3460:436/536 AML, Project 3 – Your choice

Project Description and Rubric: For this project, you'll select a **suitable dataset** and showcase your **mastery** of machine learning skills. The project is to be completed in groups of 2 to 3 students. Each group must clearly delineate and state the responsibilities of each student within the group. Each student must submit their own individual statement, reflecting their personal contributions to the project.

Rubric:

1. Problem Understanding
    a. Clearly defined problem statement
    b. Background understanding of the problem domain.
    c. Justification of the relevance and significance of the problem.
2. Data Understanding
    a. Description of dataset features and their relevance to the problem.
    b. Thorough exploration of the dataset.
    c. Handling of missing values, outliers, and data preprocessing steps.
    d. Use PCA to visualize the data.
3. Feature engineering
    a. Justification of chosen features based on relevance and importance.
    b. Explanation of feature selection methods.
    c. Justification of chosen features based on relevance and importance.
    d. Creation of new features or transformations that enhance model performance.
4. Model Selection and Training
    a. Exploration of multiple machine learning algorithms.
    b. Proper splitting of data into training, validation, and testing sets.
    c. Implementation of appropriate evaluation metrics.
5. Hyperparameter Tuning
    a. Systematic exploration of hyperparameters.
    b. Utilization of techniques such as grid search and random search.
    c. Explanation of the impact of hyperparameters on model performance.
6. Results and Analysis
    a. Presentation of model performance metrics (e.g., confusion matrix, …).
    b. Interpretation of model predictions and insights.
    c. Visualization of key findings and analysis of results.
7. Project presentation
    a. Clear and well-organized project presentation.
    b. Communicates information clearly and effectively.
    c. Proper code demo.

What to submit:

- Undergraduate students:
    1. Statement of responsibilities
    2. Your source code
    3. Project presentation (ppt slides)
- Graduate students:
    1. Project report
    2. Statement of responsibilities
    3. Your source code
    4. Project presentation (ppt slides)

Where to find a good dataset?

- https://www.kaggle.com/
- https://www.kaggle.com/competitions
- https://archive.ics.uci.edu/
- https://guides.library.cmu.edu/machine-learning
- https://www.cdc.gov/nchs/nhanes/index.htm