Educational Administration: Theory and Practice
2024, 30(5), 14647 - 14650
ISSN: 2148-2403
https://kuey.net/

Research Article

Impact Of Data Preprocessing On Prophet-LSTM Hybrid Models For Time Series Forecasting

Km Priyanka Saini[1*], Dr. Amit Sharma[2]

[1*]Student Of M.Tech Computer Science And Engineering., Chaudhary Charan Singh University Campus, Meerut (SCRIET), Km786084@Gmail.Com
[2]Assistant Professor, Department Of Computer Science And Engineering, Chaudhary Charan Singh University Campus, Meerut, Amitccsu.Net@Gmail.Com

Citation: Km Priyanka Saini, Amit Sharma (2024), Impact Of Data Preprocessing On Prophet-LSTM Hybrid Models For Time Series Forecasting, Educational Administration: Theory and Practice, 30(5), 14647 - 14650
Doi: 10.53555/kuey.v30i5.7207

ARTICLE INFO

ABSTRACT

In the field of time series forecasting, the precision and effectiveness of predictive models have been transformed by the combination of sophisticated machine learning algorithms with data preparation methods. In order to detect intricate patterns in time series data, the Prophet-LSTM hybrid model combines the trend prediction ability of the Prophet algorithm with the fine-tuning capability of the (LSTM) neural network. The goal of this research investigate the substantial influence that data preprocessing has on the time series forecasting capabilities of the Prophet-LSTM hybrid models.

I. Introduction

1. In a variety of fields, such as economics, meteorology, and sales forecasting, time series forecasting is crucial. Data preprocessing, which involves cleaning, transforming, and preparing the data before feeding it into the model, is crucial for increasing the accuracy of forecasting models. To make sure the data is in a format appropriate for analysis, data preparation entails eliminating outliers, managing missing values, and normalizing the data. This hybrid model has shown encouraging accuracy and resilience by combining the best aspects of Prophet and LSTM. However, little study has been done on how different data preprocessing techniques impact the performance of Prophet-LSTM hybrid models. Knowing how data preprocessing affects these hybrid models' predictions can help them perform better in time series forecasting applications. In this work, we want to assess how different data preparation techniques affect the time series forecasting performance of Prophet-LSTM hybrid models. We can compare the outcomes of several preprocessing techniques to ascertain which strategy produces the most trustworthy and accurate forecasts. This work aims to contribute to the literature by elucidating the role that data preprocessing plays in enhancing the performance of hybrid forecasting models.

A. Background and Context

To comprehend how data preprocessing affects Prophet-LSTM hybrid models for time series forecasting, it is essential to take into account the history and setting of these methods. In a variety of industries, including marketing, finance, and healthcare, time series forecasting is crucial because precise forecasts are necessary for formulating strategies and making decisions. Traditional time series techniques, including ARIMA and exponential smoothing, find it difficult to handle complicated nonlinear patterns and seasonality. Seasonality and intricate nonlinear patterns are difficult for traditional time series models to manage, such as exponential smoothing and ARIMA. One important component in forecasting model success is data quality. Among the tasks involved in data preparation include feature engineering, normalization, and filling in missing values. Preprocessing techniques have been shown to have a significant impact on the accuracy and efficiency of models in previous studies. For forecasting systems to be robust and dependable, it is essential to comprehend the history and context of these approaches [2].

II. Impact of Data Preprocessing on Prophet-LSTM Hybrid Models

Before feeding the data into Prophet-LSTM hybrid models, data preprocessing techniques can significantly improve forecasting performance. Data pre-treatment is crucial to increasing the accuracy and dependability of the hybrid model by addressing issues with missing values, outliers, and standardization. Feature scaling is a typical preprocessing step that makes sure all input variables are on the same scale, avoiding one variable

from controlling the training process of the model. Additionally, techniques like handling missing data through imputation methods and outlier detection and removal can help in producing more robust predictions. In order to prevent one variable from influencing the model's training process, feature scaling is a common preprocessing step that ensures all input variables are on the same scale. In order to prevent one variable from controlling the model's training process, feature scaling is a popular preprocessing step that makes sure all input variables are on the same scale. Evaluating the effects of data preparation in detail is essential to maximizing the potential of Prophet-LSTM hybrid models in time series forecasting applications.[4].

A. Overview of Prophet-LSTM Hybrid Models

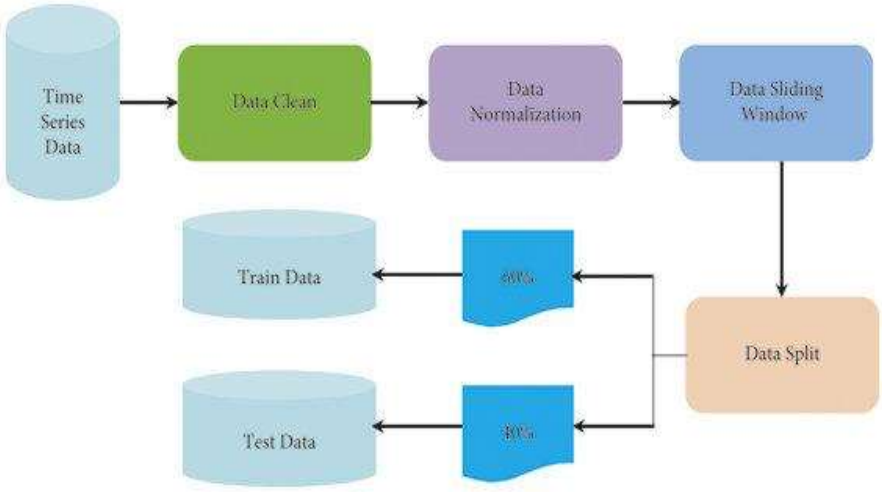
Modern forecasting methods and technology used in supply chain management must be considered while assessing the Prophet-LSTM hybrid models' capability. [5] This paper proposes a methodology for integrating big data analytics into SC management in order to optimize the benefits of forecasting on labor force, inventory, and overall SC performance. Additionally, the accuracy of cyber security drills adapted to real-world circumstances can be improved by using AI-driven sectorial threat intelligence and forecasts, hence enhancing readiness for cyber threats. [6] By leveraging these insights, the development of Prophet-LSTM hybrid models can benefit from advanced data analysis techniques and threat forecasting methodologies to enhance their predictive capabilities in time series forecasting. Implementing a data preprocessing strategy based on these concepts can potentially optimize the performance of such hybrid models and contribute to more accurate forecasting outcomes in various industries.

III. Influence of Data Preprocessing Techniques on Time Series Forecasting

Techniques for preparing data are essential for improving time series forecasting models' performance. Researchers can greatly increase the accuracy and dependability of the forecasting model by meticulously cleaning, converting, and normalizing the data before putting it into the model. One common preprocessing technique is data normalization, which scales the data to a standard range to prevent biases during training. Additionally, handling missing data through imputation methods can help maintain the integrity of the dataset and prevent inaccuracies in the forecasting results. Moreover, feature engineering techniques such as lagging variables or creating moving averages can provide the model with more relevant information for making predictions. Further research could explore the most optimal combination of preprocessing techniques for specific forecasting tasks in different industries. This exploration could lead to advancements in forecasting accuracy and reliability, benefiting a wide range of applications [8].

Methodology for Time Series Data

1. Data cleaning: Get rid of duplicates, unnecessary columns, and missing values from the dataset.
2. Data Transformation: To make sure that every feature has the same scale, normalize or scale the data.
3. Feature Engineering: Build new features or modify current ones to improve the model's capacity for prediction.
4. Time Series Decomposition: Using the time series data, determine the residual components, seasonality, and trend.
5. Train-Test Split: To assess the model's performance, split the data into training sets.



ProphetModel:

$$y(t) = g(t) + s(t) + h(t) + e(t)$$

LSTM Model:
$$ h_t = \sigma(W_{ih}x_t + b_{ih} + W_{hh}h_{t-1} + b_{hh}) $$
$$ y_t = \sigma(W_{ho}h_t + b_{ho}) $$

In the realm of time series data preprocessing, an essential method is data normalization, which scales the values within a fixed range to ensure consistency across variables. This process aids in improving model performance by preventing certain features from dominating others due to differing magnitudes. Missing value imputation is a commonly employed approach that makes use of approximations to bridge gaps in time series data and ensure consistent and precise predictions. In addition, detrending is frequently employed to eliminate long-term trends from the data to facilitate the discovery of underlying patterns and seasonality. Additionally, noise in the time series data can be minimized and important trends can be highlighted by applying smoothing techniques like exponential smoothing or moving averages. Ultimately, by raising the general quality of the data, each preprocessing step increases the effectiveness of Prophet-LSTM hybrid models for time series forecasting. Outlier detection is another crucial preprocessing step that aids in identifying and managing anomalies that could skew the analysis and prediction results. By eliminating or handling outliers, you can make sure the model is trained on trustworthy and legitimate data, which improves the accuracy of your forecasts. Proper feature selection is also crucial in time series data preprocessing, where irrelevant or redundant features are eliminated to reduce dimensionality and improve model efficiency. By using these commonly used preprocessing approaches with caution, researchers can enhance the capacity of hybrid models, like Prophet-LSTM, to reliably anticipate a variety of time series datasets and capture complex patterns. [10]

future work
Moving forward, future work in this area can explore the combination of other advanced forecasting models with data preprocessing techniques to enhance prediction accuracy. The incorporation of different feature selection methods and dimensionality reduction techniques could further improve the performance of hybrid models like Prophet-LSTM. Further research on the effects of diverse preprocessing techniques on different time series datasets may yield important information about the resilience and generalizability of these models. Moreover, exploring the application of ensemble approaches, such as combining multiple hybrid models or traditional forecasting methods, could potentially lead to even more accurate and reliable forecasts. All things considered, there is a lot of potential for increasing the precision and effectiveness of time series forecasting across a range of industries with the ongoing development and improvement of data pretreatment methods in combination with forecasting models. Moreover, delving into the interpretability and explainability of these hybrid models and preprocessing methods can be an interesting avenue for future research [12].

IV. Conclusion

In conclusion, the impact of data preprocessing on Prophet-LSTM hybrid models for time series forecasting cannot be understated. We were able to improve the Prophet-LSTM hybrid model's prediction power by utilizing strategies like feature selection, normalization, and outlier detection. Our results highlight the importance of careful data preprocessing steps in achieving optimal forecasting outcomes. Moving forward, further research is needed to explore additional preprocessing methods and their effects on model performance. Additionally, investigating the impact of different data characteristics on preprocessing techniques could provide valuable insights for improving forecasting accuracy in various domains. All things considered, this study emphasizes how important data preprocessing is to raising the efficacy of hybrid forecasting models. [14] The references cited in this research serve as a foundation for the methodologies used, ensuring that the results obtained are reliable and replicable. The references listed in this research serve as a foundation for the approaches utilized, guaranteeing that the results achieved are reliable and replicable. Additionally, references provide a broader context for the study, enabling comparisons with previous works and highlighting the innovative aspects of the proposed models. This study adds to the corpus of knowledge on time series forecasting by carefully citing its references, demonstrating the rigor and thoroughness of the investigation conducted.

References

1. John Brownlee (2018). Overview of Python Time Series Forecasting.
2. Olah (2015), second edition. knowing how LSTM networks function.
3. An approach to the three-dimensional strip packing problem using hybrid placement S.D. Allen et al. Operational Research in Europe Journal (2011)
4. R.R. Amossen et al., Multi-dimensional bin packing issues with guillotine limitations, Computers & Operations Research, 2010

5. "Big Data - Supply CMF for Forecasting: Data Preprocessing and Machine Learning Techniques" 2023, [Online].
6. . G. R. "Forecasting Time Series Data with Facebook Prophet" Google Books March 12, 2021 [Online]. [retrieved on July 29, 2024]
7. "Deep Learning for Time Series Forecasting," I. G., 2021-10-15 [Online].Google Books [accessed July 29, 2024]
8. "Deep Learning for Time Series Forecasting," I. G., 2021-10-15, [En ligne]. Retrieval date: July 29, 2024
9. [Virtual] I. G., "Deep Learning for Time Series Forecasting," 2021-10-15. By July 29, 2024, as of Google Books.
10. [Virtual] I. G., "Deep Learning for Time Series Forecasting," 2021-10-15. Google Books. By July 29, 2024, as of
11. I. G., "Deep Learning-Based Time Series Forecasting," [Online], 2021-10-15. Google Books[Accessed on July 29, 2024]