

Precificação e Classificação de Equipamentos Industriais

Projeto Final - Ciência de Dados e Inteligência Artificial





Contexto do Negócio

O Desafio

Uma empresa do setor industrial precisa desenvolver um sistema de gerenciamento de equipamentos e instalações industriais. Essa empresa forneceu dados sobre as características dos equipamentos em seu portfólio e os preços de venda realizados.

Além disso, a empresa realiza uma classificação dos equipamentos para melhor atender às diferentes demandas de seus clientes.

O objetivo é criar algoritmos capazes de:



Prever o preço de venda de novos equipamentos com base nas suas características.



Classificar a categoria da máquina a partir de suas características.

Base de Dados

A empresa forneceu uma planilha com os seguintes dados dos equipamentos:

1

Variáveis Numéricicas

- Preço de venda
- Potência (kW)
- Peso
- Durabilidade
- Garantia

2

Variáveis Categóricas

- Classe (A, B, C, D, E)
- Tecnologia aplicada
- Status operacional
- Sistema de refrigeração

3

Variáveis Booleanas

- Necessidade de energia
- Requer manutenção
- Com software integrado
- Compatível com IoT

Base de Dados

preco	potencia	durabilidade	tecnologia	peso	necessidade_energia	requere_manutencao	iot	protecao_corrente	sistema_refri	garantia	com_software	classe	status
479500	7000	3	1	2 sim	nao	sim	nao	nao	0 nao	C	sem manutenção		
349300	4900	3	1	2 nao	nao	nao	nao	nao	0 nao	C	sem manutenção		
2248143	2581	2	1	1 sim	sim	nao	nao	nao	1 S	A	revisado		
350000	9500	3	1	2 sim	nao	nao	nao	nao	3 sim	C	sem manutenção		
377300	6930	4	1	2 nao	nao	nao	N	nao	1 nao	C	revisado		
19802643	2416	2	1	1 sim	nao	nao	nao	nao	2 nao	E	em manutenção		
374500	3480	2	1	1 sim	nao	nao	nao	nao	0 sim	E	em manutenção		
626500	6000	4	1	3 sim	sim	sim	nao	N	0 sim	C	sem manutenção		
350000	3512	2	1	1 sim	nao	nao	nao	nao	1 S	D	sem manutenção		
266000	3630	3	3	2 nao	sim	nao	nao	nao	0 nao	C	sem manutenção		
840000	8875	3	1	1 sim	nao	nao	N	N	1 nao	C	em manutenção		
448000	8250	3	1	1 S	N	nao	nao	N	0 nao	C	Revisado		
581000	5828	4	1	4 sim	sim	nao	nao	nao	0 nao	C	em manutenção		
210000	3360	2	1	1 sim	nao	nao	N	nao	1 nao	D	sem manutenção		
280000	3960	3	1	1 sim	nao	nao	nao	nao	0 nao	C	revisado		
329000	2500	2	1	1 nao	nao	nao	nao	sim	0 nao	D	sem manutenção		
315000	3986	2	2	1 nao	sim	sim	nao	nao	1 nao	C	sem manutenção		
260400	2910	2	1	1 nao	nao	nao	N	nao	0 nao	E	sem manutenção		
350000	12944	3	1	1 sim	nao	nao	nao	nao	0 nao	C	sem manutenção		
252000	3000	2	1	2 sim	nao	N	N	nao	0 nao	C	revisado		
461300	4510	4	2	2 sim	N	sim	nao	nao	0 N	C	Em manutenção		
350000	2135	3	2	2 nao	nao	nao	nao	nao	0 nao	C	sem manutenção		
399000	4050	2	1	2 sim	sim	sim	N	nao	0 sim	C	sem manutenção		
420000	4600	3	2	2 sim	nao	nao	nao	sim	1 nao	B	em manutenção		
357000	3640	2	1	1 sim	nao	nao	nao	nao	0 nao	D	sem manutenção		
308000	4960	2	1	1 S	nao	sim	nao	sim	0 N	C	sem manutenção		
619500	5500	3	2	1 sim	sim	sim	nao	nao	2 sim	B	revisado		
609000	8372	3	1	3 sim	nao	nao	N	sim	2 nao	C	sem		
668500	6600	2	2	4 sim	nao	sim	nao	nao	0 sim	C	revisado		

Qualidade dos Dados

Realizei uma análise inicial que revelou desafios típicos de dados coletados ao longo do tempo:



Dados Faltantes

Algumas colunas estavam com dados inválidos



Registros Incoerentes

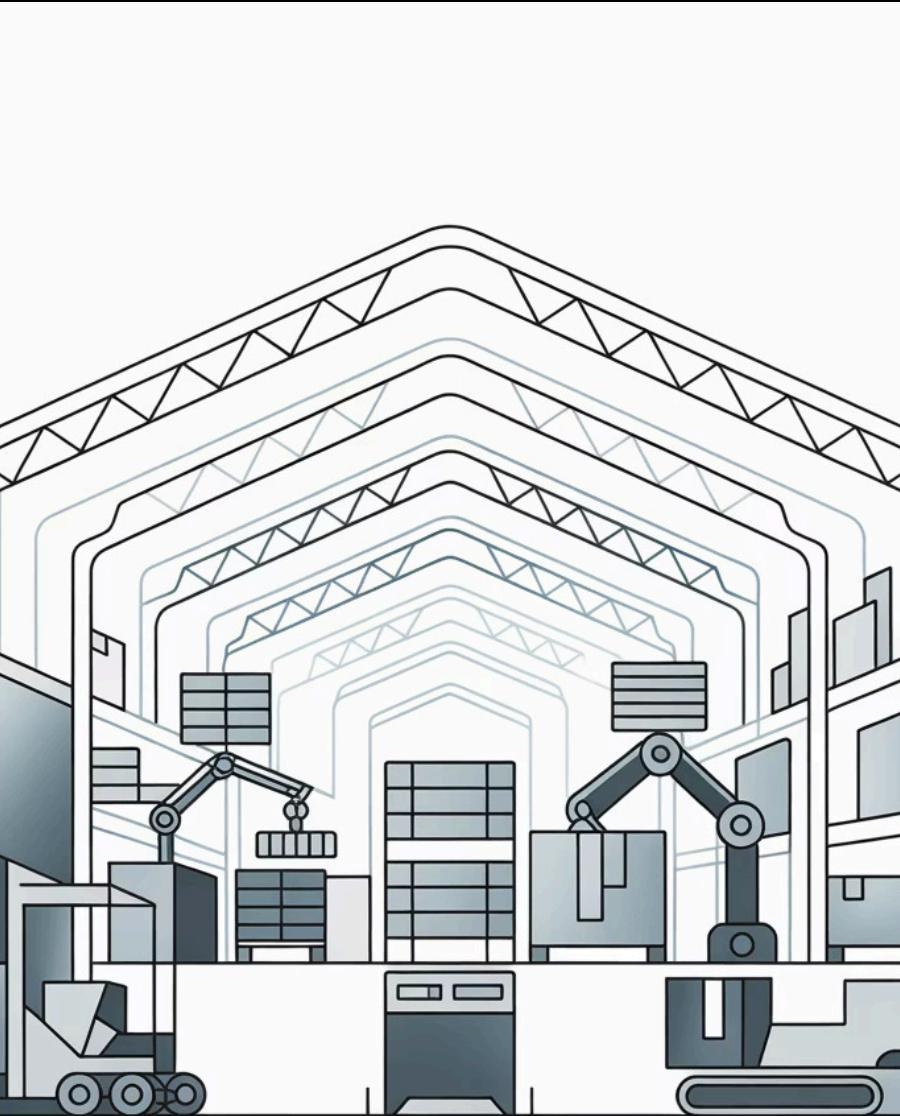
Encontradas inconsistências na especificação de equipamentos



Valores Inconsistentes

Campos com formatos variados (sim/não, S/N, 0/1)





Preparação dos Dados

01

Tratamento de Valores Ausentes

Remoção de dados específicos que não possuíam todas as colunas preenchidas

02

Identificação de Outliers

Remoção de valores atípicos que representavam erros de cadastro ou equipamentos com preços inconsistentes

03

Normalização de Strings

Padronização de textos, conversão para lowercase e remoção de caracteres especiais

04

Conversão de Booleanos

Transformação de todas variáveis sim/não para formato binário (0/1) para processamento algorítmico

05

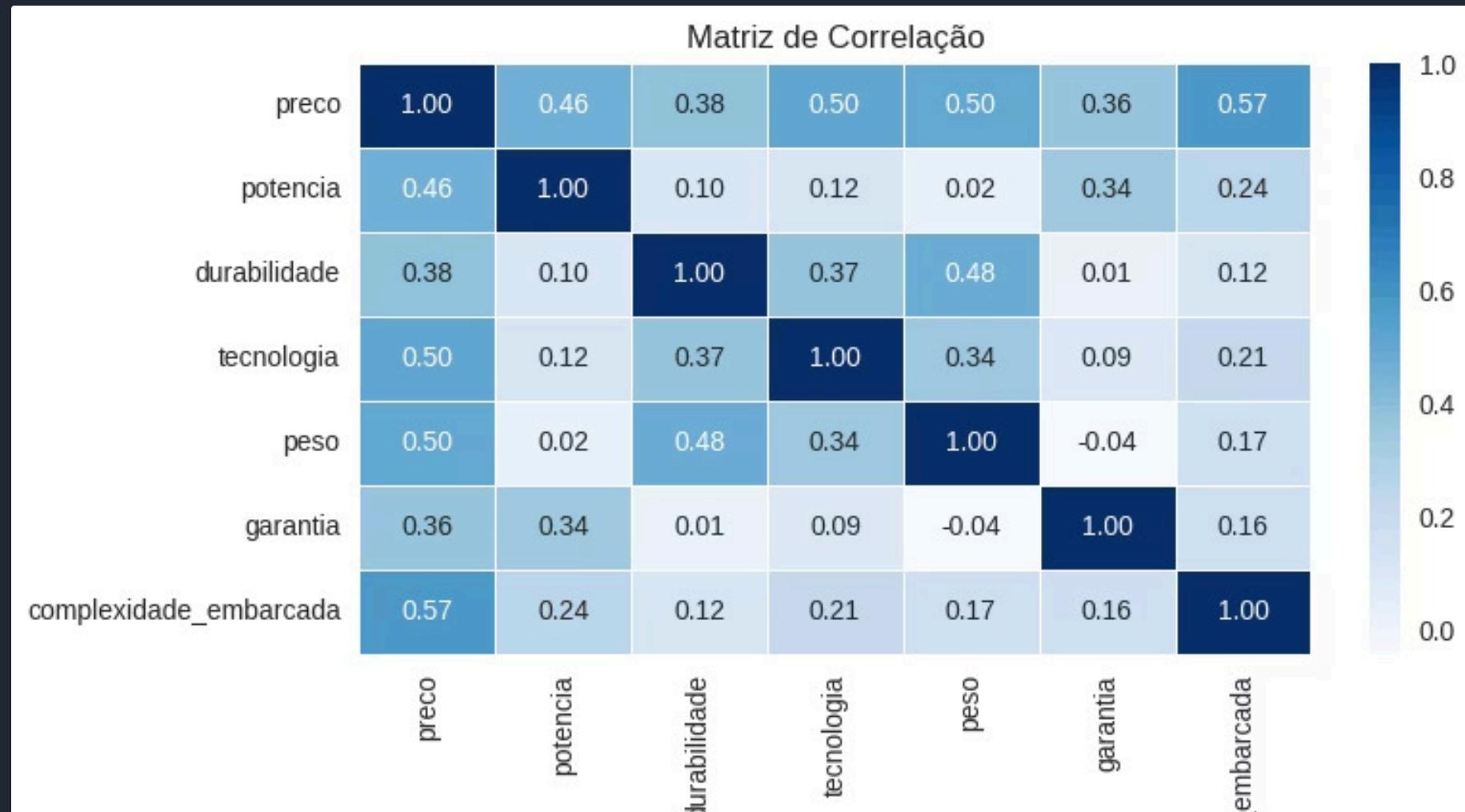
Tratamento de registros incoerentes

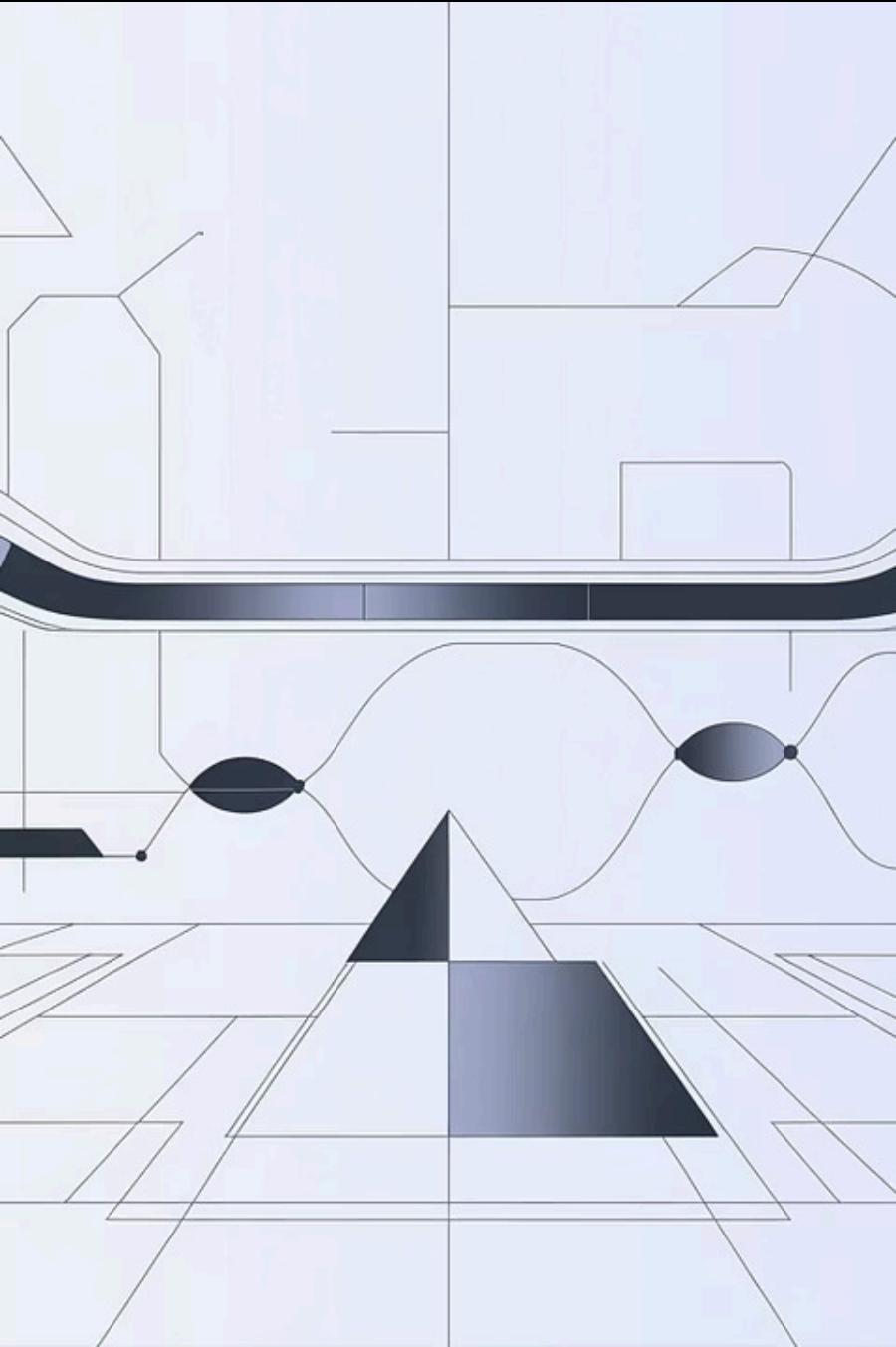
Remoção de registros com preços incoerentes

Base de Dados após preparação

preco	potencia	durabilidade	tecnologia	peso	garantia	classe	complexidade_embarcada
479500	7000	3	1	2	12	2	2
349300	4900	3	1	2	12	2	0
2248143	2581	2	1	1	24	0	2
350000	9500	3	1	2	48	2	2
377300	6930	4	1	2	24	2	0
374500	3480	2	1	1	12	4	2
626500	6000	4	1	3	12	2	3
350000	3512	2	1	1	24	3	2
266000	3630	3	3	2	12	2	0
448000	8250	3	1	1	12	2	1
581000	5828	4	1	4	12	2	1
210000	3360	2	1	1	24	3	1
280000	3960	3	1	1	12	2	1
329000	2500	2	1	1	12	3	1
315000	3986	2	2	1	24	2	1
260400	2910	2	1	1	12	4	0

Análise de Correlação





Previsão de Preço



Algoritmo Escolhido

Linear Regression - Selecionado por conta da correlação entre os dados



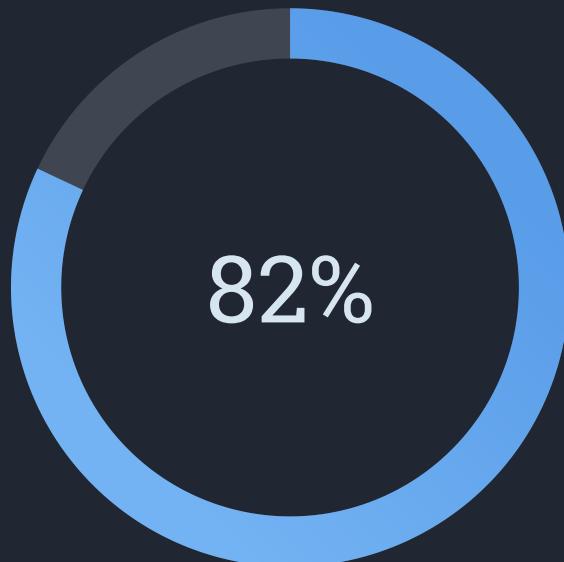
Divisão dos Dados

80% treino / 20% teste garantindo generalização adequada do modelo

Previsão de Preços

Performance do Modelo

O modelo de regressão demonstrou **boa capacidade preditiva**:

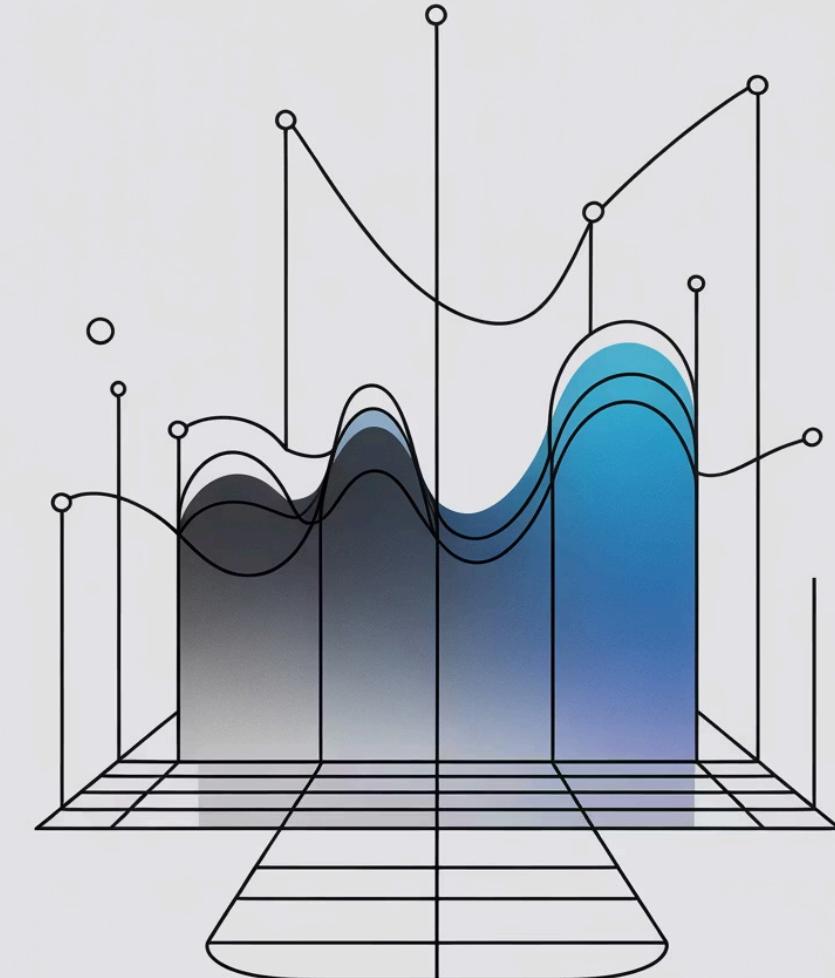


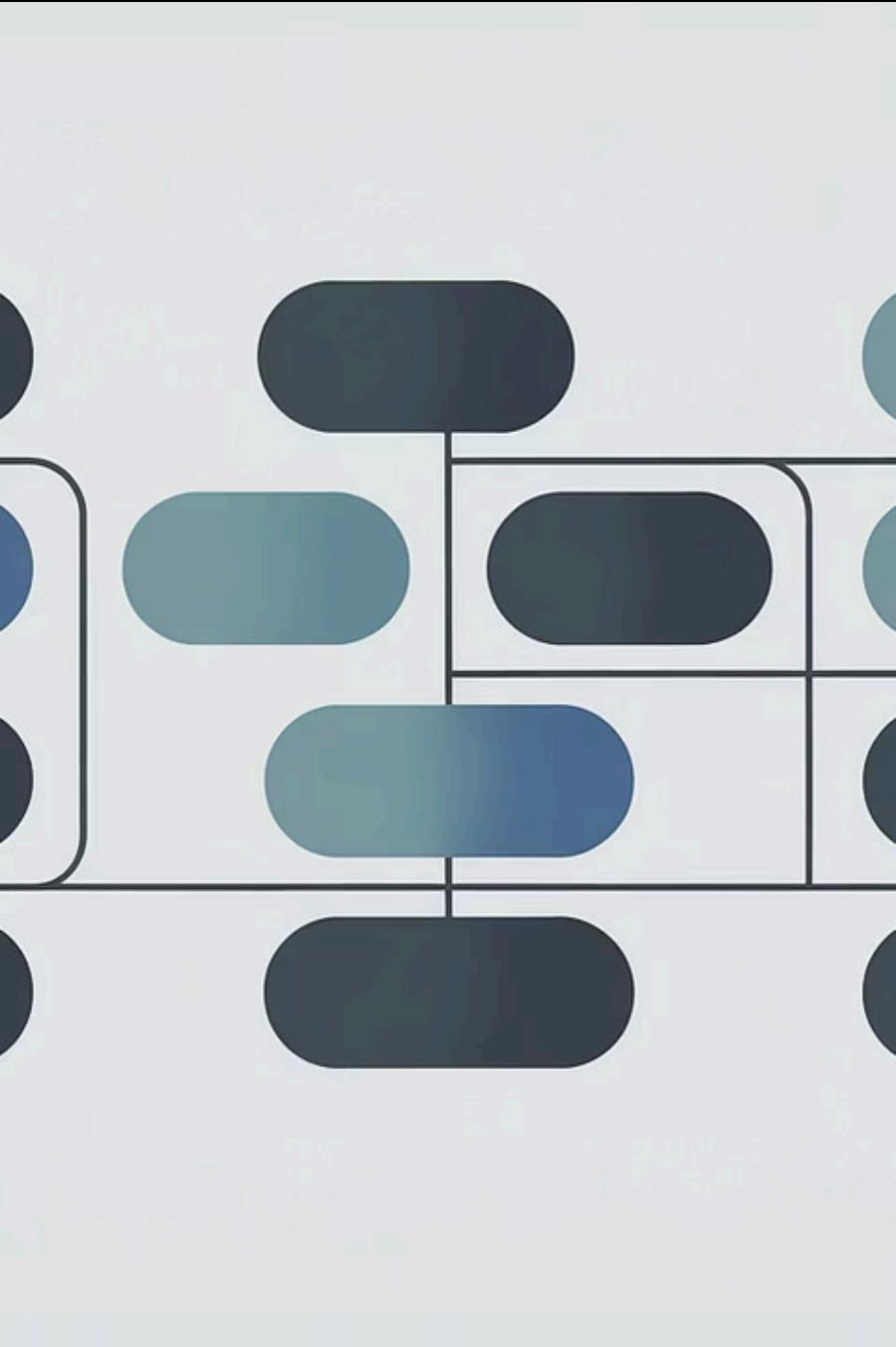
R² Score

4433.46

MAE

Erro Absoluto Médio





Classificação de Equipamentos

Random Forest Classifier

Teve o melhor desempenho com o conjunto de dados

Ao combinar múltiplas árvores de decisão, capturou padrões robustos para segmentação, permitindo uma classificação automática e mais precisa de novos equipamentos.

Classificação de Equipamentos

Performance do Modelo

O modelo demonstrou ótima capacidade preditiva:



Acurácia



Precisão



Avaliação dos Modelos



Regressão

- **MAE:** R\$ 4433,46
- **R²:** 0.82

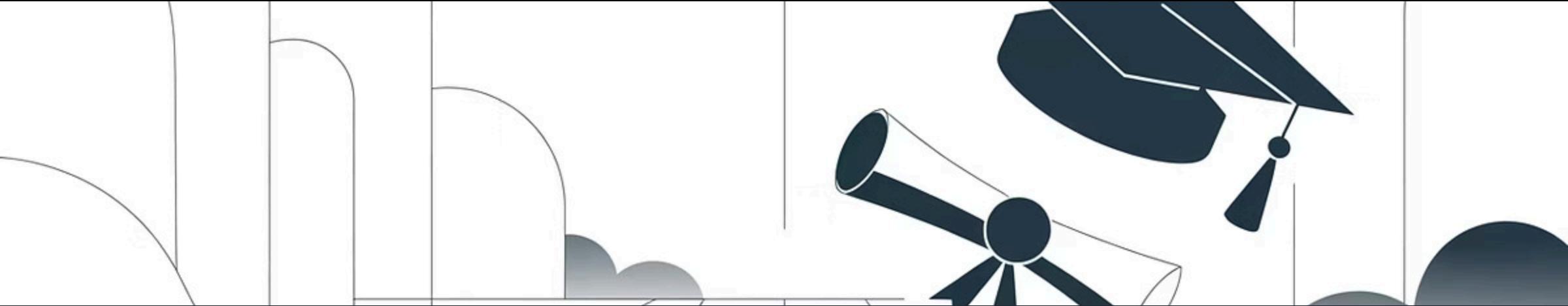
Precisão adequada para orçamentos, principalmente em máquinas com preço elevado.



Classificação

- **Acurácia:** 97%
- **Precisão:** 98%

Performance balanceada entre classes, praticamente pronto para utilização.



Próximos Passos

Auditória de Dados

Auditar o dataset inicial para corrigir erros de registro e garantir a integridade dos dados.

Otimização de Métricas

Melhorar de peso e durabilidade, segmentando por qualidade e quantidade de material utilizado.

Expansão do Dataset

Aumentar a quantidade de registros no dataset para enriquecer o treinamento dos modelos.

Conclusão

✓ Resultados Alcançados

Modelos com performance superior a 80% em todas as métricas, prontos para uso em produção

Valor para o Negócio

Redução no tempo de orçamentos e na classificação de equipamentos

Agradecimentos aos orientadores, colegas e à empresa parceira pela colaboração neste projeto.