

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO**

CAINÃ ARAÚJO PAVANI

**APLICAÇÃO DE REDES NEURAIS E FLORESTAS ALEATÓRIAS PARA
PREDIÇÃO DE PRODUTOS PERECÍVEIS**

**VITÓRIA
2022**

CAINÃ ARAÚJO PAVANI

**APLICAÇÃO DE REDES NEURAIS E FLORESTAS ALEATÓRIAS PARA
PREDIÇÃO DE PRODUTOS PERECÍVEIS**

Trabalho de Conclusão de Curso apresentado ao Departamento Do Curso de Engenharia de Produção, do Centro Tecnológico, da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Bacharel em Engenharia de Produção.
Orientador: Prof. Dr.-Ing. Renato Antônio Krohling.

**VITÓRIA
2022**

CAINÃ ARAÚJO PAVANI

**APLICAÇÃO DE REDES NEURAIIS E FLORESTAS ALEATÓRIAS PARA
PREDIÇÃO DE PRODUTOS PERECÍVEIS**

Trabalho de Conclusão de Curso apresentado ao Departamento Do Curso de Engenharia de Produção, do Centro Tecnológico, da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Bacharel em Engenharia de Produção.

Aprovado em __ de março de 2022.

COMISSÃO ORGANIZADORA

Prof. Dr.-Ing. Renato Antônio Krohling.
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Renato Elias Nunes de Moraes
Universidade Federal do Espírito Santo
Examinador Interno

Prof. Dr. Rodrigo de Alvarenga Rosa
Universidade Federal do Espírito Santo
Examinador Interno

AGRADECIMENTOS

Aos meus pais, por toda a ajuda e suporte afetivo em toda minha jornada enquanto filho e estudante e por investir e acreditar em meu potencial.

Aos excelentes professores que tive ao longo de minha vida, por terem me guiado pelo caminho do conhecimento.

Ao meu orientador, prof. Dr.-Ing Renato, por ter confiado em minha capacidade, por ter me dado direcionamento para atingir meus objetivos e pela compreensão e consideração.

Aos meus amigos, por terem me ajudado a manter o foco com calma e tranquilidade e por ter me feito acreditar no meu potencial. Em especial, à Amanda, por ter me ajudado, inspirado e motivado na vida e nos conhecimentos acadêmicos.

Aos meus colegas de trabalho, por terem me dado apoio, recursos e liberdade para utilizar meu tempo e direcionar meus esforços.

A Deus, por ter me capacitado e me sustentado, mesmo nos momentos mais difíceis.

RESUMO

A predição de demanda é essencial para o planejamento e controle da produção de uma empresa e está totalmente ligada à qualidade da cadeia de suprimento da mesma. Uma boa predição evita retrabalhos nos processos produtivos e na gestão comercial, principalmente quando se trata de bens de consumo perecíveis, a fim de se evitar perdas de produtos ao mesmo tempo em que se evita deixar as prateleiras das lojas vazias. Este trabalho tem o objetivo de aplicar métodos de predição de séries temporais para fins comparativos com métodos de predição tradicionais que são utilizados na maioria das vezes. Foram aplicadas redes neurais multicamadas e florestas aleatórias. Resultados obtidos da predição foram comparados com o método de médias móveis e se mostraram promissores.

Palavras-chave: PCP. Predição. Redes neurais. Florestas aleatórias.

ABSTRACT

Demand forecasting is essential to the production planning and control of a company and it is totally connected to the quality of the supply chain management. A good forecast avoids rework inside the production processes and commercial management, it gets an even bigger importance when it comes to perishable goods, in order to balance waste and avoid out of stock shelves on the stores at the same time. This work's objective is to apply computational models of time-series forecasting in order to compare the performance of those models against traditional forecasting models which are used most of the time in this context. Neural networks and random forest were applied. Forecasting results have been compared to the moving average method and its results proved promising.

Keywords: PPC. Forecasting. Neural networks. Random forest.

LISTA DE FIGURAS

Figura 1 - Estrutura de um neurônio artificial.	14
Figura 2 - Exemplo de arquitetura de uma rede neural RNA com uma camada de entrada, uma camada intermediária e uma camada de saída.	15
Figura 3 - Pseudocódigo do algoritmo do backpropagation.	16
Figura 4 - Representação esquemática de FA.	17
Figura 5 - Histograma da distribuição dos dados de venda (em kg) do produto 42 na loja 2. .	24
Figura 6 - Predições realizadas e dados reais do conjunto de teste.	25
Figura 7 - Histograma da distribuição dos dados de venda (em kg) dos 5 itens do experimento 2.	27
Figura 8 - Predição para o conjunto de dados de teste usando RNA e FA para o item 42.	27
Figura 9 - Histograma de distribuição dos valores de venda (em kg) dos 10 produtos selecionados em 2 lojas diferentes.	29
Figura 10 - Métricas de desempenho para RNA, FA e média móvel no conjunto de dados de treino e validação.	30

LISTA DE TABELAS

Tabela 1 - Exemplo de uma base de dados modelada com $n=3$ semanas de atraso.	19
Tabela 2 - Métricas para diferentes modelagens de dados testadas no modelo RNA.	22
Tabela 3 - Métricas para diferentes modelagens de dados testadas no modelo FA.	22
Tabela 4 - Resultados das métricas obtidas no experimento feito com um produto em uma loja.	24
Tabela 5 - Resultados das métricas obtidas no experimento feito com 5 produtos em 1 loja..	26
Tabela 6 - Resultados das métricas do produto 42 para os modelos treinados com 5 produtos na mesma loja do primeiro experimento.	26
Tabela 7 - Métricas obtidas no experimento de 10 produtos em 2 lojas.	28
Tabela 8 - Métricas obtidas para o item 42 para os modelos treinados no experimento de 10 produtos e 2 lojas.	28

LISTA DE ABREVIATURAS E SIGLAS

BI – *Business Intelligence*

FA – Florestas Aleatórias

IA – Inteligência Artificial

Kg – Quilograma

MAE – Erro Médio Absoluto

MAPE – Erro Percentual Médio Absoluto

MSE – Erro Médio Quadrático

PCP – Planejamento e Controle da Produção

RMSE – Desvio Médio Quadrático

RMSLE – Desvio Médio Quadrático Logarítmico

RNA – Redes Neurais Artificiais

UFES – Universidade Federal do Espírito Santo

SUMÁRIO

1	INTRODUÇÃO	10
1.1	CONTEXTUALIZAÇÃO	10
1.2	MOTIVAÇÃO E JUSTIFICATIVA	11
1.3	OBJETIVOS	12
1.4	ORGANIZAÇÃO DA MONOGRAFIA	12
2	METODOLOGIA	14
2.1	REDES NEURAIS ARTIFICIAIS	14
2.2	FLORESTAS ALEATÓRIAS	16
2.3	BASE DE DADOS	17
2.4	MODELAGEM DOS DADOS	18
2.5	DIVISÃO DOS CONJUNTOS DE DADOS	19
2.6	CRIAÇÃO DOS MODELOS DE PREDIÇÃO	19
2.7	MÉTRICAS DE DESEMPENHO PARA PREDIÇÕES	20
3	RESULTADOS EXPERIMENTAIS	23
3.1	EXPERIMENTO 1	23
3.2	EXPERIMENTO 2	25
3.3	EXPERIMENTO 3	28
4	CONCLUSÃO	31
	REFERÊNCIAS BIBLIOGRÁFICAS	32

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A previsão de demanda é uma das áreas de estudo do planejamento e controle da produção, ou PCP (CHAPMAN, 2006), e tem como objetivo estimar valores futuros com base em valores já registrados de uma série de dados históricos. Séries de dados que são observados em um determinado período são chamadas de séries temporais (BOX; JENKINS; REINSEL, 1980) e a previsão de demanda está diretamente relacionada ao estudo e análise de séries temporais.

Dentro do campo de estudo de uma série temporal, busca-se modelar o comportamento dessa série, e, da compreensão de quais parâmetros têm influência sobre o comportamento desta para então aplicar técnicas de predição de séries temporais, que é justamente a estimativa de valores futuros de uma série temporal.

A predição de séries temporais tem diversas aplicações nos mais variados contextos e é de suma importância, podendo agregar vantagem competitiva para empresas em seus planejamentos estratégicos (CONSTANTIN, 2016).

Além disso, a eficiência de uma cadeia de suprimentos também é influenciada diretamente pela qualidade da predição realizada (ADEBANJO; MANN, 2000) e isso se torna ainda mais evidente em casos em que a cadeia de suprimento é para o comércio de bens de consumo perecíveis. Isso ocorre devido ao fato de que para este tipo de caso, tanto a superestimativa como a subestimativa da demanda são muito prejudiciais para a empresa (HUBER; GOSSMANN; STUCKENSCHMIDT, 2017).

Quando a demanda é superestimada, ocorre uma entrega na cadeia de suprimento maior do que a demanda real e, como são bens perecíveis, essa parcela além da demanda real que foi fornecida acaba sendo descartada, trazendo prejuízo ligado ao custo de produção ou aquisição e logística daquele item.

Já com a demanda subestimada, o malefício é enxergado de forma diferente, sendo mais difícil de se quantificar. Porém, quando há falta de um produto nas prateleiras de um estabelecimento, a fidelidade do consumidor e probabilidade de futuras vendas é afetada negativamente (ZINN; LIU, 2008).

Desta forma, ainda que não seja totalmente estimável, há prejuízo tanto na perda de uma possível venda, onde o estabelecimento deixa de ganhar pela falta de produto, como um prejuízo na imagem e conexão que o estabelecimento tem com seus clientes.

Para se atingir um nível satisfatório de predição de séries temporais, deve-se investigar os dados e diferentes técnicas para identificar quais têm melhor desempenho na predição a fim de se chegar numa melhor estimativa.

Existem diversas técnicas de predição para séries temporais, desde as mais simples como média móvel (JOHNSTON; BOYLAND; MEADOWS, 1999) até técnicas pouco sofisticadas e tradicionais, como o modelo denominado autorregressivo integrado de médias móveis (ARIMA), introduzido por Box e Jenkins em 1976.

Para que um modelo tenha uma boa qualidade de predição é essencial que a tomada de decisão seja sustentada por dados (POWER, 2008), juntamente com a experiência dos responsáveis pela definição das previsões, tendo como base fatores que influenciam a demanda, tais como promoções ou sazonalidades específicas de determinados itens.

No histórico mais recente de técnicas de predição baseadas em dados, houve o desenvolvimento de modelos mais complexos, como Redes Neurais Artificiais (RNA) (GARDNER; DORLING, 1998) e o florestas aleatórias (FA) (BREIMAN, 2001).

Redes neurais artificiais (RNA) usam o algoritmo de treinamento conhecido como *backpropagation* (RUMELHART; HINTON; WILLIAMS, 1986), RNAs podem ser utilizadas em aplicações onde é necessária a manipulação de uma grande quantidade de dados, termo conhecido como *Big Data*.

Outro algoritmo amplamente utilizado é FA, baseado na combinação aritmética de saídas de múltiplas árvores de decisões que são formadas e treinadas por meio de inteligência artificial (IA) à partir de seleção aleatória de subconjuntos de dados e variáveis de decisão da base de dados.

Em geral estes algoritmos costumam apresentar bons resultados para problemas de regressão em análises de séries temporais, com resultados satisfatórios para predição de demanda com boa qualidade que sustentará a cadeia de suprimentos (HUBER; GOSSMANN; STUCKENSCHMIDT, 2017).

1.2 MOTIVAÇÃO E JUSTIFICATIVA

Para este trabalho, o tema de estudo será uma rede de padarias da região da Grande Vitória, que possui uma central de produção, com produção e distribuição diária para as lojas. A equipe de trabalho da central de produção que gerencia o PCP dos itens de produção própria da empresa e a predição de demanda da quantidade para cada item a ser entregue em cada loja. É notado no cotidiano da empresa que a

predição da demanda é o ponto de partida que vai ditar a qualidade e eficiência do sistema de PCP da empresa e da cadeia de suprimento das lojas como um todo.

Quando a predição não é bem ajustada, ajustes pontuais na produção e logística precisam ser feitos, tomando tempo da equipe de gerenciamento de produção por conta de retrabalho. Isso ocorre justamente para se reduzir ou mitigar os malefícios citados anteriormente de uma superestimativa ou subestimativa de demanda dos itens de produção própria nas prateleiras das lojas.

A superestimativa ainda tange outro ponto importante relacionado ao papel social de uma organização. Quando se trata de bens de consumo alimentício perecíveis, sabe-se que existe uma má distribuição desses recursos na sociedade, tanto em âmbito local como global. Portanto, deve-se ter uma atenção especial na otimização do consumo destes bens para que se busque uma melhoria neste ponto, ainda que com impacto pequeno na sociedade como um todo.

Atualmente, o método de predição para as séries temporais das lojas é média móvel, um método básico de predição, porém, RNA e FA podem ser promissores para predição de séries temporais de bens de consumo perecíveis (HUBER, 2019).

Neste trabalho será investigada a viabilidade da aplicação de RNAs e FA para os itens de produção própria da empresa a fim de comparação de resultados obtidos na predição.

1.3 OBJETIVOS

Neste trabalho, o objetivo consiste na aplicação de redes neurais artificiais e florestas aleatórias para predição de demanda usando dados reais de uma empresa de produtos perecíveis. Para analisar a eficácia dos algoritmos, será feita comparação dos resultados obtidos com aqueles obtidos via média móvel, que é utilizado atualmente pelo PCP da empresa.

1.4 ORGANIZAÇÃO DA MONOGRAFIA

No capítulo 2 deste trabalho a metodologia será apresentada, aprofundando a descrição da arquitetura da RNA e FA. Também será apresentada a fonte, modelagem e divisão dos dados selecionados para os experimentos. As métricas de desempenho utilizadas para os experimentos também serão apresentadas no capítulo 2.

Os experimentos realizados e seus resultados serão mostrados no capítulo 3 deste trabalho, dividido em 3 seções, uma para cada experimento e apresentação dos

seus respectivos resultados. A conclusão dos resultados obtidos será desenvolvida no capítulo 4, juntamente de recomendações para trabalhos futuros.

2 METODOLOGIA

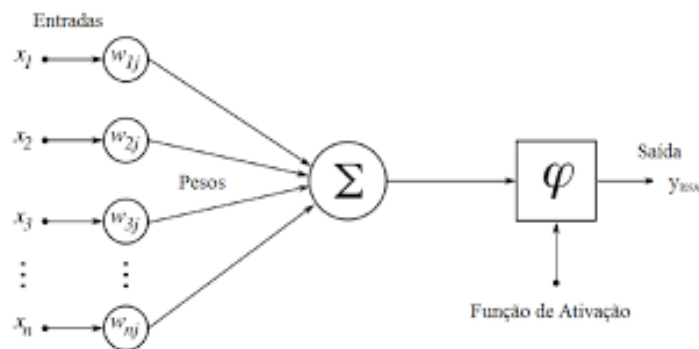
Este capítulo introduz os conceitos e as metodologias utilizadas para cada etapa na aplicação dos algoritmos e modelagem dos dados utilizados.

2.1 REDES NEURAIS ARTIFICIAIS

RNA é um modelo computacional que tem seu funcionamento inspirado no modo de transmissão de sinais nos neurônios do cérebro de seres biológicos (DE ALMEIDA BARBOSA, 2021). A arquitetura de um RNA é composta de camadas de “neurônios artificiais” denominados *perceptrons*.

Um neurônio recebe entradas encaminhadas de neurônios anteriores a ele, juntamente com os pesos definidos para cada entrada e, através de uma função de ativação, processa as entradas e pesos para emitir uma saída para a próxima camada de neurônios, conforme mostrado na Figura 1.

Figura 1 - Estrutura de um neurônio artificial.



A função de ativação é a função que, dadas as entradas naquele neurônio, as processa e fornece a saída deste neurônio, que é entrada para o neurônio seguinte. Funções que são comumente utilizadas como função de ativação são as funções degrau unitário, sigmoide e tangente hiperbólica, que são definidas pelas equações mostradas a seguir.

$$\text{Degrão unitário} = \begin{cases} 0, & \text{se } x < 0 \\ 1, & \text{se } x \geq 0 \end{cases}$$

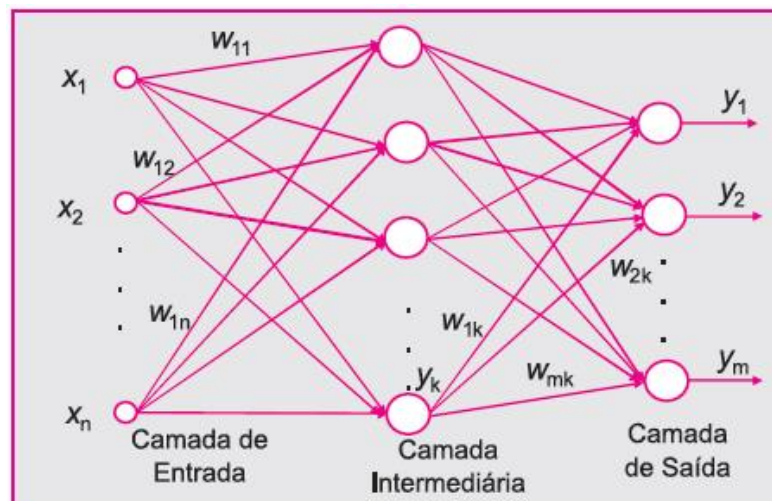
$$\text{Sigmoide} = \frac{1}{1 + e^{-x}}$$

$$\text{Tangente hiperbólica} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

A função de ativação é o agente que torna o neurônio um classificador de entradas. A função degrau unitário separa as entradas em dois grupos, definidos pelas saídas 1 ou 0. As funções sigmóides e tangente hiperbólica atribuem valores próximos a 0 para entradas negativas que têm módulo relativamente grande, valores próximos a 1 para valores de entradas positivas que têm módulo relativamente grande e atribuem valores entre 0 e 1 para valores de entrada próximos de 0.

Uma rede neural multicamadas possui uma camada de entrada, cujos neurônios vão ser alimentados pelos dados de variáveis de entrada, uma ou mais camadas intermediárias e uma camada de saída. A arquitetura de uma RNA está representada na Figura 2.

Figura 2 - Exemplo de arquitetura de uma rede neural RNA com uma camada de entrada, uma camada intermediária e uma camada de saída.



Os pesos de uma RNA são ajustados pelo algoritmo *backpropagation* (RUMELHART; HINTON; WILLIAMS, 1986), que é parte fundamental do treinamento de rede. Dentro de uma RNA algoritmo atua minimizando uma função de erro, por exemplo erro quadrático médio, entre a saída da rede, dado o padrão de entrada, e a saída esperada, ajustando os pesos da RNA para atingir a minimização.

Os pesos definidos entre a camada de entrada da RNA e as camadas intermediárias e os pesos entre as camadas intermediárias e a camada de saída são armazenados pela rede em forma de matrizes que são transformadas em vetor para se realizar o treinamento da rede, por conta da simplicidade deste tipo de formato.

O pseudocódigo do *backpropagation* é descrito a seguir.

Figura 3 - Pseudocódigo do algoritmo do backpropagation.

```

01: inicializar os pesos da rede
02: repita
03:   para cada padrão  $p$  do conjunto de treinamento
04:     apresentar padrão à rede
05:     para cada peso sináptico  $w$ 
06:       calcular  $\Delta w$ 
07:        $w = w + \Delta w$ 
08:     fim para
09:   fim para
10: até critério de parada ser satisfeito

```

O algoritmo de uma RNA possui diversos parâmetros que afetam o desempenho das predições feitas pela rede. Os principais parâmetros são número de camadas intermediárias, número de neurônios em cada uma das camadas intermediárias, o método de otimização dos pesos da rede, taxa e forma de aprendizado da rede.

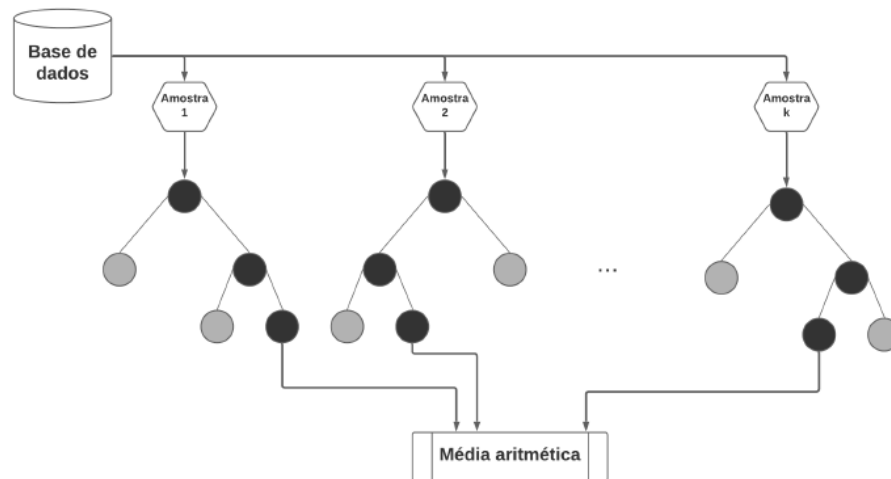
Estes parâmetros são definidos por meio de seleção no código do algoritmo, de acordo com a combinação que resulta no melhor desempenho das predições.

2.2 FLORESTAS ALEATÓRIAS

O algoritmo FA pode ser resumido como uma combinação de múltiplas árvores de regressão a partir da segmentação aleatória do conjunto de dados. Para este modelo, subconjuntos da base de dados são escolhidos aleatoriamente pelo algoritmo, assim como a escolha de subconjuntos das variáveis de decisão para cada subconjunto de dados.

Com essa seleção aleatória dos subconjuntos, árvores de regressão são desenvolvidas separadamente entre si e os resultados de cada árvore formada são combinados em forma de média aritmética (HOU; EDARA; CHANG, 2017). Uma representação esquemática do algoritmo é mostrada na Figura 3.

Figura 4 - Representação esquemática de FA.



As FA combinam a simplicidade das árvores de decisão com a flexibilidade, generalização e aleatoriedade para melhorar a acurácia da predição. O mecanismo de funcionamento de uma FA se dá pela segmentação da base de dados para a criação das árvores de decisão, selecionando de forma aleatória diferentes conjuntos de entradas e diferentes variáveis de decisão entre as árvores criadas, podendo ou não repetir conjuntos ou variáveis de decisão em diferentes árvores, de acordo com os parâmetros definidos para o algoritmo.

Desta forma, as árvores criadas possuem tamanhos e características distintas entre si, o que contribui para a generalização do algoritmo. Quando os dados de teste são usados no algoritmo, eles passam por todas as árvores geradas e as saídas de cada árvore são combinadas para gerar a saída final do algoritmo.

São parâmetros do algoritmo das FA o número de árvores, tamanho máximo da árvore, número mínimo de amostras necessárias para a divisão de um nó interno das árvores, número máximo de variáveis de decisão escolhidas para a escolha aleatória na segmentação dos dados e a permissão ou negação da repetição de entradas ou variáveis de decisões nas diferentes segmentações da base de dados.

2.3 BASE DE DADOS

O primeiro passo para a implementação do trabalho é a aquisição dos dados de vendas dos itens de produção própria nas diferentes lojas, que serão utilizados para realizar os testes de desempenho. Para tal, foi utilizado o sistema de *business intelligence*, ou BI (CHAUDHURI; DAYAL; NARASAYYA, 2011), que armazena os

dados de venda de cada item que compõe o mix de produção própria. No sistema do BI é possível selecionar o intervalo de tempo, quais segmentos de produtos e lojas se deseja fazer uso da informação e é retornado um relatório com a venda semanal de cada item para cada loja em cada semana diferente.

Para diferentes testes foram selecionados diferentes filtros, começando com testes com apenas um produto de uma loja e, na sequência, o número de itens e lojas foram aumentados para novos testes.

O horizonte de tempo selecionado foi da primeira semana de 2017 até a semana mais recente no presente momento dos experimentos realizados, sempre selecionando dados começando numa segunda-feira e terminando num domingo, para fazer a correspondência exata do total de uma semana agrupada.

Este relatório gerado pelo BI funciona como a base de dados dos experimentos, que é importada diretamente na plataforma *Jupyter*. É a partir desta base que o tratamento e modelagem dos dados são feitos para a utilização nos algoritmos investigados.

2.4 MODELAGEM DOS DADOS

Normalmente, busca-se complementar a base de dados com mais informações para o treinamento dos algoritmos. Uma das formas de se fazer isso é adicionar colunas à base de dados com informações de vendas das semanas anteriores e a diferença entre semanas anteriores (FILHO, 2022).

Com base neste princípio, testes foram realizados para uma base de dados de um único produto em uma única loja, a fim de identificar se havia uma melhoria nos resultados do treinamento entre a base de dados com o acréscimo dessas colunas e a base de dados somente com a informação de venda semanal.

Para este teste, foram feitos diferentes experimentos, com diferentes n semanas de atraso, em que as colunas criadas eram da forma “*Venda -n*” para representar a venda da semana n anterior à semana de referência daquele dado e “*Diferença -n*” para representar a diferença de venda entre as semanas n e $n+1$ anteriores à semana de referência.

Considere os dados listados na Tabela 1. Note que na linha da semana 7, seguindo a ordem das colunas temos a venda da semana 7, a venda da semana 6 (onde $n = 1$), a diferença entre a venda da semana 6 e 5 (onde $n = 1$ e $n+1 = 2$), e assim por diante, até $n = 3$.

Tabela 1 - Exemplo de uma base de dados modelada com $n=3$ semanas de atraso.

Semana	Venda	Venda -1	Diferenca -1	Venda -2	Diferenca -2	Venda -3	Diferenca -3
5	42.60	41.16	7.91	33.25	-13.32	46.57	0.50
6	62.80	42.60	1.44	41.16	7.91	33.25	-13.32
7	43.68	62.80	20.20	42.60	1.44	41.16	7.91
8	47.97	43.68	-19.12	62.80	20.20	42.60	1.44
9	50.07	47.97	4.29	43.68	-19.12	62.80	20.20

Os testes realizados foram feitos com a base de dados somente com a venda da semana de referência, bases de dados com acréscimo de colunas somente com as vendas de n semanas anteriores, para valores de $n = (3, 5, 7, 10)$ e bases de dados tanto com colunas das vendas de n semanas anteriores como a diferença entre a semana n e $n+1$ anteriores, para os mesmos valores de n citados anteriormente.

Estes testes serviram de base para a decisão de modelagem de dados para os experimentos seguintes com mais itens e mais lojas.

2.5 DIVISÃO DOS CONJUNTOS DE DADOS

Para este trabalho, a base de dados foi separada entre valores objetivo reais (\hat{y}) e variáveis da camada de entrada (x) sendo y a coluna com o valor da venda de cada semana e x todas as outras colunas da base de dados.

Além disso, é necessário separar o conjunto de dados entre dados de treino, de validação e de teste. Para isso, tomamos os 90% iniciais dos dados para conjuntos de treino e validação, deixando os últimos 10% dos dados como dados de teste. Dentro dos dados de treino e validação, o algoritmo computacional seleciona aleatoriamente 80% dos dados para treino e 20% dos dados para validação, sem ordem específica de dados.

Em seguida, os dados foram normalizados fazendo-se uso da ferramenta da biblioteca *sklearn* denominada *standard scaler* que faz a normalização dos dados automaticamente, para ficarem prontos para serem utilizados nos modelos criados na sequência.

2.6 CRIAÇÃO DOS MODELOS DE PREDIÇÃO

Para o desenvolvimento deste trabalho, foi utilizada a plataforma *Jupyter*, uma

plataforma que opera na linguagem de programação *Python* e utiliza bibliotecas e ferramentas bastante utilizadas para análise de séries temporais, tal qual implementação de algoritmos de predição de séries temporais e plotagem de gráficos.

Com os dados já divididos e normalizados, é possível criar modelos de predição usando RNA e FA. Para isso, foi utilizada a biblioteca *sklearn* presente na plataforma *Jupyter* para importação dos dois modelos.

Os algoritmos possuem alguns parâmetros. Nem sempre os mesmos parâmetros vão resultar em melhores desempenhos. Isso depende diretamente do conjunto de dados que está sendo aplicado ao modelo.

Para definição dos melhores parâmetros para o conjunto de dados de treinamento, foi utilizada uma ferramenta denominada *Grid Search* (PONTES; AMORIM; BALESTRASSI; PAIVA, 2016), que testa diferentes combinações de parâmetros e utiliza a melhor delas para o treinamento e validação do modelo.

Com os parâmetros definidos pelo *Grid Search*, o treinamento dos modelos foi realizado com os dados de treino, validado com os dados de validação para então ter o desempenho das predições do conjunto de dados de teste analisado de acordo com as métricas de desempenho.

2.7 MÉTRICAS DE DESEMPENHO PARA PREDIÇÕES

Métricas de desempenho para predições de séries temporais pode ser definida como um número que quantifica a diferença entre a predição de um valor e o valor real dentro da série de dados, também conhecido como erro de predição.

Estas métricas em geral, podem ser medidas mantendo a escala da série de dados ou através do cálculo percentual do erro em relação ao valor real esperado (HYNDMAN; KOEHLER, 2006).

Para este trabalho, serão abordadas as métricas Erro Médio Absoluto (MAE), Erro Percentual Médio Absoluto (MAPE), Desvio Médio Quadrático (RMSE) e Desvio Médio Quadrático Logarítmico (RMSLE), que serão descritas a seguir.

Para as equações a seguir, vamos utilizar y como o valor real e \hat{y} como o valor da predição. Portanto, o erro é calculado como $y - \hat{y}$.

MAE, erro médio absoluto, é uma das métricas mais simples que existem e consiste basicamente no cálculo da média da somatória do módulo do erro, descrito por:

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

O valor do MAE resulta na mesma escala dos dados utilizados na série, o que facilita sua compreensão. Portanto, o valor apontado pelo MAE é a média do valor absoluto do erro (DE OLIVEIRA, 2021).

MAPE é uma métrica similar ao MAE, porém ele compara percentualmente o erro com o valor real da série temporal fazendo-se a divisão do módulo do erro pelo módulo do valor real.

Note que, caso o valor real seja 0, ocorre uma inconsistência no valor do MAPE, por não existir operação de divisão por zero. Neste caso, a biblioteca *sklearn*, utilizada na plataforma *Jupyter* substitui o valor de y por ϵ , dado por $2,220446049250313e-16$.

A equação do MAPE é descrita por:

$$MAPE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\max(\epsilon, |y_i|)}$$

A métrica MAPE, por trazer uma abordagem percentual do erro, é considerada uma métrica com uma compreensão mais abrangente do que a métrica MAE (DE OLIVEIRA, 2021).

RMSE é uma métrica que é variação de uma outra métrica, denominada Erro Quadrático Médio (MSE). O MSE é uma métrica que se assemelha ao MAE, porém, neste caso ela tem maior peso para erros que apresentam grandes diferenças entre o valor real e o da predição, por elevar cada erro calculado ao quadrado. Porém, por causa do exponencial ao quadrado, o número do MSE não está na mesma escala do conjunto de dados da série de dados, portanto sua interpretação se torna mais difícil.

Para melhor compreensão desta ideia, surgiu então o RMSE, que basicamente é a raiz quadrada do MSE, trazendo então o número da métrica de volta à mesma escala dos dados.

O RMSE é descrito por:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

A métrica RMSE está na mesma escala do conjunto de dados dos valores reais e previstos, assim como no MAE, porém aprofunda a análise dos resultados, acentuando o prejuízo trazido de pontos previstos que estão muito fora do valor real (DE OLIVEIRA, 2021).

Por último, temos uma métrica que foi introduzida mais recentemente como uma espécie de evolução do RMSE. Trata-se do RMSLE, que é descrito por:

$$RMSLE (y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \log(\hat{y}_i + 1))^2}$$

Enquanto valores da predição distante do valor real podem extrapolar o valor do RMSE para valores muito altos, no valor de RMSLE, o efeito de predições distantes tem sua escala reduzida, porém não anuladas, trazendo uma abordagem um pouco diferente para a métrica.

Além disso, por conta das propriedades logarítmicas, o RMSLE também é visto como um erro relativo ao valor real da série temporal (SAXENA, 2019).

Estas serão as métricas utilizadas no comparativo de desempenho dos modelos de predição que serão discutidos a seguir.

3 RESULTADOS EXPERIMENTAIS

3.1 EXPERIMENTO 1

Os primeiros resultados a serem apresentados são dos experimentos iniciais com um único produto para definir a influência da adição das colunas de n atrasos na modelagem do conjunto de dados, como descrito na seção 2.4.

Para isso, serão comparadas as métricas obtidas no conjunto de teste usando RNA e FA para as diferentes modelagens que foram feitas, como apresentado nas tabelas 2 e 3.

Os parâmetros selecionados pelo *Grid Search* neste experimento para a RNA foram função de ativação sigmoide logística, taxa de aprendizado (alfa) igual a 0,0001, uma camada intermediária com 100 neurônios, taxa de aprendizado adaptativa e o método *stochastic gradient* (sgd) como *solver* do modelo.

Para FA os parâmetros foram 30 árvores de regressão, mínimo de amostras para divisão de nós igual a 8, número máximo de variáveis de decisões escolhidas para cada árvore sendo dado pelo logaritmo de base 2 do número total de recursos e com possibilidade de selecionar amostras repetidas vezes para diferentes conjuntos de amostra.

Tabela 2 - Métricas para diferentes modelagens de dados testadas no modelo RNA.

Modelagem de dados – RNA									
MÉTRICAS	$n=0$	$n=3$		$n=5$		$n=7$		$n=10$	
		s/ diferença entre n e $n+1$	c/ diferença entre n e $n+1$	s/ diferença entre n e $n+1$	c/ diferença entre n e $n+1$	s/ diferença entre n e $n+1$	c/ diferença entre n e $n+1$	s/ diferença entre n e $n+1$	c/ diferença entre n e $n+1$
RMSLE	0,2676	0,1506	0,1265	0,1180	0,1179	0,1255	0,1136	0,1054	0,0939
RMSE	15,3893	8,2396	7,2027	6,8156	6,8103	6,9352	6,6733	6,3843	5,8927
MAE	13,5000	6,7433	6,1015	5,6057	5,6650	5,5979	5,4209	4,8678	4,7023
MAPE	0,2763	0,1339	0,1127	0,1015	0,1021	0,1011	0,0959	0,0821	0,0765

Tabela 3 - Métricas para diferentes modelagens de dados testadas no modelo FA.

Modelagem de dados - FA									
MÉTRICAS	$n=0$	$n=3$		$n=5$		$n=7$		$n=10$	
		s/ diferença entre n e $n+1$	c/ diferença entre n e $n+1$	s/ diferença entre n e $n+1$	c/ diferença entre n e $n+1$	s/ diferença entre n e $n+1$	c/ diferença entre n e $n+1$	s/ diferença entre n e $n+1$	c/ diferença entre n e $n+1$
RMSLE	0,2218	0,2157	0,1760	0,1298	0,1238	0,1233	0,1189	0,1030	0,0898
RMSE	12,3120	12,3285	9,9384	7,4114	7,1420	7,0650	6,9752	6,2687	5,6448
MAE	10,4406	10,6730	8,4017	5,3964	5,3821	5,4075	5,5970	4,7038	4,4158
MAPE	0,2158	0,2123	0,1644	0,0997	0,0969	0,0988	0,0995	0,0788	0,0725

Pode-se notar que, no geral, a modelagem dos dados com o acréscimo de colunas com semanas de atraso resulta de maneira direta numa melhora de desempenho dos modelos.

Comparando as métricas da modelagem sem atraso com qualquer uma com atraso, pode-se averiguar que as métricas resultam melhor com modelos com atraso, tendo um salto significativo de melhoria entre $n=0$ e $n=3$ para o modelo RNA.

Também fica evidente que as métricas seguem uma tendência de melhoria à medida que se aumenta a quantidade n de semanas de atraso, tanto para a RNA quanto para FA, todos os modelos com $n=10$ desempenharam melhor que os modelos com $n=3$, por exemplo.

Além disso, para os mesmos valores de n , fica claro que para a maioria dos casos o melhor desempenho é quando também há a adição da coluna com a diferença entre a semana de atraso n para a semana de atraso $n+1$.

Com isso, pode-se concluir que é vantajoso modelar os dados adicionando-se tanto as colunas com vendas de semanas anteriores quanto a diferença de venda entre essas semanas e a anterior a elas, em acordo com outras experimentações analisadas (FILHO, 2021).

Apesar dos resultados indicarem que quanto mais atrasos, melhor a qualidade do modelo, por conta de demanda de processamento computacional, optou-se pela modelagem com 3 semanas de atraso, contendo colunas com a diferença entre n e $n+1$ para os experimentos realizados com mais produtos.

Com a modelagem definida, pode-se comparar agora o desempenho dos modelos RNA e FA com o desempenho da predição do método já utilizado atualmente pela empresa.

O método de predição atual no PCP da empresa é a média móvel com 3 semanas de janela de atraso, ou seja, a predição para a próxima semana é a média móvel da venda de cada item em cada loja nas 3 semanas anteriores.

Inicialmente, o experimento foi feito para um produto em uma loja, neste caso, o produto em questão é o produto de código interno 42 e a loja escolhida foi a loja de código interno 2 para comparação inicial e o número de produtos e lojas foi incrementado a cada experimento, para observação da diferença de resultado quando ocorre a inserção de mais dados para o processamento dos modelos.

Na tabela 4, estão listados os valores das métricas obtidas para cada modelo de predição com o experimento feito no conjunto de dados de teste para um produto

em uma loja.

Tabela 4 - Resultados das métricas obtidas no experimento feito com um produto em uma loja.

Métricas obtidas para experimento com um produto em uma loja			
	RNA	FA	média móvel
RMSLE	0,1265	0,1315	0,1227
RMSE	7,2027	7,4233	7,1260
MAE	6,1015	6,0971	6,1210
MAPE	0,1127	0,1136	0,1108

Neste experimento, pode-se observar que as 3 previsões tiveram métricas próximas umas das outras, porém com a média móvel ainda um pouco melhor que os outros dois modelos treinados.

Os modelos RNA e FA tiveram vantagem somente na métrica MAE, o que indica que tiveram erro absoluto, em média, menor que da média móvel, mas com mais previsões distantes dos valores reais, o que prejudicou as outras métricas para estes dois modelos.

Na Figura 4 é mostrado o histograma com a distribuição dos dados de venda deste produto e na Figura 5 o resultado visual comparativo entre os valores das previsões e os valores reais do conjunto de dados de teste.

Figura 5 - Histograma da distribuição dos dados de venda (em kg) do produto 42 na loja 2.

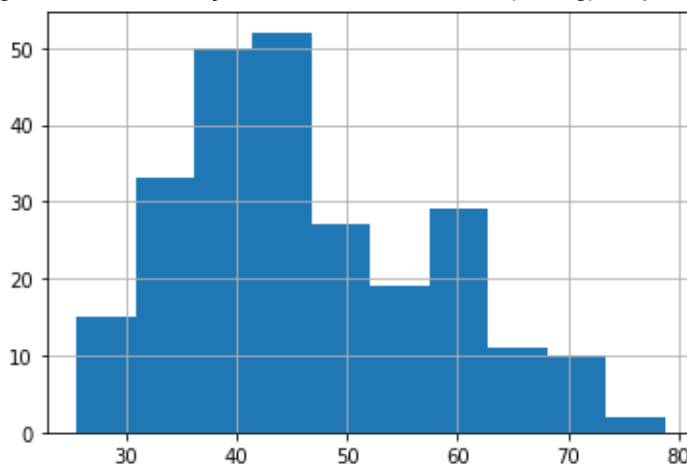
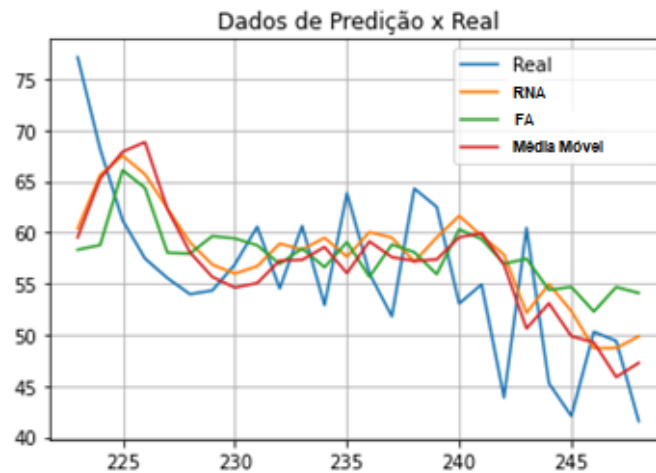


Figura 6 - Predições realizadas e dados reais do conjunto de teste.



Nota-se no gráfico da Figura 5 a proximidade dos desempenhos nas métricas sendo repetida nas oscilações que se aproximam das oscilações dos dados reais.

3.2 EXPERIMENTO 2

Para o experimento seguinte, foram utilizados dados de venda de 5 produtos para a mesma loja do experimento inicial, sendo um dos produtos o de código interno 42, para que fosse possível comparar as métricas para as predições dos 5 produtos e também verificar se, quando segmentamos os dados de teste somente para o produto 42, os modelos treinados com 5 produtos conseguiram prever melhor as vendas do item 42 quando informações de outros produtos foram agregadas ao conjunto de dados para os modelos RNA e FA.

Os parâmetros selecionados pelo *Grid Search* neste experimento para a RNA foram função de ativação sigmoide logística, taxa de aprendizado (alfa) igual a 0,0001, três camadas intermediárias com 50, 100 e 50 neurônios, respectivamente, taxa de aprendizado adaptativa e o método *stochastic gradient (sgd)* como *solver* do modelo.

Para FA os parâmetros foram 100 árvores de regressão, mínimo de amostras para divisão de nós igual a 8, número máximo de recursos sendo dado pela raiz quadrada do total de recursos e com possibilidade de selecionar amostras repetidas vezes para diferentes conjuntos de amostra.

Os resultados são apresentados nas Tabelas 5 e 6.

Tabela 5 - Resultados das métricas obtidas no experimento feito com 5 produtos em 1 loja.

Métricas obtidas para experimento com 5 produtos em 1 loja			
	RNA	FA	média móvel
RMSLE	0,1340	0,1431	0,1324
RMSE	5,4713	5,6451	5,4204
MAE	4,1744	4,3512	4,1904
MAPE	0,0991	0,1041	0,1038

Tabela 6 - Resultados das métricas do produto 42 para os modelos treinados com 5 produtos na mesma loja do primeiro experimento.

Métricas obtidas para o item 42 no experimento com 5 itens			
	RNA	FA	média móvel
RMSLE	0,1181	0,1154	0,1227
RMSE	6,4936	6,2122	7,1260
MAE	5,7719	5,4037	6,1210
MAPE	0,1100	0,1045	0,1108

Analisando-se os resultados obtidos na Tabela 6, pode-se notar que, com mais informação agregada para os modelos RNA e FA, obteve-se uma melhora em seus desempenhos quando comparados ao desempenho da média móvel.

Apesar da média móvel ainda apresentar um desempenho praticamente tão bom quanto o modelo RNA neste caso, as duas predições ficaram quase equivalente em termos de métricas para os 5 itens. O algoritmo FA obteve métricas piores que os outros 2 modelos, porém a diferença relativa é pequena.

Pelos dados apresentados na Tabela 7, tanto RNA quanto FA neste caso tiveram resultados melhores do que a média móvel, tendo o FA como melhor modelo para o produto 42 neste experimento.

Isto indica que mais informações agregadas ao conjunto de teste resultou numa melhoria de desempenho para se realizar predições, principalmente para FA.

Na Figura 6 é mostrado o histograma da distribuição dos dados de vendas dos

5 produtos deste experimento. Na Figura 7 são mostrados os resultados das previsões para o item 42 neste experimento junto dos valores reais.

Figura 7 - Histograma da distribuição dos dados de venda (em kg) dos 5 itens do experimento 2.

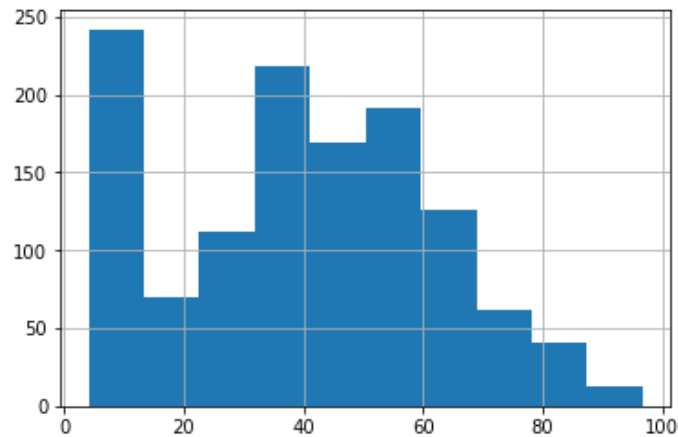
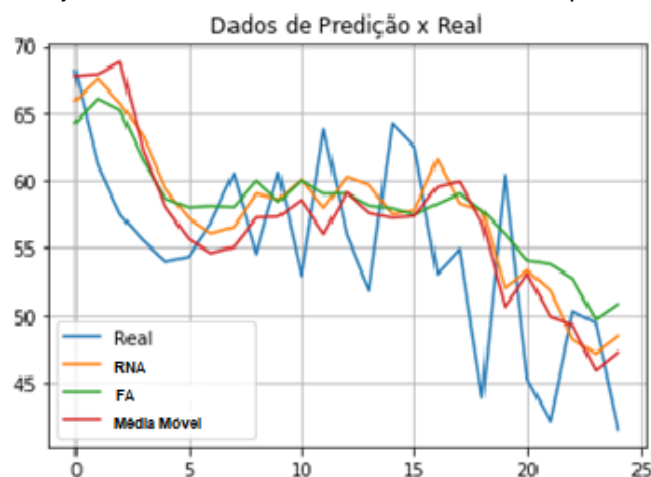


Figura 8 - Predição para o conjunto de dados de teste usando RNA e FA para o item 42.



Comparando o histograma da Figura 6 com o histograma da Figura 4 do experimento 1, nota-se que há uma concentração maior de venda no segundo experimento na faixa dos 10kg, no eixo horizontal do gráfico.

Enquanto isso, a concentração maior do histograma do primeiro experimento fica em torno dos 40kg. Esta diferença justifica o motivo das métricas MAE e RMSE do segundo experimento terem sido menores, já que estes dois indicadores apresentam números na mesma escala dos números de venda dos itens, enquanto os indicadores MAPE e RMSLE são métricas com números relativos ao valor objetivo.

3.3 EXPERIMENTO 3

O terceiro experimento foi realizado agrupando dados de vendas de 10 produtos de 2 lojas diferentes, incluindo o item 42 e a loja 2, para fins comparativos.

Os parâmetros selecionados pelo *Grid Search* neste experimento para a RNA foram função de ativação sigmoide logística, taxa de aprendizado (alfa) igual a 0,05, camada intermediárias com 100 neurônios, taxa de aprendizado adaptativa e o método *stochastic gradient* (sgd) como *solver* do modelo.

Para FA os parâmetros foram 50 árvores de regressão, mínimo de amostras para divisão de nós igual a 8, número máximo de recursos sendo dado pelo logaritmo de base 2 do número total de recursos e com possibilidade de selecionar amostras repetidas vezes para diferentes conjuntos de amostra.

Os resultados das métricas apresentados pelos modelos nos dados de teste são listados nas Tabelas 7 e 8:

Tabela 7 - Métricas obtidas no experimento de 10 produtos em 2 lojas.

Métricas obtidas para experimento com 10 produtos em 2 lojas			
	RNA	FA	média móvel
RMSLE	0,1911	0,1681	0,1427
RMSE	4,4983	4,5992	4,3809
MAE	3,2854	3,2559	3,1256
MAPE	0,1826	0,1585	0,1300

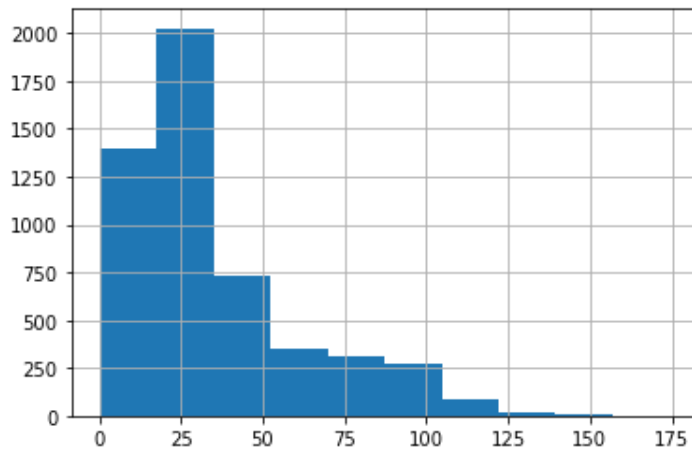
Tabela 8 - Métricas obtidas para o item 42 para os modelos treinados no experimento de 10 produtos e 2 lojas.

Métricas obtidas para o item 42 no experimento com 10 itens			
	RNA	FA	média móvel
RMSLE	0,1432	0,1499	0,1398
RMSE	7,2246	7,5428	7,0112
MAE	6,5293	6,7441	6,3553
MAPE	0,1339	0,1404	0,1313

Como pode ser observado neste experimento os modelos RNA e FA tiveram um desempenho bem inferior ao da média móvel nas métricas que têm o valor dado através de relação com o valor real da série (RMSLE e MAPE).

Para investigar o motivo, foi observado o histograma de distribuição das vendas dos 10 produtos selecionados mostrado na Figura 8.

Figura 9 - Histograma de distribuição dos valores de venda (em kg) dos 10 produtos selecionados em 2 lojas diferentes.



Comparando este histograma com os histogramas dos outros experimentos, podemos notar que a concentração dos valores de venda deste histograma está em sua maioria na faixa de 0 a cerca de 30kg, mas com alguns dados que extrapolam a faixa dos 125kg.

Isso pode ser um indicativo de que estes dados que estão muito afastados da concentração principal dos dados de venda podem ter uma influência negativa no desempenho dos modelos RNA e FA.

Um fato que contribui para este argumento são as métricas, para estes mesmos modelos treinados, somente para o item 42 e loja 2, que como já vimos possui distribuição mais concentrada no histograma de vendas. Apesar de estarem com números piores quando comparados aos da média móvel, estão bem mais próximos do que estão no contexto geral.

Na Tabela 9, são apresentadas as métricas obtidas no conjunto de treino e validação para os modelos RNA, FA e como desempenharia se a predição da média móvel tivesse sido feita para este mesmo conjunto de dados.

Figura 10 - Métricas de desempenho para RNA, FA e média móvel no conjunto de dados de treino e validação.

Treinamento e validação			
	RNA	FA	média móvel
RMSLE	0,1679	0,1175	0,1788
RMSE	5,2145	3,5416	5,5493
MAE	3,3677	2,2307	3,5280
MAPE	0,1441	0,0955	0,1478

De acordo com as métricas para o conjunto de treino e validação, o modelo RNA e, principalmente, o modelo de FA, tiveram desempenho superior ao da média móvel neste conjunto de dados. Isto também indica que um treinamento eficaz de um modelo não necessariamente irá implicar em boas previsões para novos dados.

4 CONCLUSÃO

Este trabalho propôs analisar e comparar o desempenho de modelos computacionais de predição de séries temporais com o método de predição da média móvel. A metodologia aplicada foi capaz de prover respostas para estas questões.

Redes Neurais Artificiais (RNAs) e florestas aleatórias (FA) apresentam resultados promissores comparados a média móvel em experimentos realizados, mas nem sempre são o melhor algoritmo a ser utilizado. Ficou claro que a modelagem dos dados tem total influência no desempenho final dos modelos, sendo um dos pontos mais importantes de toda a metodologia.

Apesar de nem todos os resultados terem superado o desempenho da média móvel para as predições nestes conjuntos de dados, o objetivo inicial de se realizar um levantamento comparativo de desempenho entre os dois algoritmos de predição para séries temporais pôde ser considerado atingido.

Tópicos que podem ser abordados em trabalhos futuros são a investigação comparativa de desempenho com mais semanas de atraso e com outras formas de agregar informação na modelagem dos dados, além de aplicar esta metodologia com conjuntos de dados ainda maiores para analisar o desempenho dos modelos.

REFERÊNCIAS BIBLIOGRÁFICAS

ADEBANJO, D.; MANN, R. **Identifying problems in forecasting consumer demand in the fast moving consumer goods sector**. Benchmarking: An international journal, 2000.

BOX, G., JENKINS, G. M. & REINSEL, G. C. **Time series analysis: Forecasting and control**. Prentice Hall, New Jersey, 2008.

BOX, G. E., JENKINS, G. M., REINSEL, G. C., & LJUNG, G. M. **Time series analysis: forecasting and control**. John Wiley & Sons, 2015.

BREIMAN, L. **Florestas aleatórias**. Machine learning, v. 45, n. 1, p. 5-32, 2001.

CHAPMAN, S. N. **The fundamentals of production planning and control**. Upper Saddle River, NJ, USA: Pearson/Prentice Hall, 2006.

CHAUDHURI, S.; DAYAL, U.; NARASAYYA, V. **An overview of business intelligence technology**. Communications of the ACM, 54 (8), 88–98. 2011.

CONSTANTIN, C. **The importance of sales forecasting in establishing marketing strategies**. Bulletin of the Transylvania University of Brasov. Economic Sciences. Series V, v. 9, n. 1, p. 3, 2016.

DE ALMEIDA BARBOSA, P. J. **Combinação de redes neurais LSTM para previsão de demanda na cadeia de suprimentos perecíveis**. Dissertação (Mestrado em Engenharia de Controle e Automação) – Instituto Federal do Espírito Santo, Serra, 2021.

DE OLIVEIRA JÚNIOR, C. **Números anteriores: Entendendo as medidas R², MAE, MAPE, MSE e RMSE**. 2021. Disponível em: <https://medium.com/data-hackers/prevendo-n%C3%BAmeros-entendendo-m%C3%A9tricas-de-regress%C3%A3o-35545e011e70>.

FILHO, M. **How To Predict Multiple Time Series At Once With Scikit-Learn (With a Sales Forecasting Example)**. Disponível em: <https://www.mariofilho.com/how-to-predict-multiple-time-series-with-scikit-learn-with-sales-forecasting-example/>. Acesso em: 19 dez. 2021.

GARDNER, M. W.; DORLING, S. R. **Artificial neural networks (the multilayer perceptron) — a review of applications in the atmospheric sciences**.

Atmospheric environment, v. 32, n. 14-15, p. 2627-2636, 1998.

HOU, Y.; EDARA, P.; CHANG, Y. **Road network state estimation using florestas aleatórias ensemble learning**. In: IEEE. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). [S.l.], 2017. p. 1–6. 16.

HUBER, J., GOSSMANN, A., & STUCKENSCHMIDT, H. **Cluster-based hierarchical demand forecasting for perishable goods**. *Expert Systems with Applications*, n. 76, p. 140–151, 2017.

HUBER, J. **Data-driven Decision Support for perishable Goods**. 2019. Tese (PhD) – Universität Mannheim, Mannheim, 2019.

HYNDMAN, R. J.; KOEHLER, A. B. **Another look at measures of forecast accuracy**. *International journal of forecasting*, Elsevier, v. 22, n. 4, p. 679–688, 2006.

JOHNSTON, F. R. *et al.* **Some properties of a simple moving average when applied to forecasting a time series**. *Journal of the Operational Research Society*, v. 50, n. 12, p. 1267-1271, 1999.

MACIEL, L. Dos S.; BALLINI, R.; SILVEIRA, R. L. F. **Apreçamento de opções sobre taxa de câmbio R\$/US\$ negociadas no Brasil: uma comparação entre os modelos Black e redes neurais artificiais**. *Finanças e Contabilidade. Rev. Adm. São Paulo*, 2012. Disponível em:

<https://www.scielo.br/j/rausp/a/Bnk8RHGJ57TSKbpQ66LxdJK/?lang=pt>. Acesso em: 22 jan. 2022.

PAIVA, C. **Redes Neurais Artificias: uma introdução conceitual**. 2019. Disponível em: <https://clovesadriano.medium.com/redes-neurais-artificias-uma-introdu%C3%A7%C3%A3o-conceitual-8216eab8658d>. Acesso em: 08 jan. 2022.

PONTES, Fabrício José et al. **Design of experiments and focused grid search for neural network parameter optimization**. *Neurocomputing*, v. 186, p. 22-34, 2016.

POWER, D. J. **Decision support systems: a historical overview**. In: Handbook on decision support systems 1. Springer, Berlin, Heidelberg, 2008. p. 121-140.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning representations by back-propagating errors**. *Nature*, Nature Publishing Group, v. 323, n. 6088, p. 533-536, 1986. 11

SAXENA, S. **Qual é a diferença entre RMSE e RMSLE?** 2019. Disponível em: <https://medium.com/analytics-vidhya/root-mean-square-log-error-rmse-vs-rmlse-935c6cc1802a>. Acesso em: 17 dez. 2021.

SL. HO, M. XIE. **The use of ARIMA models for reliability forecasting and analysis**. *Computers & Industrial Engineering*, v. 35, p. 213-216, 1998.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. Biblioteca Central.

Normalização de referências: NBR 6023:2002 / Universidade Federal do Espírito Santo, Biblioteca Central. - Vitória, ES: EDUFES, 2015.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. Biblioteca Central.

Normalização e apresentação de trabalhos científicos e acadêmicos / Universidade Federal do Espírito Santo, Biblioteca Central. - 2. ed. - Vitória, ES: EDUFES, 2015.

ZINN, W.; LIU, P. C. **A comparison of actual and intended consumer behavior in response to retail stockouts**. *Journal of Business Logistics*, v. 29, n. 2, p. 141-159, 2008.