Machine Learning Course Final Project

General Description

This final project is designed to assess your ability to apply machine learning techniques in a comprehensive manner. Throughout this project, you will have the opportunity to demonstrate your understanding of machine learning concepts, model development, analysis, and enhancement strategies. The project is structured to guide you through the process of developing a baseline model, analyzing existing models for insights, enhancing model performance through advanced techniques, and exploring state-of-the-art methods in machine learning. This hands-on project will allow you to showcase your skills in data preprocessing, feature engineering, model selection, and performance evaluation, culminating in a detailed report of your findings and methodologies.

Tasks Overview

1. **Baseline Model Development (20%)**
   a. **Baseline Model**: Create and describe a baseline model as per the guidelines in Task 2.
   b. **Baseline Results**: Present and discuss the model's performance metrics.
2. **Notebook Analysis (20%)**
   a. **Notebook Selection**: Choose a notebook from the provided list for detailed analysis.
   b. **Analysis Summary**: Summarize the chosen notebook's methodology, results, and conclusions.
   c. **Critique and Improvements**: Provide a critical review of the notebook's approach and suggest improvements or alternatives.
3. **Model Enhancement (40%)**
   a. **Feature Engineering**: Describe advanced feature engineering techniques used and their rationale.
   b. **Enhanced Model Development**: Explain the development process of the enhanced model, including any preprocessing or tuning efforts.
4. **Advanced Modeling (10%)**
   a. **Model Implementation**: Detail the implementation of advanced models like Random Forest and XGBoost.
   b. **Evaluation and Comparison**: Compare the models' performance and analyze differences.

5. **State-of-the-Art Techniques (10%)**
    a. **Technique Exploration**: Investigate and document advanced techniques from top Kaggle competitors.
    b. **Application and Performance**: Explain how these techniques were applied and their impact on model performance and ranking.

Please ensure your project submission is thorough, clearly articulated, and demonstrates a deep understanding of the machine learning concepts and techniques applied. This project is your opportunity to showcase your skills and creativity in solving real-world data challenges.

Submission Guidelines

- **Code**: submit the ipynb version of your project, make sure that the notebook can run by using the "Run All" command without having errors
- **Paper**: Submit your project as a PDF document, that includes a detailed description of your work notebook, please make sure that it is organized by the 5 different parts of the project
- **Presentation**: Create a short 5-minute presentation that describes the most interesting points from your paper, try to include the most innovative and important findings from your work, at the end of the presentation you will need to answer a question about your work.

Evaluation Criteria:

- **Clarity and Organization**: Your report should be clearly written and well-structured, making it easy for readers to understand your process and findings.
- **Comprehensiveness**: Cover all aspects of the tasks, providing detailed explanations of your approach, analysis, and conclusions.
- **Innovation**: Demonstrate creativity and originality in your feature engineering, model enhancement, and exploration of state-of-the-art techniques.
- **Accuracy and Performance**: Show evidence of model performance improvements through your methodologies and compare your results with baseline metrics.
- **Relative Performance Evaluation:** Final grades will be assigned based on how each submission measures up against others in the course, rewarding work that demonstrates superior innovation, analytical depth, and exceptional predictive insight.