FORECASTING AND REINFORCEMENT LEARNING STRATEGIES FOR EFFICIENT ENERGY EXCHANGE IN PEER-TO-PEER ENERGY TRADING GAME AMONG NANO/MICROGRIDS: EMPIRICAL ANALYSIS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

RABİA ŞEYMA GÜNEŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

SEPTEMBER 2023

Approval of the thesis:

FORECASTING AND REINFORCEMENT LEARNING STRATEGIES FOR EFFICIENT ENERGY EXCHANGE IN PEER-TO-PEER ENERGY TRADING GAME AMONG NANO/MICROGRIDS: EMPIRICAL ANALYSIS

Submitted by RABİA ŞEYMA GÜNEŞ in partial fulfillment of the requirements for the degree of **Master of Science in Information Systems Department, Middle East Technical University** by,

Date: _*07.09.2023*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : **Rabia Şeyma Güneş**

Signature :

ABSTRACT

FORECASTING AND REINFORCEMENT LEARNING STRATEGIES FOR EFFICIENT ENERGY EXCHANGE IN PEER-TO-PEER ENERGY TRADING GAME AMONG NANO/MICROGRIDS: EMPIRICAL ANALYSIS

Güneş, Rabia Şeyma
MSc., Department of Information Systems
Supervisor: Prof. Dr. Sevgi Özkan Yıldırım

September 2023, 88 pages

New technologies included in distributed energy systems have created solutions that allow the management of demand and generation variability in the electricity grid and the costs arising from this variability. Trade between these small grids has enabled the sale of excess energy between each other and the purchase of needed energy, thus reducing costs and system constraints. The purpose of this trade is modeled as a game of agents mentioned in reinforcement learning, enabling the creation of the market that offers those benefits from each peer. Each peer provides its electricity demand with both internal resources and other peers. The aim of this thesis is to comply with system constraints while providing the demand of each peer in this game aiming at maximum benefit. A novel Multi-Agent Reinforcement Learning model to facilitate very short-term energy trading among peers is suggested in this thesis. The key contributions of this thesis lie in incorporating very short-term load, generation, and price forecasts into the framework to enable more accurate decision-making by individual agents. To evaluate the performance of the proposed model, it is conducted extensive simulations using real-world data collected from various peers. The results compared with rule-based working agents. The experiment shows incorporating very short-term forecasts significantly enhances the ability of agents to adapt to rapidly changing conditions, thereby leading to more efficient and stable energy trading decisions. The use of very short-term forecasts empowers prosumers to make informed decisions in response to dynamic energy market conditions, ultimately contributing to increased grid reliability, energy efficiency, and sustainability.

Keywords: Energy Trading, Multi-Agent, Reinforcement Learning, Peer-to-Peer Trading, Very Short-Term Forecasts

ÖZ

NANO/MİKRO ŞEBEKELERDE EŞLER ARASI ENERJİ TİCARET OYUNUNDAKİ ETKİLİ ENERJİ TİCARETİ İÇİN TAHMİN VE PEKİŞTİRMELİ ÖĞRENME STRATEJİLERİ: AMPİRİK ANALİZ

Güneş, Rabia Şeyma
Yüksek Lisans, Bilişim Sistemleri Bölümü
Tez Yöneticisi: Prof. Dr. Sevgi Özkan Yıldırım

Eylül 2023, 88 sayfa

Dağıtık enerji sistemlerine dahil olan yeni teknolojiler şebekedeki talep ile üretim değişkenliğinin ve bu değişkenlikten doğan maliyetlerin yönetilmesine fırsat veren çözümler oluşturmuştur. Bu küçük şebekeler arasında üreten tüketicileri içeren ticaret, artık enerjinin birbirleri arasında satılmasına ve ihtiyaç duyulan enerjinin alınmasına olanak tanıyarak maliyetleri ve sistem kısıtlamalarını azaltmıştır. Bu ticaretin amacı, pek çok oyuncunun fayda sağlayacak şekilde, pekiştirmeli öğrenme yöntemlerinde bahsedilen ajanların oyunu olarak modellenmiştir. Her oyuncu, kendi elektrik talebini hem kendi iç kaynaklarından hem de diğer oyunculardan sağlamaktadır. Bu tezin amacı, bu oyunda her oyuncunun talebini maksimum fayda ile karşılayarak sistem kısıtlamalarına uymaktır. Bu çalışmada, oyuncular arasında çok kısa vadeli enerji ticaretini kolaylaştırmak için yeni bir Çoklu Ajan Pekiştirmeli Öğrenme modeli önerilmektedir. Tezin ana katkıları, çok kısa vadeli yük, üretim ve fiyat tahminlerini çerçeveye dahil ederek bireysel ajanların daha doğru kararlar almasını sağlamaktır. Önerilen modelin performansını değerlendirmek için çeşitli oyunculardan toplanan gerçek sistem verileri kullanılarak kapsamlı simülasyonlar yapılmıştır. Sonuçlar, çok kısa vadeli tahminlerin ajanların hızlı değişen koşullara uyum sağlama yeteneğini önemli ölçüde artırdığını, daha verimli ve istikrarlı enerji ticareti kararları alınmasını sağladığını göstermektedir. Çok kısa vadeli tahminlerin kullanımı, üreten tüketicilerin dinamik enerji piyasası koşullarına karşı bilinçli kararlar almasına olanak tanırken, aynı zamanda artan şebeke güvenilirliği, enerji verimliliği ve sürdürülebilirliğe katkıda bulunmaktadır.
Anahtar Sözcükler: Enerji Ticareti, Çoklu Ajan, Pekiştirmeli Öğrenme, Eşler Arası Ticaret, Çok Kısa Süreli Tahminler

To beloved my family

ACKNOWLEDGMENTS

I would like to express appreciation to my advisor, Prof. Dr. Sevgi Özkan Yıldırım, for her invaluable guidance, unwavering support, and encouragement throughout the journey of this thesis study. Her expertise, patience, and insightful feedback have been instrumental in shaping the direction of this research.

I would also like to extend my deepest gratitude to my dear family for their constant love, understanding, and encouragement. Their unwavering belief in me has been a source of strength and motivation during the ups and downs of this school life.

Lastly, a special thank you to my husband Zekeriya for being my rock and pillar of support throughout this entire thesis study. Your patience, understanding, and unwavering belief in my abilities have been the driving force behind my success.

I extend my sincere gratitude to professors, staff, and fellow students who have played a significant role in my learning and personal development in Information Systems program. Their invaluable insights, vast knowledge, and unwavering support have made a profound impact on my academic journey and growth.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A2C	Advantage Actor-Critic
ANN	Artificial Neural Networks
API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Average Model
CALA	Continuous Action Learning Automation
CDD	Cooling Degree Days
CNN	Convolutional Neural Network
DER	Distributed Energy resource
DNI	Direct Normal Irradiance
DQN	Deep Q-Networks
EMS	Energy Management System
EPİAŞ	Enerji Piyasaları İşletme A.Ş.
FALA	Finite Action Learning Automation
GNB	Generalized Nash Bargaining
GTI	Global Horizontal Irradiance
HDD	Heating Degree Days
HVAC	Heating, Ventilation, and Air Conditioning
IEA	International Energy Agency
KNN	k-Nearest Neighbors
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MARL	Multi-Agent Reinforcement Learning
MCTS	Monte-Carlo Tree Search
MDP	Markov Decision Process
MLP	Multi-Layer Perceptron
MPC	Model Predictive Control
MSE	Mean Squared Error
NMAE	Normalized Mean Absolute Error
P2P	Peer-to-Peer
PV	Photovoltaic
RBRL	Rule Based Reinforcement Learning
ReLU	Rectified Linear Units

RL	Reinforcement Learning
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SAC	Soft Actor Critic
SARSA	State-Action-Reward-State-Action
SoC	State of Charge
SVM	Support Vector Machines
USD	United States Dollar

CHAPTER 1

INTRODUCTION

1.1. Problem Statement

The global energy demand is rising at a rapid pace, with electrical energy being the fastest-growing energy source among all (International Energy Agency, 2018). According to International Energy Agency's (IEA) report dated 2018, electricity consumption will double of current demand by 2040. Countries worldwide are confronted with energy-related challenges, including decrease in fossil fuel reserves, the need for sustainable energy provision, and escalating impacts of global warming (Mackay, 2008). IEA predicts that electricity will surpass usage of other energy sources within the next 25 years, and as a result, it places significant emphasis on development and utilization of electricity (International Energy Agency, 2018). Renewable energy is critical in addressing these energy-related issues. The modern electricity grid is embracing growing presence of green energy sources such as solar, wind, and hydro, enabling integration of innovative and intelligent solutions throughout grid infrastructure. By adopting these solutions, renewable energy systems have been able to keep pace with advancements, effectively handling fluctuations in demand through time-varying generation, and concurrently decreasing expenses and reliance on main grid.

The most important of these solutions are nano/microgrids that are part of smart grid. Nano/microgrids are the parts that contain dynamics of the real grid, which makes small-scale energy supply that consists of consumers with electricity demand, solar or wind sources that produce renewable energy, and battery systems that store energy (Figure 1). Microgrids are larger and more complex, operating autonomously with mix of energy sources, while nano grids are smaller, usually connected to main grid, and rely on a limited number of energy sources. In other words, microgrids are networks of nanogrids (Nordman & Christensen, 2015). Each microgrid is obliged to main grid when it is insufficient while trying to meet its energy demand with its energy sources. Besides, it wants to benefit from this energy when it cannot store the excess energy it has. By trading with other nano/microgrids like itself, it can reduce its dependency on the grid, sell its excess energy, and buy energy for times in need of energy. In this way, it can also manage the time-dependent variability of renewable energy sources and electricity demand. The fact that it is trading with nearby nano/micro grids reduces the energy transmission loss caused by the distance from the main grid.

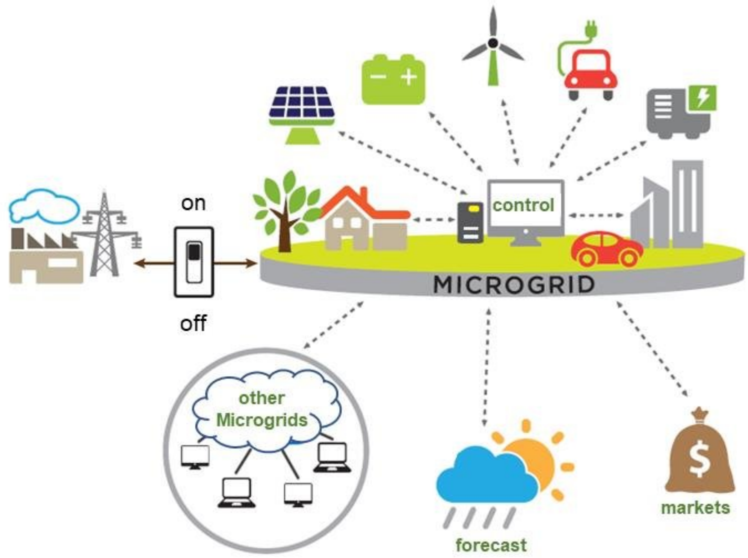

Figure 1: Components of Nano/Micro Grid (Stadler & Nasle, 2019)

It is known as peer-to-peer (P2P) energy trading concept that microgrids, which are many consumers, some of which are prosumers, buy and sell energy between each other. P2P energy trading system allows neighboring producers and consumers to exchange energy directly with each other without the need for a main grid. However, when doing this trade, each microgrid should create bids and offers according to the current battery level, expected renewable energy generation, and energy demand (Zahraoui, et al., 2021).

While each nano/microgrid tries to meet its local energy demand, it maximizes its profit from the market, while the market finds electricity prices and matches according to the offers according to the clearing mechanism. There is not only energy distribution and pricing in the market. Buyers' prioritization, physical constraints, and costs arising from the network also play an active role in optimizing the market. In establishing this optimization problem, the aim may be to minimize production, demand, transmission, and other costs or to maximize total profit.

1.2. Scope of the Thesis

The purpose of this thesis is to make each nano/microgrid self-sufficient and maximize profit while meeting the electricity demand of each nano/microgrid. In doing so, attention is paid to the following.

- The stochastic situation due to time-dependent changes in electricity demand, renewable energy generation, and electricity price is prioritized.

- The dependence of nano/microgrids on production sites is reduced by battery and trading.

- Energy losses from transmission are reduced by exchanging neighboring peers.

- Energy balance equations and system constraints are provided.

In this thesis, a reinforcement learning model is generated to optimize P2P energy trading system while achieving goals of each nano/microgrid. The aims of this thesis are:

- Formulating both the energy trading game of each peer and whole system in terms of local energy demands, battery levels, generations from renewables, and system constraints.

- Maximizing profit of each participant of P2P energy market with consideration of traded energy, battery charge/discharge, and generated energy.

In this study, it is presented pioneering simulation of P2P energy trading approach, using data specific to Turkey. The study is the first simulation of peer-to-peer energy trading approach using data specific to Turkey. Through this simulation, it is explored the feasibility and potential benefits of implementing such a system within Turkish energy landscape.

1.3. Research Questions

In this thesis study, the solutions to some questions were searched. The first is how to transfer energy among participants including maximizing their resources and utilities. The most appropriate solution to this problem is to enable the peers to use energy generation and storage resources in the most efficient way.
Another research was done to look for the answer to how to affect electricity prices, renewable energy generation, local power demand, and battery levels from energy trading policy.
One of the most critical concerns of peers entering this market is meeting the local energy demand. Peers can become participants in this market if they can meet their demands with the energy market that does P2P trading without the need for main grid. Therefore, another research question is related to the market confidence of nano/microgrids in meeting their energy demands.
While searching for answers to the questions mentioned in this thesis study, an algorithm based on forecasting and reinforcement learning was developed to provide a solution. With this algorithm, while resource allocation was performed, information is transmitted between prosumers, and profit is maximized while maintaining network and system constraints.

1.4. Outline of the Thesis

The rest of this thesis study is organized as follows: Chapter 2 explains literature review of machine learning approaches, energy forecasting, and reinforcement learning techniques for P2P energy trading systems. Chapter 3 gives the methodology of this thesis study. It covers forecasting and reinforcement learning strategies for efficient energy exchange in P2P energy trading game. In Chapter 4, the simulation of an experiment is explained consisting of P2P energy trading game among 8 prosumers. Lastly, Chapter 5 concludes this thesis study with a summary and future work on P2P energy trading game.

CHAPTER 2

LITERATURE REVIEW

In this chapter, various aspects of the machine learning discipline and its application to energy trading and forecasting are presented. The literature review is divided into three sections, each focusing on a distinct topic.

The first section is an overview of the discipline of machine learning, highlighting different learning paradigms commonly utilized in the literature. These paradigms contain supervised learning, unsupervised learning, and reinforcement learning. Moving on to the second section, the concept of energy trading and models employed in this domain are explained. Existing literature on energy trading, exploring various market structures, pricing mechanisms, and trading strategies are examined. Because of usage of supervised machine learning approaches applied in energy trading, such as demand, generation, and price forecasting, energy forecasting models are examined in the third section to identify the most effective and relevant techniques for energy trading applications.

2.1. Discipline of Machine Learning

The field of machine learning has emerged as a distinct discipline, encompassing algorithms, methodologies, and techniques that enable computers to learn from data and make predictions or decisions (Conway & White, 2012).
Machine learning can be categorized into various types based on different dimensions and approaches (Perez C. , 2019). The common types are supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. There are often overlaps and hybrid approaches, and researchers continue to develop new algorithms and techniques that push the boundaries of these categories. For instance, ensemble learning technique combines multiple models or learning algorithms to make predictions or classifications. By aggregating the predictions of individual models, ensemble methods can often achieve better performance and robustness compared to using a single model. Understanding the main types of machine learning is crucial as it enables tackling a broad range of problems and datasets effectively. Hence, the following subsections provide a concise overview of these types.

2.1.1. Supervised Learning

In supervised learning, the algorithm learns from labeled training data, where each data instance is associated with a known target or label (Perez C. , 2019). The algorithm learns to make predictions or classify new, unseen data based on this labeled training set. Supervised learning plays an integral role in applications across various industries, including predictive analysis in finance, medical diagnostics, spam filtering in email services, sentiment analysis, customer churn prediction in marketing, credit scoring in banking, recommendation systems in entertainment platforms, and social network filtering in social media platforms. Regression and classification are two primary types of supervised learning and associated some common algorithms with these types are below.

a) Logistic Regression (classification)

Logistic regression models the relationship between the input variables and the probability of belonging to a certain class. It uses a logistic function to estimate the probabilities and makes predictions based on a threshold.

b) Support Vector Machines

SVM constructs a hyperplane or set of hyperplanes to separate data points for different classes in classification or within a certain margin or tolerance in regression. It aims to maximize the margin between the classes and can handle both linear and non-linear tasks.

c) k-Nearest Neighbors

KNN classifies new instances based on their proximity to labeled instances in the training data. It assigns the class label based on the majority vote of its k nearest neighbors in the feature space.

d) Decision Trees

Decision trees create a tree-like model of decisions and their potential consequences. Each internal node represents a feature or attribute, and each leaf node represents a class label. Decision trees can handle both classification and regression tasks.

e) Random Forests

Random forests are bagging-based ensemble learning method that combines multiple decision trees. Each tree is trained on a random subset of training data, and predictions are made by aggregating the results from individual trees. Random forests improve prediction accuracy and reduce overfitting.

f) Linear Regression (regression)

Linear regression models the relationship between input variables and continuous output by fitting linear equations. It estimates the coefficients that best-fit input data using the method of least squares.

g) Neural Networks

Neural networks, such as multilayer perceptron (MLP), are powerful models composed of interconnected nodes or neurons on la. They learn complex relationships in the data through hidden layers of neurons and can handle both classification and regression tasks.

These explanations provide a brief overview of the algorithms and their characteristics. Each algorithm has its own underlying principles and specific use cases. It's important to consider the nature of the data, the complexity of the problem, and the requirements of the task when selecting the most suitable algorithm for the supervised learning problem.

2.1.2. Unsupervised Learning

Unsupervised learning involves learning patterns and structures in unlabeled data (Conway & White, 2012). The algorithm discovers hidden patterns, clusters, or relationships within the data without any specific target or label. Clustering algorithms, dimensionality reduction techniques (e.g., Principal Component Analysis), and generative models (e.g., Gaussian Mixture Models) are primary algorithms of unsupervised learning methods.

2.1.3. Reinforcement Learning

Reinforcement learning (RL) focuses on how an agent can learn to make sequential decisions through interaction with an environment to maximize a cumulative reward signal (Sutton & Barto, 2018). The agent learns through trial and error, exploring the environment, and receiving returns as rewards or punishments. It draws inspiration from the way humans and animals learn through trial and error. In RL, an agent interacts with an environment, takes actions based on its current state, and receives feedback in the form of rewards or punishments. The agent's objective is to learn a strategy that guides its decision-making process to maximize the long-term rewards it receives from the environment. To understand better, the key components of RL framework are shown in Figure 2 and explained as follows:

- Agent: The learner or decision-making entity that interacts with the environment. The agent takes actions based on its current state and receives rewards from the environment.

- Environment: The external system or world in which the agent operates. The agent explores the environment, makes decisions, and aims to improve its performance over time.

- State: Current condition of the environment at a given time.

- Action: The choices or decisions made by the agent in response to its current state. Actions affect the subsequent state and the rewards received.

- Reward: A scalar value that quantifies the desirability or quality of the agent's actions. The agent's goal is to maximize the cumulative reward over time.

- Policy: A strategy or mapping from states to actions, which guides the agent's decision-making process. The policy guides the agent in choosing actions based on its current state. The agent's goal is to learn an optimal policy that maximizes its cumulative reward in the long run.

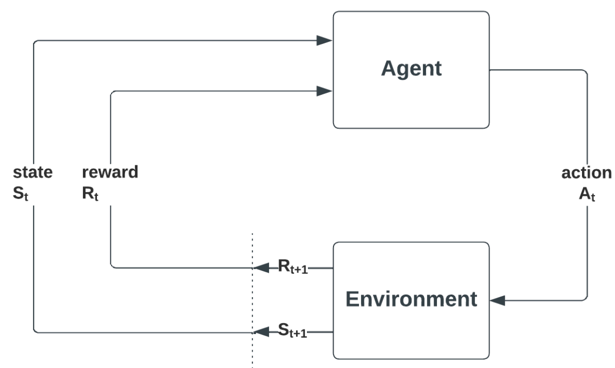

Figure 2: Reinforcement Learning Framework

The agent learns to improve its decision-making capabilities by employing various learning algorithms and optimization techniques. Popular RL algorithms include Q-learning, Deep Q-Networks (DQN), Policy Gradient methods, and Actor-Critic methods. RL has been successfully applied to a wide range of domains, including robotics, game playing, driving, and resource management. It enables agents to learn complex behaviors, adapt to dynamic environments, and make decisions in situations where explicit supervision or labeled training data is not available.

RL can be categorized into different groups based on the components used in the learning process, including value function, policy, and model. Main categories are explained in the following sections:

a) Value-Based RL

Value-based RL methods focus on estimating and optimizing the value function, which represents the expected long-term cumulative reward for an agent in a given state or state-action pair. These model-free methods aim to find the optimal value function or value function approximation without explicitly learning a policy. In value-based RL, Bellman equation plays a crucial role in updating and estimating value function. Bellman equation expresses the relationship between the value of a state or state-action pair and the values of its successor states. It is a recursive equation that consists of the immediate reward and the discounted value of the next state(s). Popular algorithms in value-based RL are Q-learning, Deep Q-Networks (DQN), and SARSA(State-Action-Reward-State-Action).

b) Policy-Based RL

Policy-based methods directly learn the optimal policy without estimating a value function. The policy represents the mapping from states to actions, leading the agent's decision-making process. These model-free methods optimize the policy by searching for policy parameters that maximize the expected cumulative reward. REINFORCE (Monte-Carlo Policy Gradient) is the most popular policy-based RL method.

c) Critic RL

Actor-critic RL methods combine value-based and policy-based approaches and are also in model-free category. They operate both a value function (critic) and a policy (actor) to learn and improve agent's behavior. Algorithms such as Soft Actor-Critic (SAC) and Advantage Actor-Critic (A2C) are model-free actor-critic methods.

d) Model-Based RL

Model-based RL algorithms explicitly learn a model of the environment. These algorithms aim to estimate the transition probabilities and rewards based on observed interactions with the environment. Model-based methods combine model learning with planning algorithms to optimize agent's behavior. Examples are Monte-Carlo Tree Search (MCTS), Model Predictive Control (MPC), and Dyna-Q.

e) Multi-Agent RL

Multi-agent reinforcement learning (MARL) is an area of research that focuses on learning in environments where multiple agents interact and learn simultaneously (Zhang, Yang, & Başar, 2021). MARL's goal is for each agent to learn its optimal policy while taking into account presence and actions of other agents in the environment.

MARL involves extending traditional RL framework to accommodate multiple learning agents. Each agent has its state, action space, and policy, and they interact with the environment concurrently. The agents' actions can affect each other and overall system dynamics, introducing the challenge of coordination, competition, or cooperation between the agents. MARL has applications in diverse domains, including robotics, multi-robot systems, traffic control, and social networks, where interactions between multiple agents play a crucial role. Moreover, this research area is used in energy trading games and each market participant acts as an agent.

2.2. Peer-to-Peer Energy Trading

Peer-to-peer (P2P) energy trading refers to the direct exchange of energy between producers and consumers within a local energy system. The term "peer" refers to individuals or entities within a local energy system who participate in trading of energy. These peers can include energy producers, consumers, or both and be interconnected distributed energy resources, and microgrids.

The objective of P2P energy trading game algorithms is to investigate a solution for meeting energy demand of peers while maximizing the profit of each. One of the review studies on peer-to-peer energy trading examined investigated studies in six different approaches (Soto, Bosman, Wollega, & Leon-Salas, 2021). These approaches are listed as algorithms, optimization, trading platforms, blockchain, simulation, and game theory. The review said that the study often covers multiple approaches. All of the studies either have optimization or algorithm approaches. In this thesis, research is conducted based on these approaches.

One of the early researches on this topic, that is in the recent past, proposes a distributed algorithm that considers grid-connected microgrid with PV and storage system and optimizes battery scheduling according to dynamic load and solar power (Raju, Sankar, & Milton, 2015). In this model, each agent uses Q-learning algorithm for this optimization of itself. For optimization of multiple agents, Multi Agent RL algorithm, namely Coordinated Q-Learning is proposed to improve on utility of battery and solar in the grid. This proposed system runs in a strategic manner for different operational scenarios to achieve the possible minimum cost. Markov Decision Process is used for the sequential decision-making strategy. All in all, this proposed Coordinated Q-Learning algorithm can handle stochastic patterns of demand and generation and reduce dependency on electricity grid. In the same year, another study proposes an optimization model whose aim is different (Guan, Wang, Lin, Nazarian, & Pedram, 2015). This study proposes a reinforcement learning technique for optimal control of the storage system of households. Optimization of energy storage systems is critical for uncertain nature of PV generation and load. To take higher performance from the method, TD(λ)-learning algorithm is used and electric bill minimization for residential households is determined as the goal in the objective function. More recently, another paper introduces peer-to-peer energy trading framework based on repeated game which gives chance randomly selection of strategy for microgrids individually for trading energy in independent market (Wang H. , Huang, Liao, Abu-Rub, & Chen, 2016). The purpose of these markets is to maximize the total revenue. To establish Nash equilibrium handling different scenarios learning automation-based algorithms were developed. This paper contributes a finite action learning automation (FALA) based model to update each participant's probability distribution for seeking the best action for each in discrete strategy space. For continuous strategy space, continuous action learning automation (CALA) base model was proposed.

Some research focuses on solutions to balance energy load and generation including grid constraints in a microgrid instead of microgrids. Three different strategies for reducing costs and increasing income for producers in small-scale distributed energy resources were proposed, namely mid-market rate, auction-based pricing, and bill-sharing strategy research (Long, et al., 2017). Each of them was evaluated on a residential microgrid with solar systems. According to the results presented in the paper, moderate level of solar penetration reduces energy costs by 30%. The reason for this reduction is diversity of demands of households in the community microgrid. This study shows that flexible demand response can be added, and these strategies can be applied to larger-scale community microgrids. Because of this study, research focused on trading in a microgrid was checked.

One of the oldest studies proposes an algorithm that plans battery charging - discharging schedule and utilization from wind turbines with two steps ahead of reinforcement learning (Kuznetsova, et al., 2013). There is a microgrid case that has demand from local consumers, renewable energy generators from wind and battery storage systems. This system is also connected to the electricity grid. The proposed approach gives opportunity for optimal actions for battery scheduling in different seasonal and weather conditions to prosumers. In other words, prosumers can learn how to handle stochastic environment of the energy management system with this method. Reinforcement learning based algorithms used in the solution of this problem are increasing according to the latest researches. To understand local energy trading behavior in a grid, motivation of recent research is modeling energy trading behavior of prosumers, that have energy storage systems, in local energy market. In other words, this paper tries to explain how prosumers choose their trading strategy considering their energy resources on hand (Chen & Su, 2018). The deep reinforcement model based on Q-learning solves Markov decision process with multiple continuous variables. This helps decision making processes in the smart energy system and increases participation of prosumers in local energy market. This paper shows how deep Q-learning local energy trading algorithm outperforms rule-based and dummy random strategies in terms of daily economic benefit of prosumers. In another study, there is an introduction of a deep reinforcement learning model for decision making problem of microgrids using peer-to-peer energy trading model in local energy market (Chen & Bu, 2019). The introduced deep model uses Q-learning approach for an hour-ahead peer-to-peer energy trading model satisfying physical constraints like utilization of energy resources, charging and discharging limits of batteries. Also, this model includes virtual penalty cost in objective function for giving power plant schedules to microgrid participants. Another energy trading game model autonomous peer-to-peer energy trading method which maximizes the prosumers' profit (Kim & Lee, 2020). It is the modification of deep Q-network based algorithm and considers electricity bill, electric energy stored, trading energy, and virtual loss. The proposed methodology is based on a long-term delayed reward that enables monthly effective learning of patterns. This paper concludes that peer-to-peer reinforcement learning based energy trading model through long-short term delayed reward gives higher profits and reduces over generation of electric energy loss. Moreover, it maximizes each prosumer's profit on noncooperative game theory.

Some approaches combine different models with reinforcement learning based algorithms. One of the recent papers deals with challenging nature of microgrids with combination of deep learning and reinforcement learning approaches (Chandrasekaran, Kandasamy, & Ramanathan, 2020). This nature consists of high penetration of solar and wind and uncertainty of customer load. For demand forecasting side, the problem gets the help of deep learning reinforcement learning model. In this study, different components of microgrids including decision-making, power and demand forecasting, prediction, and analysis have served. For each component, different deep learning-based models were compared. It is stated that reinforcement learning models are preventive and take advantage of the control of generation.

Game theory methods are also used with reinforcement learning in a dissertation (Hu, 2020). This dissertation aims to introduce autonomous distributed control system for microgrids by performing demand response and energy management systems. To satisfy this aim, three different approaches were considered. In the first approach, multi-player game, different solutions were generated according to setting of game. When the number of players increases in the game, performance of the game decreases. For the second attempt, Q-learning and Linear Reward-Inaction based reinforcement learning algorithms were tried. Q-learning algorithm failed in multi-agent environment scenario while training time of Linear Reward-Inaction algorithm was too long. The last approach was based on load-ratio learning game algorithm which solves performance deterioration of the game. According to the dissertation, this algorithm gave the best result and showed the greatest potential to satisfy the requirements of microgrid energy management system in terms of communication.

An Optimized Reinforcement Learning with Decision Tree method was proposed for energy management and economic dispatch m to select the best policy in every situation in microgrid environment without using any forecasting module (Levent, et al., 2019). The proposed model is presented as learning phase of past data in microgrid and the execution of data with dynamic decision tree model. The model in this study does not require forecasting model to solve economic dispatch problem. However, it is stated that there is a need for an approximation method to generalize behaviors of participants in the market.

Apart from the algorithm for peer-to-peer energy trading problems in a microgrid, some papers concentrated on different problems. Peer-to-peer energy trading problems among houses in a microgrid is new concept and it brings some concerns. Security of the platform exposes trading environment is critical. The recent research introduces deep learning model with blockchain based framework for smart microgrids is proposed as DeepCoin (Ferrag & Maglaras, 2019). This model gives an opportunity to exchange surplus energy between neighboring agents. Also, this protects the smart grids from security deficiencies and cyber-attacks. For these, short signatures, hash functions and bilinear pairing methods are used. This peer-to-peer energy trading model is based on practical Byzantine fault tolerance algorithm. With this algorithm, consensus between agents can be satisfied. The novel deep learning scheme uses recurrent neural networks with truncated backpropagation through time algorithm. In this study, blockchain strategy and deep learning using backpropagation through time algorithm are used together first. For evaluating the proposed model, three different datasets were used, and the results were compared with support vector machines, random forest, and naïve bayes algorithms. The results overperformed these algorithms in terms of false alarm rate.

In addition to security, grid-based constraints were investigated further in recent research. The financial and electrical perspectives of energy transactions were taken into account in a platform for peer-to-peer energy trading process (Elliott, Shanklin, Zehtabian, Zhou, & Turgut, 2020). Furthermore, how number of users in the smart grid affect the grid from sustainability and reliability side. This platform checks the considerations of prosumers such as distributed energy sources of them and connectivity of main grid for possible participants of peer-to-peer energy systems. In the algorithm used in the platform, physical constraints including voltage regulator tap change, voltage limit, capacitor bank capacity, and branch current capacity are checked before matching. The main algorithm is based on the first in first out rule for incoming orders. The simulation results of the proposed platform show that increase in number of users provides more sustainable network.

The other paper proposes a fully decentralized approach for market clearing in peer-to-peer energy market (Paudel, Sampath, Yang, & Gooi, 2020). This proposed approach considers power losses and network utilization fees throughout peer-to-peer energy trading. For network utilization fees, electrical distance between consumers and producers is calculated and then the fees are given to the model proportional to these distances. In this proposed approach, the aim is to maximize social welfare by considering network usage. Also, the model includes transformed constraints by relaxing nonconvex constraints. According to this model, electricity prices and generation/demand amounts can be calculated by satisfying all the constraints and without sharing preferences of the agents.

Instead of maximizing the total profit, an approach focuses on the total cost of consumers (Alam, St-Hilaire, & Kunz, 2019). This paper proposes an approach for minimizing total cost of energy trading among smart houses in micrograds. There is Energy Cost Optimization via Trade which is the first near optimal cost optimization model for Demand Side Management. This model considers unfair cost distribution problem with Pareto optimality. Also, the paper evaluated effects of renewables and storage concerning households. According to the study, if households have renewables and storage, energy trading for this house is necessary to minimize the costs. Also, this paper shows that storage capacity does not increase cost savings linearly. Until saturation point, if there is a renewable, cost savings can increase when storage capacity increases. In this consumer-oriented approach, the households are protected when the grid fees are too high. One research covers all different aspects of smart grids in their paper (Mengelkamp, et al., 2018). In this paper, trades between consumers and prosumers in peer-to-peer energy fashion on microgrid market were explained. With this fashion, consumers and prosumers can arrange their costs and profits within their energy community. For this purpose, seven components of microgrid energy markets with blockchain were introduced. Microgrid setup must be defined so that a sufficient number of agents, in other words, producers and consumers, should be in the market, and market access should be only on market participants. Grid connection, high-performing information system, market mechanism for day-ahead and intraday markets, pricing mechanism, trading system, and regulation should be in market structure.

The energy trading mechanism among microgrids was investigated further after checking trading in a microgrid. One paper introduces a formulation for the energy trading game using Nash equilibrium of the game according to predicted renewable energy production and energy demand, battery level, and energy trading history (Xiao, et al., 2018). Also, there is a proposition of one reinforcement learning based model that applies deep Q-network for reducing dependency on power plants and improving utility of microgrids. This paper also shows that long-distance power transmission loss is reduced in dynamic energy trading games. In another paper, direct energy trading is formulated between multiple microgrids and also utility by using generalized Nash bargaining (GNB) problem method (Kim H. , Lee, Bahrami, & Wong, 2019). This proposed GNB problem considers maximizing social welfare and distributing the income between microgrids based on their market power. The paper proposes solving the optimal power flow problem to determine amount of energy trading and to determine the market clearing price and mutual payments of the microgrids. The generalized Nash bargaining problem has been attributed to three key contributions Direct Trading Framework, Distributed Optimization Methods, and Significant Cost Reduction. While GNB minimizes the total cost and maximizes social welfare, Direct Trading Framework maximize the profit of each microgrid. Distribution optimization methods in this paper guarantee amount of energy exchange of each grid is proportional to its profit.

To understand different models based on different logic, some review papers were checked. One paper searches peer-to-peer energy trading projects in terms of their main focuses, outcomes and then compares these projects' similarities and differences (Zhang, Wu, Long, & Cheng, 2017). According to the paper, some projects only focus on the development of their business models and ignore application of these models to smaller local energy markets. However, some of these provide connections between producers and consumers and the electricity price can vary in these projects while some give importance to shortage systems. This paper concludes that the design of communication and control systems on these networks is important for enabling energy trading in a microgrid or among microgrids. The other article is a comprehensive review and analysis of peer-to-peer energy trading concept based on prominent academic papers, projects in literature, and industrial applications (Zhou, Wu, Long, & Ming, 2020). This review examines the previous studies in terms of energy trilemma which covers three critical objectives: energy security, energy equity, and environmental sustainability. In other words, what problems came and what problems were solved with peer-to-peer trading and distributed energy resources, that can share and trade energy. This paper covers journal papers published in the last five years. According to total number of papers published in a year, peer-to-peer energy trading has become popular for two years. This paper states that majority of papers in this area focused on market design, trading platforms, physical and information technology infrastructure of peer-to-peer energy trading. Also, it says that policy issues are the problem of deploying these platforms on large-scale markets. This review paper mentions papers that utilize blockchain technology to facilitate peer-to-peer energy trading solutions at different levels. Market design models such as centralization level, differentiation of products, game theoretic perspective, and market stability are included in this paper. These are very important to understand concepts in this problem and determining starting point for new study.

In recent years, studies conducted have shifted towards multi-agent reinforcement learning. One of the studies proposes agent-based transactive energy trading platform integrating energy storage systems, utilizing reinforcement learning for bidding strategies, and addressing losses with simulations in multi-microgrid systems (Nunna, Sesetti, Rathore, & Doolla, 2020). Another multi-agent based model introduced in a paper that novel P2P transactive energy trading scheme using multi-actor algorithm, including scalability and privacy challenges in efficient local energy trading while minimizing cost and peak demand reduction compared to existing methods (Ye, Tang, Wang, Zhang, & Strbac, 2021). These multi-agent approaches are needed to create economic benefits. Markov decision process bases multi-agent reinforcement learning model was presented with achieving maximum income and ensuring privacy (Qiu, Wang, Dong, Wang, & Strbac, 2022). The advantages of multi-agent reinforcement learning in terms of cost, scalability, and applicability demonstrate the feasibility of this method. Therefore, this thesis study adopts focus on multi-agent reinforcement learning for these reasons.

2.3. Energy Forecasting

Energy forecasting is a crucial task in energy systems planning, operation, management, and optimization. It involves predicting future energy consumption, generation, prices, or other relevant variables to support decision-making processes. Energy forecasting can be classified into different categories based on various factors. According to forecasting horizons category shown in Figure 3 with their usage purposes, temporal granularity is categorized as follows (Zor, Timur, Çelik, Yıldırım, & Teke, 2018):

- Very Short-Term Forecasting: Very short-term forecasting focuses on predicting energy variables in the very near future for energy purchasing activities, typically up to a few hours (Mir, et al., 2021).

- Short-Term Forecasting: This type of forecasting focuses on predicting energy variables in the near future, typically up to a few days or weeks.

- Medium-Term Forecasting: Medium-term forecasting involves predicting energy variables for a longer time horizon, ranging from a few weeks to a few months.

- Long-Term Forecasting: Long-term forecasting extends the prediction horizon to several months, years, or even decades. It aims to provide insights into future energy trends and planning scenarios.

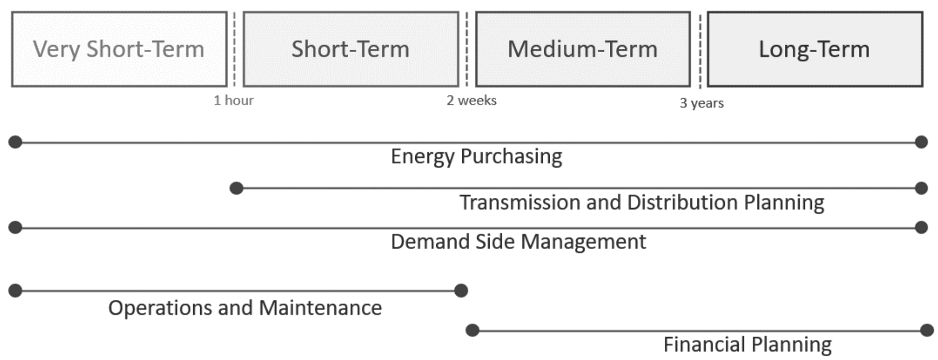

Figure 3: Forecasts' Purposes for Different Horizons

Energy forecasting can be classified according to modeling approaches as follows (Hong, et al., 2020):

- Statistical Methods: Statistical methods use historical data patterns and mathematical models to create forecasts. Some algorithms are time series analysis, regression, and exponential smoothing.

- Machine Learning: Machine learning methods, such as neural networks, random forests, support vector machines, or gradient boosting, learn patterns from historical data to create forecasts.

- Hybrid Models: Hybrid model approaches combine statistical and machine learning techniques to consolidate their respective strengths and improve forecasting accuracy.

In terms of spatial forecasting approaches, there are two different solutions in the literature:

- Point Forecasting: It provides a single predicted value for a specific energy variable at a given point in time.

- Probabilistic Forecasting: It provides a range of possible outcomes along with their associated probabilities. It quantifies the uncertainty in the forecasts and helps decision-making.

Energy forecasting techniques are also classified according to output energy variable (Hong, et al., 2020):

- Load Forecasting: Load forecasting focuses on predicting future electricity demand or consumption patterns. It helps utilities and grid operators optimize their generation and distribution plans.

- Generation Forecasting: Generation forecasting predicts the future output of energy sources, such as wind, solar, or hydro power plants. It assists in optimal resource allocation and energy trading decisions.

- Price Forecasting: Price forecasting involves predicting future energy market prices, such as electricity or natural gas prices. It helps market participants in making informed trading and investment decisions.

All in all, the classification of energy forecasting depends on the specific context, requirements, and variables of interest. Different forecasting methods and models are applied based on the classification criteria to fit certain forecasting tasks and time frames.

2.3.1. Load Forecasting

Since electricity load forecasting is important for both the management of power systems and trading in energy market, models using different approaches have been published in the last decades. One of the most basic models is autoregressive integrated moving average model (ARIMA) which is based on time series. This model was used both by giving exogenous input (ARIMAX) and adding seasonal factors (SARIMA - SARIMAX) (Mohamed, Ahmad, Suhartono, & Ismail, 2011; Nengbao, Babushkin, & Afshari, 2014). Random forest structure, which is one of the ensemble methods based on decision trees, has been preferred in some studies because it creates a large number of trees and reduces the effect of weak classifiers (Dudek, 2014). In another approach that shows the effect of ensemble methods on error reduction, gradient boosted trees were used (Taieb & Hyndman, 2014). After exploration of neural network and deep learning approaches in time series modeling and taking better results, recurrent neural networks were used with long short-term layers (Muzaffar & Afshari, 2019). Especially RNNs have become very popular in time series modeling (Connor, Martin, & Atlas, 1994). These networks are special cases of autoregressive moving average models nonlinearly. The different RNN approaches were applied with clustering similar days by Mandal (Mandal, Senjyu, Urasaki, & Funabashi, 2006). In another approach, LSTM architecture was used first time in power demand forecasting by adding weather, calendar, and time series features (Cheng, Xu, Mashima, Thing, & Wu, 2017).

Another neural network architecture used in this problem is convolutional neural networks. For this project, I checked especially two papers. I took help from them intuitively. One of them is based on CNN with transfer learning. It focuses on capturing intraday, daily, and weekly seasonal patterns using limited training data with convolutional neural network approach (Hooshmand & Sharma, 2019). In this paper, the problem is having less imbalance using an insufficient amount of training data for load forecasting. To overcome this problem, publicly available datasets are used for training the first CNN-based predictive model. In this model, there 5 convolutional layers with 32 filters of size. After each convolutional layer, batch normalization and max-pooling layers come. ReLU is the activation function of the model. There are dense and dropout layers (0.5) before the output layer. Before training the model, the normalization step comes to correct scaling differences between target and public datasets. CNN model is trained with public datasets and finetuned with available limited data of the target. This approach was tested on publicly available customers' datasets. Results were compared with SARIMA and CNN model without using transfer learning (fresh CNN). The normalized mean absolute error of the final model is 0.7 when SARIMA is 0.87 and fresh CNN is 0.84 for a month period. The approach in the paper is so meaningful because load forecasting of customers who are in a similar group like utility, house, etc. can show similar patterns.

Another paper does not use deep learning approaches. However, there is a transfer learning approach that uses the load of some cities for forecasting load of target city (Zeng, Sheng, & Jin, 2019). In this paper, TrAdaBoost algorithm was applied. The main purpose of the paper is to decrease errors in holidays. First, the source cities are selected according to the distance to target city in China. The model was compared with support vector regression model which is also applied to load forecasting problems. The mean absolute percentage error of the proposed algorithm is 2.88, while support vector machine's is 3.16. Another tree-base boosting algorithm LightGBM has been used in multiple works (Yao, Fu, & Zong, 2022; Zhou, Lin, & Xiao, 2022). They showed robust and applicable performances in regional and customer-based load forecasting.

2.3.2. Generation Forecasting

Certainly, while there are various renewable energy sources available such as wind, hydro, and geothermal, this thesis study primarily focuses on solar energy for a few key reasons. Solar power has a wide range of applications and scalability, from small rooftop installations to large-scale solar farms, making it a viable option for different energy needs (Gueymard, 2009). In Turkey, the share of solar energy production in energy generation is increasing day by day with both licensed and unlicensed solar energy investments. Due to its inherent uncertainty, accurate solar power generation forecasting is of critical importance for both the system operators and solar plant owners. Expected generation forecasts made a certain period in advance are communicated to system operator, enabling them to perform necessary planning. Additionally, solar plant owners require generation forecasts to optimize their post-production trading. The literature on solar power generation forecasting can be categorized into two modeling approaches: physical models and data-driven models that learn from historical data.

Physical models are based on calculating the radiation value on an inclined surface and determining the amount of energy that will be produced by the solar panel in the future in a deterministic manner (Hay, 1979; Gueymard, 2009; Hooshmand & Sharma, 2019). The basic equations and variables that form the basis of physical modeling are given below.

$$GTI = DNI * cos\theta + R_d * DHI + f_s * \rho * R_r * GHI \quad (1)$$

where θ is beam angle on tilted plane, R_d is diffuse transposition factor, f_s is shading factor, ρ is ground albedo, and R_r is ground reflection transposition factor.

Different models in the literature model different components of $R_d * DHI$. The formula of Hay Model (Hay, 1979) is shown below:

$$DHI[F * R_b + \frac{1+cos\beta}{2} * (1-F) \quad (2)$$

The formula of Gueymard Model (Gueymard, 2009) is shown below:

$$DHI\left[\left(1-N_{pt}\right) * r_{d0} + N_{pt} * r_{d1}\right] \quad (3)$$

The formula of Perez Model (Perez, Ineichen, Seals, Michalsky, & Stewart, 1990) is shown below:

$$DHI\left[\frac{F_1 a}{b} + (1-F_1) * \frac{1+cos\beta}{2} + F_2 * sin\beta\right] \quad (4)$$

Data-driven methods generally utilize radiation and temperature weather variables as inputs and rely on a mathematical function between historical plant productions and predicted values for future estimations using a linear or nonlinear model (Trapero, Kourentzes, & Martin, 2015). Statistical methods are primarily class that models the input and output space using a linear function. Mathematically, in the function $y = f(x)$, the f function is linear. The x variables representing the input space consist of variables from the equations (1) and lagged values of solar production. Linear regression, ARIMA(X), and SARIMA(X) models are found in the literature (Trapero, Kourentzes, & Martin, 2015; Yang, Thatte, & Xie, 2006). Machine learning regression models such as SVM, tree-based bagging/boosting models including random forest, gradient boosting, LightGBM, catboost, and artificial neural networks (Persson, Bacher, Shiga, & Madsen, 2017; Mellit & Pavan, 2010; Sobri, Koohi-Kamali, & Rahim, 2018) associate input and output space with nonlinear function. Mathematically, f function is nonlinear in function $y = f(x)$. Hybrid methods are models that combine linear and nonlinear models.

2.3.3. Price Forecasting

The need for accurate electricity price forecasting is fundamental in the energy market landscape. It serves various stakeholders, including power producers, consumers, and traders, by informing strategic bidding, investment planning, contract formulation, and risk management. With increasing investment in renewable energy sources and their variability, accurate price forecasting has become more crucial and complex (Grossi & Nan, 2019).

Over recent years, several sophisticated algorithms have been used to improve the accuracy of electricity price forecasting. Machine learning algorithms including Support Vector Machines (SVM), Gradient Boosted Trees, Artificial Neural Networks (ANN), and Random Forests have gained popularity in this domain due to their ability to capture complex nonlinear relationships and adapt to changing market dynamics (Lago, Marcjasz, Schutter, & Weron, 2021; Zhou, Wang, Wang, Wang, & Yang, 2018). LightGBM model is also used in some papers (Park, Jung, Jung, RHo, & Hwang, 2021). Additionally, time-series forecasting models such as ARIMA and GARCH have been widely used due to their ability to handle volatility in price series (Zhang, Zhang, Li, Tan, & Ji, 2019). Hybrid models combining different approaches are also emerging as a robust way to capture the best of different forecasting techniques (Yang, Ce, & Lian, 2017).

CHAPTER 3

MULTI-AGENT REINFORCEMENT LEARNING FOR PEER-TO-PEER ENERGY TRADING

3.1. Problem Statement

The new technologies that come with renewable energy solutions have enabled varying structures in demand and production depending on time, avoiding the costly architecture of main grid and providing demand with different solutions. In recent years, there has been significant accretion in small-sized distributed energy resources, such as solar panels, wind turbines, and home energy storage systems. This growth has allowed consumers to become producers of energy, giving rise to the concept of prosumers. Prosumers are individuals or entities, like homeowners or businesses, who both generate and consume energy. They are called also nano grids. On the other hand, a microgrid refers to localized energy system that operates independently or in conjunction with main power grid (Nordman & Christensen, 2015). It integrates various distributed energy resources, such as renewable generation, energy storage, and controllable loads, to supply power to a specific area. The combination of prosumers and microgrids represents a symbiotic relationship, where prosumers contribute to the resilience and sustainability of microgrids, while microgrids provide a platform for prosumers that are actively involved in energy market and maximize value of their renewable energy investments. Therefore, although the term "microgrid" is predominantly used in the thesis, the concept of "prosumer" will occasionally take its place.

Microgrids install renewable energy systems on their properties, benefiting from reduced reliance on traditional energy sources and the potential to save money and earn revenue by selling surplus energy. Thus, the transformation of renewable energy systems by microgrids fosters a more resilient, cost-effective, and environmentally friendly energy future. Furthermore, it helps the next generation energy management technique called peer-to-peer (P2P) energy trading has shown up.

3.2. P2P Energy Trading Platform

Peer-to-peer energy trading refers to energy producers and consumers can directly trade electricity with each other through a platform, without the involvement of intermediaries or traditional energy providers. P2P trading enables prosumers/microgrids to be actively involved in energy market by selling their surplus energy, contributing to a more resilient and decentralized energy system. Although microgrids create energy balance within themselves, they can gain more profit by entering the market. In markets where more than one microgrid can trade with each other, participants can buy, sell, and share energy with each other. In doing so, constraints such as energy needs, prices, distance of energy resources, or social preferences are provided.

However, trading in P2P energy market has challenges because it expects prosumers to trade their energy with little or no control from a central authority, making P2P platforms untrustworthy systems. Encouraging prosumers to work together in such an untrustworthy environment is difficult, especially in large energy systems where it is hard to understand how different factors affect energy trading decisions due to potential conflicts of interest among prosumers in the market. Managing P2P energy markets to address issues related to price and technical constraints is a complex problem requiring careful adjustment.

P2P energy trading offers several advantages. Firstly, it allows for greater efficiency and flexibility in energy market by enabling direct transactions between prosumers, bypassing intermediaries. Secondly, P2P trading promotes utilization of distributed energy resources, such as rooftop solar panels, by enabling prosumers to monetize their excess energy and sell it to other consumers in the network. This fosters a more sustainable energy ecosystem. Additionally, P2P trading enhances grid resilience as it allows for localized energy sharing, enabling communities to maintain electricity supply during grid disruptions. Lastly, through P2P trading, prosumers can sell their excess energy directly to other consumers, earning revenue that can offset their energy costs and reduce their bills. By participating in P2P trading, consumers can take greater control of their energy expenses and potentially lower their overall electricity bills.

3.2.1. Layers of P2P Energy Trading Platform

There are mainly two layers of P2P energy trading platform, including virtual and physical layers. The flow between these layers is shown in Figure 4 (Tushar, Saha, Yuen, Smith, & Poor, 2020).

Virtual Layer: The virtual layer provides an environment which is secured for deciding energy trading decisions to the platform's participants. It includes software platforms, communication protocols, and marketplaces that enable participants to connect, trade, and interact in a virtual environment. This layer facilitates the matching of energy supply and demand, price discovery, and transaction settlement. It also incorporates features such as smart contracts, blockchain technology, and data analytics to ensure efficient energy trading.

Physical Layer: The physical layer refers to the tangible components of the P2P energy trading system. It includes the physical infrastructure and assets involved in energy production, distribution, and consumption. This layer encompasses distributed energy resources (DERs) like solar panels, wind turbines, and energy storage systems, as well as the physical transmission and distribution networks that connect them. The physical layer also includes the measurement devices, meters, and sensors used to monitor energy flows and ensure accurate measurement of energy transactions.

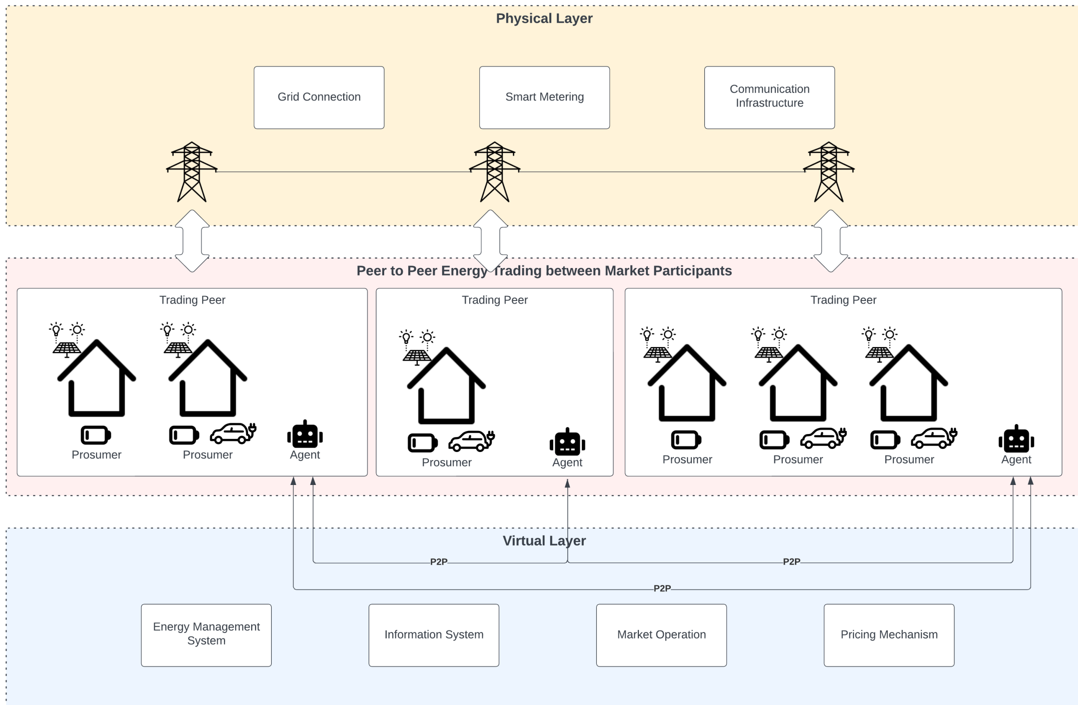

Figure 4: Layers of Peer-to-Peer Energy Trading Framework

The virtual layer and the physical layer work together to create a seamless P2P energy trading ecosystem. The virtual layer provides the digital infrastructure and tools necessary for participants to engage in energy trading, while the physical layer represents the actual generation, storage, and consumption assets that produce and consume the energy being traded. The integration of these two layers enables efficient and transparent peer-to-peer energy transactions, empowering participants and promoting the adoption of renewable energy sources.

In virtual layer, there should be some key elements:

- Information system: A well-functioning information system is vital to success of P2P energy trading platform. It serves as central component of the platform, connecting market participants, offering a suitable structure for trading, providing access to the market, and monitoring all operations. Equal access to market information, without any bias or interference, is crucial. One example of such an information system is smart contracts based on blockchain technology (Kang, Yu, Maharjan, Zhang, & Hossain, 2017).

- Energy management system: During engaging in P2P trading using a specific bidding structure, energy management system of prosumer ensures continuous supply of energy. To achieve this, EMS accesses real-time supply and demand data through smart meter of the prosumer. Using this information, EMS develops a generation and load plan for prosumers and makes bidding decisions on their behalf in the trading market.

- Market operation: P2P platform's information system plays a key role in enabling market operations, which cover market allocation, billing rules, and structured bidding. The primary objective of market operations is to facilitate an influential energy trading process by effectively matching orders and bids in real time. It is crucial to establish various market time horizons that can consistently provide adequate energy allocation throughout the entire market operation.

- Pricing mechanism: Pricing mechanisms play a crucial role in P2P trading by efficiently balancing energy supply and demand. Unlike traditional electricity markets, P2P pricing is not burdened by surcharges and taxes, allowing prosumers to set prices for their energy and maximize profits. These structures need to represent energy state within market network, with greater energy excess leading to lower prices and vice versa.

The elements in the physical layer are listed below:

- Grid connection: In P2P energy trading architecture, well-defined connection points with main grid are crucial for balancing energy generation and consumption. Smart meters can be installed at these points to evaluate network performance and energy savings. If a physical microgrid-distribution network exists, it can separate from main grid during emergencies.

- Metering: For involvement in P2P trading, each prosumer requires appropriate metering infrastructure, including a smart meter additively classical electricity meter.

- Communication infrastructure: Effective communication is vital in P2P trading for information exchange within the network. The communication structure should fit to requirements of performance, including considerations of latency, throughput, reliability, and security.

Moreover to these, market participants in P2P energy trading infrastructure are an important element. P2P energy trading requires an adequate number of market participants, and a subset of these participants should have the capability to generate energy. The purpose of P2P energy trading, such as promoting renewable energy usage or reducing reliance on main grid, has an impact on the design of pricing and market mechanisms (Tushar, et al., 2018). Additionally, it is necessary to define the specific form of energy being traded, whether it is electricity, heat, or a combination of both. For this thesis study, only electricity is considered.

3.2.2. Energy Models of P2P Energy Trading Platform Elements

A microgrid or prosumer, comprised of energy generation system, energy storage system, controlled electrical loads, electrical vehicle, and diesel generator, brings together various components to ensure a dependable power supply and maintain balance between generated and consumed electricity (Zahraoui, et al., 2021; Jiayi, Chuanwen, & Rong, 2008).

a) Loads

The electricity consumption of a building or microgrid is typically composed of various elements, including lightning, HVAC, appliances, and plugs. These elements collectively contribute to overall electricity consumption within a building or microgrid. Monitoring and managing these consumption elements are crucial for energy efficiency, load balancing, and optimizing overall energy usage. The local climate and weather conditions influence the demand for heating or cooling. Extreme temperatures may lead to higher energy consumption for maintaining comfortable indoor conditions. To overcome the effects of variability of energy consumption in P2P energy trading, robust load forecasting models should be used. The models used in the thesis study are elaborated in the following sections.

b) Solar Panels

There are different types of energy generation resources, including wind, solar, and hydro plants as renewable energy resources and natural gas, and coal plants as nonrenewable energy resources. In this study, peers are using solar panels as renewable energy resources. Using past generation characteristics of the panels, generation forecast models are implemented. The structure of these models is explained below sections.

c) Battery

Prosumers may utilize energy storage systems, such as batteries or pumped hydro storage, to store excess energy generated during periods of high production (Sarda, Lee, Patel, Patel, & Patel, 2022). This stored energy can be used when energy demand exceeds generation. Battery capacity is defined in kWh and shown with C. The unit of nominal power of the battery is kW.

$$C \geq 0 \tag{5}$$

Another parameter of the battery is capacity loss coefficient (c_{loss}). It refers to the rate at which battery's capacity decreases over time. It is a measure of battery's ability to store and deliver energy relative to its original capacity. This coefficient is influenced by various factors, including battery chemistry, usage patterns, operating conditions, and aging effects. The capacity loss coefficient provides insights into battery's performance degradation and is used in battery modeling and simulations to estimate its remaining capacity and overall efficiency. The new capacity of the battery after completion of each cycle and, complete discharge and recharge process of the battery, can be calculated by the number of finished cycles according to the below formula.

$$C_{n+1} = c_{loss} * \# \text{ of cycles (n)} * C_0 \tag{6}$$

Capacity power curve is another key parameter of battery and represents the relationship between battery's capacity and its power output (rate of energy delivery) over a given period of time. The curve illustrates the maximum power (P_t^{max} where t is time, P is power) that the battery can provide at different levels of its remaining capacity. As the battery's capacity decreases, the available power output may also decrease due to internal resistance and other factors. This relationship is important in determining the battery's performance and its suitability for various applications.

State of charge (SoC), the amount of energy stored in a battery relative to its maximum capacity, of a battery typically has upper and lower limits to ensure its safe and efficient operation.

$$DoD \leq SoC \leq SoC_{max} \tag{7}$$

where DoD is deep of discharge level of a battery.

Round trip efficiency (η_{eff}) is another measure used to evaluate the energy efficiency of a battery. It can be function depending on charging and discharging power of the battery. The relation between this power and efficiency is shown below figure.

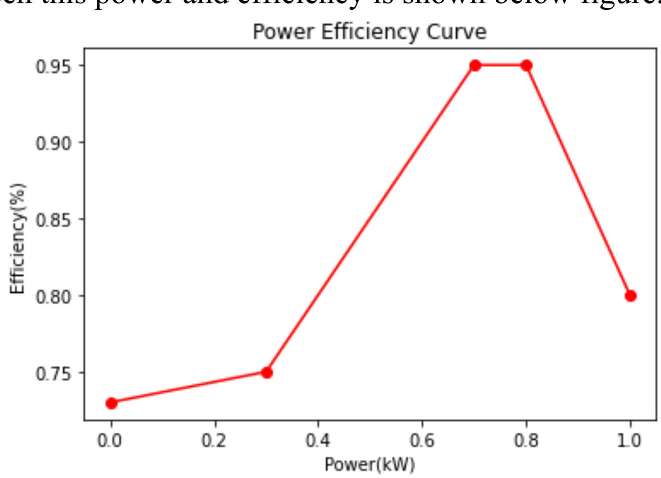

Figure 5: Power Efficiency Curve of Battery

SoC of a battery is influenced by the amount of energy charged into or discharged from the battery. When energy is added to the battery during charging, SoC increases, and when energy is withdrawn during discharging, SoC decreases (Sarda, Lee, Patel, Patel, & Patel, 2022). The relationship between charging/discharging amount and the resulting change by time t in SoC can be described by the following equation:

$$SoC_{t+1} = SoC_t + \frac{E_c}{\sqrt{\eta_{eff}}} - E_D * \sqrt{\eta_{eff}} \tag{8}$$

where E_C is charging amount, E_D is discharging amount.

3.3. Methodology

Implementation of reinforcement learning based P2P energy trading solution for composed of multiple peers that have different capacities of load, generation, and battery, is presented in Figure 6. In this solution, each peer is equipped with its reinforcement learning agent. The primary goal is to train these agents to effectively cooperate and coordinate with one another, even when they initially start with random policies and lack knowledge about the system dynamics. The focus is on optimizing energy consumption of each within the grid by monitoring the overall load profile.

The performance evaluation of the agents centers around several key energy-related metrics. These include minimizing yearly peak demand, reducing daily peak demand and total load, and the details of them are in the following sections. By achieving improvements in these metrics, the agents demonstrate their ability to efficiently manage energy consumption in the grid and enhance overall energy efficiency and stability of the grid.

Through the implementation and evaluation of agents based on these metrics, the aim is to develop effective coordination strategies that enable the peers to collectively optimize their energy usage. This approach holds the potential to contribute to development of intelligent energy management systems and enhancing sustainability.

The algorithms behind load, generation, price forecasts, and reinforcement learning parts are examined separately in the coming parts of the thesis study. This approach ensures that each algorithm's impact and influence on overall energy management system can be thoroughly explored, providing valuable insights for future improvements and optimizations.

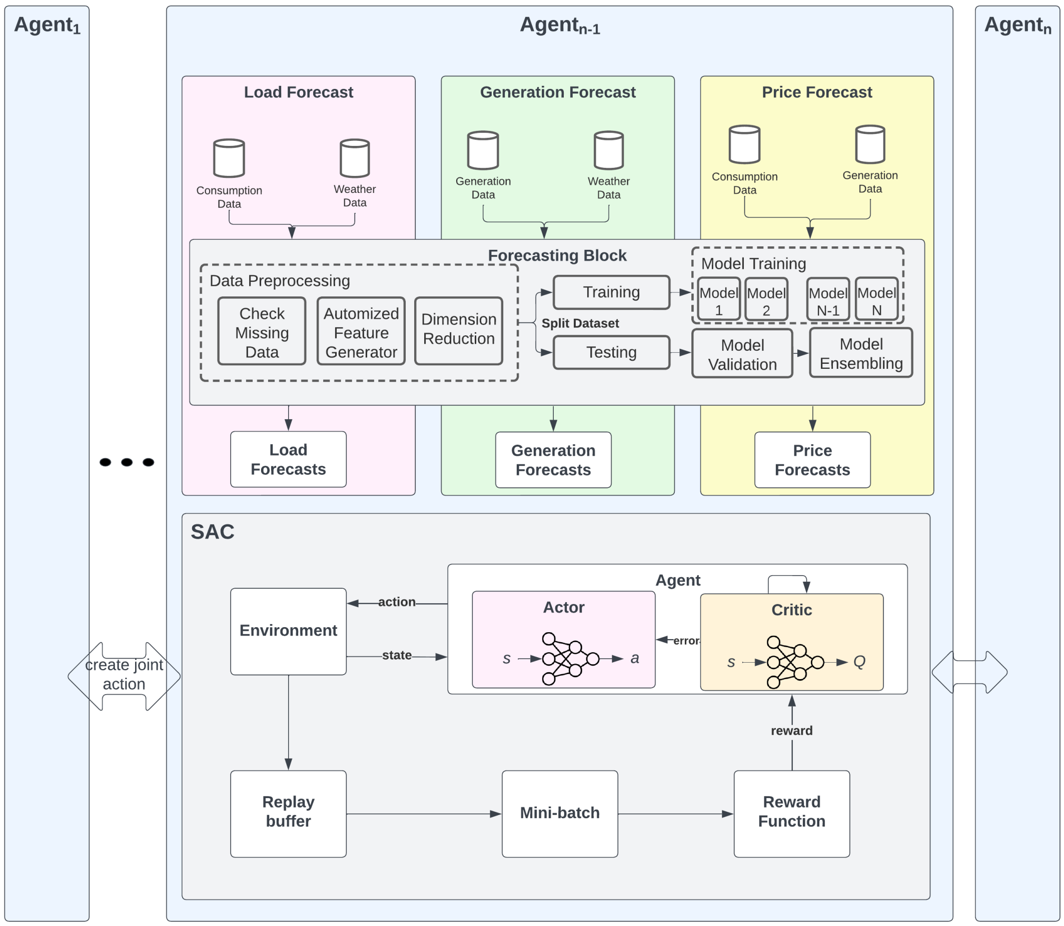

Figure 6: Framework for P2P Energy Trading Algorithm

3.3.1. Forecasting Approach

Load, generation, and price forecasting often share similar model structures due to their inherent dependencies. While each forecasting task focuses on different aspects of energy market, they often rely on similar data sources and utilize comparable techniques for modeling and prediction (Hong, et al., 2020).

One key reason for the similarity in model structure is shared reliance on historical data (Hong, Energy forecasting: Past, present, and future, 2014). Load, generation, and price forecasting models all benefit from historical data that captures past trends, seasonality, and patterns in energy consumption, generation, and market dynamics. The use of historical data allows for the identification of recurring patterns and helps in making informed predictions about future behavior. Another reason for the similarity in model structure is the common utilization of statistical and machine learning techniques. These

techniques allow for the identification of correlations, trends, and dependencies between various factors affecting load, generation, and price.
The dataset and feature engineering techniques for these datasets are explained firstly for each energy variable.

a) Load Forecasting

The main characteristics of electricity load depend on industry, weather, and holidays. The effects are analyzed using Turkey's total load dataset. Workdays and weekends show different patterns due to industry and commercial activities. Over a day, the load is the lowest level in the middle of the night (3 am - 5 am) and the highest level at noon (1 pm - 3 pm) depending on the season. The general daily trend can be understood from Figure 7. Also, by looking at this graph, the effect of air conditioning or temperature on demand can be understood. Especially in the summer months, the air conditioning effect causes the demand to increase between 15% and 20% compared to the spring months. Similar situation is observed in winter with an increase of about 10% due to the decrease in temperature.

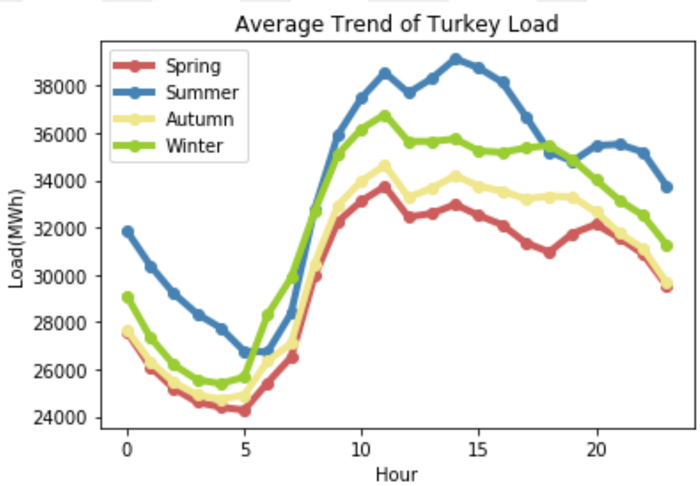

Figure 7: Daily Trend of Turkey by Hour

Within a week, the load is changing according to day type. Tuesday, Wednesday, and Thursday show the same pattern at all hours. On Mondays, the midnight load is less than on weekdays because of the activity recovery effect of industry. On Friday afternoons, the load decreases when it is compared with other weekdays because of the weekend effect. On Saturdays, not all companies and industries work. Because of that, the load changes between weekdays and Sundays. Sundays are like a holiday because of that the load is the lowest daily level in a whole week. All these patterns are shown in Figure 8.

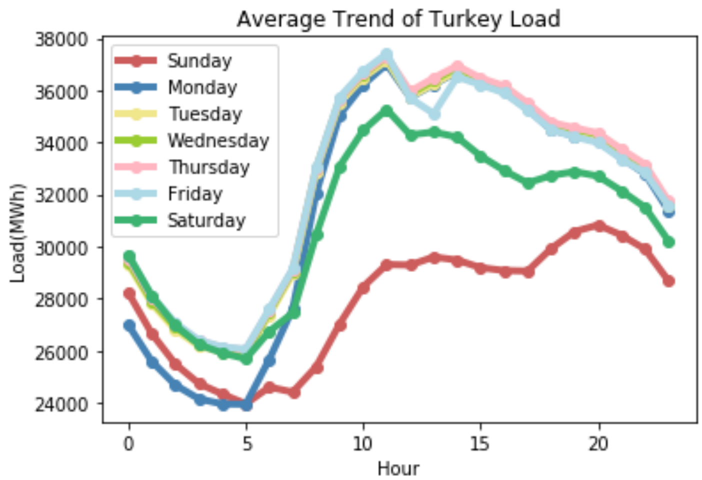

Figure 8: Daily Trend of Turkey according to Day Type

Apart from normal typical days, there are different kinds of days in each country called holidays. These atypical days are nonworking days, but they have their features separated from each other like national or religious days. Generally, national days are on fixed dates but not fixed on weekdays. In other words, day type can be changed year by year for 1 January New Year. This causes problems when generating forecasts. The habit of people changes when the holiday coincides with Friday or Monday. They can combine weekends and holidays. Moreover, if the holiday happens on Tuesday or Thursday, people do not work also on Monday or Friday. Also, the days before the holidays are announced as arefe in some holiday types such as Victory Day or Ramadan Eid. These days, people work half of the day. The eids in Turkey take a longer period. If Ramadan or Sacrifice Eid starts on Monday or Tuesday, people combine two weekends before and after eid, and this means nine days of holiday. All these different types are summarized and given to the models in Table 1.

Table 1: Holiday Types in Turkey

Holiday ID	Holiday Name
1	New Year
2	National Sovereignty and Children's Day
3	1 May Work and Solidarity Day
4	Atatürk Commemoration Youth and Sports Day
5	Arefe of Ramadan Eid
6	Ramadan Eid
7	Arefe of Eid Al-Adha
8	Eid Al-Adha
9	Arefe of Republic Day
10	Republic Day
11	15 July Democracy and National Union Day
14	Victory Day

The pattern of typical days and the pattern of holidays, atypical days, in Turkey, are compared in Figure 9. As can be understood from the figure, the pattern of the holidays is like a typical Sunday.

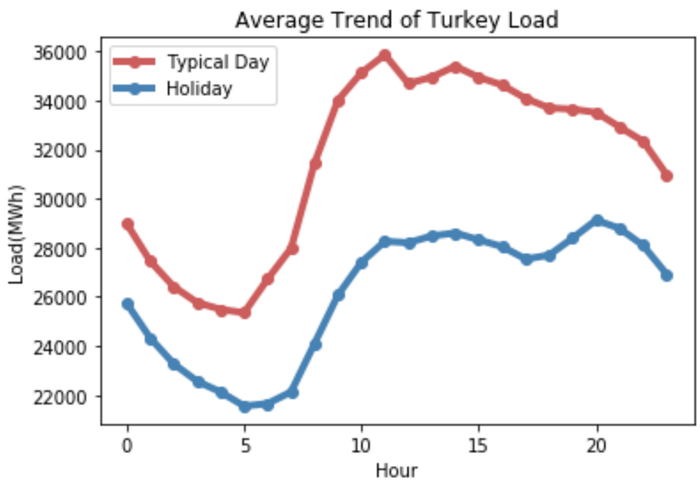

Figure 9: Daily Trend of Turkey at Holidays

In load forecasting problems, independent variables are selected according to the effects on the demand. Calendar features like weekdays, hours, and seasons; numerical weather parameters like temperature, and humidity; and holidays such as national, and religious holidays or extreme days such as lockdown affect the load. The details of all datasets can be found in Table 2.

Table 2: Summary Table of Datasets Used in Load Forecasting

Dataset Name	**Detail**	**Type**	**Description**
Load		Time Series-Quantitative	Hourly total load
Temperature	Santigrat	Time Series-Quantitative	Hourly temperature
Cloud Cover	Percentage(0-1)	Time Series-Quantitative	Hourly cloud cover percentage
Humidity	Percentage(0-1)	Time Series-Quantitative	Hourly humidity percentage
DataYear	2016,…2020	Categorical	Year of a data point
DataMonth	1,2,…,12	Categorical	Month of a data point
DataDay	1,2,…,31	Categorical	Day of a data point
DataHour	0,1,…,23	Categorical	Hour of a data point
DataWeekDay	1,2,..,7	Categorical	Weekday of a data point
DataSeason	1,2,3,4	Categorical	Season of a data point
IsHoliday	0,1	Binary	Shows the day is a holiday or not

Effective load forecasts rely on feature engineering to capture fundamental patterns of electricity demand. Including lagged values of historical load data as additional features allow model to capture autocorrelations and dependencies in electricity demand over time. Moreover, giving binary features for holidays and special days' marks is important. In load forecasts, weather data plays a significant role in understanding how external conditions influence electricity demand. Heating and cooling degree days, represent the difference between average outdoor temperatures and comfort temperature that is between

18-20 in most of the places. These values reflect heating or cooling requirements and correlate with electricity demand for heating or cooling purposes. Creating new interaction features by combining different weather variables such as temperature multiplied by humidity to represent the joint effect of weather conditions.

b) *Solar Generation Forecasting*

Solar generation forecasting is challenging due to the inherent variability and intermittency of solar energy production, which depends on factors like cloud cover, atmospheric conditions, and the position of the sun. Continuous improvements in weather data collection, advanced modeling techniques, and machine learning algorithms have significantly enhanced the accuracy of solar generation forecasts, contributing to the increased integration of solar power into the grid and better overall energy management. When developing machine learning models for solar generation forecasting, the input set typically consists of various features that influence solar energy production. The choice of input features depends on the forecast horizon (short-term, medium-term, long-term), data availability, and specific characteristics of the solar power system being modeled (Rahimi, et al., 2023). Because P2P energy trading needs short-term forecasts, the necessary features are examined according to that.

The geographical location of solar plants can affect solar irradiance levels and, consequently, power generation. This information is vital for accurate regional forecasting. The location has a significant impact on solar generation due to variations in solar irradiance and other related environmental factors. Solar irradiance is the amount of sunlight received at a particular location that varies with latitude, longitude, altitude, and local climate. Also, geographical location affects daylight hours throughout year, sun angle at which sunlight strikes the Earth's surface varies and local climate, weather patterns impact the amount of cloud cover, atmospheric conditions, and potential shading from natural features. These effects can be seen in Figure 10 shows total solar generation in Turkey for consecutive days. Solar irradiation, cloud cover, and sun angle change the amount of generated solar energy. For instance, day 2 is more cloudy and rainy day in the whole of Turkey and the amount of generated energy is low according to the installed power of solar.

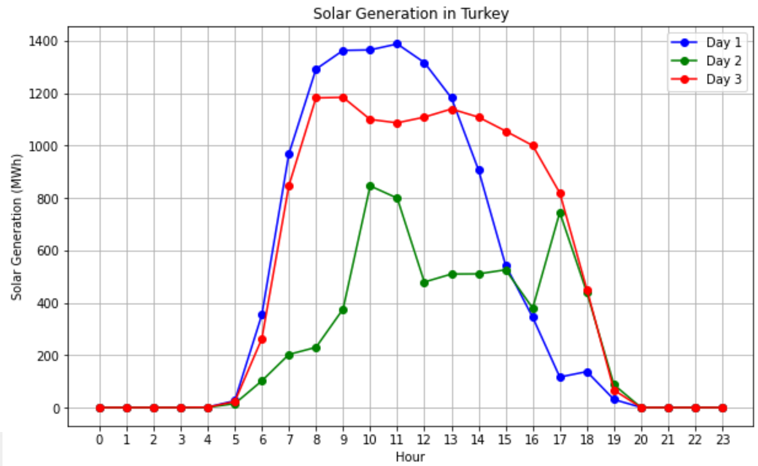

Figure 10: Solar Generation for Turkey on Consecutive Days

Because the effects of climate are radical, meteorological data and numerical weather predictions are necessary to forecast solar generation. Common meteorological features include solar irradiance with its components including direct and diffuse irradiance, ambient temperature, wind speed, relative humidity, cloud cover, and atmospheric pressure (Rahimi, et al., 2023). For short-term solar generation forecasting, incorporating real-time or near-future forecasted weather data can improve prediction accuracy.

Apart from weather data, information about the duration of daylight in a day, which varies with season and location, can be useful for understanding when solar generation will occur. Because of that, season, month, and hour features help capture temporal patterns and seasonal variations in solar energy generation.

Feature engineering plays a crucial role in improving the accuracy and performance of machine learning models for solar generation forecasts (Wu, Huang, Phan, & Li, 2022). By carefully selecting and transforming relevant features, the models can better capture the underlying patterns and relationships within the data. Calculating moving averages of solar power generation over a certain window of time can help smooth out noise and highlight underlying trends in generation. Moreover, for capturing periodic patterns in solar generation data, Fourier transforms can be used to identify dominant frequencies, which helps model daily and seasonal variations. Aggregating weather-related data, such as daily mean or maximum temperature, total precipitation, or average cloud cover, can provide a concise representation of weather conditions over specific periods. Creating interaction features between different meteorological variables, such as solar irradiance and temperature, can help capture nonlinear relationships that affect solar energy. For instance, multiplying or taking the product of temperature and solar irradiance values can improve model accuracy. The effectiveness of these feature engineering techniques depends on the dataset and machine learning algorithm being used. It's essential to experiment with different combinations of features and transformations to find the best configuration for each forecasting task. The used features and the trials are explained in the experiment section of the thesis study.

c) *Price Forecasting*

Electricity price forecasting is the process of predicting future electricity prices in energy markets. It is a critical aspect of energy market analysis, helping market participants, energy companies, grid operators, and consumers make informed decisions related to electricity purchasing, generation, and consumption (Weron, 2014). The input set for electricity price forecasting using machine learning models typically consists of historical price, demand, weather, and generation data. Time-lagged features utilize past price data as inputs to anticipate future price trends. By including lagged features, model can recognize correlations within time series data, which are fundamental for comprehending price fluctuations and patterns (Lago, Marcjasz, Schutter, & Weron, 2021). Extracting various time-related features such as weekday, month, hour, and seasonal indicators can help the model capture recurring patterns in electricity prices. Price curve with similar trends according to hours is visible in Figure 11, while variations based on weekdays are also apparent.

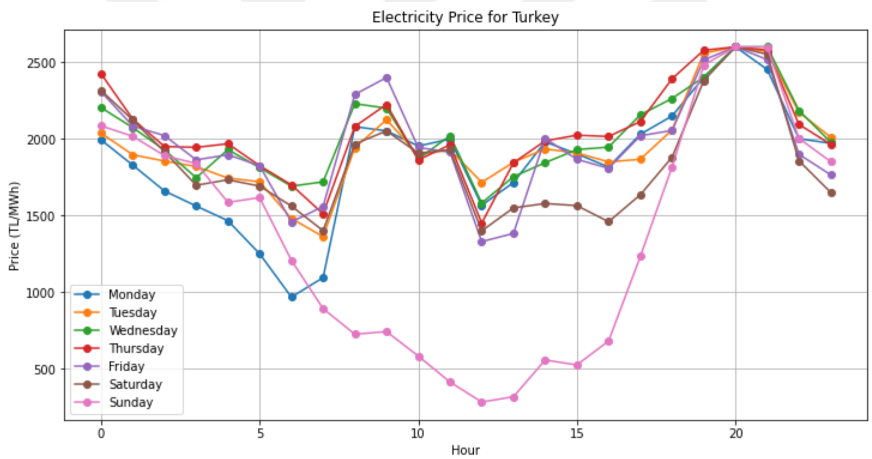

Figure 11: Daily Day Ahead Market Prices Averages of Turkey according to Day Type

Electricity demand is a significant driver of prices. Historical electricity demand data, both aggregated and at the individual consumer level, can be valuable in understanding demand-price relationship. Figure 12 shows the relationship between price and consumption. For instance, while consumption is decreasing on Sundays because of nonworking days, price is decreasing.

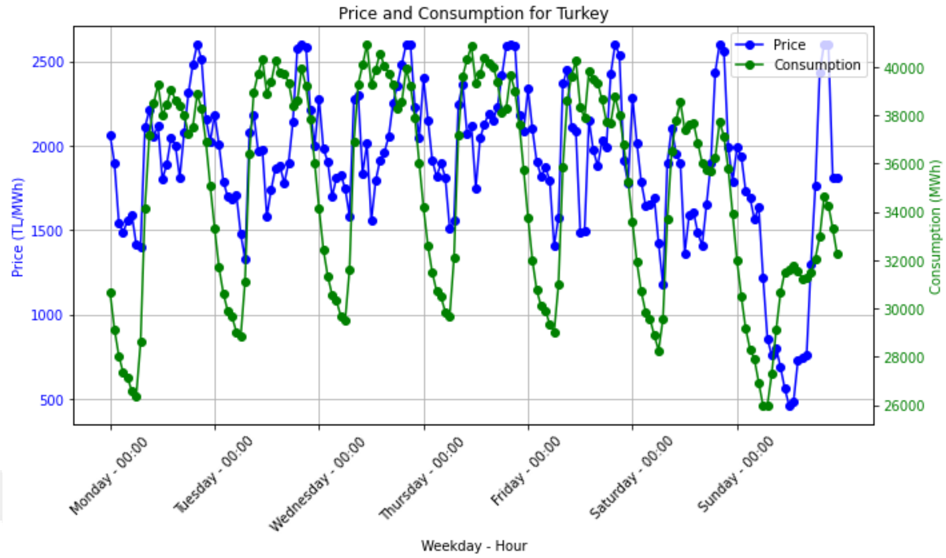

Figure 12: Day Ahead Market Prices vs. Consumption Averages of Turkey

Information about generation mix and output from various power plants including coal, natural gas, and renewables can be relevant for understanding price dynamics, especially in markets with competitive electricity generation. Renewable energy sources, like solar and wind, can suppress electricity prices during periods of high generation due to their low marginal costs. Non-renewable sources, such as natural gas and coal, are subject to fuel price fluctuations, directly increasing electricity generation costs. These effects can be seen in Figure 13.

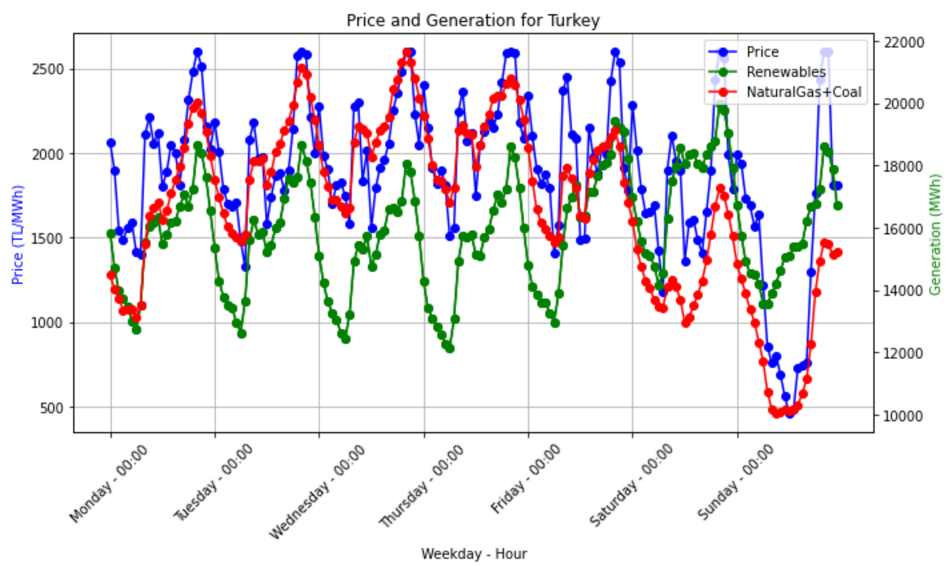

Figure 13: Day Ahead Market Prices vs. Renewables and Non-Renewables Generation Averages of Turkey

Actually, weather conditions can influence electricity consumption patterns and generation from renewable sources. Because of that, weather is directly affecting electricity prices. Including weather variables such as temperature, humidity, wind speed, and solar irradiance can help machine learning models account for weather related changes in demand and supply (Yang, Ce, & Lian, 2017).
In addition to direct features, feature engineering techniques should be applied to input datasets to improve machine learning models' accuracies. Calculating metrics such as price volatility and standard deviation can provide insights into price fluctuations. Moreover, including price differences between peak and off-peak periods can be informative for price forecasts.

d) Forecasting Algorithm

Energy forecasting using machine learning techniques involves predicting future energy consumption, demand, or generation based on historical data and relevant features. Solar generation, load, and electricity price forecasts can share the same machine learning approach because they are all related to energy domain and have similar underlying patterns and dependencies. While specific features like related weather parameters and data sources may vary for each type of forecast, fundamental principles of forecasting are consistent across them. The forecasting approach is shown in Figure 6. According to that, energy dataset should be applied to data preprocessing step to be ready for the machine learning model. Checking for missing values in the dataset and deciding on how to handle them are important to train the model with correct data. Removal of rows with missing data is preferred in the proposed approach instead of imputation because filling the missing should consider climate and holiday effects. After handling missing data, automated feature engineering techniques are applied as mentioned in previous sections to create new relevant features from raw data. If the final dataset has a large number of features, Principal Component Analysis, which is one of the dimension reduction techniques, is used to reduce the number of features while retaining important information.

The dataset is divided into training and testing sets. Training set is used to train the model while testing set is used to evaluate the model's performance on unseen data. LightGBM model is selected as a machine learning model for forecasting model (Park, Jung, Jung, RHo, & Hwang, 2021; Ju, et al., 2019; Deng, et al., 2021). It is a gradient boosting framework that utilizes decision trees as its base learners to progressively enhance model accuracy. The algorithm constructs decision trees in a top-down, greedy manner, selecting features that provide the most information gain at each node for data splitting. This recursive process continues until a stopping criterion, such as maximum tree depth or minimum samples in a leaf node, is met. By combining multiple decision trees through gradient boosting, LightGBM creates a powerful learner that achieves high accuracy. Because LightGBM is fast, scalable, robust to overfit, and needs low memory, it is preferred. For P2P energy trading model, there is a need for fast and usable model for all types of energy variables' forecasts. Hyperparameter tuning is essential for getting the best possible performance out of a LightGBM model. Built-in tuning function

LightGBMTuner is an automatic hyperparameter tuning method and a combination of random and grid search is applied. After the training process, the model is validated using a testing dataset. The state-of-art special performance metrics are calculated for each energy variable to assess its accuracy in predicting future energy values. After the model is trained and validated, it is used to make energy forecasts for future market periods.

Energy forecasting is a dynamic process that involves iterating through these steps and continually improving the model to make more accurate and reliable predictions, so the models are continuously trained for each peer.

3.3.2. Reinforcement Learning Approach

Reinforcement learning is an agent-based algorithm that learns by interacting with its environment to maximize cumulative rewards over time. The agent explores the environment, takes actions, and receives feedback in rewards (Sutton & Andrew). The goal is to find an optimal policy that maps states to actions for maximizing long-term rewards. RL can be model-free, making it suitable for unknown or complex environments. It involves estimating values of state-action pairs or learning a value function to guide decision-making. Mathematically, these problems are defined using Markov Decision Process (MDP). MDP consists of sets of states S and set of actions A, reward function $R(s, a, s')$ that maps states and actions to rewards, and transition probabilities $P(s, a, s')$ between states. The policy π maps states to actions, $\pi: S \rightarrow A$, and value function $V^{\pi}(s)$ represents the expected reward for the agent beginning in state s and pursuing policy π thereafter. The value function provides an estimate of long-term rewards agent can expect to accumulate from a given state. The policy maps states to actions, and the value function estimates the expected return for the agent beginning in a specific state and pursuing policy thereafter.

$$V^{\pi}(s) = \sum_{s' \in S} p(s' \mid s, \pi(s,a)[R(s,\pi(s,a),s') + \gamma V^{\pi}(s')] \quad \forall s \in S \tag{9}$$

where the reward taken after the following action a in state s, going to the next state s' is shown as $R(s,\pi(s,a),s')$.

The reward represents immediate feedback or desirability of state-action pair. The discount factor γ (where $\gamma \in [0, 1]$) is used to balance the importance of immediate rewards versus future rewards. When $\gamma = 1$, the agent considers future rewards to be just as important as immediate rewards. On the other hand, when $\gamma = 0$, greater emphasis is placed on immediate rewards, and the agent focuses less on future rewards.

Because there are no available environments' dynamics and transition probabilities, a model-free RL approach is used (Pong, Gu, Dalal, & Levine, 2020). Model-free RL approach learns directly from interactions with the environment, without assuming prior knowledge of transition probabilities. The agent explores the environment, takes actions, and receives feedback from rewards. It uses this experience to estimate the values of state-action pairs or learn a policy that maximizes expected cumulative rewards.

a) Q-learning

Q-learning is a widely recognized model-free RL technique in most of applications (Clifton & Laber, 2020). For environments with a small number of states, transitions can be denoted using a table which keeps state and action values known as Q-values. Each record in Q-table corresponds to state-action pair (s, a), and Q-values are updated using the following formula:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r(s, a) + \gamma \max_{a} Q(s',a) - Q(s,a)] \tag{10}$$

where, s' represents the next state, $r(s, a)$ is the reward taken by the following action a in state s, α is the learning rate that determines how much new knowledge overrides old knowledge, and γ is discount factor that balances the importance of immediate and future rewards.

After following action in a particular state, Q-values estimate the cumulative expected sum of discounted rewards, taking greedy policy from thereon. Q-learning is an off-policy method, meaning it updates its policy using historical experiences collected from different policies. These experiences, stored as state-action-reward values, are saved in the memory replay buffer. During the learning process, the buffer is sampled to iteratively update the Q-values using the Q-learning equation. By continuously updating the Q-values based on observed rewards and selecting actions that maximize the Q-values, Q-learning aims to learn an optimal policy that maximizes long-term cumulative rewards.

b) Actor-Critic Algorithm

In environments with continuous states and actions like energy trading problems, tabular Q-learning is impractical due to curse of dimensionality. The state and action spaces become too large, making it infeasible to store and update Q-values for each state-action pair individually. Actor-critic RL techniques tackle this challenge by employing neural networks to extend among state and action combinations. Neural network of actor translates states into optimal actions, whereas critic neural network reaches actions by associating them with their respective states and deriving Q-values. By leveraging neural networks, actor-critic methods efficiently handle continuous environments, enabling effective exploration and exploitation. Actor-critic methods learn three functions, including actor, critic, and value function.

Actor learns policy, which maps states to actions. Policy is typically represented by neural network that takes the state as input and outputs the parameters of a probability distribution over actions. Actor aims to maximize expected cumulative reward by selecting actions that yield high expected returns. Critic estimates soft Q-function, which measures the expected cumulative reward for a given state-action pair. Unlike traditional Q-learning, the selected Actor-Critic structure has soft Q-function that incorporates an entropy term, encouraging exploration (Haarnoja, et al., 2019). This architecture is called Soft Actor Critic (SAC), and its objective of it is not only to maximize lifetime rewards but also to maximize the entropy of policy. Entropy measures the unpredictability of a random variable. A policy with high entropy encourages exploration by assigning equal probabilities to actions with similar Q-values. It prevents the policy from repeatedly selecting a specific action that exploits inaccuracies in Q-function approximation.

SAC also learns the value function V, which estimates the expected cumulative reward starting from a given state. The value function is used to approximate state value and guide the learning process. The formula for updating the value function V(s) can be written as:

$$V(s_t) = E_{a_t \sim \pi_\theta}[Q(s_t, a_t)] + \alpha H \tag{11}$$

where H is the entropy of action distribution of policy π_θ in which state s_t and $\alpha \in (0, 1)$ is the temperature term that controls the importance of entropy.

If it is 1, then the entropy has maximum stochasticity, if it is 0 then entropy is ignored. Zero entropy gives deterministic policy and policies which have nonzero entropies are more random selections of actions. The main objective is based on a maximum entropy reinforcement learning model that tries to find optimal policy maximizing expected long-term return and entropy. The objective function is:

$$J(\pi_\theta) = E_{\pi_\theta}[\sum_{t=0}^{T-1} \gamma^t R(s_t, a_t) + \alpha H(\pi(.|s_t))] \tag{12}$$

According to the given objective function, the optimal policy can be found below the equation which seeks the highest long-term reward and entropy.

$$\pi^* = argmax_{\pi_\theta} E_{\pi_\theta}[\sum_{t=0}^{T-1} \gamma^t R(s_t, a_t) + \alpha H(\pi(.|s_t))] \tag{13}$$

The critic networks are updated according to the minimization of expected error which is calculated by differencing prediction of value network and the expected value of Q function (Haarnoja, et al., 2019).

$$J_Q = E_{(s_t, a_t) \sim D}\left[\frac{1}{2}(Q_\theta(s_t, a_t) - (R(s_t, a_t) + \gamma E_{s_{t+1}}[(V_\theta(s_{t+1})]))^2\right] \tag{14}$$

where *D* is the replay buffer.

By simultaneously updating actor, critic, and value functions, SAC learns a policy that maximizes expected cumulative rewards while also considering exploration through entropy maximization term. This combination of actor-critic architecture, maximization of entropy, and off-policy updates allows SAC to efficiently explore and learn in continuous state and action spaces.

c) Multi-Agent RL

The P2P energy trading problem involves buildings/microgrids making independent decisions regarding energy trading and load scheduling in a dynamic and uncertain environment. To address this problem, the proposed solution involves using a multi agent deep reinforcement learning method. This method enables agents to coordinate through reward sharing and mutual information sharing. Moreover, each peer can make decisions based on local observations and learn through trial and error. In this problem, a separate RL agent is assigned to each peer. The objective of these agents is to learn, starting from random policies and without prior knowledge of the system dynamics (Vazquez-Canteli, Henze, & Nagy, 2020; Wang, Li, & Zhang, 2022). The agents aim to minimize the overall load of the market, and their performance is evaluated based on cost metrics such as minimizing yearly peak demand, daily peak, and ramping rate while maximizing the average daily load factor of entire peers.

State:

The state vector s_t^n of peer n at time t consists of various variables including:

- Solar panel generation: $E_{t,PV}^n$ represents the amount of solar energy generated by peer n at time t.
- Consumption: $E_{t,L}^n$ represents the amount of consumed energy by peer n at time t.
- SoC level of Battery: $SoC_{t,B}^n$ represents the state of charge level of the battery belonging to peer n at time t.
- Buying and selling prices: p_t^B represents the buying price of electricity at time t, while p_t^S represents the selling price of electricity at time t.

Therefore, the state vector can be expressed as:

$$s_t^n = [E_{t,PV}^n, E_{t,L}^n, SoC_{t,B}^n, p_t^B, p_t^S] \tag{15}$$

This vector captures relevant information for peer n at time t, enabling it to make decisions regarding energy trading and load scheduling in the given dynamic and ambiguous environment. In the thesis study, buying and selling prices of the market are assumed to same.

Action:

The action vector a_t^n for peer n at time t consists of various variables including

- Buying energy: $a_{t,B}^n$ represents estimated energy amount to buy from other peers.
- Selling energy: $a_{t,S}^n$ represents estimated energy amount to sell to other peers.
- Charging: $a_{t,C}^n$ represents estimated energy amount to charge the battery.

- Discharging: $a^n_{t,D}$ represents estimated energy amount to discharge the battery.

Therefore, the action vector can be expressed as:

$$a^n_t = [a^n_{t,B},\ a^n_{t,S}, a^n_{t,C}, a^n_{t,D}] \tag{16}$$

These components of the action vector a^n_t capture the different decision variables involved in energy trading and load scheduling for household n at time t.

Reward:

Reward function r^n_t represents immediate benefit obtained by peer n at time t when taking action a^n_t based on state s^n_t. It can be expressed as:

$$r^n_t(s^n_t, a^n_t) = -\ E^n_t \tag{17}$$

where E^n_t is the net electricity consumption of peer n at time t.

The negative sign indicates that the reward function is typically designed to be minimized or reduced, implying that each peer aims to minimize cost or maximize utility.

Algorithm:

To overcome the challenges of P2P energy trading problem, a novel multi-agent deep reinforcement learning method is introduced. In this approach, the policy is trained by utilizing past shared state-action, which is taken and stored in replay buffer. The proposed approach is based on an actor-critic architecture. Following the policy training, peers have ability to read output values obtained from their critic networks. Based on the outputs and their individual state vectors, each peer can then make deterministic actions using their actor networks. In the algorithm, training takes place in the critic network, while execution is employed in the actor network. The agent of each peer employs its actor network to make deterministic actions based on local states. Replay buffer is used to store experiences including state, action, reward, and next state.

The collected joint state and action data from all peers are denoted as $s_t = \{s^1_t, s^2_t, \dots, s^n_t\}$ and $a_t = \{a^1_t, a^2_t, \dots, a^n_t\}$, respectively. By utilizing this multi-agent RL approach, optimal strategies can be computed for P2P energy trading, enabling efficient coordination among peers. Step-by-step algorithm is explained below:

Algorithm 1: Multi Agent Deep Reinforcement Learning Algorithm for P2P Energy Trading

for peer i in *(1, 2, …, n)* **do**
    initialize actor and critic networks
    initialize target network
    initialize replay buffer D_i
end for
for episodes e in *(1, 2, …, m)* **do**
    initialize random process for action
    observe initial state for each peer
    for time t in *(1, 2, …, k)* **do**
        for peer i in *(1, 2, …, n)* **do**
            observe current state s_t^i
            select action based on current state s_t^i
            execute a_t^i
            observe reward r_t^i
            observe next state s_{t+1}^i
            store in D_i
            take sample mini-batch from D_i
        end for
        for peer i in *(1, 2, …, n)* **do**
            update actor and critic networks
        end for
        update target network
    end for
end for

CHAPTER 4

EXPERIMENT

4.1. Experimental Setup

To simulate this peer-to-peer energy trading game, data was collected from 8 prosumers who participate in peer-to-peer energy trading network. The data was obtained from the real-time monitoring system of Inavitas[1], covering a period of one year. The prosumers comprise three commercial buildings, including shop and supermarket, and five households with different numbers of households living there. Each prosumer has solar generation installed on their rooftops, contributing to their energy production capacity. PV generation data is analyzed to understand the solar energy generation profiles of each prosumer and to identify any seasonal variations in solar generation throughout the year. Additionally, energy consumption patterns of both commercial buildings and households are studied to determine their energy demands. To facilitate the peer-to-peer energy trading process, a demand-supply matching mechanism is devised. This matching process aims to align the energy demand of households and commercial buildings with the available solar generation from their solar panels. The surplus energy generated by prosumers can then be traded with others within the network. In addition, the experimental setup includes information about battery storage systems for all 8 prosumers. Each prosumer has a battery installed, which allows them to store excess energy generated by their solar systems or store energy during off-peak hours for later use.

The experimental setup includes a realistic simulation scenario that closely mirrors the real-world peer-to-peer energy trading environment. Factors such as geographical locations, weather conditions, and real-time energy prices are incorporated into the simulation model to create a dynamic and authentic market platform.

To evaluate the performance of P2P energy trading system, various metrics are defined following sections for forecasting models and reinforcement learning approach separately. The experimental results are validated against real-world scenarios wherever possible. Sensitivity testing is conducted to understand how a system's performance changes with variations in energy demand, solar generation, and market conditions.

---

[1] www.inavitas.com

4.1.1. Dataset

The real time data of 8 prosumers were collected for P2P energy trading problem. The general features of these prosumer are shown in Table 3. Because of the physical constraints of the grid, geographically near prosumers were selected. The data was collected from the region of Urla in İzmir, Turkey. The dataset contains hourly load and generation data of each prosumer between 2022-01-01 and 2023-05-31. The expected number of data points for this period is 12384. However, some data points are missing.

Table 3: Prosumer Data Details

Name	Type	Demand(kW)	Solar Installed Power(kW)	Battery Capacity(kWh)	Charge/Discharge Efficiency of Battery(%)
commercial_1	grocery	10.4	7	9.8	95%
commercial_2	shop	13.1	9	12.6	95%
commercial_3	grocery	6.8	3	4.2	95%
house_1	house	1.1	0.3	0.4	90%
house_2	house	0.8	0.3	0.4	90%
house_3	house	0.9	0.3	0.4	90%
house_4	house	0.9	0.3	0.4	90%
house_5	house	1.2	0.3	0.4	90%

To use this dataset for forecasting load and generation, weather data was collected from Meteomatics API[2]. This weather data includes weather parameters listed below table.

Table 4: Weather parameters in Dataset

Parameter	Unit	Description
Temperature	°C	Measure of heat or coldness in atmosphere
Cloud Cover	%	Fraction of sky covered by clouds
Global Radiation	W/m²	Total solar radiation received on earth's surface
Diffuse Radiation	W/m²	Solar radiation reaching earth's surface indirectly
Direct Radiation	W/m²	Solar radiation reaching darth's surface in straight line from sun
Relative Humidity	%	Amount of water vapor in air relative to the maximum possible at the same temperature
Wind Speed	m/s	Total solar radiation received on earth's surface
Precipitation Probability	%	Likelihood of precipitation occurring at a given location
Apparent Temperature	°C	Perceived temperature that accounts for combined effects of temperature, humidity, and wind
Sun Elevation	°	Angle between horizon and center of sun

[2] https://www.meteomatics.com/en/weather-api/

a) Data Preprocessing

After the dataset collection procedure was completed, data preprocessing was done. The data was put outlier detection to identify and remove any extreme values that may significantly deviate from the typical pattern. Outliers could be due to measurement errors or unusual events. Z-score outlier detection method was used to detect problematics data in load. It is a statistical technique used to identify extreme values in a dataset (Rousseuw, 2011). It measures how many standard deviations the data point is away from the mean of the dataset. Z-score is calculated using the below formula:

$$z = \frac{x - \mu}{\sigma} \tag{18}$$

where x is the value of the data point, μ is the mean of the dataset and σ is the standard deviation of the dataset.

Typically, Z-score greater than the threshold is considered an outlier. The threshold value is usually set to 3, depending on the level of strictness desired in identifying outliers (Abdi, 2007). In this study, 3 was used as a threshold. Z-score method is widely used for identifying outliers as it is simple to implement and can be applied to various types of data. For the dataset of this study, Z-score method was applied, and the number of data points shown in the below table were detected for each peer's load. These detected outlier data points were converted to missing values.

Table 5: Number of Detected Outliers in Load Data

Name	Number of Outliers
commercial_1	5
commercial_2	25
commercial_3	121
house_1	333
house_2	406
house_3	389
house_4	224
house_5	449

Any missing values in the load and generation data were identified and addressed. Missing data points could be the result of various reasons such as communication issues, or data recording errors. Moreover, the detected outliers were converted to missing values. To fill these missing values, different approaches were used for the load and generation data.

For filling missing data in load series, data points at same hour and day of type for last two weeks of each missing data point were used. The averages of these data points were used to fill missing places. It is an approach that leverages temporal patterns of load data to estimate missing values in meaningful way. Actually, this method is useful to capture seasonality and periodicity inherent in load data. Limited time window, last two weeks of data for imputation balances need to capture recent patterns without using too distant data, which might not be representative of the current behavior of the data. The number of missing data points is shown in below table for each peer's load.

Table 6: Number of Missing in Load Data

Name	Number of Missings
commercial_1	136
commercial_2	216
commercial_3	546
house_1	792
house_2	359
house_3	927
house_4	325
house_5	365

Using realized global radiation data from weather observations for filling missing values in generation data was used for imputing missing generation values. It is based on the assumption that the amount of solar radiation received on earth's directly influences the electricity generation from solar panels. Solar generation is directly correlated with the amount of solar radiation available. On sunny days with higher global radiation, solar panels can generate more electricity, and vice versa on cloudy days with lower global radiation. The gathered global radiation has more than 0.9 correlation with each peer's generation data. A scaling factor based on the relationship between solar generation and global radiation was calculated. These factors were multiplied with related global radiation values which is at missing data's datetimes. This gives an approximation of the missing generation. The number of missing data points is shown in below table for each peer's generation.

Table 7: Number of Missing in Generation Data

Name	Number of Missings
commercial_1	2652
commercial_2	3563
commercial_3	2644
house_1	2640
house_2	2658
house_3	2641
house_4	2638
house_5	3582

b) *Data Exploration*

For analyzing characteristics of peers' load data, descriptive statistics were prepared in Table 8. According to the mean of each peer's load, commercial peers' consumptions are higher than houses. Houses have similar consumption levels according to percentiles, minimum and maximum values.

Table 8: Descriptive Statistics of Load Data

	commercial_1	commercial_2	commercial_3	house_1	house_2	house_3	house_4	house_5
count	12384	12384	12384	12384	12384	12384	12384	12384
mean	3.80	3.39	1.10	0.10	0.13	0.05	0.09	0.07
std	1.79	2.88	0.61	0.08	0.07	0.04	0.05	0.07
min	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
25%	2.35	0.47	0.57	0.05	0.09	0.02	0.06	0.03
50%	3.35	3.06	1.06	0.06	0.11	0.04	0.08	0.04
75%	5.18	5.09	1.48	0.10	0.14	0.06	0.11	0.07
max	10.43	13.11	6.87	0.42	0.40	0.25	0.30	0.39

To understand deeper underlying patterns of load data, monthly total values were plotted in Figure 14. However, monthly total load data is not sufficiently intuitive. Since seasonal climate effects could not be observed in this dataset. Other factors including holidays, and human attitudes can influence electricity consumption patterns.

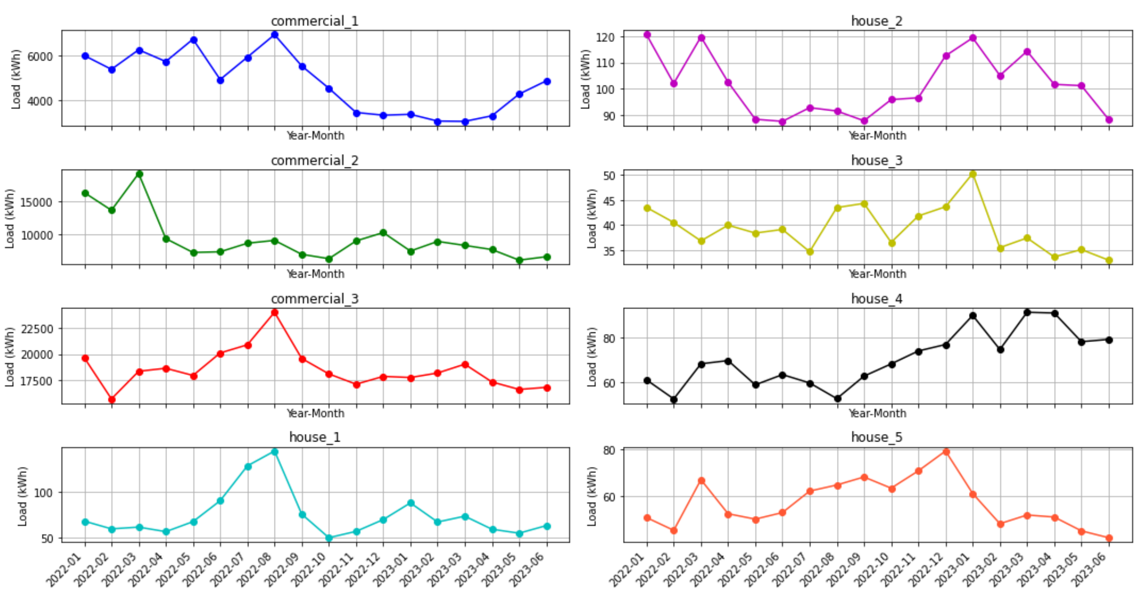

Figure 14: Monthly Total Load of Peers

Day of week and hour averages of whole load data of each peer shows weekly and daily seasonality (Figure 15). Commercial peers consume similar every day of week to provide comfortable shopping environment. Some houses consume more on weekends than on weekdays.

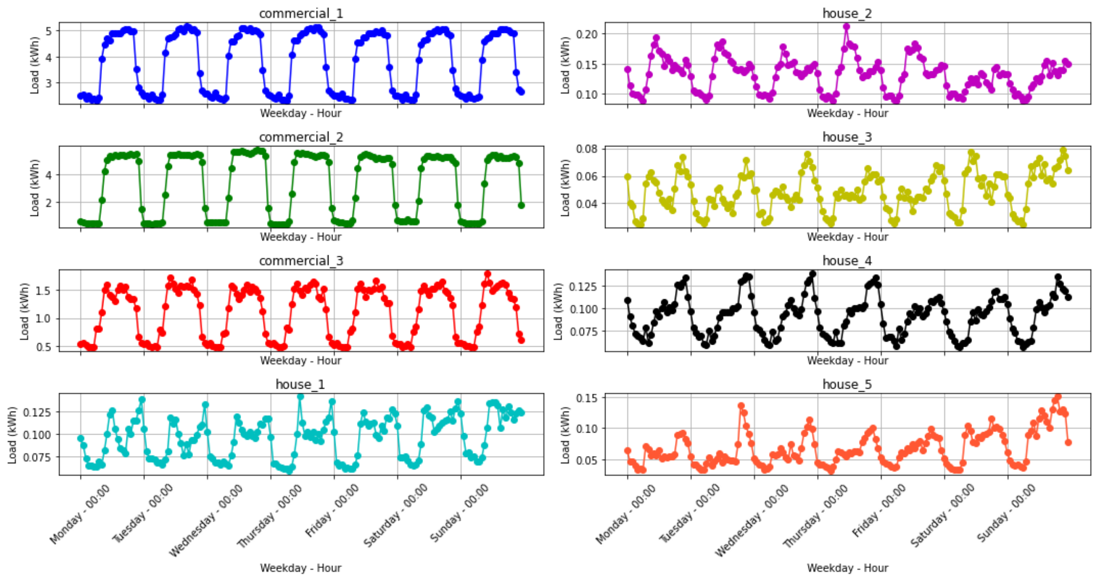

Figure 15: Average Day of Week Load of Peers

Total monthly generation was calculated and plotted in Figure 16. The generation amount increases in summer season and decreases in winter. Because peers are close to each other, patterns are similar.

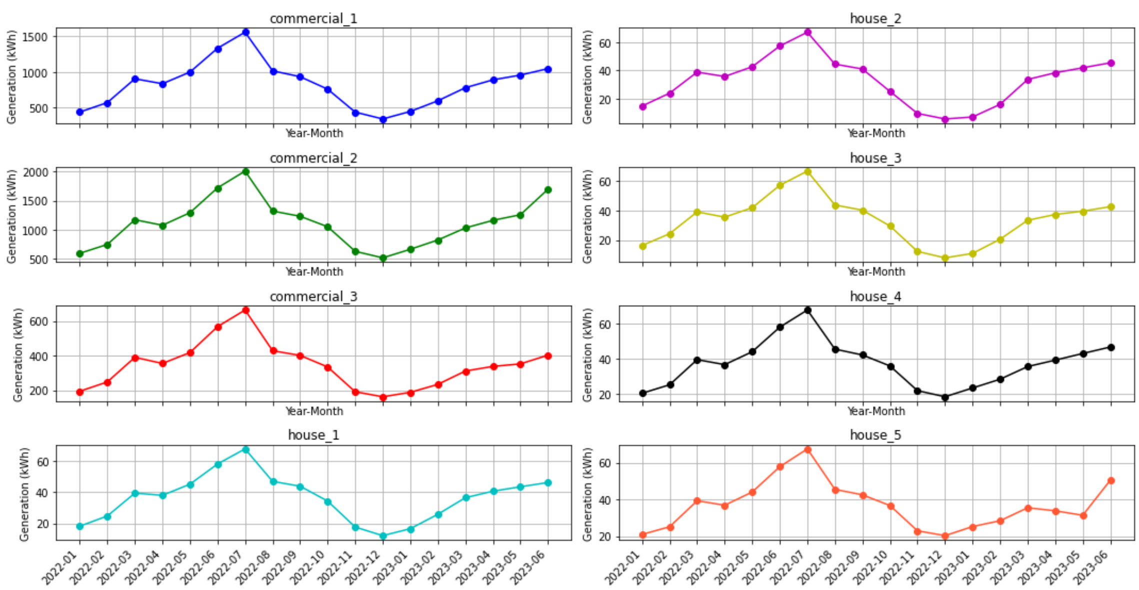

Figure 16: Monthly Total Generation of Peers

Hourly average generation of each peer's solar is shown in Figure 17. Typical solar generation can be seen in this figure for each peer's solar.

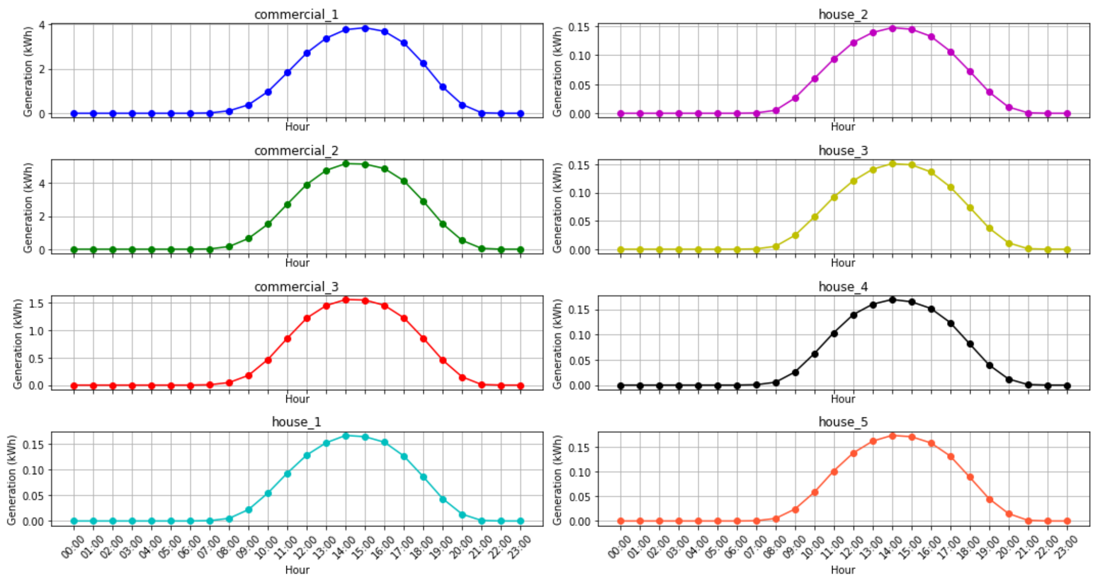

Figure 17: Average Hourly Generation of Peers

4.2. Evaluation Metrics

4.2.1. Forecasting Metrics

a) Load Forecast

Performance metrics are essential to evaluate the accuracy and effectiveness of load forecasts generated using machine learning models. The most used performance error metrics in the order of prevalence are listed below. The lowest value is better when evaluating models using these metrics. 240 academic papers are reviewed to extract this result (Nassif, Soudan, Azzeh, Attilli, & AlMulla, 2021). In the formulas, y_i shows the actual, y_i' shows the forecasted value for sample i, and n is the total number of samples.

- Mean Absolute Percentage Error (MAPE) calculates the percentage difference between forecasted load values and actual load values. It measures relative forecasting error, making it useful for comparing accuracy across different datasets or time periods (Mir, et al., 2021). When dealing with datasets that contain zero or very low actual values, it makes MAPE calculation invalid for those instances.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_i'|}{y_i} \tag{19}$$

- Mean Squared Error (MSE) calculates the average squared difference between forecasted load values and actual load values. It penalizes larger errors more than MAE, making it sensitive to outliers.

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2 \tag{20}$$

- Root Mean Squared Error (RMSE) is square root of MSE and provides a measure of average magnitude of forecast errors. It is widely used and easy to interpret.

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2} \tag{21}$$

- Mean Absolute Error (MAE) measures the average absolute difference between forecasted load values and actual load values. It provides a simple and interpretable metric for forecasting accuracy.

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - y_i'| \tag{22}$$

When evaluating load forecasts, it is essential to consider multiple metrics to gain a comprehensive understanding of the model's performance (Mamun, et al., 2020). Additionally, visualizations, such as time series plots comparing predicted and actual loads, can also aid in understanding the model's strengths and weaknesses in capturing load patterns and trends. However, to automize the structure of P2P energy trading, MAPE is used for selecting the best model for each peer.

b) Solar Generation Forecast

Evaluation metrics play a critical role in assessing the accuracy and efficacy of solar generation forecasts generated by machine learning models (Sobri, Koohi-Kamali, & Rahim, 2018). Similar to load forecasts, various performance error metrics are commonly employed to gauge the quality of solar generation forecasts. The following metrics are widely used to evaluate the performance of solar generation forecast models (Rahimi, et al., 2023):

- Normalized Mean Absolute Error (NMAE) is a variation of Mean Absolute Error that normalizes absolute errors with respect to the magnitude of installed power, allowing for relative comparison of forecast accuracy across different datasets or time periods.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_i'|}{p} \tag{23}$$

  where p is the installed power of the solar plant.

- Mean Squared Error (MSE)

- Root Mean Squared Error (RMSE)

- Mean Absolute Error (MAE)

Like with load forecasts, considering multiple metrics is essential when evaluating solar generation forecasts to gain a comprehensive understanding of the model's performance. Additionally, visualizations, such as time series plots comparing predicted and actual solar generation can aid in understanding the model's strengths and weaknesses in capturing solar generation patterns and trends.

For automating the structure of P2P energy trading, NMAE is used for selecting the best model for each peer. This metric enables comparison of forecast accuracy and facilitates the selection of the most suitable forecasting model for P2P energy trading purposes (Wu, Huang, Phan, & Li, 2022).

c) Electricity Price Forecast

In the field of electric price forecasting, the most commonly used metrics to assess the accuracy of point forecasts are Mean Absolute Error, Root Mean Square Error, and Mean Absolute Percentage Error. These metrics are defined above sections.

While MAE and RMSE are widely used, they may not always provide informative comparisons between different datasets due to their reliance on absolute errors (Lago, Marcjasz, Schutter, & Weron, 2021). Since electricity costs and profits are often linearly dependent on electricity prices, metrics based on quadratic errors such as RMSE can be challenging to interpret and may not accurately represent underlying forecasting problems for most market participants.

In most electricity trade applications, inherent risk, profits, and costs depend linearly on price and forecasting errors. Hence, linear metrics represent the risks of forecasting errors more effectively than quadratic metrics. Similarly, MAPE values become very large when prices are close to zero, regardless of the actual absolute errors, making MAPE less informative, especially during periods of low prices. However, when automating P2P energy trading, MAPE was used for selecting the best model. This is because MAPE, being an absolute percentage error metric, offers a direct measure of the accuracy of price forecasts relative to actual prices, which is particularly relevant in energy trading scenarios where forecasting errors directly impact the profits, costs, and risks associated with electricity price fluctuations (Weron, 2014).

4.2.2. Cost Functions for P2P Energy Trading Game

The cost functions are important in P2P energy trading, and they play a critical role in optimizing and guiding trading decisions between peers. P2P energy trading allows peers with renewable energy resources, such as solar panels or wind turbines, to directly exchange excess energy with one another. The cost function is an essential component that determines how much each peer’s agent strategy works (Zahraoui, et al., 2021). The most popular cost functions used in literature are listed below. These are also applied in comparison of results.

a) Net Electricity Consumption

Net electricity consumption is calculated for a peer or whole market. It represents the portion of a peer's load that is not met by their own solar generation and battery storage. This value indicates the energy requirement that a peer cannot fulfill internally. A lower value for net consumption indicates that a peer is less dependent on external sources and is more self-sufficient, meeting a larger portion of its energy needs through its resources. It is calculated using the below formula for each peer.

$$E_{net}^{n} = \sum_t max(E_{t,L}^{n} - E_{t,PV}^{n} + E_{t,C}^{n}, 0) \quad for\ peer\ n \tag{24}$$

It is calculated using the below formula for the market.

$$E_{net} = \sum_n \sum_t max(E_{t,L}^{n} - E_{t,PV}^{n} + E_{t,C}^{n}, 0) \tag{25}$$

b) Net Electricity Consumption with Negatives

Net electricity consumption with negative consumption values is calculated for a peer or whole market. It is the summation of net electricity consumption including self-generated renewable energy resources. It is calculated using the below formula for each peer.

$$E_{net}^{n} = \sum_t (E_{t,L}^{n} - E_{t,PV}^{n} + E_{t,C}^{n}) \quad for\ peer\ n \tag{26}$$

It is calculated using the below formula for the market.

$$E_{net} = \sum_n \sum_t (E_{t,L}^{n} - E_{t,PV}^{n} + E_{t,C}^{n}) \tag{27}$$

c) Electricity Cost

Electricity cost is calculated for a peer or whole market. This is the monetary cost of electricity and is calculated by the multiplication of price and net electricity consumption. It is calculated using the below formula for each peer.

$$Cost_{net}^{n} = \sum_t max(E_{t,L}^{n} - E_{t,PV}^{n} + E_{t,C}^{n}, 0) * p_t^{B} \quad for\ peer\ n \tag{28}$$

It is calculated using the below formula for the market.

$$Cost_{net} = \sum_n \sum_t max(E^n_{t,L} - E^n_{t,PV} + E^n_{t,C}, 0) * p^B_t \quad (29)$$

d) Peak Demand

Simultaneous increases in demand of all peers in the market can make it challenging to meet demand. Reducing peak demand in the market during P2P energy trading is essential in this regard. This value is calculated by taking a maximum of summation of all net consumptions of peers in the market. It is calculated using the below formula for the market.

$$E_{peak} = max(\sum_n max(E^n_{t,L} - E^n_{t,PV} + E^n_{t,C}, 0)) \; \forall t \quad (30)$$

e) Daily Peak Demand

In addition to measuring peak demand throughout all simulation horizons, it is also essential to evaluate peak demands in a daily manner. To assess this, daily peak demands are calculated. The average of these values is included in the cost function.

f) Load Factor

Stabilizing peak demand at all times is crucial for the reliability and efficiency of the system. To measure the system's stability, load factor is used. This value is calculated by subtracting the ratio of average net consumption to peak demand from 1. It is calculated using the below formula for the market.

$$E_{loadfactor} = 1 - \frac{(\sum_n \sum_t max(E^n_{t,L} - E^n_{t,PV} + E^n_{t,C}, 0))/t}{max(\sum_n max(E^n_{t,L} - E^n_{t,PV} + E^n_{t,C}, 0)) \; \forall t} \quad (31)$$

g) Ramp Cost

The increase or decrease in energy demand within a unit of time is measured by the ramp rate. Achieving sudden increases or decreases reliably is challenging. Therefore, the system's ramp cost is calculated. This value is obtained by summing differences between consecutive net electricity consumption values.

In summary, cost functions in P2P energy trading are crucial for creating a balanced, efficient, and sustainable energy marketplace (Nguyen, Peng, Sokolowski, & Alahakoon, 2018).

4.3. Results and Discussion

The simulation environment created for P2P energy trading game model based on reinforcement learning, which involves multiple agents, was evaluated separately at two main subproblems: forecasting and energy trading. The simulation data for this game belongs to 8 prosumers covering approximately 1.5 years. Recursive forecasts have been generated for the last 3 months of data, and these forecasts are used as input for energy trading game. The results of generation, load, and price forecasts were individually evaluated for each peer, and energy trading results were examined based on different cost functions.

4.3.1. Forecasting Results

a) Load Forecasts

Load forecasting is a critical input required for P2P energy trading. Before conducting very short-term load forecasts, comprehensive data exploration and preprocessing stages were completed, as outlined in previous sections. During the feature engineering phase, the focus was on creating impactful features. Calendar variables including hour, day of week, month, and season were incorporated to capture seasonality, and after applying One-hot encoding, they were transformed into binary variables (Okada, Ohzeki, & Taguchi, 2019). Moreover, holiday information was added to the input set as a binary variable to catch trend changes in load. To enhance the model's ability to capture load behavior within a very short-term forecasting horizon, lagged values were introduced. These are the most important feature sets in load forecasting problems. The same hour of last week's each day and all hours of the previous 24 hours were taken as lagged features. Additionally, beyond considering weather parameters as significant factors affecting load, new features derived from them were integrated into the input set. These features, such as relative humidity-to-temperature ratio, and apparent temperature-to-temperature ratio, were deemed valuable for improving the accuracy of load forecasting in P2P energy trading. Lastly, heating and cooling degree days were calculated. Heating Degree Days (HDD) and Cooling Degree Days (CDD) are measures used to estimate heating and cooling energy requirements for consumers based on outdoor temperatures. HDD indicates how much heating is needed when it's colder than the comfort temperature (nearly 18.15 °C in Turkey), while CDD shows cooling requirements when it's hotter than the comfort temperature (nearly 22.15 °C in Turkey). The squares of these metrics were also added to the input dataset. These features are summarized in the below table.

After dividing approximately 1.5 years of load data into two parts for each peer separately as training and test set, the test data was split into weekly intervals for generating very short-term load forecasts. For each week in the test set, a forecast was generated, and forecasted values were then incorporated into a training set for subsequent rounds of forecasting. This iterative training approach allowed the model to progressively learn from recent data, leading to improved forecast accuracy over time.

For load forecasting of 8 peers, LightGBM model was chosen as machine learning model approach due to its outstanding performance in various aspects. Its hyperparameters, including the number of estimators, feature fraction, learning rate, maximum depth, and number of leaves, were meticulously tuned to achieve optimal performance. The fine-tuning process ensured that LightGBM model was optimized to produce highly accurate load forecasts for each peer, accounting for their individual power consumption needs and diverse load profiles arising from different installed power capacities.

Table 9: Feature Space of Electricity Load Forecasting

Variable Name	**Variable Type**
Temperature	Meteorological
Apparent Temperature	Meteorological
Relative Humidity	Meteorological
Global Radiation	Meteorological
Cloud Cover	Meteorological
Wind Speed	Meteorological
Precipitation Probability	Meteorological
Is Holiday	Calendar
Month	Calendar
Day of Week	Calendar
Hour	Calendar
Season	Calendar
Relative Humidity over Temperature	Meteorological
Temperature cross Relative Humidity	Meteorological
Apparent Temperature over Temperature	Meteorological
Temperature cross Apparent Temperature	Meteorological
Heating Degree Days	Meteorological
Cooling Degree Days	Meteorological
HDD Square	Meteorological
CDD Square	Meteorological
Lag of previous week's whole days of same hour	Lagged
Lag of previous day's whole hours	Lagged

In addition to LightGBM model, benchmarking models were employed for comparison, namely ANN with 2 hidden layers, Random Forest, and Ridge Regression. These models were chosen as benchmarking models due to their widespread use and applicability in forecasting tasks. By comparing the performance of these benchmarking models against LightGBM model, a comprehensive assessment of the forecasting approaches was conducted to determine the most effective and accurate method for load forecasting in the context of P2P energy trading.

The below graph represents actual load values and load forecasts of all models for March 2023.

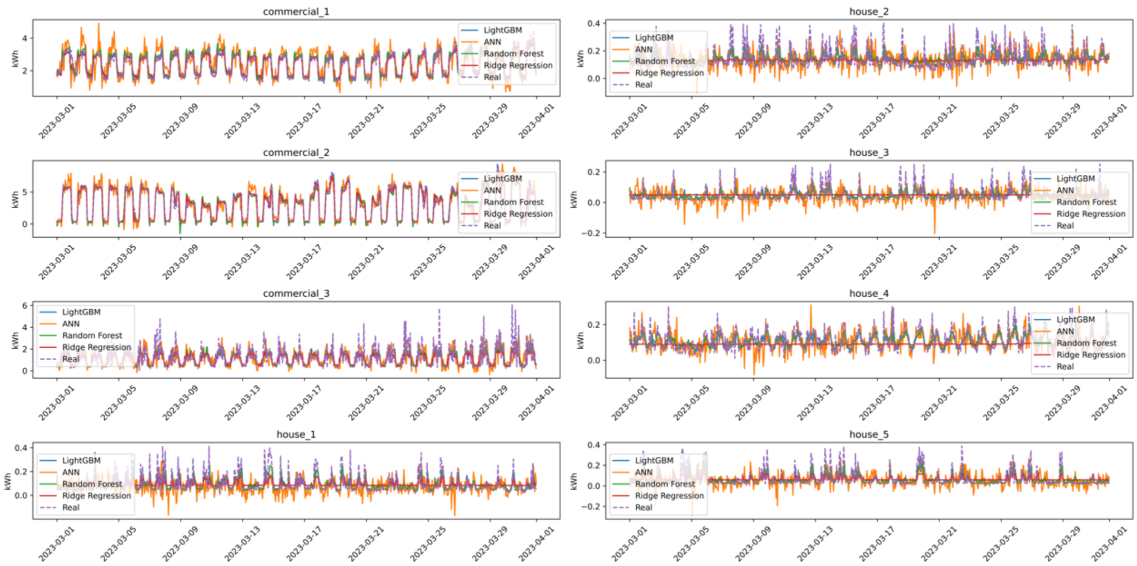

Figure 18: Electricity Load Forecasts vs. Actual Values for March 2023

April 2023 values are shown in the below graph.

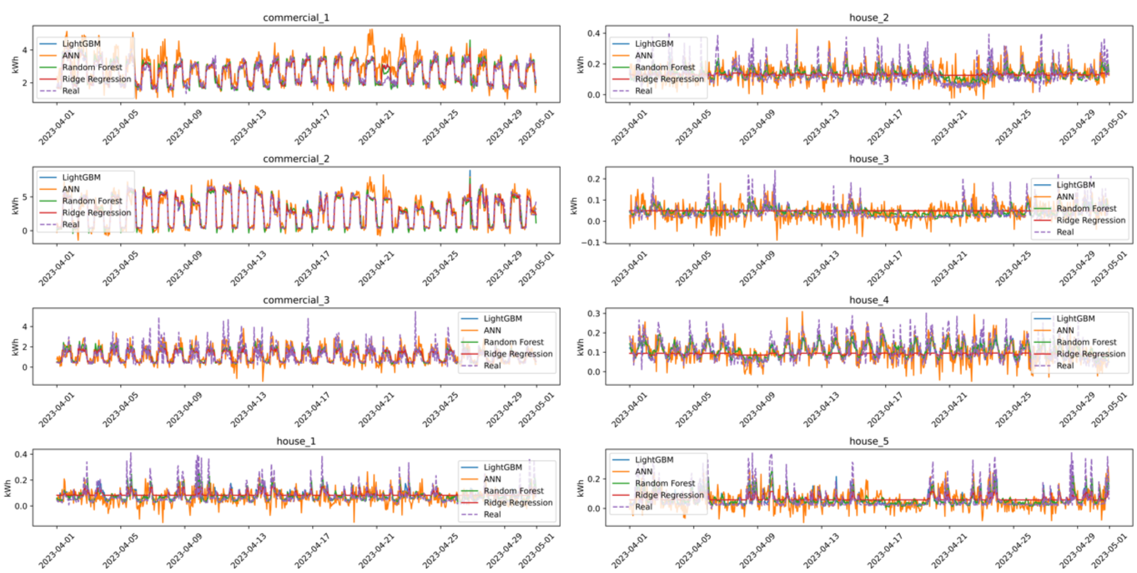

Figure 19: Electricity Load Forecasts vs. Actual Values for April 2023

The below graph shows forecasted load values for the period of May 2023 including all benchmarking models.

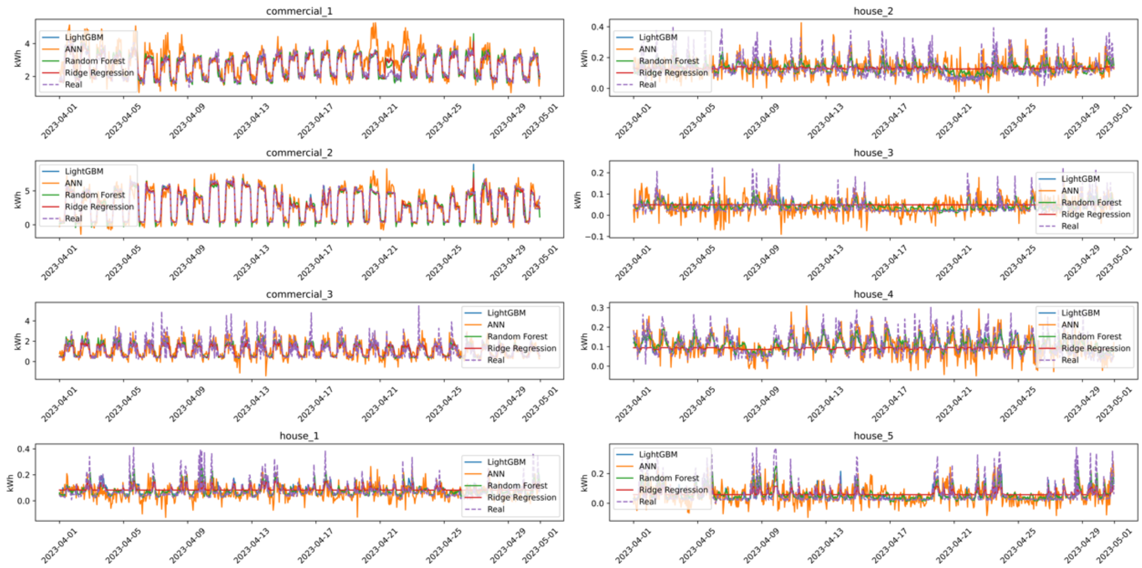

Figure 20: Electricity Load Forecasts vs. Actual Values for May 2023

In the following tables, a comparison of LightGBM model with benchmarked ANN, Random Forest, and Ridge Regression models are shown in terms of MAE and MSE. LightGBM model demonstrates better performance not only in terms of computational speed but also in accuracy compared to other models for commercial peers. MAPE values were not selected for load forecasting. The real load of the selected dataset is very small. Since MAPE involves dividing absolute error by actual value, it makes metrics sensitive to the magnitude of actual values. Therefore, MAPE provides a percentage error that is relative to the scale of realized data and is often used to understand the forecasting model's accuracy in terms of relative errors.

Table 10: Electricity Load Forecast MAE Values

Month	Peer	LightGBM	ANN	Random Forest	Ridge Regression
3	commercial_1	**0.15**	0.36	0.18	0.17
3	commercial_2	**0.30**	0.62	0.40	0.33
3	commercial_3	**0.43**	0.57	**0.43**	**0.43**
3	house_1	**0.04**	0.06	**0.04**	0.05
3	house_2	**0.04**	0.06	**0.04**	**0.04**
3	house_3	**0.02**	0.04	**0.02**	0.03
3	house_4	**0.03**	0.04	**0.03**	0.04
3	house_5	**0.03**	0.06	0.04	0.05
4	commercial_1	**0.19**	0.42	0.20	0.20
4	commercial_2	**0.32**	0.66	0.42	0.34
4	commercial_3	**0.38**	0.56	0.40	**0.38**
4	house_1	**0.03**	0.05	**0.03**	0.04
4	house_2	**0.04**	0.06	**0.04**	**0.04**
4	house_3	**0.02**	0.03	**0.02**	0.03
4	house_4	**0.03**	0.04	**0.03**	0.04
4	house_5	**0.03**	0.05	0.04	0.04
5	commercial_1	0.26	0.44	**0.23**	0.25
5	commercial_2	**0.28**	0.52	0.37	0.30
5	commercial_3	**0.16**	0.24	**0.16**	**0.16**
5	house_1	**0.03**	0.04	**0.03**	0.04
5	house_2	**0.04**	0.05	**0.04**	**0.04**
5	house_3	**0.02**	0.03	**0.02**	0.03
5	house_4	**0.03**	0.04	**0.03**	**0.03**
5	house_5	**0.02**	0.04	0.03	0.04
Total	commercial_1	**0.20**	0.41	**0.20**	0.21
Total	commercial_2	**0.30**	0.60	0.40	0.32
Total	commercial_3	**0.32**	0.45	0.33	**0.32**
Total	house_1	**0.03**	0.05	**0.03**	0.04
Total	house_2	**0.04**	0.06	**0.04**	**0.04**
Total	house_3	**0.02**	0.04	**0.02**	0.03
Total	house_4	**0.03**	0.04	**0.03**	0.04
Total	house_5	**0.03**	0.05	**0.03**	0.04

Table 11: Electricity Load Forecast MSE Values

Month	Peer	LightGBM	ANN	Random Forest	Ridge Regression
3	commercial_1	**0.043**	0.220	0.061	0.049
3	commercial_2	**0.338**	0.799	0.424	0.361
3	commercial_3	0.493	0.680	**0.473**	0.488
3	house_1	**0.004**	0.007	**0.004**	**0.004**
3	house_2	**0.004**	0.007	**0.004**	0.005
3	house_3	**0.001**	0.003	**0.001**	0.002
3	house_4	**0.002**	0.003	**0.002**	0.003
3	house_5	**0.003**	0.006	**0.003**	0.004
4	commercial_1	**0.068**	0.315	0.075	0.073
4	commercial_2	0.387	0.863	0.460	**0.362**
4	commercial_3	0.398	0.645	0.407	**0.394**
4	house_1	**0.003**	0.005	**0.003**	**0.003**
4	house_2	**0.003**	0.006	**0.003**	0.004
4	house_3	**0.001**	0.002	**0.001**	**0.001**
4	house_4	**0.002**	0.003	**0.002**	**0.003**
4	house_5	**0.003**	0.006	0.004	0.004
5	commercial_1	**0.121**	0.348	0.097	0.115
5	commercial_2	**0.247**	0.598	0.299	0.260
5	commercial_3	0.054	0.102	**0.051**	0.052
5	house_1	**0.002**	0.003	**0.002**	0.003
5	house_2	**0.003**	0.005	0.004	0.004
5	house_3	**0.001**	0.002	**0.001**	0.002
5	house_4	**0.001**	0.003	**0.001**	0.002
5	house_5	**0.002**	0.003	**0.002**	0.003
Total	commercial_1	**0.077**	0.294	0.078	0.079
Total	commercial_2	**0.323**	0.752	0.394	0.327
Total	commercial_3	0.314	0.474	**0.309**	0.311
Total	house_1	**0.003**	0.005	**0.003**	**0.003**
Total	house_2	**0.004**	0.006	**0.004**	**0.004**
Total	house_3	**0.001**	0.003	**0.001**	0.002
Total	house_4	**0.002**	0.003	**0.002**	0.003
Total	house_5	**0.003**	0.005	**0.003**	0.004

All in all, every peer has access to special their load forecasts to make logical decisions in trading. A precise load forecast is essential for effective energy trading decisions, as it allows peers to optimize their energy load and generation strategies based on forecasted load patterns.

b) *Solar Generation Forecasts*

Another input required for P2P energy trading is solar generation forecasting. Before conducting very short-term solar generation forecasts, data exploration and data preprocessing stages as described in previous sections were completed. During the feature engineering stage for solar generation forecasting, creating effective features is important. Calendar variables such as hour, month, and season were added to capture seasonality, and after applying One-hot encoding, they were used as binary variables. Given a very short-term forecasting horizon, the model's ability to capture generation behavior within a day was improved by adding lagged values. Additionally, apart from weather parameters, which are significant factors affecting generation, new features derived from them were also added to the input set. These features, as summarized in the feature space table below, include the radiation-to-temperature ratio, square of radiation, radiation multiplied by temperature, and others.

Table 12: Feature Space of Electricity Generation Forecasting

Variable Name	**Variable Type**
Temperature	Meteorological
Apparent Temperature	Meteorological
Relative Humidity	Meteorological
Global Radiation	Meteorological
Cloud Cover	Meteorological
Direct Radiation	Meteorological
Precipitation Probability	Meteorological
Diffuse Radiation	Meteorological
Sun Elevation	Meteorological
Month	Calendar
Hour	Calendar
Season	Calendar
Radiation over Temperature	Meteorological
Radiation Sqaure	Meteorological
Temperature cross Radiation	Meteorological
Apparent Temperature over Temperature	Meteorological
Lag of previous day's whole hours	Lagged

After dividing solar generation data into two parts for each peer separately, approximately 1.5 years were used for training, while the remaining data was split into weekly intervals to serve as a test set for generating very short-term solar generation forecasts. For each week in the test set, a forecast was made, and the predicted values were included in the training set for the next round of forecasting. This iterative training approach allowed the model to progressively learn from recent data and improve the accuracy of the forecasts over time.

To forecast solar generation of 8 peers, LightGBM model was selected as the preferred algorithm. The model's hyperparameters, such as the number of estimators, feature fraction, learning rate, maximum depth, and number of leaves, were tuned to achieve optimal performance. By fine-tuning these parameters, LightGBM model was optimized to produce precise forecasts for solar generation data of 8 peers including different solar plants that have different installed power.

The graph below presents forecasted solar generation values for the period of March 2023, alongside actual solar generation values for each peer. Additionally, the graph compares the results of LightGBM model with benchmarking models, which include ANN with 2 hidden layers, Random Forest, and Ridge Regression.

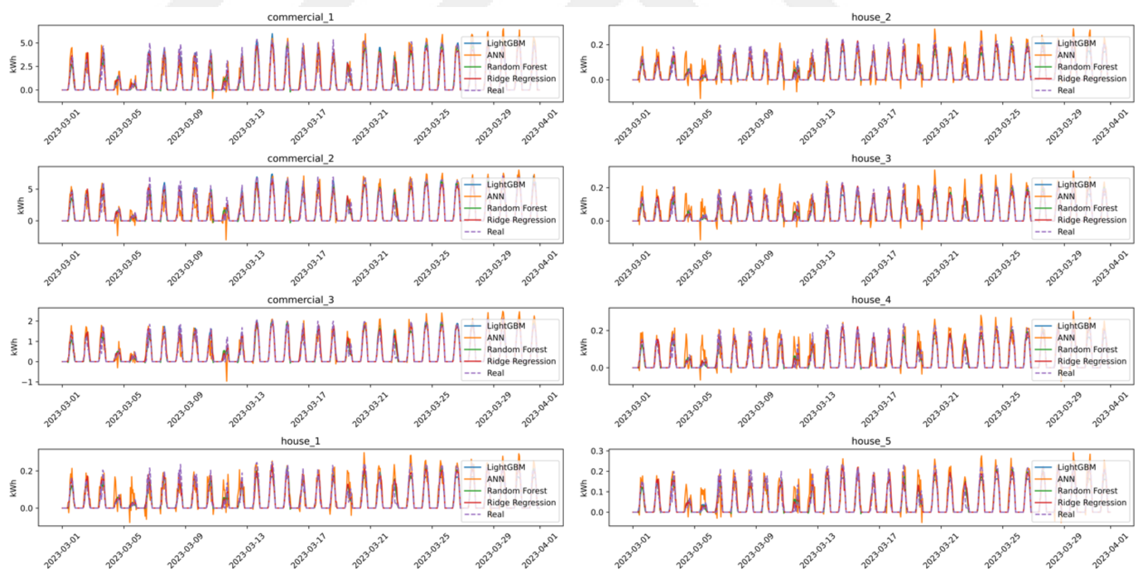

Figure 21: Electricity Generation Forecasts vs. Actual Values for March 2023

The below graph shows forecasted solar generation values for the period of April 2023 including all benchmarking models.

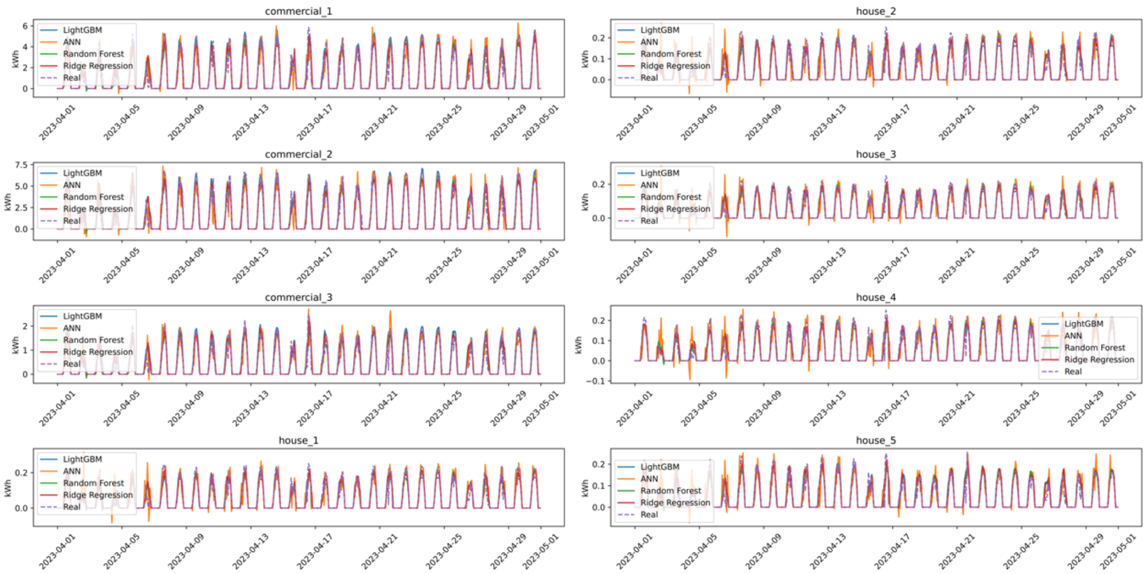

Figure 22: Electricity Generation Forecasts vs. Actual Values for April 2023

The graph depicted below displays forecasted solar generation values for the month of May 2023, encompassing all benchmarking models.

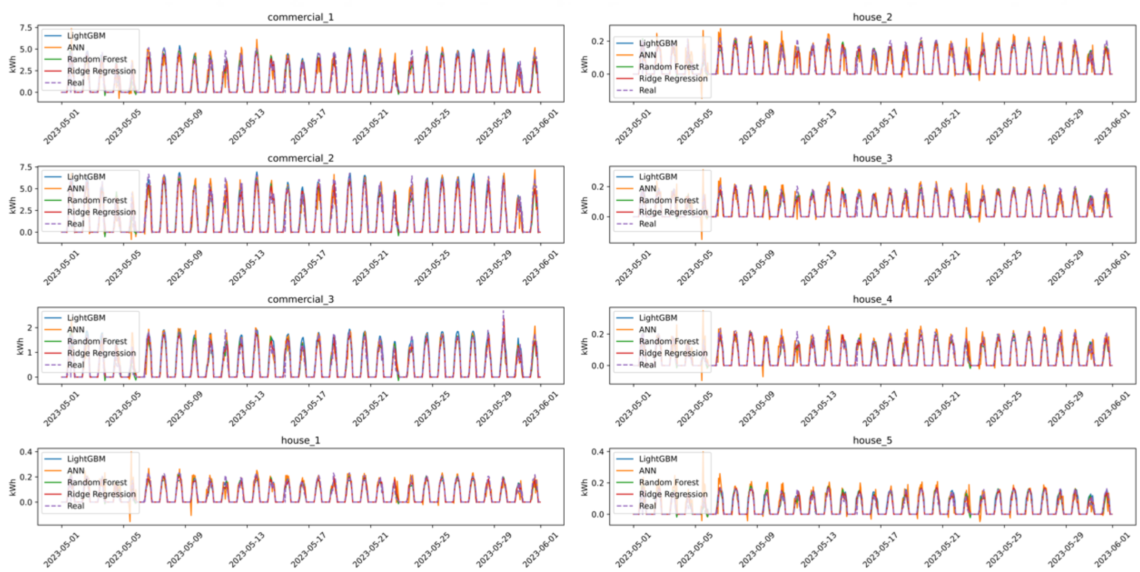

Figure 23: Electricity Generation Forecasts vs. Actual Values for May 2023

The tables below present a comparison of LightGBM model with benchmarked ANN with 2 hidden layers, Random Forest, and Ridge Regression models. LightGBM model demonstrates better performance not only in terms of computational speed but also in accuracy compared to other models for commercial peers. Because the values are so close to house peers, LightGBM model was selected for them also.

Table 13: Electricity Generation Forecast MAE Values

Month	Peer	LightGBM	ANN	Random Forest	Ridge Regression
3	commercial_1	**0.195**	0.324	0.246	0.228
3	commercial_2	**0.259**	0.429	0.314	0.299
3	commercial_3	**0.076**	0.120	0.089	0.084
3	house_1	0.013	0.019	0.012	**0.011**
3	house_2	0.012	0.019	0.011	**0.010**
3	house_3	0.012	0.020	0.011	**0.010**
3	house_4	0.011	0.018	**0.010**	**0.010**
3	house_5	0.012	0.018	**0.010**	**0.010**
4	commercial_1	**0.248**	0.356	0.313	0.308
4	commercial_2	**0.290**	0.428	0.381	0.367
4	commercial_3	0.109	0.140	0.113	**0.104**
4	house_1	0.014	0.021	0.014	**0.013**
4	house_2	0.014	0.019	**0.013**	**0.013**
4	house_3	0.013	0.019	0.013	**0.012**
4	house_4	0.014	0.020	**0.013**	**0.013**
4	house_5	0.013	0.019	**0.012**	**0.012**
5	commercial_1	**0.231**	0.357	0.290	0.277
5	commercial_2	**0.294**	0.446	0.370	0.353
5	commercial_3	0.118	0.116	0.101	**0.092**
5	house_1	0.013	0.019	0.013	**0.012**
5	house_2	0.013	0.019	0.012	**0.012**
5	house_3	0.012	0.018	**0.011**	**0.011**
5	house_4	0.013	0.018	0.013	**0.012**
5	house_5	0.012	0.017	**0.009**	**0.009**
Total	commercial_1	**0.224**	0.345	0.282	0.271
Total	commercial_2	**0.281**	0.434	0.355	0.339
Total	commercial_3	0.101	0.125	0.101	**0.093**
Total	house_1	0.013	0.020	0.013	**0.012**
Total	house_2	0.013	0.019	**0.012**	**0.012**
Total	house_3	0.012	0.019	0.012	**0.011**
Total	house_4	0.013	0.019	**0.012**	**0.012**
Total	house_5	0.012	0.018	0.011	**0.010**

Table 14: Electricity Generation Forecast RMSE Values

Month	Peer	LightGBM	ANN	Random Forest	Ridge Regression
3	commercial_1	**0.422**	0.652	0.489	0.456
3	commercial_2	**0.560**	0.911	0.633	0.605
3	commercial_3	**0.168**	0.261	0.183	0.176
3	house_1	0.024	0.036	0.023	**0.022**
3	house_2	0.023	0.035	0.022	**0.020**
3	house_3	0.022	0.037	0.021	**0.020**
3	house_4	0.022	0.035	0.021	**0.020**
3	house_5	0.022	0.034	0.021	**0.020**
4	commercial_1	**0.514**	0.690	0.575	0.572
4	commercial_2	**0.590**	0.849	0.674	0.658
4	commercial_3	0.224	0.272	0.219	**0.212**
4	house_1	**0.026**	0.039	**0.026**	**0.026**
4	house_2	**0.025**	0.036	**0.025**	**0.025**
4	house_3	0.024	0.036	0.024	**0.023**
4	house_4	0.025	0.036	0.025	**0.024**
4	house_5	**0.024**	0.036	**0.024**	**0.024**
5	commercial_1	**0.465**	0.663	0.532	0.502
5	commercial_2	**0.558**	0.814	0.649	0.610
5	commercial_3	0.213	0.218	0.196	**0.186**
5	house_1	0.023	0.037	0.024	**0.022**
5	house_2	0.023	0.036	0.023	**0.022**
5	house_3	0.021	0.034	0.021	**0.020**
5	house_4	0.023	0.034	0.023	**0.022**
5	house_5	0.022	0.033	0.018	**0.017**
Total	commercial_1	**0.468**	0.668	0.533	0.511
Total	commercial_2	**0.569**	0.859	0.652	0.624
Total	commercial_3	0.203	0.251	0.200	**0.191**
Total	house_1	0.024	0.037	0.024	**0.023**
Total	house_2	0.024	0.036	0.023	**0.022**
Total	house_3	0.022	0.035	0.022	**0.021**
Total	house_4	0.023	0.035	0.023	**0.022**
Total	house_5	0.023	0.034	**0.021**	**0.021**

Each peer's solar generation data was forecasted separately, considering their unique installed power capacities. Solar generation forecasts were tailored to each peer, taking into account their specific solar energy generation capabilities. As a result, every peer has access to personalized their production forecasts. This approach ensures that each peer can make informed decisions based on forecasted electricity solar generation.

c) Electricity Price Forecasts

Electricity price forecasting follows a similar trend as shown in Figure 6, for load and solar generation forecasts. Electricity price data was obtained from day-ahead market prices of Turkey's energy exchange, known as "Enerji Piyasaları İşletme A.Ş.(EPİAŞ)". The data was sourced from their transparency platform[3]. Its unit is TL/MWh. Since generation and load data in this study are scaled at a kilowatt level, the price values are divided by 1000 to obtain prices in TL/kWh.

Since electricity price data used for forecasting is verified and cleaned, it requires no further preprocessing. Additionally, feature data used for electricity price prediction includes periodically varying maximum limit value of price. The graph of price and maximum limit is shown below.

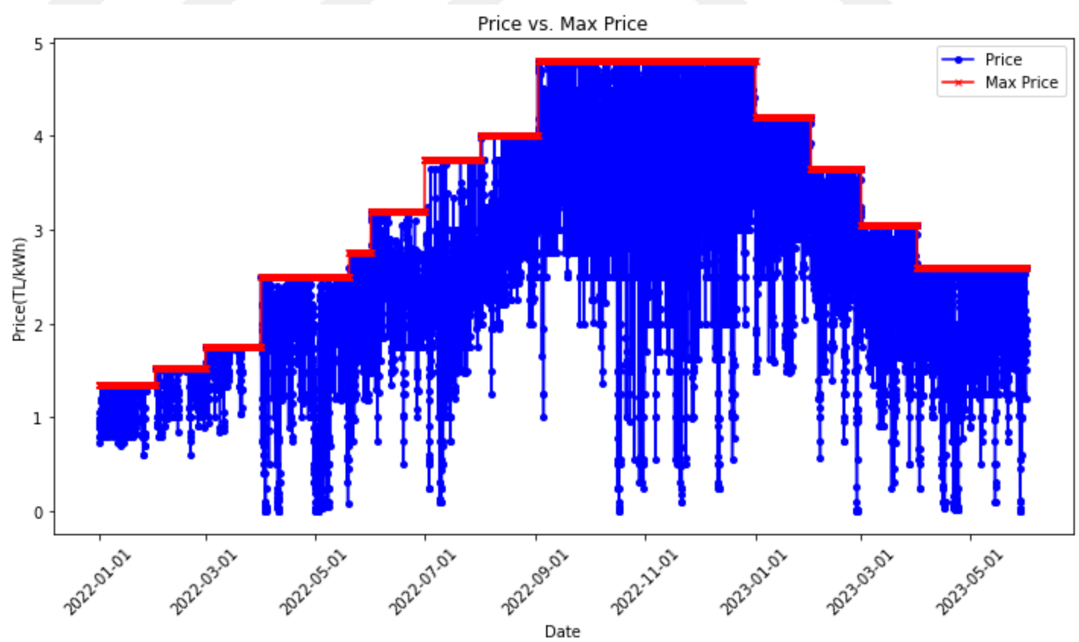

Figure 24: Electricity Prices and Maximum Price Limit

One of most significant factors influencing electricity prices is generated energy, a critical portion of which is dependent on natural gas, affected by USD exchange rate. Therefore, USD exchange rates were included in the feature set for electricity price forecasting.

[3] https://seffaflik.epias.com.tr/transparency/piyasalar/gop/ptf.xhtml

However, it is important to note that when adding this data to the dataset, the most recently disclosed USD exchange rate was considered as the basis for inclusion. Weather variables were added to the model's input set because they are significant factors affecting electricity prices, primarily through their impact on energy generation and consumption. Features like temperature, relative humidity, and radiation were put into feature engineering to generate new relevant features. Additionally, calendar variables were added to the model to capture daily, weekly, and monthly seasonal patterns present in price data. Categorical variables were converted to binary variables using one-hot encoding technique during the feature engineering process. Considering high autocorrelation in price data, lagged values were incorporated into the model, taking into account daily and weekly seasonality. This was done to better account for time dependencies present in the price data. The resulting features after the feature engineering process are listed in the table below:

Table 15: Feature Space of Electricity Price Forecasting

Variable Name	**Variable Type**
Maximum Price Limit	Price
USD	Price
Temperature	Meteorological
Apparent Temperature	Meteorological
Relative Humidity	Meteorological
Global Radiation	Meteorological
Is Holiday	Calendar
Month	Calendar
Day of Week	Calendar
Hour	Calendar
Season	Calendar
Relative Humidity over Temperature	Meteorological
Temperature cross Relative Humidity	Meteorological
Apparent Temperature over Temperature	Meteorological
Temperature cross Apparent Temperature	Meteorological
Heating Degree Days	Meteorological
Cooling Degree Days	Meteorological
HDD Square	Meteorological
CDD Square	Meteorological
Lag of previous week's whole days of same hour	Lagged
Lag of previous day's whole hours	Lagged

After splitting approximately 1.5 years of electricity price data, the first 12 months were used as the training set, and the remaining data was divided into weekly intervals to create the test set for creating short-term forecasts. Each week from the test set was used for forecasting, and the forecasted values for that week were then included in the training set. This approach allowed for a recursive training process, progressively expanding the training set closer to the current time to advance the forecasting process.

LightGBM model was chosen for price forecasting, and hyperparameter tuning was performed to optimize its performance. The hyperparameters were tuned for number of estimators, feature fraction, learning rate, maximum depth, and number of leaves parameters. The graph below depicts forecasted price values for the period of March 2023 to May 2023 alongside the actual price values. The graph shows benchmarking models with LightGBM model including ANN with 2 hidden layers, Random Forest, and Ridge Regression.

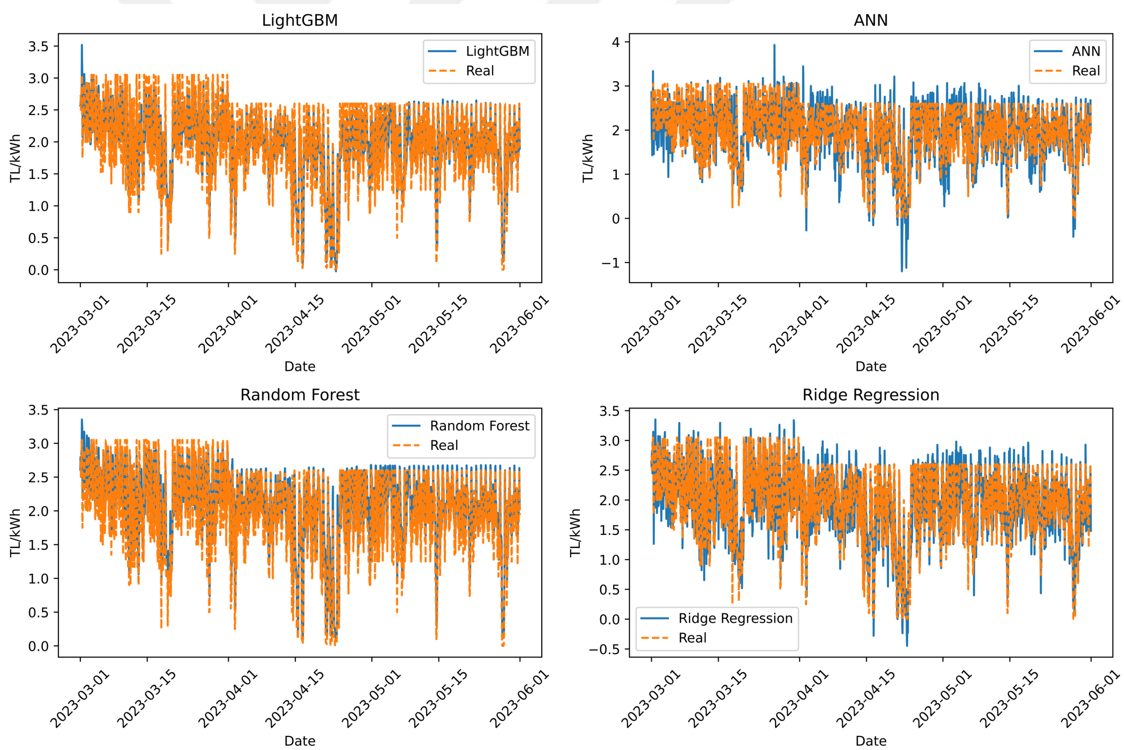

Figure 25: Electricity Price Forecasts vs. Actual Values

The results of benchmarking the developed LightGBM model against ANN with 2 hidden layers, Random Forest, and Ridge Regression models are shown in the table below. LightGBM model stands out not only in terms of computational speed but also in accuracy compared to the other models.

The results of this model are assumed to be shared across all peers, and price forecasts are considered to be publicly available in the market. This assumption means that all peers use the same model and have access to the same price forecasts when making their energy trading decisions. It allows for a fair and consistent comparison of the model's performance across all peers in the P2P energy trading game. With publicly available price predictions, each peer can make informed decisions based on the forecasted electricity prices for their production and consumption strategies.

Table 16: Electricity Price Forecast Results

		Month			
		3	4	5	Total
MAE (kWh/TL)	LightGBM	**0.26**	0.25	**0.20**	**0.24**
	ANN	0.32	0.29	0.27	0.29
	Random Forest	0.29	**0.23**	0.21	0.24
	Ridge Regression	0.30	0.27	0.26	0.28
RMSE (kWh/TL)	LightGBM	**0.12**	0.11	**0.07**	**0.10**
	ANN	0.17	0.15	0.13	0.15
	Random Forest	0.14	**0.10**	0.08	0.11
	Ridge Regression	0.15	0.12	0.11	0.13

4.3.2. P2P Energy Trading Game Results

The results of multi-agent reinforcement learning model established to simulate the trading game of P2P energy trading network, consisting of 8 peers, were evaluated in this section. After modeling solar generation, consumption, and price data in a continuous learning framework, the created forecasts are utilized as inputs for the energy trading game.
The model established in the thesis study operates in a model-free approach, meaning it works without any prior knowledge. Therefore, the model was trained for a considerable number of episodes to compare the model's results. An episode refers to a run where interactions occur between an agent and its environment. With each increasing episode, the agent learns from interactions and can develop better strategies. Replay buffer stores past states and rewards, allowing for improved exploration. However, increasing number of episodes also leads to longer computation times, and beyond certain limits, it may cause delays in real-world trading in the market.

The cost function values of the proposed multi-agent reinforcement learning model's trading, using very short-term load, solar generation, and price forecasts between March 2023 and May 2023 in a system consisting of 8 peers, are presented in the following tables. The net electricity consumption of each peer, including the values during times when net electricity consumption exceeds solar generation, and total electricity cost are evaluated here. The results of multi-agent RL model were analyzed based on the actions of agents who explored the environment during the first two months of 3-month data period. These actions were analyzed during May 2023.

Multi-agent RL (MARL) model results were taken for 10, 20, 50 and 100 episodes. Table 17 presents calculated electricity cost, net electricity consumption, and net electricity consumption with negatives for models trained with different numbers of episodes. When examining electricity cost values, it can be observed that peer-based behaviors vary as number of episodes increases. Similarly, there are changes in net electricity consumption values corresponding to these variations. Therefore, an increase in number of episodes allowed each agent to undergo more training, leading to the development of more competitive behaviors.

Table 17: Peer-based Cost Function Values of Proposed Approach for Different Number of Episodes

Cost Function	Peer	Episode 10	Episode 20	Episode 50	Episode 100
Electricity Cost	commercial_1	1.010	1.004	1.001	1.008
	commercial_2	1.031	0.989	0.991	1.000
	commercial_3	1.068	1.043	1.026	1.040
	house_1	1.036	1.042	1.079	1.059
	house_2	1.025	1.030	1.026	1.024
	house_3	1.088	1.066	1.106	1.116
	house_4	1.038	1.025	1.036	1.040
	house_5	1.074	1.106	1.099	1.099
Net Electricity Consumption	commercial_1	1.016	1.010	1.016	1.009
	commercial_2	1.043	0.998	0.997	1.012
	commercial_3	1.076	1.052	1.036	1.046
	house_1	1.042	1.052	1.089	1.065
	house_2	1.026	1.030	1.026	1.025
	house_3	1.101	1.083	1.108	1.119
	house_4	1.042	1.031	1.039	1.038
	house_5	1.072	1.107	1.099	1.102
Net Electricity Consumption with Negatives	commercial_1	1.010	1.009	1.014	1.008
	commercial_2	1.018	1.013	1.014	1.023
	commercial_3	1.025	1.021	1.016	1.017
	house_1	1.040	1.049	1.060	1.058
	house_2	1.014	1.017	1.016	1.016
	house_3	0.809	0.805	0.821	0.791
	house_4	1.018	1.020	1.025	1.023
	house_5	1.077	1.101	1.110	1.105

The impact of an increasing number of episodes is similar in individual peers and the entire system. As net electricity consumption increases, the market's electricity cost also rises. Analyzing load factor values, it is observed that in the case of 50 episodes, average consumption is less than peak demand, which indicates a more reliable state for the entire system. This effect is also reflected in peak demand and daily peak demand values. Consequently, a decreasing ramp cost value indicates that the system's sensitivity to rapid changes in electricity consumption is reduced. However, exceeding the certain threshold for a number of episodes has a negative effect on the system. Another cost associated with the number of episodes is time. In a real-time trading game, it was observed that working with 10 episodes is the most optimal in terms of time efficiency.

Table 18: Cost Function Values of Market for Different Number of Episodes in Proposed Approach

	Episode			
Cost Function	10	20	50	100
Electricity Cost	1.046	1.038	1.046	1.048
Ramp Cost	1.174	1.105	1.090	1.129
Net Electricity Consumption	1.052	1.045	1.051	1.052
Net Consumption with Negatives	1.002	1.005	1.010	1.005
Peak Demand	14.271	13.809	11.022	16.492
Daily Peak Demand	1.043	1.002	0.978	1.033
Load Factor	1.029	1.008	0.996	1.045

In addition to values in the above tables, a comparison of net electricity consumption for each peer and the market at hourly resolution with respect to episodes is presented in the graph below. According to this graph, when solar generation is limited and used solely for internal consumption on days with low generation (cloudy days), increasing number of episodes does not significantly impact net electricity consumption. This is because all agents most probably utilize solar generation energy for their internal needs, resulting in reduced trading activities. Conversely, on days with high solar generation, an increasing number of episodes leads to increased learning among agents, creating a more competitive market. As a result, the curves compared on an episode basis start to diverge.

Figure 27 depicts solar generation curve in net consumption, including battery consumption (amount of energy used to charge the battery) and total consumption. In this graph, the effect of solar generation on total consumption can be observed during the periods when battery consumption is zero. If this system were established in a geographically dispersed structure, where solar power plants' behaviors are not similar, the likelihood of identifying instances with zero battery consumption in the market would be significantly lower. House 5's consumption from May 25th to May 28th, where it almost reaches zero due to its consumption being mainly base load (refrigerator, freezer, etc.), is seen to have an impact on net consumption curve. Almost all of the solar generation was utilized for trading. Having such an automated P2P energy trading system would increase prosumers' chances of continuous gains.

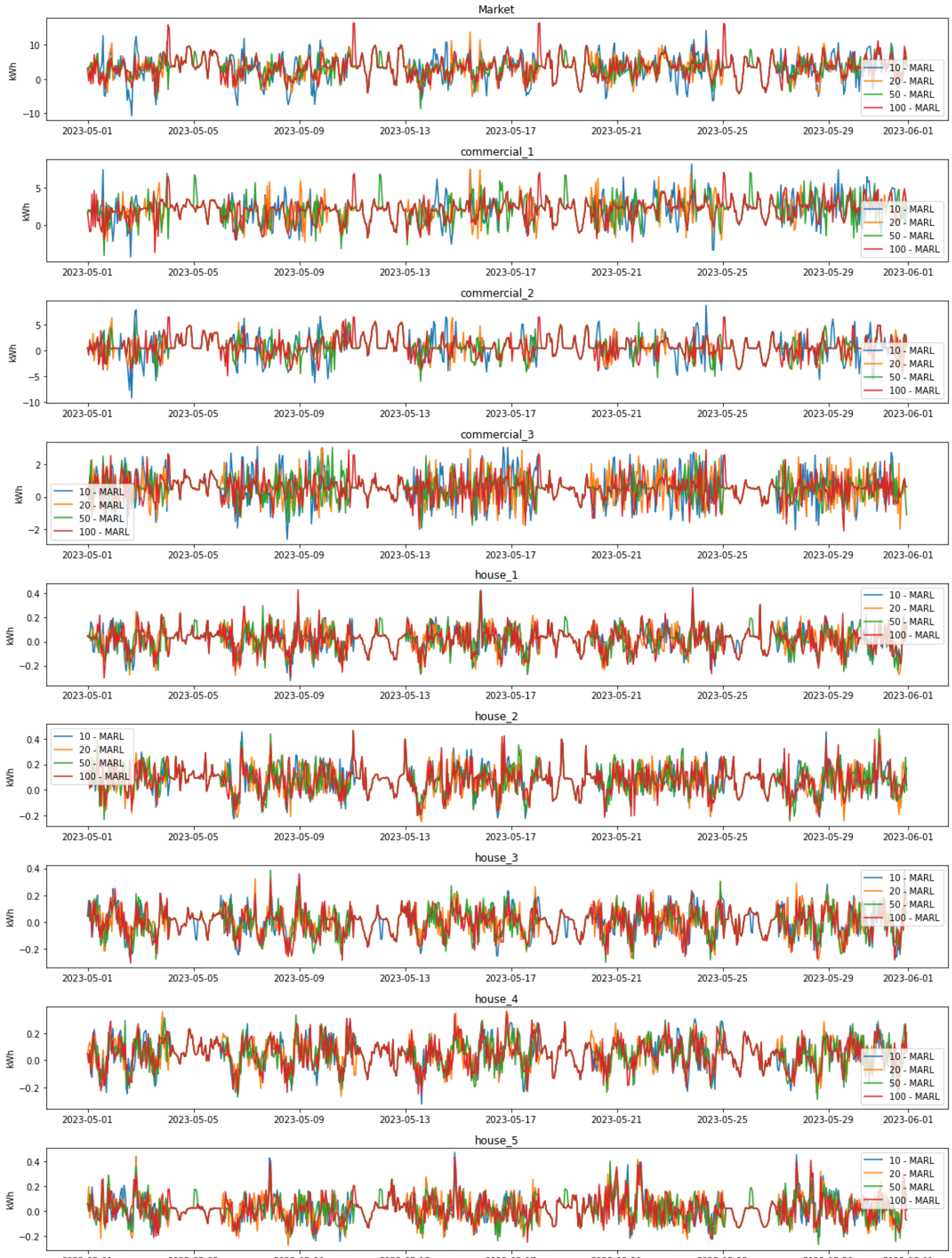

Figure 26: Net Electricity Consumption for Different Number of Episodes

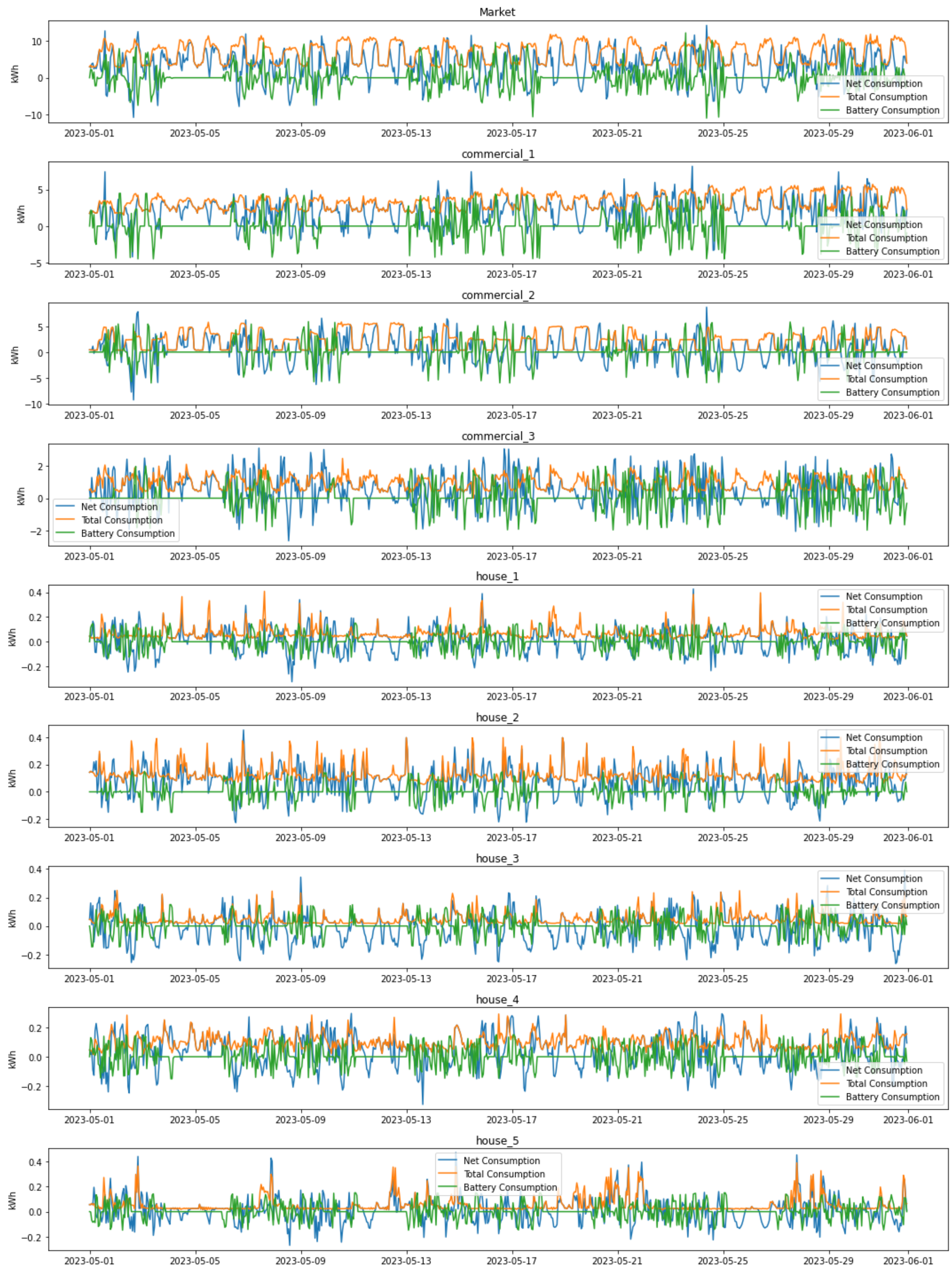

Figure 27: Net, Total, and Battery Consumption for MARL with 10 Episodes

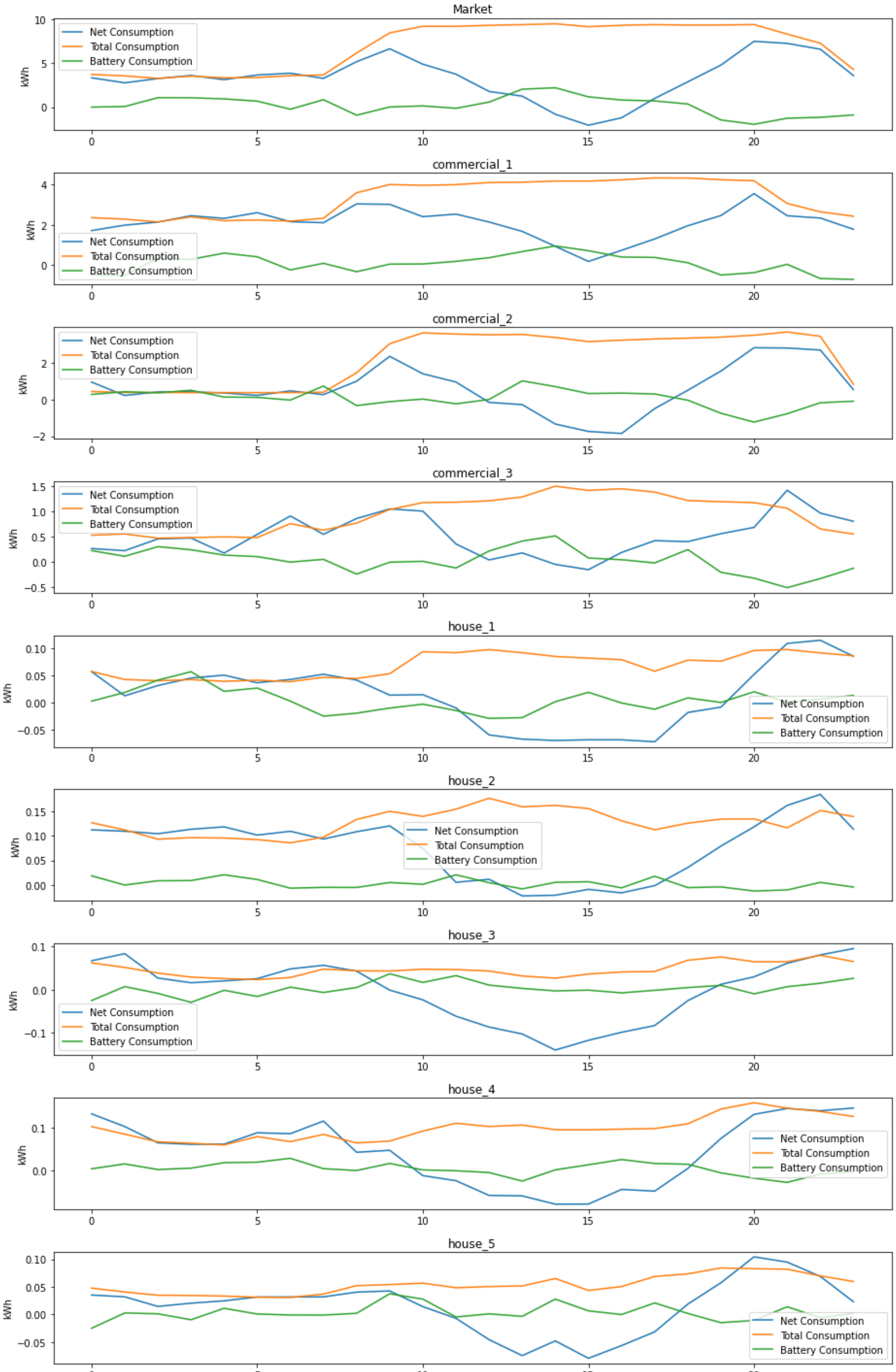

Figure 28: Hourly Averages of Net, Total, and Battery Consumption for MARL with 10 Episodes

The graph above provides hourly averages of net energy consumption, total energy consumption, and battery energy over May 2023. Typical behavior of total consumption can be observed here. The behavior of the battery also shows a tendency to charge during daytime and discharge during the night.

To compare the performance of the proposed model, rule-based model (RBRL) was constructed. In this model, agents adjust charging and discharging times and their amounts based on time of day. If the hour is between 9 AM and 8 PM, during solar generation times, the battery is charged. If the hour is before 9 AM or after 8 PM, the battery is discharged to meet consumption needs or engage in trading with the market. The charging and discharging rates are determined based on percentage ratios of battery capacity. Through tuning, it has been found that discharging a battery at a rate of 10% of its capacity and charging it at a rate of 7% of its capacity are the best options for rule-based agents. RBRL model was selected because of multiple reasons. Rule-based models are easy to understand and interpret. This makes them a useful tool for taking insights how different factors affect the agents' behaviors. They represent basic level of decision making, and this baseline helps to evaluate the added value of machine learning methods. Another important and the main reason for selecting benchmarking is people's habits. They play a significant role in energy consumption and trading decisions. Rule-based models can capture these human-like behaviors effectively.

Table 19 presents values of peer-based cost functions for both the proposed approach and RBRL. According to these values, the proposed approach has enabled agents to act in a way that preserves the interests of all peers compared to rule-based agents. The decrease in energy costs highlights the need for peers to adopt such an approach, as it benefits them collectively.

Table 19: Peer-based Cost Function Values for Proposed Approach and RBRL

	Electricity Cost		Net Electricity Consumption		Net Electricity Cons. with Negatives	
Peer	MARL	RBRL	MARL	RBRL	MARL	RBRL
commercial_1	1.010	1.102	1.016	1.121	1.010	1.044
commercial_2	1.031	1.152	1.043	1.202	1.018	1.079
commercial_3	1.068	1.098	1.076	1.124	1.025	1.047
house_1	1.036	1.220	1.042	1.257	1.040	1.121
house_2	1.025	1.080	1.026	1.097	1.014	1.035
house_3	1.088	1.469	1.101	1.522	0.809	0.525
house_4	1.038	1.134	1.042	1.160	1.018	1.052
house_5	1.074	1.298	1.072	1.334	1.077	1.216

Similarly, when considering the costs of the entire system shown below figure, it is observed that total energy consumption in the market decreases, and market players share their energy needs with each other. While both models yield similar results in peak demand, ramp cost is higher in the proposed approach. This suggests that the model may require further testing and improvements in terms of system reliability.

Table 20: Cost Function Values of Market for Proposed Approach and RBRL

Cost Function	MARL	RBRL
Electricity Cost	1.046	1.194
Ramp Cost	1.174	1.077
Net Electricity Consumption	1.052	1.227
Net Consumption with Negatives	1.002	1.015
Peak Demand	14.271	14.828
Daily Peak Demand	1.043	1.083
Load Factor	1.029	1.035

Finally, when comparing net electricity consumption for both peers and the entire market in the two models, it is evident that the proposed approach results in lower net electricity consumption on an hourly basis Figure 30. Solar energy was better utilized with the proposed method, leading to more efficient utilization of energy generated from solar sources.

Additionally, the distribution of net electricity consumption can be observed in histograms shown in Figure 29. The proposed approach exhibits net electricity consumption average closer to zero, and distribution appears to be more closely following normal distribution. However, rule-based reinforcement model shows distribution with higher net consumption, as evident in the histogram. The less skewed distribution of the proposed approach indicates a more consistent learning method.

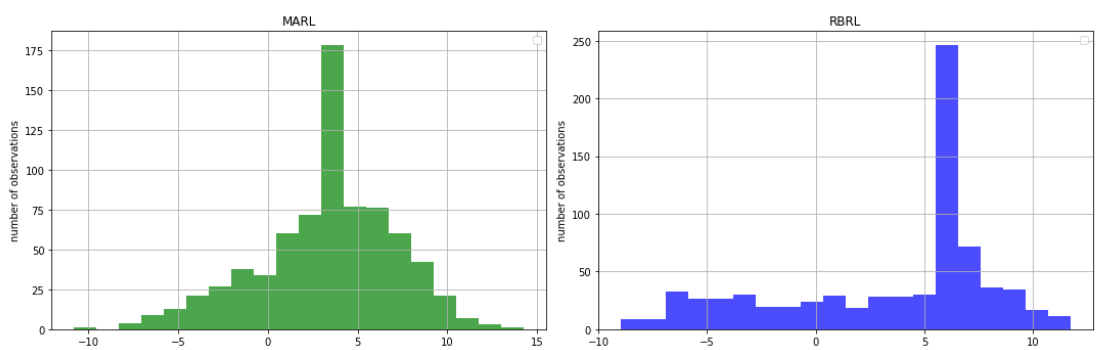

Figure 29: Histogram of Net Electricity Consumption for Proposed Approach and RBRL

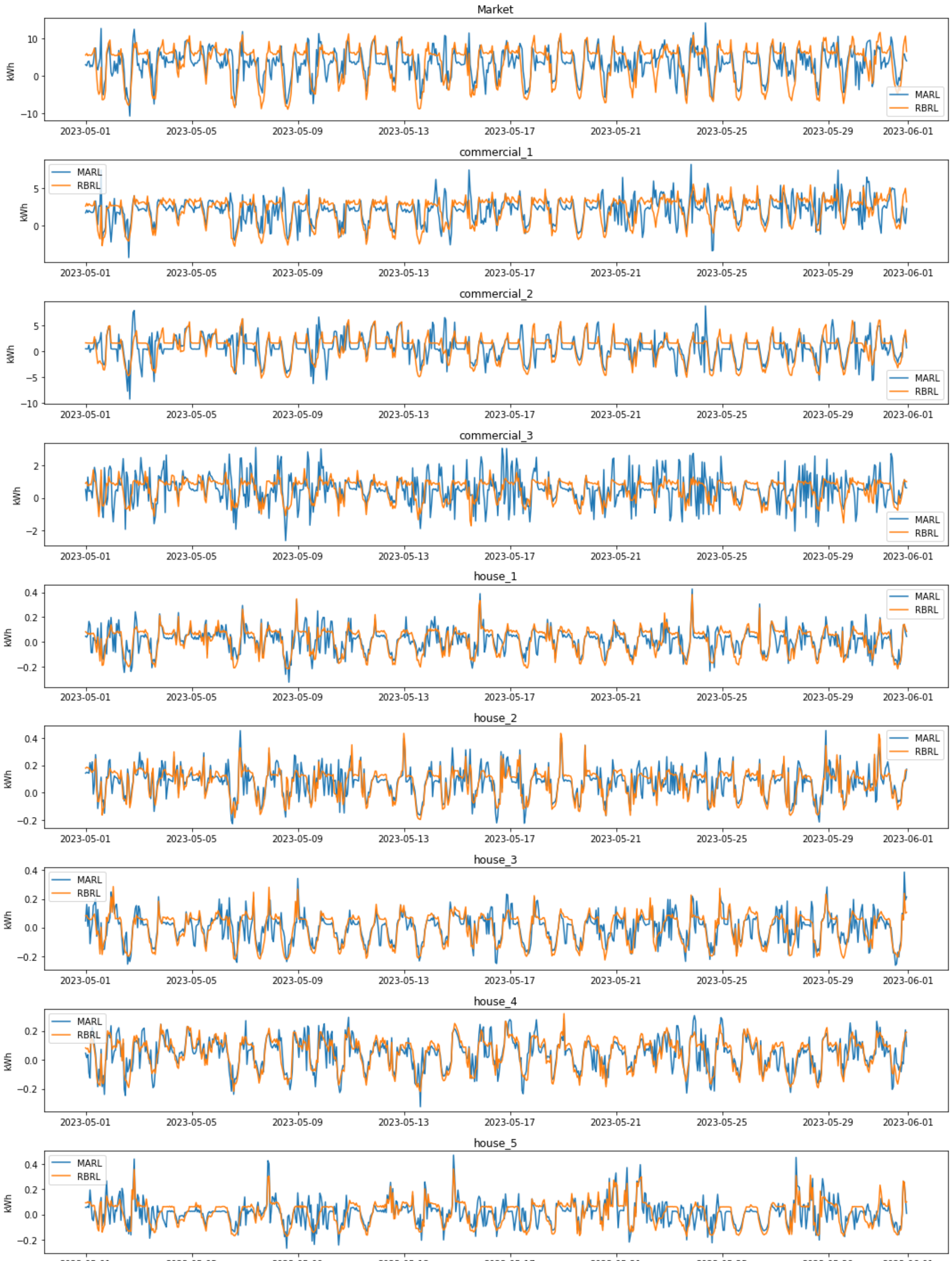

Figure 30: Net Electricity Consumption for Proposed Approach and RBRL

CHAPTER 5

CONCLUSION

5.1. Summary

The thesis study focuses on challenges posed by increasing energy demand and dependency on electrical energy in today's world. It addresses the importance of renewable energy in resolving energy-related issues and introduces peer-to-peer energy trading as a solution for local markets.

The concept of peer-to-peer (P2P) energy trading was introduced, allowing nano/microgrids to trade energy directly with each other without the need to involve main grid. The thesis aims to bring P2P energy trading system to a self-sufficient prosumer for each peer while maximizing the profit of each peer. It considers stochastic situations due to varying electricity demand and renewable energy generation, seeks to reduce prosumers' dependency on external sources, and minimizes energy losses through the management of energy sources.

Multi-agent reinforcement learning model was employed to optimize P2P energy trading game for each peer's goals. The thesis explores various research questions related to energy transfer among participants, determining electricity prices, renewable energy generation, local power load, and battery levels in trading policy. The thesis analyzes the pioneering simulation of P2P energy trading approach, utilizing data specific to Turkey. It examines the feasibility and potential benefits of implementing such a system within Turkish energy landscape.

The second chapter is a comprehensive exploration of machine learning discipline and its application in the context of energy trading and forecasting. Literature review was presented in three main sections, each addressing specific topics related to the subject matter. The first section of Chapter 2 provides an overview of machine learning, focusing on different learning paradigms commonly employed in literature. These paradigms included supervised learning, unsupervised learning, and reinforcement learning, which are essential in understanding machine learning techniques applied in energy trading. In the next section, the concept of energy trading was thoroughly explained, along with the models frequently utilized in this domain. The existing literature on energy trading was reviewed, including various market structures, pricing mechanisms, and trading strategies, providing valuable insights for energy market analysis. In the last section of Chapter 2, energy forecasting models, particularly those using supervised machine learning approaches for demand, generation, and price forecasting were searched. By examining the relevance of different techniques, this section aimed to identify the most suitable models for energy trading applications.

Chapter 3 delves into the methodology of P2P energy trading game among nano/microgrids. Before explaining the proposed approaches for forecasting and reinforcement learning, the problem was stated once more. Then, P2P energy trading platform description, enabling direct energy trading between producers and consumers without intermediaries, promoting efficiency, sustainability, and grid resilience was explained. It consists of two layers, a virtual layer facilitating energy trading decisions and a physical layer comprising tangible components like solar panels and batteries, which, along with market participants, form a seamless ecosystem. Additionally, energy models for loads, solar panels, and batteries were explained, considering load forecasting and generation forecast models.

The proposed methodology was presented in Chapter 3, it is the implementation of multi-agent reinforcement learning P2P energy trading solution for multiple peers with different capacities in terms of load, generation, and battery. Each peer is equipped with its reinforcement learning agent, aiming to optimize energy consumption and enable effective cooperation among peers, even when starting with random policies and limited knowledge about system dynamics. Before starting trading, load, generation, and price forecasts are prepared according to the proposed forecasting approach because they play crucial role in energy trading decisions. By examining load, generation, price forecasts, and reinforcement learning algorithms separately, this study aims to a provide valuable insights for future improvements and the development of intelligent energy management systems to enhance sustainability.

Chapter 4 explains the experiment conducted to evaluate P2P energy trading system including forecasting and reinforcement learning parts. Generation and load data was collected from 8 prosumers, including commercial buildings and households, over one and a half year. The experiment analyzes solar energy generation, energy consumption patterns, prices and incorporates real-world factors to create the market. The performance of models was evaluated using defined metrics and sensitivity testing.

After evaluating energy forecasts, multi-agent reinforcement learning (MARL) model and rule-based reinforcement learning (RBRL) model presented in Chapter 3 were trained and tested for the last month of data. The results were assessed based on defined cost functions such as net electricity consumption, ramping cost, and peak demand. As a result, it was demonstrated that agents who learn from their state, actions, and environment and engage in P2P trading with each other were more successful in terms of peer-based costs, system cost, and stability.

5.2. Future Work

In this thesis, presented forecasting and reinforcement learning-based efficient energy exchange strategies in peer-to-peer trading are intended to be further developed in the future. Future work can be expanded by introducing new constraints and considerations to improve P2P energy trading strategies.

Incorporating physical distance between nano/microgrids as a constraint can add realism to a trading system. Prosumers in closer proximity may have more favorable trading opportunities due to lower transmission losses and reduced transportation costs. Implementing distance-based constraints could optimize energy trading by promoting local energy exchange and reducing dependency on long-distance energy transfers.

Secondly, developing a buyer selection mechanism can enhance the efficiency and reliability of energy transactions. Prosumers could have the option to prioritize their buyers based on factors like reliability, reputation, or trading history. Implementing a robust buyer selection process would ensure that prosumers can find suitable and trustworthy partners for energy trading.

Apart from the trading side, expanding forecasting models to include wind power generation can further diversify renewable energy sources in the trading network. Accurate wind plant forecasts would allow prosumers to anticipate fluctuations in wind power generation and adapt their trading strategies accordingly.

The purpose of participating in such a market can vary for each prosumer. Indeed, prosumers may want to engage in trading for multiple purposes. Implementing multi-objective optimization techniques can enable prosumers to consider multiple criteria simultaneously, such as maximizing profit, minimizing environmental impact, and optimizing grid stability. Introducing multi-objective approaches would provide a more comprehensive analysis of tradeoffs and enable prosumers to make conscious decisions.

By incorporating these new constraints, objectives, and features, future research can advance P2P energy trading platform, making it more efficient, resilient, and sustainable. Addressing these aspects would contribute to the broader goal of fostering a decentralized and environmentally friendly energy ecosystem.

REFERENCES

Abdi, H. (2007). Z-scores. In *Encyclopedia of measurement and statistics 3* (pp. 1055-1058).

Alam, M. R., St-Hilaire, M., & Kunz, T. (2019). Peer-to-peer energy trading among smart homes. *Applied Energy, 238*, 1434-1443.

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica, 21*(4), 503-546.

Chandrasekaran, K., Kandasamy, P., & Ramanathan, S. (2020, July). Deep learning and reinforcement learning approach on microgrid. *International Transactions on Electrical Energy Systems, 30*(10).

Chen, T., & Bu, S. (2019). Realistic Peer-to-Peer Energy Trading Model for Microgrids using Deep Reinforcement Learning. *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)* (pp. 1-5). Bucharest, Romania, Romania: IEEE.

Chen, T., & Su, W. (2018, October 18). Local Energy Trading Behavior Modeling With Deep Reinforcement Learning. *IEEE Access, 6*, 62806 - 62814.

Cheng, Y., Xu, C., Mashima, D., Thing, V. L., & Wu, Y. (2017). Powerlstm: Power demand forecasting using long short-term memory neural network. *International Conference on Advanced Data Mining and Applications*, (pp. 727-740).

Clifton, J., & Laber, E. (2020). Q-Learning: Theory and Applications. *Annual Review of Statistics and Its Application*.

Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=72*, 240-254.

Conway, D., & White, J. M. (2012). *Machine Learning for Hackers.* United States of America: O'Reilly.

Deng, H., Yan, F., Wang, H., Fang, L., Zhou, Z., Zhang, F., . . . Jiang, H. (2021). Electricity Price Prediction Based on LSTM and LightGBM. *2021 IEEE 4th International Conference on Electronics and Communication Engineering (ICECE)*. Xian: IEEE.

Dudek, G. (2014). Short-Term Load Forecasting Using Random Forests. *Intelligent Systems'2014*, (pp. 821-828).

Elliott, E., Shanklin, N., Zehtabian, S., Zhou, Q., & Turgut, D. (2020). Peer-to-Peer Energy Trading and Grid Impact Studies in Smart Communities. *2020 International Conference on Computing, Networking and Communications (ICNC)* (pp. 674-678). Big Island, HI, USA: IEEE.

Ferrag, M. A., & Maglaras, L. (2019). DeepCoin: A Novel Deep Learning and Blockchain-Based Energy Exchange Framework for Smart Grids. *IEEE Transactions on Engineering Management, 67*(4), 1285-1297.

Grossi, L., & Nan, F. (2019). Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. *Technological Forecasting and Social Change*, 305-318.

Guan, C., Wang, Y., Lin, X., Nazarian, S., & Pedram, M. (2015). Reinforcement learning-based control of residential energy storage systems for electric bill minimization. *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)* (pp. 637-642). Las Vegas, NV, USA: IEEE.

Gueymard, C. A. (2009). Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications. *Solar Energy, 83*(3), 432-444.

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., . . . Levine, S. (2019). Soft Actor-Critic Algorithms and Applications. *Machine Learning*.

Hay, J. E. (1979). Calculation of monthly mean solar radiation for horizontal and inclined surfaces. *Solar Energy, 23*(4), 301-308.

Hong, T. (2014). Energy forecasting: Past, present, and future. *Foresight: The International Journal of Applied Forecasting*.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Very Short-Term Electrical Energy Consumption Forecasting of a Household for the Integration of Smart Grids. *Power and Energy*.

Hooshmand, A., & Sharma, R. (2019). Energy Predictive Models with Limited Data using Transfer Learning.

Hu, R. (2020). A Distributed Energy Management Strategy for Renewable Powered Communication Microgrid using Game Theory and Reinforcement Learning. *University of Pittsburgh*.

International Energy Agency. (2018). *World Energy Outlook 2018.* White Paper: IEA.

Jiayi, H., Chuanwen, J., & Rong, X. (2008). A review on distributed energy resources and MicroGrid. *Renewable and Sustainable Energy Reviews*.

Ju, Y., Sun, G., Chen, Q., Zhang, M., Zhu, H., & Rehman, M. U. (2019). A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting. *IEEE Access*.

Kang, J., Yu, R., Maharjan, S., Zhang, Y., & Hossain, E. (2017). Enabling Localized Peer-to-Peer Electricity Trading Among Plug-in Hybrid Electric Vehicles Using Consortium Blockchains. *IEEE Transactions on Industrial Informatics*, 3154-3164.

Kim, H., Lee, J., Bahrami, S., & Wong, V. W. (2019, July 2). Direct Energy Trading of Microgrids in Distribution Energy Market. *IEEE Transactions on Power Systems, 35*(1), 639 - 651.

Kim, J. G., & Lee, B. (2020, October 14). Automatic P2P Energy Trading Model Based on Reinforcement Learning Using Long Short-Term Delayed Reward. *Energies, 13*, 5359.

Kuznetsova, E., Li, Y.-F., Ruiz, C., Zio, E., Ault, G., & Bell, K. (2013). Reinforcement learning for microgrid energy management. *Energy, 59*, 133-146.

Lago, J., Marcjasz, G., Schutter, B. D., & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*.

Levent, T., Preux, P., Pennec, E., Badosa, J., Henri, G., & Bonnassieux, Y. (2019). Energy Management for Microgrids: a Reinforcement Learning Approach. *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)* (pp. 1-5). Bucharest, Romania, Romania: IEEE.

Long, C., Wu, J., Zhang, C., Thomas, L., Cheng, M., & Jenkins, N. (2017). Peer-to-peer energy trading in a community microgrid. *2017 IEEE Power & Energy Society General Meeting.* Chicago, IL, USA: IEEE.

Mackay, D. J. (2008). *Sustainable Energy - without the hot air.* UK: UIT Cambridge.

Mamun, A. A., Sohel, M., Mohammad, N., Sunny, M. S., Dipta, D. R., & Hossain, E. (2020). A Comprehensive Review of the Load Forecasting Techniques Using Single and Hybrid Predictive Models. *IEEE Access*.

Mandal, P., Senjyu, T., Urasaki, N., & Funabashi, T. (2006). A neural network based several-hour-ahead electric load forecasting using similar days approach. *International Journal of Electrical Power & Energy Systems, 28*(6), 367-373.

Mellit, A., & Pavan, A. M. (2010). A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Solar Energy, 84*(5), 807-821.

Mengelkamp, E., Gärttner, J., Rock, K., Kessler, S., Orsini, L., & Weinhardt, C. (2018, January 15). Designing microgrid energy markets: A case study: The Brooklyn Microgrid. *Applied Energy, 210*, 70-880.

Mir, A. A., Khan, Z. A., Altmimi, A., Badar, M., Ullah, K., Imran, M., & Kazmi, S. A. (2021). Systematic Development of Short-Term Load Forecasting Models for the Electric Power Utilities: The Case of Pakistan. *IEEE Access*.

Mohamed, N., Ahmad, M. H., Suhartono, & Ismail, Z. (2011). Improving Short Term Load Forecasting Using Double Seasonal Arima Model. *World Applied Sciences Journal*, 223-231.

Muzaffar, S., & Afshari, A. (2019). Short-Term Load Forecasts Using LSTM Networks. *Energy Procedia, 158*, 2922-2927.

Nassif, A. B., Soudan, B., Azzeh, M., Attilli, I., & AlMulla, O. (2021). Artificial Intelligence and Statistical Techniques in Short-Term Load Forecasting: A Review. *International Review on Modelling and Simulations (I.RE.MO.S.)*.

Nengbao, L., Babushkin, V., & Afshari, A. (2014). Short-Term Forecasting of Temperature Driven Electricity Load Using Time Series and Neural Network Model. *Journal of Clean Energy Technologies*.

Nguyen, S., Peng, W., Sokolowski, P., & Alahakoon, D. (2018). Optimizing rooftop photovoltaic distributed generation with battery storage for peer-to-peer energy trading. *Applied Energy*.

Nordman, B., & Christensen, K. (2015). DC Local Power Distribution with microgrids and nanogrids. *2015 IEEE First International Conference on DC Microgrids (ICDCM)*.

Nunna, H. K., Sesetti, A., Rathore, A. K., & Doolla, S. (2020). Multiagent-Based Energy Trading Platform for Energy Storage Systems in Distribution Systems With Interconnected Microgrids. *IEEE Transactions on Industry Applications*.

Okada, S., Ohzeki, M., & Taguchi, S. (2019). Efficient partition of integer optimization problems with one-hot encoding. *Scientific Reports*.

Park, S., Jung, S., Jung, S., RHo, S., & Hwang, E. (2021). Sliding window-based LightGBM model for electric load forecasting using anomaly repair. *The Journal of Supercomputing*.

Paudel, A., Sampath, L., Yang, J., & Gooi, H. B. (2020, May 27). Peer-to-Peer Energy Trading in Smart Grid Considering Power Losses and Network Fees. *IEEE Transactions on Smart Grid, 11*(6), 4727 - 4737.

Perez, C. (2019). *Machine Learning Techniques: Unsupervised Learning. Examples with Matlab.* Amazon Digital Services LLC - KDP Print US.

Perez, R., Ineichen, P., Seals, R., Michalsky, J., & Stewart, R. (1990). Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy, 44*(5), 271-289.

Persson, C., Bacher, P., Shiga, T., & Madsen, H. (2017). Multi-site solar power forecasting using gradient boosted regression trees. *Solar Energy, 150*, 423-436.

Pong, V., Gu, S., Dalal, M., & Levine, S. (2020). Temporal Difference Models: Model-Free Deep RL for Model-Based Control. *Machine Learning*.

Qiu, D., Wang, J., Dong, Z., Wang, Y., & Strbac, G. (2022). Mean-Field Multi-Agent Reinforcement Learning for Peer-to-Peer Multi-Energy Trading. *IEEE Transactions on Power Systems*.

Rahimi, N., Park, S., Choi, W., Oh, B., Kim, S., Cho, Y.-h., . . . Lee, D. (2023). A Comprehensive Review on Ensemble Solar Power Forecasting Algorithms. *Journal of Electrical Engineering & Technology*, 719-733.

Raju, L., Sankar, S., & Milton, R. S. (2015). Distributed Optimization of Solar Micro-grid Using Multi Agent Reinforcement Learning. *Procedia Computer Science, 46*, 231-239.

Rousseuw, P. J. (2011). Robust statistics for outlier detection. *Wires Data Mining and Knowledge Discovery*.

Sarda, J. S., Lee, K., Patel, H., Patel, N., & Patel, D. (2022). Energy Management System of Microgrid using Optimization Approach. *IFAC-PapersOnLine*.

Sobri, S., Koohi-Kamali, S., & Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. *Energy Conversion and Management*.

Soto, E. A., Bosman, L. B., Wollega, E., & Leon-Salas, W. D. (2021). Peer-to-peer energy trading: A review of the literature. *Applied Energy*.

Stadler, M., & Nasle, A. (2019). Planning and implementation of bankable microgrids. *The Electricity Journal*.

Sutton, R. S., & Andrew, B. G. (n.d.). *Reinforcement learning: An introduction.* 2018: MIT press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning, second edition: An Introduction (Adaptive Computation and Machine Learning series)*. Bradford Books.

Taieb, S. B., & Hyndman, R. J. (2014). gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting*, 382-394.

Trapero, J. R., Kourentzes, N., & Martin, A. (2015). Short-term solar irradiation forecasting based on Dynamic Harmonic Regression. *Energy, 84*, 289-295.

Tushar, W., Saha, T. K., Yuen, C., Smith, D., & Poor, H. V. (2020). Peer-to-Peer Trading in Electricity Networks: An Overview. *IEEE Transactions on Smart Grid*, 3185-3200.

Tushar, W., Yuen, C., Mohsenian-Rad, H., Saha, T., Poor, H. V., & Wood, K. L. (2018). Transforming Energy Networks via Peer to Peer Energy Trading: Potential of Game Theoretic Approaches. *Signal Processing*.

Uriarte, F. M. (2019, May 6). *Pecan street project field data*. Retrieved from https://sites.google.com/site/fabianuriarte/downloads

Vazquez-Canteli, J. R., Henze, G., & Nagy, Z. (2020). MARLISA: Multi-Agent Reinforcement Learning with Iterative Sequential Action Selection for Load Shaping of Grid-Interactive Connected Buildings. *Sensys*.

Wang, H., Huang, T., Liao, X., Abu-Rub, H., & Chen, G. (2016, April 27). Reinforcement Learning in Energy Trading Game Among Smart Microgrids. *IEEE Transactions on Industrial Electronics, 63*(8), 5109 - 5119.

Wang, J., Li, L., & Zhang, J. (2022). Deep reinforcement learning for energy trading and load scheduling in residential peer-to-peer energy trading market. *International Journal of Electrical Power and Energy Systems*.

Wang, W., & Huang, J. (2018). Incentivizing Energy Trading for Interconnected Microgrids. *Transactions on Smart Grid*, 2647-2657.

Wang, Z., Chen, B., Wang, J., & Kim, J. (2016). Decentralized Energy Management System for Networked Microgrids in Grid-Connected and Islanded Modes. *Transactions on Smart Grid*, 1097-1105.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 1030-1081.

Wu, Y.-K., Huang, C.-L., Phan, Q.-T., & Li, Y.-Y. (2022). Completed Review of Various Solar Power Forecasting Techniques Considering Different Viewpoints. *Energies*.

Xiao, L., Xiao, X., Dai, C., Peng, M., Wang, L., & Poor, H. V. (2018, June 19). Reinforcement Learning-based Energy Trading for Microgrids.

Yang, C., Thatte, A. A., & Xie, L. (2006). Multitime-Scale Data-Driven Spatio-Temporal Forecast of Photovoltaic Generation. *IEEE Power Engineering Society General Meeting.*

Yang, Z., Ce, L., & Lian, L. (2017). Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. *Applied Energy*, 291-305.

Yao, X., Fu, X., & Zong, C. (2022). Short-Term Load Forecasting Method Based on Feature Preference Strategy and LightGBM-XGboost. *IEEE Access*.

Ye, Y., Tang, Y., Wang, H., Zhang, X.-P., & Strbac, G. (2021). A Scalable Privacy-Preserving Multi-Agent Deep Reinforcement Learning Approach for Large-Scale Peer-to-Peer Transactive Energy Trading. *IEEE Transactions on Smart Grid*.

Zahraoui, Y., Alhamrouni, I., Mekhilef, S., Khan, M. R., Seyedmahmoudian, M., Stojcevski, A., & Horan, B. (2021). Energy Management System in Microgrids: A Comprehensive Review. *Sustainability*.

Zeng, P., Sheng, C., & Jin, M. (2019). A learning framework based on weighted knowledge transfer for holiday load forecasting. *Journal of Modern Power Systems and Clean Energy, 7*(2), 329-339.

Zhang, C., Wu, J., Long, C., & Cheng, M. (2017, May). Review of Existing Peer-to-Peer Energy Trading Projects. *Energy Procedia, 105*, 2563-2568.

Zhang, J.-L., Zhang, Y.-J., Li, D.-Z., Tan, Z.-F., & Ji, J.-F. (2019). Forecasting day-ahead electricity prices using a new integrated model. *Forecasting day-ahead electricity prices using a new integrated model*, 541-548.

Zhang, K., Yang, Z., & Başar, T. (2021). Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms.

Zhou, L., Wang, B., Wang, Z., Wang, F., & Yang, M. (2018). Seasonal classification and RBF adaptive weight based parallel combined method for day-ahead electricity price forecasting. *2018 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT).*

Zhou, Y., Lin, Q., & Xiao, D. (2022). Application of LSTM-LightGBM Nonlinear Combined Model to Power Load Forecasting. *Journal of Physics: Conference Series*.

Zhou, Y., Wu, J., Long, C., & Ming, W. (2020, July). State-of-the-Art Analysis and Perspectives for Peer-to-Peer Energy Trading. *Engineering, 6*(7), 739-753.

Zor, K., Timur, O., Çelik, Ö., Yıldırım, B., & Teke, A. (2018). Very Short-Term Electrical Energy Consumption Forecasting of a Household for the Integration of Smart Grids.