# AI/ML in the Era of Climate Change

Group 10

Project 2: Presentation

---

*Vienna, 11.01.2023*

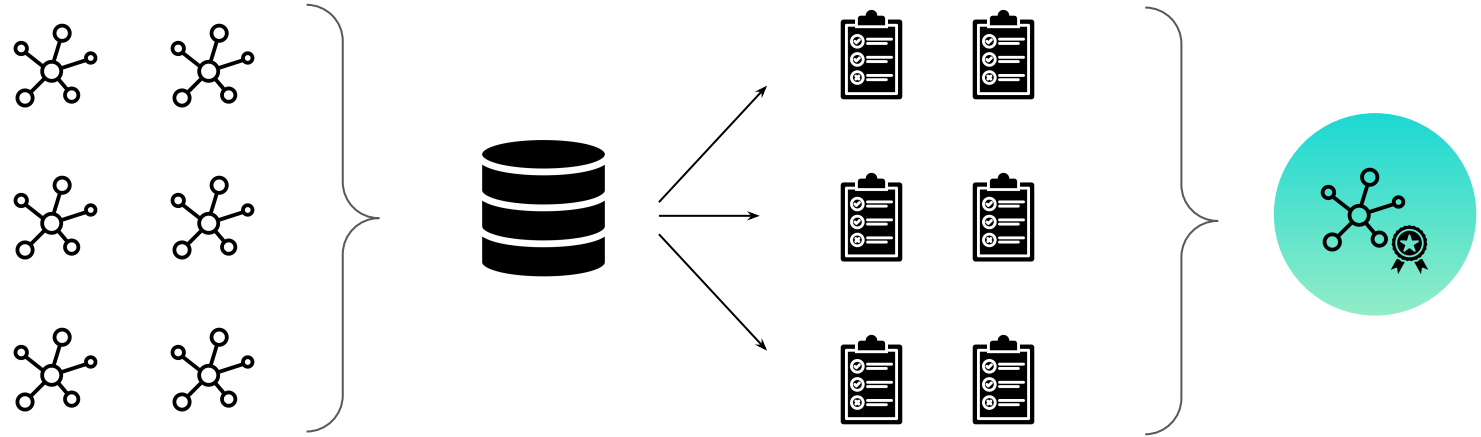
# Agenda

- Research Objective
- Experiment Results
- Conclusion

# Agenda

- Research Objective
- Experiment Results
- Conclusion

# 6 models were assessed with a data set to **find the best performing one** and optimize it



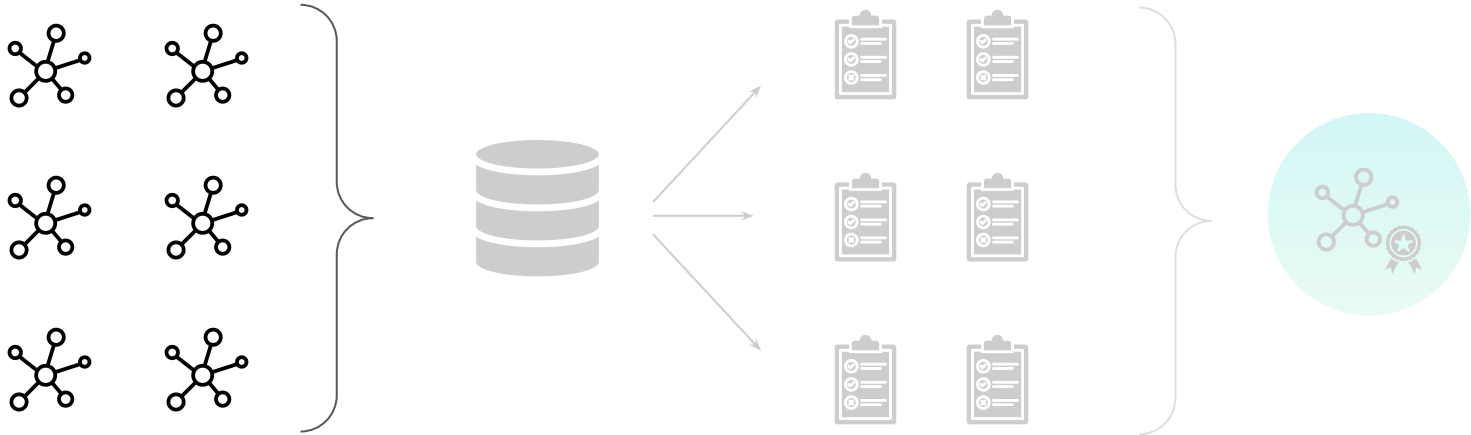
**6 Models**
for comparison

**Data**
to train and test on

**Test results**
to find best performing model

**Winning Model**
to be fully optimized

# 6 models were assessed with a data set to **find the best performing one** and optimize it



**6 Models**
for comparison

**Data**
to train and test on

**Test results**
to find best performing model

**Winning Model**
to be fully optimized

# The models were selected to cover a **wide range of regression model types**

Model 1: **Linear Regression**

```
lr = LinearRegression()
```

Model 2: **Linear SVR**

```
lsvr = LinearSVR(max_iter=100)
```

Model 3: **Random Forest**

```
rf = RandomForestRegressor(n_estimators=10)
```

Model 4: **XGBoost**

```
xgb = XGBRegressor(n_estimators=10)
```

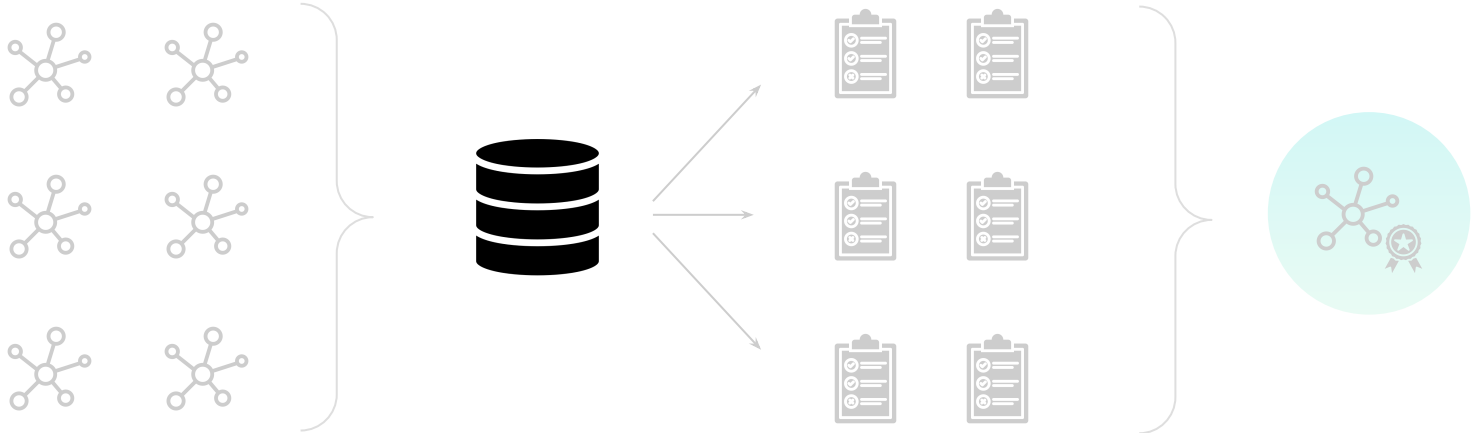
Model 5: **Neural Net**

```
nn = NeuralNet()
```

Model 6: **Deep Neural Net**

```
dnn = DeepNeuralNet()
```

# 6 models were assessed with a data set to **find the best performing one** and optimize it



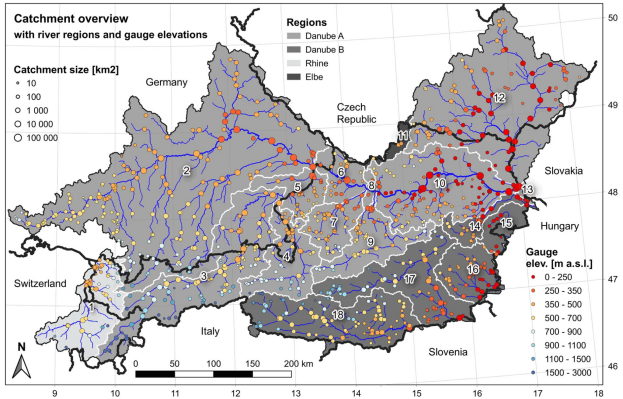
**6 Models**
for comparison

**Data**
to train and test on

**Test results**
to find best performing model

**Winning Model**
to be fully optimized

# The experiment was done using the **LamaH-CE data set**




Data:

**LamaH-CE**

(Hydrology and Environmental Sciences Data for Central Europe)

- **Total Size:** 859 gauged catchments (100 used)
- **Time span:** >35 years
- **Target variable:** precipitation

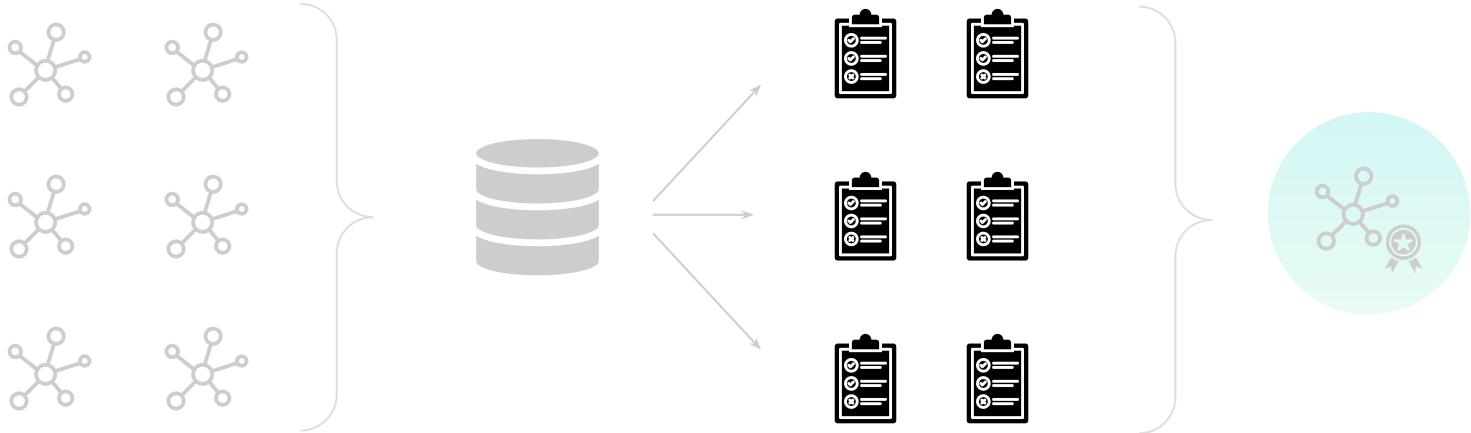
# Agenda

Research Objective

**Experiment Results**

Conclusion

# 6 models were assessed with a data set to **find the best performing one** and optimize it



**6 Models**
for comparison

**Data**
to train and test on

**Test results**
to find best performing model

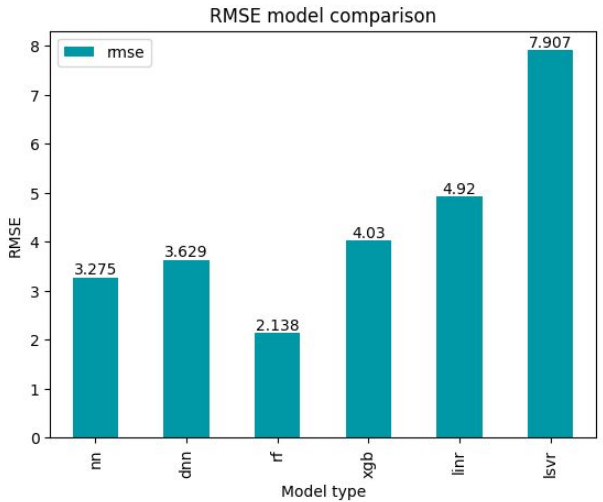
**Winning Model**
to be fully optimized

# The results show that **Model 3 (RF) scored the lowest RMSE** the test data and was therefore further optimized

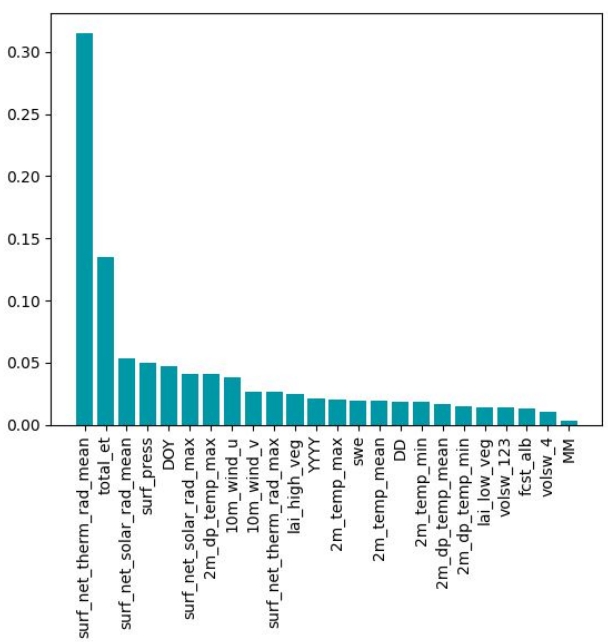
| | RMSE |
|---|---|
| Model 1: **Linear Regression** | 4.92 |
| Model 2: **Linear SVR** | 7.91 |
| Model 3: **Random Forest** | **2.14** |
| Model 4: **XGBoost** | 4.03 |
| Model 5: **Neural Net** | 3.23 |
| Model 6: **Deep Neural Net** | 3.63 |

# The results show that **Model 3 (RF) scored the lowest RMSE** the test data and was therefore further optimized

| | RMSE |
|---|---|
| Model 1: **Linear R** | 4.92 |
| Model 2: **Linear S** | 9.49 |
| Model 3: **Random** | 2.14 |
| Model 4: **XGBoos** | 4.03 |
| Model 5: **Neural N** | 3.23 |
| Model 6: **Deep Neural Net** | 3.63 |



# The random forest model shows the **heavy weight of the first two features**, followed by a long tail of the remaining ones




**Random Forest**

*Feature Importance Takeaways:*

1. Strong focus on **two features**
2. **Long tail** of remaining, less important features

# The RF Regressor was tried to further optimize using **randomized search cross-validation**

Model 7: **Random Forest**

**OPTIMIZED**

- Started with grid search, switched to **randomized search** (runtime reasons)
- Optimized for **three parameters:**
  - Parameter 1: **n_estimators**
  - Parameter 2: **max_depth**
  - Parameter 3: **bootstrap**

# The overview shows that the further optimized **random forest regressor could not be further improved**

| | RMSE |
|---|---|
| Model 1: **Linear Regression** | 4.92 |
| Model 2: **Linear SVR** | 7.91 |
| Model 3: **Random Forest** | 2.14 |
| Model 4: **XGBoost** | 4.03 |
| Model 5: **Neural Net** | 3.23 |
| Model 6: **Deep Neural Net** | 3.63 |
| Model 7: **Random Forest** OPTIMIZED | **2.14** |

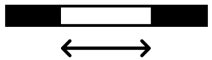
# Agenda

Research Objective

Experiment Results

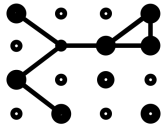
**Conclusion**

# The experiment once again showed how important model selection can be: **great differences in performance** could be observed and **the complex models did not perform best**



## Big range in performance

Different models' performance can vary greatly which once again **underlines the importance of selecting the right model**



## Simplicity (sometimes) wins

Though sometimes more complex models (such as neural nets) perform best, very **often simpler models actually work better** - and should be preferred

Thank you for your attention - **any questions?**

# Q&A