Store Demand Forecasting using Time-Series and Neural Networks

Authors:

Aravinda Raman Jatavallabha, Pankaj Manoharlal Thakur, Suraj Raghu Kumar

North Carolina State University, Raleigh

Introduction

Pivotal role in supply chain management and retail operations.

Preventing stockouts or overstocking.

Optimize inventory levels, allocate resources effectively, and enhance overall operational efficiency.

Neural Networks and time series forecasting techniques play a substantial

Goal of the Project

- Constructing a forecasting model for predicting future sales in various stores.
- Using machine learning algorithms and time series analysis techniques for accurate predictions.

Importance of Demand Forecasting

- Improved Planning
- Increased Efficiency
- Better Service
- Enhanced Performance
- Competitive Advantage

Related works

Authors	Work	Description
A. Krishna, A. V, A. Aich and C. Hegde	Sales-forecasting of Retail Stores using Machine Learning Techniques [1]	The paper compares various machine learning techniques for retail sales forecasting, finding Gradient Tree Boosting to be the most effective. It stresses the importance of accurate forecasting for businesses and the need for proper model selection and hyperparameter tuning.
Y. Ali and S. Nakti	Sales Forecasting: A Comparison of Traditional and Modern Times-Series Forecasting Models on Sales Data with Seasonality [2]	The paper compares traditional and modern time-series forecasting models for sales data, highlighting SARIMA's effectiveness in capturing fast-moving sales with seasonal patterns and LSTM's potential for further enhancement. It emphasizes the importance of accurate forecasting for business planning and the need for selecting models tailored to specific business requirements.

Related works

Authors	Work	Description
F. -C. Yuan and C. -H. Lee	Sales Volume Forecasting Decision Models [3]	The paper compares various machine learning techniques for retail sales forecasting, finding Gradient Tree Boosting to be the most effective. It stresses the importance of accurate forecasting for businesses and the need for proper model selection and hyperparameter tuning.
B. Lakshmanan, P. S. N. -V. Raja and V. Kalathiappa	Sales Demand Forecasting Using LSTM Network [4].	The paper proposes an LSTM-based sales forecasting model, demonstrating its superiority over conventional methods like MLP regressor and linear regression. It achieves 96.77% accuracy with real-time sales data from southern India, offering improved performance and addressing forecasting challenges for businesses.

Dataset Description

Data Source:

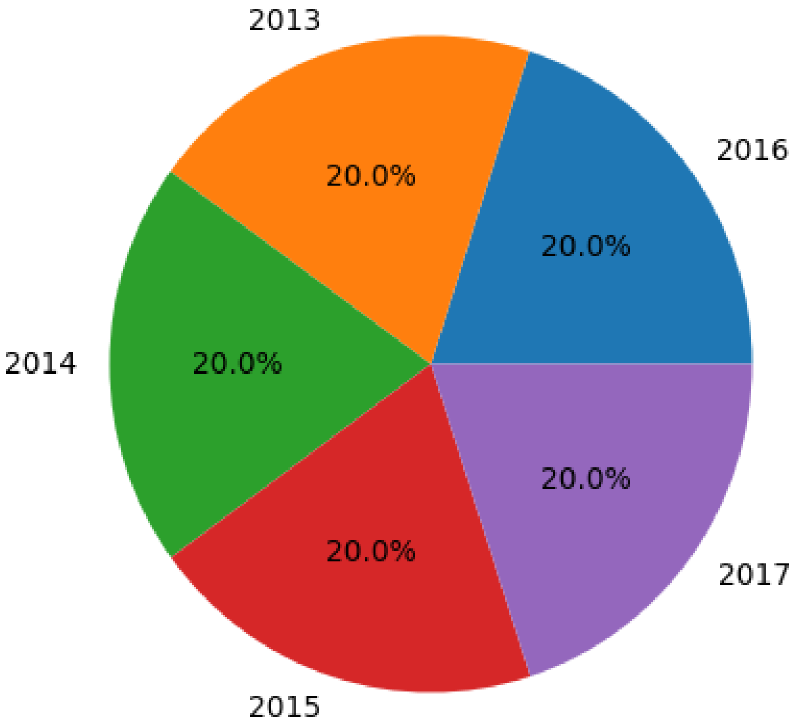
- We’ve used data obtained from the Store Item Demand Forecasting Challenge dataset from Kaggle. [5]
- It contains 4 attributes related to Date, Store ID, Item ID and Sales (no. of units sold).
- Sales were recorded across 10 different stores, for 50 different items from 01/01/2013 to 12/31/2017

Understanding the Data

DATE	STORE	ITEM	SALES
2013-01-01	1	1	13.0
2013-01-01	2	3	100.0
2013-01-02	1.	1	20.0
2013-01-02	2	3	78.0
2014-01-01	3	2	10.0
2014-01-02	5	38	28.0
2014-01-03	10	50	27.0

Data Preprocessing

- Verified absence of duplicates, ensuring data integrity.
- Transformed date format for consistency and analysis.
- Confirmed null values absence, ensuring data completeness.
- Acknowledged weak correlations with sales.
- Adopted inclusive approach, incorporating all attributes.
- Aimed to leverage interactions for enhanced model efficacy.
- Sorted data chronologically, with 75%/25% train/test split.



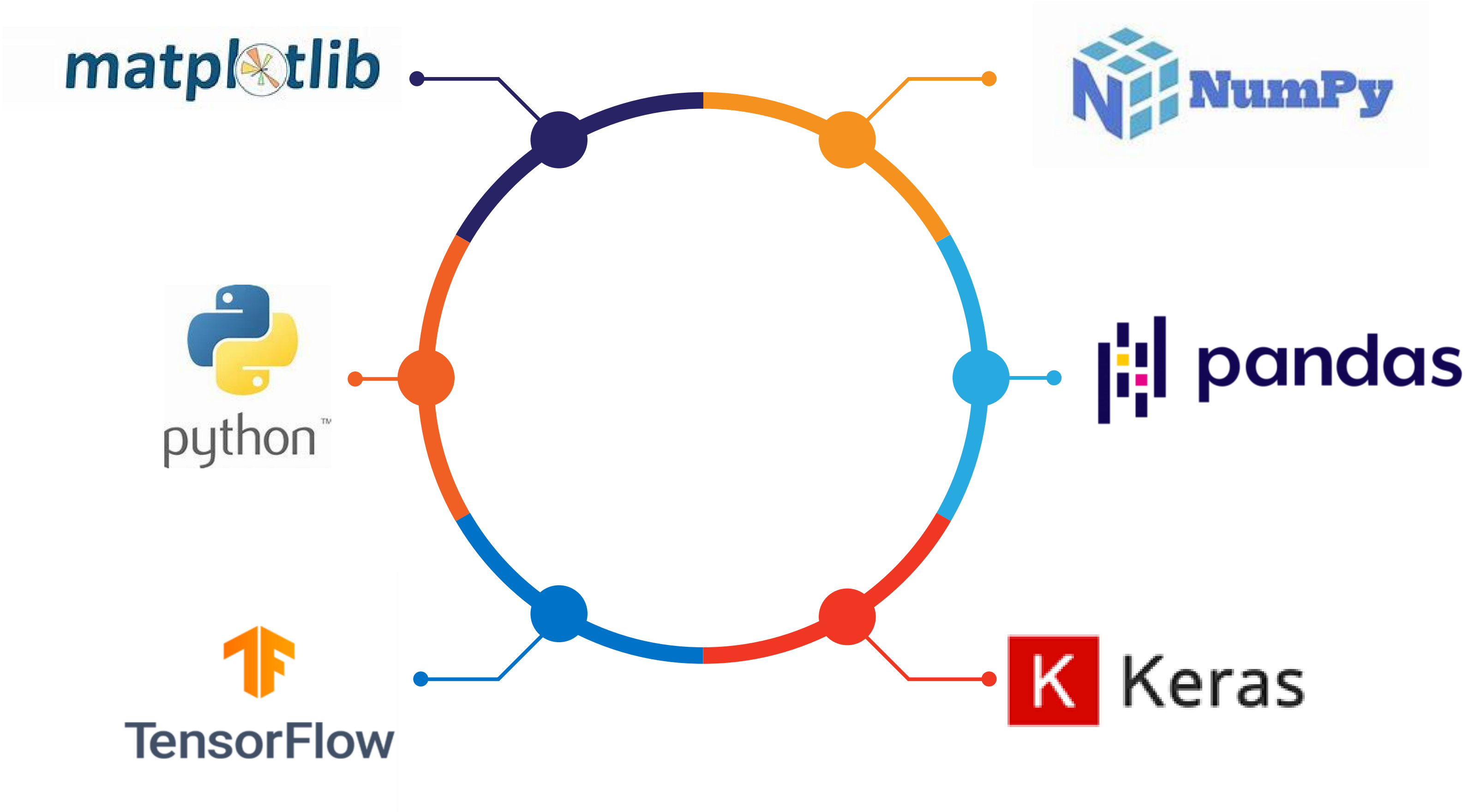
Data Distribution

Pre-Processing: Data Transformation

Independent Vars.	Description	Initial Format	Transformed format	Var. Type
DATE	Scheduled departure, local time	yyyy-mm-dd	1,2,3….100	Discrete, Ordinal
STORE	Store ID	1, 2, 3,…,10	0/1 and 1, 2, 3,…, 10 *	Discrete, Ordinal
ITEM	Item ID	1, 2, 3,….., 50	0/1 and 1, 2, 3,…, 50 *	Discrete, Ordinal

* Tried One-hot encoding, but Label Encoded Store IDs turned out better

Tools and Technologies used

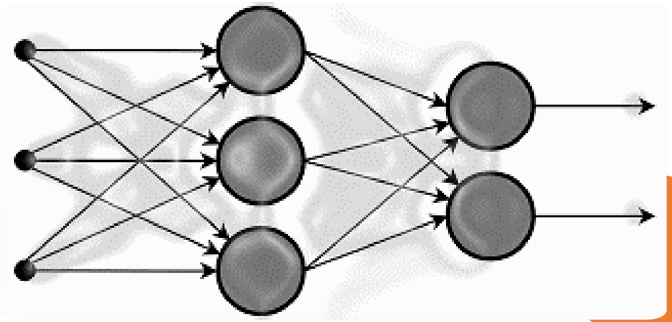


Method

- Selection of advanced time series models: Bi-LSTM, LSTM+CNN and BiLSTM+CNN chosen for their proficiency in capturing complex temporal patterns.
- Incorporation of baseline regression models: Random Forest Classifier, XGBoost, Artificial Neural Network, Convolutional Neural Networks (CNN), ARIMA and LSTM for benchmarking.
- Advantages of LSTM and Bi-LSTM: adeptness in handling vanishing gradient problems and capturing long-term dependencies.
- Mitigation of vanishing gradient issue: employment of techniques like gradient clipping and batch normalization in LSTM models.
- Expected superiority of time series models, especially LSTM and its variants, in predicting sales due to their capacity to capture temporal dependencies, offering valuable insights into machine learning approaches for sales prediction.

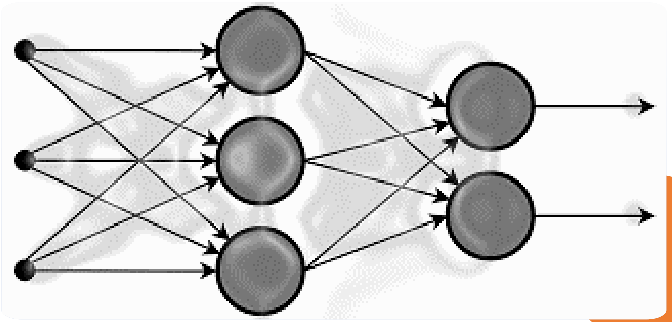
Modeling

- The baseline machine learning algorithms used in this study are:
  - **Random Forest Regressor**
  - **Convolutional Neural Networks (CNN)**
  - **XGBoost Regressor**
  - **Artificial Neural Network (ANN)**



Modeling

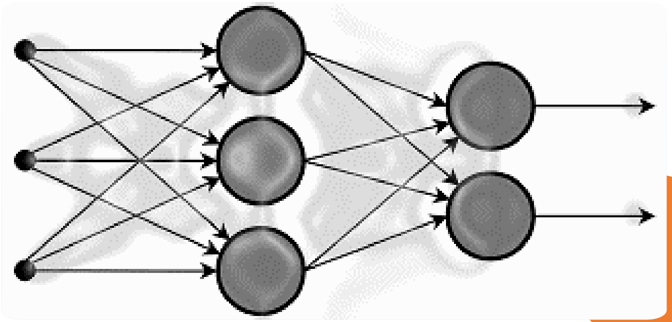
- The time-series machine learning algorithms used in this study are:
  - **Long short-term memory (LSTM)**
  - **ARIMA**
  - **Bidirectional Long Short-Term Memory (BiLSTM)**



Modeling

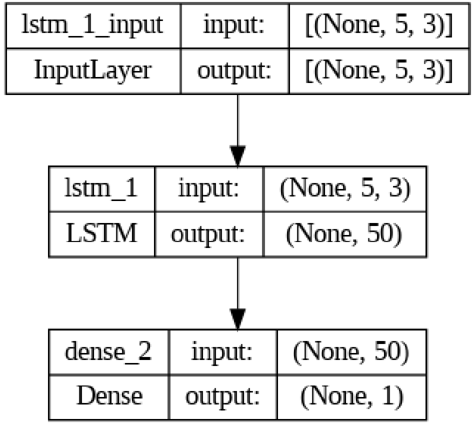
Our Final Ensemble Models

- **LSTM + CNN Hybrid Model**
- **BiLSTM + CNN Hybrid Model**



Time-Series Architecture

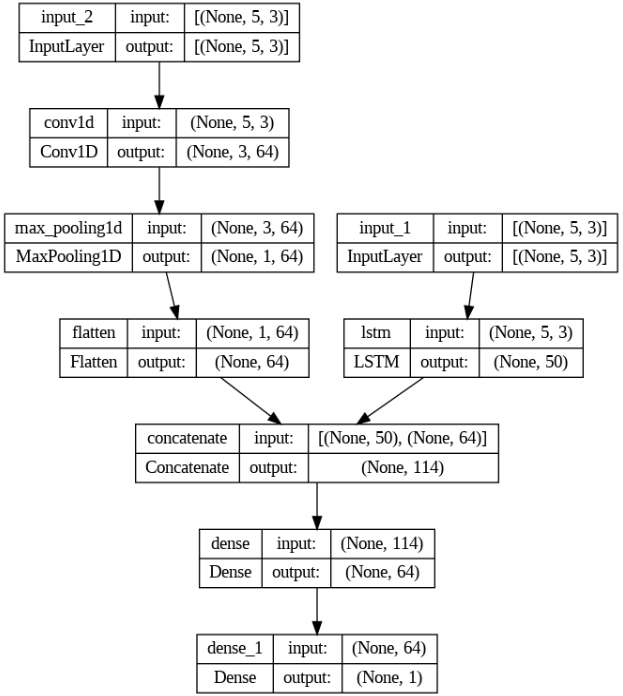
LSTM



Hand tuned batch size (256) and LSTM hidden nodes (50)
'Adam' optimizer and ReLU activation function were used

LSTM/CNN Hybrid

(Our model)



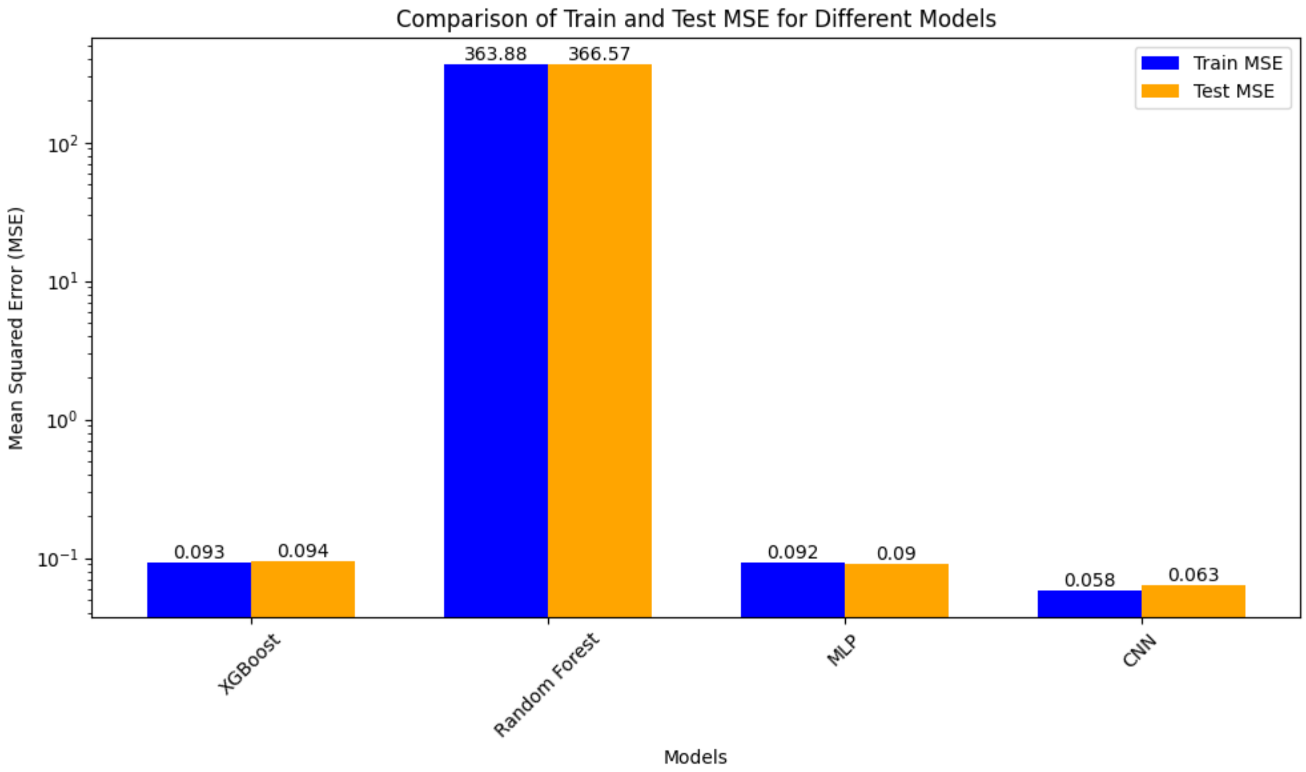
Training and Evaluation Metrics

Mean Squared Error (MSE)

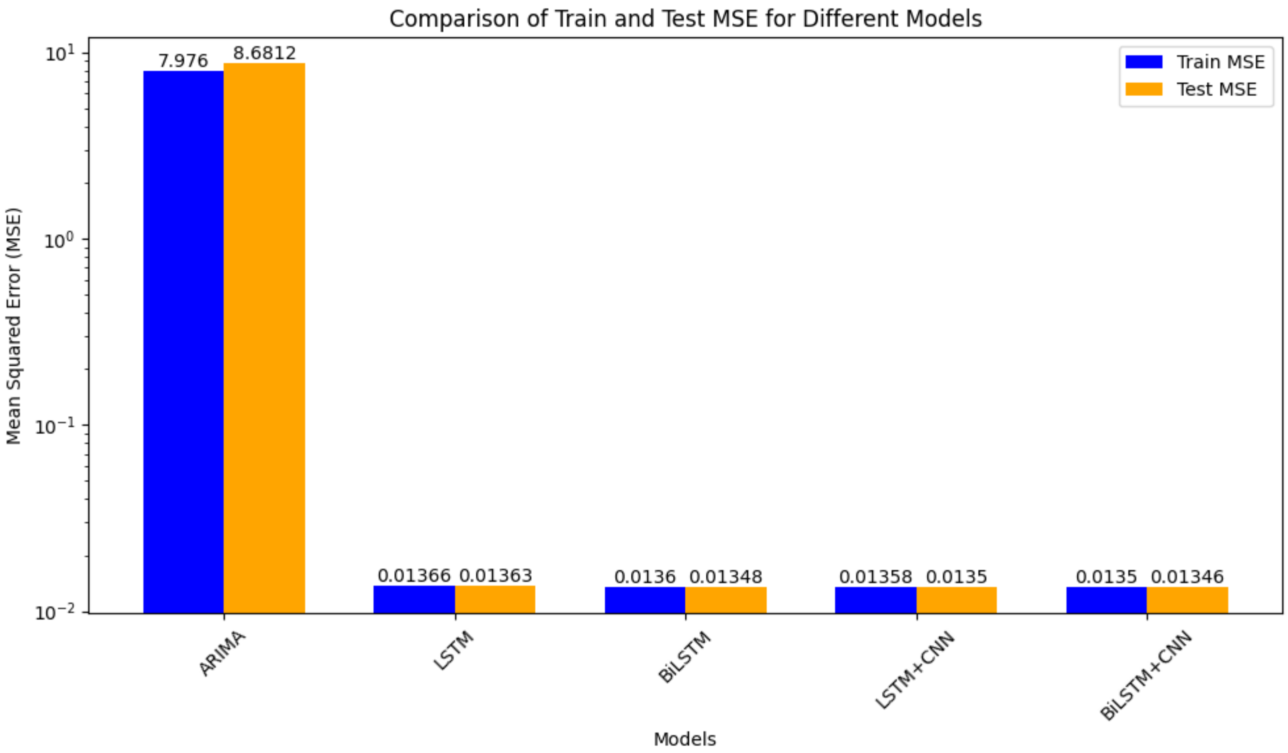
$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

- MSE quantifies the average squared magnitude of errors between predicted and actual values
- Useful for training models because it gives higher weighting to lower frequency values.

Results: Baseline Models



Results: Time-Series Models



Discussion and Future Scope

- BiLSTM+CNN model showed the best performance in sales prediction.
- BiLSTM and LSTM+CNN hybrid models followed, indicating promising results.

- Possible issues
  - **Limited Attribute Coverage**: The dataset's restriction to only three attributes—Date, Store ID, and Item ID—might have hindered the models' ability to capture the full spectrum of factors influencing sales, potentially overlooking crucial variables such as promotions, competitor activity, or seasonal trends.

  - **Lack of Granularity**: With only three attributes available, the dataset may lack the granularity needed to accurately capture the intricacies of sales patterns, potentially leading to oversimplified model representations and reduced predictive accuracy.

- In the future, advanced evaluation techniques to assess model performance in capturing temporal dynamics and other complexities of sales data can be explored in addition to obtaining more information on the dataset.

References

[[1]A. Krishna, A. V, A. Aich and C. Hegde, Sales-forecasting of Retail Stores using Machine Learning Techniques, 2018, 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 160-166, doi: 10.1109/CSITSS.2018.8768765.

[2] F. Wieland. Limits to growth: results from the detailed policy assessment tool [air traffic congestion]. In 16th DASC. AIAA/IEEE Digital Avionics Systems Conference. Reflections to the Future. Proceedings, volume 2, pages 9.2–1–9.2–8 vol.2, Oct. 1997.

[3] F. -C. Yuan and C. -H. Lee, "Sales Volume Forecasting Decision Models," 2011 International Conference on Technologies and Applications of Artificial Intelligence, Chung Li, Taiwan, 2011, pp. 239-244, doi: 10.1109/TAAI.2011.49. keywords: {Forecasting;Marketing and sales;Genetic algorithms;Support vector machines;Predictive models;Artificial neural networks;Biological cells;Sales volume forecasting;Support vector regression;Genetic algorithm;Artificial neural networks}.

[4] Lakshmanan, B., Vivek Raja, P.S.N., Kalathiappan, V. (2020). Sales Demand Forecasting Using LSTM Network. In: Dash, S., Lakshmi, C., Das, S., Panigrahi, B. (eds) Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing, vol 1056. Springer, Singapore. https://doi.org/10.1007/978-981-15-0199-9_11

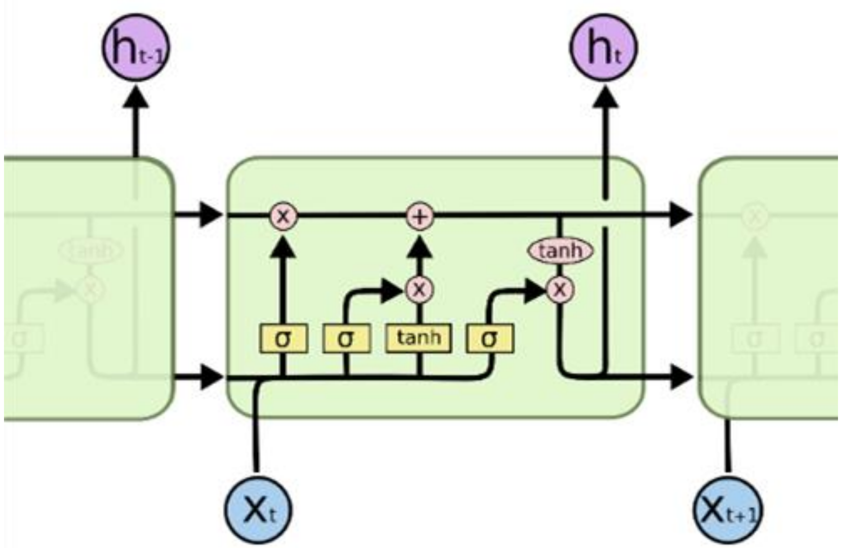
[5] https://www.kaggle.com/competitions/demand-forecasting-kernels-only

THANK YOU

APPENDIX

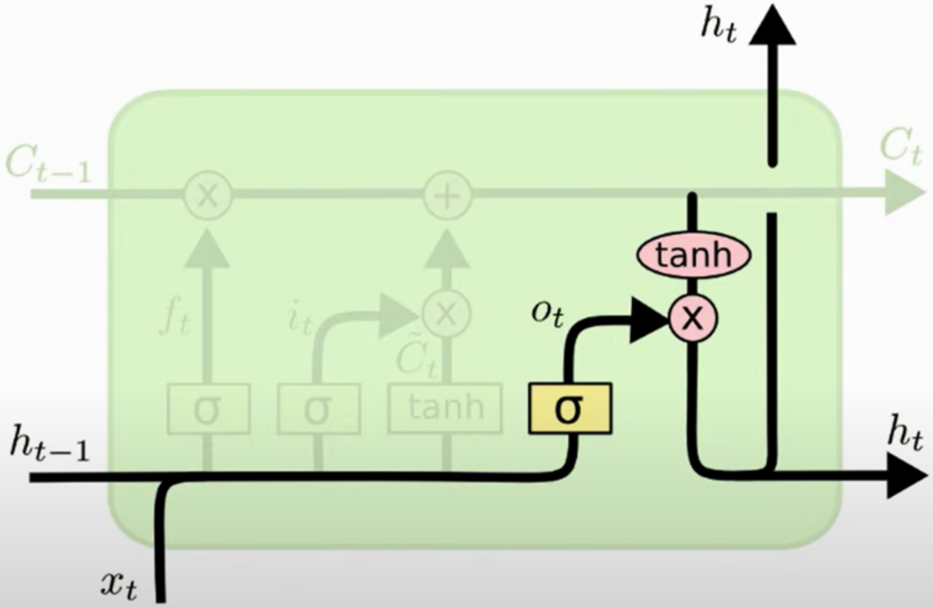
LSTM Model

- Long Short-Term Memory (LSTM) models provide a powerful framework for modeling sequential data, such as time series prediction tasks like flight delay prediction.
- LSTMs are a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem and capture long-term dependencies in sequential data



The repeating module in an LSTM contains four interacting layers. [7]

LSTM



Three Gates of LSTM Cell:

- **Input Gate** Is Cell Updated?

$$i^{(t)} = \sigma\left(W^i\left[h^{(t-1)}, x^{(t)}\right] + b^i\right)$$

- **Forget Gate** Is memory set to 0?

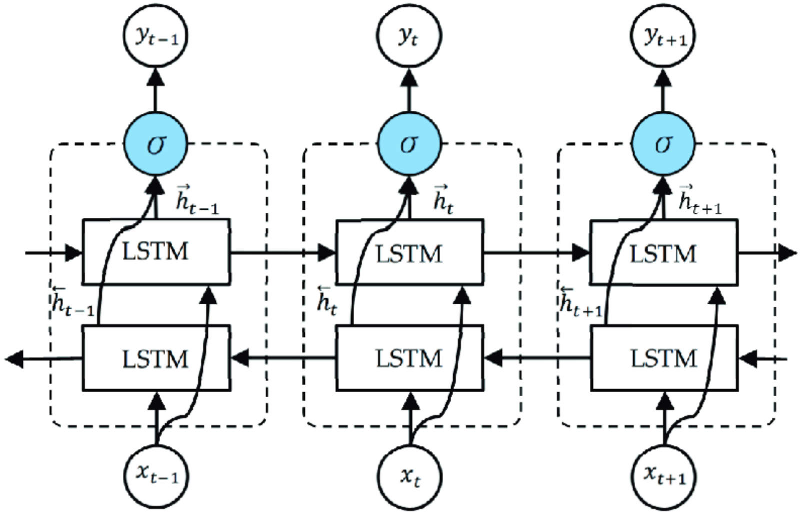
$$f^{(t)} = \sigma\left(W^f\left[h^{(t-1)}, x^{(t)}\right] + b^f\right)$$

- **Output Gate** Is current info visible?

$$o^{(t)} = \sigma\left(W^o\left[h^{(t-1)}, x^{(t)}\right] + b^o\right)$$

BiLSTM Model

- Bidirectional Long Short-Term Memory (BiLSTM) models extend the capabilities of LSTM by processing input sequences in both forward and backward directions.
- This allows the model to capture contextual information from both past and future time steps, making it particularly effective for tasks requiring a comprehensive understanding of sequence data.



The unfolded architecture of Bidirectional LSTM (BiLSTM) with three consecutive steps.