OPTUNA OPTIMIZATION BASED CNN-LSTM MODEL FOR PREDICTING ELECTRIC POWER ENERGY CONSUMPTION

MSc Research Project
MSc DATA ANALYTICS

IBIDOKUN EKUNDAYO
Student ID: X17156068

School of Computing
National College of Ireland

Supervisor: Dr Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	IBIDOKUN OLAOYE EKUNDAYO		
Student ID:	X17156068		
Programme:	MSC DATA ANALYTICS	**Year:**	2019/2020
Module:	RESEARCH PROJECT		
Supervisor:	CHRISTIAN HORN		
Submission Due Date:	17/08/2020		
Project Title:	… **OPTUNA Optimization-based CNN-LSTM Networks for Predicting Energy Consumption**		
Word Count: 7350	**Page Count:** …24……………………		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: IBIDOKUN EKUNDAYO…………

Date: 17/08/2020………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

OPTUNA OPTIMIZATION BASED CNN-LSTM MODEL FOR PREDICTING ELECTRIC POWER ENERGY CONSUMPTION

IBIDOKUN EKUNDAYO
X17156068

Abstract

Forecasting residential energy consumption using deep neural networks has been attempted in past researches. Typically, optimizing these networks relies on the operator's prior knowledge. They are also affected by the size of the search space and the tuning parameters for the model. In this research, we integrate OPTUNA, an optimization algorithm that can automatically determine different frames of hyperparameters to tune a CNN-LSTM neural network for predicting electric power energy consumption. Our research findings indicate that the proposed model can be applied as a potential alternative forecasting framework for better forecast accuracy and a broader generalization potential. The OPTUNA hyperparameter can remove mutation operations and crossover in comparison to the traditional models. To validate the potential of our proposed algorithm, we have selected the household electric power consumption dataset from the UCI machine learning public repository. Our proposed OPTUNA CNN-LSTM model explores varying degrees of optimum forecasting frameworks automatically and presented a lower MSE compared to a result of past literature that implemented the particle swarm algorithm.

1 Introduction

The consumer demand for power supply is ever increasing on a global scale due to wide scale use of machinery including electric powered large-scale machineries, technological advancement such as cloud infrastructures powered with electricity. An important phase of everyday living is a constant access to uninterrupted power which may be better managed if the electric grids are can be networked (Li et al., 2017). Comparably, gradual integration of smart meter infrastructure on a global scale has set the course for the establishment of smart grids with active power energy systems inbuilt in them (Kong et al., 2018). This implementation has carved out a niche for electrical energy forecasting on a short or long term, as it is also very necessary for efficient utilisation and management of power grids, especially for household users (Moriarty & Honnery, 2019).

Demand for electric power is impacted by some factors including weather conditions, consumer attitude and how they utilise their respective electrical household appliances. Therefore, it is necessary to consider the need to discover and develop new frameworks for a well-planned and efficient management of power consumption in residential and industrial structures that will control and manage energy demands. Different energy prediction models have been implemented over the years in the electric power domain to forecast electricity demand over hours, days, weeks, months and years so as to supply the right amount of power into electric grids improve efficient use and management of electrical grid. With this view in mind, different researchers have built forecast models to maximise the use of electrical energy and enhance power grid quality. In Khosravani et al., (2016), this research implemented a support vector machine model to run a regression analysis to forecast electricity consumption of a residence. In Ullah et al., (2018), this research implemented the adaptive neuro-fuzzy inference system (ANFIS) to forecast electricity power usage.

Recently, the deep learning framework has been proven to be more efficient as a predictive model than classical prediction algorithms because it can easily learn non-linear patterns and map non-linear functions. Many neural network models have been proposed to forecast the consumption of energy, but for practical applications, there must be an adequate network structure. Methods to look for the best fitting hyperparameters are non-trivial and complex. Also, even though neural networks can map non-linear patterns in energy consumption attributes, they may suffer from vanishing gradient problems thus loosing past information. Therefore, as the data size grows larger over time, classical deep learning networks may suffer overfitting. Currently, conventional neural networks have been further advanced by recurrent neural network and convolutional neural network. CNN extracts feature while learning the weights of attribute maps composed of layers and retains the connections between different attributes (Ronao & Cho, 2016). In similar fashion, recurrent neural networks have a hidden memory in which it stores dynamic features (Greff et al., 2017). In the prediction of energy consumption, CNN has the ability to extract input features wrapped in temporal contexts so as to reduce the dimension of the variables. The long short-term memory network (LSTM) has the ability to memorise past information in time series data over time and maps the complex non-linear energy consumption patterns. The combination of this two models CNN-LSTM architecture converts multidimensional space to predict a time series related power consumption. However, there may be a need to adjust the hyperparameters as a result of this combination.

This research therefore investigates the extent of improvement a new hyperparameter tuning algorithm OPTUNA can bring when used to optimise a CNN-LSTM approach to household power consumption prediction. Optuna is an open-source hyperparameter tuning optimization developed in 2019 which has the extensive capacity to automate the fit-trial-and-error-calculation process of hyperparameters optimization (Akiba et al., 2019). Optuna finds the best or optimal hyperparameter values using a targeted approach which can be defined in an API style. This is the first research to adopt the Optuna hyperparameter framework to in this domain of research. The objectives identified to answer the research question are;

- Carry out a series of data visualisations to understand the seasonal trends in the data.
- Investigate the effect of seasonality on the results using statistical methods such as SARIMA.
- Implement neural network backed methods to improve results.
- Evaluate the models using RMSE and further optimise the best performing approach using the OPTUNA framework.

The major contribution of this research is the use of the novel OPTUNA hyperparameter optimisation algorithm to improve the CNN-LSTM approach to prediction of household power consumption and this is the first time OPTUNA will be used in this domain.

The rest of this research is presented as follows. Section 2 discusses related work while sections 3 and 4 present the proposed methodology and implementation of the research. Sections 5 and 6 present the results and conclusion respectively.

2 Related Work

This section reviews works from past and recent relevant literature in the time series, deep neural networks and machine learning domain. Research works that have experimented power load forecasts using various procedure and models. This research review focuses on the CNN-LSTM framework and traditional time series model, the SARIMA model.

2.1 Statistical time series methods

In the study of time series and load prediction, one of the most widely used methods in past literature is Auto Regressive Integrated Moving Average (ARIMA) due to the effectiveness of the model on historical data (Mirjat et al., 2017). An autoregressive integrated moving average is a statistical time series model used t forecast future trends and have a better understanding of the time series datset. Even though the ARIMA model is known to have efficient forecast capabilities, it only works best with datasets that are stationary in nature.

Erdogdu (2007), applied the ARIMA model, Using the co-integration of the demand for electric power in Turkey and the result gained from this framework was evaluated and the result showed link with the actual data with a less tolerance score.

To forecast the electric power energy peak load in Saudi Arabia, El Desouky and Elkateb (2000) adopted ARIMA and ANN. Their research proposed both algorithms to decrease prediction errors and their results outcome showed lesser forecast errors. An ARIMA model used to perform an electricity consumption forecast in China was proposed by Yuan et al., (2016), their research applied a vector regression framework to decompose the seasonality of the data, results inferred that the model is effective for time series forecast.

To predict the outflow of electric power energy load in a hydroelectric plant, Cassiano et al., (2016) presented an ARIMA model and integrated principal compenet analysis associated with hierarchical clustering to reduce dimensionality in their data. To forecast electric power energy peak load prediction for seven buildings in a Thailand university, Kandananond (2019) used Auto-regressive integrated moving average (ARIMA). The ARIMA was applied to map the time series data to show the extra information of energy usage of the international students living In the building. This research. Also presented an Auto-regressive model (ARX) with exogenous output, their research adopted a very limited data ARX outperformed ARIMA but the constraint of a limited data may question the validity of their model. Other research has constructed a model which merged ARIMA with an exponential smoothing model (ESM) with seasonality to predict electric power energy peak load for an electric provider with unstable and dynamic growth (Barakat et al. 1992). Other researches also developed similar models have been developed to predict short term demand for electricity using minutely demand observation from a British related data (Taylor 2008). Fernández et al., (2011) presented a comparative analysis of ARIMA, ANN, SVM, multiple linear regression models to predict the short term electric peak load for a residential building.

2.2 Traditional Machine Learning Methods

In the modelling of electric power consumption prediction, machine learning methods have remained popular over the years. To mention a few, Wang et al., (2018) implemented a random forest model to forecast the usage of electrical energy for two buildings in Florida over a very short period of time (hourly consumption of electricity). Candanedo et al., (2017) experimented a comparative analysis with different models including support vector machine, random forest, gradient boosting algorithm and multiple linear regression to forecast energy consumption, the research findings concluded that Gradient boosting algorithm performed best as compared to other models. In some extant literature, conclusions that electricity consumption is impacted by human behaviour has been documented. For example one of these researchers have adopted support vector regression (SVR) to correctly forecast the electrical energy consumption for an office building in China (Wang and Ding, 2015), and also another paper documented a multiple regression machine learning framework which was used to forecast a residential buildings daily electrical energy usage (Fumo and Biswas 2015). Furthermore, Cai et al., (2019) used the data collected from 16000 residential homes from an entire region to correctly classify electric power consumption rates of residents in this region. Their research recorded the patterns of electrical consumption by applying data mining models and also adopted particle swarm k-means algorithm, to perform clustering. Also, after clustering the electricity usage patterns were split using the centres of recorded clusters, and support vector machine was applied for classification purpose.

2.3 Deep neural network methods

2.3.1 Long Short-Term Memory

Long short-term memory (LSTM) is an extension to the recurrent neural network according to Shaikh et al., (2019) is used to learn context from input sequences. They have been shown to extract and map non-linear patterns within data (Liu et al., 2017). Gugulothu and Subramanian (2019) in their work used two LSTM based algorithms (sparse ED and sparse LSTM) to determine electricity consumption. In their approach, they used a fully connected feed-forward layer that performs feature selection and dimensionality reduction of the input data. The LASSO sparsity constraint is used to introduce sparsity on the weights of the feed-forward layer. In the work conducted by Wang et al. (2016), an electric load forecasting model with 5 hidden layers of LSTM neural network was presented. In this model, to combat overfitting, dropout was added after each LSTM layer before a dense layer was used as an output layer. Similar research conducted by Yuan et al. (2019) introduced a stacked LSTM approach to forecast future load values for power generation. In their paper, they implemented a probability density prediction method which provided more information for the LSTM model to make more accurate predictions.

Kim et al. (2018) conducted further research proposing a short-term electricity load forecast created a daily consumption sequence using an LSTM. In this approach, each daily consumption pattern is converted into a representative form. This is then used as an input to pass a vector implant into the model and results showed 82.5 performance on testing data. They implemented the approach without consideration for missing values which may have posed a challenge on the results due to inconsistent time steps within the data sequence. Ermatita et al. (2019) further showed the use of parameters in the LSTM layers to train the forecasting model effectively. Scaled and normalised historical data from an Indonesian substation was used in the implementation of their research and their results indicate a high accuracy for determining consumption for the week ahead.

Wang et al. (2019) implemented an LSTM backed electricity load forecast model for a large building in Alaska. In their model architecture, they used five-layers and a min-max normalization to scale the values. Their model outperformed traditional machine learning approaches such as support vector machines and random forest when the results were compared. However successful, the did not take into consideration the lagged electricity consumption data as this might have further improved performance. Ud Din et al. (2019) used an LSTM architecture to generate 24 hours forecasts for independent bi-variate time series data. Their work also implemented an LSTM forecasting performance for deployed with different weight optimizing methods including adaptive moment estimation, stochastic gradient descent and root mean square propagation. Exogenous features alongside data features from the collected time series were used.

2.3.2 CNN-LSTM

Kim and Cho (2019) proposed a CNN LSTM framework for the prediction of power consumption, their model introduced a hyper parameter selection model particle swarm optimization algorithm to optimise the number of kernels and units in the CNN and LSTM layers respectively. Their result showed an improved root mean square error when evaluated against contemporary time series models. Similar research done by Alejandro (2020) proposed a mix architecture of convolutional neural network and artificial neural network combinations to predict energy of consumption for a French grid. The model was trained and then utilized in a real setting to provide a French energy demand prediction. The final result of experiment showed that their approach outperforms the ARIMA model it was benchmarked with using the mean absolute error

2.5 Hyper-Parameter Optimization

Depending on the dataset, electric power consumption forecast can be experimented with various neural network framework and learning process. the learning process in neural network applies a set of hyper-parameters so that adjustments can be done to tune the model performance depending on the nature of the problem. So, we may have different neural network frameworks just by tuning hyper-parameters for a particular problem and to find the best parametric combination for optimum results. The procedure of investigating hyper-parameters demands a manual engineering process which may be non-trivial and may take a lot of time. As the possibility to build the frameworks of neural networks expands, prevailing manual processes are being replaced by search algorithms. Traditional frameworks for finding hyper parameters are random search and grid search (Li et al., 2017 and Gonzalez et al., 2005). Grid search models is a machine learning method that is used for hyper parameter adjustment, this process seeks for parameter combinations within a specified range by a grid, and then systematically evaluates the neural network model. The notion of a grid search is based on a priori knowledge which determines the scope of the grid and analysing the problem, this makes it possible to search effectively. Howbeit, there is a risk that the optimal hyper-parametric combinations may be outside of the range specified by the grid. To tackle the risk of grid search, random Search as oppose to the grid search and its range is much broader to explore hyperparameters. Random search supplies statistical distributions that can randomly samples discrete values for each specified hyperparameter. When compared to the grid search, random search within a predefined range can efficiently search for the best combinations of hyperparameters. Howbeit, even though statistical notion of random sampling is proposed to be non-bias, sample selections bias may happen.

In deep neural networks, there are a number of parameters which can be tuned such as learning rate, batch size, and number of training iteration, activation function and optimizer algorithm in general. In addition, for special architectures such as CNN or LSTM, there are ore hyperparameters which can be tuned, such as in CNN, number of filters in convolutional layer, the size of the filter and the stride are common hyper-parameters. In LSTM, the number of units in the layer, whether to return sequence or not, are common hyper-parameters which should be tuned. They are not, however, just numerical values but also categorical values such as the optimizer could be a Momentum SGD or Adam. It is almost impractical to make a machine learning model do a task without tuning hyperparameters. The amount of hyperparameters tends to be high, especially in deep learning, and it is believed that the performance of neural network model relies on how the model is tuned. In this research, we propose to use OPTUNA, which is a next level hyperparameter tuning framework for parameter optimization introduced by Akiba et al., (2019). The Optuna framework is efficient for both pruning and searching procedures. This research thus proposes a novel OPTUNA CNN-LSTM model.

3 Research Methodology

This research methodology is based upon related and relevant existing literature. Therefore, some notions that have been investigated by past literature will be applied for the course of the research and the reasons for applying them will be listed therein.

This paper adapts the Knowledge discovery in databases (KDD) approach is applied. This approach outlines several stages for conducting research. These are data selection, processing and transformation, data mining, data Interpretation, data modelling and model selection and evaluation as seen in figure 1.

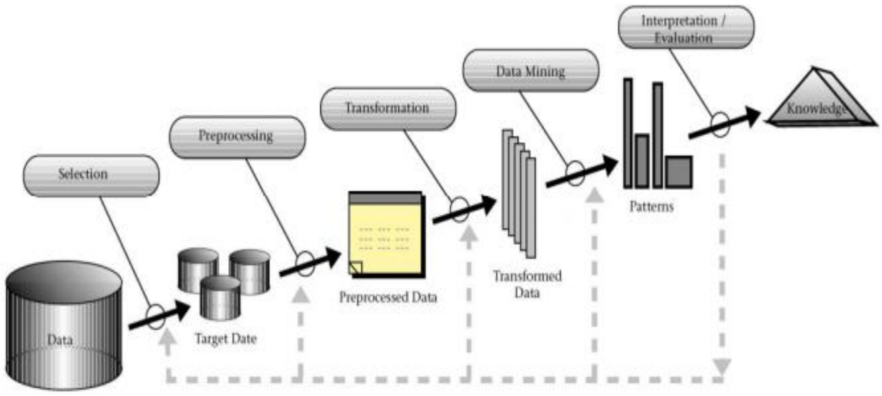

Figure 1: Steps of the proposed KDD process

3.1 Dataset

The dataset adopted for this research is an individual household electricity consumption data collected from the UCI Machine Learning Library[1]. The measurements in the dataset are collated from the smart meter readings from a residential building situated in France and it spans 47 months of minutely recorded meter readings from December 2006 to November 2010. It constitutes a multivariate time-series of electricity related variables, that may be adopted for the prediction of future electricity consumption and is a standard dataset that has been used in reviewed literature (Nishida et. al., 2017). The dataset contains 2,075,259 instances and has 9 attributes that can be leveraged to predict future electricity consumption.

3.3 Data Processing

To get the best out of the chosen dataset, it is necessary to clean and transform data into the right format. Outlined below are the steps taken to clean dataset and ensure optimum model performance. The data processing steps are outlined in the following sections.

3.3.1 Missing Values and Outliers

Outliers and missing values can impact the integrity of the research outcome and the selected dataset does appear to have missing values. There is about 2-3 days' worth of missing data values around 28th of April 2007. Critically investigating the dataset, we infer that the information supplied in the dataset is associated to the active electric power energy and In the data transformation stage, we obtain a new columns which indicates the remaining of the active current not recorded in the listed submeters and the new column was named Sub_metering_4.

3.3.2 Data Visualization

After data cleaning, to gain further understanding of the data, visualizations such as line plots for each of the data variables to see the trend across the entire time period are done. Further visualizations are done to check the distribution of the data variables. Dickey Fuller test is also

[1] https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption

conducted to test for stationary status of the data. Further visualizations are done to check the distribution of the data variables.

3.4 Data Mining

For the data mining stage of the research, different modelling approaches were adopted to model the data and predict electricity consumption. The models adopted are discussed herein.

3.4.1 SARIMA

The Seasonal Autoregressive Integrated Moving Average extends the notion of ARIMA. Unlike ARIMA which is applicable to only stationary data. SARIMA Is applicable to seasonal and non- stationary data which is the kind of data used for this experiment. To implement SARIMA for the proposed experiment, we will require selecting hyperparameters for seasonal and trend factors of the data. Furthermore, new hyperparameters will also be examined to show autoregression (AR), Differencing (I), moving average (MA). to an ARIMA model, the trend factors such as the trend autoregression order, trend difference order and trend moving average order is configured for this model. Different from the ARIMA model, the proposed SARIMA model will further configured for seasonal elements in the selected dataset including merging a global holiday dataset to improve the mode accuracy to reveal the seasonality of the data.

Therefore, where the selected hyperparameters for the proposed SARIMA are indicated, the reopresentation of the hyperparameters for the proposed model are shown as

SARIMA (Trend factor) (seasonal factor) = SARIMA(p,d,q)(P, D, Q)m.

where the time steps (m) will have an effect on the seasonal Autoregressive order (P), seasonal difference (D) and seasonal moving average (Q). For instance, our research chooses a time of 7 which represents a weekly cycle.

3.4.2 CNN-LSTM Encoder decoder model

Long short-term memory (LSTM) neural networks have shown the potential to learn features from sequential data and map the non-linear pattern discovered within data.. The research proposes a hybrid model comparable to the Encoder-Decoder LSTM (ED-LSTM) model but the Encoder layer is replaced by a 1D-CNN layer. Our proposed model will replace the encoder layer with a CNN layer and the 1D-CNN layer can be fit into the framework by supplying each variable in the dataset to the layer as different one-dimensional time-series sequence of inputs. The layer then learns an internal pattern and representation for each sequence input that is then collectively interpreted by the decoder. This research refers to this hybrid model as a OPTUNA CNN-LSTM model.

To overcome the problem of overfitting, three regularization methods, L1, L2 and dropout regularization are experimented as part of the model architecture. In the deep learning framework, dropout can improve model performance and eliminate the activation values of randomly selected neurons during the process of training which amount in no overfitting and a more robust algorithm. This restriction on the training process lets the highly complex model to map more robust features instead of depending on the predictive potential of a small group of neurons in the network.

Further hyperparameter optimization to determine the hyperparameter settings for the proposed CNN-LSTM model to achieve optimum performance is explored using the Optuna framework. The advantages offered by the Optuna hyperparameter optimization include easy specification of length of optimization duration, easy integration of results, saving computational cost by pruning low performing trials early. Other advantages include ease of implementation and integration into various machine learning frameworks hence the choice was made to use it for this research.

3.5 Model Evaluation

This section explains the evaluation methods applied to measure the performance of the models built for this research purpose.

The proposed model will consist of values for a week ahead. Our research implements a multistep prediction procedure whereby model evaluates the days of the week separately to know which of the days the model predicts better by adopting a learning curve.

It will be helpful to understand how our model predicts each days of the week. Furthermore, we adopt the residual mean squared error which is a scale dependent error to evaluate the performance of our proposed model. The scale-dependent error has error scale at the same lvel to the data (Bauer et.al; 2019) and examples are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE).

4 Implementation

The implementation process of the proposed research is presented in detail in this section of the report.

4.1 Setup

First, the host computer is a MacBook Pro with 8GB RAM, core i5 processor and 256GB SSD.. The anaconda environment and Jupyter notebook is used to run the codes to implement this research. Other libraries used include Keras, sklearn and Pandas.

4.2 Dataset

The selected dataset was retrieved from UCI public repository. The data was separated using semicolons with each row representing a minutely smart-meter recorded data. The variables contained in this dataset are described as follows:

Table 1: Dataset variables

Variable	Description
Global active power	global averaged active power per minute (in kilowatt) for the household
Voltage	Average voltage per minute

Global reactive power	This is the global averaged reactive power per minute (in kilowatt) for the household
Sub_metering_1	Electricity consumption for kitchen appliances
Sub_metering_2	Electricity consumption for laundry room.
Sub_metering_3	Electricity consumption for climate control appliances

4.3 Data Handling

4.3.1 Data cleaning

The data was downloaded as a text file and then converted to csv file before processing. For the data cleaning process, few rows with missing dates were removed as they weren't significant enough to cause an impact on the performance of the model. Also, new columns were created for the date and time and sub_metering_4 which represented household usage appliances that weren't accounted for in the recorded sub_meters (e.g TV, Radio).

4.3.2 Data visualization

To have a clear view and understanding of a multivariate time series dataset of this nature, visual plots for each variable in the dataset over 4 years were created. These plots supply pictorial information about trend and seasonality as there are predicting models that may be incompatible with seasonal data, which this data appears to be. Also, the distribution of the data is revealed with the help of visuals.

For instance, the line plot created for each data variable revealed some information on sub metering 3 which is the climate control that may not directly with the weather conditions in the month of august. Perhaps, this home may have installed new systems. While, global active power displayed some wave of seasonal influence in the data visualization.

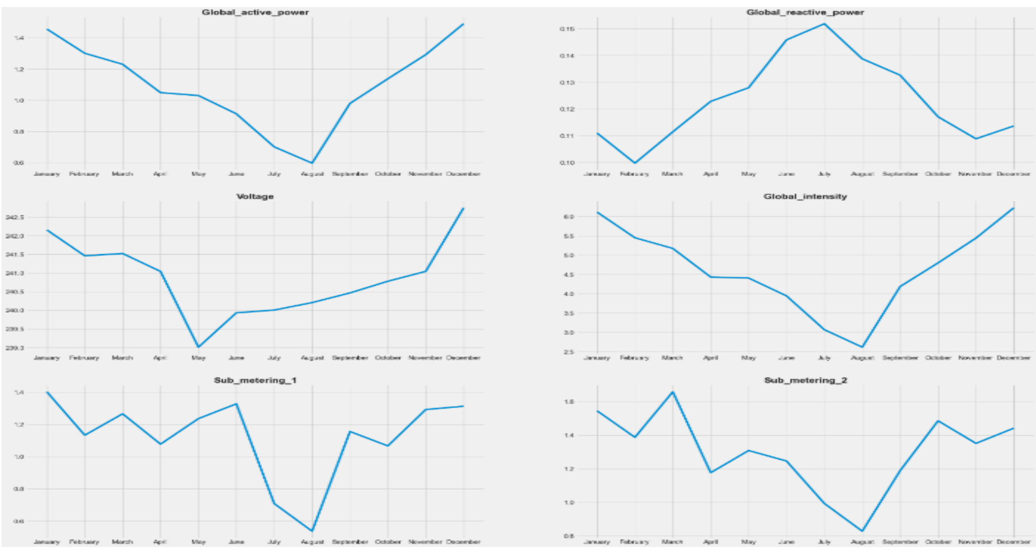

Figure 1: Average daily consumption each month in 4 years

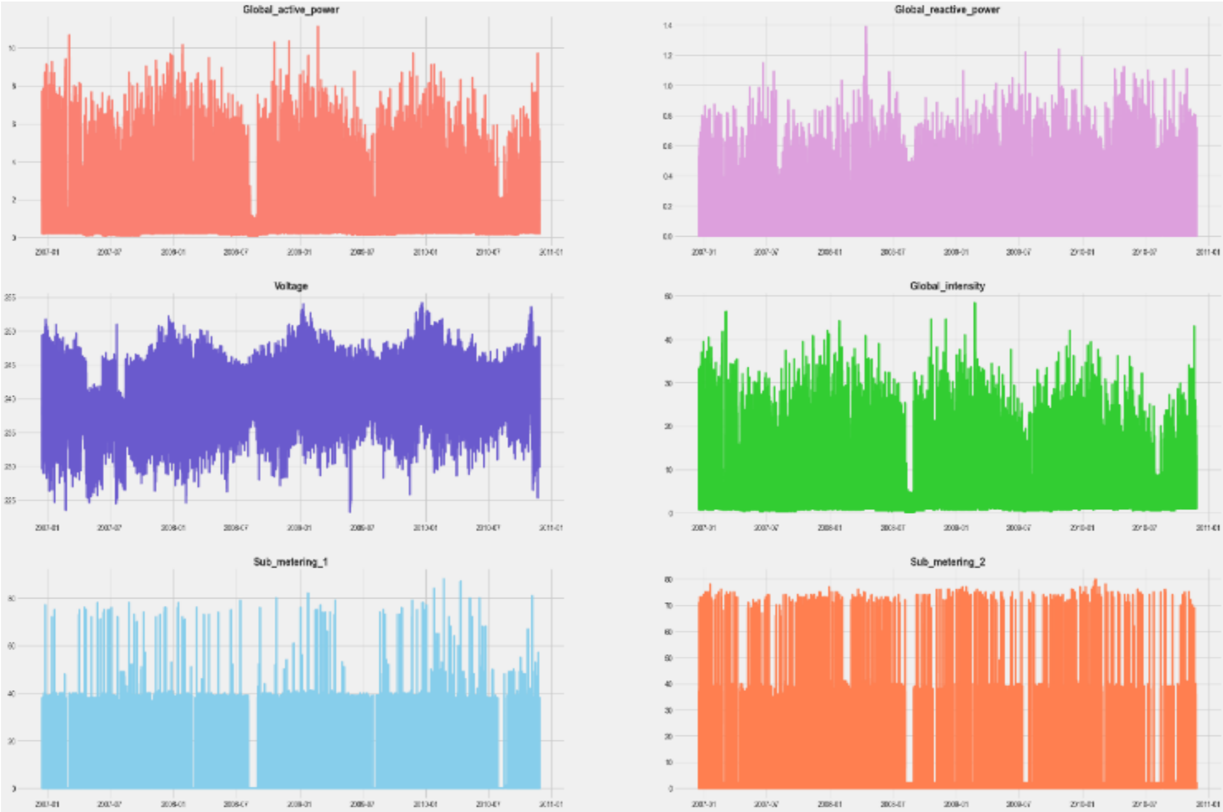

Figure 2: Average Consumption of each day in 4 years

Furthermore, density plots were created to check the distribution of each data variables and it was seen that only voltage was normally distributed while the rest were somewhat skewed as shown in the figure below. It was observed that the data distribution was skewed, and the data was scaled appropriately.

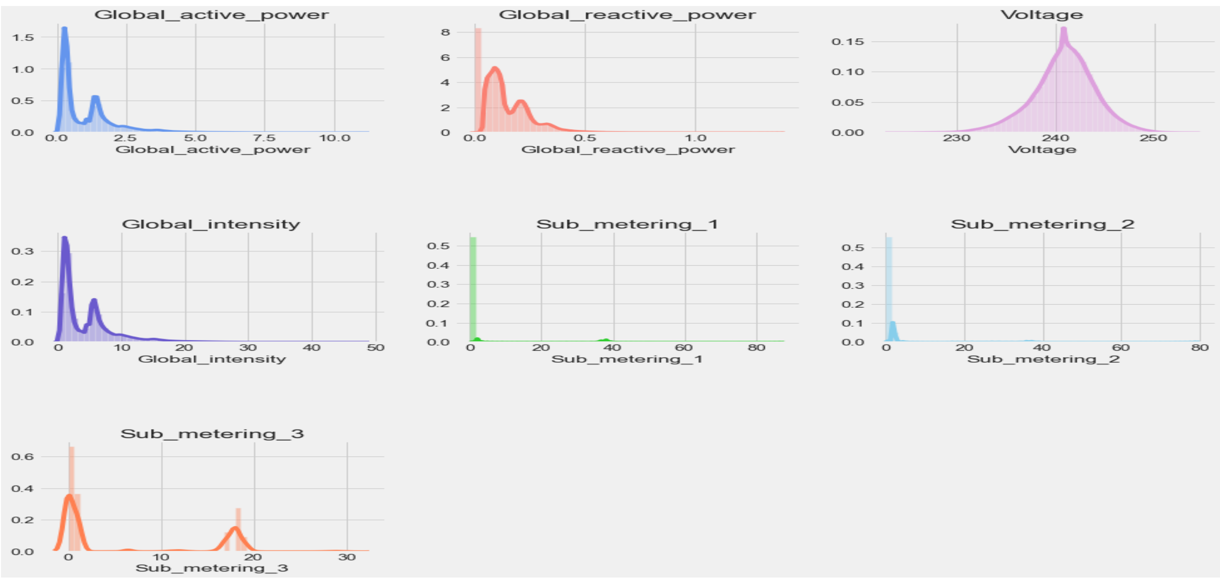

Figure 3: Data distribution

To gain further understanding of the data, we also looked at the annual periodicity in the data and It can be observed that annual periodicity exists for the Global Active Power.

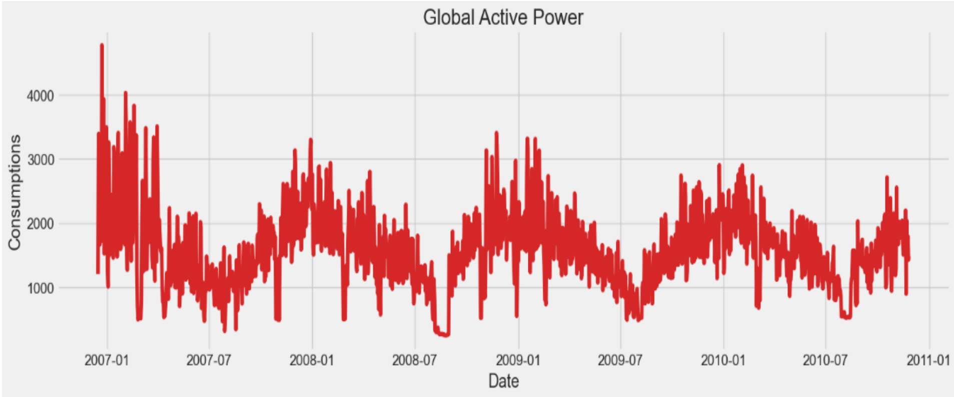

Figure 4: Global active power periodicity

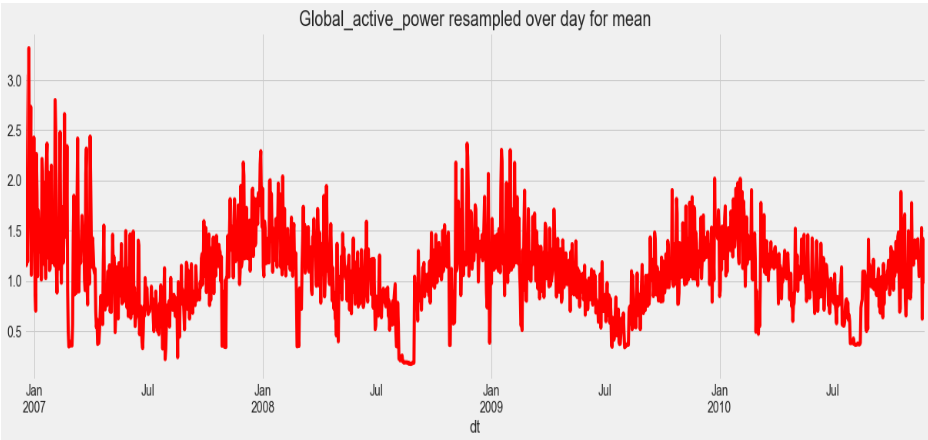

Figure 5: resampled global active power periodicity

4.4 Modelling

The main model implemented in this research is the encoder decoder 1D CNN-LSTM model aiming to predict household energy consumption. The encoder part of the model consisted of 2 1D CNN layers before a maxpooling layer. To help battle overfitting, a dropout layer is introduced after the maxpooling layer. The output vector from this layer is flattened before being passed to a repeat vector layer to repeat the vector across the entire time steps. The decoder LSTM is introduced at this stage before another dropout layer is used to further tackle potential overfitting. The model architecture culminates in double time distributed dense layers with a further dropout layer between them as is the case for time series modelling. The model architecture is illustrated in the figure below.

```
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_142 (Conv1D)          (None, 4, 128)            384
_________________________________________________________________
conv1d_143 (Conv1D)          (None, 3, 64)             16448
_________________________________________________________________
max_pooling1d_71 (MaxPooling (None, 3, 64)             0
_________________________________________________________________
dropout_6 (Dropout)          (None, 3, 64)             0
_________________________________________________________________
flatten_71 (Flatten)         (None, 192)               0
_________________________________________________________________
repeat_vector_71 (RepeatVect (None, 7, 192)            0
_________________________________________________________________
lstm_71 (LSTM)               (None, 7, 64)             65792
_________________________________________________________________
dropout_7 (Dropout)          (None, 7, 64)             0
_________________________________________________________________
time_distributed_142 (TimeDi (None, 7, 32)             2080
_________________________________________________________________
dropout_8 (Dropout)          (None, 7, 32)             0
_________________________________________________________________
time_distributed_143 (TimeDi (None, 7, 1)              33
=================================================================
Total params: 84,737
Trainable params: 84,737
Non-trainable params: 0
```

The hyperparameters used for the model are detailed in the table below.

Table: Model hyperparameters

Hyper Parameter	dtype	values
CNN Filters	int	[32, 64, 128]
Kernel size	int	[2]
Strides	int	[1, 2]
Activation	string	[relu, linear, elu, selu]
Pool size	int	[1, 2]
LSTM units	int	[32, 64]
Kernel Regularizer	string	[l1, l2]
Kernel initializer	string	[glorot_normal, glorot_uniform]
Optimizer	string	[sgd, adam, rmsprop]
Minimum Learning Rate	int	0.001
Maximum Learning Rate	int	0.1
Number of Epochs	int	100
Batch Size	int	8

4.5 Training and Testing

The chosen dataset for this research shows 47 months (4 yrs)of per minute readings of electric power energy consumption of a household. Therefore, for the training purposes of our proposed models, a 75:25 training and testing ratio was implemented to indicate the first 36 months to be used and training data and the last year which is the remaining months will be used as out test data to evaluate the performance of the forecast model. To simplify the dataset further, data is sub categorized into weeks. The reason for this is model our data to fit the purpose of this research which is the forecast electric power that will be consumed for a week ahead or for a certain day.

During training, we also defined the call back utility for the training of the neural network models which consists of optimizer for learning rate, early stopping of the training based on unchanging and stagnant loss to avoid the unusual training of the model and saving the computation time and also integrated tensor board to visualize the training losses and model architectures.

5 Results and evaluation

This section presents the results and evaluation of the conducted research. Overall, 3 experiments were conducted over 2 base model architectures. Firstly, the results for the preliminary statistical methods are presented before the results for the neural networks are presented and discussed in the following sections.

5.1 EXPERIMENT 1: SARIMA model

In this research a SARIMA model, was built for values of p=6, q=0, d=0 and the test predictio ns are shown below. The mean absolute percentage error was found to be 23.28 % and Symm etric mean absolute percentage error (SMAPE) was 20.91%.

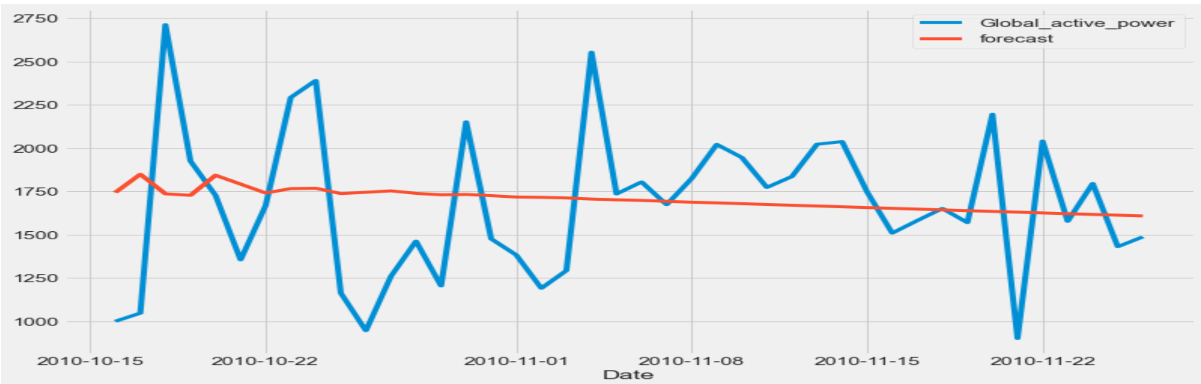

Figure 6: Test predictions for SARIMA

Next, we consider SARIMA using other features in the data, and also added the holidays data to build a more robust model for which predictions are show in the figure below. In this mode l, the mean absolute percentage error (MAPE) was found to be 16.37 % and Symmetric mean absolute percentage error (SMAPE) was 15.25% which can be considered as good drop in the error and SARIMA model is now also able to learn the fluctuation of the time series and predi ct well.

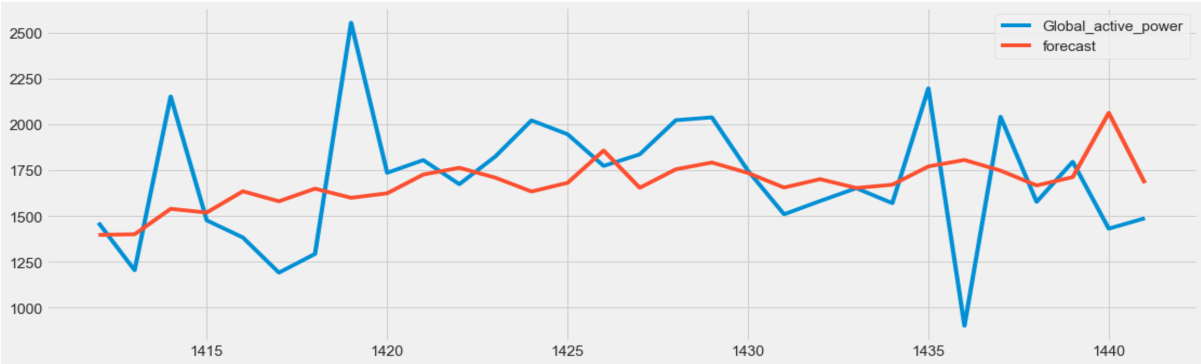

Figure 7: Test Predictions for SARIMA with all features and Holidays consideration

5.2 EXPERIMENT 2: Neural network-based approaches

We defined four different model architecture as our baseline models to test different architectures and evaluate them based on the least root mean squared error. The following were the architectures of the models.

- LSTM model
- Encoder Decoder LSTM model
- Encoder Decoder CNN-LSTM model
- Bidirectional LSTM

The results of the above four model are shown in the figures below.

Table: Neural network evaluation results

Model	RMSE
1. LSTM	0.09349
2. Encoder Decoder LSTM	0.09637
3. Encoder Decoder CNN-LSTM	0.09175
4. Bidirectional LSTM	0.09258

It can be seen from the above table that Encoder-Decoder CNN LSTM model seems to be performing the best with the lowest average RMSE of 0.09175 as compared to the other models. In the next experiment, the model hyperparameters were tuned using the Optuna framework.

5.3 EXPERIMENT 3: OPTUNA CNN-ENCODER DECODER LSTM

For the final experiment, the best performing neural network was further optimized using the OPTUNA framework. The hyperparameters are tuned for both the CNN and LSTM networks.

From the results, ‘rmsprop’ with a learning rate of 0.00128 was found to be the best optimizer for the trial which gave the least value of RMSE as 0.894. We also plotted the optimization history of the Optuna study as shown in the figure below. It can be seen that the tuned model with the best trial has a few models which are below 0.90, which was the best and least RMSE value for our baseline models.

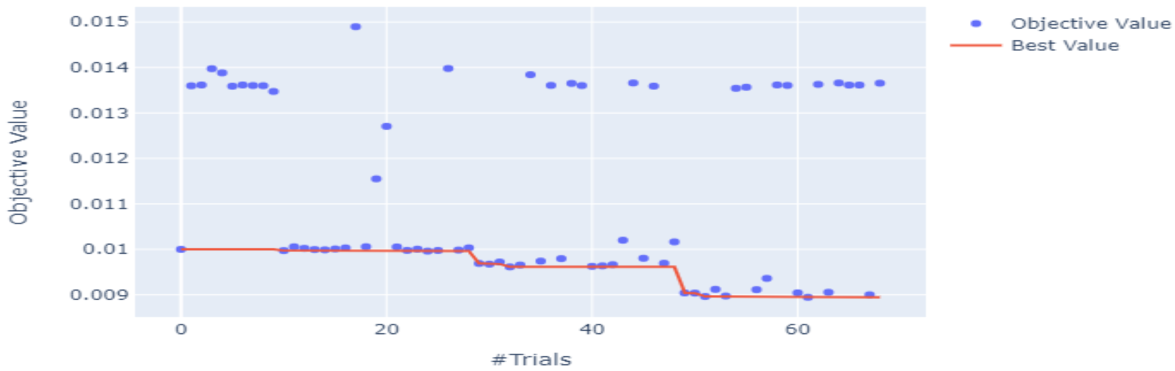

Figure 8: OPTUNA optimization history

The slice plot for Optuna study shows the distribution of the Objective value or RMSE for the different trials based on the values of the various parameters which is also a good way to understand the contribution of different hyper parameters which contribute the accurateness and robustness of the model. It is clear that 'relu' is the activation function which is fitting best and affecting the model, 128 filter in CNN are providing the least value for the RMSE, 'glorot-uniform' is the best kernel initializer and L2 regulariser is fitting the best for our encoder-decoder CNN LSTM model.

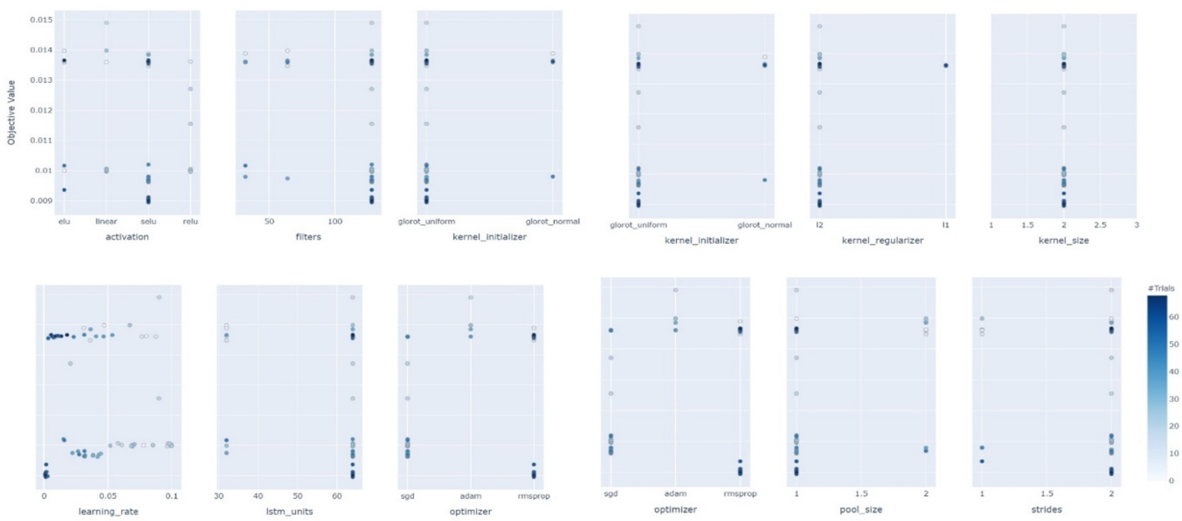

Figure 9: Optimum hyperparameters

The model architecture for the tuned model is shown in the figure followed by the training an d validation curve for the training of the tuned model shown in figure Finally, the RMSE for t he weekly forecast of the tuned model with the Optuna is shown in the figure 11. It can be sho wn that the model has outperformed all the baseline models and average RMSE of the model was found to be 0.09094.

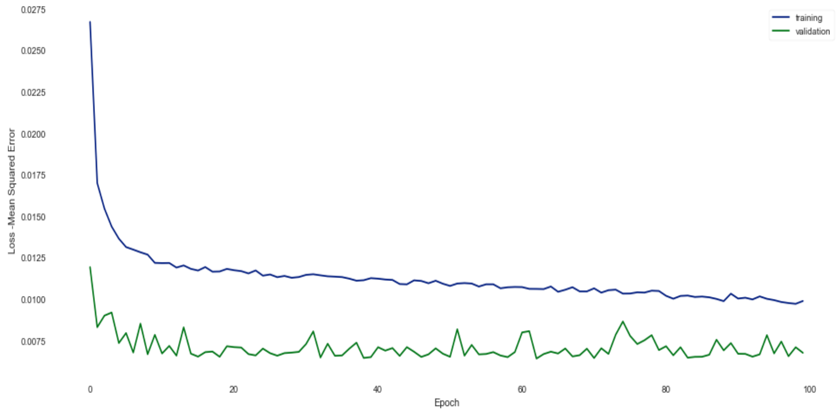

Figure 10: Training history

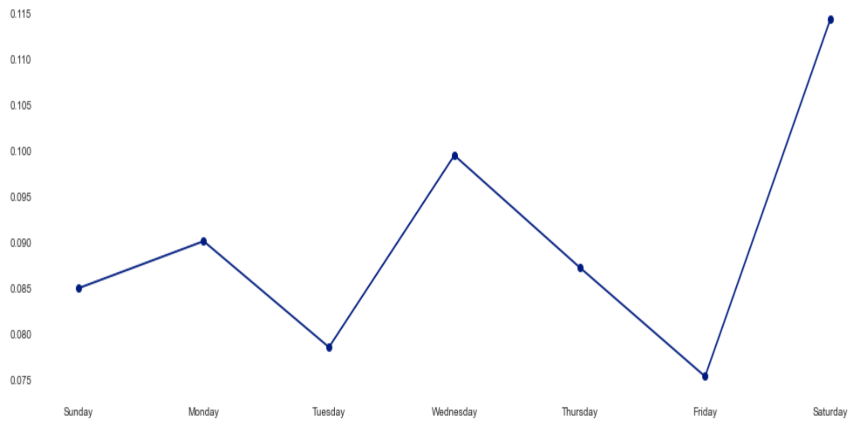

Figure 11: RMSE for the weekly forecast for tuned model

5.4 Discussion

In this research, 3 experiments are conducted aiming to predict household power consumption using different approaches. For the first experiment, a SARIMA model was used to test the effect of seasonality on the model and the residuals were further investigated to detect any hidden information. The SARIMA model had a MAPE of 23.28% and an SMAPE of 20.91%. Following the findings, this experiment was extended using a dataset with world holiday information to further neutralise the effect of the seasonality on the data and the MAPE and SMAPE scores reduced to 16.37% and 15.25% respectively.
Aiming to improve the results, further experiments are conducted using neural networks. LSTM, encoder decoder LSTMs, encoder decoder CNN-LSTM and bidirectional LSTM models resulted in RMSE of 0.09349, 0.09637, 0.09175 and 0.09258 respectively. The encoder decoder CNN-LSTM model performed best amongst the architectures considered. In the final experiment, further hyperparameter tuning is done on the best performing model using the OPTUNA framework resulting in an even lower RMSE of 0.090.
Comparing our results with the work done by Kim and Cho (2019) using particle swarm optimisation to tune the hyperparameters of a CNN-LSTM model, our approach using an OPTUNA model performed better. The results indicate that the Optuna approach is a better

algorithm for hyperparameter tuning compared to PSO. This also validates the ability of our approach to effectively predict future weekly electricity consumption. The Particle swarm Optimization algorithm a min value of 0.0879 MSE which is equivalent to 0.2964 RMSE and our model achieved a 0.090 RMSE. This shows that our OPTUNA approach performs better than the Particle Swarm Optimization algorithm which was presented in the previous work.

6 Conclusion and future work

In this research, Optuna CNN- Encoder Decoder LSTM neural network has been proposed for effective forecasting of electric power demand. To validate the proposed model, we have performed an experiment using the household electric power consumption dataset retrieved from the UCI machine learning repository and result of the experiment showed lower root mean square error in comparison to established models. We have adopted the presented model to probe for the best performing hyperparameters of the CNN-Encoder Decoder LSTM neural networks. Furthermore, we have also adopted SARIMA for the stationarity in our time series dataset to better get a true view of the dataset we are dealing with. Future research endeavour will look to adopting a new algorithm that can search for the best activation functions for converting the input sequence of the deep neural framework into the output sequence.

References

Li, Q, Peng, C, Chen, M, Chen, F, Kang, W, Guerrero, JM & Abbott, D 2017, 'Networked and Distributed Control Method with Optimal Power Dispatch for Islanded Microgrids', *I E E E Transactions on Industrial Electronics*, vol. 64, no. 1, pp. 493 - 504. https://doi.org/10.1109/TIE.2016.2598799

Kong, W., Dong, Z., Hill, D., Luo, F., & Xu, Y. 2018. Short-Term Residential Load Forecasting Based on Resident Behaviour Learning. *IEEE Transactions on Power Systems, 33*, 1087-1088.

Moriarty, P & Honnery, D 2019, 'Ecosystem maintenance energy and the need for a green EROI', *Energy Policy*, vol. 131, pp. 229-234. https://doi.org/10.1016/j.enpol.2019.05.006

Khosravani, H. R.; Castilla, M. D.M.; Berenguel, M.; Ruano, A. E.; Ferreira, P. M. 2016. "A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building." *Energies* 9, no. 1: 57.

Ullah, I., Ahmad, R. and Kim, D., 2018. A Prediction Mechanism of Energy Consumption in Residential Buildings Using Hidden Markov Model. *Energies*, [online] 11(2), p.358. Available at: <http://dx.doi.org/10.3390/en11020358>.

Ronao, CA & Cho, SB 2016, 'Human activity recognition with smartphone sensors using deep learning neural networks', *Expert Systems with Applications*, vol. 59, pp. 235-244. https://doi.org/10.1016/j.eswa.2016.04.032

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. 2017. LSTM: A Search Space Odyssey. *IEEE Transactions On Neural Networks and Learning Systems*, *28*(10), 2222-2232. https://doi.org/10.1109/TNNLS.2016.2582924

Akiba T., Sano S., Yanase T., Ohta T., and Koyama M. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD ’19). Association for Computing Machinery, New York, NY, USA, 2623–2631. DOI:https://doi.org/10.1145/3292500.3330701

Mirjat N. H., Uqaili M. A., Harijan K., Valasai G. D., Shaikh F., Waris M. 2017. "**A review of energy and power planning and policies of Pakistan**," Renewable and Sustainable Energy Reviews, Elsevier, vol. 79(C), pages 110-127.

Erdogdu E. 2007. Electricity demand analysis using cointegration and ARIMA modelling: a case study of Turkey. Energy Pol., 35 , pp. 1129-1146.

El Desouky A. A. and Elkateb M. M. 2000. Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA. IEEE Proceedings - Generation, Transmission and Distribution, vol. 147, no. 4, pp. 213-217.

Yuan C., Liu S., & Fang Z. 2016. Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. Energy, 100 , pp. 384-390.

Cassiano K. M., Junior L. A. T. , De Souza R. M., De Menezes M. L., Pessanha J. F. M., Souza R. C. Chae, Y. T. 2016. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy and Buildings Vol 111, pp 184-194.

Kandananond K. 2019. Electricity demand forecasting in buildings based on ARIMA and ARX models. In Proceedings of the 8th International Conference on Informatics, Environment, Energy and Applications (IEEA '19). Association for Computing Machinery, New York, NY, USA, 268–271. DOI:https://doi.org/10.1145/3323716.3323763

Taylor, J. W. (2008). "An evaluation of methods for very short-term load forecasting using minute-by-minute British data". International Journal of Forecasting, 24(4), pp 645–658.

Fernández, I., Borges, C. E., & Penya, Y. K. 2011. Efficient building load forecasting. In Paper presented at the International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 1–8).

Wang Z., Wang Y., Zeng R., Srinivasan R. S., and Ahrentzen. S., 2018. Random forest based hourly building energy prediction. Energy Buildings, vol. 171, pp. 11–25.

Candanedo L. M., Feldheim V., and Deramaix D. 2017. Data driven prediction models of energy use of appliances in a low-energy house. Energy Buildings, vol. 140, pp. 81–97.

Wang Z., and Ding Y. 2015. An occupant-based energy consumption prediction model for office equipment. Energy Buildings, vol. 109, pp. 12–22.

Fumo N., and Biswas M. R. 2015. Regression analysis for prediction of residential energy consumption. Renew. Sustain. Energy Rev., vol. 47, pp. 332–343.

Cai H., Shen S., Lin Q., Li, X. and Xiao H. 2019. Predicting the energy consumption of residential buildings for regional electricity supply–side and demand–side management. IEEE Access, vol. 7, pp. 30386–30397.

Shaikh H. A., Rahman M. A., and Zubair A. 2019. Electric load forecasting with hourly precision using long short-term memory networks. In Proc. 2019 International Conference on Electrical, Computer and Communication Engineering, Cox's Bazar, Bangladesh, pp.1-4.

Liu Y., Wang Y., Yang X., and Zhang. L. 2017 . Short-term travel time prediction by deep learning: A comparison of different LSTM-DNN models. In Proc. IEEE 20th International Conference on Intelligent Transportation Systems, Yokohama, Japan.

Narendhar Gugulothu and Easwar Subramanian. 2019. "Load Forecasting in Energy Markets: An Approach Using Sparse Neural Networks". In Proceedings of the Tenth ACM International Conference on Future Energy Systems (e-Energy '19). Association for Computing Machinery, New York, NY, USA, pp 403–405.

Yuan. C, Xiu. T, Lou. T . 2019. Probabilistic Long-term Load Forecasting Based on Stacked LSTM. Proceedings of the 2019 4th International Conference on Mathematics and Artificial Intelligence Probabilistic Long-term Load Forecasting Based on Stacked LSTM

Wang P., Bidong L., and Hong T. 2016. Electric load forecasting with recency effect: A big data approach". International Journal of Forecasting, Vol 32, Issue 3, pp 585-597.

Kim N., Kim M. and Choi J. K. 2018. LSTM Based Short-term Electricity Consumption Forecast with Daily Load Profile Sequences. In IEEE 7th Global Conference on Consumer Electronics (GCCE), Nara, pp. 136-137.

Ermatita I., Pahendra E., Darnila M., Sinambela M., and Fuadi W. 2019. Peak Load Forecasting Based on Long Short Term Memory. International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, pp. 137-140.

Wang X., Fang F., Zhang X., Liu Y., Wei L. and Shi Y. 2019. LSTM-based Short-term Load Forecasting for Building Electricity Consumption. IEEE 28th International Symposium on Industrial Electronics (ISIE), Vancouver, BC, Canada, pp. 1418-1423.

Ud Din A. Z., Ayaz Y., Hasan M., Khan J. and Salman M. 2019. Bivariate Short-term Electric Power Forecasting using LSTM Network. International Conference on Robotics and Automation in Industry (ICRAI), Rawalpindi, Pakistan, pp. 1-8.

Kim T., and Cho S. 2019. Particle Swarm Optimization-based CNN-LSTM Networks for Forecasting Energy Consumption. *IEEE Congress on Evolutionary Computation (CEC)*, Wellington, New Zealand. pp. 1510-1516, doi: 10.1109/CEC.2019.8789968.

Alejandro J., Dorado F., & Durán J. 2020. **Energy Demand Forecasting Using Deep Learning: Applications for the French Grid**. Energies, MDPI, Open Access Journal, vol. 13(9), pages 1-15.

Gonzalez P. A., and Zamarreno J. M. 2005 "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network,". Energy and Buildings, vol. 37, no. 6, pp. 595-601, 2005.

Li C., Ding Z., Zhao D., Yi J., and Zhang G. 2017. Building energy consumption prediction: An extreme deep learning approach. Energies, vol. 10, no. 1-20, pp. 1525.

Nishida K., Takeda A., Iwata S., Kiho M. and Nakayama I. 2017. Household energy consumption prediction by feature selection of lifestyle data. 2017 *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Dresden, pp. 235-240, doi: 10.1109/SmartGridComm.2017.8340676.

Li C., Ding Z., Zhao D., Yi J., and Zhang G. 2017. Building energy consumption prediction: An extreme deep learning approach. Energies, vol. 10, no. 1-20, pp. 1525.

Gonzalez P. A., and Zamarreno J. M. 2005. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. Energy and Buildings, vol. 37, no. 6, pp. 595-601.