National College of Ireland

Project Submission Sheet – 2020/2021

School of Computing

Student Name:	Sai Chethan Singu		
Student ID:	x18181937		
Programme:	MSCDA	**Year:**	2020
Module:	Domain Applications Of Predictive Analytics		
Lecturer:	Vikas Sahni		
Submission Due Date:	28-06-2020		
Project Title:	Predictive Analysis Of Online Shoppers Intention		
Word Count:	2953		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: Sai chethan singu

Date:

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1.Please attach a completed copy of this sheet to each project (including multiple copies).
2.**You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Abstract— **There is a huge surge in the percentage of people doing online shopping which is otherwise known as e-shopping in recent years. The continuous growth of sales indicates that e-commerce has enormous potential in the market. The main factors behind the growth include convenience, availability, time-saving and wide range of products. There is a change in the behavior of customers with respect to shopping. Due to modernization, online shopping has gathered much attention in the public. And in this internet era with flexible modes of payments and delivery were also the contributors. In the beginning, most of the people showed a tendency to do online shopping for the products that do not require physical feel. But later stages people slowly habituated to online shopping for all types of needs. Nowadays, the demand for analyzing the consumers behavior is increasing in order to predict the sales patterns so that desired revenue and hence profits can be achieved. Many companies have started real-time analysis for visitor's information who do online and offline shopping.**

I. INTRODUCTION

The main concern now, is that how to analyze the behavior of online purchasers as there will be no physical interactions unlike traditional way of buying. Research works has been done on this to analyze the patterns of online shoppers in terms of purchase. Nowadays, a new trend has evolved among the e-commerce websites that are identifying potential customers based on website browsing. The main motive behind this is to avoid losing the customers and to focus on the customers who contribute more revenue. On the other way to retain the potential visitors [1].

A recent study by the author predicts the intent of the shoppers who are likely to purchase by using data like pageview, session time and some other elements [2]. Another study proposed that the interaction between consumers and promotion channels provides the scope to predict their intent at the time of promotions [3]. A study on the behavior of university students states that the purchase intention is strongly associated with their friends, family and media [4].

The factor which is influential in driving the attention of the consumers is the pictures of products and this was revealed by a study. It was said that the product images played a major role in grabbing the focus of consumers in visiting the websites [5]. Another research work proposed that income and age are the two main factors behind the likelihood of online shopping. And should be accessible to older people also as they are not familiar with the day to day updating technology [6].

To what extent, factors like income, age, occupation and education background drives the online purchasing intention has been found out in this research work based on random sampling [7]. For the evaluation of customers shopping behavior through online, many parameters were taken into consideration and different techniques like conducting surveys, loyalty programs, promotional offers and so forth. are being used. So, with the careful monitoring of consumers behavior multinational companies are changing accordingly as per the consumer needs in order to achieve profits. One way of monitoring the behavior is that by collecting the data from the time when customers enter the website to shop until they leave the website. From that, online websites were able to target the potential customers by using target ads and recommendations based on past search history.

Scope:

This dataset analysis will help in finding potential customers and predicting the revenue. By using data elements like traffic, session information, page value, region and also other parameters were being used to analyze the behavioral patterns. These data elements were fed to machine learning algorithms to predict whether revenue can be generated from the customers. This can eventually help the company to know the customers preferences before hand and effective advertisement campaigns can be made within the budget constraints. Based on this, one can recognize or predict the willingness of customers in terms of purchasing the products or services online.

II. GOAL

1. Predicting customers online shopping intention in completing the transaction when a person visits the website and evaluating important parameters in contributing revenue.

III. ETHICAL CONCERNS IN E-COMMERCE

One of the rapidly growing industry along with technology is e-commerce. It has emerged as a wide platform for people to buy or sell products across the globe. Many ethical problems and challenges arose between buyers, sellers and vendors. Ethics plays a major role in terms of maintaining commitment and trust in the long run. Nowadays there are many e-commerce websites available, in order to differentiate them with others good ethical performances should be adopted. Dataset for this project is taken from the public domain called 'Kaggle'. This dataset was created by "Sakar" and "Kastro" in the year 2018. As the dataset is available publicly, prior consent is not required to proceed with further analysis. Generally, there are different ways of violating ethics as mentioned below.

Violations Of Intellectual Property:

Many companies have their own logo or design where most of the people recognize the companies based on them. Nowadays it became so common to replicate the famous companies brand logo and using without any consent or legal procedures in order to gain public attention. Good images, domain names and copyrighted content from the websites are prone to theft and used in other websites for their own popularity which is also unethical. In order to save money and time, some companies encourage these unethical practices.

Selling Fake Products:

Some of the online retailers sell counterfeit products as a regular practice. Labeling them as they are original branded goods and selling online and offline can be seen frequently. In some cases, companies themselves are fully aware that they are selling low quality and fake products in order to save money. It is done because people normally tend to buy branded products without noticing much of their properties.

Information Security:

One of the most crucial elements of ethical concerns is data privacy and protection. Many online websites collect customers information like their personal details, credit card or bank details, phone numbers and passwords. Companies should be aware of any spyware that can be used to watch the actions of the users and sends sensitive information to unreliable sources. Monitoring or tracking people identities, e-mails by e-commerce providers are also considered to be unethical. Setting up of fake or duplicate websites in the internet as imitation of real websites to gather sensitive information like their personal details and bank related information.

Product Quality:

Unlike the traditional way of buying in shops, online shopping completely replaces that. Earlier customers use to shop by checking the standards and quality of products or services. Now as there will be no direct contact with the retailers, those people are taking this as an advantage by selling products that are of low quality in the market.

Customer Service:

The online e-commerce websites act as a barrier between retailers and customers. Customers have to depend on online retailers in case of assistance, queries, support and for refunds. Many online retailers are offering customer services via telephone and chat to resolve customer issues and queries. There are companies where would not respond to customer emails or calls and convey false information to the customers which is considered to be unethical practices.

IV. STRATEGIES USED FOR ANALYSIS

There are various reasons why customers do not complete the transactions and return from the websites. There are many elements like eye-catching images, advertisements, unique content, flexible options play a major role in online websites. By using strategical approaches, targeted customers and expected profits can be achieved. The analysis is done not to predict the revenue but also to analyze the key factors for generating revenue in business perspective.

Social media is a key factor in targeting customers. It is a marketing strategy where online retailers concentrate on Instagram, Facebook and so forth by means of advertisements. The content from e-commerce websites is published or shared on social media platforms to promote their products and services. This will eventually target the audience in a larger number with greater visibility. Social media is one of the widely used strategies as it reduces the cost and saves time. And it is said to have higher chances of conversions of online shoppers by advertising in social media. Based on area and region, social media platforms target specific audience.

Advertising strategies improve online presence and boost sales and popularity. Initially, companies begin with advertising with Facebook Ads or Google advertisements as they are user-friendly. Advertisements are of either display or text-only content with targeted keywords. As content flows very quickly through the ads, it gained much popularity and widely used. Companies now are also analyzing the customers previous purchase history and automatically targeted ads were shown on his/her second visit. This is done to generate revenue from their second visit. They can even capture the visitors clicks in the website and predict whether a successful transaction can be made or not.

Content marketing strategy is used for building a good audience base by creating and publishing consistent and relevant content. With content marketing, companies can educate their customers by making tutorial videos, puzzles, webinars, polls and podcasts. This will increase their customer retention. Customers tend to change from one-time buying to repeat shopping in the websites. Hence company reputation and community base will grow gradually. This is a widely used strategy by the companies to boost their online visibility.

Public relations is about having good or positive image in the public and to increase their brand awareness. The reputation of the company also depends on the reachability of the products or services to the public. Nowadays companies are preferably doing some activities in order to gain publicity and with this, the relationship with customers also strengthens. Online websites, generally give offers and discounted prices to the customers so that they will feel satisfied and tend to shop again on the same website.

E-mail marketing and discounts were proved to be efficient marketing strategies in terms of generating sales. The purpose sending emails to targeted customers and promotional offers is to do remarketing with customers. This will eventually increase their brand awareness in the market. Sending blog posts, newsletters and personalized content are some of the ways used by online retailers to attract the customers. This will also help the companies to inactive customers active. By segmenting customers into different categories and marketing by means of emails, messages and discounts will increase the conversion rate. Another way used by retailers to retain customers, in the long run, is by rewarding loyal customers. Companies even send emails to those who added products in their cart list and did not proceed to payment.

Strategy Used To Analyze The Dataset

From processing of raw data to clean data, many steps have been involved. Data mining methodology was used to extract

the meaningful insights from the data. For that knowledge discovery in databases (KDD) methodology has been used. Different steps that includes inn KDD are data selection, pre-processing & transformation, applying data mining algorithms and then followed by interpretation or evaluation of the model. The data cleaning process was done by using python. Then by applying machine learning algorithms, the target variable which is revenue was predicted. To find the insights from the data, it was visualized using python and powerBI.. In figure.1, steps involved in KDD has been shown. By analyzing the data decisions can be taken accordingly which can drive the company into profits. From the visualizations, companies can check what factors are predominantly contributing to generate revenue. This model makes the businesses aware of elements that can crucial in yielding and also find out the aspects that are not contributing the revenue.

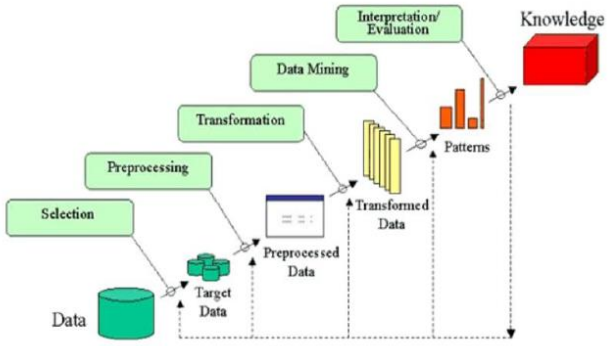

Fig. 1. Steps in KDD

V. PRELIMINARY VISUALIZATIONS

The visualizations provide a better understanding of the data. The growth and trend can be easily understood for evaluating the performance. Various factors involved in influencing the intention of online shoppers were visualized and discussed below.

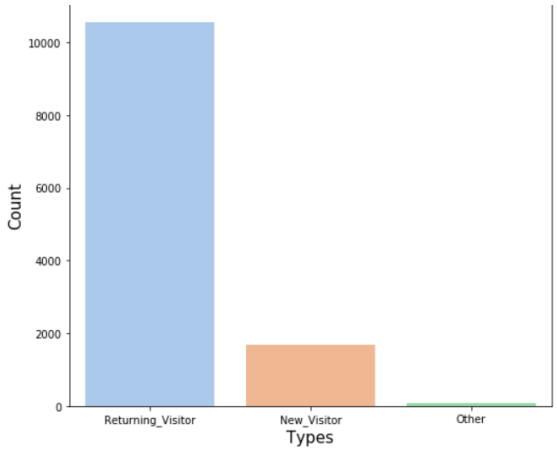

Fig. 2. Types of visitors

The bar chart in figure.2 shows the category of visitors to online websites. The count of new visitors is less than one-fourth of returning visitors or repeat customers. This shows that those who are aware of the website are visiting when compared to people who don't know much.

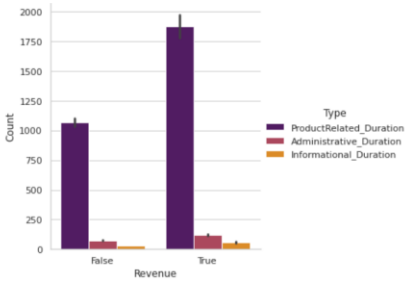

Fig. 3. Revenue by duration

The above chart denotes revenue earned form visitors based on different sections like information, administrative related and product-related. It is clearly understood that most of the companies revenue was earned on the basis of their products and services rather than for information and administrative related.

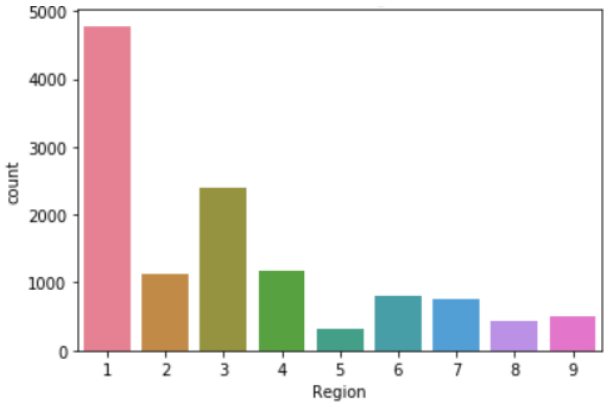

Fig. 4. Revenue generated based on regions

In figure 4, region 1 followed by region 3 produces most of the revenue when compared to other regions. Companies should prefer giving offers or discounts to the other regions to boost sales.

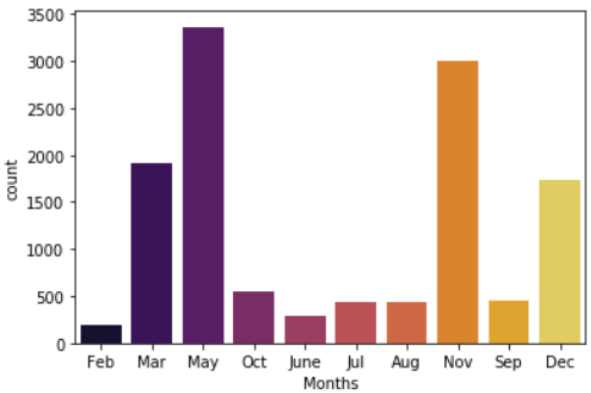

Fig. 5. Customers usage on different months

The above graph depicts the customers usage of online websites for shopping in different months. In the months of

May and November, a large number of customers did visit the website to make a purchase. Generally, there might be known festivals in those months that trigger customers to buy. Companies must also concentrate on other months to grab their attention by organizing season sales.

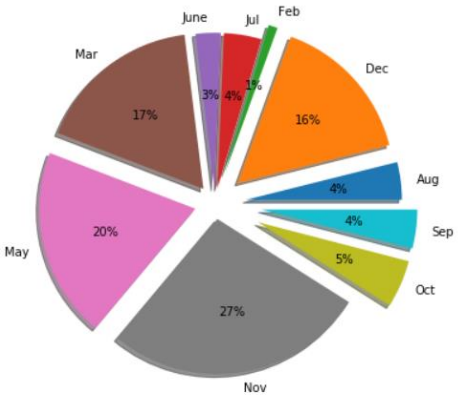

Fig. 6. Number of special days in different months

The pie chart in figure.6 signifies the number of special days with respect to months. This feature indicates that the customer is most likely to complete the transaction on website. Special days may include Mother's Day, birthdays, and so forth. In May, March and November months consist of many special days and the least contributing month is February. The months that contribute less number of special days should be focused by the companies and try to generate revenue from those months as well.

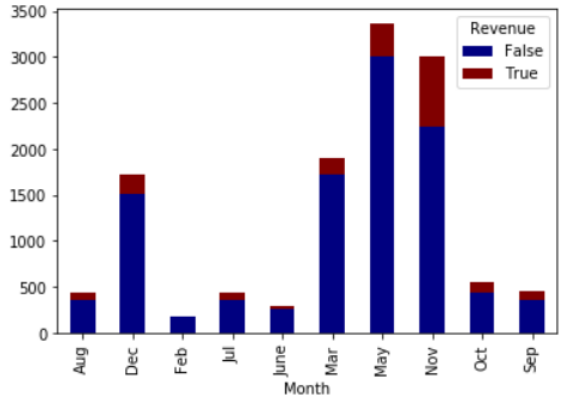

Fig. 7. Revenue generated for each month

In fig. 7, month-wise revenue generated was shown. Even though there are more visits to the website in the month of May but not many customers turned up in completing the purchase. Whereas in November, fewer people have visited the website with a higher conversion rate when compared to all other months. Like this there are many factors were companies look upon into data for analyzing and segmenting visitors to make profitable decisions.

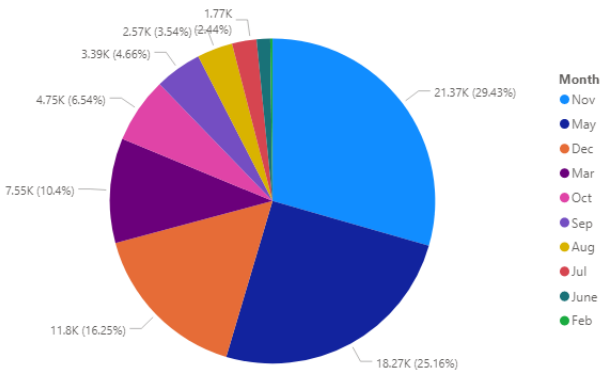

Fig. 8. Average page values in different months

Page value is one of the metrics used by Google to tell the page average value before visiting the goal page for completing the transaction. From figure.8, it was shown that in the months of May, November and December have higher page values which denotes that many visitors tend to purchase when compared to months.

VI. APPLICABLE TECHNIQUES

This model predicts whether a customer successfully purchases from the website or not. The main goal is to avoid abandoning of website by the visitors. Classification algorithms namely Random forest and Naïve bayes have been used by the author for prediction out of which Random forest gives the best performance [1]. Another research work was proposed by using deep learning that uses LSTM for predicting the effect of promotions from multiple online channels [2]. By using structural equation modelling proposed research work based on how effective the product information on shoppers purchase intention. It was proved that it has positive correlation by significantly influencing purchaser's intention. [8].

Techniques used for analysis of this dataset:

The aim of the project is to predict the shopper's intention in purchasing the product. The target column is revenue which is either True or false. From the analysis of different algorithms, this two machine learning algorithms namely Gradient boosting and Random forest were selected and going to be used for the analysis and evaluation. By using gradient boosting algorithm, accuracy can be boosted. After data preprocessing like removing improper values and correlation between variables was checked, if found highly correlated values have been removed. It was done so that independent variables are not affected. In this dataset, target column is biased towards one class that raises to class imbalance problem. Over sampling technique is one the resampling techniques used to get rid of class imbalance problem was used. Now the data is unbiased after overcoming class imbalance problem. Only the important features were selected by using recursive feature elimination technique. It

will rank the features based on how important it is in terms of contributing to target variable. This method is used to improve the accuracy and overall performance of the model.

VII. REFERENCES

[1] Baati, Karim, and Mouad Mohsil. "Real-Time Prediction of Online Shoppers' Purchasing Intention Using Random Forest." IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, Cham, 2020.

[2] Sakar, C, Polat, S.O., Katircioglu, M. et al. "Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks" Neural Comput & Applic 31, 6893–6908 (2019).

[3] Ling, C., Zhang, T., & Chen, Y. "Customer Purchase Intent Prediction Under Online Multi-Channel Promotion": A Feature-Combined Deep Learning Framework. IEEE Access, 7, 112963-112976.

[4] Y.Lima, S.Salahuddinc, A.Romle, S.Abdullahe." Factors Influencing Online Shopping Behavior: The Mediating Role of Purchase Intention" Procedia Economics and Finance, 2016.

[5] C. Changchit and R.J. Vokurka, "Factors that influencing Internet shopping behavior," Journal of Electronic Commerce, vol. 10, pp. 1- 15, 2004.

[6] Lee, R. J.Sener, I. N., & Handy, S. L. (2015). Picture of Online Shoppers: Specific Focus on Davis, California. Transportation Research Record, 2496(1).

[7] Malik, G., & Guptha, A. (2013). An Empirical Study on Behavioral Intent of Consumers in Online Shopping. Business Perspectives and Research.

[8] Sharma, R., Kaur, B., Rani, S., & Gupta, D. (2019). Recommender system: "Towards classification of human intentions in e-shopping using machine learning". Journal of Computational and Theoretical Nanoscience.