



Electronic Theses and Dissertations

2021

A Prototype for profit maximization using Apriori data mining algorithm: case of Kula Kona restaurant.

Okeyo, Seth Ouma
Faculty of Information Technology
Strathmore University

Recommended Citation

Okeyo, S. O. (2021). *A Prototype for profit maximization using Apriori data mining algorithm: Case of Kula Kona restaurant* [Thesis, Strathmore University]. <http://hdl.handle.net/11071/12936>

Follow this and additional works at: <http://hdl.handle.net/11071/12936>

A Prototype for Profit Maximization Using Apriori Data Mining Algorithm: Case of Kula Kona Restaurant

Okeyo, Seth Ouma

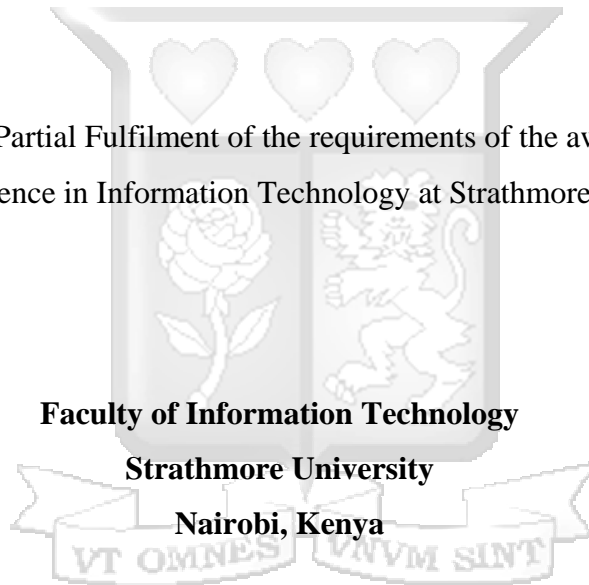
094759

A Thesis Submitted in Partial Fulfilment of the requirements of the award of the Degree of
Master of Science in Information Technology at Strathmore University

Faculty of Information Technology

Strathmore University

Nairobi, Kenya



August 2020

This Thesis is available for use on the understanding that it is copyright material and that no quotation from this thesis may be published without proper acknowledgment.

Declaration

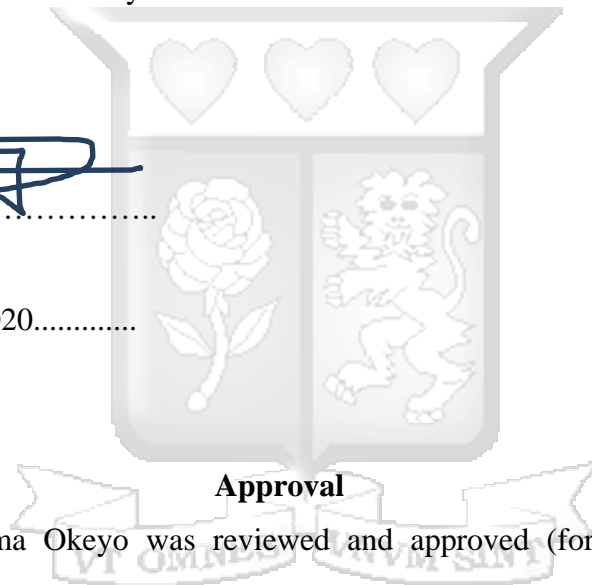
I declare that this work has never been previously submitted and approved for any award of degree by Strathmore University or any other University. To the best of my knowledge, this thesis contains no material previously written or published by someone else to the best of my knowledge except where due reference is made within the thesis.

©2020 No part of this research thesis may be produced without the authors' permission and permission of Strathmore University.

Okeyo, Seth Ouma

Signature: 

Date:26-August-2020.....



Approval

The thesis of Seth Ouma Okeyo was reviewed and approved (for examination) by the following:

Dr. Bernard Shibwabo

Senior Lecturer, Faculty of Information Technology

Strathmore University

Signature: *B.K.S*

Date:12-November-2020.....

Abstract

Most businesses use different promotional and pricing methods to improve profits, revenues, and sales volumes. For example, a restaurant manager may change prices to encourage sales of food items. Also, he or she may in a special way advertise or present advertise the items to increase customers' awareness and demand. This has become cumbersome and has made the management of such businesses difficult and has informed the decisions to develop systems with aims of coming up with a solution to this, most of the systems are complicated in nature and difficult for the users to apply, most of them are also rigid to platforms/system requirements since they were developed in old and probably unscalable platforms. The aim of this research is to formulate a prototype for profit maximization using Apriori data mining algorithm. This is achieved by applying the algorithm on existing sales knowledge bases with other given parameters, some kept constant and others varying, and the algorithm is able to determine the sales patterns using different internal and external parameters. The prototype then automatically analyzes the patterns and come up with reports and summaries which can aid in decision making and consequently profit maximization with the optimal prices of the goods, which is advantageous to both the restaurant owners and clients. The research site is Kula Kona restaurant located in the Hurlingham area within Nairobi. The research design is used to conduct the scientific study and descriptive approach to demonstrate the effects of adjustments of different variables which help understand the behavior and effects on other variables in relation to sales at a given time. The Data-driven modelling methodology was used in this model development. The methodology was ideal since it relied on retrospective data and it performed at the accuracy level of 93.71% and a mean square error of 0.039. The results were great and showed that Apriori algorithm is the best fit for this type of machine learning prototype.

Keywords: Artificial intelligence, Apriori algorithm, Data Mining, Development Platforms, Knowledge Base.

Acronyms

AI –	Artificial Intelligence
CSV –	Comma Separated Values
DFD –	Data flow Diagram
ICTs –	Information and Communication Technology
ID –	Identification
IEEE –	Institute of Electrical and Electronics Engineering
KDD –	Knowledge Discovery in Databases
MBA –	Market Basket Analysis
PoS –	Point of Sales



Operational Definition of Terms

Apriori Algorithm: An algorithm for mining frequent itemset using Boolean association rules (Agrawal & Srikant, 1994).

Artificial Intelligence: An area in Computer Science that deals with developing machines that act like human beings (Terry et al., 1994).

Data Mining: The practice of examining large pre-existing databases to generate new information (Yang & Shen, 2003).

Profit Maximization: This is a process by which a firm determines the price, input, and output levels that result into the greatest profit (Ehrenberg & Smith, 2016).



Acknowledgments

My acknowledgment goes to my research supervisor, Dr. Bernard Shibwabo, for his continuous guidance and input throughout my research and giving his time generously, to Dr. Vincent Omwenga for the great ideas he gave me towards this research, Prof. Ismail Ateya for the knowledge on databases he gave unwaveringly and to Dr. Joseph Orero for his good advice, to my family and friends for their support and encouraging ideas throughout my research.



Dedication

I give special dedication of this research to my beloved family for the immense support, encouragement and ample time they accorded to me through many obstacles to achieve my academic excellence. I sincerely give gratitude to God for His abundant favor and grace throughout my academic journey.

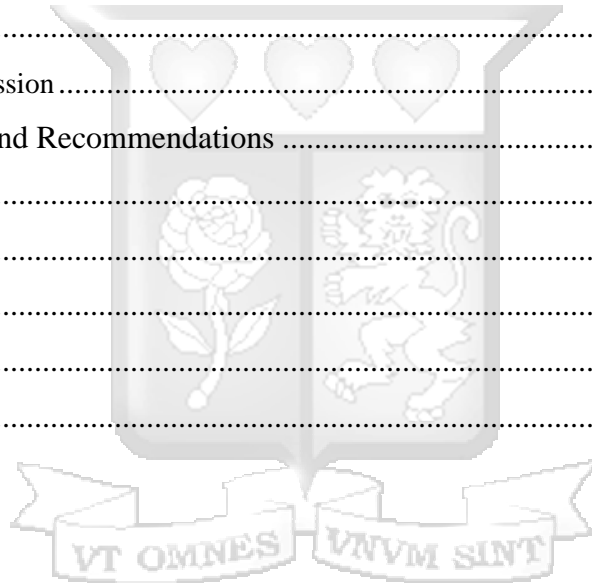


Table of Contents

Declaration.....	ii
Abstract.....	iii
Acronyms.....	iv
Operational Definition of Terms.....	v
Acknowledgments.....	vi
Dedication.....	vii
List of Figures.....	xi
List of Tables.....	xii
Chapter 1: Introduction.....	13
1.1. Background of the Study.....	13
1.2. Problem Statement.....	14
1.3. Research Objectives.....	14
1.4. Research Questions.....	15
1.5. Justification.....	15
1.6. Scope.....	16
1.7. Limitations.....	16
Chapter 2: Literature Review.....	17
2.1. Introduction.....	17
2.2. Factors Determining Profit in Restaurant Business.....	18
2.3. Profit Maximization.....	18
2.3.1. Importance of Profit Maximization.....	20
2.4. Apriori Algorithm (Associate Rule Mining).....	20
2.5. Profit Maximization Elaborations Example.....	21
2.1.....	21
2.6. Conceptual Framework.....	22
Chapter 3: Research Methodology.....	23
3.1. Introduction.....	23
3.2. Research Site.....	23
3.3. Research Design.....	23

3.4.	Data Acquisition	24
3.5.	Design and Development Approach	26
3.5.1.	Waterfall Model - Design	27
3.5.2.	Justification for Choosing Waterfall.....	30
3.6.	Data Management	30
3.7.	Data Readability and Validity	31
3.8.	Ethical Considerations.....	31
Chapter 4: System Analysis, Design, and Architecture		32
4.1.	Introduction.....	32
4.2.	Requirement Analysis	32
4.2.1.	Functional Requirements.....	32
4.2.2.	Non-Functional Requirement	32
4.3.	System Design and Architecture	33
4.3.1.	Partial Domain Model.....	33
4.3.2.	Use Case Diagram	34
4.3.3.	Use Case Basic Flows.....	35
4.3.4.	Context Diagram	37
4.3.5.	Level 0 Data Flow Diagram	37
4.3.6.	Sequence Diagram.....	38
4.3.7.	Class Diagram.....	39
4.4.	Algorithm Characteristics	40
4.4.1.	Data Source	40
4.4.2.	Data.....	41
4.4.3.	Algorithm Structure.....	41
Chapter 5: System Implementation and Testing.....		44
5.1.	Introduction.....	44
5.2.	Hardware and Software Specifications	44
5.3.	Dataset Generation and Preprocessing	45
5.3.1.	Cost of Goods	45
5.3.2.	Profit	46
5.3.3.	Change in Revenue.....	47

5.3.4.	Change in Quantity	48
5.3.5.	Change in Cost	49
5.3.6.	Marginal Revenue	49
5.3.7.	Marginal Cost	50
5.3.8.	Combined Changes.....	51
5.4.	Training Dataset creation and processing for the model.	52
Chapter 6: Discussions.....		57
6.1.	Introduction	57
6.2.	Model Representation	57
6.3.	Data Mining Algorithms for Profit Optimization	58
6.4	Model validation	58
6.4.	Conclusion on discussion.....	60
Chapter 7: Conclusions and Recommendations		61
7.1.	Conclusion.....	61
7.2.	Challenges	61
7.3.	Recommendations	62
7.4.	Future Work.....	62
References.....		64



List of Figures

Figure 2. 1: Diagram of Profit Maximization	19
Figure 2. 2: Diagrammatic Explanation of the Apriori Algorithm.	21
Figure 2. 3: Diagrammatic Representation of the Model.	22
Figure 3. 1: SDLC Stages	27
Figure 3. 2: Waterfall Phases	28
Figure 4. 1: A Partial Domain Model	34
Figure 4. 2: Use Case Diagram	35
Figure 4. 3: Context Diagram	37
Figure 4. 4: Level 0 Data Flow Diagram	38
Figure 4. 5: Sequence Diagram.....	39
Figure 4. 6: Class Diagram	40
Figure 4. 7: Pricing Method with Mathematical Model	43
Figure 5. 1: Calculating Cost of Goods	46
Figure 5. 2: Calculating Profit	47
Figure 5. 3: Calculating Change in Revenue	48
Figure 5. 4: Calculating Change in Quantity	48
Figure 5. 5: Calculating Change in Cost.....	49
Figure 5. 6: Calculating Marginal Revenue.....	50
Figure 5. 7: Calculating Marginal Cost.....	51
Figure 5. 8: Combines changes calculations.....	52
Figure 5. 9: Lift, Support and Confidence script	53
Figure 5. 10: Data optimization script	54
Figure 5. 11: Association rules.	55
Figure 5. 12: The results from the model.....	56
Figure 6. 1: Diagrammatic Representation of the Model.	58
Figure 6. 2: Steps to check the efficiency of an algorithm	58

List of Tables

Table 2. 1: Profit Maximization Elaborations Table	21
Table 5. 1: Hardware and Software Specifications.....	44



Chapter 1: Introduction

1.1. Background of the Study

The main aim of business enterprises is to make a profit, profit can be defined as a financial gain, which is a positive difference between what is earned and what is spent (McDonald & Wilson, 2016). Ideally, profit is the measure of the success of a business entity in the sense that if a business does not make a profit, then it is on the verge of its existence. Therefore, the major aim of the business is profit-making and the more the profit, the successful the business is (Grant et al., 2016; Alen & Gearan, 2011).

Many of the business operators are constantly thinking of better ways of making more profit from their businesses. They come up with different methods, for example, a person in the restaurant business will come up with daily promotions, special offers and different strategies to lure and attract more customers in his business, and the major aim is to make more profit. The process of making more profit is profit maximization (Ehrenberg & Smith, 2016).

According to (Meng & Zeng, 2016), the approach on profit maximization intends to demand customers behaviors response by management, learning in smart grid explains further the processes of profit optimization because of the variety of promotions like discounts often applied at any one time, the market complexity, and difficulties in understanding of customers' response, it is usually not obvious to accurately forecast the effectiveness of various promotional tools and price adjustments by the business and to evaluate the influence of each selected tool on the overall sales revenues. It is also even more difficult to actualize various marketing campaigns and pricing adjustments beforehand to achieve this intended business optimality.

This proposed prototype utilizes the use of data mining algorithms primarily the Apriori algorithm to carry out data mining and discover sales patterns from the Point of Sales (PoS) data. Apriori involves a certain analysis method called Market basket analysis (MBA), this method has been used in previous research studies (Kaur et al., 2016).

Market Basket Analysis identifies the Changing Trends of Market Data Using Association Rule Mining (Manpreet et al., 2016). The main aim of this analysis in marketing is

to provide the information to the seller to understand the buyer's purchase behavior which helps the seller to make the correct decisions. There are different algorithms that are available for performing market basket analysis. These existing algorithms have worked on static data and they do not incorporate data changes with time. The proposed algorithm not only mine retrospective data but also involves data mining techniques and therefore will be applicable in different sets of data. This research involves the input of PoS data, analysis of the data by application of different techniques listed above, mine the data and discover the association patterns and using these discoveries make informed decisions.

1.2. Problem Statement

Most entrepreneurs and business managers are trying different methods and models to maximize their sales (Enherberg, 2016). Profit maximization has been a problem considering different aspects determining the sales in businesses like a restaurant, mostly the different factors affecting sales are so many and making a conclusive decision on how to adjust which one of them becomes a nightmare. There is a dire need for advanced, scalable, better quality and efficient models to fill the gap in the maximization of profit. The current systems for profit maximization are mostly long processes and manual calculations and arguments with long, unorganized tables and charts which makes it more tiresome and inefficient.

The need for better systems has informed this research to come up with a model that maximizes profit by use of an algorithm on the sales data, using different factors and conditions as controls. Use of Apriori data mining algorithm should employ techniques on the sales data and by discovering of sales patterns, decisions on the variables to adjust as per given constrains can be made easily, this can enable the restaurant manager to make well-calculated decisions which can help come up with optimal sales parameters and processes which maximizes the profit.

1.3. Research Objectives

- i. To investigate the factors determining profit in the restaurant business.
- ii. To appraise algorithms, models, and architectures that can be applied in profit maximization.
- iii. To develop a model for profit maximization using Apriori data mining algorithm.

- iv. To validate the model using restaurant sales data.

1.4. Research Questions

- i. What are the factors determining the profit level in the restaurant business?
- ii. What algorithms, models and architecture can be applied in profit maximization?
- iii. How can a prototype maximization prototype be developed using the Apriori algorithm?
- iv. How can restaurant sales data be used to validate the prototype developed?

1.5. Justification

The process of profit maximization has not been easy, considering that there are several factors affecting the fluctuations in sales. These factors inform patterns in sales which in turn determine the profit at given times of sales. The goal of this research is to come up with a model that automates this process and makes it easy and understandable to take actions that help in the betterment of the returns (McDonald & Wilson, 2016). The results from this research assist restaurant managers in the process of pricing of the products in relation to various parameters that influence the price and come up with the best pricing which results in the maximization of the profit.

The proposed system uses the business influence concepts to manipulate and aggregate the data on demand to forecasting consequent demand levels for different business items. After receiving this electronic information, identifying products' plurality and product cost information, the customer price sensitivity and logical relationships between products' gross profit margins and the price sensitivity of customers are determined for the products.

After receiving the data identifying plurality pattern of products and product cost information, the sensitivity of customers to price and logical sensitivity between the gross profit margins and this customer price sensitivity, the control system then receives sales data from the storage system and generates data showing price change, including a changed price of an identified product based on the sales data received from one of the virtual and physical storage systems. The data showing price change is then transmitted by the control system the physical and virtual store systems to consequently change the price of the identified product. Thus, it is capable of price optimization and consequently, optimizes profit.

Finally, the results of this research contribute to both the local and international literature bodies and are useful to academicians and the future. It should then narrow the gap in the existing literature by providing techniques which are based on new perspectives of ways to apply Information technology methods to facilitate financial management and maximize profit.

1.6. Scope

Profit maximization is wide and covers including others, stocking and store management, product process management, marketing plan and pricing of products. This research is focused on the process of pricing to achieve optimum equilibrium on demand and supply.

The process of pricing itself involves several inputs and predictions considering other factors both internal and external, this research is majorly focused on the use of data mining algorithms retrospectively on the already existing data and over the specific time frame to discover the sales patterns which in turn helps in deciding on optimum sales and consequently maximize profits. The process of profit maximization can be applied in several business setups by the financial advisors, business managers, and other entrepreneurs, this research is specific to restaurant sales data for now.

1.7. Limitations

This research is limited by time constraints as per the academic requirements. This is mitigated by focusing the scope to the pricing process and to the restaurant specifically and by use of already existing data, the requirements for the research are easily met.

Another limitation may be financial which may limit the sources of data in case of modeling which may require tools that can be costly, it mitigated by the use of the available resources are open source.

Data availability and access may also be a limitation. To get a restaurant's sales data may not be easy from any other restaurant. This has been mitigated by requesting and acquiring necessary approvals for the sales data of Kula Kona restaurant located in Hurlingham Nairobi.

Chapter 2: Literature Review

1.8. Introduction

The data collected from a supermarket's point of sale during research in the University of Waikato which listed items purchase trends showed a pattern in which the purchases are done. Some goods are notably purchased together when the timestamp of purchase and the point of sale where the sales were made data are analyzed (Mehrotra & Khan, 2015). This led to the development of interest in the business world and further research was done on how to improve the quality of this data to be used in decision making. Several paradigms were discovered including data mining methodologies that are applied in this and used in different applications.

Data Mining, which is also known as Knowledge Discovery in Databases (KDD), is a process to find trends, patterns, correlations, anomalies in these databases which can be used in making accurate future decisions. Data Mining is an intersection of the fields of Databases, Artificial Intelligence, and Machine Learning. Market Basket Analysis also called Association Mining is the major focus of this context. Market Basket Analysis (MBA) is the discovery of relations or correlations among a set of items (Nengsih et al., 2015).

For example, in the case of a supermarket, items are clustered according to the combinations or criteria of which they are purchased together by a customer at a given time(Basket), this ordering is used in marketing in the sense that, the previous MBA data is used to arrange the items in the shelves such that a customer is persuaded to select a group of "related" items without having to make several stops in the counters (Gatzioura et al., 2015). Apriori algorithm is applied in this methodology as the model used to predict the patterns and abstract the data (Agrawal et al., 1993), using a mathematical formula which consequently helps in decision making.

Data mining process is a computation that involves exploring and discovering patterns in large datasets which are also known as big data. It is a field of computer science that combines many fields including machine learning, statistics, and data science (Piatetsky et al., 1996). The

term Data mining was derived this time from the then known term Knowledge Discovery in Databases as it was applied most by the financial community in analyzing data.

2.1. Factors Determining Profit in Restaurant Business

i. Labor Costs

This is the way time and resources are utilized towards production in the restaurant (Fritschel et al., 2013).

ii. Food Waste

This majorly focuses on the quantity of food served to customers; they might be too much to the point where too much food is thrown away as excess (Fritschel et al., 2013).

iii. Employee Theft

This can be on the side of employees, they might be selling meals and failing to record or account for; this is mostly mitigated by having a chain of service with a proper information system (Fritschel et al., 2013).

iv. Employee Turnover

Employees' personalities might be the reason for keeping some of the customers to your business, once they leave, and new ones come, some customers may find it hard to continue coming to the same place (Fritschel et al., 2013).

v. High Wait Times

How long customers have to wait on a table or to be served food at your restaurant will affect your profit, since it determines they are coming back next time (Fritschel et al., 2013).

2.2. Profit Maximization

Profit maximization is the process where a business enterprise determines the price, input and output levels that lead to the optimum profit (Pettinger et al., 2017). Maximization is a process of knowing if your business is growing or diminishing, and therefore, it should be taken with utmost seriousness. There are different ways of profit maximization most of which have over time proven to be too cumbersome and the knowledge is centralized within individuals.

Figure 2.1 illustrates Profit Maximization with an assumption in classical economics that firms seek to maximize profits.

- i. Profit = Total Revenue (TR) - Total Costs (TC). Therefore, profit maximization occurs at the gap between total cost and total revenue.
- ii. A business enterprise can maximize profits only if it produces at an output where marginal cost (MC) = marginal revenue (MR)

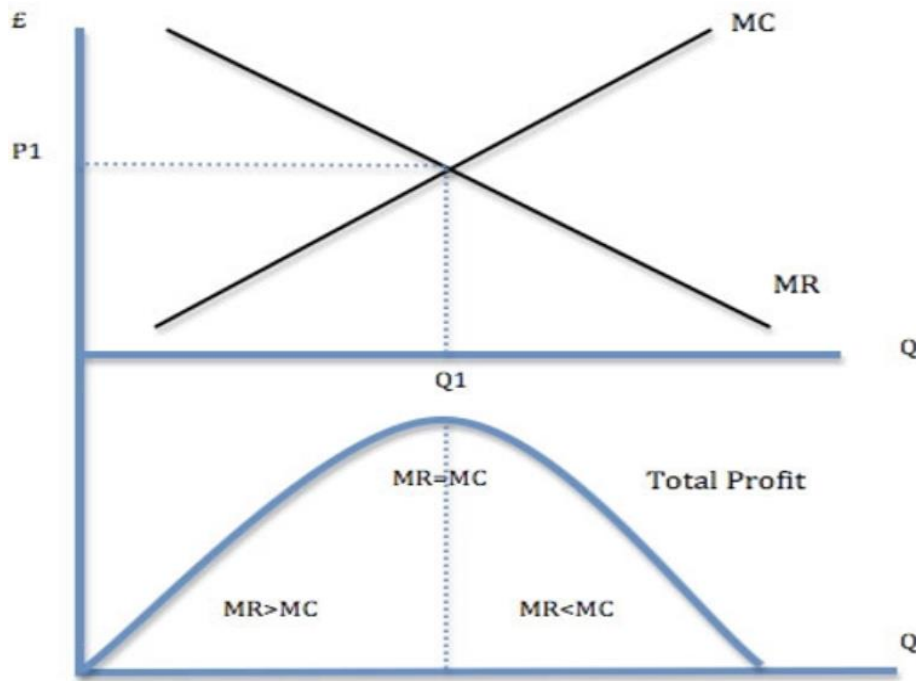


Figure 2. 1: Diagram of Profit Maximization (Pettinger et al., 2017)

To illustrate Figure 2.1 further:

- i. If the business enterprise produces less than Q_1 , marginal revenue (MR) is greater than marginal cost (MC). Therefore, for this extra revenue, the firm business more revenue than it pays in costs, and gross profit will increase.
- ii. Close to Q_1 , marginal revenue (MR) is only just greater than marginal cost (MC); therefore, there is a very small increase in gross profit, but actually, it is still rising.

- iii. However, after Q1, the marginal cost of the output becomes greater than the marginal revenue. This can mean the firm will experience a fall in its gross profit level because the cost of the extra units will be greater than revenue.

2.3.Importance of Profit Maximization

After coming of goods into the market and investments are made, profitability is key in measuring investment output by observing different criteria and key variables which consequently informs the business decisions (Kolari & Anari, 2011). Profitability does not determine the level of investments only but also the level of output from the business. Profit rates also inform the decision on the type of business investment to involve in and what factors to consider and adjust. After making capital investments, the business enterprise combines the fixed stocks of capital with various labor quantities to attain the required profit rates. Basing this research on this, we intend to come up with an automated way of profit maximization through the process of data mining and pattern analysis to come up with an up to task model.

2.4. Apriori Algorithm (Associate Rule Mining)

The Apriori algorithm is the greatest contributor in the history of association rules mining techniques; it was first discovered by Agrawal (Agrawal & Srikant, 1994). It is a straight forward approach that runs on the item tables in database, using the association rules in generation of support, lift and confidence then storing the counters of each candidate while with their frequencies. The Apriori is more efficient during the candidate generation processes for two reasons; Apriori employs a different candidate item generation method which is a new pruning technique.

These association Rules find all sets of items that have supports greater than the minimum support and then using the large sets of item to generate the desired association rules with confidence greater than the set minimum confidence. Lift of a rule is the ratio of the observed support to that expected if X and Y are independent. The typically and widely employed example of application of association rules is Market Basket Analysis (MBA) (Sayyad et al., 2016)

Figure 2.2 illustrates the Apriori Algorithm and it can be deduced that the Apriori algorithm takes uses the fact that any subset of a frequent itemset in the transaction is also a

frequent itemset. This algorithm can, therefore, reduce the number of considered candidates being by exploring the itemset whose support count is greater than the minimum support set. All infrequent itemset is pruned if its subset is infrequent.

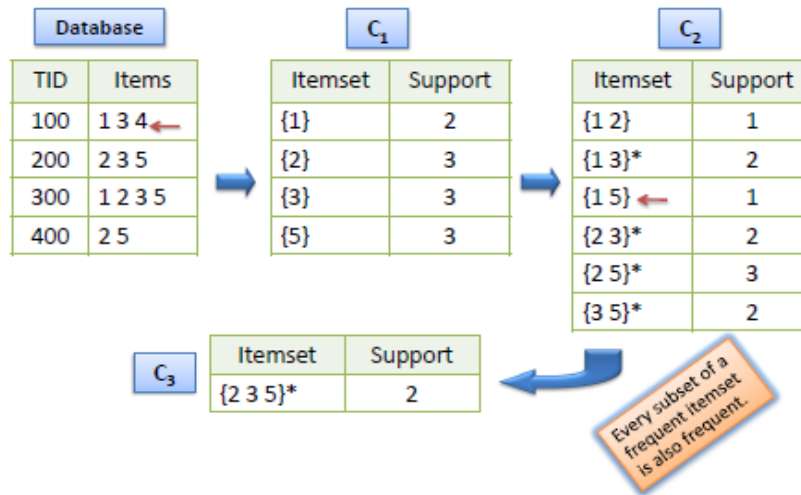


Figure 2. 2: Diagrammatic Explanation of the Apriori Algorithm (Sayyad et al., 2016).

2.5. Profit Maximization Elaborations Example

Table 2.1 elaborates on how profit maximization is calculated. The maximum profit is at the point where marginal revenue equals marginal cost, in this case, it is at the point where the quantity sold is 2 and 3 at the prices of 15 and 12 respectively. This is the point where the profit will be equal to 20.

Table 2. 1: Profit Maximization Elaborations Table (Pettinger et al., 2017)

PROFIT MAXIMIZATION EXAMPLE TABLE						
Product Quantity (Q)	Price(Pr)	Total Revenue (TR)=(Q*Pr)	Total Cost (TC)	Profit (P)=(TR-TC)	Marginal Revenue (MR)= $\Delta TR/\Delta Q$	Marginal Cost (MC)= $\Delta TC/\Delta Q$
0	-	0	4	-4		
1	18	18	6	12	18	2
2	15	30	10	20	12	4
3	12	36	16	20	6	6
4	9	36	24	12	0	8
5	6	30	34	-4	-6	10
6	3	18	46	-28	-12	12

2.6. Conceptual Framework

Figure 2.3 illustrates how the system uses different parameters like price, time of sales, foods sold together at the same time to discover the sales pattern which in turn informs the decision to be made in the price optimization process and consequently profit maximization.

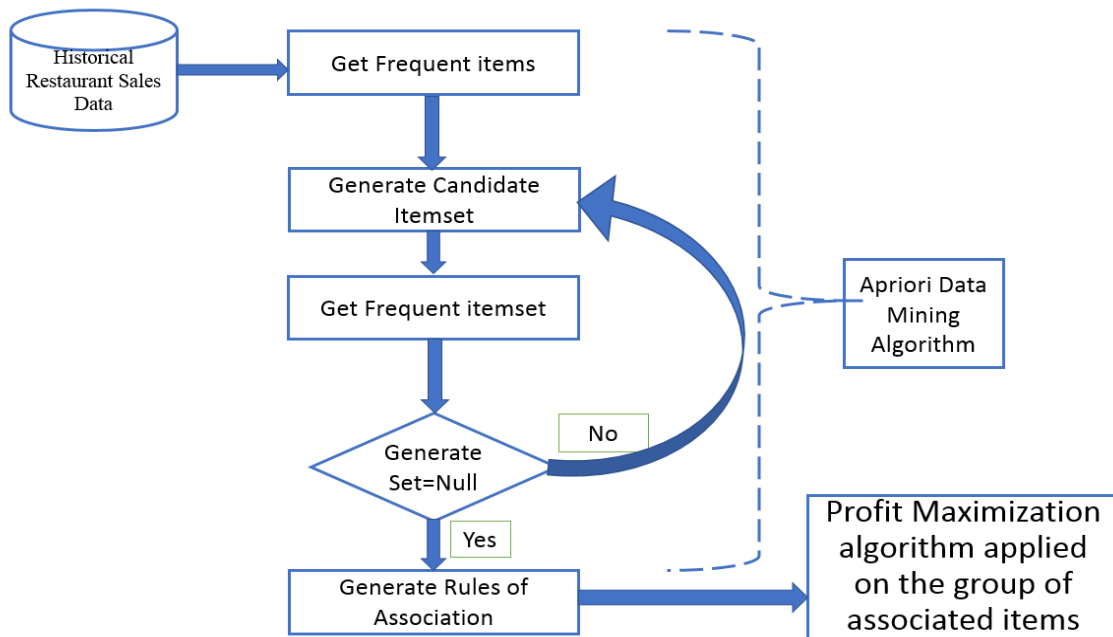


Figure 2. 3: Diagrammatic Representation of the Model.

The data is from Kula Kona restaurant's retrospective data of a time that is available in the database. The reports generated by the systems are used by the scholars as input in more research in the same field.

- i. Candidate itemset is generated by use of only the large itemset of the previous pass without database transactions considerations.
- ii. The previous pass' large itemset is joined with itself to generate all itemset whose size is higher by 1 unit.
- iii. Each generated itemset with a subset that is small is deleted. The remaining itemsets become the candidate ones.

Chapter 3: Research Methodology

3.1. Introduction

The methodology for carrying out this research putting into consideration the objectives set in chapter one is outlined in this chapter, explanation of processes and steps of this study as well as the methodologies which have been used in the related works. The research design, data collection methods, and analysis techniques, tools for developing algorithms are outlined in this chapter.

3.2. Research Site

The study uses historical sales data collected between 01st January 2017 and 31st December 2018 at Kula Kona restaurant located in Hurlingham Nairobi. The date is recorded automatically during sales in the restaurant and stored in the restaurant database located in the same facility. It majorly contains prices of goods which are food for this case at a given time-point, time of sales, the number of sales, time & date, frequency distribution of sales and sales distribution across different products.

3.3. Research Design

The research design is an orderly methodology that an analyst uses to lead a logical report. There are four principal types of quantitative research structures; graphic, correlational, semi-trial and test. The contrasts between the four types basically identify with the level of the analyst plans for control of the factors in the analysis. Following is a depiction of each sort of quantitative research design just as outline looking into the methodologies (Creswell et al., 2008).

Experimental Designs, often known as true experimentation, utilizes the logical strategy to build up the cause-impact relationships among assembly of factors in an exploration

contemplate. Specialists try to control all factors apart from the one being controlled (the autonomous variable). The impacts of the free factor on the dependent variable are collected and examined for existence of a relationship (Sullivan et al., 2005).

Data mining approach was used in the study. The built model was run on the retrospective data to find the patterns and associations. The system was validated using new cases to ascertain its performance in the real-life situation.

3.4. Data Acquisition

The Apriori Algorithm proposed for this case operates on historical data which is in this case retrospective sales data of the restaurant. Since this is the primary data for analysis, sales of different food commodities are recorded automatically by the Point-of-sales system installed in the restaurant and the payment section. This data contains the items sold, timestamp, date, and items bought together on the same bill. The algorithm was applied to discover the sales patterns to help in making decisions using the same data (Roffee et al., 2016).

Apriori algorithm is an algorithm for mining frequent itemset mining and association rule learning on transactional databases. It identifies the frequent individual items in the database and extending them to different larger item sets as long as those items appear sufficiently frequent in the database. Apriori uses a bottom-up approach, where frequent subsets are extended one by one, a step is known as candidate generation, and different groups of candidates are subjected against the data. The algorithm terminates automatically when no further successful extensions are found (Garraghan et al., 1946).

Apriori algorithm uses a breadth-first search and a Hash tree structure to count candidate itemsets efficiently. It then generates candidate itemsets of length k from itemsets of length $k-1$. It then prunes the candidates which have infrequent sub-patterns. According to the downward closure lemma, the candidate set contains all frequent k -length itemsets. After that, it scans the transaction database to determine frequent itemsets among the candidates.

The pseudo-code for the algorithm is given below for a transaction database T , and a support threshold of ϵ . The usual set-theoretic notation is employed, though note that T is a multiset. C_k is the candidate set for level k . At each step, the algorithm is assumed to generate the candidate sets from the large itemsets of the preceding level, heeding the downward closure lemma. $count[c]$ accesses a field of the data structure that represents candidate set c , which is initially assumed to be zero. Many details are omitted below, usually, the most important part of the implementation is the data structure used for storing the candidate sets, and counting their frequencies.



```

Apriori( $T, \epsilon$ )
   $L_1 \leftarrow \{\text{large 1-itemsets}\}$ 
   $k \leftarrow 2$ 
  while  $L_{k-1} \neq \emptyset$ 
     $C_k \leftarrow \{c = a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a, \{s \subseteq c \mid |s| = k-1\} \subseteq L_{k-1}\}$ 
    for transaction  $s \in T$ 
       $D_s \leftarrow \{c \in C_k \mid c \subseteq s\}$ 
      for candidates  $c \in D_s$ 
         $count[c] \leftarrow count[c] + 1$ 
     $L_k \leftarrow \{c \in C_k \mid count[c] \geq \epsilon\}$ 
     $k \leftarrow k + 1$ 
  return  $\bigcup_k L_k$ 

```

Limitations

Apriori algorithm, while historically significant in many areas of application, suffers from a number of inefficiencies and trade-offs, which have lowered its competitiveness against other algorithms. Candidate generation generates large numbers of subsets (the algorithm attempts to load up as many as possible candidate set before each scan). Bottom-up subset

exploration (essentially a breadth-first traversal of the subset lattice) finds any maximal subset S only after all $2^{|S|} - 1$ of its proper subsets.

The algorithm scans the database several times, thus reduces the overall performance. Due to this, the algorithm assumes that the database is Permanent in the memory.

Also, both the time and space complexity of this algorithm are very high: $O(2^{|D|})$, thus exponential, where $|D|$ is the horizontal width (the total number of items) present in the database.

3.5. Design and Development Approach

The Software Development Life Cycle (SDLC) model picked for this task is the Waterfall Model. The SDLC procedure has different stages to it. SDLC in Figure 3.1. is a procedure pursued by a product venture, inside a product association. It comprises an itemized arrangement depicting how to create, keep up, supplant and adjust or improve explicit programming. The existence cycle characterizes a system for improving the nature of programming and the general advancement process (Lesage & Dieter, 2009). The accompanying Figure 3.1 is a graphical portrayal of the different phases of a common SDLC.

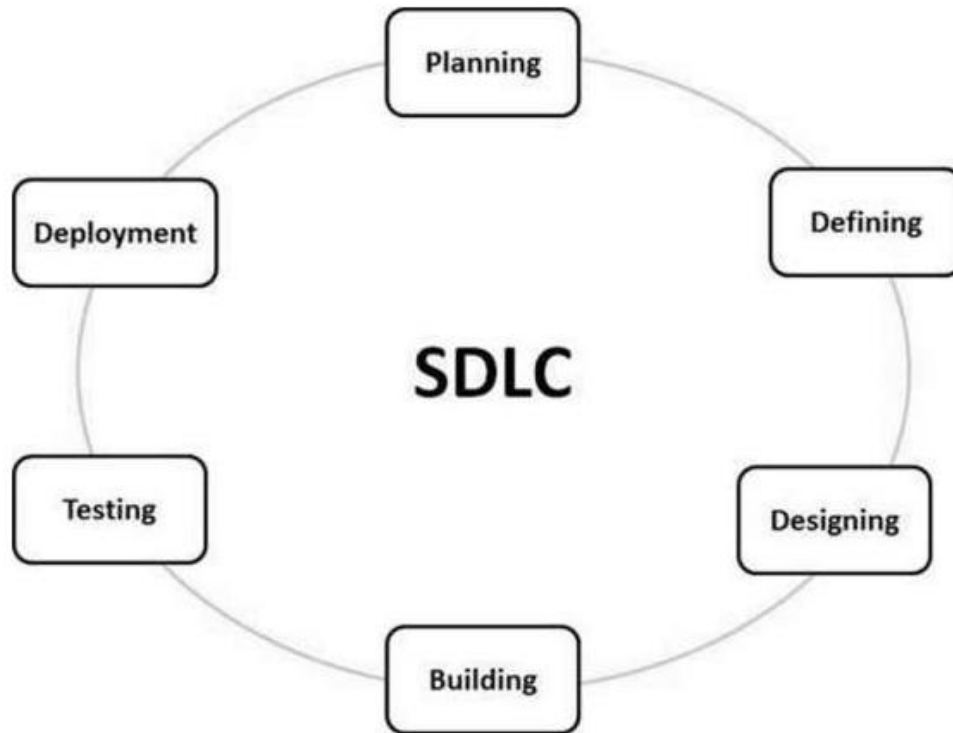


Figure 3. 1: SDLC Stages (Lesage & Dieter, 2009)

3.6. Waterfall Model - Design

Waterfall Model was the first SDLC Model to be used generally in Software Engineering which guaranteed the achievement of the tasks. In the Waterfall approach, the entire procedure of programming advancement is divided into independent states (Klein & Joh, 2010). In this Waterfall show, commonly, the result of one stage contributes to the following stage consecutively. Figure 3.2 is a representation of the different phases of the Waterfall Model.

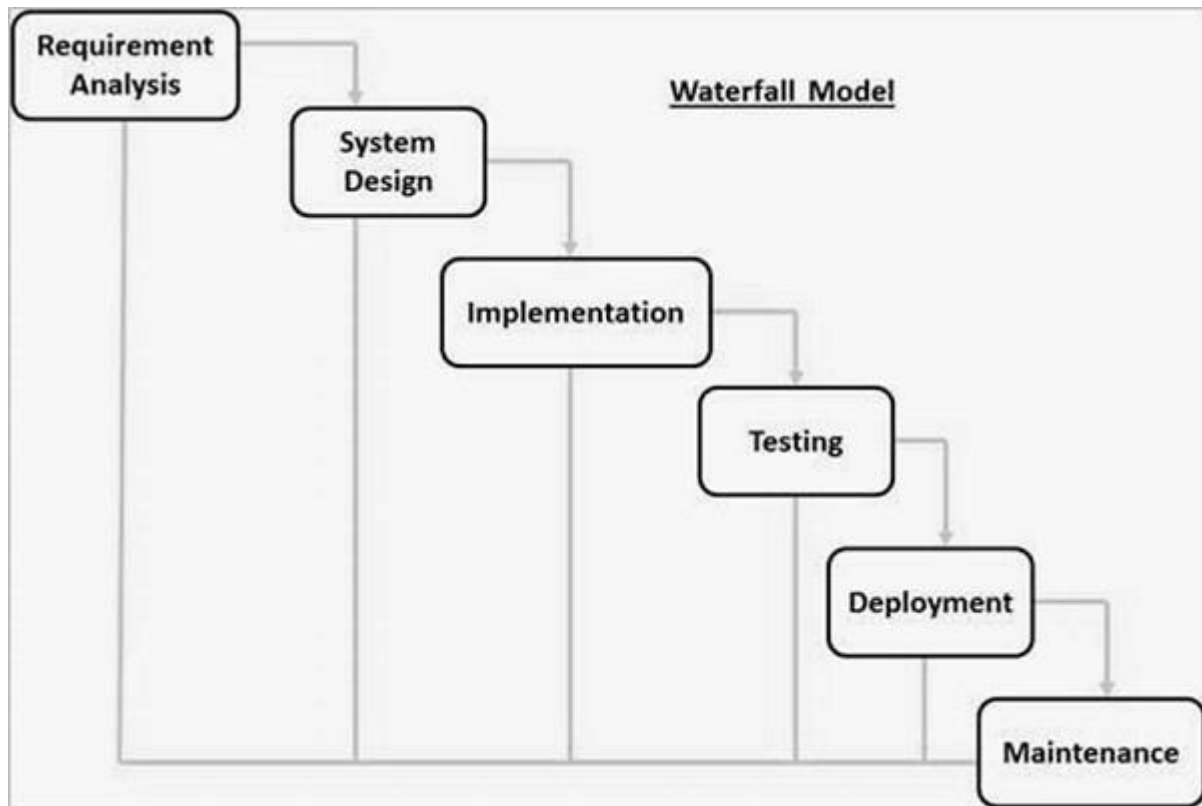


Figure 3. 2: Waterfall Phases (Barbour, 2001)

The sequential phases in the Waterfall model are:

- i. **Requirement Gathering and analysis:** All conceivable necessities of the framework to be created are caught in this stage and recorded in a prerequisite particular archive. At this point, the system development requirements such as software like python libraries and frameworks, the development computers and the data used. The data was acquired from Kula Kona restaurant upon formal request for their retrospective data. The data was then subjected to review if the variables required are all captured; the data was then cleaned and sorted out to fit the required set. Python is an off-the-shelf programming language which was freely downloaded plus all its libraries for the development of this system.
- ii. **System Design:** The prerequisite determinations made from the principal stage are contemplated at this stage and the configuration of framework is done. This configuration of framework helps in equipment indication and framework requirements

which aides in characterizing the general framework creation. The system was developed in the Python development framework, the data formatting scripts were developed first which helped in creating the required variables which were auto-calculated from the existing set by scripting formulas into python functions. Then out of this, the Apriori algorithm was developed which was the core codes for the system which maximized the profit of several items at the same time. The testing was done after developing the scripts to ensure that the output data is what was desired as the output.

- iii. **Implementation:** Using contributions from the initial plan, the framework is initially created in units, which are small projects, which are included into the following stage. Every created unit is tried for its viability, which is intended to be Unit Testing. The developed system was tested by subjecting it to a set of training data then to test data to ensure it is able to learn and imply the algorithm on the data and produce the desired results. The results were awesome, and each unit of the system was then debugged and put into use.
- iv. **Integration and Testing:** All the units created in the execution stage are coordinated into a framework in the wake of testing of every unit. Post-mix the whole framework is tried for any shortcomings and disappointments. The system was developed in modules and each of these modules was tested to confirm if it achieves the desired output, after testing each module, the modules were then integrated together into one system which is then tested as a whole system.
- v. **Deployment of the system:** Once the practical and non-useful testing is done; the item is conveyed in the client condition or discharged into the market. This system was deployed into the market where it was taken to Kula Kona restaurant and the current fixed duration data was uploaded into the system then the algorithm was tested and confirmed to be working as expected. The results were recorded and used as part of deployment documentations.
- vi. **Maintenance:** There are some few issues that come up to the customer condition. Patches are discharged to fix those issues. Additionally, to upgrade the item some better forms are discharged. Support is done to convey these adjustments in the client's

condition. The issues which came up during the deployment of the system were resolved on the spot and continuous system maintenance was done as the system was already put in to use.

All these phases are cascaded to each other in which progress is seen as flowing steadily downwards (like a waterfall) through the phases. The next phase is started only after the defined set of goals is achieved for the previous phase and it is signed off, so the name "Waterfall Model". In this model, phases do not overlap (Borgdorff & Henk, 2012).

3.6.1. Justification for Choosing Waterfall

The upsides of the waterfall model are that it thinks about departmentalization and control. A timetable can be set with due dates for each period of progress and a thing can proceed through the headway methodology exhibit arranges one by one. Improvement moves from thought, through an arrangement, use, testing, foundation, researching, and ends up at action and upkeep. Every time of progression proceeds in the exacting solicitation. The waterfall approach has the accompanying points of interest; Simple and straightforward and use, simple to oversee because of the inflexibility of the model each stage has explicit expectations and an audit procedure, Phases are handled and finished each one, in turn, functions admirably for littler ventures where necessities are (Creswell & John, 2014).

3.7. Data Management

Data is collected, put away in SQL databases; the information is then extricated into Excel exercise manuals, which are then changed over to comma-isolated esteem (CSV). The CSV documents are level records and effectively coherent by both humans and into investigative apparatuses.

Leading data management platforms enable undertakings to use Big Data from all information sources, progressively, to consider increasingly successful commitment with clients, and for expanded client lifetime esteem (CLV). Information about the executive's programming is fundamental, as we are making and devouring information at remarkable rates.

Top information the board stages give undertakings and associations a 360-degree perspective on their clients and the total permeability expected to increase profound, basic bits of knowledge into buyer conduct that give marks an aggressive edge. When data is received from Kula Kona Restaurant in this format, the data preparation scripts are then run to make the data have the desirable columns for mathematical calculations. After this, the algorithm operates on the data to give the desired outcomes which are used for decision making.

3.8. Data Readability and Validity

Information intelligibility is the marker of a proportion of interior consistency (Chall and Dale., 1995) in which a measure is coherent when diverse endeavors give similar outcomes while legitimacy alludes to the exactness of a measure (Habib, Pathik & Maryam., 2014). The eatery's business information is put away in a configuration that pursues the set down information norms which makes it simple to peruse through various data the executives programming and sharing. There is a few information control programming that guarantees the right information organizations and consistency among parameters and perceptions.

3.9. Ethical Considerations

The data used for this study is readily available in the Kula Kona restaurant servers and a due process is followed in the acquisition of the same. This involves a formal request for the data, protocol approvals, and controlled use of data by the Kula Kona management. Kula Kona management.

Chapter 4: System Analysis, Design, and Architecture

4.1. Introduction

As this research came up with a model to expand benefit utilizing Apriori calculation, it is important to clarify the key capacities it meets just like the real information source. This section centers on dissecting the ideal calculations, recipes and distinctive arithmetical prerequisites and qualities of the information that was QA utilized. In this manner, the combination of all these into a plan that manages the usage of the calculation to the expected benefit amplification is critical (Michael et al., 2012).

Analysis and design were got from known standard practices which help in understanding what the definite need is, propose suitable strategies for advancement and actualize as indicated by the information holes.

4.2. Requirement Analysis

4.2.1. Functional Requirements

- i. The system must accept data in CSV format.
- ii. The system must detect all the required raw column data.
- iii. The system must detect and eliminate empty data rows
- iv. The system must calculate and append the required columns in the dataset as per the formulas scripted.
- v. The system must display the associations and sales patterns from the dataset on the console.
- vi. The system must list all the points of maximum profit after optimization.

4.2.2. Non-Functional Requirement

4.2.2.1. Ease of Data Input

Inputting data into the system should be easy. The addition of new dataset samples requires just uploading and running the scripts. This makes the algorithm run smoothly and easily.

4.2.2.2. Robust Generalization

The system should generalize the new instances of the executions with no complexity and a reasonable bias-variance trade-off. A complex model-based system is often termed as overfitting and hardly generalizes well on new test cases.

4.2.2.3. Security

The system should be secure to prevent any unauthorized access to the model parameters and system functionalities. Any modifications or changes made to the model must be authorized by the system administrator to guarantee reliable system behavior and reliability.

4.2.2.4. Adaptation

The system should adapt gradually; constantly improving based on different exposure to different datasets. When subjected to different datasets of the same format, the system should be able to run and execute the scripts effectively and by this, it ensures the relevant response is given based on the unseen input instances.

4.3. System Design and Architecture

4.3.1. Partial Domain Model

Partial Domain Model is a system design diagram that captures the most important concepts of the proposed system. It includes entities and key objects interacting with the system. An instance of latter and former in this system context is composed of the end-user; medical expert, stained image, feature set representations and the model. In addition, the schematic also indicates the possible attributes at a high level. The Partial Domain Model is represented in Figure 4.1.

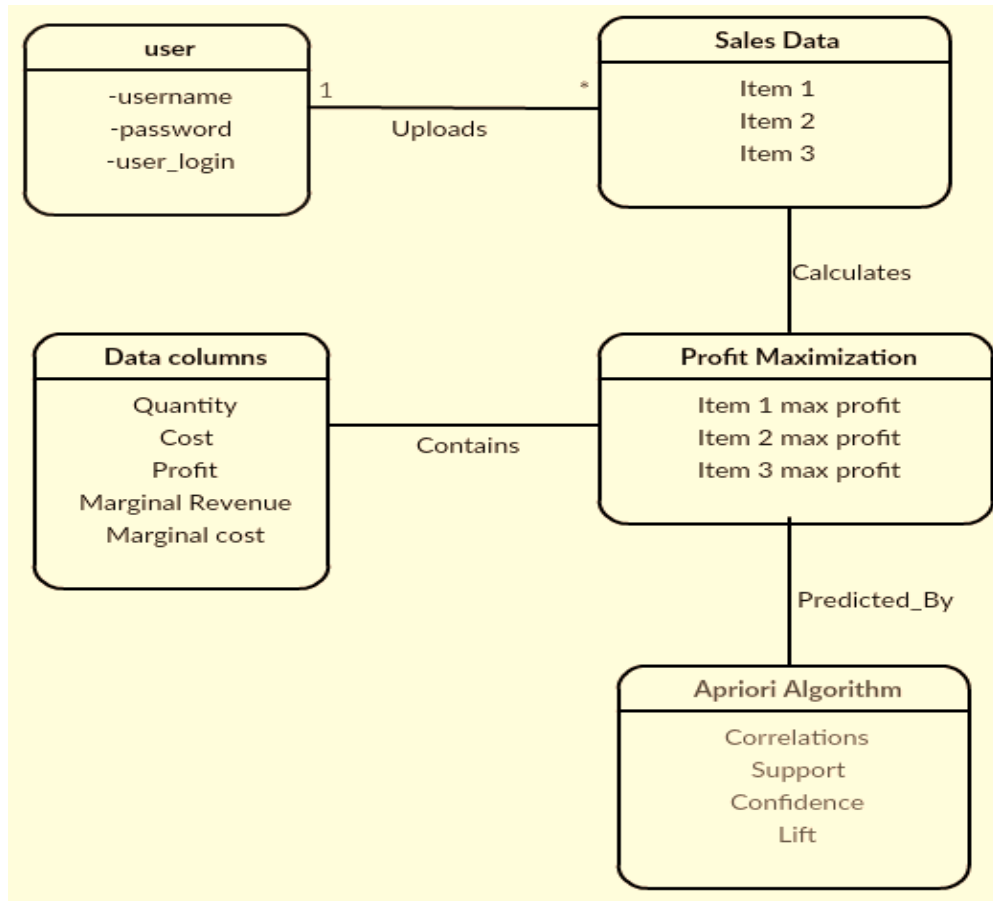


Figure 4. 1: A Partial Domain Model

4.3.2. Use Case Diagram

The use case diagram is a diagrammatic representation of the system and the actors interacting with the system. The primary actor is the restaurant financial planner/manager who is responsible for uploading the sales data into the system. The actors also interact with authenticating modules to validate his/her access to the system and generate use cases as well as the labeling process. System Use Case Diagram has been represented in Figure 4.2.

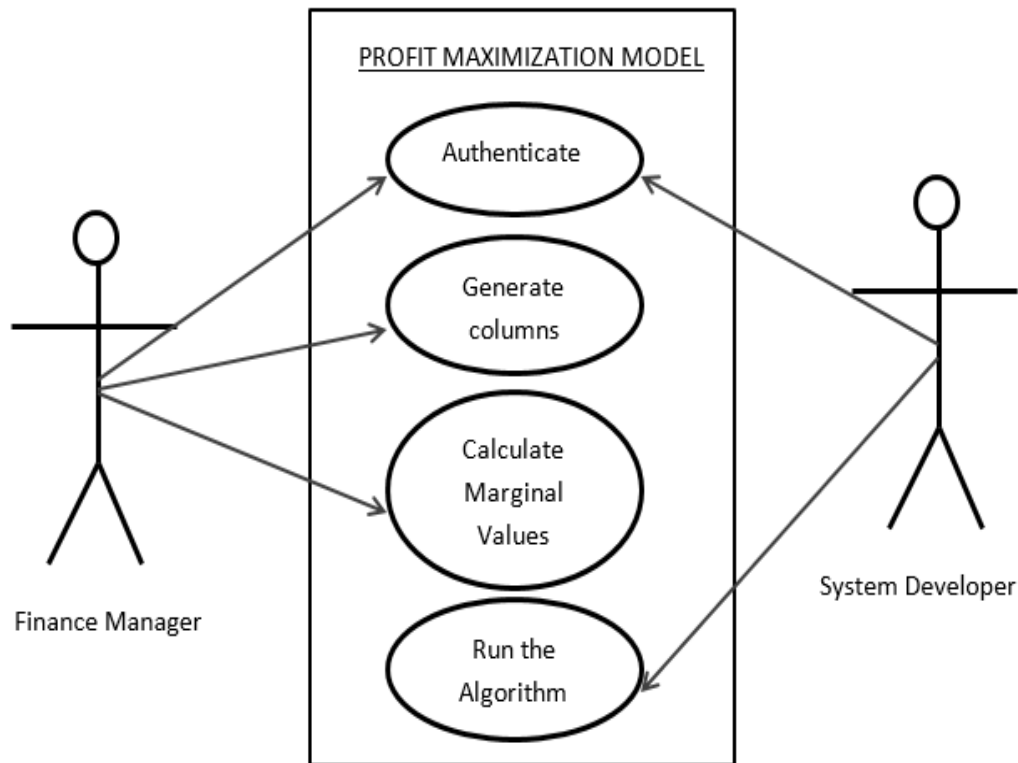


Figure 4. 2: Use Case Diagram

4.3.3. Use Case Basic Flows

This subsection highlights the success scenarios, primary actors, preconditions, and postconditions of different uses represented by Figure 4.2.

Use case: Maximize profit

Primary Actor: Finance manager

Pre-Conditions: data input must contain columns for Product, Revenue, Quantity, Cost and Gross margin.

Post-Conditions: The last two columns must be Marginal Revenue and Marginal cost respectively.

Basic Flow:

- a) The finance manager or user uploads the structured data.
- b) Data is sorted by item types.
- c) Variables required for profit maximization are calculated and appended to the dataset.

- d) The points at which profit is maximized are marked from the columns of Marginal Revenue and Marginal cost.
- e) This data is then merged together into one dataset with only maximum profit points.
- f) The data is then subjected to the Apriori algorithm which calculates Support, Confidence, and Lifts.
- g) This is then used to find the patterns of sales using Sales ID.
- h) The results are then displayed on the screen and are used in decision making.

Use case: Model update

Primary Actor: System Developer

Pre-Conditions: Restaurants' data with all the required columns must exist.

Post Conditions: All column names must be updated

Basic Flow:

- a) The developer loads model the data into memory.
- b) The developer then does a test run of the scripts on the raw data and checks the output CSV data if the columns are generated as expected.
- c) The developer then runs the Algorithm and checks the results.

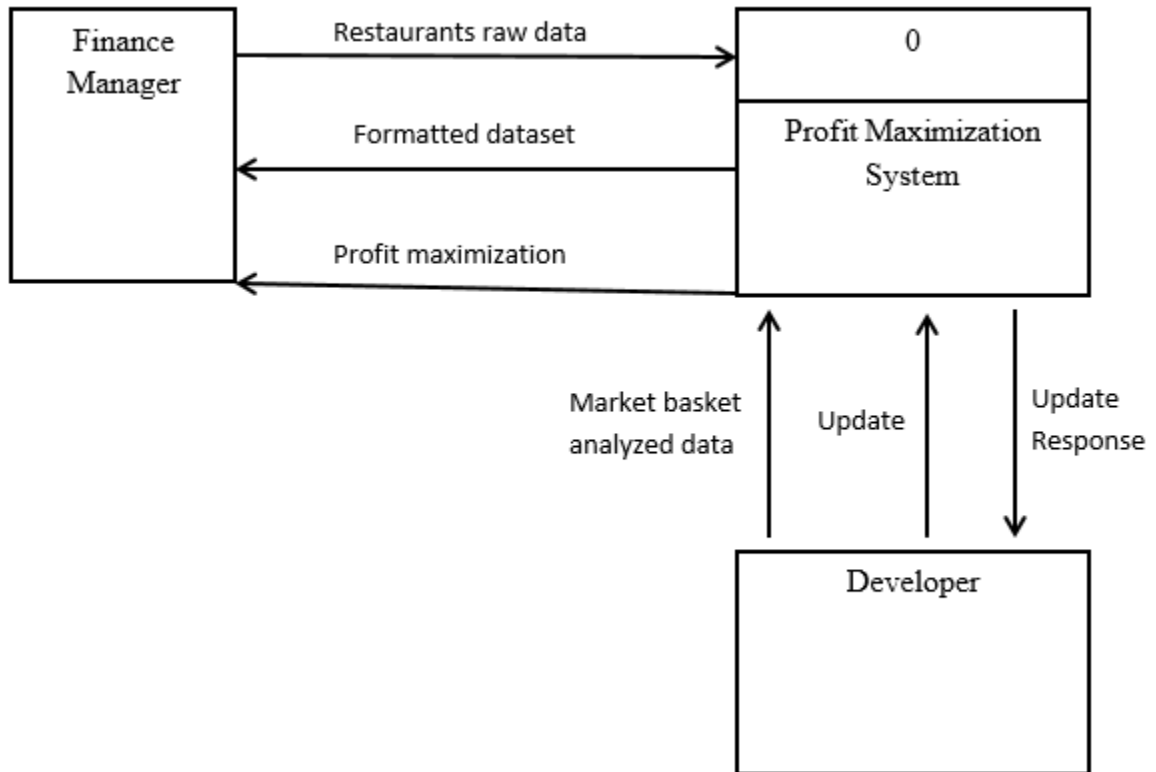
Extensions

- a) If a fails, the developer checks and update the data read code appropriately.
- b) If b fails, the developer checks the columns one by one on the dataset and identifies the missing column then modifies the codes accordingly.
- c) If c fails, the developer modifies the algorithm appropriately and re-run the model.

The Data flow diagram (DFD) diagrammatically explains how data flows in the system. It captures the external entities interacting with the system, for example, the user and the developer. It also shows all the processes involved in the system for instance uploading data, sorting, profit maximization and running the Algorithm. Lastly, the data flow diagram depicts data stores' interaction with the system processes and entities.

4.3.4. Context Diagram

The context diagram contains high-level data flow between the major processes plus the inputs and outputs involved. Figure 4.3, represents the context diagram for the profit maximization system and the interactions with the associated systems in the organization.



VT OMNES VNVM SINT
Figure 4. 3: Context Diagram

4.3.5. Level 0 Data Flow Diagram

The Level 0 DFD below depicts all the processes, data flows and data stores and the interactions involved within the system. It also shows the relationship between the system and the users. The outputs and the inputs of the system are also shown in Figure 4.4.

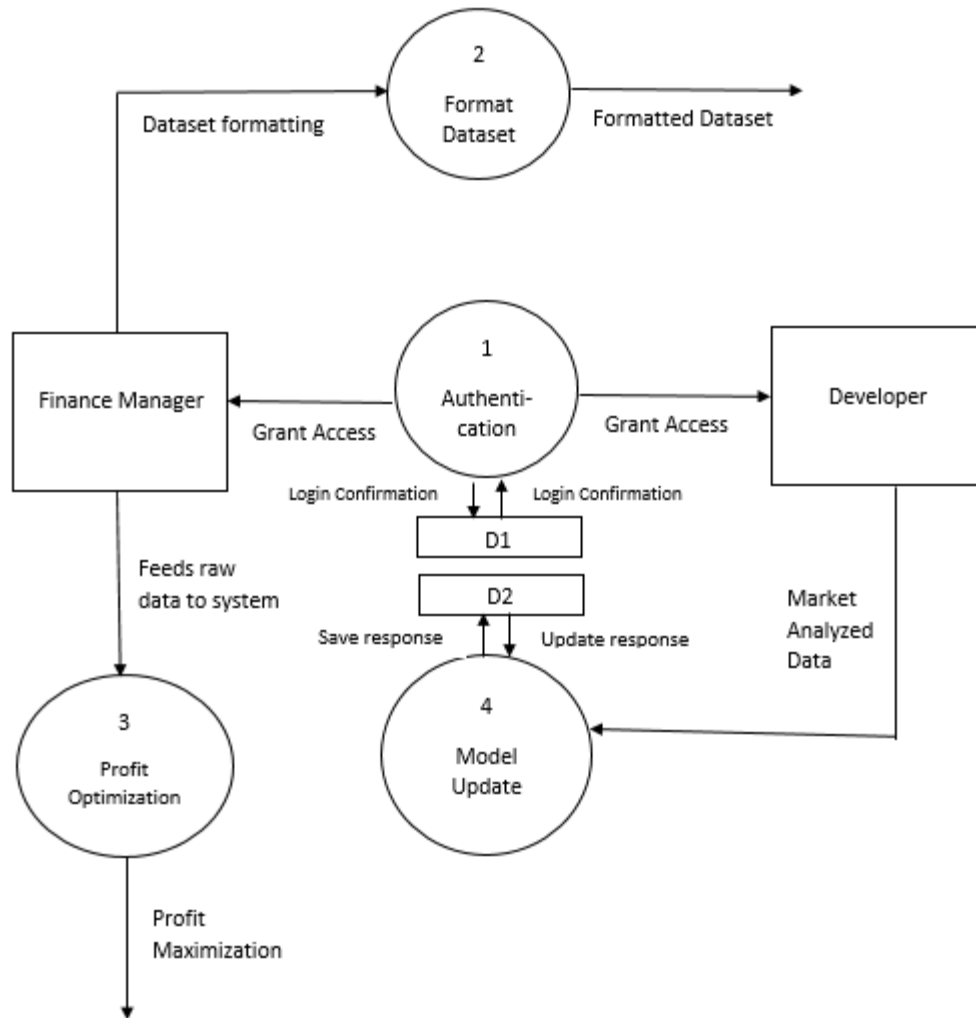


Figure 4. 4: Level 0 Data Flow Diagram

4.3.6. Sequence Diagram

The sequence diagram in Figure 4.5 shows how objects interact when arranged in a time sequence depicts the objects and classes used under this scenario and the sequence of messages exchanged between the objects used to carry out the functionality of the scenario. The system

contains seven objects namely; the user, raw data, preprocessor, train data, save data and optimize profit. Each of these encapsulates their properties and methods.

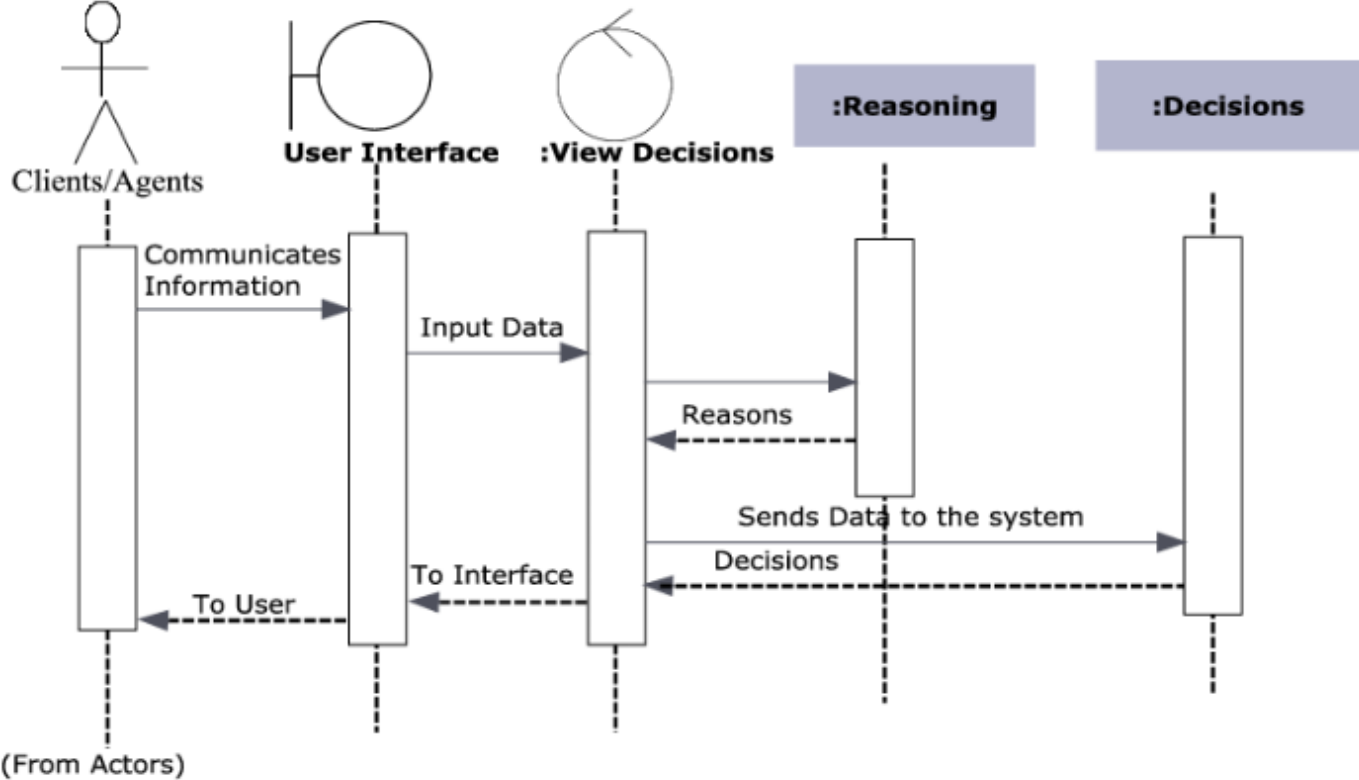


Figure 4. 5: Sequence Diagram

4.3.7. Class Diagram

A class diagram in the Unified Modeling Language is a type of static structure diagram, also known as the Design Class Diagram (DCD) that describes the structure of a system by showing the system's classes, their attributes, operations, and the relationships among objects. Figure 4.6 shows the class diagram for the profit maximization model.

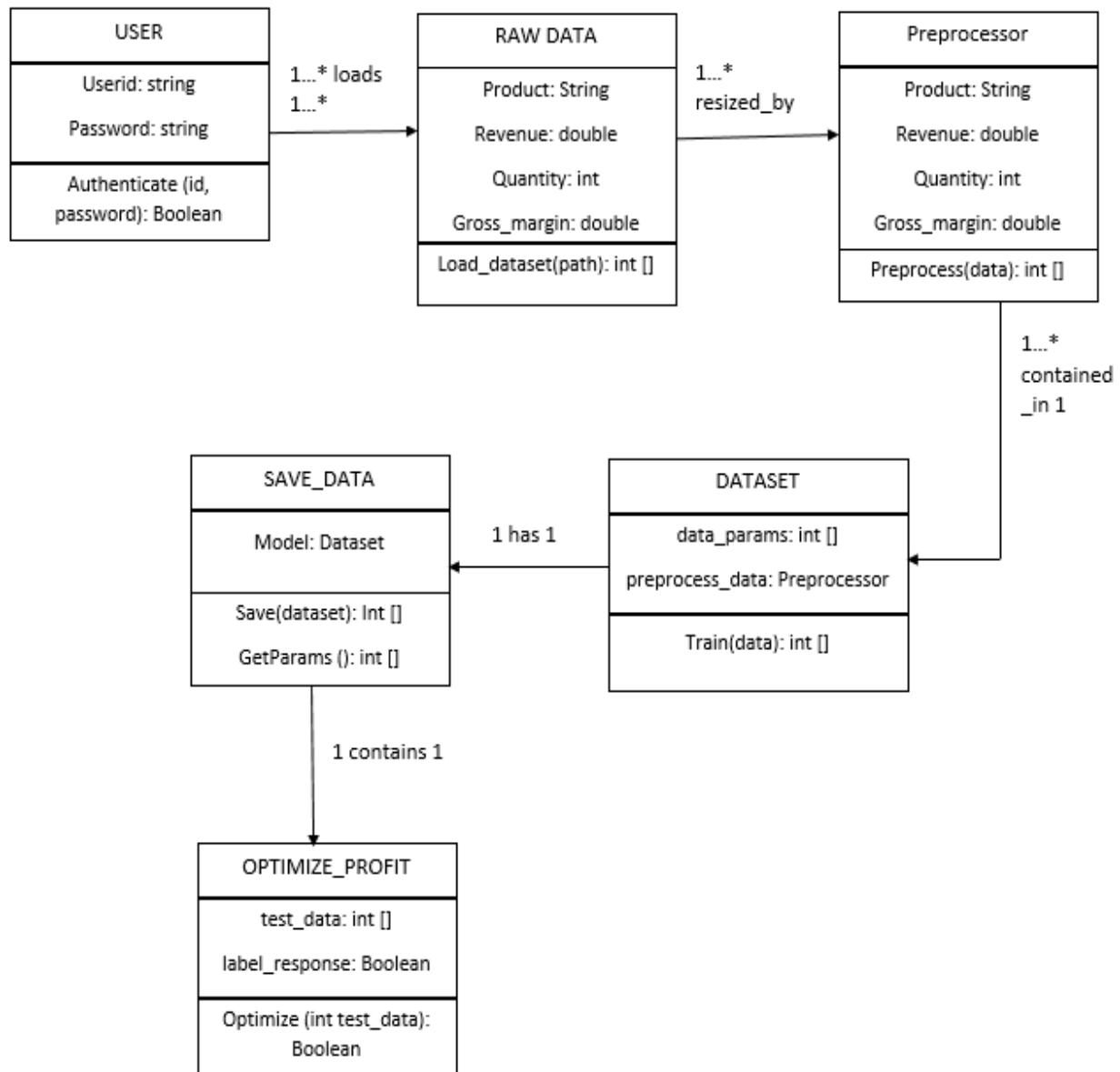


Figure 4. 6: Class Diagram

4.4. Algorithm Characteristics

4.4.1. Data Source

Data source defines where the data was got from. The data used to develop the proposed algorithm was a retrospective data collected by the Kula Kona Restaurant over the period between January 2017 and December 2017. This restaurant has a system in place which collects data and has them stored in a database, this is made possible because all the transactions going

on in the restaurant is done through the system and data is stored which is a reflection of the activities and the transactions and also include how the stocks run in the restaurant.

This data is stored in a PostgreSQL database which is where the restaurant management information system runs on. The data was received in encrypted comma-separated values file (CSV) which is one of the right formats to run Python scripts on using the Pandas library.

4.4.2. Data

The data is the actual figures and facts collected from over a period of time to be manipulated by the algorithm. The data collected contains the location where the data was collected from, the time it was collected, also included are values of measured weather conditions such as the number of sales, gross margin, and cost of goods.

To achieve the structure of the desired data for the research, several scripts are developed and run on the extracted data to calculate and add the desired columns. This is done through Python scripting which takes into consideration the formula for calculating different aspects of financial data and then appends excel columns on the dataset and adds the calculated data.

4.4.3. Algorithm Structure

The algorithm is divided into 3 main categories: required variable column calculations and addition module, profit maximization module and finally Apriori algorithm module. The first module is calculating the required columns from the available columns and appending to the dataset, for example, to calculate the marginal cost and marginal revenue of an item, total revenue and total cost must be calculated progressively across the dataset, thereafter profit is calculated, from these variables, marginal revenue, and marginal cost can be calculated.

The order of execution was:

- i. Different items were identified and clustered according to item type.
- ii. This was followed by separate calculations of marginal revenue and marginal costs of each individual item.
- iii. The data from different items were merged together into one big dataset as before but with additional fields.

- iv. From the marginal cost and marginal revenue, profit for each item was maximized at the point where Marginal Cost equals Marginal Revenue. At this point, each item has an identified point of profit maximization.
- v. Apriori algorithm is applied at this point where all the items in the dataset are subjected to the algorithm.
- vi. The algorithm then helps determine support, confidence and lift which are defined below, this help identify the patterns and association in sales between the items in the dataset.

Support: This is the percentage of transaction that contains all of the items in an itemset (e.g., pencil, paper, and rubber). The higher the support, the more frequently the itemset occurs. Rules with high support are preferred because they are likely to be applicable to more future transactions (Kohli & Suri, 2011).

Confidence: This is the probability of a transaction containing the items on the left side of the rule (as in the example above of Pencil and paper) contains as well the items on the right-hand side (as in a rubber above). Therefore, the higher the confidence level, the greater the likelihood that an item on the right-hand side is bought, the higher the return rate expected for a given set of rules (Kohli & Suri, 2011)

Lift is a probability that all the items in a rule occurring together divided by the items on the left-hand side's product probabilities and right-hand side occurring as there being no association between them. For instance, if Tea, Samosa, and Cake are sold together in 2.5% of all the transactions, Tea, and Samosa in 10% of transactions and Cake in 8% of transactions, then Lift would be calculated as $0.025/(0.1*0.08) = 3.125$. Lift of more than 1 suggests that the presence of Tea and Samosa increases the probability that Cake would also be bought in the same transaction. Overall, lift is a summary of strength of association between the goods on the left and right sides of the rule; the larger the lift, the greater the association between the two products (Kohli & Suri, 2011)

After identifying the sales patterns between the items in the dataset, and the point of maximum profits are also identified, it only takes minimum scripting to list items as grouped by the patterns and their optimum profit points and this is listed as the results of the algorithm

where a list of items is given with its calculated price in the combination of others, the data which the restaurant uses to price its products and maximize profit as show in Figure 4.7.

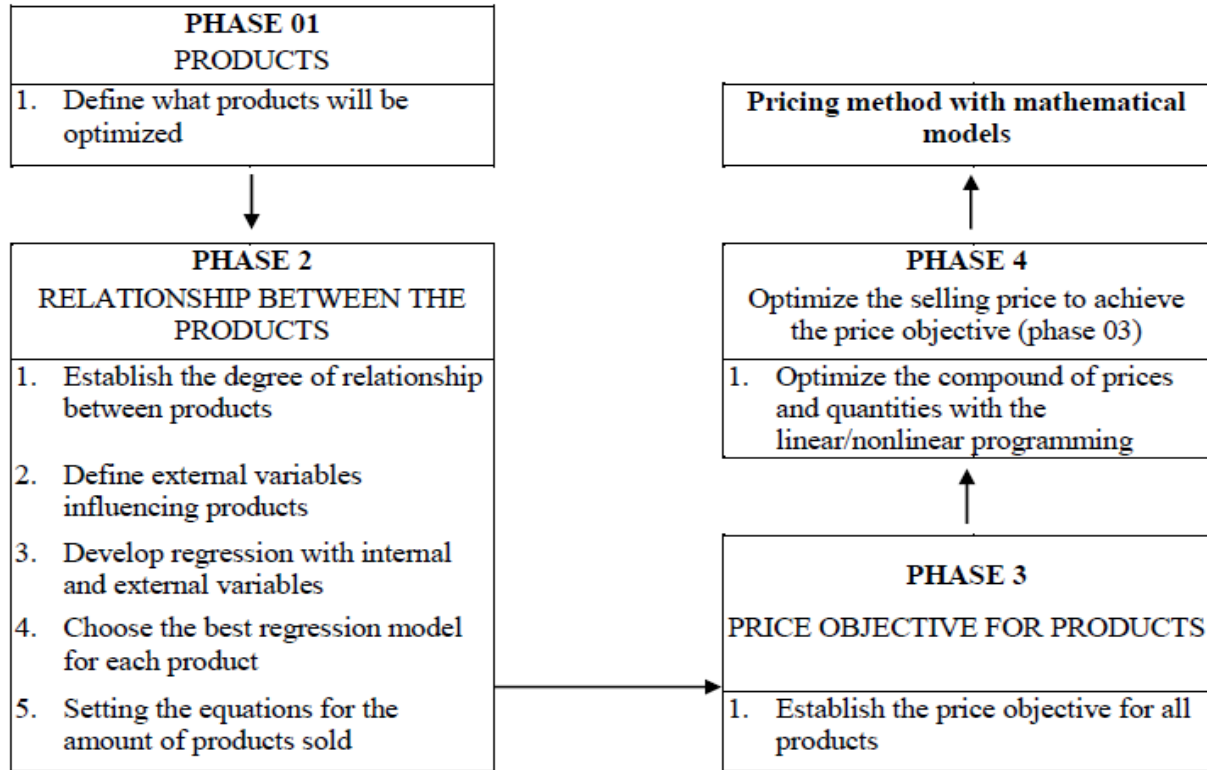
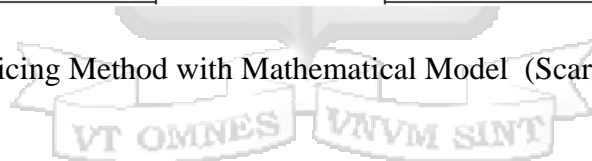


Figure 4. 7: Pricing Method with Mathematical Model (Scarpin, 2003)



Chapter 5: System Implementation and Testing

5.1. Introduction

Systems implementation is the route toward describing how the system was put into use, ensuring that the data framework is operational and used and ensuring that the data framework fulfills quality rules (i.e., quality affirmation). Framework Testing is an element of programming testing where aggregate and joined writing computer programs is attempted. The inspiration for driving this test is to survey the framework's consistency with the foreordained necessities (Han, Jiawei & Micheline, 2001). The section also entails describing the model selection based on the experiments which have been carried out, the selected model is used to build the prototype for this research work.

5.2. Hardware and Software Specifications

A software environment for a specific application could incorporate the working framework, the database framework, explicit advancement apparatuses or compiler. Hardware refers to the physical components of a PC likewise alluded to as the hardware or the gear of the PC that the product condition keeps running on. The product and equipment condition utilized for this case has been given in Table 5.1 below:

Table 5.1: Hardware and Software Specifications

ENVIRONMENT	SPECIFICATIONS
Software	Windows 7 Ultimate Operating System. Python 3.7 running on PyCharm JetBrains IDE community edition. Version 3.4 2018
Hardware	Dell computer

	64-bit core i7 4610M CPU
--	--

5.3. Dataset Generation and Preprocessing

5.3.1. Cost of Goods

The Cost of goods sold (COGS) is the conveying estimation of items sold amid a specific period. Expenses are related with specific items utilizing one of the few recipes, including explicit ID, first-in-first-out (FIFO), or normal expense. The expenses incorporate all expenses of procurement, expenses of transformation and different costs that are acquired in conveying the inventories to their present area and condition (Lovell & Michael, 1983).

In this case, the gross margin and Revenue is given in the dataset shared by the restaurant. The calculation of COGS was done by the algorithm in Figure 5.1 below using the formula referenced above; this is by multiplying gross margin by revenue then subtracting revenue and multiplying the result by negative one.

$$\text{Cost of Goods} = -1((\text{gross margin} * \text{Revenue}) - \text{Revenue})$$

```

import csv
import pandas as pd

#### calculating Cost of Goods when Gross Margin and Revenue is given####
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
o_handle['Cost_Of_Goods'] = -1*((o_handle['Gross_margin']*o_handle['Revenue'])-o_handle['Revenue'])

o_handle.to_csv('WA_Sales_Products_2012.csv', index=False)

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_colwidth', 100)
pd.set_option('display.width', None)
print(o_handle.head(89000))

```

Figure 5. 1:Calculating Cost of Goods

5.3.2. Profit

Profit is the financial gain between what is spent and what is earned from a business transaction. In this scenario, profit is calculated using the formula alluded to below of getting the difference between Revenue and Cost of goods calculated in Figure 5.2.

Profit= revenue – Cost of goods:



```

import csv
import pandas as pd

###calculating profit, given Revenue and cost of goods ###
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
o_handle['Profit'] = o_handle['Revenue']- o_handle['Cost_Of_Goods']

o_handle.to_csv('WA_Sales_Products_2012.csv', index=False)

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_colwidth', 100)
pd.set_option('display.width', None)
print(o_handle.head(89000))

```

Figure 5. 2: Calculating Profit

5.3.3. Change in Revenue

Change in income can either be certain or negative. A positive outcome indicates that the income has expanded, while a negative outcome implies that the income declined. The higher the rate, the more grounded the improvement or diminishing. Frequently, an organization puts income on its pay articulation when the cash is earned. Such activities can control the income change rate contingent upon when the income was recorded (Kantardzic & Mehmed, 2003).

This was got by getting the difference in initial revenue and the final revenue as calculated using the formula discussed in the paragraph above. Figure 5.3 is the Python content used to calculate change in revenue.

```

import numpy as np
import matplotlib.pyplot as plt
from apyori import apriori
from pandas import ExcelWriter
from pandas import ExcelFile
import csv
import pandas as pd

### Calculating Change in Revenue
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
df = pd.DataFrame(o_handle)
df['Change_Rev'] = df['Revenue'].diff(1)

o_handle.to_csv('WA_Sales_Products_2012.csv', index=False)

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_colwidth', 100)
pd.set_option('display.width', None)
print(o_handle.head(89000))

```

Figure 5. 3: Calculating Change in Revenue

5.3.4. Change in Quantity

This is the difference in quantity of goods sold in different consecutive transactions. This change is calculated progressively by subtracting the next quantity from the previous quantity as shown in Figure 5.4 using python script.

```

import numpy as np
import matplotlib.pyplot as plt
from apyori import apriori
from pandas import ExcelWriter
from pandas import ExcelFile
import csv
import pandas as pd

### Calculating Change in Quantity
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
df = pd.DataFrame(o_handle)
df['Change_Qty'] = df['Quantity'].diff(1)

o_handle.to_csv('WA_Sales_Products_2012.csv', index=False)

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_colwidth', 100)
pd.set_option('display.width', None)
print(o_handle.head(89000))

```

Figure 5. 4: Calculating Change in Quantity

5.3.5. Change in Cost

A basic idea that rouses full-lifecycle testing is the expense of progress. It demonstrates the general expense of tending to a changed necessity, either in the light of the fact that it was missed or misjudged, all through the lifecycle (Kurgan & Musilek, 2006).

The idea of change in cost is derived from what is spent in initial stocking or buying of items against what has been spent in the next stocking, which is the cost of goods, the difference is either positive or negative and this helps in predicting the trend of change in cost. Figure 5.5 shows the python script used to calculate the change in cost.

```
import numpy as np
import matplotlib.pyplot as plt
from apyori import apriori
from pandas import ExcelWriter
from pandas import ExcelFile
import csv
import pandas as pd

### Calculating Change in Cost
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
df = pd.DataFrame(o_handle)
df['Change_Cost'] = df['Cost_Of_Goods'].diff(1)

o_handle.to_csv('WA_Sales_Products_2012.csv', index=False)

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_colwidth', 100)
pd.set_option('display.width', None)
print(o_handle.head(89000))
```

Figure 5. 5: Calculating Change in Cost

5.3.6. Marginal Revenue

Minor Revenue (MR) is the extra income that will be created by expanding item deals. Mathematically, Marginal Revenue is calculated by dividing Change in revenue (calculated above) divide by Change in quantity (also calculated above). Figure 5.6 shows the python script used to calculate the Marginal Revenue.

```

import numpy as np
import matplotlib.pyplot as plt
from apyori import apriori
from pandas import ExcelWriter
from pandas import ExcelFile
import csv
import pandas as pd

### Calculating Marginal Revenue
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
df = pd.DataFrame(o_handle)
df['Marginal_Rev'] = df['Change_Rev']/df['Change_Qty']

o_handle.to_csv('WA_Sales_Products_2012.csv', index=False)

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_colwidth', 100)
pd.set_option('display.width', None)
print(o_handle.head(89000))

```

Figure 5. 6: Calculating Marginal Revenue

5.3.7. Marginal Cost

The peripheral expense is an adjustment to the absolute cost emerging when the amount delivered is increased by one unit; that is, it is the expense of creating one more unit of a decent. Naturally, minor expense at each dimension of creation incorporates the expense of any extra data sources required to deliver the following unit. At each dimension of creation and timespan being considered, peripheral expenses incorporate all costs that shift with the dimension of generation, while different costs that do not change with creation are fixed and hence have no minor expense (Azevedo & Santos, 2008).

By and by, a minor investigation is isolated into short and long-run cases, so that, as time goes on, all expenses (counting fixed expenses) become minimal. Where there are economies of scale, costs set at peripheral cost will neglect to take care of all-out costs, in this way requiring an endowment. Negligible cost valuing doesn't involve just bringing down the general dimension of costs with the guide of an endowment; with or without sponsorship it requires a radical rebuilding of estimating rehearses, with open doors for exceptionally generous enhancements ineffectiveness at basic focuses. This is simplified in the script below by change

in cost calculated above by change in quantity (also calculated above), the result is got in percentage as shown in Figure 5.7.

```
import numpy as np
import matplotlib.pyplot as plt
from apyori import apriori
from pandas import ExcelWriter
from pandas import ExcelFile
import csv
import pandas as pd

### Calculating Marginal Cost
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
df = pd.DataFrame(o_handle)
df['Marginal_Cost'] = df['Change_Cost']/df['Change_Qty']

o_handle.to_csv('WA_Sales_Products_2012.csv', index=False)

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_colwidth', 100)
pd.set_option('display.width', None)
print(o_handle.head(89000))
```

Figure 5. 7: Calculating Marginal Cost

5.3.8. Combined Changes

Figure 5.8 below is the Python script used to generate the combined changes, that is, change in Cost, change in revenue and change in Quantity. The scrip here was run on the same table to generate the required columns for the algorithm.

```

import numpy as np
import matplotlib.pyplot as plt
from apyori import apriori
from pandas import ExcelWriter
from pandas import ExcelFile
import csv
import pandas as pd

#### calculating Cost of Goods when Gross Margin and Revenue is given###
o_handle = pd.read_csv('C:\School\Yr2-3- Semester\IT Thesis\Thesis\WA_Sales_Products_2012a.csv')
o_handle['Cost_Of_Goods'] = -1*((o_handle['Gross_margin']*o_handle['Revenue'])-o_handle['Revenue'])

###calculating profit, given Revenue and cost of goods ###
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
o_handle['Profit'] = o_handle['Revenue']- o_handle['Cost_Of_Goods']

### Calculating Change in Revenue
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
df = pd.DataFrame(o_handle)
df['Change_Rev'] = df['Revenue'].diff(1)

### Calculating Change in Quantity
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
df = pd.DataFrame(o_handle)
df['Change_Qty'] = df['Quantity'].diff(1)

### Calculating Change in Cost
o_handle = pd.read_csv('WA_Sales_Products_2012.csv')
df = pd.DataFrame(o_handle)
df['Change_Cost'] = df['Cost_Of_Goods'].diff(1)

o_handle.to_csv('WA_Sales_Products_2012.csv', index=False)

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_colwidth', 100)
pd.set_option('display.width', None)
print(o_handle.head(89000))

```

Figure 5. 8: Combines changes calculations

5.4. Training Dataset creation and processing for the model

A total of 190 records were generated in the format specified for the model, this dataset contains combinations of different items sold in the restaurant and the model was trained on this then after the learning process, it applied to the main restaurant dataset which has over 80,000 records. The part of the script for training dataset is shown in Figure 5.9. The script accepts the input of the training dataset, calculates the Support, the Lift and Confidence using the Apriori Algorithm as explained in chapter 4 of this document. The script then runs to execution on each item record in the dataset, finding out the best optimum combinations.

```
In [80]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from apyori import apriori

In [81]: #apriori algo expects strings
#preprocess
#these products were recorded for 7 days
dataset = pd.read_csv('data_n.csv',header=None)
#dataset
transactions = []

for i in range(0,7501):
    transactions.append([str(dataset.values[i,j]) for j in range(0,20)])

In [82]: #training
#we shall support to test 3 products seen in a day for 7 days
#there our minimum support shall be 3*7/7501 = 0.003
#minimum confidence 0.2 how many times is rules correct ,
#chance of buying A that will also buy B % age of chance
#minimum lift -> relevant of the rule 3 enables sort the rules

rules = apriori(transactions=transactions,min_support=0.003,
                min_confidence=0.2,min_lift=3,min_length=2,max_length=2)

In [83]: results = list(rules)
```

Figure 5. 9: Lift, Support and Confidence script

The intermediate results are displayed alongside the script as in Figure 5.10. Consequently, the scripts ran and produce the desired end results which were then implemented on the test data. The three mark point; Lift, Support and Confidence are part of the inputs to the second part of the algorithm. The script then runs continuously to calculate using the algorithm the items' sales in a day and subject the sales to the Apriori algorithm as in Figure 5.10.

```

In [84]: #visualization & results computation
def inspectRules(results):
    cost = []
    rev = []
    profit = []
    number_of_purchase_per_day =3
    number_of_days_was_sample=7
    number_of_items_sold = number_of_purchase_per_day*number_of_days_was_sample
    lhs = [tuple(result[2][0][0])[0] for result in results]
    rhs = [tuple(result[2][0][1])[0] for result in results]

    for (l1,l2) in zip(lhs,rhs):
        if l1 == 'light cream' and l2 == 'chicken':
            c = 750*number_of_items_sold
            r = 950*number_of_items_sold
            p= r-c
            cost.append(c)
            rev.append(r)
            profit.append(p)
        elif l1 == 'mushroom cream sauce' and l2 == 'escalope':
            c = 900*number_of_items_sold
            r = 1000*number_of_items_sold
            p= r-c
            cost.append(c)
            rev.append(r)
            profit.append(p)
        elif l1 == 'pasta' and l2 == 'escalope':
            c = 600*number_of_items_sold
            r = 750*number_of_items_sold

```

Figure 5. 10:Data optimization script

Mining for association among items in a large dataset of a sample sales transaction is an important mining function. Based on the types of values, the association rules can be classified into two categories: Boolean Association rules and Quantitative Association rules. For this model, the Quantitative Association Rules were employed since the combinations were more than two and were not in Boolean format. The Association rules shown in Figure 5.11 were used in this optimization model.

```

it[4]: [RelationRecord(items=frozenset({'light cream', 'chicken'}), support=0.0045327289694707
37, ordered_statistics=[OrderedStatistic(items_base=frozenset({'light cream'}), items_a
dd=frozenset({'chicken'}), confidence=0.29059829059829057, lift=4.84395061728395)]),
  RelationRecord(items=frozenset({'escalope', 'mushroom cream sauce'}), support=0.005732
568990801226, ordered_statistics=[OrderedStatistic(items_base=frozenset({'mushroom crea
m sauce'}), items_add=frozenset({'escalope'}), confidence=0.3006993006993007, lift=3.79
0832696715049)]),
  RelationRecord(items=frozenset({'escalope', 'pasta'}), support=0.005865884548726837, o
rdered_statistics=[OrderedStatistic(items_base=frozenset({'pasta'}), items_add=frozenset
({'escalope'}), confidence=0.3728813559322034, lift=4.700811850163794)]),
  RelationRecord(items=frozenset({'fromage blanc', 'honey'}), support=0.0033328889481402
48, ordered_statistics=[OrderedStatistic(items_base=frozenset({'fromage blanc'}), items
_add=frozenset({'honey'}), confidence=0.2450980392156863, lift=5.164270764485569)]),
  RelationRecord(items=frozenset({'ground beef', 'herb & pepper'}), support=0.0159978669
51073192, ordered_statistics=[OrderedStatistic(items_base=frozenset({'herb & pepper'}),
items_add=frozenset({'ground beef'}), confidence=0.3234501347708895, lift=3.29199384113
49285)]),
  RelationRecord(items=frozenset({'ground beef', 'tomato sauce'}), support=0.00533262231
7024397, ordered_statistics=[OrderedStatistic(items_base=frozenset({'tomato sauce'}), i
tems_add=frozenset({'ground beef'}), confidence=0.3773584905660377, lift=3.840659481324
083)]),
  RelationRecord(items=frozenset({'olive oil', 'light cream'}), support=0.00319957339021
4638, ordered_statistics=[OrderedStatistic(items_base=frozenset({'light cream'}), items
_add=frozenset({'olive oil'}), confidence=0.20512820512820515, lift=3.114709851551957
3)]),
  RelationRecord(items=frozenset({'olive oil', 'whole wheat pasta'}), support=0.00799893
3475536596, ordered_statistics=[OrderedStatistic(items_base=frozenset({'whole wheat pas
ta'}), items_add=frozenset({'olive oil'}), confidence=0.2714932126696833, lift=4.122410
097642296)]),
  RelationRecord(items=frozenset({'pasta', 'shrimp'}), support=0.005065991201173177, ord
ered_statistics=[OrderedStatistic(items_base=frozenset({'pasta'}), items_add=frozenset
({'shrimp'}), confidence=0.3220338983050847, lift=4.506672147735896)])]

```

Figure 5. 11: Association rules.

After getting the output of the rules, that is the Lift, Support and Confidence from the algorithm, the system then goes ahead and calculates the revenue and profit. The results are then displayed as shown in Figure 5.12 for all the items and then they algorithm prioritizes the optimum profit for the items.

```

lifts = [result[2][0][3] for result in results]
return list(zip(lhs,rhs,supports,confidences,lifts,cost,rev,profit))
resultsDF = pd.DataFrame(inspectRules(results),
                        columns=['ITEM_ONE','ITEM_TWO','SUPPORT',
                                'CONFIDENCE','LIFT','COST','REVENUE','PROFIT'])

```

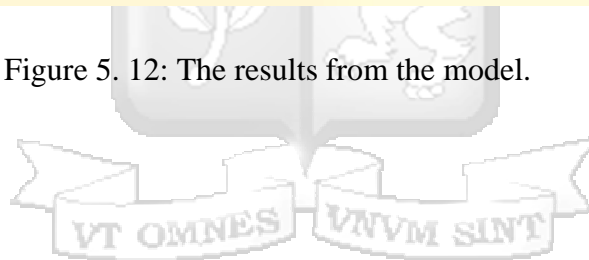
In [78]: #resultsDF

In [85]: resultsDF.nlargest(n=10,columns='LIFT')

Out[85]:

	ITEM_ONE	ITEM_TWO	SUPPORT	CONFIDENCE	LIFT	COST	REVENUE	PROFIT
3	fromage blanc	honey	0.003333	0.245098	5.164271	31500	42000	10500
0	light cream	chicken	0.004533	0.290598	4.843951	15750	19950	4200
2	pasta	escalope	0.005866	0.372881	4.700812	12600	15750	3150
8	pasta	shrimp	0.005066	0.322034	4.506672	48300	56700	8400
7	whole wheat pasta	olive oil	0.007999	0.271493	4.122410	45990	62790	16800
5	tomato sauce	ground beef	0.005333	0.377358	3.840659	25200	35700	10500
1	mushroom cream sauce	escalope	0.005733	0.300699	3.790833	18900	21000	2100
4	herb & pepper	ground beef	0.015998	0.323450	3.291994	42000	48300	6300
6	light cream	olive oil	0.003200	0.205128	3.114710	48300	56700	8400

Figure 5. 12: The results from the model.



Chapter 6: Discussions

6.1. Introduction

This chapter discusses the outcome of the model in relation to study's objectives discussed in the initial chapters. The primary objective was to develop a model for profit maximization using Apriori data mining algorithm, following the available different manual methods of profit maximization, which are less efficient and tiresome to use, The model was based on different integrated programming paradigms, because of its useful approach to optimization, when the outcome of the model is of integer or non-integer types. It accommodates both types of outcomes. In designing of the model, some of the user inputs included items sold together, the items costs and revenues.

6.2. Model Representation

Considering the idea which was intended to be used, this model has been followed to the latter and desirable results were obtained. This shows that the scientific input into the modified Apriori algorithm has born fruits and added to the frequency of the work of restaurant owners and operators who might desire to implement this system. The algorithm was able to obtain the data and follow the flow shown in Figure 6.1 to give the desired results.

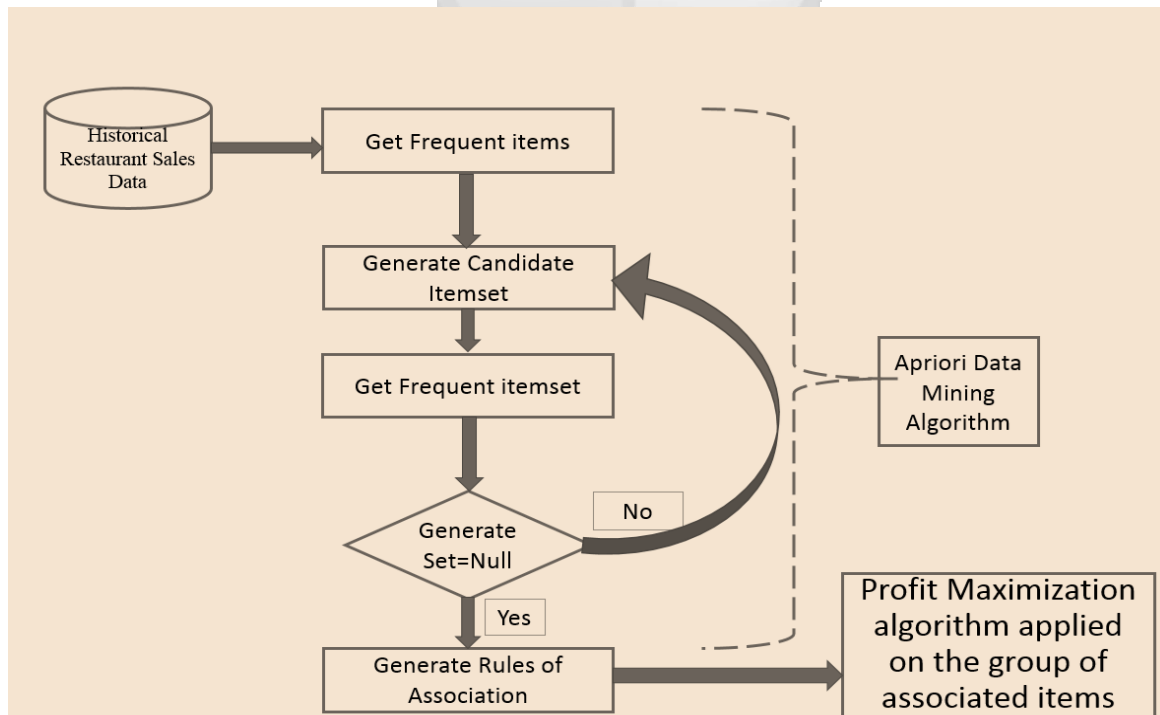


Figure 6. 1: Diagrammatic Representation of the Model.

6.3. Data Mining Algorithms for Profit Optimization

When coming up with this model for profit maximization, the steps shown in Figure 6.2 were put into consideration. The data was prepared and assurance was made that all the required input variables are present. The data was then run through the algorithm and when the results were compared to the expected outcomes, it brought about the desired output and this let to refinement of the model to make it more users friendly before running it on the actual data.

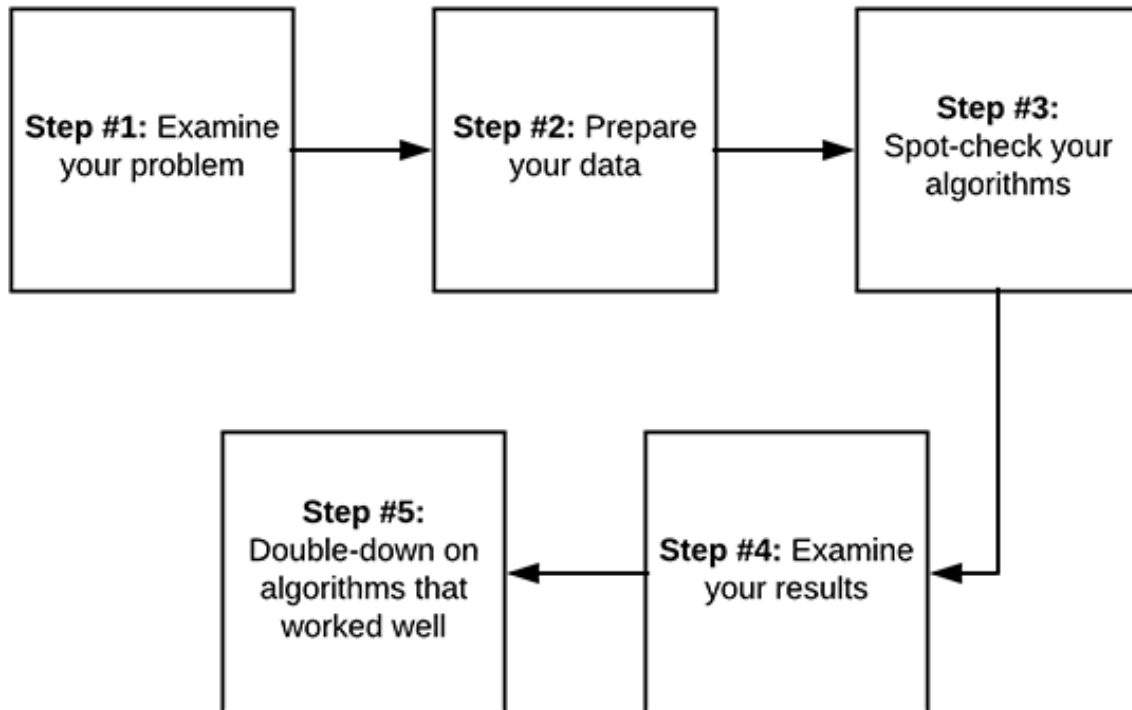


Figure 6. 2: Steps to check the efficiency of an algorithm (Fandango & Armando, 2008)

6.4 Model validation

The model selection for this research was done by following the standard laid down approaches for machine learning. Out of all the proposed models, the Apriori was the best

performing and was selected following the kind of data we had in mind and the expected outcome of the research. Since this was a supervised learning, evaluation on the performance in terms of accuracy, precision and efficiency were considered ideal. Each of these measures was exhibited in the previous segments. Accuracy is defined as the ratio of true positive to the sum of true positive and false positive, Precision on the other hand is defined as the ratio of true positive to the sum of true positive and false positive and efficiency is the ease of execution of programs and transactions by a system.

Distinctive sets of classical models were tested. Each has its own hypothetical underpinnings and changing hypothesis of learning. Bolster Vector Machine, for instance, depends on pivot misfortune and expansion of the edge of division, Naive Bayes; a probabilistic model dependent on the boost of the probability of the back probabilities molded on the earlier qualities.

Despite the fact that the model's outcomes were not exactly the same as expected, the outcomes appeared to agree with the hypothesis in the questions to answer as stated earlier. Hypothetically, numerous examinations have proposed that when there is a great deal of information then the algorithm being utilized does not have a major effect. The bigger the dataset estimate, the better the model's execution in spite of the theoretical framework.

However, the data mining algorithm that was settled for in this case was the Apriori Algorithm. This is mainly because:

- i. Apriori Algorithm is easy to understand among the association rule learning algorithms.
- ii. The resulting association rules of Apriori are intuitive and easy to communicate to an end user.
- iii. The algorithm doesn't require labeled data as it is fully unsupervised; as a result, one can use it in many different situations because unlabeled data is often more accessible.
- iv. The algorithm is exhaustive, so it finds all the rules with the specified support and confidence thus making the algorithm really efficient.

6.4. Conclusion on discussion

Following the discussions above, the model has been taken through validation processes, has been also compared to other existing models and found to have advantages over the other models developed so far. The algorithm used have different theories of machine learning and data mappings making it more efficient and easy to use, this gives it more advantages over the existing models which involves a lot of manual input and that automation is less forcing the users to include other extra work and leading to accuracy issues. The model here is fit for the restaurant users to implement and enjoy the efficiency it brings in optimizing their activities and maximizing the profit.



Chapter 7: Conclusions and Recommendations

7.1. Conclusion

The main goal of this research work was to develop a data mining tool which can aid in coming up with profit optimization in sales. The target market of this research is restaurants and it is aimed at helping restaurants in their sales so that they can optimize their chances of making profits and avoid losses at all cost, if possible.

The researcher segmented the principle objective into sub-objectives by inspecting the writing on benefit streamlining rehearses, utilization of data mining in profit optimization, set of learning algorithms and highlight designing methods which have given promising outcomes among the exploration algorithms. Numerous arrangements of established and profound learning models have been created and their execution benchmarked against measurements, for example, exactness, precision, and review. In addition to the fact that this helped the researcher select the best model but also validate the model.

Diverse data mining techniques have distinctive learning hypothetical demerits. Thus, the array of experiments carried out helped to select the best data mining algorithm. Following the tests, this Apriori algorithm proved to be best fit over the majority of the established algorithms in terms of accuracy, preciseness, efficiency and also ease of use. Therefore, it was chosen in this case to aid in profit optimization.

7.2. Challenges

Despite the ability to achieve the main objectives for this research endeavor, quite significant challenges were encountered. Luckily, these challenges did not affect the project to a large extent.

Initially, the researcher would be given incorrect data as the management of Kula Kona Restaurant was unsure of what the researcher would do with the information or whether the authorities had sent him to benchmark on how they operate.

Secondly, some of the Kula Kona Restaurant staff members were not willing to share the restaurant information due to privacy and tax policies concerns.

Thirdly, it was rather challenging to convert the information given by the staff members through word of mouth, into python scripts that can be run on an IDE and implement the data mining techniques and algorithms so as to come up with an optimum profit.

Lastly, since the research requires multiple data sources across multiple time frames so as to test the efficiency of the data mining techniques and algorithms, the researcher had to keep going to the restaurant for more data but the staff would become less and less cooperative each time.

7.3. Recommendations

- 1) This kind of research work should be done in close partnership with an accounting facility such an accounting firm. This makes it easier to further validate the model to assert the performance of the model is reflected in a real-world setting and that it actually aids in optimizing profits.
- 2) To facilitate ease of adoption of this tool, the project should be carried out liaising with end users, such as restaurants or supermarkets as the model would help these businesses in optimizing profits and reducing losses.
- 3) More tests ought to be done to improve the model's execution. The algorithm's precision can be without a doubt expanded by an expanded presentation to new marked cases from new sales subjects.
- 4) The business network ought to secure aptitude input right on time from the initiation of the undertaking and furthermore address approaching social, moral, protection concerns which are basic in the business segment.

7.4. Future Work

There is still such a great deal of work to be done. In the future, the researcher will gather more information to help lessen the error rate of the present model and to influence it to

sum up much better on concealed occurrences. This will empower the model to scale on a level plane empowering a significant number of people to access it as a service.

Working on larger datasets will require experimentation with more parallel enabled models and resource computing platforms. The researcher is also intending to automate the profit optimization process through the use of mobile devices which can transmit data to a remote server hosting the built model. This will make it possible for a huge number of people to gain access to the model as a service (MaaS).

The users will be able to initiate the monitoring procedure even without having any specialized expertise. In addition, this will make it simpler to extend the model's capacity to optimize profits from the server and transfer back the outcome remotely to the users by means of the same channel.



References

- Baghel, A., & Sharma, T. (2013). Survey on a fuzzy expert system. *International Journal of Emerging Technology and Advanced Engineering*, 3(12), 230-233
- Baldwin, D., & Scragg, G. (2004). *Algorithms and Data Structures: The Science of Computing*. Hingham, US: Charles River Media / Cengage Learning. Boston, US: Technology Research News. Contemporary Practices: Guidelines for Academic Researchers (1).
- Creswell, J. W. (2008). *Educational Research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River: Pearson.
- Dinu, M. D. (2014). Supply chain management–key factors: Viewpoint and replies on previously published articles. *Network Intelligence Studies*, 2(1), 32–37
- Eisner, E. W. (1981). "On the Differences between Scientific and Artistic Approaches to Qualitative Research". *Educational Researcher*. 10 (4): 5–9.
- Garraghan, Gilbert J. (1946). *A Guide to Historical Method*. New York: Fordham University Press. p. 168. ISBN 978-0-8371-7132-6.
- Gauch, Jr., H.G. (2003). *Scientific method in practice*. Cambridge, UK: Cambridge University Press. 2003 ISBN 0-521-81689-0 (page 3)
- Grandinetti, L., Joubert, G. R., & Kunze, M. (2015). *Advances in Parallel Computing: Big Data and High Performance Computing*. Fairfax, US: IOS Press.

Grebmer, K., Headey, D., Olofinbiyi, T., Wiesmann, D., Fritschel, H., Yin, S., Haddad, L. (2013). *2013 Global hunger index: The challenge of hunger: building resilience to achieve food and nutrition security*. Washington, US: Welt Hunger Hilfe, International Food Policy Research Institute, Concern Worldwide

Gupta, A., Mehrotra, A., & Khan, P. M. (2015). *Challenges of Cloud Computing & Big Data Analytics*. Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on (pp. 1112-1115). New Dehli: IEEE.

Habib, M. M., Pathik, B. B., & Maryam, H. (2014). *Research Methodology - Contemporary Practices: Guidelines for Academic Researchers* (1). Newcastle-upon-Tyne, UNITED KINGDOM: Cambridge Scholars Publishing.

Habib, M. M., Pathik, B. B., & Maryam, H. (2014). Research Methodology -Huang, H. C., Chang, A. Y., & Ho, C. C. (2013). *Using artificial neural networks to establish a customer-cancellation prediction model*. *Przeglad Elektrotechniczny*, 89(1b), 178-180

Hashem, H., & Ranc, D. (2016). Extending Standard MapReduce Algorithms. 2016 IEEE Second International Conference on Multimedia Big Data (pp. 159-165). Taipei: IEEE.

J. Scott Armstrong & Tad Sperry (1994). "Business School Prestige: Research versus Teaching" (PDF). *Energy & Environment*. 18 (2): 13–43. Archived from the original(PDF) on 20 June 2010. Retrieved 8 December 2011.

Karl Rexer, H. A., & Paul Gearan (2011); *Understanding Data Miners, Analytics Magazine*, (May/June 2011). INFORMS: Institute for Operations Research and the Management

Sciences (pp. 98-105).

Kaur, N., & Kang, S. (2016). *Market Basket Analysis: Identify the changing trends of market data using associate rule mining*. Elsevier: Procedia computer science. Newcastle-upon-Tyne, UNITED KINGDOM: Cambridge Scholars Publishing. on Computational Intelligence and Security, 445-449.

Kobielus, James; *The Forrester Wave: Predictive Analytics and Data Mining Solutions*, (Q1 2010), Forrester Research, (1 July 2008)

Leavy, Patricia (2015). *Methods Meet Art (2nd ed.)*. New York: Guilford. ISBN 781462519446. Biotech Business Week Editors (June 30, 2008); BIOMEDICINE; HIPAA Privacy Rule Impedes Biomedical Research, Biotech Business Week, retrieved 17 November 2009 from LexisNexis Academic

Lesage, Dieter (Spring 2009). "Who's Afraid of Artistic Research? On measuring artistic research output" (PDF). *Art&Research – A Journal of Ideas, Contexts, and Methods*. 2(2). Retrieved (14 August 2011).

Mikut, Ralf; Reischl, Markus (September–October 2011). "Data Mining Tools". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.

Nisbet, Robert A. (2006); *Data Mining Tools: Which One is Best for CRM?* Part 1, Information Management Special Reports, (January 2006)

Roffee, James A; Waling, Andrea (18 August 2016). "Resolving ethical challenges when researching with minority and vulnerable populations: LGBTIQ victims of violence, harassment, and bullying". *Research Ethics*. 13 (1): 4–22.

Shields, Patricia, and Rangarajan, N. (2013). *A Playbook for Research Methods: Integrating Conceptual Frameworks and Project Management*. Stillwater, OK: New Forums Press.

Snelson, Chareen L. (March 2016). "Qualitative and Mixed Methods Social Media Research".
International Journal of Qualitative Methods. 15 (1): 160940691562457

Solve the Transcendental Logarithm Problem. 10th International Conference

Song, J., Xu, S., Zhang, L., Pahl, C., & Yu, G. (2015). Performance and Energy

Tang, Y., & Su, S. (2014). Application of Grover's Quantum Search Algorithm to
Technology and Control, 1-12.

Text and Data Mining: Its importance and the need for change in Europe. Association of
European Research Libraries. Retrieved (14 November 2014).

UK Researchers Given Data Mining Right Under New UK Copyright Laws. Archived (June 9,
2014), at the Wayback Machine Out-Law.com. Retrieved (14 November 2014)

"Licences for Europe - Structured Stakeholder Dialogue (2013)". European
Commission. Retrieved (14 November 2014).

Kohli, C. & Suri, R. (2011). *The price is right? Guidelines for pricing to enhance profitability*.
Business Horizons, 54, 563-573.

Kopalle, P., Biswas, D., Chintagunta, P.K., Fan, J., Pauwels, K., Ratchford, B.T. & Sills, J.A.
(2009). Retailer pricing and competitive effects. Journal of Retailing, 85(1), 56-70.

Zhao, J., Tang, W. & Wei, J. (2012). Pricing decision for substitutable products with retail
competition in a fuzzy environment. Int. J. Production Economics, 135, 144-153.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified,
real-time object detection. In Proceedings of the IEEE conference on computer vision and
pattern recognition (pp. 779–788).

- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: *Towards real-time object detection With region proposal networks*. In *Advances in neural information processing systems* (pp. 91–99).
- Røe, K., Kakar, M., Seierstad, T., Ree, A. H., & Olsen, D. R. (2011). *Early prediction of response to radiotherapy and androgen-deprivation therapy in prostate cancer by repeated functional mri: a preclinical study*. *Radiation oncology*, 6(1), 65.
- Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In *Advances in neural information processing systems* (pp. 3859–3869).
- Schneider, M. J., Cyran, C. C., Nikolaou, K., Hirner, H., Reiser, M. F., & Dietrich, O. (2014). *Monitoring early response to anti-angiogenic therapy: diffusion-weighted magnetic resonance imaging and volume measurements in colon carcinoma xenografts*. *PloS one*, 9(9), e106970.
- Scholtens, T. M., Schreuder, F., Ligthart, S. T., Swennenhuis, J. F., Greve, J., & Terstappen, L. W. (2012). *Automated identification of circulating tumor cells by image cytometry*. *Cytometry Part A*, 81(2), 138–148.
- Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1), 1929–1958.
- Svensson, C.-M., Krusekopf, S., Lücke, J., & Thilo Figge, M. (2014). *Automated detection of Circulating tumor cells with naive Bayesian classifiers*. *Cytometry Part A*, 85(6), 501–511.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception Architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826).

Theilmann, R. J., Borders, R., Trouard, T. P., Xia, G., Outwater, E., Ranger-Moore, J., Stopeck, A. (2004). *Changes in water mobility measured by diffusion mri predict response of metastatic breast cancer to chemotherapy*. *Neoplasia*, 6(6), 831–837.

Thoeny, H. C., & Ross, B. D. (2010). *Predicting and monitoring cancer treatment response with diffusion-weighted mri*. *Journal of Magnetic Resonance Imaging*, 32(1), 2–16.



A Prototype for Profit Maximization Using Apriori Data Mining Algorithm Case of Kula Kona Restaurant

ORIGINALITY REPORT

24%	16%	5%	21%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Gulf College Oman Student Paper	2%
2	Tariq Lambachri, Amir Hajjam El Hassani, Emmanuel Andres. "Aligning Pattern Extraction Algorithms for the Lambda Architecture", 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), 2018 Publication	1%
3	select-statistics.co.uk Internet Source	1%
4	pdfs.semanticscholar.org Internet Source	1%
5	Submitted to Strathmore University Student Paper	1%
6	edoc.site Internet Source	1%
7	Submitted to International School of Management and Technology (ISMT), Nepal Student Paper	1%
8	Submitted to Mancosa Student Paper	1%
9	Submitted to Higher Education Commission Pakistan Student Paper	1%
10	Submitted to Study Group Australia Student Paper	1%

11	medium.com Internet Source	1%
12	Submitted to University of New England Student Paper	1%
13	www.saedsayad.com Internet Source	1%
14	Submitted to Middle East College of Information Technology Student Paper	1%
15	Submitted to University of Computer Studies Student Paper	1%
16	docplayer.net Internet Source	1%
17	Submitted to K. J. Somaiya College of Engineering Vidyavihar, Mumbai Student Paper	1%
18	gradesfixer.com Internet Source	<1%
19	www.lsi.upc.edu Internet Source	<1%
20	Submitted to King's Own Institute Student Paper	<1%
21	core.ac.uk Internet Source	<1%
22	www.ukessays.com Internet Source	<1%
23	en.wikipedia.org Internet Source	<1%
24	Submitted to Westminster International University in Tashkent Student Paper	<1%
25	bicorner.com Internet Source	<1%

26	Submitted to University of Greenwich Student Paper	<1%
27	Submitted to Indian School of Business Student Paper	<1%
28	Submitted to Softwarica College of IT & E-Commerce Student Paper	<1%
29	ascelibrary.org Internet Source	<1%
30	Submitted to Regenesys Business School Student Paper	<1%
31	wikimili.com Internet Source	<1%
32	Submitted to Government College of Engineering Aurangabad, Maharashtra State, India Student Paper	<1%
33	Submitted to The British College Student Paper	<1%
34	Submitted to The University of the South Pacific Student Paper	<1%
35	Submitted to St. Patrick's College Student Paper	<1%
36	hdl.handle.net Internet Source	<1%
37	Submitted to Cranfield University Student Paper	<1%
38	ir.jkuat.ac.ke Internet Source	<1%
39	Submitted to Universiti Tenaga Nasional Student Paper	<1%
	Submitted to Asian Institute of Technology	

40	Student Paper	<1%
41	Submitted to Tarleton State University Student Paper	<1%
42	Submitted to University of Huddersfield Student Paper	<1%
43	www.ijrcar.com Internet Source	<1%
44	Submitted to Ain Shams University Student Paper	<1%
45	Submitted to Middlesex University Student Paper	<1%
46	"Communications, Signal Processing, and Systems", Springer Science and Business Media LLC, 2020 Publication	<1%
47	www.riceclima.com Internet Source	<1%
48	Submitted to 60892 Student Paper	<1%
49	Submitted to Greenwich School of Management Student Paper	<1%
50	ceng.cankaya.edu.tr Internet Source	<1%
51	Submitted to Druid Hills High School Student Paper	<1%
52	Submitted to Majan College Student Paper	<1%
53	Submitted to University of Arizona Student Paper	<1%
54	www.savingstudentsmoney.org Internet Source	<1%

55	newsprod.abs-cbnnews.com Internet Source	<1%
56	www.slideshare.net Internet Source	<1%
57	etd.adm.unipi.it Internet Source	<1%
58	Mattia Casula. "Economic Growth and Cohesion Policy Implementation in Italy and Spain", Springer Science and Business Media LLC, 2020 Publication	<1%
59	Submitted to Universiti Utara Malaysia Student Paper	<1%
60	argouml-stats.tigris.org Internet Source	<1%
61	mafiadoc.com Internet Source	<1%
62	erepository.uonbi.ac.ke Internet Source	<1%
63	repositories.lib.utexas.edu Internet Source	<1%
64	Submitted to King's College Student Paper	<1%
65	repository.nwu.ac.za Internet Source	<1%
66	eureka.lib.teithe.gr:8080 Internet Source	<1%
67	link.springer.com Internet Source	<1%