



UNIVERSIDADE FEDERAL DE OURO PRETO
ESCOLA DE MINAS DE OURO PRETO
DEPARTAMENTO DE ENGENHARIA DE MINAS



MARCO ANTONIO CERQUEIRA DE OLIVEIRA

**Criação de Modelo Econométrico Para Previsão do Preço
Spot do Cobre**

OURO PRETO

2021

MARCO ANTONIO CERQUEIRA DE OLIVEIRA

**Building of Econometric Model to Predict Copper Spot
Price**

**Criação de Modelo Econométrico Para Previsão do Preço
Spot do Cobre**

Trabalho de Conclusão de Curso apresentado ao departamento de Engenharia de Minas da Escola de Minas de Ouro Preto, como requisito para a obtenção de título de bacharel em engenharia de minas.

Orientador: Prof. Dr. Felipe Souza Ribeiro

OURO PRETO

2021



FOLHA DE APROVAÇÃO

Marco Antônio Cerqueira de Oliveira

CRIAÇÃO DE MODELO ECONOMÉTRICO PARA PREVISÃO DO PREÇO SPOT DO COBRE

Monografia apresentada ao Curso de Engenharia de Minas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Engenheiro de Minas

Aprovada em 30 de Abril de 2021

Membros da banca

Dr - Felipe Ribeiro Souza - Orientador(a) Universidade Federal de Ouro Preto
Dr^a - Tatiana Barreto dos Santos - Universidade Federal de Ouro Preto
M. Sc - Maria Clara Martins- Universidade Federal de Ouro Preto

Felipe Ribeiro Souza, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 22/08/2022



Documento assinado eletronicamente por **Felipe Ribeiro Souza, PROFESSOR DE MAGISTERIO SUPERIOR**, em 22/08/2022, às 08:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0384576** e o código CRC **B887749C**.

AGRADECIMENTOS

Agradeço a meus pais, Nih e Artur, por sempre estarem ao meu lado, sempre me apoiarem e me incentivarem a seguir meus sonhos.

A minha irmã Debora, por ser uma inspiração para mim e por sempre me incentivar.

A todos meus amigos, que nunca saíram do meu lado e sempre estiveram lá por mim quando precisei deles.

Ao meu orientador, Prof. Dr. Felipe Ribeiro, por todo apoio e auxílio ao longo deste trabalho tão importante.

“If I have seen further, it is by standing on
the shoulders of Giants.”

Sir Isaac Newton

RESUMO

Devido ao fato de as mineradoras serem tomadores de preço, a engenharia de minas tem uma forte ligação com o mundo da economia, o que pode ser observado nos termos técnicos usados frequentemente na área como: *treatment charges*, *refining charges*, teor de corte, ponto de equilíbrio etc. Não obstante, são raras as ocasiões nas quais obstáculos na mineração são resolvidos buscando técnicas usadas com frequência pelos profissionais de finanças e economia. Por falta de modelos de previsão de preços de commodities, muitas mineradoras ficam expostas a oscilações nos preços de mercado sendo que poderia fazer *hedge* a fim de repassar esse risco para especuladores. Modelos de aprendizado de máquina são amplamente utilizados na econometria e ciência de dados e podem ser manejados para buscar prever os preços de *commodities* no futuro próximo, a fim de fazer *hedge* contra eventuais oscilações nos preços no mercado financeiro, vendendo e/ou comprando contratos futuros para otimizar operações, maximizar o lucro nas vendas de concentrados ou até como forma de captação de recursos, por meio da venda de opções. Esta dissertação concebeu um modelo econométrico que prevê de forma lucrativa o retorno do preço *spot* do Cobre negociado na *London Metals Exchange* (LME) empregando análise de componentes principais e técnicas de aprendizado de máquina como regressão de Lasso.

Palavras chave: Aprendizado de máquina; Commodities; Preço spot; Modelo Econométrico.

ABSTRACT

Given the fact that miners are price-takers, mining engineering has a strong connection with the world of the economy, which can be seen in the technical terms often used in the area such as: Treatment charges, refining charges, cutoff grade, break even etc. Nevertheless, there it is uncommon for obstacles in mining to be resolved by seeking techniques that are frequently used by finance and economics professionals. Due to the lack of models for forecasting commodity prices, many mining companies are exposed to fluctuations in market prices that can be hedge by passing the risk on to speculators. Machine learning models are widely used in econometrics and can be managed to seek to forecast commodity prices in the near future, in order to hedge against possible price fluctuations in the financial market, selling and / or buying future contracts to optimize operations, maximize profit on sales of concentrates or even as a means of raising funds, through the sale of options. This dissertation conceived an econometric model that profitably predicts the return of the spot price of copper traded on the London Metals Exchange using machine learning techniques such as Lasso regression and principal component analysis.

Key words: Machine learning; Commodities; Spot Price; Econometric Model.

SUMÁRIO

<u>1 INTRODUÇÃO</u>	1
<u>2 OBJETIVO</u>	2
<u>3 REVISÃO BIBLIOGRAFICA</u>	3
<u>3.1 Mercados de balcão</u>	3
<u>3.2 Os agentes do Mercado e Seus Objetivos</u>	3
<u>3.3 Sistemas de Venda</u>	3
<u>3.4 Conceito de Volatilidade e Hedge</u>	4
<u>3.5 Arbitragem</u>	4
<u>3.6 O Mercado de <i>Commodities</i></u>	5
<u>3.7 Métodos Econométricos</u>	5
<u>3.8 Séries Temporais e Métodos de Predição de Preços de <i>Commodities</i></u>	6
<u>3.9 <i>Machine Learning</i></u>	6
<u>3.10 <i>Python</i></u>	7
<u>3.11 CRISP-DM</u>	7
<u>3.12 Análise de Componentes Principais</u>	8
<u>3.13 Regressão de Lasso</u>	8
<u>4 METODOLOGIA</u>	10
<u>4.1 Entendimento do problema</u>	10
<u>4.2 Compreensão dos dados</u>	10
<u>4.3 Preparação dos Dados</u>	13
<u>4.4 Modelagem</u>	16
<u>4.5 Avaliação</u>	17
<u>5 RESULTADOS E DISCUSSÃO</u>	17
<u>5.1 Modelo de precificação de opções</u>	17
<u>5.2 Seleção de fatores</u>	18
<u>5.3 Seleção do <i>time frame</i> e anos utilizados no treinamento do modelo</u>	20

<u>5.4 Modelagem</u>	21
<u>6 CONCLUSÃO</u>	23
<u>REFERÊNCIAS</u>	25

1 INTRODUÇÃO

Os preços de commodities negociados nas bolsas ao redor do mundo costumam ser mais voláteis que outras classes de derivativos. Essas grandes oscilações nos preços podem ter impactos consideráveis nas receitas de empresas mineradoras e nos custos dos consumidores dessas commodities. O manejo de modelos econométricos que consigam prever, com certa destreza, os preços dessas matérias primas é essencial para a saúde financeira de uma empresa. Isso porque quedas imprevistas nos preços de commodities metálicas, como o cobre ou zinco, podem ter impactos devastadores nas empresas de mineração que são tomadores de preço, ou seja, não controlam o preço dos produtos que vendem.

Apesar da previsão dos preços ter tamanha importância para a indústria da mineração, muitos projetos consideram preços constantes ao longo do tempo. Em geral, apenas projetos mais complexos lançam mão de técnicas de análise de séries temporais, como ARMA e ARIMA, a fim de prever as oscilações nos preços de commodities, mas estas não são robustas o suficiente e não apresentam precisão satisfatória em muitos casos.

Existem modelos econométricos amplamente utilizados na previsão de preços de ativos em relação a exposição de fatores de risco como o *Capital Asset Pricing Model*, que trata de prever o retorno de um ativo, com base na volatilidade deste com a relação à do mercado, e o *Fama and French three factor model* que funciona como uma regressão linear múltipla do preço de um ativo com base no risco de mercado, risco de tamanho e risco de valor do ativo.

Com a popularização do aprendizado de máquina, um braço da mineração de dados (data mining), novas técnicas e modelos preditivos mais robustos estão ganhando popularidade na comunidade acadêmica e na indústria. Modelos, como regressão de Lasso e Ridge, que são capazes de regularizar e selecionar variáveis, conseguem fazer previsões mais precisas e de fácil interpretação.

Assim sendo, neste projeto, pretende-se testar a hipótese de que a criação de um modelo econométrico que usa regressão de Lasso para prever o preço do cobre um mês no futuro na *London Metals Exchange* (LME) pode ser útil para mineradoras como ferramenta para auxiliar no *hedge* contra fortes oscilações nos preços do cobre na bolsa. Para isso, serão selecionados diversos índices macroeconômicos como variáveis independentes para criação do modelo seguindo a metodologia *Cross-Industry standard process for data mining* (CRISP-DM).

2 OBJETIVO

O objetivo deste trabalho de conclusão de curso consistiu em propor um modelo macroeconômico usando técnicas de aprendizado de máquina e estatística multivariada, mais precisamente regressão de *Lasso* e análise de componentes principais, visando a previsão do preço *spot* do Cobre na *London Metals Exchange*.

Dessa maneira, os objetivos específicos foram os seguintes:

- a) definir os principais fatores responsáveis pelas oscilações nos preços *spot* do cobre na LME;
- b) descobrir o modelo e os parâmetros ideais para prever o preço de commodities na LME; e
- c) fomentar a implementação de modelos de aprendizado de máquina na engenharia de minas.

3 REVISÃO BIBLIOGRAFICA

3.1 Mercados de balcão

Um mercado de balcão, ou *over-the-counter*, é um mercado descentralizado onde os agentes negociam ações, derivativos, moedas, commodities ou outros instrumentos entre si, sem a intervenção de uma bolsa de valores. Estes mercados não possuem locais físicos, são sujeitos a menos regulação e as negociações são realizadas de forma eletrônica (KRAMER, 2020). Apesar de serem menos formais e negociarem diversos títulos que não são autorizados a serem transacionados na bolsa, costumam ser bem organizados, com dealers agindo como *formadores de mercado* gerando ordens de compra (*bid*) e venda (*ask*) (DODD, 2021). Devido a essas características, esse tipo de mercado é caracterizado pela presença de riscos adicionais, como: operações menos transparentes, risco de contrapartida e risco de liquidez.

3.2 Os agentes do Mercado e Seus Objetivos

Os agentes participantes no mercado financeiro podem ser divididos em duas principais categorias, traders e investidores. Ambos operam no mercado financeiro, mas com diferentes objetivos e estratégias (KUEPPER, 2019). Traders compram e vendem ativos com janelas de tempo curtas buscando lucrar com oscilações nos preços, lançando mão de variações de oferta e demanda, análise técnica e *sentimento de mercado* para tomadas de decisão. Investidores, por outro lado, buscam operações de longo prazo e fazem escolhas com base na análise fundamentalista, dando preferência a value stocks e estratégias de buy and hold.

3.3 Sistemas de Venda

Existem três principais sistemas de vendas nos mercados financeiros (quais sejam, spot, contratos futuros e opções), cada uma com características e funções próprias. O mercado de spot, também conhecido como *physical* ou *cash market*, representa um real equilíbrio de mercado, pois o preço *spot* é o preço pelo qual um ativo pode ser comprado imediatamente. Neste mercado, ativos como moedas, *commodities* e ações são negociadas com entrega imediata (SMITH, 2021).

Por contraste, contratos futuros são instrumentos financeiros derivativos – contrato entre duas partes cujo valor está atrelado a um instrumento financeiro – que possibilita a alienação ou assunção do risco de oscilação de preço do ativo subjacente. Um contrato futuro obriga o comprador a comprar (e o vendedor a vender) o instrumento subjacente em data e a preço predeterminados. Contratos futuros permitem a especulação com alavancagem e que produtores e compradores de commodities façam hedge.

Um contrato de opção proporciona ao comprador o direito de compra ou venda de uma quantidade fixa de um ativo subjacente a um preço específico dentro de um prazo delimitado. Existem dois tipos de opções: puts e calls. Uma put, ou opção de venda, proporciona ao comprador o direito de vender e, ao vendedor, a obrigação de comprar (caso o comprador opte por vender), um ativo a um determinado preço definido em prazo delimitado de comprar o ativo subjacente. Por contraste, uma call, ou opção de compra, proporia ao comprador o direito de comprar, e ao vendedor a obrigação de vender o ativo subjacente. Opções são um instrumento extremamente versátil que permite especulação com alavancagem e hedge.

3.4 Conceito de Volatilidade e *Hedge*

A volatilidade representa a dispersão (forma com que os valores estão distribuídos em torno de uma média ou média móvel) dos retornos de um ativo. A volatilidade é habitualmente representada pelo Beta do ativo, pelo desvio padrão ou variância dos retornos do mesmo. A volatilidade está ligada a grandes oscilações no mercado, independente da direção, e é usada como representação de risco. Ativos mais voláteis são considerados mais arriscados devido à incerteza e menor previsibilidade das dimensões das variações em um curto espaço de tempo, resultando em possíveis mudanças drásticas nos preços.

Hedging é uma estratégia que busca garantir certa proteção contra movimentos adversos a uma posição, tanto comprada como vendida, no mercado financeiro e, por ser uma estratégia de controle de exposição a risco, está fortemente ligado a volatilidade. A redução de risco vem à custa de uma redução do retorno esperado (MARKOWITZ ,1952). Existem diversas formas de fazer *hedge*, a exemplo da adoção de posições contrárias em ativos semelhantes, abrindo posições em contratos futuros ou com opções. O custo associado ao *hedge* depende do instrumento financeiro escolhido, por exemplo, no caso das opções, o custo vem na forma do prêmio pago ao vendedor da opção.

3.5 Arbitragem

Arbitragem é uma estratégia de *trading* que tem como finalidade usufruir das ineficiências do mercado para obter lucros. Ocasionalmente podem ser observadas discrepâncias pequenas e momentâneas entre os preços de um mesmo instrumento financeiro – ou instrumentos semelhantes – sendo negociado em diferentes mercados (CHEN, 2021). Quando essas discrepâncias ocorrem, um trader pode criar ordens de compra em um mercado e, simultaneamente, vender em outro, a fim de lucrar com a convergência dos preços (convergência estipulada pela *market efficiency theory*) (CFI, 2021). A estratégia é considerada de baixo risco, já que é uma compra e venda simultânea. Porém, é preciso prestar atenção aos

custos operacionais, porque as divergências nos preços são bastante apertadas e podem ser completamente eliminadas por custos operacionais.

3.6 O Mercado de *Commodities*

Um mercado de *commodities* é um mercado regulado e organizado onde se facilita a compra e venda de contratos standardizados (quantidade, qualidade, preço e entrega) de matérias primárias como milho, café, soja, ouro, ferro, petróleo, etc. Atualmente existem aproximadamente cinquenta principais mercados de *commodities* no mundo que possibilitam a compra e venda de aproximadamente cem *commodities* (HAYES, 2021).

Mercados de *commodities* possibilitam a existência de mercados centralizados e líquidos, os quais providenciam um ambiente para determinação dos preços spot das *commodities* e que podem ser acessados por produtores, consumidores e *end-users* de *commodities*, que não apenas usam o mercado para compra e venda de seus produtos, mas também para fazer hedge contra oscilações nos preços. As principais formas de investir no mercado de *commodities* é por derivativos (*forward*, futuros e opções), por meio dos quais produtores e consumidores transferem o risco para especuladores que estão dispostos a assumi-los por um prêmio.

Além dos produtores e consumidores, existem outros *players* nestes mercados, principalmente em tempos de volatilidade de mercado. Especuladores, investidores e *arbitrageurs*. Especuladores e *arbitrageurs* enxergam *commodities* como uma classe de ativo alternativa, e buscam posições com contratos derivativos a fim de diversificar suas carteiras. Isso se deve ao fato de as *commodities* apresentarem mais volatilidade que classes de ativos “tradicionais” e a possibilidade de abrir posições alavancadas.

3.7 Métodos Econométricos

De acordo com a enciclopédia Britannica, econometria é a análise matemática e estatística das relações econômicas frequentemente usadas como base para previsões econômicas. A econometria é majoritariamente aplicada ao estudo das correlações entre variáveis econômicas por economistas. Modelos econométricos fazem uso da estatística inferencial para analisar dados econômicos usando técnicas como distribuição de frequências, probabilidade, distribuição de probabilidade, análises de correlação, regressão simples e múltipla e análises de séries temporais (HAYES, 2020). Dois exemplos de modelos econométricos são: o *Capital Asset Pricing Model* (CAPM), que busca descrever a relação entre o risco sistemático e o retorno esperado para um ativo; e o *Fama and French Three Factor*

Model, que complementa o CAPM adicionando *size (market cap)* e *value risk* (relacionado ao valor da empresa) à equação.

3.8 Séries Temporais e Métodos de Predição de Preços de Commodities

Uma série temporal é uma sequência de dados ou observações, qualitativas ou quantitativas, coletadas em intervalos de tempo, constantes ou não, e ordenados cronologicamente (ODSC, 2021). As séries temporais podem ser classificadas como estacionárias ou não estacionárias. As estacionárias apresentam características como média, variância e autocorreção que não variam com o tempo. As não estacionárias, por sua vez, apresentam características que variam com o tempo e certos padrões como tendência, sazonalidade, ciclicidade e componente estocástico.

Análises de séries temporais são usadas para entender como um ativo ou variável econômica varia ao longo do tempo. Os métodos mais comuns de predição de preços de *commodities* são análises de séries temporais como médias móveis, *Auto-Regressive Moving Average* (ARMA) e *Auto-Regressive Moving Average* (ARIMA).

3.9 Aprendizado de máquina

Aprendizado de máquina é um braço da ciência ou mineração de dados popularmente conhecido pela sua aplicação em modelos de recomendação usado por empresas como Netflix e Google. Algoritmos de *aprendizado de máquina* usam a capacidade computacional para busca padrões em dados (HAO, 2018). A ideia por trás destes algoritmos é construir modelos que são capazes de aprender e refinar a precisão de suas previsões por conta própria, sem necessariamente ter sido programado para isso. Este aprendizado é possível porque os modelos são treinados para identificar e lançar mão de padrões em quantidades massivas de dados e, em seguida, fazer previsões com base nesses padrões.

Algoritmos de Aprendizado de máquina podem ser classificados em três categorias: aprendizagem supervisionada, não supervisionada e aprendizado reforçado. O mais popular é a aprendizagem supervisionada, em que o modelo é alimentado com dados e seus respectivos rótulos, estes últimos são os resultados esperados para esses dados. A ideia aqui é fazer com que o modelo identifique uma relação entre os dados e os resultados. Esta categoria é popularmente usada em modelos de previsão e recomendação. Por outro lado, não supervisionada só apresenta os dados, sem rótulos para que o modelo identifique qualquer forma de padrão presente nos dados. Por fim, aprendizado reforçado são algoritmos que

aprendem por tentativa e erro. O algoritmo é recompensado quando produz resultados positivos e penalizado quando fogem do objetivo.

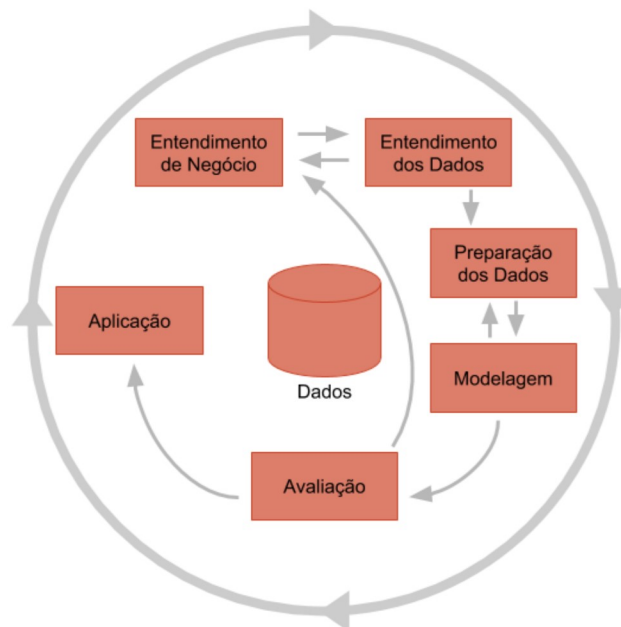
3.10 Python

Python é uma linguagem de programação de alto nível, interpretada de *script*, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte (Python, 2021). Todas essas características, junto com a sua popularidade e extensas comunidades e livros disponíveis, fazem do *Python* uma ferramenta extremamente poderosa na resolução de problemas complexos.

3.11 CRISP-DM

CRISP-DM, ou *Cross-Industry Standard Process for Data Mining*, é uma metodologia que procura apontar as seis principais fases no processo de *mineração de dados*, que pode ser descrito como o processo de extração de insights de dados com técnicas de estatística e *aprendizado de máquina*. A metodologia não só apresenta as fases, mas também a ordem na qual elas devem ser seguidas e a relação entre elas. A figura 1 destaca as fases do ciclo.

Figura 1- Ciclo de Data Mining



Fonte: <https://medium.com/@kvmoura/crisp-dm-79580b0d3ac4>

A primeira etapa no processo de mineração de dados é o entendimento do problema, onde se identifica os objetivos do trabalho, quais informações e dados estão disponíveis e quais são os riscos envolvidos. A segunda etapa é denominada de compreensão de dados. Nesta etapa é feita uma análise inicial dos dados disponíveis para entender as características dos dados disponíveis e se a procura por mais dados será necessária. A terceira etapa, a de preparação de dados, é a mais demorada de todas. Nela são definidos os formatos dos dados, derivação de novos atributos, organização e concatenação de dados.

Com os dados limpos e preparados é iniciada a quarta etapa denominada de modelagem, na qual ocorre a mineração de dados, no treinamento de modelos de aprendizado de máquina e análises estatísticas. As duas últimas etapas são a avaliação e implementação do modelo. Nelas são obtidas as métricas de avaliação do modelo e, se estas forem satisfatórias, o modelo é implementado.

3.12 Análise de Componentes Principais (PCA)

A análise de componentes principais é uma técnica amplamente utilizada na ciência de dados. Trata-se de método estatística multivariado de interdependência (ou não supervisionado), que é empregado quando se tem um número elevado de variáveis independentes a serem utilizados para treinar um modelo. Esta análise é conhecida como um método de redução de dimensionalidade, pois identifica as direções ortogonais de máxima variação. É comumente utilizada em análise exploratória dos dados como forma de identificar padrões e visualizar dados com grandes dimensões, que são difíceis de representar de forma gráfica. Também pode ser utilizada para filtrar variáveis com demasiado ruído.

O método consiste em três principais etapas: (1) a padronização, na qual se reduzem as dimensões das variáveis para garantir que todas contribuam igualmente para a análise, já que a PCA é sensível a variância das variáveis introduzidas; (2) cálculo da matriz de covariância, e (3) a computação dos autovalores e autovetor da matriz de covariância, em que são computadas novas variáveis que são ortogonais, ou independentes entre si, de forma que comporem a maior quantidade possível de informação no componente principal.

3.13 Regressão de Lasso

Regularização é uma técnica usada em aprendizado de máquina que busca generalizar um modelo com o intuito de garantir que este trabalhe bem nos dados de teste e em dados inéditos. Na prática, a regularização consiste em reduzir os erros de um modelo por intermédio de uma penalização à complexidade deste. Esta penalização é realizada mediante o

encolhimento, em outros termos, a aproximação ao valor zero dos coeficientes estimados buscando derivar um modelo menos flexível, evitando assim um possível sobre ajuste.

Modelos demasiado simples não são uteis pois não logram realizar previsões precisas (cenário de sub ajuste ou *undefitting*). Da mesma forma, modelos desmedidamente complexos são igualmente inúteis já que estes memorizam os dados de treinamento junto com seu ruído e detalhes, arruinando a generalização do modelo e resultando em previsões inexatas (cenário de sobre ajuste, ou *overfitting*).

Equação 1- Regressão de Lasso

$$\beta^{Lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^P |\beta_j| \right\}$$

Em que:

y_i : Valore real no mês i ;

β_0 : Intercepto;

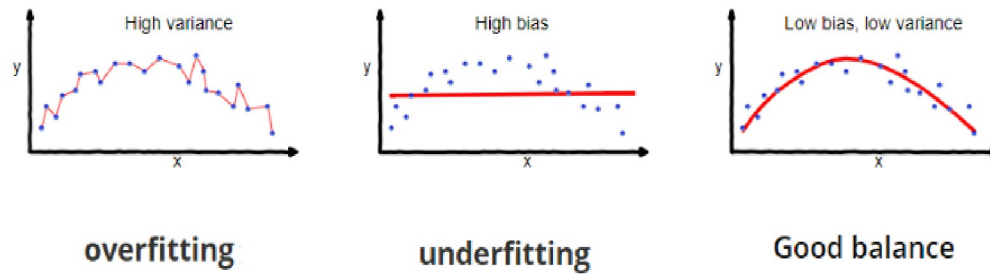
x_{ij} : Valor da variável j no mês i ;

β_j : Coeficiente da variável independente j ;

λ : Parâmetro de complexidade.

Como pode ser visto na equação 1, o método de otimização matemático utilizado para minimizar o valor de perda no ajuste de uma regressão de Lasso— *least absolute shrinkage and selection operator* –, ou regularização L1, é semelhante ao de uma regressão linear (simples ou múltipla), mas que apresenta um termo de regularização. A regularização atua elevando o viés da função por meio de uma penalização quando o viés estiver demasiado baixo (a reta traçada aos dados se encaixa muito bem com os dados de treinamento) e quando a variação estiver demasiado alta (a linha traçada não se encaixa bem nos dados de teste). A figura 2 apresenta de forma gráfica a relação entre viés e a variância e sobre ajuste e sub ajuste.

Figura 1- viés e variância



Fonte: <https://medium.com/turing-talks/turing-talks-20-regress%C3%A3o-de-ridge-e-lasso-a0fc467b5629>

A equação 2 apresenta a penalização L1 utilizada na regressão de Lasso. A tarefa desta penalização é multiplicar a soma do módulo dos coeficientes e os multiplicar por um valor de *alpha* arbitrário (o *alpha* pode ser qualquer valor entre 0 e infinito). O valor de *alpha* dependendos dados utilizados no treinamento e torna necessário o afinamento deste de forma manual, por tentativa e erro, ou de forma automatizada, lançando mão de técnicas como validação cruzada. Utilizar uma *alpha* igual a 0 resultaria na eliminação da penalização, revertendo, assim, a regressão de Lasso a uma regressão linear. Por outro lado, um valor de *alpha* muito elevado resultaria em uma reta horizontal.

Equação 2- Penalização L1

$$\lambda \sum_{j=1}^P |\beta_j|$$

O viés utilizado na regressão de Lasso tem como finalidade reduzir a variação do modelo. O acréscimo da penalização busca evitar o sobre ajuste que pode vir a ocorrer em modelos devido à pequena quantidade de dados de treinamento ou elevado número de variáveis independentes. O resultado obtido na implantação da penalização é alteração sóbria e deliberada da inclinação da reta calculada pelo modelo na direção de se obter um melhor encaixe da reta com os que ainda não foram introduzidos ao modelo, os dados de teste. Quanto maior for o parâmetro de complexidade da penalização – alfa -, menor será a inclinação da reta, podendo esta chegar a um coeficiente angular igual a zero. ‘

4 METODOLOGIA

4.1 Entendimento do problema

De acordo com a metodologia CRISP-DM, para o primeiro passo antes de se criar um modelo de aprendizado de máquina é crucial ter um discernimento do problema que se pretende modelar. Para este trabalho de conclusão de curso, foi necessário buscar entender os principais motivos responsáveis por alterar o equilíbrio entre a oferta e demanda de cobre, o que resulta em uma variação no seu preço *spot* na *London Metals Exchange*.

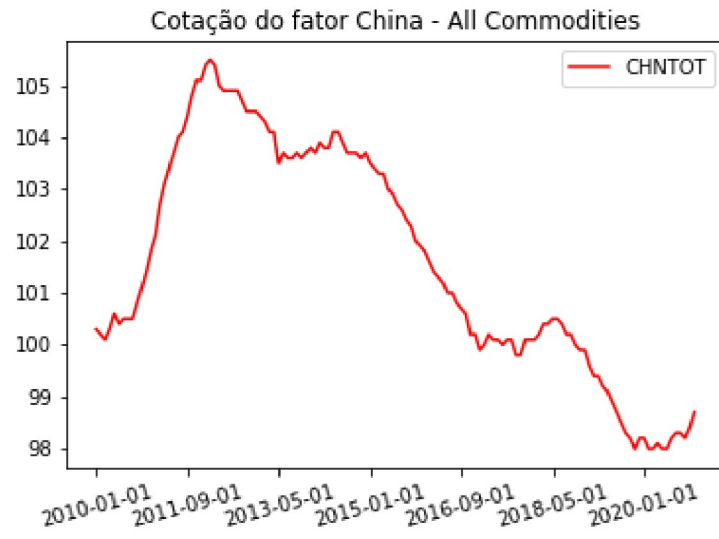
4.2 Compreensão dos dados

Este trabalho utilizou 42 fatores na criação modelo, dentre eles estão: Índices macroeconômicos, *Exchange traded funds*, índices setoriais, informações de estoques de metais etc. Todos os dados foram obtidos como arquivos individuais no formato CSV (*Comma-separated values*) de séries históricas de sites de informação financeira.

A fim de realizar análises descritivas dos fatores selecionados, lançou-se mão de um ambiente interativo de desenvolvimento *open-source* flexível, que suporta uma variedade de linguagens de programação e que é popularmente utilizado para desenvolver *workflows* na ciência de dados, computação científica e criação de modelos de aprendizado de máquina (Project Jupyter, 2021). A linguagem de programação preferida para realização do trabalho foi *Python*.

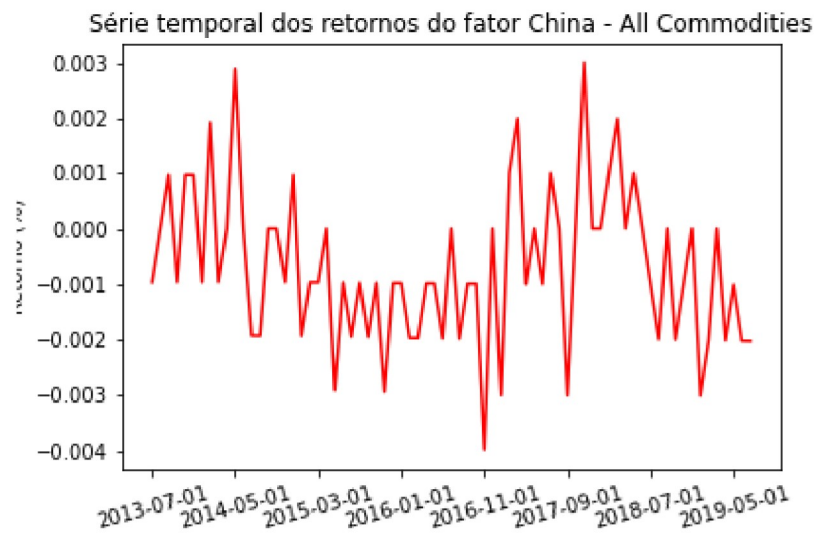
O cálculo de estatísticas descritivas dos fatores – como média, volatilidade (desvio padrão), curtose, correlação com a variável dependente e plotar as distribuições dos retornos dos fatores – foram essenciais para entender as características de cada uma das variáveis para criar uma estratégia de tratamento de dados que possibilitasse a concatenação destes em um banco de dados único que pudesse ser utilizado no treinamento modelo desejado. As figuras 3, 4 e 5 apresentam, respectivamente, a série histórica, a série histórica dos retornos e a análise de distribuição de um dos fatores selecionados, qual seja, o índice China – All commodities. A análise da distribuição dos valores é essencial, uma vez que fatores que apresentam uma distribuição normal apresentam sinais mais claros para os modelos de aprendizado de máquina.

Figura 3-Cotação do fator China - All Commodities



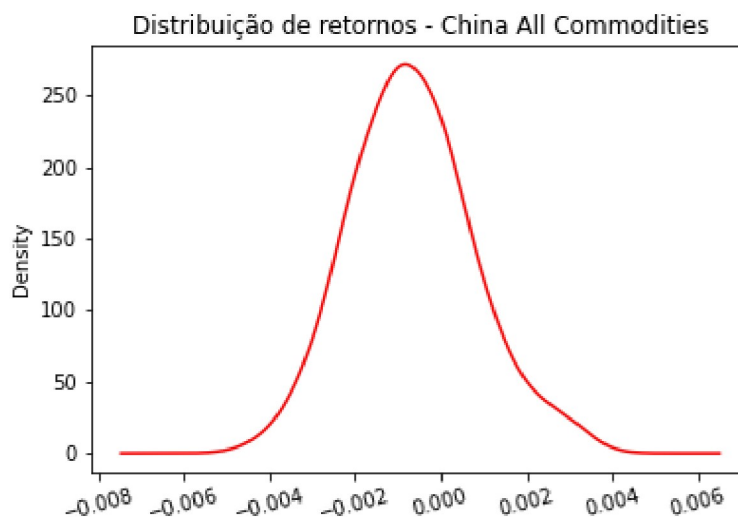
Fonte: Própria 2021

Figura 4- Série histórica dos retornos do China – All Commodities



Fonte: Própria 2021

Figura 5 - Distribuição dos Valores do fato China – All Commodities



Fonte: Própria 2021

4.3 Preparação dos Dados

Uma vez que toda fonte de dados disponibiliza suas séries históricas com uma formatação específica, foi necessária uma empreitada de tratamento de dados, popularmente conhecido como limpeza de dados. A primeira tarefa desta fase, foi verificar se os arquivos CSV possuíam informações a mais do que as requeridas. Algumas fontes disponibilizam dados referentes aos fatores – como preço de abertura, máximo e mínima, volume de negociação, etc. – que não são necessários para o modelo e que devem ser eliminados. Logo certificou-se se havia valores ausentes (valores faltando) nos dados. Neste caso, como foram usados valores mensais, não se encontraram valores ausentes devido a feriados ou imprevistos.

Em seguida, prosseguiu-se à formatação adequada dos dados. Os arquivos de cada fonte possuíam uma formatação particular, as mais evidentes foram as datas e valores numéricos. O Python utiliza pontos para os algarismos numéricos, nos quais algumas fontes utilizam vírgulas. Isso é um empecilho, pois, quando os dados possuem vírgulas, a linguagem os interpreta como texto, impossibilitando assim os cálculos. A formatação das datas é relevante, visto que estas são usadas como índices, porque os dados analisados são séries históricas. O impasse na configuração das datas resulta da forma que cada fonte as representa, onde uma utiliza o formato “dd-mm-yyyy”, considerando sempre o primeiro dia do mês, outras

usam o formato escrito resumido (Ex: Jan 2010). A tabela 1 apresenta uma fração dos dados coletados de uma das fontes após a formatação inicial

Tabela 1- Dados após primeira etapa de preparação

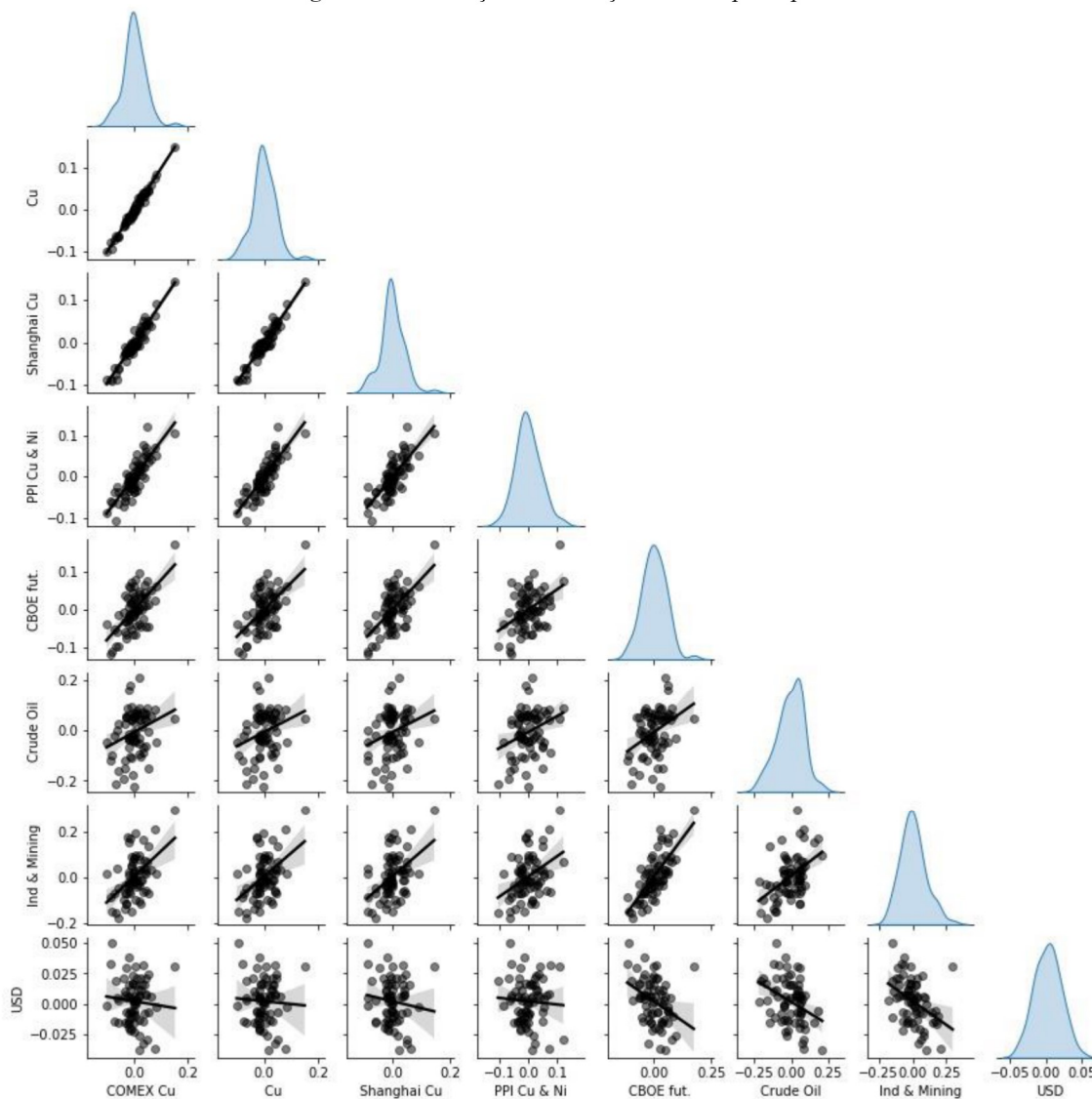
	INV - 10YB	INV - CBOE futures	INV - Crude oil	INV - DJ Ind Metals & Mining	INV - Hang Seng	INV - MSCI All- Country World Equity Index	INV - US Dollar Index
Mar 10	3,83	3,55	269,08	300,84	21239,35	307,40	81,07
Apr 10	3,66	3,34	277,97	278,55	21108,59	307,35	81,87
May 10	3,30	3,10	232,70	258,31	19765,19	277,17	86,59
Jun 10	2,94	2,94	232,49	220,83	20128,99	268,25	86,02
Jul 10	2,91	3,31	241,01	253,06	21029,81	289,75	81,54
Aug 10	2,47	3,36	221,53	244,19	20536,49	279,06	83,20
Sep 10	2,51	3,65	240,90	275,32	22358,17	305,16	78,72
Oct 10	2,60	3,73	242,93	296,10	23096,32	315,95	77,27
Nov 10	2,80	3,82	248,73	307,51	23007,99	308,38	81,19
Dec 10	3,29	4,44	269,67	358,95	23035,45	330,64	79,03

Fonte: Própria 2021

Com os dados de todos os arquivos formatados com os mesmos parâmetros, o seguinte passo foi concatenar todas as séries em um único banco de dados com base nas datas dos dados. Em seguida, realizou-se mais uma vez a eliminação dos meses para os quais havia valores ausentes, porque, ao concatenar series temporais de comprimentos diferentes, foram acrescentados valores ausentes às mais curtas. Com o banco unificado pronto, efetuou-se o cálculo da variação percentual entre os meses.

A utilização das variações mês-a-mês, ou retorno mensal, é oportuna porque estas são mais semelhantes a uma distribuição normal do que a distribuição dos valores “brutos”, isso é explicado pelo fato de muitos dos fatores utilizados apresentarem movimentos em tendências. A figura 6 apresenta um gráfico com a distribuição dos retornos dos fatores com maior correlação ao preço do cobre (diagonal), o resto do gráfico mostra a relação dos fatores entre si e com o cobre.

Figura 6: Distribuição e correlação entre as principais variáveis



Fonte: Própria 2021

A última alteração feita ao banco de dados na fase de preparação de dados foi a definição do *lag* temporal -a delonga entre uma ação econômica e a sua consequência - para os fatores. Esta etapa foi realizada de forma empírica e manual. Com o auxílio do método “.corr()” da biblioteca *Pandas* foi calculada a correlação entre a série histórica do cobre com a série do fator com diversos *lags* (1 a 5). O grosso dos fatores apresentaram uma maior correlação com um *lag* de 1 mês, motivo pelo qual, no intuito de preservar a maior quantidade de dados, foi tomada a decisão de utilizar o *lag* de um mês para todos os fatores. A tabela 2 apresenta a correlação dos *lags* do fator Dow Jones Industrial Metals and Mining com o preço do cobre.

Tabela 2: Análise de correlação para cada lag do fator Dow Jones Industrial Metals and Mining Index

	Copper Cash LME Daily Official \$ per tonne Monthly Average
Lag 1	0,3697
Lag 2	-0,1030
Lag 3	-0,0543
Lag 4	-0,0010
Lag 5	0,1108

Fonte: Fonte Própria

4.4 Modelagem

Para criação dos modelos e análises realizadas na etapa de modelagem foi empregada uma biblioteca da linguagem de *Python* própria para aplicação prática de aprendizado de máquina. O primeiro passo da criação do modelo foi a realização da análise de componentes principais. Esta etapa foi crucial para definir se e quanto os fatores deliberados para a criação do modelo logram explicar as oscilações nos retornos do preço *spot* do cobre.

Para treinar e, finalmente, testar o modelo ansiado, é boa prática elaborar uma etapa de pré-processamento composta por padronização e normalização dos dados. A normalização garante que variáveis que tenham maior dispersão não influenciem mais o modelo que as variáveis menos dispersas (o modelo apresentado utiliza o retorno dos fatores para amenizar este impacto, mas, independentemente disso, foi realizada para igualar o impacto dos fatores mais voláteis com os menos voláteis).

Em seguida, os dados foram separados em duas frações, uma parte de teste e outra de treinamento. A proporção de dados de treinamento e teste foi de 80 e 20% respectivamente. Como a única informação importante para a previsão dos preços são os fatores macroeconômicos do mês anterior, os dados de teste foram selecionados aleatoriamente utilizando a função *train_test_split()* da biblioteca Sklearn. Essa separação é crucial para evitar o sobre ajuste do modelo, pois só uma parte é usada no treinamento deste, a fim de evitar que as medidas de avaliação sejam calculadas em valores que tenham sido usados no treinamento.

A seguinte etapa foi definir os parâmetros a serem utilizados no treinamento, sendo que os principais foram: o número máximo de interações (100e5) e *alpha* (0, 00001). A última etapa consistiu em realizar o ajuste do modelo e calcular os valores previstos usando o método “fit()” do objeto Lasso criado usando a biblioteca Sklearn.

4.5 Avaliação

A biblioteca empregada na criação do modelo possui funções que, quando alimentadas com os valores previstos e valores reais (valores de teste que foram separados anteriormente ao *fit* do modelo), retornam os valores das métrica desejadas. Para este trabalho de conclusão de curso, as métricas selecionadas foram o erro médio absoluto e erro médio quadrático para avaliar os desvios dos valores previstos em relação aos valores reais, e o R-quadrado para avaliar quanto das oscilações dos retornos do cobre podem ser explicadas pelos retornos do mês anterior dos fatores. Modelos de aprendizado de máquina usam estas métricas como principais critérios de avaliação porque a finalidade destas é a *performance* das previsões em dados vistos previamente pelo modelo.

5 RESULTADOS E DISCUSSÃO

5.1 Modelo de precificação de opções

O propósito inicial deste trabalho era ajustar um modelo de precificação de opções binárias (preferencialmente o de Black-Scholes) de forma a prever mais habilmente o preço de *puts* e *calls* de cobre na London Metals Exchange. No entanto, o acesso a séries temporais de preços históricos de contratos de opções de metais confiáveis e gratuitas – ou até com preços acessíveis – mostraram-se inexistente. Tampouco foi possível ter acesso aos *options greeks* (Delta, Gama, Vega, Theta e Rho) dos contratos passados, impossibilitando assim o esforço de criar uma versão de um modelo de precificação de opções próprio para metais.

Não obstante, existem outros caminhos além das opções pelo qual a mineração pode otimizar seus processos recorrendo aos mercados financeiros. A primeira opção cogitada foi a de criar um modelo de previsão dos preços de contrato futuro para o cobre. Porém, não faz sentido econômico criar uma previsão de um contrato futuro de uma commodity, uma vez que um contrato futuro não representa um equilíbrio de mercado, e sim um equilíbrio temporário do que o mercado como um todo (*traders* comerciais, grandes especuladores e pequenos especuladores) pressupõem ser um preço justo para a commodity em um tempo futuro definido.

Por conseguinte, decidiu-se criar um modelo que busca indicar a direção e, se possível, a amplitude dos movimentos dos preços *spot* do cobre em mercados organizados. A criação de um modelo para prever o preço *spot* de uma *commodity* faz sentido econômico, visto que o preço *spot* representa, sim, um equilíbrio de mercado em um determinado tempo. Ademais, ter uma forma confiável de estimar o preço futuro de uma *commodity* permite a operação de contratos futuros. É que, se for possível saber como os preços devem estar em um determinado tempo futuro, pode-se: (i) adotar posições compradas ou vendidas em contratos futuros, dependendo da situação; ou (ii) captar recursos com opções adotando estratégias de *volatility selling*.

5.2 Seleção de fatores

Existem três principais bolsas no mundo onde se pode negociar o cobre: A *London Metals Exchange*, A *Chicago Mercantile Exchange* (CME) e a *Shanghai Future Exchange* (SHFE). Este trabalho busca prever o preço *spot* do cobre em um mês no futuro na London Metals Exchange, pois é para os contratos negociados nesta que é possível encontrar o maior número de fontes de informação livre e gratuita.

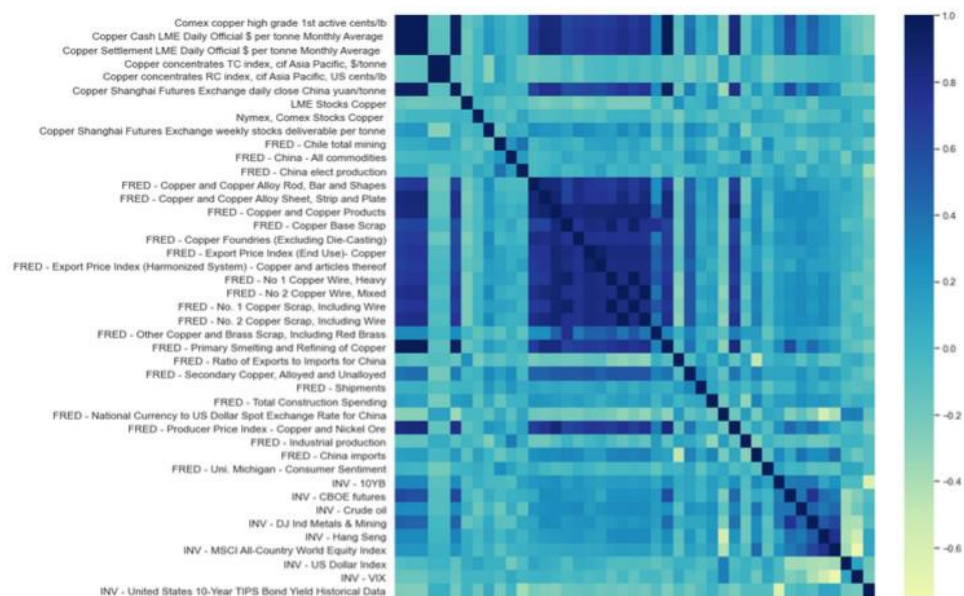
O presente trabalho buscou usar a análise fundamentalista – a qual explica as forças da oferta e da demanda por trás dos movimentos de preços no mercado – como principal guia para a seleção dos fatores macroeconômicos (variáveis independentes) utilizados no modelo. Isso porque a análise técnica pouco se importa com os motivos por trás dos movimentos, mas sim em buscar padrões nos preços passados buscando oportunidades de especulação. A análise técnica supõe que os preços descontam, ou já consideram, todas as informações disponíveis no mercado, sendo assim, uma ferramenta importantíssima na otimização da escolha do momento de entrada e/ou saída em operação no mercado.

Séries históricas de dados macroeconômicos (como sentimento do consumidor, estoques dos metais da London Metals Exchange, e relação de importação/exportação da China, entre outros) podem ser acessados gratuitamente na internet. Todavia, acesso a informações mais específicas e confiáveis, a exemplo da produção de concentrado de cobre por país, somente são encontradas em fontes pagas. Por conseguinte, este trabalho buscou usar alternativas adequadas para os fatores que não puderam ser acessados de forma gratuita, como a substituição do Li Keqiang Index – alternativa considerada menos enviesada do PIB da China – pela produção total de energia elétrica da China.

Em seguida da seleção dos principais fatores macroeconômicos disponíveis para criação de um modelo econométrico, que busca explicar e prever as oscilações do preço do cobre na LME, foi criada a matriz de correlação dos fatores que foi representada como um mapa de calor na figura 7. O mapa evidenciou que a maior parte dos fatores macroeconômicos selecionados demonstraram ter correlação média a baixa entre si, embora demonstraram ter correlações mais fortes com os preços do cobre (independente de qual bolsa).

Os fatores macroeconômicos que apresentam maior correlação com o preço do cobre na London Metals Exchange são os fatores que representam o preço do cobre em outras bolsas. Isso é de se esperar, já que a presença de *arbitrageurs* traz eficiência ao mercado, resultando no equilíbrio dos preços entre as bolsas. Fatores ligados ao preço de sucata de cobre também apresentam alta correlação ao preço do cobre, pois as siderúrgicas também usam sucata para alimentar seus fornos. Isso pode ser evidenciado pela linha mais escura na parte superior e esquerda do mapa abaixo:

Figura 7 - Mapa térmico da matriz de correlação dos fatores selecionados para criar o modelo econométrico

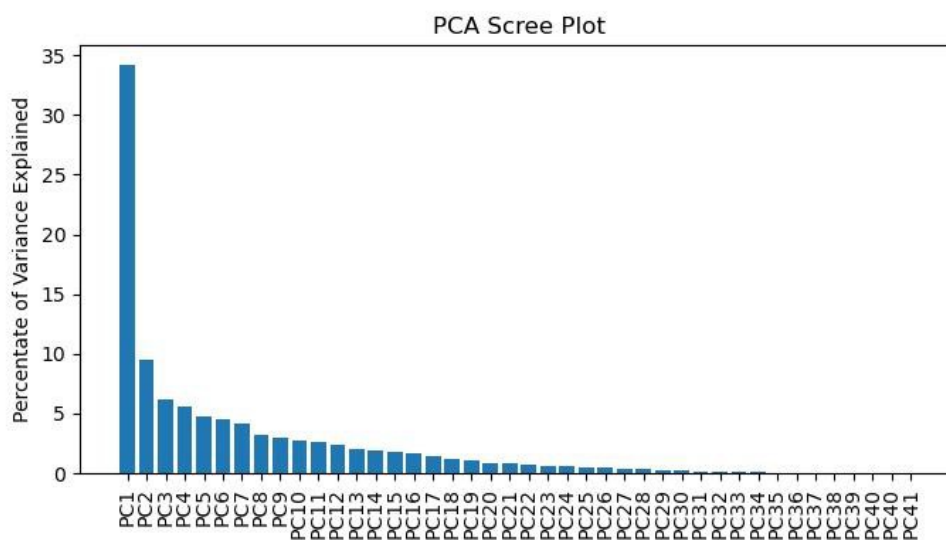


Fonte: Própria 2021

Como várias das variáveis independentes são alternativas que buscam representar os reais fatores macroeconômicos que alteram a oferta e demanda do metal no mercado, foi

necessário empregar uma grande quantidade de fatores. Por isso, em seguida a análise da matriz de correlação dos fatores macroeconômicos selecionados, realizou-se uma análise de componentes principais, a fim de assegurar que as viáveis independentes selecionadas fossem capazes de explicar a variação dos preços spot do cobre na LME. A figura 8 apresenta o gráfico de Scree da PCA realizada. Este apresenta a porcentagem das oscilações nos preços de cobre que cada componente principal logra explicar. O gráfico Scree auxilia a identificar o número ótimo de variáveis independentes. Habitualmente se selecionam as componentes principais que compõem a reta de maior inclinação pois as demais componentes contribuem menos para a previsão reduzindo assim a complexidade do modelo.

Figura 8- Análise de componentes principais



Fonte: Própria 2021

A análise de componentes principais apontou que mais de 50% das componentes principais poderiam ser eliminadas do modelo já que estas não lograram explicar em tal intensidade satisfatória as variações nos preços do cobre. Sem embargo, elegeu-se não tolher nenhuma das variáveis independentes. A PCA mostrou que é possível condensar a maior parte da informação contida nos fatores macroeconômicos em alguns poucos componentes principais, mas como o objetivo deste modelo é prever o retorno do preço do cobre, algo que requer extrema precisão para poder ser útil para fazer hedge, optou-se por preservar toda a maior quantidade de informação possível.

Considerando que as variáveis independentes foram selecionadas para tentar representar os reais fatores que não puderam ser acessados, como, por exemplo, o China Electricity que, neste caso, busca representar o Li Keqiang Index, que representa a economia chinesa (o maior consumidor de cobre do mundo e um dos maiores produtores) foi preciso manter o máximo de informação possível, já que os sinais gerados por estas para o modelo poderiam não ser tão certos. Apesar de ter determinado manter um elevado número de variáveis independentes não foi necessário tomar outra medida a fim de diminuir a complexidade -havia uma alta proporção de observações por variáveis independentes- e controlar o sobre ajuste do modelo (a regressão própria regressão de Lasso efetua uma seleção de fatores).

5.3 Seleção do *time frame* e anos utilizados no treinamento do modelo

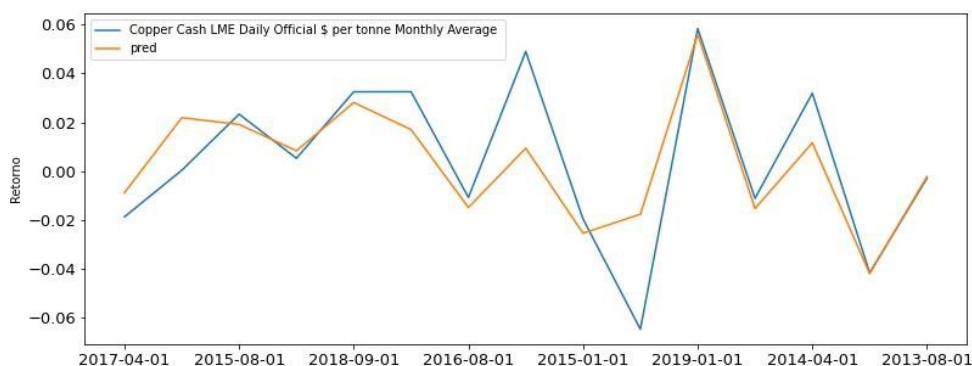
O *time frame* utilizado neste trabalho foi mensal. A escolha se deu porque a variação de preço diário do cobre no mercado financeiro apresenta demasiado ruído – oscilações nos preços devido a oferta e demanda que não são ligadas a variações nos fundamentos, mas sim a outras causas, como a especulação – o que pode ser prejudicial na criação de um modelo. Outro motivo pelo qual não foram utilizados os preços diários – nem os semanais – foi a granularidade dos dados disponíveis. A par disso, muitas informações utilizadas na criação do modelo, como, por exemplo, a produção mineira do Chile (o maior país exportador de cobre do mundo), apenas estão disponíveis de forma mensal.

Modelos supervisionados de aprendizado de máquina – como regressão de Lasso –, requerem números elevados de dados, mas, devido à dificuldade de encontrar informações confiáveis e gratuitas, foi necessário fazer um *trade-off* entre o número de fatores e a quantidade de dados utilizados no treinamento do modelo. Alguns fatores extremamente significantes na precificação da *commodity*, como o *treatment charge* e *refining charge*, só estavam disponíveis a partir do início de 2013, enquanto outros apenas estavam disponíveis referentes aos meses finais de 2019.

5.4 Modelagem

Em virtude dos preços e índices macroeconômicos apresentarem certo ruído e usar uma base de dados com séries históricas relativamente curtas (72 meses), este trabalho recorreu a regressão de Lasso – *least absolute shrinkage and selection operator* – a fim de amenizar o impacto do ruído e evitar um possível sobre ajuste do modelo aos dados. A figura 9 e as métricas da tabela 4 demonstram os valores reais (linha azul) e previstos (linha laranja). Nota-se que o modelo consegue prever a direção do movimento do preço com bastante precisão, mas não acompanha tão bem a amplitude do movimento nos dados de teste.

Figura 9- Retorno real x Retorno previsto (Alpha=0.00001)



Fonte: Própria 2021

A tabela 3 apresenta os coeficientes encontrados no ajuste do modelo para todas as variáveis independentes com a identificação da fonte dos dados (INV e FRED) . Quatro variáveis foram atribuídas coeficientes iguais a zero. Isso é de se esperar em razão de terem sido usados fatores “indiretos” além do fato da regressão de Lasso apresentar penalização (L1), o que resulta em um número reduzido de coeficientes diferentes de zero.

Tabela 3: Coeficientes

	Coef.
FRED - No 1 Copper Wire, Heavy	4,3256938
FRED - Industrial production	2,1554742
FRED - Copper and Copper Products	0,979908
FRED - Export Price Index (End Use)- Copper	0,7066812
INV - CBOE futures	0,4920909
FRED - Copper and Copper Alloy Rod, Bar and Shapes	0,4165501
INV - United States 10-Year TIPS Bond Yield Historical Data	0,3439044
Copper Settlement LME Daily Official \$ per tonne Monthly Average	0,299719
FRED - China imports	0,2532423
INV - Hang Seng	0,2471592
FRED - Copper Base Scrap	0,2387682
FRED - Shipments	0,1428772
FRED - Producer Price Index - Copper and Nickel Ore	0,1186632
FRED - Total Construction Spending	0,1157788
INV - Crude oil	0,0981298
Copper concentrates RC index, cif Asia Pacific, US cents/lb	0,0660573
FRED - Ratio of Exports to Imports for China	0,0515082
FRED - Uni. Michigan - Consumer Sentiment	0,0180346
INV - 10YB	0,00961
INV - DJ Ind Metals & Mining	0,0088555
Copper Shanghai Futures Exchange weekly stocks deliverable per tonne	0,008094
Copper Shanghai Futures Exchange daily close China yuan/tonne	0,0074648
Nymex, Comex Stocks Copper	0,0034884
LME Stocks Copper	0,0025826
FRED - No. 2 Copper Scrap, Including Wire	0
Copper concentrates TC index, cif Asia Pacific, \$/tonne	0
FRED - Primary Smelting and Refining of Copper	0
FRED - China elect production	0
FRED - Chile total mining	0,0142692
INV - VIX	0,0365359
FRED - Export Price Index (Harmonized System) - Copper and articles	0,0402211
FRED - No 2 Copper Wire, Mixed	-0,135309
FRED - Copper and Copper Alloy Sheet, Strip and Plate	0,2888911
FRED - Other Copper and Brass Scrap, Including Red Brass	-0,31691
FRED - National Currency to US Dollar Spot Exchange Rate for China	0,3424579
INV - US Dollar Index	0,3839774
FRED - Copper Foundries (Excluding Die-Casting)	0,5464078
Comex copper high grade 1st active cents/lb	0,5977857
FRED - Secondary Copper, Alloyed and Unalloyed	0,7240485
INV - MSCI All-Country World Equity Index	0,9654682
FRED - China - All commodities	3,0157343
FRED - No. 1 Copper Scrap, Including Wire	5,2910996

Fonte: Própria 2021

A tabela 4 apresenta as métricas de avaliação que foram utilizadas para avaliar modelo. Um modelo econométrico com R-quadrado de aproximadamente 40% já pode ser considerado lucrativo (WAN, 2020). Portanto, um modelo com 68,3% de R-quadrado, ou seja, que explica 68,3% da variação da variável dependente, apresenta um alto potencial de sucesso de aplicação.

Tabela 4- Métricas de Avaliação para diferentes valores de Alpha

	<u>Erro</u> <u>Médio</u> <u>Absoluto</u>	<u>Erro</u> <u>Médio</u> <u>Quadrático</u>	<u>R-</u> <u>Quadrado</u>
Regressão de Lasso Alpha = 0.001	0.02357	0.00082	0.23744
Regressão de Lasso Alpha = 0.0001	0.01717	0.00054	0.49881
Regressão de Lasso Alpha = 0.00001	0,01219	0,00034	0,68312

Fonte: Própria 2021

O erro médio absoluto e o erro médio quadrático expressam o erro médio de predição do modelo em unidades da variável de interesse. Ambas variam do 0 ao infinito e são indiferentes à direção dos erros e, quanto melhor for o modelo, mais se aproximam do 0 (Wesner, 2016). Como pode ser visto na tabela 1, o modelo apresentou erro médio absoluto e erro médio quadrático muito próximos de zero. Isso significa que os valores estimados pelo modelo ficaram próximos aos retornos reais observados. Apesar de os valores estarem muito próximos de zero, o que poderia ser interpretado como um sinal de sobre ajuste do modelo, pode-se rejeitar essa hipótese, pois as métricas foram calculadas com valores reais das séries históricas que não foram usados no *ajuste* do modelo, comprovando assim a autenticidade das métricas obtidas.

6 CONCLUSÃO

O principal objetivo desta tese – criar um modelo econométrico empregando técnicas de aprendizado de máquina – foi alcançado. No mundo das finanças, um modelo que consegue explicar 50% das oscilações nos preços de um ativo já é considerado lucrativo, e o modelo proposto neste trabalho logra explicar aproximadamente 68% das oscilações nos preços *spot* do cobre (preço que representa um real equilíbrio de mercado na principal *commodities exchange* do mundo, a London Metals Exchange). Além disso, os resultados registrados na análise de componentes principais apontam que a escolha de fatores foi adequada – apesar de muitos serem aproximações dos reais fatores por trás das variações de oferta e demanda – já que, mais

de 80% das componentes principais conseguem explicar pelo menos 1% das variações no preço do cobre no espaço de tempo usado para o treinamento e teste do modelo.

Da mesma forma, pode-se considerar que os objetivos específicos também foram alcançados. A escolha de empregar a regressão de Lasso provou-se adequada, porquanto, como foi apresentado, os valores previstos apresentam volatilidade inferior aos dos valores reais. Isso se deve à escolha de um valor de *alpha* muito próximo de 0 (0.00001).

Além disso, a presente dissertação apresentou um modelo que se mostra preciso nos dados de teste e que utiliza dados disponíveis na internet de forma gratuita que podem ser utilizados por qualquer empresa ou pessoa para fazer *hedge*, maximizar lucros operacionais e até como forma de captação de recursos que será disponibilizado de forma gratuita.

REFERÊNCIAS

ASITDUBEY, V. **An Introduction to Time-Series Analysis**. Disponível em: <<https://medium.com/analytics-vidhya/an-introduction-to-time-series-analysis-2a12d3702299>>. Acesso em: 28 de março de 2021

COMMODITY.COM. **Commodity Exchanges: The Ultimate Guide To How They Work & Why You Should Care**. Disponível em: <<https://commodity.com/trading/exchanges/>>. Acesso em: 28 de março de 2021

ENCICLOPEDIA BRITANICA. **Econometrics**. Disponível em: <https://www.britannica.com/topic/econometrics-economic-analysis>. Acesso em: 30 de março de 2021

DODD, E, **Markets: Exchange or Over the Counter**, 2021. Disponível em: <<https://www.imf.org/external/pubs/ft/fandd/basics/60-markets.htm>>, Acesso em: 12/03/2021

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **Organization for Economic Co-operation and Development, Production: Mining: Total mining: Total for Chile [CHLPRMITO01IXOBSAM]**. Disponível em: <<https://fred.stlouisfed.org/series/CHLPRMITO01IXOBSAM>>, Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **Organization for Economic Co-operation and Development, Production: Energy: Electricity: Total for China [CHNPRENEL01MLM]**. Disponível em: <<https://fred.stlouisfed.org/series/CHNPRENEL01MLM>> Acesso em: 07 Abril 2020
 FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: Metals and Metal Products: Copper and Copper Alloy Rod, Bar and Shapes [WPU10250238]**. Disponível em: <<https://fred.stlouisfed.org/series/WPU10250238>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: Metals and Metal Products: Copper and Copper Alloy Sheet, Strip and Plate [WPU10250237]**. Disponível em: <<https://fred.stlouisfed.org/series/WPU10250237>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: Special Indexes: Copper and Copper**

Products [WPUSI019011]. Disponível em: < <https://fred.stlouisfed.org/series/WPUSI019011>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: Metals and Metal Products: Copper Base Scrap [WPU102301].** Disponível em: < <https://fred.stlouisfed.org/series/WPU102301>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Industry: Other Nonferrous Metal Foundries, Except Die-Casting: Copper Foundries (Excluding Die-Casting) [PCU3315293315295].** Disponível em: < <https://fred.stlouisfed.org/series/PCU3315293315295>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Export Price Index (End Use): Copper [IQ12210], retrieved from FRED, Federal Reserve Bank of St. Louis.** Disponível em: < <https://fred.stlouisfed.org/series/IQ12210>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Export Price Index (Harmonized System): Copper and Articles Thereof [ID74].** Disponível em: < <https://fred.stlouisfed.org/series/ID74>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Industry: Material Recyclers: No 1 Copper Wire, Heavy [PCU429930429930211].** Disponível em: < <https://fred.stlouisfed.org/series/PCU429930429930211>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: Metals and Metal Products: No. 1 Copper Scrap, Including Wire [WPU10230101].** Disponível em: < <https://fred.stlouisfed.org/series/WPU10230101>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: Metals and Metal Products: No. 2 Copper Scrap, Including Wire [WPU10230102].** Disponível em: < <https://fred.stlouisfed.org/series/WPU10230102>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Industry: Material Recyclers: Other Copper and Brass Scrap, Including Red Brass [PCU429930429930214]**. Disponível em: < <https://fred.stlouisfed.org/series/PCU429930429930214>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Industry: Nonferrous Metal (Except Aluminum) Smelting and Refining: Primary Smelting and Refining of Copper [PCU3314103314101]**. Disponível em: < <https://fred.stlouisfed.org/series/PCU3314103314101>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **Organization for Economic Co-operation and Development, Ratio of Exports to Imports for China [XTEITT01CNM156S]**. Disponível em: < <https://fred.stlouisfed.org/series/XTEITT01CNM156S>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: Metals and Metal Products: Secondary Copper, Alloyed and Unalloyed [WPU102403]**. Disponível em: < <https://fred.stlouisfed.org/series/WPU102403>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **Cass Information Systems, Inc., Cass Freight Index: Shipments [FRGSHPUSM649NCIS]**. Disponível em: < <https://fred.stlouisfed.org/series/FRGSHPUSM649NCIS>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Census Bureau, Total Construction Spending [TTLCONS]**. Disponível em: < <https://fred.stlouisfed.org/series/TTLCONS>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **Organization for Economic Co-operation and Development, National Currency to US Dollar Spot Exchange Rate for China [CCUSSP02CNM650N]**. Disponível em: < <https://fred.stlouisfed.org/series/CCUSSP02CNM650N>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **U.S. Bureau of Labor Statistics, Producer Price Index by Industry: Copper, Nickel, Lead and Zinc Mining: Copper Ore and Nickel Ore [PCU2122302122304]**. Disponível em: < <https://fred.stlouisfed.org/series/PCU2122302122304>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **Board of Governors of the Federal Reserve System (US), Industrial Production: Total Index [INDPRO]**. Disponível em: < <https://fred.stlouisfed.org/series/INDPRO>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **Organization for Economic Co-operation and Development, Imports: Value Goods for China [XTIMVA01CNM664S]**. Disponível em: < <https://fred.stlouisfed.org/series/XTIMVA01CNM664S>>. Acesso em: 07 Abril 2020

FRED, FEDERAL RESERVE BANK OF ST. LOUIS. **University of Michigan, University of Michigan: Consumer Sentiment [UMCSENT]**. Disponível em: < <https://fred.stlouisfed.org/series/UMCSENT>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **CBOE Volatility Index (VIX)**. Disponível em: < <https://www.investing.com/indices/volatility-s-p-500-historical-data>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **United States 10-Year Bond Yield**. Disponível em: < <https://www.investing.com/rates-bonds/u.s.-10-year-bond-yield>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **Copper Contract COMEX**. Disponível em: < <https://www.investing.com/commodities/copper>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **Crude oil contract**. Disponível em: < <https://www.investing.com/commodities/crude-oil>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **Dow Jones Industrial Metals & Mining (DJUSIM)**. Disponível em: < <https://www.investing.com/indices/dj-industrial-metals---mining>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **Hang Seng (HSI)**. Disponível em: < <https://www.investing.com/indices/hang-sen-40>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **iShares MSCI All-Country World Equity Index (MIWD00000PUS)**. Disponível em: < <https://www.investing.com/indices/msci-world-stock>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **US Dollar Index (DXY)**. Disponível em: < <https://www.investing.com/currencies/us-dollar-index>>. Acesso em: 07 Abril 2020

FUSION MEDIA LIMITED. **United States 10-Year TIPS Bond Yield**. Disponível em: < <https://www.investing.com/rates-bonds/united-states-10-year-tips> >. Acesso em: 07 Abril 2020

MCKINNEY, W. **Data structures for statistical computing in python**. Proceedings of the 9th Python in Science Conference, Volume 445, 2010.

PEDREGOSA *et al.*, **Scikit-learn: Machine Learning in Python**, JMLR 12, p. 2825-2830, 2011.

PROJECT JUPYTER. **JupyterLab Documentation**. 2018. Disponível em: < <https://jupyterlab.readthedocs.io/en/latest/> >. Acesso em: 12 janeiro 2021.

SMITH, T, **Spot Market**, 2021. Disponível em: < <https://www.investopedia.com/terms/s/spotmarket.asp>>, Acesso em: 12/03/2021

PASSARELLI, D. **Trading Option Greeks**. Bloomberg Press, Nova York, 2008, n.1 p.xiii-xv, Acesso em: 20/03/2021

PYTHON. **Python Docs**. 2018. Disponível em: < <https://www.python.org/doc/>>. Acesso em: 12 janeiro 2021.

KUEPPER, E, **The Roles of Traders and Investors**, 2019. Disponível em: < https://www.investopedia.com/articles/basics/07/trading_investing.asp>, Acesso em: 12/03/2021