**Aluno(a): Pablo Diego de Albuquerque Pereira**

**Orientador(a): Diego Pedroso dos Santos**

**Curso: MBA em Data Science e Analytics**

**Machine Learning como Motor da Evolução em Estratégias Pair Trading no Mercado Financeiro Brasileiro.**

**Introdução**

Métodos quantitativos de “trading” já vem sendo muito utilizados, principalmente pelas grandes instituições no mercado financeiro. “Pair Trading”, também conhecido como “Long and Short” ou negociação em pares, é uma operação de arbitragem estatística. Essa técnica, utilizada na indústria financeira desde a década de 80, busca obter retornos através da identificação e exploração de ineficiências relativas de preço entre ativos correlacionados (Caldeira, 2013). A lógica fundamental é que, embora os preços dos ativos possam divergir em um dado momento, eles tendem a restabelecer a relação histórica de correlação. É um dos modelos mais populares de “trading” quantitativo, pois permite que os investidores possam lucrar tanto em mercados de alta quanto de baixa, através da compra (“long”) de ativos relativamente subvalorizados e venda (“short”) de ativos relativamente supervalorizados, sendo assim declarado como uma estratégia neutra (Ehrman, 2006), visto que também há possibilidade de negociação em pares direcionais – o que, contudo, não constitui o foco desta pesquisa.

Dentro dessa estratégia, a escolha dos ativos que serão utilizados é parte central do processo de “trading”. Vários métodos vêm sendo utilizados para identificar esses pares, sendo alguns dos mais populares: cointegração e distância. O método de cointegração, proposto por Vidyamurthy (2004), consiste em identificar pares de ativos que possuem possibilidade de cointegração. O método de distância (Gatev et al., 2006) é bem mais simples e popular no mercado, se comparado ao meio acadêmico, e consiste em encontrar ativos com alta correlação, indexar um pelo outro e buscar uma operação de retorno a média caso haja uma distorção. Apesar do grande esforço estatístico nessas abordagens, ainda ocorre muita dificuldade em identificar padrões complexos, além de uma constante necessidade de adaptação e, inclusive, subjetividade em alguns casos.

“Machine learning” (ML) tem se tornado um grande motor de mudanças nas finanças como um todo. As atuações mais divulgadas estão na área de avaliação de crédito e detecção de fraudes, onde as fintechs tiveram grandes investimentos mundiais e, no Brasil, tivemos a criação de empresas como Nubank e Creditas. Além disso, pode ser utilizado para previsão de preços de diversos ativos, incluindo imóveis e seu potencial valor de locação, que pode ser visto em empresas como a Loft e o QuintoAndar. Seu uso pode ser feito em qualquer segmento como, por exemplo, em políticas públicas, ou no combate à evasão escolar. Essas e diversas outras aplicações mostram a incrível seara que o “machine learning” pode atuar, e o quanto ele pode ser benéfico para a sociedade (Favero e Belfiore, 2024).

A sofisticação dos modelos de “trading” tem aumentado com o avanço da tecnologia e da ciência de dados. O uso de “machine learning”, por exemplo, tem sido foco de muitos estudos e até práticas de mercado para identificação dessas oportunidades (Lópes de Prado, 2018). Isso acontece porque, diferente dos modelos tradicionais de análise, ele também pode capturar relações não lineares de ativos, lidar melhor com alterações de fluxo do mercado financeiro de maneira mais assertiva e também lidar com grandes volumes de dados para identificar brechas que não seriam facilmente identificadas por modelos tradicionais (Zong, 2021). O uso de modelos não supervisionados, como os de agrupamento, pode ser uma alternativa para identificação desses pares, por exemplo. Assim como os modelos supervisionados, como redes neurais, podem ser utilizados para prever a relação entre os ativos.

**Objetivo**

O objetivo deste projeto é desenvolver e avaliar um modelo de “machine learning” não supervisionado para a seleção de ativos para operações de negociação em pares. Espera-se, com isso, otimizar o desempenho dessas estratégias no mercado financeiro brasileiro e superar os modelos tradicionais de cointegração e correlação.

**Metodologia ou Material e Métodos**

Esta pesquisa utilizará uma abordagem quantitativa para desenvolver e validar um modelo de “machine learning” para seleção dinâmica de ativos em estratégia de negociação em pares, utilizando dados históricos do mercado financeiro brasileiro. O estudo será dividido em 3 etapas principais: seleção de pares, modelagem e testes.

A etapa de seleção de pares se iniciará com a coleta de dados históricos diários de ativos da B3 por meio de fontes confiáveis, como a api do Metatrader, com um período mínimo de 5 anos. Serão buscadas informações de abertura, fechamento, máxima e mínima de ações e fundos listados em bolsa, junto com indicadores técnicos, como índice de força relativa, médias móveis, volatilidade histórica, entre outros. O pré-processamento incluirá a filtragem, para garantir que os ativos possuam liquidez suficiente para operações de negociação em pares.

A Etapa de modelagem consiste em usar técnicas de “machine learning” não supervisionados para identificar pares de ativos com potencial operacional. Serão utilizados algoritmos de “clustering”, como K-means, que agrupará os ativos em um número predefinido de “clusters”, ou agrupamentos, com base na similaridade dos seus perfis de indicadores técnicos históricos, buscando identificar grupos de ativos com padrões de comportamento semelhantes. Também será utilizado o DBSCAN para agrupar os ativos em “clusters”, com base na densidade de suas características de indicadores técnicos, com objetivo de identificar grupos de ativismo com forte interconexão e, potencialmente, remover ativos que não se encaixam claramente em nenhum grupo. Em seguida, será aplicado um algoritmo de “reinforcement learning” para otimizar a seleção de pares formados a partir dos agrupamentos, aprendendo a escolher os pares com maior probabilidade de gerar retornos positivos na estratégia de “pair trading”, considerando as relações entre os ativos previamente definidas pelo agrupamento.

A etapa de testes tem como objetivo medir o desempenho do modelo desenvolvido em relação a estratégias tradicionais, como a de cointegração. A realização de “backtests” será conduzida com dados históricos para avaliar a rentabilidade, risco e robustez em diferentes cenários. A avaliação da eficácia ocorrerá por meio da comparação dos resultados com estratégias de cointegração, para avaliar a eficácia. Esses testes serão feitos utilizando o “framework” backtrader, considerando custos de transação e “slippage”. As métricas de avaliação incluem “sharpe”, retorno ajustado ao risco, taxa de acerto, “drawdown” máximo e “sortino”.

**Resultados Esperados**

Espera-se que o modelo de “machine learning” seja capaz de agrupar os ativos de maneira eficiente e, além disso, selecionar pares de ativos com potencial de operação de “pair trading”. Espera-se que os resultados do modelo demonstrem desempenho igual ou superior em relação a estratégias tradicionais de cointegração e correlação

.

**Cronograma de Atividades**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Atividades planejadas** | **Mês** | | | | | | | | | |
| **Mar** | **Abr** | **Mai** | **Jun** | **Jul** | **Ago** | **Set** | **Out** | **Nov** | **Dez** |
| Projeto de pesquisa | X | X |  |  |  |  |  |  |  |  |
| Submissão do projeto ao PECEGE |  | X |  |  |  |  |  |  |  |  |
| Revisão bibliográfica aprofundada | X | X | X | X | X | X | X |  |  |  |
| Coleta e tratamento dos dados |  | X |  |  |  |  |  |  |  |  |
| “Feature engineering” e ”selection” |  |  | X |  |  |  |  |  |  |  |
| Treinamento do modelo de ML |  |  | X | X |  |  |  |  |  |  |
| Treinamento e ajuste dos modelos |  |  |  | X |  |  |  |  |  |  |
| “Backtesting” do modelo |  |  |  | X |  |  |  |  |  |  |
| Resultados Preliminares |  |  |  | X |  |  |  |  |  |  |
| Ajuste e Escrita do TCC |  |  |  | X | X | X |  |  |  |  |
| Redação e Revisão do TCC |  |  |  |  |  |  | X |  |  |  |
| Entrega do TCC e Defesa |  |  |  |  |  |  | X |  |  |  |
| Preparação da apresentação |  |  |  |  |  |  |  | X | X |  |
| Defesa do TCC |  |  |  |  |  |  |  |  | X | X |

**Referências**

Favero, L.P.; Belfiore, P. 2024. Manual de Análise de Dados: estatística e Machine Learning com EXCEL®, SPSS®, STATA®, R® e Python®. 2ed. LTC, Rio de Janeiro, RJ, Brasil.

López de Prado, M. (2018). Advances in Financial Machine Learning. Wiley.

Vidyamurthy, G. (2004). Pairs Trading: Quantitative Methods and Analysis. Wiley.

Whistler, M. (2004). Trading Pairs: Capturing Profits and Hedging Risk with Statistical Arbitrage Strategies. Wiley.

Ehrman, D. (2006). The Handbook of Pairs Trading: Strategies Using Equities, Options, and Futures. Wiley.

Hull, J. C. (2009). Options, Futures, and Other Derivatives. Pearson Education.

Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative value arbitrage rule. Review of Financial Studies, 19(3), 797-827.

Engle, R. F., Granger, C. W. (1987). Cointegration and error correction: Representation, estimation, and testing. Econometrica: Journal of the Econometric Society, 251-276.

Avellaneda, M., Lee, J. H. (2010). Statistical Arbitrage in the US Equities Market. Quantitative Finance, 10(7), 761-782.

Xie, W. e Wu, Y. (2013). Copula-based pairs trading strategy. SSRN Electronic Journal.

Do. B., Faff, R., Hamza, K. (2006). A new approach to modeling and estimation for pairs trading. Journal of Derivatives, 13(1), 66-87.

Gu, S., Kelly, B., Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. The Review of Financial Studies, 33(5), 2223-2273.

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc.

Sarmento, S. M., Horta, N. (2020). A Machine Learning based Pairs Trading Investment Strategy. Springer.

Caldeira, J. F. (2013). Arbitragem estatística, estratégia long-short, pairs tradingm abordagem com cointegração aplicada ao mercado brasileiro. Dissertação de Mestrado, Universidade Federal de Minas Gerais.

Huck, N., & Afawubo, K. (2014). Pairs trading and selection methods: is cointegration superior? Applied Economics, 47(6), 599–613.

Zong, X. (2021) Machine learning in stock indices trading and pairs trading. PhD thesis