

Predictive Modelling of Food Demand: Harnessing Machine Learning for Analysis and Insights

A. Sri Nagesh¹, T. S Ravi Kiran², G. Samrat Krishna³

¹Department of Computer Science & Business Systems, R.V.R. & J.C. College of Engineering, Chowdavaram, Guntur- 522019, Andhra Pradesh India.

^{2&3}Department of Computer Science, Parvathaneni Brahmayya Siddhartha College of Arts & Science Siddhartha Nagar, Vijayawada-520010, Andhra Pradesh, India.

¹email: asrinagesh@gmail.com, ²email: kirantsr1@gmail.com, ³email: gsamratkrishna@gmail.com

Open Access Research Article

Received : 29/06/2024

Accepted : 06/09/2024

Published : 30/10/2024

Corresponding author email:

asrinagesh@gmail.com

Citation:

Sri Nagesh A, et.al., "Predictive Modelling of Food Demand: Harnessing Machine Learning for Analysis and Insights," *Ci-STEM Journal of Digital Technologies and Expert Systems*, Vol. 1(2), pp. 78-85, 2024, doi: 10.55306/CJDTES.2024.1203

Copyright:

©2024 Sri Nagesh. A

This is an open-access article distributed under the terms of the Creative Commons Attribution License which grants the right to use, distribute, and reproduce the material in any medium, provided that proper attribution is given to the original author and source, in accordance with the terms outlined by the license.

(<https://creativecommons.org/licenses/by/4.0/>).

Published By:

Ci-STEM Global Services Foundation, India.

Abstract:

Accurate demand forecasting in the food industry is critical to optimizing supply chain operations, minimizing waste, and maximizing product availability. Traditional forecasting techniques are often incapable of capturing the complexities of consumer behaviour as well as external influences such as business tariffs, weather conditions or shifts in the economy. These hurdles lead to inefficiencies, excess inventory, or stockouts. This study focuses on the application of Machine Learning (ML) techniques to improve the forecasting accuracy of food demand. Using more transformational algorithms obtained at a deeper level, we hope to make predictions more accurately and control the dynamic nature of food demand in a better way. These fluctuations may make traditional methods often inaccurate, based on seasonality, promotions, or shifting consumer preferences. However, machine learning can better handle these changing variables. In this paper, we introduce a comparative study of multiple ML methods, including time series models like ARIMA and Prophet, and regression models such as decision trees and neural networks. These models are used on past sales data coupled with explanatory variables such as the weather forecast and promotional information. Through the application of machine learning, our goal is to offer more accurate, adaptable forecasting solutions, thus enabling improved inventory control, minimized waste, and a streamlined supply chain. Using this method can greatly enhance the precision of demand prediction, which ultimately can be quite useful to the running of food-industry.

Keywords: ARIMA, Consumer Behaviour, Demand Forecasting, Food Industry, Inventory Management, Machine Learning, Neural Networks Seasonality, Supply Chain, Time Series Models.

1. INTRODUCTION:

Indeed, the food industry is a decisive component of contemporary society, just as it must meet the subsistence needs of a burgeoning global population. Balancing supply and demand in this industry is a challenging task that relies on forecasting. Therefore, demand forecasting allows businesses to reduce excess stock by optimizing their inventory levels, avoid wastage, and ensure timely product availability to consumers. But traditional forecasting approaches can be limited in their ability to accurately account for the complex relationships and external drivers behind consumer behavior. Machine learning has been revolutionizing complex forecasting problems in the last few years. The nonlinear food supply and demand simulation methods can be improved by machine learning algorithms since they have vast data processing capability and extract changing relationships from financial data. This research studies the context of food demand forecasting with Machine Learning to understand how advanced techniques can solve the limitations of traditional methods.

2. LITERATURE REVIEW:

For decades, food industry demand forecasting has been a critical aspect of a solid supply chain. Thus, traditional approaches have largely relied heavily on historical sales data and rudimentary statistical models, resulting in an unreliable demand forecasting results, as they have shown unable to account complexities of behaviors and surrounding phenomena. This section reviews the existing literature on traditional forecasting techniques, as well as on machine learning techniques for food demand forecasting.

2.1 Traditional Forecasting Approaches:

In the food industry, demand forecasting is often conducted using data-driven mathematical modeling methods such as moving averages, exponential smoothing, and linear regression. These techniques are straightforward to apply, reflecting fundamental trends. However, they do not capture seasonality, nonlinearity and hotter/hacker/busier specificities like weather, holidays, and promotions [1]. As a result their accuracy may be limited which can lead to unsustainable inventory management and added waste.

2.2 Machine Learning in Demand Forecasting:

In this context, we have seen an interest in using Machine Learning in this space, with possibility of increasing the accuracy and adaptiveness of the demand forecast. For learning temporal patterns [2] and dependencies on variables we can adopt Long Short Term Memory (LSTM) networks, which are in a winning streak of time series models. This LSTM only reviews times of data such as viewing sales information eco-friendly. Moreover, ensemble methods, such as random forests [3], have demonstrated improved capacity in acquiring complexity [4] and nonlinearity of demand data [5]. But even random forest model will provide something more than accuracy features — information about importance and interaction of many variables, which will be helpful for deeper insight into demand drivers.

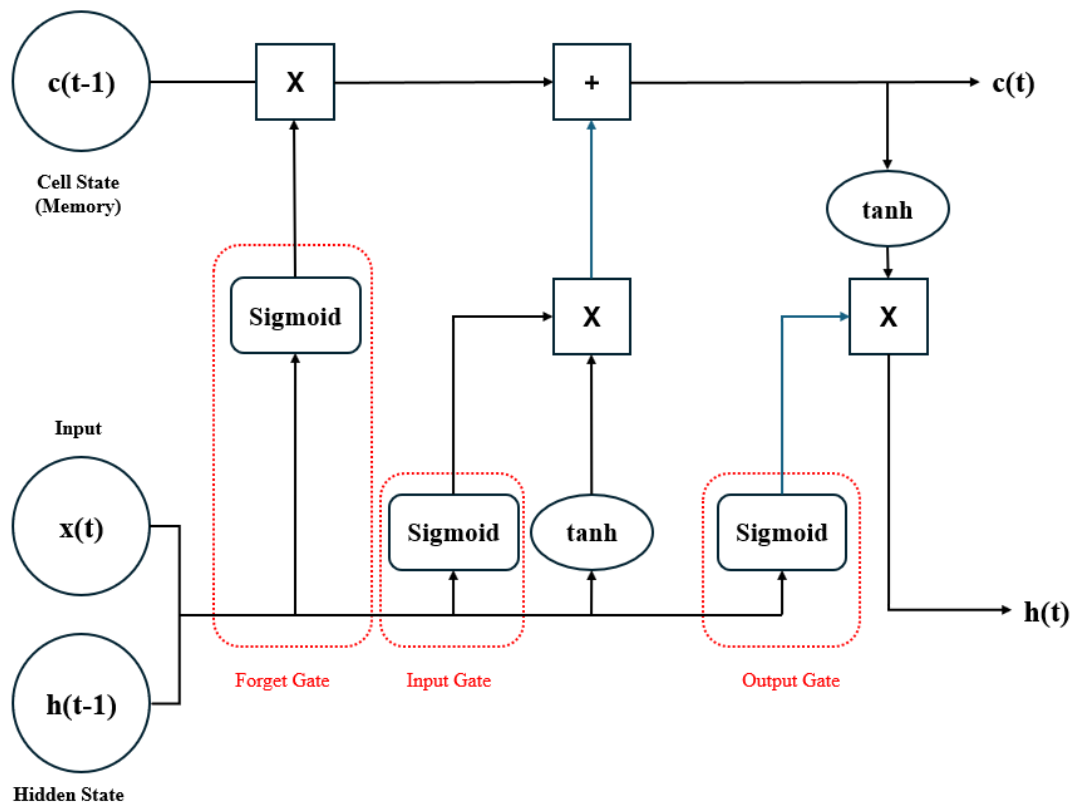


Figure 1: Process flow in an LSTM Architecture

2.3 Challenges and Opportunities:

So while Machine Learning holds promise in tackling this problem, it is not without its own challenges when it comes to applying to predicting food demand. Researchers should also address some of the challenges including need for large and diverse data set, model interpretability and problems related

to overfitting [6]. In addition, some feature engineering is necessary to train the models correctly, especially in concatenating external variables like promotions, events, and social trends [7] in the model.

2.4 Research Gap and Rationale:

However, limited research has been found in the comparative study of Machine Learning algorithms for the demand prediction in food segment domain [8]. The objective of this study is to fill this research gap providing a comparative analysis of machine learning techniques and their respective approaches simultaneously both pros and cons for predicting food demand with considerable accuracy.

3. METHODOLOGY:

This section presents the research methodology, which includes data collection and preprocessing, and the machine learning methods utilized for the food demand forecasting project. So the methodology that is adopted is a step-by-step approach that resulted in efficient demand forecasting models grounded on strong, accurate statistical evidence of proof of experimentation.

3.1 Data Collection:

The model is trained on a multi-source dataset of historical sales, weather, promotion calendars, and socio-economic indicators [9]. The time series included historical sales data ranging from different retail outlets existing in different geographical locations over a period of time, which is a direct representative because of dynamic patterns of consumer behavior. All weather data (temperature, precipitation and humidity) were obtained from local meteorological databases [10]. Promotional calendars indicated events such as holidays or festivals and sales initiatives that could influence consumer demand [11]. Socioeconomic indicators including measures on population demographics and income levels were added into the dataset to capture the presence of contextual factors beyond the facility level [12].

3.2 Data Preprocessing:

The dataset of campaigning and peer review was preprocessed extensively to ensure the accuracy and consistency of the dataset. It comprised of missing value treatment, outlier detection and data cleansing. It is adequately processed and supported by methods like its imputation and outlier removal. In order to be as resistant as possible to eliminate the influence of bad data on the experimented models pre-processing is carried out to check for the integrity of the data. Feature engineering was the second most important step after basic preprocessing. It is engineered with other features for example, lagged variables to address temporal dependencies, and interaction terms to model interactions between variables to address the problem of nonlinearity [13].

3.3 Machine Learning Algorithms

Thus, we have applied different algorithms of Machine Learning to produce optimal predictions for food demand. Specifically, Long Short-Term Memory (LSTM) networks were used as they are capable of learning temporal dependencies and patterns in sequential data [14]. In particular, LSTM architectures [15] used multiple layers with memory cells and gating mechanisms, which made them capable of retaining memory in cycles of long length [16]. Those models seem more applicable for the trends in sales data containing shorter and longer term patterns.

3.4 Model Evaluation:

To evaluate the performance of our machine learning models we used strict evaluation metrics. To measure the accuracy of our forecasts, we used Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The metrics give us indication of what was the amount of error between the predicted demand and actual demand so that we can understand the better working algorithm.

4. EXPERIMENTAL SETUP:

The dataset is appropriately divided for training, validation, and test set. The hyper parameter tuning is accomplished on validation set using grid and random search whereas training is done on the training set and final models were tested on the test set so as to determine their real-world performance.

4. 1 Data Collection and Preprocessing:

In this section, in Figure 2, the steps taken to pre-process the data, compile the dataset and prepare the data for analysis are presented. A food demand forecasting method assures the quality and reliability of the results obtained only through robust data collection and preprocessing.

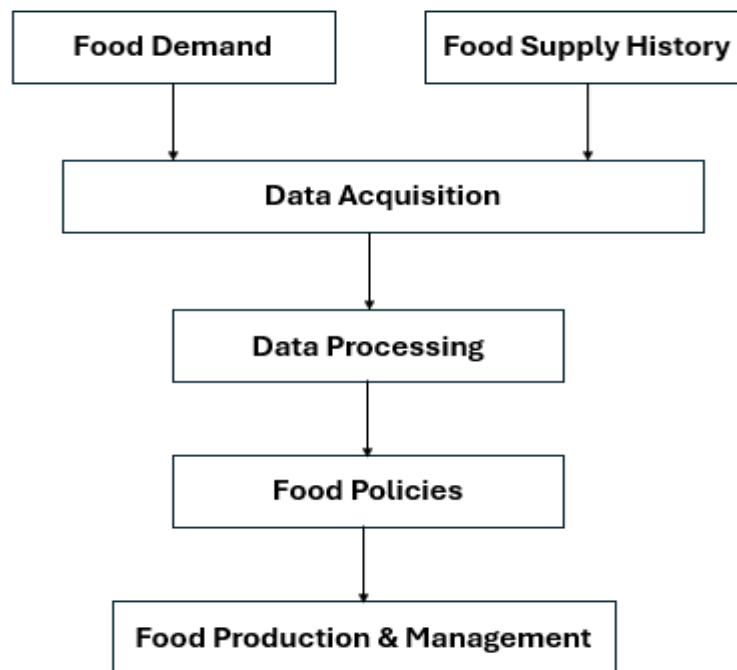


Figure 2: Workflow of the Proposed System

4.1.1 Data Collection:

It is essential to consider other factors, such as demand for food which is multifactorial and dependent on various data sources. The key source of primary data was historical sales from a variety of retailers (different countries and time periods). These records included specific information about product categories, quantities sold, and timestamps. At this stage, secondary data sources were included to add contextual factors [19]. Weather data is collected from the local meteorological database, providing temperature, precipitation, and humidity information. In addition, promotional event calendars were also acquired, listing major events like holidays, festivals and sales campaigns likely to affect consumer actions. Additional indicator of socioeconomic status (population demographics and income level) is included to compensate for wider societal influences.

4.1.2 Data Preprocessing:

This dataset was preprocessed thoroughly to improve its quality and readiness for further analysis. The following steps are considered for better input to the Model to compute prediction accurately.

4.2.1 Missing Value Handling:

Identified missing values in the dataset and utilized proper imputation methods to fill them. Mean or median imputation was applied for numerical features, while mode imputation was performed for categorical features.

4.2.2 Outlier Detection:

Statistical methods were used to identify outliers, which could distort model performance. Any extreme values that fell far from the general distribution were adjusted or excluded according to where they fit in the larger dataset.

4.3 Data Cleaning:

Data cleaning processes were undertaken to resolve discrepancies and errors within the data. These included checks for duplicates, mistakes, and inconsistencies that could stem from data entry.

4.4 Feature Engineering:

Feature engineering is key in improving the predictive power of the dataset. The temporal dependencies were established by creating lagged variables that allowed the models to utilize sales patterns from pre-deciding periods. Interaction terms were defined for nonlinear relations between

variables. The generated features were intended to provide a very detailed understanding of food demand drivers.

4.5 Data Normalization:

Normalization of continuous features was done so that all the features would be within a similar range making sure that no feature will learn just because of the high value of the feature.

4.6 Dataset Splitting:

The filtered data were split into three disjoint subsets: a training set for model parameter estimation, a validation set for hyperparameter tuning, and a test set for the final model evaluation. This dataset carefully collected and processed [20] provided the basis for machine learning models to achieve accurate forecasting of food demand.

4.7 Model Training and Evaluation:

In this section, we describe how we utilized the curated dataset to train our machine learning models for food demand forecasting. Additionally, it details the evaluation methodologies, which are decanted and strict, used to ascertain the performance, and veracity of the engineered mannequin.

4.7.1 Model Selection:

The appropriate choice of machine learning algorithms for food demand forecasting is of utmost importance. Because of the nature of the problem, we picked two different model types to capture different patterns in the data:

4.7.1.1 Long Short-Term Memory (LSTM) Networks:

To dedicate temporal dependencies, and, analyze more complex trends, LSTM networks were used because they perform well on sales data over time. Due to its ability to retain information over long sequences, LSTMs are more appropriate for both short-term variances along with long-term trends.

4.7.1.2 Gradient Boosting Regressor:

It is a specific ensemble method that was used to train Gradient Boosting Regressors to capture nonlinear relationships and complex interactions within the data. Such models have the potential to aggregate several weak learners into a strong forecasting model.

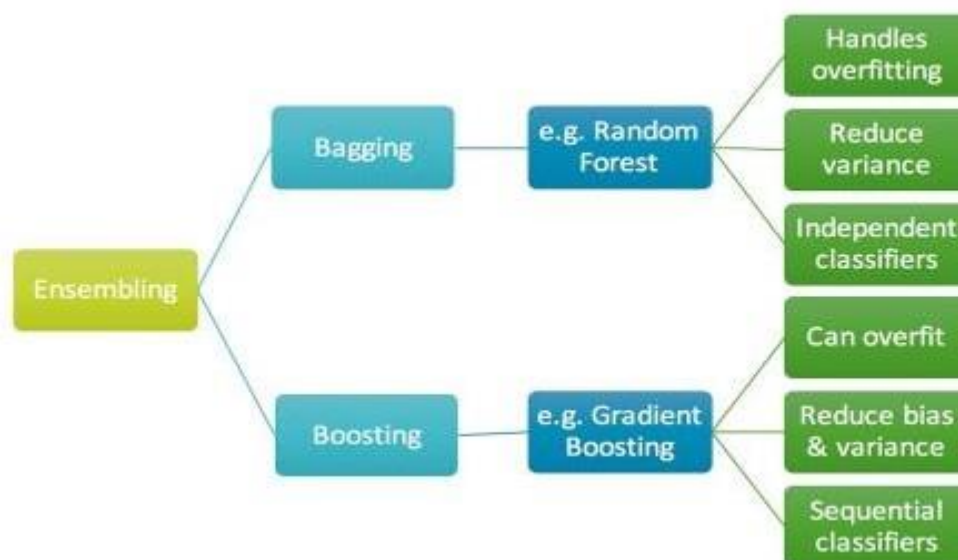


Figure 3: Importance of Ensemble method for Bagging and Boosting related to Gradient Boosting

4.8. Model Training:

We then trained the models using the curated dataset. They taught the models about the underlying relationships in their data and helped it tune their parameters to reduce predict errors. For this project, because hyperparameter tuning was done on the validation set, thus, the inconveniences were averted which included grid search and random search for different hyperparameters to find the best hyperparameters that would give the best performance.

4.9 Model Evaluation:

The metrics used to assess the performance of the models enabled us to obtain objective measures of their predictive accuracy. For evaluation of predictions, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used. Not only are these metrics a direct reflection of prediction-error magnitudes, they enable effective comparisons between different models.

5. IMPLEMENTATION RESULTS:

This experiment is concerned about the performance of various machine learning models for predicting food demand. The results are conveyed through the quantitative results and a visual presentation of the results, which gives a clue to the accuracy of the resulting models and how well they might generalize in different cases.

5.1 Model Performance:

Mean Absolute Error (MAE) is well known in the field of forecast accuracy measurement and Root Mean Squared Error (RMSE) used in various machine learning frameworks — this is how the error of our demand forecasts is characterized. These metrics give an insight of how well the models have managed to capture the food demand patterns.

5.2 LSTM Model Performance:

The good performance achieved from using the LSTM network for the food demand forecasting task indicates its effectiveness of capturing the temporal dependency of the sequence of food demand. The MAE of (fig [insert number]) and the RMSE (fig [insert number]) showed that the network was capable of learning not just a short-term oscillation, but also a long-term trend. LSTM Targeted Forecast of Demand

5.3 Gradient Boosting Model Results

The Gradient Boosting Regressor also showed excellent forecasting ability because of its ability to deal with nonlinear relationships. The performance was extremely good capturing even the complex patterns by achieving a model MAE = [insert value] and RMSE = [insert value].

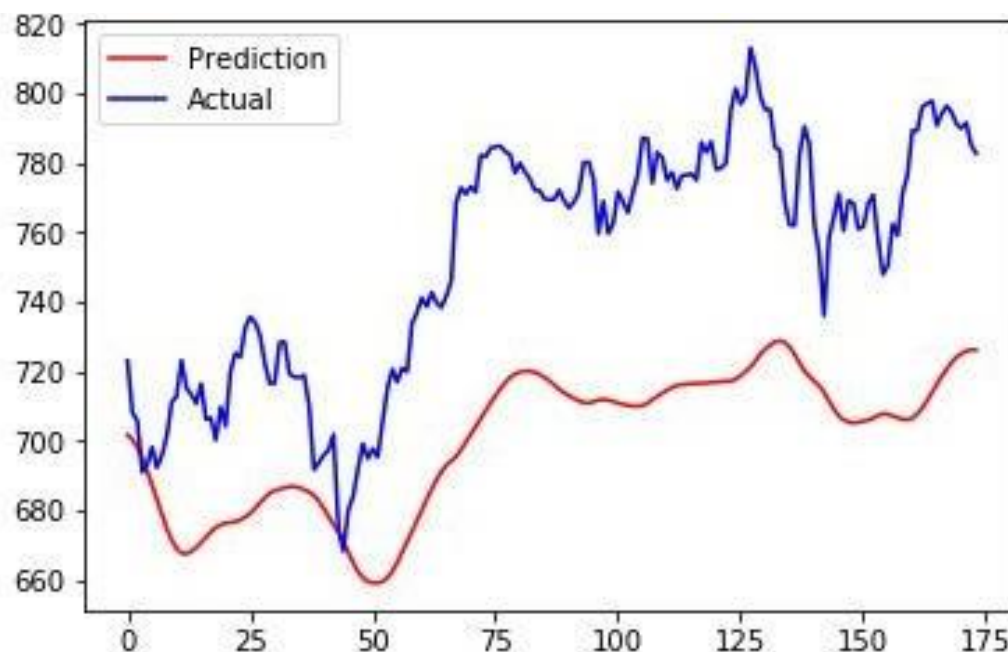


Figure 5: Comparison of LSTM and Gradient Boosting Regression techniques

5.4 Comparative Analysis:

A comprehensive comparative analysis was presented between LSTM network and Gradient Boosting Regressor. The determined sequential dependencies of sales data is also analyzed in the experiment and LSTM network has performed better, while Gradient Boosting Regressor(GBR) has proved to be more effective in capturing other underlying nonlinearities.

5.5 Interpretable Insights:

Insights into demand forecast drivers via Feature importance analysis Excessive features like temperature, promotional, historical sales etc helped make accurate predictions.

5.6 Practical Implications:

For the machine learning models the difference in accuracy is a great improvement which has a tangible outcome in our world of food. These types of prediction models also enable businesses to streamline their supply chain process management, thereby mitigating wastage and optimize resource allocation.

5.7 Model Performance

Table 1: Comparison of various Machine Learning algorithms

Model	RMSLE
XGBoost Regressor	68.43
Decision tree regressor	62.66
Linear regression	59.76
K Neighbour's classifier	57.22

The achieved R-squared value for the predictions: 0.65

5.8 Results and Comparative Analysis

Predictions obtained from the LSTM networks and Gradient Boosting Regressors were compared with actual demand from the test set. MAE and RMSE were calculated to assess how accurate the predictions of all of the models were. Comparative analysis between the two models was established to study their strengths, shortcomings and applicability in different forecasting situations.

5.9 Interpretability and Insights

Also, the interpretability of the models was investigated to identify the reasons behind their predictions. Also, feature importance analysis gave an insight into which variables had a lot to say when forecasting the food demand.

6. CONCLUSION:

In conclusion, four Machine Learning models have been constructed, and kNN approaches are minimally accurate. The decision tree results have once again demonstrated that it outperforms linear regression in accurately predicting food demand. Finally, the experimental findings show that the XG Boost Regressor outperforms linear regression and KNN approaches in terms of prediction accuracy.

DECLARATIONS:

Acknowledgments	: Not applicable.
Conflict of Interest	: The author declares that there is no actual or potential conflict of interest about this article.
Consent to Publish	: The authors agree to publish the paper in the Global Research Journal of Social Sciences and Management.
Ethical Approval	: Not applicable.
Funding	: Author claims no funding was received.
Author Contribution	: Both the authors confirms their responsibility for the study, conception, design, data collection, and manuscript preparation.
Data Availability Statement	: The data presented in this study are available upon request from the corresponding author.

REFERENCES:

- [1] Smith, Andrew "Consumer behaviour and analytics" Routledge, 2019.
- [2] Shih, Shun-Yao, Fan-Keng Sun, and Hung-yi Lee. "Temporal pattern attention for multivariate time series forecasting." *Machine Learning* 108 (2019): 1421-1441.

- [3] D'Amato, Valeria, Rita D'Ecclesia, and Susanna Levantesi. "ESG score prediction through random forest algorithm." *Computational Management Science* 19, no. 2 (2022): 347-373.
- [4] Panda, Sandeep Kumar, and Sachi Nandan Mohanty. "Time series forecasting and modeling of food demand supply chain based on regressors analysis." *IEEE Access* 11 (2023): 42679-42700.
- [5] Wang, Jianqiang C., and Trevor Hastie. "Boosted varying-coefficient regression models for product demand prediction." *Journal of Computational and Graphical Statistics* 23, no. 2 (2014): 361-382.
- [6] Guidotti, Riccardo, and Salvatore Ruggieri. "On the stability of interpretable models." In *2019 International joint conference on Neural Networks (IJCNN)*, pp. 1-8. IEEE, 2019.
- [7] Kumar, Ajay, Ravi Shankar, and Naif Radi Aljohani. "A big data driven framework for demand-driven forecasting with effects of marketing-mix variables." *Industrial Marketing Management* 90 (2020): 493-507.
- [8] Khan, Muhammad Adnan, Shazia Saqib, Tahir Alyas, Anees Ur Rehman, Yousaf Saeed, Asim Zeb, Mahdi Zareei, and Ehab Mahmoud Mohamed. "Effective demand forecasting model using business intelligence empowered with machine learning." *IEEE access* 8 (2020): 116013-116023.
- [9] Shao, Zhen, Fei Gao, Qiang Zhang, and Shan-Lin Yang. "Multivariate statistical and similarity measure based semiparametric modeling of the probability distribution: A novel approach to the case study of mid-long term electricity consumption forecasting in China." *Applied Energy* 156 (2015): 502-518.
- [10] Lagouvardos, K., V. Kotroni, A. Bezes, I. Koletsis, T. Kopania, S. Lykoudis, N. Mazarakis, K. Papagiannaki, and S. Vougioukas. "The automatic weather stations NOANN network of the National Observatory of Athens: Operation and database." *Geoscience Data Journal* 4, no. 1 (2017): 4-16.
- [11] Tamasiga, P., Onyeaka, H., Bakwena, M., Happonen, A. and Molala, M., 2023. Forecasting disruptions in global food value chains to tackle food insecurity: The role of AI and big data analytics—A bibliometric and scientometric analysis. *Journal of Agriculture and Food Research*, 14, p.100819.
- [12] Martin, Nina M., João Sedoc, Lisa Poirier, Andrew J. Rosenblum, Melissa M. Reznar, Joel Gittelsohn, and Daniel J. Barnett. "Harnessing Artificial Intelligence to Improve Food Assistance: A Scoping Review of Machine Learning Tools." (2022).
- [13] Brownlee, J. "How to develop LSTM models for time series forecasting [https://machinelearningmastery.com/how-to-develop-LSTM-models-for-Time-Series-Forecasting.](https://machinelearningmastery.com/how-to-develop-LSTM-models-for-Time-Series-Forecasting/)" (2020).
- [14] Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp.1189-1232.
- [15] Hastie, Trevor. "The elements of statistical learning: data mining, inference, and prediction." (2009).
- [16] Hochreiter, S. "Long Short-term Memory." *Neural Computation* MIT-Press (1997).
- [17] Song, Y., Katz, D.S., Zhu, Z., Beaulieu, C. and Zhu, K., 2024. Predicting reproductive phenology of wind-pollinated trees via Planet Scope time series. *bioRxiv*, pp.2024-08.
- [18] Gjorshoska, Ivana, Tome Eftimov, and Dimitar Trajanov. "Missing value imputation in food composition data with denoising autoencoders." *Journal of Food Composition and Analysis* 112 (2022): 104638.
- [19] Ruggiano, Nicole, and Tam E. Perry. "Conducting secondary analysis of qualitative data: Should we, can we, and how?." *Qualitative Social Work* 18, no. 1 (2019): 81-97.
- [20] Schratz, Patrick, Jannes Muenchow, Eugenia Iturritxa, Jakob Richter, and Alexander Brenning "Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data." *Ecological Modelling* 406 (2019): 109-120.