**Product Comparison and Analysis using Data Mining**

A Thesis

Presented to the Faculty of the

College of Computer and Information Sciences

Polytechnic University of the Philippines

In Partial Fulfilment

Of the Requirements for the Degree

Bachelor of Science in Computer Science

Averos, Christian M.

Delicano, Jobea Ann F.

Chapter 1: The Problem and Its Background

1.1 Introduction

In the fast-paced modern world, where technology progresses rampantly, we try to keep up with today's trends as well as needs. With the rise of technology, clinging unto technology makes our lives easier, and some experts say that the internet can now be consider a need along with food and shelter [Roper Starch Worldwide and America Online, Inc., 1998]. Everything is being digitalized as for our communication and work-processes, and with the use of the internet and its rapid development, delivery systems are getting more and more popular alongside of online shopping websites (OSW). With that being said, it has a great affect on us, Filipinos, especially modern Filipina buy products on the internet through OSW.

Most people all over the Philippines have access to the internet and they are still growing in numbers but there's a problem, how can we diminish the hard work process of search through many OSWs just to get the best deals? As a Filipino, I'm used to seeing my mother go through many store just to get the optimal price that fits in our budget, sometimes, we even come to a point of bargaining or asking the store to lower the price if we're going to buy bulk of their products. Users would search for many websites to compare and analyze what is the best for their searched similar products, weighing in the reviews, quantity, and even quality just to make sure that they're making a good decision upon purchasing such product. Therefore, the researchers decided to develop a system to help with troubled online shoppers (OS).

1.2 Background of the Study

Most Filipino shoppers have a hard time jumping to another store into other store just to find the best deals of their wanted product, may it be a new pair of shoes, shining jeans, or maybe a newly released branded cap. As much as we can, we bargain, bargain, and bargain. Most modern Filipino women would go to mall after mall just to buy off on-sale product which is also true for old-fashion mothers who would prefer to bargain just to get deals from their market suppliers. In today’s technological advantage, many shoppers prefer e-commerce as their mode of shopping (Kitonyi, 2017), primarily because internet is easy to access and buying is just a click away.

Filipinos like to bargain, since the old times, Filipino mothers would go to the market with a fixed money and get much more worth of their money for. Just imagine the hard work of going through and bargaining to many stores just to lessen the pay for a similar product as well as avoiding similar low quality products. This is also true to online shoppers looking for looking for a specific product in an online store, they would search many online store just to get the best deal in their opinion. They would scan for reviews of the said product and sometimes to avoid getting scammed or getting less than what you expected. One of the most common problem in OSWs is that there are scammers always present (Jean Magboo, 2014) even in the most trusted OSWs.

The researchers plan to build a system that caters in the selection of similar products in which the user has to decide to buy or not. Getting the data only from the trusted OSW to minimize cyber market scamming and increasing the satisfaction rate of the users in selecting. Filtering only the 'best' deals in terms of reviews, quantity, and popularity to get the most-likely product to be picked by the user.

1.3 Conceptual Framework

1.3.1 Conceptual Framework of the System

Figure 1.3.1 shows the conceptual framework of the System where the User’s search terms are the input data. The processes are getting the data from OSW and filtering out redundant data. The output data are the collection of best deals.

|  |  |  |
| --- | --- | --- |
| **INPUT** | **PROCESS** | **OUTPUT** |
| * User search term | * Getting data from site * Filtering out redundant data | * Best Deals |

Figure 1.3.1 Conceptual Framework of the System

1.4 Statement of the Problem

This study aims to develop a system that gets the best deal from various online shopping websites based on user’s search terms.

1. Using the a-priori algorithm, what is the accuracy of getting the relevant data from online shopping websites?
2. How should the algorithm be devised in order to get the best deals out of the listed relevant product?
3. What is the accuracy and reliability of the devised algorithm in getting the best deals?

**1.5 Significance of the Study**

The system will benefit the following people:

Online Shoppers. This study would benefit shoppers who wants to lessen their time on selecting the best deal.

Store Owners. This will benefit them by being able to search how they would price their own products.

Future Researchers. This study will be a help as a guiding reference in making a system related to e-commerce.

1.6 Scope and Limitation

1.6.1 Scope and Limitation of the System

The system will mine data from popular retailer online shopping websites and will not include second hand online shopping websites, specifically:

1. Lazada (lazada.com.ph)
2. Shopee (shopee.ph)
3. Amazon (amazon.com)
4. Zalora (zalora.com.ph)

The scope in assessing the best deals will base on the relevancy, availability, price including the shipping fee, specifications, and numerical reviews regarding the product. The product prices will be displayed in Philippine Peso.

The system will not be affiliated in selling products and will only act as a recommender system for finding the best deals.

**1.6.2 Scope and Limitation of the Study**

This study will focus on the accuracy of the finding the best deals of a product selected by the user at the top online shopping websites.

1.7 Definition of Terms

Data Mining - process of extracting information from a large sets of data.

Online shoppers - product consumers that uses Online Shopping Websites as their medium for purchasing.

Online Shopping Websites - form of e-commerce that allows users to purchase commodities over the internet with the help of a web browser.

Scammers - people who extorts or uses dirty tactics to gain information, power or money.

**References**

Kitonyi, https://www.gurufocus.com/news/490164/ecommerce-is-killing-traditional-retail

Jean Magboo, http://lifestyle.inquirer.net/157606/safety-in-online-buying-and-selling/

American Online, Inc., https://www.clickz.com/internet-becoming-necessity-to-users/72138/

[1] DATA MINING TECHNIQUES AND APPLICATIONS

Mrs. Bharati M. Ramageri, Lecturer

Modern Institute of Information Technology and Research,

Department of Computer Application, Yamunanagar, Nigdi

[2] DATA MINING: CONCEPTS, BACKGROUND AND METHODS OF INTEGRATING UNCERTAINTY IN DATA MINING

Yihao Li, Southeastern Louisiana University

Faculty Advisor: Dr. Theresa Beaubouef, Southeastern Louisiana University

[3] Use of Data Mining in Various Field: A Survey Paper May-Jun. 2014

Smita, Priti Sharma

(Student, M.Tech, Amity University)

(Assistant Professor, Amity University)

[4] Aakanksha Bhatnagar, Shweta P. Jadye, Madan Mohan Nagar” Data Mining Techniques & Distinct Applications: A Literature

Review” International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 9, November- 2012

[5] Data Mining Techniques for Marketing, Sales, and Customer Support march 2004

Michael J.A. Berry

Gordon S. Linoff

[6]Data Mining and Warehousing April 3-5, 2014

Ali Radhi Al Essa

School of Engineering

University of Bridgeport

Bridgeport, CT, United States

aalessa@my.bridgeport.edu

Bach, Christian

School of Engineering

University of Bridgeport

Bridgeport, CT, United States

cbach@bridgeport.edu

Chapter 2: Review of Related Literature and Studies

2.1 Review of Related Literature

**2.2 Review of Related Studies**

2.3 Synthesis of the Study

Data mining

In the ever expanding data collection and data storage, Data Mining (DM) has been a logical process used to search useful information through piles of data. [1] Since data sets has become more complex and larger than it was, DM was improved by different algorithms discovered and developed in the field of computer science. DM is an important task in knowledge discovery in databases (KDD). [2] KDD consists procedures including [3]:

1. selection of data from varying sources;
2. preprocessing or data reduction;
3. transformation of data;
4. identifying the desired result; and
5. interpretation and analysis to give relevant information.

There are several DM techniques developed and applied in KDD [4]:

Association. Association is one of the best known data mining technique.

In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction. For example, the association technique is used in reservation systems analysis to identify in which area customers frequently make reservations. Based on this data businesses can set up corresponding reservation counters in that area to sell more tickets and make more profit.

Classification. Classification is based on machine learning.

Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. Basically classification is used to categorize each item in a set of data into one of predefined set of classes or groups [4]. For example, we can apply classification in application that “given all past records of employees who left the company, predict which current employees are probably to leave in the future.” In this case, we divide the employee’s records into two groups that are “leave” and “stay”.

Clustering.

Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. Different from classification, clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. Consider library as an example. In a library, books have a wide range of topics available. The challenge is how to keep those books in a way that readers can take several books in a specific topic without hassle. By using clustering technique, we can keep books that have some kind of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in a topic, he or she would only go to that shelf instead of looking the whole in the whole library.

Prediction.

It is one of a data mining techniques that discover relationship between independent variables and relationship between dependent and independent variables [3]. For instance, prediction technique can be used in Library to predict books that need to be purchased for the future if we assume that the courses offered by a university are constant. Courses are independent variable, and books could be a dependent variable.

Sequential Patterns.

Sequential patterns analysis is one of data mining technique that seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships among data.

Discrimination.

Data discrimination produces what are called discriminant rules and is basically the comparison of the general features of objects between two classes referred to as the target class and the contrasting class. For example, one may want to compare the general characteristics of the customers who rented more than 30 movies in the last year with those whose rental account is lower than 5. The techniques used for data discrimination are very similar to the techniques used for data characterization with the exception that data discrimination results include comparative measures [1]

DM is currently used in wide range of industries that holds a large amount of data and is commonly combining it in other tools that can enhance the power of DM in various fields. [1] Some of these fields include the field of biological science to analyze sequential pattern in genes and identify various diseases related to it. Finance industries use DM for price prediction, stock forecasting, identify frauds and money laundering. [2] Sales industry also use DM as in different ways [5]:

1. Information broker for customer buying habits, from transaction histories to loyalty card usage. Because of this, supermarkets can predict customer behavior and act upon it for customer satisfaction and better sales.
2. Recommendation-based business that tracks customer's bought items and offer customer new items that they might also like.
3. Keeping good customers that boosts sales and avoiding fraudulent, bad customers.
4. decision support system for both business holders and/or customers[6]
5. Other applications that can build customer relationship.

Shopping

Shopping is the activity of trading goods and products for other goods and products. Simply put, exchanging a product for a more benefiting one, this can be; food, tools, money, ad etc. depending on your needs. In the early age of man, the art of exchanging goods took place in which leads to easier lives of the customers for they don't even have to hunt or gather food themselves in which fasten the growth of many communities. Markets or shopping malls have been the main place of trading, but they're often in urban areas or in the middle of the city which can be, depending where you live or just lazy, can be hard to go to.

As technology develop and as our understanding of computers and the internet as well as the remaining difficulty of going to the market place, landlines, the internet and trading companies gave birth to ticket booking and food delivery services, and online shopping websites (OSWs). Pizzas, clothes, and pizza, can be delivered in front of your doorstep in a matter of minutes by the usage of the internet by visiting such OSWs. But be wary of scammers especially in the e-shopping category as they run rampant waiting for an user to be conned.

As the danger of scammers aren't enough, one of the most commonly problem encountered by many users is the tedious work of manually comparing and searching of a particular type of product in many OSWs. Users have to go to multiple OSW just to type in the same keywords and is faced with many similar product in that OSW plus many more with the other OSWs. Reviewing the 'reviews', as well as looking in to the quantity, and the quality of the searched product come hand in hand which makes it hard for the comparison process thus making a decisive and long decision.

Fortunately, 'Comparison of ecommerce products using web mining authors', have started a way for tackling this issue way easier with the use of Data mining. Studying the HTML code structure of international OSWs such as Lazada and Amazon, and using web-crawling spider to scrape out unstructured data and organizing them to a meaningful data. They let the user input search terms that will display the comparing pair deals, successfully lessening the user's time for searching manually. Displaying it in a single-view page for users to view and pick what they desire. However, displaying and adding a table of comparison for such products still doesn't solve the problem of finding the best deals a site can offer. Finding the best deals have to account the reviews of other users, quality, quantity, and availability of the product.

Chapter 3: Research Methodology

3.1 Research Method Used

3.2 Research Paradigm

3.3 System Architecture

Figure 3.3 System Architecture

3.3.1 Preprocessing

3.3.2 Main Process

3.4 Population Frame and Sample

3.5 Description of the Respondents

3.6 Sampling Technique

3.7 Instrumentation

3.8 Data Gathering Procedure

## 3.9 Statistical Treatment

3.9.1

Chapter 4: Presentation, Analysis and Interpretation of data

The purpose of this study was to develop a systems:

4.1 Accuracy of the System

4.2 Reliability of the System

4.3 Reliability Statistics

Chapter 5: Summary of findings, Conclusion and Recommendation

5.1 Summary of findings

5.2 Conclusion

5.3 Recommendation

5.3.1 Context-Based Approach

5.3.2 Real Time Detection

# **REFERENCES**

x

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

x

# Appendix A

Appendix A. Experiment Paper

Table A.1 Experiment Paper

Table A.2 Experiment Paper with Raw Data

Appendix B

Appendix B. Implementation Report

**Introduction**

**Problem Statement**

**Respondents/Subject**

**Time frame**

Table B.1 Time Frame for Implementation

|  |  |
| --- | --- |
| Activity / Tasks | Schedule |
| **Audio Gathering** |  |
| **Testing the Accuracy of the System** |  |
| **Computing the System's Accuracy** |  |
| **Chapter 4** |  |
| **Chapter 5** |  |
| **Finalizing Document** |  |

**Participants:** Researchers **Required Tools and Equipment:** Laptop

**Implementation Procedures**

The developed system is intended to be tested in terms of Accuracy and Reliability.

**The following are done by researchers during system implementation:**

1. **Gathering Data.**
2. **Testing of the System**
3. **Checking of the output**
4. **Determining the Accuracy and Reliability of the system.**

**Illustrations and Pictures**

Photos: Development, Testing and Implementation Period

Photos: Development, Testing and Implementation Period

Photos: Development, Testing and Implementation Period

Photos: Development, Testing and Implementation Period

Appendix C

**Appendix C. Graphical User Interface**

**Graphical User Interface Screenshots**