ml3

November 3, 2024

Classification Analysis: Implement K-Nearest Neighbors' algorithm on social network ad dataset. Compute confusion matrix, accuracy, error rate, precision and recall on the given dataset. Dataset link: https://www.kaggle.com/datasets/rakeshrau/socialnetwork-ads

[1]:

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score,␣
  ↪recall_score

import warnings
warnings.filterwarnings("ignore")
```

[2]:

```
df=pd.read_csv(r"C:\Users\dell\Desktop\DMV and ML\ML␣
  ↪Datasets\Social_Network_Ads.csv")
```

[3]:

```
df.head()
```

[3]:

```
    User ID  Gender  Age  EstimatedSalary  Purchased
0  15624510    Male   19            19000          0
1  15810944    Male   35            20000          0
2  15668575  Female   26            43000          0
3  15603246  Female   27            57000          0
4  15804002    Male   19            76000          0
```

[4]:

```
x=df.iloc[:,[2,3]] # Features (Age and EstimatedSalary columns)
y=df["Purchased"] # Target variable
```

[5]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.
  ↪2,random_state=100)
```

[6]:

```
sc = StandardScaler()
x_train=sc.fit_transform(x_train)
x_test=sc.transform(x_test)
```

[7]:

```
k=5 # Number of neighbors
model= KNeighborsClassifier(n_neighbors=k)
model.fit(x_train,y_train)

y_pred=model.predict(x_test)
```

[8]:

```
confusion_matrix(y_test,y_pred)
```

[8]:

```
array([[47,  2],
       [ 6, 25]], dtype=int64)
```

[9]:

```
accuracy = accuracy_score(y_test, y_pred)
error_rate = 1 - accuracy
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)

print("Accuracy:", accuracy*100,"%")
print("Error Rate:", error_rate)
print("Precision:", precision)
print("Recall:", recall)
```

```
Accuracy: 90.0 %
Error Rate: 0.09999999999999998
Precision: 0.9259259259259259
Recall: 0.8064516129032258
```

KNN is a supervised machine learning algorithm used for classification and regression tasks. It classifies a data point based on how its neighbors are classified. When given a new data point, the algorithm identifies the ‘k’ closest points in the training set (based on a distance metric like Euclidean distance) and assigns the most common label (in classification) or the average of the labels (in regression) among those neighbors to the new data point.

[ ]:

```

```