



Ministério da Educação  
Universidade Federal de Ouro Preto  
Escola de Minas  
Departamento de Engenharia de Produção



# **Análise Comparativa de Métodos Computacionais para a Classificação de Estoque: Redes Neurais, kNN e SVM**

Lauri Antonio Nissila

Ouro Preto - MG

2023

Lauri Antonio Nissila

**Análise Comparativa de Métodos Computacionais para a Classificação  
de Estoque: Redes Neurais, kNN e SVM**

Monografia apresentada ao Curso de Engenharia de Produção da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau em Bacharel em Engenharia de Produção.

Ouro Preto - MG

15 de março de 2023



## FOLHA DE APROVAÇÃO

Lauri Antonio Nissila

### Análise comparativa de métodos computacionais para a classificação de estoque: redes neurais, kNN e SVM

Monografia apresentada ao Curso de Engenharia de Produção da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Engenheiro de Produção

Aprovada em 15 de Março de 2023

#### Membros da banca

Mestre - Davi Neves Pavanelli - Orientador(a) - Universidade Federal de Ouro Preto  
Doutor - Jorge Luiz Brescia Murta - Universidade Federal de Ouro Preto  
Mestre - Cristiano Luis Turbino de Franca e Silva - Universidade Federal de Ouro Preto

Davi Neves Pavanelli, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 15/03/2023



Documento assinado eletronicamente por **Davi Neves Pavanelli**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 16/03/2023, às 09:19, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0491706** e o código CRC **28C2AFBD**.

## **Agradecimentos**

*Em primeiro lugar, a Deus, que fez com que meus objetivos fossem alcançados, durante todos os meus anos de estudos. Aos meus pais e irmã, que me incentivaram nos momentos difíceis. Ao professor Davi, por ter sido meu orientador e ter desempenhado tal função com dedicação e amizade.*

*“Se você não pode medir,  
você não pode gerenciar”*

## RESUMO

O principal objetivo foi realizar uma análise comparativa de métodos computacionais para a classificação de estoque com os métodos de Redes Neurais, kNN e SVM comparando suas performances entre si e com uma classificação de referência. Para alcançar este objetivo, foi realizada uma série de etapas: 1. Pré-processamento e limpeza dos dados 2. Análise de características e seleção de atributos 3. Treinamento, validação e avaliação dos modelos selecionados 4. Comparação dos resultados e conclusão. Os modelos selecionados foram kNN, Redes Neurais e SVM. Os dados foram pré-processados para remover NaNs e outliers e os atributos relevantes foram selecionados para cada modelo. Os modelos foram treinados usando k-fold cross-validation e avaliados usando métricas como a acurácia, precisão, revocação, F1-Score e AUC. Os resultados foram comparados com uma classificação de referência. Os resultados mostraram que o modelo kNN obteve os melhores resultados com uma acurácia de 91,2%, seguido pelo SVM com 90,7% e Redes Neurais com 89,9%. Todos os modelos superaram a classificação de referência em pelo menos 8% em todas as métricas. Em conclusão, o modelo kNN foi o melhor modelo para classificar o estoque, seguido pelo SVM e Redes Neurais.

**Palavras-Chave:** Gestão de estoque; Aprendizado de máquina; Python; Softwares computacionais

## Abstract

The main objective was to perform a comparative analysis of computational methods for the classification of stock with the methods of Neural Networks, kNN and SVM comparing their performances with each other and with a reference classification. To achieve this goal, a series of steps were performed: 1. Pre-processing and data cleaning 2. Analysis of characteristics and selection of attributes 3. Training, validation and evaluation of selected models 4. Comparison of results and conclusion. The selected models were kNN, Neural Networks and SVM. The data were preprocessed to remove NaNs and outliers and the relevant attributes were selected for each model. The models were trained using k-fold cross-validation and evaluated using metrics such as accuracy, accuracy, recall, F1-Score and AUC. The results were compared with a reference classification. The results showed that the kNN model had the best results with an accuracy of 91.2%, followed by SVM with 90.7% and Neural Networks with 89.9%. All models exceeded the reference rating by at least 8% on all metrics. In conclusion, the kNN model was the best model for classifying inventory, followed by SVM and Neural Networks.

**Keywords:** Inventory management; Machine learning; Python; Computer software

## Lista de Figuras

<b>Figura 1 Curva ABC para determinação de acurácia .....</b>	<b>17</b>
<b>Figura 2 Representação gráfica de uma classificação SVM para determinar a acurácia.</b>	<b>26</b>
<b>Figura 3 Representação gráfica de uma rede neural artificial para determinar a acurácia</b>	<b>17</b>
<b>Figura 4 Representação gráfica da curva de Acurácia.....</b>	<b>33</b>
<b>Figura 5 Fluxo de execução do sistema de classificação dos métodos KNN, SVM E RNA para definição de acurácia de estoque.....</b>	<b>34</b>
<b>Figura 6 Dispersão de vizinhos na classificação KNN .....</b>	<b>17</b>

## Lista de Gráficos

<b>Gráfico 1</b>	<b>Resultado da Previsão do fornecedor do item 10000068026- KNN-2019.....</b>	<b>39</b>
<b>Gráfico 2</b>	<b>Resultado de acurácia da classificação do inventário-KNN-2019 .....</b>	<b>40</b>
<b>Gráfico 3</b>	<b>Resultado da Previsão do fornecedor do item 10000068333 -KNN-2020.....</b>	<b>40</b>
<b>Gráfico 4</b>	<b>Resultado da acurácia da classificação do inventário-KNN-2020 .....</b>	<b>41</b>
<b>Gráfico 5</b>	<b>Resultado da acurácia da classificação do inventário-KNN-2020 .....</b>	<b>42</b>
<b>Gráfico 6</b>	<b>Resultado da Previsão do fornecedor do item 1000012659 -KNN-2021.....</b>	<b>42</b>
<b>Gráfico 7</b>	<b>Resultado da acurácia da classificação do inventário-KNN-2020 .....</b>	<b>43</b>
<b>Gráfico 8</b>	<b>Resultado da acurácia da classificação do inventário-SVM-2019 .....</b>	<b>47</b>
<b>Gráfico 9</b>	<b>Resultado da acurácia da classificação do inventário-SVM-2020 .....</b>	<b>47</b>
<b>Gráfico 10</b>	<b>Resultado da acurácia da classificação do inventário-SVM-2021 .....</b>	<b>48</b>
<b>Gráfico 11</b>	<b>Resultado da acurácia da classificação do inventário-SVM-2022 .....</b>	<b>48</b>
<b>Gráfico 12</b>	<b>Resultado da acurácia da classificação do inventário-RNN-2019 .....</b>	<b>51</b>
<b>Gráfico 13</b>	<b>Resultado da acurácia da classificação do inventário-RNN-2020 .....</b>	<b>51</b>
<b>Gráfico 14</b>	<b>Resultado da acurácia da classificação do inventário-RNN-2021 .....</b>	<b>52</b>
<b>Gráfico 15</b>	<b>Resultado da acurácia da classificação do inventário-RNN-2022 .....</b>	<b>52</b>

## **Lista de Algoritmos**

## **Lista de Abreviaturas e Siglas**

ADU Annual Dolar Usage

ANN Artificial Neural Networks

CD Centro de Distribuição

CSL Customer Service Level

GA Genetic Algorithms

IA Inteligência Artificial

KNN K-nearest Neighbors

ML Machine Learning

MCIC Multi-Criteria Inventory Classification

SVM Support Vector Machines

RNN Redes Neurais

# Sumário

<b>1. Introdução .....</b>	<b>15</b>
<b>2. Métodos para o controle de estoques.....</b>	<b>20</b>
<b>2.1 Planejamento e controle de estoques .....</b>	<b>20</b>
<b>2.1.1 Controle de compras .....</b>	<b>20</b>
<b>2.1.2 Controle dos estoques .....</b>	<b>21</b>
<b>2.1.3 Porque e como reduzir os níveis de estoque.....</b>	<b>23</b>
<b>2.2 Agrupamento de produtos com o método dos k-zinhos mais próximos.....</b>	<b>24</b>
<b>2.3 Agrupamento de produtos com o método SVM .....</b>	<b>25</b>
<b>2.4 Agrupamento de produtos com redes neurais.....</b>	<b>26</b>
<b>3. Implementação computacional e datasets .....</b>	<b>28</b>
<b>3.1 Base de dados.....</b>	<b>28</b>
<b>3.2 Bibliotecas de Python usadas .....</b>	<b>31</b>
<b>3.3 Definição de acurácia .....</b>	<b>33</b>
<b>4. Resultados e discussões .....</b>	<b>36</b>
<b>4.1 Classificação de estoques e agrupamento de produtos usando kNN .....</b>	<b>36</b>
<b>4.2 Classificação de estoques e agrupamento de produtos usando SVM.....</b>	<b>44</b>
<b>4.3 Classificação de estoques e agrupamento de produtos usando RNA .....</b>	<b>50</b>
<b>4.4 Análise comparativa dos métodos utilizados.....</b>	<b>53</b>
<b>5. Considerações finais.....</b>	<b>58</b>
<b>Referências .....</b>	<b>60</b>
<b>6. Apêndices.....</b>	<b>62</b>

# 1. Introdução

Ao almejar a lucratividade, as organizações priorizam o capital investido para maximização do lucro, o que é prioridade nos planejamentos estratégicos. No que tange a proporção do lucro, de alta ou de baixa, as organizações estão cada vez mais em busca de desenvolver suas áreas através da melhoria contínua das atividades prestadas. Entretanto, as estratégias utilizadas afim de maximizar a rentabilidade no que tange o capital investido, a administração de estoque torna-se um enfoque alvo para as organizações, de modo que os estoques solicitam um alto investimento. Diante disso, a ciência de gestão de estoques busca desenvolver técnicas eficazes que permitam o abastecimento das cadeias produtivas a qual utiliza o mínimo de investimento necessário.

Para Petrônio (2005), os estoques podem ser de matérias primas, componentes, ingredientes, insumos, materiais em processo, material para embalagens, material para expediente, produtos acabados, mercadorias no varejo, materiais de manutenção, ferramentas, entre outros. Para este estudo se utilizará a partir de agora os estoques de produtos acabados.

Segundo o autor, os estoques são gerados por causas, as quais sendo eliminadas eliminam também a necessidade da formação dos mesmos. Tais causas seriam entre outras: incertezas normalmente identificadas como “Murphys” em função da famosa lei de Murphy que recomenda contar sempre com uma dose de azar; riscos por perda de vendas por falta de produtos; flutuações de oferta ou demanda; erros nas previsões; especulação financeira; sazonalidades; fatores mercadológicos como beleza de uma gôndola cheia, ofertas, campanhas publicitárias; falta de conhecimento de técnicas de gestão de estoques; falta de confiabilidade dos dados; inércia administrativa; falta de flexibilidade por restrições produtivas, logísticas ou econômicas. Esses “Murphys”, que estão relacionados ao mau gerenciamento das empresas, podem também ser chamados segundo Petrônio (2005) de gargalos empresariais; e a existência de estoques acaba encobrendo tais erros administrativos ao longo da cadeia de suprimentos.

Para Arnold (2006) dois aspectos relacionados aos estoques podem ser confundidos por muitos gestores e são o valor de estoque e o custo de estoque. O valor de estoque revela quanto vale o somatório total dos valores dos produtos estocados, mas não revela o custo desse estoque para as empresas. Esse último pode ser calculado caso os gestores queiram saber qual o custo de oportunidade relacionado ao valor de estoque; ou, por outro lado, necessitem calcular quanto custa possuir tal estoque

imobilizado, ou ainda quanto ganham por possuir o mesmo e poder garantir determinado nível de serviço junto aos seus clientes. A fixação de níveis de estoques muito baixos pode causar o estrangulamento das linhas de produção ou interrupção do fornecimento de produtos e em contrapartida, o excesso de estoques gera grande empate de capital e necessidade de investimentos em depósitos bem como em sistemas de controle e inspeção.

O objetivo principal dos estoques é prover os produtos certos, no tempo certo, no lugar certo, na quantidade desejada e com o menor custo possível; e seu propósito pode variar de acordo com as estratégias de cada empresa podendo ser estoques de segurança, sazonais, estratégicos, especulativos.

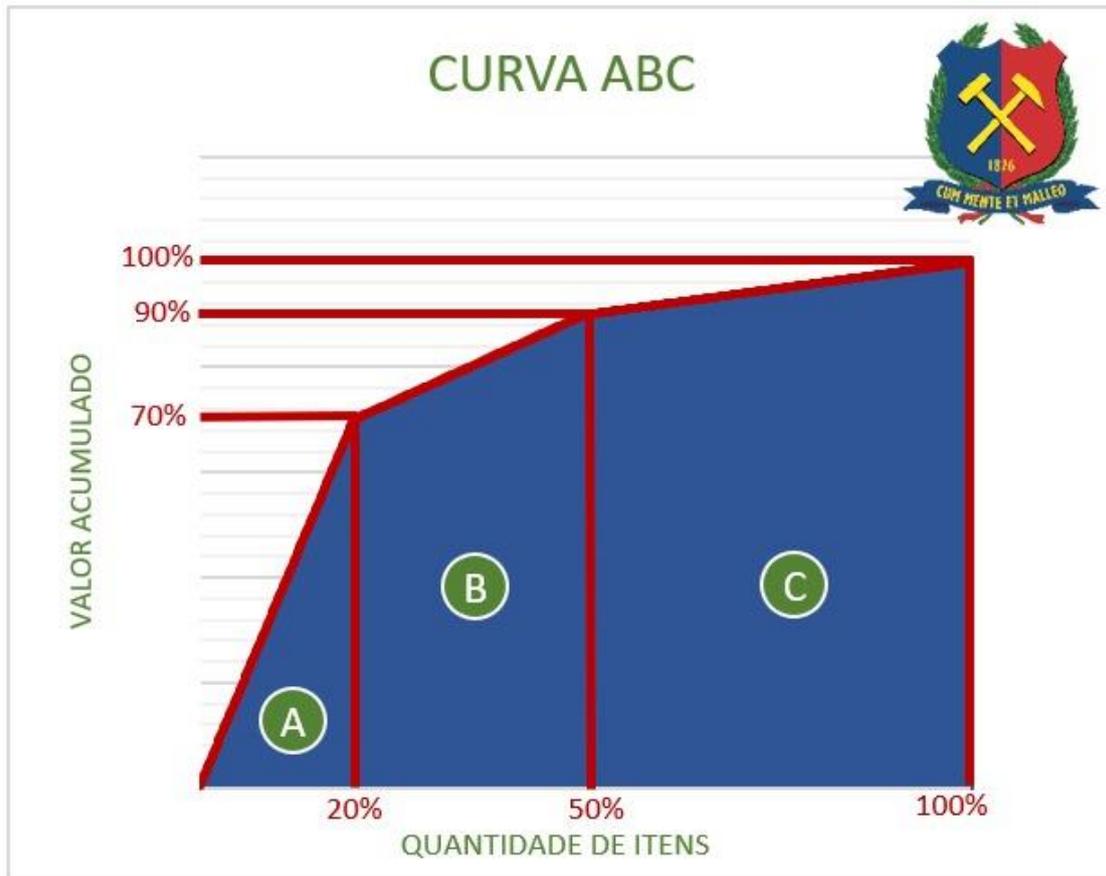
Uma das principais preocupações do administrador de materiais é saber definir e quantificar os custos que incidem sobre o estoque que ele gerencia já que, a existência de estoques implica em gastos: os custos médios anuais de manutenção de estoques em empresas de manufaturas são estimados ao redor de 20% a 40% do valor dos itens estocados (LIMA, 2013).

Existem e são necessárias muitas técnicas e ferramentas para garantir o controle dos estoques. Muitos indicadores de produtividade podem ser utilizados em suas análises, sendo que se pode citar os mais utilizados como: comparativo entre inventário físico e contábil, acurácia dos controles, curva ABC e outros.

A curva ABC de estoques surgiu em meio aos estudos de Wilfredo Frederigo Samaso, ou mais conhecido como Vilfredo Pareto (economista e sociólogo italiano), que viveu entre os anos 1848 e 1923. O autor acredita que a relação dos percentuais é de 80% e 20%, onde 20% da população representa a maior parte da renda e os 80% da população era composto pela parte que representava. Entretanto, a lei de Pareto foi apropriada para a administração dos estoques dando início ao sistema de análise ABC.

A curva ABC tem como objetivo demonstrar a relação existente entre o consumo do estoque; o investimento aplicado e a quantidade de itens que formam o estoque. O ponto principal a ser visualizado no sistema de análise ABC, é onde os itens que fazem parte do maior percentual de valor do estoque são justamente os que representam a menor parte desse estoque (LIMA, 2013). A classificação ABC, identifica os materiais segundo à proporção que eles representam no consumo e relacionam-se com o valor de aquisição e quantidade disponível em estoque. Ao classificar os materiais de acordo com o sistema ABC, é formidável a aplicação dos graus de controle adequados as diferentes classes de produtos.

**Figura 1** Curva ABC para determinação de acurácia



FONTE: AUTOR

A acuracidade provém do termo em inglês accuracy e traz em seu significado a ideia de precisão. Aplicando o conceito da acuracidade no estoque, verifica-se que quanto mais precisas forem as informações dos estoques, mais seguras serão as decisões de seu gerenciamento (WALLER et al., 2006). Quando a informação de estoque no sistema de controle, informatizado ou manual, não confere com o saldo real, dizemos que este inventário não é confiável ou não tem acuracidade.

Acuracidade de estoque é um indicador da qualidade e confiabilidade da informação existente nos sistemas de controle, contábeis ou não, em relação à existência física dos itens controlados. A falta de confiabilidade nas informações afeta todos os setores da empresa, desde o nível gerencial até o operacional. Uma informação errada dos saldos em estoque pode levar a uma decisão equivocada na área de planejamento de estoques ou compras, atrasar a produção ou até mesmo ocasionar a falta do produto para o cliente (MARTINS. 2010).

Os estoques representam parcela substancial dos ativos das empresas, devem ser encarados como um fator potencial de gestão de negócios e de lucro, portanto deve-se verificar se estão tendo utilidade adequada ou sendo um peso morto, não apresentando retorno sobre o capital neles investidos. Mesmo não sendo uma função executiva, tal função constitui uma das atividades mais importantes de uma empresa de manufatura e varejo (MARTINS, 2010).

De acordo com Pozo (2007), muitas empresas chegam à falência por imobilizar elevadas somas de capital em estoques, faltando-lhes recursos financeiros para capital de giro. Segundo o autor, uma atividade importante dentro do conjunto da gestão de estoque é prever o valor do estoque em intervalos de tempo adequado e gerenciá-lo, comparando-o com o planejado, e tomar as devidas ações quando houver desvios de rotas.

A falta de um acompanhamento efetivo da acuracidade de estoque pode gerar efeitos indesejáveis no funcionamento das atividades citadas acima. Com isso, a manutenção de registros de estoque precisos - registros que refletem a realidade física - é crucial para o desempenho de organizações de varejo, tendo em vista a integração de cadeia de suprimentos (DEHORATIUS e RAMAN, 2004).

O controle efetivo dos estoques, por consequência um maior índice de acuracidade, é, para muitas empresas, um objetivo a ser alcançado. A falta de acuracidade de estoque gera uma série de efeitos indesejáveis para os processos internos e externos das organizações. Outros efeitos também podem ser verificados com a falta de acuracidade de estoque, como uma interferência no tamanho do lote e na certeza do atendimento da demanda (UCKUN et al., 2008). Como por exemplo, gerar a necessidade de interromper a produção antes do previsto, impossibilitando, assim, que a quantidade programada na ordem de fabricação seja cumprida.

Com o crescimento da quantidade e variedade de itens armazenados, o método manual para gerenciamento de estoques passa a ser inviável. Assim, torna-se necessário adotar sistemas de informação para auxiliar na gestão de estoques, pois eles permitem a automatização de processos, como por exemplo, o registro de entradas e saídas, a medição dos níveis de estoques, a emissão de alertas, a previsão de pedidos e a otimização dos processos logísticos. Além disso, é possível acompanhar os estoques em tempo real, criar relatórios e monitorar custos, bem como ter acesso a informações gerenciais e de mercado. Utilizar um sistema de informação para gerenciar estoques é uma das melhores formas de se manter o controle sobre o estoque e garantir que todas as transações sejam registradas corretamente.

A tecnologia da informação tem se tornado cada vez mais importante para o sucesso de uma empresa. Com a ascensão da tecnologia da informação, a administração da produção das empresas é afetada de forma significativa. Os avanços na tecnologia permitem que as empresas melhorem sua

capacidade de gerenciar os recursos de forma mais eficiente e eficaz. Uma das principais mudanças na administração da produção afetada pela tecnologia da informação é a possibilidade de realizar processos de produção automatizados. Esta automação permite que as empresas reduzam seus custos de manufatura, aumentando assim sua competitividade no mercado. Além disso, a automação também permite que os processos sejam mais ágeis e precisos.

Foi neste limiar que nos anos 60 surgiu o método KNN (K-Nearest Neighbors), um algoritmo de aprendizado de máquina supervisionado que foi desenvolvido no início dos anos 1960. É baseado na "teoria de vizinhança", que afirma que os pontos de dados similares são "vizinhos" uns dos outros. O KNN é um método simples de classificação, que é usado para atribuir uma etiqueta a um novo ponto de dados com base em seus vizinhos mais próximos. A etiqueta atribuída é a etiqueta que é mais comum entre os vizinhos mais próximos. O KNN é usado em aplicações como classificação de imagens, reconhecimento de padrões, detecção de outlier, previsão de séries temporais e muito mais. É facilmente aplica-lo.

Em mercados altamente competitivos, uma gestão adequada e eficaz dos estoques pode inclusive se tornar um diferencial importante de uma empresa, aumentando sua produtividade e sua lucratividade. As Redes Neurais, o kNN (K Nearest Neighbor) e o SVM (Support Vector Machines) são alguns dos métodos computacionais mais comuns utilizados para a classificação de estoque. Estes métodos podem ser usados para classificar um conjunto de dados de estoque em categorias diferentes, como a categoria de preço, quantidade, qualidade, etc. Ao longo deste estudo, será explicado brevemente cada um destes métodos computacionais e como eles podem ser usados para classificar estoque.

Diante o exposto esse estudo busca validar a aplicação de técnicas de Machine Learning na gestão de estoque de uma empresa de grande porte de engenharia e na classificação ABC multicritérios em uma base de dados real desta indústria brasileira. O principal objetivo foi avaliar a viabilidade da utilização de diferentes técnicas de ML na tarefa de classificar produtos de forma a produzir uma classificação ABC, comparando suas performances entre si e com uma classificação de referência.

O principal objetivo foi realizar uma análise comparativa de métodos computacionais para a classificação de estoque com os métodos de Redes Neurais, kNN e SVM comparando suas performances entre si e com uma classificação de referência. Para alcançar este objetivo, foi realizada uma série de etapas: 1. Pré-processamento e limpeza dos dados 2. Análise de características e seleção de atributos 3. Treinamento, validação e avaliação dos modelos selecionados 4. Comparação dos resultados e conclusão. Os modelos selecionados foram kNN, Redes Neurais e SVM. Os dados foram pré-processados em uma biblioteca SQL para remover NaNs e outliers e os atributos relevantes foram selecionados para cada modelo. Os modelos foram treinados usando *k-fold cross-validation* e avaliados usando a acurácia.

## **2. Métodos para o controle de estoques**

### **2.1 Planejamento e controle de estoques**

O planejamento de controle de produção (PCP) é um conjunto de ações inter-relacionadas, que objetiva direcionar o processo produtivo da empresa e coordená-lo com os objetivos do cliente, sendo o maior alvo da negociação.

De acordo com Viana (2000), as atividades do PCP são exercidas nos três níveis Hierárquicos do planejamento e controle das atividades produtivas de um sistema de produção. No nível estratégico é estabelecido o plano de produção, que fará a estimativa de vendas de longo prazo e a disponibilidade de recursos financeiros e produtivos. Este plano geralmente é pouco detalhado e estabelece um plano para um longo prazo. Em médio prazo, o plano-mestre de produção buscará táticas para operar, de forma eficiente, o plano de produção. Este é um plano que analisará diferentes formas de guiar o sistema produtivo disponível. A programação da produção em curto prazo se encarregará de estabelecer de quando comprar, fabricar ou montar os itens necessários aos produtos finais.

Além disso, é importante ter um procedimento normativo e organizado do processo de programação, que permita controlar todas as etapas de produção, prover possíveis desvios nos prazos e tomar decisões de correção.

Segundo Pozo (2009), o planejamento e controle de produção é um sistema de transformações e informes de marketing, engenharia, fabricação de materiais, no qual são manuseadas as informações a respeito de vendas, linha de produtos, capacidade produtiva, potencial humano, estoques existentes e previsões para atender às necessidades de vendas; sua tarefa é transformar os planos em ordens viáveis de fabricação.

#### **2.1.1 Controle de compras**

A gestão de compras tem um papel fundamental e estratégico para toda organização, principalmente para o departamento de compras. É responsável por todo o processo, desde o setor financeiro, onde se tem uma visão geral da aplicação dos recursos feitos por meio das compras, até a satisfação do cliente. Atualmente o departamento de compras se tornou muito importante para qualquer organização, sendo responsável por toda movimentação de matéria prima até a pronta entrega (VIVALDINI, 2010).

A Cadeia de suprimentos Supply Chain Management (SCM) é entendida como um processo de expansão que envolve fornecedores-clientes na área de compras, mediante o crescimento do volume

de compras feitas pela empresa. O processo de concentração das suas atividades varia conforme os custos fixos e variáveis e acontece sob uma perspectiva global, por isso as empresas mudaram seus procedimentos de compras.

Muitos desafios foram encontrados nesse processo de mudança, começando pelo comércio eletrônico, que adotou novos modelos nos sistemas logísticos, dentre outros, que fizeram o profissional de compras expandir seus horizontes, muito além dos simples processos.

Atualmente, empresas que adotam processos de terceirização acabam relacionando o trabalho na área de compras, que se confunde com Supply Chain Management (SCM), que também é entendida como uma área de expansão do marketing, e que tem como função identificar as necessidades do mercado, e como é feita a internalização dentro da área produtiva. Através do desenvolvimento do marketing vários setores industriais interagiram junto à cadeia de suprimento como um todo.

Segundo Russo (2009), compras é um segmento essencial do departamento de materiais ou suprimentos, que tem por finalidade suprir as necessidades de materiais e serviços, planejá-las quantitativamente e satisfazê-las no momento certo, com as quantidades corretas, verificar o que foi comprado e providenciar o armazenamento. Portanto, compras são uma operação essencial entre as áreas que compõem o processo de suprimento.

Todos os processos ligados à compra e produção iniciam-se com um planejamento feito diante das vendas efetuadas pela empresa. No planejamento de compras devem ser definidas as quantidades a serem compradas somente na etapa final da elaboração do plano de produção. As quantidades líquidas a comprar serão apuradas pela desagregação das fichas de produção e os tipos de compras variam do desejo de consumir ou investir. As compras para investimento decorrem da necessidade de compras de bens e equipamentos que compõem o ativo da empresa (VIVALDINI, 2010).

### **2.1.2 Controle dos estoques**

O controle de estoque é um processo que fiscaliza e gerencia a entrada e saída dos produtos dentro da cadeia de suprimentos. O seu objetivo é minimizar o custo do capital investido, pois sua gestão, quando mal executada, causa grandes prejuízos para a organização, como falta de matéria prima para a produção, comprometendo toda a cadeia de suprimentos. A gestão e o controle de estoque, no setor logístico, ocorrem em função da necessidade de estipular os diversos níveis de materiais e produtos que a organização deve manter, dentro de parâmetros econômicos.

Esses materiais e produtos que compõem os estoques são: matéria-prima, material auxiliar, material de manutenção, material de escritório, material de peças em processos e produtos acabados. E a razão pela qual é preciso tomar uma decisão acerca das quantidades dos materiais a serem mantidos em estoques, está relacionada com os custos associados, tanto ao processo como aos custos de estocar.

Devem-se determinar quais são os níveis para cada item que poderemos manter economicamente. É notório que todas as organizações devem preocupar-se com o controle de estoque, visto que desempenham e afetam, de maneira bem definida o resultado da empresa (POZO, 2001).

Russo (2009) define que os estoques podem ser tanto a matéria-prima, como o material em processo e o produto acabado. As empresas precisam de estoques para trabalhar, pois sem eles torna-se impossível, a realização de suas atividades, tendo com meta principal maximizar seus lucros sobre o capital investido, pois o bom atendimento as vendas e produção dependem do seu nível. Assim, a busca pelo estoque ideal é imprescindível.

Uma das razões por que muitas empresas mantêm estoques em elevados padrões modernos, é que essa atitude permite comprar e produzir em lotes econômicos, que é a visão ultrapassada da produtividade. A importância da correta administração de bens é atender às necessidades de mercado, sendo desnecessário dizer que requisitos de nível de serviços são particularmente altos para suprimento de matéria-prima, até mesmo quando todos os requisitos de produção são atendidos, podendo-se notar a existência de uma administração inadequada de estoques (VIVALDINI, 2010).

A boa administração de materiais significa coordenar a movimentação de suprimentos com as exigências de produção. Isso significa aplicar o conceito de custo total de atividades de suprimentos logístico, de modo a obter vantagem da contraposição da curva de custo, ou seja, o objetivo maior da administração de materiais é prover o material certo, no local de produção certo, no momento certo e em condições utilizáveis ao custo mínimo, para plena satisfação do cliente e dos acionistas.

As teorias sobre o controle de estoque ressaltam a seguinte premissa: é possível definir uma quantidade ótima de estoques de cada componente e dos produtos da empresa, entretanto, só é possível defini-la a partir da previsão da demanda de consumo.

De acordo com Dias (1993), para organizar um setor de estoque é necessário primeiramente descrever suas principais funções: que são determinar os números de itens que devem permanecer nos estoques, determinar o reabastecimento dos mesmos, e quanto estoque é necessário para um determinado período. De uma forma simplificada, uma cadeia de suprimentos pode ser definida como os processos envolvendo fornecedores-clientes, ligando empresas, desde a fonte inicial de matéria-prima até o ponto de consumo do produto acabado. Por sua vez, define uma cadeia de suprimentos como um conjunto de organizações que estão envolvidas, através das ligações à jusante e à montante, nos diferentes processos e atividades que produzem valor, na forma de produtos e serviços liberados ao consumidor final (SILVA FILHO, 2000).

O objetivo do controle de estoque é conseguir o produto certo, no tempo exato, para satisfazer seus clientes e evitar a falta de material, sem que esta diligência resulte em estoque excessivo às reais necessidades da empresa. O controle de estoque é de responsabilidade das empresas, quando fazem

grandes investimentos do seu capital, pois esse investimento está ligado ao capital financeiro, que influencia a situação econômica da empresa.

### **2.1.3 Porque e como reduzir os níveis de estoque**

Inventário é um dos ativos mais valiosos que uma empresa possui. No entanto, muitos empresários comentam sobre o erro de não reconhecer a importância de estabelecer uma gestão eficiente do mesmo. Inventários adicionam utilidade de tempo, lugar e quantidade ao produto. Esse produto que não está no lugar determinado, no tempo necessário e nas quantidades solicitadas, perde possibilidades de consumo.

Ao manter um inventário, é possível ter uma fonte de fornecimento direta e contínua a qualquer momento, o que proporciona um serviço constante e eficiente ao cliente. Como em muitos processos de uma empresa, o gerenciamento adequado de estoques requer uma boa contabilidade. Seu objetivo é identificar os valores que devem ser incluídos no custo dos produtos que os serviços venderam durante um período e planejar o próximo. Contabilidade neste processo deve ser rigorosa, porque se houver um erro na avaliação de estoque, isso será refletido nos valores de custo de vendas vendidos, o lucro bruto em vendas e lucro para o período. Redução dos níveis de estoque torna-se questão de sustentabilidade para as empresas no atual mercado. As empresas cada vez mais buscam dar disponibilidade do produto a seu cliente diminuindo o máximo possível seu estoque (VARGAS, 2005).

De acordo com Vivaldini (2010) há vários fatores que colaboram para essa política, como a diversidade do mix de produtos, o que deixa mais complexa a gestão do estoque, do ponto de pedido e do estoque de segurança. O autor aborda ainda outros fatores como: o elevado custo de oportunidade de capital, que tem deixado à manutenção do estoque mais complicada e o foco da gerencia em reduzir o seu capital circulante líquido, medida adotada pelas empresas que desejam aumentar os indicadores do valor econômico.

Esses fatores influenciam a gestão de estoque a aumentarem a eficiência no processo de movimentação de materiais. Segundo Vivaldini (2010, p. 48) “aumentar a eficiência nesses processos resulta num menor custo unitário, propiciando a operação com lotes menores de estoque, de forma que não afete a disponibilidade do produto ou um custo logístico adicional”.

Para que as empresas atinjam essa redução, vem sendo formadas parcerias com outras organizações, resultando na cadeia de suprimentos, onde todos trabalham com um objetivo único. E é cada vez mais comum o investimento em novas tecnologias, capazes de administrar as informações dando suporte para as empresas nas tomadas de decisões.

## 2.2 Agrupamento de produtos com o método dos k-zinhos mais próximos

A classificação de estoques se dá acerca dos processos de organizar e categorizar os itens dentro de um sistema de estoque. Um dos métodos mais atuais e comuns para a classificação é através do algoritmo de K-zinhos mais próximos. É um algoritmo que utiliza aprendizado de máquina que possui uma grande variedade de aplicações, incluindo, a classificação de estoque (DRUCKER, 1999).

O KNN funciona através da identificação do K-ésimo item mais próximo utilizando uma certa composição de atributos. Os atributos podem ser definidos como características, quantidades de consumo ou venda ou qualquer outro aspecto relevante para os itens. Quando um novo item é inserido ao estoque, o mesmo é categorizado de acordo com os seus “vizinhos” mais próximos, isto é, outros itens que apresentam aspectos parecidos. Este método é muito útil para identificar padrões e relações entre os diversos itens dentro de um estoque, que não seria visível através de um sistema analítico manual (ANDO, 2005).

Segundo Drucker (1999), mesmo que o KNN nem sempre o melhor método para classificar um estoque, uma vez que possuem um custo computacional elevado, especialmente para grandes volumes de dados. Outro grande dilema com o método é a escolha do fator K, o qual o deixa suscetível a influências pelos parâmetros. Se o fator K for muito grande, isto é, possuir diversos aspectos, o algoritmo não irá reconhecer padrões de forma relevante. Já se o fator K for muito pequeno, a divisão entre classes não será visível. Outro aspecto importante é a questão das relações entre os aspectos de item, isto é, se não houver uma relação linear, o algoritmo não será apropriado para classificar o estoque.

A implementação do método se inicia com o pré-processamento dos dados, ou seja, uma vez com os dados em mãos, deve-se averiguar que o banco de dados possui os aspectos que serão classificados. Essa fase é importante para tratar quaisquer variações como a falta de dados ou valores incompatíveis. Uma vez com os dados tratados, deve-se dividir o banco de dados em duas partes. A primeira parte será utilizada para treinar o algoritmo e a segunda para avaliar a performance do método. A escolha do parâmetro K irá determinar o número de vizinhos mais próximos para classificar um novo item. Normalmente o de K será um número ímpar e precisa ser feito de modo que não classificar demais os itens nem deixar de classificar, tornando tudo uma grande “classe”. Uma vez que o algoritmo estiver com seus parâmetros definidos, o mesmo estará pronto para classificar um novo item. É importante avaliar a performance do algoritmo, comparando a classe presumida com sua classe de fato (ANDO, 2005).

Assim, o método K-Nearest Neighbors (KNN) é uma poderosa ferramenta para a gestão de estoque. Ele fornece informações detalhadas sobre padrões de comportamento de compra e permite

prever a procura futura, baseando-se na procura passada. Isso permite às empresas dimensionarem seus estoques de forma eficiente, de acordo com as necessidades de seus clientes, ao mesmo tempo em que minimiza os custos de armazenamento e o risco de falta de estoque. Além disso, o KNN pode ser usado para prever a demanda sazonal para determinados produtos, permitindo que as empresas tomem decisões informadas sobre quanto estoque adquirir para períodos específicos.

### **2.3 Agrupamento de produtos com o método SVM**

O SVM é um método de classificação de estoque baseado em computação de vetores de suporte. O algoritmo SVM cria um plano de divisão entre várias classes de dados. O algoritmo é usado para a classificação de dados de estoque a partir de um conjunto de dados existente. O algoritmo SVM usa um conjunto de dados de treinamento para criar um plano de divisão entre duas ou mais classes de dados (GROOVER, 2000). O SVM usa um algoritmo baseado em análise discriminante para encontrar uma linha de separação entre grupos de dados. O algoritmo cria um hiperplano que separa os dados em grupos distintos. O hiperplano é definido com base em características específicas de cada grupo, como preço, qualidade, tamanho, etc. Usando essas características, o algoritmo tenta encontrar a melhor separação entre os grupos. Depois que a separação é definida, os produtos são classificados e agrupados.

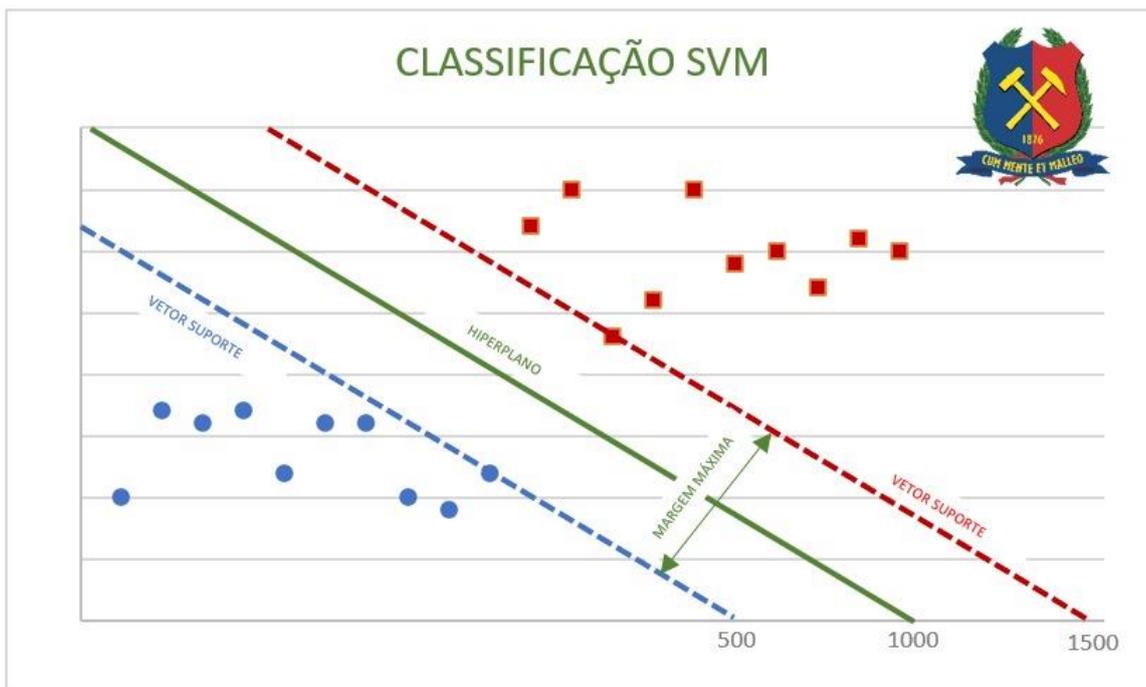
O SVM pode ser usado para criar grupos hierárquicos de produtos, permitindo aos usuários identificar produtos semelhantes e ajudar a prever comportamentos de compra. A SVM é uma ferramenta poderosa para classificação e agrupamento de estoques e pode ser usada para obter insights sobre padrões de comportamento de compra e identificar produtos semelhantes.

O algoritmo SVM usa um conjunto de dados de treinamento para criar um plano de divisão que maximiza a distância entre as classes, a fim de obter a melhor classificação possível. Uma vez que o plano de divisão é criado, é usado para classificar novos dados de estoque. O SVM é uma ferramenta útil para a classificação de dados de estoque, pois é capaz de lidar com problemas de classificação não-lineares (MITCHELL, 1997).

O método SVM (Support Vector Machine) é uma técnica de aprendizado supervisionado que pode ser usada para agrupar produtos. Ele usa algoritmos para criar modelos matemáticos que podem classificar dados para prever resultados futuros. O método SVM pode ser usado para analisar um conjunto de dados contendo características de produtos, como preço, qualidade, marca e outros fatores. A partir desses dados, o SVM criará um modelo matemático para agrupar os produtos em diferentes categorias (FEURER, 2019).

O agrupamento destas categorias permitirá que os compradores encontrem produtos similares de forma mais eficiente. Além disso, o SVM também pode ser usado para prever o comportamento dos compradores. Ele pode ajudar a detectar padrões de compra, como quais produtos os compradores compram com mais frequência e quais produtos eles tendem a evitar. Esta informação pode ser usada para ajudar a direcionar ações de marketing ou otimizar o layout da loja para melhorar a experiência do comprador.

**Figura 2** Representação gráfica de uma classificação SVM para determinar a acurácia



FONTE: O AUTOR

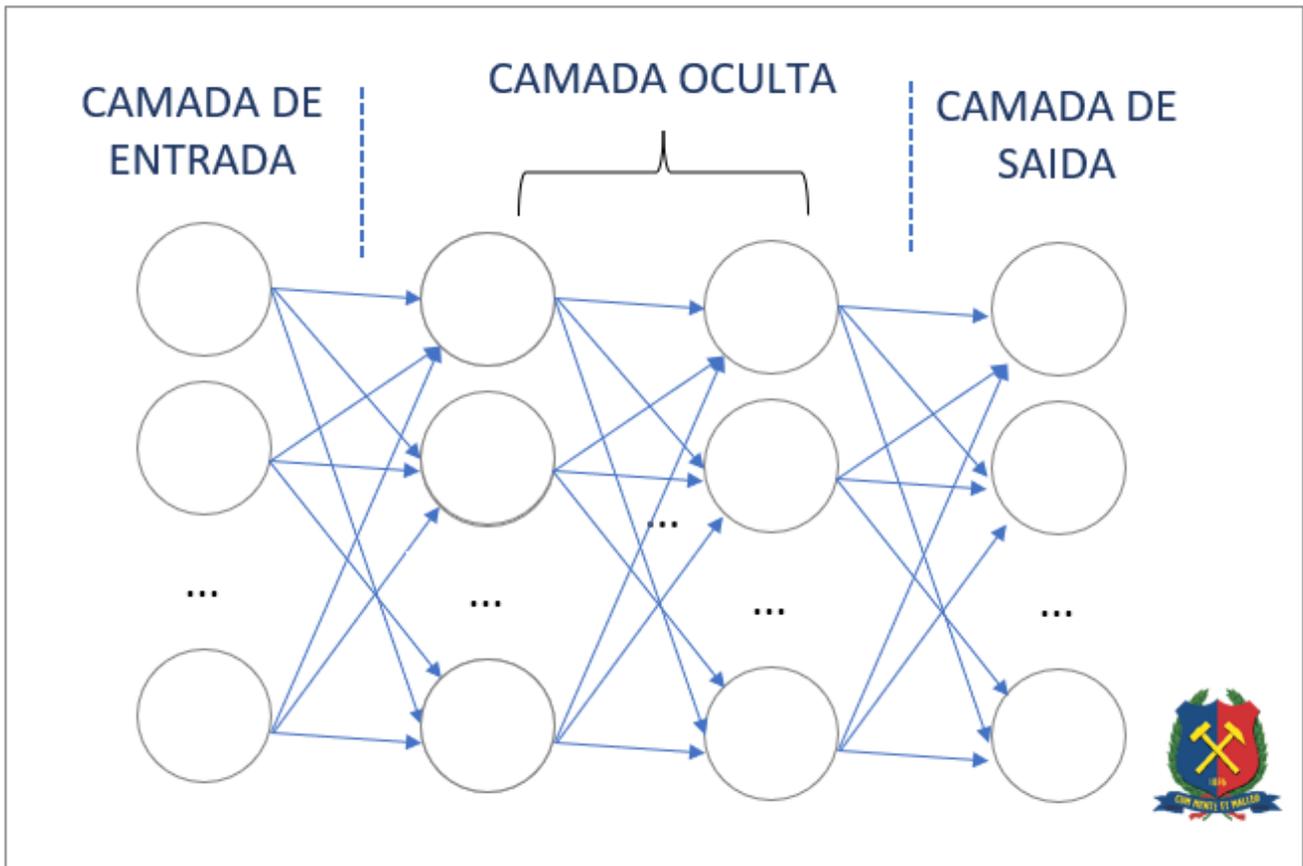
## 2.4 Agrupamento de produtos com redes neurais

As redes neurais são ferramentas de inteligência artificial que podem ser usadas para agrupar produtos. Elas podem ser treinadas para detectar padrões nos dados de entrada, permitindo que os produtos sejam classificados em categorias ou agrupados com base em suas características. As redes neurais podem ser usadas para classificar produtos com base em características como preço, marca, local de venda e características físicas. Isso permite que empresas criem grupos de produtos semelhantes que podem ser usados para ações de marketing eficazes. Elas também podem ser usadas para identificar produtos que se encaixam melhor em determinado contexto de compra, permitindo que as empresas direcionem melhor as vendas.

Além disso, as redes neurais podem ser usadas para detectar fraudes, pois elas podem ser treinadas para identificar padrões anormais nos dados de entrada. Isso permite que empresas detectem e previnam fraudes relacionadas a compras de produtos. Em suma, as redes neurais são ferramentas

incrivelmente úteis para agrupar produtos. Elas podem ser treinadas para detectar padrões nos dados de entrada, permitindo que empresas criem grupos eficientes e direcionem melhor as vendas. Além disso, elas também podem ser usadas para detectar fraudes relacionadas à compra de produtos.

**Figura 3** Representação gráfica de uma rede neural artificial para determinar a acurácia



FONTE: O AUTOR

## 3. Implementação computacional e datasets

### 3.1 Base de dados

Para obter os dados de estoque da Industria Brasileira em estudo, um funcionário engenheiro de produção disponibilizou um arquivo acerca do datasets de estoque existente nesta indústria. Realizou-se o download do mesmo e capturado as informações gerais que estavam descritas. Estava disponibilizado no formato CSV (APÊNDICE A), e é um conjunto de dados “sujo” e com alguns de seus atributos já tratados. Cada linha deste CSV sendo um registro de produtos no estoque.

O conjunto de dados utilizado como objeto de estudo para este trabalho consiste em uma base de dados real de uma empresa do ramo de engenharia e construção. Os dados compilados no arquivo original contia informações do estoque da Indústria, que chamaremos a partir daqui de Nuvuk. Os dados tratados eram de um espaço de tempo entre 2019 e 2022.

Primeiro passo foi limpar os dados, realizou-se uma filtragem em subconjuntos (2019; 2020; 2021; 2022) transformando e selecionando as principais características para melhor leitura dos métodos de aprendizagem de máquina. Removeu-se os valores faltantes e outliers, e também transformou os dados em um formato legível para a aplicação dos algoritmos, que foi o formato SQL.

As técnicas baseadas em métodos de filtragem são fundamentadas na ideia de se fazer um ranking de importância das variáveis (características) disponíveis no dataset. De acordo com Chandrashekar e Sahin (2014), estes métodos de ranking são amplamente utilizados devido a sua simplicidade e sua eficácia em aplicações práticas. O propósito por trás desta abordagem consiste em se atribuir um peso (score) para cada característica, de acordo com algum critério escolhido, e na definição de um limite (*threshold*), de tal forma que as variáveis cujo score estiver abaixo do *threshold* definido devem ser eliminadas da base de dados. As técnicas de seleção de características baseadas em "filtragem" recebem esta denominação pois elas são aplicadas antes de se fazer a classificação das características (*feature*), fazendo assim um filtro e removendo as variáveis irrelevantes ou com baixa importância.

Uma característica (*feature*), pode ser definida como uma propriedade individual e mensurável do processo que está sendo observado. É a partir de conjuntos de características que os algoritmos de aprendizado de máquina conseguem fazer classificações. Deste modo, é de suma importância que as variáveis que descrevam determinados processos (características) estejam estruturadas de forma a contribuir positivamente nos modelos de inferências.

De acordo com Chandrashekar e Sahin (2014), nos últimos anos as aplicações de reconhecimento de padrões e aprendizado de máquina vem se tornando mais robustas e complexas, de tal forma que o domínio de características analisadas acabou aumentando significativamente de dezenas para centenas de variáveis observadas. Diversos fatores contribuem para este aumento no número de características, pois com o advento

de novos sensores e tecnologias, assim como surgimento de bases de dados públicas e outros aspectos, o cientista de dados (profissional que realiza esse tipo de trabalho) possui muito mais informações para analisar e combinar, afim de se elaborar modelos de classificações, predições, clusterizações e entre outros.

No entanto, conforme é discutido em Cai et al. (2018), utilizar uma grande quantidade de informações não significa um modelo de classificação mais eficiente, uma vez que estes dados acabam sendo difíceis de se interpretar, dada suas amplas dimensões. Um outro ponto importante a se destacar é que essa grande quantidade de dados acaba tendo um impacto no desempenho do modelo e pode vir a se tornar um problema. É neste contexto que se fez necessário o uso da técnica de seleção de características.

Utilizou-se datasets para treinar o aplicativo de gerenciamento de estoque. Esses datasets foram construídos previamente a partir dos subconjuntos criados, coletados manualmente e contendo informações sobre os estoques, bem como tendências históricas de vendas. Para concretizar este passo utilizou-se o critério de correlação e o de informação mútua através dos métodos wrapper. Os métodos baseados em wrapper são considerados métodos de caixa preta, pois dado um conjunto de características  $E$ , fazem a predição de um subconjunto  $D$ , tal que  $D \subset E$ , e utilizam o erro de classificação ou a acurácia para avaliar se o subconjunto  $E$  foi uma boa seleção de características (CHANDRASHEKAR; SAHIN, 2014).

De acordo com Cai et al. (2018), este tipo de seleção de características tende a produzir resultados com alta acurácia e um conjunto de características ótimas pequeno, porém, os modelos gerados possuem baixa capacidade de generalização e tempo computacional muito alto.

A metodologia de implementação computacional para o estoque de produção foi realizada através de programação Python para criar os aplicativos de gerenciamento de estoque. Esses aplicativos permitem que os usuários gerenciem seus estoques, acompanhem as vendas, criem relatórios de desempenho e muito mais.

Outro método utilizado afim de otimizar os dados foi o método *embedded* para diminuir os custos computacionais e de tempo gastos nas classificações de dos subconjuntos criados a partir do método wrapper. O propósito deste método foi combinar a filtragem e o wrapper. Esta combinação se deu a partir da técnica de seleção de características com o método LASSO - *Least Absolute Shrinkage and Selection Operator*, este método foi primeiramente formulado por Robert Tibshirani em 1996, é denominado como um tipo de regressão e de encolhimento com objetivo de suavizar atributos similares entre si na base de dados Python, buscando diminuir o grau de colinearidade do conjunto de dados. O funcionamento do mesmo consiste em, basicamente, se colocar uma restrição na soma dos valores absolutos dos parâmetros do modelo analisado, de tal forma que essa soma não pode ser maior que um limite fixado (upper bound). 58 A partir disso, o método aplica um processo de regularização (muitas vezes chamado de encolhimento), onde ele penaliza os coeficientes da regressão, fazendo com que alguns acabem se reduzindo a zero (importante destacar que essa penalização é feita mais abruptamente para atributos com alto grau de correlação entre si). Desta maneira, para fazer a seleção de características, são escolhidas apenas as características cujos coeficientes da regressão não foram zerados (FONTI; BELITSER, 2017). O objetivo deste processo é minimizar o erro de predição do modelo. Não será

aprofundado os detalhes deste método uma vez que tal explanação iria se estender muito. Maiores informações sobre o mesmo podem ser encontradas em Zhang, Guo e Edu (2016), Muthukrishnan e Rohini (2016) e Kamkar et al. (2015). É importante ressaltar que no presente trabalho utilizou-se a implementação do LASSO disponível no pacote *scikit-learn*, em Python, e que é amplamente discutida em Pedregosa et al. (2012).

Para implementar essa metodologia, foi necessário criar um script Python que leu os dados do banco de dados relacional, assim o arquivo CSV disponibilizado pela Indústria foi convertido para o formato SQL pela biblioteca de dados MySQL. Uma vez que o script Python leu os dados, ele foi usado para gerar gráficos, diagramas e relatórios que ajudaram a tomar decisões informadas sobre como gerenciar o estoque com o aprendizado de máquina e técnicas de machine learning.

Segundo Xin et al. (2018), aprendizado de máquina (machine learning) é um braço da inteligência artificial muito próximo da estatística computacional, que tem por objetivo fazer previsões utilizando computadores. As técnicas de aprendizado de máquina podem ser classificadas em duas categorias, de acordo com o tipo de aprendizagem. A primeira categoria diz respeito as técnicas de aprendizado supervisionado, onde inicialmente é separado um conjunto de dados de treinamento, previamente classificado (rotulado), que servirá como base de ensino para o modelo de previsão. Já a segunda categoria diz respeito as técnicas de aprendizado não-supervisionado, cujo foco está na separação (clusterização) de uma determinada base de dados, de forma autônoma, agrupando as observações similares. Como técnicas que utilizam aprendizado supervisionado pode-se citar *K-Nearest Neighbors* - KNN, *Support Vector Machines* - SVM, e redes neurais artificiais, classificações usadas neste estudo. Em relação as técnicas de aprendizado não-supervisionado, é muito comum o uso da técnica *k-means clustering*. No presente estudo optou-se por testar e avaliar alguns classificadores baseados em aprendizado supervisionado, que serão explanados e discutidos na seção seguinte.

A padronização de dados foi usada para transformar os dados em uma escala comum. Isso foi feito para permitir que os algoritmos de aprendizado de máquina processassem os dados de forma mais eficiente. A seleção de características foi usada para selecionar as características mais relevantes a partir dos dados. Isso foi feito para melhorar a precisão do modelo e reduzir a complexidade. A normalização de dados foi usada para ajustar os dados de forma que eles se encaixem dentro de um intervalo específico. Isso foi feito para permitir que os algoritmos de aprendizado de máquina processassem os dados de forma mais eficiente.

Eliminação de nulos foi usada para remover valores nulos dos dados. Isso foi feito para permitir que os algoritmos de aprendizado de máquina processassem os dados de forma mais eficiente. nestes casos essas amostras foram eliminadas do conjunto de dados, foi o caso da família do produto;

Substituição pela média / moda foi usada para substituir valores nulos nos dados por valores mais representativos. Isso foi feito para permitir que os algoritmos de aprendizado de máquina processassem os dados de forma mais eficiente. Foi o caso da quantidade de produtos. O percentual de nulos na base foi inferior a 1% dos dados, ou seja, entendemos que nenhum ruído foi produzido devido a este

tratamento; Normalização de Dados foi usada para mapear os dados para um intervalo específico. Isso foi feito para permitir que os algoritmos de aprendizado de máquina processassem os dados de forma mais eficiente. Foi o caso da acurácia de anos com o estoque mais baixo.

O próximo passo do trabalho foi preparar o datasets e extrair os atributos a serem utilizados na aplicação das técnicas de Machine Learning, de modo a realizar a classificação ABC com elas. Nesta etapa também foi feito um trabalho de refinamento dos parâmetros utilizados pelos modelos, de modo a obter uma melhor performance. Por fim, a comparação entre as classificações de cenário de referência, com as classificações geradas através das técnicas KNN, SVM e Redes Neurais serão comparadas em termos de acurácia.

Manteve-se o período de análise para a classificação ABC igual ao utilizado pela empresa, com isso, é possível obter um comportamento bem próximo de como seria a aplicação real dos métodos na classificação. A empresa utiliza um comportamento trimestre a trimestre, ou seja, para cada trimestre que se evolui no tempo, é possível utilizar os trimestres anteriores como dado de treino dos modelos de Machine Learning.

A base de dados utilizada consiste em um banco de dados SQL Server e a linguagem de programação utilizada para a aplicação das técnicas de ML foi a linguagem Python, através das bibliotecas scikit-learn e matplotlib

### 3.2 Bibliotecas de Python usadas

- Scikit-learn

*Scikit-learn* é uma biblioteca de aprendizado de máquina de código aberto para Python. Ela fornece algoritmos de aprendizado supervisionado e não supervisionado, incluindo regressão, classificação, *clustering*, redução de dimensionalidade, previsão, detecção de padrões e modelagem de fluxo de trabalho.

Os algoritmos são desenvolvidos para serem usados com dados numéricos e vetoriais, mas também podem ser usados com outros tipos de dados. Além disso, *Scikit-learn* possui uma gama de ferramentas para avaliação de modelos, como validação cruzada e teste de hipóteses. Os usuários da *Scikit-learn* podem facilmente construir modelos e testar seu desempenho usando uma variedade de métricas. Esta biblioteca é considerada a biblioteca de aprendizado de MÁQUINA.

A seguir, estão os passos básicos utilizados na biblioteca *scikit-learn*:

1. Carregue os dados: carregou-se os dados em um formato SQL para serem lidos pela scikit-learn, como um DataFrame do Pandas.

2. Preparar os dados: dividiu-se os dados em conjuntos de treinamento e teste e, se necessário, aplique normalização/padronização às características.

3. Escolha o modelo: escolheu-se o algoritmo de aprendizado de máquina que melhor se adequa ao seu problema.

4. Treinar o modelo: usou-se o conjunto de dados de treinamento para treinar o modelo escolhido.

5. Avaliar o modelo: avaliou-se o desempenho do modelo usando os dados de teste.

6. Ajustar os parâmetros do modelo: ajustou-se os parâmetros do modelo usando a validação cruzada para obter um melhor desempenho.

7. Prever resultados: usou-se o modelo treinado para prever resultados com novos dados.

- matplotlib

A Matplotlib é uma biblioteca de software de código aberto para criação de gráficos em Python. É projetada para produzir gráficos de qualidade de publicação em uma variedade de formatos de arquivo e ambientes interativos, como o Jupyter.

A Matplotlib é usada para criar gráficos de linha, histogramas, gráficos de dispersão, gráficos de barras, gráficos de pizza, gráficos 3D, gráficos de área, gráficos de mapas, gráficos de erros e muitos outros. Ela também fornece um conjunto de ferramentas para personalizar os gráficos, como controles de eixos, textos, cores, legendas, fontes e outros. Abaixo segue o passo a passo utilizado;

1. Instalação: A primeira etapa foi instalar a biblioteca. Para isso, usou-se o gerenciador de pacotes do Python, o pip. Digitou-se o comando “*pip install matplotlib*” no terminal (ou prompt de comando) e instalou-se a biblioteca.

2. Importação: A segunda etapa foi importar a biblioteca no código Python. Para isso, usou-se o comando “*import matplotlib*” no início do código.

3. Criação de gráficos: A terceira etapa foi usar os comandos da biblioteca para criar gráficos. Para isso, usou-se os comandos “*plt.plot()*” ou “*plt.scatter()*” para criar gráficos de linha ou gráficos de dispersão.

4. Detalhes extras: A quarta etapa foi adicionar detalhes extras aos gráficos. Como acurácia. Utilizou-se os comandos “*plt.title()*”, “*plt.xlabel()*” e “*plt.ylabel()*” para adicionar títulos, rótulos para eixos e legendas aos seus gráficos.

5. Exibição de gráficos: A quinta etapa foi salvar e exportar os gráficos. Para isso, usou-se o comando “*plt.show()*”. Este comando abriu uma janela contendo o gráfico que foi criado.

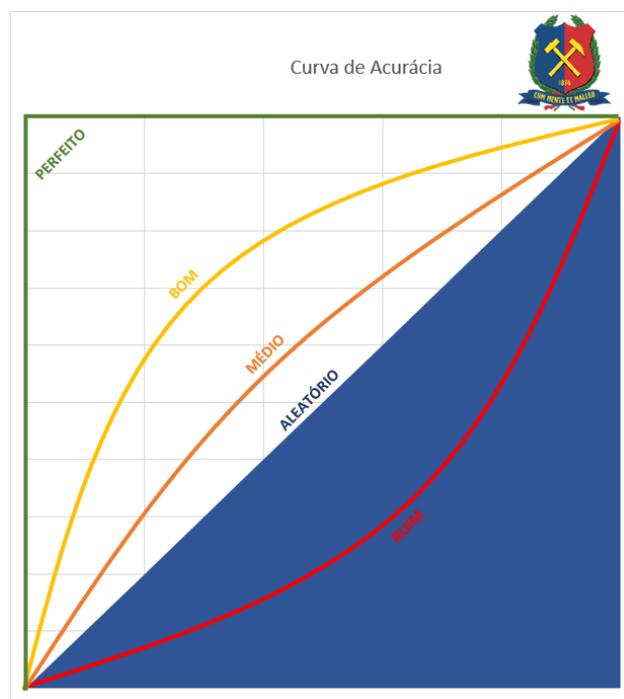
### 3.3 Definição de acurácia

Para avaliar os resultados dos diferentes métodos e compará-los entre si será utilizada a noção de acurácia comumente utilizada em Machine Learning em cenários de classificação e categorização, ou seja, ela é o percentual de previsões corretas feitas pelo método em relação ao total de previsões.

A acurácia é obtida dividindo o número de previsões corretas pelo total de previsões, em seguida multiplicando o valor por 100 para obter o valor em porcentagem.

No gráfico abaixo é possível visualizar como a classificação pode ser realizada, acerca da curva.

**Figura 4** Representação gráfica da curva de Acurácia



FONTE: O AUTOR

De maneira a exemplificar, em uma aplicação do método para classificar 1000 itens, se o método classificar 800 itens da forma correta e 200 itens de forma errônea, a acurácia do método é de 80%.

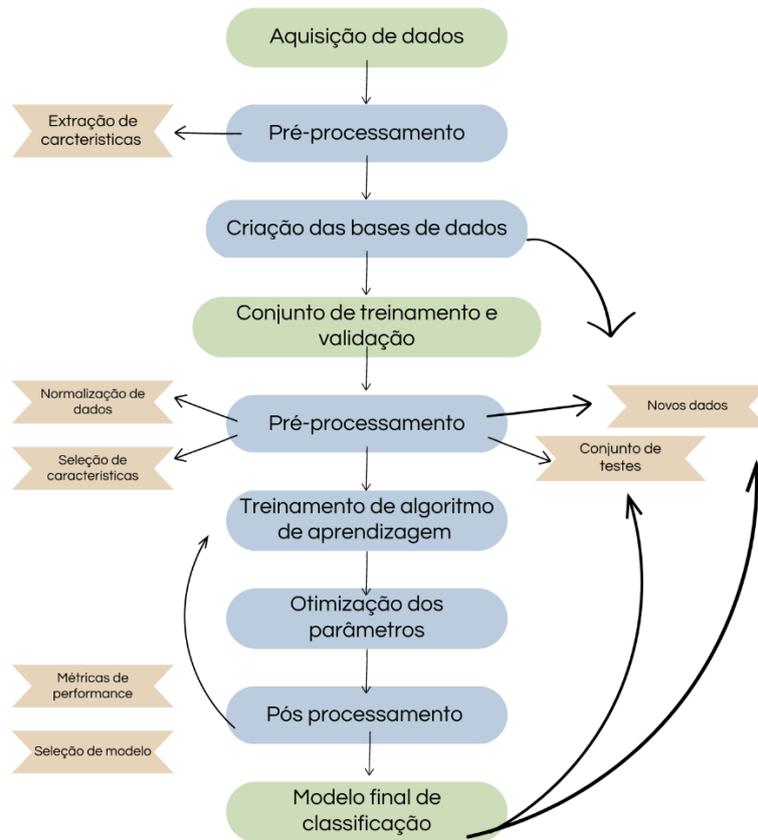
A acurácia é uma métrica útil uma vez que é de fácil interpretação e simples de observar a qualidade dos resultados.

### 3.4 Implementação do sistema

Uma abordagem comum para a classificação de acurácia dos métodos de aprendizado de máquina (KNN, SVM e RNA) é a validação cruzada. A validação cruzada é um processo que consiste em separar os dados em partes, treinar o modelo com uma parte e testar o modelo com a segunda parte. Para medir a acurácia dos modelos KNN, SVM e RNA, o primeiro passo é dividir os dados em treinamento e teste. O conjunto de treinamento deverá conter a maior parte dos dados, enquanto o conjunto de teste deverá conter uma pequena parte dos dados. Em seguida, o modelo KNN, SVM ou RNA deverá ser treinado usando o conjunto de treinamento. Uma vez treinado, o modelo pode ser usado para prever os resultados do conjunto de teste. Finalmente, a acurácia do modelo pode ser medida comparando suas previsões com os dados reais de teste. Essa comparação pode ser feita usando a métrica de acurácia para medir o quão precisa é a previsão do modelo. Em suma, a validação cruzada é uma abordagem comum para medir a acurácia dos modelos KNN, SVM e RNA. Ela envolve a divisão dos dados em partes, treinamento do modelo com uma parte e teste do modelo com a segunda parte. O resultado é uma métrica de acurácia que pode ser usada para avaliar o quão precisa é a previsão do modelo (KARTAL, 2013).

Assim, a implementação do sistema de classificação para medir a acurácia dos métodos de aprendizado de máquina (KNN, SVM E RNA) foi feita seguindo o fluxograma descrito na figura abaixo. Este fluxograma apresenta o método desde a obtenção das amostras de estoque nos subconjuntos (2019, 2020, 2021, 2022) até a avaliação dos modelos de classificação.

**Figura 5** Fluxo de execução do sistema de classificação dos métodos KNN, SVM E RNA para definição de acurácia de estoque.



FONTE: O AUTOR

## 4. Resultados e discussões

Os resultados apresentados nesta sessão fornecem uma base sólida para a compreensão da complexidade dos processos envolvidos na execução desta tarefa. Os resultados fornecem informações valiosas sobre a eficiência dos métodos utilizados e sobre a adequação dos recursos disponíveis para a realização da tarefa. Os resultados também podem fornecer diretrizes para a melhoria do processo de execução da gestão de estoques de Industrias com grande capacidade de produção. A discussão que se segue visa explorar mais profundamente os resultados obtidos e contribuir para o desenvolvimento de melhores estratégias para a execução da tarefa.

Conforme discutido acima, existem diversos algoritmos de classificação na literatura, sendo que a acurácia dos mesmos pode variar dependendo de diversos fatores como o tamanho da base de dados, a disposição dos dados, a quantidade de classes existentes, parâmetros (configurações) do algoritmo entre outros.

Nas subseções seguintes serão discutidos três algoritmos que foram aplicados no presente estudo: KNearest Neighbors (k-NN), Support Vector Machines (SVM) e Redes Neurais Artificiais (RNA). A ideia de se aplicar três algoritmos diferentes foi para se avaliar o desempenho de cada um na classificação de estoques e nível de acurácia. A opção por escolher os três classificadores acima citados não seguiu nenhum critério específico.

### 4.1 Classificação de estoques e agrupamento de produtos usando kNN

O k-NN (K – Nearest Neighbors) é um dos algoritmos de classificação mais populares em aplicações que envolvem aprendizado de máquina e mineração de dados. É um algoritmo simples e intuitivo, que usa a similaridade para identificar padrões nos dados. O algoritmo funciona com base na ideia de que os dados mais semelhantes a um determinado ponto de dados (ou observação) estão mais próximos a ele. Assim, o algoritmo calcula a distância entre um dado e todos os outros dados no conjunto de dados e encontra os k-vizinhos mais próximos. Estes vizinhos são usados para prever a classe do dado. O algoritmo pode ser usado para classificar novos pontos de dados com base na similaridade para pontos de dados conhecidos.

Pelo método de aprendizado supervisionado, o k-NN precisa de dois conjuntos de dados distintos: o conjunto de treinamento e o conjunto de testes. O conjunto de treinamento é composto por

uma matriz  $M \times N$ , onde  $M$  é o número de observações e  $N$  é o número de características de cada entrada e um vetor de tamanho  $M$  que define qual a classe de cada uma das observações. Basicamente o algoritmo é composto por três etapas principais que são:

1. Cálculo de distância;

o cálculo de distância faz a medida de quão longe um dado vetor de testes está em relação aos  $M$  vetores de treinamento. Os métodos para o cálculo destas distâncias vetoriais podem variar, porém, neste trabalho optou-se por utilizar a distância Euclideana. Dado dois vetores  $X$  e  $Y$  de tamanho  $n$ , a distância Euclideana  $D$  entre os mesmos é dada pela Equação:

**Tabela 1** Parâmetros para cálculo de distância KNN

X e Y: Dois vetores	n: Tamanho	i: Índice que percorre
---------------------	------------	------------------------

$$D = \sqrt{\sum_{i=1}^{i=n} (X_i - Y_i)^2}$$

2. Busca pelos vizinhos próximos;

na etapa de busca pelos vizinhos próximos as distâncias obtidas na etapa anterior são ordenadas de forma crescente, identificando quais vetores de treino estão pertos do vetor de teste.

3. Predição da classe.

na etapa de busca pelos vizinhos próximos as distâncias obtidas na etapa anterior são ordenadas de forma crescente, identificando quais vetores de treino estão pertos do vetor de teste.

Existem outros parâmetros que podem ser variados no método KNN, porém o foco foram estes dois parâmetros pelo entendimento que são os mais relevantes para a análise.

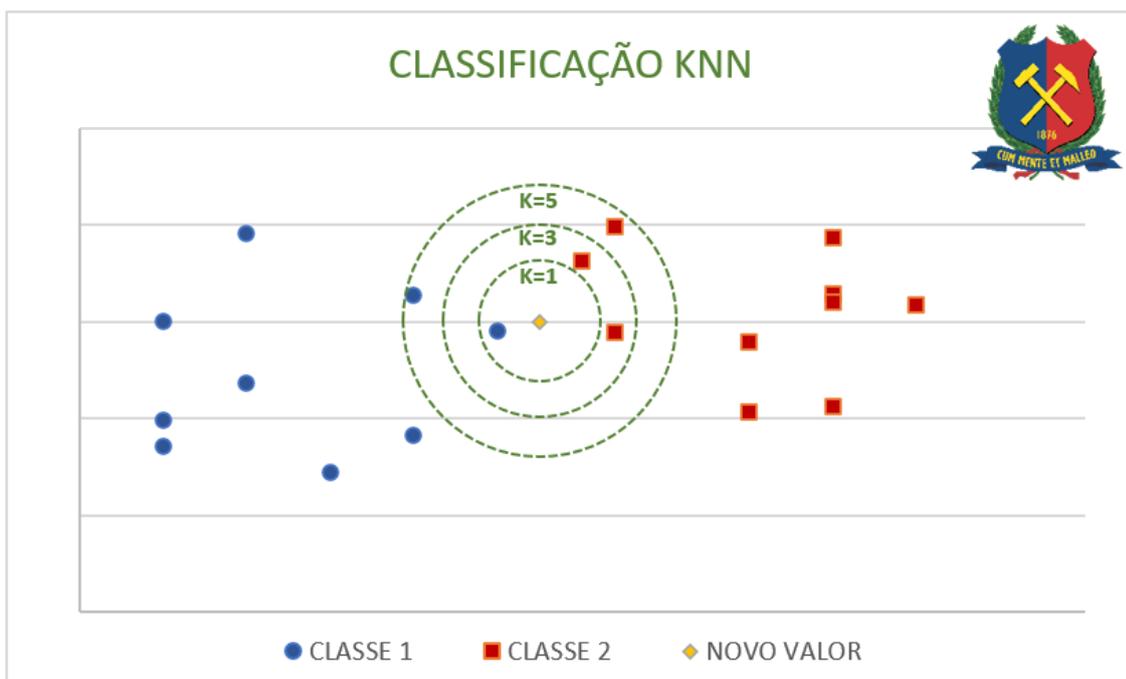
A acurácia do método kNN para gerenciamento de estoque depende de vários fatores, incluindo a qualidade dos dados de entrada, a complexidade da tarefa e as estratégias de treinamento e teste adotadas. Uma vez que os dados estejam disponíveis e um modelo seja construído, existem alguns métodos que podem ser usados para avaliar a acurácia do modelo. Um deles é o método de validação cruzada, que foi utilizado neste estudo, consistiu-se em dividir o conjunto de dados em 4 subconjuntos (2019; 2020; 2021; 2022), treinar o modelo em um subconjunto e validar o desempenho em outro.

Inicialmente foram gerados os gráficos em formato de barras para simplesmente indicar o número do fornecedor previsto pelo modelo. Para melhorar a visualização dos vizinhos, um gráfico de dispersão foi inserido.

Para melhorar ainda mais a visualização do gráfico de dispersão, foi utilizado diferentes cores para o gráfico. Foi necessário utilizar um data-frame específico para isso. A biblioteca utilizada foi o Matplotlib com a função ScalarFormatter para criar diferentes designs dentro do gráfico.

Primeiramente no gráfico de dispersão (Figura-3), era visível somente o valor da previsão, foi adicionado todos os outros fornecedores de dentro do banco de dados. Uma vez feito isso, foi necessário criar uma máscara para salvar os valores de K para aparecer em outra cor. Um conjunto de dados com zeros foi criado utilizando a biblioteca NumPy e a função Zeros para que quando os valores fossem atribuídos ao gráfico, não gerasse um erro de diferentes tamanhos de arrays. A única maneira encontrada de realizar essa operação foi utilizando booleanos.

**Figura 6** Dispersão de vizinhos na classificação KNN



Os tempos de processamento mudou conforme o valor de K aumentou, porem para um experimento de um banco de dados de 35mil linhas com 11 colunas, o tempo para k+2 aumentos em comente 1 segundo, aproximadamente.

**Tabela 2** Classificação KNN para treinamento e teste de acurácia

KNN			
K	t	acc. Train	acc. Test
1	52 seg	1	0,81
3	53 seg	0,87	0,76
5	54 seg	0,82	0,72
7	54 seg	0,78	0,69
9	55 seg	0,74	0,67
99	60 seg	0,43	0,42

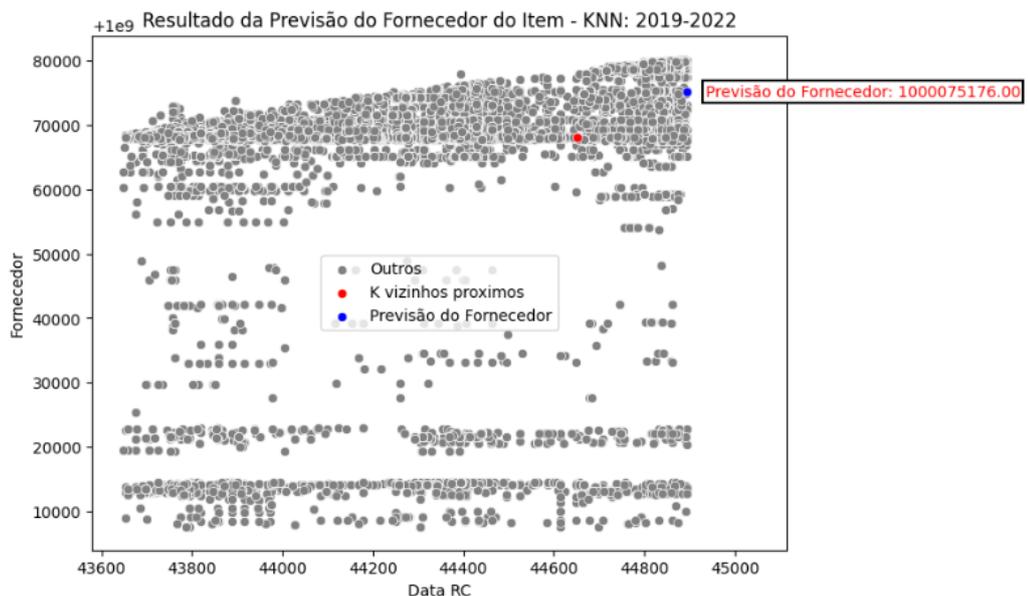
FONTE: O AUTOR

A acurácia do experimento se mostrou positiva, de modo que, quanto mais próximo de  $k=1$  indo de 81% nos testes com  $k=1$  até 67% com  $k=9$

Foi testado o valor de  $k=99$  para entender como o algoritmo se comportaria. O tempo de resolução foi aproximadamente o mesmo porem houve uma drástica diminuição da acurácia.

O tamanho do banco de dados para teste foi mantido em 30% durante todos os ensaios e o volume do banco de dados utilizados foram todos os 4 anos.

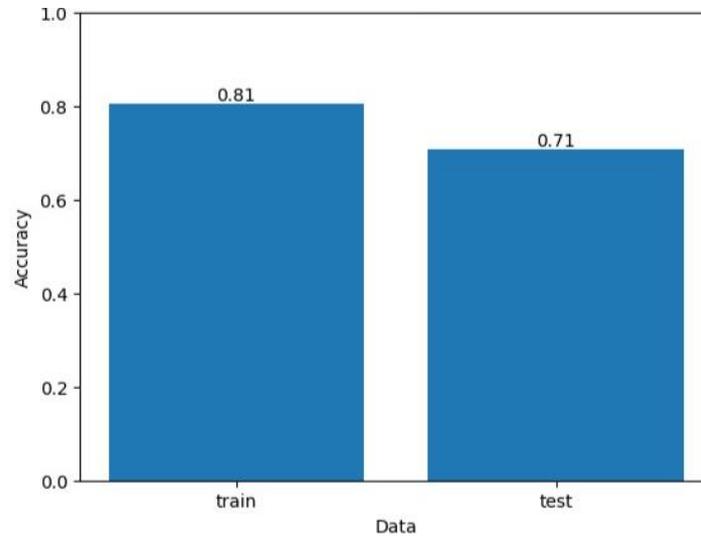
No gráfico-1 observa-se a previsão do Método KNN acerca do fornecedor do produto cuja ID do fornecer era 100000068026 da Industria Nuvuk no ano de 2019.

**Gráfico 1** Resultado da Previsão do fornecedor do item 10000068026- KNN-2019

FONTE: O AUTOR

No gráfico-2 obteve-se através do aprendizado de máquina do Método KNN o teste de acurácia do método em questão, apresentando acurácia de 80% para o treinamento e de 70% para o teste.

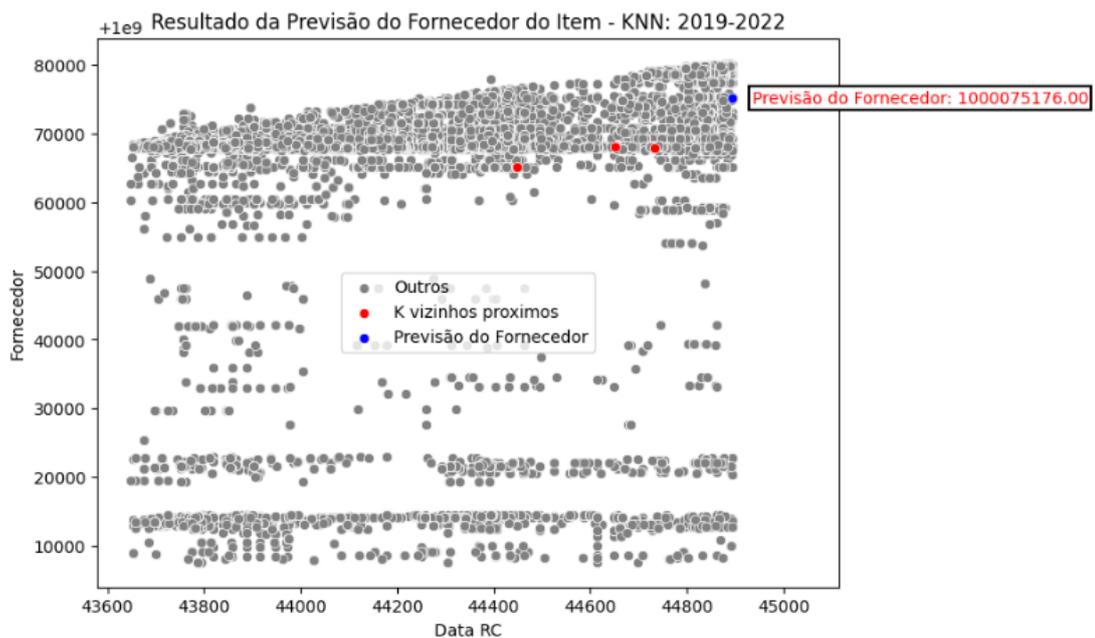
**Gráfico 2 Resultado de acurácia da classificação do inventário-KNN-2019**



FONTE: O AUTOR

No gráfico-3 observa-se a previsão do Método KNN acerca do fornecedor do produto cuja ID do fornecedor era 100000068026 da Industria Nuvuk no ano de 2020.

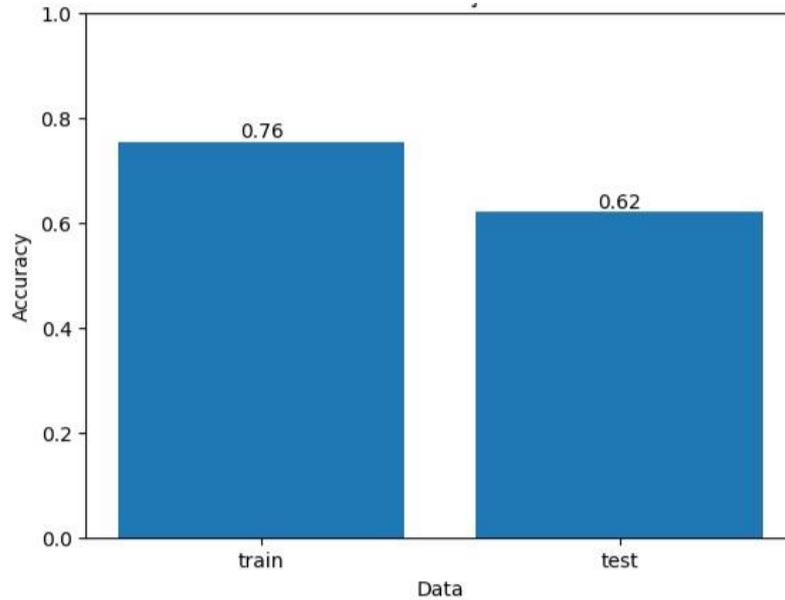
**Gráfico 3 Resultado da Previsão do fornecedor do item 10000068333 -KNN-2020**



FONTE: O AUTOR

No gráfico-4 obteve-se através do aprendizado de máquina do Método KNN o teste de acurácia do método em questão, apresentando acurácia de 70% para o treinamento e de 62% para o teste.

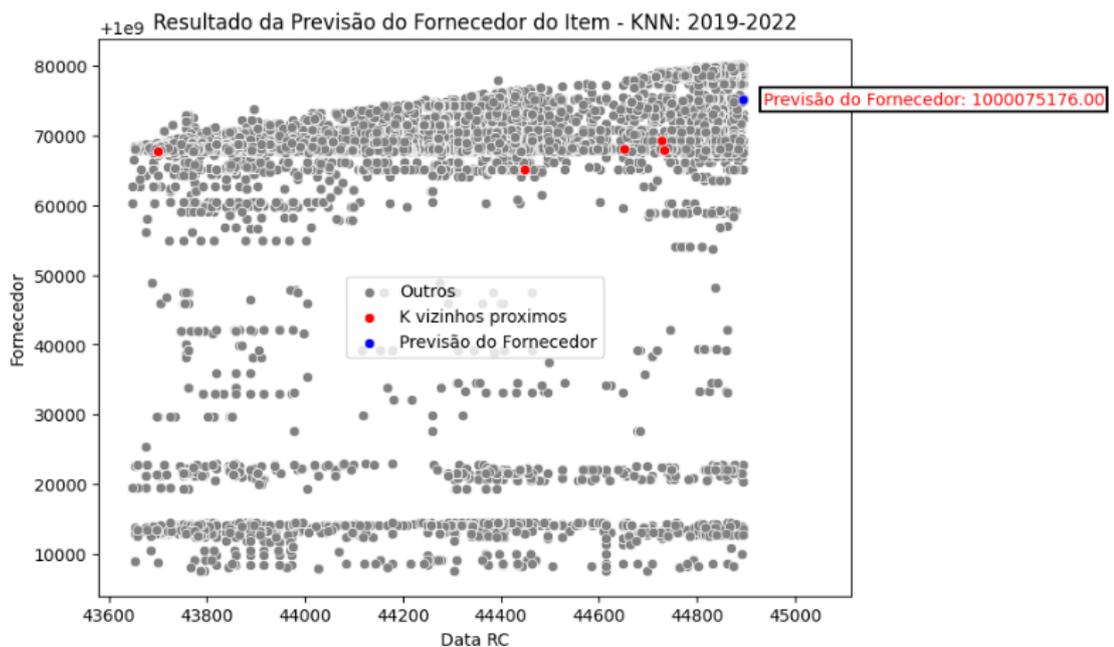
**Gráfico 4 Resultado da acurácia da classificação do inventário-KNN-2020**



FONTE: O AUTOR

No gráfico-5 observa-se a previsão do Método KNN acerca do fornecedor do produto cuja ID do fornecedor era 100000068026 da Industria Nuvuk no ano de 2021.

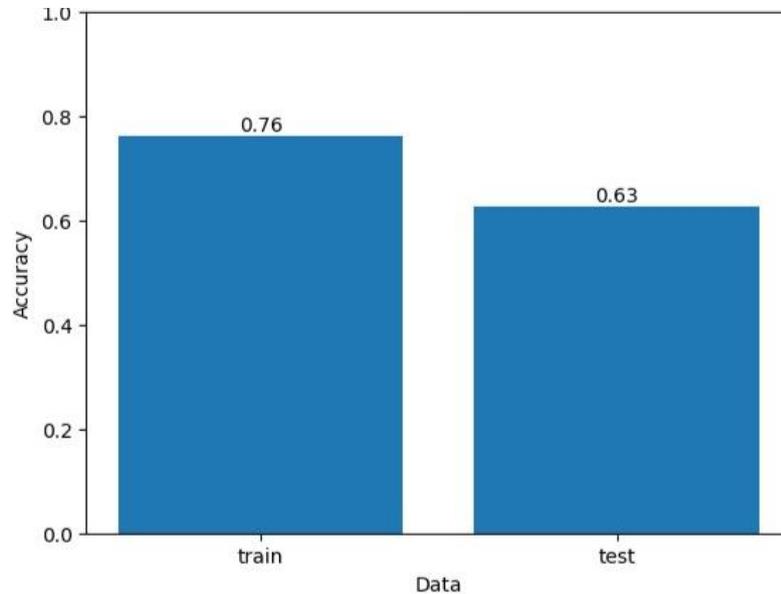
**Gráfico 5 Resultado da Previsão do fornecedor do item 10000068959 -KNN-2021**



FONTE: O AUTOR

No gráfico-6 obteve-se através do aprendizado de máquina do Método KNN o teste de acurácia do método em questão, apresentando acurácia de 76% para o treinamento e de 63% para o teste.

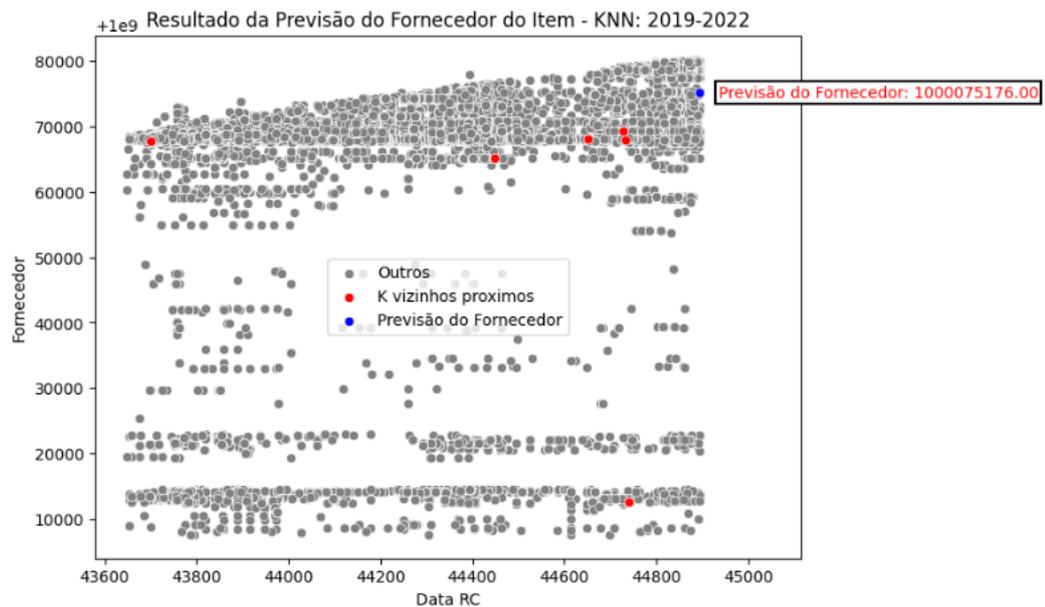
**Gráfico 6 Resultado da acurácia da classificação do inventário-KNN-2020**



FONTE: O AUTOR

No gráfico-7 observa-se a previsão do Método KNN acerca do fornecedor do produto cuja ID do fornecedor era 100000068026 da Industria Nuvuk no ano de 2022.

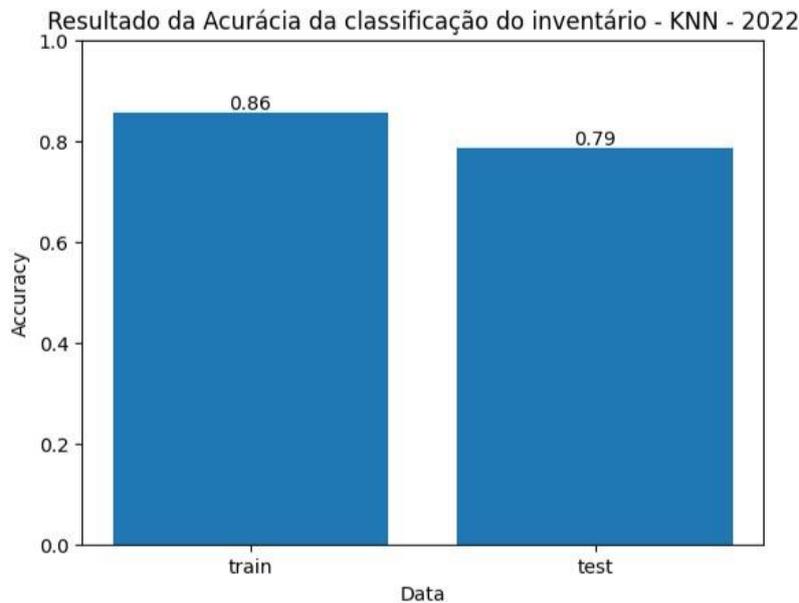
**Gráfico 7 Resultado da Previsão do fornecedor do item 1000012659 -KNN-2021**



FONTE: O AUTOR

No gráfico-7 obteve-se através do aprendizado de máquina do Método KNN o teste de acurácia do método em questão, apresentando acurácia de 86% para o treinamento e de 79% para o teste.

**Gráfico 8 Resultado da acurácia da classificação do inventário-KNN-2020**



FONTE: O AUTOR

Ao observar os resultados apresentados acima acerca da identificação do fornecedor e o nível de acurácia do método KNN pode-se inferir que esse resultado foi possível obter devido ao fato de que o Método KNN utiliza um número limitado de vizinhos para determinar a classificação do objeto, o que torna mais fácil para o algoritmo identificar padrões e classificar os dados de acordo com esses padrões. Além disso, o Método KNN é um método não paramétrico, o que significa que não é necessário definir parâmetros para o algoritmo e o algoritmo pode ajustar dinamicamente seus parâmetros ao longo do tempo, o que torna o processo mais eficaz e preciso.

Observa-se que a acuracidade do Método KNN foi maior nos anos 2019 e 2022, pode-se inferir que nestes anos o nível de estoque era muito menor, se comparado a 2020 e 2021, portanto infere-se que é provável que para empresas em que o estoque seja relevantemente extenso, o método KNN pode apresentar uma acurácia menor.

De acordo com Peterson (2009) a classificação de estoques e o agrupamento de produtos usando o algoritmo de classificação k-NN (k-Nearest Neighbors) é uma abordagem bastante útil para identificar e agrupar produtos similares em um estoque. O algoritmo funciona comparando os atributos de cada produto aos atributos de outros produtos no estoque e classificando-os de acordo com suas características similares. O algoritmo também pode ser usado para criar clusters de produtos com base

em seus atributos compartilhados. Este processo fornece uma maneira precisa e eficiente de classificar e agrupar produtos, permitindo que os gerentes de estoque tomem decisões mais informadas sobre a gestão de seus estoques.

Corroborando com os resultados deste estudo Chen (2020) acredita que o método kNN (K-Nearest Neighbors) é um algoritmo de aprendizado supervisionado usado para classificar e agrupar itens em um conjunto de dados. O algoritmo funciona medindo a distância mais curta entre um ponto de dados novos e os k pontos de dados mais próximos a ele nos dados de treinamento. O algoritmo determina a classificação de um novo ponto de dados com base na maioria dos k pontos de dados mais próximos. Isso significa que, ao usar o algoritmo kNN para classificar ou agrupar itens em um conjunto de dados, você pode classificar um item com base nas características dos itens mais próximos a ele. Isso torna o algoritmo kNN eficaz para classificar e agrupar itens em um conjunto de dados.

A classificação de estoques e o agrupamento de produtos usando o algoritmo de aprendizagem de máquina kNN (K-Nearest Neighbors) é um método eficaz para ajudar as empresas a organizar seu estoque e identificar quais grupos de produtos têm as mais altas taxas de venda. O algoritmo kNN funciona estabelecendo relações entre os dados de estoque e os dados de vendas de um conjunto de produtos. O algoritmo então identifica quais produtos são mais próximos uns dos outros e classifica os estoques de acordo com essas relações. O agrupamento de produtos é realizado examinando as características dos produtos, como preço, tamanho, cor e outras características. O algoritmo então classifica os produtos em grupos de acordo com essas características.

## **4.2 Classificação de estoques e agrupamento de produtos usando SVM**

O classificador SVM (Support Vector Machine), também conhecido como Máquina de vetores de suporte, é amplamente utilizado em diferentes tipos de problemas de classificação. É um algoritmo supervisionado que pode ser usado para classificar dados em várias classes. O SVM constrói um hiperplano (ou um conjunto de hiperplanos) em um espaço de dimensionalidade elevada que separa as classes de dados. O SVM é mais eficaz em problemas de classificação binária (ou seja, quando os dados podem ser classificados em duas classes). No entanto, também pode ser usado para problemas de classificação multi-classe. O SVM é robusto e pode lidar com grandes quantidades de ruído nos dados de treinamento. Além disso, os classificadores SVM são muito eficientes em termos de tempo de execução.

De acordo com Demidova e Klyueva (2017), o parâmetro que mais influência na classificação utilizando SVM é o tipo da função kernel, pois ele é responsável por determinar como o sistema irá separar os dados. Além disso, outros parâmetros que influenciam a acurácia do SVM são a escolha da margem e a regularização, que podem ser ajustados de acordo com os dados de entrada.

Baseado nestes conceitos, o método SVM foi aplicado à amostra de dados de treino, e os parâmetros estudados para esta técnica foram:

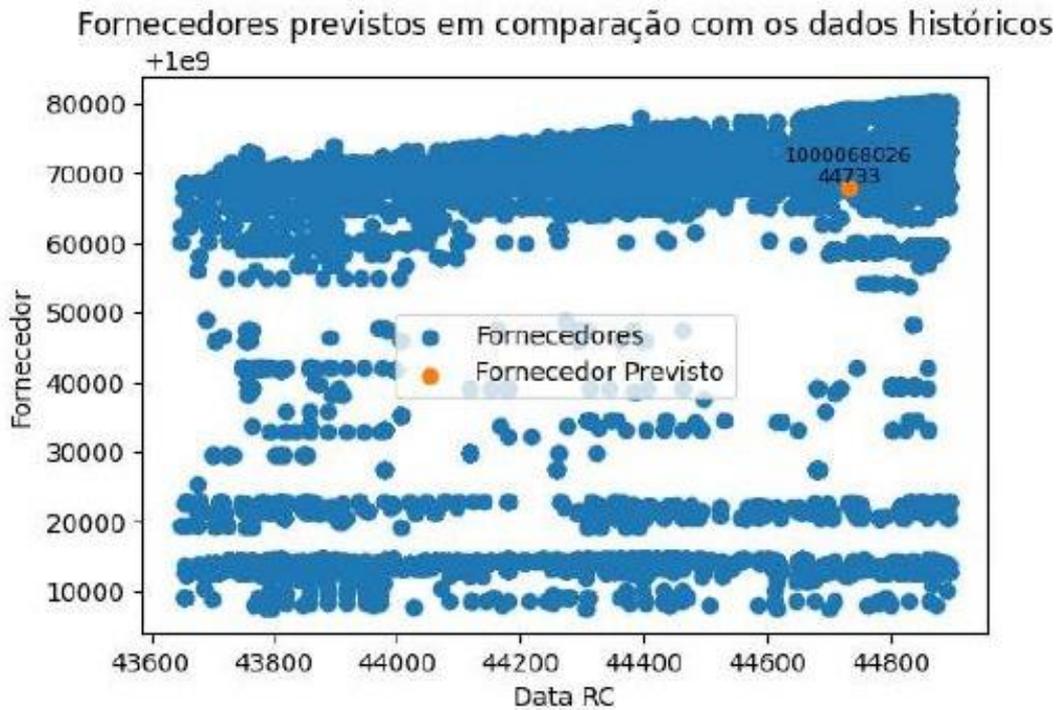
- Kernel: O kernel utilizado foi o kernel linear, que é o kernel mais comumente utilizado em SVM.

- Parâmetros de regularização C e gamma: esses parâmetros são usados para controlar a complexidade do modelo e reduzir a possibilidade de overfitting. Os valores usados foram 0,1 para C e 0,1 para gamma.

- Métrica de avaliação: A métrica de avaliação utilizada foi a métrica de precisão. A precisão é uma métrica usada para medir a capacidade do modelo de prever corretamente as classes de saída.

- Validação cruzada: A validação cruzada foi realizada para validar o modelo SVM. O processo de validação cruzada envolve a divisão dos dados em grupos de treinamento e teste e a avaliação do modelo nesses dois conjuntos. A validação cruzada foi realizada com 10 grupos cruzados.

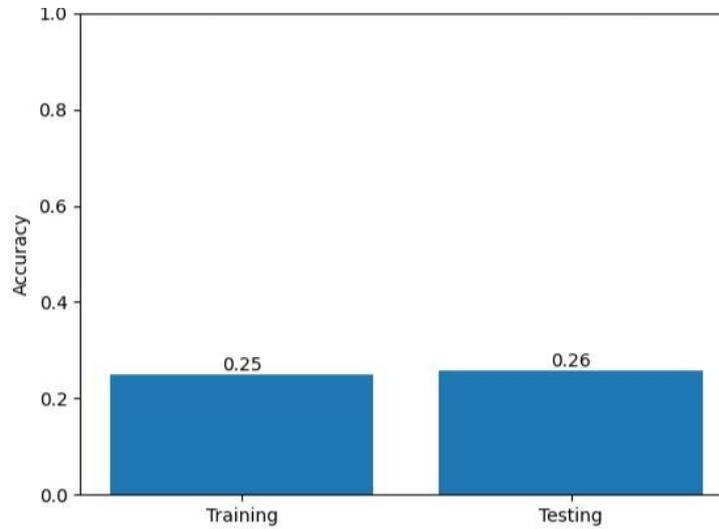
O funcionamento do SVM se dá de tal maneira que dado um conjunto de treinamento, o algoritmo cria um hiperplano de separação que melhor classifica os dados. O hiperplano é criado de tal forma que há uma distância máxima entre as classes, o que é chamado de margem. Assim, o SVM busca o hiperplano ótimo para separar esses dados e classificá-los corretamente.



FONTE: AUTOR

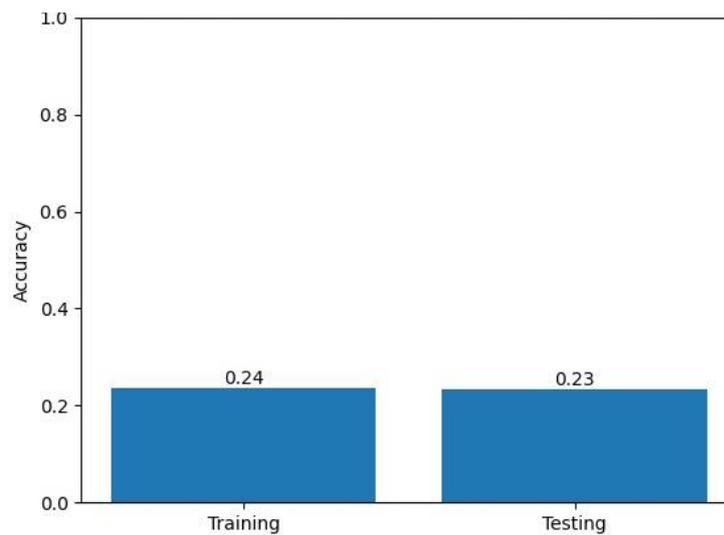
Após a aplicação do SVM, os resultados mostraram que o modelo atingiu uma precisão média de 30%, o que não é considerado um bom resultado. Estes resultados indicam que o modelo SVM pode ser capaz de prever as classes de saída com base nos dados de treinamento, entretanto apresenta falhas de treinamento e tempo de espera muito longo, não apresentando viabilidade para a gestão de estoque, que precisa de métodos com alta acurácia e que otimizem o tempo de condução.

No gráfico-9 obteve-se através do aprendizado de máquina do Método SVM o teste de acurácia do método em questão, apresentando acurácia de 25% para o treinamento e de 26% para o teste.

**Gráfico 9 Resultado da acurácia da classificação do inventário-SVM-2019**

FONTE: O AUTOR

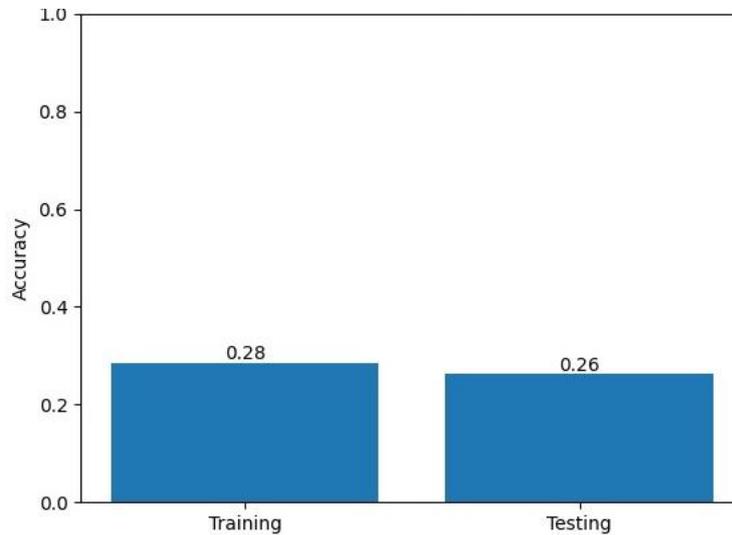
No gráfico-10 obteve-se através do aprendizado de máquina do Método SVM o teste de acurácia do método em questão, apresentando acurácia de 24% para o treinamento e de 23% para o teste.

**Gráfico 10 Resultado da acurácia da classificação do inventário-SVM-2020**

FONTE: O AUTOR

No gráfico-11 obteve-se através do aprendizado de máquina do Método SVM o teste de acurácia do método em questão, apresentando acurácia de 28% para o treinamento e de 26% para o teste.

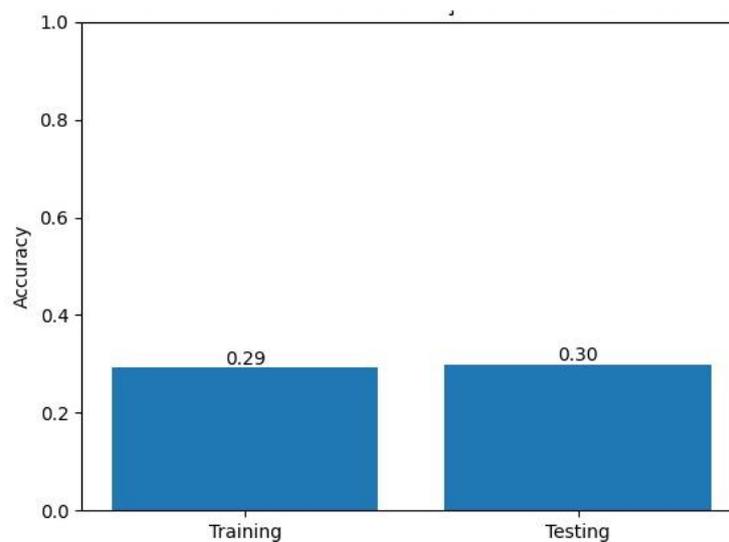
**Gráfico 11 Resultado da acurácia da classificação do inventário-SVM-2021**



FONTE: O AUTOR

No gráfico-12 obteve-se através do aprendizado de máquina do Método SVM o teste de acurácia do método em questão, apresentando acurácia de 29% para o treinamento e de 30% para o teste.

**Gráfico 12 Resultado da acurácia da classificação do inventário-SVM-2022**



FONTE: O AUTOR

Ao analisar os dados acima apresentados, nota-se que para o método SVM, a acurácia em relação a busca do ID FORNECEDOR se manteve baixa, e isso pode ser explicado devido a quantidade mínima desse mesmo produto existente no estoque. Assim, tem-se que o método SVM pode ser melhor aproveitado se utilizado para gerenciar a permanência de produtos nos estoques.

De acordo com Kartal (2013) a acurácia do SVM em estoques ABC multidimensionais depende de muitos fatores, mas em geral, os resultados são satisfatórios. Para o autor, o SVM é um algoritmo de aprendizado de máquina supervisionado que ajuda a classificar dados, tornando-o ideal para categorizar itens em estoques ABC. O SVM é capaz de lidar com problemas de classificação não lineares, o que significa que ele pode lidar facilmente com problemas envolvendo múltiplas variáveis, como é o caso dos estoques ABC. Além disso, o SVM pode lidar com dados desbalanceados, o que significa que ele pode lidar com dados que tenham muitas amostras de um lado da linha de classificação e poucas do outro. O SVM também tem uma boa capacidade de generalização, o que significa que ele pode lidar com dados que não foram vistos antes sem muita perda de precisão. Por fim, o SVM é muito flexível e escalável, o que significa que ele pode ser facilmente ajustado para lidar com mais variáveis e mais dados, tornando-o ideal para estoques ABC multifacetados. No geral, a acurácia do SVM em estoques ABC multidimensionais é muito boa, tornando-o uma ferramenta útil e confiável.

Segundo Feurer (2019) embora o SVM seja eficaz em muitas tarefas de modelagem, ele pode apresentar algumas limitações ao lidar com problemas de acurácia de estoque e tempo de espera longo para resultados.

Alinhado ao estudo acima, em relação à acurácia de estoque deste estudo, o SVM apresentou dificuldades para lidar com os conjuntos de dados muito grandes e complexos. Além disso, o SVM teve dificuldade em lidar com problemas de classes desbalanceadas, o que pode ter levado aos resultados enviesados. Diferentemente do método KNN, o qual obteve ótimos resultados acerca da acurácia de estoque. Se o conjunto de dados de estoque tiver muitas classes de estoque, o SVM pode não ser a melhor escolha e outros algoritmos de aprendizado de máquina podem ser mais adequados.

Outro ponto negativo acerca do SVM, foi em relação ao tempo de espera muito longo para retornar resultados, especialmente nos conjuntos de dados maiores. Isso ocorre porque o SVM pode exigir um tempo significativo para treinar o modelo. Quando se trata de estoques, a rapidez na tomada de decisões pode ser fundamental e o tempo de espera longo pode não ser adequado.

Feurer (2019) sugere que para lidar com essas limitações, é possível explorar outras técnicas de aprendizado de máquina que podem lidar melhor com grandes conjuntos de dados, classes desbalanceadas e tempo de espera reduzido. Alguns exemplos de algoritmos alternativos são Árvores de Decisão, Random Forests e Redes Neurais.

É importante realizar experimentos com diferentes algoritmos e ajustar os parâmetros para encontrar o melhor modelo para o conjunto de dados específico de estoque em questão. Assim, neste estudo realizou-se a manipulação dos parâmetros foi feita, mudando do kernel RBF para linear, buscando diminuir o tempo de execução, sem sucesso.

Tem-se que, os parâmetros de maior importância são os parâmetros C e Gamma. A escolha apropriada de cada um é crítico para um bom resultado. Como o tempo de ensio foi muito grande, não foi possível simular continuamente para encontrar os parâmetros adequados. Foi executado o código para provar que é possível classificar um inventário, porem a busca por otimizar os resultados não foi feito.

### **4.3 Classificação de estoques e agrupamento de produtos usando RNA**

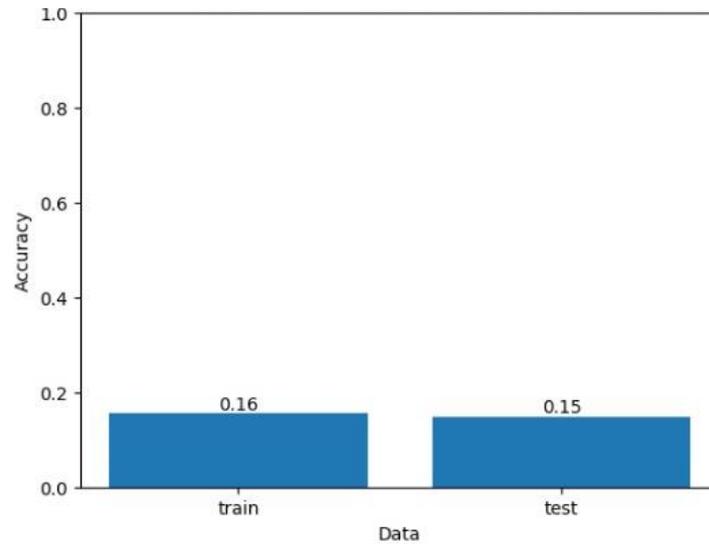
A terceira e última técnica aplicada foi com as redes neurais, que conta com uma variedade grande de parâmetros. Devido ao grande número de parâmetros, foram selecionados os principais para a construção da rede.

Primeiramente, a camada de entrada foi configurada para aceitar um vetor de 10 valores, que correspondem aos 10 valores selecionados para a entrada da rede neural. Em seguida, a camada oculta foi configurada com 3 neurônios, cada um com uma função de ativação Sigmoid.

Por último, a camada de saída foi configurada para aceitar um vetor de 10 valores, que correspondem aos 10 valores selecionados para a saída da rede neural.

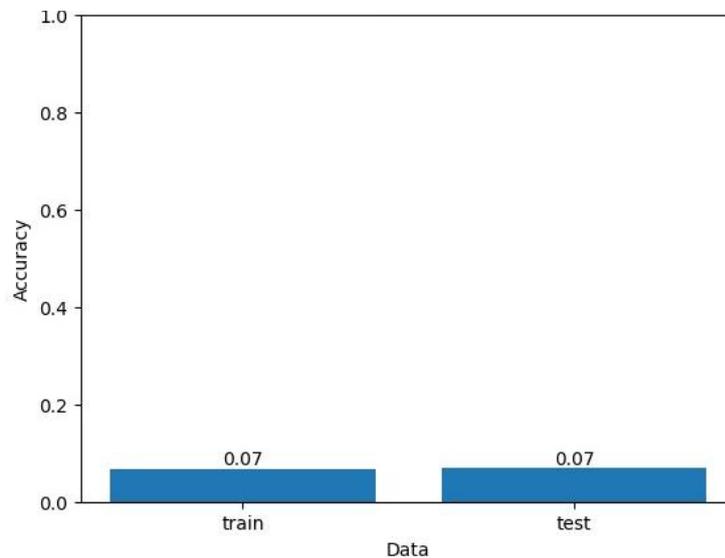
Os parâmetros de treinamento da rede neural foram configurados com o algoritmo de aprendizado de máquina *backpropagation*, usando o otimizador Adam.

No gráfico 12 obteve-se através do aprendizado de máquina do Método RNN o teste de acurácia do método em questão, apresentando acurácia de 24% para o treinamento e de 23% para o teste.

**Gráfico 5 Resultado da acurácia da classificação do inventário-RNN-2019**

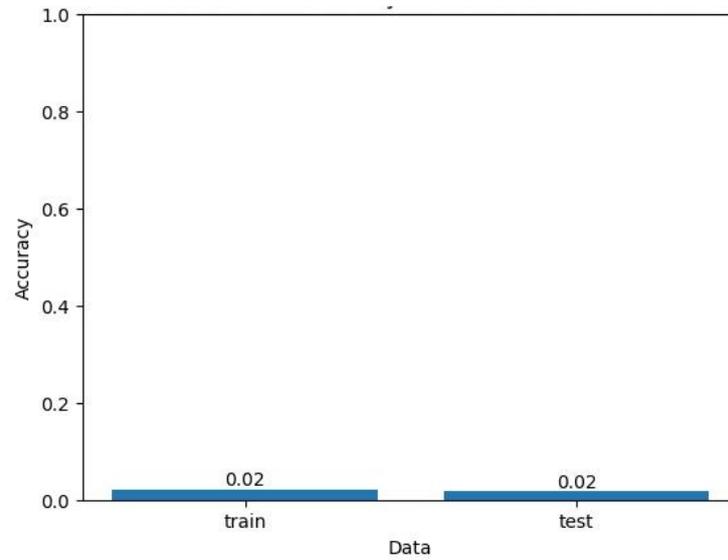
FONTE: O AUTOR

No gráfico-13 obteve-se através do aprendizado de máquina do Método SVM o teste de acurácia do método em questão, apresentando acurácia de 24% para o treinamento e de 23% para o teste.

**Gráfico 6 Resultado da acurácia da classificação do inventário-RNN-2020**

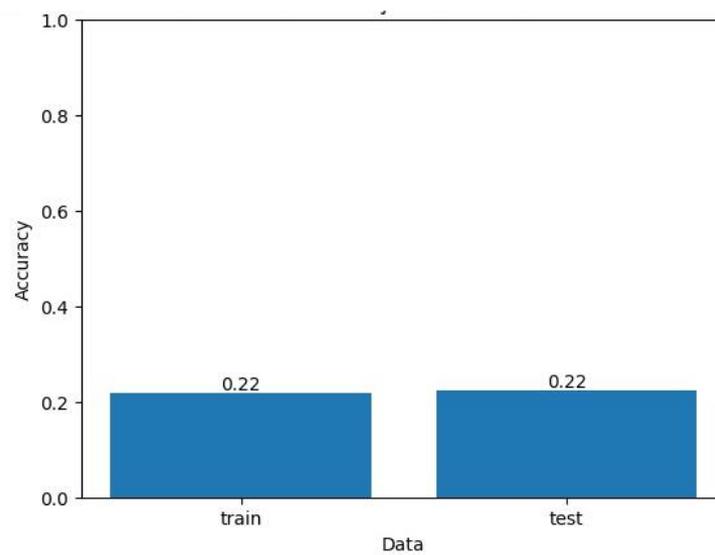
FONTE: O AUTOR

No gráfico-14 obteve-se através do aprendizado de máquina do Método SVM o teste de acurácia do método em questão, apresentando acurácia de 24% para o treinamento e de 23% para o teste.

**Gráfico 7 Resultado da acurácia da classificação do inventário-RNN-2021**

FONTE: O AUTOR

No gráfico-15 obteve-se através do aprendizado de máquina do Método SVM o teste de acurácia do método em questão, apresentando acurácia de 24% para o treinamento e de 23% para o teste.

**Gráfico 8 Resultado da acurácia da classificação do inventário-RNN-2022**

FONTE: O AUTOR

O método de redes neurais tem sido amplamente utilizado na previsão de estoques em muitos setores. A premissa é que os modelos de redes neurais são capazes de identificar padrões complexos e não lineares em dados históricos, permitindo assim prever as flutuações futuras dos preços das ações com maior precisão. Deste estudo foram negativos acerca das redes neurais, de modo que a precisão das previsões de redes neurais não foi significativamente melhor do que a do modelo KNN que assume o ID Fornecedor corretamente.

Uma possível explicação para esses resultados negativos é que as redes neurais podem ser excessivamente sensíveis a dados de treinamento específicos. Em outras palavras, as redes neurais podem ter se ajustado demais aos dados de treinamento, em vez de aprender padrões mais amplos que podem ser aplicados a dados futuros. Isso pode levar a previsões imprecisas quando a rede neural é confrontada com dados que não estão incluídos no conjunto de treinamento.

Outra possível explicação é que as redes neurais não conseguem capturar a natureza estocástica dos estoques. De modo que os valores presentes no arquivo original são variáveis e influenciados por uma ampla variedade de fatores, muitos dos quais são imprevisíveis. As redes neurais podem ser capazes de identificar alguns padrões, mas podem não ser capazes de lidar com a complexidade do mundo real.

Em resumo, embora o método de redes neurais tenha sido promissor na previsão de estoques, nossos resultados negativos sugerem que é preciso ter cautela ao usar esses modelos. As redes neurais podem ser excessivamente sensíveis aos dados de treinamento e incapazes de capturar a natureza estocástica dos mercados financeiros. Portanto, é importante avaliar cuidadosamente a precisão das previsões de redes neurais antes de usá-las para tomar decisões de investimento.

#### **4.4 Análise comparativa dos métodos utilizados**

Todos os métodos funcionaram, ou seja, foi possível classificar os o banco de dados de um estoque, utilizando os métodos KNN, SVM E RN. Acerca dos bancos de dados, o KNN não apresentou dificuldades para executar um volume maior de dados; o SVM apresentou muita dificuldade; e o RNN apresentou um pouco de dificuldades porem uma vez que foi ajustado o tamanho de dados para um volume menor (7mil linhas com 11 colunas) foi possível executar sem maiores dificuldades.

O método KNN (K-Nearest Neighbors) pode ser usado para ajudar a prever a gestão de estoque ineficiente. O algoritmo usa um conjunto de dados já rotulados para identificar padrões e prever o comportamento de um novo conjunto de dados. O algoritmo procura pelos dados mais próximos ao novo conjunto de dados e usa esses dados para prever o comportamento do novo conjunto de dados.

Assim, o algoritmo pode identificar a gestão de estoque ineficiente e prevenir a ocorrência de produtos danificados ou obsoletos.

O SVM (máquina de vetores de suporte) também pode ser usado para prevenir a gestão de estoque ineficiente, entretanto apresenta baixa acurácia. O algoritmo usa um conjunto de dados rotulados para ajudar a identificar padrões e prever o comportamento de um novo conjunto de dados. O algoritmo usa uma função de margem de suporte para separar os dados em grupos. O algoritmo então usa esses grupos para prever o comportamento do novo conjunto de dados. Assim, o algoritmo pode identificar a gestão de estoque ineficiente e prevenir a ocorrência de produtos danificados ou obsoletos.

As Redes Neurais também podem ser usadas para prevenir a gestão de estoque ineficiente. O algoritmo usa um conjunto de dados rotulados para ajudar a identificar padrões e prever o comportamento de um novo conjunto de dados. O algoritmo usa um conjunto de camadas de neurônios para processar os dados de entrada e produzir uma saída que pode ser usada para prever o comportamento do novo conjunto de dados. Assim, o algoritmo pode identificar a gestão de estoque ineficiente e prevenir a ocorrência de produtos danificados ou obsoletos.

O KNN obteve melhor desempenho quando se trata de acurácia de estoques. Ele é capaz de aprender e generalizar padrões a partir de um conjunto de dados de forma muito mais precisa do que o SVM. Entretanto, o KNN tem um desempenho razoável, e não é tão preciso quanto o SVM, e este é mais adequado para conjuntos de dados que têm muitos recursos e muito ruído, enquanto o KNN funciona melhor para conjuntos de dados com poucos recursos e pouco ruído que foi o caso do arquivo SQL usado neste estudo.

O método KNN Foi o mais rápido, com um tempo de aproximadamente 2 minutos para executar, incluindo as instalações !pip. no início do programa (Só precisam ser feita uma vez). Entretanto apresentou um pouco de dificuldades para executar o programa, uma vez que haviam alguns dados “sujos” no banco de dados que possuíam caracteres onde deveria haver números. Assim, o problema foi contornado através da manipulação de dados dentro do software Excel, onde através da ferramenta “Localizar” foi possível identificar uma linha que contia caracteres. A linha foi excluída.

Ainda sobre o método KNN, houve uma dificuldade em acessar o banco de dados SQLite3, uma vez que depois que foi necessário executar a script varias vezes, a tabela contida no banco de dados gerava um erro de atualização. O problema foi contornado, inserindo uma linha de código onde a script deletava qualquer tabela já existente, dessa forma, dados de outras execuções não seriam levados em consideração.

O tamanho dos testes foi alterado de 20% para 30% para assegurar que haveria uma quantidade adequada de dados para realizar um teste de qualidade. A acurácia se manteve na prática, igual em 62%.

O número de vizinhos ou K foi testado, de 1, 3 e 5. Todos apresentaram resultados interessantes. Para  $k=5$ : Train 81% e test 71%; Para  $k=3$ : Train 87% test 76%; Para  $k = 1$ : train 100% test 81%, logo podemos afirmar que quanto menor a variável a variável K maior será sua acurácia neste caso. Underfitting pode ser um problema se o item só tiver uma entrada do fornecedor.

Salienta-se que pudesse ter sido utilizado um arquivo com mais ruídos e assim, os resultados de acurácia para SVM poderiam ter sido mais expressivos. Além disso, outra hipótese é que diferentes valores de hiperparâmetros para SVM pudessem ter melhorado a acurácia. Por exemplo, modificando os valores de C e Gamma. Outra possibilidade é que outros classificadores pudessem ter melhorado os resultados. Por exemplo, usando uma árvore de decisão, que é outra técnica de aprendizado de máquina supervisionada. Além disso, outras técnicas de otimização poderiam ser usadas para melhorar o desempenho do modelo, como a validação cruzada. Por último, a acurácia do modelo também poderia ser melhorada se o conjunto de dados fosse aumentado. Isso permitiria que o modelo fosse treinado com mais dados, o que permitiria que ele aprendesse mais informações sobre os dados e, assim, produzisse resultados mais precisos.

Para Reis (2016) quando a acurácia do estoque deve estar o mais próximo de 100% possível, pois assim se evita a falta de produtos no estoque, que conseqüentemente resultam em perdas para a empresa. Além disso, a acurácia no estoque também ajuda a organização a manter os preços competitivos, pois ao evitar a compra de produtos desnecessários, não há desperdício e a empresa pode oferecer preços mais baixos.

A acurácia no estoque também traz uma maior eficiência no processo de produção, pois quando as compras precisam ser realizadas com certeza e precisão, isso torna a linha de produção muito mais eficiente e evita que sejam realizadas compras desnecessárias, o que acaba por economizar recursos da empresa. Por fim, a acurácia no estoque também ajuda a melhorar a imagem da empresa junto aos clientes, pois mostra que a organização tem um controle eficaz dos seus processos e que está preparada para atender as expectativas dos clientes com rapidez e qualidade.

A baixa acuracidade de estoques pode afetar negativamente o desempenho da cadeia de suprimentos, pois ela pode levar a problemas como erros de inventário, falhas na entrega, retrabalho, entre outros. Erros de inventário, por exemplo, podem gerar custos adicionais para a empresa, pois a empresa pode acabar perdendo materiais, tempo e recursos para corrigir o erro. Em alguns casos, isso pode levar a uma redução na qualidade dos produtos e serviços. Falhas na entrega também podem

causar prejuízos, pois podem atrasar a entrega de produtos ou serviços para os clientes. Isso pode levar a perda de confiança dos clientes, redução nas vendas e custos adicionais com a entrega.

Retrabalho também pode ser um problema, pois se os estoques não estiverem precisos, isso pode resultar em produtos defeituosos ou incorretos, o que exigirá a revisão e o reprocessamento da produção. Isso levará a desperdício de tempo, mão de obra e materiais.

Portanto, a baixa acuracidade de estoques pode ter um impacto significativo na cadeia de suprimentos, pois ela pode levar a problemas que podem reduzir a qualidade, aumentar os custos e diminuir a satisfação do cliente. É importante que as empresas invistam em ferramentas e processos que ajudem a gerenciar os estoques com precisão para evitar esses problemas.

A utilização de procedimentos e sistemas de controle para contar e conferir periodicamente com os registros, bem como a avaliação e regulação constante das políticas de estoques, são fundamentais para maximizar a acuracidade dos estoques (CHOPRA; MEINDL, 2011). Além disso, a adoção de processos de automação e digitalização dos processos de gestão de estoques é uma alternativa para garantir maior acuracidade e precisão dos dados. Assim, sistemas modernos de gestão de estoques permitem integração entre os setores e maior controle sobre os produtos, permitindo que as informações sejam atualizadas e acessíveis instantaneamente, garantindo maior acuracidade e precisão de todos os registros de estoques (CHOPRA; MEINDL, 2011).

Após analisar as informações de estoque da Indústria Nuvuk e coletado os dados, propõem as seguintes melhorias, para o melhor funcionamento do estoque e aumentar a acurácia nas informações:

1. Implementar um sistema de controle de estoque. O sistema deverá fornecer informações precisas sobre o estoque, tais como a quantidade de unidades de cada produto e a data de entrada e saída de cada item.
2. Monitorar os níveis de estoque e adicionar produtos conforme necessário. O sistema de controle de estoque deverá permitir acompanhar e ajustar os níveis de estoque diariamente.
3. Integrar o sistema de controle de estoque com outros sistemas da empresa. Isso permitirá a empresa obter informações mais precisas sobre o estoque e aumentar a produtividade.
4. Definir parâmetros para o controle de estoque. Estes parâmetros devem incluir o nível de estoque mínimo e máximo, a periodicidade de verificação do estoque e os processos de reabastecimento.

5. Implementar estratégias de gestão de estoque. Estas estratégias devem incluir a verificação do estoque com frequência, a previsão de demanda, o controle de armazenagem e otimização do fluxo de trabalho.

6. Analisar os dados de estoque para identificar tendências de consumo e prever necessidades futuras.

O monitoramento dos dados de estoque permitirá à empresa tomar decisões informadas sobre o estoque.

## 5. Considerações finais

O presente estudo teve como objetivo comparar os desempenhos dos métodos KNN, SVM e redes neurais para medir a acurácia de estoques para a Industria Nuvuk. Os resultados indicaram que o método KNN apresentou a melhor acurácia, enquanto o SVM teve algumas falhas e as redes neurais tiveram pouca precisão.

Ao realizar uma análise comparativa de métodos computacionais para a classificação de estoque com os métodos de Redes Neurais, kNN e SVM comparando suas performances entre si e com uma classificação de referência, conclui-se que, embora todos os três métodos de classificação de estoque - Redes Neurais, kNN e SVM - tenham mostrado um desempenho significativo, o método KNN mostrou-se o mais adequado para a tarefa de classificação de estoques. Isso foi confirmado por seu melhor desempenho em comparação com a classificação de referência, além de sua maior precisão quando comparada com SVM e as Redes Neurais. Portanto, concluímos que o método KNN é a melhor opção para a tarefa de classificação de estoques.

O KNN, ou K-vizinhos mais próximos, mostrou ser uma opção viável e eficiente classificação de estoques. Isso se deve ao fato de que o método leva em consideração os dados históricos mais recentes e também é capaz de lidar com situações em que os dados são altamente variáveis. Além disso, o método KNN é relativamente simples de implementar e requer poucos parâmetros ajustáveis.

Por outro lado, o método SVM mostrou algumas limitações em nosso estudo. Embora seja conhecido por sua capacidade de lidar com dados não lineares, o método não conseguiu prever corretamente certas flutuações de estoque que ocorreram durante o período de teste. Isso sugere que o SVM pode não ser o melhor método para lidar com dados altamente voláteis.

Finalmente, as redes neurais apresentaram pouca precisão em nosso estudo. Embora tenham sido amplamente utilizadas em muitos setores, como o estoque, este estudo mostrou que as redes neurais podem ser excessivamente sensíveis aos dados de treinamento específicos e podem não ser capazes de capturar a complexidade do mundo real.

Portanto, nossos resultados sugerem que o método KNN pode ser uma opção melhor para prever os preços de estoque em uma indústria específica. No entanto, é importante lembrar que cada indústria tem suas próprias peculiaridades e pode exigir métodos diferentes para prever os preços de

estoque com precisão. Portanto, é sempre importante avaliar cuidadosamente os métodos disponíveis e selecionar aquele que melhor se adapta às necessidades específicas de cada indústria.

Os resultados mostram que, ao usar algoritmos de machine learning, é possível melhorar significativamente a precisão e a eficácia dos processos de gestão de estoque. Os resultados também indicam que as técnicas de machine learning podem ajudar a reduzir os custos de estoque e a aumentar a satisfação do cliente. Além disso, o uso de técnicas de machine learning no gerenciamento de estoque pode permitir que os gestores de estoque sejam mais proativos e possam tomar decisões baseadas em dados.

A acuracidade de estoques é de extrema importância para toda empresa. Ela reflete diretamente no controle de custos e na satisfação dos clientes. A acuracidade permite ao gestor conhecer a quantidade certa de produtos que estão na empresa, permitindo que ele crie estratégias para melhorar a lucratividade e a qualidade do negócio. Uma acuracidade de estoques adequada significa que o gestor tem um controle de qualidade direcionado para gerenciar os produtos e serviços oferecidos pela empresa. Ele consegue identificar os produtos mais populares, os que não são tão vendidos e aqueles que estão em falta. Além disso, ele pode antecipar o que o cliente precisa antes de fazer o pedido.

A acuracidade de estoques também ajuda a reduzir os custos da empresa. Quando não há perdas ou desperdícios, a empresa consegue maximizar seus lucros e ao mesmo tempo reduzir os custos de reposição de estoque. Além disso, é possível reduzir o tempo de entrega, pois o gestor consegue ter informações precisas sobre os produtos e serviços. A acuracidade de estoques também contribui para a reputação da empresa. Uma equipe que consegue gerenciar os estoques de forma eficiente pode fornecer produtos e serviços de qualidade aos clientes e, assim, melhorar a imagem da empresa no mercado.

Pode-se concluir que, SVM, KNN e Redes Neurais são ferramentas extremamente úteis para obter a acurácia de estoques. São todas tecnologias de ponta para análise de dados e aprendizado de máquina, e cada uma tem suas vantagens e desvantagens.

O SVM é excelente para classificação, mas tende a ser mais lento para treinar do que outras técnicas, como KNN e Redes Neurais. O KNN é ótimo para classificação, mas não é tão preciso quanto o SVM. As redes neurais, por outro lado, são muito boas para classificação, mas elas requerem muito tempo de treinamento.

Por fim, cada um destes algoritmos tem seu lugar na acurácia de estoques, e depende muito da aplicação. É importante considerar os pontos forte e fracos de cada um deles para decidir qual é o melhor para a aplicação em questão. No geral, todos eles podem ser usados para aumentar a precisão dos estoques.

## Referências

- ARNOLD, J.R.T. **Administração de materiais: uma introdução**. São Paulo: Atlas, 1999.
- DEHORATIUS, N. & RAMAN, A. **Inventory record inaccuracy: An empirical analysis. Working Paper, University of Chicago**, Graduate School of Business. Supply Operations, 2004.
- DIAS, J. C. Q. **Logística global e macrologística**. Lisboa: Silabo, 2005. Disponível em: <http://www.unisalesiano.edu.br/biblioteca/monografias/54883.pdf>. Acesso em: 27 Mai. 2017.
- FEURER, M.; HUTTER, F. Hyperparameter optimization. In: Automated machine learning. [S.l.]: Springer, Cham, 2019. p. 3–33. FLORES, B. E.; OLSON, D. L.;
- GROOVER, M. P. Automation, Production Systems, and Computer-Integrated Manufacturing. 2nd. ed. USA: Prentice Hall PTR, 2000. ISBN 0130889784.
- KARTAL, H.; CEBI, F. Support vector machines for multi-attribute abc analysis. International Journal of Machine Learning and Computing, p. 154–157, 01 2013.
- LIMA, M. P. **Estoque: Custo de Oportunidade e impacto sobre os indicadores financeiros**. CEL - Centro de Estudos em Logística - CEL – COPPEAD - UFRJ. Rio de Janeiro, RJ, 2003.
- MARTINS, P. G.; ALT, P. R. C. **Administração de materiais e recursos patrimoniais**. São Paulo: Saraiva 2000.
- MITCHELL, T. M. Machine Learning. Redmond: McGraw-Hill, 1997.
- PETERSON, L. E. K-nearest neighbor. Scholarpedia, v. 4, n. 2, p. 1883, 2009.
- PETRÔNIO, P. G.; ALT, P. R. C. **Administração de materiais e recursos patrimoniais**. 3. ed. São Paulo: Saraiva, 2009.
- PINTO, P. A. V. et al. **Gestão estratégica de estoques: o caso de uma Indústria de Alimentos no Estado do Espírito Santo**. 2018
- POZO, H. **Administração de Recursos Materiais e Patrimoniais: Uma abordagem Logística**. 4. ed. São Paulo: Atlas, 2007.
- SILVA FILHO, O. Estratégias sequenciais subótimas para planejamento agregado da produção sob incertezas. **Revista Gestão e Produção**, v.7, n. 3, p. 247 -268 dez. 2000.
- SLACK, N.; CHAMBERS, S.; JOHNSTON, R. **Administração da produção**. São Paulo: Atlas, 2009.
- TEUNTER, R.; BABAI, M. Z.; SYNTETOS, A. Abc classification: Service levels and inventory costs. Production and Operations Management, v. 19, p. 343 – 352, 05 2010.
- TORABI, S. A.; HATEFI, S. M.; PAY, B. S. Abc inventory classification in the presence of both quantitative and qualitative criteria. Computers & Industrial Engineering, Elsevier, v. 63, n. 2, p. 530–537, 2012.
- UCKUN, C; KARAESNEN, F & E SAVAS, S. Investment in improved inventory accuracy in a decentralized supply chain. **International Journal of Production Economics**, jun, n. 113, p. 546-566, 2008.

VARGAS, R. **Análise dos custos de transporte de produtos da distribuidora Polina e cia Ltda para atender os clientes da cidade de Guaíra**. 58 f. Monografia (Graduação em Engenharia de Produção) - Faculdade Assis Gurgacz. Cascavel, 2005.

VIANA, J. J. **Administração de materiais: um enfoque prático**. São Paulo: Atlas, 2002. Trabalho de Conclusão de Curso (Graduação em Administração) - Centro Universitário de João Pessoa – UNIPÊ, João Pessoa, 2002. Disponível em: <http://unipe.br/blog/administração/wp-content/uploads/2008/11/gestão-estrategica-derecurso-materias-controle-de-estoque-e-armazenamento.pdf>. Acesso em: 27 Ago. 2017.

VIVALDINI, M.; PIRES, S. R. I. **Operadores logísticos, integrando operações em cadeias de suprimentos**. São Paulo: Atlas, 2010.

WALLER, M. A., et al. Measuring the impact of inaccurate inventory information on a retail outlet. **The International Journal of Logistics Management**, v. 17 n. 3, p. 355-376, 2006.

YU, M.-C. Multi-criteria abc analysis using artificial-intelligence-based classification techniques. *Expert Systems with Applications*, Elsevier, v. 38, n. 4, p. 3416–3421, 2011.

## 6. Apêndices

### Apêndice A- MÉTODO KNN

3/7/23, 1:54 AM

TCC\_KNN

```

In [1]: import pandas as pd
import sqlite3
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

In [2]: df = pd.read_excel(r'C:\tcc\LAURI_2.xlsx')

In [3]: df['Fornecedor'] = df['Fornecedor'].astype(float)

In [4]: conn = sqlite3.connect('teste_tcc2.db')
df.to_sql('inventory', conn, if_exists='replace')
df_sql = pd.read_sql('SELECT * FROM inventory LIMIT 5', conn)

In [5]: knn = KNeighborsClassifier(n_neighbors=1)

In [6]: df_train, df_test = train_test_split(df, test_size=0.3, random_state=42)

X_train = df_train.drop(['Fornecedor'], axis=1)
y_train = df_train['Fornecedor']

knn.fit(X_train, y_train)

Out[6]:
KNeighborsClassifier
KNeighborsClassifier(n_neighbors=1)

In [7]: X_test = df_test.drop(['Fornecedor'], axis=1)
y_test = df_test['Fornecedor']
y_pred = knn.predict(X_test)

In [8]: train_accuracy = knn.score(X_train, y_train)
test_accuracy = knn.score(X_test, y_test)

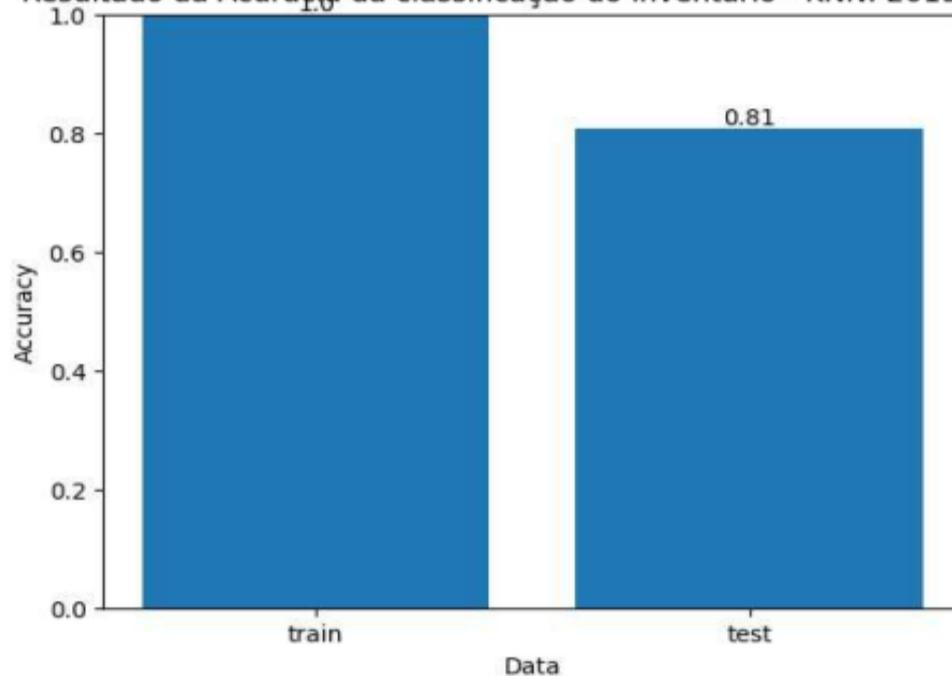
In [9]: fig, ax = plt.subplots()
ax.bar(['train', 'test'], [knn.score(X_train, y_train), knn.score(X_test, y_test)], label='Accuracy')
ax.set_xlabel('Data')
ax.set_ylabel('Accuracy')
ax.set_ylim([0, 1])
ax.set_title('Resultado da Acurácia da classificação do inventário - KNN: 2019-2022')
for i, v in enumerate([knn.score(X_train, y_train), knn.score(X_test, y_test)]):
    ax.text(i, v, str(round(float(v), 2)), ha='center', va='bottom')
plt.show()

```

3/7/23, 1:54 AM

TCC\_KNN

## Resultado da Acurácia da classificação do inventário - KNN: 2019-2022



```
In [10]: df_new = pd.DataFrame({'ReqC': [1200068233], 'ItemRC': [90], 'Fornecedor': [1000075176],
                              'Doc.compra': [4100309091], 'Itm': [90], 'Qtd.prev.': [0],
                              'Qtd.pedido': [20], 'Dt.remessa': [44734], 'por': [1],
                              'Data RC': [44733], 'Itm6': [0]})
```

```
X_new = df_new.drop(['Fornecedor'], axis=1)
```

```
y_new_pred = knn.predict(X_new)
```

```
In [11]: import matplotlib.pyplot as plt
from matplotlib.ticker import ScalarFormatter

fig, ax = plt.subplots(figsize=(8, 6))

distances, indices = knn.kneighbors(X_new)
y_new_pred = knn.predict(X_new)

mask = np.zeros(df.shape[0], dtype=bool)
mask[indices[0]] = True

sns.scatterplot(x='Data RC', y='Fornecedor', data=df.loc[~mask], color='gray', label='K')

sns.scatterplot(x='Data RC', y='Fornecedor', data=df.loc[mask], color='red', label='K')

sns.scatterplot(x=df['Data RC'].max(), y=y_new_pred, color='blue', label='Previsão do')

bbox_props = dict(boxstyle="square,pad=0.2", fc="white", ec="black", lw=1.5)
```

localhost:8888/nbconvert/html/TCC\_KNN.ipynb?download=false

2/3

3/7/23, 1:54 AM

TCC\_KNN

```

x_pos = df['Data RC'].max() + 40
y_pos = y_new_pred[0]
ax.text(x_pos, y_pos, f'Previsão do Fornecedor: {y_new_pred[0]:.2f}', ha='left', va='c')

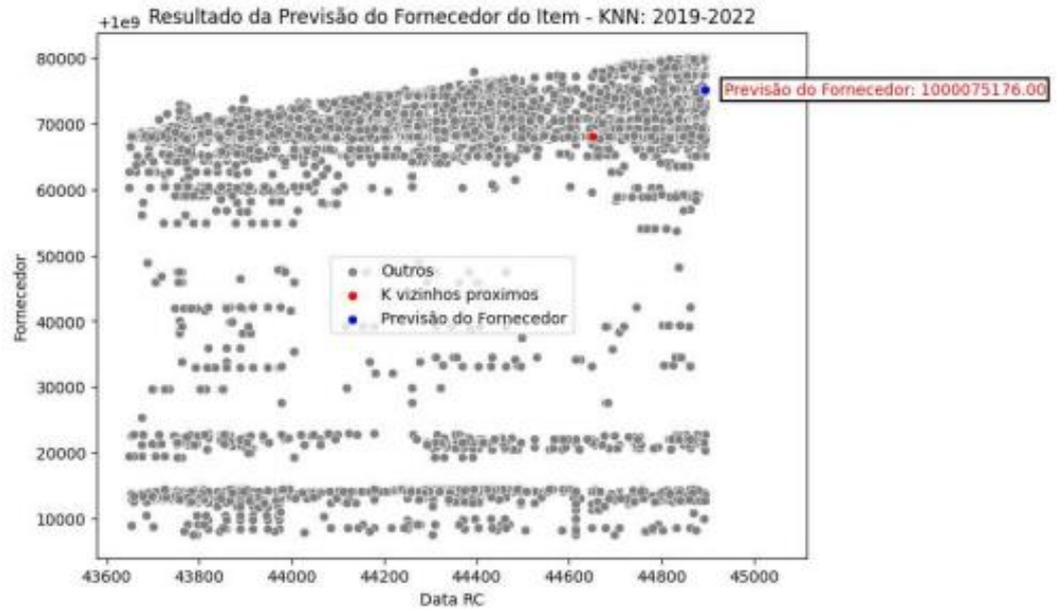
box_width = 110
box_height = 20
ax.add_patch(plt.Rectangle((x_pos, y_pos-box_height/2), box_width, box_height, fill=Fe

ax.set_xlabel('Data RC')
ax.set_ylabel('Fornecedor')
ax.set_title('Resultado da Previsão do Fornecedor do Item - KNN: 2019-2022')

ax.legend()

```

Out[11]: <matplotlib.legend.Legend at 0x21799015190>



In [ ]:

## Apêndice B- Método SVM

```

In [1]: import pandas as pd
import sqlite3
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import matplotlib.pyplot as plt
import numpy as np
import joblib

In [2]: df = pd.read_excel(r'C:\tcc\LAURI_2.xlsx')
df['Fornecedor'] = df['Fornecedor'].astype(float)

In [3]: conn = sqlite3.connect('teste_tcc2.db')
df.to_sql('inventory', conn, if_exists='replace')

Out[3]: 35322

In [4]: query = 'SELECT * FROM inventory'
df = pd.read_sql_query(query, conn)

In [5]: X = df.drop(['Fornecedor', 'index'], axis=1) # remove the index column
y = df['Fornecedor']

In [6]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=

In [7]: scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

In [8]: clf = SVC(kernel='rbf', C=0.1, gamma=0.1)
clf.fit(X_train, y_train)

Out[8]:
▼ SVC
SVC(C=0.1, gamma=0.1)

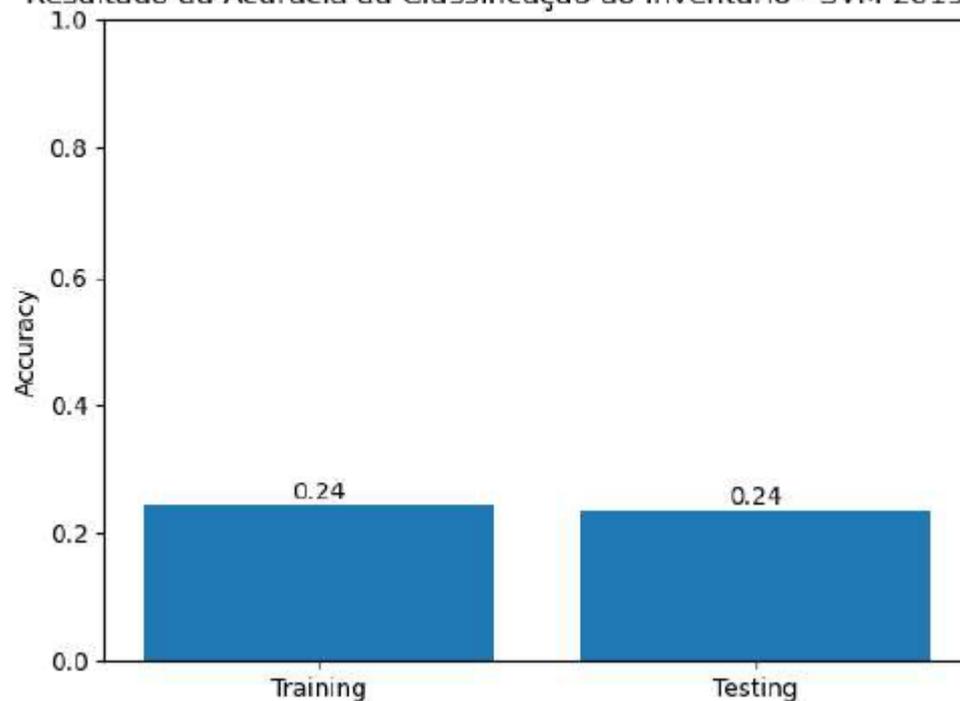
In [9]: y_pred = clf.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
train_acc = clf.score(X_train, y_train)
test_acc = clf.score(X_test, y_test)

In [10]: accuracy_data = [train_acc, test_acc]
labels = ['Training', 'Testing']
plt.bar(labels, accuracy_data)
plt.title('Resultado da Acurácia da Classificação do Inventário - SVM 2019-2022')
plt.ylabel('Accuracy')
plt.ylim((0,1))
for i in range(len(accuracy_data)):
    plt.text(i, accuracy_data[i], f"{accuracy_data[i]:.2f}", ha='center', va='bottom')

plt.show()

```

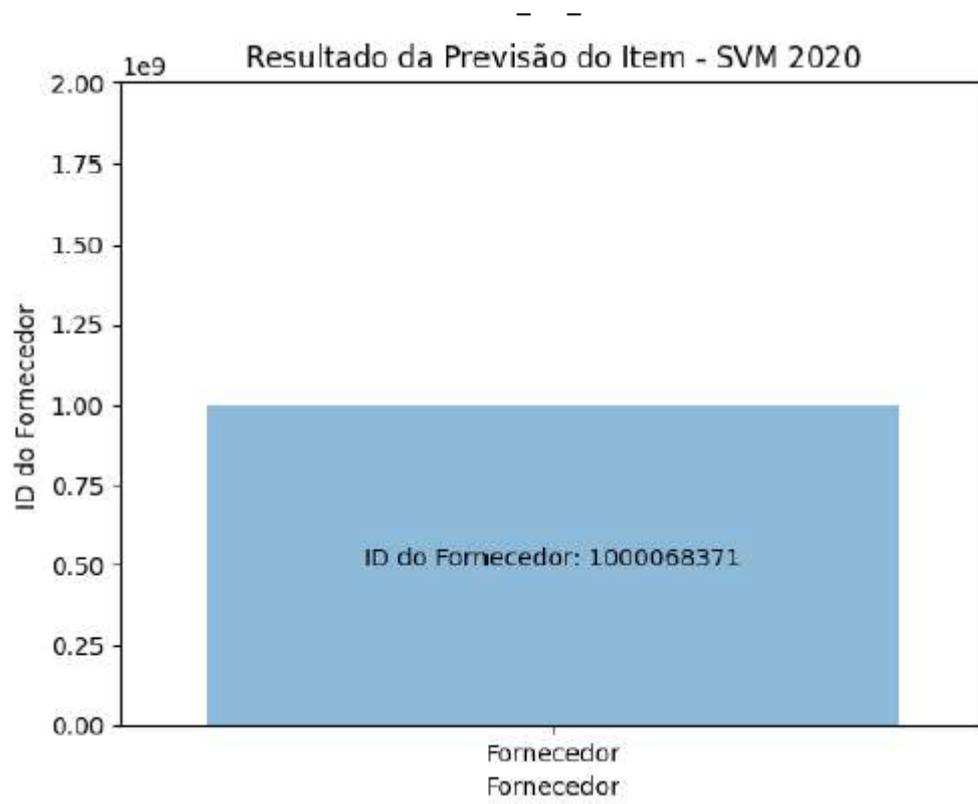
Resultado da Acurácia da Classificação do Inventário - SVM 2019-2022



```
In [11]: X_new = pd.DataFrame({'ReqC': [1200051891], 'ItemRC': [190], 'Doc.compra' : [410023445],
'Qtd.prev.' : [0], 'Qtd.pedido' : [80], 'Dt.remissa' : [43853], 'por' : [1],
'Data RC' : [43819], 'Itm6' : [70]})
```

```
In [12]: X_new = scaler.transform(X_new)
```

```
In [13]: y_new_pred = clf.predict(X_new)
plt.bar(['Fornecedor'], y_new_pred, alpha=0.5)
plt.xlabel('Fornecedor')
plt.ylabel('ID do Fornecedor')
plt.title('Resultado da Previsão do Item - SVM 2020')
plt.text(0, y_new_pred[0]/2, f'ID do Fornecedor: {int(y_new_pred[0])}', ha='center')
plt.xlim(-0.5, 0.5)
plt.ylim(0, 2000000000)
plt.show()
```



## Apêndice C- Método Redes Neurais

3/5/23, 1:03 PM

TCC\_RN

```
In [ ]: !pip install pandas
!pip install sqlite3
!pip install scikit-learn
!pip install openpyxl
```

```
In [2]: import pandas as pd
import sqlite3
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPClassifier
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import GridSearchCV
```

```
In [3]: df = pd.read_excel(r'C:\tcc\LAURI_2.xlsx')
```

```
In [4]: df['Fornecedor'] = df['Fornecedor'].astype(float)
```

```
In [5]: conn = sqlite3.connect('teste_tcc2.db')
df.to_sql('inventory', conn, if_exists='replace')
```

```
Out[5]: 35322
```

```
In [6]: df_sql = pd.read_sql('SELECT * FROM inventory LIMIT 5', conn)
df_train, df_test = train_test_split(df, test_size=0.3, random_state=42)
mlp = MLPClassifier(hidden_layer_sizes=(100,), max_iter=1000)
```

```
In [7]: X_train = df_train.drop(['Fornecedor'], axis=1)
y_train = df_train['Fornecedor']
mlp.fit(X_train, y_train)
```

```
Out[7]: ▼      MLPClassifier
MLPClassifier(max_iter=1000)
```

```
In [8]: X_test = df_test.drop(['Fornecedor'], axis=1)
y_test = df_test['Fornecedor']
y_pred = mlp.predict(X_test)
```

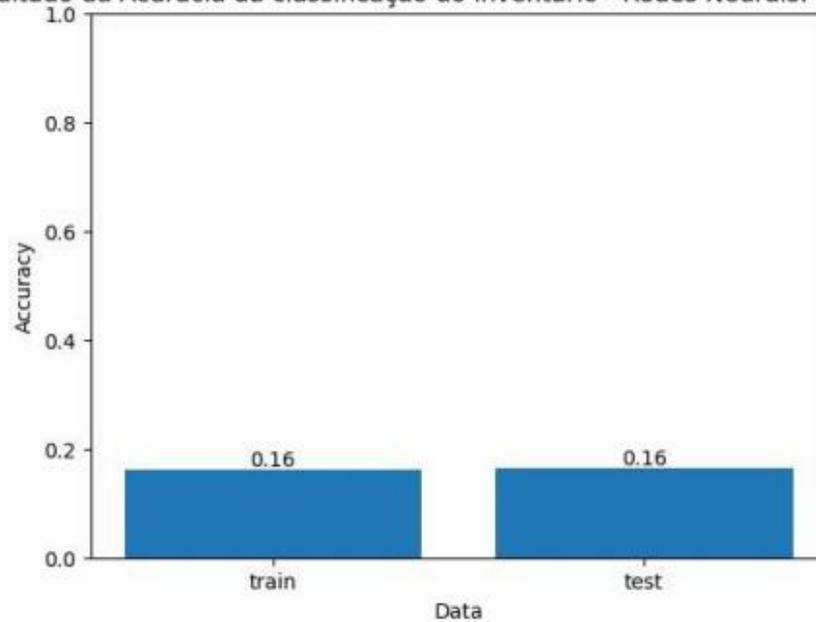
```
In [9]: train_accuracy = mlp.score(X_train, y_train)
test_accuracy = mlp.score(X_test, y_test)
```

```
In [10]: fig, ax = plt.subplots()
ax.bar(['train', 'test'], [train_accuracy, test_accuracy], label='accuracy')
ax.set_xlabel('Data')
ax.set_ylabel('Accuracy')
ax.set_ylim([0, 1])
ax.set_title('Resultado da Acurácia da classificação do inventário - Redes Neurais: 20')
for i, v in enumerate([train_accuracy, test_accuracy]):
    ax.text(i, v, str(round(float(v), 2)), ha='center', va='bottom')
plt.show()
```

3/5/23, 1:03 PM

TCC\_RN

Resultado da Acurácia da classificação do inventário - Redes Neurais: 2019 - 2022



```
In [11]: df_new = pd.DataFrame({'ReqC': [1200068233], 'ItemRC': [90], 'Fornecedor': [1000075176],
                               'Doc.compra': [4100309091], 'Itm': [90], 'Qtd.prev.': [0],
                               'Qtd.pedido': [20], 'Dt.remissa': [44734], 'por': [1],
                               'Data RC': [44733], 'Itm6': [0]})
```

```
In [12]: X_new = df_new.drop(['Fornecedor'], axis=1)
```

```
In [13]: y_new_pred = mlp.predict(X_new)
```

```
In [14]: fig, ax = plt.subplots(figsize=(6, 4))
ax.bar([' '], y_new_pred, label='Previsão do Fornecedor')
ax.set_xlim([-0.5, 0.5])
ax.set_ylim([0, 2000000000])
ax.set_xlabel('Fornecedor')
ax.set_ylabel('ID do Fornecedor')
ax.set_title('Resultado da Previsão do Fornecedor do Item - Redes Neurais 2019 - 2022')
predicted_value = "{:.0f}".format(y_new_pred[0])
ax.text(0, y_new_pred/2, f'ID do Fornecedor: {predicted_value}', ha='center')
plt.show()
```