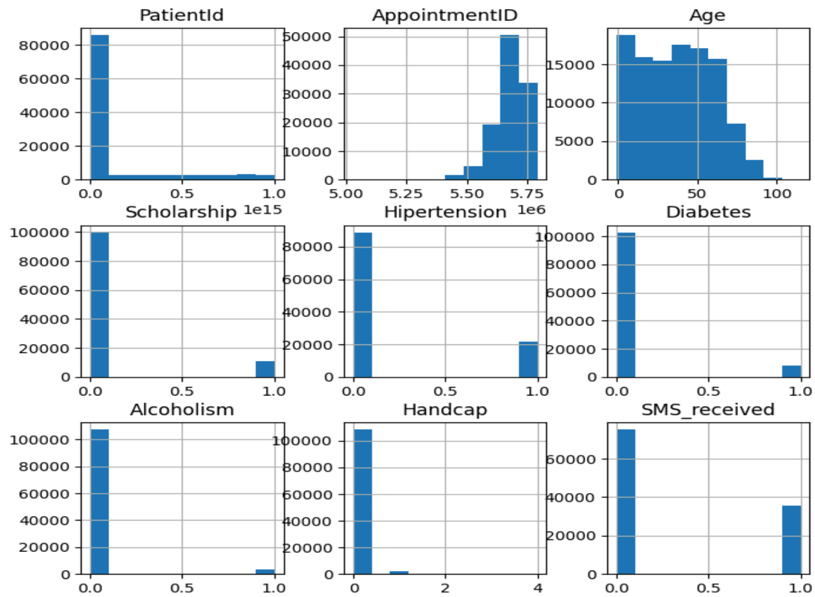
Nowroz mini-project report
Sina tayebi
400222057

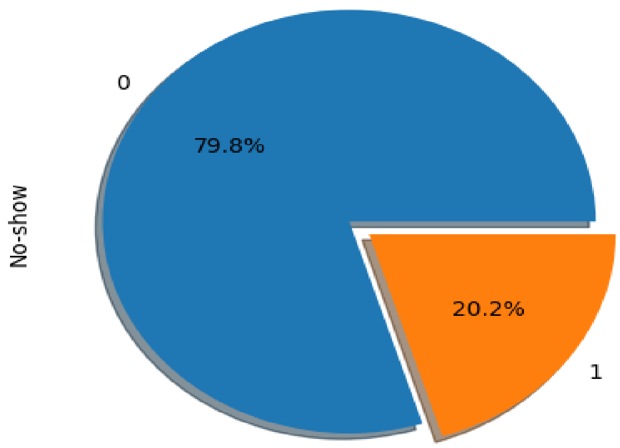
1- Dataset

The dataset includes various features such as patient age, gender, appointment date and time, scheduled date and time, neighborhood, scholarship status, and whether the patient received an SMS reminder. The target variable indicates whether the patient showed up for the appointment (1 for 'showed up' and 0 for 'no-show').

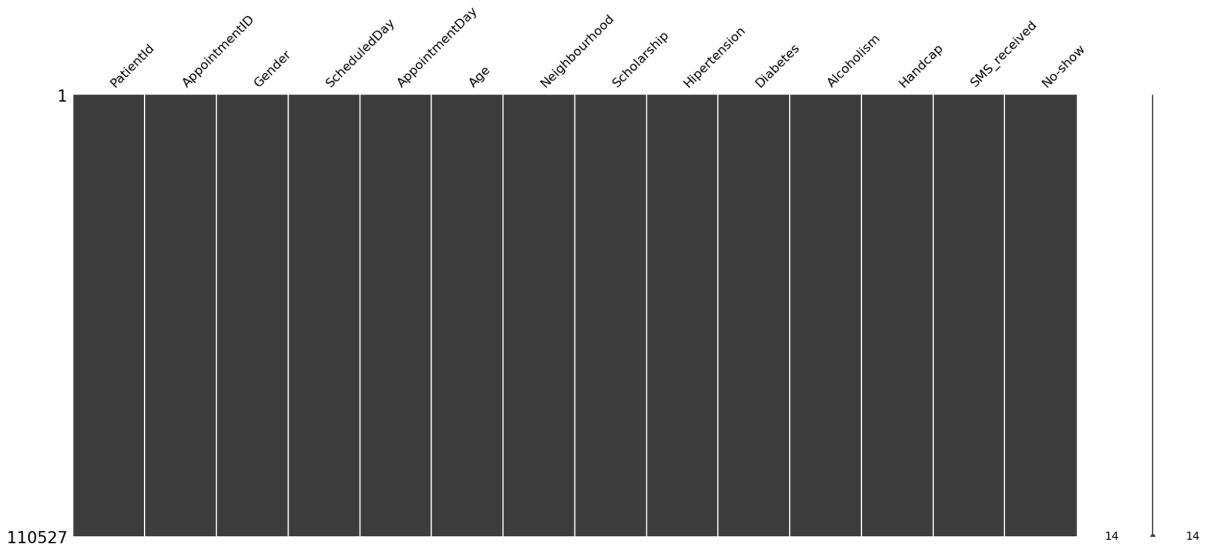
2- EDA and data preprocessing



Visualizing data distribution, shows the imbalance distribution among features.
Let us also see the pie plot of the target variable.



We can see that only 20% of target labels are in class 1. Which means we are dealing with an imbalance of data and we may need to do some balancing techniques or weighted loss functions.

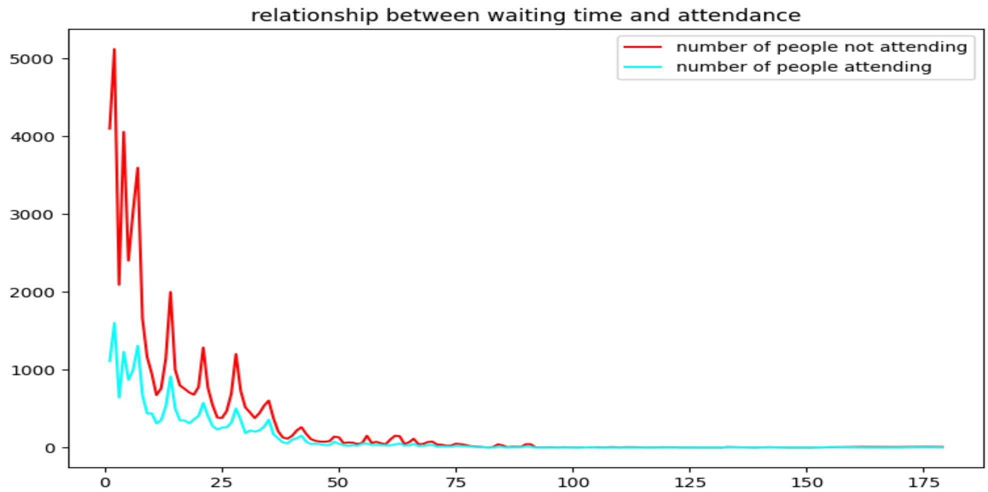


Due to the above matrix, data has no missing values.

3- Feature engineering and transformation

3-1- date/time features

We extracted the waiting time of patients for appointments by subtracting appointment date and scheduled date. It has a significant correlation with attendance of patients which we can see in the below plot.



We can see that the number of attendees decreases by increasing the waiting time.

Also we extract day of week for making a new feature which also had a significant correlation with attendances.

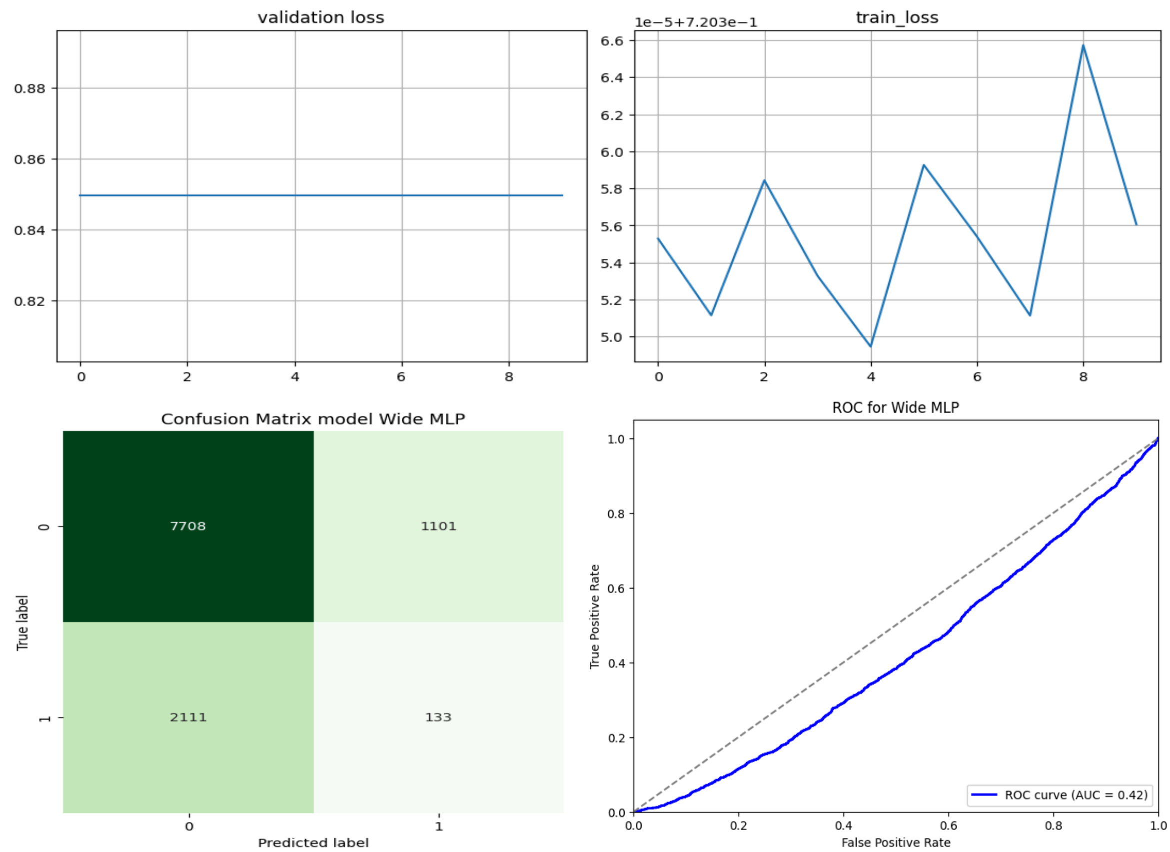
3-2- other features

Remaining features like gender encoded with one-hot encoding techniques, and Neighborhood column transformed using target encoding. We also scale numeric features like age using normalization techniques.

4- Modeling

4-1- Wide MLP

We implement a MLP with one hidden layer with 256 neurons which is quite a small network. The result of this model on 10 epochs is as follows.

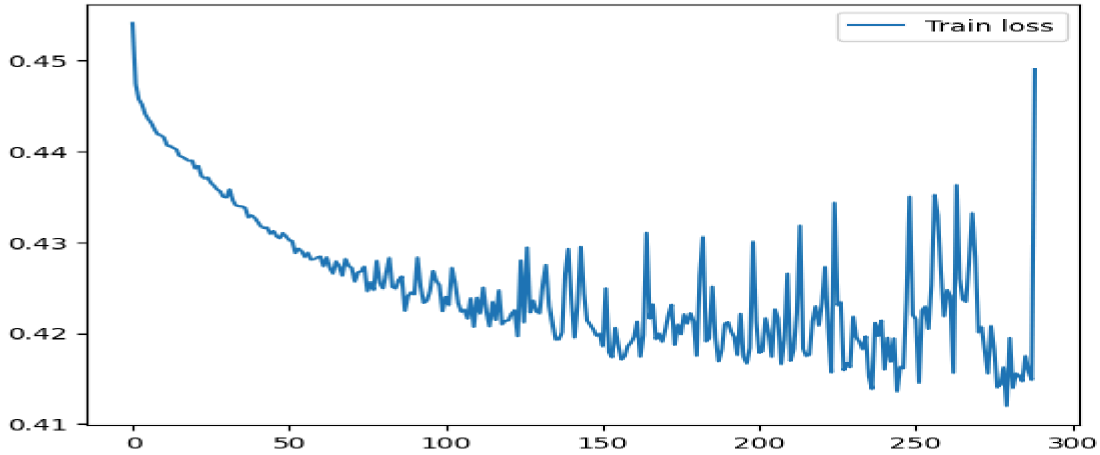


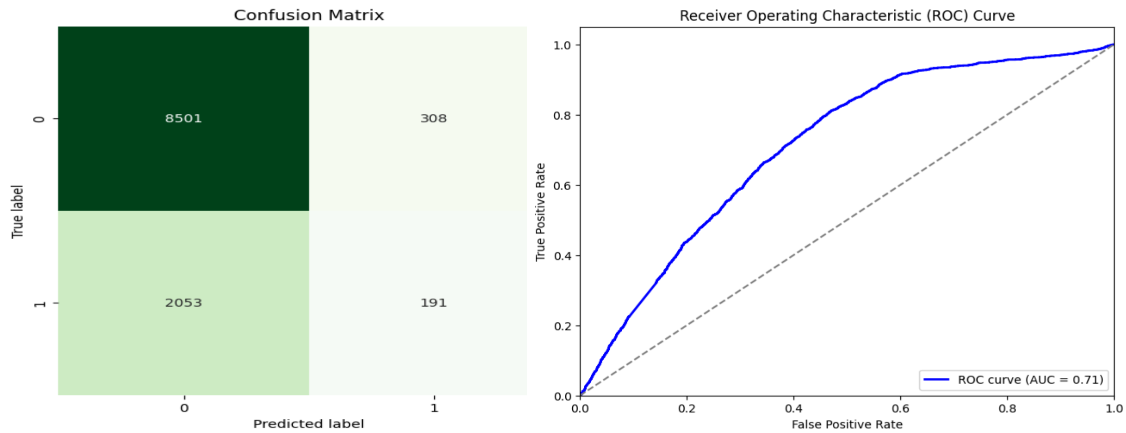
Precision of the MLP : 0.10777957860615883
Recall of the MLP : 0.059269162210338684
F1 Score of the Model : 0.07648073605520414

As we can see, the model underfitting in data and doesn't perform well. It means we need a more complex model to solve this problem.

4-2- Deep MLP

Due to the poor result of the wide MLP on the task, we designed another architecture including 3 hidden layers with 256 neurons each. We trained the model over 300 epochs.

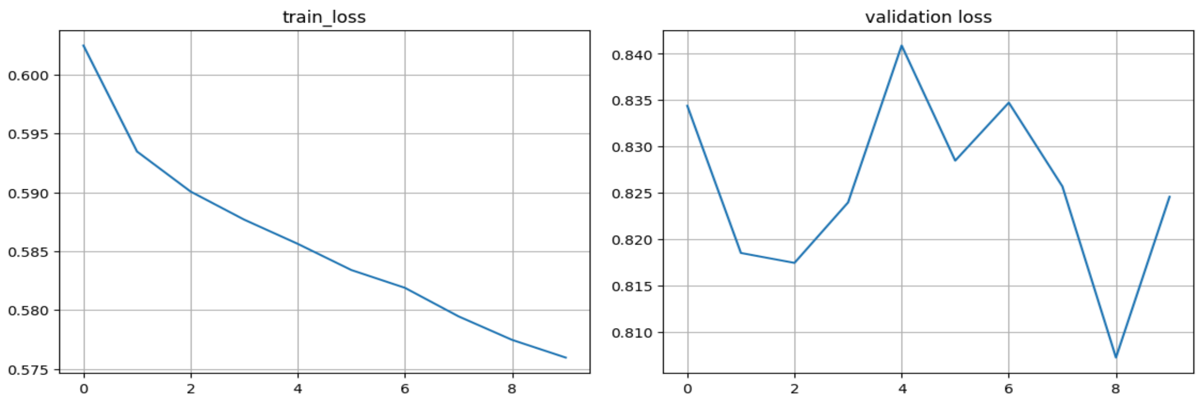


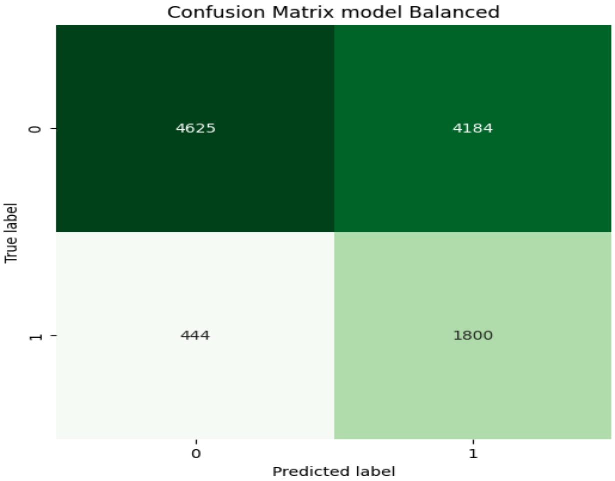


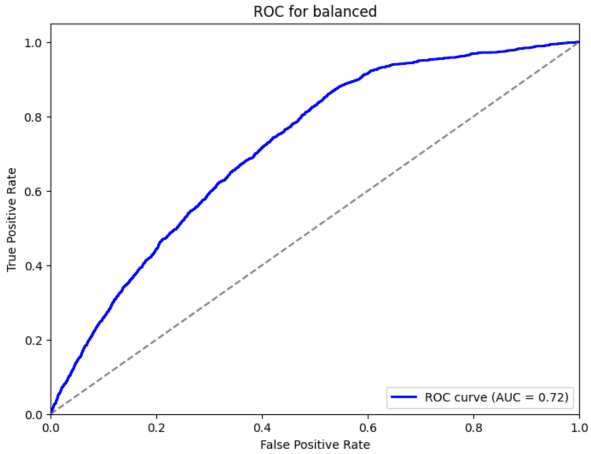
As we can see, the result wasn't sufficient enough, that's because of the imbalancing of the target variable and it requires some imbalance learning techniques to improve performance of the model. So we go ahead with deep MLP and balanced data.

4-3- imbalance learning techniques

As we saw in the EDA part, the target variable was imbalance(only 20% of labels belong to class 1). We applied the SMOTE oversampling technique to balance the data and trained the deep MLP model over 10 epochs.





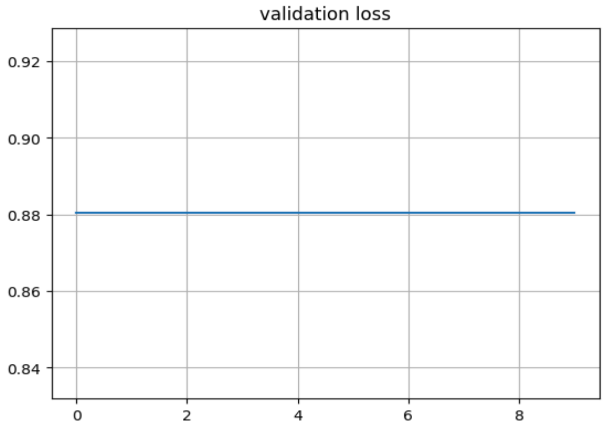


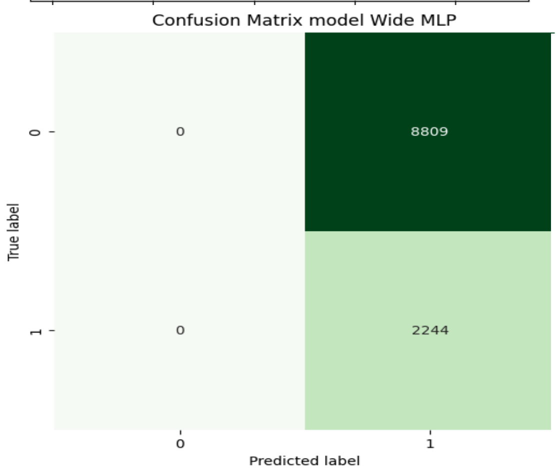
Precision of the MLP :   0.30080213903743314
Recall of the MLP   :   0.8021390374331551
F1 Score of the Model :   0.43753038405444816

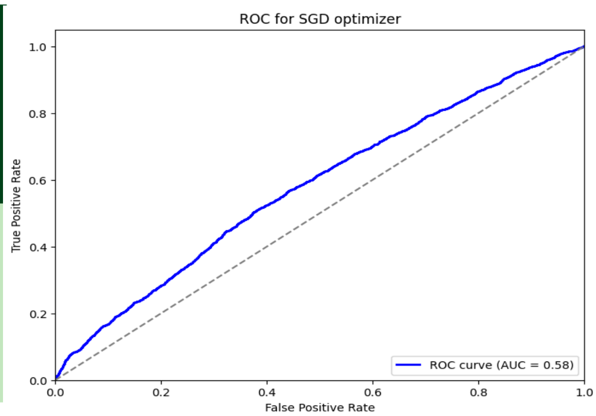
Due to the result, the oversampling makes a huge improvement in model performance.(all reported numbers belong to class 1).according to the train and validation loss plots,we can understand that with training over more epochs the performance of the model can improve.

4-4- Train model using SGD optimizer

We trained the same deep MLP architecture on the balanced data this time with SGD optimizer.





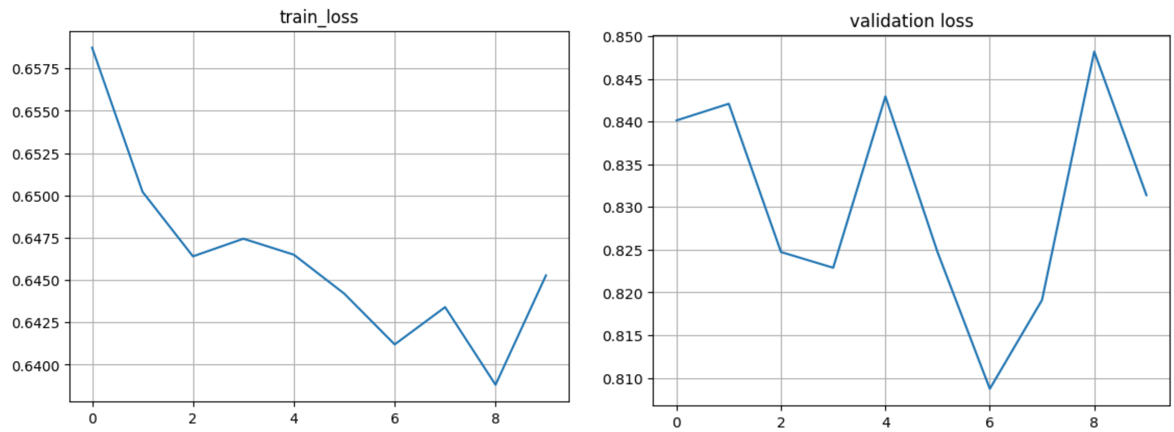


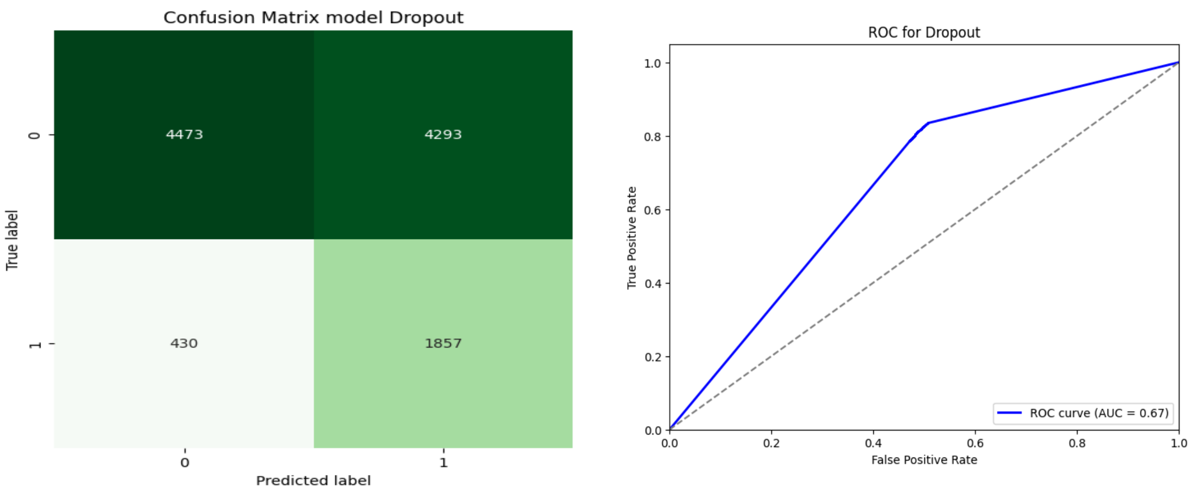
Precision of the MLP :   0.2030218040351036
Recall of the MLP   :   1.0
F1 Score of the Model :   0.337519741295029

Due to the result, the SGD optimizer needs more epochs to converge and perform well on data, so the Adam optimizer converges faster and produces better results overall.

4-5- Train with Dropout

We add 3 dropout layers for all hidden layers of our network with dropout probability equal to 0.5.



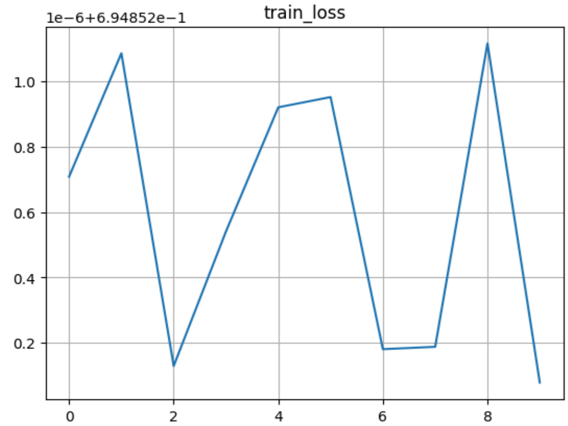


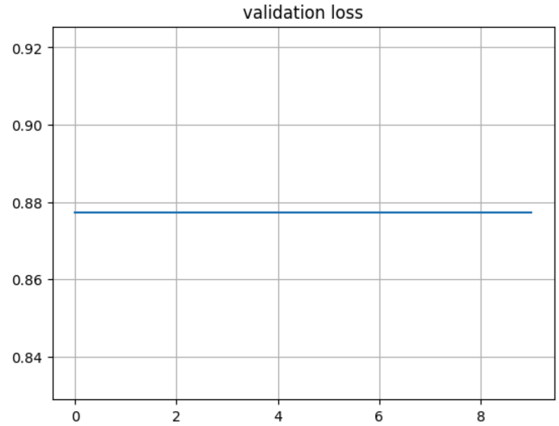
Precision of the MLP :   0.3019512195121951
Recall of the MLP   :   0.8119807608220376
F1 Score of the Model :   0.4402038639326775

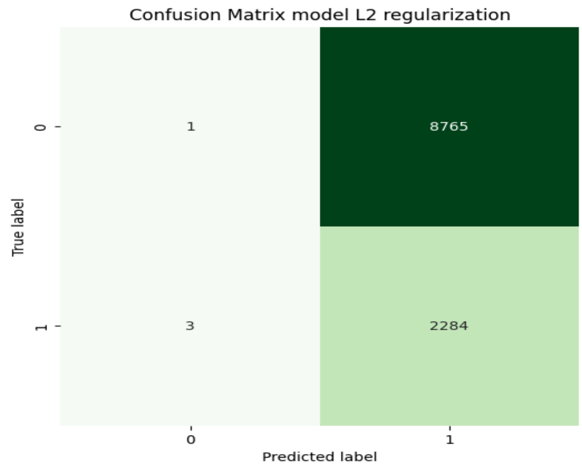
The reported result shows that the model performance improved by adding dropout. So by using dropout technique with more epochs, we are prone to get the best result.

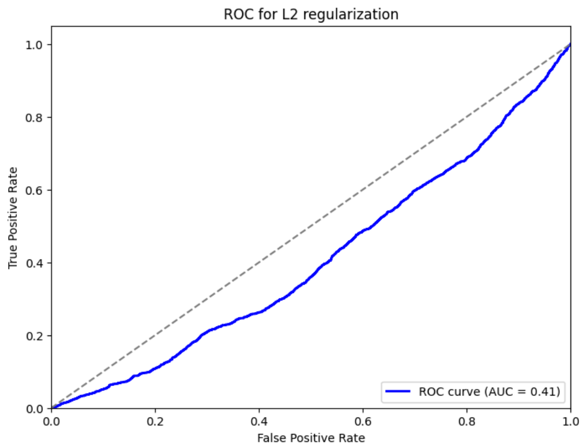
4-6- L2 regularization

Another technique to avoid overfitting is using L2 regularization to penalize weights. We used this technique by setting the weight_decay parameter of Adam optimizer equal to 0.01.









Precision of the MLP : 0.20671553986786134
Recall of the MLP : 0.9986882378662003
F1 Score of the Model : 0.34253149370125974

The reported result shows that the regularization decreases the model scores and it's not improve the model result.

5- Conclusion and future works

Due to the results of part 4, we need to use a model as complex as an MLP with 3 or more hidden layers, and appropriate balancing technique to outperform the poor result of the imbalance data and the wide MLP. the result would be better if we train the model with dropout and more training epochs. For better convergence and better results, we can use *Batch normalization* as well. On the other hand we could implement more complex models like *Bayesian neural networks* and *MC dropout* techniques to improve the result. Also we can use ensemble methods like *Stacking* to overcome the imbalance of data and lead to better results. Also more precise feature engineering and feature transformation could help us to achieve better performance.