ML - Assignment 10

The table presents the results of different parameter-tuned machine learning models, including Decision Tree, Stochastic Gradient Descent, XGBoost (XGB), Support Vector Machine (SVM), Logistic Regression, and Random Forest Classifier. These models were tuned using specific hyperparameters, and their performance is evaluated based on the F1-score.

Model	Best params	F1_score
Parameter Tuned DecisionTree Regression	{'criterion': 'gini', 'max_depth': 2, 'max_leaf_nodes': 10, 'min_impurity_decrease': 0.001, 'min_samples_leaf': 2, 'splitter': 'random'}	0.8710980392
Parameter Tuned Stochastic Gradient Descent	{'alpha': 0.001, 'loss': 'hinge', 'penalty': 'l1'}	0.8707299622
Parameter Tuned XGB	{'alpha': 6, 'eta': 0.1, 'lambda': 0.5, 'max_depth': 8, 'n_estimators': 100, 'tree_method': 'auto'}	0.8704161763
Parameter Tuned Support Vector Machine	{'C': 1, 'kernel': 'linear'}	0.8699456484
Parameter Tuned Logistic Regression	{'C': 0.5, 'max_iter': 500}	0.8695946835
Parameter Tuned RandomForest Classifier	{'bootstrap': False, 'max_depth': 2, 'max_features': None, 'max_leaf_nodes': 4, 'min_samples_split': 2, 'n_estimators': 200}	0.8669186214

Let's discuss the table in detail:

- In our dataset, we have a noticeable class imbalance, with Class 1 representing a substantial 69% of the data, while Class 0 accounts for only 32%. This imbalance adds complexity to our modeling task, as we want our models to effectively capture both classes.
- We've conducted a comprehensive evaluation of various models, each fine-tuned with specific hyperparameters to optimize their performance. Our primary focus is on the F1-score, a metric that strikes a balance between precision and recall, which is particularly crucial in situations like ours, where imbalanced classes are prevalent.
- Among the models we've explored, the DecisionTree Regression stands out as a top performer. With carefully chosen hyperparameters, it achieved an impressive F1-score of

approximately 0.871. This indicates that it effectively navigates the class imbalance and manages to capture both Class 0 and Class 1 instances.

- Another strong contender is the Stochastic Gradient Descent (SGD) model, also boasting an F1-score of around 0.871. SGD is known for its efficiency and suitability for imbalanced datasets, making it a valuable choice.
- The XGBoost (XGB) model performs quite well, with an F1-score of approximately 0.870. XGBoost's reputation for handling imbalanced data is evident in its performance.
- Meanwhile, the Support Vector Machine (SVM) model, configured with a linear kernel, maintains a competitive F1-score of around 0.870, showcasing its capability to address class imbalance.
- Our Logistic Regression model also deserves recognition, achieving an F1-score of approximately 0.870, considering the challenging data distribution.
- However, the RandomForest Classifier lags slightly behind the others, with an F1-score of around 0.867. It appears to struggle more with the class imbalance in our dataset.

In summary, our evaluation suggests that the DecisionTree Regression and Stochastic Gradient Descent models are strong contenders for our specific dataset. Still, we must also consider other factors, such as precision, recall, and the specific needs of our application when making our final model selection. Additionally, techniques like resampling or adjusting class weights may further enhance our models' performance in handling class imbalance.