Time Series Modeling with PyCaret and sktime

Nikhil Gupta

PyCaret Time Series Module – Lead Developer

Agenda

- Introduction
- Why PyCaret?
- Time Series Module
- Demo
- Q&A
- Additional Resources

About me

- **Background:** Electrical Engineer + Data Science
- **Industry:** Semiconductor
- **Work:** North America, Asia
- **Open-Source:** PyCaret, Auto_TS, sktime

Connect with me: Medium

Challenges of Time Series Modeling Lifecycle

- Time Series Modeling is an iterative process. It is very time consuming.
- Creating a seamless pipeline is hard. Managing it in production is harder.
- Focus on end-goal and solving business problems can take a backseat within small teams with increasing technical debt.
- Scalability is not just desirable, but it is very much needed.

What is PyCaret?

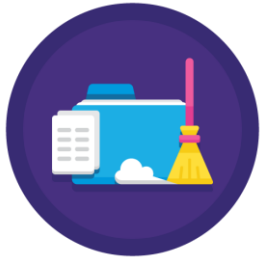
PyCaret is an open source, low-code machine learning library and end-to-end model management used to automate machine learning workflows. It is commonly used for rapid prototyping and deployment of ML pipelines.

EASY TO USE

PRODUCTIVITY TOOL

BUSINESS READY

PyCaret Features



Data Preparation

Model Training

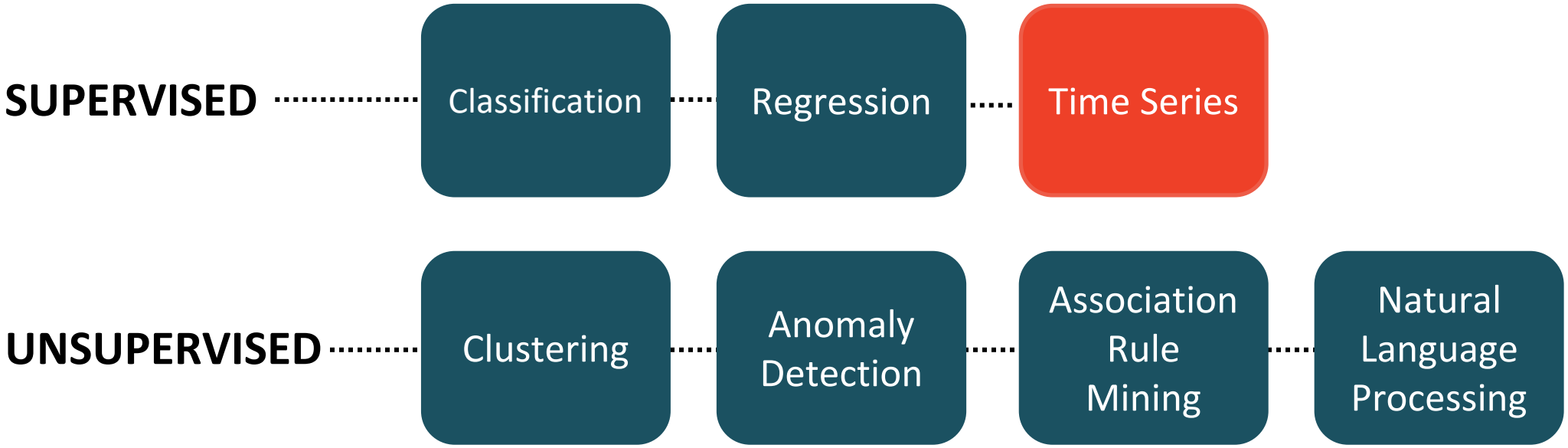
Hyperparameter Tuning

Analysis & Interpretability

Model Selection

Experiment Logging

Machine Learning use-case supported:



Time Series Module | Alpha Release

- **Data Preparation:** Train/Test Split, Cross-validation setup, Time series index handling
- **Model Training:** ~ 30 univariate models available (baseline, traditional statistical, reduced regression machine learning)
- **Hyperparameter Tuning:** Random Grid Search and Fixed Grid Search
- **Model Selection:** Available through `compare_models()`
- **Analysis and Interpretability:** Not Available in Alpha release
- **Experiment Logging:** Available through MLFlow

Time Series Module | Roadmap

- **Data Preparation:** Preprocessing (Imputation, Transformations), Time Series Feature Extraction
- **Model Training:** Multivariate Models + Additional models
- **Hyperparameter Tuning:** Refine Grid Search, add other tuning methods (Bayesian)
- **Analysis and Interpretability:** Statistical Tests, Exploratory Analysis, Model Evaluation Capabilities, Back Testing
- **Model Selection:** No additional features planned
- **Experiment Logging:** Support for additional logging libraries (Weights & Biases)

Demo

- ❑ Demo 1 – Time Series Forecasting with PyCaret and sktime | **Simple Flow**
- ❑ Demo 2 – PyCaret Time Series | **Customizing** the **Flow** and Looking **Under the Hood**

All Notebooks are available here:

https://github.com/pycaret/sktime-dev-days-2021

pip install pycaret-ts-alpha

(after 06/27/2021)

We are always looking for contributors!

Open an Issue or PR @

https://www.github.com/pycaret/pycaret

Acknowledgements

- PyCaret Time Series Module Core Development Team
  - Miguel Trejo Marrufo
  - Antoni Baum
  - Moez Ali
  - Krishnan S G

- Sktime Development Team
  - Markus Löning, Franz Király, Martin Walter

Time Series Resources

Time Series 101 — For beginners

A beginner-friendly introduction to Time Series Forecasting

Photo by Chris Liverani on Unsplash

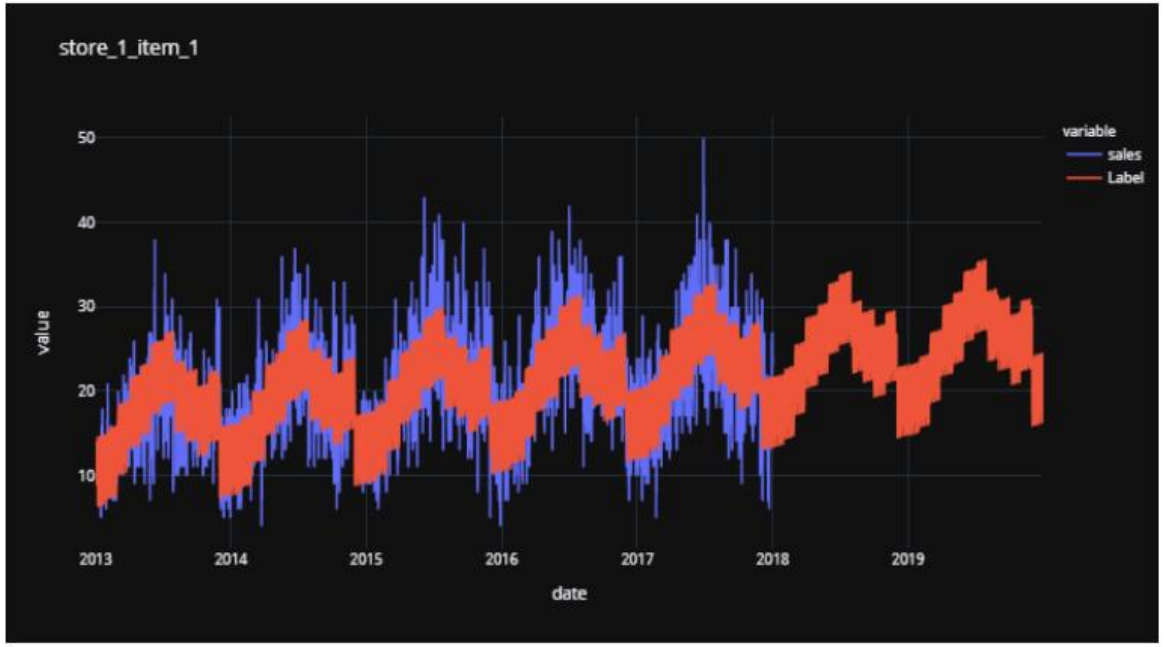
Time Series Forecasting with PyCaret Regression Module

Photo by Lukas Blazek on Unsplash

Additional Time Series Resources

Multiple Time Series Forecasting with PyCaret

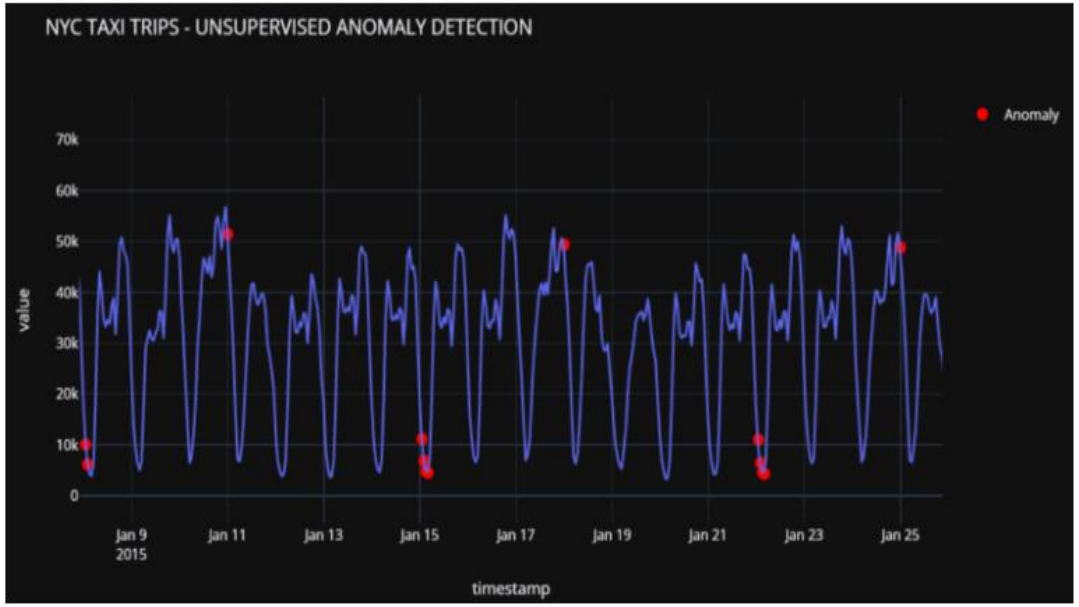
A step-by-step tutorial to forecast multiple time series with PyCaret



PyCaret — An open-source, low-code machine learning library in Python

Time Series Anomaly Detection with PyCaret

A step-by-step tutorial on unsupervised anomaly detection for time series data using PyCaret



PyCaret — An open-source, low-code machine learning library in Python

Important Links

- Official: https://www.pycaret.org
- GitHub: https://www.github.com/pycaret/pycaret
- LinkedIn: https://www.linkedin.com/company/pycaret
- YouTube: https://www.youtube.com/channel/UCxA1YTYJ9BEeo50lxyI_B3g
- Medium: https://medium.com/@moez_62905/
- Slack: https://join.slack.com/t/pycaret/shared_invite/zt-row9phbm-BoJdEVPYnGf7_NxNBP307w (expires 7/14)