International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Customer Lifetime Value Prediction for Business

Rohit Waratkar

PG Student, Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

Predicting the client Lifetime Value (CLV) of each individual client is a crucial statistic in contemporary corporate strategy since it offers information about their long-term profitability. Businesses may improve customer relationship management, optimize marketing activities, and more efficiently allocate resources by precisely projecting CLV. This study offers a data-driven method for predicting CLV that makes use of demographic information, behavioral markers, and past customer transaction data. The model uses machine learning techniques including decision trees, regression models, and ensemble methods to forecast the net profit attributable to a customer's entire future relationship, segmentation, and targeted marketing. The results demonstrate how useful CLV prediction is for promoting customer-centric tactics and raising overall company profitability.

KEYWORDS: Client Lifetime Value (CLV), Regression Models, Ensemble Methods.

I. INTRODUCTION

Through the use of the strength of predictive analytics and real-time information, hardware businesses can maximize their sales processes and improve their edge over the competitors. This backdrop provides the starting point for exploring in depth the approaches, outcomes, and connotations of deploying data analytics for improving sales performance in various branches. Sales Insights Using Data Analytics: Real-Time Tracking and Optimization of Sales. The project titled "Sales Insights Using Data Analytics: Real-Time Tracking and Optimization of Sales" focuses on leveraging data analytics to improve the sales management process within hardware companies, particularly those with multiple branches. It addresses the limitations of traditional sales tracking methods that rely on static, periodic reports, which are often insufficient for making timely and informed decisions. The system gives sales managers the ability to evaluate sales performance across several branches, pinpoint high-performing locations, and identify areas of underperformance by utilizing real-time data tracking and advanced analytics approaches. With the aid of this data-driven strategy, managers can optimize resource allocation, make proactive changes to sales strategies, and guarantee that resources are allocated to the places that need them the most. The system also integrates predictive analytics to anticipate future sales trends. allowing for better sales forecasting and more effective planning. Overall, the project aims to enable hardware companies to enhance their sales strategies, improve operational efficiency, and drive better business outcomes by optimizing sales performance in real-time. In summary, the project offers a powerful solution for sales optimization, enhancing decision-making, improving resource management, and boosting sales performance across multiple branches using the power of data analytics.

II. RELATED WORK

2.1. Recency-Frequency-Monetary (RFM) Models

- Recency: Time since the last purchase.
- Frequency: Number of purchases in each period.
- Monetary: Total spending in each period.

While straightforward, RFM models often lack the nuance to capture complex customer behaviours and may not generalize well across different industries.

2.2. Buy-Till-You-Die (BTYD) Models

BTYD models, such as the Pareto/NBD and BG/NBD, model customer purchasing behaviour by assuming customers continue to purchase until they "die" or churn. These models effectively handle non-contractual settings where churn is unobserved. However, they often assume stationarity and may not adapt well to changing customer behaviours.

III. DATA AND SOURCES OF DATA

3.1. Regression-Based Models

Machine learning regression models, including Random Forests, Gradient Boosting Machines (e.g., XGBoost, LightGBM), and Support Vector Regression, have been employed to predict CLV by capturing.

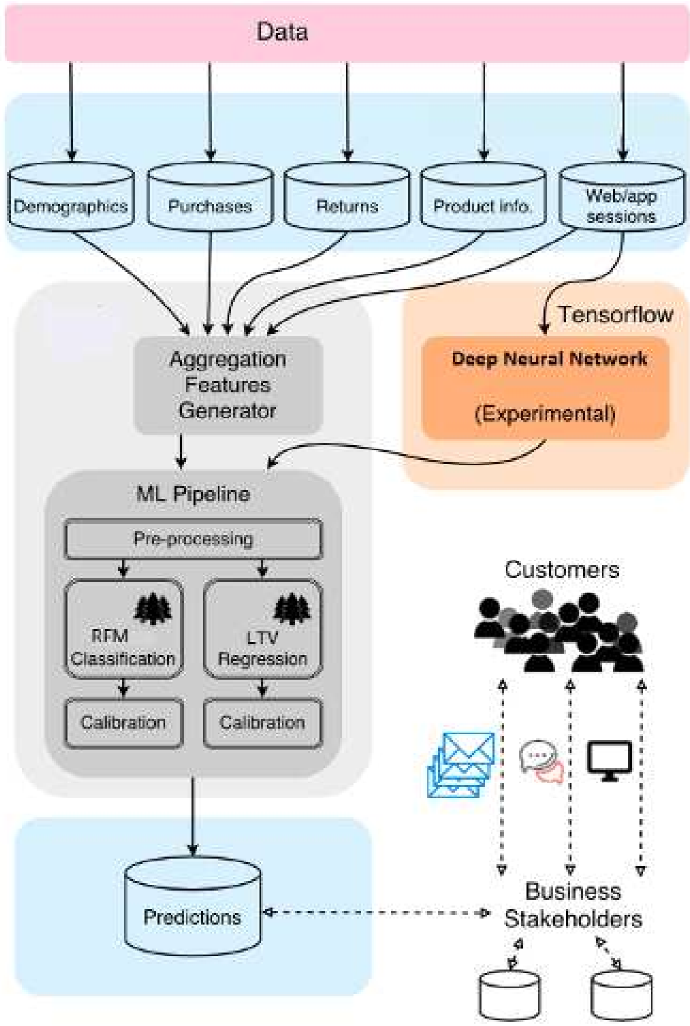


Diagram: Regression-Based Models

3.2. Ensemble Learning
Ensemble methods combine multiple models to improve predictive performance. For instance, integrating Random Forests with Neural Networks has shown enhanced accuracy in CLV prediction.

3.3. Deep Learning Approaches
Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have been utilized to capture temporal dynamics in customer behaviour. Combining LSTM with ensemble methods further enhances prediction accuracy.

IV. RESEARCH METHODOLOGY
4.1. Meta-Learning and Stacked Regression
A meta-learning-based stacked regression approach combines predictions from various models, such as bagging and boosting algorithms, to improve CLV prediction accuracy while maintaining model interpretability.

4.2. Contrastive Multi-View Framework
This framework synthesizes multiple heterogeneous CLV regressors and employs contrastive learning to mitigate data sparsity and noise, enhancing model robustness.

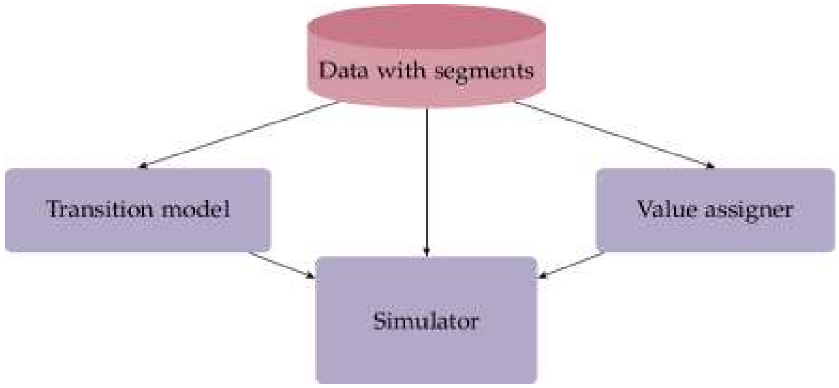
4.3. Optimal Distribution Modelling (OptDist)
OptDist addresses the challenge of modelling complex and mutable CLV distributions by adaptively selecting optimal sub-distributions for each sample, leading to improved prediction accuracy.

4.4. Uncertainty Estimation with Monte Carlo Dropout
Incorporating Monte Carlo Dropout into neural networks allows for uncertainty estimation in CLV predictions, providing confidence intervals alongside point estimates, which is crucial for risk-sensitive decision-making.

V. RESULTS AND DISCUSSION
Embedding methods, inspired by natural language processing, transform categorical variables into dense vectors, capturing semantic similarities and improving model performance. For example, ASOS utilized embeddings to enhance their CLV prediction models.

Handling High-Cardinality Categorical Variables

High-cardinality features, such as ZIP codes or product IDs, pose challenges in modeling. Techniques like target encoding and clustering have been employed to reduce dimensionality while preserving predictive power.



Evaluating CLV prediction models requires appropriate metrics:

- Mean Squared Error (MSE): Measures average squared difference between predicted and actual values.
- Mean Absolute Error (MAE): Captures average absolute difference.
- Gini Coefficient: Assesses model's discriminatory power.
- Calibration Plots: Visualize the agreement between predicted probabilities and observed outcomes.

Accurate CLV predictions inform various business strategies:

- Customer Segmentation: Identifying high-value customers for targeted marketing.
- Resource Allocation: Optimizing marketing spend based on predicted returns.
- Personalization: Tailoring offers and communications to individual customer value.
- Churn Prevention: Proactively engaging customers at risk of attrition.

VI. Acknowledgment

Integrating CLV models into real-time systems enables dynamic decision-making. Research into scalable and efficient algorithms is necessary to facilitate this integration. The field of CLV prediction has seen substantial change, adopting deep learning and sophisticated machine learning methods to grasp the complexities of consumer behavior. Notwithstanding persisting difficulties, research and technology developments could lead to more precise and useful CLV forecasts, which would eventually boost customer satisfaction and business expansion.

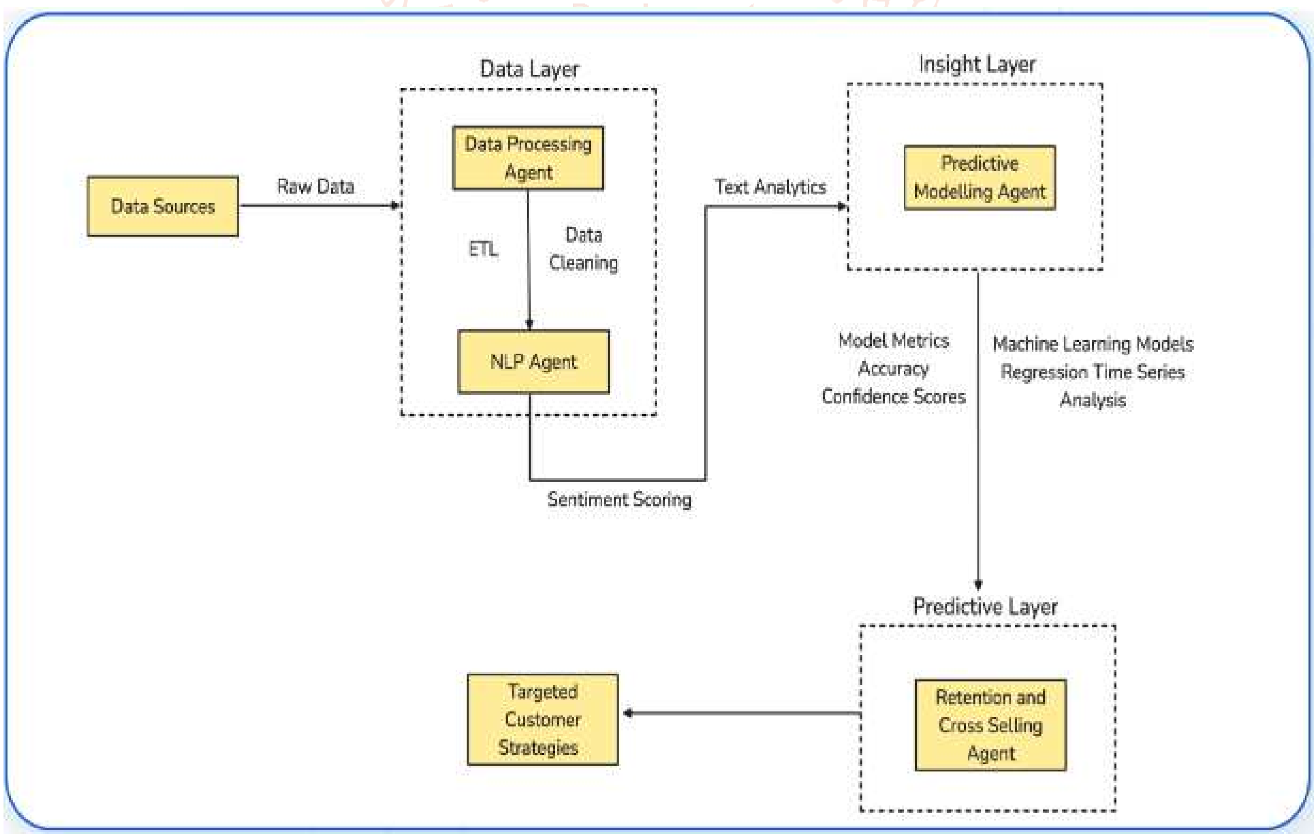


Diagram : CLV Model

VII. REFERENCES

[1] Gadgil, K., Gill, S. S., & Abdelmoniem, A. M. (2023). A Meta-learning based Stacked Regression Approach for Customer Lifetime Value Prediction. arXiv preprint arXiv:2308.08502. arxiv.

[2] Chopra, A., Iyer, P., Joshi, A., & Gupta, M. (2020). Enhancing Customer Lifetime Value Prediction Using Ensemble Learning and Long Short-Term Memory Networks. International Journal of AI Advancements, 9(4). ijoaia.com

[3] Sharma, A., Patel, N., & Gupta, R. (2022). Enhancing Customer Lifetime Value Prediction Using Random Forests and Neural Network Ensemble Methods. European Advanced AI Journal, 11(8). eaaij.com

[4] Wu, C., Li, J., Jia, Q., Zhu, H., Fang, Y., & Tang, R. (2023). Contrastive Multi-view Framework for Customer Lifetime Value Prediction. arXiv preprint arXiv:2306.14400. arxiv

[5] Weng, Y., Tang, X., Xu, Z., Lyu, F., Liu, D., Sun, Z., & He, X. (2024). OptDist: Learning Optimal Distribution for Customer Lifetime Value Prediction. arXiv preprint arXiv:2408.08585. arxiv

[6] Cao, X., Xu, Y., & Yang, X. (2024). Customer Lifetime Value Prediction with Uncertainty Estimation Using Monte Carlo Dropout. arXiv preprint arXiv:2411.15944. arxiv

[7] Wang, X., Liu, T., & Miao, J. (2019). A Deep Probabilistic Model for Customer Lifetime Value Prediction. arXiv preprint arXiv:1912.07753. arxiv

[8] Chamberlain, B., et al. (2017). Customer Lifetime Value Prediction Using Embeddings. arXiv preprint arXiv