Week 3: Feature Engineering and Selection

Tasks:

1. Feature Creation

    - Develop new features based on domain knowledge (e.g., study time per credit)
    - Create interaction terms between existing features
    - Implement polynomial features for numerical variables
    - Develop time-based features (e.g., time since last exam)

2. Feature Transformation

    - Apply log transformation to skewed numerical features
    - Bin continuous variables into categorical ones where appropriate
    - Standardize numerical features

3. Feature Selection

    - Implement filter methods (e.g., correlation analysis, chi-square test)
    - Apply wrapper methods (e.g., recursive feature elimination)
    - Use embedded methods (e.g., Lasso, Random Forest importance)
    - Perform stability selection to identify robust feature subsets

4. Dimensionality Reduction

    - Apply Principal Component Analysis (PCA)
    - Implement t-SNE for visualization of high-dimensional data

Deliverables:

1. Jupyter notebook documenting all feature engineering and selection processes
2. Feature engineering report (PDF) including:

    - Detailed explanation of each new feature created
    - Justification for feature transformations
    - Analysis of feature importance and selection results
    - Visualization of dimensionality reduction results

3. Updated dataset with engineered features
4. Python script for reproducing feature engineering steps on new data