

Projeto Final, Bootcamp Data Girls - Relatório de Análise

Este projeto tem como objetivo identificar e compreender os principais fatores que influenciam a rotatividade de funcionários (attrition) em uma empresa. Utilizando técnicas de Machine Learning, o objetivo é construir um modelo preditivo capaz de sinalizar funcionários com alto risco de saída, fornecendo à equipe de Recursos Humanos insights valiosos para a criação de estratégias de retenção mais eficazes e proativas.

Etapas da Análise:

Leitura e Limpeza dos Dados: O conjunto de dados original foi carregado e as colunas foram traduzidas para o português para maior clareza. Algumas colunas irrelevantes para o projeto foram removidas.

Análise Exploratória de Dados: A análise exploratória de dados revelou que a rotatividade de funcionários é um evento raro, criando um desequilíbrio de classes que exigiu atenção na modelagem. Os gráficos mostraram que fazer horas extras e viajar a negócios com frequência estão fortemente associados a um maior risco de saída. Embora o departamento de Pesquisa e Desenvolvimento tenha o maior número de funcionários e, consequentemente, mais casos de rotatividade, o gênero não se destacou como um fator relevante. Esses achados iniciais foram essenciais para guiar a estratégia de modelagem e a interpretação final dos resultados.

Preparação dos Dados para Modelagem: Para que o Machine Learning pudesse ser aplicado, o DataFrame (`df_pt`) foi preparado. Variáveis categóricas (Departamento, Cargo, EstadoCivil, etc.) foram transformadas em variáveis numéricas usando a técnica de One-Hot Encoding, resultando em um novo DataFrame (`df_modelo`) pronto para o treinamento. O `df_modelo` foi então dividido em conjuntos de treino e teste.

Modelagem Preditiva: Dois modelos de classificação foram treinados: **Random Forest** e **Gradient Boosting**. A escolha desses modelos se deu por sua robustez e capacidade de lidar com dados complexos. A avaliação inicial mostrou que, embora a acurácia geral fosse alta, o desempenho na classe de interesse (Rotatividade = 1) era baixo devido ao desequilíbrio de classes.

Otimização do Modelo: Para resolver o problema do desequilíbrio, a técnica SMOTE (Synthetic Minority Over-sampling Technique) foi aplicada. O SMOTE criou dados sintéticos para a classe minoritária, balanceando o conjunto de dados de treino. Após o re-treinamento, o modelo de Gradient Boosting apresentou uma melhoria significativa no recall e f1-score para a classe de rotatividade, tornando-o muito mais eficaz.

Interpretação do Modelo: O modelo otimizado foi analisado para extrair as variáveis mais importantes na sua tomada de decisão. Esta etapa final nos permitiu entender quais fatores são os mais determinantes para a rotatividade.

Principais Achados e Insights

A análise de importância das variáveis revelou que os 5 principais fatores que influenciam a rotatividade, em ordem decrescente de importância, são:

1. **NivelOpcaoAcoes:** A variável mais importante. Sugere que a participação acionária ou incentivos financeiros de longo prazo são cruciais para a retenção.
2. **EstadoCivil_Married:** O estado civil do funcionário é um fator de peso na decisão do modelo.
3. **SatisfacaoProfissional:** A satisfação com o trabalho é um dos principais preditores, validando a importância de programas de engajamento.
4. **Departamento_Research & Development:** O modelo identificou um padrão de rotatividade único e significativo neste departamento.
5. **NivelProfissional:** A posição hierárquica na empresa também é um fator importante.

Recomendações Práticas

Com base nos insights obtidos, as seguintes ações práticas são recomendadas para a equipe de Recursos Humanos:

Revisão do Plano de Opções de Ações: Analisar o programa de NivelOpcaoAcoes para entender como ele pode ser usado de forma mais eficaz para motivar a permanência de funcionários-chave.

Pesquisa de Satisfação Aprofundada: Realizar pesquisas de SatisfacaoProfissional mais frequentes e detalhadas. O modelo confirma que esta é uma área de alto impacto, e a empresa deve atuar para resolver os pontos de insatisfação identificados.

Gestão de Carga de Trabalho: Investigar a causa da alta rotatividade em funcionários que fazem HorasExtras. Uma carga de trabalho excessiva pode estar afetando o EquilibrioTrabalhoVida, um fator de risco para a retenção.

Análise Específica de Setores: Focar a atenção no Departamento_Research & Development para entender as particularidades da rotatividade. Pode haver questões específicas de carreira, gestão ou cultura que precisam ser abordadas.

Uso Preditivo do Modelo: Utilizar o modelo de Gradient Boosting treinado como uma ferramenta proativa. A cada novo funcionário, o modelo pode ser alimentado com seus dados para gerar uma pontuação de risco de rotatividade, permitindo que a equipe de RH atue de forma preventiva.

Luana dos Anjos Abreu
Estudante de Ciência de Dados