**Introdução:**

Durante a última década, o mercado passou por mudanças tornando os clientes muito mais exigentes, pois passaram a utilizar diversas outras plataformas tecnológicas como *web sites* e dispositivos *wireless* para interagir com as empresas que por sua vez começaram a investir em plataformas multicanais e em estratégias para rete-los, uma vez que um cliente consegue rapidamente absorver informações de diversos sites diferentes (RANGASWAMY; BRUGGEN, 2005).

Essa retenção tornou-se fundamental e um direcionador no gerenciamento de relacionamento com o consumidor, *Customer Relationship Managment* (*CRM*) por parte das instituições pois a deserção de clientes impacta diretamente no lucro das instituições. Portanto torna-se crucial desenvolver um modelo de rotatividade eficaz e preciso para gerenciar com eficiência o relacionamento com o cliente ([PFEIFER](https://www-sciencedirect.ez67.periodicos.capes.gov.br/science/article/pii/S1094996804701024#!); [FARRIS](https://www-sciencedirect.ez67.periodicos.capes.gov.br/science/article/pii/S1094996804701024#!), 2004; PRASHANTH; DEEPAK; MEHER, 2021).

Dentre os modelos *estado-da-arte* da literatura para previsão de rotatividade de clientes (*client churn*) o *XGBoost* demonstra grande superioridade aos demais na predição correta de um grande número de possíveis desertores (*turners)* em comparação com a regressão logística, *Support Vector Machine (SVM)*, árvore de decisão e *Random Forests* (SHARMA; GUPTA; MOHIT GOEL, 2020).

Entretanto modelo baseados em árvores não são interpretativos como modelos de abordagem tradicionais e a regressão logística é o mais balanceado tanto em precisão quanto poder explicativo, pois é possível identificar a relação causal das variáveis (X) na rotatividade dos clientes (Y), sendo muito mais interpretativo gerencialmente (HILLS JR et al., 2020).

Contudo, os modelos de regressão logística simples não são capazes de lidar com dados longitudinais que estão amplamente disponíveis nos bancos de dados das empresas devido à sua natureza assíncrona, bem como a comum existência de *missing values*, a forma encontrada por pesquisadores para adequá-los ao modelo é por meio de agregação ou retangularização transformando-os em dados estáticos. Já modelos multiníveis são capazes de lidar com os problemas de dados longitudinais (CHEN; FAN; SUN, 2012) e (JESKE; LI; WONG, 2012).

Partindo desse pressuposto o grande potencial de uma modelagem de regressão logística multinível é evidente. E em uma busca sistemática nas bases de dados relevantes da *SCOPUS*, *Engineer Village*, *IEEX* e *Web of Science* demonstraram o crescimento vertiginoso desta técnica (Figura 1), contudo é bastante utilizada nas áreas da saúde e ciências sociais (Figura 2), mas muito insipiente na área de negócios (Figura 3) demonstrando um grande potencial a ser explorado.

Figura 1 – Estudos de Regressão Logística Multinível ao longo dos anos

Interface gráfica do usuário

Descrição gerada automaticamente

Fone: Scopus, 2021

Figura 2 – Estudos de Regressão Logística Multinível por categoria de estudo

Gráfico, Gráfico de pizza

Descrição gerada automaticamente

Fone: Scopus, 2021

Figura 3 – Estudos de Regressão Logística Multinível focado na área de negócios

Interface gráfica do usuário, Gráfico

Descrição gerada automaticamente

Fone: Scopus, 2021

Quando partimos para a problemática de *client churn* um único artigo apresentado na 11ª conferência internacional de Computação Social e Media Social aborda a problemática utilizando uma técnica multinível. Contudo, o modelo utilizado por (IWATA; OTAKE; NAMATAME, 2019) era uma regressão logística hierárquica baeysiana e não uma modelagem logística baseada em Modelos Lineares Generalizados Mistos (*Generalized Linear Mixed Models* - *GLMM)* como o proposto por (JESKE; LI; WONG, 2012).

Dado este panorama e identificado o *GAP* literário, torna-se relevante a investigação de uma modelagem de regressão logística multinível baseada em *GLMM* voltada à predição de *client churn* e confrontá-la com outros modelos de classificação em *estado-da-arte*, uma vez que a regressão logística linear já é bem performática e pode ser aprimorada com efeitos aleatórios nos interceptos e nas suas inclinações, bem como a inclusão de dados longitudinais, assíncronos ou *missing values* em sua modelagem.

No contexto de dados longitudinais, os pontos de tempo em que as medidas são observadas geralmente variam entre os sujeitos. Devido a esta natureza assíncrona dos dados, é difícil usar diretamente essas medidas repetidas como recursos em regressão logística. Outro desafio com esta abordagem é como lidar com dados ausentes, um fenômeno que é comum em estudos longitudinais.

Modelos mistos são uma abordagem amplamente utilizada na modelagem de dados longitudinais (Verbeke e Molenberghs (2000)). Eles são capazes de lidar com medições assíncronas, e é bem conhecido que os parâmetros estimados e os efeitos aleatórios são robustos para faltar ao acaso (MAR) padrões de dados que são comuns em estudos longitudinais. (Jeske, Li, Wong, 2012)

dados comportamentais longitudinais estão amplamente disponíveis em bancos de dados de empresas. Como usar os dados comportamentais longitudinais para melhorar a previsão da rotatividade do cliente é um desafio para os pesquisadores. Em alguns métodos, os dados comportamentais longitudinais são transformados em dados estáticos por meio de agregação ou retangularização. (Chen, Fan, Sun, 2012)

No contexto de dados longitudinais, os pontos de tempo em que as medidas são observadas geralmente variam entre os sujeitos. Devido a esta natureza assíncrona dos dados, é difícil usar diretamente essas medidas repetidas como recursos em regressão logística. Outro desafio com esta abordagem é como lidar com dados ausentes, um fenômeno que é comum em estudos longitudinais. (Jeske, Li, Wong, 2012)

Modelos mistos são uma abordagem amplamente utilizada na modelagem de dados longitudinais (Verbeke e Molenberghs (2000)). Eles são capazes de lidar com medições assíncronas, e é bem conhecido que os parâmetros estimados e os efeitos aleatórios são robustos para faltar ao acaso (MAR) padrões de dados que são comuns em estudos longitudinais. Essas características legais do mix modelos têm motivado o desenvolvimento de vários procedimentos de classificação para dados longitudinais (ver, por exemplo, James e Hastie (2001), James (2002), Wouters et al. (2007), Luts et al. (2012)). A ideia comum usada nesses procedimentos de classificação é representar as características de sujeitos pelos efeitos aleatórios específicos do sujeito obtidos a partir de um

Contudo, modelos de regressão logística simples não são capazes de lidar com dados longitudinais que estão amplamente disponíveis nos bancos de dados das empresas, a forma encontrada por pesquisadores para adequá-los ao modelo é por meio de agregação ou retangularização transformando-os em dados estáticos. (Chen, Fan, Sun, 2012).

Entretanto se a interpretabilidade das abordagens não for uma preocupação,

A necessidade de identificar os clientes mais propensos a comutar é de alta prioridade. Foi demonstrado que uma pequena mudança na taxa de retenção pode resultar em mudanças significativas na geração de receita.

É crucial desenvolver um modelo de rotatividade de clientes eficaz e preciso para gerenciar com eficiência o relacionamento com o cliente.

Ainda assim, as condições competitivas do mercado no mundo comercial exigem que as empresas mantenham as campanhas de seus clientes atualizadas, além de conquistar novos por meio de estudos de preço, pois o consumidor é um dos fatores mais importantes nas empresas, onde suas operações e planejamentos são focados e o trabalho é feito dentro da estrutura de satisfação ou conquista de novos. ([Gulnur Oztürk Sahinkaya](https://ieeexplore.ieee.org/author/37088947433); [Damla Erek](https://ieeexplore.ieee.org/author/37088948924); [Hilal Yaman](https://ieeexplore.ieee.org/author/37088948587); [Mehmet S. Aktas](https://ieeexplore.ieee.org/author/37078476700), 2021).

Introdução:

Durante a última década, os clientes ficaram familiarizados em usar diversas outras plataformas tecnológicas como web sites e dispositivos wireless para interagir com as empresas. Isso tornou-os muito mais exigentes e as empresas mais focadas em plataformas multicanais e em estratégias para rete-los, uma vez que possibilitou a visita em vários sites diferentes. (ARVIND RANGASWAMY AND GERRIT H. VAN BRUGGEN,2005).

A retenção de clientes tornou-se fundamental e um direcionador do CRM nas empresas uma vez que a deserção dos clientes impacta diretamente no lucro. Contudo a análise da elasticidade do valor do cliente em comparação com a vida útil desse relacionamento demonstra que essa elasticidade de retenção é igual à duração financeira do ativo do cliente. ([Phillip E.Pfeifer1](https://www-sciencedirect.ez67.periodicos.capes.gov.br/science/article/pii/S1094996804701024" \l "!)[Paul W.Farris2](https://www-sciencedirect.ez67.periodicos.capes.gov.br/science/article/pii/S1094996804701024" \l "!),2004)

Ainda assim, as condições competitivas de mercado no mundo comercial exigem que as empresas mantenham as campanhas de seus clientes atualizadas e conquistem novos clientes com estudos de preços, pois o cliente é um dos fatores mais importantes nas empresas, onde suas operações e planejamentos são focados e o trabalho é feito dentro da estrutura de satisfação ou conquista de novos ([Gulnur Oztürk Sahinkaya](https://ieeexplore.ieee.org/author/37088947433); [Damla Erek](https://ieeexplore.ieee.org/author/37088948924); [Hilal Yaman](https://ieeexplore.ieee.org/author/37088948587); [Mehmet S. Aktas](https://ieeexplore.ieee.org/author/37078476700), 2021).

[10] L. Breiman, “Random forests,”Machine learning, vol. 45, no. 1, pp.5–32, 2001.

[11] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,”inProceedings of the 22nd acm sigkdd international conference onknowledge discovery and data mining, 2016, pp. 785–794

Logistic regression Support vector machine Decision trees Random forests XGBoost