**Derivativos de commodities: gestão de risco de mercado**

Renata Nunes Pereira, Rodrigo Vidal Cabral, Luiz Philipe Rangel

O presente artigo faz parte do programa do curso de Data Science da Tera, turma 02/2022 e tem como objetivo aplicar os conhecimentos adquiridos ao longo do curso.

**Introdução**

Você já ouviu falar de contratos futuros? E de opções ou swaps? Todos esses nomes são tipos de derivativos, ou seja, contratos que derivam a maior parte de seu valor de um ativo subjacente, como commodities.

Esse instrumento financeiro é muito usado pelos investidores para proteger suas operações, ou mesmo para lucrar no mercado.

O Brasil é um grande produtor e exportador de commodities. Por isso, o mercado de negociação é tão ativo e oferece muitas oportunidades tanto para empresas como para pessoas físicas.

**Contexto**

O que são Commodities?

Commodities são todas as matérias-primas essenciais que possuem baixo nível de industrialização. No Brasil, as principais commodities negociadas são agrícolas ou minerais, tais como: soja, trigo, laranja, petróleo, minério de ferro, ouro e boi.

Algumas commodities podem ser negociadas na B3 (Bolsa de Valores brasileira, BM&FBovespa) através do mercado futuro. No entanto, a maior parte das negociações de commodities são realizadas no balcão da B3.

O Mercado de Balcão da B3 é o ambiente que permite a realização de operações que não estão registradas na Bolsa de Valores. No Balcão da B3, as corretoras e distribuidoras de valores e bancos de investimento conectam pessoas que desejam comprar e vender ativos.

Como são fundamentais para o mundo todo, as commodities têm um grande peso na economia, sendo indispensáveis para nossa sobrevivência e o desenvolvimento dos países. Por isso, a negociação dos preços é realizada em escala global.

# O que são derivativos?

Derivativos são contratos que derivam a maior parte de seu valor de um ativo subjacente, taxa de referência ou índice. O ativo subjacente pode ser físico (café, ouro, etc.) ou financeiro (ações, taxas de juros, etc.), negociado no mercado à vista ou não (é possível construir um derivativo sobre outro derivativo). Os derivativos podem ser classificados em contratos a termo, contratos futuros, opções de compra e venda, operações de swaps, entre outros, cada qual com suas características.

Os derivativos, em geral, são negociados sob a forma de contratos padronizados. Isto é, previamente especificados (quantidade, qualidade, prazo de liquidação e forma de cotação do ativo-objeto sobre os quais se efetuam as negociações), em mercados organizados, com o fim de proporcionar, aos agentes econômicos, oportunidades para a realização de operações que viabilizem a transferência de risco das flutuações de preços de ativos e de variáveis macroeconômicas.

Como o nome sugere, derivativos utilizam outros ativos como referência, o dólar por exemplo, e são contratos que terão valores econômicos de acordo com a variação do valor, dentro de um determinado tempo. É o que acontece com [contratos futuros](https://www.modalmais.com.br/blog/o-que-e-contrato-futuro), por exemplo. Você negocia o contrato de um ativo baseado no preço atual, com vencimento em uma data futura, apostando na possibilidade de oscilação de valor, fato que gera rentabilidade em seu investimento.

O mercado futuro de derivativos também pode ser uma opção para quem deseja se proteger contra a alta taxa volatilidade do mercado, isto é, estão em constante variação de preço, essa operação é denominada hedge.

Essa proteção é feita a partir de uma estratégia de definição do preço futuro de uma ação ou outro ativo, tomando como base os preços e as cotações atuais do mercado.

Podemos dizer que os produtores de commodities, como milho, café e boi gordo, por exemplo, realizam a venda da produção por meio de contratos futuros. Isso quer dizer que eles fixam o preço agora, mas repassam o dinheiro, ou os próprios produtos em alguns casos, só no futuro. Essa operação ajuda a controlar o mercado, evitando que a imprevisibilidade prejudique as transações e afete todo o resultado da produção desses produtos.

# O que é o Mercado Futuro?

No mercado futuro você compra ou vende contratos com um prazo determinado no futuro. Existe um lote mínimo de contratos que você deve negociar, podendo ser sacas de café, arrobas de boi gordo, ou até mesmo pontos de índices, por exemplo.

Quando uma pessoa compra ou vende um contrato futuro ela assume uma posição nesse mercado. Assim sendo, quem está comprando lucra com a alta na cotação do ativo e perde quando há queda no preço. Bem como, quem está vendendo lucra com a queda na cotação do ativo e perde quando há alta no preço.

Quem compra e vende nesse mercado garante resultados diários com base na diferença entre o valor negociado e a cotação alcançada. Essa é uma espécie de apuração de perdas e ganhos feita pela Bolsa de Valores, conhecida como ajuste diário.

# Problema de Negócio

Baixa operação de derivativos de commodities para médias e pequenas empresas devido ao pouco conhecimento de operação no mercado.

# Impacto

Alavancar o mercado de derivativos, principalmente para proteção de pequenos e médios produtores que não possuem áreas financeiras específicas para isso ou com baixo conhecimento do mercado (ex.: cooperativas).

# Fonte dos Dados e Jupyter

Os dados utilizados foram da B3 e do Banco Central. Os dados da B3 foram fornecidos com sigilo, pois não poderíamos expor os clientes e suas transações, enquanto que os dados do Banco Central foram extraídos via Scrapping.

Os dados da B3 foram entregues em duas principais tabelas, uma “cotações e a outra “termo\_mercadoria, porém, subdividido em 671 arquivos para a primeira tabela e 713 arquivos para a segunda, todos no formato .xlsx, respectivamente. E o período de estudo foi de 26/08/2019 até 18/05/2022.

O acesso especificamente aos dados que usamos e o próprio arquivo Jupyter (programa utilizado para fazer a análise e o modelo) podem ser encontrados nos link’s abaixo:

Dados:https://raw.githubusercontent.com/seade-R/dados-covid-

Notebook: https://github.com/casjunior93/Projeto-Tera-IA---Previsao-de-obitos-por-covid-

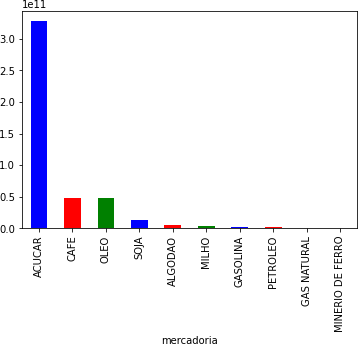
**EDA (Análise Exploratória de dados)**

Inicialmente, os DataFrames com os dados de cotações (‘cotacoes’) e mercadorias negociadas (‘termo-mercadoria’) foram unidos por meio de um “merge” e acrescidos dos dados do Banco Central.

Para melhor manuseio do DataFrame final, este foi salvo como mastertable.pkl.

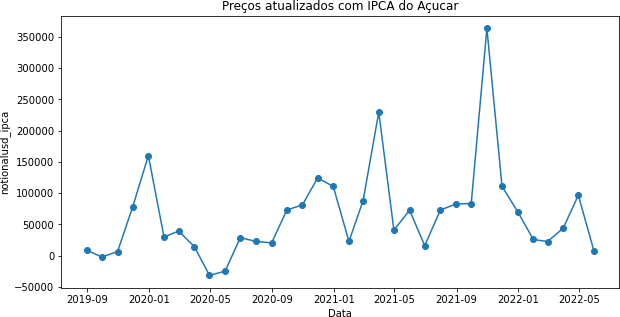
As principais commodities negociadas foram identificadas a partir da Figura 1.

Figura 1 - Principais commodities negociadas.



A partir da Figura 1 é possível observar que o açucar é a principal commodity negociada. Os preços atualizados com IPCA desta commodity pode ser verificada na Firgura 2.

Figura 2 - preços atualizados com IPCA do acúcar.



Percebendo-se que as variáveis são sazonais, ou seja, variam com o tempo, Séries temporais foi avaliada como um possível abordagem no presente estudo.

**Modelagem**

Parapredição da commodity os modelos DecisionTree, RandomForest e LGBM foram desenvolvidos (Figura 3), conforme os seguintes passos:

1. As features foram construídas utilizando o passado da base e o futuro para a variável resposta.
2. Foi definido que o target era a previsão dos próximos 30 dias.
3. O módulo TimeSeriesSplit foi utilizado para fazer a validação cruzada.
4. Foi obtido um objeto split, como em um KFold usado em validações cruzadas de problemas não temporais. Foi checado o tamanho de cada fold a cada iteração, porém foi preciso separar os dados que o modelo nunca teria acesso, mesmo na validação cruzada. Como se tinha dados temporais e se queria testar a performance no futuro, foi realizada uma separação Out of time ( “o corte”).
5. Os últimos 6 meses foram separados para teste.
6. Então, foi escolhida uma métrica e definido o modelo ideal:
   1. mean\_squared\_error ou MSE: média das diferenças entre a variável verdadeira e a predita ao quadrado. Quanto menor, melhor.
   2. mean\_absolute\_error ou MAE: média das diferenças entre a variável verdadeira e a predita em módulo. Quanto menor, melhor.
   3. root\_mean\_squared\_error ou RMSE: raíz quadrada do mean\_squared\_error. Quanto menor, melhor.
   4. r2\_score ou apenas r2: percentual da variância total da variável resposta que pode ser explicada pelo modelo além do que a média já explica. Quanto maior, melhor.
7. Em séries temporais, há uma grande preferência por uso de métricas com escala, como o MAE ou o RMSE. Para este problema, inicialmente, ufoi utilizado o MAE.
8. E ainda, foi utilizada validação cruzada de séries temporais e a métrica MAE para avaliar a performance destes modelos e fazer um gráfico.

Figura 3 - modelos DecisionTree, RandomForest e LGBM.



Pela Figura 3, observa-se que o melhor modelo com os hiperparâmetros usados é o LGBM. Então, este modelo foi escolhido para otimização.

**Seleção de variáveis: inferindo a janela ideal de defasagem**

Na grande maioria dos problemas de séries temporais, as bases de dados não são tão grandes, exceto em casos em que a captura de dados é muito alta (dados em segundos, milisegundos, etc.) Entretanto, dependendo do cenário, podemos ter bases de dados com muitas features. Assim, alguns tamanhos de base foram testadas.

Número de features com 10 defasagens: 52

Número de features com 15 defasagens: 72

Número de features com 20 defasagens: 92

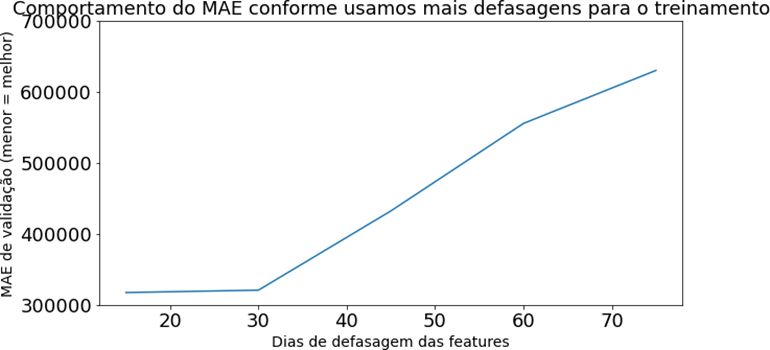
Número de features com 30 defasagens: 132

Número de features com 45 defasagens: 192

Número de features com 60 defasagens: 252

Até mesmo pensando na implantação de modelos, muitas features podem tornar o modelo pesado e complexo. Então, foi realizada uma seleção de variáveis baseada nos resultados. Para isso, foi analisado se a performance melhoria a medida em que mais features fossem utilizadas.

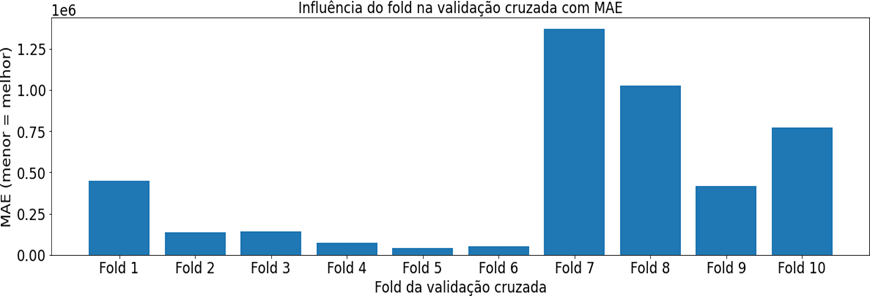
Figura 4 - Comportamento do MAE.



Pela Figura 4, é possível observar que utilizando mais features defasadas, não se obteve um ganho de performance significativo. Sendo assim, foi defiindo uma trava 15 e não 30 dias.

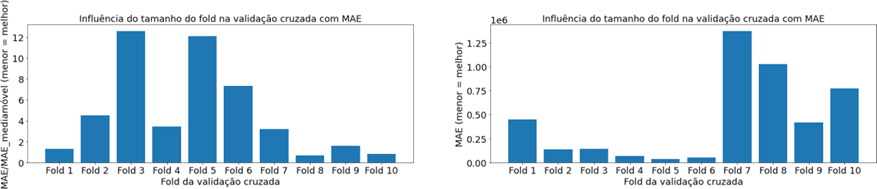
Um dos pontos mais importantes da validação cruzada de séries temporais é a escolha da métrica. Isso importa bastante, pois o tamanho do conjunto de treino varia a cada iteração. Isso foi avaliado utilizando cross\_val\_score.

Figura 5 - Influência do fold na validação cruzada com MAE (1).



Quanto menor a quantidade de dados que se tem para treinar o modelo, maior a possibilidade de underfitting, e a métrica precisa ser robusta para evitar conclusões incorretas. Uma solução simples para isso, é comparar a métrica com a métrica de um modelo de base. No caso do r2\_score, o modelo de base é uma média simples da variável resposta. Assim, foi calculado para todos os dados a média móvel do período de defasagem para o passado, e utlizada a métrica deste modelo "simples" para avaliação.

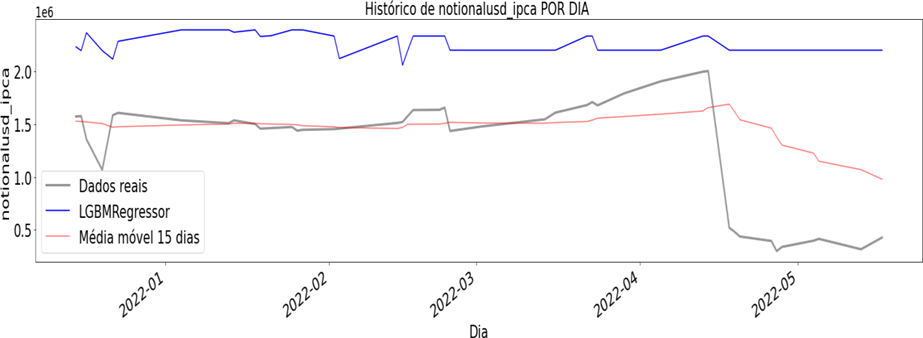
Figura 6 - Influência do fold na validação cruzada com MAE (2).



**Otimização de Hiperparâmetros e modelo final**

Agora com o modelo escolhido, os hiperparâmetros foram otimizados e plotados no modelo final. Para isso, foi utilizado o GridSearchCV específico para o LGBM. Foram plotamos os dados de treino e teste para 6 meses do conjunto de teste.

Figura 7 - Histórico de IPCA por dia com LGBMRegressor.



Pela Figura 7, verificou-se que o LGBMRegressor não propiciou uma boa performance.

Usando o Keras, uma rede neural foi construída adicionando camadas uma após a outra, com o objetivo de solucionar o problema.

Apenas para recapitular:

* A variável resposta foi construída como sendo o valor médio de notionalusd\_ipca de 15 dias seguintes.
* As features 15 anteriores foram consideradas.
* Nenhum modelo entende que essa ordem existe.

Quando se adiciona uma camada LSTM (Long Short Term Memory) à rede neural, se está dizendo que a ordem dos dados deve afetar a resposta, portanto se ganha uma nova camada de impacto em relação aos outros modelos.

Para treinar um modelo no Keras, por ser uma rede neural, foram realizados mais dois tratamentos:

* + Escala dos dados, pois não é um modelo robusto a escala .
  + Adicionar uma dimensão (para se tornar um tensor), mesmo que tenha dimensão 1, de modo em que o modelo nos diz a função custo a cada iteração.

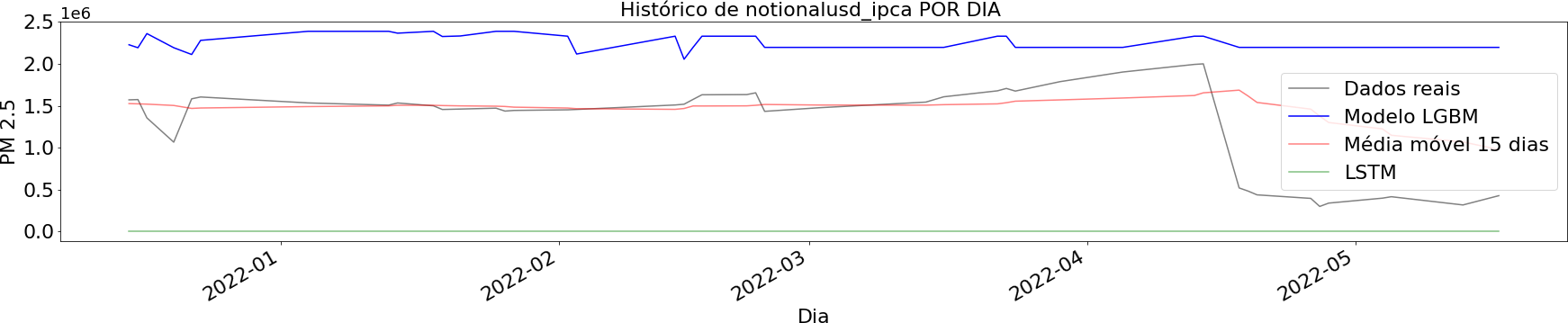
A função custo (MAE) foi plotada, no treino e no teste, ao longo das iterações (Figura 8).

Figura 8 - função custo (MAE) ao longo das iterações.

Uma imagem contendo Retângulo

Descrição gerada automaticamente

Figura 9 - IPCA por dia com LSTM.



Pela Figura 9, observamos que mesmo com esse modelo a média móvel ainda foi melhor.

# REFERÊNCIAS

HULL, J. C. Opções, futuros e outros derivativos. 9 ed. Porto Alegre: Bookman, 2016.

https://[www.investidor.gov.br/menu/Menu\_Investidor/derivativos/Derivativos\_int](http://www.investidor.gov.br/menu/Menu_Investidor/derivativos/Derivativos_int) roducao.html

https://[www.topinvest.com.br/o-que-sao-derivativos/](http://www.topinvest.com.br/o-que-sao-derivativos/) https://blog.toroinvestimentos.com.br/hedge-o-que-e