

CETM46 Assignment 1: Literature Review Report

eCommerce data product

Prepared by: Fung Yue (Lewis)

Date Prepared: June 7, 2020

<p style="text-align: center;">Assignment 1 of 2 –</p> <p style="text-align: center;">Project ID No.: 1</p>		
--	--	--

Document Information

Assignment 1:	Assignment 1		
Prepared By:	Fung Yue (Lewis)	Document Version No:	1.3
Title:		Document Version Date:	7-June-2020
Reviewed By:		Review Date:	

Distribution List

From	Date	Phone/Email
Fung Yue (Lewis)	June 7, 2020	Bh63qb@student.sunderland.ac.uk

Version History

Ver. No.	Ver. Date	Revised By	Description
1.3	7-June-2020	Fung Yue(Lewis)	CETM 46 Assignment with Ref.

Table of Contents

1	Abstract.....	4
2	Introduction	4
3	Methodology: Decision making & data problem.....	6
4	Public datasets available	7
5	Product design planning, ML models, visualization and software development	8
6	Conclusion and future work	10
7	References []	11

1 Abstract

This paper provides a literature analysis of Data Product design applications related to ecommerce. As the popularity of online shopping continues to grow, consumer intensions are present at various points of the e-commerce relationship, ranging from pre-purchase intentions to post-consumption behaviour. In the combination with best practice eg. Scrum, problem statement of dataset, machine learning selection and relevant contributions of online shopping behaviours were identified in ecommerce area is also introduced on the basis of this literature review.

2 Introduction

Ecommerce Ref.[6] is growing rapidly, and the businesses are willing to invest more on enhancing the online experience with the data product related development. In Commerce Technology Takes Off, Forrester Research authors wrote that in 1999 online business to consumer retail spending was \$20.3 billion, and by 2003 it is estimated that it would rise to \$144 billion. In 2000, global 2,500 businesses will spend 72 per cent more on ecommerce than they did in 1999. Existing sites use primitive measures, such as page views, but with the need to differentiate, the need for more serious analysis and personalisation is growing rapidly. The online data product application to business has grown explosively in the last 10 years. It provides benefits to the companies: -

1. Predication of the sales quantity
2. Find the associations among the items
3. Predication of stock price
4. Predication of network or operational issues

Ref.[8-9]

These analyses have been structured as methodology for demonstrating objective conclusions because they are based on the big data, facts according to a system logs, stock price news related article, point of sales, and registered membership system such as customer relationship management (CRM). As a result, the decision makers in business tend to rely on the “Reliable” results obtained by the data product applications of machine learning (ML) or artificial intelligent (AI) and make decision because of ML or AL predict the result in the future.

Assignment 1 of 2 – Assignment 1

Project ID No.: 1

A problem with this approach arises because the ML algorithm is often superior to humans. For example, if a customer orders a book by entering the title in the online bookshop search engine, he or she may find a list of five books recommended. However, the customer already owns two books and is already planning to purchase the other three. As a result, the system cannot provide him with any new information, because he was very sensitive to newly published books and knows more than the results of the search. We had tried relevant search results or relevant advertisements when searching online or on social platforms or streaming movies and music.

From the CRM, ML and AI perspective, customer service of the bank or company has implemented their robot – ChatBot to assist customer services helpdesk, the chess master lost in the match to AlphaGo. AI has appeared in education as well. In US, Europe, Asia, China, AI has used to help compose papers, analyse sentences, and help teachers mark points in the examination. AI also read the answers of exercises in mathematics or other sciences, so that it reports the strengths and weaknesses of students, and then provide advice to teachers in order to closely follow up with students. AI in a broad sense, is a way of imitating human thinking through a computer or computer. Of course, it must have the ability to learn by itself to solve problems, not a simple mechanical reaction. Ref. [11,19]

ML: Through some algorithms to analyse online data to get the answer, through continuous learning and massive data, get more and more accurate answers. However, the conclusions drawn by mechanical learning are sometimes wrong, and data scientist needs to correct its calculations. Deep Learning (DL): It is an evolved version of mechanical learning. Data scientist imitate human thinking methods, and through layers of artificial neural networks, let the machine decide whether its conclusions are correct. Of course, if it keeps learning repeatedly, the more data they have, the more accurate it will be. Natural Language Processing (NLP): It is to let the machine know the language of human beings and use it to communicate and receive instructions. This is very important in medical treatment, because the machine can extract important information from the patient's large amounts of data and text reports for analysis. Few examples of machine learning application that we use everyday - Smart Speakers: Amazon Echo and Google Home, Smartphones, Samsung Bixby on Samsung S8, Mobile Apps: Google Allo. Ref. [11,19]

3 Methodology: Decision making & data problem

When ecommerce comes to online data, there are two main activities we should spend the time on. Ref. Foster Provost and Tom Fawcett Published Online:13 Feb 2013 conduct a topic Data Science and Data-Driven Decision Making”.

These are the following:

Ref. [13,23]

- Periodic KPI inspections: It means getting the right "dashboard" to see if it all goes as expected, or if anything is off. Periodic KPI reviews function better when you are checking your KPIs for some form of schedule. Whether you feel it's weekly, bi-weekly, or monthly, it's important to know the time to have these maintenance "sweeps" of KPIs.
- Analysis of deep issues: If the periodic KPI inspections show that it has a problem somewhere, this is where the analysis of the deep issues comes in. This is when we dive deep into the data, attack the root of the problem, and then take action. The Deep Question Analysis is all about understanding the "Why" behind the issues.

The next question is what should we spend with the data analysis?

The problem most online shop face in today's world isn't too little data – it's getting too much data and not knowing what to do about it. The secret to data-driven decision-making is to know how to collect the right data, evaluate it for the right insights, and draw the best conclusions and strategic decisions based on the company and objectives.

1. Exactly what should be looking for

The key objective of decision taking which is guided by data? Depending on priorities, to lead us to the right conclusions and help us make the right strategic decisions for online shop. Which means the first, most critical step in the data review is to take a moment to consider what we are searching for. Think what our ambitions are, and what questions we would like to raise. Are we seeking to gain more new clients, boost retention or increasing the effectiveness of our marketing activities? This will allow us to narrow down the right data sources for evaluating and understanding which KPIs to concentrate on.

Ref.[12-14]

2. Link the sources and gather data

When we know what we're looking for, find the correct data sources and make sure we are able to pull the information we need – whether that's the ecommerce site and advertisement platforms, email marketing and automation, new sales channels, order management and delivery, consumer satisfaction and feedback, customer service and more. We can collect data by going to an individual site – but if we have an analytics platform that allows us to link several data sources, it will make it much easier to analyse and visualize the data.

3. Data Review

We need to analyse our data to determine the different metrics we need once we have linked our sources and collected the data we need. ML, Tableau and Business intelligence (BI) platform can help – adding useful visualizations and flexible comparisons over time in addition to static metrics.

4. Seek perspectives that will drive strategic decisions

We look at the data you obtained, and the research of the online shop performed, what conclusions can it draw to help direct our decision-making processes? Our analysis may have helped uncover that one of our products gain more new customers than any other – we can use that to inform future acquisition campaigns and product placements. We may find that one of the marketing channels drives more revenue-with a better profit margin-than the others. It may be adjusted in spending accordingly. Regardless of the metrics we look at and the lessons we discover, make sure that we do so on a regular basis-data-driven decision

4 Public datasets available

We found the dataset consists of feature vectors belonging to 12,330 sessions. The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period. We believed only the recent data and consider for our analysis will more relevant according to our data product design in the future.

Ref. <https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset>

5 Product design planning, ML models, visualization and software development

To be able to implement an ecommerce data product, the plan, design, visualize the dataset and software development are very important. Ref [4-5,7]

SCRUM should be applied and skill we need: -

- Analyzing the environment and its specifications
- Transfer of data to gather needed data from different sources
- Pre-processing the data to create the necessary features for the model
- Visualisation to obtain information from the data and to speak to the client
- Templates for building predictive models
- Establishing a data pipeline to access a fully automated live environment
- Development of API or web service to open a prediction service
- Checking after going live to have a "peaceful life."
- The development of the dashboard to monitor manufacturing life in both technical and business manners.

Design the product and what we found in dataset?

First of all, we knew that website design effects on consumer intension according to the key findings, website features, design and atmosphere usually have a major impact on consumer emotions and behaviour. There are cultural differences in emotional responses and perceptions in online shopping environments. Emotions are crucial to the perception of the atmospheric and characteristic characteristics of websites and affect consumer attitudes towards the site, the level of involvement and the purchase intention but those type of relationship and questions are hard to answer without data analysis. We will use appropriate data analysis and machine learning (ML) algorithms for the relationship between all variables which can be utilized in making good predictions of upcoming sales and select the best ML with a website enabled application (Webapp) for users to adjust the goal of predictions on revenue outcome. Ref. [13-14]

Assignment 1 of 2 – Assignment 1

Project ID No.: 1

1. Conceptualizing - We must identify the function that meets a need from the store owner or management. This introductory step should take place before data acquisition.
2. Data acquisition – Acquire or accumulate the dataset that corresponds to functional activities. The acquisition process needs to integrate the conceptual model to the dataset.
3. Refinement – Data refining can be achieved with tools, real-time ML and algorithmic processing of the dataset can categorise, correlate, personalise, profile and search data quickly to create meaningful models that have significant value for store owner.
4. Storage – Data storage, retrieval and processing are taking place in cloud services provider (eg. Google and Amazon) rather than on own site premises.
5. Distribution – Data product should be able to distribute via mobile devices, if Web Application content formats support, it also needs to be designed for supporting both PC and mobile browser.
6. Presentation – The user interface mattered, the easier it was to use, the more valuable it was.
7. Company Feedback – Function and features must be tested in a matter of hours with A/B or multivariate online testing approaches.
8. Target user – We must have target user before defining the data product function, eg. Webapp can be use it for the Online Store Management to predict which product should be invested or promote according to ML result.

Ref. [16-18]

Create Web Application

A web application is designed to perform a variety of functions on client and/or server request. Through the Hypertext Transfer Protocol (HTTP) network, the client makes a request, the server processes that request, and the application tells it what response to send back to browser

Front-end layer: Creates the application's appearance and feel through technologies such as HTML, CSS and JavaScript. When we develop a web app, there are a lot of things we're going to need to figure out. The target user is important, and when we want to capture the attention of users, the design of our app will need to be of high quality and provide good ML prediction and result. Before we start building Web application, here are three key tips to consider: Ref.[23-24]

- Select a framework to streamline the process.
- Follow a proven approach to evolution.
- Think about Usability and Architecture.

6 Conclusion and future work

This paper introduces a conceptual online store analysis in the sense of the data product modelling and preliminary attempt at grouping and analysing peer-reviewed articles. Related to the intensions of consumers in the context of e-commerce and the best practice of using SCRUM, ML to Webapp of forecasting online shopper intention. In addition, we summarized and analysed the key findings of the studies reviewed. Recognizing the challenge of understanding data set, online customer intension was stronger, the purpose of this paper was to increase the awareness of our data scientist provided new insights by integrating such research findings. The analysis shows that we are present at different stages of the e-commerce relationship, in particular, ML model assessment certainly isn't just the end of our research, but we could be able to plan ahead before we treat any data and use methods that are suitable for other purposes. We will go over a number of these methodologies in the future and see how they fit into the larger dataset with more dependent or independent variable comparing different model performance which will become a typical workflow in data product design. That all predictive models will normally contain errors, and that, considering the stochastic nature of dataset and algorithms, that the real job of applied machine learning is to explore the space of potential models which not limited to the one we selected and discover what a good model result on our dataset look like relative to the best result of forecast and maximize the cost of spending when promoting the online store webpage.

7 References

1. Alpaydin, Ethem. *Introduction to Machine Learning*. MIT press, 2014.
2. Beeley, Chris. *Web Application Development with R Using Shiny*. Packt Publishing Ltd, 2013.
3. Bishop, Christopher M. *Pattern Recognition and Machine Learning*. springer, 2006.
4. Cui, Ya, Xiaowei Chen, Huaxia Luo, Zhen Fan, Jianjun Luo, Shunmin He, Haiyan Yue, Peng Zhang, and Runsheng Chen. "Biocircos. Js: An Interactive Circos Javascript Library for Biological Data Visualization on Web Applications." *Bioinformatics* 32, no. 11 (2016): 1740-42.
5. Dale, Kyran. *Data Visualization with Python and Javascript: Scrape, Clean, Explore & Transform Your Data*. "O'Reilly Media, Inc.", 2016.
6. Guo, Stephen, Mengqiu Wang, and Jure Leskovec. "The Role of Social Networks in Online Shopping: Information Passing, Price of Trust, and Consumer Choice." Paper presented at the Proceedings of the 12th ACM conference on Electronic commerce, 2011.
7. Hossain, Emam, Muhammad Ali Babar, and Hye-young Paik. "Using Scrum in Global Software Development: A Systematic Literature Review." Paper presented at the 2009 Fourth IEEE International Conference on Global Software Engineering, 2009.
8. Jordan, Michael I, and Tom M Mitchell. "Machine Learning: Trends, Perspectives, and Prospects." *Science* 349, no. 6245 (2015): 255-60.
9. Langford, John, and Ruben Ortega. "Machine Learning and Algorithms; Agile Development." *Communications of the ACM* 55, no. 8 (2012): 10-11.
10. Liu, Yang, Chaoshun Zuo, Zonghua Zhang, Shanqing Guo, and Xinshun Xu. "An Automatically Vetting Mechanism for Ssl Error-Handling Vulnerability in Android Hybrid Web Apps." *World Wide Web* 21, no. 1 (2018): 127-50.
11. Marsland, Stephen. *Machine Learning: An Algorithmic Perspective*. Chapman and Hall/CRC, 2014.
12. Michie, Donald, David J Spiegelhalter, and CC Taylor. "Machine Learning." *Neural and Statistical Classification* 13, no. 1994 (1994): 1-298.
13. Murphy, Sarah Anne. "Data Visualization and Rapid Analytics: Applying Tableau Desktop to Support Library Decision-Making." *Journal of Web Librarianship* 7, no. 4 (2013): 465-76.

Assignment 1 of 2 – Assignment 1

Project ID No.: 1

14. Nootyaskool, Supakit, and Pimolrat Ounsrimuang. "Optimization Algorithm Using Scrum Process." Paper presented at the 2016 International Conference on Machine Learning and Cybernetics (ICMLC), 2016.
15. Ranganathan, Chandrasekaran, and Shobha Ganapathy. "Key Dimensions of Business-to-Consumer Web Sites." *Information & Management* 39, no. 6 (2002): 457-65.
16. Rising, Linda, and Norman S Janoff. "The Scrum Software Development Process for Small Teams." *IEEE software* 17, no. 4 (2000): 26-32.
17. Sarkar, Deepayan. *Lattice: Multivariate Data Visualization with R*. Springer Science & Business Media, 2008.
18. Schwaber, Ken, and Mike Beedle. *Agile Software Development with Scrum*. Vol. 1: Prentice Hall Upper Saddle River, 2002.
19. Suresh, K, and R Dillibabu. "Designing a Machine Learning Based Software Risk Assessment Model Using Naïve Bayes Algorithm." *TAGA J* 14 (2018): 3141-47.
20. Toth, Arthur, Louis Tan, Giuseppe Di Fabbri, and Ankur Datta. "Predicting Shopping Behavior with Mixture of Rnns." Paper presented at the eCOM@ SIGIR, 2017.
21. Wang, Lidong, Guanghui Wang, and Cheryl Ann Alexander. "Big Data and Visualization: Methods, Challenges and Technology Progress." *Digital Technologies* 1, no. 1 (2015): 33-38.
22. Wang, Yong Jian, Michael S Minor, and Jie Wei. "Aesthetics and the Online Shopping Environment: Understanding Consumer Responses." *Journal of Retailing* 87, no. 1 (2011): 46-58.
23. Wesley, Richard, Matthew Eldridge, and Pawel T Terlecki. "An Analytic Data Engine for Visualization in Tableau." Paper presented at the Proceedings of the 2011 ACM SIGMOD International Conference on Management of data, 2011.
24. Yu, Yiming, Yidan Ouyang, and Wen Yao. "Shinycircos: An R/Shiny Application for Interactive Creation of Circos Plot." *Bioinformatics* 34, no. 7 (2018): 1229-31.