**ΑΝΑΛΥΤΙΚΗ ΔΕΔΟΜΕΝΩΝ**

**Απαλλακτική Εργασία 2021**

****

Μέλη ομάδας:

Θεοδωροπούλου Χριστίνα Π18047

Μαγνήσαλης Παναγιώτης – Ταξιάρχης Π18096

Τσίγκου Σοφία – Ανθή Π18159

**Εισαγωγή**

Για την διεκπεραίωση της εργασίας χρησιμοποιήθηκε το colab της google. Υπάρχει διαφορετικό project file για κάθε ερώτημα της εργασίας. Τα προγράμματα εκτελούνται με την προσθήκη των project files και των xls αρχείων (CTG.xls, data\_akbilgic.xlsx) στο colab.

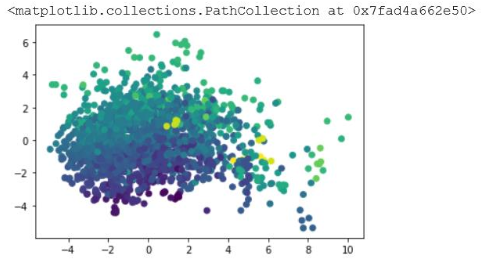
**ΜΕΡΟΣ A:**

**Ερώτημα 1**(βλ. αρχείο a\_1\_2.ipynb)

Χρήση της dropna για τις nan τιμές, αφαιρώντας τις τρείς τελευταίες εγγραφές. Έπειτα χρησιμοποιήθηκαν τα χαρακτηριστικά MIN και MAX, των οποίων οι εγγραφές διπλασιάστηκαν μέσω του pca για να μπορούμε να επεξεργαστούμε καλύτερα τα δεδομένα μας.

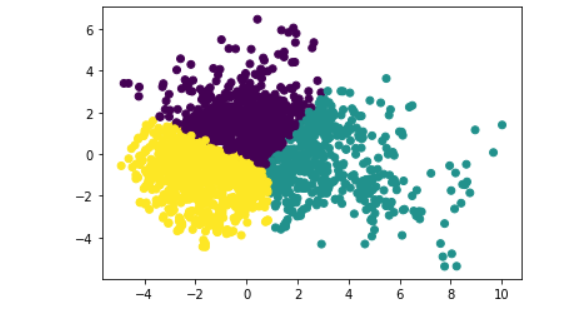
**Ερώτημα 2** (βλ. αρχείο a\_1\_2.ipynb)

Δημιουργήθηκε ένα plt διάγραμμα με τα αρχικά δεδομένα.

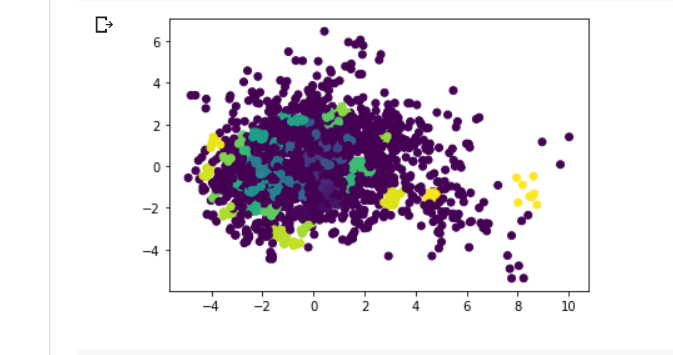


Στην συνέχεια χρησιμοποιήθηκαν τεχνικές clustering από τη βιβλιοθήκη scikit-learn.

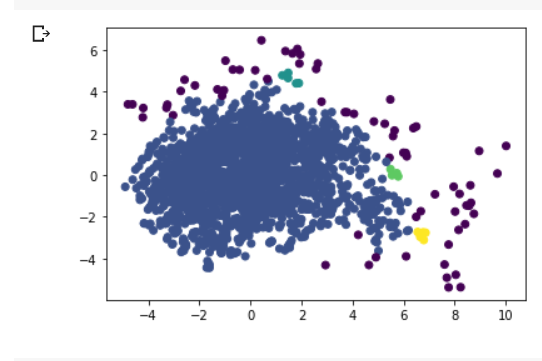
k-means με αριθμό clusters 3



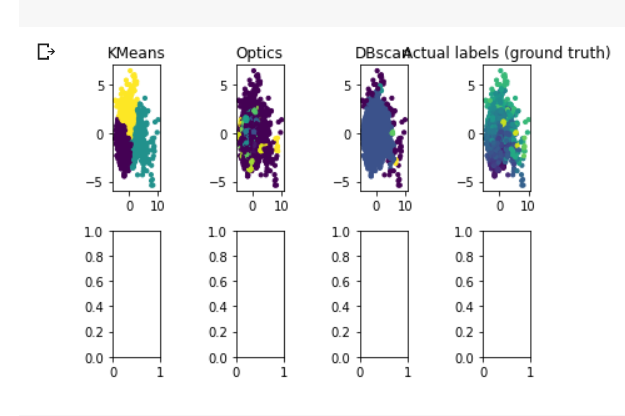
Optics με min\_samples 8



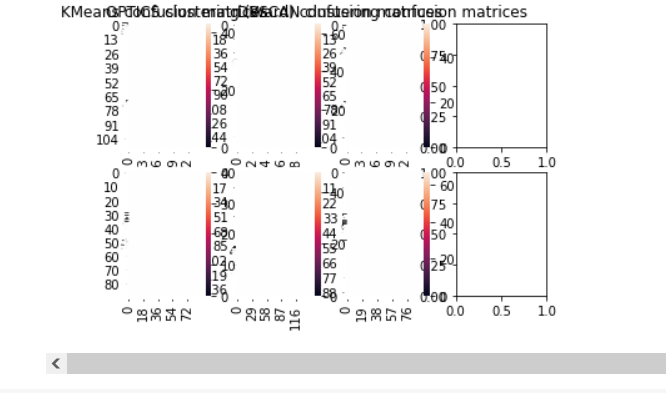
Dbscan



διαγράµµατα



confusion matrices



**Ερώτημα 3** (βλ. αρχείο a\_3.ipynb)

Χρησιμοποιήσαμε τον MLP classifier



**ΜΕΡΟΣ Β:**

**Ερώτημα 1** (βλ. αρχείο b\_1.ipynb)

Make data supervised

Εικόνα που περιέχει πίνακας

Περιγραφή που δημιουργήθηκε αυτόματα

Και σε μορφή {sample , steps , features }

**Ερώτημα 2**

1. Time-series prediction µε MLP

Η γραφική απεικόνιση και scaling των δεδομένων είναι:



Μετατρέπουμε τα δεδομένα σε supervised

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

Τέλος ετοιμάζουμε τα δεδομένα για testing(50% train και 50% testing).

Εικόνα που περιέχει κείμενο

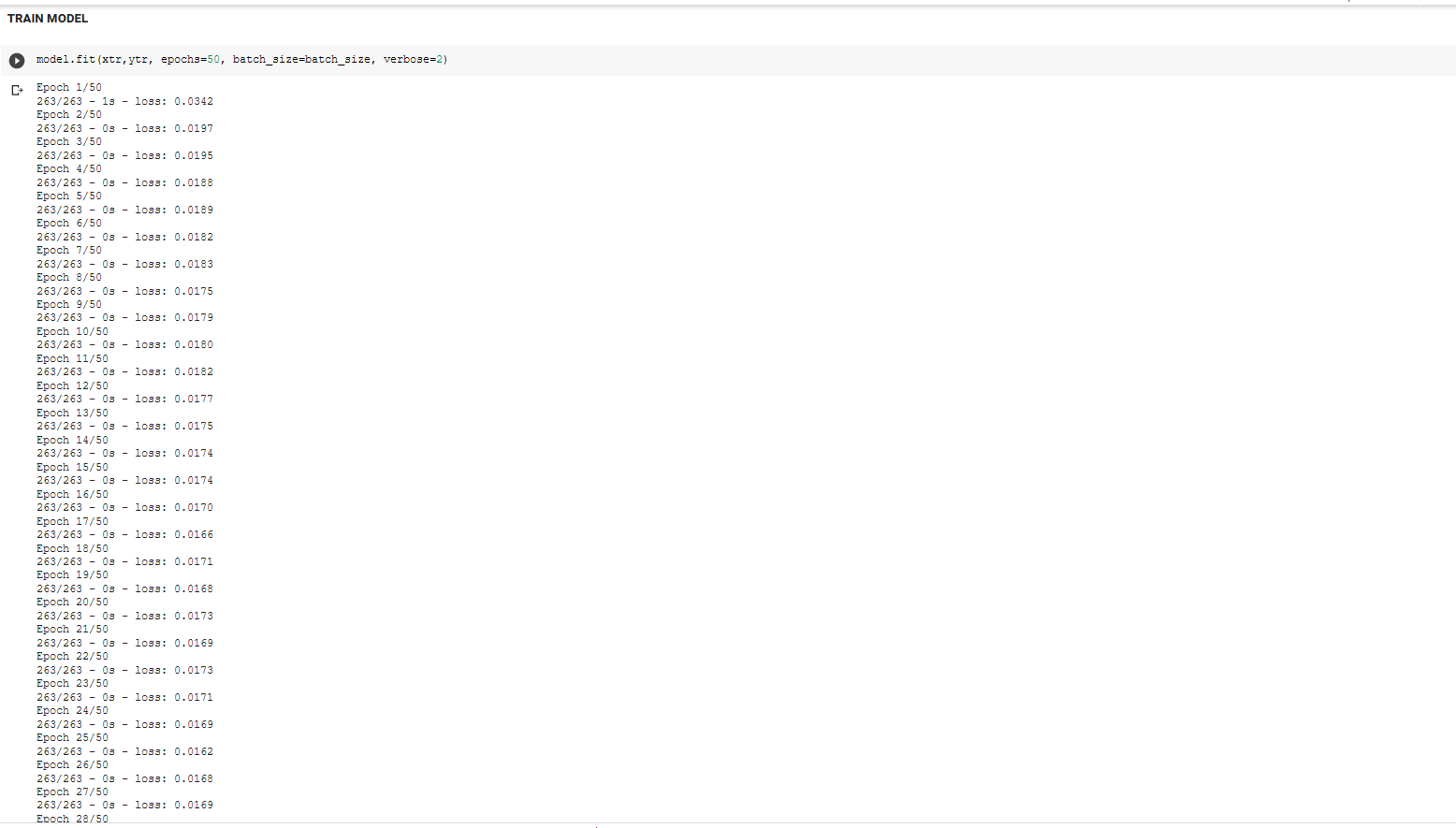
Περιγραφή που δημιουργήθηκε αυτόματα

Create NLP model

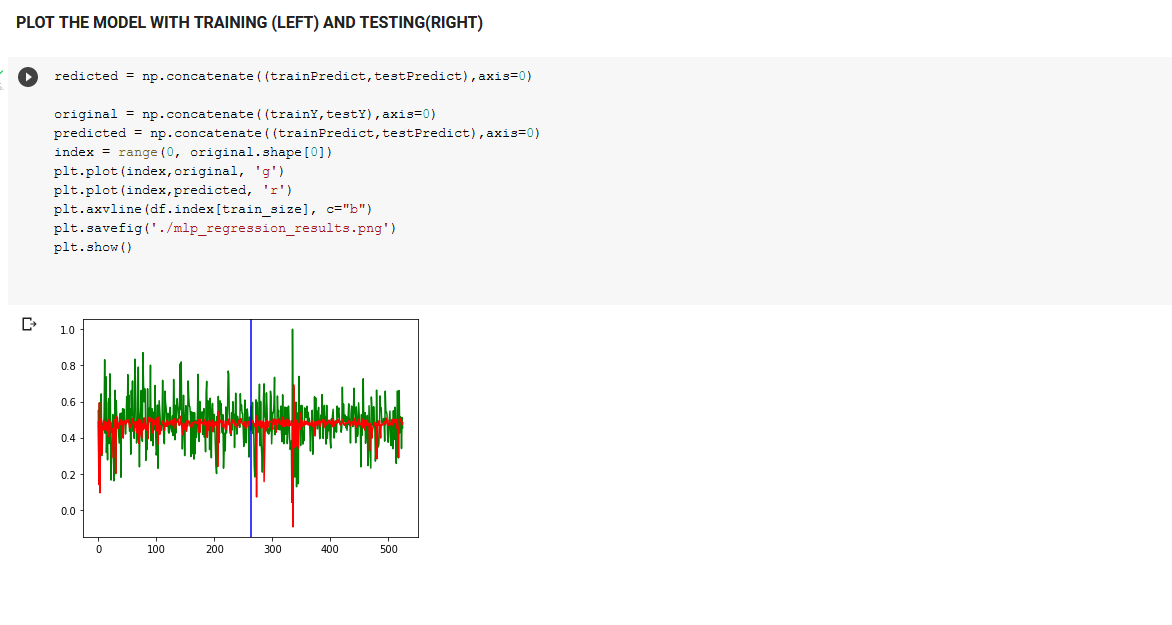
Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

Train model



Γραφική αναπαράσταση



1. Time-series prediction µε RNN

Μέχρι την προετοιμασία των δεδομένων για testing υλοποιείται το ίδιο κομμάτι κώδικα με το ερώτημα a.

Μετατρέπουμε τα δεδομένα σε 3D numpy Arrays

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

RNN model

Εικόνα που περιέχει πίνακας

Περιγραφή που δημιουργήθηκε αυτόματα

Ακολουθούμε τα ίδια βήματα training και normalization των δεδομένων όπως και στο προηγούμενο μοντέλο.

Εμφανίζουμε ένα διάγραμμα του μοντέλου όπου δεξιά είναι το testing και αριστερά το training.

