

# Customer Lifetime Value Prediction: A Machine Learning Approach

João Marcos Alves Modesto Ramos<sup>1</sup>, Fabrício A. Silva<sup>1</sup>

<sup>1</sup>Laboratório de Inteligência em Sistemas Pervasivos e Distribuídos (NESPeD-Lab)  
Universidade Federal de Viçosa (UFV)  
Florestal - MG - Brasil

{joao.m.ramos, fabricio.asilva,}@ufv.br

**Resumo.** *A previsão do Customer Lifetime Value (CLV) é de suma importância em diversos modelos de negócios. No entanto, esse cálculo varia de acordo com o contexto e o escopo do negócio. Este trabalho tem como objetivo realizar a previsão do CLV utilizando algoritmos de aprendizado de máquina e compará-lo com os principais modelos utilizados na literatura em 3 bases de dados diferentes. Durante o trabalho são descritas as escolhas e decisões tomadas para a construção dos modelos, sendo mostrado que os modelos de aprendizado de máquina apresentaram melhores resultados no cálculo do número esperado de transações, e obtiveram resultados muito semelhantes no modelo de cálculo do valor médio por transação.*

**Abstract.** *The Customer Lifetime Value (CLV) prediction is of paramount importance in various business models. However, this calculation varies according to the context and scope of the business. This work aims to predict the CLV using machine learning algorithms and compare it with existing models used in the literature in 3 different datasets. The choices and decisions taken for constructing the models are described, and it is shown that the machine learning models presented better results in the calculation of the expected number of transactions, and obtained very similar results in the model that calculates the average value per transaction.*

## 1. Introdução

Nas últimas décadas, foi percebido que é mais vantajoso tentar manter as relações com os clientes atuais, do que prospectar novos clientes. De acordo com Kotler e Keller [Sargeant 2003], os custos associados à obtenção de novos clientes podem ser cinco vezes maiores do que os custos associados com a manutenção de um bom relacionamento com clientes. Isto está fazendo com que as empresas mudem suas abordagens centradas no produto, onde o foco é vender mais produtos, para abordagens centradas no cliente [Aeron et al. 2012]. Nas abordagens centradas no cliente, foca-se em fornecer serviços que fidelizam tal cliente, com o intuito de manter uma boa relação e aumentar a retenção de clientes [Chan 2008].

Para isso, entender qual é a previsão de valor que cada cliente irá gerar para a empresa é fundamental. Baseando-se nisso, foi identificado que cada cliente possui um determinado valor financeiro atribuído a seu relacionamento futuro com a companhia. Este valor é definido como *Lifetime Value* (LTV) ou *Customer Lifetime Value*

(CLV)<sup>1</sup> [Dwyer 1997]. Ele também pode ser associado com a “vida útil” do cliente, definida como quanto tempo é esperado para que aquele cliente continue ativo e gerando receita para a empresa [Khajvand et al. 2011].

Um cálculo efetivo do CLV pode trazer diversos benefícios tanto para as companhias quanto para os seus clientes. Para as companhias, trazem benefícios como a identificação de clientes mais valiosos, uma vez que os 20% dos clientes mais valiosos são responsáveis por 80% do lucro da companhia [Ekinci et al. 2014]. Além disso, permite entender quais clientes estão em risco de *churn*<sup>2</sup>, para assim tentar trazer o interesse de volta ou identificar o motivo da evasão. Já para os clientes, os serviços podem ser ofertados de forma personalizada pela companhia de acordo com suas preferências, além de um atendimento de acordo com as suas necessidades.

Contudo, o cálculo do CLV não é trivial. Em primeiro lugar, para cada tipo de companhia, podem ser necessários parâmetros e métricas específicos. Por exemplo, no contexto não-contratual, não há uma confirmação se aquele cliente ainda é um cliente ativo, enquanto nos meios contratuais, espera-se que o cliente ficará ativo até o fim do contrato [Singh and Jain 2013]. Em segundo lugar, o cálculo do CLV é poderoso para a análise de clientes, mas sua aplicação pode ser desafiadora [Fader et al. 2005b], visto que certas implementações são complexas e/ou computacionalmente custosas. Outra dificuldade encontrada consiste na volatilidade dos relacionamentos entre os clientes, sendo que clientes que apresentam não ser lucrativos, podem se tornar clientes lucrativos [Méndez-Suárez and Crespo-Tejero 2021].

Existem diversos trabalhos na literatura que fazem da utilização de modelos probabilísticos para a previsão do CLV, como o trabalho de [Mammadzada et al. 2021], que apresenta uma aplicação de modelos probabilísticos para um banco. Outro exemplo é o trabalho de [Khajvand et al. 2011], que usa de duas soluções para estimar o CLV de uma companhia de beleza e saúde, utilizando o modelo RFM, que consiste em uma separação dos dados de acordo com a recência, frequência, e o valor monetário das compras para categorizar os clientes, e uma variação proposta pelo autor. Tem surgido nos últimos anos a utilização de técnicas de aprendizado de máquina para este cálculo, como no trabalho de [Desirena et al. 2019], que utiliza redes neurais para aumentar o CLV na indústria de seguros por meio de recomendações mais assertivas. Já o trabalho de [Sun et al. 2023] faz uso de aprendizado de máquina para medir o CLV e a segmentação com base no valor do ciclo de vida do cliente.

Apesar dos avanços, grande parte dos trabalhos fazem uso de aprendizado de máquina na etapa de segmentação dos clientes, e realizam o cálculo do CLV utilizando os modelos probabilísticos. O objetivo deste trabalho é usar o aprendizado de máquina para criar uma solução para calcular o CLV e comparar com os modelos probabilísticos, visto que, segundo o trabalho de [Temor Qismat 2020], existem poucos estudos que fazem a comparação entre os métodos probabilísticos e os modelos de aprendizado de máquina.

O principal diferencial deste trabalho é que foi feita, em três bases de dados, a utilização de modelos de aprendizado de máquina para prever o valor do CLV em conjunto com modelos probabilísticos já consolidados na literatura. Foi verificada a eficácia

---

<sup>1</sup>Esses dois termos, LTV e CLV, serão usados para indicar o mesmo conceito neste texto.

<sup>2</sup>Churn é um indicador de quando um cliente encerra o seu relacionamento com a empresa.

dos modelos utilizando aprendizado de máquina, podendo complementar ou substituir qualquer um dos modelos probabilísticos no cálculo do CLV.

O restante deste texto está organizado contando com a apresentação de trabalhos relacionados na seção 2; os materiais e métodos utilizados durante a pesquisa na seção 3; a solução abordada se encontra na seção 4, e os resultados obtidos e métodos de avaliação na seção 5; por fim, as considerações finais na seção 6.

## **2. Trabalhos Relacionados**

Foi feita a divisão entre trabalhos que são relacionados ao cálculo do CLV e em abordagens de aprendizado de máquina.

### **2.1. Cálculo do *Customer Lifetime Value***

Ao prever o CLV, há uma necessidade de diferenciar de acordo com o tipo do relacionamento entre cliente e empresa, que pode ser contratual ou não-contratual [Reinartz and Kumar 2000]. Um relacionamento contratual implica que legalmente há uma relação entre o cliente e companhia, onde a companhia geralmente tem uma expectativa de quando um cliente irá se tornar inativo. No caso não-contratual, este relacionamento não é regido por um contrato ou adesão [Reinartz and Kumar 2000]. Além deste tipo de classificação, existe outra de acordo com o período de tempo em que ocorrem as compras, sendo que se as compras forem discretas, elas só podem ocorrer em um determinado momento, e caso sejam contínuas, podem ocorrer a qualquer momento [Temor Qismat 2020].

Existe também a classificação em relação aos modelos, divididos entre modelos de comportamento do passado, e modelos de comportamento de passado-futuro. Duas características diferenciam essas categorias: a primeira diferença baseia-se no pressuposto de que os clientes sujeitos a avaliações estarão ativos no futuro. A segunda diferença é se os custos dos clientes são incluídos nos modelos ou não. O primeiro grupo de modelos faz os cálculos incluindo a taxa de ativação futura de clientes e também os custos associados ao cliente, enquanto o último grupo não os leva em consideração. Os modelos de comportamento do cliente no futuro também podem ser separados em duas categorias com base no atributo de incluir ou não o custo de aquisição do cliente [Hiziroglu and Sengul 2012].

Para ilustrar diversos contextos, existem soluções implementadas para o meio bancário, como o trabalho de [Kahreh et al. 2014], onde se investiga o papel do CLV nos benefícios da segmentação. Existem também trabalhos voltados para o varejo, como o estudo de [Khajvand et al. 2011], em que o objetivo é criar relacionamentos mais próximos e profundos com os clientes e maximizar o valor vitalício de um cliente para uma organização. Existem trabalhos para campanhas de doações, como o trabalho de [Sargeant 2003], que analisa a contribuição que o CLV pode fazer para a arrecadação de fundos.

Por fim, certos métodos do cálculo de CLV acabam se destacando dos demais. Temos, por exemplo, o Modelo de Pareto/NBD [Schmittlein et al. 1987] e o método RFM [Hughes 1996] e suas variações, como discutido a seguir.

O modelo de Pareto/NBD [Schmittlein et al. 1987] é feito para analisar uma base de clientes em cenários não-contratuais. Isto é feito através de derivações de uma série

de expressões e, entre outras coisas, (i) a probabilidade de que um cliente com um determinado histórico de transações ainda esteja ativo, e (ii) o número esperado de transações futuras para um cliente, condicionado ao seu histórico de transações. É extremamente complexo e computacionalmente intenso. Este modelo assume que o cliente está ativo por um período observado de tempo até ficar inativo. Enquanto está ativo, o cliente é retratado por uma função de distribuição *Gamma*, conhecida como distribuição de Pareto. Nela, usa-se a recência, a frequência e a duração do período de observação para prever as compras futuras de um cliente.

Baseando-se nos resultados de [Schmittlein et al. 1987], surgiu o modelo Gamma-Gamma [Fader et al. 2005a], expandindo o modelo de Pareto para permitir que seja possível estimar o valor gasto por transação no futuro. Ele assume que o valor médio segue uma distribuição *Gamma* com parâmetro de forma  $px+q$  e parâmetro de escala  $v+x\bar{z}$ , sendo que  $\bar{z}$  é a média observada do valor das transações e  $x$  o número de observações.

Surge também uma alternativa ao modelo de *Pareto*, que consiste no *BG/NBD* [Fader et al. 2005b], que tenta descrever a taxa com que clientes fazem compras e a taxa que eles deixam de consumir. Este modelo é mais simples em termos computacionais que o modelo de Pareto, e executa de maneira mais rápida, e é apresentado que os resultados são bem semelhantes na literatura, sendo uma boa alternativa.

Por outro lado, os modelos RFM [Hughes 1996] são aplicados em diversas áreas de Marketing, sendo utilizados para analisar o comportamento de um cliente, e fazer previsões baseado nisto. São muito adotados em técnicas de segmentação de clientes, e baseia-se na combinação de três métricas:

- Recência: o quão recente foi feita a última transação.
- Frequência: o quão frequentes são feitas as transações.
- Valor monetário: o quanto, em valor monetário, a transação corresponde.

## 2.2. Abordagens de Aprendizado de Máquina

Em relação à utilização de abordagens de aprendizado de máquina na previsão de CLV, há uma carência de trabalhos na área. Porém, tem-se trabalhos como o de [Sun et al. 2023], cujo objetivo é fazer a segmentação de clientes com base em algoritmos de aprendizado de máquina e modelos de análise de gerenciamento de relacionamento com o cliente e a criação de um modelo de identificação de segmentação de valor do cliente em uma relação não-contratual.

O trabalho de [Vanderveld et al. 2016] desenvolve um sistema que prevê o CLV futuro, utilizando uma implementação do *Random Forest* e o engajamento do cliente pelo e-mail e aplicativos móveis, desenvolvendo um sistema que atualiza o CLV diariamente baseando nas interações. Este trabalho também desenvolve modelos para diferentes tipos de clientes, permitindo pesos diferentes para atributos relevantes para cada tipo. Este tipo de modelo, apesar de eficaz, é ideal apenas quando as interações com o cliente são comuns, e que estas interações sejam capturadas em dados abundantes, o que não é o contexto de vários negócios.

Alternativas mais modernas, como o trabalho de [Temor Qismat 2020], utiliza modelos de aprendizado de máquina, e compara os resultados entre a mistura do Pareto/NBD em conjunto com o modelo Gamma-Gamma para categorizar seus clientes em

8 classes, de acordo com o CLV esperado usando a segmentação do RFM. Foi observado que modelos de aprendizado apresentaram resultados superiores. Porém, este trabalho cita que o modelo Pareto/NBD teria um desempenho muito melhor se comparado com um modelo de aprendizado de máquina que prevê o CLV diretamente, em vez dos segmentos.

Outra solução utilizando técnicas mais modernas, consiste no trabalho de [Desirena et al. 2019], que faz uso de redes neurais para fazer recomendações de produtos para aumentar os lucros em uma indústria de seguros, um modelo contratual. Porém, este trabalho não calcula o CLV futuro, mas apenas o CLV atual do cliente, e procura alternativas de produtos que possam aumentar o CLV atual.

Neste trabalho, é proposto um modelo de previsão do CLV do cliente baseado no passado-futuro, sem incluir os custos de aquisição. O objetivo deste trabalho é realizar uma comparação justa, e mostrar que os métodos de aprendizado de máquina possuem desempenho similar ou melhor do que os modelos probabilísticos, e que se combinados, podem produzir resultados relevantes. Diferente dos trabalhos anteriormente apresentados, busca-se criar um modelo que prediz o valor do CLV, ao invés da categoria do cliente, além de permitir que este modelo seja aplicável a contextos gerais e com dados esparsos.

### 3. Materiais e Métodos

Nesta seção são apresentados os dados utilizados neste trabalho, além da definição matemática do que é considerado o CLV, e os procedimentos para o seu cálculo.

#### 3.1. Os Dados

Para este estudo, foram utilizados três conjuntos de dados. O objetivo de testar com diversas bases de dados consiste em validar os modelos em diferentes situações com diferentes comportamentos de compras.

O primeiro conjunto de dados, chamado CDNOW, contém o histórico de compras até o final de junho de 1998 de 23.570 indivíduos que fizeram sua primeira compra na CDNOW, uma empresa que operava um site de compras on-line que vendia CDs e produtos relacionados à música. Este conjunto foi usado no artigo original do modelo Gamma-Gamma [Fader et al. 2005a].

No segundo conjunto, chamado OpenCDP, foram utilizados dados de interações dos clientes por 5 meses (outubro de 2019 a fevereiro de 2020) de uma loja online de cosméticos de médio porte, sendo esta disponibilizada na plataforma Kaggle<sup>3</sup> pelo projeto Open CDP. Cada linha nos dados representa um entre quatro eventos possíveis de interação entre o cliente e a loja. Foram utilizados somente as interações de compra, contendo 110 mil usuários únicos, realizando 159 mil compras.

O terceiro conjunto de dados, chamado Olist, contém informações de 100 mil compras de 2016 a 2018 feitos em vários locais no Brasil na plataforma Olist<sup>4</sup>, por 65 mil clientes, também disponibilizado na plataforma Kaggle<sup>5</sup>.

---

<sup>3</sup><https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-cosmetics-shop>

<sup>4</sup><https://olist.com/pt-br/>

<sup>5</sup><https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>

### 3.2. Cálculo do *Customer Lifetime Value*

A definição e cálculo do *Customer Lifetime Value* (CLV) varia entre os autores, como apresentado por [Dwyer 1997], em que o CLV representa o valor presente dos benefícios esperados subtraídos pelos encargos dos clientes. Já no trabalho de [Aeron et al. 2012], tem-se que um cliente lucrativo é definido como uma pessoa, empresa ou residência cujo a receita excede os gastos da empresa em atraí-lo, vender e servir aquele cliente ao longo do tempo. O excesso pode ser considerado como o CLV. Já em trabalhos como o de [Kahreh et al. 2014], o CLV é visto como uma métrica para medir o real valor de um determinado cliente de uma parte especial do mercado. Este conceito é antigo, sendo que uma das primeiras definições de CLV foi feita por [Kotler 1974] como o "valor presente do fluxo de lucro futuro esperado em um determinado horizonte de tempo de transação com o cliente". Porém, um conceito que permeia entre os trabalhos consiste na definição apresentada por [Dwyer 1997], em que CLV consiste nos lucros de um determinado cliente subtraídos os custos para atraí-lo.

O trabalho de [Temor Qismat 2020] definiu que, dado que  $P_{it}$  representa o valor pago pelo cliente  $i$  em um tempo  $t$ ,  $C_{it}$  representa o custo de servir o cliente  $i$  em um tempo  $t$ ,  $AC_i$  o custo de aquisição do cliente  $i$ ,  $r_{it}$  é a probabilidade do cliente estar ativo no tempo  $t$ ,  $T$  é o horizonte de tempo estimado para o cálculo do CLV e por último,  $d$  sendo a taxa de desconto, temos a seguinte fórmula:

$$CLV_i = \sum_{t=1}^T \frac{(P_{it} - C_{it}) \times r_{it}}{(1 + d)^t} - AC_i$$

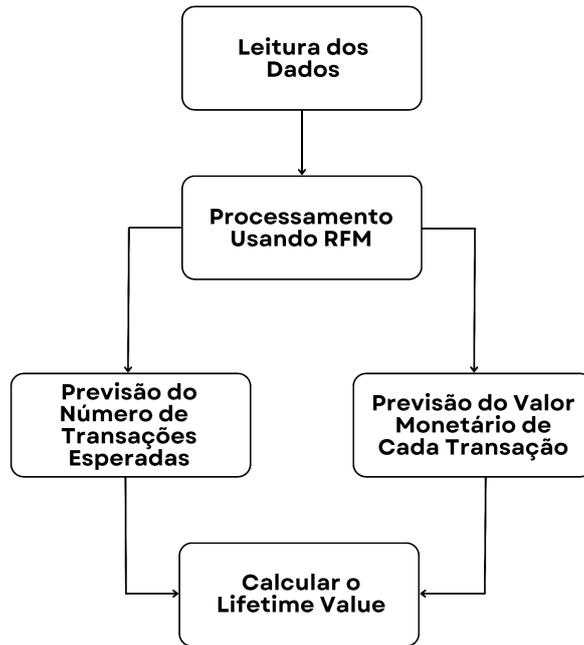
Porém, em alguns cenários os custos envolvidos não estão disponíveis, e o CLV se resume à receita esperada vinda de cada cliente. Essa alternativa é adotada neste trabalho, e o CLV é calculado como a receita deste cliente ajustada a uma taxa de desconto ao longo do tempo. Com isto, tem-se a seguinte fórmula:

$$CLV_i = \sum_{t=1}^T \frac{N_{it} \times V_{it}}{(1 + d)^t} \quad (1)$$

Onde  $N_{it}$  consiste no número de transações esperadas que aquele determinado cliente  $i$  irá realizar no horizonte de tempo  $t$ , e que  $V_{it}$  corresponde ao valor esperado por transação daquele mesmo cliente. Além disso, para todos os dados, foi considerado que a taxa de desconto será de 6%, visando simular a inflação anual, e em relação ao horizonte estimado, foi escolhido 3 meses. Com isto, podemos dividir as etapas do cálculo de acordo com o diagrama na Figura 1, que serão detalhados a seguir.

### 3.3. Processamento Usando RFM

Primeiramente, é preciso extrair métricas dos clientes usando uma das alternativas do método RFM. Neste trabalho, foi escolhido o RFMT [Ullah et al. 2023] (*Recency, Frequency, Monetary, Time*), visto que ele adiciona a quanto tempo o cliente existe na base de dados ao modelo RFM, com o objetivo de melhorar o entendimento sobre os comportamentos e a fidelidade dos clientes. Dado que um cliente  $i$  possui  $x$  transações no período



**Figura 1. Diagrama das etapas para o cálculo do CLV**

observado, e que  $z_1, z_2, \dots, z_x$  denote o valor de cada transação e que  $d_1, d_2, \dots, d_x$  denote a data de cada transação, os cálculos são feitos conforme descrito abaixo.

A recência  $R$  consiste no tempo de atividade do cliente quando ele fez a última compra. Isso é igual ao intervalo entre a primeira compra de um cliente e sua última compra na unidade de tempo escolhida.

$$R_i = d_x - d_1$$

A frequência  $F$  representa o número de compras que o cliente fez, porém, subtraí-se um do número total de compras, para que assim, clientes que são considerados *One-time-buyers*<sup>6</sup>, sejam facilmente detectados, portanto:

$$F_i = x - 1$$

O valor monetário  $M$  representa o valor médio das compras de um determinado cliente. Isso é igual à soma de todas as compras de um cliente dividida pelo número total de compras.

$$M_i = \bar{z} = \sum_{i=1}^x z_i / x$$

$T$ , oriundo de *Time*, representa quanto tempo o cliente existe até o final do tempo de observação  $d_y$ . Isso equivale à duração entre a primeira compra de um cliente e o final do período em estudo.

<sup>6</sup>Um consumidor que compra apenas um produto ou serviço de uma empresa.

$$T_i = d_y - d_1$$

### 3.4. Estimativa do número de transações

O processo de estimar o número de transações esperadas consiste em, baseado nos valores retornados do Modelo RFMT, estimar o número de transações que um determinado cliente irá realizar no horizonte de tempo esperado. Para isto, foi proposto neste trabalho um modelo usando aprendizado de máquina para realizar esta estimativa, e para comparar os resultados da solução elaborada, assim como o trabalho de [Temor Qismat 2020], foram escolhidos os modelos Pareto/NBD [Schmittlein et al. 1987] e *BG/NBD* [Fader et al. 2005b].

### 3.5. Estimativa do valor monetário

Para obter o lucro esperado por um determinado cliente, é necessário estimar o valor de cada transação que aquele determinado cliente irá realizar, baseando-se nos valores das transações passadas e os retornos do modelo RFMT.

Pensando nisto, foi proposta neste trabalho uma solução usando aprendizado de máquina, e esta foi comparada com um dos métodos mais utilizados na literatura, que consiste no modelo Gamma-Gamma [Fader et al. 2005a], que estima o valor médio monetário por transação de cada cliente.

## 4. Solução Proposta

Nesta seção é apresentada a solução proposta neste trabalho, além dos detalhamentos sobre como foi feita a separação dos dados, métricas utilizadas e escolhas feitas no processo de desenvolvimento do modelo.

### 4.1. Modelos de aprendizado de máquina

Os modelos baseados em aprendizado de máquina incorporam muitos parâmetros e atributos, opondo os modelos probabilísticos. Isto pode produzir uma melhor acurácia e melhores resultados, sendo essa a hipótese deste trabalho.

Para realizar a predição do número de transações e o valor monetário médio por transação, foram escolhidos diferentes algoritmos preditivos conhecidos na literatura, utilizando diversas metodologias para tentar encontrar determinados comportamentos entre os dados. Foram escolhidos os seguintes algoritmos: *LassoCV*, *Elastic Net*, *Kernel Ridge*, *Random Forest Regressor*, *Gradient Boosting Regressor*, *Histogram-based Gradient Boosting Regression Tree*, *LightGBM* e *Extreme Gradient Boosting Regressor*.

Em relação à otimização dos hiperparâmetros, foi utilizado o algoritmo do *Grid Search* para buscar os melhores hiperparâmetros de cada modelo. Para o algoritmo *LassoCV*, utiliza-se o número de  $\alpha \in \{100, 200, 500, 100\}$  e o número máximo de iterações  $N \in \{1000, 1500, 2000\}$ . Para o algoritmo *Elastic Net*, utiliza-se  $\alpha \in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100\}$ , o número máximo de iterações  $N \in \{1000, 1500, 2000\}$  e o parâmetro de mistura  $l_1 \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ . Para o algoritmo *Random Forest Regressor*, utiliza-se o número mínimo de folhas  $L_{Min} \in \{1, 2, 4\}$ , o número mínimo de divisões

$D_{Min} \in \{2, 5, 10\}$  e o número de estimadores  $N_{est} \in \{200, 800, 1000\}$ . Já para o *Kernel Ridge*, foi utilizado  $\alpha \in \{0.001, 0.01, 0.1, 1\}$ . Para o *Gradient Boosting Regressor*, utiliza-se o número de estimadores  $N_{est} \in \{500, 1000, 2000\}$ , a taxa de aprendizado  $L_r \in \{0.001, 0.01, 0, 1\}$ , a profundidade máxima  $P_{max} \in \{1, 2, 4\}$  e a fração de amostras  $F_{amostra} \in \{0.5, 0.75, 1\}$ . No *Histogram-based Gradient Boosting Regression Tree* utiliza-se a taxa de aprendizado  $L_r \in \{0.001, 0.01, 0, 1\}$ , a profundidade máxima  $P_{max} \in \{-1, 1, 2, 4\}$ , e o número máximo de folhas  $L_{Max} \in \{-1, 1, 2, 4\}$ . Tanto para o *LightGBM* quanto para o *Extreme Gradient Boosting Regressor* utiliza-se o número de estimadores  $N_{est} \in \{100, 500, 1000\}$ , a taxa de aprendizado  $L_r \in \{0.01, 0.05, 0, 1\}$  e a profundidade máxima  $P_{max} \in \{3, 6, 10\}$ .

## 4.2. Separação dos dados

Para realizar o treinamento dos dados é necessário fazer uma separação em treino e teste. Baseando-se nisso, foi escolhido realizar esta separação de acordo com a data das transações, sendo que as 80% transações mais antigas serão utilizadas para o treinamento e validação dos modelos, e o restante para testes, tanto para os modelos de aprendizado de máquina quanto para os probabilísticos da literatura. Estes períodos podem ser chamados também de período de calibração (treino e validação) e observação (teste). Para criar os modelos, cada semana é considerada um período de tempo, e as transações são agrupadas semana a semana.

Portanto, após a etapa do processamento usando RFM, cada linha nos dados representa um cliente para um período (i.e., semana) de tempo específico, sendo que cada uma destas representações possui seus atributos extraídos do RFMT levando em consideração os períodos decorrentes até o período especificado.

## 4.3. Métricas de avaliação

Para todos os modelos, foram utilizadas as mesmas métricas de validação, sendo que a principal consiste no Erro Quadrático Médio (*Mean Squared Error* ou MSE) - que consiste na diferença média quadrática entre os valores estimados e o valor real. Outra métrica utilizada consiste no Erro Médio Absoluto (*Mean Absolute Error* ou MAE). O erro absoluto também é utilizado, que consiste na diferença entre o valor estimado e o valor real.

## 4.4. Estimativa do número transações

Para estimar o número de transações, foram utilizados os atributos fornecidos pelo RFMT, além de qual o horizonte de tempo esperado para realizar a previsão. O alvo deste modelo consiste no número de transações que este cliente irá realizar no futuro. Para isto, utiliza-se 70% dos dados do período de calibração para treino, e os 30% restante para validação com o objetivo de definir qual dos algoritmos e respectivos hiperparâmetros obtêm um melhor desempenho. Com isto, obtêm-se o modelo que mais represente o comportamento dos dados, e os resultados estão na Tabela 1.

Todos os modelos apresentaram bons resultados, porém, o que apresentou um melhor resultado em todas as bases foi o *XGBRegressor*. Por esta razão, ele foi escolhido para realizar a predição do número de transações para as 3 bases de dados.

Modelo	MSE na Base CDNOW	MSE na Base Olist	MSE na Base OpenCDP
LassoCV	0,9227	0,0212	0,2431
ElasticNet	0.9229	0,0213	0,2428
RandomForestRegressor	0.6820	0,0206	0,2195
GradientBoostingRegressor	0.5912	<b>0,0201</b>	0,2034
HistGradientBoostingRegressor	0.6436	0,0213	0,2308
XGBRegressor	<b>0.5600</b>	0,0202	<b>0,1937</b>
LGBMRegressor	0.6207	0,0208	0,2240

**Tabela 1. Comparação de Modelos em Relação ao Erro Quadrático Médio na Predição de Transações**

#### 4.5. Estimativa do valor monetário

Para o cálculo do valor monetário por transação, foram utilizados os mesmos atributos do modelo de transações, sendo que a diferença consiste no alvo. Neste caso, a variável alvo é a média dos valores de transação total do cliente, considerando os dados do futuro. Da mesma forma que na estimativa de transações, é feita a separação dos 70% dos dados de calibração para obter o modelo que melhor representa o comportamento dos dados, sendo testados diversos modelos, e os resultados encontrados estão na Tabela 2.

Modelo	MSE na Base CDNOW	MSE na Base Olist	MSE na Base OpenCDP
LassoCV	120,1387	465,0415	238,6791
ElasticNet	120,1009	464,9450	238,4658
RandomForestRegressor	102,1443	456,3697	221,2709
GradientBoostingRegressor	<b>99,3803</b>	457,5863	222,2082
HistGradientBoostingRegressor	142,3978	875,0773	323,0098
XGBRegressor	99,7793	<b>452,3832</b>	<b>221,0846</b>
LGBMRegressor	109,5694	460,6623	292,1478

**Tabela 2. Comparação de Modelos em Relação ao Erro Quadrático Médio na Predição do Valor Monetário**

O *XGBRegressor* obteve os melhores resultados para as bases Olist e OpneCDP, tendo tido o segundo melhor resultado para a base CDNOW. Como a diferença nesta última base foi muito pequena, o *XGBRegressor* foi selecionado para as três bases.

## 5. Avaliação

Nesta seção são apresentados os resultados obtidos neste trabalho, nas situações de cálculo de avaliação do número de transações e valor monetário esperado por cada transação, além do cálculo do CLV.

### 5.1. Baselines

Para comparar os resultados, as soluções base de Pareto e *BG/NBD* foram usadas para estimar o número de transações, e a solução *Gamma-Gamma* para estimar o valor monetário. Na Tabela 3 estão descritos os parâmetros das distribuições probabilísticas obtidas após o treinamento para cada uma das bases de dados. Nos modelos de Pareto e

BG/NBD, necessitam-se de três variáveis como entrada, a recência, a frequência e o  $T$ . Já no modelo *Gamma-Gamma*, é necessário o valor monetário e a frequência de compras do usuário como entrada.

Modelo	Pareto				BG/NBD			Gamma-Gamma			
	Alpha	Beta	R	S	A	Alpha	B	R	P	Q	S
Base CDNOW	16,60	9,80	0,62	0,43	0,10	6,43	0,25	0,27	4,07	0,93	3,79
Base Olist	150,75	5,67	0,23	0,54	0,01	5,55	0,00	0,01	3,92	0,82	3,62
Base OpenCDP	11,49	48,95	0,38	0,00	0,01	4,57	0,02	0,19	3,92	0,82	3,62

Tabela 3. Parâmetros dos Modelos

## 5.2. Resultados

Pode-se observar na Figura 2 o número de transações que foram observadas no período de treino (calibração) no eixo X, e a média de transações previstas no eixo Y. É possível observar que, o modelo de aprendizado de máquina (i.e., Solução) teve um desempenho melhor, no sentido de que ele identifica o comportamento dos clientes de maneira mais semelhante aos valores reais. O mesmo é válido para o Erro Quadrático Médio (MSE) como mostra a Tabela 4.

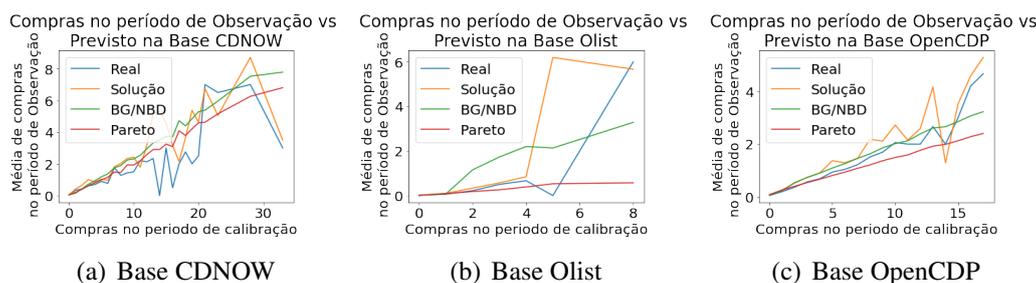


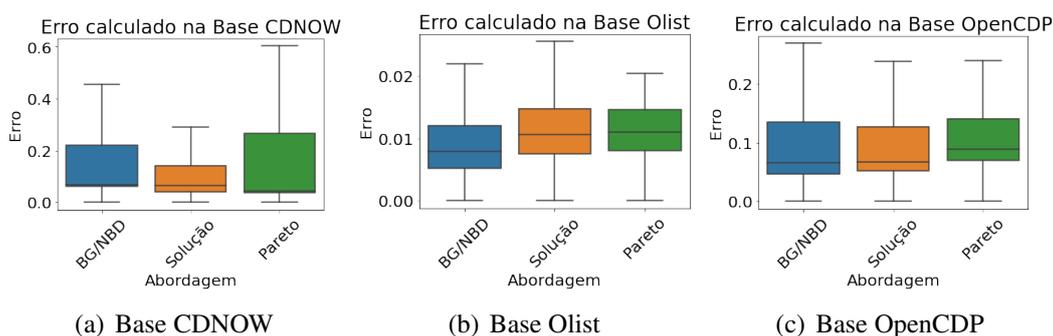
Figura 2. Transações no período de treino por compras previstas

Modelo	MSE na Base CDNOW	MSE na Base Olist	MSE na Base OpenCDP
Pareto	0,3918	0,0108	0,105
BG/NBD	0,4172	0,0113	0,106
Solução Proposta	<b>0,3730</b>	<b>0,0105</b>	<b>0,095</b>

Tabela 4. Comparação de Modelos de Predição do Número de Transações

Em relação ao erro absoluto (MAE), pode-se observar na Figura 3 que todas as abordagens conseguiram um bom resultado, mas ainda assim a solução proposta, utilizando aprendizado de máquina, se sobressaiu em relação às outras em duas das três bases.

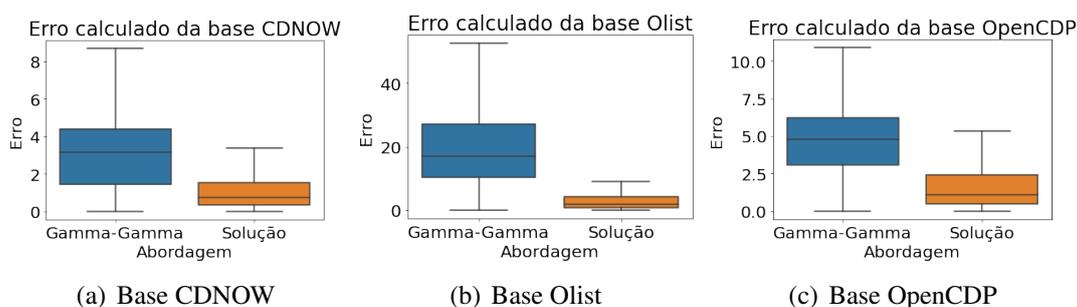
Já em relação ao valor monetário médio por transação, pode-se observar na Tabela 5, que o modelo proposto de aprendizado de máquina obteve resultados significativamente mais próximos do correto, com erros médios absolutos bem menores. Isto pode ser observado mais precisamente na Figura 4, onde o erro absoluto é mostrado com o modelo de aprendizado de máquina com resultados melhores.



**Figura 3. Erro absoluto entre os modelos de predição de transações**

Modelo	MAE na Base CDNOW	MAE na Base Olist	MAE na Base OpenCDP
Gamma Gamma	4,10299	26,88304	6,27010
Solução Proposta	<b>1,4830</b>	<b>6,6965</b>	<b>2,7155</b>

**Tabela 5. Comparação de Modelos de Predição de valor médio por transação**



**Figura 4. Erro absoluto entre os modelos de predição de valor monetário médio por transação**

### 5.3. Cálculo do CLV

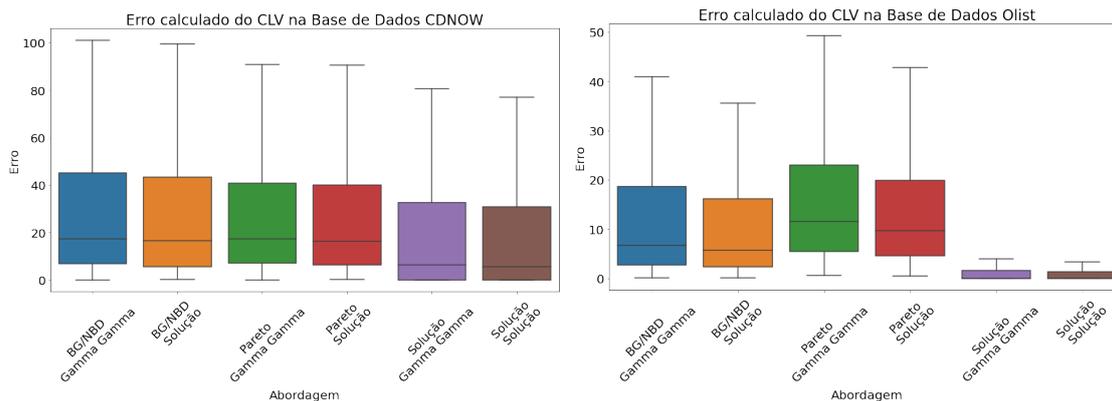
Para se obter o real valor do CLV do cliente, foi utilizado o número de transações que ocorreram no período de testes, e multiplicado pelo valor médio por transação calculado utilizando todas as transações do cliente, descontando a inflação, conforme equação 1. Foram testadas todas as seis combinações com modelos probabilísticos e de aprendizado de máquina.

Na Tabela 6 pode-se observar os resultados da utilização de cada um dos modelos de predição de transações em combinação com os de predição monetária, e os erros obtidos. Pode-se observar que, quando as duas soluções propostas utilizando aprendizado de máquina são usadas, todos os resultados foram melhores.

Pode-se observar que o erro calculado na junção da utilização de cada um dos modelos na Figura 5. Vemos que a utilização dos modelos de aprendizado de máquina proposto neste trabalho no cálculo do número de transações esperadas alcançaram um resultado significativamente melhor. Já no cálculo do valor monetário, todos os modelos apresentam resultados bem semelhantes, porém, o erro é um pouco maior quando utilizado o Gamma-Gamma.

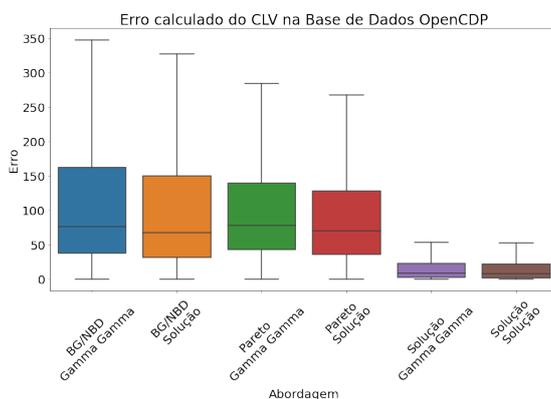
Modelo de Predição de Transações	Modelo de Predição Monetária	Base CDNOW		Base Olist		Base OpenCDP	
		MSE	MAE	MSE	MAE	MSE	MAE
BG/NBD	Gamma-Gamma	5025,7020	38,1459	15552,7530	34,0561	46915,0220	131,7316
Pareto/NBD	Gamma-Gamma	3571,3690	33,8526	3873,5680	24,5138	3025,2570	114,4087
Solução	Gamma-Gamma	2454,3811	24,3039	3313,0941	9,4711	1274,3249	18,5551
BG/NBD	Solução	4828,3215	36,6086	13705,5407	31,1360	40317,2660	120,5055
Pareto/NBD	Solução	3460,7106	32,6195	3611,1815	22,1368	25454,5611	103,7789
Solução	Solução	<b>2360,0799</b>	<b>23,7094</b>	<b>3256,7795</b>	<b>9,1423</b>	<b>1245,7343</b>	<b>18,0845</b>

Tabela 6. Comparação de Modelos



(a) Erro absoluto do CLV na base da CDNow

(b) Erro absoluto do CLV na base da Olist



(c) Erro absoluto na do CLV base do OpenCDP

Figura 5. Erro absoluto entre o cálculo do CLV

## 6. Considerações Finais

Neste trabalho, foi proposta uma solução para o cálculo do *Customer Lifetime Value* (CLV) usando aprendizado de máquina. Foi feita uma comparação da proposta com modelos consolidados na literatura. Foi possível observar que em relação ao número de transações previstas em uma situação real, os modelos de aprendizado de máquina propostos se destacam em relação aos modelos da literatura. Em relação aos modelos monetários, foi visto que ambos os modelos apresentam resultados bem competitivos, apesar dos modelos de aprendizado de máquina se sobressaírem. Já em relação ao cálculo do CLV, foi apresentado a maneira como este trabalho considera este cálculo, levando em conta o número de transações esperados, e o valor esperado por transação. Foi feito o cálculo e a solução proposta apresentou resultados mais precisos em bases reais.

As principais limitações consistem nos dados encontrados, visto que para análises mais profundas são necessário uma quantidade maior de dados, além de possuir mais atributos. Em relação a trabalhos futuros, sugere-se melhorar os modelos de aprendizado de máquina, além de testar com outras bases de dados e contextos e além de utilizar modelos diferentes. Neste trabalho foram utilizados somente os atributos fornecidos pelo modelo RFMT, por motivos de realizar uma comparação justa com os modelos probabilísticos. Porém, é importante adicionar mais atributos com o objetivo de tornar o cálculo do CLV mais preciso. Também sugere-se adicionar mais etapas, como a predição de *Churn* e/ou as interações do cliente na etapa do cálculo do CLV, visando obter uma métrica mais realista e mais próxima da realidade.

## Referências

- Aeron, H., Kumar, A., and Moorthy, J. (2012). Data mining framework for customer lifetime value-based segmentation. *Journal of Database Marketing & Customer Strategy Management*, 19(1):17–30.
- Chan, C. C. H. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert Systems with Applications*, 34(4):2754–2762.
- Desirena, G., Diaz, A., Desirena, J., Moreno, I., and Garcia, D. (2019). Maximizing customer lifetime value using stacked neural networks: An insurance industry application. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 541–544.
- Dwyer, F. R. (1997). Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing*, 11(4):6–13.
- Ekinci, Y., Uray, N., and Ulengin, F. (2014). A customer lifetime value model for the banking industry: a guide to marketing actions. *European Journal of Marketing*, 48.
- Fader, P. S., Hardie, B. G., and Lee, K. L. (2005a). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4):415–430.
- Fader, P. S., Hardie, B. G. S., and Lee, K. L. (2005b). “counting your customers” the easy way: An alternative to the pareto/NBD model. *Marketing Science*, 24(2):275–284.
- Hiziroglu, A. and Sengul, S. (2012). Investigating two customer lifetime value models from segmentation perspective. *Procedia - Social and Behavioral Sciences*, 62:766–774. World Conference on Business, Economics and Management (BEM-2012), May 4–6 2012, Antalya, Turkey.
- Hughes, A. (1996). Boosting response with rfm. *Marketing Tools*, 5:4–7.
- Kahre, M. S., Tive, M., Babania, A., and Hesani, M. (2014). Analyzing the applications of customer lifetime value (CLV) based on benefit segmentation for the banking sector. *Procedia - Social and Behavioral Sciences*, 109:590–594.
- Khajvand, M., Zolfaghar, K., Ashoori, S., and Alizadeh, S. (2011). Estimating customer lifetime value based on rfm analysis of customer purchase behavior: Case study. *Procedia Computer Science*, 3:57–63. World Conference on Information Technology.
- Kotler, P. (1974). Marketing during periods of shortage. *Journal of Marketing*, 38(3):20–29.

- Mammadzada, A., Alasgarov, E., and Mammadov, A. (2021). Application of bg / nbd and gamma-gamma models to predict customer lifetime value for financial institution. In *2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT)*, pages 1–6.
- Méndez-Suárez, M. and Crespo-Tejero, N. (2021). Why do banks retain unprofitable customers? a customer lifetime value real options approach. *Journal of Business Research*, 122:621–626.
- Reinartz, W. J. and Kumar, V. (2000). On the profitability of long-life customers in a non-contractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, 64(4):17–35.
- Sargeant, A. (2003). Using donor lifetime value to inform fundraising strategy. *Nonprofit Management and Leadership*, 12:25 – 38.
- 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.
- Singh, S. and Jain, D. (2013). Measuring customer lifetime value: Models and analysis. *SSRN Electronic Journal*.
- Sun, Y., Liu, H., and Gao, Y. (2023). Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model. *Heliyon*, 9(2):e13384.
- Temor Qismat, Y. F. (2020). Comparison of classical rfm models and machine learning models in clv prediction. Master's thesis, Norwegian Business School BI Open, Oslo.
- 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.
- Vanderveld, A., Pandey, A., Han, A., and Parekh, R. (2016). An engagement-based customer lifetime value system for e-commerce. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, page 293–302, New York, NY, USA. Association for Computing Machinery.