TIME-SERIES FORECASTING

TRADITIONAL
VS
DEEP LEARNING

EDINBURGH DATA SCIENCE MEETUP

24/04/2025

Dr. Sue Liu

Machine Learning Engineer Manager

Kingfisher Plc

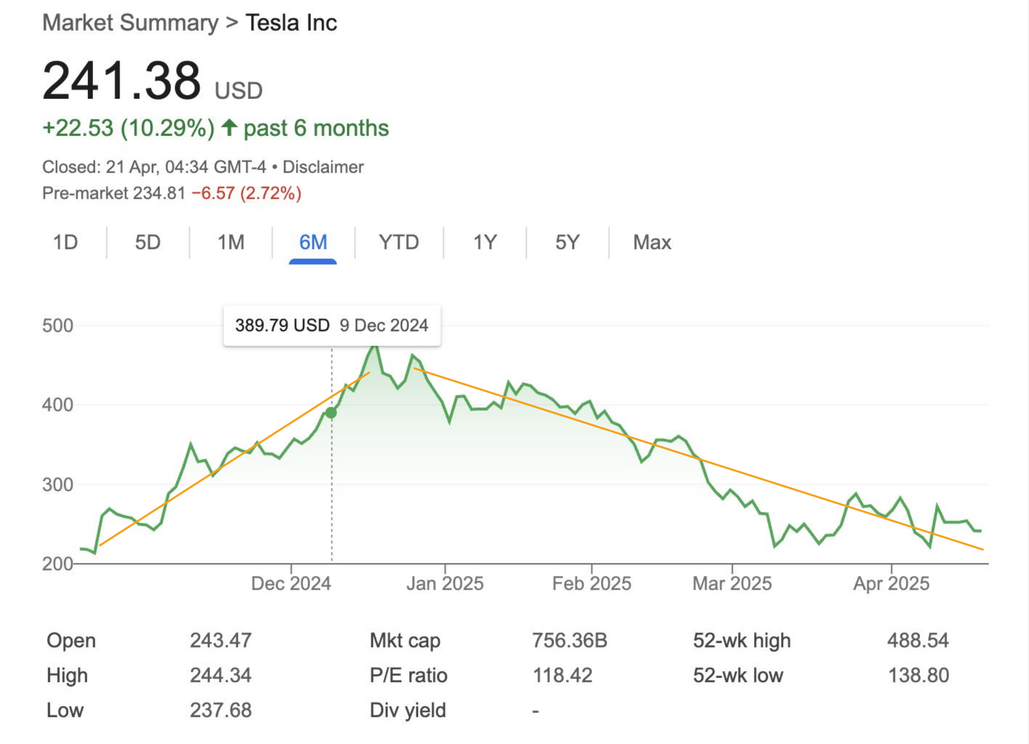
WHAT IS FORECASTING?

Predicting what is likely to happen in the future based on present and historical data

Prediction is hard, especially about the future
– Niels Bohr

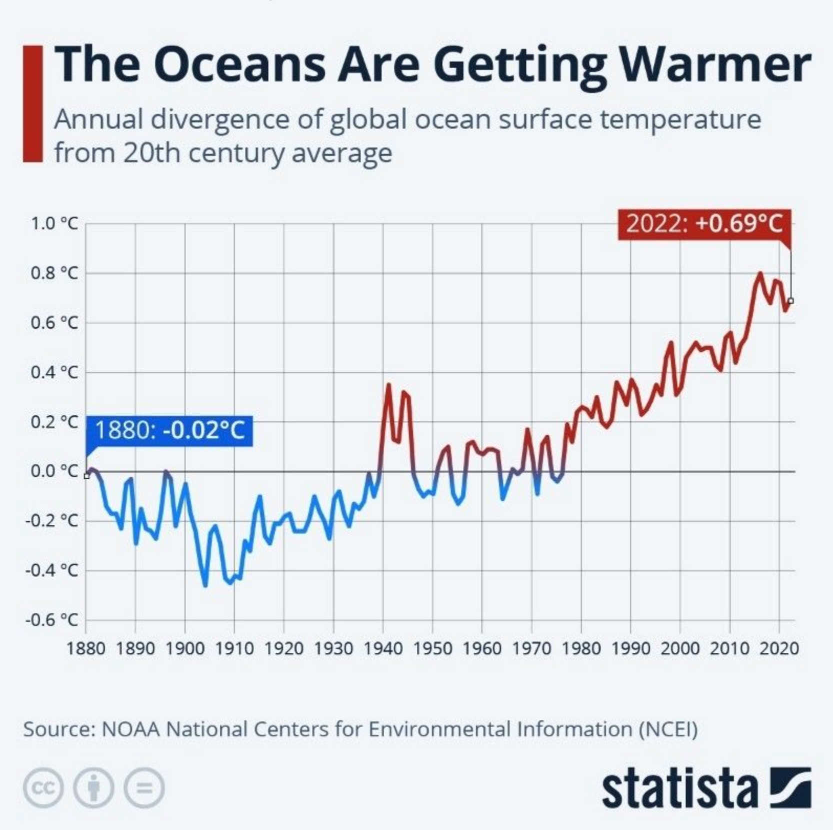
CAN WE MAKE FORECASTS?

Tesla Share Prices



Source: Google

Global average sea surface temperature



Time scale

Forecast capability depends on the persistence of patterns

UNIQUE CHALLENGES

- Time plays a crucial role
  - Difference between training and inference data

- Complex interactions with downstream decision problems
  - Long feedback cycles

- Users are typically business functions or analysts
  - Challenges for presenting results

FORECASTING: TRAINING AND EVALUATION

Loss Functions

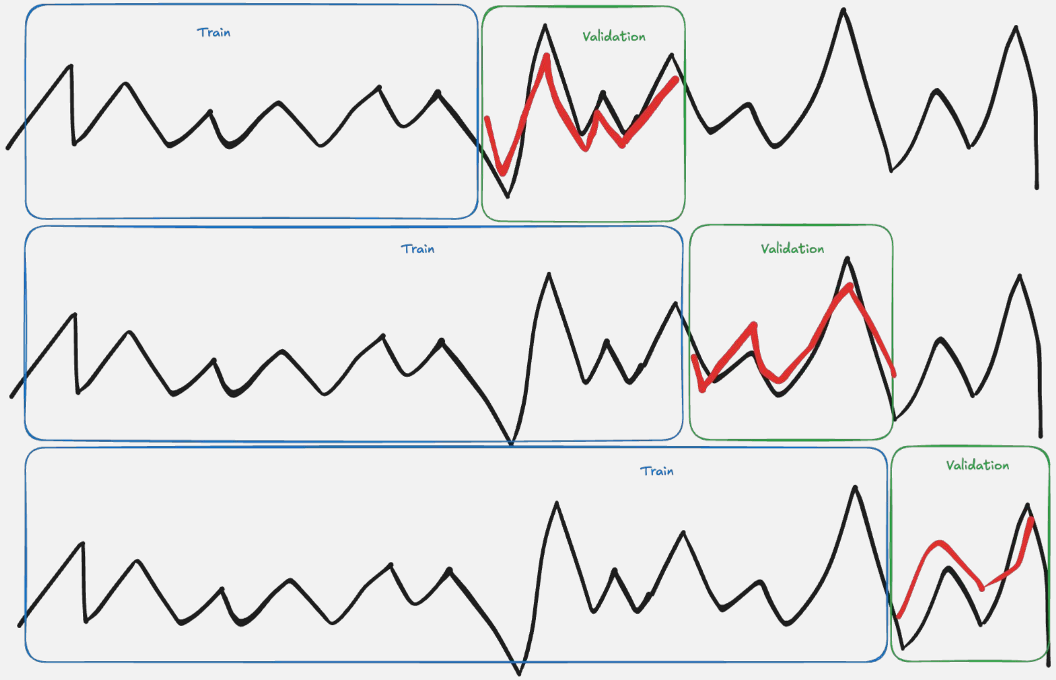
- Root Mean Squared Error (RMSE)
- Negative log likelihood
- Quantile loss
- Tweedie loss

Reporting Metrics

Symmetric Mean Absolute Percentage Error

$$\text{SMAPE} = \frac{100}{n} \sum_{t=1}^{n} \frac{|Y_t - F_t|}{(|Y_t| + |F_t|)/2}$$

Evaluation



- Training data kept contiguous
- Cross-validation using sliding windows

HISTORY OF FORECASTING METHODS

Traditional

Exponential Smoothing (ETS)

Holt (1957) Winters (1960) Gardner (1985) **Hyndman (2002)**

1950 1960 1970 1980 1990 2000 2010 2020

Kalman (1960) Hilmer (1982) Gómez (1994)

Box & Jenkins (1970)

ARIMA

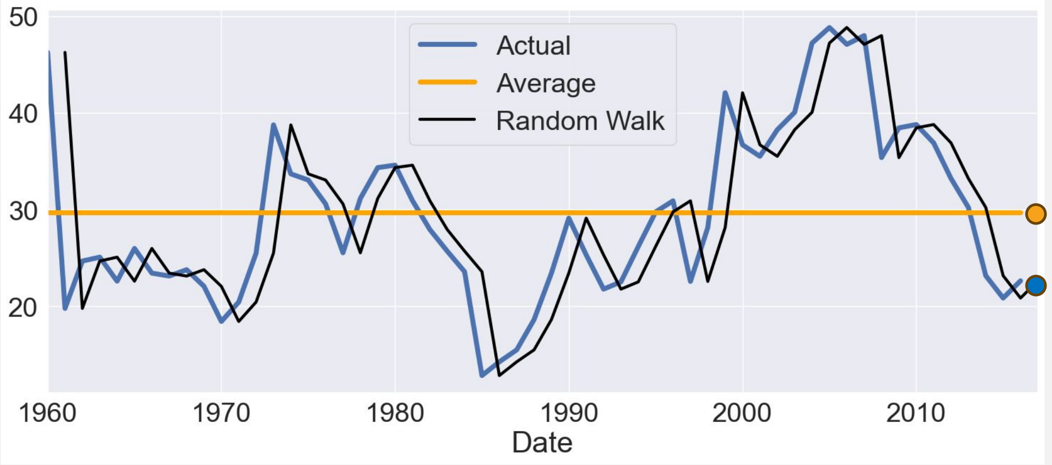
Models are over 50 years old!

EXPONENTIAL SMOOTHING

Time series $y_1, y_2, \ldots, y_T$

Naive (random walk) $\hat{y}_{T+h|T} = y_T$

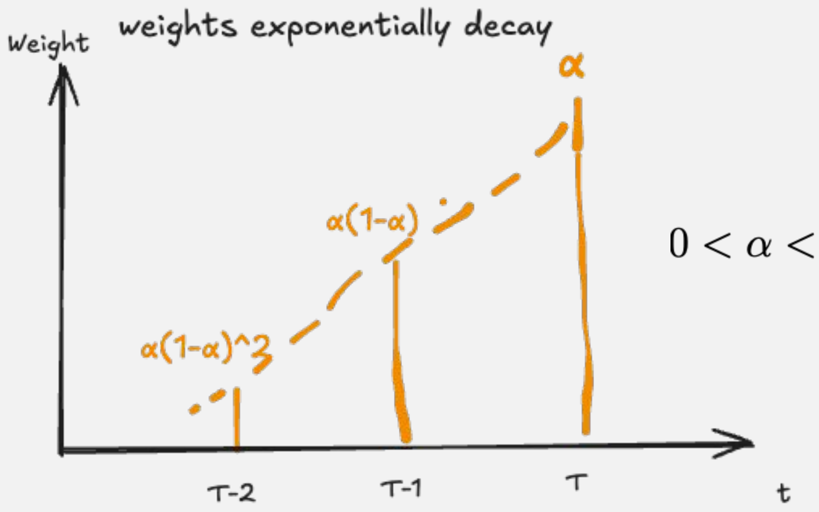
Average forecasts $\hat{y}_{T+h|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$



Exponential Smoothing

Something in between

More recent data should have more weight



$0 < \alpha < 1$

$$\hat{y}_{T+h|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots$$

EXPONENTIAL SMOOTHING

Time series $y_1, y_2, \ldots, y_T$

Simple Exponential Smoothing

Forecast equation $\hat{y}_{t+h} = \ell_t$

Level equation $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$

Model with trend and seasonality

Forecast $\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$

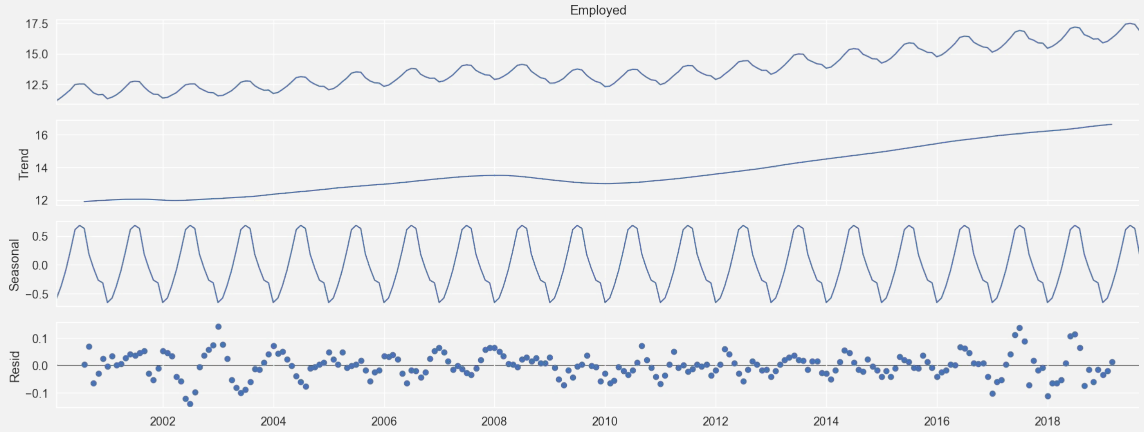
Level $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$

Trend $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$

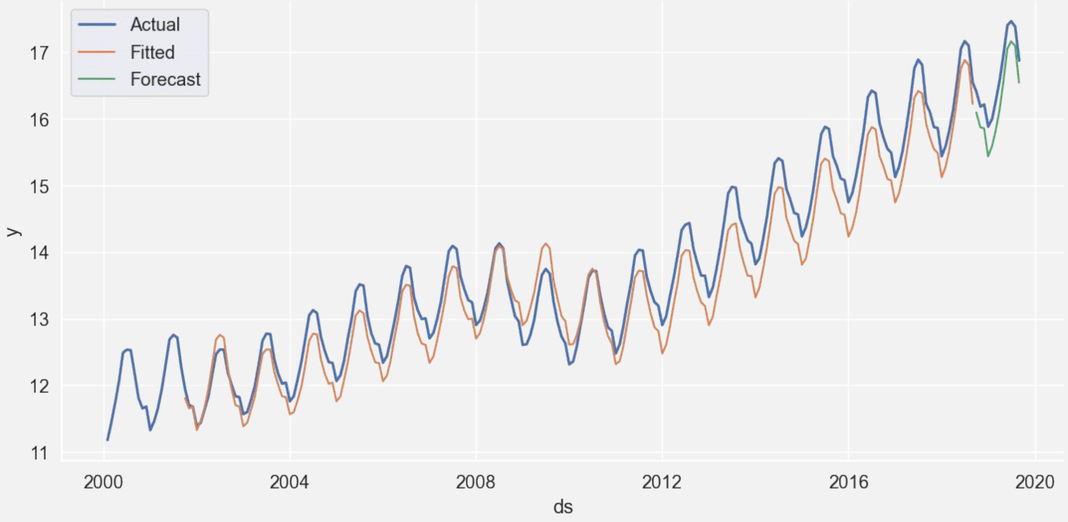
Seasonal $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},$

Winters (1960)

US Employees in hospitality industry



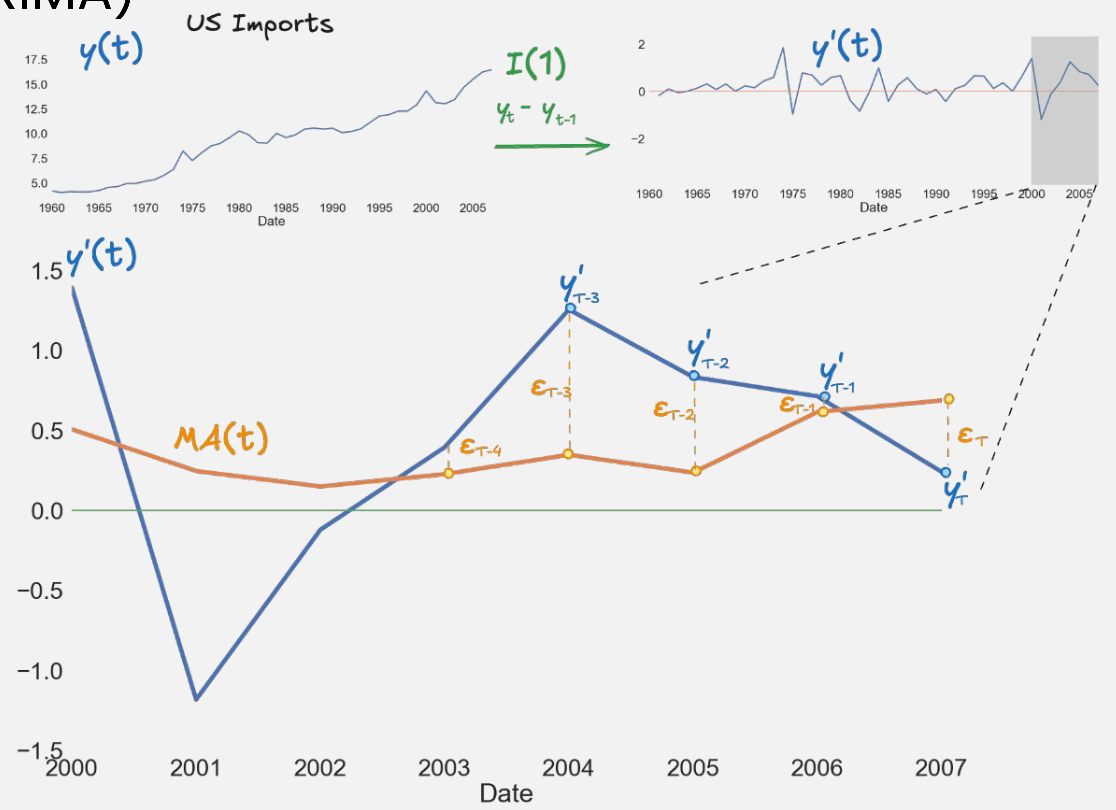
Fit and combine components



ARIMA MODELS

Autoregressive Integrated Moving Average (ARIMA)

Component	**Description**
AutoRegressive *AR(p)*	Regression using its own past *p* values as features
Integrated *I(n)*	Run differencing *n* times to make time series stationary
Moving Average *MA(q)*	Use past *q* forecast errors as features

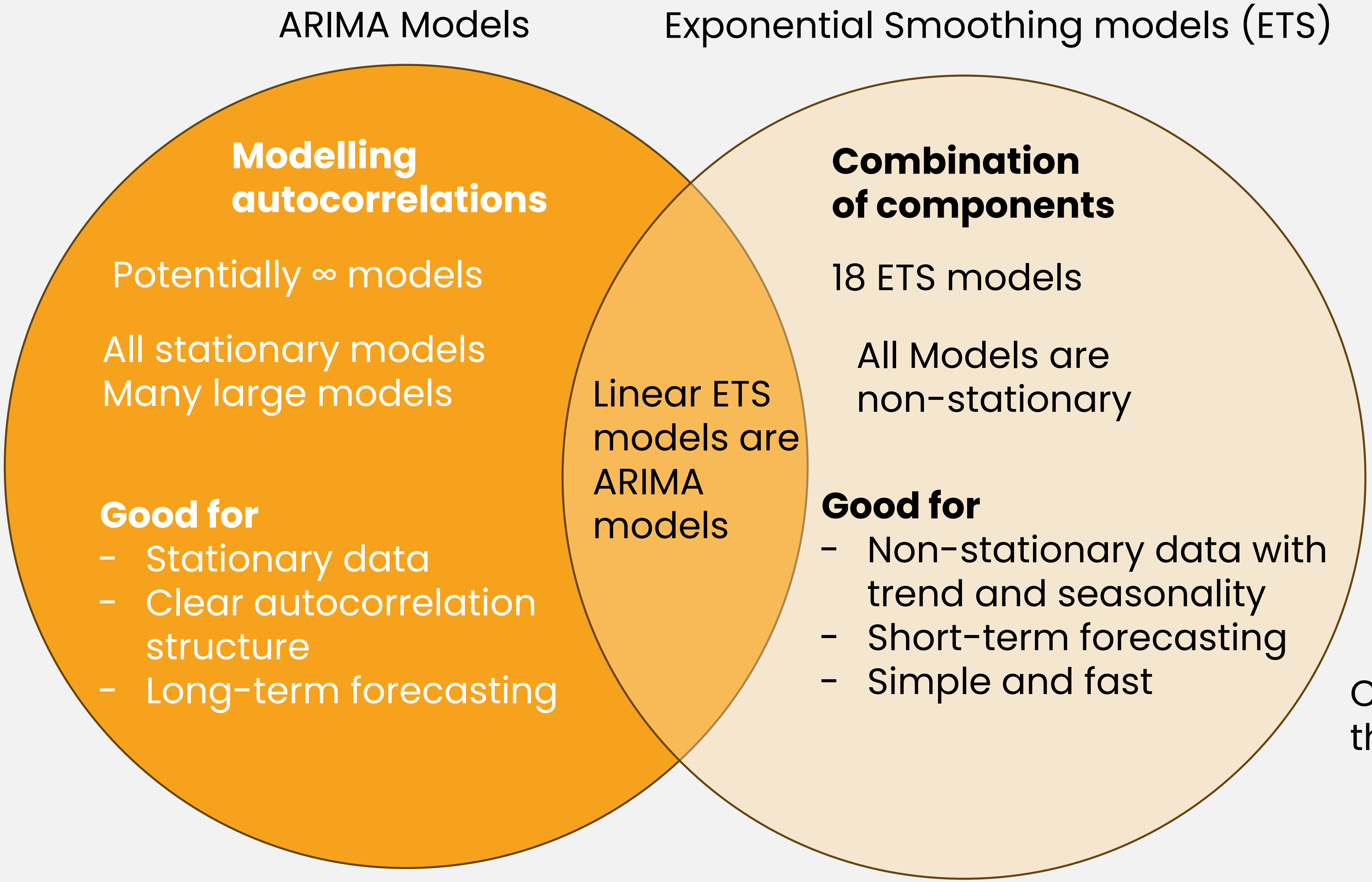


$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

AR(p) *MA(q)*

Linear model using patterns in the **autocorrelations** to forecast future values

ARIMA VS EXPONENTIAL SMOOTHING



Combine to get the best results!

FORECASTING WITH NEURAL NETWORKS

Previous "Consensus" in the Forecasting Community

Neural Networks don't work! Not enough data to fit a good NN model

M3 Competition (2000)

Forecasting challenge containing 3003 time series
Won by traditional methods

Shouldn't the successful Deep Learning models from NLP and CV just work?

YES!

M4 Competition (2018)

Forecasting challenge containing >100,000 time series
Won by a neural network, combined with a statistical method (Smyl, Uber)

DEEP LEARNING FOR FORECASTING

Challenges

Data, scaling, sample efficiency, incorporating prior knowledge

Solutions

Solve the right problems!

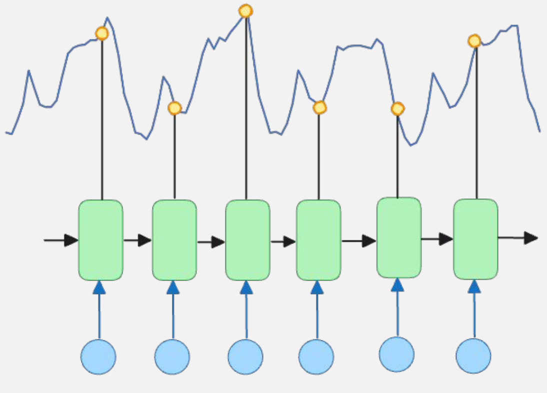
- Learn complex patterns from many time series at once
- New tools and frameworks (Nixtla, Darts)
- Novel adaptation of foundation models

Tourist trips in 9 Australian states

BASICS: MODEL STRUCTURES

One-to-one (Generative Model)

$$f : x_t \mapsto z_t$$



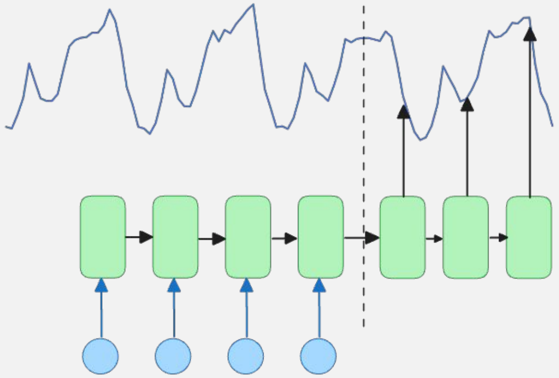
One-to-one

Training Sequence

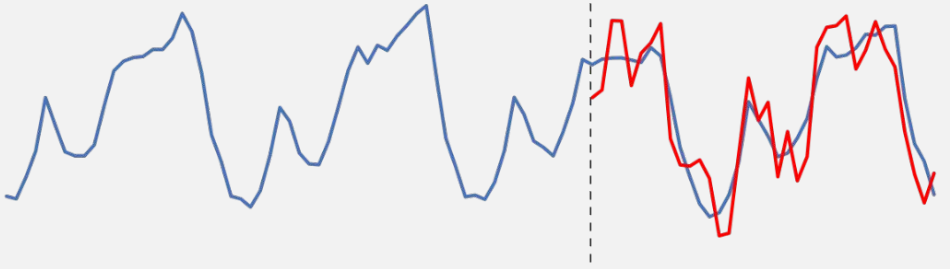
How well does the prediction reconstruct the **observed time series**?

Many-to-Many (Discriminative Model)

$$f : \{z_1, \cdots, z_{T_e}\} \mapsto \{z_{T_e+1}, \cdots, z_{T_e+T_d}\}$$



Seq2Seq (Many-to-Many)



Encoding Sequence

Decoding Sequence

How well does the prediction reconstruct the **decoding sequence** conditioned on the **encoding sequence**?

MODEL STRUCTURES COMPARISON

One-to-One

- No need to retrain for different prediction length
- Input features need to be available during prediction phase
- Autoregressive - performance decrease with prediction steps

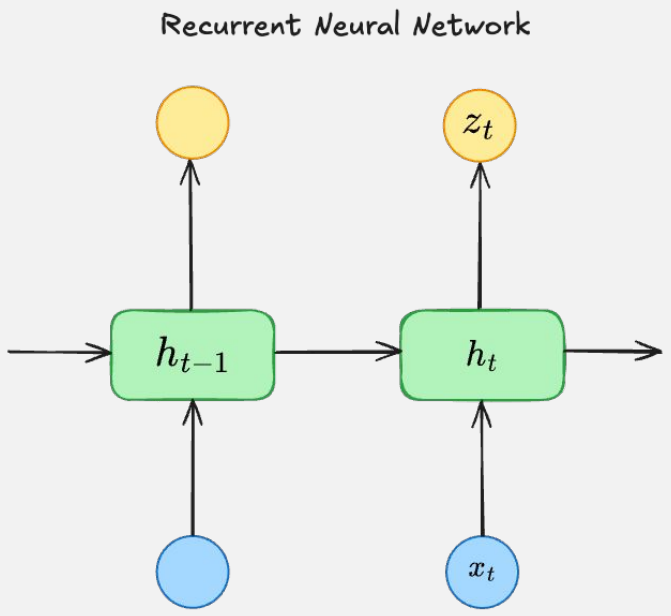
Many-to-Many

- Can have disjoint encoding and decoding features
- Needs retraining when changing the decoder length
- Generally better performance over whole prediction horizon

BASICS: RECURRENT NEURAL NETWORKS (RNN)

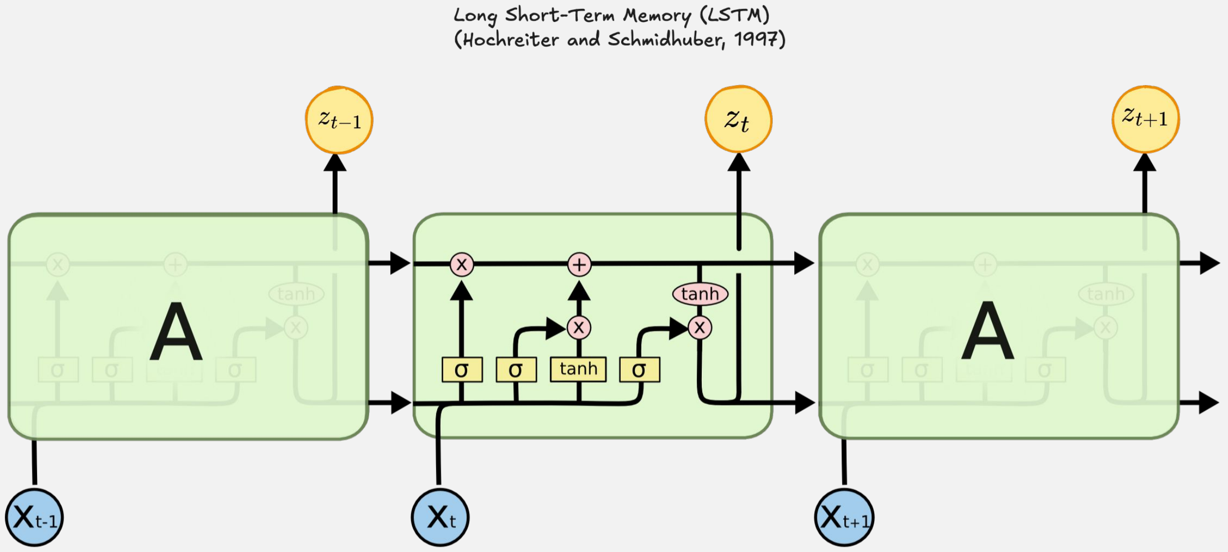
Current hidden state:

- Previous hidden state
- Input features



Can be unstable during training

Long Short-Term Memory (LSTM)

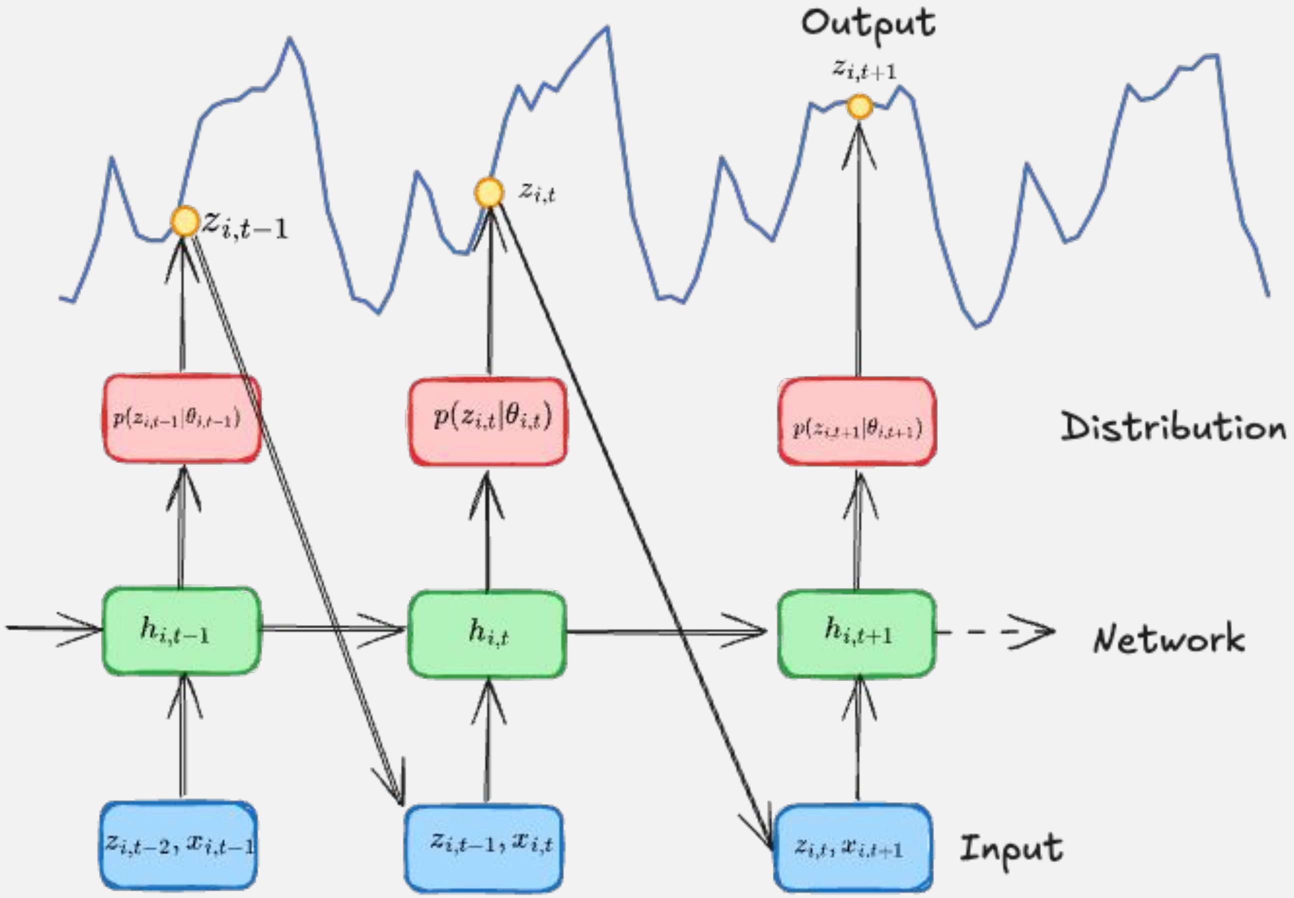


https://colah.github.io/posts/2015-08-Understanding-LSTMs/

$$C_t = \alpha_t \cdot C_{t-1} + \beta_t \times \sigma(\theta_0 h_{t-1} + \theta_1 x_t)$$

current state = forget gate x old stuff + input gate x new stuff

ONE-TO-ONE: DEEPAR (AMAZON)



- Trains a single model using multiple time series to learn global characteristics

- One-to-one
- **LSTM** network with autoregressive input
- Probabilistic forecasting
- Makes forecast through sampling
- Allows 'cold start' forecasting

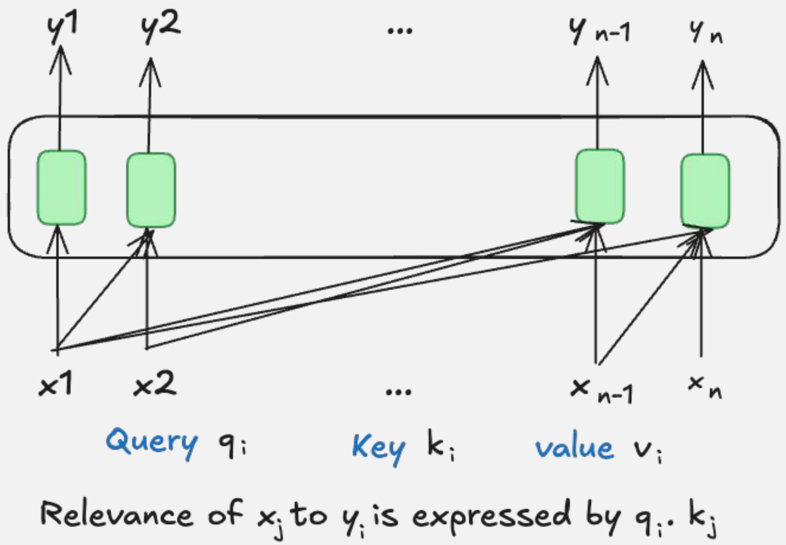
Flunkert et al. (2017)

BASICS: TRANSFORMERS

Attention Mechanism from NLP

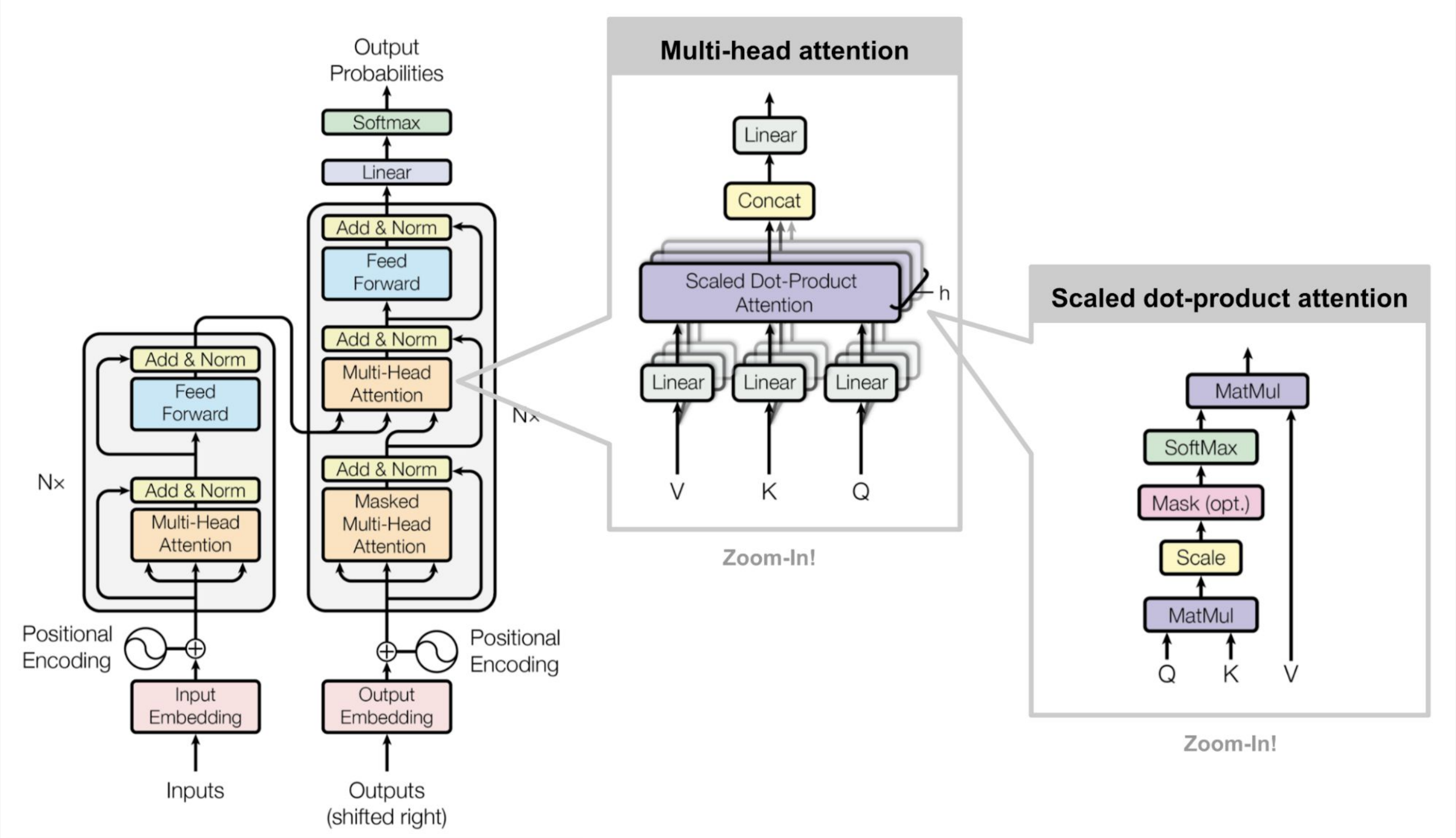
Improvement on RNNs

Self-attention



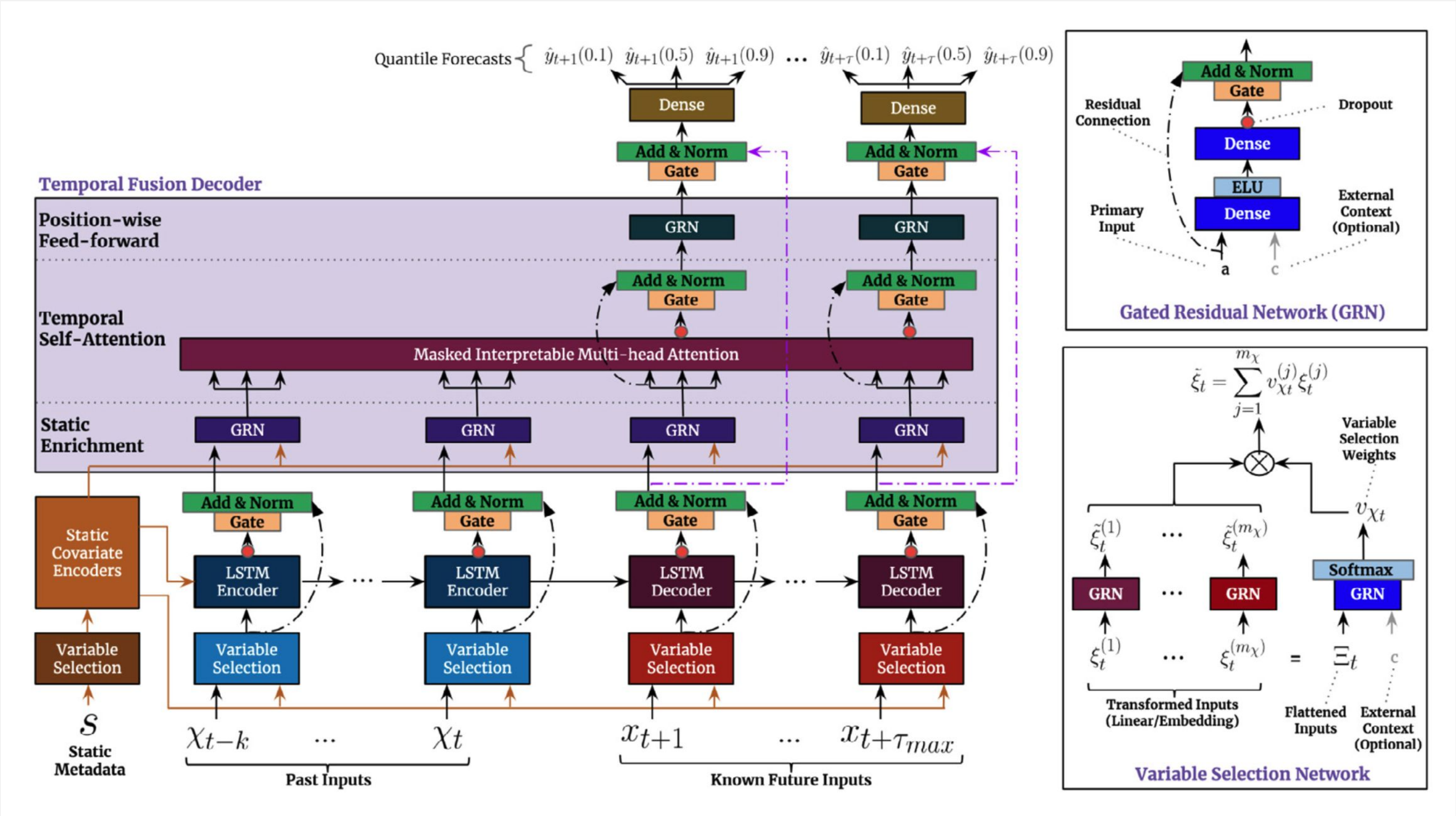
Y: $\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{n}}\right)\mathbf{V}$

Full Transformer Architecture



Vaswani et al. (2017)

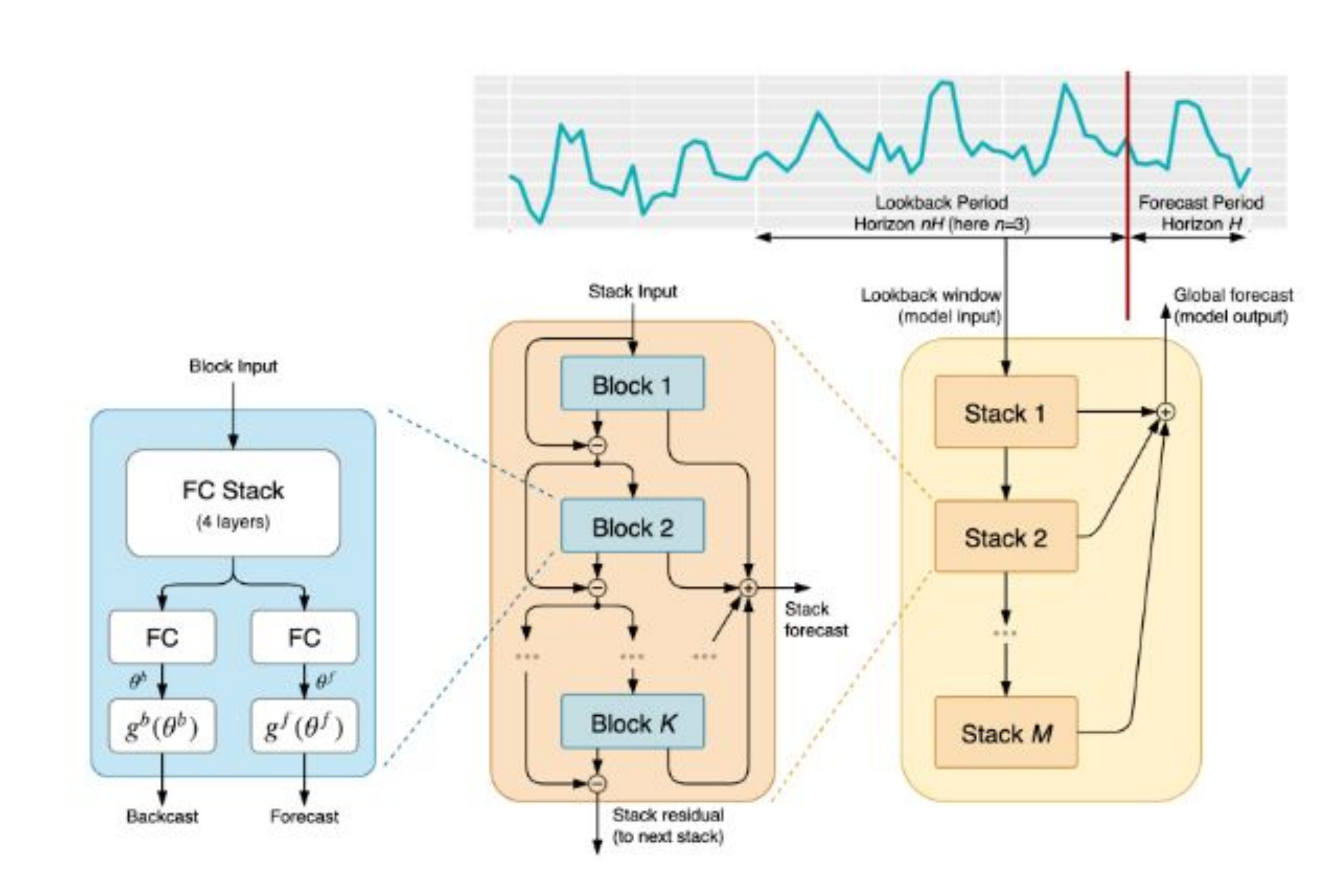
MANY-TO-MANY: TEMPORAL FUSION TRANSFORMER (GOOGLE)



- Multi-horizon forecasting
- Interpretable
- **LSTM** network for time-dependent processing
- **Multi-head attention** to integrate information from any time step
- Performs better than DeepAR on many benchmarks

Lim et al. (2021)

MANY-TO-MANY: N-BEATS (ELEMENT AI)



- Multi-step predictions
- No RNN or attention layers, faster to train
- Interpretable forecasts
- Basis for Zero-shot Transfer Learning
- First DL model to outperform all statistical approaches in the M4 competition

Oreshkin et al. (2020)

OTHERS

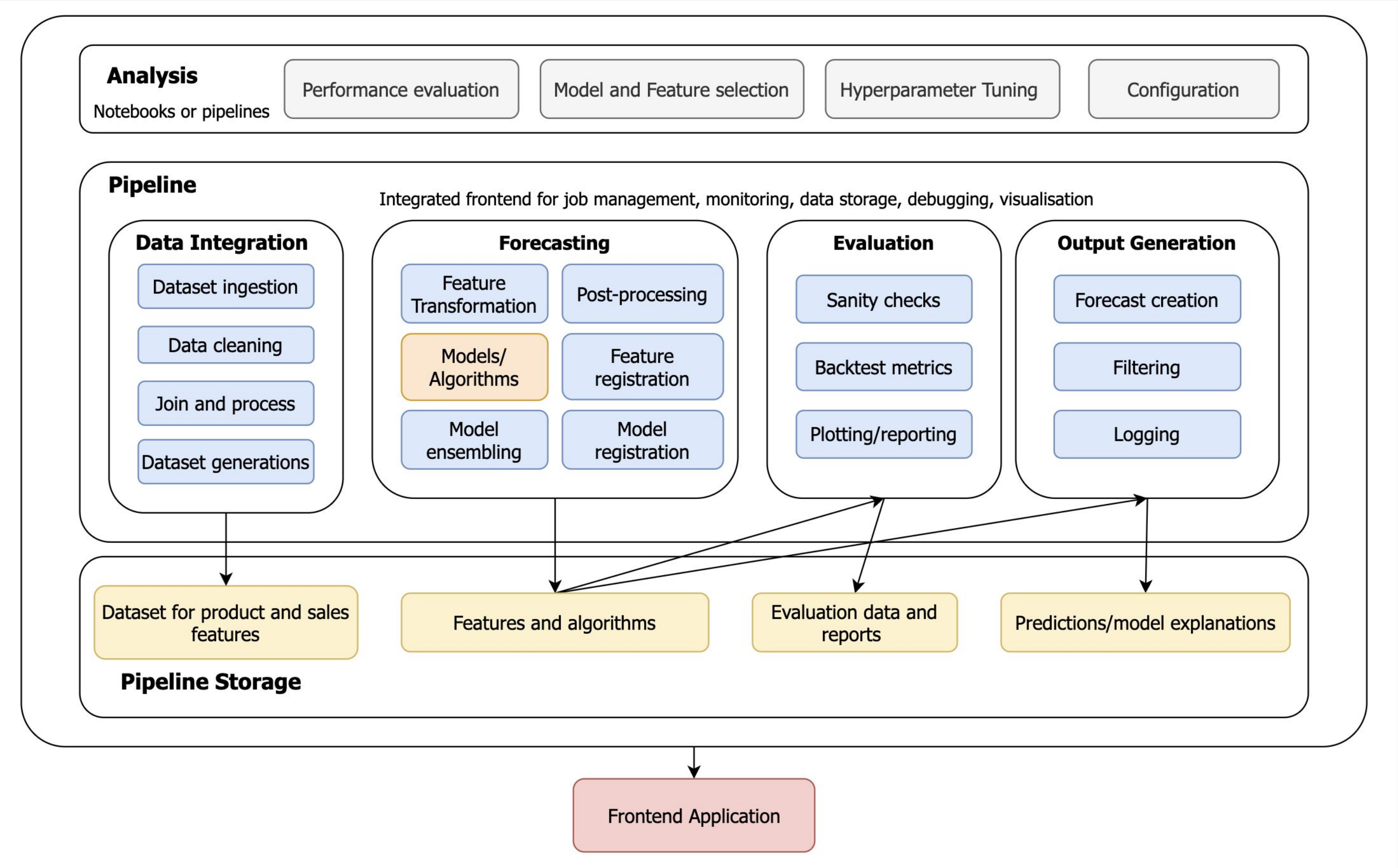
New Deep Learning models for time series forecasting are being published regularly

- NHITS
- Tiny-Time-Mixers (TTM)
- TabPFN-TS
- TSMixer
- iTransformer
- TIME-MOE
- MOIRAI
- MOMENT
- TimesFM
- TimeGPT

https://aihorizonforecast.substack.com/
Most of them implemented in the Nixtla library

FORECASTING SYSTEMS

FORECASTING SYSTEMS IN THE REAL WORLD



TRADITIONAL VERSUS DEEP LEARNING

	Traditional (Statistical) methods	Deep Learning
PROS	• Relatively easy to understand • White box – everything needs to be explicitly modelled • Embarrassingly parallel • Good performance for many use cases	• Little feature engineering needed • Learns complex patterns across time series • State of the art performance in competitions • Adopted by a surprisingly large number of companies • Constantly shifting landscape!
CONS	• Manual work by experts required • Cannot learn patterns across time series • Need pipelines to tune and maintain (imagine modelling ~100k time series) • Cannot handle cold-starts	• Little control over predictions • Costly to train • Difficult to tune hyperparameters • Infrastructure needed to serve model
	Strategic forecasting (e.g. finance, sales)	Operational forecasting (e.g. demand forecasting)

What I didn't mention: **boosting and ensemble methods** – also highly performant in M-competitions

GETTING STARTED WITH FORECASTING

Data

Makridakis Competitions (M-Competitions) 1982-

Open competitions to evaluate and compare different time series forecasting methods

GitHub: https://github.com/Mcompetitions/
Website: https://www.unic.ac.cy/iff/research/forecasting/m-competitions/

M4 (2018)

100,000 time series with different frequency

M5 (2020)

~42,000 hierarchical time series provided by Walmart

M6 (2022)

Real time financial forecasting 50 S&P 500 US stocks + 50 International ETFs

OPEN-SOURCE FORECASTING PACKAGES

Package	Language	Methods	Notes
Forecast	R	Statistical	Reference statistical forecasting package (for R enthusiasts)
Statsmodels	Python	Statistical	Python library for statistical time series modelling and analysis (not as comprehensive as R)
Prophet (Meta)	Python/R	Statistical	Out-of-the box, easy to add exogenous features. Performance variable
Nixtla	Python	Statistical/ML/ Deep Learning	State of the art deep-learning models implemented plus statistical/ML libraries
Darts	Python	Statistical/ML/ Deep Learning	Comprehensive forecasting library (re-implements models from many libraries)

Others: GluonTS, Pytorch-forecasting, etc.

REFERENCES

Textbook

Forecasting Principles and Practice (2018, 3rd Edition) Hyndman & Athanasopoulos
https://otexts.com/fpp3/

Articles

- Transformer: Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
- **DeepAR**: Salinas, David, et al. "DeepAR: Probabilistic forecasting with autoregressive recurrent networks." *International journal of forecasting* 36.3 (2020): 1181-1191.
- **N-BEATS**: Oreshkin, Boris N., et al. "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting." *arXiv preprint arXiv:1905.10437* (2019).
- **Temporal Fusion Transformers** :Lim, Bryan, et al. "Temporal fusion transformers for interpretable multi-horizon time series forecasting." *International Journal of Forecasting* 37.4 (2021): 1748-1764.

Blogs

AI Horizon Forecast: https://aihorizonforecast.substack.com/
Understanding LSTMs: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

We are hiring!

Join a team of 35 machine learning and software engineers working on exciting problems in recommender systems, search, GenAI and forecasting for retail and e-commerce!

We have a number of machine learning engineer positions open.

Feel free to reach out to me, and/or check our company pages

- LinkedIn
- Company blog on Medium