

## Ficha Técnica do Projeto

**Título do Projeto:** Previsão de Rotatividade de Funcionários com Machine Learning

**Objetivo:** Desenvolver um modelo de aprendizado de máquina supervisionado capaz de prever se um funcionário deixará a empresa.

### Ferramentas e Tecnologias:

- **Linguagem:** Python
- **Bibliotecas:** Sklearn, Pandas, NumPy, Matplotlib, seaborn, Plotly, xgboost, tensorflow
- **Plataforma:** Google Colab

**Dados:** Os dados utilizados neste projeto foram coletados pelo departamento de RH de uma empresa. A tabela de dados contém as seguintes variáveis:

- Age (Idade)
- Attrition (0 = não saiu, 1 = saiu)
- BusinessTravel (Frequência de viagens)
- Department (Departamento)
- DistanceFromHome (Distância de casa)
- Education (Nível de escolaridade)
- EducationField (Área de estudo)
- EmployeeCount (Contagem de funcionários)
- EmployeeID (ID do funcionário)
- Gender (Sexo)
- JobLevel (Nível do trabalho)
- JobRole (Função)
- MaritalStatus (Estado civil)
- MonthlyIncome (Renda mensal)
- NumCompaniesWorked (Número de empresas trabalhadas)
- Over18 (Maior de 18 anos)
- PercentSalaryHike (Aumento percentual do salário)
- StandardHours (Horas padrão de trabalho)
- StockOptionLevel (Participação em ações)
- TotalWorkingYears (Total de anos trabalhados)
- TrainingTimesLastYear (Tempo de treinamento no último ano)
- YearsAtCompany (Anos na empresa)
- YearsSinceLastPromotion (Anos desde a última promoção)
- YearsWithCurrManager (Anos com o atual gerente)

## Cronograma

### Dia 1 : Coleta e Preparação dos Dados

- Carregar os dados e explorar sua estrutura.
- Tratar valores ausentes e realizar a codificação de variáveis categóricas.

## **Dia 2: Análise Exploratória dos Dados (EDA)**

- Realizar análises descritivas.
- Criar gráficos interativos utilizando Plotly para explorar distribuições e correlações.

## **Dia 3: Desenvolvimento do Modelo (Parte 1)**

- Dividir os dados em conjuntos de treino e teste.
- Treinar um modelo básico de Random Forest e árvore de decisão para ter uma linha de base.

## **Dia 4: Desenvolvimento do Modelo (Parte 2)**

- Implementar e avaliar modelos mais avançados, como XGBoost e TensorFlow.
- Realizar a otimização de hiperparâmetros.

## **Dia 5: Avaliação e Otimização do Modelo**

- Avaliar os modelos com diferentes métricas.
- Realizar a otimização de hiperparâmetros usando GridSearchCV.

## **Dia 6: Resultados e Conclusões**

- Interpretar os resultados e comparar o desempenho dos diferentes modelos.
- Redigir as conclusões e recomendações baseadas nos resultados.

## **Dia 7: Preparação da Apresentação**

- Criar slides para a apresentação.
- Incluir visualizações e resultados do modelo.
- Ensaiai a apresentação.

# Análise de Previsão de Rotatividade de Funcionários

## Introdução

Hoje, vamos discutir como um modelo de aprendizado de máquina pode ajudar a prever se um funcionário deixará a empresa. Usamos dados de vários atributos dos funcionários para criar um modelo que nos ajude a identificar possíveis saídas.

## Dados e Modelos Usados

- **Dados:** Temos informações de 4.410 funcionários, incluindo idade, nível de cargo, salário, e muito mais.
- **Modelos Testados:**
  - **Árvore de Decisão**
  - **Random Forest**
  - **XGBoost** (o melhor desempenho)
  - **Rede Neural Sequencial**

## Desempenho do Modelo XGBoost

O modelo XGBoost teve o melhor desempenho, com as seguintes métricas:

- **Acurácia:** 99%
- **Precisão:** 99% para não saída, 100% para saída
- **Recall:** 100% para não saída, 96% para saída
- **F1-Score:** 0.98 (média para ambas as classes)

Além disso, a **curva ROC** foi excelente, com um valor de 0.998, indicando que o modelo é muito eficaz em diferenciar entre funcionários que sairão e os que ficarão.

## Curva de Aprendizado

A curva de aprendizado mostra que o modelo está aprendendo bem com o tempo, atingindo um alto nível de precisão. Isso significa que o modelo está treinado de forma eficaz e pode generalizar bem para novos dados.

## Top 10 Características Mais Importantes

Aqui estão as características que mais influenciam a previsão de saída dos funcionários:

1. **Estado Civil - Divorciado:** 6.57%
2. **Campo de Educação - Grau Técnico:** 5.53%
3. **Campo de Educação - Ciências da Vida:** 4.58%
4. **Função - Técnico de Laboratório:** 4.00%
5. **Viagem a Negócios - Não Viaja:** 3.95%
6. **Treinamento no Ano Anterior:** 3.57%
7. **Total de Anos Trabalhados:** 3.43%
8. **Anos na Empresa:** 3.22%
9. **Anos com o Atual Gestor:** 3.00%
10. **Número de Empresas Trabalhadas:** 2.93%

Essas características são fundamentais para entender os fatores que podem levar a uma saída.

### **Recomendações**

1. **Foco no Estado Civil:** Divorciados podem estar mais propensos a deixar a empresa. Considere programas de suporte para essa situação.
2. **Avaliação do Campo de Educação:** Funcionários com grau técnico têm maior probabilidade de sair. Reavalie as estratégias de retenção para esses funcionários.
3. **Política de Viagens:** Funcionários que não viajam podem ter maior taxa de saída. Reavalie a política de viagens e ofereça alternativas para engajamento.

### **Considerações Finais**

- O modelo XGBoost se mostrou eficaz na previsão de rotatividade, com uma alta taxa de acerto.
- Focar nas principais características pode ajudar a melhorar a retenção.
- Continuar monitorando e ajustando o modelo com novos dados pode aumentar ainda mais a precisão das previsões.

Se houver alguma dúvida ou necessidade de mais detalhes, estou à disposição!