

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016
PRÓ-REITORIA DE PESQUISA E POS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXVIII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS **SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2024**

Previsão das movimentações de ativos financeiros usando ensemble learning para implementação de um portfólio descorrelacionado

Gabriel Costa Baptista¹, Carlos Alberto Rodrigues²

1. Bolsista PROBIC/UEFS, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana,
email: gabriel.costa.baptista1@gmail.com
2. Orientador, Departamento de Exatas, Universidade Estadual de Feira de Santana, e-mail: carod@uefs.br

PALAVRAS-CHAVE: Machine Learning, Finanças Computacionais, Otimização

1. INTRODUÇÃO

Estudos indicam que investidores que operam por conta própria no mercado de ações tendem a obter resultados inferiores ao índice do mercado, levando muitos a abandonar essa modalidade de investimento (Barber and Odean 2000). Estratégias de investimentos, sejam elas automáticas ou manuais, estão sujeitas a períodos prolongados de perdas, conhecidos como drawdown ou redução de capital. A diversificação de ativos em um portfólio geralmente contribui para mitigar essas sequências de perdas (Chekhlov et al. 2005).

Várias técnicas de ensemble têm sido aplicadas para resolver problemas no mercado financeiro. Para entender como essa estratégia vem sendo utilizada, diversos trabalhos foram analisados. Haykin (Haykin 1999) apresenta os fundamentos matemáticos do ensemble, essenciais para a compreensão da teoria. Hastie, Tibshirani e Friedman (Hastie et al. 2009) fornecem uma abordagem teórica detalhada sobre métodos como bagging, random forest e boosting. Por outro lado, Kyriakides e Margaritis (Kyriakides and Margaritis 2019) e Liu (Liu2019) oferecem uma visão prática do ensemble learning utilizando Python, com exemplos e aplicações no mercado financeiro, além do método walk forward.

Este trabalho explora a implementação de técnicas de ensemble learning com o objetivo de realizar negociações em um portfólio composto por ativos descorrelacionados. Com base nos princípios da Teoria Moderna do Portfólio (TMP), diversos ativos do mercado de futuros brasileiro foram analisados, buscando-se aqueles com baixa correlação entre si. O objetivo dessa seleção foi mitigar os riscos associados às estratégias de negociação.

Foram aplicados os algoritmos de *ensemble learning bagging, stacking, voting e*

boosting aos dados históricos dos ativos, utilizando a biblioteca Scikit-learn em Python para prever suas movimentações. Para aumentar a robustez das previsões, adotou-se o método *walk-forward*, que permite que os modelos sejam continuamente atualizados com novos dados, adaptando-se às mudanças do mercado.

A estratégia de negociação foi desenvolvida em duas abordagens: a primeira utiliza apenas indicadores técnicos, sem considerar as previsões geradas pelos modelos de ensemble; a segunda incorpora as previsões baseadas nos indicadores de médias móveis geradas pelos modelos de ensemble. A análise dos resultados busca determinar se a inclusão dessas previsões melhora o desempenho das estratégias em comparação com a abordagem puramente técnica, avaliando a viabilidade da metodologia proposta.

2. MATERIAIS E MÉTODOS

As séries históricas utilizadas neste estudo foram obtidas pela corretora XP Investimentos, através da plataforma Metatrader 5, e incluem os seguintes ativos do mercado de futuros brasileiro: milho (CCM\$N), taxa interfinanceira (DI1\$N), mini-dólar (WDO\$N) e mini-índice (WIN\$N). O mercado de futuros brasileiro foi escolhido pela alta liquidez e diversidade de ativos, o que facilita a obtenção de ativos descorrelacionados e alavancagem.

O *Ensemble learning* é uma técnica de aprendizado de máquina que combina vários modelos de base para formar um modelo mais robusto. O princípio subjacente é que, ao reunir modelos diferentes, suas previsões combinadas podem superar o desempenho de um único modelo, melhorando a precisão e a generalização das previsões. Essa abordagem ajuda a reduzir a variância (overfitting), o viés (underfitting), e a aumentar a estabilidade do modelo.

Neste trabalho, utilizaram-se os métodos Stacking, Voting, Bagging e Boosting para a previsão dos retornos. Stacking e Voting usam individualmente classificadores treinados (base learners) para o aprendizado, enquanto o Bagging faz uso de reamostragem (bootstrapping) para o treinamento dos base learners e combina suas previsões usando o Voting. A motivação por trás do Bagging é produzir diversos base learners através da diversificação dos conjuntos de treinamento. O Boosting, por outro lado, constroi sequencialmente os base learners, corrigindo os erros dos modelos anteriores ao dar mais peso aos exemplos mal classificados em cada iteração.

Para validar os modelos, foi utilizado o método *walk-forward* (WF), que divide os dados em múltiplas janelas temporais. A cada iteração, o modelo é treinado em uma janela de dados passados e testado em dados recentes ainda não vistos. A janela de treinamento é então movida para incluir novos dados, e o modelo é ajustado. Neste estudo, foi adotada uma janela móvel de 30 candles, com cada candle representando 15 minutos.

Para mensurar a performance das previsões, foram selecionadas 4 métricas comumente utilizadas para este propósito: erro percentual absoluto médio (MAPE), erro

médio absoluto (MAE), erro quadrático médio (MSE) e raiz do erro quadrático médio (RMSE). Já para mensurar o desempenho ajustado ao risco das estratégias de negociação, foi selecionada a métrica Sharpe Ratio.

Para avaliar a eficácia das estratégias de negociação, com e sem previsões, foi realizado um backtest nos dados históricos dos ativos do portfólio. Cada ativo foi simulado independentemente, somando os resultados para calcular o rendimento total. As simulações foram avaliadas com métricas de precisão e desempenho, repetindo o processo para cada abordagem de ensemble, permitindo comparar as diferentes estratégias.

3. RESULTADOS E DISCUSSÃO

Os ativos escolhidos para o portfólio apresentam graus variados de correlação, buscando sempre se manter próximo de 0. Essa descorrelação é benéfica no contexto da TMP de Harry Markowitz, uma vez que a inclusão de ativos menos correlacionados tende a reduzir a volatilidade total do portfólio e a maximizar o retorno ajustado ao risco. Portanto, a seleção dos ativos WIN\$N, WDO\$N, CCM\$N e DI1\$N atende ao objetivo de compor um portfólio diversificado, com potencial para mitigar os riscos de mercado em estratégias de negociação.

O desempenho dos modelos de *ensemble* foi avaliado com base em métricas citadas anteriormente. Essas métricas foram calculadas para os quatro ativos (CCM\$N, DI1\$N, WDO\$N e WIN\$N) utilizando as quatro técnicas de ensemble learning: Boosting, Bagging, Stacking e Voting. A análise dos resultados com as técnicas de ensemble mostra desempenho similar entre os modelos, com pequenas variações nas métricas de erro. Embora o Boosting tenha se destacado ligeiramente em alguns casos, as diferenças não foram estatisticamente significativas. Isso está de acordo com a literatura, que sugere que modelos de ensemble costumam ter resultados comparáveis em séries temporais financeiras. Os ativos analisados apresentaram variações nas métricas devido a características próprias, como volatilidade e liquidez.

As tabelas a seguir comparam o desempenho das estratégias de trading sem e com os métodos de ensemble para os ativos. A análise dos dados revela que o uso das previsões varia pouco os resultados do Sharpe obtido. Sem previsão, os métodos de ensemble mostram diferenças de desempenho menores. A construção de uma estratégia que integre previsões de maneira eficiente representa um desafio significativo, dado que implica não apenas a seleção do modelo adequado, mas também a interpretação e a incorporação apropriada das previsões no processo decisório. A eficácia de tais estratégias pode ser amplamente influenciada pela natureza dos dados coletados, os quais podem apresentar variações que favoreçam o algoritmo utilizado, causando um overfitting.

Tabela 1. Sharpe Ratio das negociações sem predição

Ativo	Boosting	Bagging	Stacking	Voting
CCM\$N	0.16887	-0.55471	0.16887	0.1515
DI1\$N	0.20358	-2.01652	0.20358	0.14141
WDO\$N	-0.06326	-1.50728	-0.06326	-0.04377
WIN\$N	0.01214	-2.30324	0.01214	-0.16399

Tabela 2. Sharpe ratio das negociações sem predição

Ativo	Boosting	Bagging	Stacking	Voting
CCM\$N	0.16887	-0.5544	0.18959	0.1328
DI1\$N	0.14337	-2.08471	0.16929	0.12455
WDO\$N	-0.05668	-1.49234	-0.05689	-0.04348
WIN\$N	0.01145	-2.29579	0.00662	-0.14553

4. CONCLUSÃO

Este trabalho demonstrou a eficácia dos métodos de *ensemble* na previsão das movimentações dos ativos financeiros. Embora a incorporação dessas predições em estratégias possa ocasionar variações sutis no desempenho das negociações, a integração dos métodos de *ensemble* revela um potencial significativo para a otimização de portfolios, uma vez que os resultados obtidos indicam que esses métodos conseguem prever os valores dos ativos com uma precisão satisfatória, considerando suas características específicas e potenciais limitações.

5. REFERÊNCIAS

- BARBER, B. AND ODEAN, T. 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. SSRN. <https://ssrn.com/abstract=219228>.
- CAVALCANTE, R. E. A. 2016. Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, 55:194-211.
- CHEKHOV, A., URYASEV, S., AND ZABARANKIN, m. 2005. Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance*, 8(1):13-58.
- HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer.
- HAYKIN, S. S. 1999. *Neural networks: a comprehensive foundation*. Pearson, 2nd edition.
- KYRIAKIDES, G. AND MARGARITIS, K. 2019. *Hands-On Ensemble Learning with Python: Build highly optimized ensemble machine learning models using scikit-learn and Keras*. Packt Publishing.
- LIU, Y. 2019. *Python Machine Learning By Example*. Packt Publishing.
- MARKOWITZ, H. 1952. Portfolio selection. *The Journal of Finance*, 7(1):77-91.