

Business Analytics e Machine Learning Para Projetos de Data Science

Identificando e Prevendo os Fatores Que Contribuem Para a Satisfação dos Funcionários

Roberto SSoares - LfLNgLrning

[in/roberto-dos-santos-soares](#)

[Portfólio: roberto-sssoares](#)

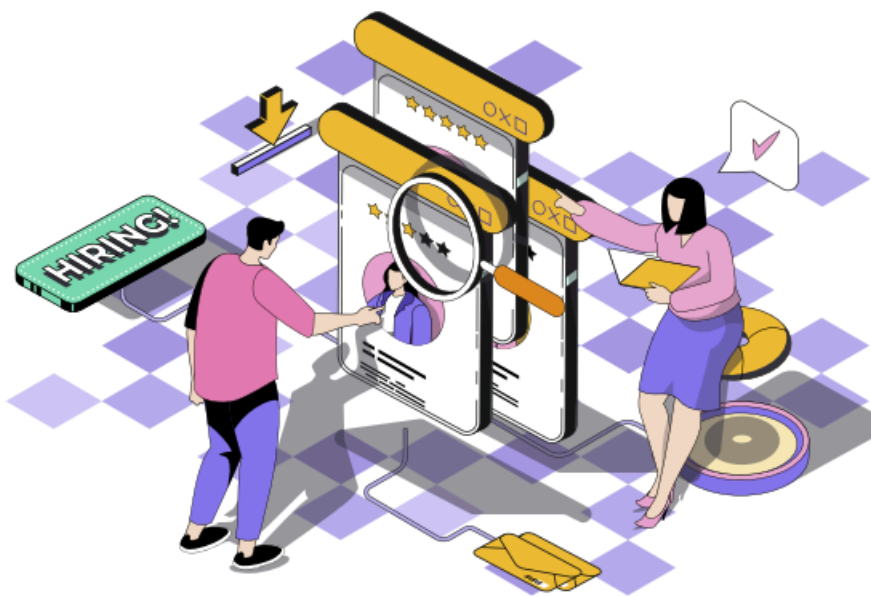
" [+] Faturamento,
[-] Custo,
[+] Qualidade de vida "
"Bruno Jardim"

Agradeço a "PoD Academy" e "DSA Academy" pelo conhecimento transmitido.
e a minha linda esposa "Elizabete" por todo apoio.

Visão Geral

Analytics em Recrutamento e Seleção

- A análise em recrutamento e seleção de pessoas envolve várias etapas fundamentais, começando pela coleta de dados relevantes.
 - Esses dados incluem informações dos candidatos, como currículos, perfis de redes sociais e resultados de avaliações, além de dados do processo de recrutamento, como tempo para preencher vagas e taxas de aceitação de ofertas.
- Dados de desempenho pós-contratação, como avaliações e feedback, também são muito importantes e devem ser considerados.
 - A coleta abrangente e sistemática desses dados estabelece a base para análises detalhadas que podem revelar insights valiosos sobre o processo de contratação.
- A análise preditiva utiliza técnicas de Machine Learning e modelos estatísticos para prever resultados futuros, como o tempo necessário para preencher uma vaga ou a probabilidade de um candidato ser bem-sucedido.
- Modelos de regressão e algoritmos de classificação são aplicados para analisar diversos fatores que influenciam o processo de recrutamento.
 - Isso ajuda as empresas a antecipar desafios e ajustar suas estratégias de recrutamento de maneira proativa.
- Ferramentas de Processamento de Linguagem Natural (PLN) são usadas para realizar análises de texto em currículos e cartas de apresentação, permitindo a identificação de habilidades e experiências dos candidatos.
- Análises de sentimentos podem ser aplicadas a feedbacks de entrevistas para extrair percepções qualitativas sobre os candidatos.
 - Essas técnicas avançadas de análise de texto complementam as análises quantitativas, proporcionando uma compreensão mais completa dos candidatos.



- Após a coleta de dados, a análise descritiva entra em ação para fornecer uma visão geral das métricas básicas, como tempo médio para preencher uma vaga e custo por contratação.
 - Visualizações, como gráficos de barras e linhas, ajudam a monitorar tendências e identificar padrões ao longo do tempo.
 - Esta etapa permite às empresas avaliar o estado atual de seus processos de recrutamento e identificar áreas que necessitam de melhorias imediatas.
- A otimização do processo de recrutamento é alcançada identificando gargalos e implementando melhorias contínuas com base em dados.
 - Ferramentas e tecnologias como sistemas de rastreamento de candidatos e plataformas de análise de dados são essenciais para gerenciar e analisar dados de forma eficiente.
- A integração de Inteligência Artificial pode automatizar tarefas e melhorar a precisão da triagem de candidatos.
 - Com essas abordagens, as empresas podem aprimorar significativamente a eficácia e a eficiência de seus processos de recrutamento e seleção, resultando em contratações de maior qualidade e maior retenção de talentos.

Modelagem Preditiva Para Turnover e Retenção de Talentos

- A modelagem preditiva para turnover e retenção de talentos é uma ferramenta poderosa que permite às empresas prever quais funcionários têm maior probabilidade de deixar a organização.
- A primeira etapa envolve a coleta e preparação de dados históricos relevantes, como registros de funcionários, avaliações de desempenho, dados demográficos, feedbacks de satisfação e histórico de promoções e aumentos salariais.
- Esses dados formam a base para a construção de modelos preditivos que identificam padrões e fatores associados ao turnover.



- A análise descritiva desses dados é essencial para entender as características dos funcionários que permanecem na empresa em comparação com aqueles que saem.
 - Essa etapa envolve o cálculo de métricas como taxas de turnover por departamento, tempo médio de permanência e a identificação de picos sazonais de saídas.
 - Visualizações de dados, como gráficos de dispersão e histogramas, ajudam a ilustrar esses padrões e fornecem insights iniciais sobre os fatores que podem estar influenciando a retenção e o turnover.
 - Com os dados preparados e as análises descritivas concluídas, a próxima etapa é a construção de modelos preditivos.
 - Técnicas de Machine Learning, como regressão logística, árvores de decisão e redes neurais, são aplicadas para criar modelos que possam prever a probabilidade de um funcionário deixar a empresa.
 - Esses modelos levam em consideração diversas variáveis, como satisfação no trabalho, engajamento, oportunidades de crescimento e equilíbrio entre vida pessoal e profissional.
 - A validação cruzada é usada para garantir que os modelos sejam precisos e generalizáveis.
 - A implementação de modelos preditivos permite que as empresas identifiquem proativamente os funcionários em risco de turnover e desenvolvam estratégias de retenção personalizadas.
 - Por exemplo, se o modelo identificar que a falta de oportunidades de desenvolvimento profissional é um fator significativo para o turnover, a empresa pode investir em programas de treinamento e desenvolvimento de carreira.
 - Da mesma forma, se o equilíbrio entre vida pessoal e profissional for um fator crítico, políticas de trabalho exigível e bem-estar podem ser implementadas para melhorar a satisfação dos funcionários.
-
- Em última análise, a modelagem preditiva para turnover e retenção de talentos não apenas ajuda a reduzir os custos associados ao turnover, mas também melhora a produtividade dos funcionários.
 - Ao entender melhor as necessidades e preocupações dos funcionários, as empresas podem criar um ambiente de trabalho mais satisfatório e engajador, aumentando a lealdade e a retenção.
 - A utilização contínua de análises preditivas e a adaptação das estratégias de retenção com base em dados garantem que as empresas mantenham uma força de trabalho estável e motivada.

Análise de Desempenho e Gestão de Talentos

- A análise de desempenho e gestão de talentos é uma prática fundamental para maximizar o potencial dos funcionários e alinhar suas habilidades e aspirações com os objetivos organizacionais.
- Tudo começa com a coleta de dados abrangentes sobre o desempenho dos funcionários, incluindo avaliações regulares, feedbacks de 360 graus, resultados de projetos e metas atingidas.
 - Esses dados fornecem uma visão holística do desempenho individual e permitem a identificação de pontos fortes e áreas de melhoria.



- A análise desses dados envolve a aplicação de métricas quantitativas e qualitativas para avaliar o desempenho dos funcionários de forma objetiva e justa.
 - Métricas comuns incluem a produtividade, a qualidade do trabalho, o cumprimento de prazos e a contribuição para a equipe.
- Ferramentas de visualização de dados, como dashboards e gráficos de tendência, ajudam a monitorar o desempenho ao longo do tempo e a identificar padrões ou desvios significativos.
 - Essa análise detalhada fornece insights valiosos para a tomada de decisões estratégicas na gestão de talentos.
- Com base na análise de desempenho, a gestão de talentos pode desenvolver planos de desenvolvimento personalizados que alinham as aspirações dos funcionários com as necessidades da organização.
 - Isso pode incluir programas de treinamento específicos, oportunidades de mentoria e projetos desafiadores que permitem aos funcionários expandir suas habilidades.
- Ao investir no desenvolvimento contínuo dos funcionários, as empresas não só aumentam a satisfação e o engajamento, mas também constroem uma força de trabalho mais capacitada e adaptável.
- A gestão de talentos também envolve a identificação de funcionários de alto potencial que podem assumir posições de liderança no futuro.
- Utilizando dados de desempenho e outros indicadores, como habilidades de liderança e capacidade de tomada de decisão, as empresas podem criar programas de sucessão para garantir que tenham um pipeline de líderes preparados.
 - Isso é crucial para a sustentabilidade a longo prazo e para manter a continuidade organizacional em tempos de mudança.
- A análise de desempenho e a gestão de talentos devem ser processos contínuos e integrados à cultura organizacional.
 - Feedbacks regulares e conversas de desenvolvimento devem ser incentivados para criar um ambiente de melhoria contínua.
- A utilização de tecnologias avançadas, como sistemas de gestão de desempenho e análise preditiva, pode melhorar significativamente a eficácia desses processos, permitindo uma abordagem mais dinâmica e adaptável à gestão de talentos.

- Assim, as empresas podem garantir que estão maximizando o potencial de seus funcionários e alcançando um desempenho organizacional superior.

Previsão e Análise de Impacto de Políticas de RH

- A previsão e análise de impacto de políticas de RH permitem auxiliar as empresas para que elas possam tomar decisões informadas e estratégicas que afetam diretamente a satisfação, retenção e produtividade dos funcionários.
- A primeira etapa nesse processo é a coleta de dados relevantes, que incluem históricos de desempenho, dados demográficos, resultados de pesquisas de satisfação, taxas de rotatividade e outras métricas de RH.
- Esses dados são fundamentais para construir modelos preditivos que ajudam a antecipar os efeitos das políticas de RH antes de sua implementação.



- A previsão utiliza técnicas de análise estatística e Machine Learning para identificar padrões e tendências nos dados históricos.
 - Modelos preditivos, como regressão linear, árvores de decisão e redes neurais, são desenvolvidos para prever resultados futuros com base em diferentes cenários de políticas.
 - Por exemplo, se uma empresa deseja implementar uma nova política de trabalho remoto, o modelo preditivo pode estimar o impacto dessa política na produtividade dos funcionários, na taxa de retenção e na satisfação geral.
- Essas previsões ajudam a empresa a entender as possíveis consequências e ajustar a política conforme necessário antes de sua implementação.
- A análise de impacto envolve a avaliação das mudanças reais após a implementação de uma política de RH.
 - Isso é feito comparando os dados reais com as previsões feitas anteriormente.
- Ferramentas de visualização de dados, como gráficos de antes e depois, ajudam a ilustrar o impacto das políticas de maneira clara e compreensível.
 - A análise de impacto também pode envolver o uso de experimentos controlados, como testes A/B, onde um grupo de funcionários é exposto à nova política enquanto outro grupo continua com a política atual.
 - Comparar os resultados desses dois grupos fornece insights adicionais sobre a eficácia da política.
- A integração contínua de feedback dos funcionários é primordial para a análise de impacto.
 - Pesquisas regulares de satisfação e feedback qualitativo ajudam a entender as percepções e experiências dos funcionários em relação às novas políticas.
- Essa abordagem permite ajustes dinâmicos e imediatos, garantindo que as políticas de RH estejam alinhadas com as necessidades e expectativas dos funcionários.
- Além disso, a utilização de ferramentas de análise de sentimento pode ajudar a captar nuances no feedback dos funcionários, proporcionando uma visão mais aprofundada dos impactos emocionais e motivacionais das políticas implementadas.

- A previsão e análise de impacto de políticas de RH devem ser vistas como um ciclo contínuo de melhoria.
 - As empresas devem constantemente revisar e atualizar seus modelos preditivos com novos dados para refinar suas previsões e análises de impacto.
- A adoção de uma abordagem iterativa permite que as empresas se adaptem rapidamente às mudanças no ambiente de trabalho e mantenham suas políticas de RH relevantes e eficazes.
- Ao utilizar análises preditivas e de impacto de forma proativa, as empresas podem criar um ambiente de trabalho mais resiliente e responsivo, promovendo a satisfação e a produtividade dos funcionários em longo prazo.

Instalando e Carregando os Pacotes

- Para atualizar um pacote, execute o comando abaixo no terminal ou prompt de comando:
 - `pip install -U nome_pacote`
- Para instalar a versão exata de um pacote, execute o comando abaixo no terminal ou prompt de comando:
 - `!pip install nome_pacote==versão_desejada`
- Depois de instalar ou atualizar o pacote, reinicie o jupyter notebook.
 - Instala o pacote watermark.
 - Esse pacote é usado para gravar as versões de outros pacotes usados neste jupyter notebook.

```
#!pip install -q -U watermark
```

```
# Imports
import sklearn
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

```
%reload_ext watermark
%watermark -a "RobertoSSoares-LfLngLrnnng"
```

Author: RobertoSSoares-LfLngLrnnng

```
pd.set_option('display.expand_frame_repr', False)
pd.set_option('display.max_columns', 500)
pd.set_option('display.max_rows', 500)
```

CRISP-DM [Business Understanding]

I. Definição e Objetivo - (Compreendendo o Problema a Ser Resolvido)

Formulado o problema junto a Área de Negócio, temos:

- Neste projeto, utilizaremos técnicas de ciência de dados e aprendizado de máquina para analisar e prever os fatores que influenciam a satisfação dos funcionários em uma empresa.
 - O objetivo é identificar as variáveis mais significativas que afetam o bem-estar e a motivação dos colaboradores, como:

- ambiente de trabalho,
 - remuneração,
 - oportunidades de crescimento,
 - e equilíbrio entre vida pessoal e profissional.
- Utilizaremos um conjunto de dados que inclui diversas características dos funcionários, como idade, tempo de empresa, avaliações de desempenho, feedbacks e índices de satisfação.
 - Através de modelos preditivos, buscaremos entender padrões e tendências que podem orientar políticas internas para melhorar a satisfação e retenção dos funcionários.
 - Usaremos um conjunto de dados que foi adaptado a partir do dataset disponível no link abaixo:
 - <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>
 - Descubra os fatores que levam à rotatividade de funcionários e explore questões importantes como
 - "mostre-me uma análise da distância de casa por função e rotatividade" ou
 - "compare a renda média mensal por educação e rotatividade".

CRISP-DM [Data Understanding]

2. Compreensão dos Dados - (Conjunto de Dados)

- Este é um conjunto de dados fictício criado por cientistas de dados da IBM.

3. Carregando os Dados

```
# Carrega o dataset
df00 = pd.read_csv('dataset.csv')
```

```
# Shape
df00.shape
```

```
(23058, 30)
```

```
# Amostra
df00.sample(5).transpose()
```

		9495	9525	22816
15131	16946			
Age		30	33	40
34	39			
Attrition	Current employee	Current employee	Current employee	
Current employee	Current employee			
BusinessTravel	Travel_Rarely	Travel_Frequently	Travel_Rarely	
Travel_Rarely	Travel_Rarely			
Department	Research & Development	Research & Development	Research & Development	
Research & Development	Sales			
DistanceFromHome	2	7		2
9	6			
Education	3	2		4
3	3			
EducationField	Life Sciences	Life Sciences	Life Sciences	
Medical	Marketing			
EnvironmentSatisfaction	3	3		3
3	2			
Gender	Female	Male		Male
Female	Female			

JobInvolvement		2	3	2
3	3			
JobLevel		2	2	1
2	3			
JobRole	Manufacturing Director	Manufacturing Director	Research Scientist	
Sales Executive	Manufacturing Director			
JobSatisfaction		4	2	3
1	2			
MaritalStatus		Single	Married	Single
Divorced	Single			
MonthlyIncome		6877	4434	2809
6893	8376			
NumCompaniesWorked		1	1	2
6	5			
OverTime		Yes	No	No
No	No			
PercentSalaryHike		24	13	14
15	11			
PerformanceRating		4	3	3
4	3			
RelationshipSatisfaction		2	4	4
4	1			
StockOptionLevel		0	1	0
1	1			
TotalWorkingYears		12	10	8
11	11			
TrainingTimesLastYear		4	3	2
3	3			
WorkLifeBalance		2	2	3
3	3			
YearsAtCompany		0	9	2
7	7			
YearsInCurrentRole		0	8	2
7	6			
YearsSinceLastPromotion		0	7	2
1	7			
YearsWithCurrManager		0	8	2
7	6			
Employee Source		Jora	Company Website	Indeed
Seek	Company Website			
AgeStartedWorking		18	23	32
23	28			

```
# Info
df00.info()
```

RangeIndex: 23058 entries, 0 to 23057
Data columns (total 30 columns):

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	Age	23058 non-null	int64
1	Attrition	23058 non-null	object
2	BusinessTravel	23058 non-null	object
3	Department	23058 non-null	object
4	DistanceFromHome	23058 non-null	int64
5	Education	23058 non-null	int64
6	EducationField	23058 non-null	object
7	EnvironmentSatisfaction	23058 non-null	int64
8	Gender	23058 non-null	object

9	JobInvolvement	23058 non-null	int64
10	JobLevel	23058 non-null	int64
11	JobRole	23058 non-null	object
12	JobSatisfaction	23058 non-null	int64
13	MaritalStatus	23058 non-null	object
14	MonthlyIncome	23058 non-null	int64
15	NumCompaniesWorked	23058 non-null	int64
16	OverTime	23058 non-null	object
17	PercentSalaryHike	23058 non-null	int64
18	PerformanceRating	23058 non-null	int64
19	RelationshipSatisfaction	23058 non-null	int64
20	StockOptionLevel	23058 non-null	int64
21	TotalWorkingYears	23058 non-null	int64
22	TrainingTimesLastYear	23058 non-null	int64
23	WorkLifeBalance	23058 non-null	int64
24	YearsAtCompany	23058 non-null	int64
25	YearsInCurrentRole	23058 non-null	int64
26	YearsSinceLastPromotion	23058 non-null	int64
27	YearsWithCurrManager	23058 non-null	int64
28	Employee Source	23058 non-null	object
29	AgeStartedWorking	23058 non-null	int64

dtypes: int64(21), object(9)
memory usage: 5.3+ MB

Este é o dicionário de dados do dataset:

- **Age:** Idade do funcionário
- **Attrition:** Tipo de atrito (demissão voluntária, etc.)
- **BusinessTravel:** Frequência de viagens a negócios
- **Department:** Departamento de trabalho
- **DistanceFromHome:** Distância de casa ao trabalho
- **Education:** Nível de educação
- **EducationField:** Área de estudo
- **EnvironmentSatisfaction:** Satisfação com o ambiente de trabalho
- **Gender:** Gênero do funcionário
- **JobInvolvement:** Envolvimento no trabalho
- **JobLevel:** Nível do trabalho
- **JobRole:** Função do trabalho
- **JobSatisfaction:** Satisfação com o trabalho
- **MaritalStatus:** Estado civil
- **MonthlyIncome:** Renda mensal
- **NumCompaniesWorked:** Número de empresas em que já trabalhou
- **Over18:** Se o funcionário é maior de 18 anos
- **OverTime:** Se faz hora extra
- **PercentSalaryHike:** Percentual de aumento salarial
- **PerformanceRating:** Avaliação de desempenho
- **RelationshipSatisfaction:** Satisfação com os relacionamentos no trabalho
- **StandardHours:** Horas padrão de trabalho
- **StockOptionLevel:** Nível de opções de ações
- **TotalWorkingYears:** Total de anos trabalhados
- **TrainingTimesLastYear:** Número de treinamentos no último ano
- **WorkLifeBalance:** Equilíbrio entre vida pessoal e profissional
- **YearsAtCompany:** Anos na empresa
- **YearsInCurrentRole:** Anos na função atual
- **YearsSinceLastPromotion:** Anos desde a última promoção
- **YearsWithCurrManager:** Anos com o gerente atual
- **EmployeeSource:** Fonte do funcionário (por exemplo, indicação)
- **AgeStartedWorking:** Idade em que começou a trabalhar

Seleção da Variável Alvo

- Objetivo do projeto: identificar os fatores que contribuem para satisfação dos funcionários.
 - Termination: como esta categoria se refere a funcionários dispensados então não usaremos para analisar satisfação do funcionário.
 - Voluntary Resignation: vamos estudar esta categoria, analisando quais fatores levaram a essa decisão.
 - Current employee: estudaremos a categoria para tentar prever uma possibilidade de saída de funcionário.

```
df00['Attrition'].value_counts()
```

```
Attrition
Current employee      19370
Voluntary Resignation  3601
Termination           87
Name: count, dtype: int64
```

- Após o estudo no dicionário de dados declaramos a variável alvo: "Agora temos uma Análise Supervisionada."

Encode da Variável Alvo e Preparação dos Dados

- Vamos converter o problema em aprendizado supervisionado, com a variável alvo sendo binária. Classes:
 - Classe 0 - Não ocorrência do evento (não pediu demissão).
 - Classe 1 - Ocorrência do evento (pediu demissão).
- Vamos analisar os resultados com base na classe 1 e compreender os fatores que influenciam a satisfação dos funcionários, ou seja, levam os funcionários a pedir demissão.
- A categoria Termination será descartada, pois a decisão de demissão nesse caso foi da empresa e não do funcionário!

```
# Filtrar o dataframe para manter apenas 'Current employee' e 'Voluntary Resignation'
df00 = df00[df00['Attrition'].isin(['Current employee', 'Voluntary Resignation'])]
```

```
# Verificar os valores únicos da coluna 'Attrition' após o filtro
df00['Attrition'].value_counts()
```

```
Attrition
Current employee      19370
Voluntary Resignation  3601
Name: count, dtype: int64
```

```
# Encode da variável alvo
df00['Attrition'] = df00['Attrition'].apply(lambda x: 1 if x == 'Voluntary Resignation' else 0)
```

```
df00['Attrition'].value_counts()
```

```
Attrition
0      19370
1      3601
Name: count, dtype: int64
```

```
# Separação das variáveis
X = df00.drop('Attrition', axis = 1)
```

```
y = df00['Attrition']
```

```
# Divisão em treino e teste
X_treino, X_teste, y_treino, y_teste = train_test_split(X, y, test_size = 0.2, random_state = 42)
```

```
# Separação das variáveis numéricas e categóricas
cat_features = X.select_dtypes(include = ['object']).columns.tolist()
num_features = X.select_dtypes(include = ['int64', 'float64']).columns.tolist()
```

Pipeline de Pré-Processamento de Variáveis Numéricas

O objetivo deste pipeline é garantir que todas as features numéricas no conjunto de dados sejam tratadas de forma consistente e apropriada antes de serem alimentadas no modelo de Machine Learning.

- **Tratamento de valores ausentes:** Substitui valores ausentes pela mediana para evitar que o modelo seja afetado por dados faltantes.
- **Normalização:** Padroniza os dados para que todas as features numéricas tenham a mesma escala, melhorando a performance do modelo e garantindo que nenhuma feature domine as outras devido à escala.

```
# Cria o pipeline
numeric_transformer = Pipeline(steps = [
    ('imputer', SimpleImputer(strategy = 'median')),
    ('scaler', StandardScaler())])
```

Pipeline de Pré-Processamento de Variáveis Categóricas

O objetivo deste pipeline é garantir que todas as features categóricas no conjunto de dados sejam tratadas de forma consistente e apropriada antes de serem alimentadas no modelo de Machine Learning.

- **Tratamento de valores ausentes:** Substitui valores ausentes por 'missing', criando uma categoria especial para valores ausentes.
- **Codificação One-Hot:** Transforma as features categóricas em uma forma que pode ser usada pelo modelo de Machine Learning, convertendo cada categoria em uma coluna binária.

```
# Cria o pipeline
categorical_transformer = Pipeline(steps = [
    ('imputer', SimpleImputer(strategy = 'constant', fill_value = 'missing')),
    ('onehot', OneHotEncoder(handle_unknown = 'ignore'))])
```

Pipeline de Modelagem

- A classe ColumnTransformer permite aplicar diferentes transformações a diferentes subsets de features.
- Isso é útil quando você precisa pré-processar colunas numéricas e categóricas de maneira distinta.

```
# Combinando os passos de pré-processamento
preprocessor = ColumnTransformer(
    transformers = [
        ('num', numeric_transformer, num_features),
        ('cat', categorical_transformer, cat_features)])
```

- O objetivo do pipeline de modelagem é combinar todas as etapas de pré-processamento e modelagem em um único fluxo de trabalho que pode ser aplicado de forma consistente aos dados de treinamento e de teste.
- Isso garante que todas as transformações necessárias sejam aplicadas corretamente e na ordem certa, facilitando a replicação e manutenção do processo.

```
# Criando o pipeline de modelagem
```

```
#modelo = Pipeline(steps = [('preprocessor', preprocessor),('classifier', LogisticRegression(max_iter = 1000))
modelo = Pipeline(steps = [('preprocessor', preprocessor),('classifier', LogisticRegression())])
```

```
# Treina o modelo
```

```
modelo.fit(X_treino, y_treino)
```

```
Pipeline(steps=[('preprocessor',
                  ColumnTransformer(transformers=[('num',
                                                  Pipeline(steps=[('imputer',
                                                                    SimpleImputer(strategy='median')),
                                                                    ('scaler',
                                                                    StandardScaler()))],
                                                                    ['Age', 'DistanceFromHome',
                                                                    'Education',
                                                                    'EnvironmentSatisfaction',
                                                                    'JobInvolvement', 'JobLevel',
                                                                    'JobSatisfaction',
                                                                    'MonthlyIncome',
                                                                    'NumCompaniesWorked',
                                                                    'PercentSalaryHike',
                                                                    'PerformanceRating',
                                                                    '...',
                                                                    'YearsSinceLastPromotion',
                                                                    'YearsWithCurrManager',
                                                                    'AgeStartedWorking ']),
                                                  ('cat',
                                                  Pipeline(steps=[('imputer',
                                                                    SimpleImputer(fill_value='missing',
                                                                    strategy='constant')),
                                                                    ('onehot',
                                                                    OneHotEncoder(handle_unknown='ignore'))],
                                                                    ['BusinessTravel',
                                                                    'Department',
                                                                    'EducationField', 'Gender',
                                                                    'JobRole', 'MaritalStatus',
                                                                    'OverTime',
                                                                    'Employee Source'])])),
                  ('classifier', LogisticRegression())])
```

```
# Previsões com os dados de teste
```

```
y_pred = modelo.predict(X_teste)
```

```
# Avalia o modelo
```

```
acuracia = accuracy_score(y_teste, y_pred)
```

```
print(acuracia)
```

```
0.8550598476605006
```

Analizando os Coeficientes do Modelo

```
# Analizando os coeficientes do modelo
coefficients = modelo.named_steps['classifier'].coef_[0]
```

```
# Nomes dos atributos
feature_names = num_features + list(modelo.named_steps['preprocessor'] \
                                     .transformers_[1][1].named_steps['onehot'].get_feature_names_out(cat_featu
```

```
feature_names
```

```
[ 'Age',
  'DistanceFromHome',
  'Education',
  'EnvironmentSatisfaction',
  'JobInvolvement',
  'JobLevel',
  'JobSatisfaction',
  'MonthlyIncome',
  'NumCompaniesWorked',
  'PercentSalaryHike',
  'PerformanceRating',
  'RelationshipSatisfaction',
  'StockOptionLevel',
  'TotalWorkingYears',
  'TrainingTimesLastYear',
  'WorkLifeBalance',
  'YearsAtCompany',
  'YearsInCurrentRole',
  'YearsSinceLastPromotion',
  'YearsWithCurrManager',
  'AgeStartedWorking ',
  'BusinessTravel_Non-Travel',
  'BusinessTravel_Travel_Frequently',
  'BusinessTravel_Travel_Rarely',
  'Department_Human Resources',
  'Department_Research & Development',
  'Department_Sales',
  'EducationField_Human Resources',
  'EducationField_Life Sciences',
  'EducationField_Marketing',
  'EducationField_Medical',
  'EducationField_Other',
  'EducationField_Technical Degree',
  'Gender_Female',
  'Gender_Male',
  'JobRole_Healthcare Representative',
  'JobRole_Human Resources',
```

```
'JobRole_Laboratory Technician',
'JobRole_Manager',
'JobRole_Manufacturing Director',
'JobRole_Research Director',
'JobRole_Research Scientist',
'JobRole_Sales Executive',
'JobRole_Sales Representative',
'MaritalStatus_Divorced',
'MaritalStatus_Married',
'MaritalStatus_Single',
'Overtime_No',
'Overtime_Yes',
'Employee Source_Adzuna',
'Employee Source_Company Website',
'Employee Source_GlassDoor',
'Employee Source_Indeed',
'Employee Source_Jora',
'Employee Source_LinkedIn',
'Employee Source_Recruit.net',
'Employee Source_Referral',
'Employee Source_Seek']
```

```
# Dataframe
coeff_df = pd.DataFrame({'Atributo': feature_names, 'Coeficiente': coefficients}).sort_values(by = 'Coeficiente')
```

```
# Display
coeff_df.head(10)
```

	Atributo	Coeficiente
22	BusinessTravel_Travel_Frequently	0.494839
32	EducationField_Technical Degree	0.275768
56	Employee Source_Referral	0.257281
46	MaritalStatus_Single	0.213240
37	JobRole_Laboratory Technician	0.213197
48	Overtime_Yes	0.183383
1	DistanceFromHome	0.160642
18	YearsSinceLastPromotion	0.157113
43	JobRole_Sales Representative	0.126106
16	YearsAtCompany	0.105078

Os coeficientes na regressão logística indicam a força e a direção da associação entre cada feature (atributo) e a probabilidade de ocorrência do evento alvo, que neste caso é a demissão voluntária (Attrition).

- Ou seja, a análise dos coeficientes de um modelo de regressão logística nos ajuda a entender a influência de cada atributo na probabilidade do evento de interesse.
 - Os coeficientes positivos indicam que, conforme o valor do atributo aumenta, a probabilidade do funcionário se demitir voluntariamente também aumenta.

Vamos interpretar os 10 coeficientes com maior valor:

BusinessTravel_Travel_Frequently (0.494839)

- Funcionários que viajam frequentemente a negócios têm uma maior probabilidade de se demitir voluntariamente.
 - Esse coeficiente é bastante significativo, sugerindo que a frequência de viagens pode ser um fator de estresse ou insatisfação.

EducationField_Technical Degree (0.275768)

- Funcionários com um diploma técnico têm uma probabilidade maior de se demitir voluntariamente em comparação com aqueles de outras áreas educacionais.
 - Isso pode indicar que esses funcionários têm mais oportunidades no mercado de trabalho ou que suas expectativas não estão sendo atendidas.

Employee Source_Referral (0.257281)

- Funcionários que foram contratados por meio de indicações (Referral) têm uma probabilidade maior de se demitir voluntariamente.
 - Isso pode sugerir que, apesar de serem indicados, eles podem não estar tão alinhados com a empresa quanto outros funcionários.

MaritalStatus_Single (0.213240)

- Funcionários solteiros têm uma maior probabilidade de se demitir voluntariamente em comparação com funcionários casados ou em outros estados civis.
 - Isso pode ser devido à maior flexibilidade e menos responsabilidades pessoais.

JobRole_Laboratory Technician (0.213197)

- Funcionários que trabalham como técnicos de laboratório têm uma maior probabilidade de se demitir voluntariamente.
 - Isso pode indicar insatisfação com a função específica ou o ambiente de trabalho.

OverTime_Yes (0.183383)

- Funcionários que fazem horas extras têm uma maior probabilidade de se demitir voluntariamente.
 - Isso sugere que o excesso de trabalho pode levar ao desgaste e à insatisfação.

DistanceFromHome (0.160642)

- Maior distância de casa para o trabalho está associada a uma maior probabilidade de demissão voluntária.
 - Longos deslocamentos podem causar cansaço e insatisfação.

YearsSinceLastPromotion (0.157113)

- Funcionários que passaram mais anos desde a última promoção têm uma maior probabilidade de se demitir voluntariamente.
 - Isso pode indicar insatisfação com as oportunidades de crescimento na empresa.

JobRole_Sales Representative (0.126106)

- Funcionários que trabalham como representantes de vendas têm uma maior probabilidade de se demitir voluntariamente.
 - Essa função pode ter alta pressão de desempenho ou falta de suporte adequado.

YearsAtCompany (0.105078)

- Quanto mais anos um funcionário passa na empresa, maior é a probabilidade de se demitir voluntariamente.
 - Isso pode indicar que, após um certo período, os funcionários podem sentir estagnação ou buscar novas oportunidades.

Conclusão:

- Os coeficientes positivos indicam que esses fatores aumentam a probabilidade de demissão voluntária.

- Entender esses fatores pode ajudar a empresa a tomar medidas preventivas, como:
 - melhorar as condições de trabalho,
 - oferecer oportunidades de crescimento
 - e minimizar a necessidade de horas extras,
- para reduzir a taxa de rotatividade voluntária.

```
%watermark -a "RobertoSSoares-LfLNgLrnnng"
```

Author: RobertoSSoares-LfLNgLrnnng

```
%watermark -v -m
```

Python implementation: CPython

Python version : 3.10.14

IPython version : 8.27.0

Compiler : MSC v.1916 64 bit (AMD64)

OS : Windows

Release : 10

Machine : AMD64

Processor : Intel64 Family 6 Model 158 Stepping 9, GenuineIntel

CPU cores : 4

Architecture: 64bit

```
%watermark --iversons
```

sklearn: 1.5.2

pandas : None

Logistic Regression

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Fim

Data Science e Machine Learning na Área de Recursos Humanos - (Outros projetos possíveis)

Diversidade e Inclusão Métricas e Modelos Analíticos

- A promoção da diversidade e inclusão (D&I) nas organizações tem por objetivo criar um ambiente de trabalho equitativo e inovador.

- o Para monitorar e aprimorar as iniciativas de D&I, as empresas devem utilizar métricas e modelos analíticos robustos.

- A primeira etapa é a coleta de dados abrangentes sobre a composição da força de trabalho, incluindo
 - o dados demográficos (gênero, raça, idade, orientação sexual, etc.),
 - o além de informações sobre contratação, promoção, remuneração e retenção.
- Esses dados formam a base para medir a diversidade dentro da organização e identificar possíveis lacunas.

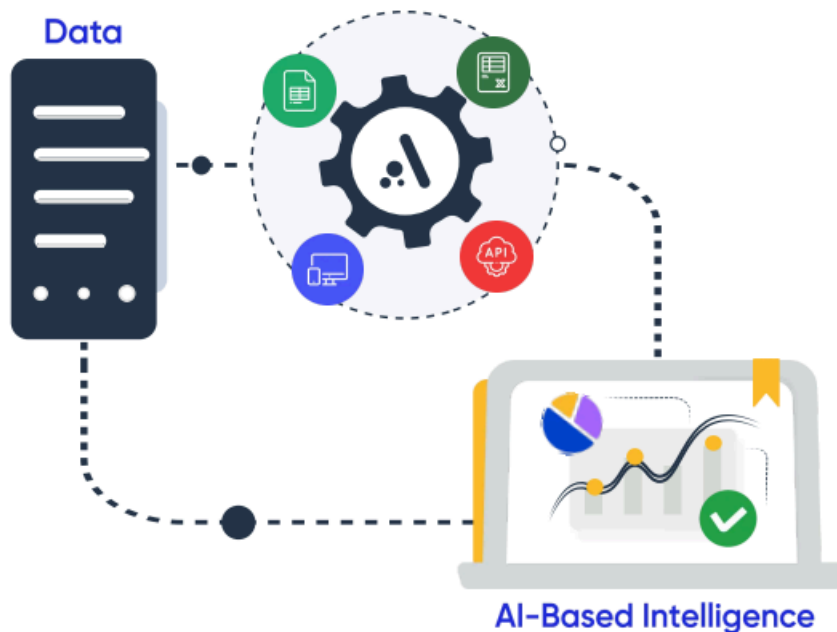


- A análise descritiva é uma abordagem inicial para entender a situação atual da diversidade e inclusão.
- Métricas básicas, como a representação de diferentes grupos demográficos em vários níveis da organização, taxas de contratação e promoção por grupo, e análises de remuneração equitativa, são essenciais.
- Ferramentas de visualização de dados, como gráficos de barras e painéis de controle interativos, ajudam a ilustrar essas métricas de forma clara, permitindo que os líderes identifiquem áreas que necessitam de atenção e melhoria.
- Para aprofundar a análise, técnicas estatísticas avançadas e modelos de Machine Learning podem ser aplicados para identificar fatores que influenciam a diversidade e a inclusão.
 - o Por exemplo, análises de regressão podem ser usadas para entender as disparidades salariais entre diferentes grupos e identificar as causas.
 - o Modelos de classificação podem prever a probabilidade de diferentes grupos serem promovidos ou deixarem a organização, ajudando a identificar possíveis barreiras ou preconceitos no processo de gestão de talentos.
- A análise de redes sociais é outra técnica útil para entender as dinâmicas de inclusão dentro da organização.
 - o Ao mapear as interações e conexões entre os funcionários, a análise pode revelar a existência de silos ou segregação entre diferentes grupos.
 - o Isso permite que as empresas intervenham de forma mais direcionada, promovendo iniciativas de integração e colaboração.
- A análise de sentimentos, aplicada a feedbacks e avaliações, também pode fornecer insights sobre a percepção de inclusão entre os funcionários, identificando áreas onde a cultura organizacional pode ser aprimorada.
- A implementação de um ciclo contínuo de monitoramento e melhoria visa auxiliar o sucesso das iniciativas de D&I.
 - o Isso inclui a definição de metas claras e mensuráveis, o acompanhamento regular das métricas de diversidade e inclusão, e a realização de pesquisas de clima organizacional para coletar feedback direto dos funcionários.

Machine Learning Para Previsão de Necessidades de Treinamento e Desenvolvimento

- O uso de Machine Learning para prever necessidades de treinamento e desenvolvimento oferece uma abordagem proativa e personalizada para o crescimento dos funcionários.

- O ideal é começar com a coleta de dados detalhados sobre o desempenho dos funcionários, histórico de treinamento, avaliações de habilidades, feedbacks de gestores e funcionários, além de metas e resultados de carreira.
- Esses dados vão auxiliar a identificar padrões e lacunas nas competências dos funcionários, estabelecendo a base para a modelagem preditiva.



- A primeira etapa analítica envolve a aplicação de técnicas de análise descritiva para entender a situação atual das habilidades e desempenho dos funcionários.
- Ferramentas de visualização de dados, como dashboards interativos e gráficos de calor, ajudam a destacar áreas de alta e baixa competência, bem como a eficiência dos treinamentos realizados anteriormente.
- Essa análise fornece insights iniciais sobre quais áreas podem necessitar de mais atenção em termos de desenvolvimento.
- Em seguida, modelos de Machine Learning, como regressão logística, árvores de decisão e redes neurais, são treinados para prever as necessidades de treinamento futuras.
- Esses modelos analisam variáveis como desempenho atual, tempo desde o último treinamento, mudanças nas responsabilidades do trabalho, e feedback de desempenho.
 - Por exemplo, um modelo pode prever que um funcionário que recentemente assumiu novas responsabilidades precisará de treinamento adicional em gestão de projetos ou liderança.
 - A validação dos modelos visa garantir precisão e relevância das previsões.
- Além de prever as necessidades de treinamento, Machine Learning também pode ser utilizado para personalizar os programas de desenvolvimento para cada funcionário.
- Sistemas de recomendação, semelhantes aos usados em plataformas de streaming, podem sugerir cursos específicos, workshops ou programas de mentoria baseados no perfil de habilidades e interesses do funcionário.
- Isso não apenas otimiza os recursos de treinamento, mas também aumenta o engajamento dos funcionários ao oferecer oportunidades de desenvolvimento alinhadas com suas aspirações e necessidades específicas.
- Finalmente, o ciclo de feedback contínuo é vital para o sucesso das iniciativas de treinamento e desenvolvimento baseadas em Machine Learning.
 - Após a implementação dos programas de treinamento recomendados, os dados de desempenho dos funcionários são continuamente monitorados para avaliar a eficácia dos treinamentos.
- Ajustes e melhorias são feitos com base nesses feedbacks, garantindo que os programas de desenvolvimento evoluam conforme as necessidades dos funcionários e da organização mudam.
- Essa abordagem dinâmica e orientada por dados garante que as empresas mantenham uma força de trabalho capacitada, engajada e preparada para enfrentar desafios futuros.

Análise de Engajamento de Funcionários e Feedback 360 Graus

- A análise de engajamento de funcionários e o feedback 360 graus são essenciais para entender e melhorar o ambiente de trabalho, promovendo uma cultura organizacional positiva e produtiva.
- A coleta de dados para essas análises envolve a utilização de pesquisas de engajamento regulares, ferramentas de feedback 360 graus, dados de desempenho e outras métricas relacionadas ao ambiente de trabalho.
- Essas informações fornecem uma visão abrangente do nível de satisfação e engajamento dos funcionários, além de identificar áreas de melhoria.



- A análise de engajamento de funcionários começa com a aplicação de pesquisas estruturadas que medem diferentes aspectos do trabalho,
 - como satisfação com o papel,
 - relacionamento com a liderança,
 - oportunidades de desenvolvimento
 - e ambiente de trabalho.
- As respostas são analisadas utilizando técnicas estatísticas para calcular índices de engajamento, identificar tendências e segmentar os dados por departamento, nível hierárquico ou outras categorias relevantes.
 - Visualizações de dados, como gráficos de barras e radar, ajudam a ilustrar os resultados e facilitam a identificação de áreas que necessitam de intervenção.
- O feedback 360 graus é uma ferramenta que fornece uma avaliação completa do desempenho de um funcionário a partir de múltiplas perspectivas: autoavaliação, feedback de colegas, subordinados e superiores.
- Esse processo gera um conjunto de dados rico e detalhado sobre as competências, comportamentos e áreas de desenvolvimento do funcionário.
 - Análises quantitativas e qualitativas dessas avaliações ajudam a identificar pontos fortes e fracos, permitindo a criação de planos de desenvolvimento personalizados.
- Ferramentas de análise de texto podem ser usadas para extrair insights de feedbacks abertos, identificando temas e sentimentos comuns.
- A integração das análises de engajamento e feedback 360 graus permite uma visão holística do ambiente de trabalho e do desenvolvimento dos funcionários.
- As empresas podem correlacionar dados de engajamento com resultados de feedback 360 graus para identificar como diferentes fatores impactam o desempenho e a satisfação dos funcionários.
 - Por exemplo, pode-se descobrir que funcionários com alto engajamento também recebem feedbacks mais positivos em áreas específicas, sugerindo práticas eficazes que podem ser replicadas em toda a organização.
- Para maximizar o impacto dessas análises, é aconselhável implementar um ciclo contínuo de feedback e melhoria.
 - Isso inclui a comunicação transparente dos resultados das pesquisas e feedbacks, o desenvolvimento de planos de ação com base nos insights obtidos e o monitoramento regular do progresso.

- Sessões de follow-up e novas avaliações ajudam a garantir que as iniciativas de melhoria estejam no caminho certo e que os funcionários se sintam ouvidos e valorizados.
 - A utilização de plataformas de engajamento e feedback em tempo real pode facilitar esse processo, promovendo uma cultura de feedback contínuo e ágil.
- A análise de engajamento de funcionários e o feedback 360 graus além de melhoram o ambiente de trabalho, também aumentam a produtividade, a retenção de talentos e a satisfação geral dos funcionários.
- Ao adotar uma abordagem orientada por dados para entender e responder às necessidades dos funcionários, as empresas podem criar um local de trabalho mais harmonioso e eficiente, impulsionando o sucesso organizacional a longo prazo.

RobertoSSoares-LfLnGLrnng

20/02/2025, 20:05:49