**Assessment 1: Research Project Proposal**

Towards a Decision Support System (DSS) using Heuristics

|  |  |
| --- | --- |
| Name: | Supervisor: |
| University ID: |  |
| Word Count: 3388 (Without References) |  |
| Date: |  |

Table of Contents

[Table of Figures 2](#_Toc69988735)

[Abstract 3](#_Toc69988736)

[Introduction 3](#_Toc69988737)

[Background 4](#_Toc69988738)

[Aims and Objectives 4](#_Toc69988739)

[Research Questions 4](#_Toc69988740)

[Ethical Considerations 5](#_Toc69988741)

[Literature Review 5](#_Toc69988742)

[A decision-support system for the design and management of warehousing systems 5](#_Toc69988743)

[A Decision Support System for Classification of Normal and Medical Renal Disease Using Ultrasound Images: A Decision Support System for Medical Renal Diseases 8](#_Toc69988744)

[Project Timeline 11](#_Toc69988745)

[Project Design 11](#_Toc69988746)

[Methodology 12](#_Toc69988747)

[Project Evaluation 13](#_Toc69988748)

[Conclusion 13](#_Toc69988749)

[References 13](#_Toc69988750)

# Table of Figures

[Figure 1 - Framework for Warehousing and Operational Issues 5](#_Toc68956673)

[Figure 2 - Results of a multi-scenario analysis. 7](#_Toc68956674)

[Figure 3 - Experimental Workflow 8](#_Toc68956675)

[Figure 4 - Dataset Description 9](#_Toc68956676)

[Figure 5 - Experiments Carried Out in this Research 9](#_Toc68956677)

[Figure 6 - Results 10](#_Toc68956678)

[Figure 7 - Gantt chart 10](#_Toc68956679)

[Figure 8 - Risk management 11](#_Toc68956680)

# Abstract

The mix of modern data mining technology and the evolution of Big Data has created a situation where enterprises that maximize their understanding of data have an advantage over their competitors. Better methods of extracting and understanding data allow for more efficient business decisions.

This research combines Machine Learning with Heuristics in the form of an algorithmic ensemble and applies them to modern businesses. It applies itself to data that governs business decisions and provides insightful feedback that can efficiently assist in making decisions.

The masses are getting increasingly aware of their surroundings and linking with the world more than ever. While this is a win, this causes demand and the global market to be more dynamic and chaotic. Keeping up with the times requires an infrastructure that can keep up with the dynamic nature of the global market. Using probability driven heuristics, this uncertainty can be captured.

The output of this research will be to analyse the global market conditions using the dataset available [here](https://www.kaggle.com/jackdaoud/marketing-data) and provide exploratory data analysis on the situation. It will also provide data-driven solutions and extensive data visualization.

**Keywords:** Heuristics, Statistical Insights, Business Data

# Introduction

Decision Support Systems (DSS) are famous apparatuses that help dynamic in an association. Its significance is likewise perceived because of its commitments in dynamic associations. DSS has been synergized with information management frameworks and have developed from before, ideas of "information handling" and Management Information Systems (MIS) to their present structure as crucial help for the business dynamic (Alyoubi, 2015). The value attributed to Business Support Systems and its ability to aid business decisions has exponentially grown over time. When support systems are employed there is a positive effect on the efficiency of business decisions. In 2011, top-performing business was found to employ analytics at double the rate of their lower-performing counterparts. Since then, the volume and value of data have only increased due to the advent of better technology. This has compounded the importance of Business Decision Support Systems to modern enterprise.

A Decision Support Systems (DSS) is a popular tool of Information Systems that supports decision-making processes. A DSS has been defined as an interactive and adaptable computer-based Information System that supports nonstructured management problems too. Through the use of DSS, decision-makers were able to find solutions to various problems. These include semi- to ill-structured problems that involved multiple attributes, goals, or objectives (Alyoubi, 2015).

The DSS gives a valuable and easy to understand apparatus for supervisors furthermore, chiefs who have no foundation and ability in programming and programming improvement however who often face warehousing framework plan and activities issues. The DSS carries out data set administration framework (DBMS) structures for information stockpiling, models and heuristic calculations and easy to use graphical UI (GUI) that empower intelligent inquiries, detailing and realistic representation (Accorsi, Manzini and Maranesi, 2014). DSS allows for a more change accommodating business infrastructure thus helping the business expand and deal with the changing target market efficiently. Using heuristics on real-time business data, DSS can model the uncertainty of the market and provide decisions to balance and counter such effects.

The proposal will proceed as follows; the introduction section will continue defining the background of the project along with the objectives this project will achieve. The questions that this research will try to provide an answer for will also be discussed along with the ethical considerations of the project and the project timeline. The next section i.e., the methodology section will discuss the processes behind how the literature review will be performed and what the project design is. It will also discuss in detail the tools and theorems used for achieving the objectives this project has. The methodology section will end with a project evaluation that will convey the modules that the project comprises of and the order in which they need to be executed.

Following the former section will be the literature review which will discuss in detail the findings of peers researching in the same domain. The literature review will summarize the methodology, experiments and findings of each research paper finalized for study in this project. The conclusion will be the last section summarizing the complete project followed by references, appendix and glossary.

## Background

Changing the infrastructure to include DSS in businesses can prove to be useful in handling the dynamicity of the consumer market. Businesses are constantly challenged by the ever-changing demands of the people and it becomes extremely hard keeping up. This research resulting in a DSS can help towards achieving relative ease in tackling the changes.

## Aims and Objectives

This research project aims to analyse the current market trend using data-driven analysis and provide data-driven solutions for tackling the market. This will be achieved by the following steps:

* Exploratory Data Analysis
  + Checking for Null values and outliers
  + Features requiring transformations
  + Features that can be derived from the current data
  + Anomaly Detection
* Statistical Analysis will be done in the form of multiple linear regressions to achieve the following:
  + Factors related to the number of store purchases
  + Comparison of US with the world regarding total purchases
  + Verifying the following hypothesis:
    - ‘People Buying Gold are conservative’
  + Relation between geography and the success of a campaign
* Data visualization
  + Most Successful marketing campaign
  + The outlook of an average customer for the company
  + Best performing products
  + Underperforming channels

## Research Questions

The following research questions help drive the motivation behind this project:

1. How to tackle the ever-changing demand of the consumer market?
2. Can machine learning be of any assistance in making adaptable decisions for a business?

## Ethical Considerations

Ethics is a complicated subject that has only become more prominent during the advent of Big Data. The UK Data Service department also provides guidelines for ethical research with specific relation to Big Data. These guidelines will form the basis for this report ethical approach. Some of the concerns that will be addressed are:

* Maintaining confidentiality in line with Birmingham City University (BCU) and DC guidelines,
* Anonymizing information that violates group privacy,
* Ensuring transparency in reasons for data collection,
* Ensuring data is only used for the direct purpose it has been requested,
* Referencing sources for all information used within the research project,
* Ensuring all data is stored in the correct location. DC information must remain on DC servers.

As this project encounters any further ethical concerns these will be met within the recommended UK guidelines and with the advice of BCU and DC supervising members.

## Literature Review

The process of reading, studying, comprehending and heeding from researches done by scientists in the same domain is called Literature Review (LR). A detailed LR helps provide a head-start to the researcher, provides contingencies for common errors, and validates the integrity of research.

### A decision-support system for the design and management of warehousing systems

This examination presents a unique choice of an emotionally supportive network (DSS) for the plan, the executives, and control of warehousing frameworks. In particular, the proposed DSS executes a top-down philosophy that considers both key stockroom plan and usable activities of the executives. The DSS can reenact the coordination and material dealing with exhibitions of a warehousing framework. Heuristic techniques also, calculations address a few basic stockroom issues, for example, the request picking measure, which is liable for 55% of the general expenses in an appropriation place. The advantages because of the appropriation of the proposed choice emotionally supportive network are summed up as a dashboard of key execution pointers (KPIs) of reality effectiveness that permit coordination suppliers, specialists, and chiefs just as academicians and instructors to confront true warehousing occurrences and to discover valuable rules for material dealing with.

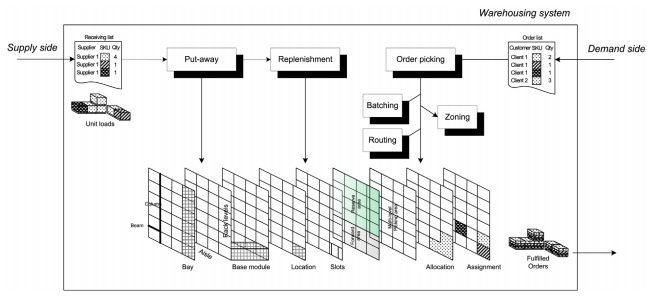


Figure 1 - Framework for Warehousing and Operational Issues

The proposed DSS executes a top-down strategy for the plan and the executives of a forward-hold OPS. This approach puts together strategies, models, furthermore, calculations in a natural successive choice to give a wide set of answers for capacity design, stockpiling allotment, and capacity task. The chief directs a succession of investigations, creating sets of elective stockroom setups to be surveyed through a consider the possibility that multi-situation reenactment. The objective execution is the minimization of the complete voyaged distance due to picking, which addresses 55% of the general distribution centre expenses.

#### Methodology

##### Layout

The proposed technique depends on the authentic stock and client interest (or the interest figures) that are expected as a contribution for a stock out hazard assessment examination that is planned to build up the necessary stockpiling limit of the warehousing framework (i.e., planned from a green-field). The reason for this progression is to set the office format through the meaning of a bunch of boundaries, like the shape factor, the number of paths, the number of inlets per passageway, the rack sizes and types, and the attributes of the unit load (i.e., the bed size or other holder arrangements). The DSS defends the entire extra room by dedicating various zones for various SKUs as far as the shape and size, which probably require explicit racks or hardware.

##### Allocation

At this progression, the chief matches the distribution results with design highlights and ultimately considers the chance to get back to the top for re-design arranging. The capacity task systems set up the suitable areas to allocate to the SKUs as per unique heuristics. The DSS requests the conduct from chosen SKUs inside the requested profile for a chose time skyline. Data on the picking measures are gathered to figure a board of measurements utilized for SKU characterization. Another significant perspective that can be considered through the proposed DSS is the connection among the SKUs that are mentioned together by clients. Associated based task arrangements can be applied to bunch SKUs that are mentioned together and appoint them to capacity areas that are near each other, to save money on the voyaging required for the picking exercises. The executed corresponded based methodology involves the accompanying three primary steps:

* Correlation investigation. The degree of relationship is by and large estimated by presenting a comparability file among the SKUs. This technique permits contrasting broadly useful similitude files, e.g., the Jaccard list and certain issue arranged issues.
* Clustering: This progression concerns the selection of various levelled bunching calculations (e.g., single linkage, complete linkage, bunch normal) and diverse closeness cut limits of a dendrogram (i.e., esteem based, percentile-based).
* Clustering Assignment. This progression processes the previously mentioned measurements (i.e., prominence, turn-over, request shutting) for each group of SKUs (e.g., the prominence of a bunch is given by the weighted whole of the prominence of the included SKUs) and to sort the bunches of SKUs likewise.

##### Multi-scenario simulation

The arrangement of choices that were recently tended to by the DSS (i.e., the format plan, allotment and task) gives an explicit arrangement for a stockroom situation. Different emphases of the DSS consider creating various stockroom situations, which vary in their design arrangement, stockpiling portion, and or on the other hand stockpiling task rules. Taking everything into account, imagine a scenario in which multi-scenario reenactment of usable exhibitions (i.e., going for set aside, renewal and picking) empowers the leader to evaluate the best answer for the stockroom plan and the executives by the minimization of the complete voyaging distance, time and cost.

#### Functionality and Design

The proposed application depends on an independent information base. Choice interaction contributions concerning employable highlights, costs, also, different boundaries are by and large dealt with by specialists in distribution centre tasks, while yields involve usable KPIs that are typically followed in reality (e.g., the pick-rate, time/ going for picking). The SQL information base engineering empowers clients to assemble, store and deal with an extremely enormous measure of information rapidly, which is gatherable by clients through unique queries. The application is coordinated around a fundamental GUI that presents all of the guideline highlights and orders to stack information or ventures furthermore, to save results. The device empowers the accompanying principal functionalities:

* Plan another warehousing framework (we call green-field).
* Import the current design (we call earthy coloured field) to play out an allotment task examination.
* Run the DSS for a total format portion task examination of a conventional stockroom zone (i.e., as per stockroom drafting).
* Union single-various stockroom zones (i.e., as per stockroom drafting) as a totalled framework.
* Carry out heuristics for capacity portion, task, single order picker-directing, request bunching.
* Build up a consider the possibility that multi-situation examination for the warehousing KPIs.

#### Database Considerations

The created DBMS addresses an interface between the information furthermore, the chief. This framework includes the preparing of an impressive measure of information which is important to depict univocally the qualities of the warehousing framework. This framework involves a social SQL engineering that is controlled by MS-Access however is rapidly replaceable by some other business DBMS (e.g., MySQL1, DB21). The information base incorporates a bunch of tables that take into account a far-reaching depiction of the framework's object of investigation through a commonplace snowflake structure. Significant primer concentrates on the brought together demonstrating language (UML) and E-R outlines are urgent to planning enlightening engineering with the instrument and to help in additional code upkeep of the changes.

#### Results

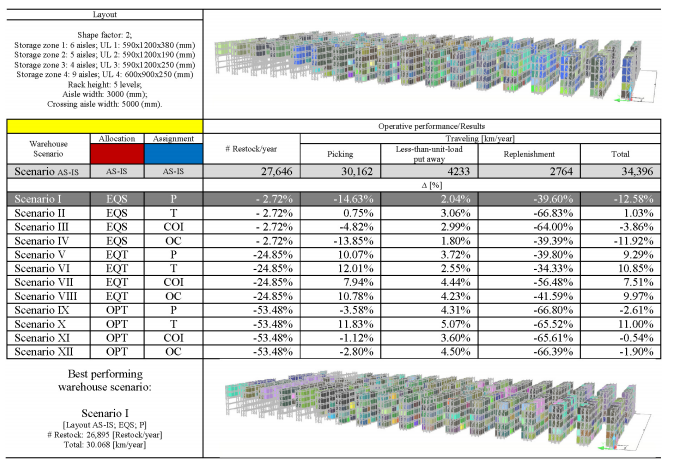


Figure 2 - Results of a multi-scenario analysis.

### A Decision Support System for Classification of Normal and Medical Renal Disease Using Ultrasound Images: A Decision Support System for Medical Renal Diseases

The current work proposes a productive choice emotionally supportive network for discovery of clinical renal infection utilizing little element space comprising of just second request GLCM measurable highlights registered from crude renal ultrasound pictures. The GLCM mean element vector and GLCM range highlight vector are processed for between pixel distanced changing from 1 to 10. These surfaces include vectors that are consolidated in different manners yielding GLCM proportion highlight vector, GLCM added substance include vector and GLCM connected component vector. The current work investigates the capability of five surface element vectors figured utilizing GLCM insights thoroughly for differential analysis between typical and MRD pictures utilizing an SVM classifier. The consequence of the examination demonstrates that GLCM range highlight vector figured with d = 1 yields the most noteworthy in general grouping precision of 85.7% with singular grouping exactness estimations of 93.3% and 77.9% for typical and MRD classes separately.

#### Methodology

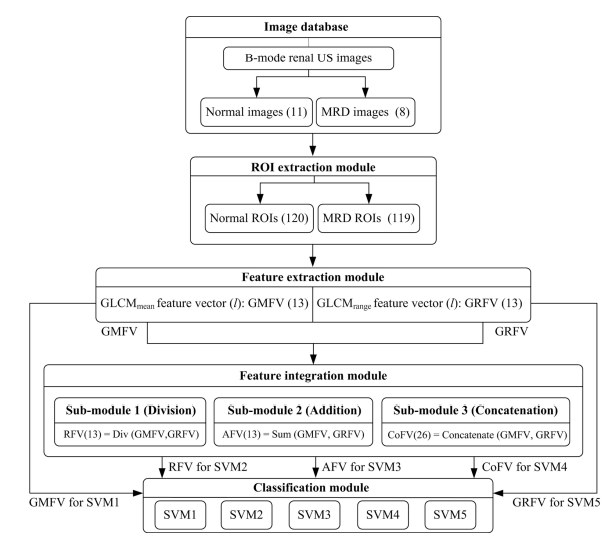


Figure 3 - Experimental Workflow

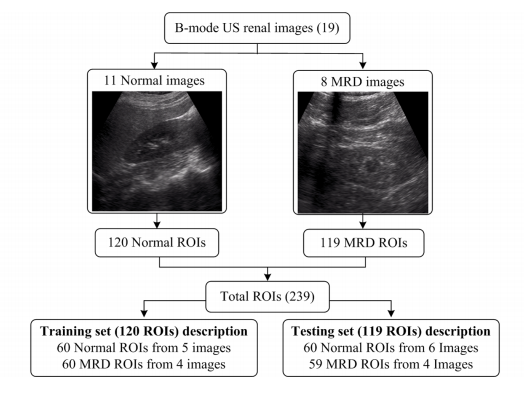


Figure 4 - Dataset Description

#### Experiments

The SVM classifier is used in this research to achieve the aim. The objective of the SVM classifier is to plan a hyperplane that groups all preparing vectors in two classes with the greatest edge. The edge is the distance between the hyperplane and the nearest information focuses on this hyperplane. In the current work, the Gaussian outspread premise work bit has been utilized for planning of information focuses from input space to higher dimensional element space. The ideal qualities for (a) part boundary γ (controls the arch of the choice limit) and (b) delicate edge steady C (increment the edge with least conceivable mistake) are needed to fabricate a productive preparing model for SVM classifier.

In the current work, the capability of various surface component vectors figured utilizing GLCM insights (GMFV, GRFV, RVF, AFV, CoFV) for various interpixel distances d ∈ {1, 2, 3…,10}, has been assessed utilizing SVM classifiers. The ideal qualities of C and γ for each SVM classifier has been acquired utilizing framework search technique with the end goal that γ ϵ {2−12, 2−11, …, 25} and C ϵ {2−4, 2−3…, 215} utilizing 10-overlay cross-approval on preparing information for example for γ =2−12 the estimation of C differs from 2−4, 2−3… 215 then for γ =2−11 the estimation of C is changed from 2−4, 2−3…, 215. In this manner for each estimation of C and γ in the predefined lattice 10-overlap cross approval preparing precision is gotten. The estimations of C and γ yielding the maximum preparing exactness are utilized to freeze the preparation model. The concealed testing information is then projected on the prepared model.

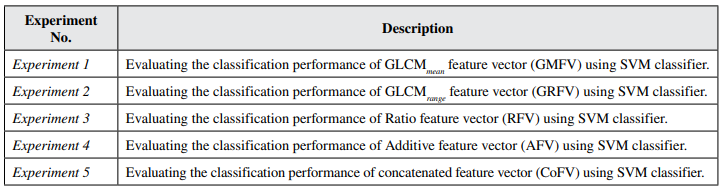


Figure 5 - Experiments Carried Out in this Research

#### Results

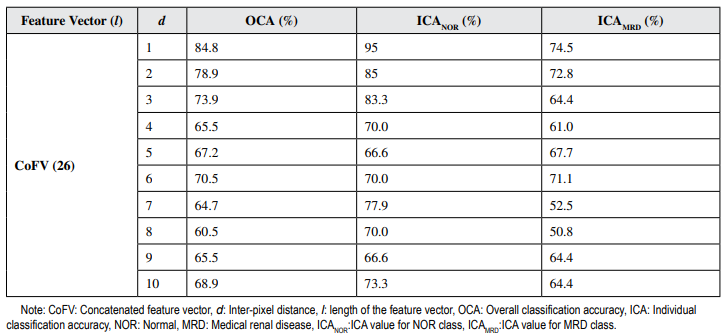


Figure 6 - Results

## Project Timeline

Projects are time-bound and the ability to meet a deadline is key to success. A project timeline lays out key project deliverables and the scope of their completion. This research project identifies time as its key resource. By efficiently allocating time to various tasks resource overload is minimized. Preventing resource overload minimizes the risk of dropping quality.

The Gantt chart is identified as a strong tool for time management. The Gantt chart designed for this project is laid out in figure 1. Tasks are laid out in chronological order on the left-hand side. The timeframe for their completion is found along the X-Axis. By sticking to this schedule, the project will be delivered promptly to a high standard.

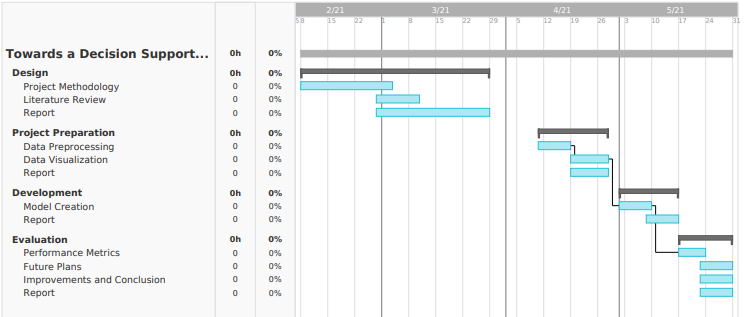


Figure 7 - Gantt chart

There may be unexpected events that cause difficulty in maintaining the project timeline. These are risks. Contingencies for some expected risks are discussed below in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Risk Impact (1-5)** | **Risk Likelihood (1-5)** | **Contingency** |
| Data Mismanagement | 5 | 1 | Each iteration of the dataset will be synced with Version Control. |
| Personal Time Mismanagement | 3 | 3 | Regular Feedback from Supervisor for Continuous development. |
| Hardware Failure | 1 | 1 | Extra Supply of RAM and SSD. |

Figure 8 - Risk management

## Project Design

This project is research using Machine Learning to learn the extent of a DSS in business applications. This will consist of experiments and data analysis. Since this is experimental research without a definitive quantitative goal, there will not be a Software Development Life Cycle (SDLC) followed. This is to allow freedom from constraints that come with an SDLC.

# Methodology

The dataset for this research comes from [here](https://www.kaggle.com/jackdaoud/marketing-data). The dataset consists of data on 2,240 customers with a dummy company XYZ. This dataset was created as a toy dataset to experiment with DSS analytical prowess. The dataset contains the features that can be classified as follows:

* Customer Profiles
  + ID
  + Year\_Birth
  + Education
  + Marital\_Status
  + Income
  + Kidhome: Number of small kids at home
  + Teenhome: Number of teenagers at home
  + Dt\_Customer: Customer since which date
  + Recency
* Product Preferences
  + MntWines
  + MntFruits
  + MntMeatProducts
  + MntFishProducts
  + MntSweetProducts
  + MntGoldProds
* Campaign Successes/Failures
  + NumDealsPurchases
  + NumWebPurchases
  + NumCatalogPurchases
  + NumStorePurchases
  + NumWebVisitsMonth
* Channel Performance
  + AcceptedCmp1
  + AcceptedCmp2
  + AcceptedCmp3
  + AcceptedCmp4
  + AcceptedCmp5
  + Response
  + Complain
  + Country

Since the overall goal is to analyse the market using the finalized dataset and provide Data-driven solutions, which is what the DSS is meant for, the choice for the Machine Learning paradigm will be regression and the algorithm used will be Linear Regression. There will be other algorithms that might be implemented and experimented with, but Linear Regression is currently the finalized candidate.

# Project Evaluation

The project will be evaluated based on the following:

* Performance metrics for the Machine Learning algorithms.
  + The performance metric used will be the Root Mean Square Error (RMSE) defined as
* The results of statistical data analysis done.
* The effectiveness of the assistance offered.

# Conclusion

DSS has been assisting businesses by supporting the business dynamic. DSS work by digesting information and data for a business scenario and provide statistical analysis on it that can crucially support the decision-making process. This research is focused on using heuristics and evolutionary approaches to achieve a DSS that can provide a much higher level of support when compared to state-of-the-art DSS.

# References

* Accorsi, R., Manzini, R. and Maranesi, F., 2014. A decision-support system for the design and management of warehousing systems. Computers in Industry, 65(1), pp.175-186.
* Alyoubi, B., 2015. Decision Support System and Knowledge-based Strategic Management. Procedia Computer Science, 65, pp.278-284.
* Sharma, K. and Virmani, J., 2017. A Decision Support System for Classification of Normal and Medical Renal Disease Using Ultrasound Images. International Journal of Ambient Computing and Intelligence, 8(2), pp.52-69.