

Πλάνο έργου λογισμικού, για την ανάπτυξη ολοκληρωμένου  
συστήματος σύστασης ταινιών με χρήση μηχανικής μάθησης

## Περιεχόμενα

ΕΝΟΤΗΤΑ 1: ΟΡΓΑΝΩΣΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΤΟΥ ΕΡΓΟΥ.....	3
1. Εισαγωγή .....	3
1.1 Τρέχουσα κατάσταση.....	3
1.2 Προηγούμενες ενέργειες .....	4
1.3 Στόχοι του συστήματος .....	4
1.4 Κατηγορίες χρηστών .....	5
1.5 Επισκόπηση του συστήματος .....	5
1.6 Ανάλυση SWOT.....	7
ΕΝΟΤΗΤΑ 2. ΟΡΓΑΝΩΣΗ ΤΟΥ ΕΡΓΟΥ .....	8
2.1 Μέθοδος ανάπτυξης.....	8
2.2 Προσωπικό, ρόλοι και υπευθυνότητες.....	13
ΕΝΟΤΗΤΑ 3. ΔΙΑΧΕΙΡΙΣΗ ΤΟΥ ΕΡΓΟΥ .....	13
3.1 Προτεραιότητες του έργου .....	13
3.2 Βασικά ορόσημα .....	14
3.3 Παρακολούθηση και έλεγχος του έργου.....	14
3.4 Διαχείριση κινδύνων.....	15
3.5 Εκτιμήσεις έργου .....	17
ΕΝΟΤΗΤΑ 4. Τεχνολογικά ζητήματα.....	18
4.1 Γλώσσες προγραμματισμού.....	18
4.2 Πλατφόρμα ανάπτυξης.....	19
4.3 Διαχείριση ανάπτυξης .....	19
4.4 Εξασφάλιση ποιότητας.....	20
4.5 Τεκμηρίωση συστήματος .....	20
Βιβλιογραφία.....	20

## ΕΝΟΤΗΤΑ 1: ΟΡΓΑΝΩΣΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΤΟΥ ΕΡΓΟΥ

### 1. Εισαγωγή

Το παρόν πλάνο έργου λογισμικού καθορίζει το τεχνικό πλαίσιο για την ανάπτυξη ενός ολοκληρωμένου συστήματος σύστασης ταινιών (movie recommendation system) για την εταιρία NTF. Σκοπός του εγγράφου αποτελεί, η περιγραφή και η κατανόηση όλων των διαδικασιών ανάπτυξης και της λειτουργίας συστήματος. Αξιοποιώντας αλγορίθμους ανάλυσης δεδομένων και μηχανικής μάθησης, το σύστημα επιδιώκει να προσφέρει στοχευμένες συστάσεις, σε ένα εύχρηστο περιβάλλον, ενισχύοντας έτσι την αφοσίωση του χρήστη στο περιεχόμενο. Μέσα από αυτό το πλάνο, αναλύονται οι τεχνικές προδιαγραφές, οι κρίσιμοι πόροι καθώς και εξάγονται πληροφορίες σχετικά με τις χρονικές και οικονομικές εκτιμήσεις της διαδικασίας υλοποίησης, που θα οδηγήσουν στην επιτυχή ολοκλήρωση του έργου.

Ειδικότερα το έγγραφο πραγματεύεται αλλά δεν περιορίζεται στα ακόλουθα:

- Γενική περιγραφή του συστήματος.
- Μοντέλο ανάπτυξης του συστήματος.
- Δραστηριότητες που θα εκτελεστούν και δημιουργία προγράμματος.
- Καθορισμός ορόσημων συστήματος.
- Δημιουργία των κατάλληλων προϋποθέσεων για διαχείριση, παρακολούθηση και έλεγχο του έργου.
- Εκτιμήσεις αναφορικά με τη διάρκεια, το μέγεθος του έργου και την απαιτούμενη προσπάθεια, παραγωγικότητα για την ολοκλήρωση του έργου.
- Τεχνολογικά ζητήματα.

Το έγγραφο αυτό θα χρησιμοποιηθεί στην ακόλουθη φάση ανάπτυξης του συστήματος:

- Προγραμματισμός του συστήματος

#### 1.1 Τρέχουσα Κατάσταση

Η NTF είναι μία από τις μεγαλύτερες εταιρίες παραγωγής ταινιών παγκοσμίως. Με μία επιτυχημένη σταδιοδρομία 30 χρόνων, αποφάσισε να ενδυναμώσει την παρουσία της στον κλάδο εισάγοντας για πρώτη φορά στην ιστορία της ένα σύστημα σύστασης ταινιών για τους θεατές της.

Βασικό της μέλημα αυτήν την χρονική περίοδο, είναι να ακολουθήσει την τάση της εποχής προς στην ένταξη της τεχνίτης νοημοσύνης για την επίλυση καθημερινών προβλημάτων. Αξιοποιώντας έτσι, τις εξελίξεις στον κλάδο της μηχανικής μάθησης, αποφάσισε να βάλει τέλος στην αναζήτηση των θεατών για ποιοτική και εύκολα προσβάσιμη ψυχαγωγία.

Έτσι εμείς, αναβαθμίζοντας την διαδικτυακή της πλατφόρμα είμαστε σε θέση να προσφέρουμε το μακροχρόνιο έργο της, στα χέρια των θεατών με το “πάτημα ενός κουμπιού”, καθώς και πολλές άλλες δημοφιλείς ταινίες. Με αυτόν τον τρόπο διατηρεί το κοινό της και εκτός

της αίθουσας του κινηματογράφου και μπορεί πλέον να είναι αναπόσπαστο μέρος της καθημερινότητας των θεατών.

## **1.2 Προηγούμενες Ενέργειες**

Το πρόβλημα που καλούμαστε να λύσουμε είναι η γρήγορη, εύκολη και στοχευμένη πρόταση ταινιών, που είναι ένα πρόβλημα κοινό σε όλους τους σύγχρονους χρήστες του διαδικτύου. Έτσι αφού και εμείς ανήκουμε σε αυτό το κοινό έχουμε την διαίσθηση να εντοπίσουμε τις ανάγκες των χρηστών. Επιπλέον με βάση την τεχνική μας εμπειρία στην δημιουργία ιστοσελίδων καθώς και στην υλοποίηση αλγορίθμων μηχανικής μάθησης διαθέτουμε την διορατικότητα να προβλέψουμε τα πιθανά προβλήματα πριν προκαλέσουν ανεπιθύμητες συνέπειες.

Επιπλέον η εμπειρία της ομάδας με την δημιουργία εφαρμογών web με το framework της React, θα αποτελέσει μεγάλο πλεονέκτημα στην παραγωγή του frontend τμήματος, του συστήματος, και είναι ο βασικός λόγος που την επιλέξαμε για την υλοποίηση.

Παράλληλα έχουμε διεξάγει έρευνες σε παρόμοια λογισμικά διαθέσιμα αυτήν την στιγμή στην αγορά, ώστε να συλλέξουμε πληροφορίες που θα μας βοηθήσουν στην αποφυγή σφαλμάτων, αλλά και τι χρειάζεται να εντάξουμε στο λογισμικό έτσι ώστε να παραμείνει ανταγωνιστικό.

## **1.3 Στόχοι του συστήματος**

Βασικό μέλημα του συστήματος αποτελεί η ευχρηστία της ηλεκτρονικής πλατφόρμας επιλογής της ταινίας. Να είναι δηλαδή ο χρήστης σε θέση να αναζητήσει ταινίες ανά κατηγορία και ανά όνομα και να μπορεί να δει λεπτομέρειες για αυτήν, που συμπεριλαμβάνουν μια σύντομη περιγραφή της πλοκής, τον σκηνοθέτη και τους ηθοποιούς που απαρτίζεται.

Επιπλέον, θα πρέπει ο θεατής να είναι ικανοποιημένος από τις προτάσεις ταινιών που του προσφέρουμε. Για αυτό ο αλγόριθμος μηχανικής μάθησης επιβάλλεται να είναι τόσο έγκυρος, όσο και γρήγορος για να κερδίσουμε την προσοχή του χρήστη. Επίσης σημαντική είναι και η αποθήκευση των αγαπημένων ταινιών του θεατή για εύκολη επαναπαρακολούθηση.

Από την πλευρά των διαχειριστών, χρειάζεται να υπάρχει ένας αποδοτικός τρόπος για να προσθέτονται ή να αφαιρούνται ταινίες στο σύστημα, ώστε να ανανεώνεται ο κατάλογος, χωρίς να υπάρχουν περιορισμοί στο πλήθος ταινιών που θα χωρούν.

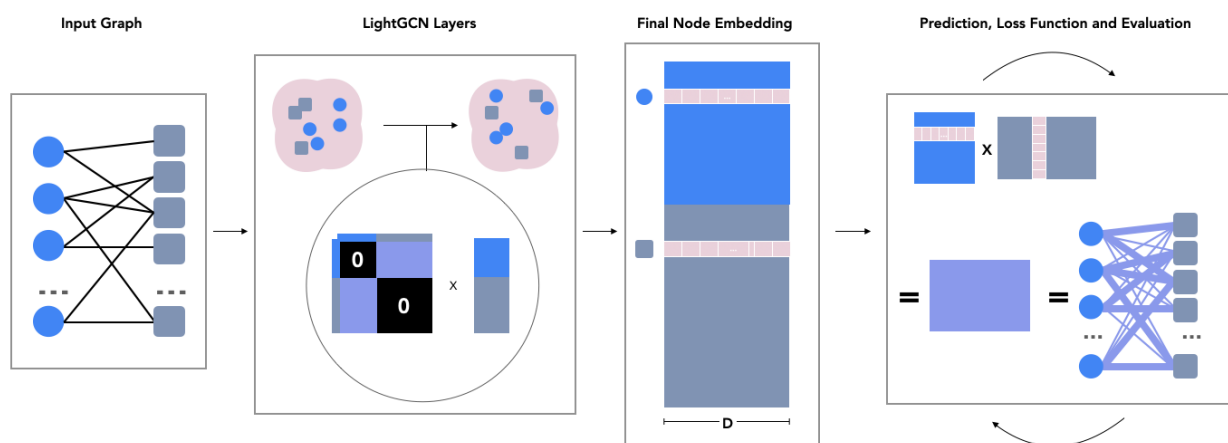
## 1.4 Κατηγορίες χρηστών

Το σύστημα απευθύνεται σε μία πολύ μεγάλη ποικιλία θεατών. Από τους ανθρώπους που θέλουν να χαλαρώσουν μετά από μία κουραστική μέρα εργασίας, μέχρι και τον φανατικό θεατή ταινιών, που ανεβάζει τις κριτικές του στα μέσα κοινωνικής δικτύωσης. Πρέπει να είναι σε θέση να προσφέρει, διασκέδαση σε άτομα κάθε ηλικίας, πράγμα που εξαρτάται τόσο από την λειτουργικότητα του συστήματος, όσο και από την ποικιλία των ταινιών που διαθέτει.

## 1.5 Επισκόπηση του συστήματος

Οι 2 αλγόριθμοι μηχανικής μάθησης αποτελούν την “καρδιά” του συστήματος, καθώς είναι υπεύθυνοι για τις προτάσεις των ταινιών.

Ο πρώτος είναι βασισμένος σε ένα νευρωνικό δίκτυο με όνομα LightGCN [1], και είναι σχεδιασμένος με γνώμονα, την απλούστευση των μαθηματικών σχέσεων μεταξύ του χρήστη και των ταινιών που έχει επιλέξει. Με αυτόν τον τρόπο επιτυγχάνει μεγάλη αποδοτικότητα (ακόμα και συγκριτικά με πιο περίπλοκους αλγόριθμους) με πολύ μικρότερο κόστος εκπαίδευσης.



Εικόνα 1: Λειτουργικό διάγραμμα του LightGCN

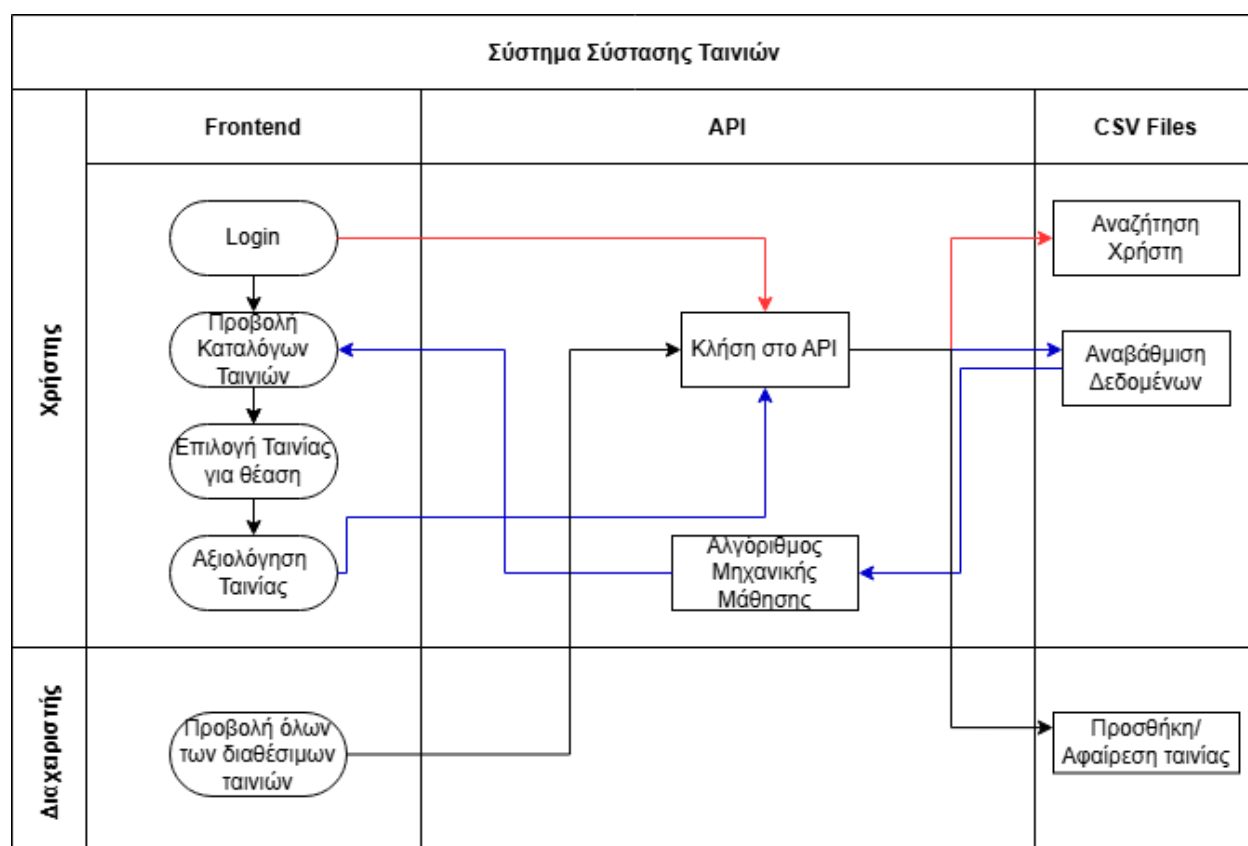
Αν και ο αλγόριθμος είναι αυτός που δίνει την λειτουργική αξία στο σύστημα, το user interface (UI) την διαδικτυακής σελίδας, είναι αυτό που στο τέλος θα κερδίσει την εμπιστοσύνη του χρήστη, θα κεντρίσει το ενδιαφέρον του και θα τον κρατήσει έτσι στην πλατφόρμα. Με ένα μινιμαλιστικό σχέδιο, που δεν “υπερφορτώνει” τον θεατή με πληροφορίες, θα προβάλει τις ταινίες που του προτείνουμε αλλά και τις πιο δημοφιλείς, θα δίνει δυνατότητες αναζήτησης ανά όνομα και κατηγορία, καθώς και θα δίνει επιλογή για αξιολόγηση των προβληθέντων ταινιών και αποθήκευσή τους.

Ο δεύτερος αλγόριθμος ονομάζεται FIRE [5] και είναι υπεύθυνος για την γρήγορη παραγωγή συστάσεων. Από το αρχικό σύνολο 10.000 ταινιών, καταλήγει για κάθε εγγεγραμμένο χρήστη σε έναν πίνακα 100 ταινιών με διαδικασία φίλτρων. Έπειτα με πολύ μικρότερο μέγεθος πινάκων ο LightGCN μπορεί να επανεκπαιδευτεί γρήγορα έτσι ώστε να αποδώσει σε ελάχιστα δευτερόλεπτα το τελικό σύνολο προτεινόμενων ταινιών για τον ζητούμενο θεατή.

Για την ομαλή λειτουργία του συστήματος χρειάζεται μία μέθοδος αποθήκευσης. Επιλέξαμε ένα σύνολο από csv αρχεία όπου το καθένα αποθηκεύει τα ακόλουθα δεδομένα:

- Τον τίτλο τους.
- Σκηνοθέτη και ηθοποιούς.
- Μία σύντομη περιγραφής της πλοκής.
- Τις κατηγορίες που ανήκουν.
- Ποιες ταινίες έχει δει κάθε χρήστης, και την βαθμολογία που τις έβαλε.

Η χρήση αυτών των αρχείων προϋποθέτει τόσο software υποδομή όσο και την αξιοποίηση server.



Εικόνα 2: Σχηματικό διάγραμμα της λειτουργικής ροής του συστήματος

## 1.6 Ανάλυση SWOT

Μέσω της ανάλυσης SWOT μπορούμε να εντοπίσουμε τις δυνάμεις, τα πλεονεκτήματα του συστήματος σε σύγκριση με τους ανταγωνιστές του, τις αδυναμίες του, τις ευκαιρίες στην αγορά και γενικά όλους τους εξωτερικούς παράγοντες που το ευνοούν και τις απειλές που θέτουν σε κίνδυνο την επιτυχία του.

<b><u>Δυνάμεις (Strengths)</u></b>  Το σύστημα μπορεί να προσαρμοστεί εύκολα σε αυξανόμενο όγκο δεδομένων και χρηστών, διασφαλίζοντας τη σταθερή λειτουργία ακόμα και σε περιόδους υψηλής ζήτησης.  Η αυτοματοποίηση μέσω ML μειώνει την ανάγκη για χειροκίνητη διαχείριση και βελτιώνει την ακρίβεια στις συστάσεις.  Το μοντέλο ML που υλοποιήσαμε δεν απαιτεί χρονοβόρα διαδικασία εκπαίδευσης, γεγονός που μειώνει το υπολογιστικό και άρα το οικονομικό κόστος.	<b><u>Αδυναμίες (Weaknesses)</u></b>  Η ακρίβεια των συστάσεων εξαρτάται άμεσα από την ποιότητα και την πληρότητα των δεδομένων. Ελλιπή ή λίγα δεδομένα μπορούν να οδηγήσουν σε λανθασμένες συστάσεις.  Το πρόβλημα “Cold Start” που αντιμετωπίζουν όλα τα συστήματα σύστασης. Δηλαδή, οι νέοι χρήστες χωρίς επαρκές ιστορικό αλληλεπίδρασης μπορεί να μην λαμβάνουν ακριβείς προτάσεις.
<b><u>Ευκαιρίες (Opportunities)</u></b>  Οι πλατφόρμες streaming έχουν αυξηθεί σε ζήτηση τα τελευταία χρόνια και πρόκειται να συνεχίσουν την άνοδο τους ειδικά στην Ελλάδα μετά την πρόσφατη νομοθεσία που αποτρέπει το κατέβασμα και την προβολή ταινιών από πολλές ιστοσελίδες.  Με την ενσωμάτωση πρόσθετων αναλυτικών εργαλείων και λειτουργιών, όπως προβλέψεις δημοτικότητας ή με την συλλογή δεδομένων από το ιστορικό του χρήστη, μπορούμε να βελτιώσουμε το σύστημα και να αξιοποιήσουμε τα δεδομένα του χρήστη για μελλοντικές δράσεις.	<b><u>Απειλές (Threats)</u></b>  Άλλες πλατφόρμες και υπηρεσίες σύστασης ταινιών, ιδιαίτερα από μεγάλους παίκτες στον χώρο του streaming, μπορεί να αποτελέσουν σημαντικό ανταγωνισμό.  Η συλλογή και διαχείριση μεγάλου όγκου προσωπικών δεδομένων ενέχει κινδύνους που σχετίζονται με την ασφάλεια και την ιδιωτικότητα, απαιτώντας αυστηρά μέτρα προστασίας.  Η ραγδαία εξέλιξη των τεχνολογιών ML και ανάλυσης δεδομένων μπορεί να καταστήσει το υπάρχον σύστημα λιγότερο ανταγωνιστικό αν δεν γίνει συνεχής ενημέρωση και βελτίωση.

## 2. ΟΡΓΑΝΩΣΗ ΤΟΥ ΕΡΓΟΥ

### 2.1 Μέθοδος ανάπτυξης

Για την ανάπτυξη του ολοκληρωμένου συστήματος σύστασης ταινιών, επιλέγεται το Επαναληπτικό (iterative) μοντέλο, που διασπά την ανάπτυξη σε μικρούς κύκλους, όπου κάθε κύκλος περιλαμβάνει σχεδίαση, υλοποίηση, δοκιμές και ανασκόπηση. Αυτό επιτρέπει γρήγορη προσαρμοστικότητα σε αλλαγές και σταδιακή παράδοση κώδικα. Κάθε επανάληψη παραδίδει ένα λειτουργικό υποσύνολο της εφαρμογής, που επιτρέπει και πολλές δοκιμές του λογισμικού.

Το επιλέξαμε γιατί δεν γνωρίζουμε ακριβώς τις απαιτήσεις του συστήματος, καθώς αποτελείται από 2 περίπλοκους αλγόριθμους, που πρέπει και να δουλεύουν μεμονωμένα και συνεργατικά. Οι απαιτήσεις, ειδικά στα αρχικά στάδια, δεν είναι ξεκάθαρες και κρίνονται απαραίτητες, λεπτομερείς δοκιμές κάθε υποσυστήματος ξεχωριστά και η προσεκτική τους συναρμολόγηση. Οι βασικές φάσεις του μοντέλου είναι:

## Επανάληψη 1: Κλάσεις & Στατικό UI

### Σχεδιασμός

- Βασικές Κλάσεις
  - App
  - Login / Sign up(επιλογή χρήστη)
  - ArxikhXrhsthScreen (όλες οι ταινίες)
  - DetailScreen (λεπτομέρειες)
  - Recommendations (τελικές προτάσεις)

### Υλοποίηση

- Conditional rendering στο App για μετάβαση από επιλογή χρήστη -> προτάσεις.
- Styling με CSS.

### Έλεγχος

- Ενεργός έλεγχος του UI μέσω expo go σε κινητό και μέσω local host σε υπολογιστή.

### Δράση

- Προσαρμογή ονομάτων φακέλων και κλάσεων αν χρειάζεται.
- Αλλαγές στη δομή όπου κρίνεται απαραίτητο



## Επανάληψη 2: Ενσωμάτωση LightGCN για Re-Ranking

### Σχεδιασμός

- Μελέτη Αλγορίθμου
  - Ανάλυση άρθρου LightGCN για κατανόηση layer aggregation και loss function.
- Δημιουργία Dataset
  - Προεπεξεργασία dataset με διαχωρισμό 80/20 train/test (train\_matrix.npz, test\_matrix.npz).
- Δομές Δεδομένων
  - Μετατροπή sparse adjacency matrix σε coo format για αποδοτική αποθήκευση δεδομένων.

### Υλοποίηση

- Preprocessing Script:
  - Κωδικοποίηση IDs, scaling ratings.
- Model Definition:
  - Κλάση LightGCN με μεταφορά μνημάτων μεταξύ κόμβων.
  - Training loop με Adam, StepLR και MSE loss.
- Αποθήκευση checkpoints και τελικό lightgcn\_final.pth.

### Έλεγχος

- Μετρικές με NDCG@K, Precision@K στο test set.
- Σύγκριση αποτελεσμάτων για διαφορετικά epochs και learning rate.
- Μέτρηση χρόνου ανά epoch.

### Δράση

- Hyperparameter Tuning.

## Επανάληψη 3: Ενσωμάτωση FIRE για Candidate Generation

### Σχεδιασμός

- Ανάλυση άρθρου LightGCN για κατανόηση layer aggregation και loss function.
- Προεπεξεργασία dataset με διαχωρισμό 80/20 train/test (train\_matrix.npz, test\_matrix.npz).

## Υλοποίηση

- Αλλαγή προτεινόμενων βιβλιοθηκών ώστε να υπάρχει συνεργασία με τον lightGCN

## Έλεγχος

- Έλεγχος νέων δεδομένων με αλλαγή userId.
- Μετρικές με NDCG@K, Precision@K στο test set.

## Δράση

- Βελτίωση λογικής για υψηλότερη απόδοση.

## Επανάληψη 4: Δημιουργία API

### Σχεδιασμός

- Ανάλυση όλων των διαφορετικών σελίδων και λειτουργιών που θα χρειαστούν.
- Διαχωρισμός στο τι θα χτιστεί με PHP και τι με Flask.

### Υλοποίηση

- Δημιουργία GET και POST μεθόδων σε PHP.
- Σχεδιασμός απαραίτητων routes στο Flask για καλύτερη σύνδεση με τους αλγόριθμους σε Python.

### Έλεγχος

- Έλεγχος κάθε route για διαφορετικά user IDs και επαλήθευση των αποθηκευμένων δεδομένων στα αρχεία csv.

### Δράση

- Αλλαγή και απλοποίηση των απαραίτητων τμημάτων.

## Επανάληψη 5: Συναρμολόγηση

### Σχεδιασμός

- Ορίζουμε ξεκάθαρα πώς το frontend αλληλεπιδρά με τις endpoints (FIRE, LightGCN), εξασφαλίζοντας σταθερή μορφή δεδομένων.
- Σχεδιάζουμε τον τρόπο με τον οποίο τα αποτελέσματα από FIRE και LightGCN θα συγχωνευτούν σε ένα ενιαίο payload προς το UI.
- Προετοιμάζουμε ένα περιβάλλον όπου μπορούν να τρέξουν όλα τα components μαζί, χωρίς να επηρεάζεται η παραγωγή.

### Υλοποίηση

- Συνδέουμε το fire και το lightgcن με το app, ώστε μετά το fetch των υποψηφίων να ακολουθεί αυτόματα το rerank και η απόδοση των τελικών προτάσεων στον χρήστη.

### Έλεγχος

- Ολοκληρωμένα tests που διασφαλίζουν τη σωστή αλληλουχία κλήσεων και τη συμβάδηση των απαντήσεων με τις απαιτήσεις του UI.

### Δράση

- Ενσωματώνουμε βελτιώσεις, διορθώνοντας edge cases και πιθανές αποκλίσεις δεδομένων.
- Αυτοματοποίηση όλων των commands για την ενεργοποίηση του pipeline σε ένα command στο terminal για εύκολη χρήση και περαιτέρω testing.

Κάθε επανάληψη ακολουθεί αυστηρά τον κύκλο Σχεδιασμός->Υλοποίηση->Έλεγχος->Δράση, διατηρώντας σταθερότητα και σαφήνεια, αξιοποιώντας το FIRE για γρήγορη γεννήτρια υποψηφίων και το LightGCN για ακριβή επανακατάταξη, με ένα σταθερό feedback loop.

## 2.2 Προσωπικό, ρόλοι και υπευθυνότητες



Εικόνα 3: Ρόλοι των μελών

## 3. Διαχείριση Έργου

Σε αυτό το κεφάλαιο θα αναλυθούν οι προτεραιότητες του έργου, καθώς και τα βασικά του ορόσημα, μαζί με την ημερομηνία τους. Στη συνέχεια, θα παρουσιαστούν οι μέθοδοι ελέγχου του έργου, καθώς και οι πιθανοί κίνδυνοι που ενδέχεται να προκύψουν, συνοδευόμενοι από τις αντίστοιχες στρατηγικές διαχείρισής τους. Τέλος, θα δημιουργηθεί το διάγραμμα Gantt, στο οποίο θα αποτυπώνονται οι εκτιμήσεις του έργου.

### 3.1 Προτεραιότητες του έργου.

Το έργο αφορά την ανάπτυξη ενός ολοκληρωμένου συστήματος σύστασης ταινιών για την εταιρεία NTF, με στόχο τη βελτίωση της εμπειρίας των χρηστών.

Μία από τις σημαντικότερες προτεραιότητες του έργου, είναι η ανάπτυξη ενός αποδοτικού και αξιόπιστου συστήματος, το οποίο θα καλύπτει πλήρως τις ανάγκες των χρηστών και της εταιρείας. Παράλληλα, είναι κρίσιμο το έργο να ολοκληρωθεί εντός των χρονικών ορίων που έχουν καθοριστεί, αποφεύγοντας τυχόν καθυστερήσεις που θα μπορούσαν να επηρεάσουν τη συνολική υλοποίηση.

Ακόμη, δίνεται ιδιαίτερη έμφαση στην ασφάλεια των δεδομένων. Η διαχείριση των προσωπικών πληροφοριών των χρηστών χρειάζεται να συμμορφώνεται πλήρως με τον Κανονισμό GDPR.

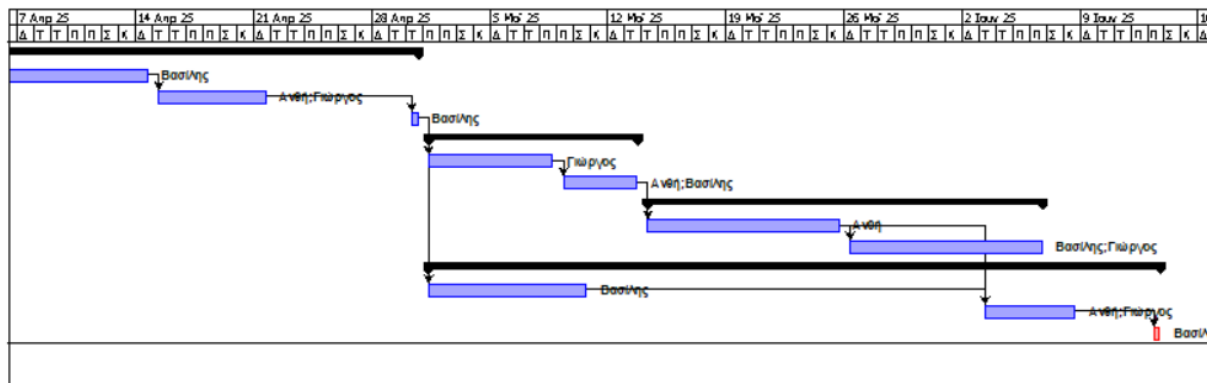
### 3.2 Βασικά ορόσημα

Τα βασικά ορόσημα του έργου, μαζί με την προσδοκώμενη ημερομηνία τους, αντικατοπτρίζονται στις παρακάτω εικόνες. Όπως αναφέρθηκε και στην Ενότητα 2.1, η μέθοδος ανάπτυξης του έργου που επιλέχθηκε είναι το μοντέλο καταρράκτη.

Στο πλαίσιο του μοντέλου καταρράκτη, κάθε φάση του έργου πρέπει να ολοκληρωθεί πλήρως πριν η ομάδα προχωρήσει στην επόμενη. Τα ορόσημα λειτουργούν ως χρονικά σημεία ελέγχου, διασφαλίζοντας ότι οι εργασίες κάθε φάσης έχουν ολοκληρωθεί σύμφωνα με τον χρονοπρογραμματισμό του έργου.

		Όνομα	Διάρκεια	Έναρξη	Λήξη	Προκάτοχοι	Ονόματα Πόρων
1		Ανάλυση Απαιτήσεων	26 days?	5/4/2025 8:00 πμ	30/4/2025 5:00 μμ		
2		Καθορισμός Στόχου	10 days?	5/4/2025 8:00 πμ	14/4/2025 5:00 μμ		Βασιλής
3		Καθορισμός Περιπτώσεων...	7 days?	15/4/2025 8:00 πμ	21/4/2025 5:00 μμ	2	Ανθή;Γιώργος
4		Έγκριση απαιτήσεων απ...	1 day?	30/4/2025 8:00 πμ	30/4/2025 5:00 μμ	3	Βασιλής
5		Σχεδιασμός Συστήματος	13 days?	1/5/2025 8:00 πμ	13/5/2025 5:00 μμ		
6		Επιλογή Αρχιτεκτονικής	8 days?	1/5/2025 8:00 πμ	8/5/2025 5:00 μμ	4	Γιώργος
7		Σχεδίαση κλάσεων	5 days?	9/5/2025 8:00 πμ	13/5/2025 5:00 μμ	6	Ανθή;Βασιλής
8		Υλοποίηση	24 days?	14/5/2025 8:00 πμ	6/6/2025 5:00 μμ		
9		Υλοποίηση κλάσεων	12 days?	14/5/2025 8:00 πμ	25/5/2025 5:00 μμ	7	Ανθή
10		Ενοποίηση κώδικα	12 days?	26/5/2025 8:00 πμ	6/6/2025 5:00 μμ	9	Βασιλής;Γιώργος
11		Ενσωμάτωση και Δοκιμές	44 days?	1/5/2025 8:00 πμ	13/6/2025 5:00 μμ		
12		Συγγραφή Περιπτώσεων...	10 days?	1/5/2025 8:00 πμ	10/5/2025 5:00 μμ	4	Βασιλής
13		Δοκιμές	6 days?	3/6/2025 8:00 πμ	8/6/2025 5:00 μμ	9;12	Ανθή;Γιώργος
14		Τελική Παράδοση Συστή...	1 day?	13/6/2025 8:00 πμ	13/6/2025 5:00 μμ	13	Βασιλής

Εικόνα 4: Βασικές εργασίες του έργου



Εικόνα 5: Διάγραμμα Gantt του έργου

### **3.3 Παρακολούθηση και έλεγχος του έργου.**

Η παρακολούθηση της προόδου του έργου θα πραγματοποιείται μέσω χρονοπρογραμματισμού, στο λογισμικό ProjectLibre, όπου έχει οριστεί η διάρκεια κάθε φάσης. Με το πέραςμα του δηλωμένου χρονικού διαστήματος της εκάστοτε φάσης, θα διεξάγεται διαδικτυακή συνάντηση μέσω Zoom, στην οποία θα γίνεται έλεγχος της δουλειάς που έχει πραγματοποιηθεί και εντοπισμός ενδεχόμενων προβλημάτων.

Σε περίπτωση που εντοπιστούν λάθη, τα μέλη της ομάδας θα συνεργάζονται ομαλά για την επίλυσή τους, ενώ παράλληλα θα προχωρούν στην διεξαγωγή των εργασιών της επόμενης φάσης του έργου. Ουσιαστικά, στόχος είναι στην επόμενη συνάντηση, όλες οι εκκρεμότητες της προηγούμενης φάσης να έχουν επιλυθεί, διασφαλίζοντας έτσι την ομαλή εξέλιξη του έργου και την ολοκλήρωση του στο σωστό χρόνο.

Επιπρόσθετα, κατά τη διάρκεια της υλοποίησης του έργου, ιδιαίτερα κατά τη συγγραφή του κώδικα, μπορεί να πραγματοποιηθούν δια ζώσης συναντήσεις σε χώρους του πανεπιστημίου, όπου τα μέλη της ομάδας θα έχουν την ευκαιρία να επιλύσουν σύνθετα τεχνικά ζητήματα σε πραγματικό χρόνο.

Συνοψίζοντας, με την παραπάνω μεθοδολογία, θα διασφαλιστεί η συνεχής επικοινωνία και η αποτελεσματική ολοκλήρωση κάθε φάσης του έργου.

### **3.4 Διαχείριση κινδύνων**

#### **Ανάλυση Κινδύνων και Στρατηγικές Αντιμετώπισης**

Το συγκεκριμένο έργο παρουσιάζει διάφορους κινδύνους που μπορούν να επηρεάσουν την ομαλή πορεία και την επιτυχία του. Οι βασικοί κίνδυνοι εντοπίζονται στο ανθρώπινο δυναμικό, τις απαιτήσεις, την τεχνολογία και την εκτίμηση του χρόνου ολοκλήρωσης.

#### **Τύπος κινδύνου: Προσωπικό**

Ένας σημαντικός κίνδυνος αφορά το προσωπικό, καθώς η απουσία βασικών μελών λόγω ασθένειας σε κρίσιμες φάσεις του έργου μπορεί να οδηγήσει σε σημαντικές καθυστερήσεις. Η πιθανότητα εμφάνισης αυτού του κινδύνου είναι μέτρια, όμως οι επιπτώσεις του θα μπορούσαν να είναι σοβαρές, επηρεάζοντας την ομαλή εξέλιξη του έργου και τη συνολική τήρηση του χρονοδιαγράμματος.

### **Διαχείριση κινδύνου:**

Δεδομένου ότι δεν είναι δυνατή η προσθήκη νέων μελών, θα γίνεται αναλυτική τεκμηρίωση των εργασιών, ώστε να υπάρχει σαφήνεια για τα καθήκοντα κάθε μέλους και να διευκολύνεται η ανάληψη εργασιών από άλλα μέλη αν προκύψει ανάγκη. Επιπλέον, οι τακτικές διαδικτυακές συναντήσεις στο τέλος κάθε φάσης, θα συμβάλουν στο να έχουν όλοι εικόνα των εξελίξεων και να μπορούν να προσαρμοστούν εάν χρειαστεί.

### **Τύπος κινδύνου: Απαιτήσεις**

Έπειτα, σημαντικός κίνδυνος αφορά τις απαιτήσεις του έργου, καθώς τυχόν προτάσεις για αλλαγές μπορεί να απαιτήσουν ριζικό ανασχεδιασμό. Ειδικά στο πλαίσιο του μοντέλου ανάπτυξης καταρράκτη, όπου κάθε φάση ολοκληρώνεται πριν την έναρξη της επόμενης, οι μεγάλες τροποποιήσεις στις απαιτήσεις θα διαταράξουν την ομαλή ροή του έργου. Η πιθανότητα εμφάνισης αυτού του κινδύνου είναι μέτρια, όμως οι επιπτώσεις θα μπορούσαν να είναι σοβαρές, προκαλώντας καθυστερήσεις.

### **Διαχείριση κινδύνου:**

Για να ελαχιστοποιηθεί ο αντίκτυπος, είναι σημαντικό να υπάρχει σαφής ορισμός των απαιτήσεων από την αρχή. Ο διαχειριστής του έργου θα αξιολογήσει την αναγκαιότητα και τον αντίκτυπο κάθε αλλαγής, λαμβάνοντας υπόψη το χρονοδιάγραμμα, τους διαθέσιμους πόρους και τις υπόλοιπες απαιτήσεις. Αν μια αλλαγή κριθεί σημαντική, θα εξεταστεί ο τρόπος ενσωμάτωσής της με τη μικρότερη δυνατή διατάραξη της ροής του έργου.

### **Τύπος κινδύνου: Τεχνολογία**

Στο τεχνολογικό επίπεδο, ενδέχεται να παρουσιαστούν προβλήματα με τμήματα λογισμικού που πρέπει να χρησιμοποιηθούν, καθώς μπορεί να περιέχουν ελαττώματα που περιορίζουν τη λειτουργικότητά τους. Αυτά τα προβλήματα μπορεί να προκύψουν λόγω παλαιότερων σφαλμάτων, έλλειψης τεκμηρίωσης ή ασυμβατότητας με τις υπόλοιπες τεχνολογίες του έργου. Η πιθανότητα εμφάνισης αυτού του κινδύνου είναι μέτρια, όμως οι επιπτώσεις του θα μπορούσαν να είναι σοβαρές.

### **Διαχείριση κινδύνου:**

Για την αντιμετώπιση αυτού του κινδύνου, απαιτείται ενδελεχής έλεγχος πριν από την ενσωμάτωση των επαναχρησιμοποιούμενων στοιχείων, ώστε να διασφαλιστεί η σταθερότητα και η συμβατότητά τους. Παράλληλα, είναι κρίσιμη η διεξαγωγή δοκιμών και η αξιολόγηση του κώδικα για τον έγκαιρο εντοπισμό πιθανών δυσλειτουργιών, καθώς και η χρήση ενημερωμένων και δοκιμασμένων βιβλιοθηκών που διαθέτουν ενεργή υποστήριξη και συχνές ενημερώσεις.

### **Τύπος κινδύνου: Εκτίμηση Χρόνου**

Τέλος, ένας από τους πιο βασικούς κινδύνους αφορά την εκτίμηση του χρόνου ανάπτυξης. Υπάρχει υψηλή πιθανότητα να έχει υποτιμηθεί ο απαιτούμενος χρόνος, γεγονός που μπορεί να οδηγήσει σε καθυστερήσεις.

#### **Διαχείριση**

#### **κινδύνου:**

Το ProjectLibre χρησιμοποιείται για τον ακριβή χρονοπρογραμματισμό του έργου, ωστόσο, παραμένει η ανάγκη για συνεχή παρακολούθηση της προόδου και πιθανές προσαρμογές του χρονοδιαγράμματος. Εάν προκύψουν τεχνικές δυσκολίες, η ομάδα πρέπει να είναι προετοιμασμένη να ανακαταναείμει τις εργασίες ή να αφιερώσει περισσότερο χρόνο στις πιο απαιτητικές φάσεις του έργου.

#### **Συμπεράσματα**

Το έργο ενέχει αρκετούς κινδύνους, οι οποίοι όμως μπορούν να αντιμετωπιστούν μέσω σωστού σχεδιασμού, έγκαιρης ανίχνευσης προβλημάτων και συνεχούς παρακολούθησης της προόδου με το εργαλείο ProjectLibre. Με την κατάλληλη διαχείριση, το σύστημα σύστασης ταινιών της NTF μπορεί να αναπτυχθεί επιτυχώς, παρέχοντας μια βελτιωμένη εμπειρία στους χρήστες.

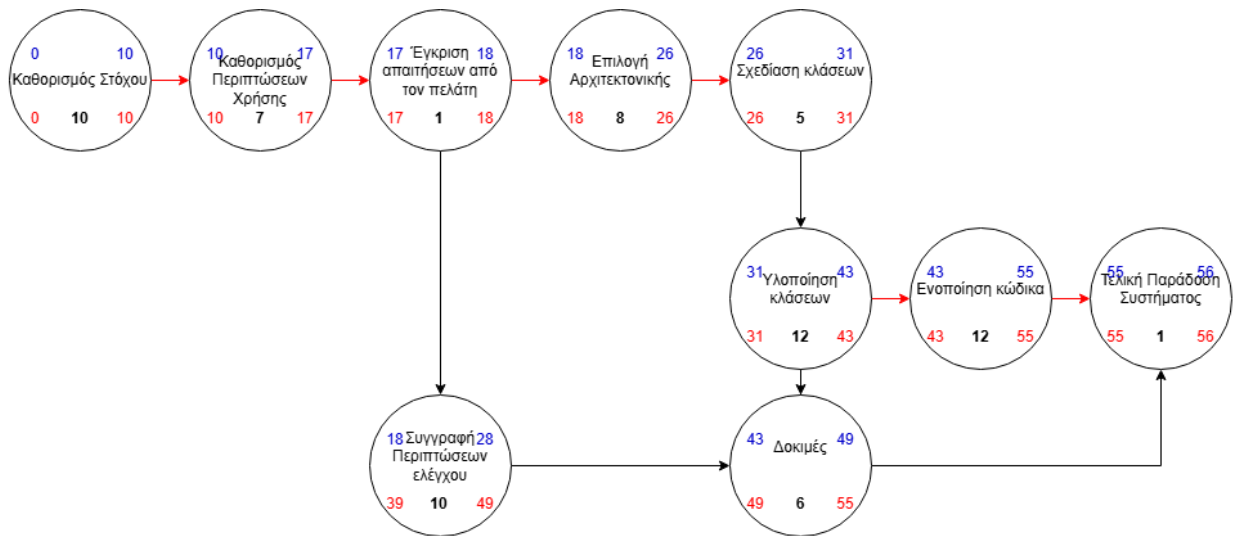
### **3.5 Εκτιμήσεις έργου**

Η εκτίμηση της συνολικής διάρκειας του έργου γίνεται βάση του διαγράμματος Gantt στο ProjectLibre. Το έργο ξεκινά στις 5/4/2025 και ολοκληρώνεται στις 15/6/2025, με συνολική χρονική διάρκεια 72 ημερών.

### **Κρίσιμο Μονοπάτι (CPM)**

Η μέθοδος του κρίσιμου μονοπατιού [2], είναι ένας αλγόριθμος για τον προγραμματισμό των δραστηριοτήτων του έργου. Καθορίζεται με τον εντοπισμό της μεγαλύτερης έκτασης των εξαρτημένων δραστηριοτήτων και τη μέτρηση του χρόνου που απαιτείται για την ολοκλήρωσή τους από την αρχή μέχρι το τέλος. Το κρίσιμο μονοπάτι του συστήματος μας αποτελείται από τις παρακάτω δραστηριότητες, των οποίων μία πιθανή καθυστέρηση θα επηρεάσει την ημερομηνία παράδοσης του έργου:





Εικόνα 6: Όλες οι εργασίες του έργου, το κρίσιμο μονοπάτι φαίνεται με τα κόκκινα βέλη

Επεξηγηματικά, τυχόν καθυστέρηση στις παραπάνω εργασίες θα καθυστερήσει ολόκληρο το έργο.

## Συμπέρασμα

Οι εκτιμήσεις δείχνουν ότι το έργο μπορεί να ολοκληρωθεί εντός του προβλεπόμενου χρονοδιαγράμματος, υπό την προϋπόθεση ότι δεν θα υπάρξουν καθυστερήσεις στις κρίσιμες δραστηριότητες. Για τη μείωση του κινδύνου καθυστέρησης, θα πραγματοποιείται στενή παρακολούθηση της προόδου.

## 4. Τεχνολογικά ζητήματα

### 4.1 Γλώσσες προγραμματισμού

Το μοντέλο μηχανικής μάθησης έχει δημιουργηθεί με την Python. Η Python αν και πολύ αργή σε σύγκριση με τις περισσότερες γλώσσες, λόγω της ευελιξίας της και της εύκολης σύνταξης της, έχει γίνει το standard στην βιομηχανία για την υλοποίηση αλγορίθμων νευρωνικών δικτύων.

Για την δημιουργία της διαδικτυακής διεπαφής επιλέξαμε το framework React του Javascript έναντι του πιο συνηθισμένου “tech stack” που περιέχει “καθαρή” Javascript, HTML και CSS. Ο βασικός λόγος για αυτήν την απόφαση είναι η εμπειρία της ομάδας με την React σε παλαιότερα projects, πράγμα που θα επιταχύνει την υλοποίηση.

Για την υλοποίηση του backend της διαδικτυακής διεπαφής επιλέξαμε τη γλώσσα PHP και, διότι η ομάδα διαθέτει ήδη εξειδίκευση, από προηγούμενα έργα, γεγονός που θα επιταχύνει την ανάπτυξη και θα διασφαλίσει σταθερή συντήρηση του κώδικα.

## 4.2 Πλατφόρμα ανάπτυξης

Η πλατφόρμα που θα αναπτύξουμε το σύστημα είναι το VS Code, σε λειτουργικό σύστημα Windows, καθώς όλα τα μέλη έχουμε εμπειρία στη χρήση του και θα μας διευκολύνει στην υλοποίηση.

Η βασική βιβλιοθήκη ανάπτυξης του νευρωνικού δικτύου είναι η Pytorch και συνοδεύεται με τις βιβλιοθήκες Pandas, για την προεπεξεργασία των δεδομένων, και την Numpy που βοηθά σε μαθηματικές πράξεις.

## 4.3 Διαχείριση ανάπτυξης

### Kaggle

Τα δεδομένα μας για την εκπαίδευση του νευρωνικού δικτύου, τα έχουμε πάρει από την ιστοσελίδα Kaggle, όπου αξιοποιήσαμε το dataset “MovieLens 100K” [4] που περιέχει 1000 χρήστες και 1700 ταινίες. Σε αυτό το dataset υπάρχουν 4 αρχεία .csv που περιέχουν όλα τα δεδομένα. Αυτά τα αρχεία είναι τα ακόλουθα:

- movies.csv, που περιέχει το id κάθε ταινίας για να τις ξεχωρίζουμε εύκολα μεταξύ τους, τον τίτλο κάθε ταινίας και τις κατηγορίες που ανήκει (πχ ταινία δράσης, κωμωδία κ.α.).
- tags.csv, περιέχει το id του χρήστη και το id των ταινιών που έχει δει καθώς και το χαρακτηριστικό που του έμεινε (tag) από την συγκεκριμένη ταινία (πχ αστεία, αληθινή ιστορία, κάποιο όνομα ηθοποιού που παίζει).
- links.csv, εδώ παρατηρούμε το id κάθε ταινίας που συνοδεύεται με τα id που είναι καταχωρημένη σε δύο από τις πιο γνωστές ιστοσελίδες αξιολόγησης ταινιών, την “IMDB” και την “The Movie Database” (TMDB) .
- ratings.csv, στο τελευταίο αρχείο, και πιο σημαντικό για τον αλγόριθμο μηχανικής μάθησης, βλέπουμε ποιες ταινίες έχει δει κάθε χρήστης και την βαθμολογία που της έβαλε με άριστα το 5. Οι συνολικές βαθμολογημένες συσχετίσεις μεταξύ χρηστών και ταινιών φτάνουν τις 100.000, από όπου παίρνει και το dataset το όνομα του. Άρα ο κάθε χρήστης έχει δει κατά μέσο όρο 100 ταινίες.

Για τους σκοπούς αυτού του συστήματος αξιοποιούμε μόνο τα movies.csv και ratings.csv.

### Github

Στη πλατφόρμα Github θα δημοσιεύονται όλα τα αρχεία του έργου συμπεριλαμβανομένου και του documentation και του κώδικα. Επιτρέπει σε οποιονδήποτε να δει και να ελέγξει όλο το ιστορικό αλλαγών μέσω Git, ενισχύοντας την εμπιστοσύνη και τη συνεργασία.

Διευκολύνει την αξιοποίηση του κώδικα, αφού όλοι οι ενδιαφερόμενοι έχουν πρόσβαση στον ίδιο χώρο εργασίας.

#### 4.4 Εξασφάλιση ποιότητας

Η εξασφάλιση της ποιότητας του συστήματος σύστασης ταινιών, περιλαμβάνει τον έλεγχο του συστήματος για να διαπιστωθεί ότι εκτελούνται ορθά οι προκαθορισμένες λειτουργίες. Για το σκοπό αυτό στη συνέχεια θα δοθούν αναλυτικά τα πλάνα ελέγχου το συστήματος.

#### 4.5 Τεκμηρίωση συστήματος

Το σύστημα θα συνοδεύεται από πλήρη τεκμηρίωση που περιλαμβάνει:

- Οδηγίες χρήσης
- Οδηγίες διαχείρισης του συστήματος
- Οδηγίες διαχείρισης της βάσης δεδομένων

#### Βιβλιογραφία

1. Xiangnan He, Kuan Deng, Xiang Wang, “[LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation](#)”
2. Jiafeng Xia, Dongsheng Li, Hansu Gu, “[FIRE: Fast Incremental Recommendation with Graph Signal Processing](#)”
3. Santiago, Jesse, “[Critical Path Method](#)”
4. Kaggle, [MovieLens 100K Dataset](#)