Project describtion document

For the classification dataset :

**General information on dataset:**

Name : diabetes\_prediction\_dataset

Number of classes : 9

Number of rows: 100K

Number of samples (training ): from 85 to 80 %

Number of samples (testing ): from 15 to 20%

**Implementation details:**

At feature extraction : all features been used (8) around 65\*8

Hyperparameters : batch\_size=32, epochs=50,learning\_rate=0.01 and number of layers

New features extracted with encoding techniques

**Result details:**





* **Confusion Matrix**:
* True Positives (TP): 580
* True Negatives (TN): 11297
* False Positives (FP): 118
* False Negatives (FN): 824

**Precision**:

Precision answers the question: “What proportion of positive identifications was actually correct?” {TP}{TP + FP}

**Recall**:

Recall answers the question: “What proportion of actual positives was identified correctly?” {TP}{TP + FN}

test Accuracy: 0.9252161979675293

train Accuracy: 0.9291