Unsupervised Clustering on Fashion-MNIST Dataset

Coursework Project

2025

Πίνακας περιεχομένων

Πίνακας εικόνων

1. Εισαγωγή

Στο πλαίσιο της παρούσας εργασίας, εξετάζεται το πρόβλημα της συσταδοποίησης (clustering) εικόνων από το σύνολο δεδομένων Fashion-MNIST, με χρήση τεχνικών μη επιβλεπόμενης μάθησης (unsupervised learning). Η ικανότητα ομαδοποίησης αντικειμένων χωρίς την ύπαρξη ετικετών (labels) αποτελεί σημαντικό εργαλείο σε πληθώρα εφαρμογών της τεχνητής νοημοσύνης και της υπολογιστικής όρασης.

Τα δεδομένα που αξιοποιούνται περιλαμβάνουν χιλιάδες εικόνες ενδυμάτων και υποδημάτων, οι οποίες ανήκουν σε δέκα διαφορετικές κατηγορίες. Η πρόκληση έγκειται στην ορθολογική αναπαράσταση των εικόνων σε μειωμένο χώρο χαρακτηριστικών, με στόχο τη βελτιστοποίηση των αποτελεσμάτων clustering και την ανάδειξη δομής στα δεδομένα.

Στόχος της εργασίας είναι:

- Να εφαρμοστούν δύο τεχνικές μείωσης διάστασης (Principal Component Analysis και Stacked Autoencoder)
- Να συγκριθούν δύο αλγόριθμοι συσταδοποίησης (MiniBatch KMeans και DBSCAN)
- Να αξιολογηθούν οι παραπάνω συνδυασμοί με βάση καθιερωμένες μετρικές ποιότητας clustering (Calinski–Harabasz index, Davies–Bouldin index, Silhouette score)
- Τέλος, να προταθεί ο καταλληλότερος συνδυασμός τεχνικών για την επίλυση του προβλήματος ομαδοποίησης στο συγκεκριμένο dataset

Η εργασία δομείται με τρόπο ώστε να παρουσιάζει με συστηματικό τρόπο τα πειραματικά αποτελέσματα, να συνοδεύεται από γραφήματα, εικόνες και πίνακες αξιολόγησης, και να προσφέρει σαφή συγκριτική ανάλυση των τεχνικών που εφαρμόστηκαν.

2. Περιγραφή dataset

Το σύνολο δεδομένων που χρησιμοποιείται στην παρούσα εργασία είναι το Fashion-MNIST, ένα ευρέως χρησιμοποιούμενο dataset εικόνων από τον χώρο της υπολογιστικής όρασης και της μηχανικής μάθησης.

Περιλαμβάνει συνολικά 70.000 εικόνες, εκ των οποίων:

- 60.000 εικόνες προορίζονται για εκπαίδευση (training set)
- 10.000 εικόνες για δοκιμή (test set)

Κάθε εικόνα:

- Έχει διαστάσεις 28x28 pixels
- Είναι μονόχρωμη (grayscale)
- Ανήκει σε μία από 10 προκαθορισμένες κατηγορίες (π.χ. μπλούζα, παπούτσι, τσάντα κ.ά.)
- Απεικονίζεται ως πίνακας επιπέδων φωτεινότητας, με τιμές κανονικοποιημένες στο διάστημα [0, 1].

Κατά την υλοποίηση, το σύνολο εκπαίδευσης διαχωρίστηκε περαιτέρω σε training και validation set, ώστε να χρησιμοποιηθεί για την εκπαίδευση νευρωνικών δικτύων (όπως ο autoencoder). Έτσι, η τελική διάσπαση των δεδομένων είναι:

- Training set: 48.000 δείγματα
- Validation set: 12.000 δείγματα
- Test set: 10.000 δείγματα

3. Προτεινόμενη Μεθοδολογία

Η μεθοδολογία που ακολουθήθηκε βασίζεται στη διαμόρφωση ενός πλήρους συστήματος μη επιβλεπόμενης μάθησης, το οποίο περιλαμβάνει προεπεξεργασία δεδομένων, μείωση διάστασης και συσταδοποίηση. Τα βασικά βήματα της διαδικασίας συνοψίζονται ως εξής:

1.Κανονικοποίηση και μετασχηματισμός εικόνων

Οι εικόνες κανονικοποιήθηκαν ώστε οι τιμές των pixels να βρίσκονται στο διάστημα [0, 1], και αναδιατάχθηκαν από 28x28 πίνακες σε διανύσματα 784 διαστάσεων.

2. Διαχωρισμός των δεδομένων

Τα δεδομένα χωρίστηκαν σε τρία σύνολα:

- Training set (48.000 δείγματα)
- Validation set (12.000 δείγματα)
- Test set (10.000 δείγματα)

Το validation set χρησιμοποιήθηκε αποκλειστικά για την εκπαίδευση του autoencoder.

3.Μείωση διάστασης (Dimensionality Reduction)

Υλοποιήθηκαν δύο διαφορετικές τεχνικές μείωσης διάστασης:

- Principal Component Analysis (PCA):Γραμμική τεχνική βασισμένη στην ανάλυση διακύμανσης των δεδομένων.
- Stacked Autoencoder (SAE):Βαθύ νευρωνικό δίκτυο τύπου autoencoder, το οποίο εκπαιδεύτηκε ώστε να συμπιέζει και να ανακατασκευάζει τις εικόνες. Μετά την εκπαίδευση, χρησιμοποιήθηκε μόνο το τμήμα του encoder για την εξαγωγή χαρακτηριστικών μειωμένης διάστασης.

4. Αλγόριθμοι συσταδοποίησης (Clustering)

Αξιολογήθηκαν δύο αλγόριθμοι clustering:

- MiniBatch KMeans: Παραλλαγή του KMeans με χρήση mini-batches για ταχύτερη εκπαίδευση.
- DBSCAN:Αλγόριθμος βάσει πυκνότητας, που δεν απαιτεί εκ των προτέρων καθορισμό αριθμού clusters.

Ο αλγόριθμος DBSCAN (Density-Based Spatial Clustering of Applications with Noise) επιλέχθηκε ως δεύτερη τεχνική συσταδοποίησης. Η επιλογή αυτή βασίστηκε στα εξής :

- Ανιχνεύει αυθαίρετου σχήματος clusters, σε αντίθεση με τον KMeans που βασίζεται σε σφαιρικές ομάδες γύρω από κέντρα, ο DBSCAN είναι κατάλληλος για πιο σύνθετες γεωμετρικές δομές, κάτι που είναι πιθανό σε δεδομένα εικόνας με πολύπλοκη κατανομή.
- Δεν απαιτεί εκ των προτέρων καθορισμό αριθμού clusters. Ιδιαίτερα χρήσιμο σε προβλήματα όπου η εσωτερική δομή των δεδομένων είναι άγνωστη. Στην περίπτωση του Fashion-MNIST, ενώ γνωρίζουμε ότι υπάρχουν 10 κατηγορίες, αυτό δεν σημαίνει απαραίτητα ότι τα clusters που προκύπτουν στον μειωμένο χώρο χαρακτηριστικών έχουν ίδια δομή ή πλήθος.
- Ανθεκτικότητα στον θόρυβο και ανίχνευση απομονωμένων σημείων (outliers). Το dataset μπορεί να περιέχει εικόνες που δεν εντάσσονται καλά σε κάποιο από τα φυσικά clusters. Ο DBSCAN έχει τη δυνατότητα να χαρακτηρίζει αυτά τα σημεία ως θόρυβο (`noise`), κάτι που βελτιώνει την ποιότητα των υπολοίπων ομάδων.

Για τους παραπάνω λόγους, ο DBSCAN αποτέλεσε μια συμπληρωματική, εναλλακτική προσέγγιση ως προς τον MiniBatch KMeans και χρησιμοποιήθηκε αποκλειστικά στα δεδομένα μειωμένης διάστασης (PCA και SAE).

5.Συνδυασμός τεχνικών & αξιολόγηση

Οι παραπάνω αλγόριθμοι εφαρμόστηκαν σε τρεις διαφορετικές εκδοχές των test δεδομένων:

- Raw data (με τιμές pixel)
- PCA-reduced data
- SAE-reduced data

Η απόδοση κάθε συνδυασμού αξιολογήθηκε με χρήση τριών μετρικών ποιότητας clustering:

- Calinski–Harabasz Index
- Davies–Bouldin Index
- Silhouette Score

Όλα τα πειράματα υλοποιήθηκαν σε περιβάλλον Google Colab, με χρήση της γλώσσας Python και των βιβλιοθηκών `scikit-learn`, `TensorFlow/Keras` και `pandas`

4. Πειραματικά Αποτελέσματα

Στην παρούσα ενότητα παρουσιάζονται τα αποτελέσματα των πειραμάτων που υλοποιήθηκαν, στοχεύοντας στη σύγκριση της αποδοτικότητας διαφορετικών συνδυασμών τεχνικών μείωσης διάστασης και αλγορίθμων συσταδοποίησης.

Συνδυασμοί που αξιολογήθηκαν

Οι παρακάτω πέντε (5) συνδυασμοί τεχνικών αξιολογήθηκαν ως προς τις μετρικές Calinski–Harabasz (CH), Davies–Bouldin (DB) και Silhouette Score:

Μείωση Διάστασης	Clustering
Καμία (Raw data)	MiniBatch KMeans
PCA	MiniBatch KMeans
SAE	MiniBatch KMeans
PCA	DBSCAN
SAE	DBSCAN

Όλες οι τεχνικές clustering εφαρμόστηκαν μόνο πάνω στο test set.

Πίνακας Μετρικών Απόδοσης

DimRed	Clustering	CH	DB	Silhouette	Clusters	DR Time (s)	Cluster Time (s)
Raw	MiniBatch KMeans	1240,08	1,99	0,14	10	0	0,31
PCA	MiniBatch KMeans	1486,8	1,94	0,17	10	3,18	0,04
SAE	MiniBatch KMeans	1238,12	2,05	0,12	10	735,08	0,04
PCA	DBSCAN	15,35	1,5	-0,34	35	3,18	1,91
SAE	DBSCAN	36,77	1,16	0,21	18	735,08	1,28

Τιμές αξιολόγησης για κάθε συνδυασμό τεχνικής μείωσης διάστασης (DimRed) και αλγορίθμου συσταδοποίησης (Clustering), βάσει των μετρικών Calinski–Harabasz, Davies–Bouldin και Silhouette.

Ο παραπάνω πίνακας συνοψίζει τα αποτελέσματα όλων των συνδυασμών τεχνικών μείωσης διάστασης και clustering που αξιολογήθηκαν στην παρούσα εργασία. Οι μετρικές Calinski–Harabasz (CH), Davies–Bouldin (DB) και Silhouette παρέχουν συμπληρωματικές πληροφορίες για την ποιότητα των παραγόμενων clusters. Ο πίνακας αυτός αποτελεί βασικό εργαλείο συγκριτικής αξιολόγησης για την εξαγωγή συμπερασμάτων ως προς την αποτελεσματικότητα και αποδοτικότητα των τεχνικών που εφαρμόστηκαν.

Γραφικές παραστάσεις & παρατηρήσεις

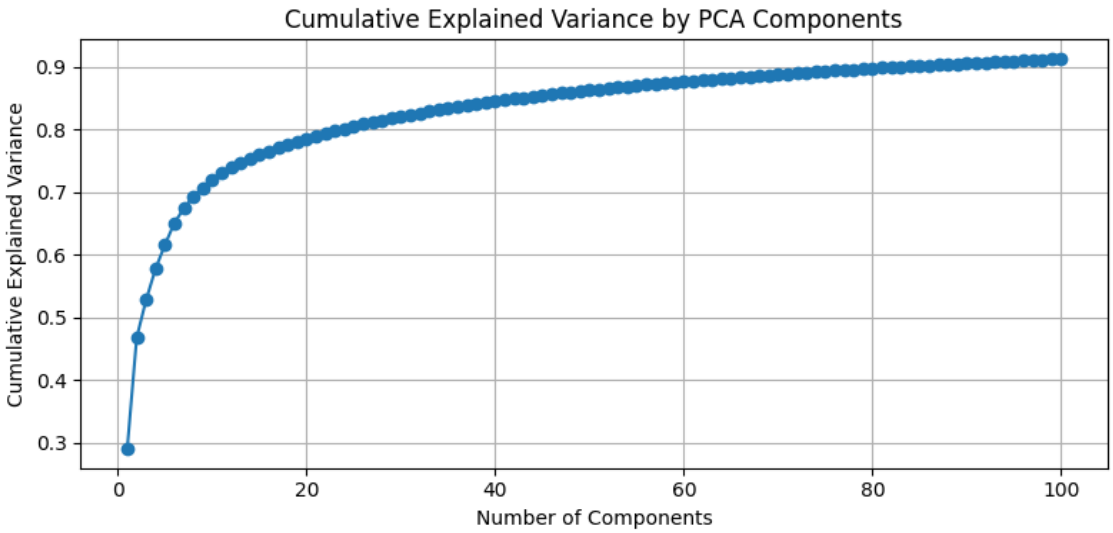

Figure 1: Διάγραμμα σωρευτικής εξηγούμενης διακύμανσης των κύριων συνιστωσών (PCA).

Το Σχήμα 1 παρουσιάζει τη σωρευτική εξηγούμενη διακύμανση των πρώτων 100 κύριων συνιστωσών της PCA. Παρατηρούμε ότι οι πρώτες ~20 συνιστώσες εξηγούν πάνω από το 70% της συνολικής διακύμανσης, ενώ οι πρώτες 100 συνιστώσες ξεπερνούν το 90%. Το αποτέλεσμα αυτό επιβεβαιώνει ότι η PCA αποτελεί κατάλληλη τεχνική μείωσης διάστασης για το συγκεκριμένο πρόβλημα, καθώς διατηρεί το μεγαλύτερο μέρος της πληροφορίας με πολύ λιγότερες διαστάσεις από τις αρχικές (784).

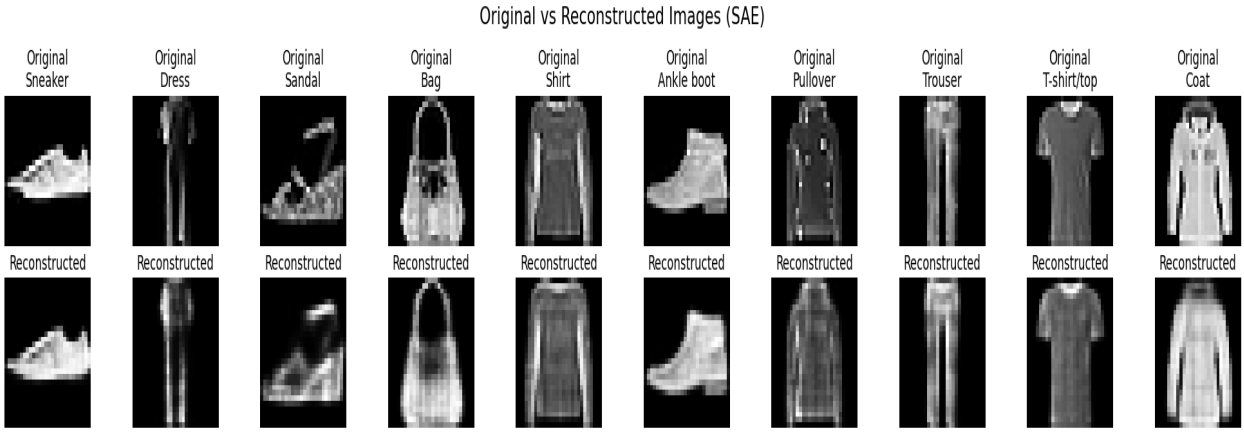

Figure 2 : Σύγκριση αρχικών και ανακατασκευασμένων εικόνων του Fashion-MNIST μέσω του Stacked Autoencoder. Από κάθε κλάση επιλέχθηκε μία εικόνα.

Το Σχήμα 2 παρουσιάζει την απόδοση του Stacked Autoencoder, μέσω της οπτικής σύγκρισης αρχικών και ανακατασκευασμένων εικόνων. Για καθεμία από τις δέκα κατηγορίες του Fashion-MNIST, απεικονίζεται η αρχική εικόνα (πάνω σειρά) και η αντίστοιχη ανακατασκευή (κάτω σειρά). Παρατηρούμε ότι ο SAE καταφέρνει να διατηρήσει τα βασικά μορφολογικά χαρακτηριστικά των εικόνων, παρότι τα δεδομένα έχουν περάσει από συμπίεση και αποκωδικοποίηση. Αυτό υποδηλώνει ότι η τεχνική κατάφερε να μάθει ουσιαστική αναπαράσταση του χώρου χαρακτηριστικών, στοιχείο κρίσιμο για την εφαρμογή clustering που ακολούθησε μετά.

Μετά την αξιολόγηση της ποιότητας ανακατασκευής του SAE, παρουσιάζονται στη συνέχεια τα διαγράμματα σύγκρισης των πέντε συνδυασμών τεχνικών ως προς τις τρεις βασικές μετρικές αξιολόγησης clustering: Calinski–Harabasz, Davies–Bouldin και Silhouette Score. Τα διαγράμματα αυτά επιτρέπουν την άμεση οπτική σύγκριση της απόδοσης κάθε μεθόδου και αποτελούν τη βάση για τη διατύπωση των τελικών συμπερασμάτων.

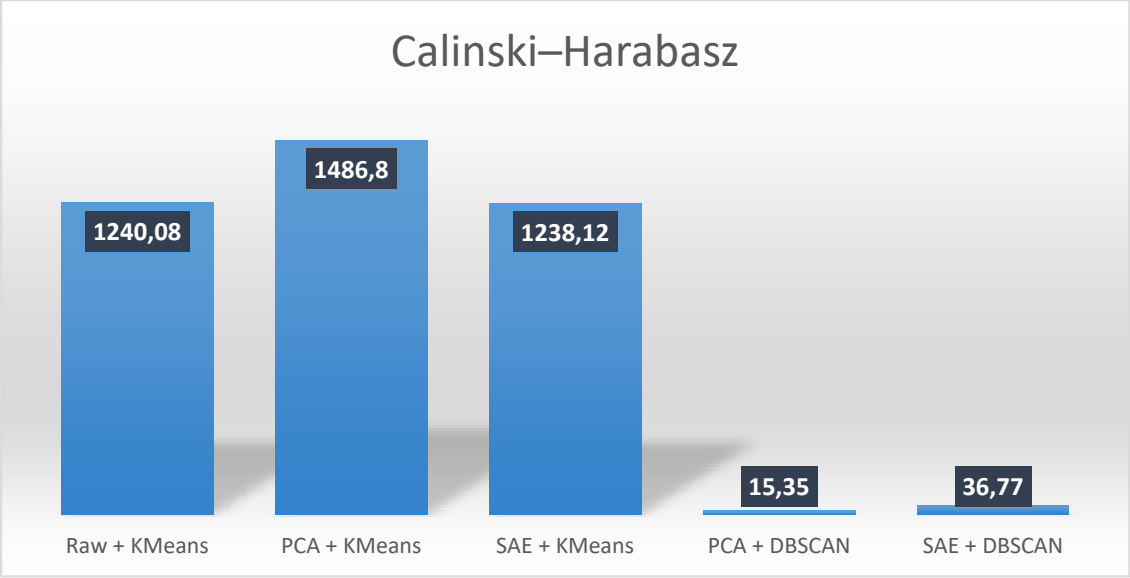

Figure 3: Τιμές του δείκτη Calinski–Harabasz για καθέναν από τους πέντε συνδυασμούς τεχνικών μείωσης διάστασης και clustering.

Το γράφημα του Σχήματος 3 παρουσιάζει τις τιμές του δείκτη Calinski–Harabasz για τους πέντε συνδυασμούς τεχνικών που αξιολογήθηκαν. Ο δείκτης αυτός μετρά την πυκνότητα και τον διαχωρισμό των clusters, με υψηλότερες τιμές να υποδηλώνουν καλύτερη απόδοση. Όπως φαίνεται, ο συνδυασμός PCA + MiniBatch KMeans πέτυχε τη μέγιστη τιμή (1486.8), ακολουθούμενος από τις μεθόδους Raw + KMeansnκαι SAE + KMeans με πολύ παρόμοιες επιδόσεις (~1240). Οι δύο συνδυασμοί με DBSCAN (σε PCA και SAE) σημείωσαν πολύ χαμηλότερες τιμές, γεγονός που υποδεικνύει μειωμένο διαχωρισμό μεταξύ των ομάδων.

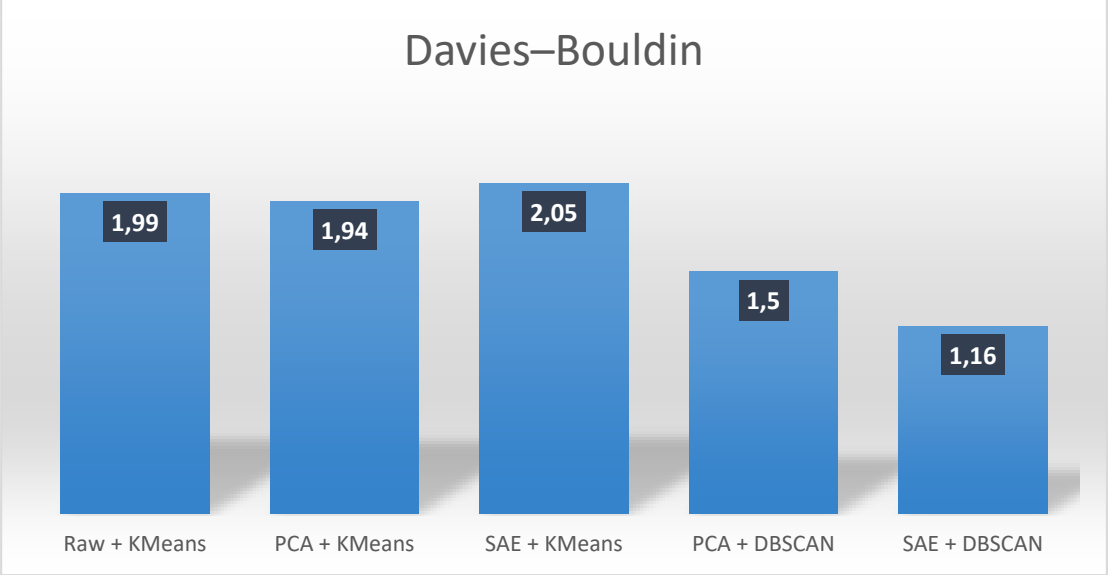

Figure 4: Τιμές του δείκτη Davies–Bouldin για όλους τους συνδυασμούς clustering.

Το γράφημα του Σχήματος 4 απεικονίζει τις τιμές του δείκτη Davies–Bouldin (DB), ο οποίος εκτιμά τη σχέση μεταξύ ενδοομαδικής συνοχής και μεταομαδικής απόστασης. Σε αντίθεση με άλλες μετρικές, όσο χαμηλότερη είναι η τιμή, τόσο καλύτερη θεωρείται η ποιότητα της συσταδοποίησης. Από το γράφημα παρατηρούμε ότι ο συνδυασμός SAE + DBSCAN επιτυγχάνει τη χαμηλότερη τιμή (1.16), γεγονός που υποδηλώνει αυξημένη αποδοτικότητα ως προς τη διαχωρισιμότητα των clusters. Το PCA + DBSCAN ακολουθεί με τιμή 1.50. Αντίθετα, οι συνδυασμοί με MiniBatch KMeans (ιδιαίτερα SAE + KMeans) έχουν υψηλότερες τιμές, με κορυφαία τη 2.05.

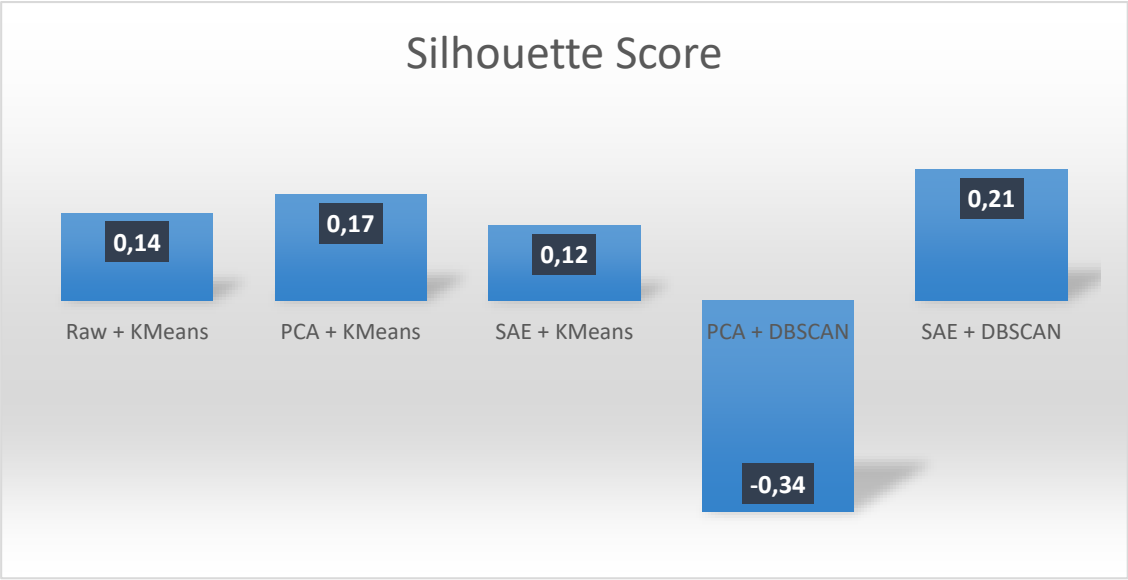

Figure 5: Τιμές του δείκτη Silhouette για καθέναν από τους πέντε συνδυασμούς clustering.

Το γράφημα του Σχήματος 5 απεικονίζει τις τιμές του Silhouette Score, μιας μετρικής που αξιολογεί πόσο καλά βρίσκεται ένα δείγμα εντός του cluster του συγκριτικά με τα γειτονικά clusters. Οι τιμές κυμαίνονται θεωρητικά από -1 έως +1, με υψηλότερες τιμές να υποδηλώνουν καλύτερο διαχωρισμό και συνεκτικότητα των ομάδων. Από τα αποτελέσματα προκύπτει ότι ο συνδυασμός SAE + DBSCAN είχε την υψηλότερη τιμή Silhouette (0.21), γεγονός που δείχνει ότι τα παραγόμενα clusters είναι σχετικά καλά διαχωρισμένα. Οι υπόλοιποι συνδυασμοί με MiniBatch KMeans κυμάνθηκαν μεταξύ 0.12 και 0.17, με το PCA + KMeans να έχει τη σχετική υπεροχή (0.17). Αξιοσημείωτο είναι το γεγονός ότι ο συνδυασμός PCA + DBSCAN εμφάνισε αρνητικό Silhouette Score (-0.34), κάτι που δείχνει αδύναμο διαχωρισμό των clusters ή κακή δομή των ομάδων στον χώρο χαρακτηριστικών. Το συγκεκριμένο εύρημα αναδεικνύει τη σημασία της επιλογής κατάλληλου συνδυασμού μεταξύ μεθόδου μείωσης διάστασης και αλγορίθμου clustering.

Αφού παρουσιάστηκαν και αναλύθηκαν οι ποσοτικές μετρικές αξιολόγησης για όλους τους συνδυασμούς τεχνικών clustering, ακολουθεί μια ποιοτική απεικόνιση των αποτελεσμάτων από δύο διαφορετικά clusters που προέκυψαν από τον συνδυασμό PCA + MiniBatch KMeans, με στόχο την οπτική εκτίμηση της εσωτερικής ομοιογένειας των ομάδων.

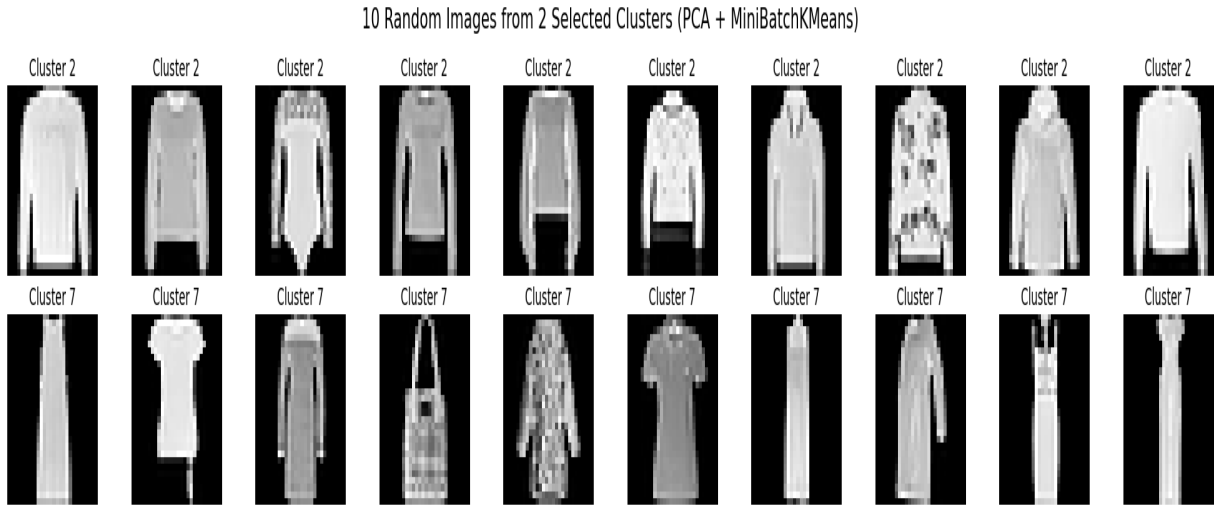

Figure 6 : Τυχαία δείγματα εικόνων από δύο επιλεγμένα clusters που προέκυψαν από τον συνδυασμό PCA + MiniBatchKMeans.

Το Σχήμα 6 απεικονίζει τυχαία επιλεγμένες εικόνες από δύο clusters του test set, όπως προέκυψαν από την εφαρμογή της μεθόδου PCA σε συνδυασμό με τον αλγόριθμο MiniBatch KMeans. Για κάθε ένα από τα δύο clusters εμφανίζονται δέκα εικόνες, οι οποίες αποδεικνύουν οπτικά τη συνοχή των ομάδων που σχηματίστηκαν. Όπως παρατηρείται, οι εικόνες εντός του κάθε cluster παρουσιάζουν ομοιομορφία ως προς την κατηγορία ή τα χαρακτηριστικά του απεικονιζόμενου αντικειμένου, στοιχείο που ενισχύει την ποιοτική αξιολόγηση της τεχνικής clustering πέραν των αριθμητικών μετρικών.

5. Συμπεράσματα

Η παρούσα εργασία υλοποίησε και αξιολόγησε μια σειρά από συνδυασμούς τεχνικών μείωσης διάστασης και συσταδοποίησης, με στόχο την ομαδοποίηση των εικόνων του dataset Fashion-MNIST χωρίς τη χρήση ετικετών. Συγκεκριμένα, αξιολογήθηκαν οι τεχνικές PCA και Stacked Autoencoder για τη μείωση διάστασης, σε συνδυασμό με τους αλγορίθμους MiniBatch KMeans και DBSCAN.

Αξιολόγηση βάσει μετρικών απόδοσης

Από την ανάλυση των αποτελεσμάτων προκύπτουν τα εξής βασικά συμπεράσματα:

- Ο συνδυασμός PCA + MiniBatch KMeans εμφάνισε την καλύτερη συνολική απόδοση ως προς τον δείκτη Calinski–Harabasz, καλή επίδοση στο Silhouette Score και ιδιαίτερα χαμηλό χρόνο εκτέλεσης, καθιστώντας τον αποτελεσματικό και αποδοτικό.
- Ο SAE + DBSCAN παρουσίασε την καλύτερη ποιότητα clustering σύμφωνα με τον Davies–Bouldin και τον Silhouette Score, όμως είχε πολύ αυξημένο υπολογιστικό κόστος λόγω του autoencoder.
- Οι συνδυασμοί με DBSCAN απέδωσαν μεταβλητά αποτελέσματα, με τον PCA + DBSCAN να έχει αρνητικό Silhouette Score, υποδεικνύοντας αδυναμία σωστής ομαδοποίησης στον συγκεκριμένο χώρο χαρακτηριστικών.

Ο συνδυασμός PCA + MiniBatch KMeans φαίνεται να αποτελεί μια καλή λύση για το συγκεκριμένο πρόβλημα, επιτυγχάνοντας καλή ισορροπία μεταξύ απόδοσης, διαχωρισιμότητας clusters και υπολογιστικού κόστους. Αντίθετα, αν δοθεί έμφαση αποκλειστικά στην ποιότητα της ομαδοποίησης, ο SAE + DBSCAN θα μπορούσε να θεωρηθεί εναλλακτική επιλογή, με το κόστος αυξημένου χρόνου επεξεργασίας.

Κανένας συνδυασμός δεν υπερίσχυσε και στις τρεις μετρικές. Το γεγονός αυτό αναδεικνύει μια βασική πραγματικότητα των προβλημάτων μη επιβλεπόμενης μάθησης: η απόδοση των αλγορίθμων εξαρτάται σε μεγάλο βαθμό από το ποια μετρική θεωρείται σημαντικότερη, αλλά και από τους περιορισμούς και τις ανάγκες της εκάστοτε εφαρμογής.

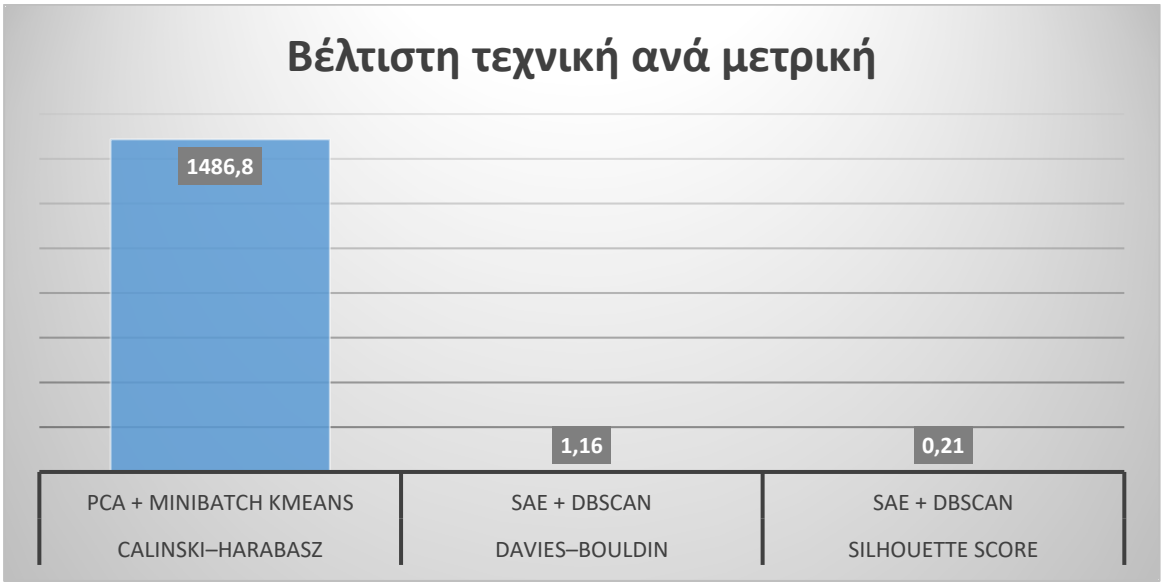

Figure 7: Βέλτιστος συνδυασμός τεχνικών clustering για κάθε μία από τις τρεις μετρικές αξιολόγησης: Calinski–Harabasz, Davies–Bouldin και Silhouette.

Το Σχήμα 7 απεικονίζει ποια τεχνική clustering σημείωσε την καλύτερη επίδοση για καθεμία από τις τρεις μετρικές αξιολόγησης: Calinski–Harabasz (CH), Davies–Bouldin (DB) και Silhouette Score. Όπως φαίνεται, ο συνδυασμός PCA + MiniBatch KMeans υπερίσχυσε καθαρά στον δείκτη CH, που μετρά τον διαχωρισμό και την συνοχή των clusters, ενώ ο συνδυασμός SAE + DBSCAN εμφάνισε τις καλύτερες τιμές τόσο για τον DB όσο και για το Silhouette Score. Η γραφική αυτή απεικόνιση ενισχύει το γενικό συμπέρασμα της μελέτης, ότι δεν υπάρχει ένας απόλυτα βέλτιστος συνδυασμός σε όλες τις μετρικές. Κάθε τεχνική προσφέρει πλεονεκτήματα ανάλογα με τη μετρική που δίνεται έμφαση, και τελικά η επιλογή πρέπει να βασίζεται στις προτεραιότητες του εκάστοτε συστήματος (π.χ. ποιότητα clustering ή αποδοτικότητα).

Ποιοτική αξιολόγηση μέσω εικόνων

Πέρα από τις αριθμητικές μετρικές αξιολόγησης, ιδιαίτερη σημασία είχε και η ποιοτική αποτίμηση των αποτελεσμάτων μέσω της απεικόνισης παραδειγματικών εικόνων από τα clusters και της σύγκρισης αρχικών με ανακατασκευασμένων εικόνων από τον autoencoder. Η ανακατασκευή των εικόνων μέσω του SAE έδειξε ότι το δίκτυο ήταν ικανό να διατηρήσει τα βασικά οπτικά χαρακτηριστικά των αντικειμένων, επιβεβαιώνοντας την αποτελεσματικότητα της μείωσης διάστασης.

Επιπλέον, η παρουσίαση εικόνων από δύο επιλεγμένα clusters που προέκυψαν από τον συνδυασμό PCA + MiniBatch KMeans αποκάλυψε σαφή εσωτερική συνοχή: τα αντικείμενα εντός κάθε ομάδας ως επί το πλείστον είχαν παρόμοιο σχήμα και κατηγορία, ενισχύοντας τη θετική ποιοτική αξιολόγηση της συγκεκριμένης μεθόδου. Αυτή η οπτική συνέπεια μεταξύ των μελών κάθε cluster αποτελεί καλή ένδειξη για την επιτυχία του αλγορίθμου, και λειτουργεί συμπληρωματικά με τις μετρικές απόδοσης. Η ενσωμάτωση των εικόνων αυτών στην ανάλυση υπογραμμίζει την αναγκαιότητα της πολυδιάστατης αξιολόγησης αριθμητικής και οπτικής σε προβλήματα μη επιβλεπόμενης μάθησης.

6. Βιβλιογραφία

[1] Fashion-MNIST dataset (Keras). Διαθέσιμο στο:
https://keras.io/api/datasets/fashion_mnist/

[2] scikit-learn: Machine Learning in Python. Επίσημη τεκμηρίωση:

https://scikit-learn.org/stable/

[3] PCA – Principal Component Analysis (scikit-learn).

https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

[4] DBSCAN – Density-Based Spatial Clustering (scikit-learn).

https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

[5] MiniBatchKMeans (scikit-learn).

https://scikit-learn.org/stable/modules/generated/sklearn.cluster.MiniBatchKMeans.html

[6] TensorFlow Keras – Autoencoder Guide.

https://blog.keras.io/building-autoencoders-in-keras.html

[7] matplotlib – Visualization library in Python. **https://matplotlib.org/**

[8] pandas – Data Analysis library. **https://pandas.pydata.org/**

[9] NumPy – Scientific Computing with Python. **https://numpy.org/**

[10] Python documentation. **https://docs.python.org/3/**