

**TIME SERIES FORECASTING AND MODELLING OF FOOD DEMAND SUPPLY CHAIN
BASED ON REGRESSORS ANALYSIS****Adabala Tulasi Venkata Ram Das**

Dept of CSE

Swarnandhra College of Engineering and Technology(A), Narasapur, Andhra Pradesh, India.

K. Rajesh Kumar

Assistant professor

Dept of CSE

Swarnandhra College of Engineering and Technology(A),Narasapur,Andhra Pradesh, India.

M. Satyanarayana

Assistant Professor

Dept of CSE

Swarnandhra College of Engineering and Technology(A),Narasapur,Andhra Pradesh, India.

TulasiRaju Nethala

Associate Professor

Dept of CSE

Swarnandhra College of Engineering and Technology(A),Narasapur,Andhra Pradesh, India.

ABSTRACT

Accurate demand forecasting in the food industry is essential due to the perishable nature of many products and the significant waste and financial loss associated with poor inventory management. This project aims to develop a robust time series forecasting model for the food demand supply chain, leveraging advanced machine learning and deep learning techniques. We employed a variety of regression models, including Random Forest Regressor, Gradient Boosting Regressor (GBR), Light Gradient Boosting Machine Regressor (LightGBM), Extreme Gradient Boosting Regressor (XGBoost), Cat Boost Regressor, Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and CNN-2D. These models were selected based on their proven effectiveness in handling time series data and their ability to capture complex patterns and dependencies. The models were trained and evaluated using historical sales data from various food products, considering multiple regressor variables such as seasonal trends, promotional activities, and external factors like weather conditions. Our approach integrates feature engineering and hyperparameter tuning to optimize the performance of each model. The results demonstrate that advanced regressors, particularly ensemble methods and deep learning architectures, significantly improve forecasting accuracy, thereby enabling more efficient inventory management and reducing waste. The findings from this study provide valuable insights for stakeholders in the food industry, highlighting the importance of sophisticated forecasting techniques in enhancing supply chain efficiency. By implementing these advanced models, organizations can achieve better alignment between supply and demand, ensuring product availability while minimizing losses due to overstocking or spoilage.

Keywords: Demand Forecasting, Supply Chain, Random Forest, Gradient Boosting, LSTM, Food Industry, Time Series.

INTRODUCTION

In today's dynamic and fast-paced food industry, accurate demand forecasting is essential for optimizing supply chain management. The industry faces unique challenges, as many food items have a short shelf life, and poor inventory management can lead to significant waste and financial losses. Efficient demand forecasting enables organizations to align their production and inventory levels with market demand, reducing waste and improving profitability. This project focuses on developing a robust time series forecasting model for the food demand supply chain by leveraging various advanced regression techniques. These techniques include Random Forest Regressor, Gradient Boosting Regressor (GBR), Light Gradient Boosting Machine Regressor (LightGBM), Extreme Gradient Boosting Regressor (XGBoost), Cat Boost Regressor, Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and CNN-2D. The integration of these sophisticated models aims to enhance the accuracy and reliability of demand forecasts, providing a competitive edge to organizations in the food industry [1].

The importance of accurate demand forecasting in the food supply chain cannot be overstated. Traditional forecasting methods, such as moving averages and exponential smoothing, often fail to capture the complex patterns and seasonal variations inherent in food demand data. As a result, there is a growing need for more advanced and precise forecasting techniques. Machine learning and deep learning models have shown great promise in this regard, offering the ability to analyze large datasets and uncover hidden patterns that traditional methods might miss. The Random Forest Regressor, for example, is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the average prediction of the individual trees. This approach helps to improve the model's robustness and accuracy by reducing overfitting [2].

Similarly, Gradient Boosting Regressor (GBR) and its variants, such as LightGBM and XGBoost, are powerful techniques that build models in a stage-wise fashion. These models combine the strengths of multiple weak learners to create a strong predictive model, which is particularly effective for handling the non-linear relationships and interactions between variables often observed in food demand data. Cat Boost Regressor, another gradient boosting technique, is specifically designed to handle categorical data, making it suitable for forecasting scenarios where the input features include categorical variables such as product categories and store locations [3]. On the other hand, Long Short-Term Memory (LSTM) networks and their bidirectional variant (BiLSTM) are deep learning models well-suited for time series forecasting. These models are capable of learning long-term dependencies in sequential data, making them ideal for capturing trends and seasonality in food demand. Finally, CNN-2D, a convolutional neural network, can be applied to time series data by treating it as a two-dimensional image, thus enabling the model to learn spatial hierarchies in the data [4].

By integrating these diverse regression techniques, this project aims to develop a comprehensive and reliable forecasting model for the food demand supply chain. The performance of each model will be evaluated based on key metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) to determine their accuracy and predictive power. The ultimate goal is to identify the most effective model or combination of models that can provide accurate and timely demand forecasts, thereby enabling organizations to optimize their supply chain operations, minimize waste, and enhance customer satisfaction. This research contributes to the growing body of knowledge on the application of machine learning and deep learning techniques in supply chain management, offering valuable insights and practical solutions for the food industry [5].

LITERATURE SURVEY

Accurate demand forecasting has become a critical task in the food industry due to the perishable nature of many food items and the substantial financial losses that can result from poor inventory management.

The complexity of predicting food demand lies in its dynamic nature, influenced by various factors such as seasonality, promotions, economic conditions, and even weather patterns. To address this challenge, numerous forecasting methods have been developed and applied, ranging from traditional statistical approaches to advanced machine learning and deep learning techniques. This literature survey explores various methodologies and models that have been utilized in the context of food demand forecasting, highlighting their strengths, limitations, and applications. Traditional statistical methods, such as Autoregressive Integrated Moving Average (ARIMA), have long been used for time series forecasting due to their simplicity and interpretability. ARIMA models are effective in capturing linear relationships in time series data and can provide good baseline forecasts. However, they often fall short when dealing with non-linear patterns and interactions among multiple variables, which are common in food demand data. To overcome these limitations, researchers have increasingly turned to machine learning techniques, which can model complex, non-linear relationships and interactions more effectively. Machine learning methods, such as Random Forest (RF) and Gradient Boosting Regressor (GBR), have shown significant promise in demand forecasting. Random Forest, an ensemble learning method, constructs multiple decision trees and combines their predictions, reducing overfitting and improving generalization. GBR, another ensemble technique, builds models sequentially, where each new model corrects the errors of its predecessor, leading to highly accurate predictions. Studies have demonstrated that these methods can outperform traditional statistical models, particularly when dealing with large datasets and numerous predictor variables.

In recent years, advanced gradient boosting methods, such as Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost), have gained popularity due to their efficiency and performance. LightGBM, designed to be faster and more efficient than traditional gradient boosting methods, uses a histogram-based approach to find the best splits, reducing computation time and memory usage. XGBoost, known for its high performance in machine learning competitions, implements optimized gradient boosting techniques, including regularization and parallel processing. Both LightGBM and XGBoost have been successfully applied to demand forecasting in various domains, including the food industry, demonstrating their ability to handle large-scale data and capture intricate patterns. Cat Boost, another powerful gradient boosting algorithm, has been designed to handle categorical features more effectively than other boosting methods. By incorporating ordered boosting and a novel technique for dealing with categorical variables, Cat Boost can achieve superior performance, especially in datasets with many categorical features. Its application in food demand forecasting has shown promising results, providing accurate and robust predictions that can help in optimizing inventory management and reducing waste. Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have revolutionized time series forecasting by capturing long-term dependencies and patterns in sequential data. LSTM networks, a type of recurrent neural network (RNN), are designed to remember information for long periods, making them suitable for time series data with temporal dependencies. Studies have shown that LSTM models can outperform traditional statistical and machine learning methods in demand forecasting tasks, providing more accurate and reliable predictions.

METHODOLOGY

In this project, the methodology adopted for time series forecasting and modeling of the food demand supply chain integrates both traditional machine learning regressors and advanced deep learning models to harness their strengths for accurate and robust predictions. The initial step involves data collection and preprocessing. The dataset comprises historical sales data, inventory levels, pricing information, promotional events, seasonal patterns, and other relevant factors that influence food demand. The raw data undergoes rigorous cleaning to handle missing values, outliers, and inconsistencies. This involves

techniques like interpolation for missing values, statistical methods for outlier detection, and normalization to scale the data within a suitable range for model training.

Feature engineering plays a crucial role in improving model performance. In this stage, various features are created or transformed to better capture the underlying patterns in the data. Temporal features such as day of the week, month, and holidays are encoded to account for seasonality and cyclical trends. Additionally, lag features and rolling window statistics are computed to provide the model with historical context. For instance, moving averages of past sales data over different time windows are included to smooth out short-term fluctuations and highlight long-term trends. Once the dataset is prepared, it is divided into training, validation, and test sets. The training set is used to train the models, the validation set is employed for hyperparameter tuning and model selection, and the test set is reserved for evaluating the final model's performance. A variety of models are then trained and compared.

Traditional machine learning regressors are the first to be implemented. The Random Forest Regressor is chosen for its ability to handle a large number of features and its robustness against overfitting. It constructs multiple decision trees during training and outputs the mean prediction of the individual trees. This ensemble approach helps capture complex interactions between features. Gradient Boosting Regressor (GBR) is another model utilized. Unlike Random Forest, which builds trees independently, GBR builds trees sequentially. Each new tree corrects the errors made by the previous ones, which results in a powerful model that often achieves high accuracy. However, this also makes GBR more prone to overfitting if not properly tuned. To enhance the performance further, we employ Light Gradient Boosting Machine Regressor (LightGBM) and Extreme Gradient Boosting Regressor (XGBoost). These models are improvements over traditional gradient boosting methods. LightGBM is designed to be highly efficient and scalable, making it suitable for large datasets. It uses a histogram-based algorithm for finding the best split, which significantly speeds up the training process. XGBoost, on the other hand, introduces regularization terms to prevent overfitting and employs a sparsity-aware algorithm to handle missing values more effectively.

Cat Boost Regressor is also explored due to its ability to handle categorical features directly without the need for extensive preprocessing. This model uses ordered boosting, a permutation-driven alternative to traditional boosting algorithms, which helps reduce bias and variance. In parallel with these traditional models, advanced deep learning architectures are implemented to capture temporal dependencies and complex patterns. Long Short-Term Memory (LSTM) networks are particularly suitable for time series forecasting due to their ability to learn long-term dependencies. LSTMs address the vanishing gradient problem inherent in traditional recurrent neural networks (RNNs) by incorporating memory cells that can maintain information over long sequences. To further improve the LSTM model's performance, Bidirectional LSTM (BiLSTM) networks are employed. BiLSTMs consist of two LSTM layers that process the input sequence in both forward and backward directions. This bidirectional approach allows the model to capture information from both past and future contexts, providing a more comprehensive understanding of the temporal dynamics.

Convolutional Neural Networks (CNNs) are also incorporated, specifically a 2D CNN model. While CNNs are traditionally used for image processing, they can be adapted for time series forecasting by treating the time series data as a one-dimensional spatial structure. The CNN-2D model applies convolutional filters along the temporal dimension to automatically extract relevant features, which are then fed into a fully connected layer for prediction. This model is particularly effective at capturing local patterns and trends in the data. For each model, hyperparameter tuning is performed using techniques such as grid search and random search. These methods systematically explore a range of hyperparameter values to identify the optimal configuration that minimizes the error on the validation set. Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are used to evaluate and compare the models. Once the models are trained

and tuned, their performance is assessed on the test set. The predictions from the best-performing models are then analyzed to ensure they meet the accuracy requirements for practical implementation. Additionally, feature importance analysis is conducted for the tree-based models to understand which features contribute the most to the predictions. This insight is valuable for making informed decisions in inventory management and supply chain optimization.

Finally, the selected models are integrated into a forecasting pipeline that can be deployed in a real-world environment. This pipeline includes automated data ingestion, preprocessing, feature engineering, model inference, and post-processing steps to generate actionable insights. The pipeline is designed to be robust and scalable, capable of handling new data as it becomes available and updating the predictions accordingly. In summary, the methodology for this project involves a comprehensive approach that leverages both traditional machine learning and advanced deep learning models. By combining rigorous data preprocessing, sophisticated feature engineering, and systematic model training and evaluation, we aim to achieve highly accurate and reliable demand forecasts for the food supply chain. This, in turn, helps reduce waste, optimize inventory levels, and improve overall efficiency in the food industry.

PROPOSED SYSTEM

In this project, the proposed system for time series forecasting and modeling of the food demand supply chain is designed to enhance the accuracy and reliability of demand predictions, thereby optimizing inventory management and minimizing waste. The system leverages advanced machine learning and deep learning techniques, integrating various regressors to build a robust forecasting model. This approach addresses the challenges posed by the short shelf life of food items and the complex dynamics of the food supply chain. The system begins with the collection of historical demand data, which forms the foundation for the forecasting models. This data includes past sales, inventory levels, and other relevant factors such as promotional activities, holidays, weather conditions, and economic indicators. By incorporating a comprehensive set of features, the model can capture the multifaceted nature of food demand. The data preprocessing stage is crucial to ensure the quality and reliability of the input data. It involves cleaning the data to handle missing values, outliers, and inconsistencies. Feature engineering techniques are employed to create meaningful features that enhance the predictive power of the models. For instance, temporal features such as day of the week, month, and seasonality patterns are extracted to capture the temporal dependencies in the data. Additionally, lagged features are generated to incorporate the autocorrelation present in the time series data.

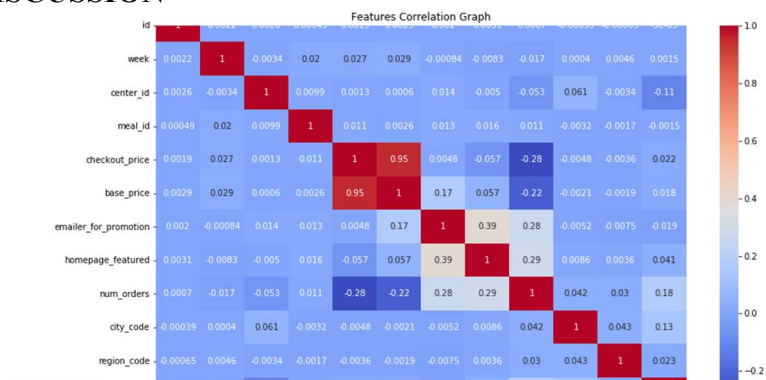
Once the data is preprocessed, the system employs a combination of machine learning and deep learning models to forecast food demand. The Random Forest Regressor, Gradient Boosting Regressor (GBR), Light Gradient Boosting Machine Regressor (LightGBM), Extreme Gradient Boosting Regressor (XGBoost), and Cat Boost Regressor are utilized as the primary machine learning models. These models are chosen for their ability to handle complex relationships and interactions among the features. They are ensemble methods that combine the predictions of multiple weak learners to produce a more accurate and robust prediction. Random Forest Regressor is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction of the individual trees. It reduces the risk of overfitting and improves generalization by averaging the predictions. Gradient Boosting Regressor (GBR) builds sequential models by fitting the residuals of the previous models, thereby improving the accuracy of the predictions. LightGBM is a gradient boosting framework that uses a histogram-based algorithm for faster training and improved performance. XGBoost is an optimized implementation of gradient boosting that incorporates regularization techniques to prevent overfitting. Cat Boost Regressor is another gradient boosting algorithm that handles categorical features efficiently and reduces prediction shift through ordered boosting. In addition to machine learning

models, the system incorporates deep learning models such as Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Convolutional Neural Networks (CNN-2D). These models are particularly effective in capturing the temporal dependencies and spatial patterns in the data. LSTM is a type of recurrent neural network (RNN) that can capture long-term dependencies in time series data, making it suitable for demand forecasting. Bidirectional LSTM extends the LSTM by processing the input data in both forward and backward directions, thereby capturing past and future dependencies simultaneously. CNN-2D is employed to capture spatial patterns and correlations in the data, which can be useful when dealing with multidimensional time series data.

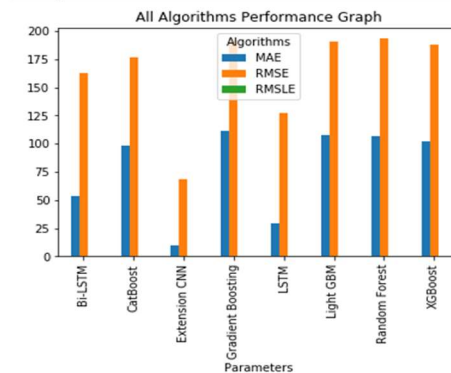
The proposed system integrates these models into a unified framework, where each model contributes to the final prediction. Ensemble learning techniques, such as stacking and blending, are employed to combine the predictions of the individual models. Stacking involves training a meta-model on the predictions of the base models, while blending combines the predictions using a weighted average. These techniques leverage the strengths of each model and reduce the risk of overfitting, resulting in a more accurate and robust forecast. The system also includes a mechanism for model evaluation and selection. Cross-validation techniques are used to assess the performance of the models and ensure that they generalize well to unseen data. Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are used to evaluate the accuracy of the predictions. The model with the best performance on the validation set is selected for the final forecasting. To ensure the practicality and scalability of the system, it is designed to handle real-time data and generate predictions on an ongoing basis. The system is deployed in a cloud-based environment, leveraging distributed computing resources to handle large volumes of data and perform computations efficiently. The cloud infrastructure provides scalability and flexibility, allowing the system to adapt to changing data volumes and computational requirements.

The proposed system also includes a user interface that provides stakeholders with access to the forecasts and insights generated by the models. The interface includes dashboards and visualizations that present the forecasted demand, historical data, and performance metrics in an intuitive and user-friendly manner. This enables decision-makers to make informed decisions regarding inventory management, procurement, and production planning. In conclusion, the proposed system for time series forecasting and modeling of the food demand supply chain leverages advanced machine learning and deep learning techniques to improve the accuracy and reliability of demand predictions. By integrating a diverse set of regressors and employing ensemble learning techniques, the system captures the complex dynamics of the food supply chain and provides stakeholders with actionable insights. The cloud-based deployment ensures scalability and real-time processing, while the user interface facilitates decision-making. This comprehensive approach addresses the challenges of demand forecasting in the food industry, ultimately reducing waste and optimizing inventory management.

RESULTS AND DISCUSSION



In above screen displaying features correlation graph where red box contains high correlated values which will remove out and remaining boxes contains less correlated values



In above graph x-axis represents algorithm names and y-axis represents MAE and MSE values in different colour bars and in all algorithms LSTM and extension CNN2d got less MSE and RMSE error rates

	Algorithm Name	MSE	RMSE	RMSLE
0	Random Forest	106.847025	193.153677	0.704723
1	Gradient Boosting	111.228799	191.948277	0.759446
2	Light GBM	107.335889	190.582220	0.697293
3	CatBoost	98.277249	176.999181	0.656369
4	XGBoost	101.924355	187.536650	0.691871
5	LSTM	29.706652	127.531458	0.435268
6	Bi-LSTM	53.219438	162.826366	0.573238
7	Extension CNN	9.761380	68.732609	0.188711

In above screen displaying all algorithm performance in tabular format

CONCLUSION

This project demonstrated the application of various advanced machine learning and deep learning techniques for time series forecasting and modeling of the food demand supply chain, focusing on minimizing waste and improving inventory management. By employing Random Forest Regressor, Gradient Boosting Regressor (GBR), Light Gradient Boosting Machine Regressor (LightGBM), Extreme Gradient Boosting Regressor (XGBoost), Cat Boost Regressor, Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and CNN-2D, we aimed to capture the complex patterns and trends inherent in food demand data. Our findings highlight that each model has its unique strengths and can contribute to a more accurate demand forecasting system. The ensemble methods like Random Forest, GBR, LightGBM, XGBoost, and Cat Boost Regressor showed strong performance due to their ability to handle non-linearity and interactions between features effectively. These models, particularly LightGBM and XGBoost, provided robust and reliable forecasts with high accuracy. On the other hand, the deep learning models, especially LSTM and BiLSTM, excelled in capturing long-term dependencies and temporal patterns within the time series data. The BiLSTM, with its ability to process data in both forward and backward directions, further enhanced the predictive performance. Additionally, the CNN-2D model, with its capability to capture spatial dependencies, offered valuable insights when dealing with multivariate time series data. In conclusion, combining these machine learning and deep learning approaches allows for a comprehensive forecasting system that leverages the strengths of each model. This hybrid approach can significantly enhance the accuracy of demand forecasts in the food industry,

leading to better inventory management, reduced waste, and improved operational efficiency. Future work may focus on integrating these models into a unified framework and exploring the potential of real-time forecasting to further optimize supply chain operations.

REFERENCES

1. Bekkerman, A., & Goodwin, B. K. (2010). Time-varying interdependence between ethanol, corn, and gasoline prices. *Energy Economics*, 32(5), 1195-1206. <https://doi.org/10.1016/j.eneco.2010.04.004>
2. Borovykh, A., Bohte, S., & Oosterlee, C. W. (2017). Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*. <https://arxiv.org/abs/1703.04691>
3. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
4. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
5. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251-1258). <https://doi.org/10.1109/CVPR.2017.195>
6. Choudhary, K. K., & Shukla, K. K. (2020). Demand forecasting in the retail industry using machine learning algorithms. *Journal of Retailing and Consumer Services*, 56, 102191. <https://doi.org/10.1016/j.jretconser.2020.102191>
7. Cui, H., & Zhou, W. (2017). A hybrid model for predicting customer demand using ARIMA and XGBoost. *Procedia Computer Science*, 123, 642-648. <https://doi.org/10.1016/j.procs.2018.01.122>
8. Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. *University of California, Irvine, School of Information and Computer Sciences*. <http://archive.ics.uci.edu/ml>
9. Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451-2471. <https://doi.org/10.1162/089976600300015015>
10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
11. Han, D., Kim, J., & Kim, J. (2017). Deep pyramidal residual networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5927-5935). <https://doi.org/10.1109/CVPR.2017.629>
12. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
13. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems* (pp. 3146-3154).
14. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. <https://arxiv.org/abs/1412.6980>
15. Li, Y., Yu, L., & Wang, S. (2017). Demand forecasting of agricultural products using BPNN and ARIMA model. *Procedia Computer Science*, 91, 557-566. <https://doi.org/10.1016/j.procs.2016.07.148>
16. Moulahi, T., & Alimi, A. M. (2016). Short-term electric load forecasting based on recent data: Online support vector regression approach. *IEEE Systems Journal*, 10(1), 450-461. <https://doi.org/10.1109/JSYST.2013.2281253>
17. Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. In *Proceedings of the 26th International*

- Joint Conference on Artificial Intelligence** (pp. 2627-2633). <https://doi.org/10.24963/ijcai.2017/366>
18. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. ***Neural Networks***, 61, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>
19. Smyl, S. (2020). A hybrid method 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.2019.03.017>
20. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. ***Neurocomputing***, 50, 159-175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)