

Análise exploratória e preditiva sobre os dados de um questionário sobre a empregabilidade de uma escola técnica no estado do Amazonas

Autor: MSc. Nicolas Sampaio Bevilaqua

1-Introdução:

O seguinte estudo foi realizado para se observar a situação empregatícia dos alunos de uma escola técnica do estado do Amazonas, no município de Manaus, por meio de um questionário realizado no mês de junho de 2021.

Devido a pandemia a situação econômica e consequentemente empregatícia se agravou no estado e no país de forma geral, dessa forma via-se a necessidade de realizar um estudo sobre a situação em alunos finalistas de cursos técnicos.

A escola em questão oferece cursos técnicos nas áreas de ambiente e saúde, gestão e negócios entre outros atendendo alunos da região metropolitana de Manaus.

Durante a análise algumas perguntas relevantes surgiram como:

- Que tipo de variáveis influem na empregabilidade dos alunos?
- A idade e gênero influenciam na obtenção de uma atividade remunerada na área?
- A formação básica, a área e/ou curso específico influem na obtenção de emprego na área de formação?

Por meio da exploração de dados via JupiterLab, utilizando a linguagem de programação em python foi possível obter conclusões sobre essas perguntas e como elas influenciam na inclusão no mercado de trabalho.

Este trabalho está dividido basicamente em duas etapas: na primeira etapa é realizada uma análise exploratória dos dados, onde são observadas as principais variáveis de estudo e em um segundo momento, são utilizadas técnicas de machine learning para obtenção de modelos para prever a empregabilidade dos alunos bem como a extração de outras informações pertinentes ao estudo.

2-Análise exploratória dos dados:

Importação de bibliotecas:

```
import pandas as pd
import seaborn as srn
import seaborn as sns
import statistics as sts
import numpy as np
import matplotlib.pyplot as plt
from sklearn.tree import export_graphviz
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import confusion_matrix, accuracy_score
```

Carregando o dataset e descrição de linhas e colunas:

```
cet=pd.read_excel('Aval_cet_f_v8.xlsx')
cet.shape()
```

(175, 14)

Dicionário de dados:

Tipo das variáveis: Todas as variáveis do estudo são categóricas, isso facilitou a limpeza dos dados, realizada previamente em Excel, utilizando os dados das modas dos atributos.

Qual sua idade: Menos de 18,19-22,23-26,27-30 e acima de 30.

Qual seu gênero: Masculino e feminino

Qual sua renda familiar: Menos de 1.000, entre 1.000 e 2.000, entre 2.000 e 3.000, entre 3.000 e 4.000, 4.000 e 5.000 e acima de 5.000.

Qual região da cidade que você mora: Norte, sul, leste, oeste, centro oeste e centro sul.

Qual sua escola: Padre Estelio Dalison, IBC Benjamin Constant e Saavedra

Qual o percentual da sua formação já foi concluída: Acima de 75%

Qual o tipo do seu curso: Especializações técnicas e Cursos técnicos

Qual o eixo tecnológico: Ambiente e saúde, gestão e negócios, segurança, informação e comunicação etc.

Qual seu curso: Radiologia, saúde bucal, administração etc.

Qual sua formação regular: Ensino médio completo, superior incompleto e superior completo.

Qual sua disponibilidade de horário para atividade remunerada: manhã, tarde, noite e sem disponibilidade

Qual sua carga horária disponível: 4h,6h,8h e sem disponibilidade.

Possui experiência na área de formação do curso: Sim e não

Variável alvo: Está exercendo atividade remunerada na área de formação do curso: Sim e não

`cet.dtypes()`

```
cet.dtypes
Qual sua idade?                object
Qual seu genero?               object
Qual sua renda familiar?       object
Qual regioao da cidade voce mora? object
Qual a sua escola?             object
Qual o percentual da sua formacao no CETAM ja foi concluida? object
Qual tipo do seu curso?        object
Qual o eixo tecnologico?        object
Qual o seu curso?              object
Qual sua formacao regular?      object
Qual sua disponibilidade de horario para atividade remunerada? object
Qual sua carga horaria disponivel? object
Possui experiencia na area de formacao do curso? object
Esta exercendo atividade remunerada na area de formacao do curso? object
dtype: object
```

Carregando o dataset e descrição de linhas e colunas:

```
cet=pd.read_excel('Aval_cetam_f_v8.xlsx')
```

```
cet.shape
```

```
(175, 14)
```

Significa que foram preenchidos 175 questionários, de uma população que não possível estimar devido falta de dados da organização em questão. A base possui 14 variáveis (colunas) as quais já foram descritas no item acima.

Visualização das 5 primeiras instancias:

```
cet.shape()
```

	Qual sua idade?	Qual seu genero?	Qual sua renda familiar?	Qual regioao da cidade voce mora?	Qual a sua escola?	Qual o percentual da sua formacao no CETAM ja foi concluida?	Qual tipo do seu curso?	Qual o eixo tecnologico?	Qual o seu curso?	Qual sua formacao regular?
0	19-22	Feminino	Entre R\$1.000eR\$2.000	Norte	Saavedra	Acima de 75% (Finalista)	Cursos Tecnicos	Ambiente e saude	tecnico em enfermagem	Ensino Medio Completo
1	Acima de 30	Masculino	Menos de R\$1.000	Norte	Saavedra	Acima de 75% (Finalista)	Cursos Tecnicos	Ambiente e saude	Radiologia	Ensino Medio Completo
2	23-26	Feminino	Menos de R\$1.000	Norte	Saavedra	Acima de 75% (Finalista)	Cursos Tecnicos	Ambiente e saude	Radiologia	Ensino Medio Completo
3	19-22	Masculino	Entre R\$1.000eR\$2.000	Norte	Saavedra	Acima de 75% (Finalista)	Cursos Tecnicos	Ambiente e saude	Radiologia	Ensino Medio Completo
4	Acima de 30	Feminino	Menos de R\$1.000	Leste	Saavedra	Acima de 75% (Finalista)	Cursos Tecnicos	Ambiente e saude	Hemoterapia	Superior Incompleto

Verificando a quantidade de valores de cada coluna:

```
cet.count()
```

```
Qual sua idade? 175
Qual seu genero? 175
Qual sua renda familiar? 175
Qual regioao da cidade voce mora? 175
Qual a sua escola? 175
Qual o percentual da sua formacao no CETAM ja foi concluida? 175
Qual tipo do seu curso? 175
Qual o eixo tecnologico? 175
Qual o seu curso? 175
Qual sua formacao regular? 175
Qual sua disponibilidade de horario para atividade remunerada? 175
Qual sua carga horaria disponivel? 175
Possui experiencia na area de formacao do curso? 175
Esta exercendo atividade remunerada na area de formacao do curso? 175
dtype: int64
```

Conforme já exposto, pelo fato de os atributos serem categóricos o processo de limpeza e tratamento de dados foi simplificado, os valores faltantes foram preenchidos com as modas dos atributos expostos acima.

```
#Distribuição estatística por variável com a MODA (Valor que mais se repete)
```

```
cet.describe().T
```

	count	unique	top	freq
Qual sua idade?	175	4	Acima de 30	72
Qual seu genero?	175	2	Feminino	121
Qual sua renda familiar?	175	6	Entre R1.000e R 2.000	77
Qual regioao da cidade voce mora?	175	6	Norte	79
Qual a sua escola?	175	3	Saavedra	142
Qual o percentual da sua formacao no CETAM ja foi concluida?	175	1	Acima de 75% (Finalista)	175
Qual tipo do seu curso?	175	2	Cursos Tecnicos	146
Qual o eixo tecnologico?	175	7	Ambiente e saude	140
Qual o seu curso?	175	16	Radiologia	45
Qual sua formacao regular?	175	3	Ensino Medio Completo	136
Qual sua disponibilidade de horario para atividade remunerada?	175	4	Manha	69
Qual sua carga horaria disponivel?	175	4	8h	68
Possui experiencia na area de formacao do curso?	175	2	NAO	118
Esta exercendo atividade remunerada na area de formacao do curso?	175	2	NAO	155

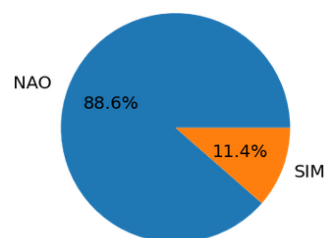
É possível verificar pela moda, que o perfil da pessoa entrevistada pode ser analisado como mulher, acima de 30 anos, renda entre 1000 e 2000 reais, morando na região norte, estudante da área da saúde, do curso de radiologia.

2.1 ANÁLISE EM AMBITOS GERAIS:

Em um primeiro momento, foram analisados os dados de forma generalista, sendo extraídos em gráficos de pizza para facilitar a compreensão. A instituição em questão não possui dados para serem utilizados como parâmetro.

2.1.1. ESTÁ EXERCENDO ATIVIDADE REMUNERADA NA ÁREA DE FORMAÇÃO DO CURSO?

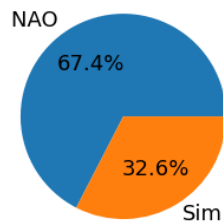
```
[14]: cet['Esta exercendo atividade remunerada na area de formacao do curso?'].value_counts().plot.pie(autopct='%1.1f%%',
figsize=(8,6
)).set_ylabel('')
```



Apenas 11,4% dos alunos estão exercendo atividade remunerada na área de atuação, um número que pode ser considerado baixo, mesmo diante das questões econômicas decorrentes da pandemia ocasionada pela COVID 19.

2.1.2 POSSUI EXPERIÊNCIA NA ÁREA DE FORMAÇÃO DO CURSO?

```
[7]: cet['Possui experiencia na area de formacao do curso?'].value_counts().plot.pie(fontsize=18, autopct='%1.1f%%',
figsize=(6,4)).set_ylabel('')
```

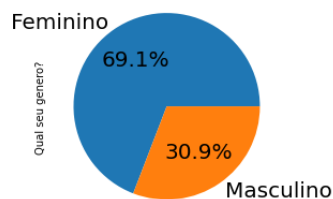


Apenas 32,6% dos alunos possuem algum tipo de experiência na área de atuação. No caso específico da instituição, parte considerável dos dados se justifica pelo estágio obrigatório supervisionado dos discentes na área de saúde, onde os alunos participam de um estágio sem qualquer tipo de ajuda de custo.

2.1.3 QUAL O SEU GÊNERO?

```
[20]: cet['Qual seu genero?'].value_counts().plot.pie(fontsize=20,autopct='%1.1f%%')
```

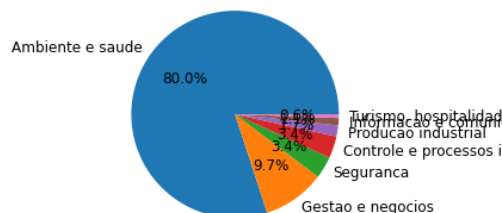
```
[20]: <AxesSubplot:ylabel='Qual seu genero?'>
```



Ratificando o exposto na seção do describe, a maior parte dos finalistas são do gênero feminino (69,1%) contra (30,9%) do gênero masculino.

2.1.4 QUAL O EIXO TECNOLÓGICO?

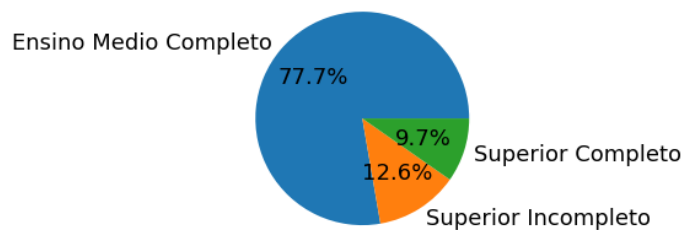
```
[23]: cet['Qual o eixo tecnologico?'].value_counts().plot.pie(fontsize=12,autopct='%1.1f%%',
figsize=(6,4)).set_ylabel('')
```



Os cursos da área da saúde correspondem por 80% dos finalistas, em 2º lugar aparecem os cursos de gestão e negócios com 9,7% e segurança com 3,4%.

2.1.5 QUAL SUA FORMAÇÃO REGULAR?

```
cet['Qual sua formacao regular?'].value_counts().plot.pie(fontsize=18,autopct='%1.1f%%',
figsize=(6,4)).set_ylabel('')
```



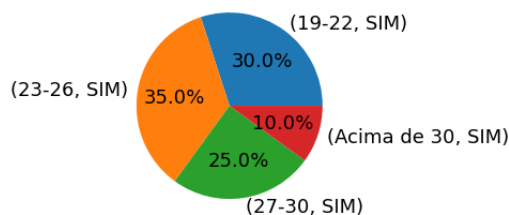
A considerável maioria dos alunos possuem apenas o ensino médio(77,7%), 12,6% declararam superior incompleto e 9,7% apresentaram o superior completo.

3 INFLUENCIA DAS VARIÁVEIS SOBRE A EMPREGABILIDADE

Nesta seção foi verificada a possível relação entre variáveis no quesito em relação a empregabilidade.

3.1 Empregabilidade e idade.

```
[51]: #INFLUENCIA DA IDADE NA EMPREGABILIDADE DOS ALUNOS
gnf1 = cet[cet['Esta exercendo atividade remunerada na area de formacao do curso?']=='SIM']
gnf1.groupby(u'Qual sua idade?')['Esta exercendo atividade remunerada na area de formacao do curso?'].value_counts().plot.pie(fontsize=18,aut
figsize=(6,4)).set_ylabel('')
```

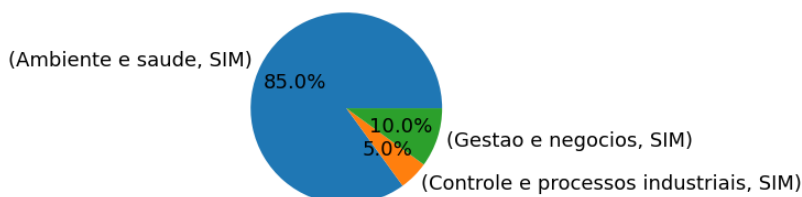


O gráfico demonstra uma possível correlação entre a idade e a empregabilidade em nível técnico. Conforme exposto pela moda na seção describe da seção dos dados, a moda da amostragem de dados é acima de 30 anos, entretanto a mesma classe foi a que apresentou o menor índice de empregabilidade.

A classe somada de 19 a 26 representam 65% do total de pessoa empregadas.

3.2 Empregabilidade e área de atuação

```
[54]: #INFLUENCIA DO EIXO TECNOLÓGICO NA EMPREGABILIDADE DOS ALUNOS
gnf1 = cet[cet['Esta exercendo atividade remunerada na area de formacao do curso?']=='SIM']
gnf1.groupby(u'Qual o eixo tecnologico?')['Esta exercendo atividade remunerada na area de formacao do curso?'].value_counts().plot.pie(fontsi
figsize=(6,4)).set_ylabel('')
```



Conforme exposto na seção describe e no percentual de alunos por área de atuação, a área da saúde possui a maior parcela dos alunos, e na participação dos finalistas exercendo atividade remunerada, a participação sobe para 85%.

3.3 Empregabilidade e cursos específicos

#Contagem de entrevistados por curso

```
x2=cet['Qual o seu curso?'].value_counts()
```

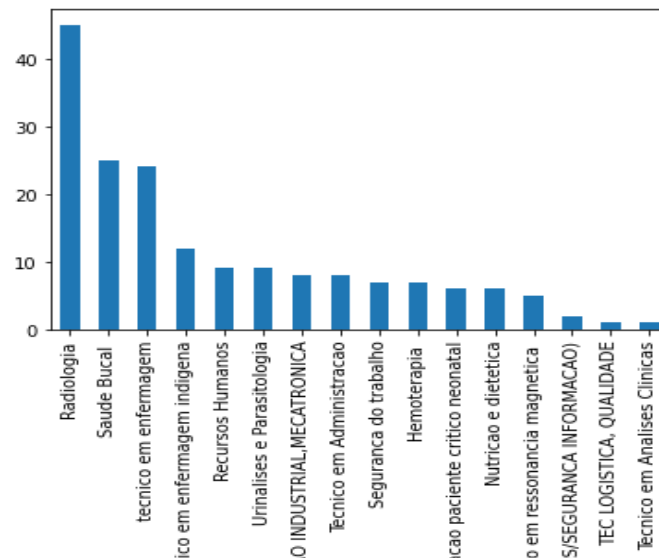
```
x2
```

Radiologia	45
Saude Bucal	25
tecnico em enfermagem	24
Tecnico em enfermagem indigena	12
Recursos Humanos	9
Urinalises e Parasitologia	9
TEC EM ELETRONICA, AUTOMACAO INDUSTRIAL,MECATRONICA	8
Tecnico em Administracao	8
Seguranca do trabalho	7
Hemoterapia	7
Especializacao paciente critico neonatal	6
Nutricao e dietetica	6
Especializacao em ressonancia magnetica	5
ESP TEC NIVEL MEDIO INFORMATICA (BANCO DE DADOS/SEGURANCA INFORMACAO)	2
TEC LOGISTICA, QUALIDADE	1
Tecnico em Analises Clinicas	1

#Contagem de entrevistados por curso no gráfico de barras

```
X3=cet['Qual o seu curso?'].value_counts().plot.bar()
```

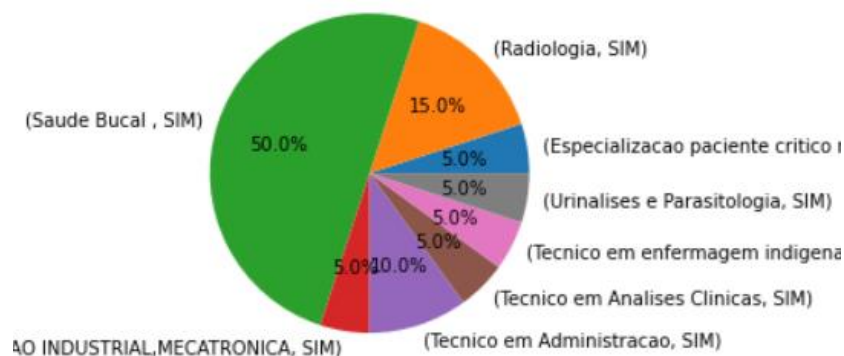
```
X3
```



```
#grafico de pizza para visualização percentual
```

```
df2 = cet[cet['Esta exercendo atividade remunerada na area de formacao do curso?']=='SIM']
```

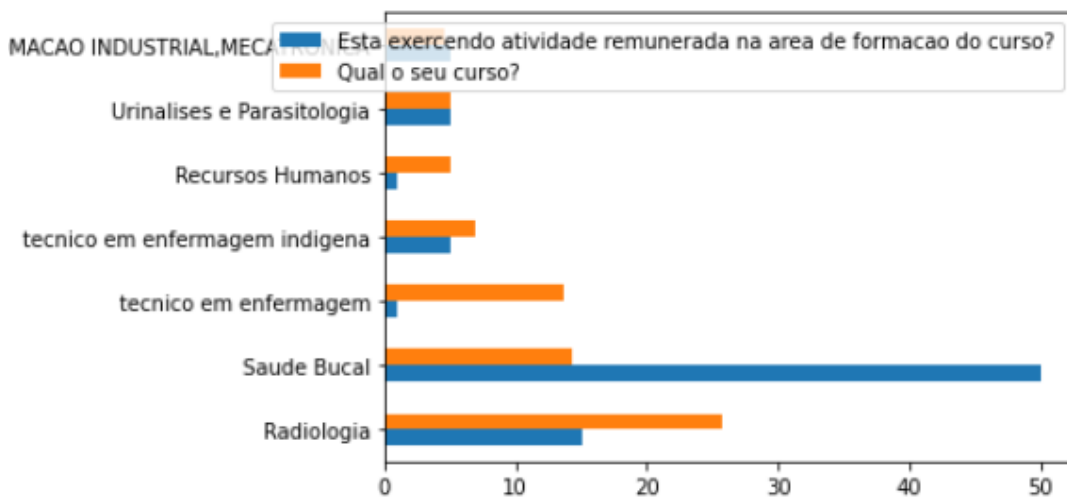
```
df2.groupby(u'Qual o seu curso?')['Esta exercendo atividade remunerada na area de formacao do curso?'].value_counts().plot.pie(fontsize=10,autopct='%1.1f%%',figsize=(8,4)).set_ylabel('')
```

Dos entrevistados, o curso de Saúde Bucal apresentou a maior participação com 50% do total de pessoas empregadas na área de atuação, seguido de radiologia (15%) e técnico de análises clínicas (10%).

Para verificar visualmente a participação percentual de cada curso plotando as barras de (%Está exercendo atividade remunerada? Vs %Qual o seu curso?)

```
atividadederemunerada = [15,50,1,5,1,5,5]
curso = [25.7,14.3,13.7,6.9,5.1,5.1,4.6]
index = ['Radiologia', 'Saude Bucal', 'tecnico em enfermagem',
        'tecnico em enfermagem indigena', 'Recursos Humanos', 'Urinalises e
        Parasitologia', 'TEC EM ELETRONICA, AUTOMACAO INDUSTRIAL,MECATRONICA']
df = pd.DataFrame({'Esta exercendo atividade remunerada na area de formacao do
curso?': actividadederemunerada,
                  'Qual o seu curso?': curso},index=index)
ax = df.plot.barh()
```



É possível identificar um destaque positivo, o curso de saúde bucal, representa 14,3% dos entrevistados, entretanto representa 50% de todos os entrevistados na categoria trabalhando na área, isso fica evidente pois a barra em azul (% atividade remunerada) é muito superior a barra em laranja(% qual seu curso?).

Existem os cursos regulares onde as barras se equivalem, como automação industrial e mecatrônica, urinalises e parasitologia e técnico enfermagem indígena.

Por últimos os destaques negativos, onde a barra laranja é superior a barra azul, onde é possível verificar os cursos de Radiologia, técnico em enfermagem e recursos humanos.

4 CONSTRUÇÃO E AVALIAÇÃO DOS MODELOS DE MACHINE LEARNING

4.1 MODELO DE ARVORE DE DECISÃO

Foi utilizada a função `DecisionTreeClassifier()` da biblioteca `sklearn` para obtenção do modelo de árvore de decisão.

```
[27]: #DIVISAO DA BASE DE TREINAMENTO E TESTE (30% testar e 70% para treinar)
      X_treinamento, X_teste, y_treinamento, y_teste= train_test_split(previsores,
                                                                    classe,
                                                                    test_size=0.3,
                                                                    random_state=0)

[28]: arvore = DecisionTreeClassifier()
      arvore.fit(X_treinamento,y_treinamento)

[28]: DecisionTreeClassifier()

[29]: #OBTENCAO DAS PREVISOES
      previsoes=arvore.predict(X_teste)
      previsoes

[29]: array(['NAO ', 'SIM', 'SIM', 'NAO ', 'NAO ', 'SIM', 'NAO ', 'SIM', 'NAO ',
            'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
            'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
            'NAO ', 'SIM', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
            'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'SIM', 'NAO ',
            'NAO ', 'NAO ', 'SIM', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
            'NAO ', 'NAO ', 'NAO ', 'NAO '], dtype=object)

[30]: #MATRIZ DE CONFUSAO
      confusao=confusion_matrix(y_teste,previsoes)
      confusao

[30]: array([[44,  5],
            [ 2,  2]], dtype=int64)

[31]: #TAXA DE ACURACIA DO MODELO PARA PREVISÃO DA EMPREGABILIDADE OU NAO DO ALUNO
      taxa_acerto=accuracy_score(y_teste, previsoes)
      taxa_acerto

[31]: 0.8679245283018868
```

Após a execução do teste e treinamento, foi obtido um modelo com 86% de taxa de acerto. Dessa forma se for simulado um perfil para o modelo prever se o indivíduo conseguira se alocar no mercado de trabalho, o retorno (sim ou não) terá uma margem de erro de 14%.

4.2 MODELO NAIVE BAYES

Foi utilizada a função `GaussianNB()` da biblioteca `sklearn` para obtenção do modelo de árvore de decisão.

```
[21]: #divisao da base de dados entre treinamento e teste (30% testar e 70% para treinar)
X_treinamento, X_teste, y_treinamento, y_teste= train_test_split(previsores,
                                                                    classe,
                                                                    test_size=0.3,
                                                                    random_state=0)

[22]: #criacao e treinamento do modelo (geracao da tabela de probabilidades)
naive_bayes= GaussianNB()
naive_bayes.fit(X_treinamento, y_treinamento)

[22]: GaussianNB()

[23]: #previsoes utilizando os registros de teste
previsoes=naive_bayes.predict(X_teste)
previsoes

[23]: array(['NAO ', 'SIM', 'NAO ', 'NAO ', 'SIM', 'SIM', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'SIM', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'SIM', 'NAO ',
          'SIM', 'NAO ', 'NAO ', 'NAO ', 'SIM', 'SIM', 'NAO ', 'SIM', 'SIM',
          'NAO ', 'NAO ', 'SIM', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO '], dtype='<U4')

[24]: #geracao da matriz de confusao e calculo de taxa de acerto e erro
confusao=confusion_matrix(y_teste, previsoes)
confusao

[24]: array([[40,  9],
          [ 2,  2]], dtype=int64)

[25]: taxa_acerto=accuracy_score(y_teste, previsoes)
taxa_acerto

[25]: 0.7924528301886793
```

Após a execução do teste e treinamento, foi obtido um modelo com 79% de taxa de acerto. Dessa forma se for simulado um perfil para o modelo prever se o indivíduo conseguira se alocar no mercado de trabalho, o retorno (sim ou não) terá uma margem de erro de 21%, consideravelmente maior que os 14% apresentados pelo modelo de árvore de decisão.

4.3 MODELO DE SELEÇÃO DE ATRIBUTOS

Para verificação dos atributos mais influentes na obtenção de um emprego da área de formação, foi aplicado a função `ExtraTreesClassifier()` do Scikit-learn.

```
[56]: #divisao da base de dados entre treinamento e teste (30% testar e 70% para treinar)
X_treinamento, X_teste, y_treinamento, y_teste = train_test_split(previsores,
                                                                    classe,
                                                                    test_size=0.3,
                                                                    random_state=1)

[57]: #criacao do modelo, treinamento, obtencao das previsoes e da taxa de acerto
svm=SVC()
svm.fit(X_treinamento, y_treinamento)

[57]: SVC()

[58]: #previsoes
previsoes=svm.predict(X_teste)
previsoes

[58]: array(['NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO ',
          'NAO ', 'NAO ', 'NAO ', 'NAO ', 'NAO '], dtype=object)

[43]: taxa_acerto=accuracy_score(y_teste, previsoes)
taxa_acerto

[43]: 0.8490566037735849

[44]: #utilizacao do algoritmo ExtraTreesClassifier para extrair as caracteristicas mais
forest= ExtraTreesClassifier()
forest.fit(X_treinamento, y_treinamento)
importancias=forest.feature_importances_
importancias

[44]: array([0.1256201 , 0.02138099, 0.09636803, 0.14772663, 0.03316694,
          0.         , 0.02405737, 0.03230122, 0.15745368, 0.04260966,
          0.06513453, 0.04520455, 0.2089763 1])
```

VALOR	ATRIBUTO
0.20543955230999741	Possui experiencia na area de formacao do curso?
0.16490198584964472	Qual regio da cidade voce mora?
0.13957702739893801	Qual o seu curso?
0.11040428149692752	Qual sua idade?
0.09478739710541238	Qual sua renda familiar?
0.06725865760208737	Qual sua disponibilidade de horario para atividade remunerada?
0.05050238815355612	Qual sua formacao regular?
0.047748902352610895	Qual sua carga horaria disponivel?
0.03571929103901822	Qual a sua escola?
0.02899321879682359	Qual tipo do seu curso?
0.028087376068212103	Qual o eixo tecnologico?
0.026579921826771553	Qual seu genero?
0.0	Qual o percentual da sua formacao no CETAM ja foi concluida?

Após a aplicação do algoritmo foi possível verificar 4 principais atributos, com uma margem de erro de 84,99%.

Com o valor de 0.20 o atributo ‘possui experiencia na área de formação do curso?’ se mostrou o mais influente na possibilidade de se conseguir um emprego na área de formação. Seguido do atributo ‘Qual região da cidade você mora’ com 0.16, ‘Qual o seu curso’ com 0.14 e ‘Qual sua idade?’ com o valor de 0.11.

Dessa forma a pesquisa corrobora o conhecimento popular da importância da experiência prévia para obter uma ocupação na área de formação. A demanda de vagas do mercado pelo curso apareceu em 3º lugar, e a idade, como exposto na seção 2, onde a classe acima de 30 anos, embora seja a maior dos entrevistados, constitui a menor parcela dos que estão trabalhando na área.

5 CONCLUSÕES

Foi possível verificar após a análise exploratória o difícil cenário da empregabilidade no momento da realização do questionário. Apenas 32,6% dos alunos reportaram ter alguma experiência na área de formação do curso e 11% estão exercendo algum tipo de atividade remunerada na área do curso.

Na relação de variáveis, não foi possível verificar influência do gênero de forma significativa (0.026) na seleção de atributos. Já quanto a idade foi possível verificar um influencia desta variável, uma vez que a moda (acima de 30 anos) consiste da menor parcela dos que estão trabalhando (10%), sendo corroborado pela seleção de atributos (4º posição). Ainda sobre a categoria seleção de atributos foi possível corroborar a importância da experiência prévia para a obtenção de um emprego.

É possível observar a necessidade de a instituição em questão analisar melhor a abertura de cursos, focando os recursos em cursos que possuem demanda no mercado de trabalho, evitando assim situações como as exibidas na tabela da seção 3 (%Está exercendo atividade remunerada? Vs %Qual o seu curso?), onde existem cursos que possuem um percentual baixíssimo proporcional contribuindo de forma negativa para os indicadores de empregabilidade.