COMPARISON ESTIMATION OF EFFECTIVE CONSUMER FUTURE PREFERENCES WITH THE APPLICATION OF AI

Veena Parihar[1], Dr. Surendra Yadav[2]
PhD Scholar, [1] Professor [2]
Career Point University, Kota, Rajasthan, India

Vivekananda Journal of Research
October, 2021, Vol. 10, Special Issue, Pg No. 133-145
ISSN 2319-8702(Print)
ISSN 2456-7574(Online)
Peer Reviewed Refereed Journal
© Vivekananda Institute of Professional Studies
http://www.vips.edu/vjr.php

Abstract

Because of shift from visiting retail stores to shopping online, customers' behavior in the context of e-commerce is becoming increasingly relevant today. It will increase consumer satisfaction and sales by promoting a more personalized shopping experience, which will lead to higher transaction rates and a competitive advantage. Clickstream and additional customer data will be used to build models for predicting consumer behavior. This research paper is based on comparing different model learning to predict a customer's purchase.
KEYWORDS*— ECG Signals, Arrhythmia, Electrocardiogram, CNNs, MIT-BIH*
Keywords: Consumer, Clickstream, e-commerce, Purchase, Transaction, Machine Learning

1. INTRODUCTION

Understanding what customers want and expect — and, hopefully, anticipating their needs — is a constant challenge for advertisers. With the advent of deep learning, artificial intelligence (AI) will make the work even simpler.

Deep learning, a branch of AI, has the ability to change the direction of marketing by assisting companies in predicting customer behaviour. It's a machine-learning technique that utilizes stacked or "deep" neural networks, which are close to those used in biological brains, to learn skills and solve complex problems faster than humans. It aids computers (or robots) in performing "human" tasks such as object perception, speech recognition, and language translation.

Deep learning allows AI to be trained to predict outcomes based on a series of inputs. It will seem easy, but it is not: Although deep learning demands less human data preprocessing than traditional machine learning methods, it does necessitate a broad data collection and a lot of computing resources. If a deep learning algorithm has access to certain main components, it can learn to reliably model human actions.

For a better comprehension, the binary classification and many machine learning algorithms can be used to solve the problem and the various algorithms can be clarified before their literature implementation is reviewed.

2. BINARY CLASSIFICATION

Data mining has numerous uses for the most general classification. The classification attempts to forecast a group target variable from a collection of input variables as a predictive analytical task. This objective variable may either be represented in separate divisions or be binary (Kotu & Deshpande,

2014). This task is a binary task, so there are two categories: purchasing, and no purchase of the goal variable. A generic correlation between the input and the target variable is taken from a labeled data set to forecast the target variable. It is then used to identify new data. Learning algorithms should adapt and generalize to the training data (P. Tan, Steinbach, & Kumar, 2005). There are various algorithms for machine learning with diverse approaches for this relationship.

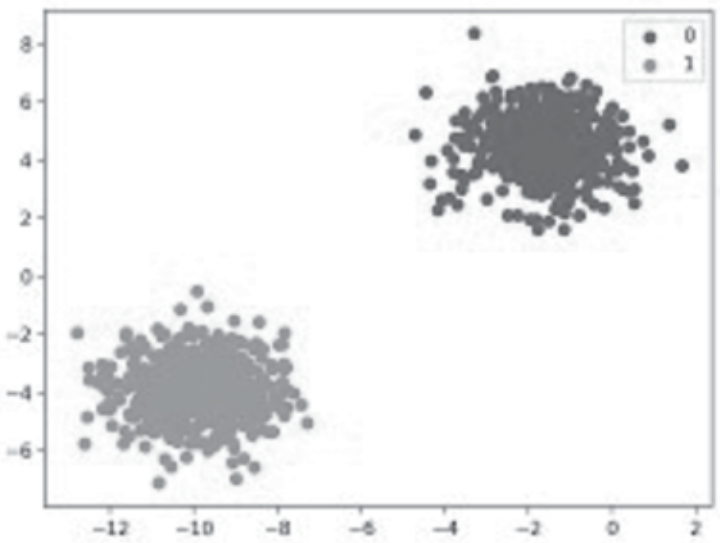


Figure 1 Binary classification

3. VARIOUS LEARNING ALGORITHMS FOR CLASSIFICATION

Vector-based approaches are the most general kind of machine learning algorithms for a binary classification problem. DTs, RFs, SVMs, LR and FNNs are part of this group. This baseline model is also part of the DT algorithm for this group. In this way, a sequence of vectors and corresponding outputs are typical to these algorithms (Heaton, 2016). They are enthusiastic models for learning where a classification process is built on the basis of a particular training dataset prior to classification by model.

On the contrary, lazy learners, like the K-nearest Neighbor (KNN) algorithm, simply store training and wait for the classification of a research data point (Han, Pei, & Kamber, 2011).

All the above are stateless master research algorithms; they have no memory and return always the same response provided the same input. However, it is complicated over time, since the various countries need to be modeled by sophisticated characteristic engineering, generating inaccuracies and increasing difficulty by increasing the amount of input features. This is well suited for certain classification activities. The ability to model sequences and patterns of extracts over time, however, can be useful for the analysis in hand, because of the click-based data used in this analysis.

Fortunately, there are stateful, memory-based models that remember prior states and extract sequential patterns without explicitly needing to create time-dependent features. RNNs and HMCs are examples of these models. These are explained below in detail in all men's algorithms, each of which is

followed by application and literary courses.

Decision Trees

DTs consist of a series of divisional conditions dividing a heterogeneous population into smaller, more even subgroups with regard to a particular variable. The objective is to create the most uniform subgroups. There are different algorithms to identify the best divisions, like Hunts, which follow a gullible strategy of best local decisions (P. Tan et al., 2005). Simple DTs benefit from convertibility into simple classification rules which are intelligible (Han et al., 2011). This understanding decreases with the widening and unbalanced models (Rokach & Maimon, 2014). Generally, DTs offer a pre-dictional speed and relatively fast learning. While the various types vary in understanding, they are still easier to understand than black box models, like FNN or SVMs (Rokach & Maimon, 2014). The disadvantages are the required characteristics engineering, the impossibility to model time streaks implicitly and the increased complexity with many category categorized trees. Methods that affect only unique DTs, called non-ensemble DTs, are often overfit and unstable with regard to noisy information (Hop, 2013).

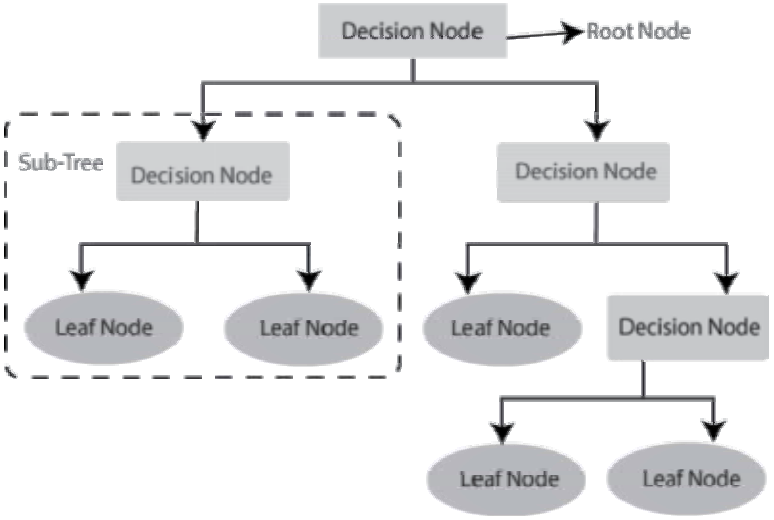


Figure 2 Decision tree

Boosted Decision Trees

Boosted D Ts, consisting of more than one DT, belong to the ensemble methods. Here a sequence of trees is constructed, in which the residual prediction of the previous tree results in each tree (Friedman, 2002). The baseline model used for this study is an example for such a method. Boosted DTs have proven to be a powerful means for predictive analysis by winning many contests in Kaggle machine learning (Kaggle, 2017) but are less understandable than simple DTs as they comprise many trees.

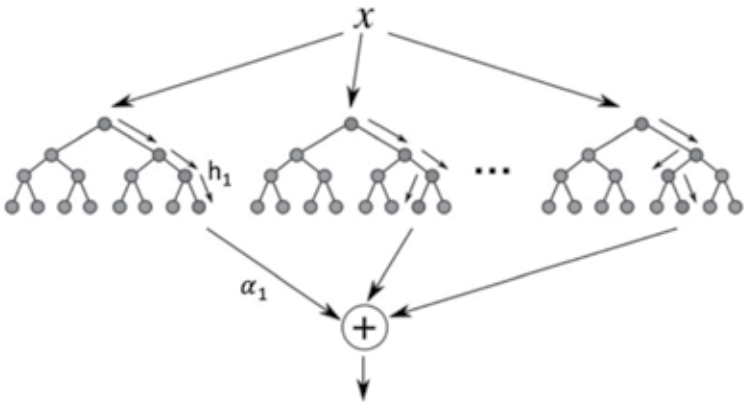


Figure 3 Boosted Decision Tree

Random Forest

Bagging is yet another example of ensembles where many large trees are suited and classified by a majority vote for the re-sampled versions of the data (Breiman, 1996). RF enhances the bagging of trees through de-correlation. A random sample of features is selected after each tree is split and they are the only ones considered for the next split. Again, the results are based on the majority vote of the individual trees. The weaknesses of non-ensemble DTs, such as robustness and fitness improve through the use of a large number of classifiers, sacking and RFs. They train quicker than boosted trees, but take longer to predict (Breiman, 2001). Bagging and RFs are nevertheless not dependent on the engineering of features and cannot model time dependencies.

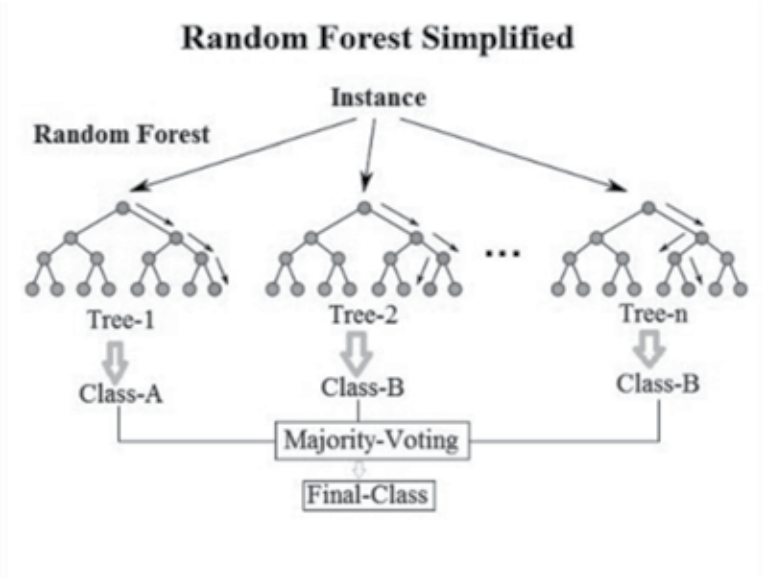


Figure 4 Random Forest

DTs demonstrate great effects with very related problems in the literature. For eg, for the session to be categorized as a buying or no buying session, Boginaetal.(2016) have different DT algorithms. A configuration was used alongside clickstream data, which improved the prediction accuracy with

additional knowledge about item sales statistics. If only press stream information is used, the nearest to our use case, the best outcomes have been found in a Bagging RepTree. The Weka method, assisted by data analysis with various techniques (Frank, Hall & Witten, 2016), showed a 0,824 precious value, a 0,808 reminder, a 0,806 F1 score, and a 0,889 ROC area (AUC).

Support Vector Machines

An SVM divides two groups by placed between them in a hyperplane. This is achieved by using one hyperplane, which varies from DTs, when after each break, a hyperplane is inserted. When the SVM senses the highest hyperplane margin, maximizing the length of the data points of both groups, where several divergent hyperplane are located. This results in increased generalization and thereby enhanced test accuracy. If a functional space is not linearly split, a kernel function is used to map data to a more dimensional space where the data becomes linearly split (Hofmann, 2006). The mapping of inputs into a broad functional field allows SVMs to model not only linear relations but also to perform non-linear classifications (Cortes & Vapnik, 1995).

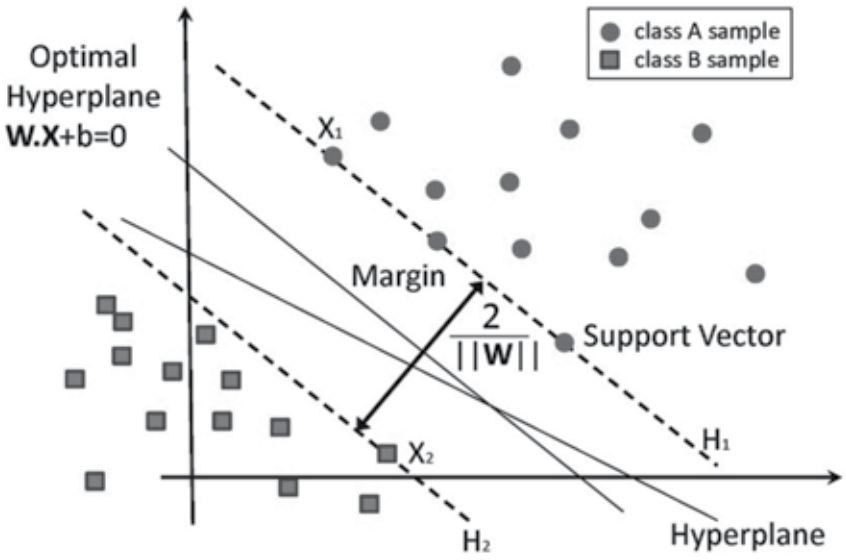


Figure 5 Support Vector Machine

SVMs display high precision and short predictive times in terms of their efficiency. It also works very well for large-scale data entry. Long preparation hours and inability to be understood are drawbacks. In addition, in addition to function design, it often includes hyper-parameter tuning, which can be complicated and time consuming (Han et al., 2011).

The analysis by Hop (2013) reveals that SVMs are not able yet do better than the FNN than the RF algorithm. Lee et al. (2015) have used an SVM to estimate the possibility of buying a given commodity. Item information was included next to the clickstream results. The three models independently were evaluated and tested on the item data, clickstream data and the cumulative data collection, in order to compare the im- portance of various data types on prediction performance.

Logistic Regression

LR is a generalized linear model class called logit regression that is used for estimating categorical target variables. LR is also called logit regression. The logistic function is obtained by means of a signmoid curve that takes values from 0 to 1. This function is modelled in linear combination of input values and coefficients.

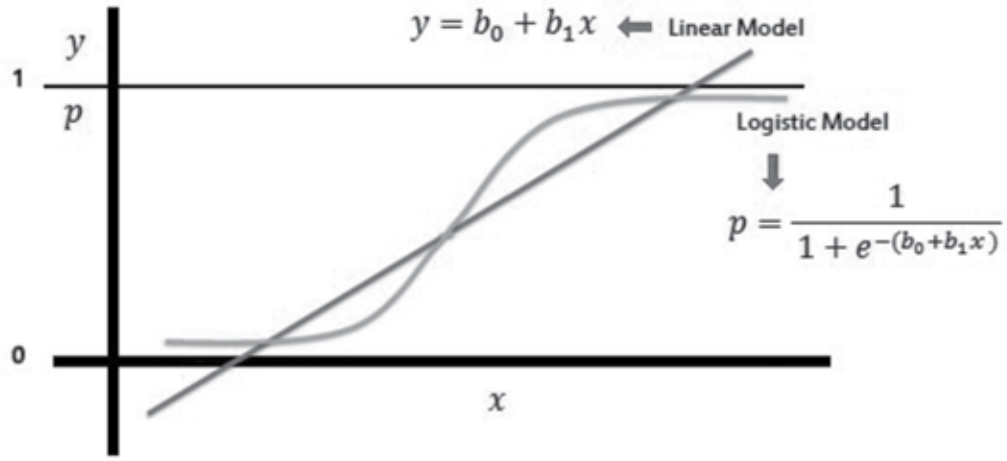


Figure 6 Logistic Regression

Each input column learns a coefficient by the maximum likelihood approximation from the training results (Hastie, Tibshirani, & Friedman, 2002). LR is very easy to predict and train time; it is therefore one of the most common binary classification ma-chine learning algorithms. However, it includes function engineering and categorical variables to be encoded. Therefore, outliers can be excluded before preparation and is therefore susceptible to noise. Moreover, in strongly coupled factors of data, LR does not suit by type.

Therefore, you should use printed cipal components for regression only, linearly uncorrelated variables (Jolliffe, 1982). The main component analysis (PCA), a process to extract the main components from a set of possible correlated variables, can help this to be achieved.

Neural Networks

The Artificial Neural Networks are one of the easiest forms. The data input flows only from the input nodes to the hidden nodes and the output nodes (Zell, 1994). Nodes are stacked where each layer node is linked to the preceding layer. Nodes are layered. Figure 2.2 demonstrates an outstanding architecture. The connections are allocated different weights that are generated by, for instance, back propagation during the learning stage. Each node has an activation feature and accordingly only burns.

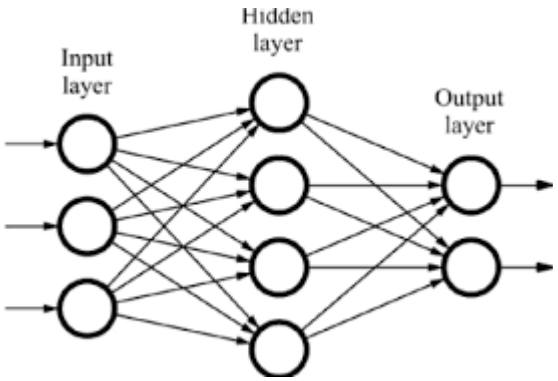


Figure 7 Neural Network

Performance is the chance of each class, which is one. An FNN can approximate any function with ample hidden units. FNNs perform a nonlinear regression from a statistical point of view. This neural networks have the downside of long stretches of preparation and little awareness. They need hyper-parameter tuning next to function engineering. However, they have shown very good literature results and generalize patterns well to unseen patterns when tolerating noisy results (Han et al., 2011).

In addition, Suchacka and Templewski (2017) could attain high precision (0.996) and reminder (0.878) of a FNN on the Clickstream Results, to foretell transactions during session except for research from the previously stated Hop (2013) with weak performance in the FNNs. For the training of the model, they used static and dynamic session data and static client data. The great discrepancy in results from Hop's paper may be that Hop uses a basic FNN with only one hidden layer and separate data is used for the training and testing of the models on both papers.

K-nearest Neighbor

As already mentioned, the KNN is a lazy learning algorithm that merely saves training data and awaits a test data point for classification (Han et al., 2011).

In an n-dimensional function field, all the stored instances correspond to the points. Range scales, most usually Euclidean, describe one point's closest neighbours. The most popular mark among its k -nearest neighbours will be given an unlabeled test data point. The benefits of this approach are its robustness to sound data and its incredibly rapid exercise tempo.

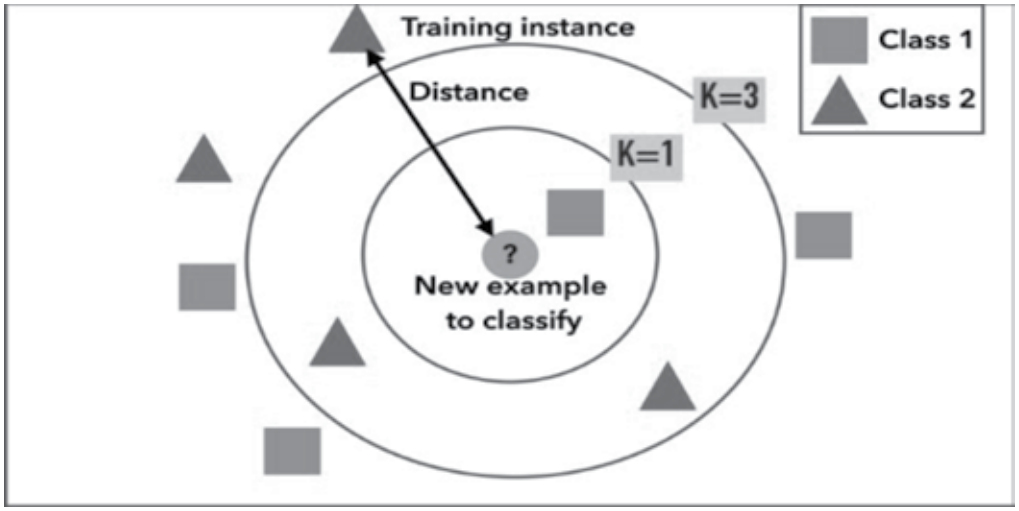


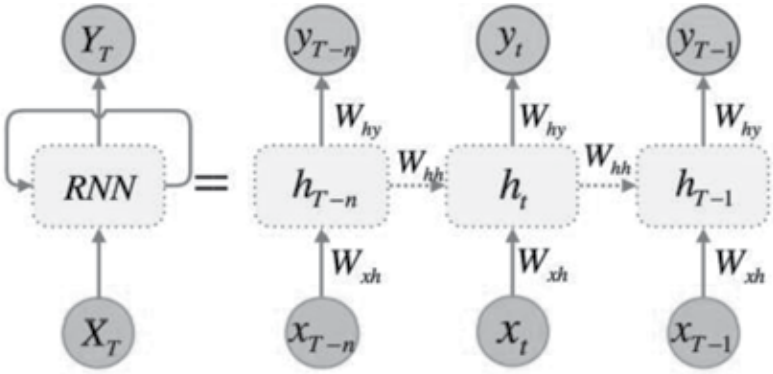
Figure 8 K nearest Neighbour

Disadvantages involve increased dimensionality complexity by irrelevant characteristics and ultimately reduced performance, stressing the importance of practical technologies, longer forecasts than the willing learning models and poor understandability and high dimensional feedback.

The KNN model was widely used in the e-commerce industry for comparison structures, in which items are recommended to web shop visitors based on their nearest neighbours' tastes. Just one study by Suchacka, Skolimowska-Kulig and Potempa (2015) could be found for classifying consumer behaviour. The study aimed at classifying web shop sessions for consumers in shopping or surfing sessions. As a model, the clickstream data scattered by sessions as an input was used as a KNN algorithm. 26,000 instructional records and 13,000 assessments have been used.

Recurrent Neural Networks

Due to the consistency of the RNNs, certain activities such as natural language processing were carried out. Their layout is similar in appearance to FNNs with added loops to add memory and machine input . For each new input, the RNN latent status is changed and revised. This state is learned and offers a wider abstraction of the input functions which makes feature engineering less required than other methods (Brownlee, 2016; Lang & Rettenmeier, 2017).



Reducing the amount of feature design not only saves time and complicates, but will deliver better results, given that feature engineering is an abstraction of the real data, which can lead to the loss of important information. By way of the RNN memory, sequential data pattern recognition is enabled, which makes it a good match for the clickflow data used in this job. RNNs are trained in time back-propagation. This is a regular back-transmission adaptation, where errors are spread over the network and weights are updated accordingly (Brownlee, 2016). When the input sequence is longer, this fault gradient disappears or explodes and thus results in a loss of information known as the disembodiment problem (Bengio, Simard, & Frasconi, 1994; Korpusik et al., 2016). The short-term memory network (LSTM) provides a solution to the disappearing gradient problem. Neurons are substituted for memory cells in LSTMs. These cells function through the input, output and status gates of the cells. These doors have a sigmoid function activated that makes the information circulate through the cell dependent (Brownlee, 2016). This keeps information in the memory cell that increases predictive precision and solves the problem of disappearing gradients (Hochreiter & Schmidhuber, 1997).

Literature Review

With today's knowledge economy and information society, e-commerce is gaining in prominence worldwide. The most popular form is business-to-customer exchange, which is traditional as online shops, which easily displaces physical shops (Suchacka & Chodak, 2016). Numbers such as an increase in German dedicated to online purchases of 3 percent in 2016 will contribute to the transition from physical to on-line shopping, with sales on e-commerce increasing by an unprecedented 12.5% to USD 58,52 billion and projected to reach USD 86 billion by the end of 2021. (Retail Ecommerce in Germany: A Major Digital Market Growing in Size and Sophistication, 2017).

Stefano Puntoni et. al (2021), This research aims to bridge these two perspectives: on one side, the authors acknowledge the value that embedding AI technology into products and services can provide to consumers. On the other side, the authors build on and integrate sociological and psychological scholarship to examine some of the costs consumers experience in their interactions with AI. In doing so, the authors identify four types of consumer experiences with AI: (1) data capture, (2) classification, (3) delegation, and (4) social.

Figi Report (2021) This report explores various challenges that consumer protection and data privacy law and regulation face with regard to big data and machine learning techniques, particularly where these are used for making decisions about services provided to consumers.

Soufyane et. al (2020) The study will determine one dimension which has tested consumers' patience for two years now. In this regard, we will conduct a quantitative study of the most important relevant dimensions in view of understanding consumers' reactions by examining their preferences. They focus also on how consumer preferences differ for locally-assembled and imported cars and how importation trends evolve

Vinicious et. al (2020), The widespread impacts of artificial intelligence (AI) and machine learning (ML) in many segments of society have not yet been felt strongly in the marketing field. Despite such shortfall, ML offers a variety of potential benefits, including the opportunity to apply more robust methods for the generalization of scientific discoveries. Trying to reduce this shortfall, this paper has four goals.

Fatemeh Safara (2020), In this paper, a prediction model is proposed to anticipate the consumers behaviour using machine learning methods. Five individual classifiers, and their ensembles with Bagging and Boosting are examined on the dataset collected from an online shopping site. The results indicate the model constructed using decision tree ensembles with Bagging achieved the best prediction of consumer behavior with the accuracy of 95.3%. In addition, correlation analysis is performed to determine the most important features influencing the volume of online purchase during coronavirus pandemic.

Asnair et. al (2019) This paper tries to propose predictive analytics to predict customer behaviour by using behavior informatics and analytics approach so that deeper insight into customer behaviour can be obtained to support predictive analysis in order to improve business decision making.

Manpreet Singh et al (2018) The present work has been carried out in a churn prediction modeling context and benchmarks four machine learning techniques against a publicly available telecommunication dataset. The results provide two important conclusions: i) Random Forest technique outperforms other basic classification models and ii) Feature Engineering plays critical role in the performance of the model.

Rajan gupta et al.(2014) This study focuses more on inventory led e-commerce companies, however the model can be extended to online marketplaces without inventories. Facilitated by statistical and machine learning models the study seeks to predict the purchase decisions based on adaptive or dynamic pricing of a product.

Yutku et al. (2012), They are aiming to find alternative methods that can match or improve the recorded highest score with more efficient use of resources. Dataset has very large number of features, examples and incomplete values. As the first step, we employ some methods to preprocess the dataset for its imperfections. Then, we compare and contrast various ensemble and single classifiers. We conclude the paper with future directions for the study.

E-exponential commerce's growth has disrupted the overall buying experience and the conventional partnerships between the purchaser and companies. This transition comes with the obstacles firms must face. These problems mean greater rivalry and a turbulent relationship between clients and dealers, since consumers are no longer identified individually and their tastes are less loyal. The key goal of modern e-commerce therefore is to draw consumers, create trust and hold them (Nakayama, 2009; Salehi, Abdollahbeigi, Langroudi, & Salehi, 2012).

Visitors to the web store have left more footprints than ever. A great deal is gathered, linked and processed for review through data mining techniques with pro-Sonal knowledge as well as click Stream data collected at any visit to the site shop.

The information gained from these analyzes will boost customer loyalty by a more effective, more interactive and increasingly customizable shopping process (Magrabi, 2016). This can, in the long term, result in a higher version rate and improved sales advantage (Hop, 2013; Suchacka & Chodak, 2016).

Customer buying behaviour, which has been studied from the dawn of electronic trading (Bellman, Lohse & Johnson 1999), has been used for several days now. For example, consumer advice (Y. Tan, Xu & Liu, 2016; Hidasi, Quadrana, Karatzoglou, & Tikk, 2016); customer group ranking for shoppers, guests, etc. (Moe, 2003; Fajta, 2014). The purchasing probability is predicted so as to give customers with a greater likelihood of purchase a higher level of service (Lo, Frankowsik & Leskovec, 2014; Korpusik, 2014).

Conclusion

As consumers' behaviour in the form of e-commerce is becoming more relevant today due to their transition from visiting retail retailers to online shopping. It will improve customer loyalty and purchases by encouraging a more customized shopping experience, resulting in better transaction rates and a competitive edge. Models for forecasting consumer behaviour can be created by using clickstream and additional customer data. This research study explores model master learning to forecast a customer's purchase. Then, by carrying out the descriptive data analysis and individual training of the models on various datasets, this research would provide an insight into the output discrepancies in models in sequential clickstreams and static consumer data using artificial intelligence.

Future Scope

Some recommendations for future search result from the aforementioned limitations. First, any follow-up study should be about running the different algorithms on a computer more powerful than on a desktop computer locally. All algorithms offer parallelisation possibilities that can also be used on a better machine. This can already help build more powerful models, by using more data and performing more thorough tuning and training courses to exploit and analyze the potential of the algorithms fully. As far as the data is concerned, it would be interesting if it results in better predictions using other types of features, such as click information. A follow-up study could therefore focus on using the available data in more detail so as to see whether performance can also be improved through an improvement of the data base instead of adjusting and explore the potential of various algorithms. In addition, future studies should distinguish between sequential and static customer data between sequential and static data.

REFERENCES

1. Rajan gupta et.al (2014) , A Machine Learning Framework for Predicting Purchase by online customers based on Dynamic Pricing, procedia computer science pp. 599-605.

2. Stefano Puntoni et. al (2021), Consumers and Artificial Intelligence: An Experiential Perspective, American Marketing Association Journal of Marketing 2021, Vol. 85(1) 131-151

3. Figi Report (2021) on Big data, Machine learning , Consumer protection and privacy pp 1-62.

4. Asnair et. al (2019), Predictive Analytics for Predicting Customer Behavior, IEEE Conference in Indonasia, pp 1-4.

5. Soufyane et. al (2020), CONSUMER PREFERENCES AND FUTURE IMPACT ON THE ALGERIAN CAR MARKET, Academy of Marketing Studies Journal Volume 24, Issue 1, pp 1-10.

6. Vinicious et. al (2020), Machine learning in marketing: overview, learning strategies, applications, and future developments pp1-73.

7. Fatemeh Safara (2020), A Computational Model to Predict Consumer Behaviour During COVID-19 Pandemic, Computational Economics pp1-14.

8. Manpreet Singh et al. (2018) , Comparison of learning techniques for prediction of customer churn in telecommunication, IEEE, conference ITNAC, pp. 1-5

9. Yutku et al. (2012), Customer Churn Prediction for Telecom Services, IEEE Conference CSESA pp 1-2.

10. Pawan Whig, S N Ahmad and Anupam Priyam," Simulation & performance analysis of various R2R D/A converter using various topologies ", International Journal Robotics and Automation, Vol 2, Issue 1 , 2018 , pp. 128 -131. ISSN: 0826-8185.(Scopus)

11. Pawan Whig and S. N Ahmad ," Novel Pseudo PMOS Ultraviolet Photo Catalytic (PP-UVPCO) Sensor for Air Purification", International Journal Robotics and Automation, 2018.(Scopus)

12. Pawan Whig and S. N Ahmad ," Novel Photo Catalytic Sensor Output Calibration Technique", SSRG International Journal of VLSI & Signal Processing (SSRG-IJVSP), 2019,vol 6 no1, pp1-10

13. Pawan Whig and Ajay Rupani ," The development of big data science to the world", Engineering Reports, 2019,vol 2 no2, pp1-7

14. Pawan Whig ," Artificial intelligence and machine learning in business", Engineering Reports, 2019,vol 2 no2, pp8-13.

15. Khyati Kathuria and Pawan Whig ," Macroeconomic Implications of the Monetary Policy Committee Recommendations: An IS-LM Framework", ACTA SCIENTIFIC AGRICULTURE, 2020,vol 4 no 2, pp1-7

16. Pawan Whig ," Novel PCS Output Calibration Technique", SSRN, 2020,vol 11 no4, pp1-7

17. Pawan Whig and Ajay Rupani ," Novel Economical Social Distancing Smart Device for COVID19", International Journal of Electrical Engineering and Technology, 2020,vol 2 no2, pp1-10 (SCOPUS)

18. Veena Parihar, "Novel Algorithm for Text Based Personality Classification", 2nd International Conference on Emerging Trends in Engineering & Applied Science (ICETEAS' 19) ISSN: 2454-4248 Volume: 5 Issue: 1

19. C. J. (1998).A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery , 2(2), 121–167.

20. Castanedo, F., Valverde, G., Zaratiegui, J., & Vazquez, A. (2014). Using deep learning to predict customer churn in a mobile telecommunication network. wiseathena.

21. Chapman, P. (1999). The crisp-dm user guide. NCR Syst. Eng. Copenhagen. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth,

22. R. (2000). Crisp-dm 1.0 step-by-step data mining guides. NCR Syst. Eng. Copenhagen.

23. Chen, Z., Fu, A. W.-C., & Tong, C.-H. (2004). Optimal algorithms for finding user access sessions from very large web logs. World Wide Web, 6, 259–279.

24. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273 – 297.

25. Craven, M. (2011). Markov chain models. Retrieved from https://cw.fel.cvut.cz/wiki/ media/ courses/a6m33bin/markov-chains-2.pdf

26. Dewey, C. (2016). Markov chain models. Retrieved from https://www.biostat Dmc 2013. (2013). Retrieved from http://www.data-miningcup.de/en/review/ goto/article/dmc-2013.html

27. Dumais, S., Jeffries, R., Russell, D., Tang, D., & Teevan, J. (2014). Understanding user behavior through log data and analysis. New York: Springer.

28. Fajta, J. (2014). Online visitor classification based on mbti model (Unpublished master's thesis). Technische Universiteit Eindhoven.

29. Feng, J. (2017). gcforest github repository. Retrieved from https://github.com/ kingfengji/ gcForest

30. 30.Frank, E., Hall, M., & Witten, I. (2016). The weka workbench. online appendix for "data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.

31. Friedman, H. (2002). Random forests. Stochastic gradient boosting, 38(4). Gelman, A., & Hill, J. (2016). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

32. Goodman, B., & Flaxman, S. (2016). European union regulations on algorithmic decision-making and a right to explanation. In ICML Workshop on Human Interpretability in Machine Learning.

33. Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques.Elsevier.