# New Energy Building Power Prediction

Team: Gradient Descent Enjoyer

Jeremias Ferrao

December 10, 2024

# Data Exploration

- Explored the dataset containing HVAC energy consumption data
- Identified key features influencing energy usage
- Visualized data distributions and correlations

# Dropping Irrelevant Features

- Removed features with low variance or high missing rates
- Excluded redundant features based on correlation analysis
- Reduced dimensionality to improve model performance

# Processing Cyclic Features

- Time features like hour and month are cyclic
- Applied sine and cosine transformations

$$\text{sin\_hour} = \sin\left(2\pi \times \frac{\text{hour}}{24}\right)$$

$$\text{cos\_hour} = \cos\left(2\pi \times \frac{\text{hour}}{24}\right)$$

- Captured cyclical patterns in energy consumption

# Dealing with Corrupt Data

- Identified missing and anomalous values
- Imputed missing data using mean and interpolation
- Removed outliers based on statistical thresholds

# Initial Approach

- Started with Linear Regression as baseline
- Achieved MAE of initial model (e.g., 500 units)
- Realized need for more complex models due to nonlinearity

# Ensemble Methods

- Utilized Random Forest and XGBoost regressors
- Ensembles help in capturing diverse model strengths
- Improved performance over individual models

# Meta Learning

- Implemented StackingRegressor with ElasticNet as meta-estimator
- Combined predictions from base models for final output
- Fine-tuned hyperparameters using Grid Search

**Code Snippet:**

```
# Optimize meta-estimator
meta_estimator = ElasticNet(**best_params_meta)
stacking_regressor = StackingRegressor(
    estimators=estimators,
    final_estimator=meta_estimator,
    passthrough=True,
    n_jobs=NUM_JOBS
)
stacking_regressor.fit(X, y)
```

# Feature Importance

- Extracted feature importances from ensemble models
- Identified top contributing features to energy consumption