

Machine Learning Lab: Naive Bayes & Decision Tree Classifier

17/11/2025

1. Naive Bayes Classifier

Dataset: SMS Spam Collection Dataset (spam.csv)

Question:

Use the SMS Spam Collection dataset to build a text classification model that predicts whether a message is spam or not using the Naive Bayes algorithm.

Tasks:

1. Clean and preprocess the text (lowercasing, stopword removal, optional stemming).
2. Convert text to numerical features using TF-IDF.
3. Train a Multinomial Naive Bayes classifier.
4. Report accuracy, precision, recall, and confusion matrix.
5. Show two examples where the model misclassified messages and explain possible reasons.

2. Decision Tree Classifier

Dataset: IRIS dataset (iris.csv)

Question:

Using the IRIS dataset, build a Decision Tree Classifier to classify flower species based on sepal and petal measurements.

Tasks:

1. Load and split the dataset into training and testing sets.
2. Train a Decision Tree using Gini and Entropy criteria separately.
3. Compare model performance for both criteria using accuracy and classification report.
4. Visualize the decision tree and explain the first two splits in simple terms.