ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

Impact Of Covid19 On Online Shopping And Review on Some Machine Learning Algorithms

Ananthalakshmi K S
Department Of Computer Application
Saintgits College Of Applied Sciences
Pathamuttom, Kottayam
al.bca1922@saintgits.org

Jilu Elsa Jacob
Department Of Computer Application
Saintgits College Of Applied Sciences
Pathamuttom, Kottayam
jiluej.bca1922@saintgits.org

Asst.prof.Jisha J Nair
Department Of Computer Application
Saintgits College Of Applied Sciences
Pathamuttom, Kottayam
jisha.jn.bca1922@saintgits.org

Abstract— *The covid-19 crisis has brought with it several changes in our society of which the changes in shopping patterns is prominent. To be more precise, the pandemic created an era of online shopping. Traditional purchasing is now replaced with E-commerce .The wide variety of products offered , low expense ,time saving ,ease of transportation are all reasons contributing to this. At the same time there are many challenges and difficulties too, like the need to sustain customer satisfaction. Consumer opinions and transactions are monitored using machine learning techniques to get more information regarding customer satisfaction, interests, shopping behaviours and threats by E-commerce Businesses. But also, there are increasing reports of unfair, confusing, falsified business habits taking place online and many scams also take advantage of the crisis. So the information regarding the users and their interests need be secured and protected in order to avoid fraud and public crisis. The changes in online shopping patterns though situational, can have long lasting effects on entire industry. We have conducted a survey on 150-200 people over the state to analyze the changes in shopping behavior throughout the epidemic .In this paper we give our reviews on papers focusing on the influence of 2019 Novel corona virus (COVID-19) on E-commerce and the changes in the demand and buying behavior of customers. Also, the paper analysis the machine algorithms like J48, Random Forest and Random Tree to find the accurate one.*

Keywords— E-commerce, Novel Corona Virus, Data Privacy, j48 algorithm ,Random forest algorithm, Random tree algorithm.

I. INTRODUCTION

Corona virus Disease or COVID-19 is a rapidly spreading infectious disease, first originated in bats or pangolins. Its first human transmission was reported to be in WUHAN, CHINA and later spread across different parts of the world. Corona viruses are a group of viruses which has the ability to infect both humans and animals and cause sickness. Covid-19 was declared as a pandemic by the World Health Organization (WHO) on March 11th, 2020. The disease is most likely to spread through close contact of within around 6 feet with one having the infection. Fever, breathlessness, cough, sore throat, headache, muscle pains, chills, loss of taste and smell, etc are all the symptoms of this disease. Older adults and people with chronic health conditions are more severe to the illness.

The covid-19 crisis caused many losses to human life worldwide, bought many changes in life pattern and even put forward many challenges to public health, food systems and world of work. Covid crisis turned people's life upside down. Social distancing and lockdowns were introduced, educational as well as business institutions were closed, people shifted their meetups to online platforms, people started working and learning from home, worshipping also turned virtual, people started purchasing online more often, health was taken into serious consideration by many, and so many other changes did also come up. Different sectors were also affected differently. Expert volumes of various sectors including agriculture and fisheries were lessened and a lot more social and economic impacts were formed. Poverty was probably increased in many regions, unemployment rate rose up, tourism and hotel industry also got affected and moreover many people even lost their dear ones during the severe pandemic. This way the changes that followed the disease end up in a very long list. Among all the other effects of covid-19, the change in shopping patterns is very notable.

Covid-19 changed consumption trends globally and in many cases for the long-term. Months of lockdown and social distancing made consumers to think differently on shopping and spending as well. People were restricted to leaving the house only for necessary items and preferred shopping online more often. Face masks, hand-sanitizers and toilet papers were of high demand and so retailers were

forced to limit its number of purchases. Customers began to place a greater emphasis on cleanliness and health, and as a result, hand sanitizers, soaps, cleaning gels, and disinfectant products were among the most popular things purchased during the epidemic. Vitamins and supplements, Detol, Lysol are also in top positions. The pandemic pushed more shoppers online and so a huge rise was experienced in e-commerce. It was found through surveys that India saw greater changes in shopping habits than other countries. Consumption of fruits, vegetables, dairy products increased over junk foods. This way, now people could buy anything just with a click of a mouse. Consumer sentiments changed and so did the myths of shopping online. Consumers who earlier refused to buy online because of their fear in connectivity, inclusion and trust now started to believe in online shopping.

The biggest challenge ever faced by any entrepreneur is of winning and keeping customers. They adopt various steps to meet this and of which knowing their customers are one of the main steps they put forward. They do this through gaining an understanding of their clients' desires, wishes, and purchasing habits. They always try to know their customers even more. The most successful entrepreneurs know about the name, age, income, hobbies, taste, dislikes, and so much about their customers. They collect such information by tracking their customers using the technologies of artificial intelligence (AI) and machine learning (ML). The gathered big data allows retailers to have a better understanding of client shopping behavior and how to attract new customers. Big data analytics is also utilized to generate customer suggestions based on their purchasing history, resulting in more personalized shopping experiences and better customer care. But while people may enjoy shopping online for making cheap deals in the comfort of their home, their financial transactions and shopping accounts could be compromised by countless prying eyes. Many online scams and fraud are found taking advantage of the Covid crisis too. The location data, bank data, personal data are all misused by many. Still, legal regulations hardly protect people privacy against this misusing. Therefore ,this raises the need for a true secure environment for consumers to share their information or data without fear of scams. In this paper, we present the data we collected through the survey that we conducted to study the changes in shopping patterns and to show the frequency of scams. We also put forward our conclusion on three machine learning algorithms to find the better of the three that can be used to ensure data privacy.

II. LITERATURE REVIEW

Privacy is a term which is hard to define. The concept of privacy has a pivotal role in this modern era. It can be found on the works of Aristotle and John Locke too. But the time-honoured definitions of privacy will not be able to handle the challenges made by the technologies, since it is formulated before the arrival of internet. Now everything is going by online, including the job, shopping etc. In which the online shopping is the current growing trend that is found to have high existence in this pandemic period. And there is a rapid growth in the number of online shopping websites which have high market potential. The important thing which is taken to consideration is that the data and . Since it is online, chances are high for the loss of data protection and privacy. According to studies the privacy can have four dimensions and they are the collection of personal information, unauthorized secondary use of personal information, errors in personal information, and the improper access to personal information. It matters when these four things are not taken well. People used to highly bother about the privacy and security and reluctant to use and share personal information in past times, but when the trends had changed and the arrival of millions of online platforms enhanced the people began to be more likely towards online. And thus it became accepted as the part of life. Trust works here. And now during this pandemic it became the only way of communication, purchasing, working, etc. So more than acceptance the needs of people worked here. A numbers of issues are going on this topic even now. It is impossible to complete even a transaction without sharing your personal data. And that is why the data privacy plays significant role in e-commerce. So it is important to gain the trust of people while coming with a new online marketing platform.

Earlier the use of online shopping was comparatively very low. People were reluctant to purchase from online since they cannot choose it by seeing or touching the product. So the trust in online products were less according to recent studies. But now everything have been changed from top to bottom. As a result, the total concepts of people were changed. Number of online shopping customers increased rapidly. During the pandemic and all, online shopping has became the only way of purchasing for people. Studies shows people will continue to stick on to online shopping even after the epidemic since it provides vast options and facilities for the consumers.

Histories shows that the efficiency of machine learning algorithms were reasonably low earlier when compared to nowadays. The hardware devices used were not much developed and advanced in the past times. However the capability of hardware devices had a great role in deciding the efficiency of algorithms. Also the basic vision on the j48 algorithm, random forest algorithm, random tree algorithm was not much advanced. The collection of data sets, analysing the algorithms, etc were merely basic. As a result efficient results were relatively small.

III. METHODOLOGY AND MATERIALS

To know more about the pattern of shopping online during covid19, we conducted a survey on 150 to 200 people of which 51.9% were male and 48.1% were females. People of the age group 20-30 were found to be more to the use of online method for shopping than any other .97.4% people uses online platform for shopping purpose and 82.5% will still prefer shopping online even after the end of covid crises. The survey also shows an increase in online purchase by 74.7% during the pandemic. Also, fashion is the category

that is shopped the most. We could also find that 12.3% had experienced some kind of scam while purchasing online. 12.3% scam being reported from a survey of 200 people is not too small. There are a huge lot of cases of scam being reported all over the world during the covid19 pandemic era as almost everything changed to online mode. Scams related to purchasing of covid vaccines, online foods that cure covid, masks that can prevent covid were noted. Also financial theft and personal data loss was located many. In the last two years of the pandemic 37% of consumers have been victims of application fraud and 38% experienced account takeovers (Survey results conducted in December 2020 of 8653 US consumers of age 18 and above).
Online scammers pretend to be legitimate and use a fake website or a fake ad on a genuine retailer site. Many of these websites offer luxury items at low prices and scam over the consumer either in method of payment or privacy and protection of information they collected. So, we need to ensure the protection of our information and money.

Different algorithms where used to solve this issue. Algorithms like Logistic regression , Naïve bayse, random forest, j48, random tree ,and so were all applied under different conditions to bring up a solution, Here we review the solutions found using j48, Random forest and random tree algorithms to find the best of the three.

A. J48 Algorithm

J48 algorithm is an open source java implementation of the C4.5 decision tree algorithm. J48 algorithm recursively splits a data set to get all the possible predictions. This algorithm uses the greedy method approach to enhance decision trees for classification. J48 algorithm generates decision trees, and it is based on certain rules to classify. Its nodes analyze the existence and the significance of every unique feature. It also has some disadvantages like over fitting, insignificant or empty branches etc. It includes the following steps:

- Step1: First the leaf is labeled with a similar class if the instances belong to similar class.
- Step2: Then, the potential data is figured and its gain is taken from the test on attribute for all attributes.
- Step3: Finally, the best attribute will be selected based on the present selection parameter.

B. Random Forest Algorithm

Random forest algorithm is a supervised machine learning algorithm which is very flexible and also easy to use. It is one of the most popular algorithm. It can be used for classification as well as regression problems in machine learning. And ensemble of multiple decision trees is what makes it a forest. When there is an increase in the number of trees, there will be an increase in the accuracy too and so prevent over fitting. Random Forest algorithm is used in different sectors like banking, medicine, marketing, land use, etc. This algorithm too has many advantages and disadvantages. It enhances accuracy of the model, it can handle large data sets with great dimensionality, it is flexible and easy to use, it is simple and diverse and reduces risk of over fitting. But this algorithm is time consuming, more complex and may require more resources too. It includes the following steps:

- Step 1: Samples are selected at random from a given data set.
- Step 2: A decision tree is constructed for each selected sample which will then be used to give the predicted result.
- Step 3: Voting for every predicted result will be done.
- Step 4 :The result with more number of votes will be selected as the final prediction result.

C. Decision tree or random tree algorithm

Decision tree and Random forest algorithms are two major widely used decision algorithms. Decision tree algorithm is also a supervised machine learning algorithm which can be operated on both classification and regression algorithms. It is like a tree with nodes in it in which the number of criteria decides the branches. Data is split into branches in them and this tree also has root nodes, leaf nodes, children nodes and so on like any other. It has many advantages like it is easy, it is fast, it can handle large data and even numerical and categorical one. But it may over fit, or show more deflections too. Thus, adding to its disadvantages. Its working includes splitting, pruning, selection of trees, entropy, and knowledge gain.

IV. DATA COLLECTION

The data set collected through the survey of 150-200 people gave us different results from which we could predict the shopping patterns of the customers during the pandemic. Here we classify the data set into different attributes like gender, age, use of online platforms, first purchase , application, category, payment, scam experience, reason, post covid shopping, satisfied mode and frequency of purchase and we could also give a description to them(table 1).

TABLE I. ATTRIBUTES AND DESCRIPTION

Attribute	Description
Gender	Specifies the gender of the customer
Age	Specifies the age of the customer
Use of online platforms	Describes if the customer has used any online platforms for shopping
First purchase	Specifies if the customer made his first purchase during covid time
Application	Specifies the most frequent application used for shopping online
Category	Specifies the mostly preferred category for online shopping
Payment	Specifies the mode of payment
Scam experience	Specifies if the customer had any scam experiences
Reason	Specifies the reason for shopping online
Post covid shopping	Specifies if the customer will continue shopping online after covid crisis
Satisfied mode	Specifies the satisfied mode of shopping
Frequency	Specifies the frequency of purchase

V. OBSERVATION

The data that was collected through the survey was analysed as above. Our reviews on papers on the efficiency of the three different algorithms namely j48, random forest and random tree gave us the results as given in the tables.

TABLEII. ACCURACY OF ALORITHMS

Algorithm	Accuracy
J48	75.84%
Random forest	76.81%
Random tree	68.59%

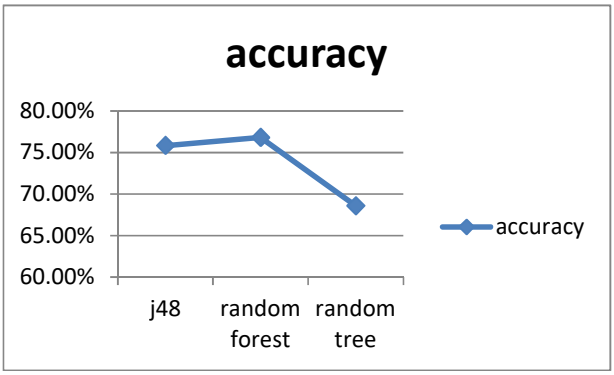

Fig.1 Figure showing accuracy of algorithms

From the papers we reviewed we could give an average accuracy value to the algorithms j48, random forest and random tree as given in the table. We ended up concluding that random forest algorithm was most efficient than others as it showed more accuracy value than other two algorithms. Therefore this algorithm proves to be most efficient one,

VI. FUTURE SCOPE

Recent studies shows that people are more likey to prefer online shopping even in the future. The study we conducted also give us similar results that 82.5% prefer shopping online in the post covid era. So the possibilities for increased scam are not small. Efficient technologies will get developed in the future to prevent scams and to enable better shopping experience online. In this information era, value of data is quite similar to the value of a human life. Hence providing more data protection schemes will enhance the trust of customers. Also it enables better prevention from the scams and fraudulent experiences.

VII. CONCLUSION

In this, paper we conducted a review study on the impact of covid19 on online shopping and the role of data protection and privacy during the pandemic. The covid crisis shifted people to E-commerce platforms which hold the risk of data privacy. For this we organised a survey among consumers of different age groups to study the behaviour of shopping patterns and scam. We could find 12.3% scams in our survey too which shows the traces of more scams occur all over the world while shopping online during the pandemic..The aim of the study was to find the behavioural change in online shopping during the covid 19 pandemic and efficient data protection algorithm among J48, Random forest and random tree for which we reviewed papers concentrating on the efficiency of these algorithms. As a result, we found that Random forest was the most efficient one as it gave more accuracy value. Therefore we conclude that Random forest is more efficient the than other two and can be applied as a solution for data privacy.

ACKNOWLEDGEMENT

We thank Kristu Jyoti College Of Management And Technology for organizing the International Conference. We also thank our HOD Prof. Ambily Merlin Kuruvila for taking initiative for us. We also like to show our sincere gratitude towards our Lecture-in-charge Asst. prof. Jisha J Nair for guiding us throughout the preparation. Above all we thank our God almighty for all his blessings.

REFERENCES

[1] J. Singh, "Real time BIG data analytic: Security concern and challenges with Machine Learning algorithm," *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, 2014, pp. 1-4, doi: 10.1109/CSIBIG.2014.7056985

[2] L. YAVUZ, A. SORAN, A. ÖNEN and S. M. MUYEEN, "Machine Learning Algorithms Against Hacking Attack and Detection Success Comparison," *2020 2nd International Conference on Smart Power & Internet Energy Systems (SPIES)*, 2020, pp. 258-262, doi: 10.1109/SPIES48661.2020.9243033.

[3] Voss, W. G. (2016). European union data privacy law reform: General data protection regulation, privacy shield, and the right to delisting. *The Business Lawyer*, *72*(1), 221-234.

[4] L. Tawalbeh, F. Muheidat, M. Tawalbeh, M. Quwaider and G. Saldamli, "Predicting and Preventing Cyber Attacks During COVID-19 Time Using Data Analysis and Proposed Secure IoT layered Model," *2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA)*, 2020, pp. 113-118, doi: 10.1109/MCNA50957.2020.9264301

[5] J. B. Prajapati and S. K. Patel, "Performance Comparison of Machine Learning Algorithms for Prediction of Students' Social Engagement," *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 947-951, doi: 10.1109/ICCMC51019.2021.9418260.

[6] J. Singh, "Real time BIG data analytic: Security concern and challenges with Machine Learning algorithm," *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, 2014, pp. 1-4, doi: 10.1109/CSIBIG.2014.7056985

[7] X. Chen, "Personal Privacy Protection in Big Data Environment Under the New Coronavirus Situation," *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, 2020, pp. 651-653, doi: 10.1109/ICISCAE51034.2020.9236848.

[8] C. Pandit, H. Kothari and C. Neuman, "Privacy in time of a pandemic," *2020 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges(51275)*, 2020, pp. 1-6, doi: 10.1109/CMI51275.2020.9322737.

[9] J. Hamm, A. C. Champion, G. Chen, M. Belkin and D. Xuan, "Crowd-ML: A Privacy-Preserving Learning Framework for a Crowd of Smart Devices," *2015 IEEE 35th International Conference on Distributed Computing Systems*, 2015, pp. 11-20, doi: 10.1109/ICDCS.2015.10.

[10] Y. Wang, R. Xu, M. Schwartz, D. Ghosh and X. Chen, "COVID-19 and Retail Grocery Management: Insights From a Broad-Based Consumer Survey," in *IEEE Engineering Management Review*, vol. 48, no. 3, pp. 202-211, 1 thirdquarter,Sept. 2020, doi: 10.1109/EMR.2020.3011054.