

Churn de Colaboradores

Como Reter Talentos com a Ciência de Dados

Cientista de Dados: Cristiane Thiel

Cliente: HR Insights

Projeto de Machine Learning

Data: Junho de 2025

O Desafio da Retenção de Talentos

4.410

Colaboradores

16%

Taxa de Churn

Por que reter talentos importa?

Cada saída gera custos com seleção, treinamento e perda de produtividade.

Mas e se pudéssemos antecipar esses desligamentos e agir antes que aconteçam?



Metodologia do Projeto

1. Tratamento e Preparação dos Dados
2. Análise Exploratória (EDA)
3. Seleção de Variáveis e Engenharia de Features
4. Divisão em Treino/Teste e Validação Cruzada
5. Treinamento dos Modelos (XGBoost e Random Forest)
6. Avaliação, Otimização e Benchmark
7. Curvas ROC, Precision-Recall e Learning Curve
8. Análise de Fairness (Sem Viés e Decisões Justas)



Dados Confiáveis = Decisões Eficientes

Para garantir a qualidade da análise e obter resultados precisos:

- Regras de negócio aplicadas aos valores nulos:
 - Correção de relações lógicas entre experiência e empresas anteriores para aumentar confiabilidade.
- Nulos sem possibilidade de correção:
 - 19 registros excluídos (0.43% do total) para evitar distorções.
- Variáveis irrelevantes descartadas:
 - EmployeeCount, Over18 e StandardHours não agregavam valor preditivo.



Quais Fatores Influenciam o Churn?

Para responder a essa pergunta, analisei as variáveis que os modelos preditivos XGBoost e Random Forest consideram mais importantes.

Perfil Demográfico e Financeiro

Idade e renda mensal aparecem consistentemente como fatores críticos, indicando que colaboradores **mais jovens** e com menor renda têm maior risco de saída.

Experiência Profissional

O tempo total de carreira é unanimidade entre os modelos, sugerindo que o **momento de carreira** influencia mais que o tempo na empresa.

Relacionamento com a Liderança

O tempo com o gestor atual, embora com pesos diferentes entre os modelos, reforça o **papel da liderança** na retenção.

Colaboradores que viajam com frequência apresentam o maior risco de saída

A taxa de churn entre quem viaja frequentemente a trabalho chega a **25%**, mais do que o triplo da taxa entre colaboradores que não viajam (8%).

O desgaste físico e emocional das viagens constantes impacta diretamente a permanência desses profissionais.

Implemente políticas de flexibilidade pós-viagem, pausas programadas e alternativas de trabalho remoto para reduzir o risco de rotatividade.



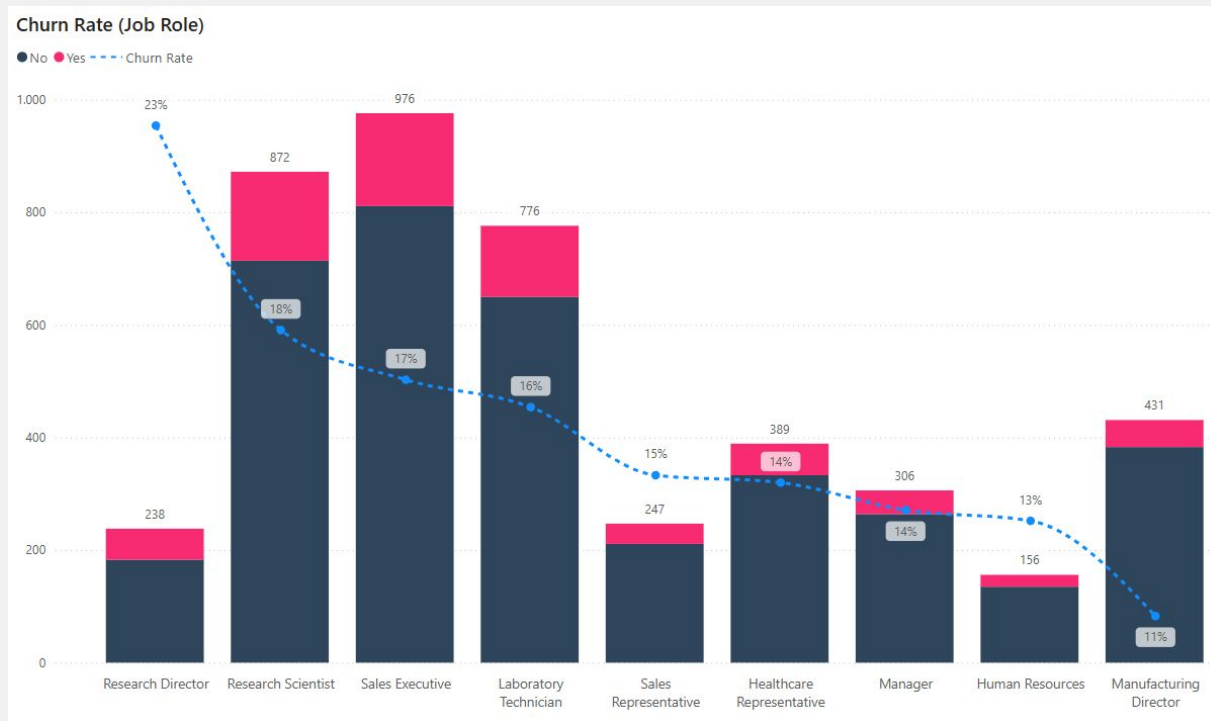
CHURN POR DEPARTAMENTO

Recursos Humanos tem a menor quantidade de colaboradores, mas apresenta a **maior taxa de churn**.

Atenção especial deve ser dada aos profissionais de **Pesquisa e Desenvolvimento**, pois perdas nessa área afetam diretamente a inovação e a competitividade da empresa.

Fortalecer a cultura organizacional e o suporte à equipe de RH, já que esse setor é estratégico para o bem-estar de toda a empresa.





Os maiores riscos de churn em cargos estão nas **funções técnicas de P&D** e nas **funções comerciais**, que combinam alta taxa e grande volume de desligamentos. Focar retenção nesses cargos é essencial para proteger a base de conhecimento, a inovação e os resultados comerciais da empresa.

A rotatividade é mais alta entre profissionais de nível júnior e pleno, quando comparados a cargos sêniores ou de liderança

Para reter esses talentos, é fundamental ir além do aumento salarial isolado.

Invista em planos de carreira estruturados, programas de mentoria, oportunidades reais de crescimento e reconhecimento contínuo.

Benefícios personalizados e acompanhamento próximo da liderança também são essenciais para fortalecer o vínculo e reduzir a rotatividade.



GARANTINDO DECISÕES JUSTAS = SEM VIÉS

A **análise de fairness** confirmou que o modelo mantém alta performance em todos os grupos demográficos, sem viés contra perfis vulneráveis.

- ❖ Jovens (18-30 anos): Recall de 94.2%
- ❖ Baixa renda (até 30k): F1-Score de 95.1%

O que isso significa na prática?

O modelo consegue prever com grande precisão quais jovens ou pessoas de baixa renda têm risco de sair da empresa, **sem cometer injustiças** ou deixar esses grupos de fora das **previsões corretas**.

Ou seja, ele não favorece ou prejudica um grupo específico.

Trata todos de forma equilibrada. Isso é importante para garantir que as decisões baseadas nesses dados sejam **justas e inclusivas**, respeitando a diversidade da equipe.



É possível prever a rotatividade antes que ela aconteça, mas qual modelo escolher?

Ambos modelos são robustos. A escolha depende da estratégia de retenção.

Cobertura Máxima? XGBoost

Eficiência de Recursos? Random Forest

- **XGBoost:** Capta mais casos de risco (maior recall)
 - Ideal para áreas e funções mais críticas como P&D e viajantes frequentes
- **Random Forest:** Mais assertivo, reduz custos com alarmes falsos (maior precisão)
 - Para os demais drivers de rotatividade, pois o volume é alto. Melhor focar apenas nos casos mais certos para otimizar ações de retenção.

Recomendação: Implementar piloto com monitoramento contínuo e revisão periódica das métricas de fairness.