CSIS 260: Mid-term and Final project

Project Overview:
In this project, you will be working with a real-world dataset to build a predictive model using machine learning techniques. The project will be divided into two phases:

Mid-Term Presentation:

- Data Preparation: Clean the dataset, handle missing values, and address any data quality issues.
- Exploratory Data Analysis (EDA): Perform exploratory data analysis to gain insights into the data. Include visualizations and statistical summaries to understand the dataset better.
- Data Preprocessing: Prepare the data for model training. This includes feature selection, encoding categorical variables, and scaling/normalization if needed.
- Data Regularization: Apply regularization techniques (e.g., L1, L2) to improve model performance.
- Model Selection: Choose at least three different machine learning algorithms for initial model training.
- Model Evaluation: Evaluate the models using appropriate metrics (e.g., accuracy, precision, recall, F1-score).
- Presentation: Prepare a mid-term presentation summarizing your progress, challenges faced, and initial results. Discuss any correlations you found and the best features identified.

Mid-Term Grading (30% of Course Grade):

- Data Acquisition and Preparation: [20 points]
- Exploratory Data Analysis: [20 points]
- Data Preprocessing and Data Regularization: [20 points]
- Model Selection and Training: [20 points]
- Model Evaluation: [20 points]

Final Presentation:.

- Hyperparameter Tuning: Optimize hyperparameters for your chosen model(s) using techniques like grid search or random search.
- Model Evaluation: Perform a comprehensive evaluation of your final model(s) on the test dataset. Compare different models and justify your choice of the best model.
- Visualization: Create informative visualizations to present your final results effectively.
- Conclusion: Summarize your findings, discuss the practical implications, and suggest possible future improvements.
- Final Presentation: Present your final results, methodology, and conclusions to the class.
- Report: Submit a detailed report that includes all aspects of the project, following a structured format.

Final Grading (40% of Course Grade):

- Hyperparameter Tuning: [20 points]
- Model Evaluation: [20 points]
- Visualization: [20 points]
- Conclusion: [20 points]
- Presentation and Report [20 points]

Dataset:
TBA: [October 18]
Link will be posted on Moodle.

Important Dates:

- Mid-Term Presentation: [November 13-15]
- Final Presentation and Report Submission: [December 11-13]

Competition:
The top three teams with the highest accuracy % and precision % (measured by mAP) will receive bonus points towards their overall course grade: 1st place will receive 5 points, 2nd place will receive 3 points, and 3rd place will receive 1 point.

Note: Please make sure to follow ethical guidelines and properly cite your data sources in your report. Plagiarism will not be tolerated.