



Pós-Graduação em Ciência da Computação

**Edvan Joaquim Soares Junior**

**Modelos de Redes Neurais Recorrentes para Previsão de Demanda de Curto Prazo**



Universidade Federal de Pernambuco  
posgraduacao@cin.ufpe.br  
<http://cin.ufpe.br/~posgraduacao>

Recife  
2020

**Edvan Joaquim Soares Junior**

**Modelos de Redes Neurais Recorrentes para Previsão de Demanda de Curto Prazo**

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

**Área de Concentração:** Inteligência Computacional.

**Orientador:** Prof. Dr. Germano Crispim Vasconcelos

Recife  
2020

Catálogo na fonte  
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S676m Soares Júnior, Edvan Joaquim  
Modelos de redes neurais recorrentes para previsão de demanda de curto  
prazo / Edvan Joaquim Soares Júnior. – 2020.  
86 f.: il., fig., tab.

Orientador: Germano Crispim Vasconcelos.  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn,  
Ciência da Computação, Recife, 2020.  
Inclui referências.

1. Inteligência computacional. 2. Séries temporais. I. Vasconcelos, Germano  
Crispim (orientador). II. Título.

006.31

CDD (23. ed.)

UFPE - CCEN 2020 - 96

**Edvan Joaquim Soares Junior**

**“Modelos de Redes Neurais Recorrentes para Previsão de Demanda de Curto Prazo”**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 12/03/2020.

**BANCA EXAMINADORA**

---

Profa. Dra. Patrícia Cabral de Azevedo Restelli Tedesco  
Centro de Informática/UFPE

---

Prof. Dr. Cleviton Vinicius Fonsêca Monteiro  
Departamento de Estatística e Informática/UFRPE

---

Prof. Dr. Germano Crispim Vasconcelos  
Centro de Informática/UFPE  
**(Orientador)**

*Dedico este trabalho à Deus, minha esposa e minha família que são a base daquilo que sou.*

## **AGRADECIMENTOS**

Agradeço a Deus por todas as maravilhas e bênçãos que Ele tem me proporcionado, pois sem a sua misericórdia nada teria sido possível. Agradeço a minha família por toda a educação e suporte que me foi dado durante toda a minha vida, sendo a minha base e meu escudo. Agradeço a minha esposa Elionai por toda a sua compreensão e apoio durante todo esse tempo, sendo peça fundamental para que eu conseguisse chegar até aqui. Agradeço a todos os amigos que fiz no decorrer dessa jornada, em especial ao meu amigo Jorge Delgado que mais uma vez me acompanha nessa nova etapa. Agradeço aos professores que foram fundamentais na minha formação como profissional e como pessoa, onde com seus ensinamentos me fizeram ter uma visão mais abrangente do mundo. Agradeço ao professor Germano por toda a sua consideração, dedicação e paciência nesses últimos dois anos, se mostrando sempre ser uma pessoa solícita e humana. Enfim, agradeço a Universidade Federal de Pernambuco pela oportunidade de estudar nessa maravilhosa instituição que tanto me orgulha e pela oportunidade de conhecer pessoas maravilhosas.

## RESUMO

A previsão de demanda consiste na atividade de gerar informações que sejam capazes de auxiliar gestores a tomarem decisões mais assertivas através de uma projeção do futuro. As técnicas de previsão de demanda basicamente se dividem entre técnicas qualitativas e quantitativas. No que diz respeito às técnicas quantitativas, a previsão de demanda utilizando séries temporais é algo bastante utilizado, onde o uso de modelos estatísticos tal como o *AutoRegressive Integrated Moving Average (ARIMA)* já são largamente explorados para essa finalidade. Acontece que esses modelos possuem algumas limitações e acabam não sendo suficientemente eficientes para algumas situações. Com a popularização da inteligência artificial e dos algoritmos de aprendizagem profunda algumas alternativas passaram a ser utilizadas na tentativa de contornar essas limitações. Dentre as classes de modelos que vêm apresentando bons resultados para este fim estão os modelos de *Recurrent Neural Networks (RNN)*, mais especificamente o *Long Short-Term Memory (LSTM)*. Nesse sentido, esse trabalho apresenta um conjunto de diferentes abordagens e resultados obtidos com a utilização do LSTM e do modelo híbrido CNN-LSTM para realizar previsões de demanda, onde para isso foram utilizados cinco *datasets* reais de uma grande empresa de bebidas do mercado nacional, que posteriormente tiveram seus resultados comparados com os resultados obtidos por modelos ARIMA e pelos modelos em operação da empresa fornecedora dos produtos. Com os resultados obtidos foi possível constatar que a utilização dos modelos baseados em LSTM's são capazes de produzir resultados satisfatórios, apresentando na maioria dos casos avaliados por esta pesquisa resultados superiores, onde em comparação com o modelo ARIMA se obteve reduções de até 75,53% nas taxas de erro, e em comparação com o modelo desenvolvido pela empresa fornecedora dos produtos essas reduções chegaram a até 84,87%.

**Palavras-chaves:** Previsão de Demanda. Séries Temporais. LSTM. CNN. ARIMA.

## ABSTRACT

Demand forecasting consists of the activity of generating information that is capable of helping managers to make more assertive decisions through a projection of the future. Demand forecasting techniques are basically divided between qualitative and quantitative techniques. With regard to quantitative techniques, demand forecasting using time series is widely used, where the use of statistical models such as AutoRegressive Integrated Moving Average (ARIMA) are already widely explored for this purpose. It turns out these models have some limitations and end up not being efficient enough for some situations. With the popularization of artificial intelligence and deep learning algorithms, some alternatives started to be used in an attempt to overcome these limitations. Among the classes of algorithms that have been showing good results for this purpose are the Recurrent Neural Networks (RNN), more specifically Long Short-Term Memory (LSTM). In this sense, this work presents a set of different approaches and results obtained with the use of LSTM and the hybrid model CNN-LSTM to make demand forecasts, where for this purpose were used five real datasets from a large beverage company in the national market, which subsequently had their results compared with the results obtained by ARIMA models and by the models in operation of the company supplying the products. The results obtained showed that the use of LSTM-based models are able to produce satisfactory results, presenting in most cases evaluated by this research higher results, where in comparison with the ARIMA model reductions of up to 75.53% in error rates were obtained, and in comparison with the model developed by the supplier of products these reductions reached 84.87%.

**Keywords:** Demand Forecasting. Time Series. LSTM. CNN. ARIMA.

## LISTA DE FIGURAS

Figura 1 – Planejamento estratégico, tático e operacional. . . . .	20
Figura 2 – Técnicas de previsão de demanda. . . . .	21
Figura 3 – Passos para previsão de demanda. . . . .	21
Figura 4 – Exemplo de série temporal. Reservas mensais de uma companhia aérea. . . . .	23
Figura 5 – Linha de tendência de uma série temporal. . . . .	24
Figura 6 – Pontos de sazonalidade de uma série temporal. . . . .	25
Figura 7 – Resíduos de uma série temporal. A imagem da esquerda apresenta uma série temporal em sua completude, e a imagem da direita mostra os resíduos extraídos desta mesma série temporal. . . . .	26
Figura 8 – Série estacionária e não estacionária. . . . .	29
Figura 9 – Gráficos da ACF e PACF. . . . .	29
Figura 10 – RNA simples com pesos e função de ativação. . . . .	31
Figura 11 – Operação de convolução. . . . .	32
Figura 12 – Operação de <i>max-pooling</i> . . . . .	32
Figura 13 – Arquitetura LSTM. . . . .	34
Figura 14 – CNN-LSTM . . . . .	36
Figura 15 – Comparação de performance de modelos. . . . .	38
Figura 16 – Comparação de performance de modelos ARIMA e LSTM. . . . .	39
Figura 17 – Comparação de performance de modelos LSTM, ARIMA e SVM. . . . .	40
Figura 18 – Comparação de performance de modelos LSTM e CNN-LSTM. . . . .	40
Figura 19 – Comparação de performance de modelos LSTM, GRU, CNN-LSTM e outros. . . . .	41
Figura 20 – Decomposição dos componentes do Produto 1. . . . .	44
Figura 21 – Decomposição dos componentes do Produto 2. . . . .	44
Figura 22 – Decomposição dos componentes do Produto 3. . . . .	44
Figura 23 – Decomposição dos componentes do Produto 4. . . . .	44
Figura 24 – Decomposição dos componentes do Produto 5. . . . .	45
Figura 25 – ACF do Produto 1. . . . .	46
Figura 26 – PACF do Produto 1. . . . .	46
Figura 27 – ACF do Produto 2. . . . .	46
Figura 28 – PACF do Produto 2. . . . .	46
Figura 29 – ACF do Produto 3. . . . .	47
Figura 30 – PACF do Produto 3. . . . .	47
Figura 31 – ACF do Produto 4. . . . .	47
Figura 32 – PACF do Produto 4. . . . .	47
Figura 33 – ACF do Produto 5. . . . .	48

Figura 34 – PACF do Produto 5. . . . .	48
Figura 35 – Aplicação do One-Hot Encode para criação de variáveis temporais . . .	50
Figura 36 – Predição com ARIMA e <i>Modelo em Operação</i> do Produto 1. . . . .	52
Figura 37 – Predição com ARIMA e <i>Modelo em Operação</i> do Produto 2. . . . .	52
Figura 38 – Predição com ARIMA e <i>Modelo em Operação</i> do Produto 3. . . . .	53
Figura 39 – Predição com ARIMA e <i>Modelo em Operação</i> do Produto 4. . . . .	53
Figura 40 – Predição com ARIMA e <i>Modelo em Operação</i> do Produto 5. . . . .	53
Figura 41 – Predição com o modelo LSTM do Produto 1. . . . .	55
Figura 42 – Predição com o modelo LSTM do Produto 2. . . . .	55
Figura 43 – Predição com o modelo LSTM do Produto 3. . . . .	55
Figura 44 – Predição com o modelo LSTM do Produto 4. . . . .	55
Figura 45 – Predição com o modelo LSTM do Produto 5. . . . .	56
Figura 46 – Predição com o modelo LSTM do Produto 1 (sem tendência). . . . .	57
Figura 47 – Predição com o modelo LSTM do Produto 2 (sem tendência). . . . .	57
Figura 48 – Predição com o modelo LSTM do Produto 3 (sem tendência). . . . .	58
Figura 49 – Predição com o modelo LSTM do Produto 4 (sem tendência). . . . .	58
Figura 50 – Predição com o modelo LSTM do Produto 5 (sem tendência). . . . .	58
Figura 51 – Predição com o modelo LSTM do Produto 1 (sem sazonalidade). . . .	59
Figura 52 – Predição com o modelo LSTM do Produto 2 (sem sazonalidade). . . .	59
Figura 53 – Predição com o modelo LSTM do Produto 3 (sem sazonalidade). . . .	59
Figura 54 – Predição com o modelo LSTM do Produto 4 (sem sazonalidade). . . .	59
Figura 55 – Predição com o modelo LSTM do Produto 5 (sem sazonalidade). . . .	60
Figura 56 – Predição com o modelo LSTM do Produto 1 (variáveis temporais). . .	61
Figura 57 – Predição com o modelo LSTM do Produto 2 (variáveis temporais). . .	61
Figura 58 – Predição com o modelo LSTM do Produto 3 (variáveis temporais). . .	62
Figura 59 – Predição com o modelo LSTM do Produto 4 (variáveis temporais). . .	62
Figura 60 – Predição com o modelo LSTM do Produto 5 (variáveis temporais). . .	62
Figura 61 – Predição com o modelo LSTM do Produto 1 (variáveis externas). . . .	64
Figura 62 – Predição com o modelo LSTM do Produto 2 (variáveis externas). . . .	64
Figura 63 – Predição com o modelo LSTM do Produto 3 (variáveis externas). . . .	64
Figura 64 – Predição com o modelo LSTM do Produto 4 (variáveis externas). . . .	64
Figura 65 – Predição com o modelo LSTM do Produto 5 (variáveis externas). . . .	65
Figura 66 – Predição com o modelo CNN-LSTM do Produto 1. . . . .	67
Figura 67 – Predição com o modelo CNN-LSTM do Produto 2. . . . .	67
Figura 68 – Predição com o modelo CNN-LSTM do Produto 3. . . . .	67
Figura 69 – Predição com o modelo CNN-LSTM do Produto 4. . . . .	67
Figura 70 – Predição com o modelo CNN-LSTM do Produto 5. . . . .	68
Figura 71 – Predição com o modelo CNN-LSTM do Produto 1 (sem tendência). . .	69
Figura 72 – Predição com o modelo CNN-LSTM do Produto 2 (sem tendência). . .	69

Figura 73 – Predição com o modelo CNN-LSTM do Produto 3 (sem tendência). . .	69
Figura 74 – Predição com o modelo CNN-LSTM do Produto 4 (sem tendência). . .	69
Figura 75 – Predição com o modelo CNN-LSTM do Produto 5 (sem tendência). . .	70
Figura 76 – Predição com o modelo CNN-LSTM do Produto 1 (sem sazonalidade).	70
Figura 77 – Predição com o modelo CNN-LSTM do Produto 2 (sem sazonalidade).	70
Figura 78 – Predição com o modelo CNN-LSTM do Produto 3 (sem sazonalidade).	71
Figura 79 – Predição com o modelo CNN-LSTM do Produto 4 (sem sazonalidade).	71
Figura 80 – Predição com o modelo CNN-LSTM do Produto 5 (sem sazonalidade).	71
Figura 81 – Predição com o modelo CNN-LSTM do Produto 1 (variáveis temporais).	72
Figura 82 – Predição com o modelo CNN-LSTM do Produto 2 (variáveis temporais).	72
Figura 83 – Predição com o modelo CNN-LSTM do Produto 3 (variáveis temporais).	73
Figura 84 – Predição com o modelo CNN-LSTM do Produto 4 (variáveis temporais).	73
Figura 85 – Predição com o modelo CNN-LSTM do Produto 5 (variáveis temporais).	73
Figura 86 – Predição com o modelo CNN-LSTM do Produto 1 (variáveis externas).	74
Figura 87 – Predição com o modelo CNN-LSTM do Produto 2 (variáveis externas).	74
Figura 88 – Predição com o modelo CNN-LSTM do Produto 3 (variáveis externas).	75
Figura 89 – Predição com o modelo CNN-LSTM do Produto 4 (variáveis externas).	75
Figura 90 – Predição com o modelo CNN-LSTM do Produto 5 (variáveis externas).	75
Figura 91 – Melhores resultados do Produto 1. . . . .	77
Figura 92 – Melhores resultados do Produto 2. . . . .	77
Figura 93 – Melhores resultados do Produto 3. . . . .	77
Figura 94 – Melhores resultados do Produto 4. . . . .	77
Figura 95 – Melhores resultados do Produto 5. . . . .	78

## LISTA DE TABELAS

Tabela 1 – Quantidade de artigos encontrados por modelo . . . . .	42
Tabela 2 – Resultados dos testes Augmented Dickey–Fuller (ADF) . . . . .	49
Tabela 3 – Resultados das previsões com ARIMA e com o Modelo em Operação .	53
Tabela 4 – Resultados das previsões com LSTM . . . . .	56
Tabela 5 – Resultados das previsões do LSTM com a remoção da tendência . . . .	58
Tabela 6 – Resultados das previsões do LSTM com a remoção da sazonalidade . .	60
Tabela 7 – Resultados das previsões do LSTM com a adição de variáveis temporais	62
Tabela 8 – Resultados das previsões do LSTM com a adição de variáveis externas	65
Tabela 9 – Resultados das previsões com CNN-LSTM . . . . .	68
Tabela 10 – Resultados das previsões do CNN-LSTM com a remoção da tendência	70
Tabela 11 – Resultados das previsões do CNN-LSTM com a remoção da sazonalidade	71
Tabela 12 – Resultados das previsões do CNN-LSTM com a adição de variáveis temporais . . . . .	73
Tabela 13 – Resultados das previsões do CNN-LSTM com a adição de variáveis externas . . . . .	75
Tabela 14 – Comparação dos resultados das previsões . . . . .	78
Tabela 15 – Resultados Comparação Produto 1 . . . . .	79
Tabela 16 – Resultados Comparação Produto 2 . . . . .	79
Tabela 17 – Resultados Comparação Produto 3 . . . . .	79
Tabela 18 – Resultados Comparação Produto 4 . . . . .	79
Tabela 19 – Resultados Comparação Produto 5 . . . . .	80

## LISTA DE ABREVIATURAS E SIGLAS

<b>ACF</b>	Função de Autocorrelação
<b>ADF</b>	Augmented Dickey–Fuller
<b>AIC</b>	Akaike Information Criterion
<b>ARIMA</b>	AutoRegressive Integrated Moving Average
<b>CNN</b>	Convolutional Neural Networks
<b>GAN</b>	Redes Adversárias Generativas
<b>GDE</b>	Gradiente Descendente Estocástico
<b>INMET</b>	Instituto Nacional de Meteorologia
<b>LSTM</b>	Long Short-Term Memory
<b>MAE</b>	Mean Absolute Error
<b>MAPE</b>	Mean Absolute Percentage Error
<b>PACF</b>	Função de Autocorrelação Parcial
<b>ReLU</b>	Rectified Linear Unit
<b>RMSE</b>	Root Mean Squared Error
<b>RNA</b>	Redes Neurais Artificiais
<b>RNN</b>	Recurrent Neural Networks

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
1.1	MOTIVAÇÃO	15
1.2	CARACTERIZAÇÃO DO PROBLEMA	16
1.3	OBJETIVOS	17
<b>1.3.1</b>	<b>Objetivo principal</b>	<b>17</b>
<b>1.3.2</b>	<b>Objetivos específicos</b>	<b>17</b>
1.4	ESTRUTURA DA DISSERTAÇÃO	17
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>19</b>
2.1	PREVISÃO DE DEMANDA	19
2.2	SÉRIES TEMPORAIS	23
<b>2.2.1</b>	<b>Componentes de uma série temporal</b>	<b>24</b>
2.2.1.1	Tendência	24
2.2.1.2	Ciclo e Sazonalidade	24
2.2.1.3	Resíduos	25
<b>2.2.2</b>	<b>Previsão de séries temporais</b>	<b>26</b>
2.3	ARIMA	26
<b>2.3.1</b>	<b>Processo de diferenciação</b>	<b>27</b>
<b>2.3.2</b>	<b>Modelo de média móvel</b>	<b>27</b>
<b>2.3.3</b>	<b>Modelo autorregressivo</b>	<b>28</b>
<b>2.3.4</b>	<b>Estacionariedade</b>	<b>28</b>
2.4	APRENDIZAGEM PROFUNDA	30
<b>2.4.1</b>	<b>CNN</b>	<b>31</b>
<b>2.4.2</b>	<b>LSTM</b>	<b>33</b>
2.4.2.1	Portão de entrada	34
2.4.2.2	Portão de esquecimento	35
2.4.2.3	Portão de saída	35
<b>2.4.3</b>	<b>CNN-LSTM</b>	<b>35</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>37</b>
3.1	PREVISÃO DE DEMANDA COM ARIMA	37
3.2	PREVISÃO DE DEMANDA COM LSTM	38
3.3	PREVISÃO DE DEMANDA COM CNN-LSTM	40
3.4	CONSIDERAÇÕES	41
<b>4</b>	<b>EXPERIMENTOS REALIZADOS</b>	<b>43</b>

4.1	ANÁLISE DOS DADOS . . . . .	43
<b>4.1.1</b>	<b>Decomposição da série . . . . .</b>	<b>43</b>
<b>4.1.2</b>	<b>Análise de estacionariedade . . . . .</b>	<b>45</b>
4.2	PRÉ-PROCESSAMENTO DOS DADOS . . . . .	49
4.3	MÉTRICAS DE ERRO . . . . .	50
<b>4.3.1</b>	<b>MAE . . . . .</b>	<b>51</b>
<b>4.3.2</b>	<b>MAPE . . . . .</b>	<b>51</b>
<b>4.3.3</b>	<b>RMSE . . . . .</b>	<b>51</b>
4.4	MODELAGEM COM ARIMA . . . . .	51
4.5	MODELAGEM COM LSTM . . . . .	54
<b>4.5.1</b>	<b>Arquitetura do LSTM . . . . .</b>	<b>54</b>
<b>4.5.2</b>	<b>Resultados com LSTM . . . . .</b>	<b>54</b>
4.5.2.1	Remoção da tendência e sazonalidade . . . . .	57
4.5.2.2	Adição de variáveis temporais . . . . .	61
4.5.2.3	Adição de variáveis externas . . . . .	63
4.6	MODELAGEM COM CNN-LSTM . . . . .	66
<b>4.6.1</b>	<b>Arquitetura do CNN-LSTM . . . . .</b>	<b>66</b>
<b>4.6.2</b>	<b>Resultados com CNN-LSTM . . . . .</b>	<b>66</b>
4.6.2.1	Remoção da tendência e sazonalidade . . . . .	68
4.6.2.2	Adição de variáveis temporais . . . . .	72
4.6.2.3	Adição de variáveis externas . . . . .	74
4.7	RESUMO DOS RESULTADOS . . . . .	76
<b>4.7.1</b>	<b>Comparações de resultados extra . . . . .</b>	<b>78</b>
<b>5</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS . . . . .</b>	<b>81</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>83</b>

# 1 INTRODUÇÃO

O objetivo deste capítulo é trazer uma visão geral do trabalho desenvolvido nesta pesquisa. Sendo assim, a Seção 1.1 descreve a motivação que se deu para o desenvolvimento desta pesquisa, na Seção 1.2 será descrito a caracterização do problema, na Seção 1.3 a apresentação dos objetivos, e por fim a Seção 1.4 que descreve a estruturação desta dissertação.

## 1.1 MOTIVAÇÃO

Manter uma empresa funcionando requer um conjunto de estratégias e decisões que muitas vezes precisam ser executadas de forma rápida e precisa. No último estudo realizado pelo SEBRAE sobre a sobrevivência das empresas (SEBRAE, 2016), foi constatado que cerca de 24% das empresas brasileiras não se mantêm ativas por mais de dois anos, chegando à mortalidade. Esse mesmo estudo indica que os principais fatores que podem incidir na sobrevivência ou mortalidade das empresas são: situação do empresário antes da abertura da empresa (tipo de ocupação anterior, experiência no ramo, motivação para abrir o negócio), planejamento do negócio, gestão do negócio e capacitação dos donos em gestão empresarial. Uma das principais falhas cometidas pelos gestores é o mal gerenciamento do fluxo de caixa, que muitas vezes tomam decisões de forma empírica e se preocupam apenas com soluções de curto prazo (MENDONÇA et al., 2017). Para que essas empresas continuem em funcionamento se faz necessário planejamento e um maior controle de contas, estoques e patrimônio (PINHEIRO; NETO, 2019).

Levando em consideração os fatores supracitados, algumas técnicas são capazes de auxiliar o gestor no planejamento e gestão do negócio, técnicas essas que associadas a expertise do gestor podem ajudar na não mortalidade da empresa. Uma das técnicas que podem auxiliar neste sentido é a previsão de demanda. A previsão de demanda traz ao gestor uma projeção daquilo que há por vir, não revelando valores com exatidão, mas trazendo valores aproximados que podem auxiliar nas tomadas de decisões (PEREIRA, 2017). Através da utilização de técnicas de previsão de demanda é possível obter subsídios que auxiliarão no planejamento dos níveis de produção, distribuição, estocagem, fluxo de caixa, contratação de pessoal, entre outros (BALLOU, 2009).

Para essa finalidade se faz necessário a utilização de dados passados, para que através deles seja possível encontrar padrões que possam projetar o futuro. Nesse sentido, um modelo de dados bastante utilizado para este fim são as séries temporais, que a literatura descreve como sendo um conjunto de observações, cada uma sendo gravada em um tempo específico (BROCKWELL; DAVIS, 2016).

No que diz respeito à previsão de séries temporais um dos modelos mais utilizados

---

é o *AutoRegressive Integrated Moving Average (ARIMA)* (WANG; LOU, 2019). Porém, com a popularização dos algoritmos de aprendizagem profunda outros tipos de modelos também passaram a ser utilizados para esta finalidade, obtendo resultados relevantes, como por exemplo a utilização de *Convolutional Neural Networks (CNN)* e *Recurrent Neural Networks (RNN)* (SELVIN et al., 2017).

Nesse sentido, a utilização de algoritmos de aprendizagem profunda pode se mostrar capaz de apresentar bons resultados no que diz respeito à previsão de demandas, dado a sua capacidade de lidar melhor com dados não lineares em comparação com modelos ARIMA que lidam melhor com dados lineares (SIAMI-NAMINI; NAMIN, 2018). Para se chegar ao objetivo dessa pesquisa redes Long Short-Term Memory (LSTM)s e CNN-LSTMs foram treinadas utilizando dados reais de uma grande empresa do ramo de bebidas, onde posteriormente tiveram seus resultados comparados com os de outros modelos, observando assim se houve ganhos de resultados com sua utilização. Após essa comparação, espera-se de forma empírica identificar quais técnicas e procedimentos apresentam melhores resultados e quais delas poderiam ser melhores aplicadas para o problema de previsão de demanda.

## 1.2 CARACTERIZAÇÃO DO PROBLEMA

É sabido os benefícios que a utilização de técnicas de previsão de demanda podem trazer para os gestores no auxílio à tomadas de decisões. Contudo, a escolha da técnica correta é fundamental para tornar esses subsídios assertivos de modo a mitigar os erros. Se tratando de conjuntos de dados em forma de séries temporais, que será o abordado dentro dessa pesquisa, o modelo ARIMA é um dos mais populares para esta finalidade. O ARIMA é um modelo estatístico que assume que os valores de uma série temporal são altamente dependentes, onde cada valor pode ser explicado pelos valores que os antecedem (PELLLEGRINI, 2000). Todavia, uma das principais desvantagens de utilização desse modelo é sua dificuldade de lidar com problemas estocásticos e não-lineares (FU; ZHANG; LI, 2016), problemas esses que não são incomuns de serem encontrados.

Com o aumento da capacidade computacional dos atuais computadores e com a evolução dos algoritmos de Redes Neurais Artificiais (RNA) outras possibilidades surgiram para lidar com este tipo de problema. Recentemente, modelos de Redes Neurais Profundas produziram resultados expressivos em várias aplicações, sendo as principais responsáveis pelos avanços significativos na área de inteligência artificial e aprendizagem de máquina. Entre as redes profundas, modelos baseados em estruturas recorrentes, com capacidade natural de memorizar informações temporais, são importantes candidatos a serem investigados para o processamento de séries temporais. Entretanto, esses modelos ainda requerem investigação e comparação com métodos tradicionais em problemas reais para estabelecer sua capacidade de previsão.

No contexto de previsão de séries temporais, um dos modelos baseados em estruturas recorrentes que tem ganhado visibilidade mostrando bons resultados é o *LSTM* (FU; ZHANG; LI, 2016; SELVIN et al., 2017). O LSTM é um modelo de RNN que possui como principal característica a capacidade de memorizar valores passados a fim de otimizar a compreensão dos valores futuros (HOCHREITER; SCHMIDHUBER, 1997). Ao contrário do ARIMA, LSTM's são capazes de lidar com conjuntos de dados não lineares.

Vale salientar que a escolha de um modelo que resolva um determinado tipo de problema não é algo trivial. Embora o LSTM apresente alguns resultados satisfatórios para alguns tipos de problemas, não é possível afirmar que ele se sobressai sobre o modelo ARIMA para todos os casos. Explorar esses modelos de diferentes formas se faz necessário para identificação das abordagens que venham trazer melhores resultados.

### 1.3 OBJETIVOS

#### 1.3.1 Objetivo principal

Esta pesquisa tem como objetivo principal investigar modelos recorrentes de aprendizagem profunda, particularmente LSTMs e CNN-LSTMs, no problema de previsão de demanda de curto prazo e compará-los com modelos já consolidados de previsão de séries temporais com um conjunto de métricas relevantes e dados reais.

#### 1.3.2 Objetivos específicos

- Investigar os processos de tratamento de séries temporais de maneira sistemática, como sazonalidade, tendências e transformações na série para realizar o tratamento de dados capazes de melhorar os resultados das previsões;
- Comparar a performance de modelos baseados em LSTMs com modelos ARIMA tradicionais e uma solução implantada por uma grande empresa distribuidora de bebidas, identificando as abordagens que possibilitaram maior minimização do erro de previsão; e,
- Investigar a potencial influência de variáveis externas, obtidas do mercado, no desempenho de previsão dos modelos LSTMs.

### 1.4 ESTRUTURA DA DISSERTAÇÃO

Os próximos capítulos desta dissertação estão organizados da seguinte maneira. O capítulo 2 apresenta a fundamentação teórica dos métodos utilizados no desenvolvimento desse trabalho.

O capítulo 3 mostra um resumo dos trabalhos relacionados a esta pesquisa, bem como as fontes de buscas utilizadas para encontrar tais trabalhos.

O capítulo 4 traz uma explanação dos experimentos realizados, explicando detalhadamente cada uma das abordagens utilizadas e apresentando os resultados obtidos por cada uma delas. Ao final traz uma discussão dos resultados fazendo a comparação entre eles e identificando quais abordagens foram mais ou menos satisfatórias no desenvolvimento da pesquisa.

Por fim, o capítulo 5 traz uma conclusão do que foi obtido pela pesquisa, bem como perspectivas de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem por objetivo apresentar a revisão bibliográfica dos conceitos e abordagens relacionados ao desenvolvimento desta pesquisa, descrevendo suas características e formas de aplicação. Primeiramente serão abordados os conceitos de previsão de demanda que é o tipo de problema em que esta pesquisa foi aplicada. Posteriormente serão abordados os conceitos de séries temporais, modelo de dados utilizado no desenvolvimento dos modelos de previsão. Em seguida iremos contextualizar sobre o modelo ARIMA, que servirá como base de comparação para os outros modelos, e por fim, discutiremos sobre modelos de aprendizagem profunda, discorrendo mais detalhadamente sobre os modelos LSTM e CNN-LSTM, sendo esses os explorados por este trabalho.

### 2.1 PREVISÃO DE DEMANDA

Diariamente dentro das empresas um conjunto de decisões precisam ser tomadas, sendo muitas vezes necessário tomá-las de maneira rápida e precisa. No ambiente de oscilação em que estão inseridas as corporações, onde existem diversas variáveis que podem impactar diretamente nas operações dessas, tomar decisões de forma empírica pode não ser o melhor caminho. Para auxiliar nesse processo, muitos gestores contam com um conjunto de ferramentas e análises que os ajudam a tomar decisões mais assertivas. Dentro desse conjunto de auxiliares está incluída a previsão de demanda.

Como sugerido pelo nome, a previsão de demanda consiste na geração de informações que auxiliam o gestor a entender o comportamento futuro de uma ou mais atividades da empresa baseado em comportamentos passados, ou seja, gerar projeções que tornem possível prever com um certo grau de assertividade o que irá acontecer e que tem impacto direto nessas atividades. A previsão de demanda está diretamente ligada aos setores de produção da empresa e funciona como base para o Planejamento Estratégico de Produção (TUBINO, 2000) o que faz com que ela tenha impacto direto com todas as outras áreas da organização. Segundo Corrêa e Giansi (2000), embora uma organização seja composta por várias áreas, a produção é o setor que põe em prática os seus objetivos, de onde se obtém o produto final que movimentam todas as outras funções da empresa. Através da utilização de previsão de demanda nos dados de produção, a empresa poderá obter diferencial competitivo através da observação do custo percebido pelo cliente, velocidade e confiabilidade de entrega, flexibilidade das saídas, qualidade dos produtos e serviços prestados (SOUZA et al., 2010). Logo, embora possa estar associada a produção, a previsão de demanda desenvolve papel importante em todo o planejamento da empresa que é composto pelo planejamento estratégico, tático e operacional (SOUZA et al., 2010).

Figura 1 – Planejamento estratégico, tático e operacional.

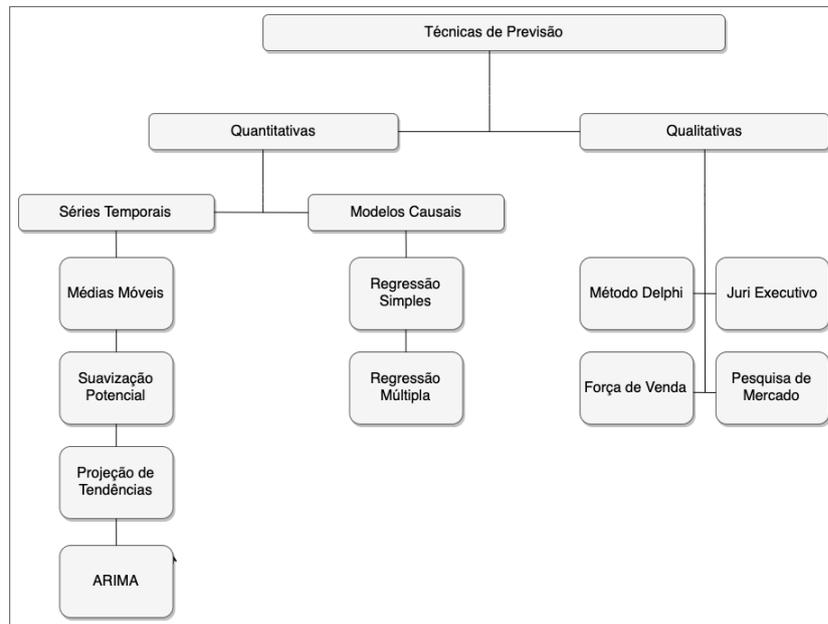


**Fonte:** <https://www.treasy.com.br/blog/planejamento-estrategico-tatico-e-operacional/>

Basicamente a previsão de demanda pode ser realizada através de três métodos: qualitativo, quantitativo e o misto que é a junção dos dois primeiros. A previsão por métodos qualitativos tem como fonte de dados a opinião e julgamento de outras pessoas uma vez que elas tenham condições de opinar, e possuem como objetivo principal transformar essas opiniões em previsões de tendências futuras (MOREIRA, 2017) (PINTO; MAZZON, 2010). Para Tubino (2000) os métodos qualitativos costumam ser mais rápidos de serem preparados, sendo mais utilizados quando não se há tempo para preparação e análise de dados passados, ou em situações de ausência desses dados.

Os métodos quantitativos, ao contrário dos método qualitativos, são desenvolvidos utilizando dados passados de demandas em busca encontrar padrões através da utilização de métodos estatísticos. Os dados utilizados com esse tipo de técnica podem ser classificados como séries temporais ou modelos causais. As séries temporais fazem a utilização de dados históricos das demandas para tentar encontrar padrões que se repitam no futuro, enquanto os modelos causais fazem a utilização de variáveis externas (e.g., PIB, clima, dados populacionais) que sejam capazes de explicar o comportamento das demandas, buscando uma correlação entre o valor da demanda e o valor dessas variáveis (PACHECO; SILVA, 2003).

Figura 2 – Técnicas de previsão de demanda.



**Fonte:** Autor

A previsão de demanda vai além da definição do método, ela requer um conjunto de pré-requisitos e processos que precisam ser bem executados para que se possa chegar ao objetivo. Se tratando de métodos quantitativos que será o método adotado por este trabalho, Wheelwright, Makridakis e Hyndman (1998) afirmam que três condições básicas são necessárias para aplicação desse tipo de método:

1. Disponibilidade de informações históricas;
2. Possibilidade da transformação das informações históricas em dados numéricos;
3. Possibilidade de repetição dos padrões observados nos dados históricos no tempo futuro.

Em seu trabalho Pellegrini e Fogliatto (2001) seis passos fundamentais para o desenvolvimento de um sistema de previsão de demanda.

Figura 3 – Passos para previsão de demanda.



**Fonte:** Adaptado de (PELLEGRINI; FOGLIATTO, 2001)

- a. *Definição do problema.* Para Pellegrini e Fogliatto (2001) essa pode ser a etapa mais complexa do processo. Nela se identifica o tipo de problema a ser solucionado,

---

definindo o tipo de previsão que se quer obter baseado nos insumos que estão disponíveis, onde para isso se faz necessário consultar outros envolvidos no processo a fim de identificar quais os tipos de dados existentes.

- b. Coleta de informações.* Essa etapa consiste na obtenção dos dados que serão utilizados. Segundo Wheelwright, Makridakis e Hyndman (1998) pelo menos dois tipos de dados devem ser utilizados, os dados estatísticos, utilizados no desenvolvimento do modelo de predição, e os dados subjetivos, que são essenciais para validação dos resultados que serão obtidos através dos modelos de predição
- c. Seleção do pacote computacional.* É comum a utilização de ferramentas computacionais para o desenvolvimento deste tipo de modelagem. Junior et al. (2010) evidencia a importância do uso de tecnologias da informação como recurso estratégico no desenvolvimento e aplicação de modelos de previsão de demanda. Sendo assim, essa etapa do processo tem por objetivo identificar quais tipos de modelos computacionais serão utilizados, buscando encontrar aquele que melhor se adequa ao tipo de problema a ser resolvido.
- d. Análise preliminar.* Essa etapa visa entender o comportamento dos dados através do agrupamento das informações e da visualização gráfica desses. O objetivo disso é tentar identificar algum tipo de comportamento anormal da série que de alguma forma venha a dificultar o processo de modelagem.
- e. Escolha e validação dos modelos.* Uma vez modelados e testados, essa etapa consiste na identificação dos modelos que obtiveram os melhores resultados. Pellegrini e Fogliatto (2001) ressaltam que a escolha do modelo deve ser baseada não apenas na precisão de predição, mas também deve se basear na sua capacidade de conhecer os aspectos que podem influenciar a demanda (e.g., promoções ou campanhas promocionais), além de conhecer as características da série temporal através das previsões de seus componentes (e.g., sazonalidade, tendências, entre outros).
- f. Verificação do sistema.* Passado todas as etapas, essa última etapa consiste na utilização do modelo desenvolvido. Nela o processo de implantação do modelo é considerado finalizado, sendo necessário apenas a realização de manutenções periódicas, que consiste na implantação de novas variáveis de interesse e revalidação do modelo.

No desenvolvimento da pesquisa foi possível observar a utilização de todos os passos definidos por Pellegrini e Fogliatto (2001), além de sua importância para que fosse possível tornar o processo mais sistematizado a fim de alcançar os resultados desejados.

Nos passos de previsão anteriormente descritos, é utilizado o modelo de dados de séries temporais para descrever o processo, sendo esse um dos modelos de dados mais utilizados para previsão de demanda, e que também foi o utilizado para o desenvolvimento

dessa pesquisa. Sendo assim, a próxima subseção tem por objetivo descrever os principais conceitos de séries temporais e suas principais características.

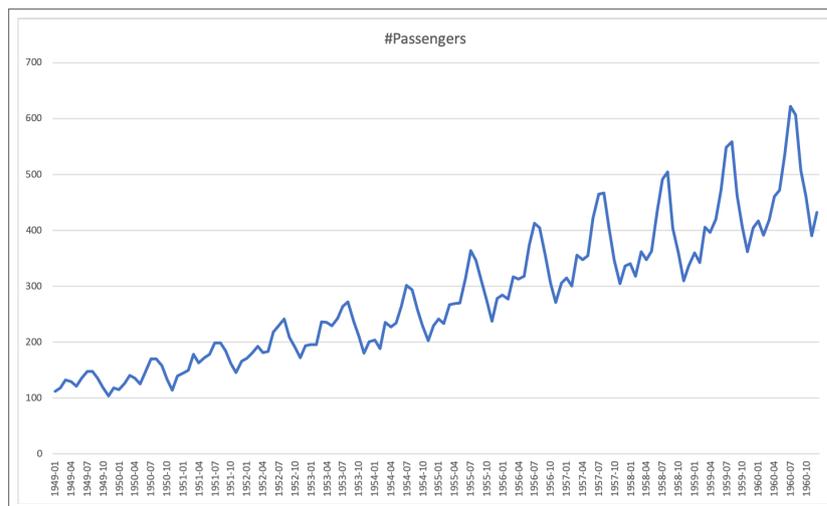
## 2.2 SÉRIES TEMPORAIS

Uma série temporal consiste em um conjunto de observações de um determinado evento. Essas observações são normalmente distribuídas em intervalos de tempo equidistantes, podendo ser representadas por valores discretos ou contínuos. Baseado nisso, uma série temporal pode ser definida como:

$$X = \{x_t \in IR | t = 1, 2, 3, 4, 5, \dots, n\} \quad (2.1)$$

onde a série temporal  $X$  é composta por um conjunto de  $n$  observações, onde cada observação  $x_t$  é distribuída sequencialmente conforme a sua ordem  $t$ .

Figura 4 – Exemplo de série temporal. Reservas mensais de uma companhia aérea.



**Fonte:** Autor

Uma série temporal é composta por quatro componentes: tendência (T), ciclo (C), sazonalidade (S) e resíduos (I). Decompor e analisar esses componentes é fundamental para entender o comportamento da série (MARTÍNEZ-ÁLVAREZ et al., 2015). Esses componentes podem ser decompostos de diversas formas, mas normalmente isso ocorre pelo modelo aditivo:

$$x_t = T_t + C_t + S_t + I_t \quad (2.2)$$

ou pelo modelo multiplicativo:

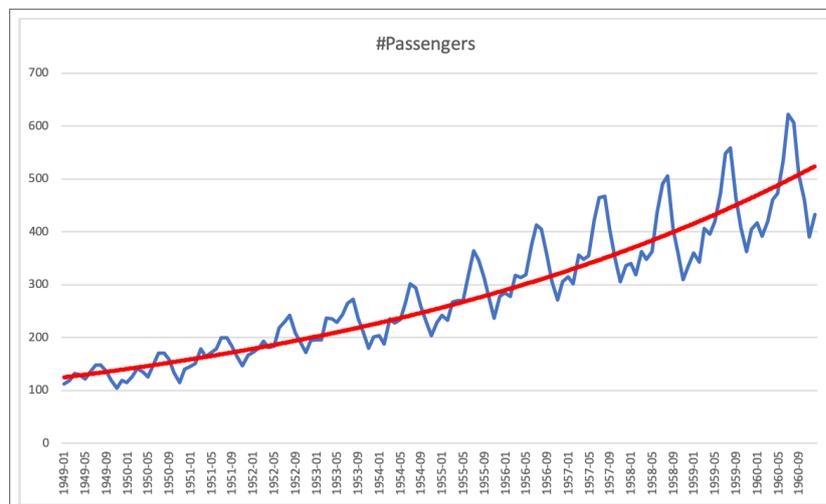
$$x_t = T_t \times C_t \times S_t \times I_t \quad (2.3)$$

## 2.2.1 Componentes de uma série temporal

### 2.2.1.1 Tendência

A tendência é responsável por retratar o comportamento de longo prazo de uma série temporal, onde através dela é possível identificar se existe uma propensão de alta ou de queda do evento observado. Segundo Reis (2015) existem três motivos principais para se identificar a tendência, sendo eles, avaliar o comportamento da série para que seja possível utilizar nas previsões, removê-la para facilitar a visualização dos outros componentes e identificar o nível da série através dos valores que ela assume caso não seja apresentado nenhum comportamento de queda ou alta. Algumas técnicas como regressão linear, médias móveis e ajuste exponencial são utilizadas para se obter a tendência se uma série (HOLT, 2004; VERBESSELT et al., 2010).

Figura 5 – Linha de tendência de uma série temporal.



Fonte: Autor

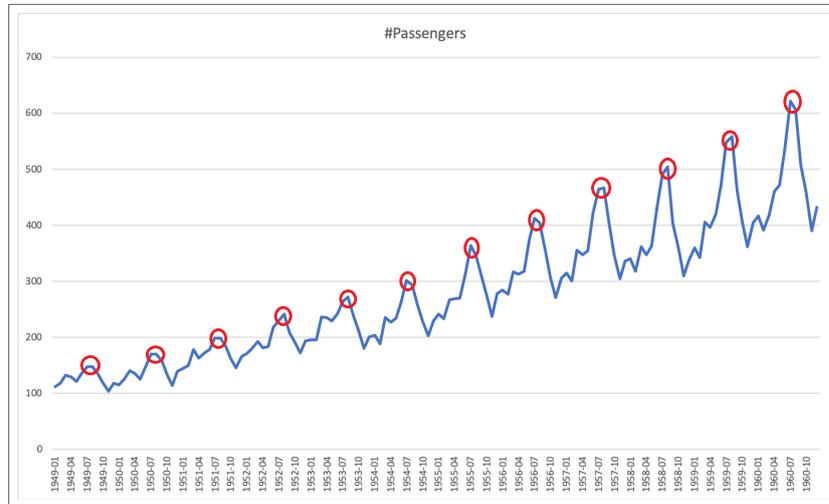
### 2.2.1.2 Ciclo e Sazonalidade

O ciclo e a sazonalidade correspondem a períodos de oscilações de altas e quedas da série temporal. A principal diferença entre esses dois componentes é regularidade do tempo em que eles são apresentados. Segundo Reis (2015) os ciclos são variações de longo prazo, com duração superior a um ano, e que não apresentam um padrão de variação regular no tempo, tornando mais difíceis de serem identificados. Reis (2015) ainda cita que embora haja uma concordância entre a comunidade científica a respeito da existência de ciclos em séries temporais, alguns autores preferem desconsiderá-los por completo, tomando apenas os outros três componentes como relevantes (MARTÍNEZ-ÁLVAREZ et al., 2015; ABBES et al., 2018; DONG; SUN, 2019).

Ao contrário dos ciclos, a sazonalidade são oscilações em intervalos de tempo menores, e que tem ocorrência em períodos regulares, podendo por exemplo, ocorrer em horas, dias

ou meses específicos.

Figura 6 – Pontos de sazonalidade de uma série temporal.



Fonte: Autor

Na Figura 6 é possível observar um conjunto de picos que se repetem regularmente dentro de um mesmo período do ano, mais especificamente em meados de julho. A análise desse componente se faz necessário para entender os eventos relacionados a variável que foi observada pela série temporal, de onde é possível identificar os coeficientes de autocorrelação dos pontos observados (KAISER; MARAVALL, 2000), importantes para modelagem da série (EHLERS, 2007).

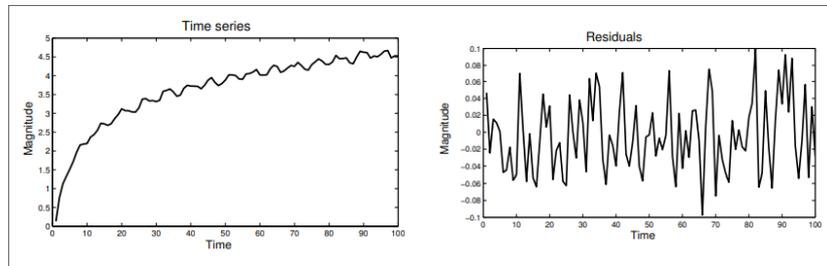
### 2.2.1.3 Resíduos

Os resíduos de uma série temporal representam os valores restantes após a remoção dos outros componentes, e é resultante de flutuações de curto prazo da série que não são sistemáticas nem previsíveis. Logo, de maneira simplificada os resíduos podem ser obtidos por:

$$I_t = x_t - T_t - C_t - S_t \quad (2.4)$$

Um dos principais objetivos da análise desse componente é identificar o grau de irregularidade da série. Uma vez que a série possui um grau de irregularidade muito alto os resíduos tendem a influenciar no movimento da série, camuflando os valores de tendência e sazonalidade, o que irá implicar em estimativas não sustentáveis (MCLAREN; ZHANG, 2010).

Figura 7 – Resíduos de uma série temporal. A imagem da esquerda apresenta uma série temporal em sua completude, e a imagem da direita mostra os resíduos extraídos desta mesma série temporal.



Fonte: Adaptado de (MARTÍNEZ-ÁLVAREZ et al., 2015)

### 2.2.2 Previsão de séries temporais

A previsão de séries temporais consiste na utilização de dados passados para prever valores futuros. De acordo com Chatfield (2000) os métodos de previsão podem ser classificados em três tipos:

- a. *Previsões de julgamento.* Esse tipo de previsão é baseado em julgamentos subjetivos, *insights*, ou qualquer outra informação quer seja julgada relevante.
- b. *Métodos univariados.* Nesse método as previsões são realizadas apenas com base nos valores de uma série única, sem a adição de qualquer outra informação.
- c. *Métodos multivariados.* Para este método a série temporal é utilizada em combinação com outras séries temporais chamadas de preditoras ou variáveis explicativas.

Para esta pesquisa foram utilizadas séries únicas, ou seja, séries capazes de serem trabalhadas com métodos de predição univariados. O modelo ARIMA é um método bastante utilizado para esse fim, e é um dos modelos que será utilizado como base dessa pesquisa. Sendo assim, a próxima seção tem por objetivo descrever as principais características desse modelo.

## 2.3 ARIMA

O ARIMA é um dos métodos mais comuns usado para a análise de séries temporais. Esse modelo especificamente pode ser dividido em três componentes distintos: autorregressão, diferenciação e média móvel. Cada um desses componentes possui uma variável que precisa ser ajustada para utilização do modelo. Esses três parâmetros correspondem a ordem do componente de autorregressão ( $p$ ), número de transformações de diferenciação ( $d$ ) e a ordem do componente de média móvel ( $q$ ).

$p$  - a ordem do componente de autorregressão é usado para realizar o ajuste de regressão linear. Devido à natureza estocástica dos modelos de séries temporais, é impossível prever

o número de elementos que formam uma localização, portanto, a escolha do valor  $p$  correto ocorre por tentativa e erro e observa-se como a linha de regressão se ajusta aos dados. Estocástico significa que os valores seguem uma distribuição de probabilidade aleatória.

$d$  - o número de transformações de diferenciação consiste no número de transformações necessárias para que a série se torne estacionária.

$q$  - a ordem do componente de média móvel se aplica ao ruído da série temporal, que não pode ser explicado por tendências ou sazonalidade. Os dados de ruído também são de natureza estocástica, portanto,  $q$  é um termo semelhante a  $p$ , mas aplicado aos resíduos da série temporal em vez das próprias observações.

Antes da introdução de modelos mais complexos, como o ARIMA, séries temporais eram analisadas sem levar em consideração a estacionariedade dos dados. No entanto, Box, Jenkins e Reinsel (1970) mostraram que um dado não estacionário poderia ser normalizado como estacionário subdividindo-o em períodos e realizando a diferenciação nas séries periódicas para descobrir uma tendência. Este é um processo fundamental na aplicação de modelos ARIMA. O reconhecimento das características dos dados a serem analisados é pertinente ao ajuste fino dos detalhes do modelo ARIMA, variando os parâmetros principais ( $p$ ,  $d$  e  $q$ ) para selecionar o mais eficaz.

### 2.3.1 Processo de diferenciação

Os modelos diferenciais transformam séries temporais não estacionárias em séries temporais estacionárias. A ordem de diferenciação mais baixa que produz flutuações constantes em torno da média (estacionária) é o valor  $d$  a ser selecionado para o modelo ARIMA. Isso pode ser determinado ajustando o modelo ARIMA com valores variáveis de 0 a  $d$ , mantendo  $p$  e  $q$  iguais a 0 até que essa característica seja observada. Quando a série temporal é sazonal, a diferenciação pode ser feita subtraindo os dados correspondentes de cada período.

### 2.3.2 Modelo de média móvel

Um modelo de média móvel (MA) baseia-se na incerteza de prever a próxima observação devido ao desconhecimento das variáveis independentes. Assumindo uma previsão de linha de base 0, os resíduos dos dados observados devem se parecer com dados ruidosos. A análise de média móvel é baseada na média ponderada de observações ruidosas anteriores (ABONAZEL; ABD-ELFTAH, 2019). Assumindo um ponto de observação  $x_t$ , a equação do modelo pode ser descrita como:

$$x_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (2.5)$$

onde  $q$  é a ordem do processo (o número de passos) e  $\theta$  são os parâmetros do processo, e  $\epsilon_t$  é o ruído branco. Determinar a ordem a partir da visualização dos dados é difícil, sendo

assim, a ordem pode ser determinada calculando a variável de autocorrelação (AC). A autocorrelação ocorre quando um valor  $x_t$  é correlacionado com um valor anterior  $x_{t-k}$ . Isso pode ser determinado encontrando a correlação de  $x_t$  nos passos 0, 1, 2, ..., até que a correlação se aproxime de 1. Essa variável de ordem é importante na combinação do modelo ARIMA.

### 2.3.3 Modelo autorregressivo

Diferentemente do modelo MA, em que consideramos os resíduos, o modelo autoregressivo (AR) calcula a média dos valores passados da série em relação a  $X_{t-p}$  (ABONAZEL; ABD-ELFATAH, 2019). O uso dos valores reais em vez dos valores de erro fornecem outra dimensão de análise. A equação do modelo pode ser descrita como:

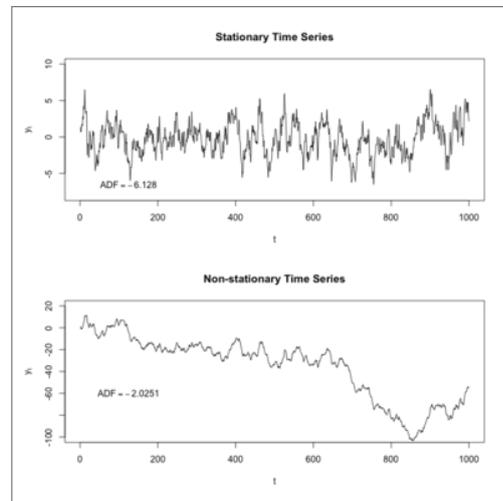
$$x_t = c + \epsilon_t + \sum_{i=1}^p \phi_i x_{t-i} \quad (2.6)$$

Onde  $p$  é a ordem do processo e  $\phi$  são os parâmetros do processo. O valor de  $p$  afeta diretamente a precisão do modelo, porque se o valor errado for escolhido, os valores discrepantes afetarão a linearidade dos valores do período. Os modelos autoregressivos são semelhantes aos modelos lineares periodizados. Ao dividir os dados em diferentes partes, as observações dos dados podem ser usadas para prever valores futuros. Variações mais complexas podem aplicar regressão logística.

### 2.3.4 Estacionariedade

Antes de usar os modelos ARIMA, a estacionariedade da série temporal precisa ser determinada. Caso contrário, transformações não lineares, como a logarítmica, deflação ou potenciação, precisam ser realizadas até que as variações aleatórias da média se tornem aparentemente simétricas. A Figura 8 mostra o exemplo de uma série estacionária e uma série não estacionária.

Figura 8 – Série estacionária e não estacionária.

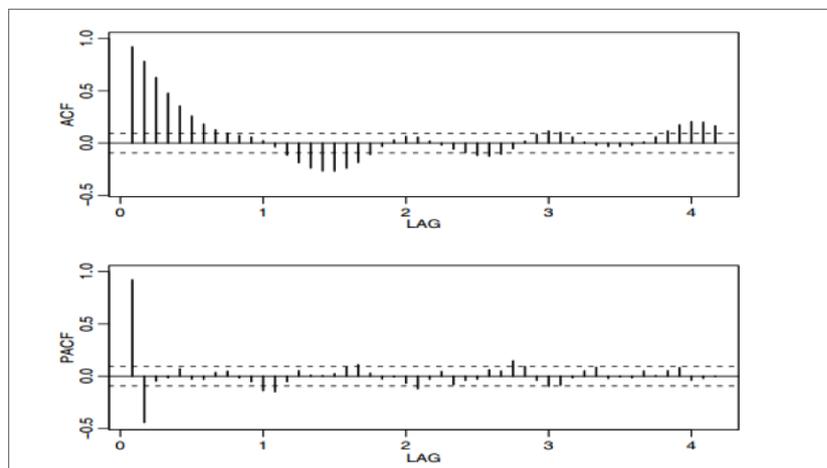


Fonte: (PALACHY, 2019)

Séries temporais são ditas estacionárias quando a média e a variância são constantes ao longo do tempo (PALACHY, 2019). Os valores das séries temporais estacionárias não dependem do tempo em que a série é observada. Às vezes, são séries que não parecem estacionárias à primeira vista, sendo essas de natureza cíclica. Uma série cíclica é aquela periodicamente estacionária, embora pareça não estacionária durante o período. É mais fácil prever e analisar séries estacionárias, e é por isso que é importante convertê-las por diferenciação. Se isso não acontecer, a diferenciação precisará ser aplicada mais uma vez, o que é conhecido como diferenciação de segunda ordem (PALACHY, 2019).

Quando uma série temporal é estacionarizada por diferenciação, a próxima etapa é determinar se os parâmetros AR ou MA são necessários para corrigir qualquer autocorrelação restante. A análise dos gráficos da Função de Autocorrelação (ACF) e da Função de Autocorrelação Parcial (PACF) revela os valores de  $p$  e  $q$  a serem usados no modelo. Exemplos de gráficos da ACF e PACF podem ser vistos na Figura 9.

Figura 9 – Gráficos da ACF e PACF.



O gráfico ACF é responsável por apresentar a correlação entre os valores da série e os intervalos entre eles. Se uma série é estacionária, o ACF cai para 0 muito rapidamente. Por outro lado, o ACF de uma série não estacionária se aproximará de 0 muito lentamente. O gráfico PACF é basicamente o mesmo que o gráfico ACF, exceto o fato dele remover o efeito em intervalos de autocorrelação mais curtos, ou seja, entre os termos  $x_t$  e  $x_{t-k}$ , onde  $k$  é o tamanho do intervalo.

Baseado na composição individual de cada modelo, a combinação dos modelos que formam o ARIMA pode ser representada pela equação:

$$x_t = c + \epsilon_t + \sum_{i=1}^p \phi_i x_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (2.7)$$

## 2.4 APRENDIZAGEM PROFUNDA

A aprendizagem profunda é o ramo do aprendizado de máquina baseada em RNAs. As RNAs usam o cérebro como fonte de inspiração para a construção de máquinas inteligentes. Consistem em várias camadas que são capazes de extrair recursos em diferentes níveis. As camadas iniciais extraem recursos de alto nível da entrada, enquanto as outras extraem recursos de nível baixo e de nível médio. Por exemplo, no caso do processamento de imagens, a primeira camada extrai formas como rostos ou objetos humanos, enquanto as camadas inferiores extraem bordas e cantos (SCHMIDHUBER, 2015).

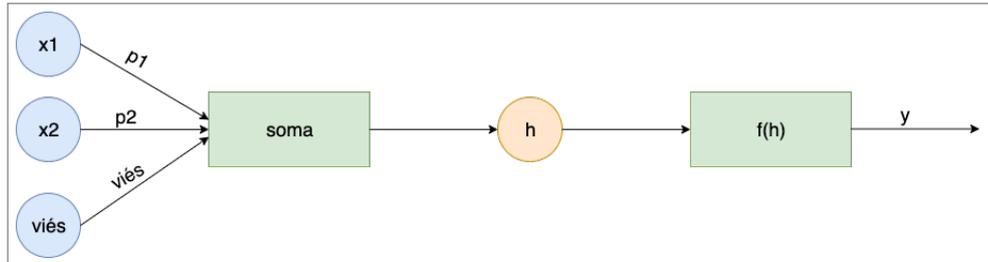
Um dos principais benefícios da aprendizagem profunda é a sua capacidade de extrair características automaticamente e sua capacidade de discriminação, especialmente nas CNNs, que extraem características através dos pesos adaptáveis do *kernel*, o que economiza o tempo de extração através do pré-processamento manual (LECUN; BENGIO; HINTON, 2015). Nesse processo o *kernel* consiste em uma matriz de pesos que é multiplicada pelas variáveis de entrada para extrair características relevantes. Esses resultados são encaminhados para o que é chamado de função de ativação, que consiste em um tipo de função que decide com base em alguma entrada multiplicada por alguma matriz de pesos (*kernel*) qual será a saída. Esses valores de saída também são chamados de "ativação". Uma função simples de ativação pode ser escrita como:

$$f(h) = \begin{cases} 0 & \text{if } h < 0 \\ 1 & \text{if } h \geq 0 \end{cases} \quad (2.8)$$

A equação acima retorna 0 se a combinação for menor que 0 e 1 se a combinação for maior ou igual a 1. Os dados de entrada são multiplicados primeiro com alguns pesos e, em seguida, o produto é passado para a função de ativação para obter os valores de saída. Os pesos são inicializados com valores aleatórios, mas conforme a rede aprende, os valores dos pesos vão sendo alterados de acordo com os valores do gradiente. O ajuste desses valores é baseado em alguns algoritmos de otimização, como o Gradiente Descendente

Estocástico (GDE) (BOTTOU, 2012) e muitos outros. Temos dois tipos de parâmetros que podem ser aprendidos no treinamento: pesos e vetores de viés, que são usados para ajustar a soma dos pesos das entradas na saída. Um exemplo de uma RNA simples com duas entradas e um viés é mostrada na Figura 10.

Figura 10 – RNA simples com pesos e função de ativação.



Fonte: Autor

### 2.4.1 CNN

O CNN é um tipo de *perceptron* de múltiplas camadas que são redes totalmente conectadas. O CNN funciona de forma que os padrões de conectividade entre diferentes neurônios se assemelham à organização do córtex visual. A principal vantagem do CNN sobre as técnicas tradicionais de aprendizado de máquina é a independência do esforço humano no processo de extração de características. O CNN emprega o nome de convolução, uma operação matemática. No CNN, usamos convolução no lugar da multiplicação da matriz.

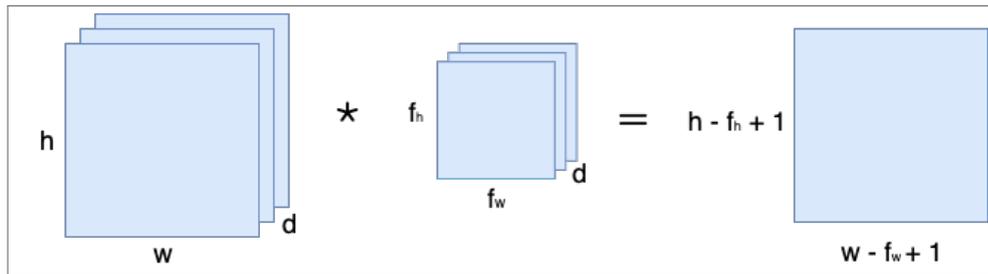
Atualmente existem diferentes arquiteturas de CNN, porém uma das arquiteturas que ganhou destaque foi a *AlexNet*. Tal arquitetura teve reconhecimento em 2012 no *ImageNet* - um desafio anual que consiste em 15 milhões de imagens com mais de 22000 categorias - onde conseguiu reduzir a taxa de erro de 0,4 para 0,1, sendo um grande feito nos campos de visão computacional e aprendizado de máquina. Tal arquitetura consiste em cinco camadas convolucionais e três camadas totalmente conectadas (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A partir de então o uso de arquiteturas baseadas em aprendizagem profunda passaram a ser mais exploradas em busca de melhores resultados na resolução de diversos tipos de problema.

Esse tipo de arquitetura é um modelo discriminativo que identifica o limite de decisão entre diferentes classes. Visto como modelos generativos, como as Redes Adversárias Generativas (GAN), que aprendem a probabilidade conjunta entre diferentes variáveis  $x$  e  $y$  (GOODFELLOW et al., 2014). A operação de convolução extrai as características por convolução de um filtro ou *kernel* na matriz de uma imagem, por exemplo. Isso consiste em:

- Uma matriz de  $(h \times d \times c)$  dimensões onde:
  - $h$ =altura

- $w$ =largura
- $c$ =canais
- Um filtro ( $f_h \times f_w \times d$ )
- Saída de um volume de dimensão  $(h-f_h+1) \times (w-f_w+1) \times 1$

Figura 11 – Operação de convolução.



Fonte: Autor

Na convolução o *stride* e *padding* são usados para realizar alterações na saída e no tamanho do campo receptivo. O *padding* é comumente usado quando o filtro não se encaixa no tamanho da matriz. O *padding* pode ser feito de duas maneiras:

- Preenchendo a matriz com 0's para se ajustar ao tamanho do *kernel*;
- Retirando os valores da matriz onde o filtro não se aplica.

O *stride* consiste na mudança dos valores de entrada. Também podemos dizer que é o número de espaços ocupados pelos filtros do kernel.

O *pooling* é usado para reduzir o tamanho espacial (linhas e colunas) da matriz para reduzir o número de parâmetros preservando as informações ali contidas. Funciona de forma independente em cada mapa de características. A operação de *pooling* comumente usada é o *max-pooling*, que seleciona o valor máximo dentro da grade.

Figura 12 – Operação de *max-pooling*.



Fonte: (KHOSLA, 2016)

Uma simples convolução 2D ( $N = 2$ ) com entrada quadrada  $i$  com tamanho de *kernel* quadrado  $K$ , taxa de *stride*  $s$  e *padding*  $p$  pode ser representada como:

$$o = i + 2(k - 1) + (k - 1)o = i + k - 1 \quad (2.9)$$

A operação de convolução basicamente incorpora essas três idéias:

- **Compartilhamento de parâmetros:** Esta parte é usada para controlar o número de parâmetros ou pesos onde eles são compartilhados entre um conjunto de características específico. No entanto, no CNN todos os pesos são inicializados aleatoriamente, para garantir a extração de características em todas as posições espaciais.
- **Representação equivariante:** Significa que embora haja transformações nas entradas, a estrutura algébrica é preservada no resultado dessas transformações.
- **Interações esparsas:** É implementado usando detectores de características ou *kernels* menores que a matriz de entrada. É utilizado um tamanho menor de kernel adequado à resolução da matriz para que seja possível extrair melhor as características.

## 2.4.2 LSTM

Muitos problemas que consistem em dados sequenciais, tais como reconhecimento de fala, tarefas relacionadas a texto, previsão de séries temporais, podem ser tratados por RNN. A RNN é um tipo de rede neural em que os nós são interconectados entre si a partir de um caminho direcionado ao longo de uma sequência temporal. Isso faz com que a RNN possa usar seu estado interno (memória) para processar entradas sequenciais. Essa capacidade da RNN de armazenar informações passadas faz com ela seja capaz de lidar bem com problemas que possuem alguma dependência de um valor atual com valores já passados.

A RNN tem o problema do *vanishing gradient* e *exploding gradient*, o que a torna menos popular e inútil em casos de entradas mais longas. No caso de sequências longas, ela não pode acompanhar as sequências de longo prazo, porque o gradiente das células iniciais se torna aproximadamente zero, deixando essas células mortas e causando perda de informações. O LSTM possui uma estrutura semelhante à RNN, mas, em vez de ter uma rede única, cada célula consiste em armazenar e remover memória. As células diferentes são conectadas umas às outras e carregam informações relevantes durante o processamento da sequência. Portanto, as informações das células anteriores são processadas pelas próximas células para superar o problema de memória de curto prazo. À medida que a célula passa informações para outras, as informações são adicionadas ou removidas ainda mais através de portões. Esses portões são redes neurais simples que decidem quais informações serão adicionadas no estado da célula.

$$f_t = \sigma(X_t * U_f + H_{t-1} * W_f) \quad (2.10)$$

$$C_t = \tanh(X_t * U_c + H_{t-1} * W_c) \quad (2.11)$$

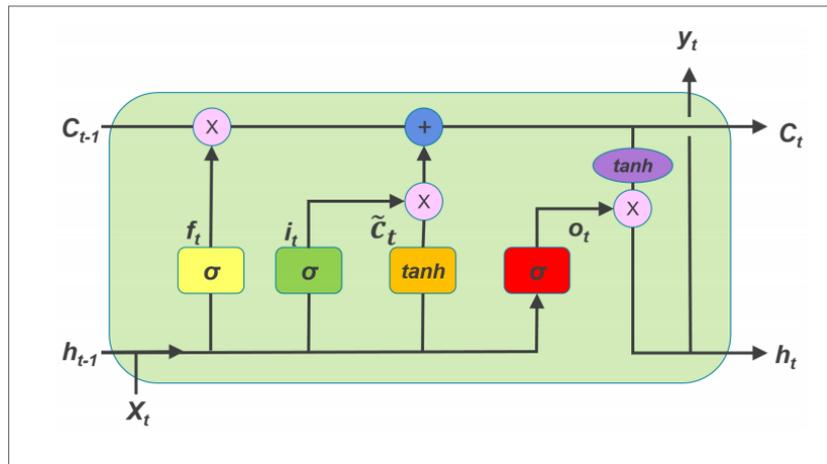
$$l_t = \sigma(X_t * U_i + H_{t-1} * W_o) \quad (2.12)$$

$$O_t = \sigma(X_t * U_o + H_{t-1} * W_o) \quad (2.13)$$

Aqui  $X_t$  é o vetor de entrada,  $H_{t-1}$  é a saída da célula anterior,  $C_{t-1}$  é a memória da célula anterior,  $H_t$  consiste na memória de célula anterior e  $C_t$  é a memória da célula atual.

Esses portões contêm funções de ativação sigmóide. A função de ativação sigmóide reduz os valores entre 0 e 1, o que ajuda a esquecer os dados, pois qualquer valor multiplicado por 0 é 0. Por outro lado, qualquer valor igual a 1 permanece o mesmo, permitindo que o portão mantenha essas informações. A rede aprende quais informações manter e quais esquecer. Uma célula LSTM basicamente consiste em entrada, saída e portão de esquecimento.

Figura 13 – Arquitetura LSTM.



**Fonte:** (ISMAIL; WOOD; BRAVO, 2018)

#### 2.4.2.1 Portão de entrada

A operação do portão de entrada é atualizar o estado da célula. As informações de entrada atual e estado anterior são passadas para a função sigmóide, que decide a importância das informações com base nos valores entre 0 e 1. As informações também podem ser passadas para a função  $\tanh$ , para comprimir valores entre -1 e 1 para regular a rede. A saída da função  $\tanh$  é então multiplicada pela saída sigmóide. A função sigmóide decide quais informações manter da função  $\tanh$ .

#### 2.4.2.2 Portão de esquecimento

Esse portão é usado para descartar informações irrelevantes e manter informações importantes. Inicialmente, as informações do estado anterior e da entrada atual são transmitidas através da função sigmóide. Fornece valores de probabilidade entre 0 e 1. Valores próximos a 0 significam menos significantes, que conseqüentemente devem ser esquecidos, e próximos a 1 significa que devem ser mantidos.

#### 2.4.2.3 Portão de saída

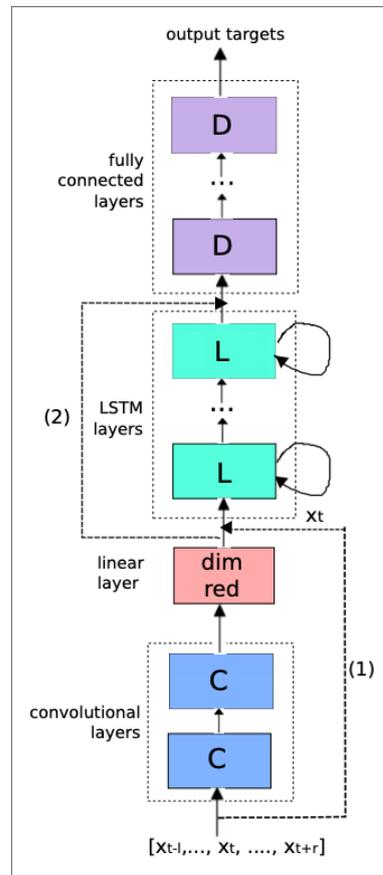
O portão de saída decide quais informações serão passadas para a próxima célula e qual deve ser o estado oculto. O estado oculto é útil para previsões. No portão de saída, primeiro passamos o estado oculto anterior e a entrada atual para a função sigmóide. Em seguida, passamos a saída da célula para a função *tanh* para descobrir quais informações devem ser incluídas na saída.

### 2.4.3 CNN-LSTM

Um CNN-LSTM é uma integração de uma camada CNN (camadas convolucionais) com uma camada LSTM. Primeiro, a parte CNN do modelo processa os dados e o resultado unidimensional alimenta um modelo LSTM. A representação visual dessa arquitetura é mostrada na Figura 14.

O modelo híbrido de CNN e LSTM conhecido como CNN-LSTM é basicamente uma arquitetura baseada em LSTM projetada para lidar com a previsão de sequências com entradas espaciais. Envolve o uso do CNN como camada inicial para a extração de características a partir de dados de entrada e esses recursos são fornecidos ao LSTM para previsão de sequência e outros problemas relacionados.

Figura 14 – CNN-LSTM



**Fonte:** (SAINATH et al., 2015)

Na Figura 14, inicialmente, a entrada é passada para a camada de convolução na qual as características importantes são extraídas. Os blocos azuis C representam as camadas de convolução. Depois, as características são comprimidas para passar para a camada LSTM. O bloco vermelho representa a camada linear. Em seguida, a saída é passada para as camadas LSTM representadas em cores azuis claras para capturar informações inter-relacionadas entre elas. Em seguida, a saída do LSTM é passada para as camadas totalmente conectadas, também conhecidas como camadas densas para previsão.

### 3 TRABALHOS RELACIONADOS

Este capítulo tem por objetivo trazer um esboço de outros estudos que estão relacionados com essa pesquisa. Nesse sentido, pesquisas relacionadas a previsão de demanda utilizando os modelos ARIMA, LSTM e CNN-LSTM foram buscadas a fim de dar embasamento ao desenvolvimento dessa dissertação.

Essa revisão de trabalhos relacionados teve como objetivo principal realizar a identificação de trabalhos considerados relevantes dentro das principais plataformas de divulgação científica. Para tal feito foram utilizadas os repositórios do *IEEE*, *Springer* e *ScienceDirect*. Para realização das buscas foram utilizados os critérios de relevância e o ano de publicação do trabalho, sendo estabelecido 2017 como ano mínimo.

#### 3.1 PREVISÃO DE DEMANDA COM ARIMA

Mesmo com o surgimento de diversas outras técnicas, o ARIMA continua sendo um modelo bastante utilizado no que diz respeito a previsão de séries temporais. Em um trabalho recente, Noureen et al. (2019) faz utilização do modelo para previsão de demanda de energia. Neste trabalho é destacado a importância de identificar a estacionariedade da série temporal utilizada, pois caso a série não seja estacionária, métodos de diferenciação terão de ser aplicados para a obtenção de melhores resultados no modelo ARIMA. Para tal identificação Noureen et al. (2019) destaca a observação da ACF e da PACF, sendo posteriormente aplicado o teste *ADF* para confirmar a estacionariedade. Embora os resultados obtidos tenham sido satisfatórios, Noureen et al. (2019) sugere a possibilidade de se estudar ainda outros tipos de modelos para serem comparados com resultados obtidos pelo modelo ARIMA, o que dá fundamento ao objetivo desta pesquisa.

Em seu trabalho Silva, Figueiredo e Braga (2019) propõem a utilização e combinação de alguns modelos para melhorar o sistema de previsão de demanda de uma empresa do ramo alimentício, que até então utilizava apenas médias móveis simples com uma janela de quatro dias para realizar suas previsões. Dentre as novas técnicas propostas para a empresa estão a suavização exponencial e o modelo ARIMA. Silva, Figueiredo e Braga (2019) frisam que embora estudos empíricos teoricamente mostrem uma superioridade do modelo ARIMA sobre outros métodos de previsão, na prática é bem comum isso não acontecer como é mostrado na figura abaixo.

Figura 15 – Comparação de performance de modelos.

2016	Observed values	Company forecasts	Exp. smooth forecasts	ARIMA	Combined forecasts
1	210170	195563	213706	217858	209042
2	182101	200109	173385	177080	183525
3	202687	197375	191788	196982	195382
4	183538	194809	196571	203926	198435
5	176603	188812	217957	172769	193179
6	148692	192680	183446	186774	187633
7	188945	194624	194643	201630	196966
8	185514	186232	186930	191736	188299
9	184927	177880	212852	217607	202780
10	182538	174445	177823	181638	177969
11	179680	174939	187624	192807	185123
12	212595	177020	186711	192316	185349
13	190533	185481	197910	203552	195648
14	194113	183165	193957	206776	194633
15	191991	189935	191605	198666	193402
16	202631	191337	204498	211532	202456
17	198675	194230	191943	199927	195367
	<i>MAD</i>	11826	11905	12301	9219
	<i>MAPE</i>	6.49%	6.60%	6.77%	5.16%
	<i>MSE</i> × 10 <sup>8</sup>	2.65	2.92	2.53	1.94

**Fonte:** (SILVA; FIGUEIREDO; BRAGA, 2019)

Como é possível ver na Figura 15 a suavização exponencial apresenta um resultado levemente melhor que o apresentado pelo modelo ARIMA. Porém o destaque vai para a combinação dos dois métodos (ARIMA e suavização exponencial) que é responsável por trazer os melhores resultados da pesquisa realizada por Silva, Figueiredo e Braga (2019). Partindo desse pressuposto, é possível observar que há espaço para a exploração de outros métodos além do ARIMA que visem melhorar os resultados das previsões.

Por fim, Ren, Chan e Ram (2017) realizam em sua pesquisa uma busca por métodos que sejam capazes de prever demandas de produtos de moda. Dentre os métodos trabalhados na pesquisa está o modelo ARIMA que é definido pelos autores como um método de previsão bom para abordagens mais rápidas, por ser intuitivo e mais fácil de aplicar. Porém, conforme Ren, Chan e Ram (2017), os modelos ARIMA são insuficientes para problemas de previsão que exigem a observação de outras variáveis além das contidas pela série temporal. Nesse sentido, Ren, Chan e Ram (2017) citam a utilização de RNA como uma classe de modelos capazes de lidar com tal limitação.

A próxima seção descreve os trabalhos relacionados a utilização do LSTM para previsão de demanda.

### 3.2 PREVISÃO DE DEMANDA COM LSTM

Uma classe de RNA que tem mostrado eficiência no contexto de previsão de séries temporais são as RNN, com destaque para o LSTM devido a sua capacidade de aprender dependências de longo prazo. Wang e Duggasani (2018) fazem a utilização do LSTM para realizar a previsão de demanda de reservas de hotéis. Dentro do trabalho Wang e Duggasani (2018), destacam o LSTM como um modelo bem adaptado para previsão de séries temporais uma vez que ela prevê o futuro com base em uma sequência de valores

passados, ao contrário de outros modelos mais tradicionais que assumem a independência das variáveis. Um outro ponto interessante apontado pelo trabalho foi a não necessidade de identificar e corrigir *outliers*, uma vez que algoritmos de RNA são capazes de lidar com dados incompletos e previamente desconhecidos (SHAW, 1992). Por fim, Wang e Duggasani (2018) descrevem o resultados obtidos pelo LSTM como satisfatórios, uma vez que comparados com outros seis resultados de algoritmos diferentes, mostraram-se superior quase todas as vezes.

Siame-Namini e Namin (2018) utilizam uma abordagem bastante semelhante a desta dissertação, realizar a comparação entre os resultados obtidos pelos modelos ARIMA e LSTM. Porém o tipo de problema aplicado é diferente, uma vez que os testes são realizados para previsão de valores no mercado financeiro, o que não é especificamente um problema de previsão de demanda. Embora não aplicado a um problema de previsão de demanda, os resultados obtidos são bem encorajadores conforme mostrado na Figura 16.

Figura 16 – Comparação de performance de modelos ARIMA e LSTM.

Stock	RMSE		% Reduction in RMSE
	ARIMA	LSTM	
N225	766.45	105.315	-86.259
IXIC	135.607	22.211	-83.621
HSI	1,306.954	141.686	-89.159
GSPC	55.3	7.814	-85.869
DJI-Monthly	516.979	77.643	84.981
DJI-Weekly	287.6	30.61	-89.356
<b>Average</b>	511.481	64.213	-87.445
MC	0.81	0.801	-1.111
HO	0.522	0.43	-17.624
ER	1.286	0.251	-80.482
FB	0.478	0.397	-16.945
MS	30.231	3.17	-89.514
TR	2.672	0.569	-78.705
<b>Average</b>	5.999	0.936	-84.394

Fonte: (SIAMI-NAMINI; NAMIN, 2018)

Como é possível ver, Siame-Namini e Namin (2018) conseguiram melhores resultados com o uso do LSTM em comparação com o ARIMA. Os resultados obtidos se mostraram superiores, com uma média de redução de erro de aproximadamente 85%.

Mesmo com os excelentes resultados mostrados por Siame-Namini e Namin (2018), não é válido afirmar que o LSTM irá sempre obter melhores resultados que o modelo ARIMA. Babich et al. (2019) mostra isso em sua pesquisa onde em um conjunto de modelos, incluindo o ARIMA, o LSTM é o que apresenta o pior dos resultados conforme é mostrado na Figura 17.

Figura 17 – Comparação de performance de modelos LSTM, ARIMA e SVM.

Method	MAPE			
	1 week	2 weeks	3 weeks	4 weeks
LSTM	0.54	0.56	0.57	0.59
SVM RBF	0.12	0.12	0.12	0.12
SVM linear	0.14	0.14	0.15	0.15
ARIMA	0.11	0.12	0.12	0.12

**Fonte:** (BABICH et al., 2019)

Essa disparidade apresentada entre os trabalhos de Siami-Namini e Namin (2018) e Babich et al. (2019) reforçam a necessidade de novos estudos de aplicação desses modelos a diferentes tipos de problemas, a fim de identificar e mapear as principais características de cada problema, podendo servir como ponto de partida para outras iniciativas.

A próxima seção descreve trabalhos relacionados a utilização do modelo híbrido CNN-LSTM para previsão de demanda.

### 3.3 PREVISÃO DE DEMANDA COM CNN-LSTM

Uma forma de tentar melhorar os resultados obtidos por um LSTM é combiná-lo com uma camada de CNN, formando então o modelo híbrido CNN-LSTM. Hu et al. (2019) utiliza esse modelo para realizar a previsão de demanda urbana de água, onde realiza a comparação dos resultados obtidos com alguns outros algoritmos. Hu et al. (2019) destaca a importância da utilização de uma camada de CNN devido a sua capacidade de extração de características que pode auxiliar na predição que será realizada de fato na camada do LSTM.

Figura 18 – Comparação de performance de modelos LSTM e CNN-LSTM.

Length (week)	5		10		15		20	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
LSTM	103.3	4.4%	97.5	3.71%	88.37	3.54%	84.1	3.25%
CNN-LSTM	100.2	4.27%	99.51	3.84%	86.35	3.42%	78.63	3.14%

**Fonte:** Adaptado de (HU et al., 2019)

Como é possível ver na Figura 18 o modelo híbrido apresentou na maioria das vezes uma leve melhora nos resultados em comparação com o LSTM, o que serve como incentivo para uma melhor exploração desse método.

Kim e Cho (2019) conseguem obter resultados ainda mais satisfatórios que os obtidos por Hu et al. (2019) com a utilização do modelo híbrido, aplicando-o na previsão de consumo de energia residencial.

Figura 19 – Comparação de performance de modelos LSTM, GRU, CNN-LSTM e outros.

Method	MSE	RMSE	MAE	MAPE
LSTM	0.7480	0.8649	0.6278	51.45
GRU	0.7432	0.8620	0.6291	51.47
Bi-LSTM	0.7235	0.8505	0.6122	50.10
Attention LSTM	0.6984	0.8357	0.5911	48.36
The proposed	<b>0.3738</b>	<b>0.6114</b>	<b>0.3493</b>	<b>34.84</b>

Fonte: (KIM; CHO, 2019)

Kim e Cho (2019) atribuem o bom resultado obtido pelo modelo híbrido a capacidade de extração de características pela camada do CNN de variáveis que afetam o consumo de energia, além da capacidade do modelo de lidar com séries de tendências irregulares que é uma característica de dados de consumo de energia.

### 3.4 CONSIDERAÇÕES

De acordo com a busca e avaliação dos trabalhos relacionados realizada, foi possível observar que o ARIMA já é um modelo bastante consolidado no contexto de previsão de demanda. Mesmo com o surgimento de outras técnicas é notório que o modelo ARIMA ainda é bastante utilizado. Porém, em muitos trabalhos que englobam o modelo ARIMA, é bastante comum o envolvimento de outros algoritmos que visam buscar uma alternativa que supere as limitações do modelo e tragam melhores resultados, onde alguns desses outros algoritmos têm apresentado resultados satisfatórios. Com a recente popularização dos algoritmos de aprendizagem profunda, algoritmos baseados em RNN têm apresentado resultados interessantes nesse contexto de previsão de demanda, com destaque para o LSTM, como é o caso do trabalho de Siami-Namini e Namin (2018). Porém nem sempre é possível obter resultados satisfatórios como mostra Babich et al. (2019).

Para entender melhor esses eventos se faz necessário a aplicação desses modelos em um número maior de diferentes problemas de previsão demanda, para que sirvam de base de conhecimento com resultados mais consolidados. Porém, atualmente, a literatura ainda carece de trabalhos relacionados a aplicação de modelos LSTM e CNN-LSTM em problemas de previsão de demanda. Em uma pesquisa geral realizada nos repositórios já citados nesse capítulo, a Tabela 1 mostra a disparidade da quantidade de trabalhos relacionados a esses modelos em comparação com trabalhos relacionados com o modelo ARIMA.

**Fonte:** Autor

Modelo	IEEE		Springer		ScienceDirect	
	Ano mínimo	Quantidade	Ano mínimo	Quantidade	Ano mínimo	Quantidade
ARIMA	1998	159	1979	502	1993	151
LSTM	2016	58	2017	46	2017	25
CNN-LSTM	2019	4	2018	3	2017	5

\*Pesquisa realizada em 30/12/2019.

Tabela 1 – Quantidade de artigos encontrados por modelo

Para realização de tal pesquisa nos repositórios foram utilizadas as seguintes condições:

- O artigo deve conter o termo "*demand forecasting*";
- O artigo deve conter o nome do algoritmo correspondente a pesquisa "(*ARIMA, LSTM, CNN-LSTM*)".

É possível que a pesquisa mostrada na Tabela 1 não traga com exatidão os resultados das quantidades dos trabalhos, porém é possível mostrar uma noção da realidade, uma vez que o padrão dos resultados é comum em todos os repositórios onde foram realizadas as buscas.

Com base nos pontos observados, é possível enxergar a relevância da pesquisa apresentada nessa dissertação não só do ponto de vista de mostrar os resultados da aplicação dessas técnicas, mas também como uma contribuição para uma área de aplicação que ainda possui poucos trabalhos relacionados.

## 4 EXPERIMENTOS REALIZADOS

Este capítulo tem por objetivo descrever de forma ordenada cada uma das etapas dos experimentos realizados na pesquisa. Sendo assim a seção 4.1 apresenta todo o processo de análise dos dados e suas características, a seção 4.2 apresenta as transformações que foram feitas nos dados para que pudessem ser usados em alguns algoritmos, a seção 4.3 discorre sobre as métricas utilizadas para avaliação dos modelos, a seção 4.4 apresenta o processo de modelagem e resultados obtidos com o ARIMA, a seção 4.5 apresenta as etapas das modelagens e resultados obtidos pelo modelo LSTM, e por fim, a seção 4.6 mostra o processo de modelagem e resultados obtidos pelo modelo híbrido CNN-LSTM.

### 4.1 ANÁLISE DOS DADOS

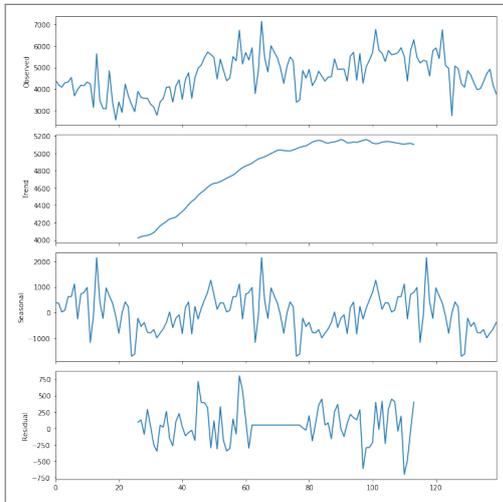
Para realização dos experimentos dessa pesquisa foram utilizados cinco conjuntos de dados reais de uma empresa do ramo de bebidas. Os dados correspondem a uma série temporal que apresenta registros semanais da produção de cinco produtos específicos, cada um com suas próprias características, o que traz mais dinamismo para o desenvolvimento dos modelos de predição. A utilização desses dados também acabou sendo um desafio de pesquisa devido a pequena quantidade de registros, sendo apenas 140 registros por *dataset*.

#### 4.1.1 Decomposição da série

Antes de realizar qualquer modelagem de predição, foi realizado uma análise através dos componentes da série a fim de buscar entender as características de cada conjunto de dados. Para realização da decomposição foi utilizada a função *seasonal\_decompose* da biblioteca *statsmodels*<sup>1</sup> em *Python*, sendo utilizado o modelo aditivo como modelo de decomposição. As Figuras 20, 21, 22, 23 e 24 mostram as séries decompostas do que chamaremos de Produto 1, Produto 2, Produto 3, Produto 4 e Produto 5. Cada uma dessas imagens apresentam a série temporal em sua completude (*Observed*), a linha de tendência da série (*Trend*), a sazonalidade (*Seasonal*) e os resíduos da série (*Residual*).

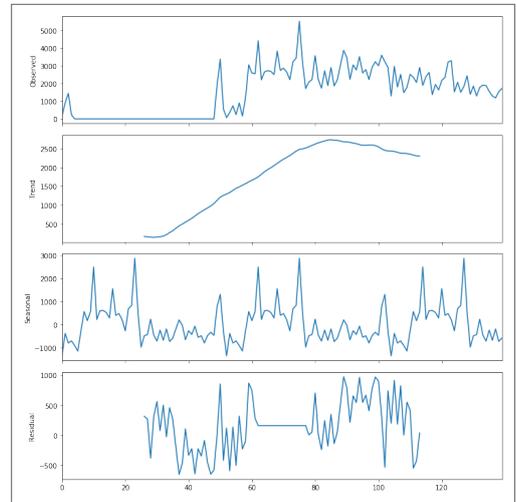
<sup>1</sup> <https://www.statsmodels.org/stable/index.html>

Figura 20 – Decomposição dos componentes do Produto 1.



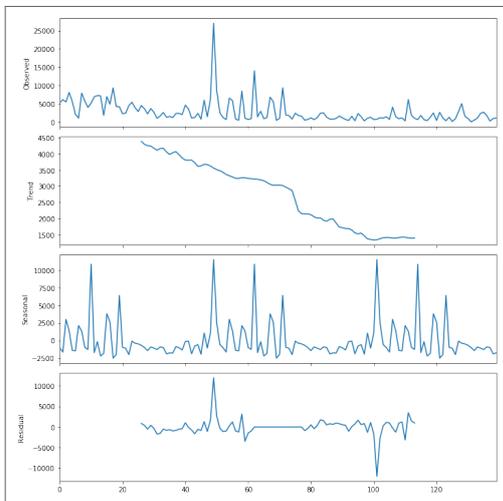
Fonte: Autor

Figura 21 – Decomposição dos componentes do Produto 2.



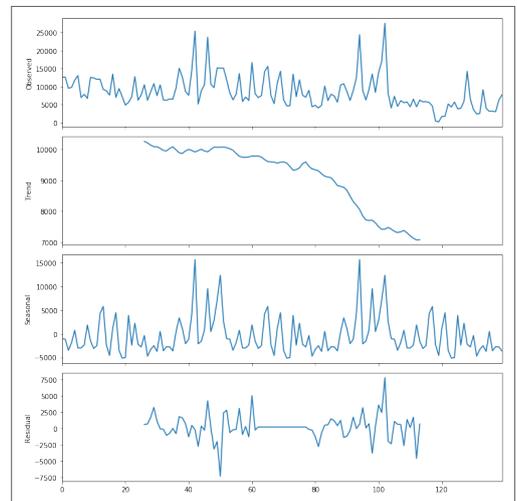
Fonte: Autor

Figura 22 – Decomposição dos componentes do Produto 3.



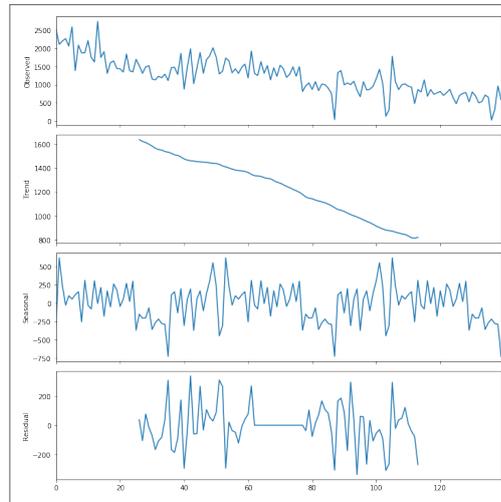
Fonte: Autor

Figura 23 – Decomposição dos componentes do Produto 4.



Fonte: Autor

Figura 24 – Decomposição dos componentes do Produto 5.



**Fonte:** Autor

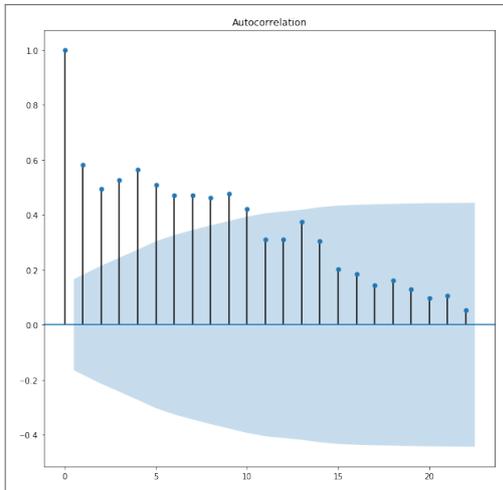
Como é possível ver nas figuras acima, todas as séries apresentam algum grau de tendência, seja ela de queda ou de alta, e apresentam algum padrão de sazonalidade. Tais eventos indicam uma possível não estacionariedade das séries, porém outros testes precisam ser realizados para chegar a essa conclusão. A próxima subseção tem justamente por objetivo trazer detalhes desses testes que foram realizados.

#### 4.1.2 Análise de estacionariedade

Para análise da estacionariedade foram analisados os gráficos da ACF e PACF das séries temporais, além da realização do teste ADF.

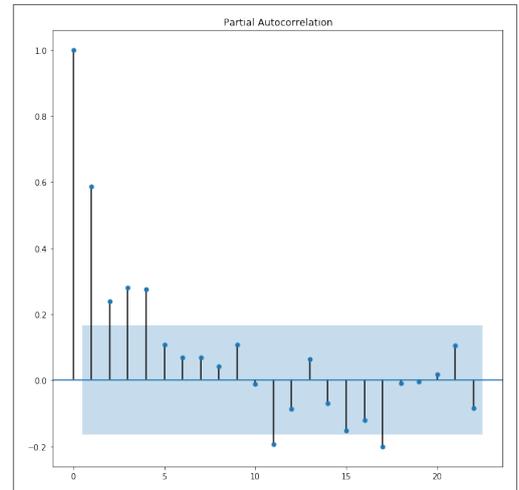
Para o Produto 1 é possível enxergar um indício de não estacionariedade da série uma vez que os valores do ACF (Figura 25) apresentam uma caída mais lenta enquanto os valores do gráfico PACF (Figura 26) apresentam uma caída mais rápida.

Figura 25 – ACF do Produto 1.



**Fonte:** Autor

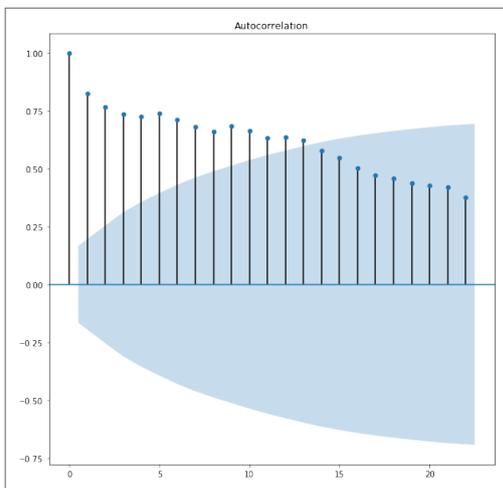
Figura 26 – PACF do Produto 1.



**Fonte:** Autor

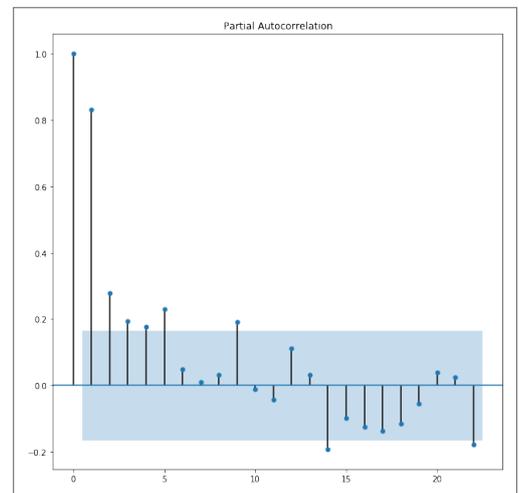
Pela análise gráfica o Produto 2 mostra claramente ser uma série não estacionária. Além da maioria dos pontos estarem fora da área de significância no gráfico da ACF (Figura 27), eles decaem muito lentamente, ao contrário dos pontos da PACF (Figura 28) que decaem de maneira mais rápida.

Figura 27 – ACF do Produto 2.



**Fonte:** Autor

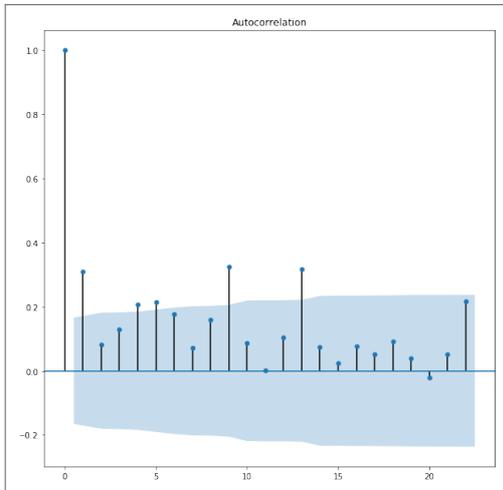
Figura 28 – PACF do Produto 2.



**Fonte:** Autor

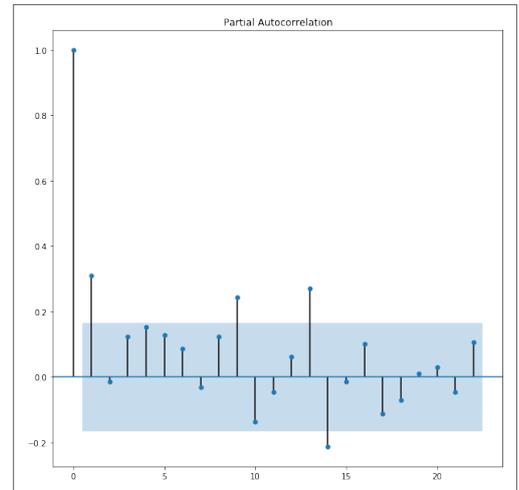
Para o Produto 3 a análise da ACF (Figura 29) não traz clareza sobre a estacionariedade da série. Como é possível ver na Figura 29, embora a maioria dos pontos estejam dentro da área de significância, não existe um comportamento claro de uma tendência de queda entre eles. Nesse caso se faz necessário a realização de outros testes para identificar a estacionariedade da série.

Figura 29 – ACF do Produto 3.



**Fonte:** Autor

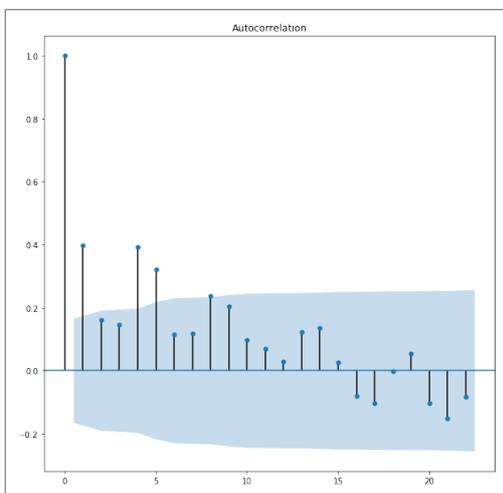
Figura 30 – PACF do Produto 3.



**Fonte:** Autor

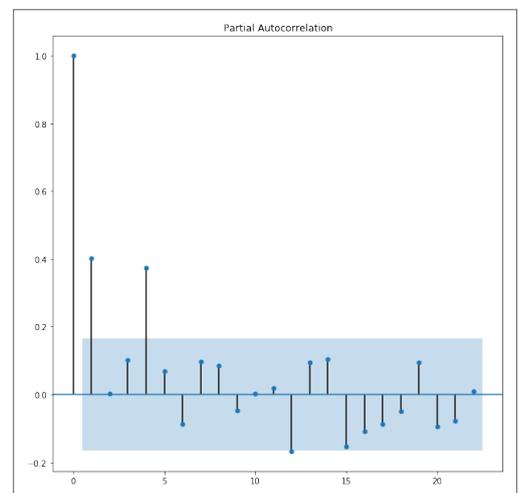
O Produto 4 apresenta sinais de estacionariedade, pois tanto no gráfico da ACF (Figura 31) quanto no da PACF (Figura 32) os pontos decaem mais rapidamente e se apresentam em sua maioria dentro da área de significância. Além disso os pontos apresentam uma tendência de queda um pouco mais acentuada.

Figura 31 – ACF do Produto 4.



**Fonte:** Autor

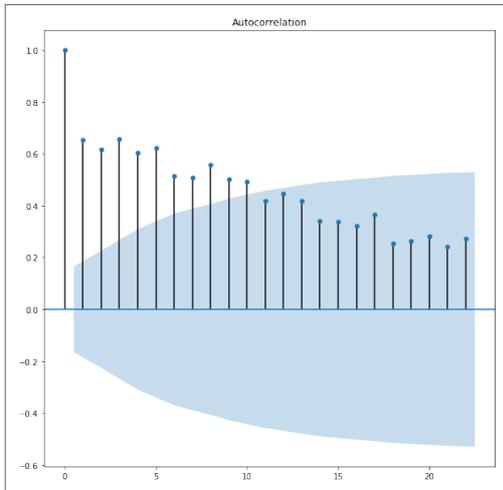
Figura 32 – PACF do Produto 4.



**Fonte:** Autor

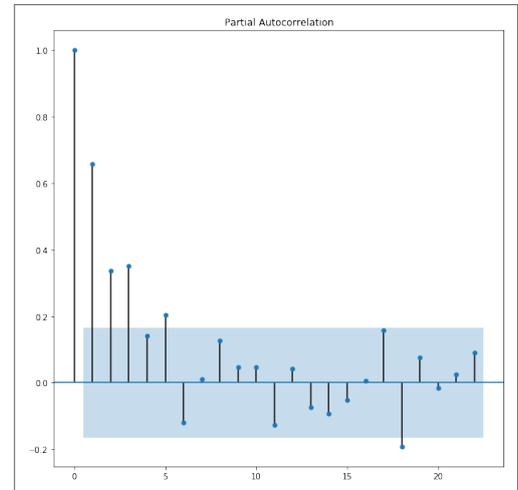
Por fim, o Produto 5 apresenta características de uma série não estacionária. Os pontos do gráfico da ACF (Figura 33) decaem de forma mais lenta, além de parte considerável dos pontos estar fora da área de significância. Enquanto de forma contrária a ACF (Figura 33), a PACF (Figura 34) apresenta uma decaída mais rápida.

Figura 33 – ACF do Produto 5.



Fonte: Autor

Figura 34 – PACF do Produto 5.



Fonte: Autor

Para reforçar o resultado da análise feita através dos gráficos, foi realizado o teste estatístico ADF. O teste ADF basicamente assume duas hipóteses (BROWNLEE, 2016):

- **Hipótese nula (H0):** Caso seja aceita, indica que a série temporal tem uma raiz unitária, significando que ela não é estacionária. Indica que existe alguma estrutura dependente do tempo.
- **Hipótese alternativa (H1):** A hipótese nula é rejeitada, sugerindo que a série temporal não possui uma raiz unitária, o que significa que é estacionária. Indica que não possui estrutura dependente do tempo.

Sua interpretação pode ser feita através do *p-value* obtido pelo teste através das seguintes condições (BROWNLEE, 2016):

- ***p-value* > 0,05:** Falha ao rejeitar a hipótese nula (H0), onde os dados têm uma raiz unitária e não são estacionários.
- ***p-value* ≤ 0,05:** Rejeita a hipótese nula (H0), o que implica que os dados não têm uma raiz unitária e são estacionários.

Com aplicação do teste ADF em cada um dos *datasets*, os seguintes resultados foram obtidos:

**Fonte:** Autor

<b>Produto</b>	<b><i>p-value</i></b>
Produto 1	0,22
Produto 2	0,52
Produto 3	0,22
Produto 4	0,03
Produto 5	0,41

Tabela 2 – Resultados dos testes ADF

Como é possível ver nos testes realizados, apenas a série do Produto 4 é estacionária. Isso implica que talvez seja a única série que não precisa passar pelo processo de diferenciação do modelo ARIMA.

## 4.2 PRÉ-PROCESSAMENTO DOS DADOS

Para que as séries temporais pudessem ser utilizadas pelos modelos LSTM e CNN-LSTM, foi necessário transformá-las em um problema de aprendizado supervisionado. Para tal finalidade foi utilizada a transformação *sliding window*, que consiste em usar valores passados da série como variáveis de entrada e o próximo valor como variável de saída. Para isso é necessário definir o tamanho da janela que representará a quantidade de valores passados que servirão como entrada para cada variável de saída. Por exemplo, suponha a série temporal

$$\left[ 10 \ 20 \ 30 \ 40 \ 50 \ 60 \ 70 \ 80 \ 90 \ 100 \right]$$

A transformação dessa série com o *sliding window* com um tamanho de janela 5 ficará

$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & y \\ \left[ \begin{array}{ccccc} 10 & 20 & 30 & 40 & 50 \\ 20 & 30 & 40 & 50 & 60 \\ 30 & 40 & 50 & 60 & 70 \\ 40 & 50 & 60 & 70 & 80 \\ 50 & 60 & 70 & 80 & 90 \end{array} \right] & \left[ \begin{array}{c} 60 \\ 70 \\ 80 \\ 90 \\ 100 \end{array} \right] \end{array}$$

É possível observar que com essa transformação, os  $n$  primeiros valores, onde  $n$  é o tamanho da janela definido, não farão mais parte do *dataset*, pois não existem valores prévios para completar o tamanho da janela.

Como já citado anteriormente, cada um dos *datasets* utilizados na pesquisa possui 140 registros. Para cada um deles a transformação de *sliding window* foi feita usando janelas de tamanho 5. Sendo assim, dos 140 registros restaram um total de 135 para utilização nos modelos LSTM e CNN-LSTM. Para o modelo ARIMA não é necessário esse tipo de transformação, sendo assim, todos os 140 registros foram utilizados.

Uma das etapas dos experimentos consistiu em utilizar os dados temporais da série para criação de novas variáveis, tais como mês, semana e estações do ano. Para isso foi utilizada a técnica *One-Hot Encode* que consiste em binarizar essas variáveis, o que implica na remoção de qualquer relacionamento ordinal entre elas, tornando-as equidistantes (CERDA; VAROQUAUX; KÉGL, 2018). Para representar dados relacionados aos meses, por exemplo, foram criadas doze novas variáveis, cada uma representando um mês do ano, onde caso o dado fosse relativo a um mês específico, a coluna referente aquele mês é preenchida com 1 e as demais colunas referentes aos outros meses é preenchida com 0. A utilização dessa técnica teve por objetivo buscar uma maior generalização dos dados tentando criar mais padrões entre eles. A Figura 35 mostra um exemplo de utilização da técnica.

Figura 35 – Aplicação do One-Hot Encode para criação de variáveis temporais

F	G	H	I	J	K	L	M	N	O	P	Q
jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
0	1	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0

Fonte: Autor

Por fim, os dados foram separados em conjuntos de treinamento e conjuntos de testes. Devido a pouca quantidade de registros, foi definido que apenas os seis últimos seriam utilizados para testes, ficando o restante para treinamento. Sendo assim, os modelos LSTM e CNN-LSTM trabalharam com 129 registros para treinamento e 6 para testes, e o modelo ARIMA trabalhou com 134 registros de treinamento e também 6 para testes.

### 4.3 MÉTRICAS DE ERRO

Para avaliação dos resultados foram utilizadas três métricas de erro diferentes: o *Mean Absolute Error (MAE)*, *Mean Absolute Percentage Error (MAPE)* e *Root Mean Squared Error (RMSE)*.

### 4.3.1 MAE

O MAE mostra o erro médio absoluto das predições, descrevendo a magnitude do erro produzido. O MAE pode ser obtido através da seguinte equação:

$$\frac{1}{n} \sum_{t=1}^n |e_t| \quad (4.1)$$

onde  $e_t$  é a diferença entre o valor real e o valor predito.

### 4.3.2 MAPE

Enquanto o MAE apresenta a magnitude do erro produzido, o MAPE possui uma interpretação mais clara, pois mostra uma porcentagem média do quão distantes os valores das predições estão dos valores reais. O MAPE pode ser obtido com:

$$\frac{100\%}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right| \quad (4.2)$$

onde  $e_t$  é a diferença entre o valor real e o valor predito, e  $y_t$  corresponde ao valor real.

### 4.3.3 RMSE

O RMSE pode ser obtido por:

$$\sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (4.3)$$

onde  $e_t$  é a diferença entre o valor real e o valor predito. A equação do RMSE é semelhante a do MAE, exceto pelo fato de ter o erro elevado ao quadrado e, posteriormente ser calculada a raiz quadrada do somatório. Tal métrica ajuda na identificação da ocorrência de resultados muito discrepantes, pois como o erro cresce quadraticamente, um erro total muito maior surgirá em comparação com o MAE.

## 4.4 MODELAGEM COM ARIMA

O processo de modelagem com ARIMA foi desenvolvido por uma empresa conceituada no mercado brasileiro no ramo de inteligência artificial, e que compartilhou os resultados obtidos da investigação de diferentes configurações dos modelos de previsão para otimização dos resultados. Entretanto, os detalhes de como a modelagem foi realizada foram obtidos e estão descritos nos próximos parágrafos.

A modelagem com ARIMA basicamente consiste em encontrar os melhores parâmetros  $p$ ,  $d$  e  $q$  que apresentem os melhores resultados de predição. A busca por esses parâmetros pode ser feita de várias formas, sendo uma delas analisando os gráficos de ACF e PACF mostrados na subseção 4.1.2, ou fazendo os testes manualmente, alterando os parâmetros um a um e observando os resultados. Essas maneiras citadas podem não ser muito

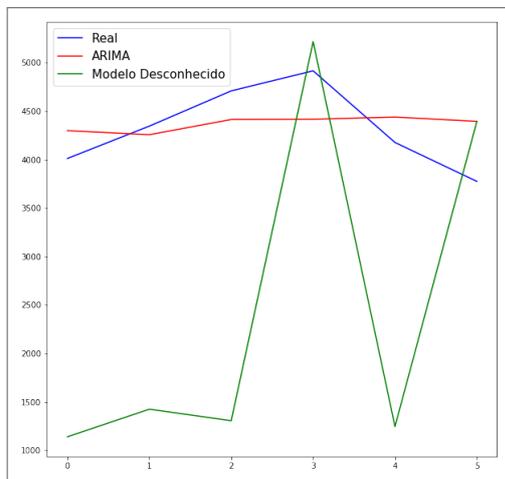
eficientes, pois podem produzir resultados sem muita precisão e podem consumir mais tempo. Processos automatizados podem ser mais efetivos, uma vez que podem apresentar resultados mais precisos e com mais agilidade.

Para modelagem do ARIMA foi utilizado o pacote *auto\_arima*<sup>2</sup> da linguagem de programação *R*. O *auto\_arima* conduz o processo de busca pelos parâmetros de forma automatizada com a utilização de diversos testes de diferenciação (*i.e.*, *Kwiatkowski Phillips Schmidt Shin*, *Augmented Dickey-Fuller* ou *Phillips-Perron*) para definir o parâmetro  $d$ , além da busca pelos parâmetros  $p$  e  $q$ . No *auto\_arima* um conjunto de parâmetros é dito o melhor quando ele atinge o menor valor para um estimador de erro. Nas modelagens realizadas foi utilizado o estimador de erro *Akaike Information Criterion (AIC)* que mede a qualidade dos modelos (OZAKI, 1977).

Além disso a empresa que desenvolveu a modelagem com o ARIMA também disponibilizou um conjunto de outros resultados desenvolvidos pela Empresa Fornecedora dos Produtos. Porém para esses outros resultados não foi possível obter detalhes do processo de modelagem, nem mesmo qual a técnica utilizada por eles. Tal modelo foi desenvolvido pela equipe interna de especialistas e modelagem do problema, sendo este o modelo atual em operação na previsão de demanda semanal. Sendo assim iremos utilizar os resultados obtidos pela Empresa Fornecedora dos Produtos apenas para fins de comparação com os outros modelos, o qual será referenciado nos experimentos como Modelo em Operação.

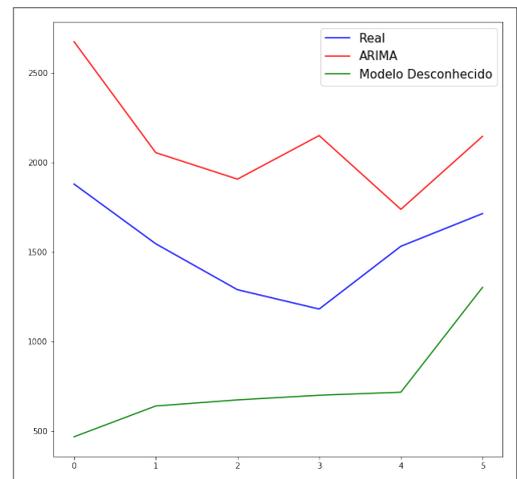
Nas Figuras 36, 37, 38, 39, 40 e na Tabela 3 são mostrados os resultados obtidos com a utilização do ARIMA pela primeira empresa, juntamente com os resultados obtidos pela segunda empresa através do Modelo em Operação.

Figura 36 – Predição com ARIMA e Modelo em Operação do Produto 1.



Fonte: Autor

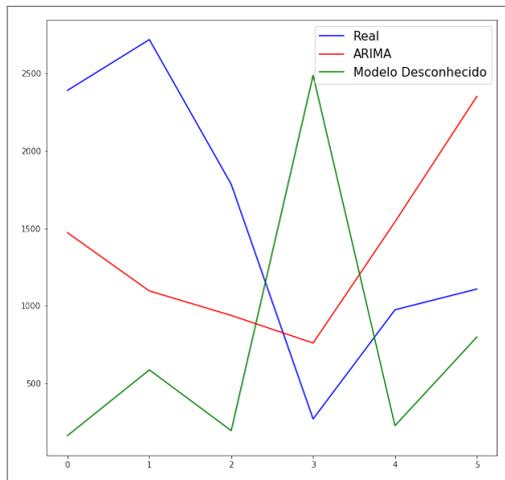
Figura 37 – Predição com ARIMA e Modelo em Operação do Produto 2.



Fonte: Autor

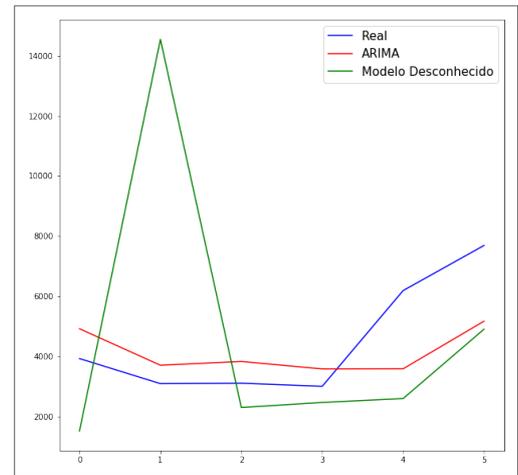
<sup>2</sup> <https://www.rdocumentation.org/packages/forecast/versions/8.10/topics/auto.arima>

Figura 38 – Predição com ARIMA e *Modelo em Operação* do Produto 3.



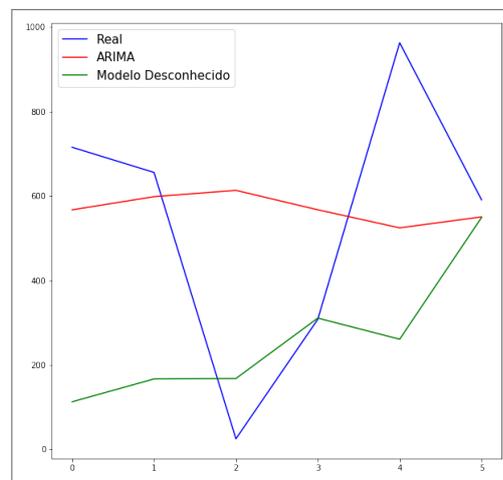
Fonte: Autor

Figura 39 – Predição com ARIMA e *Modelo em Operação* do Produto 4.



Fonte: Autor

Figura 40 – Predição com ARIMA e *Modelo em Operação* do Produto 5.



Fonte: Autor

Fonte: Autor

Produto	ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>588,2</b>	<b>40,66</b>	<b>637,87</b>	773,7	49,94	842,56
Produto 3	<b>948,04</b>	<b>82,72</b>	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	<b>1337,79</b>	<b>27,04</b>	<b>1598,53</b>	3598,22	94,9	5141,81
Produto 5	<b>255,13</b>	413,98	<b>324,4</b>	329,94	<b>133,68</b>	431,17

Tabela 3 – Resultados das predições com ARIMA e com o Modelo em Operação

Pelos resultados mostrados na Tabela 3 é possível observar uma diminuição do erro com a utilização do ARIMA em comparação com o modelo desenvolvido pela Empresa Fornecedora dos Produtos, porém não é possível discorrer sobre o porquê desse evento uma vez que a técnica utilizada é desconhecida.

Uma vez obtidos tais resultados, a etapa seguinte dessa pesquisa consistiu em explorar de diferentes formas o uso do LSTM em busca de taxas de erros ainda menores. A próxima seção tem por objetivo discorrer sobre as etapas desse processo de utilização do LSTM.

## 4.5 MODELAGEM COM LSTM

Para modelagem do LSTM foi utilizada a biblioteca *Keras*<sup>3</sup> que possui várias implementações de algoritmos de aprendizagem profunda, utilizando o *TensorFlow*<sup>4</sup> em seu *background*. A implementação dos modelos foi realizada utilizando *Python 3.6* como linguagem de programação. A próxima subseção apresenta as características da arquitetura utilizada para implementação do LSTM.

### 4.5.1 Arquitetura do LSTM

Listing 4.1 – Modelo LSTM em Python

```

1 def fit_model_lstm(X, y):
    model = Sequential()
3     model.add(LSTM(100, activation='relu', input_shape=(n_steps, n_features)))
    model.add(Dense(1))
5     model.compile(optimizer='adam', loss='mse')
    model.fit(X, y, epochs=2000, verbose=0)
7     return model

```

Como é possível ver no trecho de código acima, o modelo foi desenvolvido com a existência de duas camadas, a camada do LSTM e a camada densa de onde se obterá a saída. Na camada de LSTM foram utilizados 100 neurônios e a Rectified Linear Unit (ReLU) (ZEILER et al., 2013) como função de ativação. Para atualização dos pesos foi utilizado o otimizador *Adam* que é uma extensão do GDE e vem sendo bastante utilizado (KINGMA; BA, 2014), tendo como métrica de erro o RMSE. Por fim, foram definidas 2000 épocas para treinamento do modelo.

Uma vez definida a arquitetura e implementação do modelo, a etapa seguinte consistiu em realizar o treinamento e testar os modelos obtidos. A próxima subseção discorre sobre esses testes realizados e os resultados obtidos por eles.

### 4.5.2 Resultados com LSTM

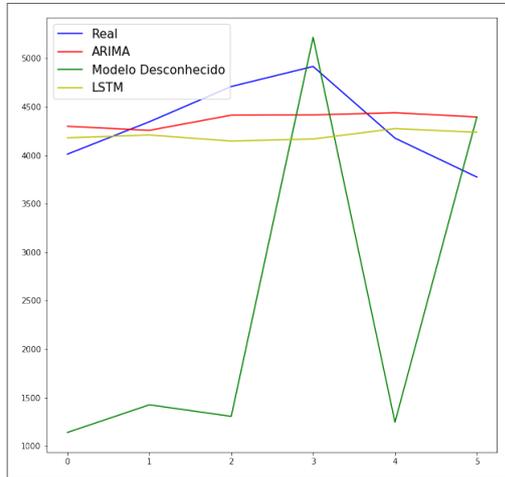
Na primeira rodada de testes com o LSTM utilizou-se apenas os dados numéricos das séries temporais, transformados pela técnica *sliding window* descrita na seção 4.2, sem a

<sup>3</sup> <https://keras.io/>

<sup>4</sup> <https://www.tensorflow.org/>

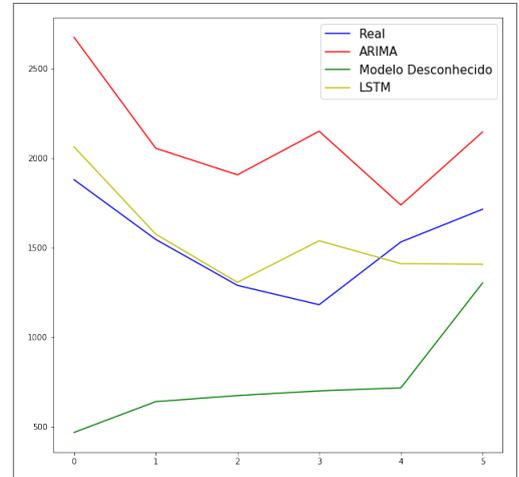
utilização de qualquer outro tipo de variável. As Figuras 41, 42, 43, 44, 45 e a Tabela 4 apresentam os resultados obtidos.

Figura 41 – Predição com o modelo LSTM do Produto 1.



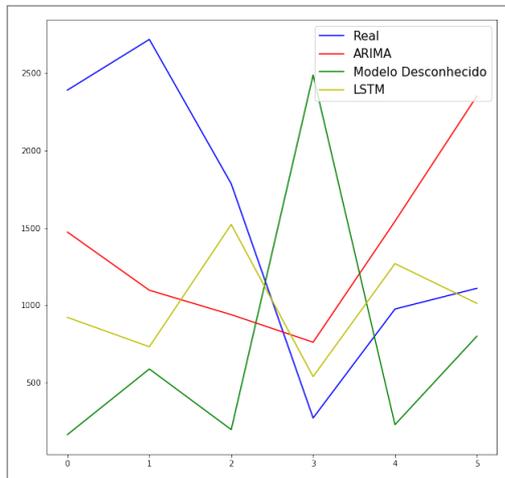
Fonte: Autor

Figura 42 – Predição com o modelo LSTM do Produto 2.



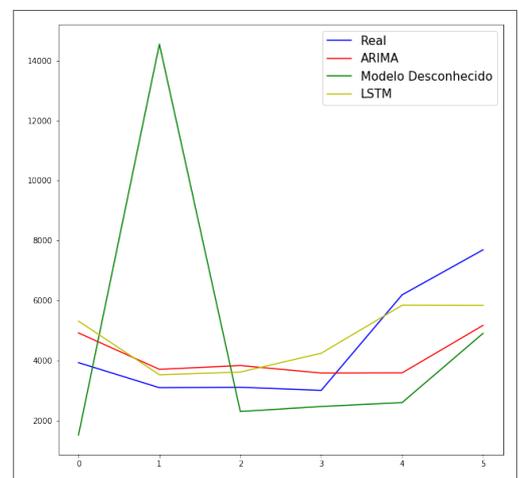
Fonte: Autor

Figura 43 – Predição com o modelo LSTM do Produto 3.



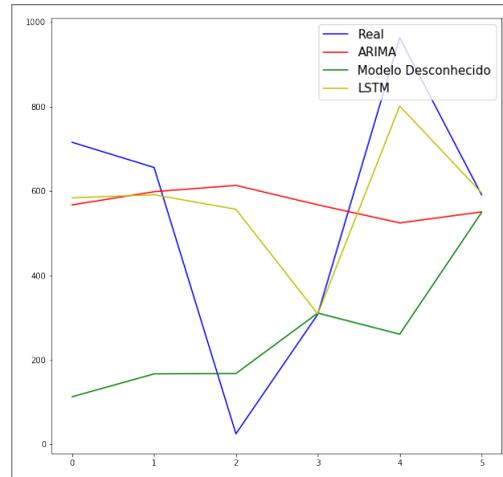
Fonte: Autor

Figura 44 – Predição com o modelo LSTM do Produto 4.



Fonte: Autor

Figura 45 – Predição com o modelo LSTM do Produto 5.



Fonte: Autor

Fonte: Autor

Produto	LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	362,78	8,18	437,22	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>169,31</b>	<b>11,52</b>	<b>212,8</b>	588,2	40,66	637,87	773,7	49,94	842,56
Produto 3	<b>729,90</b>	<b>47,81</b>	<b>1028,62</b>	948,04	82,72	1024,69	1537,3	197,13	1712,99
Produto 4	<b>960,03</b>	<b>22,71</b>	<b>1114,06</b>	1337,79	27,04	1598,53	3598,22	94,9	5141,81
Produto 5	<b>148,96</b>	356,83	<b>234,45</b>	255,13	413,98	324,4	329,94	<b>133,68</b>	431,17

Tabela 4 – Resultados das previsões com LSTM

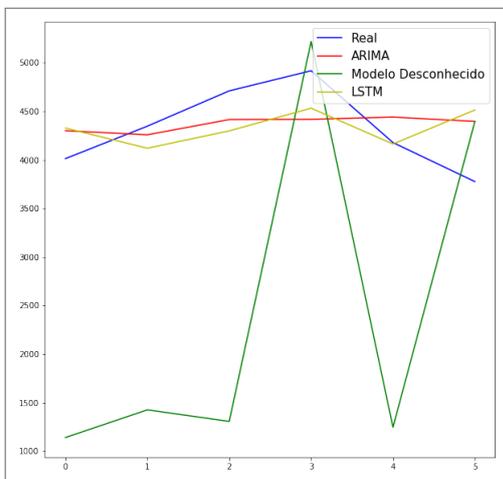
Como é possível ver nos resultados acima, já na primeira rodada foi possível obter melhorias consideráveis com a utilização do LSTM em detrimento dos resultados obtidos pelo modelo ARIMA e pelo *Modelo em Operação*. Dentre os resultados obtidos, o modelo LSTM não obteve melhorias apenas para o resultado do Produto 1 em comparação com o ARIMA. Observando a Figura 20 é possível notar que a série referente a esse produto apresenta um comportamento mais estável, onde componentes como a sazonalidade são mais claros de serem visualizados que nas outras séries. Nesse sentido, isso pode justificar o bom resultado obtido pelo modelo ARIMA que tende a fazer boas previsões com séries mais estáveis e de comportamento linear. Porém, quando olhamos para as séries que apresentam comportamentos mais irregulares, como é o caso do Produto 2 (Figura 21) que começa com valores zerados e só depois vão se ajustando, e do Produto 3 (Figura 22) que possui alguns pontos discrepantes em sua área central, o modelo LSTM apresentou uma melhor capacidade de lidar com tais irregularidades e consequentemente apresentou taxas de erros menores que as apresentadas pelos outros modelos.

Com essa primeira abordagem concluída, outras abordagens ainda com a utilização do LSTM foram investigadas a fim de encontrar outras formas de obter melhores resultados. As próximas subseções descrevem cada uma das abordagens que foram investigadas.

#### 4.5.2.1 Remoção da tendência e sazonalidade

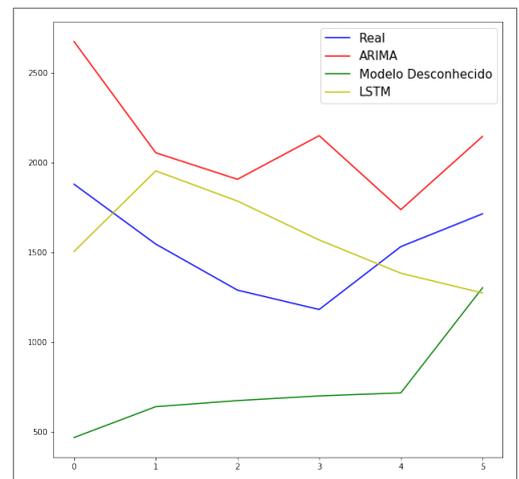
Após a primeira rodada de testes, buscou-se identificar outras formas de tentar melhorar os resultados das previsões já obtidas. Nesse sentido, foi investigado se a remoção dos componentes das séries auxiliariam na melhora desses resultados. Para tal feito foram removidas a tendência e a sazonalidade de cada uma das séries e foi refeito o processo de treinamento. A remoção foi feita subtraindo os valores da série pelos valores dos componentes que foram obtidos no processo de decomposição mostrado na subseção 4.1.1, e posteriormente o treinamento foi realizado com o resultado dessa subtração. Após obtido os resultados das previsões, foram somados a eles os valores dos componentes retirados para se obter o valor real da previsão. As Tabelas 5 e 6 e as Figuras 46, 47, 48, 49, 50, 51, 52, 53, 54 e 55 mostram os resultados obtidos através desse processo.

Figura 46 – Predição com o modelo LSTM do Produto 1 (sem tendência).



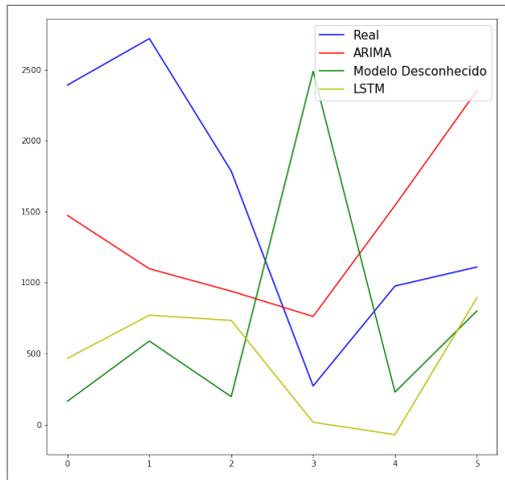
Fonte: Autor

Figura 47 – Predição com o modelo LSTM do Produto 2 (sem tendência).



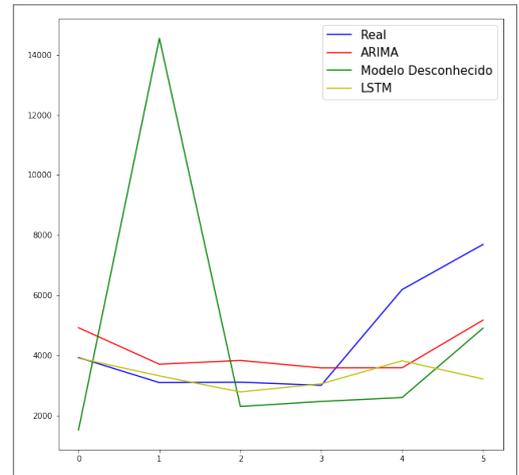
Fonte: Autor

Figura 48 – Predição com o modelo LSTM do Produto 3 (sem tendência).



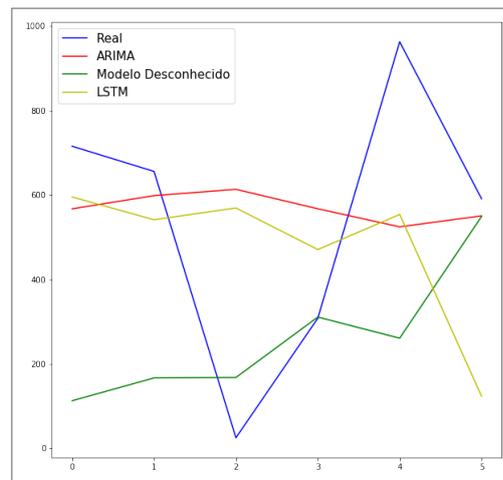
Fonte: Autor

Figura 49 – Predição com o modelo LSTM do Produto 4 (sem tendência).



Fonte: Autor

Figura 50 – Predição com o modelo LSTM do Produto 5 (sem tendência).



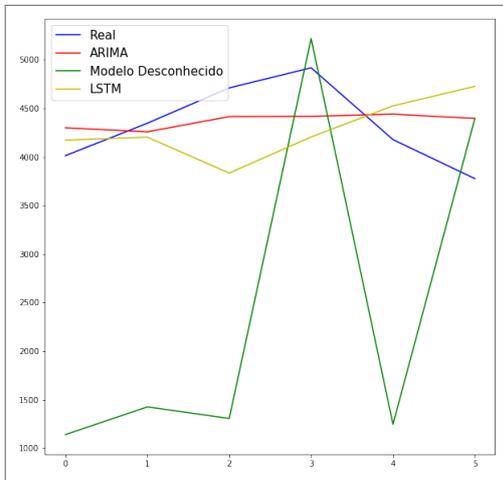
Fonte: Autor

Fonte: Autor

Produto	LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	348,38	8,31	410,72	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>376,15</b>	<b>18,78</b>	<b>391,73</b>	588,2	40,66	637,87	773,7	49,94	842,56
Produto 3	1074,29	<b>82,15</b>	1279,38	<b>948,04</b>	82,72	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	<b>1244,28</b>	<b>25,79</b>	2074,04	1337,79	27,04	<b>1598,53</b>	3598,22	94,9	5141,81
Produto 5	302,78	330,14	349,98	<b>255,13</b>	413,98	<b>324,4</b>	329,94	<b>133,68</b>	431,17

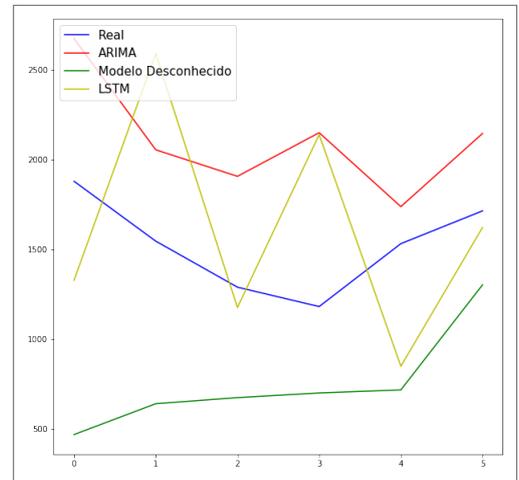
Tabela 5 – Resultados das predições do LSTM com a remoção da tendência

Figura 51 – Predição com o modelo LSTM do Produto 1 (sem sazonalidade).



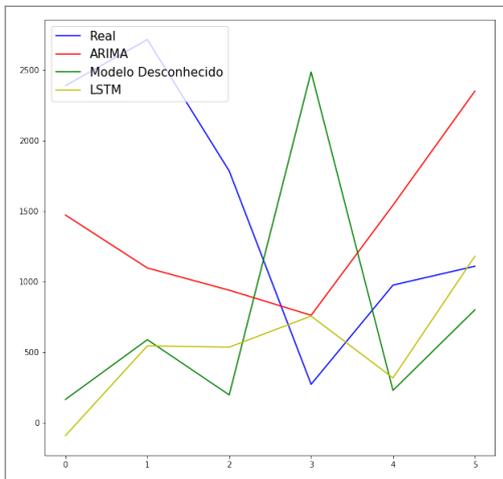
Fonte: Autor

Figura 52 – Predição com o modelo LSTM do Produto 2 (sem sazonalidade).



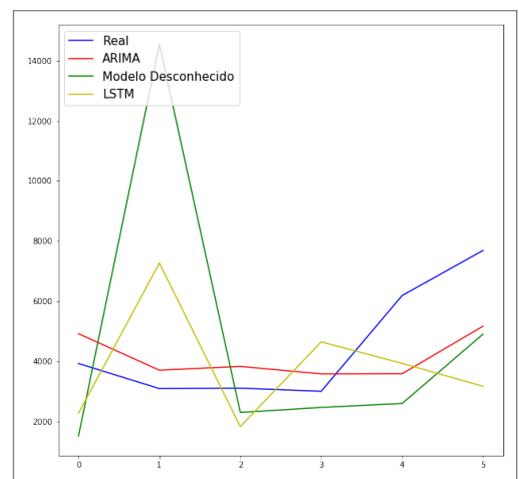
Fonte: Autor

Figura 53 – Predição com o modelo LSTM do Produto 3 (sem sazonalidade).



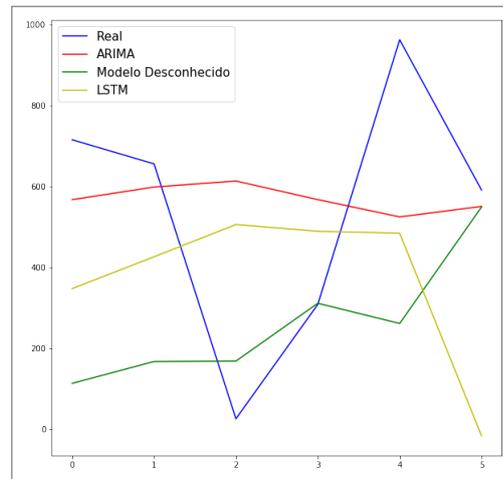
Fonte: Autor

Figura 54 – Predição com o modelo LSTM do Produto 4 (sem sazonalidade).



Fonte: Autor

Figura 55 – Predição com o modelo LSTM do Produto 5 (sem sazonalidade).



Fonte: Autor

Fonte: Autor

Produto	LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	531,78	9,12	625,26	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>574,11</b>	<b>35,59</b>	683,65	588,2	40,66	<b>637,87</b>	773,7	49,94	842,56
Produto 3	1187,44	<b>79,37</b>	1479,89	<b>948,04</b>	82,72	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	2588,54	45,47	2888,13	<b>1337,79</b>	<b>27,04</b>	<b>1598,53</b>	3598,22	94,9	5141,81
Produto 5	390,76	278,26	418,21	<b>255,13</b>	413,98	<b>324,4</b>	329,94	<b>133,68</b>	431,17

Tabela 6 – Resultados das predições do LSTM com a remoção da sazonalidade

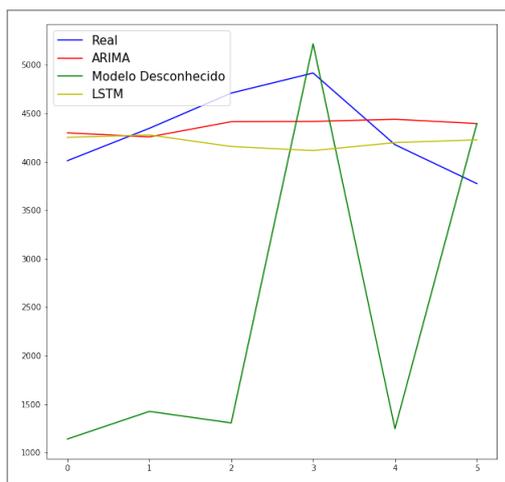
Nos resultados descritos acima é possível observar que de maneira geral a remoção dos componentes de tendência e sazonalidade não trouxeram melhorias efetivas. Quando comparamos com os resultados já obtidos anteriormente pelo LSTM sem a remoção desses componentes é notório um aumento nas taxas de erro, tornando também em vários casos esses resultados inferiores aos obtidos pelo ARIMA.

A remoção desses componentes é comumente utilizada em abordagens que utilizam o modelo ARIMA a fim de deixar a série linear para que seja possível obter um melhor aproveitamento do modelo. Nesse sentido, buscou-se identificar se esse processo de deixar a série com características lineares também seria capaz de auxiliar no processo de aprendizado do LSTM. Para os testes realizados observou-se que os resultados tiveram uma piora. Isso pode ser justificado pelo fato de que algoritmos baseados em RNA são capazes de lidar com dados não lineares, e com a remoção desses componentes é possível que tenham se perdido características que seriam importantes para o desenvolvimento do modelo, comprometendo assim os resultados das predições feitas por ele.

#### 4.5.2.2 Adição de variáveis temporais

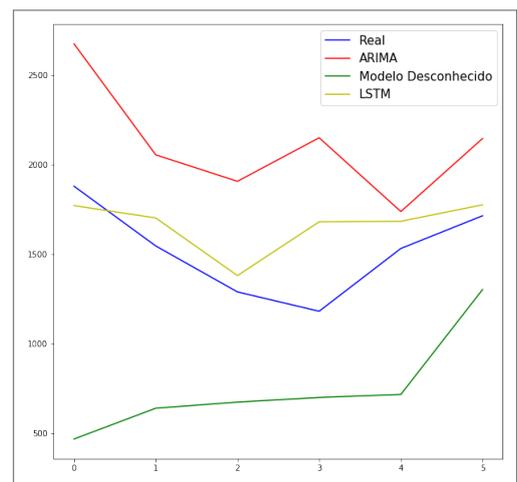
Modelos de aprendizagem profunda, como é o caso do LSTM, tendem a obter melhores resultados quando trabalhados com números maiores de dados pois são capazes de extrair representações mais complexas (NAJAFABADI et al., 2015). Como o número de dados nos *datasets* é relativamente pequeno, buscou-se incrementar a quantidade de variáveis a fim de criar novos padrões que pudessem ser identificados no processo de modelagem. Para isso foram criadas variáveis que representassem o tempo em que o dado foi registrado. Sendo assim, além das variáveis obtidas pelo processo de *sliding window*, foram adicionadas variáveis referentes ao mês, número da semana no ano e estação do ano. Foram criadas colunas com valores binários dessas variáveis utilizando a técnica de *One-Hot Encode* descrita na seção 4.2. Os resultados obtidos através desse processo encontra-se representado nas Figuras 56, 57, 58, 59, 60 abaixo e na Tabela 7.

Figura 56 – Predição com o modelo LSTM do Produto 1 (variáveis temporais).



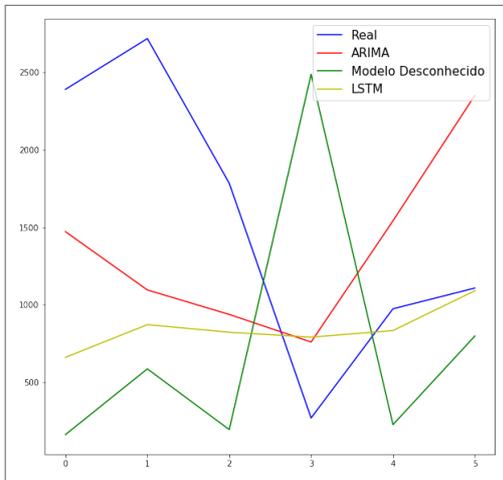
Fonte: Autor

Figura 57 – Predição com o modelo LSTM do Produto 2 (variáveis temporais).



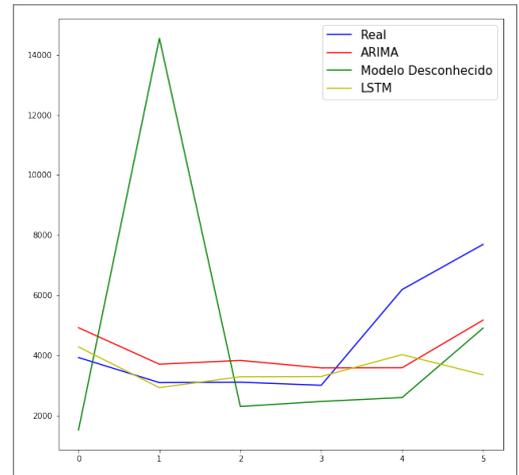
Fonte: Autor

Figura 58 – Predição com o modelo LSTM do Produto 3 (variáveis temporais).



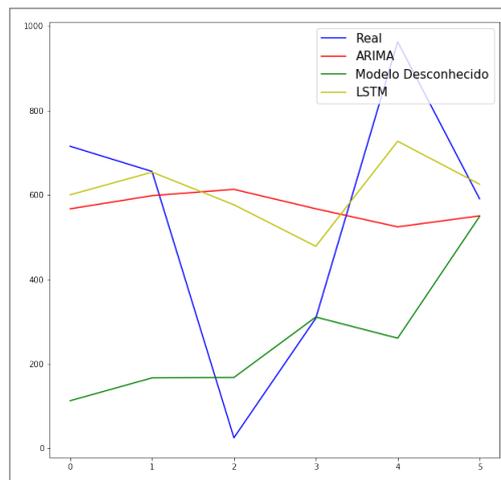
Fonte: Autor

Figura 59 – Predição com o modelo LSTM do Produto 4 (variáveis temporais).



Fonte: Autor

Figura 60 – Predição com o modelo LSTM do Produto 5 (variáveis temporais).



Fonte: Autor

Fonte: Autor

Produto	LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	355,79	<b>8,01</b>	449,45	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>178,08</b>	<b>13,13</b>	<b>231,25</b>	588,2	40,66	637,87	773,7	49,94	842,56
Produto 3	<b>869,25</b>	<b>67,03</b>	1126,47	948,04	82,72	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	<b>1248,95</b>	<b>20,2</b>	1990,07	1337,79	27,04	<b>1598,53</b>	3598,22	94,9	5141,81
Produto 5	<b>184,69</b>	379,34	<b>259,12</b>	255,13	413,98	324,4	329,94	<b>133,68</b>	431,17

Tabela 7 – Resultados das previsões do LSTM com a adição de variáveis temporais

Nos resultados obtidos é possível observar que a adição dessas variáveis não trouxeram uma melhora efetiva para todos os resultados, porém em alguns casos é possível observar alguns pequenos indícios de melhora, como são os casos do Produto 1 e do Produto 4 que tiveram o MAPE reduzido em comparação com outros resultados já obtidos nas outras etapas. Isso pode ser um indicativo de que essas variáveis podem ser melhores exploradas em outras situações. Como já citado anteriormente, algoritmos de aprendizagem profunda trabalham bem com um número grande de dados, o que não é o caso dos *datasets* utilizados. Sendo assim, é possível que as novas variáveis adicionadas não tenham tido representação suficiente para extração de padrões consistentes pelo LSTM. Logo nota-se a possibilidade de que com a utilização de *datasets* maiores, onde haja uma maior representação dessas informações, o uso de variáveis temporais pode ser uma abordagem válida de ser empregada, sendo capaz de trazer alguma melhora para os resultados.

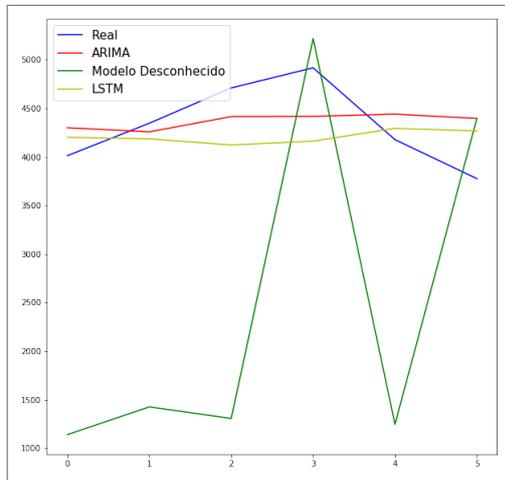
#### 4.5.2.3 Adição de variáveis externas

Ainda buscando encontrar melhores resultados para as predições, deu-se a possibilidade de utilização de novas variáveis que viessem a enriquecer o modelo. Em conversa com especialista, foi pontuado que por se tratar de dados referentes a demanda de bebidas, talvez a utilização de variáveis referentes ao clima pudessem ter algum impacto positivo na construção dos modelos. Sendo assim, buscou-se esses dados no Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) mantido pelo Instituto Nacional de Meteorologia (INMET)<sup>5</sup>. Através dessa fonte foram obtidos dados diários de temperatura, umidade e insolação que posteriormente foram utilizados para retreinar o modelo. Uma vez que o INMET fornece os dados com um padrão de tempo diário e os dados dos *datasets* são semanais, utilizou-se a média dos dias referentes a cada semana das variáveis coletadas. As Figuras 61, 62, 63, 64, 65 e a Tabela 8 mostram os resultados obtidos com a utilização dessas novas variáveis.

---

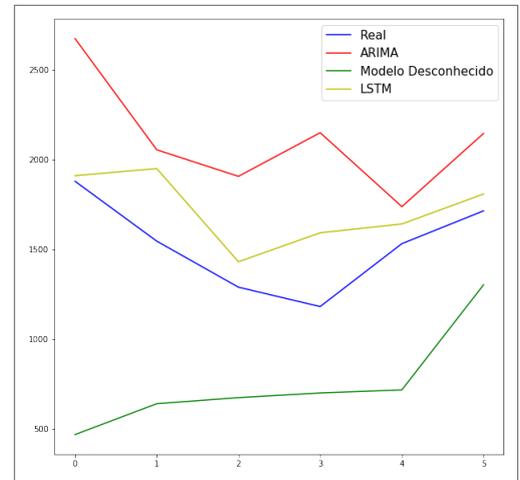
<sup>5</sup> <http://www.inmet.gov.br/portal/>

Figura 61 – Predição com o modelo LSTM do Produto 1 (variáveis externas).



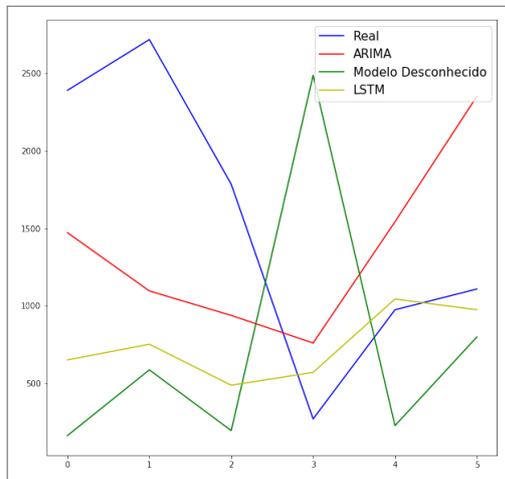
Fonte: Autor

Figura 62 – Predição com o modelo LSTM do Produto 2 (variáveis externas).



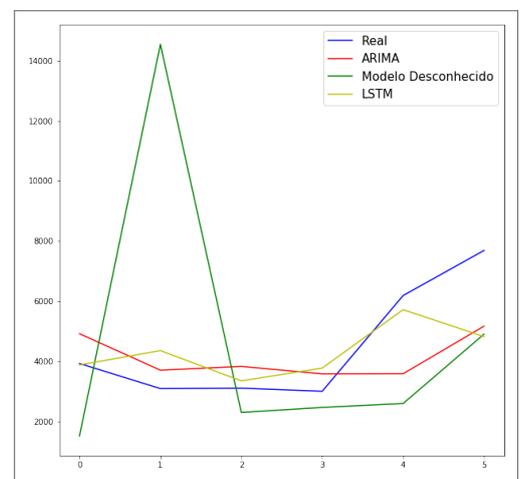
Fonte: Autor

Figura 63 – Predição com o modelo LSTM do Produto 3 (variáveis externas).



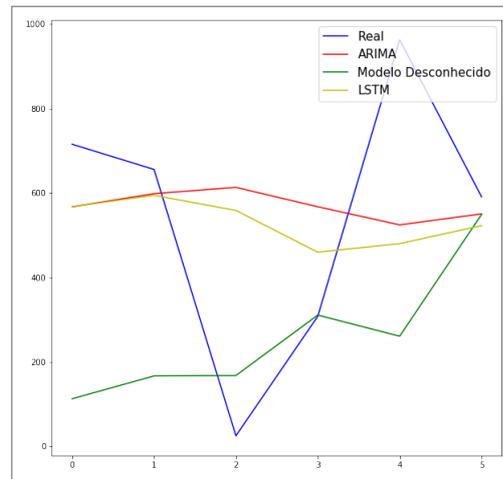
Fonte: Autor

Figura 64 – Predição com o modelo LSTM do Produto 4 (variáveis externas).



Fonte: Autor

Figura 65 – Predição com o modelo LSTM do Produto 5 (variáveis externas).



Fonte: Autor

Fonte: Autor

Produto	LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	383,16	8,67	453,03	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>198,66</b>	<b>14,38</b>	<b>249,79</b>	588,2	40,66	637,87	773,7	49,94	842,56
Produto 3	<b>917,71</b>	<b>57,94</b>	1203,2	948,04	82,72	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	<b>943,68</b>	<b>20,05</b>	<b>1335,22</b>	1337,79	27,04	1598,53	3598,22	94,9	5141,81
Produto 5	<b>240,73</b>	373,93	<b>308,29</b>	255,13	413,98	324,4	329,94	<b>133,68</b>	431,17

Tabela 8 – Resultados das predições do LSTM com a adição de variáveis externas

Assim como na adição de variáveis temporais, a adição de variáveis externas não apresentou melhorias para todos os modelos, onde apenas o modelo referente ao Produto 4 apresentou uma melhora quando comparado aos resultados já anteriormente obtidos. O problema que ocorreu com a abordagem anterior é o mesmo que pode ter ocorrido com essa, onde a falta de uma maior representação dessas variáveis devido ao tamanho pequeno dos *datasets* pode ter impossibilitado o LSTM de extrair padrões que pudessem melhorar os resultados das predições. Sugere-se que esse possa ser um dos problemas porque o uso dessas variáveis não foi ruim para todos os casos, mostrando uma viabilidade de exploração delas em outros contextos.

Nessa etapa finalizou-se os testes realizados com o LSTM em sua forma nativa. Na segunda fase dos experimentos utilizou-se o LSTM com uma camada de CNN, formando então o modelo híbrido CNN-LSTM. A próxima seção irá descrever os experimentos realizados com esse modelo híbrido e os resultados obtidos por ele.

## 4.6 MODELAGEM COM CNN-LSTM

Assim como no LSTM, para a modelagem do CNN-LSTM foi utilizada a biblioteca *Keras* utilizando o *TensorFlow* em seu *background*, todos utilizando *Python 3.6* como linguagem de programação. A próxima subseção descreve as características da implementação e da arquitetura utilizada no desenvolvimento do modelo.

### 4.6.1 Arquitetura do CNN-LSTM

Listing 4.2 – Modelo CNN-LSTM em Python

```

1 def fit_model_cnn_lstm(X, y):
    model = Sequential()
3     model.add(TimeDistributed(Conv1D(filters=5, kernel_size=1, activation='relu'),
        input_shape=(None, 1, 1)))
    model.add(TimeDistributed(MaxPooling1D(pool_size=1)))
5     model.add(TimeDistributed(Flatten()))
    model.add(LSTM(100, activation='relu'))
7     model.add(Dense(1))
    model.compile(optimizer='adam', loss='mse')
9     model.fit(X, y, epochs=2000, verbose=0)
    return model

```

No trecho de código acima é possível ver claramente a arquitetura demonstrada pela Figura 14. Primeiramente temos o modelo CNN que é composto pelas camadas de convolução 1D e *max-pooling* 1D. Na camada de convolução foi definido um filtro de tamanho 5, com um *kernel* de tamanho 1 e a ReLU como função de ativação. Na camada do *max-pooling* foi definido um *pool* de tamanho 1. Após isso foi definida a camada *Flatten* que é responsável por comprimir os resultados obtidos pelo CNN em um vetor unidimensional que servirá de entrada para o LSTM. Por fim, temos a camada do LSTM que possui as mesmas configurações já mostradas na seção anterior.

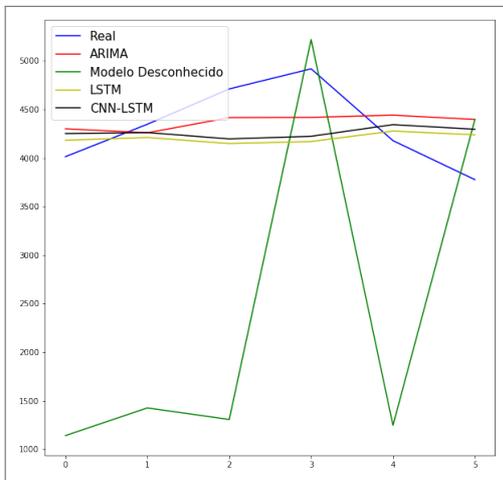
A próxima subseção apresenta os detalhes dos testes realizados e dos resultados obtidos com a utilização do modelo implementado.

### 4.6.2 Resultados com CNN-LSTM

Os testes realizados com o CNN-LSTM seguiram as mesmas etapas usadas nos testes com o LSTM. Sendo assim, cada tabela de resultados além de apresentar os resultados obtidos pelo CNN-LSTM, apresentará também os resultados obtidos pelo LSTM que utilizaram a mesma abordagem que estará sendo descrita a fim de facilitar a comparação entre os resultados, além dos resultados obtidos pelo ARIMA e pelo *Modelo em Operação*.

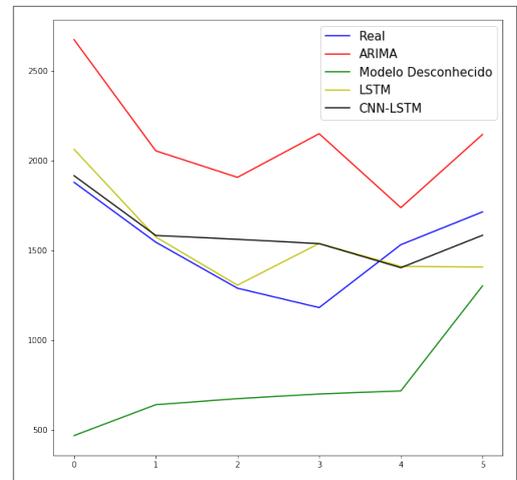
Nessa primeira etapa de testes do CNN-LSTM utilizou-se apenas os valores numéricos das séries aplicados a técnica de *sliding window* já explicada anteriormente. As Figuras 66, 67, 68, 69, 70 e a Tabela 9 apresentam os resultados obtidos através da modelagem com o CNN-LSTM.

Figura 66 – Predição com o modelo CNN-LSTM do Produto 1.



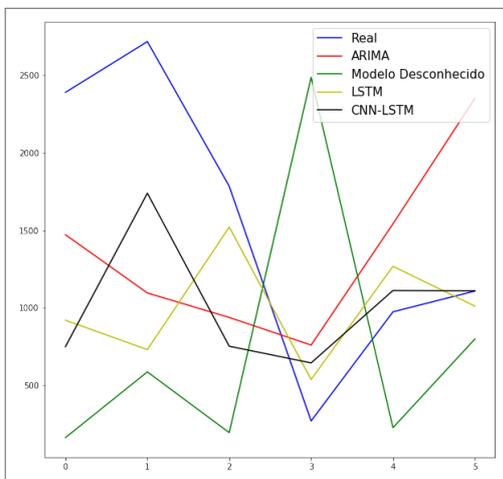
Fonte: Autor

Figura 67 – Predição com o modelo CNN-LSTM do Produto 2.



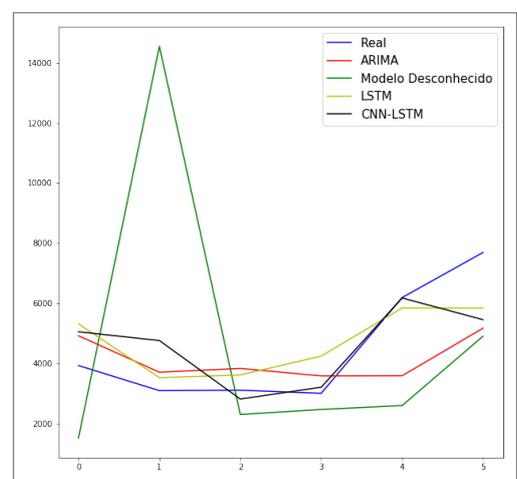
Fonte: Autor

Figura 68 – Predição com o modelo CNN-LSTM do Produto 3.



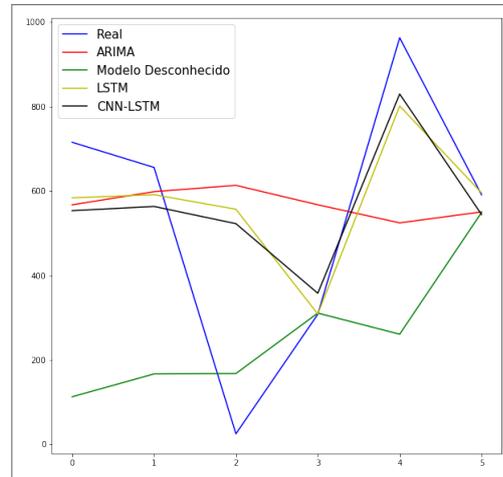
Fonte: Autor

Figura 69 – Predição com o modelo CNN-LSTM do Produto 4.



Fonte: Autor

Figura 70 – Predição com o modelo CNN-LSTM do Produto 5.



Fonte: Autor

Fonte: Autor

Produto	CNN-LSTM			LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	369,51	8,44	429,78	362,78	8,18	437,22	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>160,3</b>	11,95	<b>198,99</b>	169,31	<b>11,52</b>	212,8	588,2	40,66	637,87	773,7	49,94	842,56
Produto 3	<b>694,23</b>	52,47	<b>901,23</b>	729,90	<b>47,81</b>	1028,62	948,04	82,72	1024,69	1537,3	197,13	1712,99
Produto 4	<b>922,99</b>	<b>21,32</b>	1236,12	960,03	22,71	<b>1114,06</b>	1337,79	27,04	1598,53	3598,22	94,9	5141,81
Produto 5	163,49	339,19	<b>225,2</b>	<b>148,96</b>	356,83	234,45	255,13	413,98	324,4	329,94	<b>133,68</b>	431,17

Tabela 9 – Resultados das predições com CNN-LSTM

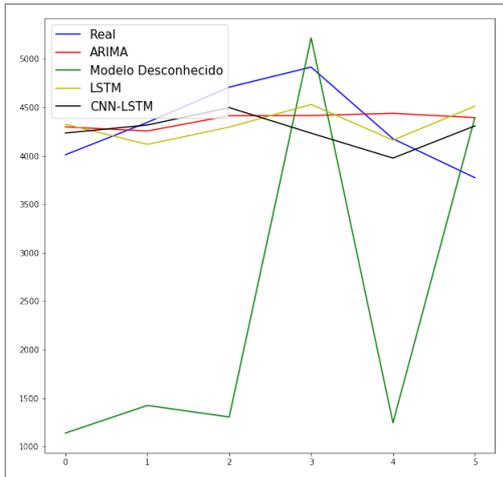
Quando observa-se os resultados acima, é notório algumas melhorias sutis nas métricas de erro com a utilização do CNN-LSTM em comparação com o LSTM. É possível ver que em quase todos os modelos houve alguma melhora em algumas das métricas, com exceção mais uma vez do modelo referente ao Produto 1 que manteve o seu melhor resultado com a utilização do ARIMA, reforçando a suposição de que o modelo ARIMA consegue obter resultados satisfatórios quando modela séries que possuem comportamentos mais estáveis. O CNN é um modelo normalmente utilizado para classificação de imagens, ou seja, é um modelo projetado para ser utilizado com um número grande de dados. Observa-se que mesmo com uma quantidade pequena de dados como a utilizada nesses testes, ele foi capaz de extrair características que passadas para a camada LSTM trouxeram algumas melhorias para os resultados. Nesse sentido, acredita-se que a utilização desse modelo em *datasets* com um número maior de registros pode ser capaz de proporcionar melhorias mais significantes.

#### 4.6.2.1 Remoção da tendência e sazonalidade

Com os detalhes já descritos na seção anterior, essa etapa consistiu na remoção dos componentes de tendência e sazonalidade das séries a fim de deixá-las com características mais

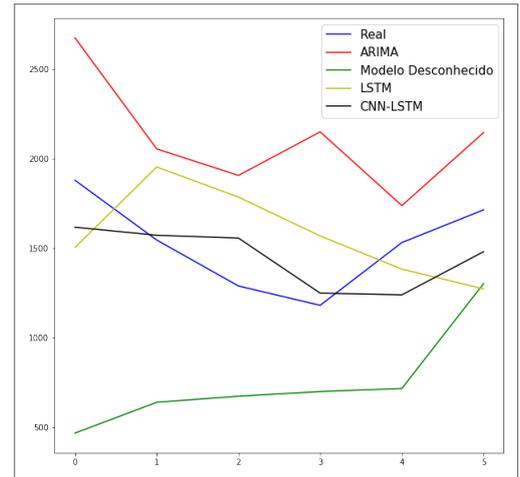
lineares, objetivando identificar se isso facilitaria o processo de aprendizagem do modelo e consequentemente obteria melhores resultados. Uma vez realizado esse processo os dados foram aplicados ao CNN-LSTM para modelagem. As figuras abaixo e as Tabelas 10 e 11 mostram os resultados obtidos com essa abordagem.

Figura 71 – Predição com o modelo CNN-LSTM do Produto 1 (sem tendência).



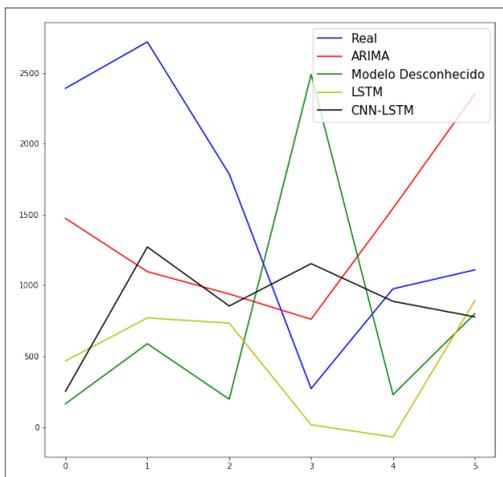
Fonte: Autor

Figura 72 – Predição com o modelo CNN-LSTM do Produto 2 (sem tendência).



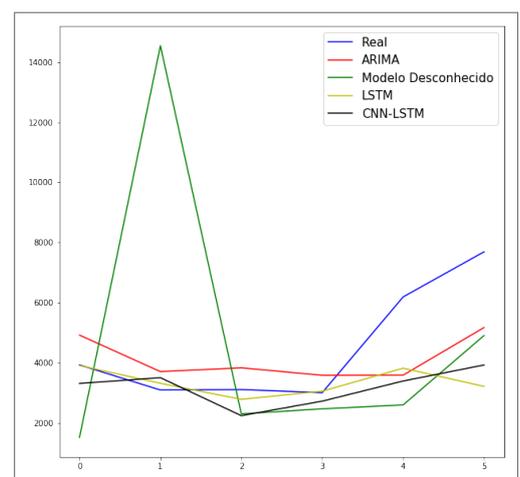
Fonte: Autor

Figura 73 – Predição com o modelo CNN-LSTM do Produto 3 (sem tendência).



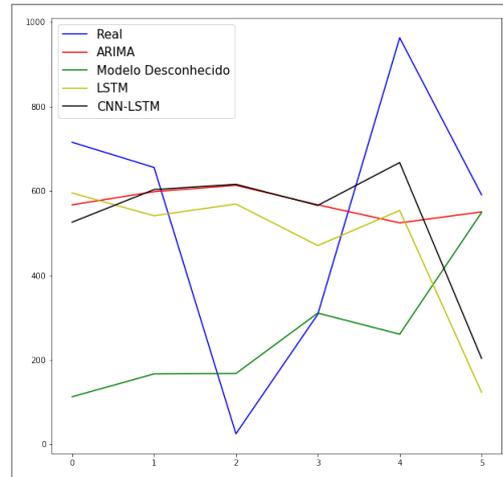
Fonte: Autor

Figura 74 – Predição com o modelo CNN-LSTM do Produto 4 (sem tendência).



Fonte: Autor

Figura 75 – Predição com o modelo CNN-LSTM do Produto 5 (sem tendência).



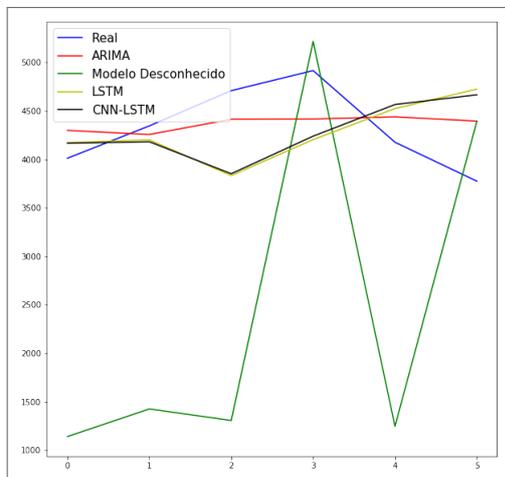
Fonte: Autor

Fonte: Autor

Produto	CNN-LSTM			LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	<b>313</b>	8,16	384,23	348,38	8,31	410,72	342,17	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>191,72</b>	<b>15,47</b>	<b>218,16</b>	376,15	18,78	391,73	588,2	40,66	637,87	773,7	49,94	842,56
Produto 3	970,28	<b>76,79</b>	1186,05	1074,29	82,15	1279,38	<b>948,04</b>	82,72	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	1455,99	29,25	1973,39	<b>1244,28</b>	<b>25,79</b>	2074,04	1337,79	27,04	<b>1598,53</b>	3598,22	94,9	5141,81
Produto 5	295,29	363,6	339,2	302,78	330,14	349,98	<b>255,13</b>	413,98	<b>324,4</b>	329,94	<b>133,68</b>	431,17

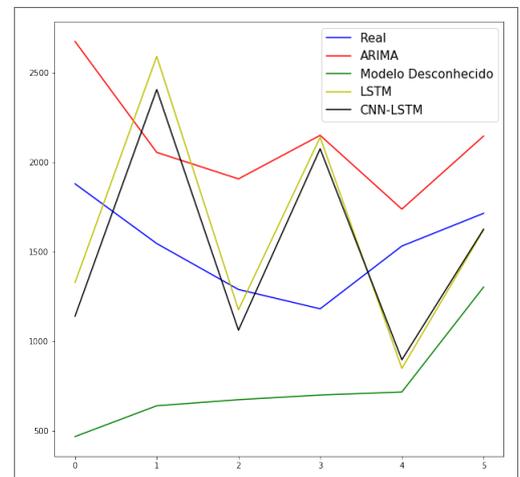
Tabela 10 – Resultados das previsões do CNN-LSTM com a remoção da tendência

Figura 76 – Predição com o modelo CNN-LSTM do Produto 1 (sem sazonalidade).



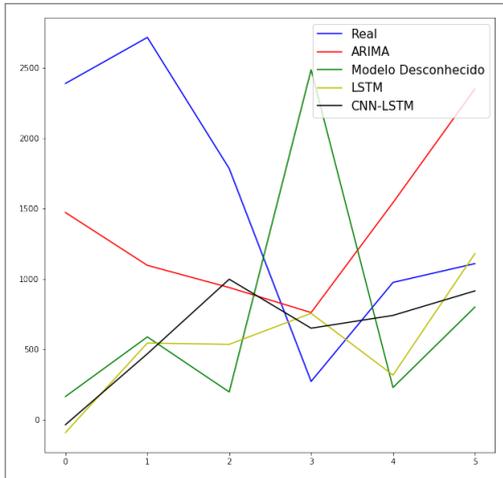
Fonte: Autor

Figura 77 – Predição com o modelo CNN-LSTM do Produto 2 (sem sazonalidade).



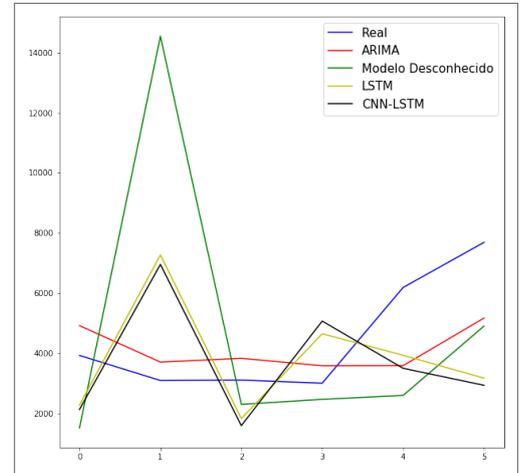
Fonte: Autor

Figura 78 – Predição com o modelo CNN-LSTM do Produto 3 (sem sazonalidade).



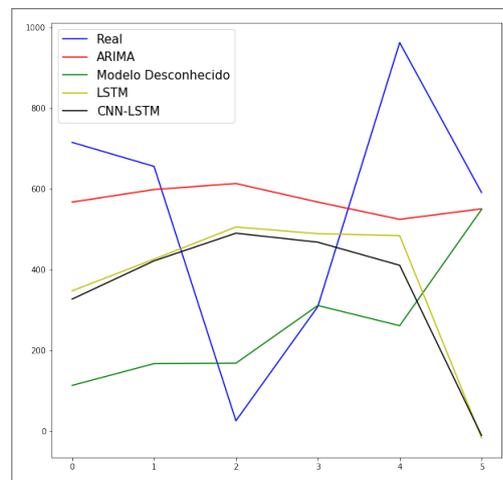
Fonte: Autor

Figura 79 – Predição com o modelo CNN-LSTM do Produto 4 (sem sazonalidade).



Fonte: Autor

Figura 80 – Predição com o modelo CNN-LSTM do Produto 5 (sem sazonalidade).



Fonte: Autor

Fonte: Autor

Produto	CNN-LSTM			LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	522,29	9,04	604,17	531,78	9,12	625,26	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	574,2	<b>34,79</b>	651,78	<b>574,11</b>	35,59	683,65	588,2	40,66	<b>637,87</b>	773,7	49,94	842,56
Produto 3	1045,73	<b>77,4</b>	1403,56	1187,44	79,37	1479,89	<b>948,04</b>	82,72	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	2781,89	47	3016,27	2588,54	45,47	2888,13	<b>1337,79</b>	<b>27,04</b>	<b>1598,53</b>	3598,22	94,9	5141,81
Produto 5	400,09	263,66	430,89	390,76	278,26	418,21	<b>255,13</b>	413,98	<b>324,4</b>	329,94	<b>133,68</b>	431,17

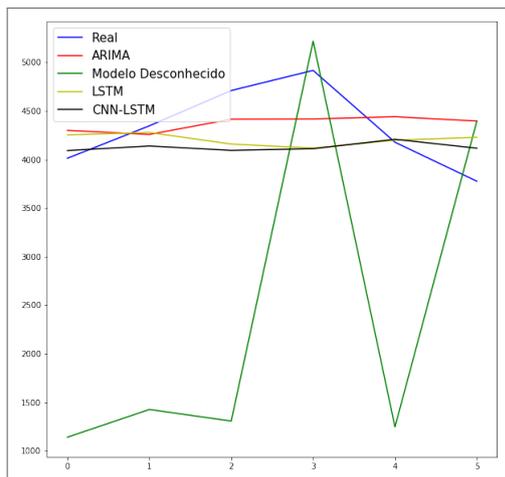
Tabela 11 – Resultados das previsões do CNN-LSTM com a remoção da sazonalidade

Com os resultados mostrados acima é possível ratificar a suposição já feita nos testes com o LSTM de que a remoção desses componentes pode não ser uma boa alternativa, uma vez que características importantes podem estar sendo perdidas e possivelmente prejudicando um melhor desempenho do modelo. Quando comparamos esses resultados com os outros resultados em que não houve a remoção desses componentes, é possível observar um aumento nos valores das métricas de erro. Porém, um fato interessante de ser observado é que quando comparamos os resultados dos modelos CNN-LSTM e LSTM que utilizaram essa mesma abordagem (Tabela 10 e Tabela 11) é possível ver em vários casos uma melhora nos resultados com a utilização do modelo CNN-LSTM, mostrando assim uma possível eficiência da utilização da camada CNN para extração de características, melhorando as previsões feitas pela camada LSTM.

#### 4.6.2.2 Adição de variáveis temporais

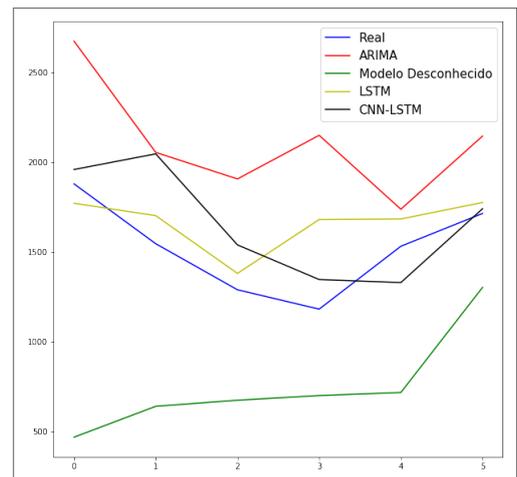
Assim como no experimentos realizados com o LSTM, para o CNN-LSTM foram utilizadas variáveis referentes ao tempo dos dados a fim de enriquecer os padrões entre eles. Foram adicionadas variáveis referentes ao mês, número da semana no ano e estação do ano. A Tabela 12 e as Figuras 81, 82, 83, 84 e 85 mostram os resultados obtidos.

Figura 81 – Predição com o modelo CNN-LSTM do Produto 1 (variáveis temporais).



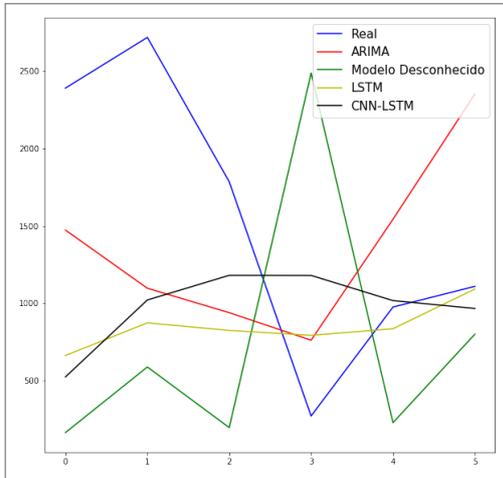
Fonte: Autor

Figura 82 – Predição com o modelo CNN-LSTM do Produto 2 (variáveis temporais).



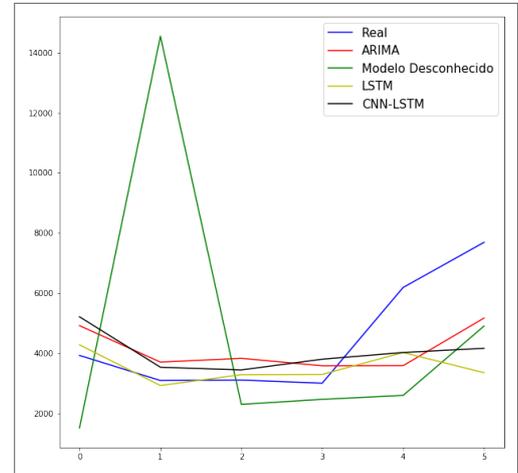
Fonte: Autor

Figura 83 – Predição com o modelo CNN-LSTM do Produto 3 (variáveis temporais).



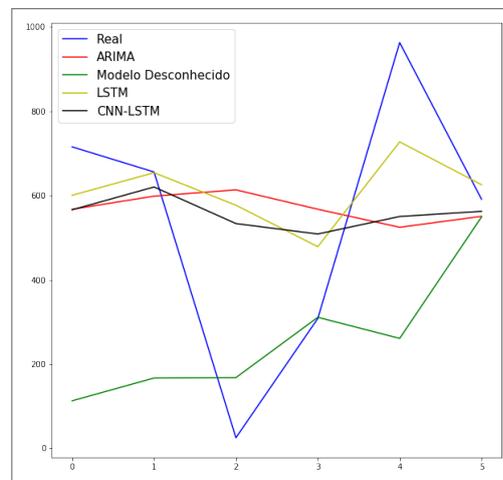
Fonte: Autor

Figura 84 – Predição com o modelo CNN-LSTM do Produto 4 (variáveis temporais).



Fonte: Autor

Figura 85 – Predição com o modelo CNN-LSTM do Produto 5 (variáveis temporais).



Fonte: Autor

Fonte: Autor

Produto	CNN-LSTM			LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	346,67	<b>7,66</b>	446,48	355,79	8,01	449,45	<b>342,17</b>	8,05	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	204,17	14,14	254,55	<b>178,08</b>	<b>13,13</b>	<b>231,25</b>	588,2	40,66	637,87	773,7	49,94	842,56
Produto 3	877,34	87,79	1124,37	<b>869,25</b>	<b>67,03</b>	1126,47	948,04	82,72	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	1424,99	27,51	1812,09	<b>1248,95</b>	<b>20,2</b>	1990,07	1337,79	27,04	<b>1598,53</b>	3598,22	94,9	5141,81
Produto 5	222,34	356,96	286,52	<b>184,69</b>	379,34	<b>259,12</b>	255,13	413,98	324,4	329,94	<b>133,68</b>	431,17

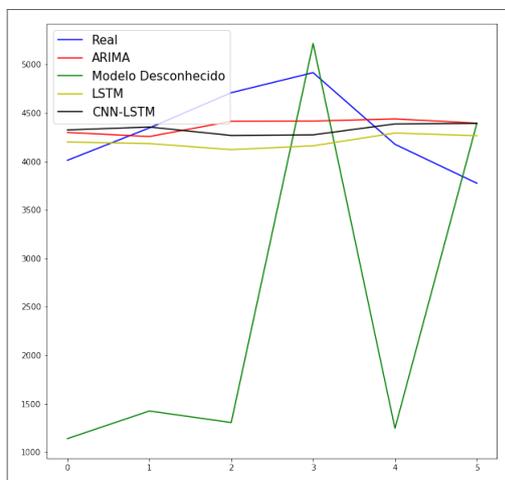
Tabela 12 – Resultados das predições do CNN-LSTM com a adição de variáveis temporais

Os resultados obtidos mostraram que a utilização dessas variáveis não trouxeram melhorias expressivas para a maioria dos modelos utilizando o CNN-LSTM, nem mesmo quando comparados com os resultados do LSTM utilizando essa mesma abordagem (Tabela 12). O único modelo que apresentou algum indício de melhora foi o modelo referente ao Produto 1 que teve o seu MAPE reduzido, o que é um fato interessante, pois até então essa foi a única abordagem apresentada que conseguiu reduzir o MAPE para esse modelo em comparação com o obtido pelo modelo ARIMA. Através dos fatos observados acredita-se que por se tratar de séries temporais, algumas características inerentes ao tempo estão de certa forma implícitas nos dados das séries, sendo assim é possível que a camada do CNN não tenha extraído das variáveis temporais características relevantes o suficiente para trazer alguma melhora para os resultados. Uma outra suposição que pode justificar uma falta de melhora dos resultados com a utilização dessa abordagem é mais uma vez a ausência de uma maior representação dessas informações, uma vez que os *datasets* utilizados são relativamente pequenos e não possuem dados suficientes para uma melhor generalização dessas informações.

#### 4.6.2.3 Adição de variáveis externas

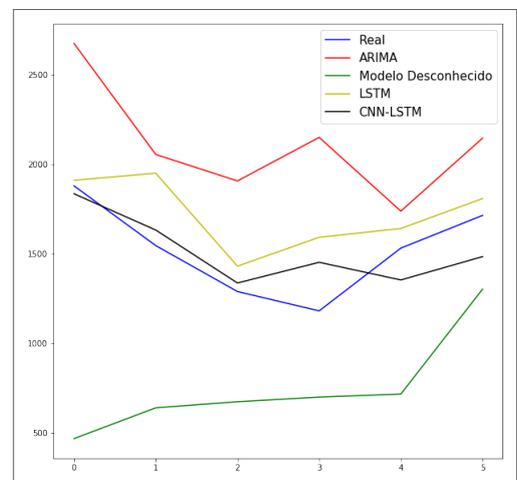
Com os detalhes já descritos na seção anterior, essa etapa do processo consistiu em utilizar variáveis externas obtidas pelo INMET a fim de enriquecer os padrões nos dados. Sendo assim foram adicionadas aos dados variáveis climáticas referentes a temperatura, umidade e insolação. As Figuras 86, 87, 88, 89, 90 e a Tabela 13 mostram os resultados obtidos pelo CNN-LSTM com a adição das variáveis externas.

Figura 86 – Predição com o modelo CNN-LSTM do Produto 1 (variáveis externas).



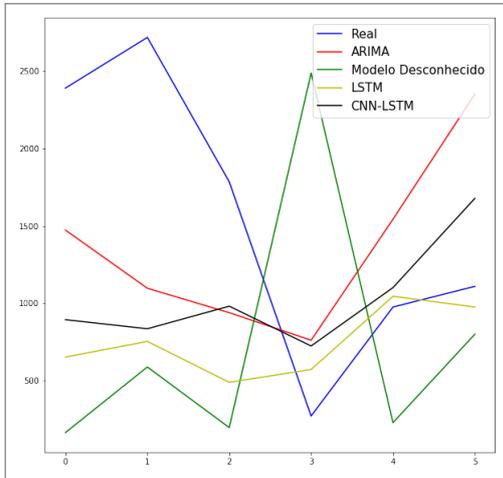
Fonte: Autor

Figura 87 – Predição com o modelo CNN-LSTM do Produto 2 (variáveis externas).



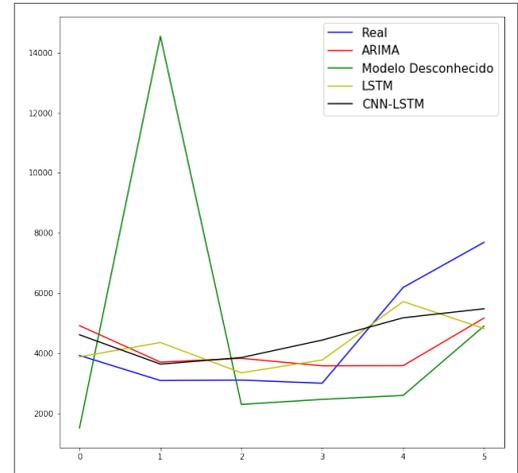
Fonte: Autor

Figura 88 – Predição com o modelo CNN-LSTM do Produto 3 (variáveis externas).



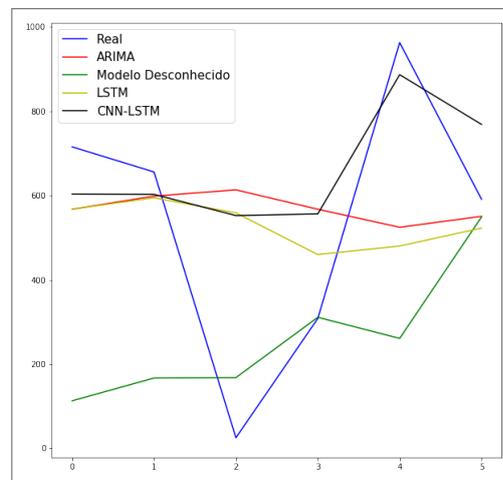
Fonte: Autor

Figura 89 – Predição com o modelo CNN-LSTM do Produto 4 (variáveis externas).



Fonte: Autor

Figura 90 – Predição com o modelo CNN-LSTM do Produto 5 (variáveis externas).



Fonte: Autor

Fonte: Autor

Produto	CNN-LSTM			LSTM			ARIMA			Modelo em Operação		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Produto 1	372,37	8,64	434,18	383,16	8,67	453,03	<b>342,17</b>	<b>8,05</b>	<b>383,01</b>	2174,85	50,62	2497,89
Produto 2	<b>143,05</b>	<b>9,95</b>	<b>168,36</b>	198,66	14,38	249,79	588,2	40,66	637,87	773,7	49,94	842,56
Produto 3	<b>888,92</b>	69	1078,92	917,71	<b>57,94</b>	1203,2	948,04	82,72	<b>1024,69</b>	1537,3	197,13	1712,99
Produto 4	1105,74	25,31	<b>1244,16</b>	<b>943,68</b>	<b>20,05</b>	1335,22	1337,79	27,04	1598,53	3598,22	94,9	5141,81
Produto 5	<b>198,95</b>	369,96	<b>255,54</b>	240,73	373,93	308,29	255,13	413,98	324,4	329,94	<b>133,68</b>	431,17

Tabela 13 – Resultados das predições do CNN-LSTM com a adição de variáveis externas

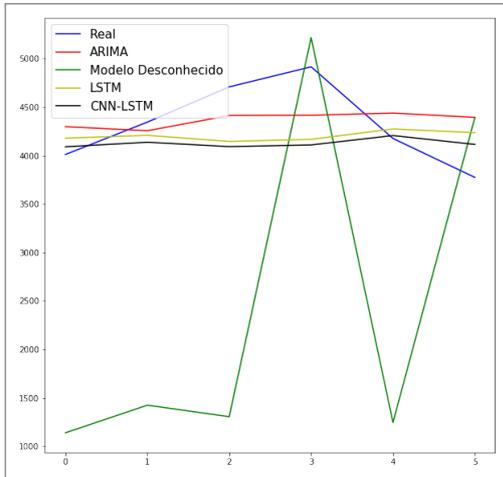
Nos resultados mostrados acima é possível observar que com a utilização das variáveis referentes a dados de clima o CNN-LSTM conseguiu obter algumas melhorias nos resultados, com destaque para o modelo referente ao Produto 2 que obteve o melhor resultado entre os resultados das abordagens já apresentadas. Parte expressiva dos resultados obtidos pelo CNN-LSTM tiveram melhorias em pelo menos uma das métricas de erro com a utilização dessas variáveis quando comparadas com os resultados obtidos pelo LSTM utilizando essas mesmas variáveis (Tabela 13). Tais observações denotam mais uma vez que a utilização da camada de CNN para extração de características pode ser benéfica para obtenção de melhores resultados.

#### 4.7 RESUMO DOS RESULTADOS

De maneira geral é possível observar que os modelos baseados em LSTM trouxeram resultados bastante satisfatórios nos experimentos realizados. Nas séries com comportamentos mais irregulares os modelos LSTM mostraram um bom desempenho em detrimento dos modelos ARIMA e do *Modelo em Operação*, mostrando sua capacidade em lidar com essas irregularidades. A utilização da camada de CNN também se mostrou efetiva em alguns casos, onde observou-se que talvez com a utilização de *datasets* com número maiores de dados a utilização dessa camada possa ser ainda mais relevante.

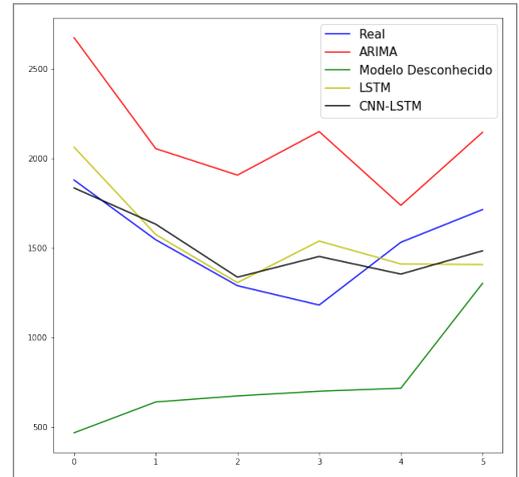
Um ponto que se pôde observar é que a abordagem de remoção dos componentes da série temporal não trouxe melhorias significativas para os resultados, onde em muitos casos eles pioraram de forma considerável. Isso pode ser explicado pelo fato de que com a remoção desses componentes a série perde características que podem ser importantes para o processo de aprendizagem dos modelos baseados em LSTM, uma vez que ao contrário do modelo ARIMA, esse modelo consegue lidar melhor com características não lineares. A utilização de novas variáveis também se mostrou um ponto positivo no desenvolvimento de alguns modelos que tiveram seus resultados melhorados. Buscando simplificar a visualização dos ganhos obtidos com as abordagens investigadas nessa pesquisa, as figuras abaixo apresentam a plotagem do melhor resultado obtido por cada um dos modelos, e a Tabela 14 traz uma comparação desses resultados, mostrando a porcentagem de aumento ou diminuição da taxa de erro em comparação com os resultados obtidos pelo modelo ARIMA e pelo *Modelo em Operação*. Para definição dos melhores resultados foram escolhidos os modelos que apresentaram o menor MAPE, com exceção do modelo referente ao Produto 5 que nesse caso em específico o valor do MAE foi utilizado para definir o modelo com a menor taxa de erro, porém o MAPE continuará sendo utilizado para realizar a comparação feita na Tabela 14. Essa mudança se deu pelo fato de que os dados de teste desse produto apresentam um valor de outlier (observar o ponto 2 do eixo x na Figura 90) que eleva o valor do MAPE, porém quando observamos o valor das outras métricas vemos que esse erro distorce a real qualidade do modelo.

Figura 91 – Melhores resultados do Produto 1.



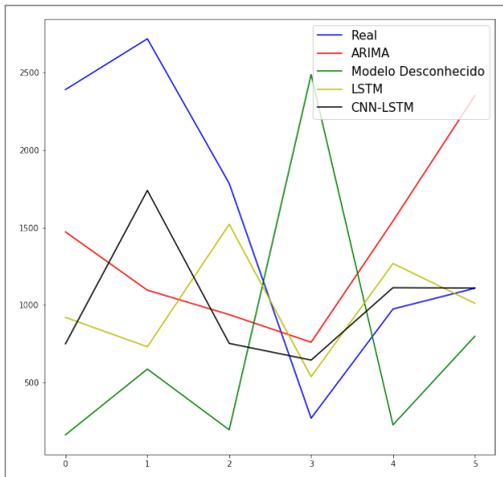
Fonte: Autor

Figura 92 – Melhores resultados do Produto 2.



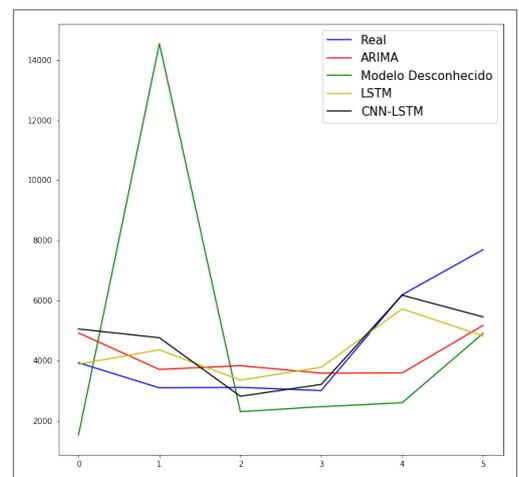
Fonte: Autor

Figura 93 – Melhores resultados do Produto 3.



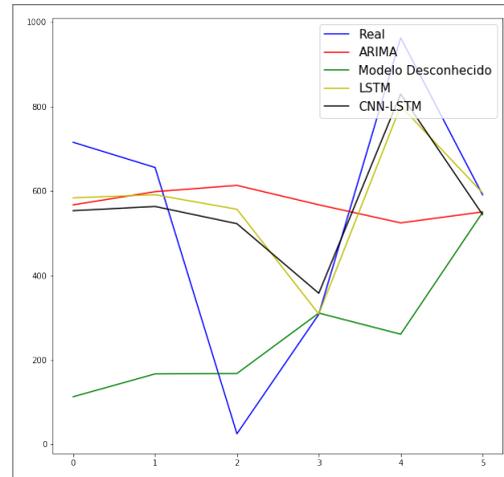
Fonte: Autor

Figura 94 – Melhores resultados do Produto 4.



Fonte: Autor

Figura 95 – Melhores resultados do Produto 5.



Fonte: Autor

Fonte: Autor

Produto	ARIMA	Modelo em Operação	LSTM			CNN-LSTM		
	MAPE	MAPE	MAPE	Redução em % (ARIMA)	Redução em % (Modelo em Operação)	MAPE	Redução em % (ARIMA)	Redução em % (Modelo em Operação)
Produto 1	8,05	50,62	8,18	+1,61%	-83,84%	7,66	-4,84%	-84,87%
Produto 2	40,66	49,94	11,52	-71,67%	-76,93%	9,95	-75,53%	-80,08%
Produto 3	82,72	197,13	47,81	-42,2%	-75,75%	52,47	-36,57%	-73,38%
Produto 4	27,04	94,9	20,05	-25,85%	-78,87%	21,32	-21,15%	-77,53%
Produto 5	413,98	133,68	356,83	-13,8%	+166,93%	339,19	-18,07%	+153,73%

Tabela 14 – Comparação dos resultados das previsões

Por fim, observado os dados acima, conclui-se que para os experimentos realizados a utilização de dos modelos LSTM e CNN-LSTM apresentaram melhores resultados em comparação com os modelos ARIMA e o *Modelo em Operação*.

#### 4.7.1 Comparações de resultados extra

Em paralelo com este trabalho, Delgado (2020) desenvolveu uma pesquisa semelhante utilizando os mesmos conjuntos de dados para treinamentos e testes. A principal diferença de abordagem utilizada por Delgado (2020) se caracteriza pela utilização de diferentes algoritmos de regressão, tais como Árvores de Decisão, *XGBoost*, entre outros. As tabelas abaixo apresentam em sua última linha os melhores resultados obtidos pelo trabalho desenvolvido por Delgado (2020), e as demais linhas mostram os melhores resultados alcançados por essa pesquisa.

<b>Algoritmo</b>	<b>MAPE</b>
ARIMA	8,05
Modelo em Operação	50,62
LSTM	8,18
CNN-LSTM	7,66
Decision Tree	6,56

Tabela 15 – Resultados Comparação Produto 1

<b>Algoritmo</b>	<b>MAPE</b>
ARIMA	40,66
Modelo em Operação	49,94
LSTM	11,52
CNN-LSTM	9,95
XGBoost	11,85

Tabela 16 – Resultados Comparação Produto 2

<b>Algoritmo</b>	<b>MAPE</b>
ARIMA	82,72
Modelo em Operação	197,13
LSTM	47,81
CNN-LSTM	52,47
MLP	70,95

Tabela 17 – Resultados Comparação Produto 3

<b>Algoritmo</b>	<b>MAPE</b>
ARIMA	27,04
Modelo em Operação	94,9
LSTM	27,05
CNN-LSTM	21,32
SGD	27,82

Tabela 18 – Resultados Comparação Produto 4

---

<b>Algoritmo</b>	<b>MAPE</b>
ARIMA	413,98
Modelo em Operação	133,68
LSTM	356,83
CNN-LSTM	339,19
SVR	110,0

Tabela 19 – Resultados Comparação Produto 5

Em 60% dos *datasets* trabalhados os modelos baseados em LSTMs conseguiram obter um melhor desempenho em comparação com os resultados obtidos por Delgado (2020). Baseado nos fatos apresentados, a utilização de LSTMs mostrou potencial para análise e previsão de séries temporais, sendo assim, a exploração desses modelos se mostrou um caminho viável na busca por melhores resultados.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

O desenvolvimento desta pesquisa teve como objetivo avaliar o desempenho de RNNs aplicadas ao contexto de previsão de demanda, mais especificamente o modelo LSTM e o modelo híbrido CNN-LSTM. Para o desenvolvimento e avaliação dos modelos foram utilizados cinco *datasets* diferentes que correspondem a dados reais de demanda de uma grande empresa do ramo de bebidas. A utilização desses *datasets* acabou também sendo um desafio para a pesquisa devido a sua pequena quantidade de registros, sendo apenas cento e quarenta por *dataset*.

Baseado na limitação do número de registros dos *datasets*, essa pesquisa também buscou explorar diferentes pré-processamentos dos dados que viessem a enriquecê-los e aumentar os padrões entre eles. Para isso foram utilizadas técnicas de remoção dos componentes das séries temporais, mais especificamente os componentes de tendência e sazonalidade, onde na maioria dos casos a utilização dessa abordagem se mostrou ineficaz, causando um aumento nas taxas de erros já anteriormente obtidas sem a utilização dessa abordagem. Nesse sentido, observou-se que a remoção de tais componentes pode ter tirado características que seriam importantes para o processo de aprendizado dos modelos de aprendizagem profunda.

Posteriormente foi explorado a adição de novas variáveis, onde primeiramente buscou-se adicionar variáveis temporais, tais como mês, número da semana e estações do ano. Essa abordagem acabou não trazendo melhorias significativas, onde foi identificado que a pouca quantidade de registros pode ter afetado a criação de padrões mais fortes que viessem a impactar positivamente o modelo. Em seguida buscou-se fazer a utilização de variáveis externas aos *datasets*, onde para isso foram utilizados dados referentes a temperatura, umidade e insolação, fornecidos pelo INMET. A utilização dessas variáveis se mostrou interessante em alguns casos onde foi possível obter algumas melhorias nos resultados. A utilização do modelo híbrido CNN-LSTM também se apresentou eficaz em algumas situações com a utilização dessas variáveis, mostrando uma boa capacidade da camada de CNN em extrair características que auxiliam na aprendizagem do modelo.

De maneira geral concluiu-se que a utilização dos modelos LSTM e CNN-LSTM trouxeram resultados significativos para os experimentos realizados nessa pesquisa. Para chegar a tal conclusão os resultados foram comparados com dois outros resultados fornecidos por duas empresas diferentes, sendo a primeira uma empresa conceituada nacionalmente no ramo de inteligência artificial, e a segunda sendo a empresa fornecedora dos produtos referentes aos dados que foram trabalhados. A primeira empresa utilizou o modelo ARIMA e a segunda não se conhece o modelo utilizado. Em comparação com os resultados obtidos pelo modelo ARIMA, os modelos desenvolvidos nessa pesquisa apresentaram reduções nas taxas de erro de até 75,53%, e em comparação com os resultados desenvolvidos pelo

modelo da Empresa Fornecedora dos Produtos essa redução chegou até a 84,87%.

Com base nos resultados apresentados, a contribuição desse trabalho se destaca na possibilidade de obtenção de resultados expressivos com a utilização das abordagens apresentadas no contexto previsão de demanda de curto prazo, além de fornecer mais uma contribuição para uma área de aplicação que possui ainda um número limitado de trabalhos relacionados.

Como trabalhos futuros e extensão dessa pesquisa destaca-se a investigação e exploração de diferentes arquiteturas dos modelos utilizados, uma vez que isso não foi abordado de maneira profunda e pode ser um bom caminho para aperfeiçoamento dos resultados. Além disso, é pretendido investigar a possibilidade de melhorias nos resultados utilizando a combinação de dados quantitativos com dados qualitativos, uma vez que durante essa pesquisa foi possível observar a utilização desse tipo de informação poderia ter ajudado nos resultados dos modelos. Outra possibilidade é a realização de experimentos com *datasets* maiores, uma vez que foi identificado que isso poderia estar comprometendo os resultados obtidos. Por fim, a investigação e exploração de novas variáveis que venham a aumentar os padrões e conseqüentemente desenvolver modelos mais robustos.

## REFERÊNCIAS

- ABBES, A. B.; BOUNOUH, O.; FARAH, I. R.; JONG, R. de; MARTÍNEZ, B. Comparative study of three satellite image time-series decomposition methods for vegetation change detection. *European Journal of Remote Sensing*, Taylor & Francis, v. 51, n. 1, p. 607–615, 2018.
- ABONAZEL, M. R.; ABD-ELFTAH, A. I. Forecasting egyptian gdp using arima models. 2019.
- BABICH, L.; SVALOV, D.; SMIRNOV, A.; BABICH, M. Industrial power consumption forecasting methods comparison. In: IEEE. *2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)*. [S.l.], 2019. p. 307–309.
- BALLOU, R. H. *Gerenciamento da Cadeia de Suprimentos-: Logística Empresarial*. [S.l.]: Bookman Editora, 2009.
- BOTTOU, L. Stochastic gradient descent tricks. In: *Neural networks: Tricks of the trade*. [S.l.]: Springer, 2012. p. 421–436.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. Time series analysis: forecasting and control holden-day san francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day 1970*, 1970.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to time series and forecasting*. [S.l.]: springer, 2016.
- BROWNLEE, J. *Stationarity in time series analysis*. 2016. <<https://machinelearningmastery.com/time-series-data-stationary-python/>>. Acessado: 30/12/2019.
- CERDA, P.; VAROQUAUX, G.; KÉGL, B. Similarity encoding for learning with dirty categorical variables. *Machine Learning*, Springer, v. 107, n. 8-10, p. 1477–1494, 2018.
- CHATFIELD, C. *Time-series forecasting*. [S.l.]: Chapman and Hall/CRC, 2000.
- CORRÊA, H. L.; GIANESI, I. G. *Just in Time, MRP II E OPT: Um Enfoque Estratégico*. [S.l.]: Editora Atlas SA, 2000.
- DELGADO, A. J. *Análise de Métodos de Regressão para Previsão de Demanda de Curto Prazo*. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de Pernambuco, 2020.
- DONG, M.; SUN, J. Partial discharge detection on aerial covered conductors using time-series decomposition and long short-term memory network. *arXiv preprint arXiv:1907.03378*, 2019.
- EHLERS, R. S. Análise de séries temporais. *Universidade Federal do Paraná*, 2007.
- FU, R.; ZHANG, Z.; LI, L. Using lstm and gru neural network methods for traffic flow prediction. In: IEEE. *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. [S.l.], 2016. p. 324–328.

- GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2672–2680.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural computation*, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.
- HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, Elsevier, v. 20, n. 1, p. 5–10, 2004.
- HU, P.; TONG, J.; WANG, J.; YANG, Y.; TURCI, L. de O. A hybrid model based on cnn and bi-lstm for urban water demand prediction. In: IEEE. *2019 IEEE Congress on Evolutionary Computation (CEC)*. [S.l.], 2019. p. 1088–1094.
- ISMAIL, A. A.; WOOD, T.; BRAVO, H. C. Improving long-horizon forecasts with expectation-biased lstm networks. *arXiv preprint arXiv:1804.06776*, 2018.
- JUNIOR, A. P. S.; SEGATTO, M.; CALIA, R. C.; FIOROTTO, J. A.; BERTON, L. T. A tecnologia da informação como suporte ao ajuste da previsão da demanda: um estudo de caso em uma empresa de bebidas carbonatadas. *Revista Produção Online*, v. 10, n. 3, p. 621–648, 2010.
- KAISER, R.; MARAVALL, A. Notes on time serie analysis, arima models and signal extraction. 2000.
- KHOSLA, S. *CNN / Introduction to Pooling Layer*. 2016. <<https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/>>. Acessado: 30/12/2019.
- KIM, T.-Y.; CHO, S.-B. Predicting residential energy consumption using cnn-lstm neural networks. *Energy*, Elsevier, 2019.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015.
- MARTÍNEZ-ÁLVAREZ, F.; TRONCOSO, A.; CORTÉS, G.; RIQUELME, J. A survey on data mining techniques applied to energy time series forecasting. *Energies*, In press, 11 2015.
- MCLAREN, C. H.; ZHANG, X. M. The importance of trend-cycle analysis for national statistics institutes. *Estudios de economía aplicada*, Asociación Española de Economía Aplicada, ASEPELT, v. 28, n. 3, p. 607–624, 2010.
- MENDONÇA, S. A. T. de; RUZZÃO, A. P. de A.; SANTOS, J. E. dos; AZADINHO, L. Z. O planejamento estratégico como ferramenta: Um estudo sobre a eficiência das micro e pequenas empresas brasileiras. *Administração de Empresas em Revista*, v. 16, n. 17, p. 50–68, 2017.

- MOREIRA, D. A. *Administração da produção e operações*. [S.l.]: Editora Saraiva, 2017.
- NAJAFABADI, M. M.; VILLANUSTRE, F.; KHOSHGOFTAAR, T. M.; SELIYA, N.; WALD, R.; MUHAREMAGIC, E. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, Springer, v. 2, n. 1, p. 1, 2015.
- NOUREEN, S.; ATIQUE, S.; ROY, V.; BAYNE, S. Analysis and application of seasonal arima model in energy demand forecasting: A case study of small scale agricultural load. In: IEEE. *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*. [S.l.], 2019. p. 521–524.
- OZAKI, T. On the order determination of arima models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 26, n. 3, p. 290–301, 1977.
- PACHECO, R. F.; SILVA, A. V. F. Aplicação de modelos quantitativos de previsão em uma empresa de transporte ferroviário. *ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, Ouro Preto*, 2003.
- PALACHY, S. *Stationarity in time series analysis*. 2019. <<https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322>>. Acessado: 30/12/2019.
- PELLEGRINI, F. R. Metodologia para implementação de sistemas de previsão de demanda. *Master's thesis in Production Engineering. Porto Alegre: UFRGS*, 2000.
- PELLEGRINI, F. R.; FOGLIATTO, F. S. Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso. *Production*, SciELO Brasil, v. 11, n. 1, p. 43–64, 2001.
- PEREIRA, P. T. do V. Previsão de demanda dos consumidores: como fazer? *Santa Catarina: SEBRAE*, 2017.
- PINHEIRO, J. F. D.; NETO, M. N. F. Fatores que contribuem para mortalidade das micro e pequenas empresas no brasil/factors contributing to mortality of micro and small enterprises in brazil. *Brazilian Journal of Development*, v. 5, n. 7, p. 11107–11122, 2019.
- PINTO, L. R. L.; MAZZON, J. A. *Proposição de um modelo de simulação de vendas: uma análise dos influenciadores das vendas da cesta de bens de consumo auditada pela Nielsen*. [S.l.]: São Paulo: Universidade de São Paulo, 2010.
- REIS, M. Análise de séries temporais. *Departamento de informática e Estatística. Universidade Federal de Santa Catarina. Disponível em: <http://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf>. Acessado em, v. 2, 2015.*
- REN, S.; CHAN, H.-L.; RAM, P. A comparative study on fashion demand forecasting models with multiple sources of uncertainty. *Annals of Operations Research*, Springer, v. 257, n. 1-2, p. 335–355, 2017.
- SAINATH, T. N.; VINYALS, O.; SENIOR, A.; SAK, H. Convolutional, long short-term memory, fully connected deep neural networks. In: IEEE. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.], 2015. p. 4580–4584.
- SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural networks*, Elsevier, v. 61, p. 85–117, 2015.

- 
- SEBRAE, S. B. D. A. À. Sobrevivência das empresas no brasil. *Coleção Estudos e Pesquisas*. Brasília: SEBRAE, 2016.
- SELVIN, S.; VINAYAKUMAR, R.; GOPALAKRISHNAN, E.; MENON, V. K.; SOMAN, K. Stock price prediction using lstm, rnn and cnn-sliding window model. In: IEEE. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. [S.l.], 2017. p. 1643–1647.
- SHAW, J. Neural network resource guide. *AI expert*, v. 8, n. 2, p. 48–54, 1992.
- SIAMI-NAMINI, S.; NAMIN, A. S. Forecasting economics and financial time series: Arima vs. lstm. *arXiv preprint arXiv:1803.06386*, 2018.
- SILVA, J. C.; FIGUEIREDO, M. C.; BRAGA, A. C. Demand forecasting: A case study in the food industry. In: SPRINGER. *International Conference on Computational Science and Its Applications*. [S.l.], 2019. p. 50–63.
- SOUZA, R. S. de; TRIERWEILLER, A. C.; WEISE, A. D.; ROCHA, R. A. da; MONTEIRO, M. I. Previsão da demanda como suporte para o planejamento e controle da produção na sigma-indústria eletro eletrônica. *Revista Ingepro-inovação, gestão e produção*, v. 3, n. 4, 2010.
- TUBINO, D. F. *Manual de planejamento e controle da produção*. [S.l.]: Atlas, 2000.
- VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. *Remote sensing of Environment*, Elsevier, v. 114, n. 1, p. 106–115, 2010.
- WANG, J.; DUGGASANI, A. Forecasting hotel reservations with long short-term memory-based recurrent neural networks. *International Journal of Data Science and Analytics*, Springer, p. 1–18, 2018.
- WANG, Z.; LOU, Y. Hydrological time series forecast model based on wavelet de-noising and arima-lstm. In: IEEE. *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. [S.l.], 2019. p. 1697–1701.
- WHEELWRIGHT, S.; MAKRIDAKIS, S.; HYNDMAN, R. J. *Forecasting: methods and applications*. [S.l.]: John Wiley & Sons, 1998.
- ZEILER, M. D.; RANZATO, M.; MONGA, R.; MAO, M.; YANG, K.; LE, Q. V.; NGUYEN, P.; SENIOR, A.; VANHOUCHE, V.; DEAN, J. et al. On rectified linear units for speech processing. In: IEEE. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.], 2013. p. 3517–3521.