

# TSFX: A Python package for time series feature extraction

Wilhelm Söderkvist Vermelin <sup>1,2\*</sup>

<sup>1</sup> RISE Research Institutes of Sweden <sup>2</sup> Mälardalen University \* These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

## Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

TSFX is a Python ([Van Rossum & Drake, 2009](#)) library for extracting features from time series data. It is inspired by the tsfresh ([Christ et al., 2018](#)) Python package with a special focus on performance. In order to achieve good performance, it utilizes Polars ([Vink et al., 2023](#)) which is a high performance DataFrame library written in Rust ([The Rust Programming Language, n.d.](#)) with Python bindings created through PyO3 ([Project & Contributors, n.d.](#)). The feature extraction functions are also implemented in Rust for performance. Compared to tsfresh, TSFX offers a conservative estimate of 10x performance, using the same set of time series features.

TSFX can be installed using pip:

```
pip install tsfx
```

TSFX can also be configured using a TOML ([TOML, n.d.](#)) configuration file (default name `.tsfx-config.toml`).

## Statement of need

Time series is a ubiquitous data modality, used in many domains such as finance, industry, meteorology, and medicine, to mention a few. As hardware to collect and store time series data is becoming increasingly affordable, the amount of available time series data is increasing in many domains. A common preprocessing step when dealing with time series is feature extraction where useful features, such as mean, variance, skewness, etc. are extracted from time series to be used in downstream tasks such as classification, regression or clustering. For large time series datasets, performance is increasingly important to enable timely data preprocessing. TSFX is made for this purpose: extracting features from large time series datasets.

## Acknowledgements

The TSFX package was developed within the [Vinnova](#) projects [DFusion](#), [TolkAI](#), and [SIFT](#).

## References

- Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series Feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). *Neurocomputing*, 307, 72–77. <https://doi.org/https://doi.org/10.1016/j.neucom.2018.03.067>
- Project, P., & Contributors. (n.d.). PyO3. GitHub. <https://github.com/PyO3>
- The rust programming language*. (n.d.). <https://rust-lang.org/>.

- 34 *TOML: Tom's obvious minimal language*. (n.d.). <https://toml.io/en/>
- 35 Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.  
36 ISBN: 1441412697
- 37 Vink, R., Gooijer, S. de, Beedie, A., Zundert, J. van, Hulselmans, G., Grinstead, C., Gorelli, M.  
38 E., Santamaria, M., Heres, D., ibENPC, Leitao, J., Heerden, M. van, Jermain, C., Russell,  
39 R., Pryer, C., Castellanos, A. G., Goh, J., Wilksch, M., illumination-k, ... Keller, J. (2023).  
40 *Pola-rs/polars: Python polars 0.16.11* (py-0.16.11). Zenodo. [https://doi.org/10.5281/](https://doi.org/10.5281/zenodo.7699984)  
41 [zenodo.7699984](https://doi.org/10.5281/zenodo.7699984)

DRAFT