PROJECT OVERVIEW STATEMENT	Project Name: Using SHAP Post-Model Explainability as a Model-Agnostic Feature Selection Technique	Student Names: Joshua Gottlieb, Martin Sichali Chunsen, Alexander Yoo

Problem/Opportunity:

While there are many feature selection techniques, many of these techniques either fail to optimize for the predictive power of the model, require iterating through the extensive feature space, or are applicable only to certain model types. We propose using SHapley Additive eXplanation (SHAP) values to capture and rank the important features learned by a model for use in feature selection due to their model-agnostic nature and their inherent focus on capturing the predictive power of each feature in the context of the particular model being evaluated. We propose a unique set of strategies for choosing the ranked features based on the combined and individual strengths of each feature as discovered by SHAP.

Goal:

Our goal is to test a variety of feature selection strategies that use global SHAP explanations to produce feature-reduced datasets for further retraining. We aim to investigate the feasibility and effectiveness of SHAP as a feature selection technique. We will use ten different binary classification datasets (five small and five medium-size datasets, publicly available from the UCI Machine Learning Repository and from Kaggle) and five different machine learning models. We will compare the effectiveness of SHAP to other feature selection techniques such as mutual information gain and correlation based feature selection using Precision-Recall Area Under the Curve (PR AUC) as our metric. An analysis of the effectiveness and computational efficiency of the techniques will be presented to investigate the feasibility of using SHAP as a feature selection technique.

The goal described above will be completed during the semester, with measurable milestones being progress meetings, presentations, and a completion of individual tasks toward the goal. All project tasks are realistically achievable by the personnel assigned to the project using local and Google Colab compute resources.

Objectives:

Outcomes:

- Perform data preparation on the datasets and hyperparameter tuning for each of the five model types for each of the ten datasets.
- Use the fitted models to compute global SHAP values for each of the datasets and each of the model types.
- Select features based on SHAP values and retrain each model using the selected features for each dataset.
- Analyze the effectiveness and computational feasibility of using SHAP values for feature selection by comparing performance metrics of the models trained on the reduced feature sets to the performance of the original models.
- Compare using SHAP as a feature selection technique to existing feature selection techniques.

Time Frame:

- October 20th: Complete training and SHAP computation for the five small datasets.
- October 27th (midterm): Complete feature selection, retraining, and performance analysis for the five small datasets, as well as the comparison of model performances using SHAP versus other feature selection techniques.
- November 3: Produce first draft of technical paper (Introduction, Literature Review, Methodology sections).
- November 10th: Complete training and SHAP computation for the five medium-size datasets.
- November 17th: Complete feature selection, retraining, and performance analysis for the five medium-size datasets, as well as the comparison of model performances using SHAP versus other feature selection techniques.
- November 24: Produce second draft of technical paper (Analysis and Results sections) and produce final poster for submission to internal Pace student project competition.
- December 1: Finalize technical paper and presentation.
- December 15: Final presentation due.

Metrics:

- PR AUC will be used as the main metric of comparison for model performance in order to compare between balanced and imbalanced classifiers. An increase or lack of change in PR AUC for models on the reduced feature set compared to the full feature set is an indicator of the effectiveness of the SHAP selection technique.
- The computation time in seconds for the SHAP explanations will be used as a point of comparison to gauge the efficiency of the process.
- The percentage of features kept will also be used to evaluate how effective the technique is at pruning irrelevant features while maintaining or improving performance.

Action:

- Perform the tasks delineated above in the "Time Frame" section by the specified due dates.

Success Criteria:

The project will be considered a success if all of the tasks delineated in the "Time Frame" section are completed as scheduled. Additional success criteria include:

- Providing a repository including the code and data so that the project results can be replicated by others.
- Clearly presenting the performance comparisons between the models trained on the SHAP-selected reduced feature sets and the models trained on the original unpruned feature sets.
- Clearly presenting the performance comparisons between the models trained on the SHAP-selected reduced feature sets and the models trained on feature sets that were reduced using other feature selection techniques.

Assumptions, Risks, Obstacles:

The following assumptions are prerequisites to the success of the project. If any of these assumptions prove incorrect, that represents a risk to the success of the project.

Assumptions that could prove to be potential risks:

- All the tasks delineated in the "Time Frame" section can be performed according to schedule. If not, this represents schedule risk.
- All the tasks delineated in the "Time Frame" section can be performed either locally or using Google Colab. If not, this represents a risk to the time line. Potential mitigations might include reducing the number of datasets or working with smaller datasets.
- Computation of SHAP values can be time-consuming and may affect adherence to the specified time line. Should this prove to be an issue, it may be necessary to only compute SHAP values for a subset of the records in the larger datasets or to drop the larger datasets entirely.

Prepared By	**Date**	**Approved By**	**Date**
Joshua Gottlieb	10/19/2025		