

K-Nearest Neighbors (KNN) Documentation

1. Introduction

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm used for classification and regression tasks. It classifies data points based on the majority class of their nearest neighbors.

2. Objective

- Develop an interactive web application using Streamlit for KNN classification.
- Provide users with the ability to upload a dataset, select features, train the model, and evaluate its performance.

3. Dataset Description

- Dataset: **A suitable dataset**
- Features: Describe the input variables
- Target: Describe the variable to predict
- Number of records and attributes

4. Implementation Details

- **Frontend**: Developed using Streamlit for user interaction.
- **Backend**: KNN implemented using `scikit-learn`.
- **Steps**:
 1. Upload the dataset in CSV format.
 2. Select features and target variables.
 3. Choose the number of neighbors (k) for classification.
 4. Train the KNN model.
 5. Evaluate the model using accuracy, confusion matrix, and classification report.

5. Results and Analysis

- Provide model accuracy and other evaluation metrics.
- Visualize decision boundaries for 2D datasets.
- Discuss the effect of different values of k on model performance.

6. Challenges and Solutions

- Managed large datasets using efficient algorithms.
- Selected optimal k using cross-validation.
- Scaled features using StandardScaler to improve accuracy.

7. Conclusion

K-Nearest Neighbors is an easy-to-implement algorithm suitable for small to medium-sized datasets. The web application allows users to experiment with different values of k and analyze model behavior.

8. References

- Scikit-learn Documentation
- Streamlit Documentation
- Dataset Source