

RELATÓRIO COMPARATIVO DOS MODELOS

1. CNN1D

Pontos fortes: Boa precisão ($\geq 91\%$ em quase todos os ativos);

Pontos fracos: Baixo recall em PETR4 (24%) e VALE3 (40%), indicando que o modelo erra ao identificar muitos casos positivos;

Rentabilidade: Baixa, *final_mult_strategy* bem inferior ao *buy & hold* em alguns casos.

Conclusão: CNN1D não é competitivo neste caso devido a alto viés e um pouco fraco para séries temporais de dependência longa.

2. LSTM

Pontos fortes: Excelente precisão (100% em todos os ativos), ótima performance em BBAS3 e CSNA3;

Pontos fracos: Recall baixo em PETR4 (91%) e em VALE3 (66%);

Rentabilidade: *final_mult_strategy* muito alto em BBAS3 e consistente em PETR4 e CSNA3.

Conclusão: Ótimos resultados, mas acaba sendo mais complexo e pesado. Recall menor pode ser um risco (perder sinais de compra e venda).

3. GRU

Pontos fortes: Equilíbrio muito bom, precisão $>94\%$ e recall muito alto (quase sempre 100%). Em BBAS3 apresentou retorno altíssimo, superando o LSTM.

Pontos fracos: Desempenho mais baixo em VALE3;

Rentabilidade: Muito alta em BBAS3 e CSNA3. Estratégia sempre supera *Buy & Hold*.

Matriz de confusão, Precision e Recall

Foram geradas **matrizes de confusão** para cada modelo, evidenciando como eles classificam cada classe corretamente ou incorretamente.

As métricas de **Precision** e **Recall** foram extraídas a partir das matrizes.

Resumo:

CNN1D: Boa precisão, mas recall mais baixo, indicando maior cautela em algumas classes.

GRU: Equilíbrio entre precision e recall, sendo o mais consistente entre as métricas.

LSTM: Recall elevado em algumas classes, porém com menor precisão, o que pode gerar falsos positivos.

Desempenho Financeiro via Backtest

Foi realizado um **backtest** simulando operações financeiras com base nas previsões dos modelos.

O objetivo foi verificar se o modelo poderia gerar **retorno financeiro positivo**.

Resumo:

CNN1D: Retornos modestos, funcionando como baseline.

GRU: Apresentou o melhor retorno financeiro líquido entre os modelos, indicando potencial de uso prático.

LSTM: Resultados variáveis, com alguns cenários apresentando desempenho inferior.

ANÁLISE FINAL

GRU escolhido como MVP

Tem melhor recall médio (identifica mais sinais relevantes).

Apresentou performance competitiva em todos os ativos, especialmente BBAS3 e CSNA3.

É mais leve e simples de treinar/manter do que LSTM.

Supera CNN1D em quase todos os aspectos.

Observações:

O LSTM pode ser mantido como modelo de comparação, principalmente para cenários em que a precisão absoluta seja mais crítica do que o recall.

O CNN1D pode ser descartado no MVP, mas vale registrar a análise mostrando que não funcionou bem para este tipo de dado (isso fortalece o relatório acadêmico).

O código apresentado traz o MVP de acordo com as análises e testes feitas pelo grupo.

RESULTADOS APRESENTADOS

GRU

	asset	model	acc_test	precision	recall	final_mult_strategy	final_mult_bh	n_trades
0	PETR4	gru	0.992424	0.985390	1.000000	8.260572	3.620412	105
1	VALE3	gru	0.956229	0.945092	0.967632	3.525248	2.244225	91
2	BBAS3	gru	1.000000	1.000000	1.000000	6201.816432	34.444169	376
3	CSNA3	gru	1.000000	1.000000	1.000000	20.401324	1.814886	83

Relatórios salvos em: /content/outputs_dl_timeseries

LSTM

	asset	model	acc_test	precision	recall	final_mult_strategy	final_mult_bh	n_trades
0	PETR4	lstm	0.954545	1.0	0.911038	8.424181	3.620412	104
1	VALE3	lstm	0.835859	1.0	0.667802	3.561733	2.244225	86
2	BBAS3	lstm	1.000000	1.0	1.000000	6201.816432	34.444169	376
3	CSNA3	lstm	1.000000	1.0	1.000000	20.401324	1.814886	83

Relatórios salvos em: /content/outputs_dl_timeseries

CNN1D

	asset	model	acc_test	precision	recall	final_mult_strategy	final_mult_bh	n_trades
0	PETR4	cnn1d	0.606061	0.926380	0.248764	1.743432	3.620412	76
1	VALE3	cnn1d	0.687710	0.915385	0.405451	1.027791	2.244225	94
2	BBAS3	cnn1d	0.920127	0.911434	0.911861	5.513963	34.444169	374
3	CSNA3	cnn1d	0.828352	0.922388	0.638430	2.033336	1.814886	90

Relatórios salvos em: /content/outputs_dl_timeseries