



Previsão de Demanda de carros no Brasil: Comparação entre os modelos convencionais e a Rede Neural Recorrente Bidirecional LSTM

Car Demand Forecast in Brazil: Comparison between Conventional Models and LSTM Bidirectional Recurrent Neural Network

Everton Vaz de Campos¹

Resumo: A previsão de demanda tem como finalidade diminuir os custos, aumentar os lucros e construir um controle sobre o gerenciamento dos seus recursos, para o atendimento das futuras demandas. Do mesmo modo que a previsão auxilia para o crescimento competitivo das empresas, permeia a dificuldade nas organizações de implementar ferramentas que sejam claras e objetivas. Dessa maneira, o modelo de rede neural tem a finalidade de usar a inteligência artificial para aprender com os dados passados e prever o futuro. Outrossim, atribuir parâmetros ou variáveis para aumentar a taxa de aprendizagem, ou seja, obter valores de previsão com maior acurácia e precisão. Uma evolução das redes neurais *FeedForward*, são as redes neurais recorrentes Bi-LSTM que consegue lembrar valores em intervalos de tempo arbitrários, regulando o fluxo de informações da rede e resolvendo o problema do desvanecimento do gradiente na rede neural recorrente. Nos comparativos com os modelos convencionais de previsão de demanda, tais como, modelo de média simples, exponencial e modelo auto-regressivo integrado de médias móveis (ARIMA), o modelo de rede neural recorrente Bi-LSTM foi o que obteve melhor acurácia (93,4%) entre todos os testes. A previsão de demanda de automóveis no Brasil para o mês de junho de 2019, com relação ao primeiro mês do ano, é de tendência de crescimento de 1,2% para o modelo de média simples e de 2,78% para o modelo de RNR Bi-LSTM.

Palavras-chave: Previsão de demanda, rede neural, aprendizagem, competitividade.

Abstract: Demand forecasting aims to lower costs, increase profits and build control over the management of its resources to meet future demands. Just as forecasting aids in the competitive growth of companies, it pervades the difficulty in organizations to implement tools that are clear and objective. In this way, the neural network model is intended to use artificial intelligence to learn from past data and predict the future. Furthermore, assigning parameters or variables to increase the learning rate, that is obtaining prediction values with greater accuracy and precision. An evolution of *FeedForward* neural networks, Bi-LSTM recurrent neural networks are able to remember values at arbitrary time intervals, regulating the information flow of the network and solving the gradient fading problem in the recurrent neural network. Compared to conventional demand forecasting models, such as the simple average exponential model and the integrated autoregressive moving average model (ARIMA), the Bi-LSTM recurrent neural network model was the one that obtained the best accuracy (93,4%) among all tests. The forecast for car demand in Brazil for June 2019, compared to the first month of the year, is for a growth trend of 1.2% for the simple average model and 2.78% for the RNR Bi-LSTM.

Palavras-chave: Sales forecasting, neural network, learning, competitiveness.

¹ MBA em Gestão da Produção, pós-graduado na Universidade estadual paulista (UNESP) e graduado em Engenharia de Produção Mecânica pela Faculdade Anhanguera de Taubaté. Email: everton.campos@unesp.br



III Simpósio Acadêmico de Engenharia de Produção

Indústria 4.0 e os Desafios para a Engenharia de Produção

30 e 31
DE OUTUBRO DE 2019
EEL-USP

1 Introdução

Araújo et al. (2018), aborda a previsão como uma ferramenta estratégica da empresa para se tornar mais competitiva perante o mercado. Ele relata que a previsão de demanda, diminui os custos, aumenta os lucros e constrói um controle sobre o gerenciamento dos seus recursos para o atendimento das demandas futuras.

Segundo Almeida (2018), o investimento em novas tecnologias para a tomada de decisão sempre estiveram presentes nas indústrias, como parte de vantagens competitivas. Esse investimento, é essencial para a empresa ter uma estratégia de compra, produção, venda, *marketing*, previsão expansiva, e fluxo de estoque.

Para Aig (2016), o ritmo da tecnologia está em rápida evolução e as demandas dos clientes exigem a comunicação e colaboração. Novas ferramentas, maiores conjuntos de dados e mudanças de expectativas dos clientes, estão exigindo o compartilhamento das informações das empresas para trabalhar em conjunto.

Segundo Almeida (2018), a rede neural pode ser aplicada em modelos de previsões para estipular situações futuras com base nos dados passados.

Segundo Ferreira et al. (2016), o modelo de rede neurais artificiais utilizado na previsão de demanda de encomendas, identificou padrões de comportamento em um ambiente logístico complexo. Ele enfatiza a importância de ter novos experimentos com redes neurais artificiais, baseados em outros problemas de previsão demanda, para a validação do modelo.

A pesquisa tem como objetivo principal fazer um comparativo entre os métodos convencionais de previsão e a rede neural artificial para prever a demanda de carros no Brasil.

2 Previsão de demanda

A previsão de demanda refere-se a uma compra futura do cliente de um determinado produto, prevendo através da extrapolação dos dados históricos de vendas. (PERERA et al., 2019).

A previsão estatística tem como objetivo, através de dados passados, prever os dados futuros; utilizando os modelos matemáticos e estatísticos.



2.1 Média Móvel Simples

A previsão de média móvel simples (MMS) utiliza um número de dados reais passados para prever o futuro. Esse método é útil se pudermos supor que os dados serão estáveis durante todo o período de previsão. (RAHARJO; HAYATI; WAHYUNINGSIH, 2016).

A equação 5 representa o modelo da média móvel simples (MMS):

$$F_T = \frac{Y_{T-1} + Y_{T-2} + \dots + Y_{T-n}}{n} \quad (1)$$

Onde:

Y_T é a Demanda real observado no período T;

F_T é a previsão para o Período T;

n é a quantidade de períodos observados.

2.2 Suavização Exponencial

Segundo Nakano et. al. (2017), uma outra opção efetiva para a previsão é o modelo de suavização exponencial.

$$F_T = \alpha Y_{T-1} + (1 - \alpha) F_{T-1} \quad (2)$$

Onde:

F_T é a previsão para o Período T;

α é a constante de suavização ($0 < \alpha < 1$);

F_{T-1} é a previsão para o período T-1.

2.3 Componente Sazonal

É usado para eliminar o efeito sazonal dos dados da série temporal. Consiste na eliminação de componentes irregulares pela média dos valores mensais ao longo dos anos pretéritos. (SHUKLA; TRIVEDI, 2017)

A componente sazonal é obtida de acordo com as seguintes etapas:

- Classificação e organização dos dados por anos, semestres, trimestres ou meses;
- Cálculo da média dos meses pretéritos;

- Divide o valor de cada mês pela média dos meses, obtendo a componente sazonal;

A Equação 7 representa a componente sazonal:

$$CS = \frac{Y_T}{\bar{Y}_T} \quad (3)$$

Desse modo, a equação de previsão sazonal fica da seguinte forma:

$$F_{TS} = CS \cdot F_T \quad (4)$$

Onde:

CS é a componente sazonal;

F_T é a previsão para o período, seja de média simples ou exponencial.

2.4 Modelo ARIMA

O modelo de Auto-regressivo Integrado de Médias Móveis (ARIMA) tem como finalidade prever valores e movimentos futuros de padrões de dados não estacionários. O mesmo autor relata que o modelo ARIMA é um modelo ARMA de generalização, ou seja, uma combinação de média móvel do modelo autoregressivo. (PEMATHILAKE et al., 2019)

A diferença entre os modelos é que o ARMA é recomendado para prever sequências aleatórias estacionárias, onde suas propriedades não mudam com o tempo; e o modelo ARIMA é indicado para as séries não estacionárias (que mostra algum tipo de tendência ou periodicidade ao longo do tempo), que a partir a diferenciação torna-se estável. (TANG et al., 2018)

De acordo com Kumar e Upadhayay (2019), o modelo ARIMA pode ser descrito como ARIMA (p, d, q), onde: 'p' é o termo autoregressivo; 'd' é o intervalo de erro de previsão futura defasado na equação de predição; 'q' é o intervalo de distinção não sazonal.

A equação 9, representa a função do modelo ARIMA:

$$(1 - \sum_{i=1}^p \varphi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^q \theta_i L^i) \varepsilon_t \quad (5)$$

Onde:

L é o operador de defasagem;

φ é o polinômio ligado ao operador autoregressivo de ordem p;

θ é o polinômio ligado ao operador de média móvel de ordem q;

ε_t é o processo de ruído branco (Média igual a zero e variância constante).



III Simpósio Acadêmico de Engenharia de Produção

Indústria 4.0 e os Desafios para a Engenharia de Produção

30 e 31
DE OUTUBRO DE 2019
EEL-USP

Segundo Lasheras et al. (2015), a identificação do componente estocástico (p, d, q) é definido em 3 (três) etapas, sendo a primeira identificado por tentativas com base em informações preliminares; a segunda, com base no modelo experimental, ou seja, os parâmetros são estimados; e a terceira, através do uso de coeficientes estimados.

Existem três tipos de estatísticas que são utilizados para definir os melhores parâmetros do Modelo ARIMA:

- *Bayesian Information Criterion* (BIC): Baseia-se na comparação de ajustes de modelo usando índice de comparação de modelos, que avalia a precisão e a medida pelo número de cada parâmetro. (RIBEIRO; VIEIRA DE PAULA, 2016)
- Akaike information criterion (AIC): Estima a quantidade de perda de um modelo de previsão, que pode ser usado em conjunto com o parâmetro BIC. Quanto menor o indicador BIC e AIC melhor será o ajuste do modelo. (DZIAK et al., 2019)
- Teste Dickey-Fuller Aumentado (ADF): Teste para a existência de uma raiz unitária para verificar a estacionariedade da série. (RAMENAH; CASIN; BA, 2018)
 - Regra de Decisão:
 - Se $\tau \geq$ Valor crítico $\tau_c \rightarrow$ Aceita H_0 (Y_t não é estacionário)
 - Se $\tau <$ Valor crítico $\tau_c \rightarrow$ Rejeita H_0 (Y_t é estacionário)
- Teste Ljung-Box: Testa todas as autocorrelações dos erros do modelo e não apenas o primeiro *lag* (intervalo de tempo). Sua hipótese nula é quando não existe autocorrelação, ou seja, a soma dos quadrados das autocorrelações é zero; do mesmo modo, quanto maior o valor, maior será a autocorrelação. O modelo deve apresentar os resíduos estacionários (média zero e variância constante). (RIBEIRO; VIEIRA DE PAULA, 2016)

2.5 Índices de Desempenho

Serve para medir a precisão e o desempenho de diferentes modelos de previsão, definindo valores para uma base comparativa. Os quatro indicadores de performance são:

- a) *Mean Absolute Deviation* (MAD) – É o desvio médio absoluto.

$$MAD = \frac{\sum_{i=1}^N |y_i - x_i|}{N} \quad (6)$$

- b) *Mean Squared Error* (MSE) – É o indicador de erro quadrático médio.

$$MSE = \frac{\sum_{i=1}^N (y_i - x_i)^2}{N} \quad (7)$$

- c) *Root Mean Squared Error* (RMSE) – É definido como a raiz quadrada do erro quadrático médio. É um dos indicadores mais utilizados para medir a acuracidade do modelo de previsão em redes neurais.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - x_i)^2}{N}} \quad (8)$$

- d) *Mean Absolute Percentage Error* (MAPE) – É a porcentagem média do erro absoluto, ou seja, o valor percentual dos desvios do modelo de previsão.

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left(\left| \frac{y_i - x_i}{x_i} \right| \right) \cdot 100 \quad (9)$$

Qualquer previsão, independente da complexidade está sujeita a erros, todavia o rastreamento e a análise dos erros fornecem informações úteis sobre a confiabilidade das estratégias de previsão. (MOHAMMED et al., 2017)

Uma outra ferramenta para medir a confiabilidade de previsão é o *Tracking Signal*, que detecta os desvios nos erros de previsão (negativo ou positivo) ao longo da série temporal. (MOHAMMED et al., 2017)

$$Tracking\ Signal = \frac{\sum_{i=1}^N |y_i - x_i|}{MAD} \quad (10)$$

Onde:

y_i é o valor real no período t, na série temporal;

x_i é a previsão no período t, na série temporal;

N é o número de pontos de dados na série temporal;

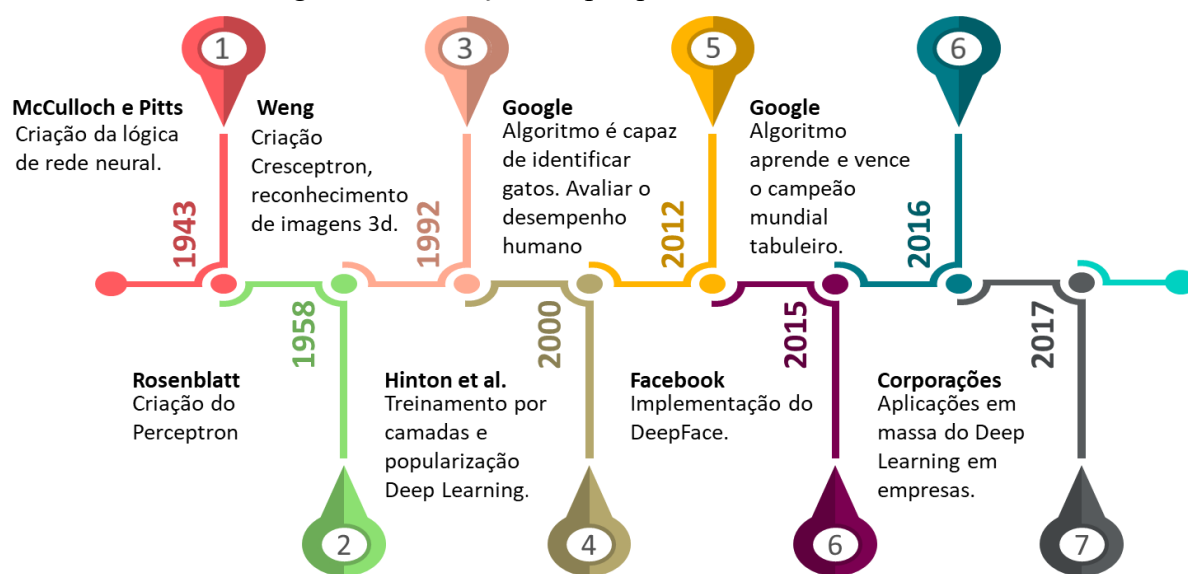
Existe um limite para esse indicador, sendo que os valores que excedem esse número, indicam a não-aleatoriedade nos erros de previsão, o que sugere que a previsão está tendo um desempenho não confiável. Os limites do sinal de rastreamento dependem do intervalo de confiança da previsão sendo comumente utilizados os valores de ± 4 . (MOHAMMED et al., 2017)

3 Rede Neural

Segundo Fleck et al. (2016), as rede neurais são inspirados na estrutura do cérebro humano, os quais possibilitam inserir os modelos matemáticos nos computadores e processá-los. Dessa forma as RNAs são capazes de aprender e tomar decisões de acordo com os treinamentos realizados e disponibilizar esse conhecimento para a melhoria de processos ou aplicações de tecnologias.

Essa evolução das redes neurais, tal como é apresentada nos dias atuais, só foi aplicado devido a tecnologia de processamento de dados. A Figura 1 ilustra a evolução das pesquisas em redes neurais, sendo que nos últimos anos as pesquisas saíram da academia e foram para as grandes corporações, dentre essas: o Google e o Facebook.

Figura 1 – Evolução das pesquisas em redes neurais.



Fonte: Autor (2019).

Desta maneira as redes neurais funcionam com a atribuição de valores a entrada, sendo transformados pelas sinapses (os pesos) e pela função de ativação (AF) dos neurônios.(MACHADO; JUNIOR, 2013)

Recebendo entradas de n neurônios, a saída é processada da seguinte maneira:

$$y_k = AF \left(\sum_{i=1}^n (y_i w_{ki}) + b_k \right) \quad (11)$$

Onde:

y_i é a saída calculada pelo neurônio i



III Simpósio Acadêmico de Engenharia de Produção

Indústria 4.0 e os Desafios para a Engenharia de Produção

30 e 31
DE OUTUBRO DE 2019
EEL-USP

w_{ki} representa o peso (sinapse) entre o neurônio i e o neurônio k

b_k é o peso entre um valor constante e diferente de zero ao neurônio k , conhecido como bias.

Vale ressaltar que se o neurônio que estiver conectado a entrada, o y_i será a própria entrada.

A convergência dos dados se dá por um processo iterativo de atualização dos pesos, chamado de treinamento. Esse treinamento estima os parâmetros dos pesos sinápticos dentre todas as camadas da rede, e é atualizado pela fórmula:

$$w(j)_{ki} = w(j-1)_{ki} + \Delta w(j)_i \quad (12)$$

Onde:

j é a iteração

w_{ki} é o peso

$\Delta w(j)_i$ é o vetor de correção ao parâmetro peso (w_{ki}) na iteração j .

As funções de ativação são primordiais para dar capacidade de aprendizado a uma rede neural. Se houvesse uma rede neural sem uma função de ativação, o mesmo é simplesmente um modelo de regressão linear, sujeitos as mesmas restrições que os modelos lineares. (GOODFELLOW; BENGIO; COURVILLE, 2016)

Segundo Goodfellow et al.(2016), as funções mais utilizadas são a Sigmóide, Tangente hiperbólica (TanH), unidade linear retificada (ReLU), unidade linear exponencial (ELU) e unidade retificada com vazamento (Leaky ReLU).

Para calcular a previsão de demanda, foi utilizado dois tipos de redes neurais: FeedForward e Recorrentes. As redes neurais *FeedForward* ou também chamadas de *perceptrons* multicamadas (MLPs), tem como objetivo que a informação flua da entrada para a saída, não havendo retorno dos dados. (GOODFELLOW; BENGIO; COURVILLE, 2016)

Já as redes neurais recorrentes é um complemento das redes FeedForward, que tem como finalidade utilizar de memórias das camadas anteriores, e armazenar nas camadas posteriores, definindo o que registrar e o que esquecer. (BRITZ, 2015)

Um dos algoritmos mais utilizados nas redes neurais - a descida do gradiente - otimiza funções mais complexas onde existem muitas dimensões. O objetivo da descida do gradiente é achar o mínimo global,

todavia para muitas funções realistas, podem haver muitos mínimos locais. Os mínimos locais devem ser evitados por meios de parâmetros e ajustes no algoritmo para obtenção do valor ótimo da aprendizagem.

Segundo Nelson (2017), as redes de *Long Short-Term Memory* (LSTM) são um tipo especial de rede neural recorrente, que tem a capacidade de resolver a questão do desvanecimento do gradiente, ou seja a dificuldade de processar grandes sequências dos dados.

O LSTM tem a finalidade de manter o fluxo do erro constante, adicionando ou removendo informações ao estado da célula, reguladas por estruturas chamadas portas (*gates*). Esses *gates* são compostos de uma camada de rede neural sigmoide e uma operação de multiplicação. (OLAH, 2015)

Segundo Rosa (2018), um Bi-LSTM é basicamente um par de LSTM que funciona do passado para o futuro, e um LSTM atrasado, que funciona do futuro para o passado. O mesmo autor, relata nesse tipo de rede, o sistema é influenciado pelas entradas passadas e futuras.

4 Resultados

A Figura 2 relata os dados acerca da quantidade de automóveis vendidos entre os anos de 2015 e 2019. A previsão da demanda realizados pelos modelos, se deu a partir de abril de 2017 a fevereiro de 2019.

Figura 2 – Quantidade de automóveis vendidos no Brasil

jan/15	fev/15	mar/15	abr/15	mai/15	jun/15	jul/15	ago/15	set/15	out/15
224.447	186.191	229.052	213.724	221.081	227.616	218.232	208.296	201.207	201.957
nov/15	dez/15	jan/16	fev/16	mar/16	abr/16	mai/16	jun/16	jul/16	ago/16
203.466	236.705	154.012	161.434	194.842	181.310	194.266	196.271	202.988	200.772
set/16	out/16	nov/16	dez/16	jan/17	fev/17	mar/17	abr/17	mai/17	jun/17
177.756	177.317	214.374	238.123	168.533	188.386	237.532	200.562	247.877	242.296
jul/17	ago/17	set/17	out/17	nov/17	dez/17	jan/18	fev/18	mar/18	abr/18
229.626	260.328	239.087	243.722	255.707	247.999	206.675	204.831	250.423	263.667
mai/18	jun/18	jul/18	ago/18	set/18	out/18	nov/18	dez/18	jan/19	fev/19
236.388	240.717	239.857	274.312	226.450	264.436	239.547	237.801	200.890	217.205

Fonte – Portal Brasileiro de Dados Abertos (2019).

4.1 Média Simples

Para o modelo de média móvel simples, o “ η ” variou de 2 a 10, assim como para o modelo de suavização exponencial o α variou de 0,1 a 0,9. Vale ressaltar, que houve a aplicação de sazonalidade para ambos os modelos.

A Figura 3 relata os valores de *Tracking Signal* (TS), a erro médio quadrático (MSE), a raiz do erro médio quadrático e o erro médio percentual absoluto (MAPE) do modelo de média móvel simples e Sazonal. De acordo com a tabela, o melhor resultado da previsão, tendo em vista os indicadores MAPE e RMSE é η igual a 3 (três).

Figura 3 – Resultados de MMS e Sazonalidade.

Média Móvel Simples e Sazonal				
η	TS	MSE	RMSE	MAPE
2	-0,7270	553.981.852	23.537	8,24%
2s	-0,0391	987.942.210	31.432	10,09%
3	0,6166	551.288.257	23.480	7,68%
3s	0,8686	1.064.242.205	32.623	10,35%
4	1,2716	555.316.917	23.565	8,07%
4s	1,5227	1.013.930.429	31.842	10,53%
5	2,2276	559.289.170	23.649	8,07%
5s	2,4151	972.106.264	31.179	10,06%
6	3,1716	577.910.044	24.040	8,35%
6s	3,1017	1.039.229.636	32.237	10,23%
7	4,5679	611.137.360	24.721	8,38%
7s	4,2152	1.067.201.033	32.668	10,34%
8	5,6383	632.552.797	25.151	8,73%
8s	5,0456	1.086.752.688	32.966	10,79%
9	6,5944	663.201.954	25.753	9,01%
9s	6,0118	1.062.318.222	32.593	10,95%
10	7,6272	687.590.832	26.222	9,08%
10s	6,9005	1.082.358.671	32.899	10,96%

Fonte: Autor (2019).

4.2 Suavização Exponencial

A Figura 4 relata os valores de *Tracking Signal* (TS), a erro médio quadrático (MSE), a raiz do erro médio quadrático e o erro médio percentual absoluto (MAPE) do modelo de suavização exponencial assim como de sua sazonalidade. De acordo com a tabela, assim como com a ferramenta “Solver” do Excel, o melhor resultado da previsão, tendo em vista os indicadores MAPE e RMSE, é α igual a 0,377, com um MAPE próximo de 7,71%.

Figura 4 – Resultados de S.E. e Sazonalidade.

Suavização exponencial e Sazonal				
Alfa (α)	TS	MSE	RMSE	MAPE
0,1	12,9683	690.203.562	26.272	9,30%
0,1 (s)	10,5747	1.197.440.490	34.604	11,81%
0,2	6,0618	554.419.151	23.546	8,31%
0,2 (s)	5,3381	1.033.160.125	32.143	10,28%
0,3	2,8096	520.039.741	22.804	7,80%
0,3 (s)	2,7246	992.657.824	31.506	9,82%
0,4	1,0303	514.988.827	22.693	7,74%
0,4 (s)	1,3026	989.062.459	31.449	9,95%
0,5	0,0469	525.893.172	22.932	7,90%
0,5 (s)	0,5322	1.001.016.760	31.639	10,22%
0,6	-0,4886	548.919.902	23.429	8,07%
0,6 (s)	0,1154	1.022.867.061	31.982	10,51%
0,7	-0,7664	582.670.028	24.139	8,38%
0,7 (s)	-0,1064	1.053.751.059	32.462	10,75%
0,8	-0,8997	626.991.621	25.040	8,66%
0,8 (s)	-0,2151	1.095.132.974	33.093	10,99%
0,9	-0,9376	682.901.524	26.132	9,02%
0,9 (s)	-0,2570	1.150.554.197	33.920	11,23%

Fonte: Autor (2019).

4.3 ARIMA

O modelo ARIMA exige que a série seja estacionária, e que não apresentem pontos bruscos de tendência nos dados. Esse comportamento explosivo da série pode inviabilizar o modelo ARIMA.

Desse modo a Figura 5 relata os dados estatísticos do teste Dickey-Fuller para a série original, com um intervalo de tempo de 10 meses. O T estatístico experimental deve ser comparado com o T estatístico tabelado, com o intuito de verificar se deve rejeitar ou aceitar H_0 , portanto, verificar se a série é estacionária ou não estacionária.

O T estatístico tabelado para o $n=50$, com o nível de significância de 5%, é dado por -1,948. Se comparado com o T estatístico experimental para a série original (Figura 5), percebe-se que o valor tabelado é muito inferior ao valor experimental, ratificando a estacionariedade da série. O valor P superior ao nível de significância de 5% corrobora com a não estacionariedade.

Figura 5 – Teste estatístico T.

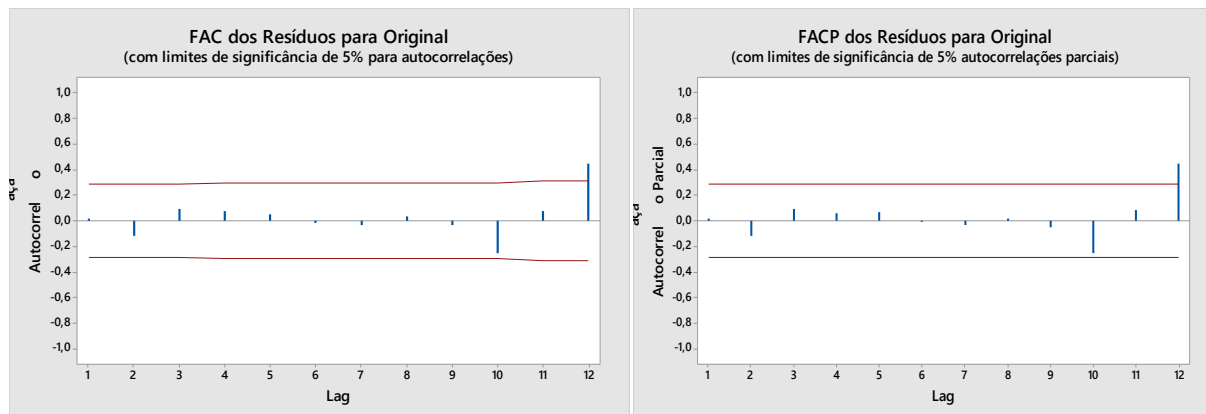
Série Original			Série com uma Diferenciação		
	T estatístico	Valor P		T estatístico	Valor P
Teste Dick-Fuller Aumentado	-0,1772	0,617	Teste Dick-Fuller Aumentado	-8,578	0
Nível 1%	-2,615		Nível 1%	-2,615	
Nível 5%	-1,948		Nível 5%	-1,948	
Nível 10%	-1,612		Nível 10%	-1,612	

Fonte: Eviews® (2019).

Após aplicar uma diferenciação na série original, o T estatístico resulta num valor inferior a -1,95, bem como um valor P significativo para a série diferenciada. Portanto, rejeita-se a hipótese nula (H_0) e determina a estacionariedade da série diferenciada.

Para confirmar a estacionariedade da série, o teste de Ljung-Box foi aplicado na série com uma diferenciação. A Figura 6 relata as funções de autocorrelação e autocorrelação parcial, onde percebe-se que os valores se tornaram estacionários, com média tendendo a zero e com uma variância constante. Conclui-se que na primeira diferenciação da série, que o grau de diferenciação (d) igual a 1, é o mais adequado, não sendo necessário continuar diferenciando a série.

Figura 6 – Autocorrelação e Autocorrelação Parcial.



Fonte: Minitab® (2019).

Para definir os parâmetros (p, q) foi realizado uma simulação de previsão automática ARIMA no software EViews 10, com nível de significância igual a 5% e com parâmetros máximos de diferenciação (2), auto regressão (4) e média móvel (4). O algoritmo mediu os valores de AIC e BIC para definir os melhores parâmetros (p, q).

A Figura 7 indica os valores de AIC e BIC para cada um dos 25 possíveis modelos. O modelo que obteve o menor índice de AIC e BIC foi o ARMA (0, 1), contribuindo para a tomada de decisão do melhor modelo.

Figura 7 – Critério de Seleção ARMA.

Model	LogL	AIC*	BIC
(0,1)(0,0)	-563.47565...	23.12145...	23.23728...
(0,2)(0,0)	-563.45727...	23.16152...	23.31595...
(1,1)(0,0)	-563.46477...	23.16182...	23.31626...
(2,3)(0,0)	-560.66493...	23.16999...	23.44025...
(2,2)(0,0)	-561.68959...	23.17100...	23.40265...
(2,1)(0,0)	-562.72903...	23.17261...	23.36565...
(0,3)(0,0)	-562.74583...	23.17329...	23.36634...
(3,0)(0,0)	-563.06350...	23.18626...	23.37930...
(2,0)(0,0)	-564.14421...	23.18955...	23.34399...
(1,2)(0,0)	-563.17519...	23.19082...	23.38386...
(3,2)(0,0)	-561.47521...	23.20306...	23.47333...
(1,4)(0,0)	-561.58086...	23.20738...	23.47764...
(4,0)(0,0)	-562.61760...	23.20888...	23.44053...
(3,1)(0,0)	-562.70736...	23.21254...	23.44419...
(0,4)(0,0)	-562.72286...	23.21317...	23.44482...
(1,3)(0,0)	-562.73114...	23.21351...	23.44516...
(4,3)(0,0)	-560.00812...	23.22482...	23.57229...
(3,4)(0,0)	-560.06391...	23.22709...	23.57457...
(4,1)(0,0)	-562.61705...	23.24967...	23.51993...
(3,3)(0,0)	-561.67320...	23.25196...	23.56083...
(4,4)(0,0)	-559.90246...	23.26132...	23.64741...
(4,2)(0,0)	-562.61693...	23.29048...	23.59935...
(1,0)(0,0)	-568.28852...	23.31789...	23.43372...
(0,0)(0,0)	-572.77668...	23.46027...	23.53748...
(2,4)(0,0)	-575.92160...	23.83353...	24.14240...

Fonte: Eviews® (2019).

De acordo com os testes realizados anteriormente, a modelagem ARIMA que obteve o melhor ajuste, foi dos parâmetros de (0, 1, 1). Os resultados obtidos dos desvios residuais, bem como da acuracidade do modelo, estão refletidos na Figura 8.

Figura 8 - Resultado da previsão ARIMA

ARIMA (0,1,1)			
TS	MSE	RMSE	MAPE
2,4066	532.180.232	23069	7,88%

Fonte: Autor (2019).

4.4 Rede Neural Artificial

Foram feitos 2 (dois) experimentos para a previsão da série temporal com a rede neural artificial, sendo essas a rede neural *FeedForward* e a rede neural recorrente Bi-LSTM. Mediante aos variados parâmetros que foram testados, os resultados que foram os mais eficientes são relatados na Figura 9.

Figura 9 – Parâmetros das redes neurais.

RNA FeedForward		RNR Bi-LSTM	
Nº de neurônios oculto	100	Nº de Camadas 1º Bi-LSTM	250
Número de épocas	1000	Nº de Camadas 2º Bi-LSTM	250
Treinamento	Bayesian Regularization BackPropagation	Épocas	120
Porcentagem Treinamento	0.80	Porcentagem Treinamento	0.87
Processamento	CPU e GPU	Processamento	CPU
Delay	2	Tamanho Lote	50
		Delay	2

Fonte: Autor (2019).

Vale ressaltar que alguns parâmetros da rede Bi-LSTM, tais como a taxa de aprendizagem e o tipo de treinamento, começam com um valor predefinido e se alteram conforme a aprendizagem do algoritmo, ou seja, existe um ajuste automático desses parâmetros e dos modelos de treinamento.

No processamento da previsão de demanda da série temporal, através do Software Matlab R2018b, os resultados de treinos e testes foram bem distintos, devido aos inúmeros parâmetros implementados. Dessa maneira, os melhores valores de previsões das redes neurais, são indicados na Figura 10.



III Simpósio Acadêmico de Engenharia de Produção

Indústria 4.0 e os Desafios para a Engenharia de Produção

30 e 31
DE OUTUBRO DE 2019
EEL-USP

Figura 10 – Resultados das Redes Neurais.

RNA <i>FeedForward</i>				RNR Bi-LSTM			
	MSE	RMSE	MAPE		MSE	RMSE	MAPE
TREINO	588.930.370	24.268	9,33%	TREINO	684.462.758	26.162	10,18%
TESTE	444.910.252	21.093	8,45%	TESTE	413.033.611	20.323	7,63%
TOTAL	573.928.274	23.957	9,24%	TOTAL	654.303.964	25.579	9,90%

Fonte: Autor (2019).

Para o modelo de RNA *FeedForward*, o software realizou a técnica de validação cruzada, particionando os dados de teste e de treino aleatoriamente. Essa técnica permite avaliar a capacidade de generalização do modelo, tornando-se mais eficaz a aprendizagem da rede neural.

Já a rede neural recorrente Bi-LSTM foi elaborado num algoritmo do Matlab R2018b, que permitiu que os testes e os treinos fossem divididos em sequência. (NEJATIAN, 2019)



III Simpósio Acadêmico de Engenharia de Produção

Indústria 4.0 e os Desafios para a Engenharia de Produção

30 e 31
DE OUTUBRO DE 2019
EEL-USP

5 Conclusão

Percebe-se que a rede neural recorrente Bi-LSTM e o modelo de média simples, obteve resultados satisfatórios na predição da demanda futura de autoveículos no Brasil, possibilitando a programação do estoque de maneira eficiente. Vale ressaltar que na previsão de teste, a rede neural recorrente Bi-LSTM obteve resultados ainda mais satisfatórios, com acuracidade de aproximadamente 93,4%, frente aos 92,3% do modelo de média simples $\eta = 3$.

A acurácia do modelo de previsão exponencial foi similar ao de média simples, porém com incongruência entre os indicadores RMSE e MAPE. Desse modo, a escolha se deu pelo indicador *Tracking Signal* com um valor de -0,81 do modelo de média simples, frente aos 2,06 do modelo de Suavização Exponencial.

A respeito da confiança das previsões, o modelo RNA FeedForward obteve um valor superior a 4, indicando que a previsão está subestimada. Desse modo, considera-se que esse modelo não foi eficaz, e que um possível ajuste nos hiperparâmetros pode ser necessário.

Com relação ao modelo ARIMA (0, 1, 1) que teve seus parâmetros ARMA ajustados pelo critério de informação Akaike e Bayesiano, bem como realizados testes, tais como Dickey-Fuller e Ljung-Box (estacionariedade); foi o quarto melhor modelo, com cerca de 92,1% de acuracidade.

A previsão de demanda de automóveis no Brasil, para o mês de junho de 2019 com relação ao primeiro mês do ano, é de tendência de crescimento de 1,2% para o modelo de média simples e de 2,78% para o modelo de RNR Bi-LSTM. Com isso, pode-se dizer que a previsão de demanda é fundamental para as tomadas de decisões da empresa, para se antecipar e ter uma boa gestão na cadeia de suprimentos, evitando gastos desnecessários ou estoques excessivos.

Dessa maneira, percebe-se que o modelo de rede neural tem uma característica distinta dos modelos convencionais, que é a capacidade de aprendizagem em função do tempo. Quanto mais o algoritmo é treinado e parametrizado de acordo com o padrão da série, maior será a acuracidade do modelo.

Para estudos futuros, a aplicação da taxa de aprendizagem das redes neurais pode ser mais aproveitada, anexando dados que são correlacionados a série. Por exemplo, na previsão de venda de automóveis no Brasil, alguns dados como estado da economia, política e variáveis internacional, podem ser correlacionados como fatores para uma previsão mais concisa. Esses dados externos servem como parâmetros de segurança da empresa, para eventuais problemas na economia ou no mercado internacional, mantendo o estoque de segurança o mais próximo possível da demanda.



III Simpósio Acadêmico de Engenharia de Produção

Indústria 4.0 e os Desafios para a Engenharia de Produção

30 e 31
DE OUTUBRO DE 2019
EEL-USP

Referências

- ALMEIDA, L. G. M. de. **Multi-Layer perceptron para previsão de demanda**. Orientador: José Jair Alves Mendes Junior. 2018. 58 f. Trabalho de conclusão de curso (Graduação em Engenharia de Produção) - Universidade Tecnológica Federal do Paraná, Curitiba, 2018.
- ARAÚJO, G. C. *et al.* Previsão de demanda e análise simplificada da gestão de estoque aplicada a uma empresa do setor alimentício. **Brazilian Journal of Production Engineering**, São Mateus, v. 4, p. 48–64, 2018.
- BRITZ, D. **Recurrent neural networks tutorial part 1: introduction to RNNs**. Disponível em: <http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>. Acesso em: 2 abr. 2019.
- DZIAK, J. J. *et al.* **Sensitivity and specificity of information criteria**. n. 814, p. 1–30, 2019. Disponível em: <http://methodology.psu.edu/media/techreports/12-119.pdf>. Acesso em: 6 mar. 2019.
- EViews®. **Software de análise econométrica**. Versão 10⁺ student version lite: eviews, 2019. Disponível em: <https://www.eviews.com>. Acesso em: 15 fev. 2019.
- FLECK, L. *et al.* Redes neurais artificiais: princípios básicos. **Revista Eletrônica Científica Inovação e Tecnologia**. Universidade Federal do Paraná, Campos Medianeira, v. 1, n. 13, p. 47–57, 2016.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Massachusetts: The Mit Press, 2016. 775 p.
- KUMAR, A.; UPADHAYAY, K. G. Price forecasting technique and methods in deregulated electricity market: an inclusive review. *In*: INTERNATIONAL CONFERENCE ON POWER ENERGY, ENVIRONMENT AND INTELLIGENT CONTROL, Gorakhpur, 2018. **Proceedings [...]**. Gorakhpur: PEEIC, p. 565–571, 2019.
- LASHERAS, F. Sanchez *et al.* Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. **Resources policy**, v. 45, p. 37–43, 2015.
- MACHADO, W. C.; FONSECA JUNIOR, E. S. da. Redes neurais artificiais aplicadas na previsão do VTEC no Brasil. **Boletim de Ciencias Geodesicas**, Curitiba, v. 19, n. 2, p. 227–246, 2013.
- MINITAB®. **Software estatístico**. Versão 18.1: Minitab, 2018. Disponível em: <http://minitab.com>. Acesso em: 11 jan. 2019.
- MOHAMMED, J. *et al.* **Performance of exponential smoothing, a neural network and a hybrid algorithm to the short term load forecasting of batch and continuous loads**. University of the West Indies, Department of Electrical and Computer Engineering, Republic of Trinidad and Tobago, p. 1–6, 2017.
- NAKANO, M.; TAKAHASHI, A.; TAKAHASHI, S. **Generalized exponential moving average (EMA) model with particle filtering and anomaly detection**. Tokyo: Elsevier, v. 73, p. 187–200, 2017.



III Simpósio Acadêmico de Engenharia de Produção

Indústria 4.0 e os Desafios para a Engenharia de Produção

30 e 31
DE OUTUBRO DE 2019
EEL-USP

NELSON, D. M. Q. **Uso de redes neurais recorrentes para previsão de séries temporais financeiras.** Orientador: Adriano Cesar Machado Pereira. 2017. 55f. Dissertação (Mestrado em ciência da Computação) - Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

NEJATIAN, Abolfazl. **Time series prediction.** MATLAB Central File Exchange, 2019. Disponível em: <https://www.mathworks.com/matlabcentral/fileexchange/69506-time-series-prediction>. Acesso em: 2 mar. 2019.

OLAH, C. **Understanding LSTM networks.** Disponível em: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Acesso em: 12 abr. 2019.

PEMATHILAKE, R. G. Hiranya *et al.* Sales forecasting based on autoregressive integrated moving average and recurrent neural network hybrid model. *In: INTERNATIONAL CONFERENCE ON NATURAL COMPUTATION*, 14., 2019. **Proceedings** [...]. Faculty of Information Technology, University of Moratuwa, Moratuwa, p. 27–33, 2019.

PERERA, H. N. *et al.* The human factor in supply chain forecasting: a systematic review. Institute of Transport & Logistics Studies, The University of Sydney, ed. Elsevier: **European Journal of Operational Research**, Sidney, v. 274, n. 2, p. 574–600, 2019.

PORTAL BRASILEIRO DE DADOS ABERTOS. **Vendas de veículos pelas concessionárias:** automóveis. Disponível em: <http://dados.gov.br/dataset/7384-vendas-de-veiculos-pelas-concessionarias-automoveis>. Acesso em: 12 fev. 2019.

RAHARJO, E. S.; HAYATI, M. N.; WAHYUNINGSIH, S. Monitoring forecasting new acceptors KB east kalimantan province using simple moving average and weighted moving average with tracking signal method. **Jurnal EKSPONENSIAL**, Mulawarman, v. 7, p. 17–22, 2016.

RAMENAH, H.; CASIN, P.; BA, M. Accurate determination of parameters relationship for photovoltaic power output by augmented dickey fuller test and engle granger method. **AIMS Energy**, AIMS Press, p. 19 - 48, 2018.

RIBEIRO, L. C.; VIEIRA DE PAULA, A. **Previsão depopulação através dos modelos Arima de Box e Jenkins:** um exercício para Brasil, 2016. Disponível em: <http://www.abep.org.br/publicacoes/index.php/anais/article/download/1067/1031>. Acesso em: 07 jan. 2019.

ROSA, Felipe Bueno da. **Study of a deep learning approach to named entity recognition for portuguese.** 2018. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Instituto de informática, Universidade Federal do Rio grande do Sul, Porto Alegre, 2018.

SHUKLA, G. K.; TRIVEDI, M. **Unit 14 seasonal component.** IGNOU, p. 33–54, 2017. <http://egyankosh.ac.in/handle/123456789/20805>. Acesso em: 12 jan. 2019

TANG, K. *et al.* Research on tool wear prediction based on deep residual network. **Association for Computing Machinery**, Chengdu, p. 73–77, 2018.