

This notebook details the results found when training machine learning models on the Mushrooms dataset (<https://www.kaggle.com/datasets/uciml/mushroom-classification/data>).

The following steps were taken:

1. Inspected the data
 - a. Removed veil-type column
 - b. Target is the first column, class
2. Encode the data
 - a. Used one-hot encoding on all columns except for target
 - b. Used label encoding for the target
3. Split the data into training/testing
 - a. The test size was = 0.05.
 - b. The following models were used:
 - i. Decision Tree
 - ii. MLP Classifier
 - iii. SVM Classifier
 - iv. Random Forest
 - v. Gradient Boosting
 - vi. K-Nearest Neighbours
4. Perform Hyperparameter Optimization
5. Perform 5-fold Cross Validation
6. Retrain Models
7. Display results

These are the final results of the notebook:

	model	accuracy	fScore	precision	recall
0	Decision Tree	1.0	1.0	1.0	1.0
1	MLP Classifier	1.0	1.0	1.0	1.0
2	SVM Classifier	1.0	1.0	1.0	1.0
3	Random Forest	1.0	1.0	1.0	1.0
4	Gradient Boosting	1.0	1.0	1.0	1.0
5	K-Nearest Neighbours	1.0	1.0	1.0	1.0

These scores are expected to be perfect due to the nature of the dataset (all categorical, easy to identify the class). Thus, the scores obtained in this notebook are an improvement over other scores or matched the other scores obtained by other sources.

The score obtained by the first source, Krishnan (2023), used two decision trees and obtained scores between 96-98%, which is lower than the decision tree of this notebook's score of 100%.

The score obtained by the second source, He & Wang (2019), used a SVM and obtained a score of 100%, which matched this notebook's SVM classifier score.

The score obtained by the third source, Demiryurek (2019), used an ANN and obtained a score of 98.46%, which is lower than this notebook's MLP Classifier score of 100%.

Github link:

<https://github.com/ZSV-PROJECT/mushrooms>

Works Cited

1. Kaggle

Krishnan, S. (2023). Mushroom classification - decision tree classifier. Kaggle. Retrieved from <https://www.kaggle.com/code/sandhyakrishnan02/mushroom-classification-decision-tree-classifier>

2. Semantic Scholar

He, X., & Wang, Z. (2019). High performance support vector machines and its applications. Semantic Scholar. Retrieved from <https://www.semanticscholar.org/paper/High-Performance-Support-Vector-Machines-and-Its-He-Wang/72200eb638423bb0810c40eb804633299bb184b9>

3. Kaggle

Demiryurek, G. (2019). Mushroom classification with ANN. Kaggle. Retrieved from <https://www.kaggle.com/code/gulsahdemiryurek/mushroom-classification-with-ann/>