**Solution Sheet**

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

I have used DecisionTree Regressor model for the task.

Decision trees are predictive models that use a set of binary rules to calculate a target value.

Each individual tree is a fairly simple model that has branches, nodes and leaves.

a decision tree learns to map data to outputs in what is called the training phase of model building.

During training, the model is fitted with any historical data that is relevant to the problem domain and the true value we want the model to learn to predict. The model learns any relationships between the data and the target variable.

After the training phase, the decision tree produces a tree similar to the one shown above, calculating the best questions as well as their order to ask in order to make the most accurate estimates possible. When we want to make a prediction the same data format should be provided to the model in order to make a prediction. **The prediction will be an estimate based on the train data that it has been trained on.**

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

I have used ARIMA model for Diuresis Time series prediction to forecast data for 27th march in part 2 problem.

An ARIMA model is a class of statistical model for analyzing and forecasting time series data.

ARIMA is an acronym that stands for **A**uto**R**egressive **I**ntegrated **M**oving **A**verage. It is a generalization of the simpler AutoRegressive Moving Average and adds the notion of integration.

This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

* **AR**: *Autoregression*. A model that uses the dependent relationship between an observation and some number of lagged observations.
* **I**: *Integrated*. The use of differencing of raw observations (i.e. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
* **MA**: *Moving Average*. A model that uses the dependency between an observation and residual errors from a moving average model applied to lagged observations.

Each of these components are explicitly specified in the model as a parameter.

A standard notation is used of *ARIMA(p,d,q)* where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

The parameters of the ARIMA model are defined as follows:

* ***p***: The number of lag observations included in the model, also called the lag order.
* ***d***: The number of times that the raw observations are differenced, also called the degree of differencing.
* ***q***: The size of the moving average window, also called the order of moving average.