Air Quality Forecast for Kampala / Uganda

AIr – we bring AI to your air

AIr – we bring AI to your air

Sebastian

Data Scientist

Background in biomedical science

Matthias

Data Scientist

Experience in Six Sigma, Project Management, Electrical Engineering

Samer

Agronomist

Experience in Remote Sensing and Data Science

Jess

Nature Conservationist

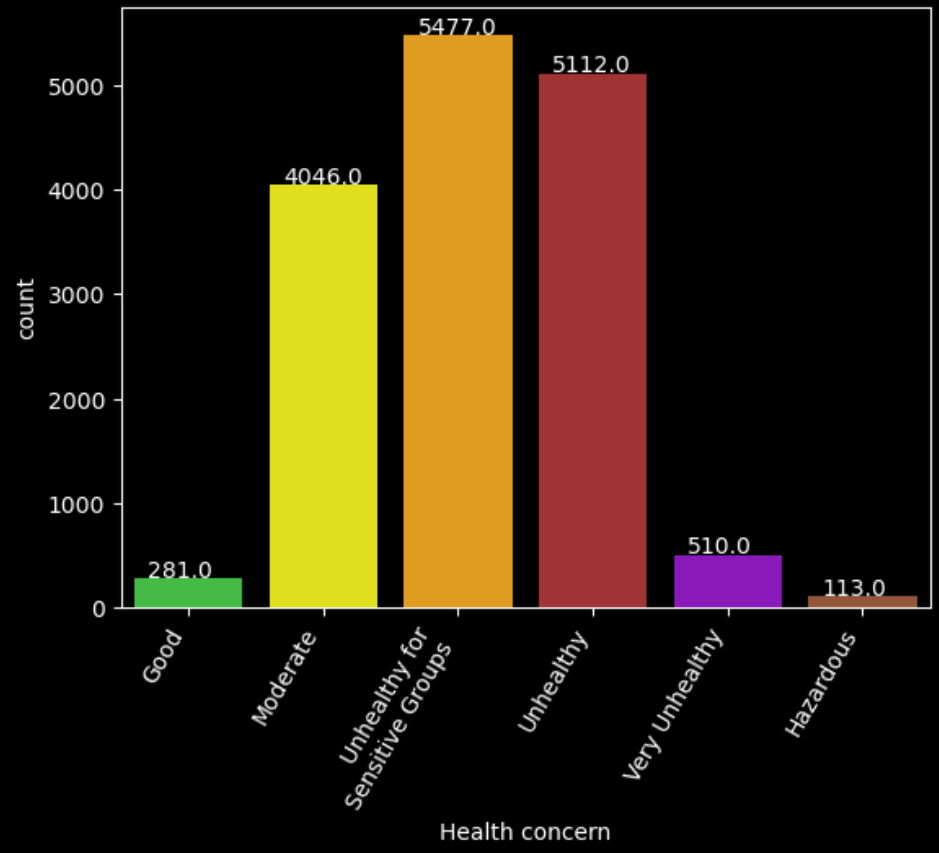
Experience in Data Science, Python, R

Problem Definition

Problem Definition

- World wide 3.2 million deaths because of air pollution (2020)
  - Uganda: Around 73.000 deaths in 2016
  - Kampala 2nd place of African cities
  - Problem is getting worse

- Root causes of high air pollution:
  - burning waste, urbanization, biomass fueled cooking, cars



Project Charter I

General information

- **Stakeholder:** AirQ
- **Challenge:** Air Quality Forecast for Kampala / Uganda

Challenge Description

- Predict air quality level exactly 24 hours after last measurement
- Using 5 days worth of hourly measured weather data from 5 different locations

Goal & Benefit

- Prediction of air pollution allows for safety alerts to be send out to citizens
- Save lives

Project Charter II

In scope	Out of scope	Deliverables
• Only 5 out of 65 sensor locations in Uganda • Prediction Model • Error Analysis • Code	• Using meta data of the sensors	• Presentation • Jupyter Notebook • Model

Data

Data set

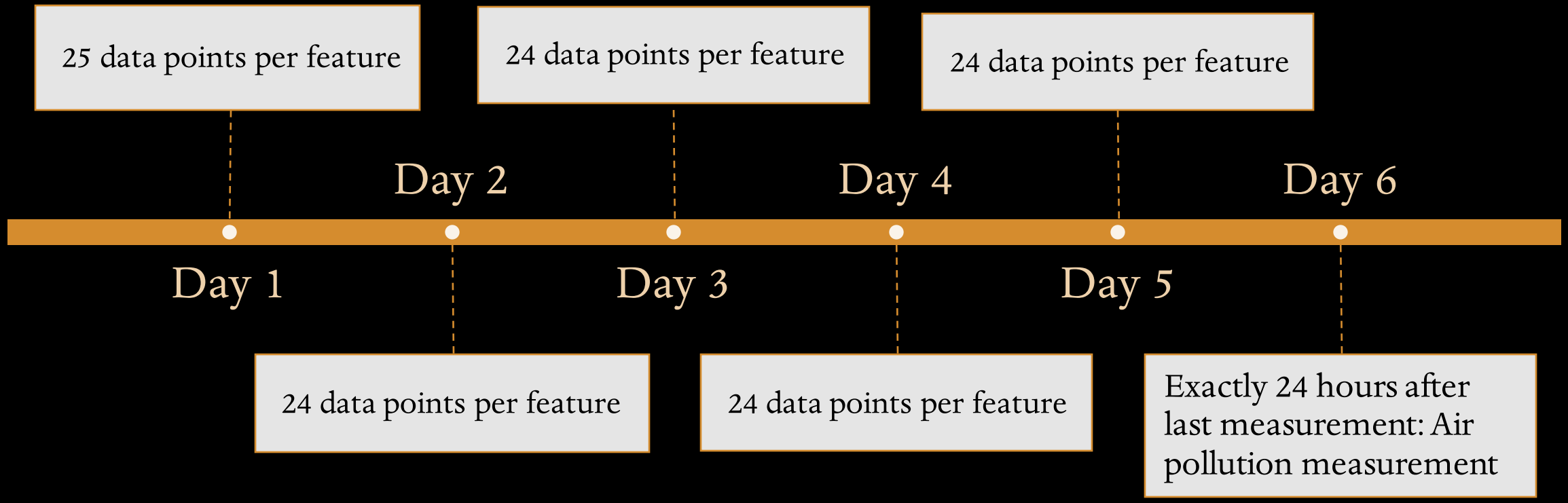
Data

- 5 sensor locations near Kampala
- 15.000 observations with 6 features
  - Temperature, precipitation, humidity, wind speed and direction, atmospheric pressure
- Cyclic data (1 per hour) for 5 days in a row (total 121 observations per feature)
- Air pollution measurement at hour 145

Limitations

- No timestamps
- No sensor meta data, e.g. geographical location
- Missing Data

Observation Structure



Modelling & Optimization

Comparison Models

Baseline model

- Linear regression
- Mean of the 121 values of temperature and wind speed
- Test set RMSE: 43.45

Best Model

- XGBoost Regressor
- Test set RMSE: 28.48

Model

Input:

Native data:

121 measurements

temperature,

precipitation,

humidity,

wind speed and direction,

atmospheric pressure

Feature engineering:

Sensor location

Daily means of all features

Model:

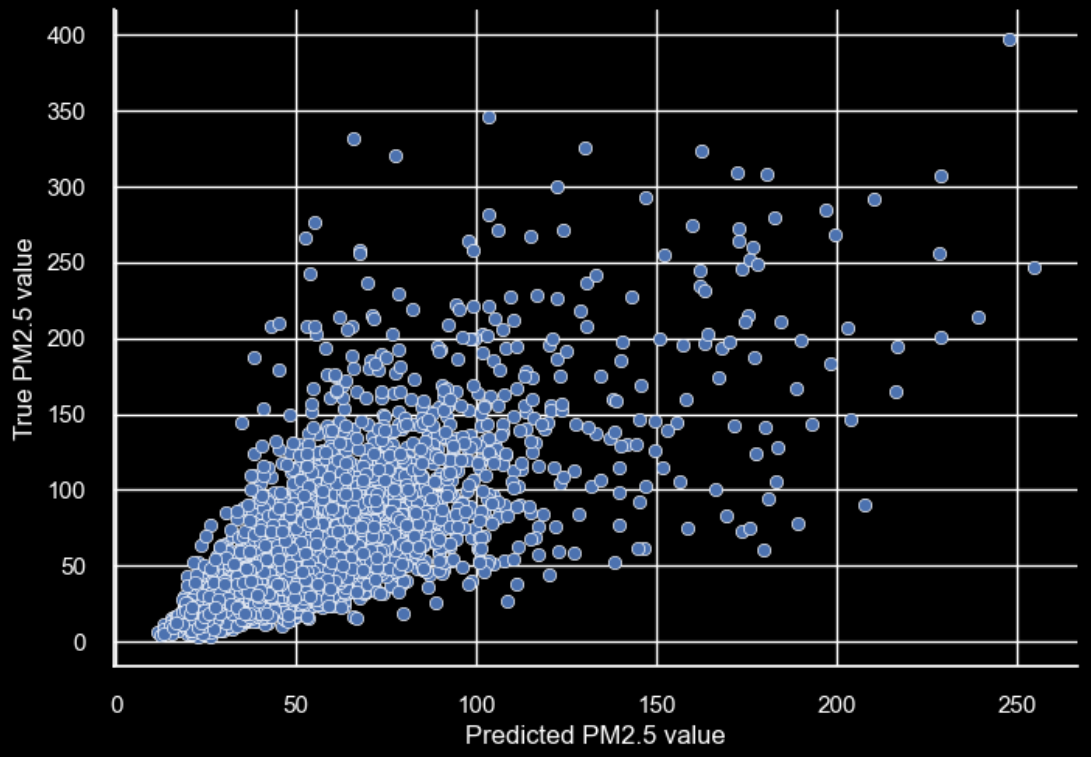
XGBoost

Predictions:

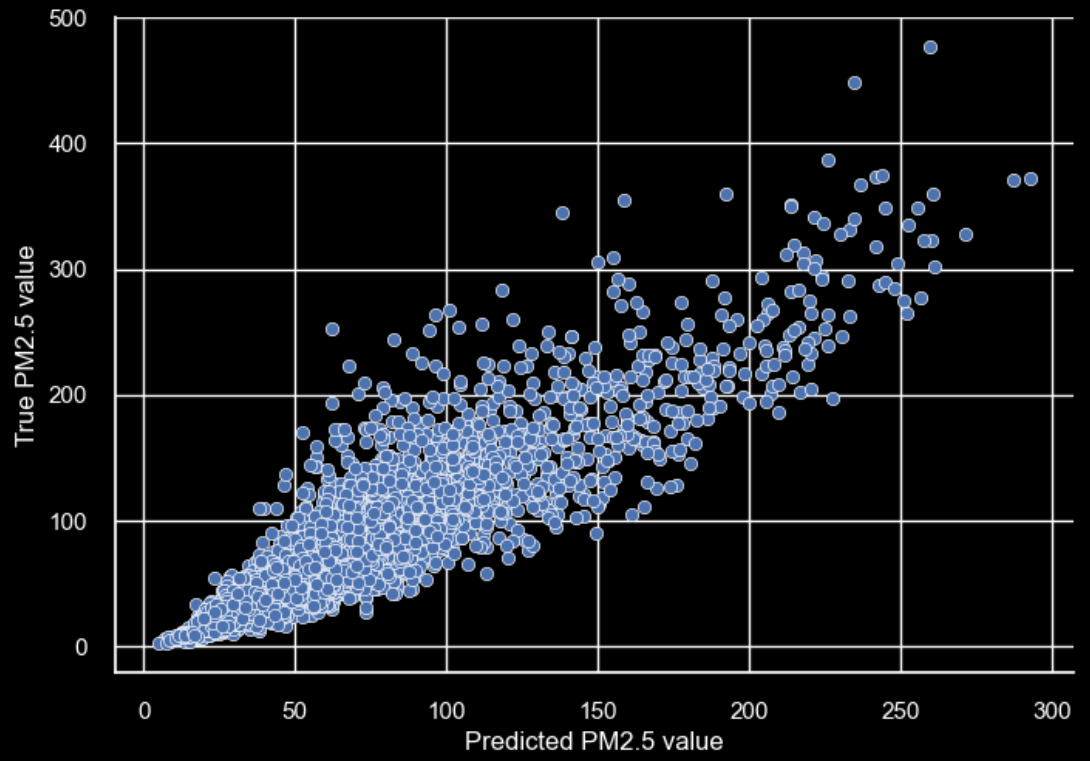
Air pollution 24 hours after last measurement

Model Evaluation

Test data set



Train data set



Conclusion and Future Improvements

Conclusion and Future Improvements

Conclusions

- Good prediction of air pollution levels possible with weather data
- Most important feature: Atmospheric pressure
- Our model can help improve health in Uganda

Future Improvements

- API access
- Include meta data of the sensors
- Generalize model for other parts of Africa or world wide

Thank you for your attention!

AIr – We bring AI to your air

Sebastian, Samer, Jess, Matthias

Backup slides

References

- https://www.atcmask.com/blogs/blog/air-pollution-in-uganda
- https://www.independent.co.ug/air-pollution-is-getting-worse/

Residual analysis