

app.py:

Συνολικά, αυτός ο κώδικας δημιουργεί μια εφαρμογή Streamlit όπου οι χρήστες μπορούν να ανεβάσουν ένα αρχείο CSV, να εκτελέσουν Quadratic Discriminant Analysis και Spectral Clustering στα δεδομένα και να προβάλουν τα αποτελέσματα διαδραστικά.

1. Ο κώδικας εισάγει τις απαραίτητες βιβλιοθήκες:

- ``streamlit``: Μια βιβλιοθήκη Python για τη δημιουργία διαδραστικών εφαρμογών ιστού.
- ``pandas``: Μια βιβλιοθήκη για τον χειρισμό και την ανάλυση δεδομένων.
- ``numpy``: Μια βιβλιοθήκη για αριθμητικές πράξεις.
- ``QuadraticDiscriminantAnalysis`` από ``sklearn.discriminant_analysis``: Μια κλάση για την εκτέλεση τετραγωνικής διακριτικής ανάλυσης.
- ``SpectralClustering`` από ``sklearn.cluster``: Μια κλάση για την εκτέλεση φασματικής ομαδοποίησης.
- ``accuracy_score`` από ``sklearn.metrics``: Μια συνάρτηση για τον υπολογισμό της ακρίβειας ταξινόμησης.

2. Ο κώδικας ορίζει διάφορες συναρτήσεις:

- ``load_data(file)``: Διαβάζει ένα αρχείο CSV και επιστρέφει τα δεδομένα ως πλαίσιο δεδομένων pandas DataFrame.
- ``run_qda(data)``: Εκτελεί Quadratic Discriminant Analysis στα δεδομένα και επιστρέφει την ακρίβεια και την καθαρότητα των συστάδων.
- ``run_spectral_clustering(data, n_clusters)``: Εκτελεί Spectral Clustering στα δεδομένα με καθορισμένο αριθμό συστάδων και επιστρέφει την καθαρότητα των συστάδων.
- ``calculate_purity(data, y_pred)``: Υπολογίζει το δείκτη καθαρότητας συστάδων δεδομένων των δεδομένων και των προβλεπόμενων ετικετών συστάδων.

3. Η εφαρμογή Streamlit δημιουργείται με τη χρήση της ``st.title()`` για να οριστεί ο τίτλος της εφαρμογής.

4. Ο χρήστης καλείται να φορτώσει ένα αρχείο CSV χρησιμοποιώντας την ``st.file_uploader()``. Το αρχείο που έχει μεταφορτωθεί αποθηκεύεται στη μεταβλητή ``uploaded_file``.

5. Όταν φορτωθεί ένα αρχείο, καλείται η συνάρτηση ``load_data()`` για να φορτώσει τα δεδομένα σε ένα DataFrame.

6. Εάν ο χρήστης κάνει κλικ στο κουμπί "Run QDA", καλείται η συνάρτηση ``run_qda()`` και υπολογίζονται η ακρίβεια και η καθαρότητα των συστάδων με τη χρήση της Quadratic Discriminant Analysis. Τα αποτελέσματα εμφανίζονται χρησιμοποιώντας την ``st.write()``.

7. Ο χρήστης μπορεί να ρυθμίσει τον αριθμό των συστάδων χρησιμοποιώντας το ρυθμιστικό της πλαϊνής γραμμής.

8. Εάν ο χρήστης κάνει κλικ στο κουμπί "Run Spectral Clustering", καλείται η συνάρτηση ``run_spectral_clustering()`` με τον καθορισμένο αριθμό συστάδων. Εκτελείται η Spectral Clustering και υπολογίζεται και εμφανίζεται η καθαρότητα των συστάδων.

createCSV.py:

Δημιουργεί ένα csv αρχείο με random numbers και το κατάλληλο formatting για να χρησιμοποιηθεί από το παραπάνω πρόγραμμα.

TO RUN THE DOCKER CONTAINER:

```
docker build -t streamlit .  
docker run -p 8501:8501 streamlit
```