Towards time-series feature engineering in automated machine learning for multi-step-ahead forecasting

Can Wang[a], Mitra Baratchi[a,], ,Thomas Bäck[a], Holger Hoos[a], Steffen Limmer[b], Markus Olhofer[b]

(a) Leiden University
(b) Honda Research Institute Europe

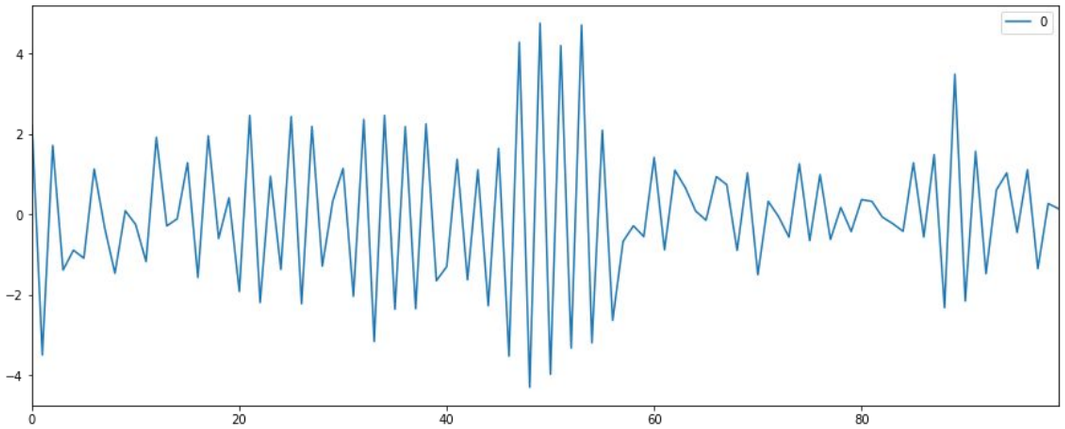
Universiteit Leiden
The Netherlands

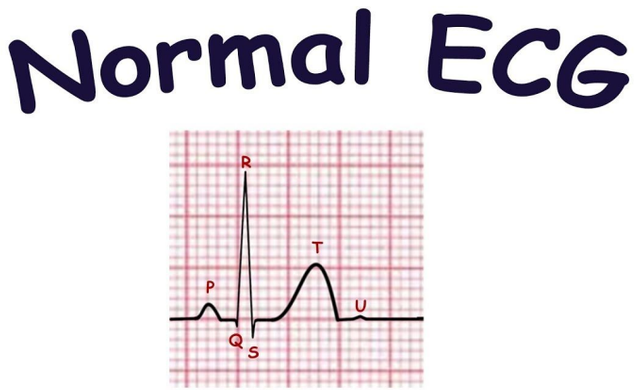
Honda Research Institute EU

Time series forecasting

Different applications of time series forecasting

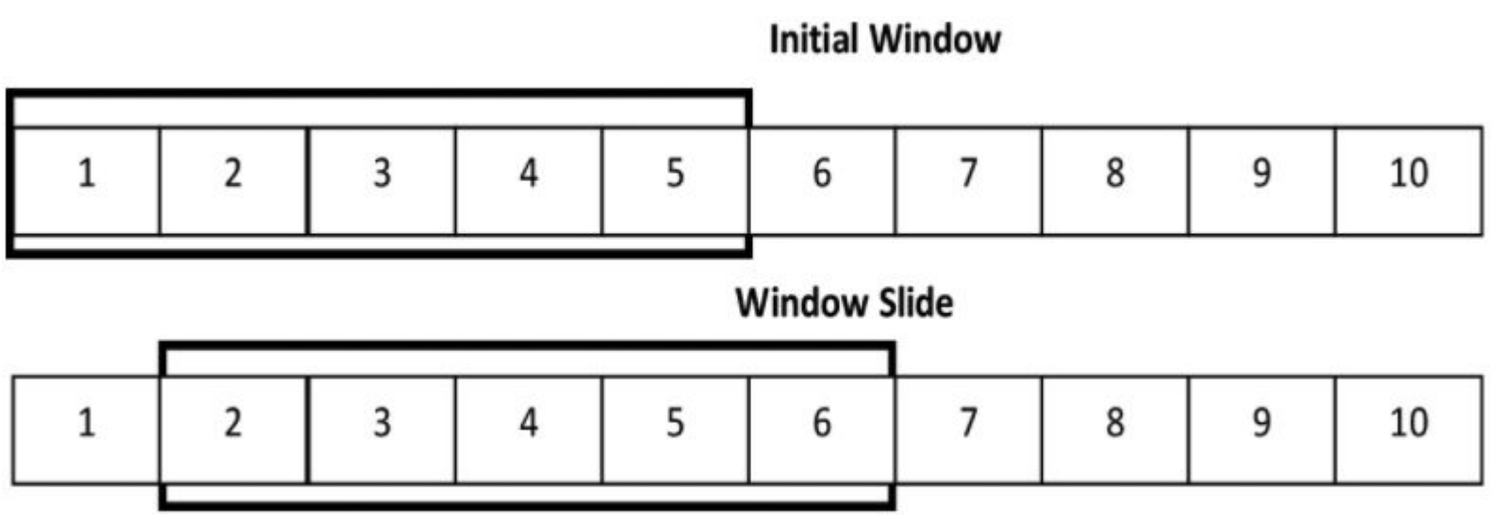
- Energy load forecasting
- Stock price forecasting
- Macroeconomics
- Load decomposition
- Human speech
- ……





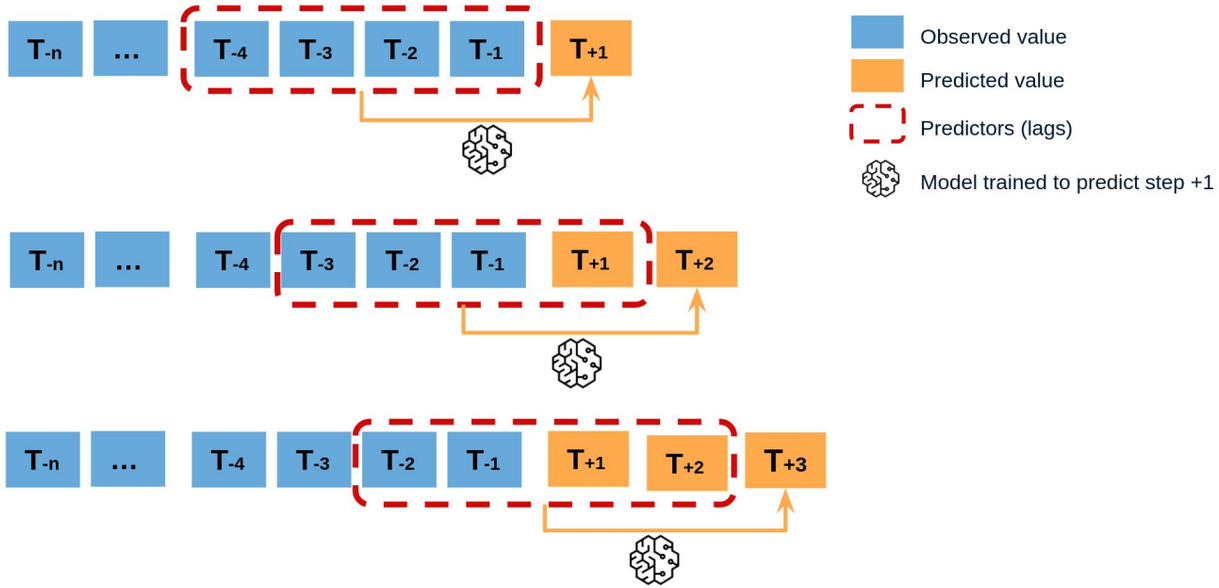
Benchmarking models and our models

- Simple baselines: Support vector machine, gradient boosting machine
- Deep net baselines: N-BEATS
- AutoML baselines: Auto-Keras, vanilla auto-sklearn
- Our newly proposed methods: auto-sklearn with automated window size selection, auto-sklearn with tsfresh features, auto-sklearn with automated window size selection and tsfresh features



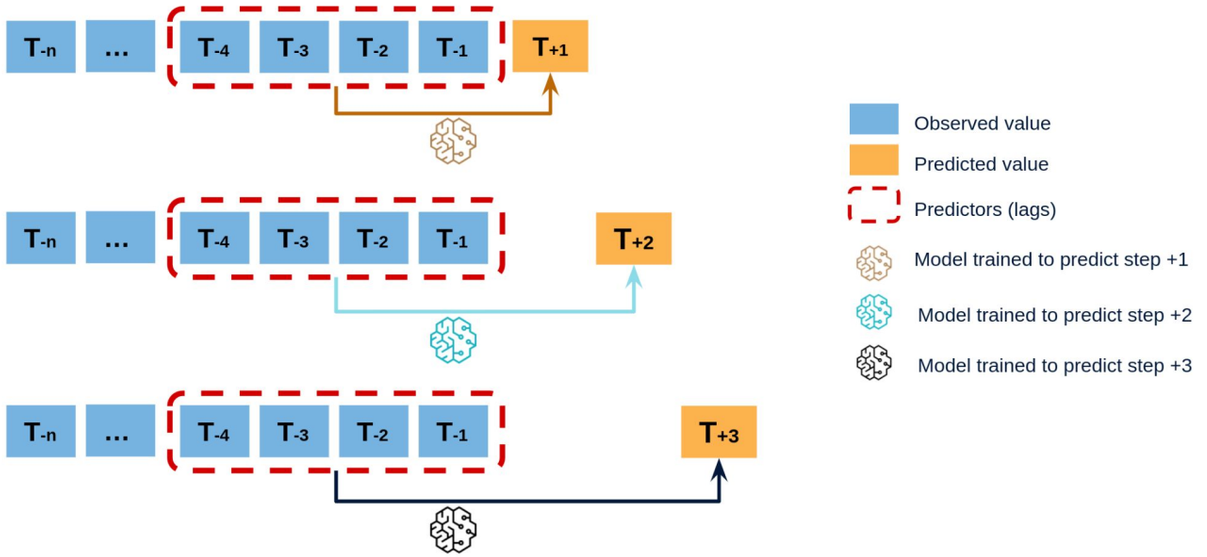
Recursive VS direct forecasting

- Recursive forecasting



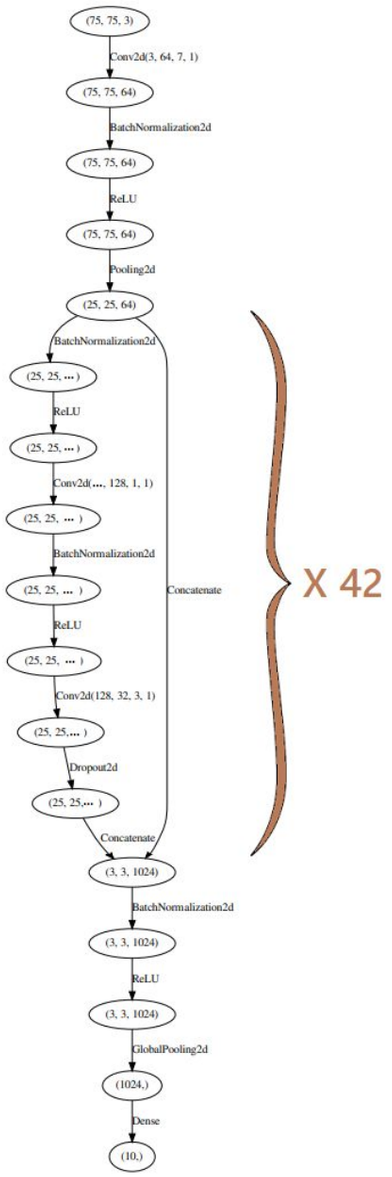
- Direct forecasting

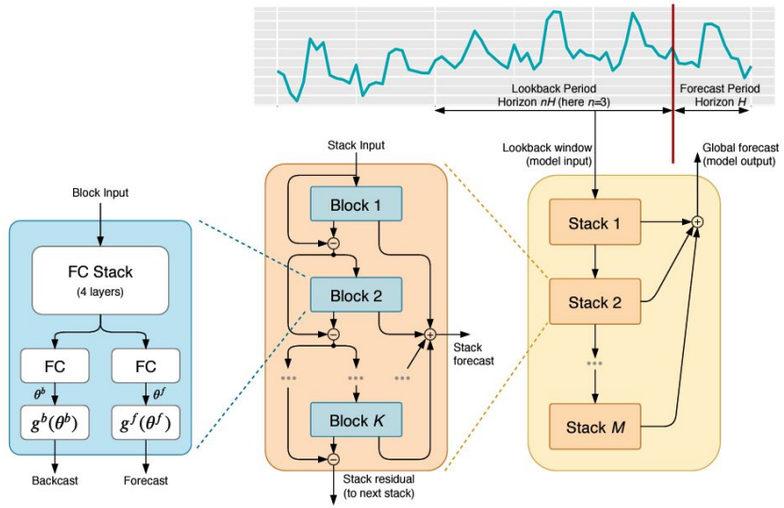
  Multiple models can be used to forecast the different time steps or use one multi-output model.

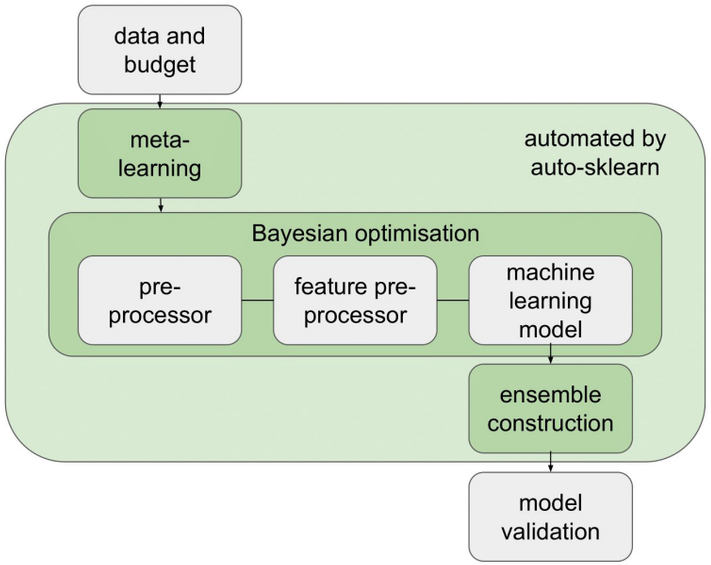


Auto-Keras, N-BEATS, and auto-sklearn

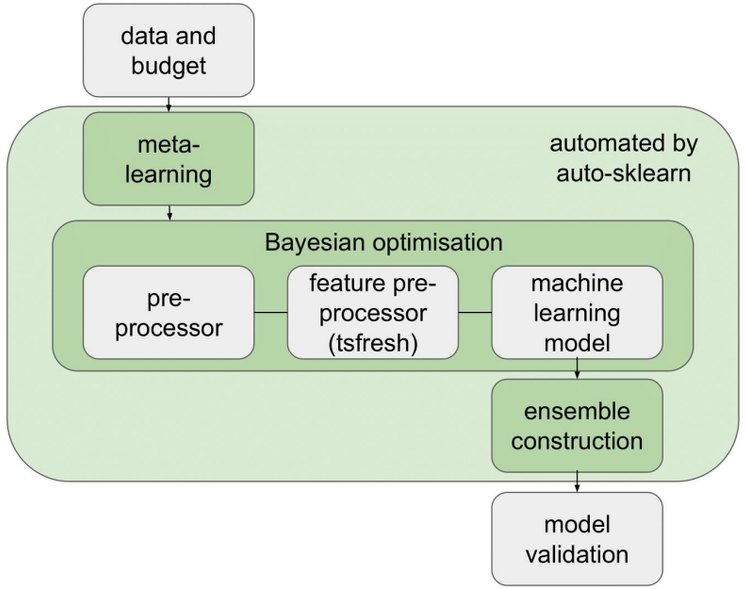
- Auto-Keras
- N-BEATS
- auto-sklearn

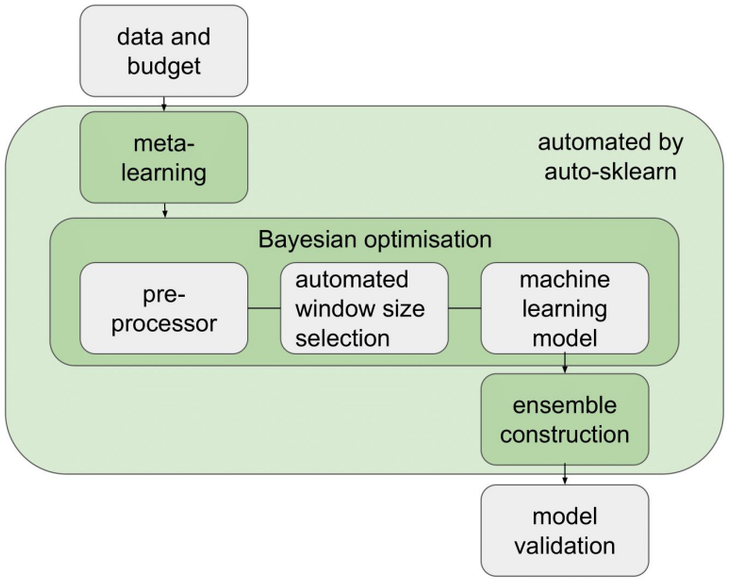


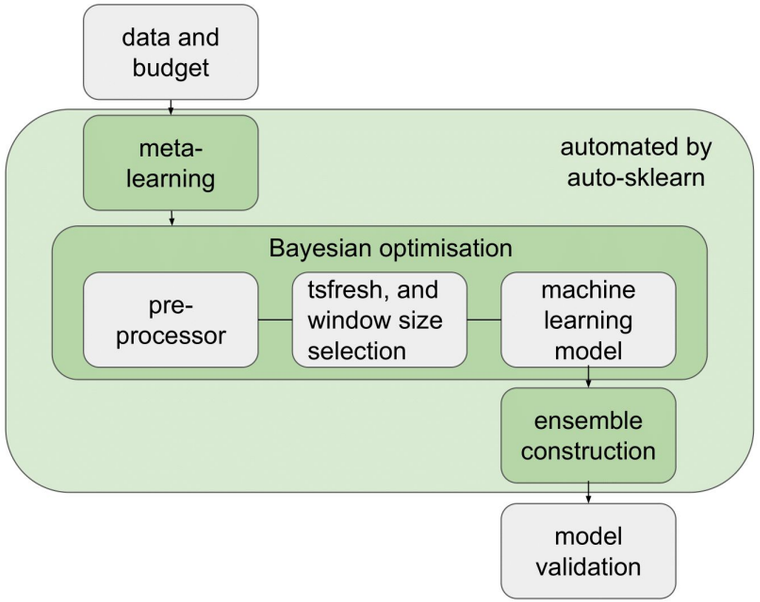




Our methods


data and budget
meta-learning
automated by auto-sklearn
Bayesian optimisation
pre-processor
feature pre-processor (tsfresh)
machine learning model
ensemble construction
model validation


data and budget
meta-learning
automated by auto-sklearn
Bayesian optimisation
pre-processor
automated window size selection
machine learning model
ensemble construction
model validation


data and budget
meta-learning
automated by auto-sklearn
Bayesian optimisation
pre-processor
tsfresh, and window size selection
machine learning model
ensemble construction
model validation

Datasets

10 Synthetic datasets:

Flow (Driven pendulum with dissipation, Duffing-van der Pol Oscillator, Driven van der Pol oscillator, Duffing two-well oscillator, Diffusionless Lorenz Attractor)

Stochastic Process (Autoregressive with noise, Correlated noise, Moving average process, Nonstationary autoregressive, Random walk)

10 Real world datasets:

Audio (animal sounds, human speech, music)

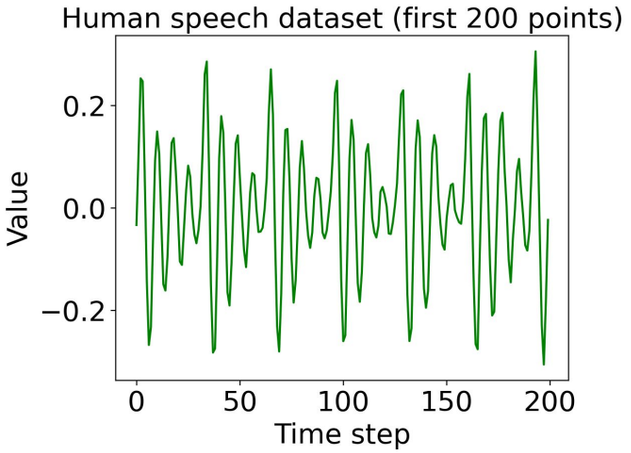
Ecology (Zooplankton growth)

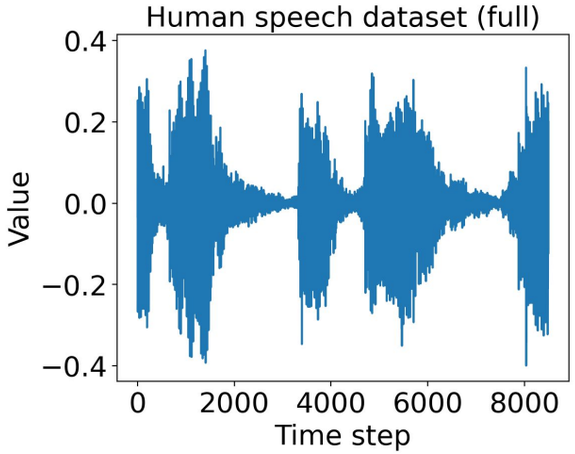
Economics (Macroeconomics, Microeconomics)

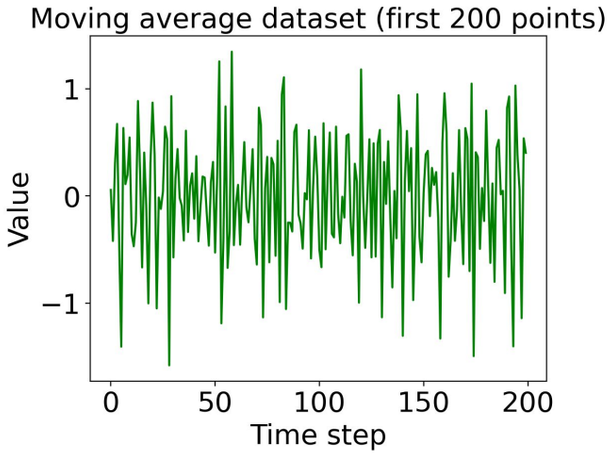
Finance (Crude oil prices, Exchange rate, Gas prices)

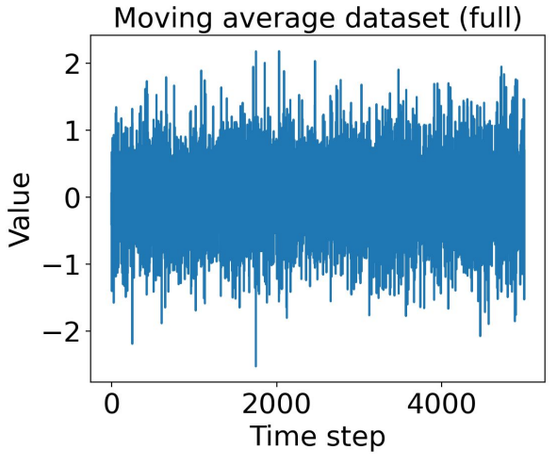
Medical (Electrocardiography ECG)

Datasets


Human speech dataset (first 200 points)
Value
0.2
0.0
−0.2
0
50
100
150
200
Time step


Human speech dataset (full)
Value
0.4
0.2
0.0
−0.2
−0.4
0
2000
4000
6000
8000
Time step


Moving average dataset (first 200 points)
Value
1
0
−1
0
50
100
150
200
Time step


Moving average dataset (full)
Value
2
1
0
−1
−2
0
2000
4000
Time step

Research questions

- How do recursive and multi-output techniques compare in terms of accuracy?
- To what extent can AutoML techniques (Auto-Keras, auto-sklearn, and our variants) beat the traditional baselines (GBM, SVM)?
- To what extent can AutoML techniques beat N-BEATS?

Results

Table 1: *RMSE* on test set acquired from traditional ML baselines. GBM-recursive, GBM-multiout, SVM-recursive, and SVM multioutput win on 6, 6, 0, and 8 out of 20 data sets respectively.

Dataset	RMSE(GBM -recursive)	RMSE(GBM -multioutput)	RMSE(SVM -recursive)	RMSE(S -multiou
Autoregre noise	**0.458890**	0.461558	0.484516	0.4594
Correlated noise	1.872176	**1.862137**	2.012916	2.0045
Lorenz Attractor	0.102323	**0.088045**	0.188223	0.1523
Pendulum	0.112041	0.104519	0.172118	**0.0353**
Driven oscillator	**0.121606**	0.124701	0.231661	0.2242
Two-well oscillator	0.033950	0.032462	0.075318	**0.0077**
Duffing oscillator	0.025830	0.021330	0.075308	**0.0137**
Moving average	0.629791	0.627176	0.641453	**0.6228**
Nonstationary	6.049796	**5.987631**	6.796246	6.4485
Random walk	**12.766561**	13.690753	30.594553	25.654
Crude oil prices	28.215008	32.909490	42.278003	**20.867**
ECG	**79.209558**	103.881034	128.420743	126.152
Exchange rate	0.006880	0.006823	0.028571	**0.0054**
Gas prices	102.819893	**100.612148**	166.021626	172.605
Human speech	0.059365	**0.054838**	0.085002	0.0576
Macroeconomics	779.515969	806.704035	713.073168	**713.363**
Microeconomics	**647.432403**	705.051879	3500.094238	3865.23
Music	0.082864	0.076047	0.068341	**0.0529**
Tropical sound	0.009468	**0.006285**	0.034925	0.0088
Zooplankton	**312.033380**	385.377067	319.839856	320.049

Table 2: *RMSE* on test set acquired from different AutoML methods including vanilla auto-sklearn (VA), our proposed variants (W, T, and WT), Auto-Keras and the state-of-the-art method N-BEATS. The accuracy of N-BEATS, Auto-Keras, VA, W, T, WT are statistically significant on 5, 0, 2, 8, 3, and 3 out of 20 data sets, respectively.

	RMSE (N-BEATS)	RMSE (Auto-Keras)	RMSE(VA)	RMSE(W)	RMSE(T)	RMSE(WT)
Autoregre noise	0.491133 ± 0.001470	0.468036	0.454026 ± 0.000502	**0.453002 ± 0.000137***	0.464252 ± 0.000286	0.463611 ± 0.000302
Correlated noise	1.949831 ± 0.012037	1.848905	1.832344 ± 0.003727	**1.822611 ± 0.0009631***	1.841418 ± 0.000878	1.843744 ± 0.000858
Lorenz Attractor	0.076379 ± 0.007143	1.248956	0.050628 ± 0.009272	**0.039705 ± 0.003994***	0.140621 ± 0.081695	0.249406 ± 0.012638
Pendulum	0.113806 ± 0.010232	0.512860	0.055285 ± 0.020322	**0.021416 ± 0.022933***	0.154815 ± 0.100197	0.177645 ± 0.048287
Driven oscillator	0.101795 ± 0.006627	0.232306	0.094895 ± 0.010350	0.085300 ± 0.005317	**0.061014 ± 0.001292***	0.149184 ± 0.003854
Two-well oscillator	**0.010395 ± 0.000897***	0.201339	0.032185 ± 0.006572	0.033103 ± 0.000724	0.019638 ± 0.001726	0.057722 ± 0.009568
Duffing oscillator	**0.004316 ± 0.001006***	0.500197	0.011301 ± 0.011301	0.010777 ± 0.000844	0.020875 ± 0.000919	0.019648 ± 0.001661
Moving average	0.662554 ± 0.002471	0.610291	0.615606 ± 0.000661	0.614169 ± 0.000080	0.609240 ± 0.000326	**0.608775 ± 0.000241***
Nonstationary	6.253612 ± 0.034291	9.888761	**5.775404 ± 0.048135***	5.783582 ± 0.012927*	6.720498 ± 0.023618	6.826646 ± 0.022398
Random walk	**3.099224 ± 0.058073***	19.385011	13.478998 ± 0.172438	13.240742 ± 0.138406	14.765951 ± 0.843100	15.411989 ± 0.368020
Crude oil prices	**32.199970 ± 0.846056***	44.305246	34.766214 ± 0.702608	34.669098 ± 0.374047	33.134671 ± 0.228864	35.373982 ± 0.209542
ECG	143.617378 ± 7.120182	149.577016	**96.577900 ± 3.258370***	102.0317220 ± 14.282081	104.228003 ± 1.6258357	103.260689 ± 2.110540
Exchange rate	0.004831 ± 0.001875	0.013859	0.006049 ± 0.000130	**0.003627 ± 0.001326***	0.006886 ± 0.000237	0.004569 ± 0.001505
Gas prices	**59.147486 ± 3.122861***	126.112217	118.033606 ± 25.754493	101.514037 ± 0.021932	109.711360 ± 4.170307	121.133289 ± 1.387935
Human speech	0.089423 ± 0.001085	0.065285	0.061140 ± 0.001988	0.061280 ± 0.002512	0.058722 ± 0.000191	**0.057920 ± 0.000288***
Macroeconomics	755.905175 ± 11.131619	711.622796	791.586127 ± 37.469871	766.036513 ± 31.082581	682.959633 ± 3.559769	**662.098158 ± 24.511173***
Microeconomics	2179.375169 ± 6643.906258	5545.979242	805.637017 ± 61.049526	**733.760632 ± 12.795080***	1295.531031 ± 34.305972	1348.425680 ± 264.961810
Music	0.078767 ± 0.001062	0.091060	0.085224 ± 0.005985	0.072422 ± 0.002119	**0.068895 ± 0.000906***	0.084715 ± 0.002951
Tropical sound	0.010038 ± 0.000203	0.010660	0.008483 ± 0.000072	**0.008407 ± 0.000013***	0.008930 ± 0.000029	0.008717 ± 0.000044
Zooplankton	306.329197 ± 2.591637	310.290963	297.091504 ± 2.118263	282.485896 ± 1.234165	**278.597102 ± 1.237304***	301.160979 ± 4.331423

Recursive VS multi-output forecasting

How do recursive and multi-output techniques compare in terms of accuracy?

we generally observe that GBM-multioutput performs better than GBM-recursive on 12 out of 20 data sets, while SVM-multioutput outperforms SVM-recursive on 17 out of 20 data sets in terms of RMSE.

As we have observed that multi-output models tend to perform better.

AutoML techniques VS traditional baselines

To what extent can AutoML techniques (Auto-Keras, auto-sklearn, and our variants) beat the traditional baselines (GBM, SVM)?

Auto-Keras beats all the traditional ML baseline models (GBM-recursive, GBM-multioutput, SVM-recursive, and SVM-multioutput) on 4 out of 20 data sets.

Vanilla auto-sklearn outperforms all the traditional ML baselines on 8 out of 20 data sets.

Our three variants show lower error than all the traditional ML baselines on 10, 5, and 5 out of 20 data sets, respectively.

The best AutoML (W) outperfoms the best traditional ML baseline (SVM-multioutput) on 14 out of 20 data sets.

AutoML VS N-BEATS

To what extent can AutoML techniques beat N-BEATS?

The best AutoML (W) outperforms N-BEATS on 14 out of 20 data sets.

Auto-Keras, VA, T, and WT beat N-BEATS on 5, 12, 11, and 10 out of 20 data sets, respectively.

Conclusions

AutoML can achieve significantly higher accuracy than the traditional ML baselines on 14 out of 20 data sets in terms of RMSE.

AutoML TS variants beat N-BEATS on 14 out of 20 data sets.

The multi-output technique tends to perform better with the same budget than the recursive technique in the multi-step TS forecasting tasks.

Future Work

Other tasks involving time-series data, such as time-series classification.

Can Wang
Email: c.wang@liacs.leidenuniv.nl