Predictive Modeling for Real-Time Customer Lifetime Value

Lalitha S., Taduvai Satvik Gupta, Aditya Thelu, Vontela Kartheek Reddy, Vubbara Chaitri Reddy
Department of Electronics and Communication Engineering, Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
s_lalitha@blr.amrita.edu , satviktaduvai@gmail.com,official.thelu@gmail.com,kartheekv15@gmail.com, vubbarachaitrireddy@gmail.com

Abstract— Customer Lifetime Value (CLV or CLTV) is the measurement of the overall net revenue of a business that may be expected from a customer. Unlike the previous studies involving segmenting customers and forecasting CLV, the focus of this work is to predict the CLV of the new customer coming in. Further, this work has been performed on the data set of an online store of a gifting company. The proposed model is based on Probabilistic models viz BG/NBD, Pareto/NBD, and Regression models viz KNN regression, decision tree, linear regression, and random forest to predict the CLV. From the experimental results, it is demonstrated that this framework builds a more optimal CLV model compared to the previous model showing an improvement of approximately 10%.

Keywords— CLV, RFM, BG/NBD, SVM, Random Forest, Decision Tree, Regression, Pareto/NBD

I. Introduction

In the post-pandemic era, most people prefer to shop online, increasing the demand for predicting the client's Customer lifetime value (CLV or CLTV). It measures the overall net income a business may expect from a consumer during their connection. It considers the consumer's initial pay, any further purchases, and the mean length of time they have been associated with the business. [1] CLV tells the marketers how much revenue a company can expect from individual customers which helps them to improve their business. After predicting the CLV we can also cluster the customers which helps the marketers to focus on the set of customers who can improve their business and enables more informed marketing and resource allocation strategies.[2] CLV also helps marketers improve their marketing and retention strategies.

The applications of CLV span diverse industries and business functions. Precise CLV predictions enable businesses to design targeted marketing campaigns, personalize products and services, and implement effective customer retention strategies. By integrating RFM analysis, companies can understand the recency, frequency, and monetary aspects of customer behavior, allowing for data-driven decision-making in sectors such as pricing optimization, resource allocation, and the design of customer loyalty programs.[3]

This work aims to improve the CLV prediction by integrating probabilistic and regression models. The motivation for this work Predicting CLV using Regression and Probabilistic Models is that it helps to find the CLV for the new customers also, and understanding CLV prediction can help companies improve marketing efforts, manage resources effectively, and build client loyalty. It helps to make data-driven decisions by providing insights into customer behavior. It enhances long-term profitability by nurturing high-value Customer relationships. It empowers businesses to offer personalized experiences and gain a competitive edge in customer satisfaction. It also helps to optimize their marketing.

The approach for forecasting the CLV involves first visualizing the dataset using various graphs to comprehend the data, and then preparing the dataset for processing, and RFM values are estimated. The CLV is predicted using probabilistic models. Then, the predicted CLV values are given to the regression models and trained. If a new sample comes the model will be able to predict the CLV.

II. Literature Survey

This section discusses the key research works related to CLV and is outlined as follows:

Yuechi Sun et al. developed a client value segmentation model using different models such as BG/NBD, Naïve Bayes, KNN, SVM, etc. Among the total customers 29% of customers with Important Value Customers, 26% of the customers are Important Development Customers and the remaining are General Customers [4].

The assessment of machine learning techniques for CLV prediction was the primary thrust of Ravi Kant et al. The study assesses and contrasts the performance of several well-known machine learning algorithms[5]. It finds that the algorithms—decision tree and random forest, in particular—performed well. [6].

Bernar Taşcı et al. study assessed the method’s efficacy by developing several models. Positive outcomes from the created prediction model prevented about 42% of actual manufacturing line issues. [7].

Marius Myburg et.al suggested a method for CLV segmentation using the RFM model. XG-Boost and K-means clustering have been used to segment the consumers with the RFM values. The method predicted CLV, with 77–78% accuracy [8].

Asiman Mammadzada et al. suggested an approach to find the CLV utilizing Gamma-Gamma Distribution, and BG/NBD, and used RFM analysis to predict the customer value of a bank. The model is trained with the data of 6 months and the next 15 days transactions are predicted. They obtained an accuracy of 92.1% [9].

Shithi Maitra et al. proposed a CLV classifier for retail businesses. This technique leverages TensorFlow-coded neural network (NN) with PostgreSQL to extract customer behaviour and utilize it as an input feature to predict profitability [10].

The empirical characteristics of CLV data in mobile games are analyzed by Arturo Valdivia et al. This technique implies that techniques such as a Multinomial-Dirichlet Bayesian model could be used to investigate features in more detail[11][12].

Rininta Rahmadianti et al. proposed a methodology that combines the k-means algorithm for customer clustering and the AHP method to calculate CLV. They identify eight customer clusters, with clusters 3, 6, and 8 representing high-value loyal customers [13].

A Segmentation method using Fuzzy-AHP and RFM analysis is suggested by Anu Gupta Aggarwal et al. For clustering, they used a k-mean clustering algorithm. Finally, the Customers are divided into 8 Clusters based on their Recent transactions, Frequency of purchases, and Monetary value [14].

Than Than Win1 et.al suggested a model that helps to predict the class of the consumer for the next year based on their CLV. The tree-based algorithm - random forest is used to educate the model. The accuracy of the model is 82.26% [15].

Yaser Hasanpour et al. suggested a clustering method using Fuzzy C-means Clustering and CRISP-DM data mining methods to preprocess the dataset. Customers are divided into 4 clusters [16].

Lavneet Singh et al. presented a mathematical model framework that combined RANSAC Response Regularization with Gradient Boost Trees. This method gives an accuracy of 80% with 0.12 MSE [17].

The key observations from the literature review include the regression method for estimating CLV and RFM, extended RFM analysis, and K-mean clustering for customer clustering. The most used dataset is the Real-time Online Retail dataset of an American company from the UCI Machine Learning Repository and the IBM Watson dataset from Kaggle. The features of both these datasets are completely different. So, the model build for one dataset cannot be compared with the other dataset. The evolution metrics accuracy, RMSE, and MSE are frequently utilized to determine the model's performance.

A few research gaps observed from the literature survey are that: the prediction performance of the model is not reported to estimate the CLV for a new test sample coming in. The Preto/NBD model is the advanced Probabilistic model in Predicting the CLV. Division of the Customers based on the CLV and RFM Values is not done. In clustering, the focus has been only on RFM values. This work focuses on predicting the CLV for the new test sample coming in and the use of the Pareto/NBD probabilistic model.

III. Proposed Work

In the development of the proposed model, a systematic step-by-step approach is followed as shown in Fig.1.

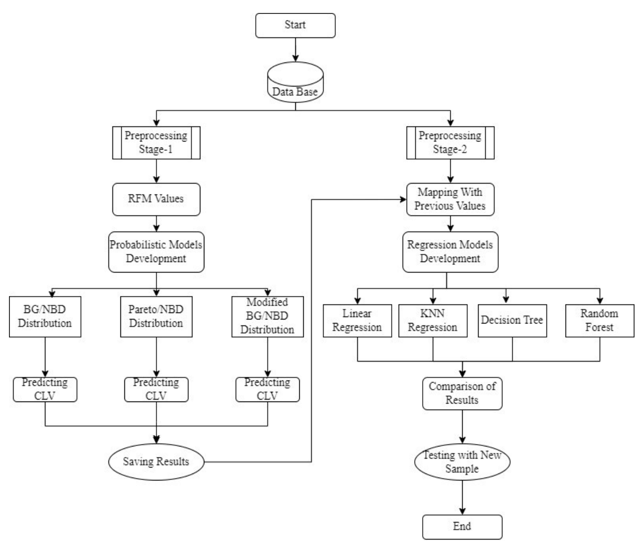

Fig.1: Flowchart of the Proposed Model

A. Data Base:

The UCI Machine Learning Repository provides the source of data for the proposed model. It is the data on the online retail sales of a US-based Company. It has transactions of 2 Financial Years. The data set contains 8 attributes specifically Stock Code, Invoice No, Quantity, Description, Date of Purchase, Price, Customer ID, and Country. It has a total of 1067371 samples and 44876 unique purchases.

B. Preprocessing (Stage-1):

The first stage of preprocessing involves the removal of null values and grouping the customers according to the Customer ID. After the Preprocessing using the Probabilistic models, the CLV is estimated using 3 different distributions and R, F, M, T values for each consumer.

C. RFM Values:

Recency denotes how long it has been since a consumer completed the transaction. Frequency indicates the regularity of a customer. The monetary worth of a customer's purchases is represented by monetary. Tenure denotes the time between the initial purchase and the final purchase.

D. Probabilistic Models Development:

The purpose of these models is to estimate the approximate overall monetary value of all the transactions a customer does in their lifetime. The dataset is divided into the calibration and holdout periods to measure the CLV and compare the performance of the model. This work uses 20 months as the calibration period and 6 months of data as the holdout period.

i. BG/NBD Distribution:

It is an Integrated probabilistic model that explains the purchasing and churning behavior of the customer. It considers the probability distribution of the behavior of the customers.

$E(Y(t) \mid X = x, tx, T, r, \alpha, a, b) = ((a+b+x-1) / (\alpha - 1)) \times (P/Q)$

$P = (1 - (\alpha+T/ \alpha+T+t)^{r+x}\ {}_2F_1(r+x, b+x; a+b+x-1))$

$Q = 1+\delta_{x>0} (a/b+x-1) (\alpha+T/ \alpha+t_x)^{r+x}$

where: (1)

- x is Customer Frequency,
- tx is Recency Value,
- T is Tenure,
- R and α Difference in rate of the transaction between consumer parameters of the gamma distribution,
- a and b are BG Parameters representing the rate of drop.
- ${}_2F_1$ is Kummer's confluent hypergeometric function.

ii. Pareto/NBD Distribution:

It differs in that a consumer can only leave immediately following a transaction. This streamlines calculation, but it has the disadvantage that a consumer cannot leave until a transaction is completed.

$E(T_x) = (r/\alpha+s) \times (s+T+\beta^{-1} - (s+t_x+\beta^{-1}) \times \exp(V)$

$V = (-(\alpha+\beta)T) \times {}_2F_1(\alpha+\beta, \alpha+\beta; (\alpha+\beta)(T+t)))$

where: (2)

- α, β and s are Pareto/NBD parameters,
- r is transaction rate.

iii. Modified BG/NBD Distribution:

It uses the Gamma distribution as the base for this distribution. First, the Gamma filter is used, and the result of this is given to the BG/NBD model[18].

E. Preprocessing (Stage-2):

In the second Preprocessing the grouping is done according to the Customer ID and the Multiple Invoice numbers are merged by adding the price of the products. Then, the data set is transposed into Customer ID, followed by the 26 months of the transactions. The previously calculated CLV will now be mapped to the transposed data according to the Customer ID.

F. Regression Models:

In this, the regression models are fit into the past data and predicted CLV. Using different regression models viz, KNN regression, linear regression, random forest, and decision tree regression the models have been trained.

G. Testing with a New Sample:

If a new test Sample comes in with 26 months of transactional data, the model will be able to predict the CLV.

IV. Evaluation Measures

The evaluation metrics used in this work to measure the model's performance are MSE, RMSE, average purchase error, and Accuracy. Accuracy is measured using 10-fold Cross-Validation[19].

$MSE = (1/n)*\sum(\text{actual} - \text{Forecasted})^2$ (3)

$RMSE = ((1/n)*\sum(\text{actual} - \text{Forecasted})^2)^{0.5}$ (4)

Avg Purchase Error = Actual – Predicted purchases (5)

$\text{Accuracy} = \frac{No.of\ correct\ predictions}{Total\ no.of\ predictions}$ (6)

V. Implementation and Results

Python 3 has been used to implement the proposed work. Stage -1 of preprocessing the data involves eliminating null values and identifying the dataset's insights. The dataset's insights are: Most of the customers are from the United Kingdom followed by Germany, EIRE & France. It is observed that Singapore has the highest average price sale followed by Norway and Malta. Most of the country's data is skewed regarding the Price feature with many higher extreme values. There are a total of 44876 unique purchases that happened. 2010 is the year in which the most transactions followed by 2011.

The first stage of preprocessing is done as shown in Table 1. As shown in the table null values have been removed and data is grouped according to the unique ID given to each customer.

Table 1: First Stage of Preprocessing

Invoice No	Stock Code	Description	Quantity	Invoice Date	Price	Customer ID	Country
489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01	6.95	**13085.0**	United Kingdom
489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01	6.75	**13085.0**	United Kingdom
489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01	6.75	**13085.0**	United Kingdom
489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01	2.10	**13085.0**	United Kingdom

Subsequently, the R, F, M, and T values for each consumer were found as shown in Table 2. These values are in accordance with the first 5 customers.

Table 2 shows the RFM values of the customers which are needed for the calculation of the CLV. It can be observed that there are customers with 0 recency and frequency and with a negative monetary value and 5942 is obtained to be the highest value in frequency, recency, and monetary value by which we can say that there are mixed types of customers.

Table 2 shows the RFM values of the customers which are needed for the calculation of the CLV. It can be observed that

there are customers with 0 recency and frequency and with a negative monetary value and 5942 is obtained to be the highest value in frequency, recency, and monetary value by which we can say that there are mixed types of customers.

Table 2: RFMT Values of each Customer

Customer ID	frequency	recency	T	monetary value
12346.0	10.0	400.0	725.0	-15.468000
12347.0	7.0	402.0	404.0	717.398571
12348.0	4.0	363.0	438.0	449.310000
12349.0	4.0	717.0	735.0	1107.172500

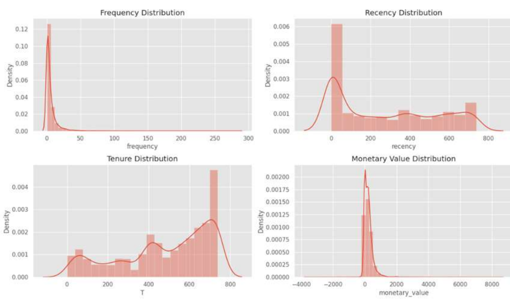

Fig.2: Distribution of RFMT Values

Fig.2 shows the distribution of the RFM values among the different customers. From the graph, it can be observed that the frequency of the customers is between 0 to 50, recency between 0 to 600, tenure is between 0 to 700 and monetary value is between 0 and 2000.Then move to the development of the First probabilistic models. Based on the transactional dataset available and RFM values of the customers applied to the BG filter the results are: a: 0.15, α:49.94, b: 2.11, r: 0.67. The estimated parameters are used to display the likelihood of a client being alive. Fig.3 shows the heat map of the probability of the customer being alive based on the R and F of the consumer.

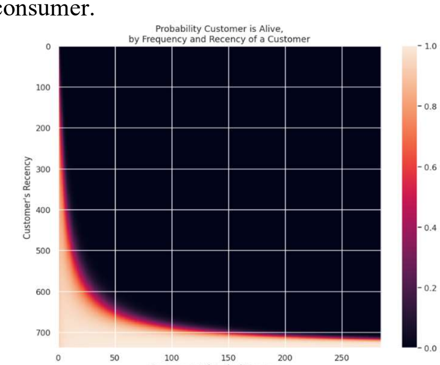

Fig.3: Probability of Customer being alive using BG/NBD Model

From fig 3 we can say that customers whose frequency is up to 250 and recency is greater than 250 have a high probability of being alive. And 80% of the customers have already churned. The data set is split into calibration and holdout categories so that the performance of the model may be observed, and the purchases are predicted for the holdout period. The calibration period is 20 months and the next 6 months purchases are predicted. Fig.4 shows the graph between the predicted Vs actual purchases in the holdout period calculated using the BG/NBD model After that, the Second Probabilistic model was developed. Based on the transactional dataset available and RFM values of the customers applied to the Pareto NBD filter the results are: α: 63.88, β: 124.23, r: 0.83, s: 0.16. Fig.5 shows the heat map of the probability of the customer being alive based on the R and F of the consumer.

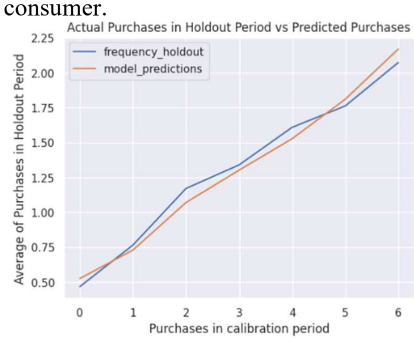

Fig.4: Actual vs Predicted Purchases in the holdout period using BG/NBD Model

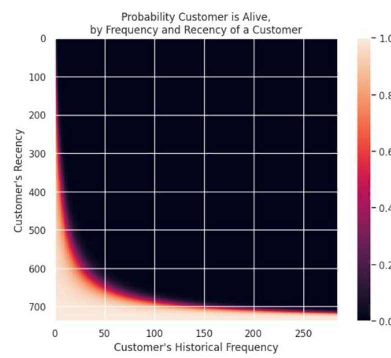

Fig.5: Probability of Customer being alive using the Pareto/NBD Model

From Fig 5 it is observed that customers with a frequency of 50 to 250 and recency with more than 250 have a high probability of being alive.

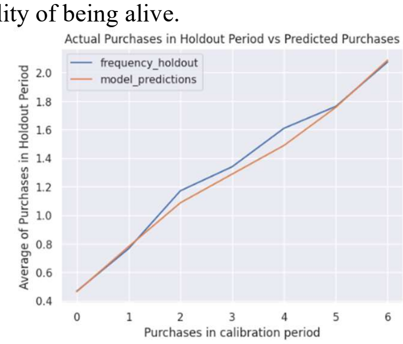

Fig.6: Actual vs Predicted Purchases in the holdout period using the Pareto/NBD Model

The graph between the expected and actual purchases during the holdout period, which was produced using the Pareto/NBD model, is displayed in Fig. 6. Comparing Fig 4 and 6 in Fig 6 it is observed that the model performance to predict the number of purchases has been improved. After that, the third Probabilistic Model was developed. It takes the base of the first model and adds a Gamma-Gamma Filter to the data. The results are: a: 0.18, alpha: 57.90, b: 2.05, r: 0.84

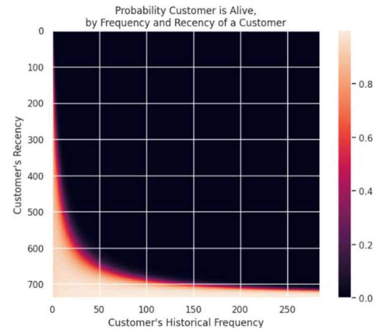

Fig.7: Probability of Customer being alive using Modified BG/NBD Model

Fig.7 shows the heat map of the probability of the customer being alive based on the R and F of the consumer. From Fig 4 it is observed that the customers with a frequency of 50 to 200 and recency with more than 300 have a high probability of being alive.

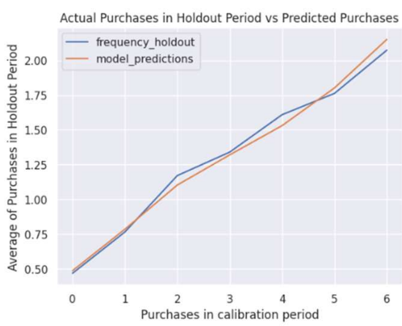

Fig.8: Actual vs Predicted Purchases in the holdout period using the Modified BG/NBD Model

Fig.8 shows the graph between the predicted Vs actual purchases in the holdout period. Comparing Fig 4, 6, and 8 in Fig 8 it is noted that the model's ability to forecast the volume of purchases had increased. So, we can say that the modified BG/NBD is predicting the purchases more accurately compared with the other 2 models. The CLV results use all 3 models and the evaluation metrics used are MSE, RMSE, and Avg purchase error.

Table 3: CLV Values Estimated

Customer ID	BG-NBD	Pareto-NBD	MBG-NBD
12347.0	7110.166375	6427.083703	7003.290033
12348.0	2512.081673	2243.170549	2522.287420
12349.0	3449.798774	3184.476671	3504.589499
12352.0	3100.010138	2760.845696	3031.895454

Table 4: Errors in the Models

	BG-NBD	Pareto-NBD	MBG-NBD
MSE Purchase Error	4.337883	4.335935	4.346083
RMSE Purchase Error	2.082758	2.082291	2.084726
Avg Purchase Error	0.411798	0.412367	0.417090

Table 4 shows the errors in the probabilistic models. MBG-NBD has less Avg purchase error. It is observed that there is not much difference when it comes to the performance of these models, but the Pareto NBD model is slightly better when it comes it minimizing the MSE & RMSE Errors.

Table 5: Second Stage of Preprocessing

MONTH_BY_YEAR	Customer ID	Apr-2010	...	Sep-2011	CLV
1	12347.0	0.00	...	0.0	7110.1663
2	12348.0	0.00	...	310.0	2512.0816
3	12349.0	1068.52	...	0.0	3449.7987
6	12352.0	0.00	...	632.5	3100.0101

In the second stage of the preprocessing, the dataset is processed as such to have the customer ID and 25 months of the transaction, and then the previously found CLV values are mapped to this data. Table 5 contains 29 columns (SI.No, Customer ID, 26-month transaction value, and CLV) showing the results after the second stage of preprocessing. After using the Supervised Regression Techniques, the model is trained, and the performance of the models has been shown below, to measure the accuracy of the models 10-fold cross-validation has been used and the RMSE values are used to find the error of the prediction. Table 6 shows the accuracy of the different regression models. Linear Regression has a high accuracy of 93.4% and decision tree has a low accuracy of 68.7%.

Table 6: Accuracy of different Regression Models

Algorithm	Accuracy
Linear Regression	93.418005
Random Forest (Estimators=150)	86.573666
Random Forest (Estimators=300)	86.562072
Random Forest (Estimators=250)	86.477140
KNN Regression(K=3)	78.822651
KNN Regression (K=2)	78.684865
KNN Regression (K=1)	78.404233
KNN Regression (K=4)	78.195173
Decision Tree	68.688595

Table 7: RMSE of different Regression Models

Algorithm	Root Mean Square Error
Linear Regression	3341.156856
Random Forest (Estimators=300)	5036.125174
Random Forest (Estimators=250)	5138.666398
Random Forest (Estimators=150)	5269.438340

Algorithm	Root Mean Square Error
KNN Regression (K=1)	5871.777964
KNN Regression (K=2)	6581.648449
KNN Regression (K=4)	6989.894130
KNN Regression (K=3)	7094.328250
Decision Tree	11514.357774

Table 7 shows the errors in the regression models. Linear regression has a low error of 3341 and the decision tree regressor has a high error of 11514. Linear Regression is observed as the best model having high accuracy and low error. Now that a new sample has been included with the transaction data of 25 months the model will be able to predict its CLV.

```
a=22251; b=6900; c=18851; d=3869; e=6448; f=3541;
new_data = np.array([a,b,c,0,0,d,0,e,f,a,b,c,0,d,e,0,f,a,b,0,0,c,d,e,f]).reshape(1,-1)
print("The CLV of the Customer is between " '${:,.0f}'.format(Lower),"and",'${:,.0f}.'.format
    "The average value the CLV is " '${:,.0f}.'.format(ExpectedValue))

The CLV of the Customer is between $306,357 and $323,570. The average value the CLV is $314,9
```

Fig.9: CLV for the new test sample.

The CLV range for the new test sample is displayed in Fig. 9, along with the average expected value of the CLV predicted by the linear regression model—which is the optimal model for CLV prediction due to its high accuracy and low error.

VI. Conclusion And Future Scope

Predicting Customer Lifetime worth is an essential challenge for companies looking to identify and optimize the worth of their clientele. A variety of methods are available for estimating the future value a customer will generate when using regression and probabilistic Models in CLV prediction. The work demonstrates machine learning's potential in Predicting the CLV particularly using Probabilistic and Regression Models and finding the CLV value for the new test sample. It also uses the Pareto/NBD model, one of the best probabilistic distributions. It helps companies to improve their marketing strategies and concentrate on high-value customers. In day-to-day usage the CLV has various applications in marketing(campaigns, segmentation), sales(prioritization, resource allocation), customer service(service differentiation, retention rate), product development, finance(investment decisions), strategic planning(partnerships and alliances) and operations(inventory management and experience management). The future scope for CLV prediction using regression, probabilistic Models, and analysis is future developments may involve the integration of more advanced machine learning algorithms. Incorporating a wider array of data sources beyond transaction histories and demographics could provide a more comprehensive understanding of customer behavior, further refining CLV predictions. And finding the months that are influencing the CLV we can predict the CLV using the data of those specific months.

References

[1] Wikipedia contributors. "Customer lifetime value." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 14 Aug. 2023. Web. 8 Oct. 2023.

[2] H. Kailash, K. Kanwar, S. Sonia and R. Kant, "Machine Learning Algorithms for Predicting Customers' Lifetime Value: A Systematic Evaluation," 2023 3rd ICACITE, Greater Noida, India, 2023, pp. 538-541

[3] Paul D. Berger, Nada I. Nasr Customer lifetime value: Marketing models and applications, Journal of Interactive Marketing, Volume 12, Issue 1,1998

[4] Yuechi Sun, Haiyan Liu, Yu Gao, Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model, Heliyon, Volume 9, Issue 2,2023,e13384, ISSN 2405-8440

[5] M. Surti, V. Shah, S. Bharti and R. Gupta, "Customer Lifetime Value Prediction of an Insurance Company using Regression Models," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-6.

[6] H. Kailash, K. Kanwar, S. Sonia and R. Kant, "Machine Learning Algorithms for Predicting Customers' Lifetime Value: A Systematic Evaluation," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 538-541.

[7] Bernar Taşcı, Ammar Omar, Serkan Ayvaz, Remaining useful lifetime prediction for predictive maintenance in manufacturing, Computers & Industrial Engineering, Volume 184,2023,109566, ISSN 0360-8352.

[8] M. Myburg and S. Berman, "Customer Lifetime Value Prediction with K-means Clustering and XGBoost," 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Istanbul, Turkey, 2022, pp. 298-302.

[9] Mammadzada, E. Alasgarov and A. Mammadov, "Application of BG / NBD and Gamma-Gamma Models to Predict Customer Lifetime Value for Financial Institution," 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, 2021, pp. 1-6.

[10] S. Maitra, M. Rakib Ahamed, M. Nazrul Islam, M. Abdullah Al Nasim and M. Ashraf, "A Soft Computing Based Customer Lifetime Value Classifier for Digital Retail Businesses," 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2021, pp. 0074-0083.

[11] A. Valdivia, "Customer Lifetime Value in Mobile Games: a Note on Stylized Facts and Statistical Challenges," 2021 IEEE Conference on Games (CoG), Copenhagen, Denmark, 2021, pp. 1-5.

[12] A. Tripathi, T. Bagga, S. Sharma and S. Kumar Vishnoi, "Big Data-Driven Marketing enabled Business Performance: A Conceptual Framework of Information, Strategy and Customer Lifetime Value," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 315-320.

[13] R. Rahmadianti, A. Dhini and E. Laoh, "Estimating Customer Lifetime Value using LRFM Model in Pharmaceutical and Medical Device Distribution Company," 2020 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2020, pp. 1-5.

[14] A. G. Aggarwal and S. Yadav, "Customer Segmentation Using Fuzzy-AHP and RFM Model," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 77-80.

[15] T. T. Win and K. S. Bo, "Predicting Customer Class using Customer Lifetime Value with Random Forest Algorithm," 2020 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 2020, pp. 236-241.

[16] Y. Hasanpour, S. Nemati and R. Tavoli, "Clustering System Group Customers through Fuzzy C-Means Clustering," 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), Tehran, Iran, 2018, pp. 161-165.

[17] L. Singh, N. Kaur and G. Chetty, "Customer Life Time Value Model Framework Using Gradient Boost Trees with RANSAC Response Regularization," 2018 IJCNN, Rio de Janeiro, Brazil, 2018, pp. 1-8.

[18] A. Chatterjee, A. Bhattacharya, S. Pal, A. Mukherjee, A. Chakraborty, D. Das, "Non-Autonomous Complex Network Architecture with Gamma Distribution", Proceeding of the conference on Networks Science (NetSci2016), June 2016, Korea

[19] R. Kumar, P. B. Pati, K. Deepa and S. Yanan, "Clustering the Various Categorical Data: An Exploration of Algorithms and Performance Analysis," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-6.