**Γειά σας, ονομάζομαι γιαννης μπουντουριδης** και θα σας παρουσιάσω την διπλωματικη μου εργασια με θέμα: Δημιουργια γενικευμενου συστηματος προτασεων με εφαρμογη σε συνολα αλληλεξαρτωμενων δεδομενων. **-> ->Ξεκινώντας, θα ήθελα** να αναφέρω κάποια εισαγωγικά στοιχεία για τα συστήματα προτάσεων που αποτέλεσαν αντικείμενο αυτής της εργασίας. **->** **Τα συστήματα προτάσεων χρησιμοποιούνται ευρέως από μεγάλες εταιρίες του διαδικτύου.** Χρησιμοποιούνται από το Youtube για να εμφανίζουν βίντεο τα οποία θεωρούν ότι είναι αρκετά πιθανό να μας αρέσουν. Χρησιμοποιούνται από το facebook για να εμφανίζουν αναρτήσεις φίλων και συναδέλφων για τις οποίες θεωρούν ότι είναι αρκετα πιθανό να μας ενδιαφέρουν. Χρησιμοποιούνται από το Spotify για να εμφανίζουν τραγούδια ή ακόμη και από το amazon για να εμφανίζουν προϊόντα τα οποία θεωρούν ότι είναι αρκετά πιθανό να πωληθούν. Γενικός τους στόχος είναι να αναγνωρίζουν τις προτιμήσεις κάθε χρήστη μέσα από την δραστηριότητα του. Αφού το κάνουν αυτό, βρίσκουν προϊόντα ή υπηρεσίας που είναι αρκετά πιθανό να αρέσουν μέσα **από τεράστιους όγκους δεδομένων.** **->Συνεχίζοντας στην ενότητα στόχοι και μεθοδολογία**, **->** **θα ήθελα να σας αναφέρω πως στόχος της διπλωματικής εργασίας** είναι η δημιουργία ενός γενικευμένου συστήματος προτάσεων που θα τροφοδοτείτε με δεδομένα και θα εξάγει προτάσεις. Η διπλωματική εργασία περιλαμβάνει δεδομένα εισόδου με ρούχα για την παροχή προτάσεων σε σχεδιαστές και δεδομένα εισόδου με ταινίες για την παροχή προτάσεων σε τηλεθεατές. **->** **Για την μεθοδολογία της διπλωματικής τα στάδια** για την δημιουργία του τελικού συστήματος προτάσεων ένδυσης και του τελικού συστήματος προτάσεων ταινιών είναι: 1) η συγκέντρωση και προ-επεξεργασία των δεδομένων 2) ο διαχωρισμός των δεδομένων σε δεδομένα εκπαίδευσης και ελέγχου 3) η δημιουργία των συστημάτων προτάσεων 4) η συνεργασία των συστημάτων προτάσεων μέσω ενός υβριδικού συστήματος και η επιστημονική τεκμηρίωση της ορθής λειτουργίας του συστήματος με χρήση ορισμένων μετρικών. **->Ακολουθούν οι περιγραφές των υλοποιήσεων**. **-> Η πρώτη υλοποίηση χρησιμοποιεί την τεχνική ανάλυσης πίνακας σε ιδιάζουσες τιμές**. Αρχικά υπάρχει ένας πίνακας ο οποίος είναι διαστάσεων M X N όπου Μ είναι οι χρήστες και Ν τα προϊόντα ( ρούχα ή ταινίες ). Κάθε θέση του αρχικού πίνακα περιλαμβάνει μια τιμή η οποία εκφράζει την αξιολόγηση του χρήστη σε αυτό το προϊόν. Αν ο χρήστης δεν έχει αξιολογήσει κάποιο προϊόν η αντίστοιχη θέση που αναλογεί στον πίνακα παραμένει κενή. Ο πίνακας αυτός αναλύεται σε 3 νέους πίνακες. Ο πρώτος περιλαμβάνει Μ σειρές και k στήλες, ο δεύτερος k σειρές και k στήλες και ο τρίτος k σειρές και N στήλες. Η ανακατασκευή του αρχικού πίνακα κρατώντας τις 𝑘 μεγαλύτερες ιδιάζουσες τιμές, εξασφαλίζει τη διατήρηση των πιο σημαντικών παραγόντων, προβλέπει βαθμολογίες σε αβαθμολόγητα προϊόντα και δημιουργεί αξιόπιστες προτάσεις. **-> Η δεύτερη υλοποίηση** **είναι ένα μοντέλο νευρωνικού συνεργατικού φιλτραρίσματος.** Το πρώτο επίπεδο αυτού του μοντέλου δέχεται ως είσοδο δύο διανύσματα. Το πρώτο διάνυσμα είναι ενός χρήστη και περιλαμβάνει όλες τις αξιολογήσεις που έχει κάνει σε κάποια προϊόντα. Το δεύτερο διάνυσμα είναι ενός προϊόντος καιπεριλαμβάνει ολες τις αξιολογήσεις τις οποίες έχει δεχθεί από ορισμένους χρήστες. Τόσο το διάνυσμα του χρήστη όσο και το διάνυσμα του προϊόντος εισέρχονται σε ενσωματομένα επίπεδα τα οποία έχουν στόχο να μειώσουν τις τεράστιες διαστάσεις αυτών των διανυσμάτων και να κρατήσουν τους k πιο σημαντικούς παράγοντες. Τα χαρακτηριστικά αυτά συνενώνονται και εισέρχονται σε ένα πολύ-επίπεδο νευρωνικό δίκτυο το οποίο έχει στόχο να αναγνωρίσει τις αλληλεπιδράσεις μεταξύ χρηστών και προϊόντων. Μετά την αναγνώριση αυτών των αλληλεπιδράσεων, εκτιμάται η βαθμολογία που θα έκανε ο χρήστης σε αυτό το προϊόν. Αν η βαθμολογία αυτή είναι υψηλή τότε το προϊόν προτείνεται από το σύστημα. **-> Η Τρίτη υλοποίηση, είναι ένα μοντέλο φιλτραρίσματος περιεχομένου με τυχαία δάση**. Το πρώτο επίπεδο αυτού του μοντέλου, δέχεται ως είσοδο τις βαθμολογίες και τα χαρακτηριστικά των προϊόντων για ένα συγκεκριμένο χρήστη. Το μοντέλο εκπαιδεύεται με τα διαθέσιμα δεδομένα. Με την βοήθεια της βιβλιοθήκης optuna, κατά την διάρκεια της εκπαίδευσης, ρυθμίζονται οι παράμετροι των τυχαίων δασών. Μετά από αυτή την διαδικασία, το σύστημα αναγνωρίζει τις προτιμήσεις του χρήστη και προβλέπει σε μη βαθμολογημένα προϊόντα αξιολογήσεις ώστε να εξάγει προτάσεις. **-> Αντίστοιχα, η τέταρτη υλοποίηση έχει ίδια είσοδο με την προηγούμενη.** Αυτή τη φορά, η αναγνώριση των προτιμήσεων του χρήστη δε γίνεται με τυχαία δάση αλλά με ένα πολυεπίπεδο νευρωνικό δίκτυο. Η διαδικασία κατά τ’ άλλα είναι πανομοιότυπη. **->** Είναι φυσιολογικό, κάποιες υλοποιήσεις να παρέχουν καλύτερες προτάσεις από άλλες. Γι αυτό, γίνεται η χρήση ενός υβριδικού συστήματος και με την βοήθεια της γραμμικής παλινδρόμησης κάθε σύστημα έχει διαφορετικό βάρος στην ψήφο του. **->**  **Πειραματικα αποτελέσματα.. -> Για τα πειραματικά αποτελέσματα χρησιμοποιήθηκαν 3 σετ δεδομένων.** Τα πρώτα δύο περιλαμβάνουν 1700 ρούχα, 100 χρήστες με 50 και 100 χιλιάδες αξιολογήσεις αντίστοιχα. Έχουν κλίμακα αξιολογήσεων από μηδέν έως δέκα και περιλαμβάνουν 10 κατηγορίες ρούχων. Το τρίτο σύνολο δεδομένων περιλαμβάνει 9000 ταινίες με 600 χρήστες και 100 χιλιάδες αξιολογήσεις. Έχει κλίμακα αξιολογήσεων από μηδέν έως πέντε και περιλαμβάνει 11 κατηγορίες ταινιών. **-> Τα πειραματικά αποτελέσματα για το σύνολο δεδομένων asos30 δείχνουν**, ότι καλύτερες επιδόσεις τόσο σε θέματα ταξινόμησης όσο και σε θέματα σφάλματος εκτίμησης, πετυχαίνει το υβριδικό σύστημα γραμμικής παλινδρόμησης. Σε θέματα ταξινόμησης το σύστημα πετυχαίνει το υψηλότερο accuracy 81%, το δεύτερο υψηλότερο recall 98%, το υψηλότερο precision 82% και το υψηλότερο score f1\_score 89%. Σε θέματα σφάλματος το σύστημα πετυχαίνει το χαμηλότερο root mean square error 2.29 και το χαμηλότερο mean average error 1.83. **->** Τα πειραματικά αποτελέσματα για το σύνολο δεδομένων asos60 δείχνουν κι αυτά, ότι καλύτερες επιδόσεις τόσο σε θέματα ταξινόμησης όσο και σε θέματα σφάλματος εκτίμησης, πετυχαίνει το υβριδικό σύστημα γραμμικής παλινδρόμησης. Σε θέματα ταξινόμησης το σύστημα πετυχαίνει το υψηλότερο accuracy 84%, το δεύτερο υψηλότερο recall 97%, το υψηλότερο precision 84% και το υψηλότερο score f1\_score 90%. Σε θέματα σφάλματος το σύστημα πετυχαίνει το χαμηλότερο root mean square error 1.97 και το χαμηλότερο mean average error 1.52. Παρατηρούμε πως συγκριτικά με το Asos30 η αύξηση των δεδομένων βελτίωσε τις επιδόσεις. **->** Τα πειραματικά αποτελέσματα για το σύνολο δεδομένων movielens100k είναι λίγο πιο περίπλοκα. Σε θέματα ταξινόμησης καλύτερη υλοποίηση είναι το φιλτράρισμα περιεχομένου με νευρωνικά δίκτυα. Η υλοποίηση πετυχαίνει υψηλότερο accuracy 88%, το υψηλότερο recall 96%, το υψηλότερο precision 90% και το υψηλότερο score f1\_score 92% Σε θέματα σφάλματος εκτίμησης καλύτερη υλοποίηση είναι το το υβριδικό σύστημα γραμμικής παλινδρόμησης. Η υλοποίηση πετυχαίνει το χαμηλότερο root mean square error 0,82 και το χαμηλότερο mean average error 0.67. **-> Ένα σημαντικό κομμάτι που πρέπει να ελέγχουμε στα συστήματα προτάσεων είναι το αν υπάρχει ισορροπία μεταξύ των προϊόντων που απορρίπτονται και των προϊόντων που προτείνονται.** Το διάγραμμα της κάλυψης μας δείχνει και στα 3 σύνολα δεδομένων μας δείχνει πως ισορροπία υπάρχει. Για το σύνολο asos30 υπάρχει 31 % των προϊόντων το οποίο είναι ήδη αξιολογημένο, 41% των προϊόντων το οποίο προτείνεται από το σύστημα και 27% το οποίο απορρίπτεται. Για το σύνολο asos60 υπάρχει 60 % των προϊόντων το οποίο είναι ήδη αξιολογημένο, 23% των προϊόντων το οποίο προτείνεται από το σύστημα και 16% το οποίο απορρίπτεται. Τέλος για το movielens-100k υπάρχει 2 % των προϊόντων το οποίο είναι ήδη αξιολογημένο, 50% των προϊόντων το οποίο προτείνεται από το σύστημα και 48% το οποίο απορρίπτεται. Ένα τελευταίο κομμάτι που πρέπει να προσέχουμε στα συστήματα προτάσεων είναι να μην υπάρχει κάποια κατηγορία του συστήματος απορρίπτεται πλήρως. Αν για παράδειγμα από την κατηγορία μαλλίνα ρούχα στο asos30 προτείνονταν μονο 3% τότε οι σχεδιαστές δε θα βλέπαν ποτέ μαλλινα ρούχα από το σύστημα. Όπως βλέπουμε όμως κάτι τέτοιο δε συμβαίνει και για τα 3 σύνολα **->**