Høgskolen i Østfold

MASTER THESIS

Data Analytics for Smart Buildings

MUHAMMAD ALI

Supervisor: Prof. Dr. Eng. Lucian Mihet

Fredrikstad,15.05.2023

Green Energy Technology

Department of Engineering at Østfold University College

Abstract

The sustainability of the planet is significantly impacted by buildings. Numerous research has been done over the past few decades to improve the performance of buildings at different stages of their lifecycles. The future structures will provide residents with increased convenience, comfort, and efficiency options. The way people live will change as technology becomes more interrelated in their daily lives and into the things and activities they use to live. Making people's lives as simple and comfortable as feasible is a future expectation of smart buildings. To enable prompt actions and better decision-making, it is necessary to analyze the huge streaming data generated and captured by smart building appliances and devices. The provision of such intelligent services will surely be made possible by machine learning and big data analytics. This research mainly focuses on using data analysis and machine learning techniques. To prepare the data for analysis, it is preprocessed, cleaned, and converted. Then, a variety of statistical and visualization methods are used to spot patterns and trends in energy usage.

Because less specific building information is needed and data-driven approaches have high computational efficiency for online applications, they have been widely used. A new big-data-driven research paradigm has emerged as a result of recent developments in information technology and data science that made it simple to acquire, store, and analyze enormous on-site measurements. To fully understand the potential of constructing operational data in energy forecasting, big data-driven methodologies are necessary. This study examines the efficiency of modern recurrent neural network-based methods for making energy predictions.

The research approach includes gathering and preprocessing data regarding the building's energy usage from the grid and solar PV. In order to get the data ready for machine learning models, exploratory data analysis and feature engineering are used. Regression models, descriptive models, and time series models are some of the machine learning models employed in the study. The investigation's findings show the ability of data analysis and machine learning to comprehend patterns of energy usage and identify variables that impact it. In particular, the analysis presents insights into the patterns of usage of energy and the effectiveness of solar PV, which could guide the creation of more effective energy management plans.

Table of Contents

List of figures

Chapter 1: Introduction

In today's world, it is widely accepted that human activities are damaging the environment and have sped up both global warming and climate change. The emissions created by the energy needed for lighting and HVAC (heating, ventilation, and air conditioning) systems in building structures have increased these environmental hazards [2]. Buildings account for a considerable portion of global energy usage. The building sector is responsible for approximately 40% of the world's carbon dioxide emissions and more than 30% of final energy consumption, according to the International Energy Agency (IEA). Energy consumption and carbon emissions are expected to continue increasing in upcoming years. Large high-rise buildings with complex structures and multiple uses have recently become more common due to urbanization's growth and advancements in building technology [1]. Building energy modeling has become increasingly difficult due to the linked impacts of the building envelope, energy systems (such as air conditioning and thermal storage), automated control systems, and climate variables. Additionally, one of the most crucial elements in the modeling of the overall energy performance of buildings is the energy-related behaviors of the occupants [1].

Smart buildings must rely on cutting-edge tools to learn, forecast, and make wise decisions to maximize comfort, reduce expense, and adapt to the needs of its occupants. Prediction, decision-making, robotics, smart materials, wireless sensor networks, multimedia, mobile computing, and cloud computing are just a few of the technologies that are covered by smart building algorithms. Buildings can cognitively handle a variety of Smart Building services, including security, privacy, energy efficiency, lighting, and maintenance, elderly care, and multimedia entertainment [3].

Building performance modeling has been significantly hampered by the ever-increasing complexity of building energy systems and the continuously improved interactions between occupant behavior and building elements. These results suggest that in the near future, increased construction activity in industrialized countries combined with ineffective energy management in older buildings will drive up energy use and amplify its negative effects. The use of more sophisticated tactics to adjust and reduce energy consumption as well as to locate alternative and sustainable energy sources is also required by changing energy costs. Data science is now applied in the field of building energy management to solve issues like:

- The analysis of building operations, equipment status, and failures to optimize operation and maintenance costs.
- The prediction of energy demand to adapt production and distribution.
- The analysis of equipment status and failures to optimize operation and maintenance costs.
- The detection of energy consumption patterns to develop specialized commercial offers and to spot fraud [2].

1.1. System Overview:

An innovative approach to sustainable and effective energy management is a smart building with PV generation and two loads—primary and secondary (see fig.1). PV panels produce power during the day and supply it to the main load. Any extra energy is also saved in the battery for subsequent use. The PV system exports any extra energy to the grid if it is still available.

When there is enough sunshine available during the day, the PV system can produce energy. The primary load is powered by the energy produced by the PV system, and any extra energy is stored in the battery system. When there is no sunshine during the night, the battery system is utilized to power the main load by storing the extra energy produced by the PV system during the day. If the PV system is not producing enough energy because of poor weather, the battery system also supplies electricity to the loads during the day.

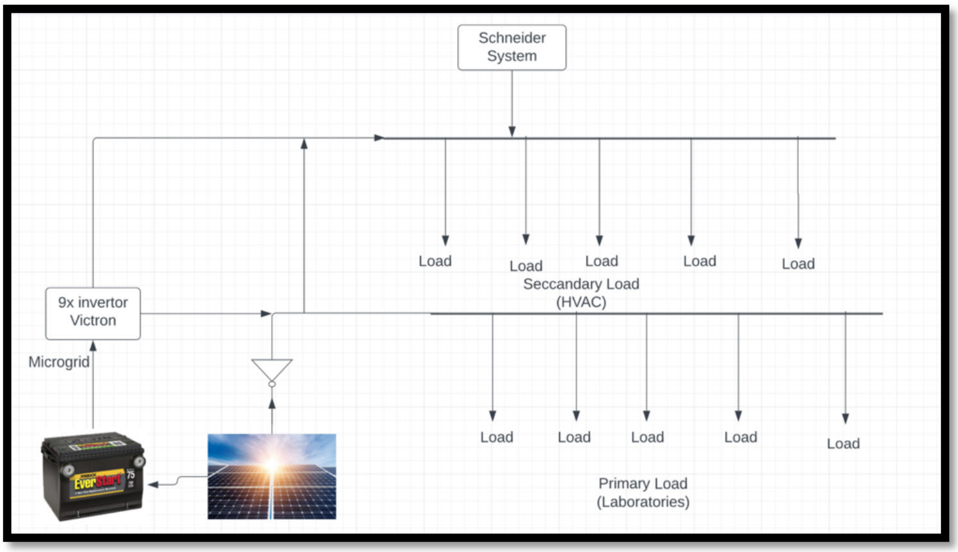

Fig. 1. Smart building with PV Generation.

The building's secondary load is directly connected to the grid, but it can also utilize the energy that the PV system exports during the day. This configuration makes sure that the PV system generates the most energy possible before switching to grid power. If there is any remaining extra energy, it is exported to the grid. The secondary load uses the PV system's exported electricity during the daytime when it is producing excess energy. At night, the secondary load draws power from the grid. Both loads get their electricity from the grid at night, when the PV system is not producing energy. The arrangement, however, permits the battery to empty and power both loads during the night if the weather forecast suggests that the next day will be sunny.

The energy generation from the PV system and the energy consumption of the primary and secondary loads are both monitored by an intelligent energy management system that is part of the Smart Microgrid integrated building system. Depending on the demand and availability for electricity, the system automatically shifts the load between the grid and the PV system. In accordance with the energy requirements of the primary load, the energy management system also handles the battery system's charging and discharging states.

1.2. The data-driven approaches:

Building energy prediction is not only a crucial element of smart buildings but also a crucial tool for evaluating the potential for energy savings during building design and refurbishment. Physics-based methods, hybrid methods, and data-driven methods are frequently used to forecast the energy performance of buildings [4]. The links between building energy consumptions and influencing elements are defined by physical approaches, which rely on physical principles and engineering experience. Usually, "white box" models are used to describe the generated models. The creation of "white box" models necessitates in-depth knowledge of building systems and is dependent on the veracity of physical hypotheses, both of which severely restrict their applicability. [5]. As opposed to this, data-driven models use inverse modeling to explain the connections between model inputs and outputs. Due to model parameters being determined based on actual building operational data, modeling processes are often more effective and adaptable [5].
In these conditions, a data-driven method is particularly interesting because it can quickly learn the building energy behavior from the building operation data and requires no prior knowledge of building and energy system designs and integrations. Advanced data mining and machine learning algorithms provide substantial technical assistance for big data analytics of building energy use. Sensing technologies and building automation systems have developed into credible sources for big data on building operations. Building energy performance modeling for building design, operation control, and policy making process is garnering growing interest [1].

In recent years, a variety of neural network models have been employed to forecast building energy usage. These AI techniques could produce extremely precise energy usage projections for time periods ranging from an hour to a week. various AI methods to calculate the heating and cooling loads of buildings. With mean absolute percentage errors around 4%, comparison results demonstrated that the ensemble technique (Support Vector Regression + artificial neural network) was the most accurate method for estimating the cooling demand and the heating load, respectively [31].

Machines need to be able to extract knowledge from the vast amount of sensory data that is being gathered from sensors and appliances, analyze it using algorithms, and then turn it into information so that they may comprehend humans more thoroughly than their surroundings.
Most importantly, such knowledge may result in new goods and services that fundamentally alter our way of life. Smart meter readings, for instance, can be utilized to more accurately forecast and balance energy use. New remote healthcare services can be created by monitoring and processing sensory input from wearable sensors affixed to patients [3].

As more smart meters are installed and more data is gathered, it becomes possible to use sophisticated data-driven models in real-world applications based IoT to learn how building energy usage patterns behave.

1.3. Data analytics:

This thesis provides the building energy consumption data analysis that is a crucial step in understanding consumption trends and increasing energy efficiency. It is feasible to understand how people use energy and spot areas for energy savings thanks to the abundance of data from numerous sources, such as smart meters, IoT devices, and weather stations.

Data from the building's energy sources, such as the grid connection, PV generation, and microgrid, are gathered to start the analysis. To ensure accuracy and consistency, the data is then pre-processed and cleansed. Trends, patterns, and anomalies in the data are found using exploratory data analysis approaches.

Machine learning techniques can be used to create predictive models of energy usage once the data has been understood. RNN LSTM models, which can account for the sequential character of the data and capture long-term dependencies, are one such method. These models may be taught to forecast future energy use based on historical data and are ideally suited for time-series analysis.

The RNN LSTM model can be trained on the measured data in a building with two loads and PV generation to forecast energy consumption for the primary and secondary loads. The model can also be trained to forecast how much electricity will be produced by the PV system and sent to the grid.

The estimated energy consumption can be used to pinpoint areas for potential energy savings and to make sure the building's energy systems are working as efficiently as possible. The model, for instance, can be used to decide when it is best to use the grid or a PV system to charge the battery.

In conclusion, building energy consumption patterns may be understood and future energy consumption can be predicted using data analysis and machine learning approaches. In buildings with numerous energy sources, such as PV generation, battery storage systems and grid connection, the application of RNN LSTM models enables accurate estimates of energy usage.

Chapter 2: Literature review

2.1. Smart buildings:

The idea of "smart buildings" refers to structures that have energy and technological systems integrated into them. The emphasis here is on resource management, occupant comfort, and energy efficiency [6]. The main challenge in smart buildings is balancing energy use requirements and occupant comfort. Air quality, visual comfort, and thermal comfort are three important factors that affect how comfortable building occupants are [7]. In this article, the authors paid particular attention to smart buildings, which experience their peak energy usage during the building's operational phase.
In addition to considering factors like weather and anticipated occupancy, the next generation of smart buildings must also be sufficiently adaptable to make the most of scheduling consumption around low energy price periods, local renewable resources, and energy storage. Production, energy storage, demand management, control, and communication are all components of a smart building's structure (see Fig. 2) [3], and all of which are managed by BEMS (Building energy management system) [8].

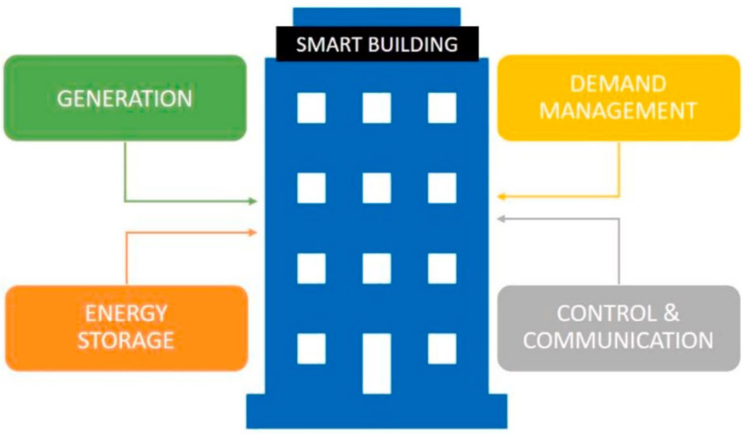

Fig. 2. Components of smart building structure.

2.1.1. Generation:

To improve the dependability and efficiency of the generation systems that result in collective, financial, and environmental benefits, traditional generation systems are now being rebuilt and transformed into intelligent grids. [9].
The definition of an intelligent building is a construction that is both energy-efficient and has a microgrid embedded into it. A little portion of the intelligent grid is known as a microgrid. Both a dispatchable load of the typical generation system and a separate generation system for small areas can be controlled using a microgrid. Industrial facilities, office buildings, and

homes have all made use of the fundamental concept of a small microgrid [9]. Microgrids powered by photovoltaic (PV) technology are increasingly becoming a major energy source, especially for household energy users.

2.1.2. Energy storage:

Storage devices, like water tanks, heat/ice storage units, and batteries, are essential for reducing energy costs in building energy systems because they can help users take advantage of time-of-use electricity rates and renewable energy sources. Two unique types of storage are used in the building to promote the progression of various sorts of resources: battery energy storage and thermal energy storage [8]. A high energy storage device known as battery energy storage serves as a cushion to store extra energy and support system functioning when necessary. The movement of thermal power is regulated by the limitations of energy storage and is dependent on the general rule of energy storage systems [10].

2.1.3. Demand management:

Demand Side Management (DSM) is a collection of connected programs that have the common goal of giving clients long-term, cost-effective access to consistent and efficient energy.
Demand response activities can play a significant role in lowering the peak load in non-residential buildings. This improves power grid effectiveness and reduces expensive energy and peak demand prices [11]. Studies on BEMS in conjunction with demand response strategies have concentrated on reducing peak demand in buildings through end-use load control, lowering home energy management costs, smart home EMS for prosumers of residential buildings, cooling and heating systems in homerooms, real-time thermal EMS for intelligent homes, and peak load reduction in a smart building [8].

2.1.4. Control & Communications:

The communication and control networks in buildings are a concern for building automation; the systems include processing units, actuators, sensors, and communication. Because of the increasing use of sensor/actuator systems, it is now more feasible for automated technology to take over instead of the tenant, allowing for the maintenance of acceptable circumstances without tenant inefficiency. Building management's primary goal is to provide a comfortable space with good energy efficiency. The temperature could be changed by using sensors. Sensors with a Dead band were developed and used to stop recurring modifications between a sensor's two situations [8].
IoT is the ability to connect and manage devices via the architecture of intelligent buildings. Numerous conceivable sensor networks speak to various applications and typically include hybrid devices, these devices can have wired or remote access. They can be working to supervise and control. Wireless communications are the most common in modern technology because they avoid the challenges of connecting devices through wired routes in large networks [12]. Wireless communications are the most common in use today because they eliminate the challenges associated with connecting devices over wired pathways in big size networks. IoT

breakthroughs led to the development of novel technologies like the Wireless Sensor Network (WSN) [13].

2.2. Building Energy Management System:

Building energy management systems (BEMS), which are computer-based programs, monitor and control a building's mechanical and electrical systems, including the lighting, power systems, heating, and ventilation.

Due to the limited supply of fossil fuels, efficient energy use and the use of cleaner energy sources become extremely important. An energy infrastructure called a "smart grid" is used in this context to manage the production, transmission, and distribution of energy more effectively. An efficient system for energy production, distribution, and consumption is made possible by a smart grid, which enables two-way communication between all parties involved. The major energy consumers are structures like offices, malls, and other infrastructure [14]. Statistics show that the residential sector uses between 30 and 40 percent of the total amount of electricity produced globally. The demand for power is rising as more users enter the market for electricity. The emergence of smart technologies like the smart grid and smart energy management systems allowed customers to actively participate in demand response programs (DRPs) and, as a result, reduce their electricity rates. [15]

The need to find novel methods to lower and optimize building energy usage has grown along with the need to reduce energy consumption levels and the continued growth of energy demand by commercial buildings.

Building energy management systems (BEMS) have become more well-known recently as a result of growing interest in reducing and conserving building energy. The building's energy requirements can be managed and controlled via the building energy management system. Building energy management systems can minimize energy consumption while also providing indoor thermal comfort, a safe and healthy atmosphere in residential buildings, including institutional, commercial, and industrial structures [16]. The fundamental objective of a BEMS is to make it possible for facility managers and building owners to track and manage the energy usage of their buildings in real-time. Data from numerous sensors and devices, including temperature sensors, occupancy sensors, and power meters, are gathered from all throughout the facility to do this. The information is then examined and used to inform choices regarding how to improve the efficiency of the building's energy systems.

Moreover, BEMS systems can be used to automate energy-saving measures including turning off lights and HVAC units in empty spaces, modifying temperature setpoints based on occupancy and weather, and planning equipment operation to take place during off-peak times when energy prices are lower. Building owners and facility managers can increase the sustainability of their structures while lowering operating expenses and energy usage by putting these techniques into practice [15].

2.2.1. Strategies of Building Energy Management System:

Based on projections of renewable generation and load demand patterns, Energy Management Systems (EMS) enable clients to accomplish their goals and those of utility providers. These systems could keep an eye on and manage how much energy was used by buildings, machinery, and businesses in accordance with various defined control logics. BEMS is a term that describes a variety of technologies used to improve a building's ability to use energy efficiently while maintaining occupant comfort inside the structure [17]. BEMS are a crucial component of an intelligent grid that gives building managers control over and management of the energy utilized in their structures, reducing demand and energy use. Both residential and non-residential structures can use BEMS with great flexibility [8]. BEMS techniques are divided into two categories: active and passive. The foundation of passive approaches is raising user energy awareness and offering future strategies to indirectly affect and lower building energy use. The infrastructure of the building's actuators and sensors is the foundation of active approaches. They rely on controlling the actuators and technology in smart buildings to reduce energy waste situations. We divided BEMS into four management techniques based on active approaches: model predictive control, demand-side management, optimization and fault detection, and diagnosis, see fig 1.3 [8].

Fig. 3. BEMS Strategies.

2.2.1.1. Model predictive control:

Model Predictive Control (MPC) can anticipate the development of a response to control requests and, by understanding how to proceed, may act appropriately to carry out the required operation. Forecasting building energy use is crucial for making better decisions on reducing CO_2 emissions and energy use.

Forecasts for building energy use have been made using three different methodologies. The term "white-box" is widely used to describe physics-based approaches.

They use a simple method based on physics calculations to describe how buildings perform in terms of energy use. Data-driven approaches, sometimes known as "black-box" approaches, mostly rely on statistical analysis and artificial intelligence to assess and estimate the building's energy usage. Grey-box methods, also referred to as hybrid methods, are those that combine white-box and black-box methodologies.

2.2.1.2. Demand side management:

Demand Side Management (DSM) is a set of measures designed to improve the user side of the energy system. It progresses from increasing energy efficiency through the use of better resources, through intelligent energy rates with drivers for specific utilization arrangements, up to contemporary continuous control of allotted energy resources [3]. Demand response (DR) and energy efficiency are two ways that are frequently recognized in DSM (see Fig. 4) [8].

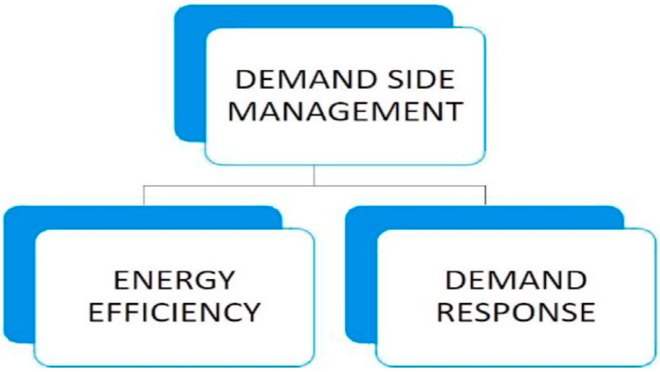

Fig. 4. Demand side management.

Energy-efficient structures can be created by fusing technologies like the production of renewable resources, energy storage systems, smart devices, and smart meters. Furthermore, it establishes the distributed generation, distributed storage, and distributed management (DSM) features of future intelligent grids.
Demand response is employed in non-residential buildings, but the energy efficiency technique is primarily used in residential and non-residential structures. The HVAC system has received the most study attention in demand response, followed by the energy efficiency strategy for appliance use and the HVAC system.

2.2.1.3. Optimization:

Building energy management systems with a stochastic approach have concentrated on increasing comfort index while using the least amount of power, effective policy measures, identifying energy consumption patterns, maximizing overall energy efficiency performance, load demand prediction of PV integrated intelligent buildings [8], and energy savings through analytics of actuators and information sources [18].
BEMS was associated with a robust approach that concentrated on efficient planning of the local energy system's constituent parts, managing multiple HVAC systems [19], controlling occupant comfort and energy use, coordinating cooling systems and individual fans, and energy use with prediction error [20].

2.2.1.4. Fault detection & diagnosis:

A building could be designed and built in an environmentally friendly and energy-efficient manner, but if the energy management system is not properly implemented, a significant amount of energy could be lost, increasing the cost of maintaining the facility. A programmed process called fault detection and diagnosis (FDD) is used to identify and isolate BEMS defects to protect a system from additional damage.

The FDD strategy for the field of BEMS can be divided into two types of techniques: knowledge-based and data-based.

Artificial intelligence is used to address FDD difficulties using data-driven approaches. With sufficient training data, the fault detection assignment is to determine if the examples of the supervisory information are similar to those of the typical training data [21]. Studies on BEMS that used a driven-based methodology concentrated on identifying unusual operating patterns and the root cause of heating system failures [22].

Knowledge-based techniques rely on experts to identify and detect flaws more successfully and reliably than the bulk of current FDD approaches, especially when analytic data is lacking or uncertain. Studies on BEMS that used a knowledge-driven methodology focused on analytical analysis for an air handling unit [23], identifying potential causes of inconsistencies for an air handling unit, and identifying and evaluating specific cooling system defects [24].

2.2.2. Building energy management system solution by Schneider Electric's eCommission:

An efficient next-generation building management system serves as a framework for the integration of building, business, device, and IoT data as well as segment-specific specialty systems, such as air quality monitoring for hospitals or room reservation systems for hotels. The ultimate goal of the BMS is to maximize occupant productivity and well-being while operating and maintaining building systems in a safe, dependable, and efficient manner. In comparison to traditional BMS systems that ASHRAE originally considered, next-generation systems have far wider scopes and capabilities. Additionally, they are simpler to use, program, scale, and update [41].

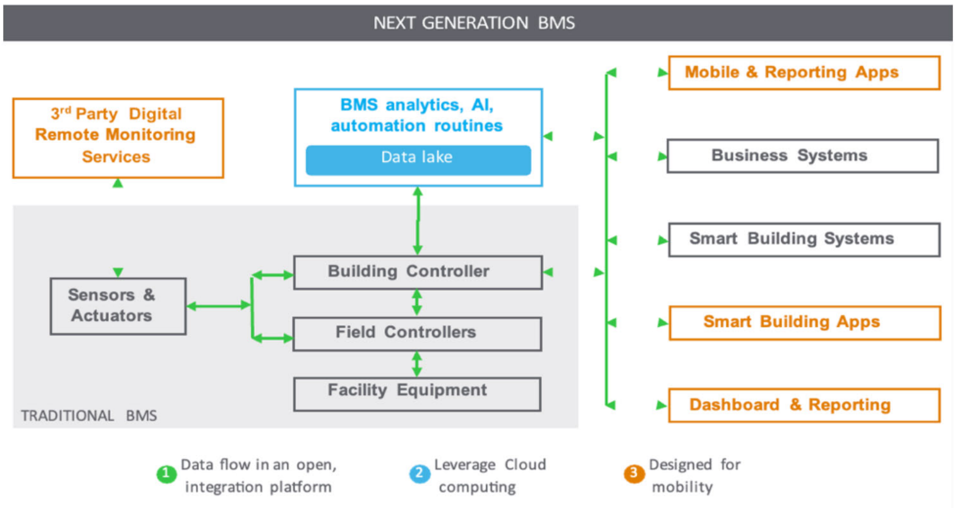

Fig. 5. An overview of the architecture of a modern BMS system that, in contrast to earlier BMS implementations, reaches from the hardware to the cloud [41].

A current, cutting-edge BMS solution spans from device sensors through building controllers (i.e., the edge) to the cloud with apps and services (See Figure 5) [41]. Next-generation systems enable knowledge reporting, the use of analytics and AI, predictive maintenance, and self-diagnosis, and they are accessible from anywhere.

This idea of a larger, more powerful BMS necessitates a design with three essential characteristics:

2.2.2.1. Based on an open, integration platform **:**

The system's ability to interact with other building subsystems, IoT devices, sensors, business processes, databases, and apps is its first crucial quality. BMSs have typically had a closed, proprietary nature due to their conventional, constrained scope. However, it goes without saying that the BMS must have the ability to gather data from and send data to systems and devices not typically integrated with a BMS if it is to perform more than just control the HVAC system. Within operations (OT), the Internet of Things (IoT), and information technology (IT), integration entails the adoption of open, standard protocols [41].

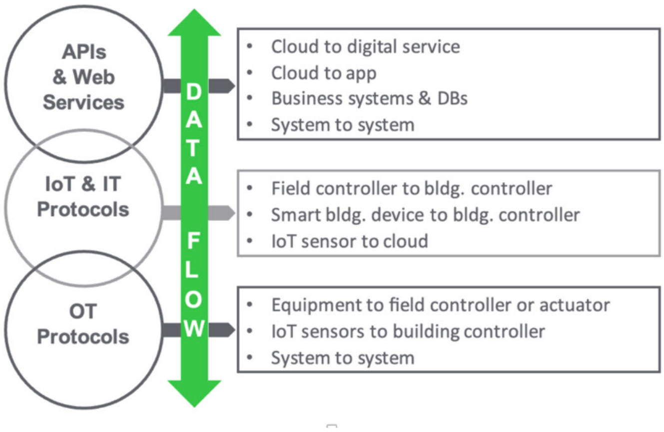

Fig. 6. Modern next-generation BMSs act as an aggregator from the device to the cloud by utilizing several open standards-based protocols [41].

2.2.2.2. Leverages cloud computing for analytics and AI-driven digital services**:**

Reduced costs, limitless scalability, enhanced resilience, mobility, simplicity of management and maintenance, and other benefits have made cloud computing a revolution in the IT sector. It is a crucial part of modern computing architecture around the globe. It also benefits the development of future building management systems. The usage of cloud computing technologies and services, whether public or private, is what we propose as being a crucial component of a next-generation BMS [41].

For a number of crucial building management capabilities and functions, such as:

• An infinite repository for storing, integrating, and backing up enormous amounts of OT a limitless repository for enormous volumes of previously segregated OT device and system data that can be stored, backed up, and integrated.

• A safe way for dependable partners or suppliers to do proactive remote monitoring of building systems and equipment as a service ("digital services").

• Facilitating "big data" analytics and training of machine-learning algorithms that produce insightful information, preventative maintenance, and control automation.

• Enables FMs to efficiently manage building fleets by providing visibility, alarm & data aggregation, and aggregated reporting across many, geographically distributed sites [41].

2.2.2.3. Designed for mobility:

Employers and building owners today desire a deeper level of interaction with their "mobile first" renters, customers, and inhabitants. For instance, people that need to access information frequently, such hotel visitors, university students, and employees, want to be able to engage with it. They anticipate and demand a certain amount of control, as well as simpler ways to execute chores. To satisfy these demands, a next generation building management system ought to provide mobile apps and services [41].

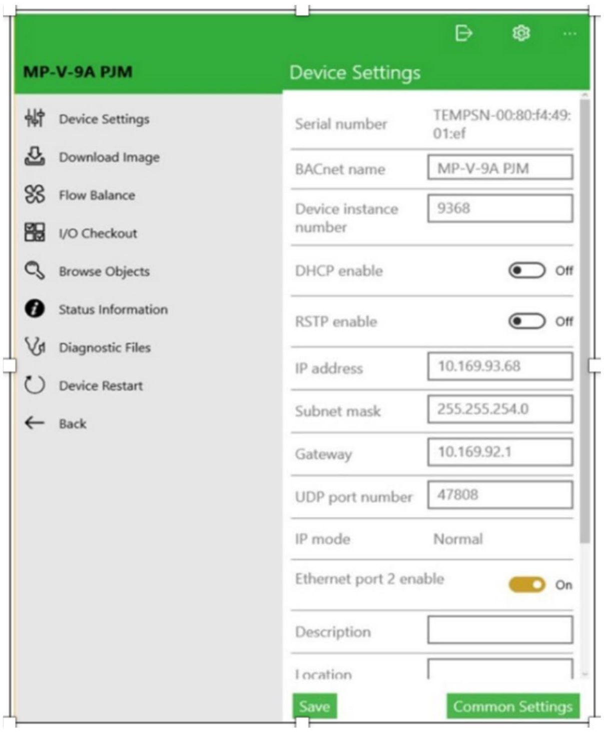

Fig. 7. A screenshot of Schneider Electric's eCommission SmartX Controllers mobile app [41].

For instance, people that need to access information frequently, such hotel visitors, university students, and employees, want to be able to engage with it. They anticipate and demand a certain amount of control, as well as simpler ways to execute chores. To satisfy these demands, a next generation building management system ought to provide mobile apps and services [41].

2.3. Microgrids:

The central module in charge of making distribution decisions is the micro grid. It also represents a small group of loads that are concentrated on a single feeder of a distribution system and can satisfy some or all of their needs using miniaturized power plants like solar cells, wind turbines, photovoltaic panels, micro-turbines, and diesel generators, respectively [14]. The load controller and other distributed energy sources are only two examples of the additional energy components that make up the microgrid. The microgrid can function in a variety of modes thanks to the integration of these two parts in the distribution network. Additionally, it makes it easier to store excess energy produced by various renewable sources so that it can be used later when demand is highest. Micro grids can be built as standalone, off-grid structures, or as a single connected grid. In a micro grid, security, efficiency, and electricity quality are three common problems. The problems are made more challenging by the changes in energy demand in smart buildings. IoT integration, however, might be able to help with these issues. In our suggested design, a centralized module capable of an integrated information flow, streamlined day-to-day operations, and analysis of the power distribution system is the true essence of using a micro grid [14].

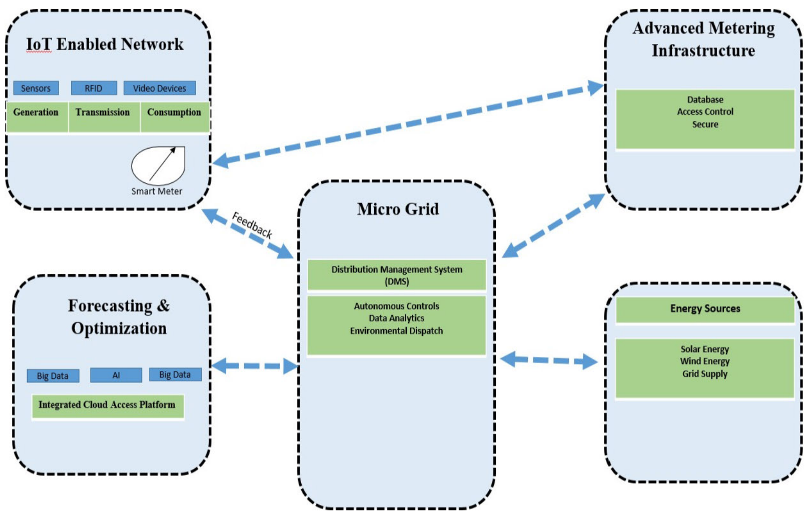

Fig. 8. Smart Microgrid integrated Smart buildings.

Forecasting and optimization make up the architecture's second part can be seen in fig.8 [14]. The microgrid's generation, storage, and energy consumption combine to produce a very

complex design. Predicting supply and demand is crucial for preserving the energy balance and delivering the necessary services.
Due to reciprocal stability, the security of several connected grids may be jeopardized. In order to retrieve the real-time state of the crucial parameters, data must be processed over an IoT enabled network using the available sensors.

The integration of the smart metering infrastructure is a key component of the concept in Fig. 8.[14]
This part is an integrated system of communication networks and smart meters that permits two-way communication between utility companies and their customers. The system performs a number of crucial tasks, including monitoring voltage, being able to connect and deactivate services, detecting tampering, identifying and isolating outages, and automatically and remotely measuring electricity use [14].

2.4. Big data analytics and Machine learning:

The vast amount of data that European buildings are producing from a variety of energy-related sources, including data from smart meters, sensors, and other Internet of Thing's devices, is posing new research problems. The purpose of this work is to describe a high-level data-driven architecture for real-time processing, data management, and interchange for buildings. Cross-domain data can be integrated with developing artificial intelligence algorithms and distributed ledger technologies thanks to this multidisciplinary big data ecosystem. A set of visualization, querying, and exploration tools, appropriate application programming interfaces (APIs) for data exchange, and a collection of customizable and ready-to-use analytical components that implement several advanced machine learning and deep learning algorithms are combined with semantically enhanced, interconnected, and multilingual repositories of heterogeneous types of data [25].

2.4.1. Big data:

The concept "big data" is used in this study to describe data sets that are so massive or intricate that they cannot be handled by conventional data-processing technologies. The "six Vs" can be used to describe big data: volume (the number of data points), velocity (the rate at which the data are generated and processed), variety (the number of different types of data), variability (the degree of consistency or inconsistency of the data), veracity (the quality of the data), and value (the results attained as a result of gathering and analyzing the data) [28].

Figure 9 [28] shows a variety of data analysis techniques that have been described in earlier works. For various purposes, data analytics employs a variety of methodologies.

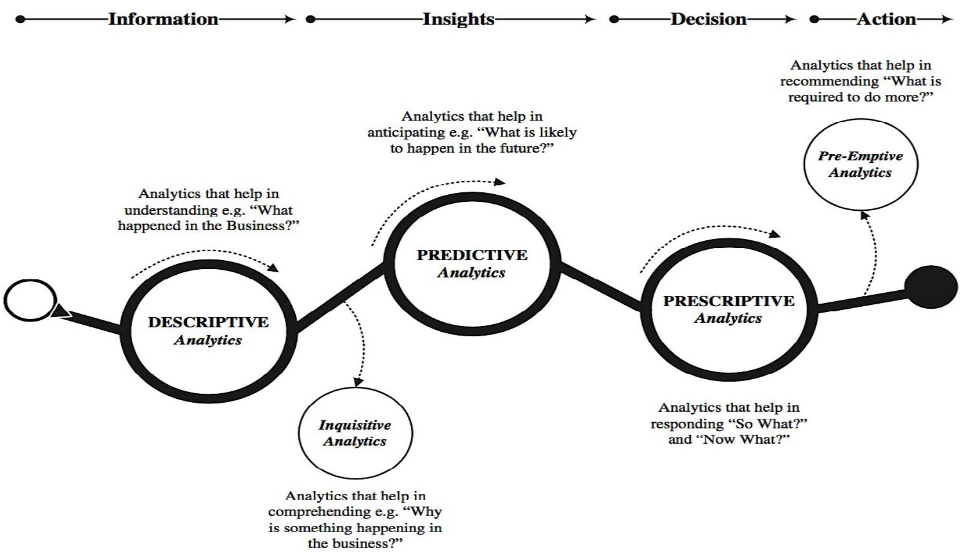

Fig. 9. Types of Data analytics.

Descriptive analytics gather and mine data to offer historical context ("what happened"). Many statistical, modeling, data mining, and machine-learning approaches are used by predictive analytics to analyze current and past data to predict the future. In order to predict potential outcomes and suggest the optimal course of action for any pre-specified outcome, prescriptive analytics uses optimization and simulation techniques [28].

2.4.2. Scalable Big Data Management:

Data analytics seeking to increase the accuracy of created algorithms and systems based on intelligent adaptive systems are fundamentally dependent on data processing through ML approaches. To extract useful but actionable information from the dataset and facilitate informed decision-making, it makes use of a variety of ML, statistical, and AI-based algorithms and models, including clustering, correlation, classification, categorization, regression, and feature extraction [26].
Based on the type of learning "signal" or "feedback" that a learning system has access to, ML approaches are further divided into supervised, unsupervised, semi-supervised, and reinforcement learning techniques. Most big data analytics and AI techniques for smart buildings are based on traditional (supervised or unsupervised) ML/DL algorithms that operate in a context-specific manner and as a result do not adequately support cross-stakeholder transfer learning, reusability of AI-based learning models, and quick cross-domain application adaptation [27].

Regression methods, lazy learners, decision trees, and support vector machines are examples of supervised learning. These techniques have been applied to the analysis of operational building data for short-, medium-, and long-term energy forecasting in various building environments [28]. The supervised learning approaches used to accurately capture the intricate interactions between input and output variables are often very complicated, such as deep neural networks, support vector machines, and decision-tree based ensembles [28]. Multiple linear regression and support vector machine approaches can reach a high accuracy with quick computation speed for buildings with complicated and erratic occupancy schedules and energy use patterns [29].

Unsupervised learning investigates the underlying relationships, correlations, and data structures.

Combining a big amount of labeled data with a comparatively small amount of unlabeled data improves learning accuracy in semi-supervised learning. The use of semi-supervised energy modeling for smart homes and buildings [28].

2.4.3. Big data-based BEMS architecture:

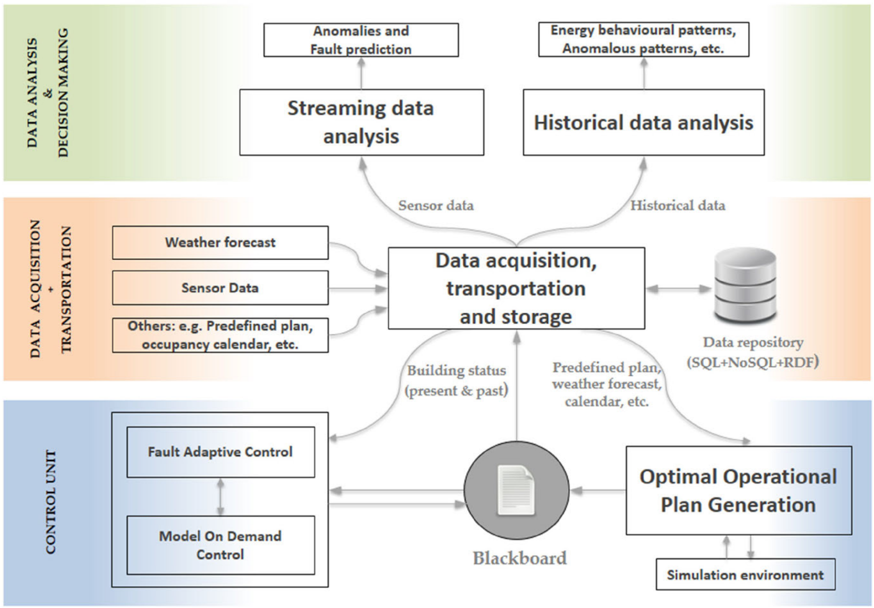

Fig. 10. Building energy Management system architecture based on big data [30].

In this part, we suggest a BEMS built on a Big Data architecture that is effective and fault tolerant. The entire system architecture used can be broken down into layers (see Fig. 10), each of which is in charge of a specific system component: data collecting and transportation, control unit, and data analysis and decision-making [30].

A) Data Acquisition Layer:

The management of all the data used by the other modules is the responsibility of the data acquisition and transportation layer.

Its main function is to establish a connection with the local BMS, retrieve the available data from it, and then make the data transparently available to the other modules. To effectively handle and store the data, data technologies are crucial. Other data handled by the data acquisition platform include energy consumption, energy demand, equipment defects, blackboard history, etc. These data are mostly supplied by the control unit layer.

In our situation, the data repositories at the pilot sites are made up of several databases that operate within a single framework. Relational, NoSQL, and RDF/OWL database management systems were used to construct this computing environment. MongoDB was specifically used to store data from sensing devices, as well as results from the control unit layer and the data analysis layer, in order to handle huge amounts of data. The optimal format to use to apply DM&ML techniques in the data analysis layer is HDFS format, which is used to store the processed data in files [30].

B) Control Unit Layer:

The daily operating plan is managed by the control unit layer, which aims to take advantage of equipment control, particularly HVAC. Model on demand control (MODC) and model predictive control (MPC) are the two components that make up this (MODC).
The MODC strategy aims to adapt the nominal control online with respect to the actual current system operating conditions, reducing the impact of potential deviations between actuals and forecasts on the ideal provided plan. MPC provides the nominal optimal plan in accordance with the weather forecast conditions [30].
Using weather forecasts, user occupancy, energy tariffs, building status, and other data regarding the anticipated building operation circumstances, MPC develops the best operational plan (OP). All this data is used to perform simulations in order to identify the ideal daily schedule that adapts to the needs of the occupants for comfort while using less energy. This produces an OP, which is transmitted to the operators using the blackboard and consists of a list of OP setpoints to be applied at specific time intervals (see Fig. 1) [30].

C) Data Analysis and Decision Making Layer:

In the last few decades, data analysis techniques have been used in the energy sector for a variety of objectives. We can specifically mention the following:
Energy demand prediction is necessary for a building to operate efficiently, along with building operation optimization, monitoring operational status and failure detection of building equipment and networks, analysis of the economic and commercial effects of user energy consumption, and detection and prevention of energy fraud.

D) Prediction of building energy load:

Energy demand, also known as energy load, is the quantity of energy needed at a specific time. Building load patterns can be difficult to detect due to numerous connected causes. Several suggestions for consumer profiling utilizing unsupervised learning, such as clustering and association rules, and classification models, such as decision trees, may be found in the literature. Anomaly detection techniques have been widely used to predict peak demand. Special consideration should be given to user behavior and how it affects energy use, and some research is being done to identify activities for load balancing and automatic energy management (heating, cooling, and lighting) [2].

E) Building operation:

Building energy management system operational data can be studied and used to extract patterns defining building operation. These patterns typically take the form of IF-THEN rules, which can aid in the creation of suggestions for control measures.
In this context, decision trees and association rules have been employed most frequently. Similar to this, other studies have investigated the relationships between control settings and energy usage, utilizing techniques like regression and neural networks, to optimize specific building components [2].

F) Economic analysis of electric consumption:

To learn and understand how and when their customers use energy, several businesses have turned to data science. Classification, clustering, and pattern analysis—most commonly using association rules—are the methods generally employed for this purpose. Classification helps to find common consumption categories. For the real-time analysis and summarizing of data on energy use, streaming data processing has also been taken into consideration [2].

Chapter 3: Methodology and Technology

The research seeks to analyze energy consumption data and predict the energy usage in a university laboratory building in Romania for the month of May 2022. The researchers made use of a dataset that had previously been acquired from the structure to accomplish this purpose. Victron Energy, an open-source web server that archives the collected data every 15 minutes, was used to collect the data. The technology also offers historical data visualization as well as real-time energy generation and consumption. This capability makes it possible for researchers to examine patterns and trends in energy usage and pinpoint variables that influence it.

The system determines how much energy is exported back to the grid and how much energy is supplied to loads by the grid. This is crucial in calculating the building's net energy use. The researchers also noted that a solar PV system had been placed on the building's roof; it supplies power to loads during the day and recharges a battery that is attached to it. This data is crucial for calculating the building's potential for renewable energy sources as well as the total amount of energy available for use.

This thesis focuses on the data analytics of building energy consumption data and Solar PV generating data. Getting insights of data that is useful for decision making and to find areas where energy efficiency can be improved. After analyzing the data machine learning techniques have been used to predict future energy consumption of building. Big data analytics technology has been used to achieve the results.

3.1. Methodology:

The machine learning approaches and data-driven statistical concepts utilized to analyze the data are briefly summarized in this section (see fig.11) [2].

Artificial Neural Networks (ANN) are a type of machine learning algorithm that can be used to make predictions based on historical data. ANN models are particularly useful for predicting energy demand as they can learn complex relationships between input variables (such as Day, temperature, etc.) and energy demand [32].

3.1.1. Data Collection:

To use an ANN model to predict energy demand, we need a dataset that contains both timestamp information and energy consumption in kilowatt-hours (kWh). The dataset should include a range of historical data points to allow the model to identify patterns and trends over time. Once we have the dataset, we can follow next steps to preprocess it , analyze and build an ANN model for energy demand prediction [32].

3.1.2. Normality testing of dataset:

To determine the dataset distribution in this research's data analysis, a normality test of the dataset for each day was carried out. By determining the dataset's kurtosis and skewness, this procedure was guided. Since normality testing typically implies that the dataset was normally distributed, it is important for model construction [32]. The normality testing is unreliable if the sample size is greater than 100. Understanding the dataset distribution, however, could offer a consequential analysis of the prediction's outcome. According to statistical analysis, kurtosis measures the peakness of the distribution, whereas skewness measures the irregularity of the probability distribution around the mean value [34].

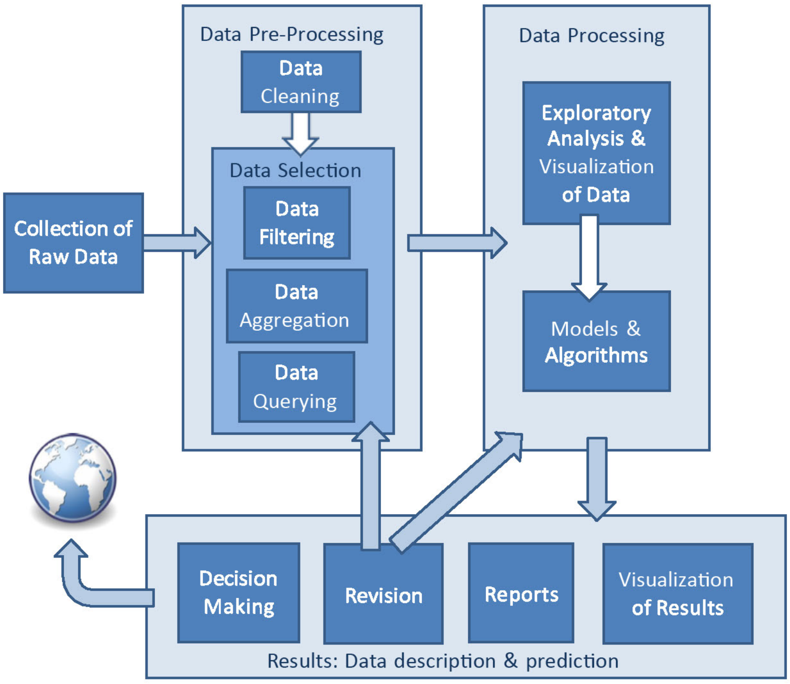

Fig. 11. Data Analytics Methodology.

3.1.3. Data pre-processing:

Data pre-processing, which typically takes a lot of time and computational capacity, is part of the machine learning preparatory process. This procedure is necessary since the dataset typically contains missing values and inconsistent value scales between features [32]. The data in this study was pre-processed using standardization and mechanics for imputed missing data, with the former utilizing the Python computer language in a Jupyter notebook. To delete, replace, or infer missing values, Python data analysis libraries like NumPy and pandas were employed. This module allows a variety of cleaning actions, such as replacing missing values with a placeholder, the mean, or another value, eliminating all rows and columns with missing values, or inferring values using statistical techniques [32].

3.1.4. Analysis and visualization of data:

- ***Descriptive statistics:*** To gain a general understanding of the data distribution and spot any outliers or anomalies, you can generate summary statistics like mean, median, standard deviation, min/max values, quartiles, etc.

- ***Data visualization:*** To visualize the distribution of the data, spot patterns or trends, and spot outliers or anomalies, you can make a variety of plots, including histograms, box plots, scatter plots, and line charts.

- ***Time series analysis:*** To find trends, seasonality, and other patterns in the data provided the dataset has a time component. This may entail employing statistical techniques like autocorrelation or Fourier analysis, as well as data visualization techniques like line charts or scatter plots.

- ***Correlation analysis:*** To determine the connections between various features and the objective variable (i.e., energy usage), correlation analysis is used. In order to do this, correlation coefficients can be computed, and the data can be shown graphically using scatter plots or heatmaps.

- ***Data exploration:*** To learn more about the connections between variables, spot correlations, and investigate various feature engineering choices, undertake exploratory data analysis (EDA).

- ***Energy efficiency analysis:*** Analysis of energy efficiency involves looking at energy usage trends to find areas where efficiency might be improved. The application of methodologies like benchmarking, energy intensity analysis, and energy performance indicators is possible.

3.1.5. Model development:

The study of interpretable machine learning is a developing area in big data analytics [37]. It tries to offer strategies and instruments to improve model interpretability while reducing model complexity [36]. Interpretable machine learning is particularly promise for the creation of smart and user-friendly building energy management systems, especially considering the practical challenges experienced by building experts when utilizing advanced supervised learning techniques as can be seen Fig.12 [35].

Machine learning approaches typically produce predictions that are more accurate than statistical techniques (such as multiple linear regression). However, there is an inherent trade-off between model interpretability and model complexity; for instance, machine learning models are "black-boxes" to users, and it is extremely challenging to understand the inference method learned [36].

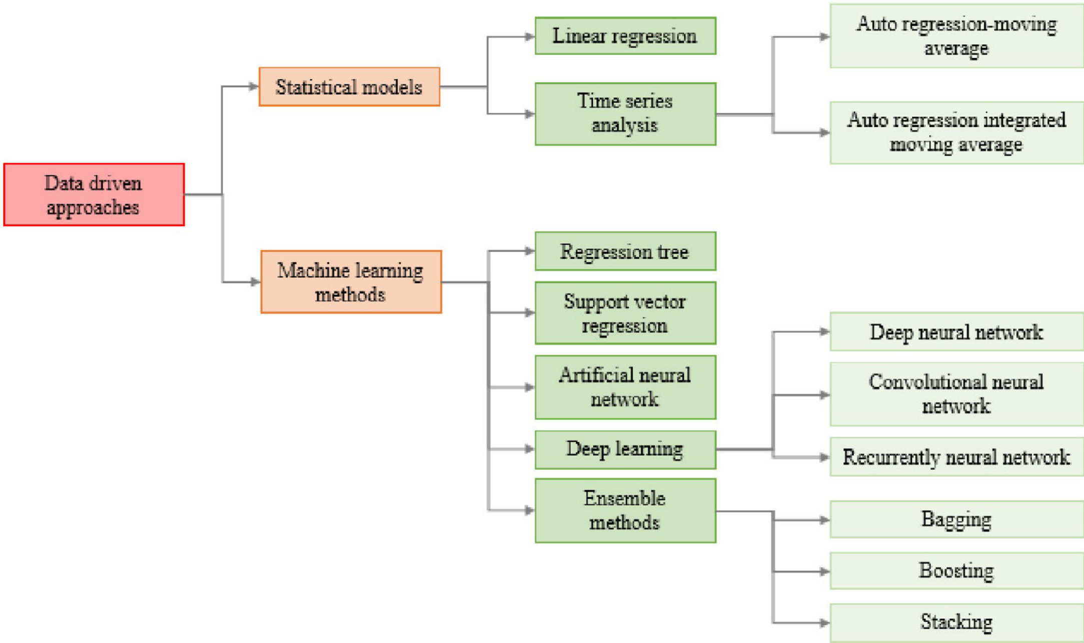

Fig. 12. Data-driven models for building energy prediction.

3.1.5.1. Artificial neural network-based methods:

The most prevalent regression algorithm for predicting building energy load is ANN. One input layer, one or more hidden layers, and one output layer make up an ANN in general. Each layer has a few synthetic neurons that are connected to the synthetic neurons in the levels above. Each connection has a weight that must be adjusted throughout the model training phase.
A typical four-layer ANN model for predicting building energy load is shown in Fig. 13 [38].

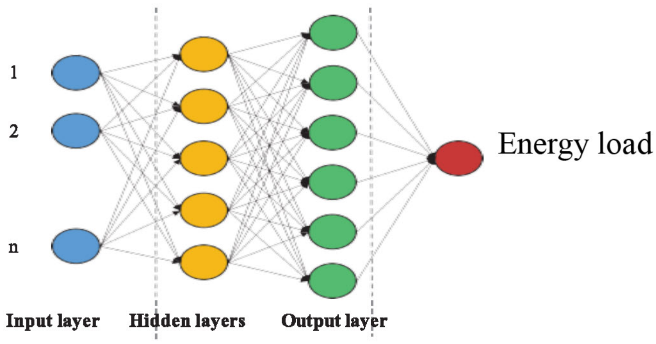

Fig. 13. illustration of a four-layer ANN-based energy load prediction method.

3.1.5.2. Generating Predictive model (training):

The training of the algorithm is the initial step ca be seen in Fig. 14 [32]. To forecast energy use, this study employed supervised machine learning techniques. The learning process was then used to enter the prepared data. The algorithm was given various feature combinations to provide a candidate for the prediction model. Data partitioning was done to divide the data into two groups—a training group and a testing group—before the data were used to build and train the model [32].

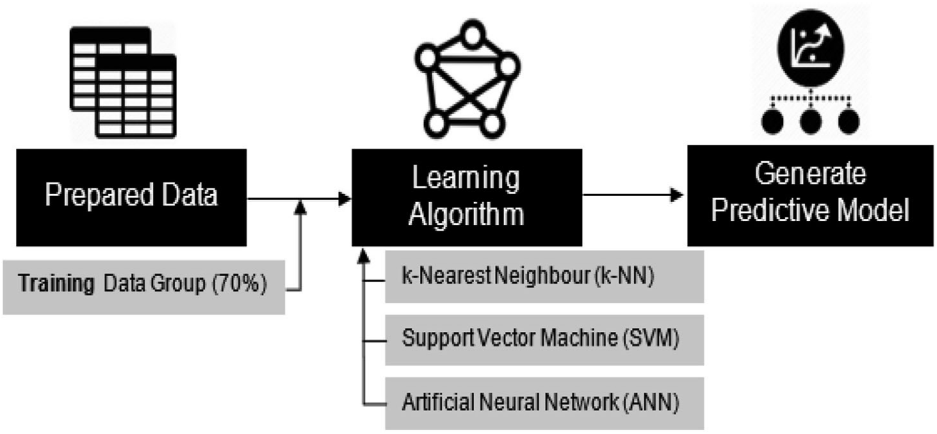

Fig. 14. Process of generating predictive model after data preparation [32].

Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) units can be used in building energy consumption data analysis using Artificial Neural Networks (ANN). RNNs are a particular kind of neural network made to deal with sequential data, such time series data. They possess the capacity to incorporate historical data into forecasts of future values. Traditional RNNs struggle to learn long-term dependencies in the data because of the problem of vanishing gradients, in which the gradients get smaller as they are backpropagated over time. LSTM units, which have the capacity to selectively retain and forget information over a longer period, were created as a solution to this problem. This makes it possible for LSTM to manage long-term dependencies in sequential data better.

3.1.5.3. Model evaluation:

The data was split into two groups before being input to the machine learning algorithm, with 70% of the dataset being utilized for training and the remaining 30% being divided into groups for testing. While the remaining data was held aside to be utilized to test the trained predictive model, the training groups of data were used to train each machine learning algorithm and produce a prediction model that could output value that fits with the recorded maximum demand data [32]. The procedure is depicted in fig. 15 [32].

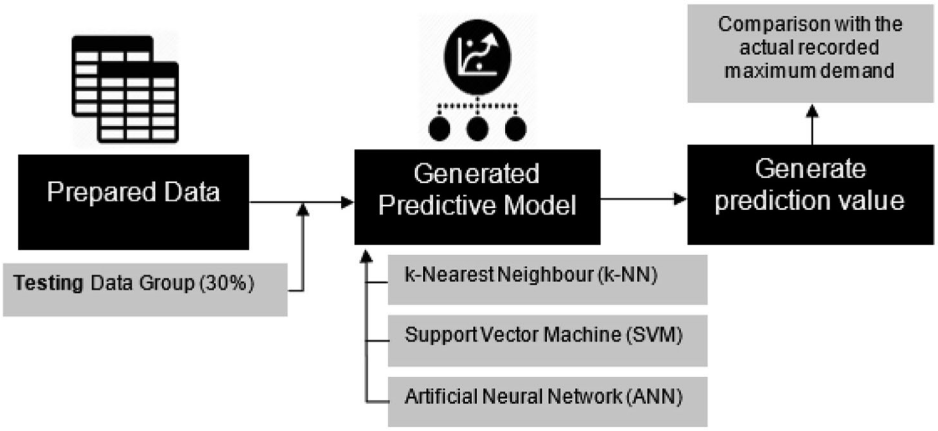

Fig. 15. Testing of the trained predictive model [32].

Results are produced by running the model on the training dataset. These results are compared to the original training data, and depending on the comparison's findings, the algorithm's various parameters can be modified to fit the training dataset [39]. The implemented algorithm, previously fitted on training data, is evaluated objectively using a validation dataset, and its essential modeling parameters are tuned to improve the model's fitting. The algorithm developed is tested on the remaining portion of the data in the third and final step, which provides a final, unbiased assessment of the modeling and predicting results. It is generally acknowledged that the model's structure and parameters shouldn't be changed in response to the findings of this last step [40].

3.2. Technology:

Technology describes the equipment, methods, and procedures used to gather, examine, and model the data. This comprises the computer hardware and software (such as programming languages, libraries, frameworks, and computational resources like CPUs, GPUs, and cloud services) used to carry out the analysis and machine learning.

3.2.1. Programing languages and libraries:

Python is a widely used programming language that is utilized for machine learning and data analysis for calculating building energy usage. It is a high-level, open-source language that is simple to read and write code in. Numerous libraries and frameworks, such as NumPy, Pandas, Scikit-learn, TensorFlow, and Keras, are available in Python that are intended primarily for data analysis and machine learning.

Python's Pandas module is a well-liked open-source tool for handling and analyzing data. It offers a robust and user-friendly set of tools for working with tabular data, allowing you to read and write data from different file formats, alter and clean data, run descriptive statistics, and visualize data.

Data on energy use can be loaded using Pandas from a variety of sources, including databases and CSV files. The library offers a variety of strong capabilities for pre-processing and cleaning the data, including eliminating missing values, filtering, grouping, and transformation.

Pandas is the best tool for evaluating building energy consumption data since it includes built-in support for time series data and is frequently collected at regular intervals. Powerful time series indexing and resampling features in the library make it simple to manipulate and aggregate time series data.

Popular Python module NumPy is widely used in machine learning and data analysis. It is the foundational library for Python's scientific computing and offers strong tools for manipulating arrays, matrices, and numerical calculations.

NumPy can be used to carry out several actions on the data, including filtering, aggregation, and statistical analysis, in the context of the analysis of building energy usage. It can be utilized, for instance, to:

- Create NumPy arrays from raw energy consumption data for straightforward modification and analysis.
- Perform operations on arrays, such as data filtering based on the amount of time or energy used.
- Determine summary statistics for various data subsets, such as mean, standard deviation, and percentile.
- Create data visualizations utilizing additional libraries, such as Matplotlib.

TensorFlow can be used to develop various machine learning algorithms to forecast future energy consumption based on historical data in data analysis and machine learning of building energy consumption. The design and training of neural networks is made simpler by TensorFlow's high-level Keras API. It is utilized to construct and train deep learning models of machine learning, including artificial neural networks, convolutional neural networks, and recurrent neural networks.

- TensorFlow enables distributed training, allowing for quicker training times on big datasets across a variety of processors and devices.
- Tensor Board: TensorFlow comes with a visualization tool called Tensor Board that may be used to view and assess how well machine learning models are performing.

A high-level neural network API called Keras was created in Python and may be used with TensorFlow, CNTK, or Theano. It offers an intuitive interface for creating and refining deep learning models, including models for analyzing data on building energy consumption.

Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, which are well-suited to the time series nature of energy consumption data, can be built and trained using Keras to develop energy consumption analysis.
For model evaluation and visualization, Keras offers a wide range of tools, including techniques for estimating loss and accuracy, visualizing model performance, and producing model predictions.

3.2.2. Hardware and software systems:

Victron Energy has a significant presence on the international market and is renowned for its reliable and high-quality goods. They additionally offer their clients technical help and training in addition to software and tools for managing and monitoring energy systems.

3.2.2.1. Victron Remote Management:

While the Energy Meter software offers real-time monitoring and administration of energy use in buildings, VRM is a cloud-based platform that lets users remotely monitor and manage their energy systems.

3.2.2.2. Hardware systems:

The hardware setup contains the following components:
Victron Battery monitors.
Victron Invertors/chargers.
Energy storage systems:
Solar panels.

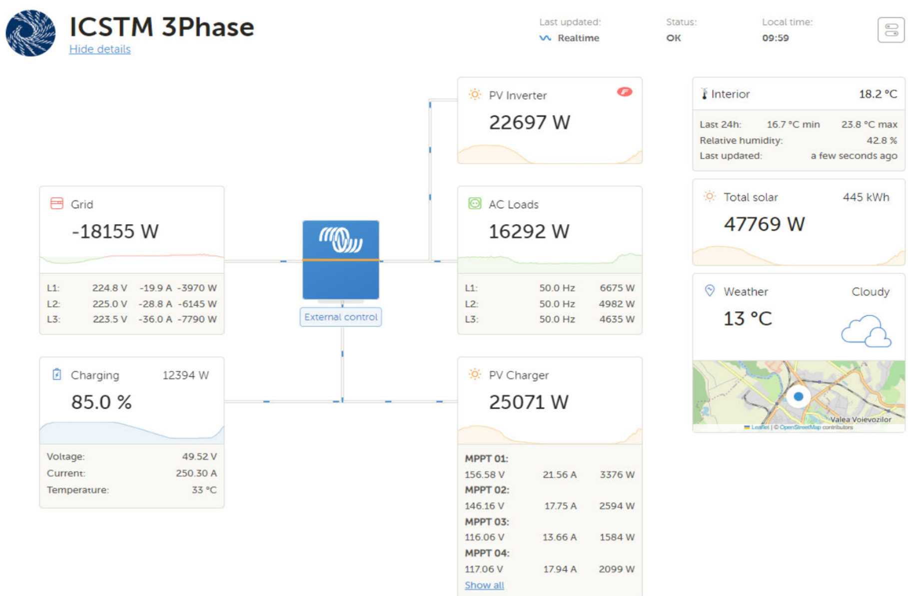

Fig. 16. Victron Remote Management system [42].

Chapter 4: Data Analysis and Results

Building energy consumption data analysis offers insights to boost energy efficiency in the building industry. In this work, we investigated the energy consumption patterns of a building using descriptive statistics, correlation analysis, and regression analysis.

The building has a two AC loads: Primary load and Secondary load. Primary loads consist of laboratories and Data center while secondary load is main HVAC systems. The main source of energy consumption is primary load. The microgrid is installed consist of solar PV and battery storage. Fig. 1 shows the AC loads with grid and microgrid connection using Victron energy system.

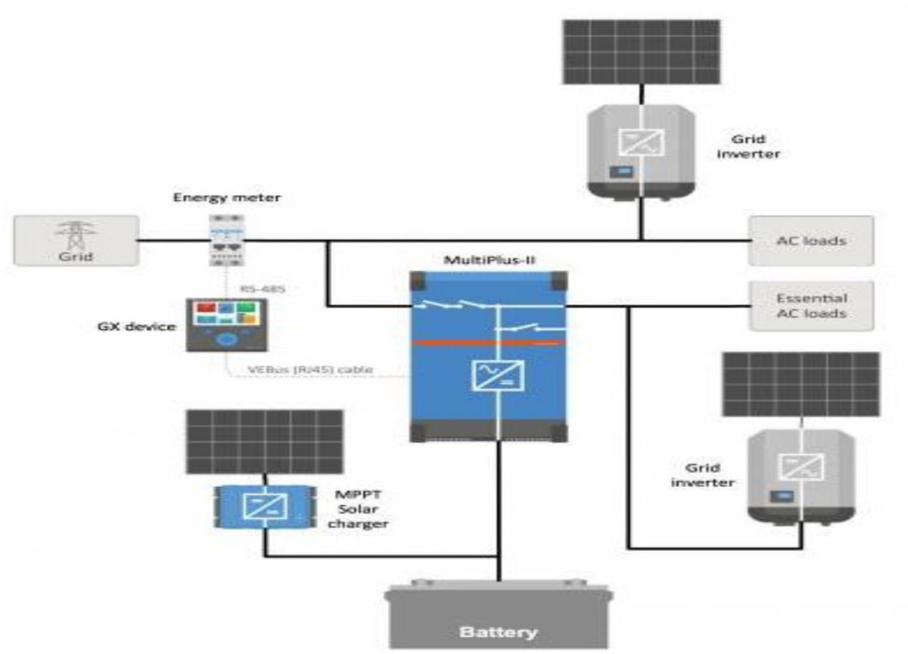

Fig. 17. Loads connected with Grid and Microgrid using Victron energy systems.

The solar PV is mainly responsible for providing energy to the primary load first and if there is an extra generation, the solar PV will charge the battery system. After that, the extra energy is provided to the secondary load and the remaining will be exported to the main grid.

Energy consumption data is measured at two different points. First energy meter is installed where the building has a connection with a main grid that has two measurements, Real energy into the load (the amount of energy consumed by both loads from the main grid) and it also stores the data about real energy out of the load (the amount of energy that is being exported to the grid from the solar PV).

The second energy meter is installed at the point where the microgrid connects with the secondary load and the main grid. The measurement includes grid to battery, grid to consumers (primary load), PV to grid (secondary load & main grid), PV to battery, PV to consumers (primary load), battery to grid and battery to consumers (primary load & secondary load).

Exploring the energy consumption data, seeing trends, and making inferences to increase energy efficiency are all part of data analysis for building energy usage. Having used a variety of data analysis methods, including descriptive statistics, visualizations, and correlation analysis.

- Firstly, we will analyze the data collected by first energy meter connected with main grid.

4.1. Distribution of datasets:

In order to spot patterns and get insights into the energy usage patterns, this part gives a visual depiction of the dataset's distribution of energy consumption values. The histogram plot displays the dataset's 'Value' column's frequency distribution. We can better evaluate the histogram's form and spot any patterns or trends in the data by comprehending the data distribution.

- This graph shows the two different datasets contained by data file. From 11000 kWh to 15000 kWh shows real energy out of the load which indicates the total energy that is exported to main grid by microgrid. From 95000 kWh from 101000 kWh shows the Real energy into the load which shows the total consumption of both loads from main grid.

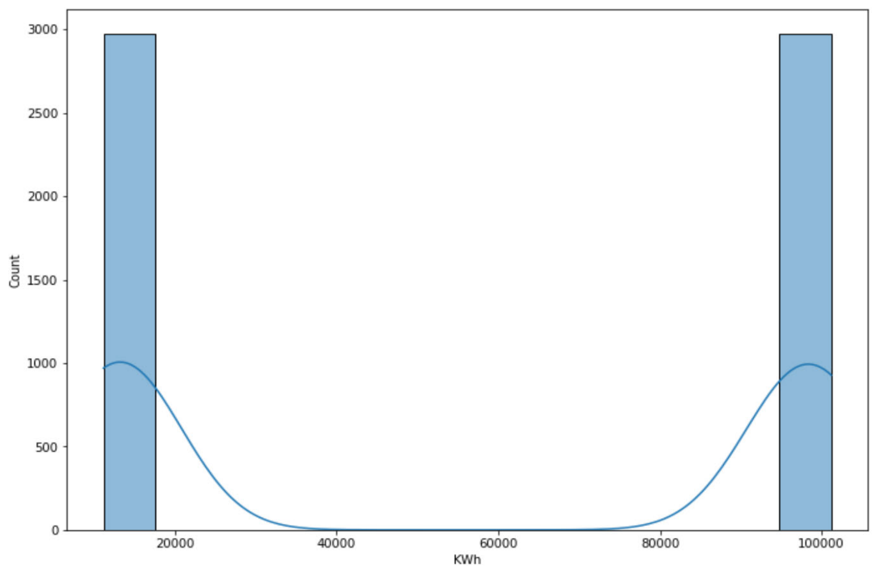

Fig. 18. Distribution of datasets.

4.1.1. Real energy into the load:

- Real energy into the load represents the total energy usage coming from the grid by both loads. The x-axis represents the time duration, and the y-axis represents the kilowatt-hour. The graph shows the usage of energy patterns over time, enabling a better understanding and management of energy consumption in buildings. This graph displays the behavior of energy consumption by loads over the indicated time frame. It demonstrates that loads use a maximum energy of 10.8 kWh and a minimum energy of 0 kWh.

```
Average energy consumption: 1.865090308184574
Maximum energy consumption: 10.890625
Minimum energy consumption: 0.0
```

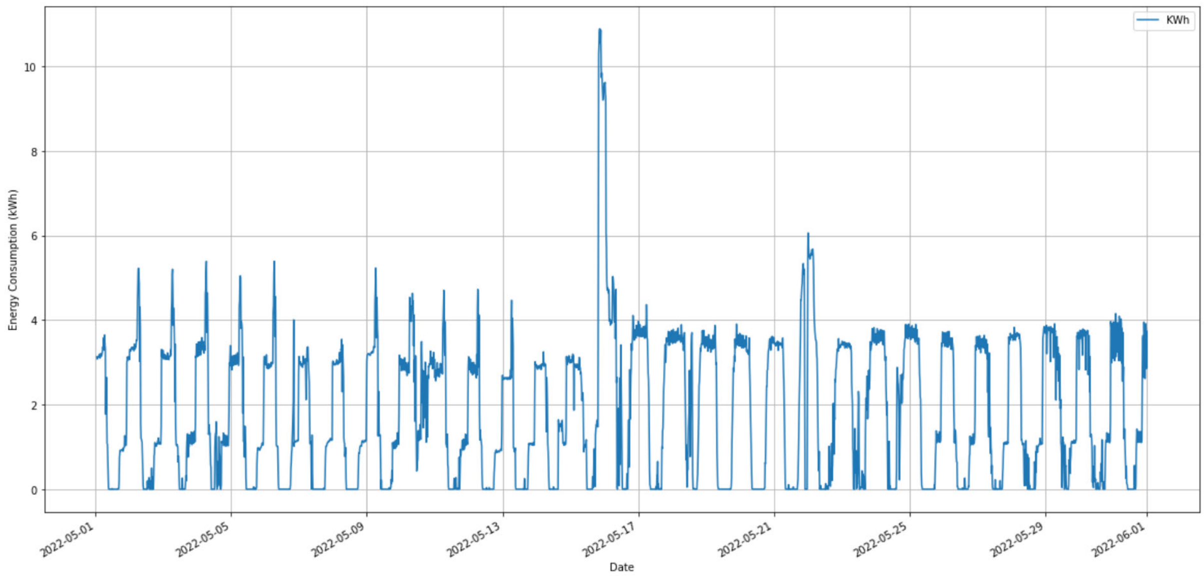

Fig. 19. Energy consumption over time into the load.

- Fig. 20 shows the y-axis in the histogram representing the count or density of observations in each bin, and the x-axis represents the range of values in the 'Value' column. The histogram displays the frequency distribution of those values. When kde=True is used, a kernel density estimate line—a smoothed curve that calculates the density of the underlying distribution—is added to the graphic. The graph shows most of the values lies between 0 kWh to 4 kWh.

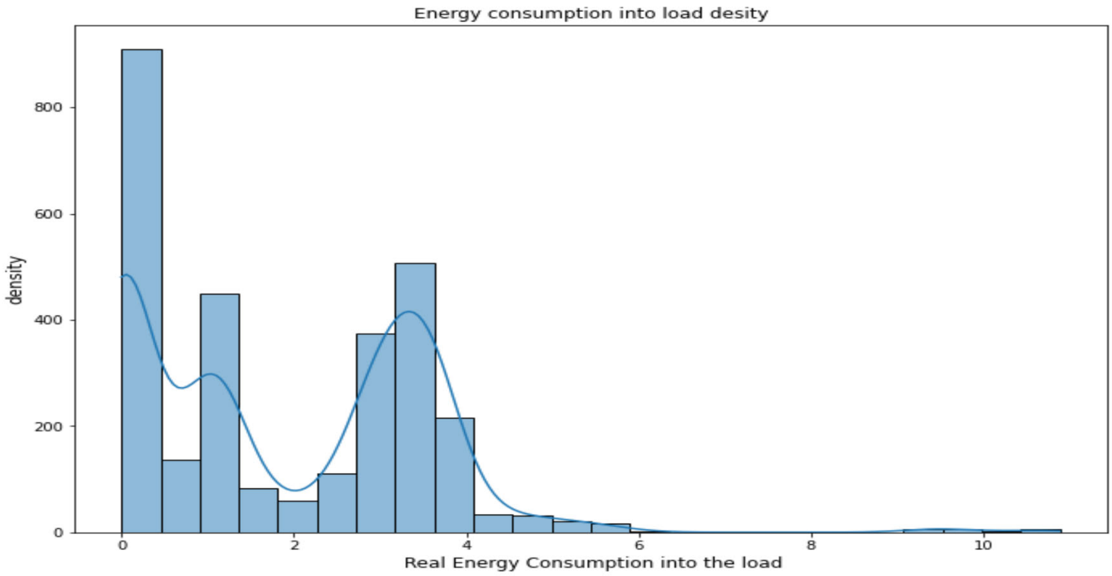

Fig. 20. Energy consumption density into the load in KWh.

- Fig. 21 shows the behavior of energy consumption according to every hour in a single day. We can see the fluctuations of energy demand between day and night mode. At night, most of the energy is used. The need for energy rises gradually in the evening and keeps rising throughout the night. After midnight to 6am in the morning, there is a spike in energy demand, followed by a sharp decline. Very little energy, often even none, is utilized by loads from the main grid after 10 a.m.

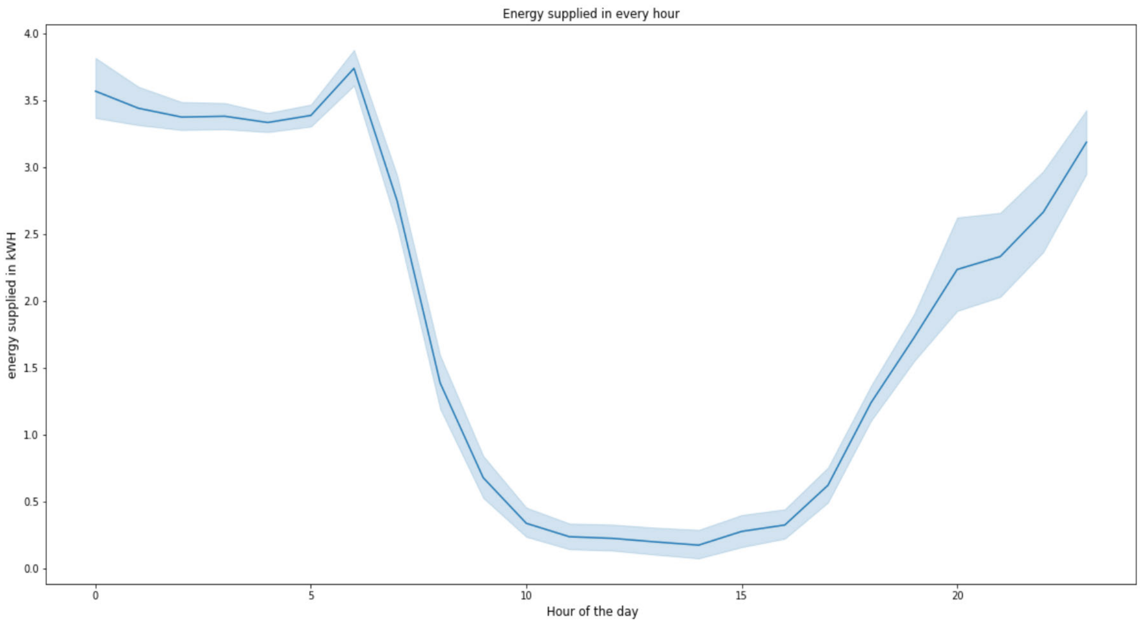

Fig. 21. Hourly energy consumption behavior for whole month.

- Fig. 22 depicts the overall quantity of energy consumed each day for the entire month. Energy fluctuates every day, but from days 9th May to 11th May a high of 250 KWh was computed, and from days 14th to 17th the greatest use of 294 KWh was recorded.

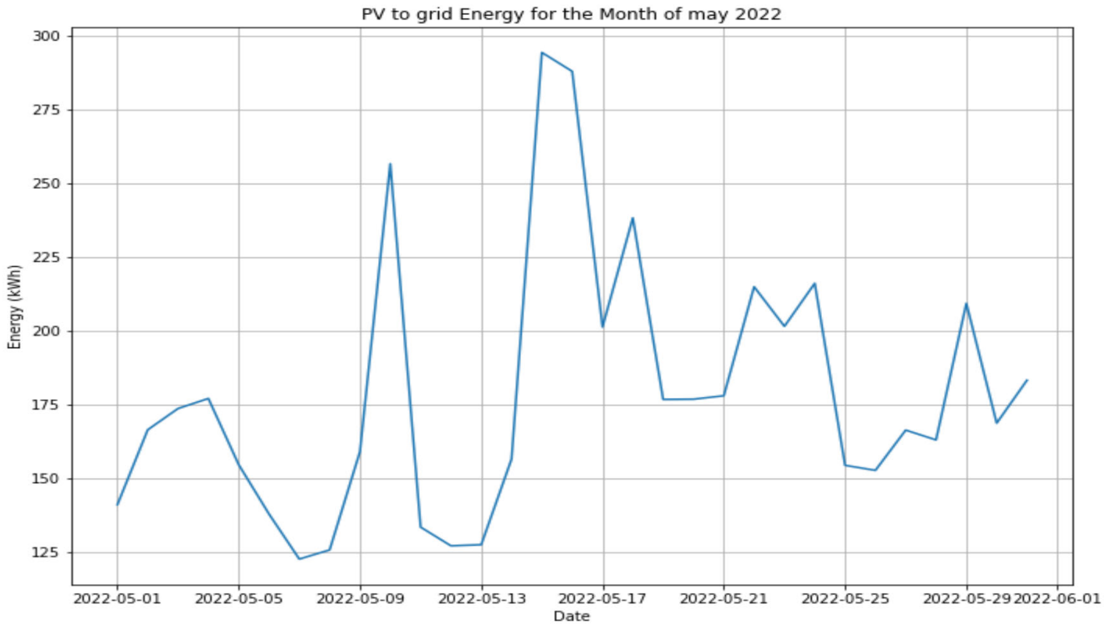

Fig. 22. Sum of energy consumed every day in whole month into the load.

- The maximum energy consumption is depicted in red in Fig. 23, while the minimal energy demand by loads is shown by the blue color line. The graph indicates that days 15 and 16 have the highest energy demand of more than 10KWh at night.

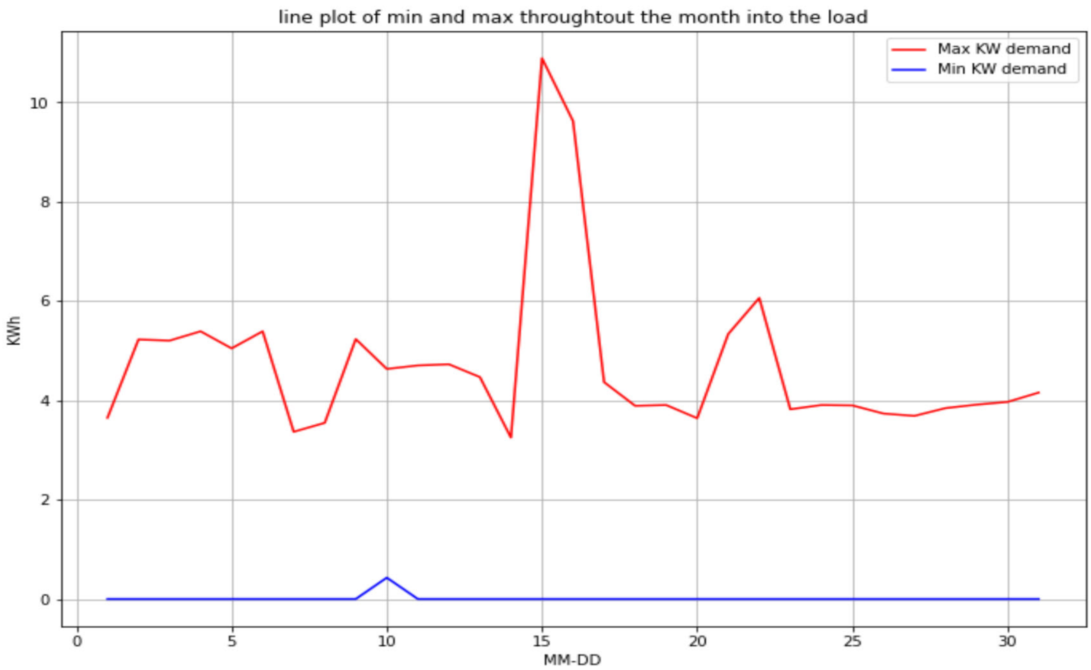

Fig. 23. Highest and lowest demand of energy.

4.1.2. Real energy out of the load:

- Real energy out of the load is the amount of energy being sent out to the grid. Kilowatt-hours are represented on the y-axis while time duration is represented on the x-axis. The findings show that while there is an excess of energy exported to the grid during the day, there is no export of energy during the night. Fig. 24 displays the behavior of energy exported by solar PV to the main grid over the indicated time frame. It demonstrates that solar PV exported maximum energy of 10.1 KWh and minimum energy of 0 kWh.

```
Average energy consumption: 1.3132588176785114
Maximum energy consumption: 10.181640625
Minimum energy consumption: 0.0
```

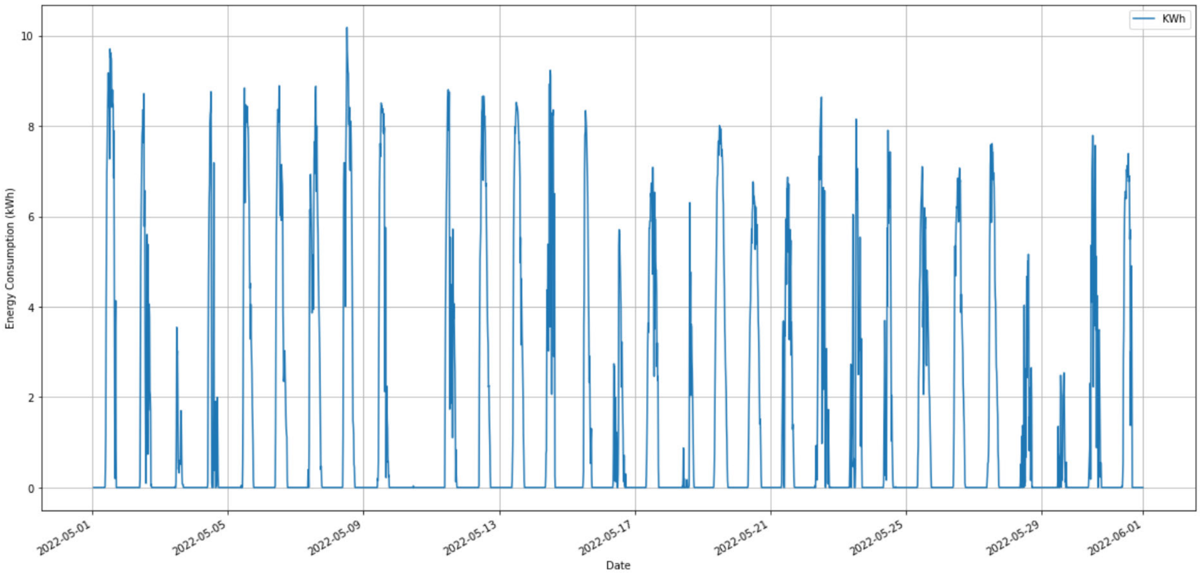

Fig. 24. Energy consumption over time out of the load.

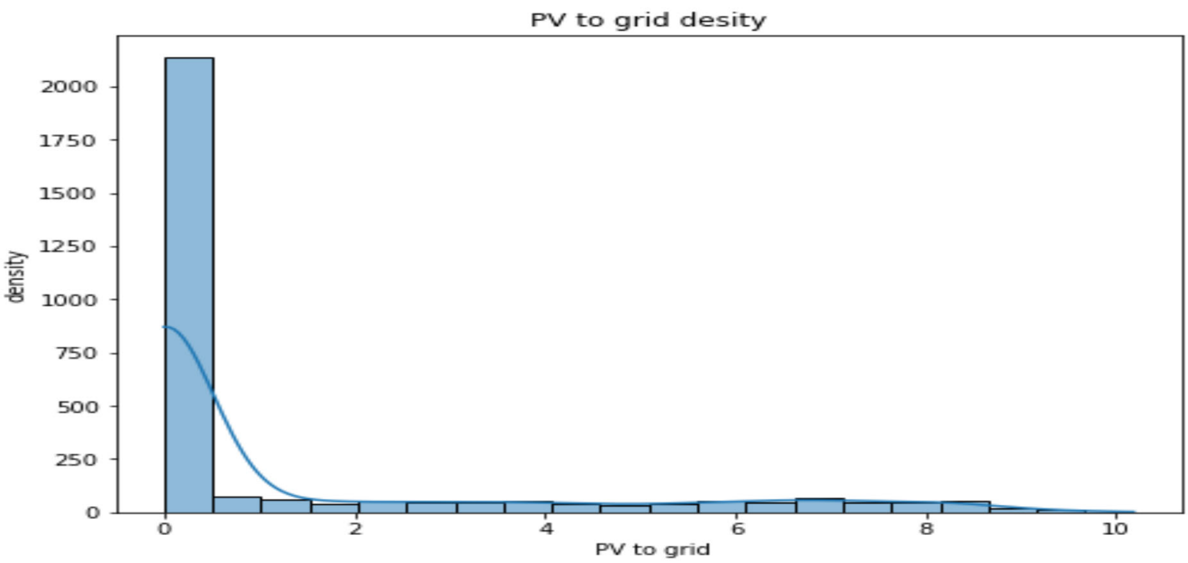

Fig. 25. Energy consumption density out of the load in KWh.

- Fig. 25 shows the y-axis in the histogram represents the count or density of observations in each bin, and the x-axis represents the range of values in the 'Value' column. The histogram displays the frequency distribution of those values. The graph shows most of the values lies between 0 KWh to 1 KWh.

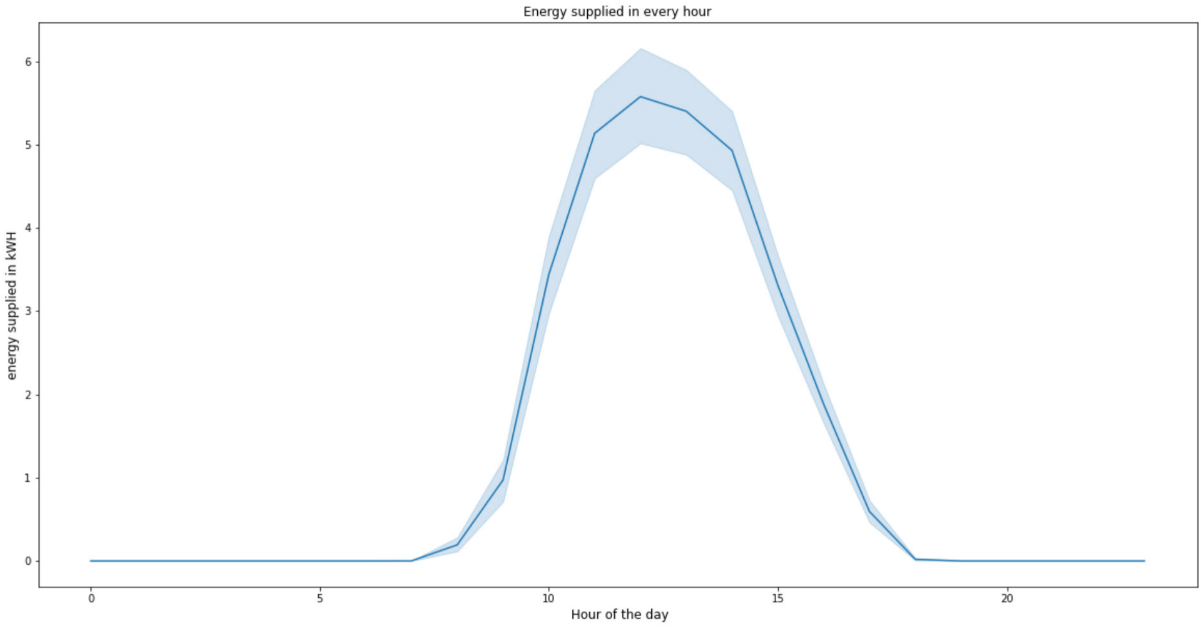

Fig. 26. Hourly energy consumption behavior for whole month.

- Fig. 26 depicts how energy usage changes throughout the course of a single day according to each hour. The variations in energy usage between daytime and nighttime operations are visible. There is no energy exported to the main grid at night. Energy export to the main grid increases significantly after 6 am. Maximum energy export occurs between 10 am and 15 pm, but after 18 pm there is no energy export to the grid.

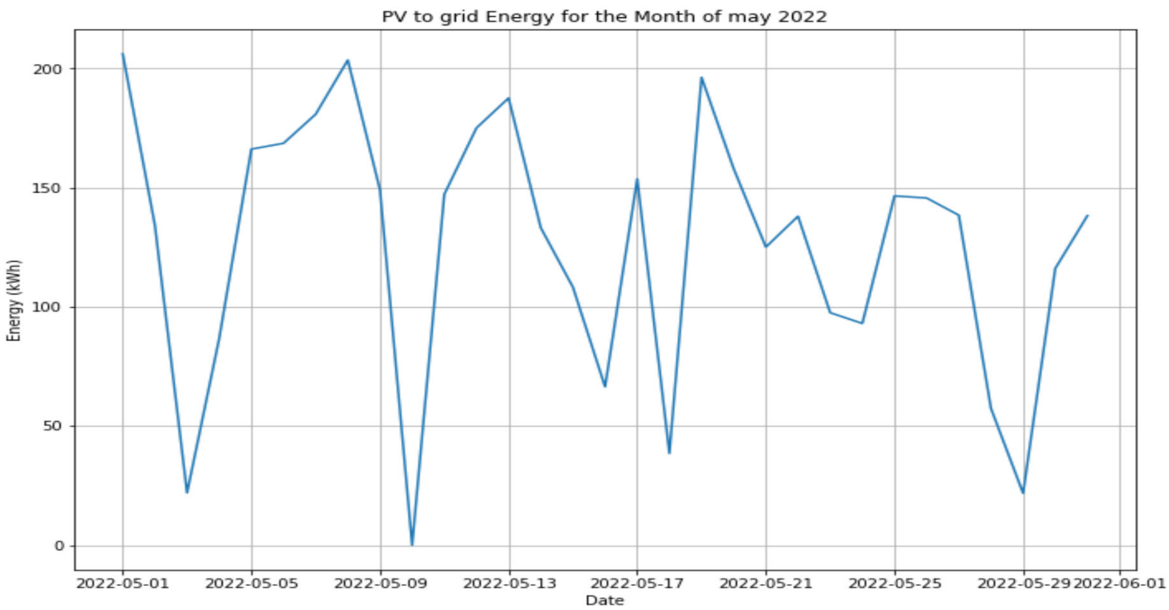

Fig. 27. Sum of energy consumed every day in whole month out of load.

- The total daily energy consumption for the entire month is shown in Fig. 27. Every day, energy levels fluctuate, however on May 10th, only 0.036 KWh were exported to the main grid, compared to 206 KWh on May 1st.

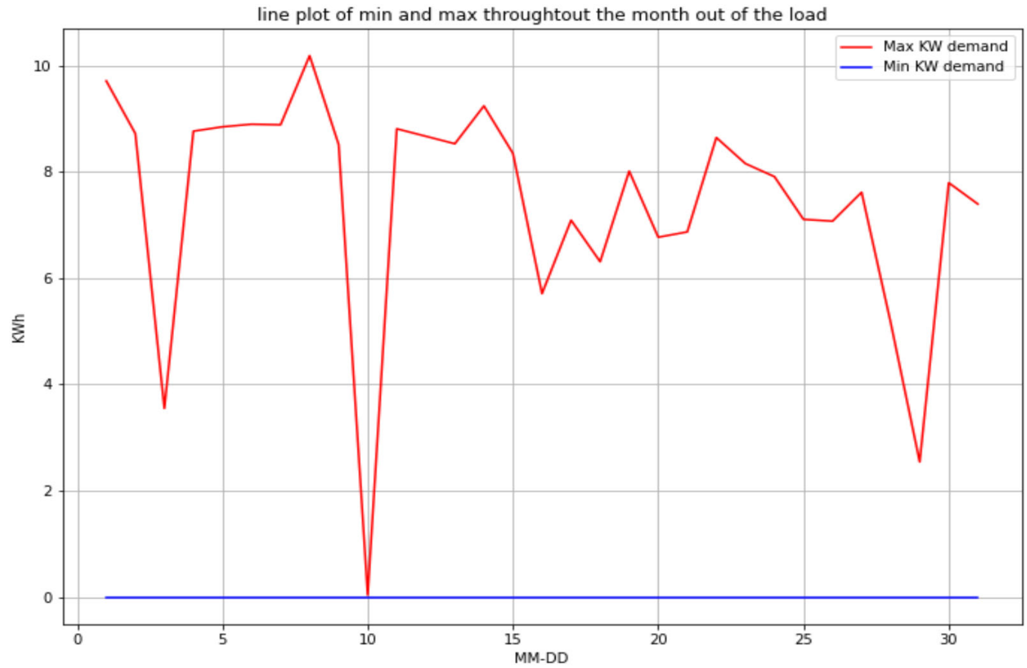

Fig. 28. Highest and lowest demand of energy.

- The maximum energy consumption is depicted in red in Fig. 28, while the minimal energy demand by loads is shown by the blue color line. The graph indicates that days 7th to 9th and the day 1st have the highest energy export to main grid of around 10kwh however, the lowest energy export is on 10th May.

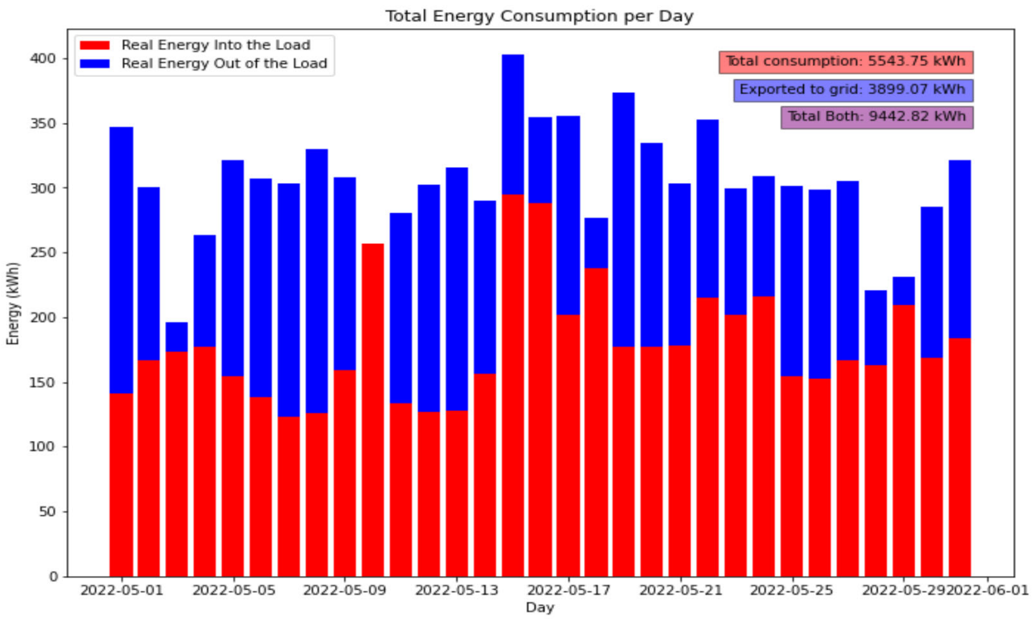

Fig. 29. Overview of grid to consumers and PV to grid.

- Fig. 29 shows the insights of the dataset measured by the first energy meter. The blue color bar displays the amount of energy supplied by solar to the grid, which is 3899 kWh, while the red color bar depicts the total energy consumed by loads, which is 5543.75 kWh.

4.2. Comparison of Real energy into the and Real energy out of the load:

The total amount of energy used by loads and the total amount of energy provided to the grid by solar PV are compared. The hourly behavior of both energy going into the load and coming out of the load will be evaluated and the results for the energy consumption by loads and energy supplied to the grid for the entire month will be compared.

- Figure 30 depicts the pattern of energy use over time in red and solar PV energy export to the main grid in green. However, Figure 31 illustrates how energy is used from the main grid at night and how solar PV exports electricity to the grid during the day. It demonstrates that there is no solar PV energy production at night, hence there is no energy export, but there is an excess of energy exported to the main grid during the day. On the other hand, the main grid supplies energy at night when PV generation is not there, and relatively little energy is used during the day because loads are receiving power from PV.

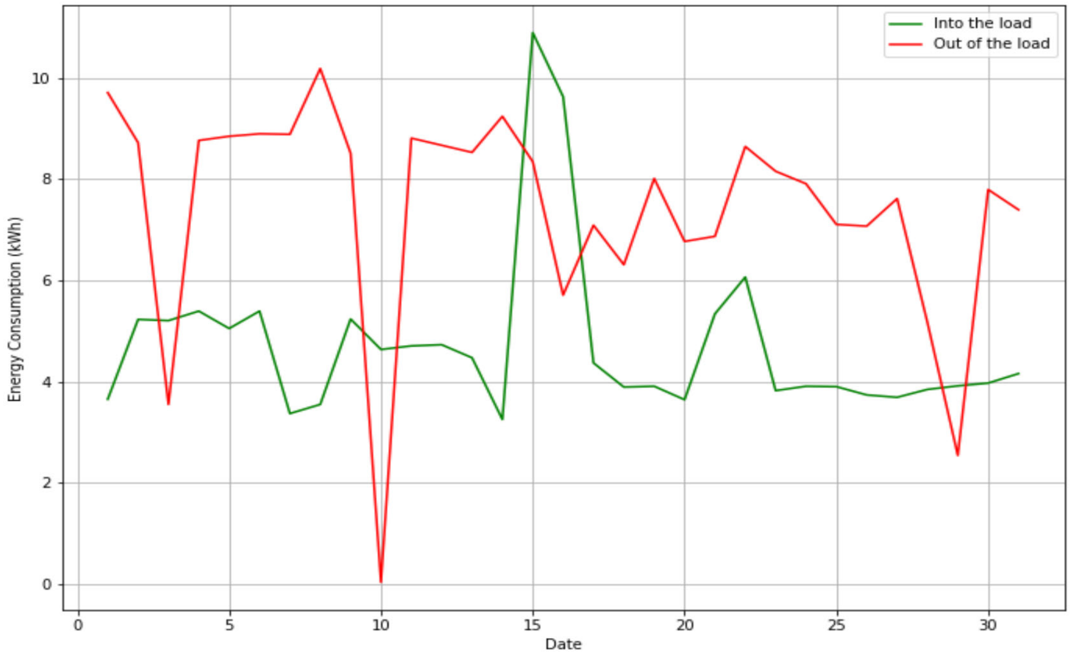

Fig. 30. Energy consumption and energy export over time.

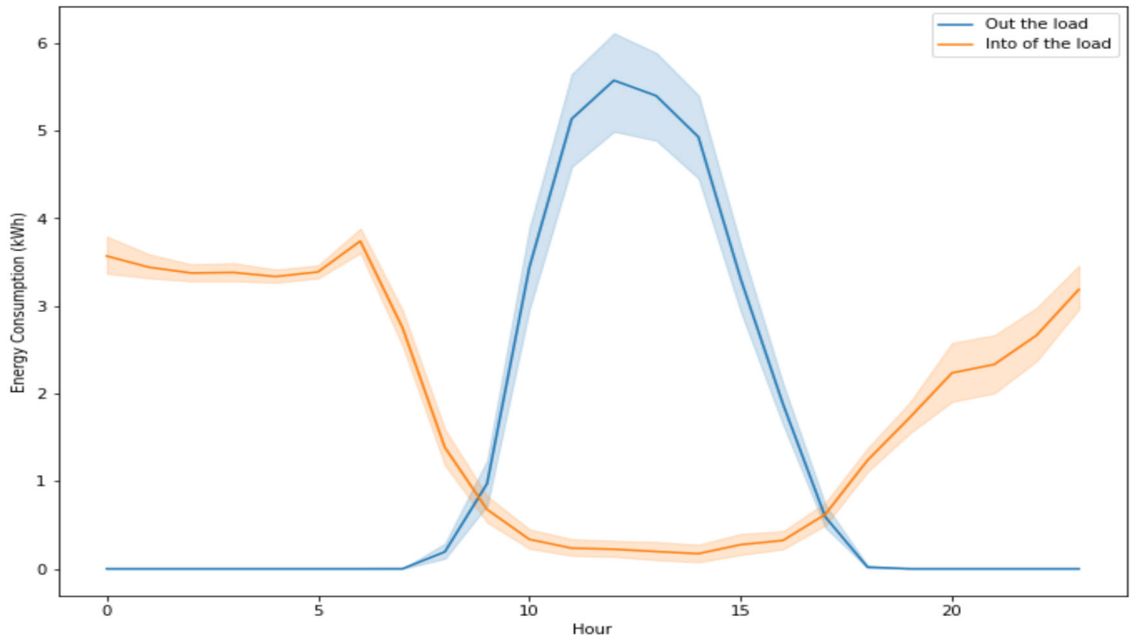

Fig. 31. Hourly behavior of energy consumption and energy export.

4.3. Energy efficiency analysis:

Energy efficiency analysis entails gathering and examining information on energy usage, detecting energy inefficiencies, and formulating plans to cut down on energy waste and increase efficiency. Energy efficiency analysis identifies the day of the week where loads consume the most energy, as well as the hours of the day where consumption is the highest and lowest.

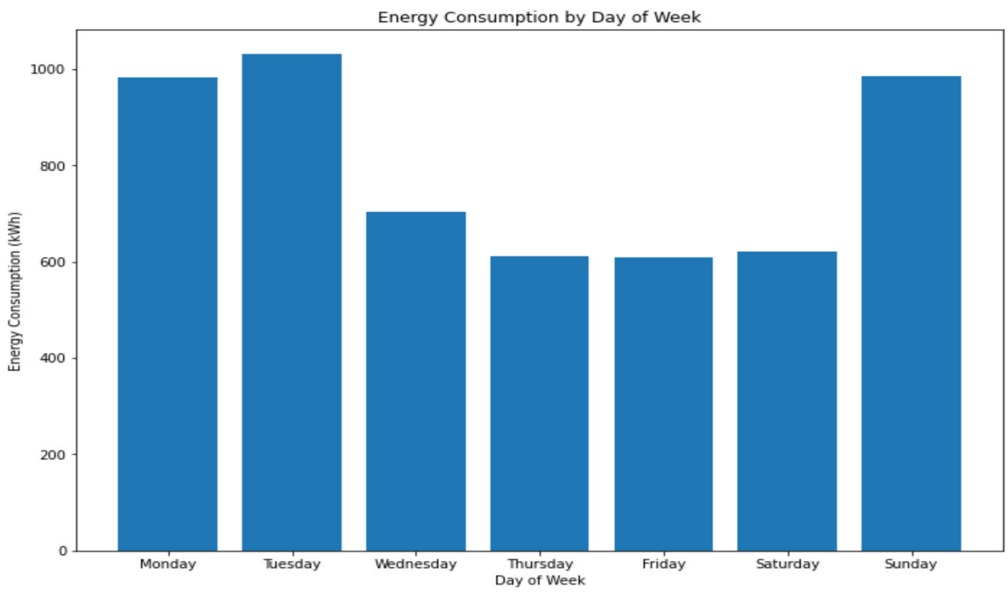

Fig. 32. Energy consumption by day of week into the load.

Fig. 32 calculated the total amount of energy consumed by each day of week. It shows that the Sunday, Monday, and Tuesday are highest energy consumption days.

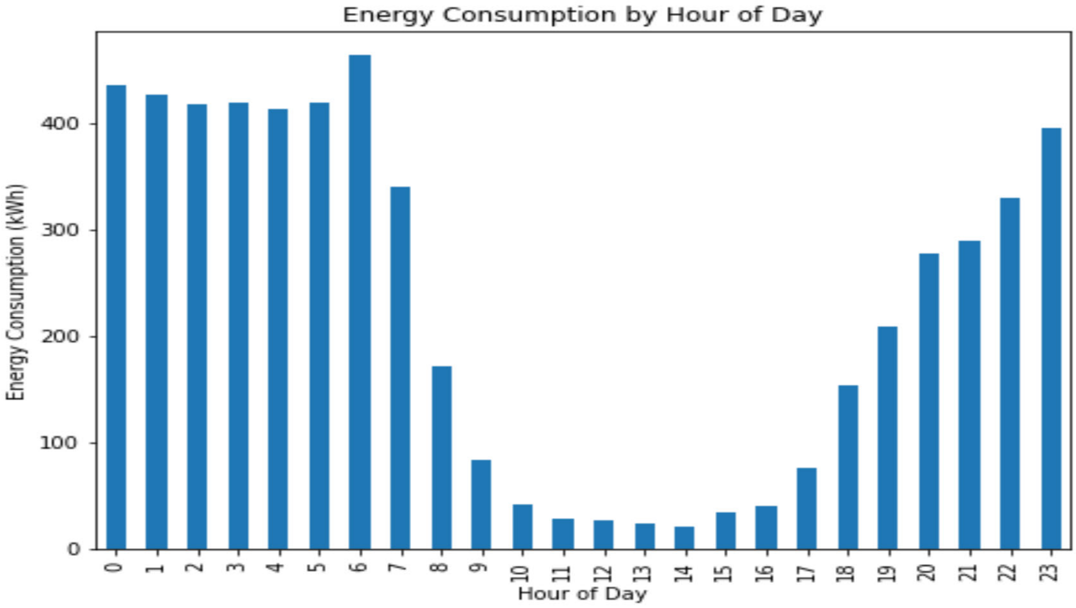

Fig. 33. Energy consumption by hour of the day into the load.

In Fig. 33 the amount of consumption is calculated according to each hour of the day. It demonstrates that consumption is higher at night and substantially lower from 9 a.m. to 17 p.m. The peak energy usage occurs around 6 a.m. during the hours of 20 p.m. to 7 a.m.

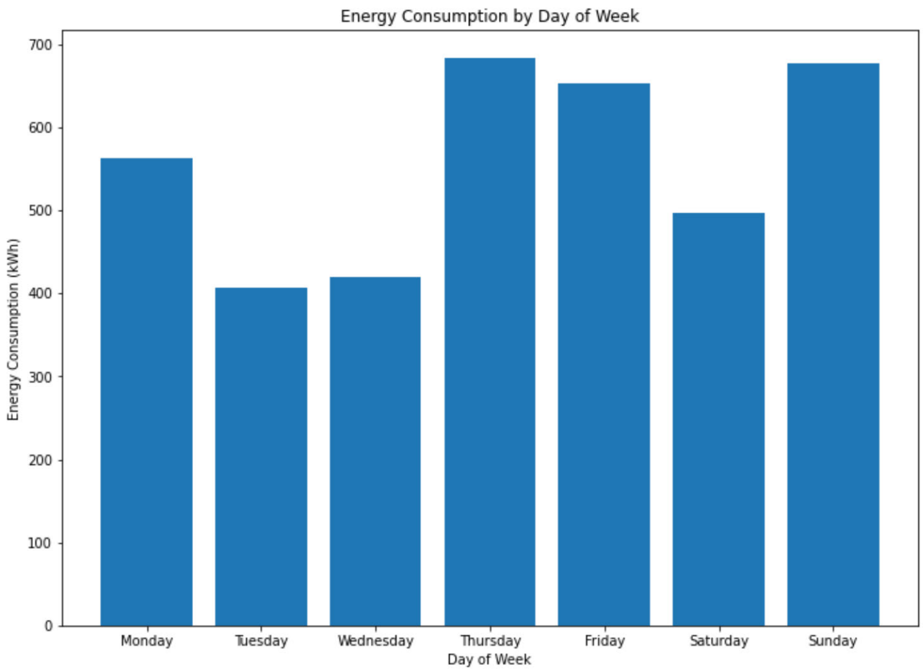

Fig. 34. Energy consumption by day of week out of the load.

Fig. 34 calculated the total amount of energy consumed by each day of week. It shows that the maximum energy is exported on Sunday, Thursday, and Friday.

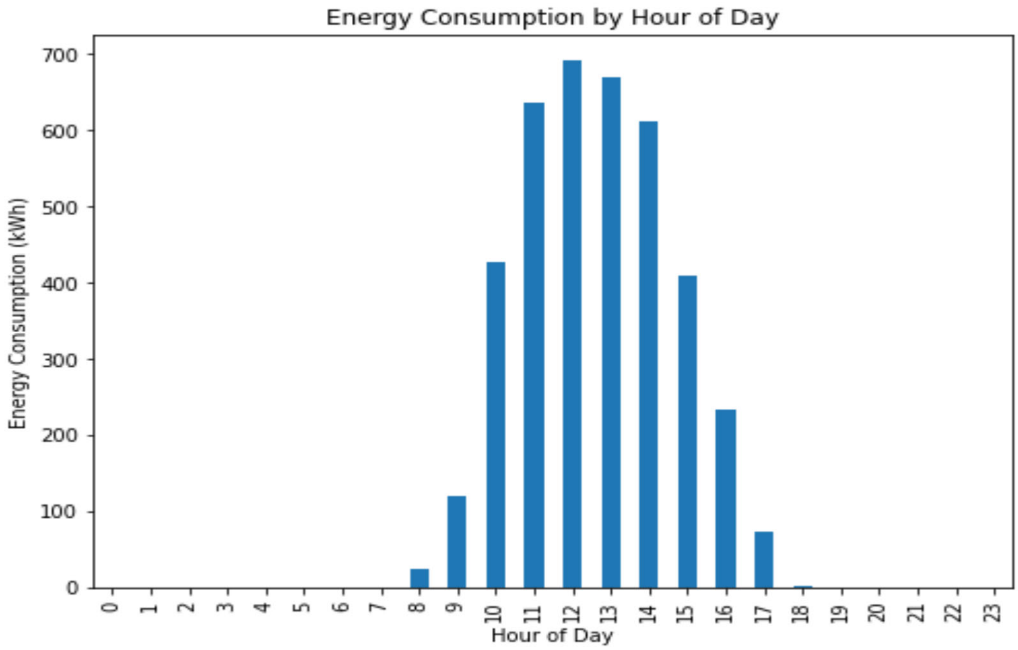

Fig. 35. Energy consumption by hour of the day out of the load.

In Fig. 35, the amount of consumption is determined for each hour of the day. It indicates that the amount of energy exported from solar PV to the grid is at its highest between 10 am and 16 pm however, that there is no energy export between 17 pm and 8 am.

- Analysis of the data from the second energy meter connected with solar PV and secondary load.

4.4. Grid support and contribution.

The grid is providing energy to the primary load and charges the battery storage system as well. This section will explain the energy usage pattern, behavior, and amount of energy provided by the grid to the primary load and battery storage system.

4.4.1. Grid to consumers:

- This graph displays the behavior of energy used by primary load from main grid over the indicated time frame. It demonstrates that primary load consumed maximum energy of 4.83 KWh during night and a minimum energy of 0 KWh during day time.

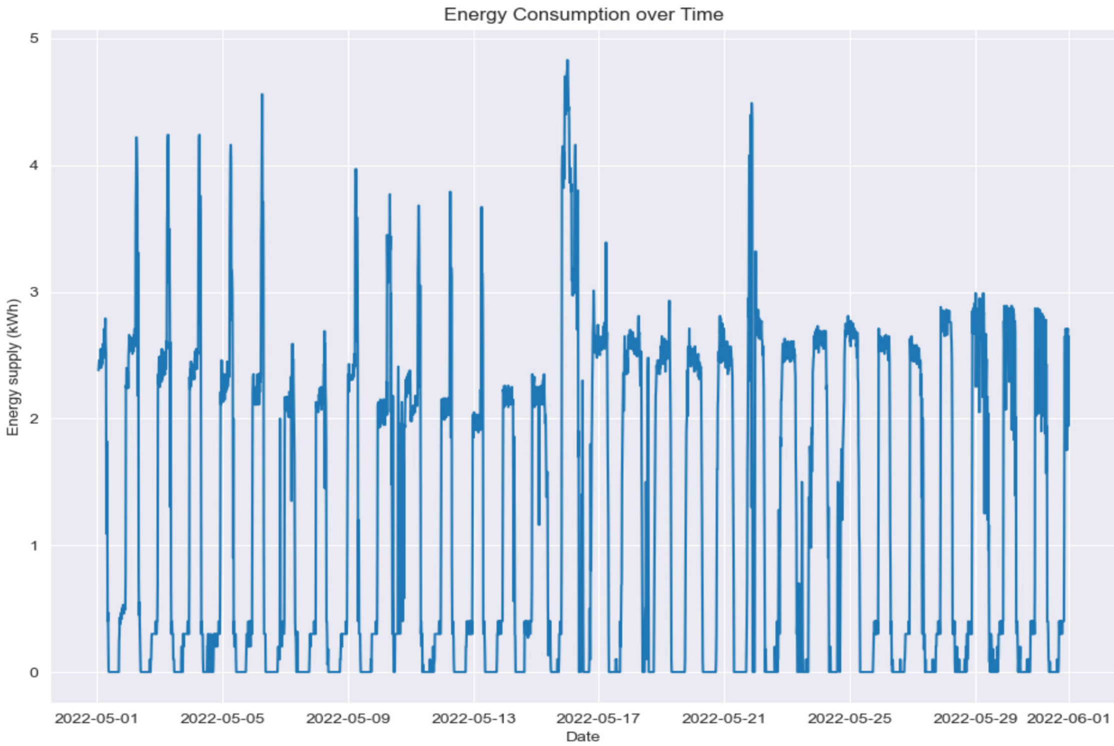

Fig. 36. Grid to consumers energy consumption.

- Fig. 37 shows the behavior of energy consumption according to every hour in a single day. There are fluctuations of energy demand between day and night mode. Most of the energy is used by primary load during night. The need for energy rises gradually in the evening and keeps rising throughout the night. After midnight to 6am in the morning, there is a spike in energy demand, followed by a sharp decline. Very little energy, often even none, is utilized by loads from the main grid after 10 a.m.

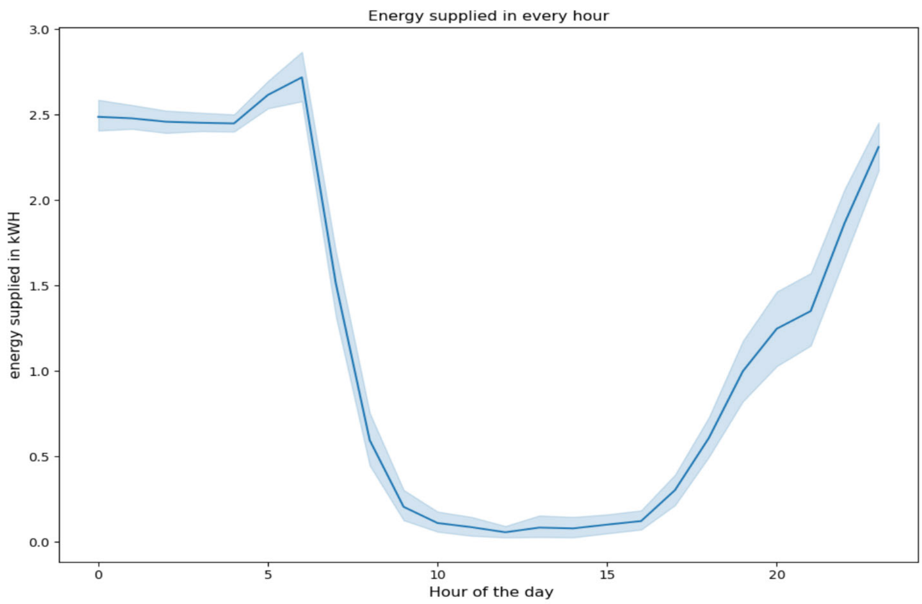

Fig. 37. Hourly energy consumption behavior for whole month Grid to consumers.

4.4.2. Grid to battery:

- The fig. 38 depicts the behavior of energy used by primary load from main grid over the indicated time frame. It demonstrates that the main grid charge battery during night which goes maximum of 2.2 kwh.

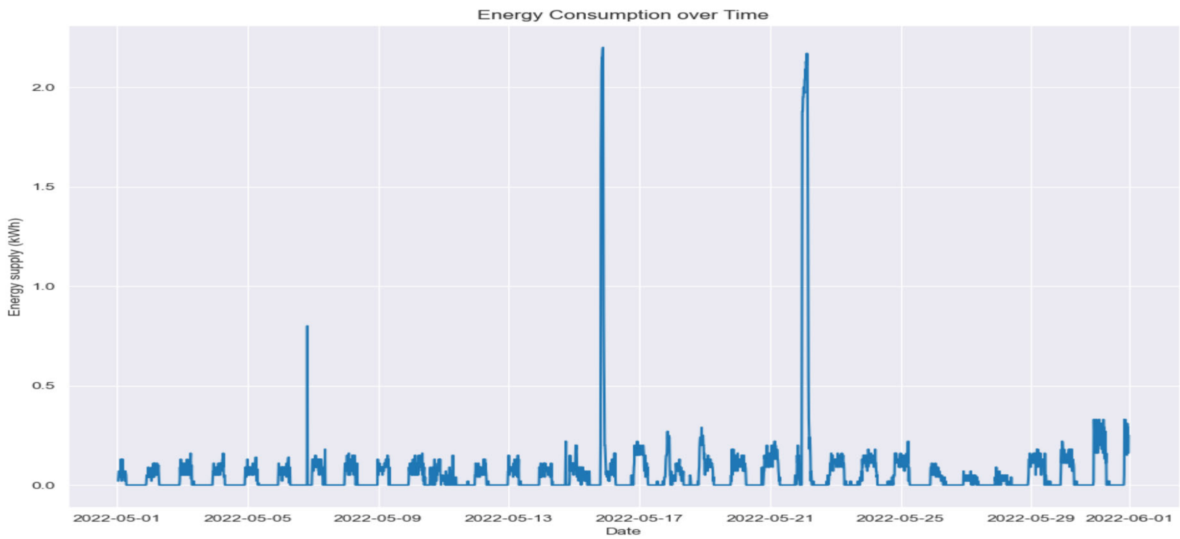

Fig. 38. Grid to battery energy consumption.

- Fig. 39 represented the behavior of energy transferred by grid to battery according to every hour in a single day. The amount of energy provided by grid to battery is very less and it only charge battery when the state of charge of battery goes less than 50%. During day battery get energy from solar and after 18pm until 6am grid is responsible to provide energy to battery.

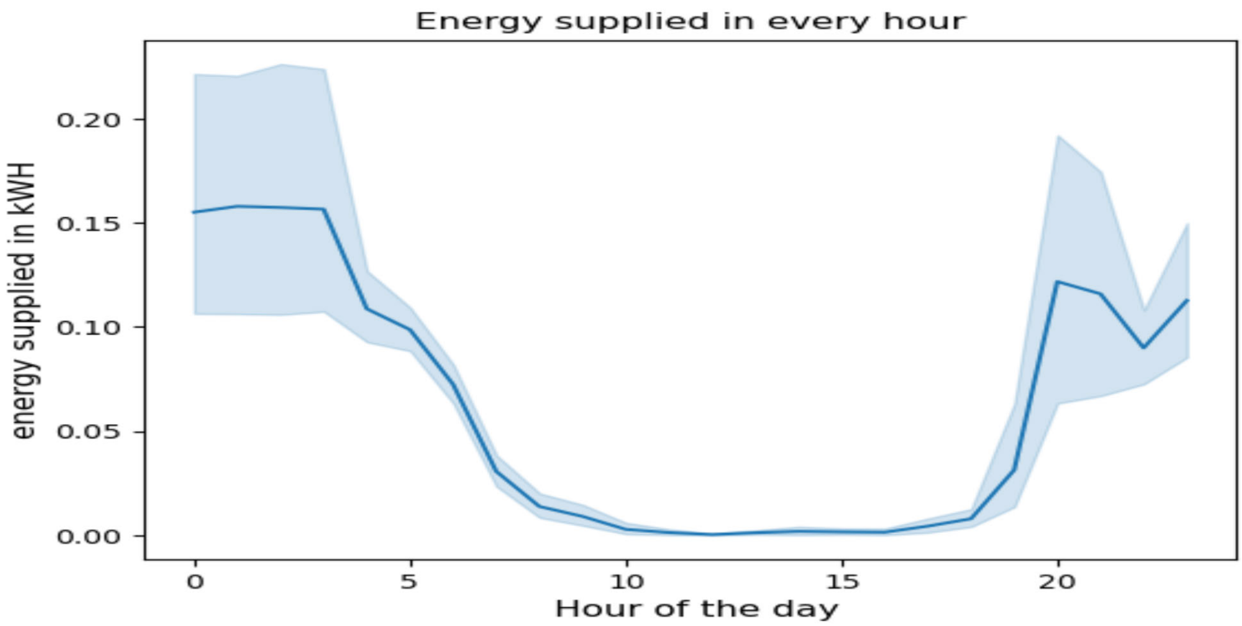

Fig. 39. Grid to battery operation mode Hourly energy consumption behavior for whole month Grid to battery.

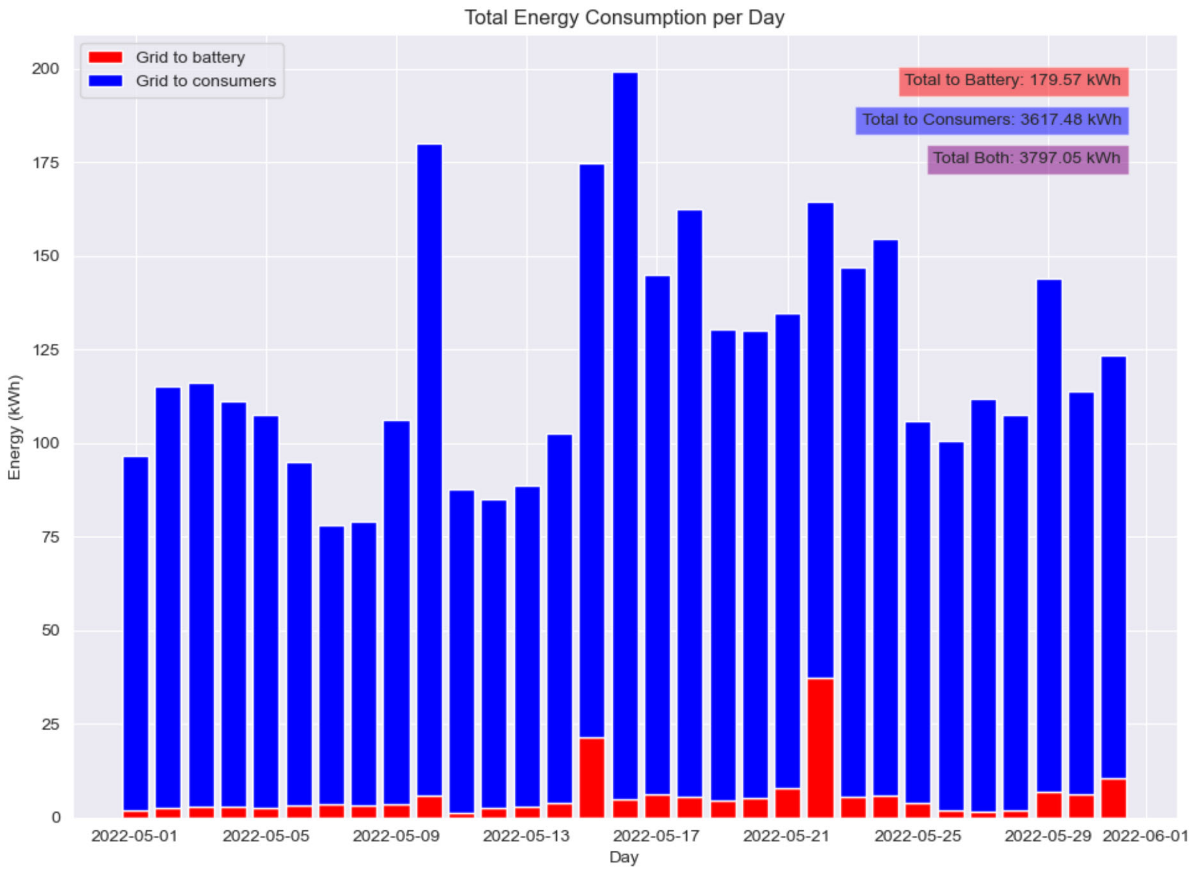

Fig. 40. Overview of grid to consumers and battery.

4.5. Solar PV Operation Modes:

Solar PV is providing energy to the primary load and charges the battery storage system as well. If there is excess amount of energy production, then the solar PV supply energy to secondary load and the extra energy is being exported to the grid. This section will explain the energy usage pattern, behavior, and amount of energy provided by the solar PV to the primary load, battery storage system and grid.

4.5.1. Solar PV to consumers:

- This graph expressed how solar PV energy used for primary load has changed over the specified time period. It shows that the primary load used a high of 5.57 KWh and a minimum of 0 KWh of energy. The highest peak demand has been noticed on 15th, 16th and 22nd of May.

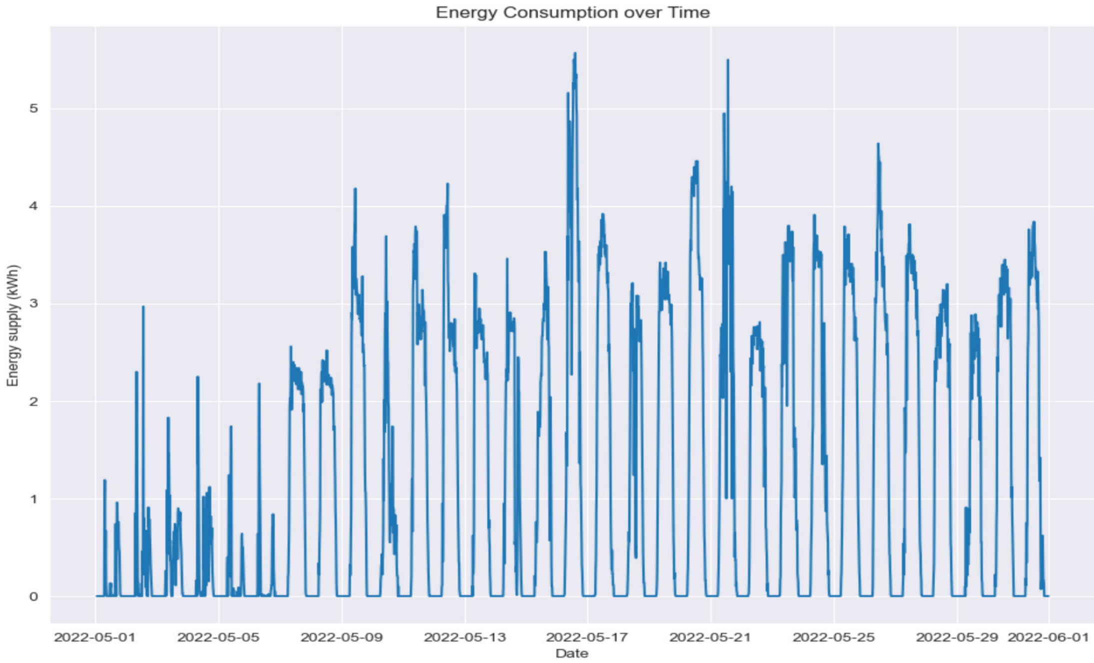

Fig. 41. PV to consumers energy consumption.

- Fig. 42 analyzed the behavior of energy supplied by PV to consumers according to every hour in a single day. There is no energy flow from PV to consumers from 20 p.m. to 5 a.m., but PV generates significant amounts of energy during the day and also supplies energy to primary and secondary loads.

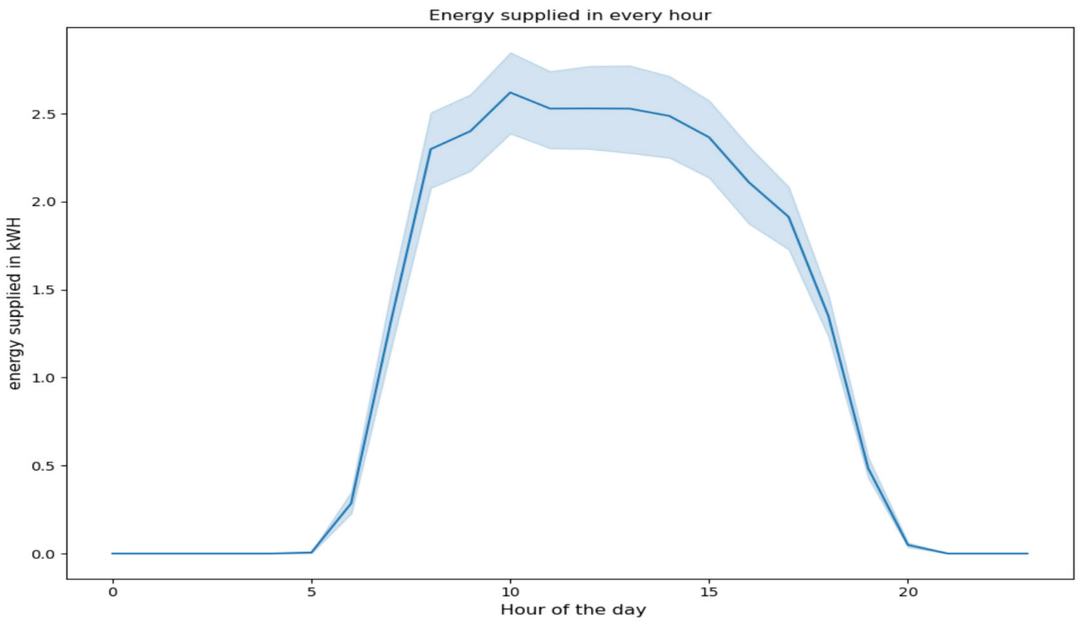

Fig. 42. Hourly energy consumption behavior for whole month PV to consumers.

4.5.2. Solar PV to grid:

- The behavior of energy exported from solar PV to the grid is described in Fig. 43. The data indicates that on May 10, relatively little energy was exported, but on May 8, the most energy was sent to the main grid and secondary loads.

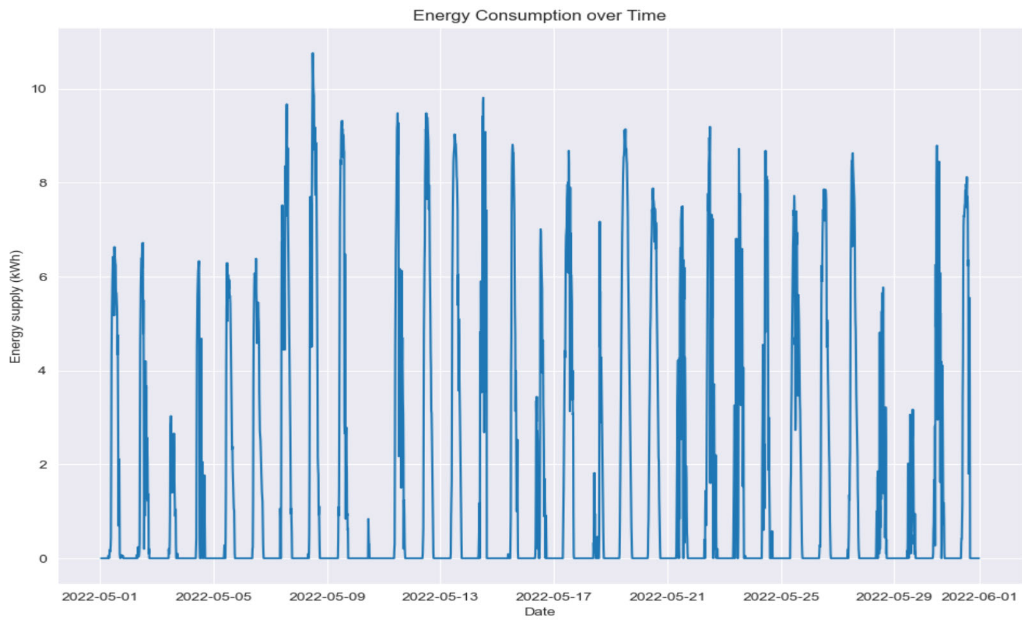

Fig. 43. Energy export from PV to grid.

- Fig. 44 shows the behavior of energy provided by PV to grid according to every hour in a single day. During night, there is no energy flow from PV to grid from 18 p.m. to 6 am, but PV generates significant amounts of energy during the day and also supplies energy to primary and secondary loads.

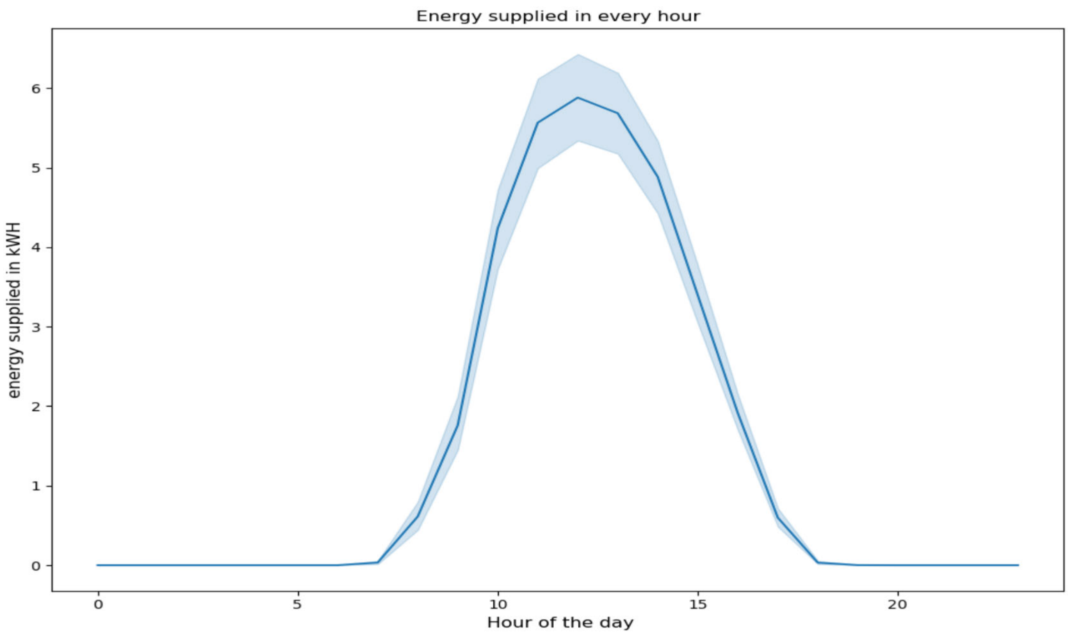

Fig. 44. Hourly energy consumption behavior for whole month PV to grid.

4.5.3. Solar PV to battery:

- The pattern of energy transmitted from solar PV to the grid is illustrated in Fig. 45. It demonstrates that solar PV was able to supply the battery with a maximum of 6.52 kwh of energy.

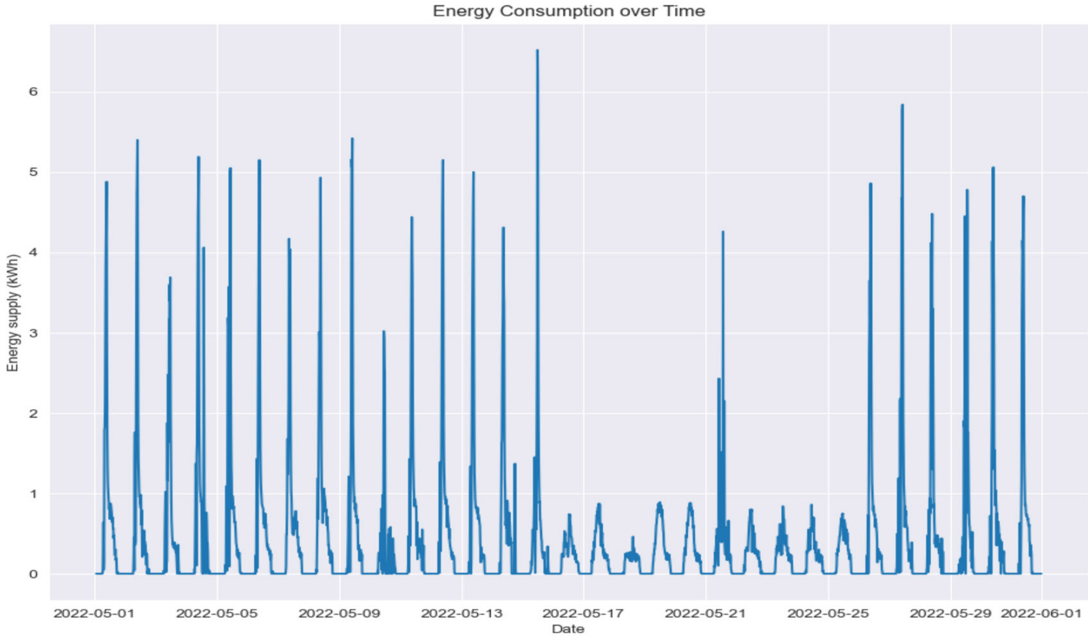

Fig. 45. Energy export from PV to battery.

- The fig. 46 shows that there is sharp increase of energy flow after 6am and reached at peak at 9am and provided maximum energy to battery from solar PV however, after 20pm until 5am there is no energy flow from PV to battery.

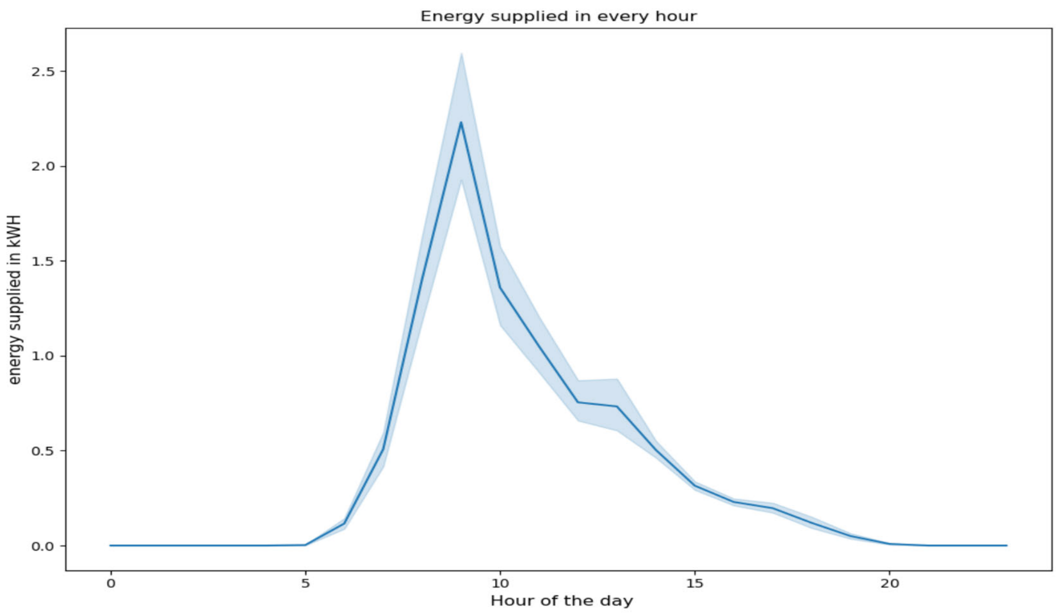

Fig. 46. Hourly energy consumption behavior for whole month PV to battery.

- Fig. 47 shows the insights of the dataset measured by the second energy meter. The blue color bar displays the amount of energy supplied by solar to the grid, which is 4290 kWh, while the red color bar depicts the total energy provided by solar PV to secondary load, which is 3362 kWh, and 1187 kWh of energy is being used by battery system from solar mentioned in the green color bar.

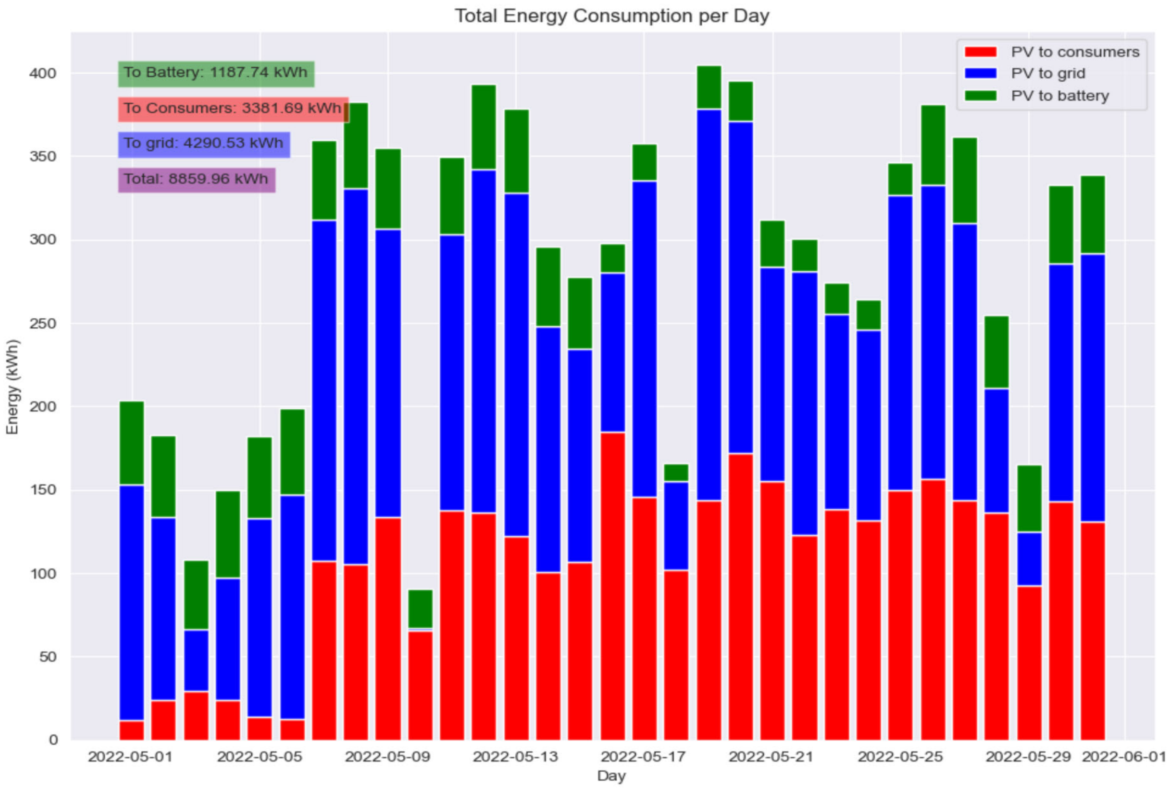

Fig. 47. Overview of PV to consumers, grid, and battery.

4.6. Battery contribution:

When demand is at its highest, battery storage systems help to power loads. When the batteries in the storage system are fully charged and the weather forecast for tomorrow calls for sunny conditions, the system begins to supply electricity.

4.6.1. Battery to consumers:

- Fig. 48 demonstrates the layout of energy transmission from battery to consumers. It reveals that the battery was capable of supplying the users with a maximum of 4.44 kwh of energy.

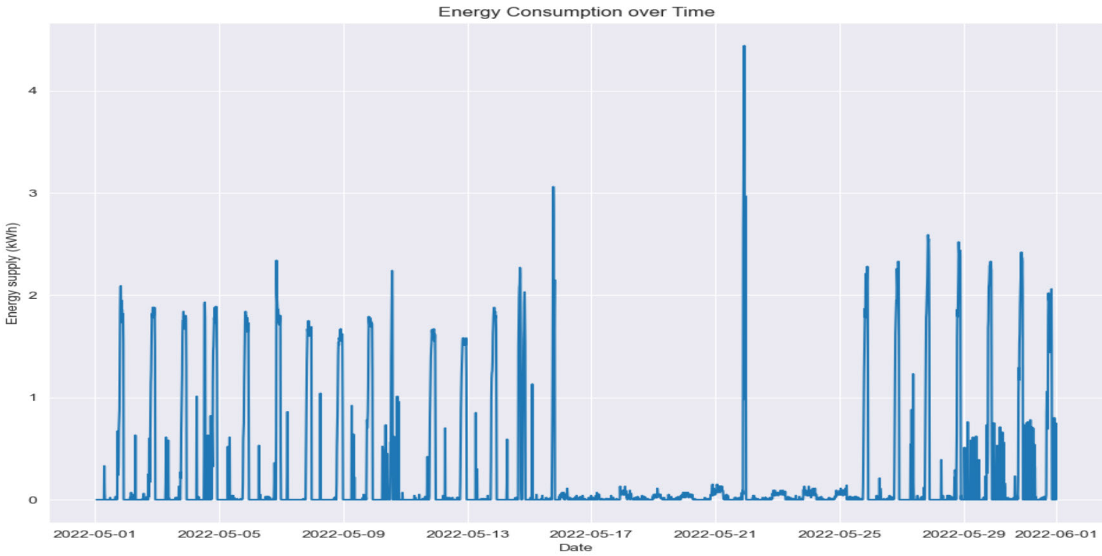

Fig. 48. Energy export from battery to consumers.

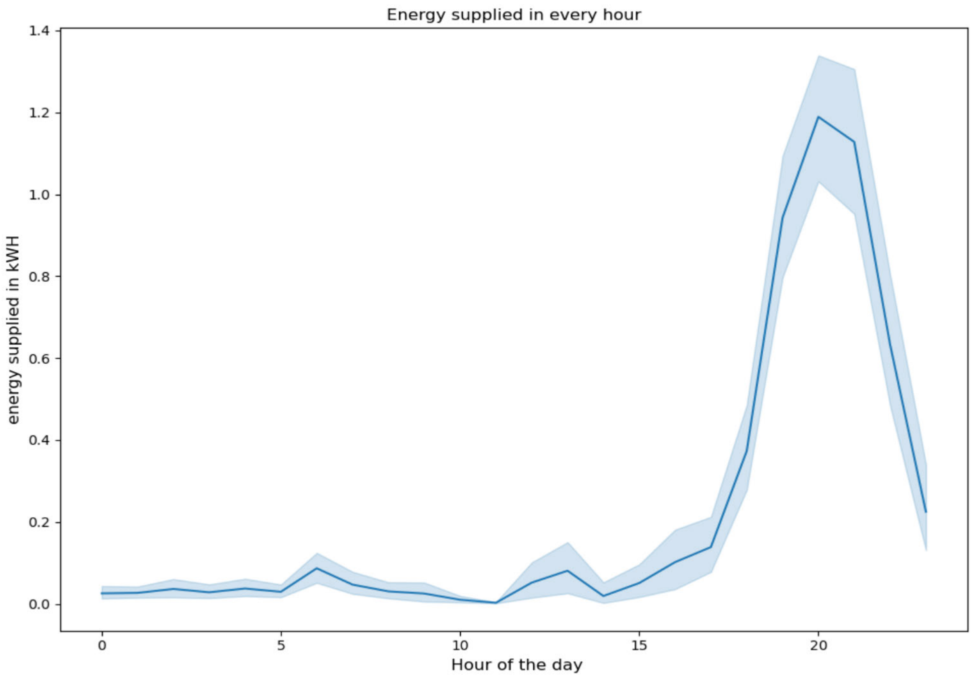

Fig. 49. Hourly energy consumption behavior for whole month battery to consumers.

- Figure 49 depicts the energy variations throughout the day. Consumers receive very little energy from the battery during the day, but after 18 o'clock the supply abruptly increases and reaches its peak at 20 o'clock.

4.6.2. Battery to grid:

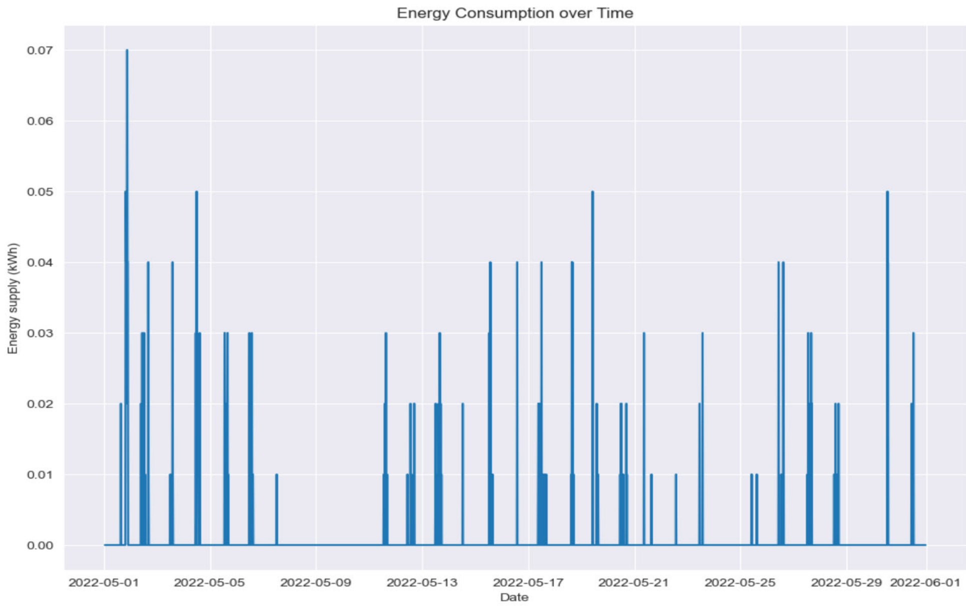

Fig. 50. Energy export from battery to consumers.

- Fig. 50 demonstrates the layout of energy transmission from battery to grid. It reveals that the battery was capable of supplying maximum 0.07 kwh of energy to the grid.

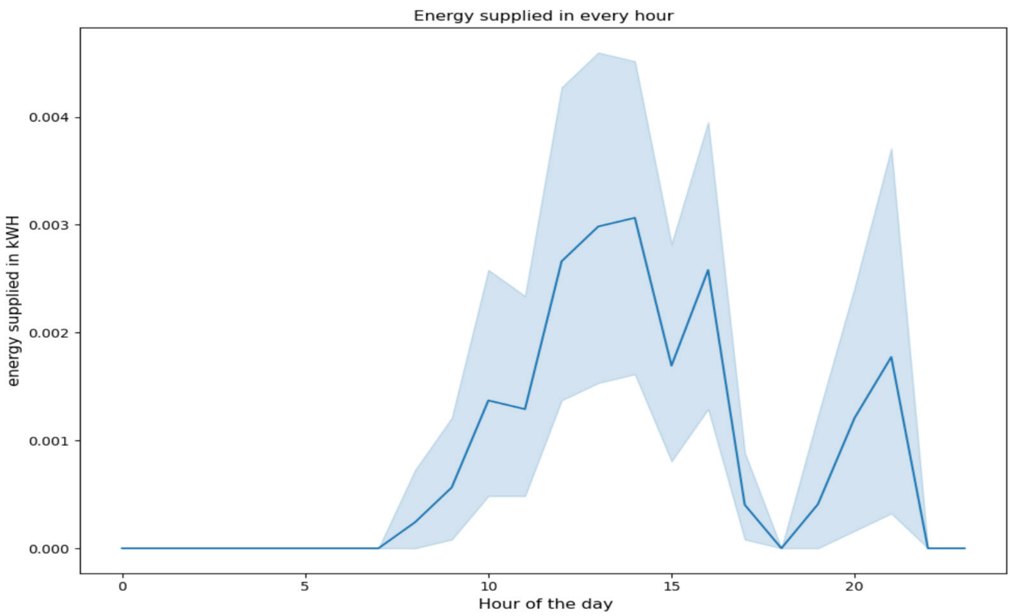

Fig. 51. Hourly energy consumption behavior for whole month battery to grid.

- Figure 51 depicts the energy variations throughout the day. Grid receives very little energy from the battery during the night after 23pm until 7am, but after 8 o'clock the supply abruptly increases and reaches its peak at 13pm. There is energy flow between 19pm to 23pm because of weather prediction. If predicted weather is sunny and more energy will produce by PV, then batteries are allowed to discharge at night as well.

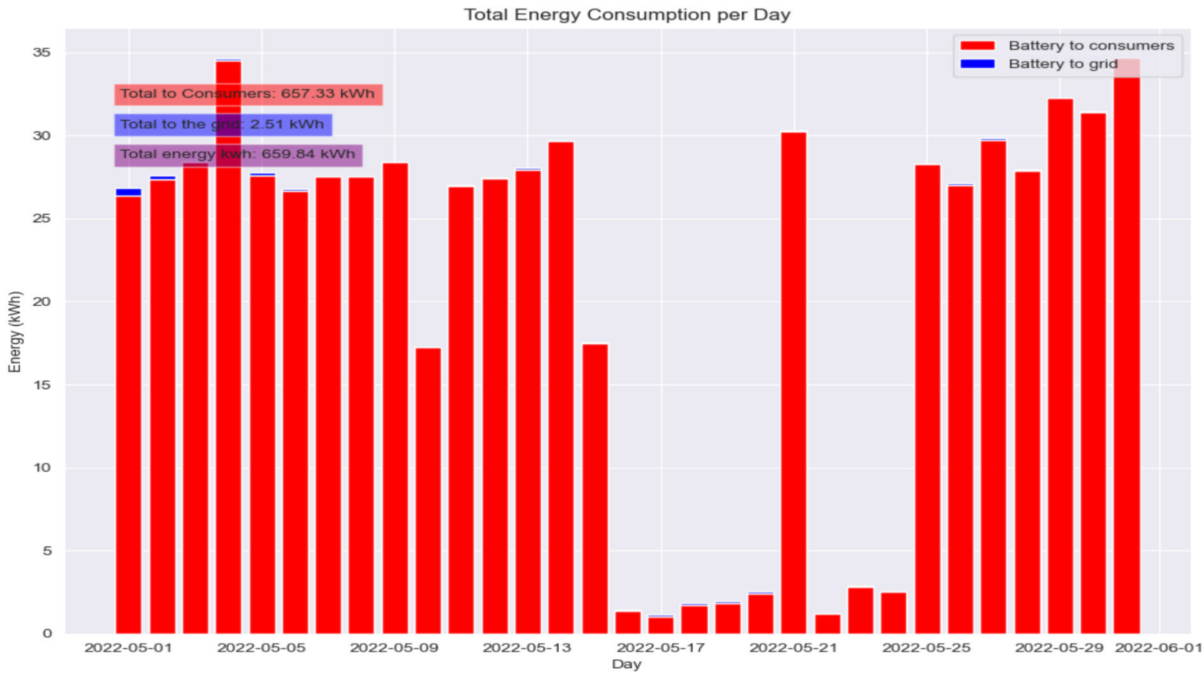

Fig. 52. Overview of battery to consumers and grid.

4.7. Building Consumption:

- Figure 53 illustrates the monthly total consumption as well as the amount of energy used by primary load from each of the three energy sources (PV, Grid, and Battery). It demonstrates that the grid, which has a capacity of 3617 kwh, provided most of the energy used by loads, but solar PV has also contributed significantly to loads with a capacity of 3381 kwh. The consumers received 657 kwh of energy from the battery.

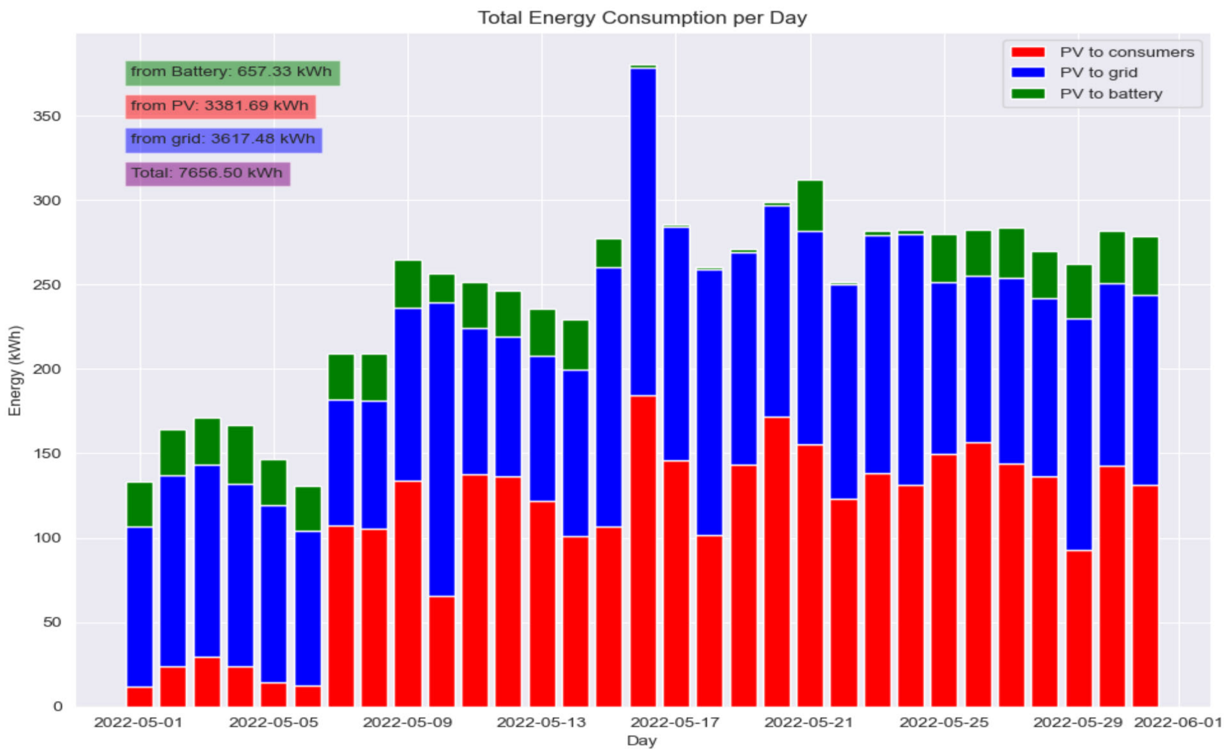

Fig. 53. Overall consumption from solar PV, Grid and Battery.

4.8. Building energy consumption prediction:

Data from the first energy meter with measurement (Real energy into the load), which is used to predict energy consumption for the next month, was utilized for training. The data is divided into a training set and a test set in the first section of the algorithm, which reads in a dataset and resamples it to a daily frequency. Subsequences of length 10 are then constructed as input features from the training set using min-max normalization, with the associated output value being the following value in the sequence.

The final step is to define an LSTM model with four LSTM layers and an output layer. For 100 epochs, the model is trained using the training set.

Concatenating the test set with the training set's final 10 values—which serve as the model's first input values—prepares it for input. Subsequences of length 10 are then constructed as input features once the test set has been normalized.

For each input subsequence in the test set, the model is used to forecast the following value in the sequence. Following an inverse transformation to restore the predicted values to their original scale, the true and predicted values are shown opposite one another on a line graph.

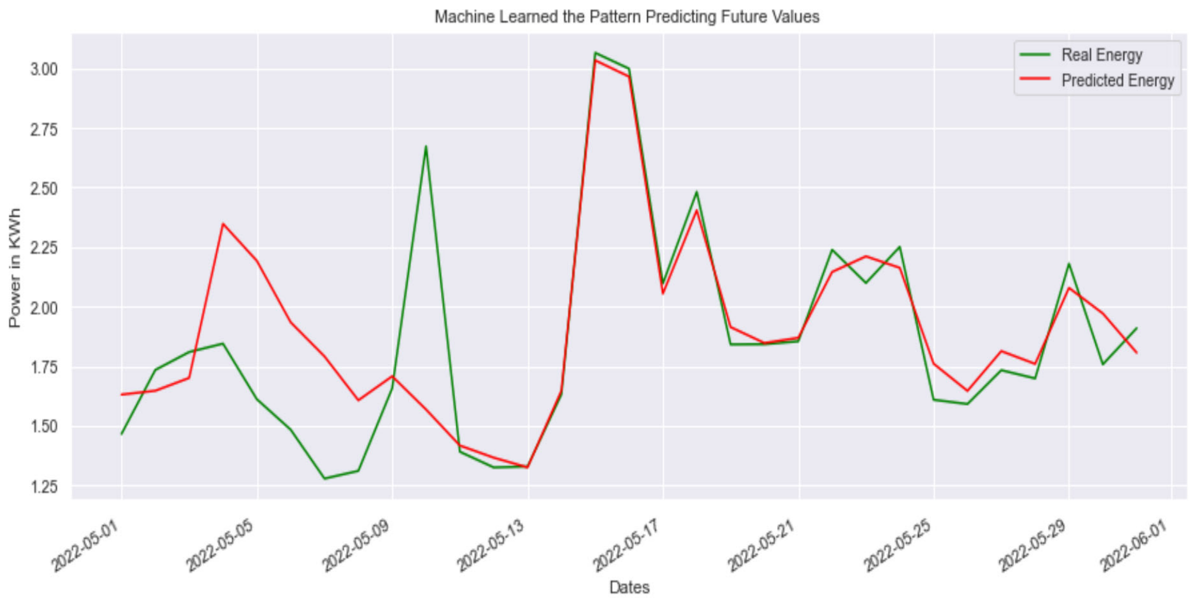

Fig. 54. Building energy prediction using RNN(LSTM).

The results show the real building energy consumption in green color. The energy line in red is what the RNN model anticipated. Although it follows the green line, the predicted line closely tracks the actual line, indicating that the model is accurate and reliable in its predictions.

4.8.1. Model performance & accuracy:

You can use a variety of assessment measures, including mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R-squared), to assess the correctness of your LSTM model. The test set's true values and predicted values can be used to produce these measures.

Better performance can be measured by lower MSE, RMSE, and MAE values as well as greater R-squared values. To enhance the performance of your model, you can tweak it and experiment with various hyperparameters. This model has low MSE, RMSE and MAE values as compared to baseline values, so there is less error in this model. In general, an R2 value of 0.5 or above is considered a good fit in most fields, but a lower R2 value may still be useful in certain cases.

Output:

Model MSE:	0.0826878127160719
Model RMSE	0.2875548864409574
Model MAE	0.17042009594857224
Model R-squared	0.5994332210468443

In this case, the model has achieved a significant improvement in all the metrics. The model's MSE, RMSE, and MAE values are all lower than the baseline, indicating that the model's predictions are closer to the actual values. The R-squared value is also higher than the 0.5, which means that the model fits the data much better.

Conclusion

In conclusion, smart building energy consumption data has the potential to be greatly improved in terms of energy efficiency and cost reduction by applying data analysis and machine learning techniques. Different data analytics techniques used to identify patterns and forecast energy usage through the study of historical data and real-time monitoring, enabling the implementation of more precise and efficient energy management systems. This may result in lower carbon emissions, less energy use, and more sustainability.

This thesis provided the pattern and trends of the transmission of energy from solar PV towards the grid, battery, loads and energy consumption from the grid to loads and battery storage system. Additionally, it provides a calculation of the amount of energy in kWh used by each source's load that helps building owners to understand how much energy they can save by using solar PV and battery storage systems.

The economic and environmental consequences of energy consumption, however, make this even more important. Identification of equipment and user consumption patterns in a smart building management system will undoubtedly result in cost savings, increased comfort, and lower pollutant emissions. The number of new businesses using data science to analyze user behavior, building infrastructure, and energy consumption data is a sign of the sector's economic significance. As a result, energy firms and information technology companies are beginning to collaborate in order to manage energy more effectively.

The study's findings demonstrate that the direct method based on recurrent models (recurrent neural network) with long short-term memory that stores information over long period of time can make the most precise predictions without substantially raising the workload. Based on past data and ongoing monitoring, machine learning algorithms (RNN) have demonstrated considerable promise in accurately predicting energy usage patterns. Building owners and managers can enhance efficiency, lower energy use, and create more efficient energy management strategies by implementing these practices.

Reference:

1. Fan, C., Yan, D., Xiao, F., Li, A., An, J., & Kang, X. (2021, February). Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches. In *Building Simulation* (Vol. 14, pp. 3-24). Tsinghua University Press.
2. Molina-Solana, M., Ros, M., Ruiz, M. D., Gómez-Romero, J., & Martín-Bautista, M. J. (2017). Data science for building energy management: A review. *Renewable and Sustainable Energy Reviews*, *70*, 598-609.
3. Qolomany, B., Al-Fuqaha, A., Gupta, A., Benhaddou, D., Alwajidi, S., Qadir, J., & Fong, A. C. (2019). Leveraging machine learning and big data for smart buildings: A comprehensive survey. *IEEE Access*, *7*, 90316-90356.
4. Lu, C., Li, S., & Lu, Z. (2022). Building energy prediction using artificial neural networks: A literature survey. *Energy and Buildings*, *262*, 111718.
5. Fan, C., Wang, J., Gang, W., & Li, S. (2019). Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. *Applied energy*, *236*, 700-710.
6. Wang, L., Wang, Z., & Yang, R. (2012). Intelligent multiagent control system for energy and comfort management in smart and sustainable buildings. *IEEE transactions on smart grid*, *3*(2), 605-617.
7. Shaikh, P. H., Nor, N. B. M., Nallagownden, P., Elamvazuthi, I., & Ibrahim, T. (2016). Intelligent multi-objective control and management for smart energy efficient buildings. *International Journal of Electrical Power & Energy Systems*, *74*, 403-409.
8. Mariano-Hernández, D., Hernández-Callejo, L., Zorita-Lamadrid, A., Duque-Pérez, O., & García, F. S. (2021). A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis. *Journal of Building Engineering*, *33*, 101692.
9. Kang, S. J., Park, J., Oh, K. Y., Noh, J. G., & Park, H. (2014). Scheduling-based real time energy flow control strategy for building energy management system. *Energy and Buildings*, *75*, 239-248.
10. Sharma, S., Xu, Y., Verma, A., & Panigrahi, B. K. (2019). Time-coordinated multienergy management of smart buildings under uncertainties. *IEEE Transactions on Industrial Informatics*, *15*(8), 4788-4798.

11. Arun, S. L., & Selvan, M. P. (2017). Intelligent residential energy management system for dynamic demand response in smart buildings. *IEEE Systems Journal*, *12*(2), 1329-1340.
12. Kifouche, A., Hamouche, R., Kocik, R., Rachedi, A., & Baudoin, G. (2019). Model driven framework to enhance sensor network design cycle. *Transactions on emerging telecommunications technologies*, *30*(8), e3560.
13. Iqbal, J., Khan, M., Talha, M., Farman, H., Jan, B., Muhammad, A., & Khattak, H. A. (2018). A generic internet of things architecture for controlling electrical energy consumption in smart homes. *Sustainable cities and society*, *43*, 443-450.
14. Mir, U., Abbasi, U., Mir, T., Kanwal, S., & Alamri, S. (2021). Energy management in smart buildings and homes: current approaches, a hypothetical solution, and open issues and challenges. *IEEE Access*, *9*, 94132-94148.
15. Shakeri, M., Pasupuleti, J., Amin, N., Rokonuzzaman, M., Low, F. W., Yaw, C. T., ... & Lai, C. W. (2020). An overview of the building energy management system considering the demand response programs, smart strategies and smart grid. *Energies*, *13*(13), 3299.
16. Jamil, M., & Mittal, S. (2017, December). Building energy management system: A review. In *2017 14th IEEE India Council International Conference (INDICON)* (pp. 1-6). IEEE.
17. Di Piazza, M. C., La Tona, G., Luna, M., & Di Piazza, A. (2017). A two-stage Energy Management System for smart buildings reducing the impact of demand uncertainty. *Energy and Buildings*, *139*, 1-9.
18. Howell, S. K., Wicaksono, H., Yuce, B., McGlinn, K., & Rezgui, Y. (2018). User centered neuro-fuzzy energy management through semantic-based optimization. *IEEE transactions on cybernetics*, *49*(9), 3278-3292.
19. Aguilar, J., Garces-Jimenez, A., Gallego-Salvador, N., De Mesa, J. A. G., Gomez-Pulido, J. M., & Garcìa-Tejedor, À. J. (2019). Autonomic management architecture for multi-HVAC systems in smart buildings. *IEEE Access*, *7*, 123402-123415.
20. Du, Y. F., Jiang, L., Duan, C., Li, Y. Z., & Smith, J. S. (2017). Energy consumption scheduling of HVAC considering weather forecast error through the distributionally robust approach. *IEEE Transactions on Industrial Informatics*, *14*(3), 846-857.
21. Zhao, Y., Li, T., Zhang, X., & Zhang, C. (2019). Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renewable and Sustainable Energy Reviews*, *109*, 85-101.

22. Gaur, M., Makonin, S., Bajić, I. V., & Majumdar, A. (2019). Performance evaluation of techniques for identifying abnormal energy consumption in buildings. *IEEE Access*, *7*, 62721-62733.
23. Deshmukh, S., Samouhos, S., Glicksman, L., & Norford, L. (2019). Fault detection in commercial building VAV AHU: A case study of an academic building. *Energy and Buildings*, *201*, 163-173.
24. Li, D., Zhou, Y., Hu, G., & Spanos, C. J. (2018). Identifying unseen faults for smart buildings by incorporating expert knowledge with data. *IEEE Transactions on Automation Science and Engineering*, *16*(3), 1412-1425.
25. Marinakis, V. (2020). Big data for energy management and energy-efficient buildings. *Energies*, *13*(7), 1555.
26. Sheng, G., Hou, H., Jiang, X., & Chen, Y. (2016). A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model. *IEEE Transactions on Smart Grid*, *9*(2), 695-702.
27. Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., ... & Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, *82*, 1027-1047.
28. Yin, R., Ghatikar, G., & Piette, M. (2019). Big-Data Analytics for Electric Grid and Demand-Side Management.
29. Zeng, A., Liu, S., & Yu, Y. (2019). Comparative study of data driven methods in building electricity use prediction. *Energy and Buildings*, *194*, 289-300.
30. Ruiz, M. D., Gómez-Romero, J., Fernandez-Basso, C., & Martin-Bautista, M. J. (2021). Big Data Architecture for Building Energy Management Systems. *IEEE Transactions on Industrial Informatics*, *18*(9), 5738-5747.
31. Mehmood, M. U., Chun, D., Han, H., Jeon, G., & Chen, K. (2019). A review of the applications of artificial intelligence and big data to buildings for energy-efficiency and a comfortable indoor living environment. *Energy and Buildings*, *202*, 109383.
32. Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2021). Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Developments in the Built Environment*, *5*, 100037.
33. Molina-Solana, M., Ros, M., Ruiz, M. D., Gómez-Romero, J., & Martín-Bautista, M. J. (2017). Data science for building energy management: A review. *Renewable and Sustainable Energy Reviews*, *70*, 598-609.

34. Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, *22*(1), 67.
35. Sun, Y., Haghighat, F., & Fung, B. C. (2020). A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy and Buildings*, *221*, 110022.
36. Fan, C., Xiao, F., Yan, C., Liu, C., Li, Z., & Wang, J. (2019). A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. *Applied Energy*, *235*, 1551-1560.
37. Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31-57.
38. Zhao, Y., Zhang, C., Zhang, Y., Wang, Z., & Li, J. (2020). A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. *Energy and Built Environment*, *1*(2), 149-164.
39. Wahid, F., & Kim, D. (2016). A prediction approach for demand analysis of energy consumption using k-nearest neighbor in residential buildings. *International Journal of Smart Home*, *10*(2), 97-108.
40. Bourdeau, M., qiang Zhai, X., Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, *48*, 101533.
41. Patrick Donovan, White Paper, 2020 ,"Three Essential Elements of Next Generation Building Management Systems (BMS)", Schneider Electric. https://www.se.com/ww/en/work/products/building-management/
42. Popławski, T., Dudzik, S., Szeląg, P., & Baran, J. (2021). A case study of a virtual power plant (VPP) as a data acquisition tool for PV energy forecasting. *Energies*, *14*(19), 6200.

Appendix

1. Code for prediction model development:

```
import keras as keras
import tensorflow as tf
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
from matplotlib import style
import sklearn
from sklearn.preprocessing import MinMaxScaler


#Import and read CSV file

df=pd.read_csv('building_consumption.csv',parse_dates=['Timest
amp','Timestamp UTC'])

columns_to_drop = [ 'Timestamp UTC', 'Source','Unit']
 # Replace with actual column names
df = df.drop(columns=columns_to_drop)


into = dataset['Measurement'] == 'Real Energy Into the Load'
df1= dataset[into]

df1['New_value'] = df1[['Value']].diff()

df1 = df1.iloc[3:]


df1["Month"] = pd.to_datetime(df1["Timestamp"]).dt.month
df1["Date"] = pd.to_datetime(df1["Timestamp"]).dt.date
df1["Time"] = pd.to_datetime(df1["Timestamp"]).dt.time
df1["Hour"] = pd.to_datetime(df1["Timestamp"]).dt.hour
df1["Week"] =
pd.to_datetime(df1["Timestamp"]).dt.isocalendar().week
df1["Day"] = pd.to_datetime(df1["Timestamp"]).dt.day_name()


df1.head(1)
```

```
# Descriptive analysis

from matplotlib import style

fig = plt.figure()
ax1 = plt.subplot2grid((1,1), (0,0))

style.use('ggplot')

sns.lineplot(x=df1["Date"], y=df1["New_value"], data=df1,
label = 'KWh')
sns.set(rc={'figure.figsize':(15,6)})

plt.title("Energy consumption in May ")
plt.xlabel("Date")
plt.ylabel("Energy in KWh")
plt.grid(True)
plt.legend(loc='upper left')

for label in ax1.xaxis.get_ticklabels():
    label.set_rotation(90)

# Sum of energy for everyday in whole month



# Group the data by day and sum the "Grid to Battery (Charge
+) [kWh]" column for each day
daily_energy_data1 = df1.groupby('Date')['New_value'].sum()

print(daily_energy_data1)
# Plot the data
plt.plot(daily_energy_data1)
plt.xlabel('Date')
plt.ylabel('Energy (kWh)')
```

```
plt.title('PV to grid Energy for the Month of may 2022')
plt.grid()
plt.show()



# Hourly behavior of energy consumption data

fig = plt.figure()
ax1= fig.add_subplot(111)

sns.lineplot(x=df1["Hour"],y=df1["New_value"], data=df1, label
= 'KWh', color = 'blue')
plt.title("Energy Consumption vs Time ")
plt.xlabel("Time")
plt.ylabel('KWh')
plt.grid(True, alpha=1)
plt.legend(loc= 'upper left')

for label in ax1.xaxis.get_ticklabels():
    label.set_rotation(90)

#Resampling of data
df1.describe()

NewDataSet = df1.resample('D').mean(numeric_only=True)
print("Old Dataset ",df1.shape )
print("New  Dataset ",NewDataSet.shape )


TestData = NewDataSet.tail(60)
Training_Set = NewDataSet.iloc[:,0:1]
Training_Set = Training_Set[:-10]


print("Training Set Shape ", Training_Set.shape)
print("Test Set Shape ", TestData.shape)

Training_Set = NewDataSet.iloc[:, 0:1]

Training_Set = Training_Set.to_numpy()
```

```
sc = MinMaxScaler(feature_range=(0, 1))
Train = sc.fit_transform(Training_Set)


X_Train = []
Y_Train = []

# Range should be from 60 Values to END

for i in range(10, Train.shape[0]):

    # X_Train 0-59
    X_Train.append(Train[i-10:i])

    # Y Would be 60 th Value based on past 60 Values
    Y_Train.append(Train[i])

# Convert into Numpy Array
X_Train = np.array(X_Train)
Y_Train = np.array(Y_Train)

print(X_Train.shape)
print(Y_Train.shape)

import keras
from keras.models import Sequential
from keras.layers import LSTM, Dropout, Dense

regressor = Sequential()

# Adding the first LSTM layer and some Dropout regularisation
regressor.add(LSTM(units =50, return_sequences = True,
input_shape = (X_Train.shape[1], 1)))
regressor.add(Dropout(0.2))

# Adding a second LSTM layer and some Dropout regularisation
regressor.add(LSTM(units = 50, return_sequences = True))
regressor.add(Dropout(0.2))

# Adding a third LSTM layer and some Dropout regularisation
regressor.add(LSTM(units = 50, return_sequences = True))
regressor.add(Dropout(0.2))

# Adding a fourth LSTM layer and some Dropout regularisation
regressor.add(LSTM(units = 50))
regressor.add(Dropout(0.2))

# Adding the output layer
regressor.add(Dense(units = 1))
```

```
# Compiling the RNN
regressor.compile(optimizer = 'adam', loss =
'mean_squared_error')

regressor.fit(X_Train, Y_Train, epochs = 100, batch_size = 32)


TestData.shape

NewDataSet.shape

Df_Total = pd.concat((NewDataSet[["New_value"]],
TestData[["New_value"]]), axis=0)

Df_Total.shape

inputs = Df_Total[len(Df_Total) - len(TestData) - 10:].values
inputs.shape

inputs = Df_Total[len(Df_Total) - len(TestData) - 10:].values

# We need to Reshape
inputs = inputs.reshape(-1,1)

# Normalize the Dataset
inputs = sc.transform(inputs)

X_test = []
for i in range(10, 110):
    element = inputs[i-10:i]
    if element.shape == (10, 1):
        X_test.append(element)
    else:
        print(f"Ignoring element with shape {element.shape}")


# Convert into Numpy Array
X_test = np.array(X_test)

# Reshape before Passing to Network
X_test = np.reshape(X_test, (X_test.shape[0], X_test.shape[1],
1))

# Pass to Model
predicted_stock_price = regressor.predict(X_test)

# Do inverse Transformation to get Values
predicted_stock_price =
sc.inverse_transform(predicted_stock_price)
```

```
True_KiloWatt = TestData["New_value"].to_list()
Predicted_KiloWatt  = predicted_stock_price
dates = TestData.index.to_list()


True_KiloWatt = TestData["New_value"].to_list()
Predicted_KiloWatt  =
predicted_stock_price[:len(True_KiloWatt)]
dates = TestData.index.to_list()


print(len(dates))
print(len(True_KiloWatt))
print(len(Predicted_KiloWatt))


Machine_Df = pd.DataFrame(data={
    "Date":dates,
    "TrueKiloWatt": True_KiloWatt,
    "PredictedKiloWatt":[x[0] for x in Predicted_KiloWatt ]
})

Machine_Df


True_KiloWatt = TestData["New_value"].to_list()
Predicted_KiloWatt  = [x[0] for x in Predicted_KiloWatt ]
dates = TestData.index.to_list()


fig = plt.figure()

ax1= fig.add_subplot(111)

x = dates
y = True_KiloWatt

y1 = Predicted_KiloWatt

plt.plot(x,y, color="green" ,label= 'Real Energy')
plt.plot(x,y1, color="red",label= 'Predicted Energy')
# beautify the x-labels
plt.gcf().autofmt_xdate()
plt.xlabel('Dates')
plt.ylabel("Power in KWh")
plt.title("Machine Learned the Pattern Predicting Future
Values ")
plt.legend()
```

2. Prediction model accuracy & error:

s

```
from sklearn.metrics import mean_squared_error,
mean_absolute_error, r2_score

# Calculate predicted values
Predicted_KiloWatt = [x[0] for x in predicted_stock_price]
Predicted_KiloWatt = [x[0] for x in
predicted_stock_price[:len(True_KiloWatt)]]



# Calculate evaluation metrics
mse = mean_squared_error(True_KiloWatt, Predicted_KiloWatt)
rmse = mean_squared_error(True_KiloWatt, Predicted_KiloWatt,
squared=False)
mae = mean_absolute_error(True_KiloWatt, Predicted_KiloWatt)
r2 = r2_score(True_KiloWatt, Predicted_KiloWatt)

# Print results
print("MSE:", mse)
print("RMSE:", rmse)
print("MAE:", mae)
print("R-squared:", r2)
```