

# Explorando métodos de explicabilidade para LSTM e GRU na previsão de série temporal do índice BOVESPA.

1<sup>st</sup> Lucas Rabelo  
*CIN; UFPE; PE, Brasil*

2<sup>nd</sup> Eliaquim Moreira  
*CIN; UFPE; PE, Brasil*

3<sup>rd</sup> Cleber Zanchettin  
*CIN; UFPE; PE, Brasil*

4<sup>rd</sup> Adriano Lorena  
*CIN; UFPE; PE, Brasil*

## Resumo

O objetivo do trabalho foi prever o índice Bovespa utilizando modelos LSTM e GRU, além de técnicas de interpretabilidade. Os modelos foram treinados com lags dos 6 e 15 dias anteriores, considerando a capacidade dessas redes neurais em lidar com dependências temporais complexas. A metodologia incluiu a análise de importância de features, LIME (Local Interpretable Model-agnostic Explanations) e Shapley Values para garantir interpretabilidade. Foram comparados com um baseline naïve e modelos ARIMA, incluindo uma abordagem híbrida. Os resultados mostraram que o modelo LSTM superou outros em métricas como RMSE, MAPE e MASE. A análise de importância de features destacou o lag 1 como crítico na previsão. As previsões de curto prazo (3 e 7 dias) foram mais precisas, especialmente com o modelo LSTM, enquanto as de 15 dias foram desafiadoras para todos os modelos. O estudo enfatiza a importância da escolha adequada de lags na construção de modelos de previsão e destaca a necessidade de considerar diferentes janelas de tempo para avaliação de desempenho.

## Index Terms

XAI, forecasting, Time-Series, LSTM, GRU

## I. INTRODUÇÃO

A série temporal do índice bovespa, que representa a evolução dos valores de ações no mercado financeiro brasileiro, mostra um aumento significativo a partir de 2014. A partir de 2015, a tendência é predominantemente positiva, com um crescimento acentuado em 2016 e 2017. Em 2018, a tendência começa a descer, mas ainda permanece acima do ponto de partida. A partir de 2019, a tendência é mais estabilizada, com flutuações menores. Em 2020, a pandemia da COVID-19 afeta o mercado financeiro globalmente, mas o índice bovespa consegue se recuperar e continuar a tendência positiva. Em 2021, a tendência é predominantemente positiva, com um crescimento acentuado em relação aos anos anteriores.

Essa análise sugere que o mercado financeiro brasileiro tem apresentado um crescimento nos últimos anos, com diversas flutuações, e alguns desequilíbrios. Portanto é importante considerar que a série temporal pode ser afetada por diversos fatores, como políticas governamentais, condições econômicas e eventos globais, e que a análise deve ser feita com cautela e considerando outros indicadores e contextos.

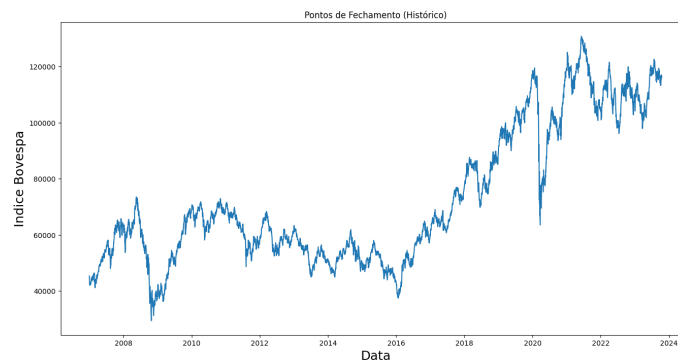


Figura 1. Série Bovespa

Neste cenário desafiador, o objetivo deste trabalho é utilizar de modelos LSTM e GRU na previsão do índice Bovespa, devido à capacidade dessas redes neurais em lidar com a variabilidade e dependências temporais complexas presentes nos dados. No entanto, também almejamos garantir interpretabilidade dos resultados do modelo com melhor capacidade de extrapolação da previsão, onde serão aplicadas técnicas de modelos agnósticos como importância de features, análise de erros do modelo, técnicas como o LIME e Valores Shapley. Além disso, a comparação com um baseline naïve ajuda a avaliar o valor agregado desses modelos em relação a abordagens mais simples, tornando a análise de séries temporais financeiras mais robusta e esclarecedora.

## II. METODOLOGIA

Os "lags" em uma série temporal se referem a observações passadas da série que são usadas como entradas para um modelo de previsão. Cada "lag" representa o valor da série temporal em um momento anterior, e a inclusão de vários "lags" permite ao modelo capturar dependências temporais e padrões históricos nos dados. Esses "lags" são usados como recursos de entrada para modelos LSTM e GRU, permitindo que esses modelos capturem as relações temporais entre os valores passados e o valor atual. Eles desempenham um papel importante na capacidade do modelo de aprender padrões e tendências da série temporal, tornando-os úteis para previsões baseadas em histórico, transformá-los em variáveis ajuda a aplicar métodos de interpretabilidade. Na tabela abaixo estão os 6 lags dos 7 dias iniciais da série do Bovespa:

Data	Close	lag1	lag2	lag3	lag4	lag5	lag6
2007-01-02	45383.0						
2007-01-03	44445.0	45383.0					
2007-01-04	44020.0	44445.0	45383.0				
2007-01-05	42245.0	44020.0	44445.0	45383.0			
2007-01-08	42830.0	42245.0	44020.0	44445.0	45383.0		
2007-01-09	42007.0	42830.0	42245.0	44020.0	44445.0	45383.0	
2007-01-10	42336.0	42007.0	42830.0	42245.0	44020.0	44445.0	45383.0

Tabela I

TABELA COM O LAGS DA SÉRIE DO BOVESPA

Os modelos de redes neurais (treinados com 300 épocas) são usados para prever valores futuros com base em uma janela de observações passadas, foi utilizada uma seed de número 22 com a função do tensorflow `randomset_seed` para garantir a reproducibilidade dos resultados, e a divisão de dados entre treinamento e validação é realizada com 80% dos dados iniciais para treinamento e 20% para validação (para que seja preservada a ordem temporal das previsões). Essa divisão é uma prática comum para avaliar o desempenho do modelo em dados não vistos, além disso, para verificar a capacidade de generalização dos modelos, as previsões foram testadas em diferentes janelas de tempo de 3, 7 e 15 dias no futuro. A normalização MinMax (para os modelos de redes neurais) foi ajustada somente nos dados de treinamento, para depois ser aplicado no conjunto de validação, garantindo que o modelo não capture informação *a posteriori* para previsão. Detalhando a arquitetura de cada modelo (na função `look_back` corresponde a quantidade de lags que são utilizadas para prever a janela futura posterior) teremos:

### - GRU:

O modelo começa com uma camada GRU com 128 unidades (neurônios). A camada GRU é uma variante das RNN que é capaz de capturar dependências temporais em dados sequenciais. O modelo é compilado usando o otimizador 'adam' e a função de perda do erro médio quadrático, que é comum para problemas de regressão.

### - LSTM:

O modelo LSTM é semelhante ao modelo GRU em termos de arquitetura, com a principal diferença sendo o uso da camada LSTM em vez da camada GRU. Assim como o modelo GRU, ele começa com uma camada LSTM com 128 unidades, que captura as dependências temporais nos dados. A entrada e saída do modelo são manipuladas da mesma forma que no modelo GRU.

### - Bidirectional LSTM:

O modelo BiLSTM é uma variação do modelo LSTM, compartilhando semelhanças em sua arquitetura, mas com uma distinção fundamental. Ambos começam com uma camada LSTM, neste caso com 128 unidades, para capturar dependências temporais em dados sequenciais. A principal diferença reside na capacidade bidirecional da camada LSTM, que processa informações em ambas as direções temporais simultaneamente. Isso permite uma compreensão mais abrangente do contexto sequencial. Assim como o modelo LSTM, o BiLSTM manipula entrada e saída de maneira similar, sendo adequado para tarefas que requerem consideração de padrões em ambas as direções temporais, como previsão de séries temporais.

### - Arima:

A função `auto_arima` do Python, proveniente da biblioteca `pmdarima`, é uma ferramenta eficaz para seleção automática do melhor modelo ARIMA para uma série temporal específica. Ela opera de forma automática, percorrendo uma combinação de hiperparâmetros do modelo, como  $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ , e  $Q$ , e utiliza critérios estatísticos como o AIC (Akaike Information Criterion) ou BIC (Bayesian Information Criterion) para avaliar o ajuste de cada combinação. O modelo ARIMA com o menor valor desses critérios é escolhido como o melhor ajuste, indicando que é a escolha mais parcimoniosa e eficaz para representar a estrutura temporal dos dados.

### A. Modelos Híbridos

Foi utilizado também um modelo híbrido, proposto por [2] como uma solução para a possibilidade de que a aproximação de modelos ARIMA a problemas não lineares complexos pode não ser adequada e o uso de redes neurais, quando empregado para modelar problemas lineares, pode produzir resultados mistos. Em geral, é difícil conhecer completamente as características dos dados em um problema real, a metodologia híbrida proposta por Zhang possui capacidades de modelagem linear e não linear, podendo ser uma boa estratégia para uso prático, pois a combinação de diferentes modelos pode capturar diferentes aspectos.

Em resumo, a metodologia do sistema híbrido proposto consiste em dois passos. No primeiro passo, um modelo ARIMA é usado para analisar a parte linear do problema. No segundo passo, um modelo de rede neural é desenvolvido para modelar os resíduos do modelo ARIMA, que seriam a parte não linear.

Uma vez que o modelo ARIMA não consegue capturar a estrutura não linear dos dados, os resíduos do modelo linear conterão informações sobre a não linearidade. Os resultados da rede neural podem ser usados como previsões de resíduos para o modelo ARIMA. O modelo híbrido utiliza a capacidade única e a força do Modelo ARIMA e do modelo de rede neural para determinar diferentes padrões. Portanto, de acordo com Zhang, pode ser vantajoso modelar padrões lineares e não lineares separadamente usando modelos diferentes e, em seguida, combinar as previsões para melhorar o desempenho geral de modelagem e previsão.

Basicamente, as previsões podem ser resumidas com a seguinte equação, onde  $\hat{L}_t$  é a parte linear e  $\hat{N}_t$  é a parte não linear da previsão:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (1)$$

### B. Métricas de Erro e Avaliação dos Modelos

Para as medidas de erro [3], foram utilizados dois tipos de erros, os erros dependentes da escala, dentre os quais temos neste trabalho o erro médio quadrático (com sigla RMSE, do inglês *Root Mean Squared Error*) e o erro médio absoluto em escala (com sigla MASE, do inglês *Mean Absolute Scaled Error*) e os erros percentuais, sendo usado o erro médio absoluto percentual (com sigla MAPE, do inglês *Mean Absolute Percentage Error*).

Para a seleção dos modelos, são escolhidos aqueles que apresentarem os menores RMSE (**Equação (2)**) e MASE (com base na **Equação (3)** e **Equação (4)**) e MAPE (**Equação (5)**).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{n}}, i = 1, \dots, N \quad (2)$$

$$MAE = \frac{\sum_{i=1}^N |x_i - \hat{x}_i|}{n} \quad (3)$$

$$MASE = \frac{MAE_{\text{model}}}{MAE_{\text{baseline}}} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|y_i - \hat{y}_i|}{|y_i|} \right) \times 100\% \quad (5)$$

A vantagem dos erros percentuais é que eles são livres de unidade e podem ser usados frequentemente para comparar diferentes conjuntos de dados. No entanto, o MAPE possui uma desvantagem por penalizar mais erros negativos, devido a essa complicação a medida foi avaliada conjuntamente com o RMSE e o MASE.

## III. RESULTADOS E DISCUSSÃO

A escolha dos lags é uma etapa importante na construção de modelos de correlação e previsão, e a escolha adequada dos lags pode significativamente melhorar a precisão das previsões. Após análise do gráfico de correlação de lags decidimos seguir com modelos que utilizem das características dos 6 e 15 dias anteriores, o autor do artigo base utilizou de 60 dias em sua slide window, no entanto apesar dos 60 lags apresentarem um valor significativo de autocorrelação com o dia base (**Figura 2**) foram utilizados os lags do 6 dias anteriores, comparadamente com os lags de 15 dias anteriores, utilizados com o intuito de capturar padrões temporais mais complexos. Além disso o gráfico de Fourier (**Figura 3**) mostra que, o mercado passa por flutuações de curto e de longo prazo com base na volatilidade de alta frequência, reforçando o comportamento caótico do mercado.

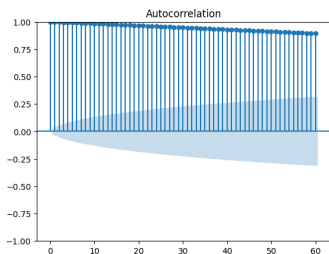


Figura 2. Autocorrelação dos Lags

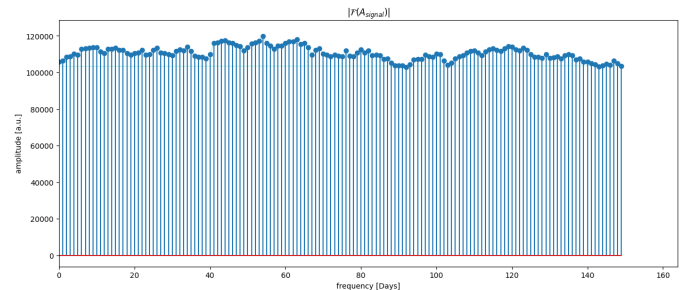


Figura 3. Decomposição de Fourier (Amplitudes)

O gráfico de decomposição STL (Seasonal and Trend decomposition using Loess) foi criado com o objetivo de proporcionar uma sensação de interpretabilidade ao modelo de previsão de séries temporais. A decomposição STL divide a série temporal em três componentes principais: tendência (trend), sazonalidade (seasonal), e resíduos (remainder). O período escolhido para a decomposição, foi 365, justificado pela necessidade de capturar padrões sazonais anuais na série temporal, como sazonalidades anuais nos dados da bolsa. Essa decomposição ajuda a identificar claramente a tendências de crescimento e estagnação do índice e os padrões irregulares na sazonalidade dos dados, o que, por sua vez, torna as previsões do modelo mais interpretáveis e úteis para entender as influências subjacentes na evolução da série temporal.

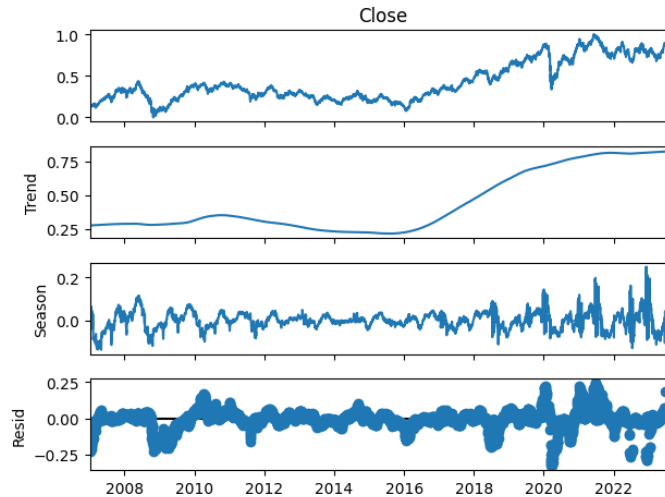


Figura 4. Decomposição STL Loess

Na construção de uma `regressionTree` para prever o índice Bovespa do próximo dia, identificamos que o lag 1 se destaca como a feature mais significativa, além disso tanto os lags 1 a 6 aparecem no rank das top 10 features importantes para previsão. Isso significa que o valor do índice Bovespa do dia anterior é um fator crítico para antecipar as variações futuras do mercado. A inclusão do lag 1 como a principal variável explicativa na árvore de regressão ressalta a importância da dependência temporal e do histórico recente do índice Bovespa na previsão de seu comportamento futuro. Portanto, ao considerar essa árvore de regressão, é evidente que o desempenho anterior desempenha um papel central na modelagem das tendências do mercado de ações, o fato de os lags recentes serem identificados como as características mais importantes sugere uma forte autocorrelação no mercado. Isso pode indicar que o mercado é influenciado por sua própria inércia a curto prazo, mais do que por fatores externos ou fundamentais.

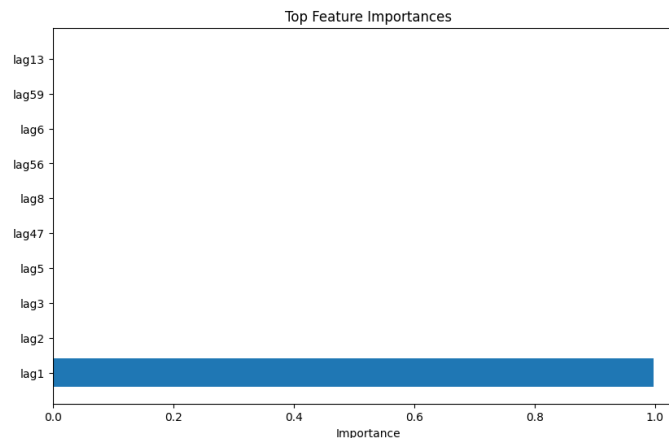


Figura 5. Feature Importance dos Lags para previsão

Nas figuras abaixo (Figuras 6, 7, 8 e 9), comparando as previsões é possível identificar que o modelo LSTM e GRU conseguem capturar bem a variação nos dados e realizar previsões bastante similares aos valores reais, a previsão é feita considerando um passo a diante, ou seja 60 dias anteriores foram utilizados para prever o próximo dia, devido a boa performance de previsão do modelo LSTM, cabe utilizar estratégias que possam melhorar a explicabilidade de suas previsões.

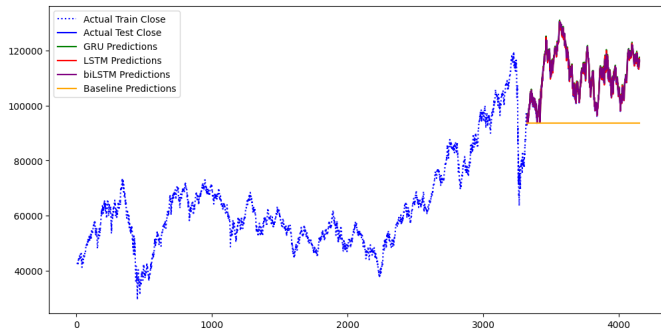


Figura 6. Previsões no conjunto de validação. (lag 6 dias)

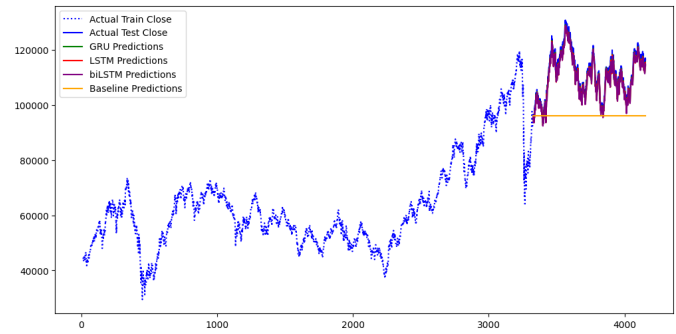


Figura 7. Previsões no conjunto de validação. (lag 15 dias)

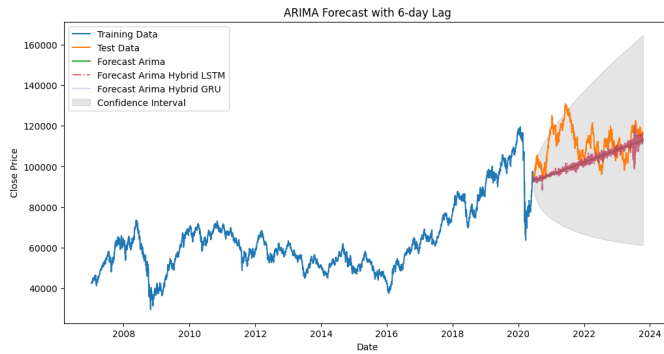


Figura 8. Previsões no conjunto de validação. (lag 6 dias)

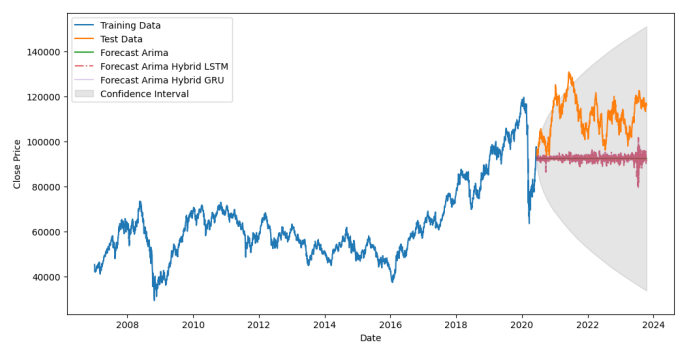


Figura 9. Previsões no conjunto de validação. (lag 15 dias)

Utilizando o lag dos 6 dias anteriores, a análise do MAPE revela que tanto o modelo GRU, o modelo LSTM e o BiLSTM apresentam alta precisão em suas previsões, com erros percentuais médios muito baixos, respectivamente de 1.03%, 1.11% e 1.07%. Isso indica que, em média, as previsões para os modelos de redes neurais estão bastante próximas dos valores reais. Por outro lado, o modelo baseline registra um MAPE de 15.43%, indicando previsões menos precisas. O modelo ARIMA apresentou desempenho superior ao baseline, mas ficou aquém dos modelos de redes neurais. Resultados semelhantes foram observados nos modelos híbridos, que tiveram desempenho inferior tanto ao ARIMA quanto às redes neurais. Ao utilizar o lag dos 15 dias anteriores como features, o modelo GRU alcançou a melhor performance no conjunto de validação. Em contrapartida, o modelo ARIMA e os modelos híbridos apresentaram desempenho inferior até mesmo ao modelo baseline, conforme indicado pelo MASE acima de 1.

Model	6 Dias Lag			15 Dias Lag		
	RMSE	MAPE (%)	MASE	RMSE	MAPE (%)	MASE
GRU	1437.32	1.03	0.06	2209.75	1.67	0.12
LSTM	1528.73	1.11	0.07	1918.77	1.42	0.10
BiLSTM	1473.23	1.07	0.07	1844.22	1.34	0.10
Baseline	19417.87	15.43	1.00	17114.87	13.16	1.00
ARIMA	13311.98	9.09	0.59	20434.37	16.43	1.24
ARIMA+LSTM	13390.75	9.16	0.60	20438.33	16.37	1.24
ARIMA+GRU	13401.97	9.17	0.60	20444.72	16.38	1.24

Tabela II

MÉTRICAS DE ERRO UTILIZANDO LAG DE 6 E 15 DIAS

Testando as previsões do modelo ajustado com base nos lags dos 6 dias anteriores, em diferentes janelas de tempo, para períodos de tempo de curto (3 dias) e médio prazo (7 dias) o modelo LSTM teve a melhor performance, tomando como base o MASE inferior ao dos outros modelos, para 3 dias, o MAPE indica que este modelo está fazendo previsões em média 1,38% distante dos valores reais, enquanto os outros modelos erram em média 1,44% ou mais. Para o período de 15 dias os modelos híbridos (que não tiveram boa performance nas previsões de médio prazo) tiveram uma melhor performance considerando o MASE e o MAPE, com o modelo ARIMA+GRU apresentando as melhores métricas, apesar disso o modelo LSTM ainda se mostrou consistente com boas métricas de erro.

Model	3 Dias			7 Dias			15 Dias		
	RMSE	MAPE (%)	MASE	RMSE	MAPE (%)	MASE	RMSE	MAPE (%)	MASE
GRU Model	2187.14	1.44	0.59	2563.87	1.89	0.74	2884.74	2.23	0.93
LSTM Model	2034.13	1.38	0.56	2433.68	1.83	0.71	2881.47	2.22	0.93
BiLSTM Model	2401.73	1.75	0.71	2801.50	2.08	0.81	2864.72	2.23	0.93
ARIMA	2693.29	2.35	0.95	2839.65	2.43	0.95	2784.06	2.34	0.97
ARIMA+LSTM	2833.77	2.26	0.93	3382.95	2.83	1.10	2910.48	2.22	0.92
ARIMA+GRU	2798.61	2.29	0.92	3375.84	2.82	1.10	2913.87	2.23	0.92

Tabela III  
MÉTRICAS DE ERRO UTILIZANDO LAG DE 6 DIAS

Testando as previsões do modelo ajustado com base nos lags dos 15 dias anteriores, todos os modelos tiveram performance piores do que o modelo baseline para todas as janelas de tempo, a maior quantidade de features pode estar capturando informação irrelevante ao modelo, piorando as previsões, ou piora pode estar sendo causada por alguns lags específicos. A análise do MASE destaca de maneira evidente a performance inferior dos modelos que incorporam componentes ARIMA (ARIMA, ARIMA+LSTM, ARIMA+GRU) em comparação com os modelos baseados em redes neurais, que também tiveram uma performance ruim em comparação ao baseline.

Model	3 Dias			7 Dias			15 Dias		
	RMSE	MAPE (%)	MASE	RMSE	MAPE (%)	MASE	RMSE	MAPE (%)	MASE
GRU	1710.47	1.47	1.42	1659.64	2.52	1.22	2084.70	1.55	1.39
LSTM	1999.22	1.65	1.27	1783.81	1.38	1.28	2323.58	1.70	1.51
BiLSTM	3867.83	3.10	2.67	3333.91	1.31	2.34	3947.74	3.02	2.69
ARIMA	2741.65	2.38	2.04	2907.73	2.49	2.29	2820.65	2.37	2.12
ARIMA+LSTM	2880.88	2.33	1.99	3417.01	2.85	2.62	2961.61	2.25	2.00
ARIMA+GRU	2861.71	2.31	1.97	3462.23	2.89	2.66	2965.13	2.25	2.01

Tabela IV  
MÉTRICAS DE ERRO UTILIZANDO LAG DE 15 DIAS

O Lag 1, com base nos valores de SHAP (Figura 10), têm o maior impacto negativo na previsão do índice do mercado de ações. Isso significa que um valor mais alto do índice hoje (lag1) está associado a um valor previsto mais baixo do índice amanhã. Isso sugere que o modelo pode estar capturando algum comportamento de reversão à média no mercado de ações, onde valores altos tendem a ser seguidos por valores mais baixos e vice-versa. Todavia, o lag1 é a feature que tem o maior impacto positivo no modelo, portanto também ajudando a capturar continuidade de altas ou baixas reforçando a importância dessa feature para compreensão dos resultados.

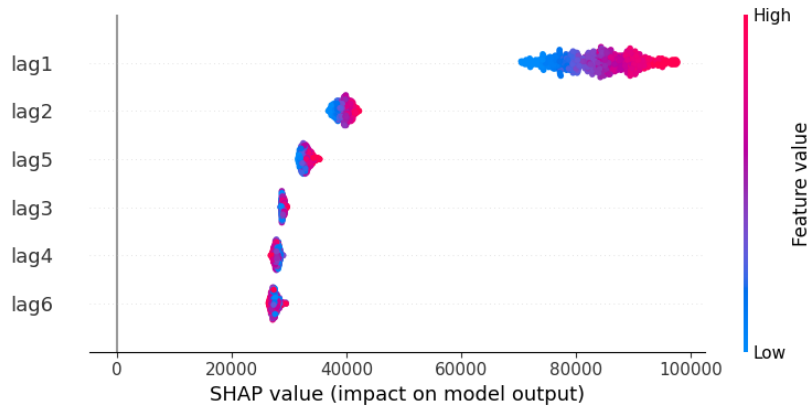


Figura 10. Beeswaarm Plot baseado nos valores SHAP (Impacto das Features)

Nas explicações locais utilizando os valores de LIME, foi selecionado um dia aleatório para avaliação mais detalhada do comportamento do modelo. Ao analisar os resultados do LIME para o dia escolhido aleatoriamente, é possível identificar como os dias anteriores afetaram o valor atual do índice bovespa, na Figura 11, para o dia específico o lag 1 (o dia anterior), o lag 2 e o lag 5 tiveram um impacto positivo na previsão do ibovespa, já os lags 3, 4 e 6 tiveram um impacto negativo, neste cenário seria importante realizar uma investigação nos dias 3, 4 e 6 dias anteriores para entender qual evento pode ter causado esse impacto negativo na previsão (por exemplo, uma IPO de uma empresa na bolsa, pode ajudar a compreender cenários em que uma queda no ibovespa nos 6 dias anteriores contribua para uma alta na previsão, mostrando uma recuperação devido a maior oferta de ações e negociações no mercado).

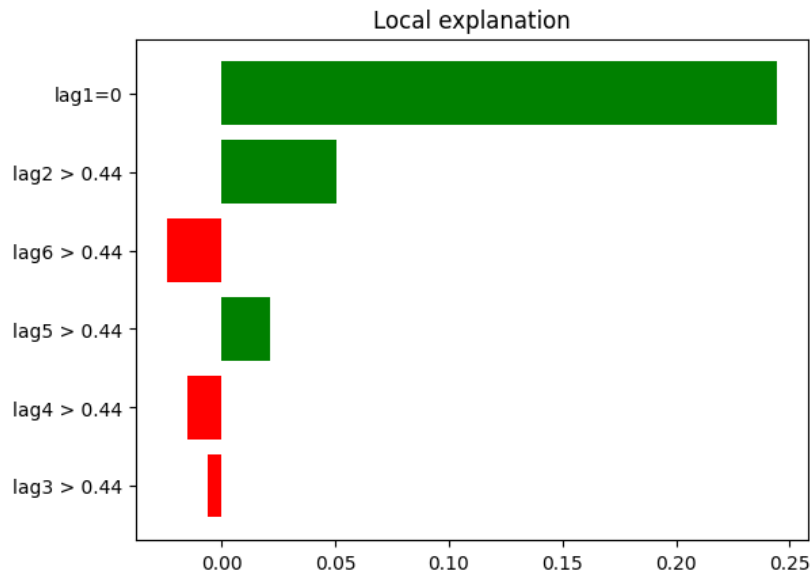


Figura 11. Explicações Locais do LIME

#### IV. CONCLUSÃO

Com base nas métricas de teste (em 3, 7 e 15 dias), é possível concluir que o modelo LSTM demonstra um desempenho superior em comparação com o modelo GRU, os modelos ARIMA e o modelo baseline. O menor valor de RMSE indica que as previsões do modelo LSTM estão mais próximas dos valores reais, enquanto o MAPE consideravelmente mais baixo demonstra que suas previsões têm um erro percentual menor. Além disso, o MASE abaixo de 1.0 destaca a capacidade do modelo LSTM de superar um modelo de previsão ingênuo (baseline). Em resumo, o LSTM, avaliando conjuntamente suas métricas, se destaca como o melhor modelo para previsões no conjunto de dados analisado, apresentou métricas consistentes no conjunto de validação (apesar de inferiores ao modelo GRU e BiLSTM), sua capacidade de extrapolação proporciona previsões mais precisas, confiáveis e escalonadas, para um período de 3 dias o modelo LSTM atinge sua melhor performance preditiva.

A ênfase na interpretabilidade de modelos LSTM para séries temporais financeiras é uma contribuição significativa. Tradicionalmente, esses modelos são considerados "caixas pretas" devido à sua complexidade. O estudo avança ao aplicar métodos agnósticos de modelo para elucidar as tomadas de decisão internas desses modelos em dados de séries temporais, trabalhos como este são pouco encontrados na literatura e em repositórios de código públicos, com esta contribuição o trabalho é um dos poucos a disponibilizar publicamente uma abordagem para o shap e o lime em séries temporais, que pode ser acessado no repositório do github neste link.

É importante notar que a interpretação local fornecida pelo LIME pode não ser totalmente representativa do comportamento global do modelo, mas é valiosa para entender as previsões em contextos específicos. Já os Shapley Values oferecem uma perspectiva diferente ao fornecer uma interpretação global das contribuições de cada lag para as previsões do modelo. É importante ter em mente que a interpretação do LIME e do SHAP para séries temporais são uma aproximação e podem não capturar totalmente a complexidade temporal. Ainda assim, ao destacar as variáveis relevantes em momentos específicos, o trabalho traz passos importantes para entender a dinâmica dos modelos em relação aos dados temporais. Para trabalhos futuros e extensão da pesquisa, dados de séries temporais multivariadas (contendo outras features, além dos lags), classificação de séries temporais (como imagens de ECG) e detecção de outliers são áreas que ainda precisam e podem se beneficiar da contribuição de métodos de explicabilidade para as previsões do modelo.

#### REFERÊNCIAS

- [1] Velarde, G.; Brañez, P.; Bueno, A.; Heredia, R.; Lopez-Ledezma, M. An Open Source and Reproducible Implementation of LSTM and GRU Networks for Time Series Forecasting. Eng. Proc. 2022, 18, 30. <https://doi.org/10.3390/engproc2022018030>
- [2] G.Peter Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, Volume 50, 2003, Pages 159-175, ISSN 0925-2312, [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0).
- [3] Hyndman, R.; Athanasopoulos, G. Forecasting: Principles and Practice. 3rd ed. Melbourne, Australia: OTexts, 2021. Acessado em: 19/05/2023. Disponível em: <https://otexts.com/fpp3/>.
- [4] Scott Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions, 2017, 1705.07874, arXiv. Acessado em: 17/12/2023. Disponível em: <https://arxiv.org/abs/1705.07874>.