37. Саветовање CIGRE Србија 2025
СИГУРНОСТ, СТАБИЛНОСТ, ПОУЗДАНОСТ И RESILIENCE ЕЛЕКТРОЕНЕРГЕТСКОГ СИСТЕМА
МУЛТИСЕКТОРСКО ПОВЕЗИВАЊЕ У ЕНЕРГЕТИЦИ И ПРИВРЕДИ
26 - 30. мај 2025, Копаоник

DOI: 10.46793/CIGRE37.D2.04

D2.04

PREDIKCIJA PROIZVODNJE ELEKTRIČNE ENERGIJE U ELEKTROENERGETSKOM SISTEMU

FORECASTING ELECTRICITY PRODUCTION IN THE POWER SYSTEM

Matija Rogić, Mileta Žarković[*]

Kratak sadržaj: Potrošnja električne energije zavisi prevashodno od potrošača a prozivodnja se vrši u elektranama koje ne mogu u realnom vremenu da isprate promenu potrošnje. Kako se električna energija ne može efikasno skladištiti, neophodno je unapred poznavati potrebe i navike potrošača. Obnovljivi izvori moraju maksimalno figurisati u prozivodnji energije a vremenske prilike od kojih zavise su slabo predvidive i sklone naglim promenama. Budući da i proizvodnja i potrošnja zavise od nebrojeno uticaja, neophodno je koristiti neuralnu mrežu za predikciju kako bi se uvrstili doprinosi svih faktora. Različite arhitekture su se pokazale uspešno u predikciji, naglašavajući rezličite kvalitete. MLP je dobra osnova u svojoj jednostavnosti. Rekurentne mreže poput LSTM-a su prirodan alat za predikciju vremenski poređanih podataka. Konvolucione mreže i Autoenkoderi briljiraju zbog svoje efikasnosti. Transformeri zbog mehanizma usmerene pažnje (attention) i velikog kapaciteta, mogu da obrade i izvuku zaključke iz ogromne količine podataka. Moguće su i hibridne arhitekture sačinjene od više različitih gradivnih blokova. U ovom radu je predstavljena primena neuralne mreže TiDE (Time-series Dense Encoder) za predikciju proizvodnje i potrošnje električne energije u Srbiji na realnim podacima iz prve polovine 2019. godine. Uz to su korišćeni i meteorološki podaci RHMZ-a za isti period. Predikcija je vršena sedam dana unapred sa učestanošću od jedan sat.

Ključne reči: *elektroenergetski sistem, predikcija, prozivodnja, neuralne mreže, Autoenkoder*

Abstract: Variable ambient conditions have a dominant influence on electricity consumption. Renewable energy sources are increasingly present in electricity generation, and the weather conditions on which they depend are prone to sudden changes and can be predicted with a certain degree of accuracy.
Since both generation and consumption depend on a large number of factors, it is desirable to use a neural network from the domain of artificial intelligence for forecasting so that the contributions of all these factors are included. Different neural network architectures have proven successful in forecasting, each emphasizing different qualities. The application of multilayer perceptron (MLP) neural networks is a good option due to its simplicity. Recurrent neural networks such as Long Short-Term Memory (LSTM) are a natural tool for predicting time-series data. Convolutional networks and autoencoders have an advantage over other

[*] Matija Rogić, wingman.ai, matija.roga@gmail.com
Mileta Žarković, Univerzitet u Beogradu -Elektrotehnički fakultet, mileta@etf.rs

architectures due to their efficiency. Transformers, thanks to the attention mechanism and their large capacity, can process and derive insights from vast amounts of data. It is also possible to use hybrid architectures composed of various building blocks.

This paper presents the application of the TiDE (Time-series Dense Encoder) neural network for forecasting electricity generation and consumption in Serbia based on real data from the first half of 2019. Additionally, meteorological data from the Republic Hydrometeorological Service of Serbia (RHMZ) for the same period were used. The forecast was carried out seven days in advance with an hourly resolution. The paper describes the algorithm for applying the TiDE neural network and the resulting errors in the time-series forecasts.

Keywords: *power system, forecasting, generation, neural networks, autoencoder*

1 INTRODUCTION

In recent years, there has been a lot of advancement in long-term time-series forecasting, namely new neural network architectures specifically designed to tackle such problems. Interestingly, many completely disparate solutions have achieved state-of-the-art (SoTA) which shows that there are many ways to achieve energy production and consumption forecasting and that the results depend on both the neural network and the data itself.

In this work we present a method of training a neural network, Time Series Dense Encoder, TiDE [1] on electricity production from different sources as well as consumption whilst utilising reversible instance normalization, RevIn [2]. Same method is applied to two datasets, first being data from 2019 in Serbia and second containing only renewable energy production and total consumption from 2020.

2 BACKGROUND

Recurrent neural networks are well suited for time-series forecasting. Nonetheless several factors need to be taken into account, as they mostly process the data sequentially. Long horizon and even longer look-back make the use of recurrent networks such as LSTM inefficient and additional techniques need to be implemented in order to preserve information about periodicity of the data. Example of a modern neural network for time-series prediction based on recurrent architecture is DeepAR [9].

Transformer neural networks are well suited for large quantities of information and are able to find relations in seemingly unconnected data. This is mostly achieved through a mechanism called self-attention, the main idea being that each individual data point is observed in the context of its peers [8]. This gives transformers another benefit of having more explicit memory than the feed-forward or even recurrent networks. Main weakness of the transformers is difficulty of training which consists of both low efficiency and tendency to overfit. Some of the most notable transformer architectures today are Informer [3], FedFormer [4], AutoFormer [5], PatchTST [6], Temporal Fusion Transformer [7].

Another equally viable option are architectures based on a simple feed-forward network. While transformers certainly have some benefits, other, simpler and more efficient networks may achieve better results on some forecasting problems. That observation is demonstrated in the work: "Are Transformers Effective for Time Series Forecasting?" [10]. Recently a few works have sparked new research in this direction, N-BEATS [11] and hierarchical version N-HiTS

[12] in particular. It would be beneficial to compare results from a network from this category for example TSMixer [13] with results given in this work.

Autoencoders are another popular option due to their efficiency, universality and excellent performance. Main idea is that input data is passed through a bottle-neck of sorts and the network is supposed to produce the same data on the output as is the input with as little loss of information as possible. That way, even unlabeled data can be used for training, since input and output data are the same. Internal representation which is often much lower dimensionality than the input data while preserving most of the information can be seen as a distilled down or compressed view of the data. This side effect is sometimes what is more important than the output itself.

3 COVARIATES

Electricity consumption, and therefore production, depend on many factors. The majority of the factories have some form of shift work, however automated the processes may be. Households vary their consumption depending on the time of day or the day of the week. Holidays represent special exceptions in consumption, and they are known in advance. Weather like severe frost or extreme drought also imply periods of increased electricity consumption. Weather changes in the environment have an impact on consumption in neighboring countries and therefore indirectly on the amount of exchange. Renewable sources are particularly vulnerable to weather conditions. Wind farms are used to the maximum when the wind is strong enough but within the permitted limits. It is necessary that the model for the prediction of production and consumption takes into consideration covariate variables.

Correlated variables or covariates can roughly be divided into past, future and static. Past covariates by definition are known only until the moment after which we make the prediction. Therefore, we use them as well as the data we operate on prediction, but their value in the future is of no importance to us. Measured values are the most common and the most obvious example of this category. Future covariates can also be some values for which we don't have exact data, for example the weather forecast, but they have a big impact on the outcome of the prediction, that is, in reality, weather conditions have a dominant influence on production from renewable sources. The group of static covariates includes holidays and fixed facts known in advance.

4 TIME-SERIES DENSE ENCODER

Given the complexity of the problem, we need a network that can get the most useful information out from all the data, either from sequences that are of interest to us for prediction or from variables that affect them. The architecture TiDE was chosen, which attributes great importance to combining covariates with predicted sequences.

In the first step, the neural network reduces the dimensionality of the input covariates. Then it merges that representation with static attributes and historical data. The data thus combined are then encoded in order to obtain an internal representation, that is, to extract the essence from the input data.

The condensed data is further decoded to return to the form with original dimensions so that they could be concatenated together.The last step is the application of a time decoder to obtain the final sequence of predictions up to event horizon.

In parallel with the described part of the network, there is a residual linear network connection that enables the propagation of historical data. This guarantees that the entire network does not miss any information in any way in the process of encoding and decoding.The residual block, which was used as a building element of the network, is shown on the right part of the picture.

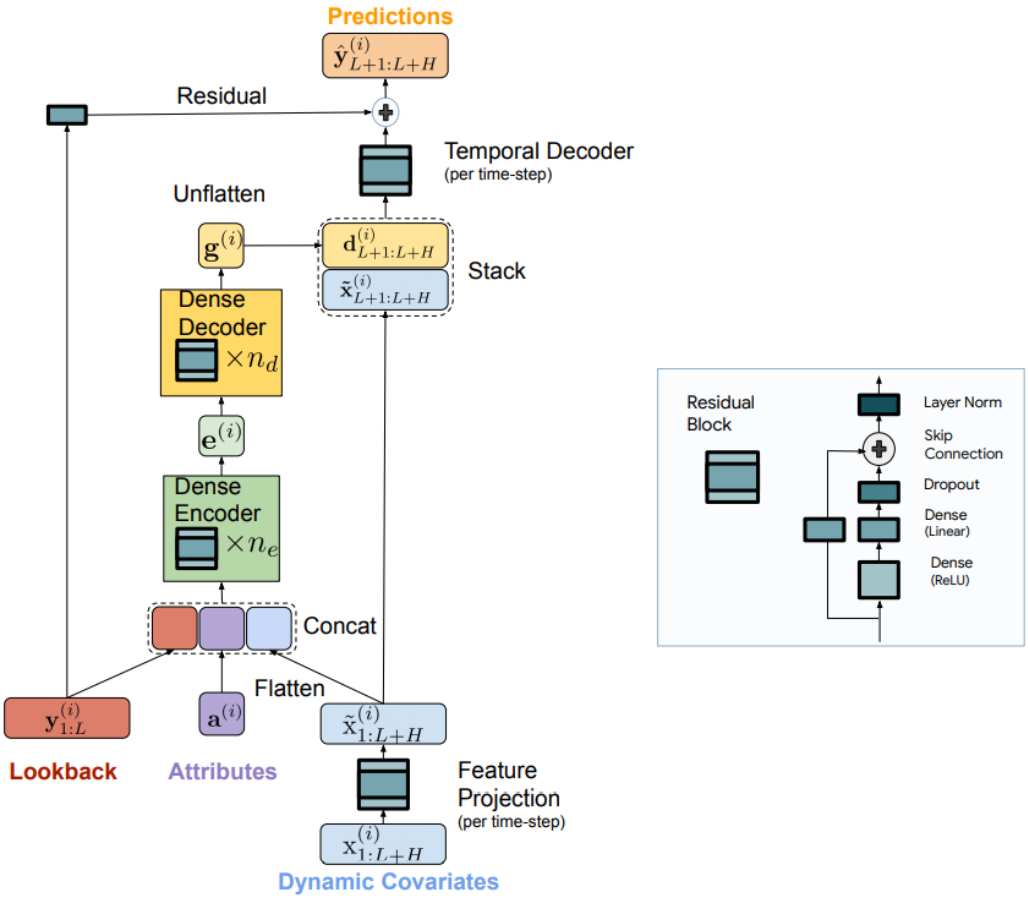

Figure 1. Overview of the TiDE neural network architecture[1]

5 TRAINING

The data were divided into sets for training, validation and test in the ratio of 80%, 10% and 10% respectively. The prediction was made seven days in advance, i.e. the horizon is 168 steps in the future. Optimization of the learning rate (learning rate) was performed using a simple exponential decreasing method with a multiplier of 0.99. The early stopping method was used to avoid overfitting.

The value was selected from the range of 0.9 to 0.999, based on multiple attempts. A value was chosen that does not reduce the precision of the results with a milder reference to the acceleration of the training itself. For the error function is the chosen method of mean square deviation, MSELoss. The technical implementation of the training was done using the darts library [14] which was built on to the popular pytorch library.

---

[1] Time-series Dense Encoder, https://arxiv.org/pdf/2304.08424

All data, when monitored in a sufficiently long interval, exhibit a distribution shift. In other words, the mean and variance are not constant. That makes it difficult in many ways to train the model and also makes the model imprecise. Reversible instance normalization was used as a solution to this problem. A simple but effective method that can be generally applied because uses two symmetric parts: the first for normalization by which the distribution of input data is translated into normal distribution and others for denormalization, which nullifies the effects of normalization and data distribution is returned back to its original form. The parameters of the transformation are variable and are adjusted in the process of training together with other weights of the network.

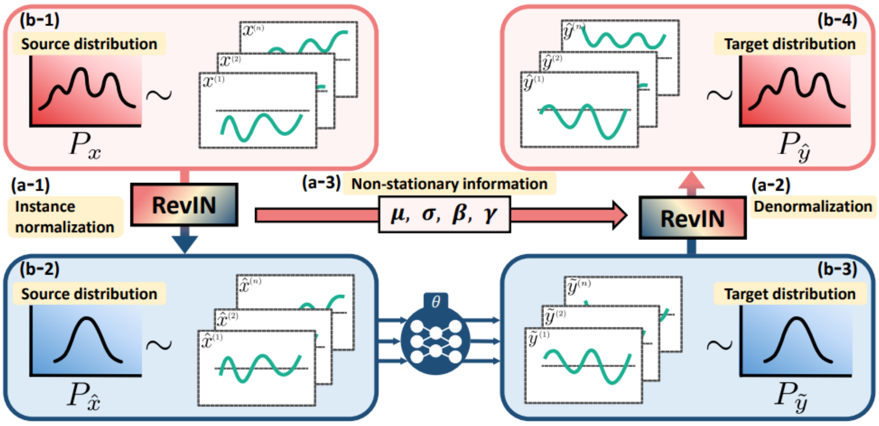

Figure 2. Working principle of the RevIN method. Beta and gamma parameters are trained.[2]

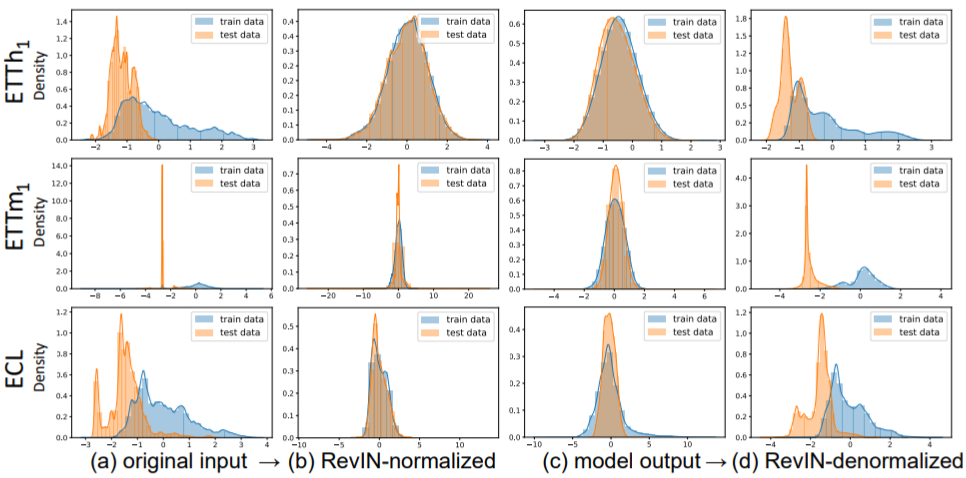

Figure 3. Graphic representation of normalization and denormalization using the RevIN method[3]. An example performed on well known datasets ETT[4] and ECL[5]

[2] Source: RevIn https://openreview.net/pdf?id=cGDAkQo1C0p

[3] Source: RevIn https://openreview.net/pdf?id=cGDAkQo1C0p

[4] Electricity Transformer Temperature, https://arxiv.org/abs/2012.07436

[5] From Numbers to Words, https://arxiv.org/abs/2401.12652

7 RESULTS

Prediction was made 168 steps into the future, which in this particular case translates to seven days with one hour interval. The next few pictures show a graphical representation of the prediction. It is observed that the model grasps the periodicity of the signals, even when there is more than one frequency, on a daily and weekly basis. Sudden changes in amplitude are not fully covered by the prediction, but rather it follows historical values. Even the high-frequency component of the signal is mostly present. Simple limitation of the model output to positive values, for quantities that can not be negative, would improve the accuracy of the prediction.

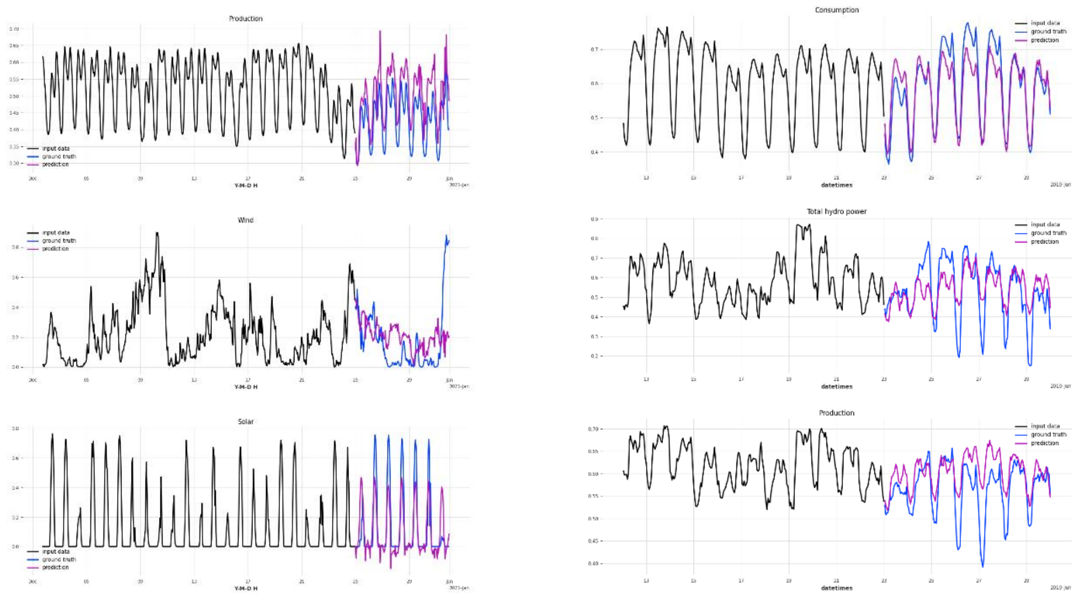

Figure 4. Prediction graphs for data from 2020 (left) and 2019(right)

Table 1. Numerical errors of prediction for data from 2020 (up) and 2019(down)

Prediction series	MAE	MSE
Production	0.074090	0.006416
Wind	0.091468	0.020600
Solar	0.156717	0.043977

Prediction series	MAE	MSE	MAPE
Hydro Power	0.082854	0.012151	21.85584
Production	0.041335	0.002922	7.921187
Consumption	0.037352	0.002148	6.132026

8 CONCLUSION

We presented an application of a state-of-the-art neural network on actual energy production and consumption data from Serbia. Although the results look promising, they leave some open questions. What additional trends could be seen if the data were collected in the period of several years or even decades? Would a higher sampling frequency reveal some phenomena that are otherwise masked? In addition to water levels in reservoirs and meteorological data, what other parameters (covariates) would be important for prediction?

The idea for improving the used architecture is reflected in the fact that a tandem encoder-decoder is symmetrical (three blocks were used, which gave the best results) and that each block encodes everything it compresses information more, reminiscent of the well-known U-Net image segmentation network [15]. Mentioned network has a direct connection at each level of the convolutional pyramid so that information can bypass the compression process if necessary. That way the maximum of useful information is extracted from each input.

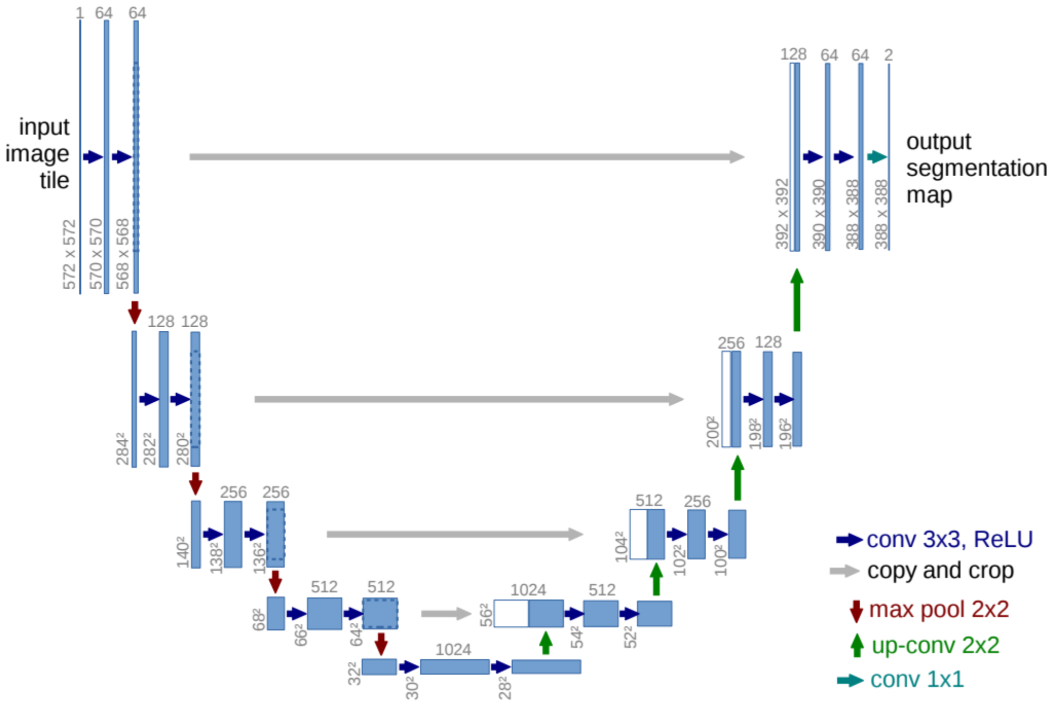

Figure 5. Overview of the U-Net neural network architecture[6].

[6] Source: U-Net https://arxiv.org/abs/1505.04597

9 LITERATURE

[1] Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan Mathur, Rajat Sen, Rose Yu (2023) Long-term Forecasting with TiDE: Time-series Dense Encoder. URL https://arxiv.org/abs/2304.08424

[2] Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, Jaegul Choo (2022) Reversible Instance Normalization for Accurate Time-Series Forecasting against Distribution Shift. URL https://openreview.net/pdf?id=cGDAkQo1C0p

[3] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. (2021) Informer: Beyond efficient transformer for long sequence time series forecasting. In Proceedings of the AAAI conference on artificial intelligence. URL https://arxiv.org/abs/2012.07436

[4] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. (2022) Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In International Conference on Machine Learning, pages 27268–27286. PMLR. URL https://arxiv.org/abs/2201.12740

[5] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. (2021) Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems, 34: 22419–22430. URL https://arxiv.org/abs/2106.13008

[6] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. (2022) A time series is worth 64 words: Long-term forecasting with transformers. International conference on learning representations. URL https://arxiv.org/abs/2211.14730

[7] Bryan Lim, Sercan O. Arik, Nicolas Loeff, Tomas Pfister (2019) Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. URL https://arxiv.org/abs/1912.09363

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin. (2017) Attention is all you need. Advances in neural information processing systems, 30. URL https://arxiv.org/abs/1706.03762

[9] David Salinas, Valentin Flunkert, Jan Gasthaus (2019) DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. URL https://arxiv.org/pdf/1704.04110

[10] Ailing Zeng, Muxi Chen, Lei Zhang, Qiang Xu (2020) Are Transformers Effective for Time Series Forecasting? URL https://arxiv.org/abs/2205.13504

[11] Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, Yoshua Bengio (2020) N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. URL https://arxiv.org/abs/1905.10437

[12] Cristian Challu, Kin G. Olivares, Boris N. Oreshkin, Federico Garza, Max Mergenthaler Canseco, Artur Dubrawski (2022) N-HiTS: Neural Hierarchical Interpolation for Time Series Forecasting. URL https://arxiv.org/abs/2201.12886

[13] Si-An Chen, Chun-Liang Li, Nate Yoder, Sercan O. Arik, Tomas Pfister (2023) TSMixer: An All-MLP Architecture for Time Series Forecasting. URL https://arxiv.org/abs/2303.06053

[14] Julien Herzen, Francesco Lässig, Samuele Giuliano Piazzetta, Thomas Neuer, Léo Tafti, Guillaume Raille, Tomas Van Pottelbergh, Marek Pasieka, Andrzej Skrodzki, Nicolas Huguenin, Maxime Dumonal, Jan Kościsz, Dennis Bader, Frédérick Gusset, Mounir Benheddi, Camila Williamson, Michal Kosinski, Matej Petrik, Gaël Grosch (2022) Darts: User-Friendly Modern Machine Learning for Time Series. URL https://jmlr.org/papers/v23/21-1177.htm

[15] Olaf Ronneberger, Philipp Fischer, Thomas Brox (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. URL https://arxiv.org/abs/1505.04597