ΔΗΜΙΟΥΡΓΙΑ ΜΟΝΤΕΛΟΥ ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΠΡΟΦΙΛ ΑΣΘΕΝΩΝ ΤΟΥ COVID-19 ΜΕ ΧΡΗΣΗ ΤΕΧΝΙΚΩΝ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ

Άγγελος Δαντσούλης, mai19012

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΣΤΟΧΟΣ

Πρόβλεψη της μελλοντικής εξέλιξης της νόσου

όσων έχουν προσβληθεί από τον ιό COVID-19, ανάλογα με τα χαρακτηριστικά τους.

Οι αλγόριθμοι συσταδοποίησης χωρίζουν το δείγμα σε υποσύνολα

Οι αλγόριθμοι κατηγοριοποίησης προσπαθούν να προβλέψουν την εξέλιξη της νόσου

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ
K-Means

Πλεονεκτήματα

- Είναι σχετικά απλός στην εφαρμογή
- Έχει κλίμακα σε μεγάλα σύνολα δεδομένων
- Εγγυάται τη σύγκλιση
- Μπορεί να ξεκινήσει τις θέσεις των κεντροειδών
- Προσαρμόζεται εύκολα σε νέα παραδείγματα
- Γενικεύεται σε ομάδες διαφορετικών σχημάτων και μεγεθών, όπως ελλειπτικά σμήνη

Μειονεκτήματα

- Επιλέγει χειροκίνητα του αριθμού των συστάδων
- Εξαρτάται από τις αρχικές τιμές
- Έχει συγκεντρωτικά δεδομένα διαφόρων μεγεθών και πυκνότητας
- Εκτελεί ομαδοποίηση ακραίων τιμών
- Έχει κλίμακα με αριθμό διαστάσεων

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ
Mean-Shift

Πλεονεκτήματα

- Δεν χρειάζεται να κάνει οποιαδήποτε υπόθεση μοντέλου
- Μπορεί να μοντελοποιήσει τις σύνθετες συστάδες που έχουν μη κυρτό σχήμα
- Χρειάζεται μόνο μία παράμετρο που ονομάζεται εύρος ζώνης
- Δεν υπάρχει ζήτημα τοπικών ελαχίστων όπως στον αλγόριθμο K-means
- Δεν δημιουργεί πρόβλημα από τα ακραία σημεία

Μειονεκτήματα

- Δεν λειτουργεί καλά σε περίπτωση υψηλής διάστασης
- Δεν έχουμε άμεσο έλεγχο του αριθμού των συστάδων
- Δεν μπορεί να κάνει διάκριση μεταξύ ουσιαστικών και χωρίς νόημα χαρακτηριστικών

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ DBSCAN

Πλεονεκτήματα

- Δεν απαιτεί από κάποιον να καθορίσει τον αριθμό συστάδων στα δεδομένα εκ των προτέρων
- Μπορεί να βρει συστάδες αυθαίρετου σχήματος
- Έχει την έννοια του θορύβου και είναι ανθεκτικός στα ακραία σημεία
- Απαιτεί μόνο δύο παραμέτρους και ως επί το πλείστον δεν είναι ευαίσθητος στη σειρά των σημείων
- Έχει σχεδιαστεί για χρήση με βάσεις δεδομένων που μπορούν να επιταχύνουν τα ερωτήματα περιοχής
- Οι παράμετροι min_samples και eps μπορούν να οριστούν από έναν ειδικό τομέα

Μειονεκτήματα

- Δεν είναι απολύτως ντετερμινιστικός: τα συνοριακά σημεία που είναι προσβάσιμα από περισσότερα από μία συστάδα μπορούν να είναι μέρος οποιουδήποτε συστάδας.
- Η ποιότητα του αλγόριθμου εξαρτάται από το μέτρο απόστασης που χρησιμοποιείται
- Δεν μπορεί να συγκεντρώσει σύνολα δεδομένων καλά με μεγάλες διαφορές στην πυκνότητα

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ
GMM using EM

Πλεονεκτήματα

- Είναι ο ταχύτερος αλγόριθμος για την εκμάθηση μοντέλων μείγματος
- Δεδομένου ότι μεγιστοποιεί μόνο την πιθανότητα, δεν θα προκαταλάβει τον μέσο όρο προς το μηδέν ή θα προκαταλάβει τα μεγέθη των συστάδων ώστε να έχουν συγκεκριμένες δομές που ενδέχεται να ισχύουν ή να μην ισχύουν

Μειονεκτήματα

- Όταν κάποιος έχει ανεπαρκώς πολλά σημεία ανά μείγμα, η εκτίμηση των πινάκων συνδιακύμανσης καθίσταται δύσκολη και ο αλγόριθμος είναι γνωστό ότι αποκλίνει και βρίσκει λύσεις με απεριόριστη πιθανότητα εκτός αν κάποιος κανονικοποιήσει τεχνητά τις συνδιακυμάνσεις
- Θα χρησιμοποιεί πάντα όλα τα στοιχεία στα οποία έχει πρόσβαση και θα χρησιμοποιεί δεδομένα που έχουν παραμείνει ή θεωρητικά κριτήρια για να αποφασίσει πόσα στοιχεία θα χρησιμοποιηθούν ελλείψει άλλων ενδείξεων

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ
Agglomerative

Πλεονεκτήματα

- Δεν απαιτείται να καθορίσουμε τον αριθμό των συστάδων
- Είναι εύκολος στην εφαρμογή και δίνει το καλύτερο αποτέλεσμα σε ορισμένες περιπτώσεις

Μειονεκτήματα

- Δεν μπορεί ποτέ να αναιρέσει αυτό που έγινε προηγουμένως
- Απαιτείται χρονική πολυπλοκότητα τουλάχιστον O (n^2 log n), όπου «n» είναι ο αριθμός των σημείων δεδομένων
- Μπορεί να έχει ευαισθησία σε θόρυβο και ακραίες τιμές
- Μπορεί να έχει μεγάλες συστάδες
- Έχει δυσκολία χειρισμού διαφορετικών μεγεθών συστάδων και κυρτών σχημάτων
- Καμία αντικειμενική λειτουργία δεν ελαχιστοποιείται άμεσα
- Μερικές φορές είναι δύσκολο να προσδιοριστεί ο σωστός αριθμός συστάδων από το δενδρόγραμμα

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΛΓΟΡΙΘΜΟΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ
Logistic Regression

Πλεονεκτήματα

- Είναι ευκολότερος στην εφαρμογή, την ερμηνεία και την πολύ αποτελεσματική εκπαίδευση
- Δεν κάνει παραδοχές σχετικά με τις κατανομές των κλάσεων στο χώρο χαρακτηριστικών.
- Μπορεί εύκολα να επεκταθεί σε πολλές κλάσεις
- Δεν παρέχει μόνο ένα μέτρο για το πόσο κατάλληλος είναι ένας προγνωστικός παράγοντας (μέγεθος συντελεστή), αλλά και η κατεύθυνση συσχέτισης (θετική ή αρνητική)
- Είναι πολύ γρήγορη στην ταξινόμηση άγνωστων αρχείων
- Έχει καλή ακρίβεια για πολλά απλά σύνολα δεδομένων
- Μπορεί να ερμηνεύσει τους συντελεστές μοντέλου ως δείκτες σπουδαιότητας χαρακτηριστικών

Μειονεκτήματα

- Εάν ο αριθμός των παρατηρήσεων είναι μικρότερος από τον αριθμό των χαρακτηριστικών, δεν πρέπει να χρησιμοποιείται
- Κατασκευάζει γραμμικά όρια
- Η υπόθεση γραμμικότητας μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών
- Μπορεί να χρησιμοποιηθεί μόνο για την πρόβλεψη διακριτών λειτουργιών
- Απαιτεί μέση ή καθόλου πολυγραμμικότητα μεταξύ ανεξάρτητων μεταβλητών

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΛΓΟΡΙΘΜΟΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ
Stochastic Gradient Descent

Πλεονεκτήματα

- Αποδοτικότητα
- Ευκολία εφαρμογής (πολλές ευκαιρίες για ρύθμιση κώδικα)

Μειονεκτήματα

- Απαιτεί έναν αριθμό υπερπαραμέτρων, όπως την παράμετρο κανονικοποίησης και τον αριθμό των επαναλήψεων
- Είναι ευαίσθητος στην κλιμάκωση χαρακτηριστικών

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΛΓΟΡΙΘΜΟΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ
Decision Tree

Πλεονεκτήματα

- Είναι σε θέση να δημιουργήσουν κατανοητούς κανόνες
- Εκτελούν κατηγοριοποίηση χωρίς να απαιτούν πολύ υπολογισμό
- Είναι σε θέση να χειρίζονται συνεχείς και κατηγορηματικές μεταβλητές
- Παρέχουν μια σαφή ένδειξη ποια πεδία είναι πιο σημαντικά για την πρόβλεψη

Μειονεκτήματα

- Είναι λιγότερο κατάλληλα για εργασίες εκτίμησης όπου ο στόχος είναι να προβλεφθεί η αξία ενός συνεχούς χαρακτηριστικού
- Είναι επιρρεπή σε σφάλματα σε προβλήματα κατηγοριοποίηση με πολλές τάξεις και σχετικά μικρό αριθμό δεδομένων εκπαίδευσης
- Μπορεί να έχει μεγάλο υπολογιστικό κόστος για την εκπαίδευση των δεδομένων

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

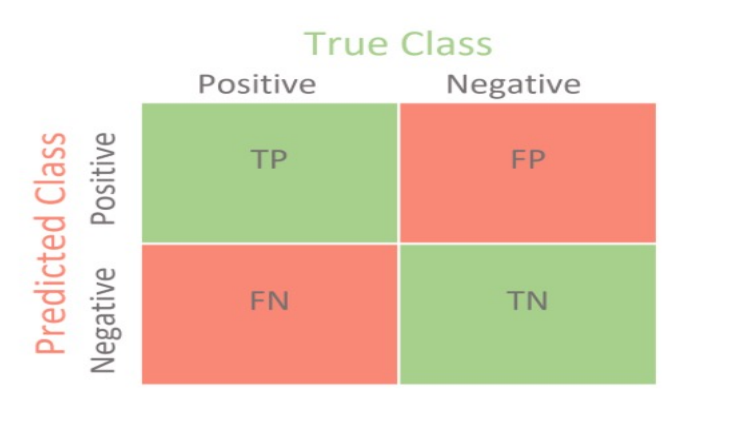
ΜΕΤΡΙΚΕΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ

Silhouette

ΜΟ Απόστασης από το Κεντροειδές κάθε Συστάδας

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΜΕΤΡΙΚΕΣ ΚΑΤΗΓΟΡΟΠΟΙΗΣΗΣ



TP = Αληθινά Θετικά

TN = Αληθινά Αρνητικά

FP = Ψευδώς Θετικά

FN = Ψευδώς Αρνητικά

Ευστοχία (Accuracy)

Accuracy = (TP+TN) / (TP+FP+FN+TN)

Ακρίβεια (Precision)

Precision = TP / (TP+FP)

Ανάκληση (Recall)

Recall = TP / (TP+FN)

F1_score

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

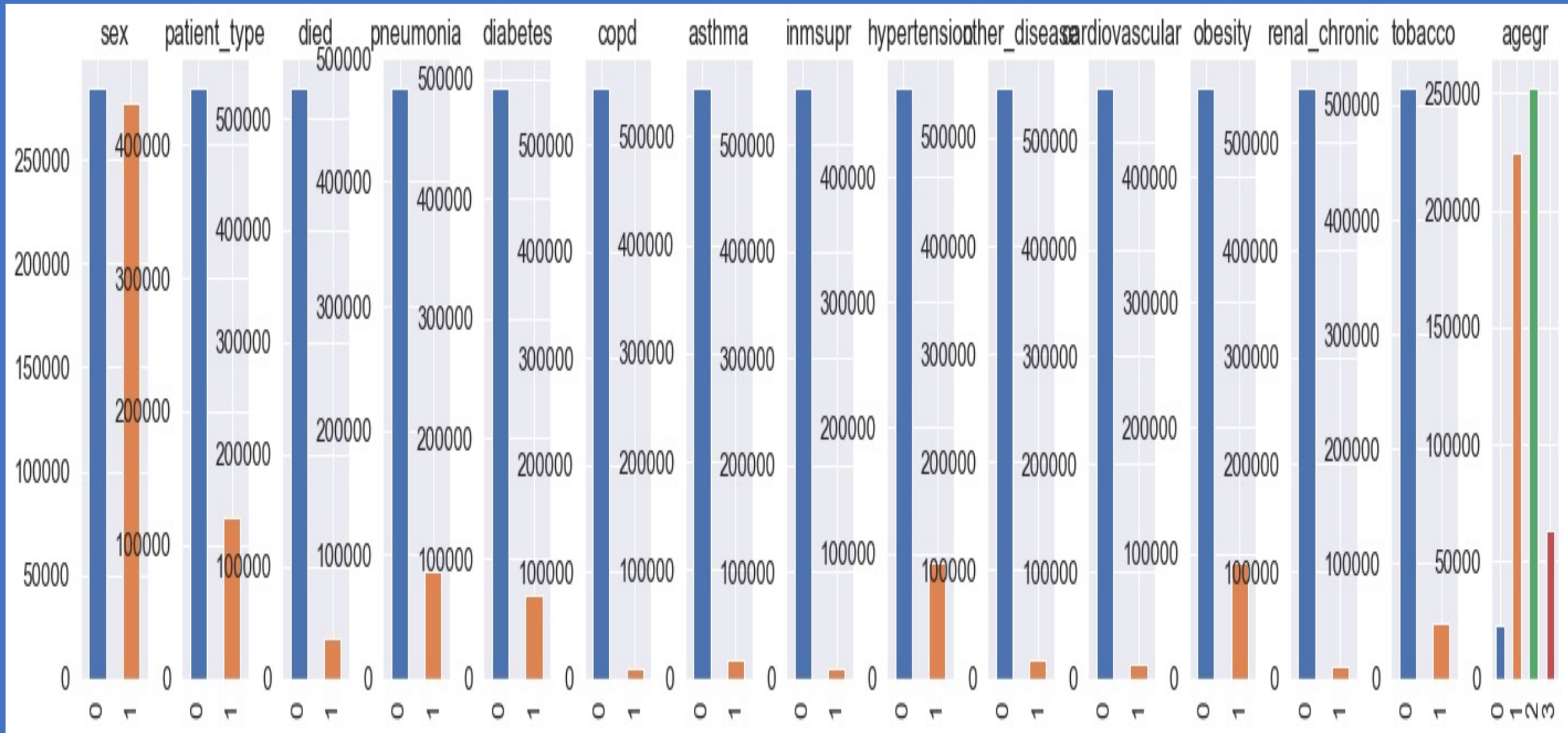
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΜΕΤΑΒΛΗΤΕΣ

Μεταβλητή	Επεξήγηση	ΤΙΜΕΣ			
		0	1	2	3
sex	φύλο	άντρας	γυναικά		
patient_type	Εντός νοσοκομείου ασθενής	ΟΧΙ	ΝΑΙ		
died	αν απεβίωσε	ΟΧΙ	ΝΑΙ		
pneumonia	πνευμονία	ΟΧΙ	ΝΑΙ		
diabetes	διαβήτης	ΟΧΙ	ΝΑΙ		
copd	ΧΑΠ	ΟΧΙ	ΝΑΙ		
asthma	άσθμα	ΟΧΙ	ΝΑΙ		
inmsupr	ανοσοκασταλόμενος	ΟΧΙ	ΝΑΙ		
hypertension	υπέρταση	ΟΧΙ	ΝΑΙ		
other_disease	άλλη νόσος	ΟΧΙ	ΝΑΙ		
cardiovascular	καρδιαγγειακό	ΟΧΙ	ΝΑΙ		
obesity	παχυσαρκία	ΟΧΙ	ΝΑΙ		
renal_chronic	χρόνια νεφροπάθεια	ΟΧΙ	ΝΑΙ		
tobacco	κάπνισμα	ΟΧΙ	ΝΑΙ		
agegr	ηλικία	0-17	18-39	40-64	65-

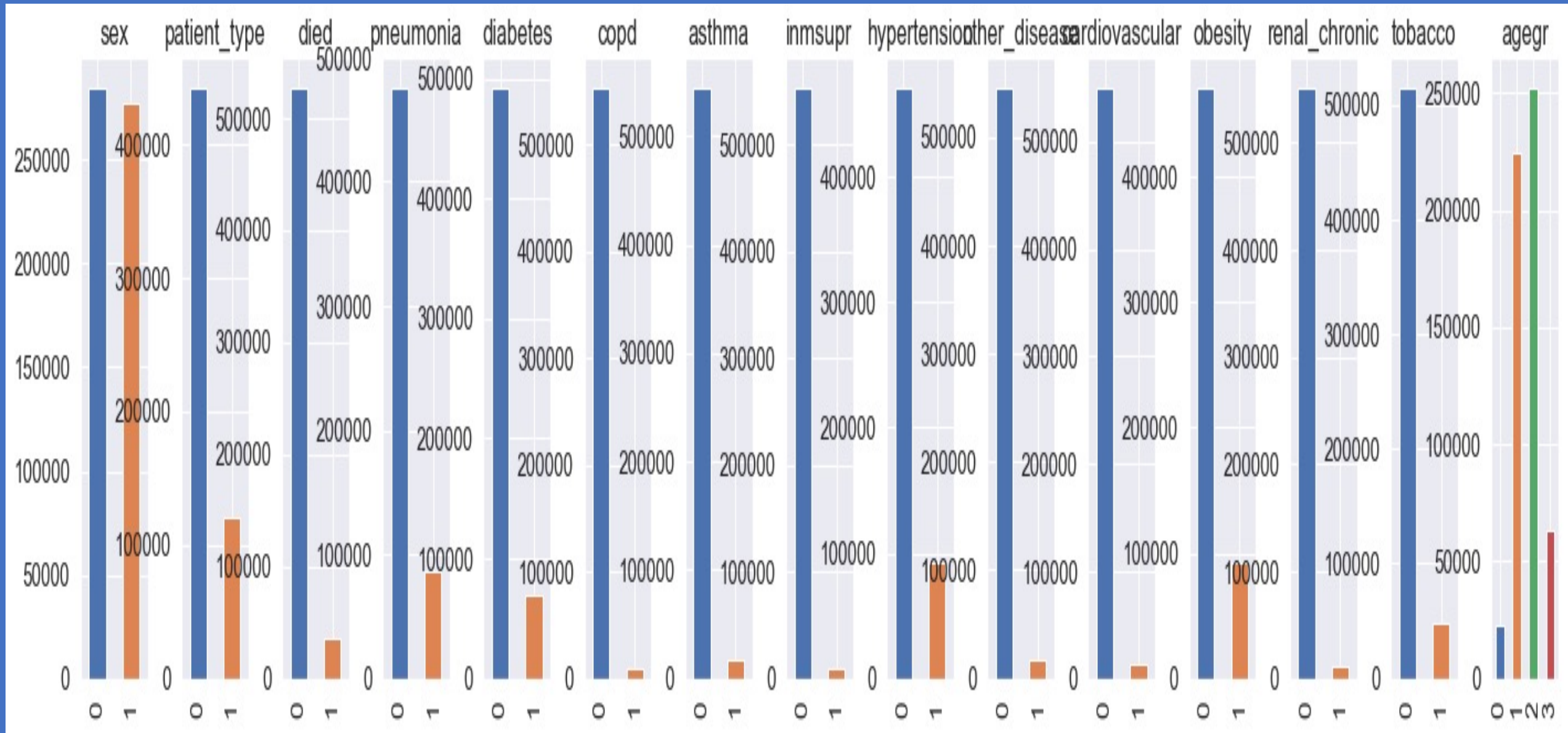
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΑΡΧΙΚΑ ΔΕΔΟΜΕΝΑ



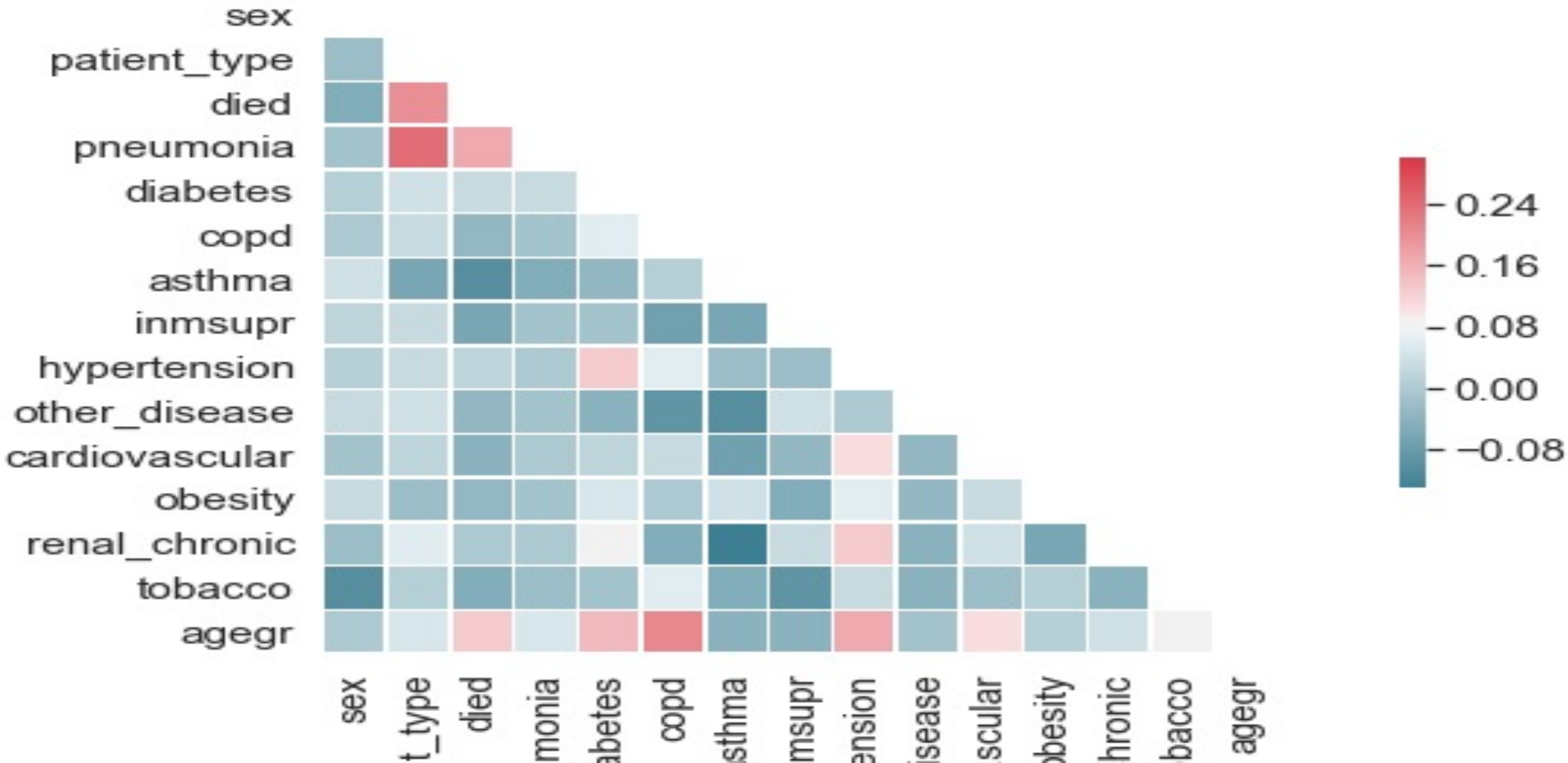
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΔΕΔΟΜΕΝΑ
ΜΕΤΑ ΤΗΝ ΑΦΑΙΡΕΣΗ ΔΙΠΛΟΤΥΠΩΝ



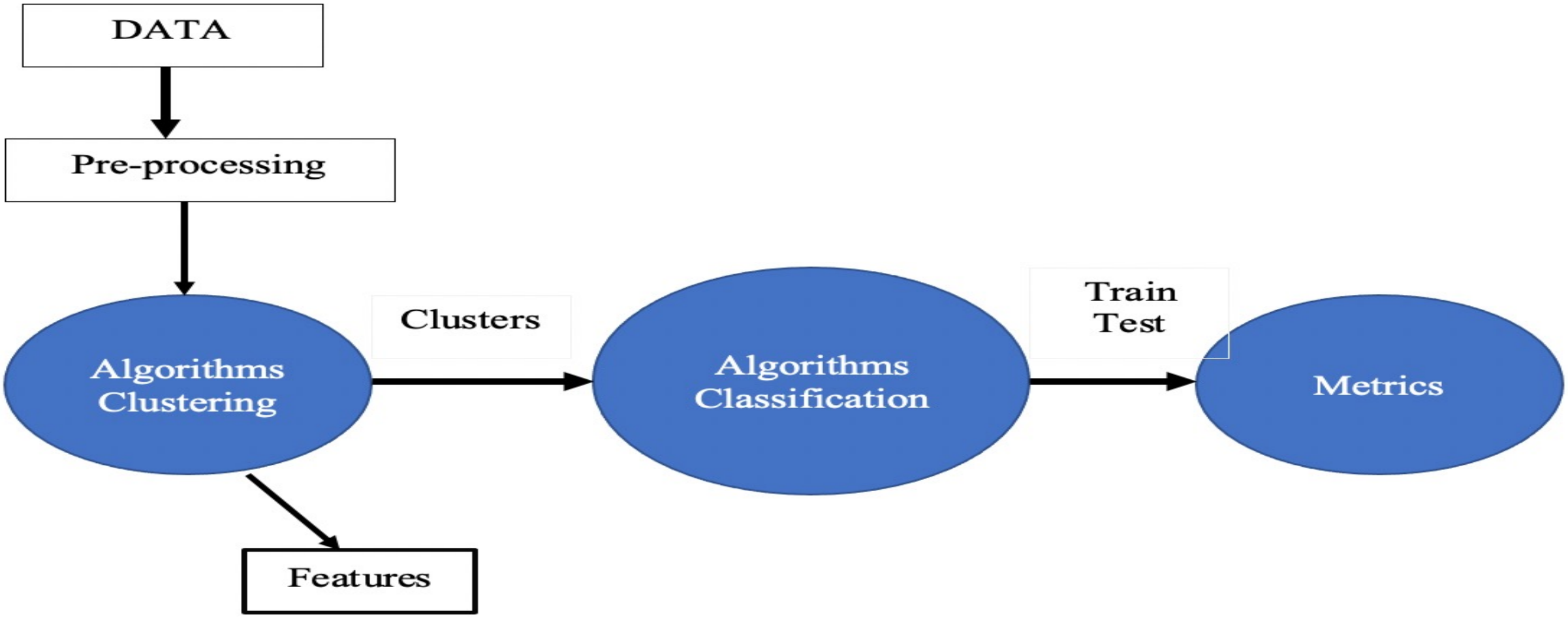
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΔΙΑΓΡΑΜΜΑ ΣΥΣΧΕΤΙΣΕΩΝ



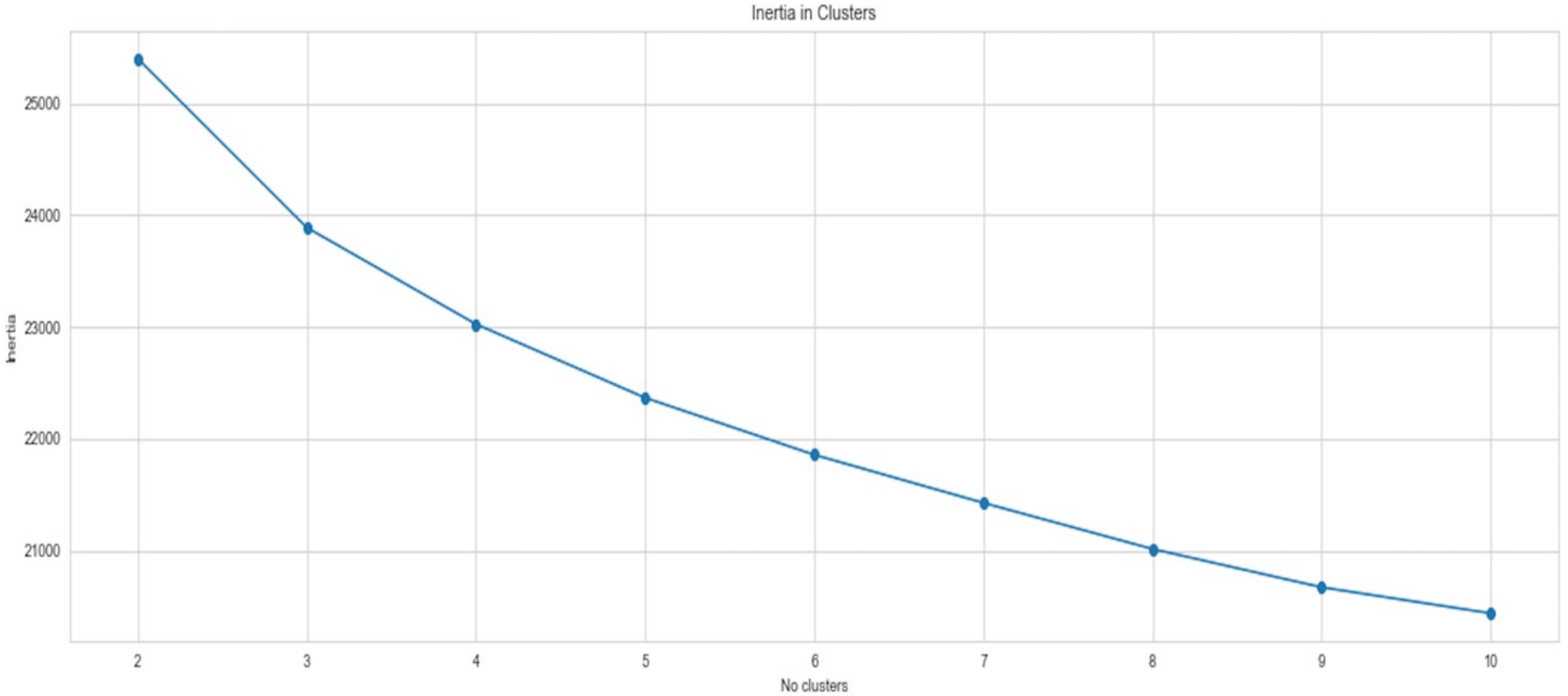
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΜΕΘΟΔΟΛΟΓΙΑ



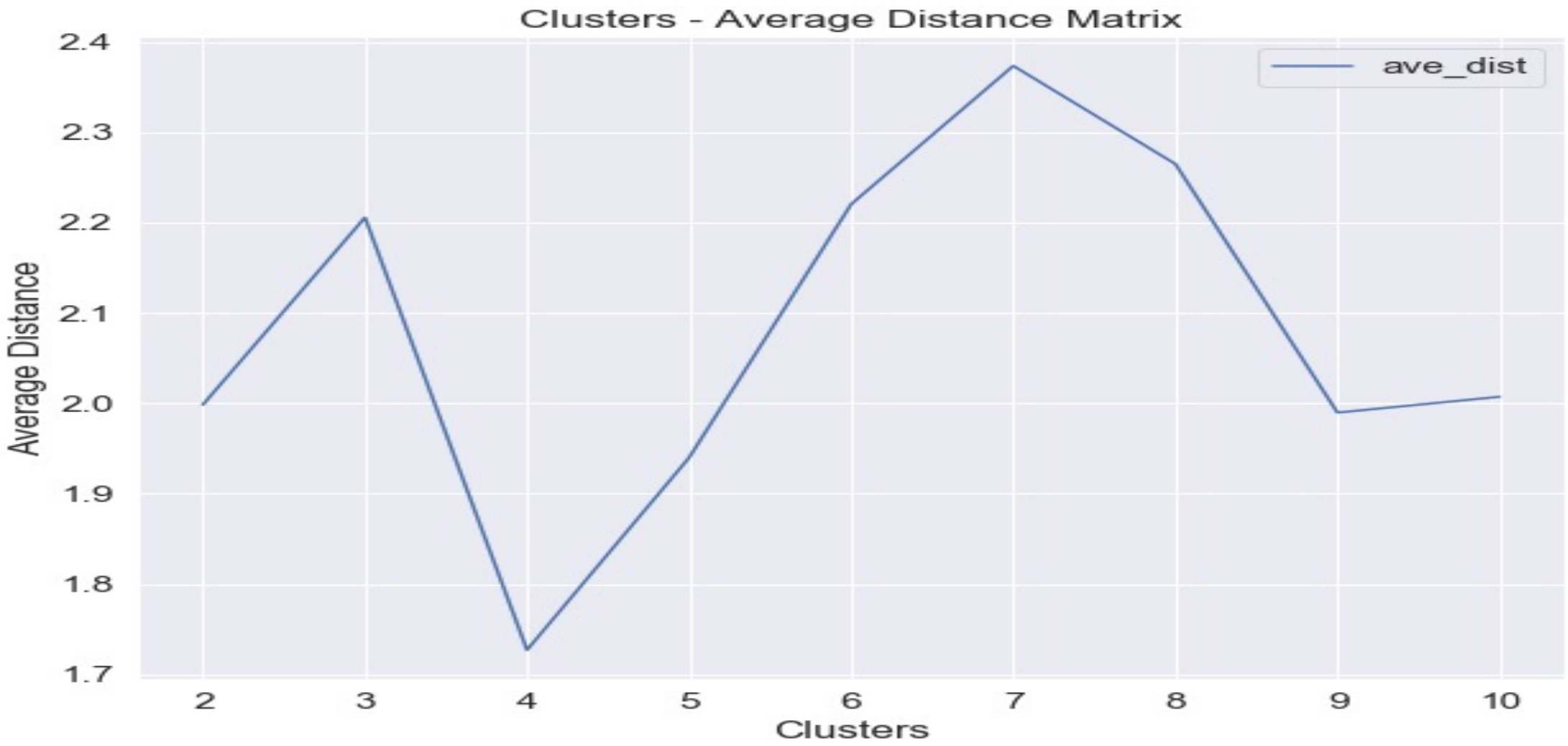
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

K-Means
Inertia



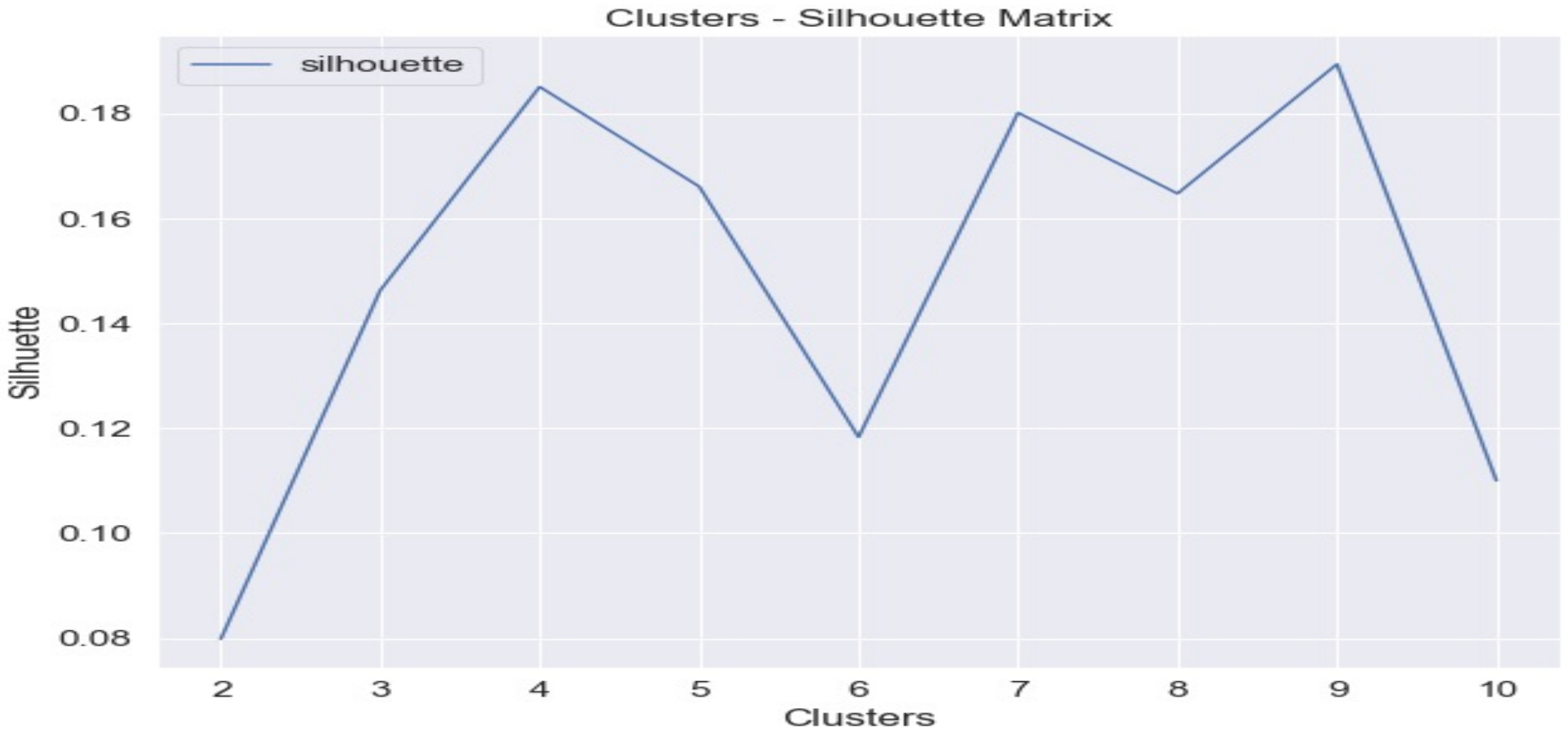
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

K-Means
ΜΟ Απόστασης



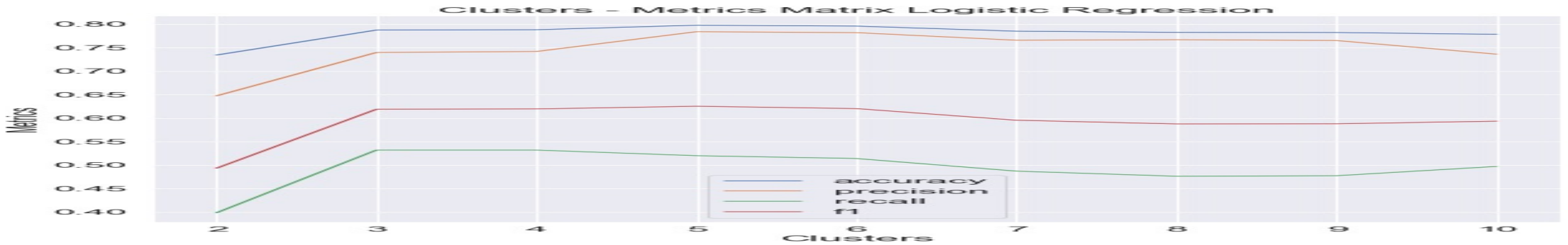
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

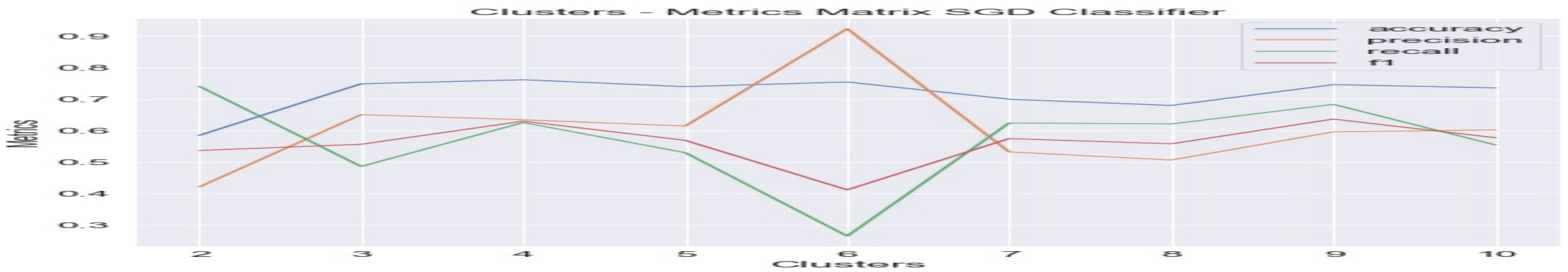
K-Means
Siihouette

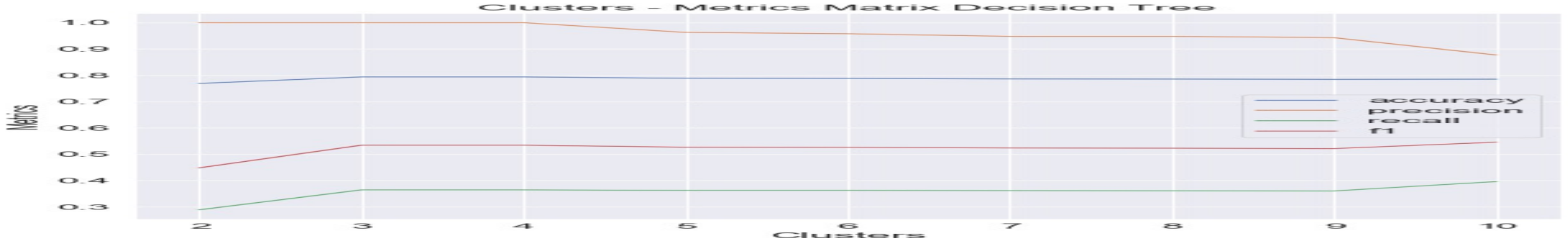


ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

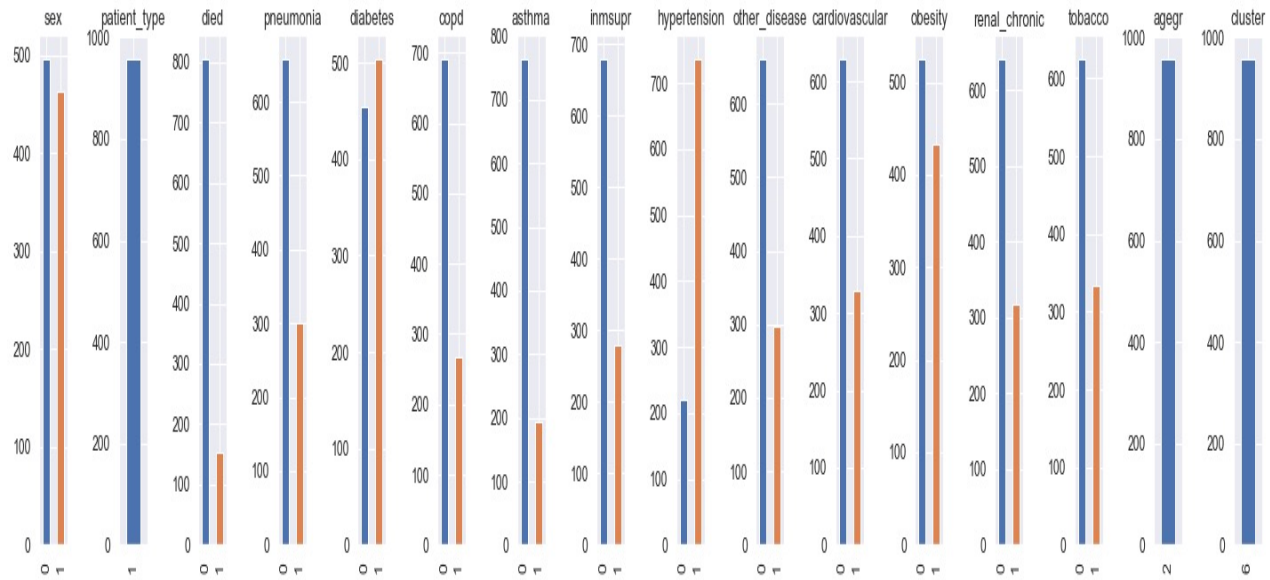
K-Means
Αλγόριθμοι Κατηγοριοποίησης

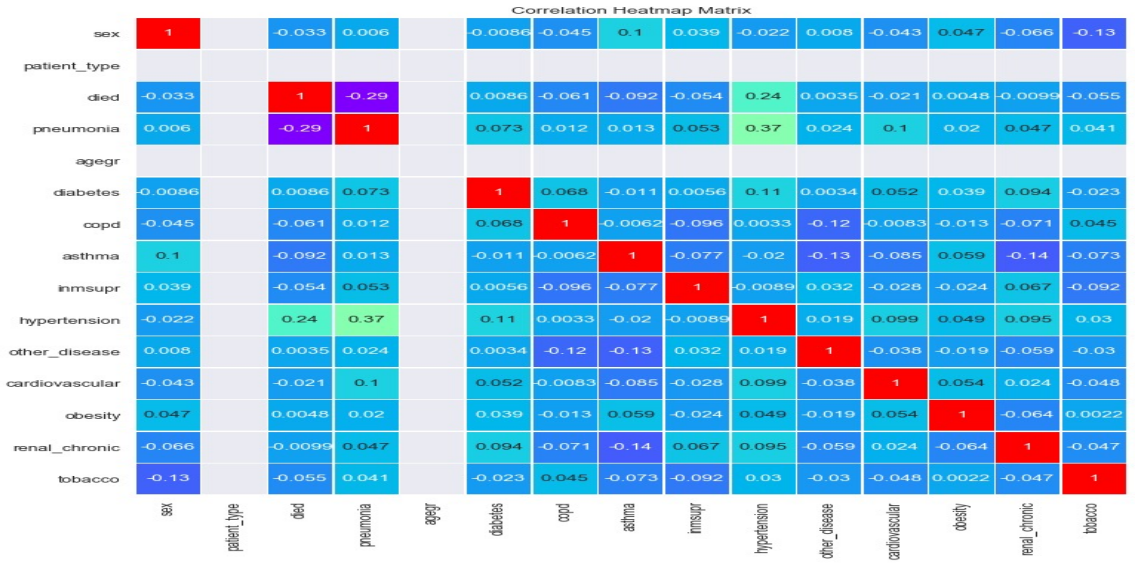






K-Means
Συστάδα 6 από 9 Συστάδες

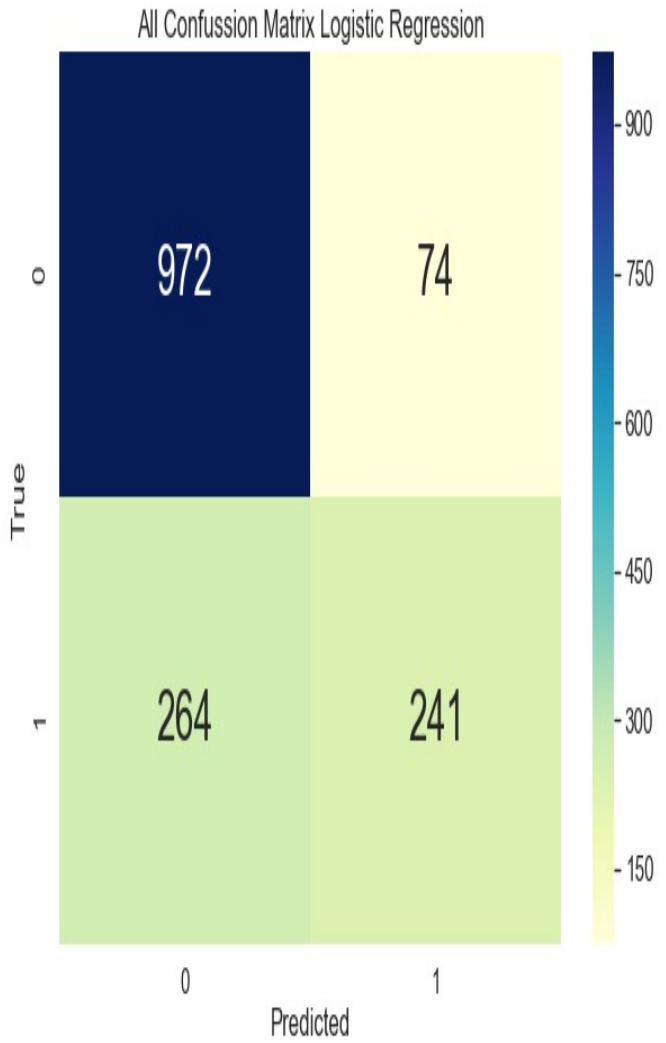


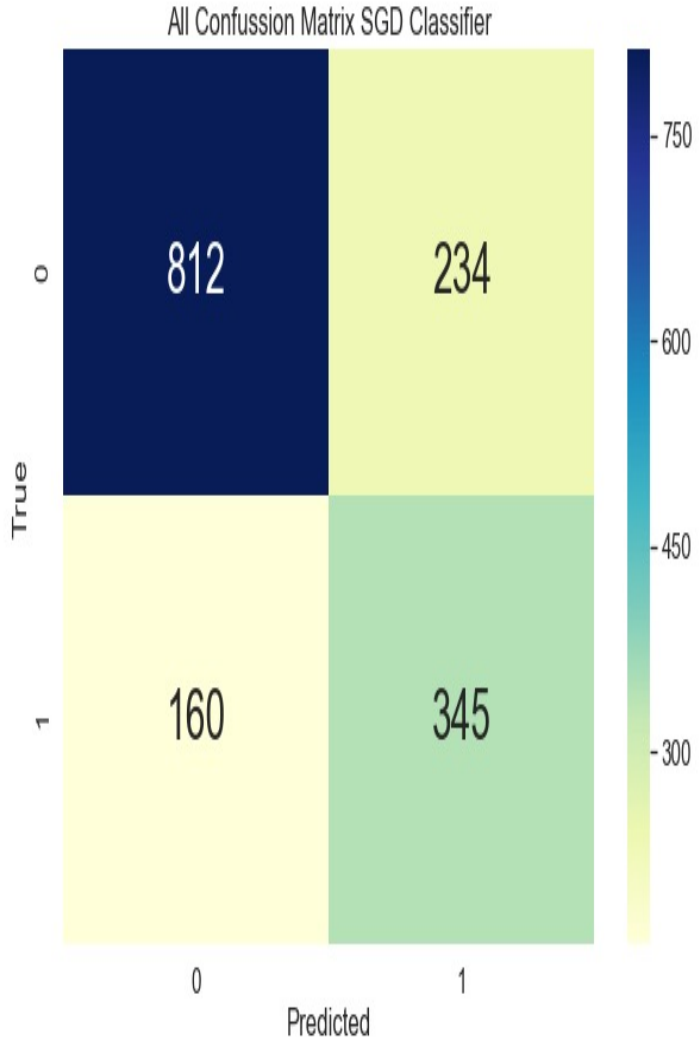


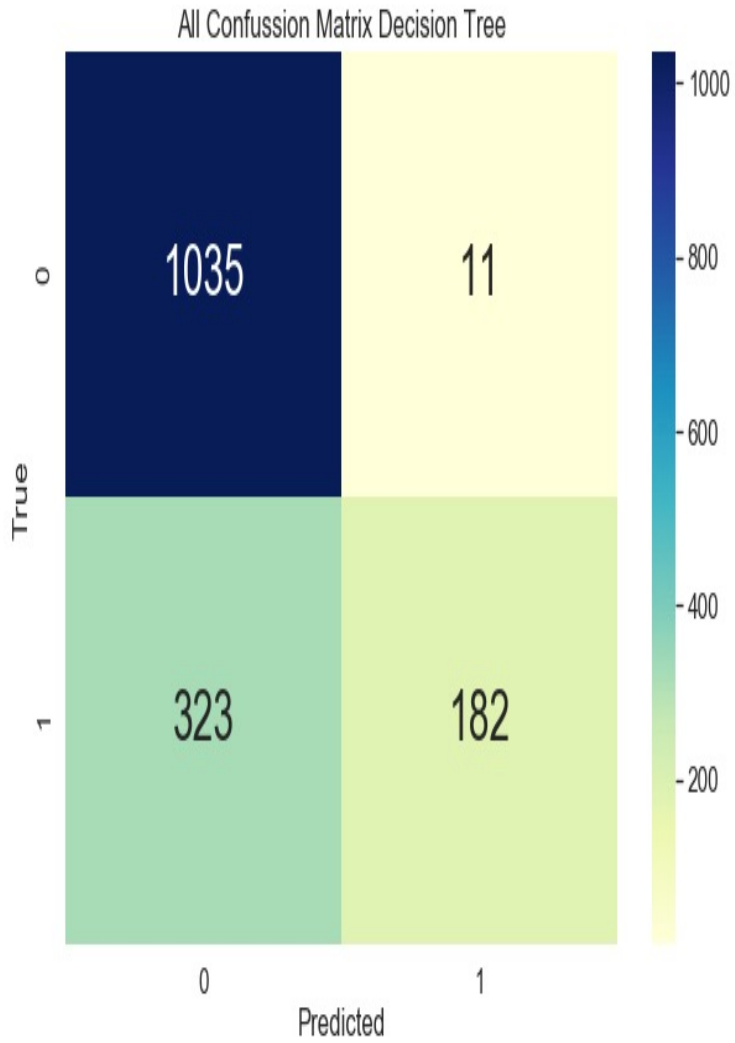
	precision	recall	f1-score
	Logistic Regression		
0	0.84	0.99	0.91
1	0.00	0.00	0.00
	0.70	0.83	0.76
	SDG Classifier		
0	0.94	0.81	0.87
1	0.42	0.71	0.52
	0.85	0.79	0.81
	Decision Tree		
0	0.84	1.00	0.91
1	0.00	0.00	0.00
	0.70	0.84	0.76

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

K-Means
ΑΠΟΤΕΛΕΣΜΑΤΑ

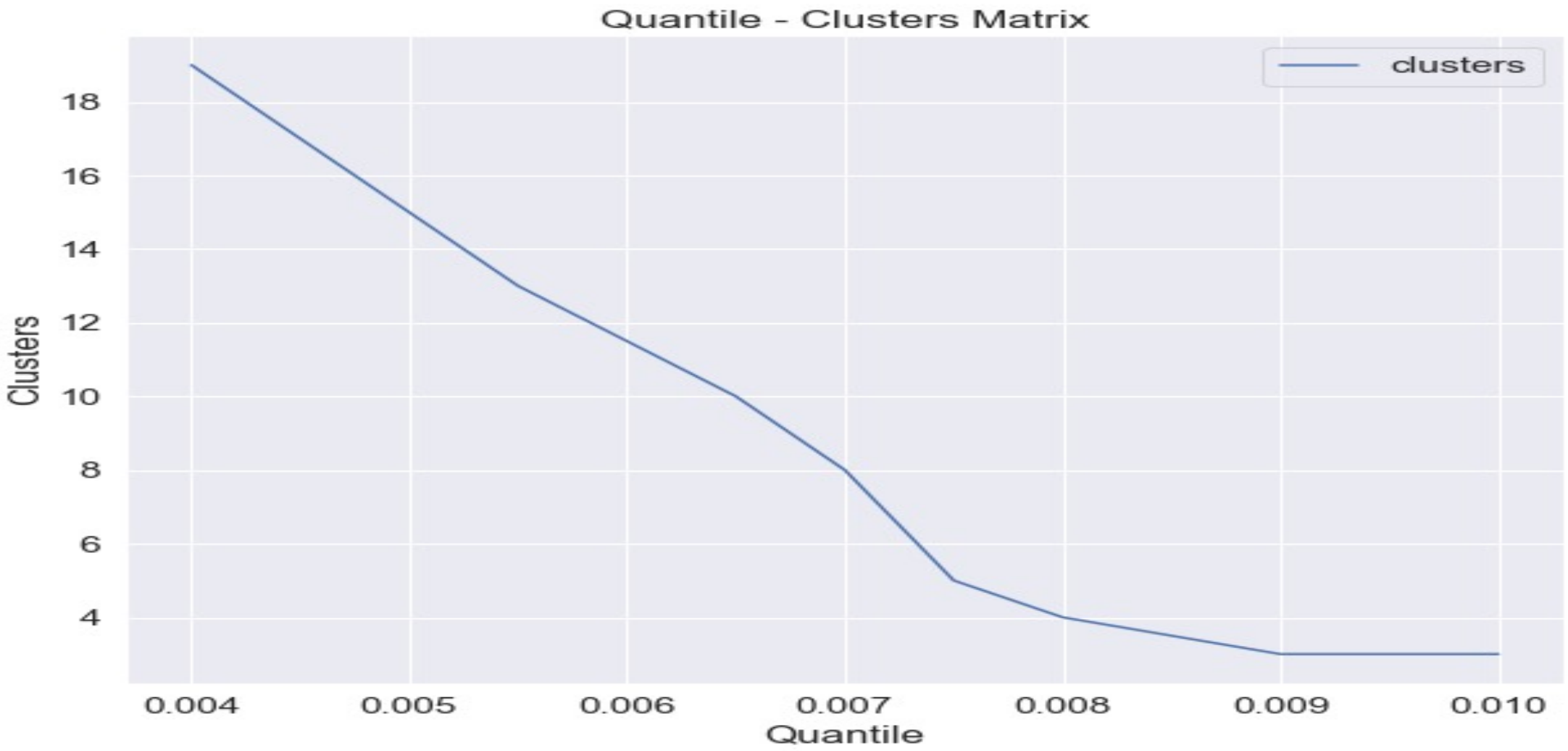






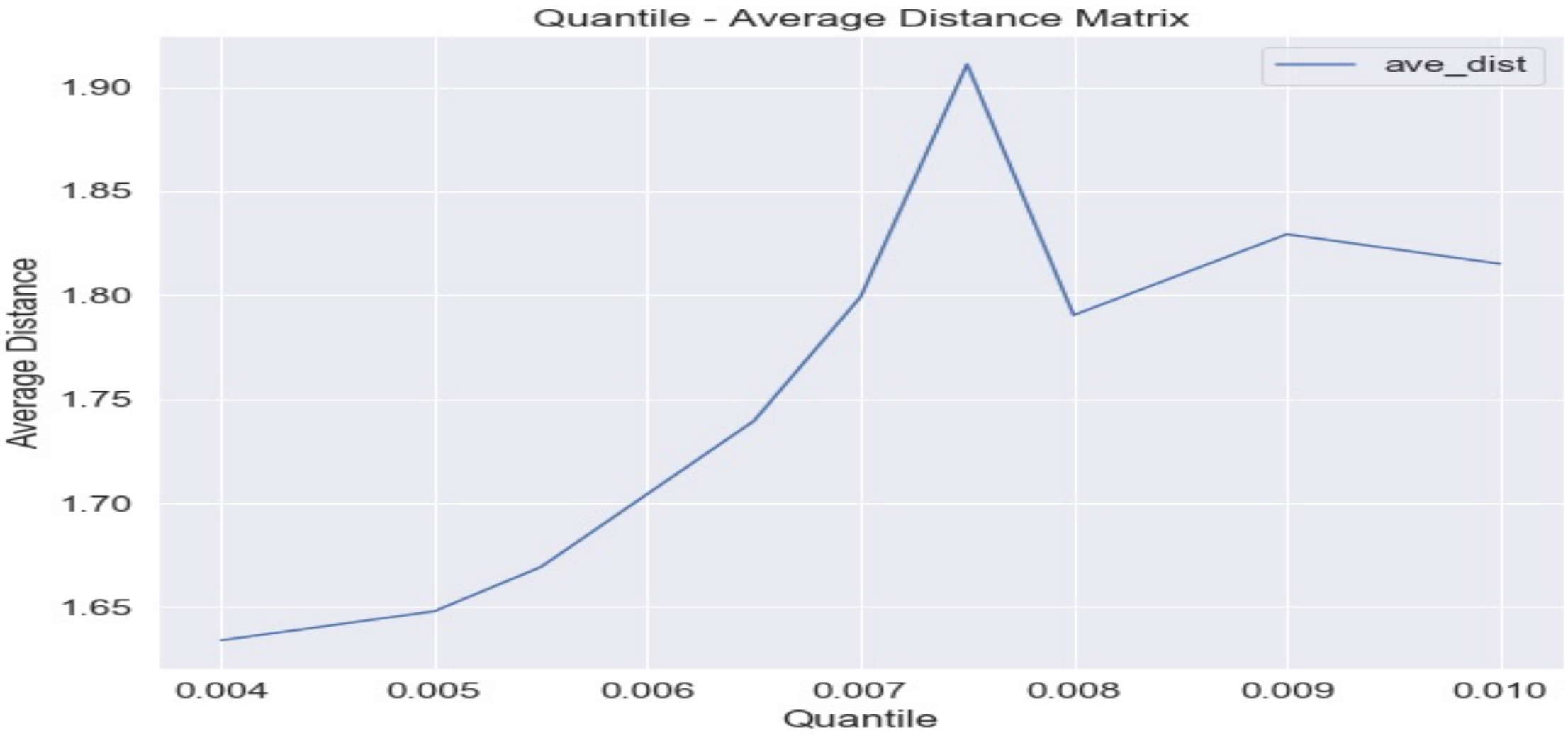
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Mean-Shift Quantile



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Mean-Shift
ΜΟ Απόστασης

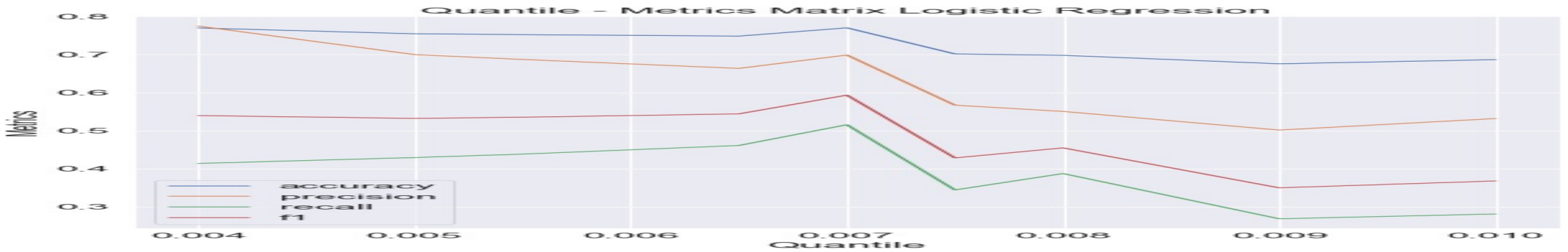


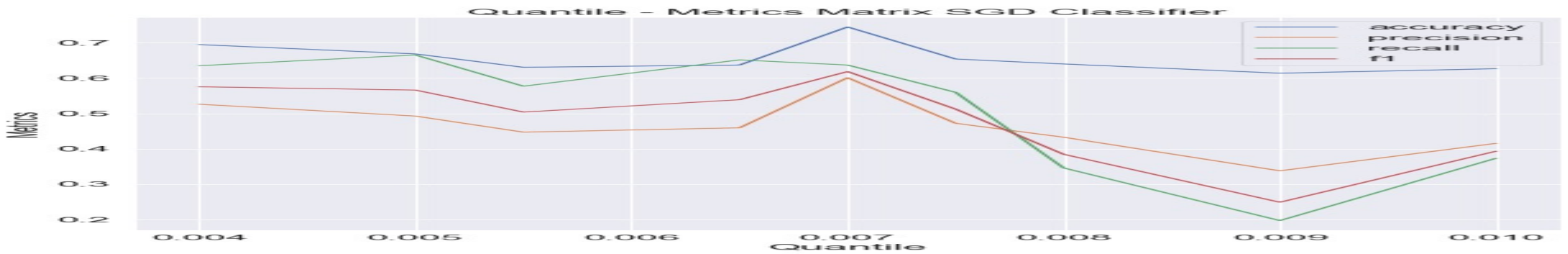
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

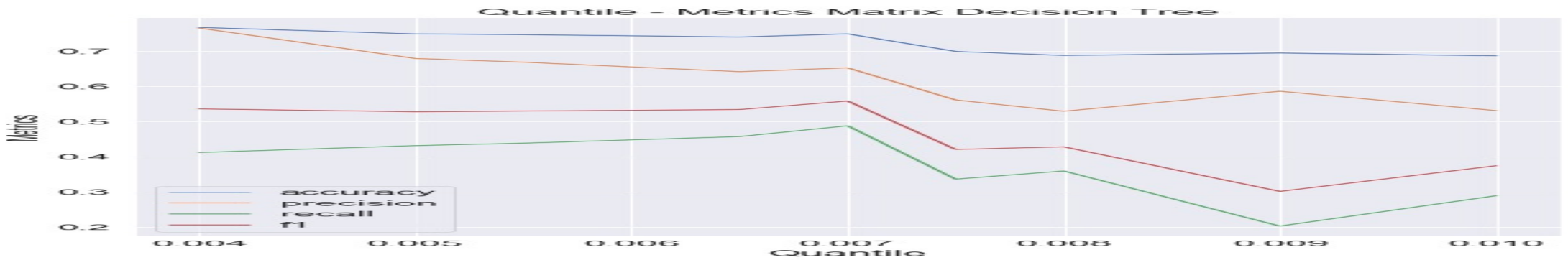
Mean-Shift Silhouette

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Mean-Shift
Αλγόριθμοι Κατηγοριοποίησης

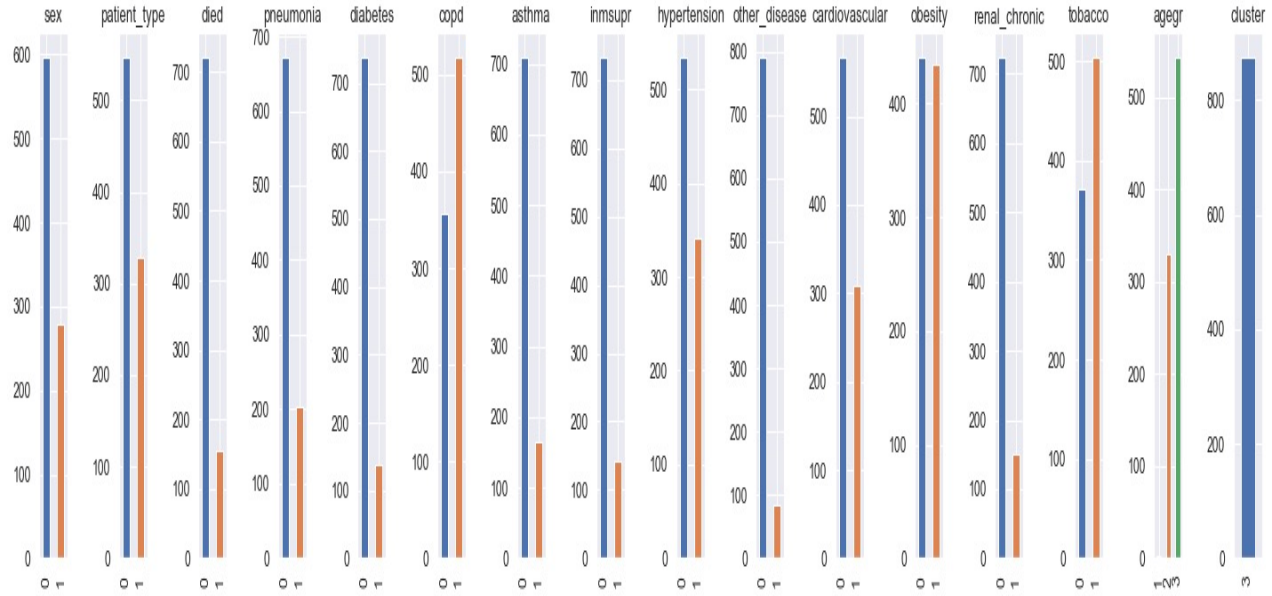


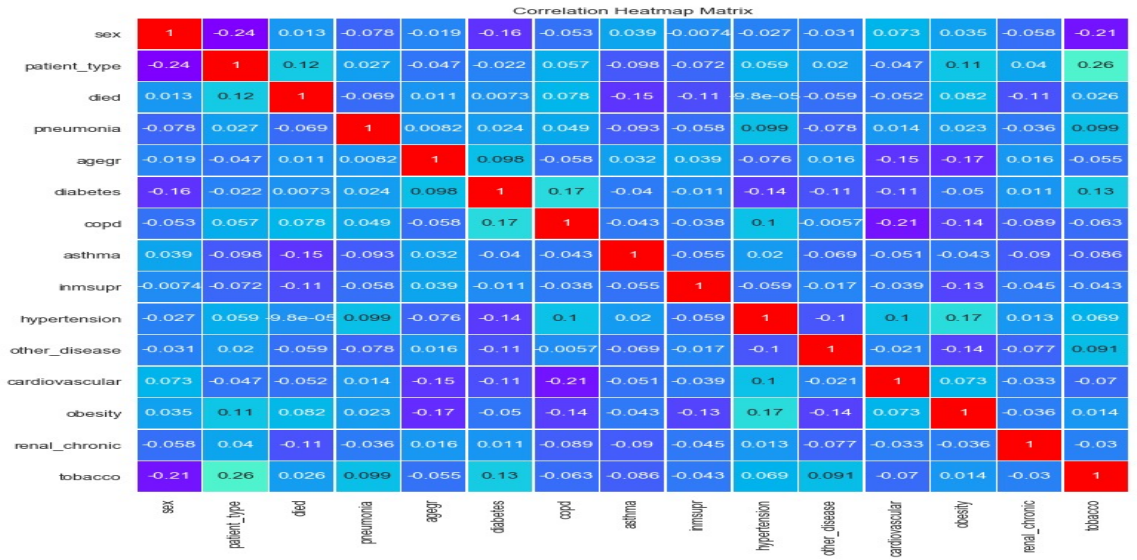




ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Mean-Shift
Συστάδα 3 από Συστάδες 8

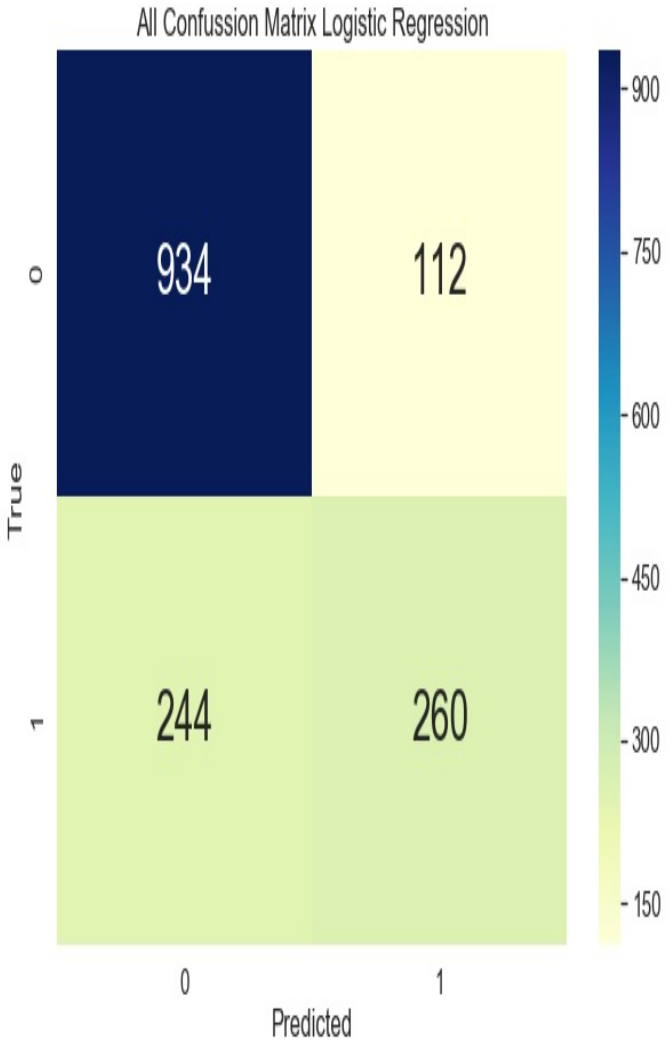


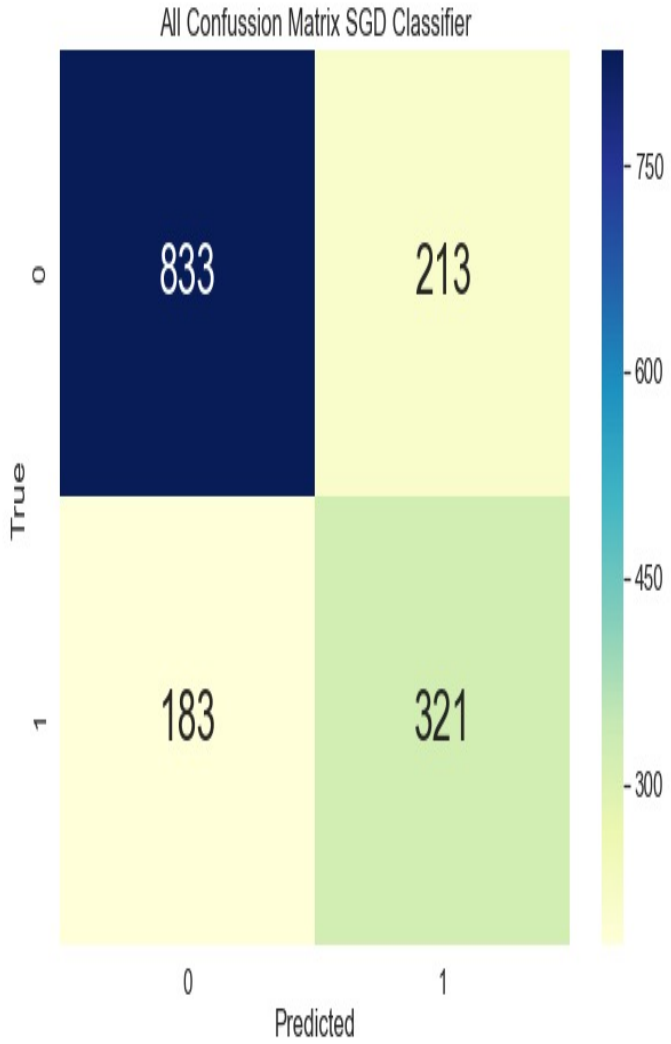


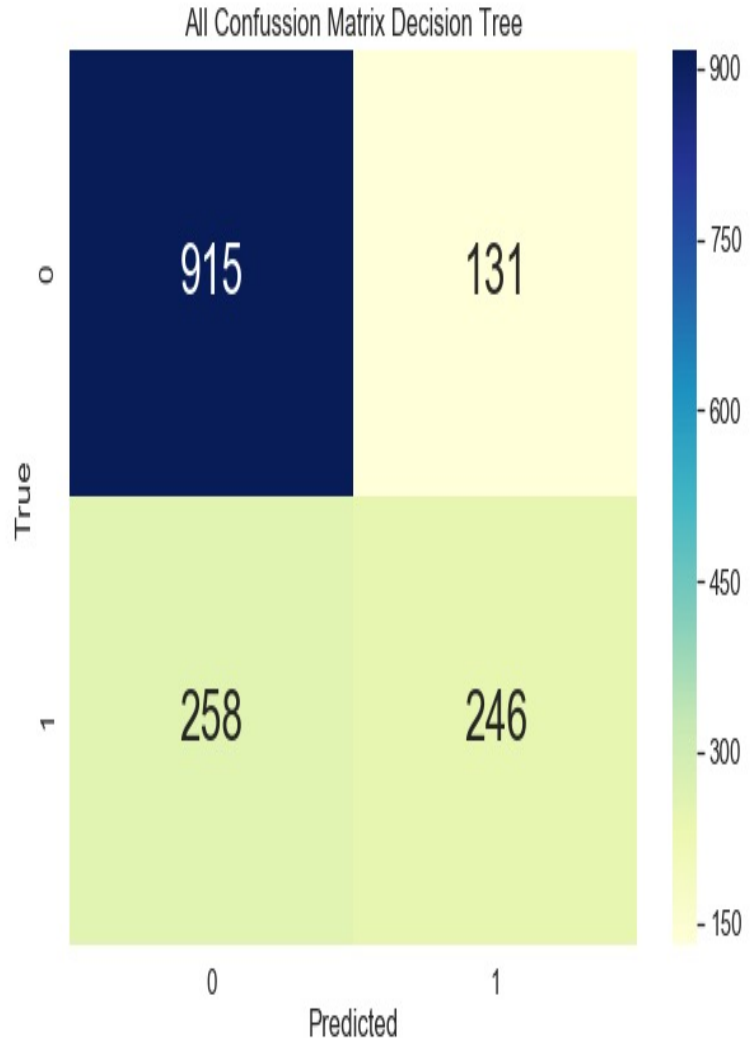
	precision	recall	f1-score
	Logistic Regression		
0	0.82	1.00	0.90
1	0.00	0.00	0.00
	0.68	0.82	0.74
	SDG Classifier		
0	0.84	0.86	0.85
1	0.28	0.26	0.27
	0.74	0.75	0.75
	Decision Tree		
0	0.82	0.99	0.90
1	0.00	0.00	0.00
	0.68	0.81	0.74

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Mean-Shift
ΑΠΟΤΕΛΕΣΜΑΤΑ

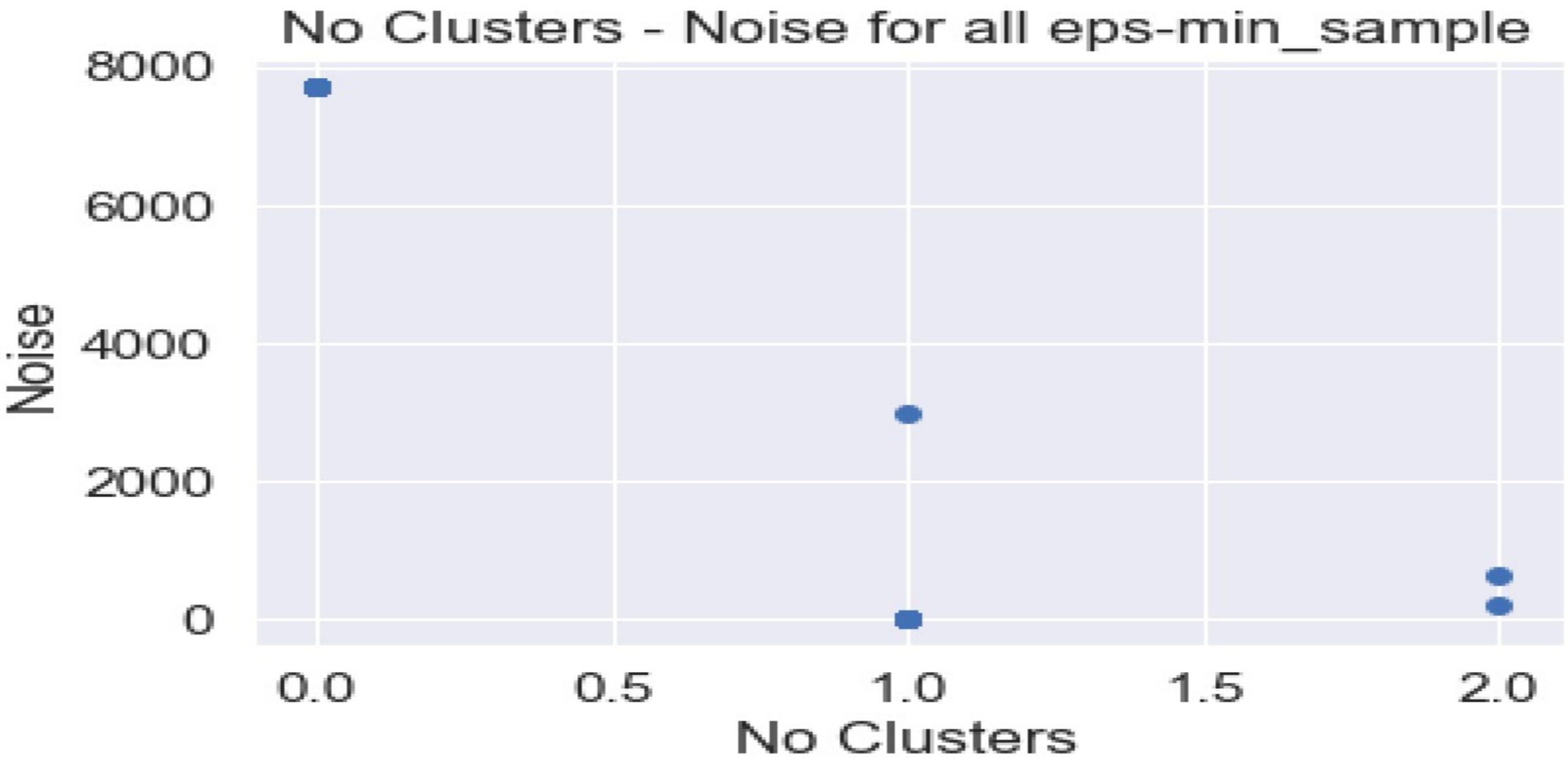






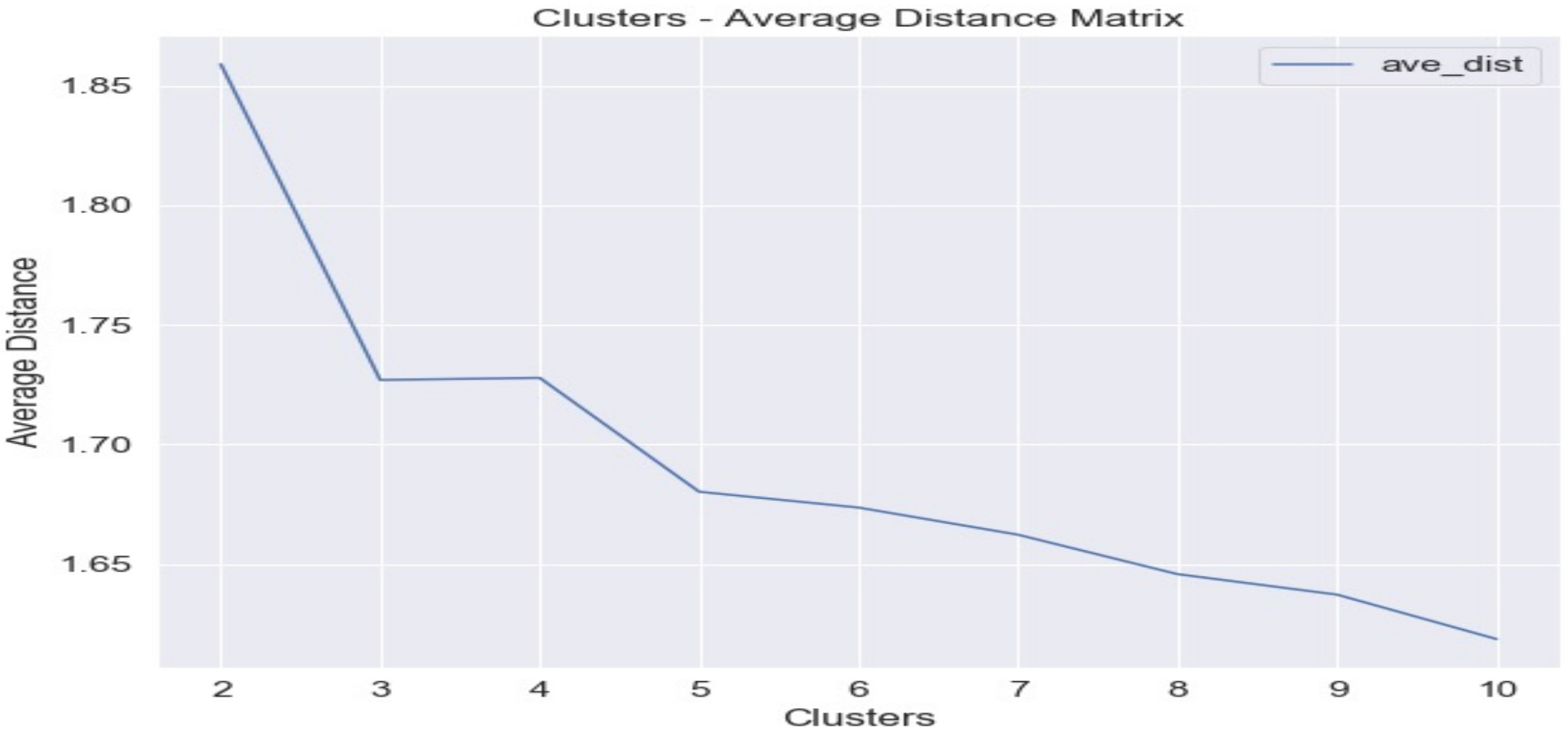
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

DBSCAN



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

GMM using EM
ΜΟ Απόστασης

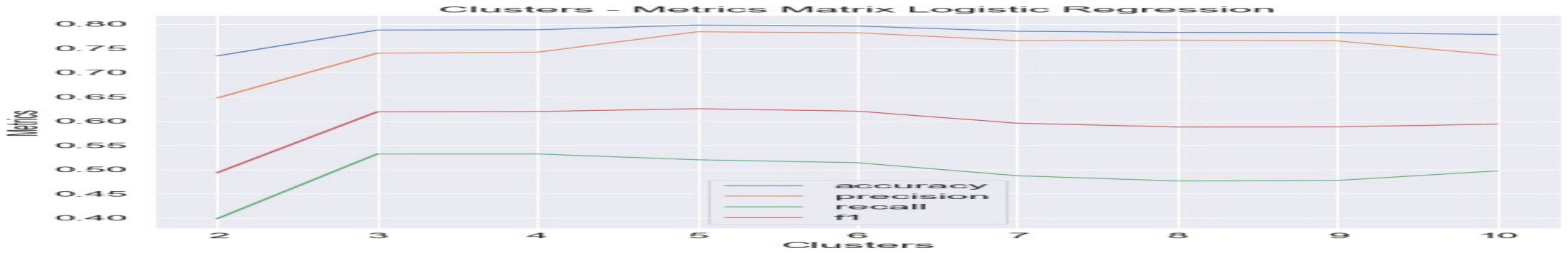


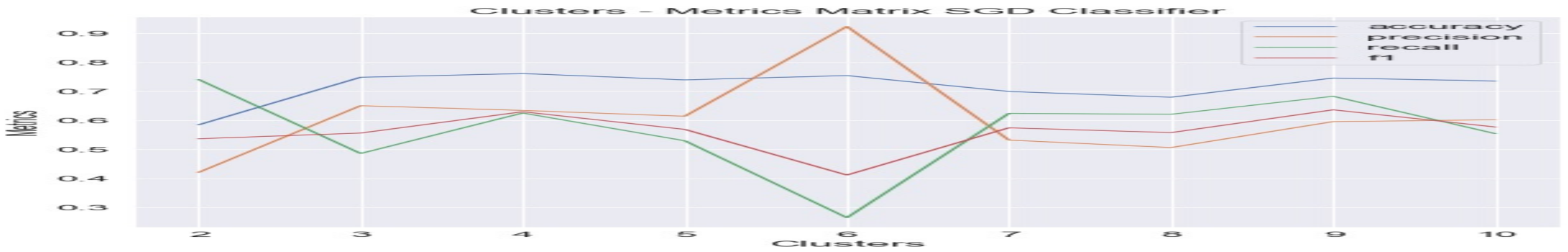
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

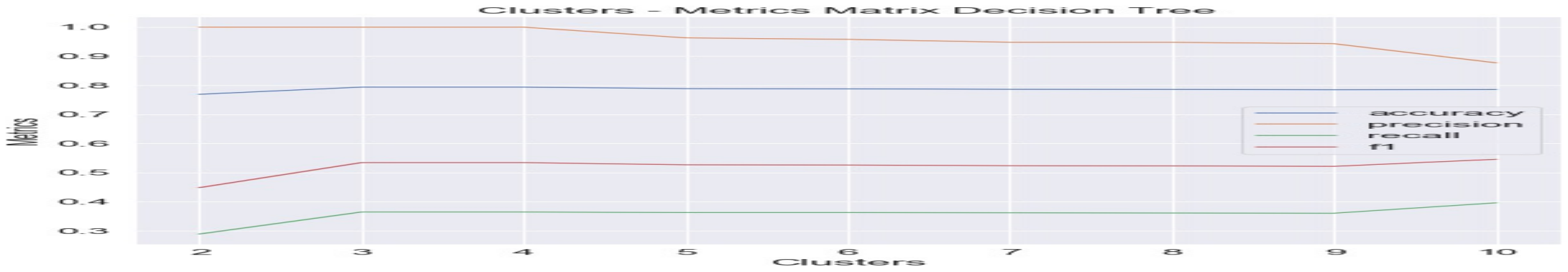
GMM using EM
Silhouette

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

GMM using EM
Αλγόριθμοι Κατηγοριοποίησης

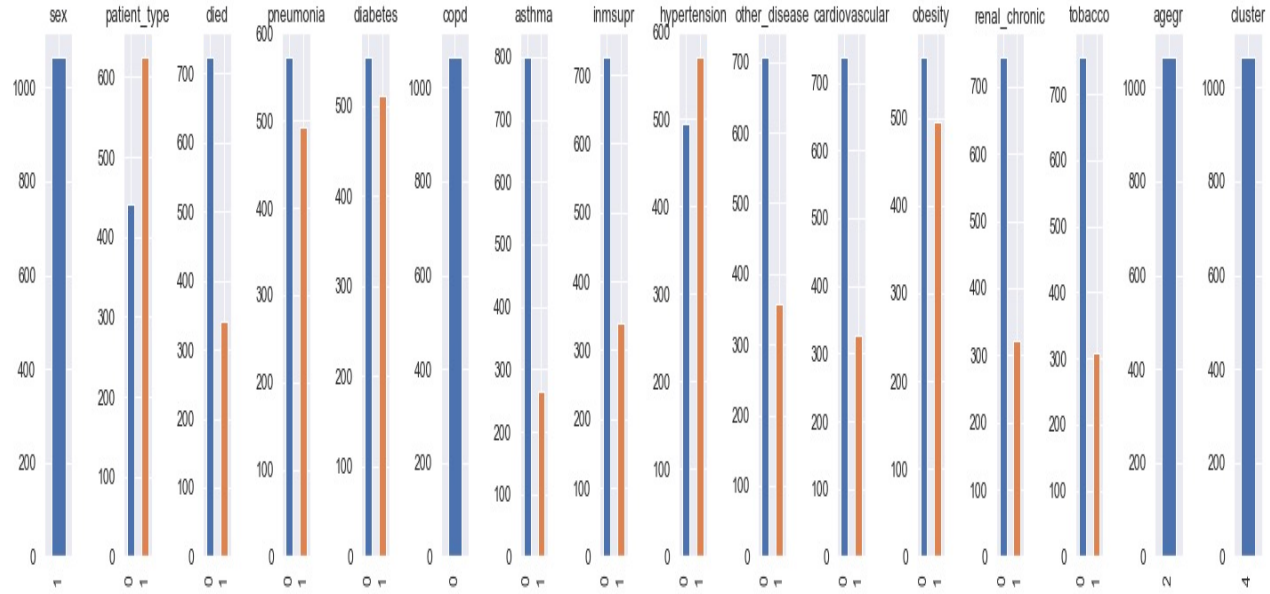


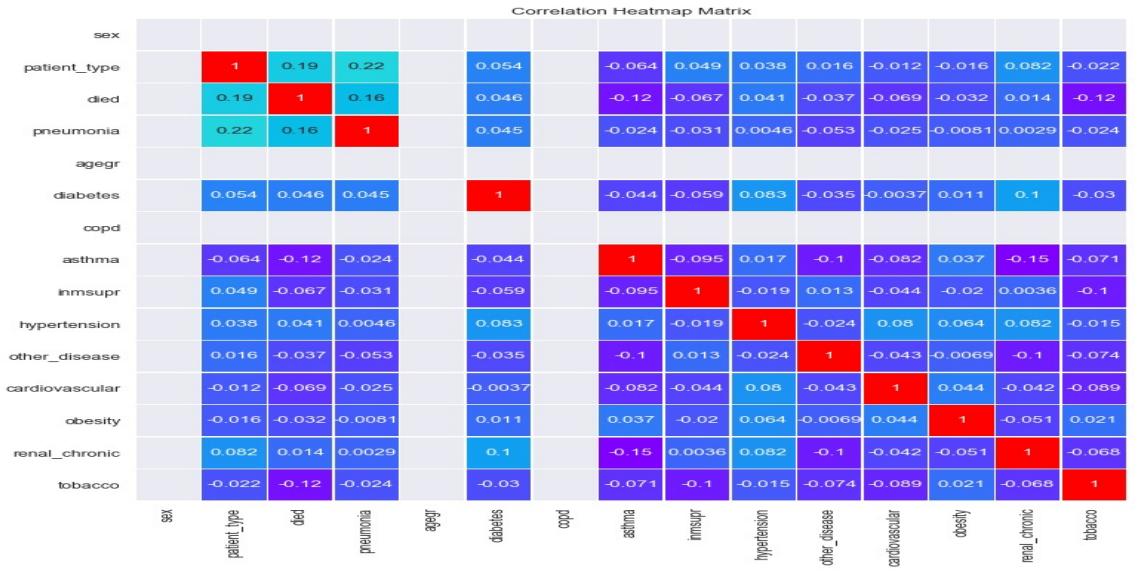




ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

GMM using EM
Συστάδα 4 από Συστάδες 9

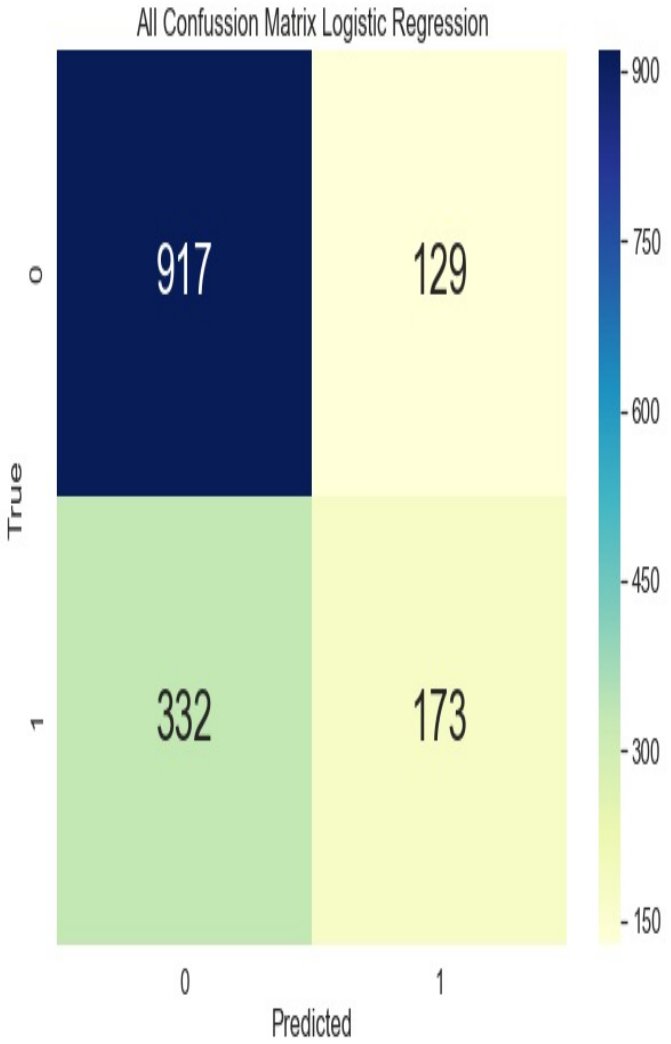


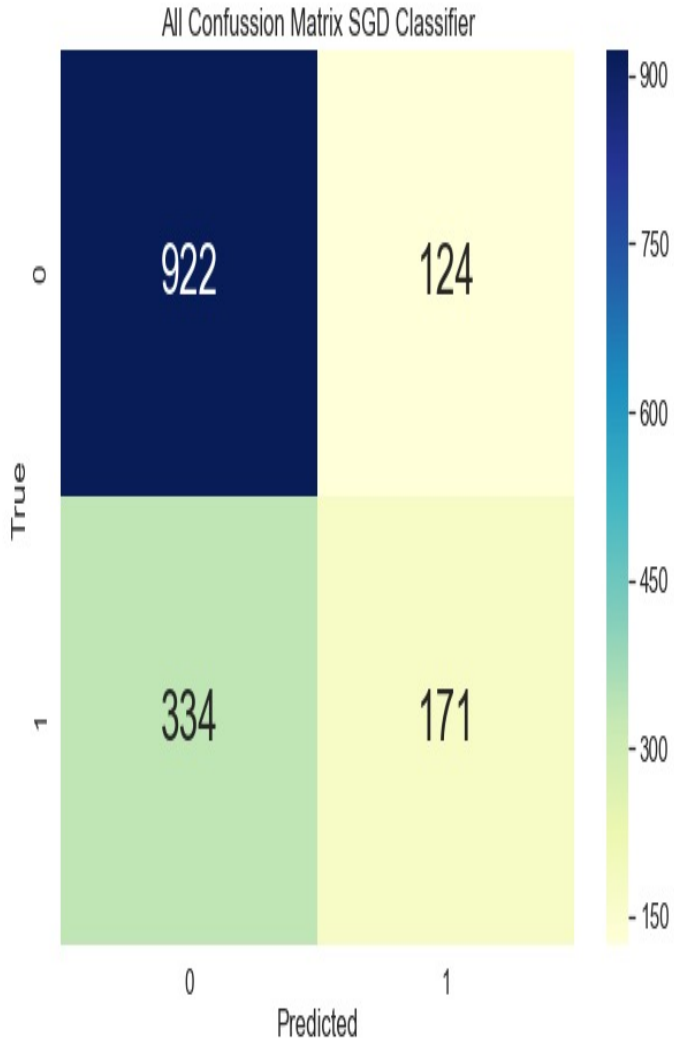


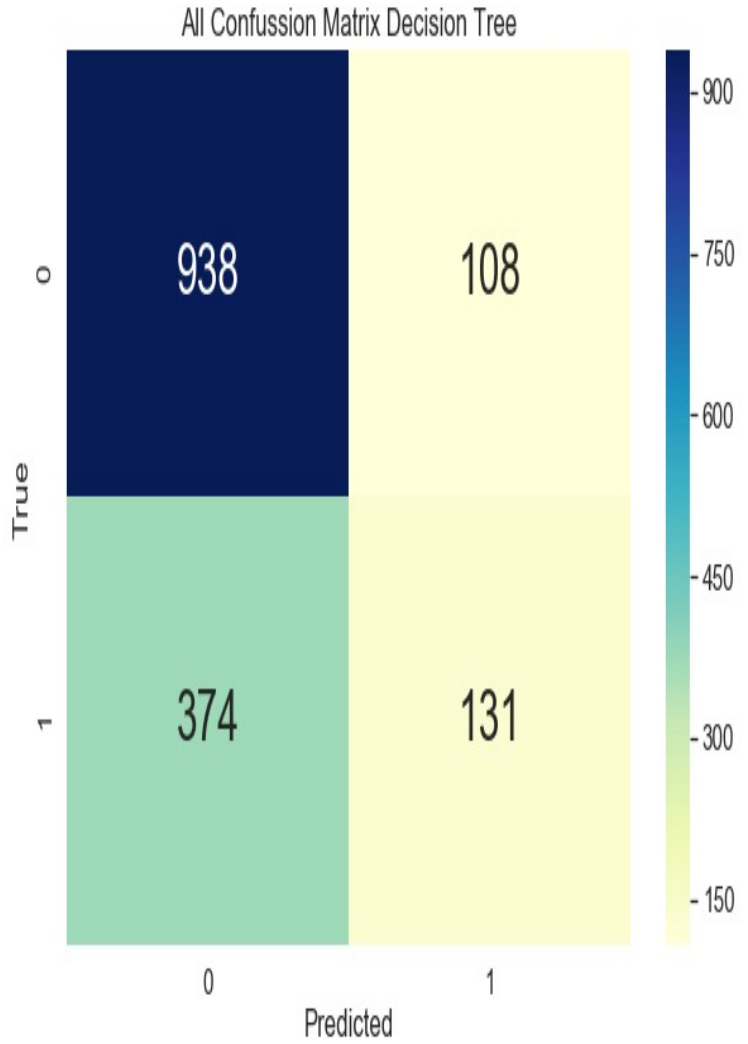
	precision	recall	f1-score
	Logistic Regression		
0	0.71	0.83	0.77
1	0.45	0.29	0.35
	0.63	0.66	0.63
	SDG Classifier		
0	0.69	0.97	0.80
1	0.50	0.07	0.13
	0.63	0.68	0.58
	Decision Tree		
0	0.68	0.95	0.79
1	0.30	0.04	0.08
	0.56	0.66	0.56

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

GMM using EM
ΑΠΟΤΕΛΕΣΜΑΤΑ

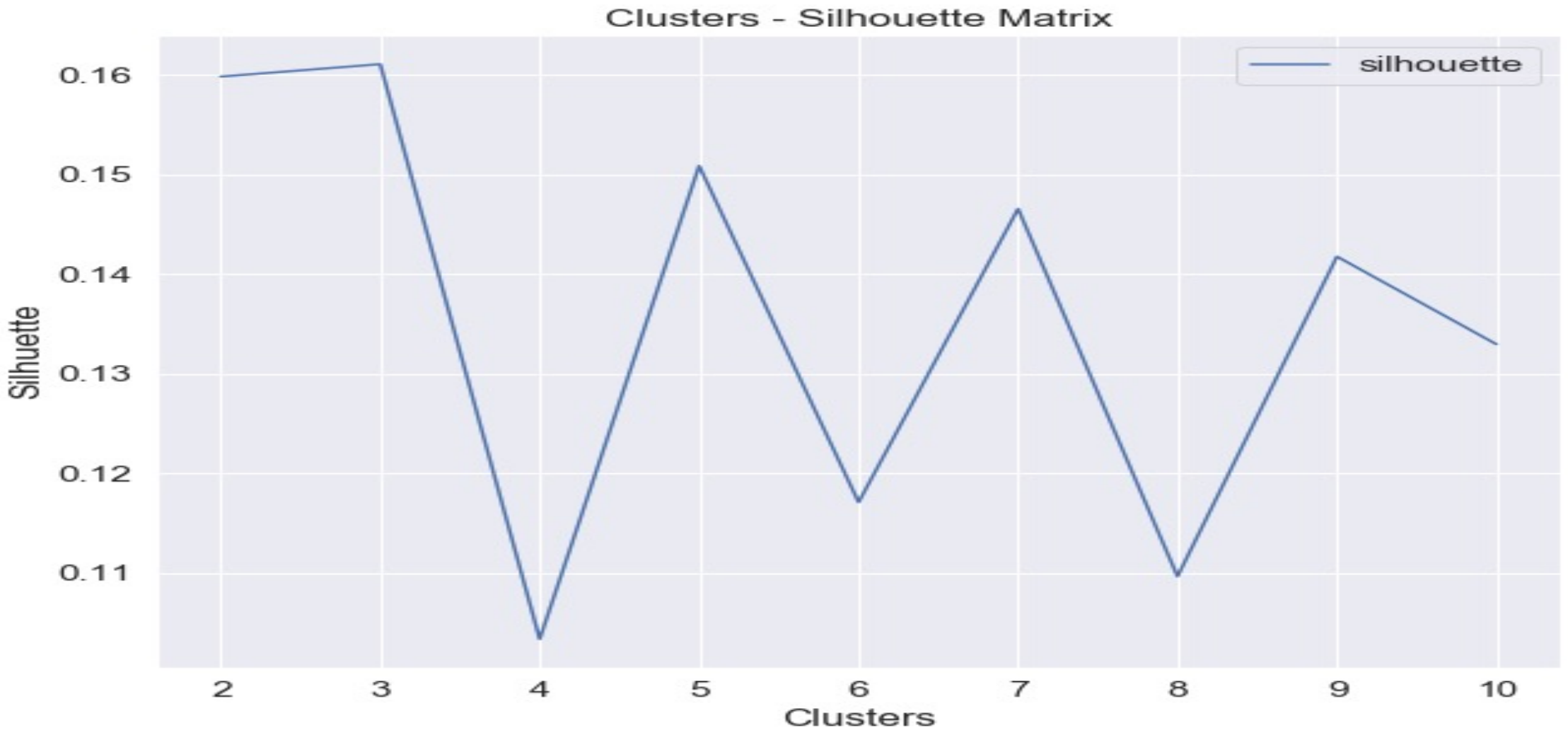






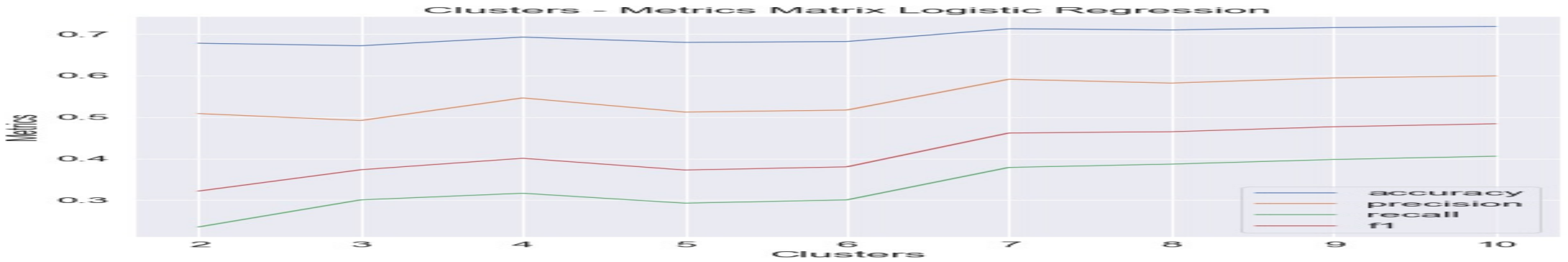
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

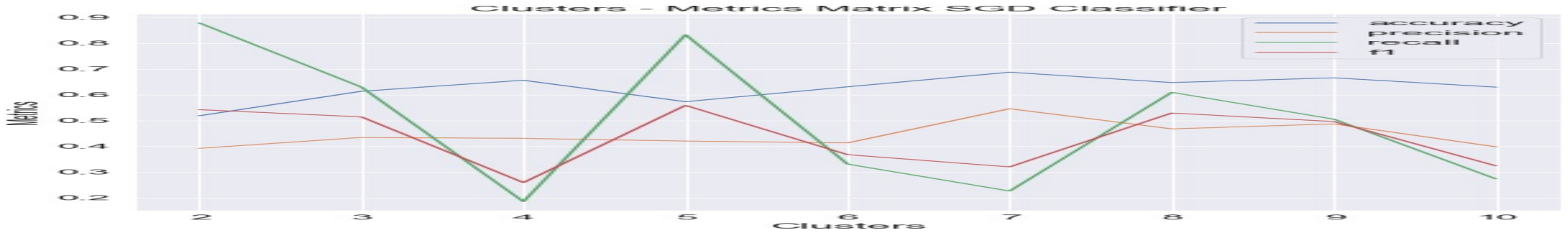
Agglomerative Silhouette

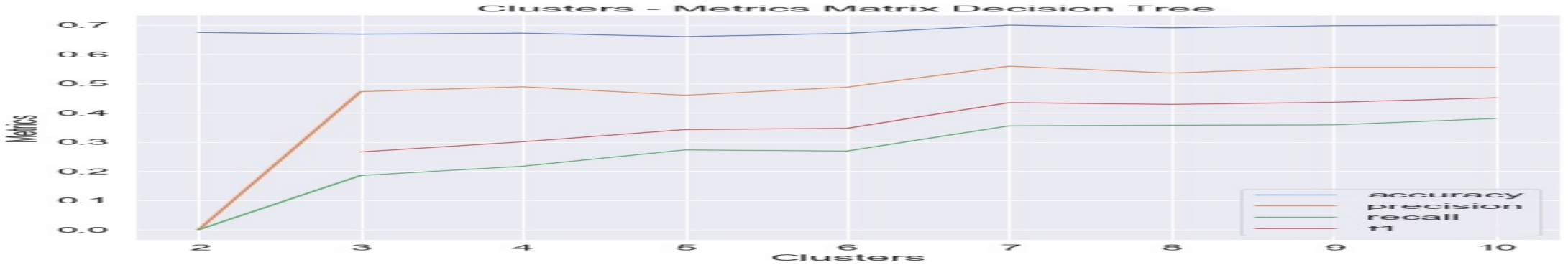


ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Agglomerative
Αλγόριθμοι Κατηγοριοποίησης

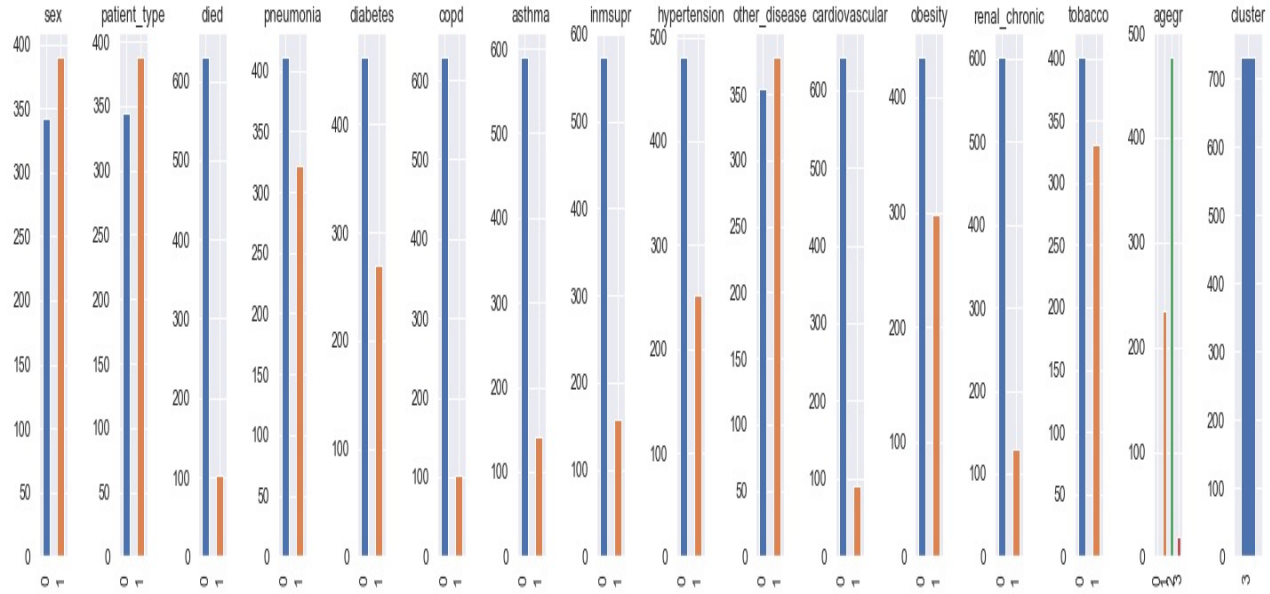


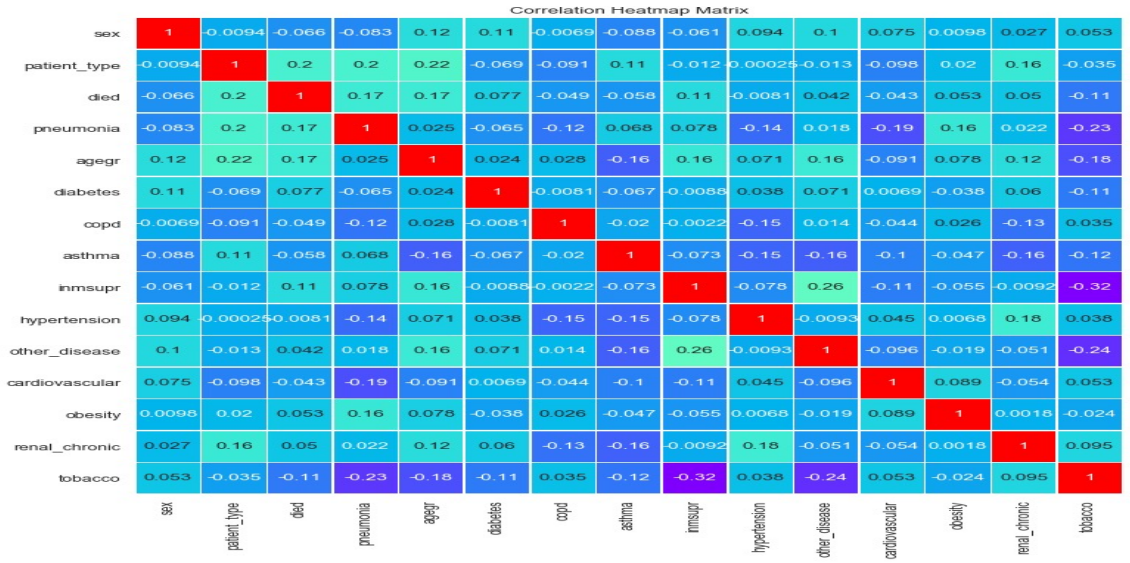




ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Agglomerative
Συστάδα 3 από Συστάδες 9

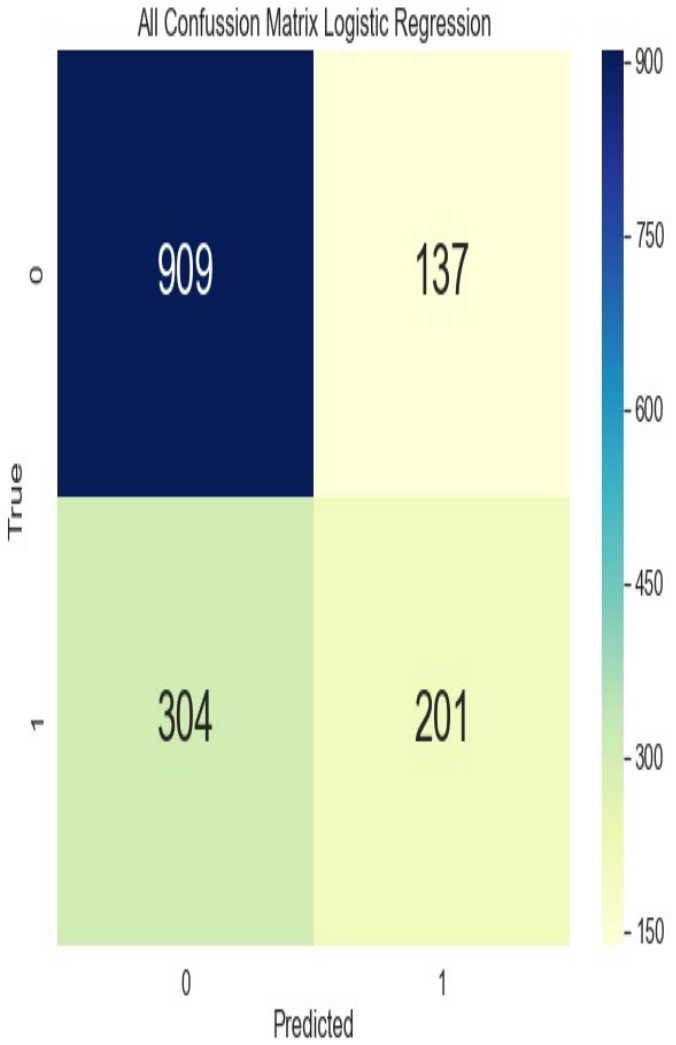


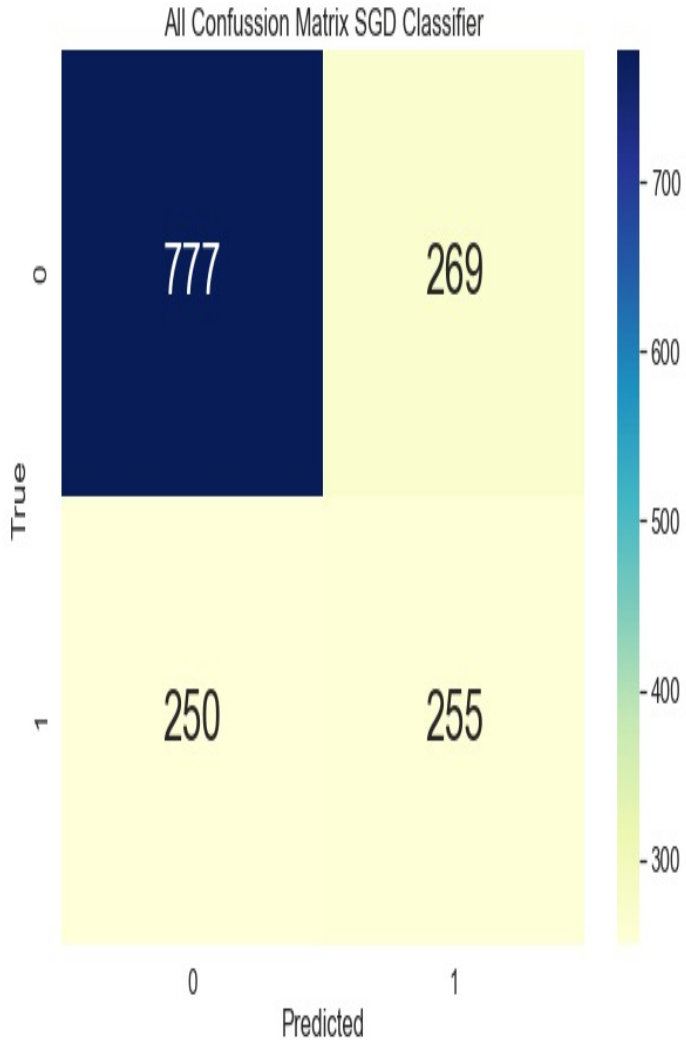


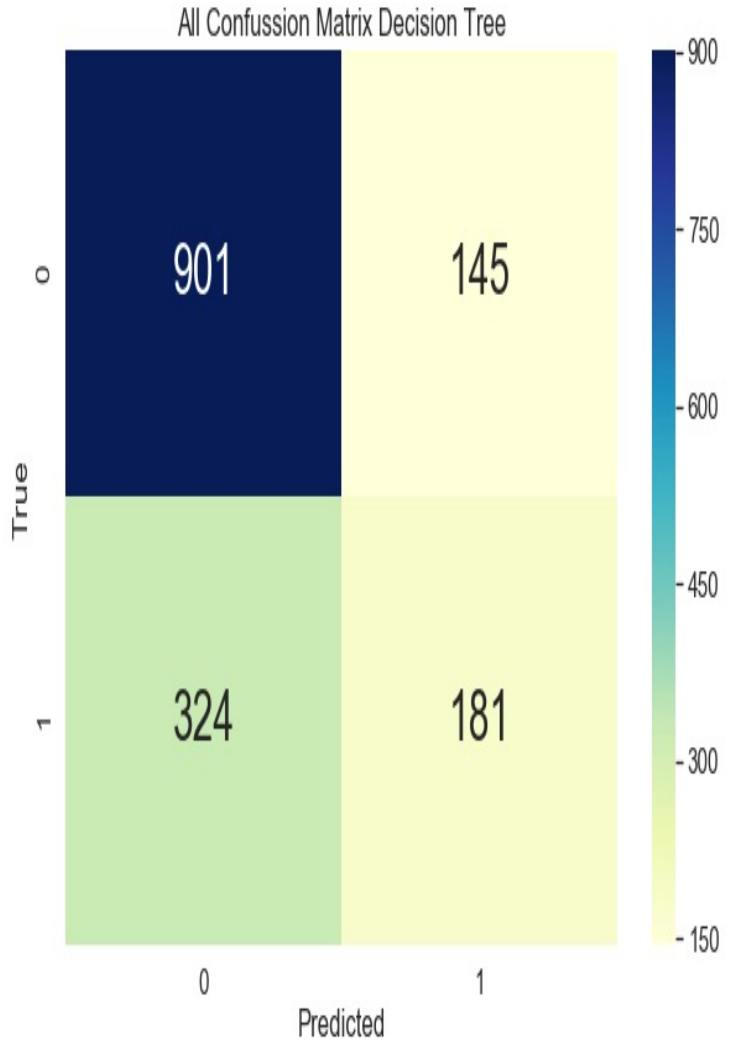
	precision	recall	f1-score
	Logistic Regression		
0	0.86	0.99	0.92
1	0.50	0.05	0.09
	0.81	0.86	0.80
	SDG Classifier		
0	0.85	0.78	0.81
1	0.12	0.19	0.15
	0.75	0.69	0.72
	Decision Tree		
0	0.86	1.00	0.92
1	0.00	0.00	0.00
	0.73	0.86	0.79

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Agglomerative ΑΠΟΤΕΛΕΣΜΑΤΑ







ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΣΥΜΠΕΡΑΣΜΑΤΑ
Πρόβλεψη Θανόντων

	Logistic Regression	SGD Classifier	Decision Tree
K-Means	241 48,1%	345 68,9%	182 36,3%
Mean-Shift	260 51,9%	321 64,1%	246 49,1%
DBSCAN	--	--	--
EM using GMM	173 34,5%	171 34,1%	131 26,1%
Agglomerative	201 40,1%	255 50,9%	181 36,1%

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΣΥΜΠΕΡΑΣΜΑΤΑ
Πρόβλεψη Μη-Θανόντων

	Logistic Regression	SGD Classifier	Decision Tree
K-Means	972 93,2%	812 77,9%	1035 99,2%
Mean-Shift	974 93,4%	833 79,9%	915 87,7%
DBSCAN	--	--	--
EM using GMM	917 87,9%	922 88,4%	938 90,0%
Agglomerative	909 87,2%	777 74,5%	901 86,4%

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΣΥΜΠΕΡΑΣΜΑΤΑ
Λάθη Προβλέψεων

	Logistic Regression	SGD Classifier	Decision Tree
K-Means	338 21,9%	394 25,5%	334 21,6%
Mean-Shift	356 23,1%	396 25,7%	389 25,2%
DBSCAN	--	--	--
EM using GMM	461 29,9%	458 29,7%	482 31,2%
Agglomerative	441 28,6%	519 33,6%	469 30,4%

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ