

Lucas Matheus da Silva - RM:550466

Análise de Attrition de Funcionários

Comentários iniciais

Este estudo utiliza uma base de dados do Kaggle sobre a demissão de funcionários (attrition) e suas diferentes características. O objetivo é prever a probabilidade de um funcionário se desligar da empresa.

Link da base de dados:

<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset?resource=download>

A variável alvo (label) do dataset é o campo **Attrition**.

1. Análise Exploratória

Distribuição de Attrition

- A grande maioria dos funcionários (83.88%) não deixou a empresa.
- Apenas 16.12% dos funcionários saíram, indicando um dataset desbalanceado.

Características Demográficas

- Idade média dos funcionários: 36.92 anos (mín: 18, máx: 60)
- Gênero: 60% masculino, 40% feminino
- Estado civil: 45.78% casados, 31.97% solteiros, 22.24% divorciados

Características Profissionais

- Departamentos:
 - Pesquisa & Desenvolvimento: 65.37%
 - Vendas: 30.34%
 - Recursos Humanos: 4.29%

- Viagens a trabalho:
 - Raramente: 70.95%
 - Frequentemente: 18.84%
 - Não viaja: 10.20%
- Horas extras:
 - Não: 71.70%
 - Sim: 28.30%

Observações Importantes

- Funcionários mais jovens tendem a ter uma taxa de attrition mais alta.
- A maioria dos funcionários possui formação em Life Sciences, seguido por Medical e Marketing.
- Funcionários que saem da empresa tendem a ter salários ligeiramente mais baixos.
- A maioria dos funcionários que fazem horas extras tendem a sair da empresa.

2. Matriz de Correlação

Relações positivas significativas

- JobLevel e MonthlyIncome: Cargos mais altos têm salários significativamente mais altos.
- TotalWorkingYears e YearsAtCompany: Funcionários com mais anos de trabalho tendem a ter passado mais tempo na empresa atual.
- PerformanceRating e PercentSalaryHike: Melhores avaliações resultam em maiores aumentos salariais.
- YearsWithCurrManager e YearsAtCompany: Funcionários que ficam mais tempo na empresa tendem a trabalhar com o mesmo gerente por mais tempo.

Relações negativas ou não significativas

- DistanceFromHome e JobSatisfaction: Pouca relação entre distância do trabalho e satisfação.
- DailyRate e YearsAtCompany: Salário diário não está relacionado ao tempo na empresa.
- PerformanceRating e Age: Avaliação de desempenho tem pouca relação com a idade.

3. Modelagem de Machine Learning

Random Forest

- Acurácia inicial: 84%
- Precisão para attrition: 50%

- Recall para attrition: 10%
- F1-score para attrition: 16%
- ROC AUC inicial: 0.76
- ROC AUC após otimização: 0.77

Regressão Logística

- Acurácia: 88%
- Precisão para attrition: 72%
- Recall para attrition: 39%
- F1-score para attrition: 51%
- ROC AUC inicial: 0.82
- ROC AUC após otimização: 0.82 (estável)

4. Análise dos Resultados dos Modelos

1. Random Forest

- Mesmo com ajuste de hiperparâmetros, não superou a Regressão Logística.
- Teve dificuldades em identificar casos de attrition (classe minoritária).

2. Regressão Logística

- Desempenho superior desde o início.
- Melhor capacidade de generalização e tratamento do desbalanceamento de classes.
- O ajuste de hiperparâmetros não trouxe melhorias significativas, indicando que o modelo já estava bem ajustado.

5. Conclusões e Recomendações

1. A Regressão Logística é o modelo recomendado para prever attrition neste conjunto de dados.
2. Fatores importantes para attrition:
 - Horas extras
 - Idade do funcionário
 - Nível salarial
 - Satisfação no trabalho
3. Recomendações para reduzir attrition:
 - Revisar políticas de horas extras
 - Implementar programas de retenção focados em funcionários mais jovens
 - Assegurar competitividade salarial
 - Melhorar a satisfação no trabalho e o ambiente de trabalho
4. Próximos passos:
 - Investigar técnicas de balanceamento de dados para melhorar a detecção de casos de attrition

- Explorar modelos mais complexos ou técnicas de ensemble
- Realizar uma análise mais profunda das features mais importantes identificadas pelos modelos
- Coletar mais dados, especialmente relacionados a casos de attrition

Esta análise fornece insights valiosos sobre os fatores que influenciam a rotatividade de funcionários e oferece um modelo preditivo para ajudar na identificação precoce de potenciais casos de attrition.