Project Report

|  |  |
| --- | --- |
| **Course Name (NICF)** | **Bundle Program – Artificial Intelligence** |
| **Product Name (Marketing & Sales)** | **Bundle Program – Artificial Intelligence** |
| **Module Name (NICF)** | **NICF-Deep Learning Foundations(SF)** |
| **Product Name (Marketing & Sales)** | **NICF-Deep Learning Foundations(SF)** |

|  |  |  |  |
| --- | --- | --- | --- |
| Student name | | Assessor name | |
| **Saminathan Renganayagi** | | **Rajendra Kissan** | |
| Date issued | Completion date | | Submitted on |
| **4/6/21** | **10/6/21** | | **11/6/21** |
|  | |  | |
| Project title | LSTM modelling | | |

|  |
| --- |
| Learner declaration |
| **I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.**  **Student signature: Date:** | |

Content

1. **Requirement Analysis**
2. **Why LSTM is a suitable solution for this scenario**
3. **How to: Import and Prepare Data into Google colab (Data Ingestion techniques )**
4. **How to: Preprocess data**
5. **How to: Model Data**
6. **LSTM Networks**
7. **Data Ingestion process: Explain the process of data Ingestion**
8. **Setting up the Data Preprocessing: Explain the Process of training, Min Max normalization and Testing (2 – 3 pages)**
9. **Python Code for Data Filtering and Transformation: Attach the Python Code used in the Project for Data Transformation**

**10.Transformed Data: Attach the transformed data as Annexure**

**11.Use ground truth dataset to generate labels for the test data**

**12.Modelling : Explain how you performed the preparation for visualization , plotting along with the output as Annexure**

**13.Implementing LSTM Model: Explain how you have implemented the LSTM Model (2 pages)**

**14.Comparing the Model Test template vs LSTM : Explain how you have refined the model along with Input, Process and Output ad explain is LSTM better or the template results are better**

**15.Model Adjustment: Explain the Forecast Output after you have adjusted he model based on actual production data**

1. Requirement Analysis

**This is a predictive maintenance project whether the engine will fail in a certain time period or not  
The dataset consists of the following three files:**

|  |  |
| --- | --- |
| **Training data:** | **It contains the aircraft engine run to failure data.** |
| **Testing data:** | **It contains the aircraft engine operating data without failure events recorded.** |
| **Ground truth data:** | **Here, the information about the true remaining cycles for each engine in testing data is available.** |

**According to the data description provided at the data source, the training data consists of multiple  
multivariate time series with cycle as the time unit, together with 21 sensor readings for each cycle**

**.**

|  |
| --- |
| **• Each time series can be assumed as being generated from a different engine of the same type. Each engine is assumed to start with different degrees of initial wear and manufacturing variation, and this information is unknown to the user.** |
| **• In this simulated data, the engine is assumed to be operating normally at the start of each time series. It starts to degrade at some point during the series of the operating cycles. This degrades the progresses and grows in magnitude. When a predefined threshold is reached, then the engine is considered unsafe for further operation.** |
| **• In other words, the last cycle in each time series can be considered as the failure point of the corresponding engine. Taking the sample training data as an example, the engine with id=1 fails at cycle 192, and engine with id=2 fails at cycle 287.** |

**The testing data has the same data schema as the training data.**

|  |
| --- |
| **• The only difference is that the data does not indicate when the failure occurs (in other words, the last time period does NOT represent the failure point).** |
| **• Taking the sample testing data, the engine with id=1 runs from cycle 1through cycle 31. It is not shown how many more cycles this engine can last before it fails.** |
|  |

**The ground truth data**

|  |
| --- |
| **• provides the number of remaining working cycles for the engines in the testing data.** |
| **• Taking the sample ground truth data shown as an example, the engine with id=1 in the testing data can run another 112 cycles before it fails.** |

**Since this is a time series data, we will use Long Short-Term Memory (LSTM) to classify weather the  
engine will fail in a certain time period or not**

TASK 1

**In this Deep Learning Foundations module, we had learned 3 neural netowork:  
-Artificial Neural Network (ANN)  
-Convolution Neural Network (CNN)  
-Recuurent Neural Network (RNN)**

**Here are their summry and comparison:**

Artificial Neural Network (ANN):

**Artificial Neural Network (ANN), is a group of multiple perceptrons or neurons at each layer. ANN is also  
known as a Feed-Forward Neural network because inputs are processed only in the forward direction.  
This type of neural networks are one of the simplest variants of neural networks. They pass information in  
one direction, through various input nodes, until it makes it to the output node. The network may or may  
not have hidden node layers, making their functioning more interpretable.**

|  |  |
| --- | --- |
| **Advantages:** | **Disadvantages:** |
| **Storing information on the entire network. Ability to work with incomplete knowledge. Having fault tolerance. Having a distributed memory.** | **Hardware dependence. Unexplained behavior of the network. Determination of proper network structure.** |

Convolutional Neural Network (CNN):

**Convolutional neural networks (CNN) are one of the most popular models used today. This neural  
network computational model uses a variation of multilayer perceptrons and contains one or more  
convolutional layers that can be either entirely connected or pooled. These convolutional layers create  
feature maps that record a region of image which is ultimately broken into rectangles and sent out for  
nonlinear processing.**

|  |  |
| --- | --- |
| **Advantages:** | **Disadvantages:** |
| **Very High accuracy in image recognition problems. Automatically detects the important features without any human supervision. Weight sharing.** | **CNN do not encode the position and orientation of object. Lack of ability to be spatially invariant to the input data. Lots of training data is required.** |

Recurrent Neural Network (RNN):

**Recurrent neural networks (RNN) are more complex. They save the output of processing nodes and feed  
the result back into the model (they did not pass the information in one direction only). This is how the  
model is said to learn to predict the outcome of a layer. Each node in the RNN model acts as a memory cell,  
continuing the computation and implementation of operations. If the network’s prediction is incorrect,  
then the system self-learns and continues working towards the correct prediction during  
backpropagation.**

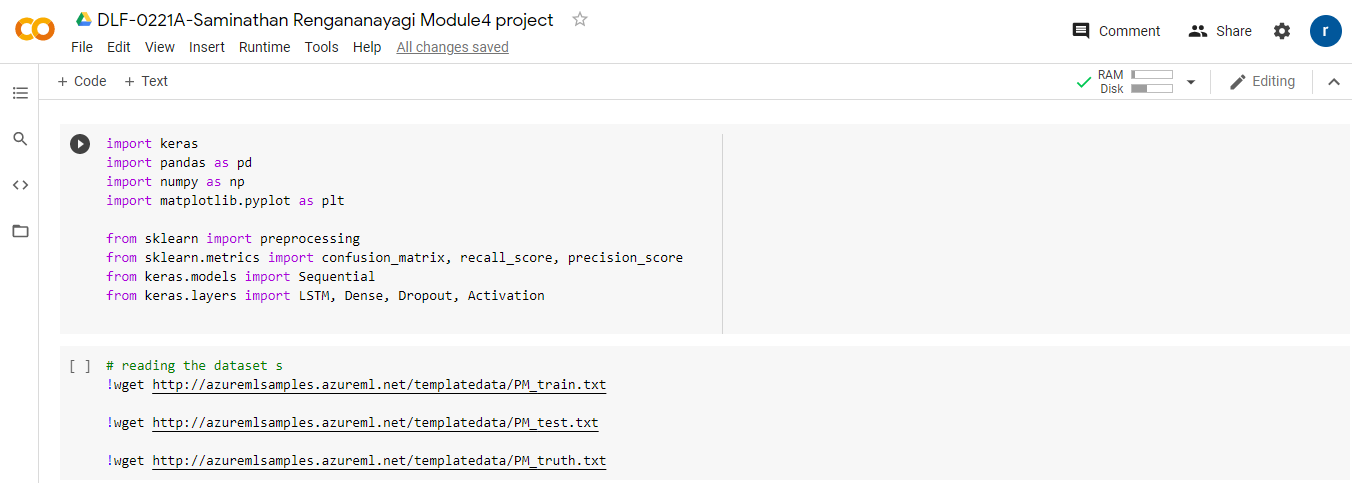
|  |  |
| --- | --- |
| **Advantages:** | **Disadvantages:** |
| **An RNN remembers each and every information through time. It is useful in time series prediction only because of the feature to remember previous inputs as well. This is called Long Short Term Memory. Recurrent neural network are even used with convolutional layers to extend the effective pixel neighborhood.** | **Gradient vanishing and exploding problems. Training an RNN is a very difficult task. It cannot process very long sequences if using tanh or relu as an activation function.** |

**Why LSTM is a suitable solution for this scenario**

**Normal feature engineering can only remember 2 or 3 rows, only LSDM can handle a  
lot, because it has memory cell.**

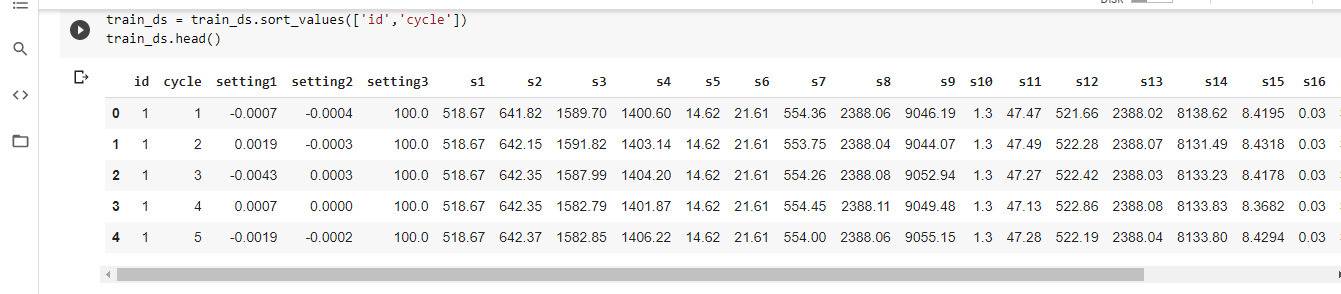
Task 2

1. How to: Import and Prepare Data into Google collab



1. How to: Preprocess data

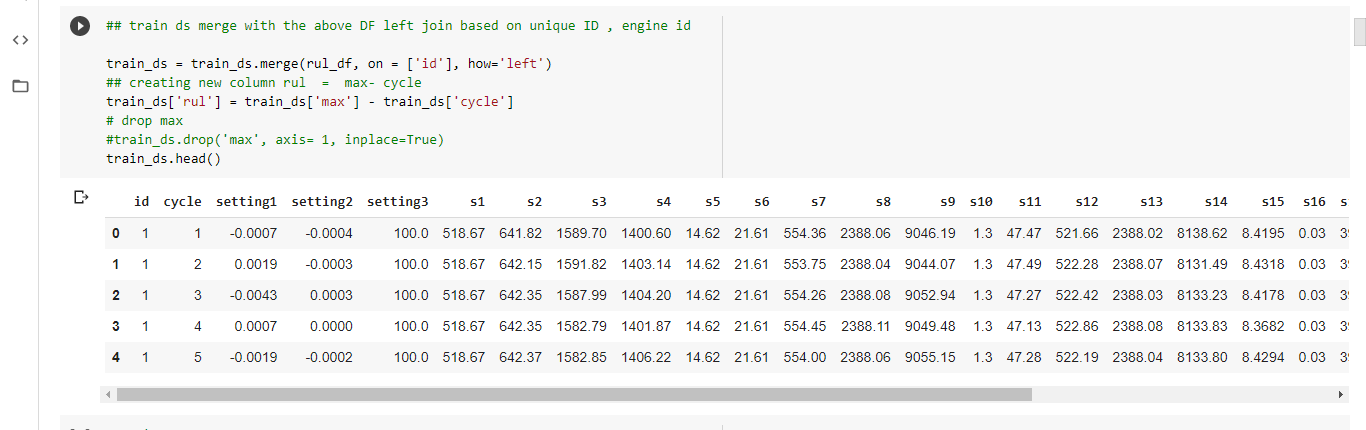




**Task 3:**

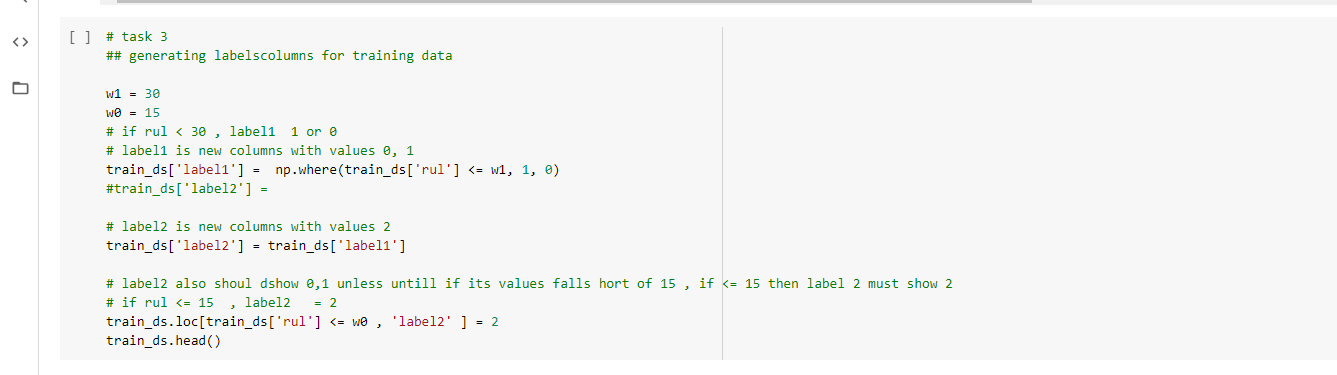


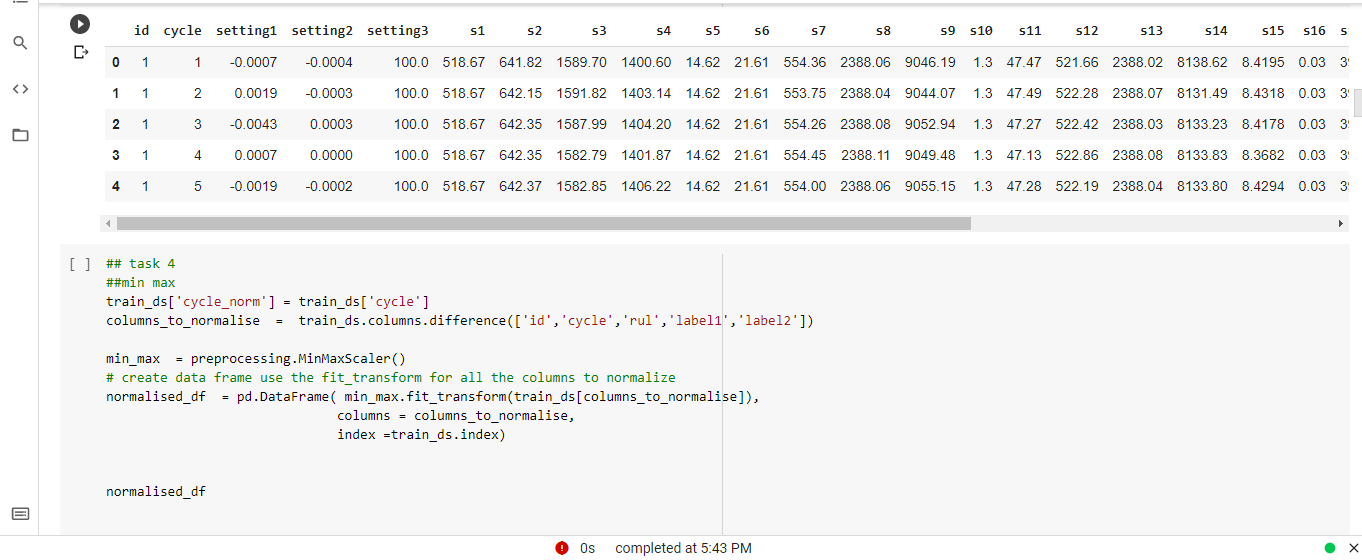
**Add the RUL columns and** **Add the RUL columns and calculate it by subtracting current cycle from max cycle**

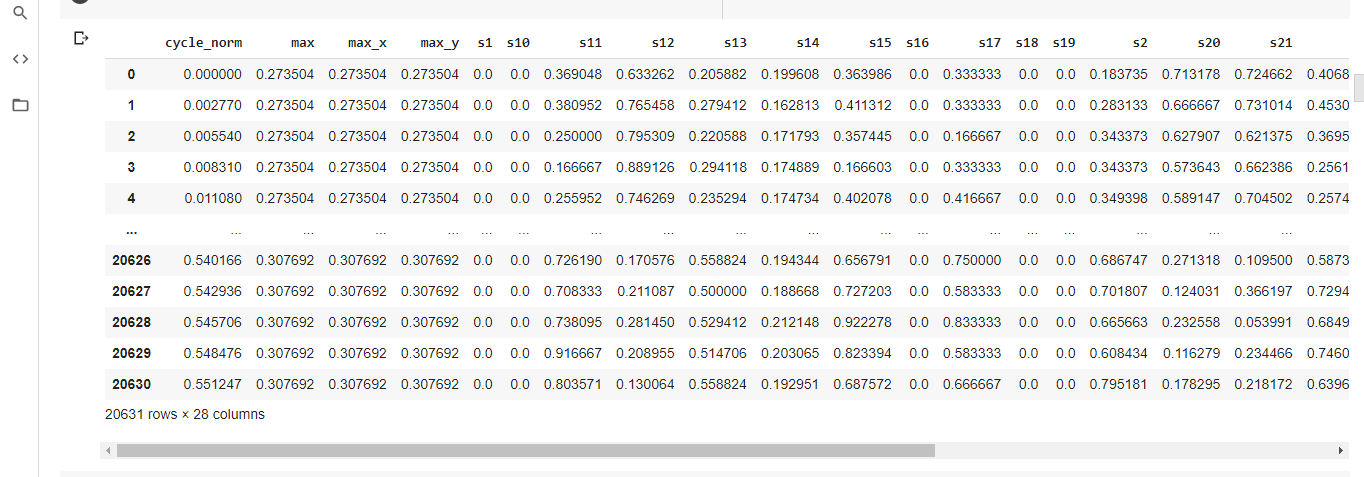


**Add label columns to show whether the engine is going**

**to fail**



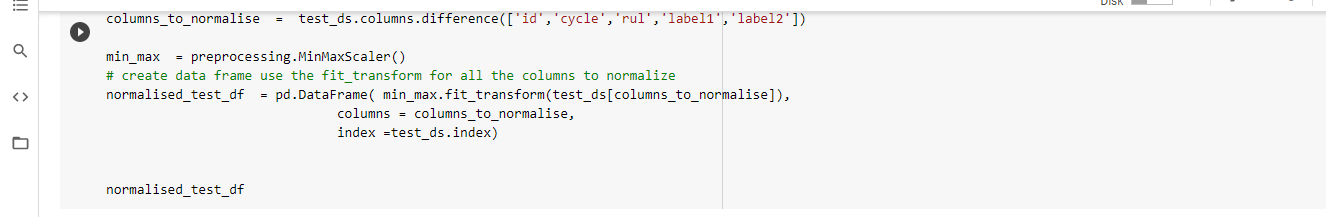


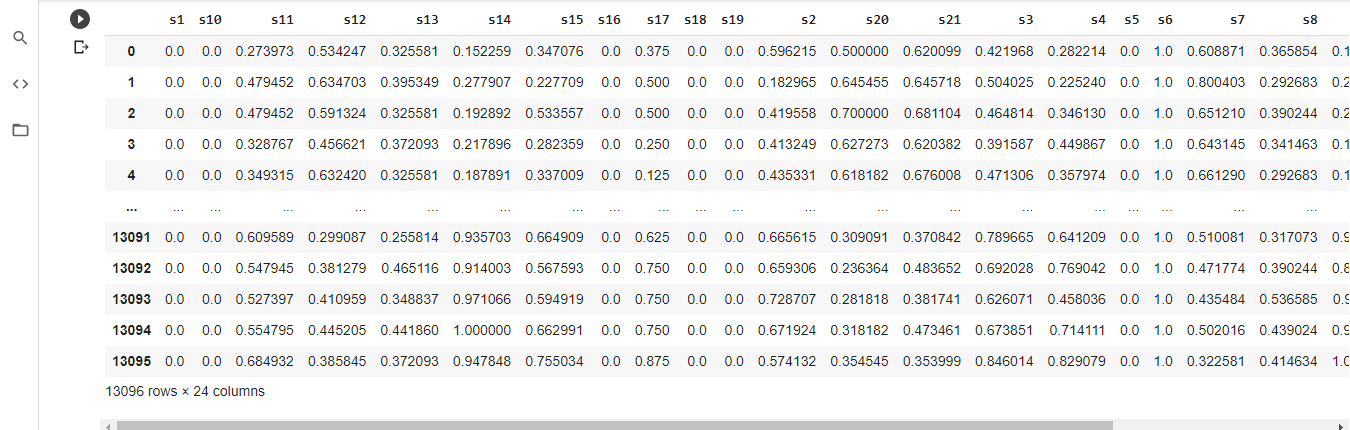


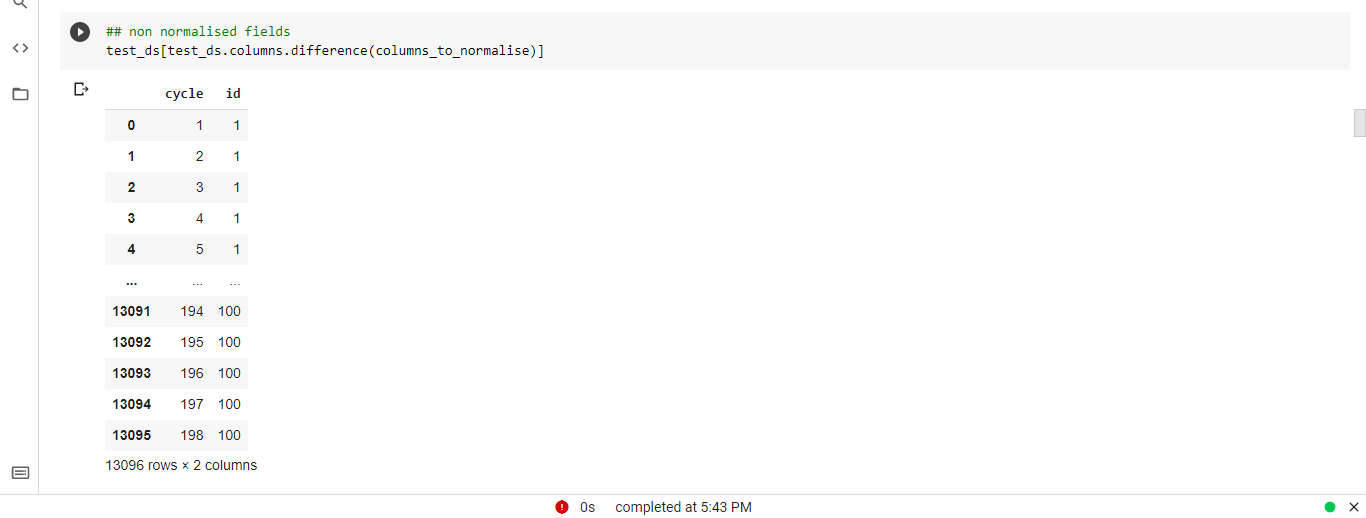


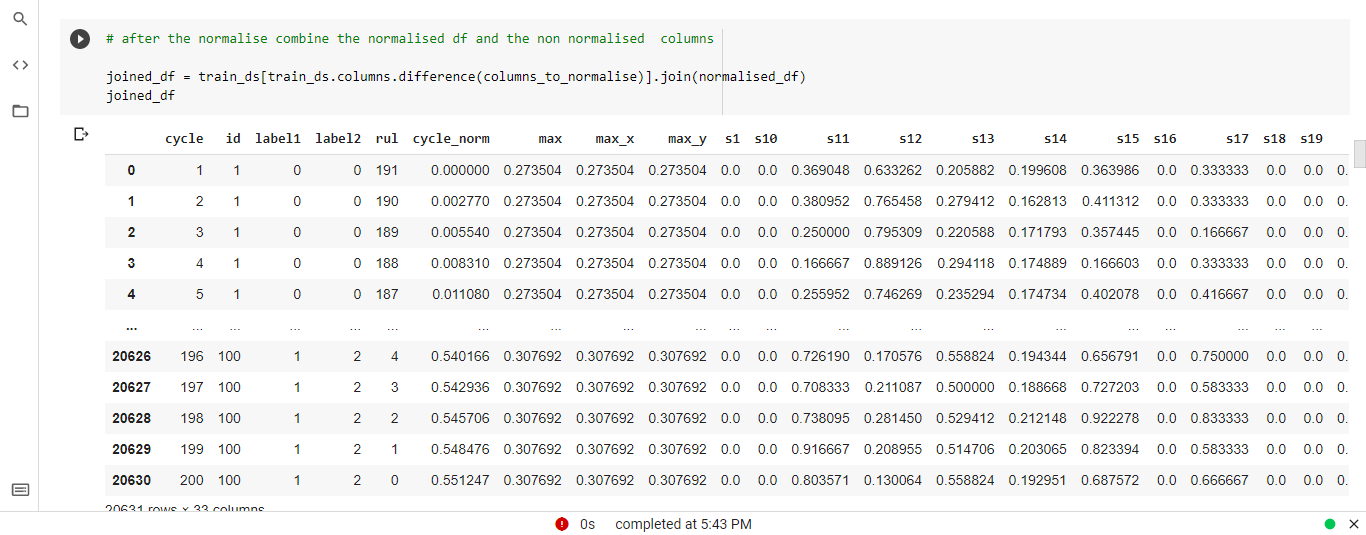
Normalize desired columns oin the non cormalized and normalized columns

into a data frame

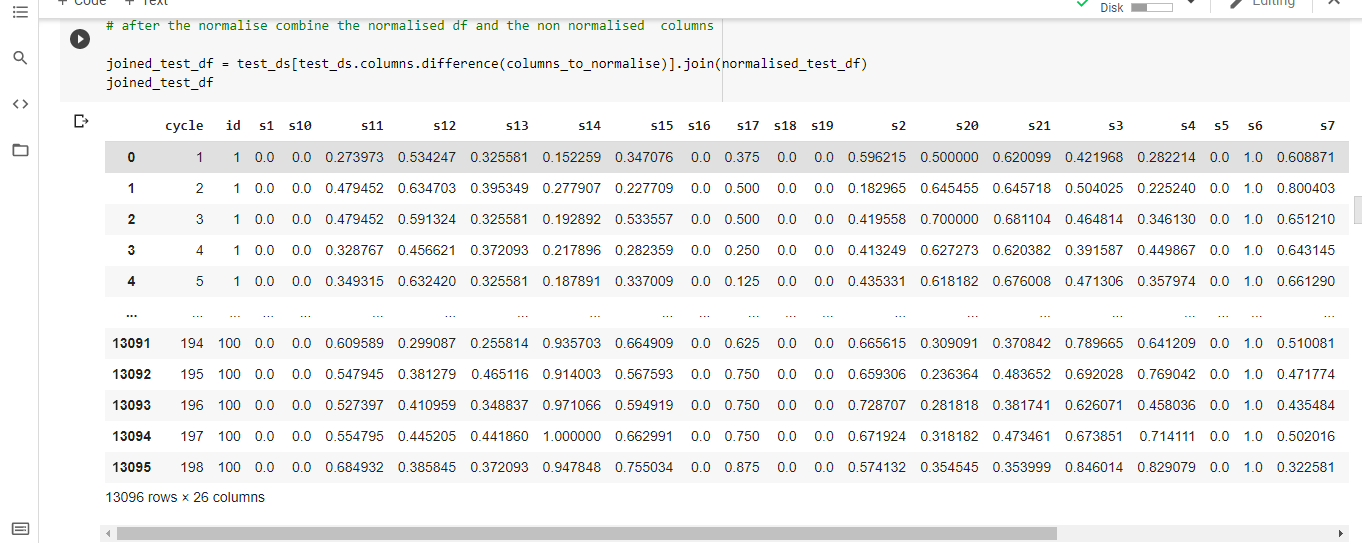


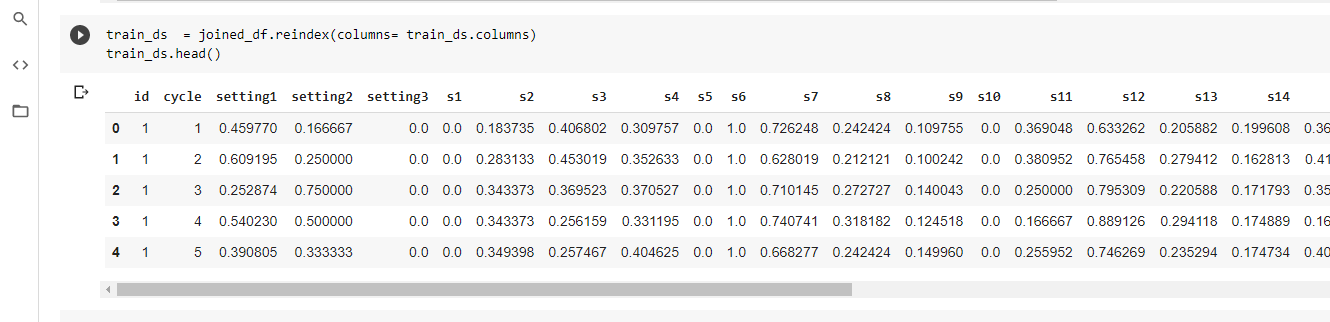


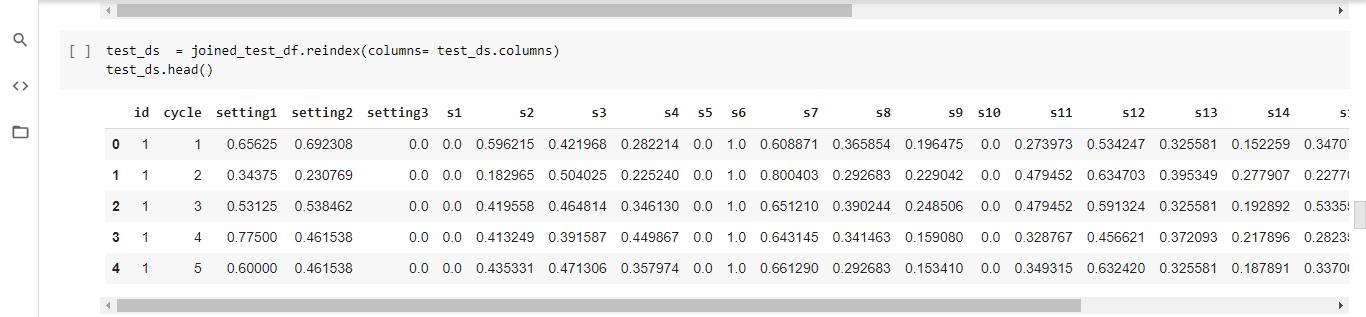




Task 4:

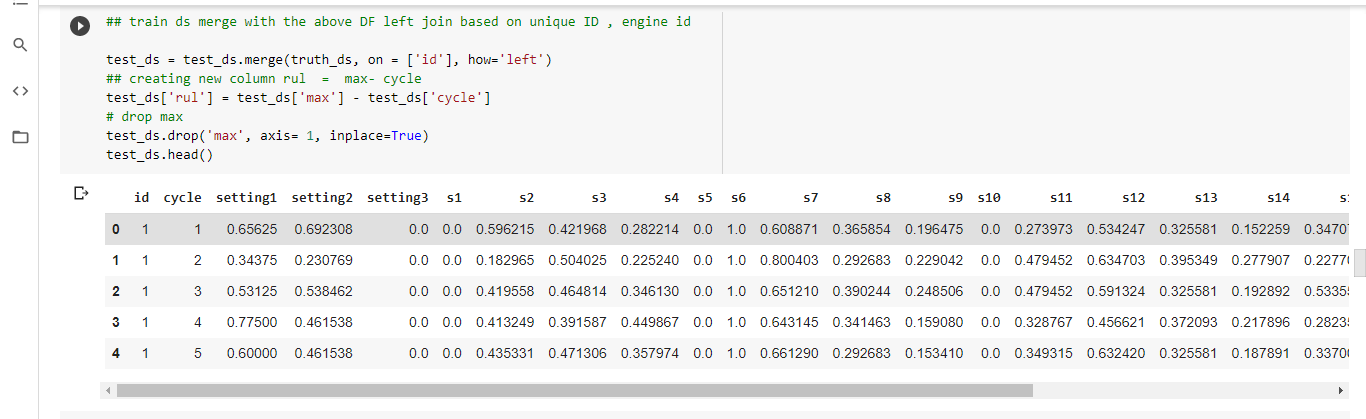




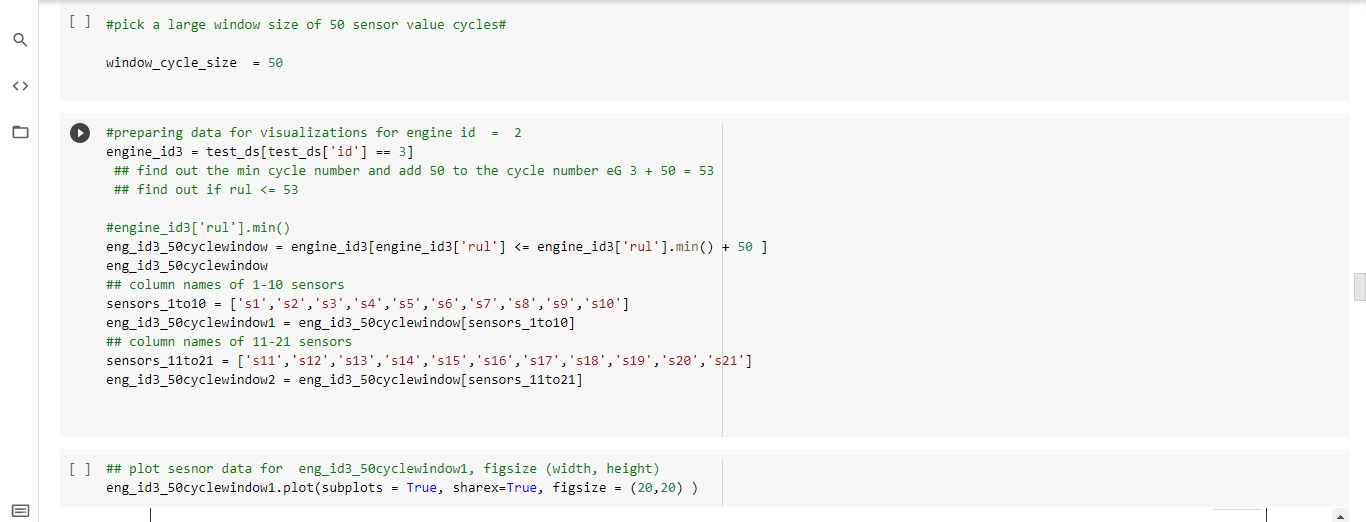


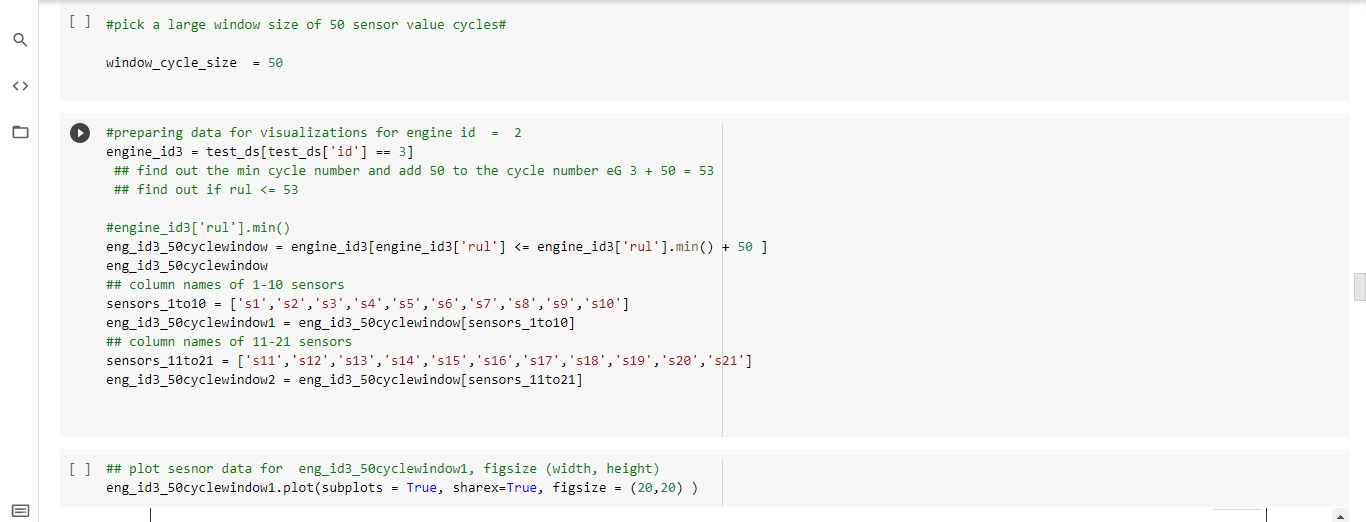


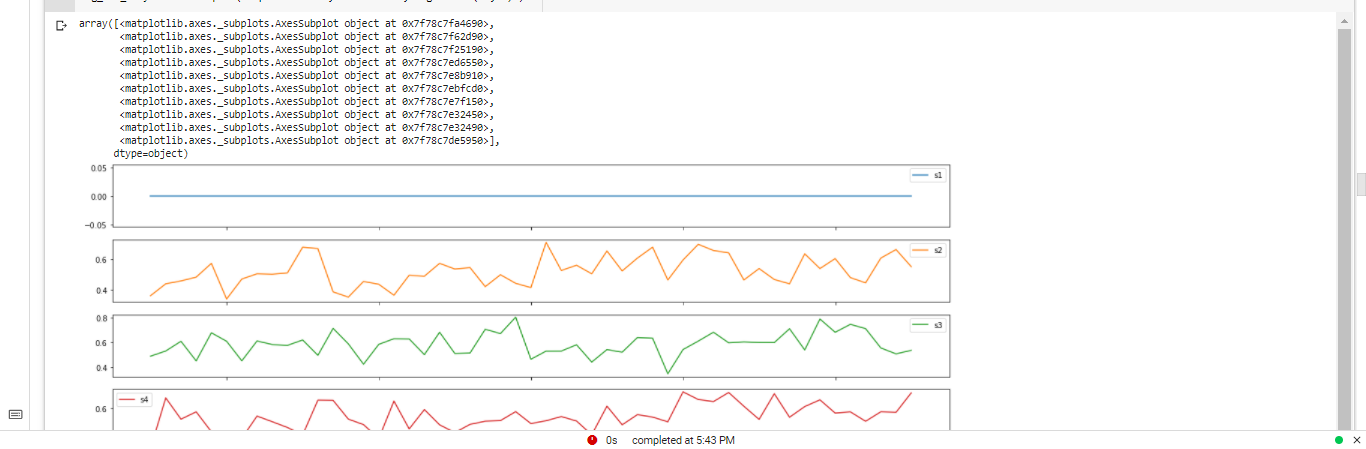


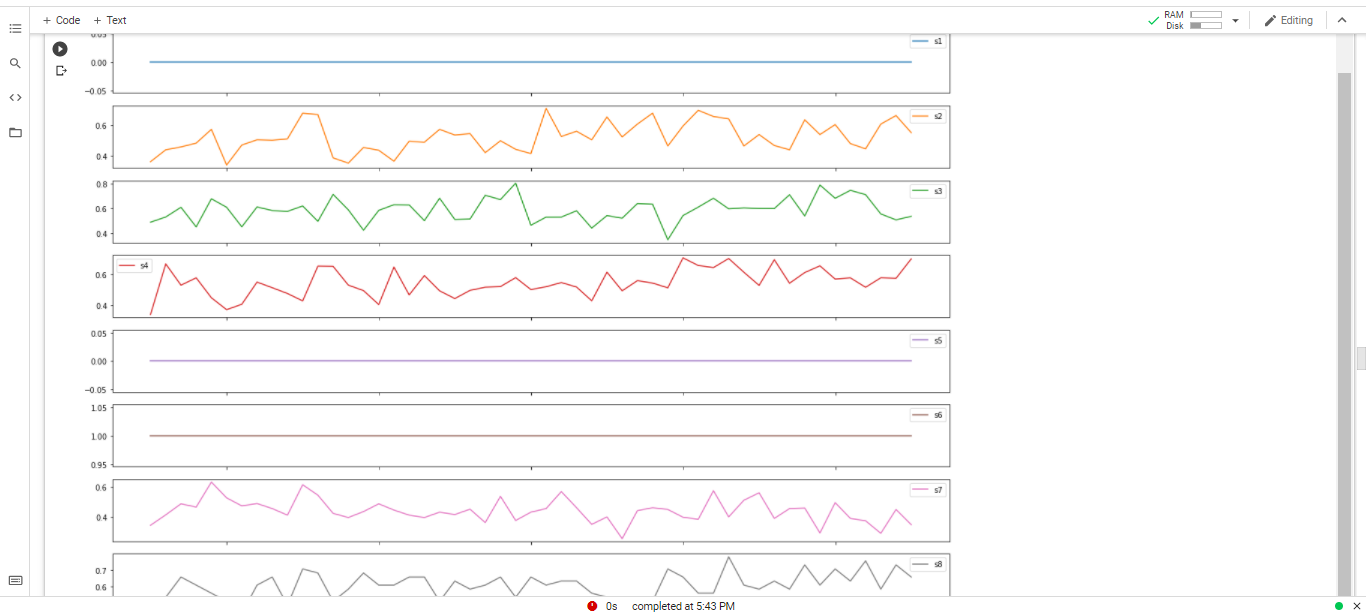


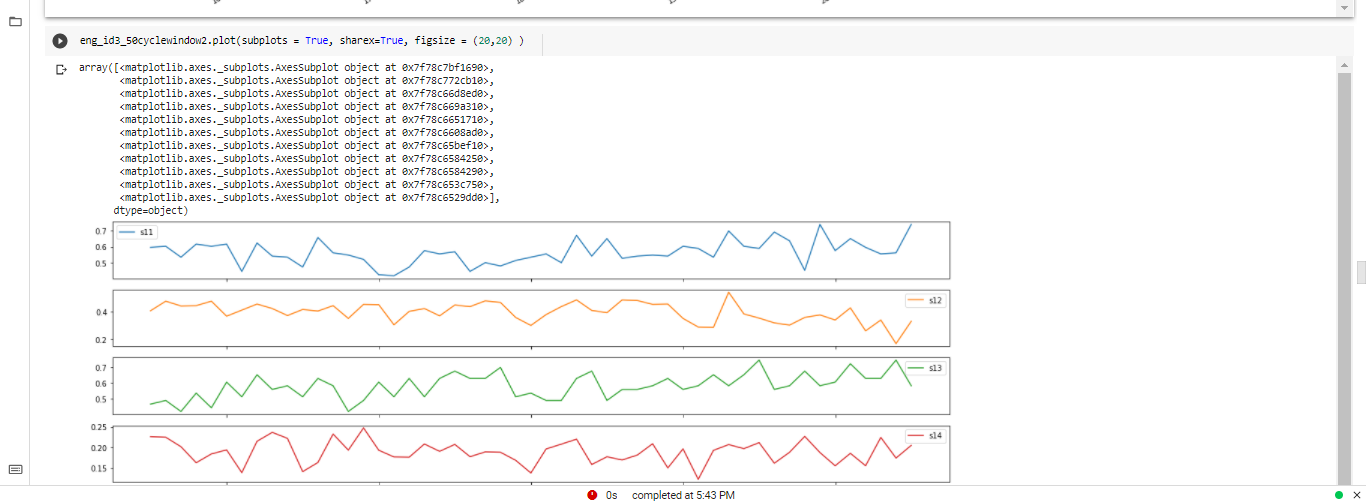
Task 5:







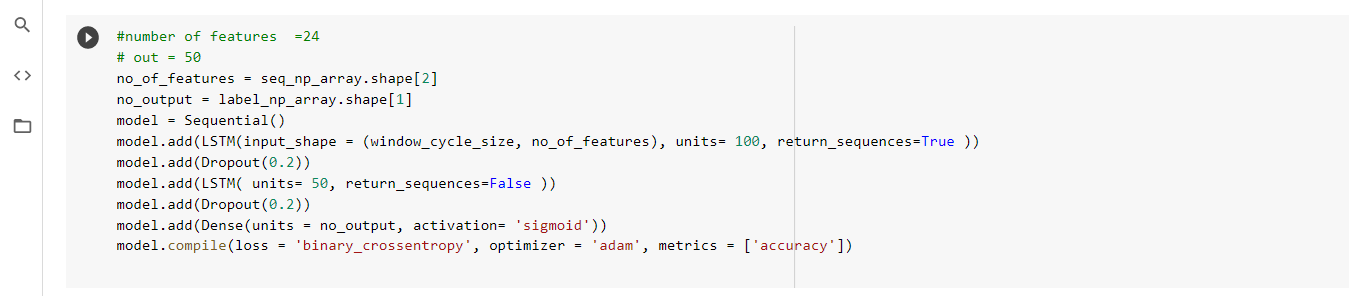


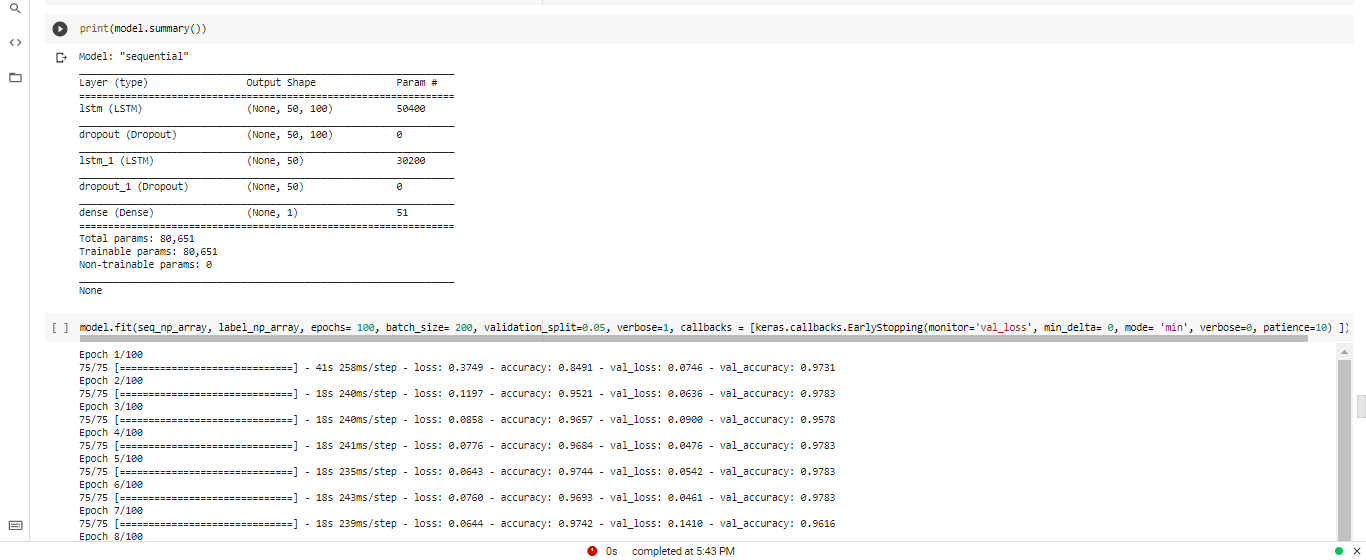


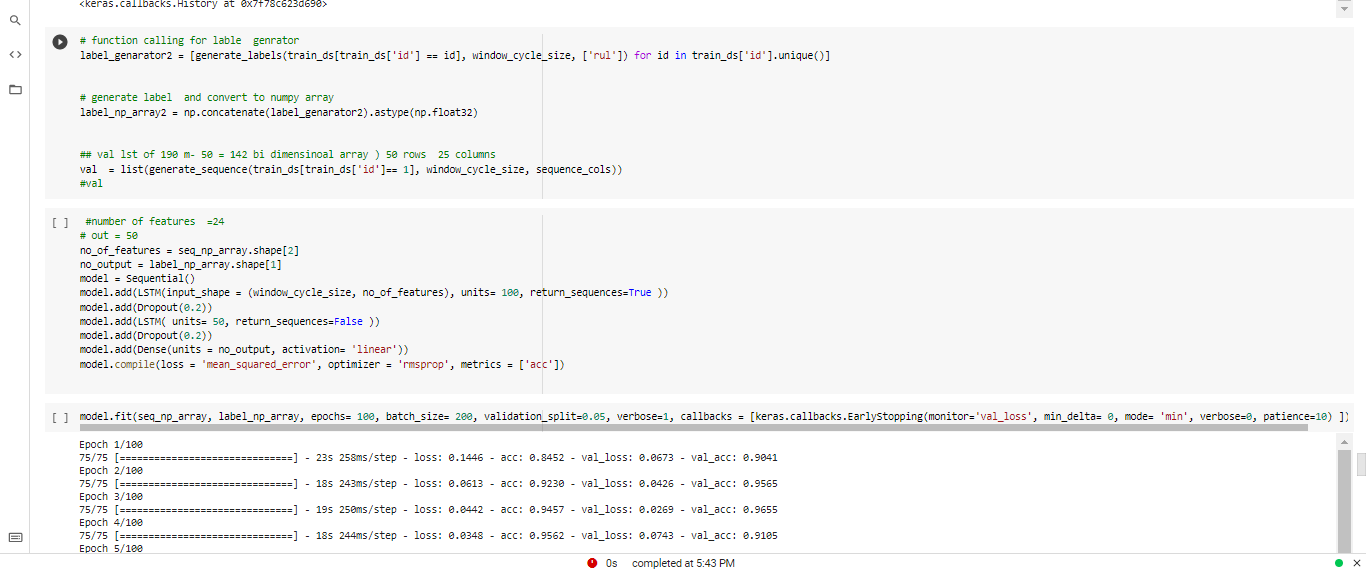


Task 6:

How to model data







Task 7

How to: Train and evaluate models

