# Dataset

In the proposed study we used the climate change dataset of 2016. We downloaded the dataset from Kaggle that was based on 420452 samples and 15 features. The first feature of the dataset was the date when the sample was collected. The downloaded dataset was a time series dataset that have the climate information with the interval of 10 minutes. In other words, the climate condition after every 10 minutes is recorded in the downloaded dataset. The climate information is also recorded in the dataset on daily basis from 2009 to 2016.

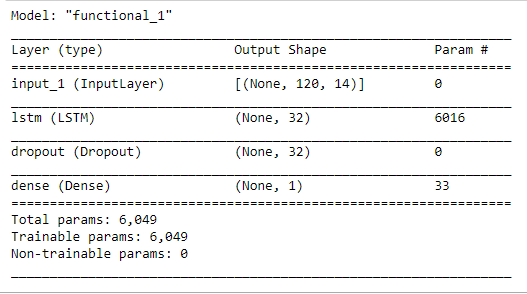
# Preprocessing

In the preprocessing phase, firstly we clean the dataset by removing the NAN and null values. To remove the NAN and null value, we replace the NAN and null values by the mean value of that column. As the dataset contain the negative and very diverse values in many features, we normalize the dataset to best fil for the proposed model training. To normalize the dataset, we subtract each value by the mean value of that feature. Further, we divide the subtracted value by the standard deviation value of that column.

# Train Test Split

After the preprocessing of the dataset, we split our dataset into different set for training, testing and validation of the model. To split the dataset into different subsets, we used the indices split technique. By using this technique, we split the dataset in training, testing and validation set with the ratio of 50%, 25% and 25% respectively. After the split of the dataset, training set contain 210,266 samples while the testing and validation set contain the 105,113 samples.

# Models

For the prediction of temperature using time series data, we used the two RNN models with LSTM and GRU. The first model (LSTM Model) was based on the simple input layer, LSTM layer and followed by the dropout layer and output layer. The dropout rate in LSTM model was set to 0.5 to train the robust model. As the proposed problem is regression problem that’s why the output layer contain the single neuron for prediction. The architecture of the LSTM model is also shown in Figure 1.

The second model (GRU Model) was based on input layer, two GRU layers followed by the dropout layer and output layer. The size of the neuron is 32 on both GRU layers with the dropout rate of 0.5 at dropout layer. Lastly, the layer was based on the dense layer with the shape of 1 for temperature prediction.

# Evaluation Measure

Different evaluation measures are available for classification and regression problems to measure the performance of the trained models. We used well known Mean absolute error to measure the performance of the trained models. Mean absolute error takes the average of absolute error that calculated by the absolute difference between predicted value and real value. We calculated the mean absolute error for our trained models by using the below formula:

# Experiment and Result

Firstly, we trained the LSTM model on 50 epochs with RMSPROP optimizer and MSE loss function. We trained the LSTM on training set and validate the LSTM on validation set. After the complete training of the model, we used the testing set to evaluate the performance of the model. We calculated the MAE Score for the 105,113 testing samples. The LSTM trained model showed the 2.56 MAE score for testing samples.

Secondly, we trained the GRU based RNN model for the prediction of temperature on time series data. GRU model was also trained on 50 epochs with same loss function and optimizer. The GRU model was also tested on test set to measure the performance of the model. We calculated the 2.40 MAE score of GRU model on 105,113 testing samples.

After analyzing the performance of the both models, we see that GRU based RNN model had the lower MAE that showed that GRU model is more accurate to predict the temperature of upcoming days.