**Thesis Proposal:**

**TIPid: Decision Support System on E-Shopping Products through Graphical Visualization and Data Mining**

Averos, Christian M.

Delicano, Jobea Ann F.

Chapter 1

THE PROBLEM AND ITS BACKGROUND

This chapter introduces the problem and the background as well the supporting information that will be used throughout the study.

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 (Pastore, 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 great effect 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.

Alongside of many frameworks for building mobile and web applications, acquired data must be represented in a formal and understandable way for users to grasp in first glance the importance of this study. Visualizing dynamic data in a user-friendly interface can make users comprehend easily the content of the acquired unstructured data sets. Therefore, the researchers decided to develop a system to help with troubled online shoppers (OS).

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 (Bolido, 2014) even in the most trusted OSWs.

Visualization is the representation of data in which is any technique for creating images, animations, and etc. to convey a message. Truly, life would be boring if there weren't many colors or shapes around. Both became one of the major tools in life for differentiating and comparing. In today's application of visualization, one of the tools it is used for, when it comes to websites and computer applications, is helping the users' on their experiences. It help them in the form of 'navigation', or even the color and theme of the website to give more feeling of appropriate based on the contents.

With the visualization as the main part of knowing, and criticizing, and differentiating can be implemented in a decision-support system in which colors and shape helps in choosing one in the items in question. 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.

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

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 and apply a visualization technique to display the best deals. Specifically, this will address the following sub-problems:

1. Using the a-priori algorithm, what is the accuracy of getting the relevant data from online shopping websites?
2. What is the accuracy and reliability of the devised algorithm in getting the best deals?
3. What is the accuracy of the displayed data based on the generated best deals?

**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.

Scope and Limitations 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.

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.

Definition of Terms

Data Mining. It is the 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.

Chapter 2

REVIEW OF RELATED LITERATURE AND STUDIES

This chapter discusses papers related to the study including Data Mining: its techniques, tools and application; visualization techniques; and shopping: physical shopping and its problems, OSWs and its problem and application of data mining in OSWs

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. (Ramageri, 2010) 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). (Li & Beaubouef, 2010) KDD consists procedures including (Smita & Sharma, 2014):

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 (Bhatnagar et al., 2012) :

Association. 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.

Classification. 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.

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.

Prediction. It is one of a data mining techniques that discover relationship between independent variables and relationship between dependent and independent variables.

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.

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. (Ramageri, 2010) 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. (Li & Beaubouef, 2010) Sales industry also use DM as in different ways (Al Essa & Bach, 2014) :

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 (Smita & Sharma (2014)
5. Other applications that can build customer relationship.

User-interface and visualization

Visualization provides user a comprehensive understanding about a data. It gives form to raw data making it coherent to the users. Using computer graphic effects, data sets can be visualized to display patterns, relationship and trends in a more advanced way. This gives users an ease to apprehend information by the use of visual reasoning rather than numerical reasoning. (Zhao, 2002) Presentation of data may involve the manipulation of graphical entities and attributes. A good data visualization must consider the effectiveness or ease of interpretation, accuracy or correct quantitative evaluation, efficiency or data redundancy removal, aesthetic or must be pleasant to the user's eye and adaptable or adjustable to serve multiple needs. According to Ward, common techniques on data visualization are following: charts, graphs, plots, maps, images, 3D surfaces and animation. Parsaye and Chignell listed the common steps followed in data visualization:

1) Numerical transformation of data by modifying the distribution

2) Data analysis to interpret data that will be used in graphical interpretation

3) Graphical interpretation by plotting the analyzed data onto graphs

4) User interaction by adding options for users to dynamically adjust mapping, zooming, panning, highlighting, et cetera.

**Shopping**

Shopping is the activity of trading goods and products for other goods and products. (Robertson, 2012) 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. According to Oxford, 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. (Bund)

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 a 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**

This chapter defines the research methods used to conduct the study. It involves the research method that will be used, paradigm, system architecture and the data gathering procedure.

**System Architecture**

Figure 3.1 shows the system architecture.

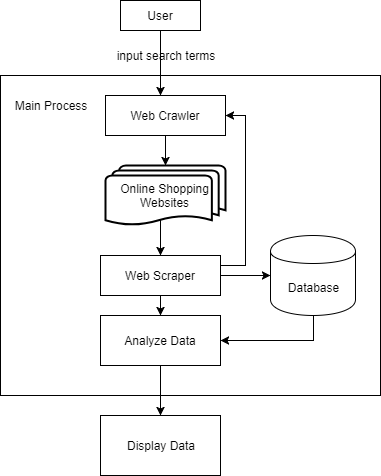


Figure 3.1

**References**

Pastore, M. (1998, December 03). Internet Becoming Necessity to Users. Retrieved from https://www.clickz.com/internet-becoming-necessity-to-users/72138/

Kitonyi, N. (2017, March 07). E-commerce is Killing Traditional Retail. Retrieved from https://www.gurufocus.com/news/490164/ecommerce-is-killing-traditional-retail

Bolido, L (2014, April 23). Safety in Online Buying and Selling. Retrieved from http://lifestyle.inquirer.net/157606/safety-in-online-buying-and-selling/

Ramageri, B. M. (2010, December). Data Mining Techniques and Applications.

Li, Y. & Beaubouef, T. (2010). Data Mining: Concepts, Background and Methods Of Integrating Uncertainty In Data Mining.

Smita & Sharma P. (2014, June). Use of Data Mining in Various Field: A Survey Paper.

Bhatnagar, Jadye, Nagar. (2012, November). Data Mining Techniques & Distinct Applications: A Literature Review.

Al Essa, Bach, C. (2014, April). Data Mining and Warehousing.

Kaidi, Z. (2002). Data visualization.

Ward, M. Overview of Data Visualization. Retrieved from www.cs.wpi.edu

Parsaye K. & Chignell M. Intelligent Database Tools & Applications.

Robertson, P. (2012, November). Robertson’s Book of Firsts: Who did what for the first time?

Oxford: Oxford University Press. (Accessed 12 August 2017) Retrieved from https://en.oxforddictionaries.com/definition/market

Bund: friends of the earth, Germany. (Accessed 12 August 2017) Shopping by bike. Retrieved from http://www.einkaufen-mit-dem-rad.de/shopping\_by\_bike.shtml