SIMPROFI Simpósio dos Programas de Mestrado Profissional 24 e 25 de novembro de 2021
Produção de Conhecimento em Programas de Mestrado e Doutorado Profissionais: Experiências e Desafios
Centro Paula Souza
GOVERNO DO ESTADO DE SÃO PAULO

Uma análise comparativa entre os métodos tradicionais e algoritmos de aprendizado de máquina para previsão de vendas no segmento varejista

Emerson Martins[1], Napoleão Verardi Galegale[2]

Resumo: As empresas varejistas, como sistemas produtivos, devem usar seus recursos de forma eficiente e tomar decisões estratégicas para obter receitas crescentes e estáveis, especialmente quando as condições de mercado estão ficando mais competitivas e com margens de lucro cada vez mais pressionadas. Desta forma, a previsão de vendas é crucial para manter a competitividade no segmento varejista, porém obter previsões imprecisas pode levar à escassez de estoque, ocasionando atrasos nas entregas e gerando insatisfação dos clientes, como também, podem elevar o estoque, aumentando o custo de armazenagem, forçando a "queima" de estoque através de campanhas promocionais, afetando diretamente a lucratividade. Prever a demanda de produtos e serviços e adequar a cadeia de suprimentos encontrando um equilíbrio sempre foi e continuará sendo um desafio no segmento varejista. Esta pesquisa tem como objetivo avaliar os principais métodos e identificar aquele que apresente maior precisão na predição de vendas. Com base em uma revisão integrativa da literatura (RIL), foram avaliados três principais métodos: séries temporais, redes neurais artificiais e algoritmos de aprendizado de máquina. Os resultados mostram que o aprendizado de máquina é mais adequado em termos de precisão, particularmente quando os modelos contêm variáveis exógenas e endógenas, além de permitir a identificação de padrões ocultos na demanda que podem ser usados para identificar tendências de mercado. Porém, em mercados com demandas constantes e poucas interferências externas, a sua utilização não se justifica pois, para estes casos, a utilização de séries temporais é mais simples e menos custosa.

Palavras-chave: Previsão de Vendas; Varejo, Aprendizado de Máquina, Séries Temporais, Sistemas Produtivos

Abstract: Retail companies, as production systems, must use their resources efficiently and make strategic decisions to obtain growing and stable revenues, especially when market conditions are becoming more competitive and profit margins are increasingly pressured. Thus, sales forecasting is crucial to maintain competitiveness in the retail segment, but obtaining inaccurate forecasts can lead to stock shortages, causing delays in deliveries and generating customer dissatisfaction, as well as increasing inventory, increasing

[1] Centro Paula Souza, emerson.martins@cpspos.sp.gov.br
[2] Centro Paula Souza, napoleao.galegale@cpspos.sp.gov.br

the cost of warehousing, forcing the “burn” of stock through promotional campaigns, directly affecting profitability. Forecasting the demand for products and services and adapting the supply chain by finding a balance has always been and will continue to be a challenge in the retail segment. This research aims to evaluate the main methods and identify the one with the greatest accuracy in sales prediction. Based on an integrative literature review (RIL), three main methods were evaluated: time series, artificial neural networks and machine learning algorithms. The results show that machine learning is more suitable in terms of accuracy, particularly when models contain exogenous and endogenous variables, in addition to allowing the identification of hidden patterns in demand that can be used to identify market trends. However, in markets with constant demands and few external interferences, its use is not justified because, for these cases, the use of time series is simpler and less costly.

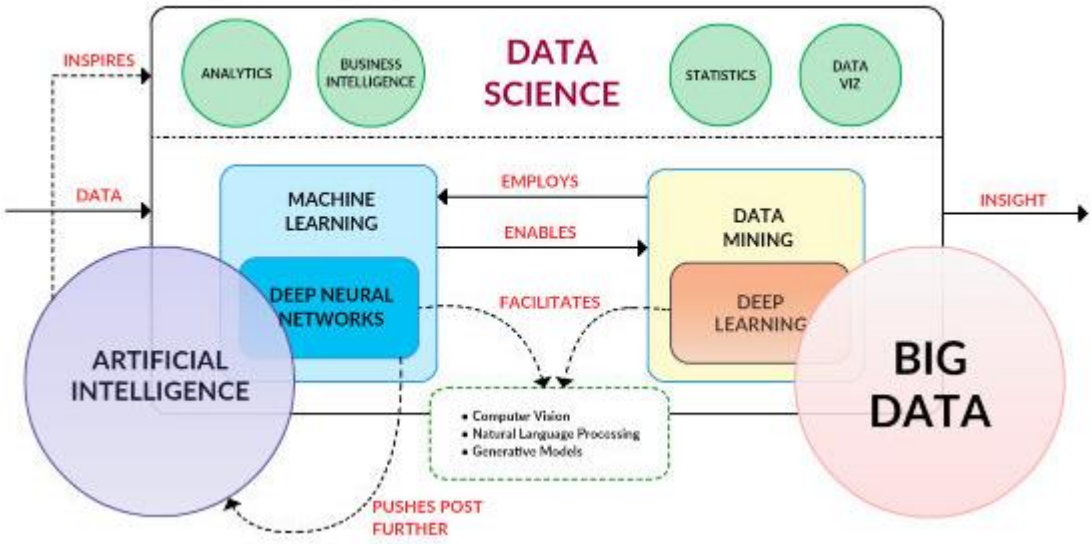
Keywords: Sales forecast; Retail, Machine Learning, Time Series, Productive Systems

1. Introdução

Nas duas últimas décadas o poder da computação evolui consideravelmente, neste contexto podemos citar: grande armazenamento de dados, processadores mais robustos, conexão mais rápida com a internet, entre outros exemplos. Problemas que pareciam ser extremamente complexos ou custosos de serem resolvidos, agora estão ao nosso alcance. Novas tendências como *Big Data*, *Cybersecurity*, Internet das Coisas (IoT) e *blockchain* surgiram, explorando conjuntamente os avanços tecnológicos mencionados acima. O IoT, que visa usar sistemas embarcados, incluindo sensores e atuadores, juntamente com a internet, para permitir o controle e o acesso imediato às informações em tempo real (Atzori, 2010; Cecchinel, 2014), representa um desses desafios, pois o relatório da “*Juniper Research*”, informa que em 2024 teremos mais de 83 bilhões de dispositivos e sensores conectados (IoT). Além disso, alguns destes dispositivos terão a capacidade de gerar quantidade de dados expressiva na ordem de *Zettabytes*, informações que podem ser valiosas para estratégia de uma empresa, portanto, a previsão de vendas não pode ignorar essas novas tendências; ela deve utilizá-la como suporte para vantagem competitiva.

Uma das técnicas recentes e populares que visa enfrentar esses novos desafios de negócios é o *Big Data Analytics* (DBA). Uma definição precisa de DBA é dada em Hofmann (2018). Em suma, é o alinhamento das técnicas de *Big Data* e *Machine Learning* (ML) para fornecer *insights* confiáveis para a tomada de decisões. A figura 1, representa a arquitetura geral e elementos em um modelo DBA. Podemos observar que o ML e *Big Data* se beneficiam uma da outra, pois podem ser acopladas para criar modelos mais completos. Além disso, o principal propósito do DBA é transformar informações em conhecimento útil.

Figura 1 – Arquitetura de um modelo *Big Data Analytics*



Fonte: Mayo, 2016

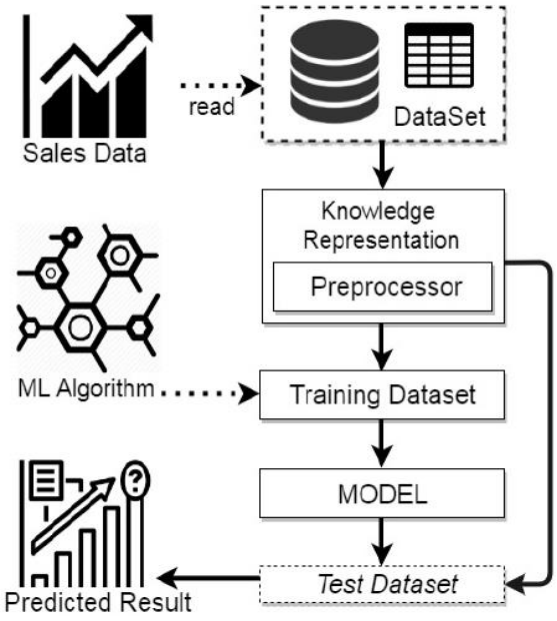
Por sua vez, análise preditiva engloba métodos que utilizam informações para criar modelos e realizar simulações que fornecerão *insights* sobre eventos futuros, permitindo que os executivos mais atentos consigam prever ações estratégicas que melhorem o desempenho da sua empresa. Por definição, os resultados obtidos através destas técnicas não são 100% precisos, pois nenhum método pode prever o futuro, sendo assim, uma boa análise preditiva é aquela que fornece os resultados mais precisos em um tempo razoável (CASTILHO *et al*, 2017).

Um dos usos comuns da análise preditiva nos negócios é a previsão de vendas – este ponto será abordado na seção 3 deste artigo – mas também há várias outras aplicações em domínios como: estimativa de custo, onde (LOYER *et al*, 2016) aplicaram técnicas de ML para estimar rapidamente o custo de fabricação de componentes de motor para aviões. Na avalição de desempenho, (FAN *et al*, 2013) utilizaram ML para estimar o desempenho da cadeia de suprimentos baseado no "*5 Dimensional Balanced Scorecard*" (5DBSC), com o objetivo de fornecer resultados rápidos e evitar avaliações tendenciosas de desempenho por parte dos gestores.

A abordagem do *machine learning* (ML) ou aprendizado de máquina pode ser descrita como o estudo de algoritmos de computador que se aprimoram automaticamente com a experiência. É tratado com uma subárea da inteligência artificial (IA). Algoritmos de aprendizado de máquina constroem um modelo baseado em dados de amostra, conhecidos como "dados de treinamento", a fim de fazer previsões ou decisões sem serem explicitamente programados para isso. Os algoritmos de aprendizado de máquina são usados em uma ampla variedade de aplicações, como filtragem de e-mail e visão computacional, onde é difícil ou inviável desenvolver algoritmos convencionais para realizar as tarefas necessárias (RASCHKA *et al*, 2017).

Os seres humanos aprendem através da experiência, usamos um processo de tentativa e erro para descobrir quais ações devem ser desencadeadas em determinadas circunstâncias. Isso nos permite fazer abstrações e construir conhecimento. O ML é de alguma forma semelhante, pode ser visto como algoritmos que têm como objetivo melhorar uma medida de desempenho, derivando automaticamente suas próprias regras e criando seus próprios modelos de decisão com base em determinadas informações (RASCHKA et al, 2017) e Mitchell (1997). Na figura 2 é demonstrado o diagrama desta arquitetura. Em ternos gerais, podemos identificar três tipos de métodos de aprendizagem: supervisionado, não supervisionado e aprendizado por reforço.

Figura 2 - Arquitetura ML



Fonte: Cheriyan *et al*, 2018

Neste contexto, o objetivo deste artigo é fornecer uma visão sobre as técnicas mais recentes de ML aplicadas à previsão de vendas no varejo por meio de uma RIL. A questão que norteou a pesquisa pode ser colocada da seguinte forma: A ML é mais adequada do que os métodos tradicionais de previsão de vendas no varejo, em termos de precisão, particularmente quando os modelos contêm variáveis exógenas e endógenas?

2. Metodologia de pesquisa

Para esta pesquisa do tipo descritiva e qualitativa, foi realizada uma RIL sobre a utilização de ML para previsão de vendas no segmento de varejo.

Foi utilizado o protocolo PRISMA-P, tal protocolo tem como objetivo apoiar os pesquisadores a melhorarem o relato de revisões sistemáticas e meta-

análises, filtrando o número de publicações com maior relevância ao tema pesquisado (MOHER *et al*, 2015).

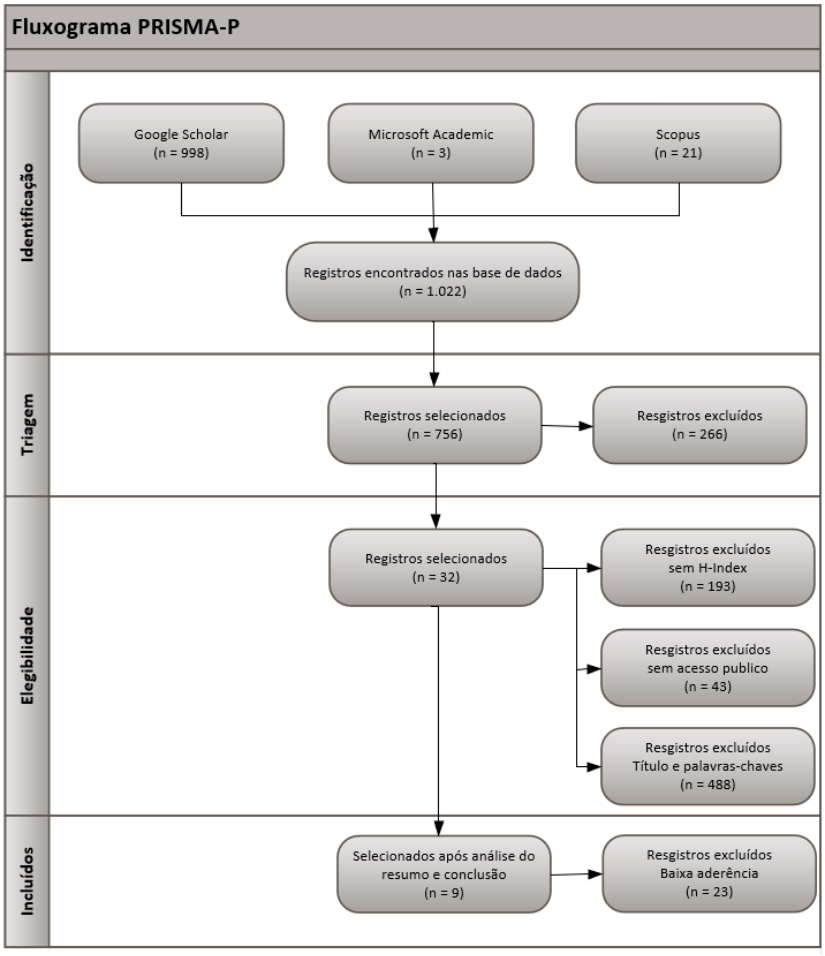
Na etapa de identificação, foi realizada busca das publicações nas bases de dados *Google Scholar*, *Microsoft Academic* e *Scopus*, com a seguinte *string* de pesquisa: ((“*Sales Forecasting*” OR “*Sales Predictions*” OR “*Predictive*”) AND (“*Machine Learning*” OR “*Algorithm*” OR “*Big Data Analytics”*) AND *Retail*”), no período de 2015 a 08/2021. Foram retornadas 1.022 publicações distribuídas conforme quadro 1.

Quadro 1 – Distribuição das publicações por base de dados

Base de Dados	Publicações	Citações
Google Scholar	998	72.226
Scopus	21	25
Microsoft Academic	3	4
Total	**1.022**	**72.255**

A figura 3 apresenta o fluxograma do processo de seleção das publicações científicas em cada uma das quatro etapas previstas pelo protocolo PRISMA-P: Identificação, triagem, elegibilidade e documentos incluídos para análise

Figura 3 - Fluxograma PRISMA-P



Fonte: Resultados da Pesquisa

Na etapa de triagem, foram removidos 87 Livros, 159 publicações classificadas como HTML e PDF, 13 Artigos de Conferência e 7 publicações classificadas como “Outros”, totalizando 266 registros excluídos, resultando 756 registros selecionados para a próxima etapa. Os critérios de inclusão e exclusão dos estudos são apresentados no quadro 2.

Quadro 2 – Critérios de inclusão e exclusão dos estudos

Critérios de inclusão
Ajuda a definir o que é predição de vendas e o impacto das variáveis exógenas e endógenas
Ajuda a categorizar os tipos de algoritmos de ML
Apresenta métricas de precisão para avaliar os algoritmos de ML
Critérios de exclusão
Artigo duplicado
Artigo completo não disponível gratuitamente
Outras literaturas, ou seja, não são artigos científicos (teses, livros, dissertações, entrevistas etc.)
Pesquisa fora do escopo de interesse

Fonte: Resultados da Pesquisa

Na etapa de elegibilidade, foram excluídos 193 artigos sem H-Index, 43 artigos cujo acesso não é público, 488 artigos cujo título e palavras chaves não tem relação com objetivo desta pesquisa.

Após análise exploratório dos 32 artigos restantes, foram descartados 23 artigos com baixa aderência ao tema da pesquisa, pois os conteúdos dos artigos não apresentaram métricas utilizadas para avaliar o desempenho dos algoritmos apresentados.

Desta forma, foram selecionadas 9 publicações para análise qualitativa conforme quadro 3.

Quadro 3 - Artigos selecionados para esta pesquisa

	Título	Autor	Ano
01	Machine learning methods for demand estimation	P Bajari, D Nekipelov, SP Ryan, M Yang	2015
02	A machine learning framework for customer purchase prediction in the non-contractual setting	A Martínez, C Schmuck, S Pereverzyev Jr	2020
03	Retail forecasting: Research and practice	R Fildes, S Ma, S Kolassa	2019
04	A deep learning approach for the prediction of retail store sales	Y Kaneko, K Yada	2017
05	Sales forecasting by combining clustering and machine-learning techniques for computer retailing	IF Chen, CJ Lu	2017
06	Intelligent Sales Prediction Using Machine Learning Techniques	S Cheriyan, S Ibrahim, S Mohanan, S Treesa	2018
07	Applying computational intelligence methods for predicting the sales of newly published books in a real editorial business management environment	Castillo, P.A., et al	2017
08	Sales-forecasting of retail stores using machine learning techniques	A Krishna, V Akhilesh, A Aich	2018
09	Comparison of different machine learning algorithms for multiple regression on black friday sales data	CSM Wu, P Patil, S Gunaseelan	2018

Fonte: Resultados da Pesquisa

3. Revisão integrativa da literatura (RIL)

Os métodos tradicionais de previsão são baseados em séries temporais, isso significa que eles são aplicados sob a hipótese de que a demanda passada pode estimar estatisticamente a demanda futura. Normalmente, esses métodos são fáceis de aplicar e apresentam bom desempenho em mercados cuja demanda é majoritariamente estável (CHOPRA *et al*, 2013). Infelizmente, temos muitos casos em que este cenário não pode ser aplicado, pois muitas vezes a demanda depende de fatores exógenos que não são efetivamente representados por valores passados. Por exemplo, serviços de transporte sob demanda como UBER ou 99 Taxi, não podem estimar sua demanda apenas contando com séries temporais, eles devem levar em conta outros elementos como condições climáticas, período do dia, dia da semana (KE *et al*, 2017).

Para satisfazer essa necessidade, outros tipos de previsão, conhecida como modelagem causal, propõem métodos que incluem elementos exógenos, como variáveis macroeconômicos, condições climáticas, estratégias de

marketing etc. (CHOPRA *et al*, 2013). Essas técnicas permitem enfrentar os limites encontrados em modelos de séries temporais. Nesse sentido, o próprio ML poderia ser considerado como um fornecedor de modelagem causal porque pode lidar com séries temporâneas, variáveis categóricas, variáveis difusas, análise de texto, imagens e outros elementos.

O ML é aplicado à previsão de vendas desde 1980 através de métodos como *Artificial Neural Networks* (ANN) (HIPPERT *et al*, 2001). Durante as últimas duas décadas, esses métodos apresentaram resultados interessantes e demonstraram algum potencial, mas vários aplicativos de pesquisa ANN não tinham validade devido a problemas de validação ou implementação (ADYA, *et al*, 1998). Esse problema, provavelmente foi causado pela escassez de dados para efetivamente treinar ANNs, pois essa técnica tem uma boa capacidade de generalização, mas precisa de muitos dados, dados bem distribuídos e tempo para treinamento. Limitações anteriores, como capacidade limitada de armazenamento, baixo poder computacional e conexões lentas com a internet, poderiam ter influenciado a relutância em usar ML em previsão de vendas.

Hoje em dia, a ML se beneficia de uma boa reputação, provavelmente porque a maioria das restrições acima mencionadas foram superadas. Graças a isso, novas aplicações foram publicadas levando a novas tendências e técnicas. Para identificar essas novas tendências e ir de encontro ao objetivo desta pesquisa, um estudo da literatura foi realizado, conforme etapas descritas na seção 4.

Como resultado foram selecionadas 9 publicações apresentadas no quadro 3. Para cada estudo, são apresentadas o objetivo da aplicação, o tipo de recurso encontrado no conjunto de dados, o método utilizado para o pré-processamento de dados e as técnicas de ML utilizadas para previsão de vendas. O quadro 4 apresenta a nomenclatura dos algoritmos utilizados nos artigos analisados e o quadro 5 o resultado da RIL.

Quadro 4 – Nomenclatura

Variáveis	
T	Variáveis de série temporal (vendas passadas)
Ve	Variáveis endógenas do modelo (preço, número do PDV, período da venda)
Vx	Variáveis exógenas do modelo (Clima, horário do dia, dia da semana (final de semana), localização espacial, sazonalidade, taxa de desemprego, taxa de inflação, PIB, tamanho populacional, renda média)
***Machine Learning* (ML)**	
RF	Random forests (based in regression trees)
SVM	Support Vector Machine
DT	Decision Trees
SR	Stepwise Regression
FSR	Forward Stagewise Regression
RR	Ridge Regression
LR	Linear Regression
LGR	Logistic Regression
Bagging	Bagging (based in regression trees)
GTB	Gradient Tree Boosting
ELM	Extreme Learning Machine
GLM	Generalized linear model
KNN	K-nearest-neighbor
LASSO	Least Absolute Shrinkage and Selection Operator / Logistic Lasso Regression
AdaBoost	Adaptive Boosting
XGBoost	Gradient Boosting
***Deep Learning* (DL)**	
ANN	Artificial Neural Network
MLK	MLK Classifier
Métricas de Precisão	
RMSE	Root Mean Squared Prediction Error
MAPE	Mean Absolute Percentage Error
MdAPE	Median Absolute Percentage Error
RMSPE	Root Mean Square Percentage Error
MAE	Mean Absolute Error
RAE	Relative Absolute Error
RRSE	Root Relative Squared Error

Métodos estatísticos e/ou de classificação	
ARIMA	Autoregressive Integrated Moving Average
ETS	Error-trend-seasonal state space formulations of exponential
HW	Holt-Winters procedure
SOM	Self-Organizing Maps
GHSOM	Growing Hierarchical Self-Organizing Map

Fonte: Resultados da Pesquisa

Quadro 5 - Aplicações de previsão de vendas com ML

Artigo	Aplicação	*Data Set*	Técnica de pré-processamento	Técnica de ML
01	Rede de supermercados, base de dados com 6 anos – 3.149 SKU e 1.510.563 transações de vendas	T, Ve	-	SR, FSR, LASSO, SVM, Bag, RF
02	Base com 10 mil clientes e 200 mil compras	T, Ve	-	GTB, LASSO, ELM
03	482 itens têxteis com 52 semanas de histórico de vendas 2 marcas de massas de 2 diferentes lojas do varejo, com dados históricos de 3 anos com vendas diárias	T, Ve, Vx	-	DT, SVM, ANN
04	Dados de três anos a partir de um ponto-de-venda (PDV) coletados entre 2002 e 2004 de um supermercado situado na região de Kanto no Japão	T, Ve	-	DL
05	Dados reais de vendas a partir de 124 pontos de venda entre Jan-2005 a Set-2009 de três grandes varejistas com domínio em produtos de informática na região de Taiwan	T, Ve	-	SVR, ELM
06	Dados reais de venda coletados de uma loja de moda por três anos consecutivos (2015 a 2017)	T, Ve	Detecção de *outlier*	GTB, DT, GLM
07	Previsão de vendas para novos livros	T, Ve, Vx	Correlation-based feature selection, Relief for atribute estimation	ELM, KNN, DT, ANN, RF, SVM
08	Conjunto de dados do Kaggle, com 8.523 entradas	T, Ve	Detecção de *outlier*, incluído valores faltantes	AdaBoost, LASSO, GTB
09	Conjunto de dados com 550 mil transações em período de *Back Friday*	T, Ve	Transformados os dados categóricos em dados numéricos	LR, MLK, DL, DT, Bagging, XGBoost

Fonte: Resultados da Pesquisa

Seguem as avaliações dos resultados apresentados no quadro 5.

No artigo [01], foram analisadas 1.510.563 transações de vendas em uma rede de supermercados com 3.149 unidades de manutenção de estoque ou SKUs (*Stock Keeping Units*), os dados foram coletados através do IRI Marketing Research através de uma licença acadêmica da universidade de Chicago. 25% dos dados foram utilizados como amostra de validação, 15% para validação e 60% para treinamento. Foi utilizada a métrica de erro de predição do quadrado médio da raiz RMSE para avaliar a precisão dos métodos utilizados, como resultado os algoritmos RF e SVM apresentaram uma precisão de 65% e 15% respectivamente, enquanto o método tradicional *Linear Regression* apresentou 6% de precisão.

[02] destaca que um dos principais desafios do varejo é diferenciar os clientes que realizaram uma compra pontual e não tem a intensão de realizar novas compras, dos clientes que possuem maior probabilidade de compras habituais, porém estão “em pausa” entre uma compra e outra. É aceito pela sabedoria empresarial e pela literatura de pesquisa que custa de cinco a dez vezes mais para adquirir um novo cliente do que reter um cliente existente (Daly, 2002; Bhattacharya, 1998). Neste estudo foi utilizado *data set* com mais de 10.000 clientes e 200.000 transações de compras, sendo que o método GTB obteve o melhor desempenho, alcançando 89% de precisão e 0,95 (AUC - *Area Under The Receiver Operating Characteristic Curve*) na previsão de compras mensais no conjunto de dados de teste.

Em [03], os autores realizam um *benchmark* através de uma revisão da literatura comparando o resultado de outros pesquisadores em vários cenários no segmento de varejo, confirmando que os modelos de série temporal têm sido muito usados para previsão de vendas agregadas no varejo, onde o *Simple Exponential Smoothing* e suas extensões, em conjunto com modelos ARIMA, têm sido os modelos de séries temporais mais empregados para previsões de vendas, porém devido a dependência de dados limitados ou uso de métricas de avaliação inadequadas, como o ajuste na amostra, alguns pesquisadores descobriram que os modelos de séries temporais padrão são às vezes inadequados para avaliar as vendas agregadas no varejo, identificando evidências de não linearidade e volatilidade na série de tempo de vendas no varejo, por exemplo, (ALON *et al*, 2001; CHU & ZHANG, 2003; KUVULMAZ, USANMAZ, & ENGIN, 2005; ZHANG & QI, 2005), eles recorreram a modelos não lineares, especialmente redes neurais artificiais. Os resultados indicam que modelos tradicionais de séries temporais com tendência estocástica, como *Simple Exponential Smoothing* e ARIMA, tiveram bom desempenho quando as condições macroeconômicas são relativamente estáveis, no entanto, quando as condições econômicas são voláteis (com rápidas mudanças nas condições econômicas), as rede neurais artificiais ANNs tem sido reivindicadas para superar os métodos lineares (ALON *et al*, 2001). Como resultado, técnicas de *Clustering* e algoritmos DT apresentaram os melhores resultados para previsão de vendas para itens novos com dados limitados de histórico. Em outro cenário foram avaliadas 2 marcas de massa em duas diferentes lojas do varejo, com

histórico de três anos de vendas diárias em produtos de promoção, os métodos SVM e ANNs apresentaram melhore resultados.

No estudo [04], foram utilizados dados de três anos a partir de um ponto-de-venda (PDV) coletados entre 2002 e 2004 de um supermercado situado na região de Kanto no Japão, com objetivo de prever o volume de vendas do próximo dia, aplicando métodos de *Deep Learning* (DP) em comparação com modelo de regressão logística LGR. Os dados das vendas foram agrupados em três categorias de acordo com os atributos dos produtos: Categoria 1 (62 atributos), Categoria 2 (569 atributos) e Categoria 3 (3.312 atributos). Os três anos de dados foram divididos para que 80% fossem utilizados para aprendizagem e 20% para verificação. Como resultado o DL foi superior ao LGR obtendo uma precisão de 86%.

No [05] foram agrupados dados reais de vendas a partir de 124 pontos de venda entre Jan-2005 a Set-2009 de três grandes varejistas com domínio em produtos de informática na região de Taiwan. Para o histórico de vendas foram considerados três produtos: Computadores (PC), Notebooks (NB) e Monitores de cristal líquido (LCD). No estudo foram comparadas três técnicas de agrupamento de dados *Clustering*: SOM, GHSOM e K-Means, juntamente com dois algoritmos de ML: SVM e ELM. Os conjuntos de dados PCs, NBs e LCDs, foram submetidos a seis modelos de previsão baseados em cluster e também de forma isolada em duas técnicas de aprendizado de máquina (*single* SVR) e (*single* ELM), ou seja, sem usar algoritmo de agrupamento. Os primeiros 88 pontos de venda (71% da amostra) são usados como amostra de treinamento, enquanto o restante 36 pontos de vendas (29% da amostra) são usados para realizar a previsão das vendas. Foram utilizados 2 critérios de avaliação para medir o desempenho de vendas: MAPE e RMSPE.

O resultado experimental, demonstrou que dos 8 modelos criados a combinação do agrupamento GHSOM com o algoritmo de aprendizado de máquina ELM forneceu desempenho superior para todos os 3 produtos selecionados.

Em [06], é destacado que a precisão na previsão de vendas proporciona um grande impacto nos negócios. As técnicas de mineração de dados são ferramentas muito eficazes na extração de conhecimento oculto de um enorme conjunto de dados para aumentar a precisão e a eficiência da previsão. As organizações enfrentam sérios desafios para identificar uma técnica de mineração de dados e uma estratégia eficaz de pré-venda (MATHEW *et al*, 2015), devido ao crescimento exponencial do volume de dados usados em transações de comércio eletrônico. Os métodos tradicionais de previsão são difíceis de lidar com uma grande quantidade de dados *Big Data,* neste contexto a técnica de mineração de dados *Data Mining* se torna um forte aliado na predição de vendas. No nível organizacional, as previsões de vendas são insumos essenciais para apoio na tomada de decisões em diversas áreas de negócio, como operações, marketing, vendas, produção, logística, estoque, finanças (fluxo de caixa). Com quase 85.000 registros a base de dados inicial considerada nesta pesquisa, foi reduzida após o pré-processamento devido a registros redundantes assim como informações irrelevantes para análise. O conjunto de dados utilizado para este artigo é baseado em uma loja de moda

com três anos consecutivos (2015 a 2017) de dados de vendas. Neste artigo foram comparados três algoritmos de ML: GTB, DT e GLM, os quais apresentaram 98%, 71% e 64% de precisão respectivamente.

[07] aborda o problema de prever vendas para produtos recém lançados ao mercado. Para este estudo foram coletados dados de vendas composto por 6000 livros de uma editora espanhola Trevenque Editorial S.L, cujo os dados foram analisados por meio do classificador SOM e duas técnicas de pré-processamento: *Correlation-based feature selection* e *Relief for atribute estimation.* Trata-se de um desafio para a indústria, pois imprimir um número muito maior de volumes do que aqueles finalmente vendidos levará a perdas, enquanto a impressão de um número adequado de cópias, otimizará as vendas e os lucros da editora. Além disso, há diversas dificuldades inerentes à previsão de vendas de novos livros, como a quantidade limitada de dados históricos ou a variabilidade do mercado (modas, sazonalidade) que tornam difícil a tarefa dos métodos preditivos. Como resultados os algoritmos DT e RF apresentaram a maior precisão e o melhor desempenho.

No artigo [08], foi comparado a precisão de diferentes algoritmos no conjunto de dados extraído do site *https://www.kaggle.com/*, onde o AdaBoost e GTB apresentaram RMSE de 1.350 e o 1.088 respectivamente, desta forma o GTB apresentou menor índice de erro, portanto melhor precisão no conjunto de dados avaliado.

No artigo [09], um conjunto de dados com 550 mil transações de vendas são coletados de uma empresa do varejo para treinar algoritmo de ML supervisionado para prever a quantidade de compra de clientes, possibilitando a criação de ofertas personalizadas conforme o perfil de consumo destes clientes no período da *Black Friday*. Na pesquisa é destacado que em algoritmos de ML, o conjunto de dados utilizado deve estar equilibrado, ou seja, todas as classes devem conter o mesmo número de amostras, caso contrário, a previsão ou classificação será tendenciosa para aquela categoria de dados em que os dados estão distorcidos, ou seja, um bom algoritmo de ML é inútil sem os dados adequados. A precisão do modelo de previsão só aumenta se os dados em que ele é construído forem sólidos, contudo, os dados do mundo real são confusos e precisam sem limpos, para ajudar nesta tratativa temos as técnicas de pré-processamento. Como conclusão é destacado que modelos complexos como redes neurais são um exagero para tratar problemas simples como a regressão, desta forma podemos utilizar modelos mais simples, juntamente com o pré-processamento para obter melhores resultados. Nos métodos de ML aplicados, o XGBoost que utiliza internamente SR e RR apresentou a melhor precisão com 2400 RMSE.

Fica evidenciado, que o ML pode ser aplicado na previsão de vendas em uma ampla gama de diferentes tipos de produtos.

[01,03,07] destacam que métodos como RF e DT fornecem um nível inigualável de interpretação, juntamente com boa precisão e tempo de computação decrescente.

A maioria dos estudos incluem entradas endógenas e os [03,07] exógenas em seus modelos, o que mostra a boa flexibilidade do ML para lidar com uma ampla gama de entradas. Além disso, [07] abordou o problema de implementar previsão de vendas em novos produtos para os quais não há dados históricos de vendas disponíveis. Para isso, foram utilizados dados históricos de outros produtos, juntamente com variáveis endógenas, como o número de semanas à venda e o preço de varejo.

Em [06,07,08,09] foram efetivamente aplicadas técnicas de pré-processamento de dados, que resultaram em modelos mais simples, menor custo computacional e boa precisão. Como estamos em uma era com alto volume de dados sendo gerado diariamente, isso provoca variáveis ruidosas e sem sentido, sendo assim, analisar as variáveis mais relevantes em aplicações do mundo real é obrigatório, desta forma conseguimos diferenciar se um modelo será ou não útil.

3.1 Comparando *Machine Learning* e os métodos tradicionais de previsão

Os modelos tradicionais oferecem enormes vantagens em termos de simplicidade e precisão, pois podem realizar a previsão de vendas em questão de segundos para vários SKUs (YU *et al*, 2011). No entanto, eles precisam ser projetados por um especialista que possa adequá-lo às necessidades da empresa. Além disso, não incluem variáveis exógenas. Os modelos de ML resolvem parcialmente esse problema porque podem incluir outros tipos de dados, como variáveis endógenas e exógenas, permitindo uma melhor representação da realidade. Além disso, as técnicas de ML que são implementadas corretamente superam a maioria dos métodos tradicionais de previsão (YU *et al*, 2011). Para avaliar se essa afirmação é válida em nosso estudo da literatura, retratado no quadro 5, o quadro 6 menciona quais foram os modelos tradicionais e de ML aplicados pelos autores e quais apresentaram melhor precisão.

Quadro 6 - Comparação do ML com métodos tradicionais de previsão de vendas

Artigo	Aplicação	Métodos estatísticos e/ou de classificação	Métrica de avaliação	Modelo ML	Modelo com maior precisão
01	Mercearia, base de dados com 6 anos – 3.149 SKU e 1.510.563 transações de vendas	Linear Regression	RMSE	SR, FSR, LASSO, SVM, Bag, RF	RF, SVM
02	10.000 clientes e 200.000 transações de compras	-	Precision, AUC	GTB, LASSO, ELM	GTB
03	482 itens têxteis com 52 semanas de histórico de vendas. 2 marcas de massas de 2 diferentes lojas do varejo, com dados históricos de 3 anos com vendas diárias	Simple Exponential Smoothing, ARIMA ARIMA, ETS e HW	RMSE, MAPE, MdAPE	DT SVM, ANN	DT, SVM, ANN
04	Dados de três anos a partir de um ponto-de-venda (PDV) coletados entre 2002 e 2004 de um supermercado situado na região de Kanto no Japão	Logistic Regression	Accuracy, Precision, Recall, F-measure, AUC	-	DL

Artigo	Aplicação	Métodos estatísticos e/ou de classificação	Métrica de avaliação	Modelo ML	Modelo com maior precisão
05	Dados reais de vendas a partir de 124 pontos de venda entre Jan-2005 a Set-2009 de três grandes varejistas com domínio em produtos de informática na região de Taiwan	SOM, GHSOM, K-Means	MAPE, RMSPE	SVR, ELM	GHSOM + ELM
06	Dados reais de venda coletados de uma loja de moda por três anos consecutivos (2015 a 2017)	-	Accuracy, Error Rate, Precision, Recall, Kappa	GLM, DT, GTB	GTB
07	Previsão de vendas para novos livros	Multiple Linear Regression	MAE, RMSE, RAE, RRSE	ELM, KNN, DT, ANN, RF, SVM	DT, RF
08	Conjunto de dados do Kaggle, com 8.523 entradas	Multiple Regression, Polynomial Regression, Ridge Regression	RMSE	AdaBoost, LASSO, GTB	GTB
09	Conjunto de dados com 550 mil transações em período de *Back Friday*	Logistic Regression, Logistic Regression, Polynomial Regression, Stepwise Regression, Ridge Regression, Lasso Regression, Elastic Regression	RMSE	LR, MLK, DL, DT, Bagging, XGBoost	XGBoost

Fonte: Resultados da Pesquisa

O fato de ML superar os modelos tradicionais no cenário de previsão de vendas não significa necessariamente que as empresas devem mudar suas ferramentas de previsão. Isso nos leva a considerar uma questão importante: quando as empresas devem investir em ML para previsão de vendas?

A previsão de vendas deve ser flexível e reativa, especialmente quando se trata de previsão de curto prazo, o que é necessário em setores como o varejo. O cálculo da previsão deve ser ágil, pois a maioria das empresas tem centenas ou até milhares de SKUs, e todas elas podem precisar de uma estimativa imediata. Se os métodos de ML não forem aplicados corretamente, podem levar horas ou até dias para treinar os modelos (HUANG *et al*, 2006). Além disso, após a fase de treinamento, esses métodos ainda são demorados, tornando-os menos adequados para previsão de vendas.

Mesmo que as técnicas de pré-processamento de dados possam reduzir substancialmente o tempo de computação, elas ainda são complexas e, portanto, custosas de estabelecer em uma empresa, pois exigem pessoas qualificadas com equipamentos adequados. Essa dificuldade está especialmente presente em pequenas empresas, onde os recursos disponíveis são limitados e os funcionários raramente têm conhecimento avançado sobre o tema. Nesse caso, se uma empresa está posicionada em um mercado estável e se a demanda histórica é suficiente para alcançar uma boa precisão na previsão de vendas com métodos tradicionais, esta empresa deve postergar a migração para técnicas de ML até que realmente identifique um valor agregado em usá-los.

Por outro lado, se uma empresa vende produtos em um mercado sujeito à constante evolução, onde é obrigatório estar na vanguarda das tendências por ser competitiva, os modelos de ML para previsão de vendas serão um ativo valioso. No entanto, não é uma mudança fácil, pois as empresas devem garantir três aspectos fundamentais: capacidade de armazenamento de dados, capacidade de processamento de dados e qualificação dos funcionários. O armazenamento de dados é fundamental, pois o ML precisa lidar com *Big Data* para um bom desempenho, e este exige grande capacidade de armazenamento, como também a capacidade de processamento de dados pode variar de modelos simples até modelos mais robustos (ZHOU *et al*, 2017); sendo assim a empresa deve considerar seus objetivos versus os recursos necessários.

Em suma, uma empresa deve optar em utilizar ML em vez de métodos tradicionais de previsão de vendas quando seu ambiente econômico realmente requer uma transformação digital, e quando a empresa consegue reunir os recursos necessários para assumir o desafio de projetar as vendas futuras.

4. Conclusão

Lidar com um alto volume de dados é um dos desafios mais comuns impostos pelas tendências modernas dos negócios. Como os dados se tornaram um dos recursos mais valiosos, os gestores de vendas estão ansiosos para extrair informações relevantes que levem a uma vantagem competitiva. Este artigo realizou um estudo de literatura para investigar a aplicação de ML em previsão de vendas como método para alcançar essa vantagem. Nove artigos de pesquisa recentes que aplicam ML em previsão de vendas foram selecionados e analisados para identificar novas tendências. Um dos achados foi que o ML amplia o alcance da predição de vendas, pois é capaz de lidar com variáveis complexas. Mais precisamente, as abordagens ANN mostraram excelente desempenho ao lidar com dados imprecisos como variáveis exógenas, enquanto DT e RF oferecem uma incrível capacidade de interpretação. Além disso, as técnicas de pré-processamento de dados provaram reduzir substancialmente a complexidade dos modelos, permitindo tanto boa precisão quanto tempo razoável de computação.

Como o processo de adaptação de uma empresa às novas tecnologias muitas vezes vem com dúvidas levantadas pela incerteza, uma comparação entre ML e métodos tradicionais de previsão foi feita com o objetivo de fornecer *insights* aos gestores que estão dispostos a implementar ML em seus processos. Os resultados deste estudo mostram que o ML é mais adequado do que os métodos tradicionais de previsão em termos de precisão, particularmente quando os modelos contêm variáveis exógenas e endógenas. Além disso, permite a identificação de padrões ocultos na demanda que podem ser usados como linha de base para identificar novas tendências de mercado.

Nos achados, fica evidenciado que a ausência de uma base de treinamento confiável, ou seja, que reflita exatamente as transações do mundo real, afeta a eficiência do algoritmo, contendo vieses que inviabilizam sua utilidade. Além disso, é constatado que em mercados com demandas constantes e poucas

interferências externas, não se justifica a utilização de ML, pois para estes casos a utilização de métodos tradicionais por meio séries temporais é mais simples e menos custoso.

Para pesquisas posteriores, um estudo semelhante deve ser realizado sobre técnicas de pré-processamento de dados, pois oferecem vantagens significativas em termos de redução de custo computacional na aplicação de modelos ML. Finalmente, abordar essas novas tendências para as pequenas empresas é vital, pois elas são numerosas, mas muitas vezes não têm meios financeiros ou experiência para implementar tecnologias disruptivas.

Referências

ADYA, M. & COLLOPY, F. (1998) - **How effective are neural networks at forecasting and prediction? A review and evaluation**. - Journal of Forecasting

ALON, I., QI, M., & SADOWSKI, R. J., (2001) - **Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods** - Journal of Retailing and Consumer Services

ATZORI, L., IERA, A., & MORABITO, G. (2010) - **The internet of things: A survey. Computer Networks**

BHATTACHARYA, C.B., (1998) - **When customers are members: Customer retention in paid membership contexts** - Journal of the academy of marketing science.

CECCHINEL, C., JIMENEZ, M., RIVEILL, M., & MOSSER, S. (2014) - **An architecture to support the collection of big data in the internet of things**. IEEE World Congress on Services

CHOPRA, S., & MEINDL, P. (2013) - ***Supply Chain Management*** - Pearson Education, Inc.

CHU, C. W., & ZHANG, G. P., (2003) - **A comparative study of linear and nonlinear models for aggregate retail sales forecasting** - International Journal of Production Economics

DALY, J.L., (2002) - **Pricing for profitability: activity-based pricing for competitive advantage. volume 11** - John Wiley & Sons.

FAN, X., ZHANG, S., WANG, L., YANG, Y., AND HAPESHI, K. (2013) - **An evaluation model of supply chain performances using 5DBSC and LMBP neural network algorithm** - Journal of Bionic Engineering

HIPPERT, H.S., PEDREIRA, C.E., AND SOUZA, R.C. (2001) - **Neural networks for short-term load forecasting: a review and evaluation** - IEEE Transactions on Power Systems

HOFMANN, E., RUTSCHMANN, E. (2018) - **Big data analytics and demand forecasting in supply chains: a conceptual analysis** - International Journal of Logistics Management
https://doi.org/10.1108/IJLM-04-2017-0088

HUANG, G.B., ZHU, Q.Y., AND SIEW, C.K. (2006) - **Extreme learning machine: Theory and applications** - Neurocomputing, Neural Networks

Juniper Research (2020). Disponível em: https://www.juniperresearch.com/press/iot-connections-to-reach-83-bn-by-2024?ch=IOT%20CONNECTIONS%20TO%20GROW

KE, J., ZHENG, H., YANG, H., AND CHEN, V. MICHAEL (2017) - **Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach** - Transportation Research

KUVULMAZ, J., USANMAZ, S., AND ENGIN, S. N., (2005) - **Time-series forecasting by means of linear and nonlinear models** - In A. Gelbukh, A. DeAlbornoz, & H. TerashimaMarin (Eds.), MICAI 2005: Advances in artificial intelligence

LOYER, J.L., HENRIQUES, E., FONTUL, M., AND WISEALL, S. (2016) - **Comparison of machine learning methods applied to the estimation of manufacturing cost of jet engine components**. -International Journal of Production Economics

MATHEW NGWAE MAINGI AND JOMO KENYATTA (2015) - **A Survey on the Clustering Algorithms in Sales Data Mining** - International Journal of Computer Applications Technology and Research

MAYO, M. (2016) - **The data science puzzle, explained**. Disponível em: https://www.kdnuggets.com/2016/03/data-science-puzzle-explained.html/2

MITCHELL, T.M. (1997) - **Machine Learning** - McGraw Hill

MOHER, D., ET AL. (2015) - **Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P)**

RASCHKA, S., AND MIRJALILI, V. (2017) - **Python Machine Learning, 2nd Ed**.- Packt Publishing, Birmingham

VAHDANI, B., RAZAVI, F., AND MOUSAVI, S.M. (2016) - **A high performing meta-heuristic for training support vector regression in performance forecasting of supply chain** - Neural Computing and Applications

YU, Y., CHOI, T.M., AND HUI, C.L. (2011) - **An intelligent fast sales forecasting model for fashion products** - Expert Systems with Applications

ZHANG, G. P., AND QI, M. (2005) - **Neural network forecasting for seasonal and trend time series** - European Journal of Operational Research

ZHOU, L., PAN, S., WANG, J., AND VASILAKOS, A.V. (2017) - **Machine learning on big data: Opportunities and challenges** - Neurocomputing