

Previsão de Tendência

IBOVESPA

Tech Challenge – Fase 2

Ferramentas: Python, yfinance, pandas, scikit-learn

Período: 2019–2025

Autor(es): Emmilly Katriny, Matheus Rodrigues



Objetivo

Desenvolver um modelo preditivo, utilizando aprendizado de máquina, para prever a tendência diária do IBOVESPA (alta ou baixa) com base em dados históricos.

Aquisição de dados

Tendência Histórica do IBOVESPA (2019–2025)



Coletados: **via yfinance**
Fonte: **IBOVESPA (^BVSP)**
Período: **2019 a 2025**

Colunas: **Abertura, Máxima, Mínima, Fechamento, Volume**

Engenharia de Atributos

Para que o modelo conseguisse entender os movimentos do **IBOVESPA**, foram criados indicadores técnicos a partir do preço e volume diário. Essas novas variáveis ajudam a capturar tendências, volatilidade e força do mercado.

Principais indicadores criados:

Variação percentual diária – mostra o ritmo de alta/baixa do mercado

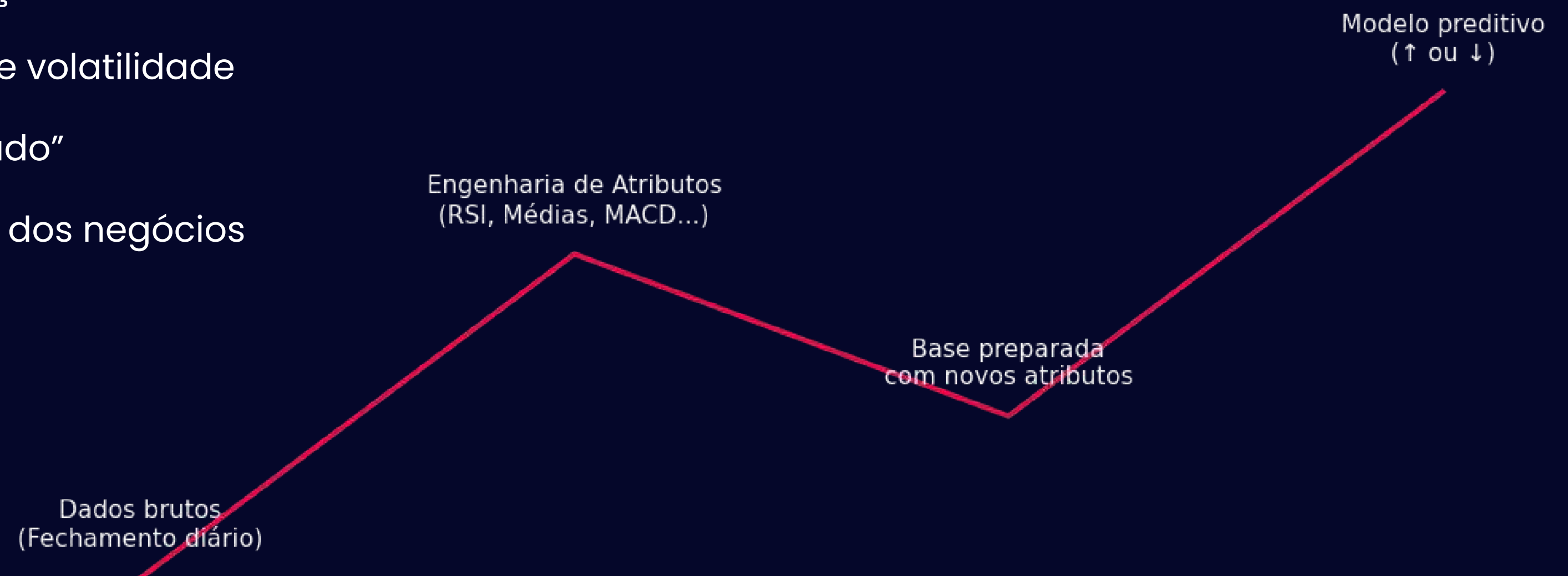
Médias móveis (5, 10, 20 dias) – revelam tendências de curto e médio prazo

RSI (Índice de Força Relativa) – mede a força das movimentações

MACD e Bandas de Bollinger – ajudam a identificar reversões e volatilidade

Lags (1 a 5 dias) – permitem que o modelo “olhe para o passado”

Z-score do volume e ATR – indicam intensidade e volatilidade dos negócios



Preparação e Modelagem

Target

1 se o preço subir, 0
se cair

Divisão

80% treino, 20% teste

**Normaliza
ção**

StandardScale

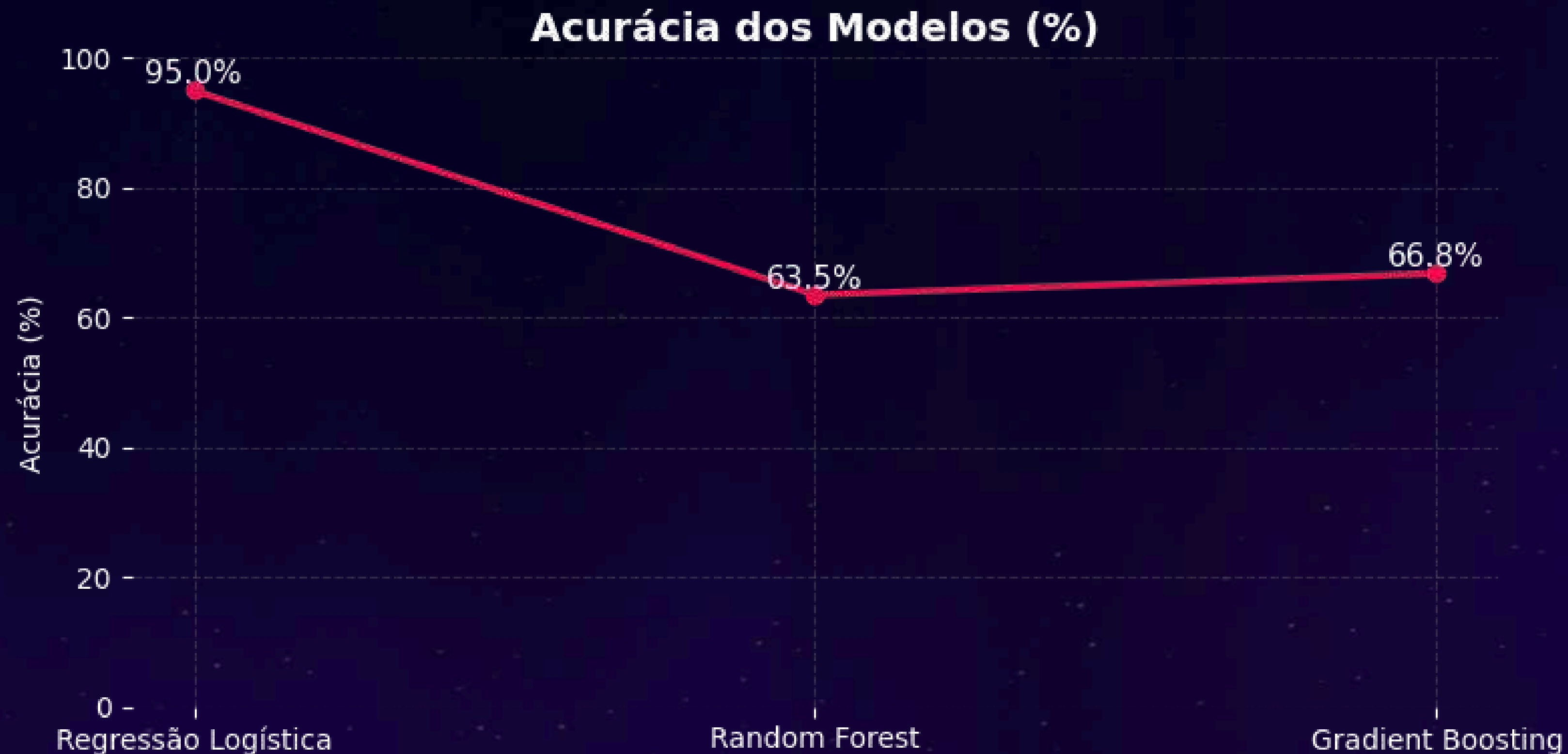
**Modelos
testados**

Regressão Logística

Random Forest

Gradient Boosting

Resultados



Melhor modelo: Regressão Logística

Interpretação: Apresentou o melhor desempenho (95%) e é o modelo mais simples e interpretável, ideal para prever tendências de alta ou baixa do índice IBOVESPA.

Justificativa Técnica: Modelo e Sequência Temporal

Modelo escolhido:

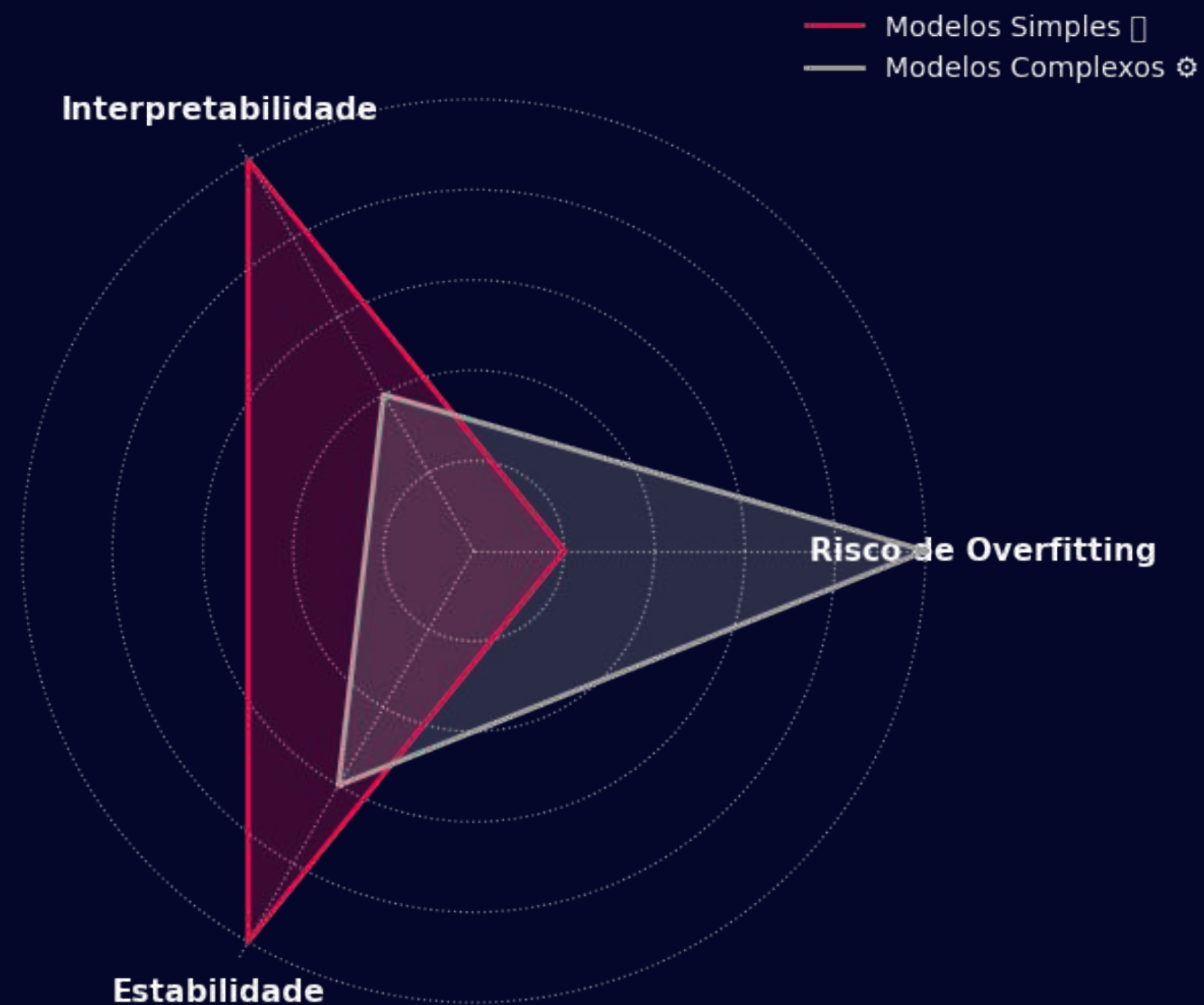
Regressão Logística → bom desempenho (95%) e fácil de interpretar.

Ideal para prever se o mercado vai subir ou cair (\uparrow/\downarrow).

Natureza sequencial dos dados:

- Os dados do **IBOVESPA** seguem uma ordem temporal.
- Foram usados lags (dias anteriores) e médias móveis para capturar tendências recentes.

Justificativa Técnica: Trade-offs e Decisão Final



Comparativo entre modelos:

Modelos Simples (Regressão Logística):

- Mais interpretáveis e estáveis
- Menor risco de overfitting
- Resultado consistente com alta acurácia (95%)

Modelos Complexos (Random Forest, Gradient Boosting):

- Capturam mais padrões, mas são sensíveis a ruídos
- Apresentaram acurácia inferior e maior risco de superajuste.

Conclusão:

Optamos pela Regressão Logística, que apresentou o melhor equilíbrio entre simplicidade, estabilidade e desempenho preditivo.

Conclusão

Resultados:

Modelo de **Regressão Logística** atingiu **95%** de acurácia na previsão de tendência (alta/baixa) do IBOVESPA.

Aplicação prática:

Pode auxiliar analistas e investidores em **decisões de compra** e venda com base em comportamento histórico.

Próximos passos:

Explorar modelos de **Deep Learning (LSTM)** para capturar dependências temporais mais complexas.

Integrar dados de **sentimento de mercado** (notícias e redes sociais) para **maior precisão** preditiva.