Bahria University
Karachi Campus

COURSE: SEL-452 DATA MINING
TERM: SPRING 2020, CLASS: BSE- 6 (A)

Submitted By:

Uzair Mehmood	51739
(Name)	**(Reg. No.)**

Hammad Shaikh	51717
(Name)	**(Reg. No.)**

Noman Ali Siddique	51777
(Name)	**(Reg. No.)**

Submitted To:

Dr. Salahudin Unar / Engr. Sobia Iftikhar

Signed   Remarks: Score:

BAHRIA UNIVERSITY, Karachi Campus)

Department of Software Engineering

Project – Spring 2020

Course Title:	**Data Mining**	**Course Code:**	CSC-452
Class:	**BSE – 6[A]**	**Shift: Morning**	
Course Instructor:	**Dr. Salahuddin Unar**	**Date: 28th May 2020**	
Lab Instructor:	**ENGR. Sobia Iftikhar**	**Max. Marks:**	**30 Marks**
Due Date:	**12th June 2020**		

Group members: Uzair Mehmood || Noman Ali Siddique|| Hammad Shaikh

---

Instructions:
1. **This is a data mining-based project.**
2. **Deadline for the submission of project on is 12th June 2020.**
3. **For this project you can use either anaconda or KNIME.**
4. **If you submit your project after the given deadline then 2 Marks will be deducted for the late submissions.**

Project criteria.   [CLO#05, 30.0 marks]

Your task for this Project is to work on the particular problem statement, identify and characterize a data set. It would be best if you have some domain experience, as this will help with data preparation. Before the implementation you must know basic requirement needed, as I mentioned below:

1. **Find a dataset with two option   [3.0 marks]**
   a. **Take a dataset of your interests the UCI Machine Learning Repository or Kaggle Repository.**
   b. **Conduct a survey on real time problem and convert into dataset file.**

I selected the second option in my project. I have used survey on real time problem and then converted into dataset file (.csv format). I then shared the google form link with friends and relatives to get their views.

You can view the survey from this link ([Form](#))

Some details of my dataset:

```python
import pandas as pd
import numpy as np
import matplotlib as plt
%matplotlib inline

df = pd.read_csv("C:\Users\iamuzairmehmood\Desktop\Depression.csv")
df.describe()
```

	Timestamp	Enter Your age	Choose Your Gender	Where do you study?	As a student, Do you think your life is stressed?	As a student, What stresses you and your life the most?	Does your stress influence your life? How?	Do you think stress is important?	What are your stress busters?	How often are you stressed?	Do you think stress management is important?
count	116	116	116	116	116	116	116	116	116	116	116
unique	115	4	2	3	2	6	5	2	6	3	2
top	2020/05/11 1:48:48 am GMT+5	19-21	Male	University	Yes	Future Plans	Yes, I suffer from weakness and headache	No	Spending time with friends and family	Sometimes	Yes, because it helps us to cope up from stress
freq	2	60	63	74	83	32	30	73	35	57	78

```python
import pandas as pd
import numpy as np
import matplotlib as plt
%matplotlib inline

df = pd.read_csv("C:\Users\iamuzairmehmood\Desktop\Depression.csv")
df.head(5)
```

	Timestamp	Enter Your age	Choose Your Gender	Where do you study?	As a student, Do you think your life is stressed?	As a student, What stresses you and your life the most?	Does your stress influence your life? How?	Do you think stress is important?	What are your stress busters?	How often are you stressed?	Do you think stress management is important?
0	2020/05/11 1:33:51 am GMT+5	22-25	Female	University	Yes	Academic Pressure	Yes, I can't sleep properly	No	Spending time with friends and family	Sometimes	Yes, because it helps us to cope up from stress
1	2020/05/11 1:34:25 am GMT+5	22-25	Female	University	Yes	Academic Pressure	Yes, I can't sleep properly	No	Spending time with friends and family	Rarely	No, because student's life is stress bearing a...
2	2020/05/11 1:34:27 am GMT+5	19-21	Male	University	Yes	Academic Pressure	Yes, I can't study properly	No	Spending time with friends and family	Rarely	Yes, because it helps us to cope up from stress
3	2020/05/11 1:35:06 am GMT+5	19-21	Male	University	Yes	Academic Pressure	Yes, I can't study properly	No	Spending time with friends and family	All the time	No, because student's life is stress bearing a...
4	2020/05/11 1:37:30 am GMT+5	15-18	Female	University	Yes	Future Plans	Yes, I suffer from weakness and headache	No	Going for outing	All the time	Yes, because it helps us to cope up from stress

2. **Describe your problem statement in detail.** [3.0 marks]

Our youth is struggling with depression and anxiety now a days. It's a very common problem among students like us now a days. My idea of conducting this survey was to determine the effect of stress and anxiety over student's life and how students are battling against it. Student's problems are my primary focus here. Secondly for them is management and rehabilitation help necessary to fight against depression or anxiety.

Students are known as the backbone of our country. We students are responsible for the future development of Pakistan. But what if we if we are not developed properly? What if we are struggling against ourselves?

Depression is the main problem here and by looking student's views we might help them eliminate depression and unusual burdens.

3. **What type of benefit you might hope to get from this implementation.** [3.0 marks]

The benefit we will get after analyzing thoroughly through algorithms and charts will be huge. What is causing those students problems and how can we help the students. How we can help students fix things side by side. We may help the youth from sacrificing themselves because of suicidal thoughts and bad habits. We might give them proper guide. We might help them focus on what's important. We might advice students to deal with those with problems with possible solutions. Moreover by understanding such data we could help a student live a better life.

4. **Identify the category of dataset either it is from classification based problem or regression based.** [3.0 marks]

Its not a dataset which contain numeric or continuous values and its not a prediction state in which we have to predict y from x on by using linear regression on graphical method.

Its a pure " classification based problem " in which we have different classes of data. We will accurately predict the target class for each case in the data by using decision trees.

5. After categories, Perform data analysis in order to check missing values.[3.0 marks]

```python
import pandas as pd
import numpy as np
import matplotlib as plt
%matolotlib inline

df = pd.read_csv("C:\Users\iamuzairmehmood\Desktop\Depression.csv")
df.apply(lambda x: sum(x.isnull()),axis=0)
```

```
Timestamp                                                   0
Enter Your age                                              0
Choose Your Gender                                          0
Where do you study?                                         0
As a student, Do you think your life is stressed?           0
As a student, What stresses you and your life the most?     0
Does your stress influence your life? How?                  0
Do you think stress is important?                           0
What are your stress busters?                               0
How often are you stressed?                                 0
Do you think stress management is important?                0
dtype: int64
```

6. What type of data mining (algorithm) you think would be relevant, Mention the name of algorithm.   [3.0 marks]

I will use "Decision Tree" algorithm on my dataset to get the graphs and accuracy.
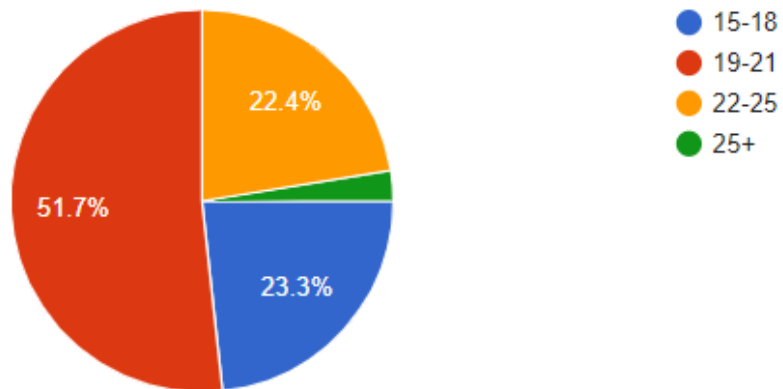
7. Apply data preprocessing steps on your dataset and implement data mining technique relevant to your dataset.   [3.0 marks]

STEPS:

1) Checking depression rate by age:
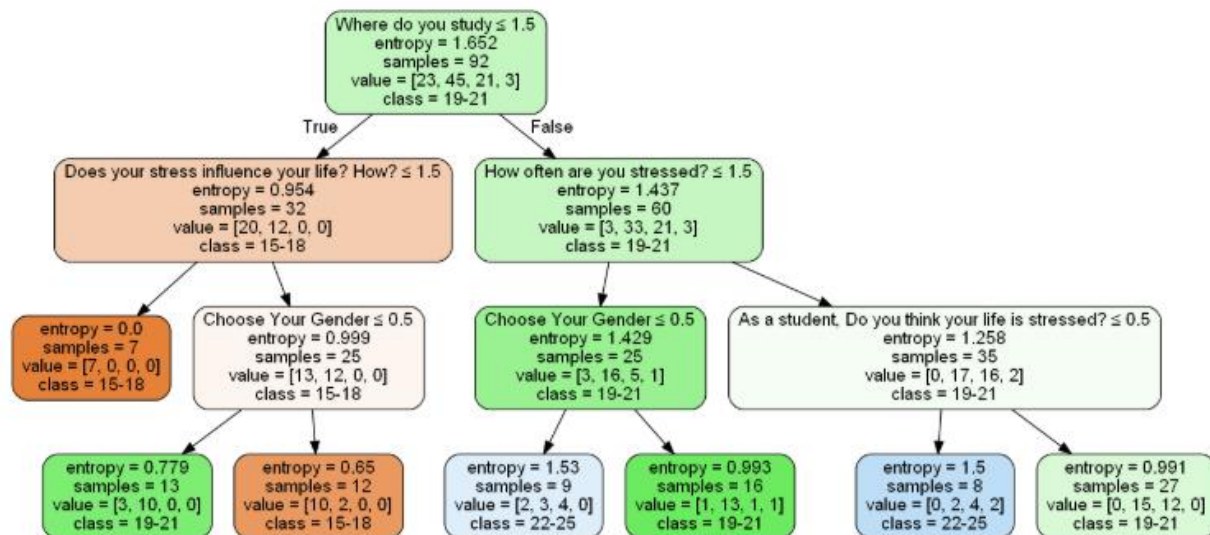
Enter Your age

116 responses

```python
import os
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Result','Choose Your Gender','Where do you study',
            'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?'
            ,'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
            'How often are you stressed?','Do you think stress management is important?']

pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Choose Your Gender','Where do you study',
                'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?',
                'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
                'How often are you stressed?','Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['15-18','19-21','22-25','25+'])
graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```
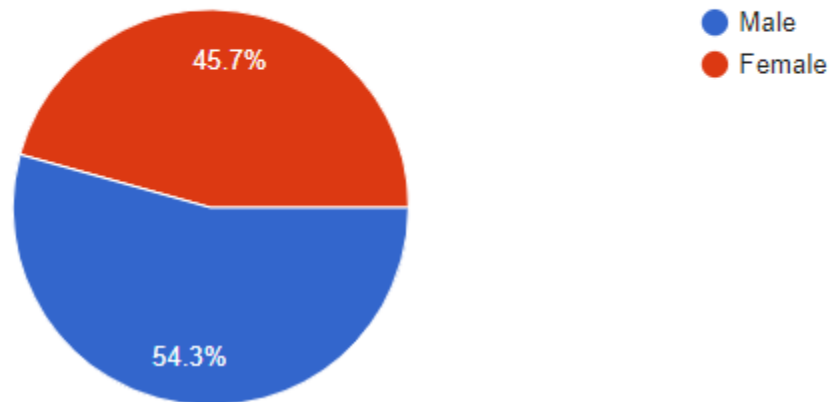
Accuracy: 58.3333333333

2) Checking depression rate by gender:

Choose Your Gender

116 responses
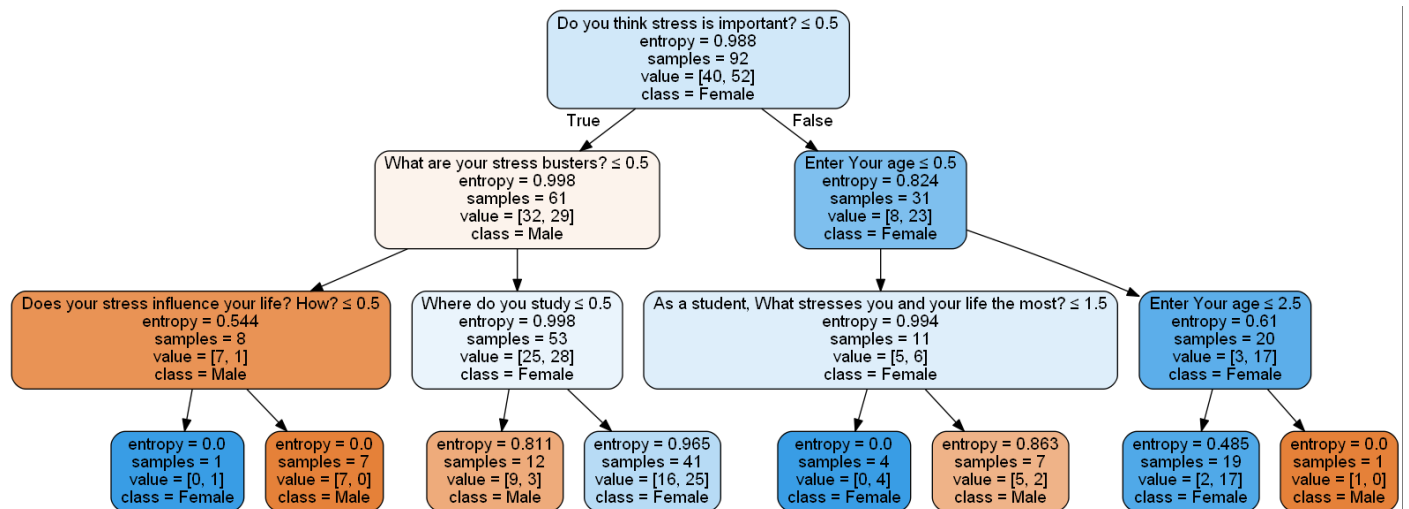

- ● Male
- ● Female

45.7%

54.3%

```
import os
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your age','Result','Where do you study',
            'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?'
            ,'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
            'How often are you stressed?','Do you think stress management is important?']

# Load dataset
pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Enter Your age','Where do you study',
                'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?',
                'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
                'How often are you stressed?','Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['Male','Female'])
graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```
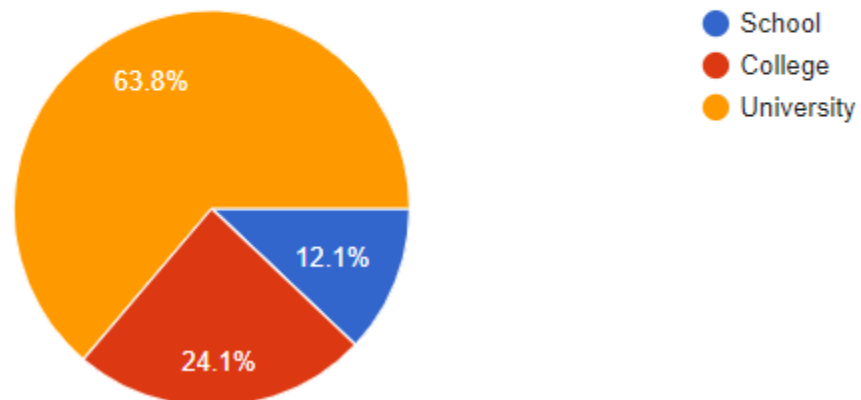
Accuracy: 66.6666666667

3) Checking depression by collecting data from various educational institutions:

Where do you study?

116 responses

- School
- College
- University

63.8%

12.1%

24.1%

```python
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your age','Result', 'Choose Your Gender',
            'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?'
            ,'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
            'How often are you stressed?','Do you think stress management is important?']

# Load dataset
pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Enter Your age', 'Choose Your Gender',
                'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?'
                'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters
                'How often are you stressed?','Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=4)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['College','University','School'])
graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```

Accuracy: 75.0

Do you think stress is i
entropy = 0
samples =
value = [40,
class = Univ

True

What are your stress busters? ≤ 0.5
entropy = 0.998
samples = 61
value = [32, 29]
class = College

Enter Your age ≤ 1.5
entropy = 0.544
samples = 8
value = [7, 1]
class = College

Choose Your Gender ≤ 0.5
entropy = 0.998
samples = 53
value = [25, 28]
class = University

entropy = 0.0
samples = 7
value = [7, 0]
class = College

entropy = 0.0
samples = 1
value = [0, 1]
class = University

As a student, What stresses you and your life the most? ≤ 0.5
entropy = 0.811
samples = 12
value = [9, 3]
class = College

As a student, What stresses you and you
entropy = 0.965
samples = 41
value = [16, 25]
class = University

entropy = 0.0
samples = 1
value = [0, 1]
class = University

entropy = 0.684
samples = 11
value = [9, 2]
class = College

entropy = 0.94
samples = 14
value = [9, 5]
class = College

entropy
sampl
value
class =

important? ≤ 0.5
0.988
= 92
0, 52]
versity

False

Enter Your age ≤ 0.5
entropy = 0.824
samples = 31
value = [8, 23]
class = University

As a student, What stresses you and your life the most? ≤ 1.5
entropy = 0.994
samples = 11
value = [5, 6]
class = University

Enter Your age ≤ 2.5
entropy = 0.61
samples = 20
value = [3, 17]
class = University

ur life the most? ≤ 1.5

entropy = 0.0
samples = 4
value = [0, 4]
class = University

What are your stress busters? ≤ 4.5
entropy = 0.863
samples = 7
value = [5, 2]
class = College

Does your stress influence your life? How? ≤ 3.5
entropy = 0.485
samples = 19
value = [2, 17]
class = University

entropy = 0.0
samples = 1
value = [1, 0]
class = College

py = 0.826
les = 27
= [7, 20]
University

entropy = 0.65
samples = 6
value = [5, 1]
class = College

entropy = 0.0
samples = 1
value = [0, 1]
class = University

entropy = 0.0
samples = 13
value = [0, 13]
class = University

entropy = 0.918
samples = 6
value = [2, 4]
class = University

4) What students think about stress:

As a student, Do you think your life is stressed?

116 responses
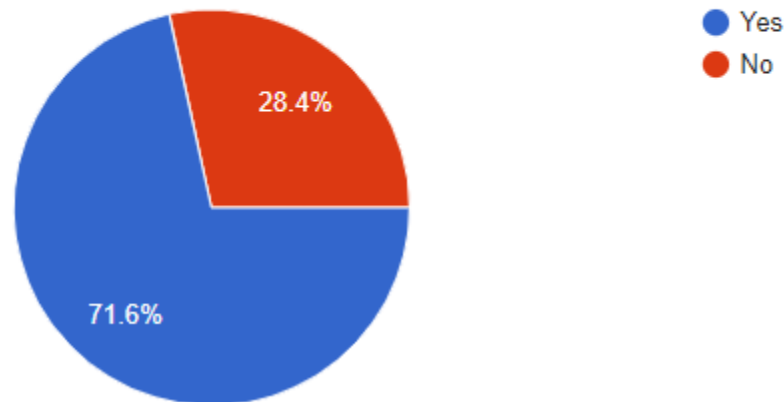
- Yes
- No

28.4%

71.6%

```python
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your age', 'Choose Your Gender','Where do you study','Result',
           'As a student, What stresses you and your life the most?'
           ,'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
           'How often are you stressed?','Do you think stress management is important?']

# Load dataset
pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Enter Your age', 'Choose Your Gender','Where do you study',
               'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
               'How often are you stressed?','Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['Yes','No'])
graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```
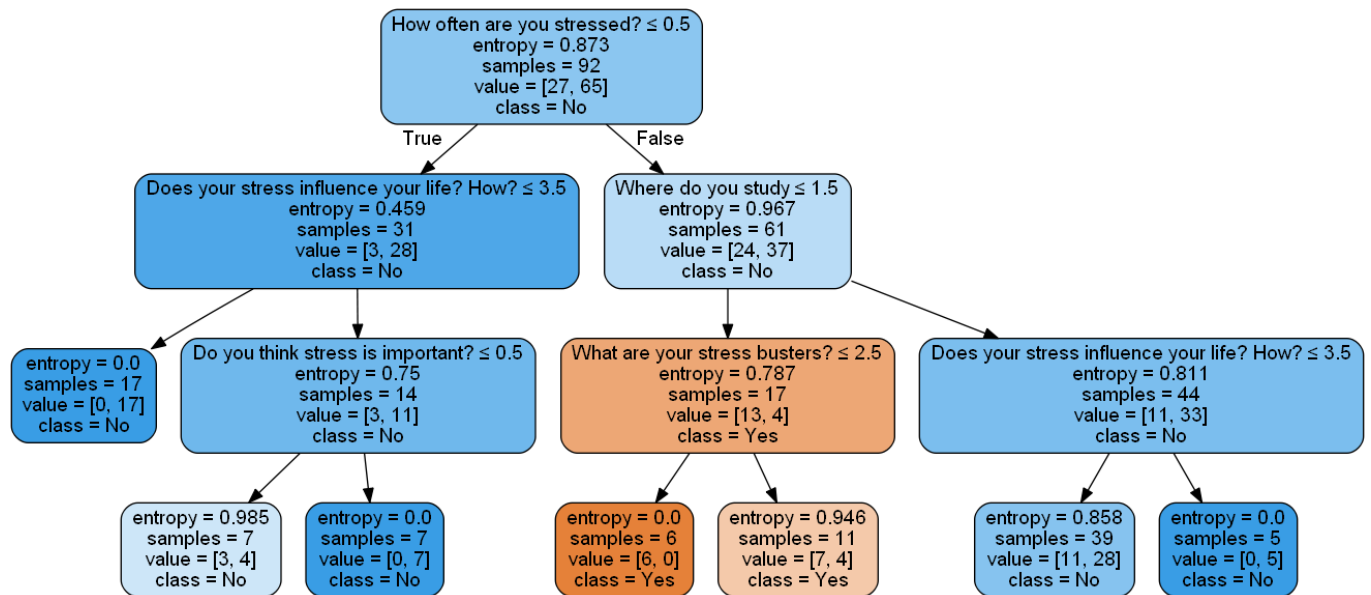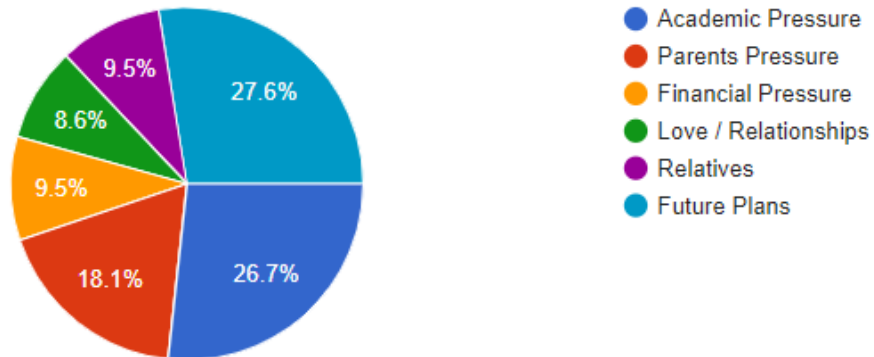
Accuracy: 70.8333333333

5) What stresses students lives:

As a student, What stresses you and your life the most?

116 responses

- Academic Pressure
- Parents Pressure
- Financial Pressure
- Love / Relationships
- Relatives
- Future Plans

27.6%
26.7%
18.1%
9.5%
8.6%
9.5%

```python
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your age', 'Choose Your Gender','Where do you study', 'As a student, Do you think your life is str
            ,'Result' ,'Does your stress influence your life? How?',
             'Do you think stress is important?','What are your stress busters?',
             'How often are you stressed?','Do you think stress management is important?']


pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Enter Your age', 'Choose Your Gender','Where do you study', 'As a student, Do you think your life is stressed?',
                'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
                'How often are you stressed?','Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['Academic Pressure','Future Plans','Parents Pressure','Financial Pressure','Love/Relationships','Relati
])
graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```
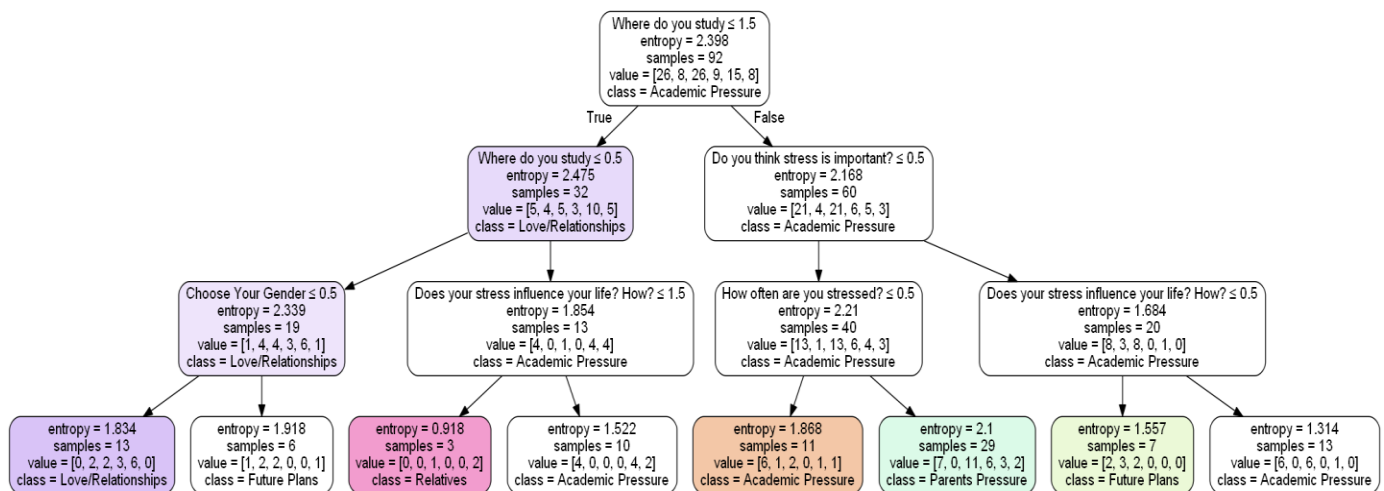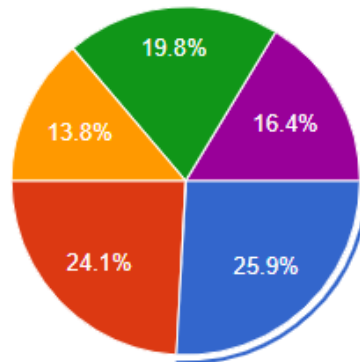
Accuracy: 29.1666666667

6) How stress disturbs students lives:

Does your stress influence your life? How?

116 responses

- Yes, I suffer from weakness and headache
- Yes, I can't sleep properly
- Yes, I can't focus on my diet properly
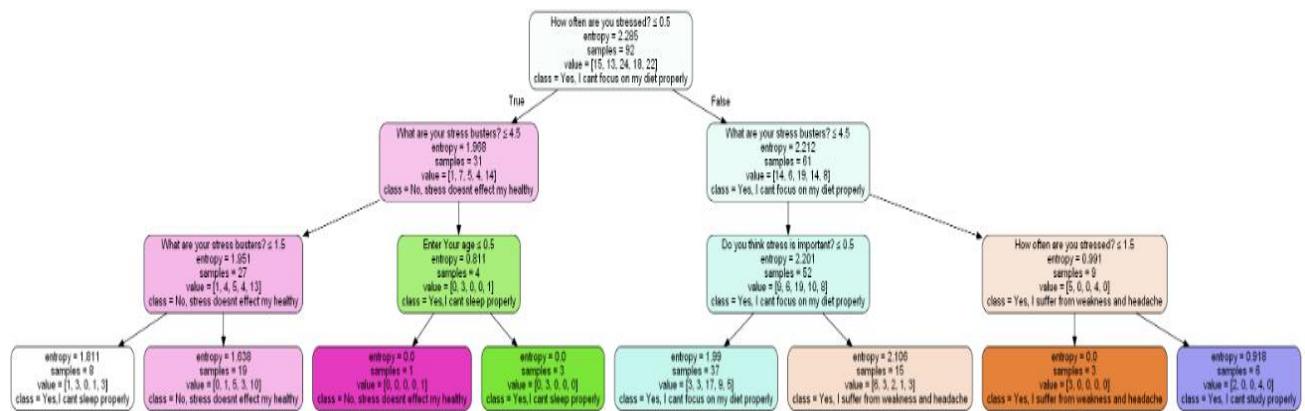- Yes, I can't study properly
- No, stress doesn't effect my health

```python
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your age', 'Choose Your Gender','Where do you study', 'As a student, Do you think your life is str
            'As a student, What stresses you and your life the most?','Result',
            'Do you think stress is important?','What are your stress busters?','How often are you stressed?','Do you think stres

# Load dataset
pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Enter Your age', 'Choose Your Gender','Where do you study', 'As a student, Do you think your life is stressed?',
                'Do you think stress is important?', 'As a student, What stresses you and your life the most?','What are your stre
                'How often are you stressed?','Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)#80% TRAINING AND 20% TESTING
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['Yes, I suffer from weakness and headache','Yes,I cant sleep properly',
                'Yes, I cant focus on my diet properly','Yes, I cant study properly',
'No, stress doesnt effect my healthy'])

graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```

```
Accuracy: 37.5
```

7) Is stress necessary from the point of view of students:

Do you think stress is important?

116 responses

- Yes — 37.1%
- No — 62.9%

```python
import os
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your Age','Choose Your Gender','Where do you study',
            'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?'
            ,'Does your stress influence your life? How?','Result','What are your stress busters?',
            'How often are you stressed?','Do you think stress management is important?']

# Load dataset
pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Choose Your Gender','Where do you study',
                'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?',
                'Does your stress influence your life? How?','What are your stress busters?',
                'How often are you stressed?','Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['Yes','No'])
graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```

Accuracy: 54.1666666667
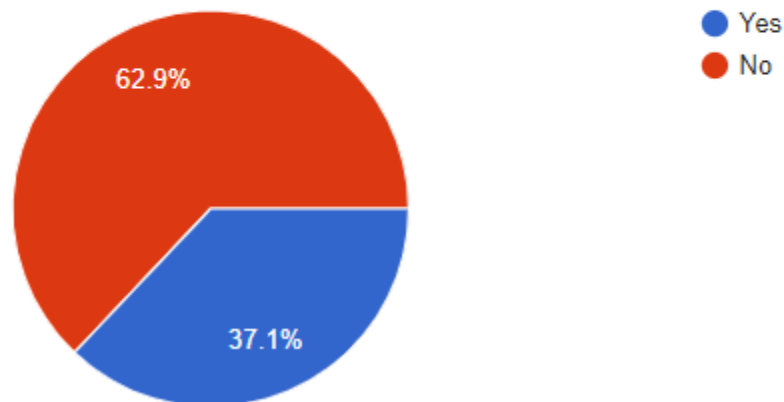
8) How students get rid of the stress:

What are your stress busters?

116 responses

- Spending time with friends and family
- Spending time alone
- Practicing your hobby
- Going for outing
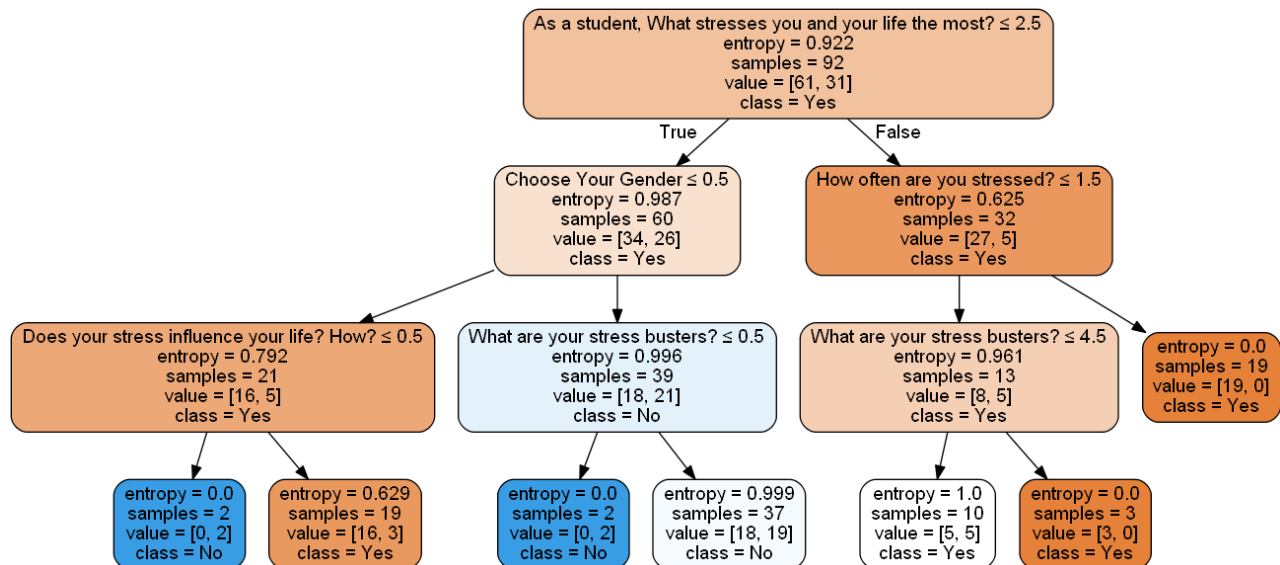- Gym and exercise
- Watching movies

```python
import os
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your Age','Choose Your Gender','Where do you study',
            'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?'
            ,'Does your stress influence your life? How?','Do you think stress is important?','Result',
            'How often are you stressed?','Do you think stress management is important?']

# Load dataset
pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Choose Your Gender','Where do you study',
                'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?',
                'Does your stress influence your life? How?', 'Do you think stress is important?',
                'How often are you stressed?','Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['Spending time with friends and family',
'Spending time alone',
'Practicing your hobby',
'Going for outing',
'Gym and exercise',
'Watching movies'])
graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```

Accuracy: 45.8333333333
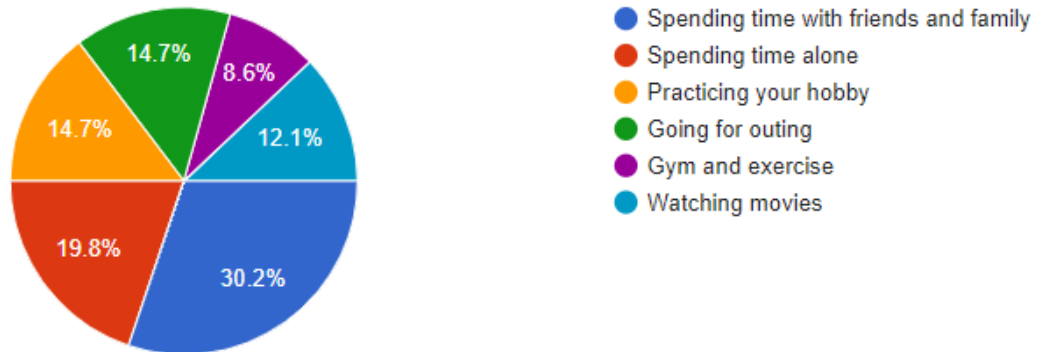
9) How much stress does we students take:

How often are you stressed?

116 responses

Legend:
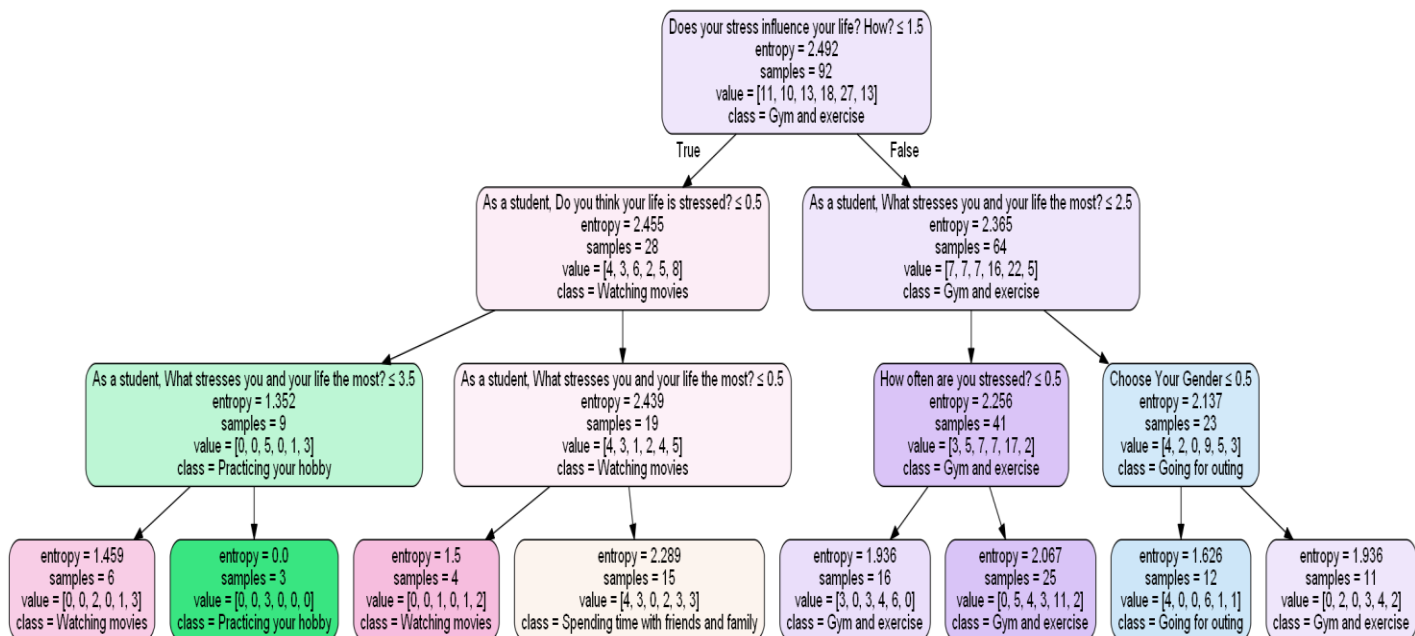- All the time
- Sometimes
- Rarely

49.1%
15.5%
35.3%

```python
import os
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your Age','Choose Your Gender','Where do you study',
            'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?'
           ,'Does your stress influence your life? How?','Do you think stress is important?', 'What are your stress busters?',
            'Result','Do you think stress management is important?']

# Load dataset
pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Choose Your Gender','Where do you study',
                'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?',
                'Does your stress influence your life? How?','Do you think stress is important?', 'What are your stress busters?',
                'Do you think stress management is important?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['All the time',
'Sometimes',
'Rarely'])

graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```

Accuracy: 50.0
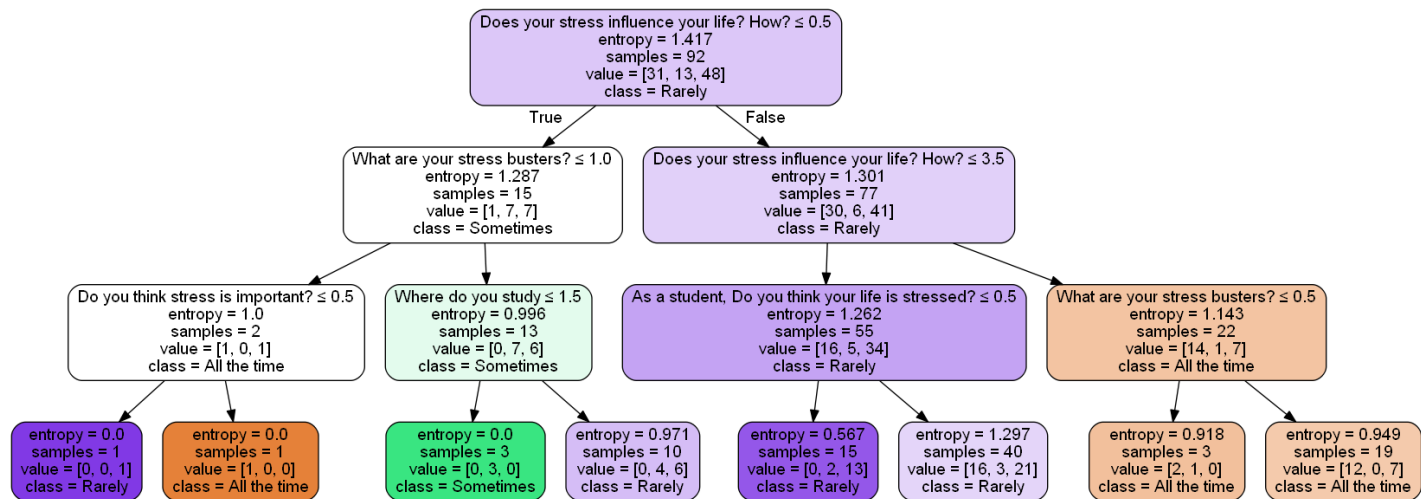
FINAL RESULT

Do we need maintenance and rehabilitation efforts to remove stress from our life?

Do you think stress management is important?

116 responses
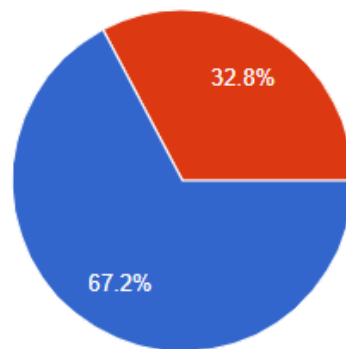
- Yes, because it helps us to cope up from stress
- No, because student's life is stress bearing age anyway

32.8%

67.2%

```
import os
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
from sklearn import preprocessing
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sn
col_names =['Timestamp','Enter Your Age','Choose Your Gender','Where do you study',
            'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?'
            ,'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
            'How often are you stressed?','Result']

# Load dataset
pima = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

pima.head()
le = preprocessing.LabelEncoder()
pima = pima.apply(le.fit_transform)
feature_cols = ['Enter Your Age','Choose Your Gender','Where do you study',
                'As a student, Do you think your life is stressed?','As a student, What stresses you and your life the most?',
                'Does your stress influence your life? How?','Do you think stress is important?','What are your stress busters?',
                'How often are you stressed?']
X = pima[feature_cols]
y = pima.Result
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
clf = DecisionTreeClassifier(criterion="entropy",max_depth=10)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test,y_pred)*100)
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,filled=True, rounded=True,special_characters=True,feature_names =
feature_cols,class_names=['Yes','No'])
graph=pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('depression.png')
Image(graph.create_png())
```

If max depth = 3:

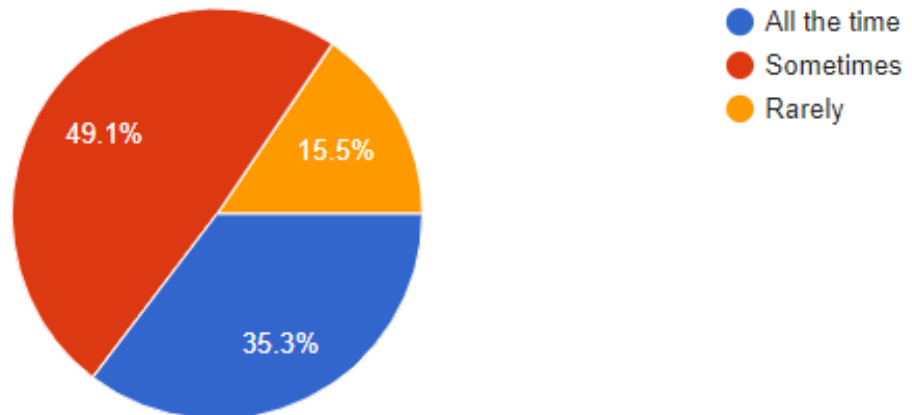

If max depth = 10:

How often are you stressed? ≤ 0.5
entropy = 0.932
samples = 92
value = [32, 60]
class = No

True | False

What are your stress busters? ≤ 4.5
entropy = 0.709
samples = 31
value = [6, 25]
class = No

Where do
entro
san
value
cl

entropy = 0.0
samples = 4
value = [0, 4]
class = No

How often are
entro
san
value
cl

What are your stress busters? ≤ 3.5
entropy = 0.764
samples = 27
value = [6, 21]
class = No

entropy = 0.0
samples = 4
value = [4, 0]
class = Yes

Enter Your Age ≤ 0.5
entropy = 0.544
samples = 16
value = [2, 14]
class = No

As a student, What stresses you and your life the most? ≤ 3.0
entropy = 0.946
samples = 11
value = [4, 7]
class = No

Does your stress influence your li
entropy = 0.811
samples = 12
value = [9, 3]
class = Yes

Does your stress influence your life? How? ≤ 2.5
entropy = 0.863
samples = 7
value = [2, 5]
class = No

entropy = 0.0
samples = 9
value = [0, 9]
class = No

Do you think stress is important? ≤ 0.5
entropy = 0.65
samples = 6
value = [1, 5]
class = No

Where do you study ≤ 0.5
entropy = 0.971
samples = 5
value = [3, 2]
class = Yes

entropy = 0.0
samples = 1
value = [0, 1]
class = No

As a student, W

Do you think stress is important? ≤ 0.5
entropy = 0.918
samples = 3
value = [2, 1]
class = Yes

entropy = 0.0
samples = 4
value = [0, 4]
class = No

As a student, Do you think your life is stressed? ≤ 0.5
entropy = 0.918
samples = 3
value = [1, 2]
class = No

entropy = 0.0
samples = 3
value = [0, 3]
class = No

entropy = 0.0
samples = 2
value = [2, 0]
class = Yes

Where do you study ≤ 1.5
entropy = 0.918
samples = 3
value = [1, 2]
class = No

entropy = 0.0
samples = 1
value = [0, 1]
class = No

entropy = 0.0
samples = 2
value = [2, 0]
class = Yes

entropy = 0.0
samples = 1
value = [0, 1]
class = No

entropy = 0.0
samples = 1
value = [0, 1]
class = No

Enter Your Age ≤ 1.5
entropy = 1.0
samples = 2
value = [1, 1]
class = Yes

entropy = 0.0
samples = 2
value = [0, 2]
class = No

entropy = 0.0
samples = 1
value = [1, 0]
class = Yes

entropy = 0.0
samples = 3
value = [3, 0]
class = Yes

entropy = 0.0
samples = 1
value = [1, 0]
class = Yes

entropy = 0.0
samples = 1
value = [0, 1]
class = No

What are

Where do y
entro
san
value
clas

entropy = 0.0
samples = 1
value = [0, 1]
class = No

Decision tree nodes:

- False
- Where do you study ≤ 1.5 / entropy = 0.984 / samples = 61 / value = [26, 35] / class = No
- How often are you stressed? ≤ 1.5 / entropy = 0.787 / samples = 17 / value = [13, 4] / class = Yes
- Does your stress influence your life? How? ≤ 2.5 / entropy = 0.876 / samples = 44 / value = [13, 31] / class = No
- entropy = 0.0 / samples = 4 / value = [4, 0] / class = Yes
- Do you think stress is important? ≤ 0.5 / entropy = 0.89 / samples = 13 / value = [9, 4] / class = Yes
- What are your stress busters? ≤ 0.5 / entropy = 0.967 / samples = 28 / value = [11, 17] / class = No
- What are your stress busters? ≤ 2.5 / entropy = 0.544 / samples = 16 / value = [2, 14] / class = No
- influence your life? How? ≤ 1.5 / entropy = 0.811 / samples = 12 / value = [9, 3] / class = Yes
- entropy = 0.0 / samples = 1 / value = [0, 1] / class = No
- entropy = 0.0 / samples = 2 / value = [0, 2] / class = No
- As a student, What stresses you and your life the most? ≤ 2.5 / entropy = 0.983 / samples = 26 / value = [11, 15] / class = No
- As a student, What stresses you and your life the most? ≤ 1.0 / entropy = 0.971 / samples = 5 / value = [2, 3] / class = No
- entropy = 0.0 / samples = 11 / value = [0, 11] / class = No
- As a student, What stresses you and your life the most? ≤ 2.0 / entropy = 0.684 / samples = 11 / value = [9, 2] / class = Yes
- As a student, Do you think your life is stressed? ≤ 0.5 / entropy = 0.918 / samples = 21 / value = [7, 14] / class = No
- Does your stress influence your life? How? ≤ 0.5 / entropy = 0.722 / samples = 5 / value = [4, 1] / class = Yes
- entropy = 0.0 / samples = 2 / value = [0, 2] / class = No
- What are your stress busters? ≤ 0.5 / entropy = 0.918 / samples = 3 / value = [2, 1] / class = Yes
- y = 0.0 / les = 3 / = [3, 0] / = Yes
- Does your stress influence your life? How? ≤ 3.5 / entropy = 0.811 / samples = 8 / value = [6, 2] / class = Yes
- entropy = 0.0 / samples = 4 / value = [0, 4] / class = No
- Enter Your Age ≤ 2.5 / entropy = 0.977 / samples = 17 / value = [7, 10] / class = No
- entropy = 0.0 / samples = 1 / value = [0, 1] / class = No
- entropy = 0.0 / samples = 4 / value = [4, 0] / class = Yes
- entropy = 0.0 / samples = 1 / value = [1, 0] / class = Yes
- What are your stress busters? ≤ 1.5 / entropy = 1.0 / samples = 2 / value = [1, 1] / class = Yes
- What are your stress busters? ≤ 1.5 / entropy = 0.592 / samples = 7 / value = [6, 1] / class = Yes
- entropy = 0.0 / samples = 1 / value = [0, 1] / class = No
- What are your stress busters? ≤ 4.5 / entropy = 0.954 / samples = 16 / value = [6, 10] / class = No
- entropy = 0.0 / samples = 1 / value = [1, 0] / class = Yes
- entropy = 0.0 / samples = 1 / value = [0, 1] / class = No
- entropy = 0.0 / samples = 1 / value = [1, 0] / class = Yes
- Where do you study ≤ 0.5 / entropy = 1.0 / samples = 2 / value = [1, 1] / class = Yes
- entropy = 0.0 / samples = 5 / value = [5, 0] / class = Yes
- Enter Your Age ≤ 0.5 / entropy = 0.918 / samples = 15 / value = [5, 10] / class = No
- entropy = 0.0 / samples = 1 / value = [1, 0] / class = Yes
- entropy = 0.0 / samples = 1 / value = [0, 1] / class = No
- entropy = 0.0 / samples = 1 / value = [1, 0] / class = Yes
- entropy = 0.0 / samples = 1 / value = [0, 1] / class = No
- Does your stress influence your life? How? ≤ 0.5 / entropy = 0.94 / samples = 14 / value = [5, 9] / class = No
- entropy = 0.918 / samples = 3 / value = [2, 1] / class = Yes
- entropy = 0.845 / samples = 11 / value = [3, 8] / class = No

By looking at the above results,

This means rehabilitation discussions, discussion meetings about student's anxiety and student support centers must be developed in our country to solve different problems that our student face in their daily life.

This will not help the student in their studies but it will also help them become a more positive person who could basically tackle any kind of hurdle in his/her life.

8. **Mention splitting criteria for training and testing purpose.** [3.0 marks]

The splitting criteria for training will be 80% and testing will be 20%

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1) # 80% training and 20% test
```

9. **Create confusion matrix and print the accuracy of your output.** [3.0 marks]

```python
# Import required Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sklearn
from sklearn.neural_network import MLPClassifier
from sklearn.neural_network import MLPRegressor
from sklearn import preprocessing
# Import necessary modules
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from math import sqrt
from sklearn.metrics import r2_score
df = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv", header=0, names=col_names)

le = preprocessing.LabelEncoder()
df= df.apply(le.fit_transform)
target_column = ['Do you think stress management is important?']
predictors = list(set(list(df.columns))-set(target_column))
df[predictors] = df[predictors]/df[predictors].max()
df.describe().transpose()
X = df[predictors].values
y = df[target_column].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=40)
# print(X_train.shape); print(X_test.shape)

from sklearn.neural_network import MLPClassifier

mlp = MLPClassifier(hidden_layer_sizes=(8,8,8), activation='relu', solver='adam', max_iter=500)
mlp.fit(X_train,y_train)

predict_train = mlp.predict(X_train)
predict_test = mlp.predict(X_test)


from sklearn.metrics import classification_report,confusion_matrix

import seaborn as sn

print(confusion_matrix(y_test,predict_test))
print(classification_report(y_test,predict_test))
print("Accuracy:",metrics.accuracy_score(y_test,predict_test)*100)
```

```
[[ 3  5]
 [ 0 16]]
           precision    recall  f1-score   support

        0       1.00      0.38      0.55         8
        1       0.76      1.00      0.86        16

avg / total       0.84      0.79      0.76        24

Accuracy: 79.1666666667
```

10. Illustrate your outcome to represent the accuracy of your output.   [3.0 marks]
      1.   **Use two different type of graph.**

Decision tree accuracy**:**

```
Accuracy: 70.8333333333
```

Multinomial NB accuracy:

```
Accuracy: 0.657142857143
```

      1.   **Use two different type of graph.**

```
In [2]:  import pandas as pd
         import numpy as np
         import matplotlib as plt
         %matplotlib inline
         df = pd.read_csv("C:\Users\iamuzairmehmood\Desktop\Depression.csv")
         df['Do you think stress management is important?'].hist(bins=100)

Out[2]:  <matplotlib.axes._subplots.AxesSubplot at 0x2abf41c07c8>
```



```
In [7]:  import pandas as pd
         import numpy as np
         import matplotlib as plt
         %matplotlib inline
         df = pd.read_csv("C:/Users/iamuzairmehmood/Desktop/Depression.csv")
         df.boxplot(column='Do you think stress is important?',by='Do you think stress management is important?')

Out[7]:  <matplotlib.axes._subplots.AxesSubplot at 0x2abf41c07c8>
```


THE END