PDF - AUTOMATED MARKET BASKET ANALYSIS SYSTEM - researchcub.info CHAPTER ONE 1.0 INTRODUCTION

Data mining isdescribed as the extraction of hidden helpful information from a collection of huge databases, data mining is also a technique that encompasses an enormousform of applied mathematics and compultational techniques like linkanalysis, clustering, classification, summarizing knowledge, regressionanalysis and so on. data mining tools predict future trends and behaviors, permitting businesses to create knowledge-driven selections. Themachine-driven, prospective analyses offered by data mining move on the farside the analyses of past events. data mining tools provides answer to business questions that were timeconsuming. They search databases for hidden patterns, finding usefulinformation that is beyond the reach of specialists.

Data mining techniquesis enforced speedily on existing package and hardware platforms to reinforce the worth of existing information resources, and might be integrated with newproduct and systems as they're brought. once enforced on high performance client/server or multiprocessing computers, data mining tools will analyze huged atabases to provide answers to questions such as, "What goods consumers tendto buy the most and goods that go alongside with it".

Coenen(2010) inhis publication" Data Mining: Past, Present and Future" discussed the historyof data mining can be dated as far back as late 80s when the term began to be used, at

least within theresearch community and diffrentiated it from sql.

Broadly datamining can be defined as as set of mechanisms and techniques, realised insoftware,

to extracthidden information from data. However, the word hidden in this definition isimportant;

By the early1990s data mining was commonly recognised as a sub process within a largerprocess called Knowledge Discovery in Databases or KDD, the most commonly useddefinition of KDD is that of Fayyad et alas "the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data." (Fayyad et al. 1996).

As such data miningshould be viewed as the sub-process, within the overall KDD process, concernedwith the discovery of hidden information". Other sub-processes that formpart of the KDD process are data preparation (warehousing, data cleaning,pre-processing,and so on) and the analysis/visualisation of results. For maypractical purposes KDD and data mining are seen as synonymous, but technicallyone is a sub-process of the other. The data that data mining techniques wereoriginally directed at was tabular data and, giventhe processing power availableat the time, computational eficiencywas of significant concern. As the

amountof processing power generally available increased, processing became less of aconcern and was replaced with a desire for accuracy and a desire to mine everlarger data collections. Today, in the context of tabular data, we have a wellestablished range of data mining techniques available.

It is wellwithin the capabilities of many commercial enterprises and researchers to minetabular

data, using softwaresuch as Weka, on standard desktopmachines. However, the amount of electronic data collected by all kinds of institutions and commercial enterprises, year on year, continues to grow and thus there is still a need for efective mechanisms to mineever larger datasets. The popularity of data mining increased significantly in the 1990s, notably with the establishment of a number of dedicated conferences; the ACM SIGKDD(specialintrest group on knowledge discovery in data) annual conference in 1995, and the European PKDD(practice of knowledge discovery in databases) and the Pacific/AsiaPAKDD(pacific asiaconference on knowledge discovery and data mining) conferencesThis increase in popularity can be attributed to advances in technology; the computer processing power and data storage capabilities available meant that the processing of large volumes of data using desktop machines was a realistic possibility. It became common place for commercial enterprises to maintain datain computer readable form, in most cases this was primarily to support commercial activities, the idea that this data could be mined often camesecond. The 1990s also saw the introduction of customer loyalty cards that allowed enterprises to record customerpurchases, the resulting datacould then be mined to identify customerpurchasing patterns. Data mining, is the method of looking into giant volumesof data for patterns using methods like classification, association rulemining, clustering, etc.. data mining is a topic that is related to topics like machine learning and patternrecognition. data mining techniques area unit the results of an extendedprocess of analysis and products development.

I am in my final year.I was bright and brilliant, my family was optimistic in me; they thought so much ofme, but I had a fault. What was my fault? I hated compiler construction. I struggled with calculations all my life. Thoughi have been lucky; I did well all thesame. However, I had to write my final exam. I searched for all Compiler constructionpast question for each year, compared, and sorted them. Guess what Idiscovered! Over 35% of the questions were repetitions. I had hit the jackpot.I carefully and thoroughly checked through the answer page. Therefore, I kepton revising only the repeated questions. Well, I have a good grade to show forthe Data Mining I performed.

There is huge amount ofdata available in Information Industry. This data is of no use until convertedinto useful information. Analyzing this huge amount of data and extractinguseful information from it is necessary.

The extraction of information is notthe only process we need to perform; it also involves other processes such asData pre-processing(Data Cleaning, Data Integration, Data Transformation) DataMining, Pattern Evaluation and Data Presentation. Once all these processes areover, we are now position to use this information in many applications such asFraud Detection, Market Analysis, Production Control, Science Exploration etc.

1.1PROBLEM STATEMENT

Through in depthresearch and observations carried on supermarket we have discorvered that retailers are willing to know what product is purchased with the other or if aparticular products are purchased together as a group of items. Which can helpin their decision making with respect toplacement of product, determining the timing and extent of promotions on product and also have a better understanding of customer purchasing habits bygrouping customers with their transactions.

This project is aimedat designing and implementing a well-structured market basket analysis softwaretool to solve the problem stated above and compare the result to that of anexisting software called WEKA.

1.2SIGNIFICANCEOF THE STUDY

1.3AIMAND OBJECTIVE OF THE STUDY

Theaim of the study is to maximize profitfor the retailers by providing betterservices to the consumers Theobjective of this study are:

·Cross-Market Analysis - Data Miningperforms Association/correlations between product sales.

·Identifying Customer Requirements -helps in identifying the best products for differentcustomers. It usesprediction to find the factors that may attract new customers.

Customer Profiling - helps to determinewhat kind of people buy what kind of products.

1.4METHODOLOGY

I.DataPre-Processing

Dueto the fact that the data we are getting is a raw data, raw data in the realworld may be incomplete it has

to bepre-processed the raw data has to go through data cleaning, data

integration, data normarlization, data reduction because without a quality data there will beno quality mining results.

Ø datacleaning: This has to do withfilling of missing values, resolving of inconsistencies in the raw data.

Ø dataintegration:combining datafrom multiple sources and generating the user with unified view of the data

Ø normarlization:normalization is used to minimize or to reduceredundancy.

Ø datareduction:reduction of the data setthat is much smaller in volume but yet yields the same analytical results

1.5 SCOPEOF THE STUDY

This scope of thestudy focuses on Babcock Ventures supermarket and the scope of this projectincludes:

1. Weaim to develop our very own market basket analysis software, which will be usedin babcock university

2. Thesoftware will exhibit a colorful GUI(graphical user interface).

3. Thesoftware will be based on Apriori .

4. Weintend to conduct a research into the various branches of science that thissoftware will be based

on, such as artificial intelligence.

5. Wewill develop a software that will eventually stand out among other data miningsoftware.

1.6 LIMITATIONOF THE STUDY

The limitations of thissoftware will include:

1.Datarestrictions: this is a major factor that stands in the way of the execution of this project. Since there is no data on households and individual consumers , we neglect such purchases.

2. Time constraints: this is also a major factor due to the fact that it can't work on a small amount ofraw

data because it tends to mislead the retailer in a nut shell this softwarewill work on large volumes of data.

AUTOMATED MARKET BASKET ANALYSIS SYSTEM

The complete project material is available and ready for download. All what you need to do is to order for the complete material. The price for the material is NGN 3,000.00. Make payment via bank transfer to Bank: Guaranteed Trust Bank, Account name: Emi-Aware technology, Account Number: 0424875728

Bank: Zenith Bank, Account name: Emi-Aware technology, Account Number: 1222004869

or visit the website and pay online. For more info: Visit https://researchcub.info/paymentinstruct.html

After payment send your depositor's name, amount paid, project topic, email address or your phone number (in which instructions will sent to you to download the material) to +234 70 6329 8784 via text message/ whatsapp or Email address: info@allprojectmaterials.com.

Once payment is confirmed, the material will be sent to you immediately.

It takes 5min to 30min to confirm and send the material to you.

For more project topics and materials visit: https://researchcub.info/ or For enguries:

info@allprojectmaterials.com or call/whatsapp: +234 70 6329 8784 Regards!!!