



Name: Sean Xander B. Aquino BSCS2A

This program uses the Abalone Dataset from the UCI Machine Learning Repository. It includes 4,177 instances that represent physical observations of abalone samples. Physical measurements of the abalones together with observed shell rings compose each data point, which functions as an indicator for age estimation. This study transformed the regression-based age prediction task into a classification problem after segmenting the “Rings” attribute into the “Age Group”: Young, Adult, and Old.

- **Young:** 0–5 rings
- **Adult:** 6–10 rings
- **Old:** 11–30 rings

After the data transformation process, the Support Vector Machine (SVM) can be applied for determining abalone maturity through classification.

The dataset features eight components among its inputs.

Categorical Data: Sex - data consisting of **M, F, I** was converted into numerical representation consisting of **0, 1, 2**.

Continuous Data:

- Length
- Diameter
- Height
- Whole weight
- Shucked weight
- Viscera weight
- Shell weight



A standardized approach with StandardScaler was applied to numerical features before training to achieve uniform scale and enhance SVM performance as well as convergence speed.

The Support Vector Machine classifier offered an ideal choice because it functions well in high-dimensional spaces and delivers non-linear relationships through kernel functions. A 5-fold cross-validation of GridSearchCV performed the hyperparameter tuning process. The following parameters were considered:

- **C (Regularization parameter):** [0.001, 0.01, 0.1, 10]
- **gamma (Kernel coefficient):** [0.1, 1, 10]
- **kernel (parameter):** was established as the 'rbf' (Radial Basis Function) since it exhibits strong functionality for non-linear models.

This reveals an opposing outcome, since general advice (from StackAbuse) suggests that lower C values enhance generalization. The model achieved increased success in class separation through elevated C values because it reduced the permitted margin tolerance between classes.

The implementation model reached 76% accuracy in testing conditions. The following performance indicators come from analysis of the classification report and confusion matrix.

- The Adult class segments achieved the highest Precision and Recall precision because it possessed balanced and abundant information.
- Data misclassifications existed in the Young category, mostly because this group was poorly represented in the sample, which affected recall performance.
- A large number of classification errors were found between neighboring age categories, since adult and young abalones appeared as similar groups. This characteristic suggests age forms a continuous spectrum.