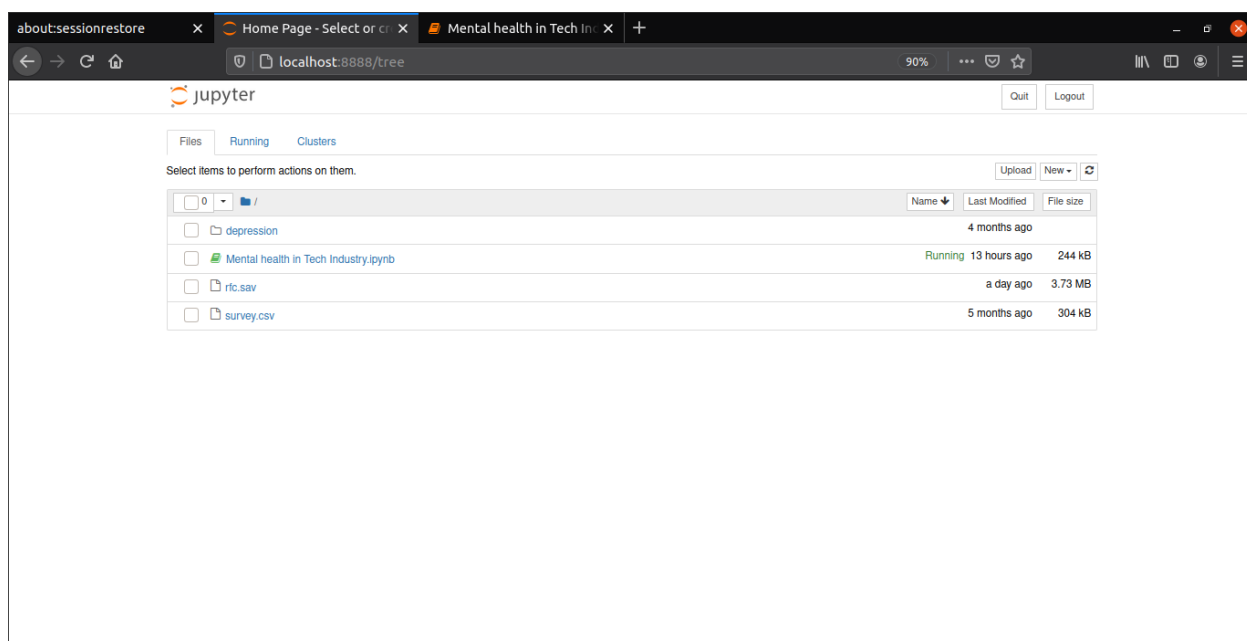


Name: Colin Mburugu

Batch: LISUMO1

Submission_date : 4/07/2021

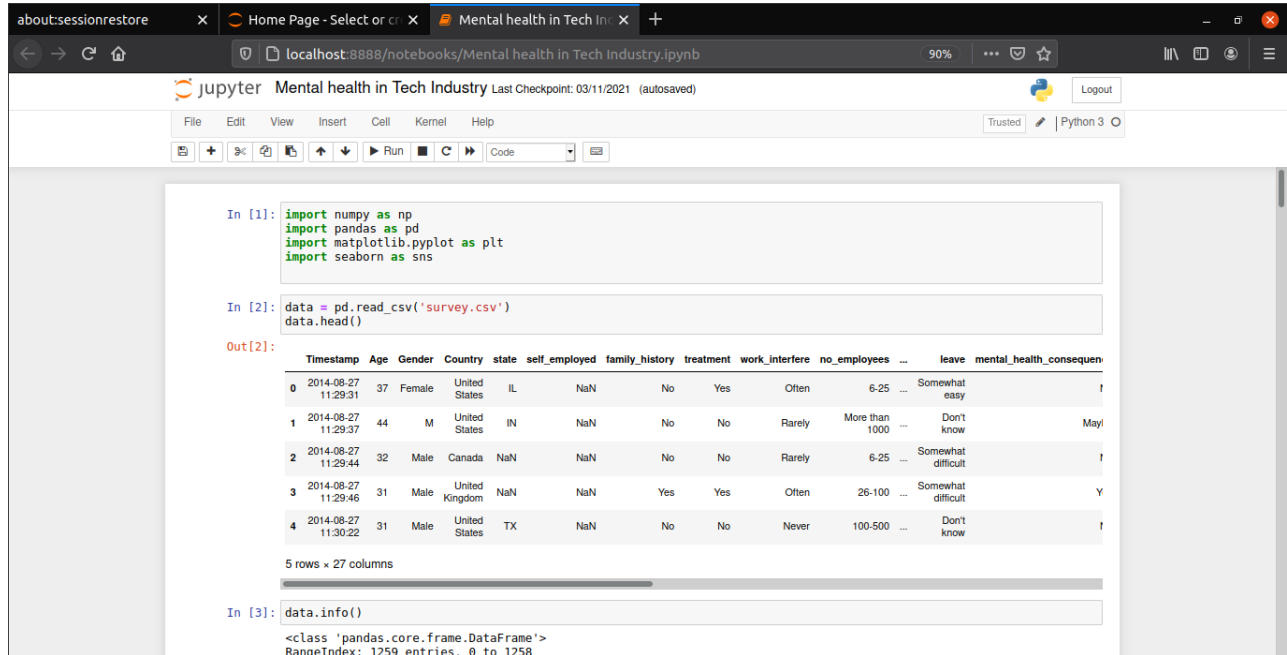
Submission_to : Data Glacier



The screenshot shows a web browser window with three tabs: 'about:sessionrestore', 'Home Page - Select or create a new environment', and 'Mental health in Tech Industry'. The address bar shows 'localhost:8888/tree'. The JupyterLab interface has a top bar with 'Quit' and 'Logout' buttons. Below the top bar are tabs for 'Files', 'Running', and 'Clusters'. The 'Files' tab is active, showing a file browser view. A message 'Select items to perform actions on them.' is displayed above a table. The table has columns for 'Name', 'Last Modified', and 'File size'. It lists four items: 'depression' (4 months ago), 'Mental health in Tech Industry.ipynb' (Running, 13 hours ago, 244 kB), 'rfc.sav' (a day ago, 3.73 MB), and 'survey.csv' (5 months ago, 304 kB).

	Name	Last Modified	File size
<input type="checkbox"/>	depression	4 months ago	
<input type="checkbox"/>	Mental health in Tech Industry.ipynb	Running 13 hours ago	244 kB
<input type="checkbox"/>	rfc.sav	a day ago	3.73 MB
<input type="checkbox"/>	survey.csv	5 months ago	304 kB

Jupyter notebook screenshots showing model creation process:



The screenshot shows a Jupyter notebook titled "Mental health in Tech Industry" with the following code and output:

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: data = pd.read_csv('survey.csv')
data.head()
```

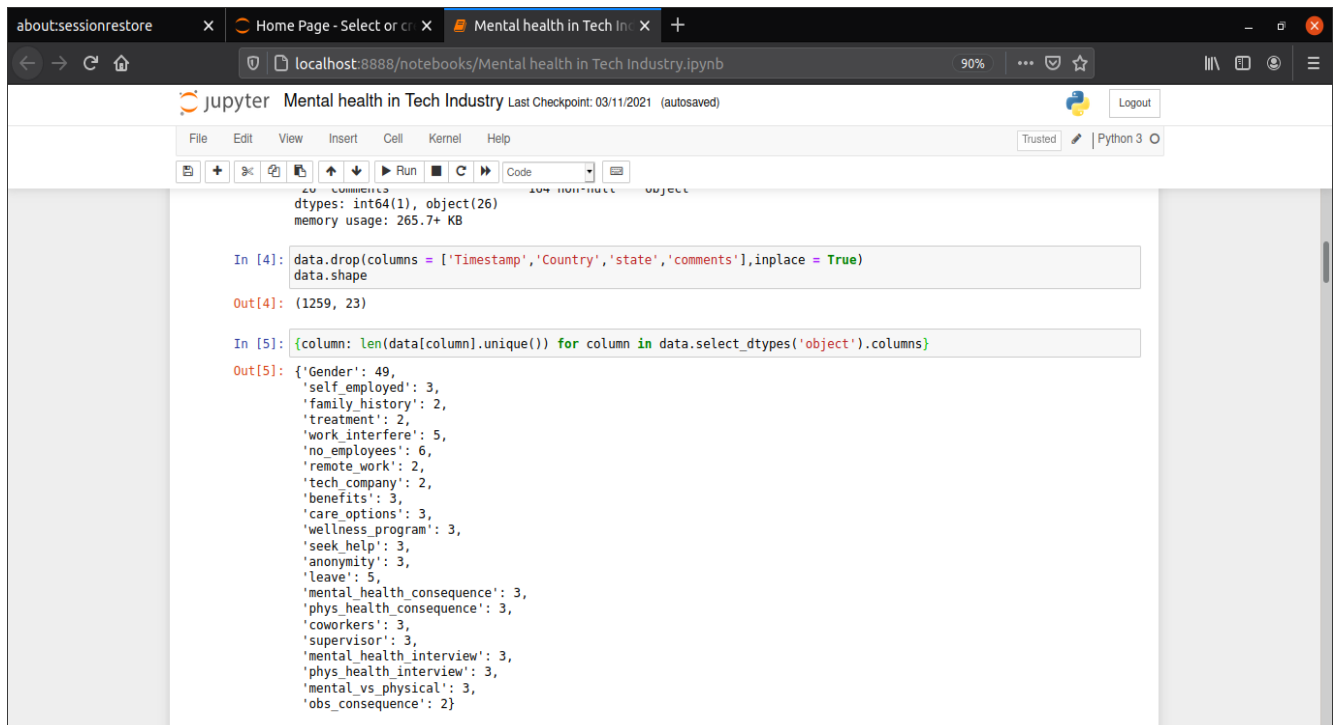
Out[2]:

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...	leave	mental_health_consequen
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	...	Somewhat easy	
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	...	Don't know	May
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	...	Somewhat difficult	
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Often	26-100	...	Somewhat difficult	Y
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	...	Don't know	

5 rows x 27 columns

```
In [3]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
```



The screenshot shows the continuation of the Jupyter notebook with the following code and output:

```
In [4]: data.drop(columns = ['Timestamp', 'Country', 'state', 'comments'], inplace = True)
data.shape
```

Out[4]: (1259, 23)

```
In [5]: [column: len(data[column].unique()) for column in data.select_dtypes('object').columns]
```

Out[5]: {'Gender': 49,
'self_employed': 3,
'family_history': 2,
'treatment': 2,
'work_interfere': 5,
'no_employees': 6,
'remote_work': 2,
'tech_company': 2,
'benefits': 3,
'care_options': 3,
'wellness_program': 3,
'seek_help': 3,
'anonymity': 3,
'leave': 5,
'mental_health_consequence': 3,
'phys_health_consequence': 3,
'coworkers': 3,
'supervisor': 3,
'mental_health_interview': 3,
'phys_health_interview': 3,
'mental_vs_physical': 3,
'obs_consequence': 2}

```
about:sessionrestore x Home Page - Select or cl x Mental health in Tech In x +
localhost:8888/notebooks/Mental health in Tech Industry.ipynb 90% ... ☆ 🏠 📄 🗑️ 🔄 📄 📄 📄 📄
Jupyter Mental health in Tech Industry Last Checkpoint: 03/11/2021 (autosaved) Logout
File Edit View Insert Cell Kernel Help Trusted Python 3
In [6]: [column: data[column].unique() for column in data.select_dtypes('object').columns]
Out[6]: {'Gender': array(['Female', 'M', 'Male', 'male', 'female', 'm', 'Male-ish', 'maile',
    'Trans-female', 'Cis Female', 'F', 'something kinda male',
    'Cis Male', 'Woman', 'f', 'Mal', 'Male (CIS)', 'queer/she/they',
    'non-binary', 'Femake', 'woman', 'Make', 'Nah', 'All', 'Enby',
    'fluid', 'Genderqueer', 'Female', 'Androgyne', 'Agender',
    'cis-female/femme', 'Guy (-ish)', 'male leaning androgynous',
    'Male', 'Man', 'Trans woman', 'msle', 'Neuter', 'Female (trans)',
    'queer', 'Female (cis)', 'Mail', 'cis male', 'A little about you',
    'Malr', 'p', 'femal', 'Cis Man',
    'ostensibly male, unsure what that really means'], dtype=object),
    'self_employed': array([nan, 'Yes', 'No'], dtype=object),
    'family_history': array(['No', 'Yes'], dtype=object),
    'treatment': array(['Yes', 'No'], dtype=object),
    'work_interfere': array(['Often', 'Rarely', 'Never', 'Sometimes', nan], dtype=object),
    'no_employees': array(['6-25', 'More than 1000', '26-100', '100-500', '1-5', '500-1000'],
    dtype=object),
    'remote_work': array(['No', 'Yes'], dtype=object),
    'tech_company': array(['Yes', 'No'], dtype=object),
    'benefits': array(['Yes', 'Don't know', 'No'], dtype=object),
    'care_options': array(['Not sure', 'No', 'Yes'], dtype=object),
    'wellness_program': array(['No', 'Don't know', 'Yes'], dtype=object),
    'seek_help': array(['Yes', 'Don't know', 'No'], dtype=object),
    'anonymity': array(['Yes', 'Don't know', 'No'], dtype=object),
    'leave': array(['Somewhat easy', 'Don't know', 'Somewhat difficult',
    'Very difficult', 'Very easy'], dtype=object),
    'mental_health_consequence': array(['No', 'Maybe', 'Yes'], dtype=object),
    'phys_health_consequence': array(['No', 'Yes', 'Maybe'], dtype=object),
    'coworkers': array(['Some of them', 'No', 'Yes'], dtype=object),
    'supervisor': array(['Yes', 'No', 'Some of them'], dtype=object),
    'mental_health_interview': array(['No', 'Yes', 'Maybe'], dtype=object),
    'phys_health_interview': array(['Maybe', 'No', 'Yes'], dtype=object),
```

```
about:sessionrestore x Home Page - Select or cl x Mental health in Tech In x +
localhost:8888/notebooks/Mental health in Tech Industry.ipynb 90% ... ☆ 🏠 📄 🗑️ 🔄 📄 📄 📄 📄
Jupyter Mental health in Tech Industry Last Checkpoint: 03/11/2021 (autosaved) Logout
File Edit View Insert Cell Kernel Help Trusted Python 3
In [28]: from sklearn.preprocessing import OneHotEncoder
    from sklearn.preprocessing import OrdinalEncoder
    from sklearn.model_selection import train_test_split
    from sklearn.compose import ColumnTransformer
    from sklearn.pipeline import Pipeline
    from sklearn.ensemble import RandomForestClassifier
    from sklearn.tree import DecisionTreeClassifier
    from sklearn.linear_model import LogisticRegression
    from sklearn.naive_bayes import GaussianNB

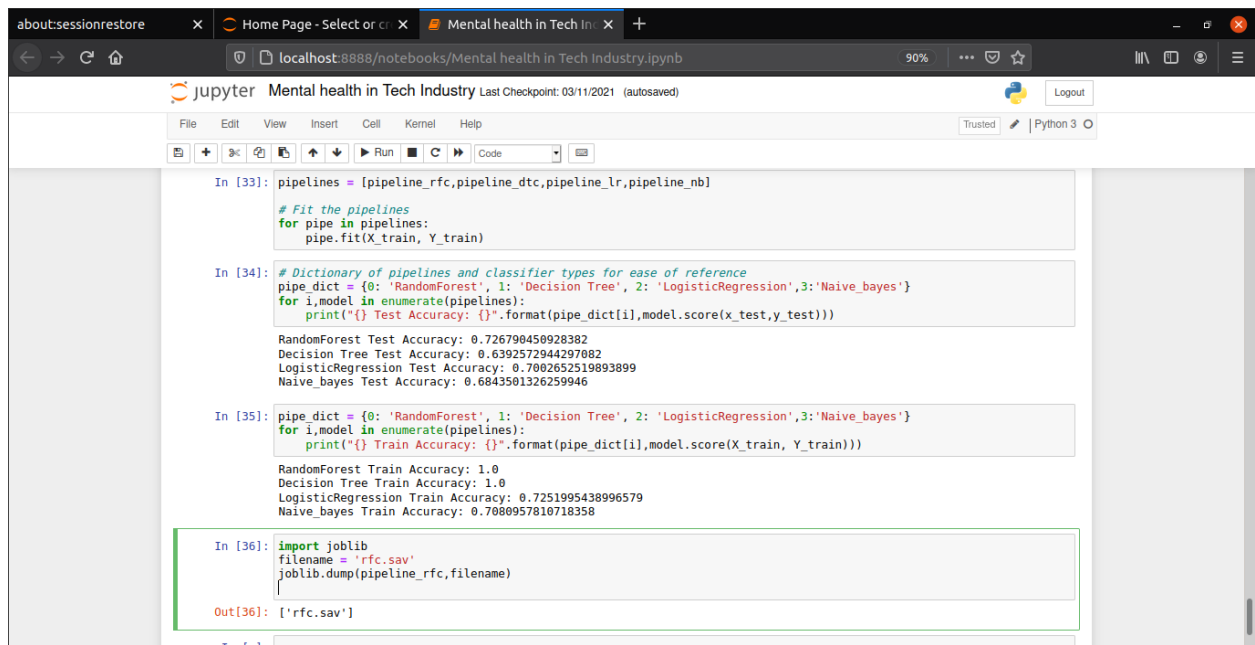
In [29]: X = data.drop(columns = 'treatment')
    y = data['treatment']
    X_train,X_test,Y_train,y_test = train_test_split(X,y,test_size = 0.3, random_state = 10)

In [30]: ct = ColumnTransformer([('enc',OrdinalEncoder(),ordinal_features),
    ('ohe',OneHotEncoder(),nominal_features)],remainder = 'passthrough')

In [31]: X_train.columns
Out[31]: Index(['Age', 'Gender', 'family_history', 'work_interfere', 'benefits',
    'care_options', 'wellness_program', 'seek_help', 'anonymity', 'leave',
    'mental_health_consequence', 'coworkers', 'mental_health_interview',
    'mental_vs_physical', 'obs_consequence'],
    dtype='object')

In [32]: pipeline_rfc = Pipeline([('transformer',ct),('rfc',RandomForestClassifier())])
    pipeline_dtc = Pipeline([('transformer',ct),('dtc',DecisionTreeClassifier())])
    pipeline_lr = Pipeline([('transformer',ct),('lr',LogisticRegression())])
    pipeline_nb = Pipeline([('transformer',ct),('lr',GaussianNB())])

In [33]: pipelines = [pipeline_rfc,pipeline_dtc,pipeline_lr,pipeline_nb]
```



The screenshot shows a Jupyter Notebook titled "Mental health in Tech Industry" with the following code and output:

```
In [33]: pipelines = [pipeline_rfc, pipeline_dtc, pipeline_lr, pipeline_nb]

# Fit the pipelines
for pipe in pipelines:
    pipe.fit(X_train, Y_train)

In [34]: # Dictionary of pipelines and classifier types for ease of reference
pipe_dict = {0: 'RandomForest', 1: 'Decision Tree', 2: 'LogisticRegression', 3: 'Naive_bayes'}
for i, model in enumerate(pipelines):
    print("{} Test Accuracy: {}".format(pipe_dict[i], model.score(x_test, y_test)))

RandomForest Test Accuracy: 0.72679450928382
Decision Tree Test Accuracy: 0.6392572944297082
LogisticRegression Test Accuracy: 0.7002652519893899
Naive_bayes Test Accuracy: 0.6843501326259946

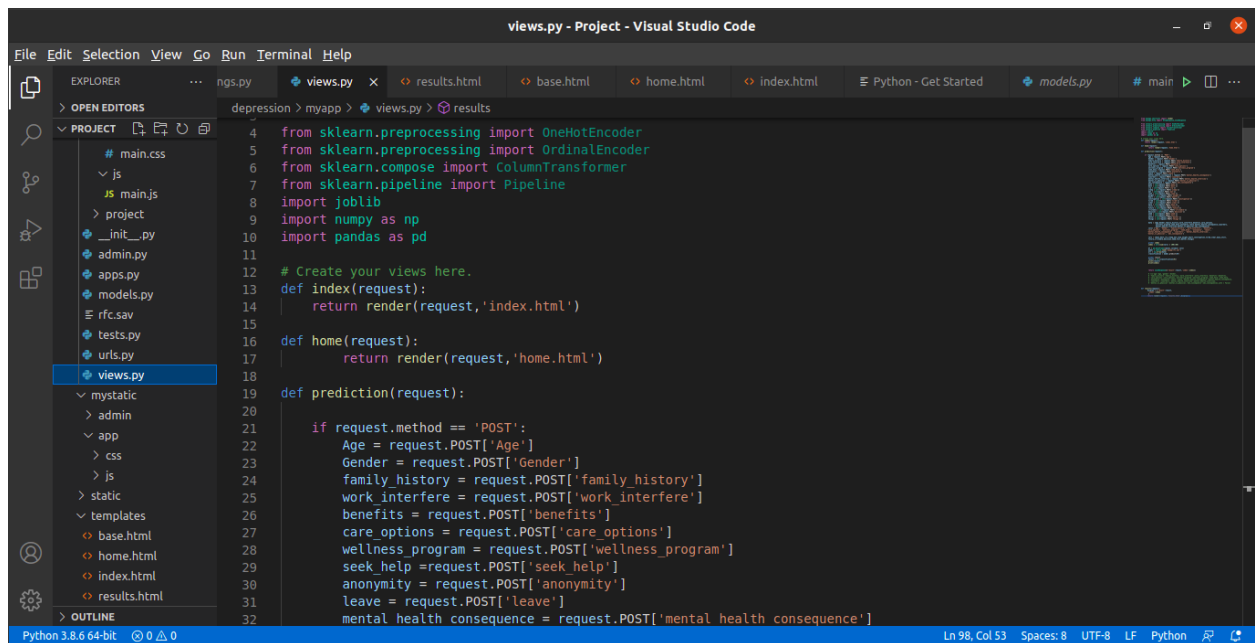
In [35]: pipe_dict = {0: 'RandomForest', 1: 'Decision Tree', 2: 'LogisticRegression', 3: 'Naive_bayes'}
for i, model in enumerate(pipelines):
    print("{} Train Accuracy: {}".format(pipe_dict[i], model.score(X_train, Y_train)))

RandomForest Train Accuracy: 1.0
Decision Tree Train Accuracy: 1.0
LogisticRegression Train Accuracy: 0.7251995438996579
Naive_bayes Train Accuracy: 0.7080957810718358

In [36]: import joblib
filename = 'rfc.sav'
joblib.dump(pipeline_rfc, filename)

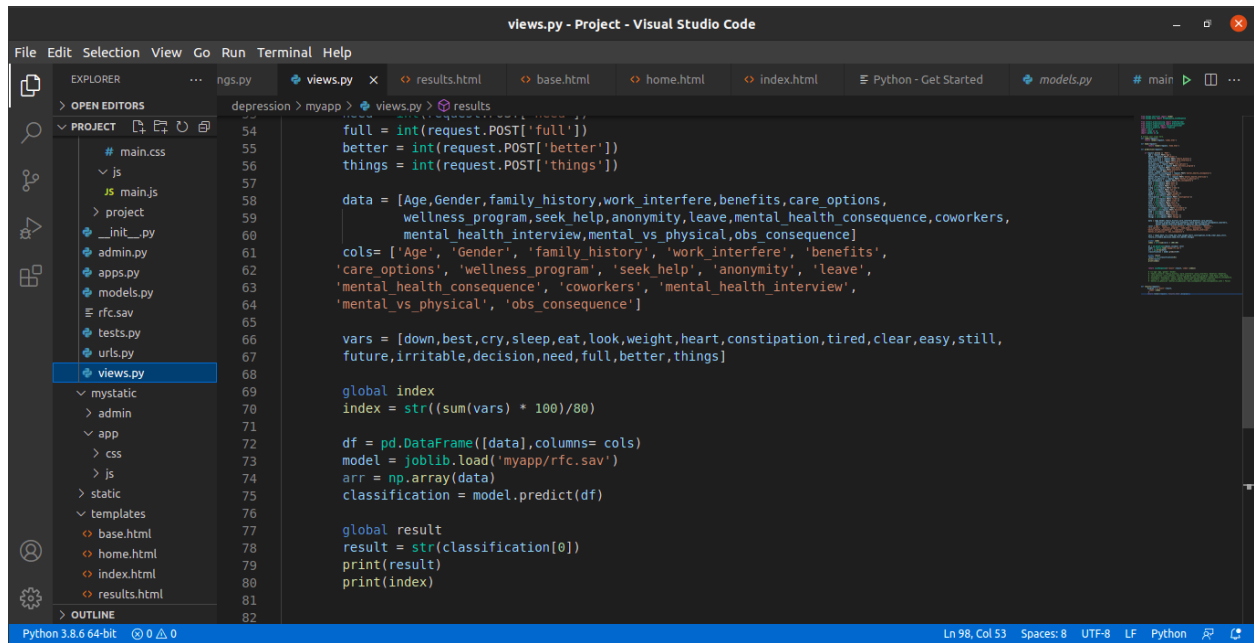
Out[36]: ['rfc.sav']
```

VS Code Screenshots showing model deployment process:



The screenshot shows a VS Code editor with a project named "views.py - Project - Visual Studio Code". The file explorer on the left shows a project structure with files like `main.css`, `main.js`, `project`, `__init__.py`, `admin.py`, `apps.py`, `models.py`, `rfc.sav`, `tests.py`, `urls.py`, and `views.py`. The `views.py` file is open in the editor, showing the following code:

```
4 from sklearn.preprocessing import OneHotEncoder
5 from sklearn.preprocessing import OrdinalEncoder
6 from sklearn.compose import ColumnTransformer
7 from sklearn.pipeline import Pipeline
8 import joblib
9 import numpy as np
10 import pandas as pd
11
12 # Create your views here.
13 def index(request):
14     return render(request, 'index.html')
15
16 def home(request):
17     return render(request, 'home.html')
18
19 def prediction(request):
20
21     if request.method == 'POST':
22         Age = request.POST['Age']
23         Gender = request.POST['Gender']
24         family_history = request.POST['family_history']
25         work_interfere = request.POST['work_interfere']
26         benefits = request.POST['benefits']
27         care_options = request.POST['care_options']
28         wellness_program = request.POST['wellness_program']
29         seek_help = request.POST['seek_help']
30         anonymity = request.POST['anonymity']
31         leave = request.POST['leave']
32         mental_health_consequence = request.POST['mental_health_consequence']
```



The screenshot shows the Visual Studio Code interface with a project named 'depression'. The Explorer sidebar on the left shows the project structure, including files like main.css, __init__.py, admin.py, apps.py, models.py, rfc.sav, tests.py, urls.py, and views.py. The views.py file is open in the editor, showing Python code for a Django view. The code handles a POST request, extracts data, and uses a machine learning model to predict a classification. The status bar at the bottom indicates the Python version is 3.8.6 64-bit.

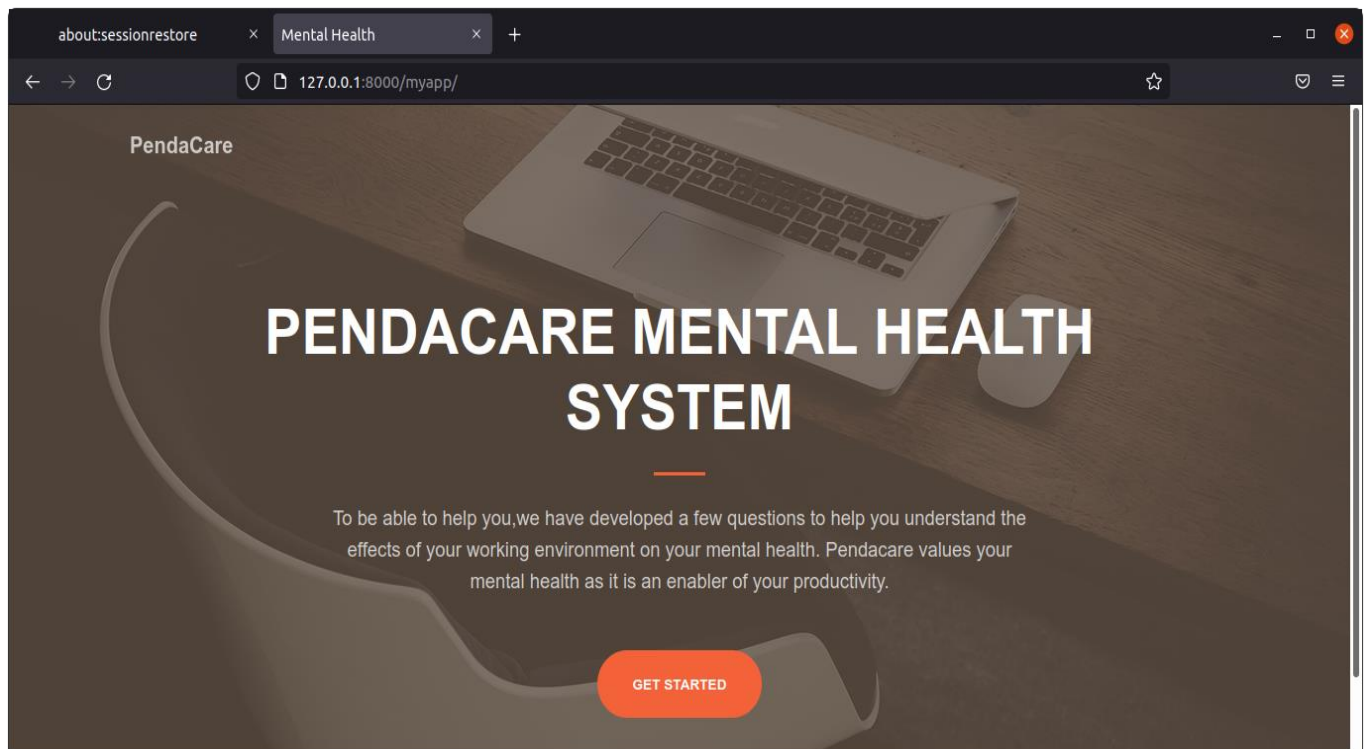
```
views.py - Project - Visual Studio Code

File Edit Selection View Go Run Terminal Help

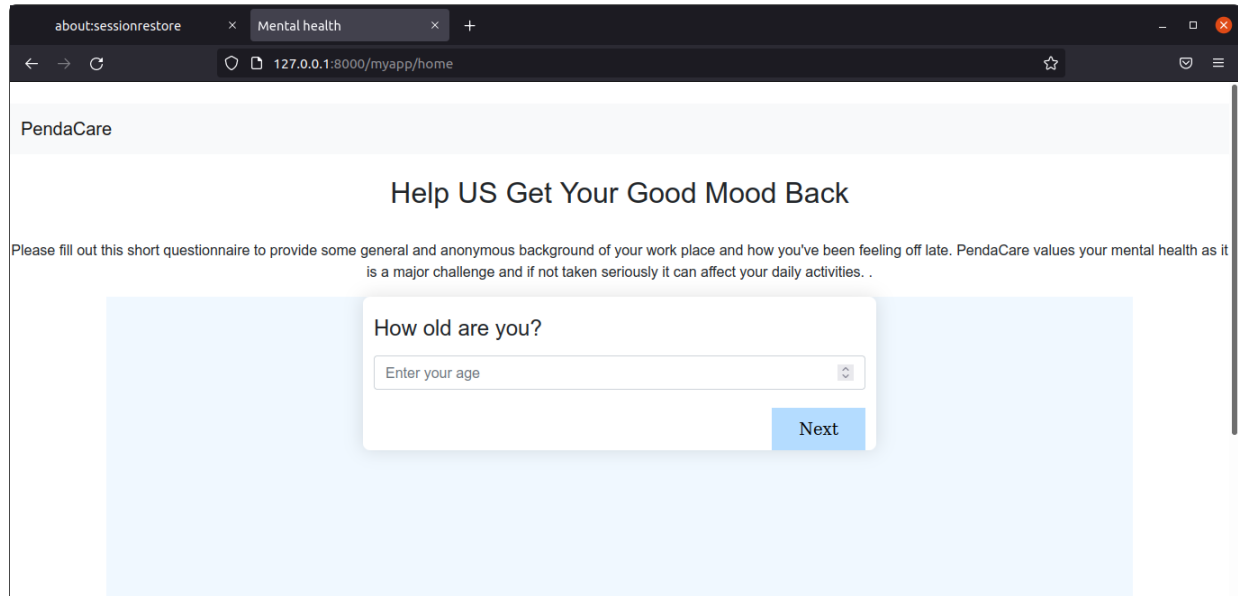
EXPLORER
> OPEN EDITORS
depression > myapp > views.py > results
54 full = int(request.POST['full'])
55 better = int(request.POST['better'])
56 things = int(request.POST['things'])
57
58 data = [Age,Gender,family_history,work_interfere,benefits,care_options,
59 wellness_program,seek_help,anonymity,leave,mental_health_consequence,coworkers,
60 mental_health_interview,mental_vs_physical,obs_consequence]
61 cols= ['Age', 'Gender', 'family history', 'work interfere', 'benefits',
62 'care options', 'wellness program', 'seek help', 'anonymity', 'leave',
63 'mental_health_consequence', 'coworkers', 'mental_health_interview',
64 'mental_vs_physical', 'obs_consequence']
65
66 vars = [down,best,cry,sleep,eat,look,weight,heart,constipation,tired,clear,easy,still,
67 future,irritable,decision,need,full,better,things]
68
69 global index
70 index = str((sum(vars) * 100)/80)
71
72 df = pd.DataFrame([data],columns= cols)
73 model = joblib.load('myapp/rfc.sav')
74 arr = np.array(data)
75 classification = model.predict(df)
76
77 global result
78 result = str(classification[0])
79 print(result)
80 print(index)
81
82

Python 3.8.6 64-bit Ln 98, Col 53 Spaces: 8 UTF-8 LF Python
```

Landing Page for the model:



Sample questions that will be interacting with the model :



The screenshot shows a web browser window with two tabs: 'about:sessionrestore' and 'Mental health'. The address bar shows '127.0.0.1:8000/myapp/home'. The page title is 'PendaCare'. The main heading is 'Help US Get Your Good Mood Back'. Below the heading is a paragraph: 'Please fill out this short questionnaire to provide some general and anonymous background of your work place and how you've been feeling off late. PendaCare values your mental health as it is a major challenge and if not taken seriously it can affect your daily activities. .'. A modal form is displayed in the center with the question 'How old are you?'. It contains a text input field with the placeholder 'Enter your age' and a 'Next' button.

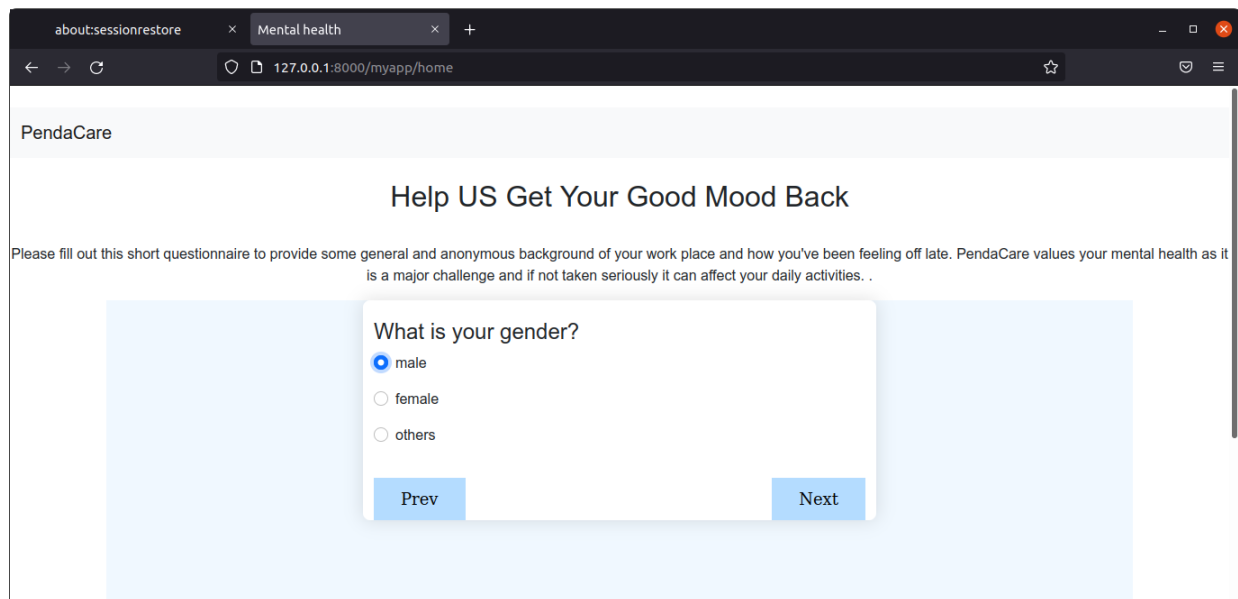
PendaCare

Help US Get Your Good Mood Back

Please fill out this short questionnaire to provide some general and anonymous background of your work place and how you've been feeling off late. PendaCare values your mental health as it is a major challenge and if not taken seriously it can affect your daily activities. .

How old are you?

Next



The screenshot shows the same web browser window as the previous one. The modal form now displays the question 'What is your gender?'. It has three radio button options: 'male' (selected), 'female', and 'others'. There are 'Prev' and 'Next' buttons at the bottom of the modal.

PendaCare

Help US Get Your Good Mood Back

Please fill out this short questionnaire to provide some general and anonymous background of your work place and how you've been feeling off late. PendaCare values your mental health as it is a major challenge and if not taken seriously it can affect your daily activities. .

What is your gender?

☒ male

☐ female

☐ others

Prev Next

about:sessionrestore x Mental health x +

← → ↻ 127.0.0.1:8000/myapp/home ☆ 🛡️ ☰

PendaCare

Help US Get Your Good Mood Back

Please fill out this short questionnaire to provide some general and anonymous background of your work place and how you've been feeling off late. PendaCare values your mental health as it is a major challenge and if not taken seriously it can affect your daily activities. .

Do you have a family history of mental illness?

☒ yes

☐ no

Prev Next

about:sessionrestore x Mental health x +

← → ↻ 127.0.0.1:8000/myapp/home ☆ 🛡️ ☰

PendaCare

Help US Get Your Good Mood Back

Please fill out this short questionnaire to provide some general and anonymous background of your work place and how you've been feeling off late. PendaCare values your mental health as it is a major challenge and if not taken seriously it can affect your daily activities. .

Do you think mental health condition interferes with your work?

☒ often

☐ rarely

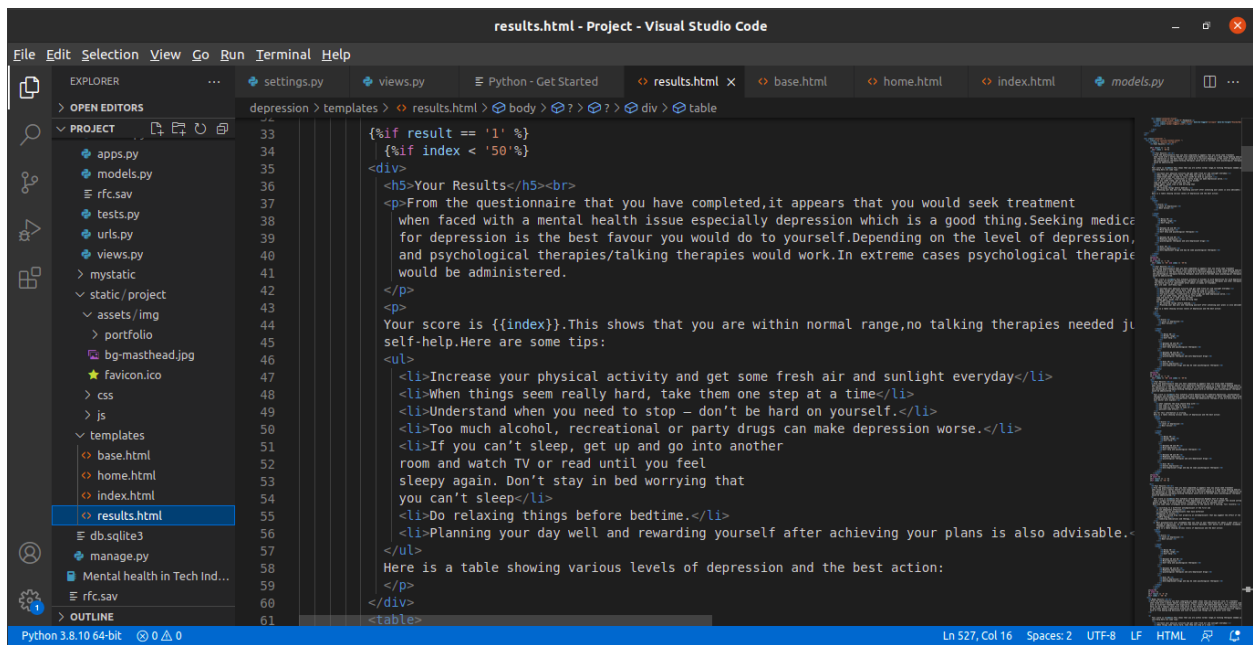
☐ sometimes

☐ never

Prev Next

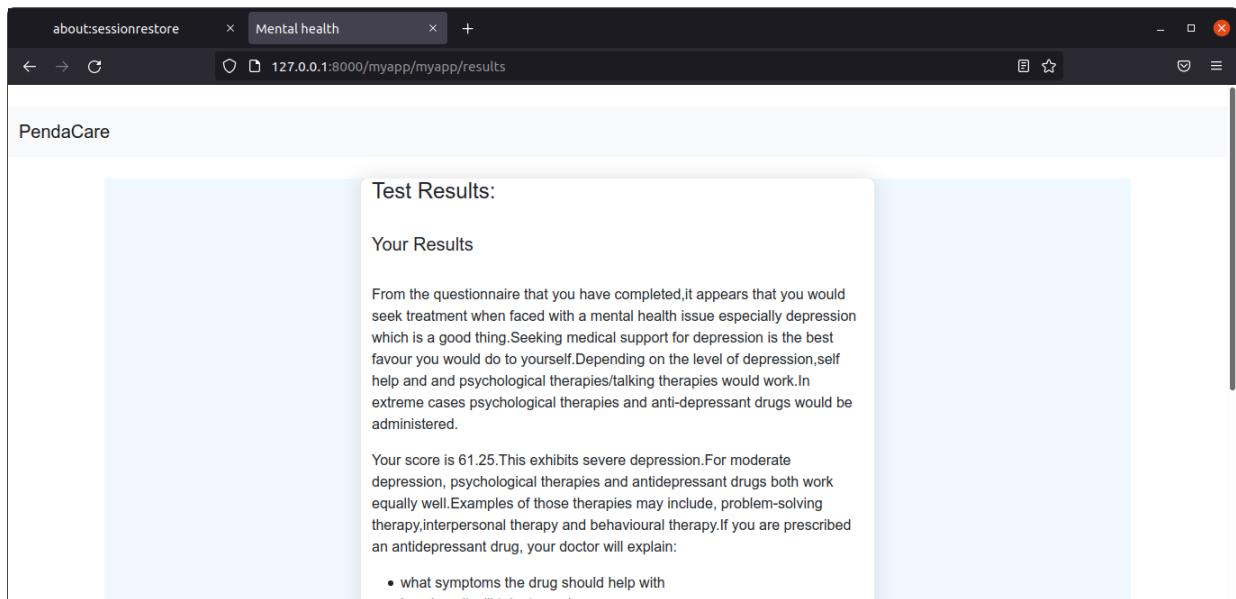
Each variable in the dataset will have such question which will be interacting with the model.

The result code based on prediction of the model



```
results.html - Project - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
PROJECT
  apps.py
  models.py
  rfc.sav
  tests.py
  urls.py
  views.py
  mystatic
  static / project
  assets / img
  portfolio
  bg-masthead.jpg
  favicon.ico
  css
  js
  templates
    base.html
    home.html
    index.html
    results.html
  db.sqlite3
  manage.py
  Mental health in Tech Ind...
  rfc.sav
OUTLINE
Python 3.8.10 64-bit 0 0 0
Ln 527, Col 16 Spaces: 2 UTF-8 LF HTML
```

The result screenshot:



about:sessionrestore

Mental health

← → ↺

🔒 127.0.0.1:8000/myapp/myapp/results

📄 ☆ 📧 ☰

Extreme cases psychological therapies and anti-depressant drugs would be administered.

Your score is 61.25. This exhibits severe depression. For moderate depression, psychological therapies and antidepressant drugs both work equally well. Examples of those therapies may include, problem-solving therapy, interpersonal therapy and behavioural therapy. If you are prescribed an antidepressant drug, your doctor will explain:

- what symptoms the drug should help with
- how long it will take to work
- how long you will need to take it
- possible side effects.

Ask for this information in writing. Here is a table showing various levels of depression and the best action:

Score	Level of depression	Best Action
Below 50	Normal range	Self help
Between 50 and 59	Mild depression	Self help and psychological therapies
Between 60 and 69	Moderate depression	psychological therapies and anti-depressant drugs
Over 70	Severe depression	Anti-depressant drugs and may be some psychological therapies