

Previsão de Rotatividade de Funcionários com Machine Learning

Projeto de Análise Preditiva de RH – Jornada de Dados / Laboratória

O Desafio da Rotatividade de Funcionários

Por que prever desligamentos é essencial para o RH estratégico?



A rotatividade gera altos custos com contratações e treinamentos.



Impacta produtividade, clima e continuidade dos projetos.



Antecipar a saída é a chave para reter talentos.



Segundo o levantamento interno, a rotatividade média foi de **16%**, afetando principalmente funcionários jovens.

Base de Dados e Processamento

Os dados foram preparados através de um processo estruturado de limpeza e transformação.

Conexão com a base de dados (rh_data.csv).

Importação

Verificação de linhas duplicadas.

Duplicados

Análise de valores discrepantes em *strings*.

Categorias

Conversão de tipos de dados

Ajuste de Tipos

Nulos

Preenchimento de nulos com a mediana.

Remoção de Ruído

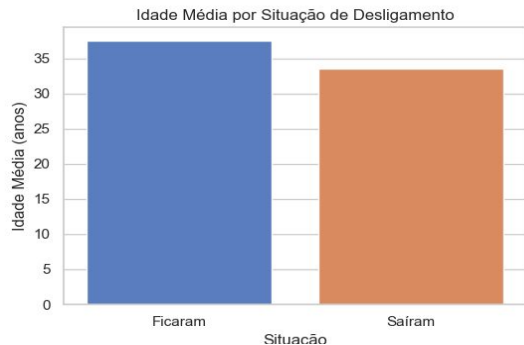
Exclusão de colunas irrelevantes

Outliers Numéricos

Deteção de valores discrepantes

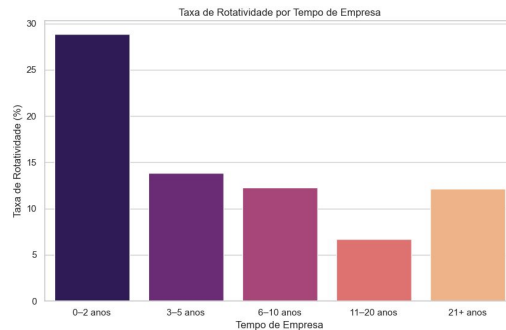
O que mais influencia a saída dos funcionários?

O que os dados revelam sobre quem sai e quem fica



Funcionários que **saíram** têm em média **33 anos**, enquanto os que **ficaram** têm **37 anos**.

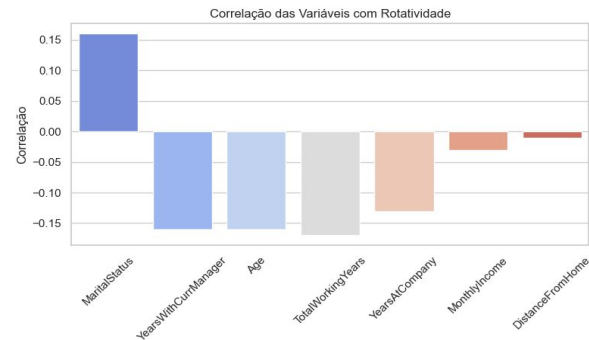
Jovens apresentam maior propensão à rotatividade.



Maior taxa de saída ocorre **nos primeiros 2 anos (28,9%)**.

Após **5 anos**, a rotatividade **cai para menos de 10%**.

Estratégias de retenção devem focar no início da jornada do colaborador.




Fatores com **maior correlação negativa**: **TotalWorkingYears**, **Age**, **YearsAtCompany**, **YearsWithCurrManager**.

Quanto maior o tempo de casa e experiência, menor o risco de saída.

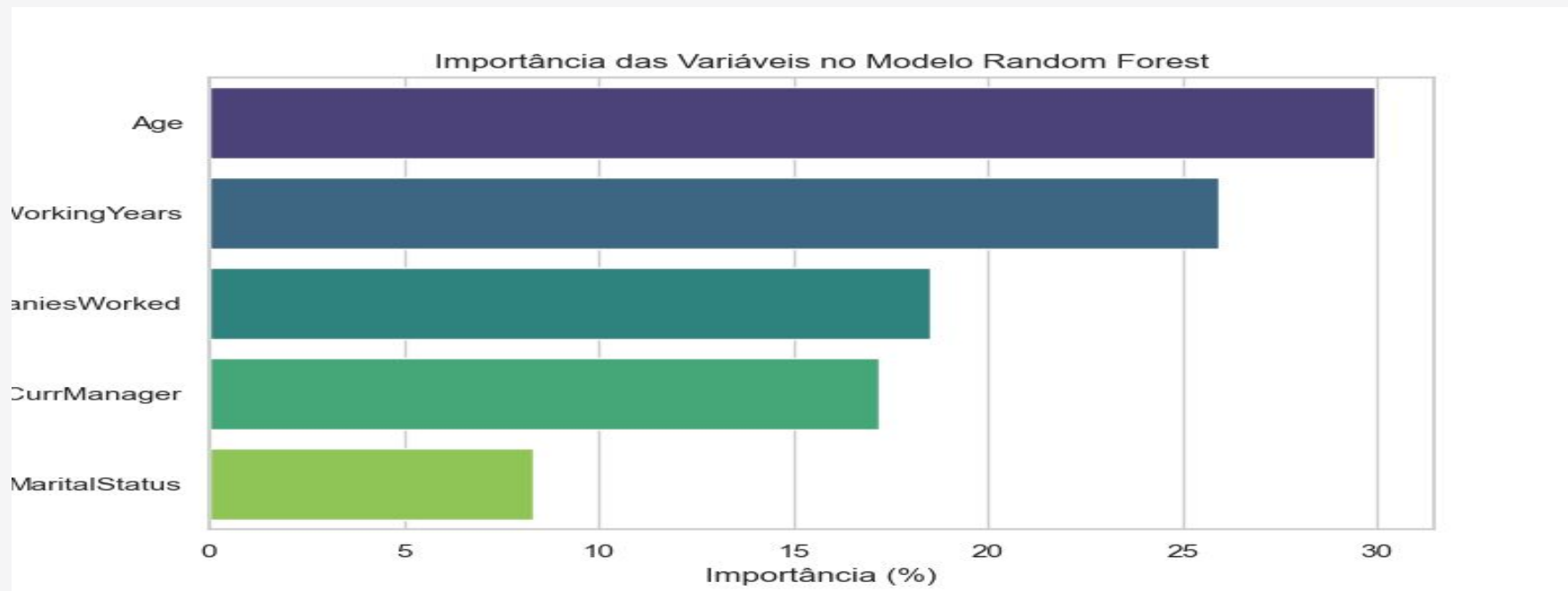
Modelos testados e desempenho

Testamos três modelos — Regressão Logística, XGBoost e Random Forest — para prever a probabilidade de desligamento.

Modelo	Acurácia	Recall	F1	ROC-AUC
Regressão Logística	0.63	0.64	0.63	0.69
XGBoost	0.96	0.97	0.96	0.99
 Random Forest	0.97	0.99	0.97	0.995

O modelo **Random Forest** apresentou melhor equilíbrio entre precisão e recall, sendo o mais indicado para uso prático no RH.

Quais fatores o modelo considera mais importantes



As variáveis mais relevantes para prever a saída são **anos de experiência total, idade, número de empresas anteriores, tempo com o gestor e estado civil**. Esses fatores mostram que tanto o histórico de carreira quanto o relacionamento interpessoal influenciam na permanência.



Principais Insights e Ações para o RH

Insight	Evidência
Jovens e novos na empresa têm maior risco de saída	Taxa de 28,9% nos primeiros 2 anos
O tempo com o gestor é fator protetor	Rotatividade de 14% ou menos após 3 anos
Funcionários com mais tempo e experiência permanecem mais	Correlação negativa (-0.16) entre idade e rotatividade
Random Forest é o modelo ideal para aplicação prática	97% de acurácia, recall 99%



A análise preditiva aplicada ao RH mostrou que é possível **prever desligamentos com alta precisão e direcionar estratégias de retenção mais inteligentes.**

O uso do modelo Random Forest pode apoiar decisões baseadas em dados, reduzindo custos e fortalecendo a cultura organizacional.