

Projeto Final da Trilha de Ciência de Dados

Análise Preditiva da Rotatividade na Data Girls S.A.

Objetivo do projeto

Na Data Girls S.A., a rotatividade de funcionários estava causando perdas e desgaste. O objetivo do projeto é criar um modelo preditivo para identificar quem tem maior risco de sair, permitindo ações antecipadas.

Etapas da análise

O processo iniciou com a exploração dos dados para entender suas características. Em seguida, os dados foram limpos e preparados, incluindo tratamento de valores ausentes e codificação de variáveis categóricas. A análise exploratória revelou padrões e distribuições relevantes.

Dois modelos foram treinados:

- 1) **Regressão Logística:** Modelo linear que permite interpretar facilmente os coeficientes das variáveis e entender seu impacto positivo ou negativo na rotatividade.
- 2) **Árvore de Decisão:** Modelo baseado em regras que identifica interações entre variáveis e oferece uma visão mais flexível das relações entre os fatores que influenciam a saída dos colaboradores.

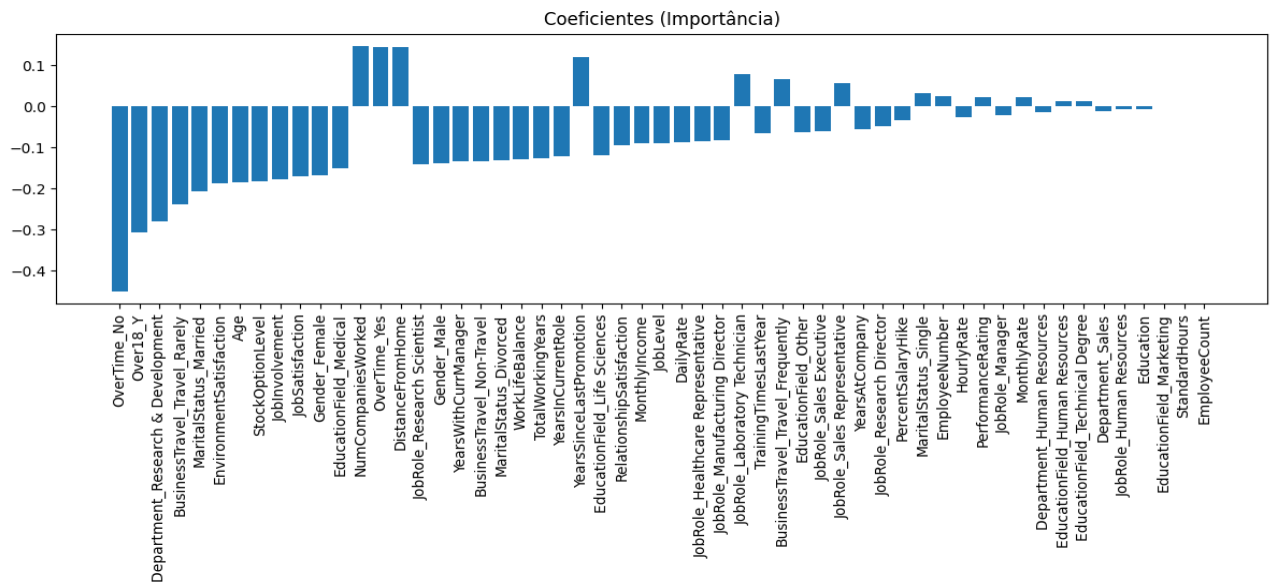
Principais achados e insights:

A análise revelou que a **Hora Extra** (OverTime) é um dos principais fatores ligados à rotatividade, mostrando que o desgaste pelo excesso de trabalho pode levar à saída dos colaboradores.

No **modelo de Regressão Logística**, percebe-se que funcionários que não fazem hora extra e que atuam nos departamentos de Pesquisa e Desenvolvimento têm menor probabilidade de sair.

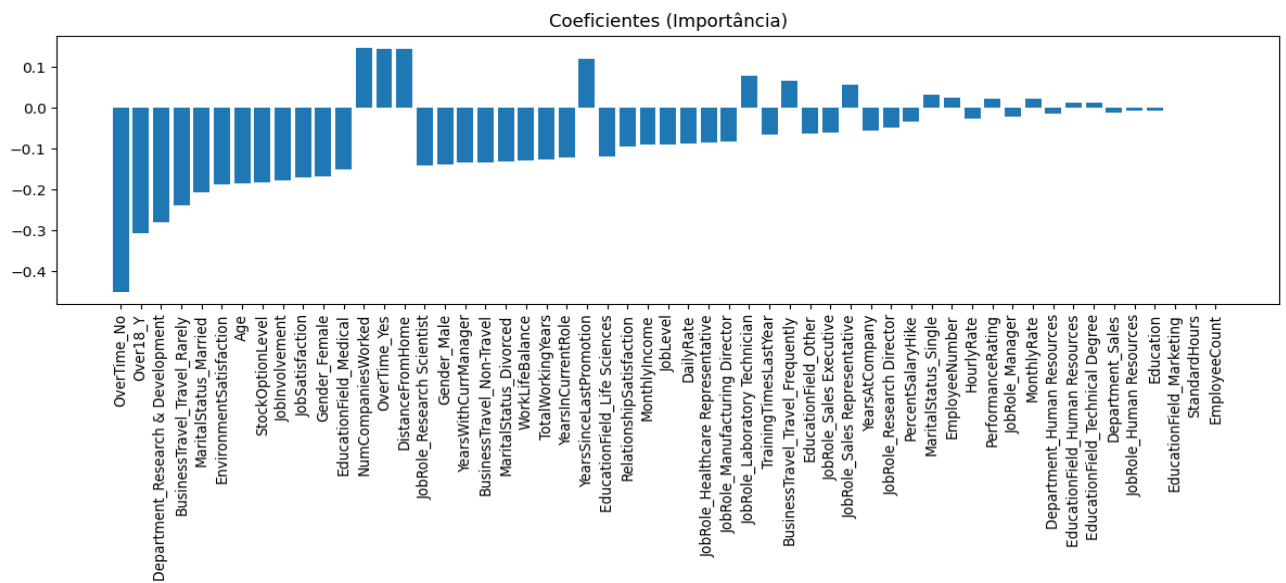
Já aqueles que trabalham em mais empresas antes, fazem hora extra, enfrentam longos deslocamentos ou não recebem promoções há bastante tempo, apresentam maior risco de deixar a empresa.

E embora a idade mostre influência, este dado é irrelevante neste caso, pois todos os colaboradores têm mais de 18 anos.



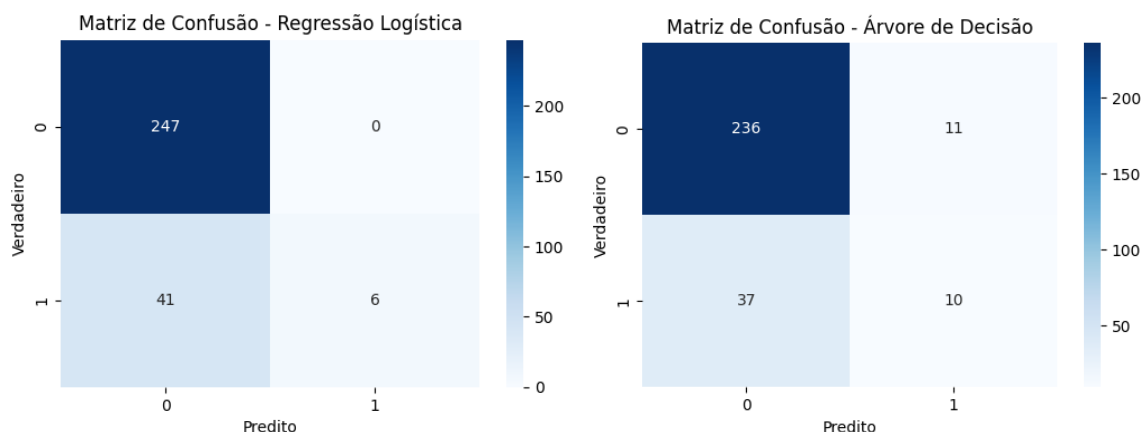
A **Árvore de Decisão** complementa essa visão, destacando que o tempo total de carreira e o salário mensal influenciam bastante a decisão do colaborador.

Funcionários com menor salário e mais anos de serviço são mais propensos a sair. Além disso, cargos específicos, como Cientista de Pesquisa, têm níveis maiores de rotatividade.



Sobre a confiabilidade dos modelos, a **Regressão Logística** evita gastos desnecessários com retenção, pois raramente classifica errado alguém como risco (falsos positivos), mas pode deixar passar pessoas que vão sair (falsos negativos).

Já a **Árvore de Decisão** identifica melhor quem realmente está em risco, mas pode gerar alertas desnecessários. Essa diferença é crucial para que a empresa escolha uma estratégia alinhada aos seus objetivos e custos.



Recomendações práticas com base nos resultados

- Reduzir a necessidade de horas extras para melhorar o equilíbrio trabalho-vida pessoal.
- Implementar políticas de reconhecimento e progressão para reter talentos.
- Ajustar remuneração e benefícios conforme o tempo de carreira e deslocamento dos funcionários.
- Escolher a estratégia de monitoramento de riscos alinhada aos custos dos diferentes tipos de erro dos modelos.

Considerações finais

Este projeto demonstrou como a análise de dados pode apoiar decisões estratégicas para reduzir a rotatividade na Data Girls S.A.

Ao identificar os fatores críticos e comparar modelos preditivos, foi possível gerar recomendações práticas que ajudam a empresa a reter talentos, otimizar recursos e melhorar o ambiente de trabalho.

A adoção dessas medidas contribuirá para um futuro mais estável e produtivo.

Embora nenhum dos modelos seja perfeito para identificar todos os colaboradores com risco de saída, a combinação dos resultados oferece uma visão abrangente e valiosa sobre o problema da rotatividade.

A escolha do modelo ideal deve considerar o custo associado aos diferentes tipos de erro:

- Se a maior preocupação for **evitar a perda inesperada de talentos** (falsos negativos), o modelo de Árvore de Decisão apresenta melhor desempenho.
- Se o foco for **minimizar gastos com retenção desnecessária** (falsos positivos), a Regressão Logística é a opção mais segura.

Além disso, fatores como o desgaste causado por horas extras, o tempo de carreira e o salário mostram-se pilares essenciais para o desenvolvimento de estratégias eficazes de retenção na Data Girls S.A.