**Solution Sheet**

1. Which model have you used for probability prediction? Explain your model.

I have tested and used multiple models to do the probability prediction. Since the variables are continuous, it is a regression task. So I have divided my approach into two parts: -

**1) ML:**

- Random Forest Regressor : This uses a forest of trees to perform a regression task

- Xgboost Regressor : This is an ensemble technique which uses boosting to improve the accuracy of previous tree based techniques. This is also used for the regression task

- Lightgbm : This is a faster, distributed version of tree based algorithms which allows multiple parameters. Specifically I have used the parameters as follows

boosting\_type': 'dart',

'objective': 'regression',

'metric': {'l2', 'l1'},

'learning\_rate': 0.05,

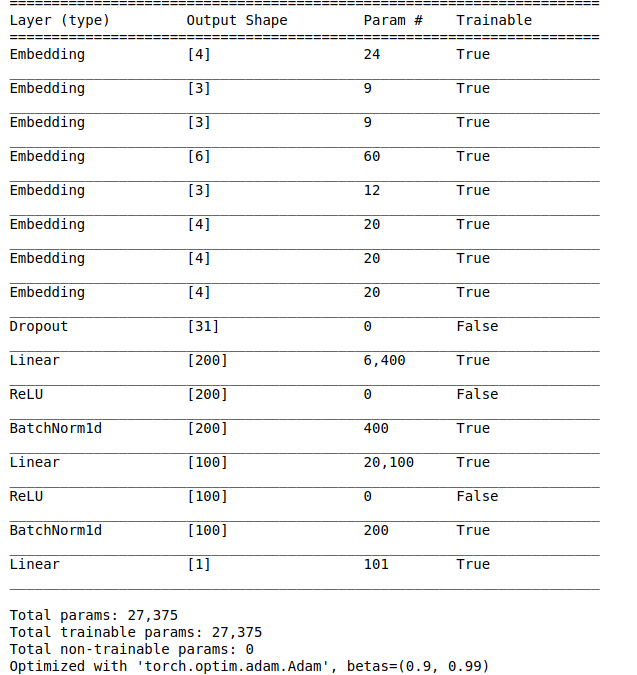
'feature\_fraction': 0.9,

'bagging\_fraction': 0.8,

'bagging\_freq': 5,

**2) DL:**

- For the Deep Learning implementation, I have used the technique of embeddings. The architecture of the network is as follows: -

-

1. Which model have you used for Diuresis Time series prediction? Explain your model.

- Since the data only has 5 columns, using an LSTM would lead to overfitting. Thus an ARIMA model has been used.

- This is a statistical technique defined as Auto Regressive Integrated Moving Average. The model goes through parts of the dataset and computes a moving average. Since time series forecasting is a regression task. It is a regressive model.