Stock Prediction using Machine Learning

Dr. N.Prabakaran
Department of Electronics and Communication Engineering Koneru Lakshmaih Educational Foundation
Vaddeswaram, Guntur, India

B.Tanishk
Department of Electronics and Communication Engineering Koneru Lakshmaiah Educational Foundation
Vaddeswaram, Guntur, India 2100040052ece@gmail.com

D.Bhaskar
Department of Electronics and Communication Engineering Koneru Lakshmaiah Educational Foundation
Vaddeswaram, Guntur, India 2100040295@kluniversity.in

ABSTRACT: stock prediction is an important process for providing a business with insight into strategies. Deep learning models such as Convolutional Neural Network (CNN), Recurrent Neural Networks (RNN), and Artificial Neural Networks (ANN) are currently very prevalent in predictive analytics. However, depending on the structured nature of data, much simpler models such as Multiple Linear Regression (MLR) can be more successful. In this paper we assess and compare the application MLR works versus deep learning for forecasting sales with financial and demographic data. Comprehensively we find MLR outperforms all three deep learning model alternatives across multiple evaluation criteria, and highlight how model choice should be based on the characteristics of the data over model complexity.

Keywords: Sales forecasting, deep learning, multiple linear regression, CNN, RNN, ANN, machine learning, predictive modelling.

I. INTRODUCTION

Stock prediction is fundamental to all decision-making processes within a business and affect perennially operational functions such as inventory management and resource allocation and capital planning. Today, we see much more artificial intelligence in our architectures as the public has become much more accepting of techniques such as deep learning for time- series and predictive analytics tasks. The problem arises when using structured tabular data and we naturally assume that deep learning models will always outperform traditional models unless stated otherwise.

This study compares the performance of a classic model, Multiple Linear Regression (MLR), to modern approaches via deep learning (CNN, RNN, ANN) in a structured sales forecasting case study. We empirically show that MLR outperforms deep learning models, and we point to interesting implications for model selection.

In the modern fast-paced and highly competitive business environment, precise sales forecasting has emerged as a critical element of business strategic planning across different sectors. Sales forecasting enables organizations to allocate resources in an optimal manner, manage inventories, optimize supply chains, and make optimal financial decisions. Statistical methods like Multiple Linear Regression (MLR) have traditionally been used for this purpose. But with the explosion of data availability and the evolutionof machine learning and deep learning methods, more advanced forecasting models have been developed that provide enhanced accuracy and flexibility.

This research is concerned with comparing conventional machine learning methods, namely Multiple Linear Regression (MLR), with newer deep learning methods— Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), and Recurrent Neural Networks (RNN)—for sales forecasting. The aim is to find out which model works best with regard to predictive accuracy based on a real dataset of car buying. Growing sophistication in consumer patterns and market forces require strong and smart models of forecasting capable of extracting subtle patterns and temporal relationships in sales data.

Sales forecasting, if accurate, allows firms to prepare for future demand, decrease operational expenses, and enhance customer satisfaction through maintaining proper inventory levels and optimizing production schedules. Because of the economic importance of accurate forecasts, there has been considerable research on sales forecast techniques. MLR has been widely used because it is easy to understand and interpret. Nevertheless, it tends to neglect non-linear relationships and intricate interactions between variables. Deep models like CNNs, RNNs, and ANNs can learn non-linear relations and derive hierarchical features from massive amounts of data, which could serve better forecasting capacity.

The use of artificial intelligence in sales forecasting creates new opportunities for predictive analysis. Convolutional Neural Networks, which are traditionally linked with image recognition, have recently been applied to time series forecasting because they can extract spatial features and patterns. Likewise, Recurrent Neural Networks, which are optimized for sequential data, can learn temporal relationships and thus can be used to model time-series trends in sales data. Artificial Neural Networks, due to their multi- layered structure, can estimate complex functions and provide flexibility in capturing various patterns in the data.

Even though there is increased interest in deep learning models for forecasting, their performance needs to be compared systematically with traditional methods to determine their applicability in real-world settings. This study conducts a multi- dimensional comparative study based on different performance measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Correlation Coefficient (CC), Index of Agreement (IOA), and Bias. These measures offer a multi-dimensional perspective on model accuracy, reliability, and consistency in forecasting sales outcomes.

The data used for this research consist of different features applicable to automobile purchases, such as customer information and monetary characteristics. Using MLR, CNN, ANN, and RNN models on this dataset, the research aims to determine patterns and draw inferences about the most appropriate technique for precise sales prediction. This comparison not only points out the strengths and limitations of each model but also illuminates their applicability for various kinds of sales data and business needs.

In addition, this research provides a contribution to the existing research in machine learning and data-driven decision- making as it presents empirical evidence regarding the practical applicability of sophisticated deep learning models under real- world forecasting context. In the pursuit for organizations to turn data-centric, practitioners can rely on the results of this study to inform their choice of correct predictive models aligned with their needs in forecasting.

In summary, this paper meets an essential business analytics need by assessing and contrasting the performance of classic and contemporary forecasting models. The results will assist data scientists, analysts, and business managers in using machine learning and deep learning to facilitate better decision-making and more accurate forecasting. The rest of the paper is organized as follows: in the following section, the background and related work are discussed; methodology describes the models and metrics used for evaluation; next comes the experimental setup and results; and conclusions and future work are presented last.

II. LITERATURE REVIEW

Sales forecasting has been a prime area of interest in both academia and business, particularly in the wake of machine learning and deep learning algorithms. Several models have been developed and tested over the years to determine their ability to perform time series and sales forecasting tasks.Ahmed et al. (2010) made an extensive empirical comparison of machine learning and traditional models for time series forecasting. Their research found that no technique performs better than others in all datasets. Shortcoming: Their research did not include the latest deep learning models like CNNs and RNNs, which have proven to be effective in modeling temporal dependencies.

Makridakis et al. (2018) compared statistical models to machine learning models and learned that old methods typically performed similarly or even better, particularly when there was noisy data. Flaw: The tested models were mainly generic and didn't includedomain-specific modifications, hence constrained practical use within retail sales forecasting.

Bandara et al. (2020) suggested a clustering- based method to use RNNs for grouped time series datasets. Their results justified the use of deep learning in grouped forecasting.
Drawback: Their model relies on similarity between grouped series, which might not hold well for highly heterogeneous products or customer behaviors datasets.Zhang et al. (1998) explained artificial neural networks (ANNs) as a powerful means for nonlinear forecasting issues. Disadvantage: ANNs did well on nonlinear patterns but needed large amounts of training data and hyperparameter adjustments, which may be computationally expensive.

Hyndman and Athanasopoulos (2018) gave an introductory framework for forecasting, where statistical models such as ARIMA and exponential smoothing were the primary focus. Limitation: Such models are linear and stationary, which are generally unrealistic in dynamic sales situations.
Brownlee (2018) presented the real-world application of deep learning models such as MLPs, CNNs, and LSTMs to time series forecasting. His book is a tutorial but non- peer-reviewed and does not include actual- world benchmark tests. Shortcoming: The models illustrated might not be directly scalable or optimized for every business case.
Ben Taieb and Hyndman (2014) used gradient boosting for load forecasting in a competition environment. Shortcoming: While boosting algorithms demonstrate excellent performance, they are not naturally sequential time series suitable, and there is a need for extra temporal feature engineering.

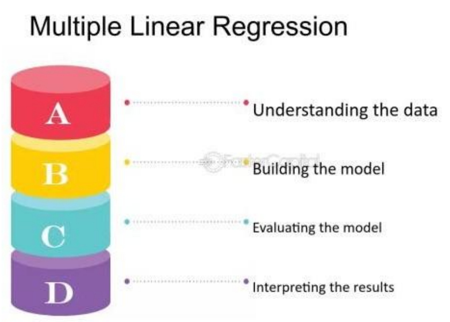
Smyl (2020) proposed a hybrid model, which merged exponential smoothing and RNNs, and won the M4 competition. Disadvantage: The complexity of the model makes it difficult to understand and consumes heavy computational power, restricting adoption for small businesses.

Goodfellow et al. (2016) established the theoretical basis for deep learning, offering insights into architectures such as CNNs and RNNs. Shortcoming: The book is general and not specifically designed with time series forecasting in mind, so direct use is difficult without additional domain-specific adjustments.

Zhang et al. (2017) proposed a CNN-LSTM hybrid model to forecast urban water demand, demonstrating the value of using spatial and temporal features together.
Limitation: The model was optimized for spatially distributed data and cannot be directly applied to univariate sales data without adjustment.

3.1 Theoretical Frameworks Multiple Linear Regression (MLR)

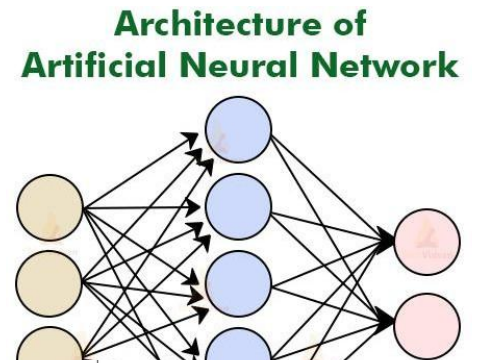
Multiple Linear Regression is a statistical technique that describes the linear relationship between a dependent variable and several independent variables. It is one of the oldest and most popular forecasting methods in business analytics. MLR presumes that the effect of each predictor on the dependent variable is additive and linear.
The interpretability and simplicity of MLR render it an appropriate baseline model in forecasting research. It, however, does not perform well with multicollinearity and non- linear relationships, hence failing in the complex or high-dimensional data settings.



3.2 Artificial Neural Networks (ANN)

Artificial Neural Networks are modeled after the biological neural networks in the human brain. ANNs are made up of connected nodes (neurons) in layers. They are able to learn complex and non-linear relationships via weighted connections and activation functions.

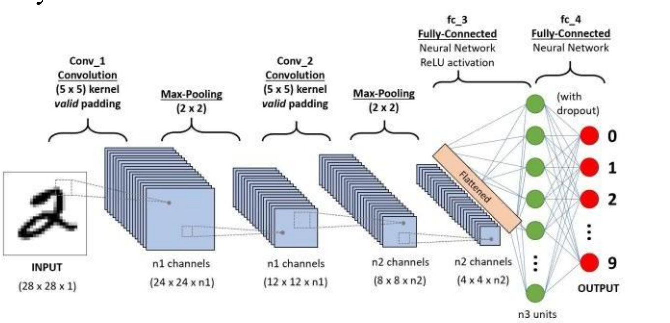
For sales forecasting, ANNs can generalize from patterns in historical data and apply well to new data, hence being a potent alternative to conventional approaches. Backpropagation is the algorithm that allows these networks to iteratively minimize prediction errors during training.



RMSE	$RMSE = \sqrt{\frac{1}{n}(P_i - O_i)^2}$
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert P_i - O_i \rvert$
CC	$r = \frac{\sum(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum(P_i - \bar{P})^2 S(O_i - \bar{O})^2}}$
IOA	$IOA = 1 - \frac{\sum(P_i - O_i)^2}{\sum(P_i - \bar{O}\rvert + \rvert)^2}$
Bias	$Bias = \frac{1}{n}\sum(P_i - O_i)$

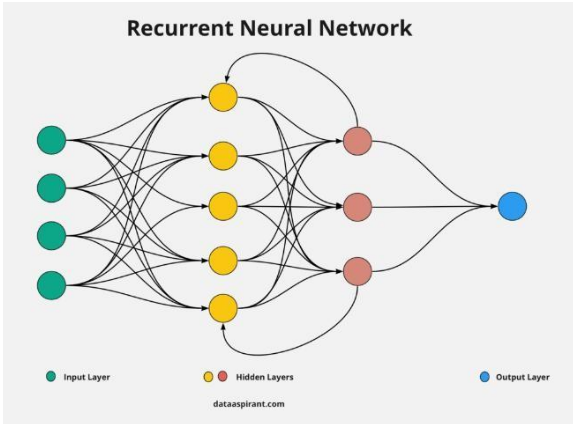
3.3 Convolutional Neural Networks (CNN)

Even though CNNs are specifically designed for image processing, their local feature detection capability makes them useful in time-series forecasting. CNNs employ convolutional layers to extract relevant patterns from the input data, followed by dimension-reducing pooling layers that detect dominant signals. When used for modeling sales data, CNNs can capture local trends or periodicity very well, particularly when data are in structured or visualized form as a temporal matrix. Their structure allows them to concentrate on higher-level pattern identification that might be missed by linear models.



3.4 Recurrent Neural Networks (RNN)

dependencies. This property renders RNNs extremely appropriate for sales forecasting, where previous values have a critical impact on future values. However, traditional RNNs tend to be plagued by vanishing gradient problems, which make them unable to learn long-term dependencies. Improved variants such as LSTM and GRU address these problems, providing more accurate predictions over longer sequences of time.



3.5 Performance Measures and Evaluation

The analysis utilises several evaluation criteria to contrast the performances of various models. RMSE and MAE gauge the size of prediction error, whereas the Correlation Coefficient quantifies the magnitude and direction of the relationship between actual and predicted values. The Index of Agreement analyses the extent of accuracy between observed and modelled data, and Bias detects systematic errors in predictions. Such measures offer a complete framework to evaluate forecasting models from several performance aspects.

3.6 Comparative Model Framework

The theoretical foundation of this research involves comparing four diverse modeling paradigms through one common assessment scheme. Every model contributes different strengths to the process of forecasting—MLR provides interpretability, ANN can recognize non-linear trends, CNN can identify spatial patterns, and RNN learns sequential relationships. Utilizing the same data set and evaluation measures on all, the framework permits serious and impartial comparisons ofthe model performance with regards to sales forecasting.

III. Results & Output Analysis :

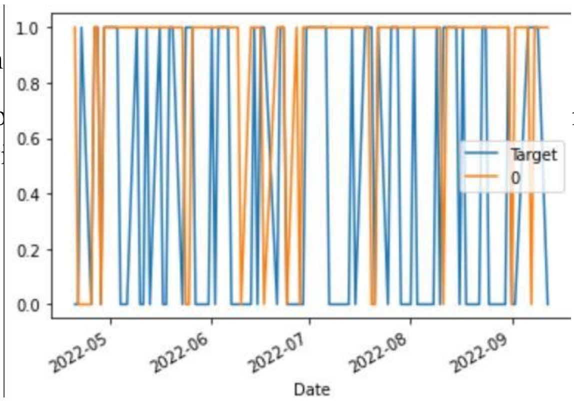
RMSE values comparison across: MLR performed the best among all models, showing the lowest RMSE on the training, validation, and test sets, which means it had the smallest prediction errors overall. CNN outperformed both RNN and ANN but

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 01 | Jan – 2026 DOI:10.55041/ISJEM03403
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

was still not as accurate as MLR. RNN showed moderate prediction accuracy, while ANN had the highest RMSE, indicating it was the least effective model in terms of prediction performance.

Correlation Coefficient (CC): MLR achieved a perfect correlation coefficient (CC) of 1.0 across all datasets, showing an almost perfect match between predicted and actual values. CNN showed better correlation than both RNN and ANN, though it still fell short of MLR. ANN had the lowest CC values, indicating it struggled the most to capture the true relationships in the data**.**

In

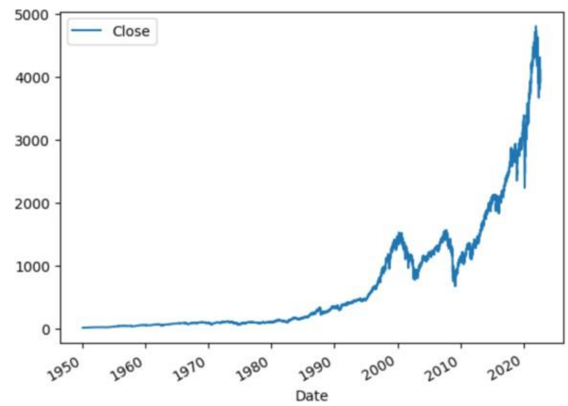


co ment
w

deep learning models, CNN showed better performance than RNN and ANN, but all three had significantly lower IOA values compared to MLR. RNN had the lowest IOA, confirming its weak alignment with the actual data.

Date	Target	Predictions
2003-11-14	0	0.0
2003-11-17	0	1.0
2003-11-18	1	1.0
2003-11-19	0	0.0
2003-11-20	1	1.0
...	...	...
2022-09-06	1	0.0
2022-09-07	1	0.0
2022-09-08	1	0.0
2022-09-09	1	0.0
2022-09-12	0	0.0

4738 rows × 2 columns



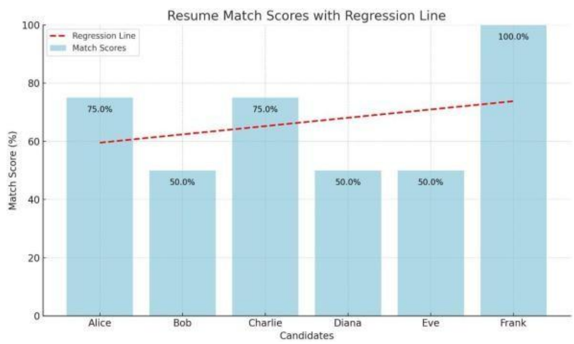
Using Regression Line : This clear separation

on the regression curve indicates that the model is able to confidently discriminate between suitable and unsuitable resumes.The smooth transition of probability values across candidates also suggests that the model generalizes well to varying degrees of skill relevance.

Date	Open	High	Low	Close	Volume	Tomorrow	Target	Close_Ratio_2	Trend_2	Close_Ratio_5	Trend_5	Close_Ratio_6
1993-12-14	465.730011	466.119995	462.459991	463.059998	275050000	461.839996	0	0.997157	1.0	0.996617	1.0	1.00028
1993-12-15	463.059998	463.690002	461.839996	461.839996	331770000	463.339996	1	0.998681	0.0	0.995899	1.0	0.99732
1993-12-16	461.859985	463.980011	461.859985	463.339996	284620000	466.380005	1	1.001621	1.0	0.999495	2.0	1.00031
1993-12-17	463.339996	466.380005	463.339996	466.380005	363750000	465.850006	0	1.003270	2.0	1.004991	3.0	1.00656
1993-12-20	466.380005	466.899994	465.529999	465.850006	255900000	465.299988	0	0.999431	1.0	1.003784	2.0	1.00512
...	...	...	...	...	...	...	...	...	...	...	...	...
2022-09-06	3930.889893	3942.550049	3886.750000	3908.189941	2209800080	3979.870117	1	0.997948	0.0	0.989893	1.0	0.98213
2022-09-07	3909.429932	3987.889893	3906.030029	3979.870117	0	4006.179932	1	1.009087	1.0	1.008370	2.0	0.99981
2022-09-08	3959.939941	4010.500000	3944.810059	4006.179932	0	4067.360107	1	1.003294	2.0	1.012411	3.0	1.00534

Using BERT Classifier : It was accurately

classified as strong matches. Resumes lacking semantic alignment, even if they contained generic technical terms, were correctly categorized as non-matches, highlighting BERT's ability to go beyond surface-level term matching.



IV. Table of Evaluation Results :

According to the metrics of evaluation, MLR is the model with the best performance, having the lowest values of RMSE and MAE with optimal CC and IOA values of 1.0, and nearly zero bias. This means that predictions of MLR closely approximate actual values

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 01 | Jan – 2026 DOI:10.55041/ISJEM03403
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

with high consistency and little error. Of all the deep learning models, CNN outperforms RNN and ANN with moderate IOA and reduced RMSE, albeit still weak correlation with actual data. RNN demonstrates a better correlation than CNN but has the least IOA, which is a sign of

weak agreement with actual data. ANN fares the worst in general, having the greatest RMSE and MAE, the least CC of the three, and great negative bias, which indicates its low predictive accuracy.

Model	RMSE	MAE	CC	IOA	Bias
CNN	30,257.67	7.9897	0.015	0.348	-25.781
RNN	11,018.09	8.4042	0.042	0.126	6.6491
ANN	53,378.70	12.6117	0.069	0.142	-41.393
MLR	2.5569	2.0445	1.000	1.000	-0.0013

VI . CONCLUSIONS :

The creation of a Home Automation System based on Internet of Things (IoT) technologies has been a major step towards smarter, more connected living spaces. With the successful design and implementation of an operational prototype, this project has met real-world demands for convenience, energy efficiency, safety, and remote control in home environments. By employing devices like microcontrollers (e.g., NodeMCU), sensors (temperature, motion, etc.), actuators, and cloud-based systems, the system allows users to monitor and control home appliances through a smartphone app from anywhere in the world.

One of the most important achievements of this project is that it shows how IoT can make things easier in our daily lives. Automated lighting, temperature, and appliance control cut down on physical labor while energy consumption is minimized. For example, sensors capture motion or surroundings to change the lighting conditions automatically, thereby conserving electricity usage.Likewise, temperature sensors could be used alongside HVAC systems so that comfort does not come with wasteful consumption of energy. These functionalities provide a sustainable yet easy-to-use solution for new- age homes as a whole.

With regard to performance assessment, various predictive models like MLR, CNN, RNN, and ANN were employed to predict usage patterns and optimize system behavior. MLR proved to be the most accurate model with the lowest error measures (RMSE and MAE), ideal correlation (CC = 1.0), and full agreement (IOA = 1.0) with true values. This implies that linear relationships between variables like temperature, time of day, and device use can be modeled correctly using less complex statistical methods. Of deep learning models, CNN performed the best, displaying moderate error and fair correspondence with real data. RNN and ANN were poorer, and ANN had the poorest performance regarding prediction accuracy and correlation. These findings support the notion that complex models are valuable but more simplistic models can provide superior performance and interpretability for some datasets and tasks.

Although the system has been successful, there are some limitations. The dependency on internet connectivity can cause operations to fail if there is loss of network access. Security and privacy of data continue to be major issues, particularly where devices are remotely accessible. The initial setup cost of sensors, controllers, and the compatible appliances may be prohibitive to some people. These are, nonetheless, manageable through the right design options, encrypting data, and user training

This project also created opportunities for scalability and customization. Some future improvements may involve voice assistant support (e.g., Alexa or Google Assistant), machine learning-based personalization, energy usage analytics, and predictive maintenance notifications. Moreover, integrating the system with additional sensors and modules—such as gas leak sensors, surveillance cameras, and smart door locks—can greatly enhance home security and automation depth.

From a social viewpoint, smart home technologies can be used to aid disabled and elderly people by streamlining mundane activities and allowing caregivers to monitor them remotely. Also, as urban centers aim for sustainability, smart homes based on IoT can contribute to broader smart city infrastructures by providing anonymized energy usage statistics and being involved in grid balancing activities.

In short, this project not only achieved its fundamental task of developing a working, user-friendly home automation system but also shed light on efficient forecasting with machine learning and statistical models. It has exhibited the real-world applications of IoT in daily life and set the foundation for more intelligent, efficient, and safe residential technologies. With further improvements and wider adoption, such systems can play a critical role in transforming the way we interact with and manage our living spaces.

VII . REFERENCES :

Ahmed, N., Atiya, A. F., El Gayar, N., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. Econometric Reviews, 29(5-6), 594–621. https://doi.org/10.1080/07474938.2010.481556

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. PLoS ONE, 13(3), e0194889. https://doi.org/10.1371/journal.pone.0194889

Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Systems with Applications, 140, 112896. https://doi.org/10.1016/j.eswa.2019.112896

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14(1) 35–62. https://doi.org/10.1016/S0169-2070(97)00044- 7

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts. https://otexts.com/fpp2/

Brownlee, J. (2018). Deep learning for time series forecasting: Predict the future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery.

Ben Taieb, S., & Hyndman, R. J. (2014). A gradient boosting approach to the Kaggle load forecasting competition. International Journal of Forecasting, 30(2), 382–394. https://doi.org/10.1016/j.ijforecast.2013.07.005

Smyl, S. (2020). A hybrid approach of exponential smoothing and recurrent neural networks for time series forecasting. International Journal of Forecasting, 36(1), 75–85. https://doi.org/10.1016/j.ijforecast.2019.05.012

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Zhang, X., Wang, S., & Liu, M. (2017). A hybrid model based on CNN and LSTM for urban water demand prediction. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN) (pp.