

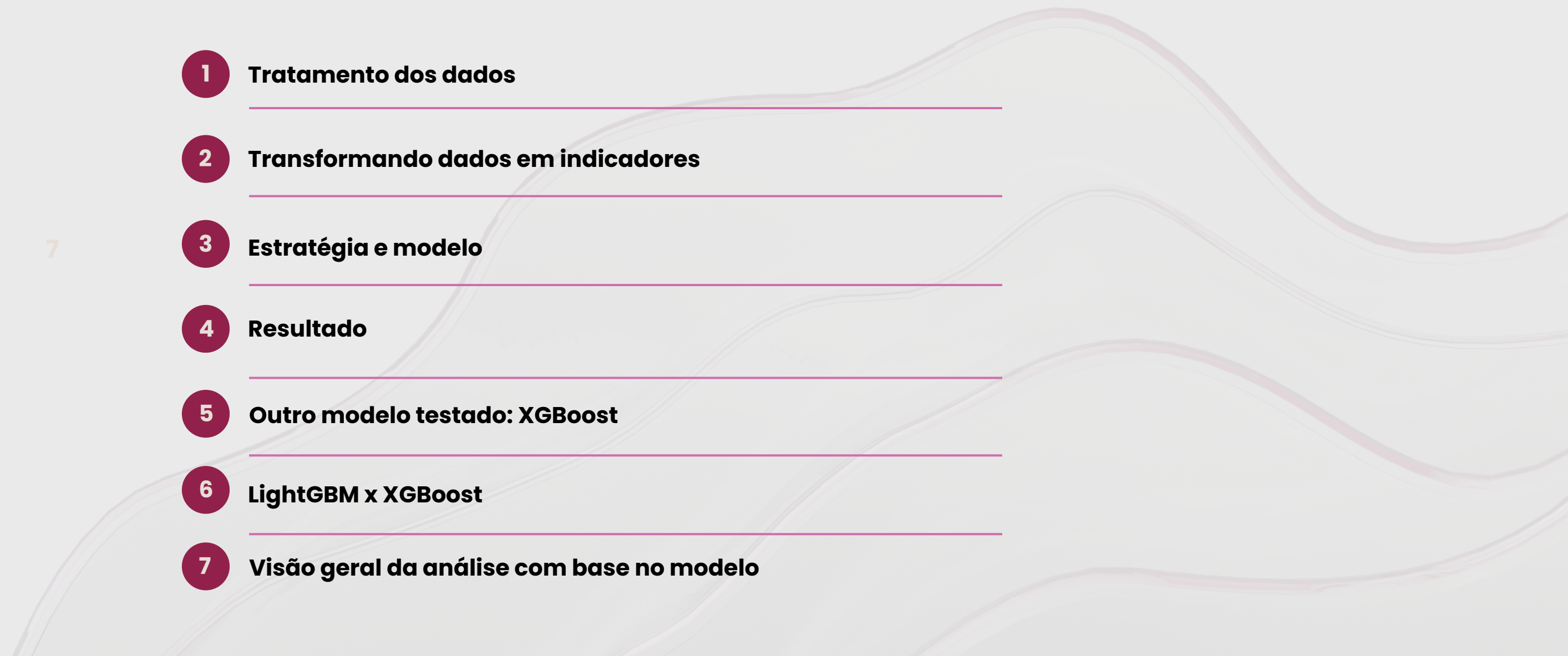
Transformando Dados em Estratégia

Previsão do IBOVESPA



TRANSFORMANDO DADOS EM ESTRATÉGIA

ANÁLISE DE MERCADO, TENDÊNCIAS E POSICIONAMENTO ESTRATÉGICO DO BRASIL

- 
- 1 **Tratamento dos dados**
 - 2 **Transformando dados em indicadores**
 - 3 **Estratégia e modelo**
 - 4 **Resultado**
 - 5 **Outro modelo testado: XGBoost**
 - 6 **LightGBM x XGBoost**
 - 7 **Visão geral da análise com base no modelo**

TRATAMENTO DOS DADOS



Utilizamos uma base histórica de **6 anos** do IBOVESPA

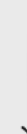


- Transformação dos dados de volume para valores numéricos (milhões e bilhões)
- Normalização dos nomes de colunas e organização da série temporal.

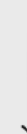


Criação da variável alvo **(target)** com base na variação do dia seguinte.

BASE 6 ANOS



- CONVERSÃO VOLUME
- NORMALIZAÇÃO
- ORGANIZAÇÃO TEMPORAL



TARGET

TRANSFORMANDO DADOS EM INDICADORES

Indicadores criados para capturar padrões de mercado:

- **Retorno diário** – Variação percentual do preço entre dias consecutivos;
- **Volatilidade Intradiária** – Medida da amplitude de oscilação dentro do dia;
- **Posição no Dia** – Relação do fechamento com mínima e máxima;
- **SMA 5 e 10** – Médias móveis simples de 5 e 10 períodos;
- **EMA 5 e 10** – Médias móveis exponenciais de 5 e 10 períodos;
- **Retorno 3 Dias** – Retorno acumulado em janela de 3 dias
- **TimeSeriesSplit** – simular ambiente real de predição, respeitando ordem temporal e evitando vazamento.



O dia da semana foi convertido em variável categórica para enriquecer a análise temporal.

ESTRATÉGIA E MODELO

Modelo escolhido: LightGBM (Gradient Boosting)

- **Alta performance em dados tabulares** – Ideal para séries financeiras com múltiplas variáveis numéricas e categóricas;
- **Eficiência e velocidade** – Treina rapidamente mesmo com grandes volumes de dados históricos;
- **Capacidade de capturar relações complexas** – Modela interações não lineares entre indicadores técnicos (médias móveis, volatilidade, retornos);
- **Ferramentas contra overfitting** – Suporte nativo a early stopping, regularização e ajuste automático de pesos em classes desbalanceadas.

Validação utilizada: TimeSeriesSplit (3 folds)

Mantém a ordem cronológica → simula previsões reais.

Vantagem Competitiva – Cada **fold** simula, de forma realista, o uso do modelo na prática: treinando com dados históricos e prevendo períodos futuros;

Prevenção de overfitting: early stopping (100 rounds)

- **Monitora o desempenho** no conjunto de validação durante o treino;
- **Interrompe automaticamente** após 100 iterações sem melhoria
- **Previne o overfitting**, evitando que o modelo memorize ruídos do treino
- Otimiza o ponto de parada, **equilibrando underfitting e overfitting**

Otimização da Decisão: Threshold (0.19)

Define o nível de confiança mínimo do modelo



aumenta a assertividade nas previsões de alta

Vantagem Competitiva – Permite **alinhar o modelo** à estratégia de negócio. Em vez de usar o padrão (50%), **ajustamos o gatilho** de decisão para que o **modelo** só sinalize "ALTA" quando a confiança na previsão é maior, **priorizando a qualidade** e a segurança das recomendações para o analista.

RESULTADO

FOLD 1

```
Fold 1
Training until validation scores don't improve for 100 rounds
[100]  valid_0's binary_logloss: 0.545205
[200]  valid_0's binary_logloss: 0.552327
Early stopping, best iteration is:
[118]  valid_0's binary_logloss: 0.541232
Acurácia Fold 1: 0.74751
```

FOLD 2

```
Fold 2
Training until validation scores don't improve for 100 rounds
[100]  valid_0's binary_logloss: 0.568809
[200]  valid_0's binary_logloss: 0.548577
[300]  valid_0's binary_logloss: 0.545009
Early stopping, best iteration is:
[259]  valid_0's binary_logloss: 0.544802
Acurácia Fold 2: 0.75083
```

FOLD 3

```
Fold 3
Training until validation scores don't improve for 100 rounds
[100]  valid_0's binary_logloss: 0.551105
[200]  valid_0's binary_logloss: 0.543698
Early stopping, best iteration is:
[180]  valid_0's binary_logloss: 0.540804
Acurácia Fold 3: 0.73422

Acurácia Final: 0.76667
```



Acurácia Final
77%

OUTRO MODELO TESTADO: XGBOOST

Modelo testado: XGBoost – Análise de Performance

Resultado por Fold:

Fold 1: 69.8% de acurácia

Fold 2: 73.4% de acurácia

Fold 3: 70.5% de acurácia

Média geral: 71.2%

Pontos Fortes:

Performance sólida – 71.2% supera baseline aleatório em +21 pontos

Capacidade de generalização – Funciona bem em dados não vistos

Robustez técnica – Early stopping preveniu overfitting efetivamente

Implementação estável – Algoritmo maduro e confiável

Pontos de Atenção:

Variabilidade temporal – Oscilação de 3.6 pontos entre folds (69.8% – 73.4%)

Consistência inferior – Maior sensibilidade a diferentes períodos de mercado

Performance geral – Ficou 3.1 pontos abaixo do LightGBM

--XGBoost--

Fold 1

[0] validation_0-logloss:0.69164

[100] validation_0-logloss:0.59797

[200] validation_0-logloss:0.57606

[300] validation_0-logloss:0.57305

[378] validation_0-logloss:0.58227

Acurácia Fold 1: 0.6981

Fold 2

[0] validation_0-logloss:0.69150

[100] validation_0-logloss:0.59694

[200] validation_0-logloss:0.56306

[300] validation_0-logloss:0.54528

[400] validation_0-logloss:0.53668

[500] validation_0-logloss:0.53333

[588] validation_0-logloss:0.53335

Acurácia Fold 2: 0.7338

Fold 3

[0] validation_0-logloss:0.69180

[100] validation_0-logloss:0.60348

[200] validation_0-logloss:0.57033

[300] validation_0-logloss:0.56085

[400] validation_0-logloss:0.56024

[500] validation_0-logloss:0.55851

[555] validation_0-logloss:0.56038

Acurácia Fold 3: 0.7045

LIGHTGBM VS XGBOOST

LIGHTGBM

Acurácia
77%

- **Impacto Financeiro:** Capta 3 em cada 4 movimentos do mercado;
- **Consistência:** Reduz risco operacional;
- **Eficiência:** Processamento mais rápido e econômico;
- **ROI estimado:** +R\$ 100.000 com ~75% de acerto.

XGBoost

Acurácia
70.5%

- **Menor Eficiência:** Requer mais recursos computacionais;
- **Maior Volatilidade:** Resultados menos estáveis;
- **Monitoramento Contínuo:** Necessita ajustes frequentes;
- **Impacto no ROI:** -R\$ 3.100 a cada R\$ 100.000 investidos.

Decisão Final

- LightGBM: escolha estratégica para o negócio;
- Maior retorno com menor risco;
- Consistência nas previsões;
- Vantagem competitiva: +3,1 pontos percentuais.

VISÃO GERAL DA ANÁLISE COM BASE NO MODELO

Com base nos resultados obtidos nas análises, nas quais validamos uma amostra dos últimos 30 dias, é possível concluir que o sucesso de um modelo preditivo se sustenta em 3 pilares:

- Precisão satisfatória nas previsões;
- Qualidade das previsões do modelo;
- Impacto financeiro positivo.

Considerando a alta taxa de acurácia do modelo e, ao mesmo tempo, o nível de eficiência limitado das previsões, é válido investir em um modelo preditivo que foque também em assertividade, mesmo que sua precisão seja menor que 75%.

É impossível obter taxas exatas sempre e com alta taxa de acertos, então o melhor cenário é equilibrar precisão e confiabilidade. Dessa forma é possível dispor de uma ferramenta tanto relevante quanto confiável nas análises do índice IBOVESPA.