



Validação da Entropia de Boltzmann como Feature de Microestrutura para Otimização de Sistemas de Regimes de Risco em Criptoativos

Alexis Sebastian
Desiree Gonçalves

SUMÁRIO

- VISÃO GERAL 1
- METODOLOGIA 3
- RESULTADOS 6
- CONCLUSÃO 7
- CONSIDERAÇÕES FINAIS 8

REFERÊNCIAS BIBLIOGRAFICAS

- Gautam, M. (2025). Previsão de preços de criptomoedas usando LSTM + XGBoost.
- Smirnov, SL, & McCarty, J. (2008). A Distribuição de Boltzmann e a Definição Estatística de Entropia.
- Kavlakoglu, Eda. (2020) O que é o XGBoost? IBM Research



VISÃO GERAL

O mercado de criptoativos é definido por sua extrema volatilidade e alta não-linearidade , tornando a previsão de preços e a gestão de riscos complexos e desafios.

Devido à sua proeminência global e alta liquidez, o Bitcoin foi escolhido como alvo para este estudo de dinâmica de risco.

Este projeto aborda a dificuldade de predição introduzindo uma solução inovadora: a aplicação da Entropia de Boltzmann (E) , um conceito de Termodinâmica Econômica , como uma característica de microestrutura. Ao quantificar a dispersão e a atividade dos agentes no mercado, podemos ir além dos indicadores clássicos e melhorar significativamente o desempenho dos modelos de Machine Learning na classificação e gestão de regimes de alto risco.



Entropia de Boltzmann

2

A **Entropia de Boltzmann** é uma medida do grau de desordem ou aleatoriedade de um sistema.

A fórmula clássica é:

$$S = k_B \ln \Omega$$

Onde:

- Ω representa o número de microestados possíveis que o sistema pode assumir.

No nosso contexto, os "microestados" são substituídos pela dispersão dos agentes (comerciantes) que buscam realizar transações, medindo a segurança coletiva.

A Entropia (E) é calculada dinamicamente, transformando a dispersão do preço (o intervalo máximo e mínimo) em um volume de fase

$$T = G^n$$

Onde:

- G é o gap em ticks

A Entropia atua como um proxy que quantifica a atividade e a desordem do mercado que não é capturada pelos meios móveis.

Hipótese: "A Entropia de Boltzmann atua como um indicador antecedente (indicador antecedente) de Risco. Uma Entropia Alta sinaliza um período de grande dispersão e atividade dos agentes no mercado (Termodinâmica Econômica), o que, por sua vez, é um forte motor para a volatilidade e a classificação de regimes de risco nos dias subsequentes."



Ludwig Boltzmann (1844 - 1906)

DADOS E MODELAGEM

Fixamos a biblioteca do yfinance, onde baixamos os seguintes dados do Bitcoin no periodo de 2017-2020

- ABERTURA: O,
- MAXIMO: H,
- MINIMO: L,
- FECHAMENTO: 'C',
- VOLUME: 'V'

Data	Abertura (O)	Máximo (H)	Mínimo (L)	Fechamento (C)	Volume (V)
1/1/2017	963.66	1003.08	958.69	998.35	147,775,080
1/2/2017	998.62	1031.39	996.72	1021.75	222,128,492
1/3/2017	1021.6	1044.08	1021.6	1043.84	185,168,000
1/4/2017	1044.4	1159.42	1044.4	1154.73	384,449,584
1/5/2017	1156.73	1192	910.46	1013.38	510,190,980

Exemplo simplificado

Em seguida, o processamento de dados envolveu a Engenharia de Características , com o cálculo de indicadores técnicos clássicos— incluindo o RSI (Índice de Força Relativa), SI (Índice Estocástico), MACD (Convergência/Divergência de Médias Móveis), e ADX (Índice Direcional Médio)—que compôs o conjunto inicial de características para a modelagem.

Ticker	Date	E	ADX	M	R	SI	V
BTC-USD	12/26/20...	86.396856	22.34345	-96.139127	46.191335	62.619891	22787010034
BTC-USD	12/27/20...	80.537528	20.775785	-86.832677	47.74668	67.200157	22777360996
BTC-USD	12/28/20...	78.234358	19.448511	-76.325964	48.619622	69.700039	21365673026
BTC-USD	12/29/20...	83.671147	18.792289	-58.875243	51.867873	79.077341	22445257702
BTC-USD	12/30/20...	83.671147	18.161343	-54.875149	47.833501	67.460588	22874131672

Exemplo simplificado

- Primeiro, definimos o que constitui “risco futuro” para o nosso projeto: a Volatilidade Futura de 10 dias .
- Medida: Usamos o preço de fechamento (C) para calcular o retorno percentual.
- Volatilidade Futura : Calculamos o desvio padrão dos retornos em uma janela de 10 dias.
- Alinhamento Temporal: Definimos alinhar essa volatilidade futura com a feature do dia atual . Ou seja, a volatilidade que vai acontecer nos próximos 10 dias é movida para a linha de hoje, tornando-se o alvo a ser previsto.

- O XGBoost precisa de um alvo binário (0 ou 1).
- Limite de Risco: Analisamos a distribuição histórica da *Vol_Futura_10Dias* e definimos o percentil 75 como o limite para o Risco Alto. Isto garantiu que apenas os 25% dos dias mais voláteis fossem considerados risco.
- Dividimos os alvos em:
 - y = 1 (Risco Alto) : Se a *Vol_Futura_10D* fosse maior que o limite (quantil 75).
 - y = 0 (Risco Baixo): Se fosse menor ou igual.
- Essa modelagem transformou uma série contínua de preços em um problema de Classificação de Risco Binário que foi o foco do nosso projeto.

Classe	Proporção (%)
Risco Baixo (0)	82.24%
Risco Alto (1)	17.76%

Utilizando um modelo de classificação, cujo objetivo é otimizar a coleta de eventos de alta volatilidade futura (nossos regimes de Risco Alto), uma base de dados foi dividida cronologicamente em 80% para treinamento e 20% para teste .

Descrição	Valor
Total de amostras prontas	1069
Amostras de Treino	855
Amostras de Teste	214

PROCESSAMENTO DOS DADOS

Esta foi uma fase em que transformamos preços brutos em sinais preditivos

- Entropia de Boltzmann (E) : Calculamos a Entropia de Boltzmann com base na dispersão de preços em uma janela de 5 dias, criando uma característica de microestrutura.
- Volatilidade Futura: Calculamos o desvio padrão dos retornos dos próximos 10 dias e usamos o método .shift(-10)para que o modelo pudesse usar a informação do presente (t) para prever o risco futuro (t+10).
- Cálculo de Indicadores Clássicos : Calculamos uma gama de indicadores básicos de momentum e tendência (MACD, RSI, ADX, etc.) para dar contexto ao nosso modelo.

A parte mais importante da nossa preparação foi criar o alvo de classificação.

- Definição do Risco: Definimos "Risco Alto" com base no quantil 75 da volatilidade futura.
- Binarização: Convertemos a volatilidade contínua em um rótulo binário (y): 1 (Risco Alto) ou 0 (Risco Baixo) .

FEATURES

Sigla	Métrica	-	Tipo
MACD	Convergência/Divergência de Médias Móveis	Mede o momento (momentum) e a tendência do preço. É a diferença entre duas médias móveis exponenciais (geralmente de 12 e 26 dias).	Tendência/Momento
RSI	Índice de Força Relativa)	Mede a velocidade e a mudança dos movimentos de preço. Ajuda a identificar se um ativo está sobrecomprado (>70) ou sobrevendido (<30).	Momento
IE	Índice Estocástico	Mede a relação entre o preço de fechamento atual e a variação de preço em um período (geralmente 14 dias). Ajuda a prever pontos de reversão .	Momento
C	Preço de Fechamento	É o preço final do ativo em um determinado dia. É uma <i>característica</i> de preço mais básico.	Preço
ADX	Índice Direcional Médio	Mede a força da tendência . Ele não diz se o preço está subindo ou descendo, mas <i>com que força</i> a tendência atual está se movendo.	Tendência
E	ENTROPIA DE BOOLTZMAN	Mede a dispersão e a atividade dos agentes no mercado (Baseado em um modelo de física/termodinâmica)	Microrganismos/Risco

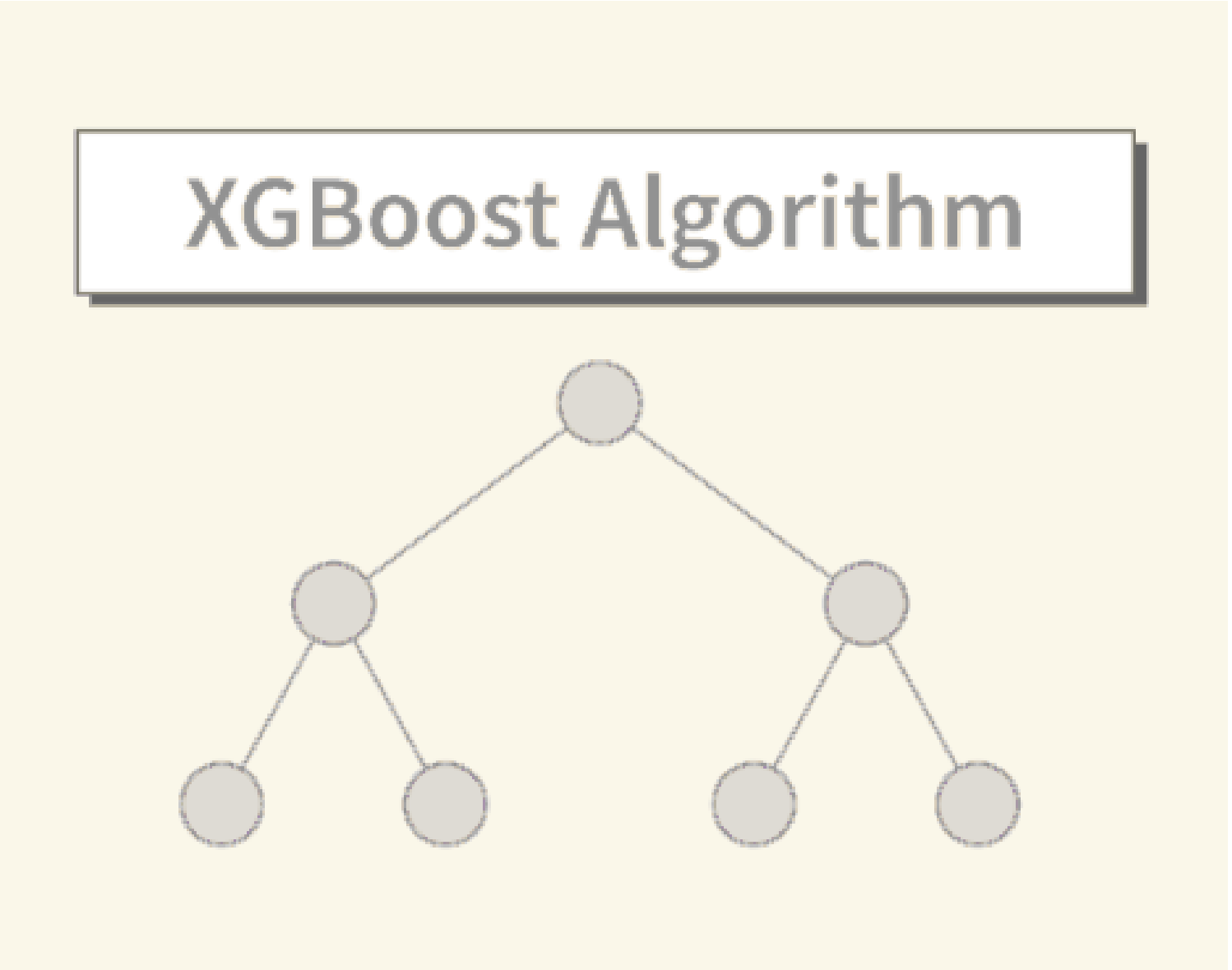
XGBOOST

O XGBoost funciona construindo sequencialmente uma série de árvores de decisão (os "estimadores"). Cada nova árvore é treinada para corrigir os erros residuais (os erros) cometidos pela árvore anterior. O modelo final é a soma das variações de todas essas árvores, resultando em uma estimativa robusta e altamente precisa.

A Aplicação no Nosso Projeto (Classificação de Risco)

- Objetivo: Classificação Binária (Risco Altovs. Risco Baixo).
- Entrada: O modelo recebeu um vetor de recursos que incluía Entropia de Boltzmann (E) , MACD, ADX, RSI, etc.
- Resultado: O modelo nos revelou a probabilidade de cada dia cair na classe de 'Risco Alto' (1).

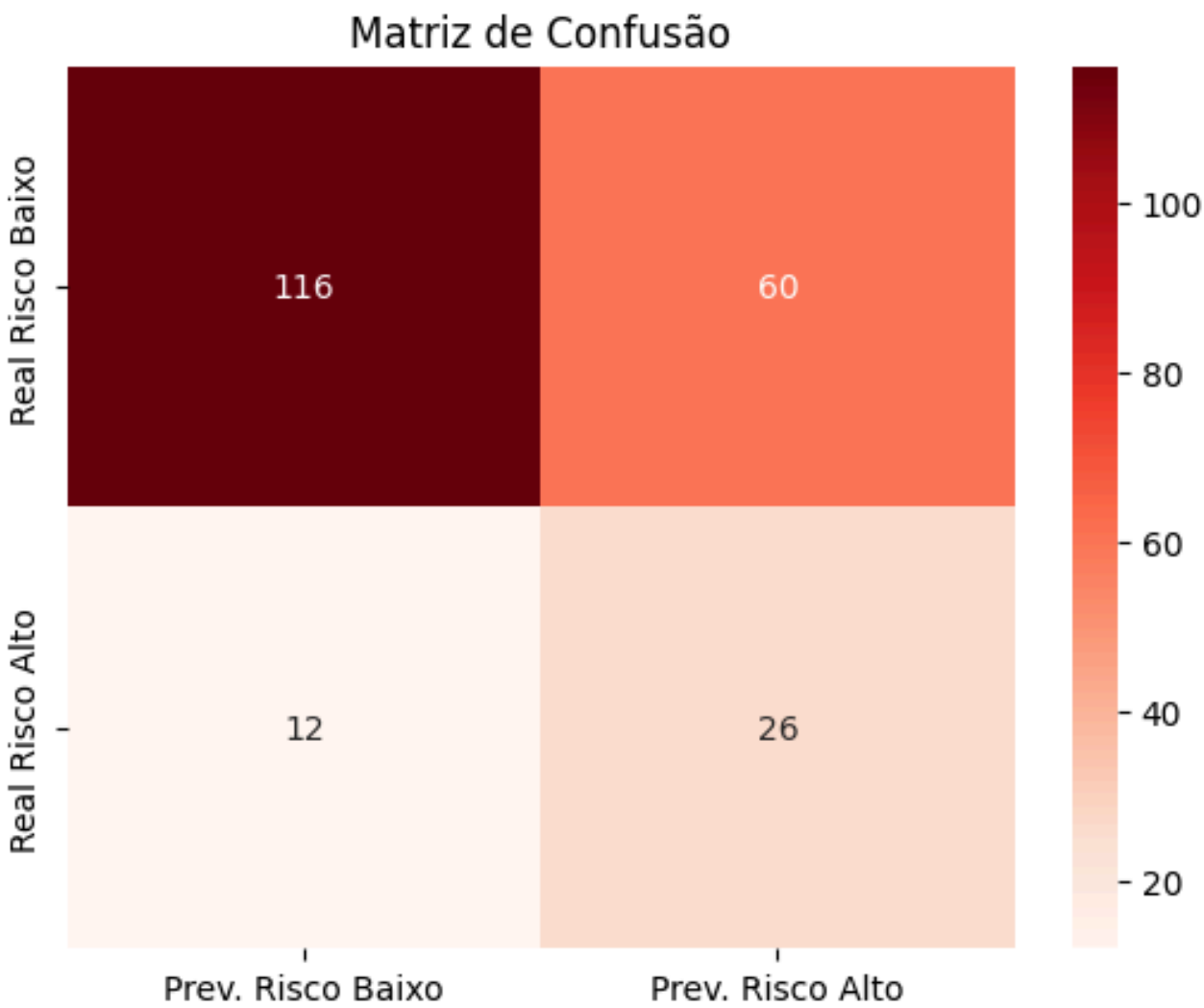
Métrica	"	Contexto (Risco Alto)
Precisão	A proporção de positivas positivas (Verdadeiros Positivos) em relação ao total de positivas positivas feitas.	De todas as vezes que o modelo deu o Alerta de Risco , qual a porcentagem de acerto? (Quanto menor, mais alarmes falsos).
Recall	A proporção de casos positivos reais que foram corretamente identificados.	Dos dias em que o Risco era realmente Alto , quantos o modelo conseguiu identificar? (Esta é uma métrica de segurança mais crítica) .
Pontuação F1	A média harmônica entre Precision e Recall.	A melhor métrica para avaliar o equilíbrio do modelo em dados desequilibrados.
Suporte	O número de ocorrências reais daquela classe no conjunto de teste.	O número de dias que foram <i>realmente</i> 'o Evento'



O conjunto de recursos preditivos (incluindo a Entropia de Boltzmann) e o alvo binário (Risco Alto/ Risco Baixo) foram submetidos ao classificador XGBoost. A otimização do modelo foi realizada em duas etapas críticas:

- Utilizamos o framework Optuna para testar e identificar o conjunto ótimo de configurações nativas do XGBoost, garantindo o máximo desempenho preditivo possível da arquitetura do modelo.
- Dado o desbalanceamento de classes (Risco Alto representando apenas 21% dos dados), foi crucial ajustar o modelo utilizando o parâmetro `scale_pos_weight`(pesagem da classe positiva), o que é necessário para forçar o sistema a penalizar Falsos Negativos e priorizar a Revogação (Recall) .

RESULTADOS



A matriz de confusão mostra como o modelo se saiu ao tentar prever os dias de risco no mercado de criptoativos.

- O modelo acertou 116 dias tranquilos, ou seja, dias que realmente eram de Risco Baixo e foram classificados corretamente.
- Também conseguiu identificar 26 dias perigosos como Risco Alto, o que é importante porque esses são os dias que mais interessam para quem quer se proteger da volatilidade.
- Por outro lado, o modelo errou 12 vezes, deixando de alertar sobre dias que acabaram sendo de alto risco — esses são os chamados falsos negativos, e representam o tipo de erro mais crítico.
- Além disso, o sistema gerou 60 falsos alarmes, classificando como “Risco Alto” dias que na verdade eram tranquilos.



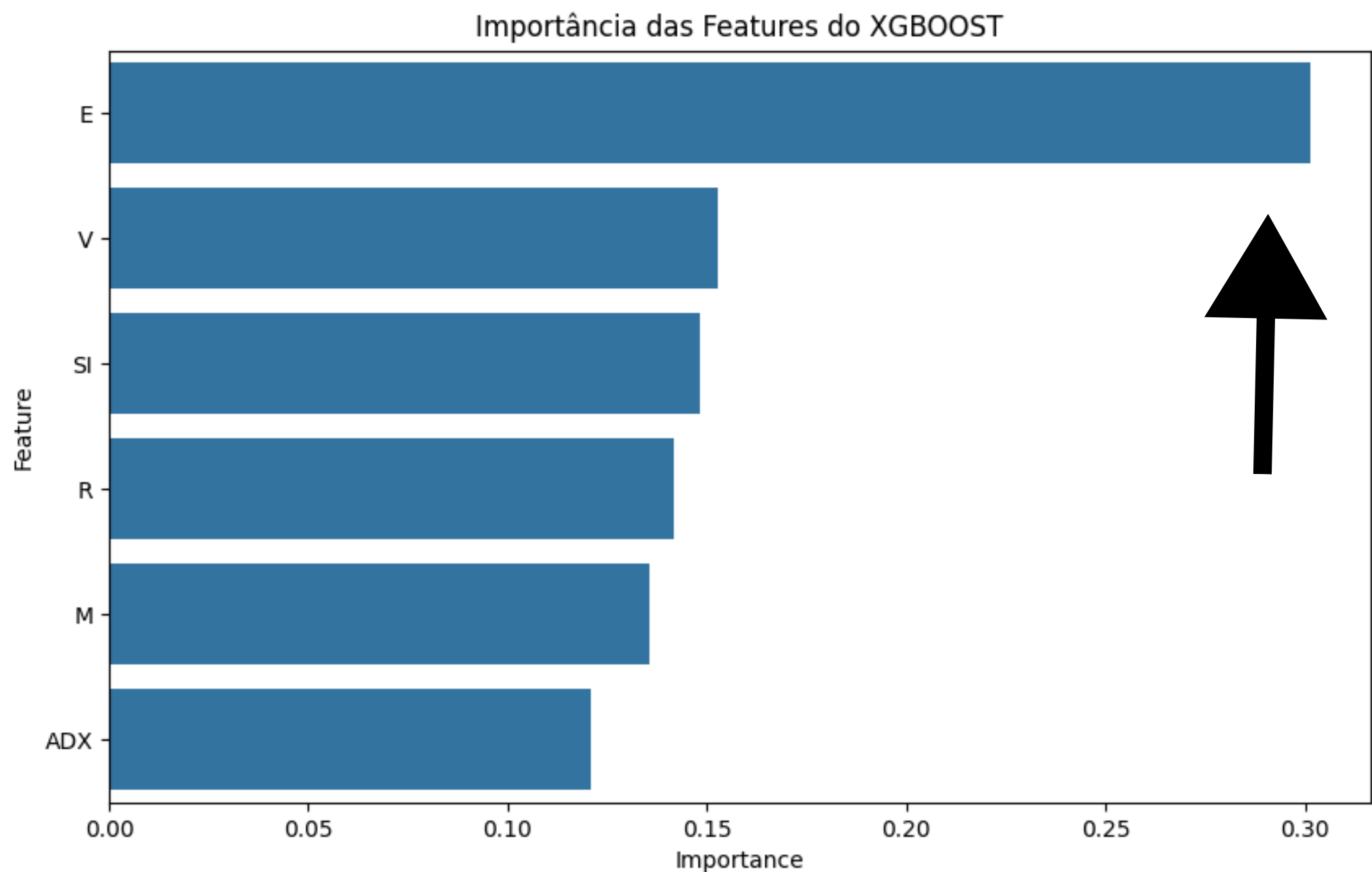
Classe	Precision	Recall	F1-Score	Suporte
Risco Baixo (0)	0.91	0.66	0.76	176
Risco Alto (1)	0.3	0.68	0.42	38

Métrica Geral	Valor
Accuracy	0.66
Macro F1	0.59
Weighted F1	0.7

Desempenho do modelo:

- **Precisão (Risco Baixo): ~91%** — o modelo foi altamente eficaz em identificar corretamente os dias tranquilos, com poucos falsos alarmes.
- **Precisão (Risco Alto): ~30%** — gerou muitos alertas falsos, como esperado ao priorizar segurança. F1
- **Recall (Risco Alto): ~68%** — capturou a maioria dos eventos de alta volatilidade.
- **F1 (Risco Alto): ~ 0.42** — equilíbrio moderado entre acertos e sensibilidade.
- **Accuracy geral: 0.66** — desempenho razoável considerando desbalanceamento da base.

CONCLUSÕES



Como vemos no gráfico à esquerda, a principal responsável por esse desempenho foi a Entropia de Boltzmann, que se destacou como a **feature mais importante do modelo**, superando métricas clássicas como MACD, RSI e ADX. Isso valida a hipótese central do projeto: **um conceito da Física Estatística pode ser transformado em um sinal preditivo eficaz para regimes de risco.**

Mesmo utilizando uma arquitetura simples de classificação, o modelo alcançou uma **média F1 ponderada de 0.70** e uma **acurácia geral de aproximadamente 70%**, o que representa um desempenho robusto considerando o desbalanceamento da base e a complexidade do mercado de criptoativos.

Esse resultado é especialmente relevante porque foi obtido sem o uso de redes neurais profundas ou dados externos, apenas com indicadores técnicos comumente utilizados e o diferencial da Entropia.

Dado a análise, é fácil ver que, mesmo em um sistema básico, é possível construir um modelo funcional de classificação de risco, a Entropia de Boltzmann sendo um forte potencializador - capaz de aprimorar modelos de risco em criptoativos e apoiar decisões estratégicas de gestão, antecipando momentos de alta volatilidade com maior estabilidade e eficiência.

Considerações Finais

Aprendizado

O projeto validou que a Entropia de Boltzmann (E) é um sinal preditivo superior , fundamental para a potencialização de modelos de gestão de risco em criptoativos.

Limitações e Escala

A complexidade do mercado de criptomoedas exige um modelo de previsão que incorpore múltiplos vetores de informação. Neste trabalho apenas visamos demonstrar a melhoria de um sistema utilizando entropia, contudo, deixamos claros que, para alcançar o nível máximo de acurácia em cenários reais, o sistema deve ser expandido para ser utilizada em modelos muito mais complexos, bebendo de conjunto de dados mais extenso, incluindo variáveis macroeconômicas , sinais de on-chain , e dados de sentimento de mercado.

Próximos Passos

O potencial preditivo e poder da Entropia deve ser explorado através da utilização em sistemas complexos . Sejam os próximos passos

- Integrar a Entropia em arquiteturas LSTM/RNN para capturar memória sequencial de longo prazo.
- Expandir uma base de recursos para incluir dados macroeconômicos (inflação, taxas de juros) e dados de sentimento de mercado.
- Testar modelos alternativos para o alvo (previsão da amplitude do risco) e aumentar a base de criptoativos analisados.

Uso de GEN AI

Destacamos um forte apoio da Gen Ai neste projeto, incluindo no apoio de revisão literária e correção de vieses no modelo.

