

Συσταδοποίηση σημάτων πρωτεϊνών

Πουλάκης Βασίλειος

Ελληνικό Ανοικτό Πανεπιστήμιο

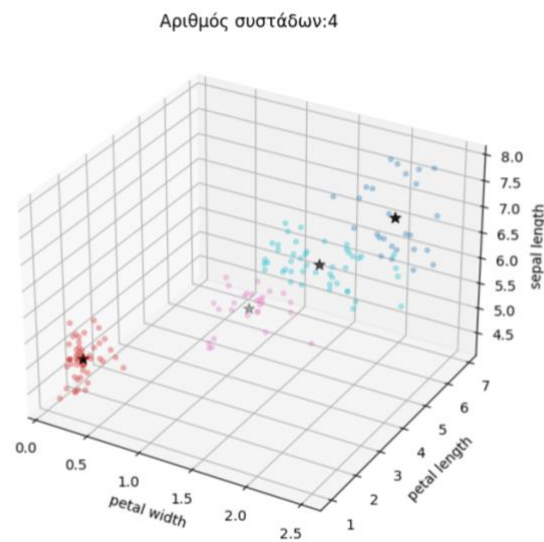
Τμήμα πληροφορικής, προγραμματισμός σε γλώσσα python

Πίνακας περιεχομένων

<i>Περίληψη</i>	<i>3</i>
<i>Αρχικός σχεδιασμός</i>	<i>4</i>
<i>Ενσωμάτωση γραφημάτων Matplotlib στο Tkinter με το FigureCanvasTkAgg</i>	<i>6</i>
<i>Ο Αλγόριθμος Kmeans</i>	<i>10</i>
<i>Υπολογισμός στατιστικών</i>	<i>12</i>
<i>Εισαγωγή Dataset.....</i>	<i>12</i>
<i>Προβολή Dataset.....</i>	<i>14</i>
<i>Αναφορές</i>	<i>18</i>

Περίληψη

Σχεδιασμός και υλοποίηση εφαρμογής μηχανικής μάθησης(συσταδοποίησης) με τη χρήση αντικειμενοστραφούς προγραμματισμού σε γλώσσα python.Η εφαρμογή υλοποιήθηκε στα πλαίσια της τελικής εξέτασης του μαθήματος ΠΛΗΠΡΟ του ΕΑΠ 2022/2023

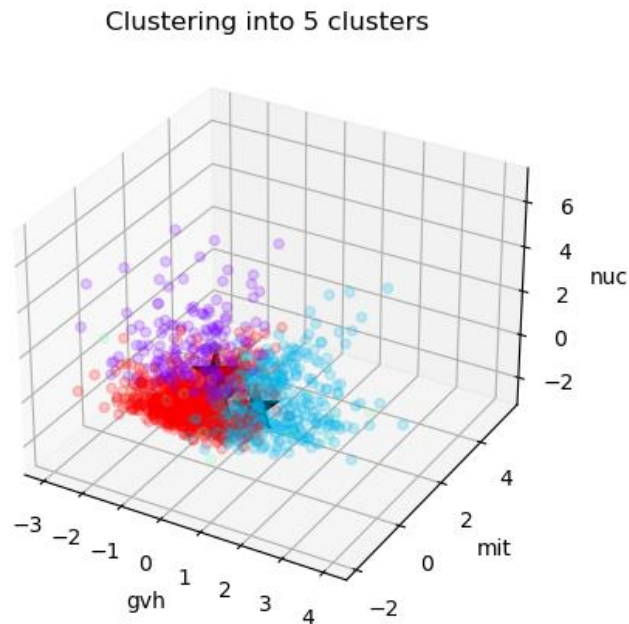


Εικόνα 1 To dataset Iris(εφαρμογή LoadanyDatat.py)

Αρχικός σχεδιασμός

Η εκφώνηση του project09 για την εκπόνηση της τελικής εργασίας στα πλαίσια του Μαθήματος προγραμματισμός σε γλώσσα Python προέβλεπε την σχεδίαση και υλοποίηση Εφαρμογής ομαδοποίησης σημάτων πρωτεϊνών. Ο χρήστης θα μπορεί μέσα από το περιβάλλον χρήσης να ορίσει τα χαρακτηριστικά που θα λαμβάνονται υπόψη για να εκτελεστεί ομαδοποίηση. Στην φάση του σχεδιασμού, σε συνεργασία με τη κ. Συρίγου αποφασίσαμε να χωρίσουμε το project σε 2 μέρη. Το πρώτο αφορούσε το κομμάτι της εξαγωγής στατιστικών από το σύνολο δεδομένων yeast.data καθώς και τη δημιουργία γραφικής παράστασης των στατιστικών αλλά και η υλοποίηση της ομαδοποίησης με τον αλγόριθμο *kmeans*.

Το δεύτερο αφορούσε το σχεδιασμό και την υλοποίηση της γραφικής διεπαφής χρήστη (Graphic User Interface) το οποίο και ανέλαβε εκείνη, πάνω στο οποίο και θα προσαρμοζόταν το κομμάτι της ‘μηχανικής μάθησης’ της εφαρμογής όπου και ανέλαβα προσωπικά.



Εικόνα 2 συσταδοποίηση σε 5 clusters 3 χαρακτηριστικών του dataset yeast

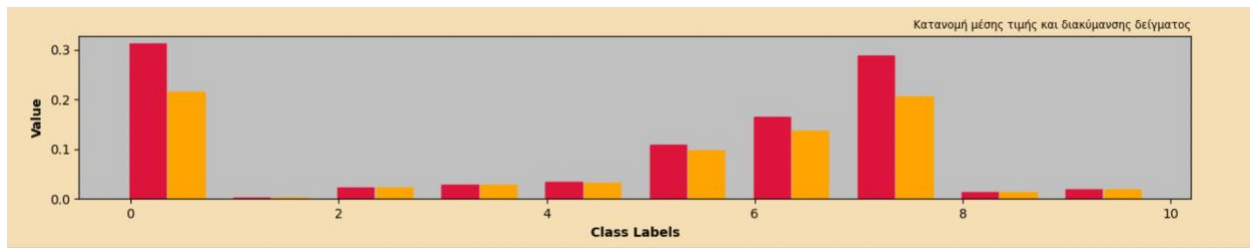
Αφού μελέτησα την περιγραφή του project09 ξεκίνησα με την χρήση *Jupyter notebook* μια πιο ‘κλασσική’ προσέγγιση του προβλήματος της επεξεργασίας του συνόλου δεδομένων yeast.data .

(Ζητούμενο 1-2)

Αρχικά με τη χρήση της βιβλιοθήκης pandas ‘διάβασα’ το σύνολο δεδομένων διέγραψα την πρώτη στήλη και μετέτρεψα τις τιμές της τελευταίας στήλης από κατηγορικού τύπου σε binary, με χρήση της μεθόδου `pd.get_dummies` όπου και προέκυψε μια νέα μορφή του συνόλου δεδομένων.

(Ζητούμενο 3)

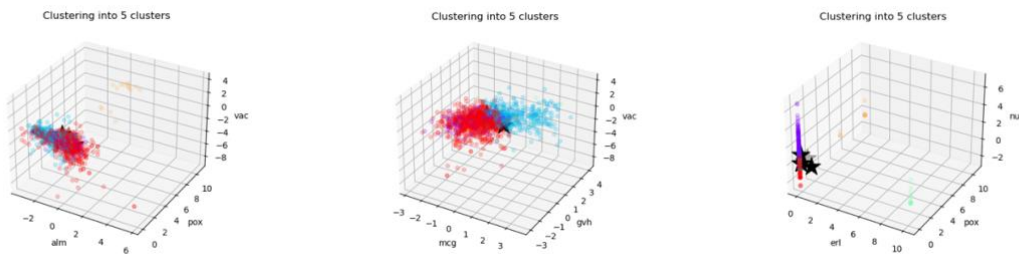
Κατόπιν, βάση αυτής της μορφής του dataset υπολογίστηκαν τα στατιστικά στοιχεία της γραφικής παράστασης που εμφανίζεται στο αρχείο `clustering.py` της τελικής εφαρμογής.



Εικόνα 3 Ραβδόγραμμα για τη μεταβλητή κλάσης Class Distribution τον Αριθμητικό Μέσο και τη Διακύμανση

(Ζητούμενο 4)

Στη συνέχεια έφτιαξα ένα μοντέλο Kmeans που εκτύπωνε *3D scatter plot* για όλους του πιθανούς συνδυασμούς ανά 3 όλων των χαρακτηριστικών με τη βοήθεια της βιβλιοθήκης *itertools*

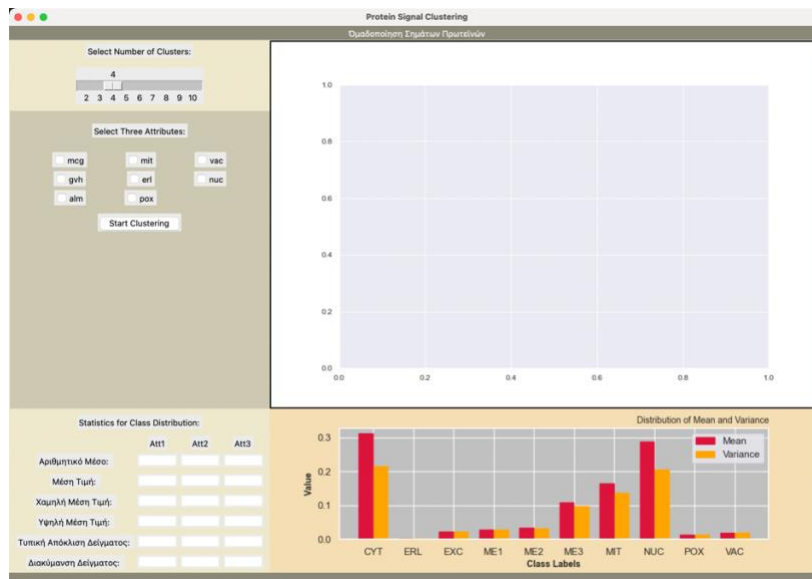


Εικόνα 4 Μερικοί από του συνδυασμούς που προέκυψαν.

Ενσωμάτωση γραφημάτων Matplotlib στο Tkinter με το FigureCanvasTkAgg

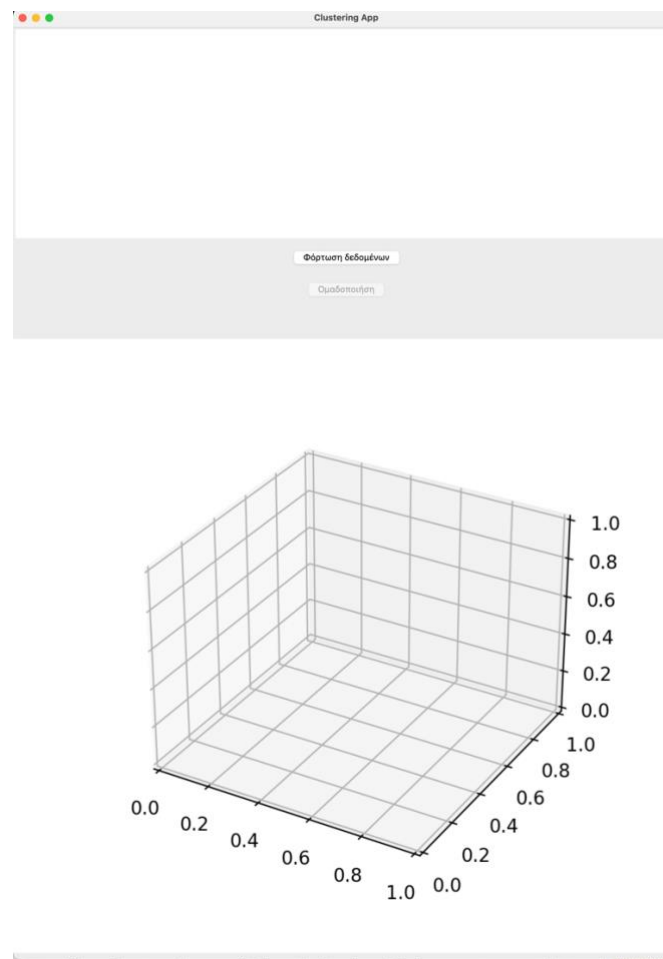
Ωστόσο η κ. Συρίγου είχε ετοιμάσει τη γραφική διεπαφή της τελικής εφαρμογής η οποία όμως δεν συνδέονταν ακόμα με το κομμάτι που είχα ετοιμάσει.

Έτσι ξεκινήσαμε μια διαφορετική προσέγγιση του προβλήματος και φτιάξαμε μια ξεχωριστή εφαρμογή `clusters.py` που αποτέλεσε και την λύση του προβλήματος της διαδραστικότητας χρήστη-εφαρμογής.



Εικόνα 5 εφαρμογή με διαχειριστή γεωμετρίας *grid* πριν την εισαγωγή του *Canvas*

Το *Matplotlib* παρέχει πολλές κλάσεις που επιτρέπουν την ενσωμάτωση γραφημάτων *Matplotlib* και χαρακτηριστικών σχεδίασης στο *Tkinter GUI*. Μία από αυτές τις κλάσεις είναι το *FigureCanvasTkAgg* που είναι ένας ειδικός καμβάς στον οποίο μπορούν να σχεδιαστούν γραφήματα *Matplotlib* και να εισαχθούν στο *Tkinter GUI* ως κανονικό γραφικό στοιχείο *Tkinter*.



Εικόνα 6 η εφαρμογή *Clustering App* με διαχειριστή γεωμετρίας *pack* και *To FigureCanvasTkAgg*.

Όταν δημιουργείται μια γραφική παράσταση ή μια εικόνα χρησιμοποιώντας το *Matplotlib*, υπάρχει στη μνήμη και πρέπει να αποδοθεί για εμφάνιση. Η κλάση *FigureCanvasTkAgg* παρέχει μια γέφυρα μεταξύ των αντικειμένων σχήματος του *Matplotlib* και του πλαισίου GUI του *Tkinter*, επιτρέποντάς σας να εμφανίζετε εικόνες *Matplotlib* σε ένα παράθυρο *Tkinter*.

Πώς λειτουργεί το `FigureCanvasTkAgg`:

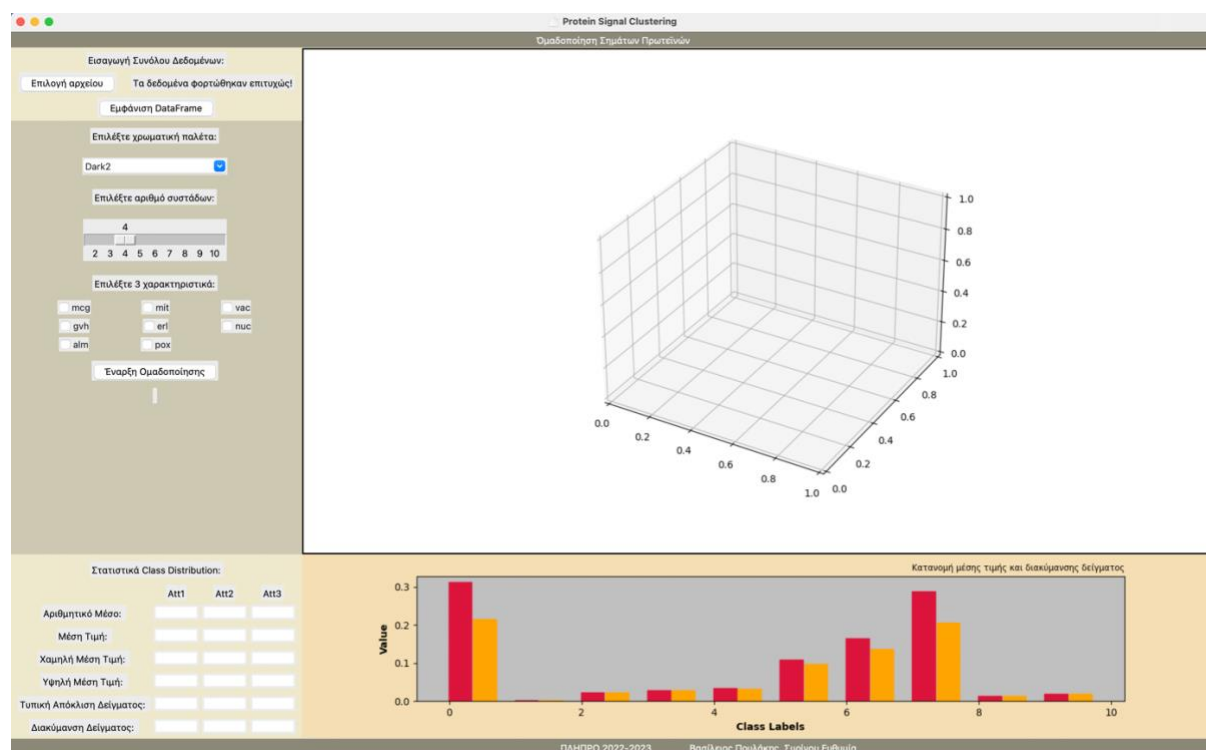
Δημιουργία σχήματος: Αρχικά, πρέπει να δημιουργούμε ένα figure *Matplotlib*. Αυτό γίνεται συνήθως χρησιμοποιώντας τη συνάρτηση `pyplot.subplots()`, η οποία επιστρέφει ένα αντικείμενο σχήματος και έναν ή περισσότερα αντικείμενα αξόνων.

Δημιουργία ενός στιγμιότυπου του `FigureCanvasTkAgg`: Αφού έχουμε δημιουργήσει ένα σχήμα, δημιουργούμε ένα στιγμιότυπο του `FigureCanvasTkAgg` περνώντας το αντικείμενο σχήμα και ένα κύριο αντικείμενο (παράθυρο) Tkinter ως ορίσματα. Αυτό δημιουργεί έναν καμβά που σχετίζεται με τη φιγούρα(*figure*).

Σχεδίαση του σχήματος: Αφού δημιουργήσουμε τον καμβά, καλούμε τη μέθοδο `draw()` στο αντικείμενο του καμβά. Αυτό θα αποδώσει τη φιγούρα στον καμβά, καθιστώντας την έτοιμη για προβολή.

Τοποθέτηση του καμβά στο παράθυρο Tkinter: Για να εμφανίσουμε το σχήμα, χρησιμοποιούμε τη μέθοδο `get_tk_widget()`. Αυτό το γραφικό στοιχείο μπορεί στη συνέχεια είτε με τον διαχειριστή *grid* είτε με τον *pack* να τοποθετηθεί στο παράθυρο του Tkinter όπως οποιοδήποτε άλλο γραφικό στοιχείο Tkinter.

Βρόχος συμβάντος Tkinter: Τέλος, με τη χρήση της μεθόδου `mainloop()` του κύριου αντικειμένου Tkinter το σχήμα Matplotlib εμφανίζεται και αλληλοεπιδρά με το χρήστη.



Εικόνα 7 Ενσωμάτωση της κλάσης `FigureCanvasTkAgg` στο GUI.

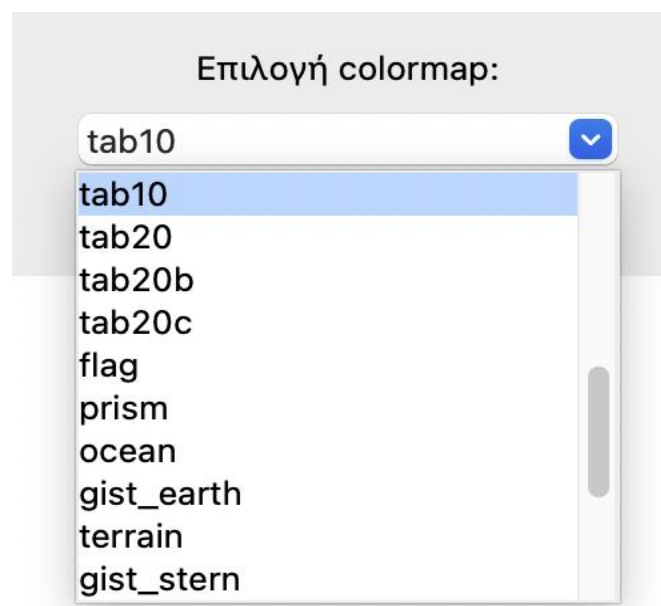
Ο Αλγόριθμος *Kmeans*

Πλέον και η μέθοδος του αλγόριθμου *Kmeans* τροποποιήθηκε στο να εκτελεί συσταδοποίηση κάνοντας *.fit* μόνο στα επιλεγμένα χαρακτηριστικά κάθε φορά που αυτά επιλέγονται από το χρήστη. (Σε αυτή τη φάση ο *Kmeans* χρησιμοποιήθηκε σαν μαύρο κουτί)

Αυτό που παραμετροποιήθηκε εκτός την επιλογή των χαρακτηριστικών ήταν ο αριθμός των συστάδων που επιλέγεται από το χρήστη καθώς και σε δεύτερη φάση η επιλογή της χρωματικής παλέτας(`cmpr`) απόδοσης των σχηματιζόμενων συστάδων.

Στο πλαίσιο της ομαδοποίησης *kmeans*, το `cmap` αντιπροσωπεύει μια παράμετρο που χρησιμοποιείται στην οπτικοποίηση δεδομένων για να καθορίσει τον χρωματικό χάρτη που αντιστοιχεί τα χρώματα διαφόρων ομάδων ή σημείων δεδομένων.

Υπάρχουν διάφοροι χρωματικοί χάρτες διαθέσιμοι στο *matplotlib*, και ο χρήστης μπορεί να επιλέξει αυτόν που ταιριάζει καλύτερα στις ανάγκες οπτικοποίησής.



Εικόνα 8Μενού επιλογής colormap χρωμάτων της βιβλιοθήκης *matplotlib*

Υπολογισμός στατιστικών

Ο υπολογισμός των στατιστικών που προβάλλονται στην εφαρμογή υλοποιήθηκε με τις αντίστοιχες μεθόδους της βιβλιοθήκης pandas('Mean', 'Median', 'Minimum', 'Maximum', 'Variance', 'Standard Deviation')

Εισαγωγή Dataset

Η περιγραφή του αρχικού project αφορούσε την εισαγωγή και επεξεργασία/συσταδοποίηση ενός dataset του yeast.data. Στην εφαρμογή 'γέφυρα' (LoadanyData) δίνεται η δυνατότητα επεξεργασίας και άλλων συνόλων δεδομένων, μεταξύ αυτών το iris.data καθώς και το

Εφαρμογή συσταδοποίησης με Kmeans										
CTG 2.xls										
DS	DP	winequality-red.csv								DP.1
.0	0.0	winequality-white.csv								0000
.0	0.0	HTRU_2.csv								0000
.0	0.0	ecoli.data								0000
.0	0.0	yeast.data								0000
.0	2.0	abalone.data								2099
.0	2.0	iris.data								2805
.0	0.0	iris.data								0000
.0	0.0	Data_Cortex_Nuclear.xls								0000
.0	0.0	BreastTissue.xls								0000
.0	0.0	BreastTissue.xls								0000
.0	0.0	CTG 2.xls								0000
.0	0.0	0.0	131.0	0.005076	0.072335	0.007614	0.002538	0.000000	0.000000	
.0	0.0	0.0	131.0	0.009077	0.222390	0.006051	0.001513	0.000000	0.000000	
.0	1.0	0.0	130.0	0.005838	0.407840	0.004170	0.005004	0.000000	0.000834	
.0	1.0	0.0	130.0	0.005571	0.380223	0.004178	0.004178	0.000000	0.001393	
.0	0.0	0.0	130.0	0.006088	0.441400	0.004566	0.004566	0.000000	0.000000	
.0	1.0	0.0	131.0	0.001524	0.382622	0.003049	0.004573	0.000000	0.001524	
.0	1.0	0.0	130.0	0.002845	0.450925	0.005690	0.004267	0.000000	0.001422	
.0	1.0	0.0	130.0	0.005055	0.160750	0.005055	0.001717	0.000000	0.000817	

Εικόνα 9Μενού επιλογής συνόλου δεδομένων της εφαρμογής LoadanyData

CTG 2 του project08.

Η επιλογή των παραπάνω συνόλων δεδομένων έγινε με βάση το μέγεθος τον αριθμό των χαρακτηριστικών ,την μορφή των χαρακτηριστικών(numerical) καθώς και τη διαθεσιμότητα.

Η εισαγωγή γίνεται με τις μεθόδους `pd.read` της βιβλιοθήκης `pandas` αλλά και την εισαγωγή των αρθρωμάτων `xlrd` και `openpyxl`(για αρχεία `xls,xlsx`).

Για την εύρεση/φόρτωση των αρχείων χρησιμοποιήθηκε το άρθρωμα `os` και πιο συγκεκριμένα το `os.path` που περιέχει μερικές χρήσιμες συναρτήσεις για τα `path` των αρχείων. Η μέθοδος `os.path.abspath()` επιστρέφει το ‘απόλυτο’ όνομα διαδρομής στη διαδρομή που έχει περάσει ως παράμετρος σε αυτήν τη συνάρτηση.

```
base_dir = os.path.dirname(os.path.abspath(__file__))
```

Προβολή Dataset

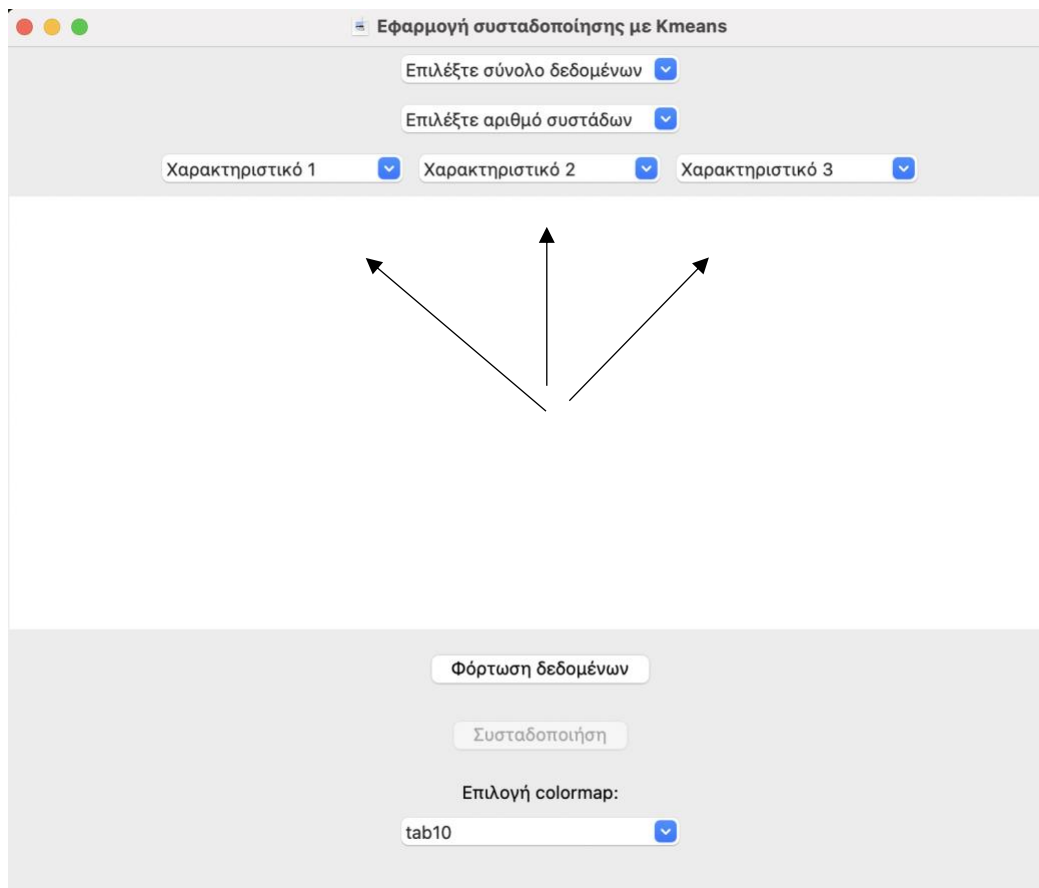
yeast.data										
	Sequence Name	mcg	gvh	alm	mit	erl	pox	vac	nuc	Class Distribution
0	ADT1_YEAST	0.58	0.61	0.47	0.13	0.5	0.00	0.48	0.22	MIT
1	ADT2_YEAST	0.43	0.67	0.48	0.27	0.5	0.00	0.53	0.22	MIT
2	ADT3_YEAST	0.64	0.62	0.49	0.15	0.5	0.00	0.53	0.22	MIT
3	AAR2_YEAST	0.58	0.44	0.57	0.13	0.5	0.00	0.54	0.22	NUC
4	AATM_YEAST	0.42	0.44	0.48	0.54	0.5	0.00	0.48	0.22	MIT
5	AATC_YEAST	0.51	0.40	0.56	0.17	0.5	0.50	0.49	0.22	CYT
6	ABC1_YEAST	0.50	0.54	0.48	0.65	0.5	0.00	0.53	0.22	MIT
7	BAF1_YEAST	0.48	0.45	0.59	0.20	0.5	0.00	0.58	0.34	NUC
8	ABF2_YEAST	0.55	0.50	0.66	0.36	0.5	0.00	0.49	0.22	MIT
9	ABP1_YEAST	0.40	0.39	0.60	0.15	0.5	0.00	0.58	0.30	CYT
10	ACE1_YEAST	0.43	0.39	0.54	0.21	0.5	0.00	0.53	0.27	NUC
11	ACE2_YEAST	0.42	0.37	0.59	0.20	0.5	0.00	0.52	0.29	NUC
12	ACH1_YEAST	0.40	0.42	0.57	0.35	0.5	0.00	0.53	0.25	CYT
13	ACON_YEAST	0.60	0.40	0.52	0.46	0.5	0.00	0.53	0.22	MIT
14	ACR1_YEAST	0.66	0.55	0.45	0.19	0.5	0.00	0.46	0.22	MIT
15	ACT_YEAST	0.46	0.44	0.52	0.11	0.5	0.00	0.50	0.22	CYT
16	ACT2_YEAST	0.47	0.39	0.50	0.11	0.5	0.00	0.49	0.40	CYT
17	ACT3_YEAST	0.58	0.47	0.54	0.11	0.5	0.00	0.51	0.26	NUC
18	ACT5_YEAST	0.50	0.34	0.55	0.21	0.5	0.00	0.49	0.22	NUC
19	ADA2_YEAST	0.61	0.60	0.55	0.21	0.5	0.00	0.50	0.25	NUC
20	C1TC_YEAST	0.45	0.40	0.50	0.16	0.5	0.00	0.50	0.22	CYT
21	PUR4_YEAST	0.43	0.44	0.48	0.22	0.5	0.00	0.51	0.22	CYT
22	PUR3_YEAST	0.73	0.63	0.42	0.30	0.5	0.00	0.49	0.22	CYT

Εικόνα 10 Προβολή του yeast dataset στην εφαρμογή LoadanyData.py

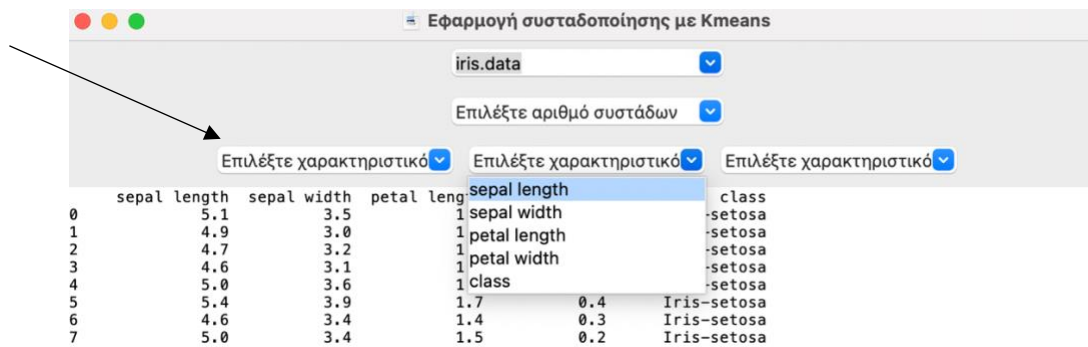
Στην εφαρμογή LoadanyData.py γίνεται χρήση του *text widget* της βιβλιοθήκης Tkinter όπου το dataset που επιλέγεται/φορτώνεται μετατρέπεται σε συμβολοσειρά χρησιμοποιώντας τη μέθοδο *to_string()*, και στη συνέχεια προβάλλεται στο πλαίσιο κειμένου *text widget* με τη μέθοδο *(self.text.insert(tk.END, data_str))*.

Επιλογή Χαρακτηριστικών(με Comboboxes)

Σε αυτήν τη βελτιωμένη έκδοση της εφαρμογής είναι ενδεικτική η ευκολία που επιλέγονται τα χαρακτηριστικά του εκάστοτε συνόλου δεδομένων για συσταδοποίηση, με τη χρήση των Comboboxes.



Εικόνα 11 Τα Comboboxes



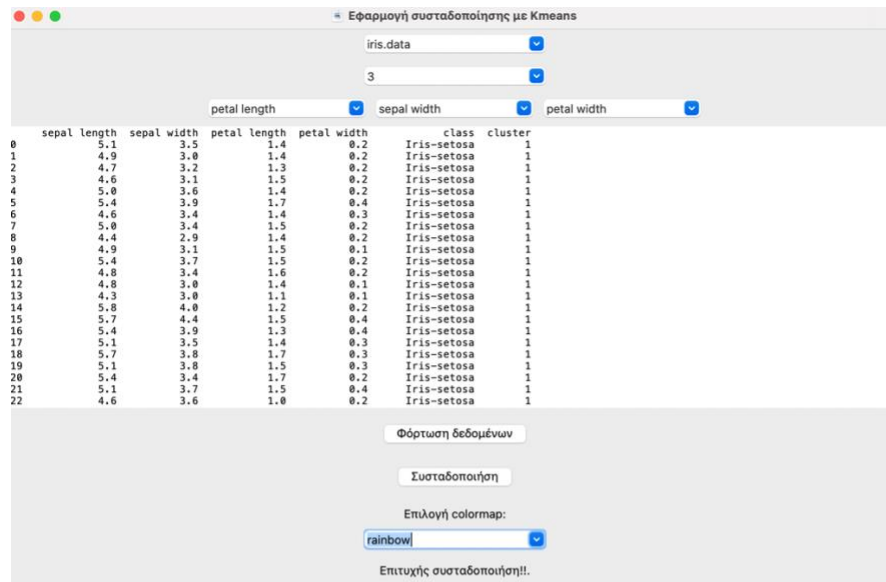
Εικόνα 12 Επιλογή χαρακτηριστικών με Comboboxes

Τα συγκεκριμένα **Comboboxes** δημιουργήθηκαν με μια επαναληπτική δομή for.

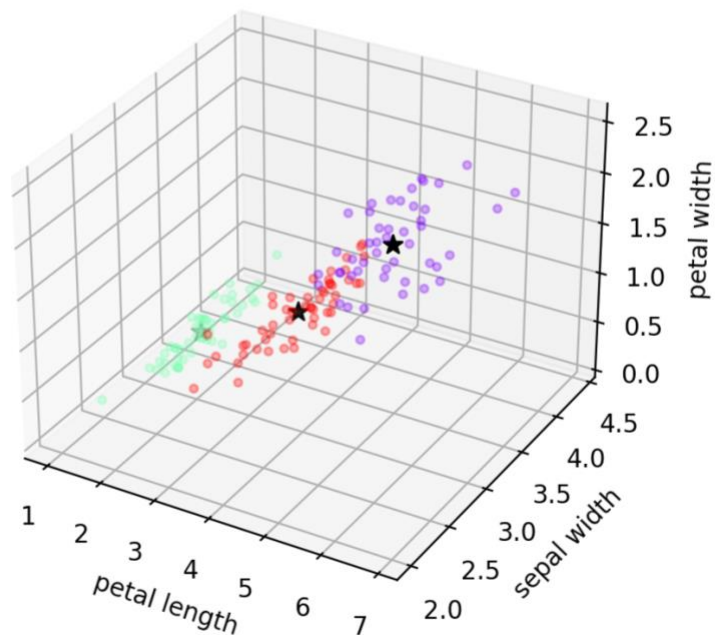
Στη συνέχεια όταν πλέον έχει φορτωθεί/επιλεγεί σύνολο δεδομένων με δεύτερη επαναληπτική δομή for, οι στήλες του συνόλου δεδομένων μετατρέπονται σε λίστα και κατόπιν γίνονται τα 'values' των *Comboboxes*.

Και τέλος με μία τρίτη επαναληπτική δομή for τα 3 επιλεγμένα χαρακτηριστικά αποθηκεύονται στη λίστα `selected_features` και τέλος με την βοήθεια ενός list comprehension γίνεται ένας έλεγχος έτσι ώστε στην επόμενη επανάληξη του βρόχου το χαρακτηριστικό που έχει ήδη επιλεγεί να μην είναι διαθέσιμο στα υπόλοιπα *Comboboxes*.

Screenshots

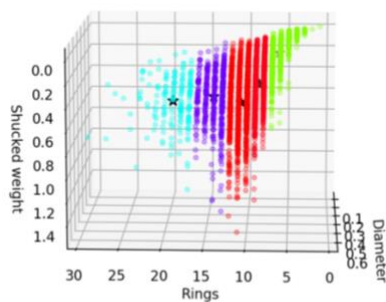


Αριθμός συστάδων:3



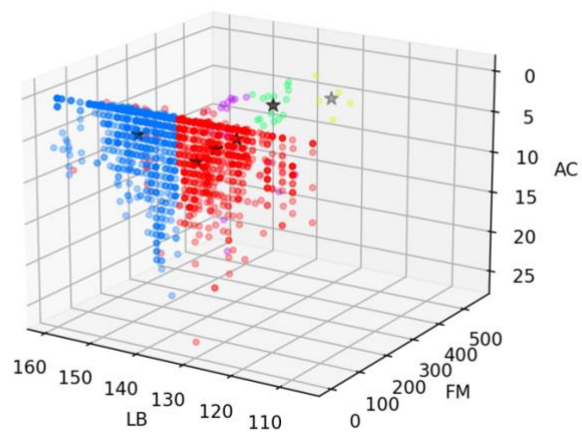
Εικόνα 13 Iris dataset

Αριθμός συστάδων:5



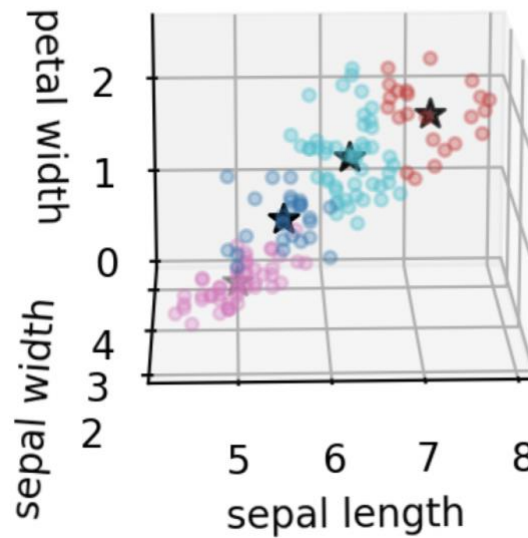
Εικόνα14 Abalone dataset

Αριθμός συστάδων:6



Εικόνα 15 CTG dataset

Αριθμός συστάδων:4



Εικόνα 16 Iris dataset

Χρόνος εκτέλεσης

Το προγραμματιστικό project εκπονήθηκε από το Φεβρουάριο του 2023 όπου και έγινε η ανάθεση μέχρι και σήμερα 14 Ιουνίου 2023. Υπολογίζεται ότι για την υλοποίηση του χρειάστηκαν γύρω στις 100 ώρες οι οποίες συμπεριλαμβάνουν τις συναντήσεις με την κ. Συρίγου κυρίως, προσωπική μελέτη και προγραμματισμό, άλλα και τις συναντήσεις καθοδήγησης με τον κ.Χαϊρή Κιούρτ.

	b	e	AC	FM	UC	DL	DS	DP	DR	LB	AC.1	FM.1	UC.1	DL.1	DS.1	DP.1	ASTV	MSTV	ALTV	MLTV	Min	cluster
0	240.0	357.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	120.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	73.0	0.5	43.0	2.4	62.0	2
1	5.0	632.0	4.0	0.0	4.0	2.0	0.0	0.0	0.0	132.0	0.005300	0.000000	0.005300	0.003190	0.000000	0.000000	17.0	2.1	0.0	10.4	68.0	0
2	177.0	779.0	2.0	0.0	5.0	2.0	0.0	0.0	0.0	133.0	0.003322	0.000000	0.003322	0.000000	0.000000	0.000000	16.0	2.1	0.0	13.4	68.0	2
3	411.0	1192.0	2.0	0.0	6.0	2.0	0.0	0.0	0.0	134.0	0.002561	0.000000	0.007682	0.002561	0.000000	0.000000	16.0	2.4	0.0	23.0	53.0	2
4	533.0	1147.0	4.0	0.0	5.0	0.0	0.0	0.0	0.0	132.0	0.006515	0.000000	0.008143	0.000000	0.000000	0.000000	16.0	2.4	0.0	19.9	53.0	0
5	0.0	953.0	1.0	0.0	10.0	9.0	0.0	2.0	0.0	134.0	0.001049	0.000000	0.010493	0.009444	0.000000	0.002099	26.0	5.9	0.0	0.0	50.0	1
6	240.0	953.0	1.0	0.0	9.0	6.0	0.0	2.0	0.0	134.0	0.001403	0.000000	0.012623	0.008415	0.000000	0.002805	29.0	6.3	0.0	0.0	50.0	1
7	62.0	679.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	122.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	83.0	0.5	6.0	15.6	62.0	2
8	120.0	779.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	122.0	0.000000	0.000000	0.001517	0.000000	0.000000	0.000000	84.0	0.5	5.0	13.6	62.0	2
9	181.0	1192.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	122.0	0.000000	0.000000	0.002967	0.000000	0.000000	0.000000	86.0	0.3	6.0	10.6	62.0	2
10	0.0	1199.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	151.0	0.000000	0.000000	0.000334	0.000334	0.000000	0.000000	64.0	1.9	9.0	27.6	56.0	2
11	57.0	1074.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	150.0	0.000000	0.000000	0.000983	0.000983	0.000000	0.000000	64.0	2.0	8.0	29.5	56.0	2
12	52.0	840.0	4.0	57.0	6.0	2.0	0.0	0.0	0.0	131.0	0.005076	0.072335	0.007614	0.002538	0.000000	0.000000	28.0	1.4	0.0	12.9	88.0	0
13	531.0	1192.0	6.0	147.0	4.0	1.0	0.0	0.0	0.0	131.0	0.009077	0.222390	0.006051	0.001513	0.000000	0.000000	28.0	1.5	0.0	5.4	71.0	0
14	0.0	1199.0	7.0	489.0	5.0	6.0	0.0	1.0	0.0	130.0	0.005838	0.407840	0.004170	0.005004	0.000000	0.000034	21.0	2.3	0.0	7.9	67.0	0
15	0.0	718.0	4.0	273.0	3.0	3.0	0.0	1.0	0.0	130.0	0.005571	0.380223	0.004178	0.004178	0.000000	0.001393	19.0	2.3	0.0	8.7	67.0	0
16	537.0	1194.0	4.0	290.0	3.0	3.0	0.0	0.0	0.0	130.0	0.006088	0.441400	0.004566	0.004566	0.000000	0.000000	24.0	2.1	0.0	10.9	53.0	0
17	1.0	657.0	1.0	251.0	2.0	3.0	0.0	1.0	0.0	131.0	0.001524	0.382622	0.003049	0.004573	0.000000	0.001524	18.0	2.4	0.0	13.9	67.0	2
18	0.0	711.0	2.0	317.0	4.0	3.0	0.0	1.0	0.0	130.0	0.002845	0.450925	0.005690	0.004267	0.000000	0.001422	23.0	1.9	0.0	8.8	59.0	2
19	7.0	1194.0	6.0	557.0	6.0	5.0	0.0	1.0	0.0	130.0	0.005055	0.469250	0.005055	0.004212	0.000000	0.000042	29.0	1.7	0.0	7.8	65.0	0
20	298.0	1192.0	0.0	304.0	4.0	2.0	0.0	3.0	0.0	129.0	0.000000	0.340045	0.004474	0.002237	0.000000	0.003356	30.0	2.1	0.0	8.5	54.0	2
21	3.0	643.0	3.0	272.0	2.0	2.0	0.0	1.0	0.0	128.0	0.004607	0.425000	0.003125	0.003125	0.000000	0.001563	26.0	1.7	0.0	6.7	57.0	0
22	538.0	1194.0	0.0	219.0	2.0	2.0	0.0	2.0	0.0	128.0	0.000000	0.333841	0.003049	0.003049	0.000000	0.003049	34.0	2.5	0.0	4.0	54.0	2
23	910.0	664.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	125.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.5	0.0	6.8	114.0	2

Φόρτωση δεδομένων

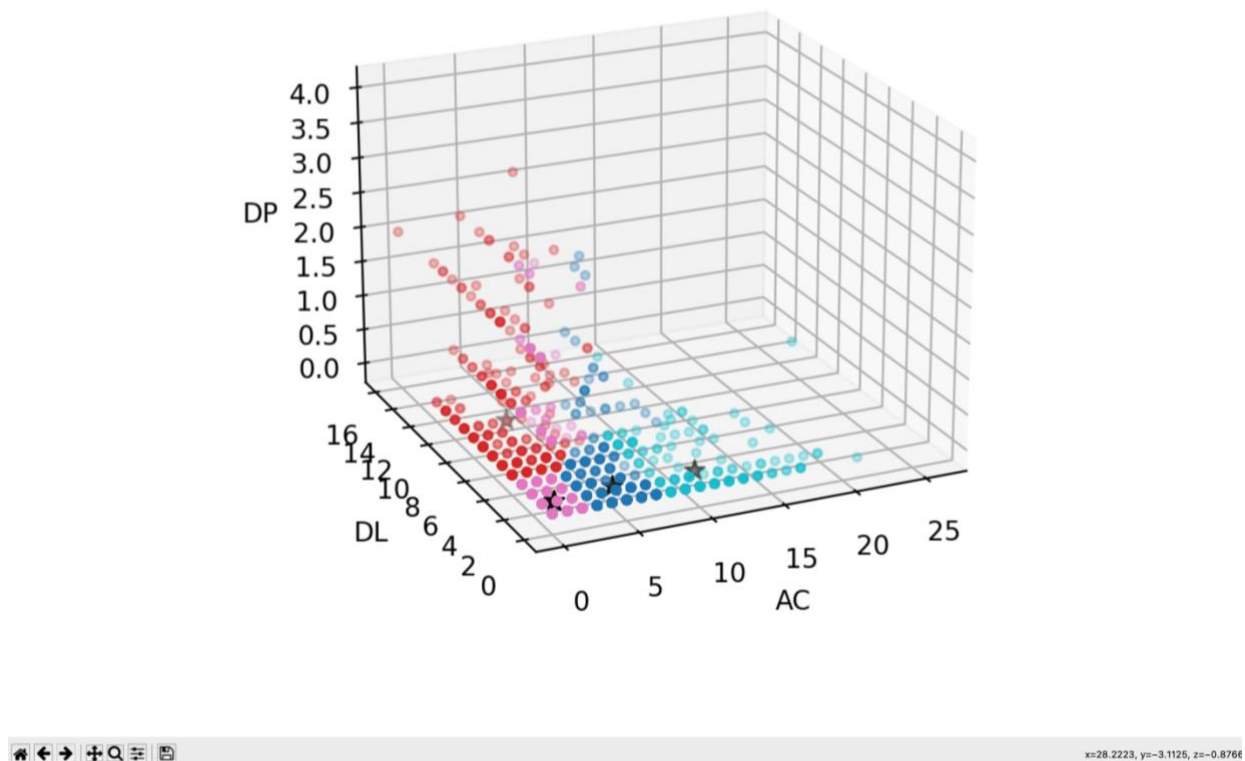
Συσταδοποίηση

Επιλογή colormap:

tab10

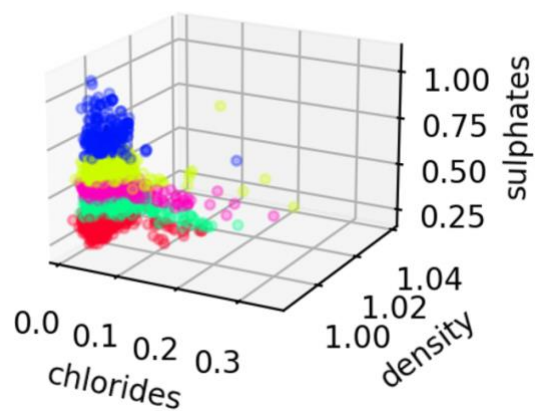
Επιτυχής συσταδοποίηση!!

Αριθμός συστάδων:4

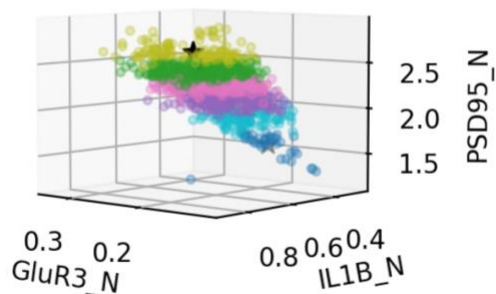


Εικόνα 17 CTG 2 dataset

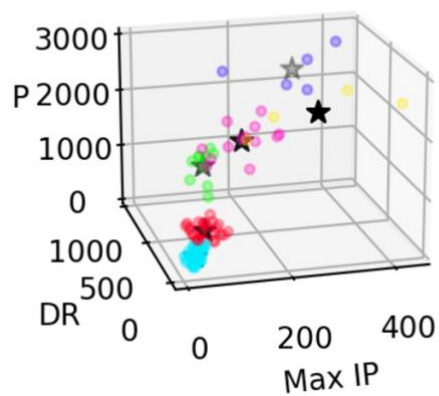
Αριθμός συστάδων:5



Αριθμός συστάδων:6

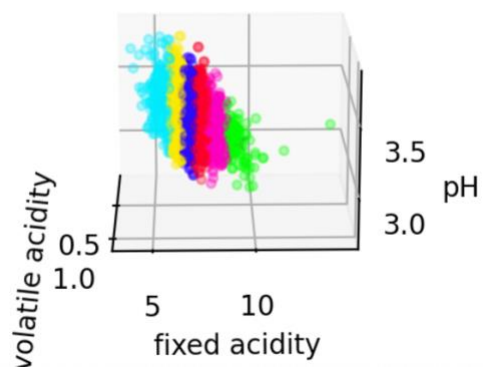


Αριθμός συστάδων:6

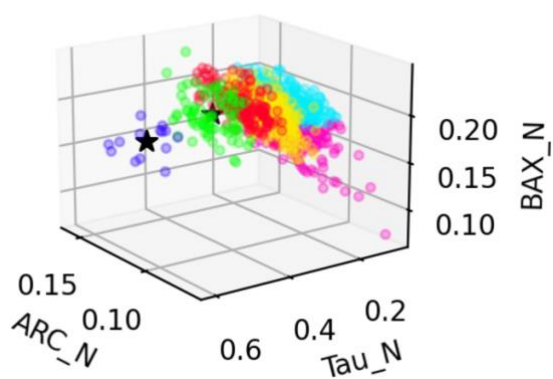


Εικόνα 18 ομαδοποιήσεις με την εφαρμογή LoadanyDataset

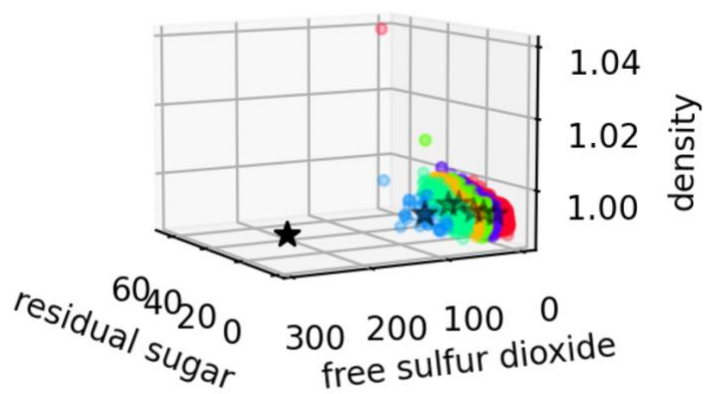
Αριθμός συστάδων:6



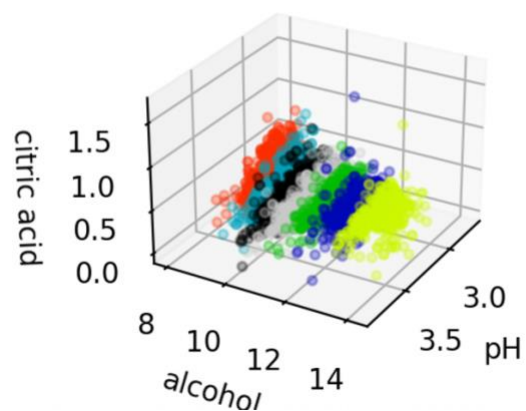
Αριθμός συστάδων:6



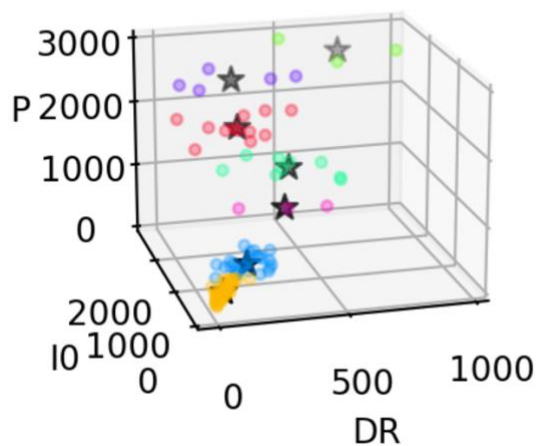
Αριθμός συστάδων:7



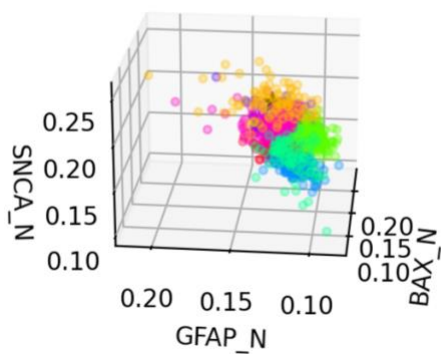
Αριθμός συστάδων:7



Αριθμός συστάδων:7



Αριθμός συστάδων:7



Βιβλιογραφία

- Αβούρης , Ν., Παλιουράς, Β., Καζιρας, Σ., Κουκιάς, Μ. (2012). Εισαγωγή στους Υπολογιστές με την Python Εκδόσεις Πανεπιστημίου Πατρών Πάτρα.*
- Βακαλή , Α., Γιαννόπουλος, Η., Ιωαννίδης, Χ., Κοιλιάς, Χ. Μάλαμας , Κ (2012). Ανάπτυξη εφαρμογών σε προγραμματιστικό περιβάλλον.*
- Raschka, S. (2015). Python Machine Learning. United Kingdom: Packt Publishing.*
- Moore, A. D. (2018). Python GUI Programming with Tkinter: Develop Responsive and Powerful GUI Applications with Tkinter.*
- Grus, J. (2015). Data Science from Scratch: First Principles with Python. Japan: O'Reilly.*
- Jones, B. K., Beazley, D. (2013). Python Cookbook: Recipes for Mastering Python 3. United States: O'Reilly Media.*
- Rossant, C. (2015). Learning IPython for Interactive Computing and Data Visualization. United Kingdom: Packt Publishing*
- Markelle Kelly, Rachel Longjohn, Kolby Nottingham,*
The UCI Machine Learning Repository, <https://archive.ics.uci.edu>

