UNIVERSITY of RWANDA

ACE-DS

The African Centre of Excellence in Data Science

MASTER'S DISSERTATION

Modeling household expenditure in Rwanda using Neural Network

---

Author:
Msc student: **Patrick REBERO**,
Registration Number:12113131

Supervisor:
Dr. Joseph NKURUNZIZA,
PHD

A Dissertation submitted in fulfillment of the requirements for the degree of Master of Science in Data Science in Actuarial Science

in the

College of Business and Economics
African Centre of Excellence in Data Science

Declaration

I declare that this dissertation entitled "Modeling household expenditure in Rwanda using Neural Network " is the result of my own work and has not been submitted for any other degree at the University of Rwanda or any other institution.

Date: July 6 ,2021

Approval sheet

This dissertation entitled Modeling household expenditure in Rwanda using Neural Network written and submitted by Patrick REBERO in partial fulfilment of the requirements for the degree of Master of Science in Data Science majoring in_Actuarial science is hereby accepted and approved. The rate of plagiarism tested using Turnitin is 15 % which is less than 20% accepted by the African Centre of Excellence in Data Science (ACE-DS).

Dr. Joseph NKURUNZIZA, PhD

Supervisor

15/8/2021

Dr. Ignace KABANO, PhD

Head of Training

Abstract

The household expenditure patterns and the factors related to them caught the attention of policymakers for social programs. Household consumption is the key component for living standards and a good economy. The objective of this research is to estimate the factors influencing household expenditure patterns concerning household residence area for each household on basis of the social-economic as well as demographic factors. The data was obtained from the Fifth Integrated Household Living Conditions Survey (EICV 5) 2016/2017. The neural network method has been used in the analysis of the data. The results on household characteristics and component expenditure, independent variables play a crucial role in discriminating the household expenditure categories i.e low and high expenditure in Rwanda. The component expenditure analysis on rural and urban residence areas revealed that there is a disparity between rural and urban residential areas. In rural residence area, the household expenditure is significantly estimated by Household assets (durable goods), food and household size while urban residence area, the household expenditure are significantly estimated by Household assets (durable goods), non-food and education cost. When the residence area is taken as an explanatory variable, the order of importance has been changed as follows the Household assets (durable goods), non-food, education cost, food, household size, and Age. All independent variables considered in this study are important in estimating household expenditure.

Keywords: **Household**, **Expenditures**,**Neural network**

Contents

List of Figures

List of Tables

Acknowledgements

The achievement of this works involves a great number of individuals, first I would like to express my sincere gratitude to Almighty God, be praised and hallowed his name, for allowing me to be alive and enable me to carry out this research. I also would like to thank my valued aunt, who plays the parental roles and assistance, late parents, late grandmother and I am thankful to my entire family who don't cease me to encourage me to go ahead in my studies.

I am grateful to the University of Rwanda, College of Business and Economics, African Centre of Excellence in Data Science which plans the research and provides us the different support and I am sincerely grateful to my academic supervisor and lecturers, especially to the management of African Centre of Excellence in Data Science for their wonderful guidance and their instruction which are programmed.

Lastly, I would like to express my special thanks to my colleagues at the University of Rwanda, College of Business and Economics, African Centre of Excellence in Data Science for their great assistance to make me more comfortable with the class's activities.

List of Abbreviations

MINALOC	**M**inistry of **L**ocal Government
MINECOFIN	**M**inistry of **F**inance and Economic Planning
NISR	**N**ational **I**nstitute of **S**tatistics of **R**wanda
GDP	**G**ross **D**omestic **P**roduct
NST1	**N**ational **S**trategies for **T**ransformation
SDGs	**S**ustainable **D**evelopment **G**oals
RWF	**RW**andan **F**ranc
OECD	**O**rganisation for **E**conomic **C**o-operation and **D**evelopment
CDF	**C**ommon **D**evelopment**F**und
VUP	**V**ision 2020 sector**P**rogram
LODA	**L**ocal **A**dministrative Entities **D**evelopment**A**gency
EICV	Integrated Household Living Conditions Survey
NGO	**N**on- **G**overmental **O**rganization
UN	**U**nited **N**ations
WHO	**W**orld **H**ealth **O**rgnaization
AI	**A**rtificial **I**ntelligence
NN	**N**eural **N**etworks
MoH	**M**inistry **of** **H**ealth
CBHI	**C**ommunity **B**ased **H**ealth **I**nsurance
UHC	**U**niversal **H**ealth **C**overage
EDPRS	**E**conomic **D**evelopment and **P**overty **R**eduction **S**trategies
ROC	**R**eceiver **O**perating **C**urve
USD	**U**nited **S**tate **D**ollar
Bn	**B**illion
H_0	**N**ull Hypothesis

Chapter 1

Introduction

In 2018, the household expenditure in the world was 48,610.6 Bn USD. Over the 48 years, this indicator's value fluctuated between 48,610.6 Bn USD in 2018 and 1,739.55Bn USD in 1970 (World Bank,2018). In Africa, Household expenditure is changing over time. It grew at a 3.9 percent compound rate since 2010 and reached 1.4 trillion USD in 2015 (Signe, 2018).

For more than 10 years, Rwanda experienced positive economic growth and social performance. The economic growth averaged 7.5 for the decade up to 2018 while the Gross Domestic Product (GDP) per capita grew at 5 percent annually (World Bank, 2018).

Now, the ambition of the country is to be a middle-Income country and a High-income country by 2035 and 2050 respectively. The way to achieve the above objectives is well defined in the National Strategies for Transformation (NST1) under the seven-year government program with a target to reinforce the sectoral strategies that are in line with Sustainable Development Goals (SDGs). Before the NST1, there were two short-run programs called Economic Development and Poverty Reduction Strategies EDPRS (2008-12) and EDPRS-2 (2013-18) respectively (MINECOFIN, 2017).

The Rwanda government effort in line with the vision of the country is to eradicate extreme poverty, malnutrition and promote socio-economic transformation by accelerating graduation from extreme poverty to household resilience. Through the social protection policies, intensive training on how to use the financial service as well as to provide the works for citizens. So that the benefits of economic growth are also taking into consideration the reduction of all forms of poverty (MINALOC, 2018).

In 2018, the household expenditure and NGOs were 6,347 billion Rwf, where the GDP is valued at 8,189 billion Rwf at current market prices, up from 5,737 billion Rwf where the GDP is valued at 7,600 billion Rwf in 2017 (NISR, 2018).

Commodities spending is a key component of economic prosperity and, as such, a primary measure of living standards. Wealth and income are available to support consumption, now and in the future (through the savings generated by earnings). The three dimensions of the broader concept of economic well-being are described by Income, consumption, and wealth, and understanding the relationship between them is crucial. Whatever else being equal, it is assumed that a person with a lower consumption level is associated with the economic well-being of a lower level rather than a person with a higher consumption level (Zivar, 2020). Factors such as electricity, cooking materials, rural and urban area, transport services, and so on are important for Household consumption. The consumption needs can be matched through income spending, through the spending of wealth, and via borrowing (OECD, 2013).

Household expenditure patterns are mainly determined by household income. The disparities between expenditure patterns are the implication of the disparities in income between the group and individual households. The expenditure patterns are influenced by many factors in respect of preference on one hand and income levels, on the other hand, the two components are affecting the expenditure pattern. It is not principle that the expenditure pattern of households differs prominently, especially in controlling income (Punt, et al., 2003).

The Rwandan government established many programs to support its citizens from extreme poverty through household resilience such as the Vision 2020 sector Program (VUP), it was established in 2008 under the Ministry of Local Government with Common Development Fund (CDF). The VUP's purpose is to make Rwanda a contribution effort to eradicate extreme poverty and strengthen household resilience (LODA, 2017).

According to the National Institute of Statistics of Rwanda (NISR), from The Fifth Integrated Household Living Conditions Survey (EICV 5) 2016/2017, the head poverty rate is obtained by making a comparison about real consumption per adult; the poverty line and the extreme poverty line were 159,375 RWF and 105,604 RWF per year respectively, this consumption per adult is used as a welfare measure. The EICV 5 survey showed that poverty and extreme poverty are 38.2% and 16 % respectively while the EICV 4 revealed that poverty and extreme poverty were 39.1% and 16.3 % respectively. The median real consumption per adult for 2014 and 2017 are respectively 187 000 RWF and 191000 RWF (NISR, 2018).

In this research, generalized weights of the neural network alongside with neural network will be applied for the estimation of significant factors influenced by household expenditure patterns in Rwanda.

1.1 Problem statement

Consumption is a prominent topic in economic literature since it is one of the primary sources of consumer welfare and value. In Rwanda, It accounts for around 78 % of Rwanda's Gross Domestic Product (NISR,2018).

The characteristics, quantities, acts, qualities, and tendencies that characterize a community's use of resources for survival, comfort, and enjoyment are represented by the household consumption expenditure trends of a country. As a result, the need to research consumption stems from the fact that consumption, along with other factors, has a significant impact on the economy. Other determinants include equitable distribution of wealth and income, as well as the availability of goods and services.

Therefore, The household expenditure offering a significant measure for cross-country comparisons, particularly in terms of well-being(OECD,2013).

In this study, We will estimate household expenditures for each residence area based on the factors examined and compare the effectiveness of the Neural Network model with the Multiple Linear Regression model in terms of estimation. This study will give background information on the elements that influence household expenditure in Rwanda. The observable demographic and socio-economic factors, which are linked with poverty alleviations. As a result, while formulating household welfare programs, policymakers should consider an allocation of resources behavior when designing household welfare programs.

The study's goal was to find answers to the following research questions:

- What are the factors that are significantly influencing the household expenditure level?
- Which do the independent variables influence for each household residence area, in estimating household expenditure patterns?
- Is there a comparison between neural network and regression models in terms of estimation?

1.2 General objective of the study

Estimation of factors influencing household expenditure patterns concerning household residence area for each household on basis of the social-economic as well as demographic factors considered.

1.3 Specific objectives of the study

- Analyze the influence of demographic and socio-economic factors on household expenditure levels.
- Investigate the influence of the independent variable's effect for each household residence area, for estimating household expenditure patterns.
- Compare neural network model with regression model in terms of estimation.

1.4 Research Hypotheses

The thesis is to test the null hypothesis that, based on the following considerations on household expenditure in Rwanda :

- H_0: There is no significant influence of demographic and socio-economic factors on household expenditure level?
- H_0: There is no influence of the independent variable's effect for each household residence area, for estimating household expenditure patterns?
- H_0: There is no difference between neural network and regression models in terms of estimation?

Chapter 2

Literature Review

The expenditure in the household is influenced by the level of income in the household, which is the main factor determining individual or household spending on consumption. But the income is not the only factor regulating the expenditure; other factors such as loans and savings have contributed to the expenditure behavior. Spending behavior is driven by the requirements to meet the necessities of the individual or the household. In addition, economic uncertainty and price changes that occur have undoubtedly influenced a household's expendable income (Nik et al., 2011).

The willingness to pay increases the elasticity demand, which is varying in respect of the family size, income, loan ratio, and the residential area so the family size played the important role in the estimation of elasticity of income and negatively down payments (Wilhelms son,2002). However, the estimation of the income is still the major issue in researches over a long period, the reason why numerous studies in underdeveloped countries used the expenditure as a proxy for annual income. Therefore, household consumption should be the better approximation of the economic situation, particularly in underdeveloped countries where the annual incomes are not consistent and steady (Azzarri et al., 2006). Further, the expenditure data is a valuable tool for assessing welfare households in developing countries (Koenker and Hallock,2001).

Household consumption is influenced by demographic and economic considerations. Household size, number of children, age of the consumer, marital status, location of residence, education level and occupation, population growth rate, and other demographic characteristics in the country are the most important demographic determinants impacting consumption expenditures(Zivar and Mubariz, 2020). The demand and desire for specific items are determined by the consumer's occupation and degree of education. Households with greater levels of education are considered to consume more fruits and vegetables than others (Mancino et al., 2004).

The household characteristic's consumption is different by regions i.e. urban and rural (Shamim and Ahmad,2007). The expenditure variations depend on socio-economic and demographic factors (Punt, et al.,2003). The household characteristics such as age and education of the household head may have a closer relationship with household spending dispersal (Zehiwot and Senapathy,2019). But the household member characteristics are not much associated with the status of the employment and its household head yields, so the household head was assumed as the one with the highest yields in the household and pretend to be linked with household expenditure (Fabrizio C. and Arthur S.,2013).

The household expenditure determinant on education by the application of multiple linear regression model to establish the relationship between the determinants of household and education expenditure and the research findings were the occupation of household head, education level of household head, the size of household and household income are playing the important role on household expenditure. However, the household head with a higher educational level, their children are more likely to get the high levels of education (Edna M., 2017).

The household expenditure on education helps to monitor, design, and implement different programs, and the education level of household heads matters the most for spending on education and it is linked to the individual level (UN,2015). And the household heads with a higher educational level, their children are more likely to get the high levels of education (Edna M., 2017), as well as households with health insurance and the salaried, are positively related to paying children education (Nkurunziza J et al.,2015).

Thus, the household education expenditure encompasses different spending, such as tuition fees, different education, and other school expenses outside of school fees such as cost of school uniforms and other school's clothes where it is applicable, books, teaching materials, transport fees, learning the software and computers. Of course, those costs are not incurred to households without school-aged children and who are attending classes (World bank,2018).

On the Other, hand, the educational attainment of household heads contributes also to household expenditure on food (Sekhampu et al.,2013). The trends of household spending on food are rising as household income rises (Koenker and Hallock,2001). But the unusual points arise while the high income related to the lower expenditure on food as they are inferior goods (Obaidullah J, 2019). There is no equilibrium between non-food and food items on household budget share. The decrease in food items, the non-food items increase, and observation made is that household is wealthier (Obisesan O et al .,2016). Take into account the household size into demographic categories by age and sex, food and non-food intake were greatly impacted by changes in spending and household size (Shamim and Ahmad ,2007).

The expenditure accompanied with the state of being healthy, thus the health expenditure i.e. out-of-pocket health care costs changing widely across countries, the affordability of health services requiring the payment, the payment capacity which is low and not able to prepayment or the health insurance in certain developing countries (WHO,2003). but in Rwanda, there is CBHI which provides people with universal and fair access to quality health services in a rural area and informal sector, where the vulnerable peoples are integrated with health systems. it supplemented the other existing Health insurance (MoH,2010), this is complying with Sustainable Development Goals (SDGs), one of the high priorities was to provide affordable health care coverage especially for the citizens of countries with low and middle incomes (UHC, 2017).

The durable goods consumption matters, consumptions of household assets regard the price and level of assets, influence positively on the household expenditure and the consumption of assets is of great significance (Hongyun H and Fan S,2020). In a rural economy, the most valuable asset is the land, with land, there is an opportunity for agriculture production for food as well as generating income (Abeeda, 2013).

The multiple linear regression for modeling by explaining social economics factors to responses in monthly expenditure in Analysis of the Factors Influencing Household Expenditure. They revealed that the social economics factors were strongly positive to household expenditure while marital status of head of household contributed negatively to household expenditure (Sekhampu et al. ,2013). The household spend more on food, next cloths and others and it found that the household size is contribute at the greatest extend to household expenditure (Zehiwot Senapathy ,2019). They worked on continuous variables while neural network performs better either on continuous variables or categorical variables with the greater predictions.

The prediction of household expenditure based on household characteristics by application of a neural network of multilayer perceptrons. Data is obtained from Household Integrated Economic Survey (HIES) 2007-08 carried out by the Federal Bureau of Statistics (FBS), Pakistan and they found that all the independent variables played an important role namely household income, household assets, and several household earners per household, sex of household head, region and literacy of household head (Ahmad Z. and Fatima A., 2011). However, this study limited to household characteristics and the residence area did not take into consideration, generalized NN.

The Changing mindset as well as social behavior to better understand household expenditure patterns, especially the poor peoples would be beneficial so that they would be to manage their spending at the household level, planning with complying the current situation as well as to overcome the unwanted of any economic uncertainties in future.

AS the income in developing countries is not steady and consistent, through the expenditure we should evaluate the consumption of goods and services, which is linked to the well-being of the population's countries. So, the research problem here is in line with the estimation of the expenditure at the household level and the influence of factors affecting the trend of households spending in Rwanda.

So far, many pieces of research on household expenditure have been carried out worldwide by using different methods, but the research with the application of the neural networks is still scanty. However, there is no research carried out in Rwanda on household expenditure. In this research the neural network algorithms with the generalized weights neural network for the Rwanda case study will be applied, this research aims to estimate significant factors associated with household expenditure patterns in Rwanda.

Chapter 3

Conceptual Framework

The conceptual framework proposed by the study discussed here, the social-economic, and Demographic factors encompass the independent variables. The Demographic factors are Age of Household Head, Education Level, Residence area, and household size while the social-economic factors are Main occupation, Health insurance, Education cost, Non-Food, Food, and household Asset. Those factors are impacting the expenditure at the household level. Alongside this conceptual framework, an attempt is made to investigate the estimation of the above-mentioned influence.

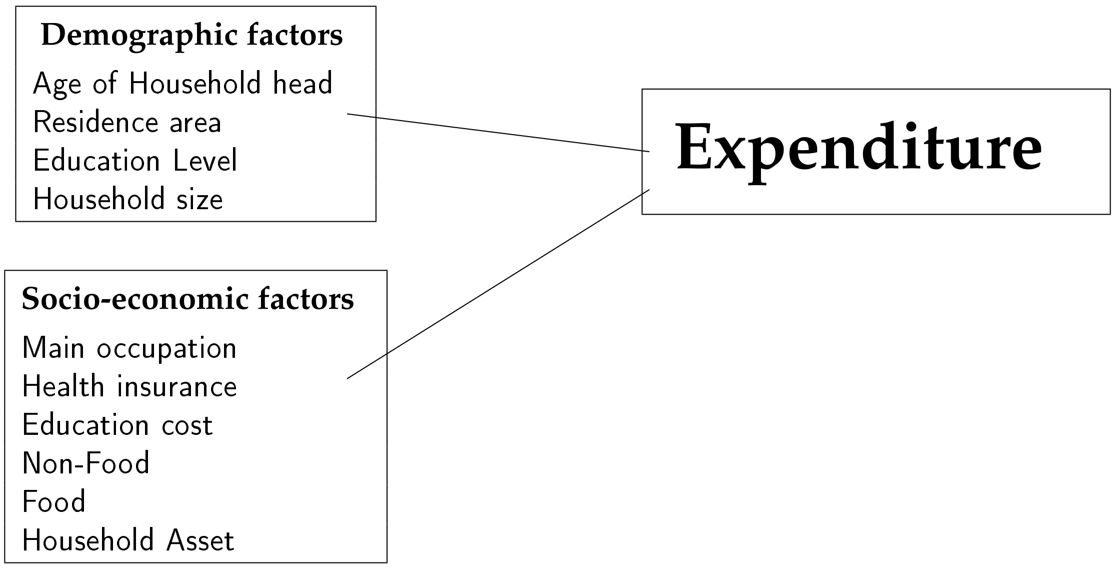


FIGURE 3.1: Conceptual framework

Chapter 4

Data and Methodology

4.1 Data Cycle

The secondary data used in this research are the ones from the Fifth Integrated Household Living Conditions Survey (EICV 5) 2016/2017 collected by the national institute of Statistics of Rwanda (NISR). The EICV covers poverty trends and wide scope of living conditions statistics. The EICV5 which is carried out by the National Institute of Statistics of Rwanda (NISR) is conducted every three years.

The data forms a key component for model creation and learning processes. The data has to be gathered, cleaned, converted, and then fed to the learning model. The data were cleaned up by removing the incomplete case (NA). The features of the data have different ranges that why we normalized the data to change the numerical common values in the data frame to a common scale.

4.2 Data cleaning

A good model and balanced data help the accuracy of the model and algorithms. Thus a data scientist spent time mostly on cleaning up the data before creating the model (Giuseppe, 2017). Checking for the total number of missing values in the entire data frame, we found that there are missing errors in the dataset. Here the process of finding the complete rows of the data frame, the complete. Cases function in R has been applied to identify the complete rows of the entire dataset (original dataset). The original dataset was 14,580 rows of data equivalent to the number of households and the 10,725 rows of data have been selected for household characteristics and household expenditure components, those missing errors appeared in continuous variables while there is no skip logic in the questionnaire.

4.3 Training and testing the model

Training and evaluating the models provide the basis for further use of predictive models in predictive analytics, numerous studies and empirical researches on the best way to split the data, 80/20 or 70/30 split are widely used (Giuseppe and Balaji, 2017). Provided a dataset of 10,725 rows of data, which include the predictive and response variables, the dataset was split into a suitable ratio (70:30) and it was split as follows 70 percent for training and 30 percent for testing.

4.4 Variables

The variables for household characteristics have been composed of the independent variables and dependent variables. The independent variables are the following: residence area which is subdivided into 3 categories residence area, households from Kigali city, households from other urban areas and households from the rural area, the health insurance and it is subdivided into 2 categories such as household members possess health insurance and household without health insurance.

The education level of the household head is made up of 4 categories, household head not attended school, household head who attended from primary 1 to primary 5, household head with primary completed, and household head with more than primary level. The fourth independent variable is the occupation status of the household head which is composed of two categories namely household head with regular occupation and household head with elementary occupation and combined with expenditure components.

The dependent variable for Household characteristics and expenditure components has been subdivided into 2 levels, depending on the median real consumption per adult for 2017, which is 191000 RWF per adult. The dependent variables are categorized into Spending below the median real consumption per adult for 2017 is Low Expenditure, the spending above the median real consumption per adult for 2017 is High Expenditure.

The variables for component expenditure are residence area, Household size, the number of household members in a household, The education cost expenditure which is the spending associated with the education, Non-Food expenditure which is the spending on non-food items such as clothes, shoes, etc., Food is spending related to food, Household assets(durable goods) which are durable household's goods and the dependent variable is household expenditure.

4.5 Neural Network

Machine learning is an Artificial Intelligent (AI) branch that helps computers program based on input data themselves. Machine learning helps AI to do problem-solving based on data. Neural Networks (NNs) are an example of algorithms used in machine learning.

Neural network (NN) training emerged from the idea of imitating the workings of the same human brain that tried to solve the problem; the brain is a complex neuron network that processes information through a multi-neuron system. Knowing the brain functions has always been challenging; but, thanks to advances in computational science, we can now artificially program neural networks (Bishop, 2006).The purpose of the whole neural network is simply to measure the outputs of all the neurons, which is a totally deterministic calculation. Essentially, NN is a set of approximations to mathematical functions.

The linear models for regression and classification are based on linear combinations of fixed nonlinear basis functions $\varphi_j(x)$ and take the form:

y(w,x)=f($\sum_{j=1}^{M} w_j\phi_j(x))f(\sum_{j=1}^{M} w_j\phi_j(x))$, (1) y= the dependent variable, which is expenditure x= the specific independent variables considered in this study, there are 10 independent variables grouped into demographic and socio-economic factors: The Demographic factors are Age of Household Head, Education Level, Residence area, and household size while the social-economic factors are Main occupation, Health insurance, Education cost, Non-Food, Food, and household Asset.

w= weights parameters, each independent variable has its weight, the weights are multiplied with their respective layers and summed up for converting into output.

Where f (.) is a nonlinear activation function in the case of classification and is the identity in the case of regression. Our goal is to extend this model by making the basis on the functions $\varphi_j(x)$depend on parameters and then to allow these parameters to be adjusted, along with the coefficients w_j.

On the other hand, there is the nonlinear combination in machine learning techniques; known as neural networks, can be described as a series of function's transformation. linear combinations of the input variables Residence area, Education Level, main occupation, health insurance, Household size, Education cost, Non-Food, Food and Household Asset $(x_1, ..., x_D)$ in the form.

$a_j = \sum_{i=1}^{D} W_{ji}^{(1)} xia_j$=$\sum_{i=1}^{D} W_{ji}^{(1)} xi + W_{jo}^{(1)} W_{jo}^{(1)}$ (2)

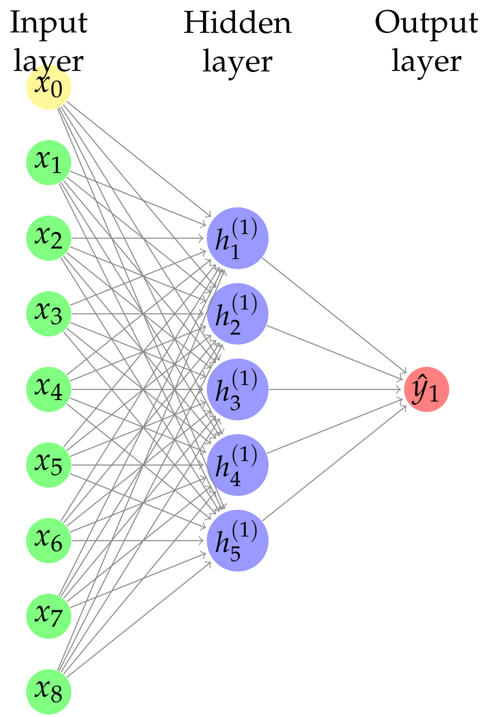
Where a_j= activation, j = 1, . . . , D, indicates that the corresponding parameters are in the first 'layer' of the network. We shall refer to the parameters as weights and the parameters $W_{jo}^1 W_{jo}^1$ as biases.

The activation function for hidden layers will be applied, and take the form of hyperbolic tangent activation: $f(x) = \frac{e^{-x} - e^{-x}}{e^{-x} - e^{-x}}$(3)

Abstracting neural network computation is done primarily through the activation functions. The activation function is a mathematical function that transforms the input into output and brings the magic of processing the neural network. Without activation function, the neural network will be like a linear function (Giuseppe C, 2017).

A neuron's output is a function of the weighted input sum plus the bias. Every neuron conducts a very simple operation involving activation if the total amount of signal obtained exceeds a threshold for activation, as shown in the following figure:

FIGURE 4.1: Neutral network architecture



(Brendan Tierney,2020)

A node layer is a line of those neuron-like switches that turn on or off as the input is fed over the NN. The output of each layer is the input of the subsequent layer at the same time starting from an initial input layer that receives the data. The hidden layer has the magic to turn the input into the desired output (Chris Nicholson, 2015).

Input layers considered for this study are 10, which are 10 independent variables grouped into demographic and socio-economic factors: The Demographic factors are Age of Household Head, Education Level, Residence area, and household size while the social-economic factors are Main occupation, Health insurance, Education cost, Non-Food, Food, and household Asset. They have the function of moving input data to hidden layers, as this layer has no activation function.

The hidden layers are 5 for household characteristics, we use one among the rules of thumb for deciding hidden layers which is the size of the number of hidden layers should be of input layers and output layers size (Heaton J.,2017). The process input layers into output layers and they build a structure multilayered the number of nodes in the hidden layer are connected by the weights, which control the strengths between neurons and determine how many input layers can affect output layers. Hidden layers are normally optimized for on interval (training set), which is approximate and are used for prediction of data set of test data with regarding the architecture chosen.

The output layer is representing a set that collects the identified features and provides the predictable output (Expenditure). According to the investigators, a multi-output neural network yields the results, which is inferior compared to a single output network (Maciel Ballini, 2008).

4.6 Generalized Weights

The Generalized Weights, a method for objects of class NN, generated typically by neural network Plots the generalized weights for one particular covariate and one response variable (Intrator and Intrator, 1993).

$$\varpi_i = \delta log \frac{\frac{0(x)}{1-0(x)}}{\delta x_i} (4)$$

Where i is the index for each covariate, 0(x) is the predicted outcome of the covariate vector. Log-odds is the connecting function of the logistic regression model. The coefficient for that covariate is the partial derivative of the log-odds function to the covariate of concern.

To train a neural network, a supervised learning method was applied and the model was verified based on root-mean-square error (RMSE) and coefficient of determination(R^2).Although NN has proven to have strong predictive power in comparison with traditional models, we apply it in the component of household expenditure with regard to the independent effect of its input variables, which are socio economic and demographic factors i.e Residence area, Household size, education cost, food , non food and Household assets.

NN has the potential to learn and model non-linear and complex interactions, which is very important since many of the relationships between inputs and outputs are both non-linear and complex in actual situations, NN can generalize after learning from initial inputs and their interactions, unseen relationships can also be derived from unseen data, allowing the model more general and predictive of unseen data. Unlike several other prediction methods, NN does not place any constraints on input variables (such as how they should be distributed). In addition, several experiments have shown that NN can better model heteroskedasticity, i.e. data with high uncertainty and non-constant variation, provided their capacity to learn hidden relationships in the data without enforcing any fixed relationships in the data.

4.7 Multiple linear regression (MLR)

To build a model for overall performance changes during preservation, multiple linear regression (MLR) was applied. The generic multiple linear regression model makes use of the following equations : y=$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \epsilon (5)$

Where y is the dependent variable, $\beta_0, \beta_1, \beta_2, ..., \beta_n$ are the regression coefficients, $x_1, x_2, x_3, \ldots, x_n$ are the independent variables and ϵ is error term.

R program and python will be used in data analysis and SPSS 25 for displaying some graphs.

Chapter 5

Results and Discussion

There is a difference in real consumption in rural and urban areas, and because disparity persists in the urban and rural areas based on household consumption (NISR,2018). The socio-economic factors and demographic characteristics are quite different, hence the analysis of household expenditure patterns in rural residence and urban residence areas separately and combined. The four neural network models have been run as follows rural, urban, rural-urban areas along with household characteristics.

The households in rural areas, are engaged in agriculture as the main economic activity of rural households in general, and the dominant vision of rural development has centered on encouraging smallholder agriculture (Ellis and Biggs, 2001; Haggblade, 2007), even if the majority of rural households are engaged in several economic activities but the agricultural still the main activities. Thus, the land is the most valuable asset in rural residential areas and an opportunity for food as well generating income (Abeeda, 2013). Most families in the rural area spend the portion of their income on food, clothes, and also family characteristics i.e. family size and age of household have relationships in rural household consumption behavior patterns (Belete et al., 1999).

In line with this research, the household expenditure in rural residential areas is estimated by household assets which are durable goods, household size, and foods, those variables are on top for estimating expenditure in the rural residential area. On the other hand, changes in family expenditure patterns will inevitably impact lifestyle disparities. Household spending and income inequality can contribute to substantial differences in their expenditure on basic needs in rural and urban areas. Depending on the degree and increase of their income, urban households prefer to increase their consumption to boost their lifestyle (Farkhanda Eatzaz, 2007). the research findings revealed that urban residential areas are estimated by the predominant variables as household assets, Education, and non-food items.

5.1 Household characteristic and expenditures components

Household characteristics are essential factors in deciding household expenditure patterns. household vary in age of Household head, Residence area, Education Level, Main occupation, Health insurance, Household size, Education cost, Non-Food, Food, and Household Assets (durable goods). Households with different characteristics are expected to have different levels of household expenditure.

Figure 5.1 showed that the receiver operating curve (ROC) of two expenditure levels namely low and high expenditure, all of these curves are in the middle to be close to the left corner. This reveals that the model is fit for the given data and the area under the curve for the high expenditure level is 0.858, the area under the curve for the low expenditure level is 0.858.

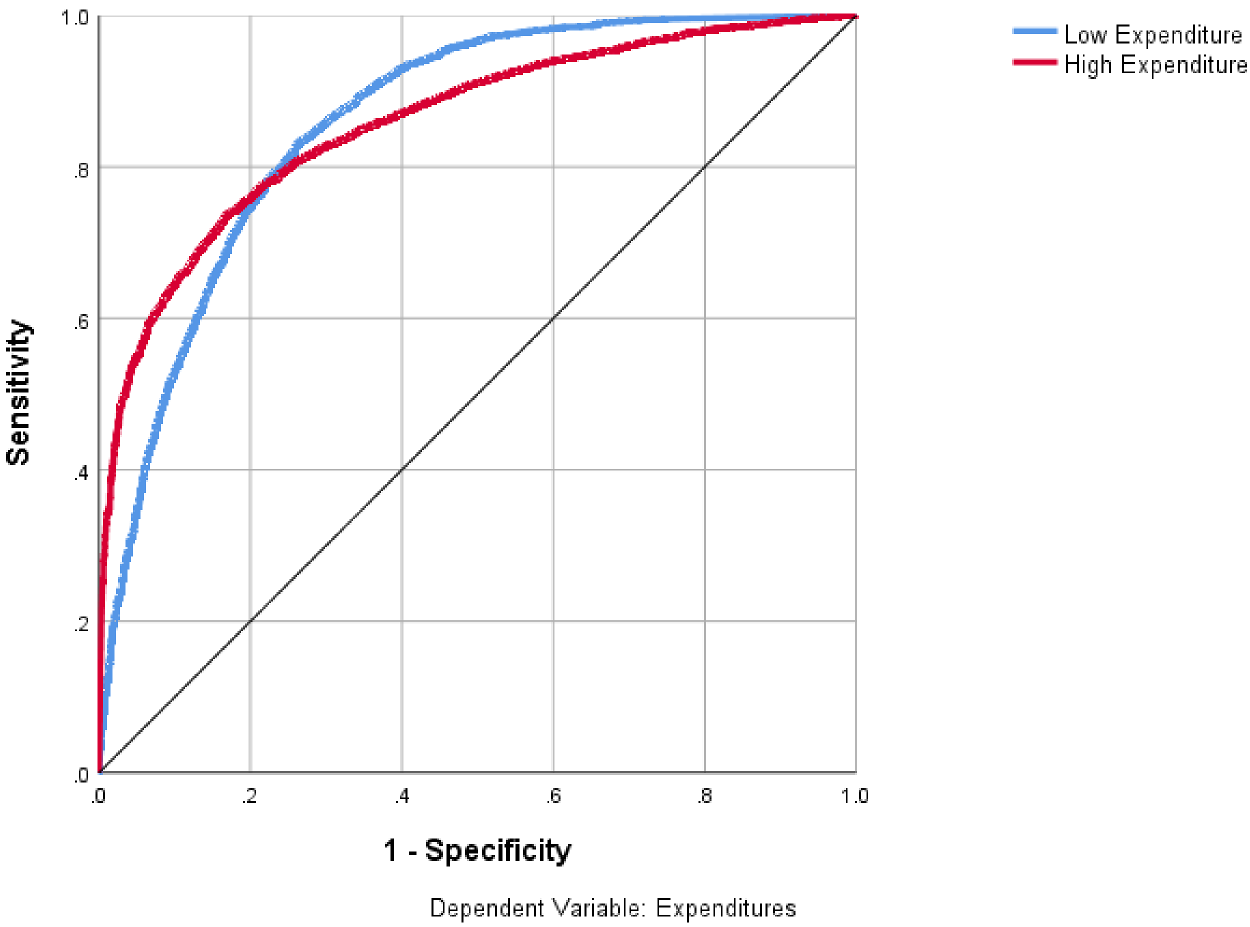


FIGURE 5.1: ROC curve
Source: An analysis based on the EICV5 dataset.

Table 5.1 showed that the most area covered is over 0.8 which indicate that the model is fit, reference made on the rule of thumb is that the accuracy of tests with Area Under the Curve between 0.50 and 0.70 is low, between 0.70 and 0.90, the accuracy is moderate and it is high for Area Under the Curve over 0.90 (Fisher, Bachmann, and Jaeschke, 2003).

Area Under Curve		
		Area
Expenditure	Low Expenditure	0.858
	High Expenditure	0.858

TABLE 5.1: Area under the Curve
Source: An analysis based on the EICV5 dataset

5.2 Component expenditures

5.2.1 Rural area model

In Figures 5.2, 5.4, and 5.6, the grey lines show positive weights while the blues lines show weights, which are negative. In other words, the grey lines indicate the increase in Expenditure while the blue lines indicate the decrease in Expenditure. The thickness of the line is commensurate with the associated magnitude of each weight. The input variables are received by the first layer (input layer), and each is connected to all the hidden layer nodes (H1 to H5). The output layers (Expenditure) are connected by all nodes of the hidden layer. Bias nodes are constant and additional into the next layer, the same as weights and each layer has a bias and they play the role of increasing the flexibility of the model to fit within the given data.

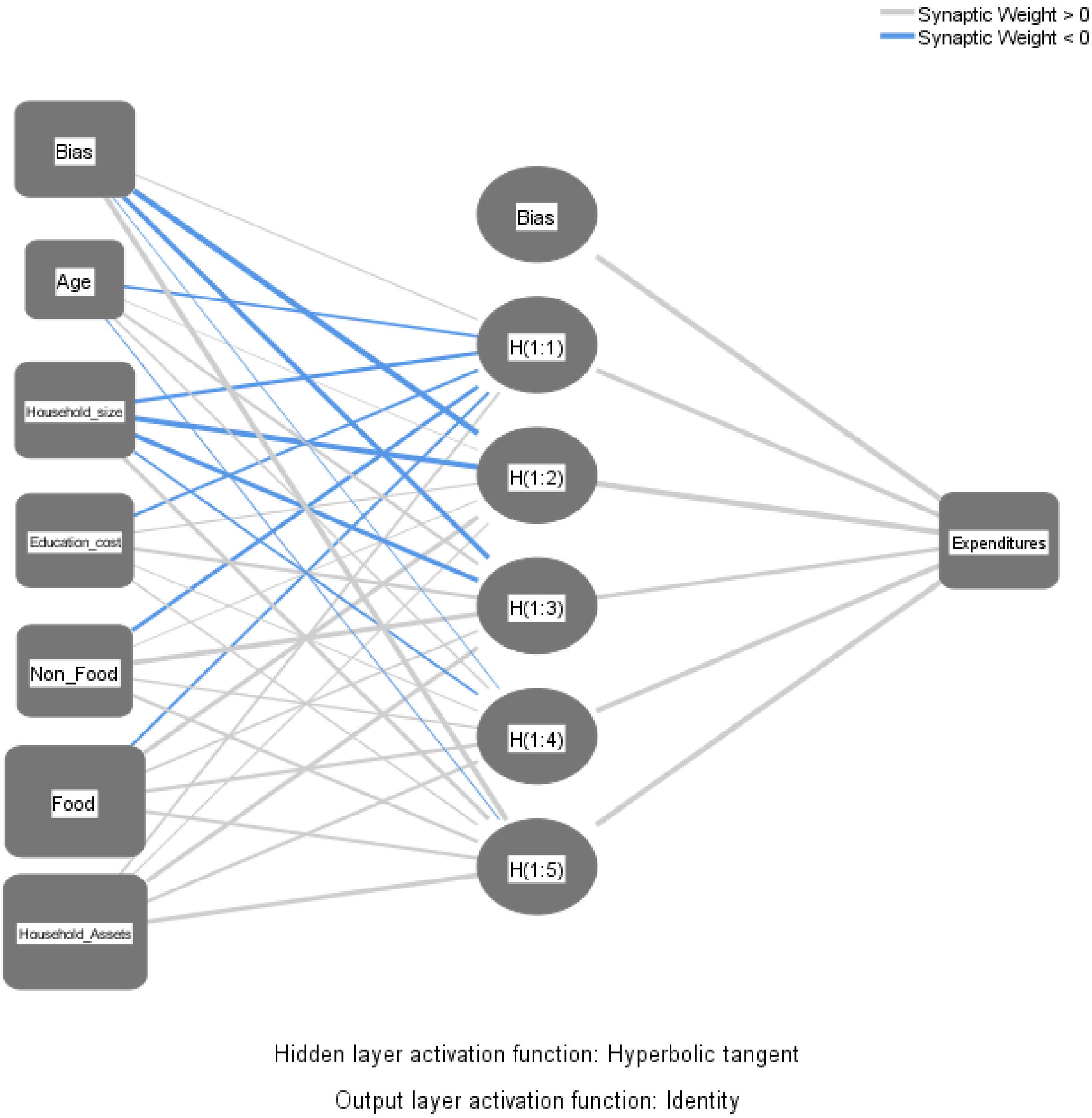


FIGURE 5.2: Neural network model for the rural area
Source: An analysis based on the EICV5 dataset.

In Tables 5.2, 5.3.and 5.4, the sign of weight shows how the weight of the input layer will affect directly the hidden layer and indirectly the expenditure of rural households. Positive weights of input layers will correlate directly to greater values of the hidden layer and vice versa for negative weights and indirectly to the output value.

From Table 5.2. the input layers of Household assets (durable goods) are associated with the positive weights while Household size, Age, Non-Food, Non-Food, and education cost associated with the negative weights to hidden layer node 1. The input layers of Household assets, Non-Food, Food, Age, and education cost are positively related to the hidden layer

Parameter Estimates							
Predictor		Predicted Hidden layer					
		H_1	H_2	H_3	H_4	H_5	Output Layer Expenditure
Input Layer	(Bias)	0.089	-2.337	-0.553	-0.008	1.044	
	Age	-0.249	0.021	0.280	0.139	-0.044	
	Household size	-0.341	-1.312	-0.472	-0.188	0.359	
	Education cost	-0.2541	0.108	0.329	0.072	0.127	
	Non Food	-0.347	0.057	0.656	0.154	0.282	
	Food	-0.250	0.448	0.163	0.341	0.291	
	Household Assets	0.236	0.097	0.441	0.288	0.481	
Hidden Layer	(Bias)						1.472
	H_1						0.6252
	H_2						2.457
	H_3						0.387
	H_4						1.007
	H_5						1.326

TABLE 5.2: neural network weight output for rural area model
Source: An analysis based on the EICV5 dataset

nodes 2, 3, and 4, while the input layer of Household size is negatively related to the hidden layer nodes 2, 3, and 4. The input layer of Food, Household assets (durable goods), Education cost, non-Food, and Household size are positively contributing to Hidden layer node 5 while Age is negatively contributing to Hidden layer node 5.

Table 5.2 revealed that the expenditure has a positive relationship with all the hidden layer nodes i.e 1, 2, 3, 4, and 5. Now, we presume that the order of importance of hidden layers is as follows $H_2 > H_5 > H_4 > H_1 > H_2$. Thus, the second, Fifth, and fourth nodes of the hidden layer are the main contributor to expenditure, where the weights of Household Assets (Durable goods) and food are 0.441,0.481, 0.288, 0.236, 0.097 and 0.448, 0.291, 0.341, -0.250, 0.163 respectively, based on those variables' weights, they are the highest contributor to these hidden layers, in line with the order of hidden layers.

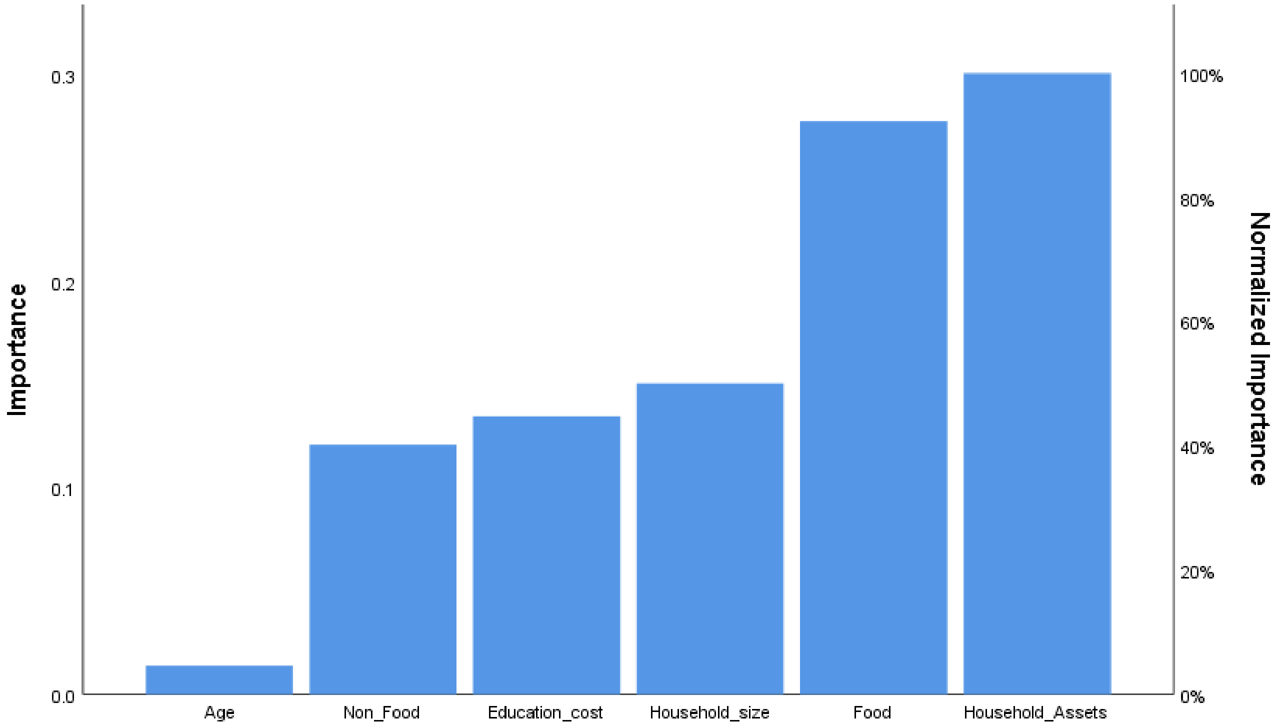


FIGURE 5.3: The relative importance of each predictor in the rural area
Source: An analysis based on the EICV5 dataset.

The above output shown in figure 5.3 is the relative importance of each predictor (components of household expenditure) of household expenditure outcomes. It reveals that the most important predictor is household assets (durable goods) (0.301), followed by food (0.27.8), household size (0.151), Education cost (0.135), non Food (0.121), and lastly Age (0.014). With the above output, we can conclude that household assets (durable goods), food, and household size have the highest importance in estimating household expenditure in rural areas.

Training	Sum Square of Error	77.427
	Relative Error	0.109
Testing	Sum Square of Error	148.298
	Relative Error	0.017

TABLE 5.3: Model Summary for Rural residence area
Source: An analysis based on the EICV5 dataset

5.2.2 Urban Area model

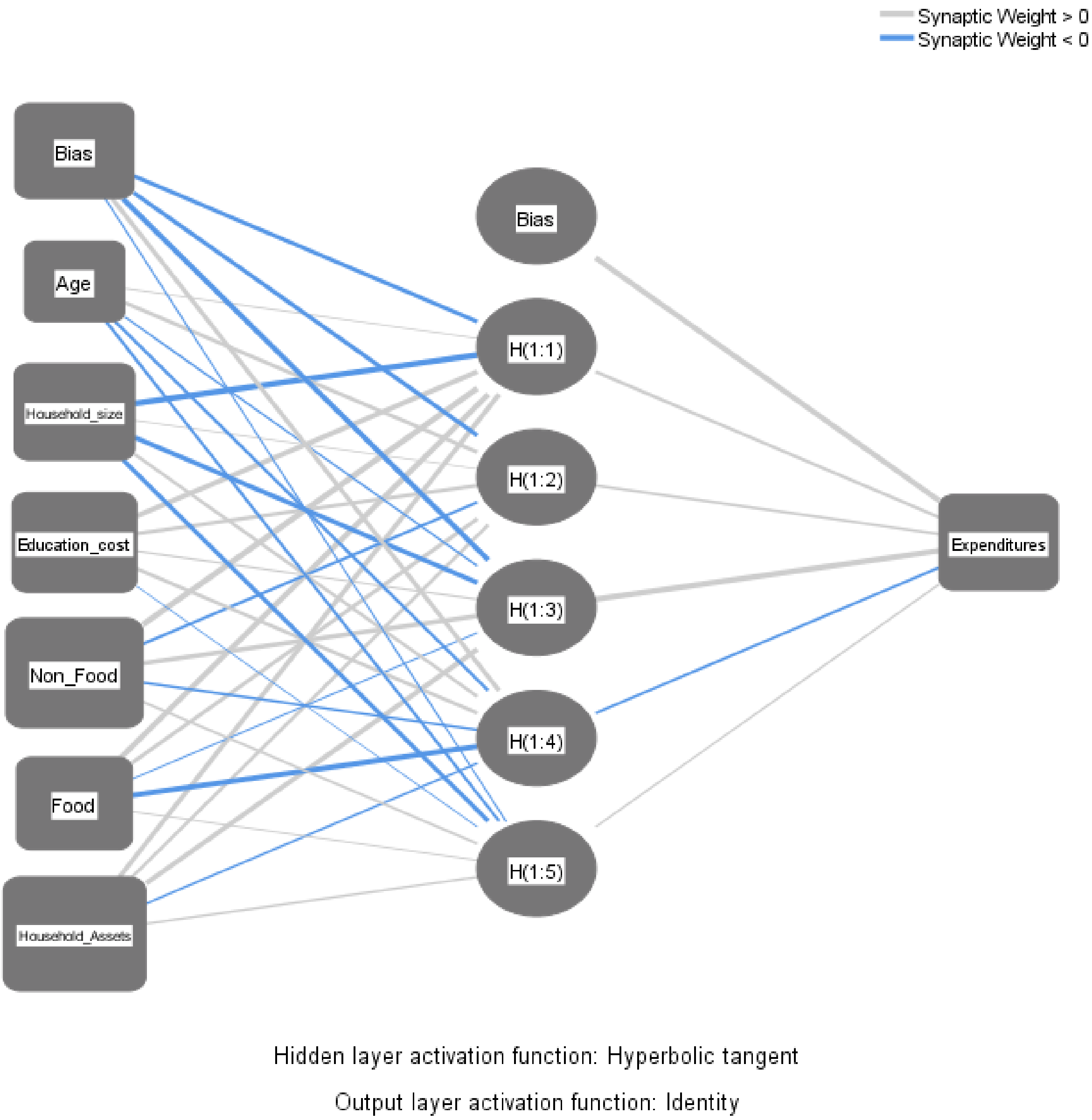


FIGURE 5.4: neural network for urban area model
Source: An analysis based on EICV5 dataset.

Predictor		Parameter Estimates					
		Predicted Hidden layer					Output Layer
		H_1	H_2	H_3	H_4	H_5	Expenditure
Input Layer	(Bias)	-0.455	-0.490	-0.774	0.575	-0.109	
	Age	0.033	0.426	-0.143	-0.295	-0.238	
	Household size	-0.924	0.021	-0.615	0.331	-0.510	
	Education cost	0.739	0.386	0.128	0.360	-0.015	
	Non Food	1.148	-0.224	0.599	-0.163	0.186	
	Food	0.859	0.442	-0.085	-0.776	0.0461	
	Household Assets	0.687	0.355	0.712	-0.153	0.163	
Hidden Layer	(Bias)						0.835
	H_1						0.325
	H_2						0.209
	H_3						0.905
	H_4						-0.209
	H_5						0.149

TABLE 5.4: neural network weight output for an urban area model
Source: An analysis based on the EICV5 dataset

From table 5.4 the input layers of Household assets, Non-Food, Food, education cost, and Age are on the top for their contribution to the hidden layer node 1 with respective own weights and are associated with the positive weights while Household size is negative to hidden layer node 1 in respect of its weight. The input layers of Household asset, Household size, Age, education cost, and Food are positively related to the hidden layer node 2, while the input layer of Non-Food is negatively related to the hidden layer node 2.

The input layer of Non-Food, education cost, and Household assets(durable goods) are the most contributors to hidden layer node 3 with its positive weight, while Food, Age, and Household size are negatively contributing to the hidden layer node 3. The input layer of Education cost and Household size are the main contributors with their respective weights while the input layers of household assets(durable goods), age, Non-Food, and Food are also positively contributing to Hidden and lastly the input layer of household assets(durable goods), food and non-food are positively contributing to hidden layer node 5 while age, household size, and education cost are negatively contributing to hidden 5. Now, we presume that the order of importance of hidden layers is as follows $H3 > H1 > H2 > H4 > H5$. The third and first nodes of the hidden layer are the main contributor to expenditure, where the weights of Household Assets (Durable goods) are 0.712, 0.355, 0.288, -0.153, 0.163, and 0.599, 1.148, -0.224,

-0.163, 0.186 respectively. based on those variables' weights, they are Leading contributors to these hidden layers, in line with the order of hidden layers.

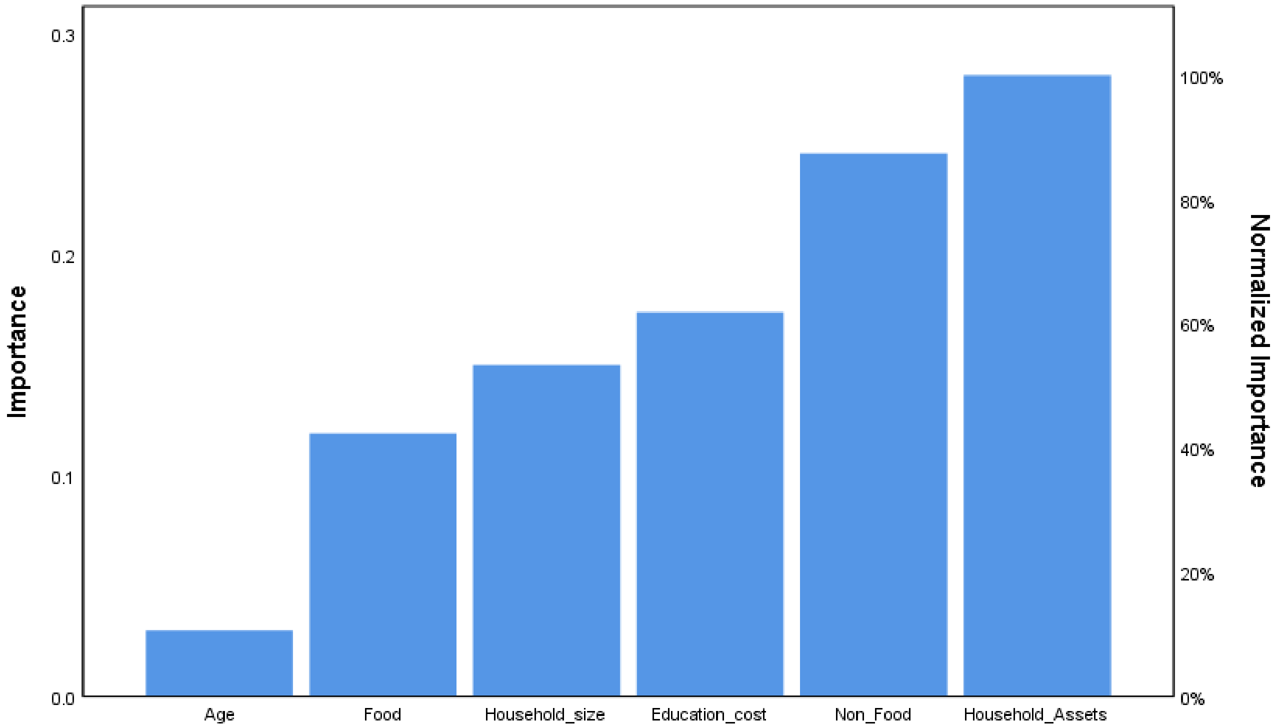


FIGURE 5.5: The relative importance of each predictor in the urban area model
Source: An analysis based on the EICV5 dataset.

The above output shown in figure 5.5 is the relative importance of each predictor (components of household expenditure) of household expenditure outcomes. It reveals that the most important predictor is household assets (durable goods) (0.281), followed by non-food (0.246), Education cost (0.174), household size (0.15), food (0.119), and lastly Age (0.03). With the above output, we can conclude that household assets(durable goods), non-food, and education costs have the highest importance in estimating household expenditure in the urban area.

Training	Sum Square of Error	104.472
	Relative Error	0.022
Testing	Sum Square of Error	14.172
	Relative Error	0.016

TABLE 5.5: Model Summary for the Urban residence area
Source: An analysis based on the EICV5 dataset

5.2.3 Single model

After running the models of rural and urban residential areas respectively, there is a difference between rural and urban residential areas in terms of household consumption as shown by table 5.2 and table 5.4. In this model, the residential area has been added as an explanatory variable as revealed by table 5.6 so that we can able to combine rural and urban residential areas. Thus, the impact of the residence variable will be analyzed on household expenditure.

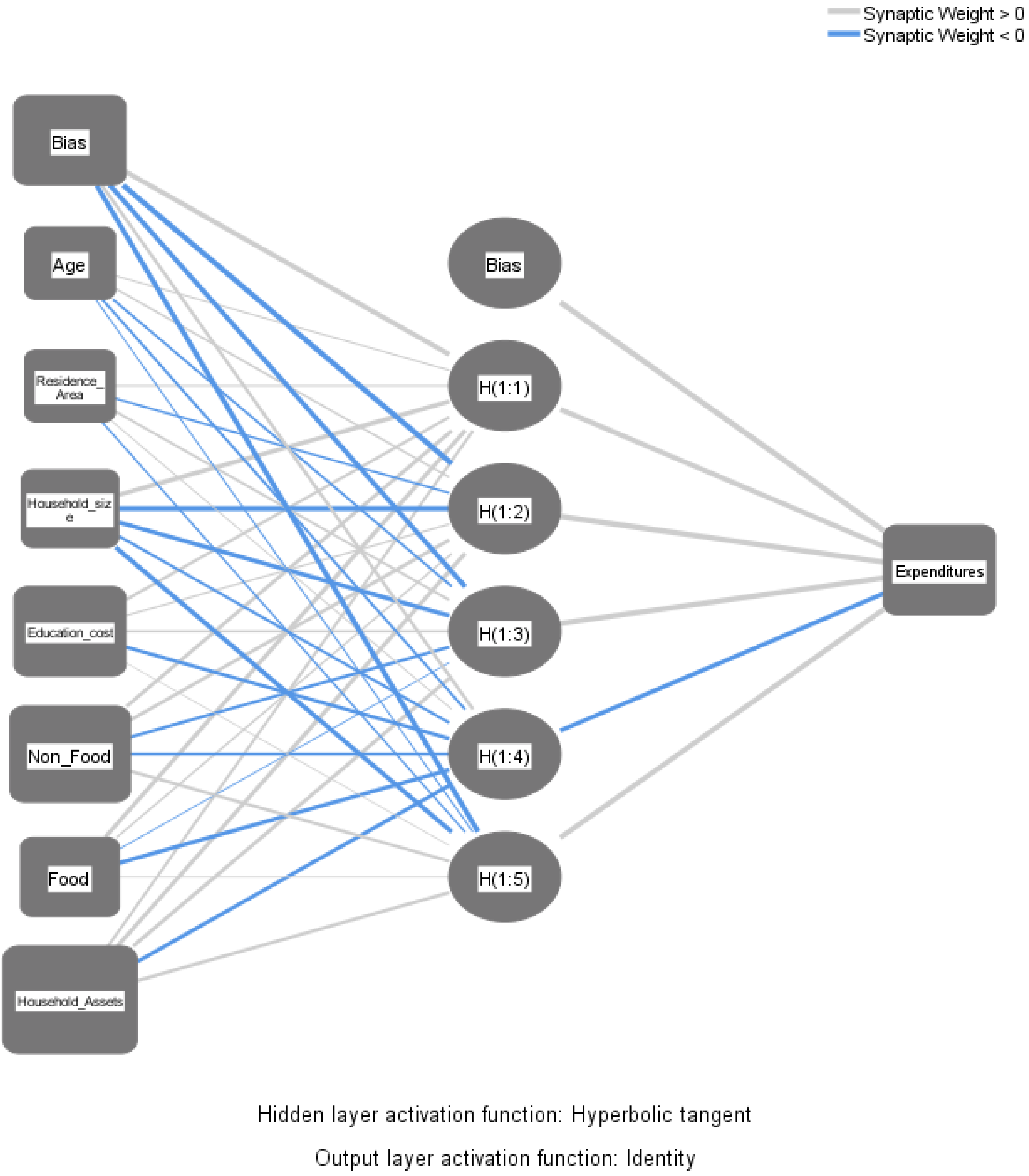


FIGURE 5.6: neural network for single model
Source: An analysis based on EICV5 dataset.

Predictor		Parameter Estimates					
		Predicted Hidden layer					Output Layer
		H_1	H_2	H_3	H_4	H_5	Expenditures
Input Layer	(Bias)	1.186	-1.611	-1.022	0.169	-1.396	
	Age	0.025	0.064	-0.076	-0.099	-0.016	
	Residence Area	0.038	-0.068	0.091	0.010	-0.050	
	Household size	0.639	-0.738	-0.462	-0.117	-0.647	
	Education cost	0.143	0.043	0.070	-0.271	0.005	
	Non Food	0.335	0.234	-0.159	-0.108	0.165	
	Food	0.421	0.054	-0.007	-0.410	0.009	
	Household Assets	0.140	0.413	0.364	-0.392	0.222	
Hidden Layer	(Bias)						3.138
	H_1						1.562
	H_2						1.849
	H_3						1.679
	H_4						-0.532
	H_5						1.566

TABLE 5.6: neural network weight output for single model
Source: An analysis based on the EICV5 dataset

Table 5.6 The input layers of Household assets (durable goods), Non-Food, Food, education cost, residence area, Age, and Household size are all positively associated with hidden layer node 1 with respective their positive weights. The input layers of Household assets, Age, education cost, Non-Food, and Food are positively related to the hidden layer node 2, while the input layer of Household size and residence area are negatively related to the hidden layer node 2.

The input layer of Residence area, education cost, and Household assets(durable goods) are the most contributors to hidden layer node 3 with its positive weight, while Food, Age, Non-Food, and Household size are negatively contributing to hidden layer node 3. The input layer of residential area is contributing positively to hidden layer node 4 while the input layers of household assets(durable goods), age, Non-Food, Food Education cost, and Household size are negatively contributing to hidden layer node 4. Lastly, the input layer of household assets (durable goods), food, non-food, and education cost is positively contributing to hidden layer node 5 with the positives weights while age, household size, and residence area are negatively contributing to hidden layer node 5. Therefore, the importance of hidden layer is ordered as follows $H2 > H3 > H5 > H1 > H4$. The first and third nodes of the hidden layer are the main contributor to expenditure, where the weights of Household Assets (Durable goods) and Non$_{F}oodare0.413, 0.222, 0.364, 0.140, -0.392and0.234, -0.159, 0.165, 0.335, -0.108respectively.basedontha$

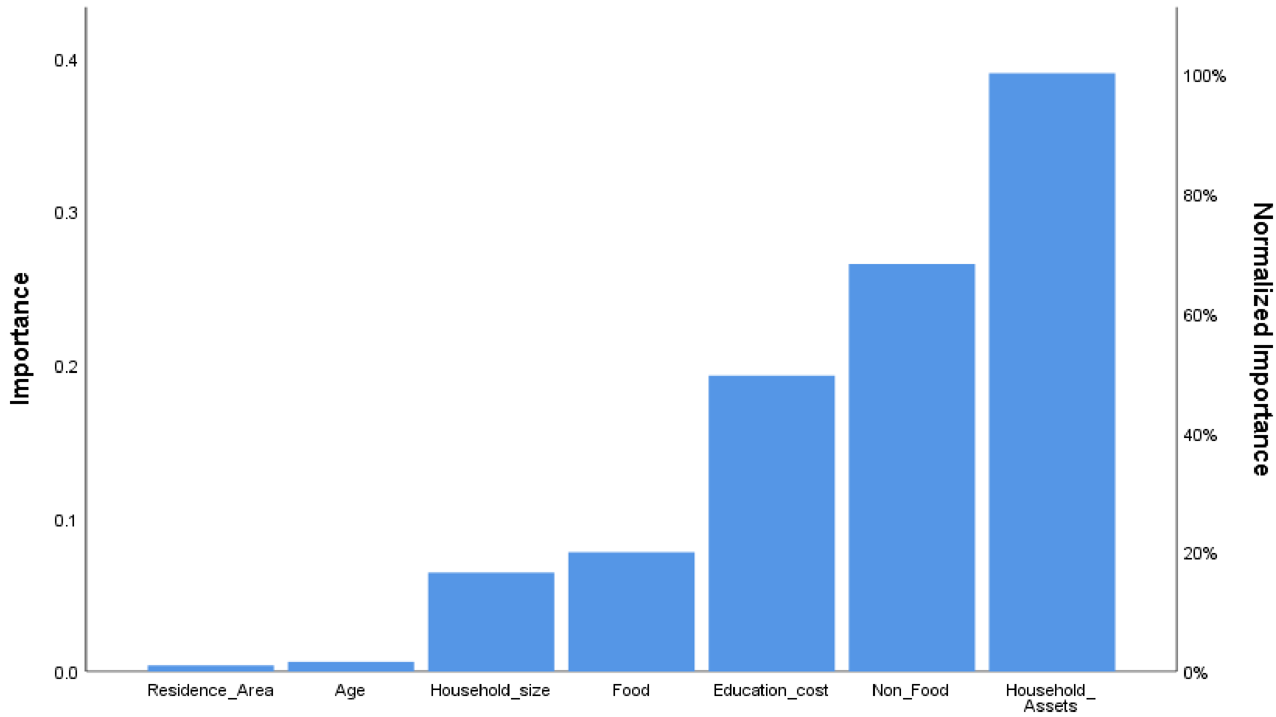


FIGURE 5.7: The relative importance of each predictor in a single model
Source: An analysis based on the EICV5 dataset.

The above output shown in figure 5.6 is the relative importance of each predictor (components of household expenditure) of household expenditure outcomes. It reveals that the most important predictor is household assets(durable goods) (0.39), followed non food (0.266), Education cost (0.193), food (0.077), household size (0.064), Age (0.006) and lastly Residence area (0.005). With the above output, we can conclude that Household assets (durable goods), non-food, education cost, food, and household size have considerable importance in estimating household expenditure in Rwanda.

The tables 5.3, 5.5 and 5.7 the relative errors for training and relative errors for testing are 0.109, 0.022, 0.108 and 0.017, 0.016, 0.026 respectively. those errors are very small; therefore, they are all very good for the prediction.

Training	Sum Square of Error	991.809
	Relative Error	0.108
Testing	Sum Square of Error	211.325
	Relative Error	0.026

TABLE 5.7: Model Summary for a single
Source: An analysis based on the EICV5 dataset

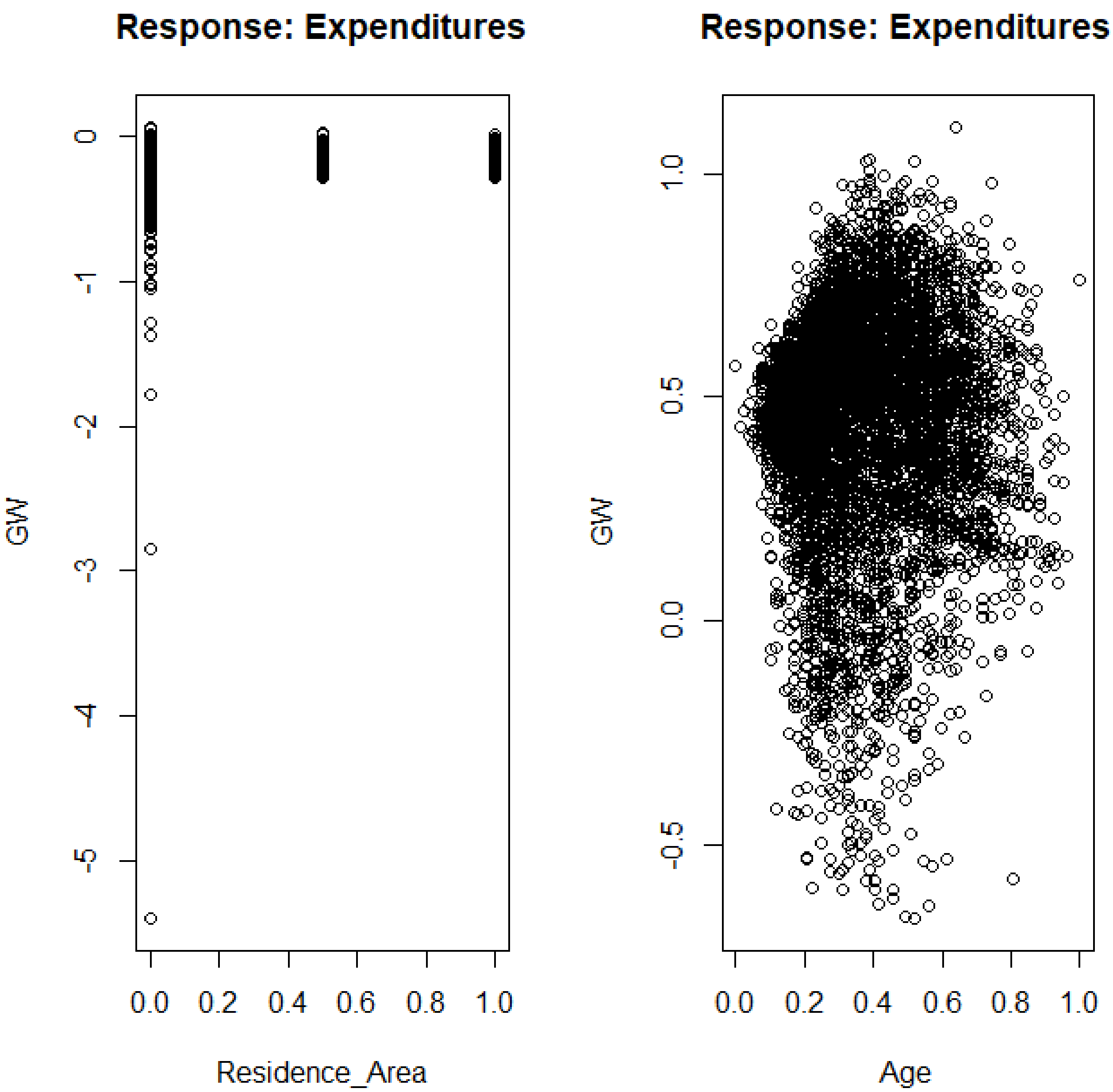
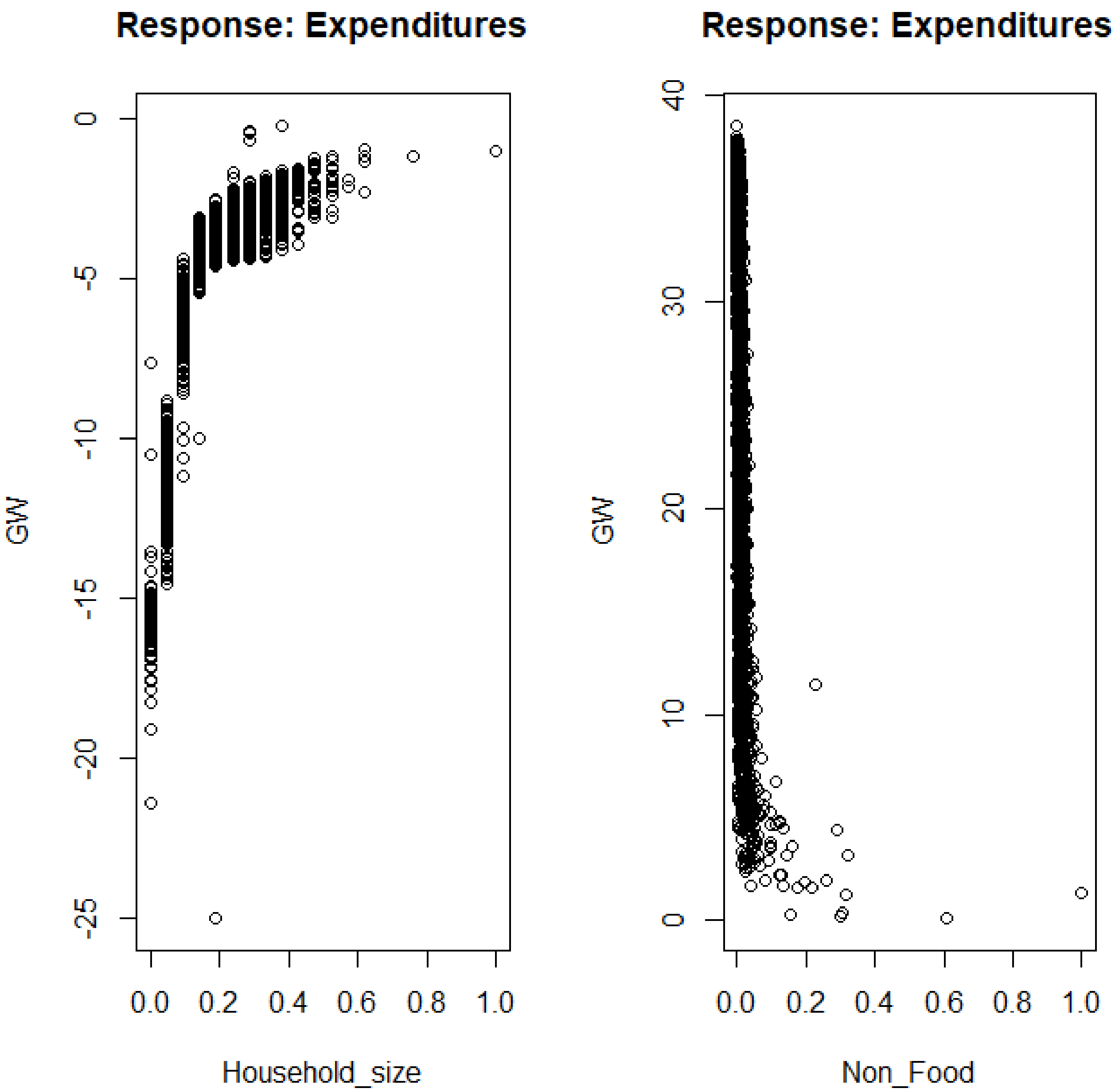
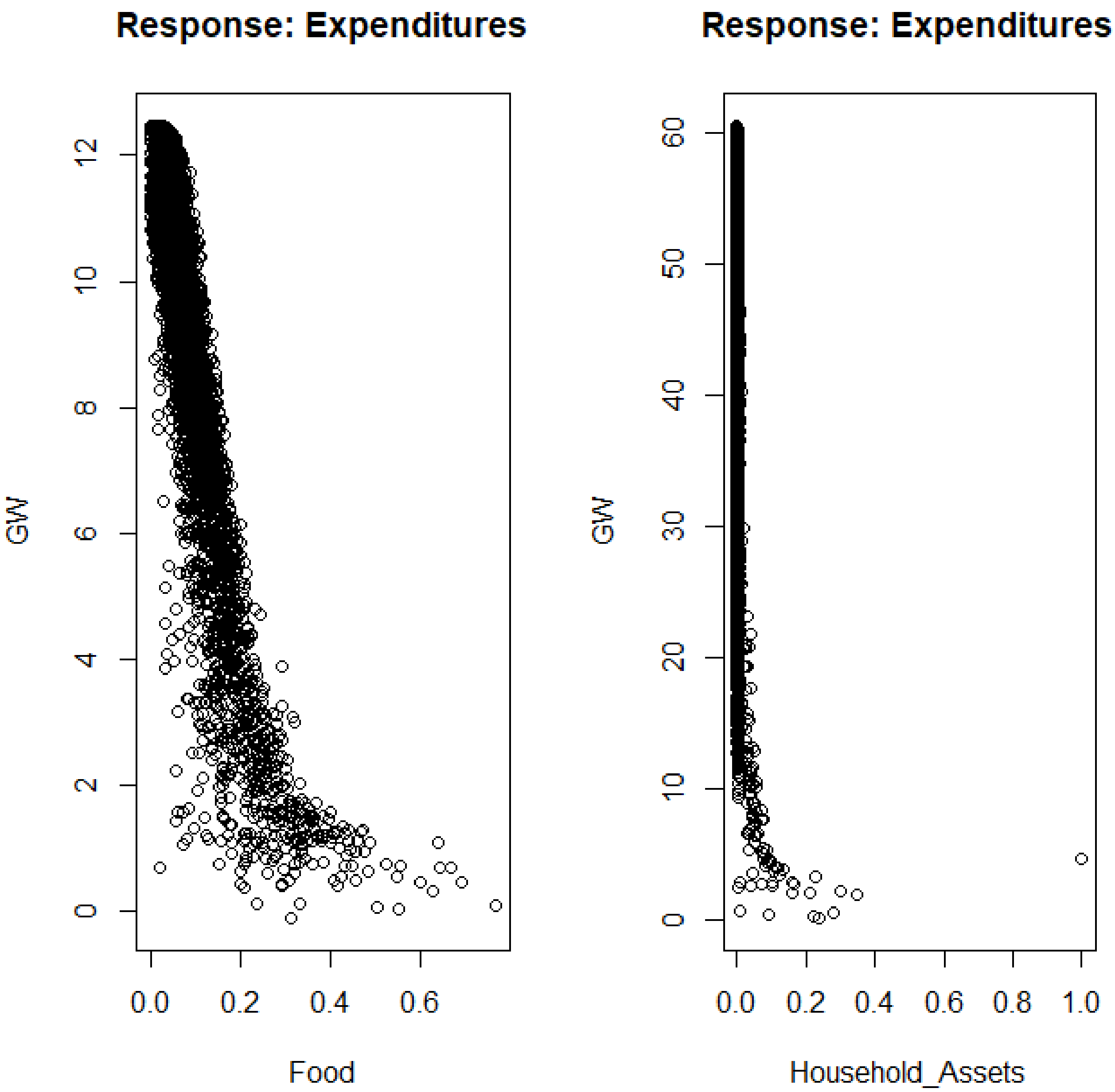


FIGURE 5.8: generalized weights for each covariate on component of household expenditure
Source: An analysis based on the EICV5 dataset.



Source: An analysis based on the EICV5 dataset.



Source: An analysis based on the EICV5 dataset.

In Figure 5.8, the variances of generalized weights for covariates Household size and Residence area appear large, indicating non-linearity of its effects, while Food, non-food, and Household asset indicating linearity of their effects. When generalized covariate weights gather about zero, the covariate does not affect the status of the outcome.

5.3 Model verification

The determination coefficient (R^2) was used to verify neural networks and regression models. The root-mean-square error (RMSE), which indicates what variation is represented by the model, was also utilized to validate the model's capacity to match. The neural network was characterized by the higher (R^2) and lower RMSE values and vice versa to the regression model. The (R^2) were 0.84 and 0.79 for the Neural network and regression respectively while RMSE were 19514 and 21720 for the Neural network and regression respectively.

Chapter 6

Conclusion and Recommendation

In this study the variables considered, are grouped into socio-economic and demographic factors namely the age of household head, residence area, Household size, education level of household head, the main occupation of the Household head, food, nonfood, household assets (durable goods) and education cost. The residence area affects household expenditure patterns in Rwanda as long as there is a discrepancy of lifestyles between rural and urban residential areas as shown by table 5.2 and table 5.4.

Neural network method was applied, the first specific objective was to analyze the influence of demographic and socio-economic factors on household expenditure levels. The results from the neural network model in figure 5.1, are discriminating the household expenditure levels.

The second specific objective of the study was to investigate the influence of the independent variable's effect for each household residence area, for estimating household expenditure patterns. The results from the neural network model in table 5.6, give the importance of household assets (durable goods). They have significantly estimated the household expenditure patterns in Rwanda, either in rural residential areas or urban residential areas so household assets(durable goods) are on top in contributing household expenditure. Moreover, analyzing the rural and urban residence area model separately as shown by tables 5.2 and 5.4, in rural residential areas the household expenditure is significantly estimated by food and household size while in the urban residence area, the household expenditure is significantly estimated by non-food and education cost.

The analysis of component expenditure revealed the order of importance of variables as it is the Household assets(durable goods), the non-food, the Household size, the education cost, the food, and the Age. while the variances of generalized weights as shown by figure 5.8 of NN for covariates Household size appear large, indicating non-linearity of its effects, while Age, Food, non-food, and Household asset indicating linearity of their independent effects on expenditure. All these component expenditures are significant factors for estimating the household expenditure in Rwanda.

So, the efficiency of neural networks models over regression models for household expenditure. Since they produced the root mean square errors and coefficient of determination which are 19514 and 0.84 respectively, while regression produced the root mean square errors and coefficient of determination which are 21720 and 0.79 respectively. This powered neural network for being the best model for estimating household expenditures.

There is limitation that should be pointed out for future researchers, this study considered rural and urban areas household expenditure. But it did not take into account the spacial analysis for better visualization region by region. Unfortunately, The spacial data was not available so the NISR should be expected to collect more information including the spacial data for further studies.The study results should be considered userful for policymakers and also the study makes contribution on methodoly for estimating household expenditure patterns.

Chapter 7

Reference

Nik, M, & Norhaslinda, A and Aslina, N (2011) Household Income and Expenditure Relationships: A Simultaneous Equation Approach Retrieved from: http://www.davidpublisher.org/Public/uploads/Contribute/5562977971a5c.pdf

Campbell, J Y, & Cocco, J, F (2004) How do house prices affect consumption? Evidence from micro data. Harvard Institute of Economic Research. Discussion Paper No. 2045. Harvard University Cambridge, Massachusetts Retrieved from: https://econpapers.repec.org/paper/redsed004/304.htm

Zurawicki, L, & Braidot, N (2005) Consumers during crisis: Responses from the middle class in Argentina. Journal of business research volume 58 (2005) 1100-1109 Retrieved from : https://www.sciencedirect.com/science/article/abs/pii/S0148296304000918

Blacklow, P, & Ray, R (2000) A Comparison of Income and Expenditure Inequality Estimates: The Australian Evidence, 1975/76 to 1993/94. The Australian Economics Review, 33(4), 317-329 Retrieved from: https://doi.org/10.1111/1467-8462.00162

World Bank (2018) Rwanda has achieved impressive development gains since the 1994 genocide and civil war Retrieved from :https://www.worldbank.org/en/country/rwanda/overview1

OECD (2013) "Household consumption", in OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth, OECD Publishing, Paris Retrieved from https://doi.org/10.1787/9789264194830-8-en

Francis, T, & Abdi, K (2014) determinants of household expenditure in Malawi , Asian academic research journal of multidisciplinary Retrieved from : https://www.academia.edu/7727625/DETERMINANTS_OF_HOUSEHOLD_EXPENDITURE_IN_MALAWI_A_SIMULTANEOUS_QUANTILE_REGRESSION_APPROACH

Koenker & Hallock, S (2001) Quantile regression: An introduction. Journal of economic perspectives Retrieved from: https://www.aeaweb.org/articles?id=10.1257/jep.15.4.143

Christopher, M Bishop (2006) Pattern and recognition machine learning Retrieved from: http://users.isr.ist.utl.pt/ wurmd/Livros/school/Bishop%20%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf

Zhongheng, Z (2016) Further insights into error function, generalized weights and others Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5009026/

National Institute of Statistics of Rwanda (2018) The Fifth Integrated Household Living Conditions Survey Rwanda (EICV5) 2016/2017 Rwanda Poverty Profile Report Retrieved from : https://microdata.statistics.gov.rw/index.php/catalog/82

National Institute of Statistics of Rwanda (2015). The Fourth Integrated Household Living Conditions Survey Rwanda (EICV4) 2013/2014 Rwanda Poverty Profile Report Retrieved from:https://microdata.statistics.gov.rw/index.php/catalog/75

Maciel, S L, & Ballini, R (2008) Design a neural network for time series financial forecasting: Accuracy and robustness analysis, pp. 1–18 Retrieved from : https://www.tandfonline.com/doi/pdf/10.1080/1331677X.2015.1081828

Muhammad, A, & Nasim, A (2012) Household Consumption in Pakistan Retrieved from: https://www.econ-jobs.com/research/43049-Estimation-of-Household-Consumtion.pdf

Houthakker, H S (1957) An International Comparison of Household Expenditure Patterns, Commemorating the Centenary of Engel's Law Retrieved from: https://www.jstor.org/stable/1905382

Edna, M (2017) The determinant of household expenditure on education Retrieved from: http://ijecm.co.uk/wp-content/uploads/2017/12/51254.pdf/

Mats, W (2002) Household expenditure for housing attributes: a linear expenditures system with hedonic prices Retrieved from https://ideas.repec.org/a/eee/jhouse/v11y2002i1p75-93.html

World Bank (2018) Measuring household expenditure on Education Retrieved from: http://uis.unesco.org/sites/default/files/documents/measuring-household-expenditure-education-2018-en.pdf/

Giuseppe, C (2017) neural network with R Retrieved from https://www.packtpub.com/product/neural-networks-with-r/9781788397872

World Health Organization (2003) Household catastrophic health expenditure: a multi country analysis, the lancet, volume 362, 2003. Retrieved from : https://www.who.int/health_financing/documents/household-health-expenditure/en/

Haag, D (2013) time use, consumption expenditures and employment status: evidence from LISS panel, the 7th Mess workshop Retrieved from: https://www.researchgate.net/publication/263803945_Time_Use_Consumption_Expenditures_and_Employment_Status_Evidence_from_the_LISS_Panel

Sumarto, S, et al (2006) Predicting consumption poverty using non-consumption indicators: experiments using Indonesia data Retrieved from : https://link.springer.com/article/10.1007/s11205-006-0023-x

Ahmad, Z, & Fatima, A (2011) Prediction of household expenditure on the basis of household characteristics, University of Gujrat, Pakistan Retrieved from : https://www.researchgate.net/publication/266736558_PREDICTION_OF_HOUSEHOLD_EXPENDITURE_ON_THE_BASIS_OF_HOUSEHOLD_CHARACTERISTICS

Nkurunziza, J, & Broekhuis, A and Hooimeijer, P (2015) Do Poverty Reduction Programmes-Foster Education Expenditure? New Evidence from Rwanda, Journal of Asian and African Studies 1–19 Retrieved from: https://journals.sagepub.com/doi/abs/10.1177/0021909615595988

Hongyun, H & Fan, S (2020) how the composition of asset portfolios affects household consumption: evidence from China based on micro data Retrieved from https://www.mdpi.com › pdf

Obisesan, O and al (2016) Food and Non-Food Expenditure Differential across Poor and Non-Poor Households in South-East Nigeria Retrieved from : https://www.ijser.org/research paper/Food-and-Non-Food-Expenditure-Differential across-Poor-and-Non-Poor-Households-in-South-East-Nigeria.pdf

Ministry of Local Government (2018) Social Protection Sector Strategic Plan (SP-SSP) 2018/19 –2023/24 Retrieved from: http://www.minecofin.gov.rw/fileadmin/templates/ documents/NDPR/Sector_Strategic_Plans/Social_protection.pdf

Obaidullah, J (2019) Engel Curve Retrieved from : https://xplaind.com/506474/engel-curve: :text=An%20Engel%20curve%20is%20a,of%20demand%20is%20consumer%20income

Zehiwot & Senapathy (2019) Determinants of Household Consumption Expenditure in Debremarkos Town, Amhara Region, Ethiopia Retrieved from: http://asrjetsjournal.org/

Ministry of Health (2010) RWANDA COMMUNITY BASED HEALTH INSURANCE POLICY Retrieved from: https://rbc.gov.rw/fileadmin/user_upload/Rwanda_Community_Based_Health_Insurance_Policy.pdf

Abeeda F, & Zahoor, A (2013) ESTIMATION OF HOUSEHOLD EXPENDITURES ON THE BASIS OF HOUSEHOLD CHARACTERISTICS BY REGION Retrieved from: https://www.researchgate.net/publication 266736651_ESTIMATION_OF_HOUSEHOLD_EXPENDITURES_ON_THE_BASIS_OF_HOUSEHOLD_CHARACTERISTICS_BY_REGION.pdf

Belete, A & Igodan, C, and Averbeke, W (1999) ANALYSIS OF RURAL HOUSEHOLD CONSUMPTION EXPENDITURE IN SOUTH AFRICA: THE CASE OF FOOD PLOT HOLDERS IN TYEFU IRRIGATION SCHEME IN THE EASTERN CAPE PROVINCE Retrieved from: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2173047_code517200.pdf?abstractid=2173047mirid=1

Farkhanda, S, & Eatzaz, A (2007) Understanding household consumption patterns in Pakistan Retrieved from:https://sci-hub.se/https://doi.org/10.1016/j.jretconser.2006.09.008

Landry, S, & Chelsea, J (2018) Africa's consumer market potential Retrieved from https://www.brookings.edu/wp-content/uploads/2018/12/africas-consumer-market-potential.pdf

OECD (2013), "Household consumption", in National Accounts at a Glance 2013, OECD Publishing, Paris Retrieved from https://doi.org/10.1787/na$_{g}lance - 2013 - 12 - en$

Zivar Z, Mubariz M (2020) Analysis of the economic factors affecting household consumption expenditures in Azerbaijan Retrieved from https://www.researchgate.net/publication/341279386

Chapter 8

Appendices

8.1 Codes

```
# Installing Packages

install.packages("NeuralNetTools")
install.packages("neuralnet")
install.packages("ppcor")
install.packages("corpcor")
install.packages("dplyr")
install.packages("haven")
install.packages("ggplot2")
install.packages("crayon")

#Required libraries
library(dplyr)
library(NeuralNetTools)
library(neuralnet)
library(ggplot2)
library(ppcor)
library(corpcor)
library(haven)


# Personal_data
Person<- read_dta(file.choose())
# Viewing number of columns and rows
ncol(Person)
nrow(Person)
# Names of columns
names(Person)

#Retrieving the columns by slicing
Person_Data<-Person[c(1,7,12,13,15,38,49)]
colnames(Person_Data)
```

```
# filtering the value with household head
Person_DatA<-Person_Data %>%
  dplyr::select(hhid,region,s1q2,s1q3y,s1q4,s3q2,s4aq2) %>% filter(s1q2==1)
nrow(Person_DatA)

#Occupation
oCcupation<-read_dta(file.choose())
oCcupation<-oCcupation[c(1,12)]
colnames(oCCupation)


# Merging Personal, Occupation and VUP_ubudehe categories
Persons_Occupation<-merge(Person_DatA,oCCupation,by="hhid")
nrow(Persons_Occupation)

#Remove duplicate row
Person_Occupation<-Persons_Occupation%>% distinct(hhid, .keep_all= TRUE)
colnames(Persons_Occupation)
nrow(Persons_Occupation)

# Ubudehe category
HHs_category<-read_dta(file.choose())
colnames(HHs_category)
nrow(HHs_category)

#Retrieving the categories  by slicing
Hhsubudehe_cat<-HHs_category[c(1,11,23)]
colnames(Hhsubudehe_cat)

# Merging Personal, Occupation and VUP_ubudehe categories
Persons_Occupation_Cat<-merge(Persons_Occupation,Hhsubudehe_cat,by="hhid")
colnames(Persons_Occupation_Cat)
nrow(Persons_Occupation_Cat)
```

Source: An analysis based on the EICV5 dataset.

```r
# EICV_Poverty
EICVPoverty_Data<-read_dta(file.choose())
colnames(EICVPoverty_Data)
#Combining water and Electricity expenses
Elect_Water<-EICVPoverty_Data$exp7+EICVPoverty_Data$exp8
View(EICVPoverty_Data)
#Columns Binding
EICVPoverty_DAta<-cbind(EICVPoverty_Data,Elect_Water)
colnames(EICVPoverty_DAta)
EICVPoverty_DATa<-EICVPoverty_DAta[c(1,7,10,19,22,25,29,38)]

nrow(EICVPoverty_DATa)

# Merging Personal, Occupation ,VUP_ubudehe categories and EICV_Poverty_Data
HHsExpenditure_Data<-merge(Persons_Occupation_Cat,EICVPoverty_DATa,by="hhid")

HHsExpenditureData<-HHsExpenditure_Data %>% distinct(hhid, .keep_all= TRUE)
colnames(HHsExpenditure_Data)
nrow(HHsExpenditure_Data)


#Slicing columns

HHsExpenditureDAta<-HHsExpenditure_Data[c(4,8:14)]
colnames(HHsExpenditureDAta)
nrow(HHsExpenditureDAta)
head(HHsExpenditureDAta)
str(HHsExpenditureDAta)
```

Source: An analysis based on the EICV5 dataset.

```r
# Renaming columns

HHs_Expenditures<-HHsExpenditureDAta%>%rename(Age=s1q3y,Education_cost = exp1,
                                        HHs_Asset=exp17,HHs_size=member,Non_Food=exp12,
                                        Food=exp15_2,Expenditures =cons1_ae)
colnames(HHs_Expenditures)
dim(HHs_Expenditures)


#Data Cleaning

#Data Cleaning by removing the missing variables
HH_Expenditures<-complete.cases(HHs_Expenditures)
Hhs_Expenditures<-HHs_Expenditures[HH_Expenditures,]

head(Hhs_Expenditures)
nrow(Hhs_Expenditures)

# Initial exporatory analysis
summary(Hhs_Expenditures)



#Partial colleration coeffiecients
pcor(Hhs_Expenditures, method = "pearson")
```

Source: An analysis based on the EICV5 dataset.

```
set.seed(123)
n.epoch=100
data = Hhs_Expenditures
max_data <- apply(data, 2, max)
min_data <- apply(data, 2, min)
data_scaled <- scale(data,center = min_data, scale = max_data - min_data)

# splitting the data into Training and Test Data
index = sample(1:nrow(data),round(0.7*nrow(data)))
Data_train <- as.data.frame(data_scaled[index,])
Data_test <- as.data.frame(data_scaled[-index,])
head(Data_train)


#Neural Network

nnetw <- neuralnet(Expenditures~Age+HHs_size+Education_cost+Non_Food+food+HHs_Asset
                  data=Data_train, hidden=c(5),act.fct = "logistic", linear.output=T, threshold=0.

nnetw$result.matrix
nnetw$net.result
nnetw$weights
plot(nnetw)
```

Source: An analysis based on the EICV5 dataset.

```
predict_test <- compute(nnetw,Data_test[,1:6])

predict_start <- predict_test$net.result*(max(data$Expenditures)- min(data$Expenditures))
test_start <- as.data.frame((Data_test$Expenditures)*(max(data$Expenditures)-min(data$Expenditures
MSE.net_data <- sum((predict_start - test_start)^2)/nrow(test_start)

Regression_Model <- lm(Expenditures~., data=data)
summary(Regression_Model)
test <- data[-index,]
predict_lm <- predict(Regression_Model,test)
MSE.lm <- sum((predict_lm - test$Expenditures)^2)/nrow(test)

MSE.net_data
MSE.lm


# General weights for each covariate
par(mfrow=c(2,2))
gwplot(nnetw,selected.covariate="Residence_Area")
gwplot(nnetw,selected.covariate="Age")
gwplot(nnetw,selected.covariate="HHs_size")
gwplot(nnetw,selected.covariate="Non_Food")
gwplot(nnetw,selected.covariate="Food")
gwplot(nnetw,selected.covariate="HHs_Asset")
```

Source: An analysis based on the EICV5 dataset.

```r
#Variable importance using Garson's algorithm

round(garson(nnetw,bar_plot = FALSE),3)
garson(nnetw)

#Relative importance using Garson's algorithm

cols <- rainbow(8)
garson(nnetw) +
  scale_y_continuous('Rel. Importance',
                     limits = c(0, 0.7)) +scale_fill_gradientn(colours = cols) +
  scale_colour_gradientn(colours = cols)
```

Source: An analysis based on the EICV5 dataset.

```r
#Categorize some variables

HHsExpenditureDATA$Education_level<-cut(HHsExpenditureDATA$Education_level,
                                        breaks=c(1, 10, 15, 16,47),
                                         labels=c(1,2,3,4))
HHsExpenditureDATA$Residence_Area<-cut(HHsExpenditureDATA$Residence_Area,
                                         breaks=c(0,1,2,6),
                                         labels=c("Kigali City","Other Cities","Rural"))
HHsExpenditureDATA$Expenditures<-cut(HHsExpenditureDATA$Expenditures,
                                        breaks=c(0,159375,Inf),
                                    labels=c("Low Expenditure","High Expenditure"))
HHsExpenditureDATA$Main_Occupation<-cut(HHsExpenditureDATA$Main_Occupation,
                                         breaks=c(0,8 ,9),
                                         labels=c(1,2))
HHsExpenditureDATA$Health_Insurance<-cut(HHsExpenditureDATA$Health_Insurance,
                                          breaks=c(1,5,6),
                                          labels=c(1,2))


head(HHsExpenditureDATA)
dim(HHsExpenditureDATA)

# Convert factors to numbers
HHsExpenditureDATA[,5] <- sapply(HHsExpenditureDATA[,5],switch,"Kigali City"=1,"Other Cities"=2,"Rural"=3)
HHsExpenditureDATA[,9] <- sapply(HHsExpenditureDATA[,9],switch,"Low Expenditure"=1,"Medium Expenditures"=2,"High Expenditure"=3)

HHsExpenditureDATA$Education_level<-as.numeric(HHsExpenditureDATA$Education_level)
HHsExpenditureDATA$Main_Occupation<-as.numeric(HHsExpenditureDATA$Main_Occupation)
HHsExpenditureDATA$Health_Insurance<-as.numeric(HHsExpenditureDATA$Health_Insurance)
head(HHsExpenditureDATA)
colnames(HHsExpenditureDATA)
str(HHsExpenditureDATA)
```

Source: An analysis based on the EICV5 dataset.

```
> summary(Regression )

Call:
lm(formula = Expenditures ~ Age + HHs_size + Education_cost +
    Non_Food + Food + HHs_Asset, data = Data_train)

Residuals:
     Min       1Q   Median       3Q      Max
-0.15458 -0.00459 -0.00170  0.00214  0.42441

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.0143373  0.0003922  36.552   <2e-16 ***
Age             0.0002100  0.0008544   0.246    0.806
HHs_size       -0.0785054  0.0015608 -50.297   <2e-16 ***
Education_cost  0.0717236  0.0060102  11.934   <2e-16 ***
Non_Food        0.2376983  0.0101351  23.453   <2e-16 ***
Food            0.1379156  0.0026710  51.634   <2e-16 ***
HHs_Asset       0.4842674  0.0084721  57.160   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01243 on 7500 degrees of freedom
Multiple R-squared:  0.7966,	Adjusted R-squared:  0.7864
F-statistic:  2391 on 6 and 7500 DF,  p-value: < 2.2e-16
```

Source: An analysis based on the EICV5 dataset.

Modeling household expenditure in Rwanda using Neural Network

ORIGINALITY REPORT

15%	10%	6%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Zhongheng Zhang. "Neural networks: further insights into error function, generalized weights and others", Annals of Translational Medicine, 2016 Publication	1%
2	Submitted to Cranfield University Student Paper	1%
3	www.isoss.net Internet Source	1%
4	vibdoc.com Internet Source	1%
5	Submitted to Higley Unified School District #60 Student Paper	1%
6	repository.tudelft.nl Internet Source	1%
7	Ching-Sheng Tu, Deng-Yi Wu, Shie-Jue Lee, Yann-Yean Su. "Supervised Classification by RBF Networks With Self-Constructing Clustering", 2018 14th International	1%

	Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2018 Publication	
8	Submitted to University of Rwanda Student Paper	1 %
9	doaj.org Internet Source	<1 %
10	www.davidpublisher.org Internet Source	<1 %
11	link.springer.com Internet Source	<1 %
12	appear.at Internet Source	<1 %
13	www.worldbank.org Internet Source	<1 %
14	www.slideshare.net Internet Source	<1 %
15	Singh, Satwinder, Puneet Mittal, and K.S. Kahlon. "Empirical model for predicting high, medium and low severity faults using object oriented metrics in Mozilla Firefox", International Journal of Computer Applications in Technology, 2013. Publication	<1 %
	Submitted to King's College	

16	Student Paper	<1 %
17	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %
18	Submitted to Westcliff University Student Paper	<1 %
19	iris.unito.it Internet Source	<1 %
20	T.J. Sekhampu, F. Niyimbanira. "Analysis Of The Factors Influencing Household Expenditure In A South African Township", International Business & Economics Research Journal (IBER), 2013 Publication	<1 %
21	eprints.soton.ac.uk Internet Source	<1 %
22	article.sapub.org Internet Source	<1 %
23	Submitted to Birla Institute of Technology and Science Pilani Student Paper	<1 %
24	www.dol.gov Internet Source	<1 %
25	mpra.ub.uni-muenchen.de Internet Source	<1 %

16	Student Paper	<1 %
17	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %
18	Submitted to Westcliff University Student Paper	<1 %
19	iris.unito.it Internet Source	<1 %
20	T.J. Sekhampu, F. Niyimbanira. "Analysis Of The Factors Influencing Household Expenditure In A South African Township", International Business & Economics Research Journal (IBER), 2013 Publication	<1 %
21	eprints.soton.ac.uk Internet Source	<1 %
22	article.sapub.org Internet Source	<1 %
23	Submitted to Birla Institute of Technology and Science Pilani Student Paper	<1 %
24	www.dol.gov Internet Source	<1 %
25	mpra.ub.uni-muenchen.de Internet Source	<1 %

26	towardsdatascience.com Internet Source	<1 %
27	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
28	docplayer.net Internet Source	<1 %
29	financetrain.com Internet Source	<1 %
30	bmcpublichealth.biomedcentral.com Internet Source	<1 %
31	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
32	m.aliran.com Internet Source	<1 %
33	iati.dfid.gov.uk Internet Source	<1 %
34	Submitted to University of Durham Student Paper	<1 %
35	Submitted to The University of the South Pacific Student Paper	<1 %
36	mafiadoc.com Internet Source	<1 %

37	www.coursehero.com Internet Source	<1 %
38	tel.archives-ouvertes.fr Internet Source	<1 %
39	A. Belete, C. O. Igodan, C. M'Marete, W. van Averbeke. "ANALYSIS OF RURAL HOUSEHOLD CONSUMPTION EXPENDITURE IN SOUTH AFRICA: THE CASE OF FOOD PLOT HOLDERS IN TYEFU IRRIGATION SCHEME IN THE EASTERN CAPE PROVINCE / ANALISE VAN VERBRUIKSBESTEDING VAN LANDELIKE HUISHOUDINGS IN SUID-AFRIKA: 'N GEVALLESTUDIE VAN KLEINBOERE IN TYEFU BESPROEIINGSKEMA IN DIE OOSKAAP PROVINSIE", Agrekon, 1999 Publication	<1 %
40	Deepak Agrawal, Christopher Schorling. "Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model", Journal of Retailing, 1996 Publication	<1 %
41	tci-thaijo.org Internet Source	<1 %
42	Kayo Santana Barros. "Treatment, by electrodialysis, of a synthetic wastewater from the cyanide-free brass electroplating and	<1 %

study on ion transport: recovery of water and chemicals.", Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA), 2020
Publication

43	www.people.cecs.ucf.edu Internet Source	<1 %
44	hdl.handle.net Internet Source	<1 %
45	www.minecofin.gov.rw Internet Source	<1 %
46	Zhibin Quan, Zhiqiang Gao, Qigui Wang, Xiao Wen, Yucong Wang, Bowang Xiao. "Rapid Residual Stress and Distortion Prediction in Cast Aluminum Components Using Artificial Neural Network and Part Geometry Characteristics", SAE International, 2014 Publication	<1 %
47	eprints.mdx.ac.uk Internet Source	<1 %
48	Mario De Luca. "Evaluation of Runway Bearing Capacity using International Roughness Index", Transportation Research Procedia, 2020 Publication	<1 %
49	Constantia Barda, Eleni Sardianou. "Analysing consumers' ‘activism’ in response to rising	<1 %

prices", International Journal of Consumer Studies, 2010
Publication

50	Edna Douglas. "Engel functions revisited", Social Indicators Research, 1982 Publication	<1 %
51	F W Agbola, Y K Saini. "EFFECT OF HOUSEHOLD CHARACTERISTICS ON THE DECISION TO CONSUME STAPLE FOODS IN SOUTH AFRICA", Agrekon, 2002 Publication	<1 %
52	agricultureandfoodsecurity.biomedcentral.com Internet Source	<1 %
53	www.utamu.ac.ug Internet Source	<1 %
54	Jennifer Batamuliza, Damien Hanyurwimfura. "A secure and efficient anonymous certificateless signcryption for Key Distribution Scheme for Smart Grid", 2020 21st International Arab Conference on Information Technology (ACIT), 2020 Publication	<1 %
55	Woon Siong Gan. "Chapter 13 Deep Learning and Artificial Neural Networks", Springer Science and Business Media LLC, 2020 Publication	<1 %
56	documentop.com	

Internet Source

<1 %

57 "Posters", The Journal of Nutrition, Health and Aging, 2009 <1 %

Publication

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		