**Predicting diabetic patients**

Problem statement

Nowadays, anyone can get outside food delivered at home by using a mobile application. If you are consuming outside food on a daily basis, then it can be harmful and it can cause diseases such as diabetes. Diabetes is one of the acute diseases that occurs because of bad food habits.

You are given a dataset that contains the details of 1 million patients who have undergone different tests and medications for diabetes. Your task is to predict if a patient requires medicines for diabetes again in the near future. Most of the features are made anonymous to protect the privacy of patients, insurance companies, etc.

You are required to predict the 'diabetesMeds' column.

**Note**: Refer to the **sample\_submission.csv** file to check the format of the submission.

Data description

**File**

* train.csv
* test.csv
* sample\_submission.csv

**Column description**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Column Label** | **Column Description** |
| 1 | encounter\_id | A calculated unique ID for each encounter with the patient |
| 2 | Patient\_id | Unique ID for each patient |
| 3 | race | Patient race |
| 4 | gender | Patient gender |
| 5 | age | Patient age |
| 6 | weight | Patient weight |
| 7 | Admission\_type\_id | The ID assigned while taking admission in the hospital |
| 8 | Discharge\_diposition\_id | The ID assigned while discharging |
| 9 | Admission\_source\_id | The ID of the physician for whom the patient got admitted |
| 10 | Time\_in\_hospital | Time spent by the patient in the hospital |
| 11 | diabetesMed | Two unique values, Yes or NO, representing if the patient needs medicines for diabetes or not |

Submission criteria

* You are required to write your code in the interface provided
* Ensure you include **dataframe.to\_csv** (**submission.csv**) to the last line of your code
* Your code runs against the hidden test data in the backend and is evaluated based on the evaluation metric in the following two phases:
  + **Compile & Run**: Evaluates your code against the public dataset available for download
  + **Submit**: Evaluates your code against the private unseen dataset

Evaluation criteria

The evaluation criteria for this problem is the weighted recall score.

Score=100∗recall\_score(actual\_values,predicted\_values,average=′weighted′)

Instructions

**To access files in the code editor in Python 3**

import pandas as pd

train = pd.read\_csv('dataset/train.csv')

test = pd.read\_csv('dataset/test.csv')

It is mandatory to write your results (data frame) into the **submission.csv** file. This can be performed as follows:

submission.to\_csv('submission.csv',index = False)

**To access the files in R**

library(readr)

train <- read\_csv("dataset/train.csv")

test <- read\_csv("dataset/test.csv")

It is mandatory to write your results (data frame) into the **submission.csv**file. This can be performed as follows:

write.table(submission,file = "submission.csv",row.names = FALSE,sep=",")