Group No -04

Group Members :

Ashakuzzaman Odree | 20301268

Sabiha Alam Chowdhury | 20301192

Fairuz Tassnim Prapty | 21101027

Syed Ashik Mahamud | 20301124

Topic

Identification of Iris Flower Varieties Through Machine Learning Techniques

RA : EHSANUR RAHMAN RHYTHM

ST : MEHNAZ ARA FAZAL

Outline

- Data Acquisition and Preparation
- Choosing the Model
- Implementation of Cross-Validation
- Utilizing the Confusion Matrix

Data Acquisition and Preparation

- Obtained from the UCI Machine Learning Repository.
- Comprises 150 samples across three Iris species: setosa, versicolor, and virginica.
- Dataset stored in the variable “Iris” for ease of reference.
- Imported using the scikit-learn toolkit.
- Split into training (60%) and testing (40%) subsets.
- Ensures stable and consistent accuracy assessments across multiple model runs.

Choosing the Model

Logistic Regression

- Logistic Regression, a statistical technique, analyzes datasets with one or more independent variables influencing the outcome.
- Primarily used for accurate data categorization based on existing information.
- Utilized for segmenting Iris flower data based on length and width attributes.
- Functions effectively, particularly with larger datasets.
- Logistic regression applied to the Iris dataset resulted in a model accuracy of 91%.

Support Vector Machine (SVM)

- SVM serves as a powerful technique for classifying datasets, whether linear or non-linear.
- It employs non-linear mapping to project training data into a higher-dimensional space.
- Each data item plotted in an n-dimensional space (n = number of features).
- Conducted classification within this transformed space.
- SVM searches for an optimal separating hyperplane in the transformed space.

K-Nearest Neighbor Classifier (KNN)

- KNN is versatile, handling both classification and regression problems in supervised learning.
- Straightforward and entirely reliant on the training dataset.
- Classifies incoming data based on similarity measured by the distance between instances.
- KNN classifier implemented using the KNeighborsClassifier(n_neighbors=3) function from the sklearn.neighbors package.
- Chosen and evaluated for its effectiveness in predicting Iris flower species.

Implementation of Cross-Validation

- Cross-validation involves reserving a portion of the dataset for validation, not used during model training.
- Utilize the reserved sample from the test set to assess the model's performance.
- A model exhibiting favorable results is deemed effective.
- Precise estimate from sample accuracy.
- Enhances model efficiency and effectiveness.
- Comparison of SVM, KNN, and Logistic Regression accuracy with and without cross-validation.
- Analysis reveals improved accuracy when cross-validation is implemented, enhancing overall model performance.

Utilizing the Confusion Matrix

- A tabular representation used to evaluate the efficacy of a classification model.
- Utilizes test data with established expected output labels.
- Provides a straightforward indication of prediction accuracy.
- Aids in pinpointing errors made by the model.
- Denotes the target, indicating the classification label for the given sample data.
- The confusion matrix contributes to the determination of the model's accuracy score.
- The confusion matrix acts as a crucial tool in assessing the accuracy and reliability of the classification model, specifically in the context of the Iris dataset.

Thank You

For Your Attention