**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 modelos 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 *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; 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 e 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 e 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.

**Referências:**

CHEN Z. Y.; FAN Z. P.; SUN M. A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. **European Journal of Operational Research,** v. 223, n. 2, p. 461-472, 2012.

HILLS W.; DANIEL W.; LU M. Y.; SCHAER O.; ADAMS S. Modeling Client Churn for Small Business-to-Business Firms. **Systems and Information Engineering Design Symposium.** p. 1-7, 2020.

IWATA M.; OTAKE K.; NAMATAME T. Analysis of the Characteristics of Customer Defection on a Hair Salon Considering Individual Differences. **Social Computing and Social Media. Communication and Social Communities,** v.2, p. 378-391, 2019.

JESKE D. R.; LI, J.; WONG V. On the effectiveness of Mixed Model Based Logistic Classifiers for Longitudional Data. **Integration: Mathematical Theory and Applications**, v.3, n.3, p. 233, 2012.

PFEIFER P. E.; FARRIS P. W. The elasticity of customer value to retention: The duration of a customer relationship. **Journal of Interactive Marketing,** vol. 18, n. 2, p. 20-31, 2004.

RANGASWAMY A.; BRUGGEN G. H. V. Opportunities and challenges in multichannel marketing: An introduction to the special issue. **Journal of Interactive Marketing,** vol. 19, n. 2, p. 5-11, 2005.

SHARMA T.; GUPTA P., NIGAM V., GOEL M. Customer Churn Prediction in Telecommunications Using Gradient Boosted Trees. **International Conference on Innovative Computing and Communications**. Advances in Intelligent Systems and Computing, vol 1059, Springer, Singapore, 2020.