Short Report on DON Concentration Prediction

1. Preprocessing Steps and Rationale:

- Handled missing values using imputation.
- Normalized hyperspectral data for consistency.
- Applied feature selection to reduce dimensionality.
- Split data into training (80%) and testing (20%) sets.

2. Insights from Dimensionality Reduction:

- PCA revealed key wavelengths contributing most to variance.
- Reduced feature set improved model efficiency.

3. Model Selection, Training, and Evaluation:

- Chose a neural network as the baseline model.
- Explored ensemble methods (stacking/boosting) for improvements.
- Used k-fold cross-validation for robust evaluation.
- Optimized hyperparameters using Grid Search.

4. Key Findings and Improvements:

- Best model achieved low RMSE and high R^2 score.
- Model explained key spectral patterns in DON concentration.
- Future improvements: Try advanced deep learning architectures, more feature engineering.