JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review of Machine Learning Techniques for Consumer Behavior Prediction Analysis

[1]Pooja Sharma, [2]Dr. Sachin Kumar

[1]Research Scholar, [2]Assistant Professor
Department of Computer Science and Engineering
Tula's Institute Dehradun, India

Abstract : Now a days, consumer behavior models are typically based on machine learning, data mining of customer data, and each model is designed to answer one question at one point in time. Predicting customer behavior is an uncertain and difficult task. Thus, developing customer behavior models requires the right technique and approach. Once a prediction model has been built, it is difficult to manipulate it for the purposes of the marketer, so as to determine exactly what marketing actions to take for each customer or group of customers. Despite the complexity of this formulation, most customer models are actually relatively simple. Because of this necessity, most customer behavior models ignore so many pertinent factors that the predictions they generate are generally not very reliable. This paper discusses various research works on consumer behavior analysis using difference machine learning, data mining techniques. The accuracy, error rate, precision is the key parameters and Python software can be used for implementation.

***IndexTerms* - Consumer, Machine Learning, Prediction, Accuracy, Error, Data Mining.**

I. Introduction

The development of Internet influenced many of our day-to-day activities. Ecommerce is one of the rapid growth areas in the Internet era. People are eager to buy products from online sites like Amazon, ebay, Flipkart etc. Online sites also provide facility for customers to write review on products they buy. These reviews help consumers and vendors for making decision on marketing strategies, and the improvement of products and services[1]. Nowadays people are very much interested to read reviews before purchasing any product and getting services. This makes areas for opinion spammers to write fake reviews to promote or to demote both products and business services. This type of activities is often referred as Review spam.

Through the studying of consumer behaviour some fundamental questions comes abroad such as:

- Why does consumer buy a product?
- How does consumer buy the product?
- How does consumes or use the product?
- How does consumer develop a product after buying it?
- How consumer exempted from the product (or his packing) after its usage?

Those questions find answers through the study of the factors that influences consumer's behaviour. Those factors are separated in four categories: social, cultural, demographical and psychological. The analysis of consumer behaviour is based on the assumption that consumers always base their decisions on a certain amount of information. This information may be divided into two categories: internal (previous experience) and external (type of product, word of mouth, etc.) According to this assumption, a company could not effectively market a product without a good understanding of the type of information consumers use to make purchasing decisions and the way in which the information is perceived and used - in other words, the decision-making processes [5].

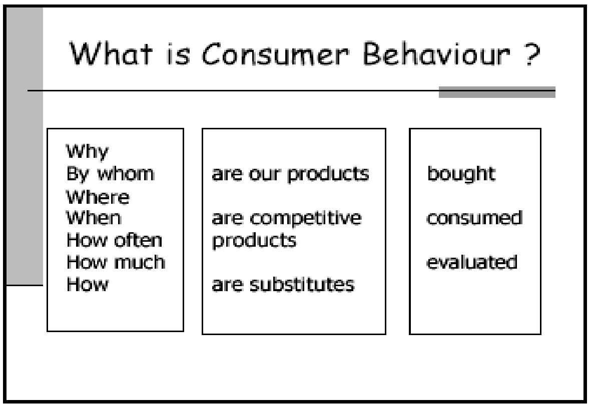

Figure 1: Consumer behaviour [Prof. Dr. Maggie Geuens, Consumer Behaviour, 1999]

The processes involved in making a decision are greatly influenced by three major types of variables: those related directly to consumers themselves; those related to the purchasing context or situation; and those concerning the products or services being considered. These three variables form the "basic triad." A large part of this chapter deals with the decision-making processes adopted by consumers and the many ways in which the information they are apt to use is actually processed.

II. BACKGROUND

V. Shrirame et al.,[1] Client produced content as surveys, appraisals, and remarks can be broke down for more prominent experiences for big business use. The examination of such buyer conduct is useful to comprehend the customer's prerequisites and foresee their future expectations towards the administration. Through this psychological examination, Online business Associations can follow the utilization and conclusions appended to their items and adopt suitable showcasing strategies to give a customized shopping experience to their purchasers, subsequently expanding their authoritative benefit. This work expects to utilize information driven showcasing instruments, for example, information representation, common language preparing, and AI models that help in understanding the socioeconomics of an association. We additionally fabricate recommender frameworks through communitarian sifting, neural organizations, and estimation examination.

E. Kafeza et al.,[2] The recognizable proof of web-based media networks has as of late been of significant worry, since clients taking an interest in such networks can add to viral advertising efforts. In this work, we center on clients' correspondence considering character as a critical trademark for recognizing open organizations i.e., networks with high data streams. We depict the Twitter Character based Open People group Extraction (T-PCCE) framework that distinguishes the most informative networks in a Twitter network chart thinking about clients' character. We at that point grow existing methodologies as a part of clients' character extraction by conglomerating information that speak to a few parts of client conduct utilizing AI strategies. We utilize a current particularity based network recognition calculation and we broaden it by embeddings a post-handling step that takes out chart edges dependent on clients' character.

M. A. Sharkh et al.,[3] The goal is to plan a Cloud application conduct forecast strategy dependent on AI indicators. Any enhancement for forecast exactness has direct effect on key execution markers for both Cloud suppliers and Cloud occupants/customers. Test results show the capability of our way to deal with improve Cloud asset planning for a Cloud server farm.

S. Shahriar et al.,[4] As the keen city applications are moving from calculated models to advancement stage, savvy transportation is one of brilliant urban communities applications and it is making strides these days. Electric Vehicles (EVs) are viewed as one of the significant mainstays of shrewd transportation applications. EVs are ever filling in fame because of their expected commitment in diminishing reliance on petroleum derivatives and ozone depleting substance outflows. Nonetheless, huge scope arrangement of EV charging stations represents different difficulties to the force matrix and public foundation. To conquer the issue of delayed charging time, the straightforward arrangement of sending additionally charging stations to increment charging limit doesn't work because of the strain on force matrices and actual space impediments. Accordingly, analysts have zeroed in on creating savvy booking calculations to deal with the interest for public charging utilizing displaying and streamlining. All the more as of late, there has been a developing interest in information driven methodologies in displaying EV charging.

J. Edmond Meku Fotso et al.,[5] This is one the principle reasons that lead to high dropout, low fulfillment and achievement rate saw in the MOOCs. Many examination work have recommended distinct, prescient and prescriptive models to address this issue, however the vast majority of these models center around foreseeing dropout, finishing as well as progress, and don't by and large give enough consideration to one of the key advance (student conduct), that precedes, and can clarify exiting and disappointment. Our examination intends to build up a profound learning model to anticipate student conduct (student connections) in the learning cycle, to prepare students and course teachers with knowledge comprehension of the student conduct in the learning cycle. We pick RNN and executed/tried the three principle models of RNN: Basic RNNs, GRU (Gated Intermittent Unit) RNNs and LSTM (Long momentary memory) RNNs. The models were prepared utilizing L2 Regularization, in view of the forecasts results, we startlingly discovered model with basic RNNs created the best execution and exactness on the dataset utilized than the other RNN designs. We had couple of perceptions, model: we saw a relationship between's video survey and test conduct and the cooperation of the student to the learning cycle.

P. A. Savenkov et al.,[6] This article examines the advancement of numerical help and programming for identifying irregular conduct of clients dependent on biometric qualities of their conduct examination. One of the difficulties in wise UBA (Client Conduct Examination) frameworks is securing of valuable data from an enormous volumes of unstructured, unparalleled information. Techniques and calculations of clever information handling and AI utilized in UBA/DSS frameworks help to take a shot at an undertaking of taking care of issues of information examination of various directivities. It is proposed a use of AI techniques in execution of versatile UBA framework. There was shaped the rundown of the main elements submitted to the contribution of the breaking down techniques during the examination. Two methodologies of recognizing strange client conduct have been proposed. The use of AI methods in clever UBA frameworks will make it conceivable to foresee data dangers and insider excfiltration of these associations ahead of time.

J. R. Goodall et al.,[7] In spite of the best endeavors of digital protection investigators, arranged figuring resources are regularly undermined, bringing about the deficiency of licensed innovation, the revelation of state privileged insights, and major monetary harms. Irregularity recognition strategies are useful for identifying new kinds of assaults and strange organization movement, yet such calculations can be hard to comprehend and trust. Organization administrators and digital examiners need quick and adaptable devices to help recognize dubious conduct that sidesteps robotized security frameworks, yet administrators don't need another mechanized apparatus with calculations they don't trust. Specialists need devices to increase their own area aptitude and to give a relevant comprehension of dubious conduct to help them decide. In this work we present Situ, a visual examination framework for finding dubious conduct in streaming organization information. Situ gives an adaptable arrangement that joins inconsistency identification with data perception.

D. Damkevala et al.,[8] This work supplies a course for utilizing the Watson AI Programming interface on IBM Cloud to do serverless information investigation utilizing AI as a help. Changing the enormous measure of information created by an association into insight should be possible utilizing progressed examination techniques, for example, utilizing an altered Mahalanobis Distance calculation for amalgamation of connection information under the domain of AI. Further refinement of connection information is finished utilizing a Multivariate Dependability Classifier model. The utilization of this high level investigation administration should be possible in a serverless way where the engineer just should be worried about how the information is broke down, i.e., scoring, cluster or stream models with a persistent learning framework without the expense of equipment whereupon to prepare those models.

Asniar et al.,[9] The advancement of the web has caused digitalization of information which opens up large information openings. Computerized information in enormous numbers leaves hints of what clients see, what they read, their inclusion and conduct, judgment, about their inclinations and inclinations to give a lot of information that can be dug for learning encounters. The huge information esteem lies in the consequences of investigation and forecasts or activities taken from the aftereffects of the examination and expectation. Prescient examination is information usage, factual calculations, and AI methods to recognize potential patterns, occasions, and practices later on dependent on chronicled information. This work attempts to propose prescient examination to anticipate client conduct by utilizing conduct informatics and investigation approach so more profound knowledge into client conduct can be acquired to help prescient examination to improve business dynamic.

F. D. Pereira et al.,[10] Numerous analysts have begun removing understudy conduct by cleaning information gathered from web conditions and utilizing it as highlights in AI (ML) models. Utilizing log information gathered from an online adjudicator, we have assembled a bunch of fruitful highlights associated with the understudy grade and applying them on a data set speaking to 486 CS1 understudies. We utilized this arrangement of highlights in ML pipelines which were improved, including a blend of a computerized approach with a developmental calculation and hyper parameter-tuning with irregular hunt. Subsequently, we accomplished a precision of 75.55%, utilizing information from just the initial fourteen days to anticipate the understudy last grades. We show how our pipeline beats cutting edge chip away at comparative situations.

M. A. Salitin et al.,[11] Associations are utilizing progressed security answers for ensure their data assets. Notwithstanding, even such high speculations, conventional security approaches neglected to ensure the organization structure against best in class assaults. New proactive ways to deal with security are on the ascent, for example, Client Element Conduct Investigation (UEBA). UEBA is a sort of online protection measures that utilizations AI, calculations, and factual examinations to distinguish constant organization assaults. This work means to evaluate the worth and accomplishment of utilizing conduct examination in making sure about the organization from not-before-seen assaults, for example, zero-day assaults. This work utilizes a methodical writing audit and self-administrated overview and meetings with accommodation inspecting of prominent organization clients and top security sellers. Review and meetings with different security specialists are used to check the self evident actuality adequacy of the arrangements dependent on conduct examination.

A. Bouhoute et al.,[12] The ongoing computerizations of vehicles, along with the improvement of sensor advances and vehicle specialized gadgets have changed the vehicles into rich wellsprings of data. The examination of information created constantly via vehicles can contribute incredibly in improving driving security and drivers comfort. Despite the fact that distinctive scientific arrangements have arisen as of late, there still exist some significant issues in driving wellbeing that we accept that were ineffectively tended to, just as assorted numerical approaches whose application in driving conduct examination is to be researched. In this work, we built up a philosophy to measure and examine vehicle produced information, with center around two investigation objectives: 1) programmed confirmation of drivers' conduct adjustment to traffic rules; and 2) perception and correlation of drivers' practices. The proposed approach is partitioned into three stages. From the outset, the reflection utilizing mathematical areas is utilized to decrease the size of the produced information.

III. CONSUMER INVOLVEMENT

All the consumer variables, consumer involvement is by far the most important. Even though researchers in this area have defined involvement in different ways over the years according to research trends popular at the time, the consensus is that the term may be understood as the feeling of importance or personal interest associated with the product in a given situation. Rothschild suggests the following definition: "Involvement is a state of motivation, arousal or interest. This state exists in a process. It is driven by current external variables (the situation; the product; the communications) and past internal variables (enduring; ego; central values) Its consequents are types of searching, processing and decision making."

Functional Risk

In terms of medical, pharmaceutical or any health related products, functional risk has the most impact on consumer behaviour. This type of risk may be defined as the possibility that the product does not meet the consumer's expectations. This risk is common in the service and health sectors, which usually do not allow consumers to test the product before buying. A consumer can, however, reduce functional risk dramatically by seeking as much information as possible on the service or drug to be bought. Pharmacist's opinion, advertising (which often reports clinical studies), or friends' opinions may also reduce functional risk. Another way to reduce functional risk is to go for "safe bets" or "sure things".

Economic Risk

This risk is the easiest to understand: the more expensive the product or the service, the more complicated the decision-making process. This relationship may be greatly attenuated by the consumer's income level. Together with functional risk, economic risk explains, at least partially, why some consumers prefer to subcontract their decision-making processes, even for OTC products, to professionals.

Psychological Risk

Psychological risk is frequently experienced in the consumption of medical products or prescription drugs. It may be defined as the risk related to the purchase or consumption of a product that does not correspond to the consumer's desired self-image. Perhaps a consumer is afraid to confront latent inner feelings and elects to not follow a prescription. Another consumer who feels physically inadequate may prefer not to purchase an orthopedic aid. Like other forms of risk, psychological risk increases the complexity of the consumer's decision-making processes. Like for other forms of risk, a professional advice is needed but not always sought.

Social Risk

Psychological risk is related to the individual consumer's self-image; whereas social risk is related to the image others have of the individual. Naturally, this risk is not present for all consumers. In fact, social risk is present only in cases in which the form of consumption is visible or the consumers are sensitive to their environment.

IV. SITUATIONAL VARIABLES

The decision-making processes, along with the related information processing strategies, are influenced by certain situational variables. The main situational variables are the period (month, day, and season) when the purchase is made, the time available to the consumer to shop for the purchase, the presence or absence or reference groups, the economic climate, and the place where the decision is made.

A. Period

The period during which a purchase is made influences the decision-making process. A snowfall in early December, for instance, encourages consumers to do Christmas shopping. Tchaikovsky's *The Nutcracker* may be a holiday season favourite, but would it be sold out or held over in July?

Time Available

The amount of time a consumer has to make a decision also influences the decision-making process adopted. If there is little time, the consumer will rely more on subordinate processes and processes based on past experience.

Reference Groups

The presence or absence of reference groups also influences the decision-making process. If a consumer is aware of signals in his or her environment and must make a decision, the presence of a reference group or person of influence will increase the tendency to use a subordinate process.

Economic Climate

The economic climate also plays an important role. If the consumer is living through a recession or is keenly aware of the economic situation, he or she will rend to use a cognitive decision-making process in which price becomes more significant.

Place

The physical environment is another element influencing the consumer's choice of a decision-making process. This last factor is especially important, since the presence or absence of affective or cognitive stimuli would determine the process used.

V. PROPOSED STRATEGY

- Load the Amazon Review Dataset from the Kaggle

In this step, the consumer review dataset will be downloaded from kaggle source. It is a large dataset providing company. Then load this dataset into the python environment.

- Visualizing the Dataset

Now open the dataset files and view the various data in term of features like product name, quantity, review, purchasing time, number of visit, add to cart etc.

- Pre-process the Dataset

Now the data preprocess step applied, here data is finalize for processing. Missing data is either removal or replace form constant one or zero in this step.

- Splitting the Dataset into training and testing

In this step, the final preprocessed of dataset is divided into the training and the testing dataset. In the machine learning, firstly the machine is trained through given dataset then it comes in tested period for remaining dataset.

- Classification Using Machine Learning Algorithm

Now apply the machine learning technique to find the performance parameters. The existing work applied several techniques and find Naïve Bayes is better method then others. In proposed method, we apply the logistic regression method and optimize the better results than other approach; According to the researchers the logistic regression method is good for optimization to enhance the accuracy.

- Performance Metrics (Accuracy,Precision,Recall,F1 - Score)

Now the performance parameters are calculated in terms of precision, recall, f-1 measure, accuracy etc by using the following formulas-

True Positive (TP): predicted true and event are positive.
True Negative (TN): Predicted true and event are negative.
False Positive (FP): predicted false and event are positive.
False Negative (FN): Predicted false and event are negative.

$$\begin{aligned} Precision &= \frac{|TP|}{|TP| + |FP|} \\ Recall &= \frac{|TP|}{|TP| + |FN|} \\ F1 &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \\ Accuracy &= \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \end{aligned}$$

VI. CONCLUSION

There are various types of consumer reviews available in the internet that increasingly affects businesses and customers. Hence it is important to detect and eliminate such fake reviews from online websites. Machine learning techniques are suitable to predict and analysis of various problems. This paper reveals several approaches used for consumer review performance measures are identified. This topic needs further research in Big Data approach to reduce the number of features and computational complexity which helps to improve the detection methods, and also consider other kinds of media such as forums, blogs etc. Still it needs to be exposed yet in this regard. Prediction model is capable to identified and review the online data of consumer reviews. Therefore need to implement and analysis of consumer review model based on machine learning. Further, implement the machine learning based methods and optimize the improved results.

REFERENCES

1. V. Shrirame, J. Sabade, H. Soneta and M. Vijayalakshmi, "Consumer Behavior Analytics using Machine Learning Algorithms," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-6, doi: 10.1109/CONECCT50063.2020.9198562.
2. E. Kafeza, A. Kanavos, C. Makris, G. Pispirigos and P. Vikatos, "T-PCCE: Twitter Personality based Communicative Communities Extraction System for Big Data," in IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 8, pp. 1625-1638, 1 Aug. 2020, doi: 10.1109/TKDE.2019.2906197.
3. M. A. Sharkh, Y. Xu and E. Leyder, "CloudMach: Cloud Computing Application Performance Improvement through Machine Learning," 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada, 2020, pp. 1-6, doi: 10.1109/CCECE47787.2020.9255686.
4. S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou and M. Nijim, "Machine Learning Approaches for EV Charging Behavior: A Review," in IEEE Access, vol. 8, pp. 168980-168993, 2020, doi: 10.1109/ACCESS.2020.3023388.
5. J. Edmond Meku Fotso, B. Batchakui, R. Nkambou and G. Okereke, "Algorithms for the Development of Deep Learning Models for Classification and Prediction of Behaviour in MOOCS.," 2020 IEEE Learning With MOOCS (LWMOOCS), Antigua Guatemala, Guatemala, 2020, pp. 180-184, doi: 10.1109/LWMOOCS50143.2020.9234363.
6. P. A. Savenkov and A. N. Ivutin, "Organizations Data Integrity Providing through Employee Behavioral Analysis Algorithms," 2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2020, pp. 1-3, doi: 10.1109/MECO49872.2020.9134271.
7. J. R. Goodall et al., "Situ: Identifying and Explaining Suspicious Behavior in Networks," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 204-214, Jan. 2019, doi: 10.1109/TVCG.2018.2865029.
8. D. Damkevala, R. Lunavara, M. Kosamkar and S. Jayachandran, "Behavior Analysis using Serverless Machine Learning," 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2019, pp. 1068-1072.
9. Asniar and K. Surendro, "Predictive Analytics for Predicting Customer Behavior," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, 2019, pp. 230-233, doi: 10.1109/ICAIIT.2019.8834571.

10. F. D. Pereira, E. H. T. Oliveira, D. Fernandes and A. Cristea, "Early Performance Prediction for CS1 Course Students using a Combination of Machine Learning and an Evolutionary Algorithm," 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), Macei??, Brazil, 2019, pp. 183-184, doi: 10.1109/ICALT.2019.00066.
11. M. A. Salitin and A. H. Zolait, "The role of User Entity Behavior Analytics to detect network attacks in real time," 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakhier, Bahrain, 2018, pp. 1-5, doi: 10.1109/3ICT.2018.8855782.
12. A. Bouhoute, R. Oucheikh, K. Boubouh and I. Berrada, "Advanced Driving Behavior Analytics for an Improved Safety Assessment and Driver Fingerprinting," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 6, pp. 2171-2184, June 2019, doi: 10.1109/TITS.2018.2864637.