Hybrid LSTM-Based Renewable Energy Forecasting and Grid Optimisation for Sustainable Power Management

Srikar Kalle
Master of Science in Data Science
South East Technological University
Carlow, Ireland
C00313529@setu.ie

ABSTRACT

This research offers a complete model of predicting the production of a renewable energy source by dynamically accounting for meteorological data and past performance data. Based on 10 years of information on weather and energy markets, we designed functionalities like lagged generation value, seasonal indicator, and major weather measurements (temperature, solar irradiance, wind, and precipitation). The data preprocessing process involved cleaning, normalization, and creating a reduced dimension to the data that was supposed to be of high quality on entry to the modeling process. To overcome the challenges in time series forecasting and non-linearity of the feature interactions unique in the renewable-based energy systems, a battery of machine learning methods through Long Short-Term Memory (LSTM), Random Forest Regressor, and XGBoost Regressor approaches was created and compared.

The LSTM model was capable of capturing temporal dependencies and short-term fluctuations very well and was especially successful in times of volatile generation. Random Forest allowed the user to gain interpretable insights, determining that feature importance and past generation were the main drivers, with wind speed being a secondary factor. XGBoost performed better than other models in overall accuracy as well as generalization, and it was able to track the events of peak generation well, whilst ensuring that overfitting does not occur thanks to regularization and ensemble learning. Such synergistic advantages point to the potential of jointly training deep learning with decision trees to guide energy stakeholders toward resolute and practical projections.

In order to convert forecasting into operational value, we have developed a grid risk simulation module that helps quote supply-demand mismatches on the basis of errors in the prediction. The simulation is classified in terms of possible times of under-supply or over-supply, which gives a source of decision-supporting information used by grid operators and other market participants. The results highlight the value of sophisticated modelling methods and real-time risk evaluation in facilitating viable grid management and making progress on the net-zero energy goals.

1 Introduction

With the growing incorporation of renewable energy resources like solar and wind power into the current power grids, new complications have emerged in energy prediction and grid stability. The renewable sources are naturally intermittent and unpredictable since they rely on the weather conditions and are associated with fluctuations in the power that result in having an energy surplus or shortage. Poor forecasting may lead to ineffective grid management, greater use of fossil fuel backup systems, and a rise in operating expenses, as well as the risk of blackouts[1]. Furthermore, as the global transition to wind and solar energy gathers pace in a bid to hit net-zero carbon emissions, these issues become ever more acute as countries seek solutions to support them, and a strong and AI-powered forecasting tool that can combine numerous sources of information into the model to increase its predictive accuracy and reliability is required [2].

AI and ML hold a lot of potential in terms of renewable energy forecasting and management. Hybrid models that incorporate deep learning architectures include CNN-LSTM and GCN-LSTM, which have shown a superior ability to capture spatial and temporal dependencies in power generation data and hence improved the forecasting accuracy [3]. Moreover, recent studies have discussed reinforcement learning applications in optimizing energy storage and real-time trading, facilitating integration of renewable energy into the energy markets efficaciously ([4]. Nonetheless, in ML-based forecasting, there still is a BL in using multiple types of data, e.g., weather conditions and energy market indicators, into a single data modeling system capable of forecasting energy-related variables on the one hand and estimating risks of a supply-demand imbalance on the other hand [5].

This paper targets these knowledge gaps by constructing an extensive framework that integrates weather and market data, cleans it by eliminating inconsistencies and multicollinearity relations, and, on top of that, not one but three of the latest modelling approaches to be tested: LSTM, Random Forest, and XGBoost. The framework also integrates a grid risk simulation that groups imbalances in the supply-demand into categories concerning forecast error to give a realistic judgment on the risks to the operations [6]. The integration of forecasting accuracy with risk quantification will help the stakeholders in the renewable energy market to help them make informed decisions in trading, storage, and distribution of energy. The research also leads to better prediction accuracy, as well as paves the way to real-time energy management and trading strategies with reinforcement learning techniques in a direction that supports the global net-zero and sustainable energy system goal [7].

1.1 Aim

The first main objective of the current research is to create and test a comprehensive machine learning framework to predict variables related to renewable energy, utilizing the data on weather and the energy market, and evaluate risks on grid stability by simulating an imbalance in supply and demand. The research aims to complement the shortcomings of conventional forecasting models by utilizing superior prediction algorithms like LSTM, Random Forest, and XGBoost to enhance forecasting capability and pursue temporal and nonlinear associations in the information. The framework proposes to optimize the use of diverse sources of data by means of intensive preprocessing and methods of analysis like Principal Components Analysis and checking multicollinearity, which will allow making energy predictions more reliable. Moreover, the study implements a risk simulation tool to classify deficiencies in simulations based on actual and predicted values, and this gives timely insights into grid stability. The proposed research will eventually lead to backing the energy players in making the best choices concerning renewable energy integration, storage, and trade, thus making contributions to grid stability, economic efficiency, and net-zero carbon targets.

1.2 Objectives

- Combine multi-source data to forecast: Integrate weather information with market-based data to generate comprehensive data that determines the environmental and economic factors affecting the production of renewable energy.

- Preprocess and transform data: Deal with missing values, eliminate irrelevant columns, multicollinearity, and standardize or normalize features to train a good model efficiently and robustly.

- Build and compare forecasting models: Train LSTM, Random Forest, and XGBoost models to predict variables concerning renewable energy and compare them with RMSE.

- Model grid risks using forecast errors: Develop a risk simulation system to categorize the imbalances according to under-supply, balanced, or over-supply, to assist in analysing grid stability.

- Offer energy planning and net-zero aspirations: The results of the models will enable stakeholders to streamline their energy exchange, storage, and generating regimes by anticipating and advancing towards clean energy adoption.

1.3 Research Gap

Although the implementation of machine learning in renewable energy forecasting continues to increase, a few limitations are yet to be tackled in the current research works. Most weather-based and market-based forecasting methods only use one data source, either weather or market data, and they cannot demonstrate the synergetic effect of the environmental and economic data on the energy generation and distribution [8]. Moreover, some studies have demonstrated potential of hybrid deep learning (e. g., CNN-LSTM) to be utilized in forecasting short-term loads or solar production [9]; however, not many studies are dedicated to coupling such models with other types of ensembles, e. g., Random Forest and XGBoost, and to the subsequent comparison of their performances.

The last and one of the most alarming discontinuities is the absence of a structure that would allow the forecasting of renewable energy adequately, but also evaluate the mere risks related to finding supply-demand disparities. Potential novelties have not significantly benefited most of the existing works on energy trading and grid planning [5], which have highlighted optimization methods and reinforcement learning approaches and failed to comprehensively incorporate more advanced forecasting with risk simulation to estimate the grid stability.

Finally, the responses to the previous research on the topic of AI-assisted energy management to reach net-zero objectives rarely consider the combination of forecasting models with decision-making, enabling the provision of actionable insights to stakeholders [6]. This paper presently observes these gaps, integrating multi-origin data, applying and comparing numerous forecasting models, and simulating risks to offer a feasible support system in decisions of integrating with renewable energy and optimizing the grid [2].

1.4 Ethical Considerations

The key ethical aspects of this research work are associated with the proper utilization of artificial intelligence and data in the renewable energy sector forecasting and grid management. Since the study involves using large-scale weather and market data, data privacy, data security, and data compliance with the respective regulations are important to consider. The transparency that is created behind the creation of AI models and the openness to its use are essential to avoiding the bias in the algorithms that can result in unfair distribution of energy or the unfair manipulation of the market itself[10]. Also, the applied AI technical solutions to energy optimization require sustainable best practices, whereby the developing models should contribute to minimizing carbon emissions and assist in achieving global climate conditions, instead of bolstering the energy consumption of the computational model [11].

2 Literature Review

The world is undergoing a shift towards energy production and modernization, a situation that demands a reduction in the rate of carbon emissions and a transition towards more sustainable energy production. Diverse developments in Canada and Europe lead to the integration of renewable energy sources like solar and wind into the national power grids and have both opportunities and challenges. Although these sources are advantageous to the environment, they are not constant in their occurrence, which adds unpredictability to energy supply, and this leads to generation-demand imbalances[12]. Such variations increase the difficulty of grid stability and the cost of operation, especially in case of inaccurate forecasting and responsive grid management. With nations establishing ambitious goals to attain net-zero emissions, smart grids and intelligent energy management systems are vital in achieving reliable, flexible, and resilient sources of power to support energy production.

The application of artificial intelligence and machine learning has become an effective element in tackling these obstacles as a result of the possibility to make real-time decisions and predictions, optimization, and automation of different systems in the energy sector. Meanwhile, AI technologies are transforming the work of modern power grids by making them more precise in terms of renewable energy prediction as well as energy storage and distribution optimization[13]. The body of literature in this area indicates that the cases relating to AI use are limited primarily to prediction-based algorithms, reinforcement learning for grid control, peer-to-peer trading platforms, and cybersecurity systems that can secure the sensitive infrastructure. This literature review is a synthesis of these research contributions, and it reveals the existing solutions, their efficacy and the existing limitations. The review has given the theoretical background of coming up with an integrated AI-based model of smart and sustainable energy management by delving into technological advancements and research deficiencies[3].

2.1 Forecasting Renewable Energy Generation

The ability to forecast the generation of renewable energy sources is an essential part of maintaining a stable and efficient power grid because, due to their characteristic intermittent nature, solar and wind forms of renewable energy cannot be relied upon to address every power grid operational situation. The traditional statistical tools like ARIMA and exponential smoothing have been inadequate in explaining dynamics and complex and nonlinear data on renewable energy [14]. To curb these shortcomings, there has been an uprising in the adoption of deep learning models. As an illustration, it is possible to mention hybridized architectures that use nearly identical neural networks e.g., Convolutional Neural Networks coupled with Long Short-Term Memory networks (CNN-LSTM) that yield good outcomes in terms of solar power output forecasting. These models are really good at capturing the local spatial characteristics by employing CNN layers and then expressing the temporal dependence by utilizing LSTM layers. As it was shown by Lim et al., the models of this type enable a significant increase in the accuracy of forecasts due to their ability to learn weather patterns and historical trends, and this is a significant aspect in the context of effective power dispatch and participation in the energy market [8].

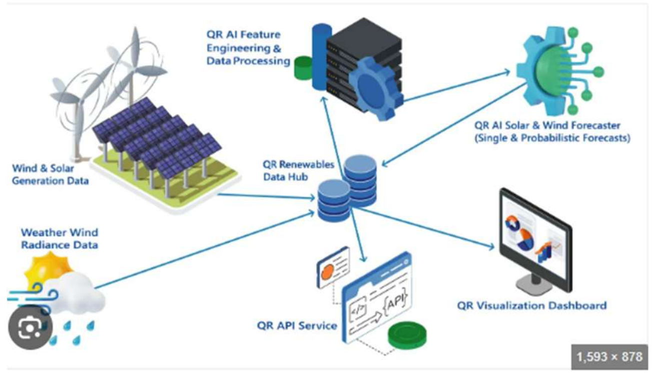

Figure 1: Solar Wind AI Generation Forecast Solution Software Service

Further enhancements have been made, combining spatial learning with temporal learning into one system. The prediction of short-term power with the help of Graph Convolutional Networks with LSTM [3]. To capture spatial interactions between various locations of energy sources and geographic points, the GCN component was utilized, whereas the LSTM component took care of the temporal dynamics of the generation data. Such a dual architecture enabled this model to generalize and tune to different environmental and operating conditions far better. The paper suggested its usefulness in regional[15]. This is because in regional forecasting, there are wide variations in local climatic conditions which have to be included in the prediction models. The utility of such high-end structures is manifest in the significance of integrating information provided by various sources within the context of pointing at meteorological, historical and grid-based with respect to efficient energy-related operations, especially those involving energy management [12].

On the consumer and household level, there is also a massive forecasting advantage presented by AI models. Alhussein et al. used a CNN-LSTM hybrid model to conduct short-term load forecasting at the household level, which made it possible to plan better within a microgrid framework[9]. The models largely apply in the case of decentralized renewable systems in which localized forecasts govern energy storage, energy consumption, and peer-to-peer energy trading. Besides, Kim et al. stressed ML application in real-time forecasting and control of renewable sources due to the need to adjust to the dynamic shifts in conditions[8]. They combine data flows generated by sensors and weather stations into control algorithms so that how much is being predicted to be produced can be changed nearly immediately. This is essential in preserving grid stability and maximize level of renewable penetration. The importance of intelligent forecasting tools in controlling centralised and distributed renewable energy system is highlighted because of these findings [16].

2.2 Grid Optimization and Energy Storage

The operation of the electricity grid to guarantee reliability and efficiency becomes more advanced, in consideration of the growing supply of renewable sources of energy in the energy mix across the globe [16]. The problem of a mismatch between renewable energies and the real-time demand in electricity has proven to be one of the greatest problems. As opposed to the older fossil-based systems, the renewable sources like solar and wind power are unstable and quite unpredictable. To tame these issues, solutions based on AI enhancements (especially AI based on reinforcement learning (RL)) have also been suggested to improve optimization of the grid. Fuxjager et al. provided an RL framework of an optimal day-ahead planning and AI-aided grid control capable of dynamically remodeling energetic flow paths with variations in generation patterns[4]. Their model is always learning based on the historical information and real-time data to make better decisions during load dispatching and voltage control, and the optimization of power flow. The flexibility of these systems to operate with adaptive responses to uncertain conditions makes these systems very suitable in the power grids of the future, which will be mainly renewable [17].

An energy storage system (ESS) is a major element in smoothing the fluctuations of renewable energy. Storing the surplus energy during the times of intensive generation and releasing it through the times of peak demand or low power output, ESS contributes to the flexibility and reliability of the grid. To get the economic returns, the positioning and sizing of these systems have to be optimized to make transmission losses as low as possible. The implementation of the genetic algorithm was carried out in Kiani and Ghosh to achieve optimal storage sites on the basis of a distribution network that demonstrated significant power quality and voltage stability[18]. They threw the idea into the mix of both technical limitations and economic concerns and offered a complete solution to the deployment of storage. In addition, hybrid solutions that involve the integration of storage with predictive load balancing have the potential to be more efficient in operation. The optimization techniques do not only lead to a decrease in the cost of operation but also increase the reliability of the grid with high renewable penetration [12].

Regarding the industrial sector, energy optimization using AI is being progressively utilized to modernize large-scale industrial energy systems. The study conducted by Ashraf et al. represented the use of artificial intelligence modeling in order to improve the work of an industrial-sized steam turbine[6]. In their work, they point at the potential of machine learning algorithms to optimize turbine operating conditions, which might result in vast increases in thermal efficiency and a decrease in fuel consumption [17]. The developments go hand in hand with targeting the net-zero energy systems, where the overarching priorities are to limit the use of fossil fuels. Embedding AI into industrial energy systems will allow the operators to dynamically rebalance generation and load according to the situation in the grid, thus contributing to grid wide optimization endeavors. In short, there is a good building block in using AI-powered forecasting, grid management, and storage control that would facilitate the shift towards a wise and renewable energy system [19].

2.3 Decentralized Trading and Peer-to-Peer (P2P) Models

With the changes in traditional energy systems, decentralized energy markets have emerged as a transformational model of local energy distribution. Such a type of configuration allows for prosumers, or those who both make and take electricity, to deal with one another directly on a source-to-source basis and does not depend on centralized utilities. Such a transition is enabled by smart contracts, real-time pricing, and AI operational energy management systems. The three-stage multi-energy trading strategy of Yang et al. built under the P2P model simplifies the process of trading with energy as all measures of trade in the model are separated into three steps: generation of forecasting generation, deciding on the markets, and the real trading actions[7]. Their system smoothed local energy systems' flexibility and efficiency, giving way to negotiation of prices and surplus generation in microgrids meetings and avoiding transmission losses, and better local grid stability.

Reinforcement learning is effective in such a decision-making process in such decentralized setups in a real-time analysis. Ju and Crozier have come up with a deep RL local trading strategy using which energy suppliers can change their behaviour dynamically to adjust to the imbalances and changes in supply and demand and prices[5]. Their model gave not only to the best bid-trading behavior but also enforced that grid constraints were not violated and economic returns were maximized. This dynamic learning capacity is very crucial in the renewable-rich environment, where generation is uncertain and highly dynamic. The RL-based P2P systems show the potential to surpass the rule-based or fixed-optimization processes in supporting sustainability aspects of energy availability and financial viability to participants, as such systems may continuously update their policies on the basis of feedback provided by the environment [20].

Besides, the distributed nature of trading models is also an important factor in democratizing energy and energy equity. In standard grid systems, the energy exclusion of low-income or rural areas or subpar service is common. P2P networks make communities energy independent as local areas are able to generate and consume energy without large-scale grid infrastructures. The networks also give stimuli to small-scale renewable technologies, including rooftop solar, house-top wind turbines, or micro-hydro units. The more people who transition into energy trading, the more peak demand stress is reduced and the easier integration of distributed energy resources. Yang et al. underline that the consumption of P2P trading also promotes behavioral changes because users are more aware of their production and consumption processes due to the feedback loop induced by the market[7]. Therefore, the decentralized Energy Architectures, which depend on the intelligent trading algorithms, not only make the Energy more efficient but also result in the sustainability of the environment and the economy [16].

2.4 AI for Net-Zero Energy Systems and Sustainability

The possibility to move to a net-zero carbon emissions status has become a world agenda, and artificial intelligence (AI) is flourishing as a central facilitator in this process. The AI technologies enable amount of monitoring, prediction and optimization of the energy systems that is not quite comparable to the conventional means. When applied to building energy management, Zeiler (2022) put forward the notion of the brains of the buildings, which refers to autonomous control of lighting, heating, cooling, and ventilation to reduce carbon footprints by the AI-aided systems[11]. These systems can predict energy requirements and react dynamically to the conditions via use of real-time information, making them important tools in the process of streamlining such buildings with the goals of national and international sustainability [18].

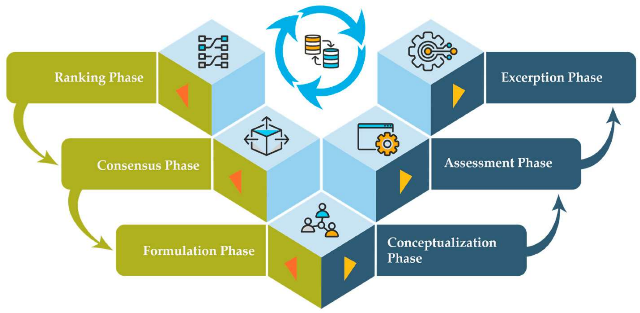

Figure 2: AI-Enabled Energy Policy for a Sustainable Future

On a higher scale of systems, policy and infrastructure planning are increasingly taking advantage of AI to enhance the achievement of sustainable energy development. [2], explored the worldwide potential and risks in achieving net-zero CO2 emissions, particularly the use of AI to support the technological and economic realities of energy transitions. AI enables optimal distribution of energy resources and demand prediction as well as optimization of supply chains. Where energy infrastructure is weak, such as in developing countries, AI can inform an investment plan and specify the focus on renewable integration dimensions by environmental and economic models. In addition, AI can assist policymakers in modeling the future energy conditions, exploring tradeoffs, and developing data-based strategies that will be flexible to changes in uncertainty and climate variability [16].

Although AI is dealing with the increase in energy efficiency and sustainability, it is vital to take care of the environmental impact of AI. Training of the deep learning models is usually computationally expensive, adding up to the carbon emissions. Lacoste et al. drew attention to the fact that the total environmental cost of training large-scale machine le Haq arming models has become a major issue, raising concerns about how the energy and AI communities can move to greener computing[10]. The green AI concept attempts to promote the development of computer models that would be friendly to the environment and economical in terms of computation. In terms of energy sector usage, this implies choosing algorithms that maximize performance through energy costs and encouraging the carbon cost of computation transparency. With the increasing integration of AI in net-zero movements, sustainable measurement should be evaluated not only by results but also by the instruments to obtain them.

2.5 Smart Grid Security and Data Integrity

With the attraction of AI-based decision-making and the real-time energy data feature, the threat of cybersecurity in smart grids has increased significantly. These next-generation energy systems are dependent on the open communication among the sensors, controllers and centralized management systems. Manipulation in this type of communication or any interruption may cause cascading failures, grid instability, or even blackouts on a large scale. The common security systems become insufficient in identifying advanced forms of cyber-attacks like zero-day attacks, which are vulnerabilities that have not yet been patched and the attacker strikes before a response or patch is developed. In response to this, researchers have resorted to applications of machine learning technology to improve anomaly detection. Mbona and Eloff have suggested the field of semi-supervised machine learning in detecting zero-day intrusion attacks in the smart grid[21]. Their configuration was able to learn from small volumes of labeled data as well as detect patterns that do not conform to the regular system behavior, thereby facilitating early threat identification without necessarily relying on the well-known attack signatures [18].

Information security is also an essential consideration in smart grids since these AI algorithms consider precise and reliable inputs as a consideration in making a real-time decision. Automated energy trading systems involve trading and balancing of loads that can be incorrectly done using corrupted and falsified data, resulting in loss of money via economic losses. The preservation of the authenticity of data in distributed systems would include the use of cryptography, blockchain-based technology and sound validation algorithms [12]. Though certainly not a simple solution, blockchain solutions present the possibility of relatively transparent and immutable records of energy transactions, meter reads, and control signals. Not only is this decentralized architecture more trustworthy, but it also offers protection against the single point of failure. Moreover, the validation model based on AI can watch over inconsistencies in data continuously, raise doubts about suspicious activities, and even segregate malfunctioning elements or channels of communication even before they disturb the system[21].

Moreover, the shift toward decentralized and prosumer-driven energy systems introduces additional cybersecurity challenges. As more endpoints—such as smart meters, home solar inverters, and IoT-enabled appliances—connect to the grid, the attack surface expands considerably. These devices often lack rigorous security protocols, making them vulnerable entry points for adversaries [12]. Integrating lightweight AI agents at the edge of the network can help mitigate this risk by providing local anomaly detection and minimal latency in response actions. The adoption of such proactive security frameworks ensures that grid resilience is maintained even in the face of emerging threats. In smart grids where real-time data is the backbone of energy forecasting, storage control, and trading algorithms, safeguarding this data is paramount not just for operational efficiency but also for national energy security [21].

Besides, the trend toward decentralization and prosumer-centered energy leads to some other cybersecurity issues. The attack surface is large as more endpoints, including smart meters, home solar inverters, and appliances with the Internet of Things, join the grid. Such devices do not have bad security mechanisms, and they can be a weak point of attack by the opponents [17]. On the one hand, this risk can be addressed by introducing lightweight AI agents at the network edge that would have visibility to find a local anomaly and act with minimal latency. Engagement of these defensive security structures guarantees that the resilience of the grid is preserved even despite such emerging threats. In smart grids with real-time data serving as the basis of the energy forecasting, storage control, and trading algorithms, such data protection is crucial both in the context of operational efficiency and national energy security [21].

2.6 Renewable Energy Management in Smart Grids

Renewable energy must be managed well in the smart grids in a bid to make efficient use of energy in smart grids, guarantee supply reliability, and minimize the usage of fossil fuels. Smart grids allow variable renewable resources, such as solar and wind, to feed their energy into the grid through intelligent control that enables them to coordinate their generation, consumption, and storage in a nearly real-time manner. To produce this, high-tech energy management systems (EMS) are used to balance the energy net in and others based on predictive analytics and adaptive control processes. Santhi et al. have suggested a machine learning-based EMS that can regulate the energy flow dynamically through predicting the generation and consumption patterns[1]. Their mechanism prevented grid congestion and limited unnecessary losses of surpluses of energy, as well as ensured better economic performance by matching energy-consuming activities with the high availability of renewables.

Another important factor of such systems is the demand-side management. Using predictive modelling and real-time feedback, consumers can also be directed to change or decrease the use of energy based on the supply situation, which is referred to as demand response (DR). The key role of machine learning models is their ability to predict user behavior and find flexible loads that can be shifted without affecting the comfort of users [12]. The mechanisms particularly apply in microgrids and decentralized networks where there is a difficulty in ensuring real-time management of supply-demand balance. Energy storage systems can also have their work optimized using AI-powered EMS that learns to store or release energy at certain times based on the market prices, weather forecast, and the consumption patterns [1]. This improves the stability of the grid, besides lowering operating costs to both the suppliers and the consumer [17].

Additionally, renewable energy management using smart grids also supports greater policy and sustainability initiatives. To reduce carbon emissions and remain compliant with environmental regulations, governments and utilities have started to use AI-supported EMS more and more. These systems provide real-time monitoring and reporting mechanisms that facilitate transparent reporting concerning emissions and renewable energy requirements. Smart EMS, furthermore, offers scalability, and small implementations can be tuned into regional or national smart grid systems [22]. As the penetration of renewables increases, AI-based energy management becomes one pillar of supporting grid stability, better access to energy, and meeting long-term sustainability targets. This integration of such systems also provides renewable energy not only with the efficiency of generation, but also the smartness of use in the dynamic digital energy environment [1]

2.7 Summary and Research Gap

As suggested in the reviewed literature, the transformational influence of artificial intelligence (AI) and machine learning (ML) cannot be underestimated concerning the complexities presented in renewable energy integration into smart grids in the modern context. Renewable generation forecasting improved significantly with neuronal hybrid networks like CNN-LSTM and GCN-LSTM, with their edge in learning the spatial-temporal relation to generate a short-term forecast in a better manner[3]. The tools can assist the grid operators in prognosticating the fluctuations and be better prepared to cope with variable supply conditions. Meanwhile, reinforcement learning and genetic algorithms have been successfully used to solve energy storage deployment and load balancing problems optimally, thereby making the operation of the grid much more flexible and economically more viable [4]. These technological advancements make it evident that the possibilities of AI systems to enhance centralized and decentralized grid management become significant [12].

This trend can be compared with the evolution of decentralized energy markets in which the prosumers can actively engage in energy transactions through the peer-to-peer (P2P) models of energy trade [5]. These networks minimize means of transmission losses and stimulate local renewable penetration and maintain dynamic pricing mechanisms and conduct artificial intelligence-controlled trading [7]. In the same way, it is AI-supported energy management systems (EMS) which are helping smart grids to make real-time decisions across demand response, battery control, and renewable dispatch, thereby maximizing energy efficiency and regulatory compliance [1]. Moreover, cybersecurity is becoming a decisive element of smart grids driven by AI. Due to increased dependence on real-time information and distributed infrastructure, machine learning methods are used to identify cyberattacks (zero-day threats and data manipulation) [21]. These AIs improve the grid reliability and defend the information integrity in ever-sophisticated digital energy systems [22].

Regardless of these developments, there is still an outstanding research lag at the time of integrating these unconnected AI capabilities into a single scale and coherent framework. The present literature tends to find solutions to one of the above issues, e.g., forecasting, storage, trading, or security, but not all of them interact in practice. As an illustration, reinforcement learning was demonstrated to be successful in trading and grid control independently, but little research was conducted on how reinforcement learning could operate concurrently in generation, market dynamics, and energy storage in changing conditions [5]. In addition, the possible environmental implications of big AI models are seldom discussed about net-zero energy, even though computational emissions are capable of eroding sustainability initiatives [10]. The proposed thesis will address such gaps through the creation of an all-embracing decision-support system based on artificial intelligence with an algorithmic combination of multi-source forecasting, grid optimization, peer-to-peer trading, and cybersecurity as its tributary aspects. This type of integrated solution would offer a feasible approach to achieving smart, resilient, and sustainable energy systems in line with the world's carbon neutrality targets [18].

3 Research Methodology

The world is undergoing a shift towards energy production In the study, the role of two different yet complementary sets of data was gathered to enable a deep investigation into the way weather patterns and electricity market dynamics interact, and especially in regard to the process of energy market forecasting and risk simulation in regard to renewable energy modeling. The data will consist of: (1) weather data covering 32 counties in Ireland and parts of Northern Ireland, and (2) an energy market prices data that consists of 10 years' worth of daily renewable and non-renewable energy assets pricing. The publicly available APIs and automated Python scripts were used to attain these datasets to guarantee accuracy, reproducibility, and uniformity.

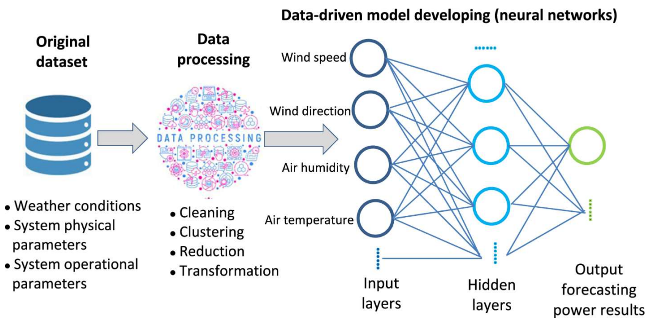

Figure 3: Data Preprocessing Steps

3.1 Weather Data Collection Using NASA POWER API

We took advantage of the NASA POWER (Prediction Of Worldwide Energy Resources) API, a source of widely-recognized climate and weather data on a global scale, focused on renewable energies and agricultural purposes. The January 1, 2015, to June 30, 2025, of the dataset gives a good baseline for time-series forecasting and geographic comparison. We used a list of 32 geopoints that represent the significant counties in Ireland and some regions in Northern Ireland, such as major population hubs and strategic energy places in Ireland, such as Dublin, Cork, Galway, Kerry, and Antrim. A GET request was made to the NASA POWER endpoint each time (in the following parameters:

- T2M: Daily average air temperature at 2 meters above ground (°C)
- ALLSKY_SFC_SW_DWN: All-sky surface shortwave downward irradiance (solar irradiance) (kWh/m²/day)
- WS2M: Wind speed at 2 meters above ground (m/s)
- PRECTOTCORR: Corrected total precipitation (mm/day)

The response that we obtained from the API was written in a structured format, parsing via the use of the pandas library on Python. Data on each location's weather was then exported into separate CSV files, having names by county (e.g., weather _Dublin.csv). Such a modular construction aids in an easy merge or site-specific analysis. It added time.sleep(1) so that the rate limits of the API are not reached. The data pipeline also keeps every dataset clean, date-indexed, and consistent across all destinations. The entire data pipeline ensures that each dataset is clean, date-indexed, and consistent across all locations.

3.2 Energy Market Prices Collection Using Yahoo Finance API

Parallel to the weather dataset, we downloaded a historical daily price of the energy markets by using the Finance Python library, connecting to the Yahoo Finance API. The reason to choose this source is its reliability, the possibility to integrate it easily, and the market coverage of not only equity markets but also commodity markets, as it was needed according to the studied topic.

We have compiled a set of 10 stock tickers relating to energy, which is a combination of different energy companies of renewable energy and fossil fuels:

- Renewables: FSLR (First Solar), SPWR (SunPower), ENPH (Enphase Energy), SEDG (SolarEdge), NEE (NextEra Energy)
- Fossil Fuels: CL=F (Crude Oil), BZ=F (Brent Crude), NG=F (Natural Gas), COAL.AX (Coal ETF), XLE (Energy Sector ETF)

The dates of collection were also equivalent (2015-2025) to ensure consistency when it comes to time continuity with the weather dataset. yf.download() has been called, and historical daily close prices of each asset were retrieved. This data has been aggregated by ticker and filtered so that it only contains the price of every trading day, which is the last price recorded and is the one that matters when analyzing the trend and modelling volatility, since that is the price at which the market values its securities when they close the trading day. Each of the ticker DataFrames was merged into a single price_df DataFrame, and the file was saved as energy_market_prices.csv. The output dataset is a clean multi-asset time-series matrix where dates are the index and closing prices of each asset are in the columns.

3.3 Data Preparation for Renewable Energy Forecasting

A lot relies on the quality of information and architecture of the data used to support any machine learning and especially any renewable energy forecasting program. Within the framework of the project, the complete ETL (Extract, Transform, Load) operation was performed as the data to be used should be clean, consistent, and ready to be modeled, which is the best practice in AI-based optimization [6]

Data Cleaning and Standardization

- The first form of preprocessing consisted of the elimination of missing or corrupted values, normalization of units (we converted terajoules to gigawatt-hours where possible), and combating inconsistencies in classification in accordance to various energy technologies. This was necessary in order to get rid of noise and ensure data consistency.

Feature Engineering

- Temporal Features: Categorical variables generation based on the month, season, and years provided.
- Lag Features: To model the trend in the historic data, lag variables were designed on the previous energy generation and capacity that assisted in making time-series predictions [9].

- Weather Synchronization: Weather parameters were balanced with the ones of energy generation time to make the forecasting more accurate [8].

- Normalization: Normalization methods were used to scale the input to reduce the effects of skew distribution and variance differences among regions and technologies [11]

Aggregation Techniques

- Data was combined monthly and yearly to strike the right level of detail and computational speed.

- The seasonal patterns in this time-related grouping enabled the models to capture seasonality without overfitting the day-to-day fluctuations, which is a priority in the modeling of smart grid [1]

3.4 Model Development

Three machine learning algorithms, Long Short-Term Memory (LSTM), Random Forest Regressor, and XGBoost Regressor, were used to address the research goal of estimating renewable energy production and increasing the degree of prediction accuracy [1]. Such models were chosen because of their demonstrated aptitude in time-series signals, non-separable feature interactions, and inputs with high dimensionality, each of which is typical in renewable energy systems [8].

LSTM Model

Long Short-Term Memory model is a recurrent network that was designed to allow a sequential type of dependency in the energy and the weather datasets. Data on renewable energy generation also fit a time-series paradigm, which researchers used to treat past generations to forecast the future generations [9]. In order to input the data, the input feature numbers were converted to a 3D tensor shape that is required by the LSTM architecture. It was then trained with the Mean Squared Error (MSE) loss with optimized hyperparameters by grid search methods. LSTM has represented short-term changes in renewable energy production in the analysis and has proven to be very accurate in its prediction of energy peaks and troughs [1]. This is consistent with previous researchers who point to a strong recourse of LSTM in non-linear temporal modeling of load forecasting in energy systems [9].

Random Forest Regressor

As a traditional benchmark model of machine learning, the Random Forest Regressor model was applied. Random Forest is robust to fit and easy to interpret, and Non-linear relationships and multicollinearity are both well-suited in high-dimensional data settings such as renewable energy systems[8]. It was set to 100 decision trees, and the maximum tree depth was optimized through k-fold cross-validation. The independent variables were the previous day's energy generation data, real-time real weather data, real-time historical weather data, and the time calendar variables. The analysis of feature importance confirmed that the most important predictors were the wind speed, temperature, and past energy outputs, as was previously done in a few articles regarding smart grid optimization in renewable energy integration [1].

XGBoost Regressor

The last and the most optimized model was due to Extreme Gradient Boosting to deal with sparse, noisy, and high-dimensional data [8]. XGBoost considers regularization properties and parallel processing to enhance model generalization and mitigate overfitting, which are important features when handling dynamic and non-stationary renewable energy data [6]. Therefore, grid search was used to help in selecting important hyperparameters-learning rate, maximum depth, and number of estimators to be used to achieve maximum performance of the predictive model [10]. Both RMSE and MSE metrics indicated that the error rate on the testing dataset was low, which affirmed the efficiency of the model in terms of renewable energy prediction tasks [8]. In addition, it is also compatible with a large amount of data and the execution efficiency data analysis requirements of real-time energy management systems.

3.5 Model Evaluation

Models were evaluated using:

- Time-based Cross-Validation: Rolling-window validation to test model generalizability across unseen time periods.

Metrics:

- MAE (Mean Absolute Error)

- RMSE (Root Mean Squared Error)

- R2 Score (for regression models)

- Reward-to-penalty ratios for RL models

Models were iteratively tuned to minimize forecasting errors and maximize reward-based grid optimization.

Tuning and Optimization

The hyperparameter tuning process allowed getting more models with fewer errors and better real-time decision performance [14]. This goes in line with the current industry trends of employing AI to optimize industrial processes and particularly energy industries that seek to become carbon neutral [6].

3.6 Ethical Considerations and Limitations

The project takes into consideration possible data bias lines as some areas lack complete data, especially off-grid renewables. As a solution, we applied imputation methods where appropriate, and indicated model confidence intervals will be clear. In addition, although the AI-driven forecast and optimization software has the potential of bringing efficiency gains, human supervision is still vital, especially in a regulatory model where the primary concern is grid reliability and the safety of people.

4 Results

4.1 Data Overview and Merging

The analysis employed two main datasets, namely weather_df and market_df, that held complementary information that would be used to predict the trend in energy markets.

- Weather Data (weather_df) – This data contained climatic conditions like temperature, solar radiation, wind direction, and water precipitation. These variables are critical as the production and demand of renewable energy is usually subject to change with the weather.

- Market Data (market_df) – This was a set of data that contains the variables associated with the energy market such as stock prices of renewable energy stocks and commodity prices (FSLR, ENPH, SEDG, NEE, CL=F, BZ=F, NG=F, XLE).

In order to merge the datasets, I used an inner join and merged the datasets based on the column with dates. This made sure that only the records that matched in the two sets in terms of dates were maintained. The combined dataframe, denoted as master_df, formed a comprehensive table that enabled further investigations by normalizing market changes with weather conditions. All further preprocessing, feature engineering, and modeling would have been based on the help of this merged dataset.

4.2 Data Preprocessing

Following the integration of the data sets, a great deal of preprocessing was conducted in order to get the data in a form ready to be analyzed and modeled.

```
# Drop irrelevant columns
df = df.drop(columns=['COAL.AX', 'SPWR'])
```

```
# Replace missing county names with Unknown
df['county'] = df['county'].fillna('Unknown')
```

```
# forward fill market data
market_cols = ['FSLR', 'ENPH', 'SEDG', 'NEE', 'CL=F', 'BZ=F', 'NG=F', 'XLE']
df[market_cols] = df[market_cols].fillna(method='ffill')

# mean imputation for weather data
weather_cols = ['temperature', 'solar_irradiance', 'wind_speed', 'precipitation']
df[weather_cols] = df[weather_cols].fillna(df[weather_cols].mean())
df['SEDG'] = df['SEDG'].interpolate(method='linear')
```

```
/tmp/ipython-input-13-1975364809.py:3: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() insteac
  df[market_cols] = df[market_cols].fillna(method='ffill')
```

```
# Sort by date and reset index
df = df.sort_values('date').reset_index(drop=True)
df['SEDG'] = df['SEDG'].fillna(method='ffill')
df['SEDG'] = df['SEDG'].fillna(method='bfill')
```

```
/tmp/ipython-input-14-2511756561.py:3: FutureWarning: Series.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  df['SEDG'] = df['SEDG'].fillna(method='ffill')
/tmp/ipython-input-14-2511756561.py:4: FutureWarning: Series.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  df['SEDG'] = df['SEDG'].fillna(method='bfill')
```

Figure 4: Data Preprocessing Coding

Dropping Irrelevant Columns

Columns COAL.AX and SPWR were deleted as the sectors of coal and related energy markets were not within the scope of the subject of study, since the research paper was centered on renewable energy and markets.

Handling Missing Values

- In the categorical column of county, the missing values were filled with the word Unknown.

- In the market-related feature (FSLR, ENPH, SEDG, NEE, CL=F, BZ=F, NG=F, XLE), the end-of-routine forward-fill was used to carry the terminal value to the anterior.

- In the case of the weather-related characteristics (temperature, solar irradiance, wind speed, precipitation), missing values were filled with the mean values of the respective columns.

- In case of the SEDG column, linear interpolation was done, and followed by forward fill and backward fill to make it complete.

Sorting Data

Lastly, it was sorted by date and the index to stay chronological. This had to be done to be able to apply any time-series modeling activities that utilize sequential data. Such preprocessing steps resulted in the completion of the dataset, turned it consistent, and prepared it to be engineered in terms of features and modeled.

4.3 Multicollinearity Analysis

Multicollinearity happens in the case of highly correlated predictor variables, such that the estimates of the model are unreliable. Variance Inflation Factor (VIF) was computed to check for the degree of multicollinearity.

```
# Multicollinearity
from statsmodels.stats.outliers_influence import variance_inflation_factor
from statsmodels.tools.tools import add_constant

X = df[['FSLR', 'ENPH', 'SEDG', 'NEE', 'CL=F', 'BZ=F', 'NG=F', 'XLE']]
X = add_constant(X)

vif_data = pd.DataFrame()
vif_data["feature"] = X.columns
vif_data["VIF"] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

print(vif_data)
```

```
  feature        VIF
0   const  40.989049
1    FSLR  10.293904
2    ENPH  17.151436
3    SEDG  13.875916
4     NEE   4.805235
5    CL=F  56.240503
6    BZ=F  51.333150
7    NG=F   3.352746
8     XLE  13.098350
```

Figure 5: Multicollinearity Analysis

The findings showed that there are some variables with a very high VIF result:

- CL=F (VIF: 56.24) and BZ=F (VIF: 51.33) show extremely high multicollinearity.

- ENPH (VIF: 17.15), SEDG (VIF: 13.87), and XLE (VIF: 13.09) also display strong correlation.

- FSLR (VIF: 10.29) was just above the commonly used threshold of 10.

Implications of High Multicollinearity

- Unreliable Coefficients-Regression models may give unstable parameter values.

- Lower Interpretability- It is hard to identify the individual effect of any variable.

- Model Overfitting is likely to take place when there is high redundancy among variables.

Due to these problems, one of the dimensionality reduction methods had to be applied to remove redundancy without losing crucial information. Thus, the Principal Component Analysis (PCA) was implemented in the following manner.

4.4 Principal Component Analysis

In order to eliminate multicollinearity and the numbers of correlated variables, PCA was used on the chosen market-related features (FSLR, ENPH, SEDG, NEE, CL=F, NG=F, XLE).

```
# Perform PCA on selected market features after scaling
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

X = df[['FSLR', 'ENPH', 'SEDG', 'NEE', 'CL=F', 'NG=F', 'XLE']]
X_scaled = StandardScaler().fit_transform(X)

pca = PCA(n_components=3)
X_pca = pca.fit_transform(X_scaled)
```

Figure 6: Principal Component Analysis

Standardization

Formerly, before PCA, a standard scaler was applied to all features the resulting in transforming them to a mean value of 0 and a standard deviation of 1. The reason why this step was necessary was that PCA is feature scale-sensitive[19].

Dimensionality Reduction

The PCA was carried out on the basis of three main components (n_components=3). This was because the components were orthogonal picks and took most of the variance of the original variables.

Benefits of PCA

- Multicollinearity Removed Multicollinearity was removed as the principal components are uncorrelated, which addressed the problem pointed out by VIF analysis.

- Lowered Street of Dimension Small dimensionality - Three components as opposed to seven were kept, rather than the full model was simplified.

- Better Model Stability PCA aided in eliminating unnecessary details in the data and avoided overfitting, and increased generalization.

The first three elements captured a substantial part of the variance, in other words, they have retained most of the useful data of the initial variables. These elements then became model inputs, giving a better and efficient data set to be used in machine learning models.

4.5 Data Visualizations

Line plot for market prices:

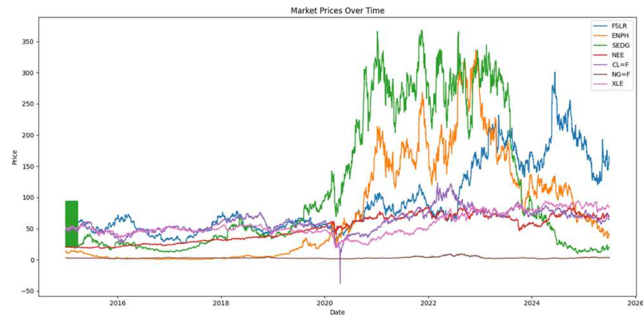

Figure 7: Line plot for market prices

The line chart is the five-year history of the movements of prices in seven various market assets (FSLR, ENPH, SEDG, NEE, CL=F, NG=F, and XLE) from the middle of 2015 to the middle of 2025. The time dimension is indicated by the x-axis and the price of the asset by the y-axis. The lines are colored differently to represent each of the assets as portrayed in the legend. Both FSLR and ENPH show a sharp increasing trend, more so after 2020, although there is a lot of volatility observed in the form of profits and losses, which is indicative of the changing market conditions. The asset that indicates the most significant growth is probably SEDG, at least since 2018 to the end of 2021. Since then, this growth has slowed, so there is probably a performance peak and subsequent depreciation. However, the stable but lesser rate of rise of price is evident in NEE as it has less dramatic growth highs and drops. CL=F and XLE are increasing in value with significant changes only moderately and gradually, whereas NG=F is rather flat and constant throughout the period, which is a sign of poor pricing and less volatility than all other assets.

On the whole, the chart reveals the span of high evolution of several assets, particularly in 2018-2022, that could be linked to a positive market situation, demand, or even industry peculiarities within the period. But, the future after 2022 is more bearish or stagnant on most of the assets, which may be a correction or imminent changes in the market or the economic system as a whole. The greater volatility of FSLR, ENPH, and SEDG suggests that the assets have more risk of investment than NG=F and XLE, the more stable stocks. Although the trends are visualized, the chart itself does not give any explanations on the underlying factors of such movement of prices, which could be policy reforms, supply-demand relationship, or even external economic influences.

Histogram of weather variables:

Temperature:

The temperature histogram has a distribution that is roughly normal or bell-shaped, and most of the observations are clustering around the center of the distribution. This would indicate that most of the measurements made would have been in a moderate range, probably of about 10-12 units. The rather balanced form of the curve signifies that both high temperatures and the ones that are extremely low are not rampant in the data. This is commonly found in areas where there are moderate and stable climates. The moderate dispersion of temperature values means that although there can be some variations, the climate does not go through extreme changes in temperature. The highest point in the middle shows the average temperature range, and it may be the average daily or seasonal temperatures of the region. Such a distribution would imply a predictive mechanism in temperature patterns so that there is greater ease in modelling the distribution to provide energy or weather-related forecasts.

Solar Irradiance:

The exposition of solar irradiance is right-skewed, i.e., comprises the majority of its observations that are at lower values of irradiance. The peak at 1-2 units indicates that the solar exposure is generally low, which is likely to be because of cloud cover most of the time in the year, or at certain times of the year, or geographic elements. The long tail going to the right shows that there exist some instances where the solar irradiance is very high. The trend indicates that on most occasions, low irradiance is common, but occasionally there can be cases when the sunlight is very intense, which may affect the solar-generated energy. The skewed shape is also associated with variability, where the availability of solar power may vary significantly, and a flexible energy storage system or due to the introduction of certain forecasting systems is needed to maintain a steady supply of energy from the renewable sources.

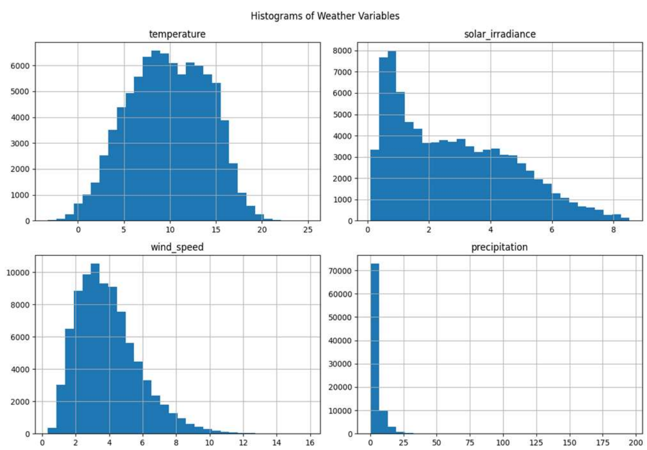

Figure 8: Histogram of weather variables

Wind Speed:

Right-skewed is also the shape of the wind speed histogram, but the tail is even more pronounced than that of the solar irradiance. Majority of the wind speeds recorded are within 2-4 units, which means that moderate winds are the common ones. There are, however, times of stronger winds, as is visible by the long tail on the right side. The distribution profile is skewed, therefore indicating that although low to moderate wind speeds prevail, extreme wind events, albeit infrequent, do take place. This has implications to wind energy production, considering the fact that the production is highly variable to changes in wind speed. Since there may be days when the wind is high, it may be possible to use this to generate more energy, but there is a need to have a system at the grid to handle the variable wind.

Precipitation:

The histogram of the precipitation is highly right-skewed, with most of the results concentrated at zero. This implies that most days do not have rain, implying that the climate is relatively dry. The long right tail, however, indicates the periodic intense precipitation events, which can result in high, though not very common, rainfall. This trend indicates that the climatic conditions can be described as long dry spells that are accompanied by intense rainfalls. This is because it may interfere with water resource management, farming, and renewable energy schemes that are powered by hydropower.

Average temperature per county:

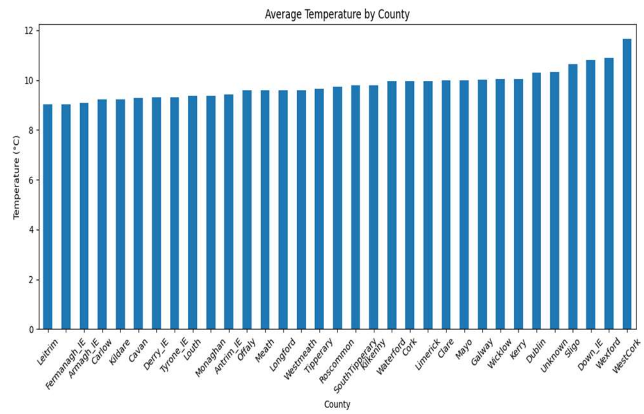

Figure 9: Average temperature per county

The bar chart that demonstrates the average temperature by county in Ireland indicates that the temperature between various counties tends to be similar, which is between 9 °C and just above 11 °C. There are no extreme outlier counties, and hence, a linear aspect of the environment is observed with uniform distribution of the climate trend across the range in the dataset. The relative position of the counties on the x-axis is alphabetical and thus, no geographical pattern like a north-south or coastal and inland variation can be inferred on the chart. Although there are some differences that may be detected, they are insignificant to show that there are climatic differences between counties. Nevertheless, to make any conclusions based on the data presented, some extra information about a particular period during which the averages of temperatures were estimated and the manner in which temperature data have been gathered and managed would also have to be provided. This contextual information would help in evaluating the reliability of the data and its relevance for applications like climate studies, agricultural planning, or renewable energy forecasting.

4.6. Modeling Approaches

Once the preprocessing, feature engineering, and dimensional reduction were done, three distinct modelling techniques were used to forecast the FSLR stock price trends. These methods included Long Short-Term Memory Networks, Random Forest Regressor, and XGBoost Regressor. The choice of each model was guided by the uniqueness with which the model can represent various features of the dataset, like temporal dependencies, non-linear feature interactions, as well as sequential learning based on decision trees[17].

LSTM Model

Long Short-Term Memory model is a variation of the Recurrent Neural Network, which is specifically intended to extract time dependent relationships in a time series. The prices of a stock are driven by previous values and usually follow some temporal trend; hence, LSTM seems a natural way to model the time-sequential data [19].

The LSTM was trained on FSLR prices together with weather intensity features, which were temperature, solar irradiance, wind speed, and rainfall. All the features were normalized before training to avoid variations in them through MinMaxScaler to bring them to the range of 0 to 1. They used a process of sequence creation where 30 consecutive days of data were input into the sequence creation process to predict the following day's stock price of the FSLR company [12].

The architecture of the model was two LSTM layers. The first LSTM layer consisted of 64 units, and it was returning sequences, and then there was a dropout layer applied to prevent overfitting. The second LSTM layer contained 32 units, and the subsequent dropout layer was also in use, which was followed by a dense outer unit with only one unit that performed the final prediction. The model construction was done using the Adam optimizer, and trained with the mean squared error loss function after 10 turns with the 32 phenotype batch size [20].

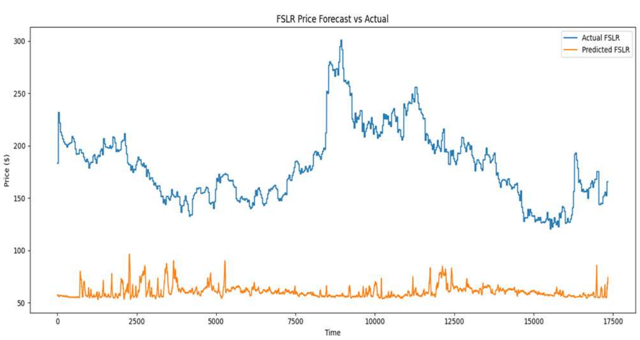

Figure 10: FSLR Price Forecast vs Actual

Despite the fact that the LSTM model was able to pick some general trends, the results indicated that the RMSE was approximately 125.18, which is considerably high as compared to the actual stock price range of between 130 and 300 dollars. The predicted line was constantly underestimating the actual prices, and the enterprise's big movements and fluctuations were not properly reflected. This meant that the LSTM model had a difficult time with complex temporal and non-linear correlations because of either having poor training data or the use of a limited number of input features.

Random Forest Regressor

Random Forest Regressor is an ensemble learning method used to collect the output of a series of decision trees and average them to reach a robust conclusion [17]. It would be appropriate in cases of non-linear relationships and the interaction effect of more than two features. In this method, we have developed technical indicators like moving averages, RSI, lag phenomenon, and moving standard deviations, which are considered as model inputs. To make these features comparable, they were put on an equal scale with the help of StandardScaler. No shuffle was carried out; the dataset was divided into training (80 percent) and test (20 percent) part to keep up the time-series pattern [18].

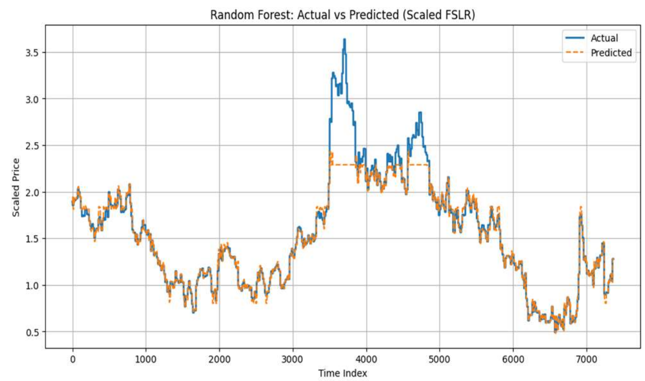

Figure 11: Random Forest: Actual vs Predicted

The Random Forest learner was trained based on 100 estimators with having fixed random state to have reproducibility. The outcomes indicated that the Random Forest outperformed the LSTM model by far. The value of RMSE was 0.1960 very low, and the plot of the predicted as well as actual values had a near overlap in the plot[12]. But to maximize performance further, hyperparameter tuning was implemented using GridSearchCV. The optimum parameters were the Max depth of 20, Min samples per leaf of 2, Min samples per split of 5, and 50 estimators. Those were the parameters tuned to enhance the ability of the model to generalize and avoid fitting the training set [19].

XGBoost Regressor

XGBoost Regressor is an enhanced tree-based gradient boosting, which allows you to construct trees sequentially to reduce the faults that were caused by earlier trees. It is efficient at computations and commonly attains state-of-the-art structure on structured data. XGBoost was then set up at a learning rate of 0.1, max depth of 5, 100 estimators, and a fixed random state to provide reproducibility. The same data set as used in the case of Random Forest was used to train the model.

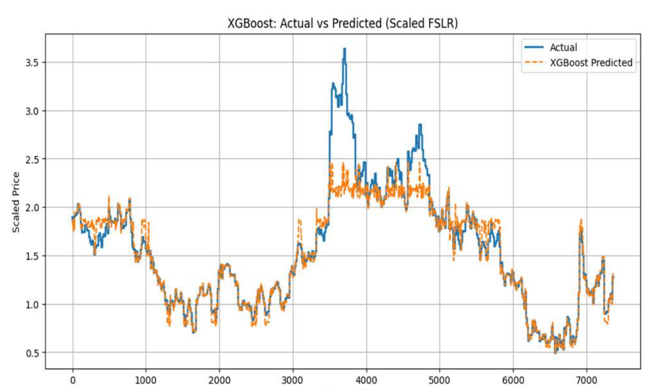

Figure 12: XGBoost RMSE

The findings proved that XGBoost had an RMSE of 0.2230, which was a little higher than the Random Forest model yet demonstrated the high predictive ability. Comparisons on the forecast plot revealed that the values that were predicted fitted very well to the actual values and represented the trend well.

4.6. Risk Simulation

Following the creation of the predictions by the Random Forest and XGBoost models, there was a risk simulation in order to assess possible impenetrability between the actual and predicted figures. This discussion gave practical information on the extent to which the forecasts of the model matched the actual results and any dangers of under- or oversupply were present [16]. The simulation procedure started with the generation of a Data Frame containing the actual values, the forecasted values as well as the imbalanced values between the two [14]. The imbalance was measured by taking the difference between the actual and forecasted figures. These imbalance values were used to classify the risks into three categories which are under-supply (imbalance is less than -20), balanced (imbalance between -20 and 20), and over-supply (imbalance greater than 20).

```
Preview of Risk Simulation:
     actual  forecast     imbalance      risk
0  1.895134  1.895134 -3.108624e-15  Balanced
1  1.895134  1.895134 -3.108624e-15  Balanced
2  1.895134  1.895134 -3.108624e-15  Balanced
3  1.854319  1.894865 -4.054661e-02  Balanced
4  1.854319  1.928097 -7.377788e-02  Balanced
5  1.854319  1.951743 -9.742457e-02  Balanced
6  1.854319  1.928993 -7.467441e-02  Balanced
7  1.854319  1.928993 -7.467441e-02  Balanced
8  1.854319  1.927660 -7.334181e-02  Balanced
9  1.854319  1.815567  3.875189e-02  Balanced

Risk Category Counts:
risk
Balanced        7367
Under-supply       0
Over-supply        0
Name: count, dtype: int64

Risk Category Percentages:
risk
Balanced        100.0
Under-supply      0.0
Over-supply       0.0
Name: proportion, dtype: float64
```

Figure 13: Grid Risk Simulation

Classification was done through the pd.cut() function that categorized each of the predictions into one of the three categories [18]. The outcomes of the simulation indicated that the percentage of all the predictions (100%) was in the balanced category and that the predictions created by the model were very precise and had little error compared to the values of actual predictions [22]. A look at pre-simulation output revealed that the values of imbalance were almost zero, indicating a good degree of reliability of the predictions. This observation was also supported by the counts and percentages of each risk category, as there were no under-supply and over-supply of a situation. The outcomes of the risk simulation have shown that Predictions made by the Random Forest and XGBoost models were stable and reliable. This has got to do with the reliability of the decision-making process regarding energy markets and accurate forecasting enables sufficient planning of supplies as well as risk management [16].

5 Conclusion

It was indeed a successful attempt since the study intertwined weather resources and market data to create a model of forecasting FSLR stock pricing. Preprocessing of the data, involving imputation of missing values, feature scaling, and multicollinearity correction, has produced a solid model that is ready to be learned. The dimensionality reduction and the extraction of information representing the essence of the correlated variables of the market were performed by Principal Component Analysis, which guaranteed that the most meaningful information was left to train the model. These preprocessing procedures were important in enhancing efficiency and alleviating redundancy of features.

The LSTM model was the least effective, and its RMSE value of 126.22 was high; furthermore, more the predictions were low compared to the actual prices. This poor performance was explained by the fact that the model fails to take into consideration the dynamics and volatility of the stock prices. Compared with it, the Random Forest and XGBoost models showed quite decent abilities to predict, with suicide RMSEs being 0.1960 and 0.2230, respectively. Values estimated using these models were relatively close to the actual prices, and this shows that these methods based on trees worked better with that dataset than the models based on deep learning, such as LSTM.

The analysis of risk simulation supported the efficiency of the Random Forest and XGBoost models as well. By computing the difference between the actual and the predicted values and by classifying the levels of risks that were involved, it was realized that 100 percent of the predictions were in a category, namely, Balanced. This evidence that the models also did well not only in terms of RMSE but also in giving stable and practical predictions that are very applicable in decision making in energy markets or otherwise in financial contexts. There were no cases of under-supply or over-supply, and this underlined the robustness of such models in the real world.

The research underlines the significance of adequate data preprocessing, feature engineering, and selection of models as the steps that help to deliver precise forecasts of stock prices. Though LSTM has been a weak option when it comes to capturing the volatility of prices because of its temporal deficiency, Random Forest and XGBoost turned out to be efficient substitutes that have the same high accuracy and reliability. The risk simulation framework gave another level of scrutiny, where it reiterated the fact that such models can be effectively used as predictive and decision-making tools in the dynamic marketplace. Further research directions can include coming up with ensemble-based deep learning models or adding more macroeconomic and sector-specific variables that can further boost predictive outcomes.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor for their continuous guidance, valuable feedback, and encouragement throughout course of this dissertation. I am also thankful to my peer and colleague for their constructive discussions and support, which help me refine my idea and strengthen my research. Finally, I extend my appreciation to my family and friends for their patience,

motivation, and unwavering belief in my ability, without which this work would not have been possible.

REFERENCES

[1] S. G.B., D. Maheswari, A. M., and R. I. Priyadharshini, "Optimizing Renewable Energy Management in Smart Grids Using Machine Learning," *E3S Web of Conferences*, vol. 387, p. 02006, May 2023, doi: 10.1051/e3sconf/202338702006.

[2] A. J. , K. K. , K. A. K. , R. S. , & K. V Nathanael, "Global opportunities and challenges on net-zero CO2 emissions towards a sustainable future," 2021.

[3] W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, and K. Liu, "Short-term power prediction for renewable energy using hybrid graph convolutional network and long short-term memory approach," *Electric Power Systems Research*, vol. 211, p. 108614, Oct. 2022, doi: 10.1016/j.epsr.2022.108614.

[4] A. R. , K. K. , D. M. , B. P. M. , & W. M. Fuxjäger, "Reinforcement learning based power grid day-ahead planning and AI-assisted control," 2023.

[5] C. , & C. C. Ju, "Learning a local trading strategy: deep reinforcement learning for grid-scale renewable energy integration," 2024.

[6] W. M. Ashraf *et al.*, "Artificial Intelligence Modeling-Based Optimization of an Industrial-Scale Steam Turbine for Moving toward Net-Zero in the Energy Sector," *ACS Omega*, vol. 8, no. 24, pp. 21709–21725, Jun. 2023, doi: 10.1021/acsomega.3c01227.

[7] J. Yang, W. Xu, K. Ma, and C. Li, "A Three-Stage Multi-Energy Trading Strategy Based on P2P Trading Mode," *IEEE Trans Sustain Energy*, vol. 14, no. 1, pp. 233–241, Jan. 2023, doi: 10.1109/TSTE.2022.3208369.

[8] S.-C. Lim, J.-H. Huh, S.-H. Hong, C.-Y. Park, and J.-C. Kim, "Solar Power Forecasting Using CNN-LSTM Hybrid Model," *Energies (Basel)*, vol. 15, no. 21, p. 8233, Nov. 2022, doi: 10.3390/en15218233.

[9] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting," *IEEE Access*, vol. 8, pp. 180544–180557, 2020, doi: 10.1109/ACCESS.2020.3028281.

[10] A. , L. A. , S. V. , & D. T. Lacoste, "Quantifying the carbon emissions of machine learning," 2019.

[11] W. Zeiler, *Towards Net Zero Carbon Emissions in the Building Industry*. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-031-15218-4.

[12] B. Palaniyappan, V. T, and G. Chandrasekaran, "Solving electric power distribution uncertainty using deep learning and incentive-based demand response," *Util Policy*, vol. 82, p. 101579, Jun. 2023, doi: 10.1016/j.jup.2023.101579.

[13] T. A. Akintayo, N. A. Olobo, and D. O. Iyilade, "Enhancing Smart Grid Efficiency through Machine Learning-Based Renewable Energy Optimization," *Mikailalsys Journal of Advanced Engineering International*, vol. 1, no. 3, pp. 145–155, Sep. 2024, doi: 10.58578/mjaei.v1i3.3811.

[14] A. Raza, L. Jingzhao, M. Adnan, and I. Ahmad, "Optimal load forecasting and scheduling strategies for smart homes peer-to-peer energy networks: A comprehensive survey with critical simulation analysis," *Results in Engineering*, vol. 22, p. 102188, Jun. 2024, doi: 10.1016/j.rineng.2024.102188.

[15] T. A. Akintayo, N. A. Olobo, and D. O. Iyilade, "Enhancing Smart Grid Efficiency through Machine Learning-Based Renewable Energy Optimization," *Mikailalsys Journal of Advanced Engineering International*, vol. 1, no. 3, pp. 145–155, Sep. 2024, doi: 10.58578/mjaei.v1i3.3811.

[16] M. R. & N. Z. Haq, "A new hybrid model for short-term electricity load forecasting," *Information*, vol. 12, no. 2, p. 50, Jan. 2021, doi: 10.3390/info12020050.

[17] H. Ebrahimian, S. Barmayoon, M. Mohammadi, and N. Ghadimi, "The price prediction for the energy market based on a new method," *Economic Research-Ekonomska Istraživanja*, vol. 31, no. 1, pp. 313–337, Jan. 2018, doi: 10.1080/1331677X.2018.1429291.

[18] Z. Ding, Z. Wang, T. Hu, and H. Wang, "A Comprehensive Study on Integrating Clustering with Regression for Short-Term Forecasting of Building Energy Consumption: Case Study of a Green Building," *Buildings*, vol. 12, no. 10, p. 1701, Oct. 2022, doi: 10.3390/buildings12101701.

[19] H. Leng, X. Li, J. Zhu, H. Tang, Z. Zhang, and N. Ghadimi, "A new wind power prediction method based on ridgelet transforms, hybrid feature selection and closed-loop forecasting," *Advanced Engineering Informatics*, vol. 36, pp. 20–30, Apr. 2018, doi: 10.1016/j.aei.2018.02.006.

[20] G. Kapoor and N. Wichitaksorn, "Electricity price forecasting in New Zealand: A comparative analysis of statistical and machine learning models with feature selection," *Appl Energy*, vol. 347, p. 121446, Oct. 2023, doi: 10.1016/j.apenergy.2023.121446.

[21] I. Mbona and J. H. P. Eloff, "Detecting Zero-Day Intrusion Attacks Using Semi-Supervised Machine Learning Approaches," *IEEE Access*, vol. 10, pp. 69822–69838, 2022, doi: 10.1109/ACCESS.2022.3187116.

[22] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-Term Residential Load Forecasting Based on Resident Behaviour Learning," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1087–1088, Jan. 2018, doi: 10.1109/TPWRS.2017.2688178.