# Leading a Carbon Neutral Future through Energy Storage

## United States vs. Other Leading Countries


**SQ21 OMSBA 5067 01 | Data Translation Challenge**

**Chequala Fuller | June 8, 2021**

## I. INTRODUCTION

In leading a carbon neutral future through energy storage systems (ESS), the United States (US) has continually progressed in its expansion of clean energy compared to other leading countries such as China (CN), Germany (DE), Spain (ES), and Australia (AU). In the energy industry, and specifically in the utility industry, we know that there is a growing number of renewable generations penetrating the electric grid, causing various issues. In approaching those issues and determining that renewable generation, although significant in providing sufficient and clean energy to consumers, does create stressors within an already aged electric network, triggering instabilities in voltage and frequency and affecting the performance and reliability of the existing grid and its infrastructure. Many electric utilities and energy companies have been exploring and implementing solutions to mitigate and, ultimately, prevent major stressors and instabilities from occurring.

One of the vital missing pieces to achieving clean, safe, and reliable energy on a grid scale, contributing to the fight for climate change, is energy storage. As energy storage technology becomes more attractive, commercially available and more affordable, I wanted to look at how the United States has progressed in its goal to modernize our grid compared to other leading countries as well as predict how the US is expected to continue to lead in this effort of integrating energy storage throughout the 21st century. Especially provided that many progressive US states have been working towards a common goal of attaining zero emissions and carbon neutral through 2035 and beyond.

## II. LITERATURE REVIEW

In performing preliminary literature research on articles targeting my proposed subject or similar, I found it challenging to gather literature that supports it from a machine learning approach. Most articles I had encountered focused on machine learning frameworks for energy consumption and energy demand forecasting but not machine learning for grid modernization or energy storage growth forecasting. To best support my subject, I opted for articles that complimented surrounding work, such as topics highlighting grid-level applications of electrical energy storage and grid modernization prioritization, then aimed to align it with my work.

For my literature review, the five articles that best complimented my machine learning data translation project included, *A Smarter Grid for Renewable Energy: Different States of Action*, *Modernizing the Grid Challenges and Opportunities for a Sustainable Future*, *Grid-Level Application of Electrical Energy Storage: Example Use Cases in the United States and China*, *A fuzzy Analytic Hierarchy Process algorithm to Prioritize Smart Grid Technologies for the Saudi Electricity Infrastructure*, and *US government works toward grid modernizatio*n. Each article seemed to provide unique perspectives to information that, with more refinement, contributes to answering my machine learning problem.

In *A Smart Grid for Renewable Energy: Different States of Action*, researchers expand on the comparative analysis of smart grid development in seven of the United States – California (CA), Illinois (IL), Massachusetts (MA), Minnesota (MN), New York (NY) and Vermont (VT). Developing, building, and modernizing the electric grid (also known as a smart grid) is jurisdictionally complex in the United States; states and regions are approaching grid modernization in different ways and this study focuses on the discussion of how renewable resources for electricity are driven amongst various states. The article concludes that the data shows connections between smart grid and renewables indirectly related to the renewables deployed in the state. Overall, the state-level differences present challenges for both federal and state-level policy makers attempting to transfer policies from one state to another. While this paper focused on the United States, many of these lessons can be generalized across various countries and across various contexts.

On a broader scale, the article of *Modernizing the Grid Challenges and Opportunities for a Sustainable Future* focuses on how several U.S. states, including California, New York, and Texas as well as countries like Germany, Spain, and Australia are aiming to adopt 30% or more high penetration renewable generation and distributed energy resources (DERs) for achieving a carbon-neutral electric grid and mitigating climate change while increasing resiliency, reliability, and demand. Progressive integration of DERs and renewables can improve efficiencies in the use of the existing grid and become part of the overall development strategy for balancing supply and demand uncertainties and risks with a variety of different resources. Moreover, it is imperative for these integrations to take place to upkeep the operation of modern and future electric distribution systems; grid modernization is a vital phase in achieving these goals.

Furthermore, my third article of *Grid-Level Application of Electrical Energy Storage* narrows its focus down to two countries with the world's largest economies and power systems – the United States (U.S.) and China. Sharing common goals of electrical energy storage (EES) application and integration, the increasing demand for specialized grid modernization in both countries continues to drive the adoption of grid-scale solutions, which present valuable technical and institutional lessons for emerging EES worldwide.

EES systems have played an increasing role in successfully integrating large amounts of variable renewable resources into existing and future electric grids. In 2015, the U.S. energy storage market grew 243% in one year on a national scale. This has been the largest growth on record. As a result, in 2016, the U.S. Senate passed the Grid Modernization Act, initiating a comprehensive, nationwide effort to enhance the future of the U.S. electric grid and resolve the challenges of integrating conventional and renewable sources with energy storage and smart grid technologies. Compared to the U.S., the pace of EES development in China is relatively slow, however, national policies continually encourage the integration of EES.

Out of the five, the only article that presented valuable algorithmic models to support the case studies covered was *A fuzzy Analytic Hierarchy Process algorithm to Prioritize Smart Grid Technologies for the Saudi Electricity Infrastructure*. Although Saudi Arabia is not considered a leading country in energy storage, the approach taken to establish a framework to develop prioritization of their grid modernization plan through fuzzy set theory and analytic hierarchy processes, provides a valuable algorithmic model that can be adopted.

Finally, the *US Government Works Toward Grid Modernization*, although lacks quantitatively supported models, showcases the progression and projection of how the U.S. is working towards supporting and adapting to grid modernization based on funding of Department of Energy (DOE) laboratories. The collaboration of government agencies, laboratories and local utilities displays reassurance as scientists, engineers, and the US government work through the challenges of grid modernization and ultimately forging ahead as the leading nation in integrating clean distributed energy resources (DERs), implementing real-time system monitoring, developing and testing advanced controls, and designing advanced cyber security technologies.

## III. METHODOLOGY

Gathering the the global energy storage dataset made available through the DOE Global Energy Storage Database (GESDB), I decided to select this dataset because I believed it would provide me with the observations and variables needed to explore the proposed machine learning problem. The DOE GESDB provides research-grade information on grid connected energy storage projects on relevant state and federal policy database and houses over 1600 storage projects worldwide. In the GESDB, there are a total of 1,687 observations and 94 variables. For the purpose of this project, I narrowed down the observations to 1,081 and the variables to 14.

With the data gathered and uploaded, I prepared the data to be trained and tested by cleaning up any anomalies or missing data points. Rather than removing any missing or 'nan' datapoints and omitting essential information, I decided to interpolate the missing data to provide complete observations for every variable and feature. Once all missing values were filled, a quick review of the dimensionality and statistical summary were performed to verify that the data rendered viable.

After preparing the data and verifying that its usable, I then visualized the data to better understand any patterns lying within the data. The first visual consists of a simple bar chart, verifying and comparing the count of ESS within each respective country. As depicted in **Figure 1**, we can see the total ESS for the US is 742, which is the highest, followed by China at 103, Germany at 97, Spain at 67, and Australia at 72.

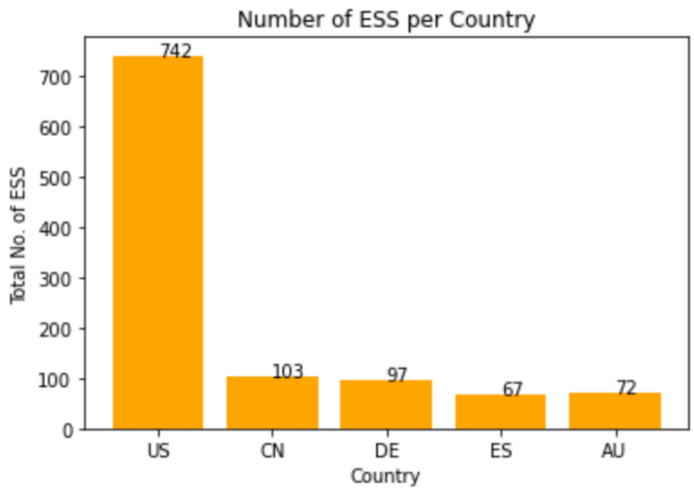



***Figure 1****: Bar chart comparing the number of ESS for each country*

Additionally, I plotted the Rated Power (kW) against its respective Commissioned Year to get a bigger picture on ESS commissioned or announced from 1900 – 2064 (see **Figure 2**). Then stacked the number of ESS based on Rated Power (kW) for each country as shown in **Figure 3**. This allowed me to see what was occurring within the patterns and what would be the best approach in selecting the models.

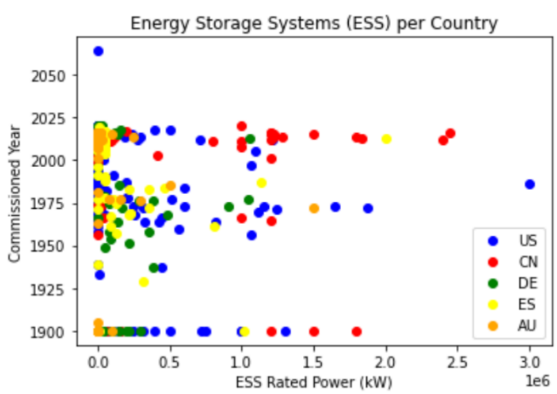



***Figure 2****: Plot of ESS rated power (kW) and their commissioned year for each country.*

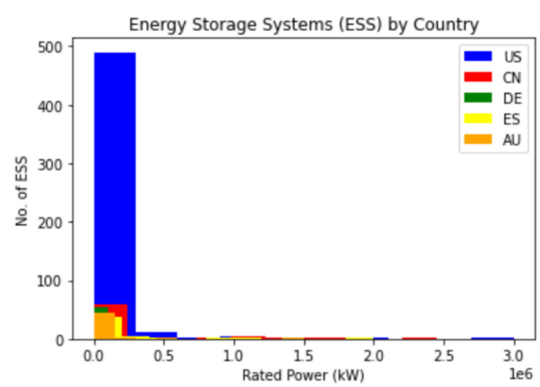



***Figure 3****: Overlapped bar chart for number of ESS and Rated Power (kW) for each country.*

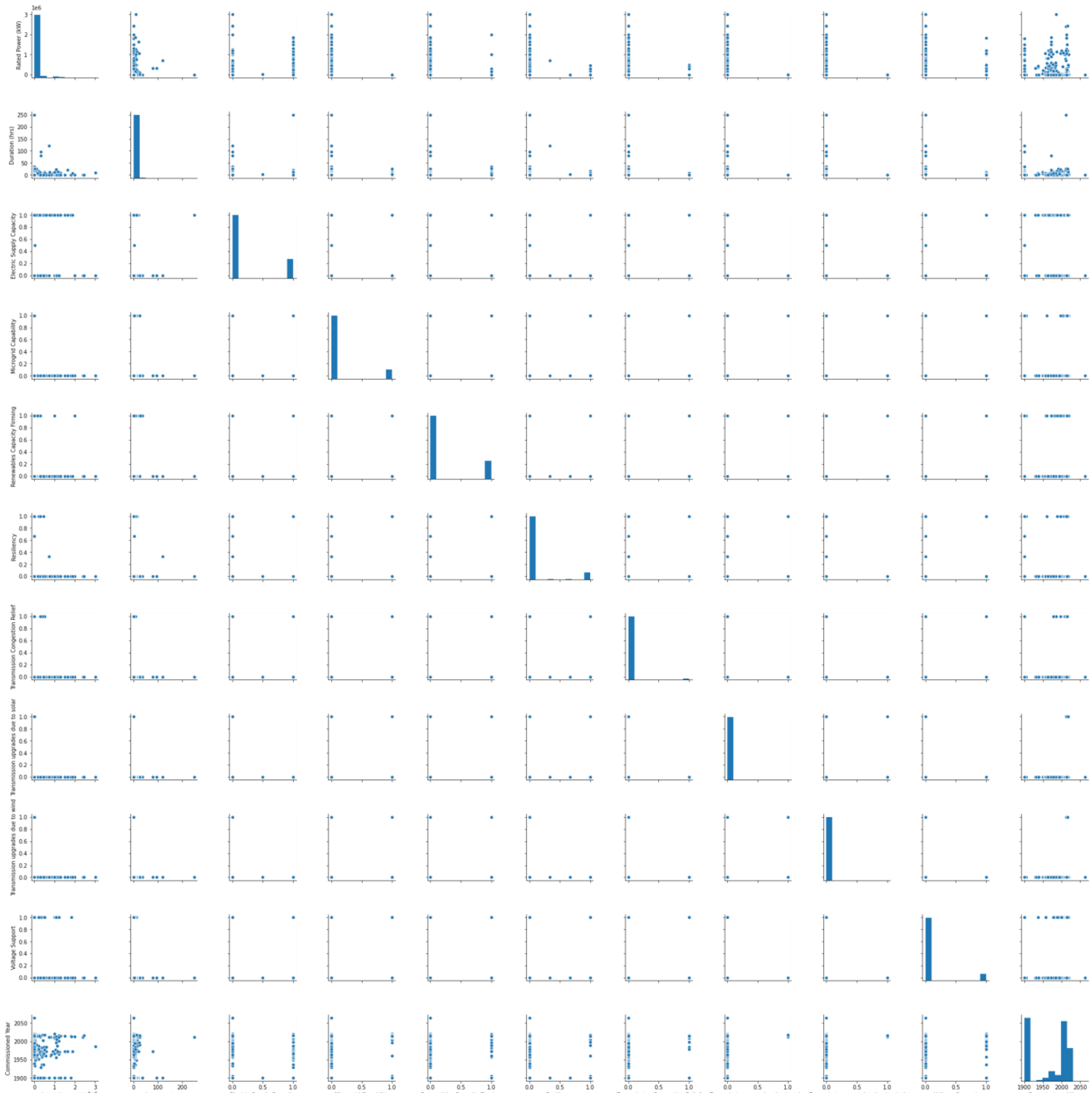



***Figure 4**: Correlation pair plots.*

Once the data was ready, the appropriate variables were assigned as the feature(s) and target variables. Next, in preparation for building the models, the dataset was dived into 70% training and 30% test. Reshaping of the training and test datasets were required before the models could be created. Lastly, depending on the results, each model was tuned and refined to achieve improvement of the outcome.

## IV. METRICS SELECTION

Feature selection was done through finding the correlation between each feature and target variable and creating a heatmap to identify the strength in relationship, which can be seen in **Figure 2** below. The higher correlations are represented with darker contrasts on the bar scale. Generally, a correlation of -1.0 shows a perfect negative correlation, while a correlation of 1.0 shows a perfect positive correlation. In this case, the correlation scale only ranges from -0.2 to 1.0, with 1.0 indicated as a deep red. It can be noticed that in **Figure 2**, the correlation of a variable with itself is 1.0 and for that reason all the diagonal values are 1.0.

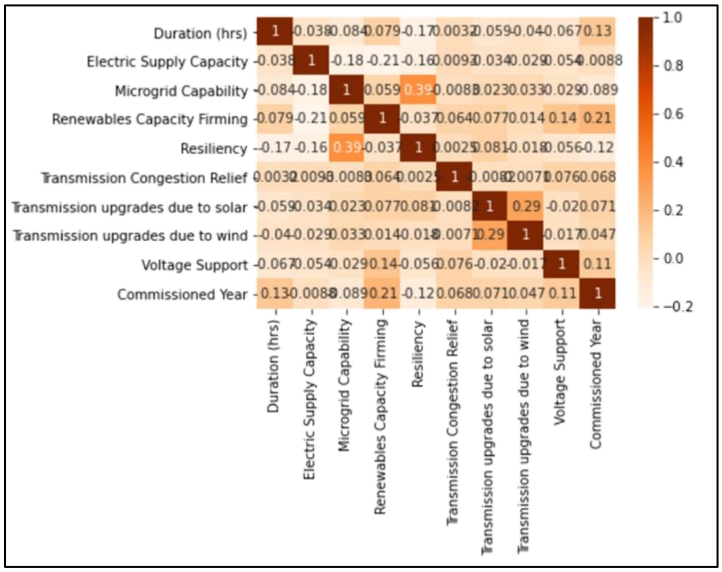



***Figure 5**: Visual heatmap to identify correlation between features and target value.*

For this dataset, the collinearity between features is considerably low with all correlations being below 0.40. However, the features with the strongest collinearity include Microgrid Capability and Resiliency with the highest correlation of 0.39. Next, we have features of Transmission upgrades due to solar and wind with the second highest correlation of 0.29. Following are the Renewables Capacity Firming and Commissioned Year (target variable) with a correlation of 0.2. Finally, although not significant, we have a small correlation between the Duration (hrs) feature and Commissioned Year (target value) of 0.13.

Since the Commissioned Year is the target value of interest, we can determine that the feature with the highest collinearity with the target variable is Renewables Capacity Firming which has a correlation of 0.21. In the energy and electric utility industry, Renewables Capacity Firming (also known as grid firming. nameplate capacity firming, or capacity firming) is widely used to address the gap due to intermittent wind, solar, and hydro-power resources to help meet grid code criteria and avoid outages or disruptions. It provides intermittent power output from renewable power generation plants, such as wind or solar, and can be maintained at a committed level for a period of time. During this period, the energy storage system smoothes the generation output and controls the ramp rate (MW/min) to eliminate any rapid voltage and power swings on the electrical grid. Understanding what this feature means supports using Renewable Capacity Firming as a selected feature variable to help answer this machine learning problem.

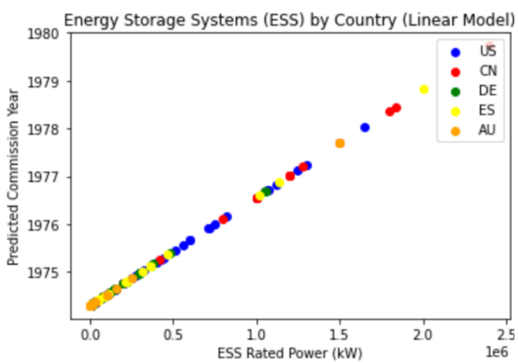
In addition, other features include Microgrid Capability, Resiliency, Transmission upgrades due to solar, Transmission upgrades due to wind and duration. Although Rated Power (kW) was feature with little to no correlation with Commissioned Year, I still want to consider it as a feature to see the results of the predictive analysis.

## V. IMPLEMENTATION/EXPERIMENT SETUP

The two models chosen are the linear regression and multilayer perceptron (MLP) regressor models. Since the goal of my project is to provide a predictive analysis for how the US is expected to continue to lead in this effort of integrating energy storage throughout the 21st century. The first model built was the linear regression model. This model was selected for its capability to work with larger datasets and linearly forecast predicted values. Secondly, the MLP regressor model was also selected for its known ability to perform well with large data sets, prediction analysis, and better accuracy.

## VI. RESULTS

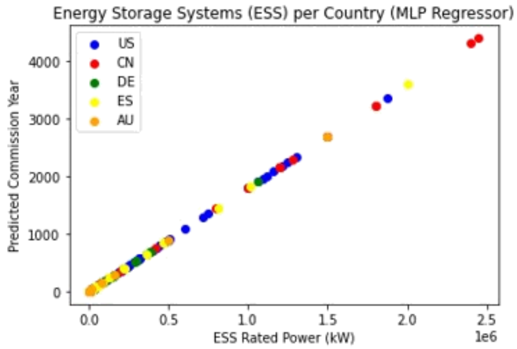
Evaluating the results of the linear regression model, it can be seen from the plot and table in **Figure 5** that there is a positive correlation between Rated Power (kW) and Commissioned Year. Though the correlation is positive, if we take a glance at the R2 scores (or coefficient of determination), we can see that both scores are considerably low for this model. This indicates that the fit for this model isn't the worst but may not be optimal for this problem,




| Description | Result |
|---|---|
| **Linear Training Error:** | 2958.40 |
| **Linear Test Error:** | 3197.44 |
| **Linear R2 Training Score:** | 0.40 |
| **Linear R2 Test Score:** | 0.06 |

***Figure 6**: Linear Regression model plot and results.*

Alternatively, the MLP regressor model, which I thought would perform better than the linear regression model, resulted in an even worse performance and fit. As it appears in **Figure 6**, the MLP regressor model, although indicating a positive correlation, results in a unacceptably poor R2 score of -1,299.75 and -1,217.06 for the training and test datasets, respectively.




| Description | Result |
|---|---|
| **MLP Training Error:** | 3,706,264.86 |
| **MLP Test Error:** | 3,670,741.00 |
| **MLP R2 Training Score:** | -1,299.75 |
| **MLP R2 Test Score:** | -1217.06 |

***Figure 7**: MLP Regressor plot and results.*

Overall, tuning and refining the linear regression model may help improve the fit and performance to better handle the data for which it is predicting. However, it is determined that the MLP regressor is not a good model to use and can be eliminated for this project.

## VII. CONCLUSION

In conclusion, the GESDB offers a solid database, however, more research and work need to be conducted to better prepare the data for building a more accurate machine learning algorithm to answer the proposed problem. Given the current datapoints we know that the US is unanimously the world leader in the number of ESS, holding 68.6% of those commissioned and announced. Following, in second place, is China at 9.5% with Germany (9%), Australia (6.7%), and Spain (6.2%) trailing closely behind.

As it relates to the performance of the models, it can be determined that, at face value, the linear regression model is a better fit compared to the MLP regressor. The linear regression shows a positive correlation between features but has a considerably low coefficient of determination. With additional tuning and refinement of the linear regression model, the model could be a better fit.

Overall, the China will integrate the largest number of grid scale capacity ESS with the US following their lead. Whereas the US is expected to continue to take the lead in the total number of ESS being integrated on the electric grid. Future work will still be required and may include looking further into each feature, their correlations to each other and how they can be more accurately represented and analyzed in future models.

# VIII. REFERENCES

Koenigs, C.; Suri, M.; Kreiter, A.; Elling, C.; Eagles, J.; Peterson, T.R.; Stephens, J.C.; Wilson, E.J. A Smarter Grid for Renewable Energy: Different States of Action. Challenges 2013, 4, 217-233. https://doi.org/10.3390/challe4020217

J. R. Aguero, E. Takayesu, D. Novosel and R. Masiello, "Modernizing the Grid: Challenges and Opportunities for a Sustainable Future," in IEEE Power and Energy Magazine, vol. 15, no. 3, pp. 74-83, May-June 2017, doi: 10.1109/MPE.2017.2660819.

Y. Zhang et al., "Grid-Level Application of Electrical Energy Storage: Example Use Cases in the United States and China," in IEEE Power and Energy Magazine, vol. 15, no. 5, pp. 51-58, Sept.-Oct. 2017, doi: 10.1109/MPE.2017.2708860.

Turki A. Alaqeel, Siddharth Suryanarayanan, A fuzzy Analytic Hierarchy Process algorithm to prioritize Smart Grid technologies for the Saudi electricity infrastructure, Sustainable Energy, Grids and Networks, Volume 13,2018,Pages 122-133,

Nekuda Malik, J. (2017). *"US government works toward grid modernization: https://gridmod.labworks.org"*. MRS Bulletin, 42(12), 874-875. doi:10.1557/mrs.2017.287