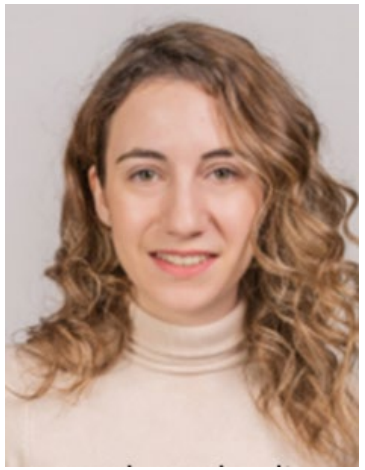
WOMEN IN DATA SCIENCE MAASTRICHT

#WiDSMaastricht2023

WiDS Datathon Maastricht 2023
Prize Ceremony

Organization Team:

Chang Sun
Postdoc
Researcher



Mado Ntekouli
PhD candidate

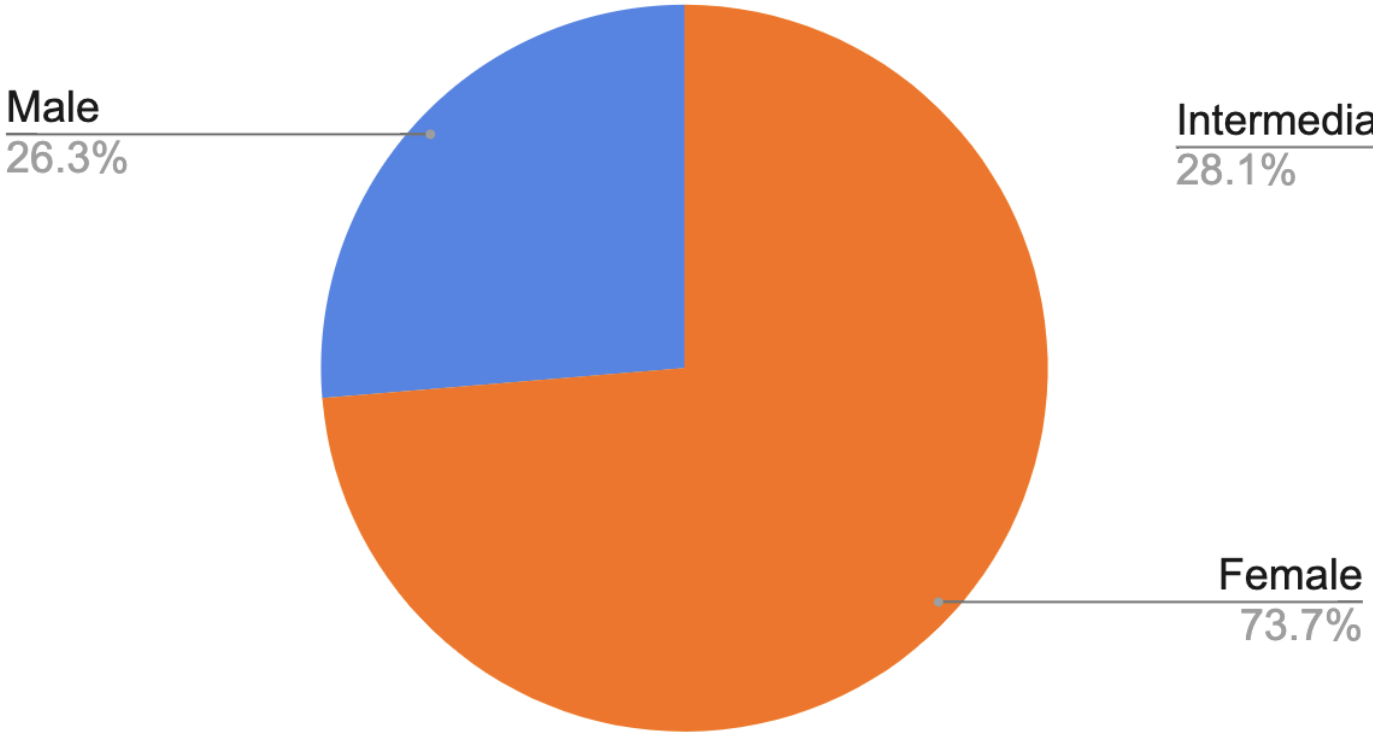
Yuyang Yan
PhD candidate

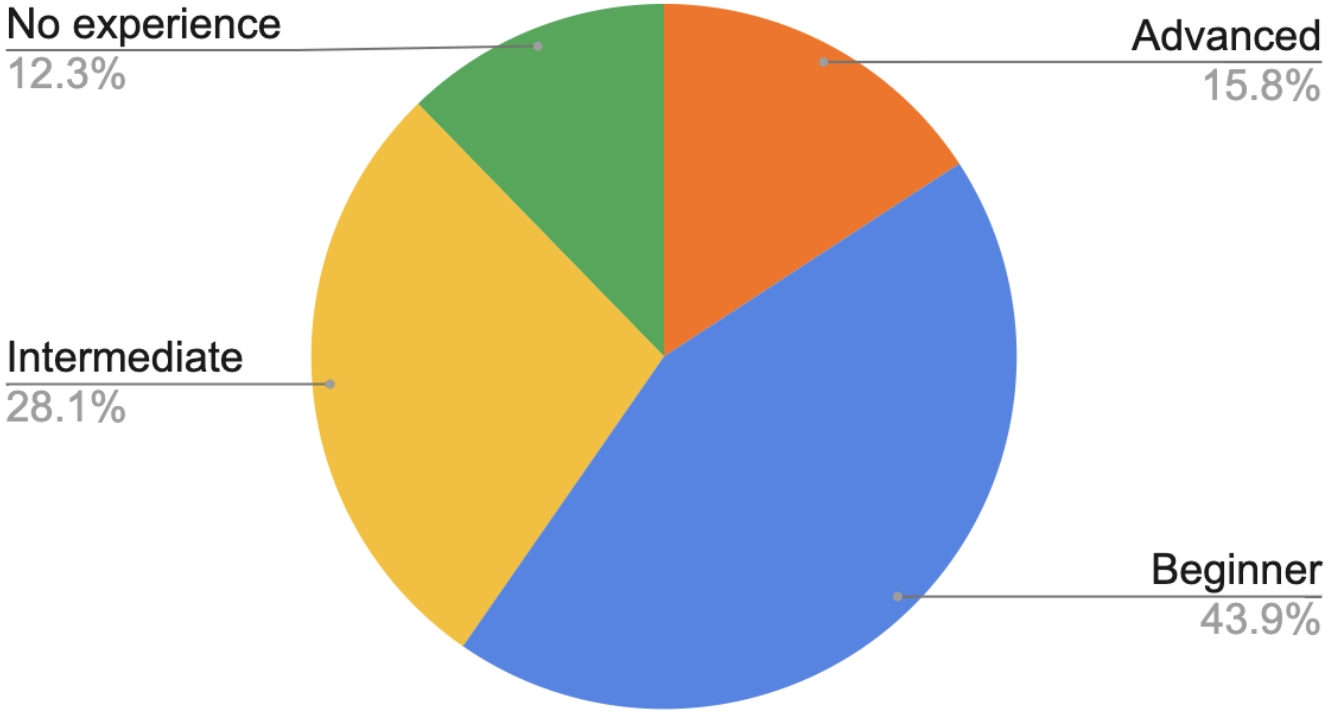
Over 40 participants
12 teams
from academia and industry

WOMEN IN DATA SCIENCE
MAASTRICHT

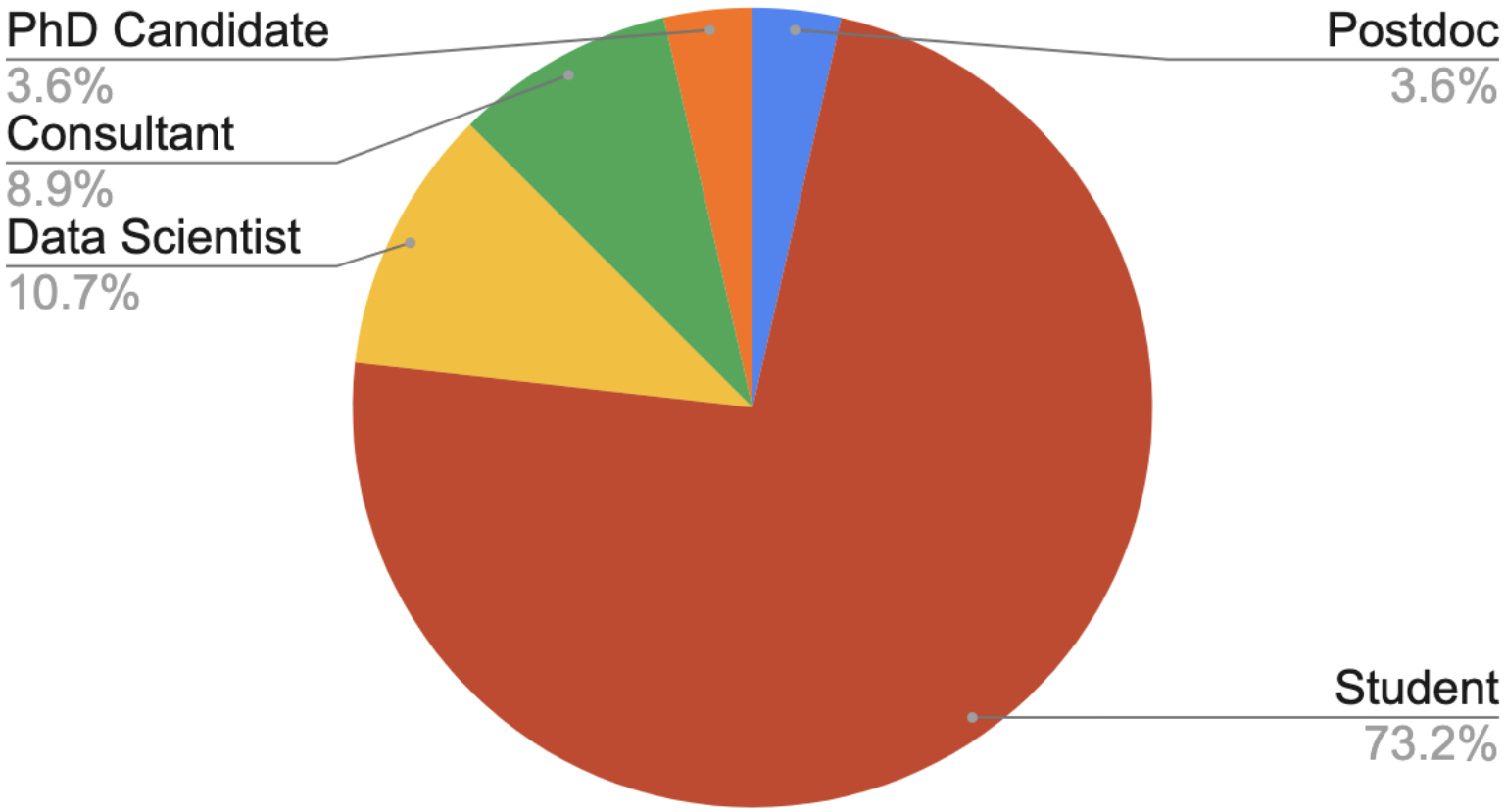
Maastricht University
Institute of Data Science

Datathon Facts - Gender and Data Science experience



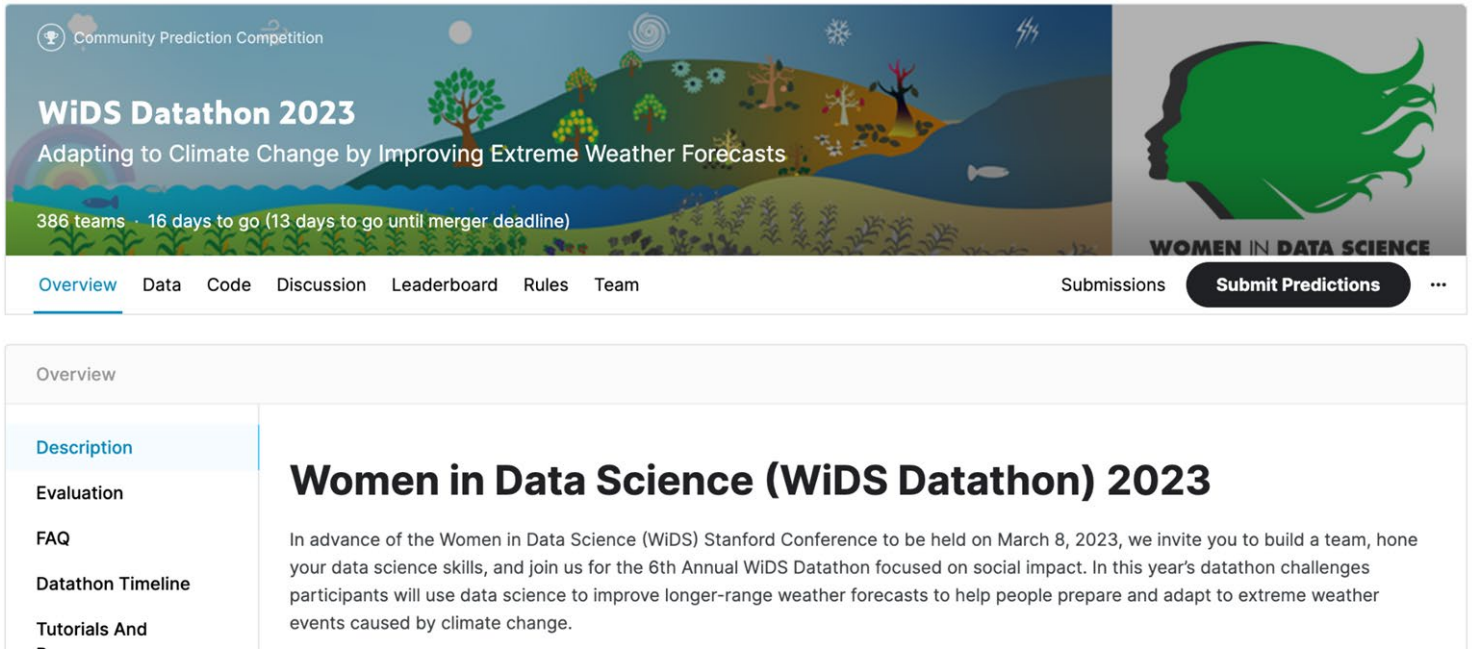


Datathon Facts - Positions



Datathon Highlights

Pre-training sessions for beginner



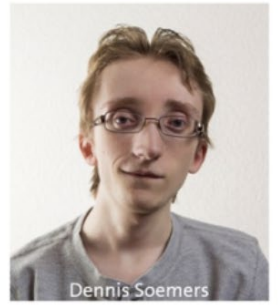
Datathon Highlights

Pre-training sessions for beginner

Support from data science experts

Stefan Meier
PhD candidate

Xu Wang
Postdoc Researcher



Dennis Soemers
PhD candidate

Thales Bertaglia
PhD candidate

Yaru Zhang
Master student

Pranav Bapat
Data Science Researcher

Datathon Highlights

- Pre-training sessions for beginner
- Support from data science experts
- Fitness and network sessions

Datathon Highlights

- Pre-training sessions for beginner
- Support from data science experts
- Fitness and network sessions
- Responsible data science pioneer team (4th Feb)

Responsible data science pioneer team (4th Feb)

WIDS MAASTRICHT DATATHON 2023

WOMEN IN DATA SCIENCE MAASTRICHT

Responsible Data Science Pioneer Team Selection Criteria

Model Performance (Point 1-3 the higher, the better)

Indicator of the accuracy of predictions.

1 point - Performance is in the top 50% of competitors

2 points - Performance is in the top 25% of competitors

3 points - Performance is in the top 5% of competitors

Originality/Creativity

Indicator of the extent to which the approach contains original or creative elements

1 point - The approach makes use of a basic machine learning pipeline

2 points - The pipeline has been adapted with state-of-the-art methods

3 points - The approach contains novel components whose performance was evaluated

Transparency and Explainability

Indicator of the ability of the team to describe the model and explain its performance

1 point - The team is able to explain the model and interpret the dataset

2 points - The team is able to explain how the model is built but not clear in reasoning

3 points - The team is able to clearly explain the model and rationale for different aspects

Bias and fairness

Indicator of the extent to which aspects of bias and fairness are addressed

1 point - Bias and fairness are minimally addressed

2 points - Issues of bias and fairness are identified

3 points - The approach actively attempts to correct for bias and/or maximize fairness

Pitch

Indicator of the overall quality of the presentation

1 - Presentation is not clear and the team is unable to answer the questions posed

2 - The team delivers a clear presentation and/or is able to answer questions posed

3 - The team delivers a clear and engaging presentation and competently answers all questions

Jury Group

Linda Rieswijk

Visara Urovi

Parveen Kumar

Yenisel Plasencia-Calaña

Today, we are going to announce…

- Pre-training sessions for beginner
- Support from data science experts
- Online Yoga and social activities
- Responsible data science pioneer team
- **Datathon Maastricht Winner**

WIDS Datathon Maastricht Winner is …

Genetic Mess

Ping Cao

Darina Obukhova

Thank you all the participants, tutors and speakers, and our partners and supports!

Maastricht University

Institute of Data Science

Maastricht University

Department of Advanced Computing Sciences

Maastricht University

UM SPORTS

Female Empowerment

Maastricht University

BISS

Maastricht University

Datathon Maastricht Winner Team

Ping Cao, Darina Obukhova

PhD student, MUMC+, Maastricht University

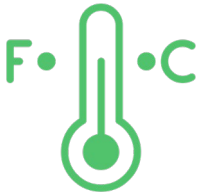
Data Overview

train_data.csv

test_data.csv

(375734, 246)

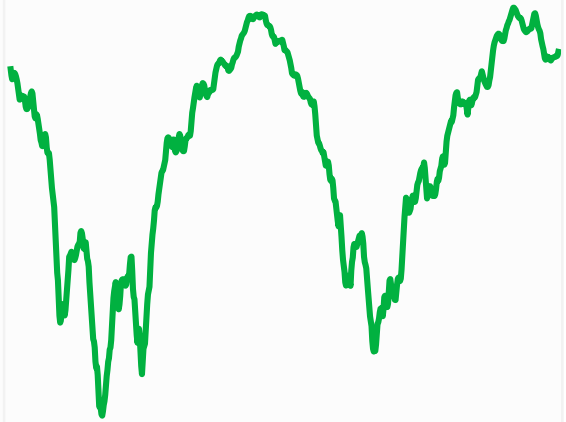
(31354, 245)



Predicting temperature over the next 14 days, for each location and start date.

Training Data

Test Data



2014 2015 2016 2017 2018 2019 2020 2021 2022

1. Data Exploring

Data structure

Missing/outlier

2. Data cleaning

Imputation

Correlation

Variance/bias

3. Feature engineering

Encoding

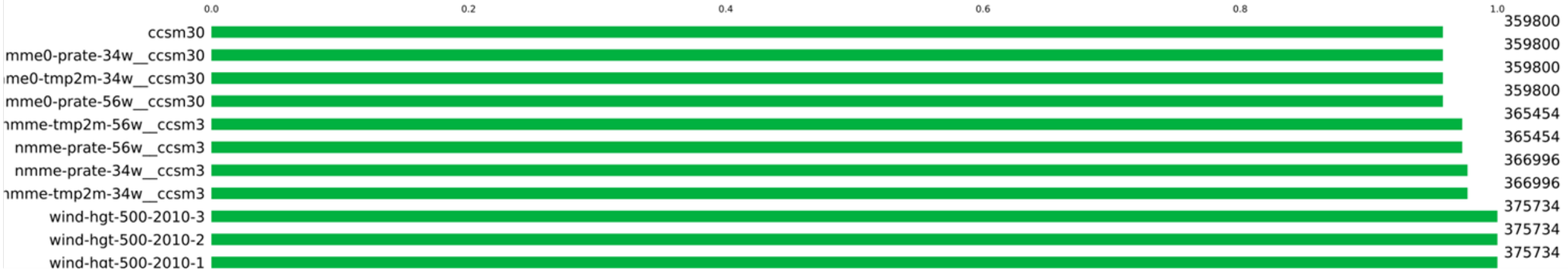
Add features

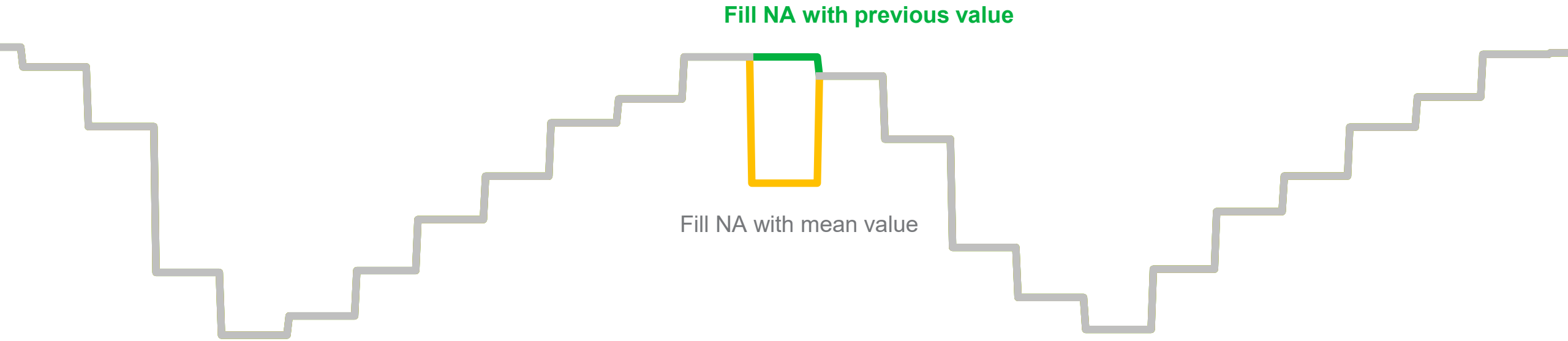
Feature selection

4. Modelling

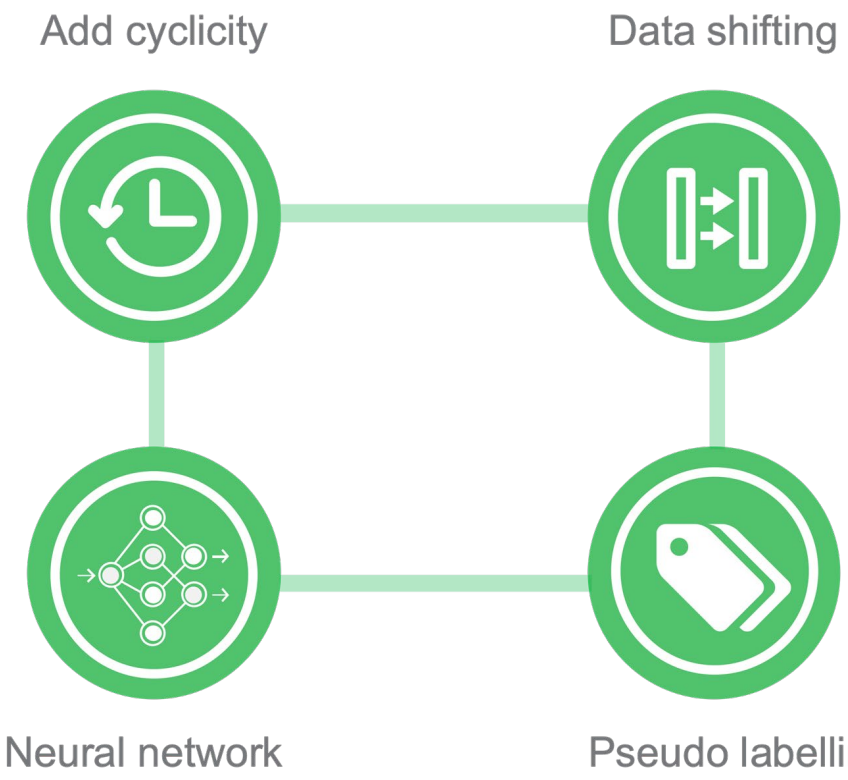
Baseline

Ensemble


0.0
0.2
0.4
0.6
0.8
1.0
ccsm30
mme0-prate-34w__ccsm30
me0-tmp2m-34w__ccsm30
mme0-prate-56w__ccsm30
mme-tmp2m-56w__ccsm3
nmme-prate-56w__ccsm3
nmme-prate-34w__ccsm3
mme-tmp2m-34w__ccsm3
wind-hgt-500-2010-3
wind-hgt-500-2010-2
wind-hgt-500-2010-1
359800
359800
359800
359800
365454
365454
366996
366996
375734
375734
375734


Fill NA with previous value
Fill NA with mean value

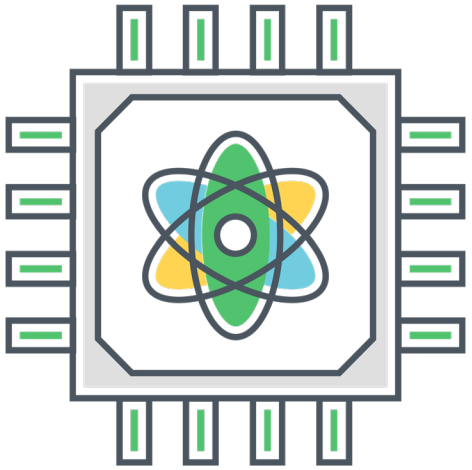
After Maastricht Datathon Workshop



RMSE 0.79

RMSE 1.02

PhD Project Relevance



Data process pipeline
Programming ability

Time series data embedding
Time series data on cell development
Data integration

Search information more
Explore the data more
Persistence

Optimism
Share ideas
Keep learning

Thank You!