American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

PREDICTIVE POWER UNLEASHED: A COMPARATIVE STUDY OF MULTIPLE LINEAR REGRESSION AND ARTIFICIAL NEURAL NETWORKS FOR NEW BUSINESS REGISTRATIONS IN THE DRC

Ms. Soo-Min Kim
Master's Degree Student in Technology Convergence, Handong Global University, South Korea

Abstract*: The level of entrepreneurship in a country is a critical indicator of economic vitality and growth. The World Bank Group introduced the "new business registered" indicator in 2006 to gauge entrepreneurship by considering various factors influencing business creation. Although research on this subject is limited, it holds significant importance, particularly for developing nations seeking economic enhancement. Identifying the factors impacting entrepreneurship and comprehending their interplay facilitates informed policy decisions for these countries.*
Prediction is a fundamental aspect of organizational decision-making, guiding responses to internal challenges. Selecting the appropriate prediction model for a specific case is a formidable challenge, influenced by numerous factors including the context, data availability, desired precision, time frame, cost-effectiveness, and analysis complexity. Information and communication technologies, especially decision support systems, offer avenues for precise predictions even in complex datasets.
As The Democratic Republic of the Congo aspires to become an emergent nation by 2030, promoting entrepreneurship is pivotal. The "new business registered" indicator is paramount for assessing growth in this sector. Leveraging ICT for accurate prediction of this metric empowers policymakers to evaluate the effectiveness of current policies.
This paper delves into the challenges of selecting the right model for decision support systems and its seamless integration into workflows, particularly in the context of enhancing entrepreneurship.
Keywords: *Entrepreneurship, Prediction Models, Decision Support Systems, Economic Development*
ICT Integration

1. Introduction

The World Bank Group, established in 2006 an indicator “the new business registered defined as the number of new limited liability corporations registered in the calendar year”. This indicator measures the level of entrepreneurship of a country by considering several factors having a possible impact on business creation in a country (World Bank Group, 2010). Though little research has been conducted on this subject, whereas particular attention should be paid to it, especially for developing countries wanting to boost their economies. Identifying factors having impact on entrepreneurship and understanding their relationship would make it easier for country decision-makers to direct their policies on this matter. (Seher Khader, 2014).

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

Prediction is one of the most important task performed in organizations. They guide decision makers on how they can address some issues inside their structures. (Douglas C. Montgomery, 2008). One of the great challenge in the prediction is the determination of the model to be adopted for a specific case. Several factors have to be considered in the choice of the prediction method such as: the context of the prediction, the availability of data, the degree of precision desired in the prediction, the period to be predated, the profit produced for the company for the use of the predictions as well as the time needed to do the analysis (Smith, 1971). Thanks to information and communication technologies, especially with the use of decision support systems, it is possible to ease the decision making process and provide very accurate prediction even on complex data. (Juneja, 2018).Since The Democratic Republic of the Congo is targeting to be an emergent country by 2030, one solution among others is to promote entrepreneurship (Forum des as, 2016). And to assess the growth in this sector, the new business registered [indicator] is the one to be considered. Thus, ICT can be used to better predict this number; this will help policymakers to assess the impact of current policies over this indicator.

The challenge raising from this is mostly about the appropriate model to use in a decision support system (solution often use in decision making process) and it integration in workflow. (Kamel, 1998)

2. Literature Review

In 2006, the World Bank Group established a single indicator which can be used to determine factors impacting private sector growth. This indicator called new business registered (numbers) is defined as the number of new limited liability corporations registered in the calendar year. It is used to measure the level of entrepreneurship in the world and shows its dependence on several factors such as the politico-economic stability of the countries, the time to register a business, the number of taxes, etc. Research has proven that a high rate of new business creation promotes competition and economic growth. Decision-makers and entrepreneurs are looking into the issue in order to boost economic growth, especially since the global economic crisis between 2008 and 2009. Also the World Bank Group and the institution Doing Business have highlighted other factors favoring the creation of new enterprises as well as the formalization of those already existing. Among these factors is the density of new business, the number and importance of taxes, good governance, that is to say, a political bunker, the use of the Internet. The data used for this analysis are from 112 countries around the world. It turns out that the economic and political climate of a country plays a big role in the dynamism of entrepreneurship and that developed countries suffer more severely a fall in the creation of new enterprise in time of crisis than the developing countries (World Bank Group, 2010).

Researches of (Seher Khader, 2014) on the same issue highlight the fact that little research has been done on this subject, whereas particular attention should be paid to it, especially for developing countries wanting to boost their economies. They believe that identifying these factors and understanding their relationship would make it easier for country decision-makers to decide which path to take, what policy to put in place to boost their economies and lift their countries from poverty.

According to (Smith, 1971)) One of the great challenges in the prediction is that of the determination of the model to be adopted for the case in presence. Several factors come into play in the choice of the prediction method such as: the context of the prediction, the availability of the data, the degree of precision desired in the prediction, the period to be predated, the profit produced for the company for the use of the predictions as well as the time needed to do the analysis. In general, there are 3 types of methods used for predictions which are the qualitative technique (which uses qualitative data such as

people's opinions, and is often used in launching a new product in a given market). Time series analysis based entirely on patterns and their changes. It relies solely on past data and the causal models that focuses on the relationship between the elements. Here too, historical data is used a lot.
(Douglas C. Montgomery, 2008)Present in his book the importance of forecasting. According to him prediction is one of the most important performed tasks in an organization. Indeed, they make it possible to visualize and anticipate the future by considering certain parameters likely to influence the functioning of organizations. Although important, predictions are not always easy to make and are a real problem for organizations. They are all the more sensitive and important because they guide the decisions that can be made by the decision-makers. There are several types (economic forecast, technologic forecast and demand forecast) and methods used (qualitative and quantitative methods) for predictions
For (Jankovic, 2016)), every business evolves in a particular environment which impacts it. He classified factors in this environment in two main categories: internal (which is related to the business organization, the human resources, organizational culture, etc.) and external (economic, politic, demography, which cannot be controlled but have an impact on the business). In his research he found that those external factors vary from cities to cities and are difficult to list.
(Reynolds, 2007) in his book entitled Entrepreneurship in the United States: The future is now talked about the complexity of predicting new business creation in the United States. To achieve this, he gathered many variables that could have an impact on the firm birth rate in the United States. Forty-one independent variables where gathered and out of them, ten has shown a statistical significance. Those factors are grouped in five categories which are: the increase in demand (population growth and high level of personal or household income), career capacity (the high level educational attainment which has a great impact on economic growth), smaller firm prevalence (which has a significant impact on population), Economic sectors volatility and diversity (regions with large volatility and important small firm has significant impact on business creation as well as diversity in sectors) and flexible government policies on employment (regulation on hiring and firing employees if minimized have a strong impact on firm creation.
Despite of his finding, he emphasized that the presence of regional factors (individuals) influence firm's creation which varies from one region to another. Roughly, the temporal instability of firm's creation as well as all other aspects such as regional and other effects make the task of prediction of the creation of new business very complex and the determination of factors even more.
Regression analysis is one of the most commonly used models in predictions, which consists in determining the relationships between variables (a dependent variable called y) and the other or the others (as the case may be) called x independent variables) ((wikipedia.org, 2018)). When it comes to the comparison between a dependent variable and another independent one speaks of simple linear regression and in the case of a dependent variable and several independent we speak of multiple linear regression which makes it possible to determine the strength of the variables independent on the dependent variable, predicting the impact of changes in independent variables on the dependent variable and finally predicting future trends and values of the dependent variable (statisticssolutions.com,
2018)

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

(Andrej Krenker, 2011)), consider them as a mathematical model which tries to simulate the functionalities and structure of biological neural networks.In his article related to Artificial Neural Networks (Sakshi Kohli, 2014) Mentioned that they are the combination (interconnection) of artificial neurons. They are organized in layers (Wisconsin university, 2007). The way artificial neurons are connected is called topology, architecture or graph of an ANN. Though these interconnection can be done in many way according to the problem to solved, there are two basic classes of architecture which are: the simple feedforward topology (or acyclic graph in which information flows from input to output in only one direction (Feedforward Neural Network)) and recurrent topology (where some of the information flow from input to output and also in the opposite direction (Recurrent Neural Network)) (Andrej Krenker, 2011)

The biological neuron which is the basic building block of a biological neural network. Neurons receive information via dendrite, process them in the soma and send the output via an axon which is in contact with another neuron by a synapse (SakshiKohli, 2014) as we can see in the Figure

Similarly, to the biological neuron, artificial neuron is also the basic building block of every ANN and as a simple mathematical model (function), artificial neurons have three sets of rules which are multiplication (at the entrance, each input is weighted, means multiply with an individual value. Higher is the weight stronger will be the input which is multiplied by), summation (in the middle, the sum of all the weighted inputs as well as bias) and activation (at the exit, all the weighted sum and bias pass through an activation function also called transfer function) (Andrej Krenker, 2011). In the figure below, the working principle of an artificial neuron is provided. This one provides how the inputs are weighted, summed and pass through an activation function to give an output.

(Zhang, 1998)in his article “forecasting with artificial neural network: the state of art emphasized that some characteristics of neural networks as the adaptability, the function mapping and the non-linearity make them suitable and enable them to provides satisfactory result in forecasting. Since there are many research made to check the performance of neural networks compared to classical methods in forecasting, finding does not conclude as to when they are better than those methods. Talking about the measure of the performances, he mentioned that the prediction accuracy is the most important measure which is defined in forecasting as the difference between the actual or the desired value and the predicted one. Since there are many accuracy measures, each has advantages and limitations as well. Among them we have the mean absolute deviation, the sum of squared error, the mean square error, the root mean square error, etc. He highlights about the measuring of their performances which can be evaluates by of that the performances of these networks can be affected by some factors and concluded that neural networks need more data and computer time for training.

(Mijwel, 2018), in his article highlights about one of the limitation of neural networks called the “black box” nature which means that nobody knows how the neural network provides the final result. Thus, features used by the system to provide the output is still unknown. This represent actually an embarrassing situation for decision makers because they have to apply some solutions (predictions, etc) without any understanding of the pattern utilized by the system but only on relying on results they provide. (wikipedia.org, 2018), talking about the training, tests and validation in neural networks emphasized about a common task in machine learning which is the study and building of algorithms capable of learning from and realize predictions on data. According to (Shah, 2017), three datasets are often used in different levels of the model's creation. First we have the training dataset (used to fit the

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

parameters), then the validation dataset (which provides unbiased assessment of a model related to the training dataset while the hyper parameters (parameter whose values are set before the beginning of the learning process), and finally the test dataset (utilized to provide an unbiased assessment of the final model adapted on the training dataset).

(Ved, 2016) in his article, mentioned about the possibilities to improve (increase) the performances of a neural network. Among them, he named the addition of hidden layers, the increase of data, the change of the activation function, etc.

In his article (Santos, 2016) demonstrates that in the context of his research, the use of neural network offers better results than those obtained using the multiple linear regression. This is confirmed in many areas in which these two methods are used like in the financial sector as well according to the researches of (Nor Mazlina Abu Bakar, 2009)) despite of its black box nature which represent the biggest limitation of these systems (Mijwel, 2018)

(Management innovation, 2008), talking about decision making defined it as the process of selecting the right way to reach an objective and (Thefreedicionnary.com, n.d.), defined the decision as the conclusion obtained after consideration or an act of deciding according to (merriam-webster.com, 2018)). Decisions are classified in two according to (Management innovation, 2008): Programmed decisions (repetitive and often very structured) and nonprogrammed decision (generally one shot decision and less structured than the programmed ones).

(Boone, (2007).), (Juneja, Decision support system, 2018)highlight the fact that those two types of decisions are applied in organizations depending on the environment in which one is found and they classified them within organization in three main categories which are strategic decision (require a lot of time and information, Executive level decisions, Direct the organization to its destiny, Risky, Non repetitive), Operational decisions (middle and low level decision related to day-to-day activities and are repetitive) and Managerial decisions (combination between top and middle management and less complex than strategic decisions)

Several authors spoke about the decision making process in different ways and as part of them (Hussung, 2017) presented in his article seven steps to this process which are: the identification of the decision (recognition of the problem and choice of the appropriate decision to address it), the gathering of information (collection of facts and data related to the problem to be addressed) , the identification of alternatives (identification of available solutions), the evidence weight (assessment of the feasibility, acceptability and desirability of solutions), the choice among alternative (selection of the best solution), the action (implementation of the selected solution.) and the review of the decision (evaluation of the decision for effectiveness). The same author highlights challenges related to this process which he gathered in three and presents them as the overloaded of information: which is a risk while gathering information since it is important to select only relevant ones by fear to utilize useless information that might mislead the result, the misidentification of the problem (this will obviously leads to an unappropriated solution) and the overconfidence in the outcome (even if the process is fully followed, one should always be careful with the result. Sometimes, the outcome does not match expectations. Thus, identifying the right path is also very important).

In his article entitled decision making with the analytic hierarchy process (Saaty, 2008)highlights the fact that in the theory of decision, making a decision becomes more and more mathematical since to do so, many parameters must be taken into account. While this is the case, having a lot of information does

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

not guarantee effective decisionmaking because it is necessary to determine in addition which of these information should be considered in each case.(Nataliya Osipova1, 2015), emphasizes that it is difficult for human being to integrate all the parameters necessary for objective decision-making given the subjectivity that attaches to his nature. With the use of these decision support systems, it is possible to integrate different data in order to obtain objective answers. For example, she highlights the importance of decision-support systems in the economic field and states that the effectiveness of such a system in this field depends on the inputs and selected indicators.

Marek J. Druzdzel and Roger R. Flynn talking about Decision support system (Flynn, 2002)), mentioned that Decision-making in organizations (companies ...) has always been a real challenge with regard to all the parameters that must be considered to this purpose. This is the reason decision-support systems has been developed and is increasingly being used to improve the quality (precision) as well as the quantity (number) of decisions within a very short time. According to (Wikipedia, 2018), Decision support systems are information systems that assist in decisionmaking process within organizations, business.

(Ralph H. Sprague, 1980) Provides in his article a brief story of decision support systems and proposed a framework for their building. Having emerged in the 1970s, they were originally called the "management decision system", a name given by Michael S. Scott Morton. These systems have developed strongly over time and have expanded in several domains for example in army, hospital, etc., touching various aspects (forecasting, human resources management, etc).

(Juneja, Decision support system, 2018) talking about these systems, emphasized that they are costly for their realization (in time as well as in resources) are designed according to the problems they are called upon to solve.

Since, end-users are not always computer experts these systems offer easy-to-use solutions and generate results that make sense to them. Their realization requires constant monitoring by the system analyst as well as the end users according to the objectives pursued. He also noticed that there are different types of decision support systems depending on the situation that needs to be resolved. The categorization of these systems is based on several criteria; For example, we have categorization based on the data provided, based on the support offered, based on the types and frequencies of decisions taken. There is also the one that takes into account the decision framework. For this one we have six kinds of decision support systems which are text-oriented DSS, database-oriented DSS, spreadsheetoriented DSS, solver-oriented DSS, rule-oriented DSS, and compound DSS. A compound DSS is the most popular classification for a DSS; it is a hybrid system that includes two or more basic structures. There is still another categorization, the one based on the technologies used. Here, we have the model-driven decision support systems, the data-driven decision support system, the decision-driven support system, but all of them are generally based on the same architecture that is composed of the user interface, the database, a model (context or representation of the situation) and knowledge.(Juneja, Decision support system, 2018)With the progress in technology, decision support systems respond to much more complex problems. Although these decision-support systems are so successful and their usefulness is increasingly recognized and accepted in different organizations, it is important to point out that they are only helping the decision and not deciding instead. Despite the numerous advantages provided by Decision support systems as the time saving, the enhancement of effectiveness, interpersonal communication, .cost reduction in decision making process, the increase

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

decision maker satisfaction due to the insight, objectivity and time used for analysis performed by these systems which is very interesting for them, those systems also have some disadvantages as the difficulty to quantify all data because they use quantifiable data and make very hard the analysis of intangible data (some data cannot be defined in numbers), the unaware of assumption by decision makers who sometimes are not aware of the data utilized by the system to propose a solution. Taking decision without consideration of uncontrollable factors may be dangerous, the monetary cost due to the realization of these systems, the devaluation of subjectivity since they promote objectivity though a good decision is not always objective. The purpose of this point is to highlight the place of a DSS in the decision making process, we see that DSSs are useful only when the problem as well as the decision are clearly defined, the decision-makers identified, the information gathered as well as the alternatives taken. It is only in the end that we can associate the DSS for a better decision. The tendency which emerges from this intensive use of these systems is that users of these systems tend to indulge in the decisions provided by these systems, forgetting that sometimes, or even more often than not, certain decisions taken have a subjective and well-founded these systems, which are only objectives, cannot predict. This is why the decision-makers who use it must always retain the power to decide without always being totally dependent on these systems, although they are oh so beneficial.

(Kamel, 1998), while talking about the integration of a Decision Support System in the strategic public sector decision making for socio-economic environment in the public administration of Egypt, highlights about the satisfactory results as output of their integration despite the emerging of challenges in decision making, in the development process of the Decision Support System and in the adoption of the new solution by the civil servants within the administration.

3. Methodology

3.1. Research design

To address this matter, from the literature related to the topic, thirteen common factors to many countries having an impact on business creation were found. Data were collected over 27 countries randomly selected on the world form different sources and Excel 2013 was used to run the regression analysis and Neural Tools 7.5 of Palisade for artificial neural networks. In the prediction process for the DR Congo case, the first step was to build a model by using all the dataset and to apply it for the targeted country. The second step was to compare their performance by the measure of the mean square error and select the most accurate.

Finally, recommendations are made on possible a possible to integrate a decision support system using this model in the Ministry of National Economy which is in charge of this matter.

3.2. Data collection.

Nine years data (from 2008 to 2016) for twenty seven countries (New Zealand, Singapore, Denmark, Republic of Korea, Hong Kong, United Kingdom, Georgia, Peru, Turkey, Costa Rica, Macedonia, Chile, Paraguay, Philippines, Democratic Republic of Congo, Australia, Finland, Germany, Austria, Malaysia, Thailand, France, south Africa, Saudi Arabia, Jordan, Jamaica, Israel) has been collected from different sources (see reference)

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

4. **Analysis.**

4.1. **Factors impacting business creation.**

Since this indicator tis impacted by other factors (Index Mundi, n.d.)through the available literature related to the subject, here below are listed some of them with definition and meaning to entrepreneurship.

N°	Name	Meaning
1	New business density	A large number of business opening will lead to competition and means there are opportunities.
2	Unemployment	High unemployment could lead to more business creation.
3	GDP	Best way to measure the economy of a country.
4	Ease of doing business.*	The lower is the number the better is the environment. (1=most business-friendly regulations)
5	Internet users	Impact of internet is obvious on business.
6	Tax payment	More tax means less friendly business environment.
7	Other taxes payable by businesses	Idem
8	Time required to start a business	Shorter is the time, friendly is the environment.
9	Inflation, consumer prices (annual %)	High inflation means very bad environment.
10	Cost of business start-up procedures	High cost means bad environment.
11	Start-up procedures to register a business.	Higher is the number worse is the environment.
12	Population density.	High density means more population, high demand.
13	Population growth (annual %)	The growth of the population means more opportunity and then business creation.

Table 1: Indicators.

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

(*) Ease of doing business: the 10 indicators compiled by Simeon Djankov ((wikipedia.org, 2018)) at the World Bank Group are:

1. Starting a business: time, procedure, cost and capital (minimum) to open a business.
2. Dealing with construction permit: time and cost procedures to build a warehouse
3. Getting electricity: all the requirement to obtain electricity for a new business permanently for a newly constructed warehouse.
4. Registering property: requirements to register commercial real estate.
5. Getting credit: credit information index.
6. Protecting investors: extent of disclosure indices,
7. Paying taxes: all about taxes (for example: number) total payable taxes as share of gross profit
8. Trading across borders: documents (numbers), cost, time to import or export.
9. Enforcing contracts: everything regarding the enforcement of debt contract.
10. Resolving insolvency: time, cost, recovering rate under bankruptcy procedures.

4.2. Model building and prediction.

4.2.1. Multiple Linear Regression (MLR)

☐ Model building with Excel 2013

After running eleven regressions on our dataset to build the model (one dependent variable and thirteen independent for twenty-seven countries) using Microsoft Excel:**97% of variations** (R^2) of the dependent variable **new business registered** are explained by the three independent variables **new business density, GDP** and **individual using the internet**.

1. The prediction formula is therefore (refer to the regression coefficient) $\mathbf{Y = 1675.197 + 5209.623.X_1 + 0.0000000215.\ X2 - 183.276X_3 + E}$

☐ Prediction:

The above formula, which is actually the model, while applied on the target country for prediction provides the following results presented in the table 3:

Table 2: Data of the Democratic Republic of the Congo.

Year	New businesses registered (number)	New business density (new registrations per 1,000 people ages 1564)	GDP (current US$)	Individuals using the Internet (% of population)
2008	534	0.01	19.2 B	0.44
2009	705	0.02	18.3 B	0.56
2010	754	0.02	20.5 B	0.72
2011	413	0.01	23.8 B	1.2
2012	819	0.02	27.5 B	1.679961
2013	565	0.016059	32.7 B	2.2

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

2014	1613	0.044433	35.9 B	3
2015	1701	0.043735	37.9 B	3.8
2016	1565	0.038851	35.4 B	6.209974

This table contains only the data related to the model built using Excel 2013.
A summary of facts and predicted values is presented in this table as well as a chart on the next page to better view the result:

Year	New businesses registered (number) Initial value	New businesses registered (number) Predicted value
2008	534	**2059.967**
2009	705	**2069.77**
2010	754	**2089.092**
2011	413	**2020.594**
2012	819	**2062.502**
2013	565	**2058.749**
2014	1613	**2129.795**
2015	1701	**2022.583**
2016	1565	**1500.878**

Table 3: Facts vs predicted values (MLR)
From this table it is obvious that there is a big difference existing between facts and predicted values. To better view the trends of each, a figure is provided below.

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

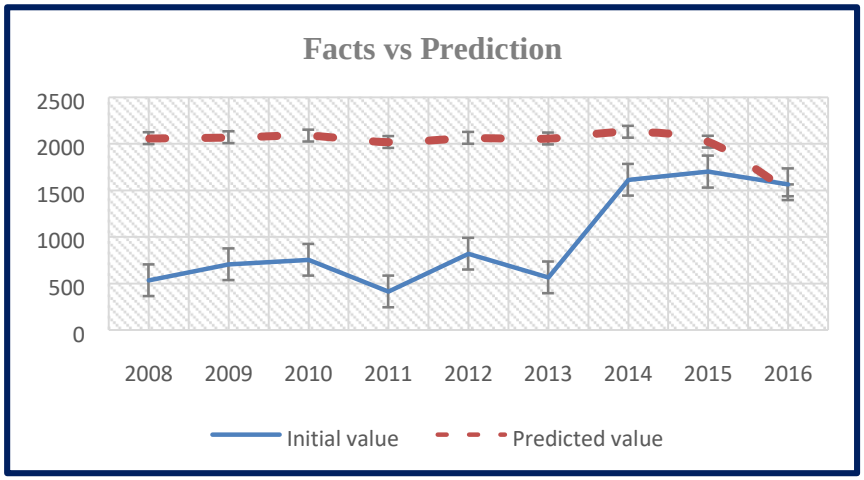

Figure 1: Facts vs Prediction

4.2.2. Artificial Neural Network

☐ Model building with Palisade Neural Tools suite 7.5 student version.

To use ANN there is a training and a testing process on the dataset. The Neural Tools of Palisade, which is used as Add-in in Excel, offers the same process and execute it automatically (80 percent trained, 20 percent tested). This software does at the same time linear and nonlinear prediction (http://www.palisade.com/GuidedT our/EN/NeuralTools/): in providing the root mean square error, which is the standard deviation of the prediction error (residual), or the measure of the difference between predicted and actual values (Stephanie, 2016). Thus, it shows how well the model performs (Lin, 2018). A lower value means better fit. (GRACE-MARTIN, 2018)

After using the above mentioned process on the 27 countries dataset to build the model, the result is as followed:

Linear Predictor vs Neural Network		
Linear Predictor		Neural network
R-square training	0.6753	-
Root mean square error (Training)	65809.23	1386.09
Root mean square error (Testing)	49818.63	3326.42

Table 4: Training and testing results with Neural Tools for 27 countries and 14 variables

With palisade, this built model is copied and applied on the case study using all the variables presented in the table below:

• Prediction.

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

New businesses registered (number)	New business density (new registrations per 1,000 people ages 15-64)	GDP (current US$)	Inflation, consumer prices (annual %)	Unemployment, total (% of total labor force) (national estimate)	Tax payments (number)	Other taxes payable by businesses (% of commercial profits)	Population density (people per sq. km of land area)	Population growth (annual %)	Start-up procedures to register a business (number)	Time required to start a business (days)	Cost of business start-up procedures (% of GNI per capita)	Individuals using the Internet (% of population)	Ease of doing business (index)
534	0.01	19.2 B	17.3	3.7	40	196.4	26.63091	3.293549	15	132.5	935.4	0.44	158
705	0.02	18.3 B	2.8	3.7	40	225.9	27.52892	3.316441	15	126.5	847.6	0.56	181
754	0.02	20.5 B	7.1	3.7	40	272.3	28.46133	3.330936	11	84.5	735.1	0.72	182
413	0.01	23.8 B	15.3	3.7	40	272.3	29.42749	3.338296	11	65.5	551.4	1.2	175
819	0.02	27.5 B	9.72	4.49	40	272.3	30.42663	3.338872	11	58.5	284.7	1.68	178
565	0.02	32.7 B	1.6	3.7	40	14.5	31.45764	3.332367	12	31.5	200.1	2.2	181
1613	0.04	35.9 B	1.2	3.7	50	14.5	32.51929	3.319176	8	16.5	30	3	183
1701	0.04	37.9 B	1	3.7	52	14.5	33.61047	3.300416	7	11.5	29.3	3.8	184
1565	0.04	35.4 B	18.2	3.7	52	14.5	34.73066	3.278534	7	11.5	29.3	6.20	184

Table 5: Data of the Democratic Republic of the Congo

Result: once the model is applied in the testing stage, the output is provided in the table below.

Neural network	
Root mean square error (Testing)	113.80
Mean absolute error	40.49
Standard deviation of the absolute error	106.35

Table 6: Testing result.

The NeuralTools showed that regression using all the variables is not possible due to the insufficient available data. On the contrary, ANN provides output as it is shown on the table11 where prediction for

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

each year is provided. The table and the chart below are providing a summary of prediction and facts. From these, it is obvious that prediction are very closed to facts.

New businesses registered (number)	534	705	754	413	819	565	1613	1701	1565
Prediction	534	705	413	413	819	565	1624.72	1689.28	1565
Years	2008	2009	2010	2011	2012	2013	2014	2015	2016

Table 7: Facts vs prediction with ANN

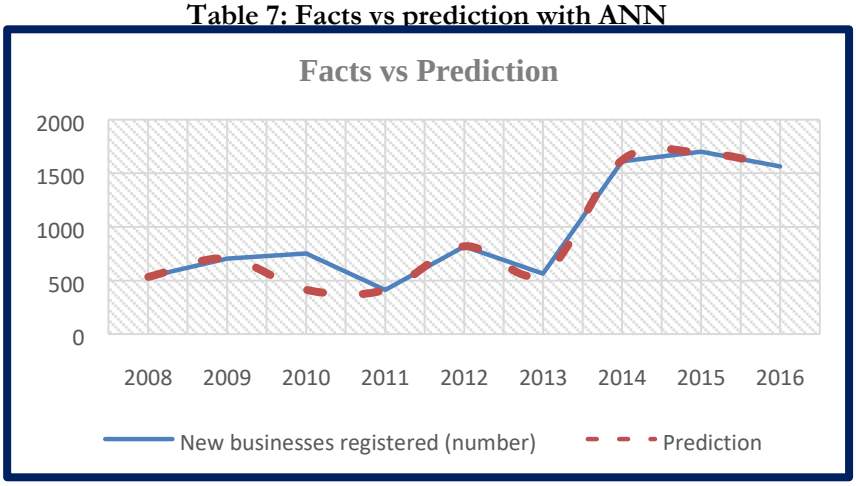

Figure 2: New business registered, Facts vs Prediction (DR Congo).

4.2.3. Prediction using three independents variables.

Results of the regression analysis show that only three variables are significant for prediction. Since this study aims at comparing the performance of each model, in this point, the same process applied for the thirteen variables will followed on the three variables for each models and the output compared.

A. ANN Model building.

Neural Networks is now applied to evaluate it performances on the three independent variables used to predict the case of the DR Congo. To perform this, the dataset for 27 countries with the four variables has to be trained and tested. Result is as followed:

Linear Predictor vs Neural Net		
	Linear Predictor	Neural Net
R-Square (Training) Root Mean Sq. Error (Training) Root Mean Sq. Error (Testing)	0.5896 73068.54 74171.08	-- 3308.54 16577.14

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

Table 8: Training and testing results. Obviously, neural networks perform better than multiple linear regression.

b. Prediction.

The model built above is now applied on the dataset of DR Congo (Table 2)

Results: the table below shows that ANN performs better training and bad testing compared to MLR.

Linear Predictor vs Neural Net		
	Linear Predictor	Neural Net
R-Square (Training) Root Mean Sq. Error (Training) Root Mean Sq. Error (Testing)	0.9819 56.89 82.77	-- 30.57 105.79

Table 9: Training and testing results (DR Congo, three variables)

This result can be better viewed in the table below which provides a comparison between the facts and the prediction and a chart which shows the trends of the two dataset.

New businesses registered (number)	534	705	754	413	819	565	1613	1701	1565
Prediction	823.40	832.52	847.49	895.87	960.11	1051.71	1246.43	1515.69	2880.45
Years	2008	2009	2010	2011	2012	2013	2014	2015	2016

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

Table 10: Facts vs predicted data with ANN (three variables)

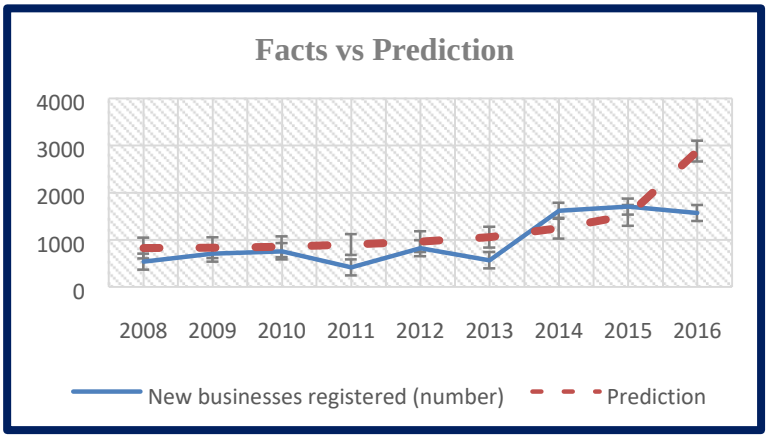

Figure 3: Facts vs Prediction with ANN for four variables; DR Congo case.

c. MLR with three independents variables. (See point 3.2.1. MLR, prediction).

4.2.4. Overall summary.

Now that the model building as well as prediction using both models has been provided, a summary of both results is provided in the table below.

	MLR	ANN	Observation
27 countries (3 or 13 variables)	Model building.		
DR Congo 13 independent variables	Impossible to predict with those variables. Only three variables can be utilized and provide bad prediction.	Good prediction (accurate results)	Better prediction with ANN (more accurate) than MLR
DR Congo 3 independents variables	Possible prediction	Possible prediction	Both are not good with those variables.

Table 11: Overall summary

4. Discussion and Recommendation

4.1. Discussion.

After working throughout the above process, results are showing that:

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

1. Multiple linear regression (MLR) since useful and effective when dealing with data showing a linearity cannot be utilized for the present purpose due to:
a. The presence of outliers and variables statistically insignificant though theses one were confirmed as very significant by the literature.
b. The bad prediction provided when compared to the facts;
c. The limitations of this model which is useful when dealing with linear data. In the present case, which is actually a complex one, the data trend is not linear and vary a lot as the time is going on.
2. Artificial Neural Network can be utilized because:
a. It offers less error compared to MLR, means more accuracy in prediction, sometimes very close to the real data;
b. It works on both linear or nonlinear data set, thus allows performance comparison;
c. It is easier to use compared to MLR;
d. It offers the possibility to upgrade the model, to add more variables, layers and data which will have the positive impact of increasing prediction's accuracy.

Since it was clearly demonstrated that it is possible to use those two models in some cases, the selection of the appropriate one should be made by comparing their performance. One of the common measure of prediction accuracy is the mean square error which is the squared difference between the predicted value and the initial one (Tutorvista.com, 2018). Thus, in the table below is provided the result of the MSE for each model used in this research and a brief explanation about each:

	Country	**DR Congo**	**Observation**
MSE	**MLR 13 VAR**	**ANN 13 VAR**	Compared to MLR, in both case, ANN has a small value which means it is more accurate than MLR. Also, using more significant variables (in this case 13) improve the performance of the model.
	Unusable	**12,920.11**	
	MLR 3 VAR	**ANN 3 VAR**	
	5,896,743	631,850.11	

Table 12: Models accuracy (MSE)

From this table it is obvious that ANN using thirteen variables is more accurate and then can be utilized to predict the number of new business registered. This does not mean that only those variables are significant for prediction but that if there are more significant variables added as input the prediction accuracy will certainly increase.

4.2. Recommendation.

This complex topic related to the prediction of new business registered can find a positive response if the artificial neural network is used as solving model. Thus, to address this issue for the targeted country, the implementation of a decision support system using this model and considering significant factors can be utilized. Decision makers will certainly find in it a tool capable of helping them in choosing the appropriate policy and assessing it possible impact in the future. However, further studies have to be conducted in order to improve the efficiency of such tool, this due to the limitations of Artificial Neural Networks.

American Journal of Information Technology and Management
Volume 10 Issue 1, January-March 2022
ISSN: 2837-1038
Impact Factor: 6.44
Journal Homepage: https://americaserial.com/Journals/index.php/AJITM,
Email: contact@americaserial.com
Official Journal of America Serial Publication

5. Limitation and Conclusion

Finding factors for this study is the result of reading articles debating on the subject. Since, each country has its own economic realities; it would have been interesting to be in the country of the research to determine those factors. Due to that limitation, only major factors found in the available literature have been used. Data used for analysis are secondary data obtained from various sources such as the World Bank open database, knoema.com a site that also provides some economic indicators, as well as Wikipedia.org from which the list of countries as well as their classification according to Doing Business's index of ease of doing business. For the recommendation of a decision support system to address this matter, only an overview of the system is provided since these systems is a long and tedious task, requiring a lot of resources (people, expertise, finances, etc.) and a close collaboration between the beneficiaries of the system as well as their designer. (Juneja, Decision support system, 2018) Not being in direct contact with the direct decision makers on the subject, it is not possible to provide a definitive solution

This study is an attempt to find a model able to predict the number of new business registered. This subject has very few literature and the outcome of our search for literature shows that there is no available paper or other sources that talk about the use of artificial neural network for this purpose. Thus, the finding must be considered with caution and subject to further examinations with regard to all the restriction mentioned above.

This research was focused on the comparison of performance between MLR and ANN to predict the number of new business registered especially for the case of the Democratic Republic of the Congo. Defined as the number of new limited liability corporations registered in the calendar year this number (World Bank's Entrepreneurship Survey and database) the new business registered is an established by the World Bank Group as one of the indicators to assess the level of entrepreneurship in a country. (Index mundi). Indeed, entrepreneurship is of a great support to the economy of nations and it plays an important role in their economic growth. Thus, having a look on this indicator to understand existing factors that may impact it will help decision makers (policies makers) to adopt the right decision to improve the economic environment favorable to entrepreneurship in the country.

After working through a process of data collection, analysis of both models, prediction for our targeted country and a performance comparison as well as a selection of the appropriate model, results show how robust is the artificial neural network in prediction compared to multiple linear regression, thus confirming it good reputation for this task. The next step will consist in building a decision support system using this model with relevant variables to be used as a tool in the decision making process related to this matter. Though there is still much to consider before launching such project, artificial neural network offers a significant light to complex subject of prediction.

Acknowledgement

I would like to acknowledge my thesis advisor, Professor **Cho Daiyon** (Handong Global University) for his guidance and wise advices throughout my master degree thesis writing which talks about the very topic discussed in this paper.

6. Appendix:

Data Sources:

1. Annual data from the World Bank's open database (nine years data for each variable),

2. Wikipedia.org where we obtained the ranking of countries for the ease of doing business index (annual data),
3. Koema.com where we have completed the annual data on the inflation rate from 2015 to 2017 (https://knoema.com/atlas/Democratic-Republic-of-the-Congo/Inflation-rate) as well as the rate of lack of employment for the years 2006 and 2007 as well as 2015 to 2017 (https://knoema.com/atlas/DemocraticRepublic-of-the-Congo/Unemployment-rate) for the Democratic Republic of the Congo.
4. Unemployment for Saudi Arabia and Jordan from https://www.statista.com/statistics/262524/unemploymentrate-in-saudi-arabia/, https://www.statista.com/statistics/385565/unemployment-rate-in-jordan/.

7. References

Andrej Krenker, J. B. (2011). Introduction to the Artificial Neural Networks. Faculty of Electrical Engineering, University of Ljubljana, 1-18.

Anthony GIUSTINI, A. B. (2016). Doing business in the Democratic Republic of Congo. Clifford Cance, 1-13.

Babs, G. (2013, 12 16). Quatre mois après son instauration, 1400 entreprises avaient déjà expérimenté le guichet unique de création d'entreprises. Retrieved from http://business-et-finances.com/: http://business-etfinances.com/quatre-mois-apres-son-instauration-1400-entreprises-avaient-deja-experimente-le-guichetunique-de-creation-dentreprises/

Boone, D. J. ((2007).). A Leader's Framework for Decision Making. . harvard business review, 1-10.

CFEF. (2016, 06 27). PDPC :DOING BUSINESS 2017, 5 REFORMES REALISEES. Retrieved from http://www.cfef.cd: http://www.cfef.cd/pdpc-doing-business-2017-5-reformes-realisees/

Douglas C. Montgomery, C. L. (2008). Introduction to Time Series Analysis and Forecasting. WIley.

Flynn, M. J. (2002). Decision Support Systems. . Encyclopedia of Library and Information Science, Second Edition., 1-15.

Forum des as. (2016, 02 23). VOICI LE PLAN DE KABILA POUR L'ÉMERGENCE DE LA RDC EN 2030. Retrieved from http://www.forumdesas.org: http://www.forumdesas.org/spip.php?article6771

GRACE-MARTIN, K. (2018). Assessing the Fit of Regression Models. Retrieved from https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/

Hussung, T. (2017, 02 10). decision making process. Retrieved from https://online.csp.edu: https://online.csp.edu/blog/business/decision-making-process Index Mundi. (n.d.). New

businesses registered (number). Retrieved from https://www.indexmundi.com: https://www.indexmundi.com/facts/indicators/IC.BUS.NREG

Jankovic, M. (2016). INFLUENCE OF EXTERNAL FACTORS ON BUSINESS. Ekonomika, 31-38.

Juneja, P. (2018). Decision support system. Retrieved from https://www.managementstudyguide.com/: https://www.managementstudyguide.com/decision-support-systems.htm

Juneja, P. (2018). Decision support systems. Retrieved from https://www.managementstudyguide.com: https://www.managementstudyguide.com/evaluating-decision-support-system-projects.htm

Kamel, S. (1998). Decision Support Systems and Strategic Public Sector Decision Making in Egypt. Institute for development policy and management, 1-20.

Lin, M. (2018). What-is-the-meaning-of-root-mean-squared-error-RMSE-in-statistics. Retrieved from https://www.quora.com/What-is-the-meaning-of-root-mean-squared-error-RMSE-in-statistics

Mapon, M. P. (2012, 11 01). Ministerial Decree. DECRET N ' 12/045 du 01/11/2012 portant creation, organisation et fonctinnement du guichet unique de creation d'entreprise. Kinshasa, Kinshasa, Democratique republic of congo: Prime minister.

Mediacongo.net. (2017, 05 21). Guichet unique: Baisse du coût de création d'entreprises en RDC. Retrieved from http://www.congoactuel.com/guichet-unique-baisse-du-cout-de-creation-dentreprises-en-rdc/

Merriam-webster.com. (2018). Decision. Retrieved from https://www.merriam-webster.com: https://www.merriamwebster.com/dictionary/decision

Mijwel, M. M. (2018, 01 27). Artificial Neural Networks Advantages and Disadvantages. Retrieved from https://www.linkedin.com/pulse/artificial-neural-networks-advantages-disadvantages-maad-m-mijwel

Nataliya Osipova1, V. (2015). Design, Development and Use of Decision Support Systems in the Study of Economic Disciplines in Higher Education. . 1-16.

Nelson, D. (2016, 05 10). How Does A Tax Base Grow? Retrieved from https://www.amarilloedc.com: https://www.amarilloedc.com/blog/how-does-a-tax-base-grow

Nor Mazlina Abu Bakar, I. M. (2009). Applying Multiple Linear Regression and Neural Network to Predict Bank Performance. International Business Research, 1-8.

Reynolds, P. D. (2007). Entrpeneurship in united state: the future is now. Miami: Springer.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. Int. J. Services Sciences, 83-98.

Sakshi Kohli, ,. R. (2014). BASICS OF ARTIFICIAL NEURAL NETWORK. International Journal of Computer Science and Mobile Computing, 745-751.

Santos, S. D. (2016). Artificial neural networks and multiple linear regression model using principal components to estimate rainfall over South America. . Nonlin. Processes in Geophys,, 1-8.

Seher Khader, R. R. (2014). Macro-Economic Factors Affecting Ease of Business. Georgia Institute of Technology, 1-16.

Smith, J. C. (1971). How to Choose the Right Forecasting Technique. Retrieved from https://hbr.org: https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique statisticssolution.com. (2018). What is Multiple Linear Regression? Retrieved from https://www.statisticssolutions.com/what-is-multiple-linear-regression/ statisticssolutions.com. (2018). What is Multiple Linear Regression? Retrieved from https://www.statisticssolutions.com/what-is-multiple-linear-regression/ Stephanie. (2016, 10 25). What is Root Mean Square Error (RMSE)? Retrieved from https://www.statisticshowto.datasciencecentral.com: https://www.statisticshowto.datasciencecentral.com/rmse/

Tutorvista.com. (2018). Mean square error. Retrieved from https://math.tutorvista.com: https://math.tutorvista.com/statistics/mean-squared-error.html

Ved. (2016 , 09 29). How to improve performance of Neural Networks. Retrieved from https://d4datascience.wordpress.com: https://d4datascience.wordpress.com/2016/09/29/fbf/ wikipedia.org. (2018, 11 01). Ease of doing business index. Retrieved from https://en.wikipedia.org/wiki/Ease_of_doing_business_index wikipedia.org. (2018, 10 8). Regression analysis. Retrieved from https://en.wikipedia.org/wiki/Regression_analysis

World Bank Group. (2010). Entrepreneurship Snapshot 2010: Measuring the impact of financial crisis on new business registration. Washington: World Bank Group.

Zhang, G. (1998). Forecasting with artificial neural networks: the state of art. Elsevier, 1-13.