AI Project Report

Submitted by:

Aravind N R

B230194EE

# Introduction

**TOPIC: Power consumption forecasting using RNN for individual household**

The AI model deals with the energy consumption forecasting using past user data. The model can be used to regulate the energy given to a power grid according to user data collected for optimized working of the grid and to reduce the losses in the grid by injecting only the required amount of power into the grid and helping to regulate the current flow. Moreover, the cost can also be optimized and the penalty calculation can be optimized using this model.

The main objective of this project is to estimate the average power required to be injected into the grid so that the consumers are able to get enough power supply and also to decrease the amount of excess power being supplied to the grid.

# Theory/Problem Statement

The given dataset contains the power consumption and meter readings for different times during a continuous time period for about 4 years. Using the data visualisation (given in the dataset description section) we can understand that the required model to be used for this data must either be a regression type, clustering type or a neural network type. Of these, the neural network is by far the most efficient model since we are able to include multiple hidden layers for getting the best results. Here, we have used LSTM (Long-short term memory) and time series analysis for forecasting the power using RNN (Recurrent Neural Networks).

RNN is mainly used for the classification of sequential (continuous) data. The difference between RNN and other neural networks is that it is able to remember the older data information provided to it and is able to find out patterns over a large area. The hidden state of RNN allows for storing the pattern details in memory and to update said patterns so as to form a system which allows for better prediction models. RNN is basically a sub-class of LSTM which refers to the ability of the model to remember and store the information given to it for later usage.

The model also used some regular power and energy calculations used normally in electrical engineering field for designing the grid. This is done in order to get the power consumed in each line and to determine the power injection to those lines.

All these techniques are used to identify the power consumption statistics for the given household and optimize the generation and minimize the losses due to overcurrent generation in the given grid. This helps both the producers and consumers by decreasing the cost for production and consumption.

# Dataset Description

Name of dataset: Individual household electric power consumption

Link: <https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption>

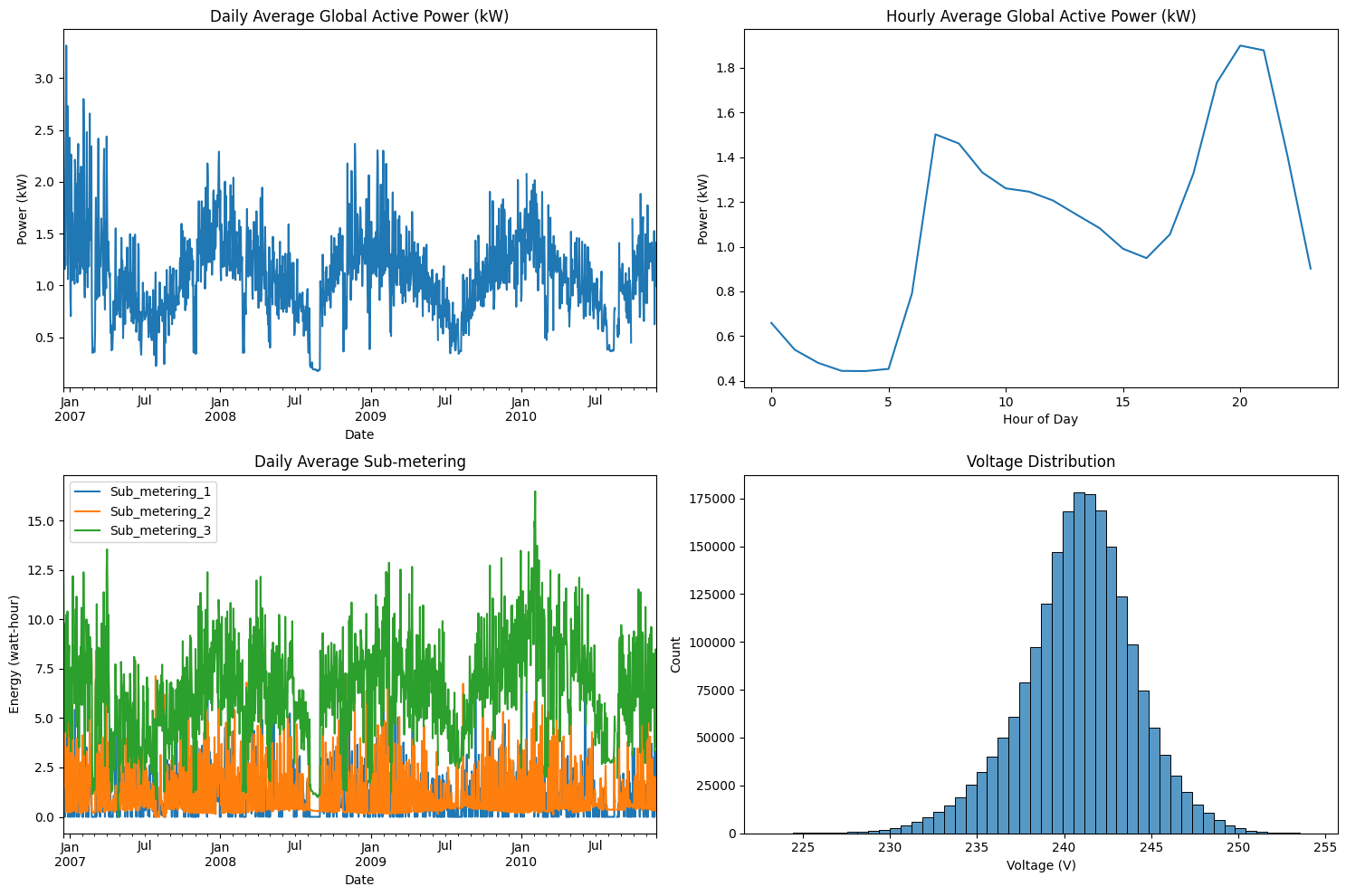
Source: UCI repository

The dataset has 9 attributes:

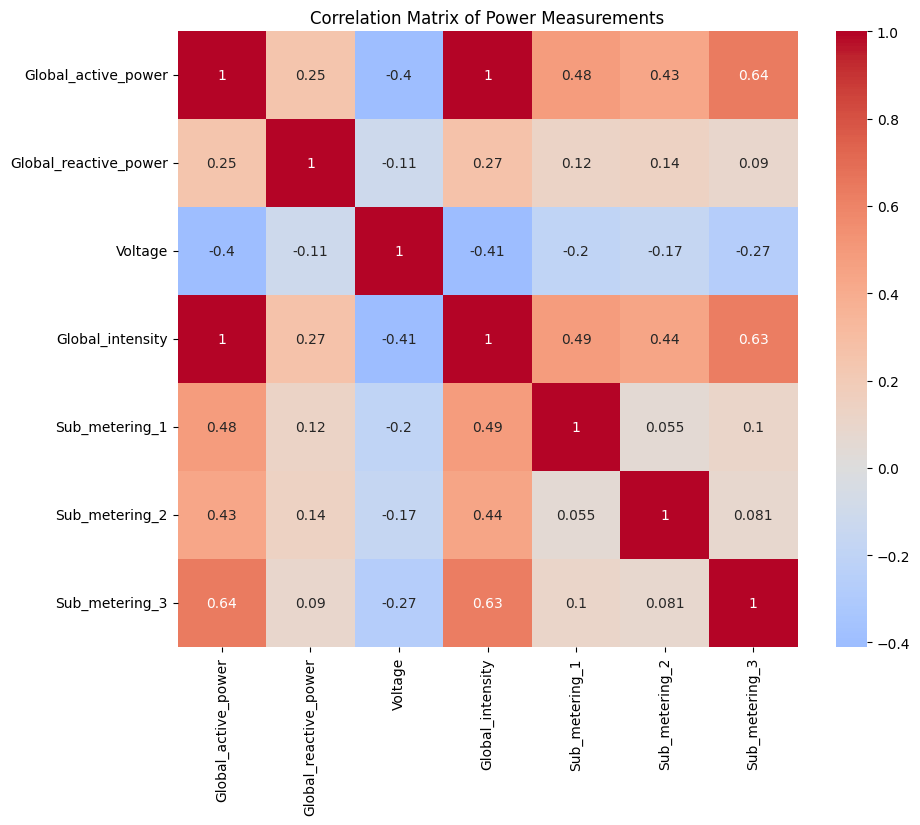
1. Date: The date for which the power consumption data was taken in DD/MM/YYYY format
2. Time: The time at which the data was recorded in HH:MM:SS format
3. Global Active power: The total amount of active power that is being used up in the household at the given date and time expressed in kW
4. Global reactive power: The total amount of reactive power being used up in the household for the given data and time expressed in kVAR
5. Voltage: The voltage measured given in volts
6. Global intensity: This gives the total current intensity in amperes (to show the current flow )
7. Submetering-1: This gives the energy consumption in the kitchen area of the household in Wh.
8. Submetering-2: This gives the energy consumption in the laundry room of the house in Wh.
9. Submetering-3: This gives the energy consumed by water heater and air conditioner in Wh

In total, the dataset has 2,075,259 measurements of electricity usage from a single household in Sceaux, France, collected at one-minute intervals from December of 2006 to November of 2010. About 1.25 percent of the total data is missing in the dataset which is characterised by blank spaces.

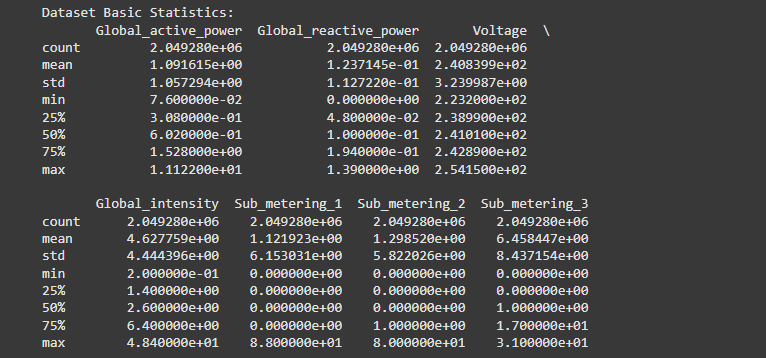
The visualisation graphs are given below



The correlation matrix is:



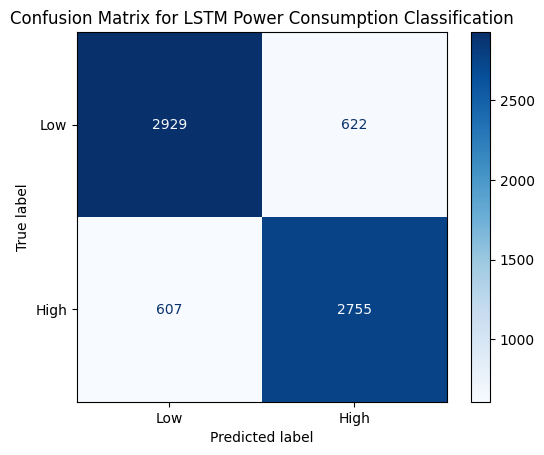
Statistics for the dataset



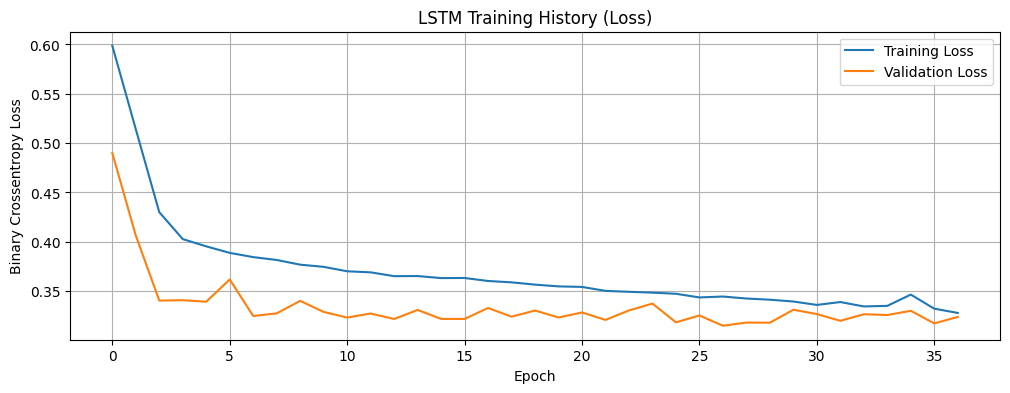
# Results

The power consumption data was trained using the 80 percent of the dataset and the tested using the rest 20 percent of the data. Of the training data, about 10 percent of it was used for validating the data.

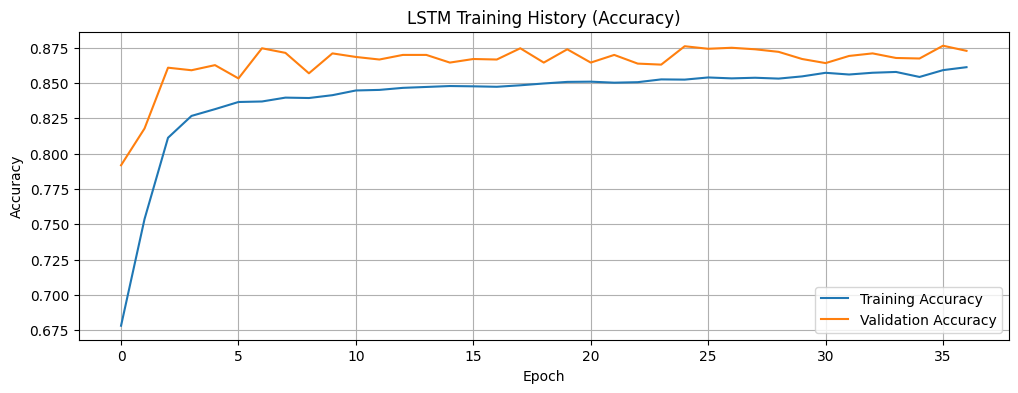
The confusion matrix for the model is given below:



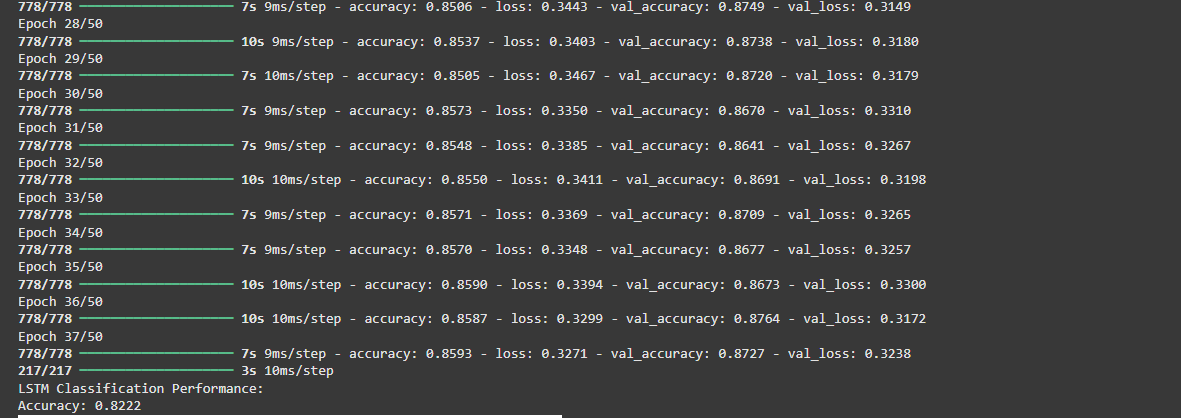
The loss vs epoch graph is given which shows the amount of value loss which occurred while the data was being trained. From the graph, it is clear that the loss was higher at first due to lesser data being available for the model and then it decreased as it is able to find the pattern of power consumption.



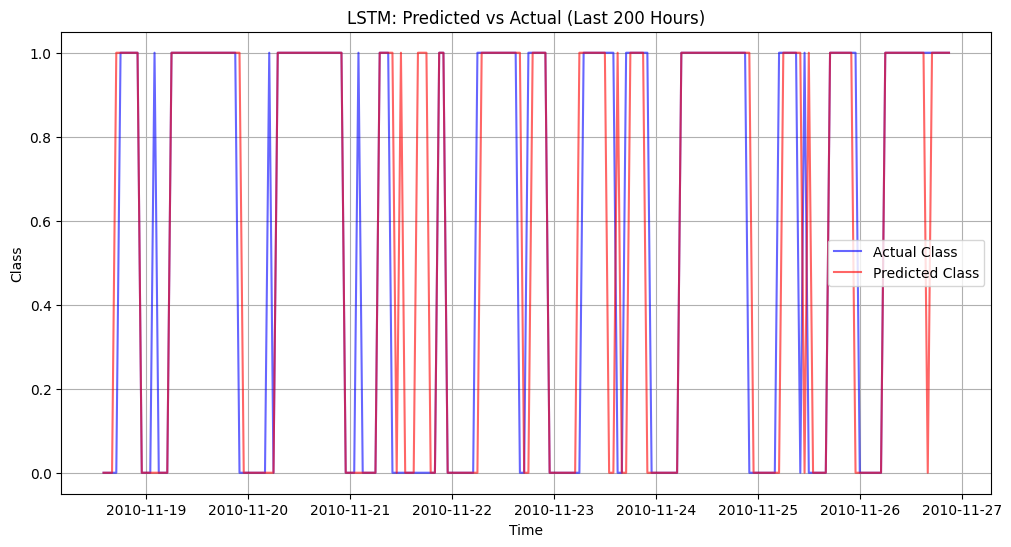
The exact same principle is true for the accuracy plot which starts off very low but increases as the training goes on since it is able to get better results by predicting the pattern.



The mean accuracy obtained for the model after training is 82.22% and the maximum accuracy obtained during training is 85.93%



Finally, the hourly prediction graph is shown which shows that actual vs true predicted values for this model.



Some of the error can be due to the fact that the dataset used for this model does not have 1.25 percent of the data. Even though, cleaning process has been done on the data, some stray data points could cause error for the model.

# Conclusion

The model using RNN is able to predict the output of the household with an accuracy of 82.22 percent. After taking into consideration the amount of data being used and the error in some data values of the dataset the accuracy for this system can be considered to be optimal for. With a bigger dataset containing more accurate data, the accuracy might be able to increase to 90 or 94 percentage which is within the permissible limit for a power grid.

The model is able to predict the output correctly is most of the cases and all the parameters can be calculated using this data.

# Code

The code is given in the url: <https://colab.research.google.com/drive/1GJwhUor9jezt_0ICSdhDC9qPH14s1hx_?usp=sharing>