Time-Series Forecasting Using Feature Based Hybrid Approach

Olashile Adebimpe.

Agenda

Introduction — Provide insight into the problem that motivated our research

Reviews — Related research that has been done to address the problem

Methods — What out hybrid methodology entails with description

Results — Experimental results

Conclusion — Research Conclusion, limitations and future works

Agenda

Introduction

Provide insight into the problem that motivated our research

Reviews

Related research that has been done to address the problem

Methods

What out hybrid methodology entails with description

Results

Experimental results

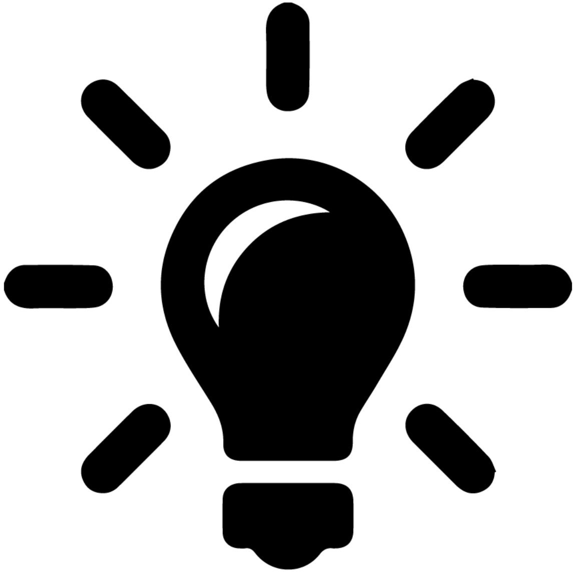
Conclusion

Research Conclusion, limitations and future works

Introduction

What is Forecasting?

*Forecasting is basically using a subjects **historical data** to build a **model** and then predicting certain outcomes in the **future based** on the same model.*



Why Forecasting?

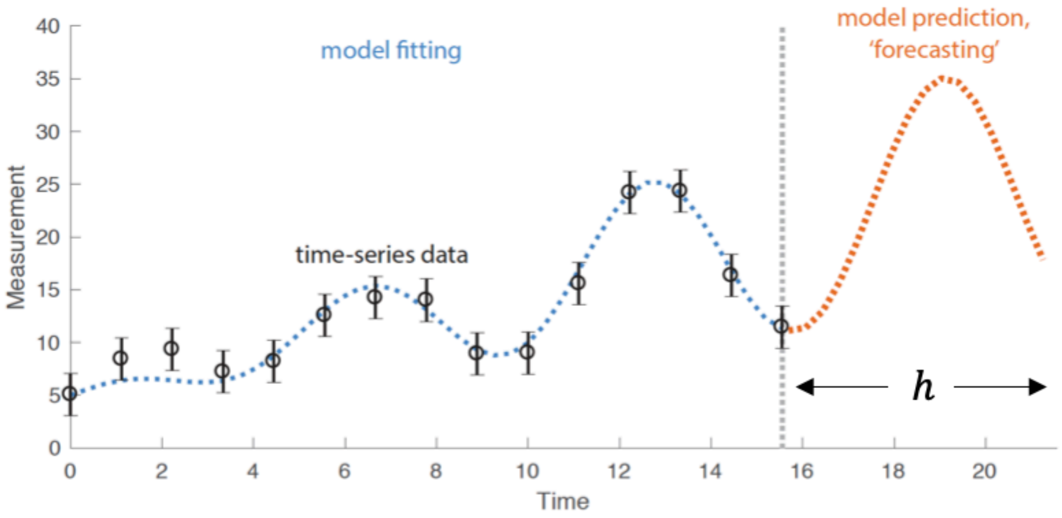
*To learn how to **improve forecasting accuracy** and how these learning can be applied to **advance the theory and practice of forecasting.***

Introduction

$$y_{t+h} = f(\langle DescriptorsOfThePast \rangle)$$

where

- $y_1, y_2, \dots, y_t$ *are past observations*
- y_{t+h} *are future values*
- h *is the forecast horizon*
- y_i *is the value of* Y *measured at time* i



What are these models made of?

Traditional or Classical Forecasting Techniques	Soft Computing-Based Forecasting Techniques
Regression Methods , Multiple Regression Methods , Exponential Smoothing	*Multilayer Perceptron (MLP) , CART Regression Trees (CART), Support Vector Regression (SVR), Neural Networks*
Fuzzy Based Forecasting Techniques	**Combinatory method**
Fuzzy Logic.	*Any combination of the other 3 methods*

Agenda

Introduction

Provide insight into the problem that motivated our research

Related research that has been done to address the problem

Reviews

Methods

What out hybrid methodology entails with description

Experimental results

Results

Conclusion

Research Conclusion, limitations and future works

Review

	Model	Pros	Cons	Reseach
Statistical/ Classical Methods	Linear Regression	Ability to handle different time series components and features. High interpretability.	Sensitive to outliers Strong assumptions.	Mbamalu (Mbamalu 1993)
	Exponential Smoothing	Ability to handle variable level,trend and seasonality components. Automated Optimization.	Sensitive to outliers Narrow confidence intervals	Brown (Brown 1959) Holts (Holt 2004)
	ARIMA(Auto Regressive Integrated Moving Avearge)	High interpretability. Realistic confidence intervals. Unbiased forecasts.	Requires more data Strong restrictions and assuptions Hard to automate	ARIMA(Zhang 2003)
Stochastic / Machine Learning Methods	Dynamic Linear Model	High Interpretability More transparent than other models Deals well with uncertainty. Control the variance of the components.	Higher holdout errors Higher training and evaluation time.	X. Yan(X. Yan 2013) X. Yan (X. Yan 2014) Frohlich(Frohlich 2004)
	Neural Network Model	Less restrictions and assumptions Ability to handle complex nonlinear patterns High Predictive power Can be easily automated	Low interpretability. Difficult to derive confidence intervals for the forcast. Requires more data. Overfitting of Data. Shortage of plenty training time. Optimal network parameters determination.	P. G. Zhang (P. G. Zhang 2003) J. Faraway (J. Faraway 2008) C. Hamzaçebi(C. Hamzaçebi, 2008)
Fuzzy Based Methods	Fuzzy Based Models	Used for forecasting problems with more than one value of attributes Employes the concept of Fuzzy Logic.	Fuzzy rules are complex to set up	E. Egrioglu (E. Egrioglu 2013)

What happened?

> *It should become clear that ML methods are not a panacea that would automatically improve forecasting accuracy. Their capabilities can easily generate implausible solutions, leading to exaggerated claims of their potentials and must be carefully investigated before any claims can be accepted*
>
> *Ahmed, Atiya, El Gayar, & El-Shishiny, 2010*

What's Next?

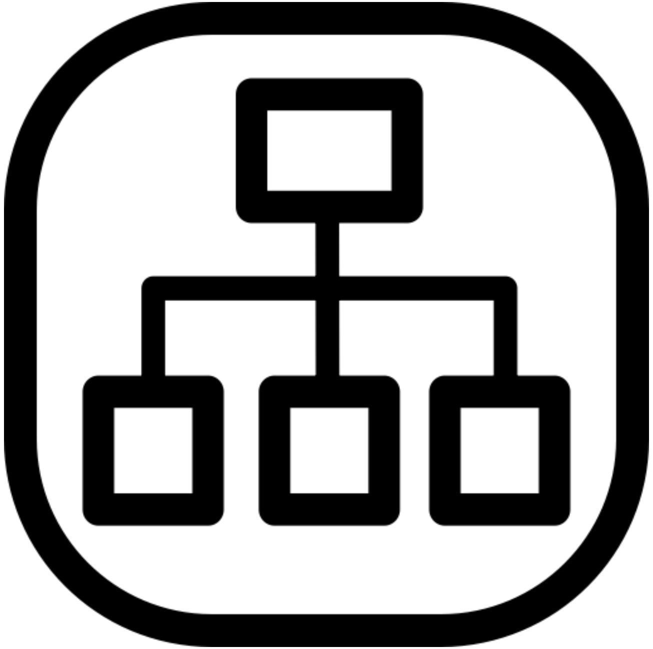
> *"Our conclusion from the M4 Competition is that the accuracy of individual statistical or ML methods is low, and that hybrid approaches and combinations of method are the way forward for improving the forecasting accuracy and making forecasting more valuable"*
>
> *Makridakis, Spiliotis, and Assimakopoulos 2018*

Our Hybrid Methodology

- *We propose a technique that decomposes time series into linear and complex systematic structures through a decomposition procedure.*

- *Employs diverse methods such as Statistical techniques and Machine Learning methods to separately model and predict linear and complex structures of the time series respectively.*

- *Our proposed hybrid model merges prediction obtained from these diverse -based models using with a goal to improve forecasting accuracy.*



Research Questions

Decomposition approach Vs Other combinatory Methods

Hybrid methodology vs other individual state-of-the-art classical approaches

How effective are Machine Learning or Classical models in hybrid methodology

Agenda

Introduction

Provide insight into the problem that motivated our research

Reviews

Related research that has been done to address the problem

Methods

What out hybrid methodology entails with description

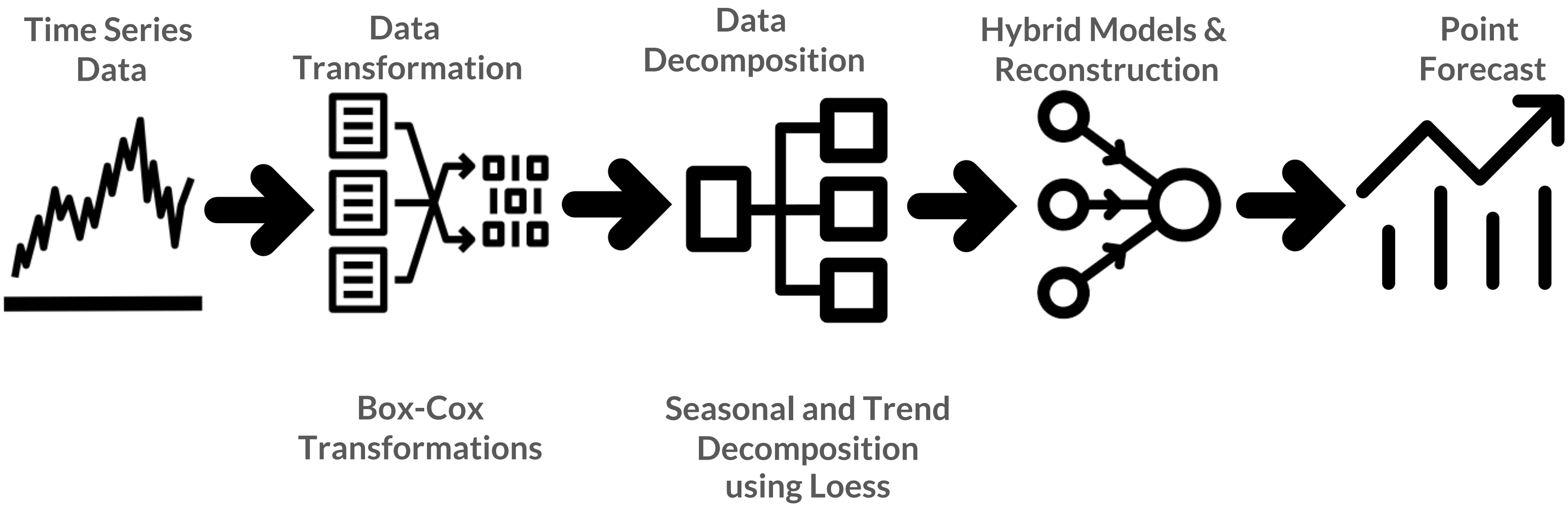
Results

Experimental results

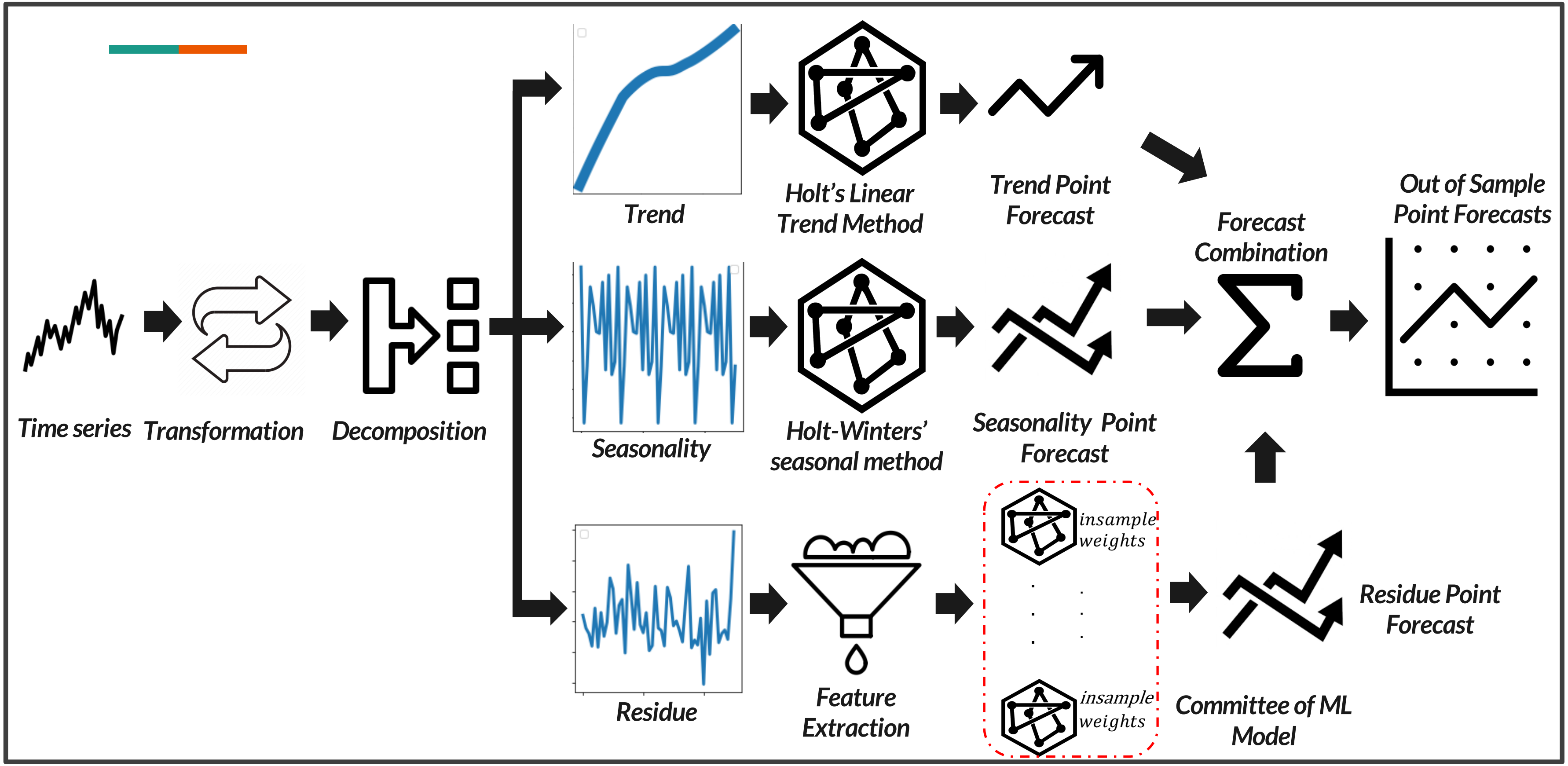
Conclusion

Research Conclusion, limitations and future works

Overview of Model Architecture

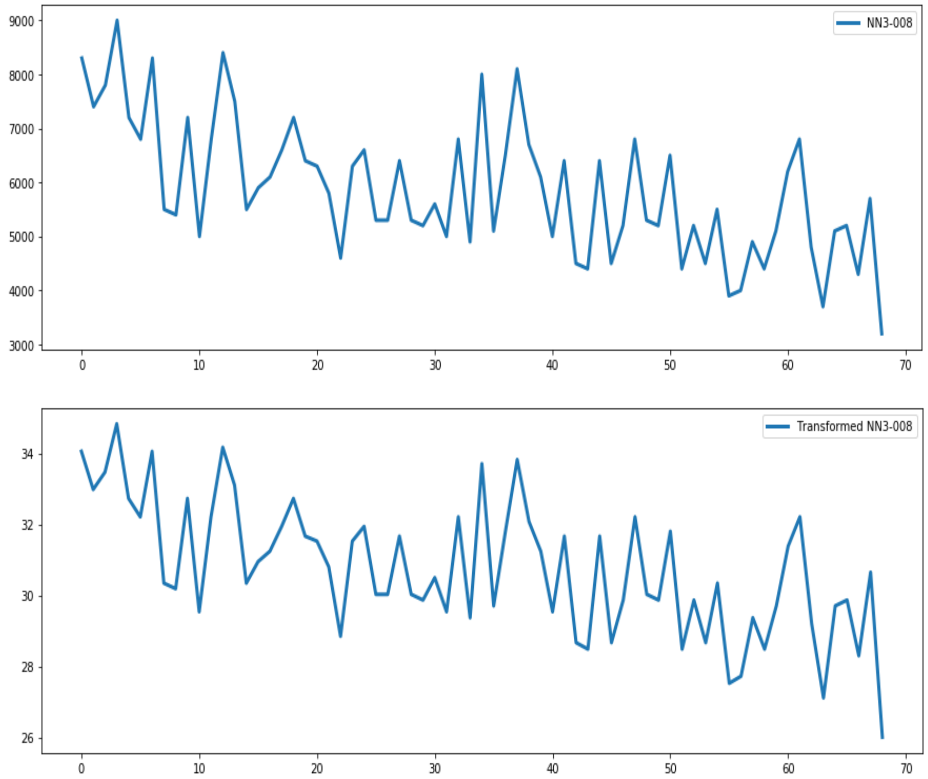


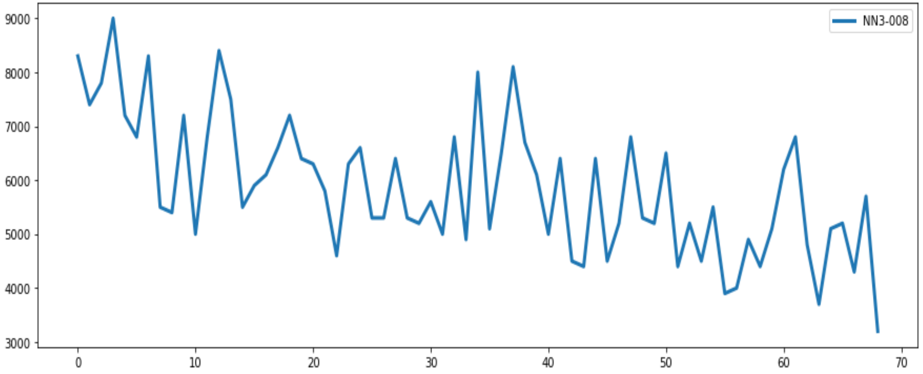
Full Model Architecture

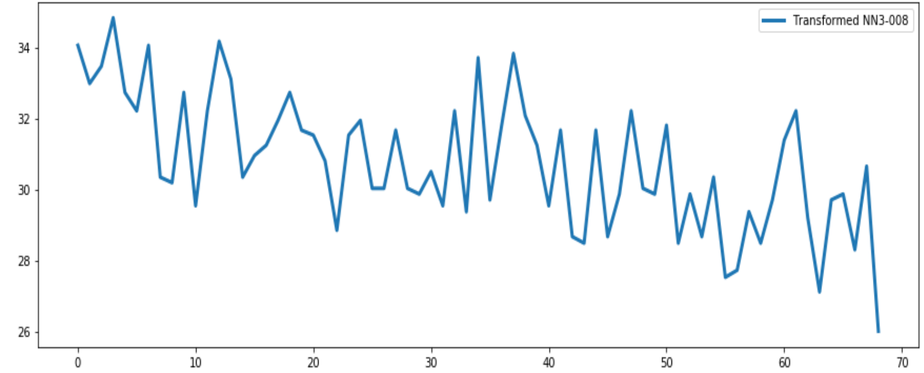


Data Transformation

- *Rescaling of historical data to a simple pattern.*
- *Example: Logarithmic transformation, Square root transformation, Power transformations and The Box-Cox transformation*





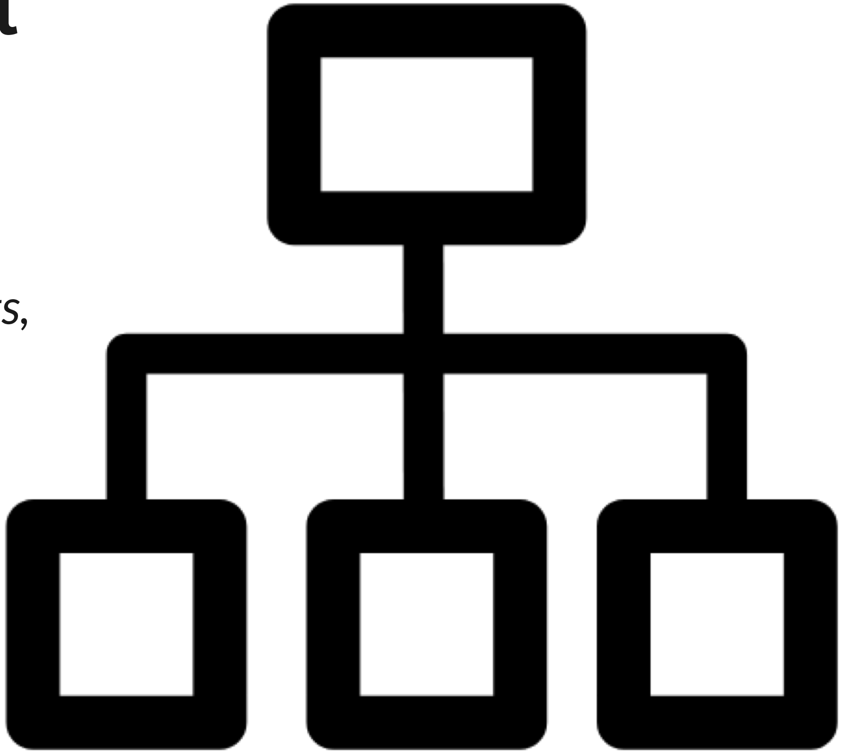


Box-Cox Transformation

- *Created by Box and Cox to stabilize the variance of a time series.*
- *Includes both logarithms and power transformation to remove non-constant variance of a variable thereby making the series look like more normally distributed.*
- *Valuable in removing spurious interactions and helps identify the factors that are really significant.*

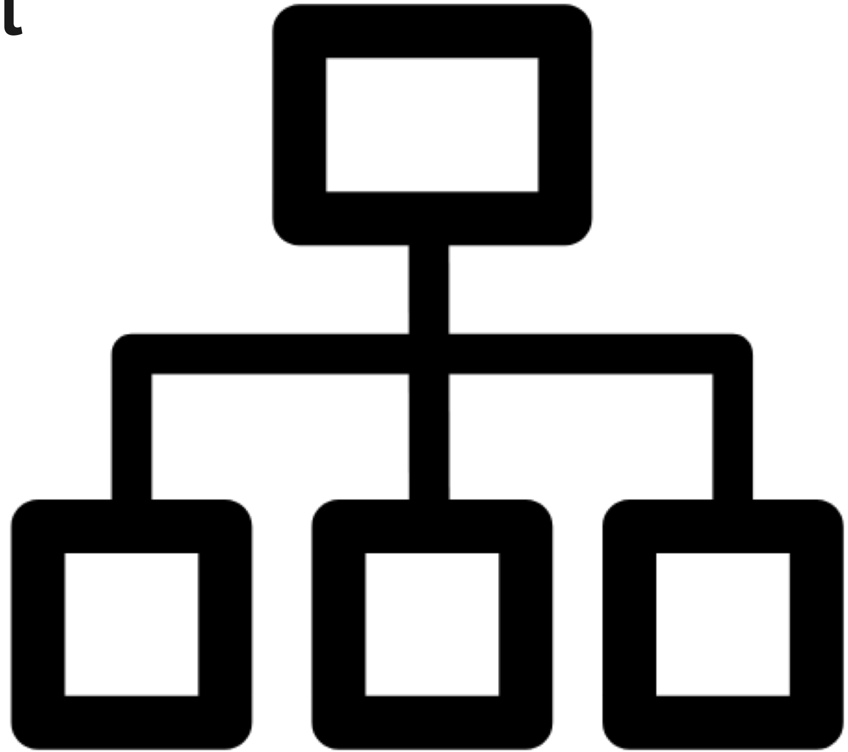
Data Decomposition as a Tool

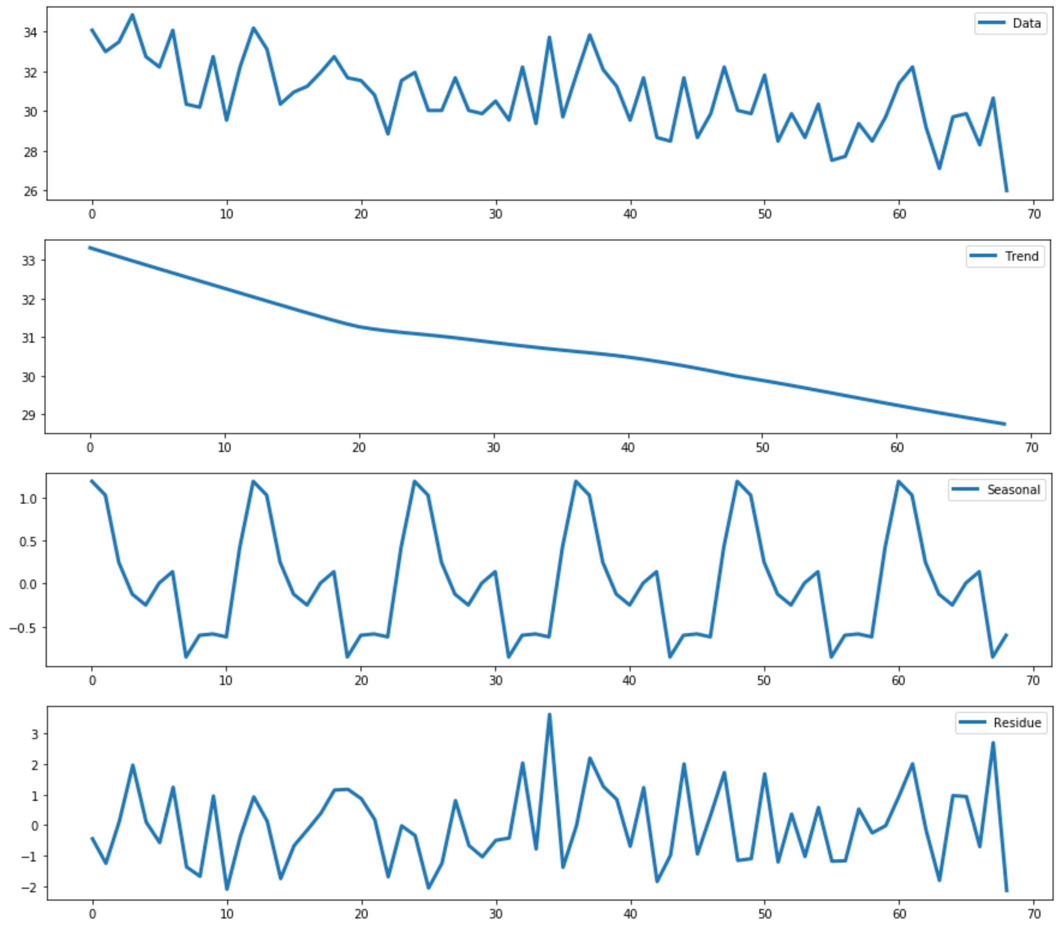
- *A mathematical procedure which transforms historical time series data into various components, where each component is representing an underlying pattern.*
- *Time series patterns:* ***Trend, Seasonal and Residue.***



Data Decomposition as a Tool

- *Helps improve understanding of the time series.*
- *Helps improve forecast accuracy.*
- *Helps exploring the historical changes over time.*





STL Decomposition

- *Seasonal and Trend decomposition using Locally Weighted Regression and Scatterplot Smoothing.*
- *Unlike classical decomposition, it handle any type of seasonality.*
- *The seasonal component is allowed to change over time*
- *Robust to outliers.*

Feature Extraction

- *With the help of a moving sliding window, set of features were extracted from the residual component of the series.*

- *Vectors of features measuring the characteristics of each time series are used to train machine learning models such as The boosting algorithm and Support Vector Regression.*

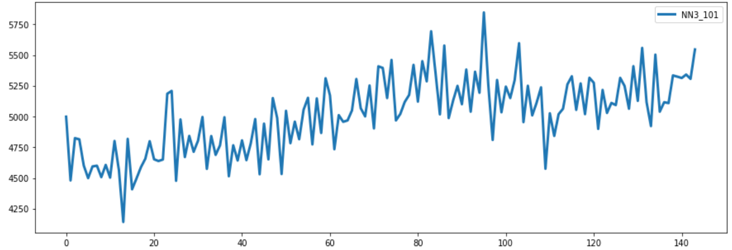
- *The feature-based approach to time series can also be used to identify the best forecasting model using a pre-trained classifier,*

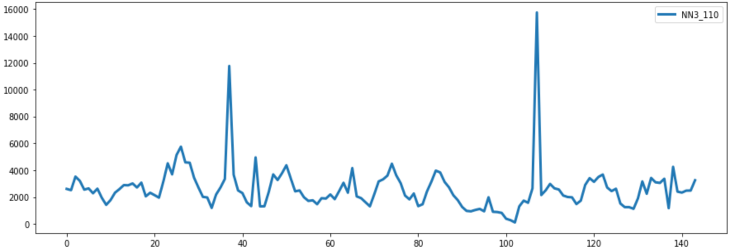
- *Implemented in the tsfresh python package*

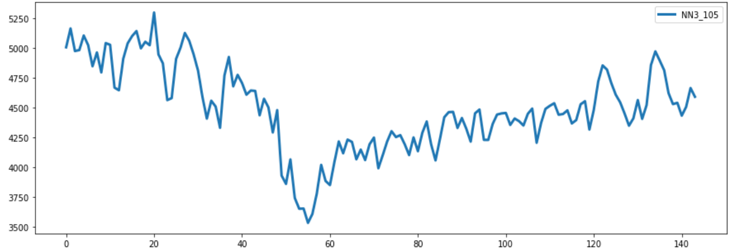
Feature
Mean
Standard Deviation
Variance
Skewness
Kurtosis
Mean change
Minimum
Maximum
Mean of absolute Change
Linear Trend
Aggregation Autocorrelation
Autocorrelation
Pairwise Correlation
Interquartile Range
…
…

Kang, Hyndman & Smith-Miles. "Visualising forecasting algorithm performance using time series instance spaces" (IJF 2017).
Talagala, Hyndman and Athanasopoulos. "Meta-learning how to forecast time series" (2018).

Evaluation: Datasets







- ***Dataset A*** *is complete dataset of* ***111*** *different monthly time series used in the NN3 forecasting competition which was drawn from homogeneous population of empirical business time series*

- ***Dataset B*** *is a subsample of* ***11*** *time series from the 111-time series and is therefore contained in the larger dataset (validation dataset)*

- ***Dataset C*** *is large subset of* ***48,000*** *real-life monthly time series used in the M4 competition mainly from the business, finance and economic world characterized by considerable seasonality, some trend and a fair amount of randomness.*

How do we Evaluate?

- ***The Absolute Error Metrics***

  Mean Absolute Error (MAE), Mean Square Error (MSE), Geometric Mean Absolute Error (GMAE).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \gamma_i|}{n}$$

- ***The Percentage Error Metric***

  Mean Absolute Percentage Error (MAPE). Symmetric MAPE (sMAPE)

$$sMAPE = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|y_i - \gamma_i|}{(|y_i| + |\gamma_i|)/_{2}}$$

- ***The Scaled Free Error***

  Mean Absolute Scale Error (MASE)

$$MASE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \gamma_i|}{\frac{i}{n-1} \sum_{i=2}^{n} |y_i - \gamma_i - 1|}$$

How do we Evaluate?

- *The Absolute Error Metrics*

  Mean Absolute Error (MAE), Mean Square Error (MSE), Geometric Mean Absolute Error (GMAE).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \gamma_i|}{n}$$

- ***The Percentage Error Metric***

  Mean Absolute Percentage Error (MAPE). Symmetric MAPE (sMAPE)

$$sMAPE = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|y_i - \gamma_i|}{(|y_i| + |\gamma_i|)/_2}$$

- ***The Scaled Free Error***

  Mean Absolute Scale Error (MASE)

$$MASE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \gamma_i|}{\frac{i}{n-1} \sum_{i=2}^{n} |y_i - \gamma_{i-1}|}$$

Agenda

Introduction

Provide insight into the problem that motivated our research

Reviews

Related research that has been done to address the problem

Methods

What out hybrid methodology entails with description

Results

Experimental results

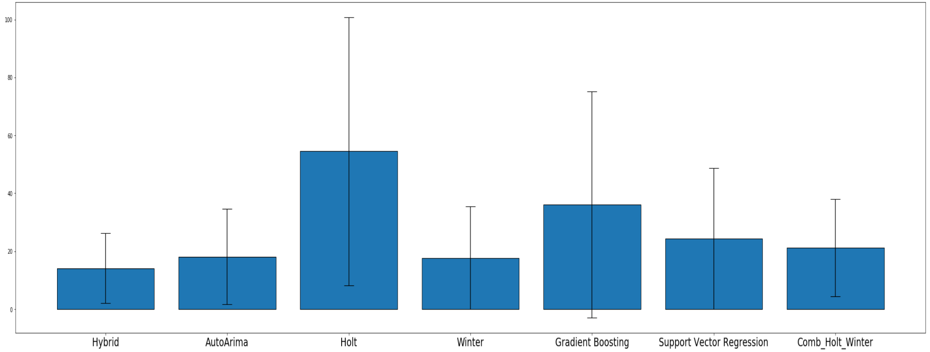
Conclusion

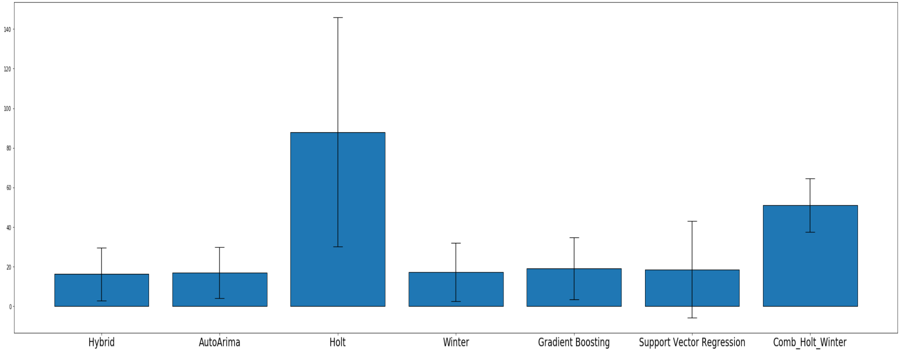
Research Conclusion, limitations and future works

Results: NN3 Datasets

Individual Methods	sMAPE	σ(sMAPE)	MASE	σ(MASE)
Our Hybrid Method	**14.17**	**12.09**	0.912	0.394
AutoArima	18.15	16.52	0.906	0.415
Holts	54.53	46.25	3.966	2.736
Winter	17.71	17.74	1.001	0.660
Gradient Boosting	36.11	39.06	1.983	1.104
Support Vector Regression	24.35	24.37	1.443	0.792
Combination Methods	**SMAPE**	**σ(SMAPE)**	**MASE**	**σ(MASE)**
Holts and Winter	21.22	16.74	1.495	0.992

Individual Methods	SMAPE	σ(SMAPE)	MASE	σ(MASE)
Our Hybrid Method	**16.28**	13.31	1.300	1.787
AutoArima	16.98	12.86	1.228	1.834
Holts	88.05	57.89	9.986	8.576
Winter	17.24	14.71	**1.172**	1.641
Gradient Boosting	19.10	15.67	1.355	1.848
Support Vector Regression	18.66	13.44	1.545	2.512
Combination Methods	**SMAPE**	**σ(SMAPE)**	**MASE**	**σ(MASE)**
Holts and Winter	51.17	48.02	3.642	3.047

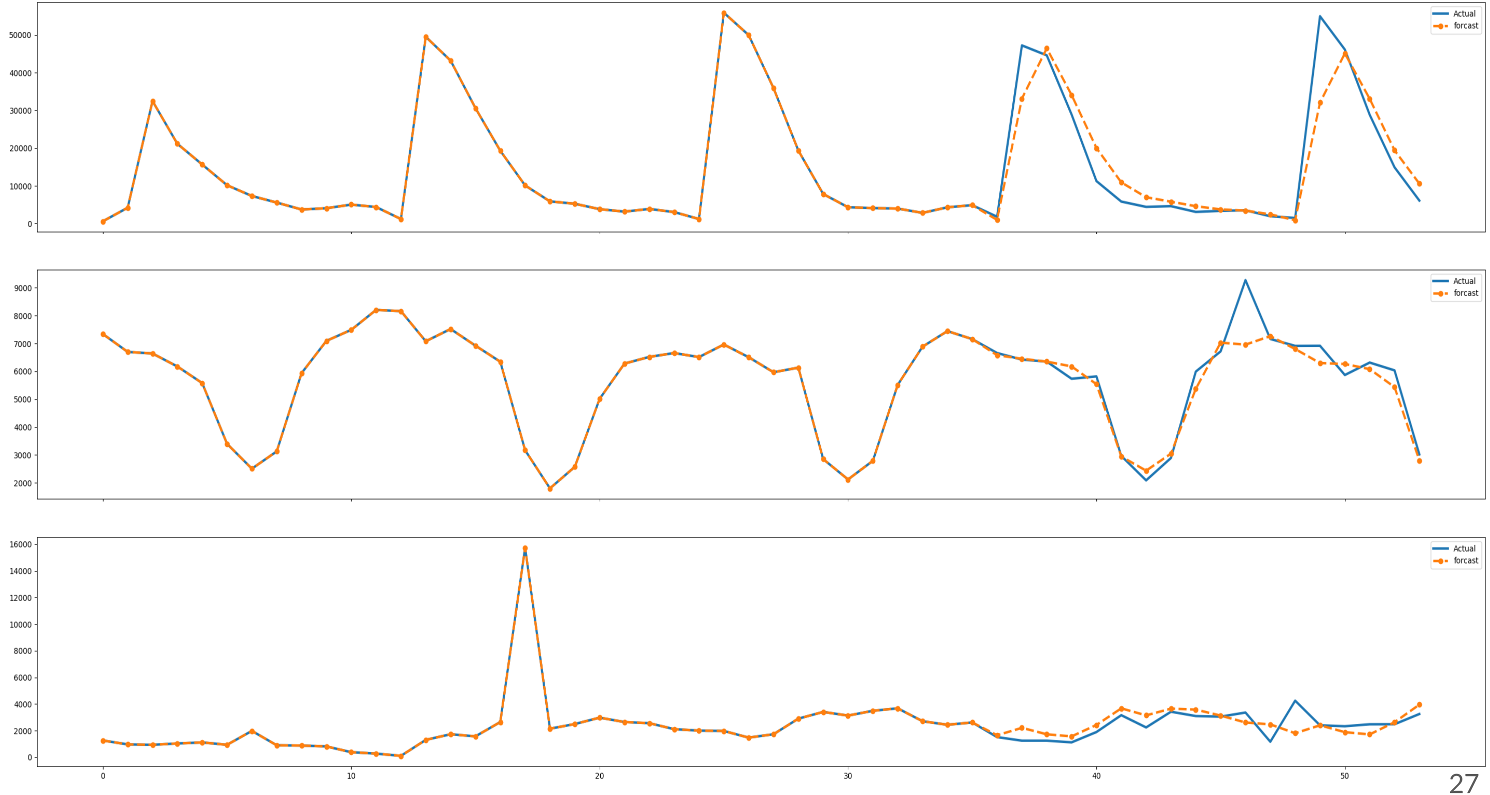




Mean and Standard deviation of NN3 11 and 111 dataset.

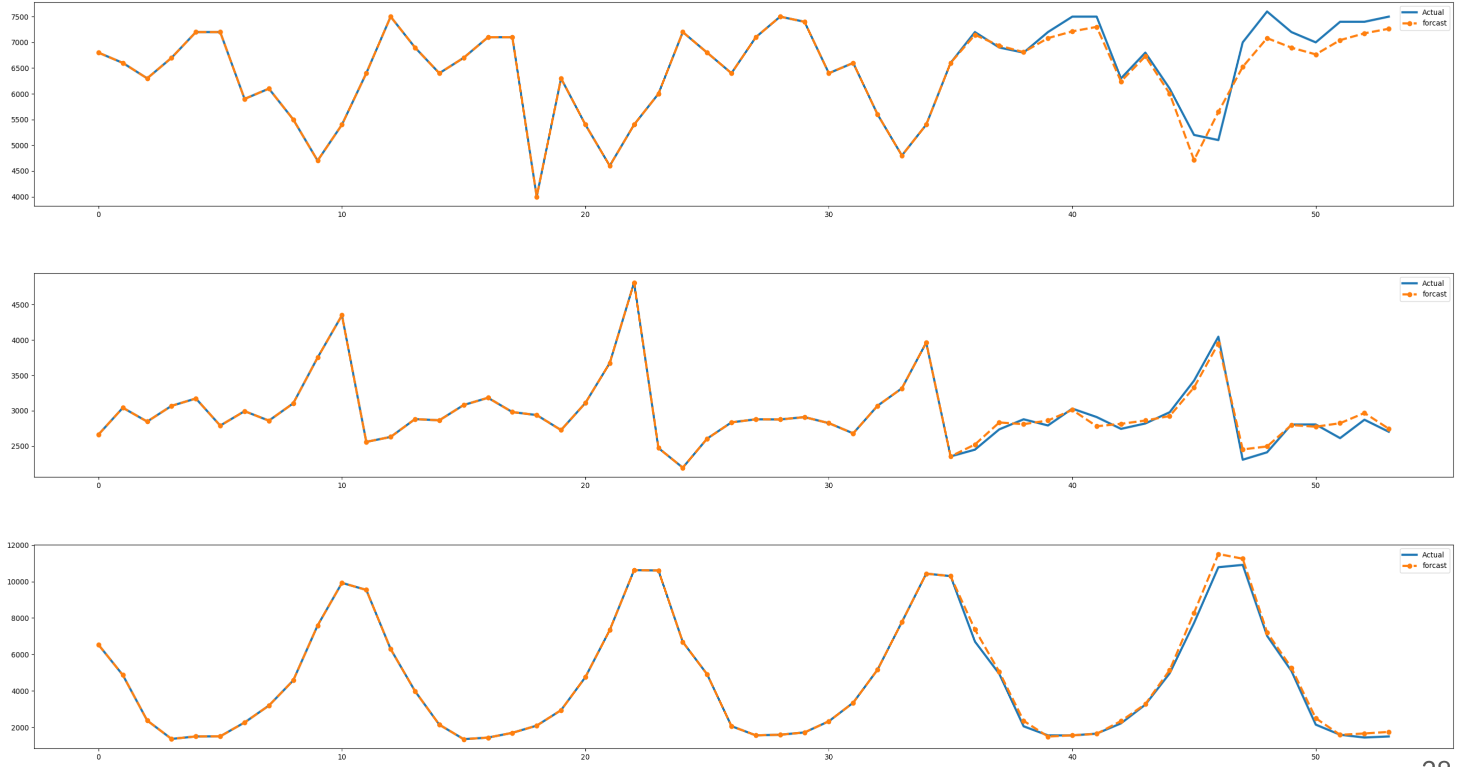
Results: NN3 Datasets

The Actual series and 18 point forecasted values



Results: NN3 Datasets

The Actual series and 18 point forecasted values

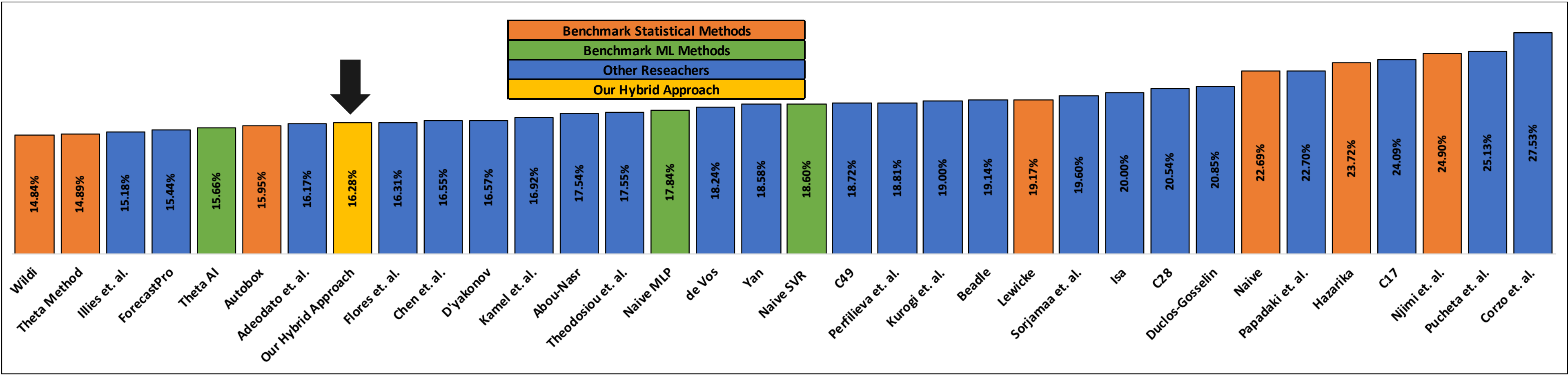


Results: NN3 Datasets

Ranked top 8 methods with the lowest sMAPE

Commercially available software

- *AutoBox (Automatic Forecasting Systems)*
- *Forecast Pro (Business Forecast Systems)*
- *ForecastX (John Galt)*
- *Autocast (Delphus)*



sMAPE of our hybrid method vs published result of the NN3 111 dataset.

Results: M4 Dataset

OWA rank, and relative errors of our proposed model and other benchmark models

Method	sMAPE	Relative sMAPE	MASE	Relative MASE	OWA	Rank
AUTOARIMA	10.767	0.822	2.354	0.755	0.789	1
ETS- Benchmark	11.342	0.866	2.614	0.839	0.852	2
Our Hybrid Method	***10.944***	***0.836***	***2.900***	***0.930***	***0.883***	***3***
WINTER Method	12.034	0.919	2.935	0.942	0.930	4
HOLT- Benchmark	12.851	0.981	2.752	0.883	0.932	5
NAÏVE- Benchmark	13.096	1.000	3.117	1.000	1.000	6
Gradient Boosting	12.409	0.948	3.340	1.071	1.009	7
RNN- Benchmark	18.316	1.399	4.171	1.338	1.368	8
AVEARGE Holts and Winter	21.714	1.658	3.993	1.281	1.469	9
MLP- Benchmark	18.673	1.426	13.011	4.173	2.800	10
HOLT Linear Method	39.305	3.001	9.372	3.006	3.004	11
Support Vector Regression	25.727	1.964	12.618	4.047	3.006	12

Results: NN3

Method	sMAPE
Decision Tree	18.00
Support Vector Machine	18.00
Zoomed ranking, best method	17.50
Neural Network	17.00
Our Hybrid Method	**16.26**
Zoomed ranking, combination	15.50

Performance comparison between our hybrid method and different meta-learning techniques

- ✓*Long Term Horizon*
- ✓*Benchmark results*
- ✓*Individual methods*
- ✓*Machine Learning Methods*
- ✓*Statistical Methods*
- ✓*Combination Methods*

Lemke, C., & Gabrys, B. (2010). Meta-learning for time series forecasting and forecast combination. Neurocomputing, 73(10–12), 2006–2016. https://doi.org/10.1016/j.neucom.2009.09.020

Agenda

Introduction — Provide insight into the problem that motivated our research

Reviews — Related research that has been done to address the problem

Methods — What out hybrid methodology entails with description

Results — Experimental results

Conclusion — Research Conclusion, limitations and future works

Summary

- *Our hybrid methods outperformed all individual predictors on average throughout this research.*

- *Our results showed that individual method alone might not have an edge in empirical studies but could stand a chance through appropriate combination method.*

- *Machine Learning methods have shown not to be very strong predictors when used alone.*

- *We showed how effective ML techniques could be harnessed to improve the forecasting strength of a forecasting model.*

Summary

- *Our hybrid methods outperformed all individual predictors on average throughout this research.*

- *Our results showed that individual method alone might not have an edge in empirical studies but could stand a chance through appropriate combination method.*

- *Machine Learning methods have shown not to be very strong predictors when used alone.*

- *We showed how effective ML techniques could be harnessed to improve the forecasting strength of a forecasting model.*

Contribution

- *A simple approach of a hybrid system without weight estimations.*

- *Effectively combined statistical technique with machine learning methods to improve the forecasting accuracy*

- *Uniquely leverage the gains of Decomposition, transformation, statistical methods, feature extraction, and machine learning methods to improve the accuracy of point forecast.*

- *A paper "**Improving the Accuracy of Time series Forecasting using Hybrid Approach**" was accepted at International Symposium on Forecasting 2019*

Limitations

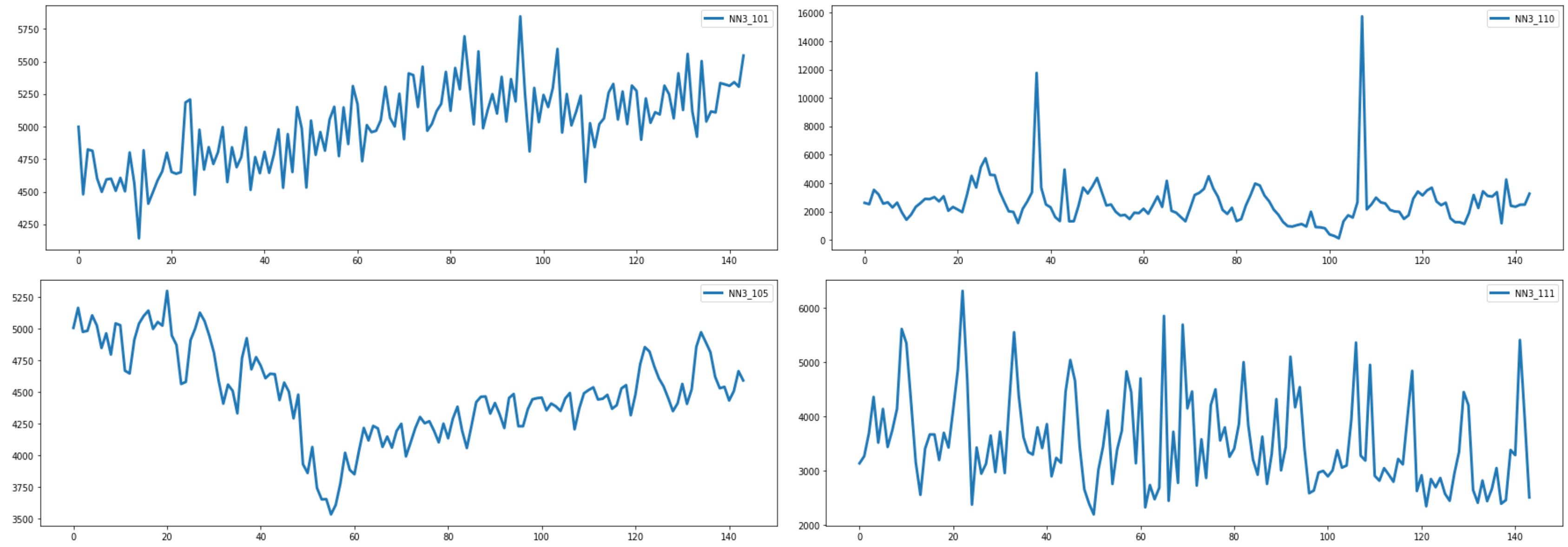
- *Short length of the observations in the Datasets.*
- *Only considered positive observations.*
- *Focus on only monthly available series.*

Future Works

- *Extend the methodology to other forms of series such daily, weekly, quarterly and yearly series.*
- *Dynamically selecting important features based on on series.*
- *Using auto adaptive parameters.*
- *Adding domain features to solve domain specific problems.*

End

Datasets



Sample Dataset from the NN3 forecasting competition

	Model	Pros	Cons	Reseach
Traditional or Classical Methods	Linear Regression	Ability to handle different time series components and features. High interpretability.	Sensitive to outliers Strong assumptions.	Mbamalu (Mbamalu 1993)
	Exponential Smoothing	Ability to handle variable level,trend and seasonality components. Automated Optimization.	Sensitive to outliers Narrow confidence intervals	Brown (Brown 1959) Holts (Holt 2004)
	ARIMA (Auto Regressive Integrated Moving Avearge)	High interpretability. Realistic confidence intervals. Unbiased forecasts.	Requires more data Strong restrictions and assuptions Hard to automate	ARIMA(Zhang 2003)
Stochastic / Machine Learning Methods	Machine Learning Model	High Interpretability More transparent than other models Deals well with uncertainty. Control the variance of the components.	Higher holdout errors Higher training and evaluation time.	X. Yan(X. Yan 2013) X. Yan (X. Yan 2014) Frohlich(Frohlich 2004)
	Neural Network Model	Less restrictions and assumptions Ability to handle complex nonlinear patterns High Predictive power Can be easily automated	Low interpretability. Difficult to derive confidence intervals for the forcast. Requires more data. Overfitting of Data. Shortage of plenty training time. Optimal network parameters determination.	P. G. Zhang (P. G. Zhang 2003) J. Faraway (J. Faraway 2008) C. Hamzaçebi(C. Hamzaçebi, 2008)
Fuzzy Based Methods	Fuzzy Based Models	Used for forecasting problems with more than one value of attributes Employes the concept of Fuzzy Logic.	Fuzzy rules are complex to set up	E. Egrioglu (E. Egrioglu 2013)