

Aplicação de Modelos Nebulosos Univariados e Multivariados na Previsão de Preços de Minério De Ferro: Um Estudo Comparativo

Hamilton Tonidandel Jr. *, Frederico Gadelha Guimarães**

* Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração, Universidade Federal de Ouro Preto e Instituto Tecnológico Vale, Ouro Preto, MG (e-mail: hamilton.junior@ufop.edu.br)

**Departamento de Engenharia Elétrica, Universidade Federal de Minas Gerais, Universidade Federal de Minas Gerais, Belo Horizonte, MG (e-mail: fredericoguimaraes@ufmg.br)

Abstract: Iron ore ranks second among the highest value items exported by Brazil in 2020, with Vale S.A. being the largest exporter of this commodity in the world. At this point, estimating the future behavior of a short-term iron ore price time series is an important tool at mining projects, especially in decision-making related to operational planning. The present work evaluates, in this context, the accuracy of the fuzzy models PWFTS (Probabilistic Weighted Fuzzy Time Series) and FDT (Fuzzy Decision Trees) in iron ore price prediction, by presenting a comparative study with the predictive models: ARIMA, Multilayer Perceptron (MLP) and Xgboost. In order to ensure the variability of input patterns, the data are distributed into subsets, using the sliding windows technique. In a multivariate context, predictor variables are selected through correlational analysis with the target time series, with emphasis on the inclusion of the iron ore surplus production from Vale S.A. Results for the univariable and multivariable modeling indicate superiority of the fuzzy models, according to RMSE and MAPE metrics.

Resumo: O minério de ferro se coloca na segunda posição dos itens de maior valor exportados pelo Brasil em 2020, sendo a Vale S.A. a maior exportadora desta commodity no mundo. Neste ponto, a estimativa do comportamento futuro da série temporal de preços de minério de ferro a curto prazo é uma ferramenta importante na elaboração de projetos de mineração, em especial na tomada de decisões relacionadas ao planejamento operacional. O presente trabalho avalia, neste contexto, a acuracidade dos modelos nebulosos PWFTS (*Probabilistic Weighted Fuzzy Time Series*) e FDT (*Fuzzy Decision Trees*) na previsão de preços de minério de ferro, ao apresentar um estudo comparativo com os modelos preditivos ARIMA, *Multilayer Perceptron* (MLP) e *Xgboost*. Para garantir a variabilidade dos padrões de entrada, os dados são distribuídos em subconjuntos, por meio da técnica de janelas deslizantes. Em um contexto multivariado, variáveis preditoras são selecionadas por meio de análise correlacional com a série temporal principal, com destaque para a inclusão do excedente produtivo de minério de ferro da Vale S.A. Os resultados para a modelagem univariável e multivariável indicam superioridade dos modelos nebulosos nas métricas RMSE e MAPE.

Keywords: Time series; Forecasting; Iron ore; Fuzzy logic; Multivariate series; Operational planning.

Palavras-chaves: Séries temporais; Previsão; Minério de ferro; Lógica nebulosa; Planejamento operacional.

1. INTRODUÇÃO

Dentre as *commodities* minerais, o minério de ferro se destaca por ser um recurso vital para a economia e para a sociedade, sendo utilizado como matéria-prima na produção do ferro-gusa, posteriormente transformado em aço bruto, que é então empregado na produção de automóveis, estruturas metálicas para construção civil, entre outros fins. Neste aspecto, o Brasil representa um importante papel, sendo um dos principais exportadores da *commodity*. De fato, o mercado de abastecimento de minério de ferro atual se concentra majoritariamente nos chamados “*big three*”, isto é, Vale S.A. – Brasil, Rio Tinto – Austrália, e BHP Biliton – Austrália (Ma and Zhen, 2020). Segundo dados da plataforma ComexStat (2020), o Brasil exportou aproximadamente 326

milhões de toneladas de minério de ferro em 2020, o que coloca a *commodity* na segunda posição dos itens de maior valor exportados pelo Brasil naquele ano. Do ponto de vista da importação de minério de ferro, a China concentra a maior parte das transações, movimentando mais de 1 bilhão de toneladas em 2020 (Trademap, 2020).

De acordo com Aminrostamkolae et al. (2017), a expectativa futura de comportamento do preço do minério de ferro tem se tornado uma questão central para a elaboração de projetos de mineração, sendo uma variável importante no planejamento operacional, especialmente para o melhor aproveitamento do ciclo de vida de uma mina em atividade, contribuindo nas tomadas de decisões da empresa mineradora. No caso de uma projeção de preços em queda,

por exemplo, as opções de atraso momentâneo na produção, desaceleração ou mesmo abandono da mina podem ser consideradas. Por outro lado, dada uma projeção de alta nos preços, é possível aplicar estratégias de aceleração, ou mesmo expansão das operações na mina.

Segundo Li et al. (2020), diferentes técnicas estatísticas e baseadas em inteligência artificial têm sido sugeridas para a previsão de séries financeiras aplicadas em *commodities*. Além dos métodos tradicionais, os modelos nebulosos também se apresentam como uma alternativa viável para previsão de variáveis diversas relacionadas ao mercado financeiro (Egrioglu et al., 2020). Destaca-se, neste ponto, o trabalho de Cheng and Chen (2018), que utiliza a modelagem por séries nebulosas com relações lógicas de alta ordem, e aplica o modelo na previsão dos preços de ouro e taxas de câmbio.

De acordo com Ewees et al. (2020), um dos maiores desafios de se elaborar um modelo preditivo com boa acurácia para uma série financeira é a escolha das variáveis predictoras externas, isto é, variáveis que podem ser correlacionadas com a série principal e agir como auxiliares na previsão, no caso de modelos multivariados. Para tanto, este trabalho contempla a análise de tais variáveis, evidenciando a correlação com a série de preços de minério de ferro e o impacto na performance dos modelos em relação à acuracidade na estimativa, diante do horizonte de previsão predefinido.

Acerca do horizonte de previsão, Beasley (2006) sugere que métodos qualitativos são mais adequados para previsões de longo prazo, ao passo que a análise preditiva de séries temporais, especialmente associadas a variáveis externas, como ocorre em modelos econômicos multivariados, é mais eficaz em um horizonte de curto prazo (*short-term*), que, na maior parte das aplicações industriais voltadas ao planejamento operacional, se trata de um intervalo de 3 a 6 meses.

Atestada, portanto, a importância da previsão dos preços de minério de ferro no plano de tomada de decisões da indústria mineradora, o objetivo deste trabalho é estudar a aplicação de modelos preditivos nebulosos, univariados e multivariados, para estimar o comportamento futuro da série temporal de preços médios de compra e venda de minério de ferro refinado com teor de 62%, em dólares por tonelada. Para tanto, é feita uma análise comparativa com o modelo estatístico ARIMA, e com outros modelos baseados em inteligência computacional (*Multilayer Perceptron* e *Xgboost*). Ao final, os melhores modelos são escolhidos com base nos parâmetros estatísticos das métricas de acuracidade RMSE e MAPE, para diferentes subconjuntos de dados de treino e teste.

2. METODOLOGIA

O procedimento metodológico utilizado neste estudo está sumarizado no fluxograma da Fig. 1.

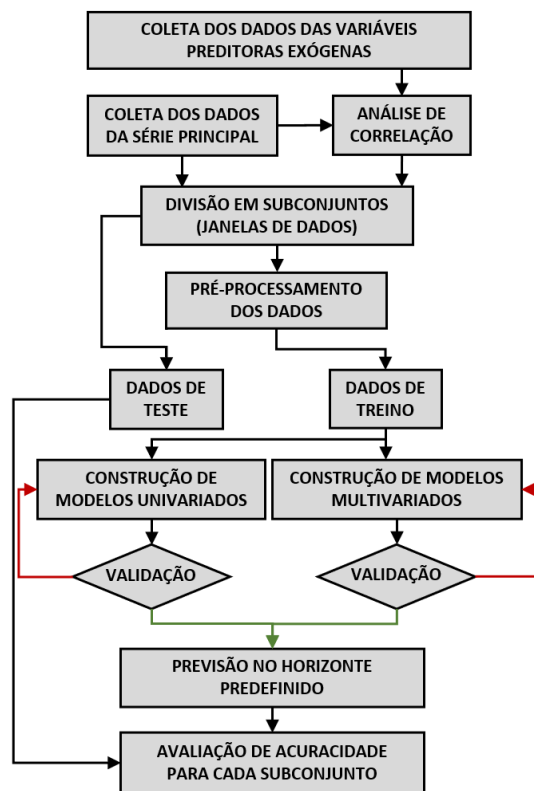


Fig. 1. Fluxograma do procedimento metodológico.

2.1 Descrição dos dados da séries temporais e análise de correlação com as variáveis exógenas

O conjunto de dados para a formação da série de preços de minério de ferro em dólares americanos por tonelada pode ser obtido de diversas fontes, gratuitamente. Para esta pesquisa, considera-se a série mensal de preços médios de contratos futuros de minério de ferro refinado com teor de 62%, em relação ao período compreendido entre julho de 2015 e janeiro de 2022. A Fig. 2 representa a série temporal em destaque.

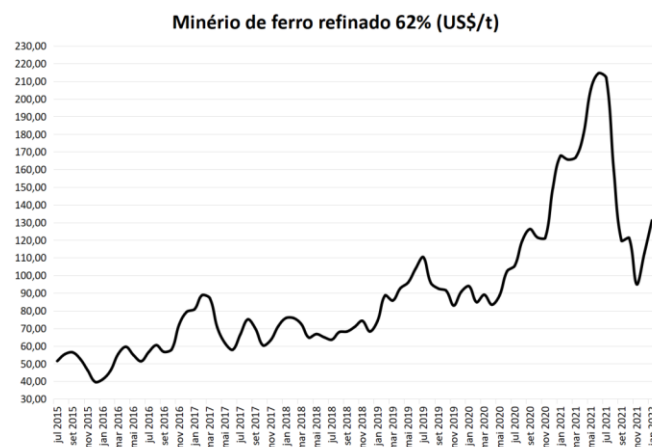


Fig. 2. Séries de preços de minério de ferro, em US\$/t.

Além do conjunto de dados da série principal, considera-se ainda os dados referentes às variáveis exógenas, que são

aplicadas aos modelos multivariados. Tomando como base os principais *datasets* sugeridos por Ewees et al. (2020), adota-se como possíveis preditoras as explicitadas na Tabela 1. Todos os conjuntos de dados referentes às variáveis exógenas são coletados também de fontes gratuitas, disponibilizadas online.

Tabela 1. Conjunto de dados avaliados.

Variáveis	Unidade	Fonte de dados
Preço do minério de ferro refinado 62% Fe	US\$/t	<i>Investing.com</i>
Preço do petróleo	US\$/barril	<i>U.S. Energy Information Administration</i>
Preço do ouro	US\$/onça troy	<i>Investing.com</i>
Preço da prata	US\$/onça troy	<i>Investing.com</i>
Preço do cobre	US\$/onça troy	<i>Investing.com</i>
Preço da sucata de ferro	US\$/onça troy	<i>Organization for Economic Co-operation and Development</i>
Produção de aço bruto na China	1000t	<i>Worldsteel Association</i>
Quant. de minério de ferro importado pela China	1000t	<i>Trade statistics for international business development</i>
Taxa de inflação na China (IPC)	%	<i>Organization for Economic Co-operation and Development</i>
Excedente produtivo de minério de ferro da Vale	Mt	<i>Vale.com</i>

Neste ponto, vale destacar que uma variável adicional é analisada: o excedente produtivo de minério de ferro da Vale S.A, assumindo a hipótese de que o excedente produtivo de minério de ferro (quantidade de minério produzida menos a quantidade vendida) possa ter alguma correlação com os preços praticados.

Os dados desta série podem ser coletados por meio dos relatórios de produção da Vale S.A., disponibilizados trimestralmente.

Para investigar a relação entre os preços de minério de ferro e as variáveis exógenas, busca-se estabelecer uma correlação que se preocupa em determinar, essencialmente, a relação entre cada grupo de duas variáveis. Para tal, utiliza-se os coeficientes de correlação de Pearson (para análise linear) e Spearman (para relações monótonas, lineares ou não), que assumem valores entre -1 e 1, tal que: Se igual a 1, existe uma correlação perfeita positiva entre as duas variáveis; se igual a -1, existe uma correlação perfeita negativa entre as duas variáveis; se igual a 0, as duas variáveis não dependem uma da outra.

É válido destacar que a elaboração dos algoritmos pertinentes à análise correlacional e às demais etapas do projeto é feita com o auxílio de pacotes e *frameworks* dedicados à inteligência computacional e à previsão de séries temporais, disponíveis para as linguagens *Python* e *R*. No caso da análise de correlação, utilizou-se a biblioteca “*stats*”, do *R*.

2.2 Séries temporais nebulosas

As séries temporais nebulosas, ou *fuzzy time series* (FTS), são métodos de previsão inicialmente propostos por Song and Chisson (1993), tomando como base a teoria dos conjuntos nebulosos (*fuzzy sets*). Inicialmente, os primeiros trabalhos conduzidos por esta técnica elaboraram o conceito de modelagem de primeira ordem, em que a estimativa pontual da série é determinada por meio de um único preditor. A ordem indica, neste caso, o número de atrasos de tempo, ou *lags*, usados no modelo (Chen, 1996). A partir do modelo de primeira ordem, Hwang et al. (1998) implementou a técnica de modelagem de alta ordem, isto é, utilizando mais de um *lag* como preditor. Desde então, os métodos FTS têm sido utilizados na previsão de variáveis em inúmeras áreas do conhecimento, com destaque para as previsões relacionadas ao mercado financeiro (Cheng and Chen, 2018).

Para este estudo, utiliza-se a biblioteca de código aberto *pyFTS*, desenvolvida pelo laboratório *MINDS Lab*, da Universidade Federal de Minas Gerais, e disponível para a linguagem *Python*, para aplicar o conceito de séries temporais nebulosas à previsão de preços de minério de ferro.

O treinamento de um modelo FTS na biblioteca *pyFTS* segue o conceito proposto por Song and Chisson (1993), tendo os seguintes passos:

- Definição e particionamento do universo de discurso em intervalos sobrepostos;
- Definição dos conjuntos nebulosos a partir dos dados da série temporal (fuzzificação);
- Criação dos padrões temporais (*fuzzy logical relationships*);
- Agrupamento dos padrões e formação das regras, que constituem o modelo FTS propriamente dito.

Partindo do modelo criado, segue-se à etapa de inferência, em que se realiza efetivamente a previsão, seguindo os passos:

- Fuzzificação do valor de entrada que se deseja prever;
- Busca por regras compatíveis no conjunto de padrões temporais organizados na etapa de treinamento;
- Defuzzificação dos dados, que transforma o resultado previsto em valor numérico.

Em relação aos métodos FTS, considera-se nesta pesquisa os modelos de múltipla ordem, isto é, em que dois ou mais *lags* são usados para a construção das regras. Os métodos avaliados são do tipo PWFTS (*Probabilistic Weighted Fuzzy Time Series*, isto é, com pesos probabilísticos no precedente e no consequente das regras), para a modelagem univariável; e do tipo FDT (*Fuzzy Decision Trees*), método que combina

lógica nebulosa às árvores de decisão, construídas pelos algoritmos ID3 (*Iterative Dichotomiser 3*), CART (*Classification and Regression Tree*) e RF (*Random Forest*). O método FDT é aplicado neste trabalho para a estratégia de modelagem multivariável, e está descrito em Silva et al. (2021).

2.3 Construção dos modelos preditivos univariados e multivariados

Inicialmente, a série de preços de minério de ferro é tratada de forma univariável, isto é, são utilizados os próprios *lags* da série para construção dos modelos. Em um segundo momento, é feita a modelagem multivariável, considerando a inclusão das variáveis exógenas que possuem correlação significativa com a série principal, conforme o procedimento descrito na seção 2.1. Em ambos os casos, os dados são divididos em 23 subconjuntos de janelas deslizantes (Fig. 3), que avançam mês a mês, e cada janela se divide em dois grupos: treino (90%) e teste (10%), sendo os dados de teste correspondentes a uma previsão trimestral, isto é, do tipo *short-term*.

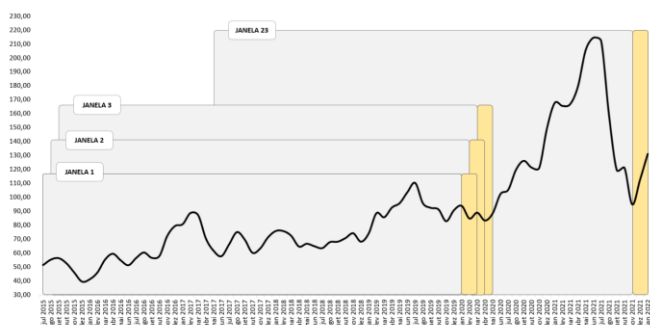


Fig. 3. Divisão dos dados em 23 subconjuntos de treino e teste.

Neste estudo, considera-se a técnica estatística ARIMA (*autoregressive integrated moving average*), em comparação às técnicas de inteligência artificial: baseadas em redes neurais artificiais, pela arquitetura multicamadas (*multilayer perceptron*), ou MLP; coletâneas de árvores de regressão, pelo algoritmo XGboost (*extreme gradient boosting*); e lógica nebulosa, por meio das *fuzzy time series*, ou FTS. A Tabela 2 explicita, sumariamente, os modelos construídos para a previsão da série de preços de minério de ferro, indicando as bibliotecas específicas utilizadas e os principais parâmetros configurados manualmente em cada caso.

2.4 Métricas de desempenho e avaliação de acurácia

Após a etapa de previsão, realiza-se um estudo comparativo dos modelos para identificar os de maior acuracidade entre os avaliados. Na literatura, diversos indicadores são encontrados. Para esta pesquisa, considera-se dois destes, o *Root Mean Square Error* (RMSE) e o *Mean Absolute Percentage Error* (MAPE), sendo o RMSE um parâmetro dependente de escala, e o MAPE um parâmetro percentual,

portanto independente de escala. Em ambos os casos, quanto menor, melhor.

Tabela 2. Modelos de inteligência computacional.

Método	Bibliotecas	Parâmetros	Sintaxe
FTS	pyFTS – Python	<i>npart</i> = n° de partições do universo de discurso; <i>order</i> = n° de lags significativos	PWFTS <i>npart</i> ; FDT_ID3; FDT_CART; FDT_RF;
MLP	nnfor, neuralnet – R	<i>hd</i> = n° de neurônios em cada camada oculta; <i>lags</i> = atrasos significativos; <i>eta</i> = taxa de aprendizado	MLP <i>hd</i>
Xgboost	xgboost, caret – R	<i>max_depth</i> = profundidade máxima de uma árvore; <i>eta</i> = taxa de aprendizado	XGBOOST

Os melhores modelos são definidos com base no desempenho das métricas em todas as janelas de previsão. Para tanto, são avaliadas as estatísticas descritivas do conjunto de RMSE e MAPE, em que são verificadas a posição central (mediana), a dispersão do erro (intervalo interquartil, ou IQR), e a presença de valores discrepantes (*outliers*).

3. RESULTADOS E DISCUSSÕES

3.1 Análise correlacional e escolha das variáveis predictoras

Tabela 3. Correlação das variáveis exógenas em relação à série de preços de minério de ferro.

Variável	Coef. Pearson	Coef. Spearman
Preço do petróleo	0,09	0,17
Preço do ouro	0,87	0,75
Preço da prata	0,77	0,55
Preço do cobre	0,51	0,46
Preço da sucata de ferro	0,34	0,33
Produção de aço bruto na China	0,76	0,87
Quant. de minério de ferro importado pela China	0,75	0,84
Taxa de inflação na China (IPC)	0,52	0,55
Excedente produtivo da Vale	-0,51	-0,50

Os coeficientes correlacionais de Pearson e Spearman para as variáveis exógenas, em relação à série temporal de preços de minério de ferro, são vistos na Tabela 3. Tomando como critério a necessidade de ambos os coeficientes serem maiores ou iguais a $|0,50|$, as variáveis descartadas dos modelos multivariados são: preço do petróleo, preço do cobre e preço da sucata de ferro.

3.2 Modelos univariados

Todos os modelos passaram pela fase de pré-processamento, com uma etapa de diferenciação e normalização na fase de treino.

Os modelos nebulosos foram construídos a partir de três variações do método PWFTS, com 10, 15 e 20 partições. Todos os modelos são de terceira ordem, uma vez que os de primeira e segunda ordem não apresentaram resultados satisfatórios. Os dados foram fuzzificados considerando uma função de pertinência do tipo triangular, com conjuntos nebulosos de tamanhos iguais.

Nos modelos MLP, considerou-se a variação do número de neurônios da camada oculta em 5, 10 e 20, tomando a taxa de aprendizado como 0,4. Os lags 1, 2, 3 e 4 foram utilizados no treinamento, e cada modelo MLP passou por 10 rodadas de treino e teste para cada janela de dados, sendo o resultado final formado pela média aritmética das 10 rodadas.

Acerca do modelo *Xgboost*, os parâmetros configurados manualmente foram “*max_depth*” = 10, e “*eta*” = 0,4. O algoritmo possui uma ferramenta de otimização interna para a escolha dos demais parâmetros. Similar ao modelo MLP, a série teve 10 rodadas de treino e teste para cada janela de dados.

Em relação ao modelo ARIMA, os coeficientes foram escolhidos por otimização interna da função “*auto.arima()*”. Os resultados da tabela são referentes aos modelos ARIMA do tipo (0,1,1).

Tabela 4. Estatística descritiva RMSE dos modelos univariados

Modelo	Mínimo	Mediana	Máximo	IQR
ARIMA	3,63	21,08	88,79	18,30
MLP5	4,69	21,35	107,43	19,57
MLP10	9,12	21,70	101,80	15,58
MLP20	8,01	19,30	85,90	13,66
XGBOOST	3,69	33,56	111,93	60,64
PWFTS10	2,69	16,98	59,52	15,72
PWFTS15	2,41	16,92	58,21	11,62
PWFTS20	2,72	16,83	58,21	13,82

A Tabela 4 e a Tabela 5 exibem as estatísticas descritivas das métricas RMSE e MAPE, respectivamente, para os 23 subconjuntos de dados de teste, que também podem ser visualizadas nos *boxplots* da Fig. 4. Nota-se baixo poder de generalização do modelo *Xgboost*, que apresentou as maiores medianas e grande dispersão tanto no RMSE quanto no

MAPE. Os modelos ARIMA e MLP obtiveram melhores resultados do que o *Xgboost*, porém evidencia-se a presença de fortes *outliers* em ambos os casos. Os modelos FTS, por outro lado, se destacam por apresentarem as menores medianas, com melhores mínimos e máximos, e baixa variabilidade, pelo IQR reduzido. A Fig. 5 ilustra a comparação entre os valores reais da série e os valores preditos por cada modelo univariado.

Tabela 5. Estatística descritiva MAPE dos modelos univariados

Modelo	Mínimo	Mediana	Máximo	IQR
ARIMA	3,11	14,66	67,73	9,89
MLP5	3,31	15,19	81,94	10,40
MLP10	1,90	14,45	77,71	7,92
MLP20	0,87	14,84	65,58	8,15
XGBOOST	3,71	21,58	48,26	20,99
PWFTS10	2,10	10,91	26,15	10,78
PWFTS15	1,29	12,60	28,18	8,22
PWFTS20	1,37	10,69	25,16	10,12

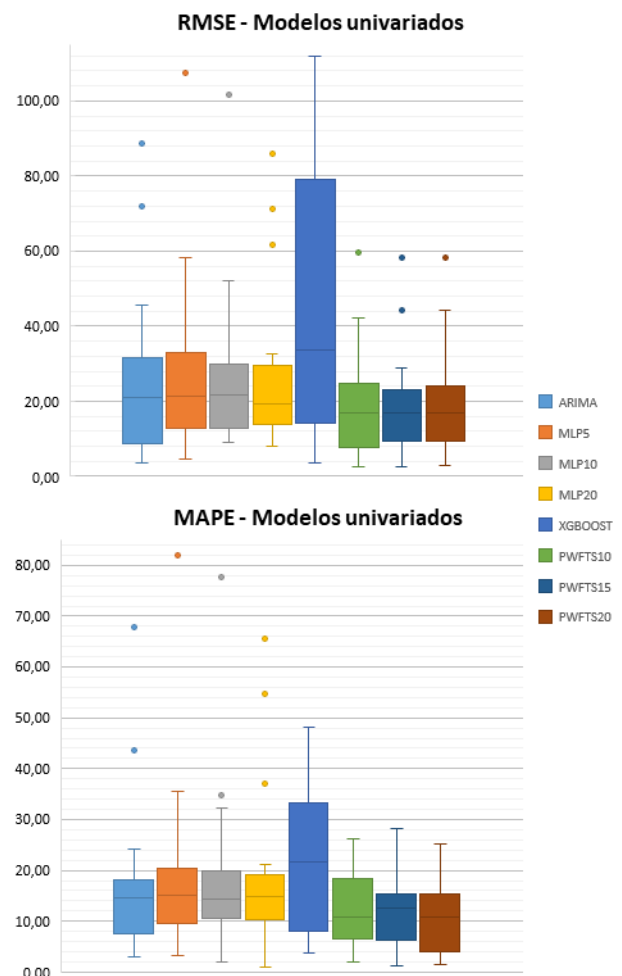


Fig. 4. *Boxplots* para os subconjuntos de RMSE e MAPE das séries univariadas.

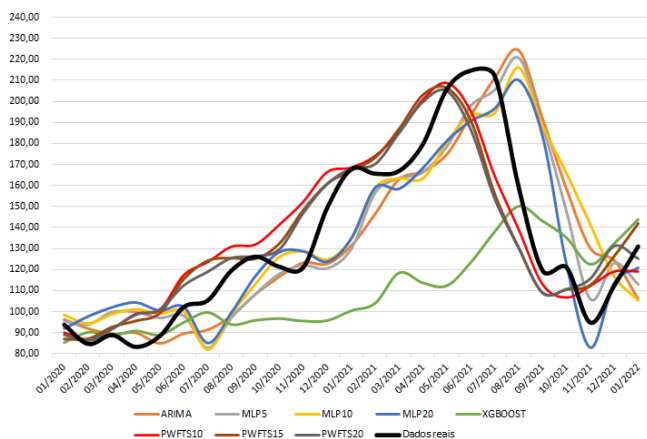


Fig. 5. Comparação entre o valor real e o valor previsto para a série de teste univariada.

3.3 Modelos multivariados

Para os modelos multivariados, a etapa de pré-processamento inclui diferenciação de primeira ordem e normalização em todo o conjunto de dados de treino.

A Tabela 6 e a Tabela 7 exibem as estatísticas descritivas das métricas RMSE e MAPE, respectivamente, para os 23 subconjuntos de dados de teste da série multivariada, que também podem ser visualizadas nos *boxplots* da Fig. 6.

Tabela 6. Estatística descritiva RMSE dos modelos multivariados

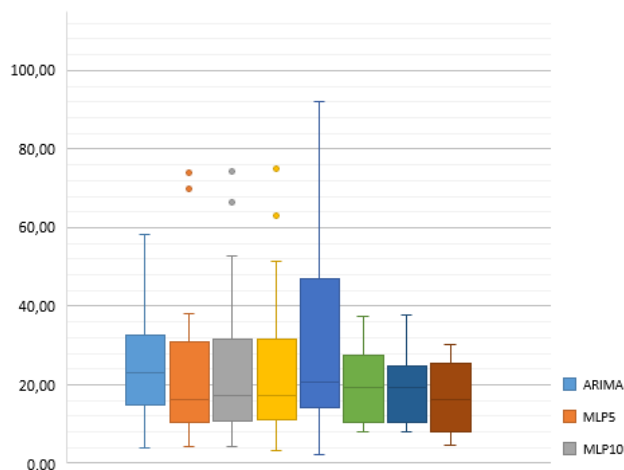
Modelo	Mínimo	Mediana	Máximo	IQR
ARIMA	3,76	23,04	58,15	16,53
MLP5	4,24	16,17	73,97	17,78
MLP10	4,19	17,34	74,34	18,80
MLP20	3,26	17,29	74,99	17,62
XGBOOST	2,13	20,82	92,06	31,01
FDTFTS_ID3	7,96	19,15	37,30	15,51
FDTFTS_CART	7,96	19,20	37,73	13,28
FDTFTS_RF	4,69	16,17	30,30	16,07

Tabela 7. Estatística descritiva MAPE dos modelos multivariados

Modelo	Mínimo	Mediana	Máximo	IQR
ARIMA	2,83	14,09	28,29	14,32
MLP5	3,88	12,00	64,15	9,92
MLP10	3,83	12,98	64,43	11,71
MLP20	2,38	12,67	64,94	7,95
XGBOOST	2,08	18,14	45,10	18,74
FDTFTS_ID3	3,71	11,94	30,25	10,63
FDTFTS_CART	3,71	12,06	30,47	8,73
FDTFTS_RF	1,97	10,50	22,17	9,04

Em comparação às séries univariadas, o modelo ARIMA multivariado apresentou maior mediana na métrica RMSE, porém com menor dispersão ao redor, e ausência de *outliers*. Os modelos MLP, todavia, apresentaram comportamento inverso, reduzindo a mediana e o valor dos *outliers*, e aumentando a dispersão. Nota-se clara melhora nas métricas do *Xgboost*, mas ainda insuficiente para competir com os demais modelos. Neste sentido, destaca-se novamente os modelos nebulosos pela ausência de valores discrepantes, especialmente o FDTFTS_CART e o FDTFTS_RF, que apresentaram ainda a menor dispersão e a menor mediana, respectivamente, em ambas as métricas.

RMSE - Modelos multivariados



MAPE - Modelos multivariados

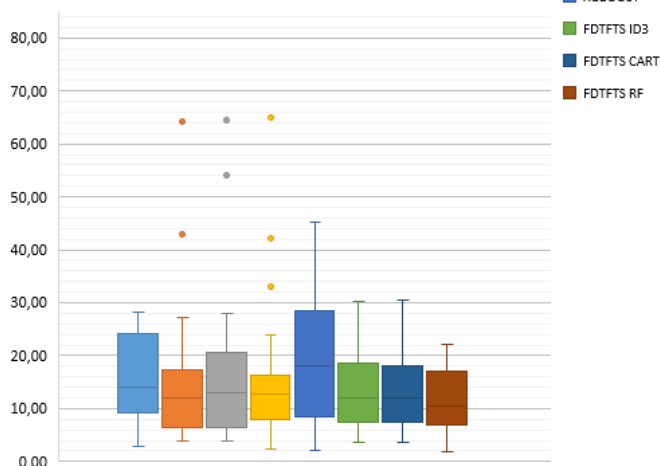


Fig. 6. *Boxplots* para os subconjuntos de RMSE e MAPE das séries multivariadas.

A comparação entre os valores reais da série e os valores preditos por cada modelo para a série multivariada pode ser vista na Fig. 7.

A Fig. 8 ilustra um recurso presente nos modelos FDT, que permite visualizar o grau de importância de cada variável exógena na construção das árvores de previsão. Nos três algoritmos, o excedente produtivo de minério de ferro da Vale S.A. aparece como a variável exógena de maior

significância, atestando a importância desta variável como preditora da série *target*.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), do Instituto Tecnológico Vale (ITV) e da Universidade Federal de Ouro Preto (UFOP).

REFERÊNCIAS

- Aminrostamkolae, B., Scroggs, J. S., Borghei, M. S., Safdari-Vaighani, A., Mohammadi, T. and Pourkazemi, M. H. (2017). Valuation of a hypothetical mining project under commodity price and exchange rate uncertainties by using numerical methods. *Resources Policy*, 52, 296-307.
- Beasley, J. E. *OR-notes*. Available at: <http://people.brunel.ac.uk/~mastjjb/jeb/or/forecast.html>. Accessed February, 21.
- Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy sets and systems*, 81(3), 311-319.
- Cheng, C. H. and Chen, C. H. (2018). Fuzzy time series model based on weighted association rule for financial market forecasting. *Expert Systems*, 35(4), e12271.
- Comexstat. *Ministério da Indústria, Comércio Exterior e Serviços. Exportação e Importação geral. Base de dados 2019-2020*. Available at: <http://comexstat.mdic.gov.br/pt/home>. Accessed December, 20.
- Egrioglu, E., Bas, E., Yolcu, U. and Chen, M. Y. (2020). Picture fuzzy time series: Defining, modeling and creating a new forecasting method. *Engineering Applications of Artificial Intelligence*, 88, 103367.
- Ewees, A. A., Abd Elaziz, M., Alameer, Z., Ye, H. and Jianhua, Z. (2020). Improving multilayer perceptron neural network using chaotic grasshopper optimization algorithm to forecast iron ore price volatility. *Resources Policy*, 65, 101555.
- Hwang, J. R., Chen, S. M. and Lee, C. H. (1998). Handling forecasting problems using fuzzy time series. *Fuzzy sets and systems*, 100(1-3), 217-228.
- Li, D., Moghaddam, M. R., Monjezi, M., Jahed Armaghani, D. and Mehrdaneh, A. (2020). Development of a group method of data handling technique to forecast iron ore price. *Applied Sciences*, 10(7), 2364.
- Ma, Y. and Zhen, W. (2020). Market fundamentals and iron ore spot prices. *Economic Record*, 96(315), 470-489.
- Silva, R. R., Caminhas, W. M. and Guimarães, F. G. (2021, July). A C4.5 Fuzzy Decision Tree Method for Multivariate Time Series Forecasting. In *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-6). IEEE.
- Song, Q. and Chissom, B. S. (1993). Fuzzy time series and its models. *Fuzzy sets and systems*, 54(3), 269-277.
- Trademap. *Trade statistics for international business development*. Available at: <https://www.trademap.org>. Accessed December 20.

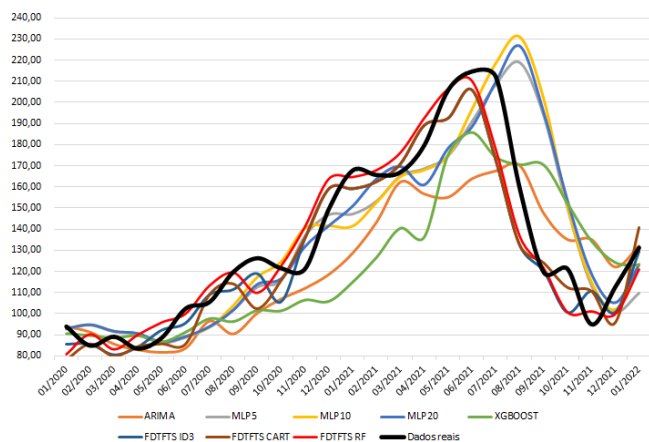


Fig. 7. Comparação entre o valor real e o valor previsto para a série de teste multivariada.

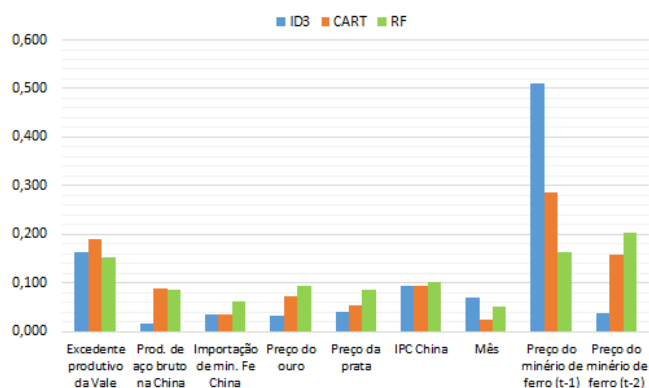


Fig. 8. Grau de importância de cada variável preditora nos modelos FDT.

6. CONCLUSÕES

A partir da realização deste estudo comparativo, foi possível observar a superioridade dos modelos baseados em lógica nebulosa na previsão de preços de minério de ferro, com destaque para os modelos PWFTS univariados de terceira ordem com 15 e 20 partições, e os modelos FDT multivariados baseados nos algoritmos CART e *Random Forest*.

Ao comparar a acuracidade dos modelos univariados e multivariados, conclui-se ainda que a inclusão de variáveis exógenas à série *target* não implica necessariamente em maior acuracidade. Contudo, a maior parte dos modelos testados aumentou significativamente o potencial de generalização após a introdução do conjunto de séries correlacionadas. Destaca-se, neste ponto, o elevado grau de significância do excedente produtivo de minério de ferro da Vale S.A. na formação dos modelos FDT.