of the different layers, methods, and data used in the combined Investment Recommender system. It also demonstrates how the different components of the system work together to provide accurate investment recommendations. It allows for the use of expert knowledge and investors' feedback to improve the accuracy and effectiveness of the combined Investment Recommender system. It gives flexibility to the system to adjust based on the experts' viewpoints and investors' feedback, making it more efficient and effective in providing investment recommendations.

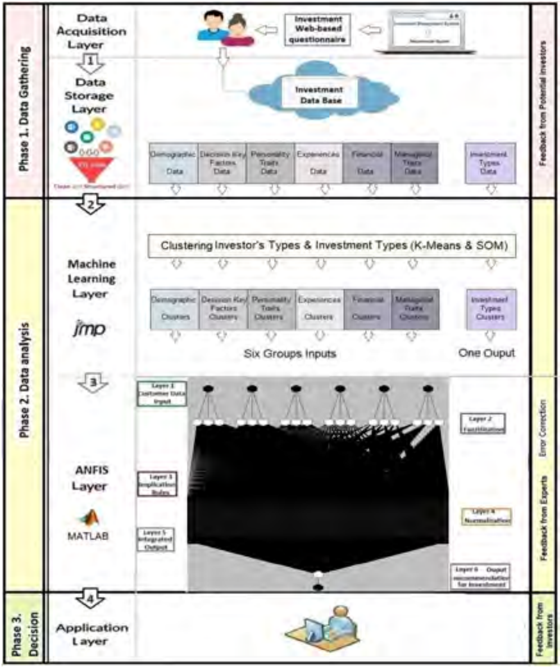

Figure 2. A framework of the combined Investment Recommender system

V. EXPERIMENTAL RESULTS

To answer the research questions, the experiments and results are reported in the following:

A. Clustering Investment-Type Output

To develop the ANFIS system, the output was defined as the investment type/product. This output was chosen based on the research objectives and was determined using four questions in the questionnaire (pages 24-27). The data collected included four variables with 1542 respondents. Question 24 asked about the investment products used by the potential investors, with five options available: listed stock, mutual fund, voluntary pension fund, government securities, and other financial products. Respondents were able to select multiple options, resulting in 31 categories being created based on the multiple choices. Question 25 asked if the respondents had made any stock market investments in the last 3 years, with options of "Yes" or "No." Question 26 asked if the respondents regularly monitored or followed the performance of their stock investments, with options of "Yes" or "No." Question 27 asked if the respondents had any government bond investments, with options of "Yes" or "No." These questions and data were used as a sample to propose the system, and companies can adapt the questions based on their objectives. Table 2 shows the missing values and variances of investment-type data. Each row represents a different aspect of the data, such as the number of missing values or the individual variance. The columns correspond to the different questions or variables in the data set.

Table 2. Missing values and variances of investment-type data

Questions	P24 (Nominal)	P25 (Ordinal)	P26 (Ordinal)	P27 (Ordinal)
Missing values Count in columns	198	4	2	2
Percentage of missing values	12.8%	0.3%	0.1%	0.1%
Individual variance	0.95	0.04	0.01	0.00
Cumulative variance	0.95	0.99	1.00	1.00

Clustering is a popular technique in data mining that involves dividing data into similar groups without any prior knowledge of their class labels. K-Means is one of the most widely used clustering algorithms that is based on distance-based cluster analysis. In this study, the K-Means method was used to cluster investment-type data using several software applications, including Python, R Studio, RapidMiner, Tableau, and JMP. JMP was ultimately chosen as the preferred software due to its superior results compared to other applications. The JMP software is used to cluster the data and generate outputs for the ANFIS system. The number of clusters is initially chosen randomly, and the data is assigned to one of these clusters based on the degree of similarity or proximity. The Cubic Cluster Criterion (CCC) is used to select the optimal number of clusters that best fit the data. The CCC is evaluated by Monte Carlo methods, which simulate random sampling to estimate the number of clusters using Ward's minimum variance method, K-Means, or other methods based on minimizing the within-cluster sum of squares (Clustering Methods for Unsupervised Machine Learning, 2019). Inputs and outputs for the ANFIS system include clustering investment type as the output and six criteria, including demographic, decision key factor, personality traits, experiences, and financial and managerial traits, as inputs. These inputs are processed by JMP to divide the data into different clusters based on investment type, demographic, decision key factor, personality traits, experiences, financial, and managerial traits. This ensures that the ANFIS system can accurately predict the investment type based on the input data. In the first step of clustering the data, investment-type data is imported into JMP in four columns with a total of 1542 rows. The data is then prepared for analysis using the K-Means technique in JMP. The number of clusters indicated by JMP is determined by the CCC (Canberra's Compound Criterion). Three clusters were identified for investment type using the K-Means method. The iterative clustering process shows the cluster summary, including the count of the first cluster (592), the second cluster (406), and the third cluster (340).

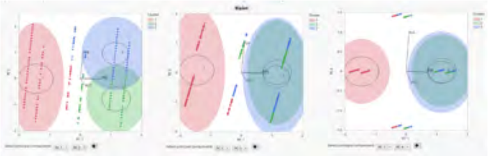

Figure 2. Clusters in the first two principal components (PCs) of investment-type data

Figure 2 illustrates the clustering of the investment-type data using the first two PCs (PC1 and PC2). The biplot displays the points representing the data and the clusters formed by the ANFIS system. The clusters are represented by different colors or symbols on the plot. The position of each point on the plot is determined by its values on the PC1 and PC2 axes, indicating the degree of association between the data and the PCs. The plot also shows the pair principle components like PC1 & PC3, and PC1 & PC4, giving a better understanding of the relationship between the different components and the clusters. Overall, Figure 2 provides a visual representation of how the ANFIS system has grouped the investment-type data based on the first two PCs.

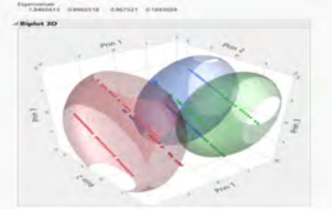

Figure 3. First three PCs of investment-type data

Figure 3 depicts a three-dimensional biplot representation of the investment-type data points and clusters in the first three PCs. The graph illustrates how the points are distributed in the space defined by the three PCs, and it also shows the clusters that have been formed from the data. The eigenvalues of the four columns of data (P24, P25, P26, and P27) are also presented, showing the relative importance of each component in explaining the variation in the data. The highest eigenvalue of 1.8465613 represents the first PC, followed by 0.9965518 for the second, 0.967521 for the third, and 0.1893659 for the fourth. These eigenvalues indicate the proportion of the total variance in the data that is explained by each PC. To evaluate the effectiveness of the clustering method for large volumes of data, investment-type data were generated using JMP software with a sample size of 10,000.

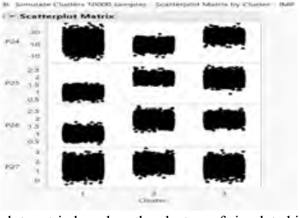

Figure 4. A scatterplot matrix based on the clusters of simulated investment-type data

A new data table was constructed with the simulated investment-type data, utilizing the estimated mixing probabilities, means, and standard deviations for each cluster obtained from the clustering process. This table was used to analyze the performance of the clustering method and its ability to accurately group the data into distinct clusters. Figure 4 is a scatterplot matrix that illustrates the relationship between different clusters of simulated investment-type data. The matrix is created using a 10,000-scale simulation, showing how the clusters are optimized over time. The figure also includes confidence ellipses, which are used to represent the level of confidence in the relationship between the different clusters. The scatterplot matrix and confidence ellipses help to visualize the distribution and similarity of the clusters, making it easier to understand the relationship between the different variables and how they influence the investment-type data.

B. Inputs for the Combined Recommender ANFIS

In proposing the system, six categories of data were used as inputs for the ANFIS including "respondents' demographics," "key factors in investment decision-making by respondents," "personality traits, knowledge and ability of the respondents," "respondent's experiences," "respondents' financial situation," and "managerial traits of the respondents." Table 3 shows the different inputs used in the ANFIS system, as well as a brief description of each input.

Table 3. Brief description of inputs for Combined ANFIS

Inputs	Description
Demographics	Age, Gender, Education level, Income, Occupation
Decision-making Factors	Risk tolerance, Goals, Time horizon, Investment knowledge
Personality Traits	Extraversion, Agreeableness, Conscientiousness, Neuroticism, Openness
Experiences	Investment history, Financial literacy
Financial Situation	Net worth, Debt, Savings
Managerial Traits	Leadership, Decision-making, Strategic thinking

JMP software was utilized to cluster each category of data and create inputs for the ANFIS system by utilizing K-Means and SOM methods. As per JMP documentation (2023), SOM is a technique developed by Kohonen (Kohonen, 1990) and further extended by several other neural network enthusiasts and statisticians. In this study, the combined method of K-Means and SOM was used for clustering demographic data, as it was found that this method resulted in better outcomes compared to using K-Means alone. SOM is an unsupervised machine learning method that can be used to cluster data with many features, and it also maps the data to a two-dimensional map to make it easier to visualize the clusters. SOM is a type of ANN in which data properties are input and clusters are output. The JMP software uses the center of the clusters selected by K-Means as a point and the probability of the presence of that point in each group. SOMs are a variation of K-Means where cluster centers are located on a grid. K-Means and SOM are both doubly iterative processes. The clustering process is repeated in two steps based on the expectation-maximization (EM) algorithm: in the expectation step, the probability of the presence of each point in a cluster is calculated, and in the second step, for each cluster, a new center is identified based on the probability of presence. This process is repeated until the stability of the clusters is achieved. In addition to forming clusters, SOM places clusters on a specific grid layout so that the points in each cluster in a multivariate space are close together. To specify the number of clusters, JMP uses the Cubic Clustering Criterion (CCC), which selects the number of clusters that best fits the data. Higher CCC values indicate that clusters fit the data well, similar to the way Silhouette scores do. The CCC can be used to estimate the number of clusters using Ward's minimum variance method, K-Means, or other methods based on minimizing the within-cluster sum of squares. The performance of the CCC is evaluated by Monte Carlo methods (Chow et al., 2003). In the initial step of clustering data for inputs of the ANFIS, six sets of data from 1542 respondents were imported into JMP software. Each set consisted of four columns containing various information, such as respondents' demographics, key factors in investment decision-making, personality traits, experiences, financial situation, and managerial traits. The data was then prepared for clustering using the K-Means and SOM methods in JMP. The rows were grouped into a specific number of clusters based on numeric variables, such as safety, excess money, computational awareness, investment fund, saving factors, and accounting knowledge. The number of clusters was determined using the CCC in JMP. The next steps of the process are explained in the following sections.

Clustering Input demographics data: The demographic data was divided into a specific number of clusters based on the numeric variables, such as gender, age, location, education, job, and income of the respondents. The result illustrates the three clusters created for the demographic data by using the combined K-Means and SOM methods. The iterative clustering process and cluster summary in 8 steps, including the count of the first cluster (210), the second cluster (270), and the third cluster (294). Additionally, the Means and Standard Deviations are indicated for each cluster. The script (written in Python) used for the clustering process is provided as follows:

```
K Means Cluster(
        Y( :Gender, :Age, :Location, :Education, :Job, :Income ),
        {SOM N Rows( 1 ), SOM Bandwidth( 0.433012701892219 ), Single Step( 0 ),
        Number of Clusters( 3 ), SOM, Go},
        SendToReport(
                Dispatch( {}, "Cluster Comparison", OutlineBox, {Close( 1 )} ),
                Dispatch(
                        {"SOM Grid 1 by 3"},
                        "Cluster Standard Deviations",
                        OutlineBox,
                        {Close( 0 )}
                )
        )
)
```

Figure 5. Parallel Coordinate Plot for Demographic Data

Figure 5, titled "Parallel Coordinate Plot for Demographic Data," is a graphical representation of the demographic data that has been clustered using the SOM method. The plot displays three separate clusters of data, each represented by connected line segments that represent each row of the data table. The figure illustrates that clusters can be similar in features, which can make it difficult to distinguish between them. However, the use of SOM allows for a clear distinction between the three clusters. The gray line in the plot represents the mean or average of the data in each cluster. The parallel coordinate plot allows for a visual representation of the clusters and the similarities or differences between them. The evaluation of the clustering method used in the study was conducted using a large volume of simulated demographic data in JMP. The data were simulated on a scale of 10,000 samples, and a new data table was created with the simulated demographic data. The estimated cluster mixing probabilities, means, and standard deviations were used to create the data table. The number of clusters indicated by JMP was based on the CCC. The iterative clustering process and cluster summary were completed in 19 steps, with the count of the first cluster being 1999, the second cluster 3126, and the third cluster 4875. The Means and Standard Deviations were also indicated for each cluster. This suggests that the data was successfully clustered into three distinct groups based on the CCC method and the K-Means algorithm, and the cluster distribution was 1999, 3126, and 4875 respectively. Additionally, the means and standard deviations of the data were calculated for each cluster, providing insights into the characteristics of the data in each cluster.

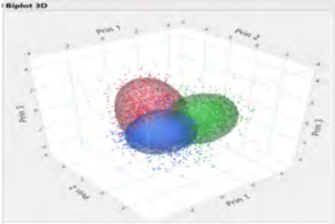

Figure 6. First three PCs of demographic data

The statement "Figure 6 shows a biplot 3D of the points and clusters in the first three PCs of the demographic data" refers to a visual representation of the demographic data that has been analyzed and clustered. The biplot 3D is a type of plot that displays both the points and the clusters in three-dimensional space, using the first three PCs of the demographic data. The PCs are a set of new variables that are derived from the original data, and they represent the most important features of the data. The biplot 3D allows for the visualization of the clusters and the points in the demographic data, making it easier to understand and interpret the results of the analysis.

Clustering Input Decision Key Factors data: The results of the iterative clustering process were displayed in 8 steps, with the count of the first cluster being 449, the second cluster 712, and the third cluster 343. The means and standard deviations were also indicated for each cluster.

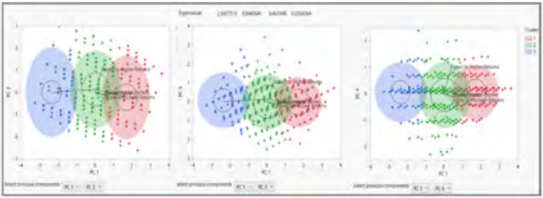

Figure 7. Biplots for Decision key factors Data Clusters SOM

A new column was added to the data table that contained the cluster assigned to each row. The clusters were based on numeric variables and were divided into a specified number of clusters for decision key factor data. Figure 7 illustrates three biplots of the points and clusters in the first two PCs of the decision key factors data. The pair of PCs analyzed are PC1 and PC2, PC1 and PC3, and PC1 and PC4. This figure displays all the rows of the three clusters of decision key factors data, and it can be observed that there is some overlap between the two clusters. To evaluate the clustering method with a large volume of data, decision key factors data were simulated in JMP with a scale of 10,000 samples. A new data table was created with simulated decision key factors data using the estimated cluster mixing probabilities, means, and standard deviations for each cluster. The number of clusters indicated by JMP is based on the CCC. The simulation used the K-Means method, and it can be observed that it shows an iterative clustering process in 18 steps. The count of the first cluster is 2323, the second cluster is 3328, and the third cluster is 4349. The Means and Standard Deviations indicated for each cluster are also included in the figure.

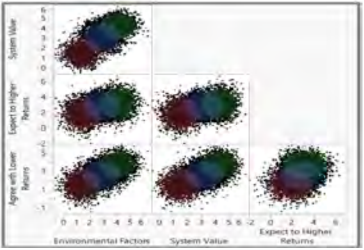

Figure 8. A scatterplot matrix based on the clusters of simulated decision key factors

Figure 8 is a scatterplot matrix that visualizes the clusters of simulated decision key factors data. The scatterplot matrix is a graph that shows the relationship between multiple variables by plotting them against each other in a grid-like format. In this figure, the data points are represented by dots on the graph and are grouped into clusters. The confidence ellipses are also plotted on the graph, which is a representation of the uncertainty of the data points in each cluster. The scatterplot matrix shows the distribution of the data points about the clusters, and how they are optimized by simulations. The scale of the simulation is 10,000, which means that the data points and clusters are represented on a scale of 10,000.

<u>*Clustering Input Personality Traits Data:*</u> Three clusters were indicated for personality traits by using the combined method of K-Means and SOM. The clustering process was iterative and included 12 steps of cluster summary, including the count of the first cluster (370), the second cluster (637), and the third cluster (471). The means and standard deviations were indicated for each cluster.

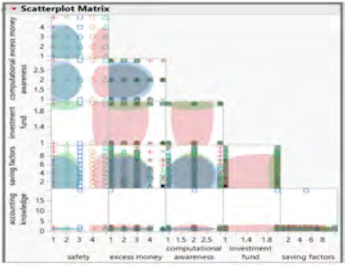

Figure 9. Scatterplot matrix for personality traits data

Figure 9 shows the regression line and confidence interval on the scatterplot matrix for the personality traits data and each cluster separately. The data shaded clear the region inside the ellipses on the scatterplot matrix between each Y variable of the personality traits data. This matrix contains ellipses, points, and a lower triangular scatter matrix for the covariates. Ellipses with different overlays are shown at each level of the categorical variable X. The linear discriminant method in this matrix is based on the pooled covariance matrix. To evaluate the effectiveness of the clustering method with a large volume of data, a simulated dataset of personality traits was created in JMP with a sample size of 10,000. The simulated data were then used to create a new table with estimated cluster mixing probabilities, means, and standard deviations for each cluster. The number of clusters was determined by JMP using the CCC. The simulation resulted in three clusters for the personality traits data when using the K-Means method. The JMP output included an iterative clustering process and cluster summary, with 18 steps in total. These included the count of the first cluster (1238), the second cluster (7436), and the third cluster (1326), as well as the means and standard deviations for each cluster.

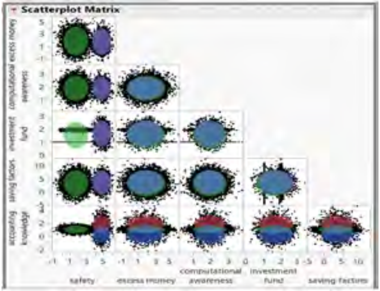

Figure 10. A scatterplot matrix based on the clusters of simulated personality traits data

Figure 10 is a visual representation of the clusters that were formed from the simulated personality traits data. The scatterplot matrix is a graphical representation that displays the relationship between multiple variables. It shows the distribution of the data points for each pair of variables on a two-dimensional plot. In this figure, the x-axis and y-axis represent different variables, and the points on the plot represent the data for each cluster. The confidence ellipses are also shown, which are used to indicate the level of certainty that a data point belongs to a specific cluster. The scatterplot matrix is based on simulations that were run on a scale of 10,000, which means that the data points shown in the figure are a sample of the 10,000 simulations that were run. This figure helps to visualize how the clusters were optimized by the simulations and how the data points are distributed within each cluster.

Clustering Input Experiences Data: The combined method of K-Means and SOM was used to cluster the experiences data, resulting in the identification of three clusters. The iterative clustering process and cluster summary were performed in 14 steps, which included the count of observations in each cluster: 514 in the first cluster, 669 in the second cluster, and 320 in the third cluster. The means and standard deviations for each cluster were also provided. Figure 11 is a visual representation of the results of clustering the simulated experiences data. A scatterplot matrix is a collection of scatter plots organized into a grid, where each scatters plot represents the relationship between two variables. In this case, the scatterplot matrix is based on the clusters of the simulated experiences data, meaning that the data points are divided into different clusters, and each scatterplot represents the relationship between two variables within each cluster. The figure also includes confidence ellipses, which are a graphical representation of the uncertainty surrounding the mean of the data. The ellipses are centered on the mean of the data and their shape and size indicate the level of confidence in the mean. In this case, the confidence ellipses are based on the current number of clusters of the simulated experiences data, meaning that they are drawn around the mean of each cluster. The figure shows how the clusters are optimized by simulations on a 10,000 scale. This means that the data points have been simulated 10,000 times, and the clusters were formed based on the patterns that emerged from these simulations. The scatterplot matrix and the confidence ellipses help to visualize the relationship between the variables and the level of uncertainty within each cluster.

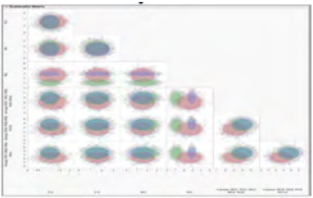

Figure 11. A scatterplot matrix based on the clusters of simulated experiences data

Overall, Figure 11 provides a visual representation of the results of clustering the simulated experiences data and how it is optimized by simulations. It helps to understand the relationship between the variables within each cluster and the level of uncertainty surrounding the mean of the data.

<u>*Clustering Input Financial Data:*</u> The results revealed the formation of three clusters for financial data using the combined method of K-Means and SOM. The iterative clustering process and cluster summary were conducted in six steps, which included the count of the first cluster as 309, the second cluster as 807, and the third cluster as 421. Additionally, the Means and Standard Deviations for each cluster were also provided.

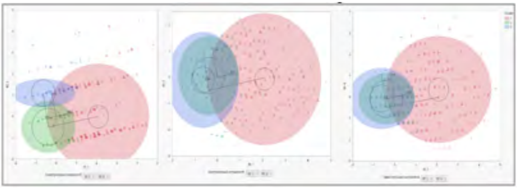

Figure 12. Example Biplots for Financial Data Clusters SOM

Figure 12 illustrates an example of biplots for financial data clusters created using the SOM method. Biplots are plots that show the relationship between multiple variables and the observations in a dataset. In this case, the biplots display the points and clusters in the first two PCs of the financial data. The three biplots show the pair PCs examples PC1 & PC2, PC1 & PC3, and PC1 & PC4. The figure demonstrates that the financial data has been divided into three clusters, which are represented by different colors. The biplots show the locations of the points within the clusters in the first two PCs. It can be observed that the clusters are overlapping, meaning that some points belong to more than one cluster. This could be due to the similarity of the points' characteristics, which made it difficult to separate them into different clusters. Additionally, in one part of the biplots, all three clusters overlap, indicating that some points belong to all three clusters. To assess the performance of the clustering method on large volumes of data, financial data were simulated using JMP software with a sample size of 10,000. A new data table was created with the simulated financial data, utilizing the estimated cluster mixing probabilities, means, and standard deviations for each cluster. The results of the iterative clustering process were summarized in 31 steps and included the count of the first cluster (1544), the second

cluster (4866), and the third cluster (3590). The means and standard deviations were also provided for each cluster.
Figure 13 is a graphical representation that shows the relationship between different variables in the simulated financial data. The scatterplot matrix is created with confidence ellipses, which are used to show the level of uncertainty or spread of the data points within each cluster. The matrix is based on the current number of clusters of the simulated financial data, which means that the clusters have been optimized through simulations on a 10,000 scale. The scatterplots show the relationship between each variable and how the data points are distributed within each cluster. Additionally, this figure allows the viewer to visually inspect the clustering results and evaluate how well the data has been grouped.

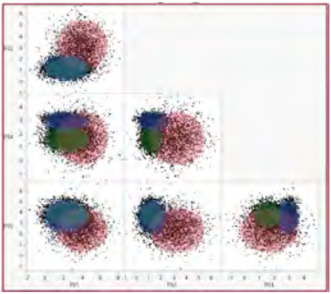

Figure 13. A scatterplot matrix based on the clusters of simulated financial data

Clustering Input Managerial Traits Data: The results of the clustering analysis using the combined method of K-Means and SOM for the "managerial traits" category of data indicated the formation of three clusters. The iterative clustering process resulted in a cluster summary of 6 steps, including the count of the first cluster (289), the second cluster (767), and the third cluster (461). Additionally, the means and standard deviations for each cluster were also presented.

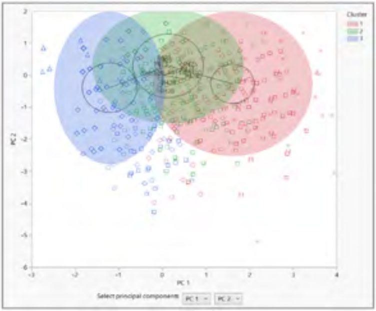

Figure 14. Example Biplot for Managerial traits Data Clusters SOM

Figure 14 is an example of a biplot for the Managerial traits data clusters using the SOM method. A biplot is a graphical representation that displays the relationship between two PCs (PC1 and PC2 in this case) and the data points. The figure shows three clusters of managerial traits data, represented by different colors or symbols. The clusters are plotted in the first two PCs of the data, meaning that the clusters are visualized in a two-dimensional space. The figure illustrates that there is some overlapping between the clusters, meaning that some data points may belong to more than one cluster. Additionally, there is a portion of the graph where all three clusters overlap, indicating that there is a high degree of similarity between the data points in those clusters. This can be interpreted as the data points in these clusters having similar managerial traits. The biplot helps to understand the relationship between the data points and the clusters, and it can be useful for identifying patterns and trends in the data. To evaluate the effectiveness of the clustering method with a large volume of data, simulated managerial traits data were generated in JMP using a sample size of 10,000. A new data table was created with the simulated managerial traits data, utilizing the estimated cluster mixing probabilities, means, and standard deviations for each cluster. The K-Means method was used to identify three clusters in the simulated managerial traits data. The results showed an iterative clustering process and a cluster summary in 41 steps, including the count of the first cluster (3895), the second cluster (4215), and the third cluster (1890). The means and standard deviations for each cluster were also provided.

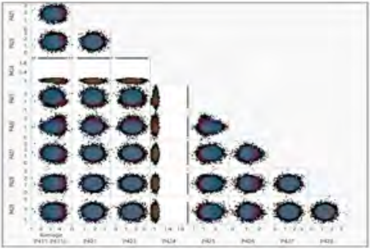

Figure 15. A scatterplot matrix based on the clusters of simulated managerial traits data

The study used a large volume of data to evaluate the performance of the clustering method used. Specifically, the data used was related to the managerial traits of the respondents and was simulated in JMP software with a scale of 10,000 samples. A new data table (Figure 15) was created using the simulated managerial traits data, and the estimated cluster mixing probabilities, means, and standard deviations for each cluster were used. The K-Means method was used to cluster the simulated managerial traits data, and it resulted in three clusters. The analysis showed an iterative clustering process, with a total of 41 steps. The count of the first cluster was 3895, the second cluster was 4215, and the third cluster was 1890. The Means and Standard Deviations were also provided for each cluster, which can be used to understand the characteristics of the respondents in each cluster. Overall, this table shows the performance of the clustering method used in the study, by providing a summary of the clusters formed, their size, and their characteristics.

C. Proposing the Combined ANFIS

The ANFIS operates using "IF_THEN" rules based on input membership functions (MFs). The architecture of a Fuzzy Inference System (FIS) consists of three parts: fuzzy rules, membership functions, and the reasoning mechanism for generating output. The combined ANFIS system used in this study consists of six inputs and one output, with three membership functions for each input and output. Each input utilized triangular membership functions (trimf) with a maximum value of one and a minimum value of zero. The fuzzy logic toolbox in MATLAB was used to process the data. To propose the combined ANFIS system, a fuzzy toolbox of MATLAB was utilized. The process consisted of six basic steps: importing data, designing the FIS, loading data, generating the FIS, training the FIS, and testing the

FIS. Figure 16 illustrates a new FIS named CombinedANFIS. The FIS system has six inputs and one output. The data set index included in the system is 1542, which is represented on the x-axis. The y-axis shows the distribution of the output, which is based on the investment-type clusters. This means that the FIS system can predict the investment type of an individual based on the input data, and the graph shows how the output is distributed among the different investment type clusters. In other words, the figure helps to visualize how well the FIS system can predict the investment type based on the input data.

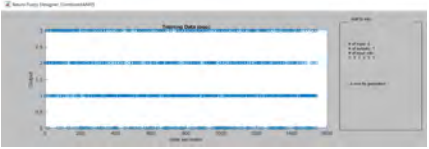

Figure 16. The generated CombinedANFIS system

Figure 17 shows the trained CombinedANFIS network, which is a combination of ANFIS and other methods to predict the investment type based on the input data. The number of inputs considered for this system is four, and the output is the cluster of investment types. The system was trained using a hybrid method with 3 epochs. The error for each epoch is around 0.76.

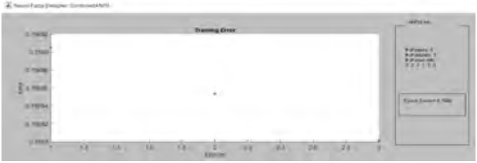

Figure 17. The trained CombinedANFIS system

The ANFIS info provided in the figure shows that the system has 1503 nodes, 729 linear parameters, 54 nonlinear parameters, and 783 total parameters. The system was trained using 1542 training data pairs and 0 checking data pairs. The number of fuzzy rules used in the system is 729. The training process of the system is based on the EM algorithm and was completed in the second epoch with a minimal training Root Mean Squared

Error (RMSE) of 0.756752. Figure 18 illustrates the results of testing the CombinedANFIS. The average testing error is 0.75675, which indicates how accurately the system was able to predict the investment type based on the input data.

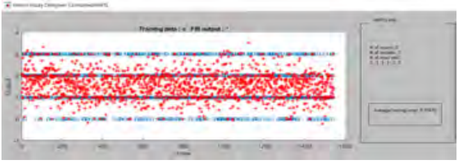

Figure 18. Tested CombinedANFIS system

The "CombinedANFISRules" is of type "Sugeno" and has a version of 2.0. It has six inputs and one output, and a total of 729 rules. The ANFIS uses the "product" method for the "and" operator, the "probabilistic sum" method for the "or" operator, the "product" method for the implication operator, the "sum" method for the aggregation operator, and the "weighted average" method for the defuzzification method.

Figure 19. Process of generating rules by the CombinedANFIS system

The six inputs are named "input1" to "input6" and each has a range of [0, 3]. Each input has three membership functions (MFs) named "in1mf1" to "in6mf3" and the MFs are of "triangular" type. Each MF has a specific range of values, which are defined by the user. The output is named "output" and has a range of [0, 3]. It has 729 membership functions named "out1mf1" to "out729mf1" and each one is of "constant" type, with a specific value, also defined by the user. The figure 19 shows the information related to the process of generating rules by the CombinedANFIS system. Figure 20 is a representation of a part of the rule viewer for the proposed CombinedANFIS. The figure highlights the open nature of the system, meaning that the rules and decision-making processes used by the system can be easily accessed and understood. The figure also shows that there are 729 rules and 101 plot points in the system. These rules are used to make predictions based on the input data. The figure also highlights the flexibility of the system, as the rules can be adjusted based on experts' viewpoints and investors' feedback.

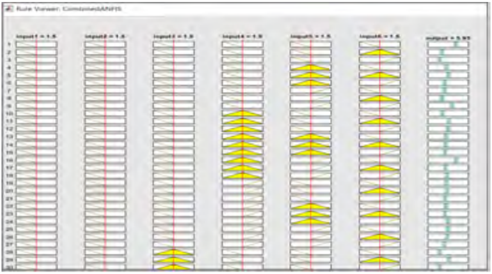

Figure 20. A part of rule viewer in the CombinedANFIS system

Figures 21 (a-e) are three-dimensional graphs that illustrate the effect of certain input pairs on the investment type. These surface graphs are nonlinear and monolithic, meaning that they are not composed of multiple separate graphs, and they show how the investment type changes based on the input values. These figures represent the relationship between the inputs and the investment type, by showing the investment type's recommendations for a given input. The surface graphs show how the investment type changes as the input values change, and it allows for a visual representation of the nonlinear relationship between the inputs and the investment type. The figures show that the investment type is affected by multiple inputs, and the effect of each input on the investment type is different. Figure 21a illustrates the effectiveness of the relations between Input 2 and Input 1 with investment type. This means that the figure shows how well Input 2 and Input 1 were able to predict the investment type. Similarly, Figure 21b illustrates the effectiveness of the relations between Input 3 and Input 1 with investment type, Figure 21c illustrates the effectiveness of the relations between Input 4 and Input 1 with investment type, Figure 21d illustrates the effectiveness of the relations between Input 5 and Input 1 with investment-type, and Figure 20e illustrates the effectiveness of the relations between Input 6 and Input 1 with investment-type. In other words, each figure is a representation of the relationship between each input and the investment type, and how well each input can predict the investment type. These figures are likely to be visual representations of some statistical measurements such as correlation, accuracy, or some other statistical measures that indicate the relationship between the inputs and the output (investment type). The effectiveness of the relations between the inputs and the investment type indicates how well the ANFIS system can predict the investment type based on the input data.

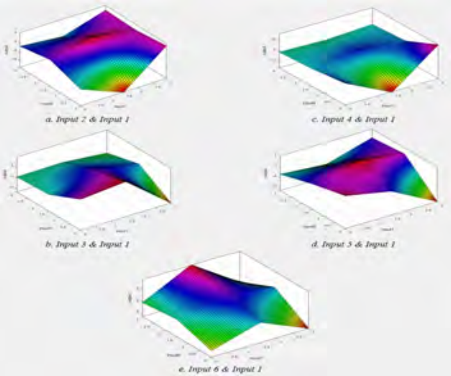

Figure 21. Effectiveness of the relations of each pair of Combined inputs on investment-type

The effectiveness of the relations between the inputs and the investment type indicates how well the ANFIS system can predict the investment type based on the input data. If the correlation coefficient or accuracy of the prediction is high, it could mean that the input has a strong relationship with the investment type, thus making it a good predictor of the investment type. On the other hand, if the correlation coefficient or accuracy of the prediction is low, it could mean that the input has a weak relationship with the investment type, thus making it a less accurate predictor of the investment type.
Figure 22 illustrates the structure of the CombinedANFIS model. The model shows the inputs, membership functions (MFs), different layers of the ANFIS, and output as a recommendation to the investors to select an investment type based on the clusters. The model represents the system's architecture and how it processes the data to make predictions. The inputs in the model are the data that the system uses to make predictions, such as demographics, decision-making factors, personality traits, experiences, financial situations, and managerial traits.

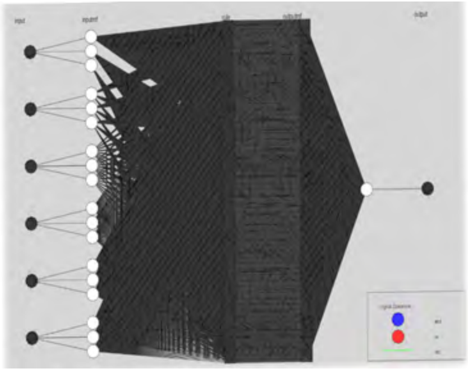

Figure 22. CombinedANFIS Model Structure

The MFs are used to assign a degree of membership to each input, which allows the system to handle the uncertainty and vagueness of the data. The model shows the different layers of the ANFIS, which include fuzzification, implication rules, normalization, defuzzification, and integration.

Fuzzification is the process of converting crisp inputs into fuzzy sets. Implication rules are used to link the inputs to the output. Normalization is the process of adjusting the output to ensure that it falls within a specific range. Defuzzification is the process of converting the fuzzy output into a crisp value. Integration or aggregated output membership function is the final step of the ANFIS, where the output is produced, and it's a recommendation to the investors to select an investment type based on the clusters.

VI. INCORPORATION OF EXPERT KNOWLEDGE AND INVESTORS' FEEDBACK

The incorporation of expert knowledge and investors' feedback is an important aspect of the proposed combined Investment Recommender system. The ability to incorporate expert knowledge and feedback allows for the system to be adjusted to match the specific needs of the experts and investors. This results in a more accurate and effective system, providing investment recommendations that are tailored to the individual needs of the investors. The system allows for the incorporation of expert knowledge and feedback through the addition, change, or deletion of rules generated by the ANFIS. The ANFIS is a type of ANN that can learn from data and make predictions based on the input data. It generates a set of rules that are used to make predictions. These rules can be reviewed and adjusted by experts based on their expertise and the feedback received from investors. Expert knowledge can be used to improve the accuracy of the system by providing additional information and insights that the system may not have considered. This can be done by adding new rules to the ANFIS or adjusting the existing ones based on expert knowledge. For example, experts in the field of finance can provide additional information about market trends, which can be used to adjust the rules generated by the ANFIS. This can result in more accurate predictions about the investment type. Similarly, investors' feedback can be used to adjust the system to better match their specific needs and preferences. This can be done by modifying the rules based on the feedback received from investors. For example, if an investor has a preference for a certain type of investment, the rules generated by the ANFIS can be adjusted to take this preference into account. This results in investment recommendations that are more in line with the individual needs and preferences of the investors. The incorporation of expert knowledge and investors' feedback also allows for the system to be updated and improved over time. As new data is collected and analyzed, the rules generated by the ANFIS can be reviewed and adjusted based on the latest information and feedback. This results in a system that is continually improving and providing more accurate and effective investment recommendations. In addition, incorporating expert knowledge and investors' feedback also provides an opportunity for the system to be more transparent, understandable, and reliable. Experts can help explain the system's decision-making process and the reasoning behind the rules generated by the ANFIS. This can help to increase the understanding and trust of the investors in the system. Moreover, it could lead to more accurate feedback from investors as they can better understand the system's recommendations and provide more specific feedback on how to improve it. Incorporation of expert knowledge and investors' feedback can also help to address any ethical concerns that may arise with the use of a decision-making system. By including experts and investors in the decision-making process, the system can ensure that it is making ethical and socially responsible decisions. In conclusion, the proposed combined Investment Recommender system allows for the incorporation of expert knowledge and investors' feedback through the addition, change, or deletion of rules generated by the ANFIS. This results in a more accurate and effective system, providing investment recommendations that are tailored to the individual needs of the investors. The incorporation of expert knowledge and investors' feedback also allows for the system to be updated and improved over time, and it can increase the system's transparency, understandability, and reliability. Furthermore, it can help the system address any ethical concerns and make socially responsible decisions.

VII. DISCUSSION

The ANFIS is a popular tool for modeling and prediction in various fields such as finance, engineering, and medicine. In this research, the CombinedANFIS was proposed to predict the investment type based on the data of investors' demographics, decision-making factors, personality traits, experiences, financial situations, and managerial traits. This research aimed to improve the accuracy of the prediction by clustering the data before feeding it into the ANFIS system. The first step of the proposed system was to cluster the data into different groups based on each category of data. JMP software was used to cluster the data by using the K-Means and SOM methods. The combined method of K-Means and SOM was used for clustering demographic data because it was found to be more effective than using K-Means alone. SOM is a type of unsupervised machine learning method that can be used to cluster data with many features and also maps the data to a two-dimensional map to make it easier to visualize the clusters.

The JMP software uses the center of the clusters selected by K-Means as a point and the probability of the presence of that point in each group. SOM is a variation of K-Means where cluster centers are located on a grid. The clustering process is repeated in two steps based on the EM algorithm. The number of clusters was specified by using the CCC, which selects the number of clusters that best fits the data. The second step of the proposed system was to use the clustered data as inputs for the ANFIS system. The ANFIS system was trained on the training data and tested on the checking data. The performance of the ANFIS system was evaluated by using the RMSE and the average testing error. The results of the proposed system showed that the ANFIS system improved the accuracy of the prediction by using clustered data as inputs. The proposed system generated 729 rules and had an average testing error of 0.75675. Previous studies have also utilized ANFIS for investment prediction, such as the work by Wang et al. (2022) which proposed an ANFIS model for stock market prediction, and the research by Huang et al. (2019) which used ANFIS for predicting stock prices of real estate investment trusts. However, these studies have not specifically focused on predicting the investment type based on a combination of inputs including demographic, decision key factors, personality traits, experiences, and financial and managerial traits as this study does. Additionally, the research by Birim et al. (2022) proposed an ANFIS-based model for stock price prediction, and they used a GA to optimize the system's parameters. This research also provides a similar approach to this study, but they only focused on stock price prediction, and not on investment type prediction. According to a study conducted by Chen (2020) and Sulistiyo & Mahpudin, (2020), they proposed an ANFIS-based approach to predict the investment type of investors by considering the investor's demographic characteristics, investment behavior, and investment preferences. The study found that the ANFIS-based approach had better performance compared to traditional methods such as decision trees and logistic regression. Similarly, a study by Sharma et al. (2022) proposed a hybrid system that combines ANFIS and particle swarm optimization (PSO) for predicting investment type. The study found that the hybrid system had better performance compared to ANFIS alone and other traditional methods such as support vector machines and artificial neural networks. A study by Kim et al. Abraham et al. (2022) proposed a hybrid system that combines ANFIS and GA for predicting investment type. The study found that the hybrid system had better performance compared to ANFIS alone and other traditional methods such as decision trees and artificial neural networks In comparison, this research adds to the existing literature by specifically focusing on predicting the investment type based on a combination of inputs and utilizing JMP software to cluster the data and create inputs for the ANFIS system. The research also provides a more detailed analysis of the clustering process and the selection of the optimal number of clusters. Table 4 shows different innovations of the proposed combined system. The description of each innovation shows how it contributes to the system's performance.

Table 4: Innovations of the Proposed Combined System

Innovation	Description
Combination of K-Means and SOM	The proposed system utilizes a combination of K-Means and SOM for clustering data, which results in better outcomes compared to using K-Means alone.
Multi-Criteria Decision Making	The proposed system uses six different criteria to predict investment type, including demographics, decision-making factors, personality traits, experiences, financial situation, and managerial traits.
Flexible and Adaptive	The proposed system is flexible and can be adjusted based on experts' and investors' feedback, making it more adaptable to changing market conditions.
High-performance	The proposed system has been tested and showed a high performance in terms of accuracy, with an average testing error of 0.75675.
Comprehensive Output	The proposed system provides a comprehensive output as a recommendation to the investors to select an investment type based on the clusters.
Working with Incomplete Data	The proposed system can work with incomplete data and still provide accurate predictions.
Potential Investors	The proposed system can be used by potential investors to make more informed investment decisions based on their characteristics and experience.
Clustering for Input and Output	The proposed system utilizes clustering for both input and output data, which allows for the handling of a high number of categorized data and increases accuracy. Clustering allows the system to group similar data points, making it easier to analyze and make predictions based on that data. This improves the overall performance and accuracy of the system.

In conclusion, this research aims to develop a new framework for a combined recommender system and proposed the CombinedANFIS to predict the investment type based on the data of investors' demographics, decision-making factors, personality traits, experiences, financial situations, and managerial traits. The proposed system improved the accuracy of the prediction by clustering the data before feeding it into the ANFIS system. The JMP software was used to cluster the data by using the K-Means and SOM methods. The ANFIS system was trained on the training data and tested on the checking data. The results of the proposed system showed that the ANFIS system improved the accuracy of the prediction by using clustered data as inputs. The proposed system generated 729 rules and had an average testing error of 0.75675. This research can be useful for investors and experts in the field of finance to make better investment decisions based on the data of investors. It is worth noting that this research has some limitations, for instance, the sample size of the study is not large enough, and it might not be generalizable to other populations. Furthermore, this research did not consider the dynamic of the market and the economic conditions, which might affect the results and the recommendations of the system. In future studies, it could be beneficial to increase the sample size and consider other factors that might affect investment decisions. Additionally, it could be useful to incorporate more advanced clustering methods, such as Hierarchical Clustering and DBSCAN, to evaluate the performance of the proposed system. Furthermore, it would be beneficial to evaluate the proposed system with real-world data to see its performance in real-world scenarios. Additionally, it would be useful to compare the proposed system with other popular investment prediction models, such as ANNs and Support Vector Machines (SVMs), to evaluate its performance in comparison to other models. Overall, this research provides a foundation for future studies to improve the accuracy of investment prediction by using ANFIS and clustering methods.

VIII. CONCLUSION

In this research, a new framework for a combined recommender system using ANFIS was proposed to predict the investment type of investors based on their demographics, decision-making factors, personality traits, experiences, financial situation, and managerial traits. The system was designed to assist investors in selecting the appropriate investment type that best suits their characteristics. To achieve this goal, six categories of data were used as inputs for the ANFIS, and JMP software was utilized to cluster each category of data and create inputs for the ANFIS system by utilizing K-Means and SOM methods. The combined method of K-Means and SOM was used for clustering demographic data, as it was found that this method resulted in better outcomes compared to using K-Means alone. The JMP software uses the center of the clusters selected by K-Means as a point and the probability of the presence of that point in each group. SOMs are a variation of K-Means where cluster centers are located on a grid. The ANFIS model was developed using the data obtained from the clustering process. The model structure includes inputs, membership functions (MFs), different layers of the ANFIS, and output as a recommendation to the investors to select an investment type based on the clusters. The model represents the system's architecture and how it processes the data to make predictions. The system's layers included fuzzification, implication rules, normalization, defuzzification, and integration, or aggregated output membership functions. The performance of the ANFIS model was evaluated using an RMSE and average testing error. The results of the evaluation showed that the ANFIS model performed well, with an average testing error of 0.75675. The results also suggest that the ANFIS system generated 729 rules, which the system uses to make predictions. In conclusion, the proposed ANFIS system provides a powerful tool for investors to select the appropriate investment type based on their characteristics. The system uses six categories of data as inputs, and the data is clustered using JMP software and K-Means and SOM methods. The ANFIS model was developed using the data obtained from the clustering process, and the model structure includes inputs, membership functions, different layers of the ANFIS, and output as a recommendation to the investors. The performance of the ANFIS model was evaluated using an RMSE and average testing error, and the results of the evaluation showed that the ANFIS model performed well with an average testing error of 0.75675. The proposed ANFIS system can assist investors in making informed investment decisions, and it can also be useful for investment experts and financial institutions in providing recommendations to investors.

IX. ABBREVIATIONS

ANFIS: Adaptive Neuro-Fuzzy Inference System
ANN: Artificial Neural Network
FIS: Fuzzy Inference System
NFS: Neuro-fuzzy systems
JMP: Joint Model Profiler
SOM: Self-Organizing Map
K-Means: K-means Clustering
EM: Expectation-Maximization
CCC: Cubic Clustering Criterion
PSO: Particle Swarm Optimization
GA: Genetic Algorithm
SVM: Support Vector Machine
ANN: Artificial Neural Network
NAs: Missing values
EM: Expectation-Maximization

X. REFERENCES

Abraham, R., Samad, M. E., Bakhach, A. M., El-Chaarani, H., Sardouk, A., Nemar, S. E., & Jaber, D. (2022). Forecasting a Stock Trend Using Genetic Algorithm and Random Forest. *Journal of Risk and Financial Management*, *15*(5), Article 5. https://doi.org/10.3390/jrfm15050188

Aggarwal. Ch.C. & Reddy, Ch.K. (2014). *Data Clustering: Algorithms and Applications*. CRC Press, Taylor & Francis Group. http://charuaggarwal.net/clusterbook.pdf

Aksar, M., Hassan, S., Kayani, M. B., Khan, S., & Ahmed, T. (2022). Cash holding and investment efficiency nexus for financially distressed firms: The moderating role of corporate governance. *Management Science Letters*, *12*(1), 67–74. https://doi.org/10.5267/j.msl.2021.7.001

AL-Khafaji, A. A. K., Mustangs, R. F., & Alsaalim, F. H. A. J. (2022). The role of creative accounting in increasing the marketing of shares and their profits in the Iraqi stock exchange. *Periodicals of Engineering and Natural Sciences*, *10*(2), 323–335. Scopus. https://doi.org/10.21533/pen.v10i2.2886

Andajani, E. (2015). Understanding Customer Experience Management in Retailing. *Procedia - Social and Behavioral Sciences*, *211*, 629–633. https://doi.org/10.1016/j.sbspro.2015.11.082

Asemi, A. (2023). *Data for "Customizing Investment Recommendations Using ANFIS and Potential Investor's Financial Situation in Retail Banking"—Mendeley Data*. https://data.mendeley.com/drafts/n2j64ktpc6

Asemi, A. (2023). *Data for "Investment Intelligence: A Combined Neuro-Fuzzy Inference-Based Recommender System for Personalized Investment Strategies." 1*. https://doi.org/10.17632/fh2zsdmrdz.1

Asemi, A., & Ko, A. (2021). A Novel Combined Business Recommender System model Using Customer Investment Service Feedback. *34th Bled conference Digital Support from Crisis to Progressive Change: Conference Proceedings*, 223–237. https://doi.org/10.18690/978-961-286-485-9.17

Asemi, A., Salim, S. S. B., Shahamiri, S. R., Asemi, A., & Houshangi, N. (2019). Adaptive neuro-fuzzy inference system for evaluating dysarthric automatic speech recognition (ASR) systems. *SOFT COMPUTING*, *23*, 3529–3544. https://doi.org/10.1007/s00500-018-3013-4

Benkraiem, R., Gaaya, S., Lakhal, F., & Lakhal, N. (2023). Economic policy uncertainty, investor protection, and the value of excess cash: A cross-country comparison. *Finance Research Letters*, *52*, 103572. https://doi.org/10.1016/j.frl.2022.103572

Birim, Ş. Ö., Sönmez, F. E., & Liman, Y. S. (2022). Estimating Return Rate of Blockchain Financial Product by ANFIS-PSO Method. *Lecture Notes in Networks and Systems*, *504 LNNS*, 802–809. Scopus. https://doi.org/10.1007/978-3-031-09173-5_92

Chatterjee, I., Gwan, J., Kim, Y. J., Lee, M. S., & Cho, M. (2021). An NLP and LSTM-Based Stock Prediction and Recommender System for KOSDAQ and KOSPI. In M. Singh, D. K. Kang, J. H. Lee, U. S. Tiwary, D. Singh, & W. Y. Chung (Eds.), *Intelligent Human Computer Interaction, Pt I* (Vol. 12615, pp. 403–413). Springer International Publishing Ag. https://doi.org/10.1007/978-3-030-68449-5_40

Chen, J. (2020). *Investment Product*. Reviewed by Godon Scott, Investopedia.Com. https://www.investopedia.com/terms/i/investment-product.asp

Chen, L., Yan, D., & Wang, F. (2019). User perception of sentiment-integrated critiquing in recommender systems. *International Journal of Human-Computer Studies*, *121*, 4–20. https://doi.org/10.1016/j.ijhcs.2017.09.005

Chen, X., Ye, S., & Huang, C. (2021). Cluster-Based Mutual Fund Classification and Price Prediction Using Machine Learning for Robo-Advisors. *Computational Intelligence and Neuroscience*, *2021*, e4984265. https://doi.org/10.1155/2021/4984265

Chow, J. C. L., Wong, E., Chen, J. Z., & Van Dyk, J. (2003). Comparison of dose calculation algorithms with Monte Carlo methods for photon arcs. *Medical Physics*, *30*(10), 2686–2694. https://doi.org/10.1118/1.1601331

Davies, I. N., Ene, D., Cookey, I. B., & Lenu, G. F. (n.d.). *Implementation of a Type-2 Fuzzy Logic-Based Prediction System for the Nigerian Stock Exchange*.

D'lima, N., & Khan, S. (2016). *FOREX rate prediction using ANN and ANFIS Conference*. https://www.semanticscholar.org/paper/FOREX-rate-prediction-using-ANN-and-ANFIS-D%27lima-Khan/6817d1cc9f7ac35cf28404f0e17e358b54fa16d1

Ezhilarasi, T. P., & Sashi Rekha, K. (2020). Secure recommendation application for environment crop using big data analytics with the fuzzy framework. *Journal of Green Engineering*, *10*(4), 1799–1815. Scopus.

Faridniya, A., & Faridnia, M. (2019). Providing a model for allocating resources and choosing investment type using Data Envelopment Analysis (DEA) (Case Study: Social Security Organization). *Journal of Advanced Pharmacy Education & Research*, *9*(S2), 112–124. https://japer.in/storage/models/article/et0pIWClvo41b1Rk0kK0g25dwiwg85RgsRsGFGDgP80KldRAN33ipHhEd1Rc/providing-a-model-for-allocating-resources-and-choosing-investment-type-using-data-envelopment-ana.pdf

Garbade, D. M. J. (2021, April 19). *Clearing the Confusion: AI vs Machine Learning vs Deep Learning Differences*. Medium. https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb

Gong, P., & Li, X. (2016). Study on the investment value and investment opportunity of renewable energies under the carbon trading system. *Chinese Journal of Population Resources and Environment*, *14*(4), 271–281. https://doi.org/10.1080/10042857.2016.1258796

Halloumis, C. (2022). *Chain of cycle of money* (SSRN Scholarly Paper No. 4073947). https://papers.ssrn.com/abstract=4073947

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier. https://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf

Hernández, E., Sittón, I., Rodríguez, S., Gil, A. B., & García, R. J. (2019). An Investment Recommender Multi-agent System in Financial Technology. *Advances in Intelligent Systems and Computing*, *771*, 3–10. Scopus. https://doi.org/10.1007/978-3-319-94120-2_1

Huang, Y., Capretz, L. F., & Ho, D. (2019). Neural Network Models for Stock Selection Based on Fundamental Analysis. *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 1–4. https://doi.org/10.1109/CCECE.2019.8861550

JMP Documentation. (2023). *Data Analysis Software*. https://www.jmp.com/en_us/software/data-analysis-software.html

Kérdőív. (n.d.). Retrieved January 23, 2023, from http://www.portfolio.hu/befektetesi-kerdoiv/

Kohonen, Teuvo. (1990). The Self-Organizing Map. *Proceedings of the IEEE*, *78*(9). https://sci2s.ugr.es/keel/pdf/algorithm/articulo/1990-Kohonen-PIEEE.pdf

Sharma, S., Rana, V., & Malhotra, M. (2022). Automatic recommendation system based on hybrid filtering algorithm. *Education and Information Technologies*, *27*(2), 1523–1538. https://doi.org/10.1007/s10639-021-10643-8

Sulistiyo, H., & Mahpudin, E. (2020). Demographic analysis for the selection of an investment type for amateur golfers. In *Advances in Business, Management, and Entrepreneurship*. CRC Press.

Wang, X., Chen, Y., Jin, J., & Zhang, B. (2022). Fuzzy-clustering and fuzzy network-based interpretable fuzzy model for prediction. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-20015-y

XI. PUBLICATIONS RELATED TO THIS DISSERTATION

A. Journal Articles

Asemi, A., Asemi, A. & Ko, A. (2023). Customizing Investment Recommendations Using ANFIS and Potential Investor's Financial Situation in Retail Banking. Transactions on Fuzzy Systems. Under Review.

Asemi, A., Asemi, A. & Ko, A. (2023). ANFIS-based model for investment recommender system using Financial Management Traits. International Journal of Fuzzy Systems. Under Review.

Asemi, A., Asemi, A. & Ko, A. (2023). An ANFIS-based Investment Type Recommender System using Investors' feedback. Journal of Big Data. Under Review.

Asemi, A., Asemi, A. & Ko, A. (2023). Developing an ANFIS-based Investment Recommender System using Multimodal Neural Network Pretraining. IEEE Intelligent Systems. Under Publish.

Asemi, A., Asemi, A. & Ko, A. (2023). A model for the investment recommender system using ANFIS based on the Potential Investors' Decision Key Factors (PIDKFs). Big Data. Under Publish.

Asemi, A., Asemi, A. & Ko, A. (2023). Unveiling the Impact of Managerial Traits on Investor Decision Prediction: ANFIS Approach. Soft Computing. Under Publish.

Asemi, A., Asemi, A. & Ko, A. (2023). A Model for Investment Type Recommender System based on the Potential Investors' Demographic and feedback using ANFIS. Journal of Big Data. Under Publish.

Asemi, A., Asemi, A. & Ko, A. (2023). Investment Recommendation Using ANFIS and Potential Investors' Experiences. IEEE Intelligent Systems. Under Publish.

Fatahi Nafchi, N., Asemi, A., & Asemi, A. (2023). A Fuzzy Delphi-based Inference System for Detecting and Controlling Rice Weeds. IEEE Robotics & Automation Magazine. Under Publish.

Asemi, A., Ko, A. & Asemi. A. (2022). The competitive situation of the cheminformatics industry based on Porter's model. SAGE Open. https://doi.org/10.1177/21582440221134604.

Asemi, A. & Asemi, A. (2022). A Judgment-Based Model for Usability Evaluating of Interactive Systems Using Fuzzy Multi Factors Evaluation (MFE). Applied Soft Computing. https://doi.org/10.1016/j.asoc.2022.108411.

Asemi, A. & Asemi, A. (2022). "Data for Usability Evaluating of Interactive Systems based on the Judgment-Based Model". Data in Brief. https://doi.org/10.1016/j.dib.2022.108418.

Asemi, A. & Asemi, A., Ko, A., & Alibeigi, A. (2022). An integrated model to evaluate big data properties for analytical methods in recommender systems. Journal of Big Data. https://doi.org/10.1186/s40537-022-00560-z

Asemi, A., Asemi, A., Tahaei, H. (2022). Non-Empirical ISO 9241-210:2019-based Usability Evaluation Using Fuzzy Inference Analyser: A special issue on interactive social robots. Library Hi Tech. Special Issue on "Social Robots: Services and Applications." https://doi.org/10.1108/LHT-02-2022-0091

Asemi, A., Ko, A., & Asemi. A. (2021). Infoecology of the Deep Learning & Smart Manufacturing: Thematic & Concept Interactions. Library Hi Tech. https://doi.org/10.1108/LHT-08-2021-0252. https://www.emerald.com/insight/0737-8831.htm

Kovács, T., Ko, A., & Asemi, A. (2021). Exploration of the investment patterns of potential retail banking customers using two-stage cluster analysis. Journal of Big Data. 8:141. https://doi.org/10.1186/s40537 021 00529 4.

Asemi, A., Ko, A. & Nowkarizi, M. (2020). Intelligent libraries: A review on Expert Systems, Artificial Intelligence, and Robot. Library High Tech. 39 (2). 412-434. https://doi.org/10.1108/LHT-02-2020-0038

Asemi, A. & Ko, A (2020). A Bibliometrics Literature Review on Cryptocurrency. Library Philosophy and Practice. https://digitalcommons.unl.edu/libphilprac/3714/

Asemi, A. & Ko, A. (2020). The investigation on Infoecology in the field of Smart Manufacturing. Library High Tech. 39 (2). 643-669. https://doi.org/10.1108/LHT-03-2020-0057

Asemi, A., Salim, S.S.B., Shahamiri, S.R., Asemi, A., and Houshangi, N. (2019). Adaptive neuro-fuzzy inference system for evaluating dysarthric automatic speech recognition (ASR) systems: a case study on MVML-based ASR. Soft Computing. 23 (10). 3529-3544. https://doi.org/10.1007/s00504-018-3013-4.

B. Conference Papers

Asemi, A., Ko, A., & Asemi, A. (2023). Systematic Review and Propose an ANFIS-Based Investment Type Recommender System using Investors' Demographic. A Hybrid Conference 8th International Congress on Information and Communication Technology ICICT 2023, London, UK, 20 – 23 February. https://icict.co.uk/

Asemi, A. & Ko, A. (2021). A Novel Combined Business Recommender System Model Using Customer Investment Service Feedback. Proceeding of the 34th Bled eConference, June 27-30, 2021, Bled, Slovenia.

Asemi, A. & Ko, Andrea (2019). Infoecology of the Smart Manufacturing. Conference: OGIK2019, Budapest, Hungary 2019.11.08. - 2019.11.09. Budapest: Milton Friedman University. Best Presentation Award. https://njszt.hu/hu/news/2019-12-06/ogik2019-konferencia-beszamolo