

e-ISSN:2582-7219



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 1, January 2024



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.54



6381 907 438



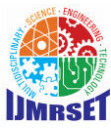
6381 907 438



ijmrset@gmail.com



www.ijmrset.com



Customer Behavior Analysis in E-Commerce using Machine Learning Approach: A Survey Applications

¹Ashutosh Sharma, ²Prof. Amit Thakur

¹Research Scholar, Department of Computer Science Engineering, Swami Vivekanand College of Science and Technology, Bhopal, India

²Assistant Professor, Department of Computer Science Engineering, Swami Vivekanand College of Science and Technology, Bhopal, India

ABSTRACT: These days, consumer behaviour models are often founded on machine learning and the data mining of customer data. Forecasting client behaviour is an unclear and tough endeavour. Thus, one has to use the appropriate strategy and strategy while constructing consumer behaviour models. Once a prediction model has been built, it is difficult to manipulate it for the purposes of the marketer in order to determine exactly what marketing actions to take for each customer or group of customers. This is because once the model has been built, it is impossible to change the variables that make up the model. While this formulation is very complicated, the majority of customer models are, in practise, rather straightforward. Because of this requirement, most customer behaviour models neglect so many essential elements that the forecasts they provide are often not particularly trustworthy. The purpose of this study is to present different research work that has been done on the analysis of consumer behaviour using various machine learning and data mining approaches.

KEYWORDS: Customer, Machine Learning, Prediction, Accuracy, Error, Data Mining.

I. INTRODUCTION

The advent of the Internet has had a significant impact on many aspects of our day-to-day lives. In today's internet-driven world, one of the most dynamic development sectors is electronic commerce. Consumers are eager to purchase goods from online retailers such as Amazon, eBay, and Flipkart, amongst others. Online sites additionally give ability for consumers to post review on things they purchase. Customers and sellers alike may benefit from the decisions made based on these assessments about marketing techniques, as well as the enhancement of goods and services [1]. Today consumers are very much eager to read reviews before buying any goods and obtaining services. This opens the door for opinion spammers, who will submit bogus evaluations in an attempt to boost or lower the reputation of businesses, goods, and services. These sort of actions is sometimes described as Review spam.

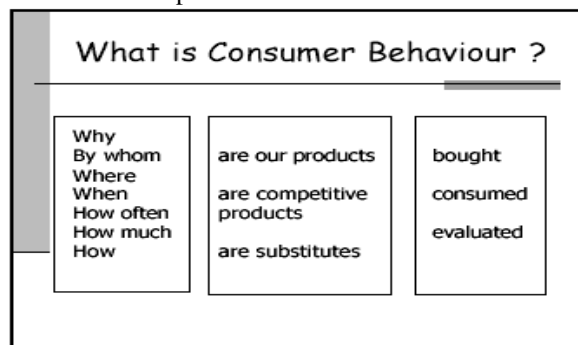


Figure 1: Customer behaviour [Prof. Dr. Maggie Geuens, Customer Behaviour, 1999]



The investigation of the elements that have an effect on the actions taken by customers yields the answers to these questions. These influences may be broken down into four distinct categories: the social, the cultural, the demographical, and the psychological.

The study of customer behaviour begins with the assumption that consumers always base their choices on a certain quantity of information before making a choice. This knowledge may be broken down into two categories: internal (gained through one's own prior experiences), and external (type of product, word of mouth, etc.) According to this presumption, a business would be unable to successfully market a product if it did not have a solid understanding of the kind of information customers rely on when making purchasing decisions, as well as the manner in which customers perceive and make use of that information—in other words, the decision making processes [5].

The processes that go into making a decision are heavily influenced by three major categories of variables: those that are directly related to the customers themselves, those that are related to the context or situation in which the purchase is being made, and those that are related to the products or services that are being considered. The "fundamental trio" consists of these three different factors. This chapter devotes a significant amount of its content to discussing the decision-making processes that are used by consumers as well as the several methods in which the information that customers are likely to employ is actually processed.

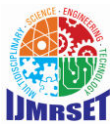
II. LITERATURE REVIEW

E. Manohar et al.,[1] The client-generated information such as surveys, evaluations, and notes may be broken out for more prominent experiences that can be used by large businesses. It is helpful to understand the requirements of the client and to anticipate their future expectations towards the administration by conducting an investigation of the buyer's behaviour such as this. With the help of this psychological research, Online Business Associations are able to monitor the applications and outcomes associated with their products, as well as adopt appropriate marketing strategies for the purpose of providing a personalised shopping experience for their customers, which ultimately results in an increase in their authoritative benefit. This study intends to leverage information driven marketing tools such as information representation, common language training, and AI models that aid in understanding the socioeconomics of an organisation. These tools will be used in this article. In addition to this, we develop recommender systems by using communitarian sorting, neural organisations, and estimate analysis.

V. Shrirame et al.,[2] Recent events have shown that the unmistakable proof of web-based media networks is a big cause for concern. This is due to the fact that customers who take an interest in such networks might contribute to viral advertising efforts. We focus on clients' communication in our study, taking their nature into consideration as an important identifier for open organisations, which we define as networks with a large volume of data streams. We illustrate the Twitter Character based Open People group Extraction (T-PCCE) framework by depicting it in a Twitter network chart. This framework identifies the most informative networks by taking into consideration the clients' personalities. After this, we expand upon previously established approaches as a part of the client character extraction process by combining data from a number of different aspects of client behaviour and using AI techniques. We make use of an existing particularity-based network recognition computation, which we then expand by including a post-handling phase that removes chart edges depending on the nature of the customers.

B. Lebichot et al.,[3] The objective is to devise a system for doing forecasts inside a Cloud application that is based on AI indicators. Any improvement made to the accuracy of the projection will have a direct impact on the key execution markers for cloud providers as well as cloud occupiers and consumers. The results of the tests demonstrate that our method is capable of dealing with and improving cloud asset planning for a cloud server farm.

S. Shahriar et al.,[4] One of the brilliant applications for urban communities that is making gains these days is smart transportation, which is one of the sharp city applications that is transitioning from the stage of calculated models to the stage of progress. Electric vehicles, sometimes known as EVs, are increasingly being recognised as a crucial component of innovative mobility solutions. Electric vehicles are gaining popularity rapidly because of the predicted contribution they will make towards reducing dependency on ozone-depleting substances and petroleum derivatives. Despite this, the widespread distribution of EV charging stations creates a number of challenges to the power matrix and the public



foundation. The basic plan of sending extra charging stations to increase charging capacity does not work as a solution to the problem of delayed charging time. This is due to the pressure placed on force matrices as well as the real space constraints that must be overcome. Similarly, experts have focused in on generating intelligent booking calculations to cope with the urge for public charging employing displaying and streamlining. All the more as of late, there has been an increasing interest in information driven techniques in showing EV charging. In light of this, researchers are attempting to identify a shopper charging personal behaviour standard that may supply pieces of information and the capacity for prescient evaluation. The cause for this article is to supply an expanded audit to the utilisation of directed and solo AI just as Profound Neural Organizations for charging conduct evaluation and expectation. In addition, both suggestions and potential directions for further research are discussed.

J. Edmond Meku Fotso et al.,[5] This is one of the primary factors that contributes to the high dropout rate, as well as the poor rate of contentment and accomplishment seen in MOOCs. The majority of these models focus on predicting dropout, finishing, and progress; however, the vast majority of these models don't generally give enough consideration to one of the key advances (student conduct), that precedes, and can explain exiting and disappointment. Many research works have recommended distinct, prescient, and prescriptive models to address this issue. Our study aims to construct a comprehensive learning model that can predict student behaviour (student connections) in the learning cycle. This model will be used to provide students and course instructors with information and understanding of student behaviour in the learning cycle. We choose RNN and executed/tried the three fundamental models of RNN: Basic RNNs, GRU (Gated Intermittent Unit) RNNs and LSTM (Long momentary memory) RNNs. The models were generated using L2 Regularization, and based on the results of the predictions, we were surprised to find that a model with simple RNNs had the greatest performance and accuracy on the dataset that was used as compared to the other RNN designs. We came to a few conclusions, one of which was that there is a connection between the video survey and test administration, as well as the level of participation of the student in the learning cycle.

P. A. Savenkov et al.,[6] This article analyses the development of numerical help and programming for recognising irregular behaviour of customers depending on biometric characteristics of their conduct evaluation. While developing sensible UBA (Client Conduct Examination) frameworks, one of the challenges that might arise is the challenge of safeguarding important data from vast amounts of unstructured, unmatched information. Methods and computations of intelligent data handling and artificial intelligence (AI) that are used in UBA/DSS frameworks aid to take on an endeavour of taking care of concerns of information analysis of varied directivities. The use of AI methods in the operation of a flexible UBA framework is an idea that has been offered. Throughout the evaluation, the examination led to the formation of the rundown of the primary components that was subjected to the participation of the breaking down approaches. There have been two different methods described for spotting unusual behaviour shown by customers. When used in intelligent UBA systems, the application of AI approaches will make it possible to anticipate data risks and to detect and prevent the theft of sensitive information by insiders at these organisations in advance.

J. R. Goodall et al.,[7] In spite of the finest efforts of digital protection investigators, organised figuring resources are routinely undermined, bringing about the lack of permitted innovation, the exposure of state protected insights, and big monetary damages. Irregularity identification algorithms are helpful for spotting new types of attacks and odd organisation movement. Nevertheless, such computations might be difficult to grasp and it may be difficult to know whether or not to believe the results. Administrators of organisations and digital auditors require rapid and flexible tools to assist in recognising suspicious behaviour that circumvents automated security systems. Nevertheless, administrators do not need another automated gadget with calculations that they do not trust. The specialists require tools that will help them improve their own area aptitude as well as provide them a relevant understanding of questionable behaviour that will assist them in making decisions. Situ is a visual assessment tool that we describe in this research for spotting questionable behaviour in streaming company information. Situ provides a versatile setup that combines the discovery of inconsistencies with the interpretation of data.

D. Damkevala et al.,[8] This article provides a tutorial on how to use the Watson AI Programming interface on IBM Cloud to do serverless data analysis with the assistance of AI. Using advanced analysis procedures, such as applying an altered Mahalanobis Distance computation for amalgamation of connection information within the realm of artificial intelligence, it ought to be able to transform the massive amount of data produced by an organisation into insight. Using a Multivariate



Dependability Classifier model, more refining of connection information is completed. The use of this high level investigation service should be possible in a serverless manner, where the engineer is only responsible for worrying about how the data is broken down, such as scoring, cluster, or stream models with a persistent learning framework, without the expense of equipment whereupon those models can be prepared. The purpose of this study is to investigate the use of serverless computer-based intelligence frameworks in the context of client behaviour analysis across varying socioeconomic conditions.

Asniar et al.,[9] The proliferation of the internet has led to the digitization of information, which has resulted in the development of several new information channels. Computerized information in large numbers leaves hints of what customers see, what they read, their inclusion and behaviour, judgement, and about their preferences and inclinations to give a lot of information that can be dug up for learning experiences. This information can be mined for learning experiences. The tremendous value of the information may be found in the findings of the research as well as any predictions or actions that are derived from the findings of the research. Using data, doing calculations based on facts, and employing AI techniques, prescient analysis is the process of identifying prospective patterns, events, and practises in the future based on information that has been documented. This paper makes an effort to propose that prescient examination should be able to anticipate client behaviour by utilising conduct informatics and an investigation approach. The goal of this is to enable prescient examination to obtain a more profound understanding of client behaviour, which would help prescient examination improve business dynamic.

F. D. Pereira et al.,[10] A large number of researchers have started eliminating student behaviour by cleaning information acquired from online conditions and using it as highlights in artificial intelligence (ML) models. This process was previously described. We have compiled a number of useful features connected with the student grade by using log information that was acquired from an online adjudicator. These features were then applied to a data collection that included information about 486 students who were enrolled in CS1 courses. We made use of this arrangement of features in enhanced machine learning pipelines, which included a combination of a computerised technique with a developmental computation and hyperparameter-tuning with irregular search. After that, we were able to predict the students' final grades with an accuracy of 75.55 percent by using data from the first fourteen days of the semester. We demonstrate how our pipeline is superior than cutting edge chip away by comparing the two in various scenarios.

M. A. Salitin et al.,[11] Organizations are increasingly turning to more advanced security solutions in order to protect their data assets. In spite of such high stakes, traditional security measures still failed to protect the organisational structure from the most sophisticated attacks. There is a rise in the use of novel preventative approaches to dealing with security issues, such as Client Element Behavior Investigation (UEBA). UEBA is a kind of internet security mechanism that uses artificial intelligence (AI), computations, and factual inspections to determine whether or not a business is being attacked continuously. The purpose of this article is to assess the usefulness and success of applying conduct examination in protecting an organisation against attacks that have not been observed previously, such as zero-day attacks. Specifically, this evaluation will focus on zero-day attacks. This article makes use of a systematic writing audit, as well as a self-administrated overview and meetings with accommodation inspections of major organisation customers and leading security sellers. Evaluation and discussions with a variety of security experts are used in order to verify the self obvious reality sufficiency of the arrangements that are reliant on conduct assessment. Analysts will go for a planned conference with merchants who are supplying solutions in order to understand the exhibition of conduct investigation based arrangements and the clear highlights of their responses while they are in the process of collecting the necessary information via an overview.

A. Bouhoute et al.,[12] The continuing computerizations of automobiles, along with the advancement of sensor innovations and vehicle specialist gadgets have converted the vehicles into rich wellsprings of data. The analysis of data that is continuously produced by cars has the potential to make significant contributions to the improvement of driving safety and the comfort of drivers. There are still a few significant problems in driving safety that we acknowledge were not effectively attended to, just as there are a variety of numerical approaches whose application in driving conduct examination is to be researched. This is despite the fact that distinct scientific arrangements have emerged in recent times. In this article, we built a philosophy to measure and examine information produced by vehicles, with the centre of our focus being on two investigation objectives: 1) programmed confirmation of drivers' behaviour adjustment to traffic rules; and 2) perception



and correlation of drivers' practises. Both of these objectives are centred around the vehicle. The technique that has been suggested may be broken down into three distinct parts. From the very beginning, the reflection that makes use of mathematical regions is used in order to reduce the total amount of information that is created.

III. PROPOSED STRATEGY

- Load the Amazon Review Dataset from the Kaggle

In this step, the customer 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, decision tree 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.

$$Precision = \frac{|TP|}{|TP| + |FP|}$$

$$Recall = \frac{|TP|}{|TP| + |FN|}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}$$

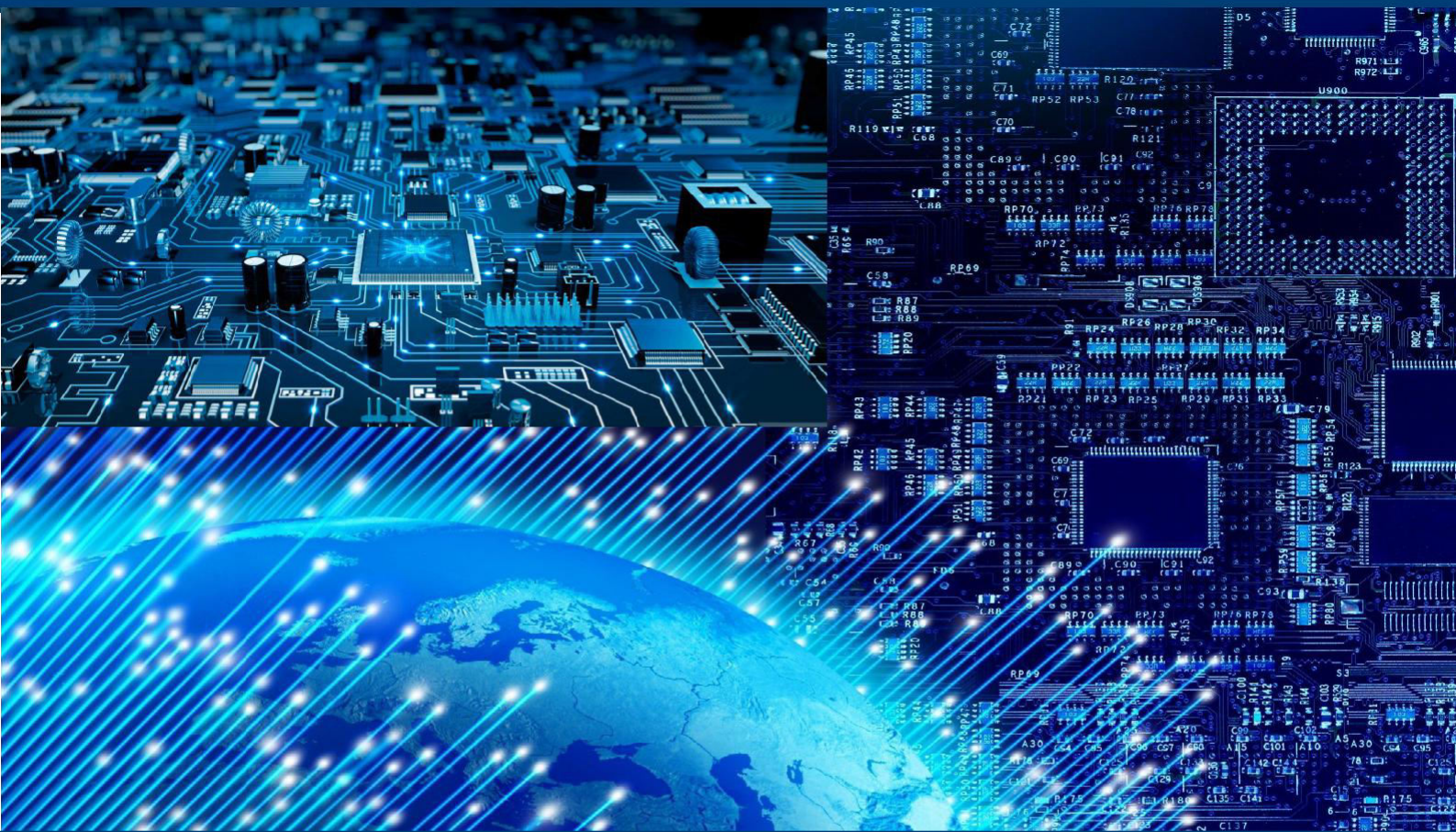
IV. CONCLUSION

There are various types of customer 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 customer 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 customer reviews. Therefore need to implement and analysis of customer review model based on machine learning. Further, implement the machine learning based methods and optimize the improved results.



REFERENCES

1. E. Manohar, P. Jenifer, M. S. Nisha and B. Benita, "A Collective Data Mining Approach to Predict Customer Behaviour," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1310-1316, doi: 10.1109/ICICV50876.2021.9388558.
2. 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.
3. B. Lebichot, T. Verhelst, Y. -A. Le Borgne, L. He-Guelton, F. Oblé and G. Bontempi, "Transfer Learning Strategies for Credit Card Fraud Detection," in IEEE Access, vol. 9, pp. 114754-114766, 2021, doi: 10.1109/ACCESS.2021.3104472.
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 (ICAIT), Yogyakarta, Indonesia, 2019, pp. 230-233, doi: 10.1109/ICAIT.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.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor
7.54

ISSN

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com