



Previsão de *short squeezes* com aprendizado incremental: um modelo ajustado ano a ano

Edvaldo Pereira dos Santos¹, Milton Luis Ribeiro Júnior², Gabriel Sifuentes Rocha³, Lucas de Angelis Oliveira⁴, Gerson Trujillo Navarro⁵, Raphael Chaves⁶, Milton Miranda Neto⁷, Francisco Louzada Neto⁸

ICMC-USP

1 Introdução

Short squeeze é um evento que ocorre quando investidores que realizaram vendas a descoberto são forçados a recomprar rapidamente as ações devido a uma alta repentina no preço, resultando em uma demanda artificial pelo papel e impulsionando ainda mais sua valorização. Apesar de ser um evento raro no mercado financeiro, é potencialmente catastrófico para investidores que estejam operando vendido durante sua ocorrência, uma vez que acarretam em grandes perdas.

O prejuízo da venda a descoberto durante esse tipo de evento pode ser ilimitado, à medida que o preço do ativo se valoriza. É possível até mesmo que o investidor perca mais que 100% do patrimônio investido, se endivide e entre em falência, uma vez que o prejuízo é proporcional à valorização do ativo. Pesquisas recentes indicam que esses eventos podem aumentar os riscos e reduzir significativamente os retornos esperados de operações vendidas, uma vez que forçam os investidores a absorver custos elevados para fechar posições, como mostrado por Schultz [1].

Apesar do impacto econômico, há uma carência de pesquisas aprofundadas sobre a previsibilidade desses eventos, tanto em mercados norte-americanos quanto europeus, como observado por Allen et al. [3]. Ainda assim, estudos indicam que o percentual de ações alugadas, as taxas de aluguel e volumes de negociação podem ser determinantes para a ocorrência de *short squeezes* [2].

Entretanto, a raridade e o caráter estocástico desse fenômeno tornam a previsão desses eventos um desafio para modelos de classificação supervisionados, devido ao desbalanceamento severo

¹edsantos99@usp.br

²milton.ribeiro9@usp.br

³gabrielsifu@alumni.usp.br

⁴lucas.angelis.oliveira@usp.br

⁵truna@usp.br

⁶raphaelchaves@usp.br

⁷milton.cm.miranda@gmail.com

⁸louzada@icmc.usp.br

das classes. Além disso, como o mercado financeiro se adapta constantemente, faz-se necessário o uso de um modelo incremental, que seja capaz de ajustar-se continuamente às novas informações, conforme descrito por Geng e Smith-Miles [4] e por van de Ven et al. [5]. Nesse contexto, a adoção de técnicas de aprendizado incremental podem proporcionar uma maior adaptabilidade ao modelo sem elevar os custos computacionais, permitindo incorporar a variabilidade temporal dos dados de mercado.

Nesse viés, este trabalho tem como objetivo desenvolver um modelo preditivo capaz de identificar a probabilidade de ocorrência de um *short squeeze* no mercado financeiro, utilizando técnicas de aprendizado de máquina incremental, balanceamento de classes e aprendizado profundo, além disto, o presente trabalho foi desenvolvido na disciplina MAI5003, utilizando a abordagem de Aprendizado Baseado em Problemas (PBL). O modelo desenvolvido será alimentado com dados de períodos consecutivos, ajustando-se continuamente à medida que novas informações se tornam disponíveis. O principal objetivo é capturar as características dinâmicas do mercado e prever esses eventos raros. A identificação antecipada de um *short squeeze* pode auxiliar investidores e gestores de fundos a otimizar suas estratégias de investimento, minimizando os riscos associados a posições vendidas.

2 Metodologia

2.1 Base de Dados

Utilizamos um conjunto de dados financeiros diários de empresas listadas em bolsa⁹, contendo cerca de 800 mil registros e 236 variáveis após o processo de limpeza. As variáveis incluem retornos, capitalização de mercado, volatilidade, beta, volume e outros indicadores relevantes para identificar a ocorrência de *short squeezes*. Os dados abrangem um período de 14 anos, de 2011 a 2024, permitindo a análise de eventos históricos e tendências cruciais para a modelagem. Vale destacar que estamos lidando com um evento raro, com uma incidência de aproximadamente 0,06%.

2.2 Pré-processamento

No pré-processamento, foram realizadas correções para lidar com dados ausentes e *outliers*. Para isso, utilizamos *forward fill* e técnicas de winsorização. As variáveis foram normalizadas para garantir escalas comparáveis de forma a se tornarem mais digestas aos modelos utilizados. Séries temporais foram transformadas em janelas deslizantes, capturando o comportamento sequencial das variáveis em formato tabular.

2.3 Modelagem

Desenvolvemos um modelo preditivo incremental com o objetivo de lidar com dados em fluxo contínuo para detectar *short squeezes*, utilizando aprendizado supervisionado. A abordagem consiste em recalibrar o modelo com os dados de períodos consecutivos. O modelo final foi ajustado ano a ano, utilizando os dados incrementais de 2011-2014 para prever 2015, e assim por diante, até 2024. O desempenho foi avaliado por métricas como AUC, Precisão e Revocação.

⁹Dados fornecidos pela empresa Constância Investimentos

Além disso, comparamos dois modelos: uma regressão logística e uma MLP incremental. Devido à raridade do evento de interesse, utilizamos a técnica de ‘class weights’ para balancear o impacto das classes nos dois modelos. Essa estratégia foi essencial para lidar com o desequilíbrio no conjunto de dados e melhorar a performance preditiva dos modelos.

3 Conclusões

A análise dos modelos preditivos para a detecção de *short squeezes* revela que o modelo MLP incremental apresentou um desempenho superior em relação à Regressão Logística. A curva ROC da MLP incremental, apresentada na Figura 1, mostra uma área sob a curva (AUC) de 0.776, significativamente superior ao valor de 0.678 obtido pela Regressão Logística.

Além disso, conforme apresentado na Tabela 1, a MLP incremental apresentou valores superiores em todas as métricas de avaliação: Precisão (0.012 vs. 0.007), Revocação (0.669 vs. 0.601), e F1-Score (0.023 vs. 0.016). Esses resultados indicam que, apesar da raridade do evento, a MLP conseguiu capturar melhor os padrões nos dados, tornando-a uma solução mais eficaz para a detecção de *short squeezes*.

Dessa forma, podemos concluir que a MLP incremental aliada ao balanceamento dos pesos entre as classes oferece uma melhora significativa no desempenho preditivo em relação à Regressão Logística, sendo a abordagem mais recomendada para a tarefa em questão.

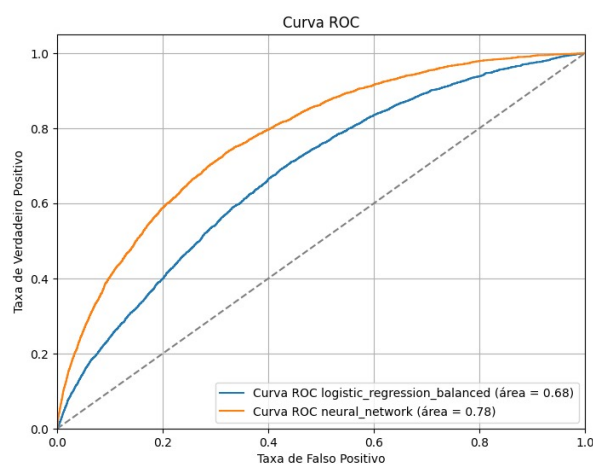


Figura 1: Curva ROC

Tabela 1: Comparação de modelos preditivos.

Modelos	AUC	Precisão (corte 0.5)	Revocação (corte 0.5)	F1-Score (corte 0.5)
Regressão Logística	0.678	0.007	0.601	0.016
MLP incremental	0.776	0.012	0.669	0.023

Referências

- [1] P. Schultz. *Short Squeezes and Their Consequences*. Journal of Financial and Quantitative Analysis, vol. 59, no. 1, pp. 68-96, 2024. doi: 10.1017/S0022109022001533.
- [2] Z. Jiang, B. Liu, A. Schrowang, and W. Xu. *Short Squeezes*. Financial Analysts Journal, vol. 80, no. 2, pp. 152-173, 2023. doi: 10.1080/0015198X.2023.2284625.
- [3] F. Allen, M. Haas, M. Pirovano, and A. Tengulov. *How Prevalent Are Short Squeezes? Evidence From the US and Europe*. Swiss Finance Institute Research Paper, no. 23-63, 2023. Available at: <https://ssrn.com/abstract=4526147>. doi: 10.2139/ssrn.4526147.
- [4] X. Geng and K. Smith-Miles. *Incremental Learning*. In: S. Z. Li and A. Jain (eds), Encyclopedia of Biometrics. Springer, 2009. doi: 10.1007/978-0-387-73003-5_304.
- [5] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias. *Three types of incremental learning*. Nature Machine Intelligence, vol. 4, pp. 1185-1197, 2022. doi: 10.1038/s42256-022-00568-3.