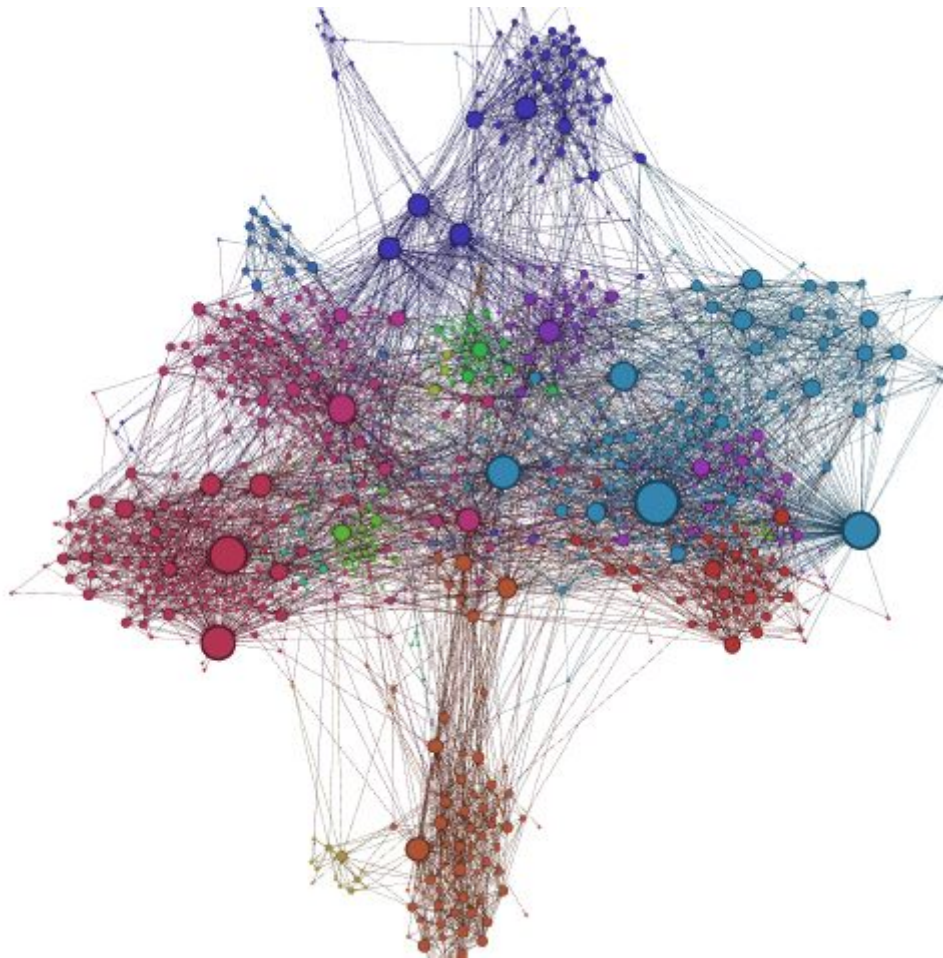


**Πανεπιστήμιο δυτικής Αττικής  
Σχολή Τεχνολογικών Εφαρμογών  
Τμήμα Πληροφορικής**



**Αναγνώριση προτύπων & Μηχανική Μάθηση  
Εργασία 4  
Σφυριδάκη Αγγελική cs151036  
cs151036@uniwa.gr**

## Περιχόμενα

Αρχεία	1
Σκοπός εργασίας	2
Επεξήγηση κώδικα	2
Αποτελέσματα με εικόνες	3
Αποτελέσματα	4
Απόδοση	5
Συμπεράσματα	5

## Αρχεία

### Scripts:

Raw: Ολόκληρος ο κώδικας.	Encoded: Ολόκληρος ο κώδικας.
Technics_with_Raw_data:	Technics_with_Autoencoder:
Εκτελεί το clustering	Εκτελεί το clustering
Test_Raw: Δοκιμάζει τα αποτελέσματα	Test_AutoEncoded: Δοκιμάζει
Autoencoder: Κάνει Encoding	τα αποτελέσματα

### InputData

Train,test data

### OutputData

Autoencoder.h5

DBSCAN.pkl

MiBatchKMeans.pkl

SpectralClust.pkl

DBSCANAE.pkl

MiBatchKMeansAE.pkl

SpectralClustAE.pkl

## Σκοπός εργασίας

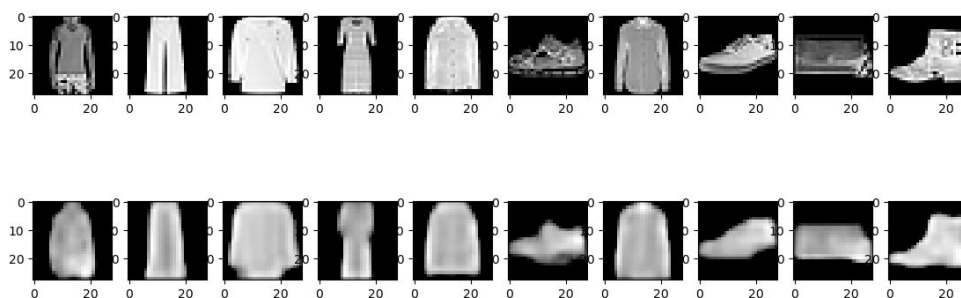
Η εργασία χωρίζεται σε δύο μέρη, όπου τα δεδομένα εκπαίδευσης στο πρώτο μέρος δεν έχουν επεξεργαστεί ενώ στο δεύτερο έχουν διαμορφωθεί με έναν encoder. Στο πέρας της εκτέλεσης συγκρίνουμε τα αποτελέσματα και καταλήγουμε στην προτιμότερη τεχνική clustering από τις 3 που χρησιμοποιήσαμε.

## Επεξήγηση κώδικα

**Raw:** Αρχικά διαβάζουμε τα δεδομένα και στη συνέχεια τα κανονικοποιούμε. Στη συνέχεια ρυθμίζουμε τις τεχνικές συσταδοποίησης και στη κάθε μία εφαρμόζουμε τα δεδομένα. Έπειτα εμφανίζουμε τα scores, ένα confusion matrix, και δειγματικές εικόνες(4) από αυτές που μάντεψε.

**Autoencoder:** Διαβάζει τα δεδομένα και τα δίνει σε ένα convolution δίκτυο με σκοπό την 'σμίκρυνση' τους έτσι ώστε να διατηρεί το ουσιώδες κομμάτι. Εκπαιδεύεται σε όλες τις εικόνες για 40 εποχές και αποθηκεύει το μοντέλο.

**Encoded:** Διαβάζει το μοντέλο που παραπάνω αποθηκεύσαμε και τα δεδομένα όπου στη συνέχεια μετατρέπει ο κωδικοποιητής. Εμφανίζει παραδειγματικές κωδικοποιημένες εικόνες και τις αντίστοιχες αρχικές. Στη συνέχεια ρυθμίζουμε τις τεχνικές συσταδοποίησης και στη κάθε μία εφαρμόζουμε τα δεδομένα. Έπειτα εμφανίζουμε τα scores, ένα confusion matrix, και δειγματικές εικόνες(4) από αυτές που μάντεψε.

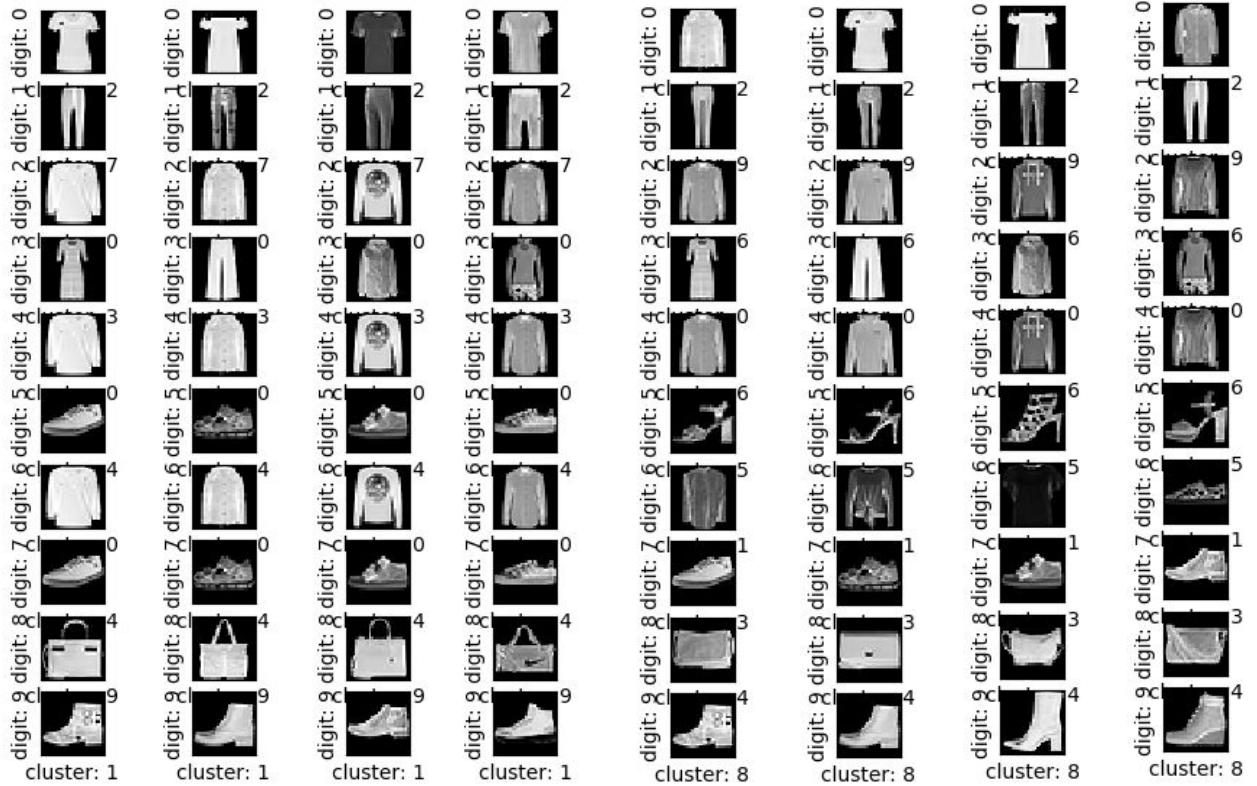


**Σημείωση:** Η τελευταία τεχνική δεν λειτουργεί καθόλου.

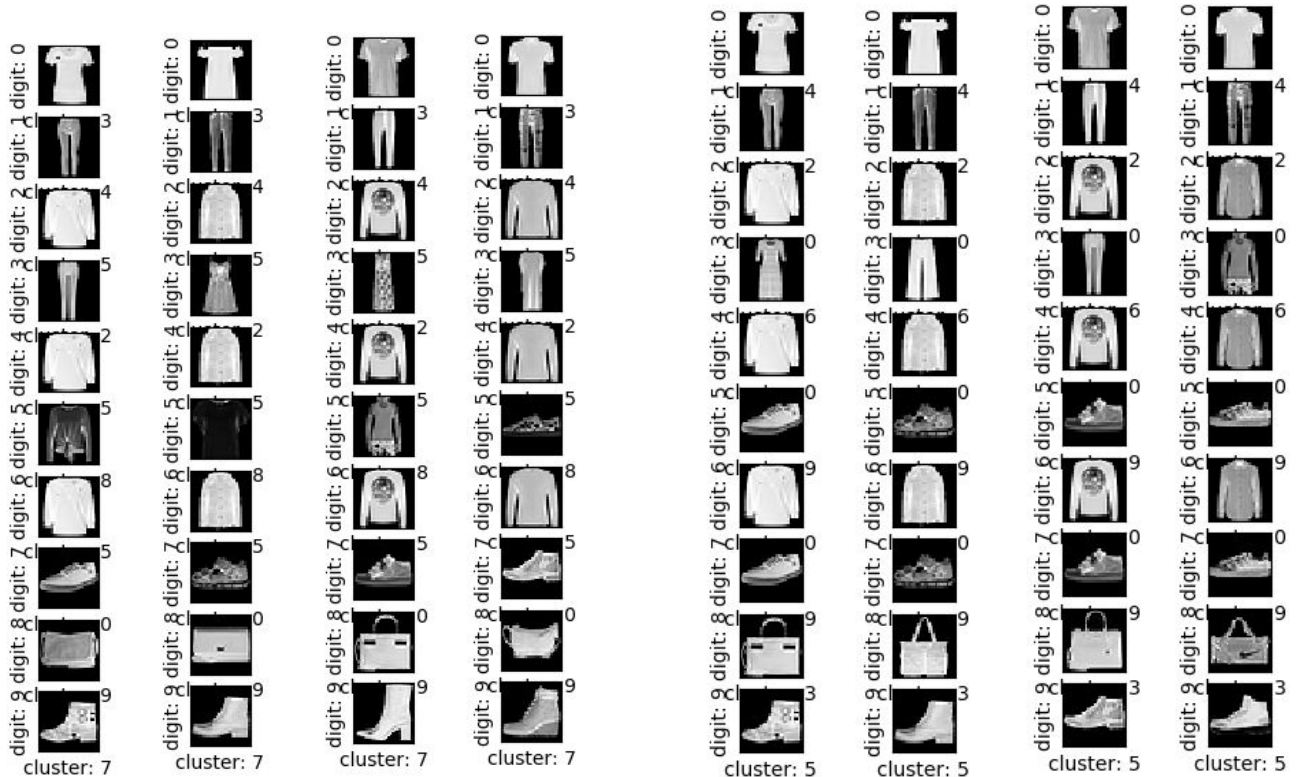


## Αποτελέσματα με εικόνες

Raw results:



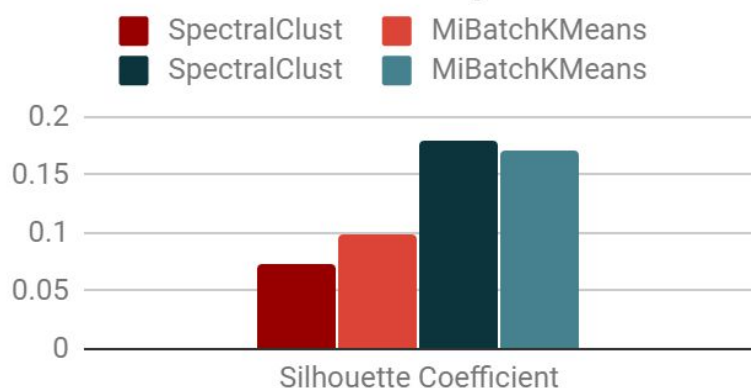
Encoded results:



## Αποτελέσματα

Τα κόκκινα αφορούν Raw data ενώ τα μπλέ encoded

### Silhouette Coeficiency



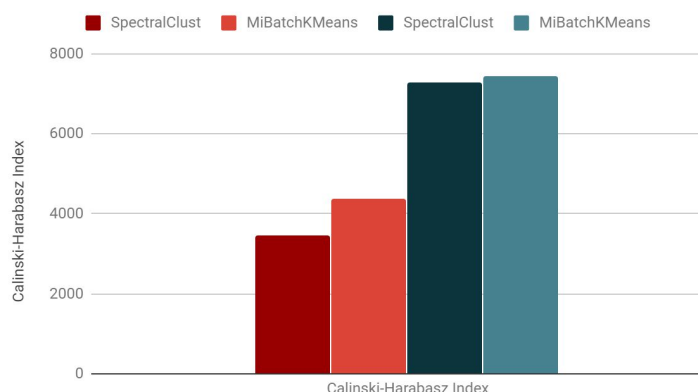
Silhouette Coefficient:

-1: incorrect 0: overlapping,

+1: highly dense clusters..

Πιο πυκνά ήταν αυτά που έχουν κωδικοποιηθεί.

### Calinski-Harabasz Index

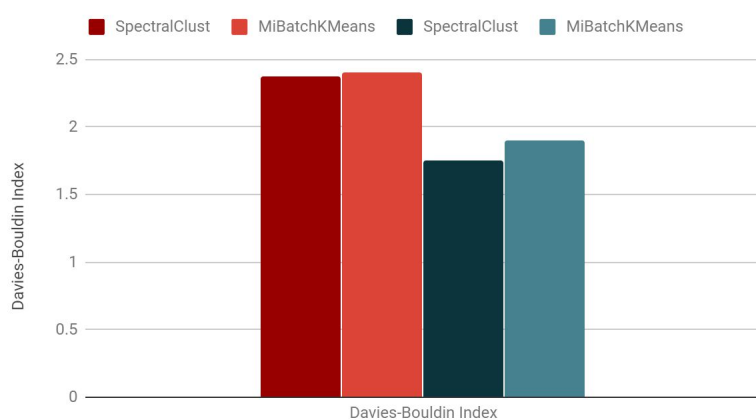


Calinski-Harabasz Index

Higher the value better the clusters

Επίσης καλύτερα clusters αποδίδει με κωδικοποίηση.

### Davies-Bouldin Index

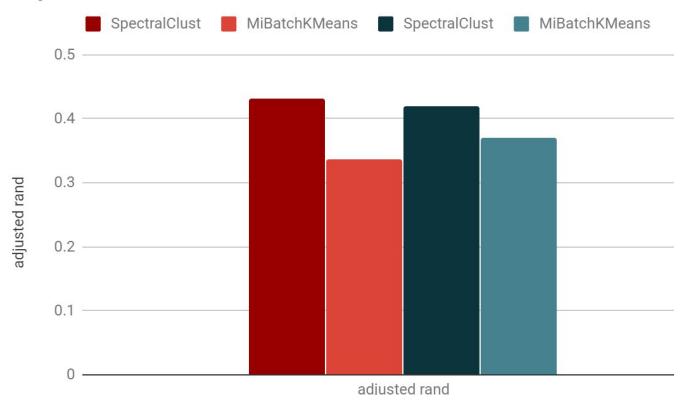


Davies-Bouldin Index

0: Lowest possible value, good partitioning.

Πιο μικρή τιμή έχουν πάλι τα κωδικοποιημένα.

### adjusted rand



Adjusted rand

Perfect labeling is scored 1.0

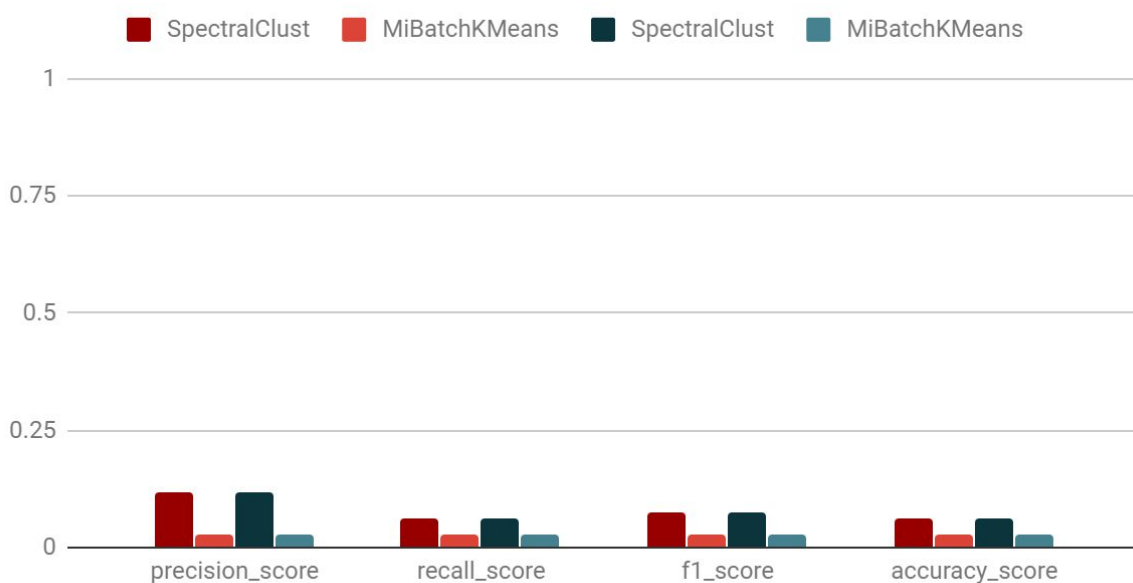
Bounded range [-1, 1]

Δείχνει σε σχέση με τα πραγματικά αποτελέσματα πως τα πήγε

## Απόδοση

Απο τα παραπάνω γραφήματα βλέπουμε πως τα κωδικοποιημένα δεδομένα έδωσε πολύ καλύτερα αποτελέσματα. Παρακάτω παρουσιάζεται γράφημα με τις αποδόσεις των συσταδοποιήσεων, όπου τα κόκκινα αντιπροσωπεύουν τα raw ενώ τα μπλε τα encoded. Εμφανής είναι η αποτυχία της εκπαίδευσης καθώς και η καλύτερη επίδοση του SpectralClust.

SpectralClust, MiBatchKMeans, SpectralClust and MiBatchKMeans



## Συμπεράσματα

Λαμβάνοντας υπόψη όλα τα προηγούμενα αποτελέσματα καταλήγουμε πως καλύτερες επιδόσεις είχε ο συνδυασμός encoded δεδομένων και της τεχνικής SpectralClust.