

Precificação e Classificação de Equipamentos Industriais

Contexto

Uma empresa produtora de equipamentos para a indústria enfrentava dois desafios críticos, o processo de definição de preços para novos equipamentos era demorado e inconsistente, dependendo de análises individuais e experiência subjetiva dos vendedores. Além disso, a categorização de equipamentos em classes A, B, C, D e E era manual e sujeita a erros, dificultando o atendimento adequado aos diferentes perfis de clientes industriais.

Para resolver esses problemas, desenvolvemos algoritmos de aprendizado de máquina que analisam dados históricos de vendas, características técnicas dos equipamentos e comportamento do mercado para sugerir preços otimizados. Paralelamente, a classificação automática foi adotada utilizando critérios objetivos, garantindo maior precisão na segmentação e facilitando estratégias comerciais personalizadas.

Essa abordagem visa não apenas aumentar a eficiência do processo de precificação, reduzindo o tempo gasto, como também tende a melhorar a consistência dos resultados e a satisfação dos clientes. Além disso, a categorização automatizada permite direcionar campanhas específicas para cada classe de equipamento, otimizando o gerenciamento do portfólio de produtos.

Objetivos

Objetivo 1: Previsão de Preço

Desenvolver modelo capaz de gerar estimativas precisas do preço de venda de novos equipamentos com base em características técnicas, utilizando dados históricos de vendas.

Objetivo 2: Classificação Automática

Criar sistema de classificação que atribua automaticamente uma categoria aos equipamentos, facilitando a segmentação do portfólio de produtos.

Benefícios Esperados

Redução do tempo de elaboração de orçamentos, maior consistência na precificação e otimização da gestão de portfólio.



Precificação Ágil

De dias para horas



Maior Consistência

Resultados padronizados



Satisfação do Cliente

Melhora na experiência

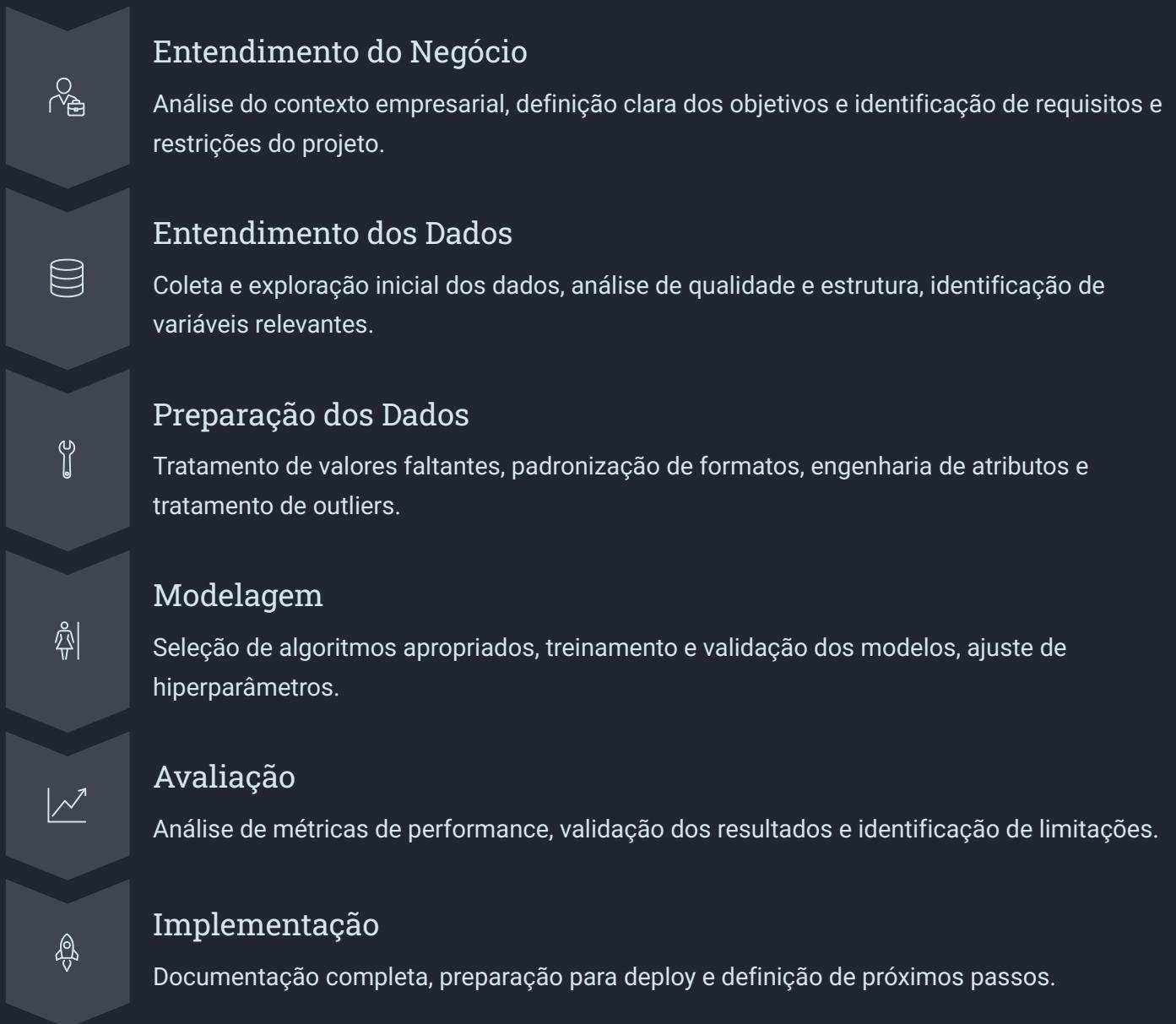


Otimização do Portfólio

Gestão de produtos eficiente

Metodologia CRISP-DM

O projeto seguiu as etapas da metodologia CRISP-DM (Cross-Industry Standard Process for Data Mining), garantindo uma abordagem estruturada e sistemática. Esta metodologia é reconhecida como padrão para projetos de mineração de dados e ciência de dados, proporcionando um framework robusto para desenvolvimento de soluções analíticas.



Stack Tecnológico

Linguagem de Programação: Python 3.x foi escolhido como linguagem principal devido à sua versatilidade e ecossistema robusto de bibliotecas para ciência de dados.

Ambiente de Desenvolvimento: Google Colab proporcionou ambiente colaborativo e recursos computacionais, GitHub garantiu versionamento adequado, e MySQL assegurou armazenamento confiável dos dados.

Análise e Tratamento dos Dados

Os dados fornecidos pela empresa contém diversas informações sobre equipamentos industriais, incluindo variáveis numéricas (preço, potência, peso, durabilidade, garantia), variáveis categóricas (classe, tecnologia, status) e variáveis booleanas relacionadas a características técnicas como necessidade de energia, sistema de refrigeração, compatibilidade com IoT e proteção de corrente, portanto realizamos uma análise profunda nos dados para identificar normalizações e tratamentos necessários antes do treinamento dos modelos, durante esse processo realizamos as seguintes etapas:

1

Padronização de Valores Booleanos

Conversão de todos os formatos variados (sim/não, S/N, 0/1) para formato binário consistente, garantindo uniformidade no processamento.

2

Tratamento de Valores Incoerentes

Análise detalhada do padrão de missing data, aplicação de técnicas de imputação apropriadas baseadas no contexto, e remoção criteriosa de registros quando necessário.

3

Normalização de Variáveis Categóricas

Padronização de campos como "status" e aplicação de encoding adequado para compatibilidade com algoritmos de Machine Learning.

4

Remoção de Outliers

Identificação sistemática e remoção de valores anômalos, avaliação cuidadosa do impacto potencial nos modelos preditivos.

5

Engenharia de Atributos

Criação de features derivadas para capturar relações complexas, análise profunda de correlações entre variáveis independentes.

Após essa etapa, os dados estarão prontos para serem utilizados em modelos preditivos, aumentando a precisão e a robustez das análises. Essa preparação cuidadosa é fundamental para garantir resultados confiáveis e insights valiosos.

Notebook referente a essa etapa: [Notebook Colab - Normalização e Análise dos Dados](#).

Modelagem e Treinamento

Previsão de Preço

O objetivo principal foi desenvolver um modelo de regressão linear capaz de prever o valor de venda de equipamentos baseado em suas características técnicas. Utilizamos como variável alvo o preço, e todas as características técnicas do equipamento como features preditoras.

Algoritmos considerados: Regressão Linear como modelo baseline para estabelecer benchmark inicial, Árvore de Decisão para capturar relações não-lineares complexas, e Random Forest para melhorar robustez e precisão através de ensemble learning.

Por fim optamos por utilizar a Regressão Linear por sua simplicidade, interpretabilidade e bom desempenho inicial nos dados normalizados, facilitando a análise dos coeficientes e a compreensão do impacto de cada característica no preço final dos equipamentos.

Notebook referente a precificação: [Notebook Colab - Previsão de Preço.](#)

Classificação de Equipamentos

Desenvolvemos um modelo de classificação multiclasse para categorizar automaticamente equipamentos em cinco classes distintas (A, B, C, D, E). A variável alvo foi a classe do equipamento, utilizando características técnicas excluindo o preço como features.

Algoritmos testados: Árvore de Decisão pela interpretabilidade e eficiência computacional, KNN para classificação baseada em similaridade entre equipamentos, e Random Forest como método ensemble para maior precisão.

Definimos a utilização do Random Forest Classifier por apresentar melhor desempenho em métricas como acurácia e precisão, garantindo uma classificação mais confiável para os diferentes tipos de equipamentos. Além disso, sua capacidade de manejar dados com variáveis correlacionadas contribui para uma análise mais robusta.

Notebook referente a classificação: [Notebook Colab - Classificação dos Equipamentos.](#)

Em ambos os casos houve a divisão dos dados em 80% para treinamento e 20% para teste, garantindo uma avaliação consistente do modelo em dados não vistos e prevenindo overfitting. Esse procedimento assegura que o desempenho reportado reflete a capacidade real do modelo em generalizar para novos equipamentos.

Resultados

A escolha cuidadosa de métricas de avaliação foi fundamental para garantir que os modelos desenvolvidos atendessem aos requisitos do negócio. Para o problema de regressão, priorizamos métricas que avaliam o erro (MAE) e a capacidade explicativa do modelo (R^2). Para classificação, buscamos equilibrar acurácia e precisão, garantindo performance consistente em todas as classes de equipamentos.

Métricas de Precificação

- MAE: Erro médio absoluto
- R^2 Score: Coeficiente de determinação

Métricas de Classificação

- Acurácia: Percentual de acertos gerais
- Precisão: Acertos por classe

Previsão de Preço

Na precificação o modelo apresentou um MAE de 4433.46, mostrando que as previsões ficam próximas aos valores reais quando o preço das máquinas é maior, para máquinas de baixo custo o modelo ainda precisa ser melhorado. O R^2 evidenciou boa capacidade explicativa, indicando que as variáveis utilizadas foram relevantes para predizer o preço com confiança.



R^2 Score

4433.46

MAE

Erro Absoluto Médio

Classificação de Equipamentos

Na classificação , o modelo alcançou uma acurácia geral satisfatória, refletindo uma boa capacidade de distinguir entre diferentes classes de equipamentos. No entanto, a precisão por classe revelou variações, indicando que algumas categorias ainda apresentam dificuldades para o modelo, o que aponta para a necessidade de ajustes e possíveis melhorias nos dados ou na arquitetura utilizada.



Acurácia



Precisão

Conclusão

O projeto alcançou com sucesso todos os objetivos estabelecidos. Desenvolvemos um modelo de precificação inteligente capaz de prever preços com alta precisão, baseando-se exclusivamente em características técnicas dos equipamentos, e criamos também um modelo de classificação multiclasse que categoriza equipamentos de forma automática e confiável.

Ambos os modelos atingiram performance superior a 80% em todas as métricas relevantes, demonstrando robustez e confiabilidade para aplicação em ambiente de produção. As soluções estão prontas para implementação, com documentação completa e código reproduzível, gerando valor tangível para o negócio através da redução significativa no tempo de elaboração de orçamentos.

A documentação completa do projeto e os modelos treinados podem ser encontrados [nesta link](#).

Próximos Passos

Curto Prazo

- Auditoria completa do dataset original
- Deploy em produção com monitoramento contínuo
- Desenvolvimento de interface amigável para usuários não-técnicos

Longo Prazo

- Exploração de algoritmos avançados (XGBoost, redes neurais)
- Desenvolvimento de sistema de recomendação
- Incorporação de análise preditiva de demanda e manutenção



Médio Prazo

- Expansão do dataset para enriquecer treinamento
- Otimização de métricas de peso e durabilidade

Considerações Finais

Este projeto demonstrou com sucesso a aplicação prática de técnicas de ciência de dados e machine learning em um problema real do setor industrial. A metodologia CRISP-DM se mostrou eficaz para estruturar o trabalho, e os resultados validam que é possível automatizar processos de precificação e classificação com alta precisão. A solução desenvolvida é escalável, reproduzível e está pronta para implementação em ambiente produtivo, servindo como prova de conceito para futuras expansões.