

# Modelos de Previsão das Taxas de Juros dos Títulos IPCA do Tesouro Direto

Sergio Luiz Rodrigues de Oliveira Junior  
Departamento de Computação  
CEFET-MG  
Belo Horizonte, Brasil  
sergio.luijzr08@gmail.com

Rogério Gomes, Bruno Santos e Fábio Silva  
Departamento de Computação  
CEFET-MG  
Belo Horizonte, Brasil  
{rogerio, bsantos, fabiorochadasilva}@cefetmg.br

**Abstract**—Lançada pelo Governo Federal brasileiro no final de 2002, a plataforma Tesouro Direto possibilita a compra de títulos públicos por pessoas físicas, caracterizando-o como um investimento de renda fixa seguro e mais rentável que a poupança. Considerando as variações que as taxas de juros desses títulos sofrem ao longo do tempo, existe a possibilidade de se obter lucros acima do acordado inicialmente com a compra/venda no momento certo. Assim, este trabalho propõe a utilização de modelos de aprendizado de máquina, mais precisamente das redes neurais MLP, CNN e LSTM, na tentativa de se criar um modelo de previsão das taxas de juros de quatro títulos pós-fixados do Tesouro Direto atrelados ao IPCA. Para isso, foram analisadas e definidas algumas *features* (características) da economia brasileira que poderiam impactar mais diretamente nas taxas do Tesouro Direto. Simulações computacionais foram realizadas com variações de presença dessas características nos conjuntos de dados para que o melhor modelo de previsão fosse identificado. Ao final, concluiu-se que a CNN apresenta um melhor resultado geral, mesmo se mostrando mais sensível à remoção de *features*. Em contrapartida, a MLP apresenta erros quase constantes, independentemente da variável que está sendo removida. Além disso, descobriu-se que o CDI, importante métrica de investimentos, impacta de forma negativa a capacidade de previsão dos modelos, devido ao fato de seu desempenho estar relacionado à variação da Taxa Selic.

**Index Terms**—Tesouro Direto, série temporal, CNN, LSTM, MLP, títulos públicos

## I. INTRODUÇÃO

Segundo a Confederação Nacional do Comércio de Bens, Serviços e Turismo [1], cerca de 66,5% da população brasileira esteve endividada durante o ano de 2020, representando um aumento de 2,8 pontos percentuais em comparação com 2019. Apesar disso, a Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais (ANBIMA) [2] mostrou, em seu último relatório de 2020, que 44% dos brasileiros possuíam algum tipo de ativo financeiro no ano de 2019, sendo que 84,2% desse total possuíam a caderneta de poupança como destino de suas economias.

Embora o número de CPFs cadastrados na B3 (bolsa de valores oficial do Brasil) tenha atingido a marca de 3,17 milhões em novembro de 2020 [3], é nítido observar, pela pesquisa da ANBIMA, que os brasileiros possuem um perfil mais conservador e, portanto, uma preferência por ativos de renda fixa.

Dentre esses ativos de renda fixa, os títulos públicos são os que melhor aliam a segurança da poupança com uma maior rentabilidade. Sendo assim e de forma a democratizar o acesso de pessoas físicas diretamente ao mercado de títulos, o Governo Federal lançou a plataforma digital Tesouro Direto no final de 2002. A captação de investimentos desses títulos públicos auxilia o Governo a financiar atividades a nível federal, tais como educação, saúde e infraestrutura [4].

Os títulos públicos disponíveis para compra no Tesouro Direto pertencem a um dos seguintes tipos: prefixados, nos quais se sabe exatamente a rentabilidade anual do investimento; e os pós-fixados, que são compostos por uma taxa predefinida no momento da compra do título, adicionada à variação da inflação (IPCA - Índice Nacional de Preços ao Consumidor Amplo) ou da taxa básica de juros da economia brasileira (Selic - Sistema Especial de Liquidação e Custódia).

Embora os ativos de renda fixa, no geral, não forneçam um retorno tão alto quanto os de renda variável, é possível obter um excelente lucro no Tesouro Direto a partir da compra de um título a uma certa taxa, com a sua subsequente venda quando a taxa de compra tiver diminuído. Com isso, torna-se relevante prever as taxas de um determinado título e, dessa forma, prever o melhor momento para realizar a sua compra e/ou venda.

Dentre esses títulos, o mais interessante para a análise que será proposta nesse trabalho é o Tesouro IPCA, já que a volatilidade da inflação e da taxa adicionada ao índice garantem alterações no valor de compra/venda, oferecendo uma excelente oportunidade de se obter lucros acima do esperado [5].

Embora existam muitos estudos sobre o funcionamento do Tesouro Direto e do mercado da dívida pública brasileira, não foram encontrados trabalhos que fazem uma análise temporal dos dados disponíveis do Tesouro Direto. O único trabalho encontrado sobre esse tópico específico foi escrito por Cedismondi [6] e visa uma previsão utilizando apenas métodos estatísticos. Alguns trabalhos que tratam de séries temporais no mercado financeiro e que serviram de base para este trabalho serão referenciados na próxima seção.

Assim, o objetivo desse trabalho é construir e comparar os modelos de redes neurais artificiais MLP (*Multi-Layer Perceptron*), CNN (*Convolutional Neural Network*) e LSTM (*long short-term memory*) aplicados na previsão das taxas de

compra do título do Tesouro Direto pós-fixados atrelados ao IPCA.

## II. METODOLOGIA

A metodologia do presente trabalho é constituída pelas etapas a seguir:

### A. Definição das features

A primeira etapa do trabalho foi entender detalhadamente o funcionamento do Tesouro Direto e a sua relação com os vários indicadores econômicos do contexto brasileiro que impactam mais diretamente as suas taxas. Em qualquer estudo do cenário econômico brasileiro, a taxa Selic e o IPCA são de extrema relevância [7]. Assim, essas duas características foram incluídas no *dataset*, bem como o CDI (Certificado de Depósito Bancário), por ser um *benchmark* utilizado como referência em retornos de investimentos.

No único trabalho diretamente relacionado com o que está sendo proposto neste artigo, Cedismondi [6] fez uma análise estatística da correlação de diversos indicadores econômicos com a variação das taxas de juros dos títulos públicos brasileiros. A partir dos resultados obtidos por Cedismondi [6] definiu-se, para o presente trabalho, o câmbio dólar/real e o Tesouro Americano (títulos públicos estadunidenses, conhecidos como *treasuries*) como características econômicas a serem utilizadas pelos modelos propostos, adicionalmente às três citadas anteriormente.

### B. Tratamento dos dados

A base de dados final utilizada neste trabalho foi construída por meio da integração de cinco *datasets* distintos, cada um referente a uma *feature* específica. Sendo assim, foi realizado um tratamento dos dados de forma que a união de todas as informações em uma única tabela pudesse ser feita com uma simples comparação de valores. Após unir todos os *datasets* em um único, as linhas contendo valores nulos foram removidas e os valores foram normalizados entre 0 e 1. A Fig. 1 mostra as taxas de compra de um desses títulos, o Tesouro IPCA com vencimento em 2024. Embora a taxa final do título esteja atrelada ao IPCA, as taxas de compras disponíveis dizem respeito apenas à parte pré-fixada dos juros.

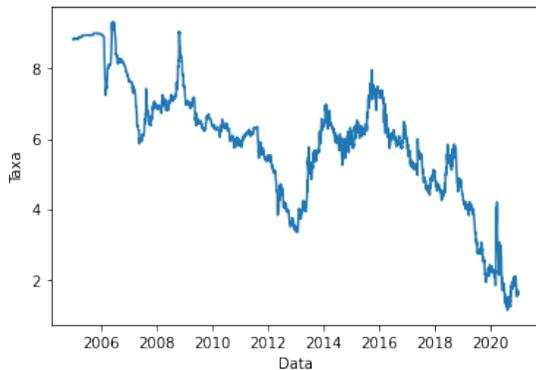


Fig. 1. Taxas de compra do Tesouro IPCA 2024 de 18/07/2005 a 14/01/2021

### C. Análise do dataset consolidado

Com o *dataset* completo, realizou-se um estudo estatístico das relações entre as variáveis escolhidas para representar o problema, antes da construção dos modelos de aprendizado de máquina. Nesse sentido, calculou-se a autocorrelação total e parcial entre as taxas de compra, a correlação entre as variáveis presentes no *dataset*, além da verificação da estacionariedade das séries. Todos os testes a serem exibidos nessa seção se referem ao Tesouro IPCA 2024, mas os demais títulos IPCA apresentaram resultados similares à série estudada e tomada como referência.

Para melhor entender a relação que uma taxa de juros de um título tem com as taxas de dias anteriores, foi realizado o cálculo da autocorrelação e da autocorrelação parcial. A primeira medida analisa a correlação de um valor no tempo  $t$  com os valores subsequentes até  $t - n$ , considerando a relação que os dias nesse intervalo possuem entre si. Em contrapartida, a autocorrelação parcial leva em consideração apenas a relação isolada de um valor no tempo  $t$  com o valor no tempo  $t - n$ .

Os resultados encontrados para o *dataset* do Tesouro IPCA 2024 se encontram nas Figuras 2 e 3. Nelas, o *threshold* da autocorrelação definido pela função empregada é exibido graficamente pela faixa azul ao redor do eixo x. Assim, pontos que se encontram acima ou abaixo desse limiar indicam uma autocorrelação relevante.

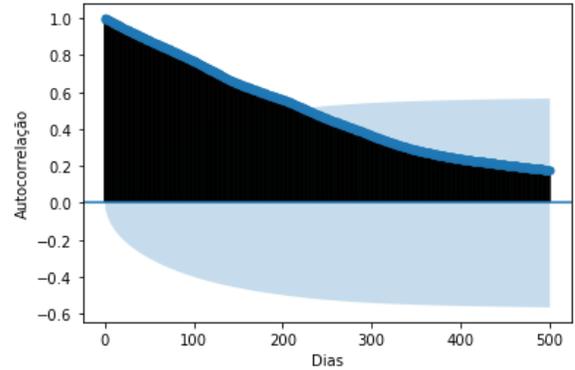


Fig. 2. Autocorrelação das taxas de compra do Tesouro IPCA 2024

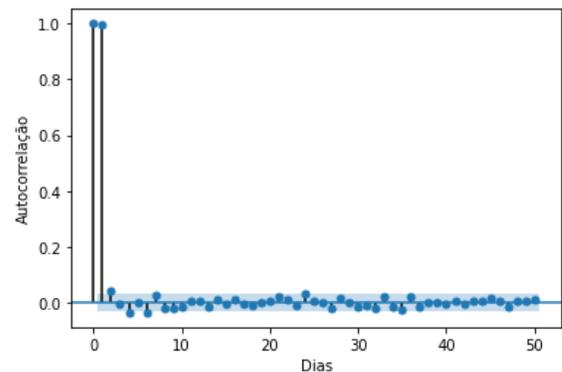


Fig. 3. Autocorrelação parcial das taxas de compra do Tesouro IPCA 2024

A partir do cálculo da autocorrelação, percebe-se que os dados passados se relacionam de forma relevante por volta de 250 dias. Esse valor coincide, aproximadamente, com a quantidade de dias úteis existentes no calendário brasileiro em um ano. Assim, pode-se concluir a indicação de uma sazonalidade anual dos dados, já que o Tesouro Direto negocia apenas em dias úteis. Para confirmar essa sazonalidade, foi feita uma decomposição dos dados, como ilustrado na Fig. 4. Apesar disso, a autocorrelação parcial demonstra que apenas o dia anterior é relevante para a previsão da taxa, já que  $t - 1$  é o único ponto acima do limiar delimitado.

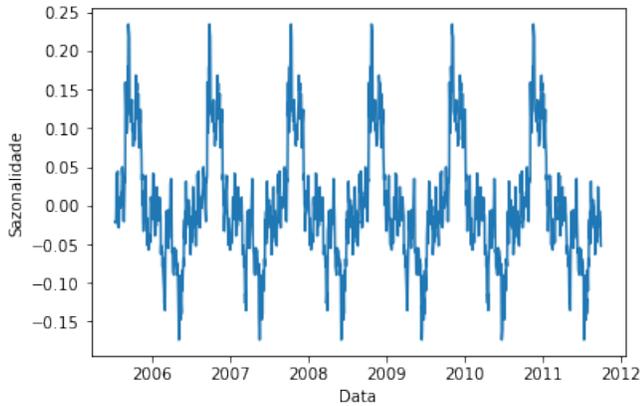


Fig. 4. Sazonalidade das taxas de compra do Tesouro IPCA 2024 de 18/07/2005 a 30/09/2011

Em seguida, calculou-se a correlação entre as *features* escolhidas, cujo resultado pode ser observado na Fig. 5. Nela, é possível comprovar a alta correlação entre a Selic e o CDI, visto que o valor das duas é definido de forma aproximada quando o mercado não espera movimentações da taxa básica pelo COPOM [8].

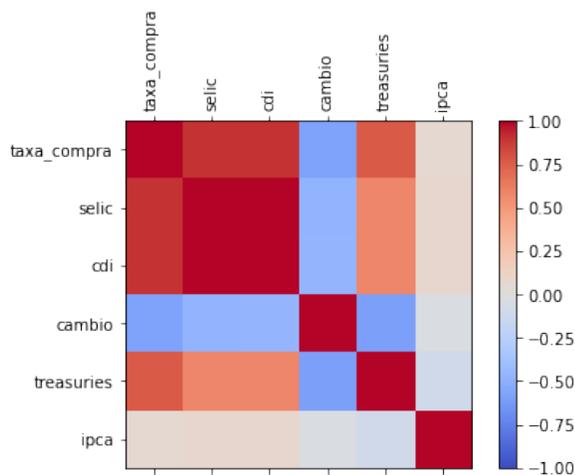


Fig. 5. Correlação entre as *features* escolhidas

Outra informação relevante é a correlação inversamente proporcional entre a Selic e a taxa de câmbio do real brasileiro

(BRL) para o dólar americano (USD). Isso ocorre pois um câmbio alto simboliza a desvalorização da moeda e, consequentemente, uma inflação alta. A medida que a inflação permanece alta por um longo período, a taxa Selic aumenta numa tentativa de diminuí-la. Assim, um câmbio alto ocorre quando a Selic está baixa, e vai diminuindo a medida que a Selic aumenta. Como tanto o CDI quanto as taxas de juros acompanham a taxa Selic, a explicação vale para as três variáveis.

É curioso notar a baixa correlação do IPCA com todos os outros indicadores, principalmente considerando que os títulos analisados possuem uma taxa variável atrelada ao próprio IPCA. Entretanto, de acordo com Fernandes et al. [9], essa baixa correlação se explica pela alteração na Selic ter um efeito tardio, de aproximadamente 20 dias, na inflação. Assim, embora funcione no longo prazo, no curto prazo, que é levado em consideração pelo cálculo de correlação, as variáveis aparentam ser descorrelacionadas.

Por fim, a estacionariedade das *features* foi analisada. Utilizando o teste de Dickey-Fuller aumentado [10], concluiu-se que todas as *features* são não-estacionárias, com exceção do IPCA. Para transformá-las em estacionárias foi feita uma diferenciação de primeira ordem. Na Fig. 6 é possível notar como as taxas são erráticas em alguns pontos, apresentando a influência que ações antropomórficas têm sob os dados. A partir dessa informação, descartou-se a possibilidade de usar modelos estatísticos no escopo do trabalho, como foi feito por Cedismondi [6], pois viu-se que o comportamento das taxas requereria modelos com maior capacidade de generalização, como redes neurais.

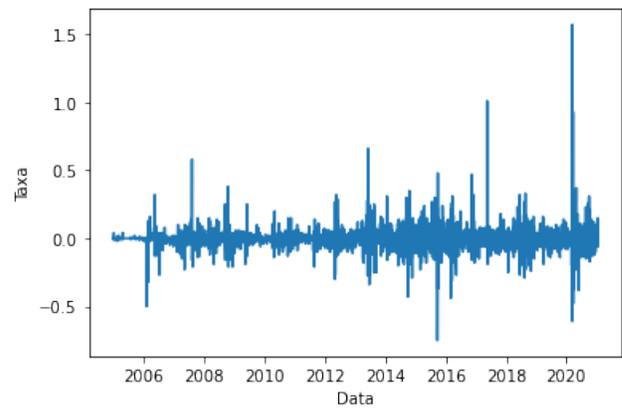


Fig. 6. Transformação das taxas de compra do Tesouro IPCA 2024 em uma série estacionária após diferenciação de 1ª ordem

#### D. Escolha dos modelos

Após a análise das características e relações existentes entre as *features* presentes na base de dados, foram consideradas três redes neurais para realizar a tarefa de previsão das taxas de compra dos títulos IPCA: *perceptron* multi-camadas (MLP, do inglês *Multi-Layer Perceptron*), rede neural convolucional (CNN, do inglês *Convolutional Neural Network*) e a rede *Long Short Term Memory* (LSTM).

O *perceptron* multi-camadas é considerado como uma das redes neurais mais simples e será utilizada neste trabalho como referência basal (*baseline*) para os demais métodos [11].

A inspiração para o uso da rede CNN veio do trabalho desenvolvido por Chen e Hongxiang [12], que implementaram uma CNN para fazer previsões no mercado de ações chinês. As redes CNN possuem a habilidade de prever dados futuros a partir do aprendizado de padrões que se repetem com o tempo. Esse aprendizado surge do uso de filtros, que são capazes de detectar padrões espaciais, muito utilizados na área de visão computacional. Como as taxas do Tesouro Direto apresentam um padrão sazonal, essa propriedade da rede poderia ser interessante para o presente trabalho.

Outra variação na implementação de uma rede neural são as Redes Neurais Recorrentes (do inglês *Recurrent Neural Networks* ou RNN), utilizadas para previsão em séries temporais [13] [14]. De forma simplificada, essa rede possui células capazes de armazenar o contexto de uma informação. A saída é modulada pelo chamado estado das células, que é definido a partir do contexto histórico das entradas para fazer as previsões. Como o *dataset* do Tesouro Direto possui dados desde 2003, o fato da LSTM ser capaz de lembrar de informações por um longo tempo a torna indicada para este trabalho. Essa rede também possui a habilidade de fazer um *dropout* nos dados, ou seja, eliminar taxas aleatoriamente de uma época a partir da probabilidade de aplicação de uma máscara binária [15].

### E. Implementação dos modelos

A implementação das redes propostas neste trabalho utilizou o *framework keras* com o conjunto de dados contendo as *features* referentes a  $n$  dias anteriores, como variáveis de entrada, e a taxa de compra do dia atual como variável de saída. Considerando a autocorrelação parcial obtida na subseção II-C, definiu-se  $n = 1$ , já que essa medida estatística indica a relevância de valores passados com relação ao valor atual.

Essa quantidade de dias passados relevantes para a previsão das taxas de compra dos títulos foi de especial importância para as redes CNN e LSTM, visto que esses modelos exigem a especificação de quantos passos temporais devem ser considerados.

Para automatizar e facilitar a escolha dos hiperparâmetros das redes neurais, foi utilizado o *framework optuna*, utilizado na otimização de hiperparâmetros em algoritmos de aprendizado de máquina [16]. Nele, possíveis valores para os hiperparâmetros são especificados e, a cada nova execução da rede, eles são atualizados (ou não) de forma a minimizar o erro quadrático médio do modelo (MSE, do inglês *Mean Squared Error*). Para o escopo desse trabalho, definiu-se como 50 o número de execuções dos modelos com variações nos hiperparâmetros do *optuna*.

## III. RESULTADOS

### A. Otimização dos modelos pelo *optuna*

A Tabela I exibe os valores finais dos hiperparâmetros escolhidos para compor o modelo definitivo de cada uma das redes. Para que pudesse ser feita uma comparação direta nas previsões realizadas, os modelos foram otimizados considerando todas as *features* presentes no conjunto de dados do Tesouro IPCA 2024. Assim, independentemente do título que está sendo previsto ou da variação realizada no *dataset*, todos os testes utilizaram o mesmo modelo de cada rede. Dos sete parâmetros, cinco são compartilhados entre todas as três redes e dois deles, filtros e *dropout*, são específicos das redes CNN e LSTM, respectivamente.

TABELA I  
HIPERPARÂMETROS OBTIDOS PELO PROCESSO DE OTIMIZAÇÃO REALIZADO PELO *optuna*

Parâmetro	MLP	CNN	LSTM
Filtros	-	128	-
Neurônios	1000	1000	100
Função de ativação	tanh	tanh	sigmoid
<i>Dropout</i>	-	-	0
Taxa de aprendizado	1.23 e-05	3.11 e-05	0.0009
Nº de épocas	400	400	200
<i>Batch size</i>	64	128	128

Pode-se observar, na Tabela I, que a rede LSTM aparenta ter uma vantagem sob as demais por ter atingido um ponto ótimo com apenas um décimo da quantidade de neurônios e com metade do número de épocas.

A Tabela II exibe a relevância, em porcentagem, de cada hiperparâmetro gerado pelo *optuna* durante o processo de otimização. Nota-se que nas três redes a taxa de aprendizado, que serve de parâmetro para o otimizador do modelo, foi a mais relevante de todas. Em todas as redes, o otimizador escolhido foi o Adam [17], que aparece na literatura como sendo um dos mais eficientes.

TABELA II  
IMPORTÂNCIA DOS HIPERPARÂMETROS OBTIDOS PARA A MINIMIZAÇÃO DO ERRO DOS MODELOS

Parâmetro	MLP	CNN	LSTM
Filtros	-	4%	-
Neurônios	0	2%	13%
Função de ativação	1%	9%	8%
<i>Dropout</i>	-	-	3%
Taxa de aprendizado	98%	81%	67%
Nº de épocas	1%	4%	3%
<i>Batch size</i>	0	0	6%

### B. Testes realizados nos modelos

As redes neurais foram implementadas e o desempenho foi avaliado pela métrica da raiz do erro quadrático médio (RMSE, do inglês *Root-Mean-Square Error*) [18]. Para a minimização do erro durante o processo de aprendizado das redes neurais, escolheu-se a métrica MSE, que utiliza a mesma fórmula anterior, porém, sem extrair a raiz. Cada um dos conjuntos de

dados testados foram divididos entre treino (60%), validação (20%) e teste (20%) [19].

Com a análise realizada na seção II-C, viu-se que a Selic e o CDI possuem uma altíssima correlação entre si, e que o IPCA possui uma correlação quase nula com qualquer uma das outras *features* escolhidas. Buscando determinar se essas descobertas afetariam o resultado final dos modelos, testes foram feitos de forma a remover essas três variáveis do conjunto de dados em todas as sete possíveis combinações de eliminação entre elas.

Para a validação do melhor modelo, foram utilizados quatro títulos: IPCA 2024, IPCA 2024 com Juros Semestrais, IPCA 2035 e IPCA 2035 com Juros Semestrais. Embora não tenham sido citados anteriormente, os títulos com juros semestrais têm por característica pagar dividendos a cada 6 meses para o investidor, ao invés de fazer um único pagamento na data de vencimento do título. A comparação entre os títulos foi realizada levando em consideração as combinações de redes neurais e de *features* que apresentaram o menor erro.

A Tabela III exibe os valores do RMSE para os testes realizados no Tesouro IPCA 2024. Nessa tabela, assim como em todas as seguintes, estão destacados os valores correspondentes à análise feita no artigo. O melhor resultado foi obtido pela rede CNN com a remoção do CDI do conjunto de dados, como mostra a previsão na Fig. 7. As outras duas redes também obtiveram os seus melhores resultados com a remoção dessa variável, mostrando que a presença do CDI gera um ruído que piora a previsão dos modelos. Na rede LSTM, o pior resultado foi obtido com a remoção das três *features*. A CNN, por sua vez, obteve o pior resultado quando o IPCA foi removido, gerando um erro quase cinco vezes maior do que o melhor caso para essa rede.

É interessante notar como, em todos os testes exibidos neste trabalho, a rede MLP apresentou um erro praticamente constante independentemente das *features* que estavam sendo removidas do *dataset*. Uma possível explicação é que, ao montar a equação que modela os dados, a MLP dá um peso pequeno a essas variáveis, resultando em um baixo impacto na previsão quando as mesmas são removidas.

TABELA III  
RAIZ DO ERRO QUADRÁTICO MÉDIO DOS TESTES REALIZADOS COM O *dataset* TESOURO IPCA 2024

<i>Feature(s)</i> removida(s)	MLP	LSTM	CNN
Nenhuma	0.126	0.128	0.138
CDI	<b>0.126</b>	<b>0.117</b>	<b>0.115</b>
IPCA	0.133	0.117	<b>0.516</b>
Selic	0.126	0.130	0.142
CDI + IPCA	0.126	0.135	0.199
CDI + Selic	0.126	0.151	0.130
IPCA + Selic	0.127	0.117	0.121
CDI + IPCA + Selic	0.126	<b>0.197</b>	0.176

Na Tabela IV têm-se os resultados dos modelos no conjunto de dados do Tesouro IPCA 2024 com Juros Semestrais. Novamente, o menor erro foi obtido na rede CNN porém, dessa vez, com a remoção das *features* IPCA e Selic. A Fig. 8 exibe a previsão realizada a partir desse modelo. Assim como no

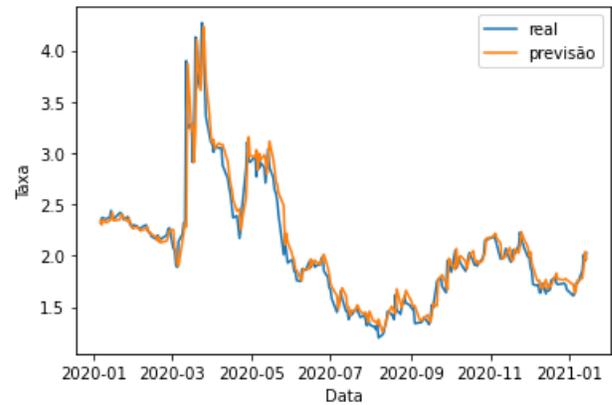


Fig. 7. Previsão do Tesouro IPCA 2024 pela rede CNN de janeiro de 2020 a janeiro de 2021

título anterior, as três redes apresentaram um bom desempenho com a remoção do CDI do conjunto de dados. Nota-se que a CNN se mostra mais sensível a alterações no *dataset*, embora a LSTM também apresente variações a cada teste.

TABELA IV  
RAIZ DO ERRO QUADRÁTICO MÉDIO DOS TESTES REALIZADOS COM O *dataset* TESOURO IPCA 2024 COM JUROS SEMESTRAIS

<i>Feature(s)</i> removida(s)	MLP	LSTM	CNN
Nenhuma	0.113	0.136	0.161
CDI	<b>0.114</b>	<b>0.113</b>	<b>0.118</b>
IPCA	0.113	0.116	0.224
Selic	0.114	0.119	0.125
CDI + IPCA	0.115	0.114	0.274
CDI + Selic	0.114	0.243	0.183
IPCA + Selic	0.114	0.115	<b>0.111</b>
CDI + IPCA + Selic	0.114	0.192	0.239

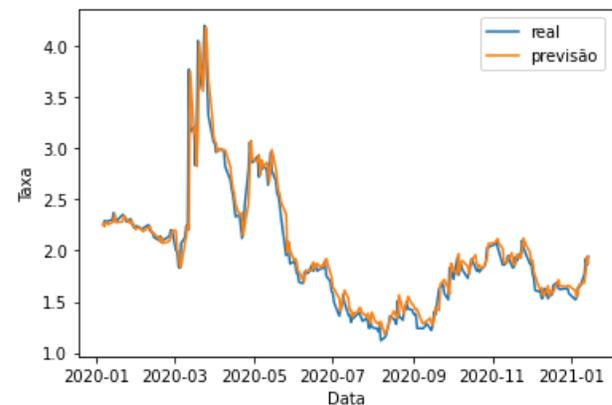


Fig. 8. Previsão do Tesouro IPCA 2024 com Juros Semestrais pela rede CNN de janeiro de 2020 a janeiro de 2021

Para o Tesouro IPCA 2035, a Tabela V mostra que, no geral, a melhor rede foi a MLP, devido a sua consistência e baixo erro. No entanto, o menor valor de erro encontrado dentre todas as redes foi obtido pela CNN, com a remoção das *features* CDI, IPCA e Selic. É interessante notar como, nesse

título, as variações nos erros é baixa, com valores próximos, tanto entre os modelos quanto entre a remoção de variáveis. Pode-se especular que, por se tratar de um título com uma data de vencimento distante (considerando a data de escrita desse artigo), ele se mostra menos suscetível a oscilações no curto prazo. A Fig. 9 mostra a previsão feita pela rede CNN.

TABELA V  
RAIZ DO ERRO QUADRÁTICO MÉDIO DOS TESTES REALIZADOS COM O *dataset* TESOURO IPCA 2035

Feature(s) removida(s)	MLP	LSTM	CNN
Nenhuma	0.106	0.139	0.112
CDI	0.101	0.116	0.108
IPCA	0.103	0.121	0.123
Selic	0.101	0.117	0.108
CDI + IPCA	0.102	0.112	0.109
CDI + Selic	0.102	0.117	0.099
IPCA + Selic	0.102	0.123	0.108
CDI + IPCA + Selic	0.103	0.116	<b>0.094</b>

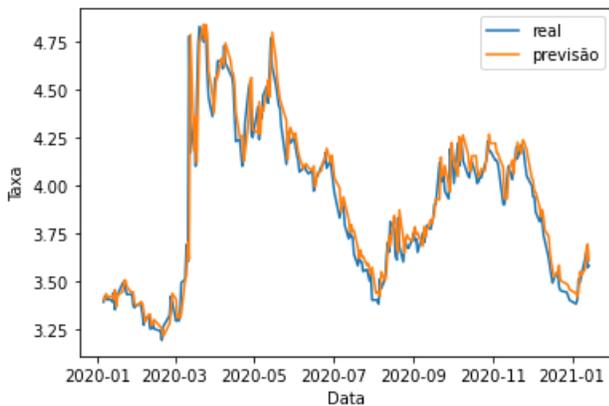


Fig. 9. Previsão do Tesouro IPCA 2035 pela rede CNN de janeiro de 2020 a janeiro de 2021

Por fim, o *dataset* do Tesouro IPCA 2035 com Juros Semestrais foi inserido nos modelos, com os erros obtidos exibidos na Tabela VI. Dessa vez, o menor erro foi obtido pela rede LSTM, tanto com a remoção apenas da Selic quanto com a remoção da Selic e do CDI. Assim como em todos os modelos anteriores, a remoção do CDI não impactou o desempenho das redes, se mostrando uma variável dispensável dentro dos conjuntos de dados na previsão da taxa de juros. O gráfico obtido a partir da LSTM com a remoção da Selic se encontra na Fig. 10.

C. Comparação dos resultados obtidos entre os títulos IPCA 2024 e IPCA 2035, com e sem juros semestrais, para cada um dos modelos de redes neurais utilizadas

Uma vez que os testes foram realizados, viu-se a necessidade de comparar os resultados entre as redes e entre os modelos. Por mais que os tamanhos dos conjuntos de dados sejam diferentes, a característica de pagamento semestral dos juros dos títulos Tesouro IPCA 2024 Semestral e 2035 Semestral instiga a realização de uma análise para determinar se essa similaridade é transmitida para os erros resultantes dos

TABELA VI  
RAIZ DO ERRO QUADRÁTICO MÉDIO DOS TESTES REALIZADOS COM O *dataset* TESOURO IPCA 2035 COM JUROS SEMESTRAIS

Feature(s) removida(s)	MLP	LSTM	CNN
Nenhuma	0.091	0.103	0.141
CDI	<b>0.091</b>	<b>0.088</b>	<b>0.106</b>
IPCA	0.093	0.088	0.163
Selic	0.091	<b>0.087</b>	0.129
CDI + IPCA	0.092	0.088	0.140
CDI + Selic	0.091	<b>0.087</b>	0.102
IPCA + Selic	0.091	0.088	0.143
CDI + IPCA + Selic	0.091	0.090	0.161

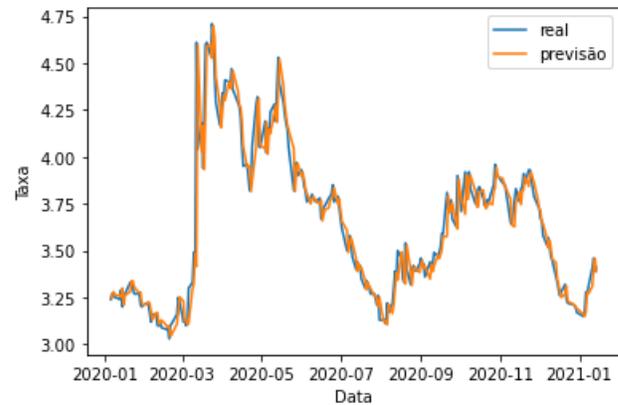


Fig. 10. Previsão do Tesouro IPCA 2035 com Juros Semestrais pela rede LSTM de janeiro de 2020 a janeiro de 2021

modelos. A mesma ideia é válida para os títulos Tesouro IPCA 2024 e 2035.

Pela Tabela VII, nota-se que o comportamento da rede MLP foi similar para o conjunto de dados dos títulos Tesouro IPCA 2024 e 2035, com um valor quase constante de erro, independente de qual *feature* estava sendo removida. Entretanto, os testes nos dois títulos não concordaram sobre qual remoção faz com que o erro aumente já que, para o Tesouro IPCA 2024, um maior RMSE foi obtido quando o IPCA foi removido, enquanto que para o Tesouro IPCA 2035, isso ocorreu quando nenhuma *feature* foi retirada do *dataset*. Apesar disso, o erro obtido nesses casos não foi discrepante dos outros resultados encontrados.

TABELA VII  
COMPARAÇÃO DA RAIZ DO ERRO QUADRÁTICO MÉDIO DA REDE MLP PARA OS *datasets* TESOURO IPCA 2024 E TESOURO IPCA 2035

Feature(s) removida(s)	IPCA 2024	IPCA 2035
Nenhuma	0.126	<b>0.106</b>
CDI	0.126	0.101
IPCA	<b>0.133</b>	0.103
Selic	0.126	0.101
CDI + IPCA	0.126	0.102
CDI + Selic	0.126	0.102
IPCA + Selic	0.127	0.102
CDI + IPCA + Selic	0.126	0.103

A Tabela VIII traz a comparação entre os mesmos títulos, porém com as execuções feitas na rede CNN. Nela, observa-se

uma maior variação entre os erros relacionada à *features* que estavam sendo removidas, embora essa variação seja menor para o Tesouro IPCA 2035. Para essa rede, o pior caso em ambos *datasets* foi obtido com a remoção do IPCA dos dados, com destaque para o alto valor de erro obtido no IPCA 2024.

TABELA VIII  
COMPARAÇÃO DA RAIZ DO ERRO QUADRÁTICO MÉDIO DA REDE CNN PARA OS *datasets* TESOURO IPCA 2024 E TESOURO IPCA 2035

Feature(s) removida(s)	IPCA 2024	IPCA 2035
Nenhuma	0.138	0.112
CDI	0.115	0.108
IPCA	<b>0.516</b>	<b>0.123</b>
Selic	0.142	0.108
CDI + IPCA	0.199	0.109
CDI + Selic	0.130	0.099
IPCA + Selic	0.121	0.108
CDI + IPCA + Selic	0.176	0.094

Como última comparação entre esses dois títulos, a Tabela IX apresenta os erros obtidos com as execuções da rede LSTM, com destaque para os maiores e menores valores. Novamente é exibida uma discordância entre os modelos com relação a remoção das *features*, além dos erros associados ao Tesouro IPCA 2035 serem mais estáveis entre os testes em comparação com o IPCA 2024.

TABELA IX  
COMPARAÇÃO DA RAIZ DO ERRO QUADRÁTICO MÉDIO DA REDE LSTM PARA OS *datasets* TESOURO IPCA 2024 E TESOURO IPCA 2035

Feature(s) removida(s)	IPCA 2024	IPCA 2035
Nenhuma	0.128	<b>0.139</b>
CDI	<b>0.117</b>	0.116
IPCA	<b>0.117</b>	0.121
Selic	0.130	0.117
CDI + IPCA	0.135	<b>0.112</b>
CDI + Selic	0.151	0.117
IPCA + Selic	<b>0.117</b>	0.123
CDI + IPCA + Selic	<b>0.197</b>	0.116

Para comparar o desempenho da rede MLP nos títulos Tesouro IPCA 2024 com Juros Semestrais e IPCA 2035 com Juros Semestrais, construiu-se a Tabela X. Assim como aconteceu com os títulos anteriores, os erros resultantes dessa rede se mostraram consistentes, independentemente da alteração feita nas variáveis de entrada. Portanto, pode-se inferir que, de maneira geral, não foi atribuída uma grande importância às

TABELA X  
COMPARAÇÃO DA RAIZ DO ERRO QUADRÁTICO MÉDIO DA REDE MLP PARA OS *datasets* TESOURO IPCA 2024 SEMESTRAL E TESOURO IPCA 2035 COM JUROS SEMESTRAIS

Feature(s) removida(s)	IPCA 2024 Sem.	IPCA 2035 Sem.
Nenhuma	0.113	0.091
CDI	0.114	0.091
IPCA	0.113	0.093
Selic	0.114	0.091
CDI + IPCA	0.115	0.092
CDI + Selic	0.114	0.091
IPCA + Selic	0.114	0.091
CDI + IPCA + Selic	0.114	0.091

*features* na construção do modelo e, consequentemente, na previsão das taxas de compra.

Em contrapartida, a rede CNN segue sendo sensível às variações realizadas nos dados que alimentam o modelo, como mostra a Tabela XI.

TABELA XI  
COMPARAÇÃO DA RAIZ DO ERRO QUADRÁTICO MÉDIO DA REDE CNN PARA OS *datasets* TESOURO IPCA 2024 SEMESTRAL E TESOURO IPCA 2035 COM JUROS SEMESTRAIS

Feature(s) removida(s)	IPCA 2024 Sem.	IPCA 2035 Sem.
Nenhuma	0.161	0.141
CDI	0.118	0.106
IPCA	0.224	0.163
Selic	0.125	0.129
CDI + IPCA	0.274	0.140
CDI + Selic	0.183	0.102
IPCA + Selic	0.111	0.143
CDI + IPCA + Selic	0.239	0.161

Como última comparação, a rede LSTM não obteve uma performance semelhante na comparação dos dois títulos com Juros Semestrais. A Tabela XII revela que, novamente, não houve uma consistência entre os modelos com relação ao valor do erro relacionado com a remoção de uma *feature*. A exceção a isso foi a exclusão do CDI que, como explicitado anteriormente, forneceu um bom resultado em ambos modelos.

TABELA XII  
COMPARAÇÃO DA RAIZ DO ERRO QUADRÁTICO MÉDIO DA REDE LSTM PARA OS *datasets* TESOURO IPCA 2024 SEMESTRAL E TESOURO IPCA 2035 COM JUROS SEMESTRAIS

Feature(s) removida(s)	IPCA 2024 Sem.	IPCA 2035 Sem.
Nenhuma	0.136	0.103
CDI	<b>0.113</b>	<b>0.088</b>
IPCA	0.116	0.088
Selic	0.119	0.087
CDI + IPCA	0.114	0.088
CDI + Selic	<b>0.243</b>	<b>0.087</b>
IPCA + Selic	0.115	0.088
CDI + IPCA + Selic	0.192	0.090

#### IV. CONCLUSÃO

Conforme investigado, viu-se que as taxas de juros dos títulos do Tesouro Direto possuem uma sazonalidade equivalente a um ano e uma autocorrelação parcial que indica a importância apenas do dia anterior para as previsões. Uma análise de correlação entre os dados também trouxe o importante *insight* da proximidade da Selic com o CDI, bem como a inesperada falta de correlação do IPCA, um dos mais importantes indicadores econômicos do Brasil, com as demais *features* escolhidas. Com o uso do *optuna*, os testes foram feitos de forma eficiente, já que a escolha dos hiperparâmetros das redes, um passo tão vital, pôde ser automatizado.

A partir desses parâmetros, as previsões e os erros associados a eles indicaram uma melhor performance da rede CNN, principalmente com a remoção do CDI. Como a Selic e CDI possuem valores próximos, pode-se concluir que a presença do CDI no conjunto de dados adiciona um nível de complexidade

e ruído que prejudica a capacidade de generalização das redes. O IPCA, embora tenha correlação quase nula com as taxas de compra dos títulos, não alterou significativamente a performance dos modelos quando foi removido, com exceção do teste da rede CNN no Tesouro IPCA 2024, que apresentou um alto erro quando essa variável foi excluída do conjunto de dados.

Além de concluir que a rede convolucional teve o melhor desempenho, é interessante destacar como ela se mostrou sensível à remoção de *features*, principalmente nos dois títulos com vencimento em 2024. Uma possível explicação é que, dada a proximidade da data de vencimento em comparação com os títulos a serem pagos em 2035, a importância de determinadas variáveis aumenta, enquanto outras menos importantes geram ainda mais ruído no aprendizado do modelo. Vale ressaltar, também, que o *perceptron* multicamadas, apresentou erros dentro de uma pequena faixa de valores em todos os títulos avaliados. Isso pode indicar uma atribuição de pesos pequenos às *features* que estão sendo removidas na equação gerada internamente pela rede, tornando as previsões mais dependentes dos valores das taxas de compra em si.

Como trabalhos futuros, propõe-se, inicialmente, a realização de um estudo mais aprofundado sobre as possíveis relações existentes entre os índices econômicos apresentados nesse trabalho e o valor das taxas de compra dos títulos do Tesouro, no que tange aos aspectos econômicos e estatísticos, bem como a realização de simulações de compra e venda desses títulos para uma análise mais robusta das previsões geradas pelos modelos propostos. A adição de modelos de análise de sentimento em notícias também traria um grande ganho para o estudo, já que os indicadores econômicos, bem como os investimentos associados a eles, são completamente influenciados por ações antropomórficas. Além disso, o presente trabalho se ateve a investigação da performance de redes neurais, sendo relevante o uso de outras técnicas de *machine learning* para a previsão das taxas de juros dos títulos do Tesouro Direto.

#### AGRADECIMENTOS

Os autores gostariam de agradecer ao CEFET-MG pelo suporte financeiro, sem o qual esse trabalho não teria sido possível.

#### REFERÊNCIAS

- [1] CNC, “Número de brasileiros endividados em 2020 foi o maior em 11 anos,” 2021. [Online]. Available: <http://cnc.org.br/editorias/economia/noticias/numero-de-brasileiros-endividados-em-2020-foi-o-maior-em-11-anos>
- [2] ANBIMA, “Raio x do investidor brasileiro - 3ª edição,” 2020. [Online]. Available: [https://www.anbima.com.br/pt\\_br/especial/raio-x-do-investidor-2020.htm](https://www.anbima.com.br/pt_br/especial/raio-x-do-investidor-2020.htm)
- [3] B. Alvarenga, “Bolsa dobra o número de investidores em 2020 e vê maior diversificação,” *Exame Invest*, 2020. [Online]. Available: <https://exame.com/invest/bolsa-dobra-o-numero-de-investidores-em-2020-e-ve-maior-diversificacao/>
- [4] M. Perlin, “A Microestrutura do Tesouro Direto: Sazonalidade do Fluxo de Ordens e o Processo de Formação de Spreads,” *Economia Aplicada*, vol. 20, no. 3, pp. 253 – 272, 2016.
- [5] Tesouro Direto, “Conheça todos os títulos do Tesouro Direto,” 2019. [Online]. Available: <https://www.tesourodireto.com.br/titulos/tipos-de-tesouro.htm>
- [6] G. N. d. S. Cedismondi, *Análise de um modelo de previsão da taxa de juros dos títulos públicos brasileiros*. São Paulo: Monografia (Graduação em Engenharia de Produção) - Escola Politécnica da Universidade de São Paulo, 2008.
- [7] E. C. Brito, E. C. d. Araújo, and E. L. Araujo, “Inter-relações entre a dívida pública e política monetária no Brasil: uma análise histórica,” *Economia e Sociedade*, vol. 28, pp. 153 – 175, abr. 2019.
- [8] R. Q. d. Amaral and J. L. Oreiro, “A relação entre o mercado de dívida pública e a política monetária no Brasil,” *Rev. econ. contemp.*, vol. 12, no. 3, pp. 491 – 517, dez. 2008.
- [9] B. Fernandes, P. R. A. Brene, A. M. M. Caravieri, A. G. Bernardelli, and C. D. Aoki, “Regime de metas de inflação no Brasil: Uma análise da correlação entre o IPCA e a Taxa Selic durante 1999 e 2015,” *Revista Espacios*, vol. 38, 2017.
- [10] R. Harris, “Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test,” *Economics Letters*, vol. 38, no. 4, pp. 381–386, 1992.
- [11] D. Munandar, “Multilayer Perceptron (MLP) and Autoregressive Integrated Moving Average (ARIMA) Models in Multivariate Input Time Series Data: Solar Irradiance Forecasting,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 1, 2019.
- [12] S. Chen and H. He, “Stock prediction using convolutional neural network,” *IOP Conference Series: Materials Science and Engineering*, vol. 435, nov. 2018. [Online]. Available: <https://doi.org/10.1088/1757-899x/435/1/012026>
- [13] F. d. A. Moura, *O uso de redes neurais artificiais na previsão de tendências no mercado de ações*. Pernambuco: Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Pernambuco, 2006.
- [14] D. Nelson, A. Pereira, and R. de Oliveira, “Stock market’s price movement prediction with LSTM neural networks,” mai. 2017, pp. 1419–1426.
- [15] G. Cheng, V. Peddinti, D. Povey, V. Manohar, S. Khudanpur, and Y. Yan, “An Exploration of Dropout with LSTMs,” ago. 2017, pp. 1586–1590.
- [16] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” pp. 2623–2631, jul. 2019.
- [17] D. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *International Conference on Learning Representations*, dez. 2014.
- [18] T. Chai and R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature,” *Geoscientific Model Development*, vol. 7, pp. 1247–1250, jun. 2014.
- [19] B. D. Ripley, “Glossary,” in *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996, p. 354.