CS23334 -Fundamental Of Data Science

240701463

K-Nearest Neighbors (KNN) Classification

Exp : 09

Date: 23-09-2025

Aim:

To implement and evaluate the **K-Nearest Neighbors** (KNeighborsClassifier) algorithm to classify the **variety** of Iris flowers based on their sepal and petal measurements.

Algorithm:

1. **Load Data:** Load the Iris dataset, check info, and verify class balance (value_counts).

2. **Prepare Data:** Separate features (X) and label (Y: variety).

3. **Split Data:** Split the data into training and testing sets (train_test_split).

4. **Train Model:** Initialize and fit the KNeighborsClassifier (n_neighbors=5).

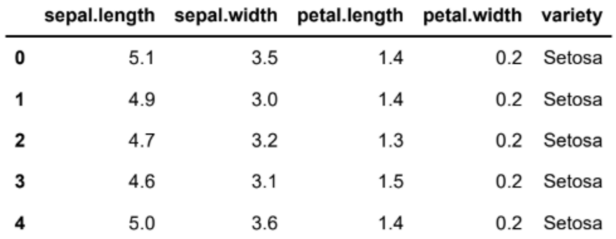
5. **Evaluate Score:** Calculate the model's accuracy score for training and testing sets.

6. **Evaluate Metrics:** Calculate and print the Confusion Matrix and Classification Report.

Code:

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
df=pd.read_csv('Iris.csv')
df.info()
df.variety.value_counts()
df.head()
features=df.iloc[:,:-1].values
label=df.iloc[:,4].values
xtrain, xtest, ytrain, ytest=train_test_split (features, label, test_size=0.2, random_state=0)
model_KNN=KNeighborsClassifier(n_neighbors=5)
model_KNN.fit (xtrain, ytrain)
print(model_KNN.score (xtrain, ytrain))
print(model_KNN.score(xtest, ytest))
confusion_matrix (label, model_KNN.predict(features))
print(classification_report (label, model_KNN.predict(features)))
```

Output:



	sepal.length	sepal.width	petal.length	petal.width	variety
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa

```
<class 'pandas.core.frame. DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
# Column Non-Null Count Dtype
--- ------ -------------- -----
0 sepal.length 150 non-null float64
1 sepal.width 150 non-null float64
2 petal.length 150 non-null float64
3 petal.width 150 non-null float64
4 variety 150 non-null object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
Setosa 50
Versicolor 50
Virginica 50
Name: variety, dtype: int64
0.9583333333333334
1.0
array([[50, 0, 0], [ 0, 47, 3], [ 0, 2, 48]], dtype=int64)
          precision recall f1-score support
Setosa 1.00 1.00 1.00 50
Versicolor 0.96 0.94 0.95 50
Virginica 0.94 0.96 0.95 50
```

```
Accuracy                 0.97 150
macro avg 0.97 0.97 0.97 150
weighted avg 0.97 0.97 0.97 150
```

Result:

The K-Nearest Neighbors (KNN) model, configured with n_neighbors=5, was successfully implemented and trained to classify Iris varieties. The model achieved a high level of accuracy, with a **Training Score of 0.958** and a **Testing Score of 1.0**. The overall accuracy reported was **0.97**. The Confusion Matrix shows that 5 Versicolor and Virginica samples were misclassified (3 Versicolor predicted as Virginica, 2 Virginica predicted as Versicolor), while the Setosa class was perfectly classified. The Classification Report confirms excellent performance across all classes, with high precision and recall values (all above 0.94).