**Forecasting Renewable Energy Using a Time Series Model: Predicting Supply Levels in New York State**

J. Michael Barbiere

Chigozie Nnodim

Estherline Pierre

Cassidy Zaruba

INFO\_588 Analytics Practicum

Professor Rashmi Jain

Final Project Report

Due: 05/06/2022

*Executive Summary*

Over the course of our Analytics Practicum class Group 5 had the opportunity to work with a company called Solv4x where we had the chance to apply different business analysis techniques to solve a current problem the company was facing. As a group we decided to title our project “Data Collection and Forecasting” which focused on determining the correlation between electricity and demand along with renewable generation such as wind and solar to study the patterns of mismatches. The main target for this project was the New York area in order to determine the amount of energy consumption in mega-wattages that was being put to use, how much green-energy was available at the moment, and finally, the forecasted levels for both metrics/parameters. Our sponsor and guide throughout the project was Rekha Sharma, the CEO of Solv4x. Rekha along with her team will work with the Canadian government to carry out the green-energy model in the United States in the near future.

The problem we focused on solving was the increased demand of energy due to the mass adoption of Electrical Vehicles in North America and more specifically, the US. This increase in demand is due to the combustion engine slowly becoming phased out. The need to manage electrical vehicles will become imperative and as good measure, the energy that is supplied to charging stations for EV’s should logically be from green sources. At the end of the project we hoped to be able to determine the availability of green energy supply EV users’ charging demands. In order to solve our business problem we created and deployed a machine learning model that was able to forecast the number of megawatt demands for the state of New York on a daily and hourly basis. Overall the solution for Solv4x will have a good amount of economical and environmental benefits to both internal and external stakeholders. For example, power supply, availability, distribution, and source type could all be predicted by the ML models in near real-time frequency. This will be beneficial because state and federal governments can then implement a maintained EV charging program that will not rely on fossil-fuel-based energy sources.

The approaches Group 5 employed consisted of technology platforms and tools such as Python, Excel, and Forecasting Models. Along with those we used Machine Learning Algorithms, predictive analysis, and corrective analysis. The machine learning algorithm we utilized was ARIMA, which allowed us to forecast the energy demands and implement predictive and corrective analyses to update and correct the forecasting model through the use of historical data.

*Organization* and problem context.

The New York area has numerous challenges when it comes to energy supply and consumption. Since there is a limited amount of green-energy supply available, and the amount of energy consumption being experienced is currently high, the large number of electric vehicles that are already in use will continue to increase (Appendix, A1). In order to manage the transition to electric vehicles, we can forecast future levels of green-energy supply and energy consumption. This will allow businesses to plan for the necessary infrastructure changes and ensure that there is enough green energy available to meet the needs of the population. When forecasting future levels of green-energy supply and energy consumption, businesses will need to understand the current scope of the problem. The New York area currently has a limited amount of green-energy supply available. Most of the area's energy comes from sources that are non-renewable, such as natural gas and coal. As a result, there is a finite amount of green energy that can be supplied to individuals charging their EV’s. Because the amount of energy consumption in the New York area is high, and there’s an increasing number of electric vehicles being used, these vehicles' battery-life will typically be supplied by non-renewable energy sources. Also, electric vehicles require a lot of energy to charge, and this energy will need to come from renewable energy sources to further prevent exacerbating climate change. Furthermore, in the New York area the majority of this energy comes from non-renewable sources, so forecasting future levels of green-energy supply and energy consumption will allow for more intelligent grid planning (e.g., location and distribution of charging stations). This modeling approach can ensure that there is sufficient green energy supply available at crucial opportunities, within the fluctuating degrees of EV consumer and broader population demands alike.

*Business Problem Statement*

As Mass adoption of Electric Vehicles (EV) becomes increasingly popular and more widespread, the demand for energy will continue to grow. This increase in demand could potentially strain the energy supply, especially if the majority of the EVs are being charged during peak hours. Additionally, the increased demand could also lead to higher energy prices. Businesses can increase the supply of green energy by increasing the number of renewable energy sources, such as solar and wind power; however, the energy-load demand will continue to outweigh the supply of these energy sources. Therefore, EV owners will frequently rely upon fossil fuels to charge their vehicles when plugging into the grid. To understand the amounts of green energy sources available for EV owners’ power-needs; Solv4x has been examining the proportion of energy-load demand that’s supplied by renewables using Machine Learning (ML) forecasting algorithms. So, looking at the New York State geographic location, our project’s main focus is to predict the trend of renewable energy supply as a proportion of overall load-demand. Having an accurate forecast in this manner will enable EV owners (i.e., or other users of Solv4x’s services like municipalities) to know optimal re-charging times; in terms of tapping into the greatest proportion of renewable energy supply, and thereby the lowest proportion of fossil fuels in the grid.

*Analytical Problem Statement*

New York will continue to have an increase in demand due to the growth of the population and the increase in the number of electric vehicles (EVs) in the area. The rise in population will emerge as an increase in energy demand for homes and businesses. The increase in EVs will result in a rise for energy for charging stations. There are many reasons for the increase in green-energy supply, however, one of the primary reasons is due to the increase in the number of solar and wind in the area. Therefore, forecasting these levels will be an integral part in supplying accurate information to consumers who wish to avoid continued reliance on fossil fuels. Smart-grid planning and regional municipalities could additionally benefit from reliable and accurate forecasts (i.e., conducted daily), and this information needs to be readily accessible for these consumers. The front-end interface therefore should be an easily navigable dashboard that’s capable of reporting forecasts in real-time to assist in each of the fields discussed above.

*Problem Solving Approach*

In collaboration with our business sponsor, Rekha Sharma, founder and CEO of Solv4x, we had searched for a renewable energy dataset that could be reliably scraped, and fed into a given Machine Learning (ML) modeling. This search involved looking for both forecasted renewable energy and forecasted energy demand (i.e., within the New York State location). After a period of a couple days, we determined these values were not publicly available; however, the New York Independent System Operator (NYISO) website was found to provide data that forecasted the [NY State total energy load](https://www.nyiso.com/real-time-dashboard). This source thus supplied half of the target formula: predicted consumer demand. But we still had required a forecasted supply of renewable energy to complete the target computations: predicting the proportion of available energy supply that’ll be renewable-energy sourced. Further, we additionally weren’t able to construct a web-scraping application to mine these data automatically (as initially proposed); because the data weren’t readily available. Solv4x had verified the dilemma at this time, and our sponsor pivoted our assigned tasks to include construction of a ML forecasting model instead. Although real-time web-scraping wouldn’t be feasible within the scope of the project now; we would still be able to manually construct a dataset from the historical data of renewable energy.

Along these lines, the dataset would have to be fed into a ML model to make predictions of future renewable energy supply that can be compared to future consumer demand predictions. Then we’ll know the proportions of demand that’s being supplied by renewable sources (i.e., the perceived business problem). So, using the historical data archives [available on the NYISO website](http://mis.nyiso.com/public/P-63list.htm), monthly archived .csv files were downloaded to a personal computer’s hard drive. Each monthly dataset folder contained daily records of energy supply (i.e., values in generated Megawatts), organized by Fuel Category. These records had been continually updated with data every five minutes, as frequently as the NYISO dashboard refreshes. Therefore, each daily dataset contained approximately 2,000 records: MW values generated by Fuel Category every five minutes for the entire day (7 Fuel Categories \* 288 Intervals = 2,016 records). Because every month typically has 30 days, and we were asked to compile two years’ worth of data (initially one year’s worth was constructed but Solv4x requested a second year be retroactively added), the finalized dataset contained over 1M records: 2016 \* 30 \* 24 = 1,471,680. The sheer size of this dataset proved to be challenging in several ways throughout the project’s lifespan, but particularly because an Excel spreadsheet “only” contains 1,045,576 rows. Therefore, some preprocessing was necessary before even loading the dataset into a Python notebook. To fit all the data required by Solv4x into a single (.csv) file, data columns were filtered by Fuel Category type, and large stacks of data were deleted from Fuel Categories that were known to be unnecessary for this project’s scope (e.g., Nuclear, Natural Gas, Duel Fuel). This provided enough space to add all the needed data for the desired Fuel Category types (i.e., Solar and Wind) for the two-year span. Ultimately, the curated dataset contains 1,023,377 records, including data for Wind and Other Renewables (i.e., Solar) that dates back to January of 2020.

\*\*\*

Once construction of the dataset was complete, the team could commence data preprocessing and model building in Python. Since Solv4x requested that we determine a forecasted proportion of total load data that’s supplied by renewable energy sources; the remaining Fuel Category types had to be filtered out. Also, although hydro-electric energy values were available, we did not include that Fuel Category in our analysis because it’s a fairly constant supply across time (business sponsor’s request). Thus, forecasted renewable energy values were stripped from the original dataset, and their values were combined to create a simple two column dataset for total renewable energy supply (one column contained the Time Stamp, the other the renewable-energy values). Despite the fact that historical data were supplied in five-minute intervals; the predicted load data are provided in hourly increments by the NYISO. So to match the sampling rate of forecasted load data, our renewable-energy values were down-sampled into hourly values (i.e., by averaging together the six 5-minute intervals for each daily hour) (Appendix-A2).

Furthermore, several plots were constructed to visually analyze the datasets at this time, including a histogram and density plot (Appendix-A3). Additional pre-processing also involved checking for missing values and dropping rows that contained any, as appropriate. The time series was then decomposed to assess trend, auto-correlation, and seasonality. Stationarity was also tested for, which is an important consideration for making forecasts with time-series data. The results indicated that there wasn’t any seasonality or inherent trends, and the time series was determined to have stationarity (i.e., the mean, variance, and covariance don’t vary with time), with a p-value of 0.000. Therefore, our dataset did not require any further preprocessing to be used in forecasting models at this time. Solv4x had requested an Auto-Regressive Integrated Moving Average (ARIMA) forecasting model, which was constructed using the ‘stats models’ package and importing the ARIMA function. To determine the best parameters for our model (i.e., p, d, and q which correspond to the Auto-Regressive, stationarity, and Moving-Average formulas, respectively), an auto-ARIMA package was additionally installed. The Auto-ARIMA was able to compare different values of p, d, and q by using a step-wise search to determine the optimal values. The results indicated that p, d, and q should be tuned to 2, 0, 2, respectively, and the results of this analysis are provided in Appendix-B1. With this information, we could finally construct the univariate ARIMA model to use with our preprocessed renewable-energy dataset.

Subsequently, the univariate ARIMA model was fit to the cleaned dataset’s values with optimal parameters, and the saved to file using a pickle technique. The model could then be loaded from file to make a one-step prediction of the next, future value of renewable energy. Because Solv4x had expressed the need for a full day’s worth of values, the prediction code was tweaked to return 24 forecasted values (one for each future hour). These values were saved as a dataset, and also plotted along a line-plot (Appendix-A4). At this time, we attempted to supply Solv4x with a rudimentary solution to their initial business query: what proportion of load-data (i.e., demand) will be supplied by renewable-energy sources? Since we had successfully forecasted renewable-energy values, and the NYISO website supplies forecasted consumer-demand levels, it was possible to simply divide our forecasted values by their forecasted values to obtain the target results. This calculation was done within the Python notebook, and a line-graph was plotted depicting the resulting trend line (Appendix-A5). Although our business sponsor was pleased with our accomplishments at this point in time, the model’s accuracy was still unverified. Since actual data that correspond with the forecasted renewable-energy values were available after that day had passed (i.e., historical data); we could mine that dataset to use in determining the model’s forecast accuracy. Therefore, a single-day’s dataset that corresponded to the forecasted values was uploaded into the Python notebook and preprocessed as appropriate. Renewable-energy supply values were summed together (i.e., from the Solar and Wind Fuel Categories), and the forecasted values that were derived from the ARIMA model were added to that dataset as a separate column. These values were then divided by each other row-wise (e.g., actual renewable-energy supply at hour 00:00/forecasted renewable-energy supply at hour 00:00, and so on); and the differences were compiled into a third column (Appendix-B2). Finally, the Differences column was averaged together vertically to arrive at an overall mean. This calculation was also done manually in Excel for verification purposes; and the resulting value was 0.696. As our target accuracy threshold was 85%, our business sponsor decided at this point to again pivot her deliverables. She had hypothesized that the unsatisfactory accuracy score was being weighted by the variability of the Wind energy supply. Therefore, the model may achieve better results (in terms of forecast accuracy) by predicting the renewable-energy sources separately, as in a multivariate ARIMA model.

\*\*\*

The Vector Auto-Regressive Moving Average (VARMAX) model is capable of making predictions for two variables along a time series; or a multivariate time series forecast. To proceed with the VARMAX prediction, the raw dataset was preprocessed as described above; however, Solar and Wind values were not combined into a single column. During model construction in this phase, the optimal p, d, q values of 2, 0, 2, respectively were again supplied. And the model was subsequently fit to the dataset, saved to file, then loaded to make both a one-step and a multi-step prediction, as performed earlier. The multi-step forecast (one day or 24 hour’s worth) was then downloaded as a dataset, and plotted in Python using the seaborn package (Appendix-A6). We again compared the forecasted values to their actual corresponding renewable-energy supply values; however, differences for both Wind and Solar Fuel Category types were obtained. Historical data were mined from the NYISO website (once available) for the six days following the end of the original dataset’s time-series. We chose to include numerous comparative samples because our testing dataset encompassed such a wide-range of weather data (i.e., across actual seasons), and because the time of year that this modeling was constructed is known to be usually temperamental (i.e., the month of March, and the changing of winter to spring).

Therefore, six different Differences columns were constructed using the aforementioned computational techniques, for each of the six daily-testing datasets. Although the Differences from Solar-energy actuals and forecasted values were computed and are available, they were not of interest to us for this component due to their high accuracy levels (i.e., mean of 1.05 across the six days). Note: our preliminary hypothesis was that Wind variability was dragging model forecasting accuracy down. So, using the six different datasets, vertical means were calculated among all Differences columns from Wind energy; these ranged from a low 0.95 to high of 3.80 on the second and fourth days, respectively. The accuracy, as compared to actual renewable-energy supply values, was very high on the second day (i.e., 95%, well above the threshold of acceptability at 85%); but was as bad as almost 4x under-estimating the actual values another two days afterwards (3.80). Therefore, we were to confirm our hypothesis that the Wind energy values were too unpredictable for the forecasting model to accuracy predict; on the fourth testing day the model accounted for only 26.31% of the actual renewable Wind energy-supply values (i.e., 1/3.80 = 0.2631).

Additionally, descriptive statistics were employed to determine the minimum and maximum values for each of the daily Differences columns. This allowed us to analyze the degree of variability that was exhibited, across hourly time intervals. The min/max values were combined into a singular dataset (Appendix-B3), transposed horizontally, and plotted using a boxplot in the Python notebook using the seaborn package (Appendix-A7). This visualization succinctly displays the ranges of accuracy (or inaccuracy) displayed by the forecasting model across each of the six novel datasets. It can be interpreted that the model’s levels of accuracy only fell within acceptable territory (i.e., within 15% of a perfect 1) on just one of the testing days (i.e., day two as described above). Although four out of the six plots have tighter ranges (i.e., and less variability; on the second, third, fifth, and sixth days); three of these fall at a mean accuracy of two or above. That signifies to us that the model was capable of displaying some degree of precision, but was routinely underestimating actual values by at least one-half. In conclusion, the group, along with our business sponsor, Solv4x, had theorized that since the ARIMA (and VARMAX) models are linear at heart, they’re predictive power will degrade the further they’re used to forecast into the future. That certainly does seem to be the case here; however, it’s also noted that accuracy levels for the subsequent fifth and sixth testing days were consecutively lower following the huge miss that our model had on the fourth day. So it might just be the nature of wind’s unpredictability, as well as the more tumultuous time of year when testing occurred, that resulted in our model’s poor performance.

*Findings and Results*

Having accurate wind power forecasting is a critical factor in power-grid planning, and reduces the risk of uncertainty associated with integrating wind into power systems (Foley et al., 2012). In the academic literature, ARMA and ARIMA models have been used successfully in forecasting of wind speed and generated power by wind farms. So, knowledge of the power-generation levels produced by wind farms is important in managing a regional power-grid. Excessive energy-supply needs to be harvested and stored if possible, and deficient supply needs compensation from other energy sources. Other undertakings have used the ARIMA model in short-term forecasting (i.e., hourly or daily values); however, the results were inferior to blackbox ML algorithms with sufficient features (e.g., Artificial Neural Network, or ANN). And a variant of the ANN (i.e., the recurrent multi-layer perceptron model, or MLP) has performed well in longer-term power supply forecasting (i.e., two to seven days out). Additionally, other blackbox models have been used in the literature to predict wind speed, wind power generation, and general wind power forecasts via turbine farms. These have included Support Vector Machine (SVM), MLP, ANN, regression trees, random forest, and other ML algorithms. Also, Numerical Weather Prediction (NWP) models have been widely used, which compute various future weather conditions based upon current observations (i.e., compared to feeding a model historical data as demonstrated herein). The NWP technique relies upon a high accuracy of modeling data, so an ensemble method has been proposed to mitigate the effect of a single NWP model. Other “mixed model” techniques have been described (e.g., combining kNN and ANN approaches) to optimize forecasting accuracy with respect to Root Mean Squared Error (RMSE), with results indicating that the combination of several models together can improve results substantially (Perera et al., 2014).

Along these lines, our findings should be carefully weighed against the potential to incorporate more sophisticated modeling techniques (i.e., combining multiple models into a potential “gray-box” solution). When using the VARMAX model to forecast wind values, our resulting RMSE was 220.98, which is fairly large given the average Wind-energy value as reported by NYISO for all days in 2020-21 as 482.14. Therefore, our model’s error-margin was nearly half the value of any given wind-supply data (220.98 / 482.14 = 0.46). Although the RMSE and calculated accuracy levels were unsatisfactory; Solv4x expressed their satisfaction in our reporting, as our group was only one in a larger network of teams. The business sponsor had expressed her intentions in using the foundational findings to construct a more complex forecasting model that would account for numerous other features (e.g., other potential input variables to enhance predictive power). It’s possible that Solv4x may combine our results with an LSTM or ANN model that was developed elsewhere; however, such an undertaking is beyond the scope of this project. It’s also noteworthy to mention that our Solar forecasting values were highly accurate, and using the ARIMA model may be an acceptable forecasting solution for states with an increased emphasis on this energy source. For example, California absorbed 15.7%, or 30 TWh, of its energy needs in 2020 from solar sources (Jacobs, 2021); compared to New York’s solar energy values, which was approximately 2.5% of supply (wind comprised 4% and hydroelectric ~12%) (US EIA, n.d.). Therefore, specific geographic considerations should also be considered when potentially implementing the model.

*Possible Extensions*

In the present context, predictions of renewable energy sources were computed to construct a rough mathematical model of renewable energy supply as a proportion of consumer demand. Solv4x was primarily interested in developing a framework for knowing the most opportune times throughout a given day to plug electric vehicles (EV) into the power-grid for re-charging purposes. Logically, the sources of energy that most EV’s tap into are those that are readily supplied by the local grid. And since no regional grid is completely reliant upon renewable energy sources at this time, most EV drivers are ultimately powering their vehicles with fossil fuels. For renewable-energy prediction, the energy sources that ML algorithms have been most frequently applied towards are wind and solar (Lai et al., 2020). Both energy sources can prove to be challenging to forecast; and this variability could lead to reluctance in readily adopting these sources to the grid (Alsharif et al., 2019). In general, wind power forecasting models typically study “stable” wind conditions, and forecasting models produce less error when predicting wind speeds versus generation of wind power (e.g., Megawatts) (Foley et al., 2012).

Also, solar radiation can be extremely variable depending, and may be influenced by several factors (e.g., geographic location, humidity, season, cloud-cover, etc.) (Alsharif et al., 2019).

Therefore, as big data computing and Machine Learning (ML) techniques continue to advance, errors in forecasting predictions will be minimized; which should lead to an increased interest in the space. ML algorithms have successfully forecasted renewable-energy levels in the past, as well as other types of renewable energy sources too. For example, Lai et al. (2020) have described the potential for both AI and mathematical models to forecast tidal energy, biomass energy, hydraulic power, and geothermal energy. Whereas these energy sources may be more predictable, wind forecasting tools will continue to be invaluable because they can provide a better understanding of the financial and technical risk that’s inherently associated with wind power production (Foley, 2012). Essentially, the more error that’s reduced in wind and solar energy forecasting, the more confidence and certainty that regional electricity markets will have in the trading/brokerage of such contracts. And this increased certainty could lead to an even greater rate of renewable energy adoption in various consumer markets.

Other potential directions that can ostensibly be taken from here include increasing the number of features, or input variables, included in the model fitting. Selection of input variables can be used in determining the architecture of the model; however, our model was fitted with only the time series values themselves as a feature. Solv4x had acknowledged this potential shortcoming during our time working with her, and had advised us that her team will be developing more complex models with additional features. To our knowledge, those developments will make use of this preliminary experiment. Finally, Sfetsos (2010) found that AI-based models can outperform their respective linear ones (e.g., ARMA, ARIMA) on mean hourly wind speed prediction; however, other research has found the opposite (i.e., Lai et al., 2020 reported that a Bayesian linear regression model was superior to AI algorithms in predicting the water levels of a reservoir). Therefore, different forecasting models themselves should be tested on the dataset, and we hope to have provided a coherent framework for doing so herein.

*Conclusion*

In conclusion, the breadth and depth of the present project’s scope has been characterized by a narrower focus of the business problem presented to us by our sponsor, Solv4x. This focus involved a heavy-handed coding component that prompted a significant amount of outside research, and hands-on learning. Several of the results have been deemed pertinent solutions to our business problem; however, shortcomings have also been recognized, particularly the accuracy levels of the completed model. The total renewable energy supply (i.e., from wind and solar sources) for NY State had been forecasted with a ~70% accuracy; however, further testing revealed that disaggregating the two energy sources resulted in significant differences of accuracy levels. That is, the wind energy predictions were far more inaccurate than those forecasted for solar energy, across several days of testing on novel datasets. To understand why, we have found substantial support of a linear ARIMA model in similar forecasting studies with both wind (Foley et al., 2012; Lai et al, 2020; Sfetsos, 2000; and Perera, 2014) and solar energy (Alsharif et al., 2019). Some differences have been noted though. Foley et al. (2012) described a 20% error reduction rate using an ARMA model to forecast hourly wind speed with a 10 hour forecast; however, that study incorporated nine years of historical data, compared with just the two years that we had compiled. Similarly, solar radiation data were fed into an ARIMA model by Alsharif et al. (2019) that encompassed 13,513 values across 444 months, or 37 years. Clearly, when constructing linear forecasting models, the more data the better. Therefore, future extensions of this project should consider assembling datasets that span at least five years worth of historical data.

Furthermore, despite strong support of the ARIMA model in the literature (see above), the effectiveness of developing a hybrid model is another popular technique (Lai et al., 2020). The potential benefits of this include decomposition of the dataset, and other preprocessing techniques used in ML models; as well as inputting multiple features, or input variables, which will aid the model’s ability to make accurate predictions (Stefsos, 2000). Because we’ve developed a linear model, the number of potential input variables was limited; however, we were able to estimate and control for potential autocorrelation (ACF) and partial autocorrelation (PACF). These factors could have an effect on the ARIMA (and VARMAX) models’ p and q parameters, but by using the ! pip install pmdarima package, our models’ parameters were fine-tuned to the optimal numbers automatically.

Therefore, the totality of this project’s scope can be assessed by looking at the average predicted renewable-energy supply versus the actual renewable-energy supply values (Appendix-A8). There we see the predicted generation of MWh at around 600 (blue bar), compared to actual values of MWh generated at 400 (light green bar). This means our forecasting model was “overly optimistic” about renewable-energy supply values. Although mathematically the forecasted values were 150% greater than actual levels (600 vs. 400); it could also be reasoned that this difference was miniscule compared to the total MWh consumption that NY State experiences daily. For example, in 2020 there were 140,406,632 MWh sold to consumers within that area (US EIA, n.d.). This equates to 384,675 MWh per day, or roughly 16,000 MWh per hour. And since solar and wind energy sources account for approximately 5% of total energy consumption in NY, we can expect that, on average, a 16,000 MWh day consists of 800 MWh of renewable energy. So, our predicted value of 600 MWh may actually have been “regressing closer to the mean'' as compared to the actual value of 400 MWh observed.

Comparatively, the difference in actual vs. forecasted values represents a broader known trend in the forecasting of renewable energy supply: fluctuations in wind energy are much more difficult to predict and account for, while representing a larger proportion of renewable energy supply. Because solar energy is seen as a relatively less volatile energy source, our findings here are more encouraging than preliminarily determined. Wind and solar energy combined can be predicted with an approximately 70% accuracy rate using a white-box algorithm that only accounts for a single input variable (i.e., time series data). And these values represent only a fraction of the overall energy consumption seen within the target consumer geographic area (i.e., New York State). Therefore, we feel that the project in its entirety has some merit that Solv4x could use to further propel its business initiatives; and we’ve been informed by our business sponsor that the work presented here has been more than satisfactory from her perspective.

Sources

Alsharif, M. H., Younes, M. K., & Kim, J. (2019). Time series ARIMA model for prediction of

daily and monthly average global solar radiation: The case study of Seoul, South Korea. Symmetry, 11(2), 240.

Lai, J. P., Chang, Y. M., Chen, C. H., & Pai, P. F. (2020). A survey of machine learning models

in renewable energy predictions. *Applied Sciences*, *10*(17), 5975.

Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and

advances in forecasting of wind power generation. Renewable energy, 37(1), 1-8. doi:10.1016/j.renene.2011.05.033

Jacobs, F.. (2021, September 28). *Electricity Generation by US State.* Big Think. Accessed

April 22, 2022 from [Electricity generation by U.S. state - Big Think Energy mix per U.S. state](https://bigthink.com/strange-maps/electricity-generation-by-state/)

NYC.gov. (n.d.). *NYC is building a clean, resilient, and affordable energy system.* Accessed

February 10, 2022, from [Energy - Sustainability (nyc.gov)](https://www1.nyc.gov/site/sustainability/our-programs/energy.page)

NYISO (n.d.). *Power Trends 2021, Annual Grid & Markets Report.* Accessed February 14,

2022, from [Power Trends - NYISO](https://www.nyiso.com/power-trends)

Perera, K. S., Aung, Z., & Woon, W. L. (2014, September). Machine learning techniques for

supporting renewable energy generation and integration: a survey. In *International Workshop on Data Analytics for Renewable Energy Integration* (pp. 81-96). Springer, Cham.

Sfetsos, A. (2000). A comparison of various forecasting techniques applied to mean hourly wind

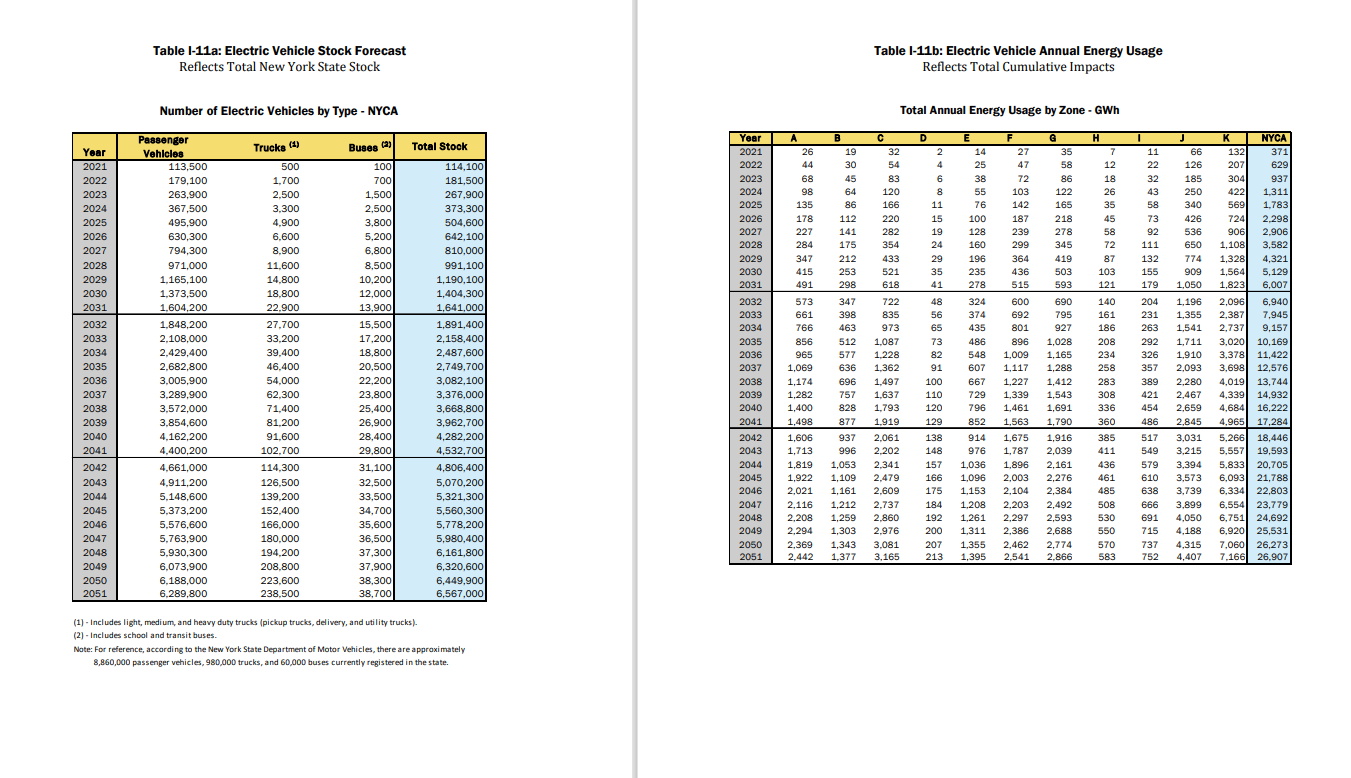
speed time series. Renewable energy, 21(1), 23-35.

U.S. Energy Information Administration. (n.d.). New York, State Profile and Energy Estimates.

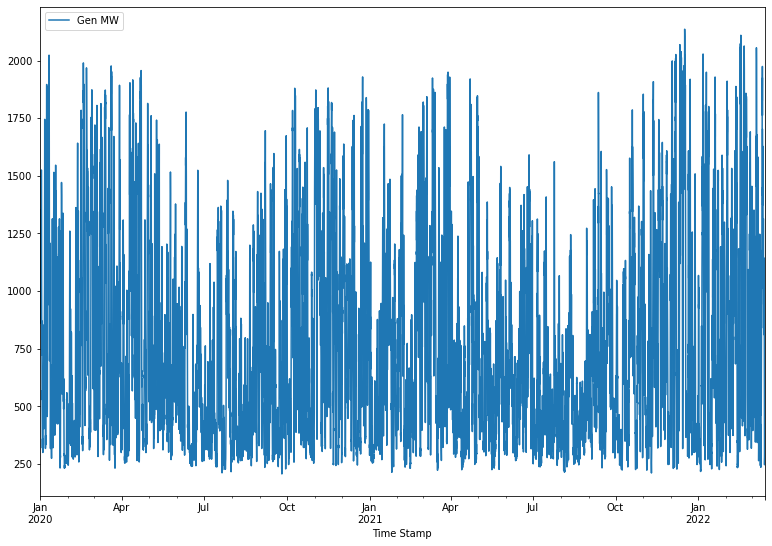
[www.eia.gov](http://www.eia.gov), Accessed April 22, 2022, from <https://www.eia.gov/state/analysis.php?sid=NY>

*Appendix – A: Graphic Charts and Visualizations*

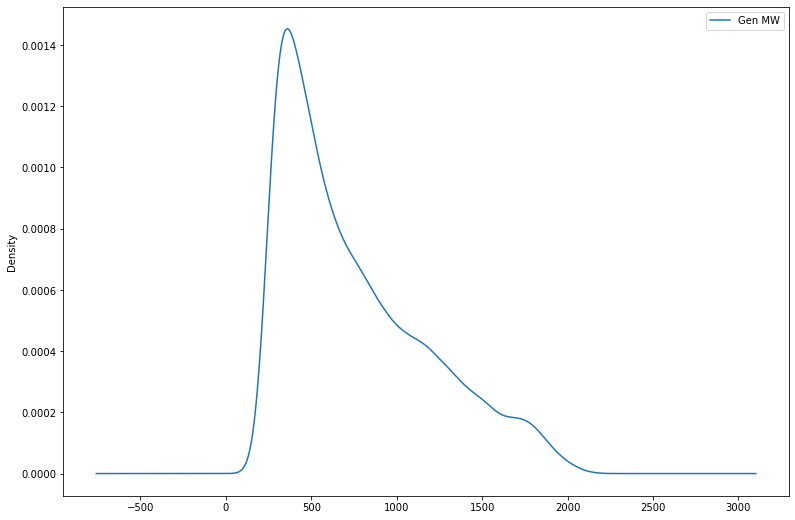
*I. A1 - Forecasting Electric Vehicle Stock*



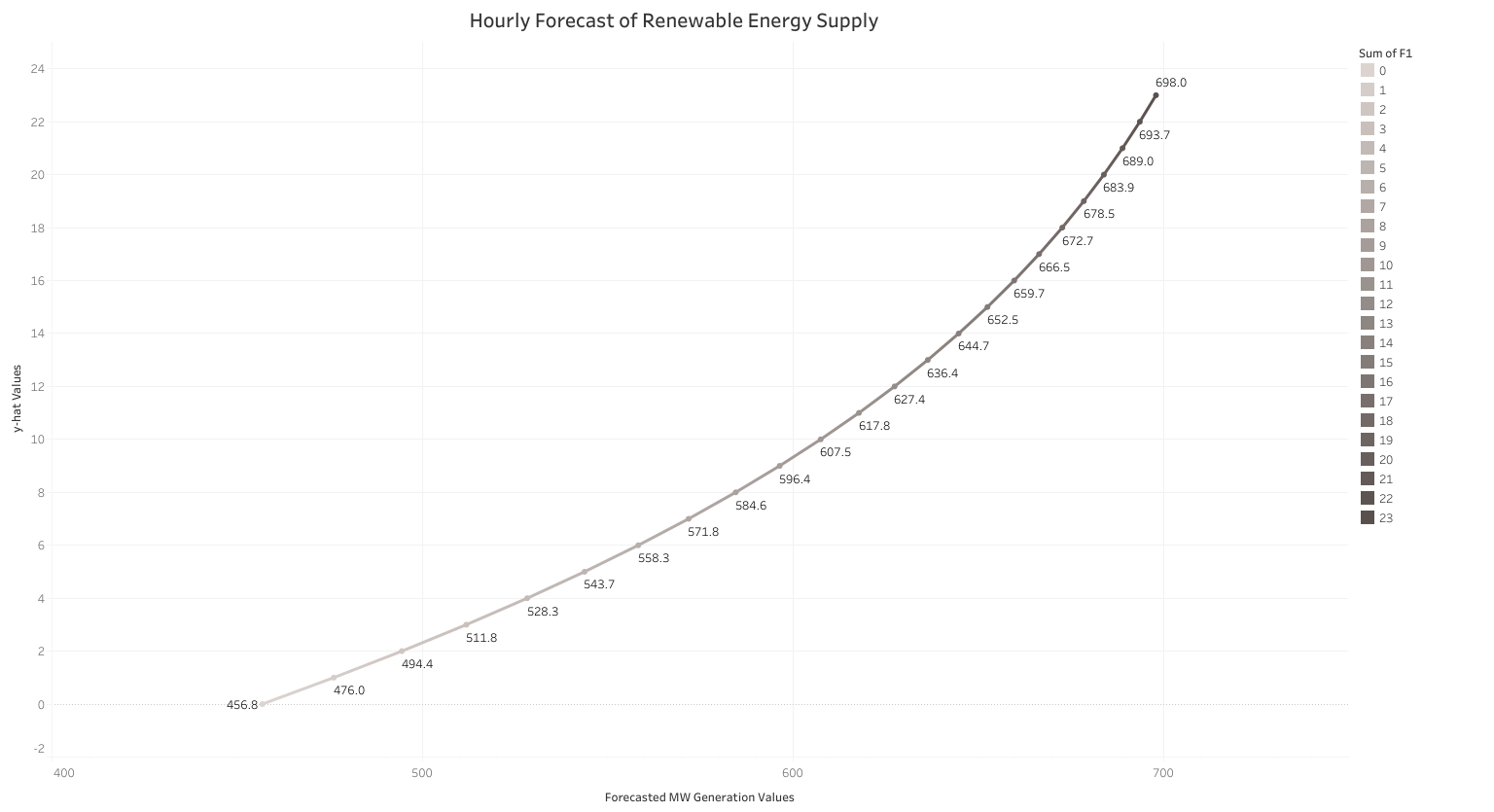
*II. A2 - Resampled Hourly Values*

**

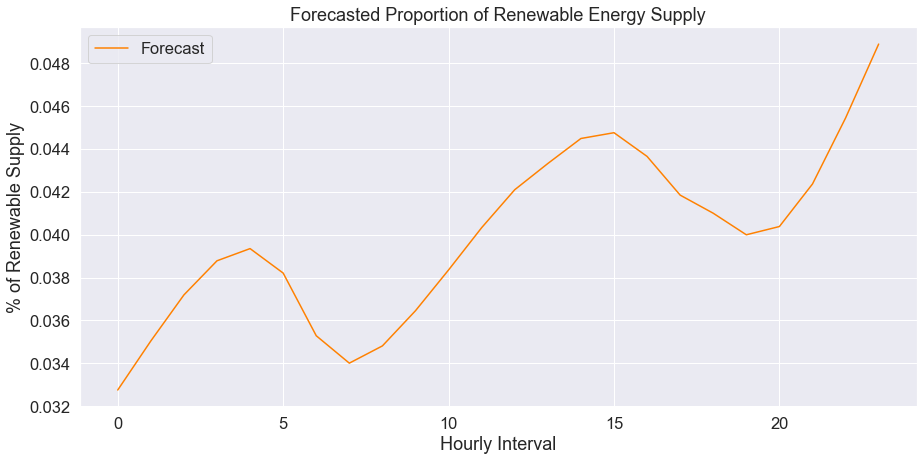
*III. A3 - Density plot of dataset*

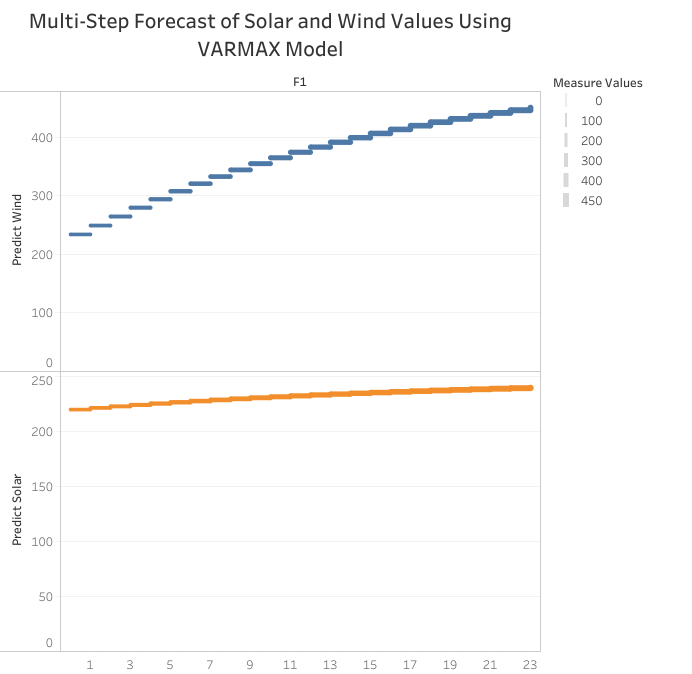


*IV. A4 - Forecasted values*

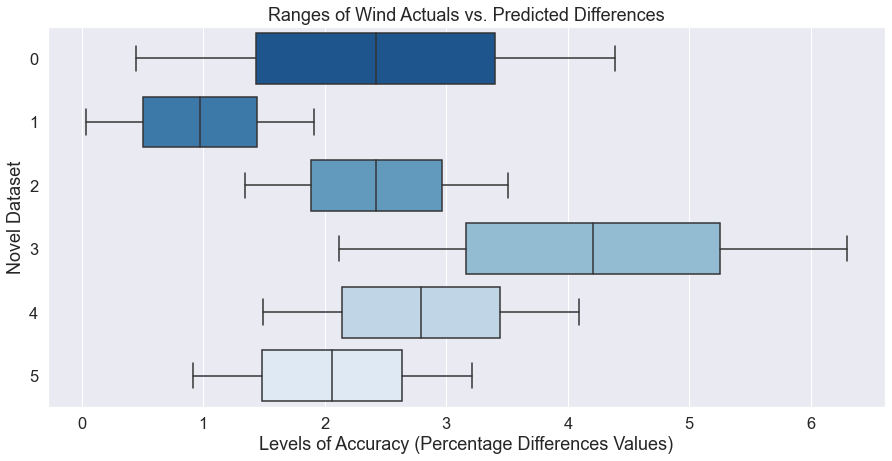


*V. A5 - Proposed solution to business problem.*

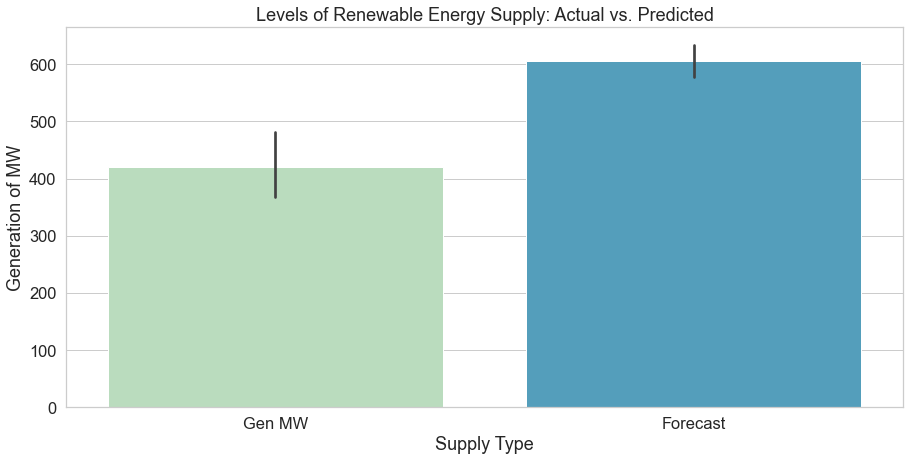
*VI. A6 - Multivariate prediction (wind and solar separate)*



*VII. A7 - Plotted MinMax descriptive statistics*

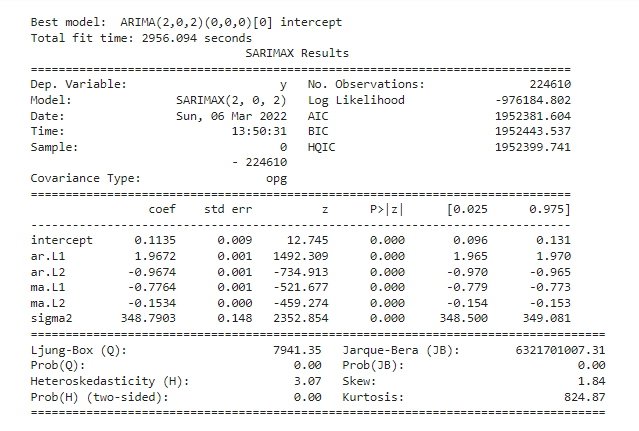


*VIII. A8 - Actual vs. Predicted values bar graph*

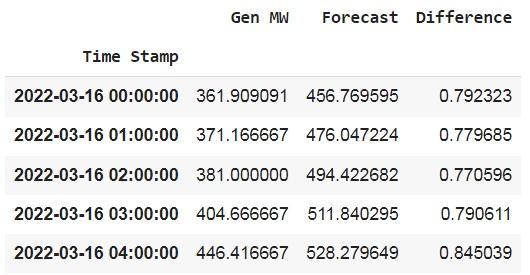


*Appendix – B: Data Tables and Other Figures*

*IX. B1 - Auto-ARIMA output*

**

*X. B2 - Differences between actual and forecasted values*



*XI. B3 - MinMax descriptive statistics*



*Appendix – C: References to Literature*

*C1- Source: Foley et al., 2012*Types and kinds of forecasting models, their developers, and some geographical locations where deployed.
