1. Understanding the Challenge and Objectives

- Challenge Background:

Reviewed the challenge’ s goal to predict urban heat island (UHI) effects using near-surface temperature data and Sentinel-2 satellite imagery.

- Stakeholder Applications:

Considered how local governments, urban planners, and decision-makers might use your model for interventions, scaling to other cities, addressing socioeconomic vulnerabilities, and managing energy demand.

2. Data Processing and Feature Engineering

- Data Collection:

Compiled the raw Sentinel-2 band data (B01 – B12) along with derived features such as NDVI, NDMI, NDBI, NDWI, UI, STI, B11/B12 ratio, impervious surface estimates, and vegetation fraction.

- Resampling and Stacking:

Discussed strategies to resample bands of differing resolutions (10m, 20m, 60m) to a common resolution, allowing for multi-channel stacking and consistent feature extraction.

3. Handling Missing Values

- Imputation:

Implemented techniques using Python (e.g., scikit-learn’ s SimpleImputer) to fill in missing values in your dataset, ensuring that your model has a complete and robust input for training.

4. Model Building and Hyperparameter Tuning

- Traditional Machine Learning Models:

Experimented with various models, including Random Forest, XGBoost, LightGBM, and CatBoost, focusing on parameter adjustments to reduce overfitting and improve generalization.

- Ensemble Approaches:

Explored stacking and voting ensemble methods—developing improved ensembles with adjusted weights and hyperparameters (e.g., lowering max_depth, reducing num_leaves, and using early stopping) to balance model complexity and performance.

5. Deep Learning Model Considerations

- Tabular Data Models:

Since dataset is tabular, I examined deep learning models tailored for such data (e.g., MLP, TabNet, TabTransformer) as alternatives or complements to traditional ML approaches.

6. Literature and Research Context

- Related Research:

Reviewed and referenced several academic papers that apply strong machine learning models to UHI prediction, including studies on downscaling land surface temperature using advanced models, deep learning for urban thermal mapping, and ensemble methods for robust UHI forecasting. This provided insights into feature engineering, model selection, and validation strategies.