Meta-learning and Data Augmentation for Stress Testing Forecasting Models

Authors
Ricardo Inácio[(1)(2)]
Vítor Cerqueira[(1)(2)]
Marília Barandas[(3)]
Carlos Soares[(1)(2)(3)]

Affiliations
[(1)] Faculdade de Engenharia da Universidade do Porto, Portugal
[(2)] Laboratory for Artificial Intelligence and Computer Science (LIACC), Portugal
[(3)] Fraunhofer Portugal AICOS, Portugal

Problem Definition

- Univariate time series forecasting.
- Using machine learning approaches based on an autoregressive methodology.
  - Time series are transformed using sliding windows and time delay embedding.
- Meta-learning and data augmentation are applied to enrich the metadataset and improve the detection of cases of large errors.

Objectives

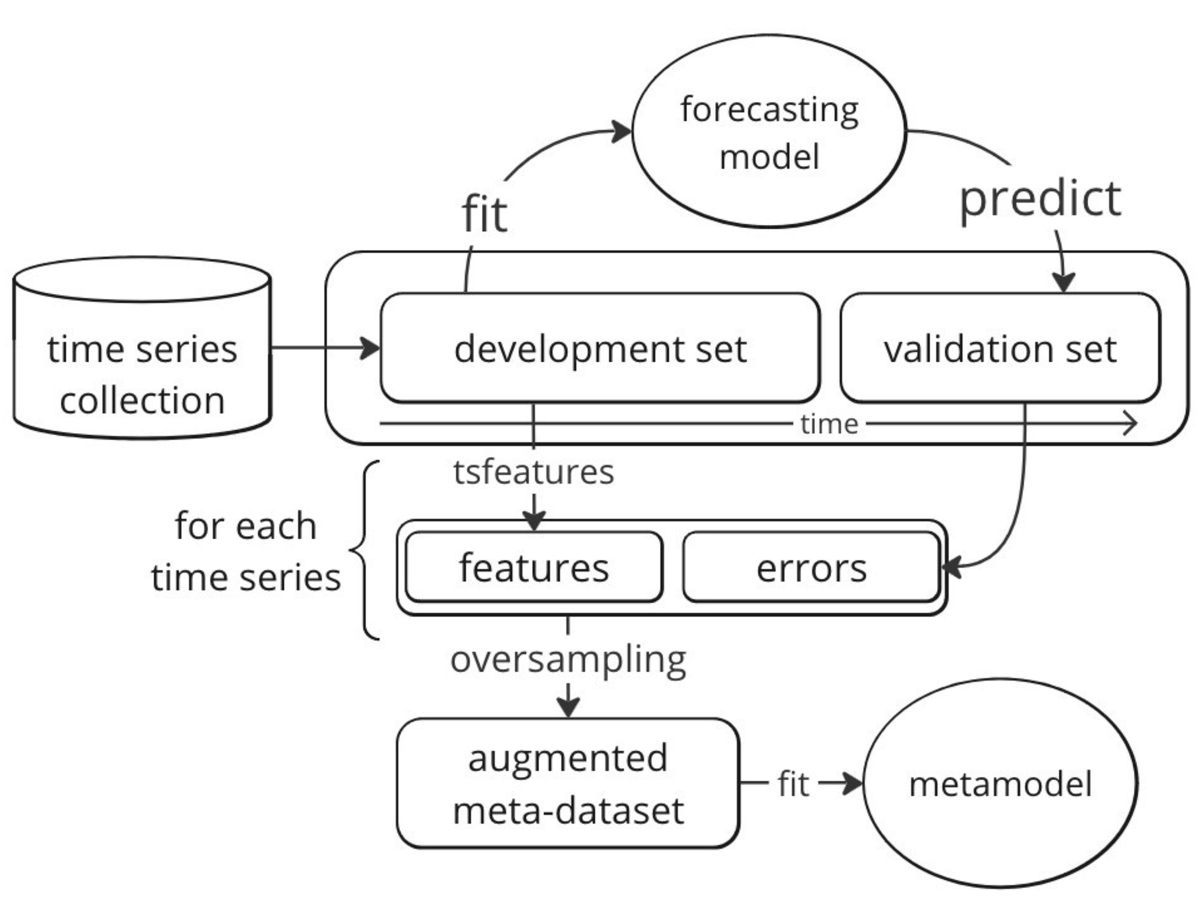
- Identify the conditions that lead to a large forecasting error.
- Model large errors based on time series structural properties using meta-learning
  - Using feature extraction (e.g. seasonal strength)
- Leverage data augmentation methods, such as oversampling, to enrich the metadataset
- Use XAI methods to understand model limitations

What is Stress Testing?

- A method that aims to identify the conditions in which a model is under stress and is prone to under-perform.
- It evaluates machine learning methods by identifying, modelling and simulating challenging scenarios.

Methodology

The workflow is split into two stages:
1. **Development Phase:** performance estimation, feature extraction, data augmentation, and meta-learning are conducted.
2. **Inference Phase:** the metamodel is applied to predict large errors.



Experimental Setup

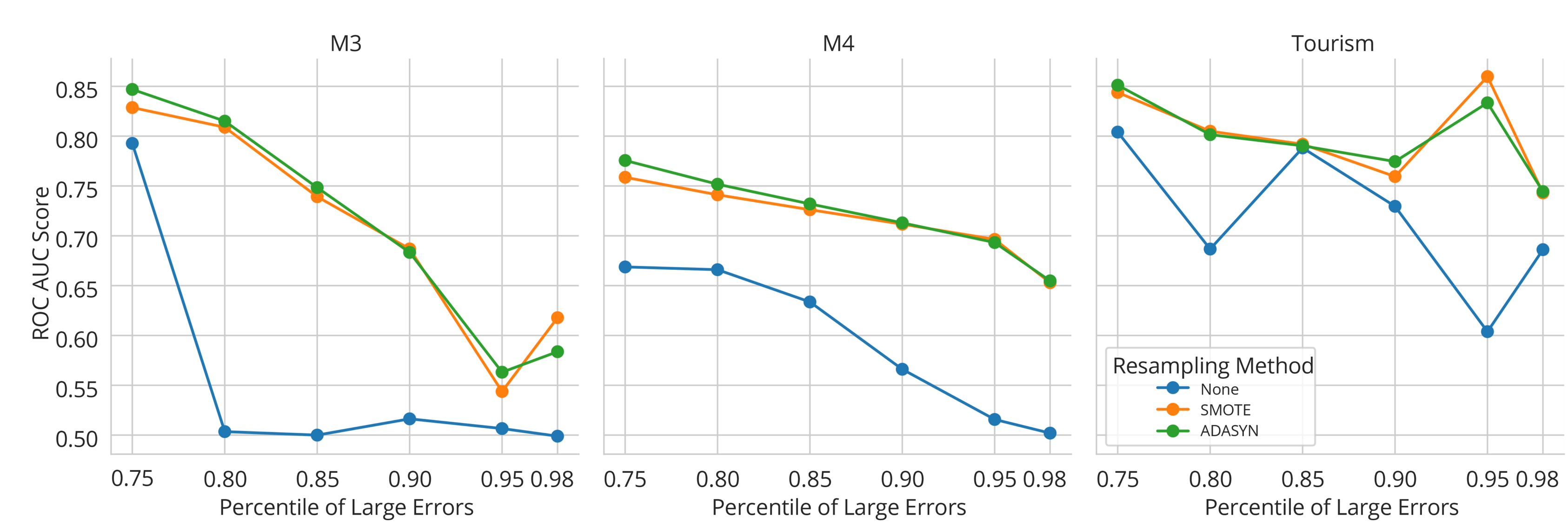
Forecasting Model:
- Built using a *lightgbm* (LGBM) regression algorithm.
- Hyperparameter tuning is automatically handled by the selected framework (*mlforecast*).
- A Seasonal Naive method is included as a benchmark, which was surpassed in every experiment, as shown below, by comparing the average *SMAPE* score in each dataset.

	Seasonal Naive	LGBM
M3 Monthly	7.98%	**7.28%**
M4 Monthly	7.17%	**6.58%**
Tourism Monthly	9.15%	**8.50%**

Metamodel:
- Also trained using a *lightgbm* algorithm.
- Implemented using an objective function tailored for binary classification.
- Hyperparameters are optimized via *grid search.*

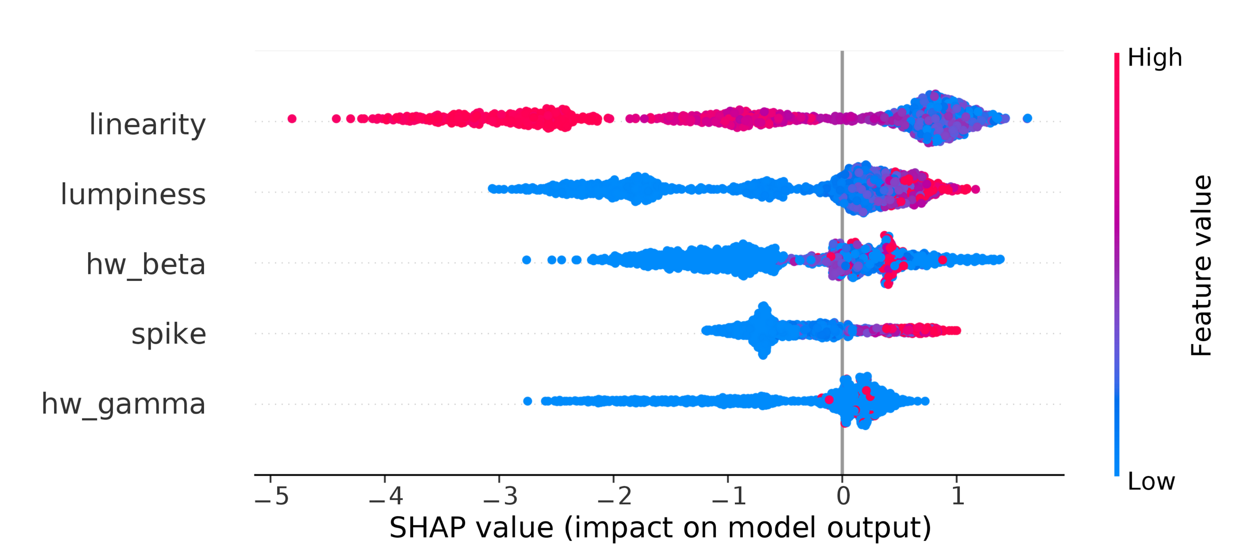
Sensitivity Analysis

Analysis of the impact of the error percentile threshold in metamodel performance, measured in AUC. For each time series collection, the performance for the base and augmented metamodels is compared regarding six levels of percentile error threshold.
- Overall, as the error percentile threshold increases, the metamodel perfomance decreases.
- Both oversampling techniques make the metamodels more robust to stress.



Metamodel Explanations

Analysis of the performance of the forecasting model based on the predictions of the metamodel:
- High *linearity* values reduces the probability of classifying a time series as large error inducing.
  - This suggests that the forecasting model may struggle in time series with a low *linearity* value.
- The opposite happens regarding *lumpiness*.



Metamodel Perfomance (AUC Scores)

- The metamodel performs the best when coupled with data augmentation.
  - Mainly with *ADASYN (in bold).*
- It struggles in two datasets (M3 & M4), when no augmentation is applied.

	ADASYN	SMOTE	No Aumgentation
M3 Monthly	**0.728**	0.678	0.516
M4 Monthly	**0.713**	0.709	0.566
Tourism Monthly	**0.787**	0.761	0.730

Acknowledgments

This work was partially funded by projects AISym4Med (101095387) supported by Horizon Europe Cluster 1: Health, ConnectedHealth (n.º 46858), supported by Competitiveness and Internationalisation Operational Programme (POCI) and Lisbon Regional Operational Programme (LISBOA 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund(ERDF) and Agenda "Center for Responsible AI", nr. C645008882-00000055, investment project nr. 62, financed by the Recovery and Resilience Plan (PRR) and by European Union - NextGeneration EU, and also by FCT plurianual funding for 2020-2023 of LIACC (UIDB/00027/2020 UIDP/00027/2020)

Related Literature

1. Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., Gamboa, H.: Tsfel: Time series feature extraction library. SoftwareX 11, 100456 (2020)
2. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research 16, 321–357 (2002)
3. He, H., Bai, Y., Garcia, E.A., Li, S.: Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE international joint conference on neural networks. pp. 1322-1328
4. Moniz, N., Cerqueira, V.: Automated imbalanced classification via meta-learning. Expert Systems with Applications 178, 115011 (2021)
5. Montero-Manso, P., Athanasopoulos, G., Hyndman, R.J., Talagala, T.S.: Fforma: Feature based forecast model averaging. International Journal of Forecasting 36(1), 86–92 (2020)