

Sistema de tomada de decisão no mercado de ações utilizando LGBM

Thiago R. S. Santos* Oswaldo L. V. Costa**

* Departamento de Engenharia de Telecomunicações e Controle, Universidade de São Paulo, SP (e-mail: thiago.ssantos@usp.br).

** Departamento de Engenharia de Telecomunicações e Controle, Universidade de São Paulo, SP (e-mail: oswaldo@lac.usp.br)

Abstract:

This work focuses on the study of the application of a machine learning model known as LGBM (Light Gradient Boosting Machine) in classifying the best buying moments of an asset in the Brazilian stock market, which works as a trading system. As predictor variables, technical market indicators calculated from the historical price series are used. To define the best buying moment of the asset, the three-barrier method is applied. Additionally, this work presents a practical approach with the inclusion of Take Profit and Stop Loss limits. The results obtained from the proposed model were compared with the Moving Average Crossover (MAC) trading strategy.

Resumo: Este trabalho se concentra no estudo da aplicação de um modelo de aprendizado de máquina conhecido como LGBM (*Light Gradient Boosting Machine*) na classificação dos melhores momentos de compra de um ativo no mercado de ações brasileiro, funcionando como um sistema de *trading*. Como variáveis preditoras, são utilizados indicadores técnicos de mercado calculados a partir da série histórica de preços. Para definir o melhor momento de compra do ativo é aplicado o método das três barreiras. Além disso, esse trabalho apresenta uma abordagem prática com a inclusão dos limites de *Take Profit* e *Stop Loss*. Os resultados obtidos com o modelo proposto foram comparados com a estratégia de *trading Moving Average Crossover* (MAC).

Keywords: Machine Learning; LGBM; Trading System; Triple-Barrier Method; Technical Market Indicators; Investment Strategies.

Palavras-chaves: Aprendizado de Máquina; LGBM; Sistema de Trading; Método das Três Barreiras; Indicadores Técnicos de Mercado; Estratégias de Investimento.

1. INTRODUÇÃO

Nos últimos tempos, a implementação de algoritmos automatizados para negociação de ações com base em métodos matemáticos têm ganhado destaque na indústria financeira. Essa tendência é influenciada pelos avanços tecnológicos, como o processamento rápido e acessível de grandes volumes de dados, o surgimento de algoritmos de aprendizado de máquina avançados e a disponibilidade de dados provenientes de diversas fontes. Isso, de acordo com Prado (2020), conspira para desencadear uma era dourada da aplicação de aprendizado de máquina na indústria financeira.

Há, atualmente, diversos estudos que exploram a aplicação de técnicas de aprendizado de máquina no mercado financeiro. Alguns desses estudos utilizam o aprendizado de máquina para solucionar problemas de regressão, onde o modelo é utilizado para prever valores de preço futuros. Outros trabalhos, incluindo o presente estudo, abordam a aplicação de técnicas de classificação para prever os melhores momentos de compra e venda de um ativo.

Neste trabalho é apresentada uma aplicação prática do método LGBM na classificação do melhor momento de compra de um ativo, identificando as oportunidades de in-

vestimento mais promissoras. Para o treinamento e utilização do modelo de aprendizado de máquina, são utilizados indicadores técnicos de mercado como variáveis preditoras. Após a realização da compra de um ativo, a posição é zerada com base em valores de *Take Profit* e *Stop Loss* pré-definidos, visando garantir o lucro e limitar as perdas nos casos de erro na classificação.

2. REVISÃO BIBLIOGRÁFICA

Um dos primeiros trabalhos a apresentar a aplicação de uma técnica de aprendizado de máquina na predição do movimento de um ativo no mercado de ações, White (1988) estudou o uso de uma rede neural na extração de padrões de uma série temporal com os valores de retorno diário. Aplicando *Support Vector Machine*, Huang et al. (2005) propôs um algoritmo para predição do movimento do mercado de ações da bolsa de valores de Tóquio. Como um estudo de comparação, Kumar and Thenmozhi (2006) apresenta o resultado da aplicação das técnicas de *Support Vector Machine* e *Random Forest*.

Além da utilização da série histórica de preços, alguns estudos têm utilizado indicadores de mercado como variáveis preditoras em modelos de aprendizado de máquina. Um exemplo disso é o estudo realizado por Naik and Mohan

(2019) na Bolsa Nacional de Valores da Índia (NSE). Eles extraíram 33 indicadores técnicos de ações listadas na NSE e utilizaram algoritmos de aprendizado de máquina para construir modelos de classificação capazes de prever movimentos do mercado de ações.

Recentemente, Frattini et al. (2022) apresentou um estudo que utilizou o mesmo modelo de aprendizado empregado neste trabalho. O estudo de Frattini et al. (2022) aplicou o modelo LGBM na criação de um indicador de tendência. Outro estudo recente na área, realizado por Mousavi Anzahaei and Nikoomaram (2022), apresentou uma comparação entre o modelo LGBM e o modelo *Catboost* na aplicação de estratégias de *trading*, no qual a estratégia que empregou o modelo LGBM obteve um retorno de 164%, em contraste com 142% alcançados utilizando o modelo *Catboost*.

Em um contexto semelhante, Wang et al. (2023) compara diferentes modelos de aprendizado de máquina, incluindo modelos de aprendizado profundo, combinando indicadores técnicos de mercado e análise de sentimentos como variáveis preditoras. A análise de sentimentos é realizada através da avaliação de menções em mídias sociais.

Os trabalhos que visam prever o movimento do mercado com a utilização de uma técnica de classificação apresentam diferentes abordagens para a rotulagem de dados. Algumas dessas abordagens incluem a comparação de um retorno com um valor fixo, ou a avaliação de máximos e mínimos da série histórica com o preço atual do ativo. Uma das principais referências na área, De Prado (2018), apresenta uma abordagem diferente e prática chamada de método das três barreiras, técnica utilizada neste trabalho.

Utilizando um caso particular do método das três barreiras, reduzindo o total de barreiras para dois, Palazzo et al. (2023) aplica o modelo de *Random Forest* para previsão do movimento no preço da criptomoeda Litecoin.

3. METODOLOGIA

Neste trabalho, a metodologia adotada segue quatro etapas principais: extração e pré-processamento dos dados, treinamento e aplicação do modelo de aprendizado de máquina, validação dos resultados do modelo e por último a validação dos resultados da estratégia de *trading*.

3.1 Extração e pré-processamento

Primeiramente foram extraídos os dados históricos de cada ativo, contendo valores de abertura O_i , fechamento C_i , alta H_i , baixa L_i e volume V_i para cada instante de tempo i . Os dados coletados representam valores diários, sendo assim, a indicação i representa um dia. Após a extração dos dados, foi realizado o cálculo dos indicadores de mercado apresentados abaixo, conforme detalhado por Colby (2003).

- (1) *Relative Strength Index* (RSI): este indicador, primeiramente apresentado por Wilder (1978), é usado para medir a força e a velocidade de um movimento de preço;
- (2) Williams ($\%R$): reflete o preço de fechamento em relação ao valor mais alto de um determinado período;

- (3) *Bollinger Band %*: quantifica a relação entre o valor de fechamento e as bandas de Bollinger e ajudam a identificar pontos de virada em uma tendência.
- (4) *Keltner Channel %*: indicador técnico que utiliza bandas de volatilidade para identificar possíveis pontos de entrada e saída no mercado financeiro;
- (5) CCI: mede a variação de preço do ativo e sua variação de preço médio;
- (6) *Moving Average Convergence Divergence* (MACD): mostra a relação entre duas médias móveis de preços, sendo uma de curto prazo EMA_s e uma de longo prazo EMA_l ;
- (7) *Force Index* (FI): é um indicador de momento ponderado por volume, calculado como a diferença entre o fechamento atual e o fechamento anterior, multiplicado pelo volume;
- (8) *Volume-Price Trend* (VPT): é baseado em um volume cumulativo contínuo que adiciona ou subtrai um múltiplo da alteração percentual na tendência do preço da ação e no volume atual, dependendo dos movimentos de alta ou baixa do investimento;
- (9) *Daily Return* (DR): é uma medida da variação percentual, do preço de uma ação de um dia para o outro;
- (10) *Daily Log Return* (DLR): é uma medida da variação percentual, no logaritmo natural, do preço de uma ação de um dia para o outro;

Definição das variáveis preditoras Com os indicadores de mercado apresentados, foram definidas as variáveis preditoras que são utilizadas no modelo de aprendizado de máquina. Para facilitar a captura de padrões e tendências no tempo, o vetor $\mathbf{x}_i \in \mathbb{R}^l$, que representa as variáveis preditoras, é composto pelos indicadores entre os instantes de tempo i e $i - 5$. Com a utilização de 10 indicadores de mercado e a utilização de 6 instantes de tempo, temos $l = 60$, ou seja,

$$\mathbf{x}_i = \begin{bmatrix} RSI_i \\ \vdots \\ RSI_{i-5} \\ \vdots \\ DLR_i \\ \vdots \\ DLR_{i-5} \end{bmatrix} \quad (1)$$

Definido o vetor \mathbf{x}_i para o instante de tempo i , é preciso relacioná-lo a variável $y_i \in \{0, 1\}$, sendo $y_i = 1$ um momento de compra.

Rotulagem dos dados Neste trabalho, a variável $y_i \in \{0, 1\}$ indica o melhor momento de compra de um ativo e é definida com base no método das três barreiras apresentado por De Prado (2018). Essa abordagem considera alguns efeitos importantes que precisam ser avaliados na criação de um sistema de *trading*. Utilizando a série temporal com o preço de fechamento do ativo, analisamos o ponto i com base na criação de três barreiras, uma barreira superior, uma barreira inferior e uma terceira barreira com base no tempo. Definidas as três barreiras temos que i representa um momento de compra ($y_i = 1$) caso o primeiro toque da série fechamento ocorra com a barreira superior. Na Figura 1 é ilustrado o caso em que o ponto

i é definido como um momento de compra. Vale ressaltar que, neste trabalho, a barreira utilizada tem o mesmo valor tanto para o *Take Profit* quanto para o *Stop Loss* que será utilizado na estratégia.

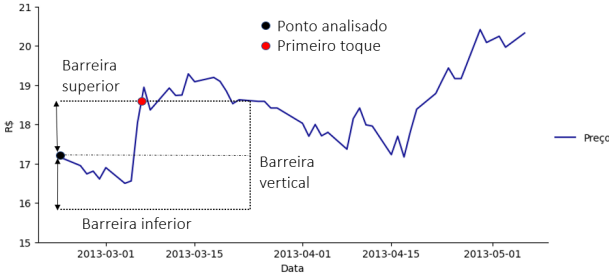


Figura 1. Representação de um ponto de compra com o método das três barreiras.

3.2 Modelo de aprendizado de máquina

Os algoritmos de aprendizado de máquina são capazes de capturar padrões com uso de dados e possuem aplicações em diversas áreas. Este trabalho se concentra na utilização do modelo conhecido como LGBM (*Light Gradient Boosting Machine*). Desenvolvido por Ke et al. (2017), o algoritmo de aprendizado de máquina LGBM é baseado em um conjunto de árvores de decisão, que utiliza a técnica de *boosting* para melhorar a precisão dos resultados.

Árvores de decisão Árvores de decisão é um tipo de algoritmo de aprendizado onde os dados são previstos por meio de uma sequência de testes lógicos definidos como nós. O treinamento de uma árvore de decisão envolve encontrar as regras de decisão que melhor se ajustam aos dados de treinamento. Usando o vetor \mathbf{x}_i como exemplo e representando uma determinada variável preditora através do índice j , a construção da regra de divisão é realizada através da condição se $x_{ij} \leq \epsilon$ ou não, onde ϵ é algum limiar e x_{ij} representa o elemento da posição j do vetor \mathbf{x}_i . A construção de uma árvore de decisão é feita pela escolha da variável e da regra associada a variável, que melhor divide um conjunto de itens em cada passo Breiman et al. (2017). Por simplificação, representamos os parâmetros j e ϵ do nó m como $\theta_m = (j_m, \epsilon_m)$, onde θ_m define a regra de divisão no nó m . Dada uma amostra de treinamento, a escolha de θ_m é avaliada com base na sua capacidade de separação dos dados no nó m , avaliada através de uma métrica definida como impureza. Representando todo o conjunto de treinamento como $\mathcal{L}_n = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, com x_{ij} o elemento j do vetor \mathbf{x}_i , e lembrando que θ_1 representa os parâmetros j e ϵ do nó 1, podemos então realizar a primeira divisão dos dados em dois grupos com tamanhos n_e e n_d :

$$\mathcal{L}_{n_d}^{direita}(\theta_1) = \{(\mathbf{x}_i, y_i) | x_{ij_1} \leq \epsilon_1\} \quad (2)$$

$$\mathcal{L}_{n_e}^{esquerda}(\theta_1) = \{(\mathbf{x}_i, y_i) | x_{ij_1} > \epsilon_1\} \quad (3)$$

onde $\mathcal{L}_{n_d}^{direita}$ e $\mathcal{L}_{n_e}^{esquerda}$ representam subconjuntos da amostra de treinamento total \mathcal{L}_n , e:

$$n = n_e + n_d \quad (4)$$

A impureza associada a divisão pode então ser calculada como:

$$G(\mathcal{L}_n, \theta_1) = \frac{n_d}{n} H(\mathcal{L}_{n_d}^{direita}(\theta_1)) + \frac{n_e}{n} H(\mathcal{L}_{n_e}^{esquerda}(\theta_1)) \quad (5)$$

onde H representa uma função de perda.

Para melhorar o desempenho das árvores de decisão, uma técnica poderosa de aprendizado de máquina é o *boosting*. Essa técnica consiste em construir várias árvores de decisão, onde cada uma delas se concentra em corrigir os erros da árvore anterior, e combiná-las para formar um modelo forte e capaz de lidar com dados mais complexos e fazer previsões mais precisas. Como afirmado por Freund and Schapire (1997), o *boosting* efetivamente reduz o viés e melhora a capacidade de generalização do modelo como um todo, tornando-o um método amplamente utilizado em aprendizado de máquina para tarefas como classificação e regressão.

Boosting Como uma aplicação da técnica de *boosting*, o *Gradient Boosting* Friedman (2002) é uma técnica avançada de aprendizado de máquina que tem se mostrado muito eficaz em muitos problemas de regressão e classificação. Essa técnica utiliza o gradiente descendente para ajustar o modelo iterativamente, minimizando o erro de previsão

O modelo utilizado neste trabalho, conhecido como LGBM é uma melhoria do algoritmo *Gradient Boosting*, que busca otimizar a eficiência computacional e a acurácia das previsões.

De modo geral, a partir de um conjunto de treino \mathcal{L}_n , o algoritmo de *boosting* busca encontrar uma função $F_M(\mathbf{x}_i)$, que pode ser representada como uma soma de M funções de aprendizado $h_m(\mathbf{x}_i)$, onde cada função de aprendizado significa uma árvore de decisão simples.

$$F_M(\mathbf{x}_i) = \sum_{m=1}^M \beta_m h_m(\mathbf{x}_i) \quad (6)$$

O parâmetro β_m , onde $0 < \beta_m \leq 1$, controla a taxa de aprendizado e é utilizado na regularização para ajustar a contribuição de cada árvore Friedman (2001). Em um problema de regressão o resultado da função $F_M(\mathbf{x}_i)$ representa a variável de resposta \hat{y}_i , enquanto em um problema de classificação binária o valor da função é então mapeado na probabilidade de $y_i \in \{0, 1\}$ pertencer a classe dado o valor de \mathbf{x}_i

$$\hat{y}_i = p(y_i = 1 | \mathbf{x}_i) = \sigma(F_M(\mathbf{x}_i)) \quad (7)$$

onde σ representa uma função sigmoid (ou logística).

Definida uma função de perda diferenciável L , a criação do conjunto de árvores ocorre de modo iterativo, na iteração m a árvore h_m busca minimizar a perda L_m dada pelo conjunto de árvores que já foram construídos no passo anterior $F_{m-1}(\mathbf{x}_i)$ conforme a equação abaixo:

$$h_m = \arg \min_h L_m = \arg \min_h \sum_{i=1}^n l(y_i, F_{m-1}(\mathbf{x}_i) + h(\mathbf{x}_i)), \quad (8)$$

Em um problema de classificação é comum o uso da função de perda *logloss* definida como:

$$l(y_i, \hat{y}_i) = -(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (9)$$

O algoritmo LGBM além de utilizar a técnica de *Gradient Boosting* foi projetado para ser distribuído e eficiente, apresentando vantagens como treinamento mais rápido, maior eficiência e menor uso de memória. Além disso, o LGBM oferece melhor precisão e suporte para aprendizado paralelo, distribuído e em GPU. Ele é capaz de lidar com dados em grande escala, tornando-se uma ferramenta atraente para a resolução de problemas de aprendizado de máquina em larga escala.

3.3 Avaliação do modelo

Neste trabalho, foram utilizadas duas métricas de avaliação do modelo de classificação: acurácia e precisão. A acurácia mede a proporção de previsões corretas em relação ao número total de previsões. A precisão mede a proporção de previsões positivas que estão corretas. É importante ressaltar que a precisão é uma métrica crucial para a avaliação do modelo, pois indicar uma compra de modo errado tende a ser pior do que deixar de indicar uma compra certa.

3.4 Avaliação da estratégia de trading

A aplicação da estratégia de *trading* consiste em realizar a compra do ativo no momento em que o modelo treinado indica a compra e manter a posição até que o *Take Profit* ou *Stop Loss* seja atingido, quando a posição de compra é então zerada. Durante o período em que a posição de compra estiver aberta, novas indicações de compra do modelo são desconsideradas. Este processo é repetido ao longo do tempo, de acordo com as indicações de compra geradas pelo modelo treinado, visando maximizar o retorno financeiro e a taxa de acerto das operações.

Para avaliar a estratégia de *trading* proposta, foram utilizadas três métricas: retorno financeiro, taxa de acerto das operações e índice de Sharpe. O retorno financeiro mede a porcentagem de lucro obtido em relação ao capital investido, enquanto a taxa de acerto das operações mede a proporção de operações bem-sucedidas em relação ao total de operações realizadas. Apresentado primeiramente por Sharpe (1966), o índice de Sharpe é uma medida que avalia o desempenho de uma estratégia de investimento em relação ao risco assumido.

Essas métricas foram comparadas com os resultados obtidos pela estratégia de *trading Moving Average Crossover* (MAC), a fim de determinar a eficácia e rentabilidade da estratégia proposta em relação a uma estratégia mais simples e amplamente utilizada.

4. DESENVOLVIMENTO

O processo de treinamento e validação foi realizado com o uso de uma técnica de janela deslizante, onde os dados

de 2011 até 2022 foram divididos em um conjunto de validação, um conjunto de treinamento e um conjunto de teste conforme Figura 2.

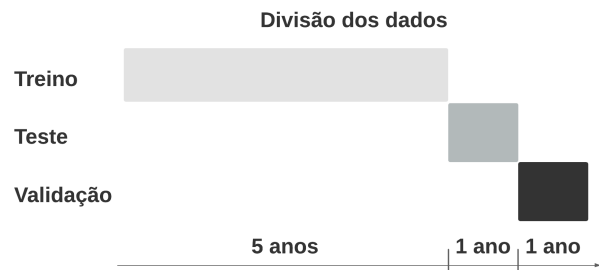


Figura 2. Representação da divisão dos dados.

Para cada ano no conjunto de validação foram utilizados os 6 anos anteriores para treinamento e teste do modelo de aprendizado de máquina. Os resultados apresentados na etapa de validação contemplam os anos de 2017 até 2022. Primeiramente os dados de alguns dos principais ativos listados na bolsa de valores do Brasil foram extraídos, como PETR4, Petróleo Brasileiro SA Petrobras Preference Shares; VALE3, Vale S.A.; ITUB4, Itaú Unibanco; BBDC4, Banco Bradesco SA Preference Shares; ABEV3, AMBEV. Os dados obtidos possuem valores diários de abertura O_i , fechamento C_i , alta H_i , baixa L_i e volume V_i . Com isso, para cada ano avaliado, os valores extraídos foram divididos de acordo com treinamento, teste e validação. Os indicadores técnicos de mercado foram calculados e o vetor \mathbf{x}_i foi obtido.

Após o cálculo do vetor \mathbf{x}_i , foi realizada a rotulagem dos dados para obtenção da variável y_i . Foi utilizado um limite de 3,0% nas barreiras superior e inferior. Esse mesmo limite foi aplicado no *Take Profit* e *Stop Loss* da estratégia de *trading*. Como barreira vertical, foi utilizado um período de 20 dias.

Para realizar o treinamento do modelo, foi utilizado o pacote LightGBM desenvolvido pela Microsoft, disponível na linguagem de programação Python. Utilizando o conjunto de treino fornecido, o modelo LGBM foi treinado com um total de 100 árvores de decisão, uma profundidade máxima de 4 e uma taxa de aprendizado de 0,06. Esse conjunto de hiperparâmetros foi escolhido com base em experimentos anteriores e ajustes finos realizados para otimizar o desempenho do modelo em relação a métricas relevantes como acurácia. O treinamento foi realizado de forma individual para cada ativo, mantendo os mesmos hiperparâmetros por questões de simplificação.

De acordo com a janela deslizante apresentada na Figura 2, após cada treinamento, os dados do conjunto de teste foram utilizados para avaliar o desempenho do modelo com uso da métrica de precisão. Quando a precisão do modelo em teste indica um maior número de falsos positivos em relação aos verdadeiros positivos, o modelo é desconsiderado, por outro lado, se a precisão for satisfatória, o modelo é aplicado no ano de validação para indicar os pontos de compra.

Com a indicação dos melhores pontos de compra, como resultado do modelo de classificação \hat{y}_i para os dados de

validação dos anos 2017 até 2022, a estratégia de *trading* foi então avaliada. Definindo $PosCompra_i \in \{0, 1\}$ como uma variável que define se existe uma posição de compra aberta no instante de tempo i , $PosCompra_i = 1$, a ordem de Compra ou Venda definida pela estratégia pode então ser expressa como:

$$\begin{cases} \text{Compra} & \text{se } PosCompra_i = 0 \text{ e } \hat{y}_i = 1 \\ \text{Venda} & \text{se } PosCompra_i = 1 \text{ e } |retorno_i| \geq lim \end{cases}$$

onde $|retorno_i|$ representa o retorno percentual obtido no instante de tempo i com relação a última operação. A variável lim , que representa o *Take Profit* e *Stop Loss*, possui o mesmo valor atribuído às barreiras superior e inferior na etapa de rotulagem dos dados.

5. RESULTADOS

Após a conclusão da simulação, os valores obtidos para o conjunto de validação, que compreendeu o período de 2017 a 2022, foram avaliados em relação à acurácia e precisão da classificação do momento de compra, bem como a taxa de sucesso das operações realizadas, conforme apresentado na Tabela 1. Esses resultados indicam que o modelo de classificação apresentou desempenhos variados para diferentes ativos.

Tabela 1. Resultado do modelo de classificação (2017 até 2022).

Ativo	Acurácia	Precisão	Taxa de acerto %	
			Modelo	MAC
ABEV3	53,82	51,51	56,60	45,95
BBDC4	53,17	54,49	54,23	41,67
ITUB4	53,90	53,40	56,92	40,54
PETR4	48,51	53,87	56,14	41,03
VALE3	47,27	56,03	55,00	55,56

Com a técnica de rotulagem utilizada o modelo de classificação possui um desafio maior em comparação a outras técnicas de rotulagem, uma vez que os rótulos indicam se, a partir de um determinado ponto, o valor do ativo ultrapassará primeiramente o limite de *taken profit* estabelecido. Isso torna a tarefa de prever o momento exato de compra mais complexa e desafiadora.

Podemos notar que a precisão do modelo de classificação está diretamente vinculada à taxa de acerto das operações realizadas. Avaliando o resultado da taxa de acerto para todos os ativos, é possível destacar um resultado superior do modelo proposto em comparação com a estratégia MAC.

A Tabela 2 apresenta o comparativo entre o índice de Sharpe e o retorno percentual para as operações realizadas com base no modelo de classificação e na estratégia MAC.

Tabela 2. Retorno % e índice de Sharpe (2017 até 2022)

Ativo	Índice de Sharpe		Retorno(%)	
	Modelo	MAC	Modelo	MAC
ABEV3	0,73	0,19	53,67	8,56
BBDC4	0,44	0,45	31,22	41,31
ITUB4	0,55	0,69	41,83	76,81
PETR4	0,67	0,27	100,01	12,00
VALE3	0,58	0,75	50,84	113,94

É possível observar que, em média, o modelo apresentou um índice de Sharpe superior ao da estratégia MAC, indicando que as operações realizadas com base no modelo tiveram um retorno ajustado ao risco melhor. Esses resultados reforçam a eficácia do modelo proposto como uma ferramenta de suporte à tomada de decisão em operações de curto prazo no mercado financeiro. No entanto, é importante destacar que o modelo também apresentou um desempenho inferior em relação à estratégia MAC para alguns ativos, o que indica que a seleção de ativos específicos pode ser um fator importante a ser considerado ao utilizar o modelo proposto.

Para ilustrar o desempenho geral do modelo, a Figura 3 apresenta o retorno acumulado de uma carteira composta com todos os ativos que foram avaliados, considerando que o valor inicialmente alocado em cada ativo seja exatamente o mesmo.



Figura 3. Retorno acumulado da carteira (%).

O resultado apresentado na Figura 3 consolida o resultado individual de cada ativo apresentado na Tabela 2. Para otimizar a carteira, seria possível alocar os recursos disponíveis durante o período em que não há posição de compra no ativo, $PosCompra_i = 0$, em outra aplicação ou mesmo aumentar a posição de compra em outro ativo da carteira. É importante ressaltar que os resultados apresentados não levam em consideração o custo de transação, que é um fator relevante a ser considerado em uma aplicação prática.

6. CONCLUSÃO

Neste estudo, foi abordada a aplicação de um modelo de classificação para o desenvolvimento de uma estratégia de *trading*. O algoritmo LGBM foi adotado como o modelo de aprendizado de máquina. Para estruturar a abordagem de classificação, utilizou-se a técnica de rotulagem conhecida como “método das três barreiras”. Como variáveis preditoras, foram empregados indicadores técnicos de mercado, os quais foram calculados com base nos valores de abertura, fechamento, alta, baixa e volume.

A avaliação da capacidade de classificação do modelo envolveu a análise dos resultados de acurácia e precisão. Além disso, para avaliar a eficácia do modelo como estratégia de *trading*, foram comparados os índices de Sharpe,

as taxas de acerto e os retornos financeiros do modelo em relação à estratégia MAC. Todos os resultados apresentados foram obtidos por meio da análise dos anos de 2017 até 2022.

Na comparação realizada entre a aplicação do modelo proposto e a estratégia MAC, ao analisarmos o resultado médio para todos os indicadores considerados, o modelo proposto demonstrou um desempenho superior, com destaque particular para a taxa de acerto.

A aplicação de aprendizado de máquina na previsão de movimentos de preços em ativos financeiros enfrenta desafios complexos. Vale destacar que o modelo de classificação apresentado é fundamentado em dados históricos, não considerando eventos externos que possam influenciar o mercado. Mesmo ao se limitar a dados passados, a adoção de um modelo de classificação com a utilização de indicadores técnicos de mercado demonstra o potencial de desenvolver estratégias mais robustas em comparação com abordagens populares também baseadas em registros históricos, como a estratégia MAC.

Em futuros estudos, seria interessante contrastar os resultados de diversos modelos de aprendizado de máquina, além de incorporar novas variáveis preditoras, como variáveis derivadas da análise de sentimentos. Além disso, seria benéfico considerar a aplicação de parâmetros de forma individual para cada ativo, como os hiperparâmetros utilizados no modelo LGBM, a fim de adaptar a abordagem às características específicas de cada ativo.

REFERÊNCIAS

- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (2017). *Classification and regression trees*. Routledge.
- Colby, R.W. (2003). *The encyclopedia of technical market indicators*. McGraw-Hill.
- De Prado, M.L. (2018). *Advances in financial machine learning*. John Wiley & Sons.
- Frattini, A., Bianchini, I., Garzonio, A., and Mercuri, L. (2022). Financial technical indicator and algorithmic trading strategy based on machine learning and alternative data. *Risks*, 10(12), 225.
- Freund, Y. and Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119–139.
- Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.
- Friedman, J.H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4), 367–378.
- Huang, W., Nakamori, Y., and Wang, S.Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & operations research*, 32(10), 2513–2522.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Kumar, M. and Thenmozhi, M. (2006). Forecasting stock index movement: A comparison of support vector machines and random forest. In *Indian institute of capital markets 9th capital markets conference paper*.
- Mousavi Anzahaei, S.M. and Nikoomaram, H. (2022). A comparative study of the performance of stock trading strategies based on lgbm and catboost algorithms. *International Journal of Finance & Managerial Accounting*, 7(26), 63–75.
- Naik, N. and Mohan, B.R. (2019). Stock price movements classification using machine and deep learning techniques—the case study of indian stock market. In *Engineering Applications of Neural Networks: 20th International Conference, EANN 2019, Xersonisos, Crete, Greece, May 24–26, 2019, Proceedings 20*, 445–452. Springer.
- Palazzo, G., Sbruzzi, E.F., Nascimento, C.L., and Leles, M.C. (2023). Predicting litecoin price movement in a pre-defined trading volume window using random forest model. In *2023 IEEE International Systems Conference (SysCon)*, 1–6. IEEE.
- Prado, M.M.L. (2020). *Machine learning for asset managers*. Cambridge University Press.
- Sharpe, W.F. (1966). Mutual fund performance. *The Journal of business*, 39(1), 119–138.
- Wang, Z., Hu, Z., Li, F., Ho, S.B., and Cambria, E. (2023). Learning-based stock trending prediction by incorporating technical indicators and social media sentiment. *Cognitive Computation*, 15(3), 1092–1102.
- White, H. (1988). Economic prediction using neural networks: The case of ibm daily stock returns. In *ICNN*, volume 2, 451–458.
- Wilder, J.W. (1978). *New concepts in technical trading systems*. Trend Research.