

Projeto de RH Analytics

Criação de modelo para predição de desligamento de funcionário

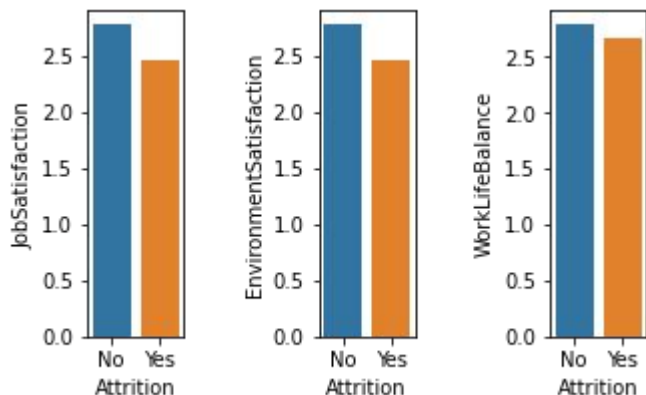
O problema

Que funcionários deixam empresas todos os dias não é segredo pra ninguém, entretanto com o avanço do aprendizado de máquina e da ciência de dados como um todo podemos tentar uma forma de tentar compreender melhor esse fenômeno

- Quais fatores influenciam os funcionários na escolha de deixar uma empresa?
- É possível construir um modelo afim de ajudar o setor de recurso humanos a prever quais funcionários estão mais propensos a deixar a empresa?
- Uma vez construído tal modelo o departamento de RH pode:
 - Tomar ações preventivas, a fim de evitar o desligamento do funcionário
 - Começar a trabalhar alguém para substituí-lo e diminuir o risco para a empresa

Iniciando a análise exploratória dos dados

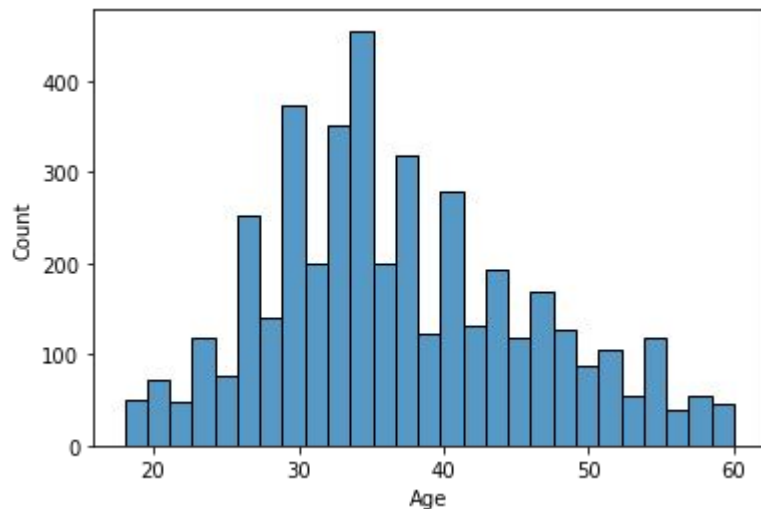
O primeiro passo consiste justamente em analisar quais variáveis são mais relevantes para determinar a saída dos funcionários.



Esse primeiro gráfico nos mostra, sem muitas surpresas que a satisfação no trabalho, bem como com o ambiente e o bom equilíbrio entre trabalho e vida pessoal reduzem a probabilidade da pessoa ser demitida.

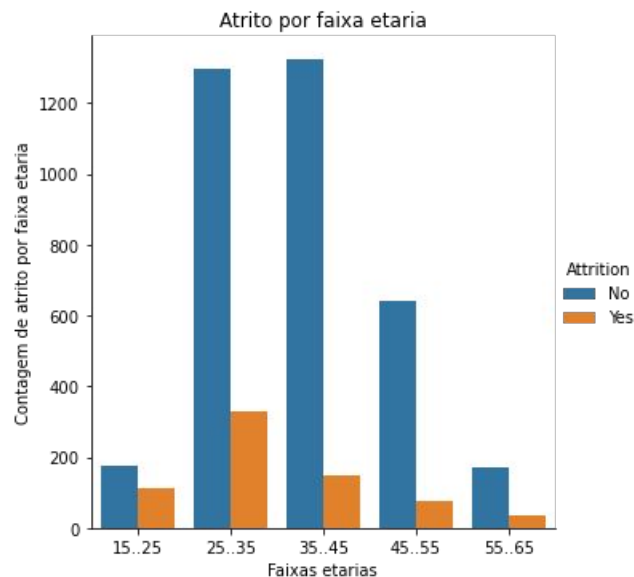
Attrition = Funcionário se desligou da empresa

Análises com a faixa etária

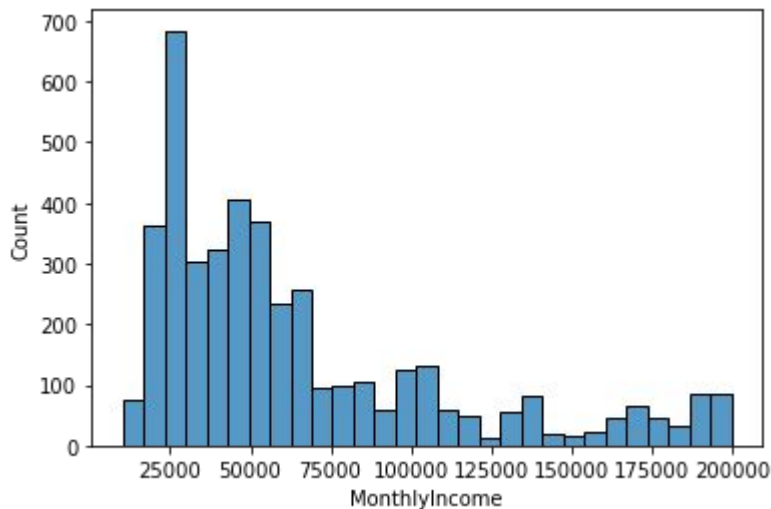


Também é importante termos um perfil geral dos funcionários, começando pela idade. Aqui vemos que a maior parte dos funcionários está entre 30 e 40 anos

Percebermos que a faixa etária entre 15 e 25 anos é a que tem a maior proporção de funcionários se desligando da empresa

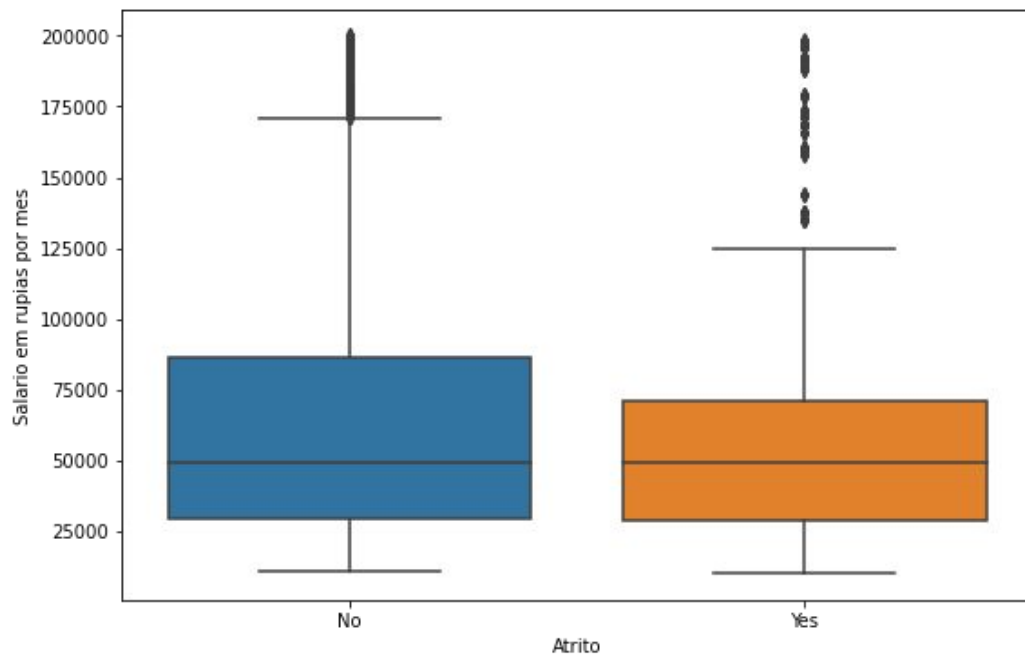


Análise envolvendo o salário dos funcionários

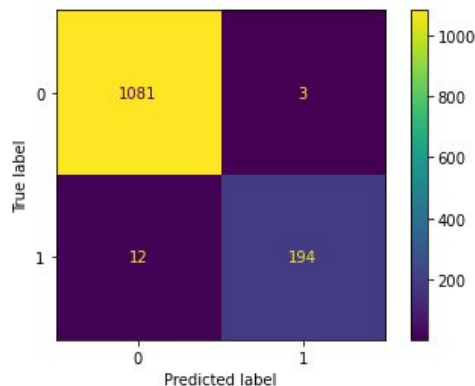


Também vemos, sem surpresas, que os funcionários com um salário um pouco maior estão menos dispostos a deixar a empresa

A distribuição dos salários (MonthlyIncome)



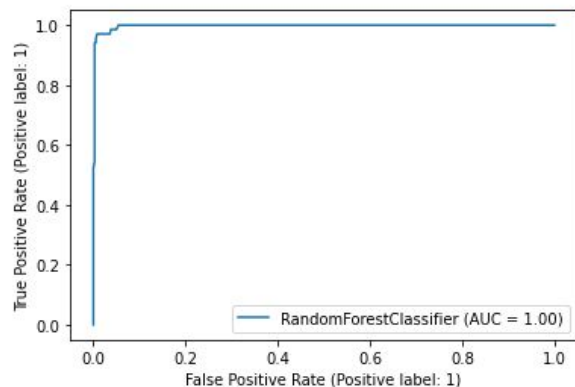
Criação do modelo e análise dos resultados



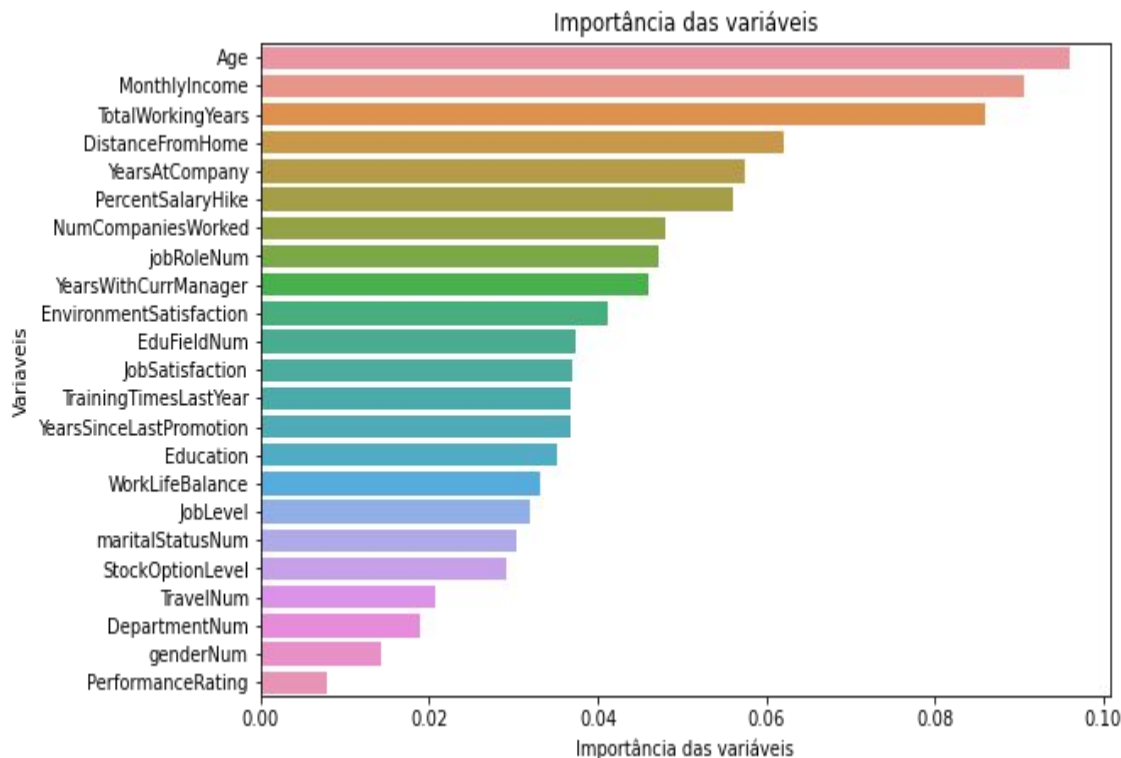
Estamos tentando prever se um funcionário deixará ou não uma empresa, ou seja, estamos tentando prever uma classificação, classificar se o funcionário deixará ou não a empresa. Para avaliar se o nosso modelo preditor realizou a sua tarefa com boa performance temos algumas métricas, entre elas:

- Matriz de confusão e
- Curva ROC

Ambas estão mostradas ao lado, nas quais foi possível ver que a performance ficou boa, estamos utilizando para isso o algoritmo Random Forest.



Analizando o resultado



Uma característica do algoritmo que foi utilizado é que o mesmo nos fornece, após a criação do modelo, quais das variáveis foram mais importantes no processo de predição.

Podemos ver claramente que os fatores que mais influenciam são:

- Idade
- Salário
- Tempo total de trabalho (também ligado à idade)
- Distância de casa

É importante lembrar que esses dados foram fornecidos a partir do modelo de machine learning no intuito de prever se o funcionário deixará a empresa ou não

Tecnologias e bibliotecas utilizadas

- Python
 - Pandas
 - Seaborn
 - Matplotlib
 - Scikit-Learn
-
- Machine learning
 - Análise exploratória de dados

Conclusão

Dados os dados fornecidos conseguimos prever com nível satisfatório de precisão se um funcionário deixará ou não a empresa.

Claro que esse tipo de informação será diferente para outro tipo de empresa que esteja, eventualmente, inserida em um contexto diferente, o que exigirá a criação de outro modelo e conseqüentemente a coleta de outro tipo de dados também.

A criação do modelo também nos permite chegar às variáveis mais importantes informação também valiosa para o pessoal de RH. Alguns podem argumentar que esse tipo de variável já é conhecido de profissionais experientes no mercado, o que não é mentira, entretanto a criação do modelo nos permite encontrar padrões que muitas vezes escapam aos olhos humanos.

Por fim conclui-se que o modelo pode ser bem útil para o setor de RH de uma empresa permitindo-o, de maneira automatizada, encontrar os profissionais de maneira mais propensos a trocar de empresa e, dessa forma, traçar um plano de ação, economizando assim custos com novas contratações ou tempo no qual os novos funcionários ainda não alcançaram produtividade total.