**Topic: Recurrent Neural Network (RNN)**

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: \_\_\_\_Uma Mahesh\_\_\_\_\_\_\_\_\_ Batch ID:** \_\_\_\_22102021\_\_\_\_\_\_\_

**Topic: Recurrent Neural Network.**

**Hints:**

1. **Business Problem**
   1. **What is the business objective?**
   2. **Are there any constraints?**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the image below:**



**2.1 Make a table as shown above and provide information about the features such as its data type and its relevance to the model building. And if not relevant, provide reasons and a description of the feature.**

**Using Python to perform the following:**

1. **Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

**3.2 Outlier treatment.**

1. **Model Building**

**4.1 Build a Recurrent Neural Network.**

**4.2 Train and test the model.**

**4.3 Briefly explain the model output in the documentation.**

1. **Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**
2. **Use Tensorflow for this assignment. Depending on your system configuration, use either Tensorflow GPU or Tensorflow CPU versions.**

**Problem Statement: -**

1. Here is the time series data [110,125,133,146,158,172,187,196,210]. Build an RNN/LSTM model to predict the next 10 digits.
2. Write down the applications of RNN.

The applications of RNN

Prediction problems

Language Modelling and Generating Text

Machine Translation

Speech Recognition

Generating Image Descriptions

Video Tagging

Text Summarization

Call Center Analysis

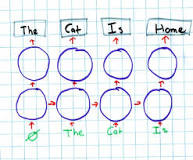
Face detection, OCR Applications as Image Recognition

1. Write about how the inputs are selected for LSTM/RNN models. Explain in terms of timesteps, samples, and features.

Time step in LSTM :

It is a standard LSTM and has two LSTM layers, each layer has **19 time steps**, and each LSTM cell contains 100 neutrons and uses the last time step value of the sequence as a training label and uses the previous time steps values as the training data

RNN in terms of samples:

[[](https://www.google.com/search?sxsrf=APq-WBuPr2mPMK87VDSsJsa1yXrLrfqbXg:1647358516586&q=What+is+sampling+in+RNN?&tbm=isch&source=iu&ictx=1&vet=1&fir=FLobmQ0p3KzfpM%252CCTEgJQH8KXgAkM%252C_&usg=AI4_-kTzD0n8ormMp-tnUGWOvXaghZ37Ng&sa=X&ved=2ahUKEwj6o8vZuMj2AhU0wzgGHbcBBjIQ9QF6BAgSEAE#imgrc=FLobmQ0p3KzfpM)](https://www.google.com/search?sxsrf=APq-WBuPr2mPMK87VDSsJsa1yXrLrfqbXg:1647358516586&q=What+is+sampling+in+RNN?&tbm=isch&source=iu&ictx=1&vet=1&fir=FLobmQ0p3KzfpM%252CCTEgJQH8KXgAkM%252C_&usg=AI4_-kTzD0n8ormMp-tnUGWOvXaghZ37Ng&sa=X&ved=2ahUKEwj6o8vZuMj2AhU0wzgGHbcBBjIQ9QF6BAgSEAE" \l "imgrc=FLobmQ0p3KzfpM)

However, when learning, the output of an RNN is a probability distribution instead of one word. **When generating text we choose only one of the words ourselves given the probabilities and feed that back into the network**. This is called sampling

1. What are the disadvantages of MLP when dealing with sequence data?

Disadvantages of MLP include **too many parameters because it is fully connected**. Parameter number = width x depth x height. Each node is connected to another in a very dense web — resulting in redundancy and inefficiency.