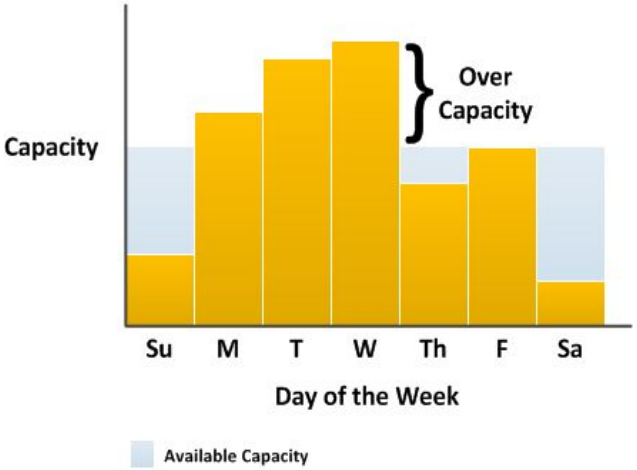
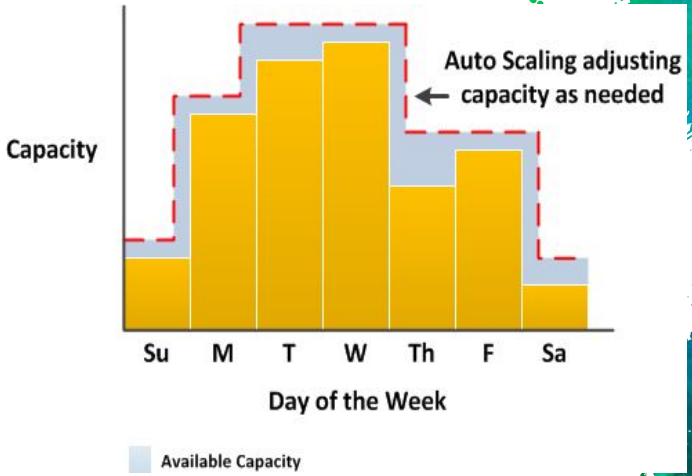

Server
Dynamic
Autoscaling
Filippo Medri
Metis 2018

Choosing the right size of CPU

- Average over the number of requests

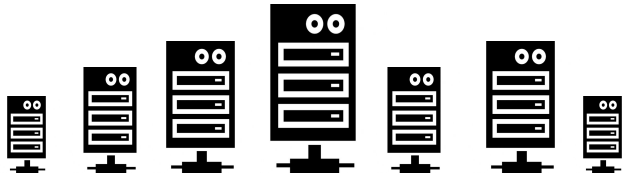


- Forecast the number of requests



Agenda

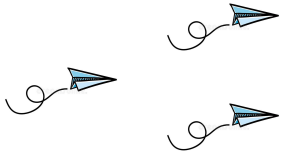
- Introduction
- Time Series
- Deep Learning
- Future Directions




1,071,840
requests over 7 months

Introduction

web logs coming from
web server..



Web Log collection

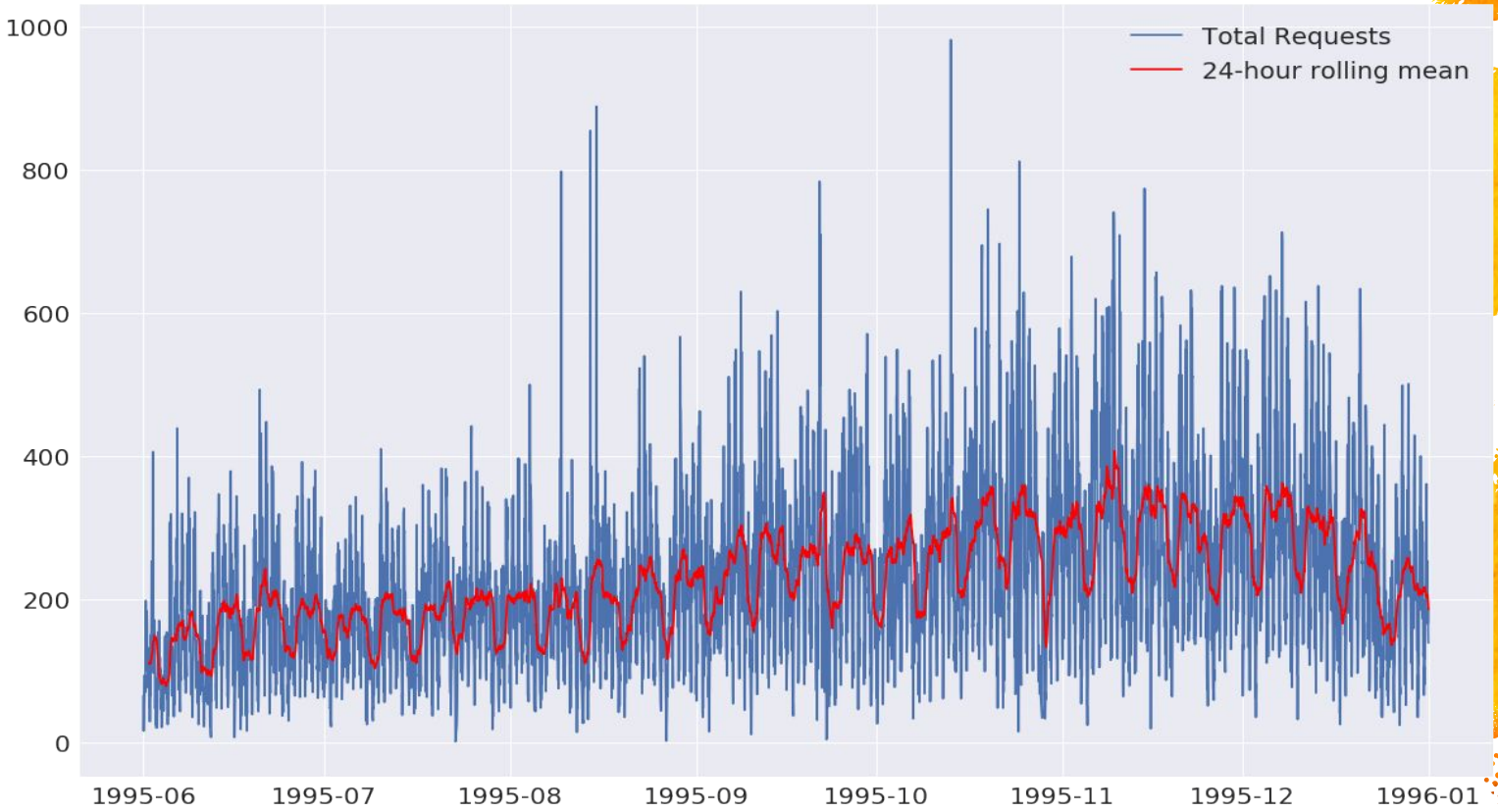
Request Trend

... forecast
1 day in advance ...

Request Forecast

CPU Scaling

Analysis of hourly requests



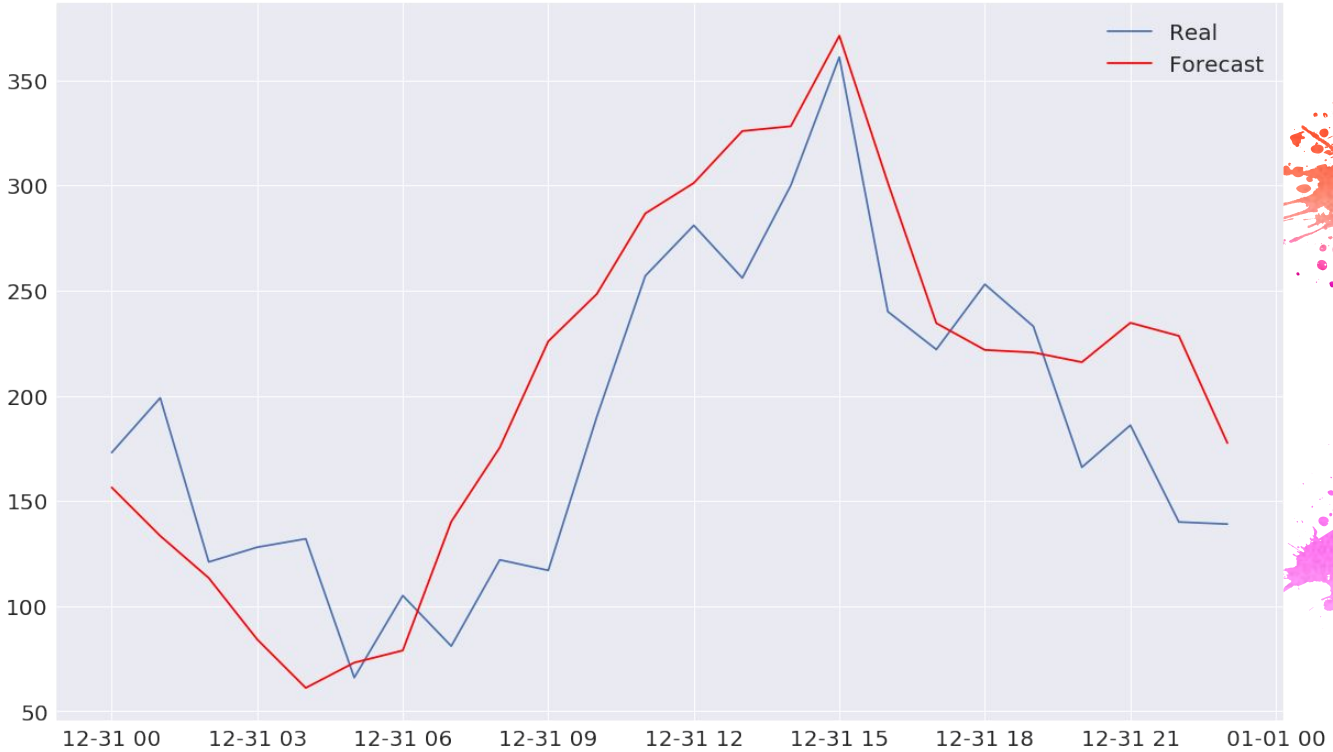
Finding the best time series model

Stationary
Autocorrelation (First 6 Lags)
Partial Autocorrelation (First 2 Lags)
Seasonality over the 24 h

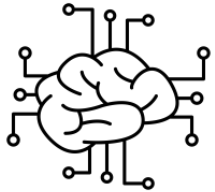
Grid Search to minimize MSE

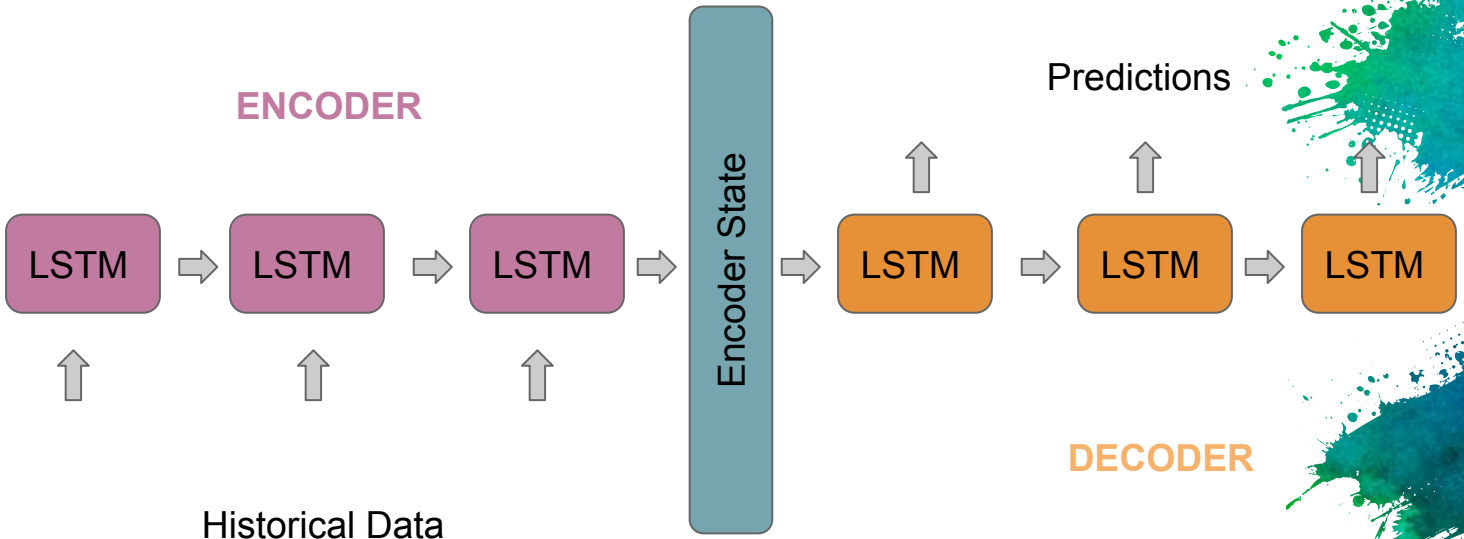
SARIMAX Model (1,1,1,0,1,2,24)

ARIMA 1 day

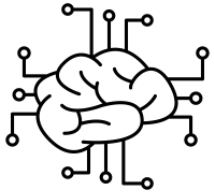

Real
Forecast
350
300
250
200
150
100
50
12-31 00
12-31 03
12-31 06
12-31 09
12-31 12
12-31 15
12-31 18
12-31 21
01-01 00

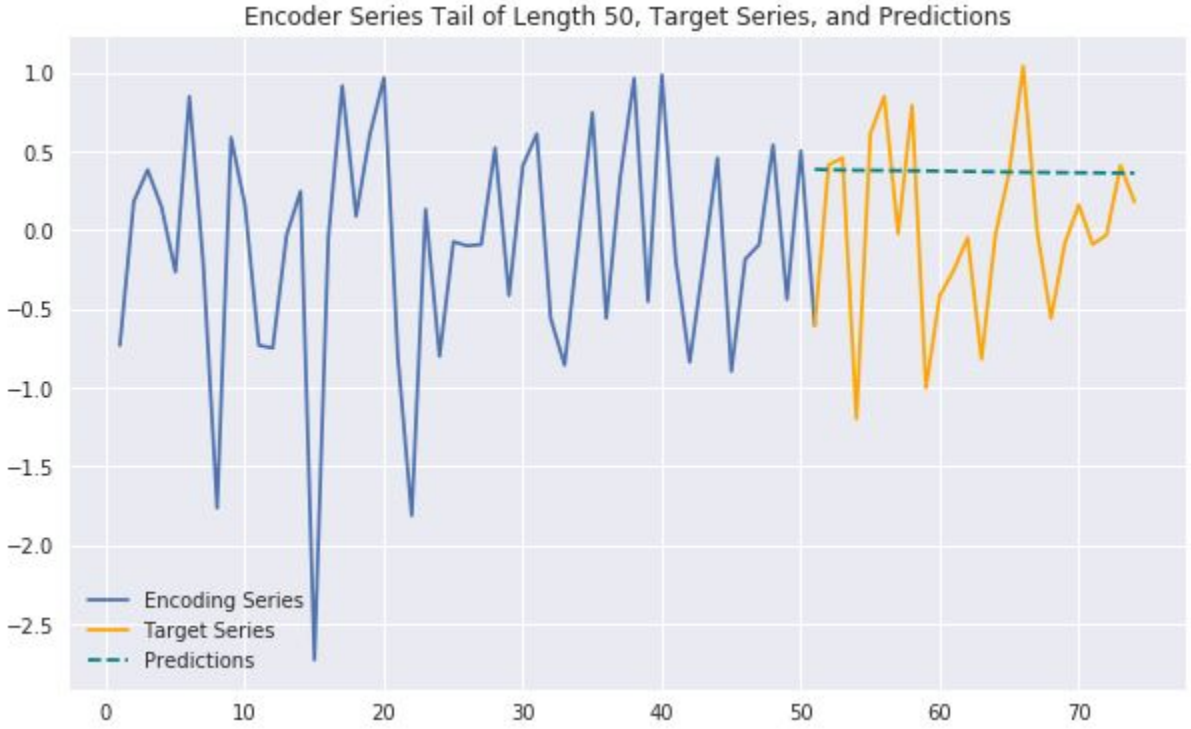
LSTM-seq2seq





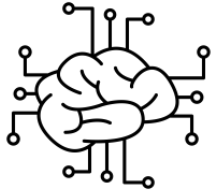
LSTM 1 day




Encoder Series Tail of Length 50, Target Series, and Predictions
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
0
10
20
30
40
50
60
70
Encoding Series
Target Series
Predictions

Future

- Larger dataset
- Multiple time-series
- Improve LSTM prediction quality



Thanks!

Any questions?

You can find me at:

@_filippo__medri

filippo.medri@gmail.com

filippomedri

filippomedri

Appendix

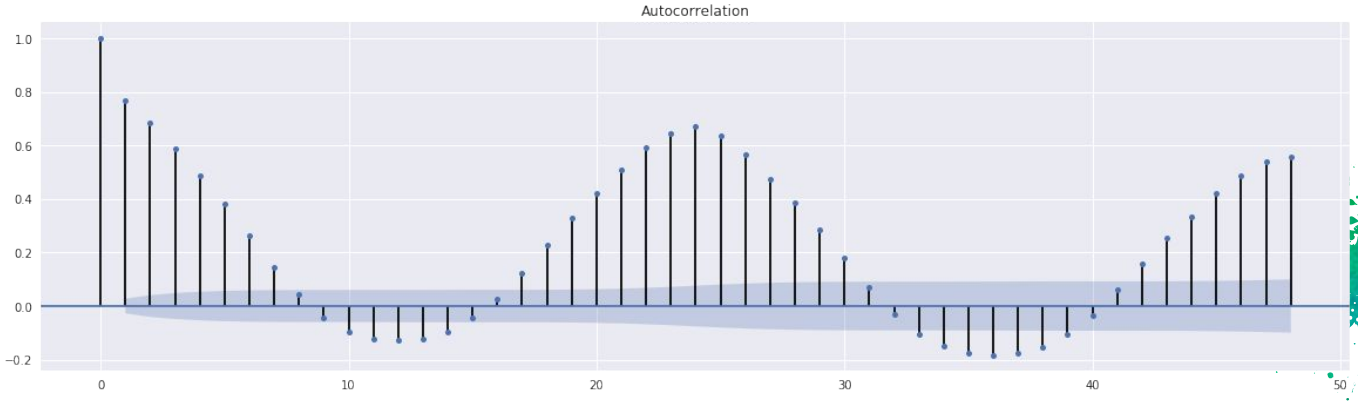
Dickey Fuller says ...

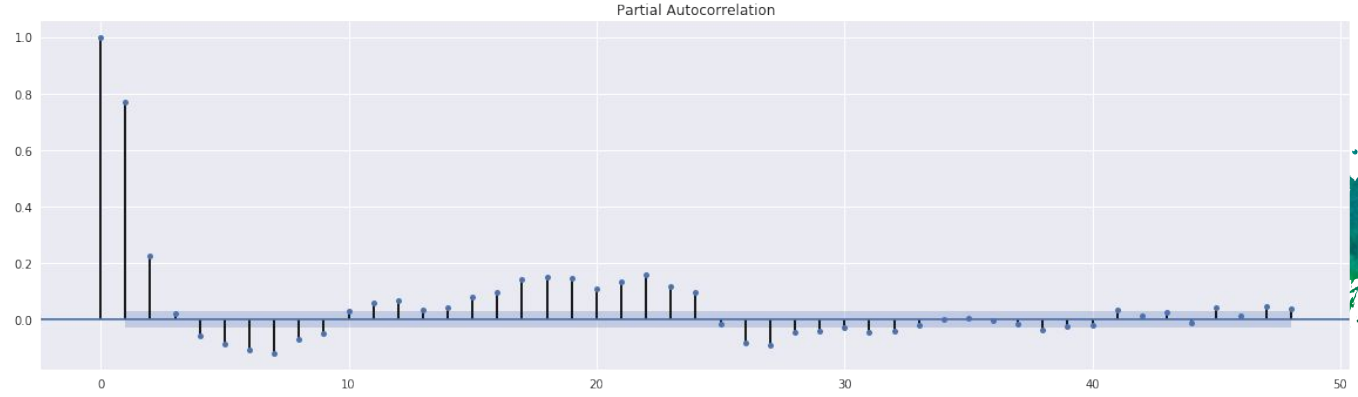
Test Statistic	p-value	#Lags Used
-6.06	**1.24e-07**	5.07e+03

Critical Value(1%)	Critical Value(5%)	Critical Value(10%)
-3.43	-2.86	-2.56


... Stationary


ACF & PACF

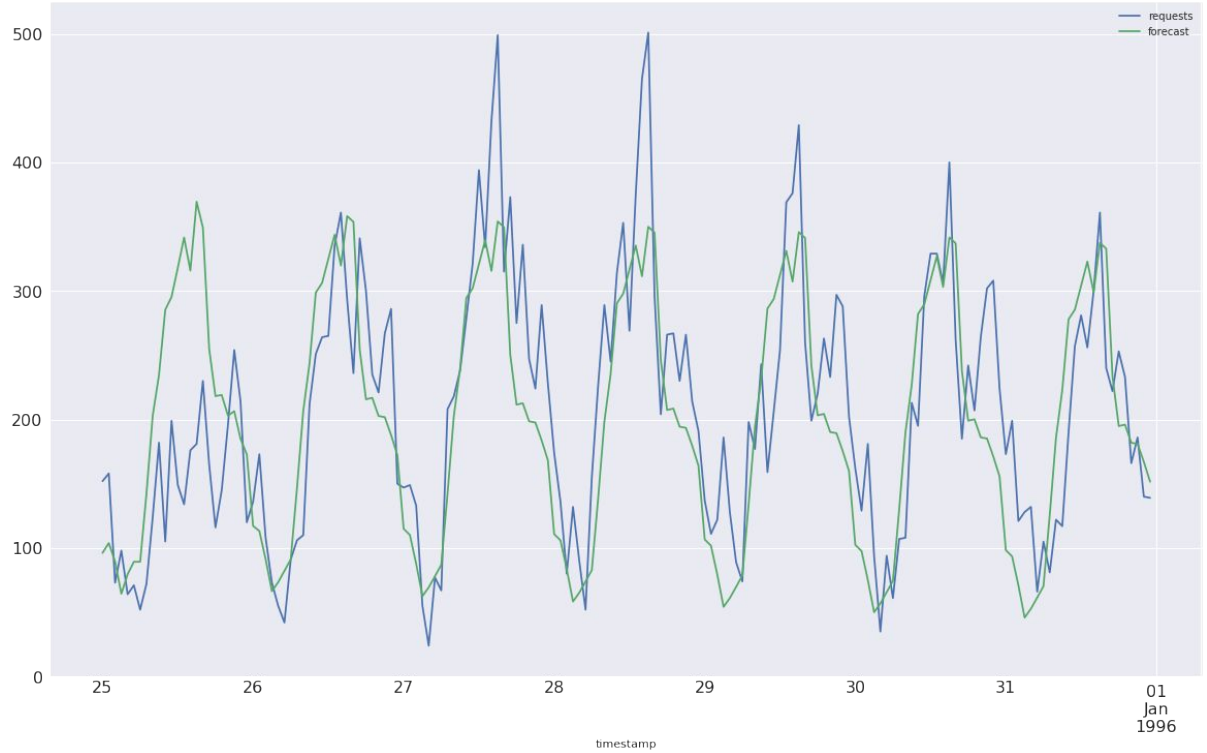

Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50


Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50

Sarimax Model	1 day MSE	1 week MSE
(1, 1, 1, 0, 1, 2)	2505.496	4936.831
(1, 1, 2, 0, 1, 1)	3669.022	5507.586
(1, 1, 2, 0, 1, 2)	3730.029	5555.638
(1, 1, 2, 1, 1, 2)	3733.279	5558.528
(1, 1, 2, 1, 1, 1)	3747.322	5758.138
(1, 1, 1, 0, 1, 1)	3807.95	7289.025
(1, 1, 1, 1, 1, 1)	3824.13	8358.719

Top 7 SARIMAX Models (24 h as seasonal)

ARIMA 1 week


requests
forecast
500
400
300
200
100
0
25
26
27
28
29
30
31
01
Jan
1996
timestamp

"If someone asks me what cloud computing is, I try not to get bogged down with definitions.

I tell them that, simply put, cloud computing is a better way to run your business."

Marc Benioff