Deep Learning Models for Time Series Forecasting: Applications in SKU Forecasting and Soft Sensing

Swapnil Sharma
Roll No : CH20B024

Guide : Dr. M. Nabil

भारतीय प्रौद्योगिकी संस्थान तिरुपति

TIRUPATI

Department of Chemical Engineering

May 10, 2024

Outline

- Introduction
- Problem Statement
- Time Series Models (Phase-I)
- Machine Learning Models
  - Light Gradient Boosting Machine (Phase-II)
- Deep Learning Models
  - Recurrent Neural Networks (Phase-II)
  - Long Short-Term Memory (Phase-II)
- Case Study
  - Case Study 1: Simulated Data (Phase-II)
  - Application 1: SKU Forecasting (Phase-I &Phase-II)
  - Application 2: Soft Sensing (Phase-II)
- Conclusion and Potential Extensions

Introduction – SKU Forecasting and Soft Sensing

- The primary challenge in SKU forecasting lies in the fluctuating and intricate factors impacting product demand, including seasonality, promotions, and external market dynamics.
- **Objective 1**: To accurately forecast SKU demand by addressing the challenges posed by factors like seasonality, promotions, and external market dynamics.
- The challenge arises in industrial processes when certain variables are hard to measure or cannot be measured due to difficulties, high costs, or impracticality.
- **Objective 2**: To estimate or infer process variables indirectly which are difficult, impractical and costly to predict without measuring them directly.
- Pointers from literature survey:
  - [Perez et al., 2022] used RNN and Transformer architecture to forecast the SKU on Corporacion Favorita dataset.
  - [Jicheng Li et al., 2023] used LSTM to predict the butane concentration at the bottom of the debutanizer column.
  - [Curreri F. et al., 2021] used RNN and LSTM for predicting H_2S and SO_2 concentration in SRU (Sulphur Recovery Unit).
- **Our focus:** To provide solutions for SKU forecasting and soft sensing by leveraging advanced deep learning techniques.

Time Series Models

Auto Regressor (AR) model

$$y_t = \sum_{i \epsilon S} \propto_i y_{t-i} + \varepsilon_t$$

Moving Average (MA) model

$$y_t = \sum_{j \epsilon S_1} \theta_j \varepsilon_{t-j} + \varepsilon_t + \mu$$

Auto-regressive Moving Average (ARMA) Model

$$y_t = \sum_{i \epsilon S} \propto_i y_{t-i} + \sum_{j \epsilon S_1} \theta_j \varepsilon_{t-j} + \varepsilon_t + c$$

Auto-regressive Integrated Moving Average (ARIMA) Model

$$(1 - \sum_{i \epsilon S} \alpha_i B^i)(1 - B)^d y_t = (1 + \sum_{j \epsilon S_1} \theta_j B^j)\varepsilon_t$$

S is set of significant lags and S_1 is the set of significant lagged error terms or white noise and the B is the backshift operator. c and b are the respective constant terms of the AR and ARMA model, μ is the mean of the time series and y_t is the SKU sales at time t. ε_t is the white noise error or residuals at time t.

PACF and ACF

- While PACF calculates the correlation after removing the influence of the intermediate lags $\boldsymbol{y}_{t-1}$, $\boldsymbol{y}_{t-2}$…… $\boldsymbol{y}_{t-k+1}$ when you calculate the partial auto correlation between $\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-k}$.
- PACF and ACF helps you to find the order of the AR and MA model respectively and also helps to find the significant lags.

Partial Auto Correlation Function

$$\phi_{kk} = \frac{\rho_k \; - (\sum_{i=1}^{k-1} \phi_{k-1,i} * \rho_{k-1})}{(1 - \sum_{i=1}^{k-1} \phi_{k-1,i} * \rho_j)}$$

where, $\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}$

$\phi_{k,j}$ is the PACF between $\boldsymbol{y}_{t-k}$ and $\boldsymbol{y}_{t-j}$. .

- There is always dependence of $\boldsymbol{y}_{t-1}$, $\boldsymbol{y}_{t-2}$ ….. $\boldsymbol{y}_{t-m}$ on $\boldsymbol{y}_t$ where m is the maximum possible lags. So, while calculating the auto correlation between $\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-k}$ it gives you the correlation after having the influence of intermediate lags $\boldsymbol{y}_{t-1}$, $\boldsymbol{y}_{t-2}$…… $\boldsymbol{y}_{t-k+1}$.

Auto Correlation Function

$$\rho_k = \frac{cov(y_t, y_{t-k})}{\sqrt{var(y_t) var(y_{t-k})}}$$

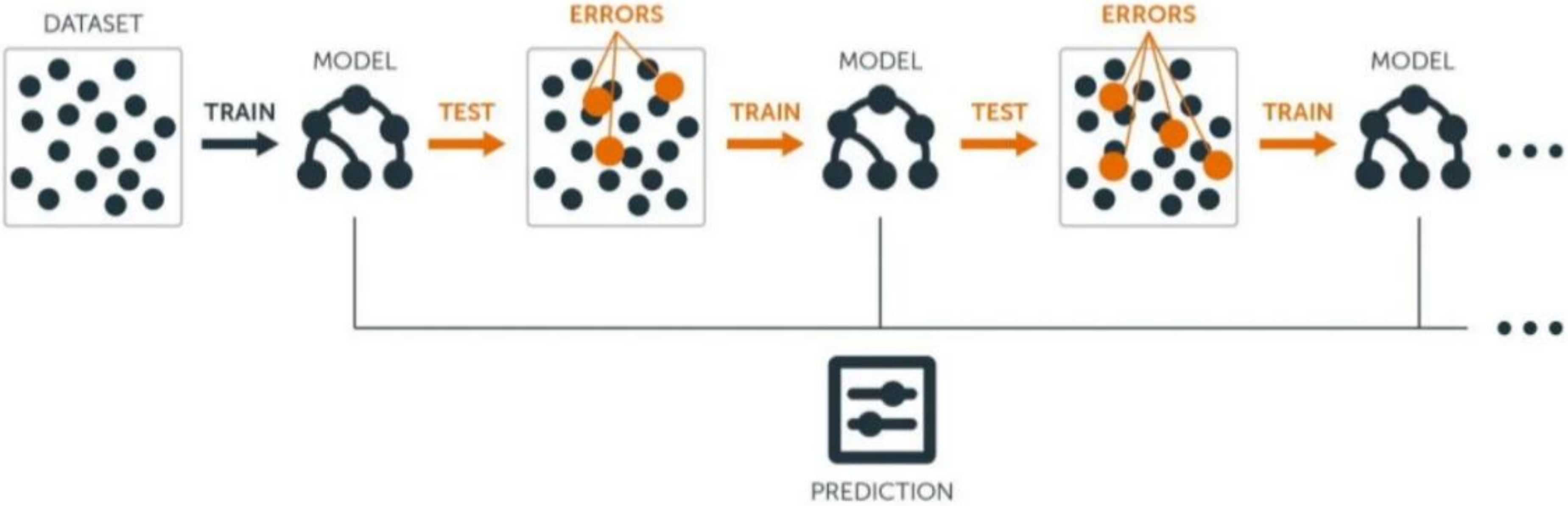
ρ_k is the ACF between $\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-k}$.

Light Gradient Boosting Machine (LGBM)

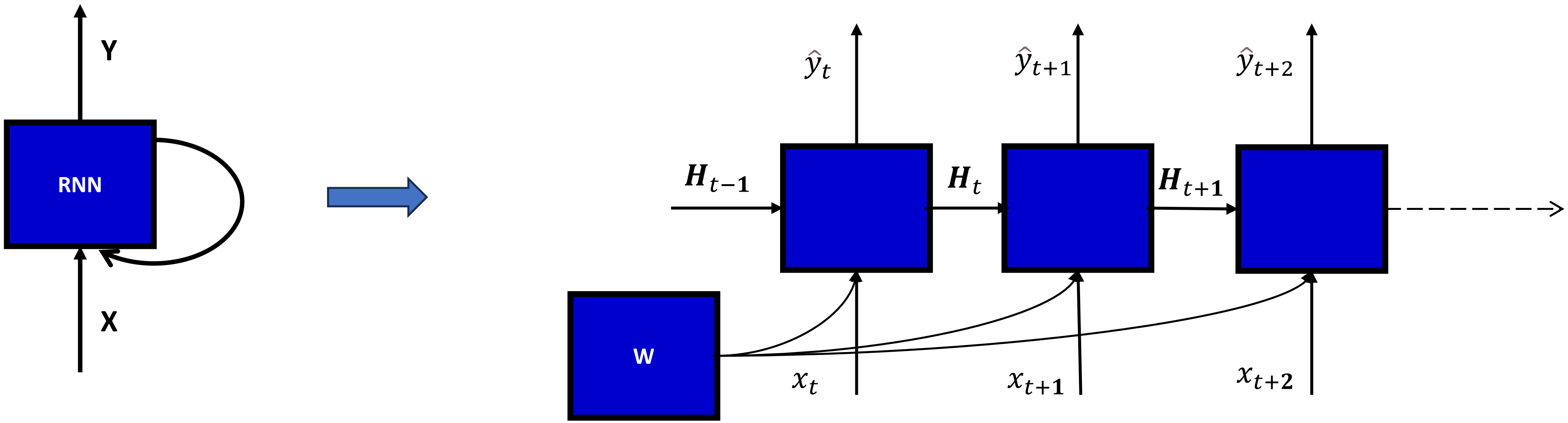
LightGBM is an ensemble learning technique. LightGBM uses three novel features: Gradient-based One Side Sampling (GOSS), Bin way of splitting and Exclusive Feature Bundling (EFB).

LightGBM has several advantages:

- Faster training speed and higher efficiency.
- Better accuracy.
- Compatibility with Large Datasets.



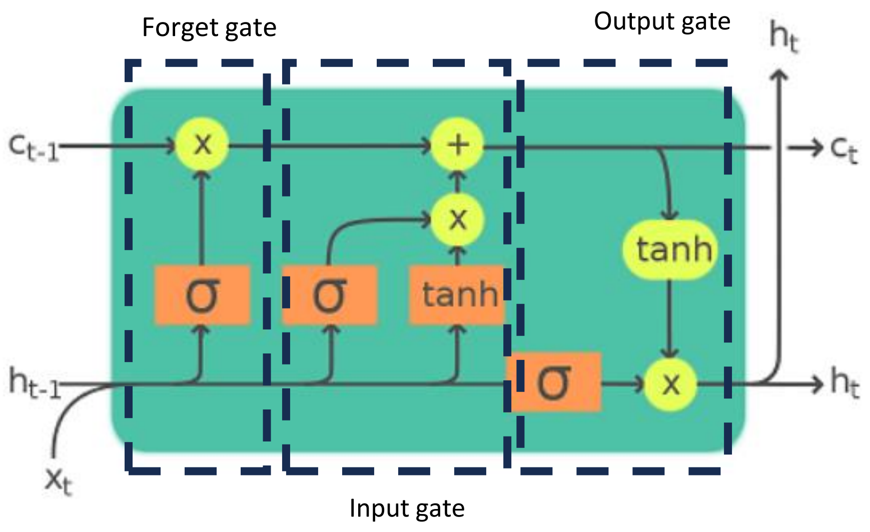
Recurrent Neural Networks (RNNs)



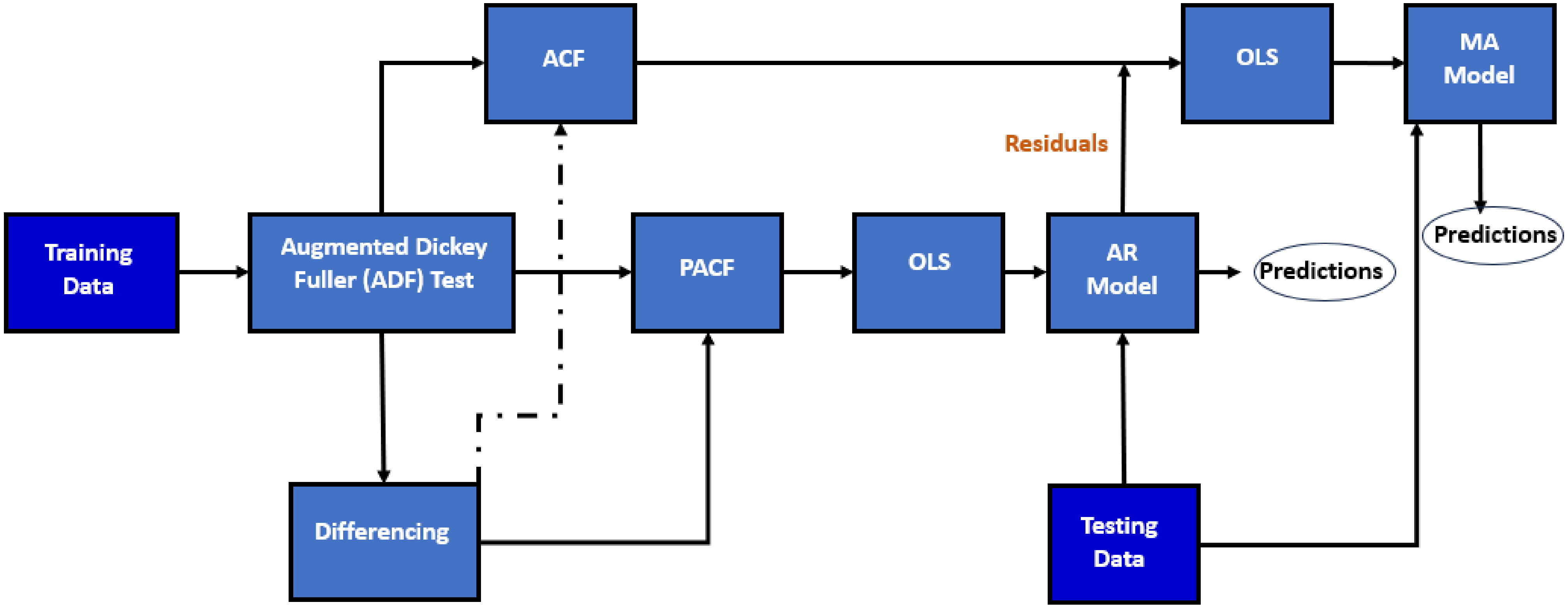
- H is the hidden layer or the hidden state.
- W_{xh} , W_{hh} and W_{hy} are the weight matrixes which are same for every hidden layer and its get updated by the backpropagation algorithm.
- $H_t = f_w(H_t, x_t)$ or $H_t = \tanh(W_{hh}H_{t-1} + W_{xh}x_t + b)$
- $\hat{y}_t = g(W_{hy}H_t + b_1)$ where g is the SoftMax for multi class classification, sigmoid for binary class classification and linear function for the regression.

Long Short-Term Memory (LSTM)

- $o_t = \sigma(Wh_{t-1} + Ux_t + b_0)$ where o_t is the **output gate.**
- $f_t = \sigma\left(Wh_{t-1} + Ux_t + b_f\right)$ where i_t is known as **forget gate**.
- $i_t = \sigma(Wh_{t-1} + Ux_t + b_i)$ where i_t is known as **input gate**.
- $\widetilde{c_t} = \sigma(Wh_{t-1} + Ux_t + b_c)$ where $\widetilde{c_t}$ is known as **cell state.**
- $c_t = f_t \circ c_{t-1} + i_t\, \widetilde{c_t}$
- $h_t = o_t \circ \tanh(c_t)$



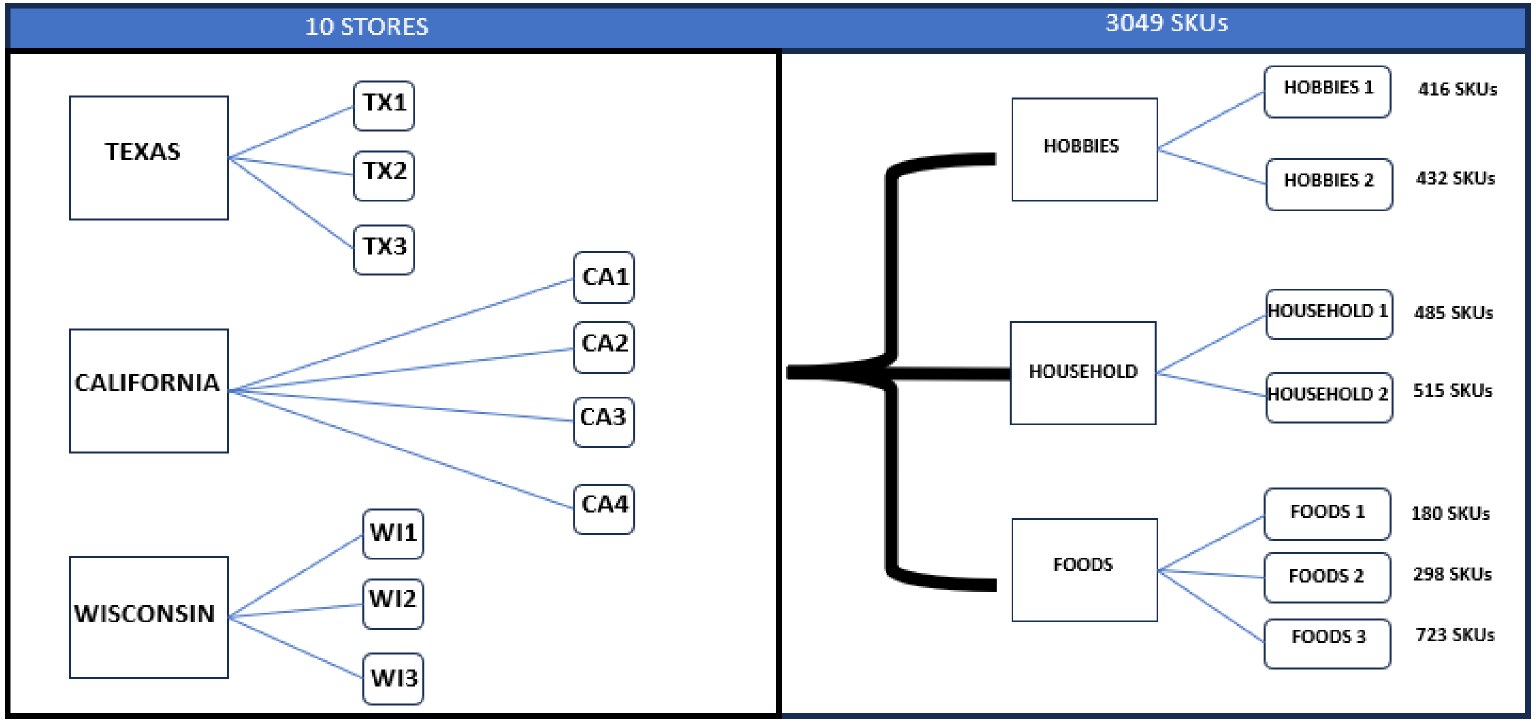
Methodology



OLS is the Ordinary Least Squares method to get the coefficients of the models. **ADF Test** helps to check whether the time series is stationary or not. Non-stationary time series will go through differencing to convert in stationary one.

Case Study and Dataset Description

- **Walmart Dataset**

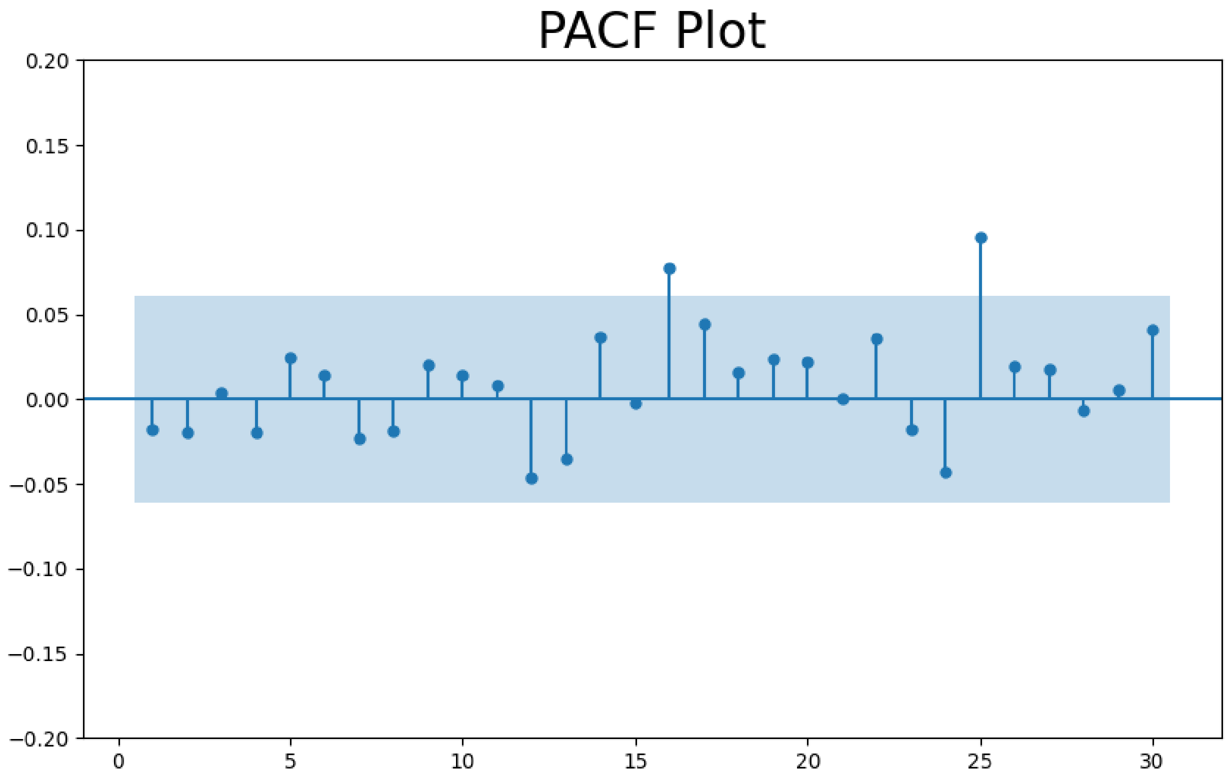


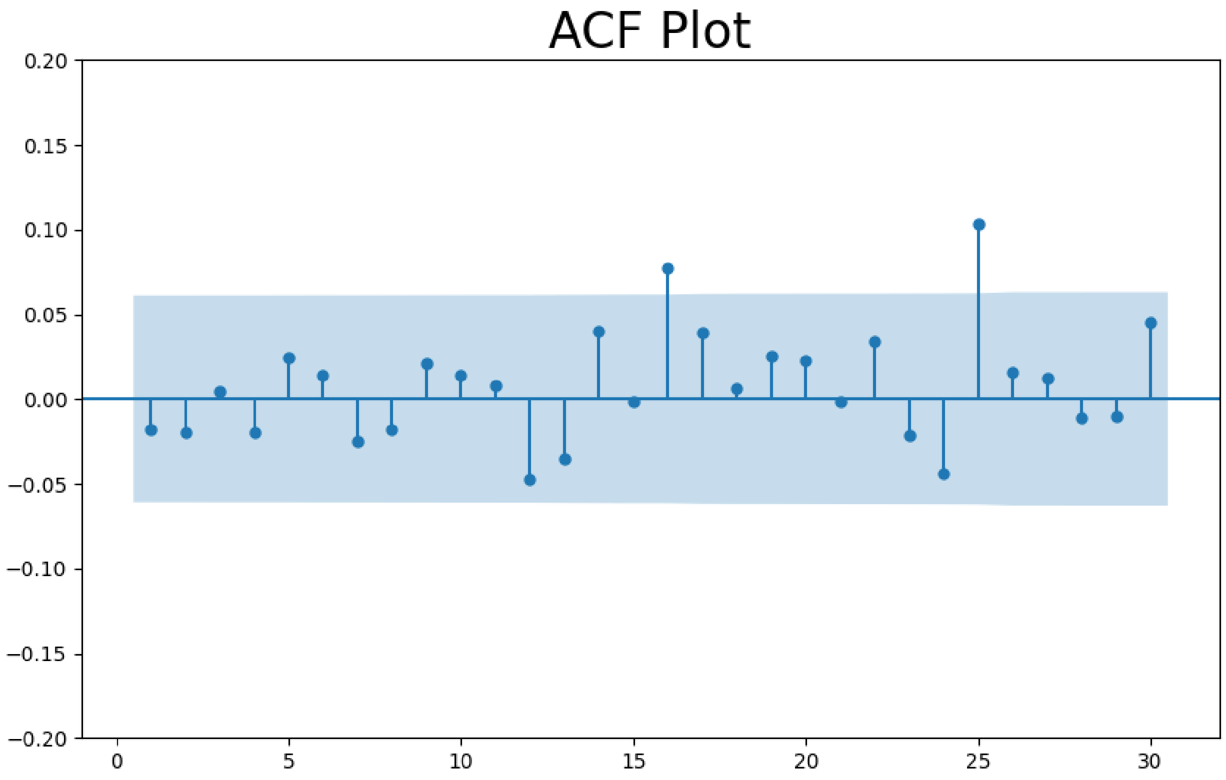
Types of Features	
ID	Special Offer
State ID	Units Sold
Store ID	Day
Dept ID	Sell Price
Category ID	Month
Holiday Name	Year
Holiday Type	

Datasets consist of all information of public holidays, promotional events and sell price of the 1941 days. 1913 days for the training and 28 days for testing. It is classic M5 Forecasting dataset of Walmart stores taken from [Kaggle.]

Results (Walmart Dataset)

The table of the performance of the respective time series models on the training set is created. The forecasting is done on the store id = "CA_2" and category id = "HOBBIES_1_001". Same has been done for the testing set.. From the ACF and PACF we got the significant lags to be t-16 i.e. of 16 days and t-25 i.e. of 25 days for the AR and for MA model errors to be considered at day t and t-25.





Results (Walmart Dataset)

Model Name	RMSE	MSE	MAE
AR	0.835	0.6975	0.665
MA	0.066	0.004	0.052
ARMA	0.065	0.004	0.052

Table 1: Performance Table of Training Set.

Model Name	RMSE	MSE	MAE
AR	0.845	0.714	0.7477
MA	0.084	0.007	0.064
ARMA	0.0865	0.007	0.066

Table 2: Performance Table of Testing Set.

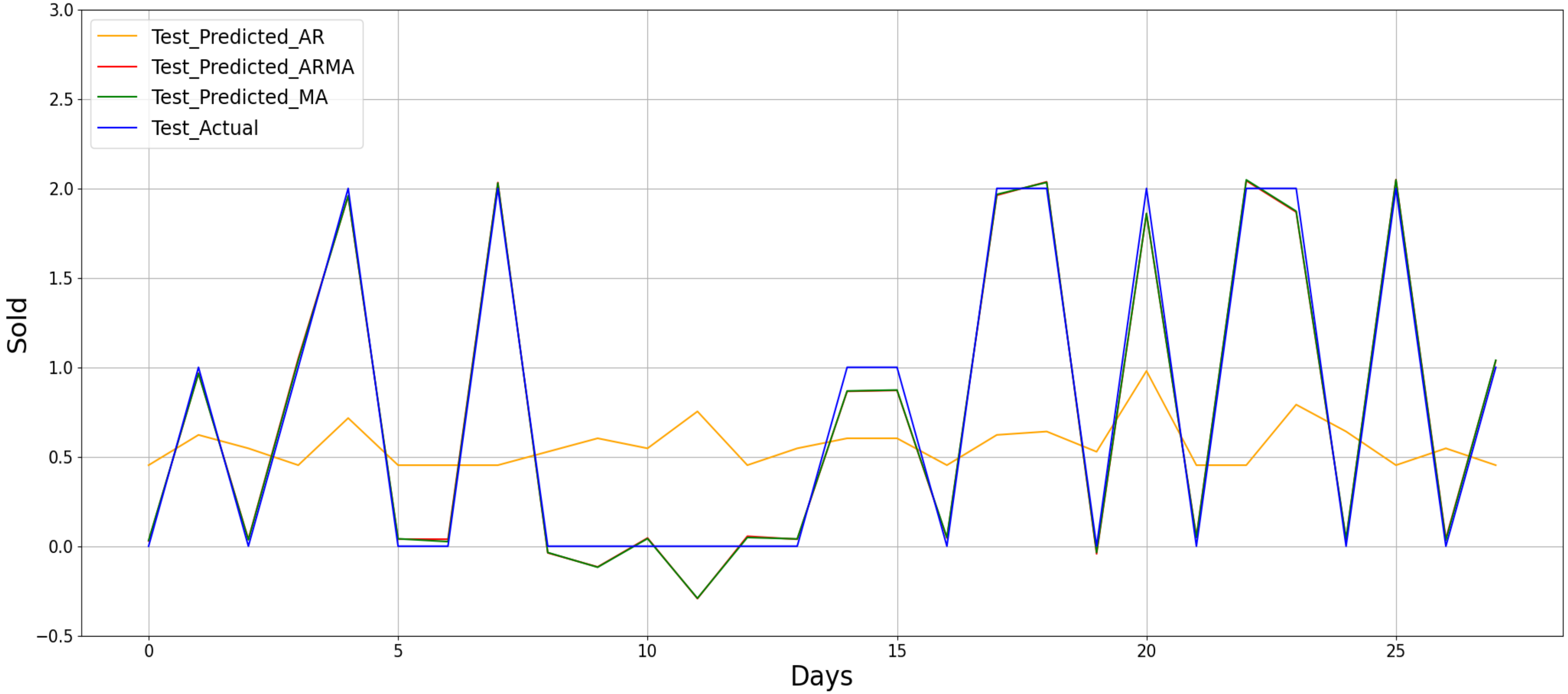
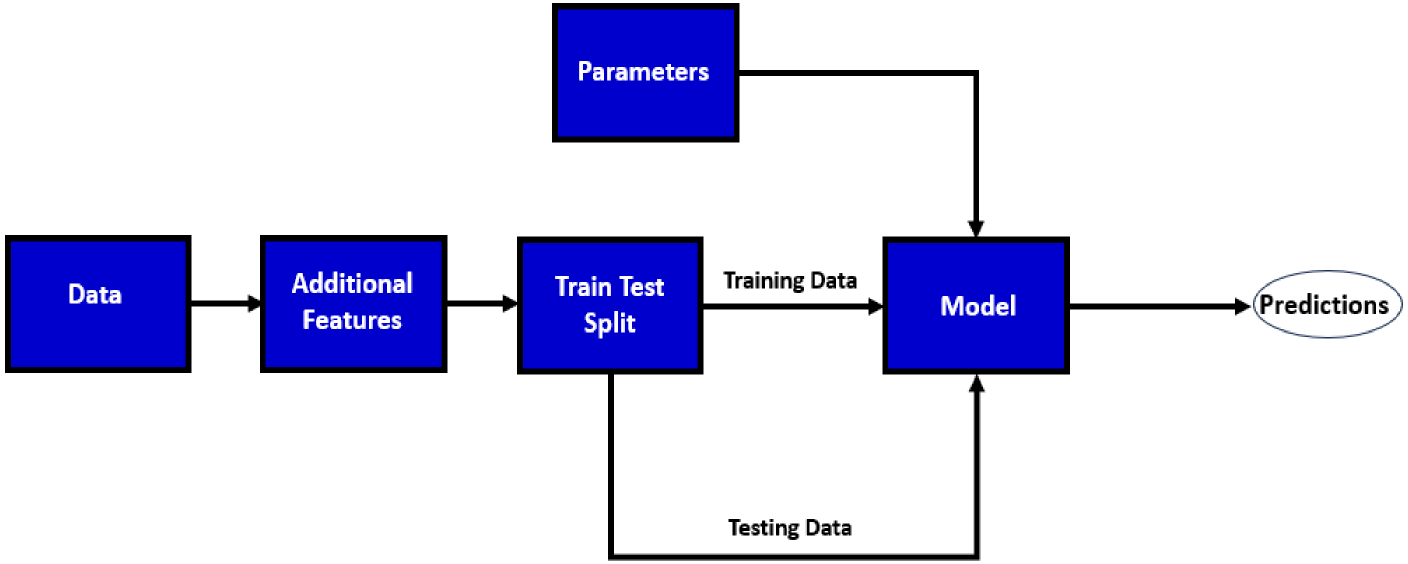


Figure: Actual vs Predicted on testing set.

Methodology for Machine Learning and Deep Learning Models



Additional features include **expanding sold mean, rolling sold mean, lags etc**. Here model can be any Machine Learning or Deep Learning model.

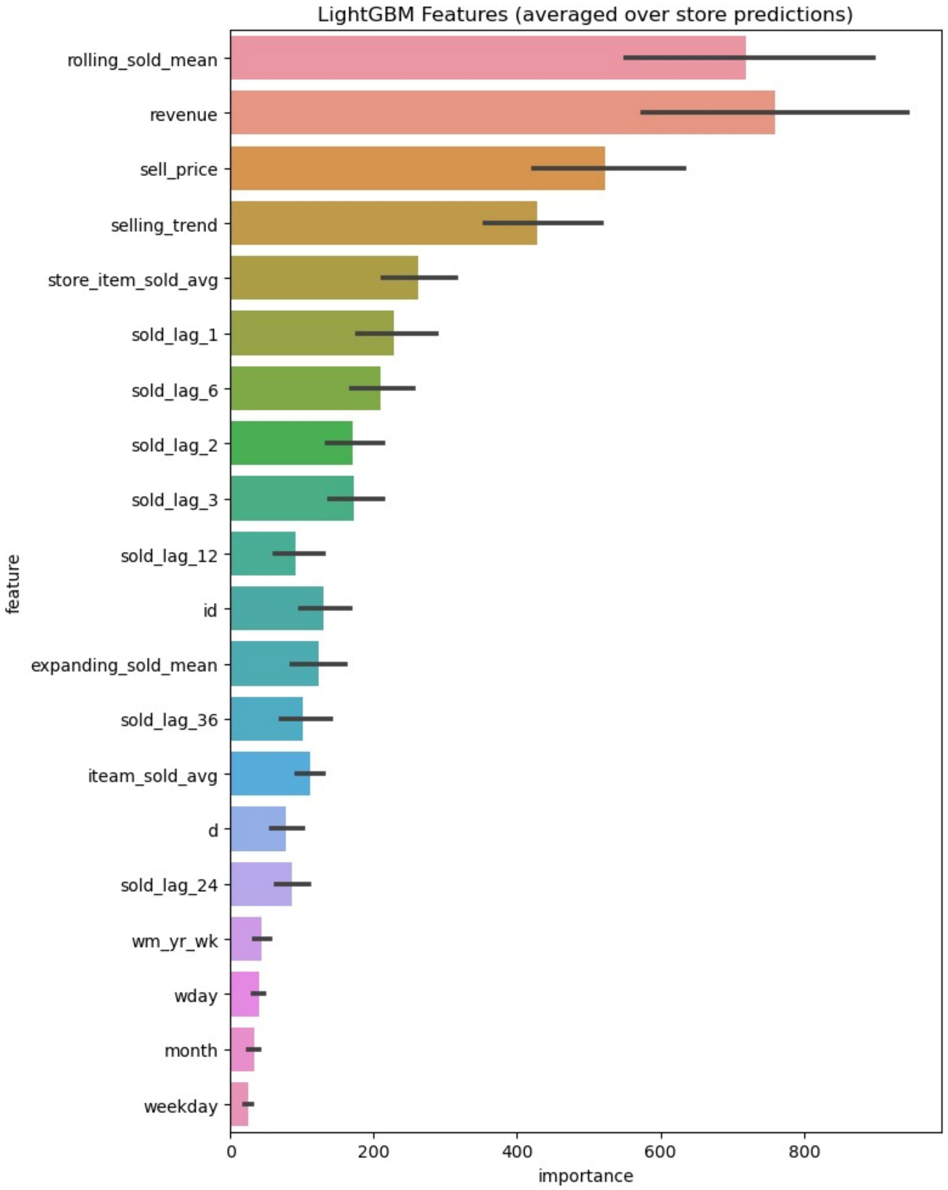
Performance Metrics of LGBM model

Level	MSE	RMSE	MAE
1	9632.951	98.147	76.468
2	1693.9228	40.922	33.0746
3	1079.2604	32.8521	29.887
4	1154.5722	33.979	29.944
5	1049.1	32.328	28.991
6	458.6033	18.151	14.673
7	197.567	10.740	8.340
8	111.929	8.35641	6.7064
9	47.739	4.9100	3.8132
10	72.600	8.521	7.44
11	34.25	5.852	4.61
12	0.105097	0.306477	0.066

Level id	Aggregation Level	Number of series
1	Unit sales of all products, aggregated for all stores/states	1
2	Unit sales of all products, aggregated for each State	3
3	Unit sales of all products, aggregated for each store	10
4	Unit sales of all products, aggregated for each category	3
5	Unit sales of all products, aggregated for each department	7
6	Unit sales of all products, aggregated for each State and category	9
7	Unit sales of all products, aggregated for each State and department	21
8	Unit sales of all products, aggregated for each store and category	30
9	Unit sales of all products, aggregated for each store and department	70
10	Unit sales of product x, aggregated for all stores/states or group by item_id	3049
11	Unit sales of product x, aggregated for each State	9147
12	Unit sales of product x, aggregated for each store	30490
Total		**42840**

Results:- Performance table on the testing set. Average over different time series in that level.

Features Importance Plot for LGBM model

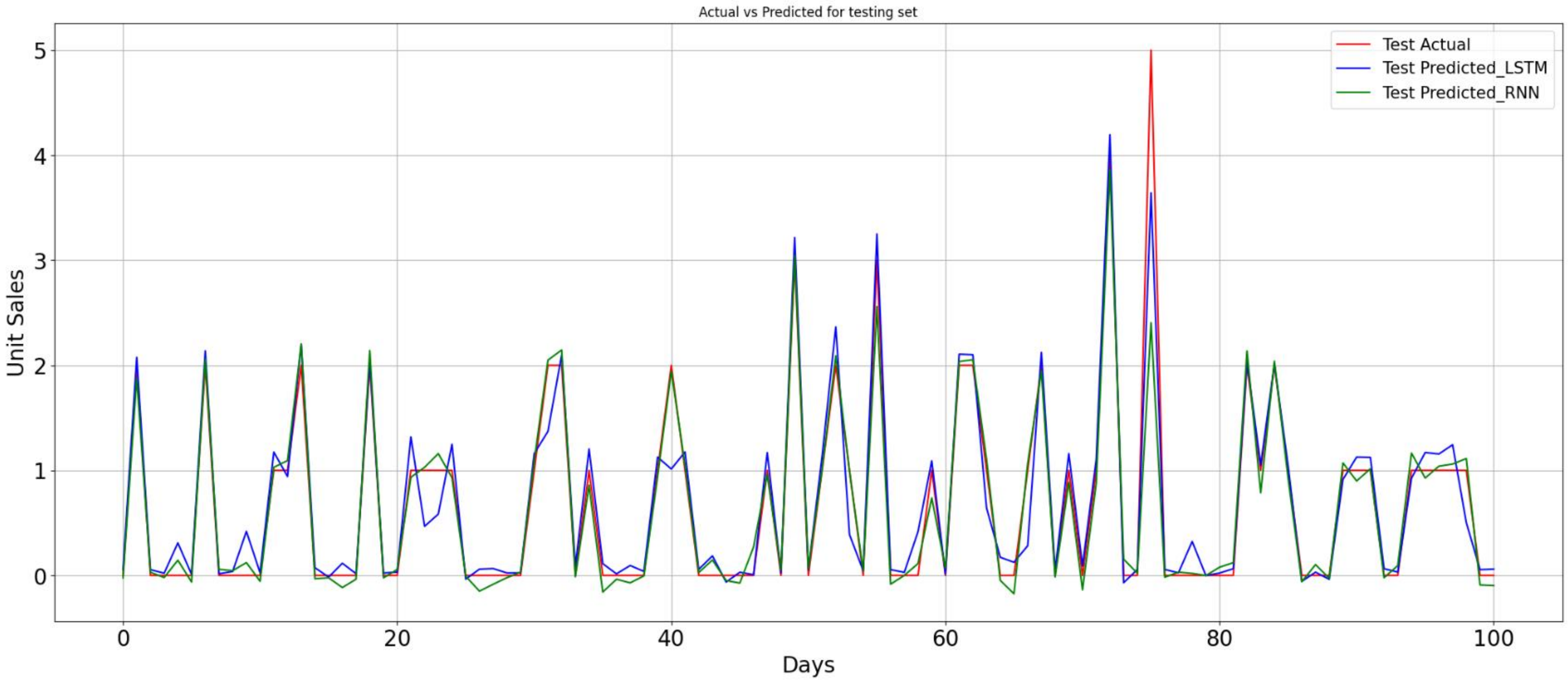


- The plot displays the relative importance of each feature in the trained LightGBM model. Features with higher importance values contribute more to the model's predictions.
- This metric simply counts how many times a feature is used to split the data across all trees.
- In splitting of node in trees, only those features are considered which give maximum difference in variance between child node and parent node.

Results and Performance Metrics of RNN and LSTM Univariate

	RNN			LSTM		
MSE	0.0287	0.0288	0.0673	0.0091	0.0120	0.100
RMSE	0.1695	0.1699	0.2595	0.0954	0.1096	0.316
MAE	0.1324	0.1317	0.1794	0.0529	0.0587	0.150
	Training	**Validation**	**Testing**	**Training**	**Validation**	**Testing**

Table: Performance Metrics for store = ‘CA1’ and item_id = ‘HOBBIES_1_001’.



Case Study and Dataset Description

- **Corporacion Favorita dataset**

  - 17 clusters
  - 15 stores
  - 16 states
  - 44 cities
  - 33 types
  - 1235 days
  - 18154 SKUs

Types of Features		
Date	Str_nbr (Store number)	Cluster (which cluster the particular store lies in with store number.)
Locale (Local holiday, Regional or National holiday)	Item_nbr (item number)	Type (Type of the store)
Locale_name (if National then Ecuador else region/state name for regional or local name	Family (of item number)	City (city of store)
Description of holiday	Class (of item number)	State (state of store)
Transferred(True Or False)	Perishable (True or False for item number)	Type of holiday(Holiday or Additional)
Daily oil price	On promotion (with respect to date, item number, store number)	
Total transactions (on particular store on particular date.)	Unit sales.	

Results (Corporacion Favorita Dataset)

The forecasting is done for the particular item_nbr = 108701 and the store_nbr = 25.

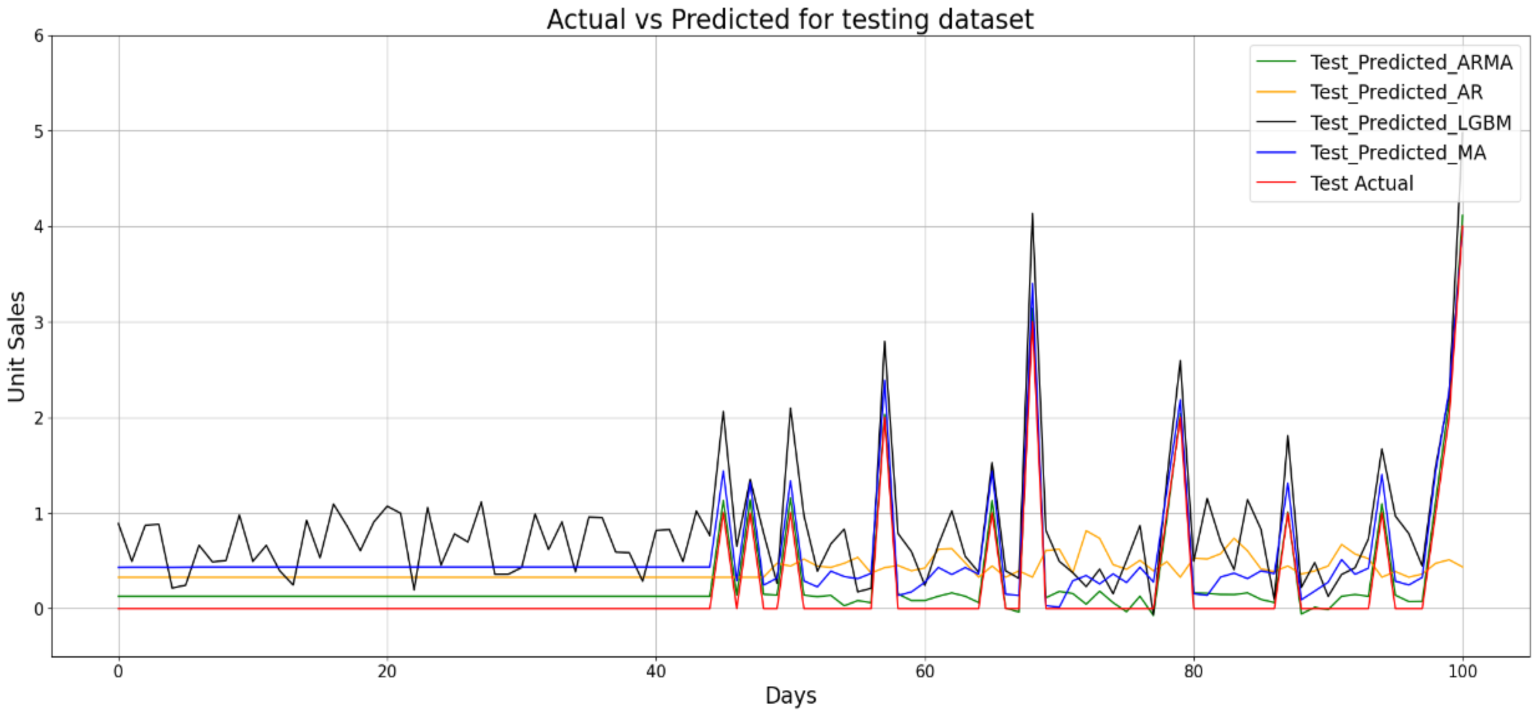


Figure : Actual vs Predicted plot for testing dataset.

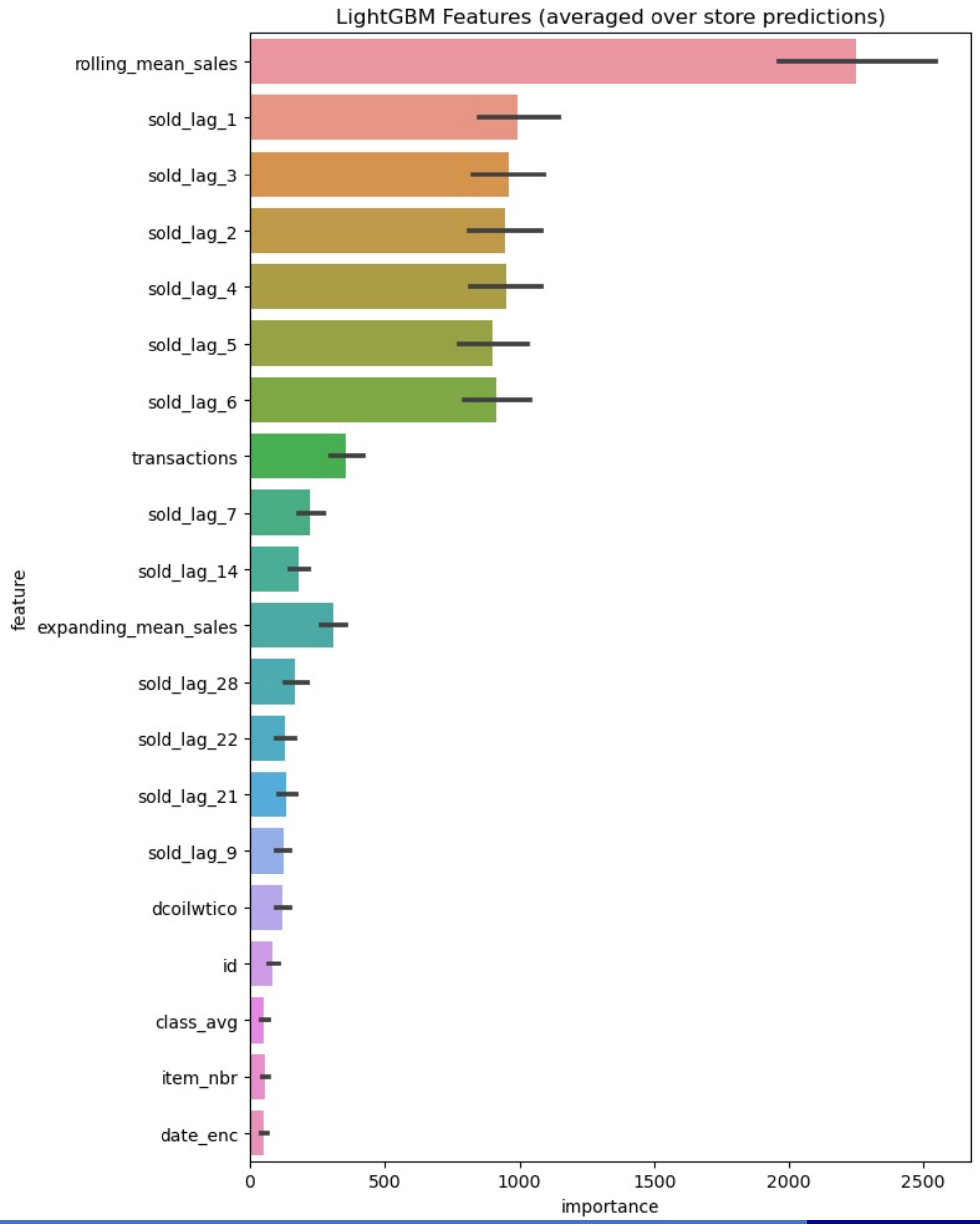
Model Name	RMSE	MSE	MAE
AR	1.077	1.1608	0.787
MA	0.323	0.104	0.249
ARMA	0.1368	0.018	0.1022
LGBM	0.204	0.041	0.154

Table : Performance metrics of testing dataset.

Model Name	RMSE	MSE	MAE
AR	0.8524	0.7266	0.6438
MA	0.308	0.0952	0.2626
ARMA	0.119	0.01436	0.099
LGBM	0.211	0.044	0.1636

Table : Performance metrics of training dataset.

Plots



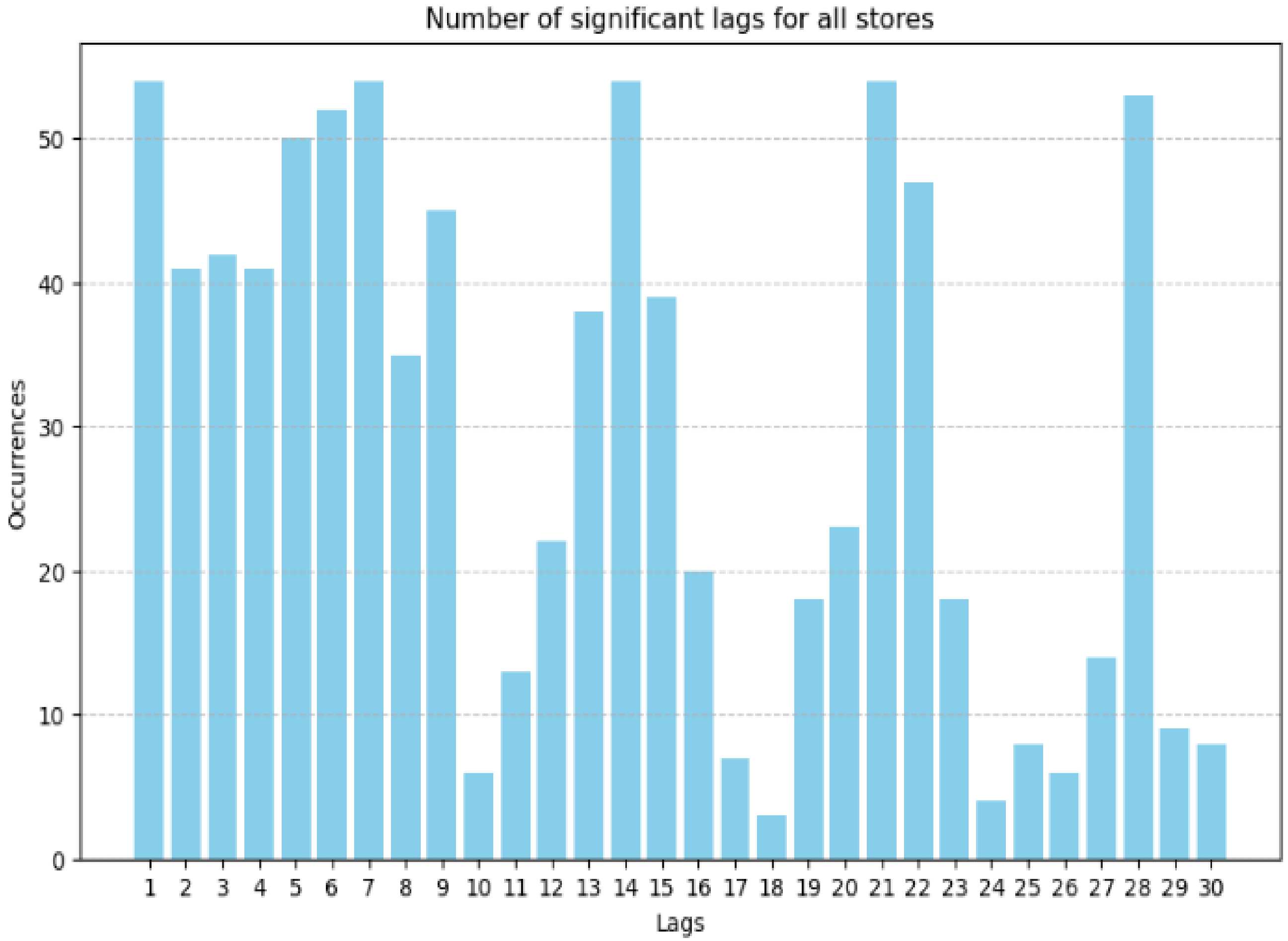


Figure : To find the most common significant lags.

Figure : Features Importance plot.

NARMA (Nonlinear ARMA) Simulated Series

	RNN			LSTM		
MSE	0.004	0.004	0.004	0.003	0.004	0.003
RMSE	0.064	0.063	0.063	0.063	0.063	0.058
MAE	0.052	0.052	0.051	0.050	0.054	0.050
	Training	**Validation**	**Testing**	**Training**	**Validation**	**Testing**

Table: Performance Metrics.

- The input signal u_t is created using the ARMA model of order (2,2).
- $u_t = 0.5u_{t-1} - 0.2u_{t-2} + 0.3\theta_{t-1} - 0.1\theta_{t-2} + \theta_t.$
- Output signal y_t is created using the NARMA model from the paper Ki H. Chon et al. (1997).
- $y_t = 0.8u_t - 0.13u_{t-2} + 0.2y_{t-1} - 0.11y_{t-3} - 0.11u_{t-1}^2 + 0.13y_{t-2}^2 - 0.18u_{t-1}y_{t-1}$

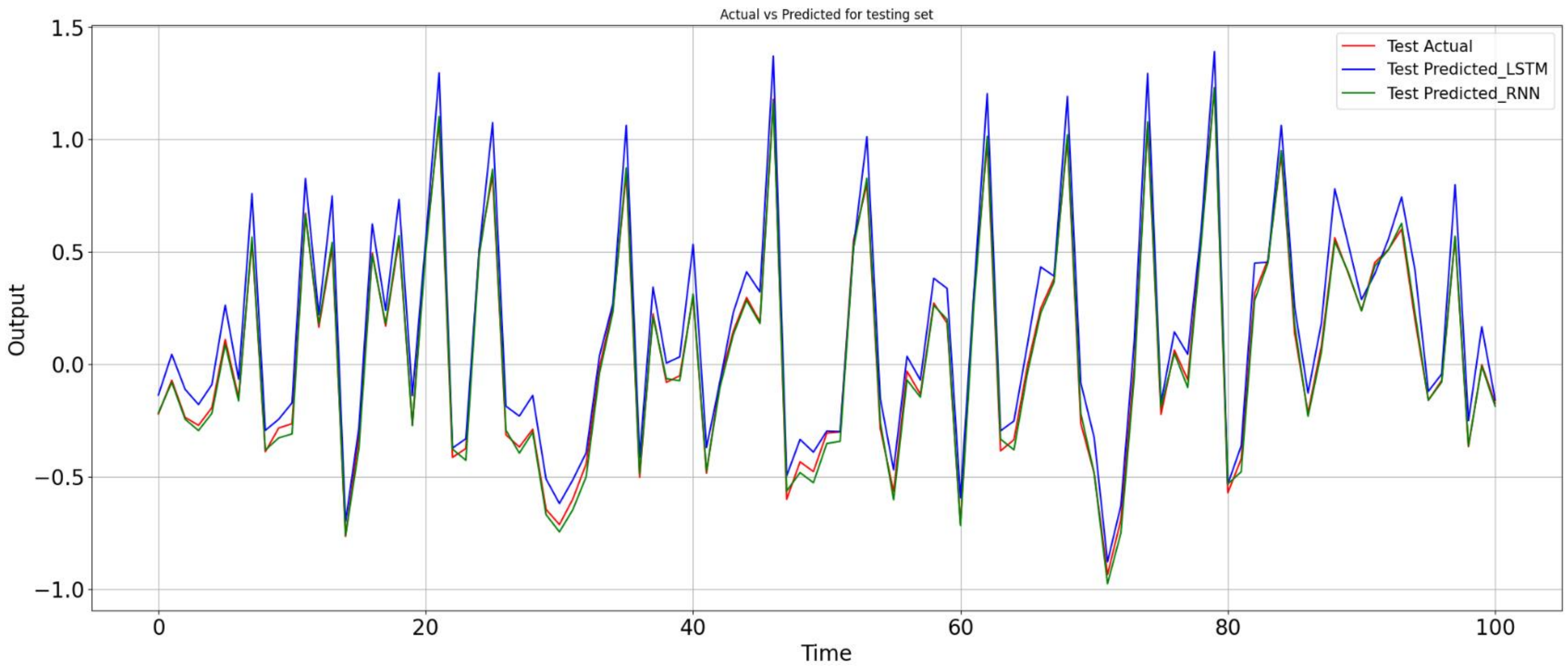
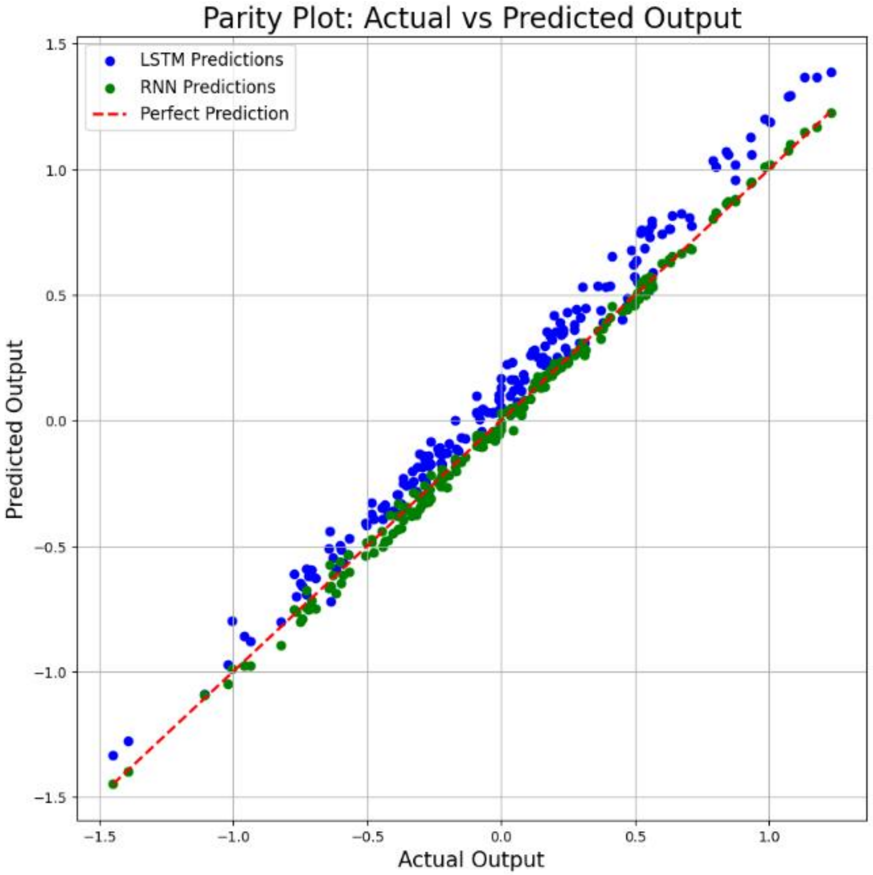


Figure :Actual vs predicted plot for testing set.



SRU(Sulphur Recovery Unit) and Debutanizer Column Dataset

- The process variables for each process are defined below:

Variables	Notation	Description
Inputs	x_1	Top temperature
	x_2	Top pressure
	x_3	Reflux flow
	x_4	Flow to next process
	x_5	6^{th} tray temperature
	x_6	Bottom temperature
	x_7	Bottom temperature
Output	y	Butane (C4) content in the debutanizer column bottom

Table: Debutanizer column variables.

Variables	Notation	Description
Inputs	x_1	MEA gas flow (Monoethanolamine)
	x_2	First air flow
	x_3	Second air flow
	x_4	Gas flow in SWS zone
	x_5	Air flow in SWS zone
Outputs	y_1	Concentration of H_2S.
	y_2	Concentration of SO_2

Table: SRU variables.

Results and Performance Metrics (SRU Dataset H_2S) of RNN and LSTM Univariate

	RNN			LSTM		
MSE	0.000192	0.00018	0.00020	0.0008	0.0008	0.00128
RMSE	0.01386	0.01356	0.0141	0.028	0.029	0.0357
MAE	0.00571	0.0054	0.0067	0.0155	0.0150	0.0182
	Training	**Validation**	**Testing**	**Training**	**Validation**	**Testing**

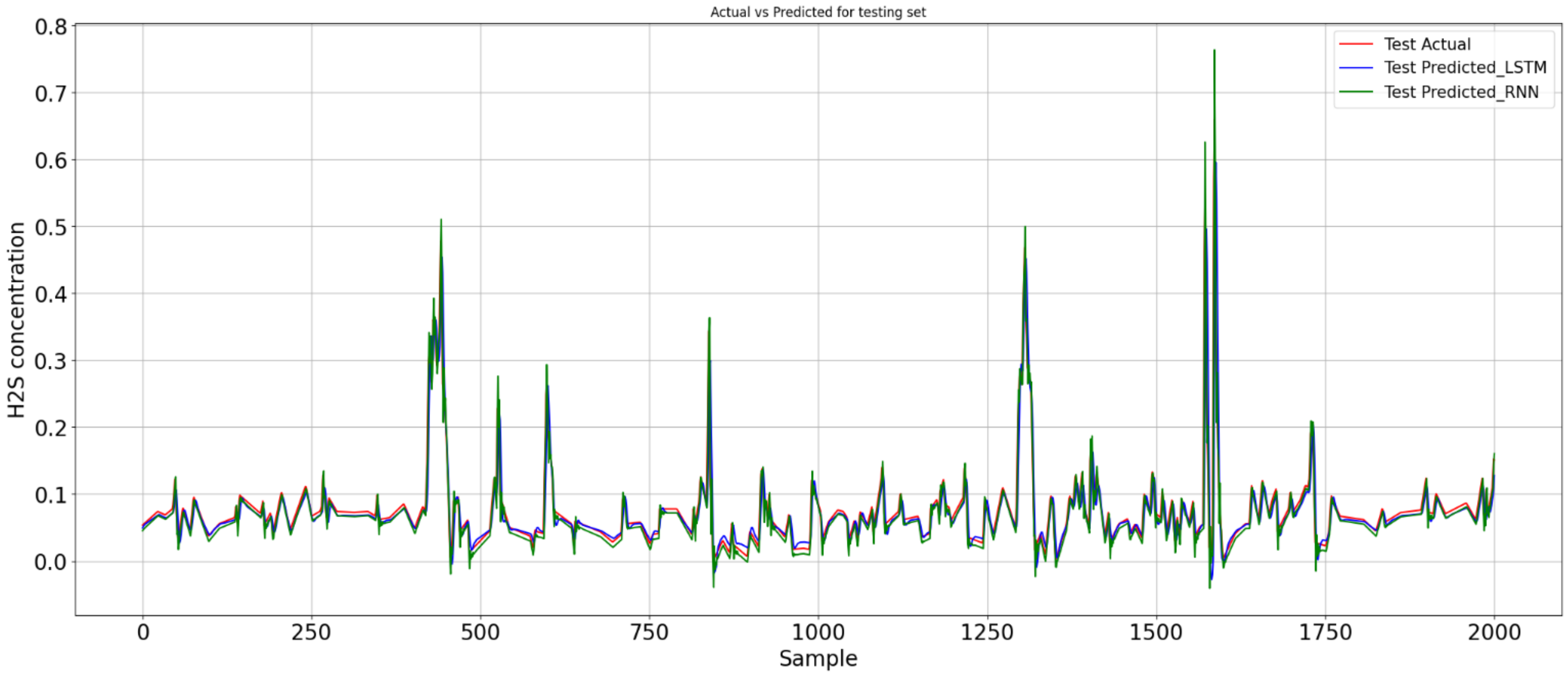
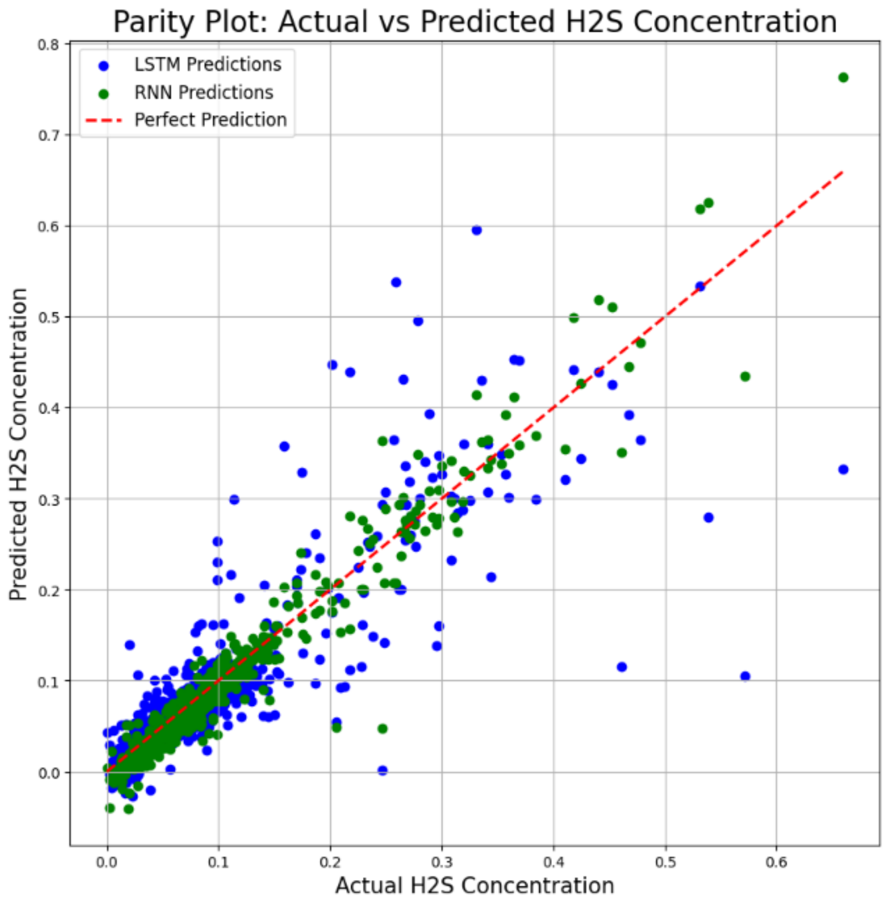


Figure : Actual vs predicted plot on testing set.



Results and Performance Metrics (SRU Dataset SO_2) of RNN and LSTM Univariate

	RNN			LSTM		
MSE	0.00032	0.00027	0.00028	0.0013	0.0008	0.0009
RMSE	0.01799	0.01653	0.01679	0.0365	0.0286	0.0312
MAE	0.0122	0.0119	0.0129	0.0233	0.0210	0.0230
	Training	**Validation**	**Testing**	**Training**	**Validation**	**Testing**

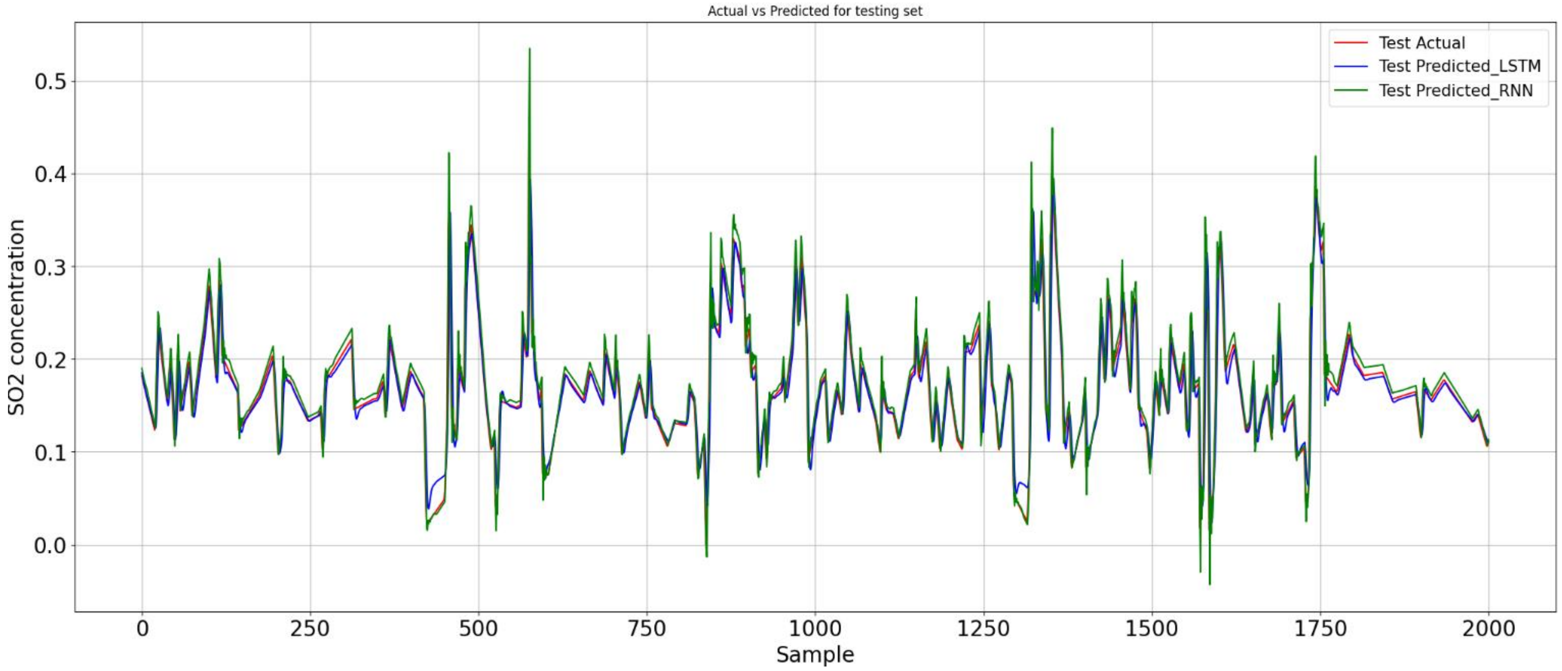
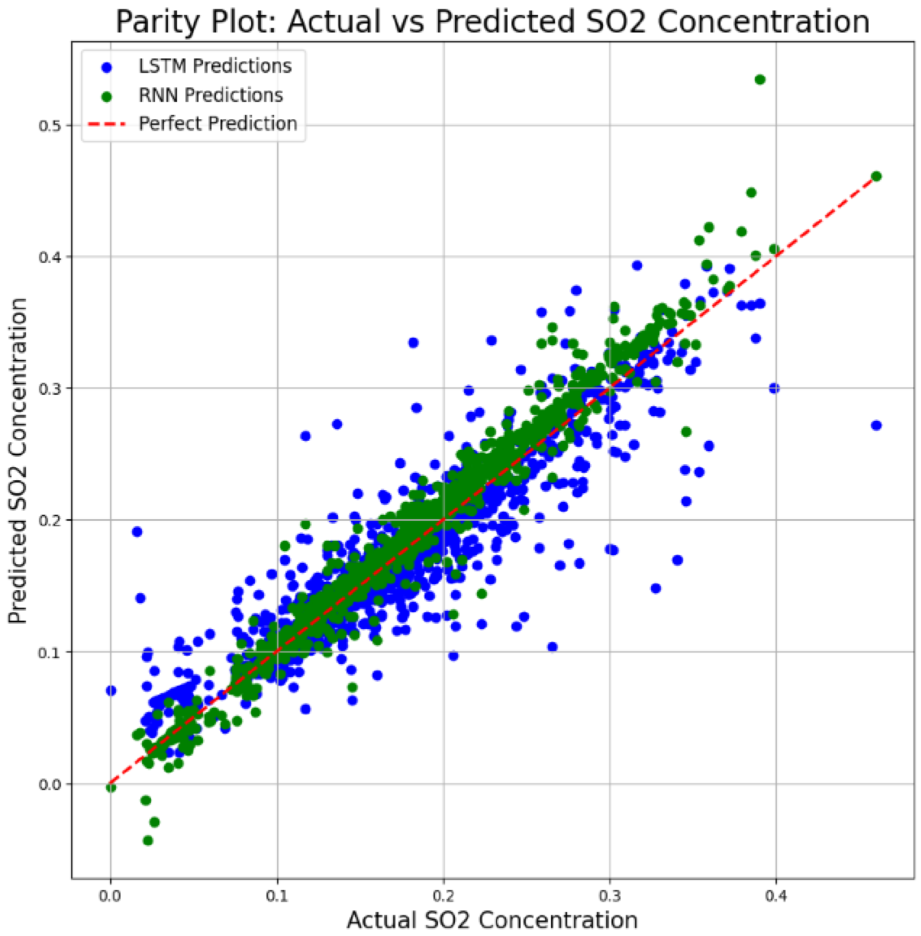


Figure : Actual vs predicted plot on testing set.



Results and Performance Metrics (Debutanizer Dataset) of RNN and LSTM Univariate

	RNN			LSTM		
MSE	0.0018	0.00256	0.0046	0.00038	0.00054	0.0007
RMSE	0.04258	0.0506	0.068	0.0196	0.0233	0.0269
MAE	0.0283	0.034	0.0545	0..0137	0.0143	0.0205
	Training	**Validation**	**Testing**	**Training**	**Validation**	**Testing**

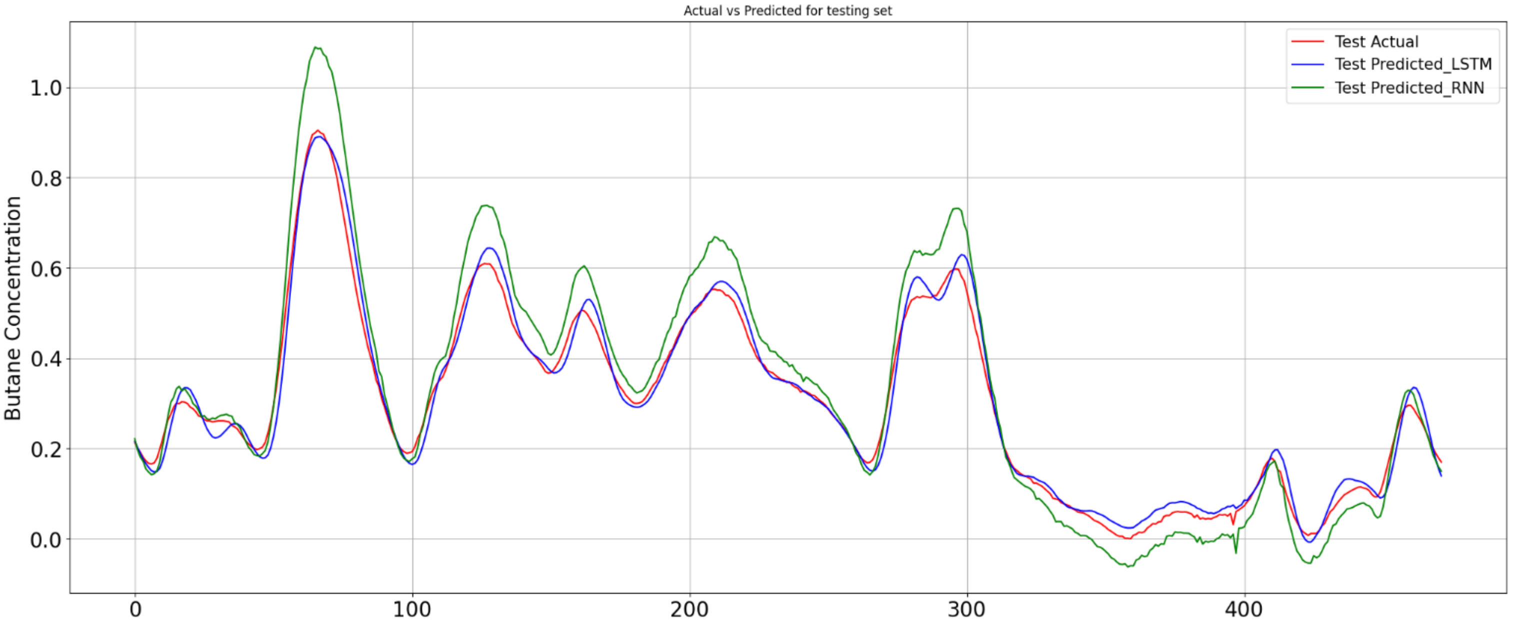
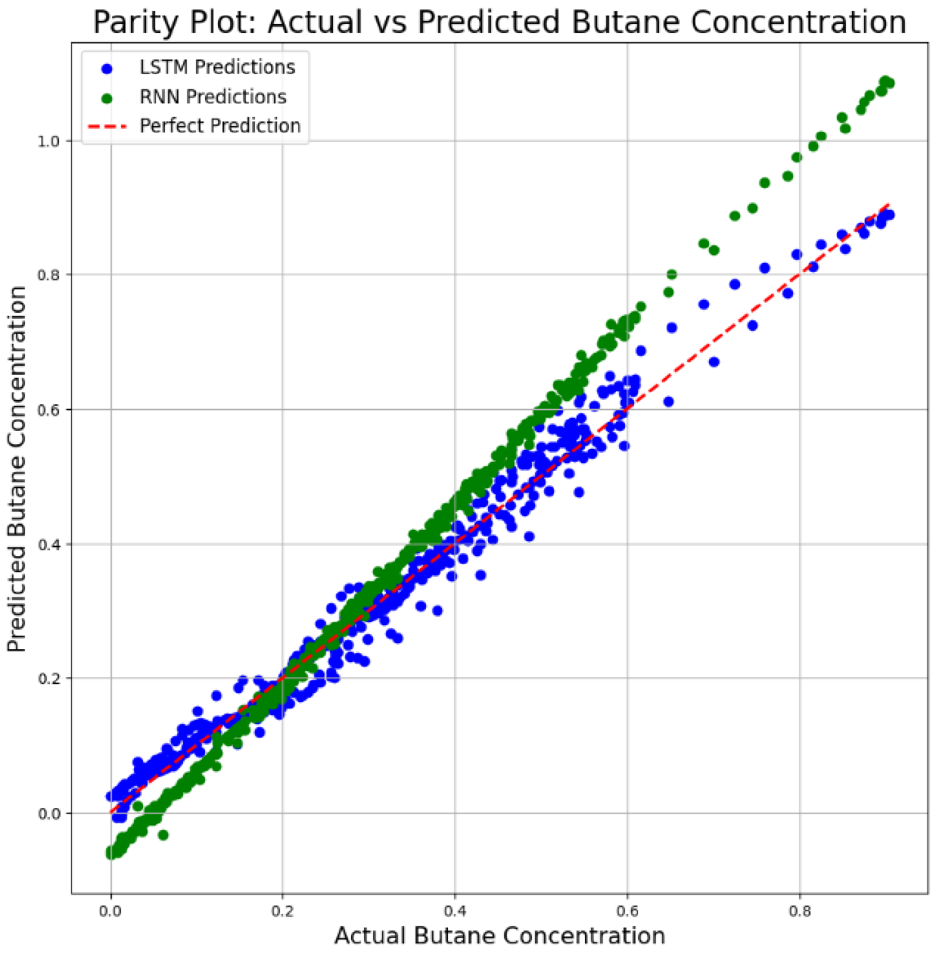


Figure: Actual vs predicted plot on testing set.



Discussions

- The decision of taking the significant lags is totally dependent on us. We always try to reduce the number of the significant lags as to reduce the computational complexity.
- We didn’t implement the ARIMA model as it is mainly used for the non-stationary time series. In all datasets we didn’t find any non-stationary time series.
- Have implemented the RNN and LSTM on Corporacion Favorita dataset but due to some missing days, any interpolation method did not give good results with any of the RNN and LSTM.
- As we created univariate RNN and LSTM models, we have considered the lags of output value as the features.

Conclusions and Potential Extensions

- Conclusions
  - AR model didn’t perform well as MA and ARMA on both Walmart and Corporacion Favorita dataset.
  - PACF and ACF are the most important concepts that will help you to get the significant lags and will help you to improve the performance of any whether it is time series model, machine learning model or deep learning models. You can include the significant lags as your features.
  - Implementing time series and univariate deep learning model is easy to implement as compared to machine learning model but have more computational complexity if want to do forecasting for large number of time series in once.
  - After accounting the effect of significant lags, the deep learning models like RNN and LSTM show very good improvement.
  - LSTM performed well as compared to RNN in soft sensing but for Walmart dataset RNN performed well.
  - Univariate deep learning model can be used to forecast the single time series however the multivariate deep learning model is capable to forecast the different time series.
  - Deep Learning models are capable to handle the non-stationary time series also.
- Potential Extensions
  - Using Transformers and Seq2Seq for SKU forecasting and soft sensing.

References

- Taghiyeh, S., Lengacher, D. C., Sadeghi, A. H., Sahebi-Fakhrabad, A., & Handfield, R. B. (2023, September)
  A novel multi-phase hierarchical forecasting approach with machine learning in supply chain management.
  Supply Chain Analytics, 3, 100032.
- Curreri, F., Patané, L., & Xibilia, M. G. (2021).
  and LSTM-Based soft sensors transferability for an industrial process.
  Sensors, 21(3), 823.
- Li, J., & Qin, S. J. (2023).
  Applying and dissecting LSTM neural networks and regularized learning for dynamic inferential modeling.
  Computers & Chemical Engineering, 175, 108264.
- Vallés-Ṕerez, I., Soria-Olivas, E., Martínez-Sober, M., Serrano-López, A. J., Gómez-Sanchís, J., & Mateo, F. (2022).
  Approaching sales forecasting using recurrent neural networks and transformers
  Expert Systems With Applications, 201, 116993.
- Ki H. Chon , Richard J. Cohen(1997, March).
  Linear and Nonlinear ARMA Model Parameter Estimation Using an Artificial Neural Network.
  IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 44, NO. 3.
- University of Nicosia, Kaggle (2020)
  Kaggle dataset link:- M5 Forecasting - Accuracy | Kaggle
- CORPORACIÓN FAVORITA, Kaggle (2018)
  Kaggle dataset link:- Corporación Favorita Grocery Sales Forecasting | Kaggle