International Journal of Computing and Artificial Intelligence
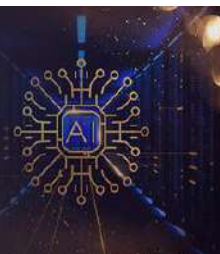
Siddhant Sharma
AKS University, Satna,
Madhya Pradesh, India

Akhilesh A Waoo
AKS University, Satna,
Madhya Pradesh, India

An efficient machine learning technique for prediction of consumer behaviour with high accuracy

Siddhant Sharma and Akhilesh A Waoo

DOI: https://doi.org/10.33545/27076571.2023.v4.i1a.59

Abstract
Customers look at ratings and feedback from other customers before deciding whether or not to buy a product. This content may take the form of favorable or negative reviews written by customers who have lately used the product in question. The Machine Learning Calculation may be able to assist us in both the visual depiction of the information as well as the factorization of the information. This investigation of customer behavior uses the Naive Bayes and Logistic Regression methods, both of which are presented in this work. The tactics based on logistic regression performed preferable than those based on other approaches. The contemporary problems are dissected, and then the contemporary solutions to those problems are presented and discussed. Afterwards, the results of the test indicate that the recommended technique has a greater accuracy as well as a higher recall and F1 score. The technique ends up being successful, with a high degree of precision on the comments. Python Spider 3.7 is the programmer that is used in order to carry out both the replication and the analysis.

Keywords: Naïve Bayes, logistic regression, precision, recall, F1 score, spider

Introduction

Machine learning (ML) is the scientific research of computations and factual models that computer systems use to do a certain activity without employing explicit directions, based on instances and derivation all things considered. It is a subfield of the field of artificial intelligence (AI). It is considered a subtype of the more general mechanised thinking. Calculations that are performed as part of machine learning generate a mathematical model that is reliant on example data. This model is referred to as "planning information," and it is used to make predictions or judgements without being explicitly modified to carry out the job [1]. Machine learning calculations are used in a broad variety of applications, such as email sorting and computer vision, in situations when it is difficult or impossible to build a traditional count for enough people carrying out the activity.

The concept of machine learning is inextricably linked to that of computational intelligence, which centres on the use of personal computers to provide projections. The study of scientific smoothing out contributes to the area of machine learning by bringing new techniques, theoretical frameworks, and application spaces to the table. Information mining is a subfield of the larger science of machine learning that focuses on the unassisted and exploratory processing of information [3, 4]. Machine learning, in the context of its application across a variety of business concerns, may also be seen as intuitive exploration.

A couple of the learning computations concentrate on developing more accurate representations of the data sources that were presented during training.
[11] Models that are examples of excellence combine head segment analysis with group analysis. Feature learning algorithms, which are also known as portrayal learning algorithms, frequently attempt to preserve the information in their information while also transforming it to such an extent that makes it useful. This is typically done as a pre-handling task prior to the operation of performing classification or expectations. This approach makes it possible to recreate the data sources by beginning with the unknown information providing dispersion. This is accomplished without giving priority to configurations that are very improbable given the conveyance in question. This eliminates the need for human component planning and makes it possible for a machine to familiarise itself with the characteristics as well as make use of them to carry out a particular activity.

Corresponding Author:
Akhilesh A Waoo
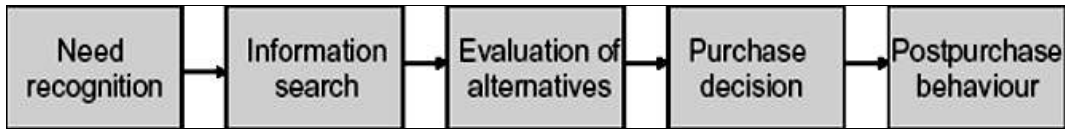AKS University, Satna,
Madhya Pradesh, India

Fig 1: Five-Stage Model of the buying process

According to this approach, customers should go through all five stages of the buying process before making a purchase. It's possible that this will be the case with high-including purchase. Customers may choose to bypass or circumvent some of these phases when making low-inclusion purchases. This model illustrates the complete set of considerations that come into play whenever a customer is faced with the prospect of making a significantly life-changing purchase.

The contentment of customers may be attributed to these three overarching components. Existing definitions allow for its identification as follows:

1. Satisfying the needs of customers is a reaction (Emotional or Cognitive)
2. The reaction is relevant to a specific area of concentration (expectations, product, consumption experience, etc.)
3. The reaction takes place at a certain moment in time (after consumption, after choice, based on accumulated experience, etc.)

The stage before any negotiations are made, during which there are assumptions made about the item, including the profits, the cost, and the availability of the item. The point in the business transaction when the customer rates the environment, the product, the kind of administration being provided, the delivery, the quality, and the change in the market. The stage that occurs after a transaction has been completed and during which the customer awaits help or guidance, the delivery of aggregate or a substitute item, and any fixes or cycles of charges.

Jain, Jay Kumar *et al*., combine ML and cyber security to talk about two distinct notions. They also talk about the benefits, problems, and difficulties of combining ML and cyber security. In addition, they explore several attacks and present a thorough analysis of various tactics in two different categories.

Jain, Jay Kumar *et al*., has discussed about the intrusion detection system, often known as IDS, is a piece of equipment or a piece of software that monitors a network or collection of devices in order to search for indications of possible intrusion. The frequency of cyber assaults has grown in recent years, and with it, the damage they do to society. The study of cyber security and the avoidance of cyber assaults, such as the use of intrusion detection as a defensive mechanism, is therefore needed. They provide a neural network approach to intrusion detection system threat prediction.

Proposed methodology

Developed the consumer behaviour exhibiting framework with a scientific method as our primary focus in order to evaluate how one kind of help compares to another in the public sphere. An accurate analysis of the content provided by customers may be useful to online business organisations in gaining information and better understanding their customers' aims and requirements. Calculations based on machine learning may assist us in plotting accurate visual renderings of this kind of customer behaviour. The planning of the framework makes use of both Naive Bayes and Logistic Regression, which are both components of machine learning classifiers.
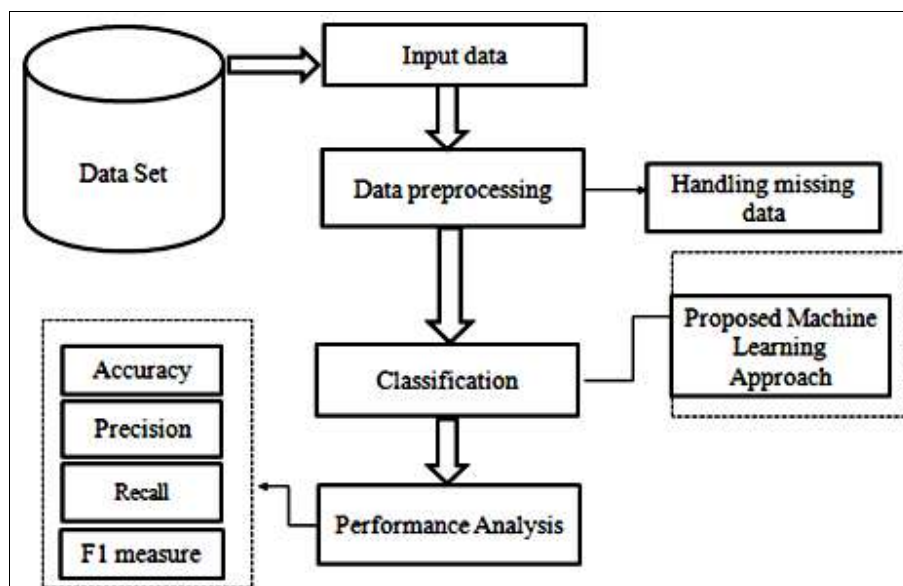

Fig 2: depicts the suggested flow chart for the process. The process of defining the input configuration and retrieving the consumer behaviour data set from the kaggle machine learning repository is the first step in the flow of work. So, before using the strategies for machine learning, you must first carry out the procedures for data pre-processing. At this stage, an example of the data is obtained; this data is sometimes referred to as the training data.

The process of defining the input configuration and retrieving the consumer behaviour data set from the kaggle machine learning repository is the first step in the flow of work. So, before using the strategies for machine learning,

you must first carry out the procedures for data pre-processing. At this stage, an example of the data is obtained; this data is sometimes referred to as the training data.

Now put the recommended strategy, which is based on the logistic regression, into action. At long last, all of the training data has been processed, and predictions of illnesses have been made. Now both the construction of the Results graph and the computation of the relevant parameters are complete.

Result and Analysis

Python Spider 3.7 is used in order to carry out the algorithm's implementation that has been suggested. The libraries sklearn, numpy, pandas, matplotlib, pyplot, seaborn, and so allow us utilise the functions that are made accessible in the spider environment for different approaches such as decision trees, random forests, naive bayes, and others.

Fig 3: displays the data set obtained from Amazon. This file contains a dataset with a total of 69000 people's information
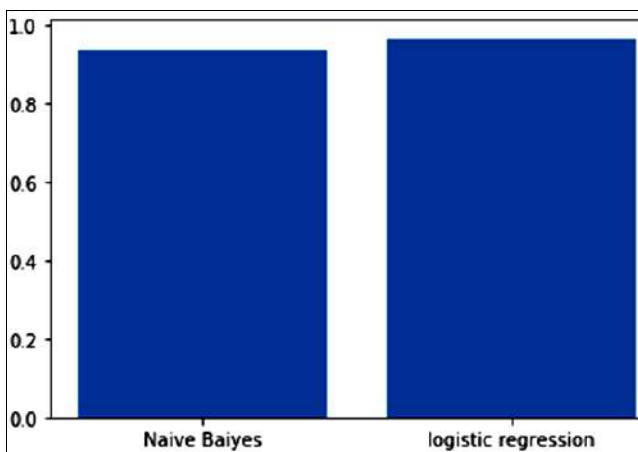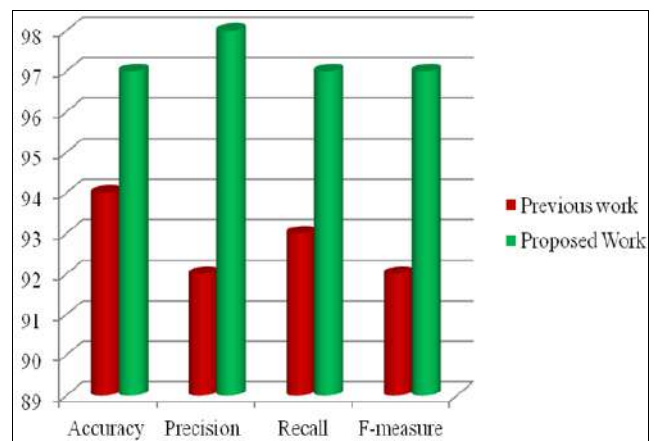

Fig 4: Accuracy comparison

Fig 5: Parameters comparison

The comparison of the results of the proposed work with the results of the previous study is shown in table 1 and figure 5. It is evident that the new technique provides an accuracy of 97%, while the old method only provided an accuracy of 94%.

Conclusions

The purpose of this article is to propose a strategy that may assist businesses in better understanding their clientele and

Table 1: Result Comparison

Sr. No.	Parameters	Previous work [1]	Proposed Work
1	Method	Naive Bayes	Logistic Regression
2	Accuracy	93.41 (94)	96.62 (97)
3	Classification error	6	3
4	Precision	92	98
5	Recall	93	97
6	F-measure	92	97

in implementing focused marketing strategies in order to expand both their clientele and their revenues. The use of sentiment analysis allowed us to assess the feelings of customers in relation to a variety of items, which in turn assisted us in analyzing the performance of the products in the market. The simulated results make it abundantly evident that the suggested method achieves an accuracy of 97%, while the old method achieved an accuracy of 94%. In the suggested method, the classification error is 3%, while in the prior technique, it was 6%. In the suggested model, the accuracy value is 98% and the F-measure is 97%, while in the prior model, those values were 92% and 92% respectively.

References

1. Shrirame V, Sabade J, Soneta H, Vijayalakshmi M. Consumer Behavior Analytics using Machine Learning Algorithms, 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India; c2020. p. 1-6. DOI: 10.1109/CONECCT50063.2020.9198562.
2. Kafeza E, Kanavos A, Makris C, Pispirigos G, Vikatos P. T-PCCE: Twitter Personality based Communicative Communities Extraction System for Big Data, in IEEE Transactions on Knowledge and Data Engineering. 2020;32(8):1625-1638.
   DOI: 10.1109/TKDE.2019.2906197.
3. Sharkh MA, Xu Y, Leyder E. Cloud Mach: Cloud Computing Application Performance Improvement through Machine Learning, 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada; c2020. p. 1-6. DOI: 10.1109/CCECE47787.2020.9255686.
4. Shahriar S, Al-Ali AR, Osman AH, Dhou S, Nijim M. Machine Learning Approaches for EV Charging Behavior: A Review, in IEEE Access. 2020;8:168980-168993. DOI: 10.1109/ACCESS.2020.3023388.
5. Edmond Meku Fotso J, Batchakui B, Nkambou R, Okereke G. 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; c2020. p. 180-184.
   DOI: 10.1109/LWMOOCS50143.2020.9234363.
6. Savenkov PA, Ivutin AN. Organizations Data Integrity Providing through Employee Behavioral Analysis Algorithms, 2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro; c2020. p. 1-3.
   DOI: 10.1109/MECO49872.2020.9134271.
7. Goodall JR, et al. Situ: Identifying and Explaining Suspicious Behavior in Networks, in IEEE Transactions on Visualization and Computer Graphics. 2019;25(1):204-214.
   DOI: 10.1109/TVCG.2018.2865029.
8. Jain, Jay Kumar, Akhilesh A Waoo, Dipti Chauhan. A Literature Review on Machine Learning for Cyber Security Issues; c2022.
9. Asniar, Surendro K. Predictive Analytics for Predicting Customer Behavior, 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia; c2019. p. 230-233.
   DOI: 10.1109/ICAIIT.2019.8834571.
10. Pereira FD, Oliveira EHT, Fernandes D, Cristea A. 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; c2019. p. 183-184. DOI: 10.1109/ICALT.2019.00066.
11. Salitin MA, Zolait AH. 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; c2018. p. 1-5. DOI: 10.1109/3ICT.2018.8855782.
12. Chauhan Dipti, Jay Kumar Jain, Sanjay Sharma. "An end-to-end header compression for multichip IPv6 tunnels with varying bandwidth. 2016 Fifth international conference on eco-friendly computing and communication systems (ICECCS). IEEE; c2016.
13. Bouhoute A, Oucheikh R, Boubouh K, Berrada I. Advanced Driving Behavior Analytics for an Improved Safety Assessment and Driver Fingerprinting, in IEEE Transactions on Intelligent Transportation Systems. 2019;20(6):2171-2184.
   DOI: 10.1109/TITS.2018.2864637.
14. Callara M, Wira P. User Behavior Analysis with Machine Learning Techniques in Cloud Computing Architectures, 2018 International Conference on Applied Smart Systems (ICASS), Medea, Algeria; c2018. p. 1-6. DOI: 10.1109/ICASS.2018.8651961.
15. Rajasekaran RG, Manikandaraj S, Kamaleshwar R. Implementation of Machine Learning Algorithm for predicting user behavior and smart energy management, 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), Pune. 2017. p. 24-30. DOI: 10.1109/ICDMAI.2017.8073480.
16. Zhang Y, Yang R, Zhang K, Jiang H. Zhang JJ. Consumption Behavior Analytics-Aided Energy Forecasting and Dispatch, in IEEE Intelligent Systems. 2017;32(4):59-63. DOI: 10.1109/MIS.2017.3121551.
17. Jain Jay Kumar, Akhilesh A Waoo. An Artificial Neural Network Technique for Prediction of Cyber-Attack using Intrusion Detection System. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN) ISSN: 2799-1172 3.02; c2023. p. 33-42.