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

Customer Behaviour Analysis in E-commerce Dataset Using Machine Learning

[1]Gaurav Vyas, [2]Prof. Chetan Gupta, [3]Dr. Kapil Chaturvedi

[1]Research Scholar, [2]Assistant Professor, [3]Associate Professor
Department of Computer Science Engineering
Sagar Institute of Research Technology & Science, Bhopal, India
vyasgaurav1985@gmail.com, chetangupta.gupta1@gmail.com, kapil.rgtu@gmail.com

Abstract : Customer choose to purchase a given item by looking at these evaluations and surveys. Such substance can be positive or negative audits made by customer who have recently utilized the item. The Machine Learning Calculation can help us to visual portrayal of the information and vectorize the information. This paper presents the Naïve Bayes and Logistic Regression etc method to investigate the customer behavior. The decision tree classification method achieved improvement in the performance parameters over others. The current issues are examined, and afterward, current answers for these issues are introduced and talked about. The simulation results show that the proposed strategy has higher precision, recall and F1 score. The strategy is end up being effective with high accuracy on remarks. The reproduction and analysis is finished utilizing the python spyder 3.7 software.

***IndexTerms* – Machine Learning, Customer Prediction Model, Decision tree.**

I. INTRODUCTION

Artificial intelligence (AI) based customer behaviour prediction models are effective working with unstructured data and identifying hidden features and similarities to make groups of data samples united by common traits. These models can also predict prices, demand, or weather. AI can make customer behaviour predictions by segmenting customers into groups, as customers with similar features will likely have similar buying behaviours. Customer demographics such as gender, age, annual income, and spending score are usually taken into account to evaluate their similarity.

Machine learning is firmly identified with computational insights, which focuses on making forecasts utilizing PCs. The investigation of scientific smoothing out conveys strategies, theory and application spaces to the field of machine learning. Information mining is a field of study inside machine learning, and focuses on exploratory information analysis through unaided learning.[3][4] In its application across business issues, machine learning is likewise implied as perceptive investigation.

Figure 1: Five-Stage Model of the buying process

A few learning calculations target finding better portrayals of the data sources gave during training.[11] Exemplary models join head segments analysis and group analysis. Feature learning calculations, likewise called portrayal learning calculations, often try to save the information in their information yet also transform it to such an extent that makes it useful, often as a pre-handling venture before performing classification or expectations. This methodology permits recreation of the data sources starting from the unknown information delivering dissemination, while not being fundamentally given to configurations that are unlikely under that conveyance. This replaces manual component planning, and permits a machine to both get comfortable with the features and use them to play out a specific task.

This model suggests that customer go through each of the five phases in purchasing an item. This might be the situation in high-including buys. In low-inclusion buys, customer may skip or converse a portion of these stages. This model shows the full scope of contemplations that emerge when a customer face a profoundly including new buy.

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

1) Customer 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-deals stage where there are the assumptions for the item, the profits, the cost and the accessibility of item. The business stage when client plate the climate, the item, the kind of administration, the conveyance, the quality and the change from the market. The after-deals stage when client anticipates the support or the advices, the substitution of item or the arrival of aggregate, fixes and cycles of charges.

II. PROPOSED METHODOLOGY

Focusing on scientific approach to assess how assistance is acknowledged in the public arena, we created customer behavior displaying framework. An exact analysis of this client created substance can be helpful to online business associations to acquire bits of knowledge and comprehend their customer' goals and prerequisites. Machine Learning Calculations can help us plot exact visual portrayals of such customer behavior. Machine learning classifiers incorporate Decision tree, Naïve Bayes, Logistic Regression are utilized in the planning of the framework.

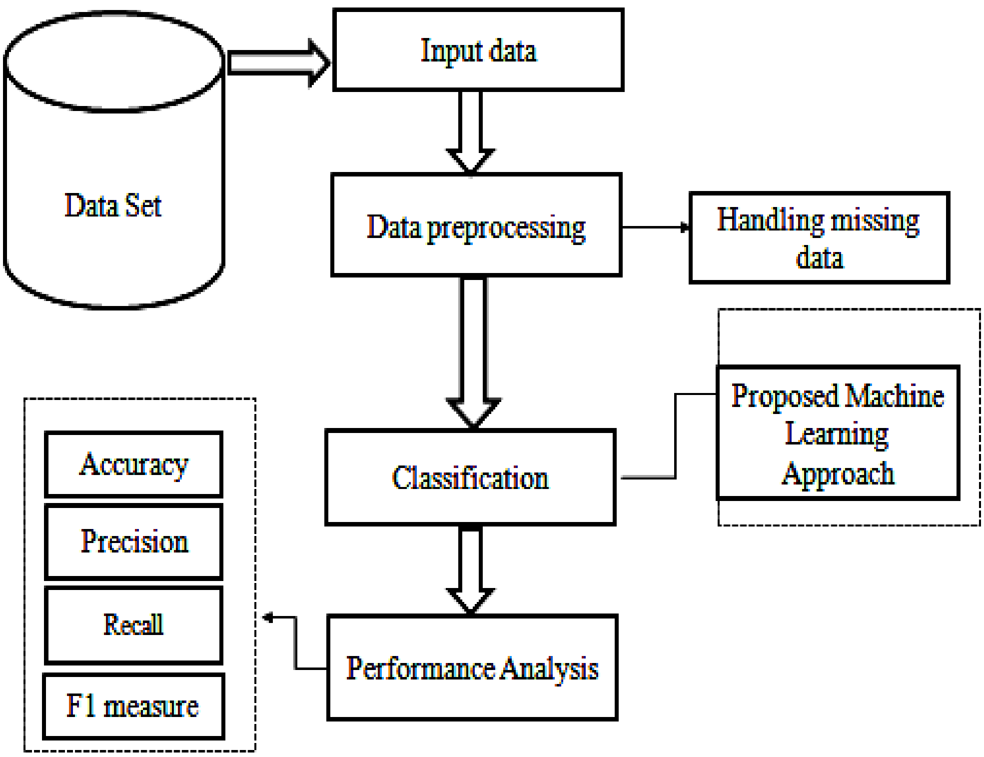


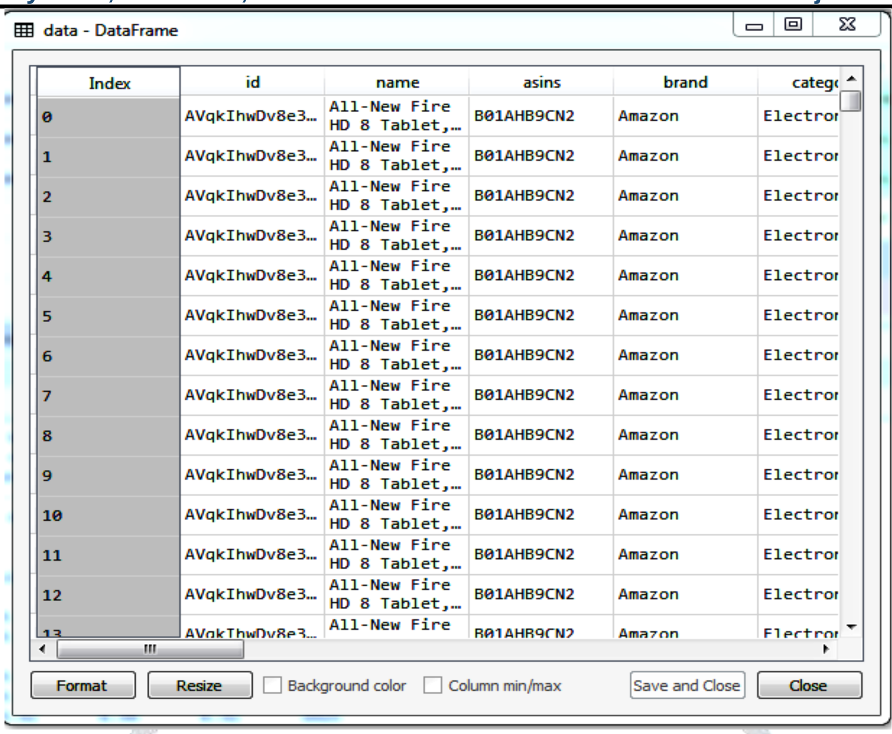
Figure 2: Flow Chart

Figure 2 is showing the proposed flow chart. The flow of work starts with to define input configuration and taken customer 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 Decision tree , logistic regression etc. 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.

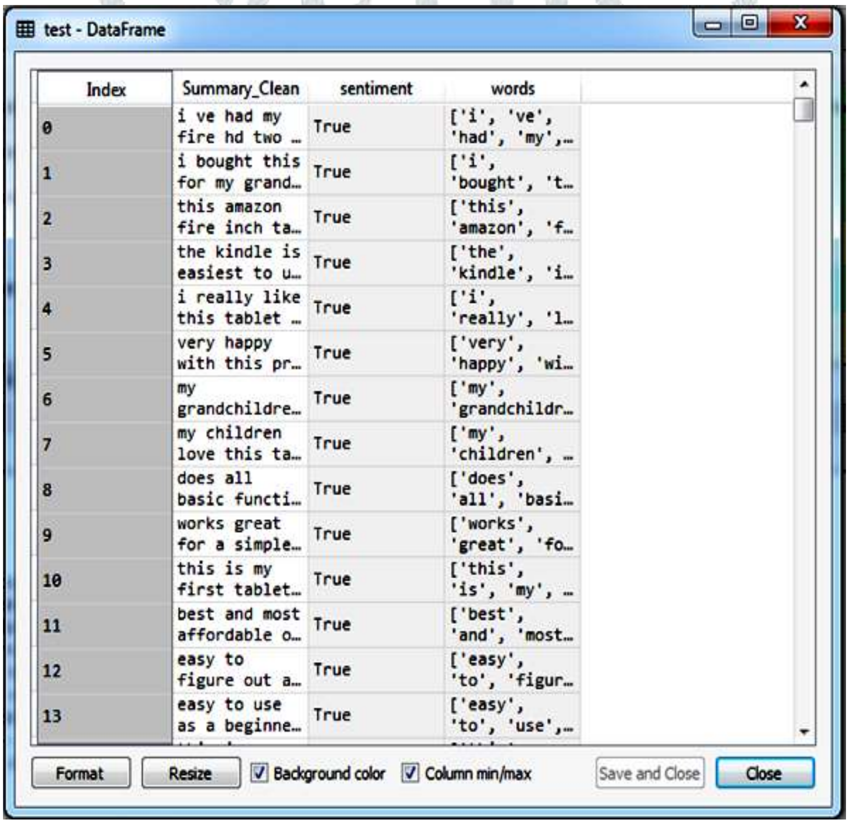


data - DataFrame

Index	id	name	asins	brand	categ
0	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
1	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
2	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
3	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
4	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
5	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
6	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
7	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
8	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
9	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
10	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
11	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
12	AVqkIhwDv8e3…	All-New Fire HD 8 Tablet,…	B01AHB9CN2	Amazon	Electror
13	AVqkIhwDv8e3…	All-New Fire	B01AHB9CN2	Amazon	Electror

Figure 3: Dataset

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



test - DataFrame

Index	Summary_Clean	sentiment	words
0	i ve had my fire hd two …	True	['i', 've', 'had', 'my',…
1	i bought this for my grand…	True	['i', 'bought', 't…
2	this amazon fire inch ta…	True	['this', 'amazon', 'f…
3	the kindle is easiest to u…	True	['the', 'kindle', 'i…
4	i really like this tablet …	True	['i', 'really', 'l…
5	very happy with this pr…	True	['very', 'happy', 'wi…
6	my grandchildre…	True	['my', 'grandchildr…
7	my children love this ta…	True	['my', 'children', …
8	does all basic functi…	True	['does', 'all', 'basi…
9	works great for a simple…	True	['works', 'great', 'fo…
10	this is my first tablet…	True	['this', 'is', 'my', …
11	best and most affordable o…	True	['best', 'and', 'most…
12	easy to figure out a…	True	['easy', 'to', 'figur…
13	easy to use as a beginne…	True	['easy', 'to', 'use',…

Figure 4: Test data

Figure 4 is showing the test data from given dataset. The Test dataset provides the gold standard used to evaluate the model. It is only used once a model is completely trained(using the train and validation sets).

train - DataFrame

Index	Summary_Clean	sentiment	words
0	i am very happy with m…	pos	['i', 'am', 'very', 'hap…
1	decent tabletperfor…	neg	['decent', 'tabletperfo…
2	this product works better…	pos	['this', 'product', '…
3	absolutely best amazon …	pos	['absolutely…
4	my son absolutely l…	pos	['my', 'son', 'absolutely'…
5	we bought these for my…	pos	['we', 'bought', 't…
6	i love this tablet it is…	pos	['i', 'love', 'this', 'tab…
7	i am very pleased with…	pos	['i', 'am', 'very', 'ple…
8	small and easy to use …	pos	['small', 'and', 'easy…
9	i love hear readers i ha…	pos	['i', 'love', 'hear', 'rea…
10	i got this for my siste…	pos	['i', 'got', 'this', 'for…
11	kindle is good i got o…	pos	['kindle', 'is', 'good'…
12	my wife has an older pap…	pos	['my', 'wife', 'has…
13	so pleased to have investe…	pos	['so', 'pleased', '…

Format | Resize | Background color | Column min/max | Save and Close | Close

Figure 5: Train Data

Figure 5 is showing the train dataset from given dataset. A training dataset is a dataset of examples used during the learning process and is used to fit the parameters (e.g., weights) of, for example, a classifier.

	0	1
0	782	144
1	121	12804

Figure 6: Confusion Matrix (DT)

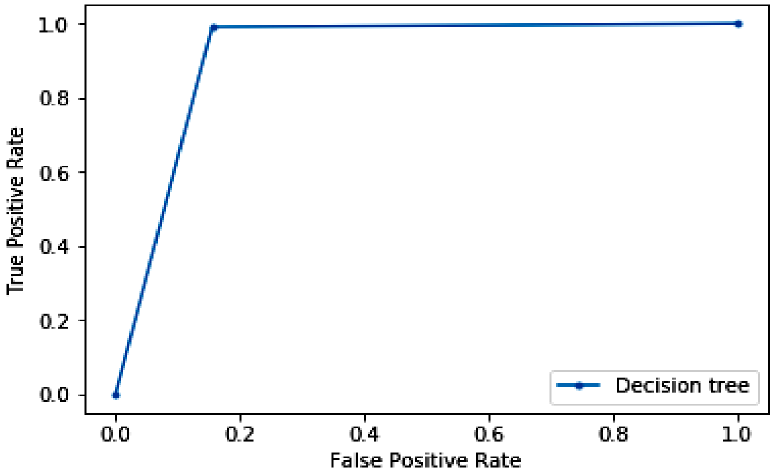


Figure 7: ROC of Decision Tree

Table 1: Result Comparison

Sr. No.	Parameters	Previous work	Proposed Work
1	Method	Collective Method	Decision Tree
2	Accuracy	94	98.08
3	Classification error	6	1.91
4	Precision	55	84.44
5	Recall	40	86.60
6	F-measure	50	85.51

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 customer' 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 98% accuracy while in previous there is 94% accuracy. The classification error is 2% in proposed while 6% in previous approach. Therefore the proposed approach gives significant better results than previous approach.

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. J. Jeong, "Identifying Customer Preferences From User-Generated Content on Amazon.Com by Leveraging Machine Learning," in IEEE Access, vol. 9, pp. 147357-147396, 2021, doi: 10.1109/ACCESS.2021.3123301.
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. X. Chen, Y. Li, J. Shimada and N. Li, "Online Learning and Distributed Control for Residential Demand Response," in IEEE Transactions on Smart Grid, vol. 12, no. 6, pp. 4843-4853, Nov. 2021, doi: 10.1109/TSG.2021.3090039.
5. S. Wu, W. -C. Yau, T. -S. Ong and S. -C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," in IEEE Access, vol. 9, pp. 62118-62136, 2021, doi: 10.1109/ACCESS.2021.3073776.
6. K. Ali and A. X. Liu, "Monitoring Browsing Behavior of Customers in Retail Stores via RFID Imaging," in IEEE Transactions on Mobile Computing, doi: 10.1109/TMC.2020.3019652.
7. E. Umuhoza, D. Ntirushwamaboko, J. Awuah and B. Birir, "Using Unsupervised Machine Learning Techniques for Behavioral-based Credit Card Users Segmentation in Africa," in SAIEE Africa Research Journal, vol. 111, no. 3, pp. 95-101, Sept. 2020, doi: 10.23919/SAIEE.2020.9142602.
8. L. Fan, J. Li and X. -P. Zhang, "Load prediction methods using machine learning for home energy management systems based on human behavior patterns recognition," in CSEE Journal of Power and Energy Systems, vol. 6, no. 3, pp. 563-571, Sept. 2020, doi: 10.17775/CSEEJPES.2018.01130.
9. Y. Yuan, K. Dehghanpour, F. Bu and Z. Wang, "A Data-Driven Customer Segmentation Strategy Based on Contribution to System Peak Demand," in IEEE Transactions on Power Systems, vol. 35, no. 5, pp. 4026-4035, Sept. 2020, doi: 10.1109/TPWRS.2020.2979943.
10. F. Zheng and Q. Liu, "Anomalous Telecom Customer Behavior Detection and Clustering Analysis Based on ISP's Operating Data," in IEEE Access, vol. 8, pp. 42734-42748, 2020, doi: 10.1109/ACCESS.2020.2976898.
11. Y. Zhang, S. He, S. Li and J. Chen, "Intra-Operator Customer Churn in Telecommunications: A Systematic Perspective," in IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 948-957, Jan. 2020, doi: 10.1109/TVT.2019.2953605.
12. E. A. E. Dawood, E. Elfakhrany and F. A. Maghraby, "Improve Profiling Bank Customer's Behavior Using Machine Learning," in IEEE Access, vol. 7, pp. 109320-109327, 2019, doi: 10.1109/ACCESS.2019.2934644.