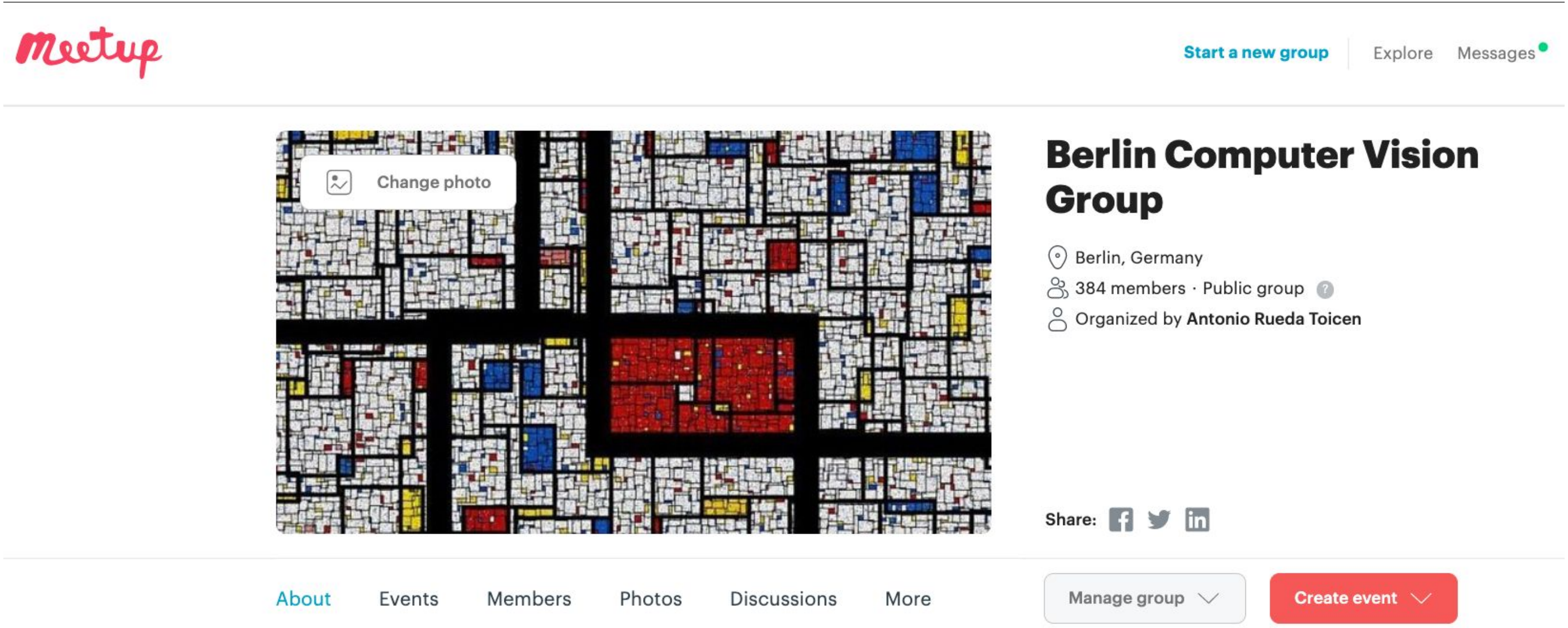
Time Series for Data Scientists

Antonio Rueda-Toicen
Data Science Retreat
September 2023

About me

- Machine Learning Engineer
  - Background in academia (computer science and bioengineering)
  - Organizer of the Berlin Computer Vision Group

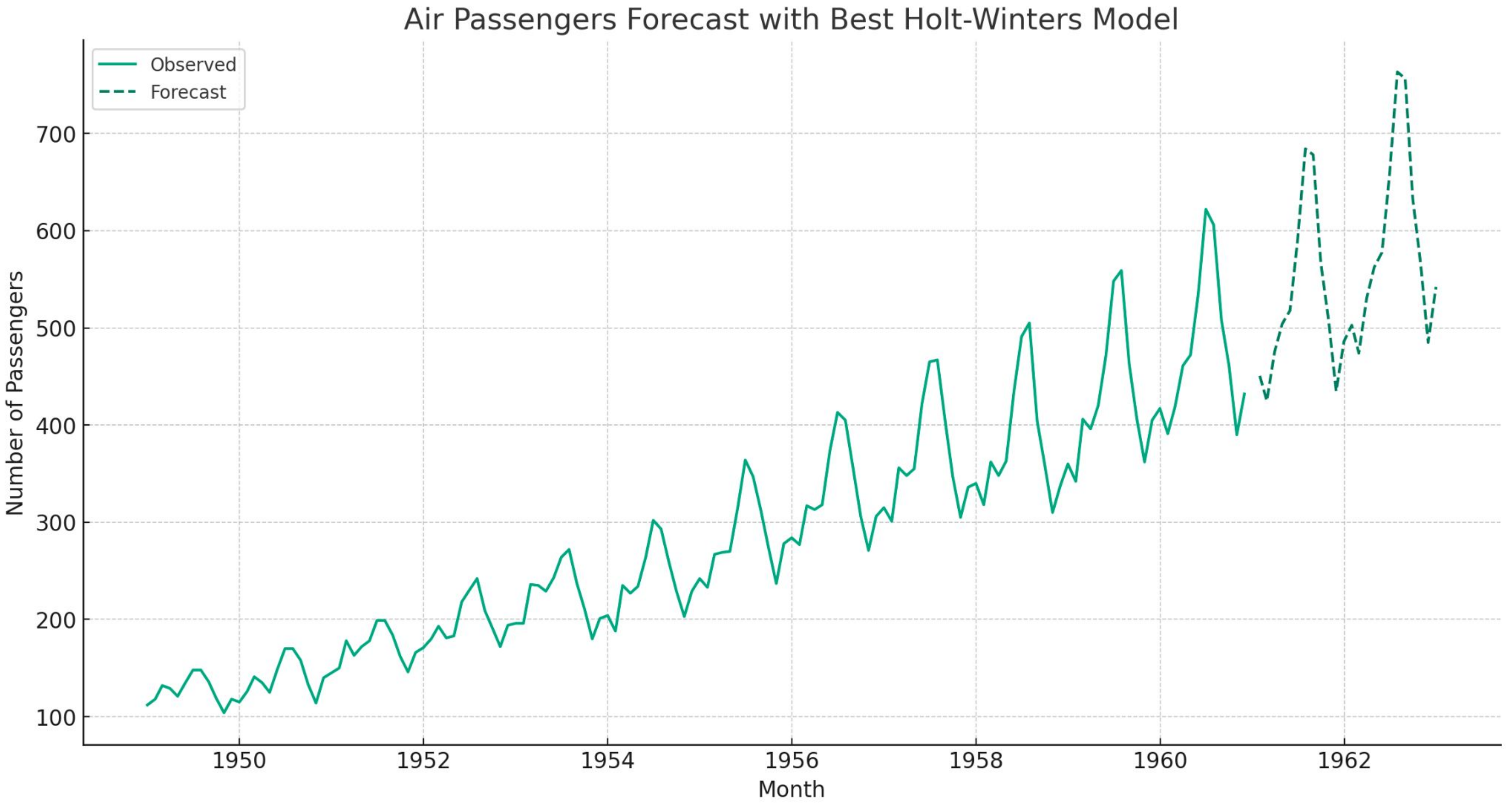


https://www.meetup.com/Berlin-Computer-Vision-Group/

Agenda

- Tasks in Time Series Analysis
- Why working with time series is different?
- How should we approach time series problems
- Properties of time series data: trend, seasonality and noise
- Visualizing time series data
- Autocorrelation and Partial Autocorrelation
- Using PyCaret and Darts
- Baseline techniques
  - Next day prediction, Moving Averages, Exponential moving averages
- Classical techniques
  - ARIMA, Holt-Winters (Exponential smoothing)
- LightGBM for time series analysis
- Facebook’s Prophet
- N-Beats
- Predicting the Standard and Poor’s 500 Index

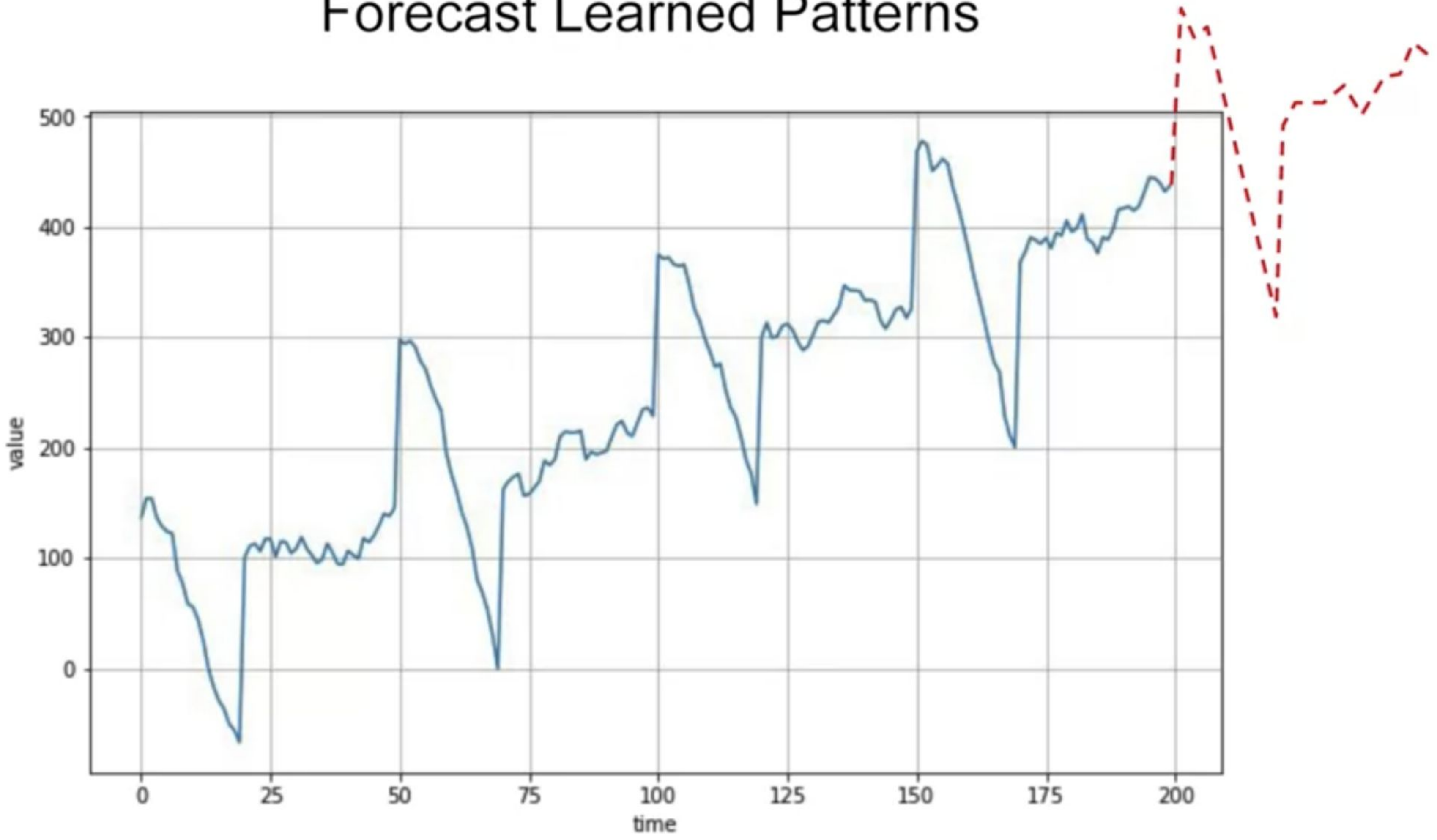
What is a time series?

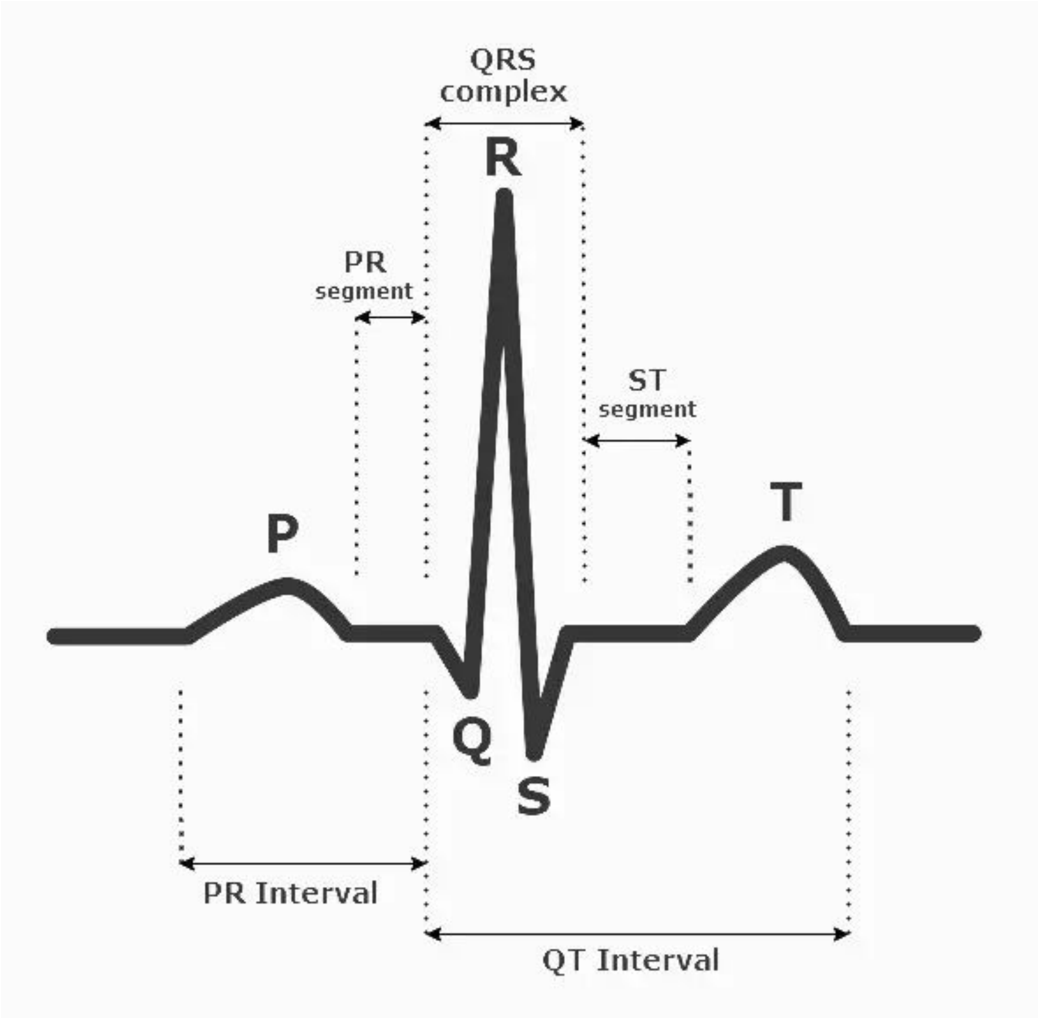


Tasks in Time Series Analysis

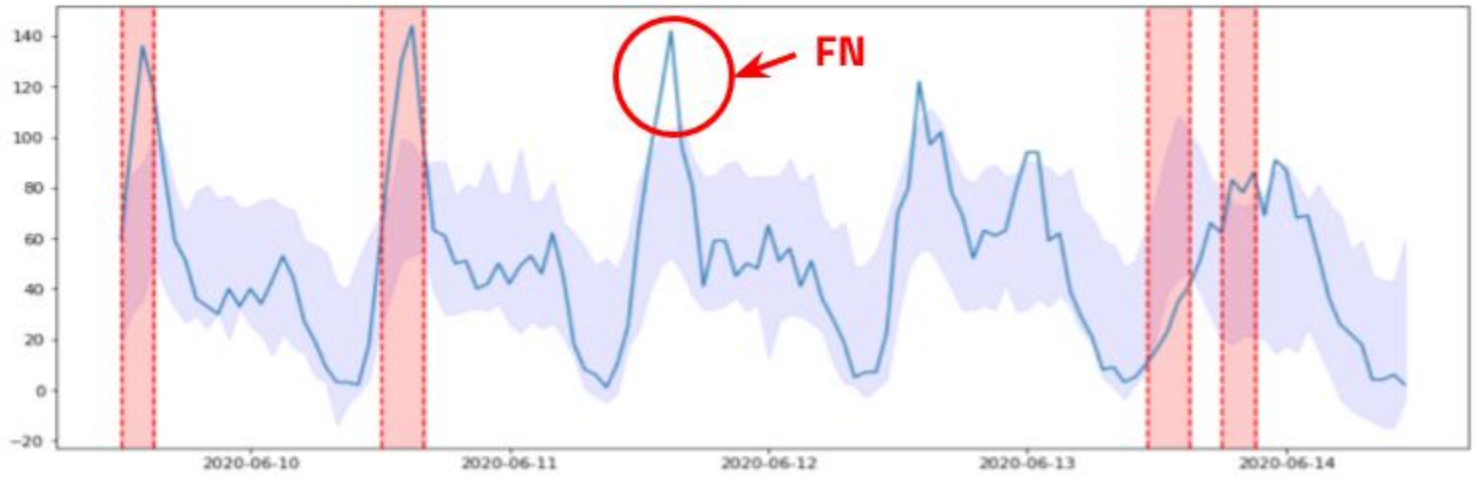
- Forecasting (‘predicting the future’ - our focus today)
- Anomaly Detection (identifying weird events)
- Classification (detecting patterns, e.g. cardiac arrest, speech recognition)

Forecast Learned Patterns





Prompt: [“Is detecting arrhythmia on an ECG a forecasting, classification, or anomaly detection problem?”]



Anomaly Detection in Time Series Analysis

Knowing the jargon

- Trend
- Seasonality
- Residual aka Noise
- Stationarity
- Autoregressive
- Autocorrelation and Partial Autocorrelation
- Differencing
- Backtesting
- Exogenous variable
- Look-ahead problem
- Multivariate vs univariate
- Recursive forecasting
- Exponential moving average
- Exponential smoothing
- LSTM
- ARIMA

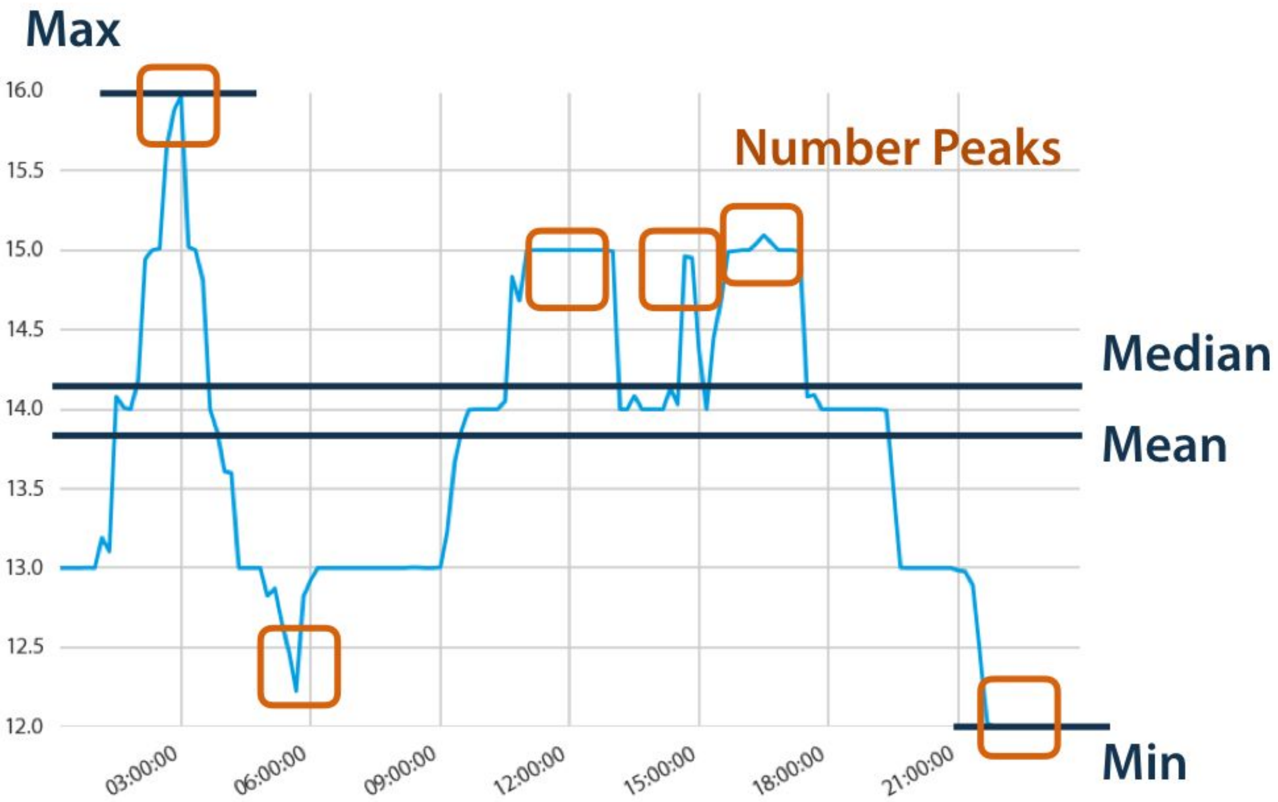
Why working with time series is different?

- We usually have as target the same feature that we use for training
- Missing values require extra care
- This creates problems with (usually unintentional) data leakage between training and testing
    - Lookahead problem - often appears in imputation and smoothing
- The “simple and traditional” methods that everybody knows from school (ARIMA and Markov models) depend on assumptions that most real data doesn’t have
    - **We torture the data into having these features through differencing**
- <u>Working with time series is one of most challenging tasks in data science</u>

Why working with time series is so difficult?

- Missing data is very common
- Low interpretability of methods
- Difficult (for humans) to spot and describe patterns in squiggly lines
- The most popular method is ARIMA which requires **a lot** of expertise and fine tuning
    - This makes ARIMA models **quite difficult to deploy in an automatic way**
- Exogenous variables
    - How can we predict events like the COVID outbreak? (Black swans)
    - How many unmeasured events contribute to our output?
- Uncomputability problem (the halting problem) [paper]
- Communicating to and keeping trust from stakeholders and teammates who often have conflicting or vested interests **is the biggest challenge**
- **It's easy to give bad predictions and it's easy to lose trust**

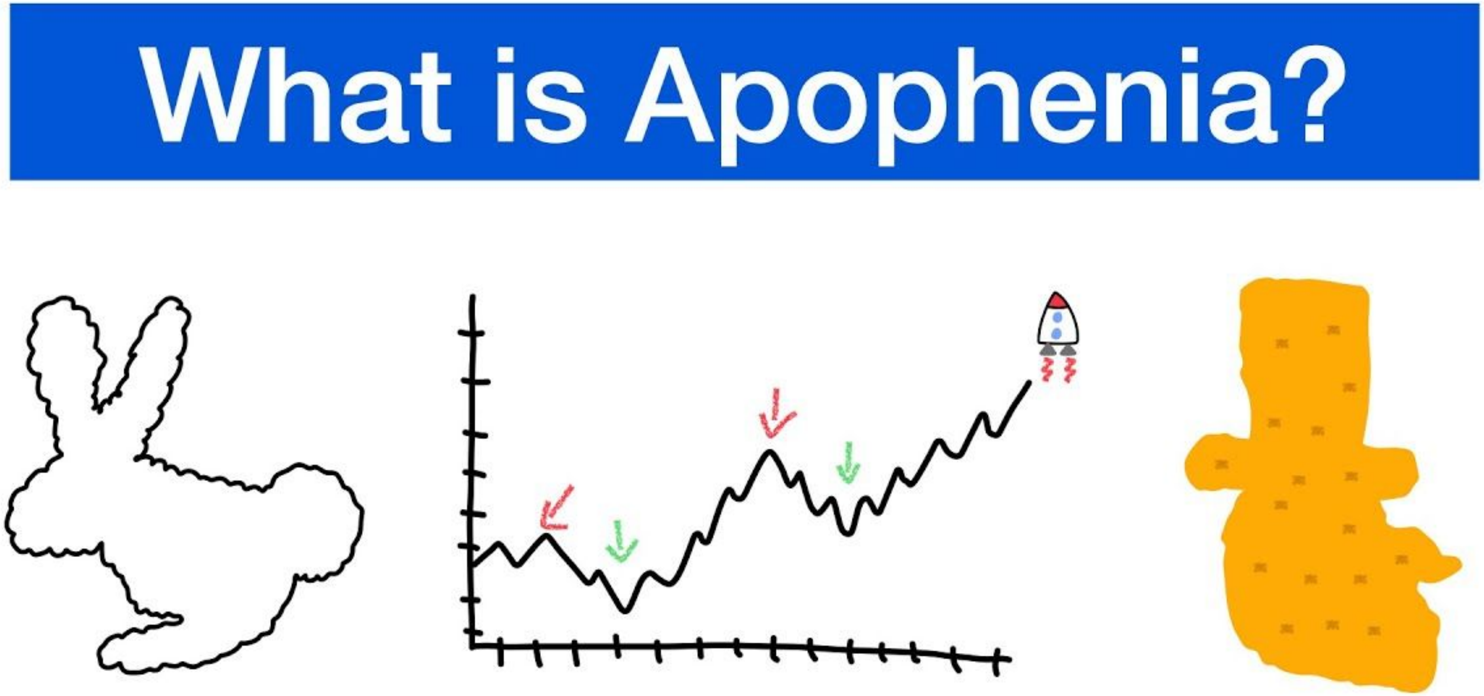
Time Series Feature Extraction



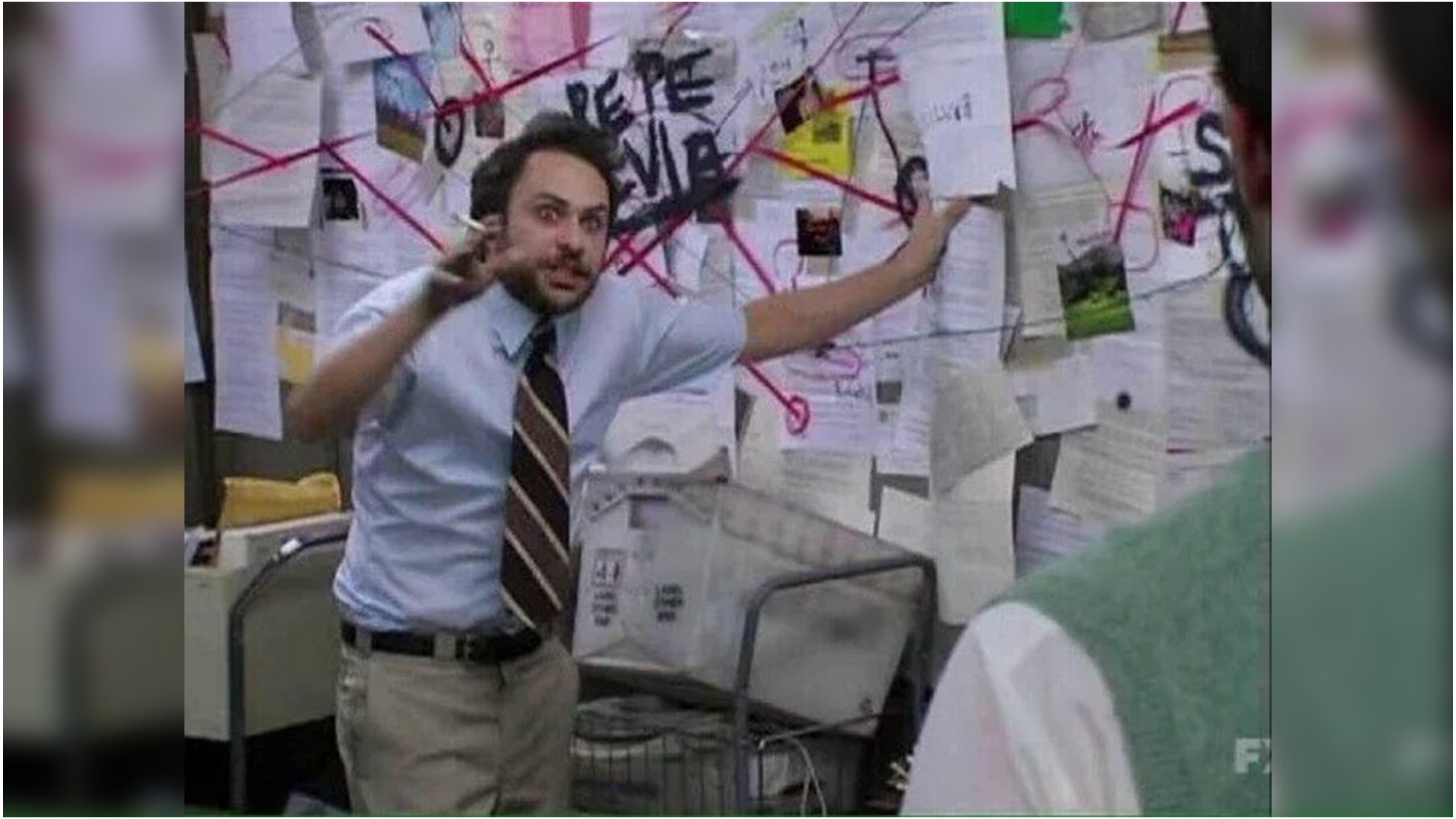
hctsa feature name	Description
Distribution	
DN_HistogramMode_5	Mode of z-scored distribution (5-bin histogram)
DN_HistogramMode_10	Mode of z-scored distribution (10-bin histogram)
Simple temporal statistics	
SB_BinaryStats_mean_longstretch1	Longest period of consecutive values above the mean
DN_OutlierInclude_p_001_mdrmd	Time intervals between successive extreme events above the mean
DN_OutlierInclude_n_001_mdrmd	Time intervals between successive extreme events below the mean
Linear autocorrelation	
CO_f1ecac	First $1/e$ crossing of autocorrelation function
CO_FirstMin_ac	First minimum of autocorrelation function
SP_Summaries_welch_rect_area_5_1	Total power in lowest fifth of frequencies in the Fourier power spectrum
SP_Summaries_welch_rect_centroid	Centroid of the Fourier power spectrum
FC_LocalSimple_mean3_stderr	Mean error from a rolling 3-sample mean forecasting
Nonlinear autocorrelation	
CO_trev_1_num	Time-reversibility statistic, $\langle (x_{t+1} - x_t)^3 \rangle_t$
CO_HistogramAMI_even_2_5	Automutual information, $m = 2, \tau = 5$
IN_AutoMutualInfoStats_40_gaussian_fmmi	First minimum of the automutual information function
Successive differences	
MD_hrv_classic_pnn40	Proportion of successive differences exceeding 0.04σ [20]
SB_BinaryStats_diff_longstretch0	Longest period of successive incremental decreases
SB_MotifThree_quantile_hh	Shannon entropy of two successive letters in equiprobable 3-letter symbolization
FC_LocalSimple_mean1_tauresrat	Change in correlation length after iterative differencing
CO_Embed2_Dist_tau_d_expfit_meandiff	Exponential fit to successive distances in 2-d embedding space
Fluctuation Analysis	
SC_FluctAnal_2_dfa_50_1_2_logi_prop_r1	Proportion of slower timescale fluctuations that scale with DFA (50% sampling)
SC_FluctAnal_2_rsrangefit_50_1_logi_prop_r1	Proportion of slower timescale fluctuations that scale with linearly rescaled range fits
Others	
SB_TransitionMatrix_3ac_sumdiagcov	Trace of covariance of transition matrix between symbols in 3-letter alphabet
PD_PeriodicityWang_th0_01	Periodicity measure of [31]

Table 1 The *catch22* feature set spans a diverse range of time-series characteristics representative of the diversity of interdisciplinary methods for time-series analysis. Features in *catch22* capture time-series properties of the distribution of values in the time series, linear and nonlinear temporal autocorrelation properties, scaling of fluctuations, and others.

https://arxiv.org/pdf/1901.10200.pdf



Apophenia Explained | Pareidolia, Confirmation Bias, & Other Pattern Errors

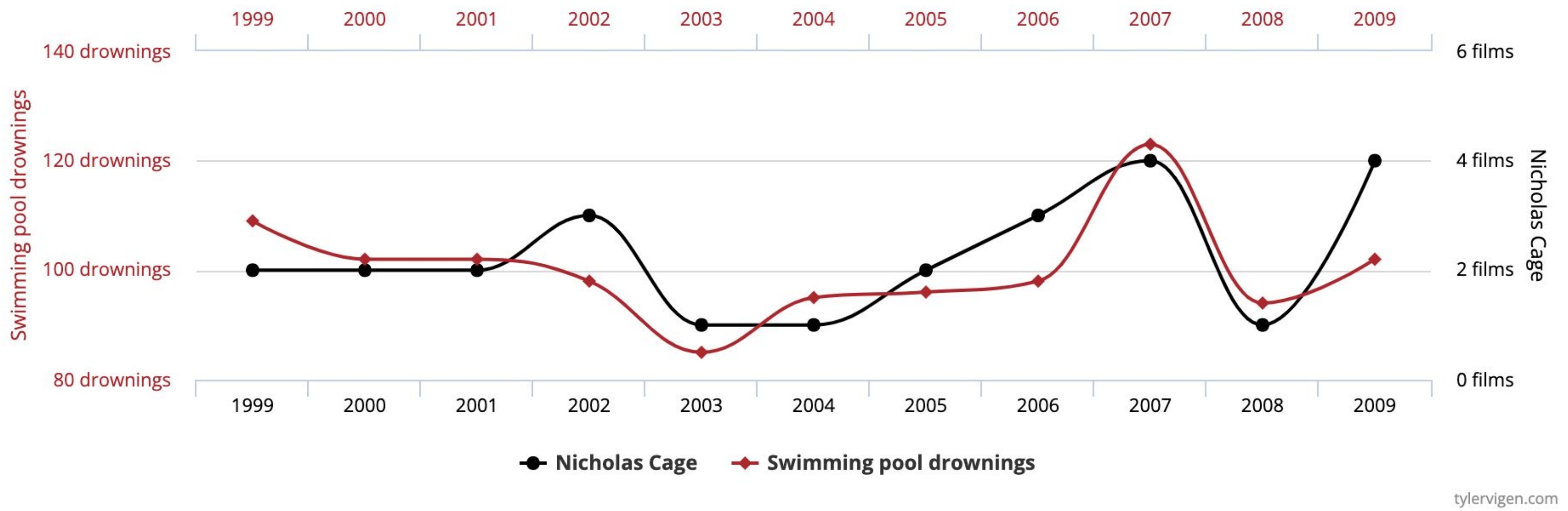


Number of people who drowned by falling into a pool

correlates with

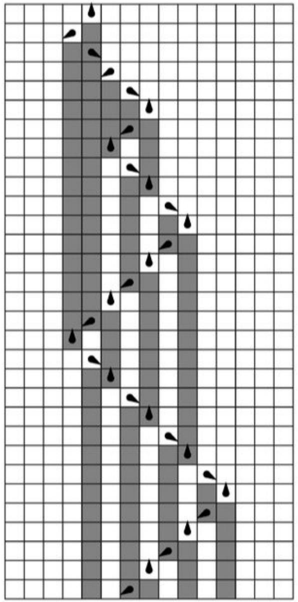
Films Nicolas Cage appeared in

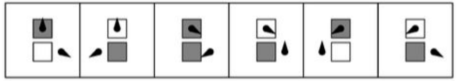
Correlation: 66.6% (r=0.666004)



Data sources: Centers for Disease Control & Prevention and Internet Movie Database

Universal Turing Machines Show Us that Perfect Prediction is Impossible

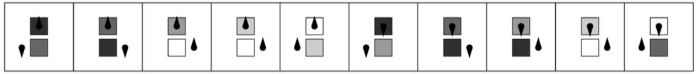




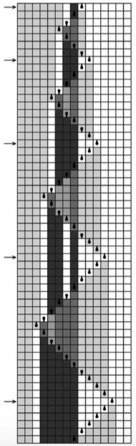
An example of a Turing machine. Like a mobile automaton, the Turing machine has one active cell or "head," but now the head has several possible states, indicated by the directions of the arrows in this picture.

https://www.wolframscience.com/nks/p78--turing-machines/

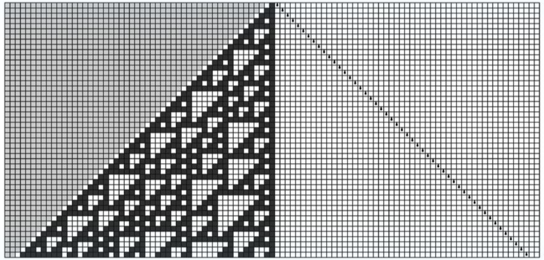
Universal Turing Machines Show Us that Perfect Prediction is Impossible



The rule for the simplest Turing machine currently known to be universal, based on discoveries in this book. The machine has 2 states and 5 possible colors.



Turing machine evolution



Turing machine evolution compressed

An example of how the Turing machine above manages to emulate rule 110. The compressed picture is made by keeping only the steps indicated at which the head is further to the right than ever before. To get the picture shown requires running the Turing machine for a total of 5000 steps.

https://www.wolframscience.com/nks/p707--universality-in-turing-machines-and-other-systems/

The Game of Life

[Inventing Game of Life (John Conway) - Numberphile]

[Life in life]

Consequences

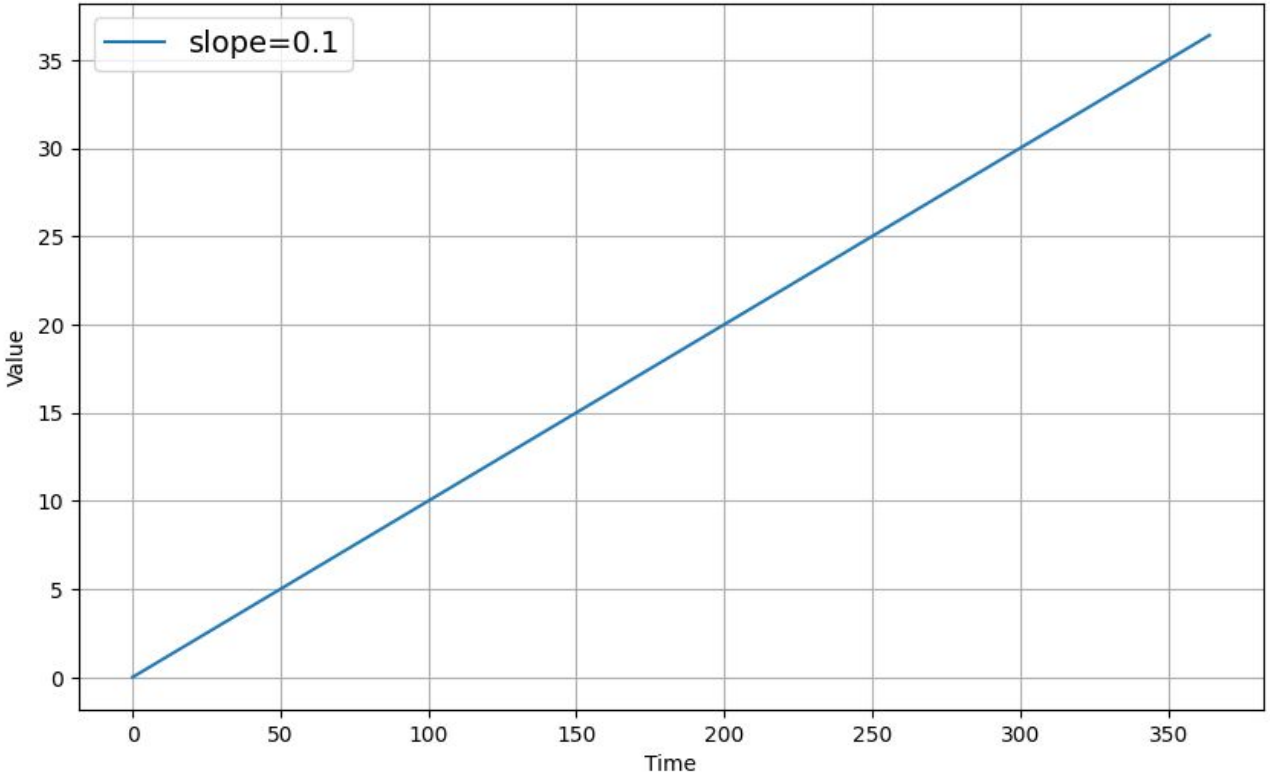
Our predictions will always have a margin of error. ***“How much will they be wrong?”*** is something that every stakeholder will ask us and we will never be able to have a definite answer **other than *the methodology that we used to generate our forecast.***

[Stephen Hawkings - The Meaning of Life (John Conway's Game of Life segment)]

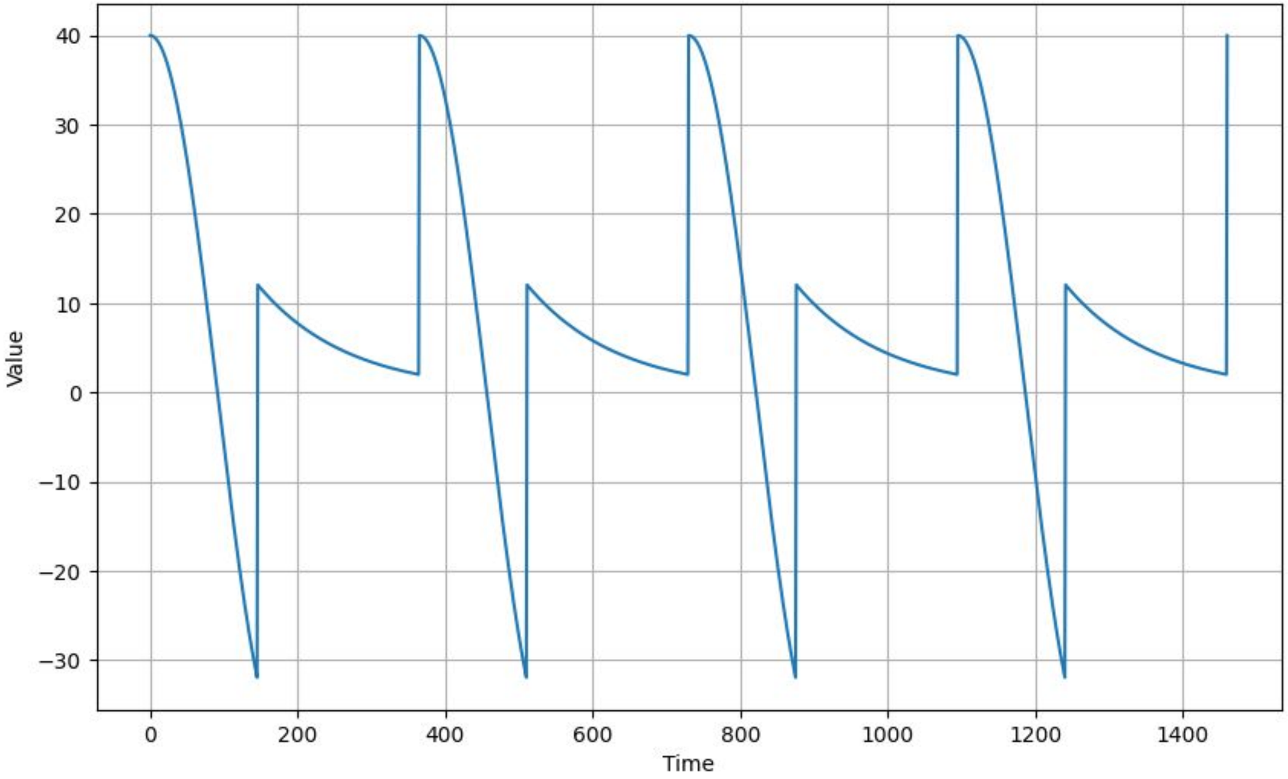
Setting the date as an index

```
data = {

    'date': ['2023-08-15', '2023-08-16', '2023-08-17'],

    'value': [10, 20, 30]

}

df = pd.DataFrame(data)

df['date'] = pd.to_datetime(df['date'])

df.set_index('date', inplace=True)
```

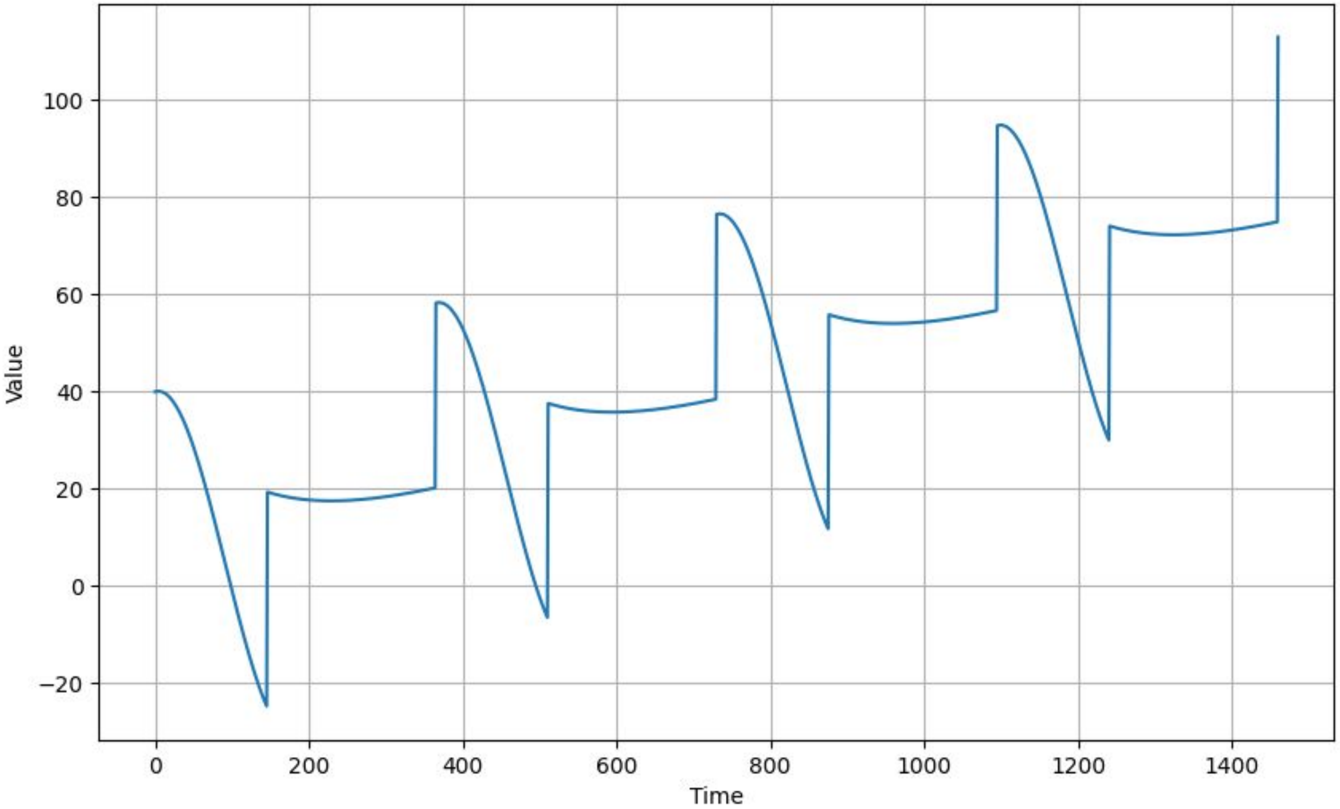
Trend


slope=0.1
Value
Time
0
5
10
15
20
25
30
35
0
50
100
150
200
250
300
350

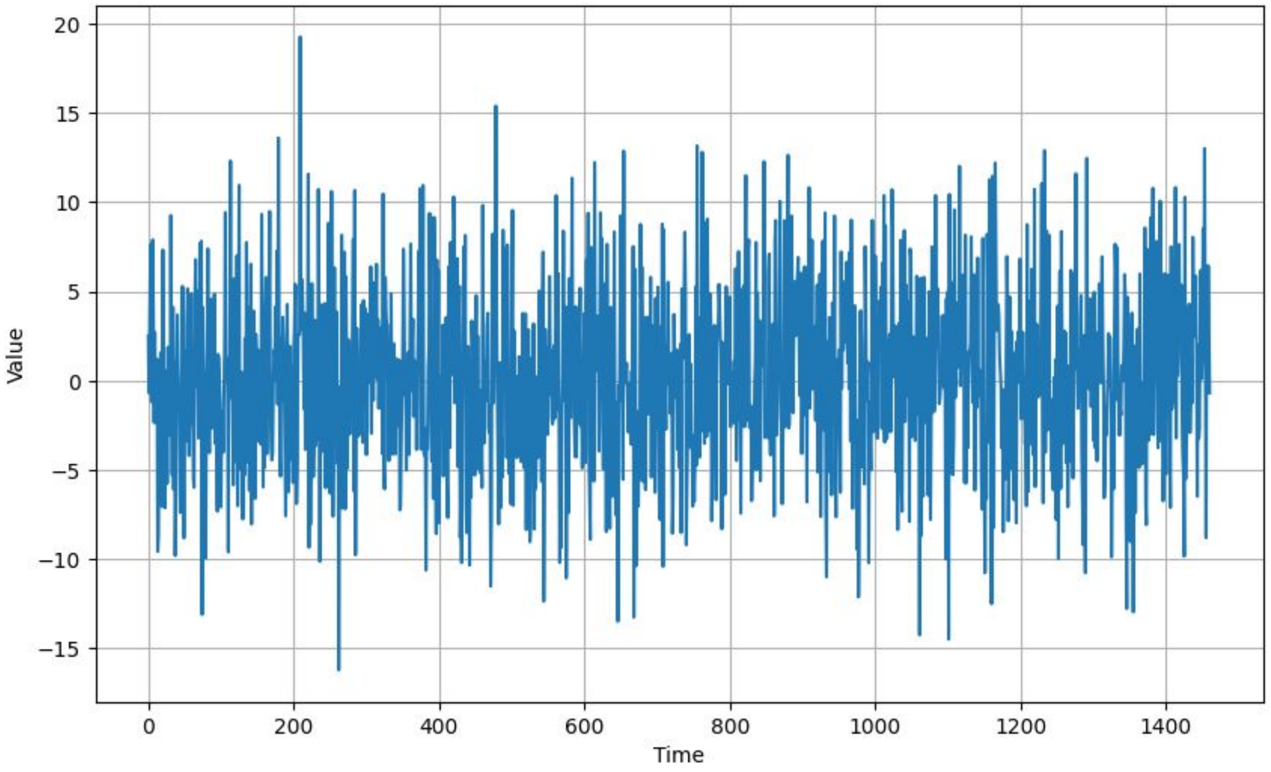
Seasonality



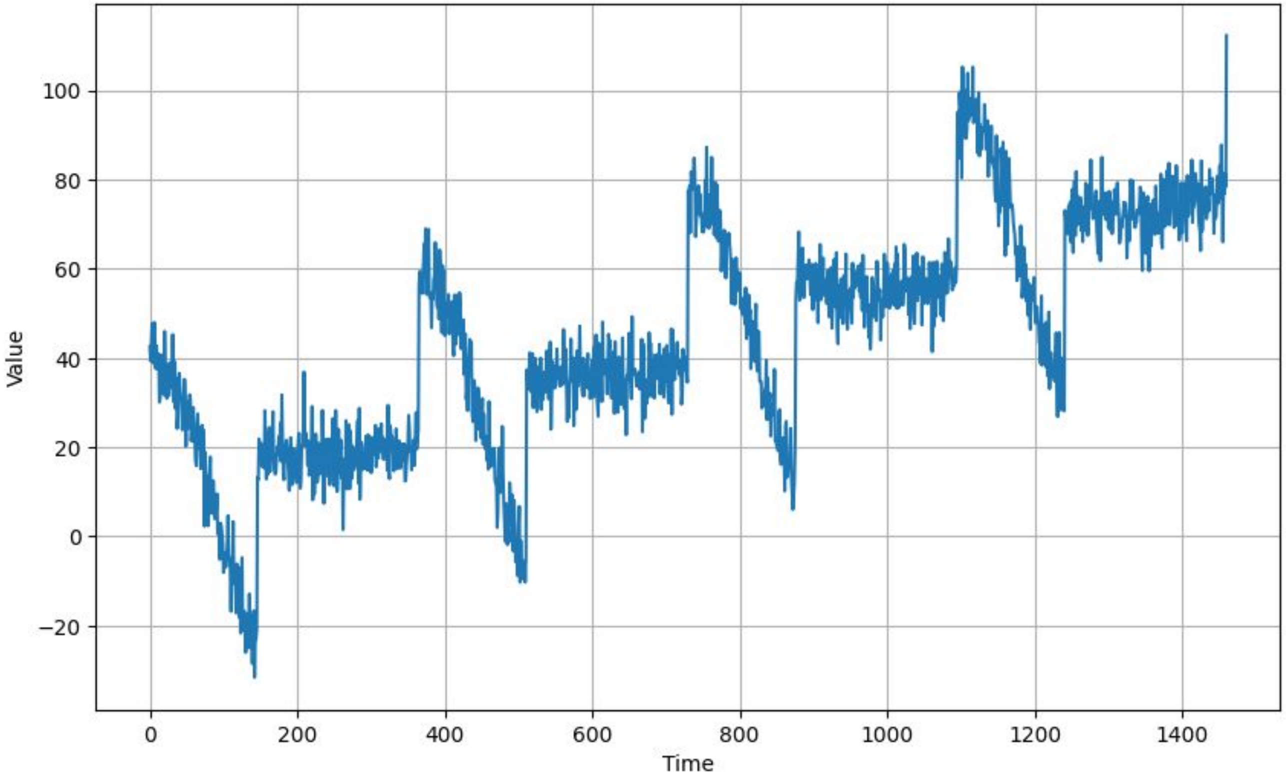
Trend and Seasonality



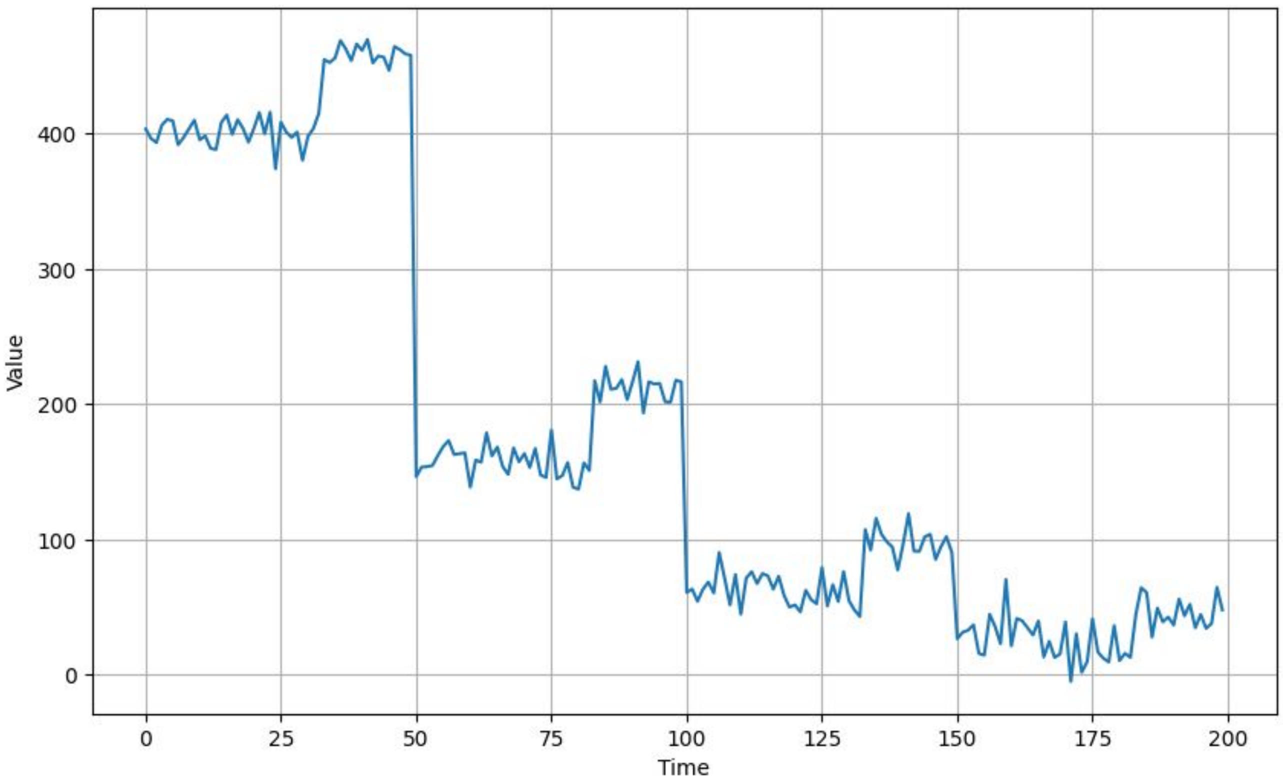
Noise



Trend + Seasonality + Noise

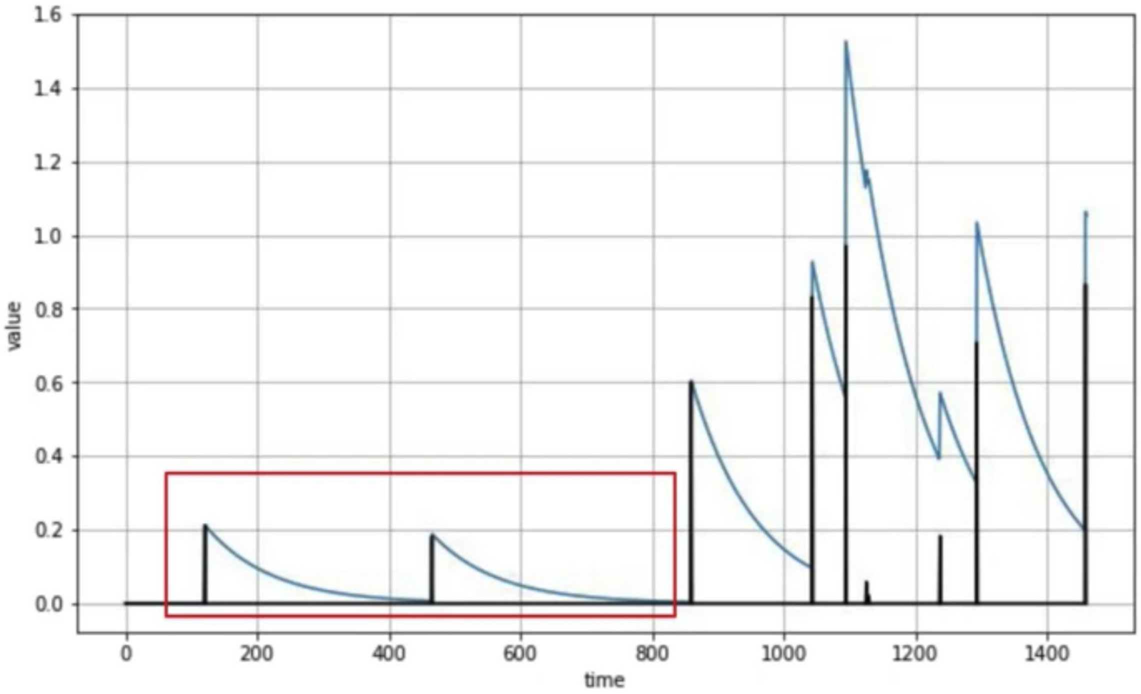


Autocorrelation

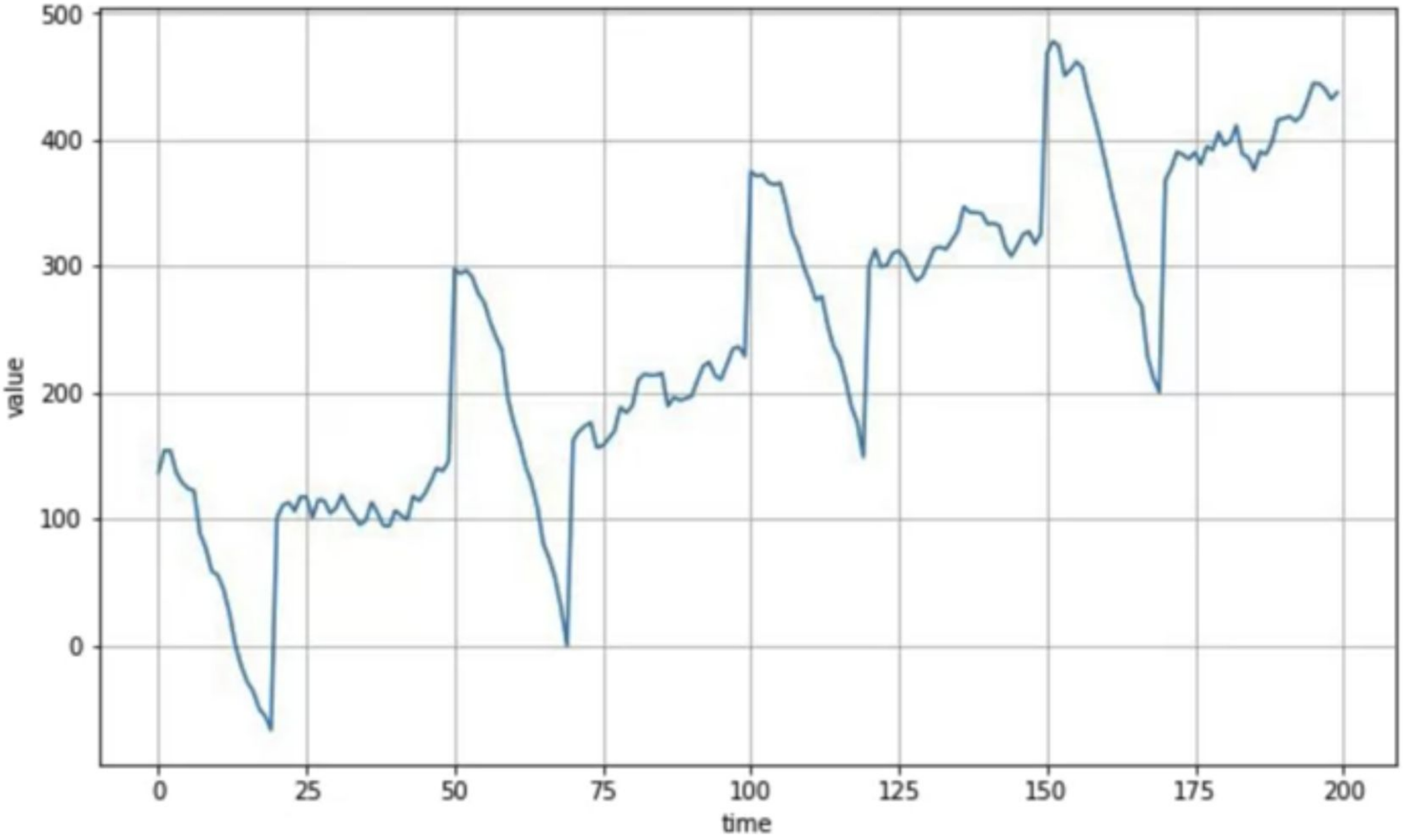


Memory and "innovations" in time series

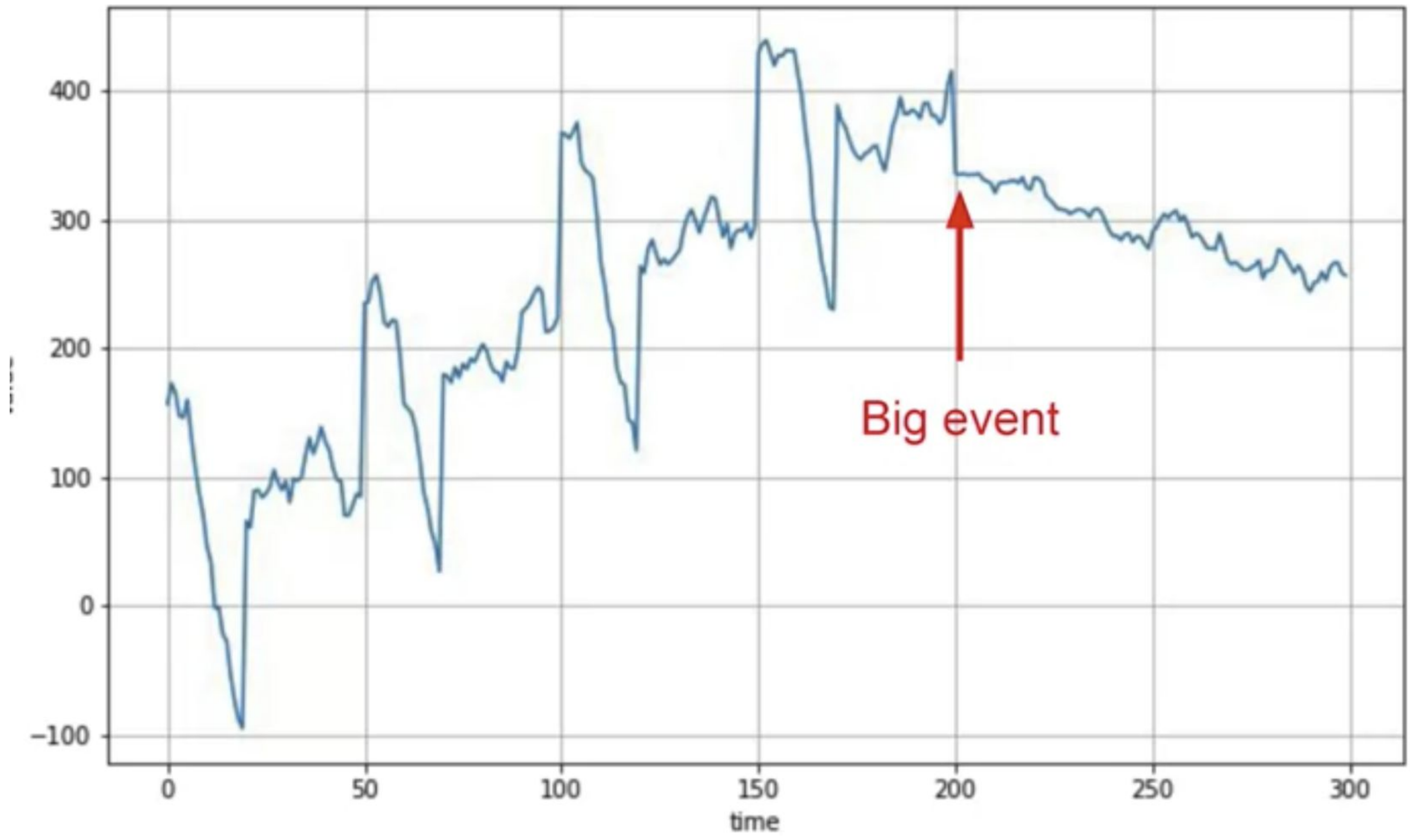
$v(t) = 0.99 \times v(t-1)$ + occasional spike



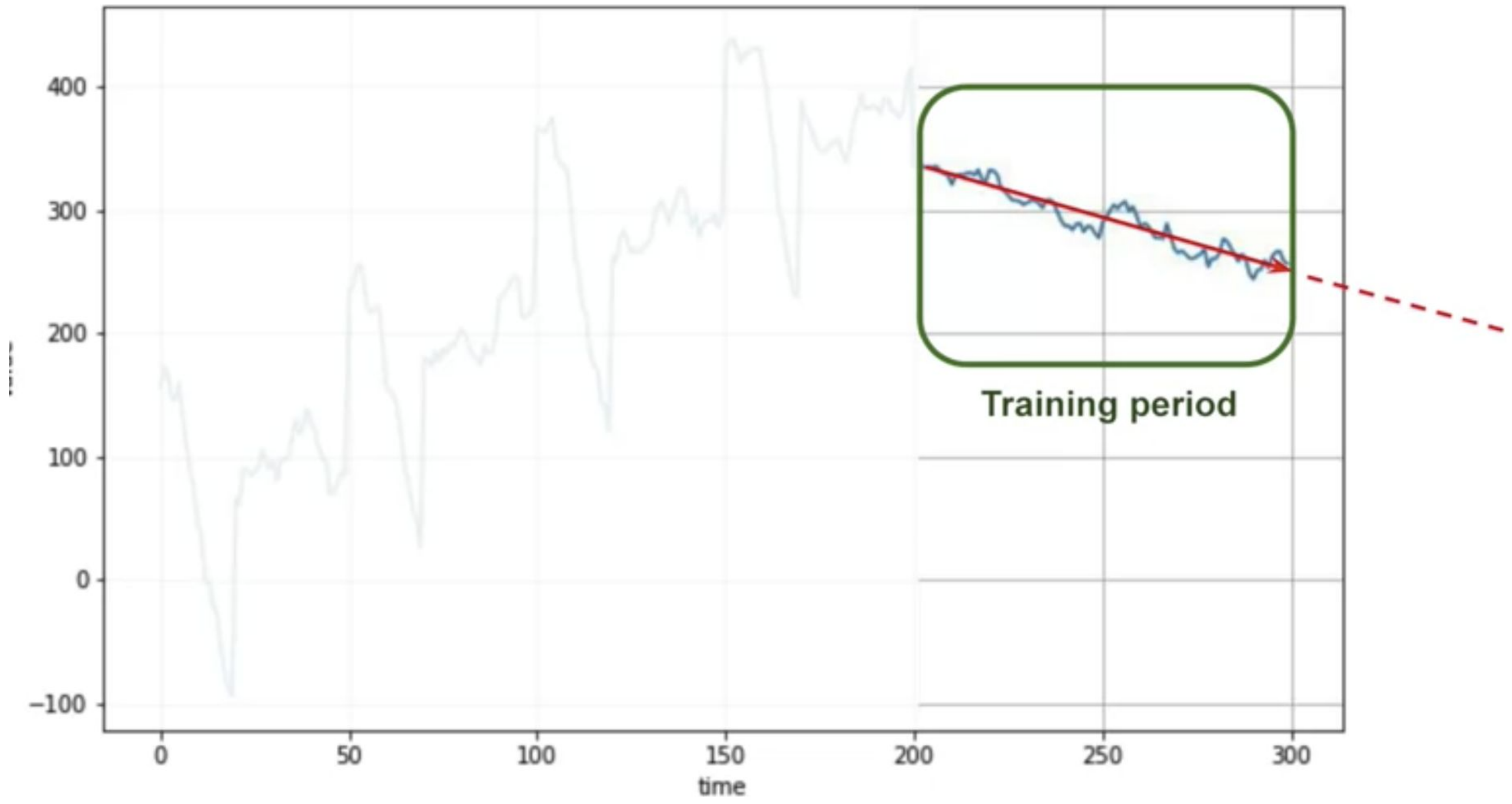
Trend + Seasonality + Autocorrelation + Noise



Non-Stationary Time Series



Non-Stationary Time Series

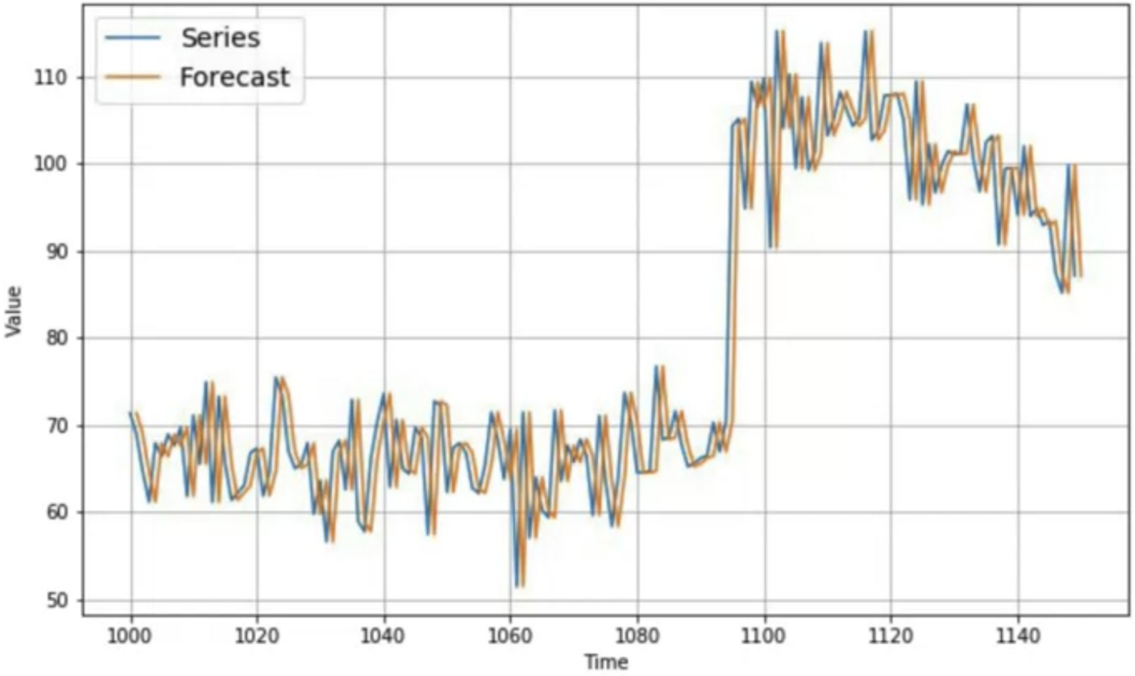


Explore time series properties with synthetic data

- Notebook

Baselines for Forecasting

Naive Forecasting



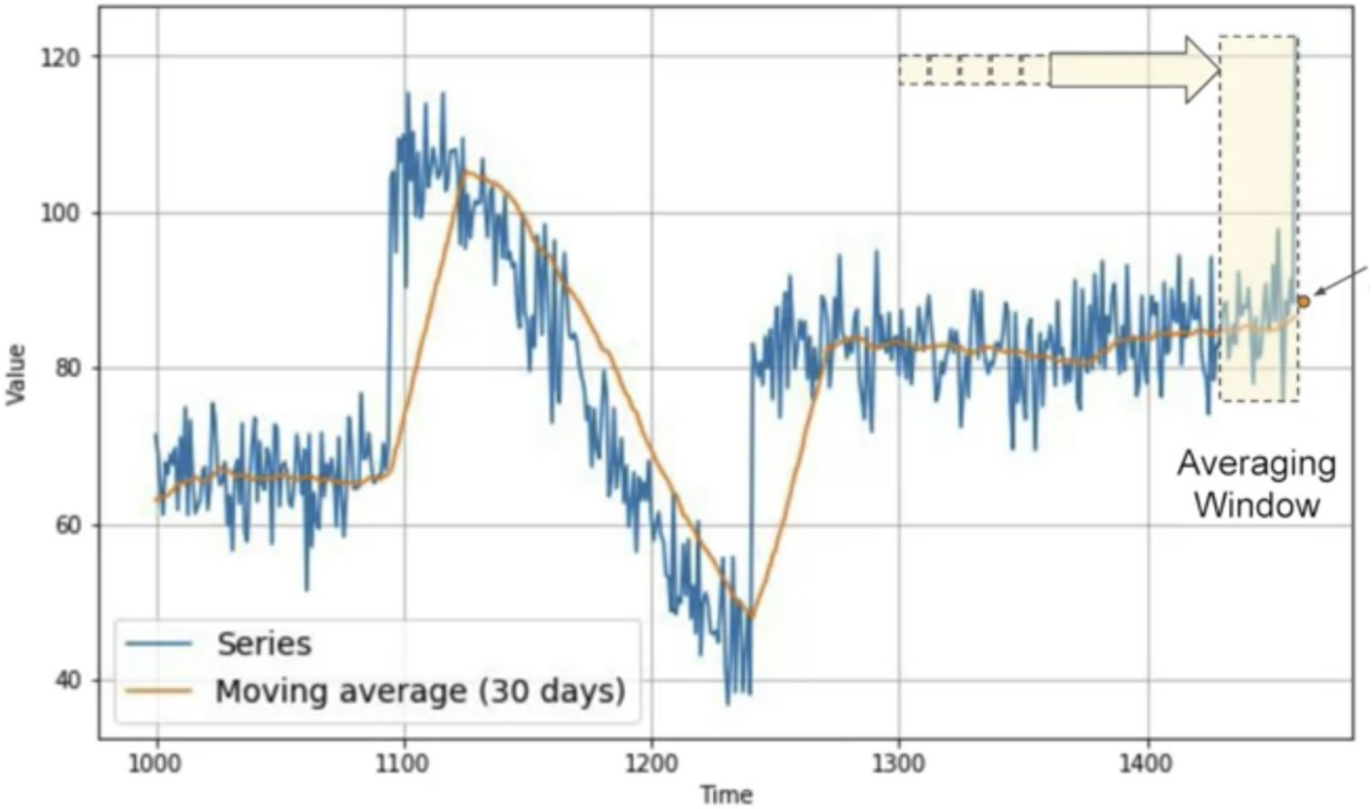
Moving averages

Pandas was invented specifically for time series analysis

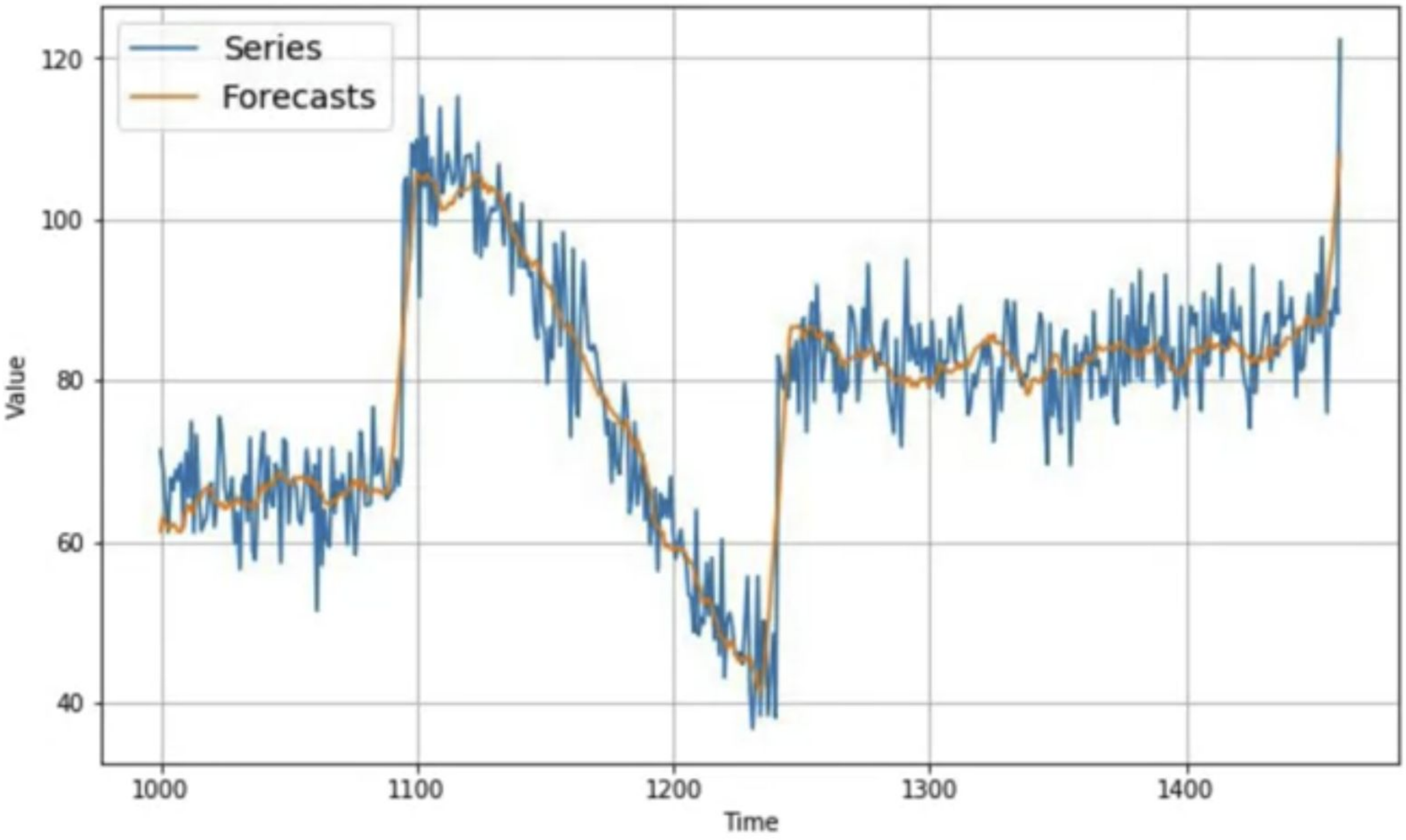
$$\bar{a}_{\mathrm{SM}} = \frac{x_n + x_{n-1} + \ldots + x_{M-(n-1)}}{M}$$

$$\bar{a}_{\mathrm{SM}} = \frac{1}{M} \sum_{i=0}^{n-1} x_{M-i}$$

Moving Average



Smoothing Both Past and Present Values

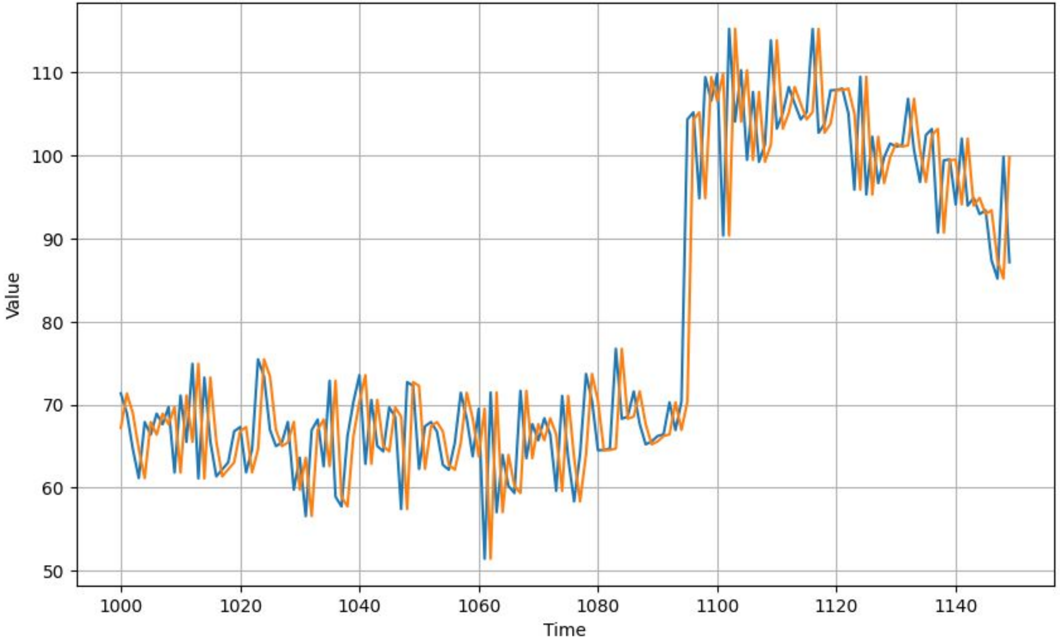


Why use moving averages?

- Remove outliers
- Smoothen out short-term fluctuations
- Highlight trends or cycles

Forecasting with Baseline Models

Notebook



A Personal Experience

At the start of this year, I was tasked with improving the forecasting model used to predict the occupancy of lockers the Vinted Go network.

There was a previous model already in production, based on Markov Chains

It had no logging, no update of probabilities, and no metrics applied to evaluate its performance


L1


zalando
zalando
zalando
free to be
free to be
Open here!
Open here!
ökologisch
PLAY
DONPASCUAL
NAVARRA

Vinted

Articles ▾ | Rechercher des articles

S'inscrire | Se connecter

Vends tes articles

? FR ▾

Femmes Hommes Enfants Maison Divertissement Animaux À propos Notre plateforme

Trouver le point le plus proche

Points

CONSIGNE, Franprix

13 PLACE D'ALIGRE, PARIS 12, FR

CONSIGNE, Franprix

25/29 AVENUE JOFFRE, SAINT MANDE, FR

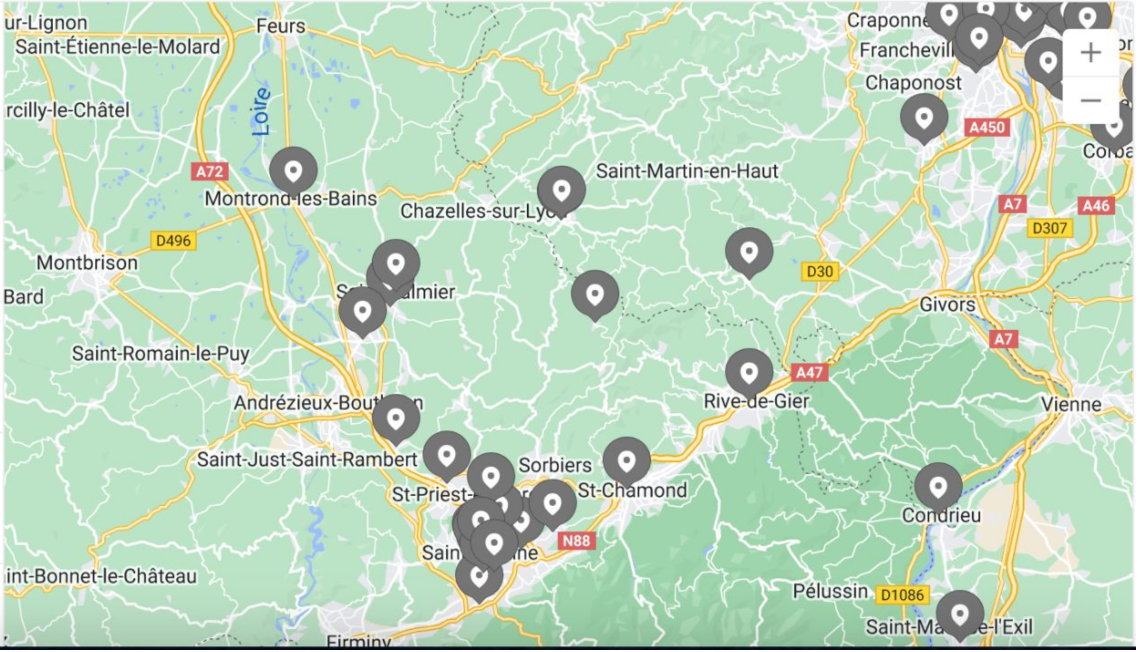
CONSIGNE, Franprix

139 AVENUE DE VERDUN, ISSY LES MOULINEAUX, FR

CONSIGNE, Franprix

7 RUE DES PETITES ECURIES, PARIS 10, FR

CONSIGNE, Franprix



VINTED ENGINEERING

These are the voyages of code tailors that help create Vinted.

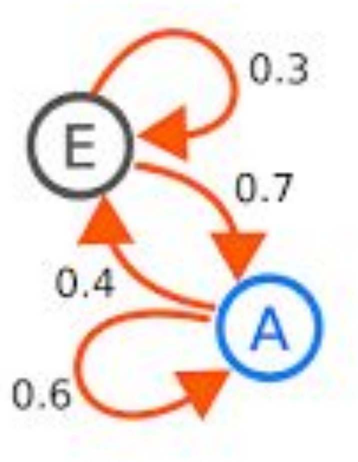
Home • Open Source • Join Us

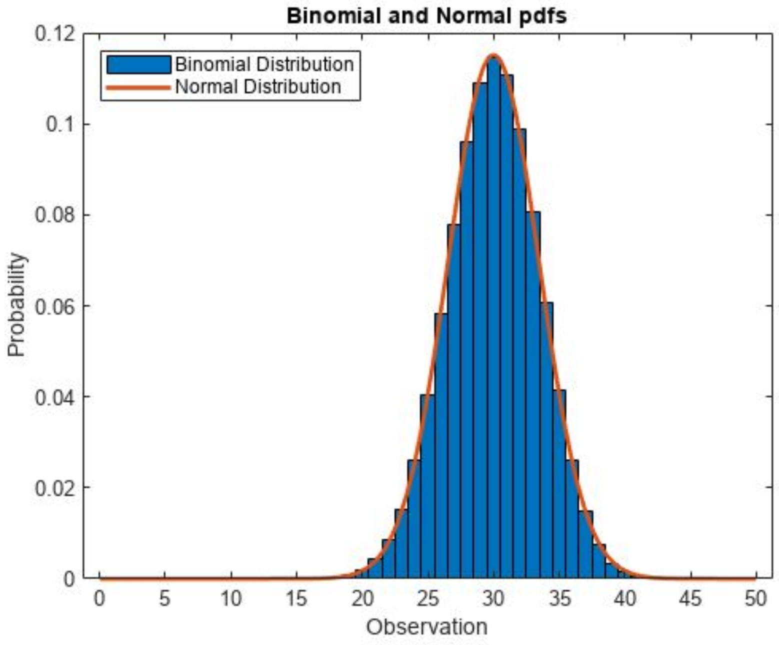
The Joy of Automation

March 4, 2023 by Simas Kučinskas and Mangirdas Skripka

https://vinted.engineering/2023/03/04/the-joy-of-automation/

Markov Chain





[How it works (notebook)]

Inflow Math

How does the script decide when to hide a locker? Well, math.

The script relies on two key metrics: occupancy rate and delivery backlog. The basic idea: When there are too many parcels waiting to be dropped off in relation to the number of locker slots available, the script hides the locker.

Our data scientists have built a mathematical model that takes these two inputs and produces a decision to hide or show the locker to our customers. Under the hood, you'll find things like Markov chains and binomial distributions. But the upshot of all that math is that the script can forecast the probability that the locker will overflow if left visible. Once the predicted probability of overflow reaches a critical threshold, the locker is hidden.

Crisis and chaos averted.

https://vinted.engineering/2023/03/04/the-joy-of-automation/

Our second step is automation. We've built a script that runs every hour and hides or shows individual lockers based on the current data. To play it safe, the manual override stayed in place.

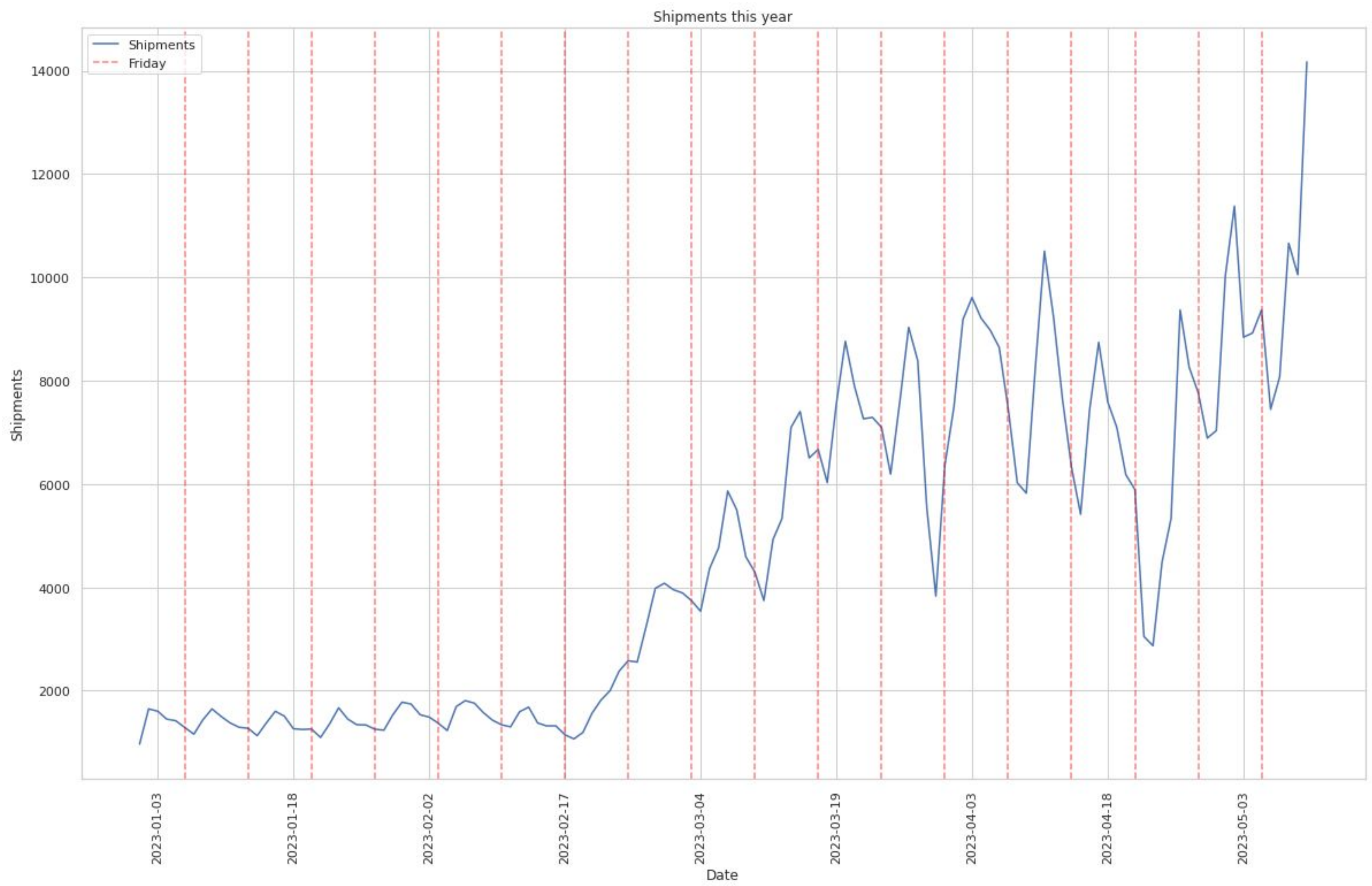
Code	Sorting center	Provider id	Address	State	Visibility	AI decision	AI control
00056	PAR1	[illegible]	Title: Vinted HQ Address: Svitrigailos g. 13, Vilnius, 03228, FR	Active	Visible	Hidden	On
00057	PAR1	[illegible]	Title: FRANPRIX Address: 32 RUE DE LOURMEL, PARIS, 75015, FR	Inactive	Visible	Visible	On
00058	PAR1	[illegible]	Title: FRANPRIX Address: 53 RUE LOSSERAND, PARIS, 75014, FR	Active	Visible	Visible	On
00059	PAR1	[illegible]	Title: FRANPRIX Address: 98-108 RUE PETIT, PARIS, 75019, FR	Inactive	Visible	Visible	Off
00060	PAR1	[illegible]	Title: FRANPRIX Address: 13 PLACE D'ALIGRE, PARIS, 75012, FR	Active	Visible	Visible	On
00061	PAR1	[illegible]	Title: FRANPRIX Address: 25 BOULEVARD GOUVION ST CYR, PARIS, 75017, FR	Active	Visible	Visible	On
00062	PAR1	[illegible]	Title: FRANPRIX Address: 13 RUE ETIENNE D'ORVES, PARIS, 92110, FR	Active	Visible	Visible	Off
00063	PAR1	[illegible]	Title: FRANPRIX Address: 126 RUE DE PICPUS, PARIS, 75012, FR	Active	Visible	Visible	On
00065	PAR1	[illegible]	Title: FRANPRIX Address: 121 RUE DU GENERAL LECLERC, PARIS, 75014, FR	Active	Visible	Visible	On
00066	PAR1	[illegible]	Title: FRANPRIX Address: AVENUE GENERAL GALLIENI, PARIS, 94340, FR	Active	Visible	Visible	On

Technical and scientific debt of the deployed model

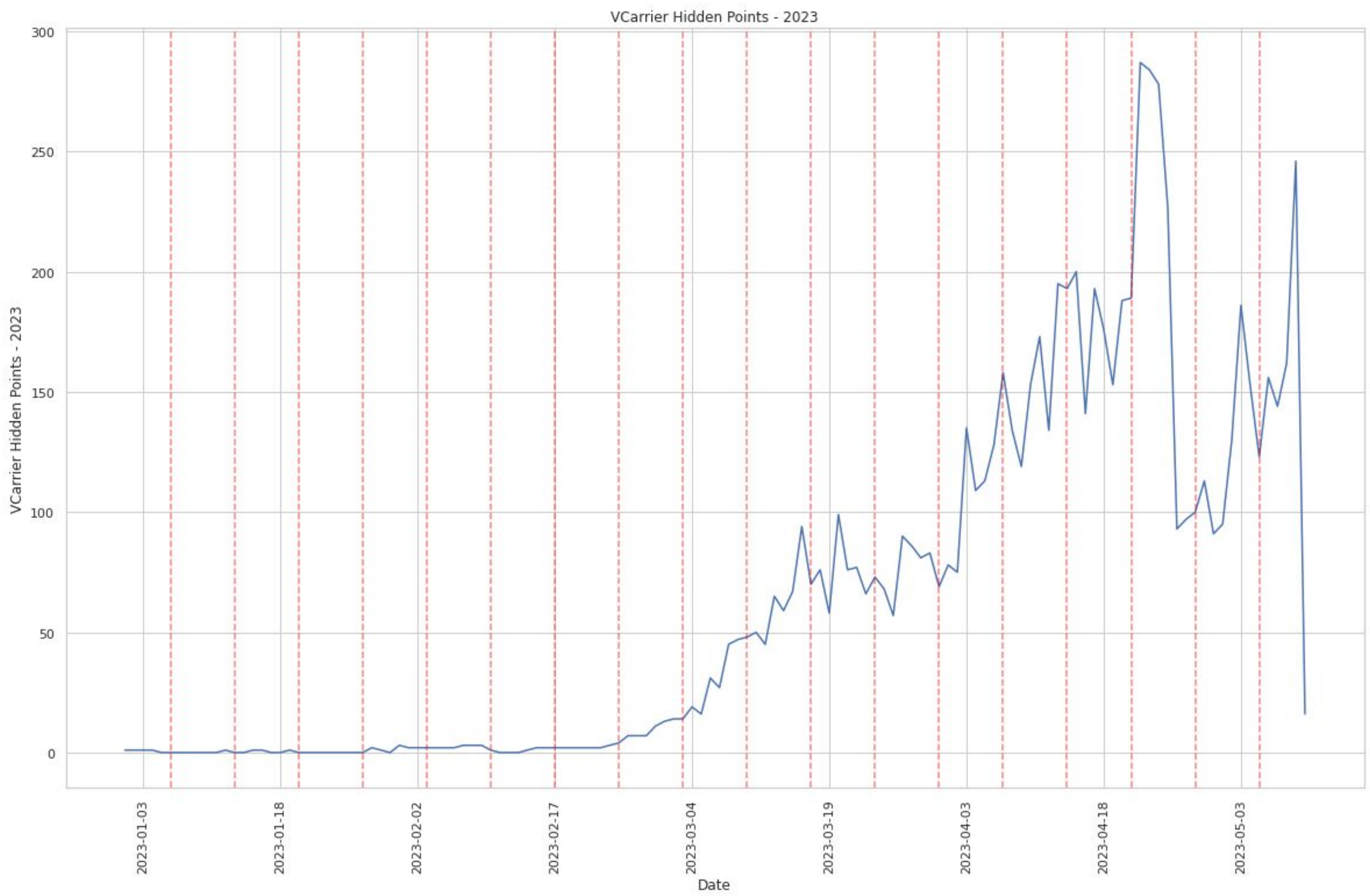
- No metrics
- No testing
- No logging
- No simpler baseline tried
- No updates to the probabilities that the model requires to be reasonable
- No account of trend or seasonality in the model
- Depended on three different parameters to define hiding threshold
  - Hiding threshold, probability of pickup, probability of drop-off
  - These probabilities were changing, the Markov chain was non-stationary
- Implemented in Ruby (yes, for real) by backend developers
- Stakeholders didn’t care about MAE and wanted “something more business oriented” (huge communication problem) but couldn’t agree on what
  - Cost Per Parcel vs Perfect Parcel Rate - no agreement within OPs team and leadership
- **This model went live immediately with huge impact to the company (crises and chaos caused)**

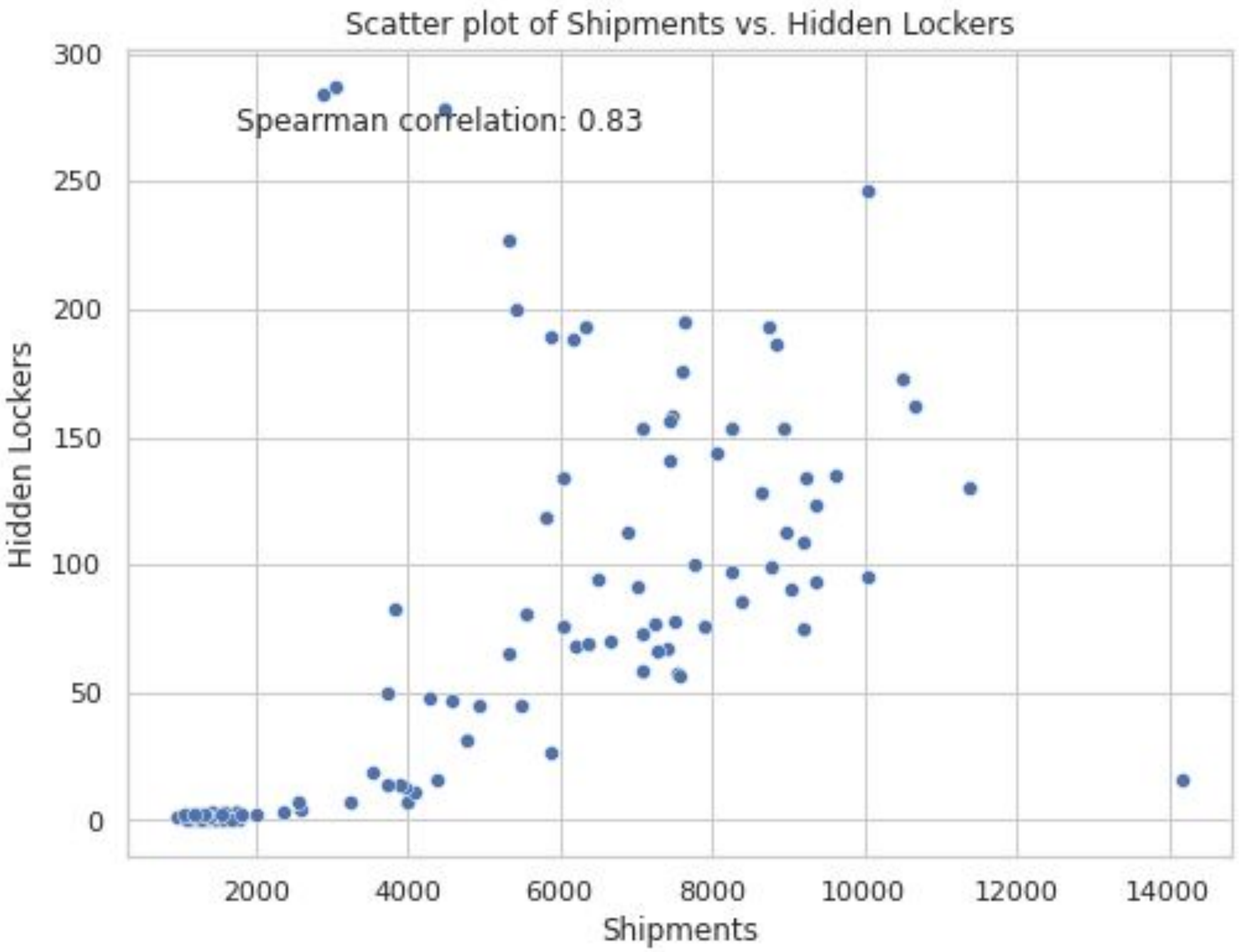
Fun Twist

People started talking about “lost transactions due to hiding” in our WBRs



Shipments this year
Shipments
Friday
14000
12000
10000
8000
6000
4000
2000
Shipments
2023-01-03
2023-01-18
2023-02-02
2023-02-17
2023-03-04
2023-03-19
2023-04-03
2023-04-18
2023-05-03
Date


VCarrier Hidden Points - 2023
VCarrier Hidden Points - 2023
300
250
200
150
100
50
0
2023-01-03
2023-01-18
2023-02-02
2023-02-17
2023-03-04
2023-03-19
2023-04-03
2023-04-18
2023-05-03
Date



Scatter plot of Shipments vs. Hidden Lockers
Spearman correlation: 0.83
Hidden Lockers
Shipments
300
250
200
150
100
50
0
2000
4000
6000
8000
10000
12000
14000

Benchmarks from May 2023 for Next Day Occupancy

Mean Absolute Error from Markov model advertised in Vinted's blog post = 17.5

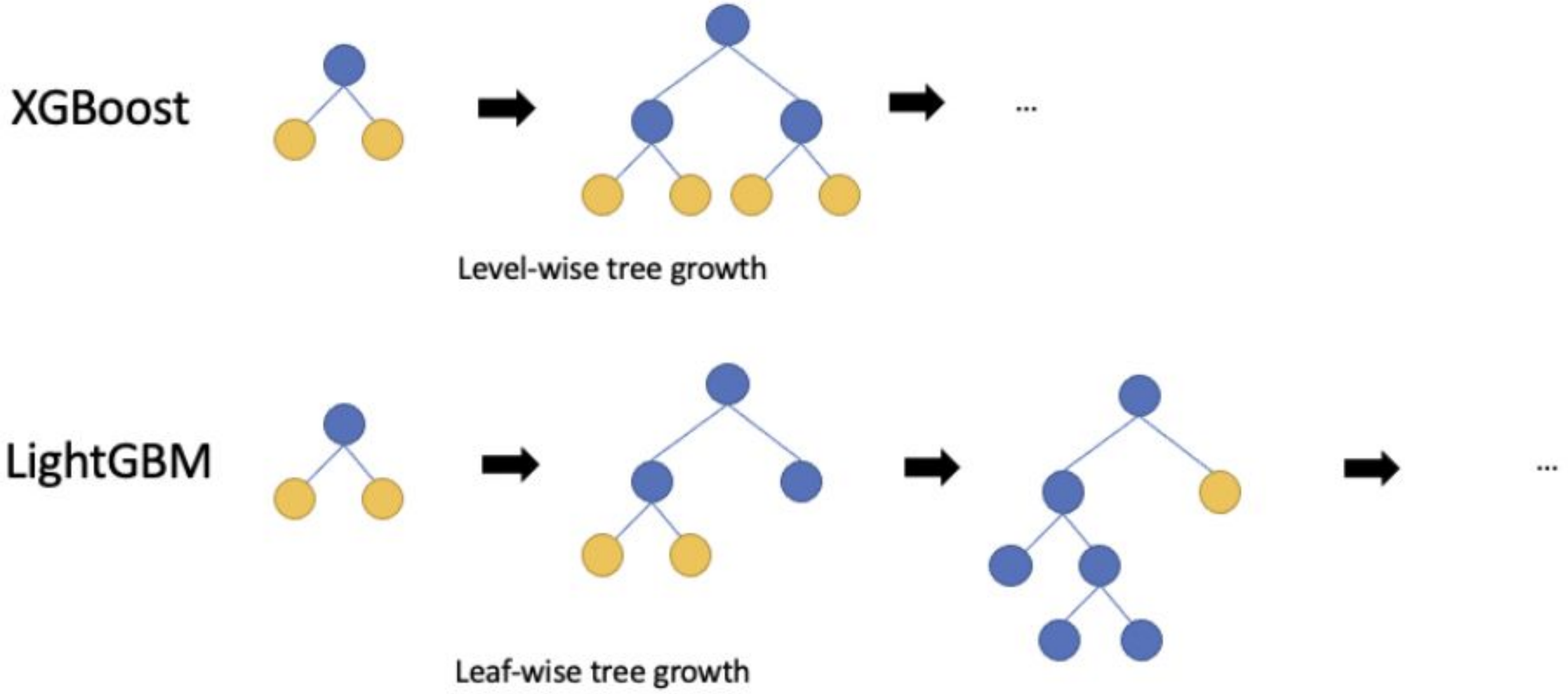
Mean Absolute Error from predicting the occupancy of next day as today's occupancy = 3.2

Y(t + 1) = Y(t) was better than the deployed Markov chain model by 5x!

Mean Absolute Error from LightGBM regression model = 5.1

What We did: Change to Mean Next Week Occupancy

- We trained LightGBM (and deployed on GCP) to predict the average occupancy of next week
- This required feature engineering
- This was better in MAE than an simple moving average **(only 2.6 vs 3.2**) on the back testing
- Main challenges were due to infrastructure and communication with Ops team
- “Lost transactions due to hiding” was a term commonly used in our WBRs

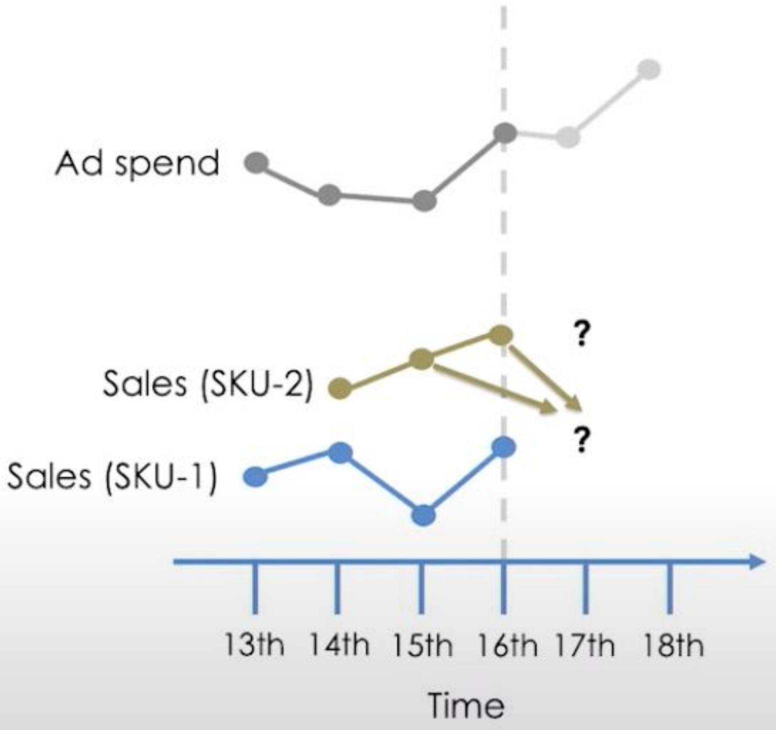

XGBoost
...
Level-wise tree growth
LightGBM
...
Leaf-wise tree growth

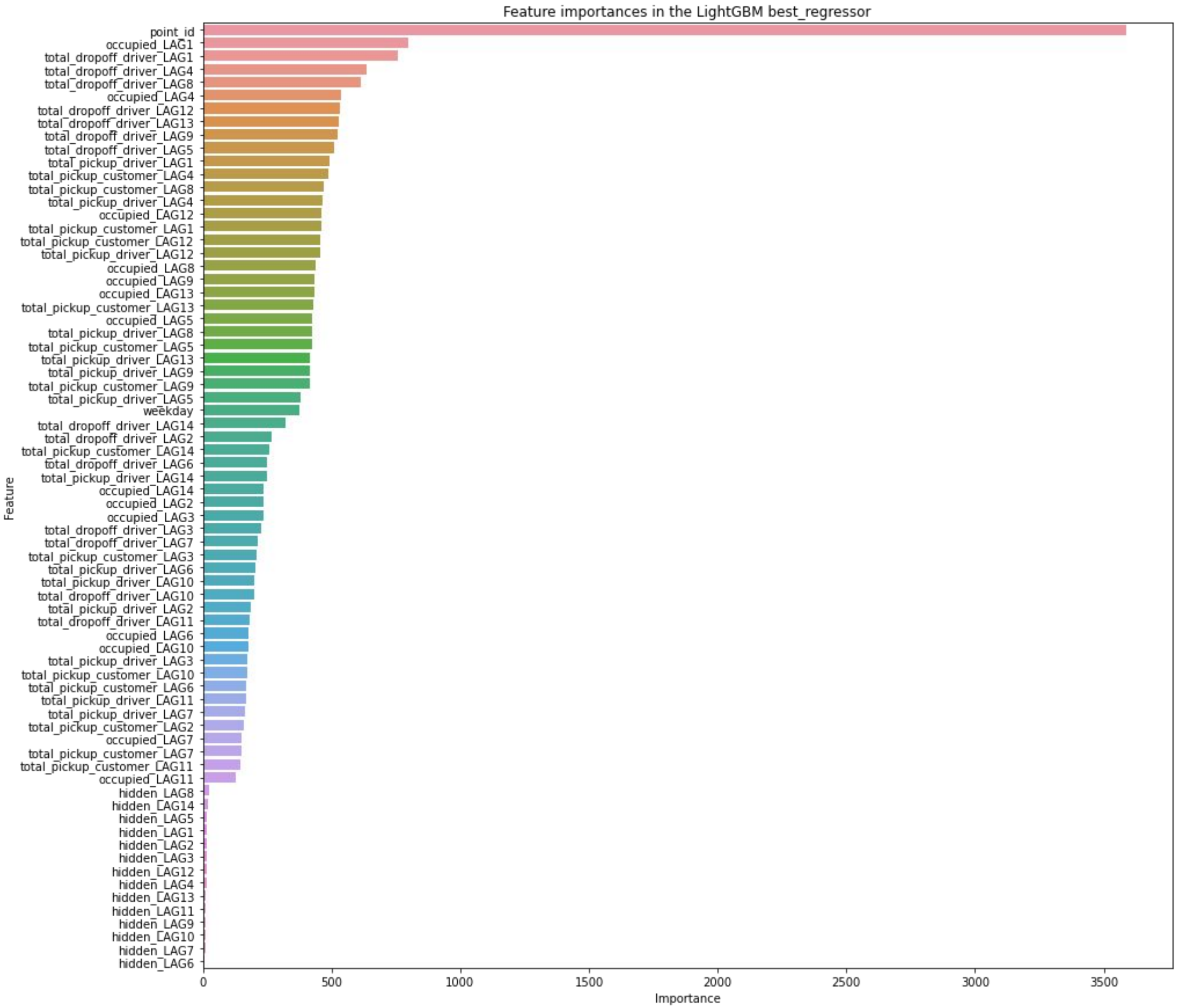
Lag features: Past values of target & features

- Past values of the target are likely to be predictive:

$$\hat{y}_t \leftarrow y_{t-1}, y_{t-2}, \dots y_{t-k}$$

- Seasonal lags good for seasonality (e.g., lag of 7 for weekly seasonality).
- Can use lags of other target time series.




Feature importances in the LightGBM best_regressor
point_id
occupied_LAG1
total_dropoff_driver_LAG1
total_dropoff_driver_LAG4
total_dropoff_driver_LAG8
occupied_LAG4
total_dropoff_driver_LAG12
total_dropoff_driver_LAG13
total_dropoff_driver_LAG9
total_dropoff_driver_LAG5
total_pickup_driver_LAG1
total_pickup_customer_LAG4
total_pickup_customer_LAG8
total_pickup_driver_LAG4
occupied_LAG12
total_pickup_customer_LAG1
total_pickup_customer_LAG12
total_pickup_driver_LAG12
occupied_LAG8
occupied_LAG9
occupied_LAG13
total_pickup_customer_LAG13
occupied_LAG5
total_pickup_driver_LAG8
total_pickup_customer_LAG5
total_pickup_driver_LAG13
total_pickup_driver_LAG9
total_pickup_customer_LAG9
total_pickup_driver_LAG5
weekday
total_dropoff_driver_LAG14
total_dropoff_driver_LAG2
total_pickup_customer_LAG14
total_dropoff_driver_LAG6
total_pickup_driver_LAG14
occupied_LAG14
occupied_LAG2
occupied_LAG3
total_dropoff_driver_LAG3
total_dropoff_driver_LAG7
total_pickup_customer_LAG3
total_pickup_driver_LAG6
total_pickup_driver_LAG10
total_dropoff_driver_LAG10
total_pickup_driver_LAG2
total_dropoff_driver_LAG11
occupied_LAG6
occupied_LAG10
total_pickup_driver_LAG3
total_pickup_customer_LAG10
total_pickup_customer_LAG6
total_pickup_driver_LAG11
total_pickup_driver_LAG7
total_pickup_customer_LAG2
occupied_LAG7
total_pickup_customer_LAG7
total_pickup_customer_LAG11
occupied_LAG11
hidden_LAG8
hidden_LAG14
hidden_LAG5
hidden_LAG1
hidden_LAG2
hidden_LAG3
hidden_LAG12
hidden_LAG4
hidden_LAG13
hidden_LAG11
hidden_LAG9
hidden_LAG10
hidden_LAG7
hidden_LAG6
Feature
0
500
1000
1500
2000
2500
3000
3500
Importance

Lessons learned

- **Running bad forecasts can be potentially catastrophic and embarrassing, with a lot of finger-pointing**

- **Trying the simplest thing first and measuring it should always be the first approach (even when inheriting a project)**

- **“What is the problem that we are trying to solve?” should always be the first question to stakeholders**

Lessons learned

- **Crunching out number is easy, crunching out number that you and others can trust and understand is hard**
- **Communicating with teammates and stakeholders was the major challenge**
- **“Null baselines” can be surprisingly difficult to beat**

The Makridakis Competitions

Spyros Makridakis

Time series is a field of data science that has ran competitions to benchmark forecast models, **20 years before Kaggle**

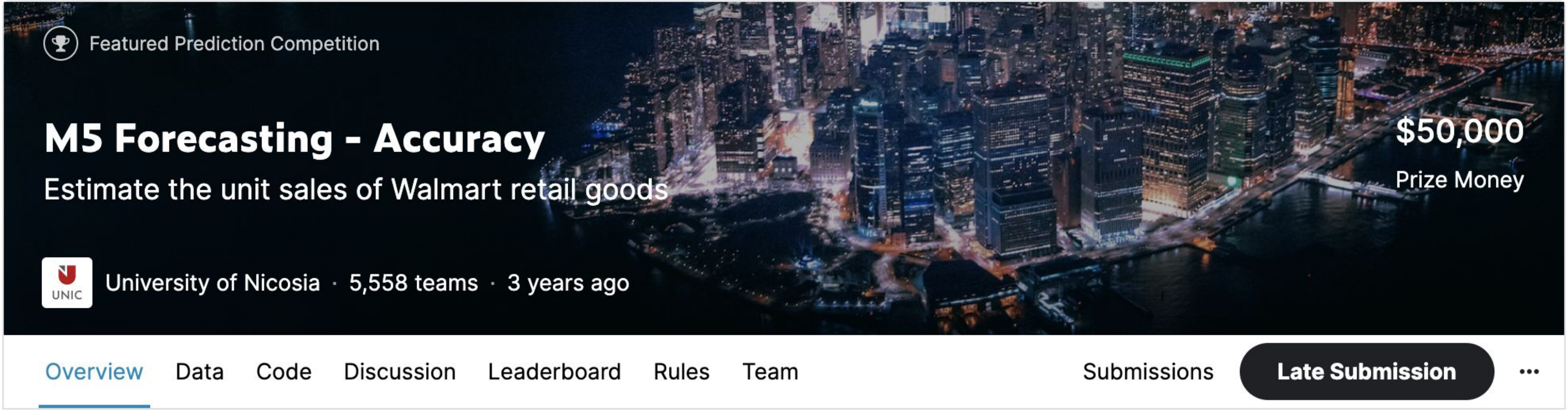
https://en.wikipedia.org/wiki/Makridakis_Competitions

The Makridakis Competitions

No.	Informal name for competition	Year of publication of results	Number of time series used	Number of methods tested	Other features
1	M Competition or M-Competition[1][5]	1982	1001 (used a subsample of 111 for the methods where it was too difficult to run all 1001)	15 (plus 9 variations)	Not real-time
2	M-2 Competition or M2-Competition[1][6]	1993	29 (23 from collaborating companies, 6 from macroeconomic indicators)	16 (including 5 human forecasters and 11 automatic trend-based methods) plus 2 combined forecasts and 1 overall average	Real-time, many collaborating organizations, competition announced in advance
3	M-3 Competition or M3-Competition[1]	2000	3003	24	
4	M-4 Competition or M4 Competition	2020[7]	100,000	All major ML and statistical methods have been tested	First winner Slawek Smyl, Uber Technologies
5	M-5 Competition or M5 Competition	Initial results 2021, Final 2022	Around 42,000 hierarchical timeseries provided by Walmart	All major forecasting methods, including Machine and Deep Learning, and Statistical ones will be tested	First winner Accuracy Challenge: YeonJun In. First winners uncertainty Challenge: Russ Wolfinger and David Lander
6	M-6 Competition or M6 Competition	Initial results 2022, Final 2024	Real time financial forecasting competition consisting of 50 S&P500 US stocks and of 50 international ETFs	All major forecasting methods, including Machine and Deep Learning, and Statistical ones will be tested	

https://en.wikipedia.org/wiki/Makridakis_Competitions

The Makridakis 5 Competition (M5 - Accuracy)



https://www.kaggle.com/competitions/m5-forecasting-accuracy

The M5 Competition

View PDF   Download full issue

ELSEVIER

International Journal of Forecasting
Volume 38, Issue 4, October–December 2022, Pages 1346-1364

M5 accuracy competition: Results, findings, and conclusions

Spyros Makridakis [b], Evangelos Spiliotis [a], Vassilios Assimakopoulos [a]

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.ijforecast.2021.11.013   Get rights and content

M5 Competition

Training Code of the Winning Solutions

Preprocessing Code of the Winning Solutions

M5 accuracy competition: Results, findings, and conclusions - ScienceDirect

M5 Forecasting- Accuracy. Forecasting is done using Xgboost… | by Jaswanth Badvelu | Towards Data Science

The M6 Competition

MOFC
Learn. Forecast. Compete. Disseminate. Excel.

ABOUT M COMPETITIONS M CONFERENCES M PUBLICATIONS M COURSES CONTACT BECOME MOFC MEMBER

6 Competition | M6 Prizes | M6 Guidelines | Participate in the M6 | Ten hypotheses | Forum

M6 forecasting competition: The Ten Hypotheses / Predictions

Spyros Makridakis (University of Nicosia), Anil Gaba (INSEAD), Ross Hollyman (University of Bath), Fotios Petropoulos (University of Bath), Evangelos Spiliotis (NTUA), Norman Swanson (Rutgers University)

1. The efficient market hypothesis will hold for the great majority of teams but this will not be the case for the top-performing ones.
2. Team rankings based on information ratios will be different from rankings based on portfolio returns or rankings based on the volatility of portfolio returns.
3. There will be a weak link between the ability of teams to accurately forecast individual rankings of assets and risk adjusted returns on investment. The magnitude of this link will increase in tandem with team rankings, on average. Additionally, team portfolios will in general be more concentrated and risky than can be theoretically justified given the accuracy of their forecasts.

https://mofc.unic.ac.cy/ten-hypotheses

The M6 Financial Forecasting Competition

The efficient market hypothesis (EMH) posits that share prices reflect all relevant information, which implies that consistent outperformance of the market is not feasible. The EMH is supported by empirical evidence, including the yearly "Active/Passive Barometer" Morningstar study which regularly finds that active, professional investment managers do not beat, on average, random stock selections. On the other hand, legendary investors like Warren Buffett, Peter Lynch and George Soros, among others, as well as celebrated firms including Blackstone, Bridgewater Associates, Renaissance Technologies, DE Shaw and many others have achieved phenomenal results over long periods of time, amassing returns impossible to justify by mere chance, and casting doubts about the validity of the EMH. It is the express purpose of the M6 competition to empirically investigate this paradox and to shed new light on the EMH by explaining the poor performance of active funds, as well as the exceptional performance of the likes of Warren Buffet, whose fund has achieved an average annual return of 20.0% since 1965, almost double that of S&P500's 10.2% annual gain during that period.

https://m6competition.com/

The M6 Competition

MOFC
Learn. Forecast. Compete. Disseminate. Excel.

7. Averaging forecast rankings (investment weights) across all teams for each asset will yield rankings (weights) that outperform those of the majority of the teams, except in cases where the very worst teams are removed from the average.
8. Teams that employ consistent strategies throughout the competition will perform better than those that change their strategies significantly from one submission point to another.
9. Submissions based on pure judgment or that rely heavily on judgment will perform worse than those based on data-driven methods, on average.
10. The top-performing teams in the forecasting challenge will employ more sophisticated methods compared to the top-performing teams in the investment challenge.

https://mofc.unic.ac.cy/ten-hypotheses/

The M6 Competition

M6 Faq Leaderboard Asset Prices My Team Hello antonio.rueda.toicen@gmail.com! Logout

Leaderboard

Pilot | Global | Quarter | Month

Position	Team	Overall Rank (OR)	Performance of Forecasts (RPS)	Rank (Forecasts)	Performance of Decisions (IR)
1	cd597d34 StanekF_STU (CERGE-EI)	5	0.15689	4	13.79121
2	38c7fc7b MP - Miguel Pérez Michaus	12	0.15648	2	3.71755
3	08986844 Peters_STU	14.5	0.15795	11	6.08679
4	b17cd734 OPPO XZ Lab	21.5	0.15970	30	6.78628
5	db21f28b 21	22	0.15729	8	1.30108

https://m6competition.com/Leaderboard

The M6 Competition

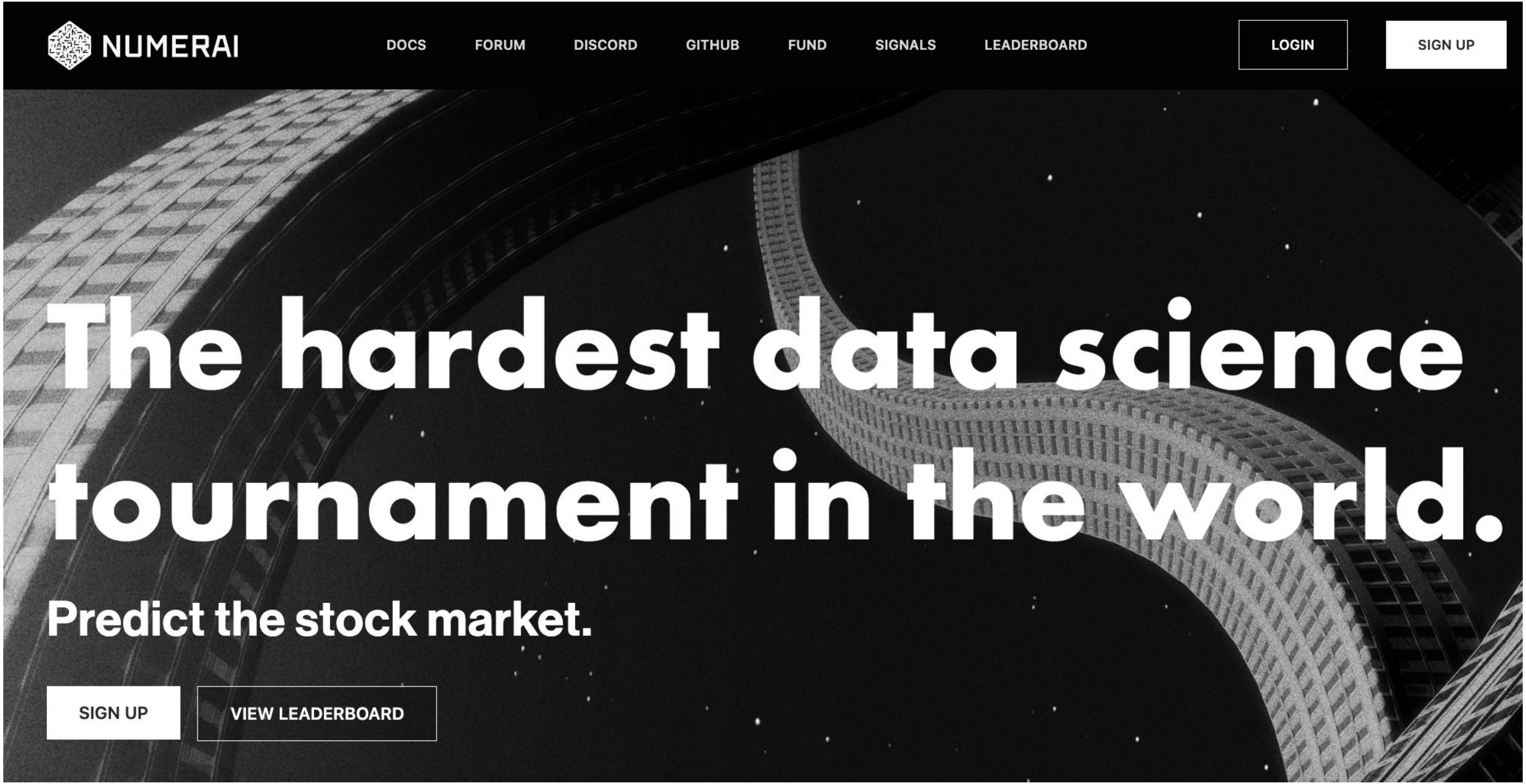
M6 Faq Leaderboard Asset Prices My Team Hello antonio.rueda.toicen@gmail.com! Logout

Leaderboard

Pilot | Global | Quarter | Month

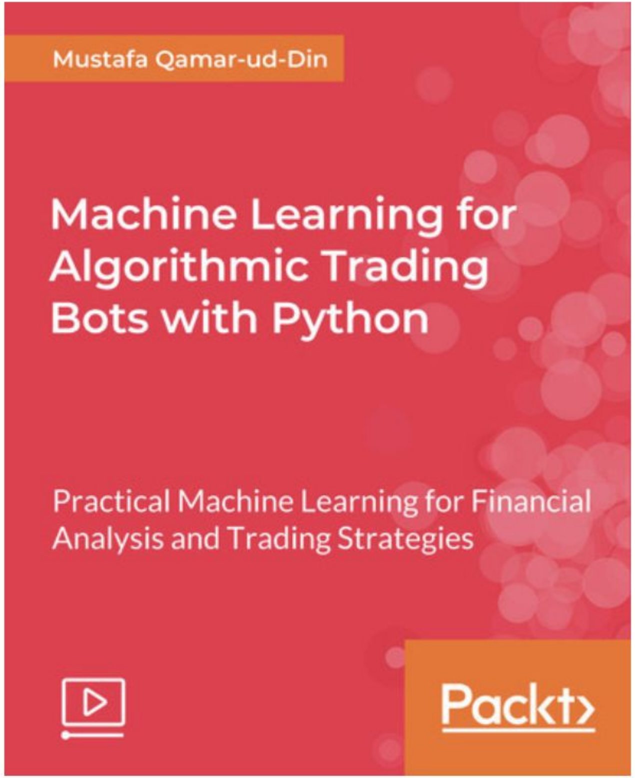
Position	Team	Overall Rank (OR)	Performance of Forecasts (RPS)	Rank (Forecasts)	Performance of Decisions (IR)
20	bbb3c357 Galloping Gargoyles	46.5	0.16389	92	32.88981
43	f8fbd1bf Selamlar H.T.(Ataforecasting)	56	0.17028	110	28.65539
48	ccb2bfb9 Quantmetry	58	0.17089	113	24.72015
15	9b7bb959 LG AI Research	41.5	0.16122	79	21.53685
82	872bdbc9 Tilefish Poele	81.5	0.30977	158	17.94236

https://m6competition.com/Leaderboard



https://numer.ai/

Time to make money!1!(?)



TIME TO COMPLETE:
4h 50m

LEVEL:
Beginner to intermediate

TOPICS:
Python

PUBLISHED BY:
Packt Publishing

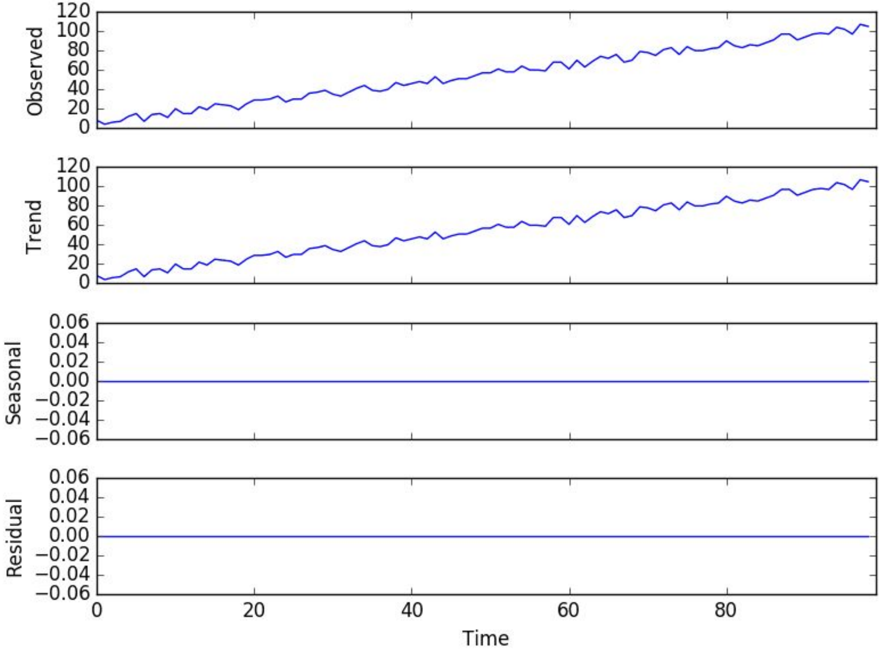
PUBLICATION DATE:
February 2019

How should we approach time series problems in 2023?

- Try simple baselines first
  - Predict the next day as the current one
  - Moving averages
  - Exponential smoothing
- Try many models at once
  - Try libraries like PyCaret and Darts first to get quick comparisons
- Only focus on classical methods like ARIMA when having a strong reason
  - Having a stubborn manager that learned ARIMA or Markov models at school
  - Most classical models are not more explainable than a gradient boosting algorithm

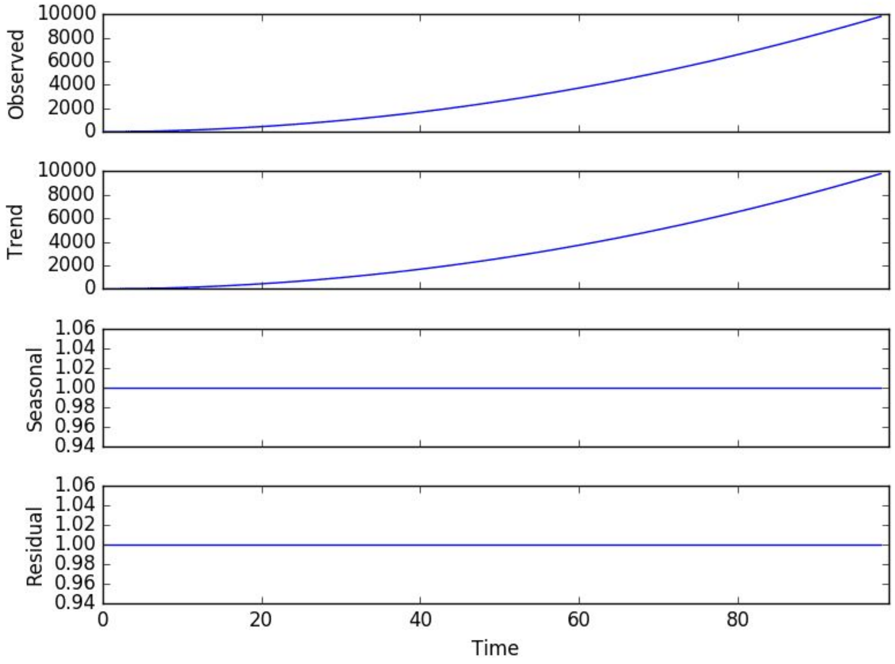
Decomposing a Time Series

[How to Decompose Time Series Data into Trend and Seasonality]



Decomposing a Time Series

[How to Decompose Time Series Data into Trend and Seasonality]



Decomposing a Time Series

How to Decompose Time Series Data into Trend and Seasonality

Colab Notebook

Exercise: change the time series from multiplicative to additive

Read this critique to the seasonal decomposition approaches

Think: how would you choose between an additive or multiplicative model?

Metrics to Measure Accuracy

MAE - targets median of the true value distribution

RMSE - targets mean of the true value distribution, sensitive to outliers

MAPE - percentage of error, popular but weird when actual = 0 or forecast > actual

SMAPE - same error when forecast < actual and viceversa [read]

Notebook 1

Notebook 2

Metrics

```
errors = forecasts - actual

mse = np.square(errors).mean()

rmse = np.sqrt(mse)

mae = np.abs(errors).mean()

mape = np.abs(errors / x_valid).mean()
```

All collaborative articles / Business Administration / **Budgeting & Forecasting**

What are the pros and cons of using MAPE (mean absolute percentage error) as a forecast accuracy metric?

Powered by AI and the LinkedIn community

If you are involved in budgeting and forecasting, you probably know how important it is to measure the accuracy of your forecasts. But how do you choose the best metric to evaluate your performance and identify areas for improvement? One common option is MAPE, or mean absolute percentage error, which calculates the average of the absolute values of the percentage errors between the actual and forecasted values. However, MAPE is not a perfect solution and has some advantages and disadvantages that you should be aware of. In this article, we will discuss the pros and cons of using MAPE as a forecast accuracy metric and how to use it effectively.

[What are the pros and cons of using MAPE (mean absolute percentage error) as a forecast accuracy metric?]

Supply Chain, Thought Leadership

Mean Absolute Percentage Error (MAPE) Has Served Its Duty and Should Now Retire

Malte Tichy, August 4, 2022 | 17 min read

Executive summary

According to Gartner (2018 Gartner Sales & Operations Planning Success Survey), the most popular evaluation metric for forecasts in Sales and Operations Planning is Mean Absolute Percentage Error (MAPE). This needs to change. Modern forecasts concern small quantities on a disaggregated level such as product-location-day. For such granular forecasts, MAPE values are extremely hard to judge and thereby disqualify as useful forecast quality indicators. MAPE also deeply misleads users by both exaggerating some problems and disguising others, nudging them to choose forecasts with systematic bias. The situations in which MAPE is suitable become increasingly rare. This is not dry theory: We simulate a supermarket that relies on a MAPE-optimizing forecast value fed into replenishment. The under- and overstocks in the fast- and slow-sellers quickly push the store out of business.

When absolute and relative errors contradict — whom should we trust?

[MAPE has served its duty]

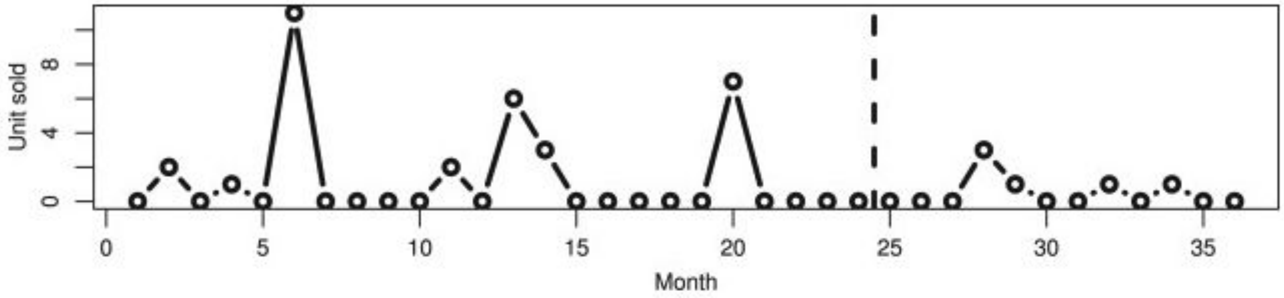
The mean absolute percentage error (MAPE) is one of the most popular measures of the forecast accuracy. It is recommended in most textbooks (e.g., Bowerman et al., 2004, Hanke and Reitsch, 1995), and was used as the primary measure in the M-competition (Makridakis et al., 1982). MAPE is the average of absolute percentage errors (APE). Let A_t and F_t denote the actual and forecast values at data point t, respectively. Then, MAPE is defined as:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - F_t}{A_t}\right|, \tag{1.1}$$

where N is the number of data points. To be more rigorous, Eq. (1.1) should be multiplied by 100, but this is omitted in this paper for ease of presentation without loss of generality. MAPE is scale-independent and easy to interpret, which makes it popular with industry practitioners (Byrne, 2012).

However, MAPE has a significant disadvantage: it produces infinite or undefined values when the actual values are zero or close to zero, which is a common occurrence in some fields. If the actual values are very small (usually less than one), MAPE yields extremely large percentage errors (outliers), while zero actual values result in infinite MAPEs. In practice, data with numerous zero values are observed in various areas, such as retailing, biology, and finance, among others. For the area of retailing, Fig. 1 (Makridakis, Wheelwright, & Hyndman, 1998) illustrates typical intermittent sales data. Many zero sales occur during the time periods considered, and this leads to infinite or undefined MAPEs.

https://www.sciencedirect.com/science/article/pii/S0169207016000121



Download : Download high-res image (133KB)

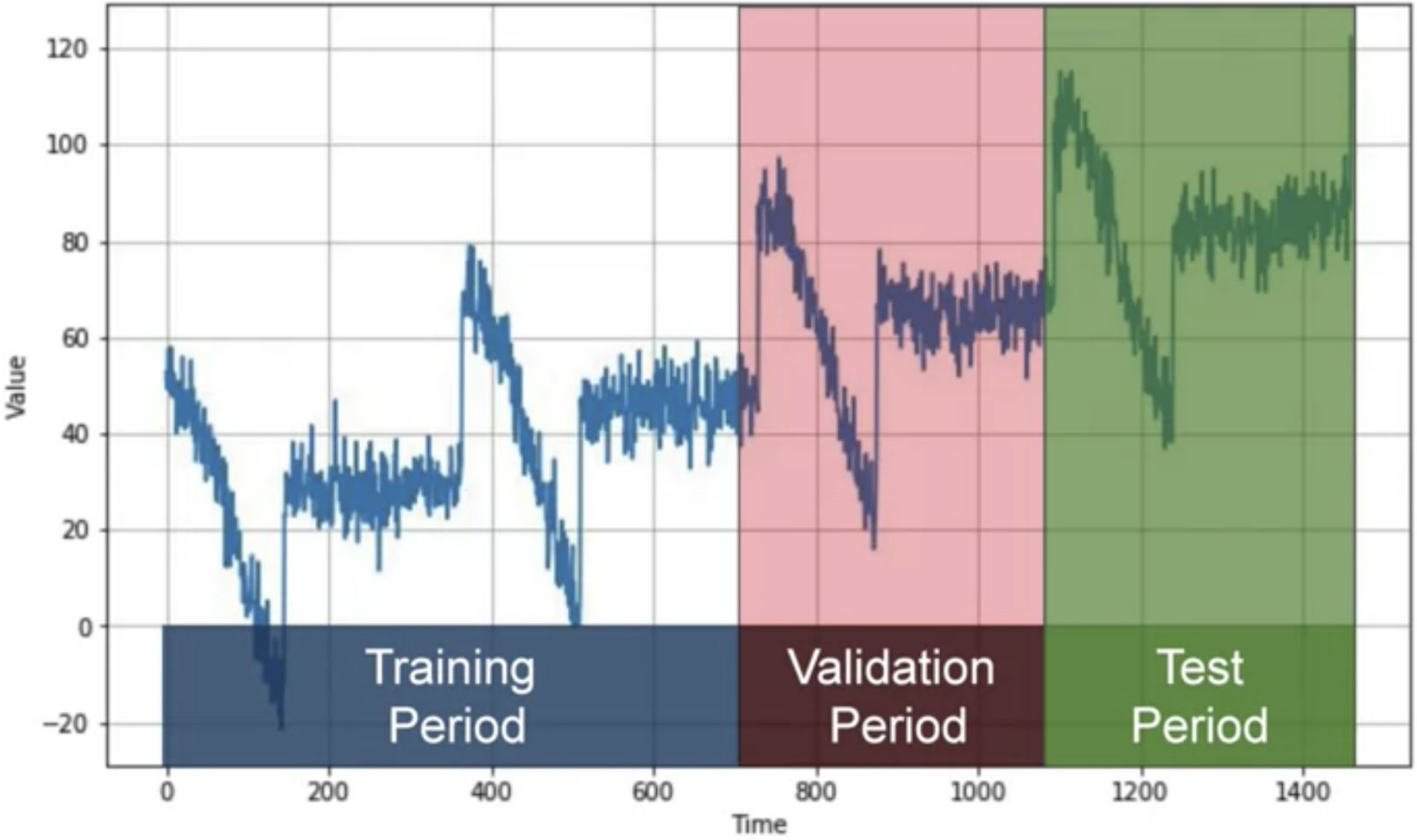
Download : Download full-size image

Fig. 1. Three years of monthly sales of a lubricant product sold in large containers. Data source: 'Product C' from Makridakis et al. (1998, Ch. 1). The vertical dashed line indicates the end of the data used for fitting and the start of the data used for out-of-sample forecasting.

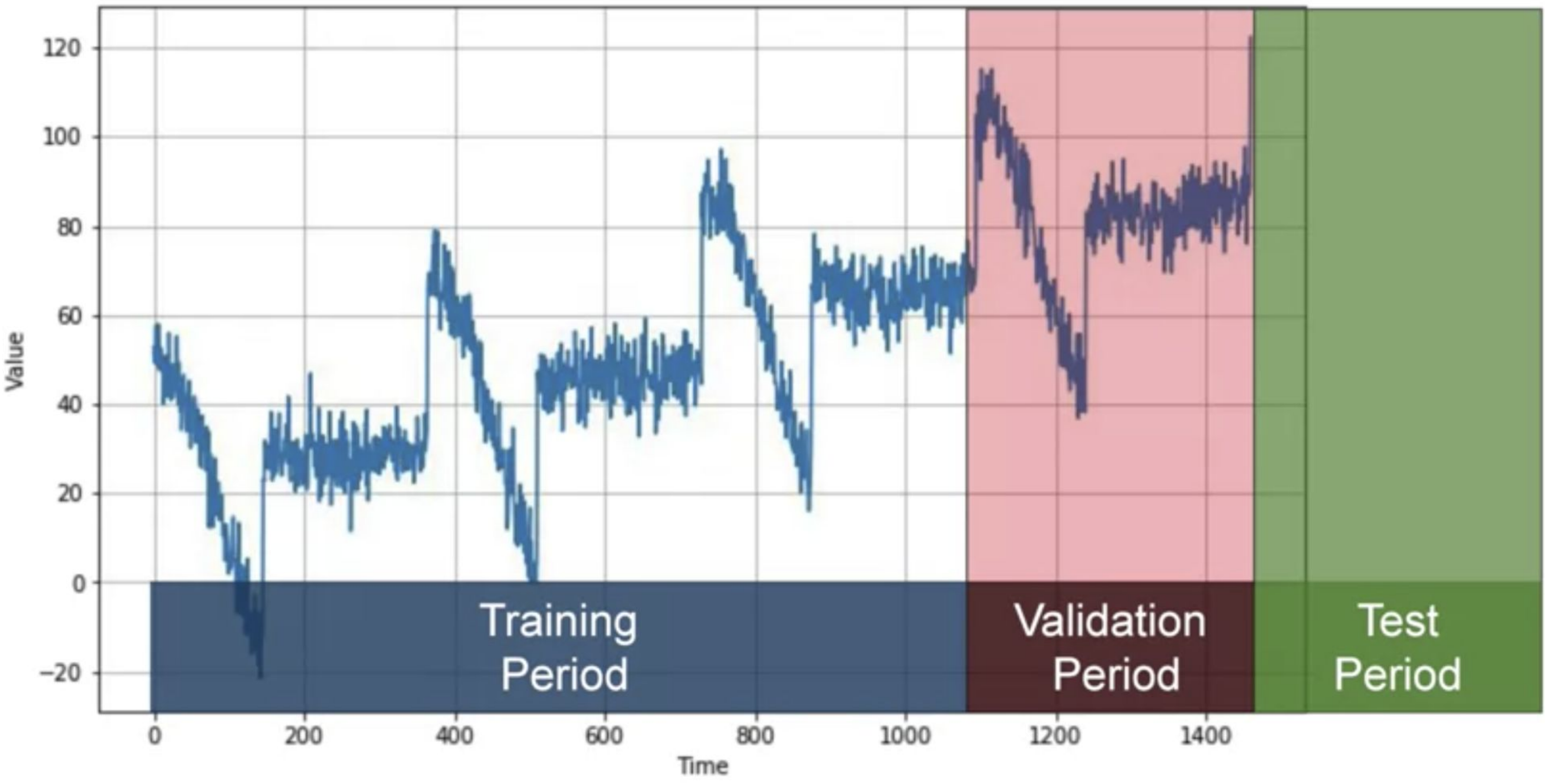
https://www.sciencedirect.com/science/article/pii/S0169207016000121

Winsorizing

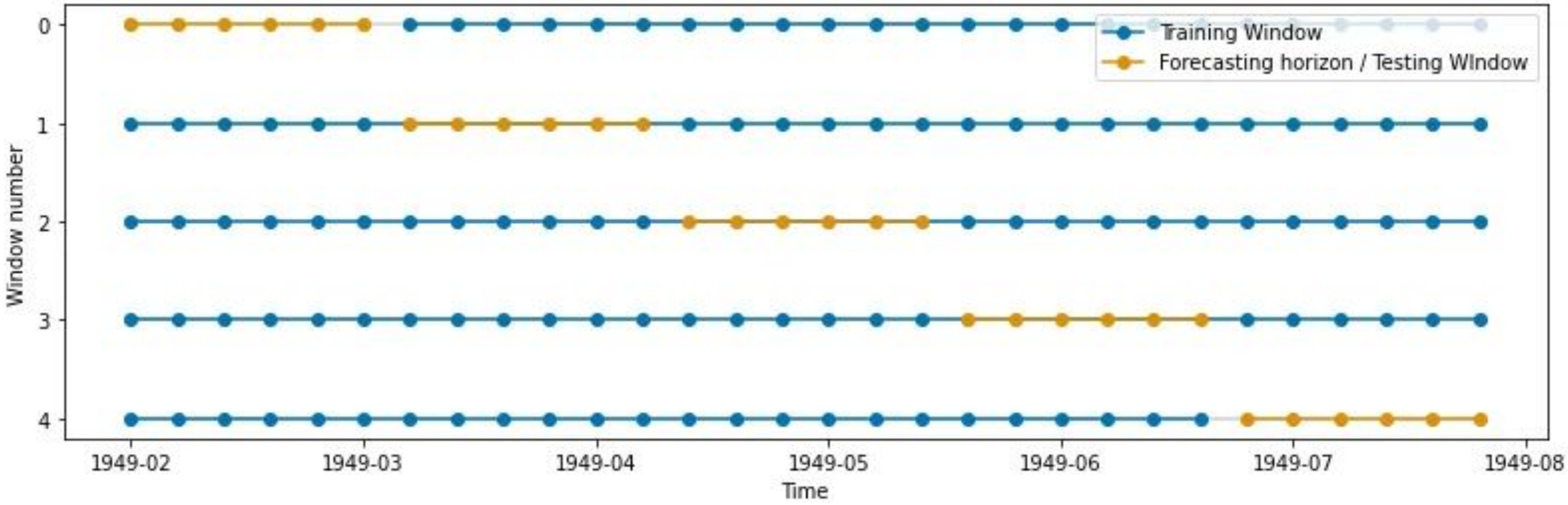
Fixed Partitioning



Fixed Partitioning

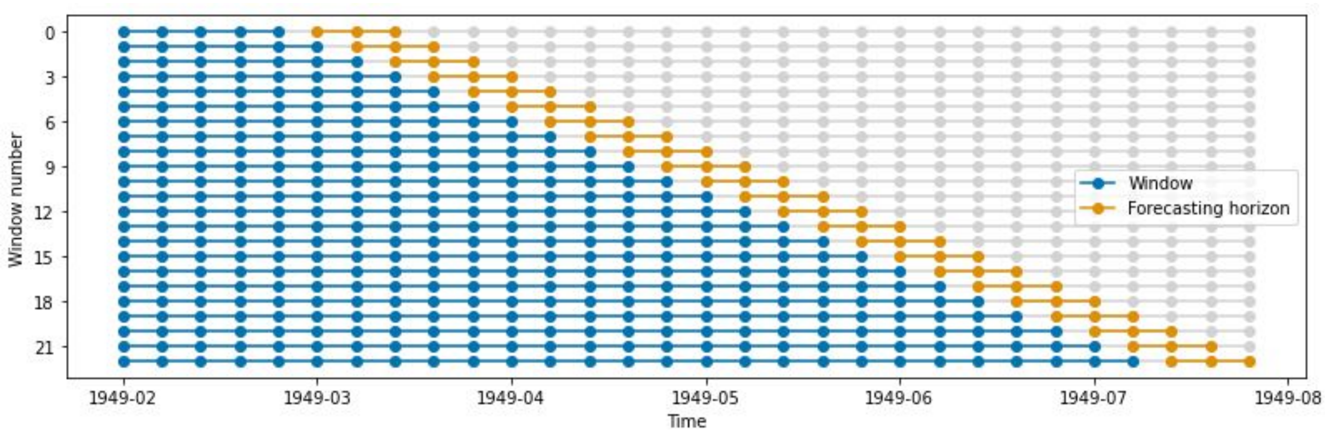


Cross-validation in time series



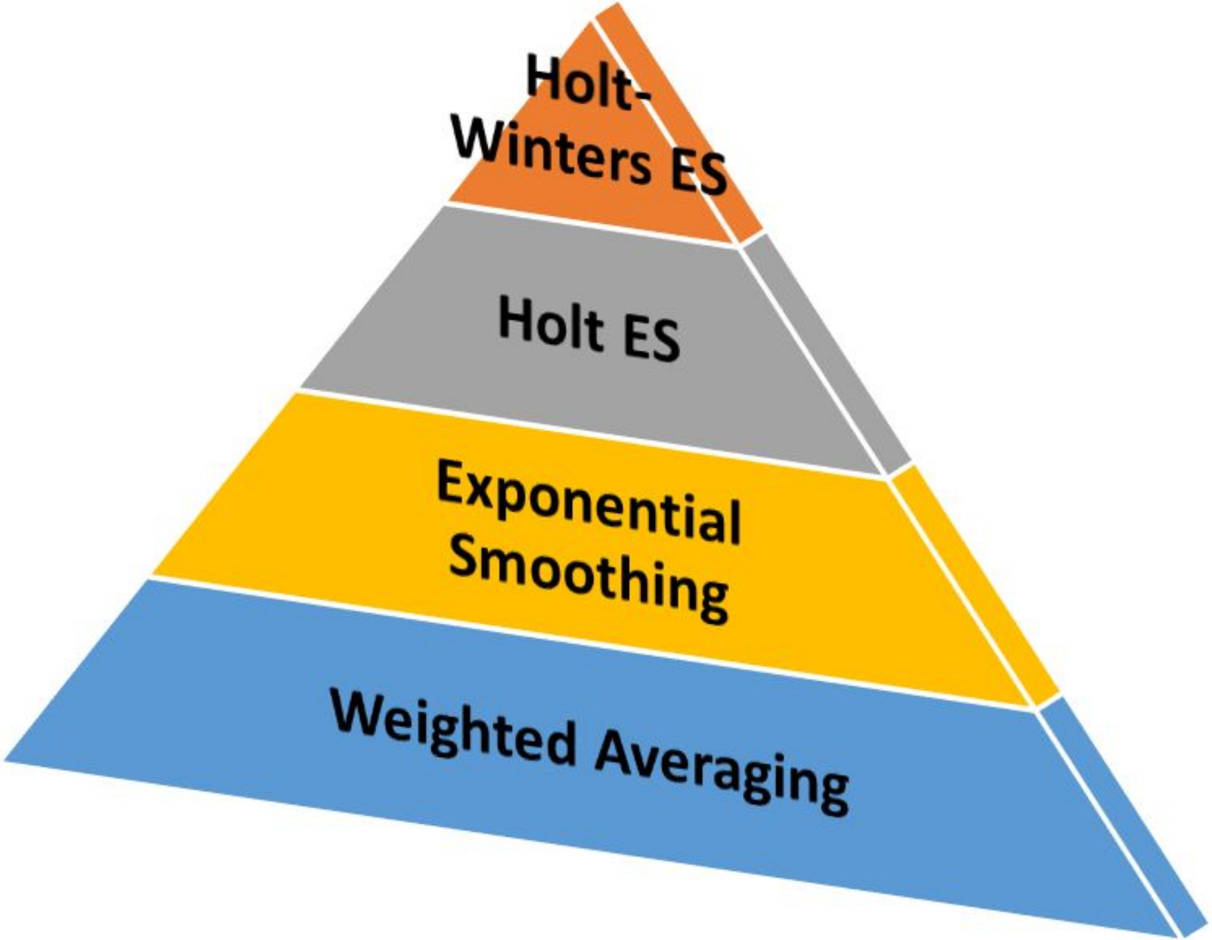
Does this look correct?

Cross-validation in time series



Train with [A]	*test with [B]*
Train with [A B]	*test with [C]*
Train with [A B C]	*test with [D]*
Train with [A B C D]	*test with [E]*
Train with [A B C D E]	*test with [F]*

[Don't Use Simple K-Fold Cross Validation on Time Series Data]


Holt-Winters ES
Holt ES
Exponential Smoothing
Weighted Averaging

Exponential Smoothing

- [Let’s ask ChatGPT about it…]()

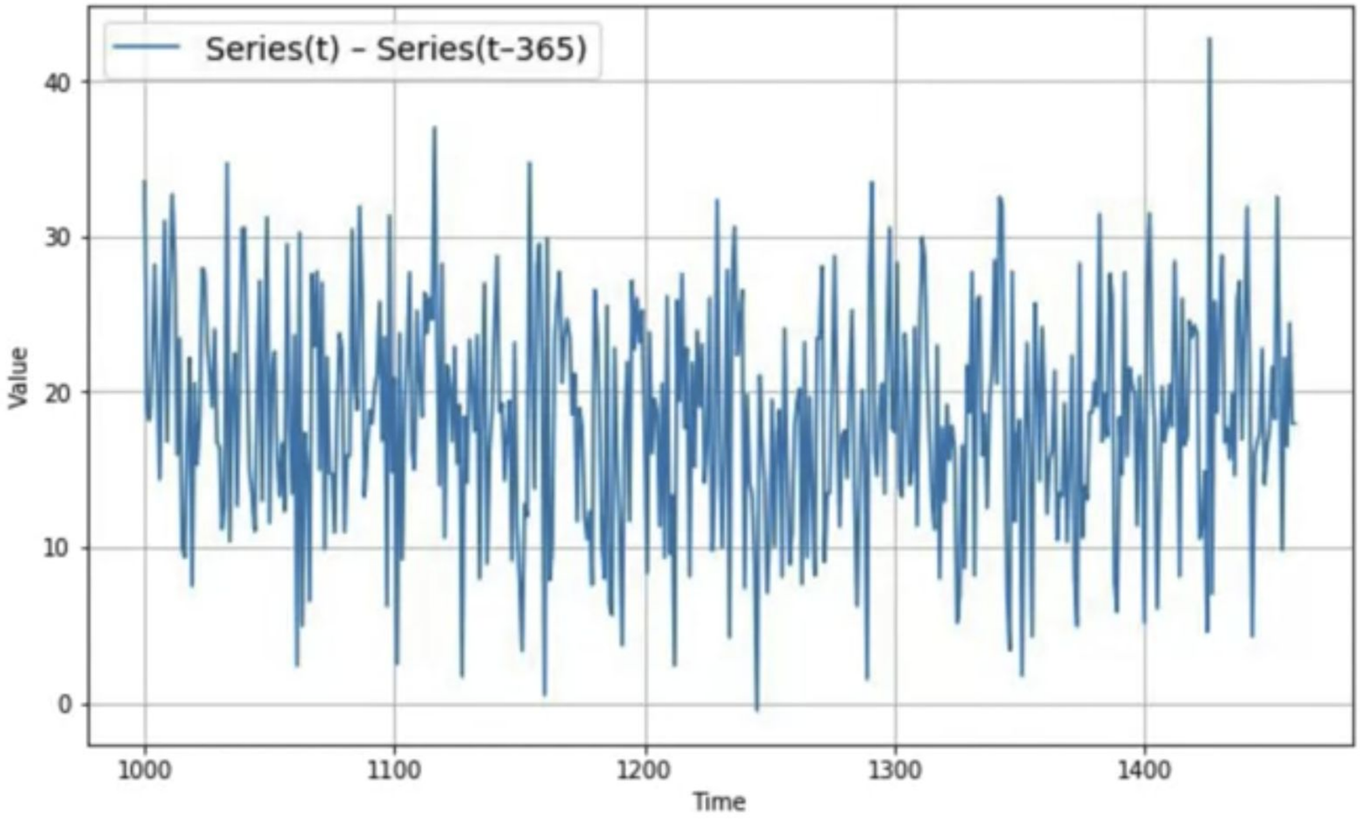
- [And then let’s ask about Holt-Winters specifically]()

- **95% of the teams that participated in the M5 Competition couldn’t beat the Holt-Winters baseline model**

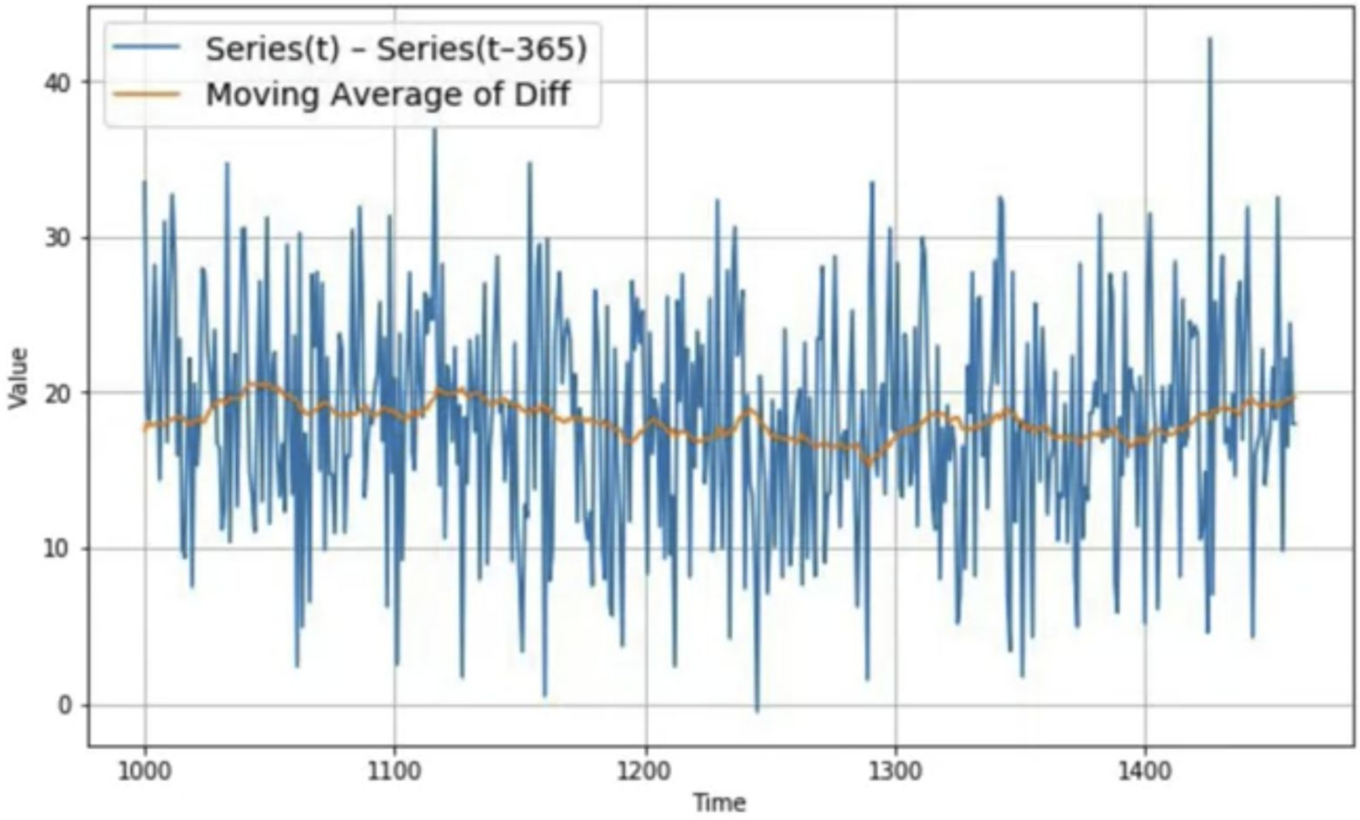
ARIMA

- [And now let’s do the same with ARIMA]()

Differencing



Moving Average on Differenced Time Series



Baseline models

- Always compare against a model that predicts t as being equal to t -1

Rolling mean and variance

https://campus.datacamp.com/courses/visualizing-time-series-data-in-python/summary-statistics-and-diagnostics?ex=5

Feature engineering

https://chat.openai.com/share/6a8e4311-73b4-48cb-a28f-882ab30296a8

Home Page

Papers

Submissions

News

Editorial Board

Special Issues

Open Source Software

Darts: User-Friendly Modern Machine Learning for Time Series

Julien Herzen, Francesco Lässig, Samuele Giuliano Piazzetta, Thomas Neuer, Léo Tafti, Guillaume Raille, Tomas Van Pottelbergh, Marek Pasieka, Andrzej Skrodzki, Nicolas Huguenin, Maxime Dumonal, Jan Kościsz, Dennis Bader, Frédérick Gusset, Mounir Benheddi, Camila Williamson, Michal Kosinski, Matej Petrik, Gaël Grosch; 23(124):1−6, 2022.

Abstract

We present Darts, a Python machine learning library for time series, with a focus on forecasting. Darts offers a variety of models, from classics such as ARIMA to state-of-the-art deep neural networks. The emphasis of the library is on offering modern machine learning functionalities, such as supporting multidimensional series, fitting models on multiple series, training on large datasets, incorporating external data, ensembling models, and providing a rich support for probabilistic forecasting. At the same time, great care goes into the API design to make it user-friendly and easy to use. For instance, all models can be used using fit()/predict(), similar to scikit-learn.

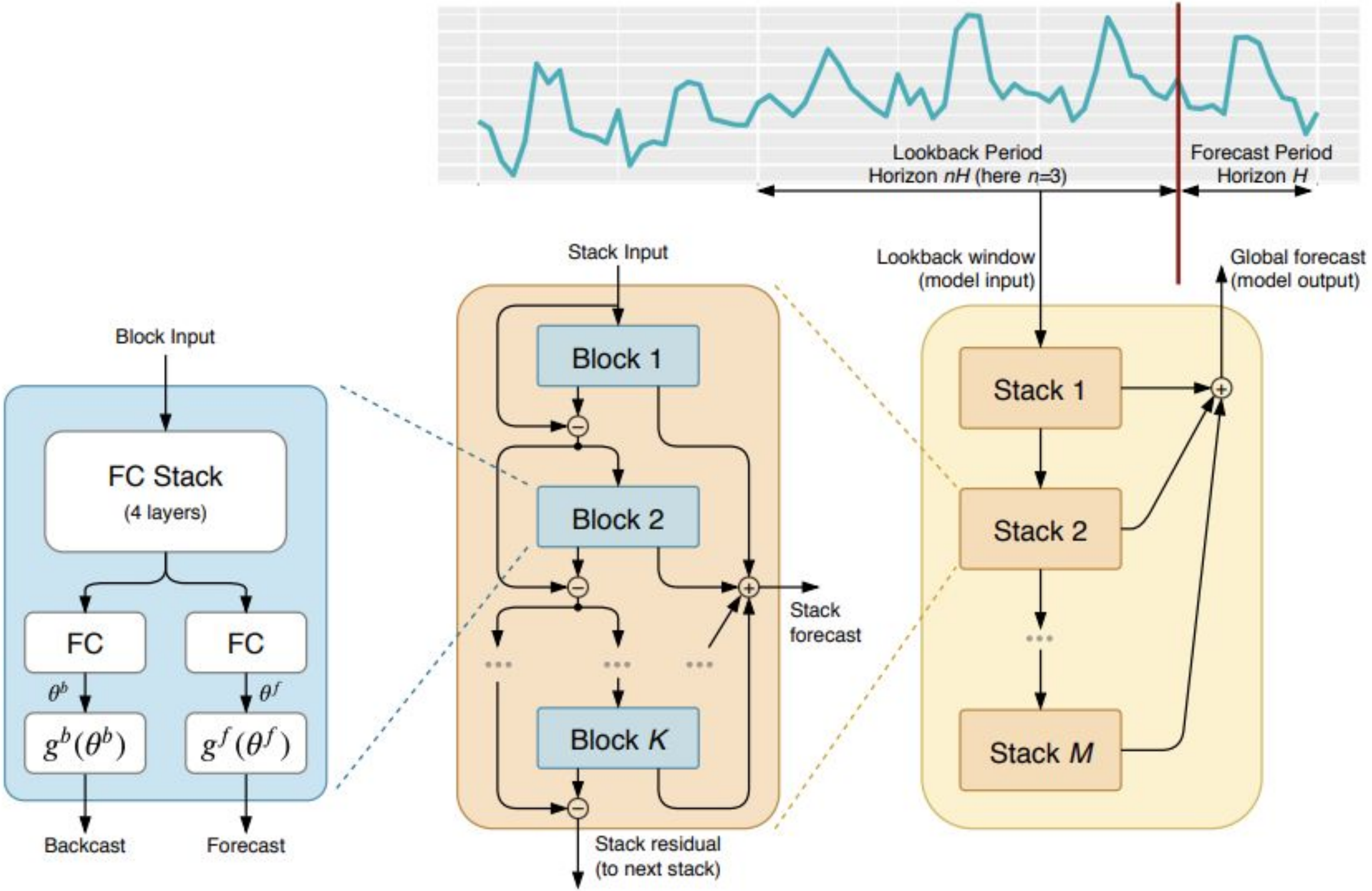
[abs][pdf][bib]   [code]

Time Series with Transformers

- Surprisingly not a thing yet
- https://huggingface.co/blog/time-series-transformers
- https://colab.research.google.com/drive/1ZV0R9XCdkxsjtCEr3Sw5Cu3sA5xITh3b

N-Beats - A Neural Network for Time Series Prediction

- https://arxiv.org/pdf/1905.10437.pdf
- Implemented in Darts



Lookback Period
Horizon nH (here n=3)
Forecast Period
Horizon H
Stack Input
Lookback window
(model input)
Global forecast
(model output)
Block Input
FC Stack
(4 layers)
FC
FC
θ^b
θ^f
$g^b(\theta^b)$
$g^f(\theta^f)$
Backcast
Forecast
Block 1
Block 2
Block K
Stack forecast
Stack residual
(to next stack)
Stack 1
Stack 2
Stack M

DARTS

Darts

Review questions

- What is seasonality?
- What is trend?
- What is autocorrelation? What is partial autocorrelation?
- What is noise?
- What is a moving average?
- What is a naive (aka “null”) prediction in the context of time series?
- Is a sound wave a good example of time series data?
- What is a non-stationary time series?
- What are the practical considerations of using a trailing moving average vs a centered moving average to smoothen a time series?
- What is a window size? How does it relate to the target?
- When we do seasonal_decompose in statsmodels, what is the name of the plot that shows noise?

Review questions

- What is differencing?
- What does it mean to detrend a time series?
- What does it mean for a time series to be stationary?
- What are the weaknesses of the ARIMA model?
- What is a theta model?
- What is a hierarchical time series?
- How does exponential smoothing work?
- How does the Holt-Winters model work?
- When deciding on a simple benchmark, should you first choose ARIMA or Exponential Smoothing?
- Which model was used by most of the winners of the M5 competition?
- What are the considerations of using MAE vs RMSE as a forecast KPI?

Review questions

- Is N-BEATS an LSTM or a stack of fully connected networks?
- How does the expanding window approach differ from regular k-fold cross validation?
- What is winsorizing?
- What are the drawbacks of using MAPE as an evaluation metric?
- When should we use MAE? When should we use RMSE?

Resources

- Feature Engineering for Time Series Forecasting - PyData 2022
- Modern Time Series Analysis - 2019 SciPy Tutorial
- Time Series with Konrad (discussion and tutorial from two Kaggle Grandmasters)
- Time series anomaly detection with PyCaret
- Gradient boosting review

Recommended videos

[LightGBM explained - PyData 2022]

[Forecasting with the FB Prophet Model]

[Time Series Forecasting with XGBoost]

[Time Series Forecasting with XGBoost - Advanced Methods]