

# Ανάπτυξη Εφαρμογής για Ανάλυση Δεδομένων

Ομαδική Αναφορά

May 29, 2024

## 1 Εισαγωγή

Η ανάπτυξη εφαρμογών για την ανάλυση δεδομένων είναι κρίσιμη για την εξαγωγή χρήσιμων πληροφοριών από τα δεδομένα. Αυτή η εφαρμογή έχει σχεδιαστεί για να παρέχει εύχρηστα εργαλεία για τη loading, την επεξεργασία και την οπτικοποίηση δεδομένων, καθώς και για την εφαρμογή αλγορίθμων machine learning.

## 2 Σχεδιασμός της Εφαρμογής

Η εφαρμογή σχεδιάστηκε χρησιμοποιώντας την αρχιτεκτονική MVC (Model-View-Controller). Το παρακάτω διάγραμμα UML απεικονίζει την αρχιτεκτονική της εφαρμογής.

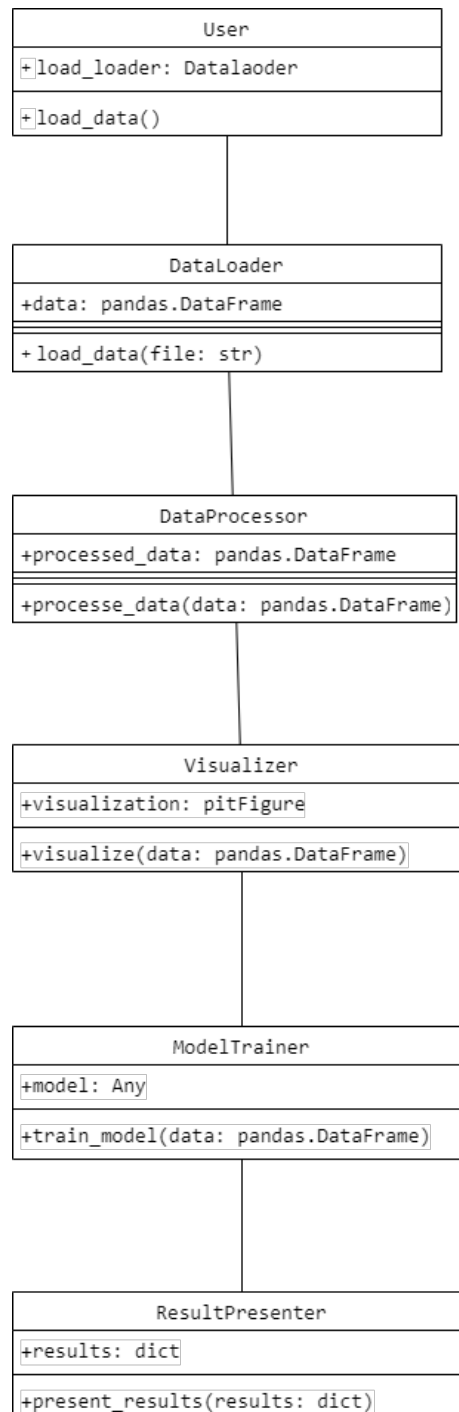


Figure 1: UML Διάγραμμα της Εφαρμογής

### 3 Υλοποίηση

Η υλοποίηση της εφαρμογής έγινε με τη χρήση της γλώσσας προγραμματισμού Python και των βιβλιοθηκών Streamlit, scikit-learn και matplotlib. Παρακάτω παρατίθεται ένα δείγμα κώδικα για τη φόρτωση και την επεξεργασία των δεδομένων.

```
import streamlit as st
import pandas as pd

# Φόρτωση δεδομένων
uploaded_file = st.file_uploader("Επιλέξτε ένα αρχείο CSV ή Excel", type=["csv", "xlsx"])
if uploaded_file is not None:
    data = pd.read_csv(uploaded_file) if uploaded_file.name.endswith('.csv') else pd.read_excel(uploaded_file)
    st.write("Data Preview:")
    st.write(data.head())
```

### 4 Αποτελέσματα

Η εφαρμογή παρέχει εργαλεία για την ανάλυση των δεδομένων μέσω 2D οπτικοποιήσεων και αλγορίθμων machine learning. Τα αποτελέσματα των αλγορίθμων παρουσιάζονται αναλυτικά, με τη σύγκριση των αποδόσεων των διαφορετικών αλγορίθμων.

### 5 Συμπεράσματα

Η εφαρμογή αυτή αποδεικνύει τη χρησιμότητα των εργαλείων ανάλυσης δεδομένων για την εξαγωγή χρήσιμων πληροφοριών. Η χρήση τεχνολογιών Python και Streamlit επιτρέπει την ανάπτυξη ευέλικτων και επεκτάσιμων λύσεων.

### 6 Ομάδα Ανάπτυξης

Η ανάπτυξη της εφαρμογής πραγματοποιήθηκε από την ομάδα μας στο πλαίσιο του μαθήματος Software Engineering. Κάθε μέλος της ομάδας είχε συγκεκριμένες αρμοδιότητες, όπως περιγράφεται παρακάτω:

- **Γεωργιάδης Ιωάννης (Π2020051):** Ανάπτυξη Κώδικα, Υλοποίηση και Τεκμηρίωση
- **Λάζαρος Γκόβας (Π2020065):** Ανάλυση Δεδομένων, Προετοιμασία και Καθαρισμός Δεδομένων
- **Βασίλης Στάθης (Π2020040):** Οπτικοποίηση Δεδομένων, Δημιουργία Διαγραμμάτων και Γραφημάτων
- **Όλοι μαζί:** Συνεργασία στον Σχεδιασμό και Δοκιμές

## 7 Κύκλος Ζωής Έκδοσης Λογισμικού

Για την ανάπτυξη της εφαρμογής μας, επιλέξαμε να χρησιμοποιήσουμε το μοντέλο Agile. Το Agile είναι ένα ευέλικτο μοντέλο ανάπτυξης λογισμικού που επιτρέπει την επαναληπτική και σταδιακή βελτίωση του προϊόντος. Ακολουθεί μια σύντομη περιγραφή του τρόπου με τον οποίο εφαρμόσαμε το Agile στη διαδικασία ανάπτυξης:

### 7.1 Στάδια του Agile

1. **Σχεδιασμός:** Στο στάδιο του σχεδιασμού, η ομάδα μας συγκεντρώθηκε για να καθορίσει τις απαιτήσεις του έργου και να δημιουργήσει το αρχικό διάγραμμα τάξεων UML. Κάθε μέλος ανέλαβε συγκεκριμένες ευθύνες για τον σχεδιασμό της εφαρμογής.
2. **Ανάπτυξη:** Η εργασία χωρίστηκε σε μικρότερα κομμάτια (sprints), με κάθε sprint να έχει διάρκεια δύο εβδομάδων. Στη διάρκεια αυτών των σπριντ:
  - Ο Ιωάννης ανέπτυξε τον βασικό κώδικα και διασφάλισε την υλοποίηση των κύριων λειτουργιών.
  - Ο Λάζαρος προετοίμασε και καθάρισε τα δεδομένα, φροντίζοντας για την ακρίβεια και την ποιότητα των δεδομένων που χρησιμοποιήθηκαν.
  - Ο Βασίλης ασχολήθηκε με την οπτικοποίηση των δεδομένων, δημιουργώντας διαγράμματα και γραφήματα για την καλύτερη κατανόηση των αποτελεσμάτων.
3. **Δοκιμές:** Μετά από κάθε sprint, η ομάδα πραγματοποίησε δοκιμές για να διασφαλίσει ότι όλες οι λειτουργίες της εφαρμογής λειτουργούσαν σωστά. Οι δοκιμές περιλάμβαναν τόσο unit tests όσο και ολοκληρωμένες δοκιμές της εφαρμογής.
4. **Παράδοση:** Στο τέλος κάθε sprint, η ομάδα παρέδωσε μια ενημερωμένη έκδοση της εφαρμογής. Η τελική έκδοση παραδόθηκε μέσω Docker για εύκολη εγκατάσταση και χρήση. Το GitHub χρησιμοποιήθηκε για τη διαχείριση του κώδικα και τη συνεργασία της ομάδας.

Η χρήση του Agile μας επέτρεψε να ανταποκριθούμε γρήγορα στις αλλαγές των απαιτήσεων και να βελτιώσουμε συνεχώς την εφαρμογή μας μέσω επαναληπτικών κύκλων ανάπτυξης.

## 8 Σύνδεσμοι

- <https://github.com/Giannisgeorgiadis/Johnapp>
- <https://www.overleaf.com/1122455446rzdmbhzhxgxybh9c6012>
- <https://www.overleaf.com/read/kqpmxswdxbygcd087e>

- <https://www.overleaf.com/project/66532bc7e52859381a3b4b95>

Για κάποιο λόγο δεν ανοίγει κανένα link απο το Overleaf που βάζω εδώ στην διεύθυνση με οποιονδήποτε τρόπο οπότε βάζω 3 link και αν τυχόν δεν ανοίξουν και σε εσάς θα ήθελα παρακαλώ να επικοινωνήσετε μαζί μας

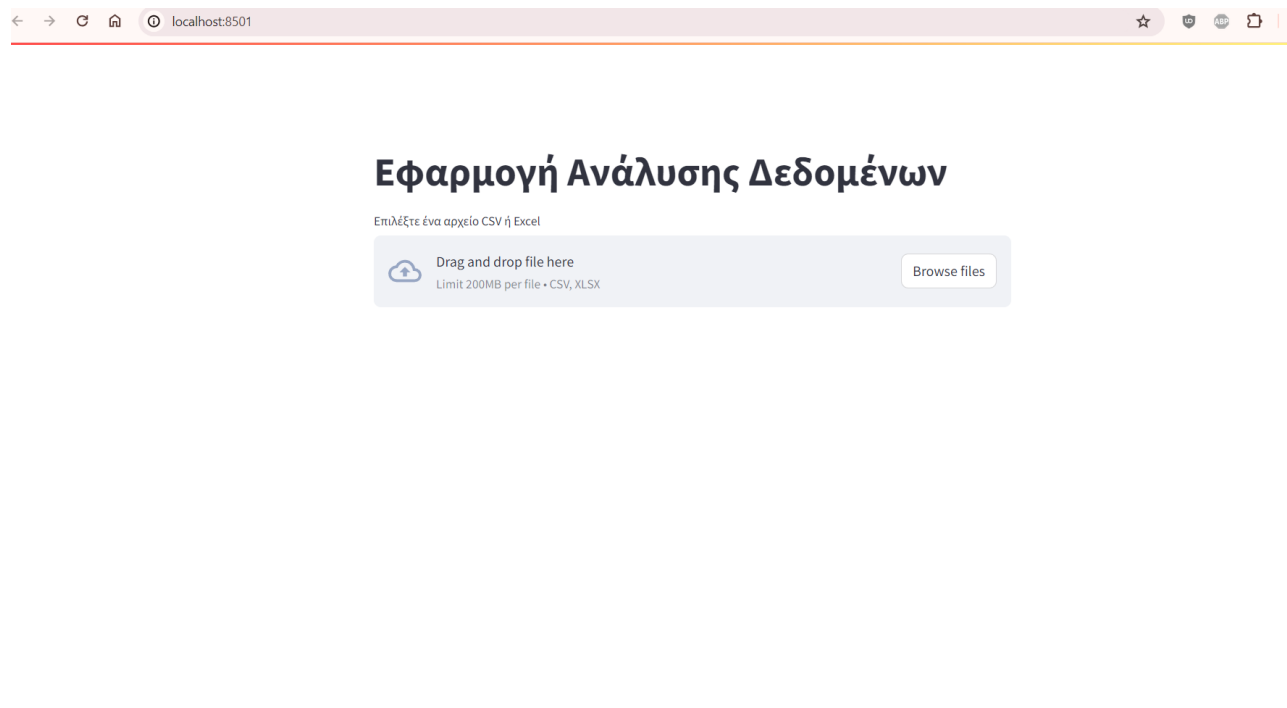



Figure 2: Screenshot λειτουργίας της Εφαρμογής


←→🔄🏠🔍localhost:8501🔍☆🛡️🌐🗂️

## Εφαρμογή Ανάλυσης Δεδομένων

Επιλέξτε ένα αρχείο CSV ή Excel

 Drag and drop file here  
Limit 200MB per file • CSV, XLSX

Browse files

 Παράδειγμα\_Λειτουργίας\_Εφαρμογής.csv 336.0B

✕

Data Preview:

	feature1	feature2	feature3	label
0	1	2.1	3.3	A
1	1.2	2	3.1	A
2	1.4	2.2	3	B
3	1.5	2.3	2.9	B
4	1.1	2.4	3.5	A

Data Info:

None

Label Column: label

Label is categorical

Processed Data:

	feature1	feature2	feature3	label
0	1	2.1	3.3	0
1	1.2	2	3.1	0

Figure 3: Screenshot λειτουργίας της Εφαρμογής

6

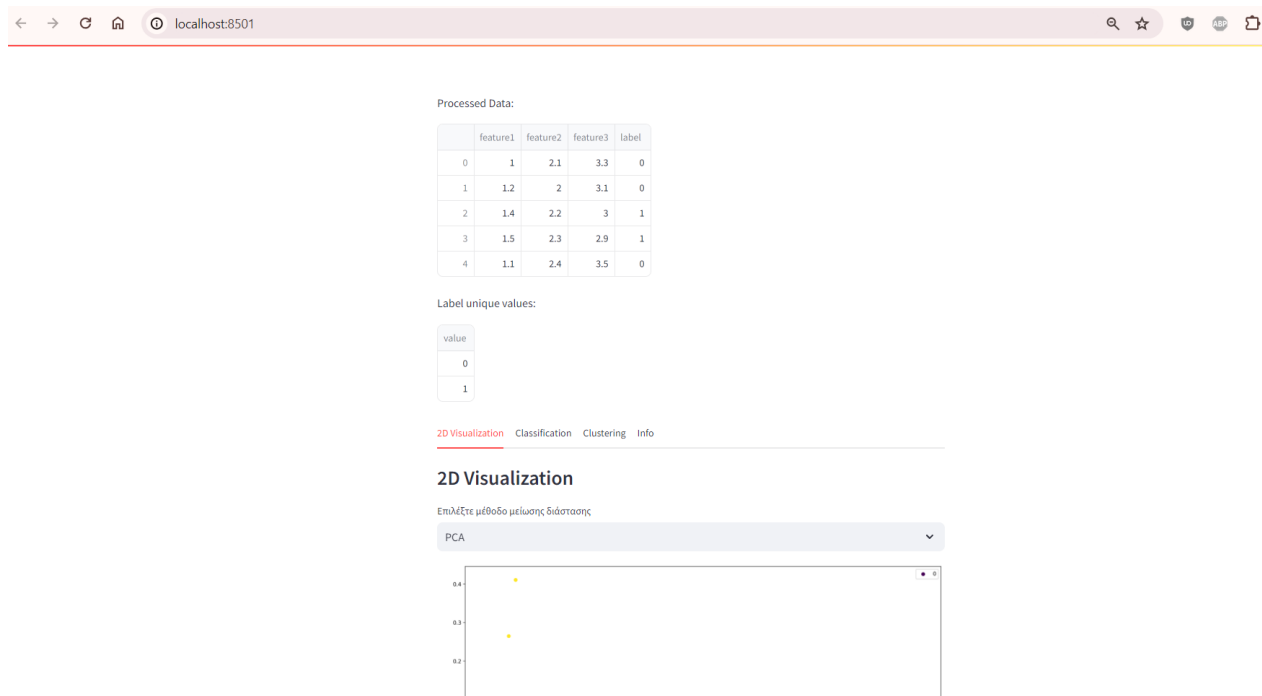


Figure 4: Screenshot λειτουργίας της Εφαρμογής

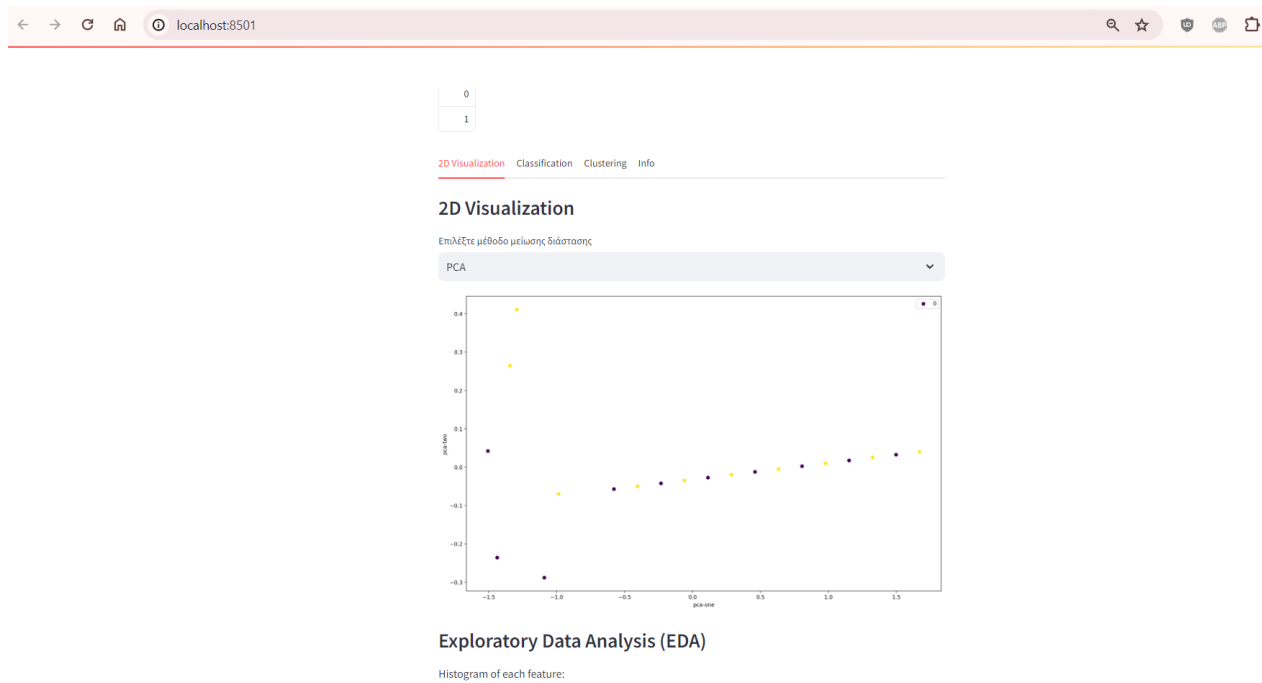


Figure 5: Screenshot λειτουργίας της Εφαρμογής



## Exploratory Data Analysis (EDA)

Histogram of each feature:

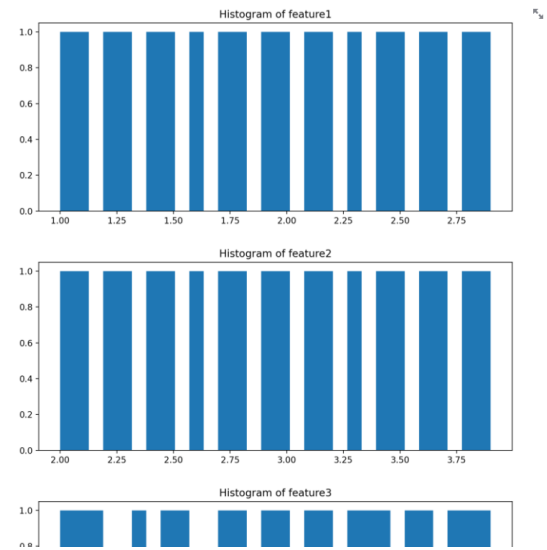


Figure 6: Screenshot λειτουργίας της Εφαρμογής

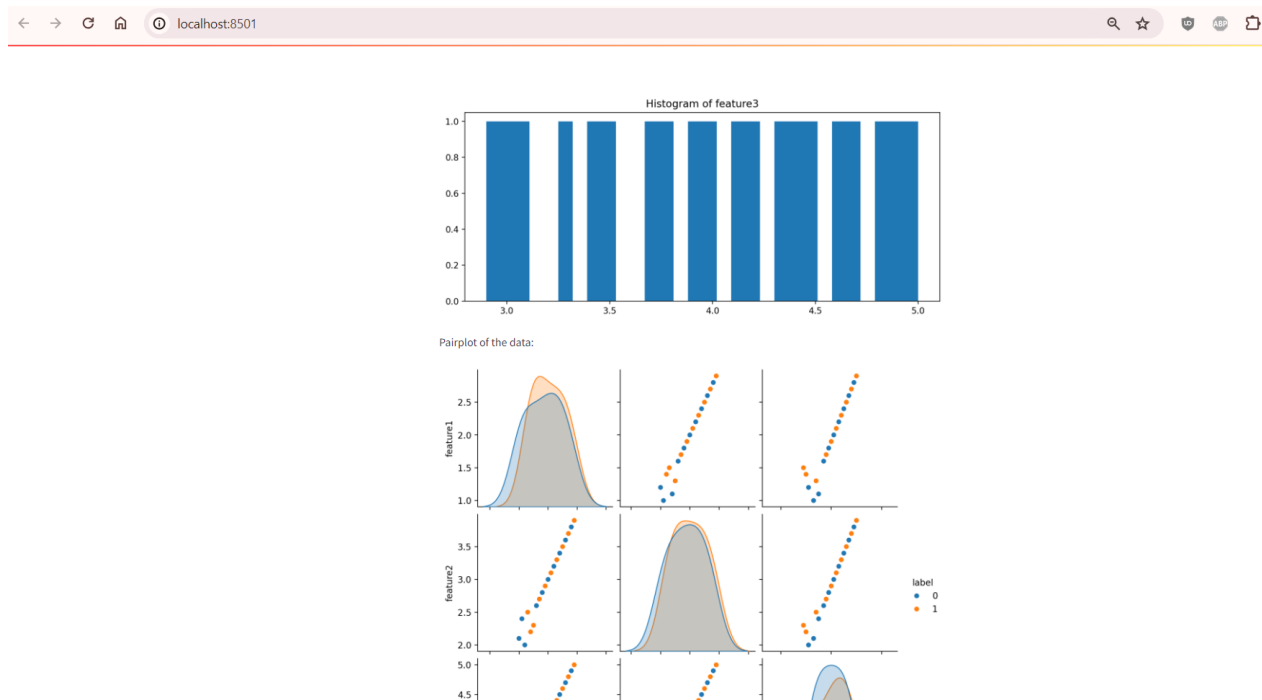


Figure 7: Screenshot λειτουργίας της Εφαρμογής

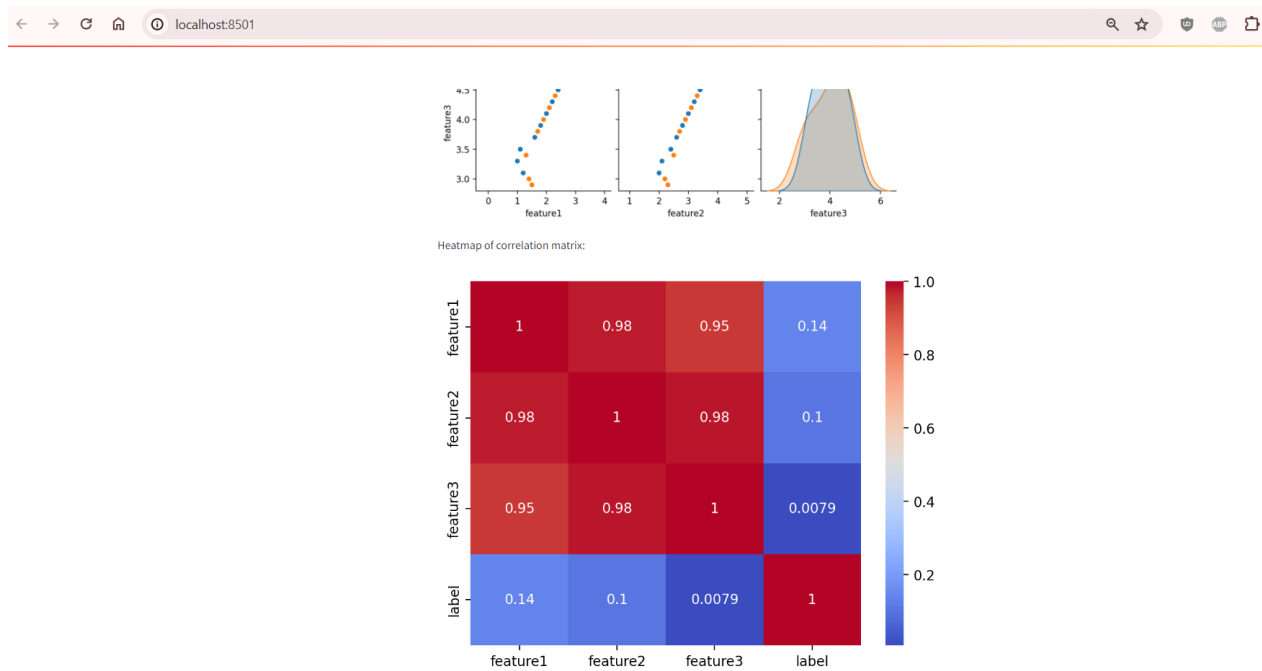


Figure 8: Screenshot λειτουργίας της Εφαρμογής

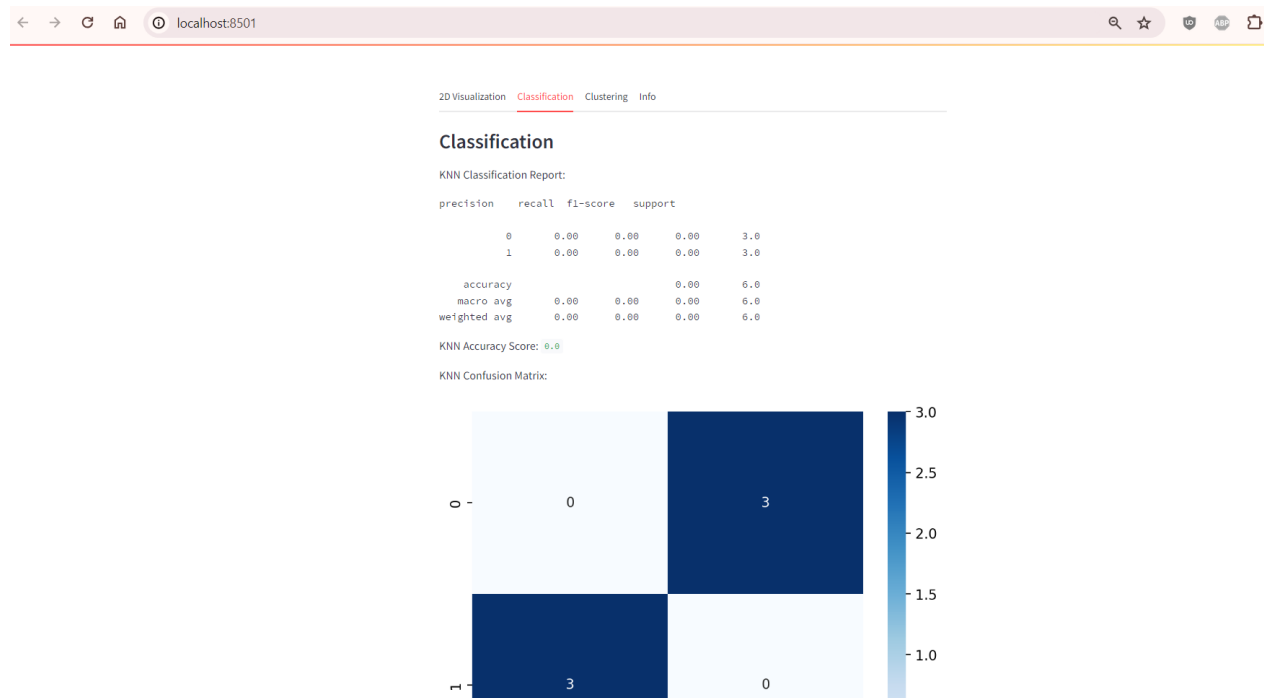


Figure 9: Screenshot λειτουργίας της Εφαρμογής



Figure 10: Screenshot λειτουργίας της Εφαρμογής

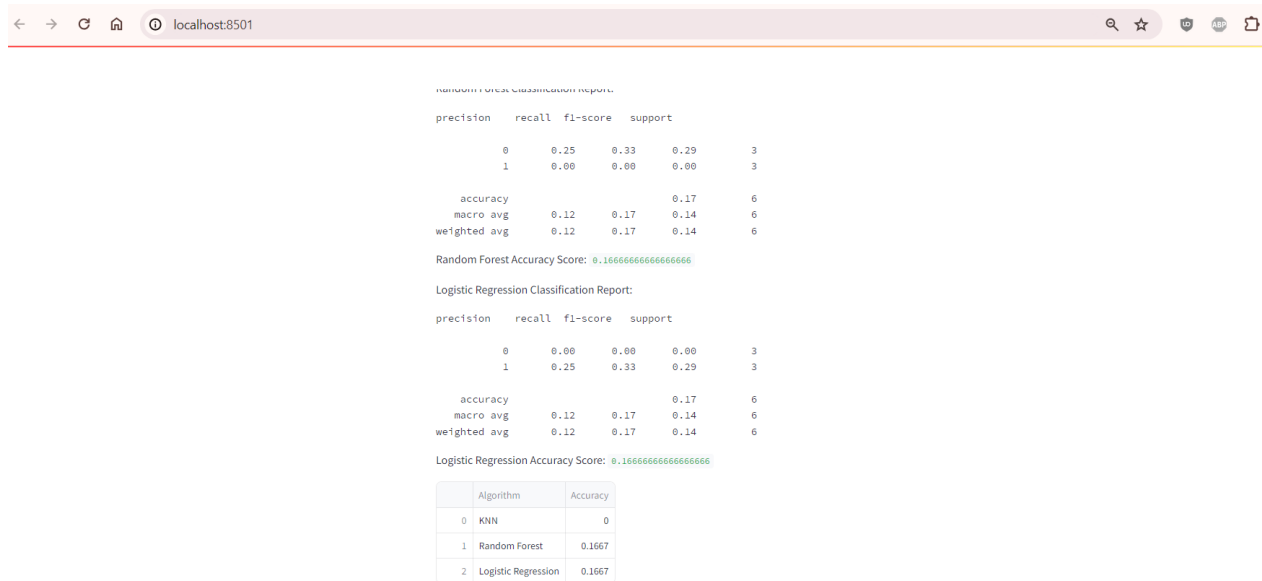


Figure 11: Screenshot λειτουργίας της Εφαρμογής

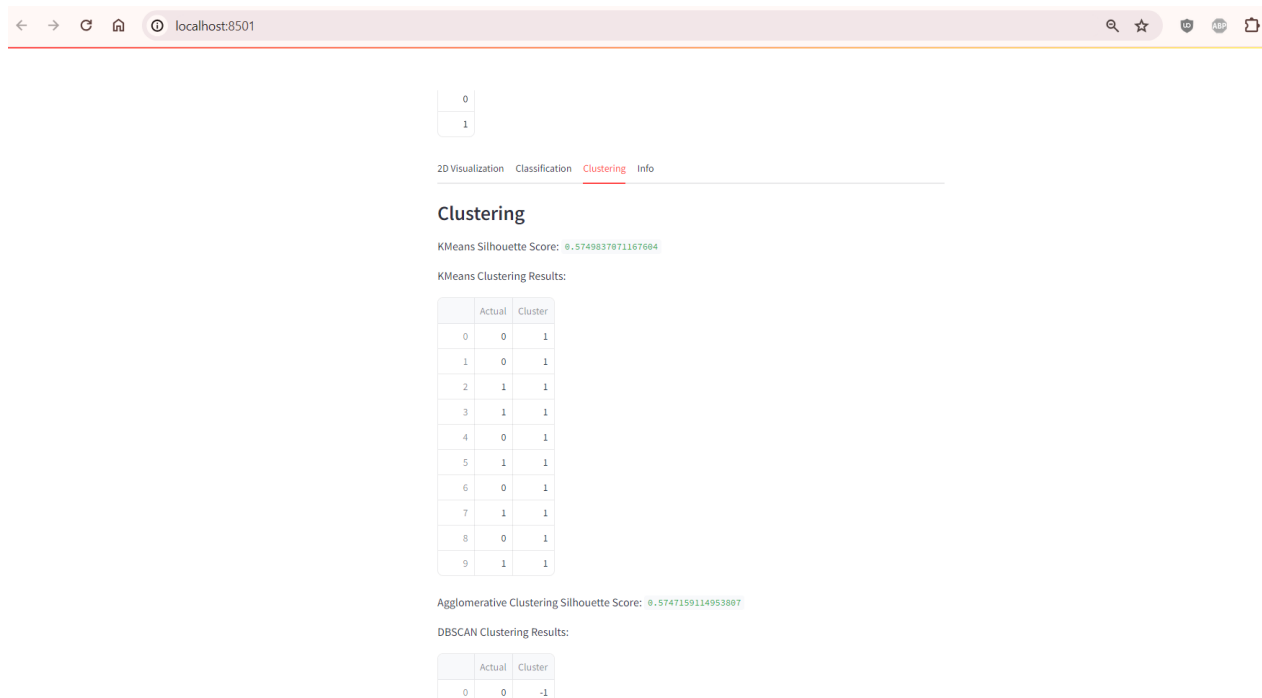


Figure 12: Screenshot λειτουργίας της Εφαρμογής

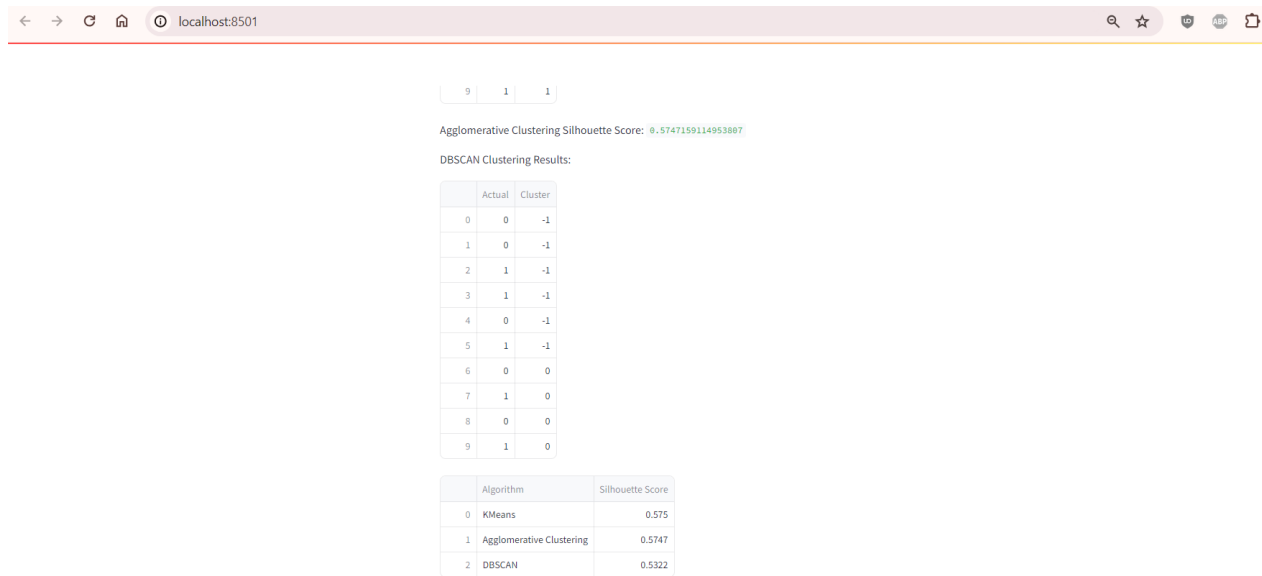


Figure 13: Screenshot λειτουργίας της Εφαρμογής



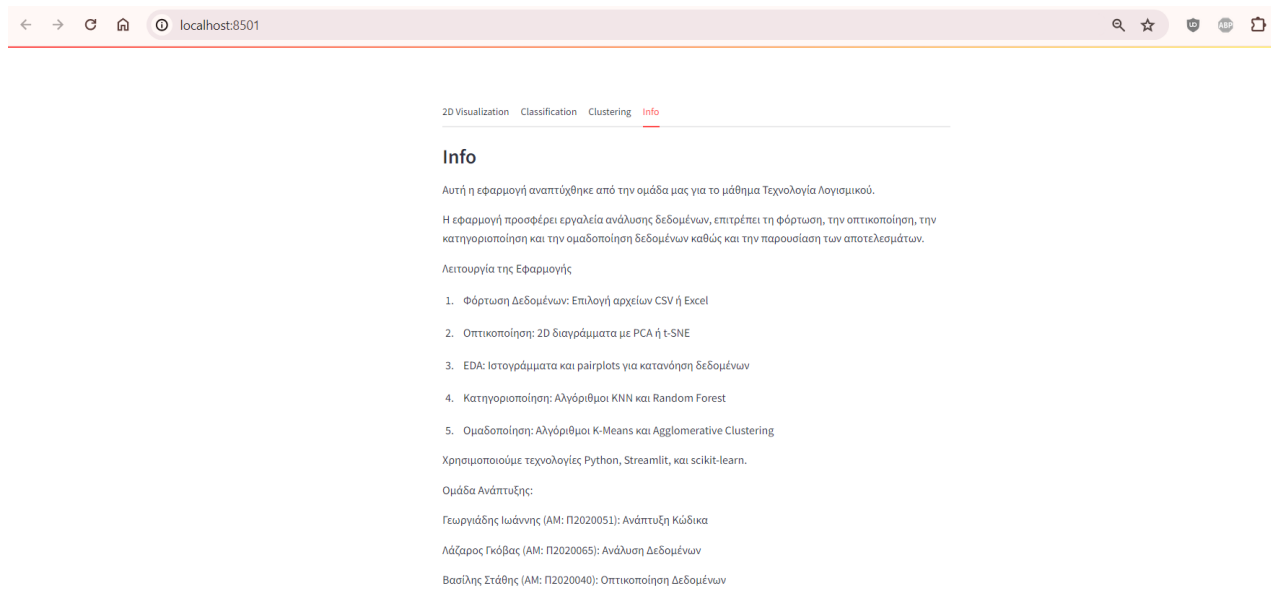


Figure 14: Screenshot λειτουργίας της Εφαρμογής