STATISTICAL AND MACHINE LEARNING MODELS COMPARISON FOR DEMAND FORECASTING

T.V. Koryakina, student 8PM1I
Tomsk Polytechnic University
E-mail: tvk26@tpu.ru

Introduction

The work of any business heavily relies on accuracy of daily demand forecasting. It dictates level of inventory holding, capacities of warehouses and production sites. This task is usually solved with well-known statistical methods. More recently, big companies saw potential in utilizing Machine Learning (ML) algorithms as alternatives to statistical approach. It has yet to be validated that ML has significant advantage. This paper will provide overview of different papers attempted to do a comparison of forecasting power of statistical and ML methods for time-series data.

Approach description

To properly compare the performance of ML methods to statistical ones, were considered the most popular forecasting methods of each kind as well as some of their variants that, according to the literature, can lead to better forecasts.

In all articles time series data was used as an input for traditional and ML models. The products were split by typical demand patterns (figure 1).

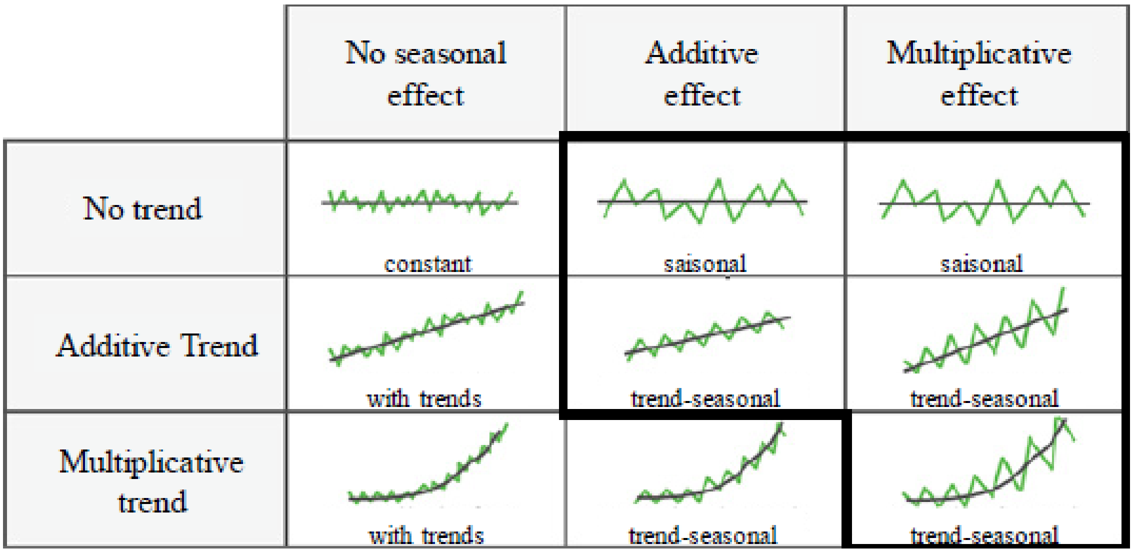

Fig. 1. Demand Patterns resulting from Trend and seasonal components [1]

I analyzed a data from three papers:

1. Machine Learning and Statistics: A Study for assessing innovative Demand Forecasting Models (2021). DOI: 10.1016/j.procs.2021.01.127
2. Comparison of statistical and machine learning methods for daily SKU demand forecasting (2020). DOI: 10.1007/s12351-020-00605-2
3. Statistical and Machine Learning forecasting methods: Concerns and ways forward (2018). DOI: 10.1371/journal.pone.0194889

To compare the individual forecasting models with each other, authors implemented every model using optimal parameters. To analyze statistical models focus was on trend and seasonality.

The ML and DL models were tested using the respective standard configuration. I took into consideration that for business besides forecast accuracy it's also crucial consider the cost of implementation of the model.

The summary table for used models and their output shown below at Figure 2.

Statistical Methods	Paper 1	Paper 2	Paper 3
Performanced criteria	The relative deviation of root-mean-square deviation, RMSE	Terms of bias (AMSE)	Symmetric mean absolute percentage error (SMAPE)
Triple exponential smoothing (ETS)	76,40%		7,19%
Average Extended (SARIMAX)	80,80%		7,34%
Simple exponential smoothing (SES)		88,10%	7,36%
Syntetos–Boylan Approximation (SBA)		75,70%	
Croston's method (CRO)		77,30%	
ML			
Multilayer Perceptron (MLP)	71,60%	81,70%	8,39%
Long-term short-term memory (LSTM)	81,90%		11,67%
Extreme Gradient Boosting (XGBoost)	53,80%		
Random Forest (RF)	43,50%	66,00%	
Support Vector Regression (SVR)		67,00%	8,88%
k-Nearest Neighbour Regression (kNNR)		68,40%	11,49%
Bayesian Neural Network (BNN)			8,17%

Fig. 2. Comparison of statistical and ML models

There is a trade-off from between the level of accuracy and complexity and cost for building a model[5]

You can see that at Figure 3. Even if we increase accuracy the technology might be not viable from deployment point of view.

Conclusion

After analyzing several papers, it can be concluded that there is no strong evidence that machine learning models perform better than popular statistical models. It shows that for some datasets accuracy can be improved over statistical models, but the same ML method can show both great and poor results depending on what kind of data we have. Having that in mind businesses should take into consideration cost for deploying a model and run several tests with both statistical and machine learning models, before making decision about deployment one.

References

1. Scheer, August-Wilhelm (1983) "Absatzprognosen." Berlin, Heidelberg, Springer-Verlag.
2. Machine Learning and Statistics: A Study for assessing innovative Demand Forecasting Models (2021). DOI: 10.1016/j.procs.2021.01.127
3. Comparison of statistical and machine learning methods for daily SKU demand forecasting (2020). DOI: 10.1007/s12351-020-00605-2
4. Statistical and Machine Learning forecasting methods: Concerns and ways forward (2018). DOI: 10.1371/journal.pone.0194889
5. Chambers, Mullick, and Smith (1971)