Preliminary results

Model Random Forest for isopropanol and water amplitude

The final model for the amplitude is a Random Forest (model 5) based on cleaned data (sensor 8 measurements removed) and with two outputs.

Features - air amplitude, air wavelength

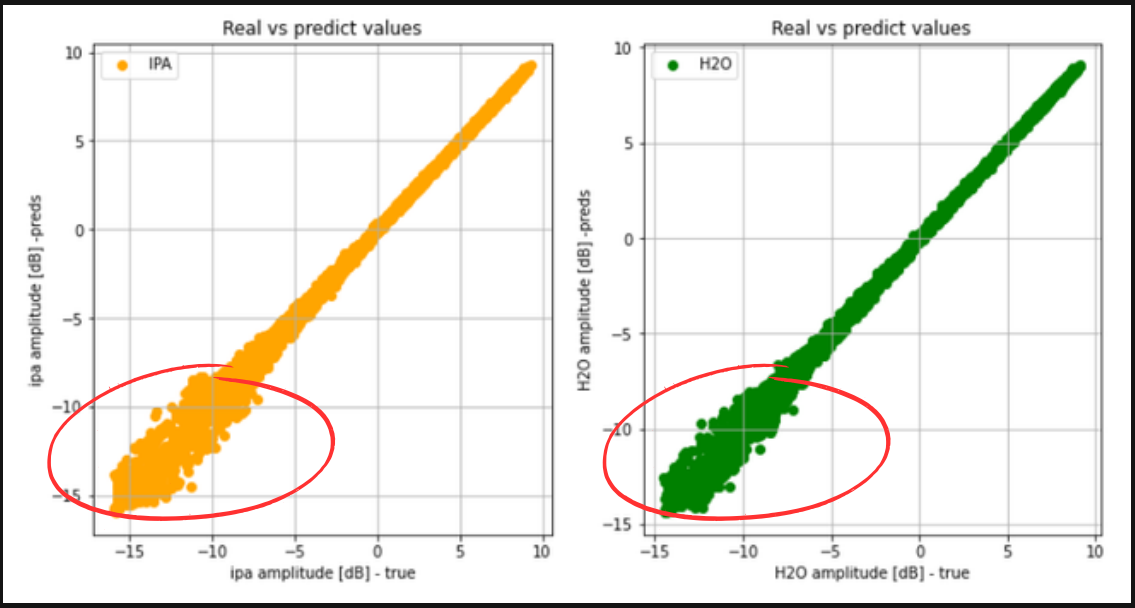
Targets - isopropanol amplitude and water amplitude

Validation set

- MAE: 0.209
- R2 score: 0.997

Test set

- MAE: 0.204
- R2 score: 0.997



From the graph of real and predicted values for Random Forest (model 5), it can be read that the greatest discrepancy of the results is for very small amplitudes (below -7). Amplitude values below -7 are the most important to create a proper wavelength vs. amplitude plot, and it is these values that are crucial for sensor verification.

What next?

The next steps are: selection of appropriate hyperparameters for the Random Forest model, saving the models, creating a simple web application (for example using Streamlit or Flask) that allows the user to enter measurements in the air and get predictions of measurements of isopropylene alcohol and water amplitude.

Michalina Hulak | email: michalina.hulak@op.pl | Github