Forecasting the Future Power Consumption of Germany in Alteryx Designer using LSTM, SARIMA, and ETS

Mayank Kumar
Department of Philosophy
University of Tuebingen
mayank.kumar@student.uni-tuebingen.de

Swapnil Jha
Department of Electronics and Communication
Maharaja Surajmal Institute of Technology
swapniljha001@gmail.com

Vikramaditya Malik
Department of Electronics and Communication
Maharaja Surajmal Institute of Technology
vikramaditya1999@gmail.com

***Abstract*—Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. Traditional time series forecasting methods such as Exponential smoothing and ARIMA with proper implementation are guaranteed to work but with recent advancements in computing capabilities, Deep Learning based forecasting techniques have significantly improved the accuracy of forecasting compared to the traditional methods. Motivated by the above, in this research we try to predict the future power consumption of the country of the Federal Republic of Germany, using traditional time series forecasting algorithms such as Seasonal Auto Regressive Integrated Moving Averages, Exponential Time Smoothening, and Deep-Learning based Recurrent Neural Network backed Long Short-Term Memory and does a comparative analysis of the predictions generated by the 3 models.**

***Keywords*—time series forecasting, power consumption, error trend seasonality, auto regressive integrated moving averages, long short-term memory networks, feedback loops**

I. Introduction

Date Time series forecasting generates insights which contribute to complex decision-making, by estimating the evolution of the given metric in the near future, which in this case is the power consumption of a country [1], this forecasting system plays a crucial role in various departments such as Power Generation [2], Importing/Exporting of Power [3], Economic Ministries [4], and dictates the self-reliance [5] of the country in the future. It can play an irreplaceable role in domains such as healthcare [6], energy management [7], and most importantly financial investments [8] to name a few.

A. Traditional Time Series Forecasting Algorithms

Traditional time series forecasting methods such as Autoregressive Integrated Moving Average (ARIMA) model [9] and Holt's Winter seasonal method [10] are generally very well understood and are theoretically guaranteed to work given that they have been properly implemented on a compatible dataset. They are mainly applicable for univariate forecasting problems which restricts their applications to many real-world complicated time series data.

B. Deep Learning Time Series Forecasting Algorithms

With the increase in computing power in the recent years, it has been shown that deep learning-based time series forecasting techniques can achieve much higher prediction accuracy than traditional approaches [11].
There are many kinds of deep neural networks being used for sequence modelling in the literary domain, and the same are applied in time series forecasting techniques [12]
While most seq2seq methods have been successfully used in many real-world data forecasting problems, they rely on generalized sequences and disregard the fact that time series is a special type of sequential data. One of the major differences is that downsampling of time series data often preserves most of the information in the data, whereas this is definitely not true for general sequential data such as textual or DNA sequence.
In fact, unlike other types of sequence data, three components characterize any time series sequence: Trend, Seasonality, and Errors (irregular components), and the former two components allow us to perform reasonable forecasting into the future with an acceptable rate of accuracy.
Motivated by the above facts, in this paper, we try to do compare exactly how well does the new Deep-Learning based Recurrent Neural Network backed Long Short-Term Memory date-time forecasting technique performs against the traditional ARIMA and ETS models on the given dataset of power consumption by the state of Bundesrepublik Deutschland by predicting the future power consumption.
The main contributions of this research are as follows:

- We created a model which can forecast the future of power consumption with an error of only 0.09 %
- We demonstrated how well different models both traditional and deep learning perform on the same dataset.

- We can use this to predict which countries might need to buy or invest in energy resources in the upcoming future.
- This research paves the path for potential research which can be done to predict more metrics such as energy surplus/deficit and imbalance in the power economics.

II. Long Short Term Memory

Long Short Term Memory networks are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter & Schmidhuber (1997) [13], They work tremendously well on a large variety of problems, unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only single data points, but also entire sequences of data.[14]

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior.

All recurrent neural networks have the form of a chain of repeating modules of neural network. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

LSTM networks are well-suited to classifying, processing and, making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. [15] LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs. [16]

III. ARIMA

An autoregressive integrated moving average, or ARIMA, is a type of statistical analysis model.

An autoregressive integrated moving average model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables.[17] The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

ARIMA which stands for Autoregressive Integrated Moving Average helps you forecast data for seasonal and nonseasonal data.

Nonseasonal ARIMA models are displayed in the terms (p,d,q) which stand for p - periods to lag for, d - number of transformations used to make the data stationary, q - lags of the error component.

- Stationary - mean and variance are constant over time vs Non-Stationary - mean and variance change over time
- Differencing - take the value in the current period and subtract it by the value from the previous period. You might have to do this several times to make the data stationary. This is the Integrated component which is d in the model terms.
- Autocorrelation - How correlated a time series is with its past values, if positive at Lag-1 then AR if negative then MA,
- Partial Autocorrelation - The correlation between 2 variables controlling for the values of another set of variables. If the partial autocorrelation drops off quickly then AR terms, if it slowly decays then MA

Seasonal ARIMA models are denoted (p, d, q)(P,D,Q)m. These models may require seasonal differencing in addition to non-seasonal differencing. Seasonal differencing is when you subtract the value from a year previous of the current value.

IV. ETS

Exponential smoothing is a technique for smoothing time series data using the exponential function. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. [18] It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality.

For simple exponential smoothing methods, the forecast is calculated by multiplying past values by relative weights, which are calculated based upon what is termed a smoothing parameter. This is the magnitude of the weight applied to the previous values, with the weights decreasing exponentially as the observations get older. [19] The formula for Forecast looks like this:

$$F = W_t Y_t + W_{t-1} Y_{t-1} + W_{t-2} Y_{t-2} + \dots + (1-\alpha)^n Y_n \qquad (1)$$

where,

t is the number of time periods before the most recent period (e.g., t = 0 for the most recent time period, t = 1 for the time period before that).

Y_t is the actual value of the time series in period t

$$W_t = \alpha(1-\alpha)^t \qquad (2)$$

where n = the total number of time periods

This model gives us a smooth line or LEVEL in our forecast that we can use to forecast the next period. Exponential smoothing is one of many window functions commonly applied to smooth data in signal processing, acting as low-pass filters to remove high-frequency noise.

V. Alteryx Designer

Alteryx Designer is a software suite that brings together all the aspects of data exploration in a single user-friendly graphical interface with hundreds of different tools for all stages of CRISP-DM[20] such as data preparation, and data blending[21], and analytical data exploration. One can utilize the aforementioned tools to create a workflow that can retrieve input from multiple sources in real-time. Workflows avail the researcher to achieve a coherent sense of documentation alongside the actual process instead of having a separate documentation report file. It would help in ensuring that the documentation is amended just as customarily as the workflow. Other salient features of Alteryx include but are not limited to live corroboration and collaboration.

One can read data from a plethora of different data sources and can even blend and manipulate on-premise and cloud data onto a single workflow. The Alteryx Server provides a venue that can be scaled according to the industry needs and usage of APIs and macros integrate into any other analytical application of choice.

A. Features of Alteryx

Alteryx uses dedicated memory data processing to enhance the efficiency in the handling of the data. This dedicated memory data processing can be used to harness the features of data warehousing without the use and technical aptitude of SQL, id est, no-code solutions for data warehousing, and data manipulation for generating business and analytical insights.

Alteryx Designer can also be used to deploy existing R and Python models directly into production workflows, enabling business teams to run analytics on live streams of data.

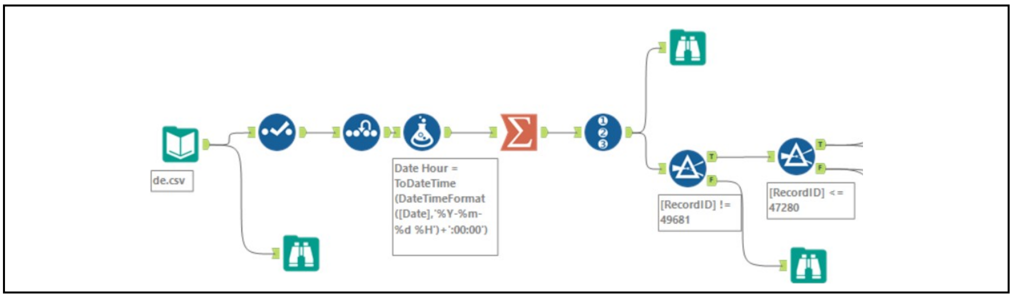

Fig. 1. Workflow for Data Preprocessing

B. Tools used in this Research [22]

- Input Data Tool - The Input Data tool can parse over a dozen different types of file formats from CSV to XML and everything in between. It can also be used to connect to a cloud server database. We employed this tool to construe the electricity consumption data from a CSV file format.
- Browse Tool - The Browse tool is a convenient tool that can execute data profiling and display it in a result window, after acquiring it from the previous tool. This tool can also be used to display graphs or other data visualization assets.
- Select Tool - The Select tool is a versatile tool that can be purposed to reorder columns, select columns, change data types, names, and many other useful things. It was called to remove 'end' and to change the data type of 'start'.
- Imputation Tool - The imputation tool as the term "Impute" would suggest is adopted for replacing missing and null values such as NaNs with the mean, median, or mode, of the specified column or even a user-specified value. This was used to impute the 7 missing values in the 'load' column using the Median value.
- Formula Tool - The formula tool is used to perform operations and calculations on an existing column or create a new column. We used this to refine the format of the 'start' column to make the machine learning more efficient.
- Summarize Tool - The Summarize tool can perform a surplus of aggregation functions such as calculating the summary, by summing or finding the minimum or maximum, grouping, counting, string concatenating, math functions, spatial object processing, and much more. This was used to convert the frequency of the dataset from 15 minutes to an hour.
- Record ID Tool - The record ID tool is used to assign a unique ID to each row in a dataset. This is especially useful when we need to maintain the linearity of the data.
- Filter Tool - The filter tool splits the data into two streams based on a conditional expression. We used this to create two datasets for training and testing.

VI. Dataset

ENTSO-E Transparency Platform gives access to electricity generation, transportation, and consumption data for the pan-European market. [23] Part of this data is provided by the various transmission system operators (TSOs) across Europe. The consumption is given in Megawatts (MW).

The dimensions of this dataset are 3x198,721 with the 3 columns in the dataset containing the start time (`start`), end time (`end`), and the number of Megawatts consumed in that given period of time (`load`), and the 198,721 records of data are provided to us in intervals of 15 minutes from January 2015 to August 2020. We proceed with checking for missing data and find 7 datapoints with missing values in the field of Power Consumption, we subsequently replace these missing records with the median value of the load from the dataset to minimize the noise produced by missing data. We remove the end column since that is not required as both start and end columns refer to the same period in time. We then change the data type of start column from string to DateTime, in order to make it more usable.

Furthermore, we create a new dataset of power consumption with the data frequency being 1 hour rather than 15 minutes. This brings the number of records in our feature matrix from 198,714 to 49,681, and then we use this new dataset to train our time series forecasting machine learning models. We create a new column called `Record

ID`, this is done because unlike other types of machine learning datasets, we cannot shuffle the data to create training and testing splits, we need to split while maintaining the linearity of the data. Thus, we use Record ID, to achieve that goal. We then proceed to divide the dataset into 2 parts from record 1 to 47,280 for training and from 47,281 to 49,681 for testing. We subsequently train the 3 models (ETS, SARIMA, and LSTM) using this training dataset and compare their outputs with the testing dataset to calculate how well the models predicted the power consumption for Deutschland for the given time period in the testing dataset.

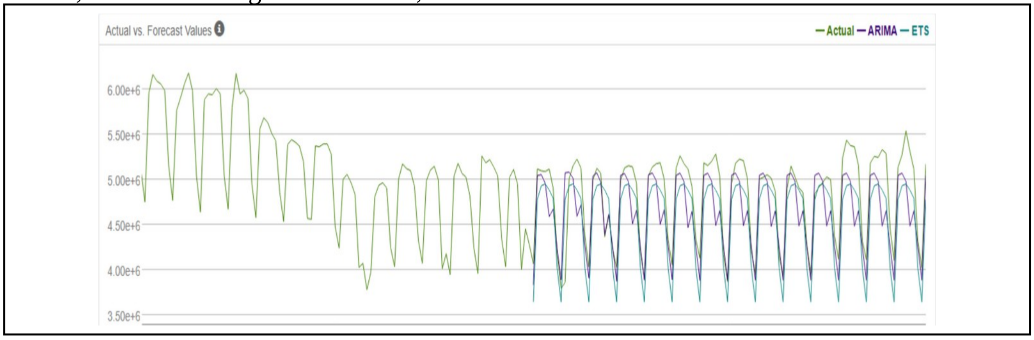

Fig. 2. Comparison chart of forecasts generated by SARIMA and ETS.

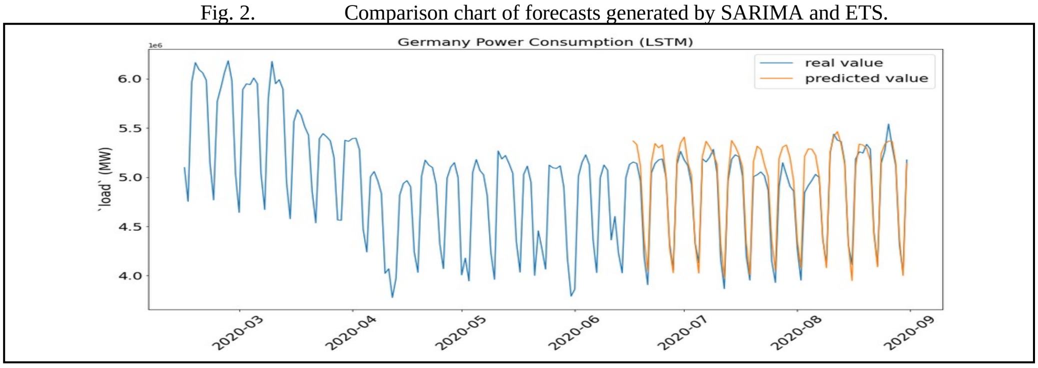

Fig. 3. LSTM Forecast along with true values.

TABLE I. ACCURACY MEASURES

Accuracy Measures:						
Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS(M, A, M)	20922.36	31400.23	26364.47	9.05	12.38	3.36
SARIMA(5,0,0)(2,1,0)[24]	20503.75	31129.03	25787.98	8.90	12.04	3.29
LSTM	02676.53	24358.74	16052.54	1.34	0.085	2.05

VII. RESULT

The result of this research is the following model.

- LSTM Model with 0.09% prediction error with configuration of [120, 80, 40, Dense].

Figure 2 shows that the ETS model is consistently underestimating the power consumption by under forecasting the amount. SARIMA is doing much better than ETS and has a lesser error margin, but it is still not totally accurate.

Figure 3 shows the LSTM forecasting model [24], as we can see, this model is head and shoulders ahead of its traditional competitors and is making much better forecasts with smaller than ever error margins.

The total Power consumption forecast of Germany for 100 days is 481.9 Terawatt.

VIII. CONCLUSION

Deep learning-based forecasting methods show promising forecasting results. In this study, we compared the results obtained of Long Short-Term Memory, a deep learning-based time series forecasting technique against traditional forecasting methods such as SARIMA [25] and ETS.

Each model is used to predict the future power consumption of Germany over a period of 100 days. Our thorough research on different date-time series forecasting techniques, with major emphasis on comparison of traditional and modern deep-learning based forecasting algorithms on a single dataset, concludes that LSTM [26] outperforms traditional approaches even on a univariate dataset, which makes it a suitable candidate for further studies to be conducted on.

This Deep learning based Recurrent Neural Network backed Long Short-Term Memory has performed over the realm of expectations we had from this algorithm and has managed to reduce the errors generated from the Manhattan Norm of distances or L1 Norm of distances, such as MAE and MAPE to a mere 2676 and 0.09% respectively.

This model can be used to measure and forecast the future power consumption of any state provided that a sufficiently sized existing training data is provided in order to make the forecast of the required period of time with an acceptable rate of error, being approximately equal to 0.09%.

Acknowledgment

This research paper and the study behind it would not have been possible without the exceptional support of our supervisor, Dr. Jasmine Chhikara. Her enthusiasm, knowledge and exacting attention to detail have been an inspiration and kept our work on track from our first encounter with data cleaning and manipulating tools like Alteryx to the final draft of this paper. Shreyansh Bhura and Aman Prakash, our fellow university mates at the Guru Gobind Singh Indraprastha University, have also looked over our transcriptions and with unfailing patience answered numerous questions about the processes with which we can optimize this comparative analysis. Sepp Hochreiter and Jürgen Schmidhuber, the main authors of the research paper behind the fantastical idea of Long Short Term Memory neural networks; shared the invaluable information on the algorithm that they had been gathering for years. We are also grateful for the insightful comments offered by the anonymous peer reviewers at IEEE. The generosity and expertise of one and all have improved this study in innumerable ways and saved us from many errors; those that inevitably remain are entirely our own responsibility.

References

[1] Yan, K., Wang, X., Du, Y., Jin, N., Huang, H., & Zhou, H. 2018. Multi-step short-term power consumption forecasting with a hybrid deep learning strategy. Energies, 11(11), 3089.

[2] Hinton, K., Baliga, J., Feng, M., Ayre, R., & Tucker, R. S. 2011. Power consumption and energy efficiency in the internet. IEEE Network, 25(2), 6-12.

[3] Kaboli, S. H. A., Fallahpour, A., Kazemi, N., Selvaraj, J., & Rahim, N. A. 2016. An expression-driven approach for long-term electric power consumption forecasting. American Journal of Data Mining and Knowledge Discovery, 1(1), 16-28.

[4] Young, D., Scharp, R., & Cabezas, H. 2000. The waste reduction (WAR) algorithm: environmental impacts, energy consumption, and engineering economics. Waste management, 20(8), 605-615.

[5] Rey-Moreno, C., Sabiescu, A. G., & Siya, M. J. 2014. Towards self-sustaining community networks in rural areas of developing countries: Understanding local ownership. In Proceedings of the 8th International Development Informatics Association Conference (pp. 63-77).

[6] Bahadori, M. T.; and Lipton, Z. C. 2019. Temporal-Clustering Invariance in Irregular Healthcare Time Series. ArXiv, abs/1904.12206.

[7] Zhou, H.-Y.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In AAAI.

[8] D'Urso, P.; Giovanni, L.; and Massari, R. 2019. Trimmed fuzzy clustering of financial time series based on dynamic time warping. Annals of Operations Research, 299: 1379–1395.

[9] Box, G.; Jenkins, G.; and Macgregor, J. 1968. Some Recent Advances in Forecasting and Control. Journal of The Royal Statistical Society Series C-applied Statistics, 17: 158–179.

[10] Holt, C. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting, 20: 5–10.

[11] Oreshkin, B.; Carpo, D.; Chapados, N.; and Bengio, Y. 2020. NBEATS: Neural basis expansion analysis for interpretable time series forecasting. In ICLR.

[12] Lim, B.; and Zohren, S. 2021. Time-series forecasting with deep learning: a survey. Philosophical Trans. of the Royal Society A.

[13] Hochreiter, Sepp & Schmidhuber, Jürgen. 1997. Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

[14] Gers, Felix A., Nicol N. Schraudolph, and Jürgen Schmidhuber. 3.Aug 2002 "Learning precise timing with LSTM recurrent networks." Journal of machine learning research: 115-14.

[15] Sherstinsky, A. 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.

[16] Box, George E. P. 2015. Time Series Analysis: Forecasting and Control. WILEY. ISBN 978-1-118-67502-1.

[17] Zhang, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

[18] Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. 2008. Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media.

[19] Billah, B., King, M. L., Snyder, R. D., & Koehler, A. B. 2006. Exponential smoothing model selection for forecasting. International journal of forecasting, 22(2), 239-247.

[20] Wirth, R. and Hipp, J., 2000, April. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1, pp. 29-40).

[21] Taranto-Vera, G. *et al.* (2021) 'Algorithms and software for data mining and machine learning: a critical comparative view from a systematic review of the literature', *Journal of Supercomputing*, 77(10), pp. 11481–11513. DOI: 10.1007/s11227-021-03708-5.

[22] https://help.alteryx.com/20221/designer/designer-tools-list

[23] Hirth, L., Mühlenpfordt, J., & Bulkeley, M. 2018. The ENTSO-E Transparency Platform–A review of Europe's most ambitious electricity data platform. Applied energy, 225, 1054-1067.

[24] Ibrahim Kovan 2021 Forecasting the Future Power Consumption of Germany using LSTM(RNN) and DNN. Towards Data Science.

[25] Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. 2003. ARIMA models to predict next-day electricity prices. IEEE transactions on power systems, 18(3), 1014-1020.

[26] Elsworth, S., & Güttel, S. 2020. Time series forecasting using LSTM networks: a symbolic approach. arXiv preprint