AN EMPIRICAL ANALYSIS CUSTOMER PREDICTIVE LIFETIME VALUE MODELING WITH HYBRID MACHINE LEARNING

PATTAPU VENKATA SANDEEP, DR. MOORAMREDDY SREEDEVI,

MCA STUDENT, DEPT. OF COMPUTER SCIENCE, S.V.UNIVERSITY, TIRUPATI
SENIOR ASSISTANT PROFESSOR, DEPT. OF COMPUTER SCIENCE, S.V.UNIVERSITY, TIRUPATI

ABSTRACT

Customer lifetime value (CLV) has gotten expanding consideration in database marketing. Undertakings can hold important customers with the right expectation of significant customers. In the writing, numerous data mining and machine learning strategies have been applied to create CLV models. In particular, hybrid strategies have demonstrated their superiorities over single procedures. Be that as it may, it is obscure which hybrid model can play out the best in customer value expectation. This position venture investigates BDA in Predictive Analysis of Customers by drawing on a deliberate survey of the writing. The task presents an interpretive structure that investigates the definitional angles, particular attributes, types, business value, and difficulties of BDA in the Predictive Analysis of Customers scene. The Project additionally triggers more extensive conversations in regards to future research difficulties and openings in principle and practice. By and large, the discoveries of the examination integrate various BDA ideas (e.g., the meaning of enormous data, types, nature, business value, and important speculations) that furnish further experiences alongside the cross-cutting investigation applications in eCommerce. The test results over a genuine case dataset show that the arrangement þ grouping hybrid methodology plays out the best. Specifically, joining two-phase choice trees gives the most noteworthy pace of precision (99.73 percent) and the least pace of Type I/II mistakes (0.22 percent/0.43 percent). The commitment of this paper is to exhibit that hybrid machine learning procedures perform superior to single ones. What's more, this paper permits us to discover which hybrid method performs best as far as CLV expectation.

Keywords:

Customer lifetime value, Data mining, Machine learning, Hybrid models, Database marketing, Customer information.

I. INTRODUCTION

Prescient CLV demonstrating figures the value of a customer over their shopping lifetime dependent on their anticipated conduct. In contrast to authentic CLV demonstrating, it endeavors to represent varieties in customer conduct. Prescient CLV demonstrating is essentially more unpredictable than chronicled CLV displaying.

Machine Learning calculations are utilized in data mining applications to recover the shrouded information that might be utilized for acceptable dynamic. Machine learning contains different methods like guideline based learning, case-based thinking, counterfeit neural system, and choice tree. Each method has its own focal points and disservices. In the past parcel of hybrid machine learning frameworks were created to bring the best from the two diverse machine learning techniques. For instance, a hybrid machine learning framework is made dependent on hereditary calculation and bolster vector machines for financial exchange forecast by Rohit and Kumkum [8] Nerijis, Ignas, and Vida built up a hybrid machine learning approach for content classification utilizing choice trees and fake neural system [8]. Sankar built up a coordinated data mining approach for upkeep booking utilizing case-based thinking and fake neural system [9] and Mammone built up a hybrid machine learning framework joining neural system and choice tree. In our exploration, we are utilizing a counterfeit neural system with case-based thinking strategies. The counterfeit neural system gives great grouping exactness contrasted with other machine learning methods like standard based and choice trees. Through this hybrid machine learning, we need to improve the order exactness of the neural system framework that might be utilized for clinical analysis.

.

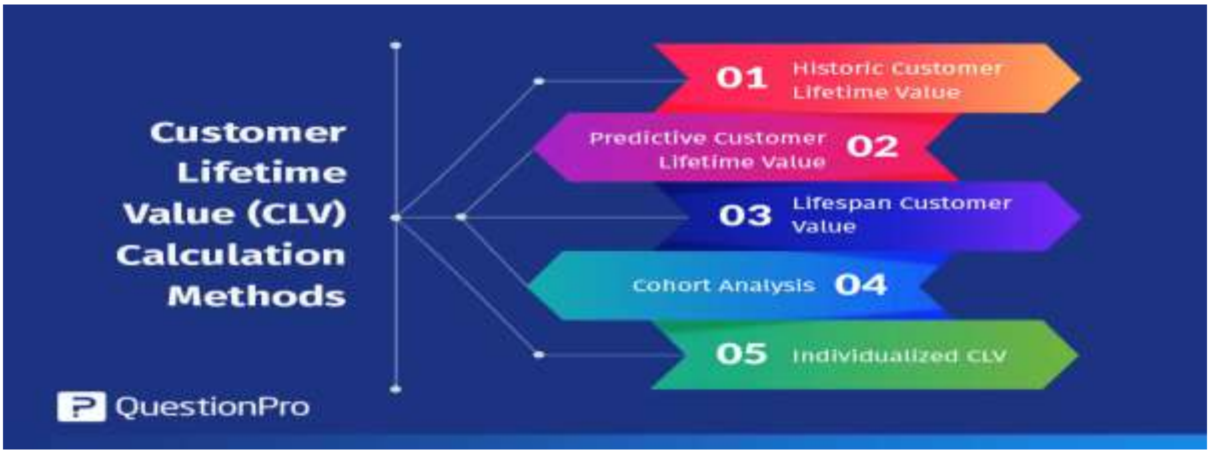


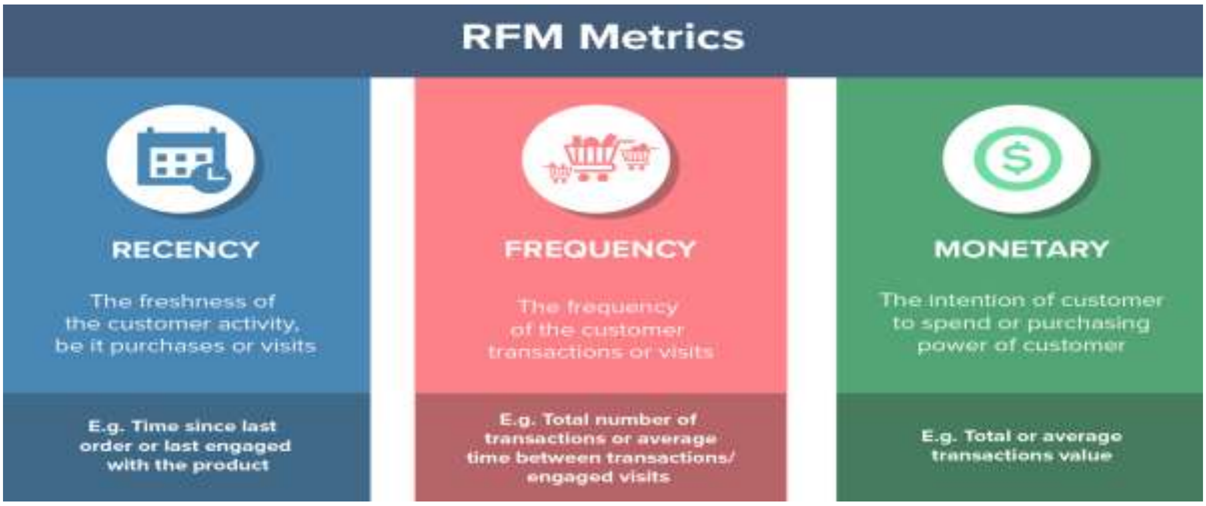
FIG : **Predictive** CLV **modeling**

Especially, a genuine customer exchange dataset was gathered from a main producer of treated steel tubes in Taiwan. We construct CLV expectation models from the dataset dependent on the notable recency, recurrence, and financial (RFM) examination (Hughes, 1997, 2005). Besides, since it is obscure which hybrid model performs better, this investigation likewise directs a relative report dependent on two-hybrid methodologies: order þ arrangement and bunching þ grouping. In particular, this investigation chooses choice trees (DT), calculated relapse (LR), and multi-layer perceptron (MLP) as the characterization procedures; and the k-implies and oneself sorting out maps (SOM) as the grouping methods.

II. RELATED WORK

2.1 The RFM model

RFM investigation numerically positions a customer in every one of these three classifications, by and large on a size of 1 to 5 (the higher the number, the better the outcome). The "best" customer would get a top score in each class.



RFM represents Recency, Frequency, and Monetary value, each relating to some key customer attribute. These RFM measurements are significant pointers of a customer's conduct in light of the fact that the recurrence and fiscal value influence a customer's lifetime value, and recency influences maintenance, a proportion of commitment Businesses that come up short on the money related perspective, similar to viewership, readership, or surfing-focused items, could utilize Engagement parameters rather than Monetary components. This outcomes in utilizing RFE – a variety of RFM. Moreover, this Engagement parameter could be characterized as a composite value dependent on measurements, for example, ricochet rate, visit length, number of pages visited, time spent per page, and so forth.

RFM factors represent these realities:

- the later the buy, the more responsive the customer is to advancements
- the all the more as often as possible the customer purchases, the more connected with and fulfilled they are
- monetary value separates squanderers from low-value buyers

III. PROPOSAL WORK

3.1 CLV forecast models

To develop the expectation models, the information factors depend on the RFM model, which are later, recurrence, and financial values. As indicated by Koch (1998), this investigation characterizes the main 20 percent RFM values as the high-value customers ("1"), and the others are low significant customers ("0") for the yield factors.

Three single classifiers, DT, LR, and MLP, are developed as the pattern models. The ten times cross-approval strategy is received for developing the preparation and testing datasets in all investigations to take care of the issue of the single inspecting on dataset causing the non-uniform circulation of the data recovery and influencing the accuracy of the arrangement. What's more, the parameters to prepare the MLP model incorporate the quantities of concealed hubs (8, 16, 32, and 64) and the learning ages (50, 100, 200, and 400).

The idea of hybrid CLV forecast models in this investigation comprises of two segments. The main part leads any of the bunching or order methods referenced in Section 2 and the second chooses a particular arrangement procedure to assemble the expectation model. In particular, the primary segment targets sifting through unrepresentative data from guaranteed (preparing) dataset. That is, the data which can't be grouped or arranged precisely by the principal part can be viewed as loud data. At that point, the remainder of the delegate data can be utilized to prepare the classifier so as to improve the characterization result.

3.2 CLASSIFICATION OF CLASSIFICATION HYBRID MODELS.

For the hybrid models by consolidating two-phase order procedures, the accurately anticipated data from the best gauge arrangement model over a given preparing set are utilized to prepare the three single characterization models exclusively. As well as can be expected be distinguished subsequent to evaluating the three single classifiers. Note that the effectively anticipated data depend on the preparation set of ten times cross-approval as opposed to the testing set. Accordingly, there are three hybrid models created, including:

(1) the best gauge model þ DT;

(2) the best gauge model þ LR; and

(3) the best gauge model þ MLP.

3.3 Clustering þ grouping hybrid models.

To join grouping and arrangement procedures, k-means and SOM are inspected separately. In particular, the value of k in k-implies is set from 2 to 5 and the guide size of SOM is set by 2 £ 2, 3 £ 2, 3 £ 3, and 3 £ 4, individually. Next, the bunching result with the most noteworthy precision by ten times cross-approval is chosen as the best grouping model. At that point the best grouping outcome (for example the precisely grouped data) by the best benchmark bunching model is utilized to prepare the three single standard models independently.

For the case of 2 £ 2 SOM, four groups are delivered spoken to by C1, C2, C3, and C4 dependent on a preparation set. As indicated by the ground truth answer of the preparation set, we can discover two out of the four bunches, which can be well "arranged" into the high value and low-value gatherings, separately. The other two bunches whose data are not very much grouped or hard to be characterized by 2 £

2 SOM are not chosen for classifier preparing.

When the best SOM is recognized, its grouped data (for example the grouping result) are utilized to prepare the three single classifiers, individually. It ought to be noticed that one explicit SOM model over the ten diverse preparing sets (by ten times cross-approval) will deliver ten distinctive grouping results, separately. That is, the data in the two delegate bunches, which can all the more likely perceive the high value and low-value bunches utilizing the ten preparing sets, are non-copied.

Conclusion

HML procedures have indicated promising outcomes in related writing. The examination target of this paper is to look at which hybrid methodology and model can perform better in the space of CLV Prediction. This paper evaluates the presentation of two significant HML approaches over a genuine case dataset. That is, arrangement þ grouping and bunching þ order hybrid models are developed for correlations. The test results show that the characterization þ arrangement hybrid methodology performs better than the grouping þ order hybrid one. In particular, in CLV expectation, utilizing DT in the principal stage to sift through unrepresentative data for the second stage classifier dependent on another DT can give the most elevated forecast exactness rate and least Type I/II mistake rates. Be that as it may, the discoveries suggest that the decision of grouping and arrangement strategies to build a hybrid model is basic. That is, only one out of every odd blend of various procedures dependent on the grouping þ order hybrid methodology can perform better than the bunching þ characterization hybrid one and single arrangement models also. In spite of the fact that our investigation shows that HML increases promising outcomes on CLV forecast, there exist some significant impediments that merit thought. To start with, this investigation inspects a few HML procedures on just a single genuine customer exchange dataset. Different datasets identified with CLV can be incorporated for future correlations. Second, the RFM investigation was considered as the premise of CLV appraisal. Some other eminent CLV models could be directed for additional assessments. Third, all the more developing order or grouping methods, for example, fluffy c-mean and bolster vector machines, can be explored to experimentally think about their exhibition.

References

1. Berger, P.D. and Nasr, N.I. (1998), “Customer lifetime value: marketing models and applications”, Journal of Interactive Marketing, Vol. 12 No. 1, pp. 17-29.

2. Blackwell, R.D., Miniard, P.W. and Engel, J.F. (2005), Consumer Behaviour, 10th ed., South-Western College Pub, Cincinnati, OH.

3. Bose, I. and Chen, X. (2009), “Hybrid models using unsupervised clustering for prediction of customer churn”, Journal of Organizational Computing and Electronic Commerce, Vol. 19 No. 2, pp. 133-151.

4. Butz, H.E. Jr and Goodstein, L.D. (1996), “Measuring of customer value: gaining the strategic advantage”, Organizational Dynamics, Vol. 24 No. 3, pp. 63-77.

5. Cheng, C.-H. and Chen, Y.-S. (2009), “Classifying the segmentation of customer value via RFM model and RS theory”, Expert Systems with Applications, Vol. 36 No. 3, pp. 4176-4184.

6. Chu, B.-H., Tsai, M.-S. and Ho, C.-S. (2007), “Toward a hybrid data mining model for customer retention”, Knowledge-Based Systems, Vol. 20 No. 8, pp. 703-718.

7. Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N. and Sriram, S. (2006), “Modeling customer lifetime value”, Journal of Service Research, Vol. 9 No. 2, pp. 139-155.

8. Hartigan, J.A. and Wong, M.A. (1979), “A K-means clustering algorithm”, Applied Statistics, Vol. 28 No. 1, pp. 100-108.

9. Haykin, S. (1999), Neural Networks: A Comprehensive Foundation, 2nd ed., Prentice-Hall, Upper Saddle River, NJ.

10. Hosmer, D.W. and Lemeshow, S. (2000), Applied Logistic Regression, Wiley, New York, NY.

11. Hruschka, H., Fettes, W. and Probst, M. (2004), “Market segmentation by maximum likelihood clustering using choice elasticities”, European Journal of Operational Research, Vol. 154 No. 3, pp. 779-786.

12. Hsieh, N.-C. (2004), “An integrated data mining and behavioural scoring model for analyzing bank customers”, Expert Systems with Applications, Vol. 27 No. 4, pp. 623-633.

13. Huang, S.C., Chang, E.C. and Wu, H.H. (2009), “A case study of applying data mining techniques in an outfitter’s customer value analysis”, Expert System Application, Vol. 36 No. 3, pp. 5909-5915.

14. Hughes, A.M. (1997), “Customer retention: integrating lifetime value into marketing strategies”, Journal of Database Marketing, Vol. 5 No. 2, pp. 171-178.

15. Hughes, A.M. (2005), Strategic Database Marketing: The Master Plan for Starting and Managing a Profitable, Customer-based

Marketing Program, 3rd ed., McGraw-Hill Companies, New York, NY.

16. Hung, C. and Tsai, C.-F. (2008), “Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand”, Expert Systems with Applications, Vol. 34 No. 1, pp. 780-787.

17. Hwang, H., Jung, T. and Suh, E. (2004), “An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry”, Expert Systems with Applications, Vol. 26 No. 2, pp. 181-188.

18. Jain, A.K., Murty, M.N. and Flynn, P.J. (1999), “Data clustering: a review”, ACM Computing Survey, Vol. 31 No. 3, pp. 264-323.

19. Kim, S.-Y., Jung, T.-S., Suh, E.-H. and Hwang, H.-S. (2006), “Customer segmentation and strategy development based on customer lifetime value: a case study”, Expert Systems with Applications, Vol. 31 No. 1, pp. 101-107.

20. Koch, R. (1998), The 80/20 Principle, Nicholas Brealey Publishing, Boston, MA.

21. Kohonen, T. (2000), Self-organizing Maps, 3rd ed., Springer, Berlin.

22. Kumar, V. and Reinartz, W.J. (2005), Customer Relationship Management: A Databased Approach, Wiley, New York, NY.

23. Lenard, M.J., Madey, G.R. and Alam, P. (1998), “The design and validation of a hybrid information system for the auditor’s going concern decision”, Journal of Management Information Systems, Vol. 14 No. 4, pp. 219-237.

24. Liu, D.-R. and Shih, Y.-Y. (2005a), “Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences”, Journal of Systems & Software, Vol. 77 No. 2, pp. 181-191.

AUTHOR PROFILE

PATTAPU VENKATA SANDEEP received Bachelor of Computer Science degree from Adi Kavi Nannaya University, East Godavari in the year of 2014-2017. Pursuing Master of Computer Applications from Sri Venkateswara University, Tirupati in the year of 2017-2020. Research interest in the field of ML BASED HUMAN ACTIVITY RECOGNITION WITH SMARTPHONE'S FOR HEALTHCARE

Dr. Mooramreddy Sreedevi, has working as a Senior Assistant Professor in the Dept. of Computer Science, S.V.University, Tirupati since 2007. She obtained her Ph.D. Computer Science from S.V.University, Tirupati. She acted as a Deputy Warden for women for 4 years and also acted as a Lady Representative for 2years in SVU Teachers Association, S.V.University, Tirupati. She published 56 research papers in UGC reputed journals, Participated in 30 International Conferences and 50 National conferences. She acted as a Resource person for different universities. Her current research focuses in the areas of Network Security, Data Mining, Cloud Computing and Big data analytics.