1 Task Description

In the region of Lechi during the last five years, the rate of depression has drastically increased by a margin of 20%. The regional government, concerned about this alarming situation, has decided to implement an early detection program to help potential patients.

In this assignment, you will use machine learning to predict the potential cases of depression in the region, by using the data available in the registry office of Lechi.

The task is formulated as a binary classification problem where you have two classes to predict- healthy or depressed. You will be evaluated with the F1-score metric.

2 Dataset Description

The dataset contains a list of inhabitants of the region with their respective data. The target attribute information is described below.

Number of instances: 15164

Number of attributes: 12

2.1 Target Class:

0- Healthy, 1- Depressed diagnosis

2.2 Attribute Information:

Item	Attribute	Type	Values
1	**age:** age	Numerical	Years
2	**workclass:** work sector	Categorical	Never-worked , Without-pay , Federal-gov , Self-emp-inc , State-gov , Local-gov , Self-emp-not-inc ,Private
3	**education:** education grade	Categorical	Preschool ,1st-4th ,5th-6th ,7th-8th ,9th ,10th ,11th ,12th ,Some-college ,Bachelors ,Masters ,Doctorate ,Assoc-voc ,HS-grad ,Prof-school ,Assoc-acdm
4	**education years**	Numerical	Years
5	**marital status**	Categorical	Widowed ,Separated ,Divorced ,Never-married ,Married-civ-spouse
6	**occupation**	Categorical	Armed-Forces ,Priv-house-serv ,Protective-serv ,Tech-support ,Farming-fishing ,Handlers-cleaners ,Transport-moving ,Machine-op-inspct ,Other-service ,Sales ,Adm-clerical ,Exec-managerial ,Craft-repair ,Prof-specialty
7	**relationship**	Categorical	Other-relative ,Wife ,Unmarried ,Own-child ,Not-in-family ,Husband
8	**race**	Categorical	Other ,Amer-Indian-Eskimo ,Asian-Pac-Islander ,Black ,White
9	**sex**	Categorical	Male, Female
10	**hours week**	Numerical	Hours
11	**country:** country of origin	Categorical	country
12	**permit:** driver license type	Categorical	A,B,C,D,M

3 Prediction File Submission

3.1 Model Submission: 06.11.2019, 12:00am

You are asked to kindly submit the following supporting information:

1. A brief description of the step by step methodology (i.e. pre-processing, visualization, training, testing, etc.) that you have followed to do the assignment, with the aim of illustrating the motivation behind your selected approach.

- File Format: .pdf
- Filename: **6-digit** student code (e.g. 123456.pdf)

2. The python code that you used to do the assignment, with comments within the code to ensure that they can be clearly understood.

- File Format: .ipynb, .py
- Filename: **6-digit** student code (e.g. 123456.ipynb or 123456.py)

The model can not be modified after the submission.

3.2 Test Publication: 06.11.2019, 14:00pm

3.3 Prediction Submission: 08.11.2019, 12:00am

You are kindly requested to strictly follow the described submission guidelines:

- File Format: .csv
- Filename: **6-digit** student code (e.g. 123456.csv)
- Column Format: 1 Column named **"target"**
- Row Format: Your predictions (0 or 1) with the **same number of rows and in the same order as the test set** provided to you (see below)

age	workclass	education	education_years	marital_status	occupation	relationship	race	sex	hours_week	country	permit
44	Local-gov	Some-college	10	Divorced	Adm-clerical	Unmarried		Female	40	United-States	B
41	State-gov	Some-college	10	Divorced	Adm-clerical	Not-in-family		Female	40	United-States	C
39	Private	Some-college	10	Divorced	Adm-clerical	Unmarried		Female	45	United-States	B
24	Private	Masters	14	Never-married	Adm-clerical	Not-in-family		Male	56	United-States	B
36	Private	11th	7	Divorced	Sales	Not-in-family		Female	40	United-States	A
22	Private	HS-grad	9	Never-married	Adm-clerical	Own-child		Female	40	United-States	B
35	Private	HS-grad	9	Divorced	Tech-support	Unmarried		Female	40	United-States	C
18	Private	Some-college	10	Never-married	Other-service	Own-child		Female	30	United-States	D
40	Private	HS-grad	9	Married-civ-spouse	Sales	Wife		Female	12	United-States	C
44	Private	HS-grad	9	Married-civ-spouse	Handlers-cleaners	Husband		Male	45	United-States	A
48	Private	Bachelors	13	Divorced	Adm-clerical	Own-child		Male	38	United-States	B
21	Private	Some-college	10	Never-married	Prof-specialty	Own-child		Female	8	United-States	B
22	?	Bachelors	13	Never-married	?	Own-child	Asian-Pac-Islander	Female	15	Taiwan	M
49	Private	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband		Male	40	United-States	B
35	Private	HS-grad	9	Married-civ-spouse	Transport-moving	Husband		Male	40	United-States	B
57	Private	HS-grad	9	Married-civ-spouse	Sales	Husband		Male	78	United-States	D
46	Local-gov	Some-college	10	Divorced	Adm-clerical	Unmarried		Female	45	United-States	M
50	Federal-gov	Some-college	10	Never-married	Adm-clerical	Not-in-family		Male	40	United-States	M
17	Private	HS-grad	9	Never-married	Other-service	Own-child		Female	20	United-States	A
29	Private	5th-6th	3	Never-married	Sales	Own-child		Female	40	El-Salvador	B
23	Local-gov	Bachelors	13	Never-married	Prof-specialty	Not-in-family		Female	40	United-States	A
49	Private	Assoc-voc	11	Married-civ-spouse	Machine-op-inspct	Wife		Female	40	Peru	M
23	Local-gov	Bachelors	13	Never-married	Prof-specialty	Own-child		Female	40	United-States	M
18	Private	12th	8	Never-married	Transport-moving	Own-child		Male	20	United-States	M
19	Private	HS-grad	9	Never-married	Sales	Own-child	Black	Female	40	United-States	A

target
0
0
0
0
1
1
0
1
0
0
0
1
0
1
0
0
1
0
1
0
1
0
0
0
0