

# Predição de Satisfação de Clientes Utilizando Modelos de *Machine Learning* em Pesquisa *Net Promoter Score*

*Predicting of Customer Satisfaction Using Machine Learning Models in a Net Promoter Score Research*

Edvaldo Acayaba<sup>1</sup>

 [orcid.org/0000-0002-7959-1733](https://orcid.org/0000-0002-7959-1733)

Rafaela Ramos<sup>1</sup>

 [orcid.org/0000-0003-0674-5807](https://orcid.org/0000-0003-0674-5807)

Thais Bispo<sup>1</sup>

 [orcid.org/0000-0003-4897-5512](https://orcid.org/0000-0003-4897-5512)

<sup>1</sup>Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil. E-mails: [essaj@ecomp.poli.br](mailto:essaj@ecomp.poli.br)

DOI: 10.25286/repa.v6i5.1763

Esta obra apresenta Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Como citar este artigo pela NBR 6023/2018: ACAYABA, E.; ALBERTINS, T.; RAMOS, R.; . Predição de Satisfação de Clientes Utilizando Modelos de Machine Learning em Pesquisa Net Promoter Score. Revista de Engenharia e Pesquisa Aplicada, Recife, v.6, n. 5, p. 65-72, Novembro, 2021.

## RESUMO

A pesquisa de satisfação do cliente é uma ferramenta amplamente utilizada pelas empresas como um importante indicador de desempenho para demonstrar a performance da organização a partir do ponto de vista do consumidor. Conhecer o nível de satisfação do cliente sobre determinado produto ou serviço pode fornecer informações relevantes para que as empresas mensurem os diferentes níveis de experiência do cliente, contribuindo para aumentar a competitividade do negócio. Uma das metodologias que tem sido mais utilizada para obter resultados oriundos de pesquisas de satisfação de clientes é o *Net Promoter Score* (NPS) que mede se um cliente recomendaria ou não uma empresa. Este artigo tem o objetivo de avaliar a possibilidade de predição do grau de satisfação do cliente por meio de indicadores de nível de serviço logístico de uma empresa que utiliza a metodologia do NPS como ferramenta de pesquisa, a partir de uma base de dados de avaliação da satisfação de clientes. Neste estudo são apresentados alguns resultados obtidos através de experimentos realizados com modelos de aprendizagem de máquina, utilizando algoritmos como *BalancedBaggingClassifier*, *DecisionTreeClassifier*, *KNN* e *K-means*.

**PALAVRAS-CHAVE:** Pesquisa de Satisfação; *Net Promoter Score*; Previsão; Aprendizagem de Máquina.

## ABSTRACT

*The customer satisfaction survey is one of the most used tools by companies as an essential index to demonstrate the organization's performance from the customer's point of view. Knowing the level of customer satisfaction about a specific product or service can provide relevant information for companies to measure the different levels of customer experience, contributing to increasing the business's competitiveness. One of the methodologies most used to obtain results from customer satisfaction surveys is the Net Promoter Score (NPS) that measures whether a customer would recommend a company or not. This article aims to assess the possibility of predicting customer satisfaction through indicators of the level of logistical service of a company that uses the NPS methodology as a research tool from a database of customer satisfaction assessments. This study presents results from experiments carried out with machine learning models, using algorithms such as *BalancedBaggingClassifier*, *DecisionTreeClassifier*, *KNN*, and *K-means*.*

**KEY-WORDS:** Satisfaction survey; *Net Promoter Score*; Prediction; Machine Learning.

## 1 INTRODUÇÃO

A qualidade de um produto ou serviço compreende diversas características, não apenas suas qualidades técnicas [1]. Características que vão além das especificações como custo, entrega, atendimento das necessidades, também agregam valor ao produto ou serviço fornecido e podem ser determinantes na forma como uma marca consegue alcançar seus clientes, satisfazê-los e fidelizá-los.

Qualquer empresa que esteja comprometida com a qualidade do que oferta tem entre seus objetivos a obtenção de dados que avaliem o grau de satisfação e a experiência obtida pelo cliente. Esses dados geram informações que quando utilizados corretamente, podem apoiar decisões na busca pela maior produtividade e competitividade das organizações. Tanto a Academia quanto o mercado dispõem de uma série de metodologias que suportam a realização de pesquisas que tem como objetivo escutar o cliente para compreender qual a percepção dele sobre o que foi adquirido.

Podem ser citadas como as metodologias mais utilizadas para avaliar a satisfação do cliente as seguintes: a Escala Likert (1932), que é utilizada para medir concordância de pessoas a determinadas afirmações onde os números da escala indicam a posição e/ou quanto às respostas diferem entre si em determinadas características ou elementos [2]; o modelo SERVQUAL (1988), que mensura a qualidade do serviço por meio da identificação de lacunas existentes entre a expectativa gerada e a percepção obtida do serviço prestado [3]; e a Análise de Sentimentos (2012), também conhecida como mineração de opinião, é o campo de estudo que analisa as opiniões, sentimentos, avaliações, atitudes e emoções das pessoas em relação a entidades como produtos, serviços, organizações, indivíduos, eventos e seus aspectos” [4].

Além destas mencionadas, outra metodologia amplamente utilizada no mercado, e que será a base para os modelos experimentados neste artigo, é o *Net Promoter Score* (NPS). Esta metodologia foi desenvolvida para mensurar a satisfação dos clientes e medir a fidelidade desses usuários em permanecer clientes e ainda recomendar ou não determinada companhia, uma vez que um cliente satisfeito, segundo esta metodologia, tem grandes possibilidades de se tornar um promotor da marca, fazendo a divulgação espontânea de seus produtos e serviços

para outros possíveis clientes. Enquanto que os insatisfeitos podem gerar o efeito inverso, interferindo inclusive na reputação de marca ao compartilharem as suas insatisfações [5].

Este artigo busca prever qual será a resposta do cliente a uma pesquisa de satisfação do NPS, a partir dos indicadores de nível de serviço logístico de uma empresa - como existência de reclamação pós-entrega, pedido atendido no prazo, disponibilidade de agenda para pedido, e pedido pendente - e a probabilidade do cliente ser um promotor, detrator ou neutro.

A aplicação de técnicas de Aprendizagem de Máquina supervisionadas é bastante útil na criação de modelos preditivos quando há disponibilidade de bases rotuladas. Neste estudo foram utilizados para treinamento do modelo os algoritmos: árvores de decisão (*DecisionTreeClassifier*), *K-Nearest Neighbor (KNN)*, *K-Médias (K-means)* e *BalancedBaggingClassifier*.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1. SATISFAÇÃO DO CLIENTE

Ao longo do tempo o conceito de qualidade passou por alguns marcos temporais que marcaram sua evolução. As classificações que predominam na literatura sobre o tema dividem a trajetória da qualidade em 04 tendências macro, a partir de algumas características apresentadas em cada época: iniciando pela Inspeção, que se restringia a verificação através de instrumentos de medição para garantir a padronização do produto; posteriormente alguns métodos estatísticos passaram a ser utilizados de forma complementar a inspeção, surgindo então o Controle Estatístico do Processo. Ambos eram utilizados para solucionar problemas já ocorridos. Um pouco mais tarde, e com uma postura mais proativa, surge a Garantia da Qualidade, onde o problema também já ocorreu, mas a solução proposta buscava impedir novas falhas no processo. E, finalmente é desenvolvido o conceito de Gestão da Qualidade Total, com uma abordagem sistêmica integrada ao planejamento estratégico das organizações em busca da melhoria contínua, com foco nas necessidades e exigências do cliente e do mercado [6].

O reconhecimento das necessidades do cliente e a busca para atender tais necessidade através da melhoria contínua de produtos e processos, poderia levar a presumir a qualidade como antecedente da satisfação, onde a satisfação

resultaria de um acúmulo de avaliações e percepções passadas de consumo, estando ambos os conceitos vinculados [7].

## 2.2. PESQUISA DE SATISFAÇÃO DE CLIENTES

Historicamente, os primeiros registros que tratam da Pesquisa de Satisfação de Clientes começam a partir da realização de uma conferência realizada em Chicago sobre o tema, em 1976, apoiada pelo *Marketing Science Institute da National Science Foundation*.

Embora esta conferência seja considerada um marco histórico para a temática, alguns anos antes, em 1972 e 1973, após a publicação de estudos realizados pelos pesquisadores Pfaff (sobre índice de satisfação do consumidor do Departamento de Agricultura Americano), e Olshavsky & Miller e Anderson, que examinaram a desconformação de expectativas e sua influência no grau de desempenho percebido, que começou a se constituir a base para muitos dos fundamentos teóricos e experimentos realizados posteriormente nesse campo [8].

Na década de 80 iniciou-se uma conscientização por parte das empresas quanto à importância de satisfazer os clientes, especialmente a partir do conhecimento da relação entre satisfação e comportamentos subsequentes à compra tais como a lealdade, recompra, propaganda favorável e pelo advento dos programas de qualidade total [9].

A satisfação do cliente possui dois principais conceitos: a satisfação específica em uma transação, como uma medida individual que avalia uma experiência particular com certo produto ou serviço, e a satisfação acumulada, que descreve a experiência total de consumo com um produto ou serviço [9] [10].

E para a medição dessa satisfação a academia e o mercado, ao longo dos anos, vem desenvolvendo as mais diversas técnicas e escalas para medição e avaliação do nível de satisfação dos clientes. A seleção das práticas de pesquisa mais adequadas para cada negócio e as análises dos resultados gerados, podem retroalimentar as organizações transformando os resultados em possíveis fatores decisivos na busca pela melhoria dos processos, E consequentemente, obter níveis de satisfação cada vez mais elevados.

## 2.3. METODOLOGIA NET PROMOTER SCORE (NPS)

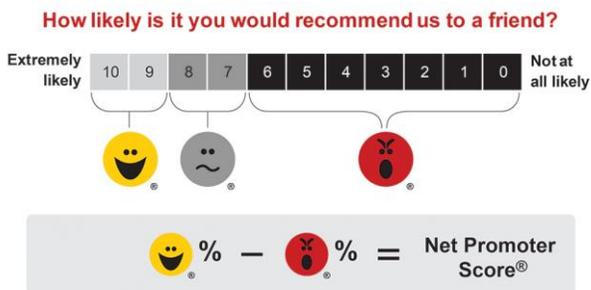
A metodologia do *Net Promoter Score* foi desenvolvida ao longo de dois pelo autor e pesquisador de negócios Frederick F. Reichheld. Esta técnica de pesquisa de satisfação foi publicada pela primeira vez em 2003 no artigo *The One Number You Need to Grow* (Um número que você precisa para crescer) na revista *Harvard Business Review*.

O *Net Promoter Score* foi desenvolvido ao longo de dois anos pelo Reichheld, e consiste numa abordagem mais simples para a pesquisa do cliente, diretamente ligada aos resultados de uma empresa, substituindo as ferramentas de pesquisas por uma única pergunta: "Qual é a probabilidade de recomendar a nossa empresa/produto/serviço?".

O cálculo é baseado nas respostas a esta pergunta, cuja pontuação é mais frequentemente realizada numa escala de 0 a 10. Aqueles que respondem com uma pontuação entre 9 e 10 são chamados de Promotores, e são considerados propensos a apresentar comportamentos que podem criar valor para o bem/serviço/negócio/marca, tais como possíveis compras adicionais, fidelidade para manter o cliente por mais tempo, fazer referências positivas para outros clientes em potencial. Aqueles que respondem com uma pontuação de 0 a 6 são identificados como Detratores, e estariam classificados como clientes menos propensos a apresentar comportamentos de criação de valor. Os clientes que respondem a pesquisa com resultados entre 7 e 8 são rotulados como Neutros (ou Passivos) pois seu comportamento pode aumentar ou diminuir o número de promotores e dos detratores.

O *Net Promoter Score* é calculado subtraindo a porcentagem de clientes Detratores da porcentagem de clientes Promotores. Os clientes Neutros contam para o número total de entrevistados, e embora não contribuam diretamente no resultado do NPS, podem impactar significativamente nos custos de serviços da empresa [5]. A Figura 1 ilustra esse modelo.

**Figura 1:** Metodologia *Net Promoter Score*.



Fonte: <https://www.netpromotersystem.com> [11].

### 3 METODOLOGIA

Esta seção descreve as metodologias utilizadas para definição dos parâmetros e para a criação dos modelos testados para este projeto.

#### 3.1 DESCRIÇÃO DA BASE DE DADOS

A base de dados analisada neste projeto apresenta avaliações realizadas pelos clientes dos serviços de logística de uma empresa do setor de construção civil, coletadas durante o período de janeiro a dezembro de 2020.

De acordo com a metodologia NPS, os clientes que responderam à pesquisa com notas entre 9 e 10 foram classificados como promotores e os que responderam com notas entre 0 e 6 receberam a denominação de detratores. Aqueles clientes que avaliaram o serviço entre 7 e 8 foram classificados como neutros. A base utilizada contém 15.911 pesquisas. A Figura 2 ilustra a distribuição da base.

**Figura 2:** Distribuição por classe.



Fonte: Os Autores.

Para este projeto foram analisadas as 05 (cinco) variáveis descritas no Quadro 1: reclamação, reclamação atendida no prazo, percentual de pedidos atendidos no prazo, percentual de disponibilidade de agenda para pedidos e

percentual de pedidos pendentes. Para análise foram considerados apenas pedidos e reclamações anteriores a data da pesquisa de satisfação realizada.

**Quadro 1:** Variáveis consideradas para os modelos testados e suas descrições.

VARIÁVEIS	DESCRIÇÃO
Reclamação	Variável numérica que corresponde a quantidade de reclamações pós-entrega antes da realização da pesquisa.
% Reclamação atendida no prazo	Variável numérica entre 0 e 1 mensurada da forma abaixo: <u>Reclamações atendidas no prazo</u> / Reclamações atendidas total
% Pedidos atendidos no prazo	Variável numérica entre 0 e 1 mensurada da forma abaixo: <u>Pedidos atendidos no prazo</u> / Pedidos atendidos total
% Disponibilidade de agenda	Variável numérica entre 0 e 1 mensurada da forma abaixo: <u>Pedidos atendidos no prazo</u> / Pedidos atendidos total A disponibilidade de agenda para cada pedido implantado considera o agendamento em até D + 3.
% Pedidos pendentes	Variável numérica entre 0 e 1 mensurada da forma abaixo: <u>Pedidos pendentes</u> / Pedidos atendidos total, Pedidos pendentes são aqueles que ultrapassam 24h de atraso.

Fonte: Os Autores.

#### 3.2 PRÉ-PROCESSAMENTO DOS DADOS

No pré-processamento observou-se a distribuição dos dados e o número de dados faltantes por atributo, demonstrados no Quadro 2.

Por critério de seleção e para simplificação do problema, a variável reclamação atendida no prazo foi retirada do treinamento, pois verificou-se que na base de reclamações constam 5.002 registros e em apenas 1.567 há a informação se a reclamação foi atendida dentro do prazo ou não, portanto não é uma variável relevante para o problema.

Além disso, as 3.534 pesquisas com dados faltantes em 03 atributos também foram desconsideradas da análise. Assim, para a análise do problema foram utilizadas as 12.374 pesquisas válidas.

**Quadro 2:** Escala e valores faltantes dos atributos.

Atributos	Escala	Valores Faltantes
Reclamação	0 a 35	0
% Reclamação atendida no prazo	0 a 14	3.435
% Pedidos atendidos no prazo	0 a 1	3.534
% Disponibilidade de agenda	0 a 1	3.534
% Pedidos pendentes	0 a 1	3.534

Fonte: Os Autores.

A base de dados utilizada possuía variáveis as quais eram medidas com escalas distintas o que dificultava o treinamento do modelo. Para ajustar diferentes escalas entre os atributos foi feita a normalização dos dados.

As variáveis foram normalizadas por meio do algoritmo *MinMaxScaler* que retorna valores num intervalo entre 0 e 1.

### 3.3 METODOLOGIA EXPERIMENTAL

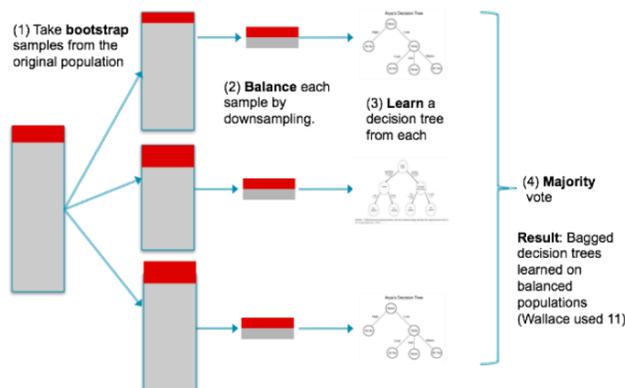
A base foi dividida em treino e teste, seguindo a proporção de 9.280 (75%) para treino e 3.094 (25%) teste. Para a escolha dos melhores parâmetros optou-se pelo algoritmo Gridsearch para cada um dos 04 modelos utilizados: *Balanced Bagging Classifier*, árvore de decisão (*Decision Tree Classifier*), *K-Nearest Neighbor* (KNN) e *K-Medias* (K-means).

Conforme Figura 2, é possível observar um desbalanceamento dos dados por classe, contendo 67% dos clientes como promotores, e 33% como neutros e detratores.

Para balancear a base de dados, optou-se por utilizar o algoritmo *Balanced Bagging Classifier*. Esse algoritmo consiste em pegar recortes da base contendo todas as classes, realizar vários balanceamentos igualitários e rodar um algoritmo nesses recortes, para então através de votação retornar um resultado a partir de amostras balanceadas [12][13]. A Figura 3 mostra o processo utilizado com a aplicação do algoritmo *Balanced Bagging Classifier*.

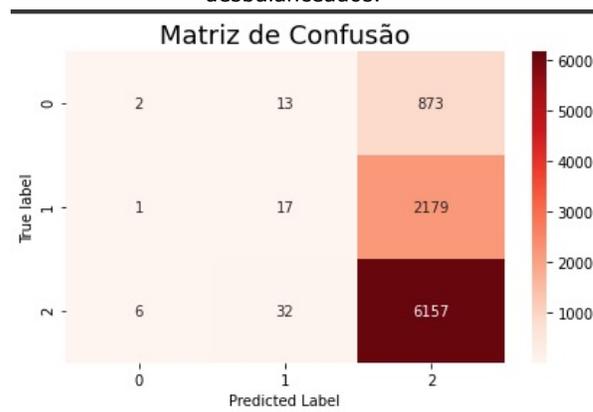
Foram testadas também outras técnicas como *oversampling* e *undersampling*, que consistem em aumentar a classe minoritária e diminuir a classe majoritária, porém os resultados foram insatisfatórios com os modelos apresentando *overfitting* e matrizes de confusão onde todos eram classificados como promotores, conforme Figura 4.

**Figura 3:** Processo realizado com a aplicação do algoritmo *Balanced Bagging Classifier*.



Fonte: [12].

**Figura 4:** Matriz de Confusão com os dados desbalanceados.



Fonte: Os Autores.

O segundo algoritmo utilizado no modelo foi a árvore de decisão que a partir de um nó raiz subdivide a base em diferentes ramificações. Cada ramificação gera um nó interno, que vem de um outro nó superior e pode gerar até dois outros nós, chamados de folhas e nós terminais. Neste algoritmo, cada folha é uma classe e contém um atributo para realização de testes que subdividem cada nó [14].

A maior vantagem de usar esse algoritmo é que ele pode usar regras de decisão e subconjuntos diferentes em diferentes estágios da atividade de classificação, que podem ser visualizadas graficamente por meio da árvore de decisão gerada.

Foi utilizado ainda o *K-Nearest Neighbor* (KNN), este algoritmo pode ser usado quando todos os valores de atributo são contínuos, contudo, pode também ser modificado para lidar com atributos

categoricos. A ideia é estimar a classificação de uma instância invisível usando a classificação da instância ou instâncias que estão mais próximas dela, em algum sentido que deve ser definido [14].

Outro algoritmo utilizado foi o K-Médias (*K-means*). A grande vantagem desse algoritmo é a facilidade de visualização dos *n* clusters contidos na base, bem como a separação inter-clusters e intra-clusters. [13]

Para escolha dos melhores parâmetros de cada modelo, foi utilizado o algoritmo *GridSearch*, que retornou os valores conforme Quadro 3.

**Quadro 3:** Melhores parâmetros retornados para os modelos experimentados.

MODELO	PARÂMETROS
DecisionTree	'criterion': 'entropy' 'max_depth': 4, 'max_leaf_nodes': 4
K-Nearest Neighbor	{'metric': 'manhattan', 'n_neighbors': 13, 'weights': 'distance'}
K-means	{'max_iter': 1000, 'n_clusters': 3}
BalancedBagging	base_estimator=DecisionTree Classifier(), sampling_strategy='auto', replacement=False, random_state=0, n_estimators=100, n_jobs=1)

Fonte: Os Autores.

## 4 ANÁLISE DOS RESULTADOS

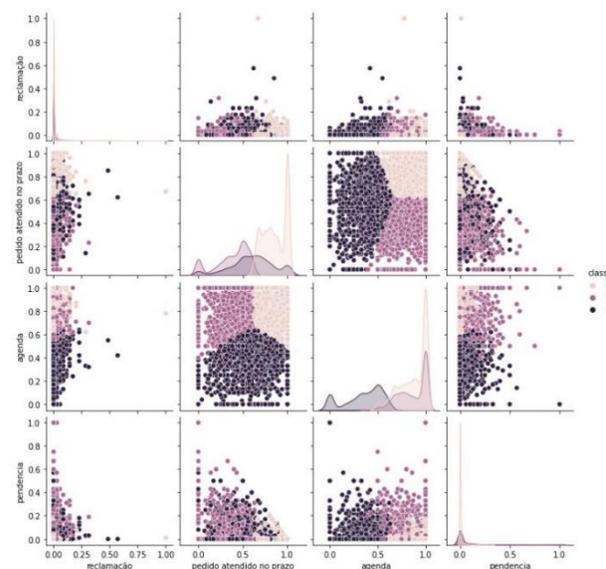
Para avaliação da distinção entre as classes promotores, detratores e neutros, foi utilizado o algoritmo *K-means* com *K*=3, sendo possível verificar as distâncias interclasses e intraclasses.

A Figura 5 ilustra a distribuição e a dificuldade de diferenciar as classes por meio dos atributos utilizados.

Também foram avaliados os modelos *Decision Tree*, *KNN* e *Balanced Bagging*, conforme parâmetros apresentados no Quadro 3.

Para todos os modelos utilizados foi setada uma mesma origem ou semente, garantindo que os resultados pudessem ser comparados do mesmo ponto de partida.

**Figura 5:** Distribuição das classes por variável utilizando o algoritmo *K-means*.



Fonte: Os Autores.

Na Tabela 1 é possível observar a acurácia obtida em cada um dos modelos. Por essa métrica o algoritmo com melhor resultado foi o *K-Nearest Neighbor*, com acurácia no treinamento de 0.789 e no teste de 0.795.

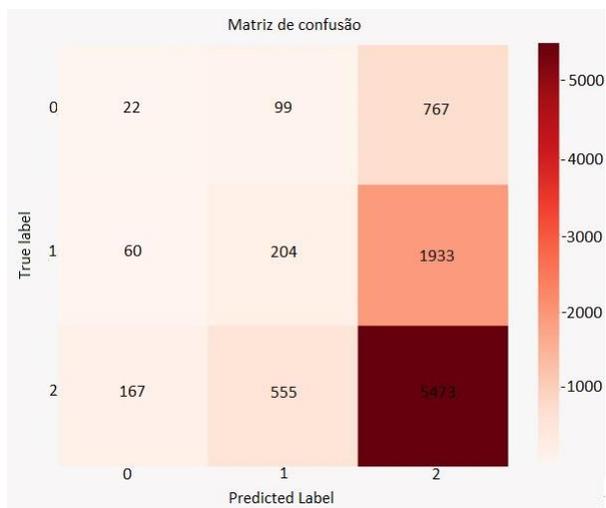
**Tabela 1:** Acurácia no treinamento e teste de cada modelo.

MODELO	TREINO	TESTE
Decision Tree	0.672	0.680
K-Nearest Neighbor	0.789	0.795
Balanced Bagging	0.599	0.593

Fonte: Os Autores.

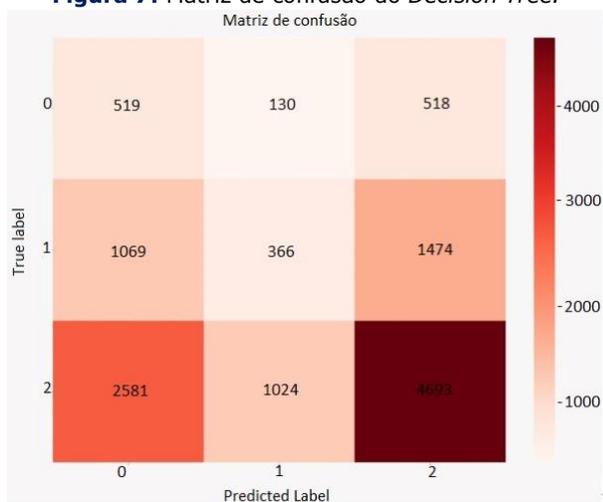
Também foi plotada a matriz de confusão de cada modelo (Figuras 6, 7 e 8) a fim de entender em quais classes o modelo acerta e em quais classes ele erra.

**Figura 6:** Matriz de confusão do *K-Nearest Neighbor*.



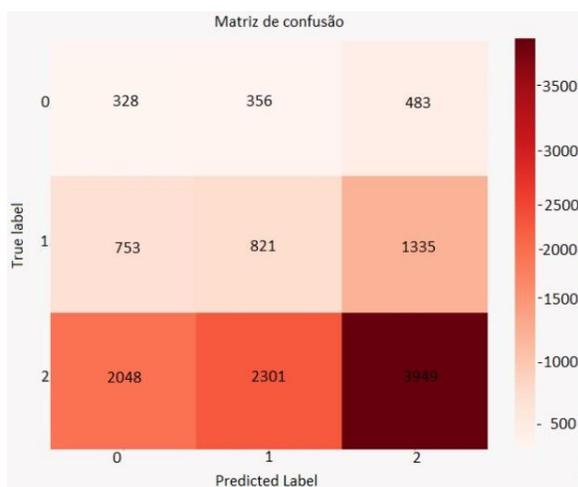
Fonte: Os Autores.

Figura 7: Matriz de confusão do *Decision Tree*.



Fonte: Os Autores.

Figura 8: Matriz de confusão do *Balanced Bagging*.



Fonte: Os Autores.

Além disso, as principais métricas de avaliação foram verificadas para cada modelo. É interessante notar que para a empresa o algoritmo é útil quando consegue prever corretamente as classes do maior número de pesquisas e que haja poucos erros entre as classes extremas (detratores e promotores). O custo de prever a avaliação de um cliente detrator como promotor é bem mais alto que de prever incorretamente um detrator como neutro.

Dos resultados obtidos, conforme as tabelas 2, 3 e 4, podemos observar que o *Decision Tree* classifica todos os promotores corretamente, porém apresenta um pior desempenho na classificação das demais classes com valores de *F-1 Score* entre 0.00 e 0.01, e *recall* entre 0.00 e 0.50. Pela matriz de confusão plotada para este modelo fica claro que o algoritmo classifica quase todos os clientes como promotores.

Já o algoritmo *KNN*, além de apresentar a melhor acurácia entre os 03 modelos (Tabela 4), também retornou valores de *precision*, *recall* e *F1-score* entre 0.68 e 0.86 para classe de promotores, enquanto o algoritmo *Balanced Bagging Classifier* apresentou desempenho menor para a mesma classe. Porém é válido ressaltar que o *K-NN* possui valores abaixo de 0.1 de *precision* e *recall*, ou seja, o modelo erra bastante nas classes detratores e neutros, o que do ponto de vista do negócio pode representar um erro grave.

Tabela 2: Métricas de avaliação do *KNN*.

KNN	PRECISION	RECALL	F1	SUP
Promotor	0.86	0.68	0.76	2667
Detrator	0.04	0.10	0.06	106
Neutro	0.09	0.20	0.13	321

Fonte: Os Autores.

Tabela 3: Métricas de avaliação do *Decision Tree*.

DT	PRECISION	RECALL	F1	SUP
Promotor	1.00	0.68	0.81	3090
Detrator	0.00	0.00	0.00	0
Neutro	0.00	0.50	0.01	4

Fonte: Os Autores.

Tabela 4: Métricas de avaliação do *Balanced Bagging*.

BB	PRECISION	RECALL	F1	SUP
Promotor	0.49	0.71	0.58	1448
Detrator	0.30	0.10	0.15	828
Neutro	0.29	0.26	0.27	818

Fonte: Os Autores.

## 5 CONCLUSÕES

A partir desse estudo, foi possível avaliar a possibilidade de predição do nível de satisfação de um cliente por meio de indicadores de nível de serviço logístico de uma empresa.

Para esse problema e em futuros estudos de base de pesquisa de satisfação é válido ressaltar que esse tipo de conjunto de dados costuma ser desbalanceado, tendo uma grande maioria de clientes satisfeitos e uma menor parcela dos pesquisados insatisfeitos ou parcialmente satisfeitos.

Embora o algoritmo *BalancedBaggingClassifier* seja recomendado para problemas de bases desbalanceadas, para a base de dados avaliada o algoritmo *KNN* retornou o melhor desempenho. A *DecisionTreeClassifier* ainda que seja uma ferramenta de fácil visualização e entendimento não foi capaz de classificar bem as diferentes classes e classificou quase todos as instâncias para a classe majoritária (promotores).

O algoritmo *K-means* também foi útil na avaliação da separabilidade entre as classes e possui fácil visualização. Por meio dele foi possível observar que as variáveis utilizadas, indicadores de nível de serviço logístico da empresa, não são capazes de diferenciar claramente um cliente detrator, neutro ou promotor. Como sugestão para novos trabalhos o uso de mais indicadores ou uma combinação desses pode apoiar na melhora do desempenho dos modelos.

Para futuros trabalhos considera-se interessante aprimorar o modelo através da combinação de diferentes algoritmos para um melhor desempenho na classificação das classes minoritárias (neutros e detratores), pois o custo de um cliente detrator previsto como promotor é muito alto para uma empresa que quer aumentar a fidelização dos seus clientes.

## REFERÊNCIAS

- [1] MIZUNO, Shigeru. **Company-Wide Total Quality Control**. Asian Productivity Organization, Tokyo, Japan, 1988.
- [2] LIKERT, R. **A Technique for the Measurement of Attitudes**. Archives of Psychology, 140: 1-55. 1932.
- [3] PARASURAMAN, A.; et al. **SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality**. Journal of F

Retailing, vol. 64, nº 1, p.12-40, New York University, Spring, 1988.

- [4] LIU, B.. **Sentiment Analysis and Opinion Mining**, Morgan & Claypool Publishers, 2012.
- [5] REICHELDT, F. F. **The one number you need to grow**. Harvard business review 81 12 (2003): 46-54, 124.
- [6] CARVALHO, Marly M. de C. Histórico da Gestão da Qualidade. In: **Gestão da Qualidade: Teoria e Casos**. Rio de Janeiro, Elsevier, 2005.
- [7] OLIVEIRA, M. C. **A relação causal entre qualidade e satisfação do usuário: Proposições para o marketing imobiliário**. UFSC, Florianópolis, 1999.
- [8] CHURCHILL, Gilbert A., and Carol Surprenant. **An Investigation into the Determinants of Customer Satisfaction**. Journal of Marketing Research, vol. 19, no. 4, 1982.
- [9] ROSSI, C. A. V.; SLOGO, L. A. **Pesquisa de satisfação de clientes: o estado-da-arte e proposição de um método brasileiro**. Revista de Administração Contemporânea, v. 2, n. 1, p. 101-125, 11.
- [10] BOULDING, W. et al. **A dynamic process model of service quality: From expectations to behavioral intentions**. Journal of Marketing Research, v. 30, p. 07-27, Feb. 1993.
- [11] NET PROMOTER SYSTEM. **Measuring Your Net Promoter Score**. Disponível em: <https://www.netpromotersystem.com/about/measuring-your-net-promoter-score>
- [12] FAWCETT, Tom. **Learning from Imbalanced Classes**. Silicon Valley Data Science, 2016.
- [13] BLASZCZYNSKI, Jerzy; STEFANOWSKI, Jerzy **Actively Balanced Bagging for Imbalanced Data**. Poznan University of Technology, Institute of Computing Science, 2017.
- [14] ZAKI, Mohammed J.; MEIRA JR., Wagner. **Data Mining and Machine Learning: Fundamental Concepts and Algorithms**. Cambridge University Press, 2020.
- [15] BRAMER, Max. **Principles of Data Mining**. Springer, London, 2007.