

**Master Thesis**

**Topic: Predicting Energy Production or Consumption of Prosumers: Leveraging the Power of Meta’s Prophet Model**

**Jibril Richter**

**ID: user15318672**

**East London University**

**School of Architecture, Computing and Engineering**

Table of Contents

[Introduction 7](#_Toc164602315)

[Background 7](#_Toc164602316)

[Research Problem 10](#_Toc164602317)

[Research question. 11](#_Toc164602318)

[Research Significance. 11](#_Toc164602319)

[Literature Review 12](#_Toc164602320)

[Prosumers: Definition and Types. 12](#_Toc164602321)

[Prosumers: Enabling Factors. 15](#_Toc164602322)

[Prosumers: Benefits and Challenges. 17](#_Toc164602323)

[Predicting Energy Behavior of Prosumers 18](#_Toc164602324)

[Related Work on Prosumers. 21](#_Toc164602325)

[Methodology. 23](#_Toc164602326)

[Data Source 23](#_Toc164602327)

[Target- Total Energy Produced/Consumed by Energy Prosumers. 31](#_Toc164602328)

[Summary Statistics of the Target Datasets 31](#_Toc164602329)

[Tools and Libraries Used 31](#_Toc164602330)

[Meta Prophet Model. 32](#_Toc164602331)

[Model Evaluation Metric: Root Mean Square Error (RMSE). 32](#_Toc164602332)

[Calculating the RMSE. 33](#_Toc164602333)

[Findings: Analyzing Datasets and Model Building. 34](#_Toc164602334)

[Data Distribution of Total Energy Produced or Consumed by Prosumers 34](#_Toc164602335)

[Analyzing Energy Production Data 35](#_Toc164602336)

[Distribution Total Energy Produced by Prosumers. 35](#_Toc164602337)

[Total Energy Produced by Prosumers Per Month. 36](#_Toc164602338)

[Total Energy Produced by Prosumers Per Day of Week. 37](#_Toc164602339)

[Total Energy Produced by Prosumers Per Hour. 38](#_Toc164602340)

[Interaction between Solar Radiation and Installation Capacity and its Relationship to Energy Production. 39](#_Toc164602341)

[Partial Autocorrelation of Energy Produced. 40](#_Toc164602342)

[Relationship between Lag 1 and Total Energy Produced by Prosumers. 41](#_Toc164602343)

[Modelling Total Energy Produced 42](#_Toc164602344)

[The Model’s Decomposition of the Components of the Energy Production Distribution of Prosumers. 43](#_Toc164602345)

[Comparing Model Prediction to Actuals. 44](#_Toc164602346)

[Distribution Total Energy Consumed by Prosumers. 45](#_Toc164602347)

[Distribution of Total Energy Consumed by per Month Prosumers. 46](#_Toc164602348)

[Energy Produced per Day of Week by Prosumers. 47](#_Toc164602349)

[Distribution of Total Energy Consumed by per Month Prosumers. 48](#_Toc164602350)

[Partial Autocorrelation of Energy Produced. 49](#_Toc164602351)

[Relationship between Lag 1 and Total Energy Consumed by Prosumers. 50](#_Toc164602352)

[Relationship between Lag 2 and Total Energy Consumed by Prosumers. 51](#_Toc164602353)

[Modelling Total Energy Consumed. 52](#_Toc164602354)

[The Model’s Decomposition Components of the Energy Consumption of Prosumers. 53](#_Toc164602355)

[Comparing Model Prediction to Actuals. 54](#_Toc164602356)

[Predicting Energy Deficit or Surplus 55](#_Toc164602357)

[Discussion and Conclusion 56](#_Toc164602358)

[Conclusion 57](#_Toc164602359)

**List of Tables**

**Table 1: Training Dataset............................................. 26**

**Table 2: Dataset on Gas Prices............................................27**

**Table 3: Dataset on clients............................................27**

**Table 4: Dataset on electricity prices............................................28**

**Table 5: Dataset on weather variables............................................29**

**List of Figures**

**Table 1: Data Distribution of Total Energy Produced or Consumed by Prosumers 34.**

**Table 2: Distribution Total Energy Produced by Prosumers 35.**

**Figure 3: Total Energy Produced by Prosumers Per Month 36.**

**Figure 4: Total Energy Produced by Prosumers Per Day of Week 37.**

**Figure 5: Total Energy Produced by Prosumers Per Hour 38.**

**Figure 6: Interaction between Solar Radiation and Installation Capacity and its Relationship to Energy Production 39.**

**Figure 7: Partial Autocorrelation of Energy Produced 40.**

**Figure 8: Relationship between Lag 1 and Total Energy Produced by Prosumers 41.**

**Figure 9: Decomposition of the Components of the Energy Production Distribution of Prosumers 43.**

**Figure 10: Comparing Model Prediction on Energy Production to Actuals 44.**

**Figure 11: Distribution Total Energy Consumed by Prosumers 45.**

**Figure 12: Distribution of Total Energy Consumed by per Month Prosumers 46.**

**Figure 13: Energy Produced per Day of Week by Prosumers 47.**

**Figure 14: Distribution of Total Energy Consumed per Month by Prosumers 48.**

**Figure 15: Partial Autocorrelation of Energy Consumed 49.**

**Figure 16: Relationship between Lag 1 and Total Energy Consumed by Prosumers 50.**

**Figure 17: Relationship between Lag 2 and Total Energy Consumed by Prosumers 51.**

**Figure 18: The Model’s Decomposition Components of the Energy Consumption of Prosumers 53.**

**Figure 19: Comparing Model Prediction to Actuals 54.**

**Figure 20: Predicting Energy Deficit or Surplus 55.**

**Abstract:**

The advancement in technology has made generating and consuming energy available to energy customers on the grid, leading to the concept of energy “prosumers”. Motivated by social and economic reasons, the number of prosumers keeps rising in the various energy markets. While prosumers bring several advantages to the energy market, prosumers present challenges to energy retailing companies. For example, many prosumers rely on non-reliable renewable energy resources to produce energy, and in some cases, either produce more than they consume or consume more than they produce. Therefore, it is vital for energy retailing companies to predict energy surplus or deficit by prosumers in order to manage resources efficiently. This project uses Meta’s Prophet model to train and predict the energy behavior of prosumers. The aim is to demonstrate how energy retailing companies could the Prophet model to manage their resources efficiently.

# Introduction

## Background

How could energy retail companies predict the behavior of energy prosumers? In this project, using an energy behavior dataset of prosumers provided by Eesti Energia, a public limited energy company in Estonia, I train Meta’s Prophet model to predict both hourly energy production and consumption of prosumers. This is to help energy retailers to predict energy surplus or deficit by these type of customers for resources management and other important decisions.

Energy consumption keeps growing, and much of the consumption is in the Americas, Europe, and Asia (Our World of Data, 2023). As of 2023, the world consumes 160,000 TWh of energy, with a significant part of it coming from oil, coal, and natural gas (Our World of Data, 2023). This is logical given the continuous increase in global population and electrification projects happening in many parts of the developing countries. The global energy demand is projected to grow as the global population grows and the need for power systems continues to be necessary.

At the same time, the demand for energy has led to environmental issues like land degradation, and the rise in average global temperatures. According to the International Energy Agency, CO2 emissions per energy supply is 54.3 tCO2. The use of fossil fuel is currently the main source of energy and is a major contributor to climate change (International Energy Agency). Efforts have been made to reduce the reliance on fossil fuels. Thankfully, technological advancement such as solar photovoltaics (PV) and wind turbines has made it possible to generate energy using solar energy and wind energy, respectively.

A corollary to this development has led to a new form of energy customers termed as “Prosumers”. The term prosumer is a blend of two words- producer and consumer; a term coined by Alvin Toffler, who introduced the producer-consumer, which is a proactive consumer, in his book The Third Wave (Toffler 1981).

According to Toffler, prosumers are a group of customers actively taking part in the process of cocreating products. Prosumers have emerged in the energy market during the last decade, thanks to the emergence of low-cost renewable energy technologies (RETs) such as solar photovoltaic (PV) panels (Toffler 1981). A broad definition of an energy prosumer recognizes the multifaceted character of the position as a consumer who also generates, sells, trades, or stores energy (Ford et al. 2016). Sustainability transformations are underway in a variety of areas. Climate change and rising energy demand have compelled us to consider more efficient and sustainable energy production and consumption methods. Decentralization, digitization, and electrification are the major breakthroughs that enable high-level energy transformation (Astarloa et al. 2017). The inclusion of small-scale variable energy sources such as solar and wind necessitates additional flexibility from the energy system, necessitating improvements in energy efficiency and demand side management (DSM) (Kotilainen, 2019).

Prosumers have been categorized in the literature by virtue of their legal status, organizational type, size, type of energy they generate, type of energy sources they use, amount of energy produced, consumed, or sold, and relationship with the power grid (e.g., European Commission 2016; Jacobs 2017; Kampman et al. 2016; Masera and Couture 2015; IEA-RETD 2014; Sajn 2016). Also, energy prosumers may be explained using at least three approaches: their function as energy market players, the energy type they create and the energy sources they utilize, and their interaction with the power system (Kotilainen, 2019).

Prosumer participation in the energy market has various advantages. Prosumers are new participants in the energy market, and their actions encourage the adoption of renewable energy sources and distributed generation. On-site power generation contributes to reducing losses during transmission and distribution, reducing the need for additional transmission and distribution capacity, increasing local community resilience in the absence of a central power system, and providing economic opportunities for individuals and communities. Furthermore, growing usage of renewable energy sources reduces emissions both worldwide and locally (Kotilainen, 2019).

Nevertheless, prosumers pose significant challenges to energy retailers. Prosumer base development can cause technical problems for the grid, resulting in quality and reliability concerns such as overvoltage, congestion, back-feeding into the circuit, instability, and system design obstacles. More importantly, the energy system could encounter technological issues when the amount of intermittent energy supplied to the grid increases dramatically. Electricity networks have typically relied on centralized energy generation from massive power facilities that produce predictable results. The energy grid must maintain stability and balance in terms of supply and demand; hence predictability is critical in system management. Stricter environmental laws make growing traditional generation capacity less economical, moving the focus to distributed generation and renewable energy sources. The inherently irregularity of the supply of renewable energy sources makes forecasting considerably more difficult.

Hence, the ability of energy retail corporations to forecast energy deficit and surplus of prosumers is crucial for business decisions. Luckily, Machine learning and Artificial Intelligence models help us to predict outputs based on data. With well trained models, data professionals are able to predict target variables. Data professionals could develop models from scratch or use powerful existing models.

For this project I used the Prophet model developed by Meta. Meta’s Prophet is a forecasting tool developed by Meta's Core Data Science team, designed to make it easier for analysts and developers to produce high-quality forecasts for time series data. It was released as an open-source project in 2017.

Forecasting energy surplus or deficit of prosumers is a time series problem as it is influenced by time features. Hence, the Meta Prophet model is ideal for the project. The model is able to decompose the components of the time series data- the trend, seasonality and datetime features. In addition, the model allows us to add other regressors that help accurately predict the target variable. Moreover, the model is fast to train and requires less computational power compared to other options like Neural Networks and Gradient Boosting Machines such as XGBoost and LightGBM. Furthermore, the Facebook Prophet is robust against outliers. It obviates the need for data transformations such as Box-Cox to normalize the data.

Therefore, this project uses Meta’s Prophet Model to forecast energy production and consumption of prosumers.

## Research Problem

Energy generation is one of the cornerstones of modern society. It powers electricity in our homes and workplace, aids in the production of goods and services, powers transportation and health industries, among other key industries. In summary, man’s ability to generate energy as and when we want is one of the key aspects of modern civilization. However, the rising temperature of the global climate, partly induced by human actions has necessitated the move away from the use of fossil fuel as a means of generating energy. Also, several advancements in technology have made it possible for individual households and organizations to generate their own energy on and off the grid to power their homes and businesses. While prosumers generally have advantages such as less reliance on the main energy grid and help reduce the demand of energy, ergo less use of fossil fuels, they have also created the problem for energy retailers. Prosumers’ ability to generate energy varies by time, especially those who rely on renewable energy such us the sun, the wind to generate power. Therefore, accurately predicting energy production and consumption of prosumers, and being able to determine energy surplus or deficit by prosumers who are on the grid becomes necessary for energy retailers to properly manage resources their resources.

Since energy production and usage depend on time features such as time of day, day of the week, and month of the year, it is prudent to model this problem as a timeseries problem. Meta’s Prophet model is a powerful time series model that could be adopted by retail organizations and other stakeholders who want to predict energy production and consumption of customers.

## Research question.

The research question of this thesis: "How can Meta’s Prophet modeling be utilized to predict energy production and consumption patterns of prosumers, and how does this capability assist energy retailers in resource planning and management for improved efficiency and grid stability?"

## Research Significance.

As many customers embrace self-generation of energy, it has become pertinent for energy retail companies to model and be able to predict energy surplus and deficit of prosumers to manage their energy resources. The aim of this research is to analyze the Eesti Energia dataset on prosumers, unearth factors that predict energy production and consumption of prosumers, and use Meta’s Prophet model to train and predict energy behavior of prosumers.

# Literature Review

## Prosumers: Definition and Types.

The general concept of a prosumer is based on the phrases- producer and consumer; the term was coined by Alvin Toffler, who originally presented the producer-consumer, also known as a proactive consumer, in his book The Third Wave (1981). Toffler defines prosumers as a group of customers who actively participate in the process of cocreating products. Essentially, the prosumer is an agent that generates part of what it consumes (Kotilainen, 2019).

Another common characterization, emphasizing the prosumer's customer function, is that of a professional consumer, which refers to a keen enthusiast or semiprofessional customer who requires professional-grade products and services, such as digital cameras, espresso coffee machines, or solar panels. The idea of prosumers has evolved further (e.g., Kotler 1986; Tapscott and Williams 2008), particularly in the context of mass customization, marketing, and, more recently, social media. In addition to prosumers, other names are used to express the same phenomenon, such as competent customers (Prahalad and Ramaswamy 2000) and working consumers (Cova and Dalli 2009).

Prosumers have emerged in the energy market during the last decade, primarily due to the emergence of low-cost renewable energy technologies (RETs) such as solar photovoltaic (PV) panels. A broad definition of an energy prosumer recognizes the multifaceted character of the position as a consumer who also generates, sells, trades, or stores energy (Ford et al. 2016).   
The European Commission defines energy prosumers as "active consumers": "a customer or a group of jointly acting customers who consume, store, or sell electricity generated on their premises, including through aggregators, or participate in demand response or energy efficiency schemes provided that these activities do not constitute their primary commercial or professional activity" (European Commission, 2016).

Energy consumers are diverse and vary in size. As a result, prosumers are difficult to categorize. However, there has been a request for a more precise definition of energy prosumers and their role in the energy system (Jacobs 2017). Prosumers have been described in the literature based on their legal status, organizational type, size, type of energy they generate, type of energy sources they use, amount of energy produced, consumed, or sold, and relationship with the power grid (e.g., European Commission 2016; Jacobs 2017; Kampman et al. 2016; Masera and Couture 2015; IEA-RETD 2014; Sajn 2016). Energy prosumers may be explained using at least three approaches: their role as energy market players, the energy type they create and the energy sources they utilize, and their interaction with the power system (Kotilainen, 2019).

Prosumers are usually categorized based on their position in energy markets; utilities and retailers consider prosumers to be key consumers. As a result, energy consumers are often divided into three categories: residential, commercial, and industrial. The exact classification of various prosumers into these groups is ambiguous, but it may be simplified (Kotilainen, 2019).

First are residential prosumers, also known as domestic prosumers. This group usually comprise of houses, apartment complexes, housing associations, cooperatives, or collectives. Residential prosumers often live in buildings with access to electricity and use it for lighting, heating, cooking, and power for domestic and household needs. Another group is Commercial prosumers, which include micro, small, and big enterprises, department shops, shopping malls, hospitals, schools, offices, and sporting facilities. Commercial prosumers utilize power, heating, and cooling for personal consumption while producing goods and services for the public.   
Finally, industrial prosumers are energy users who primarily engage in manufacturing, such as factories, mines, mills, plants, or farms (Kotilainen, 2019).

According to Kotilainen (2019), an alternative method of characterizing energy prosumers is by examining their interaction with the electrical grid, or in a comparable way, the district heating system. In this regard, prosumers can choose to be either off-grid or connected to the grid. The two primary motivations for off-grid prosumers to stay off the grid are their inability to access the electrical grid and their desire to generate enough energy to meet their needs. Off-grid prosumers are more prevalent in underdeveloped nations, rural areas, and places where the cost of energy is significantly greater than that of self-generation. Agricultural entrepreneurs in isolated villages that handle their own energy needs—heating, cooling, and power—are frequently the rural off-grid prosumers (Masera and Couture 2015).

However, the majority of prosumers remain connected to the grid and only generate a portion of the energy they need. For such a group of prosumers, energy retail companies need to be able to predict their energy production and consumption for resource management. To varied degrees, prosumers who are linked to the grid for power are dependent on it. Self-consumption can help reduce your power costs, particularly if your rates are high. However, in certain regions, prosumers have been pushed to feed all or a significant portion of their energy to the grid by attractive pricing schemes like net metering and feed-in tariffs (FITs) that it is now posing a financial risk to the utilities. Self-consumption rates are often greater among commercial prosumers than household prosumers. For instance, self-use ratios of 75–100% may be attained by commercial and manufacturing buildings in Germany and Spain (IEA-RETD 2014).

Prosumers who are linked to the grid have many options for participating in the energy markets. For instance, they can take part in demand response and flexibility programs run by energy corporations, which act as aggregators for energy service companies (ESCO). In addition to trading energy in virtual communities, grid-connected prosumers can also be members of micro-grid communities or VPPs (Kotilainen, 2019).

Energy prosumers could form communities, who can then sell whatever extra energy they generate. Energy selling can occur peer-to-peer (P2P), in virtual communities, micro-grids, or with the main power grid. In most markets, it is permitted and rewarded to feed surplus energy back into the electricity system (Ramirez et al. 2017). While P2P energy sales are generating a lot of enthusiasm, the legal and legislative environment will necessitate revisions, and technological platforms and business model preparation are still in the demonstration phase. (Martin, 2015).

## Prosumers: Enabling Factors.

Several factors enable energy consumers to become prosumers. For customers to accept new technology, the right combination of noneconomic and economic elements must exist (Rickerson et al. 2014). National factors that affect the prosumer potential include the amount of rooftop space that is available in cities. Understanding solar PV potential requires an estimation of roof area, particularly in highly populated cities like those in Southeast Asia (Byrne et al. 2015). varied geographic locations have varied levels of solar radiation, or insolation, which affects the maximum amount of solar energy that may be produced. For example, compared to countries nearer the equator, the Nordic nations receive significantly less insolation (Kotilainen, 2019).

In addition, technological advancements have facilitated energy customers to generate part of the power they consume. According to Kotilainen (2019), PV solar power has improved in efficiency and cost. Solutions for home battery storage are evolving, and costs are beginning to come down. Modern electricity networks are being transformed into "smart grids," which allow energy and information to flow in both directions and provide a significant quantity of data on production, distribution, and consumption. As a component of an automatic measurement infrastructure (AMI), smart meters oversee the measurement and control of energy and supply copious amounts of production and consumption data. This data can be leveraged to improve energy monitoring, automate prosumer management, provide remote control applications, and offer value-added services, all of which can reduce the barrier to the transition from consumer to prosumer (Kotilainen, 2019).

Moreover, different European nations are implementing smart meter rollouts at different rates and times (Zhou and Brown, 2017). Home energy management systems, or HEMS, control the available local energy resources, giving prosumers more insight into production and consumption data. New business models and advancements in the energy sector are largely made possible by digitalization. Although there is still a need for technological development in many sectors, prosumer technology is currently widely accessible and getting more inexpensive (Kotilainen, 2019).

Economic factors, such as the cost of the system (PV or otherwise), energy costs, rate structures, and possible financial advantages, are important factors for consumers to take into account before beginning prosumption. When making decisions on solar PV equipment or electric vehicles (EVs), for instance, consumers will consider the initial expenditure and estimated payback term. These are significant investments with protracted return times, even with their current price declines. Prosumers anticipate several financial benefits, such as lower energy bills and payment for the power they send into the grid, in addition to the cost-benefit analysis of their investment (Kotilainen, 2019).

## Prosumers: Benefits and Challenges.

Prosumers bring several advantages to the energy ecosystem. To begin with, engaging in consumer-driven activities increases the usage of distributed generation and renewable energy sources. Prosumers are new players in the energy markets. By lowering the need for more transmission and distribution capacity, enhancing local community resilience in the event of a central power system failure, and creating economic opportunities for both individuals and local communities, on-site power generation helps reduce losses that occur during transmission and distribution. Additionally, higher renewable energy sources use lowers emissions both locally and worldwide (Kotilainen, 2019).

Also, prosumers of energy are expected to play a variety of roles in sustainable development. The goal of sustainable development (SD) is to integrate social, environmental, and economic considerations in order to achieve long-term economic and environmental stability (e.g., Dernbach 2003). At least three approaches exist for prosumerism to support environmental sustainability: first, since most of the energy provided by prosumers is renewable, prosumerism promotes the use of RES. There are several obvious environmental advantages of DG, or microgeneration based on RES; arguably the most significant is that it contributes to a decrease in total greenhouse gas emissions. Global emissions may be prevented from rising by making investments in renewable energy sources (RES) to expand energy availability (WOE 2017).

In addition to lowering global emissions, cleaner energy decreases local particle emissions, which helps to lessen pollution-related health concerns. Second, prosumerism promotes greater energy efficiency: when energy is produced, stored, and used locally, less waste is generated. Better energy efficiency, in turn, reduces the need for energy generation (Kotilainen, 2019).

Nevertheless, prosumers emergence in the energy ecosystem presents some challenges. For instance, the energy system may face technological issues when the amount of intermittent energy supplied to the grid increases dramatically. Electricity networks have typically relied on centralized energy generation from massive power facilities that produce predictable results. The energy grid must maintain stability and balance in terms of supply and demand; hence predictability is critical in system management (Kotilainen, 2019).

Also, the emphasis has shifted to distributed generation and renewable energy sources because of stricter environmental restrictions making it less feasible to build traditional generation capacity. Because renewable energy sources are inherently intermittent, it is far more difficult to estimate their availability. Prosumer base growth can cause technical grid issues that affect quality and reliability in addition to load forecasting and capacity management. These issues include overvoltage, congestion, back-feeding into the circuit, stability, and system planning (Rickerson et al. 2014).

## Predicting Energy Behavior of Prosumers

Interest in energy demand forecasting has increased recently as attention has shifted to creating effective systems that can regulate and optimize energy production and use globally. This is due to the important environmental and economic effects of appropriately estimating energy consumption. Because electric energy cannot be stored for later use, it is now essential to generate the necessary amount of energy very instantly. To achieve the best buy prices in real-time on the global power market, a more precise hourly/daily prediction is very desirable (Petrican et al., 2018).

The energy sector is entering a new era with smart grids, which are intended to provide improved producer-consumer power generating system integration with significant advantages in the areas of security, economy, and the environment. An optimization of energy usage is essential for preserving the dependability of the power grid and preventing supply-demand mismatches since it involves two-way communication between the utility (sometimes referred to as the Distributed System Operator, or DSO) and its customers (Petrican et al., 2018). The Energy Consumption Prediction Module is an essential component of the system mentioned above as the DSO would base its adjustment choices on the module's output, with the decisions' quality being correlated with the accuracy of the predictions.

Numerous scholars have studied a number of energy demand forecasting models over the past few decades, at differing lengths of time ranging from extremely short to lengthy. Four categories have been identified by Srivatstava, Pandey, and Singh (2016): statistical methods (e.g., multiple regression (Moghram and Rahman, 1989), adoptive load forecasting, iterative reweighted least square, fuzzy logic (Liu et al, 1996), neural networks, etc.), artificial intelligence techniques (e.g., fuzzy logic, genetic algorithms, Srinivasan et al, 1998), and knowledge-based expert systems (Ho et al, 1990).

Recent research by Alani and Osunmakinde (2017) highlights the advantages and disadvantages of the most widely utilised STLF algorithms. Although a lot of work has gone into analysing previous STLF approaches using ARIMA (Autoregressive Integrated Moving Average), recently, more attention has been paid to evaluating neural networks (Yao, Zhang, and Cheng, 2000). While neural networks are better at capturing nonlinearities in the data, ARIMA operates under the premise that the observed and the future values in the time series are linearly linked. However certain variations of ARIMA models, Seasonal ARIMA (SARIMA) have been used in combinations with other models to perform energy load prediction: with SVM (Braun et al, 2014), with regression (Chikobvu and Sigauke, 2012), with neural networks (Yang, Wu, Chen, and Li, 2013).

Numerous research (Kankal et al., 2011; Ogcu, Demirel and Zaim, 2012; Birim and Tumturk, 2016; Kargar and Charsoghi, 2014) have employed neural networks to forecast energy consumption. Several authors claim that neural networks are more accurate in predicting energy use than more conventional techniques like regression or ARIMA. The authors of (Ekonomou, 2010) present a high-precision, short-term prediction model based on artificial neural networks that sends back a portion of its results. Another solution proposed in (Braun et al, 2014), where authors develop a residential building energy consumption model based on a Back Propagation ANN model, shows that the model is in line with the energy trends.

Support Vector Regression (SVR) has been suggested in some research as a workable substitute for artificial neural networks (ANNs) in the prediction of energy use (Campillo, Wallin, and Torstensson, 2012). In addition to energy consumption measures, Paudel et al. (2015) provide an SVM-based approach that makes use of climatic data and the building's approximate occupancy profile. Additionally, pertinent historical data is selected using a dynamic temporal warping method. The latter seems to retain a good accuracy level while saving a significant amount of training time. SVR models outperform ANNs because they can lessen the issue of becoming caught in a local minimum during training (Li, Gong, Li, and Sun 2005), all while maintaining high performance. In their comparison of SVM and neural network models for one-day ahead prediction, Fux et al. (2013) found that SVM-based models perform better overall since they require less parameters to be specified and are simpler to work with.

In the last years, deep neural networks gained more and more interest in the research area, being used in various domains, from computer vision to modeling and forecasting. A novel methodology for load prediction based on Deep Neural Networks is presented in (Marino, Amarasinghe, and Manic 2016). Two LSTM algorithms are compared: LSTM standard and LSTM-based Sequence to Sequence (S2S). Experiments performed at various time granularities show that the standard LSTM fails at one-minute time granularity, while the LSTM S2S performs well at both one minute and half hour time granularities. Since then, several papers have tackled the problem of STLF using this type of network (Zheng, Yuan, and Chen, 2017;

Kong et al., 2017; He, 2017; Persio and Honchar, 2017).

It seems that there is not a consensus regarding which prediction model is better, some claiming that complex regression models fail to beat the autoregressive model (Dagnely et al., 2015). Also, as already noted, some authors claim SVM as yielding the best results, while others claim that ANNs are better. This suggests that models’ accuracy is dependent on a variety of factors, an important one being the underlying dataset and its characteristics (Petrican et al., 2018).

## Related Work on Prosumers.

The rise in prosumerism has been accompanied by scholarships in energy literature. For example, Gajdziket et al (2023) explored the ways in which prosumer households who use photovoltaic panels and heat pumps could save energy. The study recommends the promotion of prosumers and encourages the use of environmentally friendly energy. Also, Rathnayaka et al (2011) introduced a new concept of managing the participation of prosumers as autonomous and intelligent goal-oriented virtual communities. According to the authors, prosumers functioning as a community could collectively increase the amount of power to be auctioned or bought thereby giving them bargaining power. The authors also studied the parameters influencing the prosumers energy behavior.

Brambati et al (2022), studied the factors that influence choosing to be a prosumer. The authors argue that several factors aid the transition from a tech driven approach to a consumer-driven approach throughout the emerging “prosumer business models.” According to the authors, factors such as concern for the environment, openness to sharing energy, impact on the cost of energy bill, and social compel people to be prosumers. Regarding the prediction of the energy behavior of prosumers, Zhoug et al (2021) points out the challenge in managing prosumers due to their diversity. The author proposes a multi-energy forecasting deep learning framework which simultaneously predicts gas, electricity, and thermal net load of local integrated energy systems. While this approach is sound, it fails to consider the computational power requirements for such frameworks.

A data professional could build a time series forecasting model from scratch and train the model and tweak its hyper-parameters to improve accuracy. However, training effective models is often challenging due to the computing resources needed. Not many companies or individuals have access to super computers with resources to be able to train highly effective models, especially when a large dataset is needed. Some corporations like OpenAI, Meta, Amazon and X have leveraged their access to powerful computers to build highly effective and scalable models such as GPT-4, LLAMA, Grok and Prophet, among others for data professionals to train with data. Hence, training existing models to make predictions has become popular among the data professional community.

This project contributes to the literature on prosumers by forecasting the energy production and consumption of prosumers to help energy retailers predict the demand and supply of energy to consumers on the grid.

Forecasting energy surplus or deficit of prosumers is a time series problem as it is influenced by time features. Therefore, this project utilizes Meta’s Prophet model. The Prophet model is able to decompose the components of the time series data- the trend, seasonality and datetime features. In addition, the model allows us to add other regressors that help accurately predict the target variable. Moreover, the model is fast to train and requires less computational power compared to other options like Neural Networks and Gradient Boosting Machines such as XGBoost and LightGBM. Furthermore, the Meta Prophet is robust against outliers. It obviates the need for data transformations such as Box-Cox to normalize the data.

# Methodology.

The aim of this project is to forecast the hourly production and consumption of energy prosumers. The goal is to help energy retailers to predict whether there is going to be an energy deficit or surplus by prosumers to be able to manage resources efficiently. To achieve this, the methodology below was used. This chapter discusses the source of the data being used, the various datasets and features in the data, the target variable and its statistic summary, the tools used for the project, Meta’s Prophet model, and Root Mean Squared Error, the evaluation metric for the models.

## Data Source

The data originates from Eesti Energia. Eesti Energia is a public limited energy company in

Estonia with its headquarters in [Tallinn](https://en.wikipedia.org/wiki/Tallinn). According to the website of the Eesti Energia, the company was founded in 1939. As of 2014, it operates in [Estonia](https://en.wikipedia.org/wiki/Estonia), [Latvia](https://en.wikipedia.org/wiki/Latvia), [Lithuania](https://en.wikipedia.org/wiki/Lithuania), [Finland](https://en.wikipedia.org/wiki/Finland), [Jordan](https://en.wikipedia.org/wiki/Jordan) and [Utah](https://en.wikipedia.org/wiki/Utah), United States. The corporation uses the brand name Enefit for international activities and Eesti Energia for operations within Estonia. The firm owns and operates mines in Eastern Estonia where oil shale, the primary raw material used in energy generation, is mined. The three primary business sectors of the Eesti Energia group are the production of shale oil, the sale and distribution of power, and the generating of electricity. The Estonian government is the owner of its shares.

The data was, however, posted on Kaggle, an online platform for learning and doing data science projects. The project uses 5 main datasets made available by the energy retailer- train.csv, gas\_prices.csv, client.csv, electricity.csv, historical.

The project splits the dataset into two main datasets- consumption dataset and production dataset. Both datasets have the same variables except that the targets differ: the consumption dataset targets the energy consumption of prosumers, and the production dataset targets the energy production of the prosumers. Thus, the difference between the energy produced and energy consumed would inform the energy retailer whether there is going to be an energy deficit or surplus.

The duration of the consumption dataset is ***2021-09-01 00:00:00* - *2023-05-29 23:00:00*,** and duration of the production dataset is ***2021-09-01 00:00:00*** - ***2023-05-29 23:00:00.*** The train energy production dataset starts from ***2021-09-01 00:00:00*** and ends at ***2022-11-20 03:00:00,*** while the test dataset starts at **2022-11-20 04:00:00 and ends at 2023-05-29 23:00:00**. Also, the train energy consumption dataset starts from ***2021-09-01 00:00:00*** and ends at ***2022-11-20 03:00:00,*** while the test dataset starts at ***2022-11-20 04:00:00*** and ends at ***2023-05-29 23:00:00.***

|  |  |
| --- | --- |
| Table 1 | **Training dataset.** |
| **Variable** | **Variable Description** |
| County | An ID code for the county. |
| Is\_business | Boolean for whether or not the prosumer is a business. |
| Product\_type | ID code with the following mapping of codes to contract types: {0: "Combined", 1: "Fixed", 2: "General service", 3: "Spot"}. |
| Is\_consumption | Boolean for whether or not this row's target is consumption or production.  datetime - The Estonian time in EET (UTC+2) / EEST (UTC+3). |
| Datetime | The Estonian time in EET (UTC+2) / EEST (UTC+3). |
| Data\_block\_id | All rows sharing the same data\_block\_id will be available at the same forecast time. This is a function of what information is available when forecasts are actually made, at 11 AM each morning. For example, if the forecast weather data\_block\_id for predictins made on October 31st is 100 then the historic weather data\_block\_id for October 31st will be 101 as the historic weather data is only actually available the next day. |
| Row\_id | A unique identifier for the row. |
| Prediction\_unit\_id | A unique identifier for the county, is\_business, |
| Target | The consumption or production amount for the relevant segment for the hour. The segments are defined by the county, is\_business, and product\_type |

|  |  |
| --- | --- |
| Table 2 | **This data set contains data on gas prices.** |
|  |  |
| Variable | Variable Description |
| Origin\_date | The date when the day-ahead prices became available. |
| Forecast\_date | The date when the forecast prices should be relevant. |
| [lowest/highest] \_price\_per\_mwh | The lowest/highest price of natural gas that on the day ahead market that trading day, in Euros per megawatt hour equivalent. |
| data\_block\_id | All rows sharing the same data\_block\_id will be available at the same forecast time. |

|  |  |
| --- | --- |
| Table 3 | **This dataset contains data on prosumers who are clients of Eesti Energia.** |
| **Variable** | **Variable description** |
| Product\_type |  |
| County | An ID code for the county. See county\_id\_to\_name\_map.json for the mapping of ID codes to county names. |
| Installed Capacity | Installed photovoltaic solar panel capacity in kilowatts. |
| Eic\_count | The aggregated number of consumption points (EICs - European Identifier Code). |
| Is\_business | Boolean for whether or not the prosumer is a business. |
| date |  |
| Data\_block\_id | All rows sharing the same data\_block\_id will be available at the same forecast time. |

|  |  |
| --- | --- |
| Table 4 | **This dataset contains dataset on electricity prices.** |
| Variable | Variable Description |
| Origin\_date | Date the electricity price prediction was made |
| Forecast\_date | Date which the forecasted price takes effect |
| Euros\_per\_hour | The price of electricity on the day ahead markets in euros per megawatt hour. |
| Data\_block\_id | All rows sharing the same data\_block\_id will be available at the same forecast time. |

|  |  |
| --- | --- |
| Table 5 | **This dataset contains weather variables.** |
| datetime | Datetime when weather was recorded |
| temperature | Temperature recorded |
| dewpoint | Dewpoint recorded |
| rain | The rain from large scale weather systems of the preceding hour in millimeters. |
| snowfall | Different from the forecast conventions. Snowfall over the preceding hour in centimeters |
| Surface\_pressure | The air pressure at surface in hectopascals. |
| cloudcover | Different from the forecast conventions. Cloud cover at 0-3 km, 3-8, 8+, and total. |
| Windspeed\_10m | Different from the forecast conventions. The wind speed at 10 meters above ground in meters per second. |
| Winddirection\_10m | Different from the forecast conventions. The wind direction at 10 meters above ground in degrees. |
| Shortwave\_radiation | Different from the forecast conventions. The global horizontal irradiation in watt-hours per square meter. |
| Direct\_solar\_radiation | Direct solar radiation recorded |
| Diffuse\_radiation | Different from the forecast conventions. The diffuse solar irradiation in watt-hours per square meter. |
| [latitude/longitude] | The coordinates of the weather station. |
| Data\_block\_id | All rows sharing the same data\_block\_id will be available at the same forecast time. |

**Steps**

Data Collection from Kaggle

Import and merge Datasets

Clean the data.

Separate the datasets into consumption and production datasets.

Separate the datasets into training and testing datasets.

Instantiate Prophet model.

Train model with train dataset

Test model with test data.

Measure model accuracy by comparing actual data with predicted data.

## Target- Total Energy Produced/Consumed by Energy Prosumers.

The target of the datasets is the total energy produced and consumed by energy prosumers in the Balkans who are clients of Eesti Energia. The clients span multiple counties, are either businesses or households, and have varied capacities of Solar VPs.

### Summary Statistics of the Target Datasets

|  |  |  |
| --- | --- | --- |
| **Statistic** | **Total Energy Produced** | **Total Energy Consumed** |
| count | 15264 | 15264 |
| mean | 5800.851650 | 30366.083487 |
| std | 12352.418858 | 9960.217975 |
| min | 0.000000 | 0.000000 |
| 1st Quartile | 42.796750 | 23725.398250 |
| 2nd Quartile | 108.764500 | 28928.608500 |
| 3rd Quartile | 3806.695500 | 36060.478750 |
| max | 89930.586000 | 63354.178000 |

## Tools and Libraries Used

1. Python Programming Language as the script language for the project
2. Jupyter Notebook for scripting the code
3. Pandas for organization of the dataset as a datafame and for analysis.
4. Numpy for statistical analysis.
5. Matplotlib for visualization.
6. Seaborn for visualization.
7. Statsmodels for autoregressive and partial autoregressive analysis.
8. FeatureEngine for extracting datetime features.
9. Prophet Model for model
10. mse from Sklearn.metrics to measure the mean squared error of the model.

## Meta Prophet Model.

The Prophet is an algorithm used to estimate time series data developed by Meta, which is capable of predicting long-term unstable trends or unseasonable data or processing missing data. Since Prophet is developed with R software and Python with open-source code, users can use this estimation model in time series problems by making changes in parameters. The algorithm of the model is:

𝑦(𝑡) = 𝑔(𝑡) +𝑠(𝑡) +ℎ(𝑡)+ℇ𝑡

Where ℇ𝑡 is the error term that is expected to be normally distributed, g(t) represents a trend, s(t) shows seasonality term and h(t) symbolizes holidays.

The model accepts one feature, datetime and the target variable. However, the features have to be renamed for the model to work- datetime feature as ***ds***and target feature as ***y.*** The model allows you to add additional features that could predict the target.

For this project, the Prophet Model would be trained to predict the energy production and consumption of prosumers.

## Model Evaluation Metric: Root Mean Square Error (RMSE).

The project uses RMSE to evaluate the trained models. RMSE is the squared root of the average of the squared differences between predicted and observed values, which in this case is the energy consumption and production of prosumers. The RMSE provides a measure of how spread out the errors are in the predicted values, with lower values indicating better model performance (Hodson, 2022).

The project uses RMSE due to several reasons. To begin with, RMSE has the advantage of being interpretable in the same unit as the target variable (e.g., kilowatt-hours for energy consumption). This makes it easy for stakeholders to understand the magnitude of the forecasting errors in practical terms. Secondly, RMSE penalizes large forecasting errors more heavily than smaller errors, which can be desirable since large errors have significant consequences, for example, financial losses due to inaccurate energy forecasts. Finally, RMSE is relatively robust to outliers in the dataset due to the squaring operation, which reduces the influence of extreme values on the overall error (Hodson, 2022).

## Calculating the RMSE.

*Step 1. Residual​=Observed Value−Predicted Value*

*Step 2. Squared Residual=(Residual​)2*

*Step 3. Mean Squared Error (MSE)=n/1​ ∑ni=1 ​(Squared Residuali​)*

*Step 4. RMSE=√MSE*

# Findings: Analyzing Datasets and Model Building.

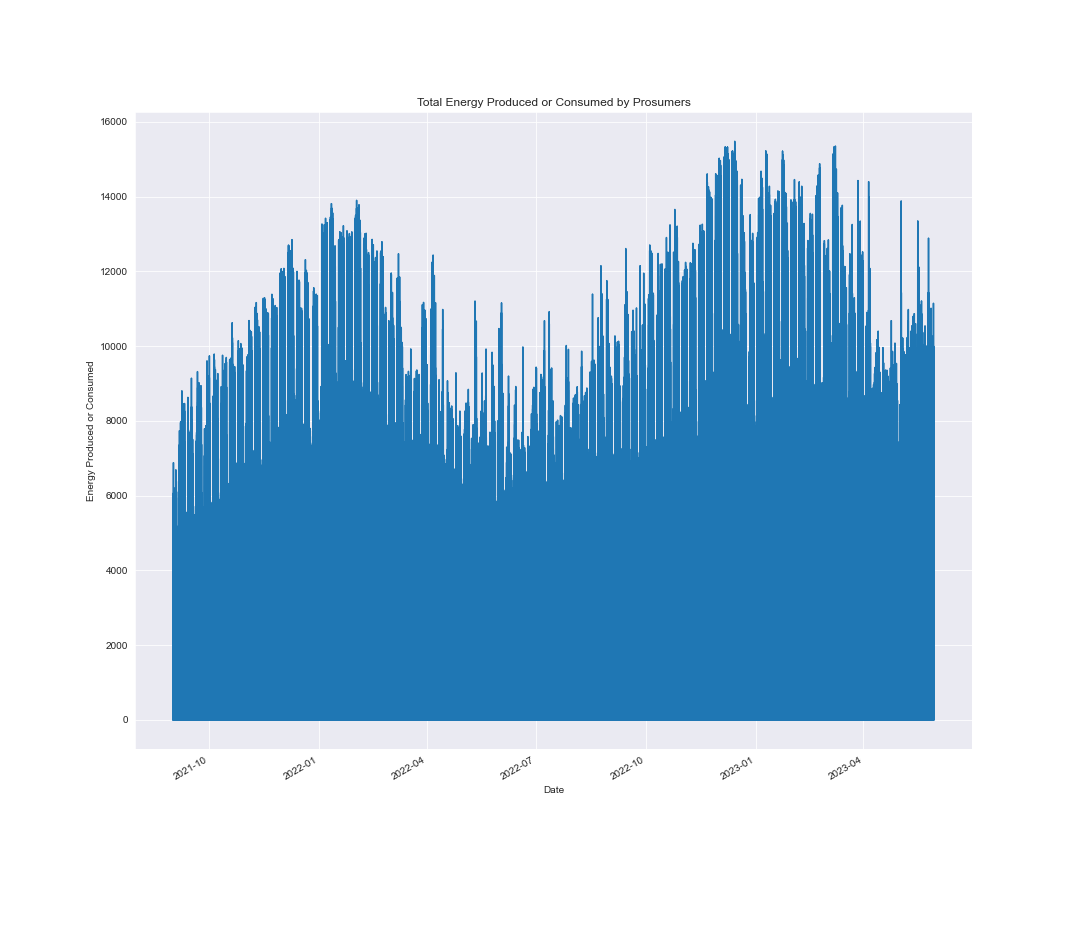
Data Distribution of Total Energy Produced or Consumed by Prosumers**.**

Fig. 1

The diagram shows the distribution of total energy produced or consumed by Prosumers. Observable pattern is noticeable. From July, there is a gradual increase in energy produced or consumed by prosumers. Right about February, the energy produced or consumed by prosumers starts to drop. This data distribution provides a general pattern of energy behavior of prosumers. The analysis below provides a detailed picture of the various datetime features and external features that predict the energy behavior of prosumers.

## Analyzing Energy Production Data

This section of the project makes a detailed analysis of the data concerning energy production of prosumers. The aim is to explore patterns in the energy behavior of prosumers.

### A graph showing a number of data Description automatically generated with medium confidenceDistribution Total Energy Produced by Prosumers.

Fig. 2

The chart above shows the distribution of total energy produced by prosumers from 2021-09-01 00:00:00 - 2023-05-29 23:00:00. As can be seen in the graph, there is an observable pattern in the production of energy. Energy production starts to increase gradually in March, hits its peak in July, and gradually starts to drop in August. Energy production is at its lowest in December to February. This is not surprising since that marks the Winter season.

### Total Energy Produced by Prosumers Per Month.

Fig. 3

The illustration above shows the energy production of prosumers per month. The graph shows that much of the energy is produced between April and July, while energy production drops from August to March. This corresponds to the summer season when there is much sunshine, and winter when the sun does not shine enough. Hence, month could predict energy production of prosumers, and thus a feature for the model.

### Total Energy Produced by Prosumers Per Day of Week.

Fig. 4

The graph above shows energy produced per day of the week of prosumers. As can be seen on the chart, the energy produced increases from Thursday and peaks on Sunday, then drops on Monday. Hence, month could predict energy production of prosumers, and thus a feature for the model.

### Total Energy Produced by Prosumers Per Hour.

Fig. 5

The graph above shows the energy produced by prosumers per hour of the day. The distribution of the data shows that the energy produced is at its lowest before sunrise (5:00am) and after sunset (8:00pm), while energy production starts to rise at 6:00am, hits its peak at noon, then gradually begins to decrease in the afternoon. The bell curve of the chart indicates that much of the energy produced occurs during the day, which makes sense since the prosumers in the dataset use Solar VP and rely on the sun to generate energy.

### Interaction between Solar Radiation and Installation Capacity and its Relationship to Energy Production.

Fig. 6

The chart above shows the impact of an interaction between the installation capacity of the Solar PV and direct solar radiation from the sun on the total energy produced by prosumers. The analysis shows a positive relationship between the energy produced and the interaction between the installation capacity of the Solar PV and direct solar radiation from the sun. Accordingly, the project adds this as an external feature for the model.

### Partial Autocorrelation of Energy Produced.

Fig. 7

The graph above is a diagram showing partial autocorrelation of energy produced by prosumers. In time series analysis, the partial autocorrelation function (PACF) measures the correlation between observations at two time points while controlling for the effect of other data points between them. It helps to identify the direct relationship between a specific time point and its lagged values, removing the influence of intermediate time points (Mestre et al., 2021). In this case, the PACF shows hourly time point of energy produced by prosumers prior to the current observed target value *(t)*.

The graph above shows the correlation between the observed time (t) and lag 1 to lag 42. As can be seen, lag 1 and lag 2 have the strongest correlation. Hence, lag 1 and lag 2 were considered as features for the model.

### Relationship between Lag 1 and Total Energy Produced by Prosumers.

Fig. 8

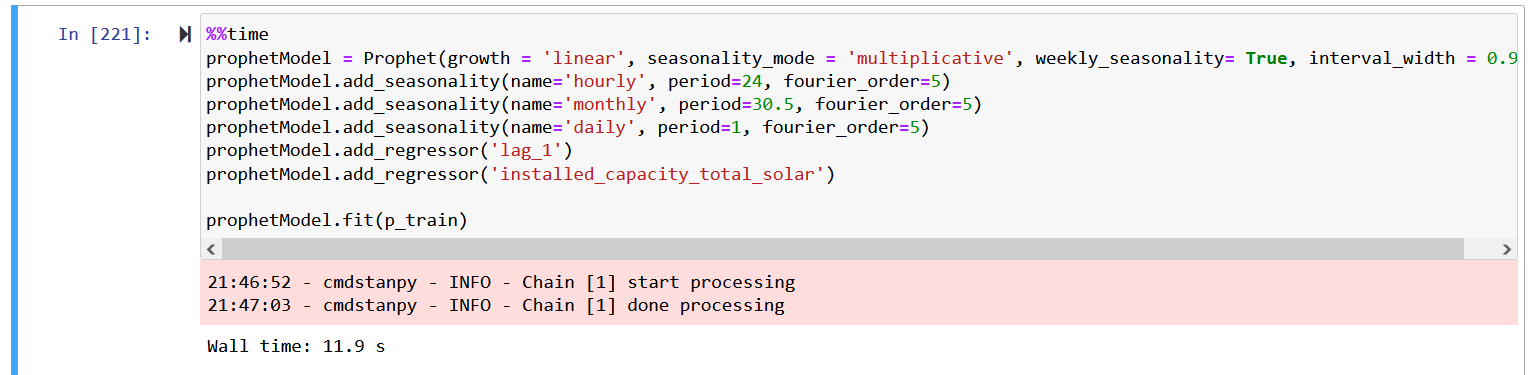
As indicated in the Partial Auto-Correlation diagram, lag 1 (t-1) of the observed value (t) has a strong relationship with the target (t). The correlation plot above confirms the observation from the Partial Auto-Correlation plot.

## Modelling Total Energy Produced

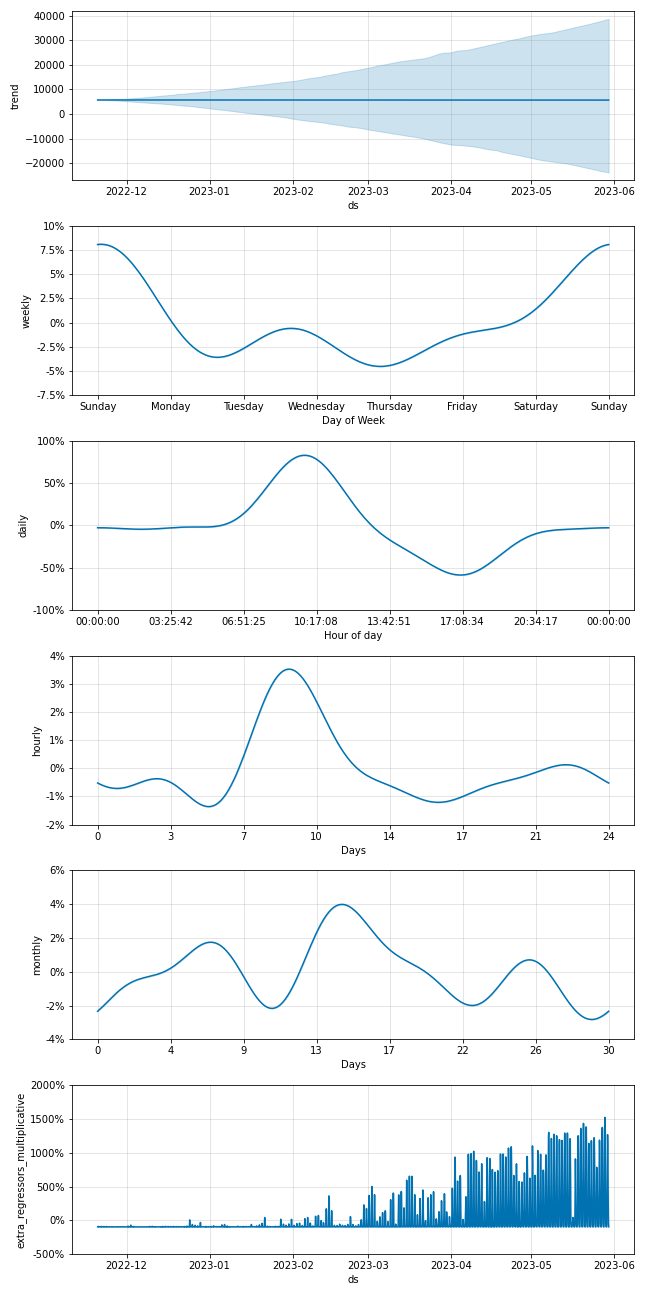
This section discusses the model building process for the total energy produced by prosumers. The project uses the Prophet model developed by Meta. The Prophet model has several hyperparameters, of which four were tuned for the training of the model- *growth, seasonality\_mode, weekly\_seasonality and interval\_width*. The growth parameter specifies the growth type of the time series. In this case, *'linear'* is chosen, indicating that the growth trend will be modeled as linear. The seasonality mode determines how seasonality is modeled. The value *'multiplicative'* suggests that seasonality will be treated as multiplicative. This means that the seasonal component will be multiplied by the overall trend to capture the seasonal fluctuations.

To add weekly seasonality, the model provides the *weekly\_seasonality* which accepts a Boolean value. The project activates the weekly seasonality and sets the confidence interval width to 95%. The Prophet model allows users to add additional seasonality. Accordingly, due to the datetime pattern observed in the data analysis section, the project adds hourly, monthly, and daily seasonality.

The *“.add\_regressor”* method of the model allows users to add external non-datetime features. Based on the exploratory data analysis, lag 1 and the interaction between installation capacity and solar radiation were added as additional features.



### The Model’s Decomposition of the Components of the Energy Production Distribution of Prosumers.



The model decomposes the various components of the dataset. In the chart above, one could observe that the model breaks down the distribution of the energy produced by prosumers into hourly, daily, weekly, and monthly seasonality, in addition to extra regressors that were added.

The model was able to capture the datetime seasonality patterns similar to the datetime feature patterns that were manually done. This makes the model a powerful tool for time series projects.

Fig. 9

### Comparing Model Prediction on Energy Production to Actuals.

Fig. 10

The model was used to test the test dataset for the energy production of prosumers which starts from **2022-11-20 04:00:00 and ends at 2023-05-29 23:00:00*.*** This was compared to the actual values. The predicted values are in blue while the actual values are in red. The black values are the previous values.

As can be seen, the model effectively predicts the values. The Root Mean Squared Error for the actual values and the predicted values is **3056.88 (***rounded to two decimal points***).**

## A graph showing a blue line Description automatically generatedDistribution Total Energy Consumed by Prosumers.

Fig. 11

The chart above shows the distribution of total energy consumed by prosumers from 2021-09-01 00:00:00 - 2023-05-29 23:00:00. As can be seen in the graph, there is an observable pattern in the production of energy. Energy consumption starts to rise gradually in August, hits its peak in January, and gradually starts to drop in February. Energy consumption is at its lowest in June to July. This is not surprising since that marks the summer season when people use less heat due to the high temperature.

### Distribution of Total Energy Consumed by per Month Prosumers.

Fig. 12

The illustration above shows the energy consumption of prosumers per month. The graph shows that less energy is consumed between April and July, while energy consumption starts to increase from August to March. Energy consumption is at its peak in January and February. This corresponds to the summer season when there is much sunshine, and winter when the sun does not shine enough. Hence, month could predict energy consumption of prosumers, and thus a feature for the model.

### Energy Produced per Day of Week by Prosumers.

Fig. 13

The graph above shows energy produced per day of the week of prosumers. As can be seen on the chart, the energy consumed is high from Monday to Friday, then drops on Saturday. Hence, month could predict energy production of prosumers, and thus a feature for the model.

### Distribution of Total Energy Consumed per Month by Prosumers.

Fig. 14

The graph above shows the energy produced by prosumers per hour of the day. The distribution of the data shows that the energy consumed is at its lowest before sunrise (5:00am). Energy consumption is at its peak during early morning, then drops during late morning. It starts to steadily increase in the afternoon and starts to drop in the evening after 8:00pm.

### Partial Autocorrelation of Energy Consumed.

Fig. 15

The graph above is a diagram showing partial autocorrelation of energy produced by prosumers. As explained earlier, the partial autocorrelation function (PACF) measures the correlation between observations at two time points while controlling for the effect of other data points between them. It helps to identify the direct relationship between a specific time point and its lagged values, removing the influence of intermediate time points (Mestre et al., 2021). In this case, the PACF shows the hourly time point of energy consumed by prosumers prior to the current observed target value *(t)*.

The graph above shows the correlation between the observed time (t) and lag 1 to lag 42. As can be seen, lag 1 and lag 2 have the strongest correlation. Hence, lag 1 and lag 2 were considered as features for the model.

### Relationship between Lag 1 and Total Energy Consumed by Prosumers.

Fig. 16

As indicated in the Partial Auto-Correlation diagram, lag 1 (t-1) of the observed value (t) has a strong relationship with the target (t). The correlation plot above shows a correlation between lag 1 and energy consumed at observed time (t) confirms the observation from the Partial Auto-Correlation plot.

### Relationship between Lag 2 and Total Energy Consumed by Prosumers.

Fig. 17

The correlation plot above shows a correlation between lag 2 and energy consumed at observed time (t) confirms the observation from the Partial Auto-Correlation plot.

## Modelling Total Energy Consumed.

This section discusses the model building process of the total energy consumed by prosumers. The project uses the Prophet model developed by Meta. The following hyperparameters were tuned- *growth, seasonality\_mode, weekly\_seasonality and interval\_width*. The growth parameter specifies the growth type of the time series. In this case, *'linear'* is chosen, indicating that the growth trend will be modeled as linear. The seasonality mode determines how seasonality is modeled. The value *'multiplicative'* suggests that seasonality will be treated as multiplicative. This means that the seasonal component will be multiplied by the overall trend to capture the seasonal fluctuations.

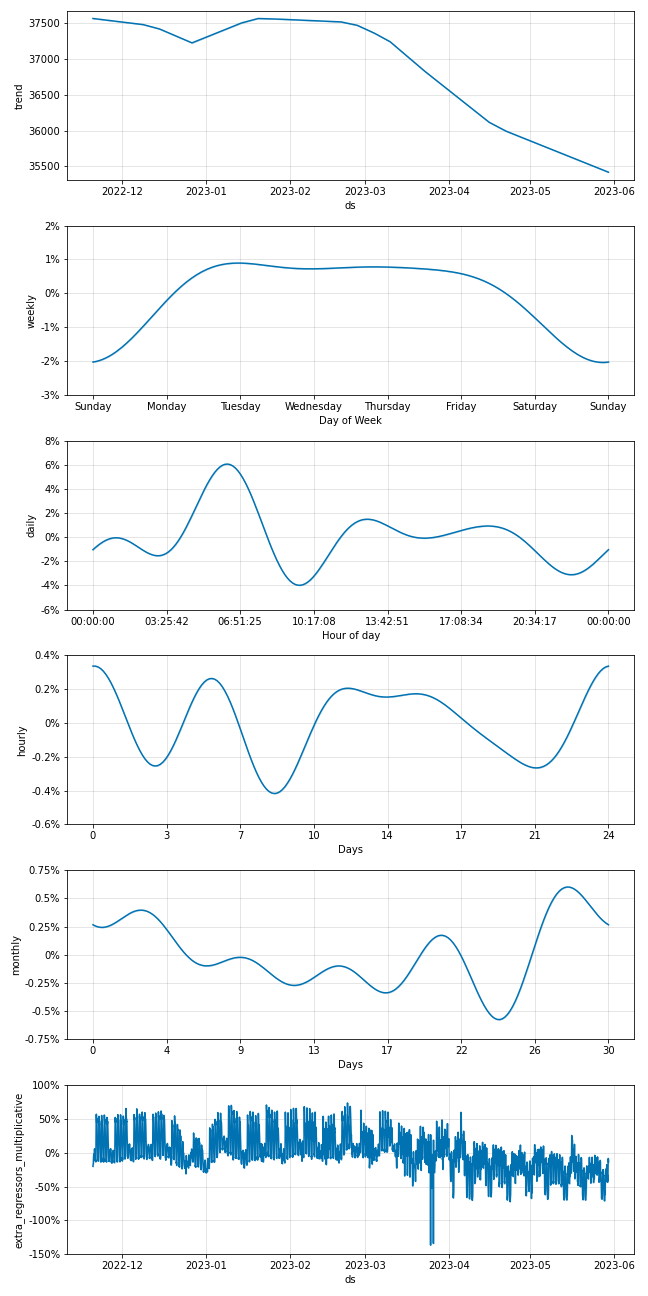
To add weekly seasonality, the model provides the *weekly\_seasonality* which accepts a Boolean value. The project activates the weekly seasonality and sets the confidence interval width to 95%. The Prophet model allows users to add additional seasonality. Accordingly, due to the datetime pattern observed in the data analysis section, the project adds hourly, monthly, and daily seasonality.

The *“.add\_regressor”* method of the model allows users to add external non-datetime features. Based on the exploratory data analysis, lag 1 and lag 2 were added as additional regressors.

A screenshot of a computer program

Description automatically generated

### The Model’s Decomposition Components of the Energy Consumption of Prosumers.

The model decomposes the various components of the dataset. In the chart above, one could observe that the model breaks down the distribution of the energy consumed by prosumers into hourly, daily, weekly, and monthly seasonality, in addition to extra regressors that were added.

The model was able to capture the datetime seasonality patterns similar to the datetime feature patterns that were manually done.

Fig. 18

### Comparing Model Prediction to Actuals.

Fig. 19

The model was used to test the test dataset for the energy consumption of prosumers which starts from **2022-11-20 04:00:00 and ends at 2023-05-29 23:00:00*.*** This was compared to the actual values. The predicted values are in blue while the actual values are in red. The black values are the previous values.

As can be seen, the model effectively predicts the values. The Root Mean Squared Error for the actual values and the predicted values is **2468.86 (***rounded to two decimal points***).**

## Predicting Energy Deficit or Surplus

Fig. 20

The diagram above predicts the energy surplus or deficit of prosumers. It is calculated by finding the difference between predicted energy produced by prosumers and the energy they consumed. The blue colored data points represent deficit, while the orange-colored data points represent surplus. From the diagram, it could be observed that during summer season, some prosumers are able to generate surplus energy while in the winter and early spring, all prosumers have energy deficit.

## Discussion and Conclusion

The goal of this project is to help energy retailers predict energy deficit or surplus of prosumers. Hence, this project uses Meta’s prophet model to predict hourly production and consumption of energy by prosumers. The difference between the predicted hourly production and consumption of energy was calculated to find energy surplus or deficit by prosumers. Consistent with the literature, for example Sun, Haghighat and Fung (2020) and Gonzalez-Vidal, Jimenez and Gomez-Skarmeta (2019), date time features like hour of day, day of week, and month of the year were observed to impact total energy produced or consumed by prosumers.

Also, lag features, precisely lag 1 (t-1) was found to predict both the energy produced and consumed by prosumers. Concerning energy production, there is a strong correlation between an interaction of installation capacity and total solar radiation and total energy produced. This is confirmed by authors like Olomiyesan et al., (2015) and Myers (2017).

The project substantiated the power of the Prophet model by Meta as it was able to capture the datetime feature patterns similar to the datetime patterns that were manually revealed from the dataset. Moreover, the model’s capability to take extra regressors beyond datetime features makes it far more useful. In this project, external features such as lags, solar radiation and installed capacity of solar PV were added to the model. This improved the accuracy of the model. Furthermore, the Prophet model is fast, and easy to use. It took 11.9 seconds and 2.4 seconds to train about 12,000 rows of data each for the energy production and consumption respectively, by prosumers.

Equipping data professionals of energy retailing companies to accurately predict energy surplus or deficit of prosumers, and their general costumers comes with several advantages, and these are highlighted in the literature. For example, knowing when prosumers will generate surplus energy allows energy retailers to better manage the electricity grid. They can anticipate periods of excess energy production and adjust grid operations, accordingly, potentially reducing wastage or the need for expensive backup power generation. More importantly, energy retailers can allocate resources more efficiently. They can adjust their supply-demand balance by incentivizing prosumers to feed surplus energy back into the grid during peak demand periods, reducing the strain on traditional energy sources.

Also, prosumers often generate renewble energy, such as solar or wind power. By accurately predicting surplus energy from these sources, energy retailers can facilitate the integration of renewable energy into the grid more effectively, supporting sustainability goals and reducing reliance on fossil fuels.

## Conclusion

In summary, this study shows how well Meta's Prophet model can forecast energy excess or deficit among prosumers by using information like lag features and datetime patterns, as well as external variables like solar radiation and installation capacity. By including further regressors and capturing these patterns, the Prophet model becomes much more accurate and useful. The literature lists a number of benefits of providing energy retailers with the capacity to precisely predict prosumer energy dynamics, including enhanced grid management, effective resource allocation, and more integration of renewable energy sources. In the end, this supports larger environmental and economic objectives by strengthening the energy ecosystem's resilience and sustainability.

**References**

AbuAl-Foul, B. M. (2012). Forecasting energy demand in Jordan using artificial neural networks. *Topics in Middle Eastern and North African Economies*, *14*.

Alani, A. Y., & Osunmakinde, I. O. (2017). Short-term multiple forecasting of electric energy loads for sustainable demand planning in smart grids for smart homes. *Sustainability*, *9*(11), 1972.

Astarloa B, Kaakeh A, Lombardi M, Scalise J (2017) The future of electricity: new technologies transforming the grid edge. World Economic forum, p 32

Birim, Ş., & Tümtürk, A. (2016). Modeling and forecasting Turkey’s electricity consumption by using artificial neural network. *American Scientific Research Journal for Engineering, Technology, and Sciences*, *25*(1), 192-208.

Brambati, F., Ruscio, D., Biassoni, F., Hueting, R., & Tedeschi, A. (2022). Predicting acceptance and adoption of renewable energy community solutions: the prosumer psychology. *Open Research Europe*, *2*.

Braun, M., Bernard, T., Piller, O., & Sedehizade, F. (2014). 24-hours demand forecasting based on SARIMA and support vector machines. *Procedia Engineering*, *89*, 926-933.

Byrne J, Taminiau J, Kurdgelashvili L, Kim KN (2015) A review of the solar city concept and methods to assess rooftop solar electric potential, with an illustrative application to the city of Seoul. Renew Sust Energ Rev 41:830–844. https://doi.org/10.1016/j.rser.2014.08.023

Campillo, J., Wallin, F., Torstensson, D., & Vassileva, I. (2012). Energy demand model design for forecasting electricity consumption and simulating demand response scenarios in Sweden. In *4th International Conference in Applied Energy 2012, July 5-8, 2012. Suzhou, China.*

Chikobvu, D., & Sigauke, C. (2012). Regression-SARIMA modelling of daily peak electricity demand in South Africa. *Journal of Energy in Southern Africa*, *23*(3), 23-30.

Cho, M. Y., Hwang, J. C., & Chen, C. S. (1995, November). Customer short term load forecasting by using ARIMA transfer function model. In *Proceedings 1995 International Conference on Energy Management and Power Delivery EMPD'95* (Vol. 1, pp. 317-322). IEEE.

Dagnely, P., Ruette, T., Tourwé, T., Tsiporkova, E., & Verhelst, C. (2015). Predicting hourly energy consumption. Can regression modeling improve on an autoregressive baseline?. In *Data Analytics for Renewable Energy Integration: Third ECML PKDD Workshop, DARE 2015, Porto, Portugal, September 11, 2015. Revised Selected Papers 3* (pp. 105-122). Springer International Publishing.

Dernbach JC (2003) Achieving sustainable development: the centrality and multiple facets of integrated decisionmaking. Indiana J Glob Leg Stud 10:247–285

Di Persio, L., & Honchar, O. (2017, November). Analysis of recurrent neural networks for short-term energy load forecasting. In *AIP Conference Proceedings* (Vol. 1906, No. 1). AIP Publishing.

Ekonomou, L. (2010). Greek long-term energy consumption prediction using artificial neural networks. *Energy*, *35*(2), 512-517.

European Commission (2016) Proposal for a directive of the European Parliament and of the council on common rules for the internal market in electricity’ (Proposal for the amendment of the electricity directive)

Fux, S. F., Benz, M. J., Ashouri, A., & Guzzella, L. (2013). Short-term thermal and electric load forecasting in buildings. In *CISBAT (International Conference on Clean-tech for Smart Cities & Buildings: From Nano to Urban Scale)*.

Gajdzik, B., Jaciow, M., Wolniak, R., Wolny, R., & Grebski, W. W. (2023). Energy Behaviors of Prosumers in Example of Polish Households. *Energies*, *16*(7), 3186.

Gonzalez-Vidal, A., Jimenez, F., & Gomez-Skarmeta, A. F. (2019). A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. *Energy and Buildings*, *196*, 71-82.

He, W. (2017). Load forecasting via deep neural networks. *Procedia Computer Science*, *122*, 308-314.

Ho, K. L., Hsu, Y. Y., Chen, C. F., Lee, T. E., Liang, C. C., Lai, T. S., & Chen, K. K. (1990). Short term load forecasting of Taiwan power system using a knowledge-based expert system. *IEEE Transactions on Power Systems*, *5*(4), 1214-1221.

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, *2022*, 1-10.

IEA-RETD, 2014. Residential Prosumers - Drivers and Policy Options (RE-Prosumers), Renewable Energy Technology Deployment. https://doi.org/10.2172/ 1163237

Jacobs SB (2017) The energy prosumer. Ecol Law Q 43:519–579. <https://doi.org/10.15779/Z38XS02>

Kampman B, Blommerde J, Afman M (2016) The potential of energy citizens in the European Union. CE Delft

Kandananond, K. (2011). Forecasting electricity demand in Thailand with an artificial neural network approach. *Energies*, *4*(8), 1246-1257.

Kankal, M., Akpınar, A., Kömürcü, M. İ., & Özşahin, T. Ş. (2011). Modeling and forecasting of Turkey’s energy consumption using socio-economic and demographic variables. *Applied Energy*, *88*(5), 1927-1939.

Kargar, M. J., & Charsoghi, D. K. (2014). Predicting annual electricity consumption in Iran using artificial neural networks (NARX). *Indian J. Sci. Res*, *5*(1), 231-242.

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE transactions on smart grid*, *10*(1), 841-851.

Kotilainen, 2019

Kotilainen, K. (2019). Energy prosumers’ role in the sustainable energy system. In *Affordable and Clean Energy* (pp. 1-14). Cham: Springer International Publishing.

Li, X., Gong, D., Li, L., & Sun, C. (2005). Next day load forecasting using SVM. In *Advances in Neural Networks–ISNN 2005: Second International Symposium on Neural Networks, Chongqing, China, May 30-June 1, 2005, Proceedings, Part III 2* (pp. 634-639). Springer Berlin Heidelberg.

Liu, K., Subbarayan, S., Shoults, R. R., Manry, M. T., Kwan, C., Lewis, F. I., & Naccarino, J. (1996). Comparison of very short-term load forecasting techniques. *IEEE Transactions on power systems*, *11*(2), 877-882.

Marino, D. L., Amarasinghe, K., & Manic, M. (2016, October). Building energy load forecasting using deep neural networks. In *IECON 2016-42nd annual conference of the IEEE industrial electronics society* (pp. 7046-7051). IEEE.

Martin R (2015) Renewable energy trading launched in Germany [WWW document]. MIT Technol Rev. [https://www.technologyreview.com/s/544471/renew able-energy-trading-launched-in-germany/](https://www.technologyreview.com/s/544471/renew%20able-energy-trading-launched-in-germany/)

Masera D, Couture T (2015) Industrial prosumers of renewable energy – contribution to inclusive and sustainable industrial development. United Nations Industrial Development Organization (UNIDO), Energy Branch, Renewable and Rural Energy Unit (RRE). Vienna, Austria

Mestre, G., Portela, J., Rice, G., San Roque, A. M., & Alonso, E. (2021). Functional time series model identification and diagnosis by means of auto-and partial autocorrelation analysis. *Computational statistics & data analysis*, *155*, 107108.

Moghram, I., & Rahman, S. (1989). Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on power systems*, *4*(4), 1484-1491.

Myers, D. R. (2017). *Solar radiation: practical modeling for renewable energy applications*. CRC press.

Oğcu, G., Demirel, O. F., & Zaim, S. (2012). Forecasting electricity consumption with neural networks and support vector regression. *Procedia-Social and Behavioral Sciences*, *58*, 1576-1585.

Olomiyesan, B. M., Oyedum, O. D., Ugwuoke, P. E., Ezenwora, J. A., & Ibrahim, A. G. (2015). Solar energy for power generation: a review of solar radiation measurement processes and global solar radiation modelling techniques.

Paudel, S., Nguyen, P. H., Kling, W. L., Elmitri, M., Lacarrière, B., & Corre, O. L. (2015). Support vector machine in prediction of building energy demand using pseudo dynamic approach. *arXiv preprint arXiv:1507.05019*.

Petrican, T., Vesa, A. V., Antal, M., Pop, C., Cioara, T., Anghel, I., & Salomie, I. (2018, September. Evaluating forecasting techniques for integrating household energy prosumers into smart grids. In *2018 IEEE 14th international conference on intelligent computer communication and processing (ICCP)* (pp. 79-85). IEEE.).

Rathnayaka, A. D., Potdar, V. M., Hussain, O., & Dillon, T. (2011, December). Identifying prosumer's energy sharing behaviours for forming optimal prosumer-communities. In *2011 International Conference on Cloud and Service Computing* (pp. 199-206). IEEE.

Rickerson W, Couture T, Barbose G, Jacobs D, Parkinson G, Chessin E, Belden A, Wilson H, Barrett H (2014) Residential prosumers: drivers and policy options (reprosumers). https://doi.org/10.2172/1163237

Sajn N (2016) Electricity ‘prosumers’. European Parliamentary Research Service (EPRS). Brussels, Belgium

Srinivasan, D. (1998). Evolving artificial neural networks for short term load forecasting. *Neurocomputing*, *23*(1-3), 265-276.

Srivastava, A. K., Pandey, A. S., & Singh, D. (2016, March). Notice of Violation of IEEE Publication Principles: Short-term load forecasting methods: A review. In *2016 International conference on emerging trends in electrical electronics & sustainable energy systems (ICETEESES)* (pp. 130-138). IEEE.

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.

Toffler A (1981) The third wave. Bantam Books, New YorkFord et al. 2016

WOE (2017) From poverty to prosperity, World Energy Outlook special report. Organization for Economic Cooperation and Development, International Energy Agency, IEA

Yang, Y., Wu, J., Chen, Y., & Li, C. (2013, January). A new strategy for short-term load forecasting. In *Abstract and applied analysis* (Vol. 2013). Hindawi.

Yao, S. J., Song, Y. H., Zhang, L. Z., & Cheng, X. Y. (2000). Wavelet transform and neural networks for short-term electrical load forecasting. *Energy conversion and management*, *41*(18), 1975-1988.

Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies*, *10*(8), 1168.

Zhou S, Brown MA (2017) Smart meter deployment in Europe: a comparative case study on the impacts of national policy schemes. J Clean Prod 144:22–32. <https://doi.org/10.1016/j.jclepro.2016.12.031>

Zhou, B., Meng, Y., Huang, W., Wang, H., Deng, L., Huang, S., & Wei, J. (2021). Multi-energy net load forecasting for integrated local energy systems with heterogeneous prosumers. *International Journal of Electrical Power & Energy Systems*, *126*, 106542.