Ministry of Higher Education and
Scientific Research
University of Diyala
Department of Computer Science

Analysis of Electrical Energy Consumption Data Set Based on Predication Models

A thesis

Submitted to the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By

Zahraa Jabbar Zghair

Supervised By

Prof. Dr. Dhahir Abdulhade Abdulah

2020 A.C 1441 A.H

بسم الله الرحمن الرحيم

وَالَّذِي هُوَ يُطْعِمُنِي وَيَسْقِينِ ﴿٧٩﴾

وَإِذَا مَرِضْتُ فَهُوَ يَشْفِينِ ﴿٨٠﴾

صدق الله العظيم

سورة الشعراء الاية (79_ 80)

Acknowledgment

First of all, praise is to GOD, the lord of the whole creation, on all the blessing was the help in achieving this research to its end.

I wish to express my thanks to my college (college of science), my supervisor, prof. Dr. Dhahir Abdulhade Abdulah for supervising this research and for the generosity, patience and continuous guidance throughout the work. It has been my good fortune to have the advice and guidance from him. My thanks to the academic and administrative staff at the Department of the computer sciences.

Zahraa Jabbar

Dedication

To ...

My father and my brother
Mohammed, may God have mercy
on them
My dear Mother
My husband Khalid for his
unlimited love, support, endurance
and encouragement
My candle, my children
My sister and my brothers
My friends
I produce this work with all my
love….

Zahraa Jabbar

(Linguistic Certification)

I certify that this research entitled "***Analysis of Electrical Energy Consumption Data Set Based on Predication Models***" was prepared by ***Zahraa Jabbar Zghair*** and was reviewed linguistically. Its language was amended meet the style of English language.

Signature:

Name: **Prof. Dr. Luma Ibrahim Al-Barazenji**

Date: 23 / 11 / 2020

(Supervisor's Certification)

We certify that this research entitled "***Analysis of Electrical Energy Consumption Data Set Based on Predication Models***" was prepared by ***Zahraa Jabbar Zghair*** Under our supervisions at the University of Diyala Faculty of Science Department of Computer Science, as a partial fulfillment of the requirement needed to award the degree of ***Master of Science in Computer Science.***

(Supervisor)

Signature:

Name: **Prof. Dr. Dhahir Abdulhadi Abdulah**

Date: 23 / ١١ /2020

Approved by University of Diyala Faculty of Science Department of Computer Science.

Signature:

Name: **Assist. Prof. Dr. Taha Mohammad Hassan**

Date: 23 / 11 /2020

(Head of Computer Science Department)

Abstract

The rising population growth and the number of electrical appliances used day by day, leads to an increase in the consumption of electrical energy, hence the demand for electricity, leading to strain on electricity suppliers. Because there are many factors that affect electricity consumption, the use of the smart meter technology enables us to obtain massive amounts of data around the clock, this facilitates predicting power consumption and energy management regulation.

In this study, a model was proposed to predict the energy consumption of a single household in the short term (one day, one week) and medium term (one month). The proposed model consists of four stages: data collection, pre-processing stage, the prediction stage and the performance evaluation stage. Data set was collected through a single house smart meter for model validation and results analysis. Then, a deep learning machine Long Short-Term Memory LSTM, and well-known machine learning algorithms Support Vector Regression SVR, K-nearest neighbor KNN and Naive Bayes applied it to pre-processed data to predict energy consumption for one day, one week and one month. They are compared using statistical measures: Mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) for performance measurement of these machine learning algorithms.

These statistical measurement values indicate that the performance of proposed model LSTM is better than K-NN, SVR and Naïve Bayes for predicting energy for one day, one week and one month on the data given. The results of the model LSTM are MAE 0.183, MAPE 18.324 and RMSE 0.244 for one day power consumption prediction, MAE 0.145, MAPE 15.182 and RMSE 0.179 for one-week power consumption prediction. MAE 0.145, MAPE 14.018 and RMSE 0.166 for one-month power consumption prediction. It is clear that LSTM model is capable of predicting the consumption of electricity in the short term (one day, one week) and medium term (one month) with high accuracy.

List of Contents

List of Abbreviations

Abbreviation	***Description***
Adam	*Adaptive Moment Optimizer*
AMI	*Advanced Metering Infrastructure*
ANN	*Artificial Neural Network*
DNN	*Deep Neural Network*
GP	*Genetic Programming*
KNN	*K-nearest-neighbor*
KWh	*Kilowatt-hours*
LSTM	*long Short-Term Memory*
MAE	*Mean Absolute Error*
MAPE	*Mean Absolute Percentage Error*
MR	*Multiple Regression*
PDF	*Probability Density Function*
QoS	*Quality of service*
RBF	*Radial Basis Function*
RMSE	*Root Mean Square Error*
RNN	*Recurring Neural Network*
SG	*Smart grid*
SMEs	*Small and medium size enterprises*
SVM	*Support Vector Machine*
SVR	*Support Vector Regression*

List of Figures

List of Tables

List of Algorithms

Chapter One

Introduction

Chapter one

Introduction

1.1 Overview

Consuming power is an everyday practice for people around the world that is done with no real care. Because of the rapid economic development and increasing population growth, the previous decades witnessed a steady high demand for energy usage and thus consumption. As a result, demand on power exceeded the generation capacity leading to difficulties to meet that high demand in some part of the world. The management of energy consumption problem is too big to deal with the losses caused by the growing consumption patterns [1].

The rapid increase in energy consumption requires an accurate expectation of the distribution of electricity consumption [2]. In order to accurately predict the use of electricity, it is necessary to track electricity consumption. Therefore, Advanced Metering Infra-structure AMI was introduced. AMI leads to a large amount of energy consumption data. AMI data is used to predict energy consumption. The prediction helps make decisions about energy distribution from the national grid. Accurate forecasting of electricity consumption can prevent unplanned power outages [3].

The smart grid (SG) is considered one of the most important applications of the Internet of Things. It is an integrated data communication network that is used to collect and analyze data via transmission lines and distribution substations as well as the final consumer throughout the electricity network, and an expectation of consumption can be obtained through this data provided to energy suppliers for strategies Efficient energy management [4].

The very crucial part of the smart grid to register power consumption all over the power grid at the end user side, is the smart meter that collect data on an hourly basis or less and feedback such data to suppliers. Smart meters allow contact between the meter and the central network possible in two ways. The collection of detailed data in an interval of 15 mints or less will assist power suppliers as well as consumers to have a comprehended view of the consumption patterns via data analytics, which has become an important part of the industry research and development field [5]. The data (meter data analytics) sent by smart meters are analyzed for the purposes of:

- Utilization of usage patterns to help in decision making pertaining to purchases.
- Making power consumption predictions using via previous consumption patterns.
- Maintaining efficient power supplies in cooperation with consumers.
- Finding out illegitimate grid connections.
- Compare and correct the performance of meter service providers, to reduce unpaid bills and better maintenance decision to help keeping the grid on [1].

Electricity consumption is a time-dependent attribute. Therefore, there are approaches that use time series to build the model to predict electricity consumption. Availability of past information leads to solutions based on time series analysis since it reflects the time-dependent variations [6].

The forecasts for electricity consumption have been identified as short term (hourly to one week), mid-term (one week to one year), and long term (more than one year) forecasts [7].

Time-series analysis techniques are addressed using conventional approaches and artificial intelligent-based approaches (Artificial Neural Network (ANN), Deep Neural Network (DNN), Multiple Regression (MR), Support Vector Machine (SVM), Genetic Programming (GP)). There are many challenges for mid-term and long-term electricity consumption forecasting [3].

This thesis presents treatment of increased electricity consumption through proposed energy consumption prediction model that use four approaches, along Short-Term Memory (LSTM), a Support Vector Regression (SVR), a K-nearest-neighbor (KNN) and a Naive Bays through forecast electricity consumption for short-term (one day, one week), mid-term (one month). The reason for using these algorithms is that they are a great fit for our problem, since electricity consumption is constantly variable.

1.2 Related Works

There is a group of studies and researches in this field that dealt with methods of predicting power consumption through the use of smart meter data and the most important are:

- **Zheng et al,2017 [8].** The use of a long-term-short-term memory (LSTM)based repetitive neural network (RNN) has been proposed to address the short-term electrical load prediction problem, using a long-term electricity consumption data set. And compare it in the following methods: SARIMA (Autoregressive Integrated Moving Average) which is a seasonal moving average model with integrated automatic regression, NARX (Nonlinear Autoregressive Network with Exogenous inputs) which is a nonlinear neural network model with external input, SVR (Support Vector Regression) which is a very popular model in financial time series prediction and NNETAR (Neural NET work Auto Regression) which is an automatic neural network model to predict single-variable time series with single-layer hidden and lagging inputs. Two evaluation criteria were used as a measure of performance: root mean square error (RMSE) and mean relative absolute error (MAPE) between real values and prediction results. Results showed that LSTM outperforms all other methods with the best expected time series. the results for LSTM are RMSE =0.0702 and MAPE =0.0535.

- **Quek et al, 2017 [9].** Proposed a short- term forecasting method that applies Naïve Bayes Classification (NBC) machine learning technique on easily available input parameters are considered as the continuous-valued data such as instantaneous power, outdoor temperature, panel temperature, on-site irradiance and time of the day to predict the overall energy generated by photovoltaic cells which are installed in distributed region in the next 15-minute period. Based on rule-based inferences, continuous-valued data is converted into categorical-valued data. Categorical valued data is represented as class labels like 'very high', 'high', 'medium', 'low', 'very low'. Historical test data of an existing photovoltaic system located in Singapore is used to evaluate the accuracy of the NBC forecasting method and the comparison demonstrates that the proposed method is able to achieve a forecasting accuracy of over 68 percent.

- **Fayaz et al,2018 [10].** Proposed a methodology for predicting energy consumption in apartment buildings. The proposed method consists of four different layers, namely data acquisition, pretreatment, forecasting, and performance assessment. For experimental analysis, they collected real data from four multi-storied apartment buildings. This data is collected as input to the acquisition layer. In the pre-processing layer, several data cleaning schemes have been published to remove anomalies from the data. In the prediction layer, a Deep extreme Learning Machine (DELM) is used to predict energy consumption. In addition, the use of the adaptive neuro-fuzzy inference system (ANFIS) and the Artificial Neural Network (ANN). A different number of hidden layers, different hidden neurons, and different types of activation functions have been used in DELM to achieve the optimal DELM structure for predicting energy consumption. In the performance appraisal layer for the comparative analysis of three prediction algorithms, mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) were used. The results indicate that DELM performed significantly better than ANN and ANFIS for predicting energy for one week and one month on the data given. The results of the algorithm DELM are MAE 2.0008, MAPE 5.7077and RMSE 2.2451 for one-week prediction. MAE 2.3347, MAPE 6.5464 and RMSE 2.6864 for one-month prediction.

- **Gokgoz F. et al,2018[11].** In this study, models were presented based on deep neural networks, especially long short-term memory algorithms LSTM to predict renewable energy loads with a short-term forecasting horizon, by using data models from the mechanism to support renewable energy resources in Turkey. With an accuracy of one hour between January 2016 and December 2017. Creating 432 different models by changing the cell number of layers and leakage. Instead of SGD Stochastic gradient descent (random gradient) the “adaptive torque estimation” algorithm used for training as a gradient-based optimizer. It performed better than SGD in terms of speed in convergence and lower error rates. Absolute mean error (MAE) and square mean error (MSE) were used to compare model performance. Of the 432 models, five results for MAE were 0.66, 0.74, 0.85, and 1.09.

- **Zhang et al,2018 [12].**In this study, the "support vector regression" (SVR) modeling approach was used to predict the consumption of the individual electric family applied to daily and hourly data for the use of electricity for fifteen households from 2014 to 2016, and by using different methods to divide the dataset into a subset of training and testing for families that are similar in electricity consumption over time, as the successive division on the basis of time works better than randomly sampled data. As for families that lack regularity in the hourly electricity use, then randomly sampling data and using 20% of them as a sub-data set test outperforms the existing approach on the time, since the accuracy of daily data achieved the results of forecasting the best hourly data for all households. Using mean absolute percentage error (MAPE) for one of the fifteen households, the daily forecast is 12.78 and the hourly forecast is 23.31, and it drops to 22.01 (per hour) if only weekdays are calculated for the same the family.
- **George et al, 2018 [13].** Proposed an analytical model describing energy consumption by using energy profiles, which gives energy to the consumer over a period of time, to conduct quantitative analysis using smart meters section. This consumer of the same type assembly and the number of devices together section. This use K-Means algorithm and k-nearest neighbor classification, the value of the account consumption of the most efficient and average consumption within each cluster on a monthly basis and used these accounts to compare individual user consumption, the use of the data set containing the monthly electricity consumption for each apartment for one year.

- **Kim and Cho,2019 [14].** Proposed model that combine the convolutional neural network and long short-term memory (CNN-LSTM). The proposed model extracts spatial and temporal features to actively predict energy consumption for the individual household electric power consumption dataset. Long-term, mid-term, short-term forecasting and real-time forecasting were considered by aggregating energy consumption in units of minute, hourly, daily, and weekly. Linear regression and LSTM models were used to compare experimental results. To evaluate performance, mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE) and mean absolute error ratio (MAPE) were used. The proposed method achieves higher performance than linear regression and LSTM, with RMSE = 0.6114,0.5957, 0.3221, 0.3085 respectively for energy consumption per minute, hourly, daily and weekly.
- **Adewuyi et al,2020 [15].**They applied three models of deep education which are (MLP), (CNN) and (LSTM) to predict electricity demand in the short term by using consumption data at the university in addition to data on the effect of weather on loads in the tropics, and compared them with each other using RMSE,MSE and MAE to measure the accuracy of the prediction As it was measured during the stages of testing, evaluation and training on different epochs 100, 80, 60, and 40, the results showed that the LSTM model outperformed the rest of the models. It was among its results in the test scale of the epoch 80, RMSE=2.46, MSE=0.45 and MAE=2.44.
- **Solyali D., 2020[16].** In this study, the techniques of artificial neural network (ANN), the Adaptive Neuroscience System (ANFIS), multiple linear regression (MLR), and the support vector machine (SVM) were used to predict the electrical load in Cyprus. Historical data were used to show the use of electricity for the period 2016-2017

with long and short- term analysis in Cyprus, and the parameters were temperature, humidity, population, electricity price per kilowatt hour, gross national income per capita, solar radiation. The results indicated that the support vector regression (SVR) is relatively superior to other models, as it showed lower prediction errors (4.34%, 4.49%) and root mean square error (RMSE) (25.43, 26.44) for long-term prediction. In the short-term term, artificial neural network (ANN) techniques showed better results than other techniques with lower prediction errors (0.97% and 1.67%) and root mean square error (RMSE) (7.67, 14.91).

Table (1.1): Related Works Summarizations

No.	year	Author	technique	accuracy
1	2017	Zheng et al. [8]	LSTM	RMSE=0.0702 MAPE =0.0535
2	2017	Quek et al. [9]	NBC	68%
3	2018	Fayaz et al. [10]	DELM	RMSE=2.2451 MAE =2.0008 MAPE= 5.7077
4	2018	Gokgoz F. et al. [11]	LSTM	MAE=0.66
5	2018	Zhang et al. [12]	SVR	MAPE=12.78
6	2018	George et al. [13]	K-Means KNN	———
7	2019	Kim and Cho [14]	CNN-LSTM	RMSE=0.3221
8	2020	Adewuyi et al. [15]	LSTM	RMSE=2.46 MSE=0.45 MAE=2.44
9	2020	Solyali D. [16]	SVR ANN	prediction errors =4.34%, RMSE=25.43 prediction errors =0.97% RMSE=7.67

1.3 Problem Statement

The rapid increase in human population and development in technology have sharply raised power consumption in today's world. Since electricity is consumed simultaneously as it is generated at the power plant, it is important to accurately predict the energy consumption in advance for stable power supply. Peak demand is a problem that the power industry has ever faced as it requires more cost-effective and efficient procedures rather than adding more generators. And since accurate electricity consumption forecasts are of utmost importance in energy planning, they provide strong support for effective energy demand management. This work demonstrates the possibility of using the best model to obtain the best predictive energy consumption. This will reduce the gap between consumers and energy facilities so that they can Communicate more efficiently.

1.4 Aim of The Thesis

This thesis aimed to predict energy consumption using smart meter data through a case study on a single house and choosing a good forecast model for predicting energy consumption.

1.5 Objective

To contribute to reducing energy consumption, changing people's opinion a little smarter during the day in order for a better distribution of energy consumption. and prevent unplanned power outages.