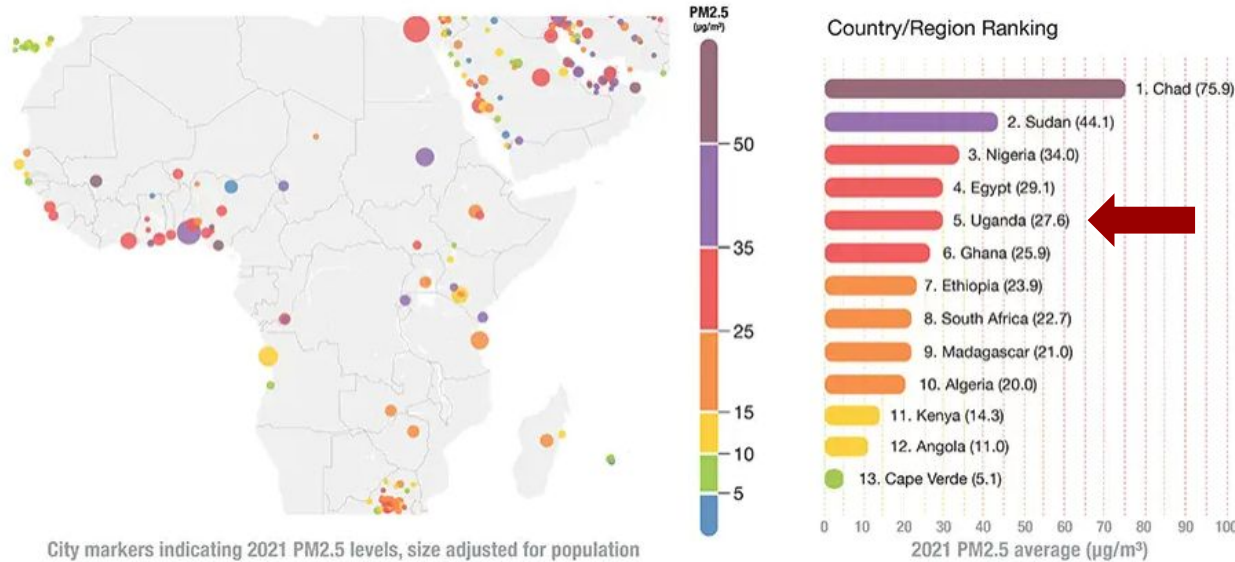

AirQo Ugandan Air Quality forecast

05.01.2023

Markus Baden, Humam Abuobayda, Jana Täumer

AirQo - Air Pollution in Uganda



source: <https://www.iqair.com/fr/world-air-quality-report>

AirQo - Air pollution

- sources of air pollution: traffic, fossil fuel burning
- measure of air quality: fine particles or PM_{2.5}
- health risks (respiratory disease, heart disease and stroke)
- physical activity outside is not recommended when air pollution is high



AirQo - Air pollution

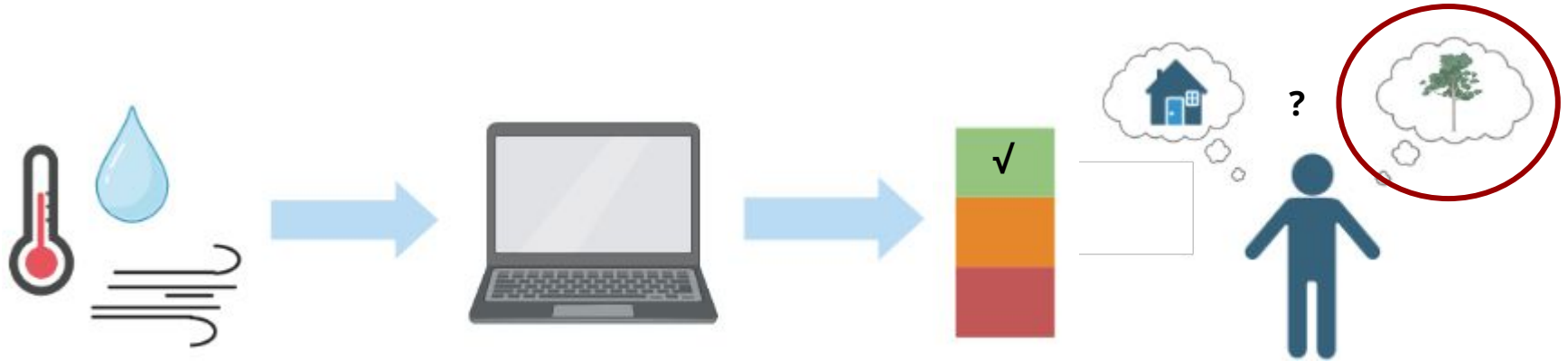
Health risk depends on:

- PM25 concentration
- health status

Health Concern	PM _{2.5} (µgm ⁻³)	Precautions
Good	0 - 12	None
Moderate	13 - 35	Unusually sensitive people should consider reducing prolonged or heavy exertion
Unhealthy for Sensitive Groups	36 - 55	Sensitive groups should reduce prolonged or heavy exertion
Unhealthy	56 - 150	Everyone should reduce prolonged or heavy exertion, take more breaks during outdoor activities
Very Unhealthy	151 - 250	Everyone should avoid prolonged or heavy exertion, move activities indoors or reschedule
Hazardous	250 +	Everyone should avoid all physical activities outdoors.

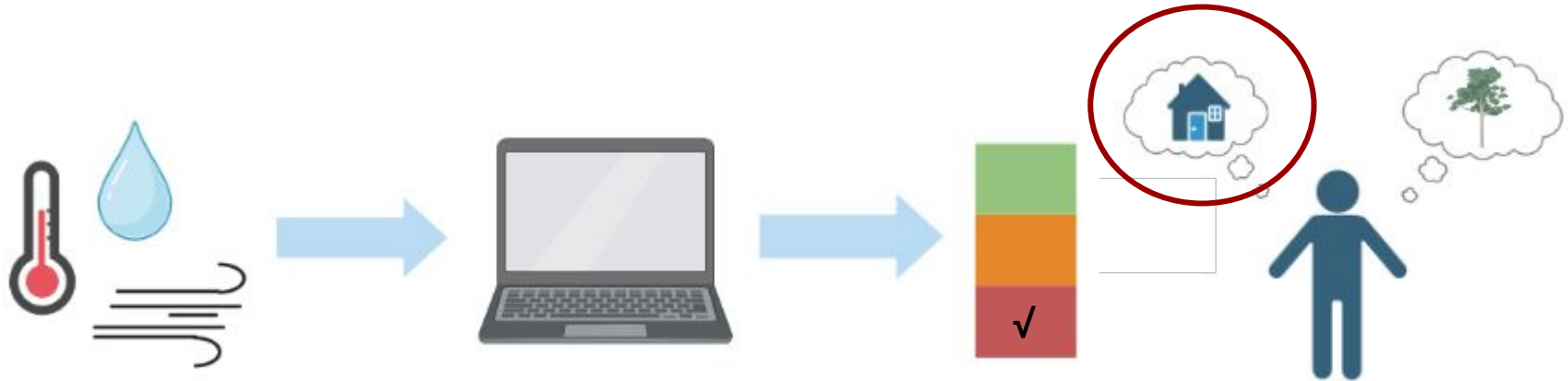
AirQo - Goals

- make precise, inexpensive forecast of air quality using weather data
- derive recommendations for public (safety alerts)
- for 65 locations across Uganda



AirQo - Goals

- make precise, inexpensive forecast of air quality using weather data
- derive recommendations for public (safety alerts)
- for 65 locations across Uganda



AirQo - Dataset

What data we have:

- 5-day measurements of weather data including
 - Wind speed and direction
 - Precipitation
 - Temperature
 - Air pressure
 - Relative humidity
 - hourly measurements
- Target: air pollution (PM_{2.5} value) 24 hours after 5-day period

Problems during data analysis:

- Missing data → delete or impute missing values
- A lot of features → aggregating means of 6-hours time intervals
- No time stamps
- Not much information about the sensors

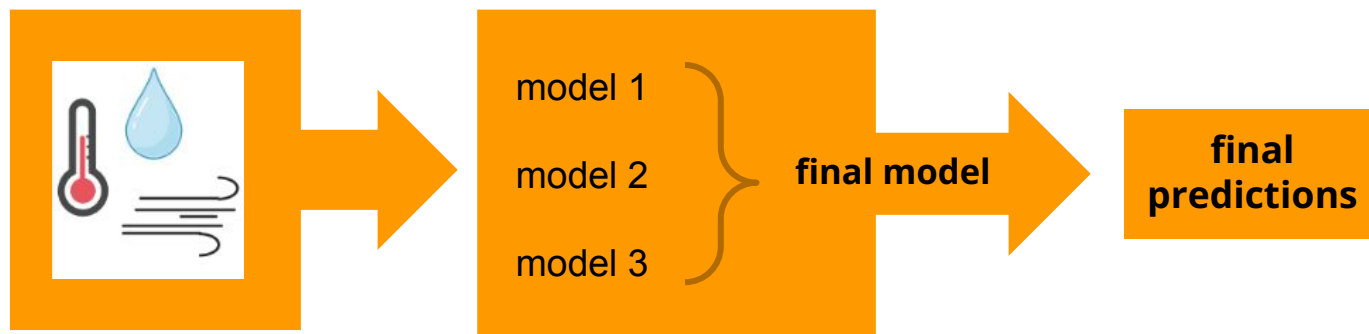
AirQo - Model

- We tested 7 different models to find the best solution
- RMSE (Root-Mean-Squared-Error)
 - measure of how far our predictions are from reality
 - small values are better than high values
- Further feature engineering
- Grid search to improve parameters

Model	Performance (RMSE)
Baseline model	44.1
DecicionTreeRegressor	40.18
AdaBoostRegressor	43.49
SVR	39.35
SGD Regressor	37.67
RandomForest Regressor	26.74 → 25.97 → 25.80
KNeighbors Regressor	31.09 → 29.85 → 26.77
XGBoost Regressor	29.17 → 25.96 → 25.58

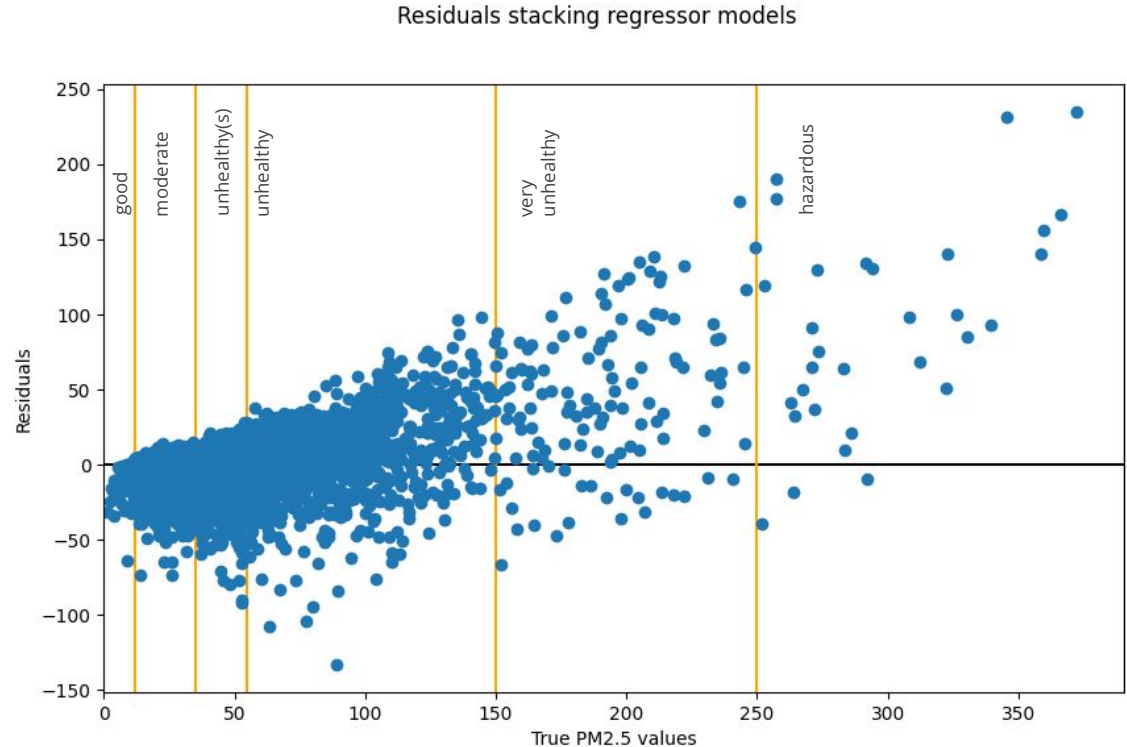
AirQo - Final Model

- Combination of the three best models by using a *stacking model*
 - individual models “vote” on the final output
 - output model is a linear regression model
 - this allows further improvement of the predictions
- → Final RMSE score: **24.48**



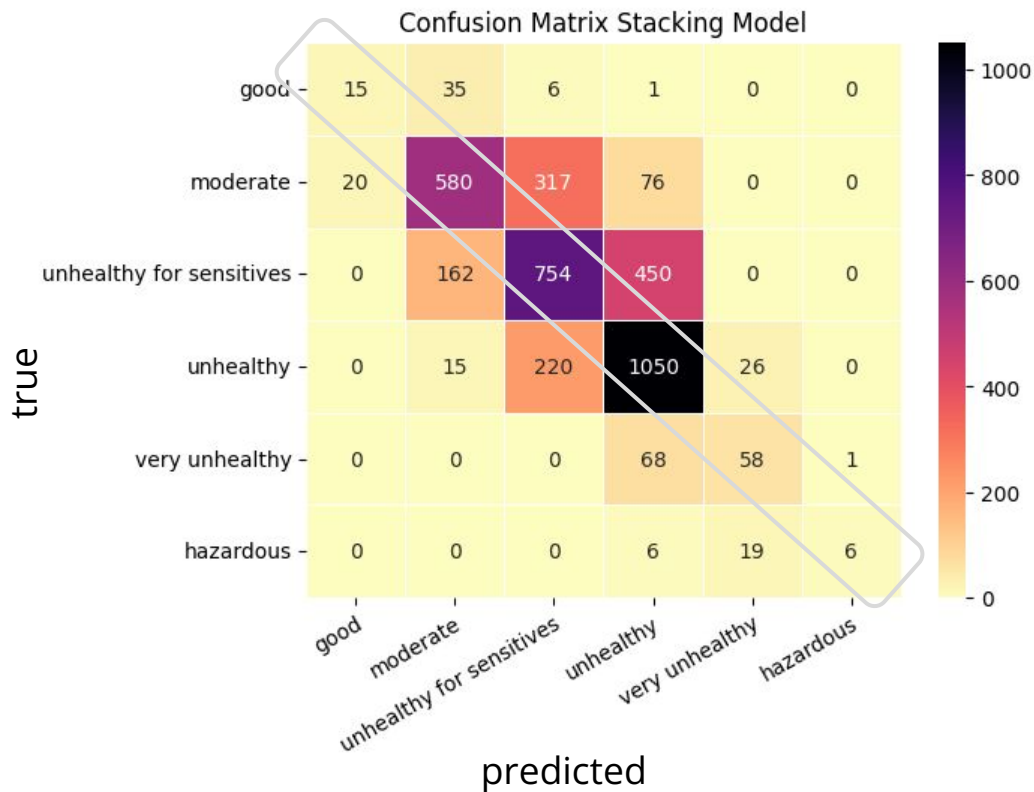
AirQo - Model evaluation

- Residual plot shows differences between predictions and actual values
- model is more accurate for smaller PM_{2.5} values → hazard categories are narrower here
- no outliers in positive area (they would be predicting better conditions than they actually are)



AirQo - Model evaluation

- When the predictions are grouped by hazard categories the results can be shown in a confusion matrix
- Correct predictions are in the diagonal cells
- Majority of the false predictions is only wrong by one category



AirQo - Ideas for future model improvement

- Try classification algorithms instead of regression
- Correct wind direction by sensor orientation
- Check performance on all 65 sensors

AirQo - Conclusion

Is it feasible to predict air quality with weather sensor data?

- Low cost sensors are sufficient to predict air pollution
- Further model improvement could make predictions better
- This concept could be extended to other African countries

Thank you for your attention

Do you have questions?

