Machine Learning Module

Presented by
Adam
Basil
Khaled

Objectives:

- To perform unsupervised learning tasks such as clustering and outlier detection on the Iris dataset.
- To build baseline models using supervised learning techniques and compare multiple algorithms.
- To tune the best-performing model and evaluate it against an ensemble of algorithms.

Project Overview

- **Unsupervised Learning: Clustering and outlier detection on the Iris dataset.**
- **Supervised Learning: Building baseline models, comparing multiple algorithms, tuning the best-performing model, and evaluating it against an ensemble of algorithms.**

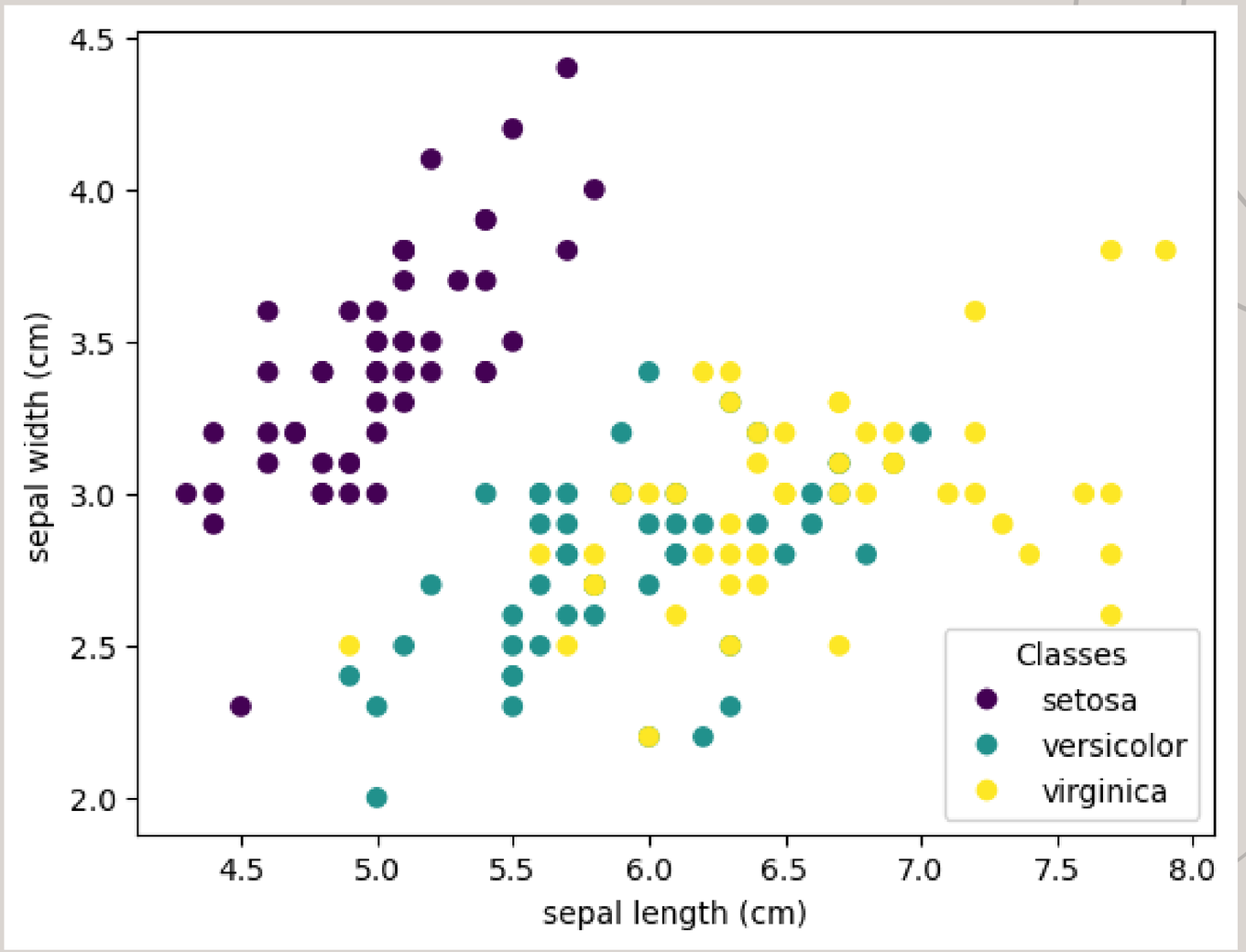
Dataset Overview

- The Iris dataset contains 150 samples of iris flowers.
- It comprises four features: sepal length, sepal width, petal length, and petal width.
- The target variable specifies iris species (setosa, versicolor, or virginica).

Data Preprocessing

- Loaded the Iris dataset.
- Performed data exploration and visualization.
- Checked for missing values and handled them.
- Split the dataset into features and target variables.

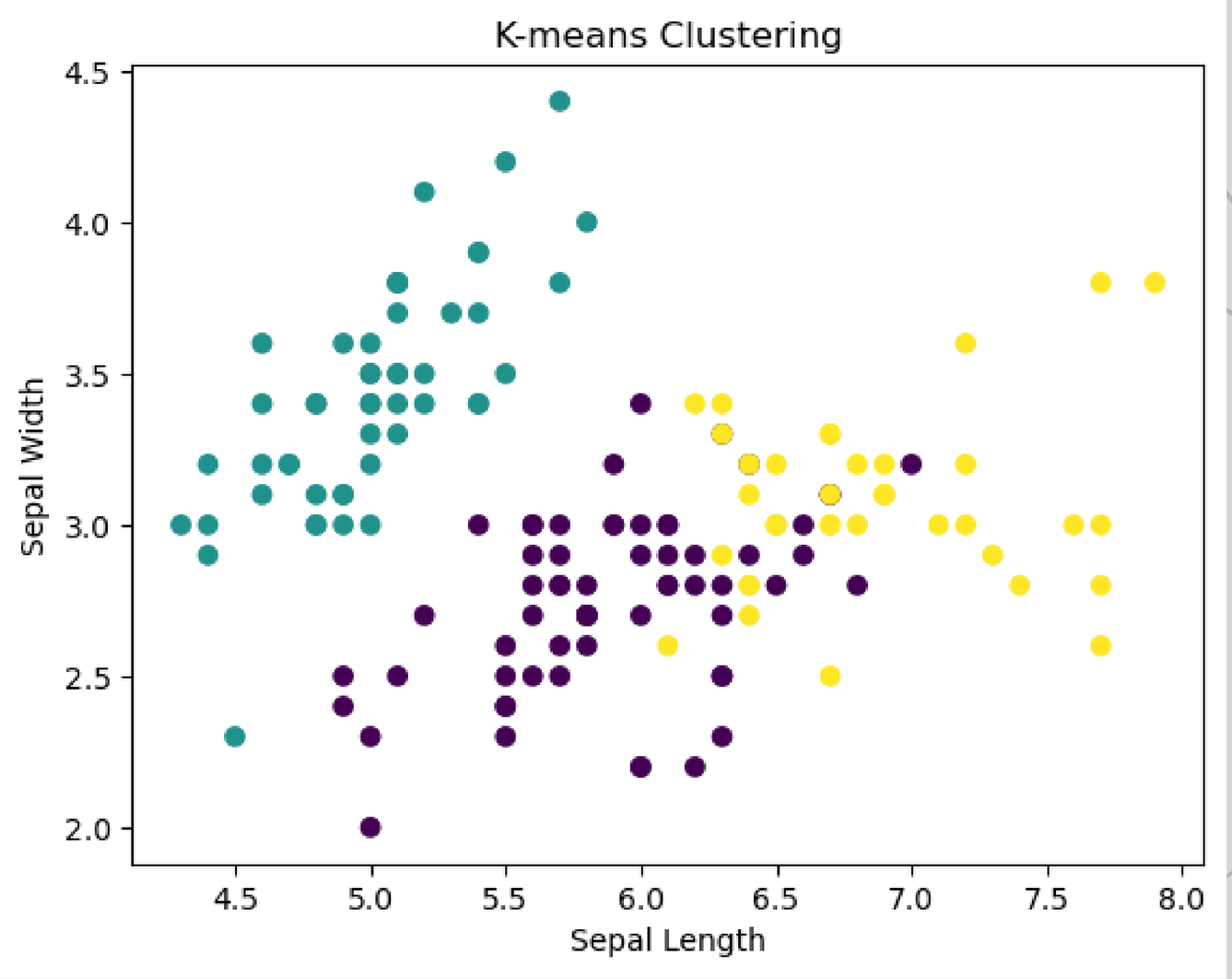
Data Preprocessing



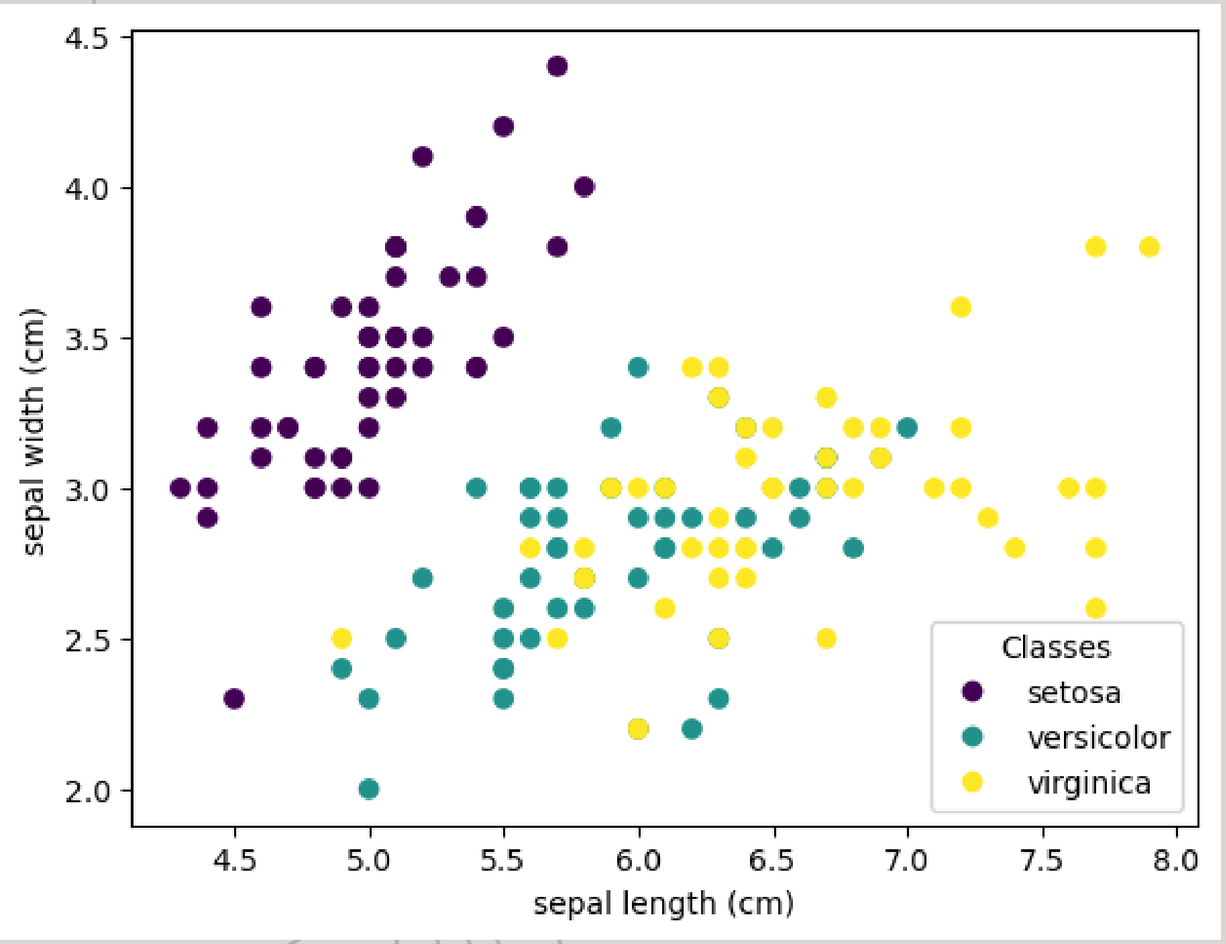
Unsupervised Learning:

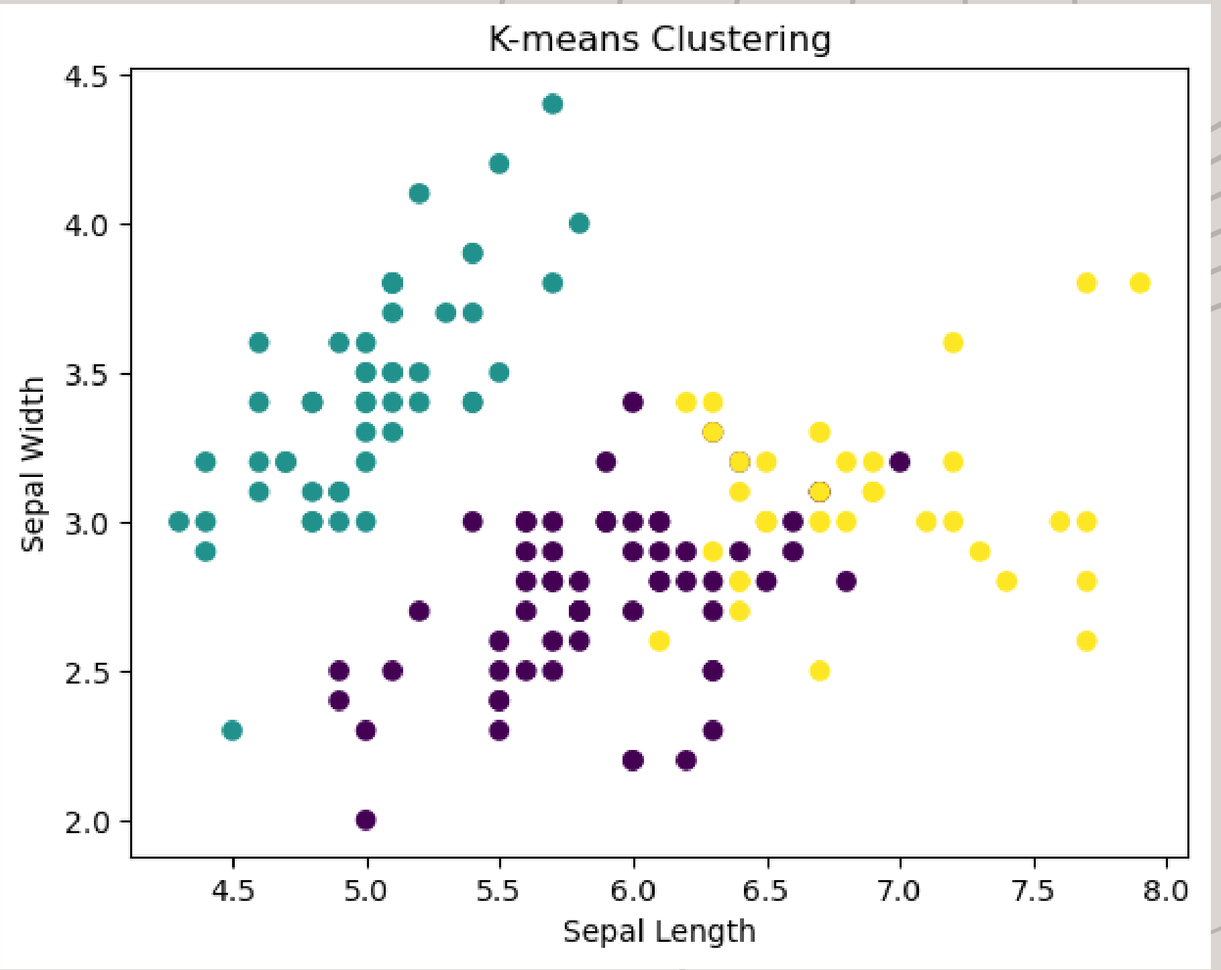
- Applied K-means clustering algorithm to cluster the data.
- Visualized the clusters.
- Detected outliers using appropriate techniques such as isolation forest or DBSCAN.
- Evaluated the clustering results.

Unsupervised Learning: Clustering and Outlier Detection

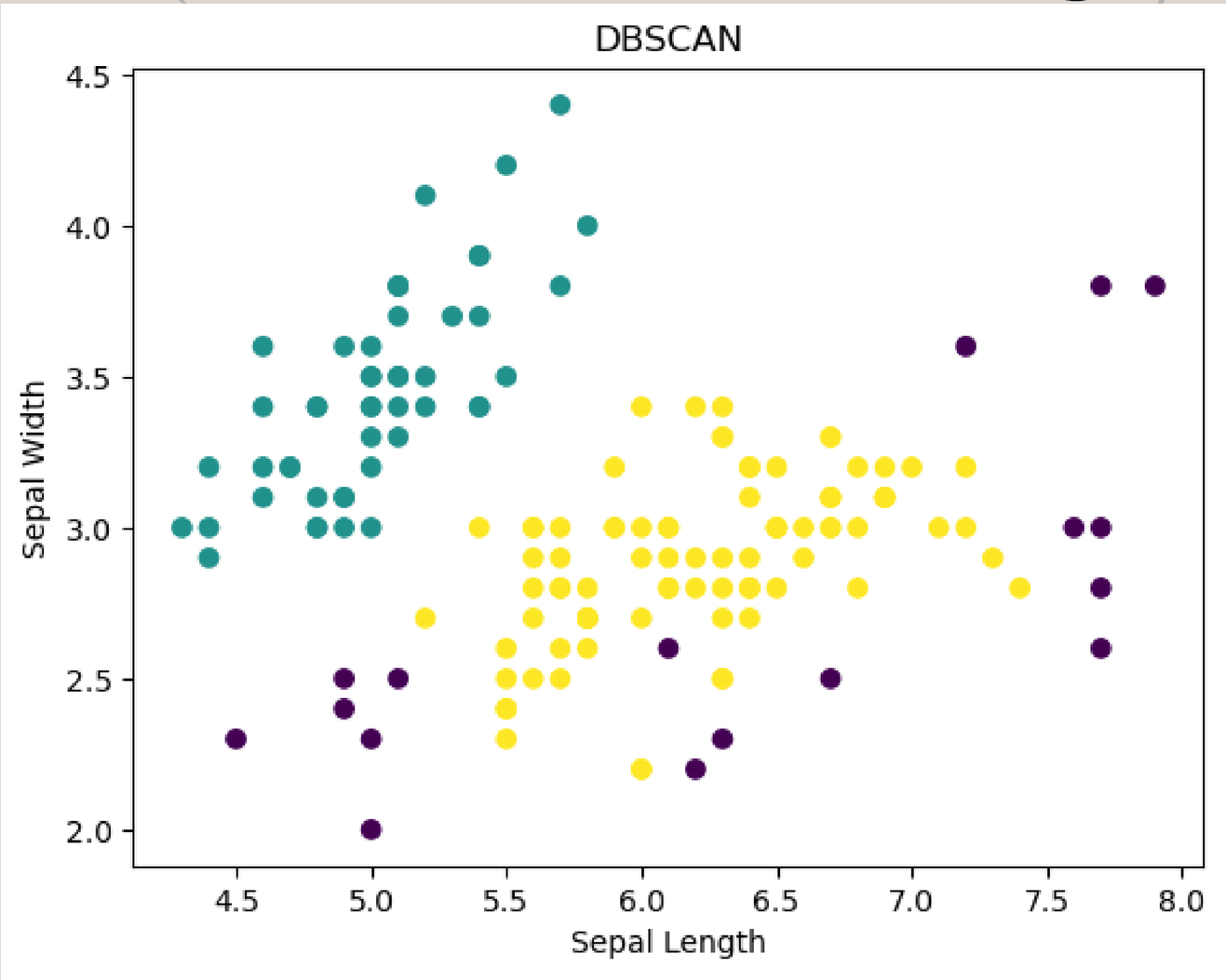


Unsupervised Learning: Comparison to actual values





Unsupervised Learning:



Supervised Learning: Baseline Model

- Chose an appropriate evaluation metric based on classification.
- Split the dataset into training and testing sets.
- Built a baseline model (e.g., logistic regression or decision tree) using default parameters.
- Evaluated the baseline model's performance.

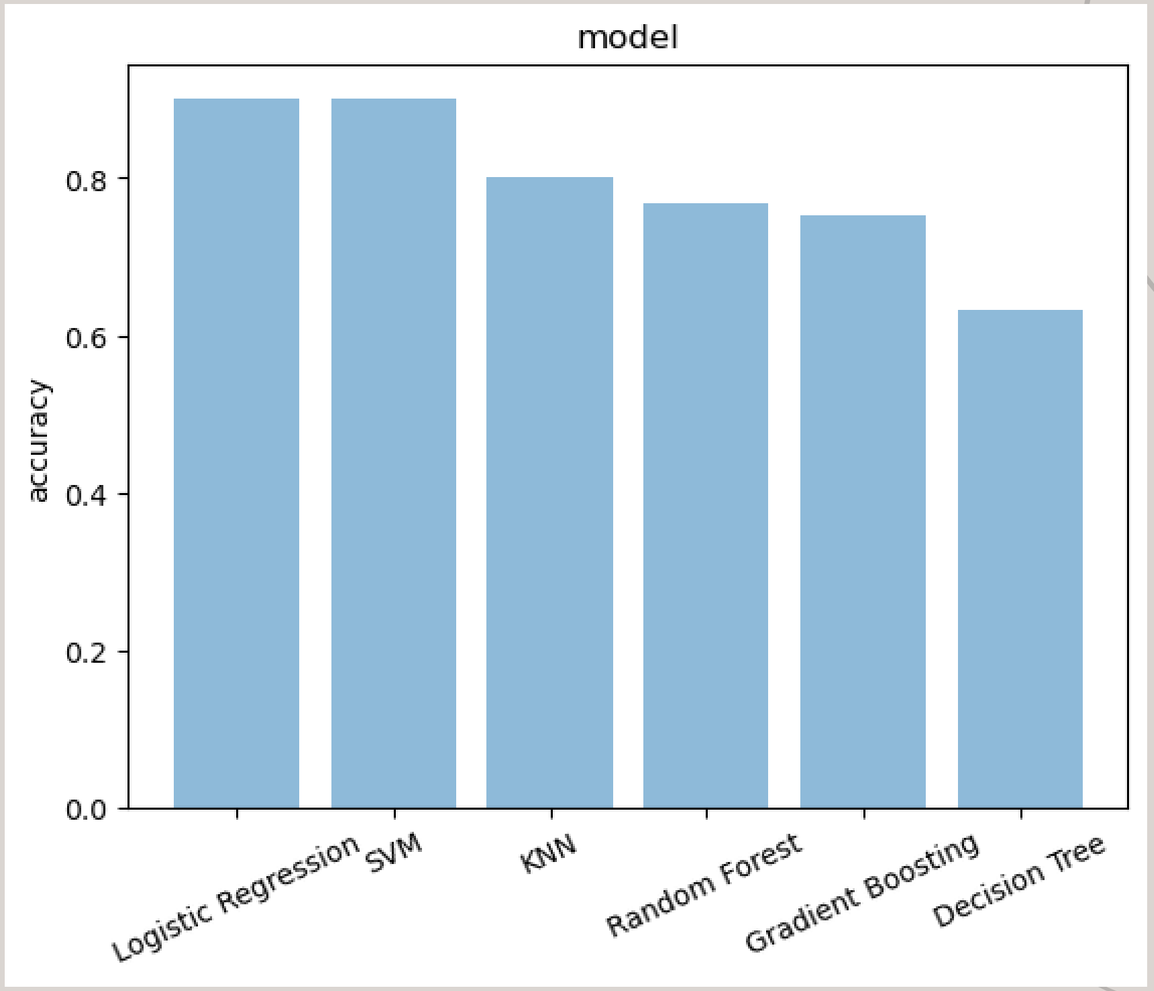
Supervised Learning: Baseline Model

```
Baseline Model Accuracy (Logistic): 0.9
Baseline Model Accuracy (Decision Tree): 0.6333333333333333
```

Model Comparison

- Selected 3–4 machine learning algorithms suitable for the problem.
- Implemented each algorithm and evaluated its performance using cross-validation.
- Compared the performance of algorithms based on evaluation metrics.
- Selected the best-performing algorithm.

Model Comparison



Model Tuning and Ensemble

- Performed hyperparameter tuning on the best-performing algorithm using Grid Search or Random Search.
- Evaluated the tuned model's performance.
- Implemented an ensemble of the top-performing algorithms and compared its performance with the tuned model.

Model Tuning and Ensemble

```
SVC 90.0
LogisticRegression 90.0
KNeighborsClassifier 80.0
VotingClassifier 93.33333333333333
```

Thank You