Retail Sales Time Series

Objective:

Develop a predictive model using a deep learning framework (TensorFlow or PyTorch) to forecast future sales from historical time series data. You are to select a deep learning algorithm of your choice to accurately predict future time steps.

Dataset Explanation

Dataset Name: Retail Sales Time Series
Format: CSV

Description:

The dataset consists of monthly sales data from a major retail chain for the years 2015 to 2020, including:

- Month: Month of the sales data record.
- Sales: Total sales value in USD.

Research 1:

Vanilla RNNs (Recurrent Neural Networks) have some limitations that make them less than ideal for sales forecasting tasks:

1. Short-Term Memory: Vanilla RNNs struggle with long-term dependencies in data. This is a problem for sales forecasting because past sales data can influence future sales, but these influences might not be immediately preceding periods. Vanilla RNNs tend to focus heavily on the most recent information, neglecting potentially valuable patterns from further in the past.

2. Vanishing/Exploding Gradients: During training, RNNs propagate gradients back through the network to update weights and biases. In Vanilla RNNs, these gradients can either vanish (become very small) or explode (become very large) as they travel back through the network. This makes it difficult for the network to learn long-term dependencies effectively.

Here's a breakdown of these limitations:

- **Vanishing Gradients:** If gradients become very small as they travel back through the network, earlier layers barely get updated during training. This hinders the network's ability to learn patterns from sequences with long-term dependencies.

- **Exploding Gradients:** Conversely, if gradients explode, they can overwhelm the network, making it unstable and leading to nonsensical predictions.

Research 2:

LSTMs (Long Short-Term Memory) are a type of Recurrent Neural Network (RNN) that excel in tasks like sales forecasting due to their ability to address key shortcomings of vanilla RNNs:

LSTMs specifically address these limitations, making them well-suited for sales forecasting:

- **Internal Memory Cell:** LSTMs have a special internal cell that controls the flow of information. It can remember values for extended periods, allowing the network to capture long-term dependencies in sales data.
- **Gating Mechanisms:** LSTMs use gates (forget gate, input gate, output gate) to regulate information flow within the cell. These gates determine what information is remembered, forgotten, and used by the network. This helps prevent vanishing/exploding gradients, enabling the network to learn from both recent and past sales trends.

Benefits of LSTMs in Sales Forecasting:

- **Improved Accuracy:** By capturing long-term dependencies, LSTMs can model complex relationships between past sales data and future sales, leading to more accurate forecasts.
- **Seasonality and Trends:** LSTMs can effectively capture seasonal patterns and identify long-term trends in sales data, which is crucial for accurate forecasting.
- **External Factors:** LSTMs can be incorporated with additional features like marketing campaigns, holidays, or economic indicators. The model can then learn how these factors influence historical sales data and use that knowledge for future predictions.

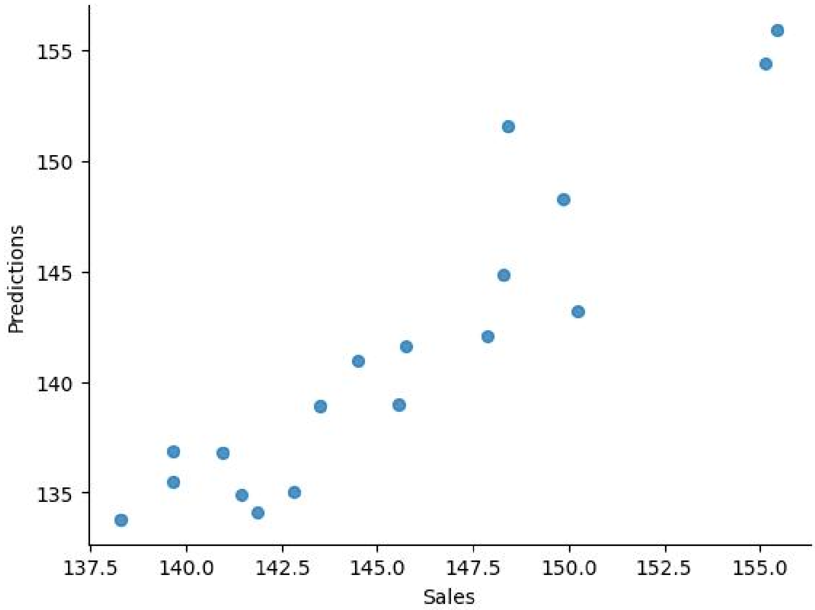
Model: "sequential_1"

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 100)	40800
dense_1 (Dense)	(None, 1)	101

Total params: 40901 (159.77 KB) Trainable params: 40901 (159.77 KB) Non-trainable params: 0 (0.00 Byte)

Results:

Date	Sales	Predictions
2020-12-30 07:00:00	155.4529014	155.89523995319615
2020-12-30 08:00:00	155.1607117	154.39717185404723
2020-12-30 09:00:00	148.4241211	151.53158438999628
2020-12-30 10:00:00	149.8670627	148.2541465603917
2020-12-30 11:00:00	148.2827763	144.84135293824218
2020-12-30 12:00:00	145.7584813	141.59229772298423
2020-12-30 13:00:00	143.4909799	138.9401000934799
2020-12-30 14:00:00	139.6391891	136.9035688751502
2020-12-30 15:00:00	139.6659947	135.50797338619572
2020-12-30 16:00:00	141.4313437	134.93233622054612
2020-12-30 17:00:00	138.2937305	133.78773208591207
2020-12-30 18:00:00	141.8590945	134.13167065470563
2020-12-30 19:00:00	142.8270109	135.06038463923537
2020-12-30 20:00:00	140.9467795	136.81923565104998
2020-12-30 21:00:00	145.5613237	138.97032971512562
2020-12-30 22:00:00	144.4840706	140.96004050825434
2020-12-30 23:00:00	147.8902907	142.06430890706923
2020-12-31 00:00:00	150.2234874	143.1963872673127



Root Mean Square Error (RMSE): 4.867772251833289
Mean Absolute Error (MAE): 4.3651717757494595
Coefficient of Determination (R2): 0.028258787136851482

Observation:

The model can able achieve prediction but not as expected.
The accuracy is so poor.

Research 3:

In this research I found out a new model prediction sales for both long term and short term.

The model is introduced in April 2024 reference: "https://arxiv.org/abs/2404.05192"

Model Name : ATFNet (Adaptive Time-Frequency Ensembled Network for Long-term Time Series Forecasting)

Key Advantage:

- It uses Ensemble learning method, which mean it uses another model to obtain more feature.

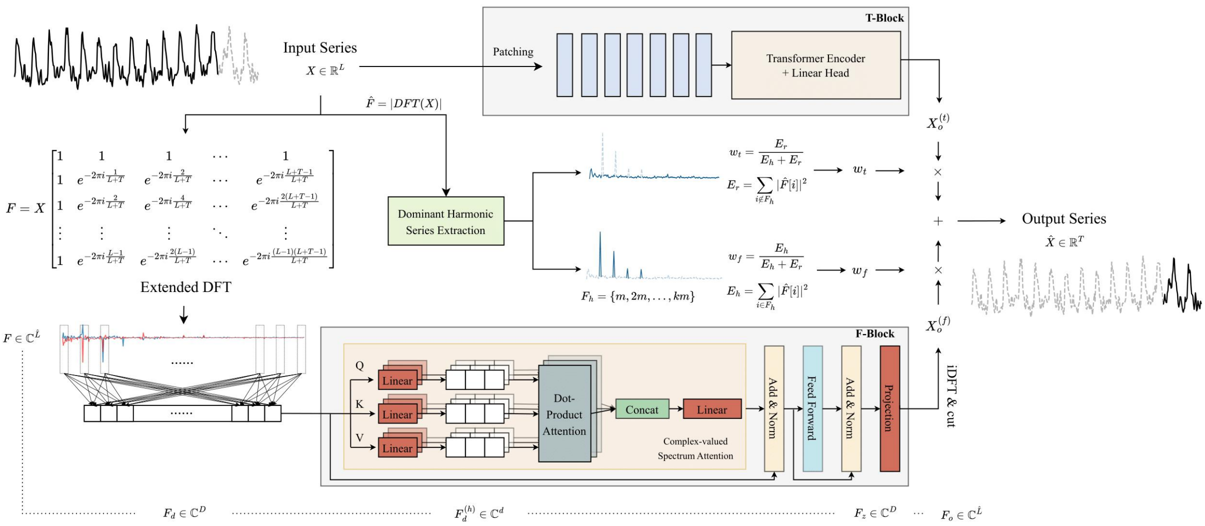
- Two difference blocks (T-Block and F-Block) are used one is for Long Term and another one for Short term

- **T_Block (Time-domain processing)**: This block processes the raw time-series data directly. It is generally better suited for capturing **short-term dependencies and local patterns** in the data, such as recent trends and Immediate fluctuations.

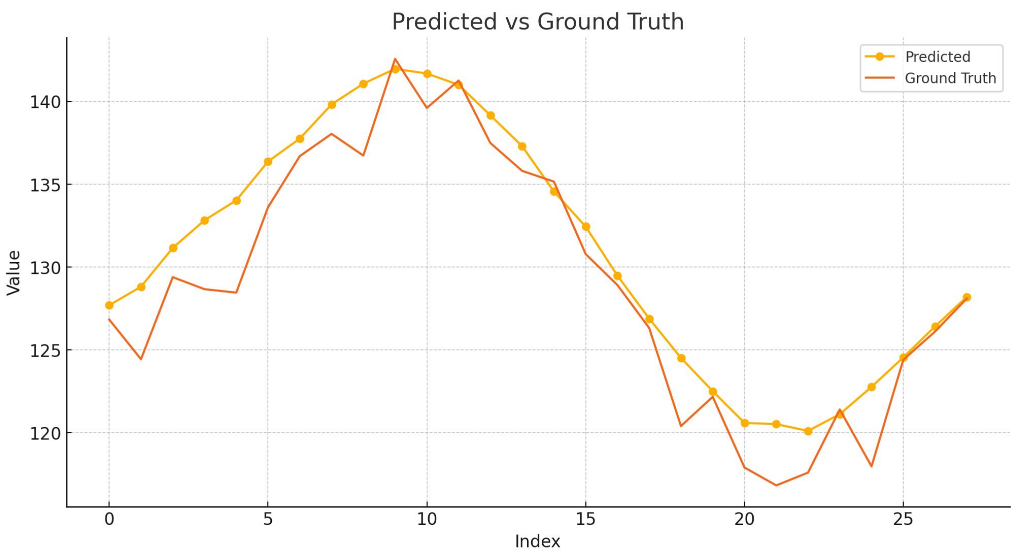
- **F_Block (Frequency-domain processing)**: This block transforms the time-series data into the frequency domain (e.g., using Fourier Transform). It is typically used to capture **long-term dependencies and periodic patterns** that may not be easily discernible in the time domain.

Training parameter:

Sequencial length	:	96
Prediction length	:	96
Batch size	:	256
Frequency	:	h (hours) for T-Block
Lerning Rate	:	variable
Epoch	:	10 (no further improvement after this)
Early Stop	:	3



Results:



Predicted	Ground Truth
127.696754	126.8330968
128.8166	124.4380766
131.15747	129.393595
132.82845	128.6661867
134.03091	128.4597117
136.37825	133.6155216
137.76137	136.702109
139.82475	138.046735
141.07838	136.7356531
141.97379	142.570293
141.68964	139.6106288
141.0087	141.2663003
139.16296	137.4916819
137.30432	135.8114636
134.56682	135.1615538
132.44444	130.7874734
129.48404	128.9172434
126.88926	126.3224552
124.517876	120.4010079
122.50751	122.1670162

mae: 0.087463624547 48068
mse: 0.011959103036423507
rmse: 0.10935768393864012

This model chosen by comparing with other latest high accuracy model for other datasets.

For our dataset, directly chosen this best model because ATFNet

Models		ATFNet		iTransformer		RLinear		PatchTST		Crossformer		TiDE		TimesNet		DLinear		SCINet		FEDformer		Stationary		Autoformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.385	0.403	0.386	0.405	0.386	0.395	0.414	0.419	0.423	0.448	0.479	0.464	0.384	0.402	0.386	0.400	0.654	0.599	0.376	0.419	0.513	0.491	0.449	0.459
	192	0.423	0.431	0.441	0.436	0.437	0.424	0.460	0.445	0.471	0.474	0.525	0.492	0.436	0.429	0.437	0.432	0.719	0.631	0.420	0.448	0.534	0.504	0.500	0.482
	336	0.436	0.452	0.487	0.458	0.479	0.446	0.501	0.466	0.570	0.546	0.565	0.515	0.491	0.469	0.481	0.459	0.778	0.659	0.459	0.465	0.588	0.535	0.521	0.496
	720	0.530	0.520	0.503	0.491	0.481	0.470	0.500	0.488	0.653	0.621	0.594	0.558	0.521	0.500	0.519	0.516	0.836	0.699	0.506	0.507	0.643	0.616	0.514	0.512
	Avg	0.443	0.452	0.454	0.448	0.446	0.434	0.469	0.455	0.529	0.522	0.541	0.507	0.458	0.450	0.456	0.452	0.747	0.647	0.440	0.460	0.570	0.536	0.496	0.487
ETTh2	96	0.276	0.332	0.297	0.349	0.288	0.338	0.302	0.348	0.745	0.584	0.400	0.440	0.340	0.374	0.333	0.387	0.707	0.621	0.358	0.397	0.476	0.458	0.346	0.388
	192	0.326	0.368	0.380	0.400	0.374	0.390	0.388	0.400	0.877	0.656	0.528	0.509	0.402	0.414	0.477	0.476	0.860	0.689	0.429	0.439	0.512	0.493	0.456	0.452
	336	0.316	0.373	0.428	0.432	0.415	0.426	0.426	0.433	1.043	0.731	0.643	0.571	0.452	0.452	0.594	0.541	1.000	0.744	0.496	0.487	0.552	0.551	0.482	0.486
	720	0.399	0.438	0.427	0.445	0.420	0.440	0.431	0.446	1.104	0.763	0.874	0.679	0.462	0.468	0.831	0.657	1.249	0.838	0.463	0.474	0.562	0.560	0.515	0.511
	Avg	0.329	0.378	0.383	0.407	0.374	0.398	0.387	0.407	0.942	0.683	0.611	0.550	0.414	0.427	0.559	0.515	0.954	0.723	0.436	0.449	0.526	0.516	0.450	0.459
ETTm1	96	0.306	0.354	0.334	0.368	0.355	0.376	0.329	0.367	0.404	0.426	0.364	0.387	0.338	0.375	0.345	0.372	0.418	0.438	0.379	0.419	0.386	0.398	0.505	0.475
	192	0.340	0.380	0.377	0.391	0.391	0.392	0.367	0.385	0.450	0.451	0.398	0.404	0.374	0.387	0.380	0.389	0.439	0.450	0.426	0.441	0.459	0.444	0.553	0.496
	336	0.379	0.402	0.426	0.420	0.424	0.415	0.399	0.410	0.532	0.515	0.428	0.425	0.410	0.411	0.413	0.413	0.490	0.485	0.445	0.459	0.495	0.464	0.621	0.537
	720	0.426	0.434	0.491	0.459	0.487	0.450	0.454	0.439	0.666	0.589	0.487	0.461	0.478	0.450	0.474	0.453	0.595	0.550	0.543	0.490	0.585	0.516	0.671	0.561
	Avg	0.363	0.392	0.407	0.410	0.414	0.408	0.387	0.400	0.513	0.495	0.419	0.419	0.400	0.406	0.403	0.407	0.485	0.481	0.448	0.452	0.481	0.456	0.588	0.517
ETTm2	96	0.174	0.262	0.180	0.264	0.182	0.265	0.175	0.259	0.287	0.366	0.207	0.305	0.187	0.267	0.193	0.292	0.286	0.377	0.203	0.287	0.192	0.274	0.255	0.339
	192	0.231	0.301	0.250	0.309	0.246	0.304	0.241	0.302	0.414	0.492	0.290	0.364	0.249	0.309	0.284	0.362	0.399	0.445	0.269	0.328	0.280	0.339	0.281	0.340
	336	0.294	0.345	0.311	0.348	0.307	0.342	0.305	0.343	0.597	0.542	0.377	0.422	0.321	0.351	0.369	0.427	0.637	0.591	0.325	0.366	0.334	0.361	0.339	0.372
	720	0.386	0.402	0.412	0.407	0.407	0.398	0.402	0.400	1.730	1.042	0.558	0.524	0.408	0.403	0.554	0.522	0.960	0.735	0.421	0.415	0.417	0.413	0.433	0.432
	Avg	0.271	0.328	0.288	0.332	0.285	0.327	0.281	0.326	0.757	0.611	0.358	0.404	0.291	0.333	0.350	0.401	0.571	0.537	0.304	0.349	0.306	0.347	0.327	0.371
traffic	96	0.375	0.262	0.395	0.268	0.649	0.389	0.462	0.295	0.522	0.290	0.805	0.493	0.593	0.321	0.650	0.396	0.788	0.499	0.587	0.366	0.612	0.338	0.613	0.388
	192	0.393	0.271	0.417	0.276	0.601	0.366	0.466	0.296	0.530	0.293	0.756	0.474	0.617	0.336	0.598	0.370	0.789	0.505	0.604	0.373	0.613	0.340	0.616	0.382
	336	0.407	0.278	0.433	0.283	0.609	0.369	0.482	0.304	0.558	0.305	0.762	0.477	0.629	0.336	0.605	0.373	0.797	0.508	0.621	0.383	0.618	0.328	0.622	0.337
	720	0.437	0.298	0.467	0.302	0.647	0.387	0.514	0.322	0.589	0.328	0.719	0.449	0.640	0.350	0.645	0.394	0.841	0.523	0.626	0.382	0.653	0.355	0.660	0.408
	Avg	0.403	0.277	0.428	0.282	0.627	0.378	0.481	0.304	0.550	0.304	0.760	0.473	0.620	0.336	0.625	0.383	0.804	0.509	0.609	0.376	0.624	0.340	0.628	0.379
electricity	96	0.130	0.227	0.148	0.240	0.201	0.281	0.181	0.270	0.219	0.314	0.237	0.329	0.168	0.272	0.197	0.282	0.247	0.345	0.193	0.308	0.169	0.273	0.201	0.317
	192	0.149	0.245	0.162	0.253	0.201	0.283	0.188	0.274	0.231	0.322	0.236	0.330	0.184	0.289	0.196	0.285	0.257	0.355	0.201	0.315	0.182	0.286	0.222	0.334
	336	0.165	0.264	0.178	0.269	0.215	0.298	0.204	0.293	0.246	0.337	0.249	0.344	0.198	0.300	0.209	0.301	0.269	0.369	0.214	0.329	0.200	0.304	0.231	0.338
	720	0.200	0.295	0.225	0.317	0.257	0.331	0.246	0.324	0.280	0.363	0.284	0.373	0.220	0.320	0.245	0.333	0.299	0.390	0.246	0.355	0.222	0.321	0.254	0.361
	Avg	0.161	0.258	0.178	0.270	0.218	0.298	0.205	0.290	0.244	0.334	0.252	0.344	0.193	0.295	0.212	0.300	0.268	0.365	0.213	0.327	0.193	0.296	0.227	0.338
weather	96	0.162	0.216	0.174	0.214	0.192	0.232	0.177	0.218	0.158	0.230	0.202	0.261	0.172	0.220	0.196	0.255	0.221	0.306	0.217	0.296	0.173	0.223	0.266	0.336
	192	0.199	0.251	0.221	0.254	0.240	0.271	0.225	0.259	0.206	0.277	0.242	0.298	0.219	0.261	0.237	0.296	0.261	0.340	0.276	0.336	0.245	0.285	0.307	0.367
	336	0.247	0.288	0.278	0.296	0.292	0.307	0.278	0.297	0.272	0.335	0.287	0.335	0.280	0.306	0.283	0.335	0.309	0.378	0.339	0.380	0.321	0.338	0.359	0.395
	720	0.313	0.337	0.358	0.347	0.364	0.353	0.354	0.348	0.398	0.418	0.351	0.386	0.365	0.359	0.345	0.381	0.377	0.427	0.403	0.428	0.414	0.410	0.419	0.428
	Avg	0.230	0.273	0.258	0.278	0.272	0.291	0.259	0.280	0.259	0.315	0.270	0.320	0.259	0.286	0.265	0.317	0.292	0.363	0.309	0.360	0.288	0.314	0.338	0.382
solar	96	0.182	0.239	0.203	0.237	0.322	0.339	0.234	0.286	0.310	0.331	0.312	0.399	0.250	0.292	0.290	0.378	0.237	0.344	0.242	0.342	0.215	0.249	0.884	0.711
	192	0.201	0.263	0.233	0.261	0.359	0.356	0.267	0.310	0.734	0.725	0.339	0.416	0.296	0.318	0.320	0.398	0.280	0.380	0.285	0.380	0.254	0.272	0.834	0.692
	336	0.206	0.268	0.248	0.273	0.397	0.369	0.290	0.315	0.750	0.735	0.368	0.430	0.319	0.330	0.353	0.415	0.304	0.389	0.282	0.376	0.290	0.296	0.941	0.723
	720	0.209	0.274	0.249	0.275	0.397	0.356	0.289	0.317	0.769	0.765	0.370	0.425	0.338	0.337	0.356	0.413	0.308	0.388	0.357	0.427	0.285	0.295	0.882	0.717
	Avg	0.199	0.261	0.233	0.262	0.369	0.355	0.270	0.307	0.641	0.639	0.347	0.417	0.301	0.319	0.330	0.401	0.282	0.375	0.291	0.381	0.261	0.278	0.885	0.711

Conclusion:

- The model performs very well and the accuracy level also high.
- The use of attention mechanism and combination of time domain and frequency domain to obtain both long and short term features also the model uses ensemble learning method improved the model performance in prediction