

---

## 🎓 Projeto Final da Trilha de Ciência de Dados:

### Predição de Turnover (Rotatividade) de Funcionários com Machine Learning

---

#### 👋 Introdução e Boas-Vindas

Parabéns, participante do **Bootcamp [RE]Start – Trilha de Cientista de Dados**! Após semanas de aprendizado intenso – da leitura de dados até a construção de modelos complexos de machine learning. Chegou a hora de aplicar tudo o que você aprendeu neste **projeto** prático, completo e alinhado às demandas do mercado.

Neste desafio, você será responsável por conduzir todas as etapas de um projeto real de ciência de dados com base no dataset **“IBM HR Analytics Attrition & Performance”**, disponível no Kaggle (<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/data>). O objetivo é construir um modelo preditivo que identifique padrões de rotatividade (*attrition*) e ajude a empresa fictícia **Data Girls S.A** a prever quais colaboradores têm maior propensão a sair da organização.

Este projeto é a sua oportunidade de colocar a mão na massa com um problema de negócio real, desenvolver seu raciocínio analítico, praticar técnicas de modelagem e apresentar resultados com impacto. Vamos juntas? 💜

---

#### 📁 Etapas orientadoras do Projeto

##### 1. Leitura e Exploração Inicial dos Dados

##### 2. Limpeza e Preparação dos Dados

##### 3. Análise Exploratória (EDA)

##### 4. Modelagem Preditiva

- Treine pelo menos **dois modelos supervisionados** diferente;
- Realize tuning dos hiperparâmetros;
- Compare os modelos com base em métricas de classificação.

##### 5. Avaliação dos Modelos

- Analise os erros do modelo (falsos positivos, falsos negativos) e o impacto para o negócio;
- Avalie a importância das variáveis nos modelos.

##### 6. Relatório/Apresentação de Resultados

- Elabore uma documentação explicando todo o pipeline desenvolvido (storytelling):
  - Objetivo do projeto

- Etapas da análise
- Escolha do(s) modelo(s)
- Principais achados e insights
- Recomendações práticas com base nos resultados

---

### Perguntas Norteadoras de Negócio

- Quais características mais influenciam na rotatividade de funcionários?
- Qual perfil de colaborador tem maior propensão a sair da empresa?
- Um modelo de machine learning pode prever com boa precisão a saída de um funcionário?
- Que ações a empresa pode tomar com base nessas previsões?

---

### Opcional – Bônus

Implemente uma **explicação do modelo** usando ferramentas como **SHAP** ou **LIME** para mostrar de forma interpretável do porquê o modelo toma certas decisões. Isso pode ajudar a traduzir os resultados para o RH de forma mais clara.

---

### Critérios de Avaliação

Critério	Peso	Detalhamento
Pipeline completo	20%	Leitura, limpeza, EDA, modelagem e avaliação realizados com clareza
Qualidade da modelagem	25%	Modelos adequados, tuning e validação bem aplicados
Interpretação dos resultados e insights	20%	Boa análise de métricas, perguntas norteadoras respondidas com profundidade e embasamento
Recomendações	10%	Ações práticas com base nos achados
Organização do código e documentação	15%	Código limpo, comentários claros e relatório bem estruturado
Criatividade	10%	Criatividade no desenvolvimento da solução
Bônus	+10%	Desenvolveu o opcional e incluiu na documentação

Critério	Peso	Detalhamento
TOTAL	100% (110%)	NOTA 10 (podendo alcançar 11)

---


#### Prazo e Entrega

- **Data limite:** 10/08/2025, às 23h59
  - **Entregáveis:**
    - Notebook (.ipynb)
    - Documentação em PDF ou no README do GitHub
    - Colocar todos os entregáveis no GitHub e submeter o link no formulário de entrega (<https://forms.gle/Ln4CBzFUM9jfLybQA>).
- 

#### Dicas Finais

- Utilize o acesso as aulas para apoiá-la!
  - Discussão ao decorrer da documentação focada em Storytelling.
  - Confira os outros projetos presentes na página do dataset no Kaggle.
  - Troque ideias com as outras participantes no Discord.
- 

Estamos muito orgulhosas da sua jornada até aqui! Este é o momento de consolidar seus aprendizados e brilhar com um projeto de impacto. Explore, questione, teste, avalie e, acima de tudo, divirta-se com os dados!

Com carinho, 

Equipe [RE]Start e Liderança Data Girls

---