Assignment - 5

The comparative analysis of XGBoost and AdaBoost classifiers on the mushroom dataset reveals significant insights into their respective performance characteristics. XGBoost demonstrates superior classification capabilities, achieving a training accuracy of 96.32% (0.9632) and test accuracy of 90.08% (0.9008), while AdaBoost shows notably lower performance with 77.98% (0.7798) training accuracy and 75.73% (0.7573) test accuracy. Despite this substantial accuracy gap of approximately 15 percentage points in favor of XGBoost, both algorithms exhibit remarkably similar training times (XGBoost: 1.07 seconds vs AdaBoost: 1.06 seconds), indicating that XGBoost's enhanced predictive power comes at virtually no additional computational cost. The accuracy differential between training and test sets (6.24% for XGBoost versus 2.25% for AdaBoost) suggests that while XGBoost may exhibit slightly more overfitting, it still maintains significantly better generalization performance overall. These results strongly position XGBoost as the preferred algorithm for this classification task, particularly given its ability to deliver substantially higher accuracy (14-15 percentage points better) with nearly identical training duration.

```
Model Comparison Summary:
+----------+--------------------+--------------------+--------------------+
|  Model   | Training Accuracy  |   Test Accuracy    | Training Time (s)  |
+----------+--------------------+--------------------+--------------------+
| XGBoost  | 0.9632336655592469 | 0.9007971656333038 | 1.0722897052764893 |
| AdaBoost | 0.7798449612403101 | 0.7573073516386183 | 1.0555365085601807 |
+----------+--------------------+--------------------+--------------------+
```

The minimal time difference of just 0.02 seconds makes XGBoost the unequivocal choice for applications where accuracy is prioritized, as it provides dramatically improved classification performance without compromising computational efficiency. This performance advantage is particularly noteworthy given that both models were evaluated under identical conditions with the same feature space reduced through PCA. The results underscore XGBoost's robustness in handling the mushroom

classification task, where it not only achieves higher absolute accuracy but also maintains a favorable accuracy-to-training-time ratio compared to AdaBoost. These findings have important implications for production deployment scenarios, suggesting that XGBoost should be the algorithm of choice when working with similar biological classification problems where both accuracy and reasonable training times are important considerations.