**Product Recommender System based on the demand of the customer in the University Store of AUP using Predictive and Association Mining**

A thesis presented to the Faculty of

Information Technology Department

College of Business

Adventist University of the Philippines

In Partial Fulfilment

of the Requirements for the Course

Capstone Project

Lino Asmolo Jr.

John Enoch Ortega

Jayson Tindoy

December 12, 2019

**Chapter 1**

**Introduction**

**1.1 Project Context**

“Data, I think, is one of the most powerful mechanisms for telling stories. I take a huge pile of data and I try to get it to tell stories.” – Steven Levitt

Every day that passes by. Electronic gadgets or machines input thousands and thousands of data. Most of those data are just being deleted every end of the day without even looked upon at. But there is process that can be used to make those everyday data useful that can help in our everyday Living

Data Mining is process used to turn raw data into useful information. By using software to look for patterns in large batches of data, business can learn more about their customers/users to develop more effective marketing strategies, increase sales and decrease costs.

It may be the 21st Century but still most companies or corporations use manual survey to gain information on how they can do better marketing strategies, increase sales and decrease costs. Though manual survey is still effective, it is not that efficient given that it takes days, weeks, months or even years.

With data mining, it’s a different story, for it uses already stored data and with those already stored data it creates patterns and algorithms which can be used as useful information not just for the companies or corporations, but also to everyone.

**1.2 Description of the Project**

University Store (UV) of AUP was established back then at the time when PUC moved from Baesa to Puting Kahoy, Silang, Cavite. Since at that time, the exact location where it is located right now is still the same. Situated along the road where customers passing by could easily see it. Before reaching Boy’s Dormitories, the building can be seen. With its strategic location, University store would easily attract customers.

UV is a two-stories building with around 30 working students that assist the department in its daily transactions. Headed by Mr. Arlene Fajardo and supported by his six staffs. The department provides AUP campus primary needs like rice, vegetables and fruits. Personal stuff products like shampoos, toothpaste, bath soap and skin care products. Also, it sells things like t-shirts, bags, shoes, socks and many more. Roaming inside the building, going upstairs where the school supplies located. UV also provided a small section that caters Postal documents.

According to Mr. Macaraig, current accountant of UV, based on previous years of the departmental operation of selling products to University and other walk-in customers, UV is gaining an income.

**1.3 Purpose of the Project**

The AUP University Store is still a growing business that sells goods and other things to the students and to the public. We are hoping to enhance the marketing and sales capability of the AUP University Store by using data mining to help the store runners interact with the customers through the data mined and patterns and algorithms it has created.

Listed below are the problems and issues the AUP University Store are facing which hinders the marketing and sales capability of the store.

1.

2.

**1.4 Project Objectives**

**1.4.1 General Objective**

The main objective of the project is to develop a (proper system) and (system) for AUP University Store in AUP Silang, Cavite wherein the marketing and sales capability will be enhanced and to promote good customer service

**1.4.2 Specific Objective**

1. Develop a system that collects everyday data in every store transaction

2. Develop a system/software that reads and transforms the collected data into patterns and algorithms

3. Develop a software with computerized marketing and sales plan that uses the collected data, patterns and algorithms

4. Enhancing the marketing and sales capability of the store by using the developed system and software

**1.5 Scope and Limitations of the Project**

The (system) is a web based and software system that can be used in any computer that runs in Windows 7 to Windows 10. The proposed system will use HTML, CSS, PHP, jQuery, framework, JavaScript, MySQL, server as the database of the system

The proposed system covers the display of the patterns and algorithms that the system has made using the collected data. The system also has a monitoring system to keep track on every transaction made in the store that also be used as raw data for the system to process.

The proposed system also displays any statistical data by using the processed collected data that can be used for better marketing and sales planning.

**1.6 Significance of the Project**

**Business Marketing and Sales Planning**

This system can be best used for marketing and sales planning for it uses credible data and accurate mathematics through statistics

**Store Business**

This proposed system can be used to monitor what the customers like to by the most and what to buy the least. With this useful info, the store runner will know what to sell and what not to sell

**Business Planning**

This proposed system can be used globally for it is web based and with that in mind, aspiring entrepreneurs around the globe can use this system to plan their business carefully by using and studying the information statistically and technically

**Chapter II**

**Review of Related Literature**

**2.1 Technical Background**

**2.2 Review of Related Foreign Systems**

**2.2.1 Recommender System for Journal Articles Using Opinion Mining and Semantics**

Building a system that would recommend relevant journals, articles, books and publications to author/s would greatly affect researchers’ studies where they could gather significant information related to their topics. Recommender systems have done its part on filtering sources as to what preferences searcher wants. With the rapid increase of published research papers like journals or articles, one may find difficult to choose of which one research papers is appropriate to read upon and collect information, that in return it would consume much time looking for it. A good choice of information means a good result of research paper that would impact authors work. As a result, the authors decided to develop a journal recommendation system, this system will suggest journals and help authors to decide and to choose suitable papers.

On developing the system, they meet difficulties like to which journals will be collected and to where this gathered data will be stored. Also, how they can do a better and much accurate recommendation of journals to users as well as to what technique would fit for the recommender system to use. The authors find a way of solving these problems by incorporating user-given opinions to their experiences of reading, subscriptions and words that are used in that journal in their data mining. By gathering this valuable information about the users, they believe that they can solve the issues and provide a significant and appropriate journal to the users.

Two frameworks were proposed in this system to gather data in which will produce recommendations of journals on the latter part. The first is User Opinion Analysis where data comes from the survey given to the user with regards to their individual comments of reading the journal and subscription of it. Surveys were done through online using google forms and offline via hard copies of survey form. Different tools were used to processed data gathered from the survey to get good information about the journal. The result of this analysis will determine whether they see user if it is helpful or otherwise not. The second is Semantic Similarity based Approach. This mainly means gathering the description of the Journals provided. Each detailed information of research paper like title of the research, keywords provided, abstract stated were collected and considered as an input for recommendation as well as the frequency of publication and citation score and any other related to journal. With the recommendations as a result of the second framework and the recommendations from the user opinion were combined. Using a term to term relationship, both was analyzed to find similarity of terms.

In conclusion, authors were able to get good recommendation of journal using a combination of user experiences on journal and a detailed information of each journal. Also, they found out that submitting research to a journal archive like “Big Data Research” is a good decision to make since it rates 3 in their query which can make a good impact in the authors side. They also wanted to conduct a similarity check to tools namely Hadoop and Spark to compare if their recommendations of journals are accurate and relevant to the user.

**2.2.2 An Improved Collaborative Filtering Based Recommender System Using Bat Algorithm**

Finding the exact information becomes an issue since there are lots of it published and uploaded in the web. To give solution to this matter, Recommender Systems were introduced that can suggest items and help user improved their searching experiences to their item preferences. In this research, the authors proposed a recommender system that is solely based on Bat Algorithm to improve its collaborative filtering of information.

The idea behind collaborative filtering is that recommending items based on similar suggestions of users. Using this filtering technique is a good choice since it has something to do with users and items relationship. Yet, drawbacks can always happen like when a new item or user is created, provided information of item or user is quite little. Recommender systems work with information gathered from the user and items itself that it suggests then, and if there is not enough information is available, that's becomes an issue. But researchers were not left behind with this problem, techniques were created to address it and minimize the effect with the ability to render quality recommendations.

In line with this filtering, heuristic techniques were also incorporated to optimize quality searches of item features that produce a better clustering of users having similar weight of features. Swarm intelligent techniques like Artificial Bee Colon (ABC) Bat Algorithm (BA) and Cucko Search are techniques used to improve searches. The idea of developing this system that proponents want to achieve are as follow: they can provide customized suggestions for every user and optimize recommendations by incorporating techniques mentioned above in their collaborative filtering process. Among the three techniques, proponents used the Bat Algorithm as a result of the experiments done that BA performs well and has provided better results of recommendations. In the review of related literature, the authors found out that collaborative filtering when used with swarm techniques could bring difficult situation to achievable results, and with this, the authors tried to use Bat Algorithm “to generate future weights” with the Pearson Correlation Coefficient to see similarity between users.

Bat Algorithm (BA) comes from the idea of a bats in their method of looking for food. Moment by moment eat bat sends sound wave which reflected when it hits any object. Based on the reflected waves, the bat will know how close or far the subject is and find better solution how to catch it. In relation to a better recommendation, bat algorithm keep on producing optimize solution for a better recommendation.

In conclusion, the authors prove that using Bat Algorithm generates an accurate recommendation, also parallel experiment were done with the mentioned algorithm to measure its effectivity and results were true, BA performs well with a 6.9% better than ABC does. The authors are also looking forward in their future development to categorize results of recommendation into bad, good and average, so a better recommendation can be achieved. Also, they want for the future research to try social networking to learn user’s behavior and traits.

**2.2.3 Product recommendation based on shared customer’s behavior**

The system was developed by Fatima Rodrigues and Bruno Ferreira from Polytechnic of Porto in 2016 to provide accurate recommendation of products. The recommender was implemented to a company that operates over 30 years and sells products like perfumes, make-ups and body care products with about 250,000 customers and 110,000 products in a given period time from 2012-2014 purchases. Also, this recommender is to suggest products to the loyal customers only.

In a selling environment where there are thousands of products available, issues would normally arise to the company how the can cater the needs of every customer by recommending relevant products to them, what products will be bought by the customer on the next purchase or store visit to see the availability of products on stocks. Giving a closer look how to resolve these issues will help company to keep intact their customers to their products which in the long run would normally increase sales and by doing so, they can gain new customers who are attracted to the treatment the company gives.

In this project, the authors developed a recommender system to solve the problem mentioned above. The recommender system will be used to understand the preferences of customers of the product that they put into their market basket. Understanding the customers market basket will benefit company in gaining more profit.

In this recommender system, the authors look on how they can implement it to make recommendations, what products would be considered that is worth to be recommended since there are thousands of it, how they can group the similar customers with the same preferences and identify the relationship between each products that are purchased by similar customers at the same time.

To implement the recommender system, the authors present the following concept of how it was made: first is to use the Customer Lifetime Value by making evaluation of their shared behavior shown towards the products on the recency, the last time that the customer purchase a product; frequency, how many times of product purchase in a given period on time; and monetary, the amount of money that was spent during that time. Second is clustering using partition algorithm, this will determine customer with the same life time value and to where group it belongs. To achieve groupings, objects should be like each other object which also creates a division from those customers with different attributes. Third is an association rule mining technique is done which identifies relationship between items in a transaction. This literally means that on a purchase, if product “A” is bought, product “B” will be bought as well. And to identify those relationship, Apriori was used, a popular algorithm that is known for collecting association rules.

In conclusion, using recency, frequency and monetary to group customers, the authors were able to generate recommendation best and appropriate product to those specific group of customers. The authors also added that doing recommendation on a specific time enables to generate recommendation of products on seasonal basis. In general, they said that the methods they used in recommender system made an excellent recommendation.

Rodriques, F & Ferreira, B (2016)

**2.2.4 Tourism recommender system using Case Based Reasoning Approach**

**(Case Study: Bandung Raya Area)**

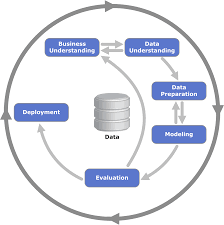
The system was developed by Bamban D. Fatmawatie and Z K A Baizal, from School of Computing, Telkom University, Bandung, Indonesia in 2019. This

**Chapter 3**

**Methodology**

Many IT projects comes up with failure, this is because they don’t follow the rules of software engineering process and by not having a model to follow in their projects. Having a model in your IT project has a huge effect in developing a system. By following the steps on your chosen model, you’ll be guided on which steps to do, you perform first and you need to consistent to your model.

There are many types of model today that we can use to our project. We adopt an Agile methodology for the carrying out of data mining and the model we were going to use is the, which stands for “Cross Industry Standard Process for Data Mining”. This has been proven for the construction of data mining.

****The methodology's assumption is the willingness to make the data mining process reliable and usable for people with few field skills but with a high level of business knowledge. The approach offers a six-phase structure that can be replicated as in a loop to update and refine the prediction model:

* Business Understanding
* Data Understanding
* Data Preparation
* Modelling
* Evaluation
* Deployment

CRISP-DM Process Diagram

**3.1 Requirement Analysis**

Requirements Analysis is the mechanism by which users ' requirements for an app to be constructed or updated are established. Assessment of specifications includes all the activities conducted to address the needs of various stakeholders. Specifications assessment involves evaluating the code or device specifications, recording, validating and handling them. Documentation, action ability, measurability, testing, traceability, help in identifying business opportunities and are defined to facilitate system design.

In developing a system, it’s very important to have a good communication between the proponent and the user for us to have a persistence analysis to business to complete the task. It is necessary to know the specific business transaction in every company to conceptualize the activities and to determine the section of the process that we need to improve by applying the trends of technology.

**3.1.1 System Analysis**

1. The raw data of their database from POS will be connected to the Product Recommender System to mine the existing data and generate a potential report about the data.
2. After mining the data, we will use the Association rule mining and Predictive techniques to determine its characteristic and it will be used for generating a report and which part is the sales is high.
3. When the system finish analyzing the data, I will generate a report to each section that will help them for improving and preparing their product efficiently and accurate.
4. Also, it will generate a summarized report to the management enable them to see the growth and build some idea how they will market their product.
5. And display real time graph sale of the POS database.

**3.1.2 Data Flow Diagram**

D1

Merchandising

Section

Raw Data

Summary Report

Sales Report

POS Database

Accounting

Section

Store Management

0v

University Store

Product Recomender System

Purchase Order Info

Graph Report

Predicted Product Trend by Month

Product Order Info Based on Predicted Products

Predicted Products by Month

Predicted Products by Week

Predicted Products by Day

**3.2 Requirement Documentation**

All IT project must have a requirement documentation so that you can keep track the changes and have explanation on every single module in your system. The documentation will be the basses of the company how it’s really work and how the system build. It is the basses of the IT department for the maintenance of the system if there is any malicious issue of the system and also, the security of the system.

For each document as well as the complete project documentation, documentation shall provide a basis for quality, standardization and history. The documentation must also be well arranged, readable and adequate. It also substantiated by the two essential documentation functions: ensuring that project requirements are met and establishing traceability regarding what was done, who did it, and when it was done.

The Product Recommender System is programed to mine the raw data of the existing database to generate into a useful report that can improve the company specifically in the section of merchandising to predict some specific product that the demand is high in particular day, week, or month. For the convenient and efficient way of preparing the product and less hassle to the management.

The system has three section: Accounting, Merchandising, and the Management. The accounting section can see the sales report and take the approval of the purchase order that came from the merchandising section. In merchandising section, where the predicted product appear, the merchant will prepare the predicted product to be sell and to be order. And the management section will see the real time sell growth and summary report. By the help of this system, both section has a benefits of saving time to prepare what to order, what to sell, and what is the lead sells growth of the company.

The system has three major accounts: administrator account, accounting account, merchandising account. Each account has specific rule and deferent function to perform. Administrator account have access to all module in the system and has control to the all rights in the user specifically in add user module where this system has some confidential information and not all can have access to this system to remain the privacy of the company and only the management must have the account. In accounting account, only modules that are related to his area will be shown like sales report module, purchase order module, and etc. And also, in merchandising account, this account will see the predicted product modules where the specific product are suggested to prepare or to order.

3.2.1 Login Module

3.2.2 Dashboard Module (Administrator account)

3.2.3 Dashboard Module (Accounting account)

3.2.4 Dashboard Module (Merchandising account)

3.2.5 User Account Module

3.2.6 Notification Module

3.2.7 Predicted Product module

3.2.8 Sales Report Module

3.2.9 Purchase Order Module

3.2.10 Summary Report Module