



**NOVA**

**IMS**

Information  
Management  
School

# MGI

---

**Mestrado em Gestão de Informação**

Master Program in Information Management

**Competitive Intelligence Algorithm**

**for PESTLE Analysis**

A Decision Support System Application module

Elson Neves Simões

Project Work presented as partial requirement for obtaining  
the Master's degree in Information Management

NOVA Information Management School  
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

**NOVA Information Management School**

**Instituto Superior de Estatística e Gestão de Informação**

**Universidade Nova de Lisboa**

COMPETITIVE INTELLIGENCE ALGORITHM FOR PESTLE ANALYSIS

**by**

**Elson Neves Simões**

Project Work report presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence.

**Advisor: Mauro Castelli**

**Co Advisor: Luis Madureira**

November 2019

## **DEDICATION**

To my mother Isabel, to whom I will be forever grateful for giving me the most precious thing I have, my education and my principles.

To my brothers Alexandre and Lucas of whom I am so proud.

To my best friend and girlfriend Catarina Baptista, thank you for all your support and patience with me, you make everything easier.

And to all my friends which I am very lucky to have in my life.

## **ABSTRACT**

In today's competitive environment, board directors of companies strive to make the right decisions that will put them ahead of their competitors. However due to the volatility of the company's external environments any decision has its inherent risks. This thesis aims to apply the Python Programming Language to develop a specific module of a Decision Support Tool that intends to predict market movements, trading position and potential, and operations orientation. The module that will be the scope of this thesis will consist in the creation of a predictive model for countries political factors based in public databases with political indexes all over the globe. This module refers to the first factor to consider in a PESTLE analysis.

## **KEYWORDS**

DSS; Competitive Intelligence; PESTLE; Python; Neural Networks

# INDEX

|   |    |
|---|----|
| 1. INTRODUCTION .....   | 10 |
| 2 LITERATURE REVIEW .....   | 11 |
| 2.1 DIKIW Hierarchy.....  | 11 |
| 2.1.1 DATA.....   | 12 |
| 2.1.2 INFORMATION.....  | 13 |
| 2.1.3 KNOWLEDGE.....  | 14 |
| 2.1.4 INTELLIGENCE.....   | 15 |
| 2.1.5 WISDOM.....   | 15 |
| 2.2 COMPETITIVE INTELLIGENCE .....  | 16 |
| 2.3 ARTIFICIAL NEURAL NETWORKS .....  | 19 |
| 2.4 NEURAL NETWORK ARCHITECTURES .....                                      | 22 |
| 2.5 PESTLE .....  | 23 |
| 2.6 COMPETITIVE ANALYSIS TOOLS .....  | 27 |
| 2.6.1 Porter’s Five Forces Analysis .....                                   | 28 |
| 2.6.2 SWOT Analysis.....  | 30 |
| 2.6.3 PESTLE Analysis.....  | 31 |
| 2.7 IMPORTANCE OF PESTLE ANALYSIS .....                                     | 33 |
| 2.8 IMPORTANCE OF AUTOMATING PESTLE ANALYSIS .....                          | 36 |
| 2.9 IMPROVEMENT OF PESTLE ANALYSIS USING NEURAL NETWORKS .....              | 38 |
| 2.10 DECISION SUPPORT TOOLS .....   | 39 |
| 2.11 PYTHON PROGRAMMING AND ITS APPLICATION TO DECISION SUPPORT TOOLS ..... | 41 |
| 3 METHODOLOGY .....   | 43 |
| 3.1 DATA SOURCES.....   | 43 |
| 3.2 DATA MANIPULATION.....  | 44 |
| 3.3 NEURAL NETWORK CONSTRUCTION AND TRAINING .....                          | 46 |
| 3.3.1 Defining predictors and targets.....                                  | 47 |
| 3.3.2 Define the Neural Network model .....                                 | 48 |
| 3.3.2.1 Comparing performances .....  | 49 |
| 3.3.3 USAGE OF THE MODEL_2 TO PREDICT THE 2018 INDEXES .....                | 49 |

|        |   |    |
|--------|---|----|
| 4      | RESULTS AND DISCUSSION .....                                  | 50 |
| 5      | CONCLUSIONS .....   | 52 |
| 6      | LIMITATIONS AND RECOMMENDATIONS FOR FUTURE DEVELOPMENTS ..... | 53 |
| 7      | Bibliography .....  | 54 |
| 8      | Appendix .....  | 58 |
| 8.1    | Code screenshots .....  | 58 |
| 8.1.1  | Worldwide Governance Indicators Manipulation .....            | 58 |
| 8.1.2  | Index of Economic Freedom .....                               | 59 |
| 8.1.3  | Transparency International .....                              | 60 |
| 8.1.4  | Doing Business Law Library .....                              | 61 |
| 8.1.5  | General Manipulation .....                                    | 62 |
| 8.1.6  | Validation of the predictors shape .....                      | 63 |
| 8.1.7  | ADAM TESTING .....  | 64 |
| 8.1.8  | SGD TESTING .....   | 65 |
| 8.1.9  | Comparing performances of model_1 and model_2 .....           | 66 |
| 8.1.10 | Results of the model_1 (Adam with lr=0.000015) .....          | 67 |
| 8.1.11 | Results of the model_2 (SGD with lr=0.000015): .....          | 68 |

## LIST OF FIGURES

|   |    |
|---|----|
| Figure 1 DIKIW schema .....   | 12 |
| Figure 2 Formation of Information and Data (from Liew, 2007) .....  | 14 |
| Figure 3 Intelligence Cycle (from Garcia-Alsina et al., 2016) .....   | 17 |
| Figure 4 The competitive intelligence process (from Pellissier & Nenzhelele, 2013) .....  | 18 |
| Figure 5 Mammalian neuron (Ajith Abraham, 2005).....  | 20 |
| Figure 6 Architecture of an artificial neuron and a multi-layered neural network (Ajith Abraham, 2005)<br>.....                                 | 21 |
| Figure 7 Different types of activation functions: (a) threshold, (b) piecewise linear, (c) sigmoid and (d)<br>Gaussian (Jain et al., 1996)..... | 21 |
| Figure 8 A typical feed-forward network architecture. (Jain et al., 1996).....  | 22 |
| Figure 9 Backpropagation steps (Jain et al., 1996) .....  | 23 |
| Figure 10 The format of a systemic PEST analysis diagram (HO, 2014).....  | 27 |
| Figure 11 Data manipulation schema .....  | 45 |
| Figure 12 Input and target schema.....  | 45 |
| Figure 13 Neural network construction with KERAS.....   | 46 |
| Figure 14 Screenshot of the final data structure .....  | 47 |
| Figure 15 Comparison graphic of the validation scores of model_1 and model_2 .....  | 49 |
| Figure 16 Screenshot of the prediction's execution .....  | 49 |
| Figure 17 Power BI map with the top 10 of the predictions.....  | 51 |

## LIST OF TABLES

|  |    |
|--|----|
| Table 1 Von Neumann computer versus biological neural system (Jain et al., 1996) ..... | 19 |
| Table 2 Detailed sub-factors of the PESTLE framework.....                              | 26 |
| Table 3 Political Indexes data sources .....   | 44 |
| Table 4 Years of Predictors and Targets .....  | 47 |
| Table 5 Prediction Values.....   | 50 |
| Table 6 Comparison between 2017 real values against 2018 predictions .....             | 50 |



## LIST OF ABBREVIATIONS AND ACRONYMS

|               |  |
|---------------|--|
| <b>AI</b>     | Artificial Intelligence                    |
| <b>ANN</b>    | Artificial Neural Network                  |
| <b>BI</b>     | Business Intelligence                      |
| <b>CI</b>     | Competitive Intelligence                   |
| <b>DSS</b>    | Decision Support Systems                   |
| <b>ICT</b>    | Information and communication technologies |
| <b>IoT</b>    | Internet of Things                         |
| <b>GP</b>     | Genetic Programming                        |
| <b>NN</b>     | Neural Network                             |
| <b>PESTLE</b> | Marketing framework                        |

# 1. INTRODUCTION

Nowadays companies strive to beat their competitors in many ways; mergers and acquisitions; exports; research and development and many other possible moves with the purpose of beating their competitors and be the best in their industry. However, there is an always inherent risk associated with any move that pursues a profitable outcome, and that risks are mainly due to the speculation board directors need to do to take the right decisions. Those speculations are far from being simple ones, the variables they need to consider are uncountable going from political speculation to technological, economic, social considerations and so on.

This thesis is part of a project that has the intent to diminish the uncertainties of a company when it comes to decision making with the development of a decision support system software that will provide board directors with the framework and data analysis from open sources databases and any information available online to weight and reduce as much as possible any speculation in any decision they could make for the future of their companies.

Considering the complexity of such software the scope of this thesis it will be focused in only one of the modules of the software that consists in a Competitive Intelligence Algorithm based in a framework used by marketers to analyse and monitor the macro-environmental factors that have an impact on an organization, the PESTLE analysis.

It is a self-explanatory framework hence its name is built through the initials of each factor it considers for making the analysis of the external environment of a company: P stands for the political factors; E regards the economic factors; S stands for social factors; T stands for technological factor; L for legal factors and E for environment.

The programming language chosen to build this, and the following modules, is python and the reasons to be this language are related with its efficiency and simplicity and for having an active community and being part of the academia world. For this specific module it will be used an artificial intelligence algorithm known as neural network to meet the de desired outcome.

## 2 LITERATURE REVIEW

### 2.1 DIKIW Hierarchy

“Data is fool’s gold. Insight is pure gold” (Shepherd M. 2018)

Considering that the main purpose of the DSS application, and consequently this module is to get insights from data, meaning to transform data into wisdom, for companies to reduce their risks when it comes to take decisions that can have great financial impact, it is imperative that I start by exposing the origin of the framework that explains the process of that transformation.

The first mention of a framework that exposes the transformation from Data to Information and from Knowledge to Wisdom was not mentioned within the Knowledge Management domain nor Information Science but in poetry (Sharma, 2008). The poet T.S. Eliot was the first to mention the “DIKW hierarchy”. In 1934 Eliot wrote in “The Rock”: (1).

“Where is the Life we have lost in living?

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?”

Cleveland (1982) in his article “Information as Resource” uses Eliot’s hierarchy as a starting point to explain the basic terms of the DIKW hierarchy.

In Anthony Liew (2013) adds another layer to the hierarchy, Intelligence, giving origin to the DIKIW hierarchy that result in the schema below.

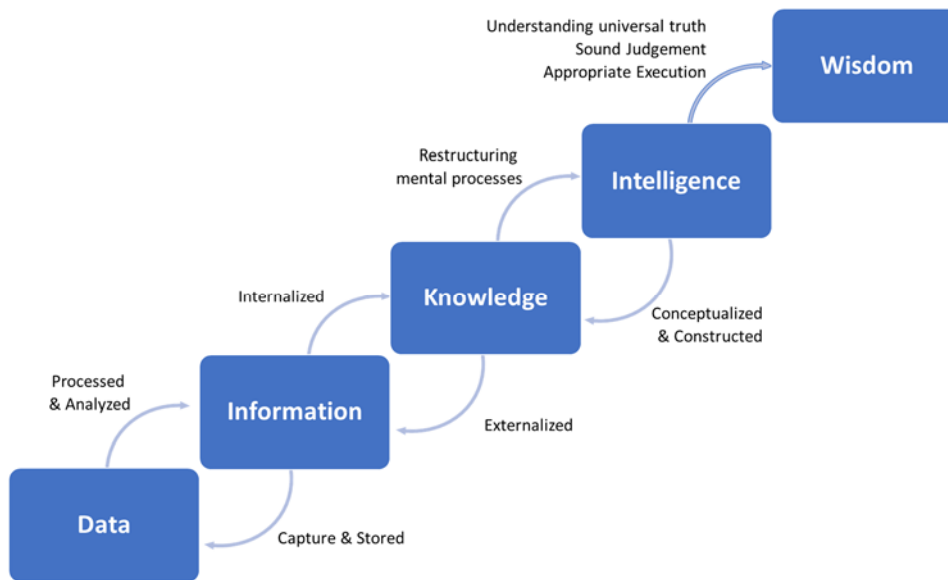


Figure 1 DIKIW schema

## 2.1.1 DATA

Data is a set of values of subjects with respect to qualitative or quantitative variables.

The Latin word data is the plural of datum, “(thing) given,” neuter past participle of *dare* “to give”.<sup>1</sup>

Taking in consideration that data is the foundation of wisdom it is curious that the origin of the word itself means “to give” when nowadays data can give us infinite insights about any scientific and non-scientific domains. In June 2018, 55.1% (4,208 million of users) of the world population had access to Internet.<sup>2</sup>

Due to the constant technology evolution and with the constant increase of internet users, data generation has given place to what is called Big Data. Today we produce around 2.5 quintillion bytes of data each day<sup>3</sup> and with the rise of IoT this number will only increase the generation of data from infinity of connected devices.

Big data can be divided in four dimensions: Volume, Velocity, Variety and Veracity.<sup>4</sup> As per Kenneth Neil Cukier and Viktor Mayer-Schoenberger in their book “Big Data: A Revolution That Will Transform How We Live, Work, and Think”, “Causality won’t be discarded but it is being knocked off its pedestal as the primary foundation of meaning. Big data turbocharges non-casual analyses, often replacing casual investigations.” This

observation made by the authors is exactly what the globality of this project intends to address, provide to the C-suite of a company a DSS application that will put aside any casual analyses but instead give them a data-driven measurable risk of their management decisions.

## 2.1.2 INFORMATION

Today there is no consensus across disciplines about a single definition of information. That lack of consensus resides for example in the different definitions about information that a physician a poet and a data scientist can have. Through the development of this thesis the definition of Information will be carried as an Information Science approach.

The origin of the word Information comes from the junction of the Latin words *in* and *forma* that adapted to English are the words *into* and *shape* respectively. Therefore, information is no more than structure unstructured data, to give form to it.

“Information is contained in descriptions, answers to questions that begin with such words as who, what, when, where, and how many.”(Ackoff, 1989). Information tells us “about” something else. Random data wouldn’t be information, it would be “noise: random fluctuations that obscure or do not contain meaningful data” (<https://en.oxforddictionaries.com/definition/noise>).

Information is data presented in a context that yields meaning. The way data is structured is inherent to the context of its structuring. (Bates, 1999) One aspect that this project attempts to overcome is to avoid misleading conclusions through miscomprehension of data itself because one’s context.

“Information is a message that contains relevant meaning, implication, or input for decision and/or action. Information comes from both current (communication) and historical (processed data or ‘reconstructed picture’) sources. In essence, the purpose of information is to aid in making decisions and/or solving problems or realizing an opportunity.” (Liew, 2007).

The following diagram illustrates the relationship between data and information:

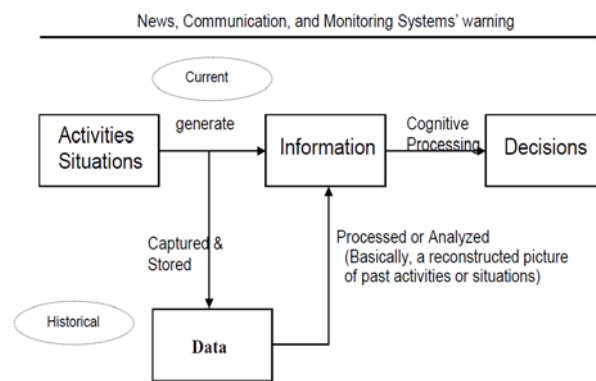


Figure 2 Formation of Information and Data (from Liew, 2007)

### 2.1.3 KNOWLEDGE

As per the Cambridge Dictionary, Knowledge is the awareness, understanding, or information that has been obtained by experience or study, and that is either in a person’s mind or possessed by people generally.

Knowledge “is information combined with experience, context, interpretation, and reflection. It is a high-value form of information that is ready to apply to decisions and actions.” (Brodie & Brodie, 2009) as cited in Davenport & Prusak, 1998) (Brodie & Brodie, 2009).

“is know-how and is what makes possible the transformation of information into instructions. Knowledge can be obtained either by transmission from another who has it, by instruction, or by extracting it from experience” Ackoff (1989).

Knowledge can be considered tacit knowledge or explicit knowledge. Tacit knowledge can only be transferred through socialization, shared experiences or practices. Explicit knowledge is tacit knowledge that is possible to be articulated to explicit knowledge and can be transferred through any type of information. (Nonaka Hirotaka, 1995).

In “Understanding Data, Information, Knowledge and Their Inter-Relationships”, A. Liew refers that the purpose of knowledge in the context of business is to create or increase value for the enterprise and all its stakeholders.

#### **2.1.4 INTELLIGENCE**

Merriam-Webster dictionary defines intelligence as the ability to learn or understand or to deal with new or trying situations. It also defines intelligence as the ability to apply knowledge to manipulate one’s environment or to think abstractly as measured by objective criteria. Ackoff in “Data to Wisdom” defines intelligence as the ability to increase efficiency.

Intelligence can be also defined by “the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioural sub-goals that support the system’s ultimate goal.”(Albus, 1991). The definition of Intelligence by James S. Albus can be directly applied to the corporate intelligence that major companies try to achieve. Their main goal is to increase their probabilities of success by taking the appropriate actions.

#### **2.1.5 WISDOM**

“Wisdom deals with values. It involves the exercise of judgment. Evaluations of efficiency are all based on a logic that, in principle, can be programmed into a computer and automated.” (Ackoff, 1989). “Wise people exhibit both intelligence and creativity in one form or another, and an emerging wisdom derived in part from intelligence and creativity. This entails balance between change (creativity) and stability (intelligence) in societal activities and human behaviours when wisdom progresses i.e. making head ways in a given field in its quest for new knowledge and understanding. (Sternberg, 2001)” (Liew, 2013)

Wisdom is considered by almost all authors as the only characteristic that can still not be produced by machines do to the humanity associated with it. Liew, 2013 refers to wisdom as a balance between creativity and intelligence and in our present even if

machines can acquire intelligence, they cannot have pure creativity. So Ackoff, 1989 also considers that wisdom is the associated with the exercise of judgement, with the capability of understanding what is wrong or right. Once again, even if machines can decide judgments based on purely facts, there are cases where judgements go behind the facts. For the mentioned reasons, and because I share the opinion of those two authors, I will not allocate more time in the understanding and search for replicate wisdom through algorithms.

## 2.2 COMPETITIVE INTELLIGENCE

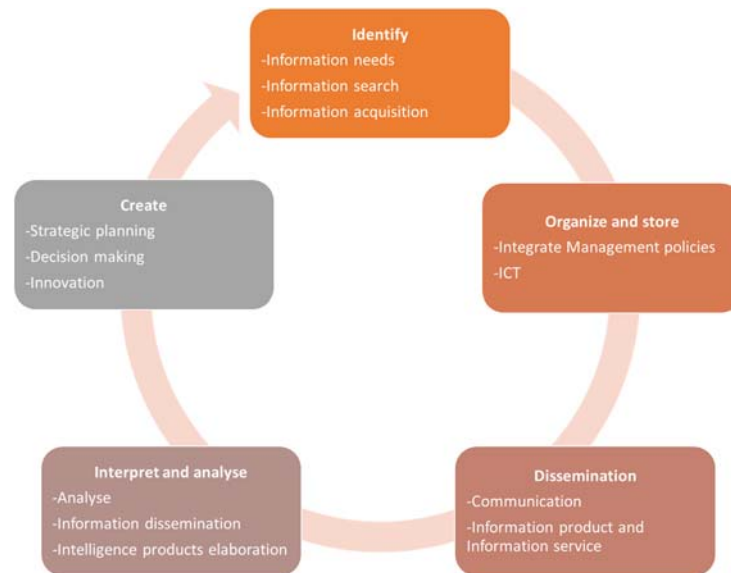
In this era of markets globalization organizations need to consider themselves as being part of a very volatile environment without any boundaries to protect them to the infinite external variables that can affect their business. The main purpose of CI is to provide actionable insights based on all available information to make all the external environment variables of a company the clearest and evident as possible to managers and consequently have a positive impact in their strategic management. (Heppes & Du Toit, 2009).

Ci is the “collection, transmission, analysis and dissemination of publicly available, ethically and legally obtained relevant information as a means of producing actionable knowledge [...] for the improvement of corporate decision making and action”. (Garcia-Alsina et al., 2016)

Heppes & Du Toit in their paper "Level of maturity of the competitive intelligence function" considers that “only by developing actionable implications does a CI program have the opportunity to create a competitive advantage and truly deliver value.”

Garcia-Alsina et al., 2016 in “Competitive intelligence theoretical framework and practices” expose the “*Intelligence cycle*” a framework proposed by Choo (2002,2006) as a starting point to analyse the five phases of the intelligence cycle.

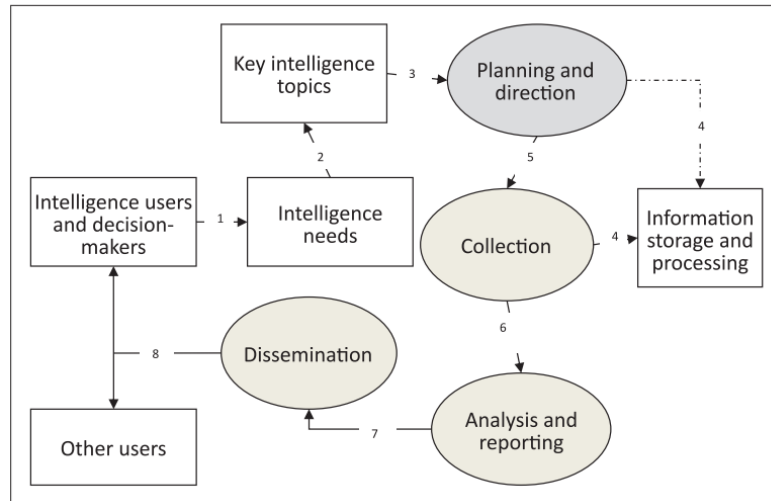




**Figure 3 Intelligence Cycle (from Garcia-Alsina et al., 2016)**

The first phase of the Intelligence Cycle, **Identify**, consists in identify and collect the organization’s information needs. In the second phase, **Organize and store**, the information identified in the first step is integrated and managed through the company informatic and communication technologies available. As per the third step, **Dissemination**, it is the phase that structures the information obtained to give sense to the collected weak signals. In the following phase **Interpret and analyse**, the analyses made from the previously disseminated information are also disseminated and give origin to intelligence products elaboration. Finally, in the **create** stage an organization defines its strategic plan and their decision making based on the intelligence products they got being a natural booster of innovation (Garcia-Alsina et al., 2016).

Pellissier and Nenzhelele in “Competitive intelligence theoretical framework and practices”, expose a more detailed framework for the competitive intelligence process as per the next figure.



**Figure 4 The competitive intelligence process (from Pellissier & Nenzhelele, 2013)**

According to Pellissier and Nenzhelele (2013), organizations are struggling with stiff competition due to globalization. Therefore, companies are looking for tools to counter competition or gain competitive advantage. One of the tools involves competitive intelligence. It is important to note that competitive intelligence has evolved from military theory, economics, marketing, information science, and strategic management (Pellisier and Nenzhelele (2013)). As a profession, competitive intelligence must follow a prescribed process. The process revolves around seven steps demonstrated in figure three. The stages include identification of intelligence needs and key topics, planning and direction, collection, information processing, analysis, dissemination, and identification of new intelligence needs. Specifically, the first step involves ascertainment of the decision needs and narrowing of the intelligence needs to relevant intelligence topics. Planning and directions include formulations of plans to guide the fulfilment of the intelligence needs of the decision-makers. After planning, information is collected from the external environment through various methods such as surveys, observations, or questionnaires. However, collection of information is done ethically and legally. Processing of information includes capturing and storing data. This data is then analysed to generate actionable intelligence. This actionable intelligence is distributed or disseminated to the decision-makers. Lastly, the last stage involves identification of new intelligence needs (Pellisier and Nenzhelele (2013)).

## 2.3 ARTIFICIAL NEURAL NETWORKS

The study of artificial neural networks (ANNs) is one of the two major branches of artificial intelligence. The other one is expert systems. (Kalogirou, 2000). Artificial Neural Networks are result of the developer’s intent to reproduce the biological neural system architecture. The motivation behind that intent is to try to reach the same performance as a human when effortlessly solve complex perceptual problems.(Jain et al., 1996). This discrepancy of performance resides in the differences between the biological neural systems. architecture and the von Neumann architecture detailed in table 1 (Jain et al., 1996)

|                              | Von Neumann computer   | Biological neural system  |
|------------------------------|--|---|
| <b>Processor</b>             | <ul style="list-style-type: none"> <li>• Complex</li> <li>• High speed</li> <li>• One or a few</li> </ul>                            | <ul style="list-style-type: none"> <li>• Simple</li> <li>• Low speed</li> <li>• A large number</li> </ul>                           |
| <b>Memory</b>                | <ul style="list-style-type: none"> <li>• Separate from a processor</li> <li>• Localized</li> <li>• No-content addressable</li> </ul> | <ul style="list-style-type: none"> <li>• Integrated into processor</li> <li>• Distributed</li> <li>• Content addressable</li> </ul> |
| <b>Computing</b>             | <ul style="list-style-type: none"> <li>• Centralized</li> <li>• Sequential</li> <li>• Stored programs</li> </ul>                     | <ul style="list-style-type: none"> <li>• Distributed</li> <li>• Parallel</li> <li>• Self-learning</li> </ul>                        |
| <b>Reliability</b>           | <ul style="list-style-type: none"> <li>• Very vulnerable</li> </ul>  | <ul style="list-style-type: none"> <li>• Robust</li> </ul>  |
| <b>Expertise</b>             | <ul style="list-style-type: none"> <li>• Numerical and symbolic manipulations</li> </ul>   | <ul style="list-style-type: none"> <li>• Perceptual problems</li> </ul>   |
| <b>Operating environment</b> | <ul style="list-style-type: none"> <li>• Well-defined, well-constrained</li> </ul>   | <ul style="list-style-type: none"> <li>• Poorly defined, unconstrained</li> </ul>   |

Table 1 Von Neumann computer versus biological neural system (Jain et al., 1996)

To have a clear perception of what programmers pretend to replicate is important have a superficial understanding of the biological neuron itself.

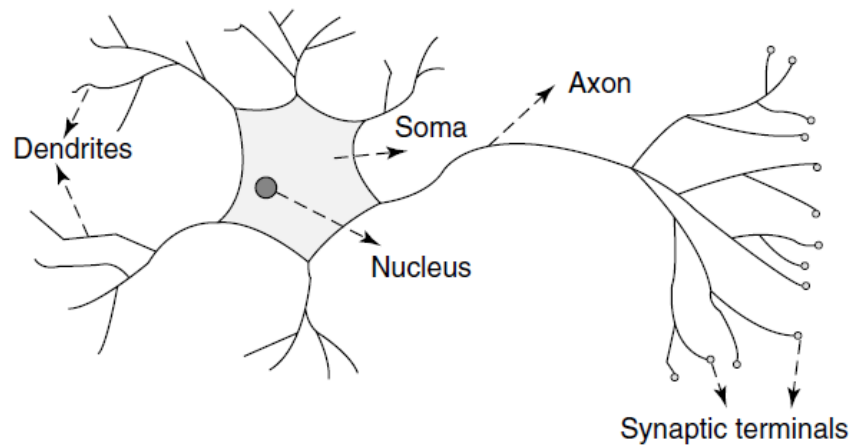


Figure 5 Mammalian neuron (Ajith Abraham, 2005)

“Each neuron is a cell (Figure 4) that uses biochemical reactions to receive, process, and transmit information. Treelike networks of nerve fibres called dendrites are connected to the cell body or soma, where the cell nucleus is located. Extending from the cell body is a single long fibre called the axon, which eventually branches into strands and sub-strands, and is connected to other neurons through synaptic terminals or synapses. The transmission of signals from one neuron to another at synapses is a complex chemical process in which specific transmitter substances are released from the sending end of the junction. The effect is to raise or lower the electrical potential inside the body of the receiving cell. If the potential reaches a threshold, a pulse is sent down the axon and the cell is ‘fired’.” (Ajith Abraham, 2005)

A typical artificial neuron can be represented as in the picture bellow:

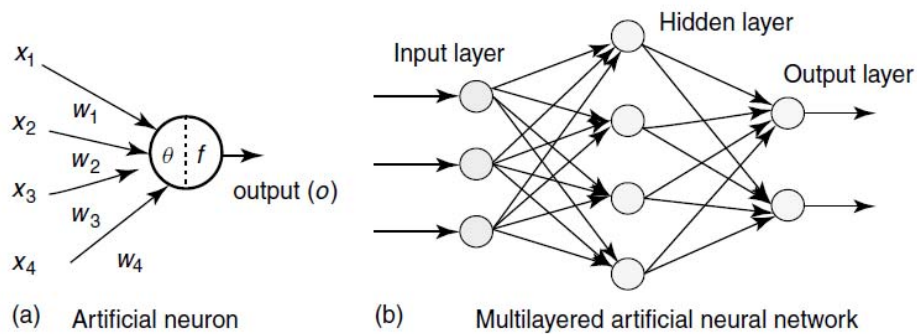


Figure 6 Architecture of an artificial neuron and a multi-layered neural network (Ajith Abraham, 2005)

Referring to the figure 5, the signal flows from the inputs  $x_1, \dots, x_n$  in a unidirectional way indicated by the arrows and results in the output signal (O). The signal O is given by the following mathematical equation:

$$O = f(net) = f \left( \sum_{j=1}^n w_j x_j \right)$$

where  $w_j$  is the weight vector, and the function  $f(net)$  is referred to as an activation (transfer) function.

There are for types of activation functions:

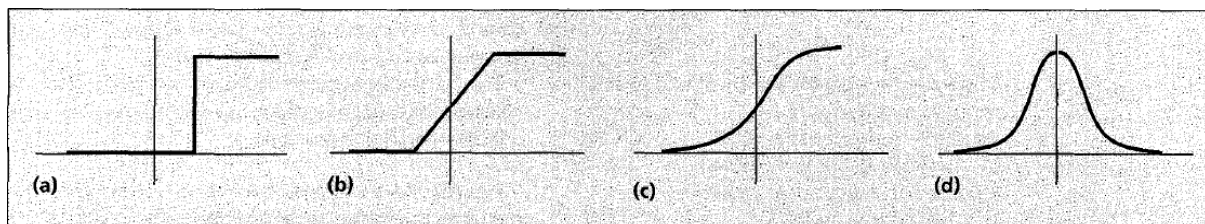


Figure 7 Different types of activation functions: (a) threshold, (b) piecewise linear, (c) sigmoid and (d) Gaussian (Jain et al., 1996)

It is important to refer that the sigmoid function is a popular activation function for the backpropagation neural network. (Kim & Ingoo Han, 2000) The backpropagation neural

network will be explained in detail in a further point. The variable net is defined as a scalar product of the weight and input vectors,

$$net = w^T x = w_1 x_1 + \dots + w_n x_n$$

where T is the transpose of a matrix, and, in the simplest case, the output value O is computed as

$$O = f(net) = \begin{cases} 1 & \text{if } w^T x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

where  $\theta$  is called the threshold level; and this type of node is called a linear threshold unit.

## 2.4 NEURAL NETWORK ARCHITECTURES

There are three types of neuron layers in a basic architecture: input, hidden and output layers. (Ajith Abraham, 2005) Please refer to the figure 7.

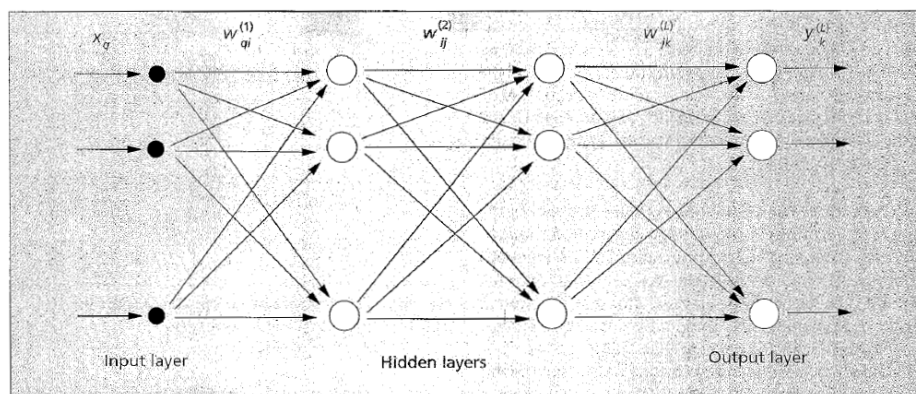


Figure 8 A typical feed-forward network architecture. (Jain et al., 1996)

Figure 7 represents a feed-forward network on which input signals flow from the input units to the output units strictly in a feed-forward direction. However, when constructing a neural network to execute predictive analyses the backpropagation architecture is the one applied. The main reason for it is the fact the mathematical relationships between the various variables are not specified. The neurons learn from the

examples fed to them. (a.T.C. Goh, 1995). Jain et al., 1996 in “Artificial Neural Networks: A Tutorial” explain step-by-step the backpropagation algorithm in the following figure:

**Back-propagation algorithm**

1. Initialize the weights to small random values.
2. Randomly choose an input pattern  $\mathbf{x}^{(u)}$ .
3. Propagate the signal forward through the network.
4. Compute  $\delta^L$  in the output layer ( $\alpha_i = y_i^L$ )

$$\delta_i^L = g'(h_i^L) [d_i^u - y_i^L],$$

where  $h_i^L$  represents the net input to the  $i$ th unit in the  $L$ th layer, and  $g'$  is the derivative of the activation function  $g$ .

5. Compute the deltas for the preceding layers by propagating the errors backwards;

$$\delta_i^l = g'(h_i^l) \sum_j w_{ji}^{l+1} \delta_j^{l+1},$$

for  $l = (L - 1), \dots, 1$ .

6. Update weights using

$$\Delta w_{ji}^l = \eta \delta_i^l y_j^{l+1}$$

7. Go to step 2 and repeat for the next pattern until the error in the output layer is below a prespecified threshold or a maximum number of iterations is reached.

Figure 9 Backpropagation steps (Jain et al., 1996)

## 2.5 PESTLE

The PESTLE is a framework that is used by businesses to support their strategic analysis activity.

The well-known PEST analysis it is self-explanatory as each letter represents one of the pillars of the framework that stands for Political (P), Economic (E), Social (S) and Technological (T) analysis on the external business environment.(HO, 2014).

The choice of the PESTLE framework instead of the PEST it is because of the consideration of two extra factors that are of major importance nowadays, the legal (L) and









The diagram presented in figure 10 is a systemic PEST analysis diagram. In the case of a PESTLE analyse the two extra factors should also be considered.

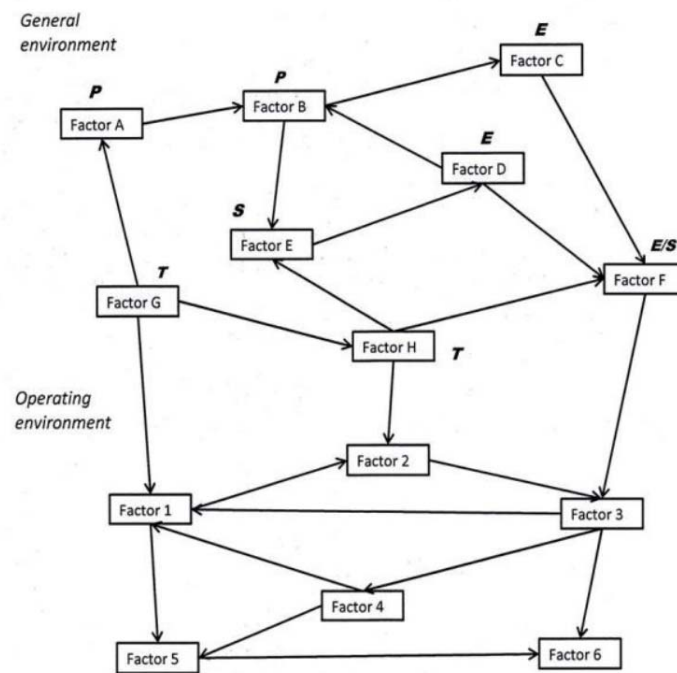


Figure 10 The format of a systemic PEST analysis diagram (HO, 2014)

The PESTLE execution consists in the analysis of possible impacts of the constituent factors and their implication to the organization itself. The purpose of the algorithm is to gather as much data insights as possible to analyse the impacts and consequent implications to the organization automatically and with a much more data-driven result.

## 2.6 COMPETITIVE ANALYSIS TOOLS

Aithal (2017) suggests that the completions within an industry makes a company to dedicate significant part of their time, resources, efforts, and smartness to formulate several strategies to out beat other businesses operating within their industry.

Understanding the competition within the industry is essential in determining the impact of such competition on individual company performance and the general performance of the industry (Aithal, 2017). Competitive analysis refers to an activity undertaken by a company to evaluate its competitors. It is a strategic technique seeks to identify the strengths and weakness of the competitors.

This information is used to improve the operation of the respective company to make its product and services unique or differentiate itself from the competitor. Most importantly, assessing the strengths and weaknesses of the competitors positions the company strategically to predict how they will react to certain actions and be able to gain a competitive advantage. Therefore, competitive analysis is an integral part of the company's strategic plan.

Notably, the competitors are placed in strategic groups according to their level of competition in terms of market share. In the strategic groups, competitors are listed with their products and services, marketing assumptions, and objectives, growth capacity and pattern, financial muscles (revenues), past and current strategies, cost structure, strengths and weakness, opportunities and threats, and size (number of employees and space). Thus, competitive analysis is important in keeping an eye on the competitors in aim to anticipate changes in the market, identify new trends and successful tactics, and maintain the cutting edge of what is working throughout the company's niche undertaking this evaluation. Several competitive analysis tools are used to analyse competitors. They include porters' five analysis forces, SWOT analysis, and PESTLE analysis (Aithal, Shailashree, and Kumar, 2017). Mainly, these tools are instrumental in understanding the external and internal environmental factors that may affect negatively or positively a company in a competitive market.

### **2.6.1 Porter's Five Forces Analysis**

According to Burns and Dewhurst (2017), Porter's five forces is a model that identifies and assesses the five competitive forces that structure industry and helps in the determination of the strengths and weaknesses of the companies operating in the same market. In other words, the model is a framework is used to evaluate an organization's

competitive environment. Michael E. Porter (a Harvard Business School professor) devised this model (Burns and Dewhurst, 2017). As a competitive analysis tool, competitive assessment can be done using the guidelines of Porter's five forces analysis including threat of new entrants to the market, threat of substituent services or products to the market, bargaining power of customers, the bargaining power of the suppliers in the market, and the competition among rivalries (Aithal, 2017).

Specifically, competition in the industry refers to the establishment of the number of competitors and their ability to weaken other companies. It is important to note that the bigger the number of organizations along with an equivalent number of products and services they provide, the smaller the power of an organization. Therefore, it means that the competition is high and vice versa. Secondly, the threat of new entrants to an industry affects the company's power in the market. The easier it is for a new company to enter the market, the more the company's position is weakened and vice versa.

Thirdly, the bargaining power of suppliers determines the cost of inputs or materials. Fourthly, the bargaining power of customers determines the prices of the outputs in the market. Lastly, substitute products or services are a threat to a company when available weakening the power of the company. Therefore, Porter's five forces analysis is essential in enabling an organization to adjust its strategy to utilize the resources according to the competition in the market effectively. In addition, to the five forces, Porter developed generic strategies that can be used for competitor's strategies analysis (Aithal, 2017b). These strategies include cost leadership, product or service differentiation, cost focus, and focus differentiation amongst others (Aithal, 2017). However, according to Mekic and MKekic (2017), Porter's model does not follow the pace of technology and globalization and its effect on the development of a competitive environment, but it is still a viable tool in analysing completion.

## 2.6.2 SWOT Analysis

Gurel and Tat (2017) substantiate that SWOT analysis is a tool used in strategic management in organizations. Most importantly, SWOT analysis is instrumental in building a competitive strategy. According to the system approach, organizations exist as one through their interactions with their environments and are made of several sub-systems. Thus, businesses operate in two environments. They include internal and external environments. It is crucial to analyse these environments for good strategic management processes. This practice of assessing the organization and its environment is known as SWOT analysis (Gurel and Tat, 2017). Mainly, SWOT analysis stands for strengths, weaknesses, opportunities, and threats. Through the conduction of external analysis, an organization identifies the important opportunities and threats in its competitive environment. Additionally, the external analysis, the company examines the likeliness of competition evolution and its implications to the threats and opportunities (Gurel and Tat, 2017). In other words, external analysis involves the identification of threats and opportunities. On the other hand, internal analysis is crucial in the identification of organizational strengths and weaknesses. Furthermore, external analysis as part of a SWOT evaluation enables an organization to understand some of the resources and capabilities that can be resourceful in beating completion (Gurel and Tat, 2017).

Therefore, strengths and weaknesses are internal attributes and factors affecting an organization. On the contrary, threats and opportunities are external attributes and factors that affect a company. Thus, SWOT analysis provides insightful information that is essential in matching the resources and capabilities of the organization to the competitive environment in which it operates (Gurel and Tat, 2017). However, SWOT analysis has some structural ineffectiveness such as a lack of prioritization of alternatives and extractable strategies (Menga, Dan, Lu, and Liu, (2015). In fact, Khan, Hossain, and Rabbi (2018) posit that studies have demonstrated the negative influence of SWOT analysis on strategic decision-making.

### 2.6.3 PESTLE Analysis

It is important to note first that PESTLE analysis is used in collaboration with the other analytical tools mentioned above (Porter's Five Analysis and SWOT Analysis) to give a clear understanding of the competition and associated external and internal factors (Birsan, Shuleski, and Cristea, 2016). Sammut-Bonnici and Galea (2015) insist that PESTLE analysis is more effective than other tools. This is because PESTLE analysis is used in conjunction with other tools such as SWOT and Porter's five forces analysis (Sammut-Bonnici and Galea, 2015). Therefore, the focus is put on PESTLE analysis as a competitive analysis tool. PESTLE analysis is a model or framework used to analyse the external environment of the organization. PESTLE is an acronym standing for political, economic, social, technology, legal, and environmental factors affecting an organization. The primary objective of the analysis is to provide a general picture of the main environmental drivers that should be considered when defining and implementing competitive strategies (Birsan, et al., 2016).

Specifically, political factors involve government interventions within the economy or an industry. In other words, political factors revolve around the government's influences or power to control organizations. They include political instability or instability, tax policy, corruption, trade restrictions, tax policy, and export and importation policy among others as shown in table 2. These factors require analysis when evaluating a competitive environment (B2U, 2018).

Economic factors refer to the determinant of the performance of the economy. These factors include exchange rates, interest rates, disposable incomes of consumers, economic growth, unemployment rates, and other factors shown in table 2. Notably, these factors can have a direct or indirect long-term effect on the company because it influences the bargaining and purchasing power of the consumers changing the demand and supply forces of the economy. Subsequently, the organizational prices of the products and services are affected (B2U, 2018).

Social factors surround the dimension of the overall environment representing the demographic characteristics, values, culture, and traditions of the people within which the company operates (Team FME, 2013). These include population growth trends, income

distribution, lifestyle, career concerns, age distribution, health and safety concerns, beliefs, culture and attitude barriers among others demonstrated in table 2. They are crucial in formulating and implementing marketing decisions to counter competition.

Technological factors revolve around the innovations surrounding technology that can influence the operations of the company and its market positively or negatively. These factors include technology incentives, research and development, technological changes, level of automation, level of innovation, and technological awareness across the market among others indicated in table 2. It is important to mention that these technological factors influence decisions such as entering certain industries, launching products, and outsourcing certain production functions. By considering these factors, an organization can predict technological changes and develop a technology that would not become obsolete soon and be able to remain competitive.

Environmental factors are increasingly becoming important due to the rising scarcity of raw materials, carbon footprint targets, and pollution targets established by any regime (B2U, 2018). These factors surround the ecological and environmental aspects such as climate change, global warming, environmental offsets, weather, and climate, which affect industries (Team FME, 2013). Most importantly, the increasing awareness of the possible impacts of climate change is influencing how companies operate and the products or services they provide. This has contributed to organizations getting involved in practices such as social responsibility and sustainability. Notably, the notion of corporate social responsibility has gained currency since the 1960s (Wang, Tong, Takeuchi, and George, 2016). Corporate social responsibility refers to a business taking responsibility to the society and the broader group of its stakeholders beyond its shareholders (Wang, Tong, Takeuchi, and George, 2016). Schrempf-Starling, Palazzo, and Phillips (2015) agree that organizations are increasingly held accountable for activities up and down their value chains but beyond their conventional corporate limits.

Lastly, legal factors are somehow related to political factors, which include regulations such as employment law, labour law, consumer protection law, patent and copyright laws, company laws, health and safety laws, and environmental laws among others (Sammut-Bonnici and Galea, 2015). Thus, an organization needs to be aware of such laws to conduct its operations ethically and legally. If a business is involved in international



trade, it becomes important to adhere to the international laws touching on importation and exportation of commodities. Most importantly, companies need to keep up with the potential changes in legislation and the effect it can have on their operations in the future. According to B2U (2018), the organization should at least be guided by a legal advisor or attorney when it comes to such matters.

## **2.7 IMPORTANCE OF PESTLE ANALYSIS**

PESTLE analysis is an important tool in the growth of any business because it involves the environmental scanning element of strategic management (Gupta, 2013). Thus, whether the growth is an expansion of a product line or launching a new branch of the company in a different location, this tool is instrumental in making plans and strategies. Specifically, PESTLE analysis assists businesses in understanding major changes in social, political, environmental, technological, economic, and legal factors. More so, it helps the organization in strategic management to expand (Gupta, 2013). Some of the advantages of PESTLE analysis include a deeper understanding of the business environment, alertness development, exploitation of opportunities and gaining of competitive advantage.

Mainly, PESTLE analysis assists an organization to gain an in-depth understanding of the external factors affecting its operations or activities. For example, when it comes to the creation of a new product, PESTLE analysis heightens the strategically thinking and deeper understanding of several factors such as safety and health laws, consumer protection laws, new technological patterns, and consumers' beliefs, attitudes, tastes and preferences among others. Furthermore, this competitive tool helps organizations to identify and contain potential threats that may be caused by external factors. For example, in the mobile manufacturing industry, a change in technology such as the development of new mobile features is a threat to a company if it does not upgrade to the innovation changes in its mobile brand. More so, PESTLE analysis is crucial in helping companies exploit opportunities that come into being because of external factors. For instance, the lower interest rates in the real estate market is an economic factor that creates an opportunity to intensify the

demand for housing since the housing loans are affordable due to the decreased interest rate.

Specifically, when it comes to the consideration of the political factors, according to Sammut-Bonnici and Galea (2015), it is evident that firms benefit from scanning and evaluating the political environment in which they operate. Regular monitoring of the political factors becomes paramount in guaranteeing that a firm's strategy is in line with the context of policy. Vital changes in the political setting can bring important changes in the structure and operations of a market or industry (Sammut-Bonnici and Galea, (2015).

Additionally, the economic factors mentioned above demonstrate the significance of the need for constant surveillance of the major economic factors that have an impact on an industry. Sammut-Bonnici and Galea (2015) substantiate that it is not an easy task to go through economic maze and to understand how the events are likely to unfold. Thus, businesses may wish to limit themselves to evaluate factors that have a direct and critical effect on their operation rather than the wide spectrum.

As far as social factors are concerned, changes in demographic data of the consumers can have an impact on many parts of an economy (Sammut-Bonnici and Galea, 2015). Many developed nations are struggling with the challenges of an aging population because the baby boomers are more than sixty years old. Notably, an old population necessitates additional demand for healthcare, medicine, social support services and decreases the demand for commodities associated with a young population such as technological products and education (Sammut-Bonnici and Galea, 2015). Mainly, the old generations have more time on leisure activities and are financially stable. Therefore, they provide the right mixture of customer spending. Hence, a scan of the social factors enhances the firm's ability to forecast the pressures that are likely to be made by various stakeholders on policy decisions of government that would, in turn, influence the company's operations (Sammut-Bonnici and Galea, 2015).

According to Sammut-Bonnici and Galea (2015) from a business point of view, technology can create a competitive advantage through a wide range of factors. These include cost-effectiveness, cheap production, enhanced access to customers, improved branding, improved product quality, and a high degree of business intelligence. Particularly, the spread of the internet and e-commerce has contributed to the elimination of

intermediaries, a process called disintermediation. For instance, in the tourism and travel industry, tourists have become their own travel agents. This is because they book and plan their travel on an online platform known as the business-to-consumer that had been previously limited to business-to-business participants of the supply chain. Furthermore, most of the retailers are experiencing intense competition as they are facing rivalry from online stores. However, technology has led to the birth of a new set of intermediaries in a process called re-intermediation. Examples of the new intermediaries include eBay, Amazon, and Expedia among others. These new forms of intermediaries are advantageous when compared to others because they are integrated with social networks that provide new channels of marketing a product or service. It is important to note that these channels enable consumers to provide their feedback on their interaction with the product or service. The feedback is essential in the provision of value in supply chain operations. These opinions of the customers are important in attracting new customers because they create a positive image of the organization (Sammut-Bonnici and Galea, 2015).

Therefore, PESTLE analysis is paramount in providing strategic planners with a guideline that increases their understanding of the external environment. However, it is not intended to restrain the analysis in a straitjacket (Sammut-Bonnici and Galea, 2015). The importance of the analysis is realized when the interdependencies among the various elements of the tool are established. For instance, an upsurge of government taxes is likely to affect or reduce the disposable income of the consumers. On the contrary, the provision of favourable incentives for an industry is likely to attract more investment that creates employment and wealth. Additionally, a growing understanding of the environment by the citizens is likely to pressure the government to formulate and implement laws that protect the environment. Sammut-Bonnici and Galea (2015) recommend that this framework should be used judiciously because the evaluation of the economic, political, environmental, legal, technological, and social factors can become an endless activity with no value to the business. The main principle of the tool is that only factors with a direct impact on the business operations or industry functionality should be evaluated to enable the gain of competitive advantage (Sammut-Bonnic and Galea, 2015).

## 2.8 IMPORTANCE OF AUTOMATING PESTLE ANALYSIS

According to Al-Nady and Alharaizah (2014) technology has contributed to the development of databases that help in the organization and analysis of the information reported to a company from all sources such as mobile phones and websites through the SPSS, a technological tool used for data analysis. Mainly, the proliferation of new technology has impacted the general business and production process. Technology has been effective in monitoring environmental changes especially the ones affecting the business efficiencies, existing infrastructure, or development of new structures or products (Al-Nady and Alharaizah, (2014).

Al-Nady and Alharaizah (2014) established that environmental scanning and communication technology paves the way for sustainable competitive advantage. In other words, the success of any organization depends on a combination of environmental scanning and communication technology as an approach to sustainable competitive advantage. In this context, competitive advantage is gained through the analysis of the external environment, which provides essential and specific information to managers to place industrial development patterns and come up with the development strategies (Al-Nady and Alharaizah, 2014). Attaining competitive advantage is the most challenging issue facing organizations across all sectors. This issue has led to the establishment of resource and knowledge-based theories that investigate the association between the core capabilities and resources. Citing Porter, Al-Nady and Alhairazah (2014) suggest that the five types of competitive advantage include low-cost production, differentiation, and focus. Particularly, a company enjoys a competitive advantage if it can deliver quality products at a cheaper rate compared to competitors. These quality products should translate into high profitability. Therefore, competitive advantage can be defined as the ability or advantage a firm has over its competitors because the company wants the gap between understood value and cost of the commodity to be greater than the competition.

Specifically, Al-Nady and Alhairazah (2014) found out that suppliers face several challenges especially uncertainties in the business environment. If the supplier does not conduct an environmental analysis or PESTLE analysis and uses technology, the application

of competitive advantage is almost of no important effect. Therefore, failing to take time to think and conduct a PESTLE analysis and use the right technology can result in increased risks and errors, expensive mistakes, and wrong understanding of the market. The role of the scanning environment through communication technology is essential in identifying and maximizing opportunities, minimization of threats especially during planning and developing strategies of a firm. It assists in forecasting future factors, which interact in the business environment. Automation of the process saves time by fast-tracking the workflow process. More automated filing systems save space, time and costs used in printing, and paper costs. Most importantly, the use of computer systems enables instant implementation of corrections.

According to Gallupe and Jiang (2015), the three pitfalls of an ineffective PESTLE analysis include insufficient analysis, unreliable technology, and lack of focus on decision-makers' real demands. The business environment is getting significantly complex, a large volume of information flows drowning firms, and many of them feel the pressures of lack of capacity in face of too many sources of information. To tackle the process, automation of PESTLE analysis has been very effective. It involves the use of powerful information technology tools to scan the environment and make decisions. Therefore, business analytics is almost at the top of the list of information technology investments for many organizations (Gallupe and Jiang, 2015). Business analytics refers to a set of approaches, tools, and organizational procedures used in combination with one another to collect data, analyse information, and predict outcomes out of the problem solutions. Notably, business analytics tools are essential in empowering firms through the generation of business insights that reflect a deep understanding of a given situation that can benefit the business (Gallupe and Jiang, 2015).

However, several organizations undertake that an insight capability comes by default through investing in and implementing business analytics instruments. This is not the case according to Gallupe and Jiang (2015). It is easy for managers to access information from a dashboard but is hard to get the right information or to analyse the information to acquire accurate insights. Encouraged by the gap between business analytics tools and efficient business insight delivery, Gallupe and Jiang proposed the need for knowledge integration, which is significant in the generation of business insights. Therefore, in addition to gaining

sustainable competitive advantage, integration of business analytics and knowledge integration on an environmental analysis process contributes to the generation of crucial business insights

## **2.9 IMPROVEMENT OF PESTLE ANALYSIS USING NEURAL NETWORKS**

As mentioned earlier, an artificial neural network is a branch of artificial intelligence. Neural networks can improve PESTLE analysis by identifying the quality of the product. In other words, the neural network can help in the determination of the perceived quality product by customers. This is because according to Kutilek, the evaluation of quality products can be possible through an automatic cognitive process. The cognitive system must apply the methods of artificial intelligence, as are neural networks (Kutilek). The analysis of the quality of a product relies on the described interpretation of vague descriptions. Kutilek defines the numerical description at the interval  $\langle -3, 3 \rangle$  with the intervals of one. The method of artificial intelligence will define the crisp value of the quality of the product with the defined range. The information source to the neural network system will be the data collected by human experts, the internet, and the business system. Nevertheless, Kutilek uses normalization of all neural network input data to the same grid interval to maintain information objectiveness.

From the noted pre-requisite, it is obvious that the proposed neural network will have one output neuron in the output layer. The number of input variable quantities defined by a human expert will determine the aggregate number of the input neurons in the input layer. The total number of layers in the neural network must be optimized according to the given number of output and input neurons. By considering quality as characteristics representing the level of customer's satisfaction, the input data is used in blocks of identification for recognition of information such as the ratio of sales returns, ration of sold goods, information about goods from customers and internet visitors, evaluation and reviews of the products in the professional and customized media. This information about the product will be input data subjected to the neural network. The neural network will analyse the input data from the external sources. Mainly, according to the structure of the

neural network and the data, it will determine the output variable demonstrating the value of the product quality (Kutilek). Lastly, neural networks improve PESTLE analysis by assisting in the identification of market share in the same process used to identify the quality of the product. Identification of market share and product quality is important especially in the analysis of economic and social factors affecting a business.

## **2.10 DECISION SUPPORT TOOLS**

Decision support tools or system is a component of the management support system. The decision support systems are designed to facilitate flexibility in the decision-making process for improving effectiveness resulting in efficiency in the organization. They provide the management with data and guidelines to carry out depth analysis such as PESTLE analysis to improve analytical decision-making procedures revolving around structured and semi-structured decisions. Therefore, decision support tools are designed to assist the users to make effective decisions by guiding them through clear decision-making stages and providing the chances of several outcomes generated from different options (Rose et al., 2016). Anwar and Ashraf (2014) substantiate that decision support systems were developed to enhance the quality of decisions.

The design of the decision support system is composed of three layers or elements involving the database management system, model base management system, and conversation management. Notably, the database management system serves as the data storage for the decision support system. In other words, it stores or saves a great amount of data linked with the applications and issues for which the database in the decision support system has been created and it improves the systematic access of data. On the other hand, model base management is integral in the provision of freedom from models that are used in the decision support systems from the application the user is utilizing. The main function of this component is to transform data stored in the database management system into useful information, which can be instrumental in the process of making decisions. Lastly, the conversation management helps in the creation of a model and provision of a rich environment to enable interaction with the overall system like acquiring

knowledge of hidden patterns and recommendations. The objective of this component is to enhance the conversation abilities of the user with the system and enable operators to enjoy the maximum benefits of the system. Therefore, these components come together to form the decision support system.

According to Rose et al. (2016), several decision support tools are available for different users. However, the recommendations of the decision-making tools will depend on the users' inputs and they can suggest an optimal decision path. Gibson et al. (2017) agree that there are many examples of decision support tools employed to analyse information to assist in decision-making. It is important to mention that decision support tools are numerous including communication-based tools, a model-based tool, a data-oriented tool, a knowledge-based tool, and a document-based tool. Specifically, the communication-based tool is based on communication among team members. Communication is facilitated through video conferencing, messages, and emails. Secondly, model-based helps the user in coming up with solutions to complex problems. Thirdly, the data-oriented tool employs database and warehouse to request data concerning different situations and problems. Managers and members of the staff mostly use this tool. Fourthly, a document-based tool is used to search the websites or specific documents. Lastly, knowledge-based tools are applicable where the information may not be visible to the user beforehand. This tool identifies the hidden trend among data records, management suggestions, and services. Several techniques are used to collect knowledge for the decision-making process in a personal or group decision support system (Anwar and Ashraf, 2014).

Decision support tools are useful in different sectors such as education, medicine, agriculture and business among others (Anwar and Ashraf, 2014). Specifically, in the past years, a decision support system in the business field has acquired an increasing amount of value among the users and the researchers due to several success cases that have reported notable progress in the performance of the organization. Decision support systems are integral in helping business personnel in the formulation and implementation of policies and in utilizing the technical means to aid the professionals in decision-making. Mainly, decision support tools are an informational framework associated with the mathematical models, informational databases and user-friendly environment to communicate recommended decisions to the managers and the users (Anwar and Ashraf, 2014). It can be deduced that



decision making in the decision support tool is diverse of rational information or management information system because it provides information and knowledge to the users' databases and reports and delivers answers according to the user's needs or questions.

## **2.11 PYTHON PROGRAMMING AND ITS APPLICATION TO DECISION SUPPORT TOOLS**

According to Miller, Ranum, and Anderson (2019), python is writing language used to develop programs. In other words, Python is a high-level programming language like Java or C++. It is crucial to mention that python was originally developed in 1991 by Guido van Rossum (Chitra, Aswini, and Geetha, 2017). The primary objective behind its development was to emphasize code readability and its syntax enables programmers to describe concepts in few lines of code. In other words, the simple syntax rules of python make it easy for the programmer to keep the codes readable and application maintainable (Mindfire Solutions, 2017). Mindfire Solutions (2017) posit that according to the latest TIOBE Programming Community Index, python is one of the most used programming languages in 2017. Python, a multi-purpose and high-level programming language, is used in the development of several applications and allows the programmer to concentrate on the core functions of the application by handling common programming tasks.

Other reasons why Python is the most preferred programming language include several programming paradigms. Particularly, like any other programming language, python supports multiple programming paradigms. It supports object-oriented and structured programming (Mindfire Solutions, 2017). Additionally, python supports several concepts in functional oriented programming. Furthermore, python is compatible with important systems and platforms. In other words, python is compatible with many operating systems. It also allows the user to run similar codes on many platforms without recompilation. Therefore, the user does not have to recompile the code after making changes. This feature

makes it easy for the programmer to initiate adjustments to the code without increasing the time of development.

Additionally, python contains a large and robust library that enables the user to select from a variety of modules according to his or her needs. Notably, every module allows the user to increase functionality to the python application without incorporating additional code. For example, when developing a web application using python, the developer can employ given modules to implement web services, execute string operations, control operating system interfaces or work with internet protocols (Mindfire Solutions, 2017). Furthermore, the programmer can collect information concerning various modules by surfing through the python standard library documentation (Mindfire Solutions, 2017).

In addition, python's popularity is heightened by the fact that it has various open-source tools and frameworks (Mindfire Solution, 2017). Mainly, as an open-source programming language, python assists the programmer to reduce costs involved in software development. The user can use the various open-source framework sources, development tools, and libraries to reduce coding time without compromising his or her budget. More so, the programmer enjoys the alternatives presented from the open-source framework and development tools to suit his or her needs (Chitra et al., 2017). For example, one can simplify and speed up the designing of web applications through frameworks such as Flask, CherryPy, Django, Bottle, and Pyramid (Mindfire Solutions, 2017).

Furthermore, the python has gained popularity due to its ability to simplify complicated software development. As mentioned earlier, Python is a multi-purpose programming language. Thus, a developer can deploy python to come up with web applications and desktop applications. Additionally, Python can be used to design complicated numeric and scientific applications. Notably, python comprises of features that catalyse data visualization and analysis (Mindfire Solutions, 2017). Therefore, a programmer or writer can utilize the data analysis features to establish big data solutions without putting much time and effort. On the other hand, data visualization features provided by python help developers in the visualization and presentation of information effectively and appealingly. Most importantly, python users use it to complete artificial intelligence and natural language processing tasks (Mindfire Solutions, 2017 and Chitra et al., 2017).

Lastly, python is used to establish a prototype of the software application in reduced time. The software application is built directly from the prototype by refactoring the code. More so, python eases the task of performing code and testing simultaneously through the adoption of the test-driven development style. Therefore, the writer of the program can easily perform the required tests before developing the code and use the tests to evaluate the application code constantly. Most importantly, the tests can be used to identify the conformance of the application to predetermined requirements according to the source code.

Through the features mentioned above, python programming is used in the development of decision support systems. For example, Sim, Ban, Tim, Sethi, and Loh (2017) developed a computerized patient support system to aid in the management of diabetes. The system generates laboratory results associated with diabetes care including the lipid and renal panel along with other tests (Sim et al., 2017). Specifically, the python programming language was applied to develop the various decision support system tools including the alert module, color-coding module, and the interactive graph module (Sim et al., 2017).

### **3 METHODOLOGY**

#### **3.1 DATA SOURCES**

Considering that this development will only address the Political module, and its purpose is to predict how well a country would perform political-wise based in performance indexes of previous years, the ideal data sources need to have the most countries and historical indexes as possible. For this development only public data sources were used. The following Data Sources described were the ones that amongst them had the biggest range of historical data, from 2013 to 2017, for 108 countries.

| PESTLE           | FACTORS                 | Data Sources                    | Description   |
|------------------|-------------------------|---------------------------------|---|
| <b>POLITICAL</b> | Political Conditions    | Worldwide Governance Indicators | The Worldwide Governance Indicators (WGI) project reports aggregate and individual governance indicators for over 200 countries and territories for six dimensions of governance:<br>- Voice and Accountability<br>- Political Stability and Absence of Violence<br>- Government Effectiveness<br>- Regulatory Quality<br>- Rule of Law<br>- Control of Corruption  |
| <b>POLITICAL</b> | Political Conditions    | Index of Economic Freedom       | The Wall Street Journal and The Heritage Foundation measure ten variables and rank over 180 countries. The Index evaluates government policies and economic conditions in 186 countries. Since its inception in 1995, the Index, an annual cross-country analysis by The Heritage Foundation, has tracked the progress of economic freedom around the globe and measured the real impact of advancing economic liberty. |
| <b>POLITICAL</b> | Political Conditions    | Transparency International      | TI publishes an annual Global Corruption Report, a Bribe Payers Index, and a Corruption Perceptions Index, ranking countries by the degree to which corruption exists among public officials.   |
| <b>POLITICAL</b> | Business Law and Policy | Doing Business Law Library      | An online collection of business laws and regulations. Links to official government sources of laws regarding labour, taxes, trade, property, banking, companies, civil procedure, and bankruptcy. Part of the Doing Business project of the World Bank.  |

Table 3 Political Indexes data sources

### 3.2 DATA MANIPULATION

Data Manipulation and data treatment, it is a crucial step when developing any machine learning algorithm, especially in order to avoid the “trash-in trash-out” rule. If we use wrong or poor-quality data to train any machine learning model it will have a direct impact into the quality of the predictions.

First, as represented in the figure below, for each source it was calculated the average of all the factors considered by that data source per year. As example, for the World Governance Indicators (WGI) the average of indexes such as Voice and

Accountability; Political Stability; Absence of Violence; Government Effectiveness; Regulatory Quality and others was to each year figured in the WGI database.

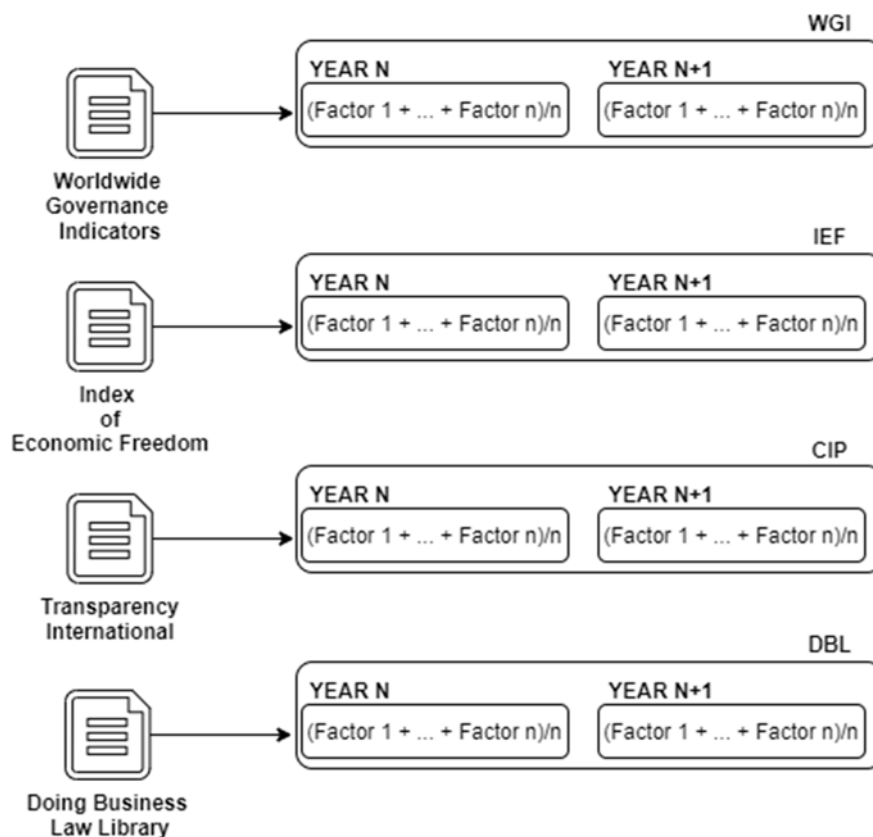


Figure 11 Data manipulation schema

Secondly, as the following image, the average of the indexes obtained from the four different databases of a respective year were used as input and the average of the indexes obtained of the next year used as target which is the result to be predicted.

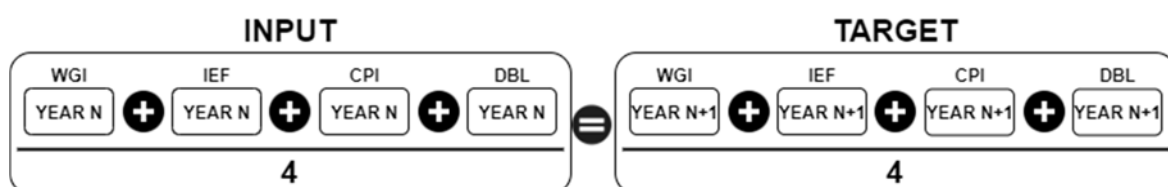


Figure 12 Input and target schema

### 3.3 NEURAL NETWORK CONSTRUCTION AND TRAINING

To build the neural network the deep neural network python library KERAS was used. Keras provides different modules to build and explore the creation of Neural Networks. Modules such as neural layers, cost functions, optimizers, initialization schemes, activation functions and more can be combined in an easy way to create new models. Therefore, a neural network model can be built as per the following steps (Heller Martin, 2019):

```
import keras
from keras.models import Sequential
from keras.layers import Dense

#Create Sequential model with Dense layers, using the add method
model = Sequential()

#Dense implements the operation:
#    output = activation(dot(input, kernel) + bias)
#Units are the dimensionality of the output space for the layer,
#    which equals the number of hidden units
#Activation and loss functions may be specified by strings or classes
model.add(Dense(units=64, activation='relu', input_dim=100))
model.add(Dense(units=10, activation='softmax'))

#The compile method configures the model's learning process
model.compile(loss='categorical_crossentropy',
              optimizer='sgd',
              metrics=['accuracy'])

#The fit method does the training in batches
# x_train and y_train are Numpy arrays --just like in the Scikit-Learn API.
model.fit(x_train, y_train, epochs=5, batch_size=32)

#The evaluate method calculates the losses and metrics
#    for the trained model
loss_and_metrics = model.evaluate(x_test, y_test, batch_size=128)

#The predict method applies the trained model to inputs
#    to generate outputs
classes = model.predict(x_test, batch_size=128)
```

Figure 13 Neural network construction with KERAS

### 3.3.1 Defining predictors and targets

In order to train the model, the data needs to be arranged into two different categories, predictors X and targets Y.

In the Data Manipulation section, all the manipulations were made to achieve the needed structure.

```
In [50]: political.head()
Out[50]:
```

|   | wgi_x     | index_x | cpi_x | dbll_x | Y         |
|---|-----------|---------|-------|--------|-----------|
| 0 | 45.038530 | 65.2    | 31    | 90.19  | 60.043712 |
| 1 | 25.663619 | 49.6    | 36    | 73.74  | 45.703716 |
| 2 | 15.442524 | 47.3    | 23    | 53.82  | 34.471750 |
| 3 | 39.831935 | 46.7    | 34    | 72.58  | 46.611185 |
| 4 | 46.092087 | 69.4    | 36    | 93.39  | 60.450627 |

Figure 14 Screenshot of the final data structure

As explained the indexes of a year (YEAR N) were used as predictors and the indexes of the year after (YEAR N+1) were used as target :

| PREDICTORS (_x) | TARGET (Y) |
|-----------------|------------|
| 2013            | 2014       |
| 2014            | 2015       |
| 2015            | 2016       |
| 2016            | 2017       |

Table 4 Years of Predictors and Targets

### 3.3.2 Define the Neural Network model

The example of Figure 13 is from a classification problem and this development is a descriptive one. Hence, when compiling the loss function that will be used will be the mean-squared-error and the optimizer of the weights will be choose by evaluating the performance of SGD, stochastic gradient descent, against Adam, adaptive moment estimation, which is an extension to stochastic gradient descent.

Stochastic gradient descent maintains a single learning rate (termed alpha) for all weight updates and the learning rate does not change during training. As per Adam authors, Adam combines the advantages of two other extensions of stochastic gradient descent:

- Adaptive Gradient Algorithm (AdaGrad) that maintains a per-parameter learning rate that improves performance on problems with sparse gradients (e.g. natural language and computer vision problems).
- Root Mean Square Propagation (RMSProp) that also maintains per-parameter learning rates that are adapted based on the average of recent magnitudes of the gradients for the weight (e.g. how quickly it is changing). This means the algorithm does well on online and non-stationary problems (e.g. noisy).

To define the right learning rate for the optimizers both were executed with following learning rates: 0.000001; 0.00001; 0.000015 and for both the lowest loss was obtained with a learning rate of 0.000015.



### 3.3.2.1 Comparing performances

To validate which was the best optimizer two models were trained, model\_1 using the ADAM optimizer and model\_2 using the SGD optimizer.

Due to the range of the Validation score and the discrepancy between model\_1(red) and model\_2(blue), val\_loss: 3429.4519 against val\_loss: 13.7325 respectively, the graph gives the wrong perception that the model two starts with losses really close to zero when indeed it starts with around 13 and finishes with a val\_loss of 1.9023

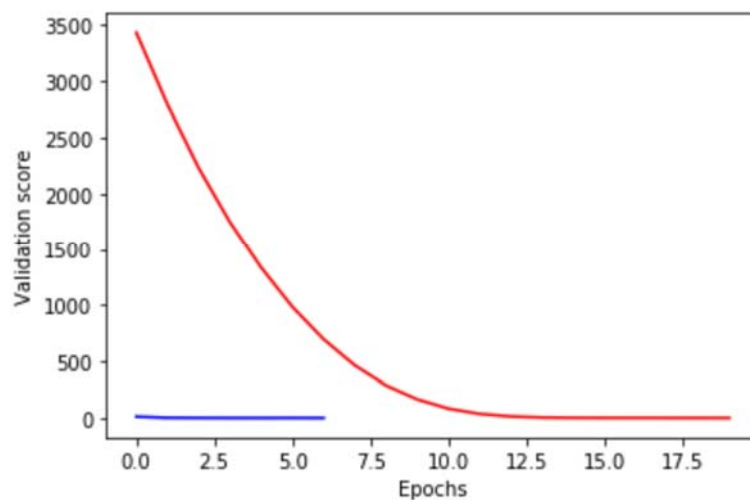


Figure 15 Comparison graphic of the validation scores of model\_1 and model\_2

### 3.3.3 USAGE OF THE MODEL\_2 TO PREDICT THE 2018 INDEXES

Considering that the model\_2 had the lowest validation score, it is the one being used to predict the 2018 indexes by using the indexes of 2017 as predictors.

```
In [95]: path='/users/Elson/Desktop/TESE/PEST/01P/p_2017.xlsx'
p_2017= pd.read_excel(path)

In [105]: p_2017=p_2017.loc[:,['2017_wgi','2017_index','2017','2017_db11']]

In [106]: predictor_2018=np.array(p_2017)

In [109]: prediction=model_2.predict(predictor_2018)

In [111]: prediction=pd.DataFrame(prediction)

In [112]: prediction.to_excel("/users/Elson/Desktop/TESE/PEST/01P/prediction.xlsx")
```

Figure 16 Screenshot of the prediction's execution

## 4 RESULTS AND DISCUSSION

Considering the loss value obtained in the model used to predict, which was loss of 1.9023 the predictions obtained are far from being trustworthy. The top 10 countries as per the predictions obtained with the model\_2 are the following:

| Country     | Predictions |
|-------------|-------------|
| New Zealand | 91.63370514 |
| Singapore   | 88.39645386 |
| Canada      | 87.66313171 |
| Switzerland | 87.02622986 |
| Norway      | 86.59942627 |
| Sweden      | 86.07897949 |
| Australia   | 85.94795227 |
| Finland     | 85.85818481 |
| Netherlands | 85.72468567 |

Table 5 Prediction Values

Hence there is no data available from the year 2018 the possibility of doing a comparison of the predictions with real data is impossible. However, when comparing with the data from 2017 the trend of the top 10 countries is almost the same.

| 2017 - REAL |             | 2018 - PREDICTED |             |
|-------------|-------------|------------------|-------------|
| Country     | Real Values | Country          | Predictions |
| New Zealand | 92.75463086 | New Zealand      | 91.63370514 |
| Singapore   | 89.50503246 | Singapore        | 88.39645386 |
| Canada      | 88.46305147 | Canada           | 87.66313171 |
| Switzerland | 87.99027132 | Switzerland      | 87.02622986 |
| Norway      | 87.66101293 | Norway           | 86.59942627 |
| Sweden      | 87.1332298  | Sweden           | 86.07897949 |
| Finland     | 86.96527822 | Australia        | 85.94795227 |
| Netherlands | 86.60231473 | Finland          | 85.85818481 |
| Australia   | 86.48155923 | Netherlands      | 85.72468567 |

Table 6 Comparison between 2017 real values against 2018 predictions

To make the data visualization easier and fast to perceive, excel of predictions was imported to PowerBI. Through the map, visualization filtered for the top 10 countries with best indexes having as result the image below:

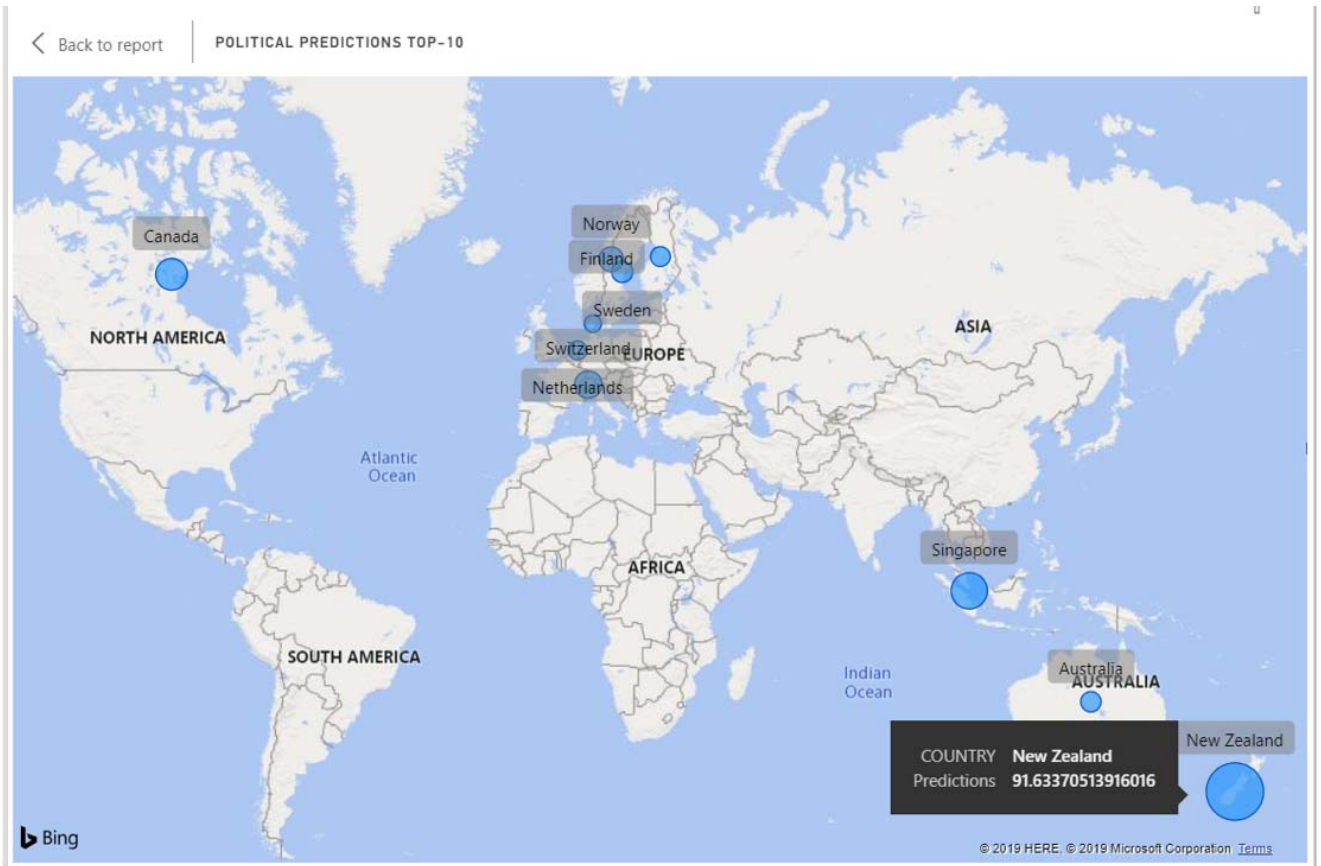


Figure 17 Power BI map with the top 10 of the predictions

## 5 CONCLUSIONS

The efficiency of a neural network is directly proportional to the dataset used to build it. Considering the range of years used, 2013 to 2016, to train this model and the high values of validation loss obtained the predictions obtained from this model are far from being correct. However, considering this exercise, any company that would have funds to buy the proper datasets and to invest in a construction of a model with more sophisticated social mathematics for machine learning it would benefit undoubtedly of having such asset in their information systems.

A live stream of input data defined by the analysts could give a worldwide accurate picture of the political situation of each country in a live dashboard allowing decision makers to take any needed actions in case of any sudden change in any of the countries they would have or wanted to conduct business.

## **6 LIMITATIONS AND RECOMMENDATIONS FOR FUTURE DEVELOPMENTS**

As mentioned previously, the major limitations of this development were the lack of open databases with the data needed to build the model and the quality of data of the databases gathered. In order to have a proper return of investment in this type of developments, the major investment should be in the acquisition of trustworthy databases and the construction of a web-scraping algorithm that could also be used to generate live indexes of political factors worldwide.

## 7 Bibliography

- Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, (16), 3–9.  
<https://doi.org/citeulike-article-id:6930744>
- Ajith Abraham. (2005). Artificial Neural Networks. In *Handbook of Measuring System Design*.
- Albus, J. S. (1991). Outline for a theory of intelligence. *IEEE Transactions on Systems, Man and Cybernetics*, 21(3), 473–509. <https://doi.org/10.1109/21.97471>
- a.T.C. Goh. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9, 143–151. [https://doi.org/10.1016/0954-1810\(94\)00011-S](https://doi.org/10.1016/0954-1810(94)00011-S)
- Bates, M. J. (1999). The invisible substrate of information science. *Journal of the American Society for Information Science*, 50(12), 1043–1050.  
[https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:12<1043::AID-ASI1>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1097-4571(1999)50:12<1043::AID-ASI1>3.0.CO;2-X)
- Brodie, I., & Brodie, L. (2009). A Knowledge-Information-Data Concept Model for Engineering Education. *Australasian Journal of Engineering Education*, 15(3), 137–144. <https://doi.org/10.1080/22054952.2009.11464030>
- Garcia-Alsina, M., Cobarsí-Morales, J., & Ortoll, E. (2016). Competitive intelligence theoretical framework and practices. *Aslib Journal of Information Management*, 68(1), 57–75. <https://doi.org/10.1108/MRR-09-2015-0216>
- Heller, Martin (2019, January 28). What is Keras? The deep neural network API explained. <https://www.infoworld.com/article/3336192/what-is-keras-the-deep-neural-network-api-explained.html>
- Heppes, D., & Du Toit, A. (2009). Level of maturity of the competitive intelligence function: Case study of a retail bank in South Africa. *Aslib Proceedings: New Information Perspectives*, 61(1), 48–66. <https://doi.org/10.1108/00012530910932285>
- HO, J. K.-K. (2014). Formulation of a Systemic PEST Analysis for Strategic Analysis. *European Academic Research*, 2(5), 6478–6492. <https://doi.org/10.1038/embor.2012.155>
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31–44. <https://doi.org/10.1109/2.485891>
- Kalogirou, S. A. (2000). Applications of artificial neural-networks for energy systems. *Applied Energy*, 67(1–2), 17–35. [https://doi.org/10.1016/S0306-2619\(00\)00005-2](https://doi.org/10.1016/S0306-2619(00)00005-2)

- Kim, K., & Ingoo Han. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, 19, 125–132. [https://doi.org/10.1016/S0957-4174\(00\)00027-0](https://doi.org/10.1016/S0957-4174(00)00027-0)
- Liew, A. (2007). Understanding Data, Information, Knowledge and Their Inter-Relationships. *Journal of Knowledge Management Practice*, 7(2), 1–16.
- Liew, A. (2013). DIKIW : Data , Information , Knowledge , Intelligence , Wisdom and their Interrelationships. *Business Management Dynamics*, 2(April 2013), 49–62.
- Nonaka Hirotaka, I. & T. (1995). *The Knowledge-creating company*. Oxford Uni(New York). <https://doi.org/10.3414/ME13-02-0041>
- Sharma, N. (2008). *The Origin of Data Information Knowledge Wisdom (DIKW) Hierarchy*. (January), 4.
- Google (2019). Definitions obtained from a Google search for “information etymology” as of January 2019.
- Aithal, P. S. (2017a). An effective method of developing business case studies based on company analysis. *International Journal of Engineering Research and Modern Education*, 2, 16-27.
- Aithal, P. S. (2017b). Industry analysis-the first step in business management scholarly research. *International Journal of Case Studies in Business, IT and Education*, 2, 1-12
- Aithal, P. S., Shailashree, V., and Kumar, P. M. (2016). Application of ABCD analysis framework on Private University System in India. *International Journal of Management Sciences and Business Research*, 5, 159-170.
- Al-Nady, B., and Alharaizah, A. (2014). Environmental scanning and communication technology as a gate for the sustainable competitive advantage of sanitary ware (bathtub) suppliers in the kingdom of Saudi Arabia. 13. 10.24297/ijct. v13i3.2764
- Anwar, N., and Ashraf, I. (2014). Significance of decision support systems. *International Journal of Current Engineering and Technology*, 4, 2740-2742.
- B2U. (2018). Scanning the Environment: PESTLE Analysis. Retrieved from <https://www.business-to-you.com/scanning-the-environment-PESTLE-analysis/>

- Birsan, A., Shuleski, D., Cristea, V. C. (2016). Practical Approach of the PEST Analysis from the Perspective of Territorial Intelligence. [PDF File]. Retrieved from <http://stec.univ-ovidius.ro/html/anale/ENG/2016/2016-II-full/s3/4.pdf>
- Burns, P., and Dewhurst, J. (2016). *Small Business and Entrepreneurship*. Macmillan International Higher Education.
- Chitra, P., Iswarya, K. H. A., Geetha, K. G. (2017). Efficient clinical decision support system. *International Journal of Engineering and Science*, 7, 5862-5872.
- Gibson, F. L., Rogers, A. A., Smith, A. D. M., Roberts, A., Possingham, H., McCarthy, M., Pannell, D. (2017). Factors influencing the use of decision support tools in the development and design of conservation policy. *Environmental Science and Policy*, 70, 1-8.
- Gupta, A. 2013. Environment & pest analysis: an approach to external business environment. *International Journal of Modern Social Science*, 2, 34-43.
- Gurel, E., and Tat, M. (2017). SWOT analysis: a theoretical overview. *The Journal of International Social Research*, 40, 995-1006.  
<https://pdfs.semanticscholar.org/48f9/b57dc3b69ae597850d27e2c7545256c480a5.pdf>
- Jiang, J., and Gallupe, B. R. (2015). Environmental Scanning and Business Insight Capability: The Role of Business Analytics and Knowledge Integration [PDF File]. Retrieved from
- Khan, R., Hossain, K. Rabbi, M. I. (2018). Strategic environment analysis of the Islamic Bank. *Journal of Case Studies in Management*, 9, 28-41.
- Kutilek, P. Using artificial intelligence for the support of strategic management. *Cybernetics*, 6, 1-23.
- Mekic, E., Mekic, E. (2014). Supports and Critiques on Porter's Competitive Strategy and Competitive Advantage. Retrieved from  
[https://www.researchgate.net/publication/266967720\\_Supports\\_and\\_Critiques\\_on\\_Porters\\_Competitive\\_Strategy\\_and\\_Competitive\\_Advantage](https://www.researchgate.net/publication/266967720_Supports_and_Critiques_on_Porters_Competitive_Strategy_and_Competitive_Advantage)
- Miller, B. D., Ranum, D. L., Anderson, J. (2019). *Python Programming in Context*. Jones and Bartlett Learning.



- Mindfire Solutions. (2017). Python: 7 Important Reasons Why You Should Use Python. Retrieved from <https://medium.com/@mindfiresolutions.usa/python-7-important-reasons-why-you-should-use-python-5801a98a0d0b>
- Pellissier, R. & Nenzhelele, T.E., 2013, Towards a universal definition of competitive intelligence. *SA Journal of Information Management*, 15. <http://dx.doi.org/10.4102/sajim.v15i2.559>
- Rose, D.C., Sutherland, J. W., Parker, C., Winter, M. M., Morris, C., Twinning, S.,...Dicks, V. (2016). Decision support tools for agriculture: towards effective design and delivery, *Agriculture Systems*, 149, 165-174.
- Sammut-Bonnici, T., and Galea, S. (2015). *PESTLE Analysis*. John Wiley and Sons Ltd.
- Schrempf-Stirling, J., Palazzo, G., Phillips, R. A. (2015). Historic Corporate Social Responsibility. *Academic of Management Review*, 41. <https://doi.org/10.5465/amr.2014.0137>
- Sim, W. L.L., Ban, K. H. K., Tan, W. T., Sethi, K. S., Loh, P. T. (2017). Development of a clinical decision support system for diabetes care: a pilot study. *PLoS One Journal*, 12. [10.1371/journal.pone.0173021](https://doi.org/10.1371/journal.pone.0173021)
- Team FME. (2013). PESTLE Analysis: Strategy Skills [PDF File]. Retrieved from <http://dl.icdst.org/pdfs/files/2d89fe2521e3327388a4e68bbf59a72b.pdf>
- Wang, H., Tong, L., Takeuchi, R., George, G. (2016). Corporate Social Responsibility: An overview and new research. *Academy of Management Journal*, 59. <https://doi.org/10.5465/amj.2016.5001>

## 8 Appendix

### 8.1 Code screenshots

#### 8.1.1 Worldwide Governance Indicators Manipulation

```
In [29]: import pandas as pd
import pandas as pd
xls=pd.ExcelFile('/users/Elson/Desktop/TESE/PEST/01P/wgidataset_v0.xlsx')
wgi_1 = pd.read_excel(xls, 'VoiceandAccountability',skiprows=13)
wgi_2 = pd.read_excel(xls, 'PoliticalStabilityNoViolence',skiprows=13)
wgi_3 = pd.read_excel(xls, 'GovernmentEffectiveness',skiprows=13)
wgi_4 = pd.read_excel(xls, 'RegulatoryQuality',skiprows=13)
wgi_5 = pd.read_excel(xls, 'Ruleoflaw',skiprows=13)
wgi_6 = pd.read_excel(xls, 'ControlofCorruption',skiprows=13)

In [30]: list_anos=list(wgi_1.columns)
list_anos_rank=['Unnamed: 0']

In [31]: for i in list_anos:
if ".3" in str(i):
list_anos_rank.append(str(i))

In [32]: print (list_anos_rank)

['Unnamed: 0', '1996.3', '1998.3', '2000.3', '2002.3', '2003.3', '2004.3', '2005.3', '2006.3', '2007.3', '2008.3', '2009.3', '2010.3', '2011.3', '2012.3', '2013.3', '2014.3', '2015.3', '2016.3', '2017.3']

In [33]: wgi_1_rank=wgi_1.loc[1:,list_anos_rank].dropna(axis='rows')
wgi_2_rank=wgi_2.loc[1:,list_anos_rank].dropna(axis='rows')
wgi_3_rank=wgi_3.loc[1:,list_anos_rank].dropna(axis='rows')
wgi_4_rank=wgi_4.loc[1:,list_anos_rank].dropna(axis='rows')
wgi_5_rank=wgi_5.loc[1:,list_anos_rank].dropna(axis='rows')
wgi_6_rank=wgi_6.loc[1:,list_anos_rank].dropna(axis='rows')

In [34]: rename_list=['COUNTRY']
for i in list(wgi_1_rank):
if ".3" in i:
rename_list.append(i[0:4])

In [35]: wgi_1_rank.columns = rename_list
wgi_2_rank.columns = rename_list
wgi_3_rank.columns = rename_list
wgi_4_rank.columns = rename_list
wgi_5_rank.columns = rename_list
wgi_6_rank.columns = rename_list

In [36]: wgi_1_rank.set_index('COUNTRY',inplace=True)
wgi_2_rank.set_index('COUNTRY',inplace=True)
wgi_3_rank.set_index('COUNTRY',inplace=True)
wgi_4_rank.set_index('COUNTRY',inplace=True)
wgi_5_rank.set_index('COUNTRY',inplace=True)
wgi_6_rank.set_index('COUNTRY',inplace=True)

In [37]: wgi_total=wgi_1_rank+wgi_2_rank+wgi_3_rank+wgi_4_rank+wgi_5_rank+wgi_6_rank

In [38]: wgi_avg=wgi_total/6

In [39]: wgi_avg.head()

Out[39]:
```

|             | 1996    | 1998    | 2000    | 2002    | 2003    | 2004    | 2005    | 2006    | 2007    | 2008    | 2009    | 2010    | 2011    | 2012    | 2013    | 2014    | :    |
|-------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|------|
| COUNTRY     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |      |
| Afghanistan | 1.76112 | 2.56899 | 1.60398 | 3.86213 | 7.13263 | 8.11849 | 5.81987 | 5.32095 | 5.06325 | 3.93951 | 3.17241 | 3.80809 | 3.85192 | 6.21565 | 6.2939  | 7.99324 | 8.6  |
| Albania     | 28.209  | 28.4821 | 29.5304 | 35.6196 | 35.3932 | 37.7156 | 34.6148 | 36.2842 | 41.0733 | 44.1242 | 46.0373 | 45.372  | 43.7703 | 43.6175 | 45.0385 | 49.9426 | 51.1 |
| Algeria     | 16.4898 | 15.7295 | 15.5154 | 23.8165 | 24.6949 | 26.6778 | 32.6603 | 28.3087 | 25.5429 | 24.4158 | 22.6411 | 23.644  | 22.6784 | 23.3818 | 25.6636 | 22.361  | 21.  |
| Andorra     | 89.6598 | 90.0586 | 89.3628 | 90.4566 | 90.7752 | 90.592  | 90.4622 | 89.2803 | 91.3071 | 91.217  | 90.8709 | 90.2432 | 91.9698 | 91.7329 | 91.3387 | 88.9784 | 89.  |
| Angola      | 8.63976 | 4.32699 | 3.62298 | 7.7985  | 10.1604 | 9.641   | 11.7369 | 12.5099 | 12.9679 | 14.7988 | 15.458  | 15.4489 | 14.4954 | 17.023  | 15.4425 | 15.5534 | 15.  |

```
In [40]: wgi_avg.to_excel("/users/Elson/Desktop/TESE/PEST/01P/wgi_v1.xlsx")
```

## 8.1.2 Index of Economic Freedom

```
In [2]: import pandas as pd
xls=pd.ExcelFile('/users/Elson/Desktop/TESE/PEST/01P/index2013_v0.xls')
IEF_2013 = pd.read_excel(xls, 'IEF_2013')
IEF_2014 = pd.read_excel(xls, 'IEF_2014')
IEF_2015 = pd.read_excel(xls, 'IEF_2015')
IEF_2016 = pd.read_excel(xls, 'IEF_2016')
IEF_2017 = pd.read_excel(xls, 'IEF_2017')
IEF_2018 = pd.read_excel(xls, 'IEF_2018')
IEF_2019 = pd.read_excel(xls, 'IEF_2019')
```

```
In [3]: IEF_2013=IEF_2013.loc[:,['Country Name','2013 Score']].dropna(axis='rows').set_index('Country Name')
IEF_2014=IEF_2014.loc[:,['Country Name','2014 Score']].dropna(axis='rows').set_index('Country Name')
IEF_2015=IEF_2015.loc[:,['Country Name','2015 Score']].dropna(axis='rows').set_index('Country Name')
IEF_2016=IEF_2016.loc[:,['Country Name','2016 Score']].dropna(axis='rows').set_index('Country Name')
IEF_2017=IEF_2017.loc[:,['Country Name','2017 Score']].dropna(axis='rows').set_index('Country Name')
IEF_2018=IEF_2018.loc[:,['Country Name','2018 Score']].dropna(axis='rows').set_index('Country Name')
IEF_2019=IEF_2019.loc[:,['Country Name','2019 Score']].dropna(axis='rows').set_index('Country Name')
```

```
In [4]: dfs=[IEF_2013,IEF_2014,IEF_2015,IEF_2016,IEF_2017,IEF_2018,IEF_2019]
```

```
In [5]: mergedDf_1 = IEF_2013.merge(IEF_2014, on='Country Name')
mergedDf_2 = mergedDf_1.merge(IEF_2015, on='Country Name')
mergedDf_3 = mergedDf_2.merge(IEF_2016, on='Country Name')
mergedDf_4 = mergedDf_3.merge(IEF_2017, on='Country Name')
mergedDf_5 = mergedDf_4.merge(IEF_2018, on='Country Name')
mergedDf_6 = mergedDf_5.merge(IEF_2019, on='Country Name')
```

```
In [6]: IEF_SCORE=mergedDf_6
IEF_SCORE.columns=['2013','2014','2015','2016','2017','2018','2019']
```

```
In [7]: IEF_SCORE.head()
```

Out[7]:

|              | 2013 | 2014      | 2015      | 2016  | 2017      | 2018 | 2019 |
|--------------|------|-----------|-----------|-------|-----------|------|------|
| Country Name |      |           |           |       |           |      |      |
| Albania      | 65.2 | 66.892240 | 65.650034 | 65.92 | 64.353003 | 64.5 | 66.5 |
| Algeria      | 49.6 | 50.843830 | 48.881858 | 50.06 | 46.528581 | 44.7 | 46.2 |
| Angola       | 47.3 | 47.703602 | 47.885804 | 48.94 | 48.515521 | 48.6 | 50.6 |
| Argentina    | 46.7 | 44.553346 | 44.138922 | 43.77 | 50.409088 | 52.3 | 52.2 |
| Armenia      | 69.4 | 68.861097 | 67.125092 | 66.96 | 70.306017 | 68.7 | 67.7 |

```
In [8]: IEF_SCORE.to_excel("/users/Elson/Desktop/TESE/PEST/01P/index_v1.xlsx")
```

```
In [ ]:
```

## 8.1.3 Transparency International

```
In [8]: import pandas as pd
        cpi=pd.read_excel('/users/Elson/Desktop/TESE/PEST/01P/CPI_FullDataSet.xlsx','CPI Timeseries 2012 - 2018')
```

```
In [9]: cpi.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 26 columns):
Country                180 non-null object
ISO3                   180 non-null object
Region                 180 non-null object
CPI score 2018         180 non-null int64
Rank 2018              180 non-null int64
Sources 2018           180 non-null int64
Standard error 2018    180 non-null float64
CPI score 2017         180 non-null int64
Rank 2017              180 non-null int64
Sources 2017           180 non-null int64
Standard error 2017    180 non-null float64
CPI score 2016         176 non-null float64
Sources 2016           176 non-null float64
Standard error 2016    176 non-null float64
CPI score 2015         168 non-null float64
Sources 2015           168 non-null float64
Standard error 2015    168 non-null float64
CPI score 2014         174 non-null float64
Sources 2014           174 non-null float64
Standard error 2014    174 non-null float64
CPI Score 2013         176 non-null float64
Sources 2013           176 non-null float64
Standard error 2013    176 non-null float64
CPI Score 2012         175 non-null float64
Sources 2012           175 non-null float64
Standard error 2012    175 non-null float64
dtypes: float64(17), int64(6), object(3)
memory usage: 36.6+ KB
```

```
In [2]: list_cpi_score=['Country']
        for i in list(cpi.columns):
            if "CPI SCORE" in i.upper():
                list_cpi_score.append(i)
```

```
In [3]: print(list_cpi_score)

['Country', 'CPI score 2018', 'CPI score 2017', 'CPI score 2016', 'CPI score 2015', 'CPI score 2014', 'CPI Score 2013', 'CPI Score 2012']
```

```
In [4]: cpi_score=cpi.loc[:,list_cpi_score]
        cpi_score.columns = ['COUNTRY', '2018', '2017', '2016', '2015', '2014', '2013', '2012']
        cpi_score=cpi_score.dropna(axis='rows')
        cpi_score.set_index('COUNTRY',inplace=True)
```

```
In [6]: cpi_score.head()
```

```
Out[6]:
```

|                | 2018 | 2017 | 2016 | 2015 | 2014 | 2013 | 2012 |
|----------------|------|------|------|------|------|------|------|
| <b>COUNTRY</b> |      |      |      |      |      |      |      |
| Denmark        | 88   | 88   | 90.0 | 91.0 | 92.0 | 91.0 | 90.0 |
| New Zealand    | 87   | 89   | 90.0 | 91.0 | 91.0 | 91.0 | 90.0 |
| Finland        | 85   | 85   | 89.0 | 90.0 | 89.0 | 89.0 | 90.0 |
| Singapore      | 85   | 84   | 84.0 | 85.0 | 84.0 | 86.0 | 87.0 |
| Sweden         | 85   | 84   | 88.0 | 89.0 | 87.0 | 89.0 | 88.0 |

```
In [ ]: cpi_score.to_excel("/users/Elson/Desktop/TESE/PEST/01P/cpi_v1.xlsx")

In [ ]:
```

## 8.1.4 Doing Business Law Library

```
In [1]: import pandas as pd |
        dbll=pd.read_excel('/users/Elson/Desktop/TESE/PEST/01P/dbll_v0.xlsx', 'All Data')

In [6]: dbll_score=dbll.loc[:,['Economy', 'DB Year', 'Score-Starting a business']]

In [7]: dbll_score.head()
Out[7]:
```

|   | Economy     | DB Year | Score-Starting a business |
|---|-------------|---------|---------------------------|
| 0 | Afghanistan | 2005    | 83.59                     |
| 1 | Afghanistan | 2006    | 83.19                     |
| 2 | Afghanistan | 2007    | 84.17                     |
| 3 | Afghanistan | 2008    | 80.29                     |
| 4 | Afghanistan | 2009    | 83.44                     |

```
In [8]: dbll_score = dbll_score.pivot(index='Economy', columns='DB Year', values='Score-Starting a business')

In [10]: dbll_score.head()
Out[10]:
```

|                     | DB Year | 2004  | 2005  | 2006  | 2007  | 2008  | 2009  | 2010  | 2011  | 2012  | 2013  | 2014  | 2015  | 2016  | 2017  | 2018  | 2019 |
|---------------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| Economy             |         |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |      |
| Afghanistan         | NaN     | 83.59 | 83.19 | 84.17 | 80.29 | 83.44 | 87.35 | 87.78 | 87.90 | 88.31 | 91.55 | 90.95 | 90.47 | 90.35 | 82.55 | 92.04 |      |
| Albania             | 63.68   | 67.28 | 67.25 | 69.04 | 68.21 | 82.00 | 87.24 | 87.29 | 87.60 | 90.19 | 90.34 | 91.73 | 91.69 | 91.43 | 91.49 | 91.58 |      |
| Algeria             | 68.11   | 70.53 | 71.96 | 72.40 | 72.48 | 73.32 | 73.50 | 73.18 | 73.53 | 73.74 | 73.61 | 74.57 | 76.59 | 78.04 | 77.86 | 78.07 |      |
| Angola              | 13.09   | 13.09 | 31.84 | 33.47 | 34.94 | 45.71 | 52.04 | 46.46 | 51.28 | 53.82 | 55.63 | 57.15 | 77.48 | 79.67 | 80.09 | 80.52 |      |
| Antigua and Barbuda | NaN     | NaN   | 80.27 | 81.34 | 81.42 | 81.45 | 81.69 | 81.52 | 81.34 | 81.53 | 81.58 | 81.61 | 81.65 | 81.66 | 81.69 | 81.74 |      |

```
In [7]: dbll_score.to_excel("/users/Elson/Desktop/TESE/PEST/01P/dbll_v1.xlsx")
```

## 8.1.5 General Manipulation

```
In [24]: import pandas as pd
path='/users/Elson/Desktop/TESE/PEST/01P/'
wgi= pd.read_excel(path+'wgi_v1.xlsx')
index= pd.read_excel(path+'index_v1.xlsx')
cpi= pd.read_excel(path+'cpi_v1.xlsx')
dbll= pd.read_excel(path+'dbll_v1.xlsx')
index.columns=['COUNTRY', '2013', '2014', '2015', '2016', '2017', '2018', '2019']
dbll.columns=['COUNTRY', '2004', '2005', '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017']

In [25]: wgi_year=list(wgi.columns)[1:]
index_year=list(index.columns)[1:]
cpi_year=list(cpi.columns)[1:]
dbll_year=list(dbll.columns)[1:]

In [26]: common_years=['COUNTRY']
year=2012
while year<=2019:
    if str(year) in wgi_year and str(year) in index_year and str(year) in cpi_year and str(year) in dbll_year:
        common_years.append(str(year))
    year+=1

In [27]: print(common_years)

['COUNTRY', '2013', '2014', '2015', '2016', '2017']

In [28]: wgi=wgi.loc[:,common_years].set_index('COUNTRY')
index=index.loc[:,common_years].set_index('COUNTRY')
cpi=cpi.loc[:,common_years].set_index('COUNTRY')
dbll=dbll.loc[:,common_years].set_index('COUNTRY')

In [6]: merge_1=wgi.merge(index, how='left', on='COUNTRY', left_on=None, right_on=None,
    left_index=False, right_index=False, sort=True,
    suffixes=('_', 'wgi', '_index'), copy=True, indicator=False,
    validate=None)

merge_2=merge_1.merge(cpi, how='left', on='COUNTRY', left_on=None, right_on=None,
    left_index=False, right_index=False, sort=True,
    suffixes=('_', '_cpi'), copy=True, indicator=False,
    validate=None)

merge_3=merge_2.merge(dbll, how='left', on='COUNTRY', left_on=None, right_on=None,
    left_index=False, right_index=False, sort=True,
    suffixes=('_', '_dbll'), copy=True, indicator=False,
    validate=None)

In [7]: all_scores=merge_3.dropna(axis='rows')

In [8]: all_scores.shape

Out[8]: (108, 20)

In [9]: all_scores

Out[9]:
```

|                | 2013_wgi  | 2014_wgi  | 2015_wgi  | 2016_wgi  | 2017_wgi  | 2013_index | 2014_index | 2015_index | 2016_index | 2017_index | 2013 | 2014 | 2015 | 2016 | 2017 |
|----------------|-----------|-----------|-----------|-----------|-----------|------------|------------|------------|------------|------------|------|------|------|------|------|
| <b>COUNTRY</b> |           |           |           |           |           |            |            |            |            |            |      |      |      |      |      |
| Albania        | 45.038530 | 49.942607 | 51.079613 | 50.997510 | 52.601864 | 65.2       | 66.892240  | 65.650034  | 65.92      | 64.353003  | 31.0 | 33.0 | 36.0 | 39.0 | 38.0 |
| Algeria        | 25.663619 | 22.361032 | 21.714218 | 21.308866 | 21.383205 | 49.6       | 50.843830  | 48.881858  | 50.06      | 46.528581  | 36.0 | 36.0 | 36.0 | 34.0 | 33.0 |
| Angola         | 15.442524 | 15.553398 | 15.480901 | 15.879068 | 16.202002 | 47.3       | 47.703602  | 47.885804  | 48.94      | 48.515521  | 23.0 | 19.0 | 15.0 | 18.0 | 19.0 |
| Argentina      | 39.831935 | 36.301393 | 38.548166 | 49.208349 | 52.260352 | 46.7       | 44.553346  | 44.138922  | 43.77      | 50.409088  | 34.0 | 34.0 | 32.0 | 36.0 | 39.0 |
| Armenia        | 46.092087 | 41.491411 | 42.046072 | 41.434624 | 41.521752 | 69.4       | 68.861097  | 67.125092  | 66.96      | 70.306017  | 36.0 | 37.0 | 35.0 | 33.0 | 35.0 |

```
In [15]: all_scores.to_excel("/users/Elson/Desktop/TESE/PEST/01P/p_scores.xlsx")

In [30]: p_scores= pd.read_excel(path+'p_scores.xlsx')

In [31]: p_scores['2013_y']=(p_scores['2013_wgi']+p_scores['2013_index']+p_scores['2013']+p_scores['2013_dbll'])/4
p_scores['2014_y']=(p_scores['2014_wgi']+p_scores['2014_index']+p_scores['2014']+p_scores['2014_dbll'])/4
p_scores['2015_y']=(p_scores['2015_wgi']+p_scores['2015_index']+p_scores['2015']+p_scores['2015_dbll'])/4
p_scores['2016_y']=(p_scores['2016_wgi']+p_scores['2016_index']+p_scores['2016']+p_scores['2016_dbll'])/4
p_scores['2017_y']=(p_scores['2017_wgi']+p_scores['2017_index']+p_scores['2017']+p_scores['2017_dbll'])/4

In [32]: p_scores.head()

Out[32]:
```

|   | COUNTRY   | 2013_wgi  | 2014_wgi  | 2015_wgi  | 2016_wgi  | 2017_wgi  | 2013_index | 2014_index | 2015_index | 2016_index | ... | 2013_dbll | 2014_dbll | 2015_dbll | 2016_ |
|---|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|------------|------------|-----|-----------|-----------|-----------|-------|
| 0 | Albania   | 45.038530 | 49.942607 | 51.079613 | 50.997510 | 52.601864 | 65.2       | 66.892240  | 65.650034  | 65.92      | ... | 90.19     | 90.34     | 91.73     | 9     |
| 1 | Algeria   | 25.663619 | 22.361032 | 21.714218 | 21.308866 | 21.383205 | 49.6       | 50.843830  | 48.881858  | 50.06      | ... | 73.74     | 73.61     | 74.57     | 7     |
| 2 | Angola    | 15.442524 | 15.553398 | 15.480901 | 15.879068 | 16.202002 | 47.3       | 47.703602  | 47.885804  | 48.94      | ... | 53.82     | 55.63     | 57.15     | 7     |
| 3 | Argentina | 39.831935 | 36.301393 | 38.548166 | 49.208349 | 52.260352 | 46.7       | 44.553346  | 44.138922  | 43.77      | ... | 72.58     | 71.59     | 72.39     | 7     |
| 4 | Armenia   | 46.092087 | 41.491411 | 42.046072 | 41.434624 | 41.521752 | 69.4       | 68.861097  | 67.125092  | 66.96      | ... | 93.39     | 94.45     | 94.45     | 9     |

```
5 rows x 26 columns

In [33]: p_scores.columns

Out[33]: Index(['COUNTRY', '2013_wgi', '2014_wgi', '2015_wgi', '2016_wgi', '2017_wgi',
        '2013_index', '2014_index', '2015_index', '2016_index', '2017_index',
        '2013', '2014', '2015', '2016', '2017', '2013_dbll', '2014_dbll',
        '2015_dbll', '2016_dbll', '2017_dbll', '2013_y', '2014_y', '2015_y',
        '2016_y', '2017_y'],
        dtype='object')

In [109]: p_scores.to_excel("/users/Elson/Desktop/TESE/PEST/01P/p.xlsx")
```

## 8.1.6 Validation of the predictors shape

```
In [51]: import pandas as pd
import keras
from keras.layers import Dense
from keras.models import Sequential
from keras.optimizers import SGD
import numpy as np
%matplotlib inline
from keras.callbacks import EarlyStopping
import matplotlib.pyplot as plt

In [52]: path='/users/Elson/Desktop/TESE/PEST/01P/political_indexes.xlsx'
path_orig='/users/Elson/Desktop/TESE/PEST/01P/political.xlsx'
political= pd.read_excel(path)

In [53]: political.head()

Out[53]:
```

|   | wgi_x     | index_x | cpi_x | dbl_x | Y         |
|---|-----------|---------|-------|-------|-----------|
| 0 | 45.038530 | 65.2    | 31    | 90.19 | 60.043712 |
| 1 | 25.663619 | 49.6    | 36    | 73.74 | 45.703716 |
| 2 | 15.442524 | 47.3    | 23    | 53.82 | 34.471750 |
| 3 | 39.831935 | 46.7    | 34    | 72.58 | 46.611185 |
| 4 | 46.092087 | 69.4    | 36    | 93.39 | 60.450627 |

```
In [54]: predictors=np.array(political.loc[:,['wgi_x','index_x','cpi_x','dbl_x']])
target=np.array(political.loc[:, 'Y'])

In [55]: predictors.shape

Out[55]: (432, 4)
```

## 8.1.7 ADAM TESTING

```
In [80]: def get_new_model(input_shape=input_shape):
model = Sequential()
model.add(Dense(400, activation='relu', input_shape = input_shape))
model.add(Dense(400, activation='relu'))
model.add(Dense(400, activation='relu'))
model.add(Dense(1))
return(model)

# Create list of learning rates: lr_to_test
lr_to_test = [.000001, .00001, .000015]

# Loop over learning rates
for lr in lr_to_test:
    print('\n\nTesting model with learning rate: %f\n%lr )

    # Build new model to test, unaffected by previous models
    model = get_new_model()

    # Create SGD optimizer with specified learning rate: my_optimizer
    my_optimizer = keras.optimizers.Adam(lr=lr)

    # Compile the model
    model.compile(optimizer=my_optimizer, loss='mean_squared_error')

    # Fit the model
    model.fit(predictors, target, epochs=20)
```

Testing model with learning rate: 0.000001

```
Epoch 1/20
432/432 [=====] - 2s 4ms/step - loss: 3868.0837
Epoch 2/20
432/432 [=====] - 0s 168us/step - loss: 3827.3575
Epoch 3/20
432/432 [=====] - 0s 167us/step - loss: 3786.6056
Epoch 4/20
432/432 [=====] - 0s 163us/step - loss: 3746.3507
Epoch 5/20
432/432 [=====] - 0s 168us/step - loss: 3706.6546
Epoch 6/20
432/432 [=====] - 0s 168us/step - loss: 3667.6542
Epoch 7/20
432/432 [=====] - 0s 167us/step - loss: 3628.5436
Epoch 8/20
432/432 [=====] - 0s 163us/step - loss: 3589.8309
Epoch 9/20
432/432 [=====] - 0s 164us/step - loss: 3551.1799
Epoch 10/20
432/432 [=====] - 0s 166us/step - loss: 3513.1326
Epoch 11/20
432/432 [=====] - 0s 168us/step - loss: 3475.1666
Epoch 12/20
432/432 [=====] - 0s 165us/step - loss: 3437.7772
Epoch 13/20
432/432 [=====] - 0s 164us/step - loss: 3400.2705
Epoch 14/20
432/432 [=====] - 0s 161us/step - loss: 3363.1754
Epoch 15/20
432/432 [=====] - 0s 162us/step - loss: 3326.4348
Epoch 16/20
432/432 [=====] - 0s 163us/step - loss: 3290.2506
Epoch 17/20
432/432 [=====] - 0s 160us/step - loss: 3253.9064
Epoch 18/20
432/432 [=====] - 0s 161us/step - loss: 3217.4263
Epoch 19/20
432/432 [=====] - 0s 168us/step - loss: 3181.4234
Epoch 20/20
432/432 [=====] - 0s 167us/step - loss: 3145.3188
```

Testing model with learning rate: 0.000010

```
Epoch 1/20
432/432 [=====] - 2s 4ms/step - loss: 4252.1557
Epoch 2/20
432/432 [=====] - 0s 176us/step - loss: 3770.0390
Epoch 3/20
432/432 [=====] - 0s 183us/step - loss: 3325.8139
Epoch 4/20
432/432 [=====] - 0s 178us/step - loss: 2911.9924
Epoch 5/20
432/432 [=====] - 0s 176us/step - loss: 2533.6670
Epoch 6/20
432/432 [=====] - 0s 179us/step - loss: 2187.6037
Epoch 7/20
432/432 [=====] - 0s 178us/step - loss: 1870.9017
Epoch 8/20
432/432 [=====] - 0s 178us/step - loss: 1581.4261
Epoch 9/20
432/432 [=====] - 0s 177us/step - loss: 1325.3999
Epoch 10/20
432/432 [=====] - 0s 174us/step - loss: 1098.4974
Epoch 11/20
432/432 [=====] - 0s 180us/step - loss: 894.9062
Epoch 12/20
432/432 [=====] - 0s 173us/step - loss: 719.5123
Epoch 13/20
432/432 [=====] - 0s 177us/step - loss: 565.3451
Epoch 14/20
432/432 [=====] - 0s 249us/step - loss: 434.5363
Epoch 15/20
432/432 [=====] - 0s 276us/step - loss: 322.4651
Epoch 16/20
432/432 [=====] - 0s 236us/step - loss: 233.0061
Epoch 17/20
432/432 [=====] - 0s 185us/step - loss: 161.1216
Epoch 18/20
432/432 [=====] - 0s 178us/step - loss: 107.2027
Epoch 19/20
432/432 [=====] - 0s 177us/step - loss: 69.4397
Epoch 20/20
432/432 [=====] - 0s 263us/step - loss: 43.7931
```

Testing model with learning rate: 0.000015

```
Epoch 1/20
432/432 [=====] - 2s 4ms/step - loss: 4235.3580
Epoch 2/20
432/432 [=====] - 0s 182us/step - loss: 3636.8319
Epoch 3/20
432/432 [=====] - 0s 179us/step - loss: 3087.5278
Epoch 4/20
432/432 [=====] - 0s 183us/step - loss: 2592.5642
Epoch 5/20
432/432 [=====] - 0s 182us/step - loss: 2145.3701
Epoch 6/20
432/432 [=====] - 0s 180us/step - loss: 1745.0755
Epoch 7/20
432/432 [=====] - 0s 177us/step - loss: 1384.9786
Epoch 8/20
432/432 [=====] - 0s 181us/step - loss: 1062.1385
Epoch 9/20
432/432 [=====] - 0s 177us/step - loss: 778.2376
Epoch 10/20
432/432 [=====] - 0s 273us/step - loss: 544.4986
Epoch 11/20
432/432 [=====] - 0s 338us/step - loss: 355.3278
Epoch 12/20
432/432 [=====] - 0s 336us/step - loss: 221.0617
Epoch 13/20
432/432 [=====] - 0s 169us/step - loss: 127.0719
Epoch 14/20
432/432 [=====] - 0s 155us/step - loss: 65.7963
Epoch 15/20
432/432 [=====] - 0s 154us/step - loss: 31.3032
Epoch 16/20
432/432 [=====] - 0s 169us/step - loss: 14.1665
Epoch 17/20
432/432 [=====] - 0s 158us/step - loss: 7.5178
Epoch 18/20
432/432 [=====] - 0s 156us/step - loss: 5.1369
Epoch 19/20
432/432 [=====] - 0s 160us/step - loss: 4.4370
Epoch 20/20
432/432 [=====] - 0s 161us/step - loss: 4.0774
```

Considering the lowest loss: 4.0774 the selected learning rate for ADAM is **0.000015**



## 8.1.8 SGD TESTING

```
In [81]: def get_new_model(input_shape=input_shape):
model = Sequential()
model.add(Dense(400, activation='relu', input_shape = input_shape))
model.add(Dense(400, activation='relu'))
model.add(Dense(400, activation='relu'))
model.add(Dense(1))
return(model)

# Create list of learning rates: lr_to_test
lr_to_test = [.000001,.00001,.000015]

# Loop over learning rates
for lr in lr_to_test:
    print('\n\nTesting model with learning rate: %f\n\n%lr )

    # Build new model to test, unaffected by previous models
    model = get_new_model()

    # Create SGD optimizer with specified learning rate: my_optimizer
    my_optimizer = SGD(lr=lr)

    # Compile the model
    model.compile(optimizer=my_optimizer,loss='mean_squared_error')

    # Fit the model
    model.fit(predictors,target,epochs=20)
```

Testing model with learning rate: 0.000001

```
Epoch 1/20
432/432 [=====] - 2s 4ms/step - loss: 2494.6489
Epoch 2/20
432/432 [=====] - 0s 154us/step - loss: 740.4357
Epoch 3/20
432/432 [=====] - 0s 143us/step - loss: 160.3113
Epoch 4/20
432/432 [=====] - 0s 143us/step - loss: 27.5597
Epoch 5/20
432/432 [=====] - 0s 150us/step - loss: 6.4860
Epoch 6/20
432/432 [=====] - 0s 147us/step - loss: 3.7141
Epoch 7/20
432/432 [=====] - 0s 182us/step - loss: 3.3153
Epoch 8/20
432/432 [=====] - 0s 159us/step - loss: 3.1996
Epoch 9/20
432/432 [=====] - 0s 152us/step - loss: 3.1254
Epoch 10/20
432/432 [=====] - 0s 145us/step - loss: 3.0596
Epoch 11/20
432/432 [=====] - 0s 159us/step - loss: 2.9980
Epoch 12/20
432/432 [=====] - 0s 155us/step - loss: 2.9439
Epoch 13/20
432/432 [=====] - 0s 165us/step - loss: 2.8887
Epoch 14/20
432/432 [=====] - 0s 142us/step - loss: 2.8421
Epoch 15/20
432/432 [=====] - 0s 167us/step - loss: 2.7943
Epoch 16/20
432/432 [=====] - 0s 149us/step - loss: 2.7514
Epoch 17/20
432/432 [=====] - 0s 162us/step - loss: 2.7137
Epoch 18/20
432/432 [=====] - 0s 156us/step - loss: 2.6793
Epoch 19/20
432/432 [=====] - 0s 145us/step - loss: 2.6422
Epoch 20/20
432/432 [=====] - 0s 153us/step - loss: 2.6086
```

Testing model with learning rate: 0.000010

```
Epoch 1/20
432/432 [=====] - 2s 4ms/step - loss: 315.6740
Epoch 2/20
432/432 [=====] - 0s 148us/step - loss: 2.7184
Epoch 3/20
432/432 [=====] - 0s 152us/step - loss: 2.5375
Epoch 4/20
432/432 [=====] - 0s 151us/step - loss: 2.4840
Epoch 5/20
432/432 [=====] - 0s 156us/step - loss: 2.3886
Epoch 6/20
432/432 [=====] - 0s 159us/step - loss: 2.3193
Epoch 7/20
432/432 [=====] - 0s 155us/step - loss: 2.3314
Epoch 8/20
432/432 [=====] - 0s 155us/step - loss: 2.2504
Epoch 9/20
432/432 [=====] - 0s 155us/step - loss: 2.2491
Epoch 10/20
432/432 [=====] - 0s 168us/step - loss: 2.2204
Epoch 11/20
432/432 [=====] - 0s 147us/step - loss: 2.1798
Epoch 12/20
432/432 [=====] - 0s 148us/step - loss: 2.1810
Epoch 13/20
432/432 [=====] - 0s 159us/step - loss: 2.1242
Epoch 14/20
432/432 [=====] - 0s 156us/step - loss: 2.1271
Epoch 15/20
432/432 [=====] - 0s 150us/step - loss: 2.1566
Epoch 16/20
432/432 [=====] - 0s 246us/step - loss: 2.1269
Epoch 17/20
432/432 [=====] - 0s 240us/step - loss: 2.0835
Epoch 18/20
432/432 [=====] - 0s 210us/step - loss: 2.1251
Epoch 19/20
432/432 [=====] - 0s 153us/step - loss: 2.0702
Epoch 20/20
432/432 [=====] - 0s 171us/step - loss: 2.0201
```

Testing model with learning rate: 0.000015

```
Epoch 1/20
432/432 [=====] - 2s 4ms/step - loss: 375.2380
Epoch 2/20
432/432 [=====] - 0s 145us/step - loss: 2.3295
Epoch 3/20
432/432 [=====] - 0s 147us/step - loss: 2.3146
Epoch 4/20
432/432 [=====] - 0s 158us/step - loss: 2.2782
Epoch 5/20
432/432 [=====] - 0s 155us/step - loss: 2.2184
Epoch 6/20
432/432 [=====] - 0s 147us/step - loss: 2.1702
Epoch 7/20
432/432 [=====] - 0s 157us/step - loss: 2.1754
Epoch 8/20
432/432 [=====] - 0s 150us/step - loss: 2.1444
Epoch 9/20
432/432 [=====] - 0s 155us/step - loss: 2.1152
Epoch 10/20
432/432 [=====] - 0s 158us/step - loss: 2.0675
Epoch 11/20
432/432 [=====] - 0s 167us/step - loss: 2.0518
Epoch 12/20
432/432 [=====] - 0s 154us/step - loss: 2.0990
Epoch 13/20
432/432 [=====] - 0s 168us/step - loss: 2.0630
Epoch 14/20
432/432 [=====] - 0s 273us/step - loss: 2.0827
Epoch 15/20
432/432 [=====] - 0s 345us/step - loss: 2.0166
Epoch 16/20
432/432 [=====] - 0s 243us/step - loss: 2.0603
Epoch 17/20
432/432 [=====] - 0s 152us/step - loss: 1.9826
Epoch 18/20
432/432 [=====] - 0s 156us/step - loss: 1.9918
Epoch 19/20
432/432 [=====] - 0s 149us/step - loss: 1.9954
Epoch 20/20
432/432 [=====] - 0s 158us/step - loss: 1.9787
```

Considering the lowest loss: 1.9787 the selected learning rate for SGD is also **0.000015**

## 8.1.9 Comparing performances of model\_1 and model\_2

In [83]:

```
# Define the early_stopping_monitor to stop epochs when there is no more improvements with delay of two epochs
early_stopping_monitor = EarlyStopping(monitor='val_loss',patience=2)

# The input shape to use in the first hidden layer, which is 4, one per datasource
input_shape = predictors.shape[1:]

# Create model_1
model_1 = Sequential()

# Add the first, second, and third hidden layers
model_1.add(Dense(400,activation='relu',input_shape=input_shape))
model_1.add(Dense(400,activation='relu'))
model_1.add(Dense(400,activation='relu'))

# Add the output layer
model_1.add(Dense(1))

# Define Adam optimizer with learning rate of 0.00001
opt_1=keras.optimizers.Adam(lr=0.000015)

# Compile model_1
model_1.compile(optimizer=opt_1,loss='mean_squared_error')

# Create the new model: model_2
model_2 = Sequential()

# Add the first, second, and third hidden layers
model_2.add(Dense(400,activation='relu',input_shape=input_shape))
model_2.add(Dense(400,activation='relu'))
model_2.add(Dense(400,activation='relu'))

# Define SGD optimizer with learning rate of 0.000001
opt=SGD(lr=0.000015)

# Add the output layer
model_2.add(Dense(1))

# Compile model_2
model_2.compile(optimizer=opt,loss='mean_squared_error')

# Fit model 1
model_1_training = model_1.fit(predictors, target, epochs=20, validation_split=0.1, callbacks=[early_stopping_monitor])

# Fit model 2
model_2_training = model_2.fit(predictors, target, epochs=20, validation_split=0.1, callbacks=[early_stopping_monitor])

# Create the plot
plt.plot(model_1_training.history['val_loss'], 'r', model_2_training.history['val_loss'], 'b')
plt.xlabel('Epochs')
plt.ylabel('Validation score')
plt.show()
```

## 8.1.10 Results of the model\_1 (Adam with lr=0.000015)

```
Train on 388 samples, validate on 44 samples
Epoch 1/20
388/388 [=====] - 2s 6ms/step - loss: 3607.4285 - val_loss: 3429.4519
Epoch 2/20
388/388 [=====] - 0s 239us/step - loss: 2957.6667 - val_loss: 2788.5750
Epoch 3/20
388/388 [=====] - 0s 244us/step - loss: 2385.7892 - val_loss: 2226.3682
Epoch 4/20
388/388 [=====] - 0s 245us/step - loss: 1893.6034 - val_loss: 1742.9021
Epoch 5/20
388/388 [=====] - 0s 194us/step - loss: 1466.3073 - val_loss: 1334.3215
Epoch 6/20
388/388 [=====] - 0s 196us/step - loss: 1107.6111 - val_loss: 983.9957
Epoch 7/20
388/388 [=====] - 0s 195us/step - loss: 803.7171 - val_loss: 696.7345
Epoch 8/20
388/388 [=====] - 0s 200us/step - loss: 556.2684 - val_loss: 465.1693
Epoch 9/20
388/388 [=====] - 0s 196us/step - loss: 362.7449 - val_loss: 288.8918
Epoch 10/20
388/388 [=====] - 0s 195us/step - loss: 218.9140 - val_loss: 166.0699
Epoch 11/20
388/388 [=====] - 0s 196us/step - loss: 121.3406 - val_loss: 84.0123
Epoch 12/20
388/388 [=====] - 0s 190us/step - loss: 58.8809 - val_loss: 37.8426
Epoch 13/20
388/388 [=====] - 0s 199us/step - loss: 26.0732 - val_loss: 15.2113
Epoch 14/20
388/388 [=====] - 0s 196us/step - loss: 10.9920 - val_loss: 6.1404
Epoch 15/20
388/388 [=====] - 0s 196us/step - loss: 5.4964 - val_loss: 3.3483
Epoch 16/20
388/388 [=====] - 0s 194us/step - loss: 4.0103 - val_loss: 2.7054
Epoch 17/20
388/388 [=====] - 0s 194us/step - loss: 3.6507 - val_loss: 2.5092
Epoch 18/20
388/388 [=====] - 0s 197us/step - loss: 3.5136 - val_loss: 2.3650
Epoch 19/20
388/388 [=====] - 0s 196us/step - loss: 3.3695 - val_loss: 2.2190
Epoch 20/20
388/388 [=====] - 0s 198us/step - loss: 3.2358 - val_loss: 2.0891
```

## 8.1.11 Results of the model\_2 (SGD with lr=0.000015):

```
Train on 388 samples, validate on 44 samples
Epoch 1/20
388/388 [=====] - 2s 6ms/step - loss: 3607.4285 - val_loss: 3429.4519
Epoch 2/20
388/388 [=====] - 0s 239us/step - loss: 2957.6667 - val_loss: 2788.5750
Epoch 3/20
388/388 [=====] - 0s 244us/step - loss: 2385.7892 - val_loss: 2226.3682
Epoch 4/20
388/388 [=====] - 0s 245us/step - loss: 1893.6034 - val_loss: 1742.9021
Epoch 5/20
388/388 [=====] - 0s 194us/step - loss: 1466.3073 - val_loss: 1334.3215
Epoch 6/20
388/388 [=====] - 0s 196us/step - loss: 1107.6111 - val_loss: 983.9957
Epoch 7/20
388/388 [=====] - 0s 195us/step - loss: 803.7171 - val_loss: 696.7345
Epoch 8/20
388/388 [=====] - 0s 200us/step - loss: 556.2684 - val_loss: 465.1693
Epoch 9/20
388/388 [=====] - 0s 196us/step - loss: 362.7449 - val_loss: 288.8918
Epoch 10/20
388/388 [=====] - 0s 195us/step - loss: 218.9140 - val_loss: 166.0699
Epoch 11/20
388/388 [=====] - 0s 196us/step - loss: 121.3406 - val_loss: 84.0123
Epoch 12/20
388/388 [=====] - 0s 190us/step - loss: 58.8809 - val_loss: 37.8426
Epoch 13/20
388/388 [=====] - 0s 199us/step - loss: 26.0732 - val_loss: 15.2113
Epoch 14/20
388/388 [=====] - 0s 196us/step - loss: 10.9920 - val_loss: 6.1404
Epoch 15/20
388/388 [=====] - 0s 196us/step - loss: 5.4964 - val_loss: 3.3483
Epoch 16/20
388/388 [=====] - 0s 194us/step - loss: 4.0103 - val_loss: 2.7054
Epoch 17/20
388/388 [=====] - 0s 194us/step - loss: 3.6507 - val_loss: 2.5092
Epoch 18/20
388/388 [=====] - 0s 197us/step - loss: 3.5136 - val_loss: 2.3650
Epoch 19/20
388/388 [=====] - 0s 196us/step - loss: 3.3695 - val_loss: 2.2190
Epoch 20/20
388/388 [=====] - 0s 198us/step - loss: 3.2358 - val_loss: 2.0891
```

