TRABALHO FINAL DE GRADUAÇÃO – JUNHO/2019
UNIVERSIDADE FEDERAL DE ITAJUBÁ
ENGENHARIA DE CONTROLE E AUTOMAÇÃO

PREVISÃO DE DEMANDA E ESTOQUE UTILIZANDO ABORDAGENS DE *ADVANCED ANALYTICS*

Anthony Carlleston de Lima[1]

Orientador: Prof. Carlos Waldecir de Souza[1]
Coorientador: Me. Adriano Henrique Rossette Leite[2]
[1]Instituto de Engenharia de Sistemas e Tecnologia da Informação (IESTI)
[2]Itaú Unibanco S.A.

***Resumo* - Este artigo apresenta uma comparação de modelos preditivos clássicos de *Autoregressive Integrated Moving Average* (ARIMA) e Moving Average (MA) com o modelo preditivo de aprendizado de máquina *Random Forest Regressor* aplicados em análise da série temporal das vendas de uma empresa de automação comercial nos últimos 5 anos. Com o objetivo de realizar a previsão das vendas diminuindo a quantidade de produtos em estoque e os riscos de obsolescência dos mesmos. Utilizou-se a linguagem de programação Python e suas principais bibliotecas de ciência de dados como: *Scikit-learn*, *Keras*, *FastAI* e *StatsModels*. Com a utilização dos modelos de aprendizado de máquina, foi possível realizar a previsão das vendas em um período de 4 meses com uma acurácia de 98%, comprovando a efetividade do modelo de aprendizado de maquinas.**

Palavras-Chave: Ciência de Dados, Aprendizado de Máquina, *Random Forest*, *Autoregressive Integrated Moving Average* (ARIMA), Métricas, *Root Mean Square Error*, *Mean Absolute Error*, Pearson *Correlation*, Previsão, Estoque.

I. INTRODUÇÃO

A indústria 4.0 vem ganhando grande destaque em diferentes áreas da engenharia e consiste em realizar a aquisição de dados, analisar padrões e aplicar técnicas de inteligência artificial para que máquinas, plantas industriais e sistemas de atuação possam tomar decisões e se adaptar de maneira inteligente. Com isso, é possível (1) identificar padrões dos conjuntos de dados, (2) aumentar a eficiência da planta industrial, (3) melhorar o consumo energético, otimizar os processos, diminuindo o número de erros no produto, falhas dos equipamentos, bem como, (4) evitar abordagens errôneas feita pelos operadores. Essa tendência aborda tópicos envolvendo *Internet of Things*, Cibersegurança, *Cloud Computing*, Manufatura Avançada, *Machine-to-Machine* (M2M), *Big Data*, entre outras (CAPGEMINI, 2016).

Técnicas de Inteligência Artificial e Analise de dados possui como objetivo: extrair padrões, tendências e interpretações que possam gerar a otimização e traçar melhores caminhos de forma estocástica para auxiliar nas tomadas de decisões (FM2S, 2016). Esse conceito tem sido aplicado a diferentes áreas como: finanças (auxilio na tomada de decisões), médica (classificação de doenças), industrial (manutenção preditiva), logística (identificar melhores rotas), estoque (previsões de demanda), agricultura (otimização do plantio) entre outras (Aquarela, 2015).

O campo da análise de dados vem crescendo de acordo com a utilização de novas tecnologias como a *Advanced Analystics*, *Machine Learning* e a própria *Big Data*, onde diferentes algoritmos e técnicas são implementados para aumentar a acurácia das previsões estocásticas (KHAL,2017).

Um dos grandes desafios do setor de revenda de automação comercial está em aumentar as vendas para reduzir o tempo de estoque de cada produto. Os produtos são importados de diferentes partes do globo e levam de 45 a 60 dias para chegar. Devido ao tempo da logística, os vendedores acabam perdendo algumas vendas, pois os clientes geralmente preferem os produtos a pronta entrega. Um dos grandes receios é fazer encomendas sem solicitações de compras, de tal forma que esses produtos acabam ficando estocado, gerando obsolescência do produto, prejuízo financeiro e perda de espaço no estoque.

O propósito deste artigo foi utilizar o campo da ciência de dados para (1) criar hipóteses, (2) obter os dados, (3) definir o modelo de aprendizado de máquina e (4) comparar com algoritmos clássicos de previsão. As vantagens da previsão das vendas trariam a empresa uma

grande otimização de espaço e lucro, além de diminuir o tempo de negociação dos vendedores com os clientes.
Para que haja maior confiabilidade nos algoritmos de *Machine Learning* utilizados, foi necessário realizar um estudo aprofundado de análise de dados por séries temporais para identificar padrões, tendências, variações cíclicas, variações irregulares e sazonalidades das vendas, consolidando o entendimento do *dataset* que será utilizado.
O *Dataset* utilizado consiste em aproximadamente 7 mil dados de vendas de diferentes tipos de produtos dos últimos 5 anos como: data da venda, localização do cliente no território nacional, nome do cliente, o produto vendido, a quantidade vendida, a quantidade em deposito e o vendedor que realizou a revenda.
Estudos de técnicas de *Machine Learning* realizados sobre variáveis continuas são feitos com base em regressões estatísticas, uma técnica necessária para esses modelos é a *Random Forest Regressor* que será comparada nesse artigo com técnicas de previsão estatísticas clássicas ARIMA e MA. As métricas utilizadas para comparação serão *Pearson Correlation*, *Root Mean Square Error*, *Mean Absolute Error* e a acurácia.

II. Estudo da Série Temporal das Vendas

Para uma boa abordagem e análise de dados é fundamental entender perfeitamente o *dataset*, suas distribuições, correlações, identificar seus *outliers*, sazonalidade e tendências, logo faz-se necessário a análise da série temporal com diferentes tipos de agrupamento para realizar e detectar alguma das características citadas acima.
A análise de séries temporais tem como principal objetivo realizar a previsão (Box & Jenkins, 1994). Neste artigo, a abordagem escolhida para a análise do agrupamento diário, a fim de ter o maior número de observações possível que serão utilizadas nos modelos de previsão.

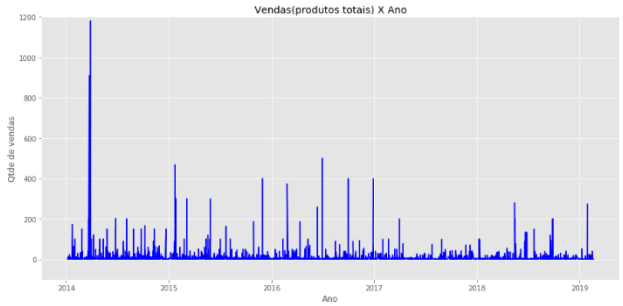


Fig.1 – Série temporal com agrupamento diário.

De acordo com a Figura 1, podemos perceber que há uma grande quantidade de picos de vendas, caracterizadas como *Outliers*, porém não há nenhuma sazonalidade e tendência identificadas a olho nu. Somente com métodos estatísticos poderemos definir a estacionariedade da série.
Os *outliers* se dão devido à grande quantidade de produtos vendidos em projetos e prevê-los é um grande desafio, podendo acabar influenciando diretamente a interpretação dos resultados dos modelos. Para realizar uma abordagem com o menor erro possível foi necessário substituir o *outlier* do ano de 2014 pela média das vendas, devido ao possível impacto causado em nossa análise.

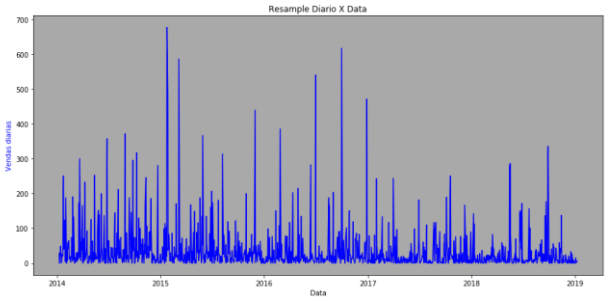


Fig 2 – Série temporal sem o outlier de 2014.

Assim a abordagem desse trabalho será em cima da série temporal da Figura 2, apesar de também apresentar outros *outliers* o pico de venda mais fora do padrão foi removido e com isso podemos continuar nossa análise.

II.1 Estacionariedade

Uma série temporal é estacionária somente se todos os valores de $t_1, t_2,..., t_n$. A distribuição conjunta de $(Y_{t1},Y_{t2},\dots , Y_{tn})$ é a mesma distribuição conjunta de $(Y_{t1-j},Y_{t2-j},\dots , Y_{tn-j})$ para todo o j (John Wiley & Sons, 2009). A importância de saber se uma série é estacionária ou não está em aplicar técnicas de previsão clássica de Box & Jenkins, para uma série não estacionária faz-se necessário a aplicação da diferenciação, deixando-a estacionária (Box & Jenkins, 1994).
Aplicando o teste estatístico *Augmented Dickey-Fuller* para verificar a estacionariedade da série temporal, temos:

ADF Statistic: -31.825421
p-*Value*: 0.000000

Em teoria, o p-*value* é uma forma continua de medir evidencias, mas na realidade é caracteristicamente interligado a altamente significante, marginalmente significante e não estatisticamente significante em níveis convencionais, com cortes entre p ≤0.01, p ≤0.05 e p >0.1 (Gelman, 2012).
Um *p-Value* menor que 0,05 significa que podemos rejeitar o teste de hipótese nula H_o, o que significa que a série possui relação com sua série defasada no tempo (*lag*) e não há raiz unitária. Quanto mais negativo o ADF mais podemos ter certeza de rejeição da hipótese nula (Fuller, 1976).
O p-*value* é uma probabilidade, sob a hipótese de nenhum efeito (hipótese nula), de obter um resultado igual ou mais extremo do que o qual foi realmente observado (Fisher, 1925).
Em suma, apesar de haver uma rejeição da H_o por *Augmented Dickey-Fuller*, significando que a série temporal é estacionária, essa análise é estocástica e não garante com toda certeza a estacionariedade da série podendo haver diferenciação no modelo ARIMA

aumentando a correlação da previsão, como veremos na seção III.

II.2 Métricas de comparação dos modelos preditivos

Como a ideia fundamental desse artigo é comparar o melhor modelo preditivo para as vendas, foi necessário definir as métricas que serão usadas na comparação, foi escolhido RMSE, R, MAE e a acurácia do total das vendas que serão explicadas abaixo para um melhor entendimento das comparações e suas finalidades.
A métrica RMSE, chamada de Raiz do Erro Médio Quadrático é uma das métricas mais empregada para aferir, a qualidade do ajuste dos modelos de previsão. É constituída pela formula a seguir:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{Previsto_i} - \mathrm{Atual_i})^2}{N}} \qquad (1)$$

A métrica RMSE possui a mesma unidade dos valores utilizados em sua formula, assim somente é possível utilizar para comparações que possuem o mesmo universo de discurso.
O Coeficiente de Pearson, conhecido como coeficiente de correlação (R) é a medida do grau de associação do modelo previsto com o modelo atual, ou modelo de teste. Significando que se R ≅ 1 possui dependência linear ascendente e correlação fortemente significativa e R ≅ -1 dependência linear descendente e correlação fortemente significativa. É descrita pela formula a seguir

$$R = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^{n}(x_i - \bar{x})^2][\sum_{i=1}^{n}(y_i - \bar{y})^2]}} \qquad (2)$$

Onde R indica a interdependência dos valores previsto e dos atuais.
A métrica Mean Absolute Error (MAE) é a medida da diferença entre duas variáveis. Em nosso caso, dos valores previstos em relação aos atuais. O cálculo se dá pela seguinte formula:

$$MAE = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N} \qquad (3)$$

Assim como RMSE, a unidade do erro calculado pelo MAE é a mesma em relação ao seu universo de discurso.

A Acurácia é a aproximação entre os valores reais utilizados no conjunto de teste e os valores previstos pelos modelos. Seu calculo consiste pela razão entre o menor valor e maior valor da comparação.

$$\text{Acurácia} = \frac{Valor\ Previsto\ pelo\ modelo}{Valor\ real\ de\ teste} \qquad (4)$$

Sendo o resultado da acurácia uma porcentagem, podendo ser empregado na comparação entre modelos com universos de discursos diferentes.

III. Modelo Auto-regressivo de médias moveis (ARIMA) e modelo de média móvel (MA)

A realização da análise de processo temporal pelo método Box & Jenkins é representado por um grupo de processos estocásticos denominados modelos ARIMA onde, cada momento t, existe um conjunto de valores que a série pode assumir, aos quais estão associados de maneira estocástica a ocorrência dos mesmos valores (Tápia, 2000).
O modelo ARIMA é um composto dos modelos Auto-regressivo de médias (AR) junto com o modelo médias móveis (MA) mais a diferenciação (I), para deixar a série temporal estacionária, assim temos sua representação matemática:

$$\varphi(B)\,\nabla^2 Z_t = \theta(B) a_t \qquad (5)$$

sendo:
φ(B)- Representa o operador auto-regressivo (AR) de ordem p.
θ(B)- Representa o operador médias móveis (MA) de ordem q.
a_t- Ruído branco.
d- Representa o número de diferenças.
∇ =1− B representa o operador diferença.

Utilizando a análise das funções de autocorrelação (ACF) e de autocorrelação parcial (PACF) da série temporal com seus *Lags* para determinar os parâmetros utilizados no modelo ARIMA (Box & Jenkins, 1994), temos:

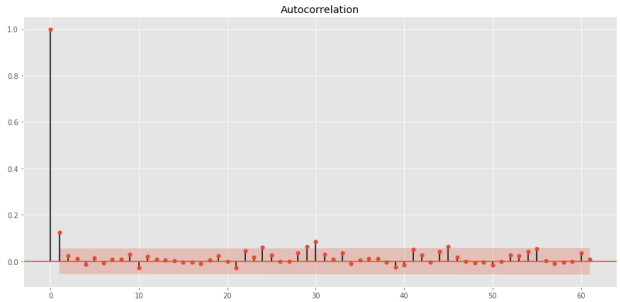


Fig 3 – Autocorrelação (ACF).

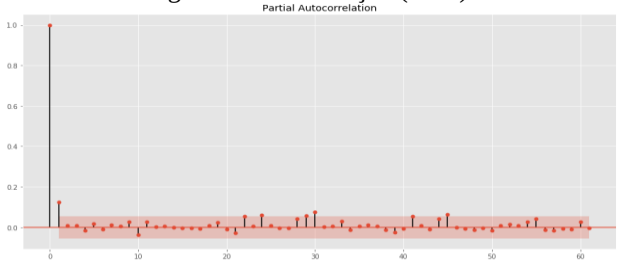


Fig 4 – Autocorrelação Parcial (PACF).

A análise de ACF e PACF é feita observando quais *lags* são influentes para nossos modelos, assim de acordo com Box & Jenkins temos AR (2) e MA (2), pois são os números de *spikes* significantes nessas funções. Utilizando as métricas preestabelecidas, temos:

MAE: 22.274086
RMSE: 47.717937
Corr: 0.112309

Com o intuito de melhorar nossa precisão e correlação do modelo ARIMA, utilizamos um algoritmo automatizado que calcula o Critério de informação de Akaike (AIC), onde foi definido utilizando o critério de possui o menor AIC, o modelo ARIMA (3,1,3), como resultado temos:

MAE: 21.970874
RMSE: 47.764666
Corr: 0.134832

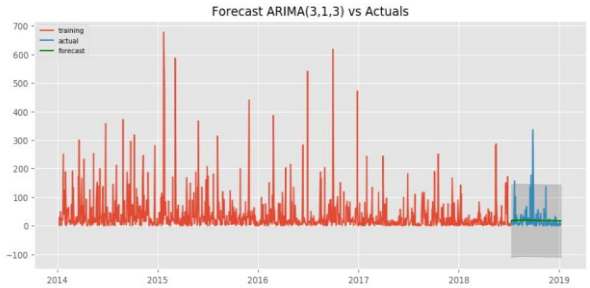


Fig 5 – Série temporal da previsão por ARIMA (3,1,3).

A partição do *dataset* foi definido como 90% dos dados para treinamento do modelo ARIMA e 10% dos dados para o teste do mesmo, essa fragmentação foi utilizada para todos os modelos nesse artigo.

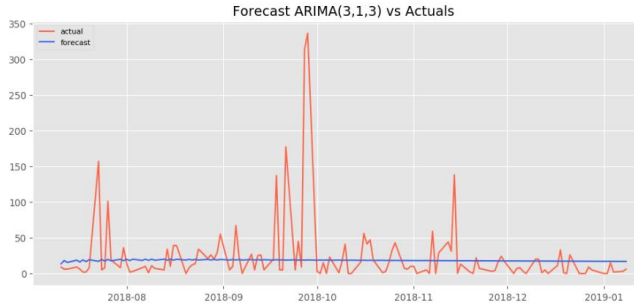


Fig 6 – Abordagem na parte de treino e previsão do modelo ARIMA (3,1,3).

Apesar de graficamente o modelo não possuir uma boa correlação com a partição de teste, realizando a somatória da *forecast* temos 2363, sendo 3033 a somatória do espaçamento de teste, ou seja, dos dados reais, assim esse modelo alcançou 78% de acurácia em relação as vendas do período de agosto de 2018 a janeiro de 2019.
A quesito de curiosidade e comparação utilizamos um modelo puramente composto por média móvel (MA), utilizando a mesma abordagem AIC, obtivemos o melhor modelo como MA (2).

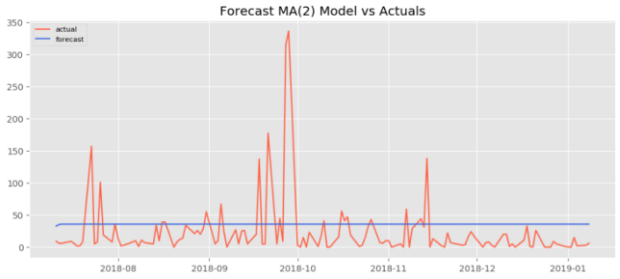


Fig 7 - Abordagem na parte de treino e previsão do modelo MA (2).

Embora graficamente a correlação ficou parecida com o modelo ARIMA, o modelo MA (2) ficou com métricas e abordagens piores, como resultado temos:

MAE: 31.757981
RMSE: 49.197161
Corr: 0.031471

III.1 Modelo ARIMA com série temporal Winsorized

A abordagem nesta subseção foi realizar a técnica de *Winsorization* na série temporal, essa técnica atua limitando os valores extremos, diminuindo a variação da série com a finalidade de melhorar as previsões diminuindo os erros e aumentando a correlação da previsão.
O parâmetro limite da *winsorization* utilizado foi ajustado para 3%, ou seja, uma *winsorization* de 94%, que são calculados utilizando os limites superiores e inferiores.
Logo nossa série temporal sem as *outliers* foi representada pela Figura 12 abaixo.

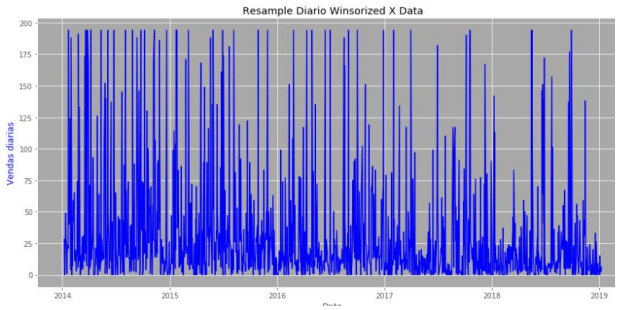


Fig 8 – Série temporal Winsorized.

Assim nossa série temporal apresenta um limite de 194 unidades, possuindo uma constância em seus valores devido a substituição das outliers pela média de vendas.
Aplicando as técnicas ACF e PACF como vista na subseção anterior temos:

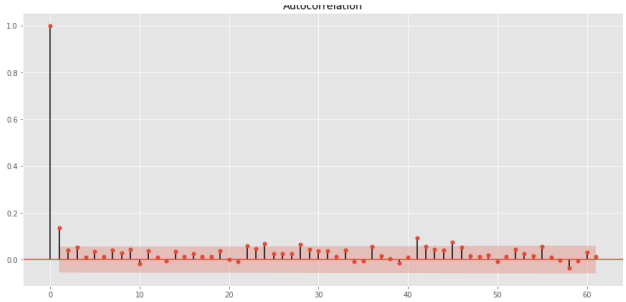


Fig 9 – Autocorrelação da Série Temporal Winsorized.

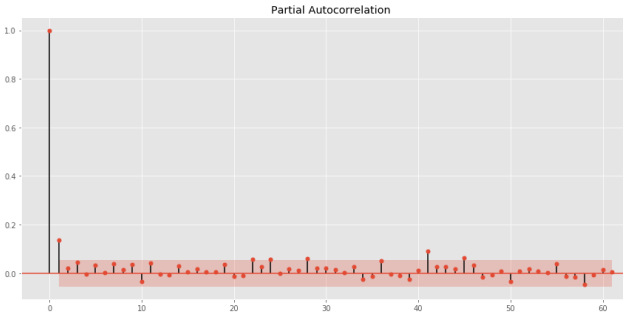


Fig 10 – Autocorrelação Parcial da Série Temporal Winsorized.

É possível perceber que o modelo possui relação com sua defasagem no tempo identificando que o modelo não está sob diferenciado e nem super diferenciado garantindo o procedimento de aplicar o modelo de previsão ARIMA.
De acordo com Box & Jenkins devemos considerar somente os picos significativos, ou seja, que ultrapassam o intervalo de correlação das funções ACF e PACF para estimar os coeficientes do modelo, logo como melhores parâmetros iniciais AR (2) e MA (2). Temos como resultado do menor AIC, utilizando o algoritmo de encontrar a melhor combinação dos parâmetro p,d,q para o modelo ARIMA, temos o ARIMA (2,1,4), assim temos.

MAE: 19.797902
RMSE: 36.212122
Corr: 0.133885

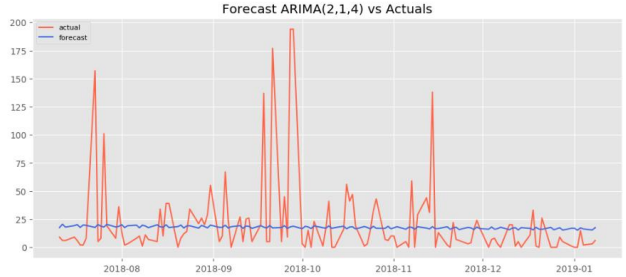


Fig 11 – Abordagem na parte de treino e previsão do modelo ARIMA (2,1,4) série temporal *winsorized*.

Não houve melhora em relação a correlação do modelo. Devido a redução das métricas MAE e RMSE podemos constar ganhos em nossa previsão, porém parte desses ganhos podem ser oriundos da mudança do universo de discurso, onde a somatória da previsão foi de 2301 unidades, sendo o total das vendas para o modelo sem *outliers* de 2771 unidades, logo a previsão corresponde a 83% de acurácia da série temporal *winsorized*.
A quesito de comparação foi utilizado um modelo puramente Média Móvel, onde MA (2):

MAE: 27.352589
RMSE: 37.773515
Corr: 0.036208

Assim temos a somatória dos produtos no período de 4188 unidades, com uma acurácia de 72%, porém devemos descartar a hipótese de utilizar modelo que sobrestimam as vendas devido ao problema de estoque definido na Seção I desde artigo.

IV. Modelo de Aprendizado de Máquina Random Forest Regressor

Para definir a técnica de inteligência artificial escolhida para a previsão do estoque foi levado em consideração os tipos dos dados analisados, a quantidade de dados disponíveis para o treinamento, o objetivo final e o tipo de aprendizado de máquina.
Devido a quantidade de dados ser pequena e do tipo continuas, foi definido a necessidade de um modelo regressivo ao invés de um classificador. Por não termos um cluster disponível para a otimização de hiperparâmetros e algoritmos mais custosos em âmbito de processamento, o melhor modelo de *Machine Learning* que foi encontrado e definido foi o *Random Forest Regressor*.
As árvores de decisão são fáceis de entender e interpretar, obtendo uma previsão completamente transparente, podem lidar com diferentes tipos de variáveis como as continuas e categóricas, apesar de ao mesmo tempo encontrar a árvore perfeita ser uma tarefa computacionalmente difícil (Joel Grus, 2015).

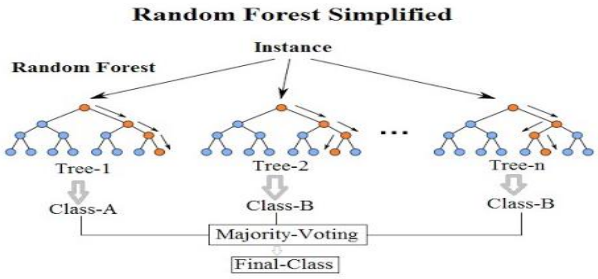


Fig 12 – *Random Forest* Simplificada.
Fonte: KOEHRSEN, Will. Random Forest Simple Explanation, 2017.

Após realizado um estudo de inteligência de negócio internamente na empresa para definir quais índices que poderiam influenciar nas vendas dos produtos, foram determinados 10 tipos: Dólar, Varejo, faturamento industrial e os 6 tipos de indicadores de crescimento industriais (Bebida, Maquina, Metal, Têxtil, Transformação, Alimentícia) que foram utilizadas como nossas variáveis características para o modelo.

Os dados foram encontrados em diferentes sites como IBGE e sistema industrial, links na seção Referências.

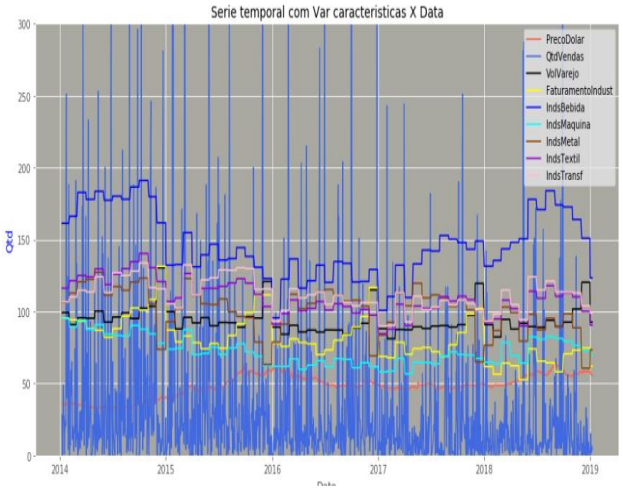


Fig 13 – Série temporal com variáveis características.

III.1 Identificação da importância das variáveis características

Os primeiros passos realizados foram identificar a importância das variáveis características e seus impactos sobre nossa variável alvo, onde realizamos 4 estudos: (1) a importância das variáveis características padrões, (2) as variáveis características defasadas em 30 dias, (3) 60 dias e (4) 90 dias.

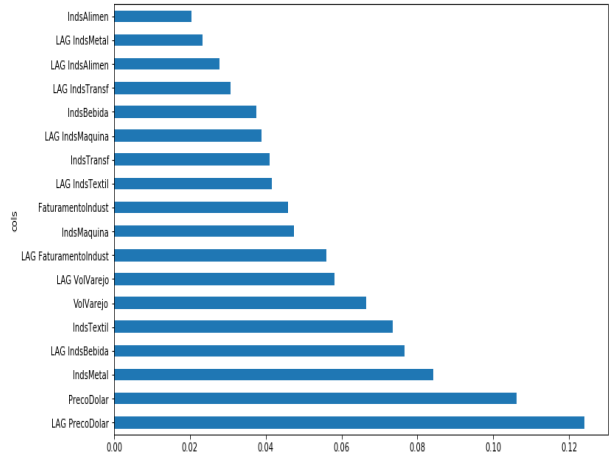


Fig 14 – Impacto das variáveis características defasadas de 30 dias.

Como resultado das variaveis caracteristicas defasadas mais importantes, temos: o indicador industrial de metalurgica *lag* de 30 dias, o preço do dolar com *lag* de 30 dias, volume de varejo lag de 60 dias e preço do dolar com lag de 60 dias. Compondo o conjunto de 13 variaveis independente que mais agregam e impactam nosso modelo.

III.2 Aprendizado de Máquina: *Random Forest Regressor*

Após definido nossa variável dependente e independentes na Seção anterior realizamos a otimização dos hiperparâmetros da nossa arvore. Essa tarefa foi realizada após uma varredura de 30 horas pela técnica de *tuning Grid Search*. A técnica consegue definir: o número de arvores, split mínimo, mínima quantidade de folhas, profundidade máxima e número máximo de variáveis características utilizadas no modelo. Apesar de ser uma técnica computacionalmente mais custosa, por realizar uma varredura completa de todas as combinações, foi definida por garantir que os hiperparâmetros ditos como resultado são os melhores dentro dos conjuntos definidos para a aplicação inicial da técnica.

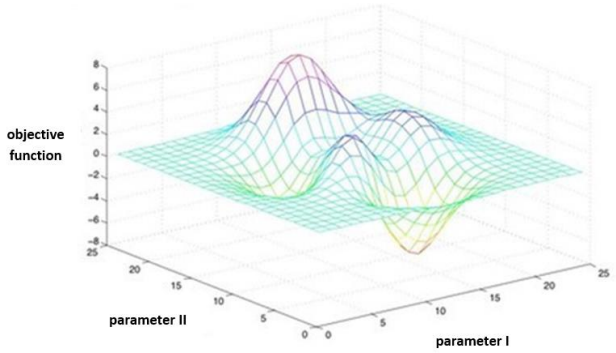


Fig 15 – Grid Search.
Fonte: DANGETI, Pratap. Statistics for machine learning. Packt Publishing Ltd, 2017.

Os hiperparâmetros utilizados foram: n_*estimators*=10, min_samples_*split* =14, *random_state=409*, *min_samples_leaf*=12, *max_features*=8, *max_depth*=100.

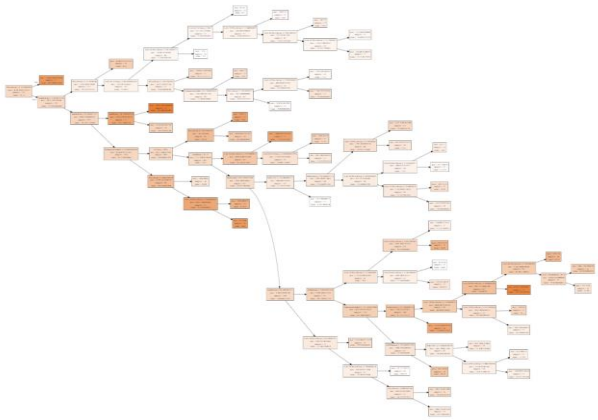


Fig 16 – Floresta Aleatória do modelo.

Apesar das métricas serem próximas aos outros modelos preditivos, o que chamou atenção foi a somatória das vendas previstas nesse modelo, obtivemos como previsão 2992 unidades sendo que a realidade tivemos 3033 unidades, como podemos perceber tivemos 98,6% de acerto se tratando da somatória no período, então:

MAE: 25.451409

RMSE: 50.232907
Corr: -0.245729

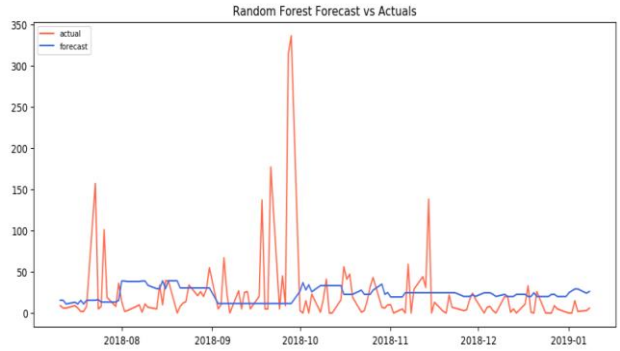


Fig 17 –Previsão da Random Forest Regressor.

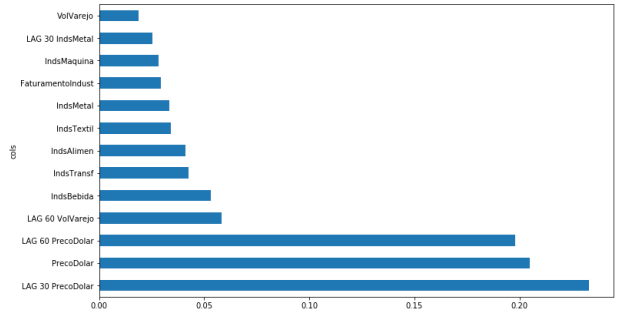


Fig 18 – Importância das variáveis característica para o modelo

Devido ao elevado grau de acurácia do modelo, podemos analisar a importância das variáveis características e como elas influenciam nas vendas. A variável que mais possui impacto nas vendas segundo nosso modelo de aprendizado de máquina é o preço do dólar com uma defasagem de 30 dias, além de constar uma perda de importancia das outras variáveis em relação aos preços do dolar. A linearidade da relação entre as vendas com o preço do dolar defasado em 30 dias possui uma correlação inversamente proporcional.

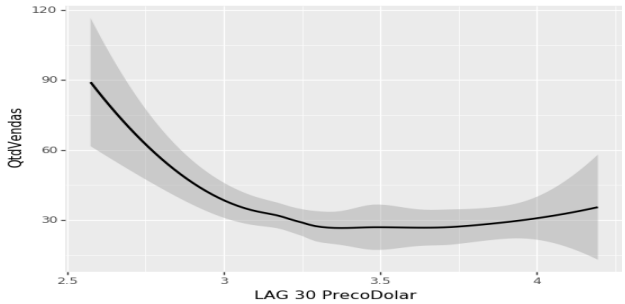


Fig 19 – Relação entre as vendas e o preço do dólar defasado em 30 dias.

III.2 Random Forest Regressor para série temporal Winsorized

Nessa subseção foram realizados os mesmos procedimentos com o objetivo de comparar os dois modelos de *random forest regressor* com e sem outliers. A importância das variáveis e seus impactos tiveram mudanças, sendo o preço do dólar defasado em 60 dias a variável que mais possui correlação com o modelo *winsorized*.
As métricas desse modelo indicam uma fraca correlação com a partição de teste, no qual os erros ficam bem próximos dos modelos preditivos clássicos.

MAE: 20.956382
RMSE: 37.248192
Corr: -0.209614

Assim como o primeiro modelo de *Random Forest* a correlação também ficou significantemente fraca, apesar do modelo possui um RMSE e MAE superior aos outros modelos de previsão com série temporal *Winsorized*, a somatória da previsão foi de 2477 unidades, sendo a somatória do real sem as outliers de 2771 unidades, uma acurácia de 89,4%.

V. CONCLUSÃO

Os resultados das métricas dos modelos preditivos gerou um curioso resultado em relação as comparações dos erros e do objetivo deste projeto. Os modelos clássicos obtiveram os menores erros e uma acurácia relativamente boa, sendo que os modelos de inteligência artificial obtiveram a maior acurácia e correlação com os dados de teste.

Tabela 1 – Comparação dos modelos Preditivos

Série Temporal	Modelos	Métricas			
		MAE	RMSE	R	Acurácia
Normal	ARIMA	21.971	47.764	0.1348	78%
	MA	31.758	49.197	0.0315	72%
	Random Forest	25.451	50.233	-0.2457	98%
Winsorized	ARIMA	19.798	36.212	0.1339	83%
	MA	27.352	37.774	0.0362	72%
	Random Forest	20.993	37.248	-0.2096	89%

Tendo como principal objetivo deste projeto realizar a previsão das vendas, os modelos aplicados se mostraram efetivos, porém se tratando da identificação de sazonalidades, tendências e picos de vendas ambos modelos falharam nesse proposito.
Os modelos de aprendizado de máquina conseguiram uma boa previsão em relação a quantidade de produtos que devem ser solicitados para o estoque.
A partição de teste utiliza dados dos últimos 4 meses da série temporal, sabendo que os produtos importados levam 2 meses para chegar, obtivemos um ganho significativo devido a essa previsão a curto prazo, como principal ganho os modelos nos trará uma melhora em relação ao tempo

que os produtos ficam parados, tendo como certeza que as vendas acontecerão em um prazo de 4 meses.
Como trabalhos futuros, iremos validar o modelo com dados recentes, aplicar o estudo de forma individual para cada produto e encontrar variáveis características com um elevado grau de importância para nosso modelo.

V. AGRADECIMENTOS

Agradeço a Deus por toda saúde, sabedoria para a realização deste artigo e por toda força que me foi concedida durante minha jornada acadêmica. Minha eterna gratidão ao Adriano Henrique Rossette Leite que com seu conhecimento e análise critica me guiou sabiamente pelo fascinante mundo da Ciência de Dados que até então era uma área inexplorada para mim. Ao professor Carlos Waldecir de Souza por aceitar esse desafio. A minha família por compreender meus momentos de ausência e me ajudar da melhor maneira que podiam durante todos os desafios enfrentados em minha graduação.

VI. REFERÊNCIAS

CAPGEMINI. Industry 4.0 - The Capgemini Consulting View: Sharpening the Picture beyond the Hype. 2016. Disponível em: <https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/07/capgemini-consulting-industrie-4.0_0_0.pdf>. Acesso em: 03 mar, 2019.

DAVENPORT, Thomas P.; COHEN, Don.; JACOBSON, Al. Harvard Business Review: Competing on Analytics, p.49-60, 2005.

CAMHI, Jonathan - EDGE COMPUTING IN THE IoT: Forecasts, key benefits, and top industries adopting an analytics model that improves processing and cuts costs, 2016.

WILEY, JOHN – Introduction to econometrics, 2009.

HURWITZ, Judith; KIRSCH, Daniel - Machine Learning for Dummies, 2018.

WILSON, H. JAMES – Human + Machine: Reimagining Work in the Age of AI, 2018.

MCKINNEY, WES – Python para Análise de Dados, 2018.

GRUS, JOEL – Data Science do Zero, 2016.

Dados do volume de varejo. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/comercio/9227-pesquisa-mensal-de-comercio.html?=&t=o-que-e> Acesso em: Mar, 2019.

Dados de indicadores industriais. Disponível em <http://www6.sistemaindustria.org.br/gpc/externo/estatisticaAcessoSistemaExterno.faces>. Acesso em: Mar, 2019.

Aishwarya Singh, An Introduction to Random Forest using the fastai Library, 2018.

Aishwarya Singh, An Intuitive Guide to Interpret a Random Forest Model using fastai library. Disponível em: <https://www.analyticsvidhya.com/blog/2018/10/interpret-random-forest-model-machine-learning-programmers/>. Acesso em: Mar, 2019.

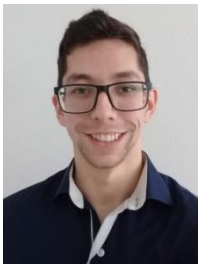
Aishwarya Singh, Building a Random Forest from Scratch & Understanding Real-World Data Products. Disponível em:<https://www.analyticsvidhya.com/blog/2018/12/building-a-random-forest-from-scratch-understanding-real-world-data-products-ml-for-programmers-part-3/> Acesso em: Abr, 2019.

Will Koehrsen, Hyperparameter Tuning the Random Forest in Python. Disponivel em:<https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74>. Acesso em: Abr, 2019.

MACKINNON, James G.; Numerical Distribution Functions for Unit Root and Cointegration Tests, 1996.

FULLER, W. A.; Introduction to Statistical Time Séries, 1976.

BIOGRAFIA:



Anthony Carlleston de Lima
Nasceu em Pouso Alegre (MG), em 1994. Serviu o Exército Brasileiro no ano de 2013, onde foi promovido a cabo em 8 meses, deixando o serviço militar obrigatório com certificado de honra ao mérito. Ingressou na Universidade Federal de Itajubá em 2015 no curso de Engenharia de Controle e Automação. Durante sua trajetória acadêmica, participou da Equipe Uairrior de Guerra de Robôs, Equipe Robok de Futebol de Robôs, Iniciação cientifica projetando um controlador Fuzzy e Neuro-Fuzzy para um sistema Aeroestabilizador e Projeto Embaixador na empresa francesa Thales *Group*.