



ADVANCED BUSINESS FORECASTING MET PYCARET (AUTOML)

Verbeter forecasting met advanced forecasting methodes...

...zonder de intimiderende code





TIES DE BONT

Data Analytics Consultant



CONTENT

1. Wat is machine learning, AI, Deep learning?
2. Wat is tijdsreeks forecasting?
3. Wat is (automated) machine learning?
4. Wat is PyCaret?
5. Forecasting met PyCaret
6. Model Development Proces in ML
7. Demo: PyCaret in actie
8. Wrap-up



WAT IS MACHINE LEARNING, AI, DEEP LEARNING?





Playgr

We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published

An artificially intelligent first author presents many ethical questions—and could upend the publishing process

By Almira Osmanovic Thunström

Write

Mars

Merc

Plane

distir

thin a

valley

AUTHOR



Almira Osmanovic Thunström is an organizational developer at the Department of ePsychiatry at Sahlgrenska University Hospital in Sweden and a doctoral researcher at the Institute of Neuroscience and Physiology at Gothenburg University working on research regarding use of artificial intelligence and virtual reality in psychiatry. You can follow her at @augmentedrobot

Submit

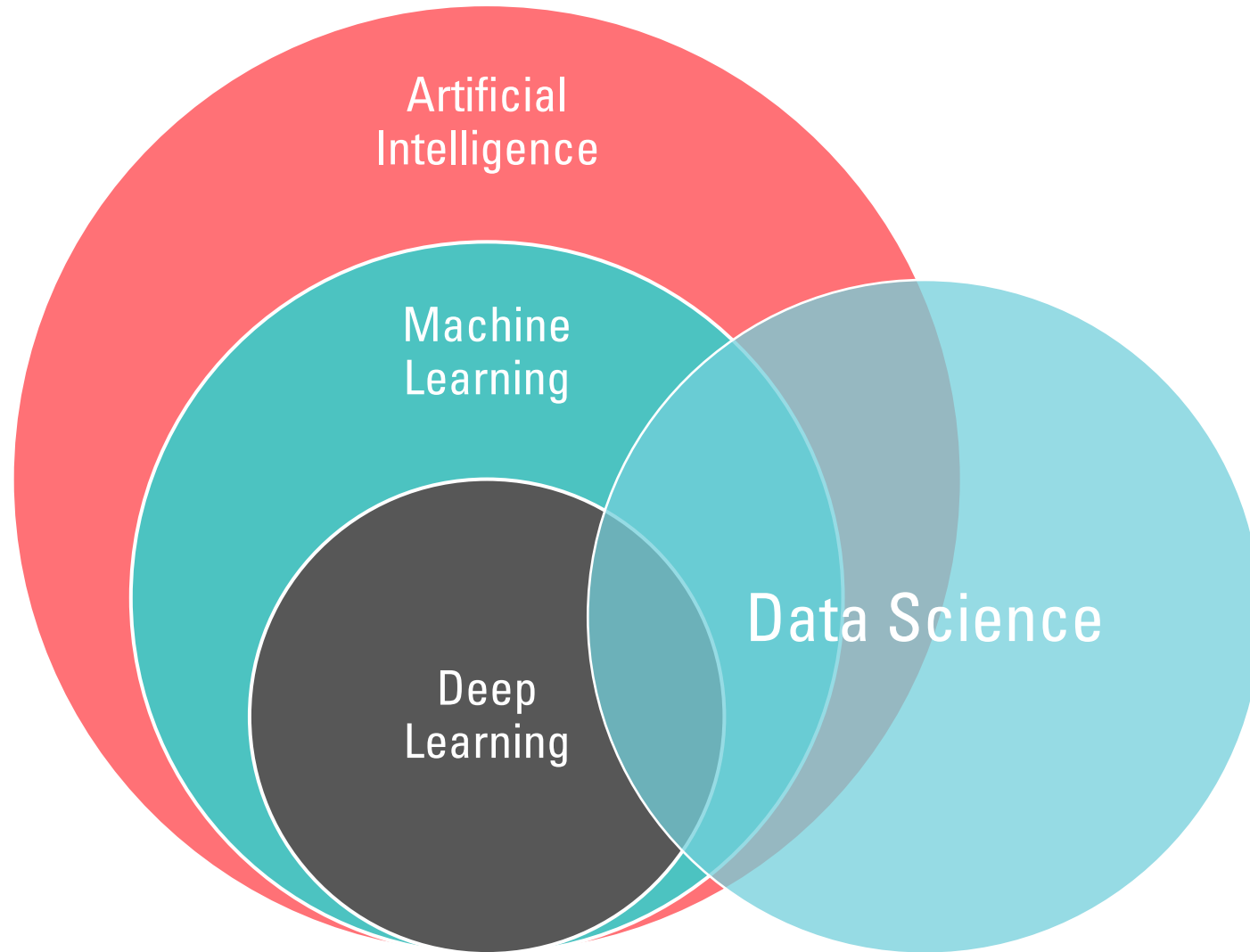
On a rainy afternoon earlier this year, I logged into my OpenAI account and typed a simple instruction for the research company's artificial-intelligence algorithm, GPT-3: *Write an academic thesis in 500 words about GPT-3 and add scientific references and citations inside the text.*

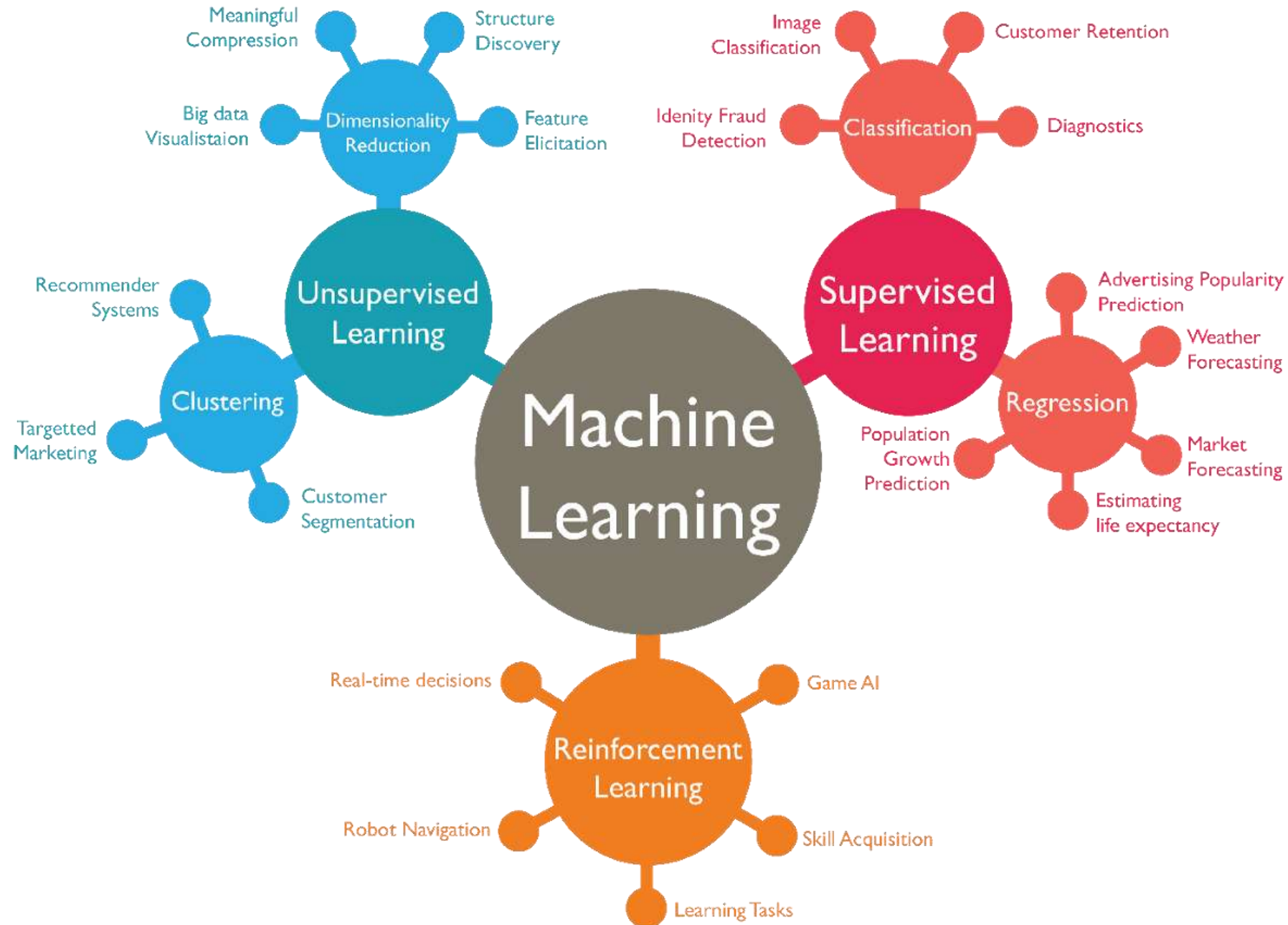
As it started to generate text, I stood in awe. Here was novel content written in academic language, with references cited in the right places and in relation to the right context. It looked like any other introduction to a fairly



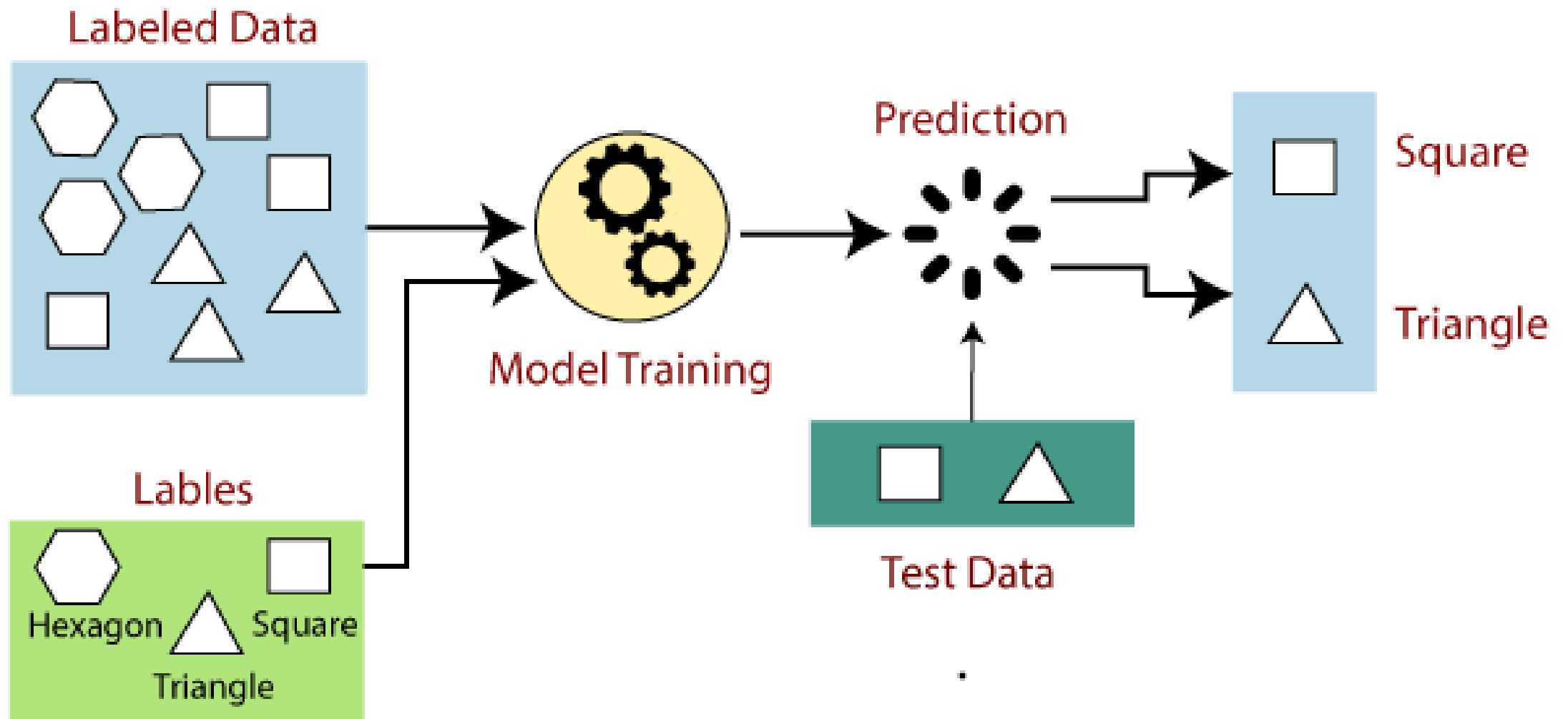
DALL-E 2







SUPERVISED LEARNING



—

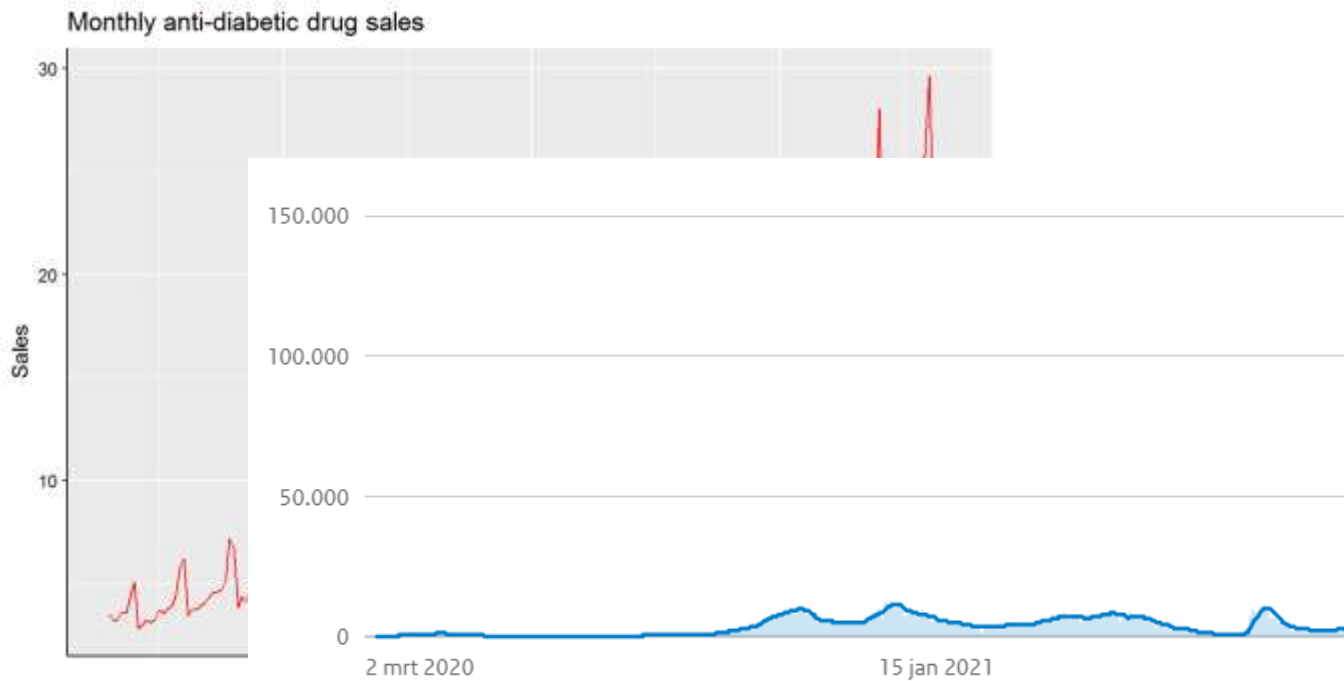
—

WAT IS TIJDSREEKS FORECASTING?

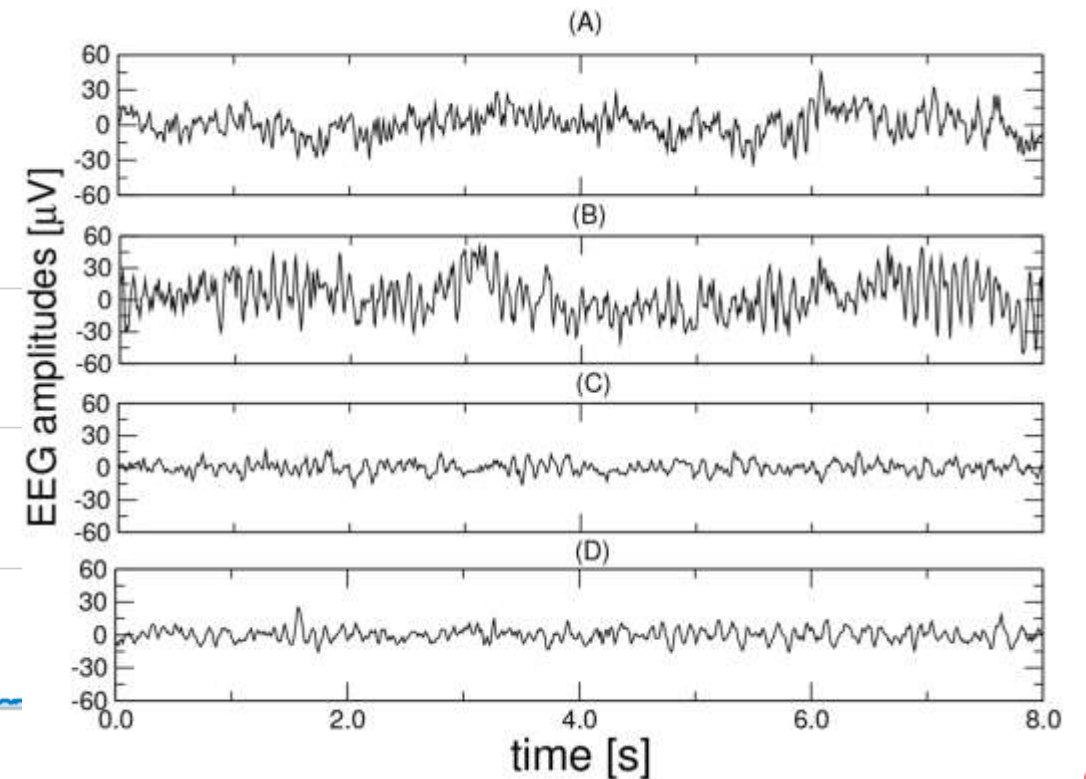


WAT IS EEN TIJDSREEKS?

- Een **tijdsreeks** is een set aan **data gespreid over een tijd-as**. Denk aan:
 - Aantal sales diabetes medicijn per maand.
 - Aantal COVID-19 gevallen per dag.
 - EEG signaal per milliseconde.



Source: <https://rpubs.com/JMFlin/246488>

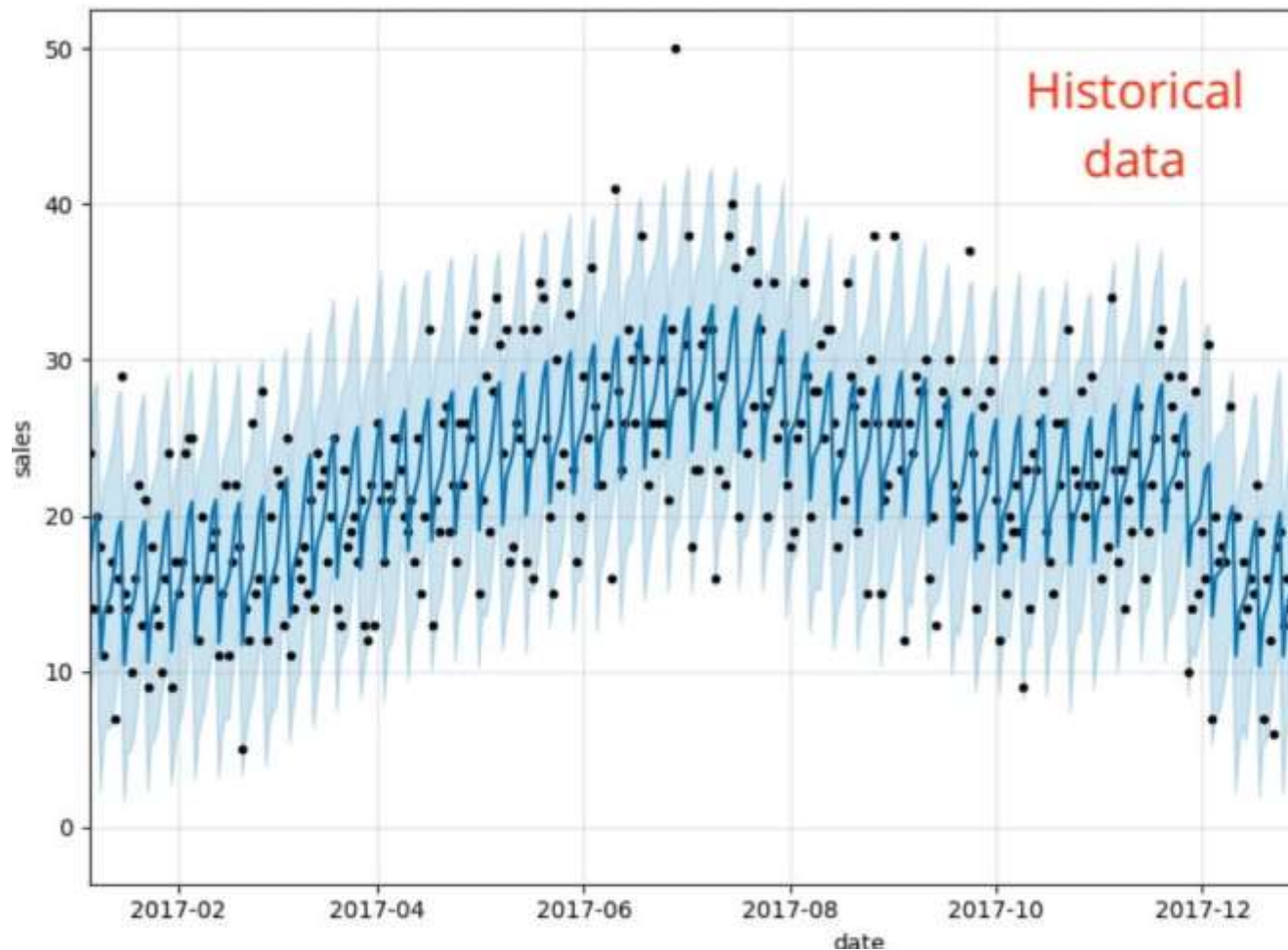


Source: RIVM

Source: <https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0231169.g002>

WAT IS TIJDSREEKS FORECASTING?

- Het **voorspellen** van **toekomstige & onbekende** data punten.



Source: <https://vitalflux.com/different-types-of-time-series-forecasting-models/>



BUSINESS FORECASTING

*“Business forecasting is het proces van **voorspellen van toekomstige ontwikkelingen** in de business **gebaseerd op analyses van trends in de verleden en huidige data.**”*

Twee types:

- **Kwalitatief**
 - Relevante managers voeren **handmatig** verwachte waarden in voor budget forecasts / verwachte product sales / margins
- **Kwantitatief**
 - Forecasts worden **automatisch** gegenereerd op basis van:
 - Generieke **ingebouwde forecasting modellen**, of
 - **advanced forecasting modellen**



BUSINESS FORECASTING

*"Business forecasting is the process of predicting future developments
in business based on analysis of trends in past and present data."*

Task types

- Qualitative
 - Relevant managers **manually input** predicted values for budget forecasts / expected product sales / margins
- Quantitative
 - Forecasts are generated **automatically** by
 - general **built-in forecasting models**, or
 - **advanced forecasting models**

Forecasting Type	Process	Reliability/Accuracy
Kwalitatief	Handmatig	Hoog
Kwantitatief: ingebouwd	Automatisch	Laag
Kwantitatief: Advanced Modellen	Automatisch	Hoog*

* Er van uitgaande dat we data hebben



WAT IS (AUTOMATED) MACHINE LEARNING?



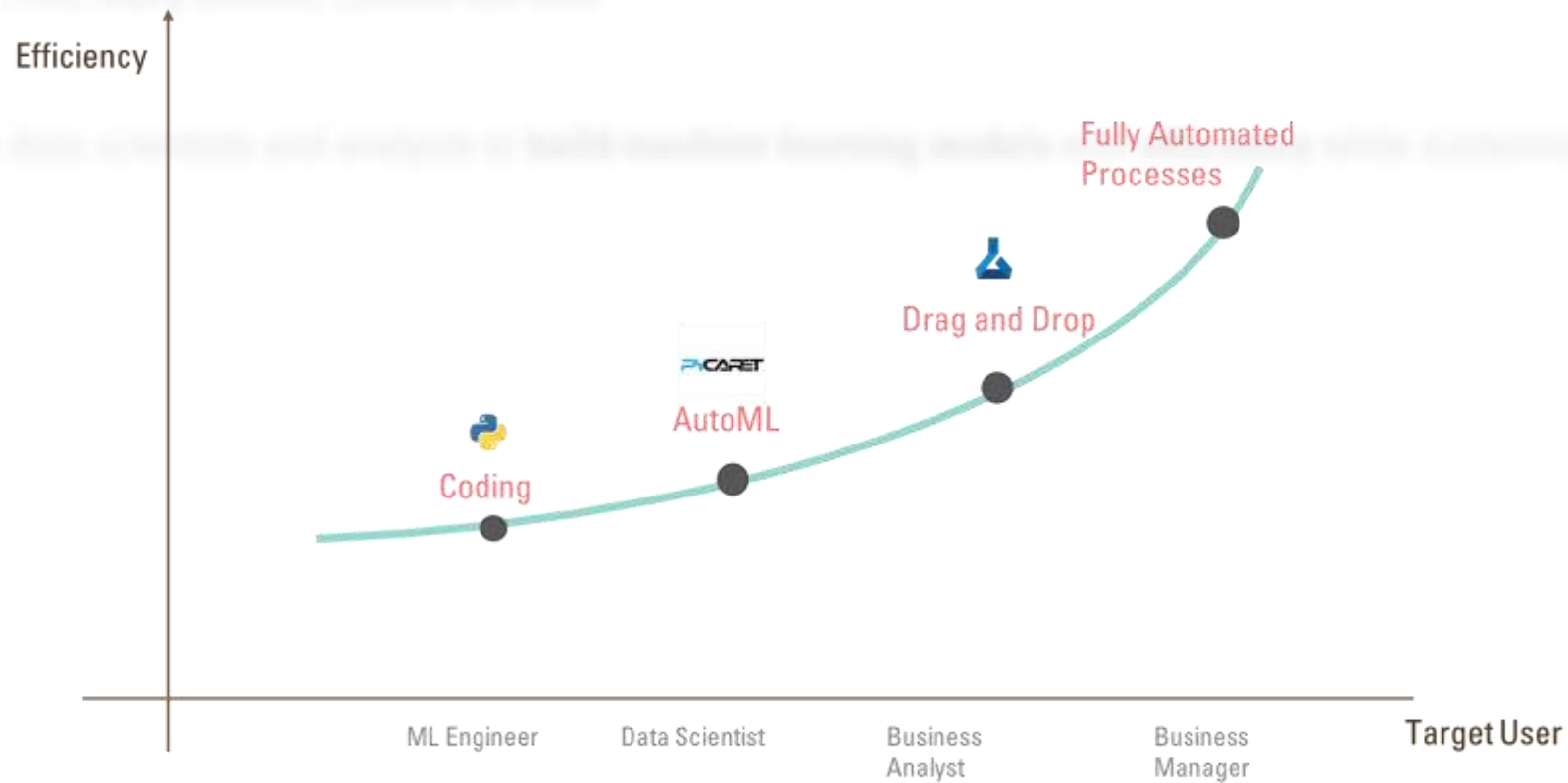
WAT IS (AUTOMATED) MACHINE LEARNING?

- AutoML is het **automatiseren** van het tijdrovende en iteratieve proces van **machine learning**
- Voor verschillende machine learning projecten, presteren andere type modellen het beste
 - Oplossing: train veel verschillende modellen en kies het beste
- AutoML helpt data scientists en analysten om **efficiënt machine learning modellen te bouwen** zonder significant kwaliteitsverlies



WAT IS (AUTOMATED) MACHINE LEARNING?

AutoML is vooral een tool for citizen data scientists



WAT IS PYCARET?



WAT IS PYCARET?



- Python is... een 'general-purpose' programmeer taal
- PyCaret is...

PYCARET IS DEMOCRATIZING MACHINE LEARNING 🙌

PyCaret empowers anyone to build low-code, powerful, end-to-end machine learning solutions. There are tons of ways for you to get started.



Wat het **echt** is...

- een open-source, low-code machine learning library in Python



WAT IS PYCARET?

- Wat kun je met **PyCaret**?
- **Auto ML** voor:



Image source: [Ali, Moez].

FORECASTING MET PYCARET



FORECASTING MET PYCARET

- Forecast wordt gemaakt op basis van data uit het verleden
- Hoe betrouwbaar is onze forecast?
 - Door performance van model te evalueren voor gebruik

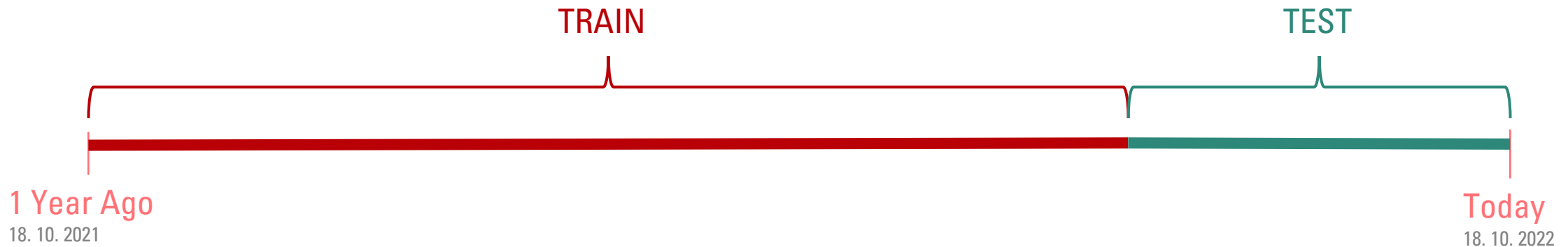
1 Year Ago
18. 10. 2021

Today
18. 10. 2022



FORECASTING WITH PYCARET

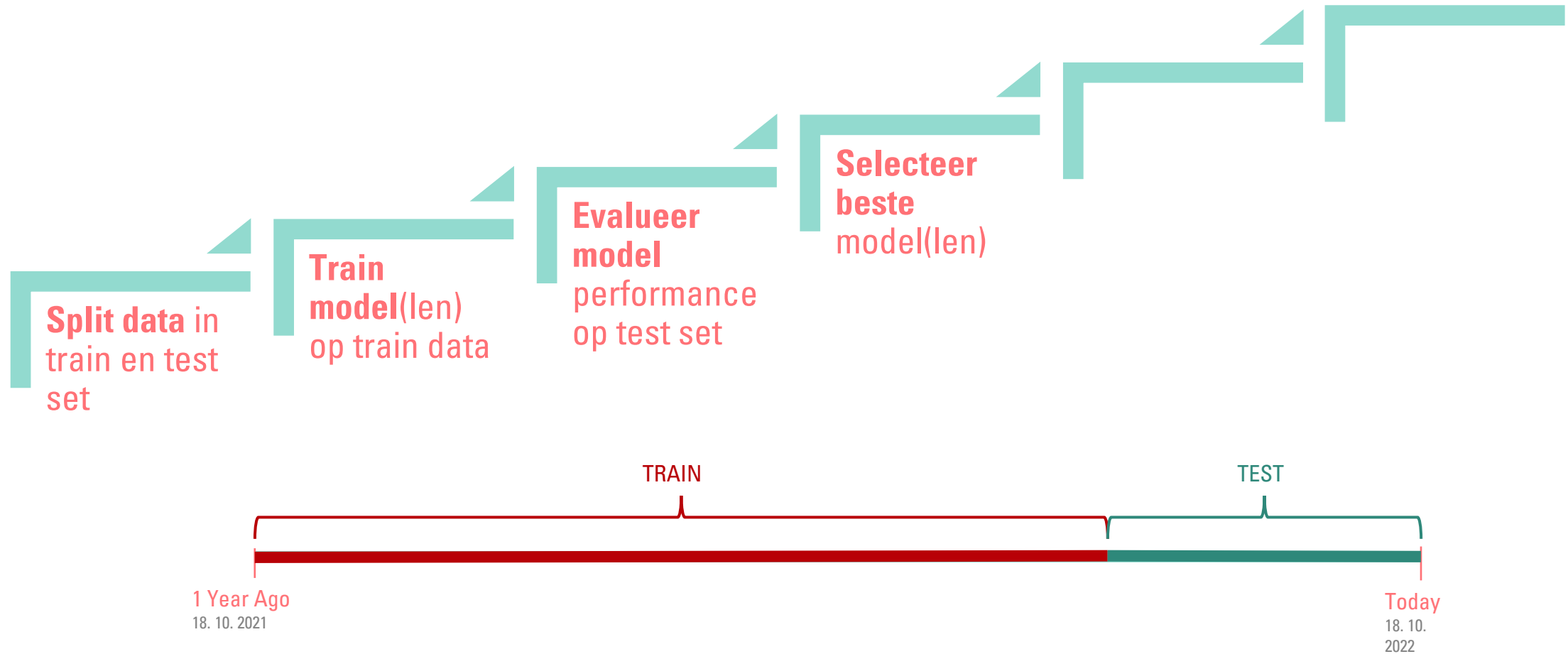
- Forecast wordt gemaakt op basis van data uit het verleden
- Hoe betrouwbaar is onze forecast?
 - Door performance van model te evalueren voor gebruik
- Splits historische data in **train** en **test data**
 - Gebaseerd op de **train** data maken we een **voorspelling**
 - **Vergelijk de voorspelling** met de **test** data, zodat we weten hoe accuraat onze voorspelling zal zijn



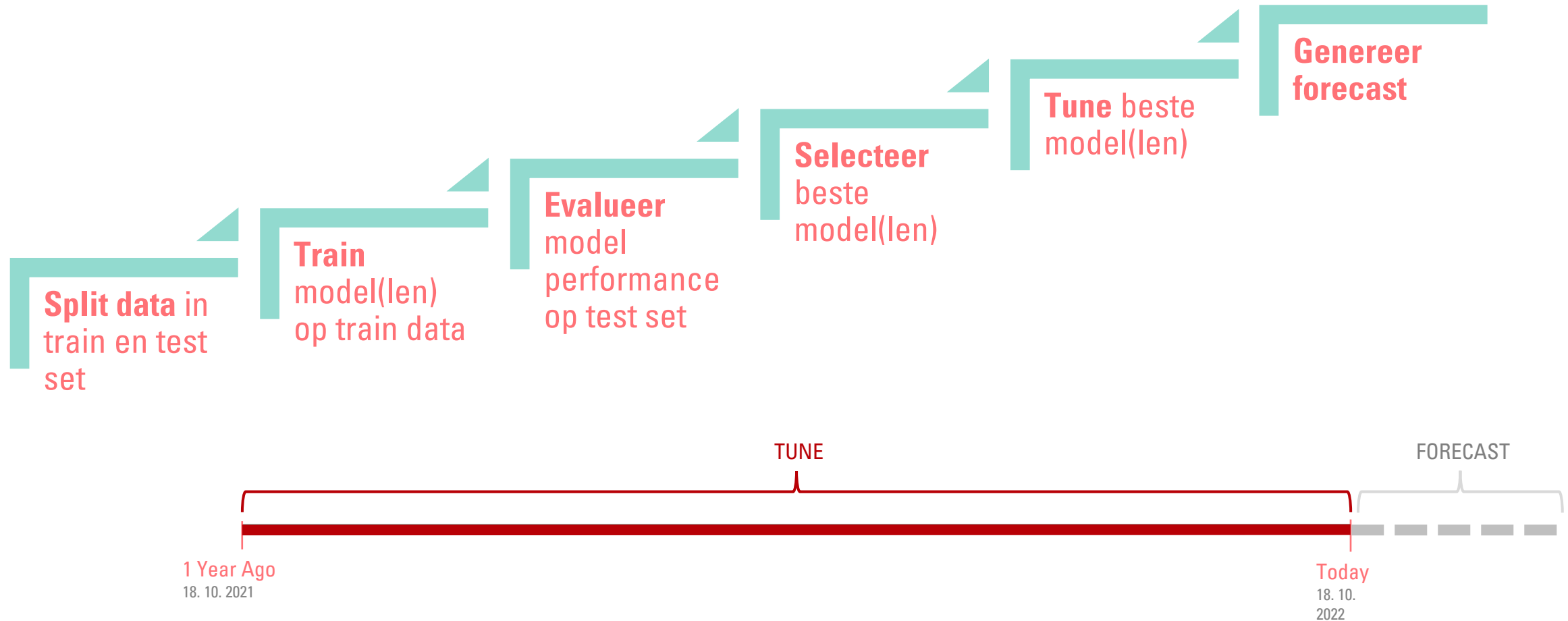
ML MODEL DEVELOPMENT PROCES



ML MODEL DEVELOPMENT PROCES



MODEL DEVELOPMENT PROCES

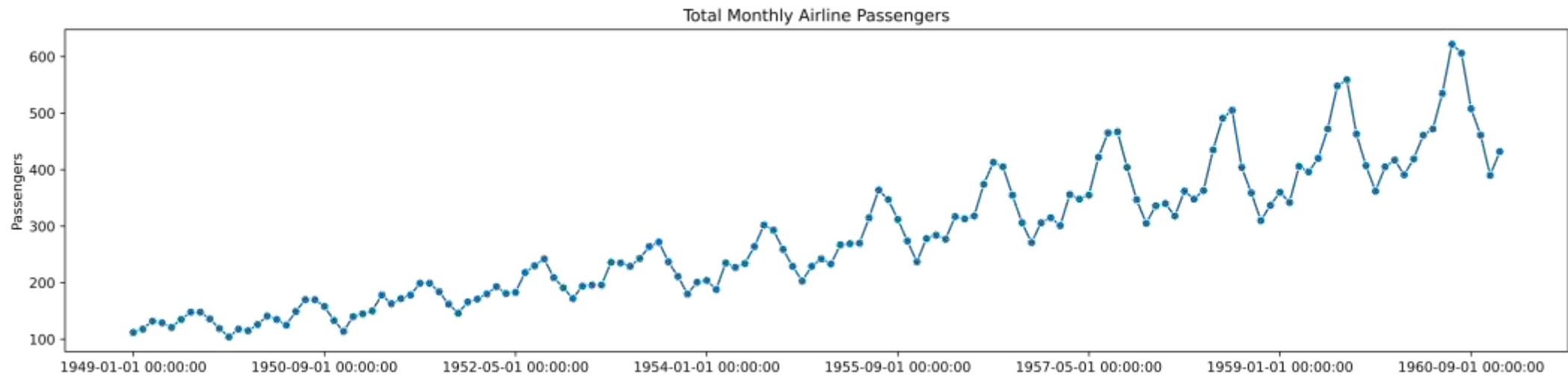


DEMO: PYCARET IN ACTIE



CONTEXT (DATASET)

- Een van de meest populaire forecast datasets.
- **Maandelijksse totalen** van **passagiers** van een U.S. **airline** van **1949** tot en met **1960**



STAP 1: IMPORTEER PYTHON LIBRARIES

```
# Pandas
import pandas as pd

# PyCaret Library
from pycaret.time_series import TSForecastingExperiment

# OPTIONAL: Plotting Library
from sktime.utils.plotting import plot_series
import matplotlib.pyplot as plt; plt.rcParams['figure.figsize'] = (20, 10)
```



STAP 2: IMPORTEER DATA

```
data = pd.read_csv('AirData.csv', index_col = 'Date')  
data.index = pd.to_datetime(data.index)  
data.head()
```

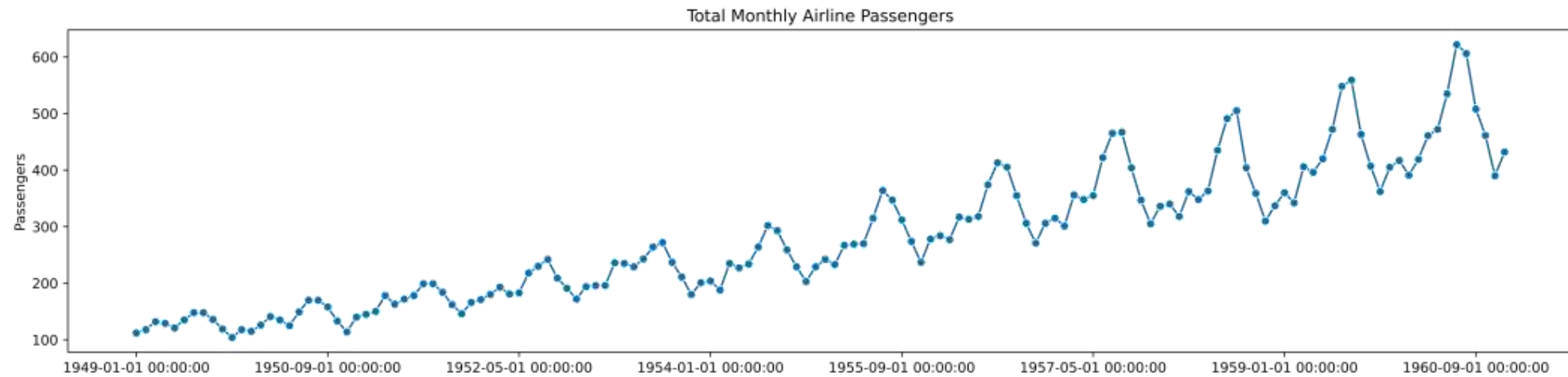
Passengers	
Date	
1949-01-01	112
1949-02-01	118
1949-03-01	132
1949-04-01	129
1949-05-01	121



STAP 3: BEGRIJP JE DATA!

- Elk project begint met wat visualisaties!

```
y = data['Passengers']  
_ = plot_series(y)
```



STAP 4: EXPERIMENT SETUP: MODEL DEFINITIES

- Ons "Experiment" gereed maken. We stellen de volgende zaken in: de forecast horizon (fh); hoe vaak we ons model willen valideren/evalueren; ...

```
exp = TSForecastingExperiment()  
exp.setup(data=y, fh=12, fold=3, session_id=42)
```

- Forecast horizon: 1 jaar (= 12 maanden)
- 3 testing folds
- Session ID: 42

	Description	Value
0	session_id	42
1	Target	Passengers
2	Approach	Univariate
3	Exogenous Variables	Not Present
4	Original data shape	(144, 1)
5	Transformed data shape	(144, 1)
6	Transformed train set shape	(132, 1)
7	Transformed test set shape	(12, 1)
8	Rows with missing values	0.0%
9	Fold Generator	ExpandingWindowSplitter
10	Fold Number	3
11	Enforce Prediction Interval	False
12	Seasonal Period(s) Tested	12
13	Seasonality Present	True
14	Seasonalities Detected	[12]
15	Primary Seasonality	12



WE KUNNEN MEERDERE MODELLEN TEGELIJK KIEZEN!

- In deze stap worden alle mogelijke modellen getraind en geëvalueerd, waarna de best presterende worden geselecteerd

```
best_baseline_models = exp.compare_models(n_select=3)
```

Our 3 best models

	Model	MASE	RMSSE	MAE	RMSE	MAPE	SMAPE	R2	TT (Sec)
{	exp_smooth	0.5716	0.5997	16.7767	19.7954	0.0422	0.0427	0.8954	0.0567
	ets	0.5865	0.6156	17.2147	20.3200	0.0434	0.0439	0.8901	0.0700
	et_cds_dt	0.6666	0.7255	19.6620	24.0121	0.0490	0.0489	0.8465	0.1067
	huber_cds_dt	0.6813	0.7866	20.0334	25.9670	0.0491	0.0499	0.8113	0.0333
	arima	0.6830	0.6735	20.0069	22.2199	0.0501	0.0507	0.8677	0.1267
	lr_cds_dt	0.7004	0.7702	20.6084	25.4401	0.0509	0.0514	0.8215	0.6167
	ridge_cds_dt	0.7004	0.7703	20.6086	25.4405	0.0509	0.0514	0.8215	0.1400
	lar_cds_dt	0.7004	0.7702	20.6084	25.4401	0.0509	0.0514	0.8215	0.0267
	en_cds_dt	0.7029	0.7732	20.6816	25.5362	0.0511	0.0516	0.8201	0.1233
	lasso_cds_dt	0.7048	0.7751	20.7373	25.6005	0.0512	0.0517	0.8193	0.0267
	br_cds_dt	0.7112	0.7837	20.9213	25.8795	0.0515	0.0521	0.8144	0.0267
	knn_cds_dt	0.7162	0.8157	21.1613	26.9700	0.0521	0.0529	0.7811	0.0500
	auto_arima	0.7181	0.7114	21.0297	23.4661	0.0525	0.0531	0.8509	1.2933
	dt_cds_dt	0.7651	0.8455	22.5331	27.9902	0.0559	0.0554	0.7884	0.0400
	rf_cds_dt	0.7765	0.8746	22.8282	28.8529	0.0549	0.0560	0.7721	0.1267
	gbr_cds_dt	0.7938	0.9214	23.3422	30.4055	0.0567	0.0574	0.7498	0.0400



STAP 5: MODEL TUNING

- De 3 geselecteerde worden getuned en 'geblend' tot een definitief model.

```
best_tuned_models = [exp.tune_model(model) for model in best_baseline_models]  
mean_blender = exp.blend_models(best_tuned_models, method='mean')
```

	cutoff	MASE	RMSSE	MAE	RMSE	MAPE	SMAPE	R2
0	1956-12	0.3722	0.4769	10.8692	15.6100	0.0265	0.0271	0.9207
1	1957-12	0.7114	0.7128	21.7505	24.1995	0.0603	0.0582	0.8466
2	1958-12	0.5475	0.5561	15.6432	18.0752	0.0355	0.0364	0.9269
Mean	NaT	0.5437	0.5819	16.0876	19.2949	0.0408	0.0405	0.8981
SD	NaT	0.1385	0.0980	4.4534	3.6112	0.0143	0.0130	0.0365

3 test folds
Evalueer performance voor alle 3
Toon gemiddelde error



STAP 6: VOORSPELLEN!

- Met dit model maken we een voorspelling

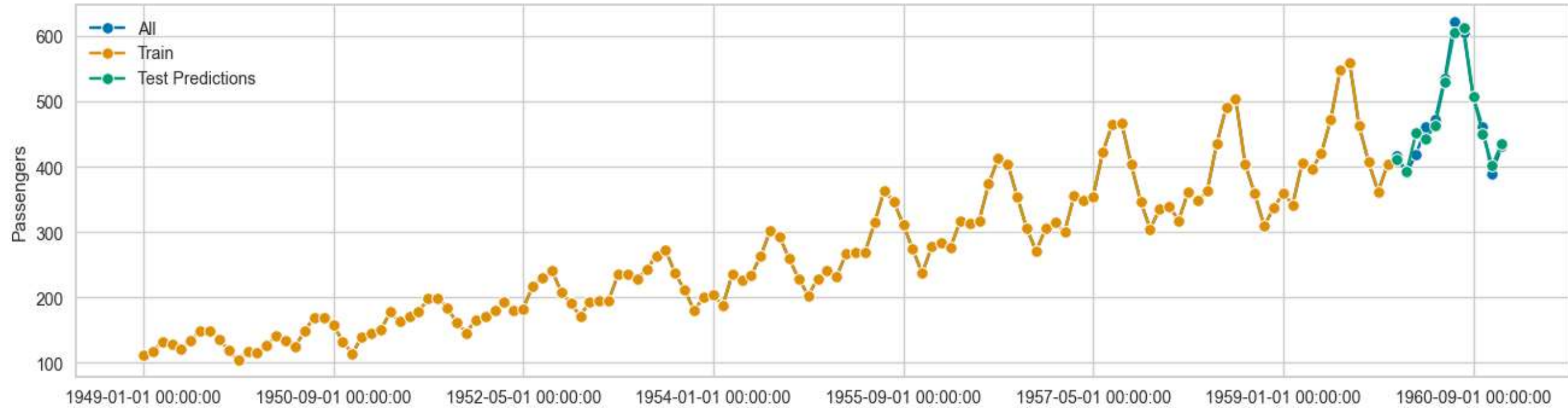
```
y_pred = exp.predict_model(mean_blender)  
y_pred.index = y_pred.index.to_timestamp()
```

Date Index Voorspelling

1960-01-01	407.0120
1960-02-01	387.2591
1960-03-01	444.9259
1960-04-01	433.2175
1960-05-01	454.2434
1960-06-01	524.0066
1960-07-01	594.9387
1960-08-01	606.2521
1960-09-01	498.2756
1960-10-01	441.9263
1960-11-01	390.5063
1960-12-01	425.5572



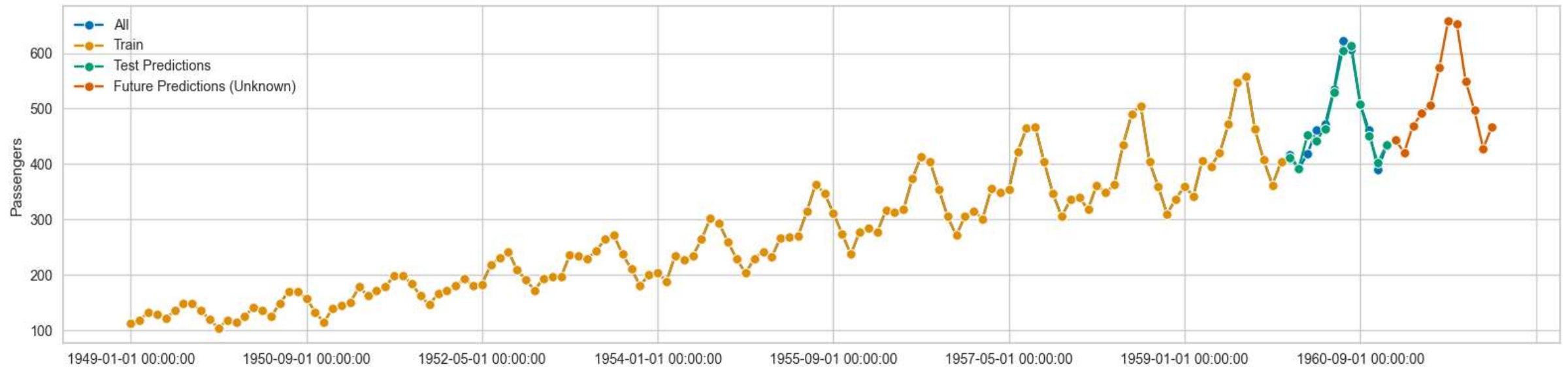
VISUALIZE VOORSPELLINGEN



STAP 7: MAAK VOORSPELLING VOOR AANKOMEND JAAR (TOEKOMST)

```
final_model = exp.finalize_model(mean_blender)

y_pred_future = exp.predict_model(final_model, verbose=False)
y_pred_future.index = y_pred_future.index.to_timestamp()
```



WRAP-UP



SAMENVATTING

- AutoML is: het **automatiseren** van het tijdrovende en iteratieve proces van **machine learning**
- PyCaret is: een **open-source, low-code** machine learning library in **Python**
 - Voor **automated supervised** en **unsupervised machine learning**
- Waarmee je **snel** en **makkelijk** aan de slag kunt met data science





Ties de Bont

+31 6 40534881

ties.de.bont@rockfeather.com

<https://www.linkedin.com/in/ties-de-bont-04232419b/>

