UNIVERSITY OF AGDER

Prediction of Electricity Usage Using Convolutional Neural Networks

Author:
Martin Hansen

Supervisors:
Ole-Christoffer Granmo
Per-Oddvar Osland

This Master's Thesis is carried out as a part of the education at the University of Agder and is therefore approved as a part of this education. However, this does not imply that the University answers for the methods that are used or the conclusions that are drawn.

University of Agder, 2017
Faculty of Engineering and Science
Department of Information and Communication Technology

This page is left intentionally blank.

Abstract

Convolutional Neural Networks are overwhelmingly accurate when attempting to predict numbers using the famous MNIST-dataset. In this paper, we are attempting to transcend these results for time-series forecasting, and compare them with several regression models. The Convolutional Neural Network model predicted the same value through the entire time lapse in contrast with the other models, while the Multi-Layer Perception through Machine Learning model performed overall best. Temperature variables are directly related to power consumption, but the weights from the power consumption values from 1, 2, 3, etc hours before the forecasting revealed to be dominating the temperature weights.

Acknowledgements

I want to thank Professor Ole-Christoffer Granmo for all of his contributions to this project. This would not have been possible without his expertise. I also want to thank Per-Oddvar Osland from Agder Energi Nett for the provision of this challenging project, and for his continuing support and assistance. I am truly grateful for the opportunity to work together with several of the employees and consultants at Agder Energi Nett. They deserve honorable mention as this project would have never been possible without them.

List of people involved in the project

Full Name	Company
Ole-Christoffer Granmo	University of Agder
Morten Goodwin	University of Agder
Sigurd Brinch	University of Agder
Per-Oddvar Osland	Agder Energi Konsern
Sindre Hellvik	Agder Energi Konsern
Erland Kolstad	Agder Energi Konsern
Tor Magne Lindeberg	Agder Energi Teknologi
Sam P. Rajarathinam	Agder Energi Teknologi
Anirban Kundu	Agder Energi Teknologi
Truls Samuelsen	Agder Energi Teknologi
Gitte E. Aasen	Agder Energi Nett
Thom Oscar Lauritsen	Agder Energi Nett
Per Gøran Gergerud	Agder Energi Nett
Signe M. K. Gusdal	Agder Energi Nett
Sabine Claire French	Agder Energi Nett
Margrete Kollstrand	Agder Energi Nett
Arne Gliddi	Agder Energi Nett
Sara Zachariassen	Agder Energi Nett
Bjørn Malde	Egde
Per H. Knudsen-Baas	Egde
Bjarte olsen	Egde
Vegar Kristiansen	Egde
Svein Eliseussen	Egde
Kenneth Pedersen-Rise	Edge
Morten Njåstad Bråten	Bouvet
Jan Waage	Bouvet
Øystein Tveitå	Bouvet
Arild Andaas	Bouvet
Bente.L. Mortensen	Bouvet
Tonje Salgado	Bouvet

List of people involved in the project

Full Name	Company
Kirsten Skådinn	Systek
Are Tysnes	Systek
Øysten Rose	Systek
Tonje Klykken	Systek
Rune Ødegård	Evry
Ole Jonny Gangsøy	Evry
Hege Christensen	Evry
Rikard Johansson	Greenbird
Frode Sætre	Kamstrup

Contents

List of Figures

List of Tables

1 Introduction

Electricity consumption forecasting is a difficult task because of the multiple stochastic variables which are introduced when attempting to map human behavior. People are consuming electricity in a manner which suits their lifestyle, but most people share the same type of usage through out the day. These are everyday household facilities such as showers, ovens, stove, lights, and entertainment systems. We are able to estimate the consumption value of people with a range of uncertainties, because there exists several patterns which is found in human behavior. These patterns can be tracked down to be:

- Requirement of external heat sources for cooking, showering, laundry and increase of temperature.
- Requirement of external lighting for visibility
- Requirement of recharging electrical vehicles.
- Requirement of other equipment or facilities which requires electricity.

There is an interesting approach in classifying the patterns above in order to understand when people require electricity. The classification could potentially provide with information over the households, and could be utilized in order to predict future events based on the patterns above. We will therefore attempt to introduce temperature values to the regression models which we are using in this thesis. The idea is to purposely inject the temperature values and explore how the algorithms will react to the values. The best case scenario would occur if the algorithms was able to highly utilize the temperature value to estimate the correct consumption value based on the patterns. We will implement and compare several regression techniques in this thesis. They are: Linear Regression, Support Vector Machine for Regression, Multi-layer Perception, and Convolutional Neural Network.

1.1 Background

Deep learning is a subfield of Machine Learning, which is a subfield of Artificial Intelligence. The field of deep learning is quickly expanding, which results in an increase of deep learning architectures. [1] Machine learning is often used for classification problems, as well as used to develop pattern recognizing. One of the main challenges of utilizing Machine Learning in general is the access to enough available data in order to gain experience. Machine Learning also struggles with overfitting as the model is likely too dependent on the separating line used for classification, and therefore have an overall higher error rate. [2] The Deep Learning techniques on the other hand are extending machine learning techniques by increasing the complexity of the network. The complexity is introduced by additional hidden layers in the network, and gives the network more information to process in order to avoid the stagnation of no more learning. Deep Learning techniques does experience stagnation in the same matter as machine learning, but has achieved a higher learning percentage than traditional machine learning models with three layers. Deep Learning requires a vast amount of data to be able to function as expected, as the hidden layers used in this technique requires proper weights. This thesis will attempt to test popular popular regression techniques compared to Multilayer Perception and Deep Learning networks by using real household consumption data. The data is retrieved from the new smart meters, which makes the consumer able to regulate his or her own consumption rate. Agder Energi Nett reached out to the University of Agder with the purpose of researching new techniques for prediction of electricity usage. They have also provided with all of the user data used in this thesis.

Electricity generated by the power plants has to be equal to the demand of consumption. The load on the power network is uncontrolled as the difference between generation and demand is ever changing. The smart meters which are currently in deployment from Agder Energi Nett has granted the consumer the ability to control his or her consumption rate. This is possible because of the technology in the smart meters, as they display the amount of power consumption of the household. This is also an advantage for Agder Energi Nett as the demand is changing from a uncontrolled to a slightly less uncontrolled variable as demand becomes more predictable. Agder Energi is required to maintain a voltage grid between 230 V +/- 10%, due to accessibility and government regulations. Agder Energi wishes to provide with enough network capacity to satisfy the changing needs of the consumers. The information gained from the smart meters can potentially be used to track down unique patterns in households. Due to the current enrollment of the smart meters, the information of current network capacity is not available and will not be researched in this thesis. However, forecasting the network capacity could potentially lead to a new thesis.

1.2 State of the Art

This thesis will look at one Deep Learning technique, two regression techniques and one machine learning network. While there exists multiple techniques for implementing Deep Learning networks, we wanted to test if it was possible to replicate the same results with a Deep Learning network which is not usually meant for time series forecasting. Recurrent Neural Networks with Long-Short Term memory is a popular approach in classifying datasets which are heavily dependent on the closest measurements values. The motivation for using Deep Learning techniques is to explore if they are able to produce finer results than other models. The State of the Art techniques will be thoroughly discussed with implementation in the method section of this thesis. Convolutional Neural Networks have achieved overwhelming result when attempting to classify the digits zero to nine based on the MINST dataset. We will attempt to push these results and test a new method where we change the input from image to time series values and have the Convolutional Neural Network attempt to predict the correct consumption value instead the correct digits. This will be an exciting experiment because the network requires proper training where the real life dataset are convoluting both through consumption per hour and temperature.

1.3 Problem Statement

We want to compare Convolutional Neural Networks to other regression techniques by their performance and prediction accuracy, using measurement data from Agder Energi Nett. The main challenge will be to convert the input data from picture to time series classification in order to build a Convolutional Neural Network.

1.4 Literature Review

The usage of Convolutional Neural Networks with time-series is quite a new concept in contrast with applying classification with image recognition. The Convolutional Neural Networks are able to predict 1.3 million high resolution images into a total of thousand different classes with a error rate at 18.97 percent. [4]

In this thesis, we will compare and contrast the results to another thesis made at University of Agder. The paper is a former master's thesis which was conducted in 2012.[16] They attempted to apply Gaussian Process techniques in machine learning in order to see if the power consumption forecasting was an improvement over regular regression techniques. The results generated from the Convolutional Neural Network is compared to some of the same regression techniques which was applied in the former thesis. The Convolutional Neural Network results will also be compared to other regression techniques which is made from legacy classifiers such as Support Vector Machine due to its success in other pattern recognition tasks. [24]

A similar type of work was done by Høverstad et al. 2015.[8] They did not directly attempting to forecast short-term loads through Convolutional Neural Networks, but propose an interesting approach. They suggest a three-stage model which consists of preprocessing, forecasting, and postprocessing. The purpose is to simplify and automate the analysis and estimation forecasting models. In this thesis, a similar three-stage step approach will be used for the prediction model.

An interesting attempt to develop a convolutional neural network was made recently by Borovkh et al. 2017.[6] They managed to design a Convolutional Neural Network forecasting model for conditional time series. The time series did not need long historical time series, but did outperform regular neural network models, which usually consists of the three main layers: Input, Hidden and Output layer.

There is also another paper by Sainath et al. 2013 which applies Convolutional Neural Networks to speech recognition.[23] This paper is not directly related to this thesis, but given the success of converting image to speech recognition indicates a possibility of converting image to time series recognition. They find that Convolutional Neural Network is able to offer an improvement of 4-12% to compared to Deep Neural Networks. They also have an interesting thought process in how Deep Convolutional Neural Networks benefits from each hidden layers as the the number of these does not clearly reveal whether it's benefital or overcomplicating the network.

1.5 Research Questions

How does the Convolutional Neural Network compare to regression techniques and standard Neural Networks by computational and prediction accuracy?

The thesis will mainly focus on the four models which was mentioned earlier. The models will only use data which have been generated by people. The essence of the idea is that no data has been generated by estimation programs, which results in a cleaner dataset. The dataset will be more correlated to the patterns which we hope to track down. The models will then attempt to forecast for n hours into the future and explore the strengths and weaknesses of the models, based on how far into the future they attempt to forecast for.

1.5.1 Subordinated Research Questions

- **Will the Convolutional Neural Network be able to forecast electricity usage?**

  The method chosen to attempt to solve this question is by selecting a series of measurements which should be related to one another. These measurements are then inserted as input nodes, and are used to train the network by inserting back-propagation for each layer in order to optimize per layer instead of per network. The data which we will test which are both available and suitable as part of neural nodes is: Periodic attributed such as previous hour/day value, weekend, quarter, Day of week, Day of year. We have also added temperature (min, average and max) to the dataset.

- **Which of the techniques will have the best performance and accuracy based on 1, 12, 24 and 72 hour forecasting?** The previous master thesis found that each of the predictions was heavily weighted by time-series values such as h(t-1), h(t-2), and d(t-1) where 'h' is a variable for hours, while 'd' is day of prediction. I want to see if the time-series values found in the dataset from Agder Energi is equally utilized as it was in the previous thesis.

- **Which of the techniques will handle the event of the single peak best?** There is exactly one astronomical real consumption rate value which is distinguishing itself from the other values. The peak value alone is not important by itself, but the models may be affected through the weights when they first have experienced such an event. We are curious to see how the regression techniques and Convolutional

Neural Networks are handling the peak value. The most interesting part will be to check if the peak value will disrupt the entire, some, or none of the network.

- **Will further future predictions resolve in a higher root square mean error?** We will expect the models to struggle more the longer it attempts to forecast into the future. The results will be reported and compared through all regression techniques to see if the trend continues, and to see which of the techniques which have the lowest overall root square mean error.

- **How will other attribute such as weather affect the forecasting model?** In our dataset, there are several attributes which acts as a supply of information which does not necessary needs to be correlated to train the techniques. They are all listed in the next chapter. We are curious to see which of them are more beneficial. We have added temperature values to the mean consumption rate for all stations per hour dataset. While temperature has a direct correlation to an increase of power consumption, we are curious to see if the techniques are benefiting from this correlation.

1.6 Solution Overview

The first approach will be to study Convolutional Neural Networks ability to forecast electricity usage. We will also determine whether any of the variables are related to one another, which is done by comparing the results from different datasets. We mentioned the injection of temperature values. The models should then be able to make an assumption based the available data within the dataset. This process will be continued further, by adding more and more information to the dataset which will ultimately reveal any relationships which may represents the patterns in electricity usage.

1.7 Contributions

In this sub section, we will go through the essential work done in this thesis.

- Answering the research questions
- Produced a script which cleans the dataset
- Managed to produce results with Convolutional Neural Network using time-series input
- Found Multilayer Perception to have several stronger results than Linear Regression and Support Vector Machine for Regression.

The thesis will strongly focus on answering the research questions and will be its main contributions. The purpose of the thesis is to research new techniques which may turn out to be superior compared to other techniques. In order to do so, we have designed a script which effectively cleans out any corruption in the dataset, which will be explained in the next chapter.

1.8 Thesis Outline

The outline of this thesis will be organized in a total of nine chapters, including this chapter as introduction. The second chapter will introduce the data to the reader, with the pre-process phase of the dataset. The pre-process phase is essentially a clean-up of corruptions in the dataset. The following chapters three, four, five and six will thoroughly explain how Linear Regression, Support Vector Machine for Regression, Multilayer Perception and Convolutional Neural Network works respectively, and how we choose to apply them in this thesis. Each of the four chapters represent the traditional "method" chapter, but has been extended from one to four chapters. The next chapter, chapter 7, is explaining the experimental setup which is used to answer the research questions in this thesis. We will also present the results in this chapter, but analyze and discuss the results in chapter 8. The final chapter, chapter 9, will summarize the results and conclude the thesis with further work.

2 Datasets

The dataset which we are using in this project are extracted from the user power consumption from Agder Energi Nett, which is obtained from the newly installed smart meters. The data which is generated has either five or six unique attributes depending on the time period it was generated. The enrollment of the smart meter program started first in summer of 2016, and is currently in progress, and will last until 01.01.2019. This thesis would have benefited more if it took place after the enrollment of the program as more information would have been available. It would be interesting to test the models with additional data, but as the data is not currently available, this could open up to a new thesis in the future. The enrollment consist of exchanging old meters with the smart meters in order for households to be able to track their own consumption. The meters are available for extraction of data when installed at any type of household, which means the amount of available data which may be used for input is increasing as more meters are installed. The full potential of the consumption data gathering from the costumers is available when the enrollment is complete, which is as mentioned, in 01.01.2019 (where delays may occur).

2.1 Introduction to dataset

The selection of information used in this thesis was determined by using a analysis platform to insert the data collected from the smart meter to gain a visualized overview of the average consumption.

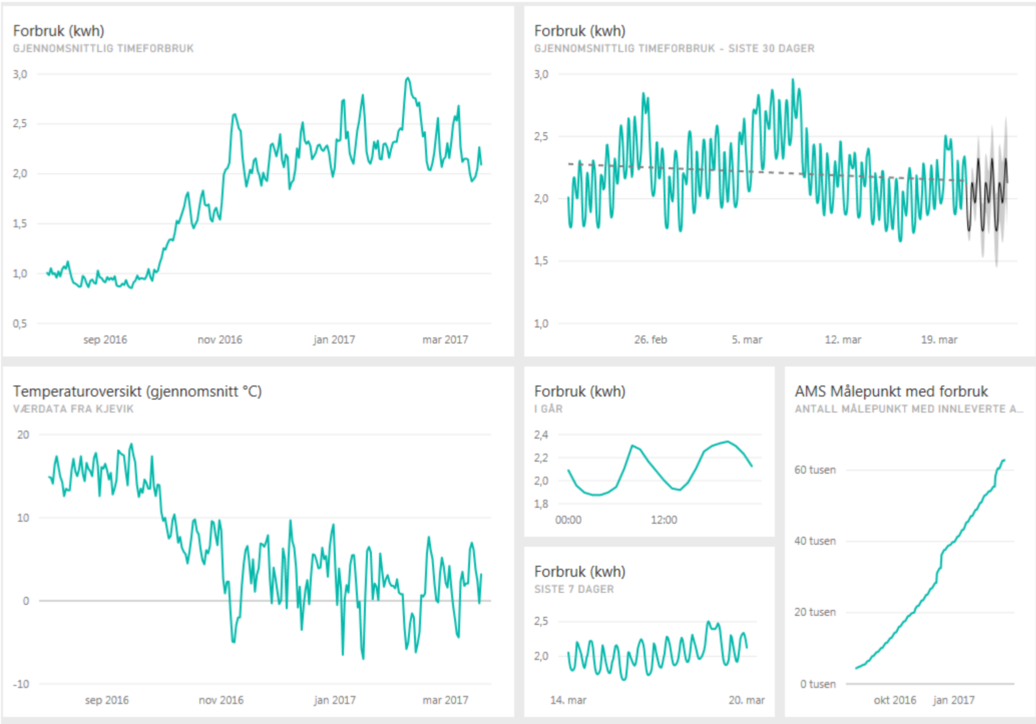

Figure 1: Overview over dataset using Microsoft Power Bi

The figure is visualizing average rates of several data, starting from September 2016 and will be updated every six hour into the future. As access to this platform is restricted, the figure will explain some of the non-confidential measurements. The top left figure is the average consumption rate per hour of all measurements, while the bottom left is the average temperature from one area. The temperature measurements have been extracted from a database found at yr.no, a online Norwegian weather prediction site. The average temperatures in the figure was extracted from only one location, which is called Kjevik, Vest-Agder, Norway.

The top right figure, reveals an overall higher average consumption rate starting from the middle of September and has stabilized at approximately between 2.0-2.5 kw/h compared to an average of of 1.0 kw/h. This is an increase of 200-250% consumption rate during the winter season. When looking at the average temperatures from Kjevik, which is located in the bottom left figure. The temperatures are stable at an average of 15 degrees Celsius during the months previous to September 2016, but drops to an average of 2 degrees Celsius starting at the middle of September to March 2017. The obvious explanation for this matter would be that people have a trend to heat up their household if the temperatures should drop.

We will attempt to train a Convolutional Neural Network to be able to execute regression predicaments to find a linear relationship between the average consumption rate and the average temperature rate, given how they both seem to be related to one another based on the visualization presented. To avoid spending excess time on scripting the exact temperature to each MRID entry, every measurement will be compared to one location which is located central to the area of these households.

```
ReadingType;Value;TimeStamp;MRID;QualityType;TimeStampLocal
0.0.7.4.1.0.12.0.0.0.0.0.0.0.0.3.72.0;1.75;2016-11-19T06:00:00Z;707057500077967665;;2016-11-19T07:00:00
0.0.7.4.1.0.12.0.0.0.0.0.0.0.0.3.72.0;1.96;2016-11-19T07:00:00Z;707057500077967665;;2016-11-19T08:00:00
0.0.7.4.1.0.12.0.0.0.0.0.0.0.0.3.72.0;1.79;2016-11-19T08:00:00Z;707057500077967665;;2016-11-19T09:00:00
0.0.7.4.1.0.12.0.0.0.0.0.0.0.0.3.72.0;2.2;2016-11-19T09:00:00Z;707057500077967665;;2016-11-19T10:00:00
```

Figure 2: Dataset input

From the graphics displayed, there are a total of 6 attributes. They are displayed in the first column and is divided by semicolon. The attributes are: *ReadingType*, $Value$, *TimeStamp*, $MRID$, *QualityType* and *TimeStampLocal*.

These attributes are not self-explanatory, so we will briefly describe them. *ReadingType* grants us information about the ID and the area of the network station which are gathering data from the households. The $Value$ attribute contains the estimate of is estimating the amount of consumption in kw/h format for a given time period. *TimeStamp* reveals the starting point of the measurement, while the *Mrid* attribute is the costumers user identification. This attribute is classified to be confidential information. Therefore, the number displayed in the picture above has been changed in order to conceal the identity of the consumer. *QualityType* is used for classifying the entry with the method of how the data was created. This attribute will be explained more in "Chapter 2.4.3 - Replacement entries". The final attribute, *TimeStampLocal* contains ending time of entry measurement and the date.

We have already access to several attributes which are not needed in our dataset. We will therefore continue to work with the available data in order to create a optimized dataset. The new and fresh dataset should save time for the models as it more beneficial to clean the dataset from the start compared to cleaning the dataset every time we attempt to test with it. There are several entries which we have determined to be corruption for the training of the Convolutional Neural Network and the other models, and we will therefore continue this thesis with explaining why the data is corrupted and how we have removed it in Chapter 2.4 - Cleanse of the Dataset.

2.2 Typical energy consumption pattern for private households

The Deep Learning networks benefits from more available data, and could potentially utilize the added information as consumption can drastically change from minute to minute within the hour. When looking on the statistics of consumption of costumers, the average consumer has a high energy consumption rate early in the morning, which decreases fast over the regular work hours from 09:00 to 15:00. The consumption rate increases when returning from work up until the average time for going to sleep, from which it decreases again during the night. It is relatively simple to find the average pattern of human behavior based on the statistics as most of the Norwegians share the same daily life when speaking of living arrangements. There are however plenty of inconsistencies from person to person, and the total inconsistency is also ever changing. This makes the perspective of the thesis more interesting as we are attempting to have the Deep Learning Networks attempt to learn human behavior.

While the pattern is astonishingly inconsistent in regular households, the behavior is more stable in Norwegian cottages compared to Norwegian households. The reason is that consumers are rarely living long at the cottage compared to households. The statistics reveal the consumption rate to be constant for any time the consumer is not present at the cottage. The rate continues to be constant with some irregularities which is commonly found in regular households. I expect the Deep Learning Networks to have an easier time predicting the consumption rate in cottages compared to households as in theory we are limiting human behavior for the Deep Learning network by asking it to predict less, easier and similar consumption activity.

When looking at cottage consumption, it is more reasonable to expect the machine to do exceptionally well when predicting the consumption when the owner is not present, as it is constant and the training will consist of many entries of the same consumption rate on a long period of time. It is therefore expected to fail during the first hours of when the owner is present, unless there are behavioral patterns for certain visiting periods such as weekends or vacations. There are numerous ways of enabling a pattern for the networks to predict correctly, as an area collectively can be used to predict if certain owners will be present during specific events such as weather, popular date of time or shared holidays.

2.3 Limitations

As mentioned earlier, the datasets which is currently available has been extracted from a period of September 2016 to February 2017. The best case scenario would be to explore this thesis in approximately 5-10 years when every AMS-meter is installed and has been measuring data over several seasons and holidays. The current available data could therefore be a potential limitation for the thesis as any Deep Learning networks require an vast amount of data. If the data should be insufficient, which will resolve in the Deep Learning networks to decline the expected learning rate. the thesis will conclude with predicting human behavior requires additional data, or the algorithms are not suitable, or the behavior is too random and is therefore not possible to predict.

2.4 Cleanse of the Dataset

This sub chapter will focus on how Agder Energi Nett is gathering the data, and how to trim the dataset as some of the attributes are not needed. This is an essential part of the thesis as the current raw data is not optimized for Deep Learning networks. Due to these irregularities, the first period of this thesis will be spent on removing unwanted entries from the dataset. The reason is mainly based on the relation between testing the performance level of Deep Learning networks and the data which is used for training the network. The thesis will briefly describe the different forms of irregularities found in order to creating the dataset, where most of them are related to communication error.

2.4.1 Duplicate entries

The first kind of data corruption we have in our dataset is duplicate entries. They occur whenever the GSM mesh network fail to upload the correct power consumption. This means that the network station have gathered any percentages of the total power consumption, but not all of it. When training a deep learning network, it is important to keep the dataset consistent as duplicates will provide with the same information twice. This is because the network are learning from two entries, which contradicts one another. We prefer to have pieces of data compared to none of it, but in the case of duplicate entries, the latter contains more information than the first entry. When examining the duplicates entries, both will contain the same attributes and the same numbers within. Except for the *value* attribute in one of the entries, which will usually have a higher number. This is because one of the entries contains more information than the other. During the clean-up phase of the the duplicate entries, the script checks for duplicates accepts the entry with the highest value, which is always the second entry after sorting.

2.4.2 Empty entries

The second kind of data inaccuracy is datasets which contains no entries. They are blank CSV-files. There have been some investigation on to why these occur. The issues could occur during interruptions of the measuring meters or during the broadcasting of the mesh network. In any case, the empty CSV-files are removed from the dataset.

2.4.3 Replacement entries

The third irregularity is datasets which contains replacement entries. The data entry in case the wireless mesh network are not able to upload the consumption rate to the local extraction points. The result of this service provides us with duplicate entries as the households are estimated based entirely on the previous week consumption. This is not an accurate representation of the real power consumption data, but the entries are marked in the *QualityType* attribute. This makes it an easy process of cleaning, as the program will determine if the entries contain any *QualityType* information, and discard them if this is the case.

2.4.4 Weather measurement restrictions

The weather report are provided by the services from yr.no. The available data is only provided for a 24 hour basis, which means there are a lack of information. The available data per day is divided into three categories; the maximum temperature, the mean temperature and the lowest temperature. We have opted to simplify the process by adding the same values for each hourly entry, but this results an inaccurate representation of weather in the dataset.

3 Linear Regression

Linear Regression is a technique for modeling a relationship between input variables which are used for prediction of a output value. The function plotted is linear, and is widely used in predicting market prices, electricity consumption, and temperature forecasting. This type of regression is among the simplest, while yielding useful results. In this thesis, we will try out both simple linear regressions which opts only one correlating input value, and multiple linear regression in which opts in several. In this case, as we saw a direct correlation between power consumption and temperature level in "Figure 1 - Overview over dataset using Microsoft BI", which will be utilized as the multiple correlating input value. As mentioned earlier, the temperature data is split among highest value, mean value and lowest value. All three will be added in testing the dataset with the multiple linear regression model.

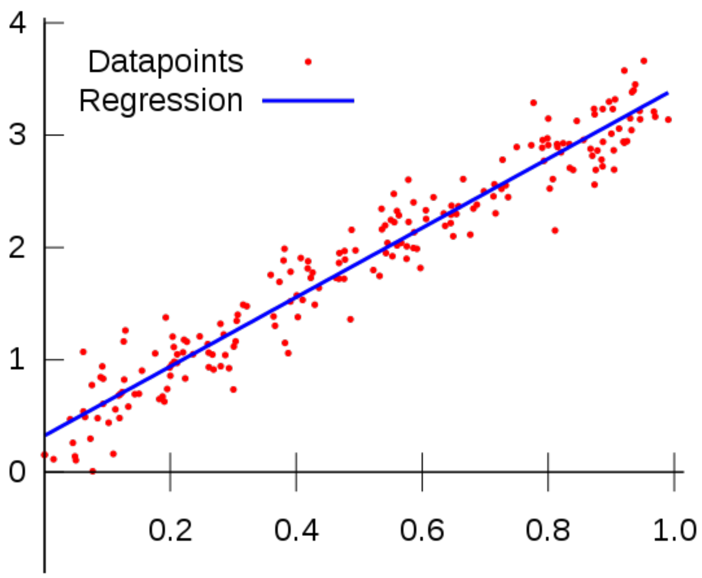

Figure 3: Linear Regression example [7]

In this particular case, we have a simple linear function in two-dimensional space which represents the trend of the dataset. While this does not represent our dataset, this trend could essentially reveal a trend of increase in power consumption. We can clearly see that the vast majority of the datapoints have a significant y value compared to the regression line in our figure. With this fact, we have represented a optimization problem in which we want to minimize the squared error from the datapoints. In order to achieve this, we can implement a higher dimensional function such as represented in the next figure.

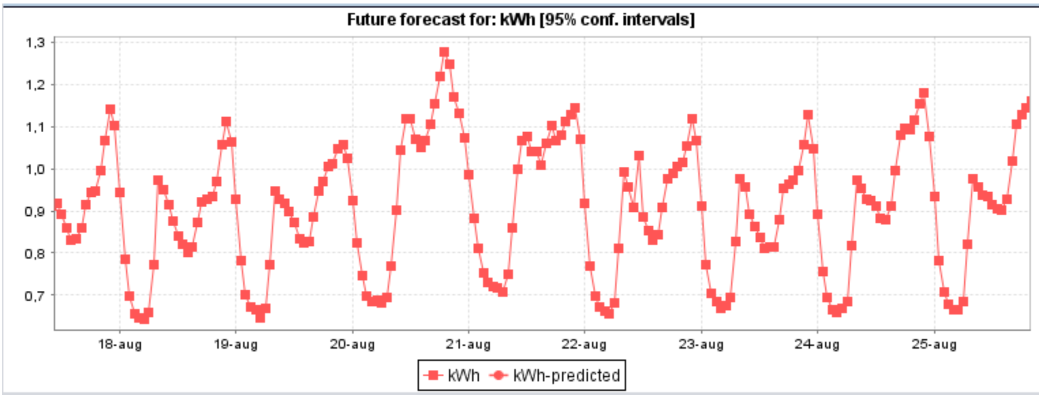

Figure 4: Higher Dimensional Linear Regression Model

The equation of the line between is given by two constants w_k, a_K from the dataset. These two constants is the basis for calculating the weighting when attempting to predict for any training instances. The purpose of the weighting is to reduce the squared error.

$$x = w_0 + w_1 a_1 + w_2 a_2 + ... + w_k a_k \tag{1}$$

The sum of squared error on our training data is given to be

$$\sum_{i=i}^{n} \left(x^i - \sum_{j=o}^{n} w_j (a_j)^i \right)^2 \tag{2}$$

3.1 Linear Regression Model

In order to be able to handle weighted instances in our linear regression model, we have opted to use the Akaike Information Criterion for model selection. The reason for using this measurement is due to asymptotically equivalent to cross-validation and great for forecasting.

Definition: *Suppose that we have a statistical model Given the statistic model M of some data x. Let k be the number of estimated parameters in the model. Let $\hat{L}$ be the maximized value of the likelihood function for the model; i.e. $L = (P(x|\hat{\theta}, M)$, where θ are the parameter values that maximize the likelihood function. Then the AIC value of the model is the following: [10]*

$$AIC = 2k - 2ln(\hat{L}) \tag{3}$$

The selection of models which is measured with the Akaike Information benefits most of those with the least AIC (Akaike Information criterion) value. The main reason for this matter is because the value is considered to fit the model, and therefore the measurement will improve. The measurement will accolade parameters which fits the model, while despises parameters which are overfitting. The parameters which are fitting are usually improving the model which is the primary goal. Ultimately, we want to explain the pattern which consist of explanatory/independent variables as the strength is found in the model selection of regression functions. [19]

3.2 Implementation

The input variables which are used in the implementation consist of a total of 59 unique variables. This is a huge amount of variables for a linear regression model, but the importance of each variable is unique from one another. This is represented in from of weights which we just discussed. The purpose is to acknowledge each of the variables, but in the end, the next power consumption forecasting should reflect the previous hour consumption more than for instance 18 or 5 hours behind. The variables are as follows:

Input = {kWh, Hour, DayOfWeek, DayOfMonth, NumDaysInMonth, Weekend, Month, Quarter, Timestamp-remapped(1 - 3), LagkWh(1 - 24 hours), Timestamp-remapped LagkWh (1 - 24 hours)}

Algorithm 1 Linear Regression Model for n hour prediction

1. Select input data for features and n hours

2a. Set the percentage of training/testing data from the available dataset

2b. Set the weights for each of the input variables

2c. Generated the output for the last step for training set

2d. Generate the regression model using AIC selection

2e. Predict for Power consumption for t + steps out

2f. Set the t-1 of timestamp/consumption to t

3. Repeat for n hours, return to step 2.

4. Return the output value for the predicted power consumption based on h hours

The weights of the input variables are chosen based on their relevance, in which the current power consumption are heavily weighted by the previous hour consumptions. We will create a model for conducting predictions over n hours, and apply prediction for the next hour prediction based on the previous power consumption for the the forecasting. The implementation allows us to implement linear regression to the data points for the next hour.

4 Support Vector Machine for Regression

Support Vector Machine is considered to be a supervised machine learning method which utilizes learning algorithms in order to compute regression analysis models or classification for two or more classes. In this thesis, we want to use Support Vector Machine as a regression technique to forecast power consumption values. The goal is to find a function which has the most deviation from the actual values for all training data, but at the same time be as flat as possible. [3] In this matter we do not bother about the errors as long as they are within the range of the confidence interval. Any deviation from these error are not accepted. The machine learning technique does this by taking two critical numbers which are called support vectors, and they are attempting to make the maximum separation between the classes. The classes are divided by the perpendicular bisector which are generated by the support vectors. [9]

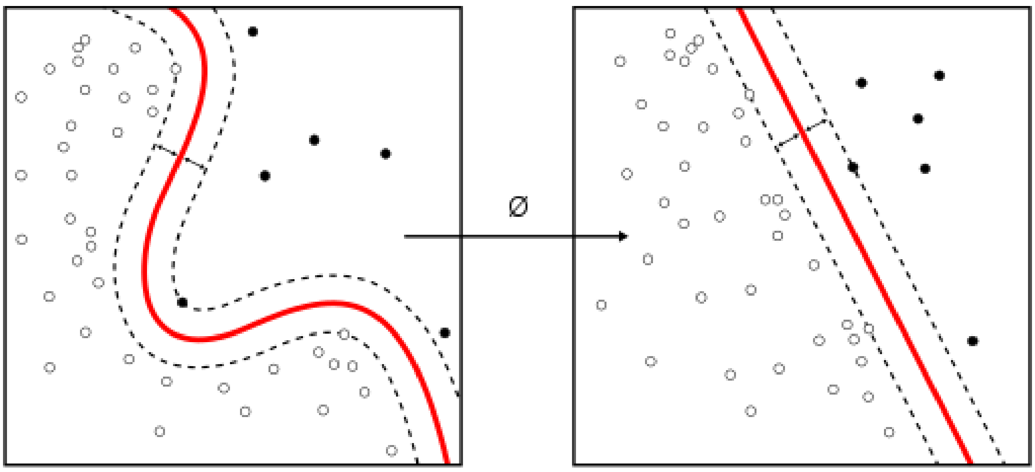

Figure 5: Classic Support Vector Machine Illustration [5]

In our thesis, the graphs will be displayed in two dimensions, but it is possible to extend the support vector classification to multiple dimensions in order to achieve optimized results. The definition of the hyper-plane which are separating the classes is given by the following equation:

$$x = b + \sum \alpha_i y_i a(i) * a \tag{4}$$

The equations sums over the support vectors which are vector products. The hyper-plane does not bother with the rest of the numbers which are placed inside the classes. It only cares about the support vectors. In most cases, and in our thesis, the Support Vector Machine is not linearly separable, but it is still possible to define them into two separate classes by utilizing "Kernel tricks" which makes the boundaries more complex. [22]. This makes it possible to divide the classes by a non-linear plane, and makes it resilient to overfitting.

4.1 Support Vector Machine Model

In our thesis, we have implemented Sequential Minimal Optimization in order to to solve the training problems which is found in Support Vector Machines. The reason is because it is more effective than solvers of quadratic programming problems, but demands more processing time. The worst case performance is the time complexity$O(n^3)$ [20]. The optimization problem in Sequential Minimal optimization is defined as:

Definition: *Consider a binary classification problem with a dataset* $(x_1, y_1), ..., (x_n, y_n)$, *where* x_i *is an input vector and* $y_i \epsilon -1, +1$ *is a binary label corresponding to it. A soft-margin support vector machine is trained by solving a quadratic programming problem, which is expressed in the dual form as follows:*

$$\max \sum_{i_1}^{n} \alpha_i - \frac{1}{2}$$
$$\sum_{i_1}^{n} \sum_{j_1}^{n} y_i y_j K(x_i x_j) \alpha_i \alpha_j,$$
subjected to:
$$0 ¡= \alpha_i <= C$$
for i = 1,2,..., n.
$$\sum_{i_1}^{n} y_i \alpha_i = 0$$

(5)

where c is an SVM hyperparameter and $K(x_i, x_j$ *is the kernel function, both supplied by the user, and the variables* α_i *are Lagrange multipliers.* [20]

Sequential Minimal Optimization are using heuristics to divide the training problem into several bulks which are then solved though analyzing. Within this method, we can use the algorithm to learn the parameters. We are achieving forecasting by using the methods for predictions using a risk function which derives the minimization problem of regression by looking at the empirical error such as root squared mean error, mean absolute error and correlation coefficient. [17]

We summarize the chapter with the implementation of the Support Vector Machine for Regression algorithm which is used in this thesis.

Algorithm 2 Support Vector Machine for Regression

1. Build the classifier

2. Do for each training example

2a. Calculate Support Vectors

2b. Minimize the training by using Sequential Minimal optimization

2c. Generate the perpendicular bisector

2d. Define the two classes and the optimal separation

2e. Estimate the regression based on empirical error rates

2f. Graph Confidence Interval (Range of classes).

3.. Continue for n hour prediction.

4. Continue for n hour prediction.

5 Multilayer Perception Method

Before we introduce the multilayer perception method, we will go through a quick introduction on how general neural networks are built up as multilayer perception is an extension to the most simplest artificial neural network which currently exists, the Single-Layer perception network. Please note that the Multi-Layer perception network is the method which is applied in this thesis, but we introduce the Single-Layer Perception and the general idea of Artificial Neural Networks for completeness of this thesis and for a greater understanding of how Multi-Layer perception works.

The artificial Neural Network aspires to recreate the neural connection which is found in living individuals. The structure of the connection is based on a high connectivity between the neurons where each each of the neurons are connected to one another by synapses. These synapses represents the weighting process of the neural network as they send electrical signals to the next neuron. The simple processing units such as these neurons/nodes consists of several input information. There are several different types of neurons in individuals such as: Sensory neurons, motor neurons and inter neurons. However, the nodes in the artificial neural network is considered to be the same type of neurons, but have been artificially designed with regular configuration in mind. The reason for adapting these networks is because they contain a structure which can learn from experience based on knowledge. This introduces a world of possibilities in the field of forecasting, as future events can be predicted based on previously known information.[13]

The Multilayer perception is based on a non-cycled, or feedforward artificial neural network model. The model is different from a recurrent neural

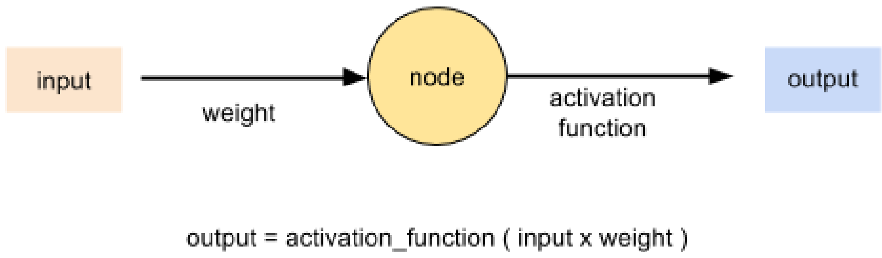

Figure 6: Single Artificial Node [21]

network, where the difference is in the connections between the nodes. In a feedforward network, the weights from each of the other nodes on the same level of the layer are never updated, but are instead pushed forward into the hidden layer. This means that the input variables will constantly be pushed forward to the output nodes. The model is therefore quite simple, but the algorithm suffers from limitations due to this matter. Multilayer Perception includes a method called backpropagation in order to conduct supervised learning for training the network.

In the figure above, we can see the network consist of three layers in total with a input layer, hidden layer and a output layer. The weights, w, is updating the value in each of the circular nodes in the next layer. This network does not use backpropagation to update the weights in the hidden layer based on the other nodes in the hidden layer.

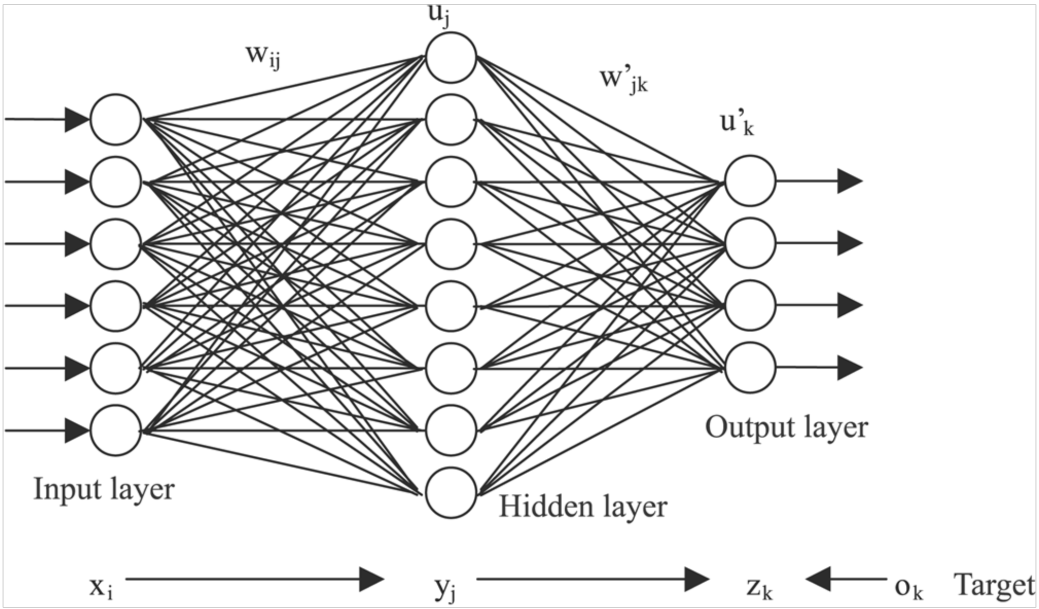

Figure 7: Feed-forward artificial neural network [11]

5.1 Backpropagation

In this sub-chapter, we will go through how backpropagation works and is implemented in our experimental setup. The main reason for implementing backpropagation is to enhance the training of the artificial neural network by committing to a two step cycle. These two steps are propagation and weight updating. As we can see in Figure 5: Feed-forward artificial neural networks, we can see that neither of the nodes in the same layer are connected to each other by any means, and while this simplifies the network,we lose some of the optimization opportunities. Backpropagation takes advantage of an input vector and will propagate it for each node and layer until it reaches the output layer. We are then comparing the output value to a optimize output value by using a simple loss function.

The loss function works by comparing the input and output value to their differences after the input value has been propagation. The function are mapping the values of the variables to accommodate a cost to the propagation. It requires mainly two assumptions in order to work. The first is to write it as a function which may represent the output from the artificial network, and the second is to write it as an average $E = \frac{1}{n}\sum E_x$ for the training data, x, and the error function E [18].

Algorithm 3 Training of a Neural Network with three layers

1. Setup sigmoid network weights

2. Do for each training example

2a. Set prediction to be output of the network

2b. Set actual to be output of the updated output

2c. Measure the error by 2a - 2b for the output values

2d. Calculate difference in weight, w_i, for all weights from hidden to output layer

2e. Calculate difference in weight, w_h, for all weights from input to hidden layer

2f. Update the weights.

3. Check if output value is satisfied, repeat 2 if not.

5.2 Multilayer Perception Model

To implement the the neural network, we use the Multi-Layer Perception network with the same input variables as used with linear regression, where each node contains the weights and each attribute. We have designed the network to have three layers where the first input layer contains 58 input neurons. This is connected to a hidden layer which contains half of the input neurons as default, which are then forwarded to the last layer with one output neuron. There are a total of 84 sigmoid neurons which are used for backpropagation and the associated data.

6 Convolutional Neural Network

We mentioned earlier that Convolutional Neural Networks have achieved tremendous results in image and speech recognition with a small dataset in other papers. Due to these results, we want to test if Convolutional Neural Networks are able to forecast power consumption values based on earlier results in other fields. The idea is to convert the input from time series variables to the same format used in the speech or pixel input. This is a tremendous challenge as converting the data to the appropriate input is necessary in order to be able to forecast the power consumption values.

We selected Convolutional Neural Networks because this is a new, successful and exciting Deep Learning Network. It does also have several similarities with the Multilayer Perception network as they both are Feed-Forward Networks. The main difference between Convolutional Neural Networks used in this thesis and the Multilayer Perception network is the addition of multiple hidden layers. The purpose of adding more layers is to add more complexity to the network. This should increase the potential of reducing the overall error rate compared to the other models discussed in this thesis. Both of the techniques are feed-forward networks which means that the models are not updating the other weight in the same state, but both of them have implemented a technique to support the other weights in each state of the hidden layer. Multilayer Perception utilizes Backpropagation algorithms, while Convolutional Neural Network uses mapping.

The figure explains how the Convolutional Neural Network is mapping the pixels from the image. The image in this example has 24x24 pixels, where each pixel contains information used to classify and train the algorithm. The

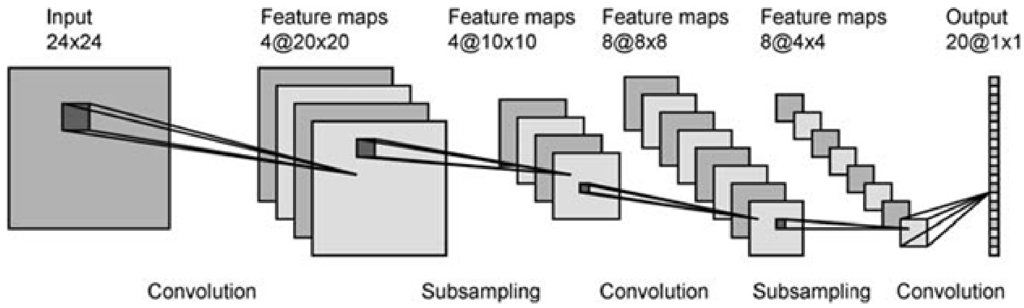

Figure 8: Generic Convolutional Neural Network using image recognizing. [14]

pixel values are connected to several feature maps which works in the same was as the feed-forward Multilayer Perception method. We will attempt to replace the information found in each pixel with the dataset from Agder Energi to see if its able to predict or recognize a power consumption value. The replacement information will consist of timestamp, power consumption, and temperature values. The feature maps are mapped in a similar way as back-propagation works, but are instead placed in a three dimensional figure which are used to classify the image. The three dimensional figure will perform a joint regression in order to forecast the image based on the defined classes. There will typically be some significant pixel values, which are more favored based on training. These types of values could potentially reflect the previous hourly timestamp values in our thesis.

6.1 Implementation of Convolutional Neural Networks

We have adapted the Convolutional Neural Network with a height of 28x28 in desired height and weight of the input values. We have also chosen to simplify the model by using a single channel. Traditional Deep Learning networks requires a vast amount of data in order to solve the complexity which occurs through the feature maps. This could potentially lead to a limitation for the Deep Learning Network as the network is not able to learn due to lack of enough input. By simplifying the model, we are attempting to reduce the overall complexity.

The figure below represents the results from this method, in which it will only forecast for the number **2.4566** kw/h which is a reasonable forecast for the months between November to March. The model we have designed in this thesis are attempting to forecast the power consumption value for the next hour, but are using the same configuration as the other three models.

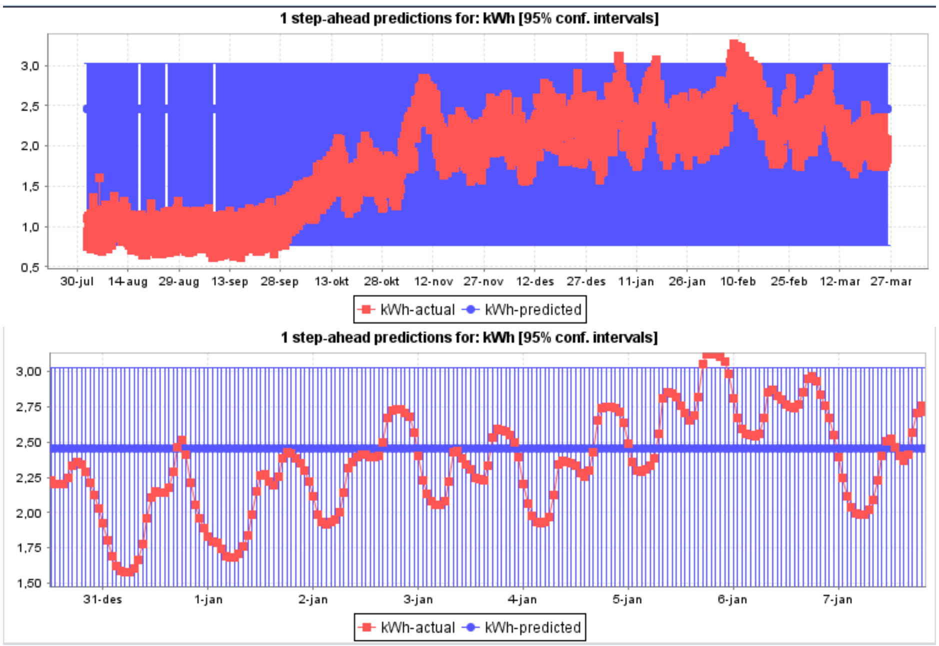

Figure 9: Convolutional Neural Network, 1 Hour Prediction

The results are divided into two graphs, where both are part of the same results. The Convolutional Neural Network are clearly struggling when predicting the next hour power consumption. The confidence interval at 95 % is too large when it comes to uncertainty and does not cover the start months as the range does not satisfy the actual values. We can also learn that the model is attempting to predict more often than one hour, despite using the same configurations as the other models. The Convolutional Neural Networks also displays several gaps in forecasting as it skips some hours of the prediction, but does continue after a random time limit. There are no consistencies or patterns found in the gaps of the results.

There could be several outcomes to why the Convolutional Neural Network are attempting to predict the same value. We will now explain some of the possible outcomes which may have affected the network. The first outcome could be that the model is not accepting the time series input. It is able to produce results for only one hour, but are forecasting the same power consumption for the entire season. This indicates that some of the data are weighted by the Convolutional Neural Network, and could potentially become over-weighted. The model could also contain defects in which the model are not able to shift the previous hour values. This means that the model does attempt to forecast for the next hour, but are always using the power consumption values from the same time series. Moving forward with the thesis, we will continue to compare the other three models in order to answer the research questions, which means that this model will be slightly disregarded. We will also present some error rates in order to gain more information of the results found here.

7 Experimental Results

In this paper, we will go through the results from each of our experiments based on the previous four techniques which have already been presented in this thesis. The models will be using a setup which are equal for every model. The evaluation of the models will present a double image for each model for each n hour prediction. The top image will display the entire time-line, while the bottom image is scoping from January 1st to 7th in order to compare the confidence intervals by each other. Gaining a perspective of the numbers may be challenging for these graphs, but they are instead presented in the next chapter. The weighting of the input are randomly selected by "random" sigmoid functions for each input, and are changed by minimizing the error rate. This is continued until the error rate is satisfactory, and is changeable.

7.1 Experimental Setup

7.1.1 Dataset:

We begin with how we have defined the dataset which are used for the experiments. In the beginning of the thesis, the purpose was to opt in a dataset which required a big-data solution, however this turned out to be a large enough limitation, due to the lack of support for storing confidential information. The dataset was CSV-based, where approximately 95% compressed to around 5 GB of data. While this is estimated to be around 500 GB, the volume of the data was simply too great for handling the training phase. Instead, we have opted to use the total mean of consumption per hour from 01.08.2016 to 25.03.2017, which is around 6000 hours of entries.

7.1.2 Performance Evaluation:

We will evaluate the performance of the predictors by the following measurement standards:

- MAE = Mean Absolute Error
- RMSE = Root Mean Squared Error
- MAPE = Mean Absolute Percentage Error

$$MAE = \left| \frac{Actual - Predicted_i}{Actual_i} \right| \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=i}^{n}(Actual_i - Predicted_i)^2}{n}} \tag{7}$$

$$MAPE = \frac{100\%}{N} \sum_{i=i}^{n} = \left| \frac{Actual_i - Predicted_i}{Actual_i} \right| \tag{8}$$

7.2 Evaluation of the predictors

We will now present the result based on forecasting for 1, 12, 24, and 72 hours using all four algorithms and present the results by the three empirical measurement standards. The order of the algorithm is: Linear Regression, Support Vector Machine for Regression, Multilayer Perception with Machine Learning, and then Convolutional Neural Networks. In order to not make the graphs confusing as there a total of 24 graphs, we will present them all at the time for each hourly prediction, then discuss the findings. Each of the figures contains two images. The first image represent the confidence interval at 95% as requested from Agder Energi Nett, then we will present a scope of the first day in December. This is because the graphs may look similar when looking at the entire graph as a whole, instead of sections. The purpose of the graphs is not to be directly analyzed, but rather give the reader an overview of how the trend continues and to be able to compare the graphs by each other. The numbers will instead by listed in the next chapter.

There are multiple options of displaying the graphs as we have multiple instances of time. The thesis will suffer if it contains more graphs than displayed. In case there would be interest in looking at different instance of time, which is not presented in the thesis, please request the appropriate scope.

The purpose of confidence intervals are to gain a frequency of estimations of true values. The range of the interval should satisfy 95% of the numbers, which means there are significance level of 0.05.[15] The presented graphs will have an actual value, and a predicted value. The color blue represents predicted values, while the color red represents the actual estimation. By comparing the range of the confidence interval, we can assure that 95 % of the forecasts will hit within this range. This will gain us a statistical advantage as we are not able to precisely estimate the value, but have an overall idea of the range which the forecasting will end up in.

We have used a Java application called Weka, to be able to produce the results in this thesis. For completeness of the paper, the Weka application can be read more about here. [12]

7.2.1 Forecasting 1 Hour in the future

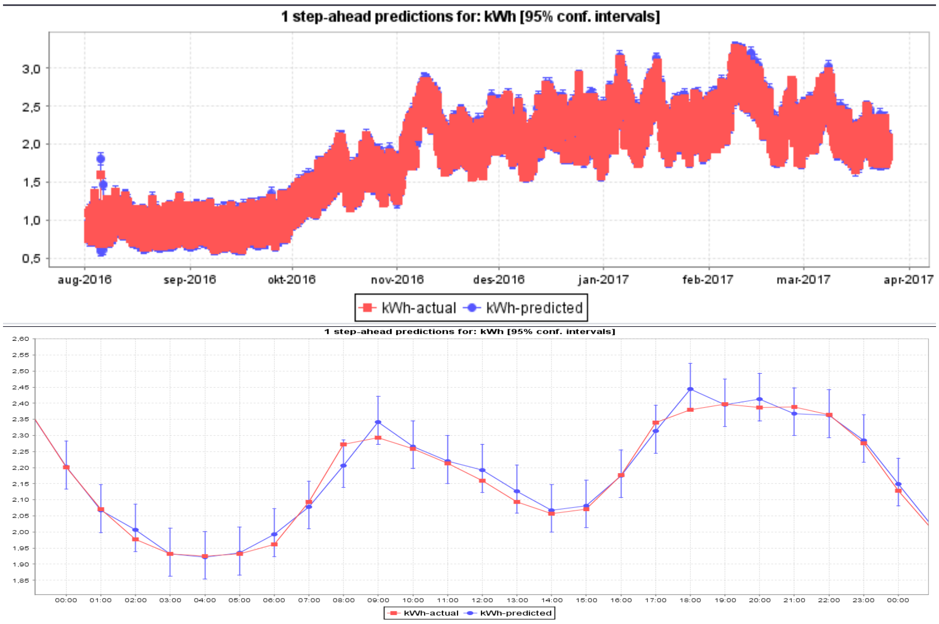

Figure 10: Linear Regression, 1 Hour Prediction

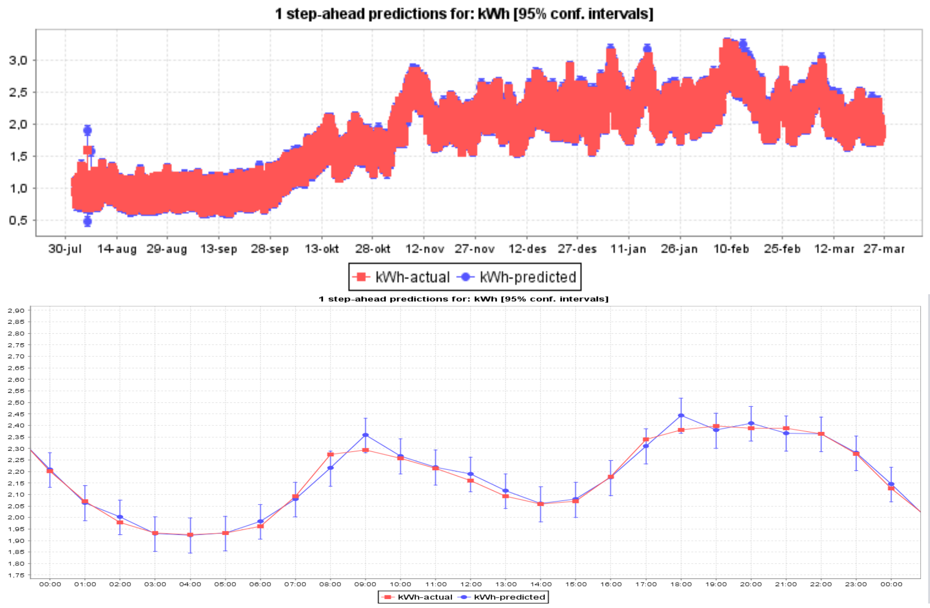

Figure 11: Support Vector Machine for Regression, 1 Hour Prediction

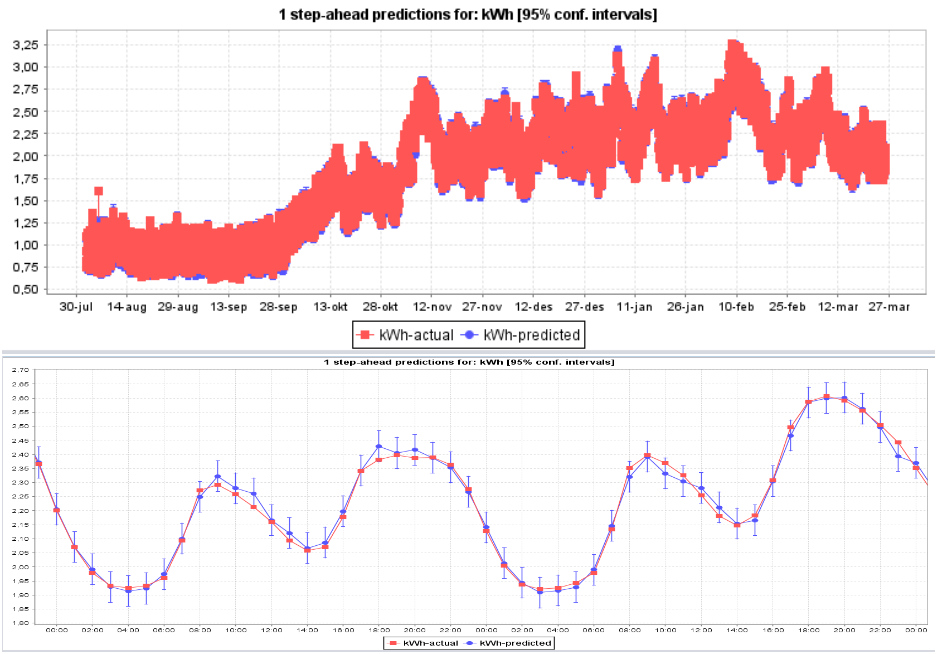

Figure 12: Multilayer Perception, 1 Hour Prediction

The algorithms are attempting to forecast based on the next hour, which means that the weights are heavily based on the previous hour. We can clearly see that all of the techniques are able to handle the predictions quite well, with a lowest range of confidence intervals based on the 12, 24 and 72 hour ahead predictions. The Linear regression yields a range of approximatly 0.15 kWH, while Multilayer Perception has a range of 0.1 kWh. These results does partially answer the research question: **Which of the techniques will have the best performance and accuracy based on 1, 12, 24 and 72 hour forecasting?** The forecasting from 12, 24 and 72 hours will also, though combined, answer this research question.

7.2.2 Forecasting 12 Hours in the future

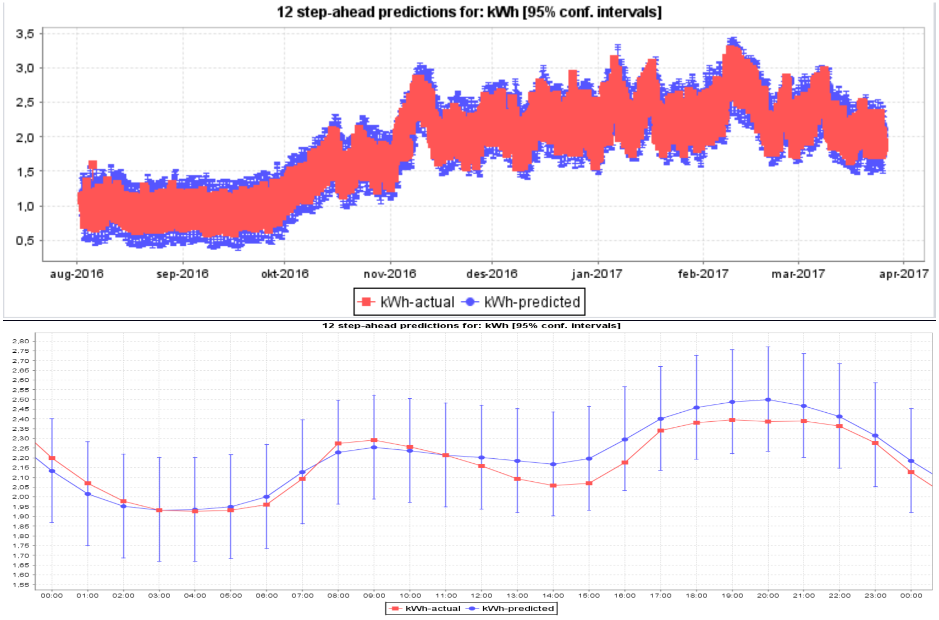

Figure 13: Linear Regression, 12 Hours Prediction

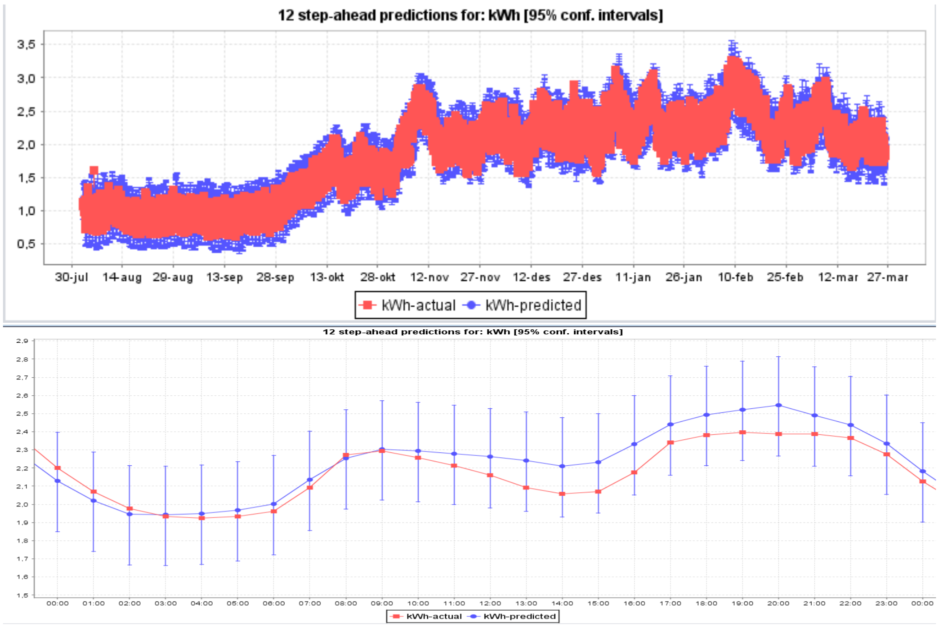

Figure 14: Support Vector Machine, 12 Hours Prediction

Multilayer Perception with Machine Learning

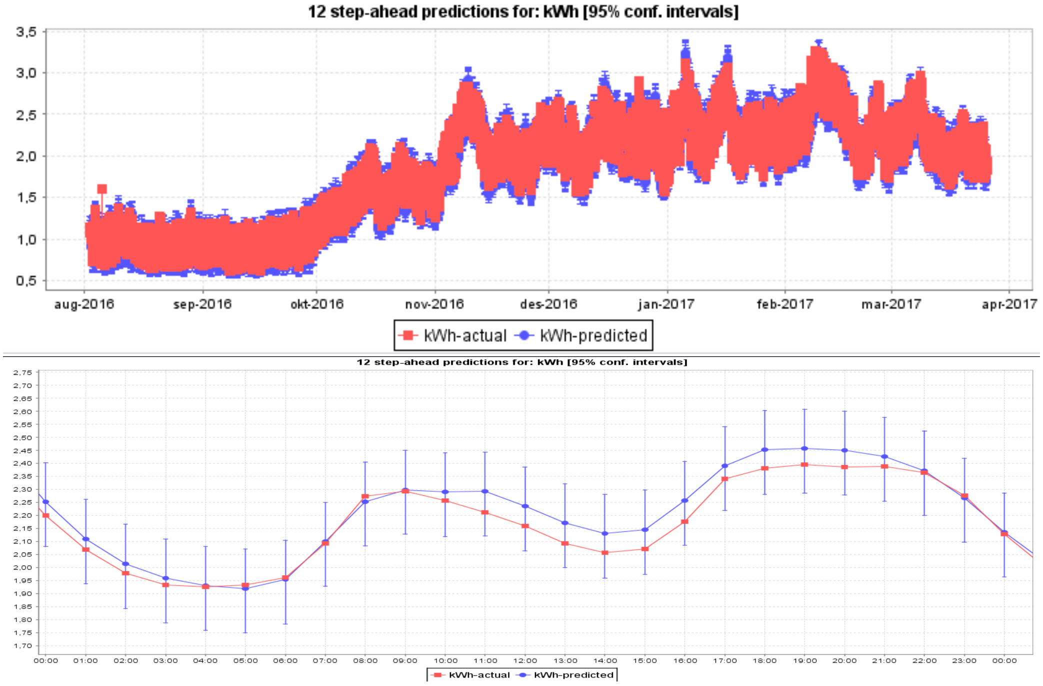

Figure 15: Multilayer Perception, 12 Hours Prediction

We have now changed the forecasting to twelve hours, which have made the range of the confidence interval significantly higher than the results from one hours, as seen in the figures. The range of the interval is still yielding satisfactory results with Linear Regression at 0.5 kWh, Support Vector Machine at 0.55 kWh and Multilayer Perception at 0.35 kWh. It seems like the Support Vector Machine is struggling more than the Multilayer Perception and Linear Regression techniques. Remember this is several models predicting from the current hour, and forecasting tweleve hours ahead. They are therefore predicting with weights were the previous value from the current step which it is forecasting from. This means it is forecasting for t_{12}, by using the weights from t_{-1}, t_{-2}, t_{-n}.

7.2.3 Forecasting 24 Hours in the future

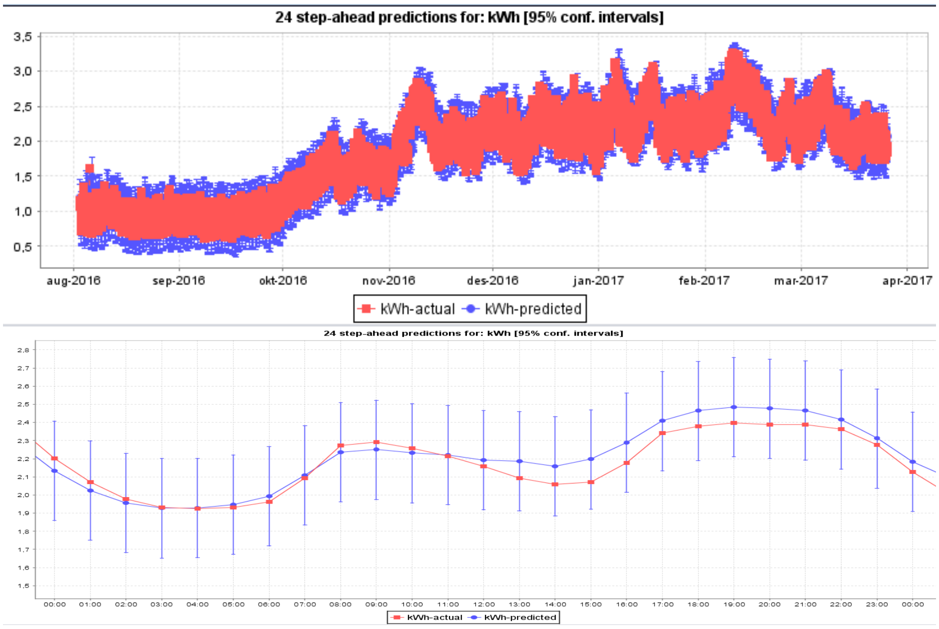

Figure 16: Linear Regression, 24 Hour Prediction

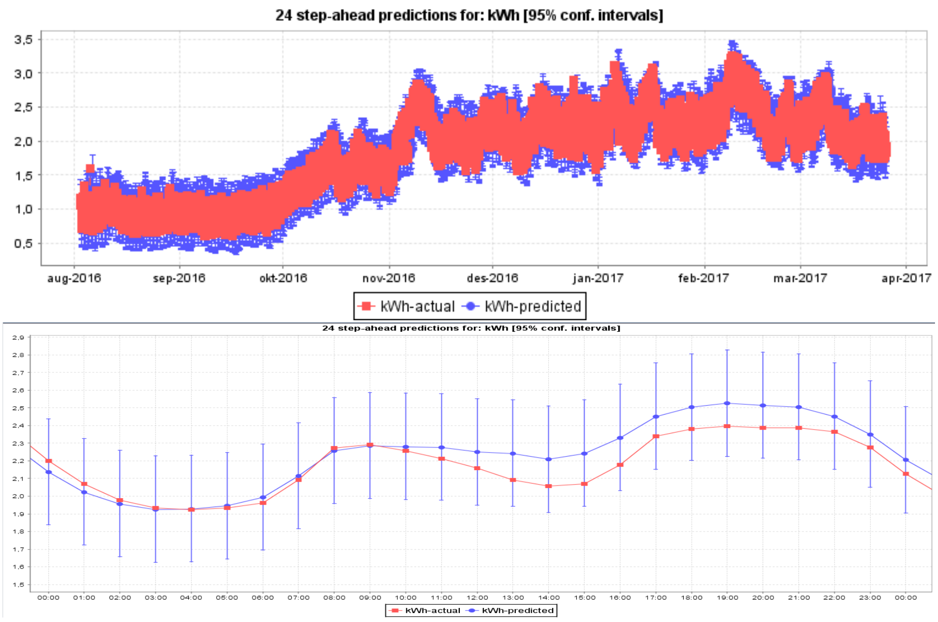

Figure 17: Support Vector Machine for Regression, 24 Hour Prediction

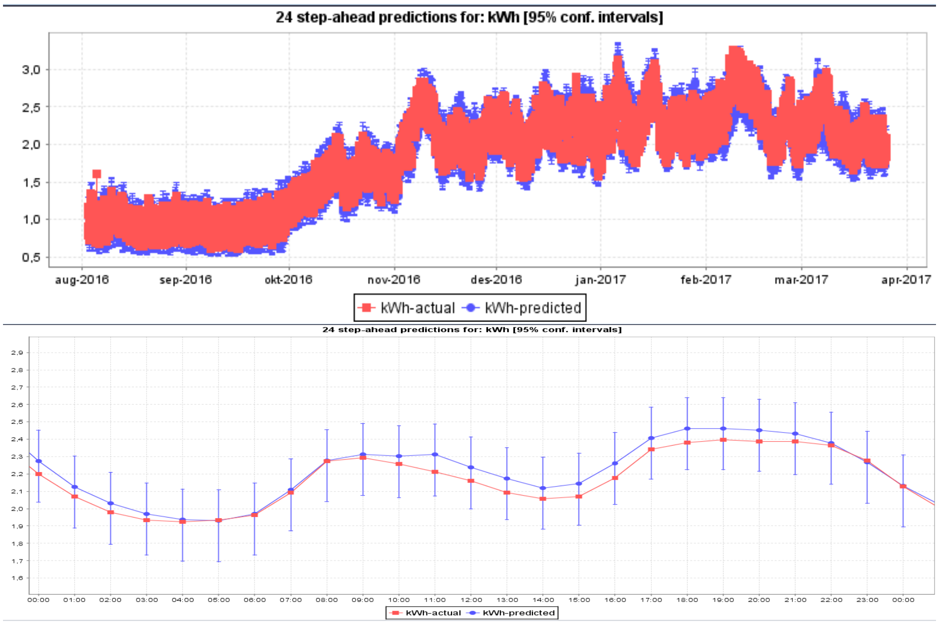

Figure 18: Multilayer Perception, 24 Hour Prediction

The trend of an increased confidence interval is significantly increasing in both Linear Regression and Support Vector Machine. The complexity in the Multilayer Perception technique seems to have yielded a significant advantage over the other techniques, but the range of the confidence interval is still within reasonable numbers. To clear up any confusion, we are expecting the forecasting per hour to hit the range by an accuracy of 95%, but if the range of the interval should yield from 0 kWh to 4 kWh in our training set, then a wild guess is just as satisfactory as the confidence interval. The length of the confidence interval is approximately 0.5 kWh in both Linear Regression and Support Vector Machine, while the length in Multilayer Perception is approximately 0.35 kWh

7.2.4 Forecasting 72 Hour in the future

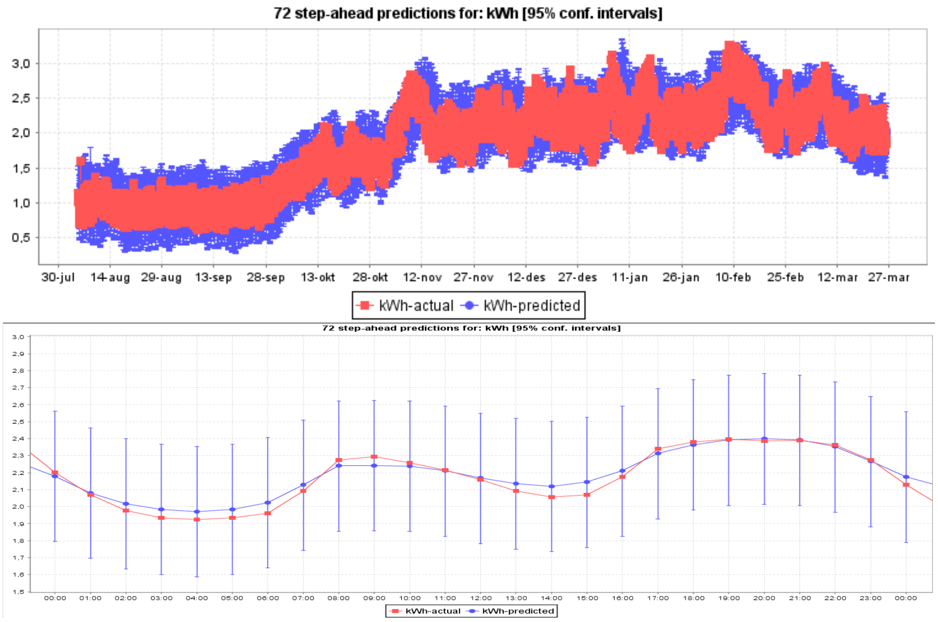

Figure 19: Linear Regression, 72 Hours Prediction

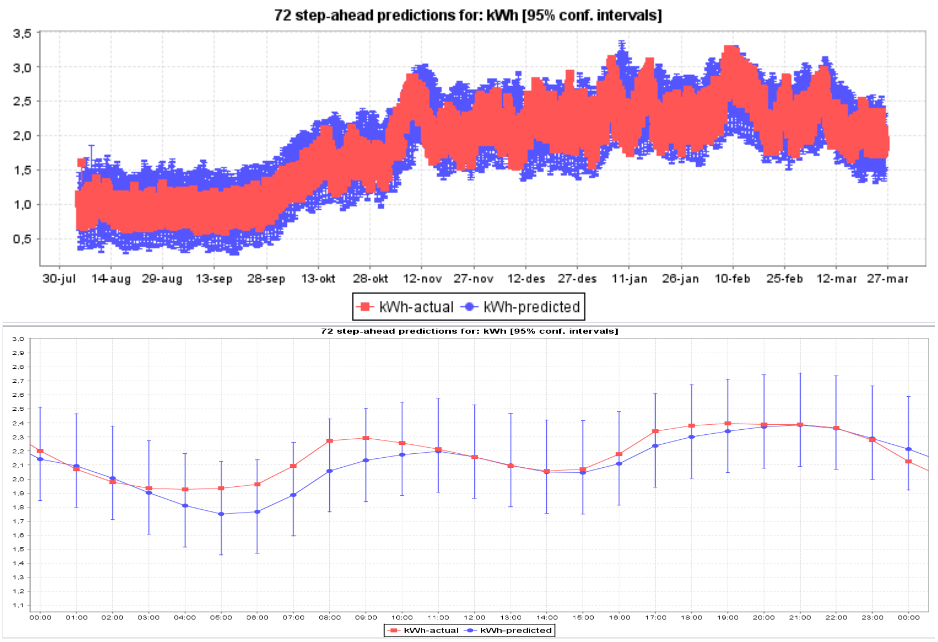

Figure 20: Support Vector Machine for Regression, 72 Hours Prediction

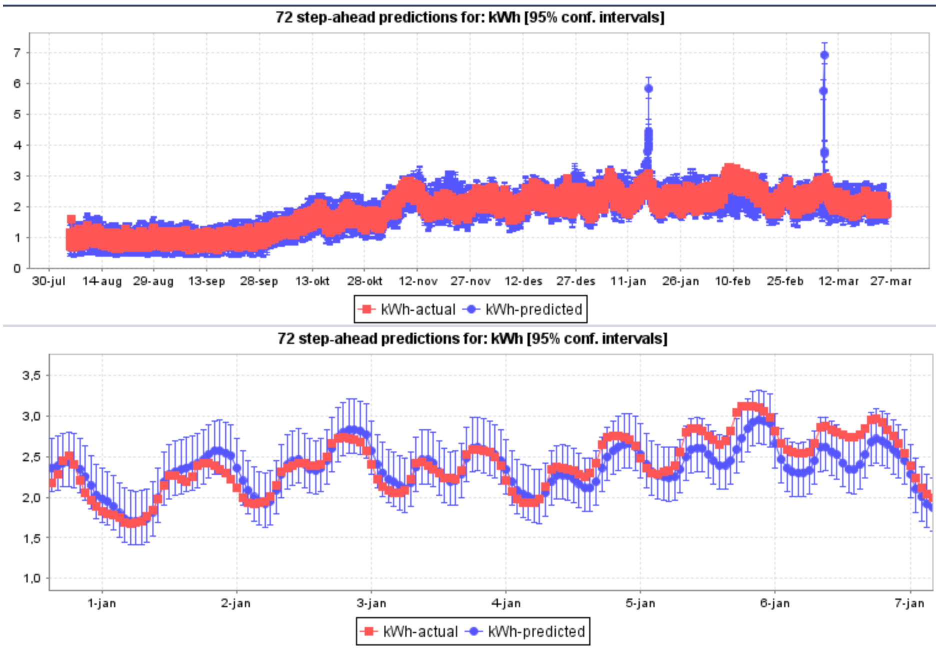

Figure 21: Multilayer Perception, 72 Hours Prediction

The final forecasting reveals that the range of the interval is increasing the longer we forecast. The Multilayer Perception model have some awkward predicaments for January and March, but represent the same results as Linear Regression model and Support Vector Machine with a range of 0.75 kWh. I expected Multilayer Perception to perform better based on the 24 hours predictions, but it seems the forecasting is starting to become affected by forecasting too far into the future, but this is just speculation. These results could also change with additional entries in the dataset.

7.3 Peak value experiment

The four main techniques used in this thesis are Linear Regression, Machine Learning through Multilayer Perception, Support Vector Machine for Regression and Convolutional Neural Networks, as mentioned in the previous chapters. In the following experiment we will attempt to see which of the following techniques are able to handle best a real value which are vastly higher compared to the mean of all values. The purpose of this experiment is to discover how each of the techniques behaves when predicting such a value, as this value may be highly disruptive for the training phases. The value is not artificial by any means, but it there is a chance the measurement system has reported a false value. All four techniques are forecasting for the next hour, which means the previous hour values should be significant in forecasting this abnormality. The dataset consists of total mean value per hour from 01.08.16 to 25.03.17. Which means the algorithms have 6000 hours to train with.

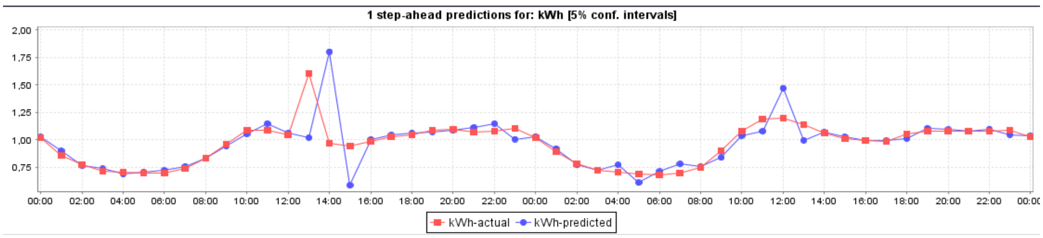

Figure 22: Linear Regression Experiment

The first figure is by using linear regression technique.There are three noteworthy results to report from this technique. Linear Regression is able to predict the next estimated value with almost no difference from the actual value as seen in the figure. It is not able to deal with the abnormality, but this is highly expected as this is the first and only value which truly distinguish itself from the other values. We are essentially asking the prediction model to predict a behavior which has never occurred previously. This is because the dataset only has one out of nearly 6000 values to learn from. Therefore, the prediction made for the 13:00 value is expected to non-predictable based on the number of similar events.

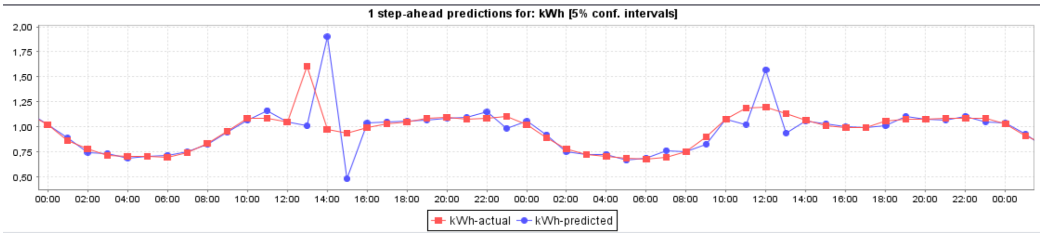

Figure 23: Support Vector Machine Experiment

The interesting results are shared among three other future prediction post 13:00 which occurs when attempting to predict future values. The affected predictions happens at 14:00 and 15:00 and 12:00 the next day. The figure reveals that the prediction at 13:00 has been increased from around 1 kw/h to around 1.625 kw/h, an increase of 0.625 kw/h over one hour. This has caused the linear regression technique to think the prediction at 14:00 will follow the same trend as the previous hour, and will therefore predict an even higher estimation at 15:00. This value is peaking over 1.75 kWh, when the mean of all values are less than 1kw/h. The technique struggles with dealing with the abnormality as it continues to estimate wrong. It continues to copy the trend which is a large decrease of power consumption, which means the algorithm will estimate the power consumption to largely decrease over the next hour. While this is essentially true as there were only one high consumption rate, it estimates the next value to be dramatically less than the real value. When finally attempting to predict the 16:00 power consumption, it looks back to a stable trend between from 14:00 to 15:00 were both values are a little under 1 kw/h.

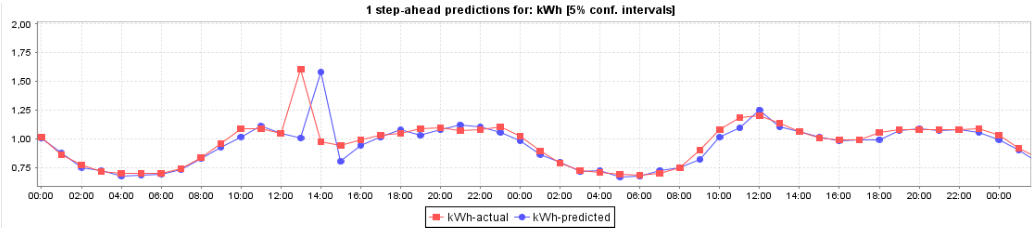

Figure 24: Multilayer Perception Experiment

The interesting estimation value is rather resolved around the 12:00 prediction which occurs the next day from the day with the abnormality. It appeared as the events of the irregularity had finally been stabilized after the trend was normalized, but it continues to struggle when attempting to estimate the time-series 11:00 to 13:00, the next day. The first and final prediction are similar in a way that they estimate both values to be less than the real value, while the middle prediction is following the same behavior found 23 hours earlier, an hour earlier. The algorithm does not appear to experience more problems after this prediction.

8 Summary of Experiment Results

We will go through the research question(s) and attempt to answer them based on the results from the prediction models which has been used in this thesis.

- **Will the Convolutional Neural Network be able to forecast electricity usage?**

  The results found with the Convolutional Neural Network forecasting did only present forecasting of 2.45 kWh across the entire dataset. There are several possible outcomes such as high end stagnation where the network is not trained enough to be able to predict other values. The confidence interval does not yield a satisfactory result so far, but could potentially be improved by using other models or providing additional data.

Empirical measurements // Method	LR	MLP	CNN
Correlation coefficient	0.9270	0.9259	-0.0193
Mean Absolute Error	0.1936	0.1953	0.6807
Root Mean Squared Error	0.2369	0.2413	0.8719
Root Absolute Error	35.8 %	36.2 %	131.2 %
Root Relative Squared error	27.5 %	38.2 %	141.6 %

Table 1: Models comparing using different empirical measurements

- **Which of the techniques will have the best performance and accuracy based on 1, 12, 24 and 72 hour forecasting?**

  The graphs presented in the previous chapter found Multilayer Perception to do overall best and kept the lowest range of the confidence interval. We also found out that all three did equally well when we were forecasting at 72 hours ahead, but this was not expected according to the trend which was found. The difference in results from 12 and 24 hours also look similar, which can indicate that forecasting of either is similar based on weights. The lower the range, the better results.

Hours // Method	Linear Regression	Support Vector Machine	Multi-Layer
1 Hour	0.15 kWh	0.15 kWh	0.1 kWh
12 Hour	0.5 kWh	0.55 kWh	0.35 kWh
24 Hour	0.5 kWh	0.5 kWh	0.35 kWh
72 Hour	0.75 kWh	0.75 kWh	0.75 kWh

Table 2: Confidence interval range of method aligned with hours.

- **Which of the techniques will handle the event of the peak consumption value best?**

  We predicted the astronomical peak value with one hour forecasting using the methods presented. We found out that both Linear Regression and Support Vector Machine struggled during the events after the peak, while Multilayer Perception did not become affected other than the next two hours. Linear Regression and Support Vector Machine did both yield similar results as they both estimated the first hour forecasting to be higher, while the second hour forecasting to be lower. They both did also forecast too low at 11:00 and 13:00, the next day. The 12:00 value turned out to be too high compared to the rest of the predicaments. This means that Multi-Layer Perception handled the events of the peak consumption best.

■ **Will future predictions have a higher root mean squared error of current predictions?**

Hours // Method	LR	MLP	SVM
1 Hour	0.0375	0.0284	0.0383
2 Hours	0.0666	0.0449	0.0672
3 Hours	0.877	0.0538	0.0880
4 Hours	0.1018	0.0574	0.1017
5 Hours	0.1114	0.0592	0.1103
6 Hours	0.1181	0.0608	0.1164
7 Hours	0.1228	0.0619	0.1205
8 Hours	0.1255	0.0623	0.1227
9 Hours	0.1265	0.0627	0.1235
10 Hours	0.1273	0.0641	0.1247
11 Hours	0.1285	0.0658	0.1265
12 Hours	0.1298	0.0671	0.1286
13 Hours	0.1311	0.0682	0.1306
14 Hours	0.1320	0.0689	0.1321
15 Hours	0.1324	0.0696	0.1329
16 Hours	0.1329	0.0708	0.1337
17 Hours	0.1334	0.0717	0.1343
18 Hours	0.1334	0.0722	0.1347
19 Hours	0.1332	0.0727	0.1348
20 Hours	0.1330	0.0734	0.1349
21 Hours	0.1327	0.0741	0.1350
22 Hours	0.1327	0.0749	0.1352
23 Hours	0.1326	0.0754	0.1352
24 Hours	0.1331	0.0765	0.1354

Table 3: Root mean Squared Error, 1 to 24 hours ahead

Multilayer Perception has a lower root mean squared error than Linear Regression and Support Vector Machine. The two latter are showing similar results. All of the techniques have an increase of root mean squared error thus farther we forecast, but Multilayer Perception provides the best results.

- **How will other attributes such as weather affect the forecasting model?**

  In chapter 1.6, we saw there was a direct correlation between the temperature and power consumption rate. This indicates that learning or weighting the temperature could potentially be used to improve the model, however, the results are exactly similar when testing with and without temperature data across all forecasting methods. This means that the previous timestamp consumption values are vastly more preferred than the temperature values. The power consumption values are too significant to the temperature values.

9 Conclusion

We have attempted to test four different types of prediction models with real life power consumption data from Agder Energi. We have also tested and compared different types of computer based prediction such as standard prediction models, machine learning methods, and deep learning networks. We learned that Multilayer Perception with Machine Learning presented better results than the classic Linear Regression model and Support Vector Machine for Regression model. This includes both in range of confidence intervals and having the lowest root squared mean error. The addition of temperature entries did also not improve the models, as the weight of the previous timestamp values was too important. The Convolutional Neural Network was not able to compare with the standard models in this thesis, but improvement to the model and the dataset could make it viable.

Future work: We mentioned earlier that the previous master thesis [16] had data which lasted for three complete years, which is a larger dataset used in this thesis. There would be an opportunity to test with a larger dataset as it becomes available through the smart meter program. We could also look into the Convolutional Neural Network model to see if it can be improved as the results presented in this thesis did not reflect the results which has been found in other classification problems such as image recognizing. We also mentioned looking into other types of data to experiment with as soon as it becomes available.

References

[1] URL: https://en.wikipedia.org/wiki/Deep_learning.

[2] URL: https://en.wikipedia.org/wiki/Overfitting.

[3] Bernhard Schölkopf Alex J. Smola. *A tutorial on support vector regression.* 2003. URL: http://link.springer.com/article/10.1023/B:STCO.0000035301.49549.88.

[4] Geoffrey E. Hinton Alex Krizhevsky Ilya Sutskever. *ImageNet Classification with Deep Convolutional Neural Networks.* 2012. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.

[5] Zirguezi Alisneaky. *Kernel Machine.* 2013. URL: https://commons.wikimedia.org/wiki/File:Kernel_Machine.svg.

[6] Cornelis W. Oosterlee Anastasia Borovykh Sander Bohte. *Conditional Time Series Forecasting with Convolutional Neural Networks.* 2017. URL: https://arxiv.org/pdf/1703.04691.pdfl.

[7] Berland. *Illustration of linear regression on a data set.* 2007. URL: https://commons.wikimedia.org/wiki/File:LinearRegression.svg.

[8] Helge Langseth Boye A. Høverstad Axel Tidemann. *Short-Term Load Forecasting With Seasonal Decomposition Using Evolution for Parameter Tuning.* 2015. URL: http://ieeexplore.ieee.org/abstract/document/7042772/.

[9] *Data Mining with Weka (4.5: Support vector machines).* 2013. URL: https://www.youtube.com/watch?v=WVkD-jURBDg.

[10] *Definition for AIC.* 2017. URL: https://en.wikipedia.org/wiki/Akaike_information_criterion.

[11] *Feedforward ANN.* 2015. URL: http://www.extremetech.com/wp-content/uploads/2015/07/NeuralNetwork.png.

[12] Witten Frank Hall. *The Weka Workbench*. 2016. URL: http://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf.

[13] Michael Y. Hu Guoqiang Zhang B. Eddy Patuwo. *Forecasting with artificial neural networks:: The state of the art*. 1998. URL: http://www.sciencedirect.com/science/article/pii/S0169207097000447.

[14] *How are filters and activation maps connected in Convolutional Neural Networks?* URL: https://stats.stackexchange.com/questions/180850/how-are-filters-and-activation-maps-connected-in-convolutional-neural-networks.

[15] *How does confidence interval work?* 2017. URL: http://www.stat.osu.edu/~calder/stat528/Lectures/lecture21_6slides.PDF.

[16] Girma Kejela. *Short-term Forecasting of Electricity Consumption using Gaussian Processes*. 2012. URL: https://brage.bibsys.no/xmlui/handle/11250/137547.

[17] Kyoung-jae Kim. *Financial time series forecasting using support vector machines*. URL: http://www.sciencedirect.com/science/article/pii/S0925231203003722.

[18] *Loss function, backpropagtion*. 2017. URL: https://en.wikipedia.org/wiki/Backpropagation.

[19] MARC J. MAZEROLLE. *: Making sense out of Akaike's Information Criterion (AIC): its use and interpretation in model selection and inference from ecological data*. 2007. URL: https://pdfs.semanticscholar.org/a696/9a3b5720162eaa75deec3a607a9746dae95e.pdf.

[20] *Sequential minimal optimization*. 2017. URL: https://en.wikipedia.org/wiki/Sequential_minimal_optimization.

[21] *Single Node*. 2015. URL: http://makeyourownneuralnetwork.blogspot.no/2015/01/the-workings-of-neural-node.html.

[22] Jitendra Malik Subhransu Maji Alexander C. Berg. *Classification using intersection kernel support vector machines is efficient*. 2008. URL: http://ieeexplore.ieee.org/abstract/document/4587630/.

[23] Brian Kingsbury Tara N. Sainath Abdel-rahman Mohamed. *Deep convolutional neural networks for LVCSR*. 2013. URL: http://ieeexplore.ieee.org/abstract/document/6639347/.

[24] Goodwin Tufteland Ødesneltvedt. *Optimizing PolyACO Training with GPU-Based Parallelizatione*. 2016. URL: http://link.springer.com/chapter/10.1007/978-3-319-44427-7_20.