Luleå University of Technology
Department of Computer Science, Electrical and Space Engineering
PERCCOM Master Program

Master's Thesis in
Pervasive Computing & COMmunications for sustainable development

Rohan Nanda

A BAYESIAN APPROACH FOR FORECASTING HEAT LOAD IN A DISTRICT HEATING SYSTEM

2015

Supervisors: *Professor Christer Åhlund* (Luleå University of Technology)
Dr. Saguna Saguna (Luleå University of Technology)
Dr. Karan Mitra (Luleå University of Technology)

Examiners: *Professor Eric Rondeau* (University of Lorraine)
Professor Jari Porras (Lappeenranta University of Technology)
Associate Professor Karl Andersson (Luleå University of Technology)

This thesis is prepared as part of an European Erasmus Mundus programme PERCCOM - Pervasive Computing & COMmunications for sustainable development.

Co-funded by the Erasmus+ Programme of the European Union

This thesis has been accepted by partner institutions of the consortium (cf. UDL-DAJ, nº1524, 2012 PERCCOM agreement).

Successful defense of this thesis is obligatory for graduation with the following national diplomas:

- Master in Complex Systems Engineering (University of Lorraine)
- Master of Science in Technology (Lappeenranta University of Technology)
- Degree of Master of Science (120 credits) –Major: Computer Science and Engineering, Specialisation: Pervasive Computing and Communications for Sustainable Development (Luleå University of Technology)

ABSTRACT

Luleå University of Technology
Department of Computer Science, Electrical and Space Engineering
PERCCOM Master Program

Rohan Nanda

A Bayesian Approach for Forecasting Heat Load in a District Heating System

Master's Thesis

2015

83 pages, 35 figures, 17 tables.

Examiners: *Professor Eric Rondeau* (University of Lorraine)
Professor Jari Porras (Lappeenranta University of Technology)
Associate Professor Karl Andersson (Luleå University of Technology)

Keywords: energy; optimization; district heating; Bayesian network; forecasting; machine learning

The growing population in cities increases the energy demand and affects the environment by increasing carbon emissions. Information and communications technology solutions which enable energy optimization are needed to address this growing energy demand in cities and to reduce carbon emissions. District heating systems optimize the energy production by reusing waste energy with combined heat and power plants. Forecasting the heat load demand in residential buildings assists in optimizing energy production and consumption in a district heating system. However, the presence of a large number of factors such as weather forecast, district heating operational parameters and user behavioural parameters, make heat load forecasting a challenging task. This thesis proposes a probabilistic machine learning model using a Naive Bayes classifier, to forecast the hourly heat load demand for three residential buildings in the city of Skellefteå, Sweden over a period of winter and spring seasons. The district heating data collected from the sensors equipped at the residential buildings in Skellefteå, is utilized to build the Bayesian network to forecast the heat load demand for horizons of 1, 2, 3, 6 and 24 hours. The proposed model is validated by using four cases to study the influence of various parameters on the heat load forecast by carrying out trace driven analysis in Weka and GeNIe. Results show that current heat load consumption and outdoor temperature forecast are the two parameters with most influence on the heat load forecast. The proposed model achieves average accuracies of 81.23% and 76.74% for a forecast horizon of 1 hour in the three buildings for winter and spring seasons respectively. The model also achieves an average accuracy of 77.97% for three buildings across both seasons for the forecast horizon of 1 hour by utilizing only 10% of the training data. The results indicate that even a simple model like Naive Bayes classifier can forecast the heat load demand by utilizing less training data.

ACKNOWLEDGEMENT

I would like to thank Erasmus Mundus Master PERCCOM program and European Commission for providing funding to carry out this Master thesis. I had a good learning experience in my thesis work. I learnt a new domain of machine learning algorithms and their application to the specific problem of forecasting heat load in a district heating system. I am thankful to my supervisor Prof. Christer Åhlund, for supervising and guiding me from the start of my Masters. He gave me continuous feedback during weekly meetings for my thesis work. I also thank my supervisors Dr.Saguna Saguna and Dr. Karan Mitra for providing their in-depth knowledge and expertise to discuss the issues I faced during my thesis work. They were always available to discuss and give their opinions.

I would like to thank Prof. Eric Rondeau, Prof. Karl Andersson, Prof. Jari Porras and other professors in the PERCCOM consortium for giving me this opportunity to perform quality research during the Masters program. I also thank my two friends who also did their Master thesis at LTU, Ngo Manh Khoi and Baptiste Louis for helping me out on small issues during my thesis work.

I would like to thank my family members for always being there for me.

Lastly, I would like to thank Astrid, for making my stay memorable in Skellefteå.

Skellefteå, May 19, 2015

Rohan Nanda

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS AND SYMBOLS

EU	European Union
DHS	District heating system
CHP	Combined heat and power
BN	Bayesian network
ICT	Information and communications technology
OrPHEuS	Optimizing hybrid energy grids for smart cities
HVAC	Heating, ventilating, and air conditioning
DH	District heating
DC	District cooling
MLP	Multilayer perceptron
k-NN	k-Nearest Neighbours
EWD	Equal width discretization
AUC	Area under curve
EFD	Equal frequency discretization
EMD	Entropy minimization discretization
LD	Lazy discretization
BEMS	Building energy management system
WEKA	Waikato environment for knowledge analysis
GeNIe	Graphical network interface

1 INTRODUCTION

The energy demand in cities is continuously increasing due to the growing urban population. The European Union (EU) has set some targets to address this growing energy demand and to reduce carbon emissions. This chapter introduces the problem domain and highlights the importance of reducing energy consumption in buildings. We discuss the motivation to reduce and optimize the heating consumption in buildings. Then we discuss the thesis scope and the research challenges faced during the thesis work. We further explain the objectives and contribution of the thesis.

1.1 Introduction

The EU is committed to become an energy efficient and low carbon economy by setting climate and energy targets for the year 2020. These targets have the following three objectives [1]:

- achieving a reduction of 20% in the EU greenhouse gas emissions as compared to 1990 emission levels
- increasing the share of renewable energy to 20% of the total EU energy consumption
- improving the EU's energy efficiency by 20%

Buildings account for 40% of the total energy consumption and 36% of the total CO_2 emissions [2]. Therefore, reducing energy consumption in buildings can lead to reduction in CO_2 emissions. The use of renewable sources of energy in the buildings and reduced energy consumption can ensure security of energy supply for a long term. The technological advances in the ICT sector have the potential to reduce the CO_2 emissions from buildings by 15% [3].

District energy systems offer the advantage of reusing the waste energy with CHP(Combined heat and power) systems and thus reducing carbon emissions [4]. CHP systems provide electricity to the commercial and residential buildings. The waste heat from the electricity generation process is used to provide heat energy to the residential and commercial buildings by transporting heated water in pre-insulated pipes through a district heating network. It is estimated that on a global scale, district heating reduces existing CO_2 emissions by 3-4% [5]. The EU acknowledges the fact that cogeneration[1] has not been used to its full potential for energy savings. It is

[1] the process of simultaneous generation of heat and electricity in a power station

important to promote cogeneration for supplying heat and power with the added advantage of saving energy and reducing carbon emissions [6].

This thesis work is a part of the OrPHEuS project [7], which aims at improving the interaction among different energy grids like thermal grid, power grid and gas grid. This will be achieved by developing control strategies for simultaneous energy efficiency and energy savings among multiple energy grids, called as hybrid energy grids. In this project, the energy system setup in two cities, Ulm(Germany) and Skellefteå(Sweden) will be utilized for making control strategies for hybrid energy grids.

1.2 Research motivation

In the EU, space heating contributes to 68% of the total household energy consumption. The total heating end use(including space and water heating) contributes to 80% of the total household energy consumption as shown in Figure 1 [8].

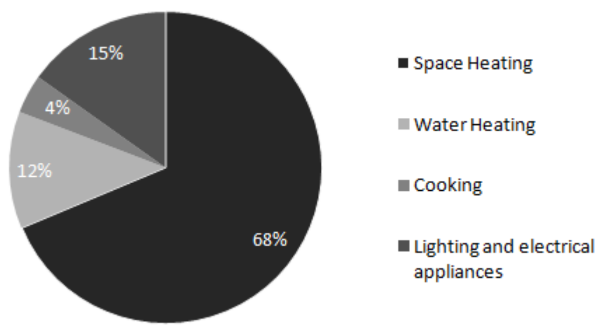

Figure 1. Household energy consumption by end-use in the EU in 2009 [8]

Therefore, it is imperative to focus on reducing and optimizing the heating consumption in buildings to achieve the targets of 2020, set by the EU. The high operational costs for operating and maintaining the energy plant also make it necessary to optimize the production and distribution of energy [9].

There is a need to forecast the energy demand to optimize the energy production process. The energy forecast provides an estimate of the future energy demand. The heat load demand varies throughout the day. Forecasting the energy demand(in our case, heat load demand) provides information about the energy resources needed to satisfy the future energy demand. A reliable heat load forecast thus leads to energy savings as the heat production unit is not generating excess heat. This prevents heat losses and improves the efficiency of the heat production unit.

Heat load forecast also provides information about the peak hours when the heat load demand is maximum. This enables the energy provider to be prepared for the peak hour usage. To manage the peak hour load, the CHP plant can be equipped with a thermal storage. At the time of low heat demand, the excess heat produced from the CHP plant can be stored in a thermal storage like a heat accumulator as shown in Figure 2. The stored heat energy from the heat accumulator can be discharged to the grid at the time of peak demand [10]. This process optimizes the operation of the CHP plant and also satisfies the peak heat load demand. The heat load forecast system provides information about the low and high heat load demand which enables the CHP plant to produce adequate heat energy as per the requirement. This reduces the excess heat energy production and leads to further optimization of the CHP plant.

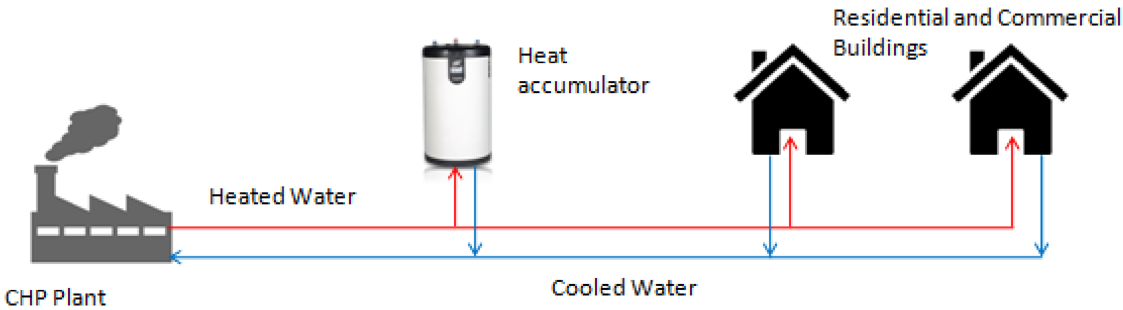

Figure 2. Heat accumulator used with CHP plant

An important issue pointed out in [11] is related to the delay in providing heating to the residential and commercial buildings. The production and supply of heat energy through a district heating network can take several hours. For example in the district heating grid in Skellefteå, the delay in the control loop is approximately 4-6 hours. Sometimes, this may lead to a situation where the supplied heat is excess or insufficient, due to the fast variation in the outside temperature. This problem can be resolved by a precise forecast of heat load demand.

1.3 Thesis scope

The thesis work utilizes the district heating data collected non-intrusively from three residential buildings located in the city of Skellefteå, Sweden. These residential buildings include multi-family apartments. Each building is equipped with a substation which connects the heat production network of the CHP plant to the internal heating system of the building. The location of the three buildings and the CHP plant is represented in the map shown in Figure 3. The weather forecast data containing the outdoor temperature is also recorded for each residential building. Thus, the district heating and weather forecast data consist of a set of parameters which may affect the heat load consumption in buildings. These parameters consist of district heating operational parameters like supply temperature, return temperature, flow rate and outdoor temperature. The recorded heat load consumption in the data, is the aggregated sum of space heat load and water thermal load.

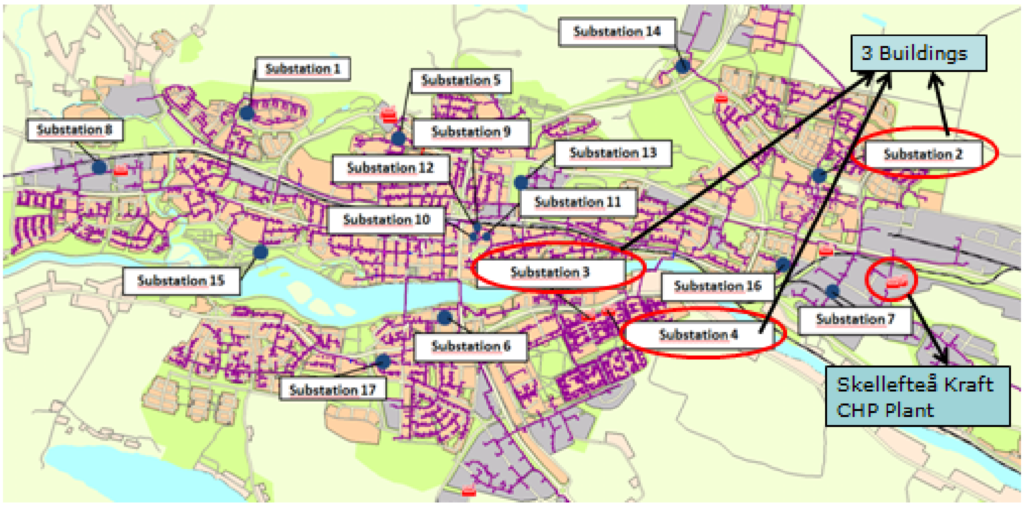

Figure 3. Three residential building substations and CHP plant in the city of Skellefteå, Sweden

There is a need to develop a model for seasonal forecast of heat load, due to significant variation in the outdoor temperature between seasons [12]. Since district heating is mostly in demand during the cold weather, we choose to develop a seasonal heat load forecast model for winter and spring seasons. The data considered for winter season is from 22 December 2013 to 28 February 2014. The data considered for spring season is from 1 March 2014 to 30 April 2014.

Some supervised machine learning techniques have been used for forecasting heat load consumption in buildings by utilizing weather and district heating parameters [13] [14]. Some of the commonly used machine learning methods are support vector machines, multiple linear regression, multilayer perceptron etc. However, Bayesian Network Classifiers [15] have emerged to be competitive with existing classifiers used for supervised learning models. They are based on probabilistic reasoning and thus incorporate uncertainty. Bayesian networks utilize probability theory to incorporate uncertainty by representing the conditional dependencies between different nodes of the network. Thus, they can model the uncertainty due to the presence of several influencing parameters on the heat load consumption. They offer a graphical model derived from the physical relationships between different nodes of the network. Bayesian classifiers also provide the added advantage of adding expert knowledge to the model and the ability to perform prediction and diagnosis simultaneously. These advantages provide motivation for this thesis work to build a model using Bayesian network by utilizing relevant district heating and weather forecast parameters over a period of two seasons.

1.4 Research challenges and objectives

The major research challenges identified in this thesis work include:

Large number of factors affecting energy consumption in buildings: Some of the factors affecting the energy consumption of a building include weather, occupant behaviour, the physical and thermal properties of the materials used in construction and HVAC(Heating, Ventilating, and Air Conditioning) system [16]. The study in [17] categorizes the factors affecting the heat load consumption into two classes: internal and external. The external factors include outdoor temperature, solar radiation, wind speed, wind direction etc. The internal factors are related to the district heating system and include supply and return water pressure, supply and return water temperature, the difference of supply and return temperature and circular flow. The presence of a large number of factors, make the problem of forecasting heat load a formidable research challenge. This thesis work addresses the challenge of identifying the parameters which have the most influence on the heat load forecast.

Forecast models need to scale for a large number of buildings in a city: Skellefteå Kraft as Sweden's fourth largest energy producer [18], has a district heating system which supplies heating to approximately 5000 buildings in the city. Since district heating data can be collected for only a limited number of buildings, the forecast models can be developed only for those particular buildings. The research challenge is to scale the heat load forecast from a few buildings to a metropolitan scale of a city consisting of 5000 buildings. It is a significant research challenge in the district heating grid. Within the scope of this thesis work, we are interested to study the heat load forecast for individual buildings. Scaling the heat load forecast from one building to a large number of buildings is not within the scope of this thesis.

Build an efficient forecasting model using less amount of training data: It is very expensive to equip 5000 buildings with sensors for collecting district heating and weather data. Also, storing, processing and analysing such a large amount of data to develop heat load forecasting models is a challenging task. Therefore, the challenge is to build a good forecasting model using the least amount of training data from a particular building.

The aim of this thesis is to investigate the performance of the Bayesian approach to develop heat load forecasting models. The objectives of the thesis are defined as follows:

- Study the impact of several parameters on the heat load forecast by developing probabilistic machine learning model based on a Bayesian network.
- Identify the parameters which have the most influence on the heat load forecast.

The major research question addressed in this thesis:

- Which parameters have the most influence on the heat load forecast and can they be used to forecast the heat load using the Bayesian approach?

1.5 Research contribution

The major contribution of this thesis work is to develop heat load forecast models for a district heating system using the Bayesian approach. A Bayesian network was developed for forecasting the heat load across two seasons for three residential buildings, in the district heating system of Skellefteå, Sweden. The impact of both district heating operational parameters (supply temperature, return temperature, flow rate, difference of supply and return temperature) and external parameters (like outdoor temperature) on the heat load forecast was studied over the horizons of 1, 2, 3, 6 and 24 hours across both winter and spring seasons.

Our results show that the proposed heat load forecast model achieved average accuracies of 81.23% and 76.74% for a forecast horizon of 1 hour in the three buildings for winter and spring seasons respectively. The model also achieves an average accuracy of 77.97% for the three buildings across both seasons for the forecast horizon of 1 hour by utilizing only 10% of the training data. Our results indicate that in case of a large number of buildings and large amount of data our model would be suitable for heat load forecast. We identified that outdoor temperature forecast and current heat load consumption are the two parameters with most influence on the heat load consumption.

1.6 Thesis outline

The rest of the thesis report is organized as follows: Chapter 2 presents the background on district heating systems and an in-depth related work on the techniques used for forecasting heat load demand at the consumption side. Chapter 3 discusses the background knowledge on Bayesian networks and our contribution of developing a heat load forecasting model for district heating systems using the Bayesian approach. Chapter 4 describes the results and analysis of the heat load forecast model on three residential buildings across two seasons. Chapter 5 presents the conclusion and future research directions.

2 BACKGROUND AND RELATED WORK

The previous chapter provides the necessary motivations, challenges and objectives within the scope of this thesis work. This chapter focusses on the background of district heating systems and the related work on heat load forecasting techniques. Section 2.1 provides the background knowledge about district heating systems and combined heat and power plants. It also discusses the heat load variation at the consumption side of the district heating system. Section 2.2 discusses the related work on heat load forecasting techniques at the consumption side of the district heating system. Section 2.3 follows up with a discussion about the related work. Section 2.4 discusses the motivation of choosing the Bayesian approach for forecasting heat load in a district heating system.

2.1 District heating systems

District energy systems have shown the potential to reduce CO_2 emissions and increase energy efficiency [5, 19]. The production and distribution of heat from a central plant rather than distributed plants, provides both ecological and economical benefits [19]. District energy systems may consist of district heating (DH) or district cooling (DC) systems. In this section, we will focus only on district heating systems.

District heating system is an infrastructure comprising of three major components [19], as shown in Figure 4:

1. **Heat production unit:** The heat production unit is a centralized source of heat energy. It may use a boiler or a heat accumulator or a CHP plant or any combination of these for heat energy production. The production unit can be powered by fossil fuels like natural gas or coal, geothermal energy, city garbage incinerators or any combination of these.
2. **Distribution network:** The distribution network comprises of the field-insulated and pre-insulated pipes which supply the hot water to residential and commercial buildings in the district. The return pipes transport the cooled water from the consumption unit to the production unit.
3. **Heat consumption unit:** The heat consumption unit comprises of the buildings where heating is required. Each building has substations working in parallel for distributing the hot water to different users in the building. The substations are equipped with heat exchangers, which transfer heat from the pre-insulated pipes to the internal heating network

of the building, as shown in Figure 5. The heat exchangers use the heated water to heat the radiators and tap water in the buildings. The cooled water from the building is then returned to the district heating plant to be reheated. The whole network forms a closed loop.

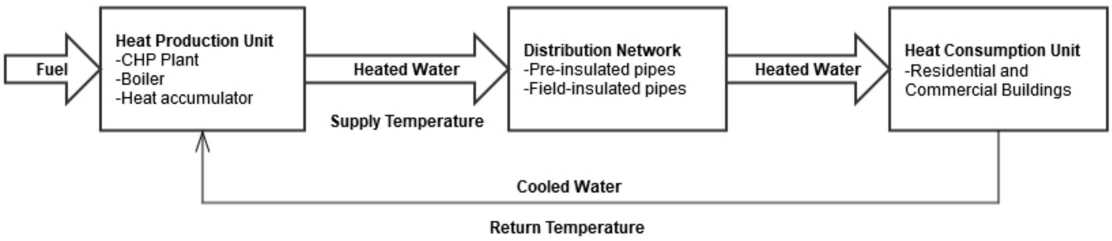

Figure 4. District heating block diagram [19]

2.1.1 Heat load in district heating systems

The heat load in district heating systems is the aggregate of the heat load consumption at the consumer side and the heat loss during distribution. Therefore, the estimated heat load demand at the production side should incorporate the heat loss occurring during the distribution phase. This relationship is explained by the following equation, discussed in [20].

$$\sum_{p=1}^{P} Q_{production} = Q_{loss} + \sum_{c=1}^{C} Q_{consumption} \tag{1}$$

Here, $Q_{production}$ is the total heat load produced at the production side. Q_{loss} is the heat loss during the distribution of heated water and $Q_{consumption}$ is the heat load consumption at consumer side. $Q_{consumption}$ varies constantly and depends on weather conditions, time of day and pressure applied from the production side. The heat distribution in the district heating grid is affected by the hydraulic and thermodynamic properties of the system [20].

$Q_{production}$ is dependent on four independent factors [12]:

1. The valves in heating radiators, ventilation air heating systems and the hot water taps.
2. The valves at the substations controlling the flow rate. They maintain constant temperature of hot water and supply temperature depending on the outside temperature.

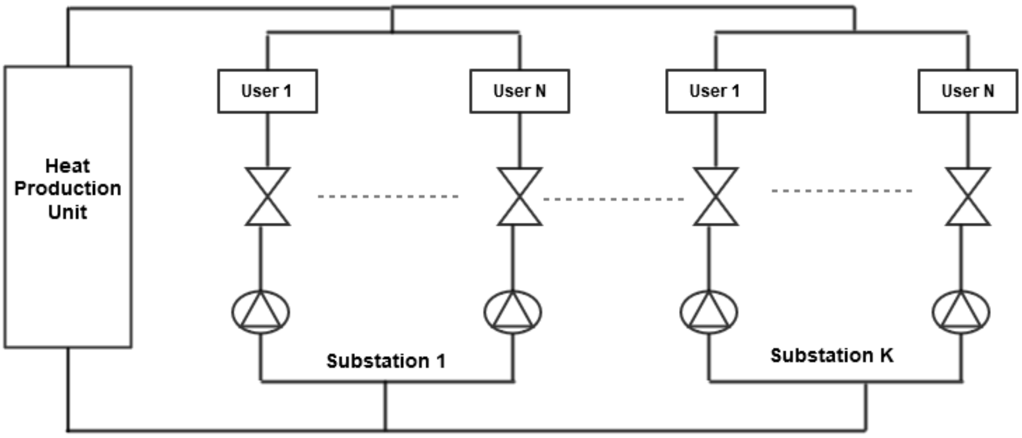

Figure 5. Internal heating network of the building [19]

3. The differential pressure control at the production side kept at a set point

4. The supply temperature at the production side. The supply temperature depends on the outdoor temperature.

From the points 1 and 2 it is easy to conclude that the heat load at production is highly dependent on the heat consumption at the consumer side. Therefore, heat load variation at consumer side results in heat load variation at the production side [12]. In order to optimize the district heating system, it is imperative to focus on the optimization of both production and consumer side. An accurate forecast of the heat load at consumption side provides valuable information about the heat load demand to the production side. The production side, in turn does not need to produce excess heat energy. This leads to energy savings. It also reduces the heat losses in the grid and lowers down the return temperature. Ultimately, the efficiency of the district heating network increases [21]. Therefore, forecasting the heat load at consumer side becomes a necessity to estimate the heat load demand at the production side and to achieve optimization in the district heating grid. Figure 6 describes the advantages of forecasting heat load through a flowchart.

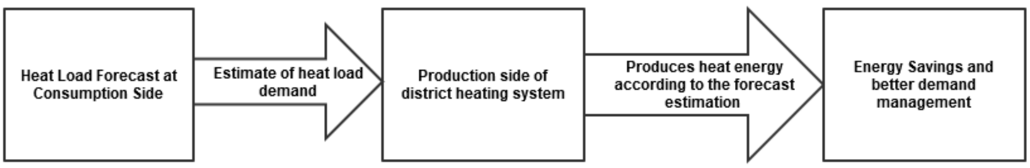

Figure 6. Advantage of forecasting heat load at the consumer side

The consumers can change the heat demand at the consumption side in two ways [19, 12]:

1. Constant temperature difference and variable flow rate: The temperature difference is the difference between supply temperature and return temperature. The consumer can increase the heat demand by increasing the flow rate. In this case, the increase in heat demand propagates to the production side at a very high speed, around 1000 m/s.

2. Variable temperature and constant flow rate: In this case, the consumer can increase or decrease the temperature difference to vary the heat demand. The change in the heat demand reaches the production side at a speed equal to the flow rate of water in the district heating system, around 1-3 m/s.

In case 1, it takes only a few seconds for the change in heat demand to propagate to the production side. While in case 2, it takes hours for the heat demand to propagate to the production side [12] in large district heating systems. Then the production plant alters the heat load according to the change in demand and the new heat load is again propagated at the speed of the flow rate to the consumption side. This explains the 4-6 hours delay in the control loop of the district heating systems.

2.1.2 Study of heat load variation at district heating consumption side

As explained in Section 1.4, forecasting heat load at consumer side is a challenging task. This is because the heat load consumption in buildings is affected by different kind of factors like internal (supply temperature, return temperature, the difference of supply and return temperature, supply and return water pressure), external (outdoor temperature, solar radiation, wind speed and wind direction), the occupant behaviour, the physical and thermal properties of materials used in building construction and HVAC [17, 16]. Forecasting the load at consumer side is sometimes avoided due to the high stochastic pattern in the heat load [22]. This stochastic nature of the heat load at consumer side can be captured by building different individual models [22].

Gadd et al. [12] discuss seasonal and daily heat load variation in a district heating system. Heat load variations between different seasons is due to large differences in outdoor temperatures(for example between winter and summer) and heating comfort indoors. There are some parameters which lead to heat load variations in different seasons. Wind can suddenly increase the heat load demand due to air leakage, as the warm air is replaced by cold air. Solar radiation decreases the heat load demand by increasing the temperature of the outer walls. The outer walls and windows act as a green-house by decreasing the flow of heat from inside of the building to outside. The occupant behaviour during different seasons also leads to heat load variations. In winters, people stay in their houses most of the time and consume more heating and hot water. On the other hand, during summer holidays many people go on vacations and heating and hot water are hardly consumed [12].

Occupant behaviour also seems to be one of the key reasons for daily heat load variation. This behaviour can be categorized as individual and collective occupant behaviour. The usage of hot water during shower is an example of individual occupant behaviour [12]. The usage of heating and hot water in an office during working hours is an example of collective behaviour. The absence of people in the office during night and weekends adds to the variation in the heat load. The lower outdoor temperature at night and the decrease in solar radiation with the daytime generate heat load variations everyday.

The daily variations in the heat load demand increase the cost of heat production due to the use of expensive fuels during the peak load demand [23]. The variation in the heat load demand has different consequences on the district heating production unit and the consumers. If a situation arises where the heat produced is insufficient then all consumers are not affected in the same manner [12]. The supply temperature is maximum when the hot water leaves the production unit. In the distribution network of pre-insulated pipes the supply temperature decreases gradually with distance. The houses near the production unit receive water with a high supply temperature and their heating demands are satisfied. However, the houses at the boundary of the district heating network receive water with much lower supply temperature and thus will receive very less or no heating in case of insufficient heat production. To handle these kind of situations, the district heating production unit generally has to produce excess heating energy to ensure heat supply to all the consumers in the district heating network.

Some strategies help in handling the heat load variations. Heat storage strategies in the district heating network can be a viable solution. By heating the water to a supply temperature higher than required, heat energy can be stored in the district heating network. Utilizing building masses along with heat storage in the heating radiator pipeline system serves a good short term heat storage solution for daily heat load variations and heat production failure. This solution is also cost efficient as compared to constructing expensive heat storage accumulators [23].

2.1.3 Combined heat and power plants

Cogeneration, also called Combined Heat and Power (CHP), is the process of simultaneous generation of heat and electrical energy in a power station from a single fuel [19]. In traditional power plants, a large amount of fuel's energy content is lost in the form of waste heat discharged by the plant. These power plants are capable of converting only 35% of the fuel's available energy into electrical energy. This results in a low efficiency and incomplete utilization of the fuel's energy. The energy losses during transmission and distribution of electricity to the user further add to the problem.There is a need to minimize the heat losses and make electricity generation more efficient. This can be achieved by developing on-site and near-site power generation plants in the form of CHP systems.

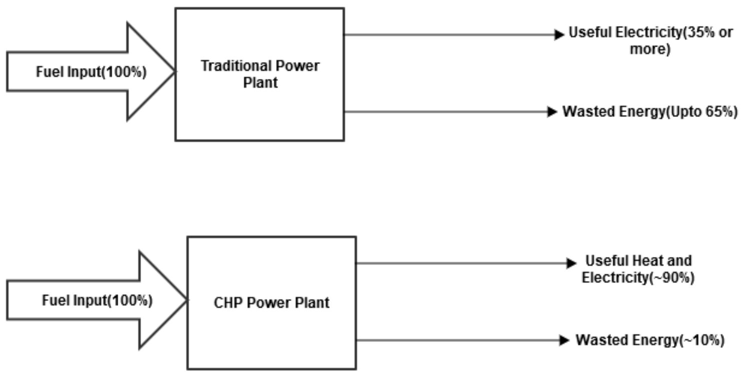

Figure 7. Energy efficiency comparison of CHP and traditional power plant [19]

The CHP systems can utilize upto 90% of the fuel for production of useful heat and electricity. This considerably increases the efficiency of utilization of energy and lowers the cost of operation. CHP systems utilize the waste heat from the electricity generation process and use it to provide heating. Fossil fuels, natural gas or renewable sources like biomass can be used in CHP plants. The steam produced by burning the fuel is used to rotate the turbine to produce electricity. The remaining heat is then collected in a heat recovery boiler and is used to heat the water [24]. The heated water is transported to the households and industries through the district heating network. Cogeneration provides an efficient way of power generation which leads to energy optimization and reduction of carbon emissions.

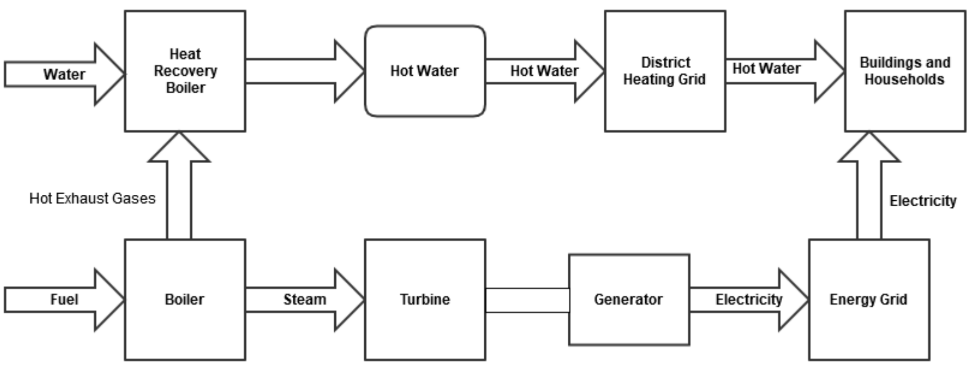

Figure 8. CHP Plant Operation [24]

2.2 Techniques for forecasting heat load in district heating systems

This section reviews the related work regarding the heat load prediction in buildings. A survey presented in [25] summarizes various types of classifications used for building energy estimation methods in the literature. The author further proposes a high level classification comprising of statistical, hybrid and engineering methods for building energy estimation.

Another broad classification discussed in [25] defines two types of approaches for energy estimation in buildings. The forward approach utilizes the equations modelling the physical behaviour of the system to predict the energy demand. These models are subject to the availability of building design data [26]. They require details of building descriptions, properties of building materials, building geometry etc. This information is easy to extract for the newly constructed buildings. However, it is really challenging to retrieve this information for old or existing buildings. The data-driven approach utilizes the data containing the records of input and output variables which govern the performance of the system [25]. Data-driven methods are dependent on the availability of data collected from buildings in the district heating network. These techniques are not dependent on the building design data [26]. However, they are highly dependent on the quality and quantity of the available district heating data. An advantage of data-driven techniques is that the district heating data can be collected for both old and new buildings, as it is not dependent on building design. The data can be collected intrusively or non-intrusively depending on the situation. In the domain of thermal load forecast, data-driven techniques have the advantage of identifying and discovering models from large datasets [14]. Additionally, they offer the flexibility of updating the existing model when new data arrives [14].

Since the objective of the thesis is to forecast the heat load by utilizing the district heating data collected from three residential buildings, it is imperative to look into the data driven techniques in this section.

The work discussed in [14] used a wide range of data mining algorithms to forecast the steam load in a building by utilizing weather forecast parameters like outdoor air temperature, humidity, solar radiation, barometric pressure, wind speed, rain gauge and wind position. Some of the algorithms used were classification and regression trees, random forest, support vector machines, multi-layer perceptron (MLP), MLP ensemble and k-nearest neigbour (k-NN). The steam load forecast model is built on the steam consumption and weather data from 2004 to 2007. The authors use correlation coefficients and boosting tree algorithm to remove irrelevant parameters influencing the steam load forecast and to reduce the dimensionality of data [27]. The boosting tree algorithm builds a sequence of trees and each tree learns instances misclassified by previous trees on the basis of prediction error [14]. The authors observed that reducing the number of input parameters helps in achieving a stable prediction accuracy and reducing the variance. The dataset of one year was divided into 4 parts, comprising of 3 months each to take into account the effect of seasons. The authors concluded that MLP ensemble method performs the best on mean absolute error and mean absolute percentage error metrics [14].

The research work presented in [28] discusses the importance of short-term thermal load forecast in district heating systems. The short-term thermal load forecast makes it possible for cogeneration plant operators to respond promptly to unforeseen random events like exceptionally high load demand. Artificial neural networks are used for forecasting because of easy input data selection and good convergence rates. Input parameters like outdoor temperature, pressure and flow were utilized for heat load forecast up to a horizon of 3 hours. The model was developed for a building complex in a university campus in Poland. The mean absolute percentage error of the model lies in the range of 3-5 %. The prediction accuracy drops with the increase in the forecast horizon. The author suggests that the accuracy achieved using neural networks reflects the heating characteristics of the building. The major disadvantage of using neural networks is requirement of a large training set and a good correlation between the input and output parameters of the network [28].

Another work in [21] also signifies the importance of short-term heat load forecast for controlling the operation of a district heating network. The authors used support vector regression for heat load prediction for one heating substation for horizons of 15, 30, 45 and 60 minutes. District heating operational parameters like supply temperature, return temperature and flow rate were utilized along with outdoor temperature, current load and historical load as input parameters for the heat load forecast. The heating substation had irregular supply of heat in the night. The model developed through support vector regression was unable to capture this effect. The authors addressed this issue by adding a dummy variable to define the state of the district heating operation. Addition of this variable helped in improving the prediction accuracy of the model. The prediction accuracy decreased with the rise in the prediction horizon. This observation is similar to the work presented in [28]. It was also observed that historical values of heat load improved the prediction accuracy. On the other hand, historical values of outdoor temperature decreased the prediction accuracy to a great extent. The major challenge while using support vector machines is the choice of a kernel function [29] and adjusting the values of two constants by the user [26].

The work presented in [13] discusses the performance comparison of four supervised machine learning methods for forecasting heat load in residential buildings. These methods include: support vector regression, regression tree, feed forwards neural network and multiple linear regression. The authors studied the effect of internal and external factors [17] on the heat load forecast and concluded that internal factors have little impact on a 24 hour horizon for load forecast. Support vector regression achieved the best accuracy. The authors studied the heat load forecast for 1, 3, 6, 12, 18 and 24 hour horizons [13]. In most cases, the forecast accuracy drops with the increase in horizon from 1 to 18 hours. This finding is similar to the techniques discussed in [28] [21]. However, the increase in forecast accuracy from 18 hour to 24 hour forecast horizon justifies the presence of a periodic daily pattern in the heat load consumption in buildings. This finding illustrates that the model is able to detect the daily consumption pattern which could be useful to study the effects of occupancy and user behaviour on the heat load forecast.

The research work explained in [30] focusses on using a distributed approach to optimize the heat load consumption at consumer side of the district heating system. The author proposes a multi-agent system to optimize the space heating consumption at substations. The software agents are deployed at a computer and can communicate with a substation. These agents were successfully able to reduce the household heating consumption by 15% if it reached a certain threshold. This strategy achieves local heat energy optimization for a particular substation. The authors further proposed a methodology to reduce the heating consumption by setting a global threshold for two substations. In this case, the agents at a particular substation first reduce their local consumption and later request consumption reduction from other substations. This multi-agent approach proves to be successful in reducing the peak consumption. The major drawback is that just reducing the heating consumption by 15% results in the temporary shutdown of the heating radiator because the return temperature of water remains high. The applicability of this approach is probably not suitable for domestic hot water consumption because it would decrease the comfort level of the consumers. However, this technique offers a choice to the district heating grid operator to make control decisions to optimize heat consumption in substations locally. But it needs to be tested on a bigger scale to gain more knowledge about its merits and demerits [30].

An online machine learning approach [31] for heat load forecasting in 16 single-family houses utilizes weather forecast and local climatic parameters. The local climate parameters include ambient temperature, wind speed and global radiation. The heat load signal is disaggregated into space heat load and hot water heat load. This is done to separate the high varying hot water heat load from a more slowly varying space heat load. Thus, both heat loads are forecasted independently. The absence of indoor temperature is accounted for by the addition of a recurring daily pattern to model the behaviour of occupants. This includes addition of parameters like time of day, working days and weekends. Adaptive linear time-series models representing the physical characteristics of building heat dynamics and climate variables were developed utilizing the data collected from various households. A forward selection approach was used where an input parameter is added to the model at each step, eventually obtaining a model with the maximum prediction accuracy. The forecasting results of a particular house indicate that the huge difference between the heat load consumption in night and day was difficult to be captured in the model. This huge difference in night and day heat consumption can be attributed to the manual control of heating by the occupants. The model was able to forecast correctly the hot water load in case of a regular patter of consumption. However, an irregular consumption of hot water was not properly captured in the model. The majority of forecast errors were due to high variation in the load consumption, changing occupant behaviour and uncertain weather forecasts over a long horizon period [31].

The work presented in [32] studied the influence of various parameters like building envelope thermal resistance, glazing surface and distribution on the façade, heat loss area and heated volume, air change rate and indoor heating set point temperature on the heat consumption in buildings. These parameters were found to be correlated with the heating demand in buildings. The authors developed energy prediction and weather modules for the buildings using building simulation software. The building simulation software was used to construct a dataset for developing regression models. The continuous values of building energy parameters and the prior knowledge about the influence of input parameters on load consumption offers an ideal case for applying regression models for forecasting heat load demand. To simplify the model and reduce the prediction errors, the authors finalize three input parameters which affect the heat consumption. These parameters include building global heat loss coefficient, south equivalent surface and temperature difference between indoor heating point and sol-air temperature. The model is validated on the on-site monthly data from 17 blocks of flats with different orientations and thermal characteristics. The authors computed the values of input parameters from the collected data and obtained an average error of 20.2% for the 17 buildings tested. This research work illustrates that by using simple regression models and simulated data, it is possible to predict the heat load consumption in real world buildings with a good accuracy. Also, utilizing building energy parameters from building simulation software and using them with weather parameters provide a good way of using both simulation and statistical models for achieving heat load forecast [32].

Grzenda et al. [20] discussed the importance of predicting the heat load consumption at consumer-side for providing the necessary data for hydraulic calculations in district heating system. These hydraulic calculations include flow, pressure and temperature in the district heating system. The authors identified the heat consumption profiles for various consumers from the monthly billing database by applying a self-organizing map network. The data collected in this way is divided into two datasets: group dataset and global dataset. The group dataset consists of average heat load consumption from the consumers belonging to that group, average heat load consumption by all the consumers, outdoor temperature, day of week and time of day. The global dataset contains all the parameters contained in the group dataset except the average heat load consumption from the consumers belonging to a particular group. Since the global dataset contains average data from all the consumers it has less variation in consumption as compared to the group dataset. Both datasets belong to the same time period to ensure unbiased comparison of the prediction models. An evolutionary construction of multilayer perceptron has been used for training the prediction model. The prediction error on the testing dataset of group models was 31% lower than global models. This is because it is difficult to achieve a global prediction for all consumers by training only a subset of the total consumer profiles. This study makes an interesting case of grouping the substations with similar heat consumption into groups. This helps in spanning the heat load prediction to a large number of substations and building a heat load profile for each group [20].

Grosswindhager et al. [22] used an autoregressive integrated moving average model for modelling the system heat load at production side in the district heating network. This technique assumes that the future heat load can be forecasted by a linear combination of past values in the time series. The authors use the autoregressive integrated moving average to model the time series of heat load with a seasonal pattern. The model is embedded in the framework of state space models and forecasting of heat load is carried out using Kalman Recursions. The authors believe that their choice for autoregressive models is motivated by the fact that for short term forecasting, the influence of weather forecast is captured in the heat load time series. Therefore, a univariate seasonal autoregressive integrated moving average model is considered sufficient for a short term heat load forecast for 12 to 24 hours horizon. The forecast results over a period of one day show 4.4% mean absolute percentage error for forecasting. The accuracy for forecasting was improved by adding the real values of outdoor temperature (not forecast ones) as a piecewise linear function [22]. This approach can be used to develop more complex models which can capture variations in heat load due to weather, user behaviour and other factors which have an impact on the time series of the heat load.

Vlachopoulou et al. [33] discuss the importance of forecasting the load consumption in a smart grid environment for energy providers and distributors. Smart grids utilize demand response strategies which focus on providing dynamic energy supply to the changing energy demand in the grid. The authors propose a dynamic Bayesian network for forecasting the aggregate end-use water heat load consumption in residential buildings. Dynamic Bayesian networks offer the advantage of relating the evolution of a set of variables over time, in a temporal analysis. The proposed dynamic Bayesian network has been built on the simulated data produced by GridLAB-D [34] simulation software. Bayesian networks are discussed in detail in Section 3.1.2 of the next chapter.The authors utilize expert knowledge to develop the structure of the dynamic

Bayesian network comprising of the parameters influencing water heat load consumption in residential areas. The input parameters modelled in the dynamic Bayesian network include outdoor temperature, solar radiation, time of day, season, water heater efficiency, hot water usage and thermostat set point temperature. The data obtained from GridLAB-D has a resolution of 5 minutes. The Bayesian model uses two time slices in the network. The parameters like water heater efficiency and thermostat set point temperature which remain constant over time are only used in the first time slice. The discretization of continuous input variables was carried out using expert knowledge and experimentation with the dynamic Bayesian network. The model was trained on the simulated data of 1000 houses in a residential area from December to March. One week in February was excluded for testing the model. The hourly heat load forecast was computed using two methods: hard forecasting and soft forecasting. The hard forecast is obtained by selecting the load class with the highest probability of classification. The soft forecast is obtained by computing the average of various load class values weighing by their probabilities. The average forecasting error was approximately 50kW. The soft forecasting technique produced mildly better results than the hard forecasting technique. The proposed Bayesian network provides a methodology to model the physical relationship between various parameters influencing the heat load. However, the authors do not provide much clarity about the forecasting accuracy and commonly used metrics for measuring the prediction errors. Also, the model still needs to be validated on the real world data [33].

2.3 Discussion

In the literature review, we have discussed the use of various machine learning techniques for forecasting the heat load in a district heating network. These include artificial neural networks, support vector machines and regression models. Other techniques discussed for estimating the heat load are adaptive linear time series models, autoregressive integrated moving average models and dynamic Bayesian networks. A distributed multi-agent technique which focusses on reducing the heat load consumption at consumption side was also discussed. Table 1. provides the advantages and disadvantages of using these machine learning techniques.

Some interesting conclusions which can be derived from this literature review:

1. The forecast accuracy decreases with the increase in the forecast horizon. However in some cases, the accuracy increases with a 24 hour forecast, indicating the presence of daily heat load pattern.

2. Most of the forecasting models utilize outside temperature as the major weather forecast

parameter which affects the heat load demand.

3. Many forecasting models try to model the effect of user behaviour on the heat load consumption by adding input variables like time of day, day of week and month of day.

4. Different machine learning and statistical approaches have been used for forecasting the heat load in a district heating system. Each technique has its own limitations and benefits. The choice of use depends on the specific load forecasting problem and the specific outcome of the problem.

5. Apart from the internal and external factors [17] influencing the heat load consumption in buildings, building design parameters like global heat loss coefficient, south equivalent surface and temperature difference between indoor heating point and sol-air temperature also have an impact on the heat load consumption [32].

Table 1. Advantages and disadvantages of some machine learning techniques used for forecast [26, 35, 36]

Machine Learning Techniques	**Advantages**	**Disadvantages**
Multiple linear regression [13]	Simple forecasting technique suitable for input and output variables with strong linear relationship	Requires large amount of training data Not suitable in case of non-collinearity in data
Artificial neural network [28][13][20]	Can detect complex non-linear relationships between input and output variables	Have limited ability to identify possible causal inference between input and predictor variable Requires greater computational resources
Support vector machine [14] [21][13]	Less prone to overfitting because it uses structural risk minimization principle for optimization problem	Determining the kernel function
Bayesian networks [33]	Deal with uncertainty caused by scarce and sparse data Easily extensible: new evidence can be added	No universally accepted method to design network from data Can only exploit causal influences that are recognized by the person who designed the network

2.4 Choice of a forecasting technique : The Bayesian approach

In smart energy grids, the buildings are equipped with smart meters to measure the energy consumption. Smart meter readings at consecutive 15 minutes interval for each household or commercial building generate large amount of data. It is estimated that, it could result in a 3000 fold increase in the production of data as compared to the monthly data recordings [37]. This data explosion leads to the problem of analysis of a large amount of data for analysing the user's energy consumption patterns.

The CHP plants in Skellefteå supply heating to around 5000 buildings. We have obtained the heating consumption data from 17 buildings at a resolution of 1 minute. If the heating consumption data is collected from all the 5000 buildings at 1 minute resolution, then it will be very challenging to analyse this large amount of data for forecasting the heat load demand. Therefore, there is a need to choose a machine learning technique which can use least amount of training data to build accurate forecast models for estimating the heat demand. Some Bayesian techniques have shown to outperform regression techniques [38] in case of less training data. Also the research work presented in [39] concludes that Naive Bayes algorithm achieves very similar predictive accuracy as support vector machines, C4.5 and C4.4 Decision trees. Also the average area under curve (AUC) for Naive Bayes was similar to support vector machines and C4.4 decision trees [39].

Naive Bayes model is a simple Bayesian model as compared with complex and computationally intensive models like support vector machines and artificial neural networks. Yet, it has been known to show similar prediction accuracies with the likes of Decision trees and support vector machine models [39]. It also has the advantage of outperforming logistic regression in case of less training data [38]. Bayesian networks also have advantage over neural networks. While training neural networks it is difficult to understand if all the domain specific knowledge present in data is being utilized or not. Also, it is difficult to estimate what impact do certain parameters have on the predictor variable. However in Bayesian networks the impact of each parameter on the predictor variable can be easily estimated. Thus, Bayesian networks can guarantee to utilize all the input parameters and features present in data [36]. All these factors motivate us to choose a Bayesian approach for forecasting heat load in this thesis work.

2.5 Summary

In this chapter, we discussed the theory and background knowledge about district heating systems. We studied the operation of district heating systems and combined heat and power plants.

We also examined the reasons for heat load variation in buildings due to several parameters. A detailed literature review about various heat load forecasting techniques was also discussed. Based on the research challenges identified in the first chapter and the discussion in Section 2.4 we choose the Bayesian approach for developing heat load forecast models for a district heating system.

3 BAYESIAN NETWORK FOR HEAT LOAD FORECAST

In the previous chapter, we explored the background knowledge in district heating systems. We observed that forecasting the heat load consumption in buildings is imperative for energy optimization in district heating systems. Further, we investigated various machine learning techniques used in the domain of heat load forecasting. In this chapter, we investigate the theory and application of Bayesian networks. We also discuss about the discretization techniques for the available continuous data. Finally, we present the proposed model for heat load forecast in a district heating system with the application of Bayesian networks.

3.1 The Bayesian approach

Statistical study requires collecting, analysing and interpreting data to make inference about one or more physical processes that give rise to a data model [40]. A statistical model is built using the data to make inferences about some parameters which describe certain characteristics of the data. There are two statistical approaches for making inferences, namely the frequentist and the Bayesian approach. The major difference between these two approaches is the way of interpretation of probabilities. The frequentist approach interprets probability as a limit of frequency in a fairly large number of trials. This approach is generally applied to events which are infinitely repeatable [40]. For example, if a fair coin is tossed an infinite number of times, both heads and tails will occur with equal frequency. The probability of each event will be $1/2$ in the long run of infinite number of coin tosses. In the Bayesian statistical approach, the probability of an event is described as measure of the degree of belief of a parameter assessing the uncertainty of a particular event. This probability can be assigned to an event (whose outcome is uncertain) and including the events which are not repeatable [40]. In this way, Bayesian statistical approach provides a methodology to assess the outcome of uncertain and non-repeatable events. For example, assessing the probability that the government will increase the health care budget in the next meeting.

3.1.1 Bayes' theorem

Bayes' theorem provides a mathematical explanation to update the existing beliefs about an event by considering the new evidence. It is used for probabilistic inference [41]. To understand Bayes' theorem we first discuss conditional and prior probability in brief.

Conditional probability of an event *B*, provided that event *A* has already occurred is represented as *P(B|A)*. It is defined as belows:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{2}$$

Prior probability of an event *A* describes how likely is the occurrence of *A* without any evidence. It is represented as *P(A)*. With Bayes' theorem we can estimate the conditional probability of event *A*, given event B has already occurred, *P(A|B)* as follows [42]:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{3}$$

Thus, Bayes' theorem relates the prior probability and conditional probability of two different events for probabilistic inference. For many real world problems it is quite costly in terms of computation costs to apply Bayes' theorem. However, with the presence of a visual representation like Bayesian networks and availability of software tools to model them, it has become less complex to apply Bayesian statistics to real world problems. In the next sub-section, we discuss Bayesian networks. Bayes' theorem is also discussed again in sub-section 3.1.4.

3.1.2 Bayesian networks

A Bayesian network is a directed acyclic graph which models the probabilistic relationships between the random variables [41].

- **Node:** Each node of the graph represents a random variable. The random variable can be continuous or discrete.
- **Edge:** An edge is a directed link that connects two nodes. The edge represents the conditional dependency between the two nodes.

Consider two random variables X and Y represented by the Bayesian network shown in Figure 9. The arrow from node X to node Y , indicates that X is parent of Y. This arrow indicates that X has a direct influence on Y. This implies a causal relationship between two random variables. The conditional probability distribution of node Y can be represented by $P(Y|X)$. So, a random variable X_i has a conditional probability distribution $P(X_i|Parents(X_i))$ that expresses the effect of the parent node on the node X_i.

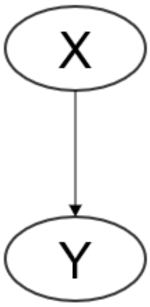

Figure 9. A simple Bayesian network consisting of two random variables

3.1.3 Conditional independence in Bayesian networks

It is very important to understand the conditional independence property of nodes to grasp the probabilistic nature of the Bayesian networks. A node in a Bayesian network is dependent only on its parent nodes. In simple terms, a node is conditionally independent of all the other nodes in the network, except the parent nodes [41]. Consider the Bayesian network shown in the Figure 10.

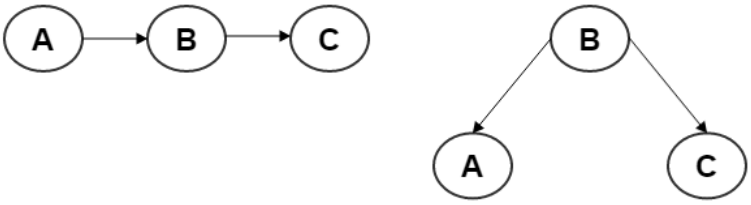

Figure 10. Nodes A and C are conditionally independent

The random variables A and C are conditionally independent given the variable B. This relationship is expressed by the following equation [43].

$$P(A|B,C) = P(A|B) \tag{4}$$

3.1.4 Modelling probability distribution in Bayesian networks

We consider a set of random variables $x_1, ..., x_n$ represented by X_i in a Bayesian network. The joint probability distribution of the network is computed as the product of all the conditional probabilities specified in the Bayesian network as illustrated by the equation below [41]:

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i)) \tag{5}$$

Here $parents(X_i)$ represent the parent node of node x_i. Each node in the network has a conditional probability table. We further explore this concept with the following example. There are two events which can cause the grass to be wet : the sprinkler or rain. Further, the use of sprinkler and the occurrence of rain depend on weather being cloudy or not. This situation is represented by the Bayesian network in Figure 11 [44]. We consider the random variables *C, S, R, W* for cloudy, sprinkler, rain and wet grass respectively. By applying Eq. 5 to this problem, we can compute the joint probability of all the nodes in the Bayesian network as follows [44]:

$$P(C, S, R, W) = P(C) * P(S|C) * P(R|C) * P(W|S, R) \tag{6}$$

This equation is written after considering the conditional independence property.

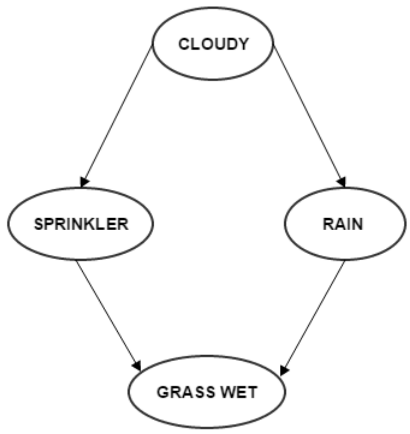

Figure 11. A Bayesian network example[44]

After constructing the Bayesian network, it is feasible to do probabilistic inference using Bayes' rule. For this we need to go thorough the following terminologies used in Bayes' rule.

Prior probability: The prior probability associated with an event is the degree of belief assigned to that event in the absence of any evidence. The prior probability is meant to attribute uncertainty to the occurrence of that event. A prior probability should only be used when there is no evidence available. When new information becomes available then it is useful to consider the conditional probability of the event given the new evidence. For example, the prior probability of the weather being cloudy, rainy or sunny can be defined by the following equation where

0.5, 0.3 and 0.2 are the respective probabilities of weather being cloudy, rainy and sunny [41].

$$P(Weather) = [0.5, 0.3, 0.2] \tag{7}$$

Posterior Probability: The posterior probability associated with an event is the conditional probability assigned to it after taking into account the available evidence [41]. Consider a random variable *R* which takes a particular value given the evidence E. The prior probability of *R* in absence of any evidence is given by $P(R)$. The posterior probability of *R* given the evidence *E* is $P(R|E)$.

The **likelihood function** is defined as the probability of the evidence, E given the random variable, R. It is represented by $P(E|R)$. The **marginal likelihood**, $P(E)$ gives the prior probability of the evidence.

Bayes' rule states that the posterior probability is a consequence of the prior probability and the likelihood function for the evidence observed from the data [45].

$$posterior\,probability = \frac{likelihood * \; prior\,probability}{marginal\,likelihood} \tag{8}$$

Further, using the notations we can re-write this equation [45] as follows:

$$P(R|E) = \frac{P(E|R) * P(R)}{P(E)} \tag{9}$$

Now, coming back to the example discussed in Figure 11, we can infer the probabilities of sprinkler and rain making the grass wet, given the evidence that the grass is wet. In this case,
Evidence: Grass is wet, W
Random variable 1: Sprinkler, S
Random variable 2: Rain, R

The important point to consider here is that due to conditional independence, Grass Wet variable is not dependent on Cloudy variable.

The probability of rain making the grass wet is given by:

$$P(R|W) = \frac{P(W|R) * P(R)}{P(W)} \tag{10}$$

Similarly, the probability of sprinkler making the grass wet is computed by the following equation:

$$P(S|W) = \frac{P(W|S) * P(S)}{P(W)} \tag{11}$$

So, in this section we learnt about Bayesian networks and how they can be utilized for making inferences about some parameters. In the next section, we discuss about representing uncertainty in Bayesian networks.

3.1.5 Representing uncertainty in Bayesian networks

Uncertainty is generally classified into two types: epistemic and aleatory uncertainty [46]. Epistemic uncertainty arises due to the lack of knowledge. It may refer to the limited knowledge we have about a particular system. It is possible to reduce epistemic uncertainty by gaining more knowledge about the system. For example, a person can be uncertain about the population of Stockholm, but his uncertainty can be reduced by referring to a credible reference. Aleatory uncertainty arises due to randomness. It is a part of natural processes of observation. The outcomes of tossing a coin and rolling a die are examples of aleatory uncertainty. Aleatory uncertainty is irreducible because of the variability among the underlying variables [47].

These uncertainties can be represented in the Bayesian network by modelling the probability distribution as per the Bayes rule discussed in the previous section. The prior probability is the degree of belief assigned to an event in absence of evidence. This implies prior probability refers to the uncertainty arising due to lack of knowledge. Thus, prior probability describes epistemic uncertainty. In statistics, aleatory uncertainty is defined as a probability model describing the observations given the parameter. This probability is represented by the likelihood function which defines the probability of the evidence given the parameter. Thus, likelihood function represents the aleatory uncertainty [48].

3.2 Application of Bayesian networks for heat load forecast in district heating systems

In order to apply Bayesian networks for the forecast of heat load at the consumption side of the district heating system, it is indispensable to study the parameters influencing the heat load. After studying these parameters we will model them in a Bayesian network to compute the heat load forecast.

3.2.1 Heat load consumption dataset

The Bayesian model developed in this thesis work, is based on the heat load consumption data of the district heating system of Skellefteå, Sweden. Skellefteå Kraft is a major energy producer in Sweden and supplies heating to around 5000 substations in Skellefteå city and outskirts.

Seasons: Heating is primarily used in cold weather conditions when the outdoor temperature is really low. For this reason we choose to develop our model for winter and spring seasons. From the dataset available to us, we take the duration from 22 December 2013 to 28 February 2014 for the winter season. The duration from 1 March 2014 to 30 April 2014 is considered for the spring season.

Buildings: We consider three residential buildings which have multi-family apartments . We refer these buildings as Building A, Building B and Building C. The substations at these buildings are deployed with sensors which collect the aggregate heat load consumption at an interval of each minute.

Available parameters influencing the heat load: The sensors deployed at the buildings record the district heating operational parameters, which include supply temperature, return temperature and flow rate. The outside temperature is also recorded from the on-site temperature sensor. The energy meters at the substation calculate the heat load from the DHS operational parameters according to the following equation [49]:

$$Q_{substation} = c * mt' * (T_s t' - T_r t')\, dt' \tag{12}$$

Here $Q_{substation}$ is the heat load consumption at a particular substation. c is the specific heat of the liquid in the district heating distribution network(mostly water), m is the flow rate, T_s is the supply temperature, T_r is the return temperature [49]. From this equation, it can be observed that the heat load consumption also depends on flow rate, supply temperature, return temperature and the difference of supply and return temperature. We refer these parameters as DHS operational parameters. The difference between supply and return temperature is referred as T_{delta}. From the experts in Skellefteå Kraft we learnt that the heat load demand at the production side is computed by considering the hourly outdoor temperature, T_{out}. Therefore we consider outdoor temperature as a key parameter influencing the heat load demand. The fact that outdoor temperature varies over different seasons, further motivates the need for a seasonal load profiling model for residential heat load consumption.

The variation of heat load consumption with outdoor temperature for Building A is shown in Figure 12. and Figure 13. Some key observations from these two figures are:

- The heat load consumption decreases with the increase in outdoor temperature. This is understandable. When the weather becomes comparatively hot, the residents tend to use less heating in the households.
- For a particular value of outdoor temperature, there is a range of heat load consumption. The range is much wider for mid ranged values of temperature, and is narrower towards extreme ranges of temperature. This implies that insignificant variation in outdoor temperature does not have a significant impact on heat load consumption. However, a huge variation in outdoor temperature, sometimes causes significant variation in the heat load consumption.

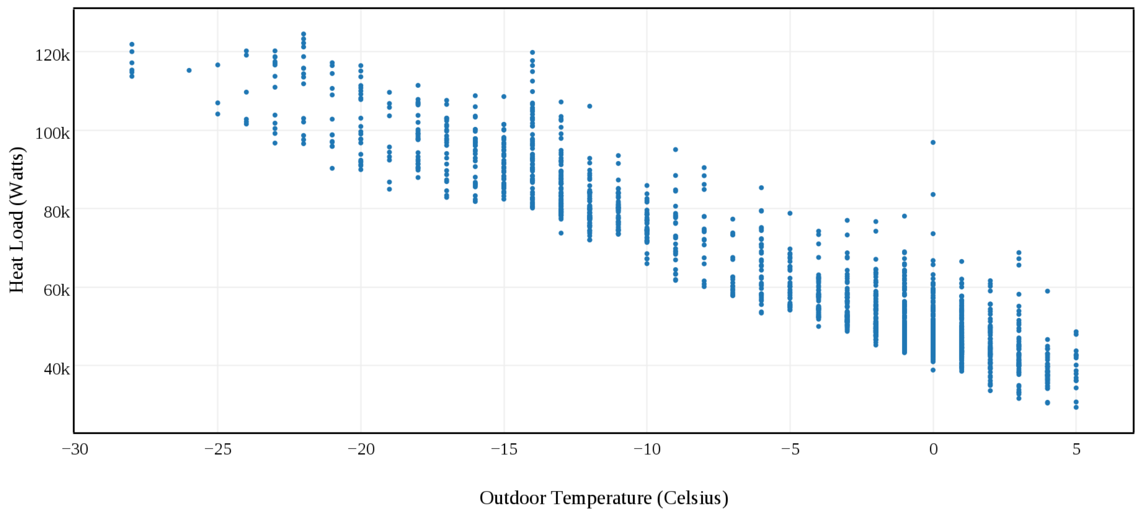

Figure 12. Heat Load variation with outside temperature during winter season for Building A

Additional parameters: In order to study the effect of user behaviour and daily load patterns, we add hour of day and day of week as two additional parameters influencing the heat load forecast. The list of all the available parameters is shown in Table 2 below. They are categorized into 3 domains: DHS operational parameters, weather forecast parameters and behavioural parameters.

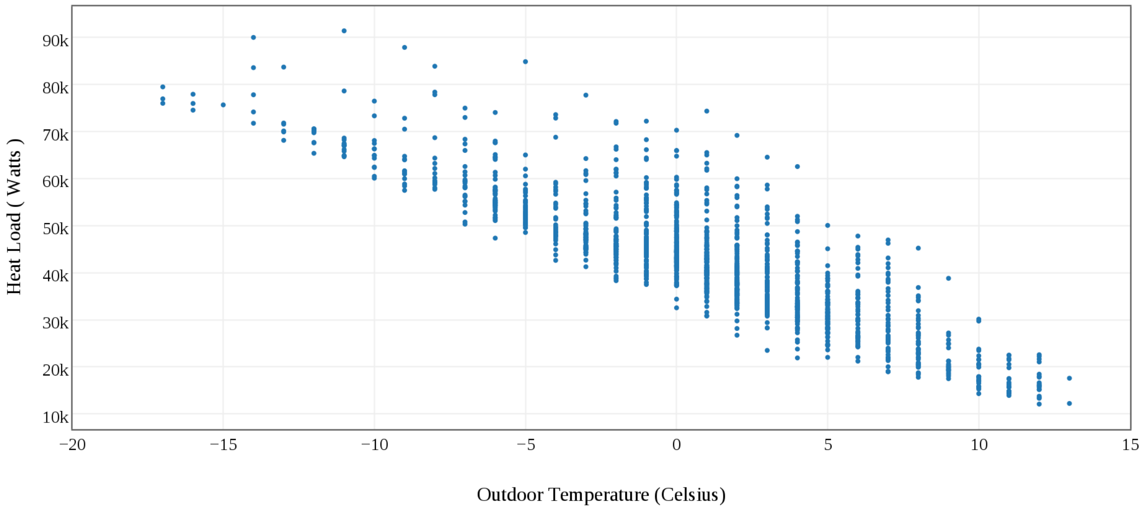

Figure 13. Heat Load variation with outside temperature during spring season for Building A

Table 2. Parameters considered for forecasting heat load

DHS Operational Parameters	**Weather Forecast**	**Behavioral Parameters**
Supply temperature(T_s) Return temperature(T_r) Flow rate(m) Difference of supply and return temperature(T_{delta})	Outdoor Temperature(T_{out})	Hour of Day(H_d) Day of Week(D_w)

Overview of the heat load consumption:

The heat load consumption for all three buildings during Winter and Spring seasons is plotted in Figure 14 and Figure 15. During the winters, the month of January shows the highest consumption, indicated by the peaks in Figure 14. We also observe that the heat load consumption in Building A is significantly lower than Building B and C. During the Spring season, the heat load consumption reaches its peak demand for a short time during the month of March. Here also, the heating consumption in Building A is much lower than the other two buildings.

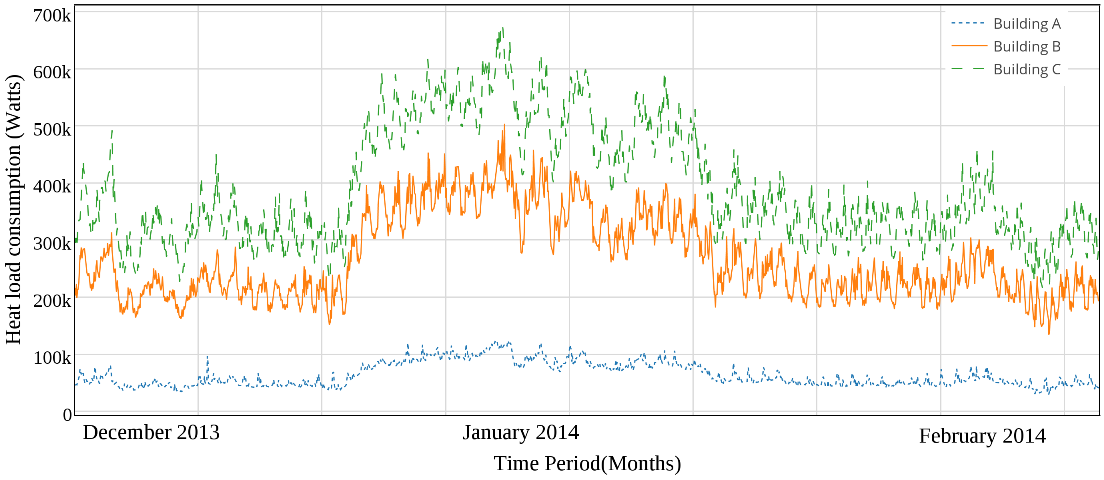

Figure 14. Heat load consumption in Winter Season in three buildings

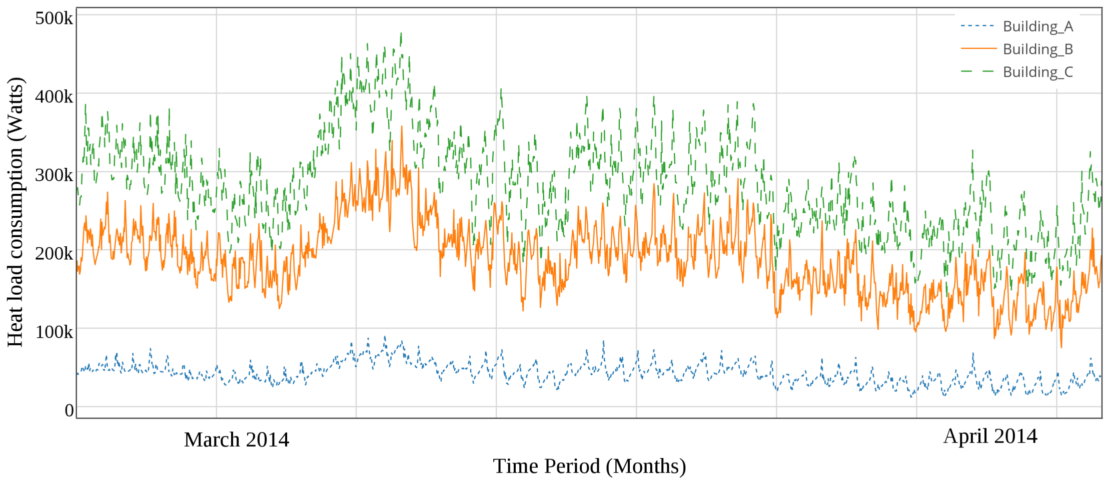

Figure 15. Heat load consumption in Spring Season in three buildings

3.2.2 Anomaly in district heating data

District heating data collected from sensors is subject to instrumentation faults which results in outliers in the thermal load consumption [49]. Such faulty values recorded in the heat load consumption dataset can lead to incorrect heat load forecast and also higher costs for consumers. It is difficult to detect these faults due to the large size of the district heating network, large amount of data and lack of fault detection functionalities in data collection and instrumentation systems [49]. Some of the common faults occur due to incorrect configuration of sensors, malfunction in temperature measurement sensors, valves and flow meters, reset of meters due to a blackout or other situations and leakage in heat transfer pipes.

Limit checking is one of the methods used to address the issue of faults in the district heating systems. This method checks whether a recorded quantity is within the bounds that are defined to be acceptable according to the physical properties of the district heating system. It is a useful method to detect common faults in district heating substations. One example of a possible limit checking scenario is that the return temperature should not be higher than the supply temperature. Failure of this test may imply a fault in one of the two sensors recording supply and return temperature. Alarms can be raised when a limit check fails at a particular substation to notify the district heating management [49]. Some sophisticated methods use entropy as a measure to detect anomalies in the thermal energy meter data. These methods are generally used to detect abnormal quantization in the thermal consumption which is characterized by a poor precision [49].

3.2.3 Naive Bayes classifier

After analysing the various parameters influencing the heat load consumption, we assume that the simplest way to observe the influence of each parameter on the heat load forecast is to assume conditional independence between different parameters. In this way, it is useful to study the influence of each parameter independently on the heat load forecast. However, it is also possible to model more complex relationships between different parameters in the Bayesian network. The objective of the thesis is to estimate the heat load forecast using the available parameters. We are not interested to study the dependencies between influencing parameters. Therefore, we limit the scope of this thesis by assuming conditional independence between different parameters influencing the heat load consumption.

In a Naive Bayes classifier [50], all features are conditionally independent given the class label. The presence of a particular attribute of a class is not related to the presence or absence of other attributes [42]. Due to this conditional independence property, each attribute contributes to the classification result independently and equally [51]. Naive Bayes classifier has been used in spam detection and document classification. It has been known to show faster training and quick decision making in comparison with other classifiers, due to its simple design. A Naive Bayes network is a simple Bayesian network which consists of only one parent node and one or more child nodes. There is a strong assumption of conditional independence among the child nodes with respect to the parent node. We consider the Naive Bayes classifier shown in Figure 16 [15]. Here, node *C* is the parent node. Nodes $A_1, A_2, ...A_n$ are the child nodes. The classifier utilizes the training data to learn the conditional probability of each attribute A_i given a class label *C*. Then it uses the Bayes' theorem to calculate the probability of *C* given each attribute $A_1, A_2, ...A_n$. Thus, it classifies *C* by predicting the class with the highest posterior probability [15]. Naive Bayes classifiers can be easily modelled as a Bayesian network as explained in section 3.1.4.

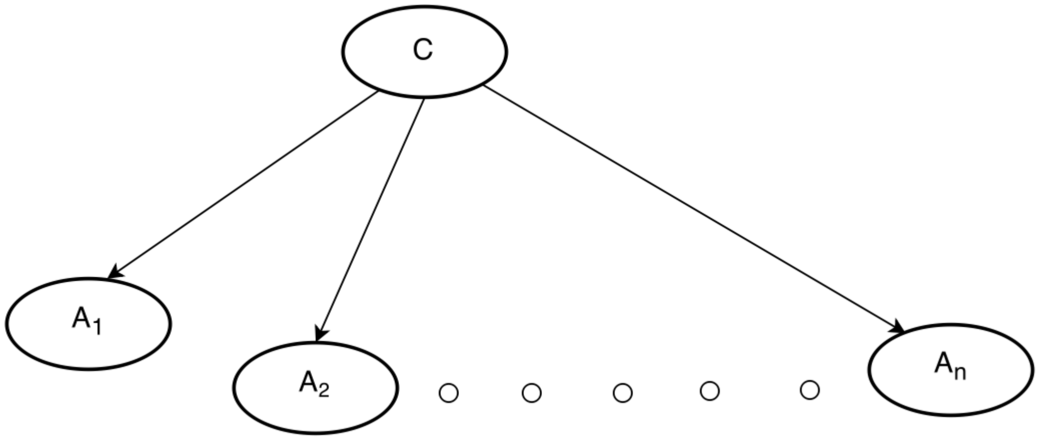

Figure 16. Naive Bayes classifier[15]

The conditional probability of the class label *C* given an attribute A_1 is defined by:

$$P(C|A_1) = \frac{P(C) * P(A_1|C)}{P(A_1)} \tag{13}$$

Similarly, the joint probability distribution of the Naive Bayes network, while considering the conditional independence is given by:

$$P(C, A_1, ..., A_n) = P(C) * P(A_1, ..., A_n|C) \tag{14}$$

$$P(C, A_1, ..., A_n) = P(C) * P(A_1|C) * .... * P(A_n|C) \tag{15}$$

Lets consider that the class label has *k* possible outcomes such that $k = \{1, ..., K\}$, then joint probability distribution function can be written as follows:

$$P(C_k, A_1, ..., A_n) = P(C_k) * \prod_{i=1}^{n} P(A_i|C_k) \tag{16}$$

For the given class label C_k Naive Bayes chooses the hypothesis with the maximum probability. This function forms the base of the classification of C_k.

$$y = \arg \max_{k \in \{1,...,K\}} P(C_k) * \prod_{i=1}^{n} P(A_i|C_k) \tag{17}$$

3.2.4 Inference in Bayesian networks

The proposed Naive Bayes classifier uses diagnostic probabilistic inference method to estimate the heat load forecast. Diagnostic inference is also called bottom-up inference. It uses an effect to inference a cause [41]. For example in the Bayesian network shown in Figure 17 we use diagnostic inference to estimate the posterior probability of the heat load given the outdoor temperature, *P(HL|T)*.

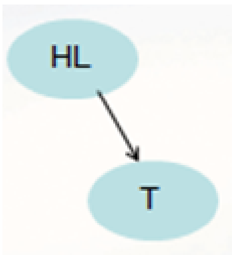

Figure 17. Diagnostic inference in Bayesian network

The posterior probability *P(HL|T)* is computed using the Bayes' rule as follows:

$$P(HL|T) = \frac{P(HL) * P(T|HL)}{P(T)} \tag{18}$$

3.3 Discretization

While learning a Bayesian network there are two ways to deal with continuous variables [52]. In the first method, the numeric attributes are assumed to follow a particular family of probability distribution. For example, continuous values associated with a class are assumed to follow a normal distribution. Thus a Gaussian Naive Bayes model can be developed for the continuous values of input variables. The Gaussian distribution may provide a suitable approximation to many real world distributions but, it is certainly not the best approximation in all cases [53]. The second method deals with the discretization of the continuous variables and then learning a Bayesian network over the discretized variables [52]. Discretization is a process of converting continuous attributes to a set of discretized intervals. For example, consider the interval [150,200]. Here 150 and 200 are continuous values but the range is discrete.

3.3.1 Discretization techniques for Naive Bayes classifier

Dougherty et al. [54] showed that the performance of Naive Bayes classifier was better when continuous attributes are discretized than when they are assumed to follow a Gaussian distribution. This can be attributed to the fact that discretization techniques do not make assumptions about the type of probability distribution of the numeric attributes [55].

The histogram of heat load consumption of Buildings A, B and C is shown in Figures 18, 19 and 20 respectively. This indicates that heat load consumption in all three buildings do not fit a normal distribution curve. Therefore, we decided to discretize the heat load and other influencing parameters for learning the Naive Bayes network.

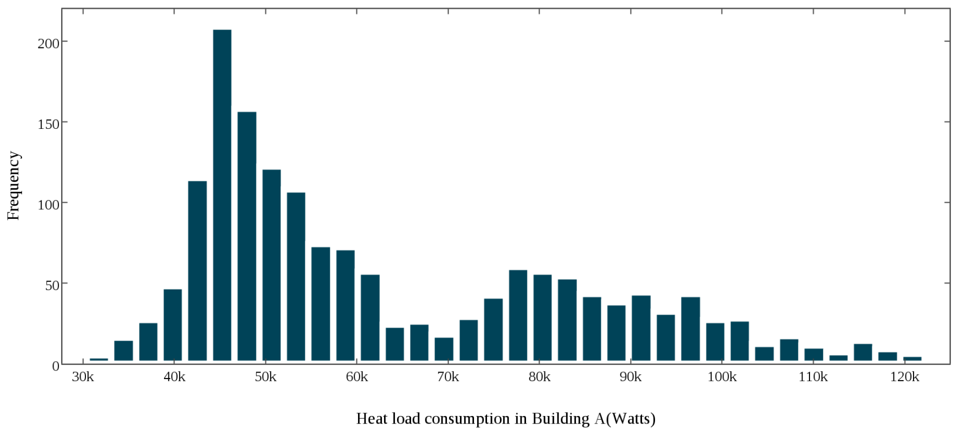

Figure 18. Histogram of the heat load consumption in Building A during Winter season

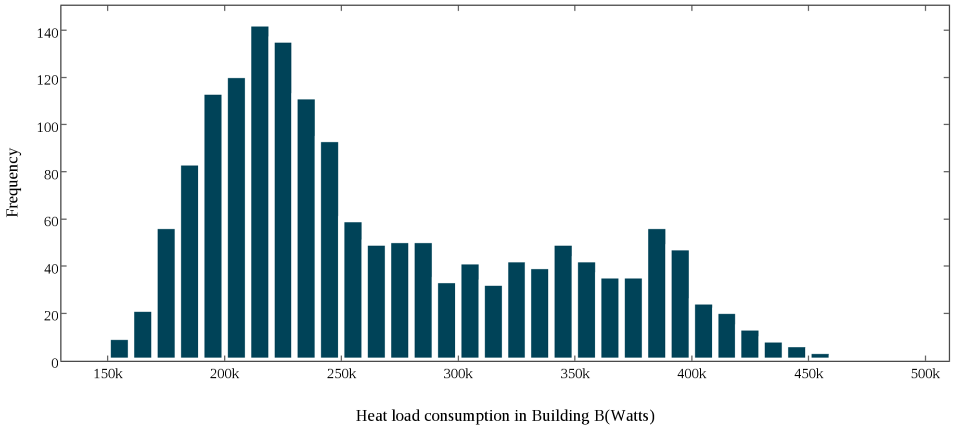

Figure 19. Histogram of the heat load consumption in Building B during Winter season

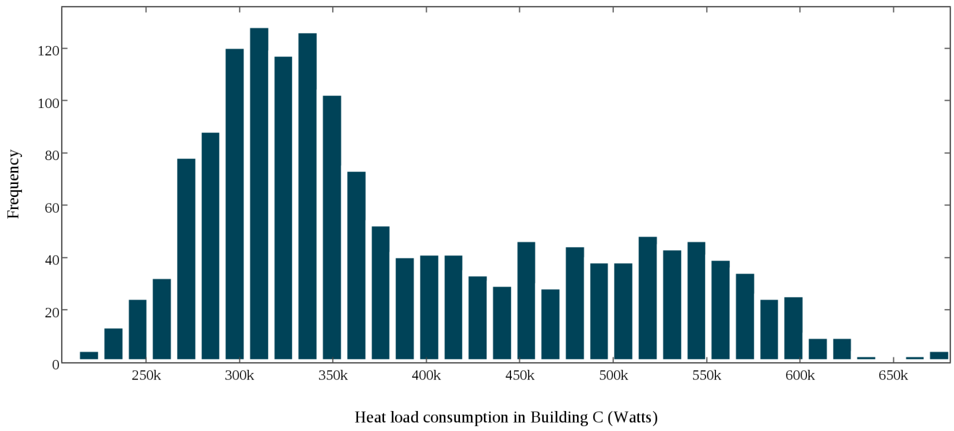

Figure 20. Histogram of the heat load consumption in Building C during Winter season

The work in [55] compares different discretization techniques for Naive Bayes with the objective of reducing the classification error.

In the upcoming sections, we discuss two less complex and commonly used discretization techniques with Naive Bayes classifier and provide a very concise review of some other complex discretization techniques. We further justify our choice of choosing a particular technique. Consider a numeric attribute X_i in a dataset with n training instances. Let v_{min} and v_{max} be the minimum and maximum values of X_i [56].

3.3.2 Equal width discretization (EWD)

Equal width discretization is one of the simplest methods for obtaining discretized values from continuous ones [54]. In this method, the continuous values corresponding to the numeric attribute are first sorted from v_{min} to v_{max}. After sorting, the range of continuous values from v_{min} to v_{max} is divided into *k* intervals of equal width, where *k* is a user supplied parameter. The width *w* of each equal sized interval is given by the following equation:

$$w = (v_{max} - v_{min})/k \tag{19}$$

A cut-point is a real value in the range of continuous values that divides the range into two intervals. For instance, an interval of continuous values from [x,z] can be partitioned into [x,y] and (y,z] where the continuous value *y* is the cut-point [57]. The various cut-points in case of equal width discretization are v_{min} + *w*, v_{min} + *2w*, v_{min} + *3w*,...., v_{min} + *(k-1)w*.

This discretization technique is used for each numeric attribute independently. It does not make make use of class information and hence it is an unsupervised discretization method. This method is less complex and easy to implement. Research has shown that it works very well with Naive Bayes classifier [56].

3.3.3 Equal frequency discretization (EFD)

In this method also, the continuous values corresponding to a numeric attribute are sorted from v_{min} to v_{max}. After sorting, the range of continuous values from v_{min} to v_{max} is divided into k intervals, such that each interval contains approximately equal number of instances. Here, *k* is a user supplied parameter. Each interval has *n/k* instances. Equal frequency distribution is suitable for a large training dataset and has shown good performance with the Naive Bayes classifier [55].

3.3.4 Other discretization techniques

Entropy minimization discretization (EMD) [58] is a supervised discretization technique which selects some candidate cut points by computing the mid point of every successive pair of sorted continuous values. Further, it selects the cut point with the minimum entropy among all the candidate cut points by using a binary discretization.

Lazy discretization (LD) approach [59] delays discretization until classification time. For each continuous value it selects two cut-points, such that the value is in the middle of the two chosen cut-points. The interval width can be chosen using equal width discretization (EWD) or entropy minimization discretization (EMD) technique.

3.3.5 Choice of a discretization technique

The dataset being considered for forecasting the heat load contains a lot of numeric attributes like outdoor temperature, supply temperature, return temperature, heat load, difference of supply and return temperature and flow rate. Discretization of all these parameters is a challenging task because each building has its own unique dataset of all these parameters. Therefore, discretization needs to be carried out for each building separately. Manual discretization of several parameters is a highly time consuming task. Supervised discretization techniques sometimes require expert knowledge or use of more sophisticated schemes like EMD and LD. There is a need to reduce the discretization overhead on the users for making discrete states of different parameters for each building. It is beneficial to utilize a discretization technique which is easy to deploy across a large number of buildings. Ease of use becomes an important factor for discretization when we are dealing with the data from around 5000 buildings at a metropolitan scale. With unsupervised discretization techniques the user provides k, the number of states, as an input parameter and the resulting discrete states are obtained without any manual intervention.

The complexity and performance of a discretization technique are the two parameters which play a key role in the choice of a particular technique, especially while dealing with a large dataset. EWD and EFD depend on the sorting of the number of training instances, n. Hence they have a time complexity of the order $O(n \log n)$. EMD also first sorts the n, training instances. This complexity of this operation is $O(n \log n)$. Further, for each candidate cut-point the probabilities of each of m classes are computed [56]. This operation requires a complexity of order $O(mn \log n)$ which is higher than the complexity of sorting. Hence the overall complexity of EMD is of order $O(mn \log n)$. LD performs discretization separately for every test instance. This leads to a complexity of $O(nl)$ where n is the number of training instance and l is the number of test instances.

Table 3 provides a comparison of the complexity of various discretization techniques discussed in the sections above [56]. This table illustrates the fact that EWD and EFD have a lower complexity than LD and EMD. Therefore it is preferable to choose EWD or EFD considering their lower complexity and unsupervised nature. EMD and LD are supervised schemes and have a higher complexity as compared to EWD and EFD. This means with an increase in the number of variables for discretization or increase in the size of training data, EMD and LD will take more time to discretize than EWD and EFD. Therefore, it is more preferable to choose a discretization technique with less complexity, considering our problem domain.

Table 3. Comparison of discretization techniques in terms of complexity [56].

Discretization Technique	**Complexity**
EWD	$O(n \log n)$
EFD	$O(n \log n)$
EMD	$O(mn \log n)$
LD	$O(nl)$

The work in [56] concludes that with Naive Bayes classifier EWD and EFD techniques achieve very similar performances. They also observed that with the increase in the size of training data, EWD performs better than EFD with Naive Bayes classifier. From this study and our analysis EWD seems to be the best choice for discretization with a Naive Bayes classifier.

3.3.6 A clustering based approach for discretization

Clustering provides an interesting way to look at the problem of discretization. Since we need a discretization technique which can scale to any building in a city comprising of 5000 buildings, it is very difficult to utilize a supervised method of discretization. In a dataset, the clustering techniques search for similar data points and group them into clusters. The objective is to minimize the distance between the data points within the cluster and maximize the distance between clusters [60]. Choosing the right number of clusters is the key factor in clustering techniques.

We choose k-means [61] for discretization because of its simplicity and unsupervised nature. It has also been identified as one of the top 10 algorithms in data mining [62]. The main objective of k-means is to divide the dataset into *k* clusters, where *k* is fixed in advance. In our case we take *k=5*, due to simplicity and having the same number of states for capturing the heat load variation in different buildings. Consider a numeric set of data points X ranging from *i* to *n*, such that $X = \{x_i | i = 1, ..., n\}$. The algorithm is explained in the following steps:

1. Initially, *k* data points are chosen randomly from the dataset. These initial seeds represent the centroids of each cluster. Let these initial *k* centroids be represented by the set, $C = \{c_j | j = 1, ..., k\}$

2. For each point in X we find the nearest centroid in *C*. Each data point x_i is allocated to its nearest centroid [63]. This is done by computing the Euclidean distance between each data point x_i and each centroid c_j. x_i is assigned to the cluster with which it has the minimum Euclidean distance. The first iteration concludes when each data point is assigned to some centroid. Now we have a set of *k* clusters, with each cluster having a unique centroid.

3. For each cluster, the centroid is re-computed by taking the mean of all data points x_i belonging to a particular cluster. Thus we have a new set of centroids in *C* for each cluster.

4. Steps 2 and 3 are repeated again and again until the centroid for each cluster is fixed. The algorithm converges when the centroid of each cluster is fixed and do not change.

The k-means algorithm aims at minimizing the squared error function represented by the objective function below [63].

$$\sum_{j=1}^{k}\sum_{i=1}^{n}||x_i - c_j||^2 \tag{20}$$

Here, $||x_i - c_j||^2$ represents the distance between a particular data point x_i and a centroid c_j.

In applying k-means to the district heating dataset available to us, we choose a fixed number of clusters for each attribute as it is less complex to deploy across a large number of buildings. We set *k=5*. We apply k-means to all the DHS operational parameters and weather forecast parameters individually for discretization. Since the behavioural parameters like hour of day and day of week are discretized categorical attributes, they do not need to be discretized. K-means outputs the discrete values for each parameter. In terms of discretization, each cluster represents a discrete state with a well-defined range. For example, a *cluster0* corresponds to a discretized state, State *A* with a range of *[a,b]* where *a* and *b* are the continuous attributes belonging to the dataset, such that *a<b*. The next cluster, *cluster1* corresponds to State *B* with a range *[c,d]* where *c>b*.

3.4 Proposed model for forecasting heat load

Figure 21 represents the proposed heat load forecasting model. The DHS operational parameters, heat load consumption and outdoor temperature are collected at a per minute interval by the sensors deployed at the substations in different buildings. This data is collected together and converted to a resolution of per hour, to take into account the fact that the delay in the control loop of Skellefteå Kraft is 4-6 hours. K-means and EWD are used to convert continuous data into discrete states. All continuous parameters like heat load, outdoor temperature, supply temperature, return temperature, flow rate, difference of supply and return temperature are converted into discrete states with well-defined ranges. The discretized parameters serve as an input to the Naive Bayes classifier which classifies the forecasted heat load to a particular state with the highest probability. The output value of the heat load forecast serves as input to the production grid of Skellefteå Kraft. The knowledge about heat load demand for the future helps in optimizing the district heating production grid. The load forecast can also be provided as an input to building energy management systems (BEMS) to estimate the building energy demand, and can also be used to develop localized energy saving strategies for the buildings.

For modelling the Bayesian network we use the notations for the various parameters as shown in Table 4. We consider the current time is t and the load has to be forecasted for $t + h$ hours into the future. Here h is the forecast horizon.

Table 4. Parameters and notations for the proposed model

Parameters	**Notation**
Current supply temperature	T_s
Current return temperature	T_r
Difference of supply and return temperature	T_{delta}
Current flow rate	m
Current outdoor temperature	$T_{out(t)}$
Outdoor temperature forecast at $t + h$ hours	$T_{out(t+h)}$
Hour of Day	H_d
Day of Week	D_w
Current heat load	HL_t
Heat load forecast at $t + h$ hours	$HL_{(t+h)}$

Now we discuss the four cases for computing the heat load forecast by modelling a Naive Bayes network for each case. Each Naive Bayes network studies the influence of some parameters on the heat load forecast. The behavioural parameters, hour of day and day of week are included in each Naive Bayes network, to accommodate the influence of user behaviour on the heat load forecast.

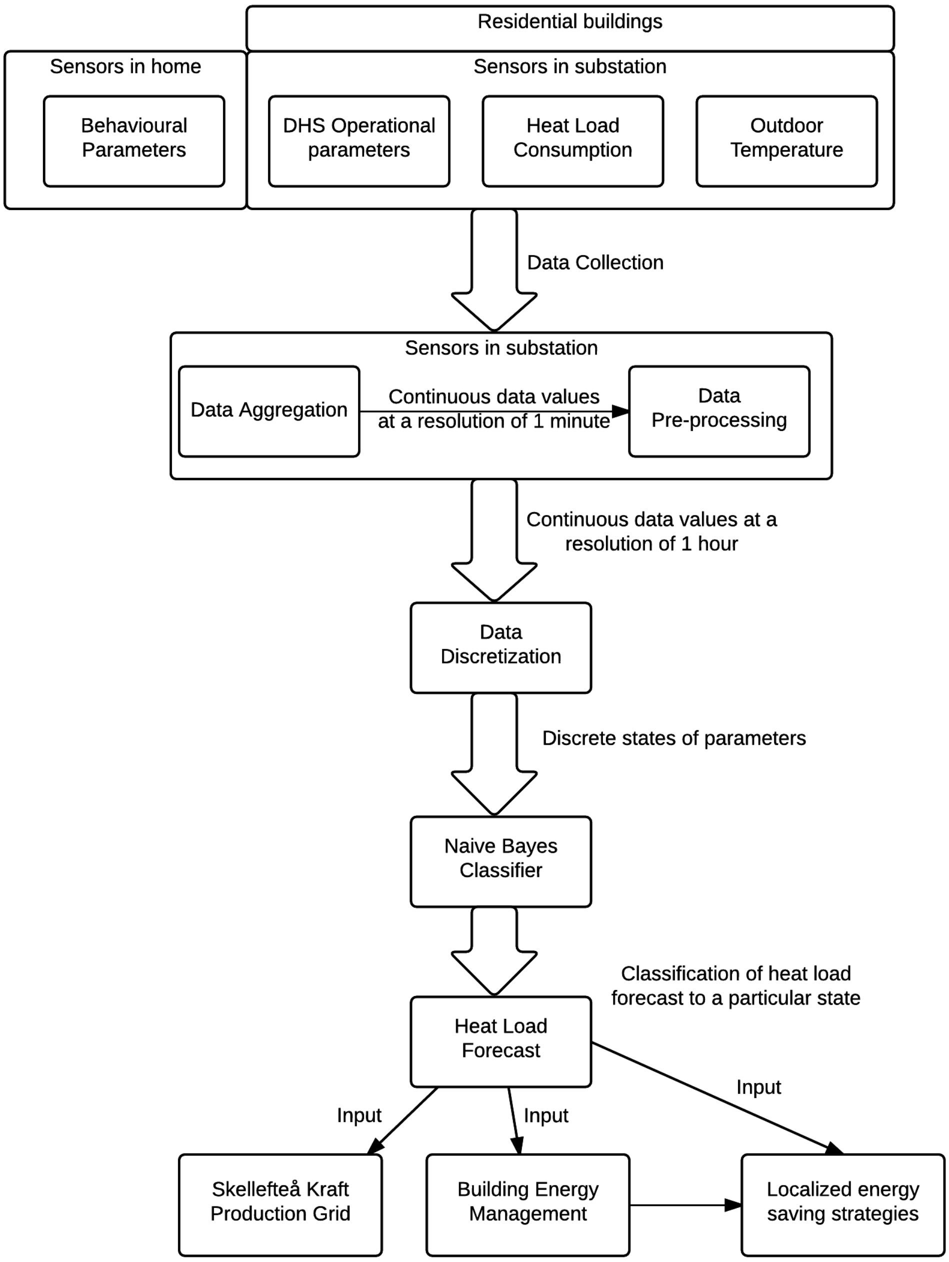

Figure 21. The proposed model for heat load forecasting

3.4.1 Naive Bayes network for DHS operational parameters

The Naive Bayes network considering the influence of DHS operational parameters on the heat load forecast is shown in Figure 22. This Bayesian network classifies the state of the heat load forecast by considering the influence of supply temperature, return temperature, flow rate and difference of supply and return temperature. The joint probability distribution function for the

DHS operational parameters and the heat load forecast can be written as follows:

$$\begin{aligned} P(H_d, D_w, T_s, T_r, T_{delta}, m, HL_{(t+h)}) = P(H_d|HL_{(t+h)}) * P(D_w|HL_{(t+h)}) * \\ P(T_s|HL_{(t+h)}) * P(T_r|HL_{(t+h)}) * P(T_{delta}|HL_{(t+h)}) * P(m|HL_{(t+h)}) * P(HL_{(t+h)}) \end{aligned} \tag{21}$$

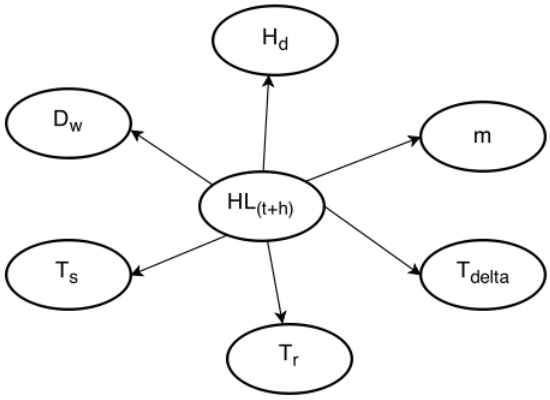

Figure 22. Naive Bayes network for heat load forecast using DHS operational parameters

3.4.2 Naive Bayes network for weather forecast parameters

We consider outdoor temperature forecast as the parameter influencing the heat load forecast. The Naive Bayes network representing this relationship is shown in Figure 23. The joint probability distribution function for the outdoor temperature, $T_{out(t+h)}$, behavioural parameters and the heat load forecast can be written as follows:

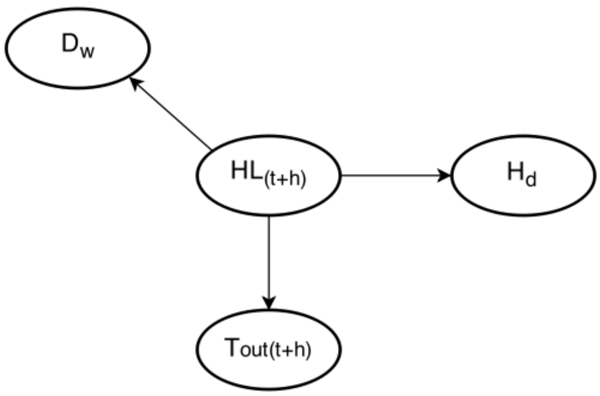

Figure 23. Naive Bayes network for heat load forecast using outdoor temperature forecast

$$P(H_d, D_w, T_{out(t+h)}, HL_{(t+h)}) = P(H_d|HL_{(t+h)}) * P(D_w|HL_{(t+h)}) * \\ P(T_{out(t+h)}|HL_{(t+h)}) * P(HL_{(t+h)}) \quad (22)$$

3.4.3 Naive Bayes network for combined influence of DHS operational parameters and weather forecast parameters

In this section, our objective is to evaluate the influence of DHS operational parameters and outdoor temperature forecast on the heat load forecast. The Naive Bayes network considering the influence of DHS operational parameters and outdoor temperature forecast on the heat load forecast is shown in Figure 24. The joint probability distribution function for this case can be written as follows:

$$P(H_d, D_w, T_s, T_r, T_{delta}, m, T_{out(t+h)}, HL_{(t+h)}) = P(H_d|HL_{(t+h)}) * P(D_w|HL_{(t+h)}) * \\ P(T_s|HL_{(t+h)}) * P(T_r|HL_{(t+h)}) * P(T_{delta}|HL_{(t+h)}) * \\ P(m|HL_{(t+h)}) * P(T_{out(t+h)}|HL_{(t+h)}) * P(HL_{(t+h)}) \quad (23)$$

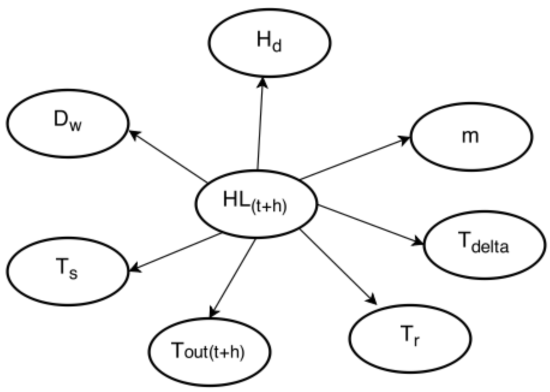

Figure 24. Naive Bayes network for heat load forecast using DHS operational parameters and outdoor temperature forecast

3.4.4 Naive Bayes network for current heat load consumption and weather forecast parameters

In this case, we consider the current load consumption $HL_{(t)}$ and outdoor temperature forecast $T_{out(t+h)}$ as the parameters influencing the heat load forecast. The Naive Bayes network representing this relationship is shown in Figure 25. The joint probability distribution function for the outdoor temperature, $T_{out(t+h)}$, behavioural parameters and the heat load forecast can be written as follows:

$$P(H_d, D_w, T_{out(t+h)}, HL_{(t)}, HL_{(t+h)}) = P(H_d|HL_{(t+h)}) * P(D_w|HL_{(t+h)}) * \\ P(T_{out(t+h)}|HL_{(t+h)}) * P(HL_{(t)}|HL_{(t+h)}) * P(HL_{(t+h)}) \tag{24}$$

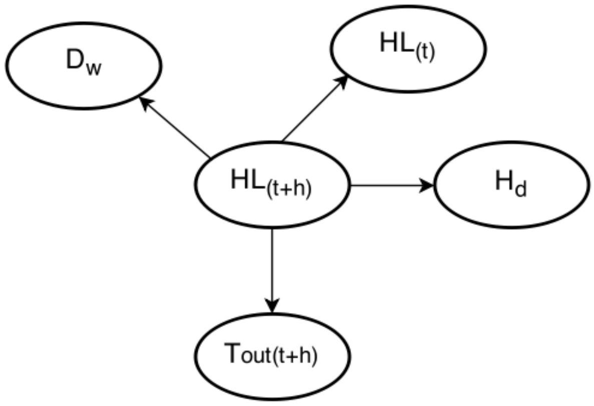

Figure 25. Naive Bayes network for heat load forecast using outdoor temperature forecast and current heat load

3.5 Summary

In this chapter, we discussed the theory of Bayesian networks and learnt how to model probability distributions in Bayesian networks. We studied the important characteristics of the heat load consumption dataset and identified the parameters to be considered for the heat load forecast. We discussed the Naive Bayes classifier and its application to develop the heat load forecasting model. Further, we discussed various discretization techniques and justified the choice of choosing equal widths discretization and k-means clustering. Then we discussed the proposed model while considering the effect of various parameters on the heat load forecast. In the next chapter, we evaluate the proposed model and discuss the results.

4 IMPLEMENTATION AND RESULTS EVALUATION

In the previous chapter, we discussed the proposed model for computing the heat load forecast by considering DHS operational parameters, outdoor temperature forecast and behavioural parameters. In this chapter, we study the results of the heat load forecast using our proposed model in three residential buildings over a period of winter and spring seasons. We evaluate the model by studying the influence of several parameters on the heat load forecast through four cases. We then analyze the results and discuss the possible energy savings by using the forecast information provided by our model.

4.1 Implementation

We compute the heat load forecast for all three buildings over a horizon of 1, 2, 3, 6 and 24 hours. We learned from the experts in Skellefteå Kraft that the delay in the district heating control loop is 4-6 hours. Therefore, we compute hourly forecasts upto 6 hours. The load forecast for the next 24 hours was computed to observe the heating demand in each building for the next day. The model was evaluated on winter and spring seasons. The duration of winter season was chosen from 22 December 2013 to 28 February 2014. The duration of spring season was chosen from 1 March 2014 to 30 April 2014. The trace-driven analysis and model evaluation was carried out in Weka 3.6 [64]. The packages of Naive Bayes and K-means implemented in Weka were utilized for this thesis work. GeNIe [65] was used to carry out EWD discretization for various continuous attributes. For discretization of continuous attributes the number of states was fixed to 5 for all attributes as it was less complex to deploy across a large number of buildings. The model was evaluated using 10 folds cross validation for both winter and spring seasons.

4.2 Results

In this section, we study the results of the heat load forecast in the three buildings over a period of winter and spring seasons by studying the four cases of the proposed model. The behavioural parameters as shown in Table 2. are added to all the four cases.

For evaluating the proposed classification model, it is necessary to compute the accuracy of classification. The accuracy of classification highlights the ability of the model to correctly predict classes and also to differentiate the classes. To get the complete picture of the performance of the classifier, it is imperative to study the confusion matrix [66]. We consider a classifier which is trained to distinguish between States A, B and C. We assume that there are a total of 70 instances including 30 instances in $StateA$, 20 instances in $StateB$ and 20 instances in $StateC$. One possible confusion matrix for such a classifier is shown in Table 5. The actual states are represented in the rows while the predicted states are represented in the columns. From the confusion matrix it can be observed that out of 30 instances of $StateA$, the classifier predicted 25 instances correctly. 3 instances of $StateA$ were classified as $StateB$ while the remaining two instances were classified as $StateC$. Similarly, out of 20 instances of $StateB$, 18 were predicted correctly by the classifier as belonging to $StateB$. For $StateC$, the classifier predicted 15 instances correctly out of the total 20 instances. Therefore the accuracy of classification in simple terms is computed by dividing the total number of correct classifications by total number of instances. The correct number of classifications for each state is represented along the diagonal of the confusion matrix.

$$Accuracy\ of\ classification = \frac{Number\ of\ correct\ classifications}{Total\ number\ of\ instances} \tag{25}$$

$$Accuracy\ of\ classification = \frac{25 + 18 + 15}{70} = 0.8285 \tag{26}$$

Therefore, the classification accuracy for this example is 82.85%. We compute the accuracy for the states of heat load in our model in a similar fashion. Since we are computing the accuracy of the heat load forecast, we will refer it as forecasting accuracy.

Table 5. Example of a confusion matrix

$Actual/Predicted$	$StateA$	$StateB$	$StateC$
$StateA$	25	3	2
$StateB$	1	18	1
$StateC$	2	3	15

4.2.1 Case I : Influence of DHS operational parameters on heat load forecast

In this case, we observe the influence of DHS operational parameters (shown in Table 2) on the heat load forecast. The accuracies of the heat load forecast across both winter and spring seasons for all four cases is presented in Figures 26 and 27. The results of forecasting accuracy for Case I indicate that the accuracy of heat load forecast decreases with the forecasting horizon for a particular building. This can be explained by Eq. 12. The magnitude of the heat load consumption at a substation depends on the current values of DHS operational parameters which are used for forecasting heat load using our trained Bayesian network. With the increase in the forecasting horizon, the dependency of the heat load forecast decreases on the current values of DHS operational parameters, which results in the decrease in accuracy. It can also be observed that Building A achieves the highest accuracy with EWD in both winter and spring seasons for all horizons.

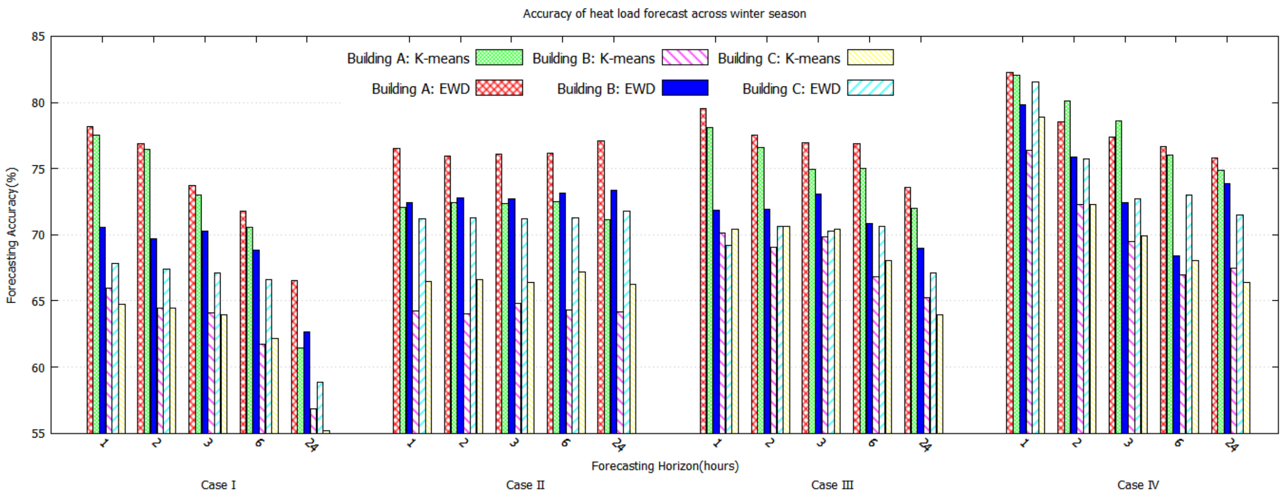

Figure 26. Accuracy of heat load forecast across winter season.

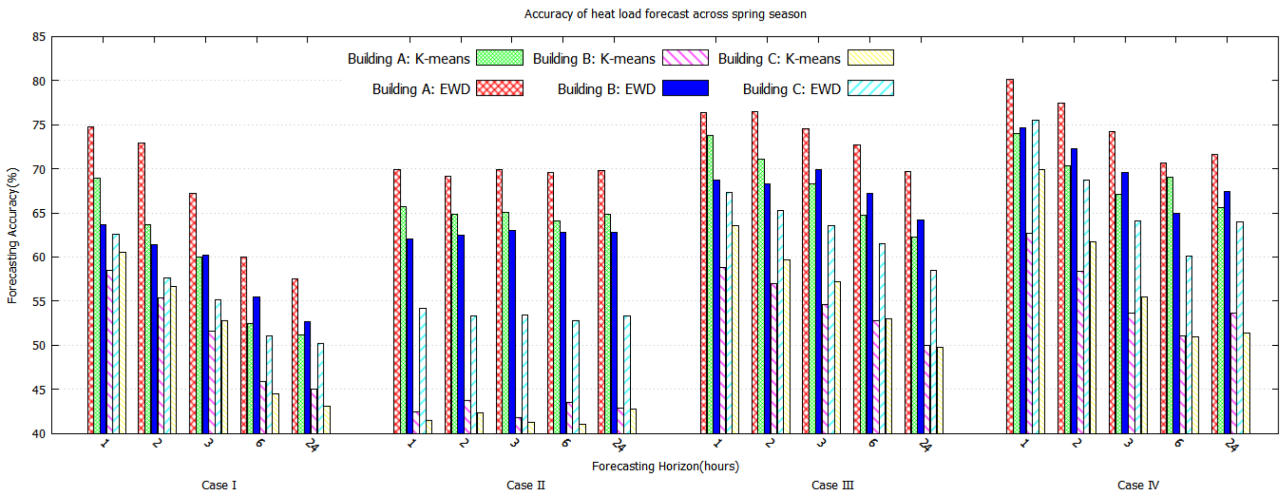

Figure 27. Accuracy of heat load forecast across spring season.

4.2.2 Case II : Influence of outdoor temperature forecast on heat load forecast

In this case, we compute the heat load forecast by training the model with the outdoor temperature forecast and behavioural parameters as shown in Table 2. In both winter and spring seasons, the accuracy of the heat load forecast across various horizons almost remains constant for a particular building. This maybe due to the fact that the heat load forecast at different horizons has similar dependency on the outdoor temperature forecast. Building A achieves the highest accuracy with EWD in both winter and spring seasons for all horizons.

4.2.3 Case III : Influence of DHS operational parameters and outdoor temperature forecast on heat load forecast

In this case, we study the influence of DHS operational parameters and outdoor temperature forecast on the heat load forecast. Figures 26 and 27 show the variation of the forecasting accuracy across different horizons for both winter and spring seasons respectively for this case. In the winter season, there was no clear trend in the accuracy of load forecast across various horizons. In the spring season, the accuracy decreases with the increase in forecasting horizon in majority of the cases for all buildings. Building A achieves the highest accuracy with EWD in both winter and spring seasons for all horizons.

4.2.4 Case IV : Influence of current heat load consumption and outdoor temperature forecast on heat load forecast

In this case, we study the impact of the current heat load consumption and outdoor temperature forecast on the heat load forecast. During the winter season, the accuracy of forecasting $HL_{(t+1)}$ is higher than the accuracy of forecasting heat load for other horizons. This indicates that the current heat load consumption has a strong influence on the heat load forecast. However, with increasing forecasting horizon, the dependency of heat load forecast on the current heat load decreases which results in a decrease in the forecasting accuracy. In the spring season, in most of the cases, the forecasting accuracy increases from 6 hour to 24 hour horizon. This is because of the user activity pertaining to a daily routine, which influences the heat load forecast.

4.2.5 Utilizing less training data for $HL_{(t+1)}$ forecast

In this section, we present the best case results for forecasting heat load for the next hour, obtained by using EWD and Naive Bayes classifier in Case IV(using outdoor temperature forecast and current heat load consumption). Figure. 28 shows the accuracy of $HL_{(t+1)}$ forecast for different percentages of training data for Buildings A, B and C over the period of both seasons. We observe that we achieve a good accuracy by just using training data between 10-20% with our Bayesian network. These results indicate that the proposed model would be effective in forecasting the heat load with less training data. There are around 5000 substations in the city of Skellefteå. The smart meter data collected from these substations at a high resolution poses the challenge of analysis of a large amount of data to forecast the heat load demand [37]. Our results indicate that in case of a large number of buildings and large amount of data, our model would be suitable for heat load forecast.

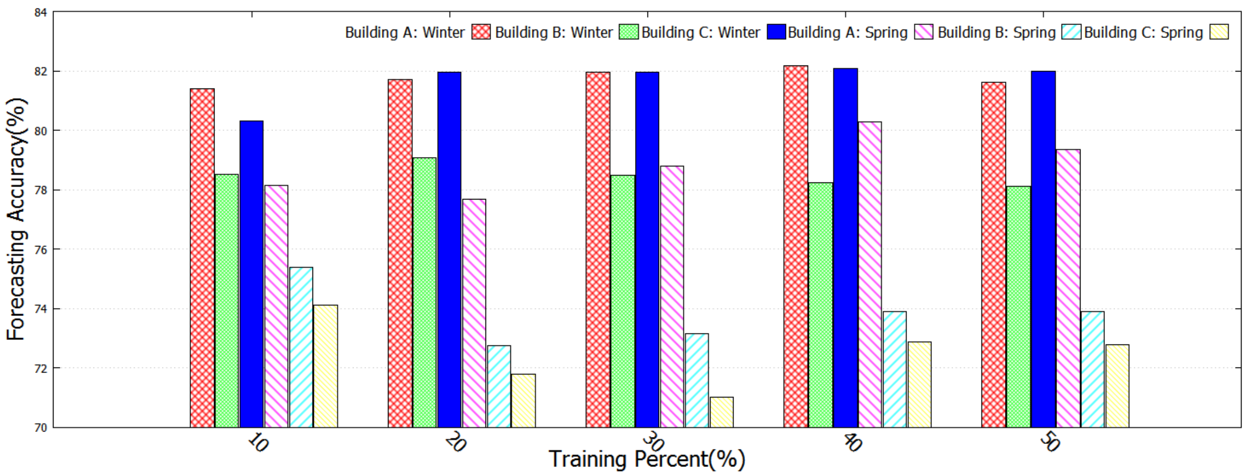

Figure 28. Forecasting accuracy for $HL_{(t+1)}$ with different percentages of training data using EWD and Naive Bayes in Case IV.

4.3 Analysis of results

In this section, we discuss the inferences obtained from the results. First of all, it is observed that each building is different due to different heat load consumption in each building during winter and spring seasons. This is represented in Tables 6 and 7. Each building exhibits a stochastic heat load demand depending on the various factors which influence the heat load consumption.

We observed that Building A achieves a higher accuracy for heat load forecast over all horizons across both seasons, in all four cases. This can be attributed to the fact that Building A has much less variation in heat load consumption in both seasons as compared to Building B and C as shown in Table 6 and Table 7. This also implies that in case of less heat load variation, the proposed model is able to forecast the heat load demand with a higher accuracy.

Figure 29. shows the average forecasting accuracies for Cases I, II, III and IV across both seasons for Buildings A, B and C by using EWD and Naive Bayes. Case IV has the highest average accuracy among all the cases for all forecast horizons except $HL_{(t+6)}$. Case III has a higher average accuracy than Case IV for $HL_{(t+6)}$ horizon. However the difference is less than 1%. Therefore, we observe that for EWD and Naive Bayes, Case IV achieves the highest accuracies for all buildings across both seasons.

Figure 30. shows the average forecasting accuracies for Cases I, II, III and IV across both seasons for Buildings A, B and C using K-means and Naive Bayes. Case IV has the highest average accuracy among all the cases for all forecast horizons except $HL_{(t+3)}$. Case III has a higher average accuracy than Case IV for $HL_{(t+3)}$ horizon. However the difference is only 0.17%. Therefore, we observe that for K-means and Naive Bayes, Case IV achieves the highest accuracies for all buildings across both seasons.

From the results we conclude that Case IV achieves the highest average accuracies across both seasons for all three buildings for various forecasting horizons with both discretization techniques. This implies that current heat load consumption and outdoor temperature forecast are the two parameters with most influence on the heat load forecast. Also the accuracies of Case III were second only to Case IV for both the discretization techniques. In comparison of Case I, Case II and Case III we observe that Case III achieves a higher accuracy most of the times across various horizons for both discretization techniques across both seasons.

The proposed model achieves best case average accuracies of 81.23% and 76.74% for a forecast horizon of 1 hour ($HL_{(t+1)}$) in the three buildings for winter and spring seasons respectively in Case IV using EWD. The model also achieves an average best case accuracy of 77.97% using EWD for three buildings across both seasons for the forecast horizon of 1 hour ($HL_{(t+1)}$) by utilizing only 10% of the training data.

We further analyse the results of Naive Bayes classification by considering both discretization techniques used for Case IV. We have observed that in few cases the forecasting accuracy achieved by using EWD and K-means is very similar while in most cases, the accuracy achieved by EWD is higher than the one achieved by K-means. The results from Figures 29 and 30 also indicate that average accuracies achieved by using EWD were higher than the ones achieved by K-means. We further investigate this issue by considering two results. In Result I, both EWD

Table 6. Heat load variation in Buildings A, B and C during winter season

Heat Load	Minimum(Watts)	Maximum(Watts)	Mean(Watts)	Standard Deviation(Watts)
Building A	29385	124533	63703	21124
Building B	134135	503548	267335	73485
Building C	213141	675061	389995	102743

Table 7. Heat load variation in Buildings A, B and C during spring season

Heat Load	Minimum(Watts)	Maximum(Watts)	Mean(Watts)	Standard Deviation(Watts)
Building A	12088	91425	42133	13268
Building B	74438	358408	188629	45278
Building C	135600	482268	283675	62801

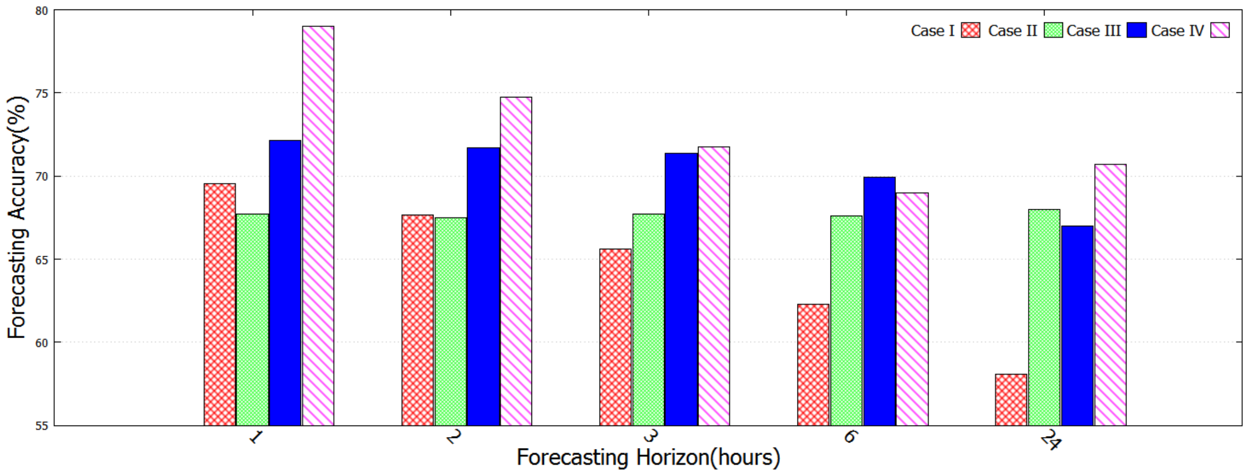

Figure 29. Average forecasting accuracy(%) for all four cases across both seasons for Buildings A, B and C using EWD discretization and Naive Bayes.

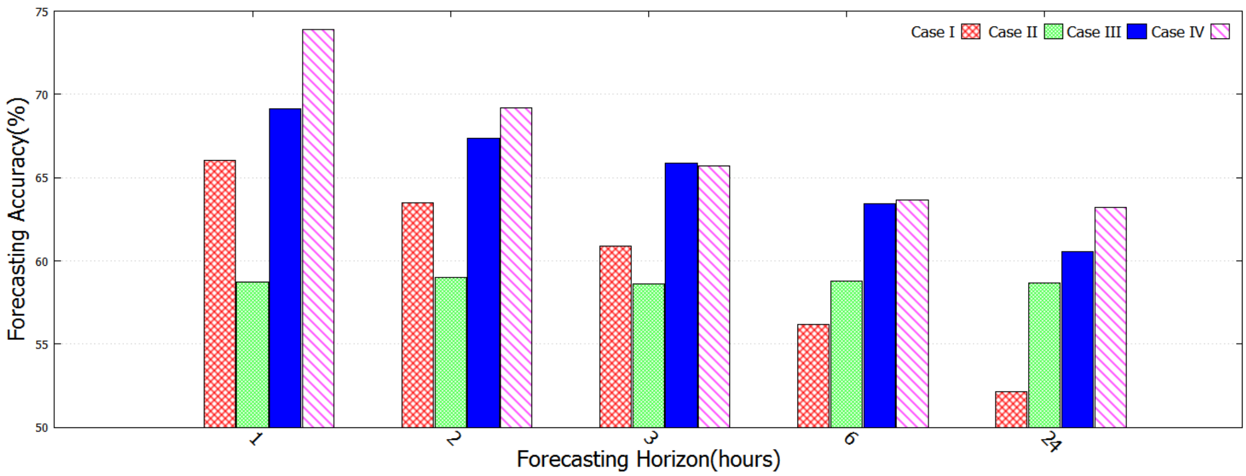

Figure 30. Average forecasting accuracy(%) for all four cases across both seasons for Buildings A, B and C using k-means clustering and Naive Bayes.

and K-means achieve similar forecasting accuracy for $HL_{(t+1)}$. In Result II, EWD achieves a much higher accuracy than K-means for $HL_{(t+1)}$ forecast. These results and their analysis is further explained in detail below.

Result I: $HL_{(t+1)}$ forecast for Building A during Winter Season

First of all we observe the discrete states of heat load obtained from EWD and K-means. Table 8. shows the discrete states of heat load represented by the ranges, counts and widths obtained by K-means clustering. Table 9. presents the discrete states of heat load represented by the ranges, counts and widths obtained by EWD. It can be observed that the states obtained in EWD have the same width. The width of each state is not the same in K-means.

Table 8. Discrete states of heat load for Building A during winter season obtained by K-means

K-means Heat load states	Range(Watts)	Count	Width
$StateA$	29385-51315	682	21930
$StateB$	51366-67705	380	16339
$StateC$	68350-85358	275	17008
$StateD$	85570-101070	217	15500
$StateE$	101368-124533	102	23165

Table 9. Discrete states of heat load for Building A during winter season obtained by EWD

EWD Heat load states	Range(Watts)	Count	Width
$StateA$	29385-48410	531	19030
$StateB$	48426-67436	526	19030
$StateC$	67485-86471	299	19030
$StateD$	86475-105235	231	19030
$StateE$	105718-124533	69	19030

Table 10 and Table 11 present the confusion matrix obtained from Naive Bayes classifier by using K-means and EWD discretization respectively. Both discretization techniques produce a unique set of states according to their underlying algorithm. Now, we compute the accuracy of heat load forecast for both the techniques by observing their confusion matrices. The total number of instances can be calculated by obtaining the sum of all entries in the matrix. In this case, the total number of instance equal 1655. The accuracy is computed by obtaining the sum of the correctly predicted classes and dividing it by the total number of instances.

$$Accuracy\ with\ K-means\ = \frac{604+277+226+176+75}{1655} = 82.0544\% \quad (27)$$

$$Accuracy\ with\ EWD\ = \frac{446+429+239+192+56}{1655} = 82.2961\% \quad (28)$$

Table 10. Confusion Matrix for heat load states obtained after Naive Bayes classification while using K-means for Building A during winter season

$Actual/Predicted$	$StateA$	$StateB$	$StateC$	$StateD$	$StateE$
$StateA$	604	76	1	0	0
$StateB$	83	277	20	0	0
$StateC$	0	22	226	27	0
$StateD$	0	1	24	176	16
$StateE$	0	0	0	27	75

Table 11. Confusion Matrix for heat load states obtained after Naive Bayes classification while using EWD discretization for Building A during winter season

$Actual/Predicted$	$StateA$	$StateB$	$StateC$	$StateD$	$StateE$
$StateA$	446	84	0	0	0
$StateB$	69	429	28	0	0
$StateC$	1	22	239	37	0
$StateD$	0	1	25	192	13
$StateE$	0	0	0	13	56

We observe that both the techniques achieve similar accuracy and all five states have been predicted and classified with a good precision by the Bayesian Network. The number of incorrect classifications is relatively less as compared to correct classifications, as observed from the confusion matrix. This result is further strengthened by the good accuracy values obtained for both discretization techniques.

Result II: $HL_{(t+1)}$ forecast for Building B during Spring Season

We observe the discrete states of heat load obtained from EWD and K-means. Table 12. and Table 13. present the discrete states of heat load represented by the ranges, counts and widths obtained using K-means and EWD respectively. Table 14. and Table 15. present the confusion matrices obtained from Naive Bayes classifier by using K-means and EWD techniques respectively. Now we compute the accuracies of heat load forecast for both techniques by observing their confusion matrices. The total number of instances in this case are 1463. Now we compute the accuracy for both the techniques.

$$Accuracy\ with\ K-means\ = \frac{287+219+212+95+104}{1463} = 62.6794\% \tag{29}$$

$$Accuracy\ with\ EWD\ = \frac{465+438+90+96+3}{1463} = 74.6411\% \tag{30}$$

We observe that EWD achieves a much higher accuracy than K-means. This can be explained from the confusion matrix of K-means and Naive Bayes in Table 14. We observe that the classifier is able to classify each state individually but it faces difficulty to be precise in classification while separating the states. For example, 68 instances of $StateE$ are misclassified as belonging to $StateB$. This indicates that though the classifier is able to predict State E by assigning majority of instances to State E, it fails to separate $StateE$ from $StateB$ in many instances. Similar kind of classification results for different states result in an overall lower accuracy.

In case of the classification results in EWD, we observe similar results like K-means. Though the classifier assigns majority of the instances to the correct state, it also assigns quite a lot instances to wrong state, which result in a comparatively lower accuracy when compared with the results of Building A in winter season. We also observe that $StateE$ is wrongly predicted by the classifier. This is probably due to the limited number of training points for this state.

Table 12. Discrete states of heat load for Building B during spring season obtained by K-means

K-means Heat load states	Range(Watts)	Count	Width
$StateA$	171446-206345	449	34899
$StateB$	136181-171305	354	35124
$StateC$	206470-250390	352	43290
$StateD$	250638-358408	136	107770
$StateE$	74438-135945	173	61507

With the discretization of just three buildings in two seasons, it would be unfair to compare K-means and EWD in terms of discretization and the resulting forecasting accuracy. We need to test these discretization techniques on a large number of buildings to come up with some kind of conclusion about their usability in the domain of discretization of heat load consumption.

Table 13. Discrete states of heat load for Building B during spring season obtained by EWD

EWD Heat load states	Range(Watts)	Count	Width
$StateA$	131436-187913	584	56794
$StateB$	188071-244491	572	56794
$StateC$	244911-301416	139	56794
$StateD$	74438-130983	148	56794
$StateE$	301711-358408	21	56794

Table 14. Confusion Matrix for heat load states obtained after Naive Bayes classification while using K-means for Building B during spring season

$Actual/Predicted$	$StateA$	$StateB$	$StateC$	$StateD$	$StateE$
$StateA$	287	86	72	1	2
$StateB$	88	219	1	0	46
$StateC$	105	2	212	33	0
$StateD$	1	0	40	95	0
$StateE$	1	68	0	0	104

Table 15. Confusion Matrix for heat load states obtained after Naive Bayes classification while using EWD for Building B during spring season

$Actual/Predicted$	$StateA$	$StateB$	$StateC$	$StateD$	$StateE$
$StateA$	465	83	0	35	0
$StateB$	105	438	28	1	0
$StateC$	1	44	90	0	4
$StateD$	52	0	0	96	0
$StateE$	0	1	17	0	3

Our results indicate that K-means performs well and at par with EWD in buildings with less variation of heat load consumption. However in buildings with higher variation of heat load consumption, K-means did not perform as well as EWD.

4.3.1 Analysis of forecasting accuracy

In this section, we analyse and explain the reasons for achieving the accuracy with regards to Case IV in detail. We choose Case IV because it achieves the highest average accuracy across both seasons. We use GeNIe [65] for explaining the rationale behind achieving a certain threshold of accuracy. We used outdoor temperature forecast and current heat load consumption as the two influencing parameters in Case IV. We have discussed in section 3.2.1 that there is high variation of heat load for mid-ranged values of outdoor temperature (temperature values which are not extreme). However for extreme values of temperature (values in range of minimum or maximum in the dataset), there is significantly less variation in heat load consumption. This factor highly influences the accuracy of the heat load forecast. We consider Building B in the spring season to explain the effect of this observation. The five states of temperature obtained from K-means is shown in the Table 16. We consider $temp0$ and $temp4$ states of the outdoor temperature as extreme states. Thus $temp0$ and $temp4$ have comparatively less variation in heat load as compared to $temp3$, $temp2$ and $temp1$. For the heat load we consider the same state representation as shown in Table 12.

Table 16. Discrete states of outdoor temperature for Building B during spring season obtained by K-means

K-means outdoor temperature states	Range(Celsius)
$temp0$	-13.45 to -5.59
$temp3$	-5.56 to 0.212
$temp2$	0.244 to 4.95
$temp1$	4.96 to 10.32
$temp4$	10.47 to 19.28

Figure 31. shows the Naive Bayes network for Case IV. It is clear that when the evidence is set as $temp0$, heat load has the highest probability of belonging to $StateD$ by a big margin. Figure 32. indicates that when outdoor temperature evidence is set to the state $temp4$, then heat load has the highest probability of belonging to $StateE$ by a fairly big margin. This indicates that for the temperature values belonging to the extreme states, the probability of heat load forecast belonging to a particular state is unambiguous. In this case, the classifier is able to predict the heat load state correctly with a higher accuracy.

We further estimate the probabilities of heat load forecast when outdoor temperature belongs to mid range states like $temp3$, $temp2$ and $temp1$. When we set the evidence for outdoor temperature to one of the mid range states we observe that the probabilities of heat load forecast belonging to a particular state is really close. This can be observed in Figures 33 -35. There is not a clear winner in terms of the state to which the heat load is more likely to belong. Therefore the classification of the heat load states is conflicted due to two or three states having similar probabilities for one state of outdoor temperature. Thus, the classifier predicts a state and assigns majority of instances to it, but the number of instances assigned to other states is also close. This leads to classification errors and ultimately lower accuracy. This characteristic of the proposed model is due to the variation of heat load with outdoor temperature as captured in the dataset

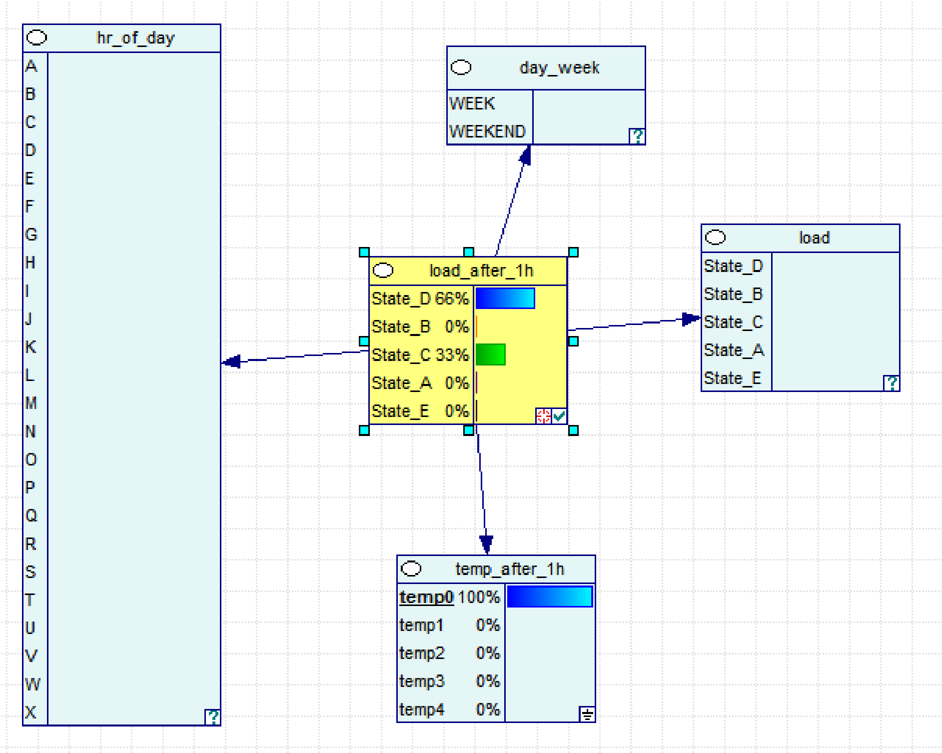

Figure 31. Probability of heat load forecast belonging to a particular state when evidence is set as $temp0$

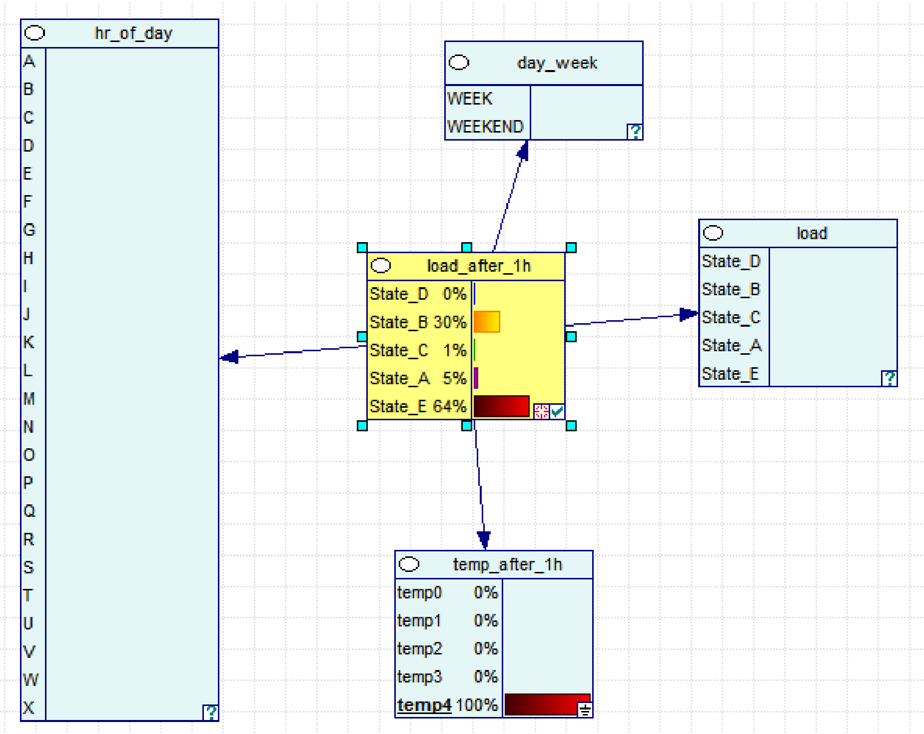

Figure 32. Probability of heat load forecast belonging to a particular state when evidence is set as $temp4$

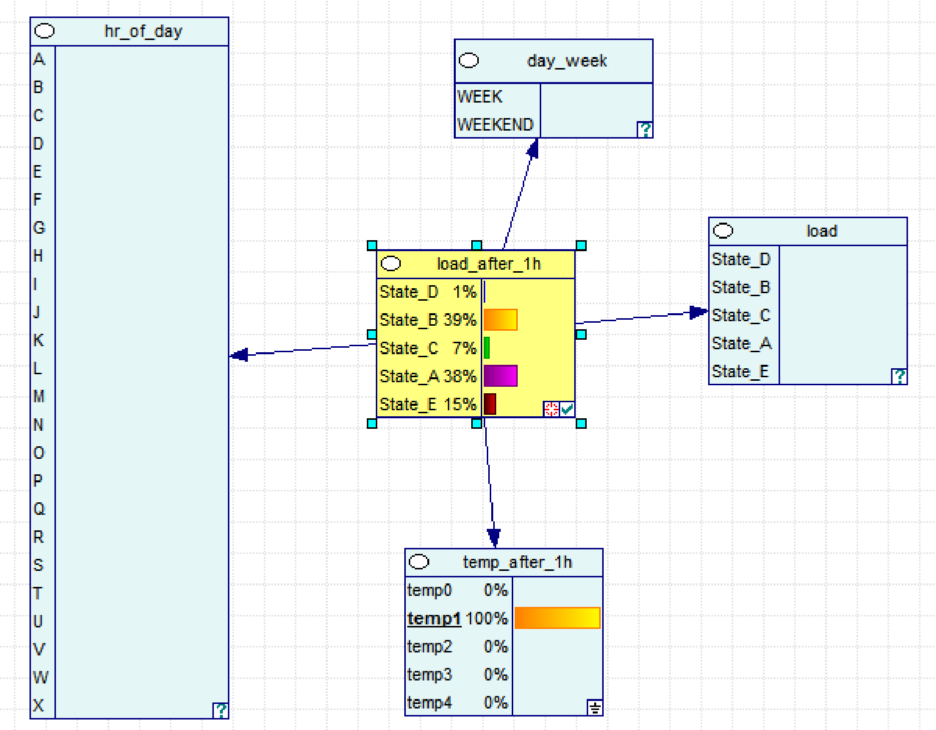

Figure 33. Probability of heat load forecast belonging to a particular state when evidence is set as $temp1$

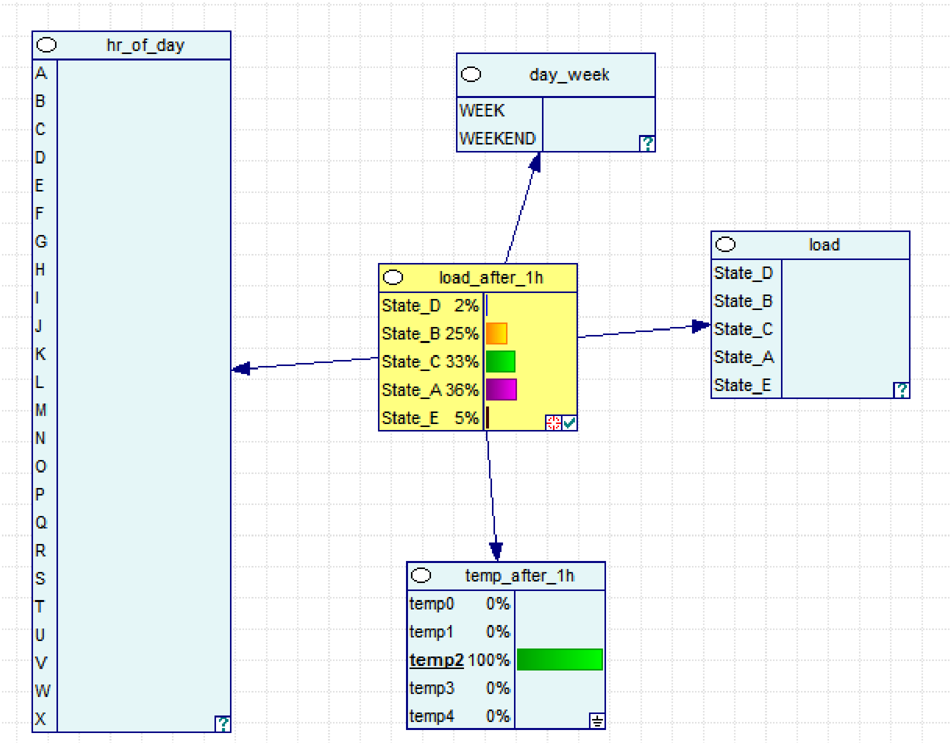

Figure 34. Probability of heat load forecast belonging to a particular state when evidence is set as $temp2$

used for training the model.

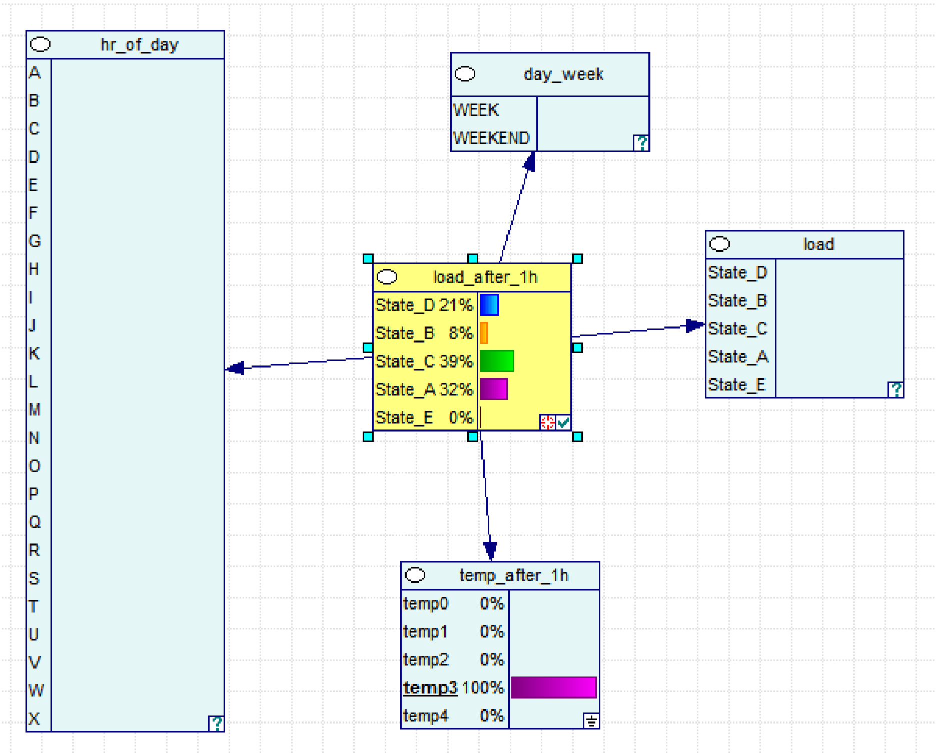

Figure 35. Probability of heat load forecast belonging to a particular state when evidence is set as $temp3$

4.4 Energy savings estimate from heat load forecast

In this section, we discuss the possible energy savings for each of the three buildings by using the knowledge of the heat load forecast provided by our model. From the results of the heat load forecast we obtain the best case forecasting accuracy for $HL_{(t+1)}$. Therefore, we estimate the energy savings for the buildings for the forecast horizon of one hour $HL_{(t+1)}$ to illustrate the benefits of using a forecasting model for energy savings.

To estimate the energy savings for the buildings we make a couple of assumptions. We assume that each building has been assigned a maximum heat load limit per hour, HL_{max} considering the physical features and number of occupants residing in the building. We also assume that in absence of the knowledge of heat load forecast the energy producer needs to supply this maximum heat load HL_{max} for the building for each hour to satisfy the heat energy demand of the building.

The maximum heat load consumption for each building during both winter and spring season can be observed from Table 6 and Table 7. We consider Building A during winter season to compute the energy savings. We observe that maximum heat load limit per hour HL_{max} is 124,533 Watts. The model achieves a forecasting accuracy of 82.29 % for $HL_{(t+1)}$ for Building

A during winter season using EWD and Naive Bayes. The discrete states of heat load are represented in Table 9. The number of correct and incorrect predictions for each state is represented by the confusion matrix shown in Table 11. State A is correctly predicted 446 times, State B 429 times and so on. When the model predicts a particular state for the next hour, the heat energy needed to be supplied by the energy company is the maximum value of the particular state so as to satisfy the heating demand for that particular hour. We compute the energy savings in case of correct predictions by the model for each state. This is achieved by subtracting the maximum heat load of a particular state from the maximum heat load limit HL_{max} of the Building A. This difference is multiplied by the number of times that particular state is correctly predicted by the model(number of correct predictions of each state also represents the number of hours during which that state was predicted)

Total energy savings for Building A during winter season

$$\begin{aligned} &= 446 * (124533 - 48410) + 429 * (124533 - 67436) + 239 * (124533 - 86471) + \\ &\qquad 192 * (124533 - 105235) + 56(124533 - 124533) \\ &= 71,247,505 \quad Watts \end{aligned}$$

Heat energy spent in absence of heat load forecast information = Sum of all elements of confusion matrix(for both correct and incorrect predictions) * HL_{max}

$$\begin{aligned} &= 1655 * HL_{max} \\ &= 1655 * 124,533 = 206,102,115 \quad Watts \end{aligned}$$

Therefore, percentage of energy savings due to correct forecasts = $71247505/206,102,115 * 100 = 34.56\%$

Thus, with the correct prediction of heat load forecast energy savings of around 35 % can be obtained in Building A during the winter season. However, the accuracy of the forecast model was 82.29 %. In cases of incorrect predictions of state of heat load forecast, the proposed model both overestimates and underestimates the heat load consumption in buildings. Due to the presence of both overestimation and underestimation scenarios with several instances, it is computationally intensive to consider all these cases and compute their impact on the energy savings. However, the energy savings obtained from the correct number of predictions highlight the fact that our model leads to significant energy savings in buildings. The table below shows the percentage of energy savings obtained from the correct forecasts in all three buildings across both seasons.

Table 17. Energy savings for $HL_{(t+1)}$ forecast for correct predictions

Building	Energy savings in Winter Season(%)	Energy savings in Spring Season(%)
Building A	34.56	37.15
Building B	31.99	29.73
Building C	28.71	25.88

4.4.1 Energy savings and sustainability

In the previous section, we computed the energy savings for each building by using the knowledge provided by the heat load forecast. The heat load forecast tells the production company the amount of heat energy it needs to produce for a particular period of time. Thus, the energy production company does not need to produce excess heat energy which makes them energy efficient and leads to energy savings. It also lowers the consumption of fuel needed to produce heat energy. This leads to efficient utilization of natural resources. The idea of sustainable development is to use the available resources efficiently to meet the needs of the present generation without compromising the needs of the future generation. Therefore, the heat load forecast information provided by our model helps in achieving sustainability. By using the heat load forecast information buildings will consume less energy and this will also decrease the greenhouse gas emissions. Further, energy savings also lead to financial savings for the energy production company. This ultimately also leads to reduced costs for the building occupants. These energy efficient buildings thus contribute to sustainable and energy efficient cities.

4.5 Summary

In this chapter, we discussed the results of the heat load forecast for the proposed model. We evaluated and analysed the performance of the model over two seasons and different horizons in three residential buildings. By computing the average accuracies across both seasons we observed that current heat load consumption and outdoor temperature forecast are the two parameters with the most influence on the heat load forecast. We also investigated the classification errors and explained the reason behind the resulting accuracy of the model. We also observed that model achieves a good accuracy for forecasting the heat load for the next hour by using less training data. We estimated the energy savings using the heat load forecast information for $HL_{(t+1)}$ in all three buildings. In the next chapter, we conclude the thesis work and discuss the future research directions.

5 CONCLUSION AND FUTURE WORK

In this chapter, we discuss the conclusion and future work related to this thesis work. The limitations of the Bayesian approach are also discussed. The objective of the thesis was to develop heat load forecasting models using the Bayesian approach. The heat load forecast in buildings was computed by studying the impact of several parameters on the heat load consumption by developing a machine learning model based on a Bayesian network.

5.1 Conclusion

This thesis work presented a Bayesian approach for forecasting the heat load in residential buildings in a district heating system. We studied that forecasting the heat load consumption helps in optimizing the heating production. The district heating operation was studied to identify the parameters influencing the heat load consumption in residential buildings. The parameters identified for heat load forecast included DHS operational parameters (supply temperature, return temperature, flow rate, difference between supply and return temperature), outdoor temperature forecast, behavioural parameters (hour of day and day of week) and current heat load consumption. We considered the Bayesian inference methodology due to its advantage of assessing the probability of uncertain and non-repeateable events. We modelled a Naive Bayes network and used two discretization techniques for converting continuous attributes to discrete states, in our proposed model. The forecast model was built by utilizing the realistic district heating data over a period of 4 months across winter and spring seasons from three residential buildings in Skellefteå, Sweden. Heat load forecasting was performed for horizons of 1, 2, 3, 6 and 24 hours to consider the effect of the district heating control loop and daily heat load consumption pattern.

Our results indicate that the current heat load consumption and outdoor temperature forecast are the two most important parameters influencing the heat load forecast. In this case, our model achieves average accuracies of 81.23% and 76.74% for a forecast horizon of 1 hour, $HL_{(t+1)}$ in three buildings for winter and spring seasons, respectively. We also observe that the combined influence of DHS operational parameters and outdoor temperature forecast on the heat load forecast is more significant than their individual influence. Further, by utilizing only 10% of training data, our model was able to achieve an average accuracy of 77.97% for the three buildings across both seasons with forecast horizon of 1 hour, $HL_{(t+1)}$. We observed that the forecasting accuracy in Building A was higher than in Buildings B and C because of less variation of heat load in Building A. We analysed the results of the forecasting accuracy by diagnosing the probabilities of various states of heat load forecast by setting evidence to particular states of outdoor temperature. We concluded that classifier errors are mostly due to the high variation in heat load consumption for the same values of outdoor temperature. We also conclude that by utilizing the Bayesian approach we were able to forecast the heat load for the next hour with a good accuracy by just utilizing the 10-20 percent of the training data. We believe from our results, that in case of a large number of buildings and large amount of data, our model would be suitable for heat load forecast. The estimated energy savings in all three buildings show that the heat load forecast information predicted by the model plays a key role in making buildings sustainable and energy efficient.

5.2 Limitations of the Bayesian approach

Though Bayesian networks offer a number of advantages, they also have some limitations. In this section, we discuss the limitations of the Bayesian approach. Bayesian networks use probability distribution function to represent uncertainty. However, the probability distribution function is not enough to represent uncertainty as it is not able to show the ignorance or uncertainty of the function itself [67]. The research work in [67] lists some of the approaches which extend the Bayesian approach by utilizing more than one value to represent the uncertainty and thus incorporating ignorance. One of the methods uses a probability interval where probability is not defined as a specific value but as an interval. The width of the interval indicates the ignorance of the system [67]. Another method uses a frequency value and a confidence value to represent the uncertainty. The confidence value is used to indicate the ignorance of the system [67]. Dempster-Shafer theory is also used to represent ignorance [41]. This theory has some conceptual differences from the Bayesian approach. It does not require prior knowledge for the computation of evidence for a proposition. In absence of prior knowledge, the Dempster-Shafer theory assumes ignorance. However, in Bayesian approach prior knowledge is required [68].

As discussed in the last paragraph, the Bayesian approach requires using a prior probability distribution. In cases when prior knowledge is vague or not available, it becomes a challenging task to specify a prior probability distribution. The prior knowledge also depends on the domain expert. Different experts can have a different opinion about prior knowledge which may lead to different prior distributions. This ultimately leads to different inferences through Bayesian analysis. When a sufficient sample size or dataset is available then prior distributions don't influence the final outcome of the Bayesian analysis [69].

There is also a limitation concerning the design of the Bayesian network. The designed Bayesian network can only represent the causal influences which are identified by the person designing the network [36]. The complexity of the Bayesian network also depends on the design. A Bayesian network with a large number of random variables and several dependencies will have a high complexity. On the other hand, modular Bayesian networks representing a particular system component can also be developed. Modular Bayesian networks lower the complexity of the entire system and can be later integrated into a single Bayesian network [70].

5.3 Future work

In this section, we discuss the future research directions which could provide more insights into the heat load forecast and help to improve upon the existing work. Some of the possible future works concerning this thesis is discussed in the following subsections.

5.3.1 Heat load disaggregation

The objective of this thesis work was to forecast the aggregated heat load demand in residential buildings. The aggregated heat load consumption is a sum of space heating and hot water consumption. The work presented in [71] discusses the advantage of separating the hot water heat load from the space heat load. Generally space heating consumption varies gradually whereas the hot water consumption varies at a faster rate depending on the user behaviour. The splitting of the heat load consumption offers the advantage to forecast both the loads independently by studying their individual patterns of consumption. Some of the methods used for disaggregation include distributed direct sensing, single-point sensing and intermediate sensing methods [72]. The future research work can focus on forecasting space heating load and hot water load separately. Individual forecasts may be more precise and accurate and this would possibly result in achieving a better overall forecasting accuracy than the aggregated method proposed in this thesis work.

5.3.2 Detailed study of weather parameters

The work presented in this thesis utilized only outdoor temperature as a weather parameter for studying the heat load forecast. The work in [14] and [33] have used weather parameters like outdoor temperature, humidity, solar radiation, barometric pressure, wind speed, rain gauge and wind position, for estimating heat load forecast. The study in [12] explains how solar radiation and wind influence cause heat load variation in the buildings. In the future work, it would be interesting to study the influence of various weather parameters on the heat load forecast subject to the availability of weather data.

5.3.3 Study of building characteristics and occupant behaviour

In the future work, the influence of building characteristics on the heat load forecast can be studied. The work presented in [32] studied the influence of building orientation, building envelope thermal resistance and thermal characteristics of materials on the heat load forecast. A study on the influence of consumer behaviour and building characteristics on the heating consumption may provide valuable insights about identifying the relevant parameters [73].

5.3.4 Bayesian diagnosis

As part of future work, a Bayesian diagnosis model can be developed on top of the existing Bayesian forecasting model. A Bayesian model utilizing probabilistic inference for both forecasting and diagnosis can provide insights into the study of various parameters influencing the heat load. Such a model would be able to diagnose the dependencies and relationships among the influencing parameters.

REFERENCES

[1] Decision no 406/2009/ec of the european parliament and of the council of 23 april 2009 on the effort of member states to reduce their greenhouse gas emissions to meet the community's greenhouse gas emission reduction commitments up to 2020.

[2] EU Parliament. Directive 2010/31/eu of the european parliament and of the council of 19 may 2010 on the energy performance of buildings, 2010.

[3] Molly Webb. Smart 2020: enabling the low carbon economy in the information age, a report by the climate group on behalf of the global esustainability initiative (gesi). *Global eSustainability Initiative (GeSI), Tech. Rep*, 2008.

[4] Kody M Powell, Akshay Sriprasad, Wesley J Cole, and Thomas F Edgar. Heating, cooling, and electrical load forecasting for a large-scale district energy system. *Energy*, 74:877–885, 2014.

[5] White paper on district heating and district cooling solutions in an environmental perspective. Technical report, Logstor, 2007.

[6] EC Directive. Directive 2004/8/ec of the european parliament and of the council of 11 february 2004 on the promotion of cogeneration based on a useful heat demand in the internal energy market and amending directive 92/42/eec. *Official Journal of the European Union*, pages 50–60, 2004.

[7] Samuel Idowu, Christer Ahlund, and Olov Schelen. Machine learning in district heating system energy optimization. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, pages 224–227. IEEE, 2014.

[8] Bruno Lapillonne, Carine Sebi, K Pollier, and N Mairet. Energy efficiency trends in buildings in the eu. *Lessons from the ODYSSEE/MURE project. ADEME, Supported by Intelligent Energy Europe. Online at: http://www. odysseeindicators. org/publications/PDF/Buildings-brochure-2012. pdf*, 2012.

[9] Johan Kvarnström, Jakob Liljedahl, and Erik Dotzauer. Forward temperatures and production planning in district heating systems. *10th International Symposium on District Heating and Cooling*, 3(5), 2006.

[10] Elnaz Abdollahi, Haichao Wang, Samuli Rinne, and Risto Lahdelma. Optimization of energy production of a chp plant with heat storage. In *Green Energy and Systems Conference (IGESC), 2014 IEEE*, pages 30–34. IEEE, 2014.

[11] Spyridon Provatas. An online machine learning algorithm for heat load forecasting in district heating systems. Master's thesis, Blekinge Institute of Technology, 2014.

[12] Henrik Gadd and Sven Werner. Daily heat load variations in swedish district heating systems. *Applied Energy*, 106:47–55, 2013.

[13] Samuel Idowu, Saguna Saguna, Christer Ahlund, and Olov Schelén. Forecasting heat load for smart district heating systems: A machine learning approach. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, pages 554–559. IEEE, 2014.

[14] Andrew Kusiak, Mingyang Li, and Zijun Zhang. A data-driven approach for steam load prediction in buildings. *Applied Energy*, 87(3):925–933, 2010.

[15] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[16] Hai-xiang Zhao and Frédéric Magoulès. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6):3586–3592, 2012.

[17] Meiping Wang and Qi Tian. Application of wavelet neural network on thermal load forecasting. *International Journal of Wireless and Mobile Computing*, 6(6):608–614, 2013.

[18] Wind. energy for the future. Technical report, Skellefteå Kraft, 2008.

[19] İbrahim Dinçer and Calin Zamfirescu. District energy systems. In *Sustainable Energy Systems and Applications*, pages 389–429. Springer, 2012.

[20] Maciej Grzenda and Bohdan Macukow. Heat consumption prediction with multiple hybrid models. In *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, pages 1213–1221. Springer, 2009.

[21] Milan Protić, Shahaboddin Shamshirband, Mohammad Hossein Anisi, Dalibor Petković, Dragan Mitić, Miomir Raos, Muhammad Arif, and Khubaib Amjad Alam. Appraisal of soft computing methods for short term consumers' heat load prediction in district heating systems. *Energy*, 82:697–704, 2015.

[22] S Grosswindhagera, A Voigtb, and M Kozeka. Online short-term forecast of system heat load in district heating networks. *Tsp*, 1:2, 2011.

[23] S. Werner L.C. Olsson Ingvarson. Building mass used as short term heat storage. In *The 11th International Symposium on District Heating and Cooling*, 2008.

[24] How does microchp technology work in homes and businesses? [online]. Available at: https://www.renewableenergyhub.co.uk/how-does-microchp-work-in-a-home-or-business.html. Accessed: 2015-03-31.

[25] Nelson Fumo. A review on the basics of building energy estimation. *Renewable and Sustainable Energy Reviews*, 31:53–60, 2014.

[26] Aurélie Foucquier, Sylvain Robert, Frédéric Suard, Louis Stéphan, and Arnaud Jay. State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, 23:272–288, 2013.

[27] Michel Verleysen and Damien François. The curse of dimensionality in data mining and time series prediction. In *Computational Intelligence and Bioinspired Systems*, pages 758–770. Springer, 2005.

[28] Krzysztof Wojdyga. Predicting heat demand for a district heating systems. *International Journal of Energy and Power Engineering*, pages 237–244, 2014.

[29] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[30] Fredrik Wernstedt. Multi-agent systems for district heating management. *Blekinge Institute of Technology, Department of Software Engineering and Computer Science, Ronneby, Sweden Licentiate Series*, (2003):07, 2003.

[31] Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term heat load forecasting for single family houses. In *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pages 5741–5746. IEEE, 2013.

[32] Tiberiu Catalina, Vlad Iordache, and Bogdan Caracaleanu. Multiple regression model for fast prediction of the heating energy demand. *Energy and Buildings*, 57:302–312, 2013.

[33] Maria Vlachopoulou, George Chin, Jason C Fuller, Shuai Lu, and Karanjit Kalsi. Model for aggregated water heater load using dynamic bayesian networks. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2012.

[34] DP Chassin, K Schneider, and C Gerkensmeyer. Gridlab-d: An open-source power systems modeling and simulation environment. In *Transmission and Distribution Conference and Exposition, 2008. T&# x00026; D. IEEE/PES*, pages 1–5. IEEE, 2008.

[35] Karan Mitra, Arkady Zaslavsky, and Christer Ahlund. Context-aware qoe modelling, measurement and prediction in mobile computing systems. *IEEE Transactions on Mobile Computing*, 99(PrePrints):1, 2014.

[36] Scott McCloskey. Probabilistic reasoning and bayesian networks [online]. Available at: http://www.cim.mcgill.ca/ scott/RIT/researchPaper.html, 1999. Accessed: 2015-04-28.

[37] Rajiv Ranjan. Modeling and simulation in performance optimization of big data processing frameworks. *Cloud Computing, IEEE*, 1(4):14–19, 2014.

[38] John Halloran. Classification: Naive bayes vs logistic regression. Technical report, Technical report, University of Hawaii at Manoa EE 645, 2009.

[39] Jin Huang, Jingjing Lu, and Charles X Ling. Comparing naive bayes, decision trees, and svm with auc and accuracy. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pages 553–556. IEEE, 2003.

[40] John W Stevens. What is bayesian statistics? Hayward Medical Communications, April 2009.

[41] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, Inc. and Dorling Kindersley Publishing, Inc., second edition, 2009.

[42] Abraham Otero. Bayesian networks. University Lecture, Universidad del Turabo, Available at :http://biolab.uspceu.com/datamining/pdf/BayesianNetworks.pdf, 2007.

[43] Finn V Jensen. Bayesian networks basics. *AISB quarterly*, pages 9–22, 1996.

[44] Kevin Murphy. A brief introduction to graphical models and bayesian networks. [online]. Available at: http://www.cs.ubc.ca/ murphyk/Bayes/bnintro.html, 1998. Accessed: 2015-05-02.

[45] Kevin Murphy. A brief introduction to bayes' rule. [online]. Available at: http://www.cs.ubc.ca/ murphyk/Bayes/bayesrule.html, 1998. Accessed: 2015-05-03.

[46] Tony O'Hagan. Dicing with the unknown. *Significance*, 1(3):132–133, 2004.

[47] Glen B. Alleman. Both aleatory and epistemic uncertainty create risk., May 2013. Accessed: 2015-07-08.

[48] Pasi Jylänki. Bayesian inference. University Lecture. Aalto University., 2008. Accessed: 2015-07-08.

[49] F Sandin, J Gustafsson, and J Delsing. Fault detection with hourly district energy data. *Stockholm: Swedish district heating association*, 2013.

[50] Pat Langley and Stephanie Sage. Induction of selective bayesian classifiers. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, pages 399–406. Morgan Kaufmann Publishers Inc., 1994.

[51] Daniela Xhemali, Chris J Hinde, and Roger G Stone. Naive bayes vs. decision trees vs. neural networks in the classification of training web pages. 2009.

[52] Nir Friedman, Moises Goldszmidt, et al. Discretizing continuous attributes while learning bayesian networks. In *ICML*, pages 157–165, 1996.

[53] George H John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995.

[54] James Dougherty, Ron Kohavi, Mehran Sahami, et al. Supervised and unsupervised discretization of continuous features. In *Machine learning: proceedings of the twelfth international conference*, volume 12, pages 194–202, 1995.

[55] Ying Yang and Geoffrey I Webb. Discretization for naive-bayes learning: managing discretization bias and variance. *Machine learning*, 74(1):39–74, 2009.

[56] Ying Yang and Geoffrey I Webb. A comparative study of discretization methods for naive-bayes classifiers. In *Proceedings of PKAW*, volume 2002. Citeseer, 2002.

[57] Huan Liu, Farhad Hussain, Chew Lim Tan, and Manoranjan Dash. Discretization: An enabling technique. *Data mining and knowledge discovery*, 6(4):393–423, 2002.

[58] Keki B Irani. Multi-interval discretization of continuous-valued attributes for classification learning. *In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 1022–1027, 1993.

[59] Chun-Nan Hsu, Hung-Ju Huang, and Tzu-Tsung Wong. Implications of the dirichlet assumption for discretization of continuous variables in naive bayesian classifiers. *Machine Learning*, 53(3):235–263, 2003.

[60] Daniela Joiţa. Unsupervised static discretization methods in data mining. *Titu Maiorescu University, Bucharest, Romania*, 2010.

[61] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.

[62] Xindong Wu, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, S Yu Philip, et al. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, 2008.

[63] Matteo Matteucci. A tutorial on clustering algorithms. *See at: http://home. deib. polimi. it/matteucc/Clustering/tutorial/html/index. html*, 2008.

[64] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[65] Marek J Druzdzel. Smile: Structural modeling, inference, and learning engine and genie: a development environment for graphical decision-theoretic models. In *AAAI/IAAI*, pages 902–903, 1999.

[66] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[67] Pei Wang. The limitation of bayesianism. *Artificial Intelligence*, 158(1):97–106, 2004.

[68] James C Hoffman and Robin R Murphy. Comparison of bayesian and dempster-shafer theory for sensing: a practitioner's approach. In *SPIE's 1993 International Symposium on Optics, Imaging, and Instrumentation*, pages 266–279. International Society for Optics and Photonics, 1993.

[69] Siva Sivaganesan. Criticisms/disadvantages of bayesian approach. University Lecture. University of Cincinnati., 2004. Accessed: 2015-07-08.

[70] CA Pollino and C Henderson. Bayesian networks: A guide for their application in natural resource management and policy. *Landscape Logic, Technical Report*, 14, 2010.

[71] Peder Bacher, Henrik Madsen, Henrik Aalborg Nielsen, and Bengt Perers. Short-term heat load forecasting for single family houses. *Energy and Buildings*, 65:101–112, 2013.

[72] Jon Froehlich, Eric Larson, Sidhant Gupta, Gabe Cohn, Matthew Reynolds, and Shwetak Patel. Disaggregated end-use energy sensing for the smart grid. *IEEE Pervasive Computing*, 10(1):28–39, 2011.

[73] Olivia Guerra Santin, Laure Itard, and Henk Visscher. The effect of occupancy and building characteristics on energy use for space and water heating in dutch residential stock. *Energy and buildings*, 41(11):1223–1232, 2009.