# Meta-learning for time series forecasting in application to ATM load time series

**Presented by:**

Ilia Zherebtsov   Makar Korchagin   Rinat Prochii   Folu Obidare   Nikita Burtsev

Moscow, 2024

Skoltech

# What is ATM forecasting?



ATM forecasting - is the process of predicting how much cash will be needed at an ATM in the future,
i.e. is predicting future values of a time series


Importance of ATM forecasting:

- Minimizing cash outages (empty ATMs)
- Optimizing cash management
- Strategic planning
- Improving Security Measures

# Problem statement

Team 9

- Time series are very complex data and as a result a suitable model must be used to forecast

- Picking a suitable model is **time consuming** and **computationally expensive** - especially with many choices of models and large number of time series to predict.

- We use meta-learning to predict the best suited model to predict time series saving a lot of time.

# Meta-Learning Technique



There is a dataset with results of applying every model to every time series

Extract best results for each time series from the the dataset and then extract features from every time series

Automate model selection for each individual time series by matching time series features with appropriate forecasting models, aiming to optimize forecast accuracy by selecting the most suitable model

# Related work: Meta-learning approach



## Feature Forecasting model averaging (FFORMA)

T. Talagala, Rob J Hyndman, G. Athanasopoulos. Meta-learning how to forecast time series // Journal of Forecasting/03.01.2020
DOI: 10.1016/j.ijforecast.2019.02.011

## Feature Forecasting model performance prediction (FFORMPP)

Thiyanga S. Talagala∗ , Feng Li† , Yanfei Kang. FFORMPP: Feature-based forecast model performance prediction // Journal of Forecasting/08.2021
DOI: 10.1016/j.ijforecast.2021.07.002

## Feature Forecasting model selection (FFORMS)

T. Talagala, Rob J Hyndman, G. Athanasopoulos. Meta-learning how to forecast time series // Journal of Forecasting/09.02.2023
DOI: 10.1002/for.2963

# Dataset and time series overview

Team 9

## DataFrame

| Time | Metric | TimeSeries | Split | Model |
|---|---|---|---|---|
| train and forecast | MAE<br>MSE<br>RMSE<br>MASE<br>RMSSE<br>MAPE<br>SMAPE | name of time series | test or validation | 24 models in total |

## Time series archive

3 folders with time series

each time series has 2 columns: date and the ATM load value

in total 874 files

# Data preprocessing: 1st step

Team 9

## Time Series Contecanation

from many time series

- danish_atm_daily_0
- danish_atm_daily_1
- danish_atm_daily_2
- danish_atm_daily_3
- danish_atm_daily_4
- danish_atm_daily_5
- danish_atm_daily_6

In total 874 time series

to three tables with all time series

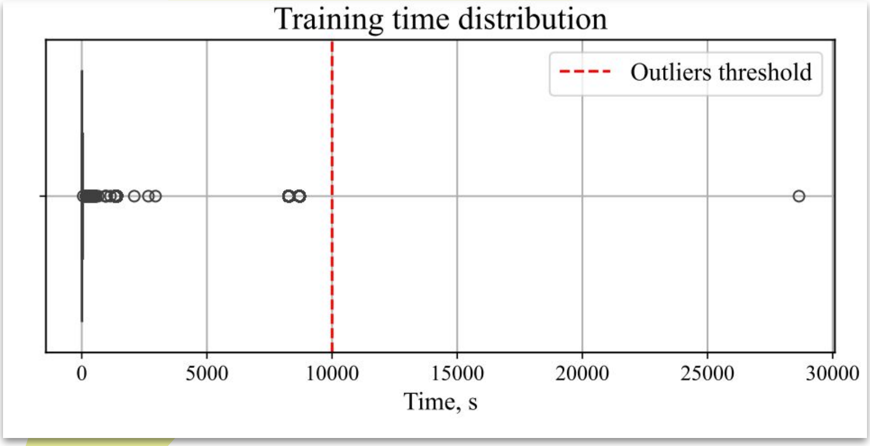
| danish_atm_daily_0 | danish_atm_daily_1 | danish_atm_daily_10 | danish_atm_daily_100 |
|---|---|---|---|
| 68 | 126 | 73 | 62 |
| 130 | 0 | 102 | 50 |
| 142 | 0 | 100 | 51 |
| 101 | 144 | 95 | 37 |
| 120 | 112 | 83 | 34 |
| 124 | 119 | 107 | 32 |
| 100 | 93 | 86 | 22 |
| 94 | 79 | 62 | 30 |
| 88 | 94 | 102 | 36 |
| 98 | 108 | 83 | 42 |

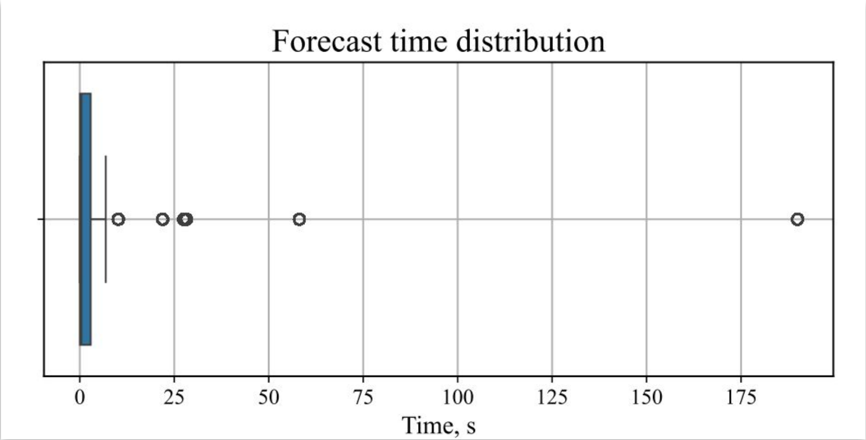
Missed values was filled using linear interpolation

# Data preprocessing: 2nd step



## Removing time outliers

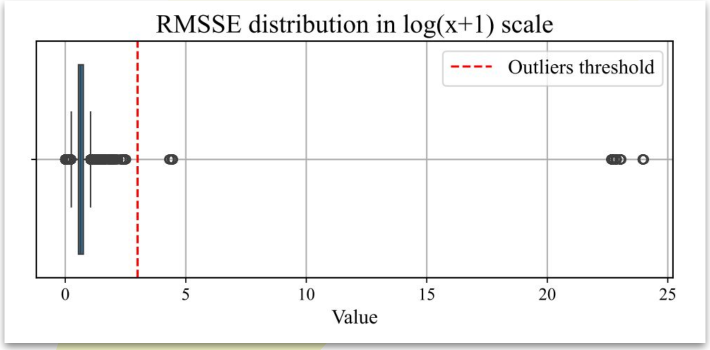


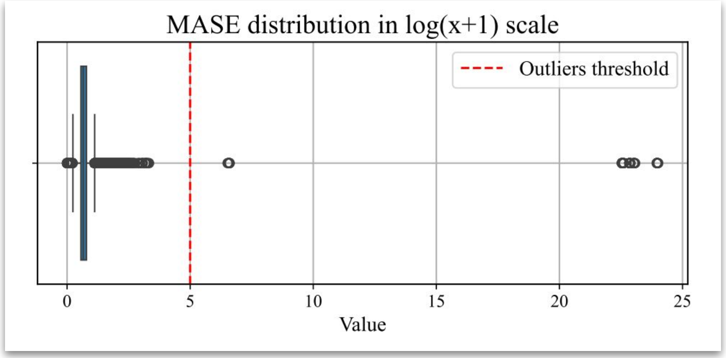



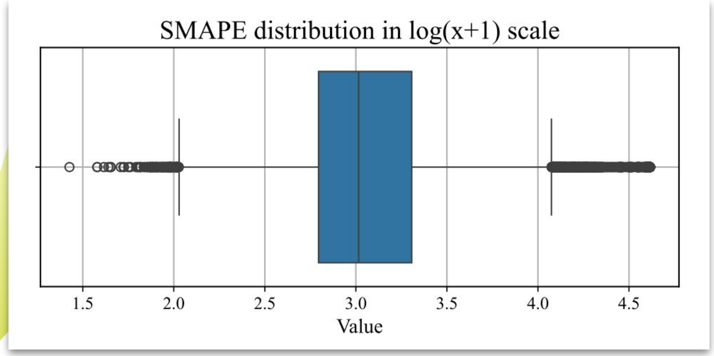
# Data preprocessing: 2nd step

Team 9

## Removing metric's outliers




MAPE was not considered because of being ill-conditioned for some cases

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|Y_t - \hat{Y}_t|}{|Y_t|}$$

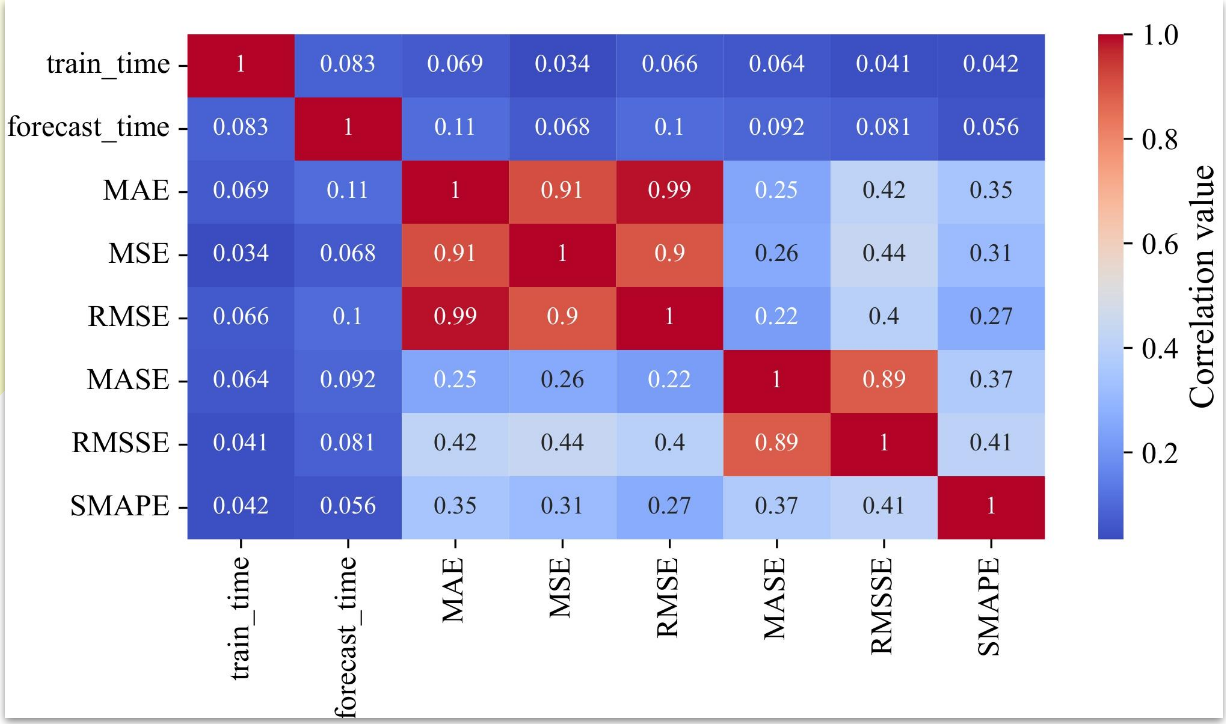


Skoltech

# Data preprocessing: 3rd step



## Finding correlations between metrics



# Feature extraction



| | tsFresh library | getML library | Manual search: tsFeatures library |
|---|---|---|---|
| **Advantages** | • automatic feature extraction process<br>• wide range of features | • automatic feature extraction process<br>• wide range of features<br>• scalability | • feature extraction control<br>• interpretable |
| **Limitations** | • high dimensionality<br>• features might lead to redundancy and overfitting<br>• some features difficult to interpret | • needed special structure of data<br>• uninterpretable features | • not automated |

# Choosing metric for meta-learner

Team 9

Metric for ATM forecasting should has following qualities:

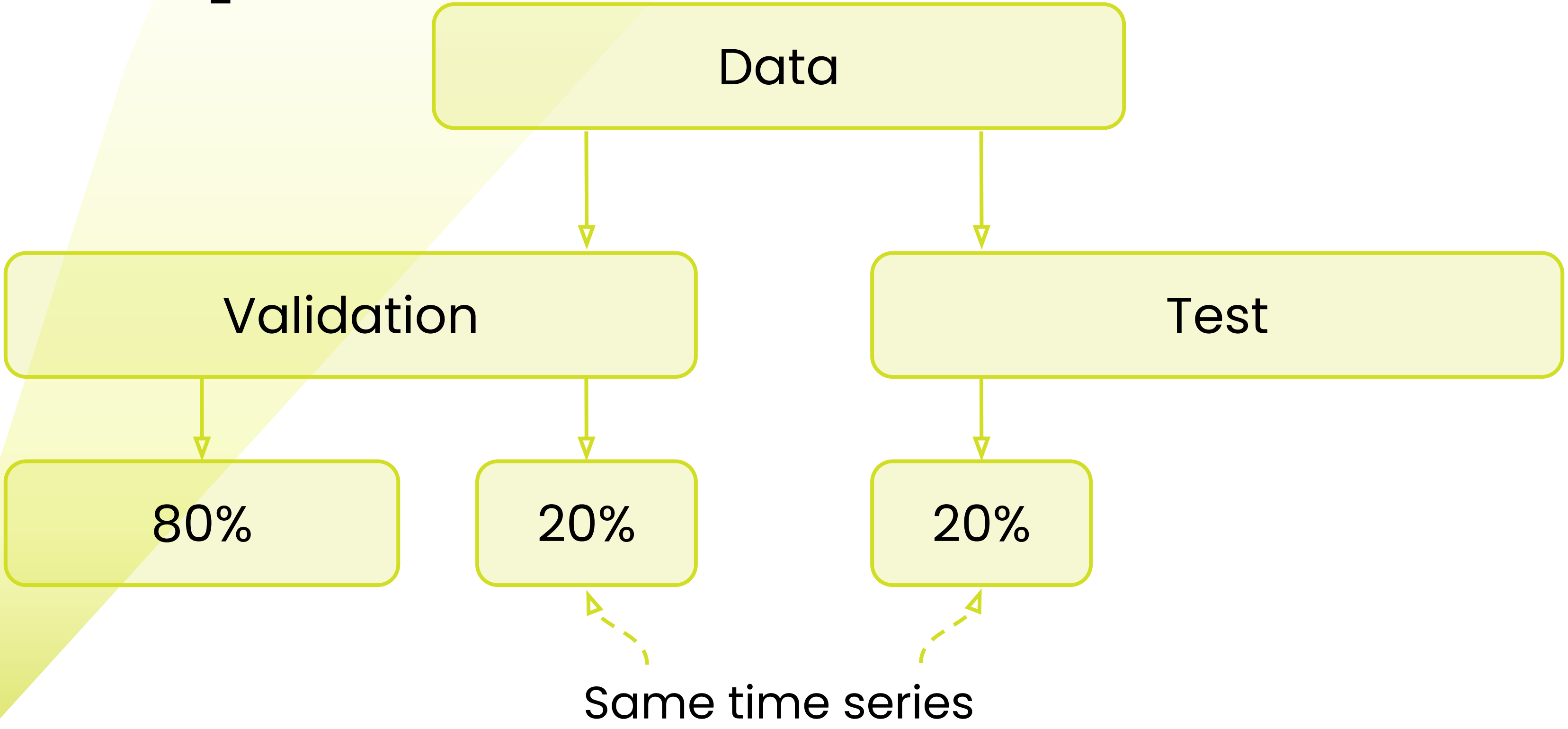
- Not sensitive to scale of time series
- Interpretability
- Sensitive to large errors
- Incorporate forecast horizon

***RMSSE*** (Root Mean Squared Scaled Error)

$$\text{RMSSE} = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{\frac{1}{h}\sum_{t=1}^{n}(Y_t - Y_{t-1})^2}}$$

$Y_t$ - actual value of the time series at time *t*
$\hat{Y}_t$ - forecasted value of the time series at time *t*
*n* - total number of observations in the time series
*h* - forecast horizon

# Setup



# Baseline Experiments



| | Exp 1 | Exp 2 | Exp 3 |
|---|---|---|---|
| Model | TFTTuningObjective_gl | TFTTuningObjective_gl | N/A |
| Accuracy | 14.94% | 13.22% | 15.52% |
| Lost rate | 10.92% | 10.92% | 0 |
| RMSSE | 0.84 | 0.97 | 0.89 |

*accuracy - percentage of test it is correct to say that this model is the best
*lost rate - percentage of test we don't have results of this model
*RMSSE - average RMSSE

# Classifier Experiments

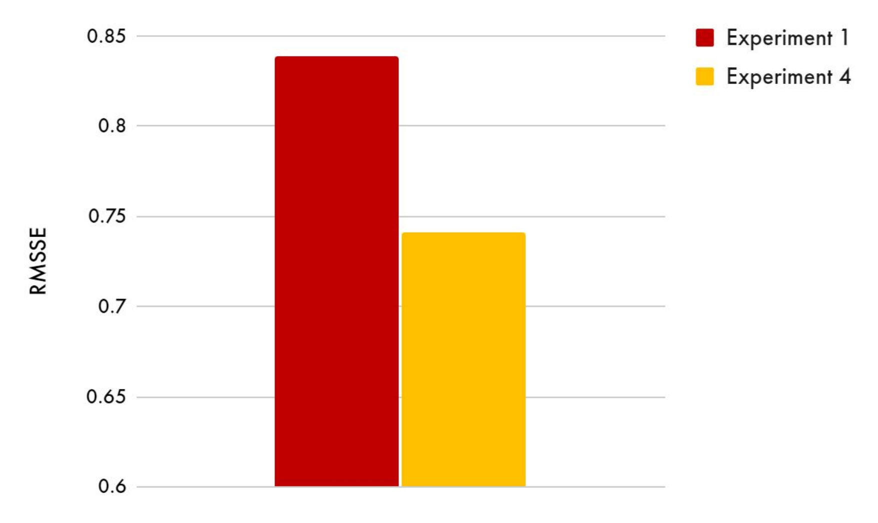


| | Classifier 1 | Classifier 2 |
|---|---|---|
| Accuracy | 24.14% | 24.14% |
| Lost rate | 0.57% | 0% |
| RMSSE | 0.74 | 0.91 |

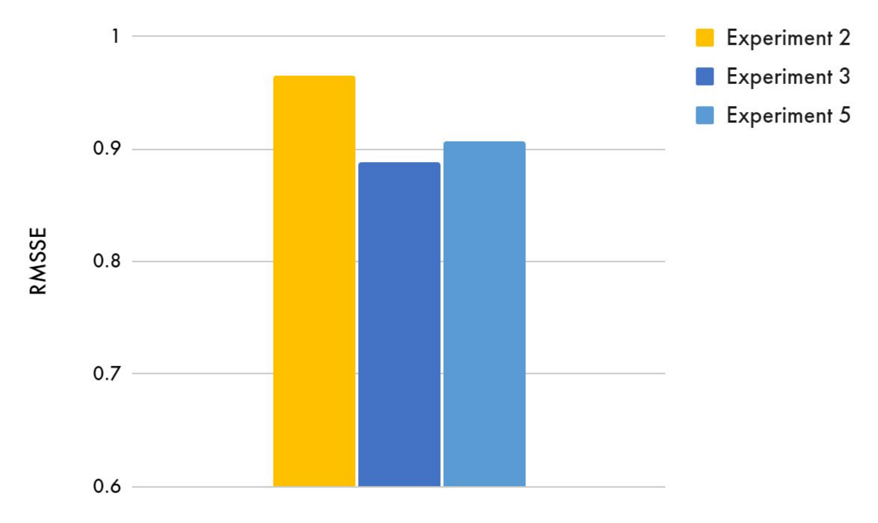
*accuracy - percentage of test it is correct to say that this model is the best
*lost rate - percentage of test we don't have results of this model
*RMSSE - average RMSSE

# Results





0.85
0.8
0.75
0.7
0.65
0.6
RMSSE
Experiment 1
Experiment 4

1
0.9
0.8
0.7
0.6
RMSSE
Experiment 2
Experiment 3
Experiment 5

# Conclusion



- Successfully built the **meta-learner** that **can outperform blind guess** of the best model for time-series forecasting
- **Results** of our solution are **very close to the ideal solution**, only 0.02 difference in RMSSE
- **Models works pretty fast**: it took only one second to be trained

# Our team

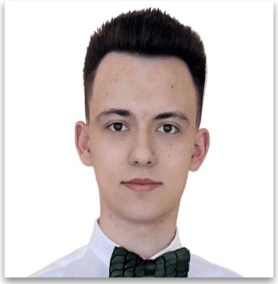
Team 9

**Nikita Burtsev**

Preparing presentation
Visual representation

**Rinat Prochii**

Feature extraction



**Ilia Zherebtsov**

Conducting experiments

**Makar Korchagin**

Data preprocessing

**Folu Obidare**

Literature review

# Thx!

Skoltech

# Feature extraction



## tsFresh library

Advantages:
- automatic feature extraction process
- wide range of features

Limitations:
- high dimensionality
- features might lead to redundancy and overfitting
- some features difficult to interpret

## getML library

Advantages:
- automatic feature extraction process
- wide range of features
- scalability

Limitations:
- needed special structure of data
- uninterpretable features

## Manual search: tsFeatures library

Advantages:
- feature extraction control
- interpretable

Limitations:
- not automated

# Setup



Test set of time series

| 80% | 20% |
| --- | --- |

Validation set of time series

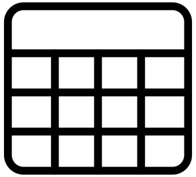
| 80% | 20% |
| --- | --- |

# Dataset and time series overview



**DataFrame**

| Time | Metric | TimeSeries | Model |
|---|---|---|---|
| train and forecast | MAE<br>MSE<br>RMSE<br>MASE<br>RMSSE<br>MAPE<br>SMAPE | name of time series | name of the model |



**Time series archive**

| nn5 | mipt_alpha | danish_atm |
|---|---|---|
| 111 files | 113 files | 650 files |