# **Energy Market Time-Series Forecasting**

Amrutha Gabbita, Ridvan Kanca, Sahana Reddy, Alexander Scarcelli, Matthew Lanham

Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

agabbita@purdue.edu; rkanca@purdue.edu; [reddy47@purdue.edu](mailto:reddy47@purdue.edu); ascarcel@purdue.edu;

lanhamm@purdue.edu

## **ABSTRACT**

The objective of this study is to identify and demonstrate the most optimal methods for forecasting energy market behavior and prototype algorithms using energy grid data in relation to various zones in the state of New York, United States of America. This study has potential benefits such cost reduction and optimal resource allocation for business. We conducted a comprehensive analysis with the data from different regions of New York collected by New York Independent System Operator (NYISO) during the years 2019 and 2020. Exploratory data analysis was done to examine the features of the dataset alongside with feature engineering which allows us to test the impact of exogenous variables such as temperature, weather conditions, and historical pricing data on energy prices. Both traditional statistical models including ARMA and SARIMAX and advance machine learning algorithms such as XGBoost and neural networks including RNN, LSTM, and Transformers were tested. Moreover, we used ensemble techniques to combine predictions from multiple models to enhance accuracy and robustness. We evaluated the models based on Mean Absolute Percentage Error (MAPE). The results from our experiments indicated that STL and ARIMAX ensemble model emerged as the most effective model, with a significant balance between accuracy and short processing time.

**Keywords:** *Forecasting, Energy, time series, machine learning*

## **INTRODUCTION**

Effectively forecasting the energy grid is a pressing issue in the global industry today. Because it will improve the efficiency of the power grid and minimize the effects of negative factors and costs if we can balance the demand and load during peak hours (Balouch et. al, 2022). Quantifying the energy grid load by geography is challenging because electrons are not tracked directly from the generator to the customer connection. Therefore, the grid is similar to an undirected graph leading to the black box effect. This black box effect makes accurately forecasting grid behavior such as swings in demand, load, and surplus, challenging, especially when it is desirable to forecast realistic behavior in a focused region, for example, by zip code.

Since the beginning of the 1990s research has been done on the demand side management of energy such as local control, direct control, and distributed control. Moreover, studies on forecasting energy demand have long been one of the most popular research areas in energy management (Balouch et. al, 2022). Traditional prediction techniques have limited advantages hence machine learning algorithms have been started to be used often because of their ability to learn from existing data and reveal neat forecasts (Benti et al., 2023). However, we cannot say that machine learning-based prediction methods are perfect without any drawbacks. Existing research mentioned that machine learning models also have some limitations that can affect their performances. For example, limited data accessibility can circumscribe the implementation of advanced machine-learning models in the energy field (Yao et al.,2023). Therefore, the objective of this study is to identify and demonstrate the most optimal methods for forecasting energy market behavior and prototype algorithms using energy grid data in relation to various zones in the state of New York, United States of America. According to existing literature it improved management of energy demand by enhancing energy efficiency and flexibility brings several benefits to the electric grid, such as lowered power generation capacity, reduced maintenance and operational cost, more reliable services with lower costs, and equipment and voltage control (Langevin et al., 2021). Indeed, several business benefits come to fruition as this objective is met. Two key benefits are cost reduction and optimized resource planning. By improving the ability to accurately forecast energy market behavior, it will be easier to eliminate unnecessary processes and resources in areas of low demand. Likewise, it will be easier to identify bottlenecks and identify the areas which need additional support. In addition, resources can be reallocated from region of low demand to regions of high demand to better handle the entire energy grid load.

Our analytical approach consists of the following methods: leverage opensource algorithms such (S)ARIMA(X), XGBoost, Meta’s Prophet, ETS, and relevant neural network architectures like RNN, LSTM, or Transformers to simulate and predict energy. After considering several models, the model yielding the most accurate results upon validation and having a reasonable run time will be proposed. It is important to use a model that predicts accurately but can also run quickly and efficiently so that the industry will be able to use the algorithm in real-time to generate forecasts promptly. Thus, the relevant information business decisions.

The remainder of this paper is organized into the following sections. Section 2 is a literature review on various criteria and methods used for energy grid forecasting. Section 3 presents the proposed methodology and a discussion of criteria formulation. Various models and algorithms are formulated and tested in Section 4. Section 5 demonstrates the performance of our optimal models. The paper concludes with Section 6 which is includes findings and implications as well as future research proposals.

## **LITERATURE REVIEW**

Considering the massive amounts of data available surrounding the electricity market and its attributes, achieving precise price forecasting often demands extensive research and data collection. Moreover, bearing in mind the vast amount of existing literature that attempts to tackle the issue of energy price forecasting, crafting an accurate model that meets established standards requires a considerable process of feature engineering and parameter fine-tuning. However, given the vast amount of available literature relating to price forecasting, the challenge now lies in identifying and incorporating novel approaches to this issue. this review aims to pinpoint any distinctive methodologies or techniques in recent electricity price forecasting literature that resonate with our dataset. These studies distinguish themselves from others in the field by either introducing novel perspectives or integrating unconventional processes to enhance existing practices. Our research begins by reviewing a meta-analysis that selects the most important features from previous studies.

Given the vast amount of time series forecasting techniques and approaches, being able to review a meta-analysis that performs feature selection and compares the accuracy of the most popular models to each other was an important starting point that allowed us to focus on a select few models. In such a study, Jacob N. Silver incorporates three sets of exogenous features into his models and compares the results to previous benchmarks (Silver, 2016). Each feature set contains externally collected or generated features that were popularly used in previous studies. The first set of variables encompasses a range of factors, including weather conditions, neighboring region import and export prices, load, price fluctuation, and volatility, alongside lagged and current price, and load data. In contrast, the second set of variables focuses on lagged values of price and load, emphasizing their volatility and fluctuations. Finally, the third set of variables primarily directs its focus toward forecasted values of price and load, providing an alternative array of exogenous variables for comprehensive analysis. After incorporating each of these feature sets into a wide variety of common time series models, he identified the ARIMA model as the most robust for predicting day-ahead hourly prices when augmented with exogenous variables that include forecasted future values of price and load. These models, when combined with carefully selected exogenous factors, typically yield Mean Absolute Percentage Error (MAPE) rates of 4% or lower. This level of accuracy not only aligns with but often exceeds the performance of previous studies across different markets, which have reported MAPEs ranging from 6% to 10% for ARIMA models. While his meta-analysis was able to identify key sets of exogenous variables that must be included to benchmark against previous results, he leaves room for improvement by incorporating said feature selection techniques into more non-conventional approaches.

In a different take on traditional feature selection, Patil, Deshmukh, and Agrawal improve on previous study results by instead incorporating a clustering approach to classify data for their models (Patil, Deshmukh, & Agrawal, 2017). A k-means clustering algorithm was applied to a NYISO market price dataset to identify distinct clusters based on the day of the week. Subsequently, the k-nearest neighbors (k-NN) algorithm was utilized to further classify these clusters according to month type classification. This approach allowed for the creation of customized ARIMA models for each cluster group, enabling more targeted and accurate forecasting within different temporal and market contexts. Notably, certain combinations of these clusters demonstrate remarkably low Mean Absolute Percentage Errors (MAPEs) in comparison to previous studies that utilized similar models without incorporating clustering techniques. While effective on a select subset of features, this study failed to incorporate a wider set of variables for clustering and classification, leaving room for future improvement.

H. Varshney, A. Sharma, and R. Kumar focus on utilizing a hybrid approach for price forecasting, incorporating a combination of singular spectrum analysis and an artificial neural network (Varshney, Sharma, & Kumar 2016). This technique significantly reduces the time required to train the model while still surpassing the accuracy of traditionally used time series models. The SSA was applied to process the data series into two distinct sub-series: the first representing the trend and a portion of the oscillating components, while the second encapsulated the remaining oscillating components excluding noise. Subsequently, a neural network architecture was constructed comprising an input layer and a hidden layer housing 15 neurons. While the hybrid Singular Spectrum Analysis (SSA) technique offers potential solutions for real-time applications, further investigation is warranted, particularly in the realms of feature selection and processing of exogenous variables. Additionally, addressing fluctuations in detected peaks through the integration of discrete-time models with the hybrid approach could enhance the accuracy and stability of electricity price forecasting, ultimately contributing to more efficient and resilient energy markets.

In another unconventional approach, Zahid was able to outperform traditional neural networks and other traditional models by using multiple machine learning techniques at different forecasting stages, combining historically successful forecasting techniques in hopes of improving model accuracy (Zahid et al, 2019). Feature selection and extraction were conducted using XG-Boost (XGB), Decision Tree (DT), Recursive Feature Elimination (RFE), and Random Forest (RF) algorithms. Enhanced Convolutional Neural Network (ECNN) and Enhanced Support Vector Regression (ESVR) models were utilized as classifiers, with Grid Search (GS) employed for parameter tuning to enhance classifier performance while mitigating the risk of overfitting. Specifically, for NYISO data, the process involved feature extraction using RFE, feature selection through a combination of attribute importance calculated by XG-Boost and DT, parameter tuning via cross-validation and GS, and prediction using ESVR and ECNN models. For future research, further enhancement of the classifiers could provide more optimal results. This includes exploring different classifiers with meta-heuristic techniques to improve accuracy.

Neupane takes a unique approach by utilizing different feature selection techniques and ensemble models (Neupane, Lee Woon, Aung, 2017). This study introduces a wrapper method for feature selection, enabling the automatic training and updating of algorithms to select the most suitable feature set for each specific algorithm. Furthermore, the study introduces two ensemble models, the Fixed Weight Method (FWM) and the Varying Weight Method (VWM), which iteratively evaluate the weights of selected learning algorithms ("experts"), with final predictions based on assigned weights. Additionally, they introduce a fallback mechanism to address fluctuations and aggregate demand response effects, ensuring prediction accuracy within a desirable range. Lastly, the study conducts comprehensive evaluations of the proposed model across various datasets and experimental configurations, demonstrating that the ensemble model automatically selects tailored features and experts, effectively capturing trends, seasonality, and patterns in energy prices. The algorithms involved in this method are used to generate predictions. The algorithm that has recently exhibited superior performance in the preceding days is ultimately chosen. While this study produces one of the best-performing models in terms of accuracy, there still exists potential for the incorporation of supplementary exogenous features, such as oil and gas prices, electricity generation modalities, etc., aimed at refining forecasting accuracy. Furthermore, subsequent studies may concentrate on integrating features to model dynamics associated with the smart grid, including demand response and load balancing.

In a 2017 study conducted by JP González and San Roque, a non-traditional Hilbertian ARMAX (Autoregressive Moving Average with exogenous inputs) model is tailored for forecasting a functional time series, specifically focusing on electricity price data (Gonzalez, San Roque, & Alonso Perez, 2018). This approach adopts a linear regression framework, wherein functional parameters interact with functional variables. These variables encompass lagged values of the series (autoregressive terms), past observed innovations (moving average terms), or exogenous variables. This enables the estimation of moving average terms within functional time series models. While this study proposes the use of a novel approach, further work is needed to meet the current standards of accurate results. One such area involves extending intervention analysis to the functional framework, enabling the modeling of sudden changes in the time series under consideration. This extension could entail the incorporation of dummy intervention variables of various natures to capture the impact of events on the response time series.

Future research in electricity price forecasting holds significant importance, particularly as more regions move towards deregulating and liberalizing their wholesale electricity supply. Improved forecasts can have far-reaching implications, enhancing market efficiency by enabling smarter generator scheduling, making Demand Response programs more viable, and facilitating the greater integration of renewable energy sources while simultaneously enhancing the efficiency of legacy fossil fuel plants. Ultimately, more accurate electricity price forecasts have the potential to elevate the overall performance of both the grid and the market, providing invaluable insights for policymakers to make informed decisions.

**Table 1**

*Summary of literature review*

|  |  |  |  |
| --- | --- | --- | --- |
| **Study** | **Best Model** | **MAPE** | **Novelty** |
| Jacob N. Silver, 2016 | ARIMAX | 4.14 | Focuses on finding the best combination of exogenous variables from past literature |
| Varshney, Sharma, & Kumar 2016 | SSA-NN | 4.73 | Combination of singular spectrum analysis and artificial neural network reduces the time required to train and performs better compared to other NN models |
| Patil, Deshmukh, & Agrawal, 2017 | ARIMA with Clustering | 4.1 | Classifies electric price data using K-mean and k-NN data mining techniques rather than by using a calendar |
| Neupane, Lee Woon, Aung, 2017 | VWM | 3.94 | Propose two different strategies for selecting each hour’s expert algorithm from the set of participating algorithms/ Final ensemble shows better results over ARIMA |
| Zahid et al, 2019 | ESVR | 4.2 | Utilizes unique machine learning techniques at different forecasting stages, outperforming benchmark schemes |
| Our Study | **Similiarities**: Forecasts LBMP using best found models and exogenous variables similar to previous literature  **Novelty**: Focus on hourly price prediction through feature experimentation and a combination of unique machine learning techniques | | |

## **DATA**

In collaboration with a national defense contractor, our team was provided a dataset spanning from January through March 2019 and January through December 2020 which contained metrics on generators in NYISO. This data can be found on the NYISO website, nyiso.com. The dataset has six columns: Time Stamp, Name, PTID, LBMP ($/MWHr), Marignal Cost Losses ($/MWHr), and Marginal Cost Congestion ($/MWHr). Please refer to Table 2 below for the data dictionary.

**Table 2**

*Data Dictionary*

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Description** | **Type** | **Example** |
| Time Stamp | This column records the date and time of a record | String | 1/1/2019 0:00 |
| Name | Name identifier of each unique generator (Location\_Type) | String | 59TH STREET\_GT\_1 |
| PTID | This column contains unique identifiers assigned to each generator | Int | 25648 |
| LBMP ($/MWHr) | “Locational Based Marginal Pricing”. The price of electricity at a specific location on the grid, calculated per megawatt-hour | Float | 25.57 |
| Marginal Cost Losses ($/MWHr) | least bit marginal price. The lowest price at which electricity can be sold in the wholesale market at a given time | Float | 1.07 |
| Marginal Cost Congestion ($/MWHr) | The costs associated with congestion in the power grid. | Float | -14.97 |

The target variable for the predictive models will be LBMP, Location Based Marginal Pricing, measured in dollars per megawatt-hour. LBMP is the chosen target variable sine s all regions have price (LBMP or day ahead, or similar) available in datasets. Thus, the price’s volatility can be used as an effective surrogate demand/capacity signal. Outliers from the dataset have not been removed since spikes in LBMP can have great significance in modeling. In addition, since most of our data dates from 2020, our team has the unique challenge on accounting for lockdowns in the energy forecasting models.

**METHODOLOGY**

Our team has experimented with several models and methods to obtain the optimal final model. These steps are outlined in Figure 1 below. First, the primary data had to be consolidated from multiple files. Then the data was preprocessed and cleaned, particularly the Timestamp column, to aggregate the data to a usable hourly format. The data was aggregated using different parameters including mean, median, and standard deviation. The data was aggregated by mean to capture outliers since outliers were important in predicting spikes in the data and would therefore need to be included. The aggregation by median was used as a more robust measure. After processing the data, our team conducted an exploratory data analysis. To cross-validate, our team has partitioned the data with a 70/30 split. This will ensure that the model has sufficient data to train on, while not compromising the amount of validation data, so the model is not overtrained.

The team then considered feature engineering and other exogenous variables including temperature, weather, lagged LBMP, and rolling mean. Then, the data was scaled so that all features would be on a similar scale which prevents specific features from dominating during model training. The models our team has focused on are ARMA, SARIMAX, ARIMAX, and Neural Networks. SARIMAX was compared with and without K-means clustering, while all other models will only be evaluated without clustering. Our team also experimented with different resampling methods, comparing resampling with all data points versus not resampling and other resampling windows. The primary statistical performance measure used to obtain the optimal model is Mean Absolute Percentage Error (MAPE). Our team’s objective was to obtain a model with a MAPE less than five percent. MAPE was chosen as the primary metric because it is expressed as a percentage of the average error relative to the actual observed values and is easily interpretable. MAPE is also useful in goal setting and can be easily used in business settings to express targets for forecast accuracy. Other performance measures considered are Test RMSE and Mean Squared Error (MSE).

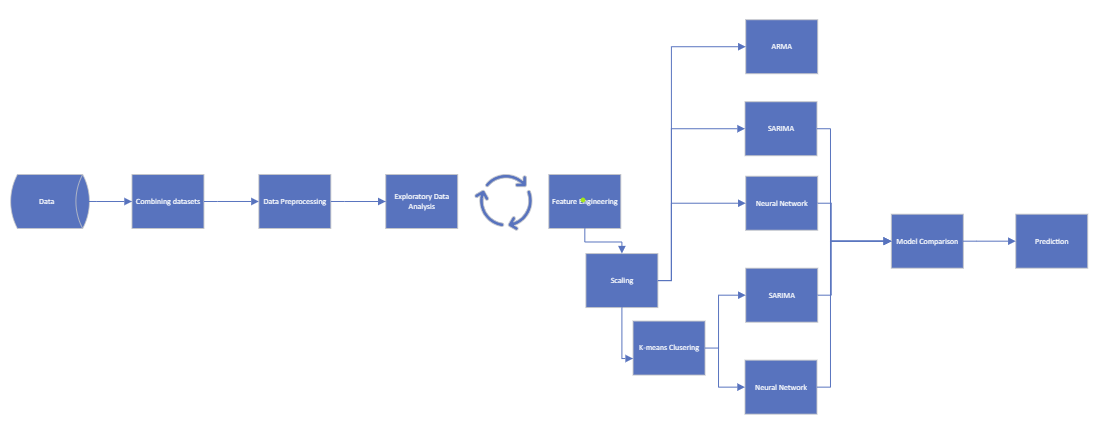
When performing feature engineering and selection, we focused on including factors cited by the US Energy Information Administration as highly influential on electricity prices. Firstly, we focused on collecting data to generate fuel features. According to the EIA, increases in fuel prices (especially for natural gas) may occur during periods of high electricity demand or when supply disruptions occur. As a result, higher fuel prices may result in higher electricity generation costs (U.S. Energy Information Administration, 2016). Per the EIA’s recommendation, our team collected the daily spot and contract prices of popular fuels. These included natural gas, uranium, and petroleum prices published by the EIA. Additionally, we were able to collect monthly power plant and transmission and distribution system costs published on the EIA and NYISO data pages. The final two features we focused on were weather data and regulations. According to the EIA, “Extreme temperatures can increase demand for heating and cooling, and the resulting increases in electricity demand can increase fuel and electricity prices” (U.S. Energy Information Administration, 2016). We collected a wide variety of publicly available weather data, including the rainfall, temperature, and wind speed of the different NYISO regions by mapping out weather stations in the specific NYISO regions. Once these stations were identified, we were able to utilize Iowa State University’s national ASOS network to retrieve the weather data for these stations during 2020. Finally, we were able to find different regulation data published by the NYISO, including any generation constraints during the year 2020.

In addition to our previously generated lagged and rolling mean variables, we combined these EIA-recommended features into the 3 hourly aggregated datasets (mean aggregation, median aggregation, and standard deviation aggregation). We then ran our models utilizing three different methods of feature selection and compared the results. For our initial models, we attempted to forecast the hourly LBMP without any exogenous variables. This allowed us to gain a better understanding of the inherent patterns and trends in the data. Additionally, these models gave us our baseline results to measure improvement with our exogenous variable models. For our second round of models, we only included the exogenous variables provided by the client. This provided us with another baseline to determine if we were improving upon our results by incorporating our own generated features in these models. However, date and time information had to be extracted from the Time Stamp manually. For our final round of models, we used a SelectKBest features selection method using the f\_regression as the scoring function. We also performed a grid search on our models, allowing us to correctly adjust each parameter to optimize results. The utilized features in all 3 rounds can be seen in the table below (Table 3). The data dictionary of the utilized generated features can be seen in Table 4. In each of these models, the Time Stamp was set as the index.

For each feature selection method, the features were tested on each dataset (mean aggregated, median aggregated, and standard deviation aggregated), using each model (AR/ARMA, ARIMAX, and SARIMAX). As expected, the more complex models (SARIMAX/ ARIMAX) performed better on the 3 datasets. Furthermore, we were able to show the inclusion of our generated features significantly improved the MAPE in all models used. However, while the models performed well on the mean and median aggregated data, all models performed poorly on the standard deviation aggregation. It seemed we were able to capture the overall central trend of the data but were failing to accurately forecast the spikes in prices. Furthermore, our best model (SARIMAX) performed the best on all three sets, but also took the longest to run. To address these issues, we incorporated an ensemble method utilizing STL and ARIMAX to improve accuracy and runtimes. STL (Seasonal and Trend decomposition using Loess) decomposition is a method used to separate a time series into its seasonal, trend, and remainder components. It utilizes a technique called locally weighted scatterplot smoothing (Loess) to estimate these components. Utilizing STL allowed us to more accurately capture the spikes present in the standard deviation dataset. We utilized this in combination with ARIMAX instead of SARIMAX to improve upon our best model’s runtime. All models and results are covered below.

**Figure 1**

*Analytics Workflow*



**Table 3**

*Features Utilized in Models (Target Variable Highlighted)*

|  |  |  |
| --- | --- | --- |
| **No Exogenous Runs** | **Client Features** | **Team Generated Features** |
| LBMP ($/MWHr) | LBMP ($/MWHr) | LBMP ($/MWHr) |
|  | Marginal Cost Losses ($/MWHr) | Marginal Cost Congestion ($/MWHr) |
|  | Marginal Cost Congestion ($/MWHr) | LBMP\_lag\_1h |
|  | Hour | LBMP\_lag\_2h |
|  | Dayofweek | LBMP\_lag\_3h |
|  | Month | LBMP\_lag\_22h |
|  | Season | LBMP\_lag\_23h |
|  |  | LBMP\_lag\_24h |
|  |  | LBMP\_lag\_2d |
|  |  | rolling\_mean\_3h |
|  |  | rolling\_mean\_12h |

**Table 4**

*Data Dictionary of Used Generated Features*

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Description** | **Type** | **Ex** |
| Hour | Indicates the hour of the day from 0-23 with 0 indicating midnight. | Int | 22 |
| DayOfWeek | Indicates the day of the week from 0-6 with 0 indicating Monday. | Int | 6 |
| Month | Indicates the month of the year from 1-12 with 1 indicating January. | Int | 1 |
| Season | Indicates the season from 1-4 with 1 indicating winter. | Int | 2 |
| LBMP\_lag\_# | The LBMP lagged for the previous 1, 2, 3, 22, 23, 24, and 48 hours. | Float | 20.45 |
| Rolling\_mean\_# | Rolling mean of LBMP for past 3 and 12 hour. | Float | 22.70 |

## **MODEL(s)**

ARMA model stands for Autoregressive Moving Average and was the least favored and least effective model that was tested. ARMA models are useful for series that are stationary and are not time series dependent. For this reason, the ARMA model is a poor fit for energy consumption forecasting since energy consumption is time series dependent and is cyclical and seasonal in nature. The ARMA model also does not account for exogenous variables which can greatly influence model accuracy as was noted in the other models which were tested. Advantages of the ARMA model include providing flexible framework to capture various autocorrelation structures within time series and its efficiency due to using only a limited number of parameters. However, due to its limitations in data with seasonal trends, it is not an ideal model for energy forecasting.

Formula: Xₜ = c + φ₁Xₜ₋₁ + φ₂Xₜ₋₂ + ... + φₚXₜ₋ₚ + εₜ + θ₁εₜ₋₁ + θ₂εₜ₋₂ + ... + θ\_qεₜ₋\_q

Where:

* Xₜ represents the value of the time series at time 't'.
* c is an optional constant term.
* φ₁, φ₂, ..., φₚ are the autoregressive coefficients.
* Xₜ₋₁, Xₜ₋₂, ..., Xₜ₋ₚ represent past values of the time series (lags).
* εₜ represents the error term (white noise) at time 't'.
* θ₁, θ₂, ..., θ\_q are the moving average coefficients (rolling means).
* εₜ₋₁, εₜ₋₂, ..., εₜ₋\_q represent past error terms.

ARIMAX model stands for Autoregressive Integrated Moving Average with exogenous variables and is a form of the ARIMA model which includes exogenous variables. This model was highly favored in our literature review. Advantages of this model include flexibility to model data with trends, seasonality, and external influences which fits well with the trends of energy consumption. By using exogenous variables, model forecasting can be greatly improved. Limitations of ARIMA models include the risk of overfitting due to multiple parameters and external variables as well as high complexity. The exogenous variables that showed the most promising results were lagged LBMP variables and rolling means.

Formula: *Xt* =*c*+*ϕ*1 *Xt*−1 +*ϕ*2 *Xt*−2 +...+*ϕp* *Xt*−*p* +*θ*1 *ϵt*−1 +*θ*2 *ϵt*−2 +...+*θq* *ϵt*−*q* +*β*1 *Z*1,*t* +*β*2 *Z*2,*t* +...+*βk* *Zk*,*t* +*ϵt*

Where:

* 𝑋*t* represents the value of the time series at time 't'.
* *c* is a constant term (optional).
* 𝜙1,𝜙2,...,𝜙𝑝*ϕ*1 ,*ϕ*2 ,...,*ϕp* are the autoregressive coefficients representing the effect of past values on the current value.
* 𝜖𝑡 is the error term (white noise) at time 't'.
* 𝜃1,𝜃2,...,𝜃𝑞*θ*1 ,*θ*2 ,...,*θq* are the moving average coefficients representing the effect of past errors on the current value.
* 𝛽1,𝛽2,...,𝛽𝑘*β*1 ,*β*2 ,...,*βk* are the coefficients associated with the exogenous variables.
* 𝑍1,𝑡,𝑍2,𝑡,...,𝑍𝑘,𝑡*Z*1,*t* ,*Z*2,*t* ,...,*Zk*,*t* represent the exogenous variables at time 't'.

SARIMAX stands for Seasonal Autoregressive Integrated Moving Average with exogenous variables and is an advanced form of the ARIMA model. This model proved to be promising in our experiments and was highly favored in our literature review. One of the advantages of the SARIMAX model is its ability to capture the seasonality of data and handle those complex patterns over time. This is particularly useful when handling data that is cyclical, like energy consumption. SARIMAX also incorporates external factors like exogenous variables which can greatly improve forecast accuracy. Disadvantages of the SARIMAX model include its high sensitivity to parameters which can lead to overfitting, requirements for a large amount of data, and high computational intensity. These models can also be complex to interpret compared to simpler models. Different tuning parameters were used including different exogenous variables, rolling windows, K-means clustering, and optimization methods such as Nelder-Mead and Powell. Our best SARIMAX model used K-means clustering with five clusters and included the exogenous variables Marginal Cost Losses ($/MWHr), Marginal Cost Congestion ($/MWHr), LBMP lagged by 2 hours, and rolling mean of 3 hours and using all data for resampling.

Formula: *Yt* =*μ*+*ϕ*1 *Yt*−1 +…+*ϕp* *Yt*−*p* +*θ*1 *εt*−1 +…+*θq* *εt*−*q* +*ϵt* +*β*1 *X*1,*t* +…+*βk* *Xk*,*t*

Where:

* 𝑌𝑡 represents the value of the time series at time 't'.
* 𝜇 is the intercept term (constant).
* 𝜙1,…,𝜙𝑝*ϕ*1 ,…,*ϕp* are the autoregressive coefficients representing the effect of past values on the current value.
* 𝜃1,…,𝜃𝑞*θ*1 ,…,*θq* are the moving average coefficients representing the effect of past errors on the current value.
* 𝜀𝑡−1,…,𝜀𝑡−𝑞*εt*−1 ,…,*εt*−*q* are the lagged forecast errors.
* 𝜖*t* is the error term (white noise) at time 't'.
* 𝛽1,…,𝛽𝑘*β*1 ,…,*βk* are the coefficients associated with the exogenous variables
* 𝑋1,𝑡,…,𝑋𝑘,𝑡*X*1,*t* ,…,*Xk*,*t* represent the exogenous variables at time 't'.

Long Short-Term Memory(LSTM) is a type of recurrent neural network (RNN) architecture designed for processing and predicting sequential data, such as time series. Unlike traditional feedforward neural networks, LSTM networks have feedback connections that enable them to capture dependencies and patterns over time. We decided to employ an LSTM model due to its ability to capture long-term dependencies and temporal patterns. However, LSTMs can be computationally intensive, which is a problem we ran into with our LSTM model. Given our goal of forecasting hourly LBMP, our LSTM models took well over an hour to run, rendering them useless for hourly forecasting purposes.

Autoregression (AR) is a statistical model where the value of the current observation is modeled as a linear combination of its past values, also known as lagged values. AR models are simple, easy to interpret, and provide a straightforward framework for understanding how past observations influence future ones. Utilizing an AR model allowed us to efficiently capture short-term dependencies in the data. However, their assumption of linearity limits their accuracy. The AR models we employed served as good baseline results to compare some of our more complex models.

Formula: *Xt* =*c*+*ϕ*1 *Xt*−1 +*ϕ*2 *Xt*−2 +...+*ϕp* *Xt*−*p* +*ϵt*

Where:

* 𝑋𝑡 represents the value of the time series at time 't'.
* 𝑐 is an optional constant term.
* 𝜙1,𝜙2,...,𝜙𝑝*ϕ*1 ,*ϕ*2 ,...,*ϕp* are the autoregressive coefficients representing the effect of past values on the current value.
* 𝑋𝑡−1,𝑋𝑡−2,...,𝑋𝑡−𝑝*Xt*−1 ,*Xt*−2 ,...,*Xt*−*p* represent past values of the time series.
* 𝜖𝑡is the error term (white noise) at time 't'.

Ensemble models are machine learning models that combine the predictions of multiple individual models to produce a final prediction. The main goal of ensemble models is to improve the accuracy, robustness, and generalization ability of the prediction compared to a single model. The ensemble model used is an STL and ARIMAX ensemble. Advantages include improved accuracy compared to any single model, robustness to overfitting, and flexibility. Disadvantages include its complexity to interpret and computational expense. The method used for the STL and ARIMA ensemble model was de-seasoning LBMP by extracting the seasonal component of STL decomposition, then point forecasting by running the ARIMAX model on adjusted data, and re-seasonalizing by adding the seasonal component back for final forecast results.

Simplified Formula: *Yt* =*Y*^trend +*Y*^seasonal +*Y*^ARIMAX +*ϵt*

Where:

* 𝑌𝑡 represents the observed value of the time series at time 't'.
* 𝑌^trend represents the estimated trend component obtained from the STL decomposition.
* 𝑌^seasonalrepresents the estimated seasonal component obtained from the STL decomposition.
* 𝑌^ARIMAXrepresents the predicted values obtained from the ARIMAX model, including any exogenous variables.
* 𝜖𝑡represents the error term.

All of the models described above were trained using hourly data from NYISO for the year 2020. There was a 70/30 data split across all models to ensure sufficient testing and validation data. Data used in the ARIMAX ensemble models and SARIMAX models were aggregated for computing efficiency. Data for SARIMAX models were aggregated on the timestamp using the mean of all data columns to capture outliers. Data for the ARIMAX and ensemble models were aggregated on the timestamp using the mean, median, and standard deviation of all data columns to capture outliers (mean), capture a robust parameter (median), and a parameter for variability (standard deviation).

## **RESULTS**

Overall, the most successful model observed is the STL and ARIMAX ensemble model using the aggregated standard deviation of the data. Other promising models included variations of AR, ARIMA, SARIMAX models and neural networks. The main metric used to determine the optimal model was mean absolute percentage error (MAPE) along with RMSE and MSE. Run time of the model was also considered with a lower run time being more favorable. The models forecast hourly LBMP ($/MWHr), so any run time longer than one hour would render the model useless for business applications. Please see Table 3 below to see a summary of the models tested.

**Table 5**

*Summary of the models tested.*

|  |  |
| --- | --- |
| **Model** | **MAPE (%)** |
| STL and ARIMAX Ensemble | 5.2932E-14 |
| Basic SARIMAX | 4.77 |
| AR | 0.051 |
| ARIMA | 0.127 |
| SARIMAX with Clustering and Exogenous Variables | 2.41 |

We observed a mean absolute percentage error (MAPE) of 5.29\*10-14 which suggests a very strong model and is well below our goal MAPE of 5%. The RMSE and MSE were near zero as well indicating strong performance. Using aggregated mean and aggregated median data also yielded low MAPE, RSME, and MSE values. This model featured one of the shortest run times of less than one second making it convenient to use in real-world modeling scenarios. Since the models forecast hourly LBMP ($/MWHr) any model taking longer than an hour to run would be rendered unusable, therefore the short run time of this model makes it successful. Please refer to Figure 2 and Table 3,4,5 below.

**Figure 2**

*This figure shows the plot between predicted and actual LBMP ($/MWHr) from January through December 2020 for the STL and ARIMAX ensemble model.*

A screen shot of a graph

Description automatically generated

**Table 6**

*This table shows the success metrics for the STL and ARIMAX ensemble model using aggregated standard deviation data.*

|  |  |
| --- | --- |
| **Runtime** | <1 Second |
| **RMSE** | 3.3529E-15 |
| **MSE** | 1.1242E-29 |
| **MAPE** | 5.2932E-14 |

**Table 7**

*This table shows the success metrics for the STL and ARIMAX ensemble model using aggregated mean data.*

|  |  |
| --- | --- |
| **Runtime** | **<1 Second** |
| **RMSE** | **3.1428E-14** |
| **MSE** | **9.877E-28** |
| **MAPE** | **1.3004E-13** |

**Table 8**

*This table shows the success metrics for the STL and ARIMAX ensemble model using aggregated median data.*

|  |  |
| --- | --- |
| **Runtime** | **<1 Second** |
| **RMSE** | **3.1428E-14** |
| **MSE** | **9.877E-28** |
| **MAPE** | **1.3004E-13** |

Autoregressive (AR) models showed very promising results particularly when using aggregated mean and median data as opposed to aggregated standard deviation. Although not as strong as the STL and ARIMAX ensemble model, the run time was also less than a second and featured MAPE values of less than 5% when using aggregated mean and median. Please refer to Figure 3 and Tables 6 through 8 below.

**Figure 3**

*This figure shows the plot between predicted and actual LBMP ($/MWHr) from January through December 2020 for the AR model.*

A graph of blue and orange lines

Description automatically generated

**Table 9**

*This table shows the success metrics for the AR model using aggregated mean data.*

|  |  |
| --- | --- |
| **Runtime** | <1 Second |
| **RMSE** | 0.129 |
| **MSE** | 0.017 |
| **MAPE** | 0.051 |

**Table 10**

*This table shows the success metrics for the AR model using aggregated median data.*

|  |  |
| --- | --- |
| **Runtime** | <1 Second |
| **RMSE** | 0.253 |
| **MSE** | 0.064 |
| **MAPE** | 0.411 |

**Table 11**

*This table shows the success metrics for the AR model using aggregated standard deviation data.*

|  |  |
| --- | --- |
| **Runtime** | <1 Second |
| **RMSE** | 2.637 |
| **MSE** | 6.952 |
| **MAPE** | 34.813 |

ARIMA models also showed strong with low run times and low MAPE though not as successful as the STL and ARIMAX ensemble model. We observed particularly high MAPE when using aggregated standard deviation data but the MAPE using both aggregated mean and median data was well below the threshold of 5%. Please refer to Figure 4 and Tables 9 through 11 below.

**Figure 4**

*This figure shows the plot between predicted and actual LBMP ($/MWHr) from January through December 2020 for the ARIMA model.*

A graph of orange and blue lines

Description automatically generated

**Table 12**

*This table shows the success metrics for the ARIMA model using aggregated mean data.*

|  |  |
| --- | --- |
| **Runtime** | 1.3 Minutes |
| **RMSE** | 0.108 |
| **MSE** | 0.012 |
| **MAPE** | 0.127 |

**Table 13**

*This table shows the success metrics for the ARIMA model using aggregated median data.*

|  |  |
| --- | --- |
| **Runtime** | 36 Seconds |
| **RMSE** | 0.237 |
| **MSE** | 0.056 |
| **MAPE** | 0.393 |

**Table 14**

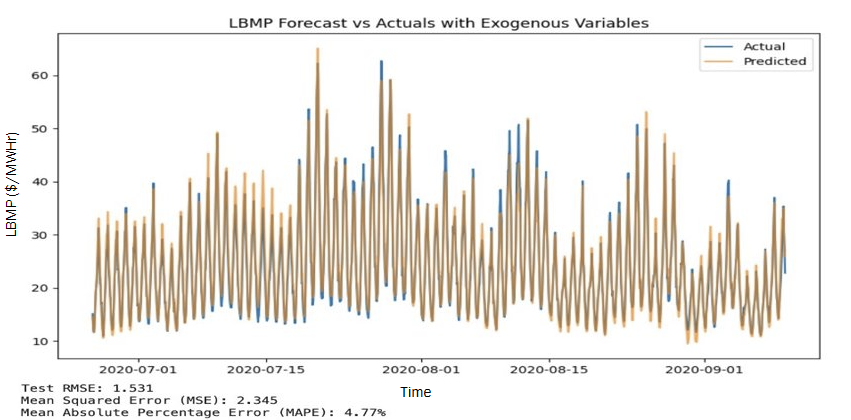
*This table shows the success metrics for the ARIMA model using aggregated standard deviation data.*

|  |  |
| --- | --- |
| **Runtime** | 17 Seconds |
| **RMSE** | 1.463 |
| **MSE** | 2.142 |
| **MAPE** | 17.414 |

Other promising models include SARIMAX models both with and without K-means clustering. K-means clustering using five clusters resulted in a small difference in Mean Absolute Percentage Error. By initially only using only the exogeneous variables Marginal Cost Losses ($/MWHr) and Marginal Cost Congestion ($/MWHr). Without using a rolling window or additional exogenous variables, the baseline MAPE was 4.77% which satisfied our goal MAPE of less than 5%. Please refer to the Figure 5 below.

**Figure 5**

*This figure shows the plot between predicted and actual LBMP ($/MWHr) from July through September 2020 for the SARIMAX model with no clustering or rolling window.*



Results from feature engineering showed the most impactful exogenous variables were lagged LBMP variables and rolling mean variables. In particular, for SARIMAX, the exogenous variables which yielded the most promising results were Marginal Cost Congestion ($/MWHr), Marginal Cost Losses ($/MWHr), rolling\_mean\_3h ($/MWHr), and LBMP\_lag\_2h ($/MWHr). K-means clustering did not make a significant difference in MAPE or run time with SARIMAX models using these exogenous variables. The run times for the SARIMAX models were longer than the run time for the best ARIMAX ensemble model at 48.16 seconds with clustering and 22.03 respectively. Still, the run times were relatively quick. Please refer to Figures 6 and 7 and Tables 12 and 13 below.

**Figure 6**

*This figure shows the plot between predicted and actual LBMP ($/MWHr) for March 2020 for the SARIMAX model with exogenous variables Marginal Cost Congestion ($/MWHr), Marginal Cost Losses ($/MWHr), rolling\_mean\_3h ($/MWHr), and LBMP\_lag\_2h ($/MWHr) and K-means clustering.*

A graph of a graph

Description automatically generated

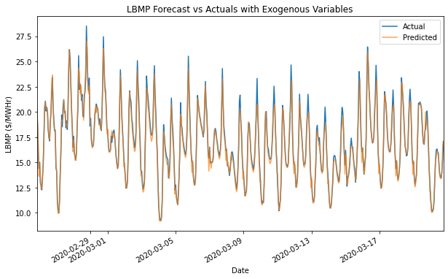
**Table 15**

*This table shows the success metrics for the SARIMAX model with exogenous variables Marginal Cost Congestion ($/MWHr), Marginal Cost Losses ($/MWHr), rolling\_mean\_3h ($/MWHr), and LBMP\_lag\_2h ($/MWHr) and K-means clustering.*

|  |  |
| --- | --- |
| **Runtime** | 48.16 seconds |
| **RMSE** | 0.581 |
| **MSE** | 0.338 |
| **MAPE** | 2.41% |

**Figure 7**

*This figure shows the plot between predicted and actual LBMP ($/MWHr) for March 2020 for the SARIMAX model with exogenous variables Marginal Cost Congestion ($/MWHr), Marginal Cost Losses ($/MWHr), rolling\_mean\_3h ($/MWHr), and LBMP\_lag\_2h ($/MWHr) without K-means clustering.*



**Table 16**

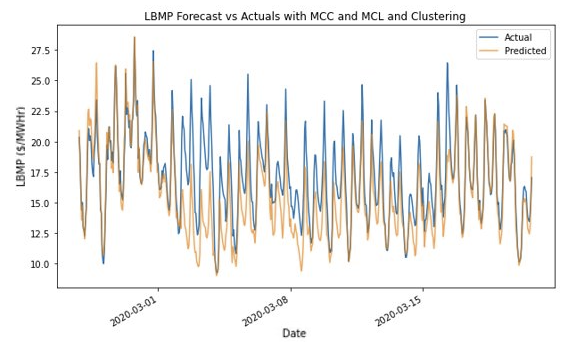
*This table shows the success metrics for the SARIMAX model with exogenous variables Marginal Cost Congestion ($/MWHr), Marginal Cost Losses ($/MWHr), rolling\_mean\_3h ($/MWHr), and LBMP\_lag\_2h ($/MWHr) and K-means clustering.*

|  |  |
| --- | --- |
| **Runtime** | 22.03 seconds |
| **RMSE** | 0.565 |
| **MSE** | 0.319 |
| **MAPE** | 2.28% |

Implementation of a rolling window also affected the results of the SARIMAX model. Rolling windows of 24 hours, 48 hours, 72 hours, and 168 hours were tested. All rolling windows less than 168 hours yielded insufficient data to run the model. The rolling window models were tested only using the exogenous variables Marginal Cost Losses ($/MWHr) and Marginal Cost Congestion ($/MWHr). The resulting MAPE both with and without K-means clustering yielded higher MAPE scores than SARIMAX models using the same exogenous variables without the rolling window. The MAPE of the SARIMAX model with the rolling window with K-means clustering was 9.65% and the MAPE of the SARIMAX model without K-means clustering was 10.73% as compared to baseline SARIMAX model using the same exogenous variables without clustering or a rolling window with a MAPE of 4.77%. Please see Figure 8 and 9 and Table 14 and 15 below.

**Figure 8**

*This figure shows the plot between predicted and actual LBMP ($/MWHr) for March 2020 for the SARIMAX model with exogenous variables Marginal Cost Congestion ($/MWHr) and Marginal Cost Losses ($/MWHr) with K-means clustering and a rolling window of 168 hours.*



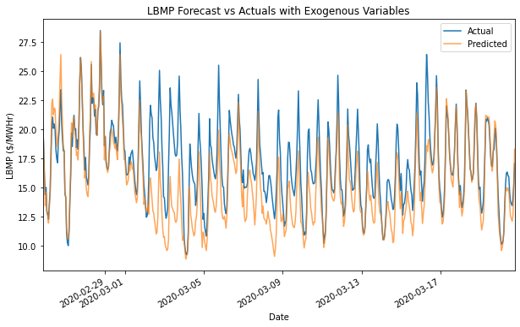
**Table 17**

*This table shows the success metrics for the SARIMAX model with exogenous variables Marginal Cost Congestion ($/MWHr) and Marginal Cost Losses ($/MWHr) with K-means clustering and a rolling window of 168 hours.*

|  |  |
| --- | --- |
| **Runtime** | 33.4 seconds |
| **RMSE** | 2.348 |
| **MSE** | 5.511 |
| **MAPE** | 9.65% |

**Figure 9**

*This figure shows the plot between predicted and actual LBMP ($/MWHr) for March 2020 for the SARIMAX model with exogenous variables Marginal Cost Congestion ($/MWHr) and Marginal Cost Losses ($/MWHr) without K-means clustering and a rolling window of 168 hours.*



**Table 18**

*This table shows the success metrics for the SARIMAX model with exogenous variables Marginal Cost Congestion ($/MWHr) and Marginal Cost Losses ($/MWHr) without K-means clustering and a rolling window of 168 hours.*

|  |  |
| --- | --- |
| **Runtime** | 16.15 seconds |
| **RMSE** | 2.501 |
| **MSE** | 6.254 |
| **MAPE** | 10.73% |

## **CONCLUSIONS**

In this study, we explored various machine learning models to forecast the energy market behavior. Especially, we focused on energy grid data across different regions of New York State. Our main aim was to improve the management of energy demand that may lead to significant business benefits such as cost reduction and optimal resource allocation.

We employed a comprehensive methodological framework in our study by leveraging advanced machine learning algorithms such as SARIMAX, ARIMAX, ensemble models and numerous neural network architectures like RNN, LSTM, or Transformers. In our evaluations, we both observed their accuracy and computational efficiency to ensure their practicality in possible real-world scenarios.

The results from our experiments indicated that STL and ARIMAX ensemble model emerged as the most effective model, with a significant balance between accuracy and short processing time. This model addresses seasonal trends in energy consumption data and also incorporates exogenous variables to improve the forecasting accuracy.

As a result, accurately forecasting the energy consumption can massively affect the strategic planning and efficiency of energy suppliers. Moreover, by forecasting peaks and downs in energy demand, energy firms better manage their resources. This better management practices could lead reducing operational costs and could help lower electricity prices for customers.

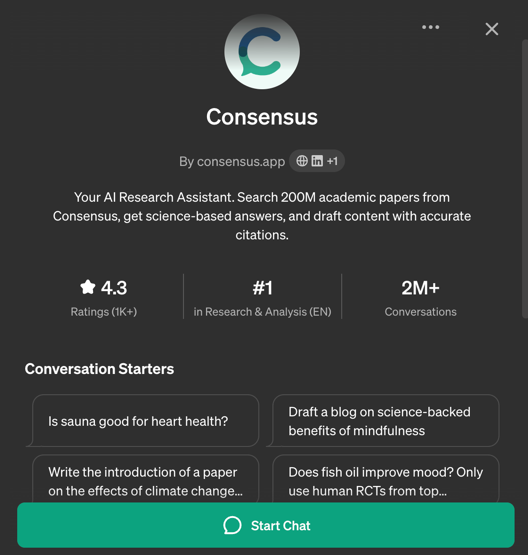
**AI RESEARCH TOOL REFLECTION**

Our team experimented with multiple methods when conducting research for our literature review including AI research tools. Two tools our team found useful in our research were Consensus, Scholar GPT, and Research Rabbit.

Consensus is an AI-driven search engine geared toward academic and research pursuits. It can streamline and simplify the search for relevant literature, datasets, and scholarly articles by understanding the intent and nuances of research queries. This platform lends itself to the needs of researchers and scholars across many disciplines. This makes it easier for researchers to locate credible sources and the latest research. See Figure 9 below for the interface.

**Figure 10**

*This figure shows the Consensus interface.*



Scholar GPT acts as a research assistant for scholarly purposes, utilizing Google Scholar and other credible academic sources to offer precise and extensive data across various scholarly subjects. It is designed to aid students, researchers, and academic professionals in locating pertinent scholarly articles, papers, and publications, streamlining the academic research process.

Research Rabbit is a platform designed to help researchers, academics, and anyone involved in scholarly work to discover academic papers, literature, and connections between various research topics more efficiently. It serves as a tool for literature discovery, enabling users to explore a web of research papers that are interconnected by citations, themes, and authors. The goal of Research Rabbit is to make the process of literature review and research exploration more intuitive and insightful, helping users to uncover relevant studies, identify key authors in a field, and understand the broader context of their research interests. Key features of Research Rabbit include mapping research connections, discovery, collaborative tools, and integration with academic databases. Please refer to Figure 10 below of the Research Rabbit interface.

**Figure 11**

*This figure shows the Research Rabbit interface for finding similar work.*

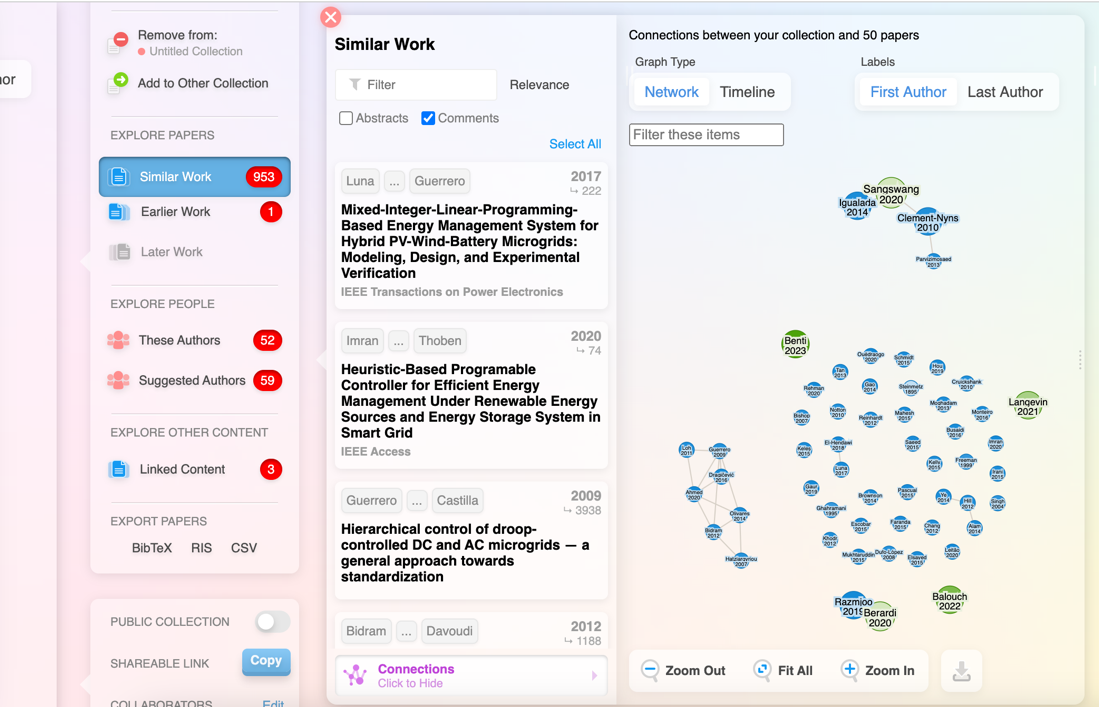


Figure 10: This figure shows the Research Rabbit interface for finding similar work.

Our team noticed several benefits when using these AI research tools. First, these tools improved research efficiency greatly. This saved the team time by delivering more relevant search results based on the true intent of the query and the researcher's profile. The tool delivered very specific results and papers on the topics searched. The tools also suggest where to look for in a paper for relevant information based on our search criteria. Another helpful feature was that when given the title of the paper, the tool can tell what topics are likely covered in the paper. In addition, the tools can access links of the paper as well. However, there were some disadvantages with these tools as well. For instance, Consensus will not summarize specific metrics from the entire paper, only the summary section. Though this issue was slightly improved in Scholar GPT the issue still persisted.

**REFERENCES**

Balouch, S., Abrar, M., Abdul Muqeet, H., Shahzad, M., Jamil, H., Hamdi, M., Malik, A. S., & Hamam, H. (2022). Optimal Scheduling of Demand Side Load Management of Smart Grid Considering Energy Efficiency. *Frontiers in Energy Research*, *10*, 861571. <https://doi.org/10.3389/fenrg.2022.861571>

Benti, N. E., Chaka, M. D., & Semie, A. G. (2023a). Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects. *Sustainability*, *15*(9), 7087. <https://doi.org/10.3390/su15097087>

Gonzalez, J. P., Roque, A. M. S., & Perez, E. A. (2018). Forecasting Functional Time Series with a New Hilbertian ARMAX Model: Application to Electricity Price Forecasting. *IEEE Transactions on Power Systems*, *33*(1), 545–556. <https://doi.org/10.1109/TPWRS.2017.2700287>

Neupane, B., Woon, W., & Aung, Z. (2017). Ensemble Prediction Model with Expert Selection for Electricity Price Forecasting. *Energies*, *10*(1), 77. <https://doi.org/10.3390/en10010077>

Patil, M., Deshmukh, S. R., & Agrawal, R. (2017). Electric power price forecasting using data mining techniques. *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, 217–223. <https://doi.org/10.1109/ICDMAI.2017.8073513>

Silver, J. N. (2016). Forecasting New York State Day-Ahead Electricity Spot Price using Time-Series Models. *Doctoral dissertation, Brandeis University, College of Arts and Sciences*.

U.S. Energy Information Administration. (2016). Prices and factors affecting prices - U.S. Energy Information Administration (EIA). Eia.gov; U.S. Energy Information Administration. <https://www.eia.gov/energyexplained/electricity/prices-and-factors-affecting-prices.php>

Varshney, H., Sharma, A., & Kumar, R. (2016). A hybrid approach to price forecasting incorporating exogenous variables for a day ahead electricity Market. *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, 1–6. <https://doi.org/10.1109/ICPEICES.2016.7853355>

Yao, Z., Lum, Y., Johnston, A., Mejia-Mendoza, L. M., Zhou, X., Wen, Y., Aspuru-Guzik, A., Sargent, E. H., & Seh, Z. W. (2022). Machine learning for a sustainable energy future. *Nature Reviews Materials*, *8*(3), 202–215. <https://doi.org/10.1038/s41578-022-00490-5>

Zahid, M., Ahmed, F., Javaid, N., Abbasi, R., Zainab Kazmi, H., Javaid, A., Bilal, M., Akbar, M., & Ilahi, M. (2019). Electricity Price and Load Forecasting using Enhanced Convolutional Neural Network and Enhanced Support Vector Regression in Smart Grids. *Electronics*, *8*(2), 122. <https://doi.org/10.3390/electronics8020122>