Johb Fritz Ekollo

DATA-DRIVEN APPROACH USING MONTE CARLO METHOD AND ANALYTICAL HIERARCHY PROCESS FOR DECISION-MAKING

School of Technology and Innovations
Master's thesis in Industrial Systems Analytics Programme

Vaasa 2022

UNIVERSITY OF VAASA
School of Technology and Innovations
Author: Johb Fritz Ekollo
Title of the Thesis: Data-Driven Approach using Monte Carlo Method and Analytical Hierarchy Process for Decision-Making.
Degree: Master of Science in Industrial Systems Analytics
Programme: Industrial Systems Analytics
Supervisor: Professor Tommi Sottinen
Year: 2022
Pages: 129

ABSTRACT :

Information is a fundamental factor affecting the efficiency of decision-makers, explicitly the capacity of decision-making. Outcomes require to be analyzed and controlled from the datasets for deriving the precise value of the expertise. Hence, decision-makers are required to gain useful perspectives on the use of outcomes, which constantly evolving of high volume, velocity, variety, and value using data-driven approach analytics.

Decision support systems (DSS) supply information as analytic results to decision-makers, which vary during their proceedings period. Thereby, decision support systems are limited in terms of the capacity to provide useful information based on the data sources available.

The thesis aims to perform and improve different statistical approaches (Monte Carlo, and decision tree) and analytical hierarchy process (AHP) in the final decision-making by avoiding several biases during the contextual analysis of datasets realized from the decision support system. The technique (DSS) is an analytical tool used to modelling different uncertainties in data analysis and provides the ability for executive managers to access the correct and strategic decision-making from an analytical approach done in real-time. Therefore, the use of Monte Carlo method and analytical hierarchy process in the decision-making process appear as the principal vector capable to control the process and coordinate the entire system constituted to solve the issue.

Finally, a data-driven approach provides a real benefit to project management for correct decisions in terms of time and profit. Likewise, the method provides the correct approach to the implementation of different analytical tools in decision-making.

Furthermore, considerable effort is required for planning, creating, deploying, maintaining, and continuously improving domain-specific big data processes for each company. Therefore, data-driven approaches used in the project management appear objective and rational for several technology possibilities to generate different concept analytics from the artificial intelligence.

KEYWORDS: Data-driven approach, Decision support systems, Statistical analysis, Monte Carlo, Decision-making, Multi-criteria decision-making, Analytical hierarchy process.

Contents

Figures

Tables

Abbreviations

g/kWh	Gram Per KiloWatt Hour
N.m	Newton-Meter
obr/min	Revolution Per Minute
AHP	Analytical Hierarchy Process
CI	Consistency Index
CO	Carbon Monoxide
CO2	Carbon Dioxide
EN	European Committee for Standardization
FN	False Negative Error
FP	False Positive Error
ICE	Internal Combustion Engine
MADM	Multi Attribute Decision-Making
MAE	Mean Absolute Error
MODM	Multi Objective Decision-Making
MSE	Mean Squared Error
NOx	Nitrogen Oxide
OLS	Ordinary Least Squares
PC	Principal Component
R25	Rapeseed 25% Diesel 75%
R50	Rapeseed 50% Diesel 50%
R75	Rapeseed 75% Diesel 25%
RI	Random Index
RMSE	Root Mean Squared Error
RPM	Revolution Per Minute / Rotational Speed or Frequency
RSE	Root Square Error
RSS	Residual Sum of Squares
S25	Swine 25% Diesel 75%
S50	Swine 50% Diesel 50%

S75	Swine 75% Diesel 25%
T25	Turkey 25% Diesel 75%
T50	Turkey 50% Diesel 25%
T75	Turkey 75% Diesel 25%
THC	Total HydroCarbons
TN	True Negative Error
TP	True Positive Error
TSS	Total Sum of Squares

1 Introduction

Digital technologies significantly changed the path businesses are developed and run, by necessitating the development of new technologies and a diverse set of applications. The volumes of data are readily accessible, as storage capacity has risen dramatically and data gathering methods have altered. Every second, the insights are produced by new data generated from a variety of sources. The data appear in different forms such as structured, unstructured, and semi-structured. Thus, to extract value from those new data, new techniques for storing and analyzing it are required. Moreover, companies use those different procedures as possible to extract information coming from volumes of data produced. Consequently, companies and consumers have access to further innovations and gadgets, resulting in the creation and collection of more data in many categories. ((Brunswicker, et al., 2015; Mohemad et al., 2010; Berger, 1985; Elgendy & Elragal, 2014).

A particular user today possesses a laptop, tablet, smartphone, and other devices, each of which provides quantities of precious data. Each data holds a varied volume of diversity and important velocity that is challenging to manage with current technologies. Every data entails another form of data-driven rationale suitable for its size, quality, and fast innovation, along with various storage and processing systems. (Elgendy, et al., 2021; Brunswicker, et al., 2015; Chang, et al, 2014).

Moreover, these volumes of data-driven should be appropriately inspected or evaluated, to obtain meaningful and pertinent information. These quantities of data-driven approach should be appropriately examined, for significant and related data to be removed. Thus, with the rising interest in using this approach and getting the benefit of its possibilities, companies are looking for basic results and protocols for data-driven control. (Elgendy & Elragal, 2014; Chang, et al, 2014).

Consequently, the research proposal of this thesis involves "How to combine a data-driven approach from decision support systems using Monte Carlo method and analytical hierarchy process to influence the decision-making procedure?".

The aim of this thesis is to design a model combining statistical analysis and analytical hierarchy processes capable of influencing the decision-making process from data-driven approaches, structured by different conceptual analyses such as the concept of decision trees.

Using this approach, decision-makers should be capable of including statistical skills in the decision-making procedure to reduce uncertainties and, therefore, ensure decision effectiveness. Furthermore, the structure joins distinctive significant parts of data-driven analytics necessarily the data analytics lifecycle as significant support and structure including the expected devices for different planned to various decision-making steps. Both key factors (data-driven approach and decision support systems) possess a large spectrum of research studies related to the data-driven approach and the efficient role they perform in the decision support process. Other elements appeared as significant aspects and efficient approaches for algorithms used by predictive analytics in decision-making. (Alter, 1980; Brunswicker, et al., 2015; Elgendy & Elragal, 2014; Berger, 1985; Chang, et al, 2014).

Statistical analysis is a significant constituent in the analyses of the data-driven by the decision support systems. It appears as the process of collecting, exploring, presenting, and analyzing datasets to identify underlying models and trends. Thus, statistical analysis is a component of business intelligence that entails the gathering and examination of datasets and the reporting of trends. This technique employs datasets to eliminate any bias when evaluating data. Furthermore, this approach is regarded as a scientific mechanism capable of illustrating, informing, and acting on the decision-making formulated. (Berger, 1985; Alter, 1980; Brunswicker, et al., 2015; Durcevic, 2019).

The notions of nonlinear dynamical systems applied in the decision support systems conferred to be valuable for analyzing distinct systems based on the datasets and offered new methods for time series analysis. A realistic pattern of a decision-making occurrence takes into consideration some unpredictabilities possible, from the realities interesting, should be anticipated in advance but will show an inherent variance that should be highlighted by the same model. This concept is often achieved by enabling the model to

be probabilistic and concerns different probability models of natural events from the posterior examination of these models, which requires specific knowledge of fundamental probability theory. (Franck et al., 2001; Elgendy & Elragal, 2014; Mohemad et al., 2010; Singh, 2015).

The Monte Carlo method is a dynamic framework defined with an enormous scope of operations for exceptionally nonlinear characteristics. The main function of the algorithmic system is the significant asset for the distributive operation with a considerable range related to the probability of uniform distribution belonging to the objective of the context initially defined.

In summary, the examination of intelligent decision support systems merging the Monte Carlo method and the analytical hierarchy process presents exceptionally both hypothetical and functional benefits and appears effective based on the analyses of decision support systems shown.

The Monte Carlo method proposes a computing technique that, in contrast to classical programming, requires a comprehensive algorithmic definition based on statistical predictions and prior uncertainties regarding the consequences that influence decision-making. Furthermore, the Monte Carlo method and the analytic hierarchy process provide an inductive method for collecting, storing, and applying to experimental testing. Therefore, the obvious interest in decision support systems appears direct and apparent. Overall, these elements constitute some of the fundamental considerations for improving decision support systems.

As a result, we observe two main types of statistical analysis as follow descriptive or normative and inference-based modelling. For this thesis, we will limit our inspection study to the inference model by using advanced statistical models as assumptions to predict events from the uniform probability distribution, enable to generate specimen data and found from the dataset predictions given of the possible outcomes for the decision-making.

1.1 Background of Study

The rising complexity and ambiguity correlated with the data describe the decision-making circumstances by increasing the capacities of variables linear models associated with the use of advanced statistical analysis requiring relationships between the quantitative models by the decision-makers. (Power, 2014; Oreški & Begičević, 2018).

Big data has emerged as a common intriguing and significant tendency in the previous decade, with enormous potential to transform the approach organizations coordinate data that are valuable to their customers, and future users into effective business concepts. Organizations might waste the capacity to get useful significant insights and data, by handling various analyses to give new benefits and possibilities to collect data. Thus those companies need data to develop new products and ensure customer satisfaction. (Papadimitriou & Yu, 2006; Zhu & Shasha, 2002).

Nowadays high-speed internet and technology progress improve the quality of data to become available, by allowing various new techniques of data gathering, storage abilities, and access to massive volumes of information. Nevertheless, data is required to be minutely collected to obtain relevant datasets, and information storage must be significantly inexpensive. (Elgendy & Elragal, 2014; Power, 2014; Roya et al., 2018).

The decision situation describes the complex uncertainty observed with the increment of datasets during the decision-making process. Likewise, this situation depicts the different abilities of the system capable of enhancing nonlinear connections between multiple variables can be carried. Moreover, decision support systems offer such insights to decision-makers with varying degrees of effectiveness throughout their history of use. Therefore, the availability of datasets shows a significant constraint on what decision support systems could harm the decision-making applied in the real context. (Sauter; 2005; Pick & Weatherholt, 2013; Elgendy, et al., 2021).

Overall, this thesis offers a structure that incorporates the significant elements required to assure the aspect and pertinency of data evaluated by decision support systems while

giving the advantages of experiences produced over time from previous judgments and positive suggestions.

1.2 Statement Problem

Recent technological advances offer various options to support the decision-making method by using decision support systems, which have emerged as one of the most significant management methods in the current modern environment. Thus, one of the most prevalent management systems in the modern environment related to decision support systems uses the data-driven approach entirely based on the datasets, which supports decision-makers by easing different stages of data extraction and ability management. (Alyoubi, 2015; Oreški & Begičević, 2018).

The current development of massive internet datasets generally named the data-driven approach, provides new possibilities to enhance decision-making procedures through the use of advanced decision support systems. Furthermore, decision support systems appear as data systems having a variety of forms and variants depending on their application context and predicted results. Data-driven approach including its emphasis on statistical analysis has been considered further significant in recent years by dint of the increasing involvement in handling large amounts of data. (Elgendy & Elragal, 2014; Kopáčková & Škrobáčková, 2006).

Decision-makers entail a solid knowledge of different factors that influence decision outcomes. Thus, it explores data from a website using keywords and employs a set of handling and offers improved for collection and analysis. This study intends to provide fundamental knowledge of data-driven decision-support systems and to assist the subject of decision-support systems. (Oreški & Begičević, 2018; Pick & Weatherholt, 2013).

As a result, the data-driven approach refers to these complex datasets in terms of size and are difficult to process from further analysis than utilizing the standard data storage and processing approaches.

1.3 Objective of Study

The objective of this thesis is to provide the analysis of a data-driven approach for decision-makers and managers from the implementation of potential algorithms applied in the decision support systems using the fundamentals of statistical analytics, Monte Carlo , and the Analytical Hierarchy Process (AHP). Likewise, this idea annotated intends to assist researchers and concerned who are curious and working on this topic, to examine deeply previous and recent experiments that cover in detail the key features of the data-driven approach.

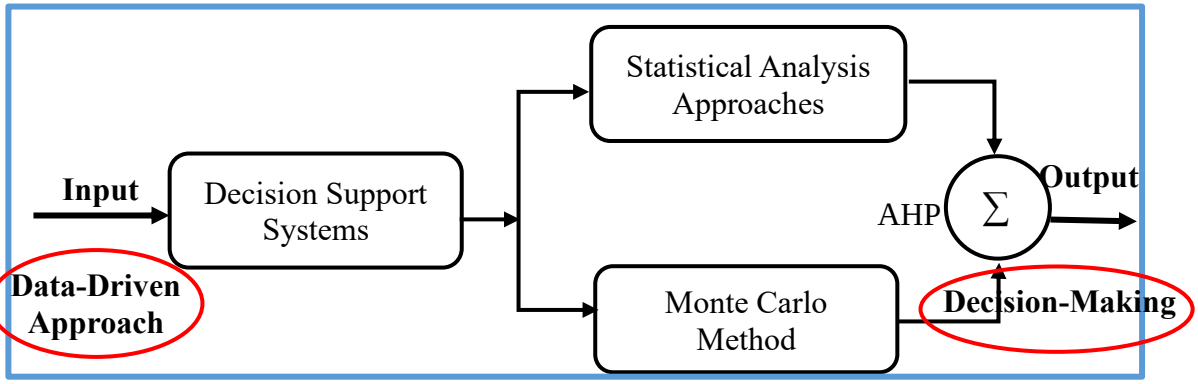


Figure 1. Conceptual Block used from a Data-driven Approach to Decision-making.

The decision-makers currently confront complex circumstances through the decision-making process, including the massive development of energy technologies and increment of information which establishes an intricate business climate wherein common decision-making methods are different to use because inadequate and poorly defined information throughout the decision-making process will result in feeble and ineffective decisions. (Power, 2014; Kopáčková & Škrobáčková, 2006).

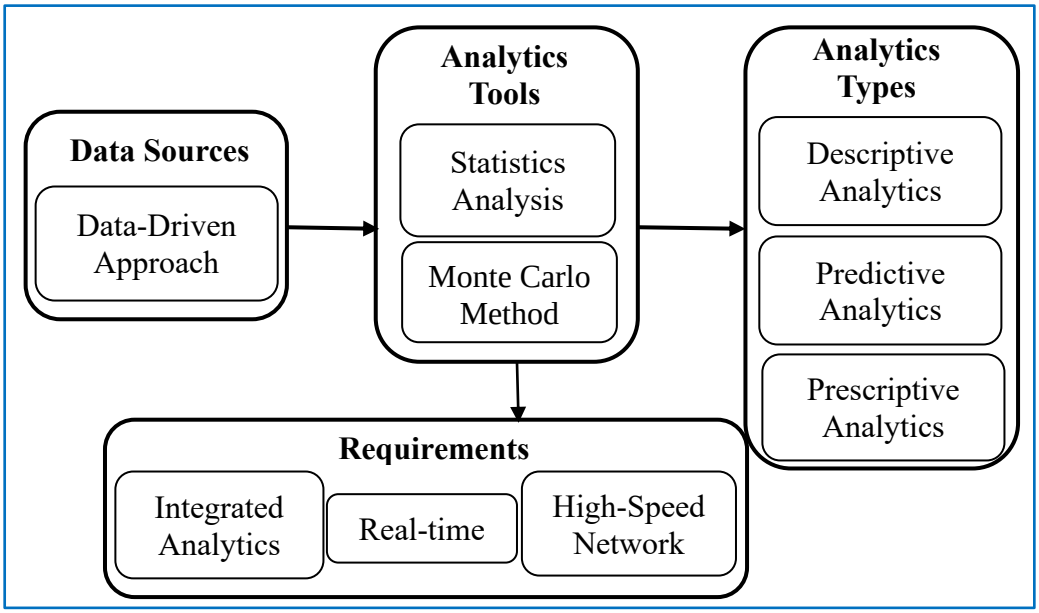


Figure 2. Conceptual Taxonomy of Analytics Methods.

The new decision-making requires enhanced methods of clearness and secrecy. This approach is illustrated by the decision support systems that appeared as human and management frameworks that relate to management knowledge and the conform technologies devices for simulation and datasets technology to maintain the mechanisms of decision-making methods. (Mohemad et al., 2010; Zhu and Zhang, 2009).

However, the method used in this thesis is defined to increase systematically efficiency, performance and reduce costs by enhancing the precision of outcomes to fulfill stakeholder expectations and meet the requirements of the markets to reach project management maturity. Therefore, the outcome described above appears as follows:

- The approach emphasizes various strategies highlighted such that, predictive analytics is used for regression analyses in the examination of information of the datasets, by extracting insights for creating predictive model, anticipating possible market standards, and deciding while support is needed to enhance the company competitivity.
- The technique could significantly reduce fraud and uncertainties by predicting suspect action to enhance sales, and customer satisfaction and choices.

As a result of better analytical methodologies for large datasets, new options have been developed that may improve how decision-makers assess issues by making judgments through data systems. In fact, this thesis examines currently achieved studies and both data-driven approaches and decision support systems to find obvious components and significant variables to the issues highlighted, from the efficient and feasible approaches to improve their combined utilization in decision-making. (Zhu and Zhang, 2009).

By using statistical analytics applied to decision support systems for data-driven approach, this method adduces essential and precious data to the decision-making confronted with hazardous or uncertain options in specific circumstances from the predictive analytics with the identification of datasets examined and from the extraction of information for developing predictive patterns using different regression analyses as the simplified technique. (Mohemad et al., 2010; Singh, 2015; Durcevic, 2019).

This approach implies the potential datasets used as hypotheses of external predictors for enhancing the significant decision-making processing. Likewise, this prism offers a large data analytics structure enabling better decision-making, which will supplement recent data processing approaches. Therefore, the key factor annotating the enhancement related to the advance of speed, and memory volume of computers promotes in real-time the enormously increased size of effective datasets through distributed networks of devices and high-speed telecommunication coming from cloud servers, and other sensors installed in the system.

1.4 Research Questions

By defined decision-making as a data-driven method that implies gathering data based on observable priorities, evaluating trends and evidence from these observations, and using them to improve procedures and actions that support the industry in various ways. The data-driven approach used in the decision support systems of the decision-making aims to increase efficiency and performance and reduce costs by enhancing

efficacy in the precision of project management outcomes. (Singh, 2015; Elgendy & Elragal, 2014; Durcevic, 2019).

This research can be streamlined and viably lead to success (the better decision is fulfilled when a company adduces enough information used for the evolution of datasets, provides data structures to improve decision support for guidance and decision support systems), and proficiently uses predictive analytics, optimization techniques, and analytic hierarchical method used as different key factors for an asset of dynamic methodology ready to impact final decision making. (Pick & Weatherholt, 2013; Malik, 2005; Elgendy & Elragal, 2014; Alter, 1980).

Data collected over a given period is named time series for the regular data used in specific tasks applied in different decisions. From statistical analysis and Monte Carlo method have been used to solve the stochastic process and time series issues from different predictions emitted to enhance the final decision-making. Monte Carlo method emerge as an important time-series optimization method and have a non-linear trend between regression and interpolation. From input and output of continuous function comprise approximated polynomials that fit the model of datasets. (Malik, 2005; Mushtaq, et al., 2017; Berger, 1985; Roya et al., 2018).

Thus, this concept mentioned above follows different steps to reach the final stage by handling different issues such as annotated below:

- How to integrate data-driven analytics into the decision-making process?
- How Monte Carlo method can be used by computation methods to raise the efficiency of decision support systems from the statistical analysis?
- What kind of multi-criteria decision-making (MCDM) methodology indicates different parallel processes that use the Monte Carlo method?

Consequently, the study research describes the understanding of the use of a data-driven approach in decision-making for energy technology projects, by mapping different statistical analysis tools and MCDM methodology used in the decision support

systems and clarifying the data analytics approach applied to various decision-making processes.

1.4.1 Research Hypothesis

The investigation study is handled with the guidance of a hypothesis, which is the approximative solution with the relationship between previous studies and techniques used, and the contribution of the Monte Carlo method. Besides, the use of this type of research offers a more optimal view in the analysis and summarization of data from the quickness of the results in the execution of the information to avoid any bias. (Alyoubi, 2015 ; Franck et al., 2001).

This study appears to underlie the benefit of the statistical analysis combined with the Monte Carlo method for the predictive aspect analysis from the appropriate methodology (Analysis Hierarchy Process), which consists to emphasize minimize costs by better predicting future needs and improving outcomes of the company.

1.5 Assumptions and Limitations

The study of decision support systems merged with statistical analytics and Monte Carlo method to produce a deeply analytical and functional value from the selected MCDM methodology, to enhance all analysis aspects coming from the decision support system for the support and prioritization of company decision-making by increasing visibility of the concept, development and achievement projects from different techniques of mature levels. The limitations are neither to measure the contribution, performance of the use of a data-driven approach from the MCDM methodology for decision-making. However, the scope is to focus on the concept of selected execution tools and methodologies (statistical analysis, Monte Carlo method, and AHP) that support the whole process of performing the prediction of decision-making in the field of energy technology.

1.6 Thesis Structure

However, this procedure is utilized to enhance the early analysis of decision-making from different predictive analysis tools. Thus, the structure of this thesis is organized in the following steps:

In section 1, different structures and meanings are mentioned as background, objective, research questions, hypotheses and limitations are given to provide detailed information.

In the following section 2, data-driven approaches are introduced to describe the review parts of decision support systems, statistical analysis concepts and the Monte Carlo method participate in the analysis study related to choosing the appropriate decision for decision-making from the analysis tools used.

From the previous analysis, section 3 focuses on the methodology used in the action research based on the theoretical background.

Section 4 consists of testing the case study from the statistical concepts defined and the methodology developed in the previous sections to explore the different processes and precise knowledge necessary for the specific use of predictive analysis to solve the different tasks mentioned in each content.

In addition, section 5 entails the results evaluated during the case analyses from different perspectives and achievements obtained. Finally, future work thoughts are stated.

2 Literature Review

2.1 Introduction

This chapter details the overview of data science, data-driven approach, decision support systems, and decision-making by analyzing and outlining theoretically the previous literature reviews from the introduction of different analysis tools used toward managing this thesis favorably.

2.2 Overview of Data Science

Data science is defined as the science of studying data (named data-centric) from different concepts such that data product (a tool that utilizes data to support companies improve decision-making and processes across decisions executed with information obtained from business objectives defined to drive the adoption of the business), which is defined as data deliverable for enabling to discover, predict and suggest information into decision-making from various models comprising paradigms that use different tools and systems. (Cao, 2017; Cervone, 2016; Sarker et al. 2020).

Commonly, data science is considered to be a domain that incorporates characteristics of statistics and computer science. Thus, data science studies the concepts for the implementation of the idea to unify statistics, data analysis, and their associated procedures allowing to generate valuable information and analyze current possibilities from data available. This is a whole of fundamental ideas that support and conduct the deductive extraction of concepts and knowledge from datasets, where a data product is a data deliverable that can be a discovery based on the prediction which is initially proposed the insight into a decision-making thought model. (Provost & Fawcett, 2013; Mushtaq, et al., 2017; Pick & Weatherholt, 2013).

In general, data science is described as a pack of basic principles that facilitates carrying information and understanding from data, where the basic principles of causal analysis

must be known. Fundamentally, data science applies different principles, methods, and approaches for understanding phenomena from the examination of data. Therefore, a large amount of data analysis is traditionally analyzed within the area of statistics as fundamental to data science. At this level, there exists a fundamental structure to data-analytic thinking based on the essential principles of knowledge of the data. (Khan & Ayyoob, 2018; Karpatne et al., 2017).

Overall, data science is a generic term that incorporates different areas such as data analytics, scientific methods, deep learning, and different disciplines related to statistics and mathematical analysis to extract information, and insights from the datasets and transform them into actionable business strategies. Data science involves methods and approaches for understanding different phenomena from the examination of data which are data analysis and data analytics.

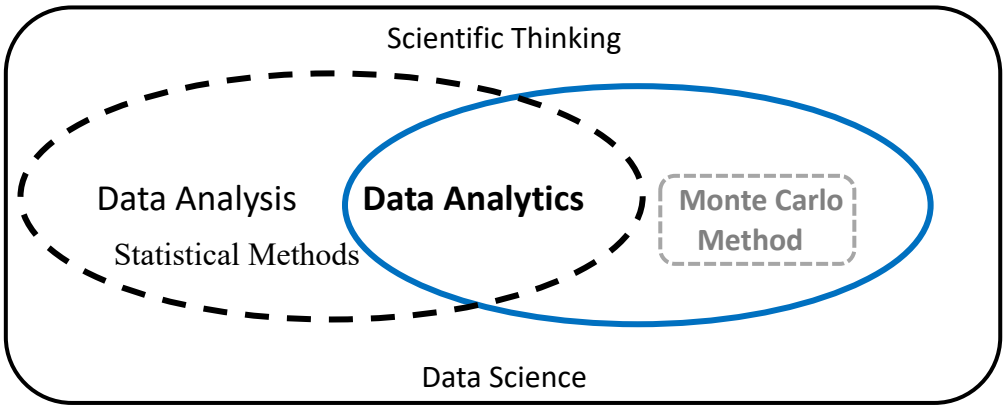


Figure 3. Conceptual Framework of Data Science, Data Analysis, and Data Analytics.

2.2.1 Overview of Data Analysis and Data Analytics

Data analysis is described as the science of analyzing raw data to derive inferences about the data used to enhance decision-making in a variety of businesses. Regardless, data analysis is a long-established subject that has seen incredible growth in many aspects of knowledge emphasizing the necessity for tools that evaluate data and create efficient and appropriate decisions. Thus, data analytics examines a wide range of data to give insights into concepts and ideas through the use of advanced statistics, computational

intelligence, and language processing data. Therefore, data analytics refers to the methods and activities aimed at collecting and analyzing data in the process of extracting useful information. (Sarker et al. 2020; Rizk & Elragal, 2020; Elgendy & Elragal, 2014;Provost & Fawcett, 2013).

On the one hand, data analytics findings are used to determine significant areas of risk, enhance business requirements by evaluating technique efficacy, and impact business choices. Similarly, the applications employed in several areas are characterized by their capacity to convert large volumes of data into insights for operational choices. Data analysis is an ongoing area of study that seems to have a considerable influence on technological and research domains where huge and complicated data sources must be analyzed. Thus, data analytics refers to large data properties such as volume, velocity, variety, and authenticity. Process analytics is concerned with the methodologies required for big data gathering, incorporation, modification, and examination to get insights from large data. (Keim et al., 2008; Brunswicker, et al., 2015; Rizk & Elragal, 2020; Elgendy & Elragal, 2014; Cervone, 2016; Karpatne et al., 2017).

Big data contains a large size and appears heterogeneous and complex from its structure, which improves day-by-day with the progress of technology applied in the use of big data. Thus, data analytics is the use of improved analytical techniques to perform on large amounts of specific information, and improved analysis is an understanding repository comprised of different types of techniques (advanced statistics and mathematics analysis). (Franck et al., 2001; Zhu & Shasha, 2002).

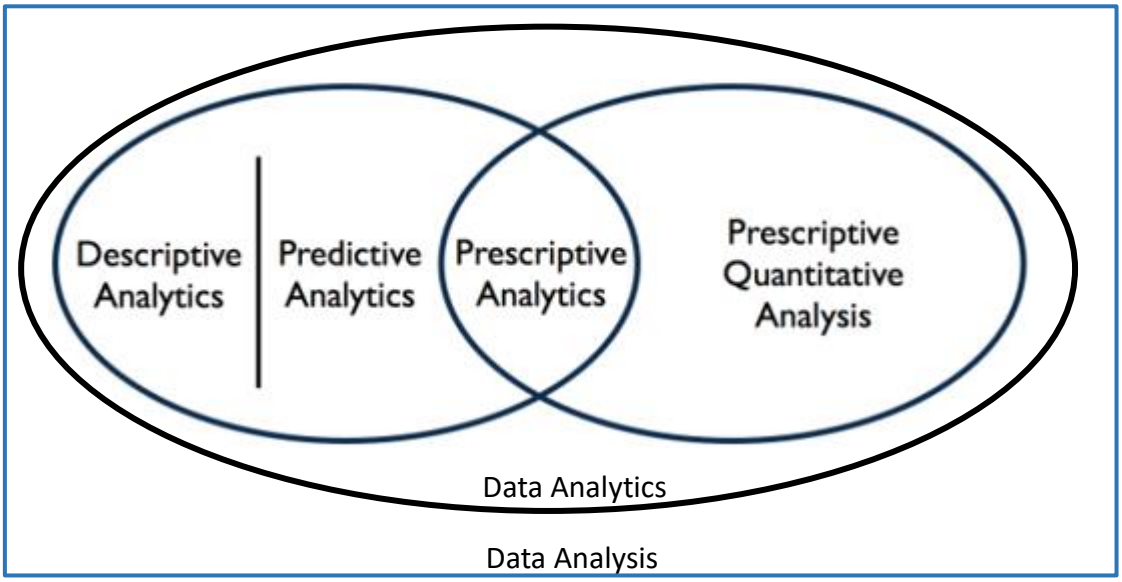


Figure 4. Conceptual Framework of Data Analysis.

In reality, those methods are used to explore and determine together data involved in the original data as a root to extract knowledge while developing analytic models, along with those findings based on **descriptive analytics** (analysis of historical data of the changes that appeared in a company by recapitulating previous data from the statistical models used to sales and procedures strategies related to a past range of appropriate business data), **predictive analytics** (the analytical technique used to identify and analyze data for the previous data to extract insights allowing to create predictive models, which anticipate potential inventory and reduce risks by predicting uncertain activity by making effective decisions), and **prescriptive analytics** involves the collection of information to optimize returns and efficiency, by gathering the amount of data from a variety of descriptive and predictive sources and using it in the decision-making procedure. (Rizk & Elragal, 2020; Iqbal, 2021; Elgendy & Elragal, 2014; Hariri et al., 2019; Han et al., 2011).

The prescriptive analytics focuses on advanced analytics instead of data monitoring, which evaluates decisions and results following an event. Then, prescriptive analytics aids in the integration of a company's data allowing it to make informed judgments and support outcomes for profitable business decisions.

This method emphasizes information to maximize overall returns and profitability, by gathering data from various descriptive and predictive sources and involving it in decision-making).

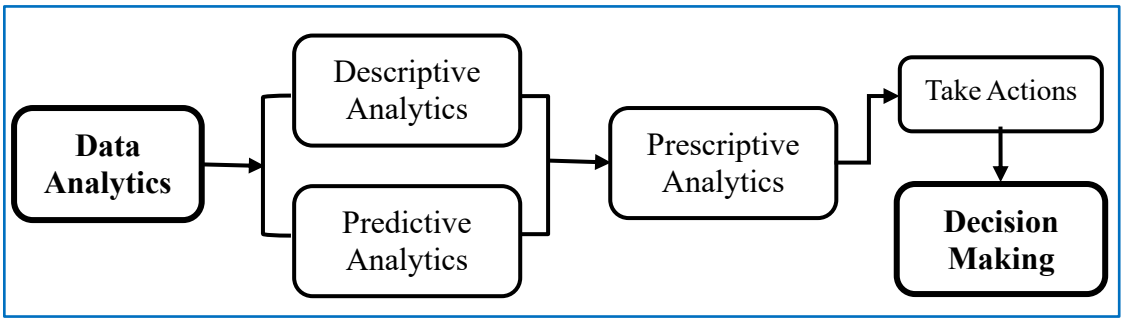


Figure 5. Conceptual Model of Data Analytics.

The conceptual model uses different analyses from data analytics to select, identify, and collect the descriptive data analyzed from the extraction of predictive information for enhancing prescriptive insights generated for decision-making.

As a result, Data analytics is applied to the analytical techniques to the whole of data to enable decision-making and to solve business problems. Consequently, the perspective for using data science in this thesis consists of providing a certain understanding of the traditional decision while using different methods such as advanced statistics and mathematics analysis to improve decision-making in the framework of problems of extracting useful knowledge from data initially treated. (Rizk & Elragal, 2020; Elgendy & Elragal, 2014; Silahtaroğlu & Yılmaztürk 2019).

The data science concept based on big data supports the data-driven approach to improve understanding and identification of business problems and their performance from various processes annotated below. In this thesis stage, we will limit our review study to the data-driven approach obtained from the data analytics concepts.

2.3 Data-Driven Approach

Information is an important part of impacting decision-makers' effectiveness, especially the reliability of their decisions. Nowadays, companies have access to amounts of data for analysis. Data is seen as the sheer material of the twenty-first century, and abundance is carried with today's 15 billion internet-connected items. As a result, solutions must be researched and developed to extract and manage value and information from the databases given. Using data analytics as procedures, concepts, and approaches for analyzing facts through data analysis, decision-makers should be capable of extracting significant insights in the amount of volume comprising the velocity and value from fast-changing data. (Rizk & Elragal, 2020; Fan et al., 2015; Hariri et al., 2019; Elgendy & Elragal, 2014).

The data-driven approach alludes to the process of making decisions obtained from data analysis instead of perception. This concept can solve big data from algorithms related to hardware equipment to act with the volume of data. Hence, the data-driven approach is described as the amount of data that exceeds technology's ability to carry, organize and analyze effectively as the data size grows alongside the evolution of ICT technologies. (Zikopoulos & Eaton, 2011; Kaisler et al., 2012).

A data-driven approach is the continuous collection, analysis, evaluation, and performance of data to the use of analytics or principles and techniques to attain notified decisions. However, decision technology may be strengthened by using big data methods and technology to analyze large integrated information rather than sample sizes to investigate. The data-driven approach is generally based on findings of a fusion of the decision-intuition maker's and experience's data analysis. As a result, the data-driven approach refers to specifying challenges from the issues, then determining strategic goals and success factors to improve and estimate alternatives, and ultimately prioritizing any of these decided options. (Mandinach, 2012; Hariri et al., 2019; Elgendy & Elragal, 2014; Provost & Fawcett, 2013).

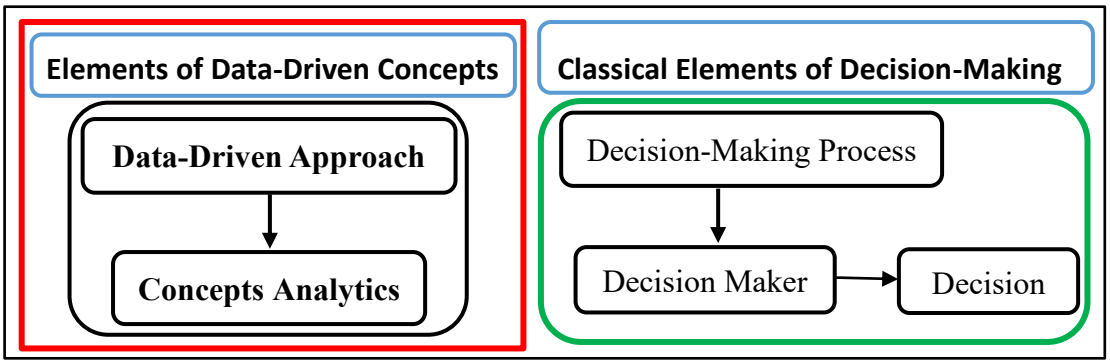


Figure 6. Conceptual Model of Data-driven Approach. (Elgendy et al., 2021).

The data-driven approach is a fundamental stage in data analytics that allows efficient data gathering, integration, and analysis to enhance the reliability, complexity, and accuracy of objective thinking and ultimate decision. Investigating quantities of data can generate descriptive significance of data by summarizing and explaining its recent or ancient activities, from the predictive value emanating from future predictions based on historical data and prescriptive value, which recommend optimal courses of action and describe the impacts. (Cao & Duan, 2015; Hariri et al., 2019; Strand & Syberfeldt, 2020).

Furthermore, the data-driven approach is appropriate to conduct better informed, capacity of judgments since a greater understanding concerning data analytics, linkages between factors, and the resultant insights can contribute to improving decision-making.

A data-driven approach can assist in addressing limited rationality difficulties, which relate to the limitations of the human mind's cognitive skills, as well as a deficiency of accessible data or the incapacity to comprehend enormous volumes of such information required to make an appropriate decision. Although data analytics does not always make critical and strong decisions, rather incremental decisions that organize, categorize, connect, and classify, their result may be utilized as insight for decision-makers can make effective decisions depending on the validity of newer knowledge and correlations. (Cao & Duan, 2015; Diakopoulos, 2016; Elgendy et al., 2021; Power, 2016).

Nowadays, decision theory based on a data-driven approach aims to assist decision-makers in digital settings in making optimal data-driven decisions. The demand for

theory development in research is guided by the necessity in practice. Nonetheless, this method is emphasized from the foundations of traditional decision theories, which are in favour of a data-driven approach, by interrogating and examining their limitations in describing data-driven accomplishments. Furthermore, if decision-makers follow the suggestions and analytics previously mentioned above. Therefore, real-time decision support can assist individuals in making reasonable and probable decisions to occur in goal accomplishment having positive effects on the company's growth. (Gregor, 2006; Grover & Lyytinen, 2015; Elgendy et al., 2021; Strand & Syberfeldt, 2020).

The idea of the data-driven approach is created on classical decision theory, which develops beyond previous assumptions to introduce new solutions capable of supporting future decisions from a scientific study in this field. It is mentioned that the data-driven approach relies on datasets and statistical analytics. Hence, an essential series of questions related to relevant factors (data turning and insight knowledge) raise for the improvements in technology from the evolvement of data analytics, which tends to focus on the new concept for decision-making under various circumstances.

2.3.1 Overview of Decision Theory

Decision theory is the study offering various possibilities and options for making a decision. Thus, decisions appear far from straightforward related to ideas and concepts, that surround complicated areas of study and discussion for decades of multidisciplinary research. The aim of decision theory is always focused on making reasonable decisions. It is a methodical investigation of decision-makers goal-directed, non-random behaviours and activities in situations when several alternatives or methods of procedures are applicable. (Peterson, 2011; Elgendy et al., 2021; Hansson, 1994).

Decision theory focuses on the outcome of decisions as judged by pre-determined standards on the means-ends principle. However, decision theories are generally defined as normative or descriptive. Normative decision theory aims to provide prescriptions for what decision-makers should accomplish logically. Moreover, a normative decision

theory is an approach concerning how judgments should be performed or the circumstances that should exist to achieve reasonable decision-making. Therefore, the choice issue is a circumstance in which a decision-maker must select between a collection of feasible actions that are influenced by circumstances beyond the decision-maker's control and result in a range of effects with positive or negative profits. (Kalantari, 2010; Gigerenzer & Gaissmaier, 2015; Peterson, 2011; Simon, 1959; Frantz & Simon, 2003).

Recent improvements in instinctive decisions and options determine different cognitive methods needed between models and explanations known as the normative and rational formals. The descriptive decision approach is an empirical domain that strives to describe and indicate how individuals make decisions. Then, this approach considers decisions in reality as emotional or rational, in which descriptive and normative decision theories are distinct. Therefore, the mechanisms of traditional decision theory are insufficient to help automate decision-making on complex and practical problems with unpredictable choices or decision alternatives, where the revealing hypothesis can be developed. Qualitative decision theory aims to support automation by improving qualitative procedures that complement and enhance quantitative decision-making capabilities to solve decision-making tasks. (Simon, 1959; Elgendy et al., 2021; Graboś, 2004; Kahneman, 2003; Gigerenzer & Gaissmaier, 2015; Doyle & Thomason, 1999).

The scope of the decision theory applied in this thesis is the interaction between human reflection and data processing investigated and conserved in the memory that is used by the patterns assigned to solve different paradigms of subjects. Then, the use of tools and models is needed to avoid bias and recurrences issued by the assumptions examined in the decision-making. In brief, this part of the study will be more developed in the last section of this chapter named decision-making.

2.3.2 Data-Driven Approach used in Decision Theory

The use of decision theory leads to a thinking process that derives from human decisions, and different evaluations of possibilities given the available facts in real-time. The result of the decision process is defined as a selection of actions aimed at fulfilling a certain target. In time-sensitive contexts, decision processes are put in additiontal strain, as the time available for analysis and selection becomes a key variable. Investigation shows large volume, data-driven approach is an essential part of the fundamental requirement for the creation of decisions in support of judgment for punctual activities. (Holsapple, 2008a; Natter, et al., 2010; Gaynor et al., 2005).

Decision-makers in the twenty-first century frequently access large volumes of widely varied information from considerable sources. Nowadays, challenges arise from data and information overload when decision-makers are at the point of implementing critical decisions. As the scope of this information becomes wider and requires more analysis, decision-makers use a more suitable decision model. Thus, decision-makers are led to use a data-driven approach to collect and analyze more data and information within the most reasonable time, even relatively shorter periods. Then, the time available to make a decision has been shortened in real-time from various tools allowing to optimize the final decision-making. (Hansson, 1994; Kannan, 2013; Singh et al., 2012; Holsapple, 2008b; Gigerenzer & Gaissmaier, 2015; Doyle & Thomason, 1999).

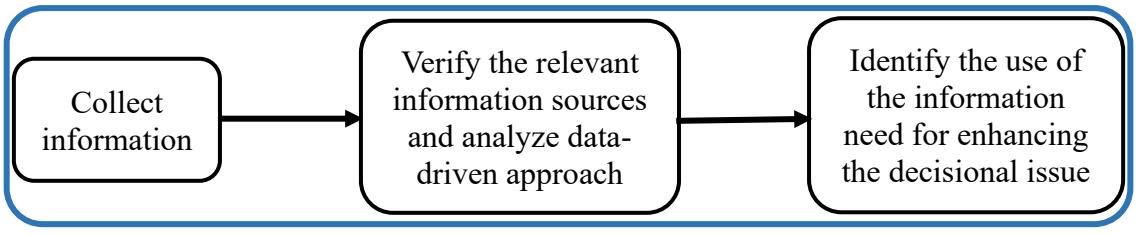


Figure 7. Evaluation of Different Alternatives of Outcomes used in Decision Theory.

The data-driven approach contributes to evaluating the impact of a decision theory on the fundamental issue and its surroundings comprising immediate and continuous advantages or repercussions. Nonetheless, the decision theory is founded on understanding the issue, then an appraisal of different alternatives founded on the

predicted result in handling the issue. Hence, decision theory is critical thinking of the connection of the variables related to the factual choice of the decision process from a data-driven approach. Finally, this method can be prescriptive or unstructured, evaluative or qualitative, and recurrent or periodic. (Baron, 2012; Simon, 1959; Von Krogh & Voelpel, 2006; Heisenberg, 2005).

Decision theory evaluates different alternatives of results coming to the data-driven approach, by using automated decision support systems as a new technique of decision support to collect, analyze, and enhance the information needed by human decisions taken on a company's efficient decision. The support method used for decision improvement from the data-driven approach will be widely developed in the next subtitles **2.4** and **2.7.2**.

2.4 Decision Support Systems

A decision support system is a human-computer interactive decision-making system that helps decision-makers use unsubstituted information and models to find solutions to unstructured and semi-structured issues by focusing on efficient decision-making procedures.

Information systems studies have explored and created decision support systems for over 40 years, implying that the notion of an interactive computer-based system that assists organizations in effective decision-making has existed since computers were widely utilized. Decision support systems are a specific subset of computerized information systems that support business and organizational decision-making activities. A developed decision support system is a dynamic software-based system and subsystems that assist decision-makers in gathering relevant information from data and personal expertise to identify and solve issues and make final judgments. (Power, 2002; Singh et al., 2012; Pick & Weatherholt 2013; Mitchell, 2013; Sharda, 1988).

The decision support system describes a data system that adduces analytical modelling and data. Thus, this description is an incorporated whole of computer tools connected, giving a decision-maker to communicate directly from a computer to acquire relevant information that can assist semi-structured and unstructured organizational decision-making. The decision support system is the capacity to support practically the organization in the decision-making process by including capabilities developed in the computerized field. In general, this technique refers to a type of computer-based information system that contains data to support and facilitate- based innovations for the decision-maker in selecting appropriate options to solve an issue. (Singh et al., 2012; Black & Stockton, 2009; Neiger & Churilov, 2008; Wang & Kumar, 2005; Sprague & Carlson, 1982).

From this postulate, a decision support system appears as a type of data system that assists in decision-making and issue resolution in difficult circumstances. The applications of the decision support system are intended to the decision process to support the decision-maker to achieve appropriate decisions, by using a comprehensible frontend to identify suitable data for collection and combination of facts coming from different sources to enhance the reliability of judgments. (Black & Stockton, 2009; Shim et al., 2002; Cummings, 2004a).

A decision support system provides data digitally that contains an intelligent framework or cognitive computing. Normal data that a decision support application collects and introduces for getting access to all data resources, including inheritance and social information sources, relative information figures, projected figures based on new information or suppositions, and the consequences of various judgment options, given previous involvement with a particular situation. Therefore, the decision support system is developed to intend and support Managers, staff, and manufacturers in making decisions by providing access to data and investigation instruments for modelling data and making quality judgments based on defined objectives. An active technique related to a decision support system handles the data-driven approach and openly demonstrates major solutions to particular issues posed. (Singh et al., 2012; Frey, 2006;

Guitouni, 2009; Gigerenzer & Gaissmaier, 2015; Doyle & Thomason, 1999; Burstein & Holsapple, 2008; Shim, et al., 2002).

DSS Types	User Groups		Purpose	Enabling Technology
	Internal	External		
Data-Driven Approach	Managers, Staff and Manufacturers	Customers and Suppliers	Decision Analysis	Analytical Tools

Table 1. Conceptual Taxonomy of the Decision Support System. (Power, 2002).

The data-driven approach emphasizes gathering data that is then analysed to require the decision-maker's demands. The data can be internal, external, and in several formats coming from user groups. Thus, the use of a data-driven approach to a type of data such as spreadsheets and database records is used to make judgments and change the data to optimize strategies. The technique uses computer storage and processing technologies that provide data retrieval and investigation. Decision support systems provide specific technical needs of hardware, software, and methods to assist a whole of decision-makers in their decision-making. This method focuses on handling the internal and external corporate data in time series and real-time operation. (Holsapple & Whinston, 1996; Neiger & Churilov, 2008; Pick, 2008; Kasunic & Anderson, 2004).

The factors required to accomplish this technique are the consistency and transmission of information. Thus, the structure of the organization is given by each decision-maker, as well as a whole of decision-aiding models that use the information adapted for the decisions constructed by each decision-maker. Therefore, the internet and its computer-based applications collaborate with leaders of different social layers to provide data and information to aid in their decision-making process. Structural systems expand the abilities of the internet to increase its influence by promoting effective and participative decision-making. (Black & Stockton, 2009; Neiger & Churilov, 2008; Carlsson & El Sawy, 2008; Shim, et al., 2002).

Information systems that enhance decision-making can be defined from different viewpoints. A data-driven information system supports high-level executives in making educated judgments that are not defined ahead of time or evolve constantly. In addition, a data-driven approach is typically promoted and integrated with certain information systems. Thus, the data-driven approach relies on procedural decision support systems and management information systems to gather the data required to carry out semi-structured or unstructured judgments, although the time data-driven can be a basis for higher-managers support structures. (Black & Stockton, 2009; Wang, et al., 2008; Morris et al., 2004).

The climb of areas using amounts of data has introduced a variety of possibilities for enhancing decision-making with the use of data-driven approaches that are sustained by findings from real-time information. The data-driven approach often concentrates on acquiring access to different datasets internal and external and adjusting various methods to examine the data to recent tendencies and assumptions appropriated to problems confronted by decision-makers. Data-driven approaches are data systems through amounts of data that permit decision-makers to retrieve pertinent information. Consequently, systems collected and stored ancient information in data warehouses which were subsequently examined from data analysis methods. (Kasunic & Anderson, 2004; Wang, et al., 2008; Pick, 2008; Morris, et al., 2004; Singh et al., 2012).

The focus of the implementation of the decision support system is to identify and analyze individual effectiveness, expand problem-solving by enhancing social communication skills, improve organizational control that generates new evidence with the support decisional and encourage exploration of decision-maker that reveals new approaches to decisional thinking related to external factors and reflecting the evolution decision and organizational of the company.

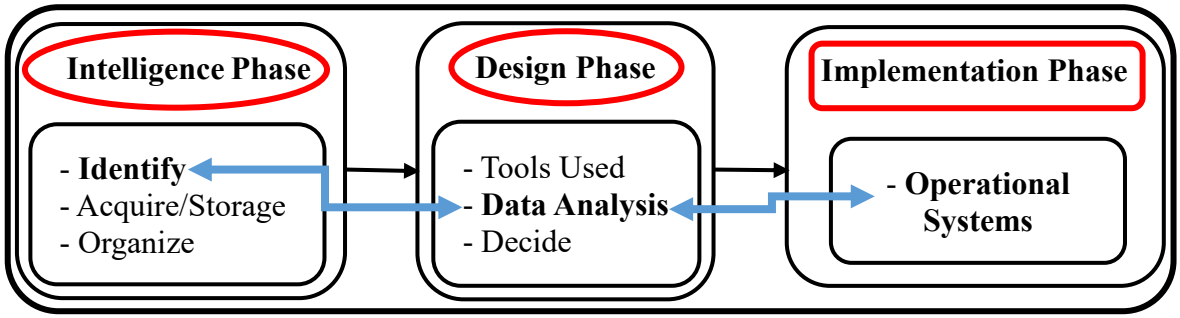


Table 2. Conceptual Taxonomy Development of Decision Support Systems.

2.4.1 The Framework Development of Decision Support Systems

The framework development of a decision support system is a conceptual framework used to construct, evaluate, and demonstrate the procedure used to manage decision-making based on data analytics. The technique is conceived to identify and analyze the problem and define the goals of a solution based on the modification integrated of the experiments utilizing the appropriate information for the needs of the company. Therefore, the approach is intended to connect data analysis tools, architectures, and insights built at various stages called phases (intelligence phase, design phase, and implementation phase) of the decision-making procedure. (Peffers, et al., 2008; Fayyad et al., 1996; Gigerenzer & Gaissmaier, 2015; Doyle & Thomason, 1999; Fisher et al., 2012).

2.4.2 Intelligence and Design Phases

The demonstration is provided in the type of real-time experimental results to give a useful context for framework analysis, which is performed by observing data analytics across the decision-making process. The data sources require to be discovered and collected from different datasets handled and transferred by the end-user.

Intelligence Phase		
Identify	**Acquire /Storage**	**Organize**
-Transactional/Operational Data - Relational Data - Text - Sensor Data	**High-Speed Network**	
	Storage Infrastructure	In-memory Data Management

Table 3. Conceptual Taxonomy of Intelligence Phase. (Elgendy & Elragal, 2016).

The intelligence phase collects data from internal and external sources to identify issues and opportunities. However, this phase is used to assess the experimental environment, which is the connection between identifying data, acquiring and organizing data, and extracting significant information by using the high-speed network. Therefore, this phase consists of testing the perspective and visualization from the storage infrastructure and data management, which are performed by identifying different data-driven approaches. (Aouadni & Rebai, 2017; Elgendy & Elragal, 2016; Shim, et al., 2002).

Design Phase		
Tools Used	**Data Analytics**	**Decide**
Advanced Analytics • Classification • Regression Monte Carlo method	- Descriptive Analytics - Predictive Analytics - Prescriptive Analytics	- Take Actions - Decision Making

Table 4. Conceptual Taxonomy of Design Phase. (Elgendy & Elragal, 2016).

The design framework is used to conceive the process for implementing a data-driven approach from the decision support systems. Thus, this stage specifies the model used, data analysis procedures and decisions, that combine advanced statistics and Monte Carlo method for each step of taking the decision. Descriptive analytics, predictive

analytics and prescriptive analytics are used in statistical analysis to provide value and insight from an integrated technique to assess and evaluate data analytics related to the decision domain. The model used is to evaluate and perform the findings to help and get insights in making the essential decision. (Wang, et al., 2008; Holsapple & Whinston, 1996; Fisher et al., 2012).

2.4.3 Implementation Phase

The implementation phase covers different solutions from operational systems ranging from the data-driven approach as input to the decision support system as output, which is examined together with statistical analysis and Monte Carlo method to provide common decision-making. However, statistical analytics use the decision support system as a key technical system that treats the data-driven approach by involving issues related to the effectiveness and ability to control the data sources from the standard method applied to data analysis. Therefore, the decision support system appears as instrumental in supporting decision-makers in handling, analyzing, and managing structured, semi-structured, and unstructured decision contexts to help decision-makers by improving the efficiency of the company decision-making from different recurrent problems. (Aouadni & Rebai, 2017; Elgendy & Elragal, 2016; Shim, et al., 2002; Fisher et al., 2012).

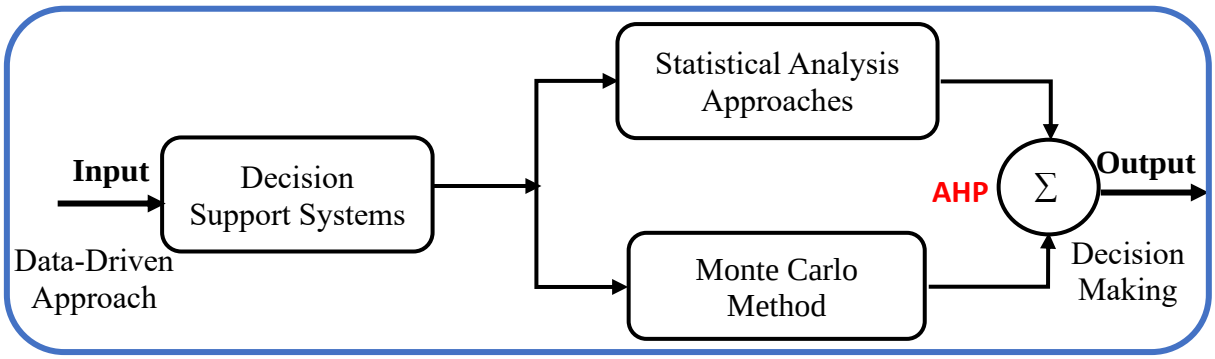


Table 5. Conceptual Block of Implementation Phase. (Elgendy & Elragal, 2016).

Finally, decision-makers take into account the possible advantages of using big data analytics to concentrate, categorize, and evaluate prospective opportunities and

valuable insights connected to a company. Advanced analytical tools and approaches used for the implementation phase after the achievement of decision support systems will be widely developed in the next title.

2.5 Statistical Analysis

Statistical analysis is the process of gathering and analyzing data to detect patterns and insights in structured, unstructured, and semi-structured information that may be used for research interpretations from the data techniques given for the creation of surveys and reports. The use of statistical information containing variables, entities, and events, enables the study of complicated analyses that incorporate advanced options from the structure, interpretation, and presentation of information. This technique represents the entire analytical process, which comprises the data preparation and execution from the analysis and testing data. (Vapnik, 1995; Bishop, 2006; Hastie et al., 2001; Taiwo, 2010).

Statistical analysis is defined as a technological instrument that uses datasets to handle decision-making from a range of statistical topics that cover analytical thinking, cluster analysis, statistical inference, prediction, fault diagnosis, and error analysis. The statistical concept introduces a theoretical estimate based on the evaluation of different variables complex that allow a contextual analysis from various measurements and solutions obtained of each specimen data. The approach evaluates each specific data specimen in a collection that depicts data in a cross-sectional manner from a cross-sectional illustration, which simplifies less sophisticated methods used for the assumptions given to assess initial data. Therefore, statistical analyses use numbers to reduce bias and provide the light with the most possible to the final results obtained during the analytical process. (Bishop, 2006; Mitchell, 2006; Hastie et al., 2001; Witten et al., 2011; Elmusrati, 2020).

Statistical analysis is a method constituted for analyzing connections between various attributes and datasets implicating the dependent variable from the events of interest and the impacts of independent variables mentioned as high uncertainty for predictors

or regression analysis destined for evaluation of prediction results as illustrated in Figure 8 below.

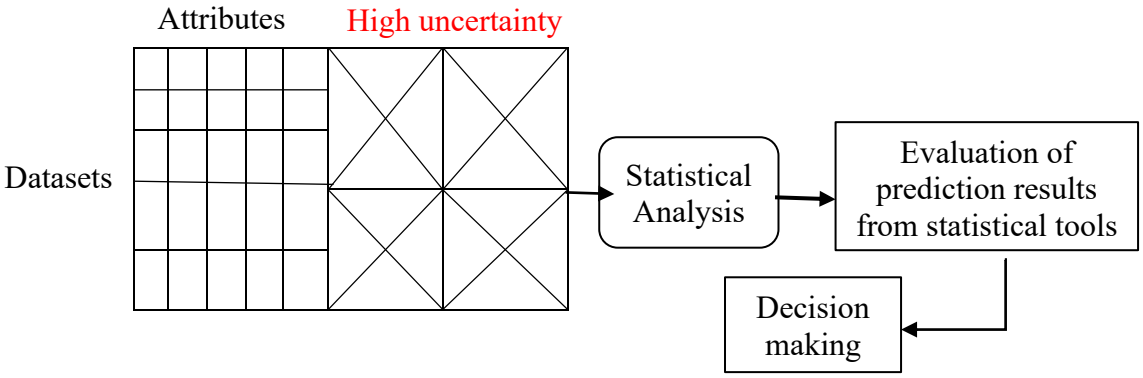


Figure 8. Schematic Summary of Statistical Analysis Methods used for Decision-making (Elmusrati, 2020).

Statistical analysis is an advanced method for analyzing connections between input variables denoted by the symbol *X* implicating the impacts of independent variables, features and variables as predictors or regression analysis and output variables denoted by the symbol *Y* having the response and dependent variables from the events of interest. Assume that we examine a quantitative response *Y* and *p* with various predictors X_1, X_2,…., X_p. We suppose that there are some connections between *Y* and $X = (X_1, X_2,\ldots,X_p)$, which having for the general form:

$$Y = f(X) + \varepsilon. \quad (1)$$

Where f is a fixed but unknown function of $X_1,\ldots,X_p$, and ε is a random error term, which is independent of *X* and has a mean of zero. As result, *h* portrays the systematic data that *X* adduces concerning *Y*.

The technique alludes to a whole of procedures for estimating *h* that enable extracting conceivable quantities of data that are adequate to comprehend the connections among the variables and parameters held in the information (descriptive analysis) to afford precious appraisals and predictions or estimations of the results (predictive analysis) along with sensible confidence. The schematic described in the figure as illustrated in

Table 4 is divided into various statistical analysis tools and represents different statistical analysis methods. (Bishop, 2006; Kourou et al., 2015; Elmusrati, 2020; Hastie et al., 2001).

Statistical analysis alludes to a whole of procedures for evaluating *f* as key theoretical ideas that appear in estimating *f* as evaluation tools used for the given assessments. Assume a set of inputs *X* and the output *Y*. The error term averages to zero are defined from the prediction *Y* such that:

$$\hat{Y} = \hat{f}(X)\,. \tag{2}$$

where $\hat{f}$ is an estimation for *f*, and $\hat{Y}$ is a resulting prediction for *Y*.

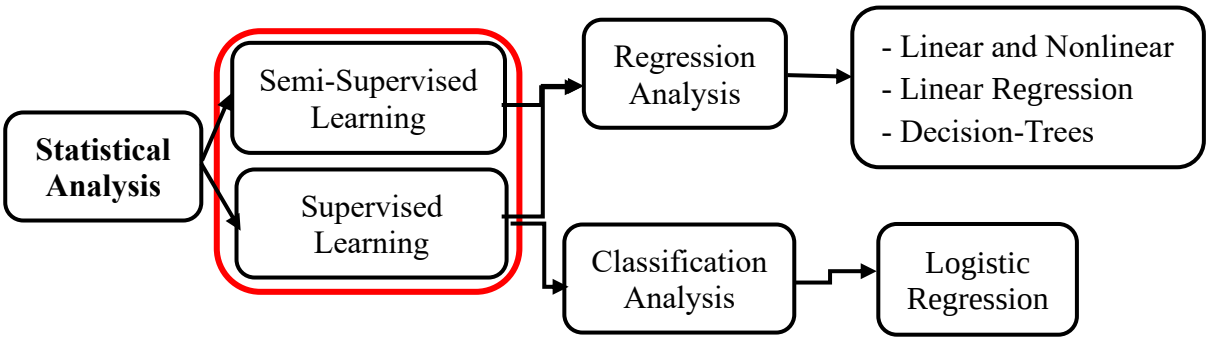


Figure 9. Schematic of Statistical Analysis Methods used for Decision-making (Elmusrati, 2020).

As shown in **Figure 9**, statistical analysis is a vast field of mathematics, for this purpose we limit ourselves to the use of appropriate formulas and theorems falling only to the content of this thesis.

2.5.1 Overview of Supervised Learning and Unsupervised Learning

A supervised learning method is defined from the training data utilized to represent the input data and associated known variables to the expected output, which is the quest for suitable algorithms to perform common hypotheses from outward provided instances involving predictions to expected instances. (Hastie et al., 2001; Kotsiantis, 2007).

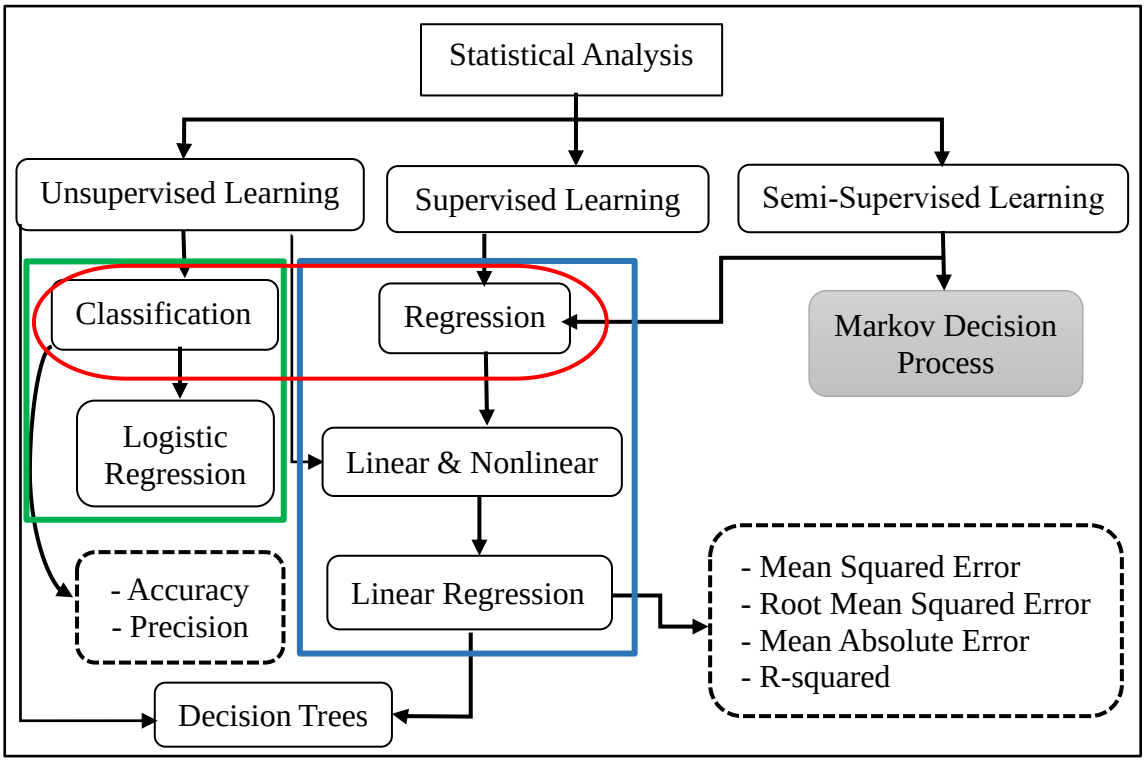


Figure 10. Useable Algorithms in Statistical Analysis Methods.

Supervised learning aims to create a suitable system that examines the interconnection between the input and the output by anticipating the output of the specific system applied to the novel inputs. Thus, the acquired mapping carries to the category of the input data by taking a limited collection of disjunct values that portray the tags of the input. In contrast, the output accepts continuous values by conducting input from a regression that illustrates the connection between input and output often provided by learned-model specifications. (Kotsiantis, 2007; Kourou et al., 2015; Elmusrati, 2020 Shalev-Shwartz & Ben-David, 2014).

A learning system requires an estimation procedure to acquire the specifications that are not straight functional or effective from training samples. For illustration, the algorithms used by the training data for supervised learning involve supervised data for the input data across the training labeled information from various forms of unsupervised learning. (Kourou et al., 2015; Elmusrati, 2020; Hastie et al., 2001).

In this case, the output is generated following the comprehensive and appropriate training of the input training data the predicted output whereas the desired or originally known output comes from the initial information named the target or intended output.

The prediction error is the variance between the desired output and the predicted output which generally notifies the choice to adjust the pattern to improve effectiveness and precision. For illustration, supervised learning refers to a learning pattern for obtaining the connection data used in a system from a specific whole of paired input and output specimens. (Haykin, 1998; Kotsiantis, 2007; Shalev-Shwartz & Ben-David, 2014).

Therefore, the supervised learning algorithms that treat further details of supervised classification comprise the junction of regression analysis and classification analysis. The input information is the tag of the output considered as supervision, while the training specimen of input-output is targeted training data. (Haykin, 1998; Hastie et al., 2001).

Conversely, the unsupervised learning method is a methodology applied to the Monte Carlo method to which the input data used in the learning phase are not labelable, nor classifiable, and finally have no notion of the output in the learning phase. The concept is to classify or collate the input information into clusters of common predicates. However, the model does not determine an output but rather examines training data related to connections and structured clusters from the patterns focused on the common attributes. (Kourou et al., 2015; Elmusrati, 2020; Haykin, 1998; Kotsiantis, 2007; Bishop, 2006).

2.5.2 Classification and Regression

Classification assignments focus to categorize or classify data into a finite set of classes. Input variables are classified using output variables within one of the potential output classes. Then, a linear classifier can be highly difficult to determine classes from a set of data. Likewise, a big challenge is finding a precise boundary between classifiers that are expected to generate noise, avoid corruption, and minimize bias in the data. (Ayer et al., 2010; W. Kim et al., 2012; Kourou et al., 2015).

The regression task aims to examine observed input-output relationships to obtain a precise model. The training data are utilized to improve the prediction model. The learned model can be used following the training procedure to evaluate data that does not appear in training and external datasets by interpolation and extrapolation respectively. (Kotsiantis, 2007; W. Kim et al., 2012; Kourou et al., 2015).

This process provides the real efficiency and adaptability required to adjust or adapt the model.

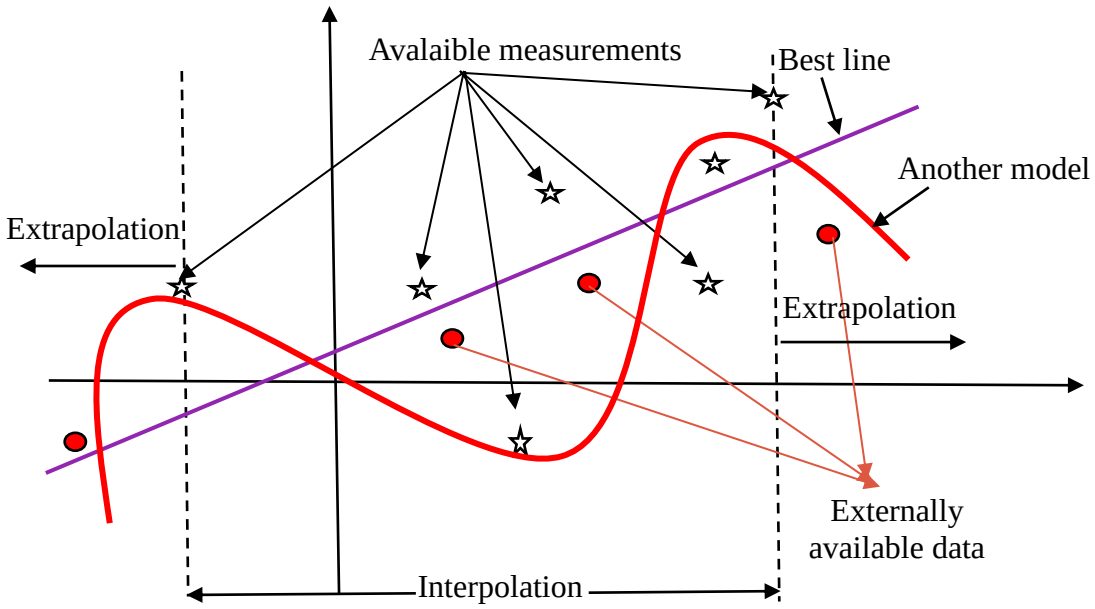


Figure 11. The concept of Interpolation and Extrapolation in Regression Algorithms (Elmusrati, 2020).

The regression model divides the available information into training and testing sets, which are then evaluated. Additionally, the prediction of real-valued variables can be referred to a regression problem between interpolation and extrapolation. The input data is classed through the identification of hidden patterns. However, this approach is similar to the fact that the classification method follows a non-existent target output called a label through which the model can learn independently of unsupervised and supervised learning (regression & classification). (Bartholomai & Frieboes, 2018; Bishop, 2006; Kourou et al., 2015).

A predictive model is developed in the process of classifying data into predefined classes in which two errors called training and generalization errors can occur. Therefore, the misclassification of test data is referred to generalization error compared to training error attributed to misclassification of training data. (Elmusrati, 2020; Kotsiantis, 2007; Bishop, 2006).

Nowadays, the development of energy technologies uses mathematical tools including analytics and algorithms to evaluate, analyze, and perform critical tasks to determine the effective solution.

2.5.3 Logistic Regression for making Predictions

Let's consider the problem of predicting a response using different predictors. By referring to the simple linear regression equation as follows:

$$Y = \ log\left(\frac{p(X)}{1-p(X)}\right) + \varepsilon \ = \beta_0 + \sum_{p=1}^{k} \beta_p X_p + \varepsilon, k \in \mathbb{N} \quad (3)$$

Where $X = \left(X_1, \ldots\ldots, X_p\right)$ are $\mathcal{P}$ predictors of the linear regression using the coefficients β_p given to compute the default probability from the accuracy measurement estimated by the standard error ε. Using the integration of both sides of the simplified equation (3):

$log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$,we obtain the simplified exponential form of equation:

$$\int log\left(\frac{p(X)}{1-p(X)}\right) = \int \beta_0 + \beta_1 X \implies \frac{p(X)}{1-p(X)} = e^{\beta_0 + \beta_1 X} \quad (4)$$

Then, the logistic regression can be expressed as the probability of default from the equation (4) as follows:

$$\frac{p(X)}{1-p(X)} = e^{\beta_0 + \sum_{p=1}^{k} \beta_p X_p} \implies \mathcal{P}(X) \ = \frac{e^{\beta_0 + \sum_{p=1}^{k} \beta_p X_p}}{1 + e^{\beta_0 + \sum_{p=1}^{k} \beta_p X_p}}, p \in \mathbb{N} \quad (5)$$

Where $\beta_0, \ldots\ldots, \beta_p$ are the unknown regression coefficients and $\mathcal{P}(X)$ is the predicted probability model of default.

2.5.4 Linear and NonLinear Model

A simple linear model is defined by the relationship between *X* and *Y* expressed in the form $Y = f(X) + \varepsilon$. However, we assume that *f* = *Y* is approximated a linear function resulting from the previous relation defined as follows $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ and generally reformulated:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon, k \in \mathbb{N} \quad (6)$$

Where *f* is an unknown function, ε is a mean-zero and an unobserved random error, Y and X_i are respectively the expected values and β_0, β_1 are the parameters. The coefficients are the intercept (constant) term β_0 and the slope β_1. Therefore, given the unknown coefficients θ_0, and θ_1 are estimated from training data as described in the equation below

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1} x \,, k \in \mathbb{N}. \quad (7)$$

The parameters $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are computed to predict the observed values Y based on given X – value that the prediction is indicated by $\hat{Y}$. Therefore, the sample training data with observations (x_1, y_1),...., (x_n, y_n) is estimated by the coefficients β_0, and β_1 from the ordinary least squares (OLS) which is annotated in the following equation:

$$(\widehat{\beta_0}, \widehat{\beta_1}) = arg\ min_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - (\beta_0 - \beta_1 x_i))^2 \quad (8)$$

With the solutions of $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are exprimated.

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (9)$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x} \quad (10)$$

where, $\bar{x}$ and $\bar{y}$ are the sample means of *x* and *y*.

The prior knowledge of each model is determined by the choice of variables *X*. Preferably, the input variables are causal variables concerning *Y*. However, correlation rather than causality may be necessary to achieve model accuracy.

The correlation approach is a type of rational autoregressive model intended to carry the causal data to be modeled from the previous values of the quantity. In particular, correlation analysis and general multivariate methods in statistical analysis methods can also be used when qualitative knowledge is inaccessible.

Nonlinear structural models provide a less restrictive definition of application to linear dynamic systems compared to linear models.

$$Y = \mathcal{F}(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m) + \varepsilon, k \in \mathbb{N}. \quad (11)$$

Where ε is a random variable called model error independent of expected values $x_1,...,x_m$. The expression $\mathcal{F}$ is a non-linear function called linear operator.
However, the resolution of this issue occurs when the model is implemented at a certain operational point, allowing linearization of the dynamic differential equation. Additionally, the linear structure has the ability to transform inputs, demonstrating the non-linear character of changing inputs rather than determining parameters.

2.5.5 Linear Regression

Given the regression coefficients in a linear regression model by minimizing

$$Y = \beta_0 + \sum_{p=1}^{k} \beta_p X_p + \varepsilon, k \in \mathbb{N}. \quad (12)$$

The linear regression souscribes in different model selection (subset, feature and variable) which refer to the selection of the appropriate subset from the *p* variables that predict and capture variability in *Y*.

$$Y = log\left(\frac{p(X)}{1-p(x)}\right) + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon, p \in \mathbb{N} \quad (13)$$

The validation set approach uses the prediction technique to identify the best predictors by dividing the sample into a test set and a validation set, which estimate the best subsets of explanatory variables from the sizes 1, 2, ..., p used for the estimation dataset to find the MSE (minimum test set).

Generally, the technique is performed from the size *k* of the best set of predictors selected ($k = N$) based on the validation approach, while the final selection of the *k* predictors is performed from the full sample and the estimated regression corresponding. Therefore, predictors in the final model may differ from those of the validation best predictors but maintain entirely their numbers.

The regularization known as shrinkage refers to fitting one of the different models with all *p* variables by shrinking the coefficient estimates towards Zero to decrease variance. Typical approaches use principal component (PC) and partial least squares, which involve projecting the *p* predictors to a lower dimensional space called Dimension reduction.

$$RSS = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2, n \in \mathbb{N} \quad (14)$$

Ordinary Least Squares estimate Regression coefficients by minimizing RSS which depends on Ridge regression estimates $\hat{\beta}_\lambda^R$ are obtained by minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 = RSS + \lambda\sum_{j=1}^{p}\beta_j^2 \quad (15)$$

where $\lambda \geqslant 0$ is a tuning parameter determined by the term $\lambda\sum_{j=1}^{p}\beta_j^2$ called a shrinkage penalty which gets the smaller closer β_j's are to Zero.

Ridge regression seeks optimal fitting finding coefficients for OLS that minimize the RSS from the optimum coefficients shrunken towards Zero which is the penalization of large coefficients taken at the same time. Therefore, the tuning parameter $\lambda \geqslant 0$ controls the relative impact of two terms:

$$\lambda \geqslant 0, \begin{cases} \lambda \geqslant 0 \; leads \; to \; OLS \\ \lambda \to \infty \; drives \; the \; coefficients \; estimates \; towards \; Zero \end{cases}$$

Where, ridge regression over OLS is rooted in the bias-variance trade-off.

The validation set approach uses the prediction technique to identify the best predictors by dividing the sample into a test set and a validation set, which estimate the best subsets of explanatory variables from the sizes *1, 2, ..., p* used for the estimation dataset to find the mean squared error MSE (minimum test set). The choice and selection of suitable analytical model predictors focus on the significant predictors of a particular model, which are classified into two varieties, such as model-free and model-based. We limit our study to model-free techniques to operate input selection without depending on the efficiency of the constructed analytical models.

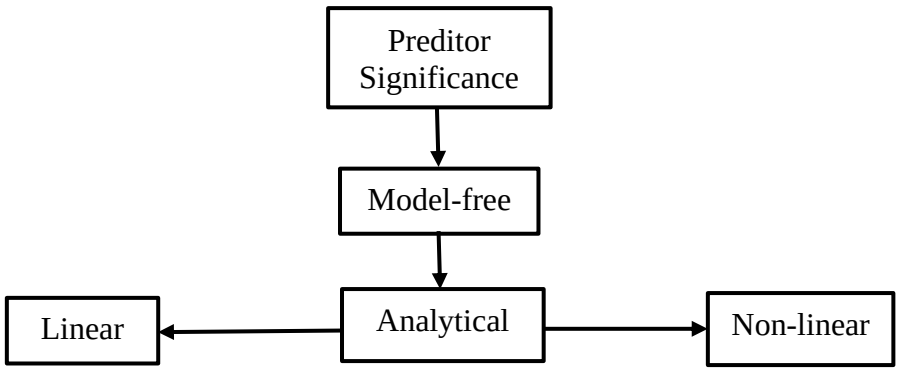


Figure 12. Schematic of Analytical Approach to determining Optimal Predictors Model (Elnusrati, 2020).

The model-free approach is the analytical method where the chosen model predictors are arbitrarily based on deep domain knowledge or expertise. In contrast, the analytical approach to input selection uses a statistical tool of correlation between model

predictors and target variables. (Hagan et al., 1995; Strogatz, 2001; Maier et al., 2010; Samarasinghe, 2006).

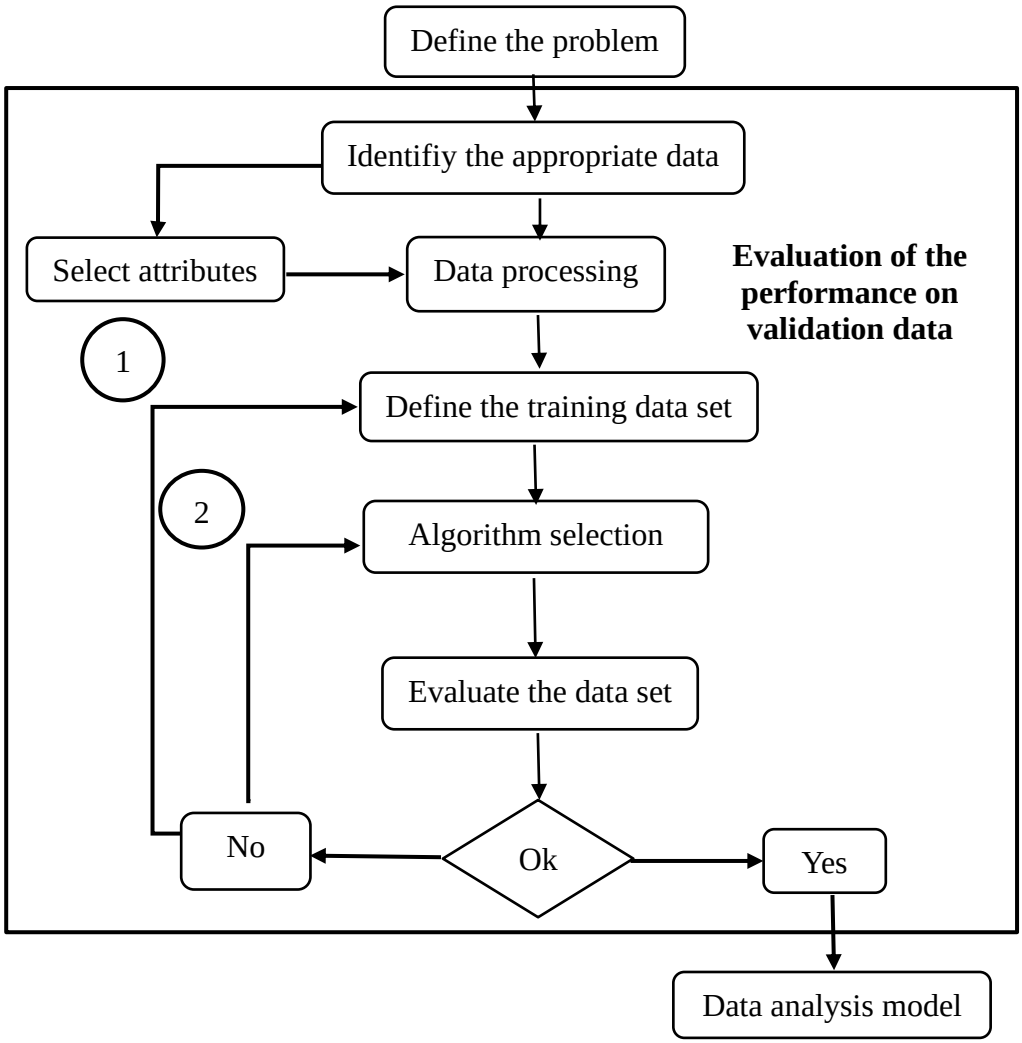


Figure 13. Flow Chart of the Validation Data Performance Evaluation Process.

The analytical approach uncovers linear data results in the removal of useful predictors that are related to the predicted variables in a nonlinear method as shown in Figure 12. The optimal predictors approach is usually accomplished by cross-correlation. (Maier et al., 2010; Samarasinghe, 2006).

We assume the predictor space from the dataset of values are $X_1, X_2, ..., X_p$ into J distinct and $R_1, R_2, ..., R_J$ are non-overlapping regions of the training observations.

For a given observation:

$X = x, \begin{cases} \text{if } x \in R_1 \text{ predicts a value } \alpha \\ \text{if } x \in R_2 \text{ predicts a value } \beta \end{cases}$, and the regions can have any shape.

We divide the predictor space into high-dimensional regions to simplify different cases of interpretation from the resulting predictive model by minimizing the given RSS as follows the equation above:

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2, n \in \mathbb{N} \quad (16)$$

Each observation made from the prediction in the region R_j is the average of the response values for the training observations in R_j.

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon, k \in \mathbb{N}. \quad (17)$$

As illustrated in **Figures 10 and 12**, the notions of statistical analysis used precisely in this chapter show the correlation between the different techniques used for the selection of decision-making and the choice of the methodological approach applied which will take effect in the implementation part as demonstrated in **Chapters 3** and **4**.

Data analysis concepts are used for prediction and anticipation based on a historical and empirical data-driven approach. Thus, quantitative models are used to handle issues arising from the data-driven approach used for decision support systems that conclude parameters such as selection, description, and construction to help a decision maker to make a decision. Therefore, analytical studies information from the data to identify linear or non-linear trends and provide specific predictions. (Henry, 1858; Maier et al., 2010; Strogatz, 2001; Pascual & Dunne, 2006).

2.5.6 Monte Carlo Method

The Monte Carlo method applies to the problem of computing the value required to calculate the weights of the criteria crucial for finding optimum solutions, which

numerical outputs are obtained using a wide variation of algorithms by performing a random selection several times from a specified probability distribution.

The variability of the average values is rather obtained by error each time a Monte Carlo calculation is carried out. Furthermore, the estimated mean differs from Monte Carlo calculations based on the number N of iterations in each Monte Carlo calculation performed.

Let's consider $\mathcal{X}_i, i \in \{1,2,\ldots, N/N \in \mathbb{N}\}$ denote the individual randomly generated values.

$$\sigma_{\mathcal{X}}^2 = \frac{\sigma\overline{\mathcal{X}}^2}{N} \tag{18}$$

Where $\overline{\mathcal{X}}$ represents the mean values of the given sample, N is the sample size of iterations, $\sigma\overline{\mathcal{X}}^2$ is the variance of the given mean and $\sigma_{\mathcal{X}}^2$ represents the variance when the Monte Carlo calculations are performed by N iterations.

However, the previously calculated mean and standard deviation values lead to the calculation of the indicated root mean square error (RSE) value, which appears when simulating N iterations of a sample size in each Monte Carlo computation. Furthermore, the criteria weights randomly belong to the uniform probability distribution.

Let consider $\mathcal{X}_i \in [a, b]$, The standard deviation describes the estimation of the sample size when the error is less than the required threshold for the uniform probability distribution.

$$\sigma_{\mathcal{X}}^2 = \frac{(b-a)^2}{N} \tag{19}$$

Let assume $A = \{A_1, A_2, \ldots, A_k \,/\, k \,\epsilon\, \mathbb{N}\}$ defined as the set of k alternatives under consideration, $C = \{C_1, C_2, \ldots, C_n \,/\, n \,\epsilon\, \mathbb{N}\}$ represents the set of n criteria and $W = \{W_1, W_2, \ldots, W_n \,/\, n \,\epsilon\, \mathbb{N}\}$ is the vector of criteria weights with $W_1 =]a, b[\ for\ b > a > 0$.

The weights of the criteria approaches are determined randomly from interval $[a, b]$ using the Monte Carlo technique. In addition, the evaluation of alternatives is carried out based on the criteria defined by following the dominant interactions between the evaluation and ranking of alternatives.

$$U(A_i) = \sum_{j=1}^{n} f_{ij} w_j \, , n \in \mathbb{N}, f_{ij} \in \mathbb{R} \quad (20)$$

Where f_{ij} represents the evaluation of the $i-th$ alternative under $j-th$ criterion, w_j is the criteria weights and $U(A_i)$ is utility function or cardinal of an alternative i which $F = (f_{ij})$ represents the decision matrix.

Monte Carlo analysis involves generating the appropriate random numbers to estimate the behaviour of random variables based on various probability distributions. In addition, the Monte Carlo method uses uncertainty distribution techniques to verify the accuracy and absolute similarity of performance uncertainty.

The set of datasets is used to introduce the approach of training algorithm from Monte Carlo simulation to examine the nonlinear method by the input-output relationships of the data to confirm the performance of the test results in an efficient method. Then, the values of mean square error and square error issued from the average weight values improve the selection of random data taken from the input-output data relationship of the weighted sum values, while representing the performed value of the potential result.

2.5.7 Decision-Trees Concept

The decision tree is a mathematical model using intuition to interpret data from estimation and probability calculations to produce decision-making conclusions. However, the decision-making process is mapped from the diagram using different options and scenarios followed by potential outcomes. Then, the decision tree structure represents a hierarchical tree constructed during data analysis to predict a continuous results variable using the input characteristics for the regression analysis.

In addition, the decision analysis uses the decision tree model in the decision-making processes to assess the probable effects and impacts of different selections and actions by providing an intuitive understanding based on the factors impacting events and outcomes from various options.

Decision trees are applied to both regression and classification problems described in tree-based methods which involve stratifying or segmentting the predictor space into several simple regions. However, the set of splitting rules used to segment the predictor space can be summarised in a tree, these types of approaches are known as decision-tree methods. Then, tree-based methods are useful for interpretation and applied to the supervised learning approaches in terms of prediction accuracy. (Bartholomai & Frieboes, 2018; Bishop, 2006; Kourou et al., 2015).

The nodes indicate decision points determined by events resulting from selections made to achieve outcomes. Then, the root node denotes the original decision taken at the starting point of the tree connected to lower-level nodes indicating decisions or events resulting from additional branches.

Furthermore, the terminal nodes reflect the endpoints of the tree through their final decisions or outcomes. The branches connected to the nodes are options or potential outcomes. In brief, each branch is a decision impacted by an event that contributes to various directions in the decision-making process.

These methods grow multiple trees by combining to yield a single consensus prediction. However, the large number of trees often combine resulting in improvements in prediction accuracy for loss interpretation. Besides, the regions of the tree analogy are known as terminal nodes and the points along the tree are referred to internal nodes where the predictor is split space. Then, the structural complexity is intrinsically challenging to interpret due to different non-linear interactions that are non-randomly distributed and modify dynamically in response to the original system. (Strogatz, 2001; Pascual & Dunne, 2006; Bartholomai & Frieboes, 2018; Elnusrati, 2020).

Decision tree principles are used to select the next path to follow at each point of decision, which illustrates the conclusions of the decision-making process indicating the outcome. Each result can carry implied probabilities or likelihoods corresponding to evaluations.

2.6 Decision-Making

Decision theory is a distinct area of study, with research focusing on a wide range of relevant issues, which is the examination of options to make a judgment. Thus, decisions derive from the complicated topics focused on various disciplines, which are oriented to surrounding approaches turning around the passionate discussion. However, decision theory has focused on appropriate decision-making from the methodical examination of the main objective, non-random behaviours, and actions of decision-makers during events or contexts in which an option could be chosen. In addition, the decision problem is the circumstance in which a decision maker determines what to do from a set of possible activities that are impacted by events beyond the decision control to contain multiple results having advantages or disadvantages over time. Therefore, decision theory often concentrates on judgment results as assessed from the predecided parameters and indicated objectivities. (Hansson, 1994; Peterson, 2011; Elnusrati, 2020; Janssen et al., 2017; Simon, 1977; Hansson, 2011; Pomerol & Adam, 2004).

Decision theories describe normative or descriptive decision approaches that produce suggestions from the required judgment of intuitive thinking issued. Thus, a normative decision approach is a concept based on decisions made from requirements fulfilled for effective decision-making. In contrast, the alternative theory is characterized by hypotheses castigated concerning human judgment from the rational theory that illustrates the concept of psychological judgment established in the baselessness of certain unsuitable values. Therefore, recent psychological researches on innovative thinking analyze spontaneous decision and empirical judgment to determine distinct cognitive processing, which capacities require various descriptions of phenomena from

conventional concepts issued by normative theories. (Peterson, 2011; Hansson, 1994; Frantz, 2003; Gigerenzer & Gaissmaier, 2015; Kahneman, 2003).

Descriptive decision theory is an experimental area that attempts to understand and anticipate how individuals execute judgments. The empirical tests demonstrated that human behaviour was incompatible with normative methods that serve as including basis for descriptive decision concepts. However, the descriptive theory also implies that critical decisions in real life from different analyses can be both rational and non-rational. Therefore, descriptive and normative judgment models are distinct disciplines that may not even interact to some extent. (Peterson, 2011; Bell et al., 1988; Simon, 1977; Frantz, 2003).

Recent researches tried to expand the ideas related to decision theory and information theory by implementing them in computers with the rise of advanced. The emphasis is on computer decision-making procedures based on an evolutionary perspective on advanced statistics that provide both human thinking and identical data processing scanned and stored as models in memory, then used to create assumptions and conclusions. In contrast, some systems may reproduce or outperform human judgment from problem-solving abilities more experimented with set up for the occasion. Therefore, the level of cooperation between humans and computers underlines impacts on decision-making requires investigation further. (Graboś, 2004; Doyle & Thomason, 1999; Buchanan & O'Connell, 2006; Pomerol & Adam, 2004).

Classical decision theory for this thesis is based on the statistical concept of a judgment procedure, the condition of numerical hypotheses is often difficult to use in practice. The mechanisms used in classical decision theory have not been demonstrated completely acceptable or sufficient for assisting efforts to optimize decision-making in the area of advanced statistics and Monte Carlo method, in even complex and practical circumstances with unpredictable choices related to judgment options in contexts where the basic presumptions vary.

The qualitative model of decision-making is based on statistical operations using an analytical tool that conducts various frameworks of motivated research. Qualitative decision theories attempt to adduce suitable assistance for automation by creating qualitative and mixed models and approaches to enhance the quantitative decision theory's capacity to manage the scope of decision-making procedures. (Bell et al., 1988; Hansson, 1994; Simon, 1977; Buchanan & O'Connell, 2006; Janssen et al., 2017).

Decision-making is focused on a topic broadly examined in the scope of the descriptive decision theories widely developed in the subtitles 2.3.1 and 2.3.2, given the significance of thinking in management, and managerial decision support systems. Nevertheless, various types of judgments and decision-makers vary in the continuing question study based on who, where, what, when, why, and how of decision-making. Therefore, different theories revolve around the key factors defined by the significant concepts mentioned such as the decision-making process, the decision-maker, and the decision. These elements are used to support managerial decision-making and are extensively developed in the following subsections.

2.6.1 Decision-Making Process

The decision-making process is progressive and implicates the analysis based on the selection and examination of ideas coming from the main topic. This structured method follows the decision-making process sequentially and includes knowledge, design, selection, and evaluation of implementation. Thus, intelligence collects the facts and data relevant to the choice and judgment, whereas the method treats different options to decide eventual solutions and whether they could satisfy the expected purposes. Nonetheless, the decision-making procedures depend on the established organization for making structured or unstructured judgments specific to the complexity of the decision problem observed. Therefore, different suitable selections related to the decision are difficult to execute if either of the steps is omitted to complete the option of the alternatives considered. (Buchanan & O'Connell, 2006; Langley et al., 1995; Frisk & Bannister, 2017; Pomerol & Adam, 2004).

An efficient decision is achieved through a methodical procedure with perfectly specified elements that follow a series of organized steps. The proposed stages of the process begin by ordering the problem as conventional or exceptional, followed by describing the problem, defining the solution to this issue and the limit requirements, concluding what is correct as opposed to what is adequate to fulfill the limit criteria, changing the choice into action, and assessing the decision by verifying its validity and efficiency opposed to the current sequence of occasions. Therefore, unstructured decisions refer to decision procedures that have never been met before and do not have a specific set of organized answers. (Drucker, 1967; Grover & Lyytinen, 2015; Intezari & Gressel, 2017; Lyytinen et al., 2020).

As illustrated in **Figures 6 and 7,** decision-making is a linear procedure based on the formula of thinking first applied to the decision-making process. However, the decision-making process is a rational approach and persistent defined in several steps that consist of defining, diagnosing, designing, and deciding among different found solutions to a raised problem. The primary goal of decision-making is to gather pertinent information about the problem to be studied and then generate all potential options by analyzing the significant effects of the conclusion based on selecting several optimal choices. (Mintzberg and Westley, 2001; Kalantari, 2010).

Decision questions are imprecise and indistinct for which there is no pre-characterized method or ideal arrangement. In any case, human instinct is based on knowledge and empirical judgment that can serve as the ground for decision-making. Nevertheless, decision-making does not necessarily follow organized or pre-characterized steps but can be based on a mixture of information, evidence, and intuitions. Moreover, the relevance of each step of the decision-making process can change depending on the situation, the period, the strategy adopted, and the effect of the consequences of the decision. Mintzberg & Westley, 2001; Intezari & Gressel, 2017; Provost & Fawcett, 2013).

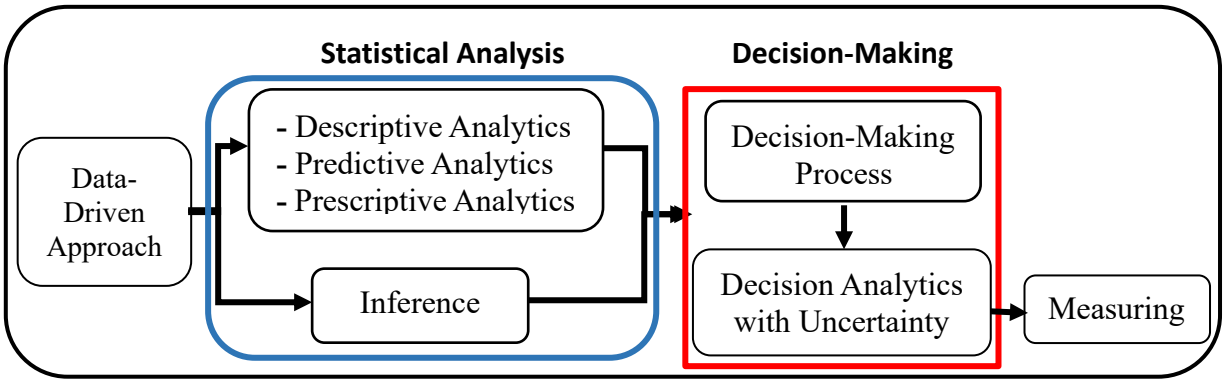


Figure 14. Diagram of the Data-driven Approach applied to the Decision-making Process.

A data-driven approach to the decision-making process consists to the organized collection, analysis, evaluation, and processing of data while using analytical methods and techniques to gain informed decisions. However, the approach begins with the identification of problems and challenges, followed by the definition of strategic purposes and critical success factors, and finally the evaluation and analysis of one of these options.(Mandinach, 2012; Elgendy & Elragal, 2014; Provost & Fawcett, 2013).

Descriptive value is created by gathering data and explaining previous or recent circumstances, predictive value is created by making predictions based on previous data, and prescriptive value offers optimal driving routes and describes the implications. However, the approach is to guide and contribute meaningfully to quality judgments based on understanding data, analyses, connections between factors, and resulting information to improve the quality of decisions. In addition, a data-driven approach to decision-making addresses the challenges of bounded rationality, which refers to the limitations of the cognitive abilities of the human mind, the lack of available data, or the lack of knowledge inability to operate large amounts of possible information to reach an ideal conclusion. (Gregor, 2006; Cao & Duan, 2015; Strand & Syberfeldt, 2020; Janssen et al., 2017).

The method of analysis used in the decision-making process enables strategic or strong decisions to be made based on information organized, combined, filtered, and used as input data allowing decision-makers to make appropriate decisions based on the

availability of data and renewed relationships. While the decision-making model applies to data, analytics, and all three elements of traditional decision-making, additional high-level assumptions about the theoretical links between data and behaviour social are required. Therefore, decision-makers act on the suggestions, and the analysis enables the computerization of decision support from reasonable choices that lead to objective and excellent results. (Simon, 1997; Janssen et al., 2017; Grover & Lyytinen, 2015; Diakopoulos, 2016; Power, 2016).

Human thinking and data processing algorithms are identical in that they collect and analyze data from different models by storing the models in memory, and then use the models to formulate predictions or hypotheses. Then, the following subsections provide some suitable cases by proposing a requirement for the theory into practice.

2.6.2 The Decision Maker's Approach

The decision maker is the user who utilizes the decision-making process to make a choice based on the reasoning used in the current data for the given decision. Thus, the gap between the reality of computational intelligence capability and the constraints of human logic versus artificial computation will always exist in computerless rationality. Therefore, the decision-makers appear irrational by their mental capacities on the assertion of excellent control of the environmental aspects. (Kalantari, 2010; Simon, 1959; Peterson, 2011).

The decision maker chooses the main answer thought acceptable rather than trying to achieve an unfeasible perfect solution when presented with an option. However, limited rationality is used to determine the concept that rationality in humans is limited by human estimation. Finally, decision theory expects decision makers to understand that the problem to be solved can identify an effective problem and obtain the necessary data and resources to solve the problem. (Mintzberg, 1975; Power, 2016; Simon, 1997; Kalantari, 2010; Janssen et al., 2017).

The decision-makers never hold a clear vision of their difficulty expressed in the form of an investigation for an acceptable compromise. A solution to the issue is constantly limited in time compared to the available resources. In addition, a reasonable decision-maker allows decisions to be achieved by utilizing any of the decision-making procedures based on the context and circumstances. (Kahneman, 2003; Simon, 1959; Bell et al., 1988).

Nonetheless, the traditional approach considers that the decision maker selects between specific and defined options to manage the decision, each of the options having given and defined effects. Hence, this is no longer possible once human discernment and understanding mediate between the decision-maker and the circumstance. (Tsoukiàs, 2008; Kalantari, 2010; Simon, 1959; Peterson, 2011).

Determining the effects is a monotonous and complicated task in the method of selection and the options sought rather than initially given, especially since the information of the decision maker regarding the computational environment is much less than an apparent estimate applied compared to the actual step of a problem. In addition, computers solve situations using heuristics with the examination of available means to demonstrate relevant intelligence from flexible objectives and circumstances.(Hansson, 1994; Mintzberg, 1989; Frantz, 2003; Simon, 1959; Janssen et al., 2017).

Nonetheless, human intelligence enables the computational resources determined by the decision maker and the instrument to implement effective investigative procedures to create viable solutions. Decision hypotheses are perceived in computing terminology, and then numerical computation terms or binary notation give a technique to model complicated human decision procedures. The human decision-maker clarifies the object of the concept of decision theory. (Janssen et al., 2017; Frantz, 2003; Hansson, 1994; Simon, 1977).

Recent advances in specific decision and selection have distinguished cognitive systems that need different models and interpretations than structured prescriptive and rational

partners. The emphasis is on decision-making procedures resulting from technological innovations and the rise of Monte Carlo method develops the fundamentals of decision theory and information theory by applying them to 'intelligent' users and tools. While classical decision theories depend on numerical representations of decision procedures and numerical concepts are complicated to apply in practice. (Gigerenzer & Gaissmaier, 2015; Kahneman, 2003; Bell et al., 1988; Tsoukiàs, 2008).

Traditional methods of decision theory have proven to be completely inadequate to continue supporting efforts to optimize decision-making in the area of Monte Carlo method, whose complex and realistic situations are the expectations for unexpected decision selections from basic assumptions that are likely adjusted. (Simon, 1977; Frantz, 2003; Janssen et al., 2017).

Nevertheless, research in advanced statistics and analytical hierarchy process is guided by different models and procedures focused on the descriptive interpretation of decision-making from qualitative and quantitative methods. These tools provide the ability to extend qualitative procedures and hybrid interpretations to handle the range of decision-making tasks with the decision support of automation. (Graboś, 2004; Doyle & Thomason, 1999).

Nowadays, decision-makers are facing complex circumstances in decision-making procedures with the evolution of technologies and the increase of data developing a complicated business surrounding in which decisions is challenging to perform with conventional methods because inefficient and bad decisions depend on incomplete and inadequate data used over the decision-making process. Nonetheless, the decision-maker considers the importance of the computer to manage complex tasks quickly with the involvement of different tools as support or significant assets to find solutions in the decision-making process. (Sauter, 2005; Bell et al., 1988; Mohemad et al., 2010; Zhu and Zhang, 2009).

Advanced statistics appears as catalysts for the decision support used in the data-driven approach to reach the appropriate decision based on the classical methods applied. While the human decision-maker remains unique in the decision-management role.

2.6.3 The Decision Approach

The decision is the consequence of an analyzed, evaluated, and reflected path by the decision-maker between the decision-making process and the choice of the appropriate option. Thus, the decision maker is limited by its mental abilities and external factors related to unlimited rationality that the ideal decision cannot be achieved. Nonetheless, the decision-maker creates an optimized model of discernment, by considering the constraints and the surrounding conditions to manage the reflection with the search for an adequate decision. (Kalantari, 2010; Pomerol & Adam, 2004; Simon, 1977; Frantz, 2003).

The traditional model of rationality expects information on the pertinent other options, results, probabilities, and an anticipated shock-free world. It appears essential to make the distinction between this direct and excellent comprehension of the words limited compared to the normal words and the significant words for the others. Moreover, part of the relevant data is obscure or needs to be evaluated, so that the circumstances of rational decision theory are not fulfilled, which makes it unsuitable for optimal thinking and requires theories, heuristics and different assumptions for the extent of the circumstances. (Grover et al., 2020; Sauter, 2005; Gigerenzer & Gaissmaier, 2011; Janssen et al., 2017).

Quality is one of the significant features of a decision involving the speed, accuracy, and correctness of the decision. Validity and reliability are seen as characteristics of value and quality applied to quantitative decisions. Previous information investigations utilize shows by indicating that the nature of the information impacts the quality of decisions. Moreover, the quality of the decision does not only depend on the information likewise

the procedure for collecting the information and handling it. (Sauter, 2005; Mohemad et al., 2010; Bell et al., 1988; Zhu and Zhang, 2009).

Therefore, the reason for the decisions is based on the significant change through technical evolution over the years The human mind is supported by the computational capacity and basic analytics of computers that are entirely dependent on machines and algorithms to automate decisions and improve the use of analytics to extract confidential information from data-driven approach and modify the decisions. (Ho, 2017; Janssen et al., 2017; Gregor, 2006; Grover & Lyytinen, 2015).

The data-driven approach used in decision-making is based on the analysis of data analysis around different examined elements of traditional decision-making developed and mentioned in subtitles 2.7.1 and 2.7.2, in which the advanced statistics are used to improve or manage the decision-making issue.

Recent articles show that a significant set of studies continues to operate in this field in the absence of an original and effective theory, which requires new, more effective, authentic tools directed toward the recently explored theories on the conceptual connections between information and artificial intelligence. (Gigerenzer & Gaissmaier, 2011; Alvesson & Sandberg, 2013; Lyytinen & Grover, 2017).

Managers are confronted with an excess of appropriate and filtered information which are issued from the irrelevant information with a modern view on the information overabundance concept. Nowadays, information overload reflects the emergence of the use of data analysis techniques to solve problems of information excess. (Power, 2016; Lyytinen & Grover, 2017; Kalantari, 2010).

The relevant information is the procedure of finding and incorporating data analysis from critical thinking based on visualizing the result of information filtering. However, the use of a data-driven approach enhances decision-making on a distinct scale and offers new methods to achieve various solutions by improving previous decisions based on the actual requirements of managers and the expectations of customers. (Akash, et al., 1999; Triantaphyllou, et al., 1998; Mohamadabadi, et al., 2009).

In conclusion, our thesis study is applied according to the specific methodology based on a particular case annotated in the next chapter, where we try to develop different key vectors whose roles are to contribute to decision-making from these different techniques mentioned earlier in this chapter.

3 Research Methodology

Recent studies realized from different calculations by a fast means involve the analytical approach to provide possible solutions to solve issues related to emissions of gas and fuel consumption. A predictive nonlinear model is applied to an internal combustion engine consists of predicting by simulation the average of fuel consumption due to different operational parameters of the fuel engine. (Shafiee, 2015; Pohekar & Ramachandran, 2004; Wang & Tolk, 2008).

The algorithmic approach of Monte Carlo method is used to develop prediction models and gas emission performance of different blends of biodiesel and diesel fuel depending on the operating circumstances and fuel properties. A nonlinear-based technique is used to model and estimate engine efficiency emissions (NOx, CO, and CO_2) of biodiesel and diesel, based on experimental test data realized from internal combustion engines using direct injection diesel. (Manjunatha, 2012; Jafarmadar, 2015; Bhaskar et al., 2014; Manieniyan & Sivaprakasam, 2013).

An internal combustion engine (ICE) is a heat-using engine in which combustion fuel and oxidant are introduced into a combustion chamber to be transformed into a fundamental element of the flow circuit of the engine fluid. Then, the thermal dilatation and high-pressure gases generated by combustion have a direct impact on specific motor elements used for an internal combustion engine. Therefore, this part of ICE is not included in the main thesis. We limit ourselves to the objectives initially assigned and defined in subtitle **1.4** of **Chapter 1**.

The methodology applied to this study aimed to evaluate the different risks observed by sthe choice of fuel consumption of engines from an internal combustion engine, that generate maximum CO, CO_2 or NO_x gas emissions and higher temperatures due to engine power while polluting the environment. Then, to determine based on the criteria taken into consideration during the various experiments carried out, the choice of the optimal fuel retained for use. Furthermore, this study is operated by the multi-criteria decision-

making (MCDM) method to solve the problem based on the data collected from different experiments realized by the operation of motor control.

3.1 Research Background

Different studies are increasingly focusing on the conversion of energy from renewable resources for which the sector of fuel production is making significant technological progress. Then, the challenge for energy sectors is to reduce CO_2 (Carbon Dioxide) emissions produced from fossil fuels by expanding the production of widely used bioethanol, while analyzing the different solutions available to meet the high needs for biofuels. (Štirbanović, et al., 2019; Shafiee, 2015; Mousavi-Nasab & Sotoudeh-Anvari, 2017).

The decision-making developed previously in **Chapter 2**, is applied in the data-driven approach based on complex questions posed by the MCDM methodology used and capable of making decisions from the appropriate choices made in terms of efficiency and potentiality to find the real means created by the decision support system that applies in different areas of expertise such as the energy sectors, which uses big data in integration with a Monte carlo method. However, the AHP approach analyzes the dataset resulting from the effect of exhaust emissions used in the engine system to provide the performance and a better understanding of the choice of fuel consumption, which is used in engine operation while significantly reducing exhaust emissions of the engine internal combustion from the observed analysis of engine temperature.

Nowadays, the energy problems related to global warming as well as different evolutions in the energy system leading to the concentration of CO_2 in the atmosphere have contributed to increasing various solutions for producing green gas emissions and the use of alternative green energy from renewable energies to reduce the content of harmful gases present in nature.

3.2 Multi-Criteria Decision-Making (MCDM) Methods

Multi-Criteria Decision Making (MCDM) is a multifaceted and well-organized procedure designed to address decision-making challenges in many domains and find the most appealing solution while taking into account all pertinent criteria. MCDM methods develop a set of framework-based criteria for selecting approaches and ranking different techniques. (Haddad & Sanders, 2018; Shafiee, 2015; Pohekar & Ramachandran, 2004).

MCDM approaches are considered an important and effective methodology for analyzing and solving difficult decision-making problems in the field of internal combustion engines using different types of fuels for their operation. In selecting the appropriate MCDM methods, it becomes necessary to evaluate the criteria of each procedure from a critical comparison to evaluate the techniques and tools used to determine the energy production technology in terms of analysis and improvement based on the criteria characteristics previously selected for the fuel used despite the complexity of combustion emissions. (Haddad & Sanders, 2018; Jato-Espino, et al., 2014; Al-Najjar & Alsyouf, 2003; Siksnelyte, et al., 2018).

An MCDM technique can be written in the following decision matrix form:

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{mn} \end{bmatrix} \; m, n \in \mathbb{N}. \tag{21}$$

and

$\{a_i\}\, i = 1,2,3,\ldots\ldots, m$

$\{c_j\}\, j = 1,2,3,\ldots\ldots, n$

$\{w_j\}\, j = 1,2,3,\ldots\ldots, n$

Where, $\{a_i\}\, i$ represent the appropriate alternatives. $\{c_j\}\, j$ represent the attributes (criteria). X_{ij} denotes the performance score of the *i*th alternative; and $\{w_j\}\, j$ denotes the weight (importance) of the *j*th attribute.

MCDM performs solutions to issues including conflicting criteria and different objectives in energy decisions for the energy sector by taking into account uncertainty and using various methods for problems such as weighted averages. Then, the MCDM approach incorporates improved correlation analysis based on multiple objectives, which values the optimization of different uncertainties, to solve various problems related to the energy sector. (Shafiee, 2015; Pohekar & Ramachandran, 2004; Al-Najjar & Alsyouf, 2003; Siksnelyte, et al., 2018; Manjunatha, 2012).

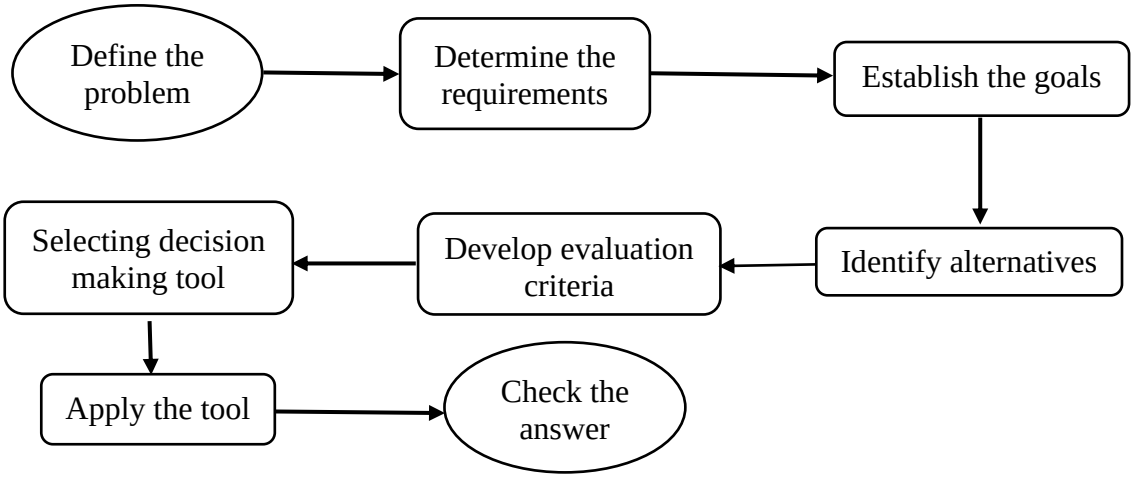


Figure 15. General Multi-Criteria Decision Making (MCDM) Process.

The energy sector represents a critical evolution that takes into account the probative results obtained from energy technologies. This approach provides viable solutions to improve energy management techniques while reducing carbon emissions to achieve sustainable development. However, the optimization objective is to choose conventional intelligence algorithms to solve the multiple problems created using the integrated approach combining algorithms and reasoning based on engine combustion evidence while proposing solutions from the efficiency approach applied to the energy sector. (Anand, et al., 2006; Jato-Espino, et al., 2014; Bhaskar, 2014; Akincilar & Dagdeviren, 2014; Mousavi-Nasab & Sotoudeh-Anvari, 2017).

Decision-making is a complex process that the decision-maker incurs entirely to obtain the final expected result. However, the decision-making process is the daily routine or task used by business leaders as an effective tool capable of analyzing different

perspectives from various data sets while selecting the appropriate criteria to achieve the goals and desired objectives in the face of the decision-making problems posed. (Shahsavari & Khamehchi, 2018; Ho, et al., 2010; Janic & Reggiani, 2002; Mulliner, et al., 2016).

As the subtitle illustrates in Figures 15 and 17 above, MCDM methods are procedures typically used in renewable and sustainable energy while including energy resources for energy decision-making. As a reminder, the MCDM method does not indicate or prescribe how decisions will be made. it simply guides the choice of a logical and reasonable decision based on a selection of different criteria or sub-criteria and alternatives available when carrying out the different experiments carried out. in other words, the MCDM method is used to help with decision-making and not to decide when a possible problem arises. (Akash, et al., 1999; Triantaphyllou, et al., 1998; Simanaviciene & Ustinovichius, 2010; Mohamadabadi, et al., 2009; Albayrak & Erensal, 2004).

The MCDM structure includes at least three distinct levels as illustrated in Figure 18 and explained in Figure 19: The upper level corresponds to the general objective of the problem, the decision is influenced at the intermediate level by the criteria or attributes, and the lower level consists of the decision choices which represent competing alternatives. The process of developing this structure certainly serves to define more clearly all aspects of the decision, in addition to recognizing the relationships between the elements taken in common due to their different characteristics capable of resolving the problem posed. (Ho, et al., 2010; Janic & Reggiani, 2002; Mulliner, et al., 2016; Štirbanović, et al., 2019; Mohamadabadi, et al., 2009).

3.2.1 Multi-Attribute Decision-Making (MADM) Methods

MCDM methods are potential instruments for studying and analyzing complex real problems due to the increasing importance of providing the inherent capacity to appraise different alternatives so that the possibility of selecting the appropriate

alternative may be examined in greater depth for their ultimate implementation. (Shafiee, 2015; Pohekar & Ramachandran, 2004; Ho, et al., 2010).

MCDM methods are generally classified into two different groups (MADM and MODM) based on several parameters of the problem posed during the different experimental phases realized. These methods are applied according to the characteristics of the decision issues.

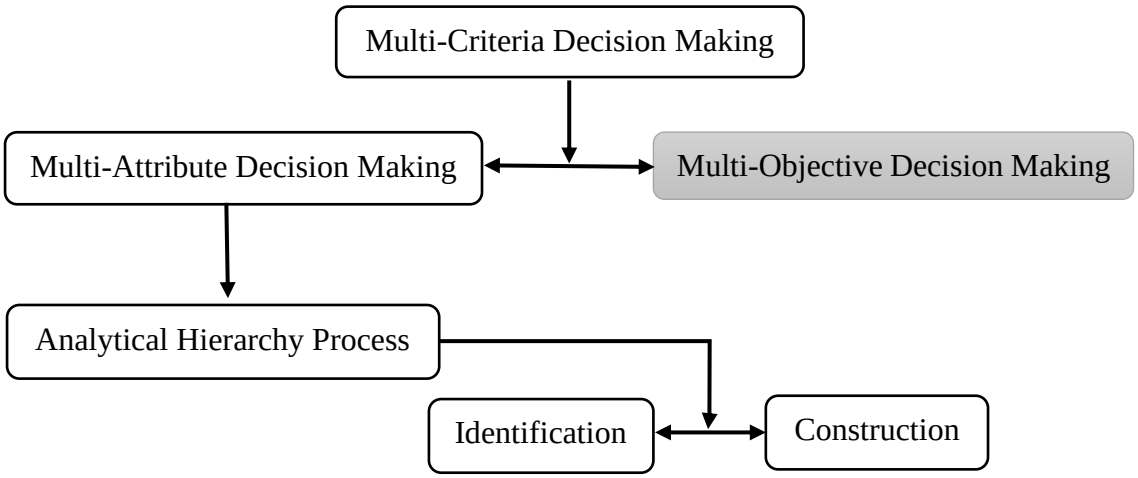


Figure 16. Diagram of the Multi-Criteria Decision Making (MCDM) Methods.

MADM (Multi-Attribute Decision Making) assigns decisions to problems with an implicit objective for a defined discrete decision space, namely a finite number of alternatives and specific attributes. MODM (Multi-Objective Decision Making) is not within the scope of our study subject. then, we cannot consider it at this methodology step (Albayrak & Erensal, 2004; Simanaviciene & Ustinovichius, 2010).

The performance of MCDM methods consists of integrating different techniques to improve the reduction of uncertainties related to multiple risks to make reliable decisions in the selection of renewable energy technologies. This approach takes into account the influencing factors that depend on the algorithms (**Monte Carlo** method, **Classification** and **Regression** techniques) used to extract the criteria, sub-criteria, and alternatives in the dataset to arrive at the decision-making. (Shahsavari & Khamehchi, 2018; Ho, et al., 2010; Janic & Reggiani, 2002; Mulliner, et al., 2016).

MCDM methods referred to the **Figure 17** depend on the effectiveness of the algorithm used and the criteria found to improve the performance by extracting different sub-criteria from the obtained results of the criteria weighting which are influential factors to be used in decision-making by reducing uncertainties and biases.

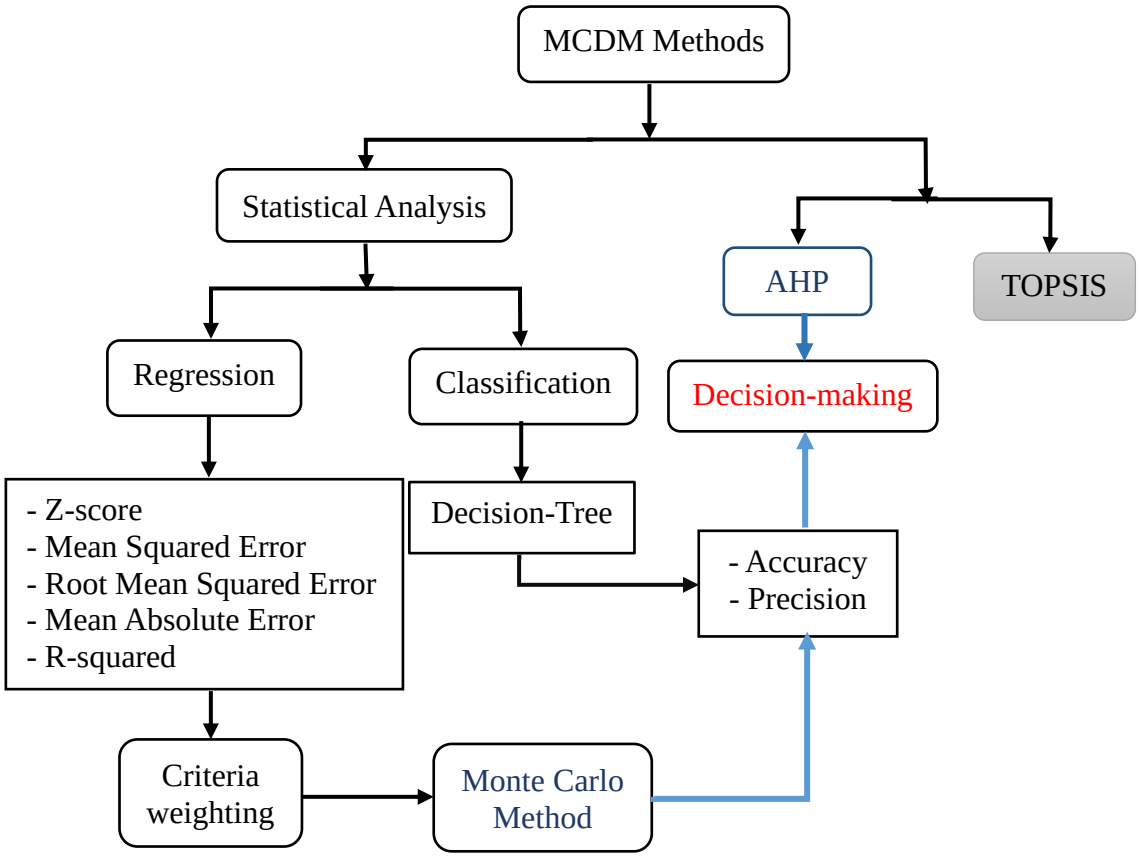


Figure 17. Description of the Multi-Criteria Decision Making (MCDM) Approach.

AHP is used by decision-makers to solve complex problems with the ability to decompose any problem into a hierarchical structure, clearly identifying the main objective, criteria, and sub-criteria that influence decision-making. The AHP approach likewise determines the weighting criteria and ranks the alternatives. (Janic & Reggiani, 2002; Mulliner, et al., 2016).

3.2.2 Classification and Regression Techniques

Classification and regression techniques are necessary to accurately ensure the model used from the evaluation realized by different metrics to perform the tasks.

Classification is a form of supervised learning technique that utilizes the characteristics of an instance to predict categorical values or classes. Then, the classification algorithm is referred to as a function of mapping input attributes to a probability distribution for output classes. The process consists to training a model on a dataset with previously labeled observations and using that model to classify new and unknown situations into one of the predetermined categories.

Two calculation techniques are considered to deal with the classification method:
Accuracy is a calculation model obtained by dividing the percentage of predictions and the number of correct predictions by the total number of predictions.

$$Accuracy = \frac{TP+TN}{P+N} \tag{22}$$

Precision is a calculation model performed by dividing the percentage of positive predictions and the number of true positives by the total number of positive predictions.

$$Precision = \frac{TP}{TP+FP} \tag{23}$$

where *TP* is a true positive error, *TN* is a true negative error, *FP* is a false positive error, *FN* is a false negative error, and (*P, N*) numbers of positive and negative classifications.

Regression is a form of supervised learning technique that predicts defined continuous values based on the mapping function of input characteristics to output values. Then, the procedure involves to estimating correlations between dependent variables or criteria variables and a range of independent variables or predictors. Furthermore, the dataset is processed with the target variable expressed as continuous data.

Different calculation techniques are considered to deal with the regression method:

The mean square error (*MSE*) computes the average of the squared variations between the predicted values and the observed values of the variable, with the best model obtained by the lowest *MSE* values.

$$MSE = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 , n \in \mathbb{N}. \quad (24)$$

where, y is the observed values of the variable, $\hat{y}$ is the predicted values of the *i*th observation and n is the number of predictors generated from a sample of variable data points.

The Root Mean Squared Error (*RMSE*) is a measure that quantifies the standard deviation of the prediction errors. It is calculated from the square root of the Mean Squared Error (*MSE*). The model performance is improved by the indication of Lower *RMSE* values in comparison to *MSE* values.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} , n \in \mathbb{N}. \quad (25)$$

Mean absolute error (*MAE*) is a measure that quantifies the average absolute difference between predicted values and true values.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}, n \in \mathbb{N}. \quad (26)$$

where y_i is the predicted values and x_i the true value. both values are considered weight factors.

The *R*-squared (coefficient of determination) or residual standard error is an estimate of the standard deviation of the error term ε, which quantifies the fraction of the variance of the target variable taken into account by the equation model (15).

$$RSE = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} , n \in \mathbb{N}. \quad (27)$$

and

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}, n \in \mathbb{N}. \quad (28)$$

Let define the equations TSS and RSS as follows previously the equation (15):

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2, n \in \mathbb{N}. \quad (29)$$

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2, n \in \mathbb{N}. \quad (30)$$

Where *TSS* is the total sum of squares which measures the total variance in the response y. *RSS* is the residual sum of squares that measures the number of unresolved variabilities after running the regression. Then, the equation (29) is the square root of equation (30) divided by (n - 2).

$$RSE = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \sqrt{\frac{1}{n-p-1}RSS} = \sqrt{\frac{RSS}{n-2}}, n \in \mathbb{N} \quad (31)$$

- For $0 \leq R^2 \leq 1, \begin{cases} R^2 = 0 -> \text{No association} \\ R^2 = 1 -> \text{Perfect fit} \end{cases}$

Then, the better model fit is indicated by higher R-squared values while *RSE* measures the lack of fit.

Therefore, the objective of using statistical techniques at this stage is to identify different criteria by applying the Analytical Hierarchy Process (AHP) model to evaluate the appropriate decision, which selects the criteria, sub-criteria, and alternatives relevant to the situation from the data set collected during the implementation phase carried out.

3.3 Analytical Hierarchy Process (AHP) Methods

The Analytic Hierarchy Process (AHP) is a mathematical and specific technique for analyzing complicated problems at many levels. However, AHP is one of the most commonly used techniques in real energy consumption circumstances, taking into consideration the financial cost and environmental impact of production, which are

chosen as the primary criteria. (Saaty, 1980; Shafiee, 2015; Pohekar & Ramachandran, 2004).

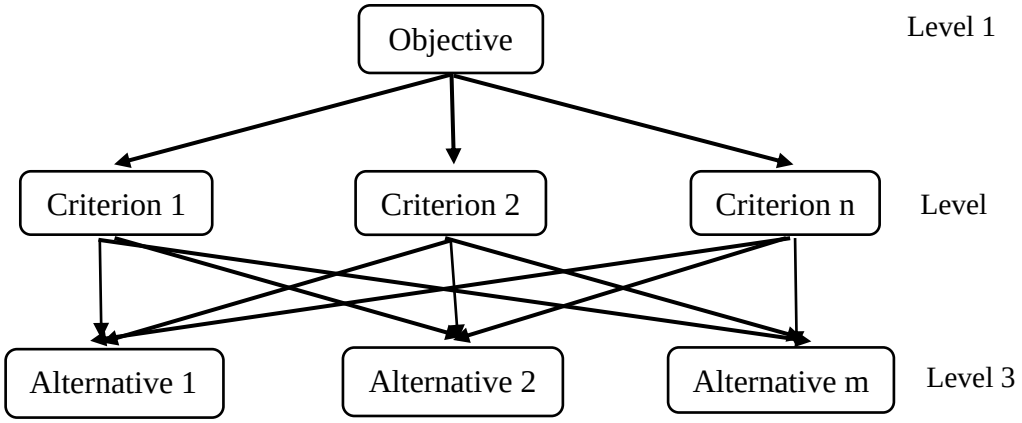


Figure 18. AHP model Hierarchical Trees.

The AHP makes it possible to select and analyze different criteria from criteria weighting to improve the efficiency and quality of production based on the most logical results obtained in the field of renewable energies. Then, this method allows decision-makers to choose techniques to deal with a complex problem based on different results obtained during various phases of experimentation. (Shahsavari & Khamehchi, 2018; Janic & Reggiani, 2002; Saaty & Tran, 2007).

The AHP is used in this study to analyze a defined fuel and gas production strategy and impacts that consider financial, environmental, safety, and pollution effects to evaluate related issues throughout the experimental selection process.

3.3.1 Identification and Selection Criteria

Different stages of the AHP methods used in this thesis consist of the following four levels: The first phase identifies the context problem's purpose in the main criteria including risks, costs, and opportunities. The next phase is to split different objectives into three sub-criteria, such as environmental impacts, enhancing social-economic and technical innovation by selecting appropriate techniques for evaluating combustion and emission performance. Furthermore, the third phase divides the major criterion into multiple sub-

criteria and provides more details by describing each sub-criteria. Finally, this phase summarizes the three levels of criteria and sub-criteria chosen from the analytical process hierarchy (AHP) model, while evaluating several possibilities for consideration in future research.

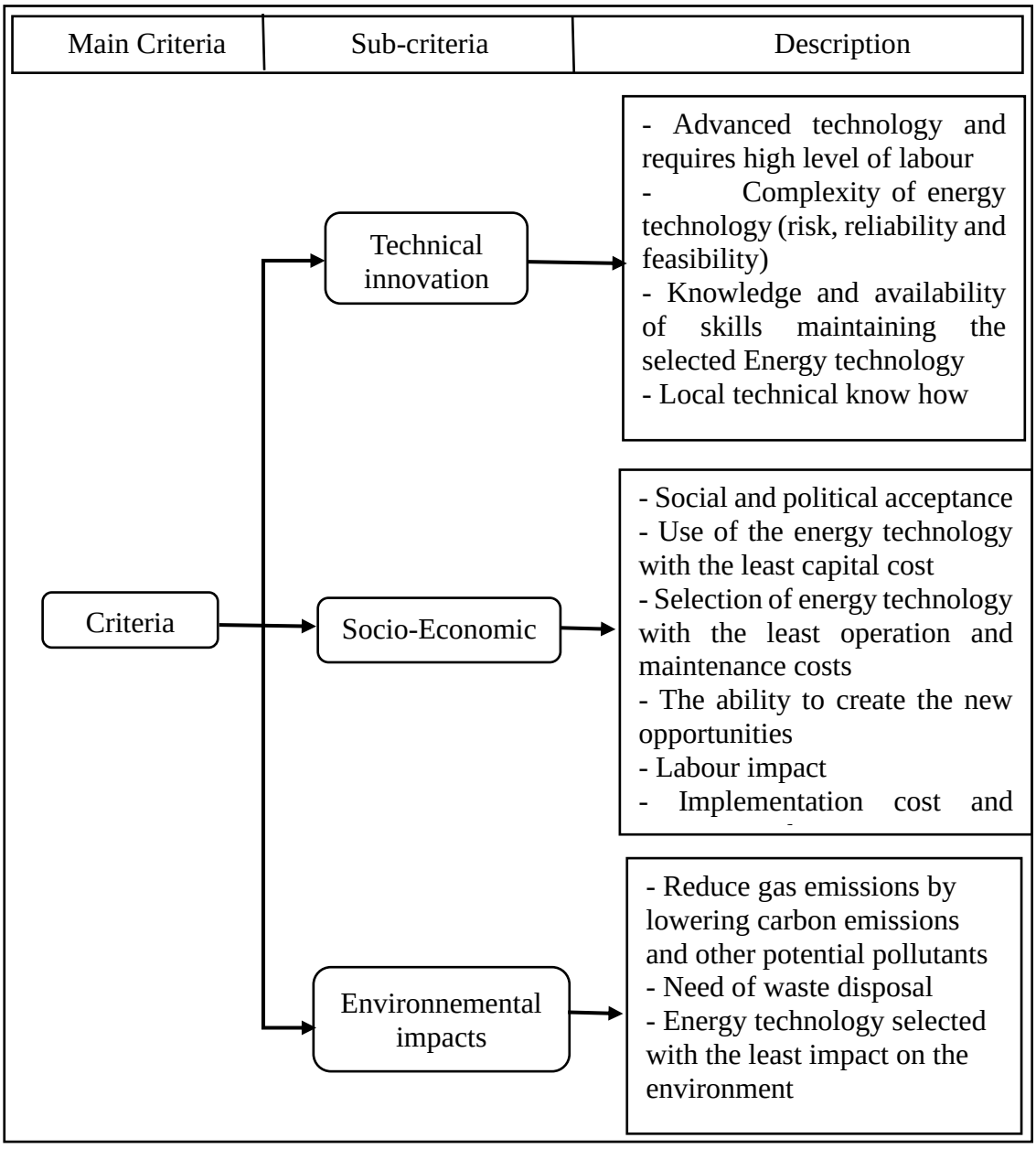


Figure 19. Criteria, Sub-criteria and Description in the AHP Model.

3.3.2 Construction of the AHP Approach

The AHP method is the fundamental pairwise comparison used by the decision maker to develop the objective of the judgment matrix from the quantitative measure whose criteria weighting is selected for the evaluation of each requested step. (Saaty, 1980; Saaty & Tran, 2007).

Considering $C = \{C_i \ / \ i = 1, 2, \ \ldots\ldots, n; \ n \in \mathbb{N}\}$ is the set of criteria related to the level of influence based on the upper level of specified criteria.

$$A = (a_i)_{nxn} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix}, n \in \mathbb{N}. \quad (32)$$

The pairwise comparison on n criteria is a (n x n) evaluation matrix A. Each element $a_{ij}(i, j \ = \ 1,2,\ldots\ldots, n)$ is the quotient of the weights of the criteria. Then, the pairwise comparison represents the square two-to-two and reciprocal matrix of A of equation (32).

$$A_{nxn} = \prod_{i,j=1}^{n} a_{ij} a_{ij+1} = [a_{11} a_{12} \ldots a_{1n} a_{2n} \ldots a_{n1} a_{n2} a_{nn}], n \in \mathbb{N}. \quad (33)$$

Where, A_{nxn} is the original matrix and a_{ij} are the relative weights. The normalized relative weights are computed by the sum of each column S_j divided by each relative weight a_{ij} by the sum of that column.

$$Norm_{nxn} = \left[\frac{a_{11}}{S_1}\frac{a_{12}}{S_2} \ldots \frac{a_{n1}}{S_1}\frac{a_{n2}}{S_2} \ldots \frac{a_{nn}}{S_n}\right], n \in \mathbb{N}. \quad (34)$$

Where $S_j = \ \sum_{i=1}^{n} a_{ij}$, with $j = \ \{1, \ldots\ldots, n\}, n \in \mathbb{N}$.

We assume that the sum of equation (33) can be interpreted from

$$w = [w_1, w_2, \ldots, w_n] = \left[w_1 = \sum_{j=1}^{n} a_{1j}, w_2 = \sum_{j=1}^{n} a_{2j} \ \ldots w_n = \sum_{j=1}^{n} a_{nj}\right], n \in \mathbb{N} \quad (35)$$

where *W* is the principal eigenvector of relative weight established from the normalized principal eigenvector (priority vector). The relative weight vector is determined by the

accurate eigenvector (w) that corresponds to the largest eigenvalue $A_w = \sum_{j=1}^{n} a_{ij} w_j = \lambda_{max} * w_i$, which is calculated from the degree of inconsistency.

$$\lambda_{max} = \sum_{i=1}^{n} \lambda_{max} w_i = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ij} \right) w_j = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} w_j = \prod_{i,j=1}^{n} S_j w_i \, , n \in \mathbb{N} \quad (36)$$

Where λ_{max} is the principal eigenvalue of a positive reciprocal pairwise comparison matrix obtained from the input judgment's consistency $S_j w_i$. However, this value is obtained by measuring the degree of inconsistency.

$$CI = -\frac{\sum_{i=2}^{n} \lambda_i}{n-1} = \frac{\lambda_{max} - n}{n-1}, n \in \mathbb{N}. \quad (37)$$

and

$$CR = \frac{CI}{RI} \quad (38)$$

Where *CI* is a matrix of comparisons noted as the consistency index, *RI* is the average random consistency index, n is the rank of the matrix and *CR* is the consistency ratio.

The use of the final consistency ratio (*CR*) helps in evaluating the adequacy judgment of the evaluation's consistency. Then, the *CR* is computed from the ratio between the random index (*RI*) and the expected value of *CI* depending on the order of matrices. The acceptable superior limit of *CR* is exactly the number 0,1 reached.

Table 6. Random Index (Saaty, 1980).

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49

However, the evaluation judgment is repeated to ameliorate consistency from the matrix A of (32). Each comparison matrix evaluates the consistency index by defining the principal criteria and sub-criteria based on local weights. (Saaty, 1980; Saaty & Vargas, 1991; Saaty & Tran, 2007).

Table 7. Cut-off Consistency Index (Saaty, 1980).

n	2	3	4	5	6	7	8	9	10
Acceptable (10%)	0	0.058	0.090	0.112	0.124	0.152	0.141	0.145	0.149
Tolerable (20%)	0	0.116	0.180	0.224	0.248	0.264	0.282	0.290	0.298

The global weight of the sub-criteria equation is evaluated by multiplying the local weight of each sub-criteria by the overall weight of the principal criteria concerned. (Saaty & Vargas, 1991; Saaty & Tran, 2007).

The following chapter will analyze the case study based on the content of Chapters 2 and 3, which takes into account the decision-making process used in the data-driven approach and involves the use of the appropriate methodological techniques during experiments on engine fuels to address challenges the biodiesels to meet the challenges of biofuels.

4 Implements and Results

4.1 The Case Study

The thesis phase aims to analyze the data collected on the fuels used from their properties and gaseous emissions resulting of several experiments carried out with the internal combustion engine, for which different characteristics of the engine used are taken into account during the experiments carried out.

The adjustment characteristics used by the internal combustion engine during experimental tests are:

- Rotational speed or frequency (RPM)
- Moment or torque (Load)
- Engine power
- Average effective brake pressure
- Brake fuel conversion efficiency
- Brake-specific fuel consumption
- Volumetric efficiency
- Fuel quantity
- Air quantity and air temperature.

However, the different types of fuel used in the experimental tests are the pure diesel and various pure biodiesels such as Swine biodiesel (S75, S50, S25), Turkey biodiesel (T75, T50, T25), Rapeseed biodiesel (R75, R50, R25).

The composition of each biodiesel sample used in the experiments is defined as follows:

- Swine 75% diesel 25%
- Swine 50% diesel 50%
- Swine 25% diesel 75%

The gases emitted by the internal combustion engine are total hydrocarbons (THC), carbon monoxide (CO), carbon dioxide (CO_2) and nitrogen oxide (NO_X). In addition, the

temperature rise in the combustion chamber causes an unsatisfactory occurrence in significant circumstances, including the increase in nitrogen oxide.

The fuel properties respecting the different values of pure biodiesel (Swine, Turkey and Rapeseed) following EN-14214 standards are measured from the ambient temperature inside the engine compartment enclosure which plays an important role in the experiments.

Experimental modelling data was received from Dr. **Kamil Duda** and Prof. **Maciej Mikulski** based on internal combustion engine experiments conducted respectively at University of Warmia and Mazury in Olsztyn, Poland and University of Vaasa, Finland.

The results of the data obtained during the experimental tests are collected from Excel files and analyzed from two downloaded software tools (**Analyze Data** and **Data Analysis**), which are contained in the Microsoft Excel environment. Then, the fuel and gas emissions data gathered are examined, processed and analyzed the data content using algorithmic computing based on the statistical analysis formulas and Excel queries in real-time.

4.2 Analysis Procedures

The analysis procedure aims to study the available data on gas emissions collected from different biodiesel, diesel and blend fuels used in internal combustion engine experiments from the appropriate calculations, methods and tools to solve the problems regarding the choice of suitable fuels with lower cost consumption and capable of reducing the production capacity of harmful gases and promoting the performance and speed of the engine power.

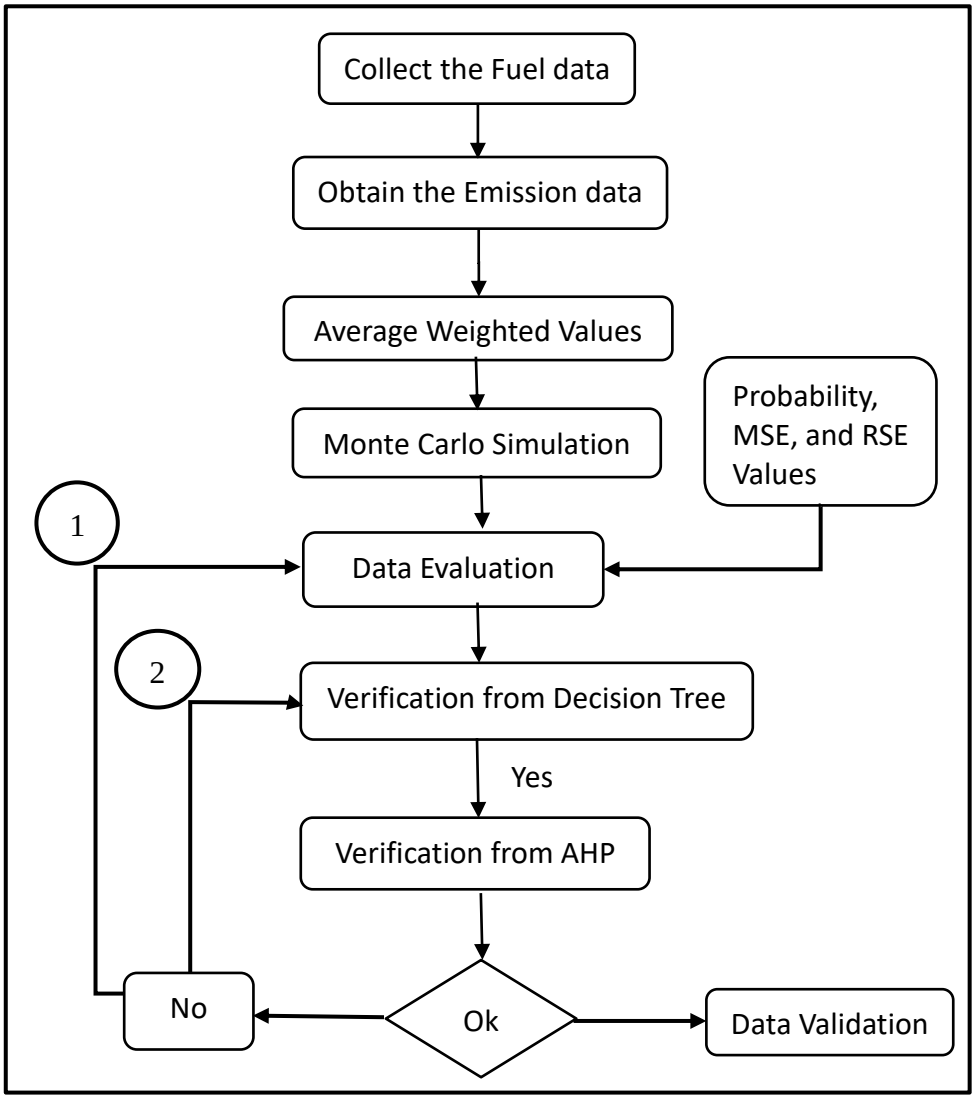


Figure 20. Flow Chart of the Evaluation Data Performance Process.

Figure 20 above illustrates the implementation process from the collected data obtained during the internal combustion engine experiments to the data validation after the analysis performed on different average weight, RSE and consistency ratio values. Then, the selection of the appropriate fuel is led by the indication of better engine emission performance.

Table 8. Emissions Results from the Experiments performed for RPM = 1500.

Fuel	RPM	Load	THC / dry	CO	CO2	NOX
	[obr/min]	[N m]	[g/kWh]	[g/kWh]	[g/kWh]	[g/kWh]
S75		50	0,7687	4,0421	983,06	7,4617
T75			0,8139	4,3342	985,03	7,7102
R75			0,8050	4,3614	994,08	7,9233
S50			0,6781	3,3781	969,41	7,3251
T50			0,7522	3,6746	1019,29	8,1169
R50			0,7693	3,8136	1030,48	8,1937
S25			0,7402	3,8015	948,34	6,7281
T25			0,7375	4,2511	946,65	6,9264
R25			0,7889	3,9342	991,08	7,3282
Diesel			0,8287	3,6373	1014,52	7,6497
S75		100	0,3249	1,3021	878,53	7,4090
T75			0,3665	1,2532	846,29	7,1367
R75			0,3345	1,2506	848,19	6,9880
S50			0,3443	1,2667	838,88	6,9421
T50			0,3690	1,2423	859,52	7,1692
R50			0,3684	1,3419	868,71	7,1125
S25			0,3919	1,3807	876,10	6,3784
T25			0,3875	1,2493	885,26	7,2707
R25			0,3922	1,4566	860,16	6,4126
Diesel			0,3841	1,4068	845,62	6,3513
S75	1500	150	0,2439	0,7961	802,94	9,6299
T75			0,2790	0,8211	842,42	9,8298
R75			0,2361	0,7415	744,64	8,1072
S50			0,2303	0,7062	744,74	8,6539
T50			0,2560	0,7660	762,43	9,2090
R50			0,2530	0,7650	771,11	9,0673
S25			0,2701	0,7516	793,41	8,2430
T25			0,2599	0,7234	760,87	8,3580
R25			0,2594	0,7763	767,71	8,3924
Diesel			0,2392	0,7501	761,19	5,8559
S75		200	0,1719	3,1423	783,00	10,0497
T75			0,1943	3,2269	781,33	10,1286
R75			0,1856	3,2642	758,11	10,0872
S50			0,1960	3,2552	752,09	9,7222
T50			0,2060	3,2933	750,66	10,2289
R50			0,2041	3,4098	752,45	9,8755
S25			0,2094	2,8079	763,54	9,7311
T25			0,2180	2,6240	728,73	9,7657
R25			0,2121	2,8465	737,91	9,6787
Diesel			0,2115	2,7234	756,89	9,5355

Table 9. Emissions Results Data from the Experiments performed for RPM=3000.

Fuel	RPM	Load	THC / dry	CO	CO2	NOX
	[obr/min]	[N m]	[g/kWh]	[g/kWh]	[g/kWh]	[g/kWh]
S75		50	1,2941	9,7831	1324,70	9,6679
T75			1,4126	10,4541	1314,52	9,3778
R75			1,2712	9,7554	1323,57	9,4327
S50			1,0232	9,1167	1341,28	8,7158
T50			1,1508	9,2041	1333,80	9,3536
R50			1,1268	9,7999	1349,51	9,0028
S25			1,3461	10,1600	1402,85	9,2904
T25			1,1774	9,3157	1345,41	9,9260
R25			1,3138	9,7018	1356,70	8,9172
Diesel			1,1098	8,5522	1379,08	9,2399
S75		100	0,4158	1,4821	941,29	7,4466
T75			0,4714	1,4638	957,26	7,6173
R75			0,4395	1,5129	964,45	7,6120
S50			0,4014	1,5008	977,06	7,6451
T50			0,4994	1,4960	973,23	7,9988
R50			0,4317	1,5161	965,27	7,8129
S25			0,4824	1,5823	994,26	7,8102
T25			0,4895	1,5413	969,74	7,8074
R25			0,4569	1,4866	938,97	7,4652
Diesel			0,4380	1,4082	886,02	7,0579
S75	3000	150	0,3395	1,4798	953,29	8,6403
T75			0,3797	1,4288	922,73	8,2498
R75			0,3491	1,2526	935,88	8,1976
S50			0,3286	1,3870	928,49	7,9521
T50			0,3594	1,3736	933,65	8,5177
R50			0,3386	1,2395	920,17	7,9982
S25			0,3850	1,5710	938,60	8,0791
T25			0,4080	1,5853	939,19	7,9704
R25			0,3851	1,5549	929,77	8,0425
Diesel			0,3402	1,6886	894,08	7,4084
S75		200	0,2843	1,2959	879,98	10,3171
T75			0,3293	1,4014	920,92	11,2065
R75			0,3024	1,3442	897,45	10,3442
S50			0,2957	1,4110	891,97	9,9494
T50			0,3588	1,5281	922,81	10,9080
R50			0,3149	1,4509	888,25	10,1629
S25			0,2715	1,9638	884,93	9,6116
T25			0,2920	2,0900	926,23	10,2689
R25			0,2764	1,9671	898,04	9,6462
Diesel			0,2435	2,2771	825,28	8,3760

The first step of the implementation consists to normalize each emission variable using different calculation methods separately, such as the difference between the minimum and maximum, the mean, the standard deviation and the Z-score.

However, we calculate the weighted sums and the maximum of each weight from the different emission variables where each weight is equal to 1. Then, we use the weighted sum method to select the right emission variables correlated with other fuels such as biodiesel, diesel and blended fuel following the hierarchy.

So far, we establish the ranking of fuel options based on different weighted sums previously obtained from the normalized emission variables that are used separately from the data points (RPM, Load). We repeated this step 2, 4 or more times after each iteration to accumulate the different fuel option rankings.

We formulated the conclusions based on the accumulated rankings that the fuel with the lowest cumulative ranking score is the optimal fuel in terms of random emission weights obtained from the values of different means and standard deviations for each combination (RPM, Load). In addition, the optimal fuel is determined and correlated respectively to each combination (RPM, Load).

The second step uses the Monte Carlo method to determine the weights combined with the values of the means and standard deviations from the average optimality method in the weight space, at least randomly with a uniform probability distribution or sampled from a multifaceted regular network.

Another approach used consists of iterating each emission variable separately by varying one weight from 0 to 1 while keeping the other weights equal, and maintaining the optimal visualization possible with the weight space for each emission variable.

4.3 Data Analysis

The objective of this exercise is to compute the difference between the maximum (higher value X_2) and the minimum (lower value X_1) of each emitted gas (THC/dry, CO, CO_2 and NOx) taken separately and belonging respectively to its column from (RPM, Load) combination of each gas point.

The mean is calculated based on the average of the previous results obtained from the maximum and minimum values. Then, the standard deviation is evaluated from the variability of the data set representing each data point taken respectively in each column of emitted gases.

The Z-score is calculated based on the ratio of each range of maximum and minimum values obtained minus the average value of the combination of the combination (RPM, Load) of each gas emitted together divided respectively by the standard deviation value of each gas emitted by the data point

In addition, the minimum and maximum, mean, standard deviation and Z-score values are evaluated to perform the normalization of each engine emission variable by taking into account the sum and average of the weighted values of biodiesel and blend taken separately from each different combination (RPM, Load). of points defined in the tables above.

The below two tables show different values or results obtained from the calculation of the normalization of each gas emission respectively of each combination of points (RPM, Load) taking into account the variation of different load values (50, 100, 150 and 200), which play a key role in adjusting the engine emissions during the experiments.

Table 10. Means, Standard Deviations and Z-scores for each Emission with RPM=1500.

Gas [g/kWh]	RPM [obr/min]	Load [N m]	Max-Min	Mean (μ)	SD (σ)	Z-Score
THC/dry			0,1506	0,7683	0,0443	-13,9559
CO		50	0,9833	3,9228	0,3249	-9,0462
CO2			83,8314	988,1923	28,2727	-31,9870
NOX			1,4656	7,5363	0,4808	-12,6260
THC/dry			0,0673	0,2527	0,0243	-7,6258
CO		100	0,2143	1,315	0,0771	-14,2677
CO2			46,3808	860,726	15,9581	-51,0301
NOX			1,0577	6,9171	0,3926	-14,9234
THC/dry	1500		0,0486	0,2527	0,0154	-13,2486
CO		150	0,1149	0,7597	0,0335	-19,2436
CO2			97,7893	775,1476	30,0369	-22,5509
NOX			3,9738	8,5346	1,1118	-4,1024
THC/dry			0,0461	0,2009	0,0141	-11,0076
CO		200	0,7858	3,0593	0,2797	-8,1298
CO2			54,2742	756,4695	16,8592	-41,6505
NOX			0,6934	9,8803	0,2296	-40,0076

For the load equal to 200, the z-score values perform worse compared to other z-score values taken in account from different load values respectively in the table above. This approach indicates that a particular Z-score value is lower than any mean value of the same load value following the empirical rule applied in statistics used to predict probabilities and normal distributions.

Table 11. Means, Standard Deviations and Z-scores for each Emission with RPM=3000.

Gas [g/kWh]	RPM [obr/min]	Load [N m]	Max-Min	Mean (μ)	SD (σ)	Z-Score
THC/dry			0,3894	1,2226	0,1229	-6,7778
CO		50	1,9019	9,5843	0,5495	-13,9804
CO2			88,3305	1347,1431	26,9964	-46,6290
NOX			1,2102	9,2924	0,3548	-22,7793
THC/dry			0,0981	0,4526	0,0327	-10,8494
CO		100	0,1741	1,499	0,0461	-28,7252
CO2			108,2421	956,7554	29,7215	-28,5488
NOX			0,9408	7,6273	0,2629	-22,7793
THC/dry	3000		0,0795	0,3613	0,0265	-10,6429
CO		150	0,4491	1,4561	0,1472	-6,8422
CO2			59,2156	928,779	15,5979	-55,7489
NOX			1,2318	8,1056	0,3381	-20,3286
THC/dry			0,1153	0,2969	0,0322	-5,6402
CO		200	0,9812	1,6729	0,361	-1,9162
CO2			100,9458	893,5862	29,1704	-27,1728
NOX			2,8305	10,0791	0,7783	-9,3136

For the load equal to 150, the z-score values perform worse compared to other z-score values taken in account from different load values respectively in the table above. This approach indicates that a particular Z-score value is lower than any mean value of the same load value.

Therefore, the z-score values obtained in the two tables are different and alternate in terms of decrease following various adjustments of the (RPM and load) during the emission experiments. The lower z-score values, the higher the mean and standard deviation values, as explained in the table.

Table 12. Weighted Sum and Average Weight Values for each Emission with RPM=1500.

Gas [g/kWh]	RPM [obr/min]	Load [N m]	All Weighted Sum	Weighted Sum	Average Weight
THC/dry	1500	50	7,6827	1	0,1302
CO			39,2281	1	0,0255
CO2			9881,9231	1	0,0001
NOX			75,3633	1	0,0133
THC/dry		100	3,3633	1	0,2730
CO			13,1501	1	0,0760
CO2			6607,2599	1	0,0001
NOX			69,1705	1	0,0145
THC/dry		150	2,5268	1	0,3958
CO			7,5973	1	0,1316
CO2			7751,4764	1	0,0001
NOX			85,3465	1	0,0117
THC/dry		200	2,009	1	0,4978
CO			30,5934	1	0,0327
CO2			7564,6946	1	0,0001
NOX			98,8031	1	0,0101

For the load equal to 50, the average total weight values performed and equal to **0,1709** are less than the other average total weight values considered respectively from different load values in the table above. This approach indicates that the lower average total weight values have the highest total weighted sum values with **10004,1972** relative to any total weighted sum value taken from the different load values in the table. One of the different average weight values of the emitted gases (THC/dry, CO, CO_2 and NO_X) is equal to **0,0001**, this implies that different curves of the emitted gases converge towards a precise point of the reference frame, which is defined in the graph.

By selecting each emission variable respectively to the range of emitted gases that corresponds to the nature of the gas and belonging to the same load, the method is used to calculate the weighted sum of each emission gas from the different emissions.
Additionally, each weighted sum value of each emitted gas of different weight values is equal to 1 as shown respectively for each combination (RPM, Load) following the hierarchy of load ranges taken between (50, 100, 150 and 200) as shown results indicated in the two tables above.

Table 13. Weighted Sum and Average Weight Values for each Emission with RPM=3000.

Gas [g/kWh]	RPM [obr/min]	Load [N m]	All Weighted Sum	Weighted Sum	Average Weight
THC/dry			12,2259	1	0,0818
CO		50	95,8432	1	0,0104
CO2			13471,43	1	0,0001
NOX			92,924	1	0,0108
THC/dry			4,526	1	0,2209
CO		100	14,9901	1	0,0667
CO2			9567,55	1	0,0001
NOX	3000		76,2734	1	0,0131
THC/dry			3,8132	1	0,2768
CO		150	14,561	1	0,0687
CO2			9296,85	1	0,0001
NOX			81,0561	1	0,0123
THC/dry			2,9687	1	0,3368
CO		200	16,7294	1	0,0598
CO2			8935,86	1	0,0001
NOX			100,7907	1	0,0099

For the load equal to 50, the average total weight values performed and equal to **0,1031** are less than the other average total weight values considered respectively from different load values in the table above. This approach indicates that the lower average total weight values have the highest total weighted sum values with **13672,4231** relative to any total weighted sum value taken from the different load values in the table. One of the different average weight values of the emitted gases (THC/dry, CO, CO_2 and NO_X) is equal to **0,0001**, this implies that different curves of the emitted gases converge towards a precise point of the reference frame, which is defined in the graph.

The average weight is the outcome obtained from the ratio of each weighted sum and all weighted sums. Then, the calculation is carried out separately for each emitted gas as shown by different results obtained from different weight values in the different tables.

Table 14. Ranking of Fuels from the Weighted Sums with RPM = 1500.

Fuel	RPM [obr/min]	Load [N m]	Weighted Sum	Average Weight	Rank
S75			0,4016	0,2495	5
T75			0,4184	0,2512	9
R75			0,4217	0,265	10
S50			0,3697	0,2399	1
T50		50	0,4024	0,248	6
R50			0,4103	0,2509	8
S25			0,3785	0,2389	2
T25			0,3921	0,2475	3
R25			0,4005	0,2492	4
Diesel			0,4048	0,2598	7
S75			0,3969	0,2466	4
T75			0,3968	0,2383	1
R75			0,386	0,2426	2
S50			0,3881	0,2518	3
T50		100	0,3987	0,2457	5
R50			0,4064	0,2484	8
S25			0,406	0,2562	7
T25			0,4088	0,258	9
R25			0,4105	0,2554	10
Diesel	1500		0,4019	0,258	6
S75			0,4177	0,2595	9
T75			0,4423	0,2656	10
R75			0,3821	0,2401	3
S50			0,3816	0,2476	2
T50		150	0,4084	0,2517	8
R50			0,4065	0,2485	7
S25			0,4047	0,2554	6
T25			0,3942	0,2488	4
R25			0,4022	0,2502	5
Diesel			0,3602	0,2312	1
S75			0,3935	0,24446	3
T75			0,40799	0,24495	8
R75			0,40141	0,25227	6
S50			0,40179	0,2607	7
T50		200	0,41296	0,25452	10
R50			0,41247	0,25217	9
S25			0,39543	0,24954	5
T25			0,38945	0,2458	1
R25			0,39413	0,2452	4
Diesel			0,39086	0,25091	2

The ranking of each fuel depends on the result obtained from the calculation of each weighted sum taken respectively in each load range varying between (50 to 200), which shows the different fuels used separately to achieve the emitted gas produced during the experiments while collecting appropriate data from the analysis calculation for the improvement of combustion and engine consumption.

In addition, the ranking of each fuel varies depending on each RPM, load and weighted sum value taken separately. We observed that the first ranks occupy the heads of the table ranking according to their performances carried out in weighted sum value and average weight value are (S50, T75 and T25) biodiesels and pure diesel, as shown in the table above.

Table 15. Ranking of Fuels from the Weighted Sums with RPM = 3000.

Fuel	RPM [obr/min]	Load [N m]	∑ of Weights	Average Weight	Rank
S75			0,4103	0,2604	8
T75			0,4231	0,2583	10
R75			0,4055	0,2577	7
S50			0,3722	0,2433	1
T50		50	0,3898	0,2392	3
R50			0,3915	0,2508	4
S25			0,4202	0,2536	9
T25			0,4002	0,2404	5
R25			0,4054	0,2501	6
Diesel			0,3818	0,2465	2
S75			0,3868	0,2454	2
T75			0,4017	0,2452	7
R75			0,3986	0,2533	5
S50			0,3912	0,2557	3
T50		100	0,4167	0,2557	9
R50			0,3998	0,2562	6
S25			0,4185	0,2525	10
T25			0,4147	0,2491	8
R25			0,3961	0,2444	4
Diesel	3000		0,3759	0,2427	1
S75			0,4047	0,2568	7
T75			0,4043	0,2468	6
R75			0,3845	0,2443	3
S50			0,3842	0,2511	2
T50		150	0,3993	0,245	5
R50			0,3765	0,2412	1
S25			0,4151	0,2505	9
T25			0,4212	0,253	10
R25			0,4126	0,2545	8
Diesel			0,3977	0,2568	4
S75			0,37406	0,23737	1
T75			0,40894	0,24966	8
R75			0,38527	0,24479	3
S50			0,38249	0,24999	2
T50		200	0,42369	0,26	9
R50			0,39302	0,2518	4
S25			0,40322	0,24334	6
T25			0,42881	0,25757	10
R25			0,4069	0,25102	7
Diesel			0,3936	0,2541	5

In general We note a different scenario with a slight change compared to the previous table observed. The first ranks occupy the heads of the table ranking according to their performances carried out in weighted sum value and average weight value are biodiesels (S50, R50, S25) and pure diesel, as shown in the table above.

Therefore, a significant change occurs from number 3 in the ranking where different species constantly exchange their previously occupied positions based on the high values of the weighted sums and the average weight shown in both tables.

The Monte Carlo approach used in this exercise is a conceptual algorithm for solving for mean and standard deviation values by iterating 2000 times random samples obtained from the previous calculation of mean and standard deviation results, which provide estimated solutions from different means and standard deviations analyzed to predict appropriate results while using the given data.

Table 16. Monte Carlo and Cumulative Value Simulation with RPM = 1500.

Gas [g/kWh]	RPM[obr/min]	Load [N.m]	Mean	St Deviation	Monte Carlo	Cumulative
		50	0,1	0,0058	0,09796	0,54358
		100	0,1	0,0066	0,10138	0,54241
THC/dry		150	0,1	0,0061	0,0938	0,53871
		200	0,1	0,007	0,09587	0,53769
		50	0,1	0,0083	0,08914	0,54702
		100	0,1	0,0059	0,09403	0,54046
CO		150	0,1	0,0044	0,09832	0,54032
		200	0,1	0,0091	0,09606	0,53923
	1500	50	0,1	0,0029	0,0979	0,53882
		100	0,1	0,0019	0,10435	0,53865
CO2		150	0,1	0,0039	0,09946	0,54115
		200	0,1	0,0022	0,10552	0,54043
		50	0,1	0,0064	0,10687	0,54086
		100	0,1	0,0057	0,10141	0,54396
NOX		150	0,1	0,013	0,08909	0,53493
		200	0,1	0,0023	0,09628	0,54096

Then, the algorithms used to compute the Monte Carlo approach, cumulative and probability are the formulas expressed in Excel queries that determine each weight value

or cumulative (0 < weighted value ≤ 1) of the emitted gas from a uniform probability distribution from the mean and standard deviation values obtained previously associated with a random sample.

In addition, we use the probability by running the Monte Carlo algorithm repeatedly for exactly 2000 iterations while testing the minimum, maximum, mean and standard deviation values, which are bounded above zero and considered to satisfy the previous values obtained in the various calculations performed.

Table 17. Monte Carlo and Cumulative Value Simulation with RPM = 3000.

Gas [g/kWh]	RPM [obr/min]	Load [N m]	Mean	St Deviation	Monte Carlo	Cumulative
		50	0,1	0,010054851	0,09653035	0,5363042
		100	0,1	0,007220149	0,09111889	0,5388666
THC/dry		150	0,1	0,007329779	0,11279398	0,5391409
		200	0,1	0,010846841	0.08466881	0,5394463
		50	0,1	0,005733429	0,10312163	0,5350979
		100	0,1	0,003076987	0,09357952	0,5401215
CO		150	0,1	0,010107588	0,09594221	0,5445808
		200	0,1	0,021578483	0,09117869	0,538618
	3000	50	0,1	0,002003972	0,10291817	0,5392984
		100	0,1	0,003106484	0,09912841	0,5406243
CO2		150	0,1	0,001677939	0,09935511	0,539753
		200	0,1	0,003264419	0.099599253	0,5389524
		50	0,1	0,003818239	0,09957867	0,5427156
		100	0,1	0,003446864	0,10417405	0,5409112
NOX		150	0,1	0,004171623	0,10322021	0,540268
		200	0,1	0,007721706	0,09228205	0,5408579

We calculate the Monte Carlo approach from the mean, standard deviation values and a random sample for each fuel used during the experiment. We create a simulation with 2000 iterations directly connected to the result which is related to the Monte Carlo algorithm. When we run the Monte Carlo algorithm from the previously obtained mean and standard deviation values, the simulation performs 2000 iterations calculating from different values obtained by the Monte Carlo approach and shows the simulation results in the results column.

The calculation values of MSE, RMSE and RSE for each emitted gas are obtained from the average of each emitted gas taken separately from the entire range of adjusted values of the RPM. It is based on the average of each gas contained in the whole composition of all emitted gases, which are diffused by the biodiesels and diesels used in each experiment carried out.

However, the average of each emitted gas is calculated based on the sum of the values produced by each gas in each result obtained during each experiment performed. Then, the average square value of each emitted gas is raised to the power of 2 to obtain the mean square of each emitted gas value, which is the key factor in calculating the mean square error (MSE) of each emitted gas in the average of each gas produced during the propagation of the emission gases.

Four scenarios are offered for each load value to calculate different MSE, RMSE, and RSE values for each emitted gas. These calculations are handled by different formulas of MSE, RMSE, and RSE defined and mentioned in subtitles 2.5.5 and 3.2.2. Then, different appropriate formulas used to solve these operations are extracted from their Excel queries respectively.

Table 18. MSE, RMSE and RSE Values of THC/dry Gas with RPM = 1500.

Fuel	RPM [obr/min]	Load [N.m]	Average of THC/dry [g/kWh]	Mean Square of THC/dry	MSE_THC/dry
S75	1500	50 to 200	**0,1952**	0,0381	**0,0673**
T75			**0,2961**	0,0877	
R75			0,2769	0,0767	
S50			0,2285	0,0522	RMSE_THC/dry
T50			0,2620	0,0687	**0,2595**
R50			0,2634	0,0694	
S25			0,2546	0,0648	
T25			0,2444	0,0597	RSE_THC/dry
R25			0,2709	0,0734	**0,5094**
Diesel			0,2872	0,0825	

The average of THC/dry is calculated from the entire column of emitted gas THY/dry which is combined with these different types of biodiesels and diesel mixed and ignited

during the experiment with an internal combustion engine to finally produce various gases. The lower and upper mean THC/dry values are 0.1952 and 0.2961 respectively and correspond to the overall mean values of S75 and T75 fuels taken over the entire THC/dry column. The mean square error (MSE_THC/dry), root mean square error (RMSE_THC/dry) and mean square error (RSE_THC/dry) values shown in Table 18 are 0.0673, 0.2595 and 0.5094 respectively.

Table 19. MSE, RMSE and RSE values of CO Gas with RPM = 1500.

Fuel	RPM [obr/min]	Load [N.m]	Average of CO [g/kWh]	Mean Square of CO	MSE_CO
S75	1500	50 to 200	0,3946	0,1557	**0,2220**
T75			0,5469	0,2991	
R75			0,5392	0,2908	
S50			0,1138	0,0129	RMSE_CO
T50			0,2305	0,0532	**0,4712**
R50			0,2675	0,0715	
S25			0,4474	0,2002	
T25			0,7435	0,5527	RSE_CO
R25			0,6126	0,3753	**0,6864**
Diesel			0,4570	0,2089	

The lower and upper mean CO values are 0,1138 and 0,7435 respectively and correspond to the overall mean values of S50 and T25 fuels taken over the entire CO column. The mean square error (MSE_CO), root mean square error (RMSE_CO) and mean square error (RSE_CO) values shown in Table 19 are 0,2220, 0,4712 and 0,6864 respectively.

Table 20. MSE, RMSE and RSE Values of CO_2 Gas with RPM = 1500.

Fuel	RPM [obr/min]	Load [N.m]	Average of CO2 [g/kWh]	Mean Square of CO2	MSE_CO2
S75	1500	50 to 200	79,8856	6381,7133	**12336,5485**
T75			91,5795	8386,8011	
R75			103,1476	10639,4318	
S50			105,3574	11100,1822	RMSE_CO2
T50			134,2796	18031,0157	**111,0700**
R50			**136,5970**	18658,7357	
S25			95,8536	9187,9194	
T25			104,2086	10859,4328	RSE_CO2
R25			121,7825	14830,9891	**10,5390**
Diesel			123,6498	15289,2644	

The lower and upper mean CO_2 values are 79,8856 and 136,5670 respectively and correspond to the overall mean values of S50 and T25 fuels taken over the entire CO_2 column. The mean square error (MSE_CO_2), root mean square error (RMSE_CO) and mean square error (RSE_CO_2) values shown in Table 20 are 12336.5485, 111,0700 and 10.5390 respectively.

Table 21. MSE, RMSE and RSE Values of NOx Gas with RPM = 1500.

Fuel	RPM [obr/min]	Load [N.m]	Average of NOX [g/kWh]	Mean Square of NOX	MSE_NOX
S75	1500	50 to 200	-0,5935	0,3522	1,5211
T75			-1,3734	1,8861	
R75			-1,2986	1,6864	
S50			-1,2405	1,5388	RMSE_NOX
T50			-1,0221	1,0448	1,2333
R50			-1,0260	1,0526	
S25			-1,7310	2,9964	
T25			-1,2670	1,6052	RSE_NOX
R25			-1,3658	1,8654	1,1106
Diesel			-1,0877	1,1830	

The lower and upper mean NOx values are -1,7310 and -0,5935 respectively and correspond to the overall mean values of S25 and S75 fuels taken over the entire NOx column. The mean square error (MSE_NOx), root mean square error (RMSE_NO_x) and

mean square error (RSE_NO_x) values shown in Table 21 are 1.5211, 1.2333 and 1.1106 respectively.

Table 22. MSE, RMSE and RSE Values of THC/dry Gas with RPM = 3000.

Fuel	RPM [obr/min]	Load [N.m]	Average of THC/dry [g/kWh]	Mean Square of THC/dry	MSE_ THC/dry
S75	3000	50 to 200	0,36302	0,13178	0,17301
T75			0,50729	0,25734	
R75			0,43071	0,18551	
S50			0,32417	0,10508	RMSE_ THC/dry
T50			0,38889	0,15123	0,4159
R50			0,35649	0,12709	
S25			0,49441	0,24444	
T25			0,40820	0,16662	RSE_ THC/dry
R25			0,46115	0,21266	0,6449
Diesel			0,38514	0,14834	

For the RPM equal to 3000, the lower and upper mean THC/dry values are 0.32417 and 0.50729 respectively and correspond to the overall mean values of S50 and T75 fuels taken over the entire THC/dry column. The mean square error (MSE_THC/dry), root mean square error (RMSE_THC/dry) and mean square error (RSE_THC/dry) values shown in Table 22 are 0.17301, 0.4159 and 0.6449 respectively.

Table 23. MSE, RMSE and RSE Values of CO Gas with RPM = 3000.

Fuel	RPM [obr/min]	Load [N.m]	Average of CO [g/kWh]	Mean Square of CO	MSE_ CO
S75	3000	50 to 200	2,95109	8,70896	12,04244
T75			4,03135	16,25182	
R75			3,74677	14,03825	
S50			3,45841	11,96060	RMSE_ CO
T50			3,43616	11,80721	3,4702
R50			3,72018	13,83974	
S25			3,74519	14,02645	
T25			3,21761	10,35303	RSE_ CO
R25			3,40696	11,60736	1,8629
Diesel			2,79839	7,83100	

The lower and upper mean CO values are 2.95109 and 4.03135 respectively and correspond to the overall mean values of S75 and T75 fuels taken over the entire CO column. The mean square error (MSE_CO), root mean square error (RMSE_CO) and mean square error (RSE_CO) values shown in Table 23 are 12.4244, 3.4702 and 1.8629 respectively.

Table 24. MSE, RMSE and RSE Values of CO_2 Gas with RPM = 3000.

Fuel	RPM [obr/min]	Load [N.m]	Average of CO2 [g/kWh]	Mean Square of CO2	MSE_ CO2
S75	3000	50 to 200	153,58598	23588,65326	39898,9737
T75			181,69605	33013,45573	
R75			192,21683	36947,30794	
S50			205,06427	42051,35494	RMSE_ CO2
T50			190,63540	36341,85595	199,7473
R50			203,78953	41530,17419	
S25			241,27172	58212,04251	
T25			191,62620	36720,59973	RSE_ CO2
R25			197,27364	38916,89058	14,1332
Diesel			227,30465	51667,40208	

The lower and upper mean CO_2 values are 153.58598 and 241.27172 respectively and correspond to the overall mean values of S75 and S25 fuels taken over the entire CO_2 column. The mean square error (MSE_CO_2), root mean square error (RMSE_CO_2) and

mean square error (RSE_CO_2) values shown in Table 24 are 39898.9737, 199.7473 and 14,1332 respectively.

Table 25. MSE, RMSE and RSE Values of NOx Gas with RPM = 3000.

Fuel	RPM [obr/min]	Load [N.m]	Average of NOX [g/kWh]	Mean Square of NOX	MSE_NOX
S75	3000	50 to 200	-0,12145	0,01475	0,2504
T75			-0,80859	0,65382	
R75			-0,57240	0,32764	
S50			-0,64982	0,42227	RMSE_NOX
T50			-0,53987	0,29146	0,5004
R50			-0,69735	0,48630	
S25			-0,22521	0,05072	
T25			-0,10986	0,01207	RSE_NOX
R25			-0,48737	0,23753	0,7074
Diesel			0,08819	0,00778	

The lower and upper mean NO_x values are -0.80859 and 0,08819 respectively and correspond to the overall mean values of T75 and Diesel fuels taken over the entire NO_x column. The mean square error (MSE_NO_x), root mean square error (RMSE_NO_x) and mean square error (RSE_NOx) values shown in Table 25 are 0.2504, 0.5004 and 0,7074 respectively.

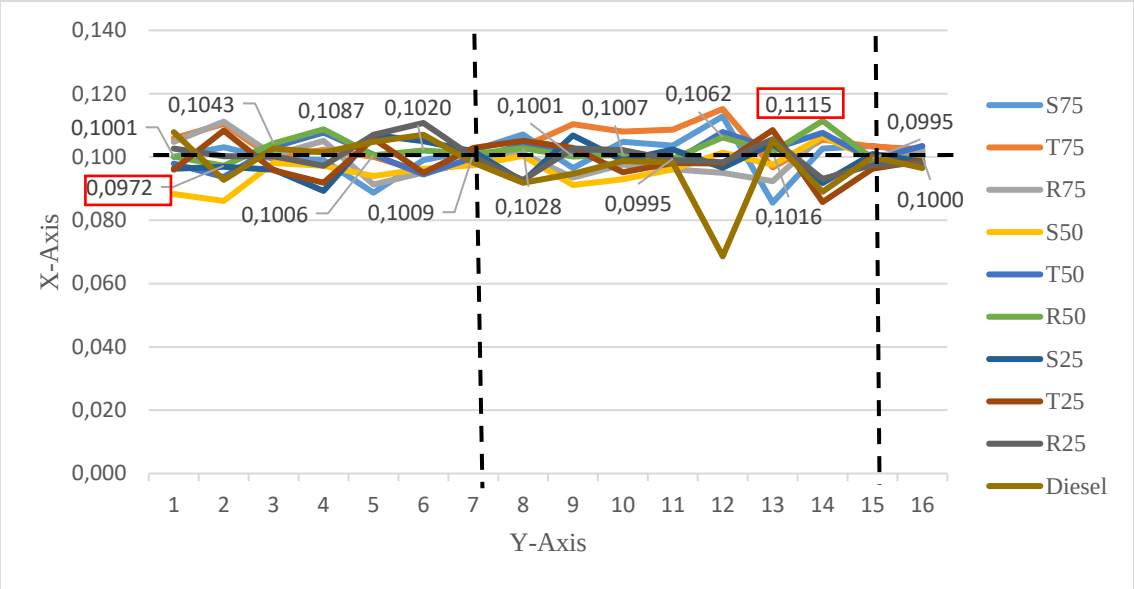
For RPM's equal to 1500 and 3000, the best-performing gases are respectively both THC/dry about the lower values of MSE, RMSE and RSE obtained during the internal combustion engine experiment.

4.4 Results and Analysis

We have completed various tables relevant to the case study while defining various computational techniques and methods performed to solve the tasks mentioned in subtitle 4.1 from the analysis procedures previously used in subtitles 4.2 and 4.3.

Then, the phase is more focused on the results obtained in the calculations carried out from the data collected in the previous subtitle 4.3, which will be used for a detailed analysis to determine the most appropriate fuel for the recommended consumption based on the different gases produced during the experiments carried out in the internal combustion engine.

Table 26. Weighted Sums and Normalized Weights for each Emission with RPM = 1500.



For the RPM equal to 1500, R50 biodiesel appears strongly correlated, stable and more consistent with less curvature above the 0.1 point of the y-axis compared to the other fuels, except for the more visible upper point of 0.1115. Then, the value 0.1 on the Y-axis considered the reference is the linear stability scale of this graph used for observing the shape of the different curves that run along the line

Most of the observed points of the R50 curve are close to 0.1 points of the ordinate except some points have values bigger than 0.1 like (0.1087, 0.1028, 0.1062 and 0.1048…) and are considered as the highest points of the R50 curve.

Within the graph range comprised between 7 to 15 at the X-axis, all fuel curves convergent to these points and represent a frame. In the range of the graph on the X-axis, all fuel curves converge to two points 7 and 15 respectively, because the weight values are equal to 0.0001. Both points represent the limits of the frame where the

progression and shape of the curves are observed. These points represent the evolution and shape of the observed curves due to the decrease in the number of predictions appearing with significant uncertainties, and leads to increased overlap of the curves.

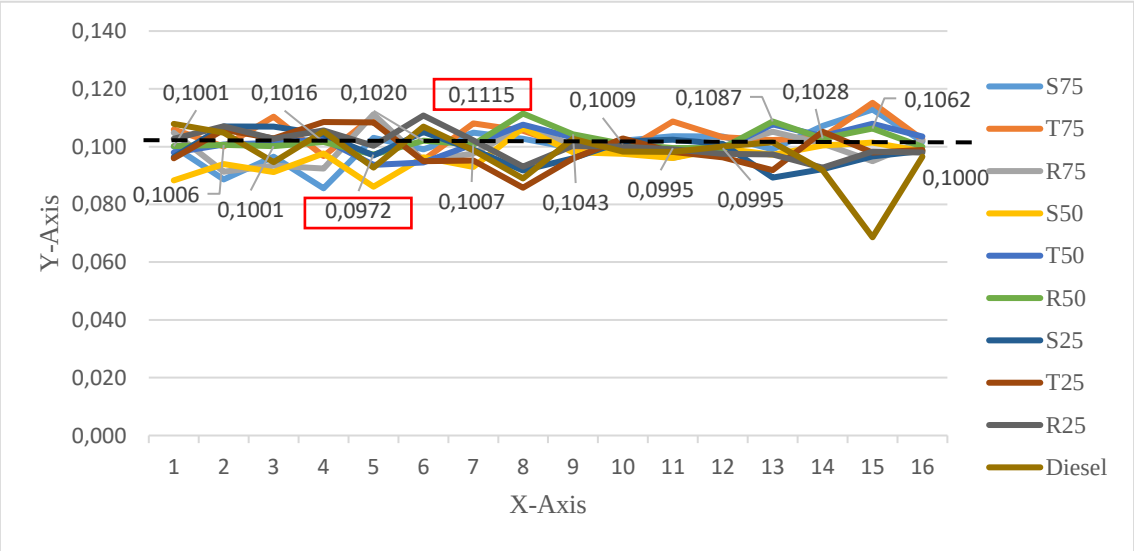
Table 27. Weight Values of CO Gas from the different Fuels used for the Experiments with RPM = 1500.

Gas [g/kWh]	All Weighted Sum	S50	T50	R50
CO	30,5934	0,1064	0,1076	0,1115
CO	39,2281	0,0861	0,0937	0,0972
CO	7,5973	0,0930	0,1008	0,1007
CO	13,1501	0,0963	0,0945	0,1020

Both point values (0,0972 and 0,1115) of the R50 curve converge respectively through both X-axis points, are used to identify the origin of two point values belong to the R50 curve from the weight values of the emitted gases CO shown in Table 27 above.

We observe the concept of interpolation used to predict the point values of emitted gases from the concentration of fuel curves on the 0,1 value Y-axis within a frame range [7 to 15] designed in the graph, which is described a regression model.

Table 28. Weighted Sum Fuel Values for RPM equal to 1500.



For the RPM equal to 1500, R50 biodiesel appears strongly correlated, stable and more consistent with less curvature above the 0.1 point of the y-axis compared to the other fuels, except for the more visible upper point of 0.1115. Then, the value 0.1 on the Y-axis considered the reference is the linear stability scale of this graph used for observing the shape of the different curves that run along the line

Most of the observed points of the R50 curve are close to 0.1 point of the ordinate except some points have values bigger than 0.1 like (0.1016, 0.1020, 0.1062 and 0.1087) and are considered as the highest points of the R50 curve.

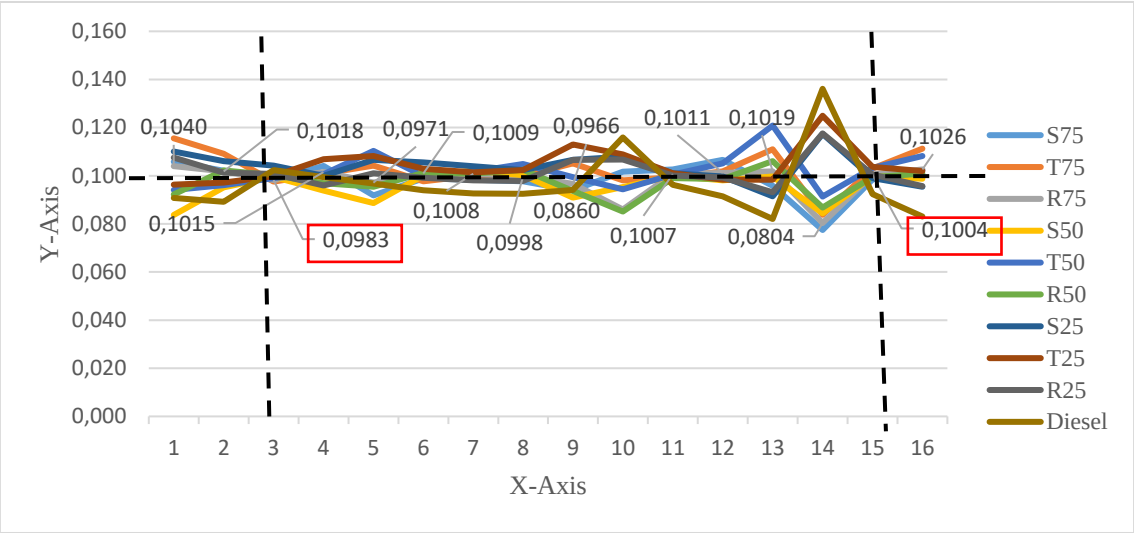
Within the graph range comprised between 1 to 16 at the X-axis, all fuel curves diverge throughout various points comprises in the frame which the progression and shape of the curves are observed. These points represent the evolution and shape of the observed curves due to the decrease in the number of predictions appearing with significant uncertainties, and leads to increased overlap of the curves.

Table 29. Weight Values of CO Gas from the different Fuels used for the Experiments with RPM =1500.

Gas [g/kWh]	All Weighted Sum	S50	T50	R50
CO	30,5934	0,1064	0,1076	0,1115
CO	39,2281	0,0861	0,0937	0,0972
CO	7,5973	0,0930	0,1008	0,1007
CO	13,1501	0,0963	0,0945	0,1020

Both point values (0,0972 and 0,1115) of the R50 curve converge respectively through both X-axis points, are used to identify the origin of two point values belong to the R50 curve from the weight values of the emitted gases CO shown in Table 29 above.

Table 30. Weighted sums and Normalized Weights for each Emission with RPM = 3000.



For the RPM equal to 3000, R75 biodiesel appears strongly correlated, stable and more consistent with less curvature above the 0.1 point of the y-axis compared to the other fuels, except for the more visible upper point of 0.1040. Then, the value 0.1 on the Y-axis considered the reference is the linear stability scale of this graph used for observing the shape of the different curves that run along the line

Most of the observed points of the R75 curve are close to 0.1 points of the ordinate except some points have values less than 0.1 like (0.0971, 0.0966, 0.0860 and 0.0804) and are considered as the smallest points of the R75 curve.

Within the graph range comprised between 3 to 15.2 at the X-axis, all fuel curves convergent to these points and represent a frame. In the range of the graph on the X-axis, all fuel curves converge to two points 3 and 15.2 respectively, because the weight values are equal to 0.0001. Both points represent the limits of the frame where the evolution and shape of the curves are observed due to the increase in the number of predictions appearing with the significant reduction in uncertainties, and leads to reduced overlap of the curves.

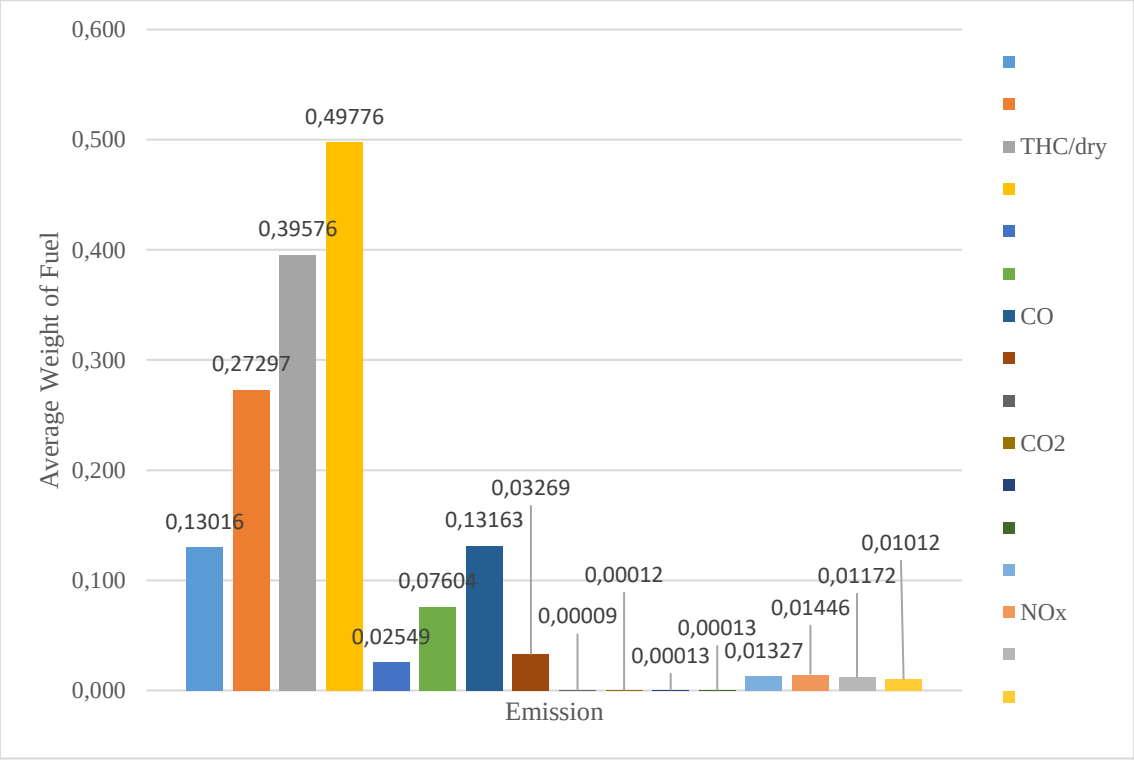
Table 31. Weight Values of CO2 Gas from the different Fuels used for the Experiments with RPM =1500.

Gas [g/kWh]	All Weighted Sum	S75	T75	R75	S50	T50	T25
CO2	8935,8624	0,0985	0,1031	0,1004	0,0998	0,1033	0,1037
CO2	9295,8509	0,1026	0,0993	0,1007	0,0999	0,1004	0,1010
CO2	13471,4313	0,0983	0,0976	0,0983	0,0996	0,0990	0,0999
CO2	9567,5544	0,0984	0,1001	0,1008	0,1021	0,1017	0,1014

Both point values (0,0983 and 0,1004) of the R75 curve converge respectively through both X-axis points, are used to identify the origin of two point values belong to the R75 curve from the weight values of the emitted gas CO_2 shown in Table 31 below.

We observe the concept of interpolation used to predict the point values of emitted gases from the concentration of fuel curves on the 0,1 value Y-axis within a frame range [3 to 15.2] designed in the graph, which is described a regression model.

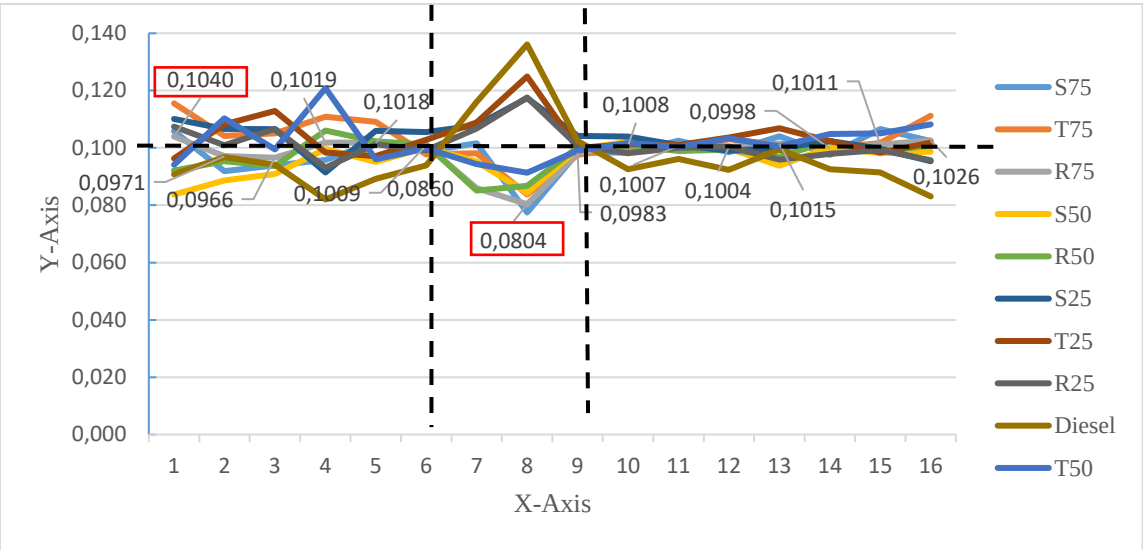
Table 32. Average Weight of Fuels with the RPM is equal to 1500.



Following the average weight values of each gas emitted during the experiments carried out, it appears that CO_2 gas performed better with lower weight values (0.0009 to 0.00013) compared to the other gases also used during these experiments.

Then, the THC/dry gas obtained has the highest mean weight values (0.27297 to 0.49776) compared to the different gases. It is worth mentioning that THC/dry gas was the most polluting despite the presence of other more toxic and harmful gases in the environment such as CO and NOx, which are less performed with lower average weight values (0.01012 to 0.01446) for NOx.

Table 33. Weighted Sum Fuel Values with the RPM is equal to 3000.



For the RPM equal to 3000, R75 biodiesel appears strongly correlated, stable and more consistent with less curvature above the 0.1 point of the y-axis compared to the other fuels, except for the more visible upper point of 0.1040. Then, the value 0.1 on the Y-axis considered the reference is the linear stability scale of this graph used for observing the shape of the different curves that run along the line

Most of the observed points of the R75 curve are close to 0.1 points of the ordinate except some points have values less than 0.1 like (0.0910, 0.0966, 0,0998, 0.0983 and 0.0860) and are considered as the smallest points of the R75 curve.

Within the graph range comprised between 6 to 9.2 at the X-axis, all fuel curves convergent to these points and represent a frame. In the range of the graph on the X-axis, some fuel curves converge to the point 6, because the weight values are equal to 0.0001. Otherwise, all fuel curves converge to point 9,2. Both points represent the limits of the frame where the evolution and shape of the curves are observed due to the increase in the number of predictions appearing with the significant reduction in uncertainties, and leads to reduced overlap of the curves.

Table 34. Weight Values of CO Gas from the different Fuels used for the Experiments with RPM = 3000.

Gas [g/kWh]	All Weighted Sums	S75	T75	R75	S50
CO	16,7294	0,0775	0,0838	0,0804	0,0843
CO	95,8432	0,1021	0,1091	0,1018	0,0951
CO	14,5610	0,1016	0,0981	0,0860	0,0953
CO	14,9901	0,0989	0,0976	0,1009	0,1001

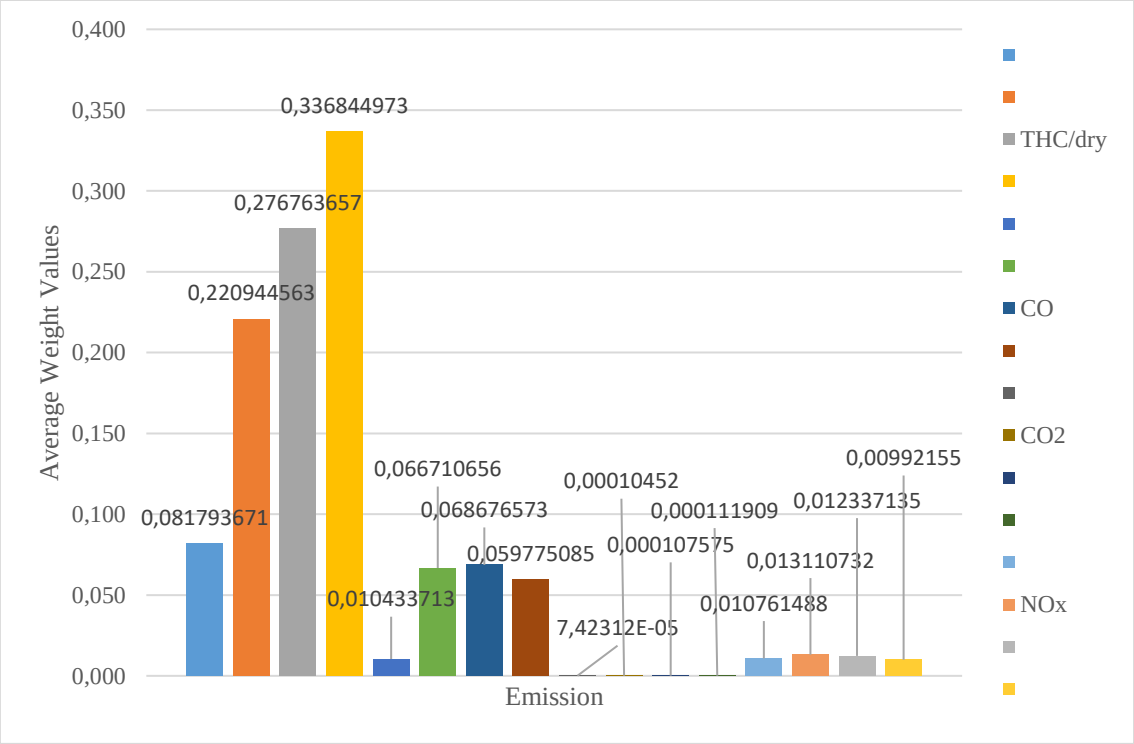
Tables 34 and 35 are derived from calculations of weighted sum values and average weight values performed with RPM equal to 3000.

Table 35. Weight Values of THC/dry Gas from the different Fuels used for the Experiments with RPM = 3000.

Gas [g/kWh]	All Weighted Sum	S75	T75	R75	S50
THC/dry	12,2258848	0,1058	0,1155	0,1040	0,0837
THC/dry	2,9687247	0,0958	0,1109	0,1019	0,0996
THC/dry	3,6131912	0,0940	0,1051	0,0966	0,0909
THC/dry	4,5260222	0,0919	0,1041	0,0971	0,0887

Both point values (0,0804 and 0,1040) of the R75 curve converge respectively through both X-axis points, are used to identify the origin of two point values belong to the R75 curve from the weight values of the emitted gases CO and THC/dry shown in Table 35 above.

Table 36. Average Weight for Fuel with the RPM equal to 3000.



Following the average weight values of each gas emitted during the experiments carried out, it appears that CO_2 gas performed better with lower weight values (7.4231E-05 to 0.000112) compared to the other gases also used during these experiments.

Then, the THC/gas obtained has the highest mean weight values (0.081794 to 0.336845) compared to the different gases. It is worth mentioning that THC/dry gas was the most polluting despite the presence of other more toxic and harmful gases in the environment such as CO and NOx, which are less performed with lower average weight values (0.009922 to 0.01311) for NOx.

4.5 Statistical Process using Monte Carlo Method and AHP Analysis

The Monte Carlo method is used to predict more possible values derived from the calculation of different mean and standard deviation values performed in the table, which are randomized by 2000 iterations in the simulation column to select approximate

values close to the estimated values appearing in the results table column and be verified with other previously collected values.

Then, the weight values obtained and selected from the Monte Carlo simulation will be compared with the result obtained from the AHP analysis performed.

4.5.1 Monte Carlo Method

We performed various mean and standard deviation calculations to obtain the Monte Carlo results from both load values 1500 and 3000. However, we created a simulation with 2000 iterations related to the result column and we chose the first value of the Monte Carlo column.

Table 37. Monte Carlo Simulation of Fuel Values.

Monte Carlo	Cumulative	Simulation	Outcome
90,467742	2,55095E-28	1	-1,114345
141,519978	1	2	2,889358
-12,846134	1	3	6,023494
-4,304667	2,4724E-15	4	6,335798
16,338580	1	5	-4,510138
19,756545	0,999999965	6	-4,848658
63,296137	1	7	14,562751
-140,288205	1	8	7,564713
-22,924170	1	9	5,993694
-52,890265	1	10	1,312272
0,355659	1	11	-11,331374
-2,909698	0,063949281	12	-11,941458
15,050954	2,02812E-51	13	-1,690076
62,521811	1	14	11,480334
28,909397	1	15	-5,265950

Then, we calculated the operation to get the results shown in the results column which are connected to the table displaying the mean, standard deviation Monte Carlo probability, MSE, RMSE and RSE values.

Table 38. MSE, RMSE, RSE and Probability Simulation Values with RPM = 1500.

Probability	Mean	St Deviation	MSE	RMSE	RSE
48 %	-42,28486	1185,12970	1788,00922	42,28486	6,50268
52 %	1,23717	1139,31351	1,53060	1,23717	1,11228

Simulation results obtained by different RPMs 1500 and 3000 respectively.

Table 39. MSE, RMSE, RSE and Probability Simulation Values with RPM = 3000.

Probability	Mean	St Deviation	MSE	RMSE	RSE
48 %	-3,48496	1134,08097	12,14498	3,48496	1,86681
52 %	13,48254	1122,93835	181,77888	13,48254	3,67186

The mean, standard deviation and probability values obtained from the simulation results of 2000 iterations are selected and used to be verified by the different MSE, RMSE and RSE values previously collected in Table 40 below, which is the synthesis from the different Tables 18 to 25.

Table 40. Summary of MSE, RMSE and RSE Collected Values from Tables 18 to 25.

Gas [g/kWh]	RPM [obr/min]	MSE	RMSE	RSE
THC/dry	1500	0,067319069	0,259459186	0,509371364
CO		0,222035472	0,471206401	0,686444754
CO2		12336,54854	111,0700164	10,53897606
NOX		1,521098186	1,233328093	1,110553057
THC/dry	3000	0,173009828	0,415944501	0,644937595
CO		12,04244261	3,470222271	1,862853261
CO2		39898,97369	199,7472745	14,13319761
NOX		0,250432855	0,500432668	0,707412657

Following different probabilities, MSE, RMSE, and RSE values were obtained by simulating 2000 iterations. In the table above, we have found the approximate values that refer to the emission. However, we realized that the CO emission is slightly higher than the NOx emission $(1{,}1105 \leq RSE \leq 1{,}8628)$, based on the different calculations and results obtained in Tables 23, 25, 32, 34 and 36, it appears that the S75 and T75 fuels perform better than the other fuels during the experiments carried out.

4.5.2 Decision Tree Concept applied from Monte Carlo Results

The decision tree is built based on the probabilities and RSE values obtained from different Monte Carlo simulation results obtained from 2000 iterations, which the values are verified and confirmed from these different probabilities and RSE values are collected and summarised in Table 40 above.

Table 41. Summary of MSE, RMSE and RSE Collected Values of different RPMs.

Decision Making	RPM [obr/min]	Probability	MSE	RMSE	RSE
Decision 1	1500	52%	1,5210982	1,2333281	1,1105531
Decision 2	3000	48%	12,042443	3,4702223	1,8628533

Construction of a decision tree from the Monte Carlo simulation values, in which decision making, RPM, decisions, probabilities and RSE values are considered in the process and belong to Table 41 above.

Table 42. The Decision Tree Conceptual Table based on Table 41.

	Decision-making	
RPM	1500	3000
Probability	0,52	0,48
Decision 1	1,5211	12,0424
Decision 2	1,1105	1,8628

Table 42 is the decision tree model based on the summary of Table 4, which uses RPM, probabilities, MSE and RSE values in the conceptual composition.

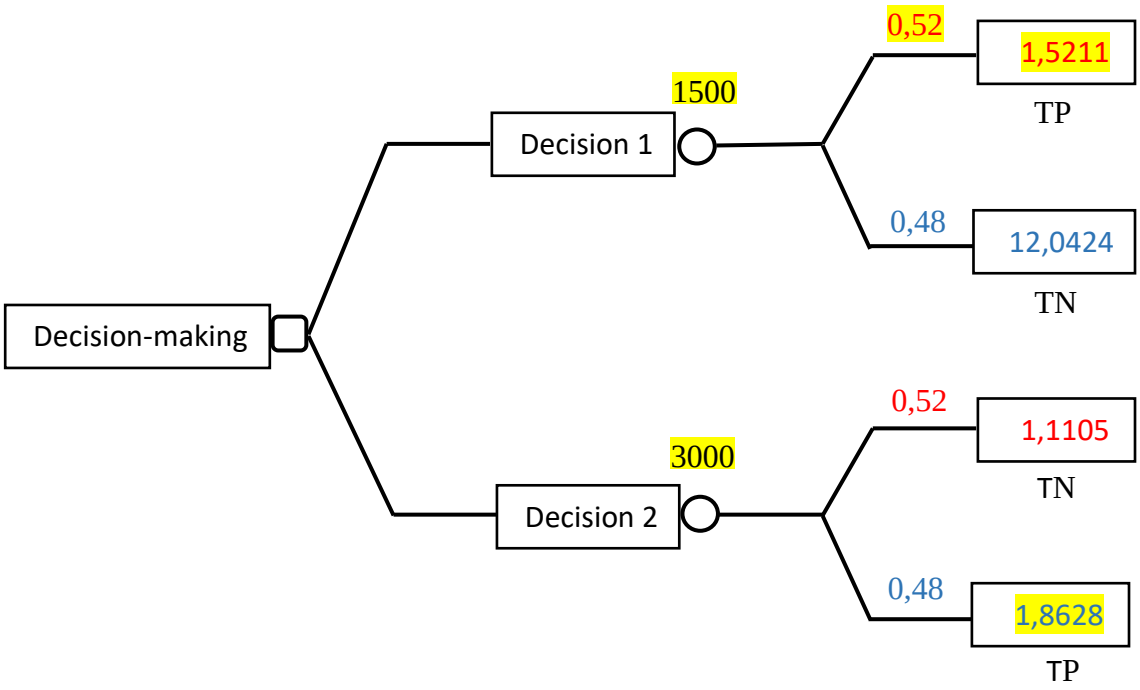


Figure 21. Decision-tree designed from the content of Table 41.

Following subtitle 3.2.2 regarding the percentage of predictions and the number of correct predictions, we conclude that decision 1 with a probability equal to **0.52** and decision 2 with an RSE equal to **1.8628** are perfect predictions for the previously chosen S75 fuel, based on different rankings observed in Tables 14 and 15, as well as different calculations performed in Tables 23, 25, 32, 34 and 36. Furthermore, the verification of the different values performed in Table 40 confirms the choice of S75 fuel.

4.5.3 AHP Analysis

We define the pairwise comparison scale model used for AHP analysis from the numerical scale and explanation shown in Table 42 below.

Table 43. Pairwise Comparison Scale Model applied for the AHP Analysis.

Numerical Scale	Explanation
1	Equal
3	Moderately more important
5	Strongly more important
7	Thought or demonstrated to be much more important
9	Demonstrated to be of much more important

However, we listen to the scenario and description based on the pairwise comparison scale model that belongs to Table 44 below.

Table 44. Scenario and Description for the Pairwise Comparison Scale Model.

Scenario	Description
A	All 1's
B	Engine Emission parameters at 0

Then, we built the engine emissions table in order to calculate the sum of emissions from the numerical scale and the different scenarios previously defined for each column of the table.

Table 45. Sum of each Engine Emission Column from different Scenarios.

		Engine Emissions			
		1	2	3	4
		THC/dry	CO	CO2	NOX
1	THC/dry	1,00	0,20	0,33	0,11
2	CO	5,00	1,00	3,00	0,20
3	CO2	3,00	0,33	1,00	0,14
4	NOX	9,00	5,00	7,00	1,00
	Sum of Columns	18,00	6,53	11,33	1,45

The table values are realized using Pascal's triangle model. Thus, the sum of each emission column is calculated from the sum of the emissions belonging to the gas column. However, the normalized pairwise matrix is performed based on each emission in the column as shown in Table 44 divided by the sum of the column to which it belongs.

Table 46. Normalized Pairwise Matrix for Engine Emissions.

	THC/dry	CO	CO2	NOX	Criteria Weight Value
THC/dry	0,056	0,031	0,029	0,076	0,048
CO	0,278	0,153	0,265	0,138	0,208
CO2	0,167	0,051	0,088	0,098	0,101
NOX	0,500	0,765	0,618	0,688	0,643

We calculated the weight value of the criteria for each emission from the sum of the values for each emission belonging to each row indicated in Table 44 below.

Table 47. Consistency of Criteria Weight of Fuel Values.

	THC/dry	CO	CO2	NOX	Weighted Sum Value
THC/dry	0,048	0,042	0,034	0,071	0,195
CO	0,240	0,208	0,303	0,129	0,880
CO2	0,144	0,069	0,101	0,092	0,406
NOX	0,432	1,041	0,707	0,643	2,823

The consistency of the criteria weight value is obtained from the multiplication of each emission in Table 44 by each criteria weight value belonging to the emission row.

Table 48. Consistency Ratio of Fuel Values.

Weighted Sum	Criteria Weights	Weight Ratio	Lambda Max λ	4,174009
0,195	0,048	4,057214	Consistency Index	0,058003
0,880	0,208	4,224904	Consistency Ratio	0,064448
0,406	0,101	4,020822	With $n = 4$	
2,823	0,643	4,393094		

The lambda maximum $\boldsymbol{\lambda}$ is obtained from the average of the weight ratio, which is calculated from the ratio of the weighted sum value to the criterion weight value of each emission row. Then, the consistency index belongs to the factor of the lambda maximum value minus the rank of matrix n divided by the factor of rank of matrix minus one. The value of the consistency ratio is obtained from the ratio of the consistency index and the

random index as shown in Table 6. Furthermore, the illustration of AHP analysis is demonstrated in more detail in subtitle 3.3.2.

Following the value of the consistency ratio obtained from the calculation of the different emission weight values shown in Table 46 and compared to the different average weight values for each emission shown in Tables 12 and 13. It appears that the value of the consistency ratio is equal to **0.0645** considered as a weight value and close to **0.0667** the weight value of CO gas with the RPM equal to **3000** respecting the acceptable cut-off consistency index at 10% equal to **0.090** for $n = 4$ indicated in Table 7. This confirms the choice of S75 fuel already approved by the previous verifications carried out in the various subtitles 4.5.1 and 4.5.2.

4.6 Study Limitation

The implementation analysis is based on the data collected from the experimental modelling data performed on the internal combustion engine, whose study of engine characteristics and fuel properties are not considered in our main study of data calculation.

However, these features are the key vectors that play a major role in adjusting the engine during machine operation in terms of speed, engine power and performance related to energy deployment to ensure considerable efficiency during the experiments performed.

Furthermore, the technical characteristics of the engine and the composition of the fuel influence the adjustment of engine maintenance and the performance of the operations carried out are not considered in the AHP process during the implementation analysis.

4.7 Findings

Some results of weighted average calculations carried out present uncertainties due to the weakness values obtained during the experiments. The statistical models used in the

data analysis show deficiencies at certain points of the graph with average weighted values extremely low, which makes it difficult to built up and describe the shape of the curve and to provide a favourable opinion on its progression.

Furthermore, this situation reduces predictions while creating more uncertainties that harm the interpretation of the shape of each plotted curve. Nevertheless, a considerable error is observed from the moment when the curves are not separated from each other in its interpretation, and leads to increased overlap of the curves.

Definitely, a lack of precision is observed due to some low weight values revealing uncertainties while creating biases in the repetition of these data, which constantly appear in terms of probability capable of influencing with more precision the exact analysis of the calculations carried out by the Monte Carlo method associated with the AHP.

4.8 Recommendation for Process Improvement

The analysis experiments could be repeated using large amounts of data to perform the calculation values of gas and fuel weights from neural networks in other software environments like Matlab or R-Studio with perfect analysis of the plotted curves while taking into account other data of the engine technical characteristics and fuel composition.

5 Conclusion

The data-driven approach is the essential part integrated into the decision-making process, which uses different calculation techniques and methodologies that are performance vectors influencing the decision at several levels from the different defined stages.

The statistical techniques used during the implementation phase play a powerful role in analyzing the different data sets collected from the descriptive phase by extracting biases while reducing uncertainties and improving the solutions to provide processing tasks related to the timeliness and knowledge of executing the perfect goal defined by saving time and reducing the costs of information management, which include decision making.

The observed performance in decision-making is based on different MADM or MODM methodologies taken into account to emphasize the different calculations and analyses performed by the statistical concept to evaluate different data in predicting events. Then, these concepts integrate the structured process capable of modifying or solving the problems of the descriptive model considered more effective which uses algorithmic techniques to make precise predictions during the analyses carried out with complex data sets presenting multiple variables.

The integration of the Monte Carlo method into the AHP provides the appropriate visualization in the optimal configuration approach capable of extending the use of analytical processes in which different performed simulations illustrate the powerful interaction of the analyses to examine the solutions in close detail from multi-criteria perspectives during the experiments. Therefore, this approach will optimize future research with considerable support based on the multiple opportunities offered by neural networks capable of developing a large classification of fuel ranking related to engine characteristics and fuel composition from big data collected during internal combustion engine experiments.

References

Albayrak E. & Erensal Y. C. (2004). Using analytic hierarchy process (AHP) to improve human performance. An application of multiple criteria decision making problem, Journal of Intelligent Manufacturing, Vol. 15, pp. 491-503.

Akash B. A., Mamlook R., and Mohsen M. S. (1999). Multi-criteria selection of electric power plants using analytical hierarchy process. Electr. Power Syst. Res., vol. 52, no. 1, pp. 29–35.

Akincilar A. & Dagdeviren M. (2014). A hybrid multi-criteria decision-making model to evaluate hotel websites. Int. J. Hosp. Manag., vol. 36, pp. 263–271.

Alyoubi B.A. (2015). Decision support system and knowledge-based strategic management. *Procedia Computer Science*, Vol. 65. pp.278–284.

Alter S. (1980). Decision support systems: current practice and continuing challenges. Addison-Wesley Publ.

Al-Najjar B. & Alsyouf I. (2003). Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making. Int. J. Prod. Econ., vol. 84, no. 1, pp. 85–100.

Aouadni I. & Rebai A. (2017). Decision support system based on genetic algorithm and multi-criteria satisfaction analysis (MUSA) method for measuring job satisfaction. Annals of Operations Research. Vol. 256, pp. 3–20.

Berger J.O. (1985). Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

Black, P., & Stockton, T. (2009). Basic steps for the development of decision support systems. Decision Support Systems for Risk-Based Management of Contaminated Sites. NY: Springer US. pp. 1 - 27.

Brillinger D.R. (1981, 2001). Time Series: Data Analysis and Theory, 2nd ed. San Francisco: Holden-Day. Republished in 2001 by the Society for Industrial and Applied Mathematics, Philadelphia.

Brunswicker S., Bertino E., & Matei S. (2015). Big Data for Open Digital Innovation - A Research Roadmap. *Big Data Research*, Vol. 2, pp. 53-58.

Burstein F., & Holsapple C. W. (2008). Handbook on Decision Support Systems 1: Basic Themes. International Handbooks on Information Systems. Heidelberg: Springer.

Cao L. (2017). Data science: A comprehensive overview. ACM Comput. Surv. Vol.50(3), pp. 1–42.

Carlsson S. A. & El Sawy O. A. (2008). Decision Support in Turbulent and High-Velocity Environments. Handbook on Decision Support Systems Heidelberg: Springer. Vol. 2, pp. 3- 17.

Cervone H.F. (2016). Informatics and data science: an overview for the information professional. *Digital Library Perspectives*, Vol. 32 No. 1, pp. 7-10

Chang R.M., Kauffman, R.J., Kwon, Y. (2014). Understanding the Paradigm Shift to Computational Social Science in the Presence of Big Data. *Decision Support Systems,* Vol. 63, pp. 67-80.

Durcevic S. (2019). Why Data Driven Decision Making is Your Path to Business Success. *Business Intelligence.*

Elgendy N. & Elragal A. (2014). Big Data Analytics: A Literature Review Paper. Advances in Data Mining: Applications and Theoretical Aspects. *Springer International Publishing.* Vol. 8557, pp. 214-227.

Elgendy N. & Elragal A. (2016). Big Data Analytics in Support of the Decision-Making Process. Procedia Computer Science, Vol. 100, pp. 1071–1084.

Elgendy N., Elragal A., Päivärinta T. (2021): DECAS: A Modern Data-Driven Decision Theory for Big Data and Analytics, Journal of Decision Systems. Taylor & Francis Group.

Elmusrati, M. (2020). Lecture notes: Machine learning course. University of Vaasa (Moodle).

Fan S., Lau R., Zhao J.L. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. Big Data Research, Vol.2, pp. 28-32.

Fayyad U, Piatetsky-Shapiro G, Padhraic S. (1996). Form Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, pp. 37-54.

Fewell S. & Clark T. (2005). Organisational interoperability: evaluation and further development of the OIM model. International Command and Control Research and Technology Symposium (ICCRTS). Edinburgh, pp. 2-42.

Fisher D., DeLine R., Czerwinski M., Drucker S. (2012). Interactions with Big Data Analytics. ACM Interactions, Vol.19 (3), pp.50 - 59.

Frank R. J, Davey N., and Hunt S.P. (2001). Time Series Prediction and Neural Networks. Journal of Intelligent and Robotic Systems, Vol. 31(1-3), pp. 91-103.

Frantz R. & Simon H. (2003). Artificial intelligence as a framework for understanding intuition. Journal of Economic Psychology, Vol. 24(2), pp. 265-277.

Frey J. W. (2006). What Does Distributed Operations Mean for Joint Air Fire Support? (Master's Thesis) US Naval War College, Department of Joint Military Operations. Defence Technical Information Centre.

Hansson S.O. (1994). Decision theory. A brief introduction. In Department of philosophy and the history of technology. Stockholm: Royal Institute of Technology.

Gaynor M., Seltzer M., Moulton S., & Freedman J. (2005). A dynamic, data-driven, decision support system for emergency medical services. Computational Science - ICCS 2005, 5th International Conference Atlanta GA, USA Proceedings. Vol. II, pp. 703-711

Gigerenzer G. & Gaissmaier W. (2015). Decision making: Nonrational theories. International encyclopedia of the social & behavioral sciences. Elsevier 2nd. Vol. 5, pp. 911–916.

Graboś R. (2004). A qualitative model of decision making. in International Conference on Artificial Intelligence: Methodology, Systems, and Applications, Berlin, Heidelberg: Springer, pp. 480–489.

Henry Gray, (1858) Anatomy : descriptive and surgical, 1st edition.

Gregor S. (2006). The nature of theory in information systems. MIS Quarterly, Vol. 30(3), pp. 611–642.

Grover & Lyytinen. (2015). New state of play in information systems research: The push to the edges. MIS Quarterly, Vol. 39(2), pp. 271–296.

Gorener, A. (2012). Comparing AHP and ANP: An application of strategic decisions making in a manufacturing company. International Journal of Business and Social Science, 3(11), 194–208.

Guitouni A. (2009). A time sensitive decision support system for crisis and emergency management. Information systems and technology panel (IST) symposium, pp. 13-15.

Haddad M. & Sanders D. (2018). Selection of discrete multiple criteria decision-making methods in the presence of risk and uncertainty," Oper. Res. Perspect., vol. 5, pp. 357–370.

Han J., Pei J., Kamber M. (2011). Data Mining: Concepts and Techniques. Amsterdam: Elsevier.

Hansson S. O. (2005). Decision theory. A brief introduction. Royal Institute of Technology, Department of Philosophy and the History of Technology. Stockholm: Royal Institute of Technology.

Hariri R. H., Fredericks E. M., Bowers K. M. (2019). Uncertainty in Big Data Analytics: Survey, Opportunities, and Challenges. Journal of Big Data. Vol. 6(1), pp.44.

Hastie T., Tibshirani R., Friedman J. H. (2001). The elements of statistical learning, Data mining, inference, and prediction. New York: Springer Verlag.

Haykin, S. (1998). Neural Networks: a Comprehensive Foundation. 2nd. Prentice Hall PTR.

Heisenberg D. (2005). The institution of 'consensus' in the European Union: Formal versus informal decision-making in the Council. *European Journal of Political Research*. Vol. 44 (1), pp. 65-90.

Ho W., Xu X., and Dey P. K. (2010). Multi-criteria decision-making approaches for supplier evaluation and selection: A literature review. Eur. J. Oper. Res., vol. 202, no. 1, pp. 16–24.

Holsapple. (2008a). Decisions and Knowledge. In F. Burstein , & C. W. Holsapple, Handbook on Decision Support Systems. Berlin - Heidelberg: Springer. Vol. 1, pp. 21-53.

Holsapple. (2008b). DSS architecture and types. In F. Burstein, & C. W. Holsapple, Handbook on Decision Support Systems.Berlin - Heidelberg: Springer. Vol.1, pp. 163-189.

Holsapple C.W. & Whinston A.B. (1996) Decision Support Systems: A Knowledge-Based Approach, West Publishing, Minneapolis/St. Paul.

Iqbal H. Sarker. (2021). Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. Springer Nature Singapore Pte Ltd. Vol. 2:377.

Jafarmadar S. (2015). A comparative analysis of two neural network predict ions for performance and emissions in a biodiesel fuelled diesel Engine. International Journal of Automotive Engineering Vol. 5, Number 2.

Janic M. & Reggiani A. (2002). An application of the multiple criteria decision making (MCDM) analysis to the selection of a new Hub Airport. Eur. J. Transp. Infrastruct. Res. EJTIR, 2, no. Mcdm.

Jato-Espino D., Castillo-Lopez E., Rodriguez-Hernandez J., and Canteras-Jordana J. C. (2014). A review of application of multi-criteria decision-making methods in construction. Automation in Construction, vol. 45. pp. 151–162.

Kahneman D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. American Economic Review, Vol. 93(5), pp.1449–1475.

Kaisler S., Amnour F., Alberto J. (2012) “Big Data: Issues and Challenges Moving Forward”, 46th IEEE international conference on system science, Wailea, Maui, HI, USA, pp. 7-10.

Kalantari B. (2010). Herbert a. Simon on making decisions: Enduring insights and bounded rationality. Journal of Management History, Vol. 16(4), pp. 509–520.

Kannan K. (2013). Big Data Analytics. IBM Course - Lecture Series . Jodhpur, IND: IBM Research Labs, pp. 17-18.

Karpatne A., Atluri G., Faghmous J.H., Steinbach M., Banerjee A., Ganguly A., Shekhar S., Samatova N., Kumar V. (2017). Theory-guided Data Science: A New Paradigm for Scientific Discovery from Data. IEEE Trans Knowl Data Eng. Vol. 29(10), pp. 2318–31.

Kasunic M., & Anderson W. (2004). Measuring systems interoperability: Challenges andopportunities. Carnegie-Mellon University, The Software Engineering Institute. Pittsburgh: Carnegie-Mellon University.

Keim D., Andrienko G., Fekete J.D., Görg C., Kohlhammer J., Melançon G. (2008). Visual Analytics: Definition, Process and Challenges. Springer, LNCS. pp.154 -175.

Khan M. & Ayyoob M. (2018). Big data analytics evaluation. Int J Eng Res Comput Sci Eng (IJERCSE). 5(2): 25–8.

Kopáčková H. & Škrobáčková M. (2006). Decision Support Systems or Business Intelligence: What Can Help in Decision Making? Scientific Papers of the University of Pardubice. Series D, Faculty of Economics and Administration. Vol. 10, p.6.

Kotsiantis, S. (2007). Supervised Machine Learning: A Review of Classification Techniques, Informatica Journal Vol. 31 pp. 249-268

Mandinach E.B. (2012). A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice. Educational Psychologist. Vol. 47(2), pp. 71–85.

Manieniyan V. & Sivaprakasam S. (2013). Artificial Neural Network Based Modelling for Vibration Characteristics of Diesel Engine Using Bio-Diesel. International Journal of Advanced Research in Computer Proceedings of the Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud). IEEEXplore part number: CFP20OSV-ART; ISBN: 978-1-7281-5464-0

Manjunatha R., Badari Narayana P., Hemachandra Reddy K. and Vijaya Kumar Reddy K.. (2012). Radial basis function neural networks in prediction and modeling of diesel engine emissions operated for biodiesel blends under varying operating conditions. Indian Journal of Science and Technology, Vol. 5 No.3. ISSN: 0974- 6846.

Mohamadabadi H. S., Tichkowsky G., and Kumar A. (2009). Development of a multi-criteria assessment model for ranking of renewable and non-renewable transportation fuel vehicles. Energy, vol. 34, no. 1, pp. 112–125.

Mohemad R., Hamdan A.R., Othman,Z.A. and Noor N.M.M. (2010). Decision Support Systems (DSS) in Construction Tendering Processes.

Morris E., Levine L., Meyers C., Place P., & Plakosh D. (2004). System of Systems Interoperability (SOSI): Final Report. Software Engineering Institute. Pittsburgh: Carnegie- Mellon University.

Mousavi-Nasab S. H. & Sotoudeh-Anvari A. (2017). A comprehensive MCDM-based approach using TOPSIS, COPRAS and DEA as an auxiliary tool for material selection problems. Mater. Des., vol. 121, pp. 237–253.

Mulliner E., Malys N., and Maliene V. (2016). Comparative analysis of MCDM methods for the assessment of sustainable housing affordability. Omega, vol. 59, pp. 146–156.

Mushtaq M. F., Akram U., Tariq A., Khan I., Zulqarnain M., & Iqbal U. (2017). An Innovative Cognitive Architecture for Humanoid Robot. *International Journal of Advanced Computer Science and Applications*. Vol. 8(8), pp. 60–67.

Natter, M., Ockerman, J., & Baumgart, L. (2010). Review of Cognitive Metrics for C2. International T&E Association Journal. Vol. 31, pp. 179-209.

Neiger D. & Churilov L. (2008). Process-Based Decision Support. Handbook on Decision Support Systems Berlin-Heidelberg: Springer. Vol. 2, pp. 211-237.

Oreški D. & Begičević Redep N. (2018). Data-driven decision-making in classification algorithm selection. Journal of Decision systems. Taylor & Francis Group.

Papadimitriou S. & Yu P. S. (2006). Optimal multi-scale patterns in time series streams. In SIGMOD, pp. 647–658.

Peffers K, Tuunanen T, Rothenberger M.A, Chatterjee S. (2008). A Design Science Research Methodology for information systems research. Journal of Management Systems, Vol. 24 (3), pp. 45 - 77

Peterson M. (2011). Decision theory: An introduction. International encyclopedia of statistical science. Springer (Ed.), pp. 349–356.

Pick R. A. & Weatherholt N. (2013). A Review on Evaluation and Benefits of Decision Support Systems. The Review of Business Information Systems (Online), Vol. 17(1), pp.7.

Pick, R. A. (2008). Benefits of decision support systems. In F. Burstein , & C. Holsapple, Handbook on Decision Support Systems 1 (pp. 719-730). Berlin-Heidelberg: Springer.

Pohekar S. D. & Ramachandran M. (2004). Application of multi-criteria decision making to sustainable energy planning - A review," Renew. Sustain. Energy Rev., vol. 8, no. 4, pp. 365–381.

Power, D. J. (2002). Decision support systems: concepts and resources for managers. Westport,CT: Quorum Books.

Power D. J. (2014). Using Big Data for Analytics and Decision Support. Journal of Decision Systems. Vol. 23(2), pp. 222-228.

Provost F. & Fawcett T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*. Vol. 1(1), pp. 51–59.

Rizk A. & Elragal A. (2020). Data science: developing theoretical contributions in information systems via text analytics. J Big Data. Vol. 7(1), pp.1–26.

Roya A., Cruz a R.M.O., Sabourina R., Cavalcanti G.D.C. (2018). A Study on Combining Dynamic Selection and Data Preprocessing for Imbalance Learning. Neurocomputing 286 pp:179–192.

Saaty T. L. (1980). The Analytic Hierarchy Process. McGraw-Hill, New York.

Saaty, T. L., & Tran, L. T. (2007). On the invalidity of fuzzifying numerical judgments in the Analytic Hierarchy Process. Mathematical and Computer Modelling, 46(7–8), 962–975.

Saaty T. L. & Vargas L. G. (1991), Prediction, Projection and Forecasting, Kluwer Academic, Boston.

Sarker I.H., Hoque M.M., Uddin M.K., Alsanoos, T. (2020). Mobile Data Science and Intelligent Apps: Concepts, AI-Based Modeling and Research Directions. Mob Netw Appl. pp.1–19.

Sauter, V.L. (2005). Competitive intelligence systems: qualitative DSS for strategic decision making, *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, Vol. 36(2), pp.43 - 57.

Shafiee M. (2015) Maintenance strategy selection problem: An MCDM overview. J. Qual. Maint. Eng., vol. 21, no. 4, pp. 378–402.

Shahsavari M. H. & Khamehchi E. (2018). Optimum selection of sand control method using a combination of MCDM and DOE techniques," J. Pet. Sci. Eng., vol. 171, no. January, pp. 229–241.

Shalev-Shwartz S. & Ben-David S. (2014). Understanding Machine Learning From Theory to Algorithms.

Sharda R., Barr S., McDonnell J. (1988). Decision support systems effectiveness: a review and an empirical test, Management Science. Vol. 34 (2), pp. 139 - 159.

Shim J. P., Warkentin M., Courtney J. F., Power D. J., Sharda R., & Carlsson C. (2002). Past, present, and future of decision support technology. (J. R. Marsden, Ed.) Decision support systems, Vol. 33 (2), pp: 111-126.

Siksnelyte I., et al. (2018). An overview of multi-criteria decision-making methods in dealing with sustainable energy development issues. Energies, vol. 11, no. 10.

Silahtaroğlu G. & Yılmaztürk N. (2019). Data Analysis in Health and Big Data: A Machine Learning Medical Diagnosis Model based on Patients' Complaints. Commun Stat Theory Methods. pp.1–10.

Simanaviciene R. & Ustinovichius L. (2010). Sensitivity analysis for multiple criteria decision-making methods: AHP, TOPSIS and SAW. Procedia - Soc. Behav. Sci., vol. 2, no. 6, pp. 7743–7744.

Simon, H.A. (1959). Theories of decision-making in economics and behavioral science. The American Economic Review, Vol. 49(3), pp. 253–283.

Singh H. (2015). Introduction to Project Management Analytics: A Data-Driven Approach to Making Rational and Effective Project Decisions.

Singh R., Sharma M., Sharma S., Kaushal S. K. (2012). Management Information System, Published by Kalyani Publishers.

Sprague R.H. & Carlson E.D. (1982). Building Effective Decision Support Systems, Prentice-Hall Inc., Englewood Cliffs, NJ.

Štirbanović Z., Stanujkić D., Miljanović I., and Milanović D. (2019). Application of MCDM methods for flotation machine selection," Miner. Eng., vol. 137, pp. 140–146.

Taiwo, O. A. (2010). Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), PP. 3- 31.

Triantaphyllou E., Shu B., Sanchez S. N., and Ray T. (1998). Multi-Criteria Decision Making: An Operations Research Approach. Electronics, vol. 15, pp. 175–186.

Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer-Verlag.

Wang J. & Kumar A. (2005). A framework for document-driven workflow systems: International Conference on Business Process Management, Springer, Berlin, Heidelberg, September, pp.285–301.

Wang W. & Tolk A. (2008). The levels of conceptual interoperability model: applying systems engineering principles to M&S. Society for Computer Simulation International, San Diego: pp. 168.

Zhu D.T. & Zhang H. (2009). Study on expressway meta-synthesis management decision support system. Chinese Control and Decision Conference, IEEE, June, pp. 4951- 4956.

Zhu Y. & Shasha D. (2002). Statstream: Statistical monitoring of thousands of data streams in real-time. In VLDB, pp. 358–369.

Zikopoulos P. & Eaton C. (2011). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw-hill Education.