INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

An Efficient Machine Learning Technique for Prediction of Consumer Behaviour with High Accuracy

Sandeep Sahu[1], Dr. Avinash Sharma [2]

Research Scholar, Dept. of CSE, Millennium Institute of Technology and Science, Bhopal, India[1]

Professor, Dept. of CSE, Millennium Institute of Technology and Science, Bhopal, India[2]

ABSTRACT: Consumers decide to buy a given product by looking at these ratings and reviews. Such content can be positive or negative reviews made by consumers who have previously used the product. The Machine Learning Algorithm can help us to visual representation of the data and vectorize the data. This paper presents the Naïve Bayes and Logistic Regression technique to analyze the consumer behavior. The logistic regression techniques optimized better performance than others. The existing problems are analyzed, and then, current solutions to these problems are presented and discussed. The experimental results show that the proposed method has higher precision, recall and F1 score. The method is proved to be effective with high accuracy on comments. The simulation and analysis is done using the python spyder 3.7 software.

KEYWORDS: Naïve Bayes, Logistic Regression, Precision, Recall, F1 score, spyder.

I. INTRODUCTION

A Machine learning (ML) is the scientific study of algorithms and statistical models that PC systems use to play out a specific task without using unequivocal instructions, depending on patterns and deduction instead. It is seen as a subset of computerized reasoning. Machine learning algorithms construct a numerical model based on sample information, known as "preparing information", so as to make predictions or decisions without being expressly programmed to play out the task.[1] Machine learning algorithms are used in a wide assortment of applications, such as email sifting and PC vision, where it is troublesome or infeasible to build up a traditional calculation for adequately playing out the task.

Machine learning is closely identified with computational statistics, which focuses on making predictions using computers. The study of scientific streamlining delivers methods, hypothesis and application domains to the field of machine learning. Information mining is a field of study inside machine learning, and focuses on exploratory information analysis through unsupervised learning.[3][4] In its application across business problems, machine learning is also alluded to as prescient analytics.

Several learning algorithms target discovering better representations of the inputs gave during training.[11] Classic examples incorporate head components analysis and cluster analysis. Highlight learning algorithms, also called representation learning algorithms, often endeavor to preserve the data in their information yet additionally transform it such that makes it useful, often as a pre-processing step before performing classification or predictions. This procedure allows reconstruction of the inputs originating from the unknown information producing distribution, while not being necessarily devoted to configurations that are implausible under that distribution. This replaces manual element designing, and allows a machine to both get familiar with the features and use them to play out a specific task.
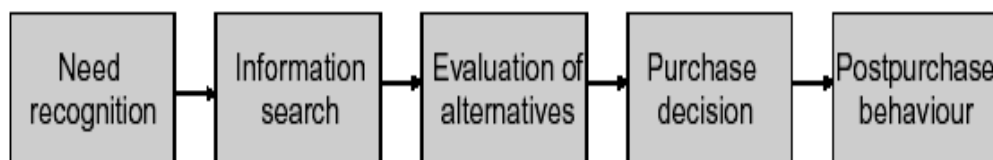

Figure 1: Five-Stage Model of the buying process

This model implies that consumers pass through all five stages in buying a product. This may be the case in high-involving purchases. In low-involvement purchases, consumers may skip or reverse some of these stages. This model shows the full range of considerations that arise when a consumer face a highly involving new purchase.

Consumer satisfaction provided by three general components. It can be identified in extant definitions:

1) Consumer satisfaction is a response (Emotional or Cognitive)

2) The response pertains to a particular focus (expectations, product, consumption experience, etc.)

3) The response occurs at a particular time (after consumption, after choice, based on accumulated experience, etc).

The pre-sales stage where there are the expectations for the product, the profits, the price and the availability of product. The sales stage when customer trays the environment, the product, the type of service, the delivery, the quality and the redress from the market. The after-sales stage when customer expects the support or the advices, the replacement of product or the return of sum, repairs and processes of charges.

II. PROPOSED METHODOLOGY

Focusing on scientific way to estimate how a new service is accepted in society, we developed consumer behavior modeling framework. An accurate analysis of this user-generated content can be helpful to e-commerce organizations to gain insights and understand their consumers' intentions and requirements. Machine Learning Algorithms can help us plot accurate visual representations of such consumer behaviour. Machine learning classifiers include Naïve Bayes, Logistic Regression are used in the designing of the system.
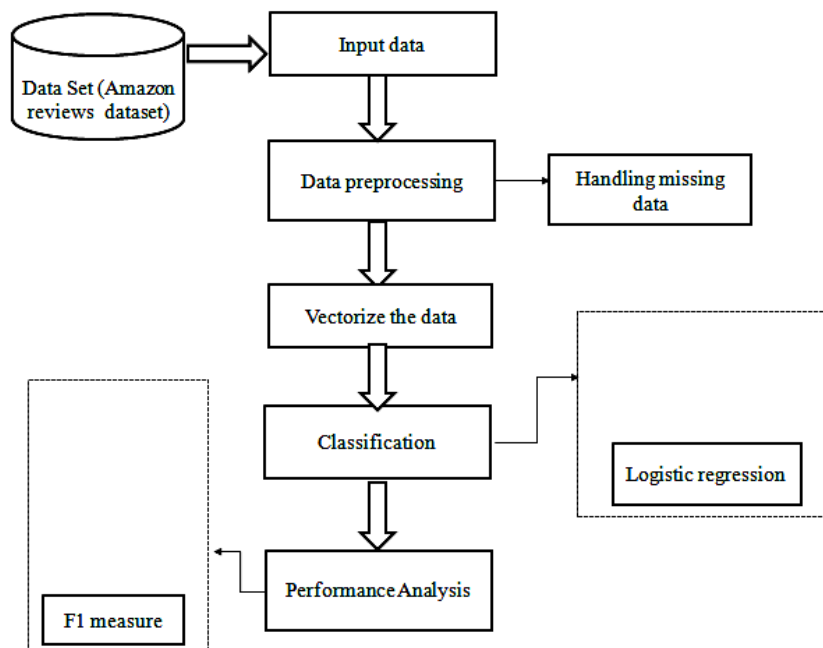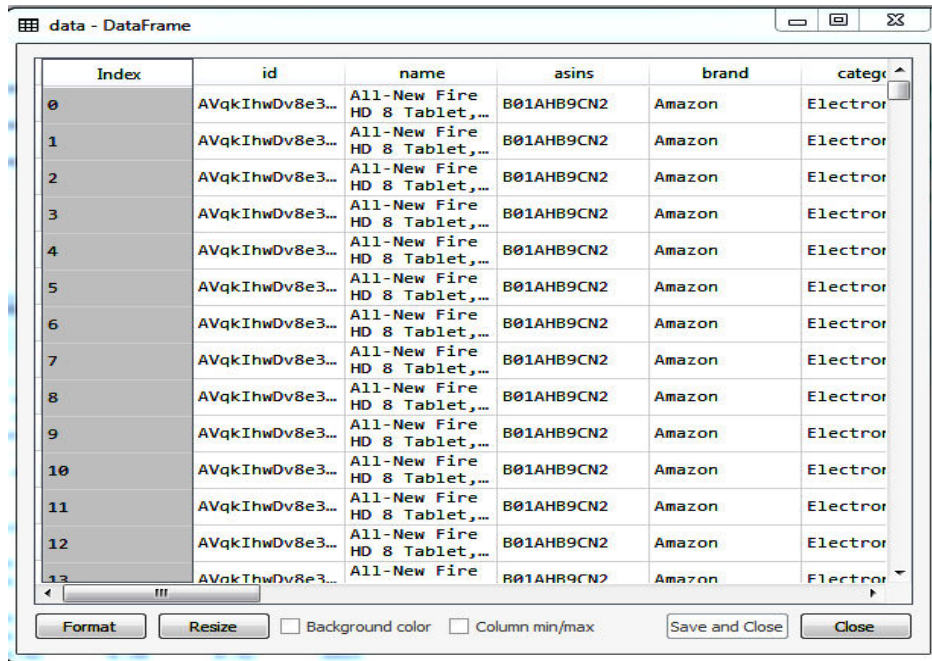

Figure 2: Flow Chart

Figure 2 is showing the proposed flow chart. The flow of work starts with to define input configuration and taken consumer behavior data set from the kaggle machine learning repository. Now before apply machine learning techniques, firstly apply the steps for data pre-processing. The sample of data is taken in this step, it is also known as training data.

Now apply proposed approach based on the logistic regression. At last all training data is process and give predication of diseases. Now, Results graph generation and calculation of necessary parameters is done.

III. RESULT AND ANALYSIS

The implementation of the proposed algorithm is done over python spyder 3.7. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library helps us to use the functions available in spyder environment for various methods like decision tree, random forest, naive bayes etc.

Figure 3: Dataset

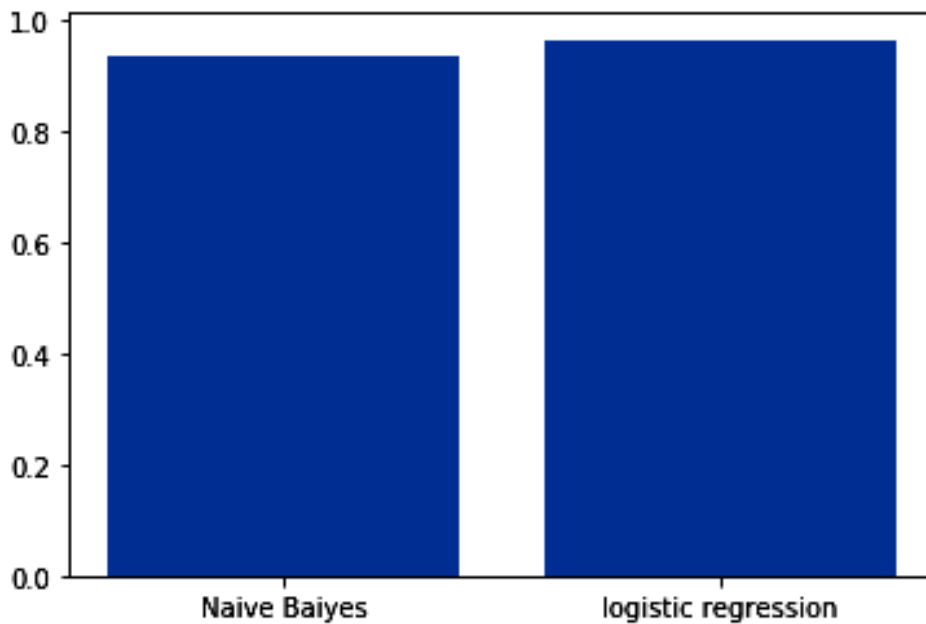Figure 3 is showing the amazon data set. Total 69000 person dataset given in this file.

Figure 4: Accuracy comparison

Table 1: Comparison of Proposed work with previous work

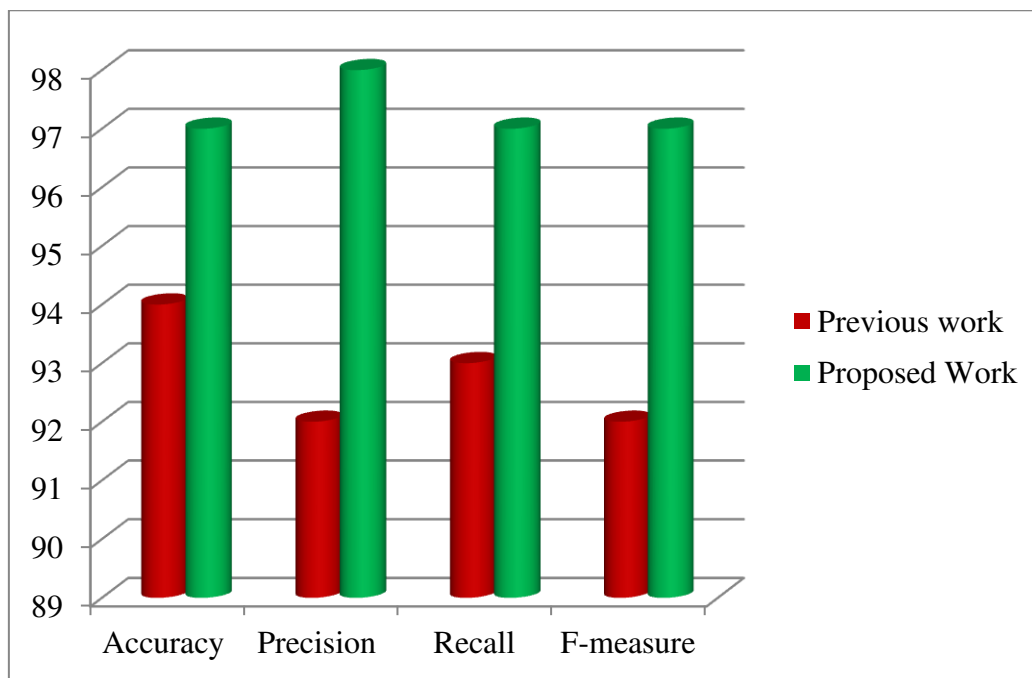| Sr. No.	Parameters	Previous work [1]	Proposed Work
1	Method	Naive Bayes	logistic regression
2	Accuracy	93.41 (94)	96.62 (97)
3	Classification error	6	3
4	Precision	92	98
5	Recall	93	97
6	F-measure	92	97

Figure 5: Parameters comparison

Table 1 and figure 5 is showing that the comparison of proposed and previous work results. It is clear that proposed method gives 97% accuracy while in previous there is 94% accuracy

IV. CONCLUSION

This paper presents an approach to help the organizations in knowing their customers and incorporating targeted marketing techniques to increase their customer base and profits. Sentiment analysis helped us evaluate consumers' sentiments related to various products which in turn helped us analyze the product's performance in the market. It is clear from simulated results that proposed approach gives 97% accuracy while in previous there is 94% accuracy. The classification error is 3% in proposed while 6% in previous approach. The precision value is 98% and F-measure is 97% in proposed while previously it 92% and 92% respectively.

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.

[13].   M. Callara and P. Wira, "User Behavior Analysis with Machine Learning Techniques in Cloud Computing Architectures," 2018 International Conference on Applied Smart Systems (ICASS), Medea, Algeria, 2018, pp. 1-6, doi: 10.1109/ICASS.2018.8651961.

[14].   R. G. Rajasekaran, S. Manikandaraj and R. Kamaleshwar, "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, pp. 24-30, doi: 10.1109/ICDMAI.2017.8073480.

[15].   Y. Zhang, R. Yang, K. Zhang, H. Jiang and J. J. Zhang, "Consumption Behavior Analytics-Aided Energy Forecasting and Dispatch," in IEEE Intelligent Systems, vol. 32, no. 4, pp. 59-63, 2017, doi: 10.1109/MIS.2017.3121551.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING