

**Carolina Martins Crispim**

**Comparação entre os Modelos SARIMA e  
os Métodos de Machine Learning para  
previsão de demanda**

Niterói - RJ, Brasil

25 de julho de 2022

**Carolina Martins Crispim**

**Comparação entre os Modelos  
SARIMA e os Métodos de Machine  
Learning para previsão de demanda**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dra. Jessica Quintanilha Kubrusly

Co-Orientador(a): Prof. Dra. Patrícia Lusié Velozo da Costa

Niterói - RJ, Brasil

25 de julho de 2022

**Carolina Martins Crispim**

**Comparação entre os Modelos SARIMA e os  
Métodos de Machine Learning para previsão  
de demanda**

Monografia de Projeto Final de Graduação sob o título  
*“Comparação entre os Modelos SARIMA e os Métodos de  
Machine Learning para previsão de demanda”*, defendida por  
Carolina Martins Crispim e aprovada em 25 de julho de 2022,  
na cidade de Niterói, no Estado do Rio de Janeiro, pela banca  
examinadora constituída pelos professores:

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Jessica Quintanilha Kubrusly**  
Departamento de Estatística – UFF

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia Lusié Velozo da Costa**  
Departamento de Estatística – UFF

---

**Prof. Dr. Jony Arrais Pinto Junior**  
Departamento de Estatística – UFF

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Mariana Albi de Oliveira Souza**  
Departamento de Estatística – UFF

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

C932c Crispim, Carolina Martins  
Comparação entre os Modelos SARIMA e os Métodos de  
Machine Learning para previsão de demanda / Carolina Martins  
Crispim ; Jessica Quintanilha Kubrusly, orientadora ;  
Patrícia Lusié Velozo da Costa, coorientadora. Niterói,  
2022.  
34 f. : il.

Trabalho de Conclusão de Curso (Graduação em  
Estatística)-Universidade Federal Fluminense, Instituto de  
Matemática e Estatística, Niterói, 2022.

1. Previsão de demanda. 2. Aprendizado de máquina. 3.  
Série temporal. 4. Produção intelectual. I. Kubrusly,  
Jessica Quintanilha, orientadora. II. Costa, Patrícia Lusié  
Velozo da, coorientadora. III. Universidade Federal  
Fluminense. Instituto de Matemática e Estatística. IV.  
Título.

CDD -

# Resumo

O Aprendizado de Máquinas é um tema em alta na área da estatística e outras áreas correlacionadas, está presente em quase todos os fóruns e congressos despertando a vontade de utilização dessa ferramenta nas pessoas e empresas. Uma forma de aplicar Aprendizado de Máquinas é realizando previsão de séries temporais, com expectativa de se obter bons resultados. Sendo assim, este trabalho se propõe a analisar a performance de dois métodos de Aprendizado de Máquinas, Floresta Aleatória e LSTM, em relação aos modelos de Séries Temporais SARIMA, aplicado à 84 séries temporais de produtos reais, para realizar previsão de demanda. Considerando como medida de erro o MAPE, ambos métodos de Aprendizado de Máquinas tiveram uma performance melhor que os modelos SARIMA, porém o método que obteve menor erro para a maioria das séries foi a Floresta Aleatória. De fato, os métodos de Aprendizado de Máquinas podem obter uma melhor acurácia na previsão de demanda, gerando valor para as empresas, porém, deve-se ressaltar que o modelo SARIMA, por ser probabilístico, possui capacidade de gerar mais informações sobre as séries temporais.

Palavras-chave: Previsão. Séries Temporais. Florestas Aleatórias. Redes Neurais. LSTM.

# Sumário

**Lista de Figuras**

**Lista de Tabelas**

<b>1</b>	<b>Introdução</b>	p. 9
	Revisão Bibliográfica . . . . .	p. 9
	Objetivo e Organização . . . . .	p. 10
<b>2</b>	<b>Materiais e Métodos</b>	p. 12
2.1	Análise de Séries Temporais . . . . .	p. 12
	Conceitos de Séries Temporais . . . . .	p. 12
	Média Móvel Integrado Autoregressivo Sazonal . . . . .	p. 14
2.2	Métodos de Aprendizado de Máquinas . . . . .	p. 16
	Árvore de Decisão . . . . .	p. 16
	Redes Neurais . . . . .	p. 19
2.3	Método de Comparação . . . . .	p. 21
2.4	Materiais . . . . .	p. 22
<b>3</b>	<b>Análise dos Resultados</b>	p. 23
3.1	Análise detalhada de apenas uma série . . . . .	p. 23
3.2	Análise de todas as séries . . . . .	p. 28
<b>4</b>	<b>Conclusões</b>	p. 30
	<b>Referências</b>	p. 32

**Apêndice 1 - Descrição das Séries Históricas**

p. 33

**Apêndice 2 - MAPE dos Modelos**

p. 34

# Lista de Figuras

1	Figura de um exemplo de árvore de decisão retirada do livro de James et al. (2017) . . . . .	p. 18
2	Rede Neural Artificial de um neurônio . . . . .	p. 20
3	Rede Neural Artificial . . . . .	p. 21
4	Histórico de Vendas do Produtos A14 no período de janeiro de 2016 até junho de 2021 . . . . .	p. 24
5	Função de autocorrelação (FAC) e função de autocorrelação parcial (FACP) da série A14. . . . .	p. 25
6	Série temporal diferenciada. . . . .	p. 26
7	Função de autocorrelação (FAC) e função de autocorrelação parcial (FACP) da série diferenciada. . . . .	p. 26
8	Previsões da Série A14 . . . . .	p. 27

# Lista de Tabelas

1	Comportamento da FAC e FACP para alguns processos . . . . .	p. 15
2	MAPE dos Modelos para a série A14 . . . . .	p. 28
3	Porcentagem de vezes que o modelo venceu . . . . .	p. 28
4	Comportamento do MAPE entre os Modelos e do Melhor Modelo . . .	p. 29
5	Porcentagem de vezes que o modelo venceu: SARIMA x FA . . . . .	p. 29
6	Porcentagem de vezes que o modelo venceu: SARIMA x LSTM . . . . .	p. 29
7	Tabela com a descrição das séries . . . . .	p. 33
8	Tabela com o MAPE dos modelos para cada série . . . . .	p. 34

# 1 Introdução

Toda empresa precisa se planejar diante da demanda de seus bens e serviços. A movimentação de recursos para atender uma certa demanda pode gerar custos, os quais deseja-se minimizar. Quando se aloca mais recursos que o necessário, tem-se o custo do excesso, e quando se aloca menos recursos que o necessário, tem-se o custo da falta. Uma das etapas do planejamento de demanda que contribui diretamente para a diminuição desses custos, é a etapa de previsão de vendas.

Se a empresa soubesse com exatidão o quanto venderia nos próximos meses, os custos de excesso e falta seriam nulos, uma vez que se alocaria recursos suficientes para suprir a demanda. Sabe-se que o cenário descrito anteriormente é impossível, porém, usando as informações das vendas passadas alinhadas com as informações de mercado, é possível obter uma boa estimativa da demanda futura. Sendo assim, tenta-se identificar padrões nas séries históricas de vendas passadas, a fim de prever a demanda futura.

Para se obter uma previsão a partir da série histórica, são comumente usados os modelos e métodos matemáticos. Os modelos são capazes de captar comportamentos das séries históricas como sazonalidade, tendência e nível, e baseados nos padrões encontrados, realiza-se a previsão. Apesar destes modelos atingirem o objetivo de prever a demanda, existe um debate sobre como os métodos de Aprendizado de Máquinas podem ser úteis para realizar previsão e estão sendo sugeridos como uma alternativa aos modelos tradicionais.

## Revisão Bibliográfica

Apesar dos métodos de Aprendizado de Máquinas (AM) serem comumente usados para problemas de classificação e regressão, alguns métodos também são aplicáveis à séries temporais. Na literatura, existem vários artigos que apresentam a utilização de AM para previsão de série temporal. Nesta seção, serão apresentados dois artigos que utilizaram métodos de AM para realizar previsão de série temporal, e além disso,

compararam a acurácia com o modelo de séries temporais chamado Média Móvel Integrada Autorregressiva (ARIMA).

Em Siami-Namini, Tavakoli e Namin (2018), compara-se os modelos ARIMA com o método de *deep learning* desenvolvido para realizar previsão de série temporal, chamado de *Long Short-Term Memory* (LSTM). O estudo foi aplicado à 12 séries temporais mensais econômicas e financeiras, todas com pelo menos 300 observações, entretanto, para treinar os métodos foram usados 70% dos dados. A métrica utilizada para medir a acurácia dos métodos foi a Raiz do Erro Quadrático Médio (*Root Mean Squared Error* - RMSE). Essa métrica compara o erro entre a previsão e o dado real para cada valor do conjunto de teste. A conclusão foi que o LSTM teve uma previsão mais acurada que o modelo ARIMA, ou seja, o LSTM obteve entre 84% e 87% de erro a menos em relação ao ARIMA.

Em Kane et al. (2014), compara-se o desempenho do modelo ARIMA com o método de regressão de AM chamado de Floresta Aleatória para previsão de gripe aviária no Egito. A série histórica utilizada é de surto de H5N1, do período de 2005 até 2007. Para o método de regressão Floresta Aleatória, além dos valores de surtos passados, foram consideradas como variáveis regressoras a temperatura e a umidade. A métrica utilizada para comparar a acurácia das previsões foi o Erro Quadrado Médio (*Mean Squared Error* - MSE). O método de Floresta Aleatória obteve menores valores de MSE em relação ao modelo ARIMA, concluindo que para previsão de gripe aviária, o método Floresta Aleatória performa melhor que o modelo tradicional ARIMA.

## Objetivo e Organização

A busca por modelos e métodos que geram uma boa previsão, ou seja, aquele que obtém uma melhor acurácia, é amplamente estudada na literatura, de tal forma que são aplicadas a diferentes cenários. No presente estudo, serão comparados os resultados das previsões geradas pelos modelos tradicionais SARIMA, e pelos métodos de Aprendizado de Máquinas LSTM e Floresta Aleatória, no cenário de previsão de vendas. Os modelos serão aplicados em cerca de 80 séries históricas mensais da venda de produtos do setor alimentício, como esses dados são de uma empresa privada, é necessário manter sigilo sobre as informações dos produtos.

Para cada série será eleito o modelo vencedor que apresentar o menor erro no período de teste, e ao final, será apresentado quais das abordagens obteve uma maior taxa de sucesso. A métrica de erro utilizada será o MAPE (Erro Médio Absoluto Percentual), como essa métrica é percentual, não é levada em consideração a escala da série, portanto,

é possível comparar a acurácia de um determinado modelo para um série que vende em torno de 100 produtos por mês, com uma série que vende em torno de 10000.

No Capítulo 2, de Materiais e Métodos, serão apresentados os conceitos dos modelos e métodos usados neste trabalho, bem como o material de estudo que são as séries temporais. Já no Capítulo 3, de Análise dos Resultados, será apresentada uma breve descrição do conjunto de dados utilizado, e logo depois, todas as análises e resultados obtidos para esses dados. Por fim, no Capítulo 4 de Conclusão, serão reforçados todos os resultados relevantes necessários para atingir o objetivo central deste trabalho.

## 2 Materiais e Métodos

Neste capítulo serão apresentados os conceitos de Análise de Séries Temporais com foco na modelagem através do modelo SARIMA (Média Móvel Integrado Autoregressivo Sazonal) e os dois métodos de Aprendizado de Máquinas que serão apresentados como alternativa a modelagem tradicional. Logo após, terá uma seção sobre a métrica que será aplicada para comparação dos resultados de acurácia de cada modelo aplicado, e por fim, será apresentado o material de estudo utilizado neste trabalho.

### 2.1 Análise de Séries Temporais

Uma das principais abordagens das Análises de Séries Temporais é tentar entender como o comportamento passado pode afetar o futuro, e esse será o foco do estudo realizado neste trabalho, ou seja, realizar previsões. Nesta seção, serão apresentadas as definições formais de Séries Temporais, e logo em seguida, o modelo SARIMA, comumente utilizado para realizar previsão.

#### Conceitos de Séries Temporais

Uma Série Temporal é um conjunto de observações ordenadas no tempo. A principal característica de uma série é o fato das observações em instantes próximos serem correlacionadas, e o estudo que se propõe a analisar essas correlações é chamado de Análise de Séries Temporais. Existem alguns outros aspectos importantes de uma série histórica que são conhecidos como nível, tendência e sazonalidade.

O nível é o valor da série no instante inicial. Já a tendência é o comportamento que a série está tendendo a ter com o passar o tempo, podendo estar crescendo, decrescendo ou sem tendência (apenas com variações em torno do nível). A sazonalidade é um fenômeno que ocorre com frequência em intervalos de tempo bem demarcados, a maioria das vezes

são comportamentos previamente esperados, como por exemplo acontecimentos marcados pelas estações do ano.

Uma definição mais formal e estatística de uma série temporal pode ser dada considerando o conceito de processos estocásticos. As definições a seguir foram retiradas do livro do Morettin e Toloi (2004).

**Definição 2.1** *Seja  $T$  um conjunto arbitrário. Um processo estocástico é uma família  $Y = \{Y_t; t \in T\}$ , tal que, para cada  $t \in T$ ,  $Y_t$  é uma variável aleatória.*

**Definição 2.2** *Série temporal ou trajetória do processo é um conjunto de observações  $y = \{y_t; t \in T\}$ , sendo  $T$  uma coleção de instantes sequenciais e ordenados no tempo. Mais ainda, é a realização particular de um processo estocástico.*

Um conceito essencial para identificação do processo estocástico da série temporal é a autocorrelação entre os pontos da série, para isso, é importante primeiro entender a função de autocovariância. A função de autocovariância mede a dependência linear entre a variável aleatória de interesse em dois instantes e sua definição formal adaptada do livro do Shumway e Stoffer (2017) é apresentada a seguir.

**Definição 2.3** *A função de autocovariância é definida como o produto do segundo momento*

$$\gamma(t, t+h) = COV(Y_t, Y_{t+h}) = E[(Y_t - \mu_t)(Y_{t+h} - \mu_{t+h})] \quad (2.1)$$

para todo  $h = 0, \pm 1, \pm 2, \dots$

A função de autocorrelação (FAC) denotada por  $\rho$  mede a correlação da variável de interesse em 2 instantes de tempo.

**Definição 2.4** *A função de autocorrelação (FAC) é definida como*

$$\rho(t, t+h) = \frac{\gamma(t, t+h)}{\sqrt{\gamma(t, t)\gamma(t+h, t+h)}} \quad (2.2)$$

para todo  $h = 0, \pm 1, \pm 2, \dots$ . Note que  $\rho(t, t+h) = 0$  quando as variáveis nos instantes  $t$  e  $t+h$  não forem mais correlacionadas. Além disso, tem-se que  $-1 \leq \rho(t, t+h) \leq 1$  e que  $\rho(t, t+h) = \rho(t+h, t)$ .

E por fim, a função de autocorrelação parcial (FACP) mede a correlação condicional entre  $Y_t$  e  $Y_{t+h}$  supondo conhecidas as variáveis de menor defasagem

$(Y_{t+1}, Y_{t+2}, \dots, Y_{t+h-1})$ . Ambas funções indicam o quanto uma observação sofre influência de outra e são informações necessárias no processo de modelagem de uma série temporal.

Outro conceito muito importante que é uma premissa do modelo que será visto a seguir, é o conceito de estacionariedade. Uma série pode ser fortemente ou fracamente estacionária, para este trabalho, uma série fracamente estacionária é condição necessária e suficiente. Uma série  $Y_t$  é dita estacionária quando seus pontos flutuam aleatoriamente em torno da sua média constante  $\mu$ . Portanto, uma série com tendência, ou seja, que cresce ou decresce ao longo do tempo é dita não estacionária. A definição de estacionariedade fraca adaptada do livro do Shumway e Stoffer (2017) é apresentada a seguir.

**Definição 2.5** *Uma série temporal fracamente estacionária,  $Y = \{Y_t; t \in T\}$ , é um processo de variância finita tal que*

- a função da média,  $\mu_t$  é constante e não depende do tempo  $t$ , e
- a função de autocovariância,  $\gamma(t, t+h)$ , depende apenas de  $t$  e  $t+h$  através da diferença  $|t+h-t|$

Para algumas abordagens em séries temporais é necessário que a série seja estacionária, portanto é aplicada uma transformação nos dados para torna-la estacionária. A transformação usual é a diferenciação denotada por  $\Delta Y_t = Y_t - Y_{t-1}$ , as vezes é necessário realizar mais de uma diferenciação até que a série se torne estacionária.

## Média Móvel Integrado Autoregressivo Sazonal

O modelo de Média Móvel Integrado Autoregressivo Sazonal (SARIMA) é baseado na ideia que todo ponto da série pode ser explicado por uma função dos  $p$  valores passados (parte autorregressiva do modelo). Além disso, o ponto também pode ser explicado por uma combinação linear de  $q$  variáveis aleatórias  $\epsilon_t$  onde  $\epsilon_t \sim N(0, \sigma_\epsilon^2)$  (parte Média Móvel do modelo). Este modelo pressupõe que a série seja estacionária, sendo denotado por  $d$  a quantidade de diferenciações necessárias para que a série se torne estacionária.

Ademais, o SARIMA também possui uma parte sazonal denotada por  $s$ , onde  $s$  pode ser, por exemplo, 12 representando uma sazonalidade mensal. Essa parte sazonal do modelo também pode ser explicada por um polinômio com uma parte autorregressiva de ordem  $P$  e uma parte Média Móvel de ordem  $Q$ . É possível que a sazonalidade da série

não seja estacionária. Portanto também é necessário diferenciar a parte sazonal do modelo  $D$  vezes até que se torne estacionária.

Os modelos são denotados da seguinte forma:  $SARIMA(p, d, q) \times (P, D, Q)_s$  cuja equação é dada por:

$$\begin{aligned} \Delta^D \Delta^d Y_t = & \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} + \\ & \Phi_1 Y_{t-1} + \dots + \Phi_P Y_{t-P} - \Theta_1 \epsilon_{t-1} - \dots - \Theta_Q \epsilon_{t-Q} \end{aligned} \quad (2.3)$$

em que  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_P$  e  $\Theta_1, \dots, \Theta_Q$  são os parâmetros a serem estimados do modelo, e  $\Delta^D$  e  $\Delta^d$  denotam as diferenciações necessárias para tornar a série estacionária.

Alguns casos particulares do SARIMA ocorrem quando a série não necessita de todas as partes desse modelo como, por exemplo, quando uma série não é sazonal podemos usar apenas a parte não sazonal do modelo denotado apenas por  $ARIMA(p, d, q)$ , e nesse caso,  $D = 0$  e os parâmetros  $\Phi_1, \dots, \Phi_P$  e  $\Theta_1, \dots, \Theta_Q$  são iguais a 0. Quando não há necessidade de diferenciar a série, ou seja, a série é estacionária e o modelo não tem sazonalidade, tem-se um  $ARMA(p, q)$  com  $d = 0$ ,  $D = 0$  e parâmetros  $\Phi_1, \dots, \Phi_P$  e  $\Theta_1, \dots, \Theta_Q$  iguais a 0. Quando o modelo não tem parte Média Móvel ou não tem parte autorregressiva tem-se um  $AR(p)$  existindo apenas os parâmetros  $\phi_1, \dots, \phi_p$  e um  $MA(q)$  com parâmetros  $\theta_1, \dots, \theta_q$ , respectivamente.

As ordens do modelo  $p, q, P$  e  $Q$  podem ser identificadas analisando a FAC e FACP. Plota-se um gráfico das funções de autocorrelação que é chamado de correlograma, os picos de correlação significativos indicam de quais ordens aquele processo se originou. O comportamento usual da FAC e FACP para cada um dos processos está listado na Tabela 1. Além das ordens, um correlograma com decaimento exponencial lento, ou seja, muitos picos altos para as defasagens, indica que a série não é estacionária. E ainda, picos em torno de defasagens sazonais como por exemplo, 12, 24, 36, ... indicam que a série é sazonal.

Tabela 1: Comportamento da FAC e FACP para alguns processos

Processo	FAC	FACP
AR(p)	Decaimento	Truncada no lag q
MA(q)	Truncada na defasagem p	Decaimento
ARMA(p,q)	Decaimento a partir da defasagem p	Decaimento a partir da defasagem q
$AR(P)_s$	Decaimento	Truncada na defasagem $Q_s$
$MA(Q)_s$	Truncada na defasagem $P_s$	Decaimento
$SARMA(P, Q)_s$	Decaimento a partir da defasagem $P_s$	Decaimento a partir da defasagem $Q_s$

A estimação dos parâmetros deste modelo pode ser feita por máxima verossimilhança, e este é o método utilizado neste trabalho. A função  $L$  de máxima verossimilhança, é a função dos parâmetros a serem estimados dada as observações, logo, a sua forma depende do processo a ser estimado. O livro do Shumway e Stoffer (2017) possui toda explicação da estimação dos parâmetros através da máxima verossimilhança. Além disso, há uma série de restrições para os parâmetros  $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_P$  e  $\Theta_1, \dots, \Theta_Q)$  deste modelo, além das propriedades de causalidade e invertibilidade que também podem ser vistas com mais detalhes no livro do Shumway e Stoffer (2017). Estes modelos são bastante utilizados para previsão em diversos ramos e são considerados robustos e eficientes para realizar previsão.

## 2.2 Métodos de Aprendizado de Máquinas

O Aprendizado de Máquinas é um método que permite a criação de modelos que aprendem automaticamente com padrões e comportamento dos dados. Nesta seção serão apresentados os dois métodos de Aprendizado de Máquinas que serão avaliados nesse estudo, como métodos de previsão alternativos aos modelos tradicionais para séries temporais vistos na seção anterior.

### Árvore de Decisão

Antes de apresentar o método da Floresta Aleatória (*Random Forest*) é preciso entender um algoritmo mais simples conhecido como Árvore de Decisão, que retorna uma previsão para a variável desejada baseada em outras variáveis. Esse algoritmo é utilizado tanto para problemas de classificação, quando a variável resposta é qualitativa, quanto para regressão, quando a variável resposta é quantitativa. Nesse estudo, por se tratar da previsão de pontos para uma série histórica, será revisado apenas o método para regressão.

Considerando um problema de Regressão em que  $Y$  é a variável resposta e  $X_j$  as covariáveis,  $j = 1, 2, \dots, p$ , a ideia principal da Árvore de Decisão, aplicada à regressão, é separar o plano dos dados em diversas regiões e a previsão é dada pela média dos pontos daquela região em que o ponto que se deseja prever pertence. Cada separação que se faz é representada por uma folha da árvore (ou nó), sendo a primeira separação chamada de raiz, e esse conjunto de regras de divisões é a chamada Árvore de Decisão. A seguir é apresentado um pseudocódigo de como esse algoritmo funciona. Esse passo a passo foi

retirado do trabalho de conclusão de curso de Porto (2021) e adaptado para representar o algoritmo da construção de uma Árvore de Decisão para Regressão.

Seja  $\mathcal{X} \subset \mathbb{R}^p$  o conjunto de todas as  $n$  observações de todas as  $p$  covariáveis. Ou seja, se  $\mathbf{x}_i \in \mathcal{X}$  então  $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})$ ,  $i = 1, \dots, n$ , onde  $x_{i,j}$  é a  $i$ -ésima observação da covariável  $X_j$

1. Definimos um critério de parada  $C$  que, em geral, é o número máximo de nós ou número mínimo de observações em cada nó.
2. Considerando todas as covariáveis  $X_j$  e todos os pontos de corte  $s$  que separam o espaço das observações nas regiões  $R_{1j,s} = \{\mathbf{x}_i \in \mathcal{X} \mid x_{i,j} < s\}$  e  $R_{2j,s} = \{\mathbf{x}_i \in \mathcal{X} \mid x_{i,j} \geq s\}$ , buscamos então os valores de  $j$  e  $s$  que minimizam:

$$RSS_{j,s} = \sum_{i|\mathbf{x}_i \in R_{1j,s}} (y_i - \hat{y}_1)^2 + \sum_{i|\mathbf{x}_i \in R_{2j,s}} (y_i - \hat{y}_2)^2, \quad (2.4)$$

sendo  $\hat{y}_k = \sum_{i|\mathbf{x}_i \in R_{k,j}} \frac{y_i}{n_k}$  e  $n_k$  o número de elementos em  $R_{k,j}$  e  $y_i$  o valor da  $i$ -ésima observação para a variável resposta.

3. Sejam  $j^*$  e  $s^*$  os valores que minimizam  $RSS_{j,s}$ , defina  $\mathcal{X}_1 = RSS_{1j^*,s^*}$  e  $\mathcal{X}_2 = RSS_{2j^*,s^*}$ . Veja que  $\mathcal{X}_1$  e  $\mathcal{X}_2$  formam uma partição de  $\mathcal{X}$ :  $\mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}$  e  $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$ .
4. Verifique se o critério de parada  $C$  foi atingido para o conjunto de pontos  $\mathcal{X}_1$ , caso positivo o algoritmo terminada, caso contrário, faça  $\mathcal{X} = \mathcal{X}_1$  e retorne para o passo 2.
5. Verifique se o critério de parada  $C$  foi atingido para o conjunto de pontos  $\mathcal{X}_2$ , caso positivo o algoritmo terminada, caso contrário, faça  $\mathcal{X} = \mathcal{X}_2$  e retorne para o passo 2.
6. Considere as regiões  $R_1, R_2, \dots, R_f$  definidas no algoritmo e uma nova observação das variáveis preditivas:  $\mathbf{x}_{novo} = (x_1, x_2, \dots, x_p)$ . Suponha  $\mathbf{x}_{novo} \in \tilde{R}$ , sendo  $\tilde{R}$  uma das regiões definidas acima. Atribuiremos como previsão para  $y_{novo}$  a seguinte expressão:

$$\hat{y}_{novo} = \sum_{i|\mathbf{x}_i \in \tilde{R}} \frac{y_i}{\tilde{n}}, \quad (2.5)$$

sendo  $\tilde{n}$  o número de elementos em  $\tilde{R}$ .

A Figura 1, retirada de um exemplo de James et al. (2017), ilustra o processo da árvore e a separação no plano cartesiano.

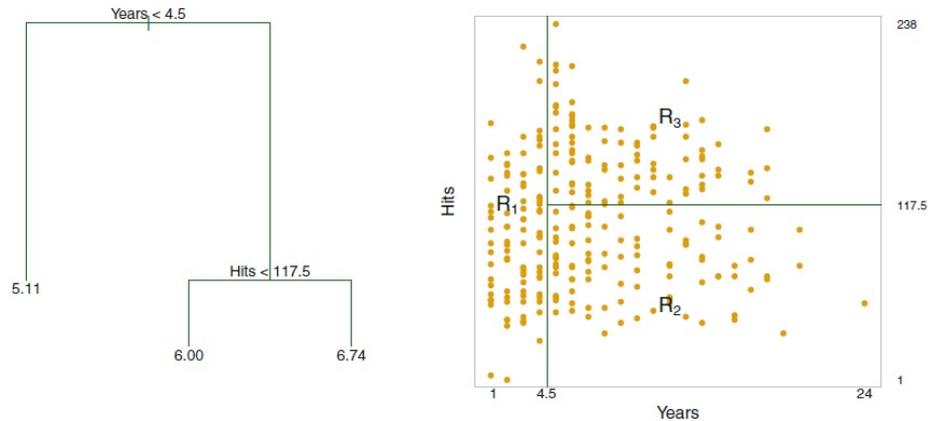


Figura 1: Figura de um exemplo de árvore de decisão retirada do livro de James et al. (2017)

Note que uma árvore é construída de acordo com o banco de dados que ela recebe, ou seja, a estrutura da árvore tem relação direta com os dados de treinamento. Tendo em vista que dados diferentes geram árvores diferentes, o algoritmo de Floresta Aleatória tenta contornar essa questão. Portanto, esse método utiliza várias árvores de decisão e consegue obter um aumento significativo na acurácia em relação ao método que utiliza apenas uma árvore.

A principal ideia do método é criar muitas árvores usando diferentes amostras e diferentes subconjuntos das covariáveis. Usa-se a técnica de reamostragem conhecida como *bootstrap* (James et al. (2017)) para sortear a amostra. As covariáveis são sorteadas de forma aleatória e sem reposição. É comum sortear  $v = \lfloor \sqrt{p} \rfloor$  (parte inteira da raiz de  $p$  variáveis). Segue o algoritmo da construção de uma Floresta Aleatória:

1. Utiliza-se *bootstrap* (técnica de reamostragem com reposição), para selecionar uma amostra do conjunto de dados do mesmo tamanho da amostra original.
2. Selecionam-se, de forma aleatória e sem reposição,  $v$  covariáveis.
3. Criamos a árvore de decisão usando a amostra e as  $v$  covariáveis selecionadas nos passos 1 e 2.
4. Repetimos os passos 1,2 e 3  $Q$  vezes, criando  $Q$  árvores.

Dado uma nova amostra o valor previsto para  $y_k$  é a média das  $Q$  previsões para essa observação considerando as  $Q$  árvores da floresta.

Para aplicar um algoritmo de Floresta Aleatória em uma série temporal, é necessário organiza-la de forma estrutural para uma regressão. Desta forma, as covariáveis são definidas como sendo as observações passadas ao ponto  $y_t$ , ou seja, para uma determinada observação  $y_t$ , as covariáveis serão  $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ . Suponha uma série com 6 observações  $y_1, y_2, \dots, y_6$  e  $p = 3$ , tem-se que  $y_6$  será explicado por  $(y_5, y_4, y_3)$ ,  $y_5$  por  $(y_4, y_3, y_2)$ , e  $y_4$  por  $(y_3, y_2, y_1)$ .

A quantidade de covariáveis (neste caso, pontos passados)  $p$  será definida usando o método de Validação Cruzada  $k$ -fold. Esse método consiste em separar a amostra de treino em  $k$  subconjuntos, em que  $k - 1$  são usados para treino e o outro é usado para teste. O valor  $p$  que obtiver um menor erro em todos os  $k$  subconjuntos será definido como o número de covariáveis a serem utilizadas para treinar o modelo. Para essa validação, o erro considerado é o Erro Quadrático Médio (*Mean Square Error* - MSE) dado por:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\text{Previsão}_i - \text{Real}_i)^2, \quad (2.6)$$

onde  $n$  é o número de observações em cada partição da amostra de treino.

## Redes Neurais

A Rede Neural é uma técnica de processamento de informação inspirada no sistema nervoso biológico. Esse sistema é composto pelos neurônios que são conectados entre si através de sinapses, e estas conexões se alteram de acordo com estímulos externos. As informações entre neurônios são transmitidas por meio de sinais elétricos ou químicos, ou seja, um neurônio recebe sinais de outros, processa e propaga um sinal para outros neurônios formando assim, uma rede interconectada. É importante ressaltar que o sistema é capaz de processar informações paralelamente.

O algoritmo de Rede Neural simula o comportamento do sistema neural biológico, sendo assim, é composto por unidades (nós) conectadas por pesos, que representam os neurônios e as sinapses, respectivamente. Cada nó recebe de entrada informações dimensionadas pelos pesos (conexões que ligam os nós) e dentro do nó realiza-se uma soma e é aplicada a chamada função de ativação retornando como saída um valor. A Figura 2 representa uma rede neural com apenas um nó, onde  $x_1, \dots, x_n$  são os valores das variáveis de entrada,  $w_1, \dots, w_n$  são os pesos. Se deseja-se estimar um valor para

a variável  $Y$ , então o resultado dessa Rede Neural com apenas um nó, apresentada na Figura 2, é dado por:

$$\hat{y} = f \left( \sum_{i=1}^n w_i x_i \right), \quad (2.7)$$

sendo  $f$  uma função chamada de função de ativação, que depende da natureza do problema.

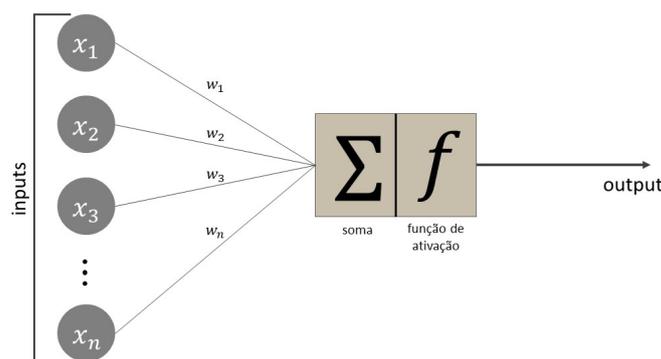


Figura 2: Rede Neural Artificial de um neurônio

Os pesos de cada conexão são os parâmetros a serem estimados, e a função de ativação é definida pelo usuário de acordo com as necessidades do problema. O conjunto de nós é chamado de camada, e como mostra a Figura 3 de uma Rede Neural mais complexa, existem as camadas de entrada, camadas ocultas e camadas de saída. As funções de ativação comumente usadas são: Sigmoides, Linear, ReLu e Softmax. Mais detalhes sobre essas e outras funções podem ser encontrados no livro do Charu (2018) bem como as particularidades da Rede Neural complexa.

As redes neurais podem ser aplicadas em vários problemas diferentes, e cada uma tem suas particularidades. Neste estudo será aplicado a Rede Neural LSTM (*Long Short-Term Memory*) que é um tipo de Rede Neural Recorrente (RNN), mais detalhes sobre as Redes Neurais Recorrentes podem ser encontrados em Charu (2018) no Capítulo 7.

A rede LSTM é formada por um conjunto de sub-redes conectadas recorrentemente permitindo que as informações persistam, ou seja, a informação da construção dos pontos anteriores são levadas em consideração para a construção dos próximos. Desta forma, se fazem ideais para processar dados em sequência, pois possuem uma arquitetura que aprende dependências ao longo do tempo. Além do livro Charu (2018) a tese do Kawakami (2008) se aprofunda nos detalhes da rede LSTM.

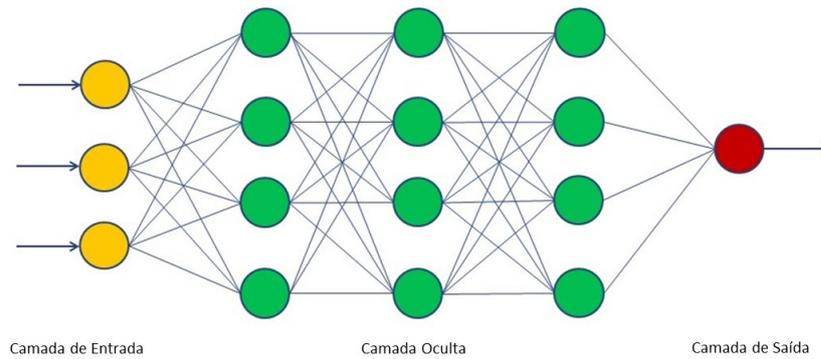


Figura 3: Rede Neural Artificial

## 2.3 Método de Comparação

Todos os modelos e métodos vistos nesse estudo serão aplicados em cada uma das 84 séries históricas. Para identificar o melhor modelo de uma série, serão deixados de fora da modelagem os 12 meses mais recentes desta série. Essas observações serão usadas para calcular o erro da previsão dos modelos, e esse conjunto de observações será chamado de Período de Teste. A medida de erro que será usada para analisar a performance dos modelos, é o Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error (MAPE)*) que é dado pela seguinte expressão:

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{\text{Real}_i - \text{Previsão}_i}{\text{Real}_i} \right| \times 100\%, \quad (2.8)$$

onde  $m = 12$  é o número de observações no período de teste para cada uma das séries.

Calculado o MAPE no período de teste para cada modelo, será eleito como o melhor modelo para a uma determinada série histórica aquele que obteve menor MAPE. Depois de modelar todas as séries e escolher o melhor modelo para cada uma delas, os resultados serão analisados com uma ideia de competição. Portanto, para as 84 séries modeladas, será identificada a porcentagem de vezes que cada um dos três modelos (SARIMA, Floresta Aleatória e LSTM) obteve uma melhor Acurácia.

## 2.4 Materiais

Este estudo é aplicado ao planejamento de demanda, portanto, antes de descrever o banco de dados, serão apresentados dois conceitos importantes. O primeiro é o conceito de demanda *sell in*, que significa que as vendas são feitas da indústria para os pontos de venda. Existe também a demanda *sell out*, que é todo tipo de venda feita diretamente para o consumidor final, porém, nesse estudo, só serão analisadas séries de vendas *sell in*.

O outro conceito importante é o de agrupar informações de vendas dos produtos por pertencerem a uma mesma família, categoria, ou por semelhança no consumo. É comum que as empresas agrupem a venda de um conjunto de produtos que fazem sentido serem analisados e planejados juntos. Por exemplo, se uma empresa fabrica garrafas de 500ml nas cores preta e branco, pode-se agrupar as vendas de ambas apenas como venda de garrafas.

Sendo assim, serão usadas 84 séries históricas reais de uma empresa do setor alimentício. Cada série apresenta a demanda *sell in* de um único produto ou um conjunto de produtos diferentes pertencentes a uma mesma família. As séries  $Y_T$  são mensais podendo ter tamanho  $T$  diferentes, mas maioria das séries possuem pelo menos 5 anos de histórico. A tabela com a descrição de cada série temporal pode ser encontrada no Apêndice 1.

## 3 Análise dos Resultados

Neste capítulo serão apresentados todos os resultados obtidos neste estudo. Como os dados usados são reais de uma empresa de nível nacional, algumas informações sobre os produtos foram ocultadas. Todas as análises presentes neste trabalho foram feitas no Programa R (R Core Team, 2014), entretanto, para aplicar o modelo LSTM foi usado um pacote que utiliza a linguagem Python, sendo necessário ter o Python instalado na máquina.

A fim de entender o processo de previsão que será realizado em cada umas das séries históricas, na Seção 3.1, serão apresentados os passos e detalhamento das análises para a série A14 (vide Apêndice 1). Além disso, apenas para essa série, foi realizado uma modelagem manual SARIMA utilizando as informações de FAC e FACP com o intuito de confrontar o resultado da função `auto.arima` do R que modela automaticamente um modelo SARIMA. Já na Seção 3.2, serão apresentados todos os resultados obtidos para as outras 83 séries históricas de demanda.

### 3.1 Análise detalhada de apenas uma série

O histórico de demanda dos produtos A14 é apresentado a seguir na Figura 4. A série começa em janeiro de 2016 e termina em junho de 2021. Sendo assim, os meses de julho de 2020 até junho de 2021 foram definidos como dados de teste, ficando fora da modelagem. Nota-se que a série possui picos no final de cada ano, e baixa de venda no início de cada ano. Isso ocorre, muito provavelmente, por conta do ponto de venda acreditar que venderá muito ao final do ano e que não pode faltar o produto em seu estabelecimento. Como o produto não vende como o esperado, os pontos de vendas acabam acumulando estoque, o que explica a baixa de demanda no início de cada ano. Portanto, essa série possui uma sazonalidade bem demarcada ao final e início de cada ano.

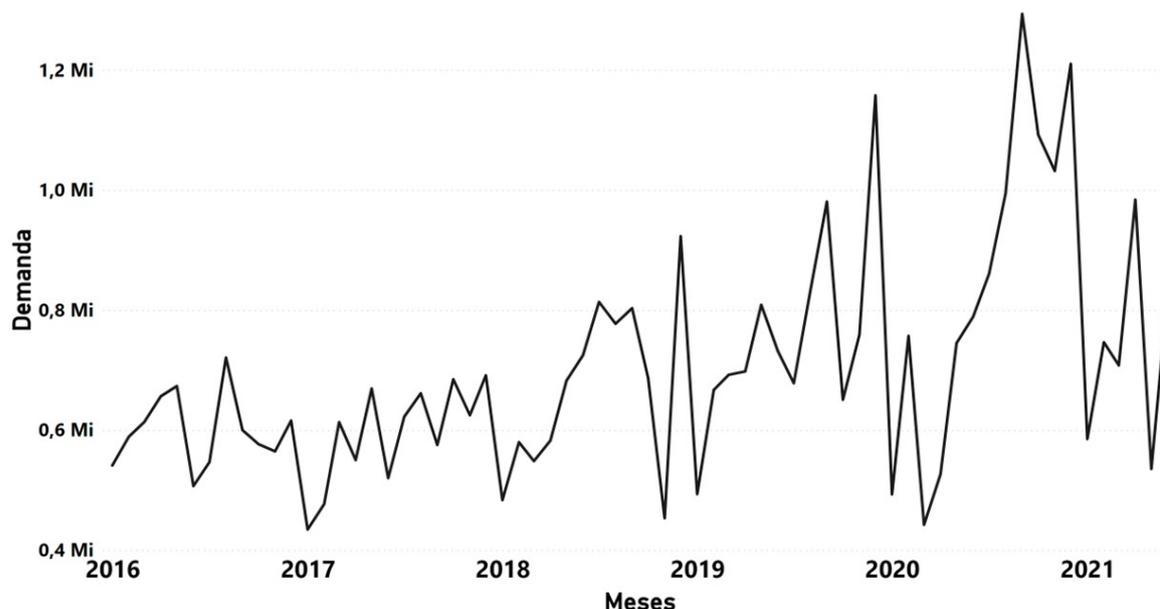


Figura 4: Histórico de Vendas do Produtos A14 no período de janeiro de 2016 até junho de 2021

A primeira modelagem foi feita utilizando a função `auto.arima` do pacote `forecast` que ajusta diferentes modelos SARIMA para uma dada série temporal. A função cria uma grade de valores para os parâmetros  $p, d, q, P, D$  e  $Q$ , considerando  $p \leq 5, q \leq 5, P \leq 2, Q \leq 2$  e  $p + q + P + Q \leq 5$ , e estima os demais parâmetros desconhecidos em cada caso. Ela seleciona o melhor modelo com base no AIC (Critério de Informação de Akaike) que é uma métrica que mede a qualidade e simplicidade do modelo, sendo que menores valores de AIC indicam um melhor modelo. Usou-se essa função na série temporal A14 e o melhor modelo, segundo essa função, foi um  $SARIMA(0, 1, 1) \times (1, 0, 0)_{12}$  com um AIC de 1395,49. A função `auto.arima` não estima os parâmetros  $p, d, q, P, D$  e  $Q$ . Sendo assim, pode ser que não seja ajustado o modelo para a melhor combinação destes parâmetros. Para investigar melhor isso, fez-se uma segunda análise da série A14. Desta forma, estes parâmetros serão avaliados através das funções FAC e FACP.

A FAC e FACP, que podem ser vistas na Figura 5 não estão muito informativas, porém a série possui tendência como visto na Figura 3.1, e possui tendência também nos picos sazonais, portanto, foi aplicada uma diferenciação comum e mais uma diferenciação sazonal a fim de torná-la estacionária. Com a série estacionária vista na Figura 6, foi analisado novamente a FAC e FACP que se encontram na Figura 7. Nota-se que a FAC

e a FACP estão truncadas na primeira defasagem, indicando um modelo com  $p = 1$  e  $q = 1$ , e como não há defasagens sazonais truncados, a princípio não houve necessidade de adicionar os parâmetros  $P$  e  $Q$ . No final, o modelo ajustado manualmente foi um  $SARIMA(1, 1, 1) \times (0, 1, 0)_{12}$  que obteve um AIC de 1084,41, menor que o AIC encontrado na função auto.arima

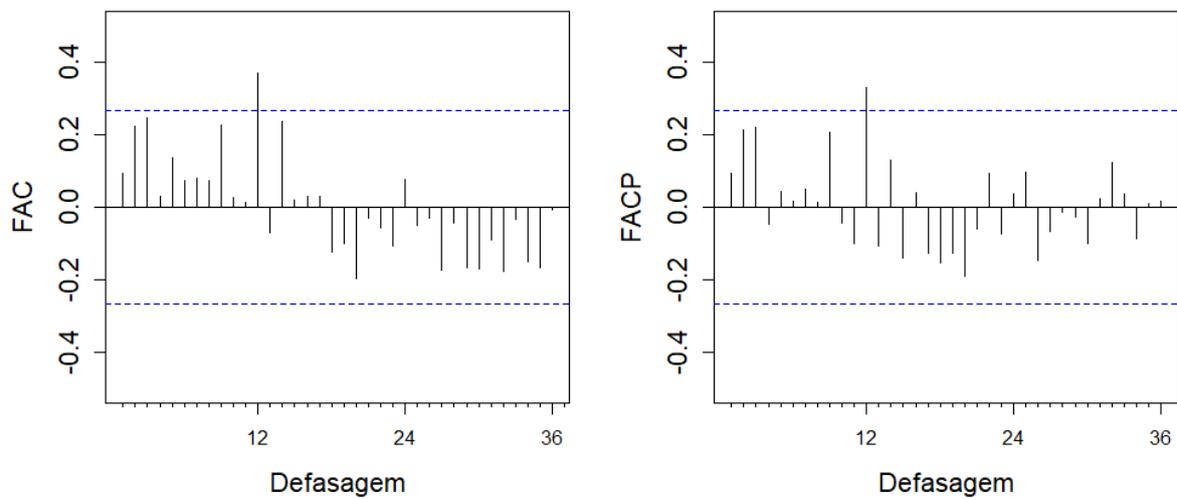


Figura 5: Função de autocorrelação (FAC) e função de autocorrelação parcial (FACP) da série A14.

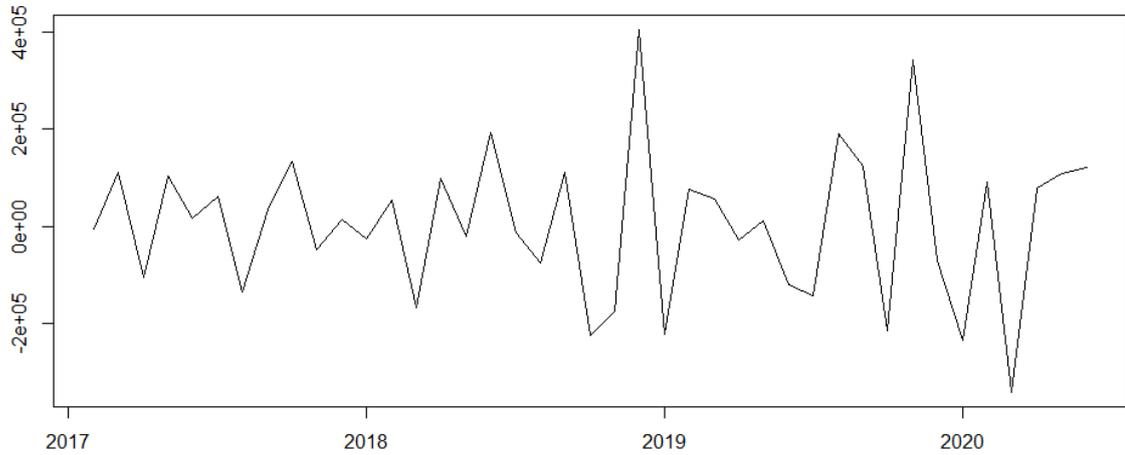


Figura 6: Série temporal diferenciada.

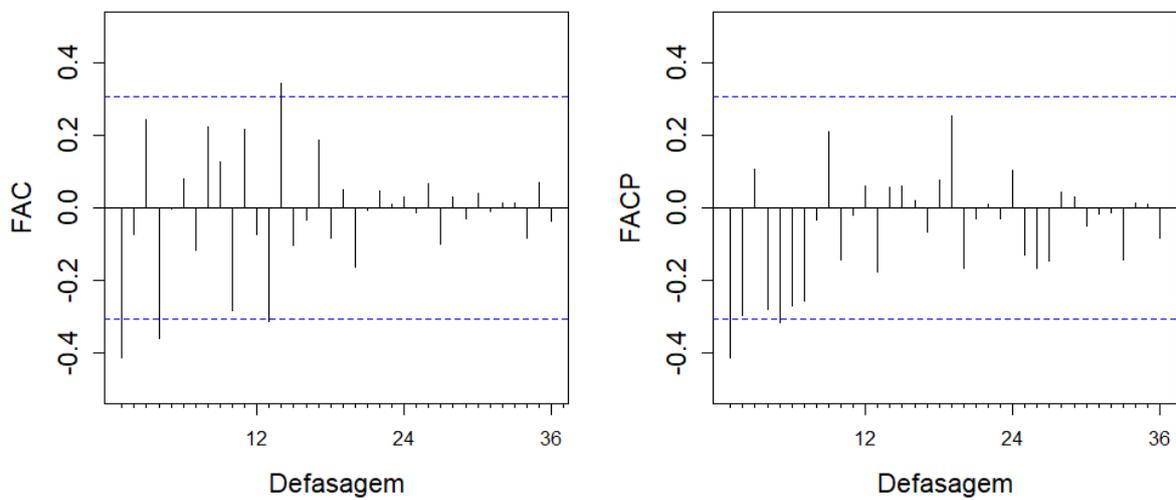


Figura 7: Função de autocorrelação (FAC) e função de autocorrelação parcial (FACP) da série diferenciada.

Para utilizar o método de Floresta Aleatória aplicado a uma série temporal, as covariáveis  $x_{i,p}$  são definidas como as observações passadas ao ponto  $y_i$ , sendo assim, é necessário escolher quantas observações passadas, ou seja, quantas  $p$  covariáveis serão consideradas na regressão. Com o intuito de escolher a melhor quantidade  $p$  de covariáveis,

foi aplicado o método *k-fold* de validação cruzada com  $k = 5$  para  $p = 3, 2, \dots, 12$ , e, para esta série o  $p = 3$  foi escolhido como sendo aquele que apresentou menor Erro Quadrático Médio. Após a escolha do  $p$ , o modelo foi treinado utilizando 500 árvores e considerando todo período de treino e realizou-se a previsão para os 12 pontos do período de teste.

No método LSTM também é necessário definir quantos pontos atrás serão considerados na modelagem. Entretanto, como o método preserva a ordem temporal, não é possível aplicar um método de validação cruzada. Por isso, para escolher entre a quantidade de pontos  $p = 3, 2, \dots, 12$  que serão consideradas como passado do ponto  $y_i$ , foi deixado de fora do conjunto de treino (que possui 54 observações) os últimos 12 meses para serem usados como conjunto de teste para cada um dos valores de  $p$ . O  $p = 12$  foi escolhido pois apresentou menor MSE nesses 12 meses. Depois de escolher a quantidade de pontos ideal que o método deve considerar, realizou-se o treinamento do modelo para os 54 meses do conjunto de treino e feita a previsão para 12 meses futuros.

A Figura 8 contém todas as previsões dos modelos SARIMA, SARIMA manual, Random Forest e LSTM. Observa-se que o SARIMA manual consegue acompanhar melhor os picos do período de teste, entretanto, o SARIMA automático também consegue obter uma boa previsão para os picos. Entre os dois métodos de Machine Learning, o LSTM aparenta ter uma performance melhor que o Random Forest, conseguindo acompanhar a sazonalidade. Na Tabela 2, verifica-se que o modelo que obteve uma melhor acurácia no período de teste foi o SARIMA manual, porém, o SARIMA manual e LSTM obtiveram um MAPE muito próximo.

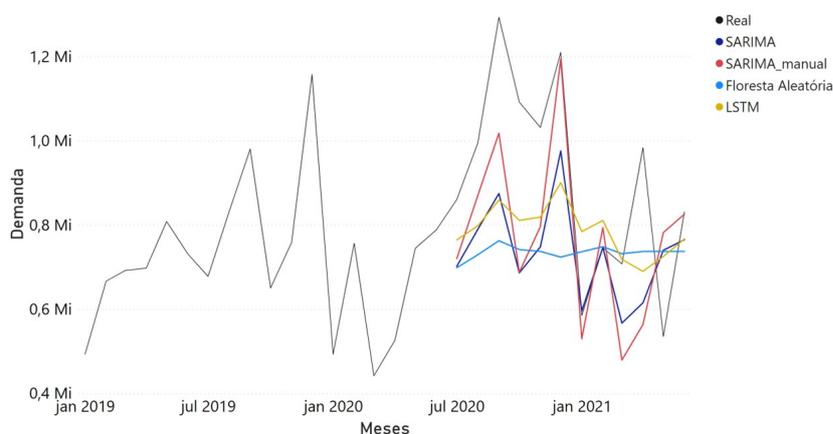


Figura 8: Previsões da Série A14

Tabela 2: MAPE dos Modelos para a série A14

Modelo	MAPE
SARIMA	21,8%
SARIMA manual	<b>20,7%</b>
Floresta Aleatória	24,3%
LSTM	21,1%

De modo geral o erro entre os modelos ficaram próximos entre 20% a 24%, e este erro é considerado baixo no ponto de vista de mercado. Nota-se que apesar do método de Floresta Aleatória entregar uma previsão quase que constante (próximo a média da série) e não acompanhar os picos, ela ainda possui um erro baixo. Isso ocorre muitas vezes porque a série tem um comportamento inesperado e se manter na média garante um melhor resultado. Na Figura 8 existe um pico positivo no final da série que todos os outros modelos previram como sendo um pico de baixa, porém a Floresta Aleatória que se manteve razoavelmente constante ficou mais próxima do real.

## 3.2 Análise de todas as séries

Na prática, para uma empresa com grande portfólio, é inviável analisar e realizar previsão de cada produto manualmente todos os meses, sendo assim, nesta seção as 83 séries restantes foram modeladas automaticamente sem nenhuma análise prévia. Os mesmos passos realizados para a série A14 na Seção 3.1, foram feitos para as 83 séries, exceto pela modelagem manual do SARIMA. O resultado final se encontra na Tabela 3. O modelo que venceu mais vezes, ou seja, aquele que obteve o menor erro no conjunto de teste para a maioria das séries, foi a Floresta Aleatória, já o SARIMA e LSTM empataram.

Tabela 3: Porcentagem de vezes que o modelo venceu

Modelo	% de vezes que ganhou
SARIMA	29%
Floresta Aleatória	42%
LSTM	29%

Na Tabela 4 é apresentado o comportamento do erro para cada modelo em relação às séries que foram modeladas. O chamado de Melhor Modelo é o comportamento do menor erro para cada série. Nota-se que o SARIMA obteve em média o maior erro, assim como a mediana e 3° quartil que também são consideravelmente maiores que os outros métodos. Os métodos LSTM e Floresta Aleatória possuem MAPE com comportamento parecido em relação as séries, tanto o mínimo, 1° quartil, mediana e 3° quartil são muito próximos.

Sendo assim, foi feita a comparação separada do SARIMA com a Floresta Aleatória (FA) e SARIMA com o LSTM vistos nas Tabelas 5 e 6. O SARIMA perdeu em ambos os casos, essa análise corrobora com os resultados obtidos pelos artigos citados na Revisão Bibliográfica vista no Capítulo 1.

Tabela 4: Comportamento do MAPE entre os Modelos e do Melhor Modelo

Modelo	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Média
SARIMA	10,5%	24,1%	37,7%	51,7%	88,6%	39,9%
Floresta Aleatória	10,9%	21,1%	29,2%	39,0%	81,2%	32,0%
LSTM	10,4%	21,5%	30,3%	41,5%	89,9%	34,1%
Melhor Modelo	10,4%	16,9%	24,5%	31,7%	81,2%	27,5 %

Tabela 5: Porcentagem de vezes que o modelo venceu: SARIMA x FA

Modelo	% de vezes que ganhou
SARIMA	39%
Floresta Aleatória	61%

Tabela 6: Porcentagem de vezes que o modelo venceu: SARIMA x LSTM

Modelo	% de vezes que ganhou
SARIMA	45%
LSTM	55%

É válido ressaltar, neste processo de modelagem automática, a importância de treinar vários modelos diferentes para cada série e escolher o melhor modelo para cada uma delas. Analisando o comportamento do MAPE do Melhor Modelo na Tabela 4, tem-se que a média diminui bastante quando é considerado o menor erro. Além disso, observando o 3º quartil do Melhor Modelo, 62 séries possuem um MAPE menor ou igual a 31,7%, que é um erro razoável dado que as séries foram modeladas automaticamente sem considerar nenhum tipo de variável externa, apenas o seu comportamento ao longo do tempo. Sendo assim, o ideal para realizar uma previsão fora da amostra, ou seja, uma previsão para o futuro, é selecionar o modelo que obteve menor erro no conjunto de teste para cada uma das séries, e aplica-lo em toda a série histórica para realizar a previsão futura.

O MAPE de cada modelo para cada série se encontra no Apêndice 2. Apesar do SARIMA ter perdido pra maioria das série, em algumas ele obteve um ganho muito maior em relação aos outros dois métodos. Por exemplo, pra série A21 (vide Apêndice 2) o SARIMA obteve um erro de 16,6% enquanto a Floresta Aleatória e o LSTM obtiveram erros de 39,3% e 24,5%, respectivamente. Uma possível continuação deste trabalho é entender se existe alguma característica nas séries que fizeram com que um método obtivesse um melhor desempenho em relação aos outros.

## 4 Conclusões

Os métodos de Aprendizados de Máquinas, de modo geral, tiveram boa performance. O método LSTM por apresentar características de dependência temporal, consegue acompanhar melhor o comportamento de uma série sazonal do que o método de Floresta Aleatória, por esse motivo, dependendo do problema, ele pode ser um método mais adequado para série temporal. Apesar de aplicações diferentes, os estudos revisados no Capítulo 1 e este trabalho, mostram que os métodos de Aprendizados de Máquinas são promissores e podem agregar pontos de acurácia quando se fala de previsão em relação aos modelos tradicionais SARIMA.

Entretanto, é importante ressaltar que os modelos SARIMA, por serem probabilísticos, são capazes de gerar não somente a previsão pontual, mas também uma previsão intervalar. Portanto, esses modelos podem trazer mais informações sobre a variabilidade da previsão através do intervalo de confiança, e essa informação pode ser muito valiosa para tomar decisões durante o processo de planejamento de demanda. Através do intervalo, pode-se saber o valor máximo e mínimo que a demanda de um determinado mês. Os outros métodos, por outro lado, geram apenas estimativas pontuais.

Uma conclusão complementar que este trabalho obteve foi sobre a comparação entre o modelo SARIMA da função `auto.arima` com o modelagem manual realizada através de análises de correlação. A modelagem manual do SARIMA obteve um AIC de 1084,41 menor que o AIC do modelo ajustado pelo `auto.arima` de 1395,49. Isso significa, uma vez que a função `auto.arima` se propõe a procurar dentre as combinações de  $p, d, q, P, D$  e  $Q$  aquela que possui menor AIC, que ela é limitada e não conseguiu encontrar o mesmo resultado que a modelagem feita manualmente. Entretanto, a função `auto.arima` não é dispensável, já que em relação ao MAPE, ambos modelos obtiveram um erro próximo, o SARIMA manual de 20,7% e o SARIMA automático de 21,8%.

Por fim, a principal conclusão deste trabalho é que os dois métodos de Aprendizado de Máquinas, Floresta Aleatória e LSTM são bons para realizar previsão de demanda,

uma vez que, juntos, ganharam em 71% das séries estudadas. É claro que empresas de setores diferentes possuem séries de demandas diferentes, porém, este trabalho estimula outras empresas a considerarem utilizar Aprendizado de Máquinas como uma alternativa para realizar previsão de demanda e conseguir um ganho de acurácia.

# Referências

- CHARU, C. A. *Neural Networks and Deep Learning: A Textbook*. [S.l.]: Springer, 2018.
- JAMES, G. et al. *An introduction to statistical learning with application in R. 7. Baski*. [S.l.]: Springer Yayinlari, sy, 2017.
- KANE, M. J. et al. Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks. *BMC bioinformatics*, BioMed Central, v. 15, n. 1, p. 1–9, 2014.
- KAWAKAMI, K. *Supervised sequence labelling with recurrent neural networks*. Tese (Doutorado) — Technical University of Munich, 2008.
- MORETTIN, P. A.; TOLOI, C. M. d. C. *Análise de séries temporais*. [S.l.: s.n.], 2004.
- PORTO, P.  
*Métodos Estatísticos de Classificação: Abordagem Aplicada ao Diagnóstico de Casos de Câncer de Mama* — Universidade Federal Fluminense, 2021.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <http://www.R-project.org/>.
- SHUMWAY, R. H.; STOFFER, D. S. *Time series analysis and its applications*. [S.l.]: Springer, 2017.
- SIAMI-NAMINI, S.; TAVAKOLI, N.; NAMIN, A. S. A comparison of arima and lstm in forecasting time series. *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, p. 1394–1401, 2018.

# APÊNDICE 1 – Descrição das Séries Históricas

Tabela 7: Tabela com a descrição das séries

Série	Quant. Produto	Início da Série (aaaa-mm-dd)	Final da Série (aaaa-mm-dd)	Série	Quant. Produto	Início da Série (aaaa-mm-dd)	Final da Série (aaaa-mm-dd)
B01	24	2016-01-01	2021-06-01	B22	4	2016-01-01	2021-06-01
A01	27	2016-01-01	2021-06-01	A22	4	2016-01-01	2021-06-01
B02	1	2016-01-01	2021-06-01	B23	3	2016-01-01	2021-06-01
A02	1	2015-10-01	2021-06-01	A23	3	2016-01-01	2021-06-01
B03	9	2016-01-01	2021-06-01	B24	2	2016-01-01	2021-06-01
A03	9	2016-01-01	2021-06-01	A24	2	2016-01-01	2021-06-01
B04	2	2016-01-01	2021-06-01	B25	1	2016-01-01	2021-06-01
A04	2	2016-01-01	2021-06-01	A25	1	2016-01-01	2021-06-01
B05	4	2016-01-01	2021-06-01	B26	9	2016-01-01	2021-06-01
A05	4	2016-01-01	2021-06-01	A26	9	2016-01-01	2021-06-01
B06	37	2016-01-01	2021-06-01	B27	4	2016-01-01	2021-06-01
A06	41	2015-10-01	2021-06-01	A27	5	2015-10-01	2021-06-01
B07	41	2016-01-01	2021-06-01	B28	1	2016-01-01	2021-06-01
A07	41	2016-01-01	2021-06-01	A28	1	2015-10-01	2021-06-01
B08	11	2016-01-01	2021-06-01	B29	5	2016-01-01	2021-06-01
A08	12	2016-01-01	2021-06-01	A29	5	2016-01-01	2021-06-01
B09	3	2016-01-01	2021-06-01	B30	1	2016-01-01	2021-06-01
A09	3	2016-01-01	2021-06-01	A30	1	2016-01-01	2021-06-01
B10	1	2016-01-01	2021-06-01	B32	3	2016-01-01	2021-06-01
A10	1	2016-01-01	2021-06-01	A32	3	2016-01-01	2021-06-01
B11	1	2016-01-01	2021-06-01	B33	2	2016-01-01	2021-06-01
A11	1	2016-01-01	2021-06-01	A33	2	2016-01-01	2021-06-01
B12	2	2016-01-01	2021-06-01	B34	102	2016-01-01	2021-06-01
A12	2	2016-01-01	2021-06-01	A34	111	2016-01-01	2021-06-01
B13	2	2016-01-01	2021-06-01	B35	4	2016-01-01	2021-06-01
A13	2	2016-01-01	2021-06-01	A35	4	2016-01-01	2021-06-01
B14	2	2016-01-01	2021-06-01	B36	2	2016-01-01	2021-06-01
A14	2	2016-01-01	2021-06-01	A36	2	2016-01-01	2021-06-01
B15	2	2016-01-01	2021-06-01	B37	1	2016-01-01	2021-06-01
A15	2	2016-01-01	2021-06-01	A37	1	2015-10-01	2021-06-01
B16	19	2016-01-01	2021-06-01	B38	12	2016-01-01	2021-06-01
A16	19	2016-01-01	2021-06-01	A38	12	2016-01-01	2021-06-01
B17	2	2016-01-01	2021-06-01	B39	2	2016-01-01	2021-06-01
A17	2	2016-01-01	2021-06-01	A39	2	2016-01-01	2021-06-01
B18	3	2016-01-01	2021-06-01	B40	2	2016-01-01	2021-06-01
A18	3	2016-01-01	2021-06-01	A40	2	2016-01-01	2021-06-01
B19	18	2016-01-01	2021-06-01	B41	120	2016-01-01	2021-06-01
A19	19	2016-01-01	2021-06-01	A41	127	2016-01-01	2021-06-01
B20	2	2016-01-01	2021-06-01	B42	7	2016-01-01	2021-06-01
A20	2	2016-01-01	2021-06-01	A42	7	2016-01-01	2021-06-01
B21	1	2016-01-01	2021-06-01	B44	1	2016-01-01	2021-06-01
A21	1	2016-01-01	2021-06-01	A44	1	2016-01-01	2021-06-01

## APÊNDICE 2 – MAPE dos Modelos

Tabela 8: Tabela com o MAPE dos modelos para cada série

Série	SARIMA	Floresta Aleatória	LSTM	Série	SARIMA	Floresta Aleatória	LSTM
B01	21.7 %	20.8 %	14.3 %	B22	24.4 %	17.8 %	30.4 %
A01	33.3 %	29.2 %	18.8 %	A22	25.3 %	31.2 %	30.3 %
B02	31.8 %	29.2 %	35.6 %	B23	45.9 %	26.4 %	42.1 %
A02	70.3 %	67.1 %	65.7 %	A23	36.1 %	30.3 %	47.1 %
B03	41.4 %	38.4 %	43.5 %	B24	30.9 %	35.4 %	29.2 %
A03	19.7 %	24.5 %	21 %	A24	45.5 %	46.8 %	39 %
B04	28.3 %	26 %	31.1 %	B25	31.8 %	43.4 %	48.9 %
A04	20.1 %	17.5 %	24 %	A25	10.5 %	22.2 %	13.1 %
B05	35.3 %	20.3 %	60.2 %	B26	39.5 %	19.1 %	25 %
A05	18.1 %	13.6 %	14.2 %	A26	26.1 %	20.5 %	20.8 %
B06	30.4 %	24.8 %	24 %	B27	81.3 %	81.2 %	89.9 %
A06	14.2 %	24.2 %	22.1 %	A27	88.6 %	81.1 %	88.8 %
B07	39.4 %	22 %	30.1 %	B28	17.7 %	26 %	23.4 %
A07	12.8 %	11.6 %	10.4 %	A28	39.6 %	30.4 %	34.3 %
B08	25.3 %	14.1 %	16.2 %	B29	46.6 %	10.9 %	18.2 %
A08	15.1 %	42.5 %	36 %	A29	47.3 %	15.9 %	26.1 %
B09	55.3 %	35.9 %	20.8 %	B30	40.2 %	37.6 %	39.1 %
A09	81.4 %	32 %	27.2 %	A30	11.8 %	21.6 %	20.5 %
B10	32.5 %	22.1 %	34.8 %	B32	37.7 %	19.7 %	20.4 %
A10	46.7 %	34.4 %	30.8 %	A32	32.3 %	33.9 %	46.7 %
B11	30.6 %	20.5 %	23.6 %	B33	46.2 %	13.4 %	19.1 %
A11	11 %	19.3 %	14.2 %	A33	51.7 %	26.1 %	35 %
B12	55.5 %	45.7 %	42.6 %	B34	12.5 %	14.3 %	16.9 %
A12	68.1 %	41.3 %	55.4 %	A34	24.5 %	25.3 %	13.3 %
B13	17.2 %	20.8 %	18.5 %	B35	25.8 %	26.7 %	26.8 %
A13	34.6 %	54.7 %	29.5 %	A35	29.4 %	31 %	31.3 %
B14	23.9 %	27.9 %	28 %	B36	38.8 %	39.8 %	63.4 %
A14	21.8 %	24.3 %	21.1 %	A36	86.7 %	40.1 %	44.2 %
B15	15.4 %	19.2 %	17.2 %	B37	51.7 %	58.8 %	59.3 %
A15	10.9 %	14 %	14 %	A37	57.3 %	61.1 %	54.6 %
B16	41.6 %	17.5 %	16.2 %	B38	48 %	29.5 %	33.6 %
A16	77.6 %	43.1 %	40.9 %	A38	82.2 %	30.9 %	43.7 %
B17	53.3 %	21.4 %	28.1 %	B39	35.8 %	40.6 %	29.6 %
A17	82.7 %	25.7 %	33.8 %	A39	68.9 %	72.2 %	88.7 %
B18	73.4 %	24.5 %	24.9 %	B40	10.7 %	33.8 %	28.8 %
A18	88.4 %	39 %	25.5 %	A40	16.5 %	15.1 %	11.9 %
B19	53.7 %	43.3 %	37.9 %	B41	40.2 %	23 %	31.5 %
A19	45.9 %	42.8 %	61.5 %	A41	48.8 %	48.4 %	40.4 %
B20	78.3 %	33.6 %	25.9 %	B42	42.9 %	22.5 %	45.3 %
A20	42.6 %	43.1 %	31.5 %	A42	53.2 %	45.8 %	87.1 %
B21	16.4 %	26.8 %	32.6 %	B44	20.5 %	36.9 %	33.6 %
A21	16.6 %	39.3 %	24.5 %	A44	52 %	53.5 %	51.4 %