**Week 2: Data Preprocessing and Exploratory Data Analysis (EDA)**

**Tasks Completed:**

1. **Data Cleaning**:
   - Dropped irrelevant columns: `Infill Pattern`, `Elongation`, `Roughness` (low correlation with tensile strength).
   - Encoded categorical variable: `Material` (PLA → 0, ABS → 1).
   - Handled missing values: Removed 5 outlier entries with extreme tensile strength values (>100 MPa).
2. **EDA**:
   - Generated correlation matrix (Figure 1) using Seaborn:
   - Strong positive correlation: Layer height vs. Roughness (0.77).
   - Moderate correlation: Wall thickness vs. Tensile strength (0.34).
   - Negative correlation: Nozzle temperature vs. Tensile strength (-0.39).
   - Visualized feature distributions using histograms and boxplots.
3. **Dataset Preparation**:
   - Split data into training (80%) and testing (20%) sets.
   - Standardized features using `StandardScaler` for Linear Regression.

**Challenges Faced**:

- Skewed distribution in wall thickness (most entries between 1–2 mm).
- Multicollinearity between nozzle temperature and bed temperature (r = 0.55).

**Outcomes**:

- Finalized preprocessed dataset (merged Kaggle + research data: 250 entries).
- Drafted **Dataset Discussion** and **Methodology** sections.

**Tasks Planned for Week 3**:

1. Train baseline regression models (Linear, AdaBoost, XGBoost, Gradient Boosting, Random Forest).
2. Evaluate model performance using $R^2$ and MSE metrics.
3. Conduct 5-fold cross-validation to assess generalizability.