

LTVHub: Uma Ferramenta Modular para Cálculo do Customer Lifetime Value com Suporte a Múltiplos Modelos

Maria Eduarda de Pinho Braga¹, João M. A. M. Ramos^{1,3},
Fabrício A. Silva^{1,3}, Linnyer B. R. Aylon^{2,3}

¹Universidade Federal de Viçosa (UFV)

Laboratório de Inteligência em Sistemas Pervasivos e Distribuídos (NESPeD-Lab)
Florestal - MG - Brasil

²Universidade Estadual de Maringá (UEM) Maringá - PR - Brasil

³Manna Team

{maria.e.braga, joao.m.ramos, fabrício.asilva}@ufv.br

lbruiz@uem.br

Abstract. *The Customer Lifetime Value (CLV) is an essential metric for identifying the most valuable customers, enabling comprehensive profit estimations. Furthermore, it allows companies to tailor their services to meet customer expectations, thereby improving the quality of the relationship between the consumer and the business. Despite its advantages, CLV is not widely adopted, and applying it across different contexts presents several challenges, particularly regarding the data required. The goal of this work is to provide a tool¹ that allows CLV estimation for different users in a visually intuitive, modular, extensible, and flexible manner, delivering a robust prediction of the expected CLV.*

Resumo. *O Customer Lifetime Value (CLV) é uma métrica essencial para identificar clientes mais valiosos, possibilitando estimativas abrangentes de lucro. Além disto, permite que a empresa forneça um serviços de acordo com as expectativas de seus clientes, melhorando a qualidade do relacionamento entre consumidor e o negócio. Apesar de suas vantagens, o CLV não é uma métrica utilizada em larga escala, e a aplicação de um contexto ao outro apresenta uma série de dificuldades, principalmente em relação aos dados utilizados. O objetivo deste trabalho é fornecer uma ferramenta¹ que permita a estimativa do CLV para diferentes usuários de maneira visualmente intuitiva, modular, extensível e flexível, fornecendo uma ótima estimativa do CLV esperado.*

1. Introdução

O *Lifetime Value (LTV)* ou *Customer Lifetime Value (CLV)* é uma métrica amplamente reconhecida na literatura de marketing e gestão de relacionamento com o cliente [Berger and Nasr 1998] [Jain and Singh 2002] [Venkatesan and Kumar 2004] [Popa et al. 2021]. Ela permite identificar quais clientes são mais valiosos para a empresa, que respondem positivamente às ações promocionais, gerando maior lucratividade ao longo do tempo. Um cálculo acurado do *CLV* possibilita compreender a dinâmica

¹O vídeo de demonstração da ferramenta LTVHub está disponível em: <https://youtu.be/S-fO2e8nPxi>.

do relacionamento com o cliente, estimando com mais precisão os ganhos esperados e a possível redução de encargos, como custos de atendimento e comunicação.

Apesar de sua relevância teórica e potencial estratégico, as ferramentas que permitem o cálculo do *CLV*, como as fornecidas pela UpGrowth, MétricaHub, Asllan Maciel, WebEngage, WebFX e Upollo, ainda são bastante limitadas, ao exigirem que o usuário insira manualmente métricas previamente calculadas como o ticket médio, a taxa de retenção, margem de lucro, dentre outras. Isso pressupõe que o usuário já tenha feito uma análise prévia dos dados, o que dificulta o acesso ao *CLV*. Além disso, essas ferramentas geralmente retornam o *CLV* médio de forma agregada, considerando a empresa como um todo, dificultando, como consequência, a adoção prática e a exploração plena do potencial analítico do *CLV*.

Dito isso, o presente trabalho tem como objetivo construir um sistema distribuído que permita o uso de diferentes algoritmos para o cálculo do *CLV*, com uma interface intuitiva para possibilitar futuros modelos de aprendizado de forma simples e modular.

2. Fundamentação Teórica

Este estudo baseia-se em [Qismat and Feng 2020] e dá continuidade ao trabalho de [Ramos and Silva 2024], que propôs uma abordagem genérica para predição do *CLV* usando atributos de transações. Aqui, desenvolve-se uma ferramenta que automatiza todo o fluxo de análise.

2.1. RFM

O método *RFM* (Recência, Frequência e Valor Monetário) é amplamente utilizado para o cálculo do *CLV*. Seu nome provém das três variáveis centrais que compõem sua lógica de análise. São elas:

- Recência: Tempo desde a última compra. Quanto menor, mais recentemente o cliente realizou uma compra.
- Frequência: Número de compras no período. Quanto maior, mais frequentemente o cliente realizou compras.
- Valor Monetário: Gasto médio do cliente. Quanto maior, mais o cliente gastou em suas compras.

2.2. Modelo Pareto/Distribuição Binomial Negativa (NBD)

Modelo de método *Bayesiano* que prevê o número esperado de compras futuras, focando na frequência e sem considerar o valor monetário [Schmittlein et al. 1987]. O Pareto modela a probabilidade do cliente estar ativo, através de uma distribuição exponencial, cuja taxa de abandono segue uma distribuição *Gamma*.

2.3. Modelo Beta-Geométrico (BG/NBD)

Similar ao Pareto/NBD, mas mais eficiente. Prevê a frequência de compras e o momento de abandono, é relevante para relações com frequência instável [Fader et al. 2005].

2.4. Modelo Gamma-Gamma

O modelo *Gamma-Gamma* [Fader and Hardie 2013] é uma extensão do modelo *Pareto/NBD* [Schmittlein et al. 1987], que atribui um valor monetário a cada compra futura, além de estimar a probabilidade de o cliente continuar ativo.

2.5. Modelos de Aprendizado de Máquina

Modelos de *ML* incorporam múltiplas variáveis, gerando, potencialmente, uma maior precisão. Para a modelagem com *Aprendizado de Máquina*, o LTVHub utiliza uma abordagem comparativa para selecionar o algoritmo de maior desempenho. Os modelos são treinados e, em seguida, avaliados com base na métrica do Erro Quadrático Médio (MSE), que quantifica a distância entre valores previstos e reais. O modelo com menor erro é escolhido para a predição final. Os algoritmos avaliados incluem: LassoCV, ElasticNet, Random Forest Regressor, Gradient Boosting Regressor, Histogram Gradient Boosting Regressor, LightGBM Regressor (Light Gradient Boosting Machine) e Extreme Gradient Boosting (XGBoost).

3. Arquitetura

O *Back-End* do sistema foi desenvolvido em *Python*, utilizando o framework *Flask* para expor uma interface de comunicação eficiente com o *Front-End*. Além disso, sua lógica interna foi estruturada com base no padrão de projeto Cadeia de Responsabilidade, no qual as tarefas são processadas em etapas sequenciais: cada etapa realiza uma operação específica e encaminha o resultado à próxima. Essa abordagem favorece a modularidade, a reutilização e a manutenção do código.

O *Front-End* foi implementado usando React, que proporciona maior dinamismo e responsividade na construção de componentes reutilizáveis.

Uma descrição breve da arquitetura do sistema pode ser encontrada na Figura 1.

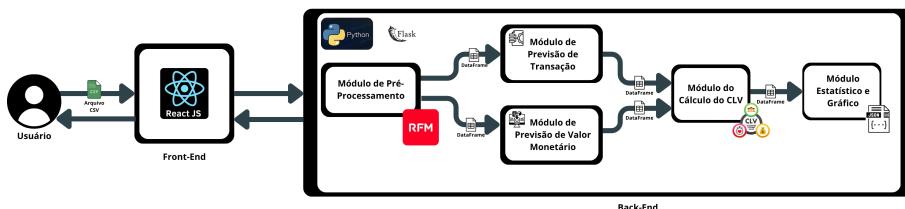


Figura 1. Arquitetura LTVHub.

Front-End: Consiste na parte do código que permite a interação do usuário. Aqui foi definida uma interface intuitiva e facilmente extensível. O sistema foi projetado com foco na extensibilidade e na facilidade de integração de novos modelos. Para isso, utiliza um arquivo, no qual o desenvolvedor pode adicionar novos modelos à interface de forma simples e direta, informando apenas os dados essenciais para seu uso. Essa abordagem elimina a necessidade de modificações complexas no código-fonte, promovendo maior flexibilidade e agilidade no desenvolvimento e na manutenção do sistema.

Na sequência, detalham-se os módulos responsáveis pelo funcionamento do *Back-End* do sistema.

Módulo de Pré-Processamento: Esta etapa recebe os dados do *Front-End* e realiza a limpeza, tratamento de nulos e transformação dos atributos. Atualmente, são permitidas as transformações para o *RFM* e para o *RFMT* [Ullah et al. 2023], este último incluindo a métrica T (tempo desde a primeira compra), que permite diferenciar clientes novos e antigos. Embora ambos sejam métodos tradicionais para calcular o *CLV*,

limitam-se a considerar apenas o histórico de compras. Para superar essa limitação, o presente artigo combina o RFM com modelos preditivos para estimar o *CLV* futuro.

Módulo de previsão de transações: Neste módulo, são definidos os modelos utilizados para estimar o número médio de transações por cliente até a data da análise. Para isso, são utilizados os dados processados no módulo anterior. A ferramenta conta com os modelos probabilísticos *Pareto* [Schmittlein et al. 1987] e *Beta Geométrico* [Fader et al. 2005], além dos modelos de aprendizado de máquina desenvolvidos no trabalho de [Ramos and Silva 2024].

Módulo de previsão do valor monetário: Se baseia em fazer as estimativas do valor monetário médio por transação. Pode-se utilizar tanto o *Gamma-Gamma* quanto o modelo de aprendizado de máquina.

Módulo do cálculo do CLV: Baseia-se nas informações calculadas nos módulos de previsão, como frequência de compra e valor monetário previsto, para realizar o cálculo do *CLV*. O *CLV* é composto pela multiplicação da frequência prevista de transações pelo valor monetário médio previsto, ajustando o resultado por uma taxa de desconto e pelo número de períodos desejados. O modelo permite calcular o *CLV* tanto para dados de calibração, quanto para dados de treino e teste, garantindo que valores negativos sejam ajustados para zero. Assim, o módulo fornece uma estimativa do *CLV* futuro esperado de cada cliente.

Módulo Estatístico e Gráfico: Responsável por consolidar os dados, individuais e agrupados, calculados nos módulos anteriores, e prepará-los para a visualização. Essas informações são estruturadas para envio ao *Front-End*, as métricas enviadas incluem o *CLV* previsto, frequência e valor monetário, bem como a segmentação dos clientes com base nos atributos do RFM, seguindo o trabalho de [Qismat and Feng 2020] (Tabela 1).

Tabela 1. Classificação dos clientes com base no modelo RFM

| Tipo de Cliente | Característica | Estratégia Recomendada |
|--|--|--|
| Cliente de Alto Valor | Compra recente, alta frequência e alto volume. Principais consumidores. | Elevar a status VIP, oferecer serviços personalizados e alocar mais recursos. |
| Cliente de Desenvolvimento Estratégico | Compra baixa frequência, mas alto valor unitário. Potenciais atacadistas/corporativos. | Oferecer pontos de membro e descontos para aumentar fidelização e retenção. |
| Cliente em Fase de Proteção | Não compra recentemente, mas alta frequência e volume passados. | Enviar atualizações (novos produtos/serviços) para incentivar retorno e consumo. |
| Cliente de Retenção Crítica | Não compra recentemente, baixa frequência, mas alto valor médio. | Enviar ofertas personalizadas, atualizações e promoções para evitar perda. |
| Cliente de Valor Geral | Compra alta frequência, mas baixo volume. | Introduzir novos produtos/serviços/funcionalidades para aumentar volume e engajamento. |
| Cliente em Desenvolvimento | Compra baixa frequência e baixo volume. Novos ou pouco engajados. | Oferecer serviços adicionais, apresentar novos produtos e incentivar aumento de consumo. |
| Cliente de Retenção Geral | Não compra recentemente, mas alta frequência passada, com baixo volume. | Apresentar novos produtos para estimular consumo. |
| Cliente Perdido | Não compra, baixa frequência e baixo volume. | Utilizar promoções para tentar reconquistar. Priorizar outros segmentos de recursos limitados. |

4. Interface e Utilização

Após todo o processamento no *Back-End*, o sistema retorna as informações para a interface, permitindo que o usuário analise o comportamento de seus clientes, tanto no

passado quanto no futuro, inserindo apenas o conjunto de dados das transações. Ao iniciar a aplicação, é necessário fornecer um arquivo no formato *.csv* com os registros das transações a serem analisadas. Esse arquivo deve conter, obrigatoriamente, a coluna de identificação única dos usuários, a coluna da data e a coluna do valor das transações. Após a inserção do arquivo de dados, é necessário realizar a configuração dos modelos no sistema. Nessa etapa, o usuário deve indicar os identificadores das colunas obrigatórias, selecionar o modelo de predição de frequência e o modelo de predição monetária a serem utilizados, além de definir o período de previsão, especificado em número de semanas, para o cálculo do *CLV* futuro. Essas informações são enviadas ao *Back-End*, que realiza o processamento, executa as predições e retorna, para cada cliente, as métricas calculadas, como *CLV*, frequência, valor monetário e segmentação, além de dados agregados para análise por grupos.

No *Front-End*, são apresentadas ao usuário duas seções principais para a visualização e interpretação dos dados obtidos a partir do modelo de previsão de *CLV*: a página “Clientes” e a página “Estatísticas”. Ambas são acessadas no *Front-End* e oferecem as perspectivas de análise individual e agregada. Na página de “Clientes”, os resultados do modelo são apresentados de forma detalhada para cada cliente individualmente. Esta página permite buscar clientes por ID ou segmento, filtrar a lista de clientes pelos seus tipos e ordenar os clientes por ID, *CLV*, frequência esperada, valor por transação ou categoria (conforme detalhadas na Tabela 1).

Além disso, para cada cliente são exibidas informações essenciais, como o seu tipo, a frequência prevista, o valor médio por transação e o valor estimado de *CLV*. Ao final, também é possível acessar a página de detalhes do cliente, que fornece mais informações sobre como lidar com o perfil daquele cliente, além de exibir seu histórico de compras em formato tabular e gráfico.

A página “Estatísticas” apresenta uma visão agregada e comparativa do comportamento dos clientes por meio de gráficos interativos. Ela permite identificar padrões gerais e avaliar o desempenho dos diferentes segmentos. As visualizações incluem:

- Gráficos de setores: mostram a distribuição percentual dos clientes, valor monetário total e frequência de transações por tipo de cliente, facilita a identificação dos perfis mais valiosos.
- Gráfico de barras horizontais: apresenta o ranking de *CLV* total por tipo de cliente, destacando os perfis com maior valor estimado.
- Interatividade: permite ocultar ou exibir categorias específicas via legenda interativa, possibilita análises comparativas mais focadas.

5. Conclusão

Este trabalho apresentou o LTVHub, uma ferramenta modular, extensível e intuitiva para cálculo do *Customer Lifetime Value*, que utiliza diferentes modelos de aprendizado de máquina para previsão da frequência e valor monetário. A arquitetura desenvolvida facilita a integração de novos modelos e garante uma interface amigável e completa para os usuários finais. Em conclusão, a aplicação automatiza todo o processo de transformação e predição dos dados, até a visualização dos resultados, contribuindo, assim, para a maior adoção e aplicação da métrica do *CLV*. Para trabalhos futuros, pretende-se incorporar novos algoritmos. Além disso, busca-se expandir a ferramenta para fornecer ao usuário um

retorno mais detalhado sobre o desempenho dos modelos de predição escolhidos, apresentando suas principais métricas. Desse modo, será possível realizar uma comparação mais precisa entre os modelos existentes.

Agradecimentos

Os autores gostariam de agradecer o apoio do Ministério da Ciência, Tecnologia e Inovações (MCTI), da Equipe Manna, da Fundação Araucária, da Softex, do CNPq (número do projeto 421548/2022-3), da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e da CAPES.

Referências

- Berger, P. D. and Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, 12(1):17–30.
- Fader, P. S., Hardie, B. G., and Lee, K. L. (2005). “counting your customers” the easy way: An alternative to the pareto/nbd model. *Marketing science*, 24(2):275–284.
- Fader, P. S. and Hardie, B. G. S. (2013). The gamma-gamma model of monetary value. *Marketing Science Institute Working Paper*, 2:1–9.
- Jain, D. and Singh, S. S. (2002). Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing*, 16(2):34–46.
- Popa, A.-L., Sasu, D. V., and Tarcza, T. M. (2021). Investigating the importance of customer lifetime value in modern marketing – a literature review. *Annals of the Faculty of Economics*, 30(2):410–416.
- Qismat, T. and Feng, Y. (2020). Comparison of classical rfm models and machine learning models in clv prediction. Master thesis, BI Norwegian Business School, Oslo, Norway. GRA 19703, Master of Science.
- Ramos, J. and Silva, F. (2024). A solution for predicting the customer lifetime value of different market segments. In *Anais do XII Symposium on Knowledge Discovery, Mining and Learning*, pages 81–88, Porto Alegre, RS, Brasil. SBC.
- Schmittlein, D. C., Morrison, D. G., and Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science*, 33(1):1–24.
- Ullah, A., Mohmand, M. I., Hussain, H., Johar, S., Khan, I., Ahmad, S., Mahmoud, H. A., and Huda, S. (2023). Customer analysis using machine learning-based classification algorithms for effective segmentation using recency, frequency, monetary, and time. *Sensors*, 23(6):3180.
- Venkatesan, R. and Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, 68(4):106–125.