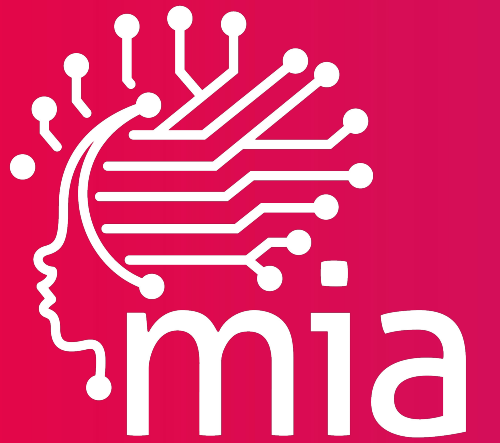


Desafio - Claudia dos Anjos

#BootcampMIA2022 #SomosMIA

Fatos curiosos

Claudia dos Anjos

- Trilheira
- Guia de Turismo
- Dona de Hortinha
- Apicultora
- Adora doramas
- Aprendendo coreano

Atividade do Bootcamp

README.md

mia
mulheres em inteligência artificial

Descrição

Bootcamp de Dados do Mulheres em IA.

Aulas

- 1 - Introdução à Ciência de Dados
- 2 - Introdução à Engenharia de Dados
- 3 - Análise Exploratória de Dados
- 4 - Visualização de Dados
- 5 - Regressão
- 6 - Classificação (parte 1)
- 7 - Classificação (parte 2)
- 8 - Clustering
- 9 - Ensemble

CLAUDIAANJOS

Saída de funcionários de assistência médica

Desafio

Os dados desse desafio são sintéticos e baseados no conjunto de dados do IBM Watson para saída de funcionários. As funções e os departamentos dos funcionários foram alterados para refletir o domínio da saúde. O seu desafio é utilizar esse conjunto de dados para predizer quando uma pessoa sairá da área médica (attrition).

Dataset

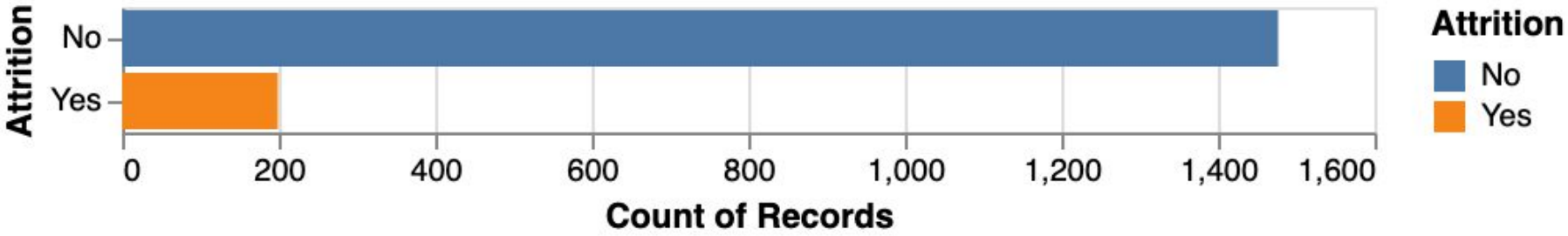
EmployeeID	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement	JobLevel	JobRole	JobSatisfaction	MaritalStatus
1313919	41	No	Travel_Rarely	1102	Cardiology	1	2	Life Sciences	1	2	Female	94	3	2	Nurse	4	Single
1200302	49	No	Travel_Frequently	279	Maternity	8	1	Life Sciences	1	3	Male	61	2	2	Other	2	Married
1060315	37	Yes	Travel_Rarely	1373	Maternity	2	2	Other	1	4	Male	92	2	1	Nurse	3	Single
1272912	33	No	Travel_Frequently	1392	Maternity	3	4	Life Sciences	1	4	Female	56	3	1	Other	3	Married
1414939	27	No	Travel_Rarely	591	Maternity	2	1	Medical	1	1	Male	40	3	1	Nurse	2	Married
1633361	32	No	Travel_Frequently	1005	Maternity	2	2	Life Sciences	1	4	Male	79	3	1	Nurse	4	Single
1329390	59	No	Travel_Rarely	1324	Maternity	3	3	Medical	1	3	Female	81	4	1	Nurse	1	Married
1699288	30	No	Travel_Rarely	1358	Maternity	24	1	Life Sciences	1	4	Male	67	3	1	Nurse	3	Divorced
1469740	38	No	Travel_Frequently	216	Maternity	23	3	Life Sciences	1	4	Male	44	2	3	Therapist	3	Single
1101291	36	No	Travel_Rarely	1299	Maternity	27	3	Medical	1	3	Male	94	3	2	Nurse	3	Married
1430504	35	No	Travel_Rarely	809	Maternity	16	3	Medical	1	1	Male	84	4	1	Nurse	2	Married
1196281	29	No	Travel_Rarely	153	Maternity	15	2	Life Sciences	1	4	Female	49	2	2	Nurse	3	Single
1207951	31	No	Travel_Rarely	670	Maternity	26	1	Life Sciences	1	1	Male	31	3	1	Other	3	Divorced
1080660	34	No	Travel_Rarely	1346	Maternity	19	2	Medical	1	2	Male	93	3	1	Nurse	4	Divorced
1420391	28	Yes	Travel_Rarely	103	Maternity	24	3	Life Sciences	1	3	Male	50	2	1	Nurse	3	Single
1337254	29	No	Travel_Rarely	1389	Maternity	21	4	Life Sciences	1	2	Female	51	4	3	Therapist	1	Divorced
1421335	32	No	Travel_Rarely	334	Maternity	5	2	Life Sciences	1	1	Male	80	4	1	Other	2	Divorced
1262683	22	No	Non-Travel	1123	Maternity	16	2	Medical	1	4	Male	96	4	1	Nurse	4	Divorced
1599218	53	No	Travel_Rarely	1219	Cardiology	2	4	Life Sciences	1	1	Female	78	2	4	Administrative	4	Married
1634788	38	No	Travel_Rarely	371	Maternity	2	3	Life Sciences	1	4	Male	45	3	1	Other	4	Single
1334559	24	No	Non-Travel	673	Maternity	11	2	Other	1	1	Female	96	4	2	Therapist	3	Divorced
1402945	36	No	Travel_Rarely	1218	Cardiology	9	4	Life Sciences	1	3	Male	82	2	1	Other	1	Single
1705718	34	No	Travel_Rarely	419	Neurology	7	4	Life Sciences	1	1	Female	53	3	3	Other	2	Single
1880078	21	No	Travel_Rarely	391	Maternity	15	2	Life Sciences	1	3	Male	96	3	1	Other	4	Single
1801346	34	No	Travel_Rarely	699	Maternity	6	1	Medical	1	2	Male	83	3	1	Other	1	Single
1142160	53	No	Travel_Rarely	1282	Neurology	5	3	Other	1	3	Female	58	3	5	Other	3	Divorced
1142062	32	Yes	Travel_Frequently	1125	Maternity	16	1	Life Sciences	1	2	Female	72	1	1	Other	1	Single
1789756	42	No	Travel_Rarely	691	Cardiology	8	4	Marketing	1	3	Male	48	3	2	Nurse	2	Married

Entendendo nosso dataset

- Temos 1676 linhas e 35 colunas;
- Apenas dois tipos de variáveis: int64 e object;
- Não temos valores faltantes e duplicatas;
- Verificamos a quantidade de valores únicos de cada coluna;
- Verificamos a correlação das variáveis.

Attrition

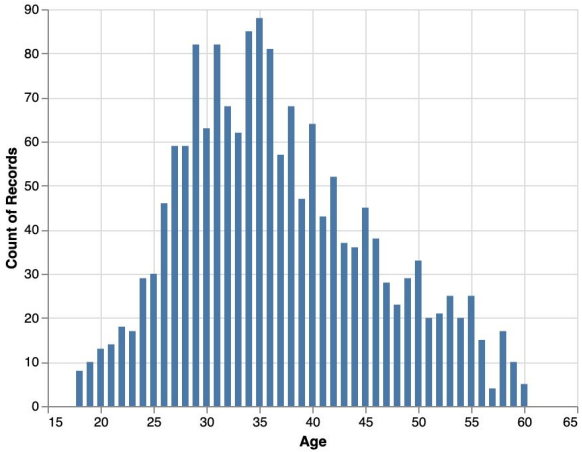
Quantidade de funcionários que saíram da empresa.

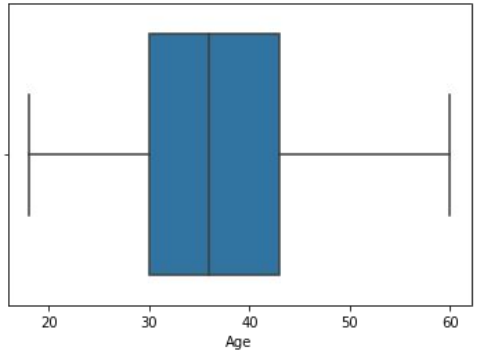


Legenda: Attrition: Saída; No: Não; Yes: Sim; Count of Records: Quantidade.

Idade

Quantidade de funcionários pela idade.

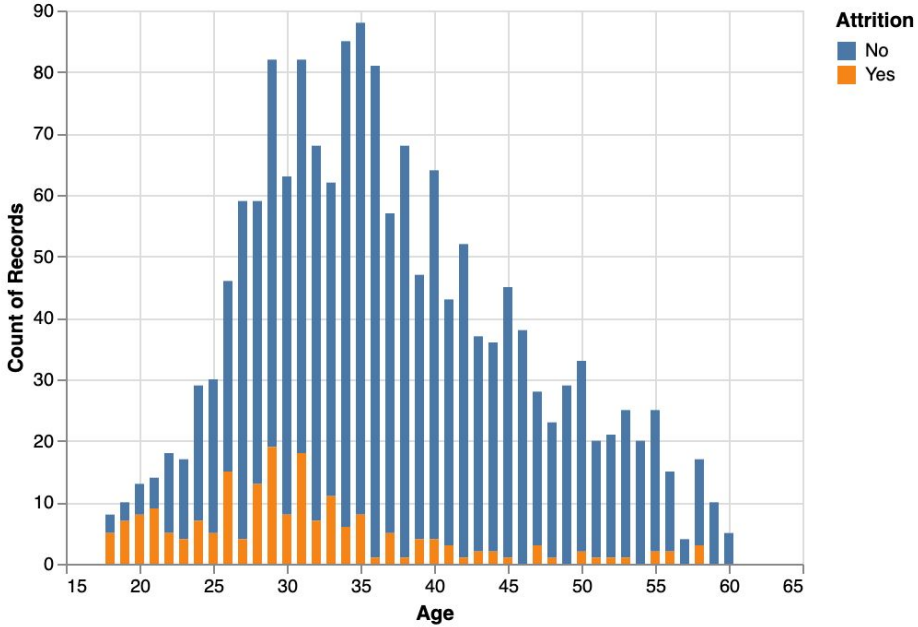




Média 36.86 anos
DP 9.12

Legenda: Age: Idade, Count of Records: Quantidade.

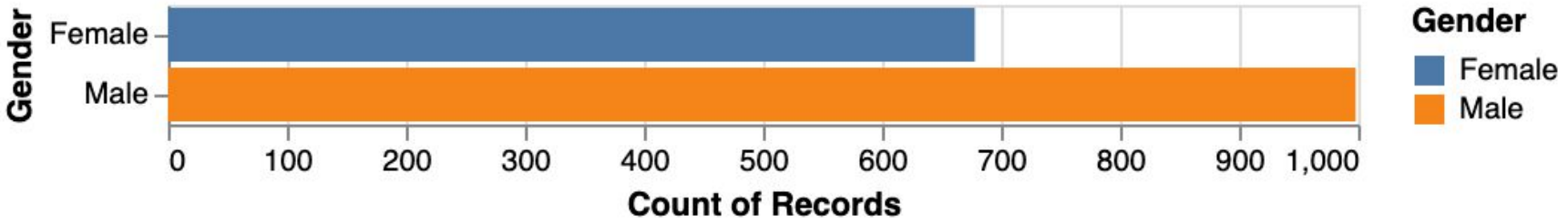
Idade



Legenda: Age: Idade, Attrition: Saída; No: Não; Yes: Sim; Count of Records: Quantidade.

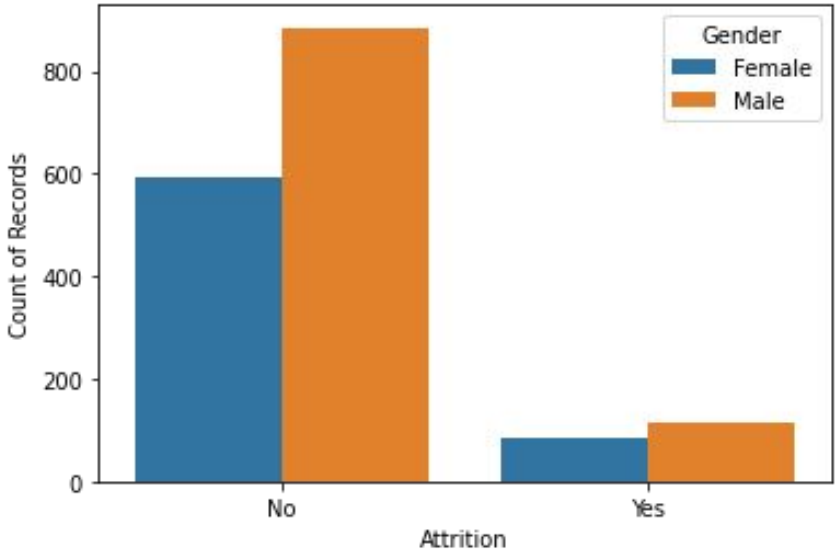
Gênero

Quantidade de funcionários pelo gênero.



Legenda: Gender: Gênero, Female: Feminino, Male: Masculino, Count of Records: Quantidade.

Gênero



Legenda: Gender: Gênero, Female: Feminino, Male: Masculino, Attrition: Saída; No: Não; Yes: Sim; Count of Records: Quantidade.

Construindo o modelo

Conversão das variáveis categóricas

MIA

```
transformer = make_column_transformer(
    (OneHotEncoder(), ['OverTime']), remainder='passthrough')

transformed = transformer.fit_transform(df)
transformed_df = pd.DataFrame(transformed, columns=transformer.get_feature_names())
```

Treinamento

MIA

```
from sklearn.model_selection import train_test_split

df = transformed_df

X = df.drop(["Attrition"], axis=1)
y = df["Attrition"]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=44)
```

Modelo: Random Forest

MIA

```
from imblearn.ensemble import BalancedRandomForestClassifier

rf_model = BalancedRandomForestClassifier(n_estimators=30, max_features="auto",
random_state=44, oob_score=True)
rf_model.fit(X_train, y_train)
```

Importância das features

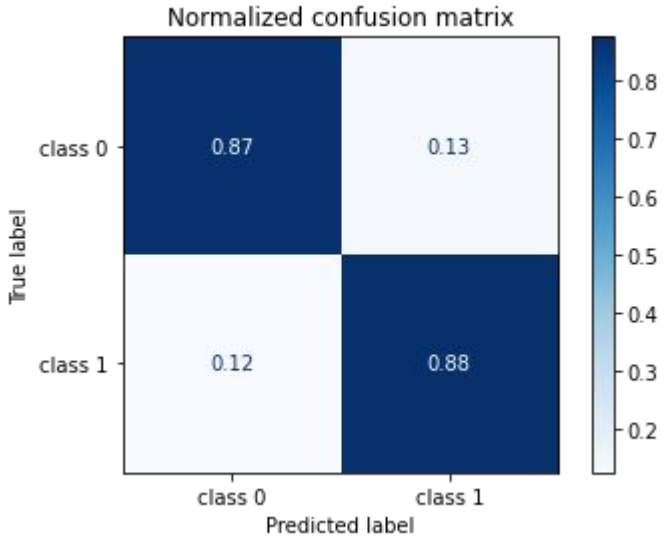
MIA

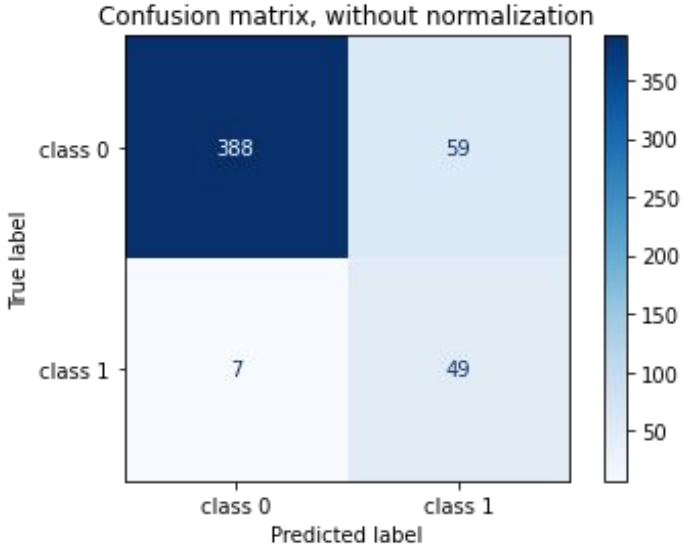
```
importances = rf_model.feature_importances_
columns = X.columns
i = 0

while i < len(columns):
  print(f" A importância da feature '{columns[i]}' é {round(importances[i] * 100, 2)}%.")
  i += 1
```

Features mais importantes

- Age
- MonthlyIncome
- TotalWorkingYears
- YearsAtCompany

Matriz de Confusão





Métricas

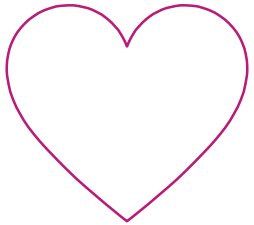
MIA

```
from sklearn.metrics import classification_report

target_names = ['class 0', 'class 1']
print(classification_report(y_true, y_pred, target_names=target_names))
```

Métricas

MIA

	precision	recall	f1-score	support
class 0	0.98	0.87	0.92	447
class 1	0.45	0.88	0.60	56
accuracy			0.87	503
macro avg	0.72	0.87	0.76	503
weighted avg	0.92	0.87	0.89	503



Agradecimentos