



## Ανακάλυψη Γνώσης από Βάσεις Δεδομένων

PP2: Recommender System

Νικόλαος Σακελλάρης

ics21096

Θεσσαλονίκη, 2024

## Περιεχόμενα

Εισαγωγή .....	3
Προ-επεξεργασία και Data Reduction .....	3
Recommender Script.....	4
Πείραμα 1.....	8
Πείραμα 2.....	11
Πείραμα 3.....	12
Πείραμα 4.....	14
Πείραμα 5.....	15
Μέσος όρος αποτελεσμάτων.....	17
Αρχεία.....	20

## Εισαγωγή

Στόχος της εργασίας ήταν η υλοποίηση και η αποτίμηση ενός συστήματος συστάσεων συνεργατικού φιλτραρίσματος αντικειμένου-αντικειμένου για ταινίες με χρήση του συντελεστή **pearson** ως μέτρο ομοιότητας.

Η υλοποίηση έγινε σε γλώσσα προγραμματισμού Python. Χρησιμοποιήθηκαν τα δεδομένα του MovieLens και συγκεκριμένα το αρχείο **ratings.csv** του μικρού συνόλου των 100.000 εγγραφών.

Υλοποιήθηκαν συνολικά 4 συναρτήσεις πρόβλεψης οι οποίες περιλαμβάνουν:

- Σταθμισμένο μέσο όρο
- Σταθμισμένο μέσο όρο με προσαρμογή της μέσης βαθμολογίας του χρήστη και αφαίρεση του bias
- Σταθμισμένο μέσο όρο με στάθμιση που βασίζεται στο πλήθος των κοινών χρηστών που έχουν βαθμολογήσει τα αντικείμενα
- Σταθμισμένο μέσο όρο με στάθμιση που βασίζεται στη διακύμανση των βαθμολογιών του κάθε αντικειμένου.

Τα κοντινότερα N αντικείμενα επιλέγονται με την ομοιότητα Pearson.

Το πρόγραμμα χωρίζει τα δεδομένα σε σύνολο εκπαίδευσης  $T = 80\%$  και σύνολο ελέγχου  $100 - T = 20\%$ . Στη συνέχεια εμφανίζει σε κάθε περίπτωση: α) το μέσο απόλυτο σφάλμα (MAE), β) τη μέση ακρίβεια (macro average precision) και γ) τη μέση ανάκληση (macro average call). Για τον υπολογισμό των μέτρων αποτίμησης μια ταινία θεωρείται σχετική αν ο βαθμός της είναι  $\geq 3$ .

## Προ-επεξεργασία και Data Reduction

Λόγω του μεγάλου όγκου των δεδομένων (100.000 εγγραφές) κρίθηκε αναγκαίο να χρησιμοποιηθεί αλγόριθμος data reduction για την εξοικονόμηση χρόνου και την βελτιστοποίηση της απόδοσης του τελικού μοντέλου.

Το αρχείο **filter-ratings.py** περιλαμβάνει κώδικα ο οποίος:

- Αρχικά, τροποποιεί τη στήλη timestamp και την μετατρέπει σε αναγνωρίσιμη μορφή date-time.
- Στη συνέχεια, εφαρμόζει μια απλή μορφή data reduction κρατώντας 1 κάθε 5 εγγραφές. Μειώνει έτσι τον συνολικό όγκο των δεδομένων στο 20% του συνολικού (από 100.000 σε 20.000).

Τα αποτελέσματα αποθηκεύτηκαν στο αρχείο *ratings-reduced.csv*.

## Recommender Script

Για σταθερό σύνολο εκπαίδευσης **80%**, σύνολο ελέγχου **20%** και για τις τιμές **N = (5,10,15,20,25)** το πείραμα πραγματοποιείται **5 φορές** συνολικά γίνεται σύγκριση των **4 συναρτήσεων πρόβλεψης** και τα μέτρα αποτίμησης αναφέρονται ως οι μέσοι όροι όλων των εκτελέσεων.

Οι συναρτήσεις πρόβλεψης και ο κώδικας εκτέλεσης του πειράματος περιέχονται στο αρχείο *recommender.py*. Αναλυτικά έχουμε τις εξής συναρτήσεις:

### 1. Σταθμισμένος μέσος όρος:

```
def predict_weighted_average(self, user_ratings, item_index, N):
    # Get N most similar items
    similar_items = np.argsort(self.similarity_matrix[item_index])[-N:]

    # Filter out items without user ratings
    valid_similar_items = similar_items[user_ratings[similar_items] != 0]

    # Calculate weighted average prediction
    numerator = np.sum(self.similarity_matrix[item_index,
valid_similar_items] * user_ratings[valid_similar_items])
    denominator = np.sum(np.abs(self.similarity_matrix[item_index,
valid_similar_items]))

    if denominator == 0:
        return 0 # Avoid division by zero

    prediction = numerator / denominator

    # Replace NaN with 0
    return np.nan_to_num(prediction)
```

### 2. Σταθμισμένος μέσος όρος με προσαρμογή της μέσης βαθμολογίας του χρήστη και αφαίρεση του bias:

```
def predict_weighted_average_adjusted(self, user_ratings, item_index, N):
    # Get N most similar items
    similar_items = np.argsort(self.similarity_matrix[item_index])[-N:]

    # Filter out items without user ratings
    valid_similar_items = similar_items[user_ratings[similar_items] != 0]

    # Calculate adjusted weighted average prediction
    user_avg = np.mean(user_ratings)
```

```

        numerator = np.sum((self.similarity_matrix[item_index,
valid_similar_items] * (user_ratings[valid_similar_items] - user_avg)))
        denominator = np.sum(np.abs(self.similarity_matrix[item_index,
valid_similar_items]))

        if denominator == 0:
            return user_avg # Return user's average rating if denominator is 0

        prediction = user_avg + (numerator / denominator)

        #Replace NaN with 0
        return np.nan_to_num(prediction)

```

### 3. Σταθμισμένος μέσος όρος με στάθμιση που βασίζεται στο πλήθος των κοινών χρηστών που έχουν βαθμολογήσει τα αντικείμενα:

```

def predict_weighted_average_common_users(self, user_ratings, item_index,
N):
    # Get N most similar items
    similar_items = np.argsort(self.similarity_matrix[item_index])[-N:]

    # Filter out items without user ratings
    valid_similar_items = similar_items[user_ratings[similar_items] != 0]

    # Calculate weights based on the number of common users
    common_users_count = np.sum(self.ratings[:, valid_similar_items] != 0,
axis=0)

    # Check if there are common users
    if np.any(common_users_count):
        weights = common_users_count / np.max(common_users_count) #
Normalize weights
    else:
        # If there are no common users, assign equal weights to all items
        weights = np.ones_like(common_users_count) /
len(common_users_count)

    # Calculate weighted average prediction with common users-based
weights
    numerator = np.sum(self.similarity_matrix[item_index,
valid_similar_items] * user_ratings[valid_similar_items] * weights)
    denominator = np.sum(np.abs(self.similarity_matrix[item_index,
valid_similar_items]) * weights)

    if denominator == 0:
        return 0 # Avoid division by zero

    prediction = numerator / denominator

```

```
# Replace NaN with 0
return np.nan_to_num(prediction)
```

Η στάθμιση του κάθε γείτονα ορίζεται από τον αριθμό των κοινών χρηστών που έχουν βαθμολογήσει τόσο το αντικείμενο αυτό όσο και το γειτονικό αντικείμενο. Όσο περισσότεροι κοινόι χρήστες τόσο μεγαλύτερο το βάρος στάθμισης. Το βάρος κάθε γείτονα κανονικοποιείται και διαιρείται με τον μέγιστο αριθμό των κοινών χρηστών ανάμεσα σε όλους τους γείτονες. Αν δεν υπάρχουν κοινόι χρήστες για τα συγκεκριμένα αντικείμενα, ανατίθεται σε κάθε γείτονα το ίδιο βάρος στάθμισης.

```
common_users_count = np.sum(self.ratings[:, valid_similar_items] != 0, axis=0)
```

```
if np.any(common_users_count):
```

```
weights = common_users_count / np.max(common_users_count)
```

```
else:
```

```
np.ones_like(common_users_count) / len(common_users_count)
```

#### 4. Σταθμισμένος μέσος όρος με στάθμιση που βασίζεται στη διακύμανση των βαθμολογιών του κάθε αντικειμένου:

```
def predict_weighted_average_variance(self, user_ratings, item_index, N):
    # Get N most similar items
    similar_items = np.argsort(self.similarity_matrix[item_index])[-N:]

    # Filter out items without user ratings
    valid_similar_items = similar_items[user_ratings[similar_items] != 0]

    # Calculate weights based on the variance of ratings
    weights = np.var(self.ratings[:, valid_similar_items], axis=0)

    # Calculate weighted average prediction with variance-based weights
    numerator = np.sum(self.similarity_matrix[item_index,
valid_similar_items] * user_ratings[valid_similar_items] * weights)
    denominator = np.sum(np.abs(self.similarity_matrix[item_index,
valid_similar_items]) * weights)

    if denominator == 0:
        return 0 # Avoid division by zero

    prediction = numerator / denominator

    # Replace NaN with 0
    return np.nan_to_num(prediction)
```

Το βάρος στάθμισης κάθε γείτονα καθορίζεται από τη διαφορά των βαθμολογιών για αυτό το γειτονικό αντικείμενο. Όσο μεγαλύτερη η διαφορά, τόσο μεγαλύτερο το βάρος στάθμισης. Το βάρος υπολογίζεται χρησιμοποιώντας τη διαφορά των βαθμολογιών για κάθε γειτονικό αντικείμενο. Αν η διαφορά είναι 0, μια πολύ μικρή τιμή προστίθεται για να αποφευχθεί η διαίρεση με το 0.

```
weights = np.var(self.ratings[:, valid_similar_items], axis=0) + 1e-9
```

Το πείραμα εκτελέστηκε συνολικά 5 φορές για τις διάφορες τιμές του N. Για κάθε πείραμα τηρήθηκε ακρίβεια 5 δεκαδικών ψηφίων και υπολογίστηκε ο μέσος όρος όλων των συναρτήσεων καθώς και ο μέσος όρος κάθε τιμής αποτίμησης (MAE, μέση ακρίβεια, μέση ανάκληση).

Ο κώδικας παράγει αποτελέσματα ξεχωριστά για κάθε συνάρτηση τα οποία έχουν την εξής μορφή κατά την εκτύπωση στην κονσόλα:

```
Experiment for N=5 (1/5):
Original Weighted Average (MAE): 3.5127958172812326
Original Weighted Average Precision: 0.4667534655285932
Original Weighted Average Recall: 0.4969693922130952
Weighted Average with Adjustment (MAE): 3.4662397927839144
Weighted Average with Adjustment Precision: 0.4667534655285932
Weighted Average with Adjustment Recall: 0.4969693922130952
Weighted Average with Variance (MAE): 3.513066993091251
Weighted Average with Variance Precision: 0.4667534655285932
Weighted Average with Variance Recall: 0.4969693922130952
Weighted Average with Common Users (MAE): 3.5130334562535612
Weighted Average with Common Users Precision: 0.4667534655285932
Weighted Average with Common Users Recall: 0.4969693922130952

Experiment for N=10 (2/5):
Original Weighted Average (MAE): 3.4166116974494285
Original Weighted Average Precision: 0.5134238464679461
Original Weighted Average Recall: 0.5022638602259987
Weighted Average with Adjustment (MAE): 3.3646190420186404
Weighted Average with Adjustment Precision: 0.5134238464679461
Weighted Average with Adjustment Recall: 0.5022638602259987
Weighted Average with Variance (MAE): 3.4166966946299504
Weighted Average with Variance Precision: 0.5134238464679461
Weighted Average with Variance Recall: 0.5022638602259987
Weighted Average with Common Users (MAE): 3.4166083871941417
Weighted Average with Common Users Precision: 0.5141234456251413
Weighted Average with Common Users Recall: 0.5024005095837467
```

## Πείραμα 1

Ο παρακάτω πίνακας δείχνει τα αποτελέσματα της κάθε συνάρτησης κατά την εκτέλεση του 1<sup>ου</sup> πειράματος για τις τιμές N που ορίστηκαν, και στη συνέχεια υπολογίζεται ο μέσος όρος τους:

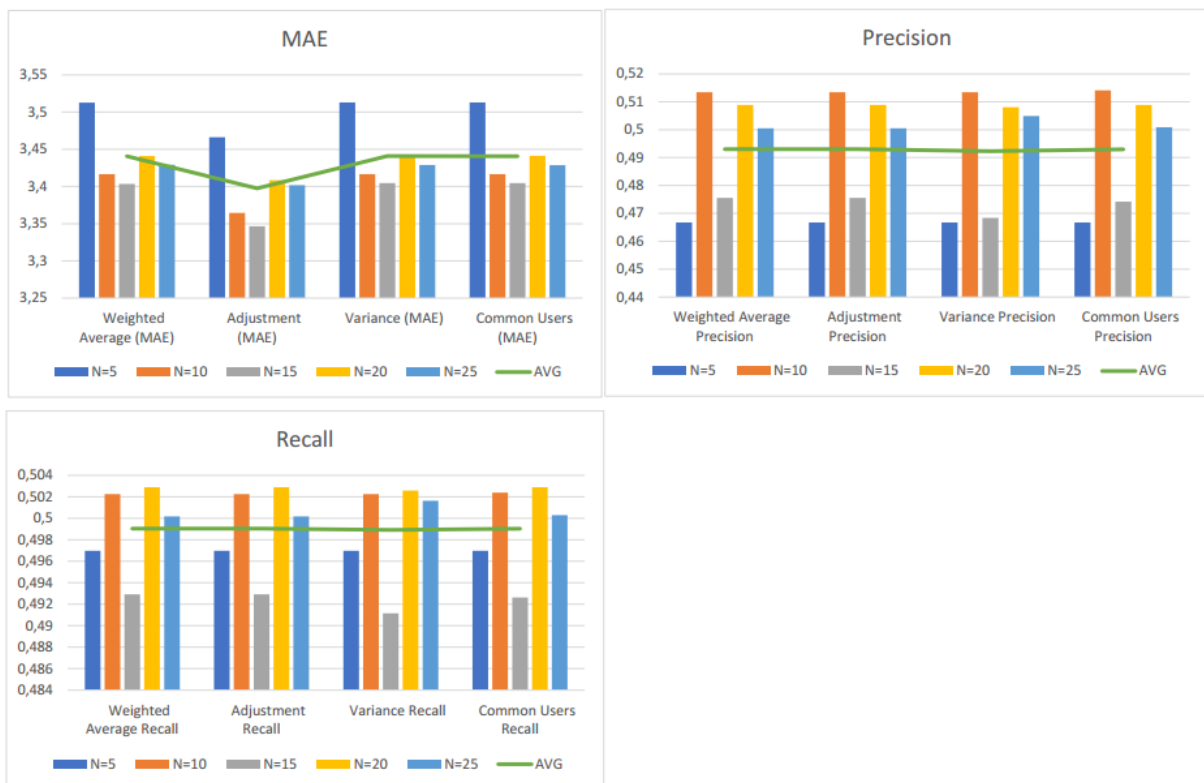
Experiment 1	N=5	N=10	N=15	N=20	N=25	AVG
Weighted Average (MAE)	3,5128	3,41661	3,40356	3,4411	3,42915	3,440644
Weighted Average Precision	0,46675	0,51342	0,47554	0,50887	0,5005	0,493016
Weighted Average Recall	0,49697	0,50226	0,49292	0,50289	0,50017	0,499042
Adjustment (MAE)	3,46624	3,36462	3,34658	3,40838	3,40168	3,3975
Adjustment Precision	0,46675	0,51342	0,47554	0,50887	0,5005	0,493016
Adjustment Recall	0,49697	0,50226	0,49292	0,50289	0,50017	0,499042
Variance (MAE)	3,51307	3,4167	3,40456	3,44157	3,42893	3,440966
Variance Precision	0,46675	0,51342	0,46838	0,50803	0,50488	0,492292
Variance Recall	0,49697	0,50226	0,49115	0,50259	0,50163	0,49892
Common Users (MAE)	3,51303	3,41661	3,40443	3,44136	3,4288	3,440846
Common Users Precision	0,46675	0,51412	0,47421	0,50887	0,5009	0,49297
Common Users Recall	0,49697	0,5024	0,49262	0,50289	0,5003	0,499036

Average Total MAE	3,429989
Average Total Precision	0,492824
Average Total Recall	0,49901

Όπου *Weighted Average* η συνάρτηση **Original Weighted Average**, *Adjustment* η συνάρτηση **Weighted Average with Adjustment**, *Variance* η συνάρτηση **Weighted Average with Variance** και *Common Users* η **Weighted Average with Common Users**.

Οι τιμές των συναρτήσεων αναπαρίστανται γραφικά και στα ακόλουθα γραφήματα:





### Παρατηρήσεις:

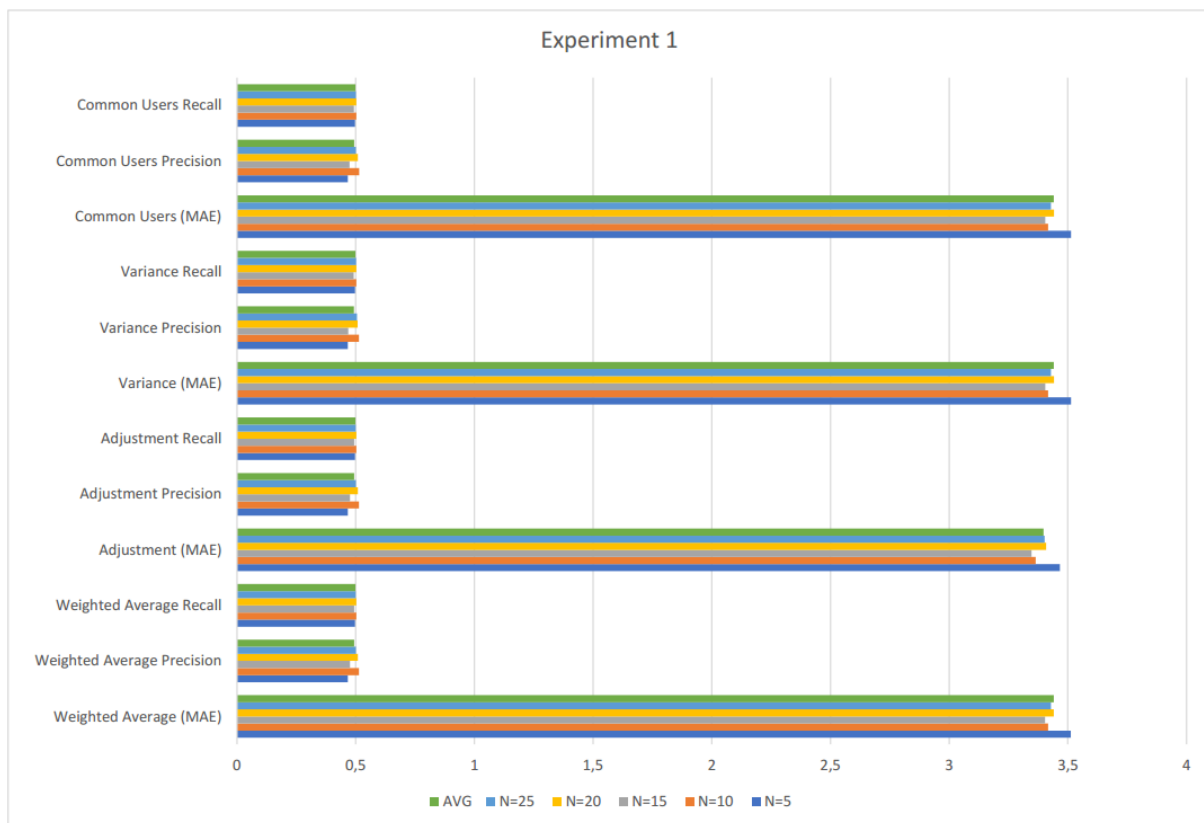
Με βάση τις τιμές που βλέπουμε παρατηρείται σχετική ισορροπία στα αποτελέσματα.

Αναλυτικά οι τιμές για το **MAE** κυμαίνονται από **3,35** έως **3,5** που συναντάται για **N=5**.

Για το Precision βλέπουμε μικρή και πάλι διαφορά στην απόκλιση των τιμών, με τις περιπτώσεις **N=10,20,25** να βρίσκονται κοντά στο **0,5** ενώ για **N=5,15** να μην ξεπερνάνε το **0,47**. Η μέση τιμή βρίσκεται κοντά στο **0,49**.

Για το Recall έχουμε τιμές που διαφέρουν πολύ λιγότερο μεταξύ τους και γενικά κυμαίνονται από **0,49** μέχρι **0,5**.

Το μέσο συνολικό **MAE** είναι **3,42** ενώ τα **Precision** και **Recall** είναι **0,49**.



Συνολικά στο παραπάνω διάγραμμα παρατηρείται ότι οι τιμές των Precision και Recall δεν διαφέρουν σημαντικά τόσο μεταξύ των συναρτήσεων όσο και μεταξύ των διαφορετικών κάθε φορά τιμών N. Αντίθετα το MAE παρουσιάζει μεγάλη διαφορά σε σχέση με τις άλλες δύο μετρικές αλλά οι τιμές των συναρτήσεων είναι και πάλι αρκετά κοντά με το μοναδικό που ξεχωρίζει ελάχιστα να είναι το N=5.

Με ανάλογο τρόπο θα μπορούσαν να παρατηρηθούν αντίστοιχες λεπτομέρειες και για τα πειράματα 2-5. Για τα οποία τα διαγράμματα δεν διαφέρουν σημαντικά, ενώ οι μετρικές MAE, Precision και Recall κυμαίνονται σχεδόν ακριβώς στις ίδιες τιμές.

## Πείραμα 2

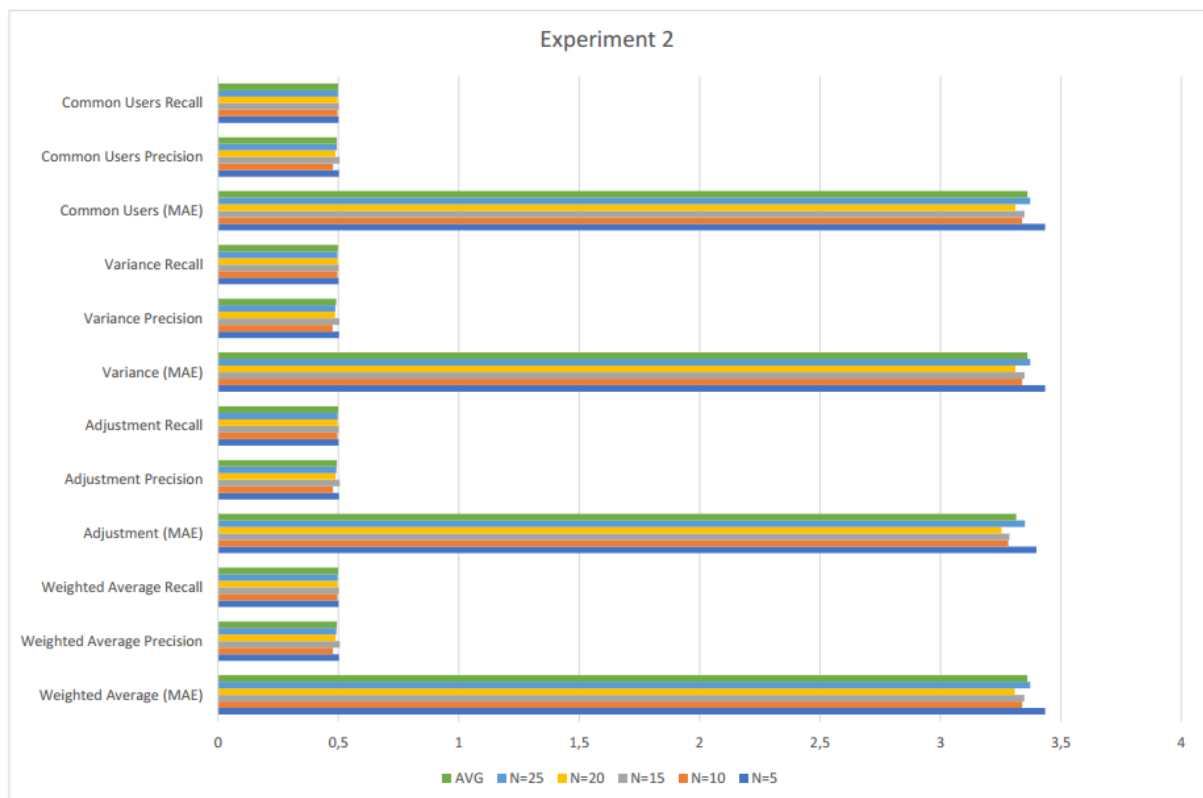
Αποτελέσματα 2<sup>ου</sup> πειράματος:

Experiment 2	N=5	N=10	N=15	N=20	N=25	AVG
Weighted Average (MAE)	3,43442	3,33855	3,34778	3,30835	3,37248	3,360316
Weighted Average Precision	0,50252	0,47768	0,50556	0,48786	0,49152	0,493028
Weighted Average Recall	0,50014	0,49623	0,50126	0,49736	0,49737	0,498472
Adjustment (MAE)	3,3982	3,28171	3,28608	3,25328	3,35022	3,313898
Adjustment Precision	0,50252	0,47768	0,50556	0,48786	0,49152	0,493028
Adjustment Recall	0,50014	0,49623	0,50126	0,49736	0,49737	0,498472
Variance (MAE)	3,43435	3,33884	3,34826	3,31082	3,37245	3,360944
Variance Precision	0,50252	0,47582	0,50395	0,48556	0,48739	0,491048
Variance Recall	0,50014	0,49597	0,50088	0,49691	0,49613	0,498006
Common Users (MAE)	3,43438	3,33891	3,34837	3,31027	3,37228	3,360842
Common Users Precision	0,50252	0,47768	0,50449	0,48711	0,49329	0,493018
Common Users Recall	0,50014	0,49623	0,501	0,49721	0,49794	0,498504

Average Total MAE	3,349
Average Total Precision	0,4925305
Average Total Recall	0,4983635

Σχετικά γραφήματα:





## Πείραμα 3

Αποτελέσματα 3<sup>ου</sup> πειράματος:

Experiment 3	N=5	N=10	N=15	N=20	N=25	AVG
Weighted Average (MAE)	3,40514	3,52458	3,40739	3,47534	3,38893	3,440276
Weighted Average Precision	0,5053	0,48481	0,48011	0,50346	0,50914	0,496564
Weighted Average Recall	0,50048	0,49785	0,4952	0,50097	0,50319	0,499538
Adjustment (MAE)	3,33879	3,46933	3,37243	3,44456	3,34713	3,394448
Adjustment Precision	0,5053	0,48481	0,48011	0,50346	0,50873	0,496482
Adjustment Recall	0,50048	0,49785	0,4952	0,50097	0,50304	0,499508
Variance (MAE)	3,40512	3,52483	3,40789	3,47554	3,38847	3,44037
Variance Precision	0,50186	0,48481	0,47813	0,50999	0,50954	0,496866
Variance Recall	0,50016	0,49785	0,49481	0,50276	0,50335	0,499786
Common Users (MAE)	3,40508	3,52477	3,40766	3,47556	3,38841	3,440296
Common Users Precision	0,5053	0,48481	0,4788	0,5067	0,50873	0,496868
Common Users Recall	0,50048	0,49785	0,49494	0,50187	0,50304	0,499636

Average Total MAE	3,428848
Average Total Precision	0,496695
Average Total Recall	0,499617

## Σχετικά γραφήματα:



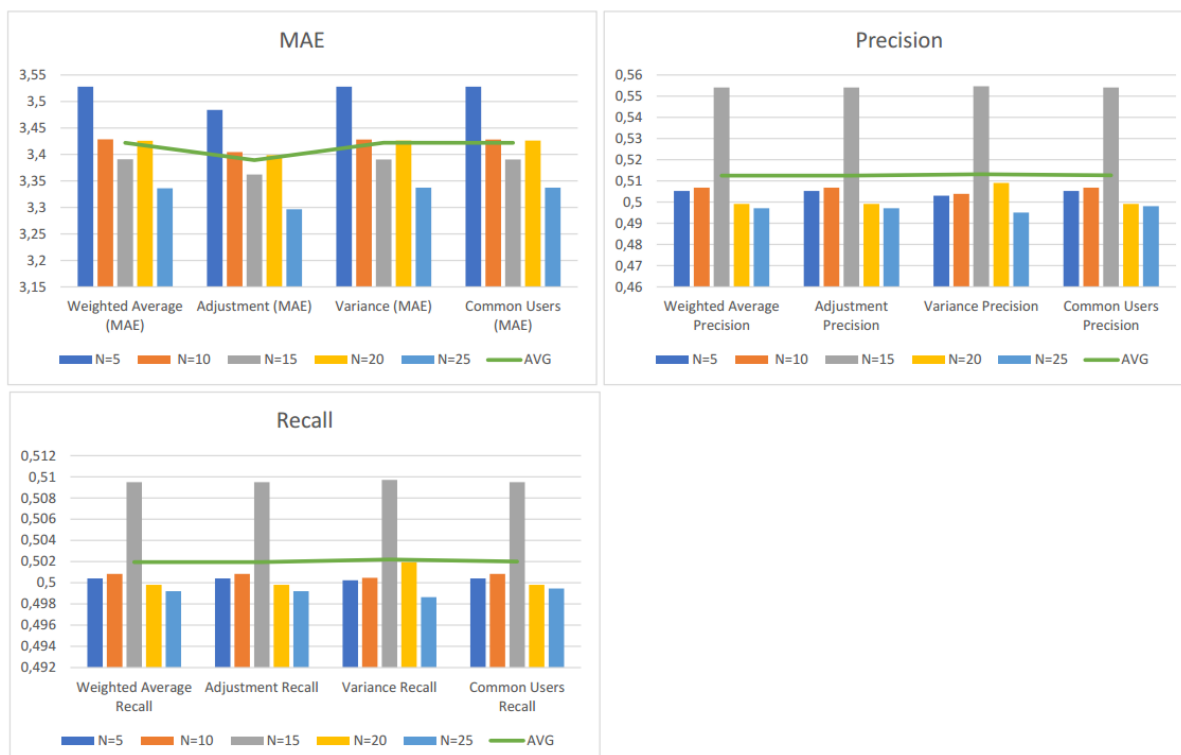
## Πείραμα 4

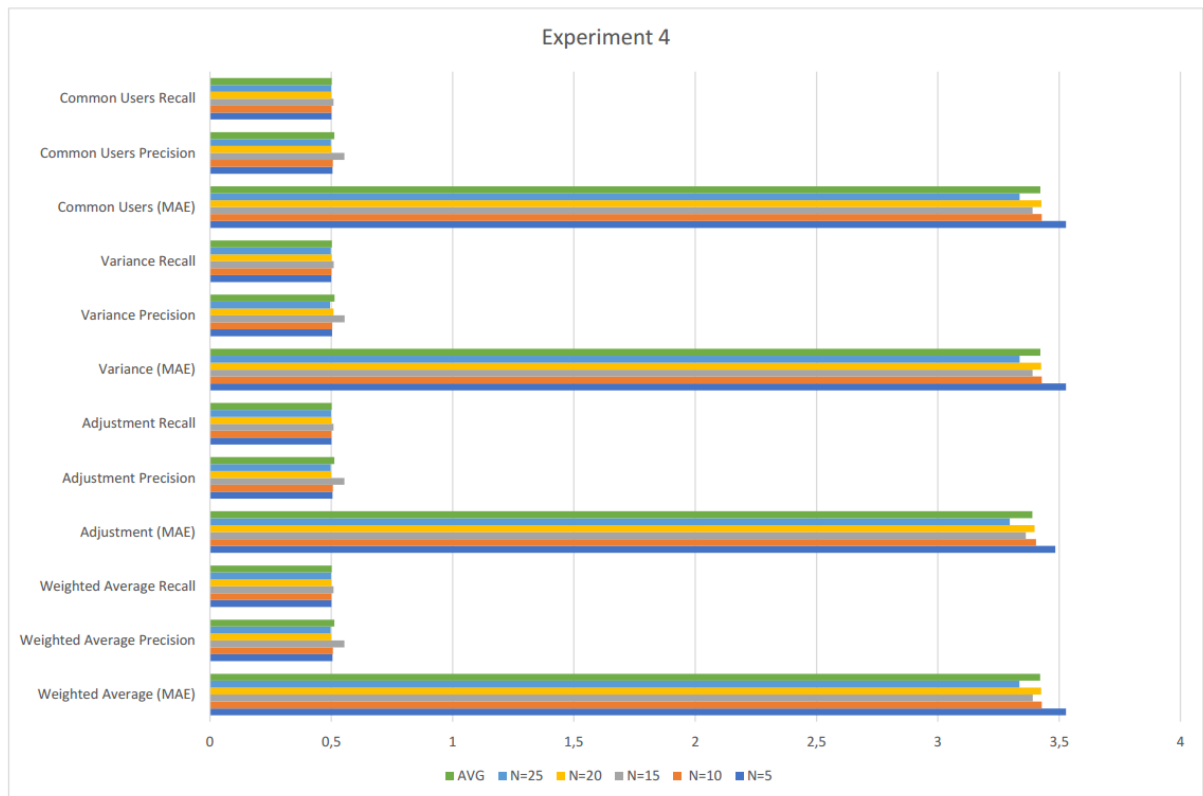
Αποτελέσματα 4<sup>ου</sup> πειράματος:

Experiment 4	N=5	N=10	N=15	N=20	N=25	AVG
Weighted Average (MAE)	3,52795	3,42848	3,39101	3,42564	3,33612	3,42184
Weighted Average Precision	0,5053	0,5068	0,55409	0,49919	0,49717	0,51251
Weighted Average Recall	0,50041	0,50085	0,50949	0,49982	0,49921	0,501956
Adjustment (MAE)	3,48417	3,40434	3,36231	3,39867	3,29672	3,389242
Adjustment Precision	0,5053	0,5068	0,55409	0,49919	0,49717	0,51251
Adjustment Recall	0,50041	0,50085	0,50949	0,49982	0,49921	0,501956
Variance (MAE)	3,52792	3,4283	3,39066	3,42621	3,33742	3,422102
Variance Precision	0,50303	0,5039	0,55465	0,50907	0,49513	0,513156
Variance Recall	0,50023	0,50047	0,50971	0,50195	0,49864	0,5022
Common Users (MAE)	3,52792	3,4283	3,39064	3,42633	3,33714	3,422066
Common Users Precision	0,5053	0,5068	0,55409	0,49919	0,49808	0,512692
Common Users Recall	0,50041	0,50085	0,50949	0,49982	0,49947	0,502008

Average Total MAE	3,413813
Average Total Precision	0,512717
Average Total Recall	0,50203

Σχετικά γραφήματα:





## Πείραμα 5

Αποτελέσματα 5<sup>ου</sup> πειράματος:

Experiment 5	N=5	N=10	N=15	N=20	N=25	AVG
Weighted Average (MAE)	3,5128	3,41661	3,40356	3,4411	3,42915	3,440644
Weighted Average Precision	0,46675	0,51342	0,47554	0,50887	0,5005	0,493016
Weighted Average Recall	0,49697	0,50226	0,49292	0,50289	0,50017	0,499042
Adjustment (MAE)	3,46624	3,36462	3,34658	3,40838	3,40168	3,3975
Adjustment Precision	0,46675	0,51342	0,47554	0,50887	0,5005	0,493016
Adjustment Recall	0,49697	0,50226	0,49292	0,50289	0,50017	0,499042
Variance (MAE)	3,51307	3,4167	3,40456	3,44157	3,42893	3,440966
Variance Precision	0,46675	0,51342	0,46838	0,50803	0,50488	0,492292
Variance Recall	0,49697	0,50226	0,49115	0,50259	0,50163	0,49892
Common Users (MAE)	3,51303	3,41661	3,40443	3,44136	3,4288	3,440846
Common Users Precision	0,46675	0,51412	0,47421	0,50887	0,5009	0,49297
Common Users Recall	0,49697	0,5024	0,49262	0,50289	0,5003	0,499036

Average Total MAE	3,429989
Average Total Precision	0,492824
Average Total Recall	0,49901

## Σχετικά γραφήματα:





## Μέσος όρος αποτελεσμάτων

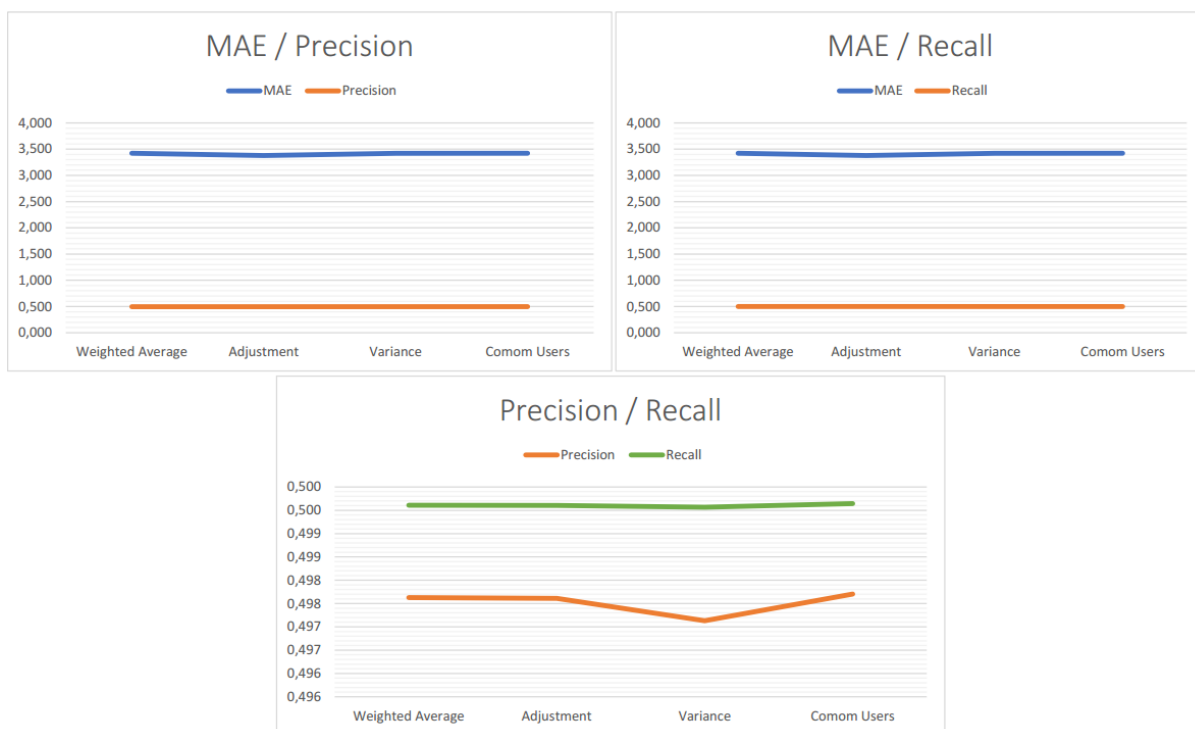
Για λόγους απλοποίησης της παρουσίασης των δεδομένων έγινε στρογγυλοποίηση σε 3 δεκαδικά ψηφία. Τα αποτελέσματα της στήλης AVG (Average) του κάθε πειράματος συνυπολογίστηκαν και προέκυψε η στήλη Total Average καθώς και οι συνολικές τιμές για τα MAE, Precision και Recall.

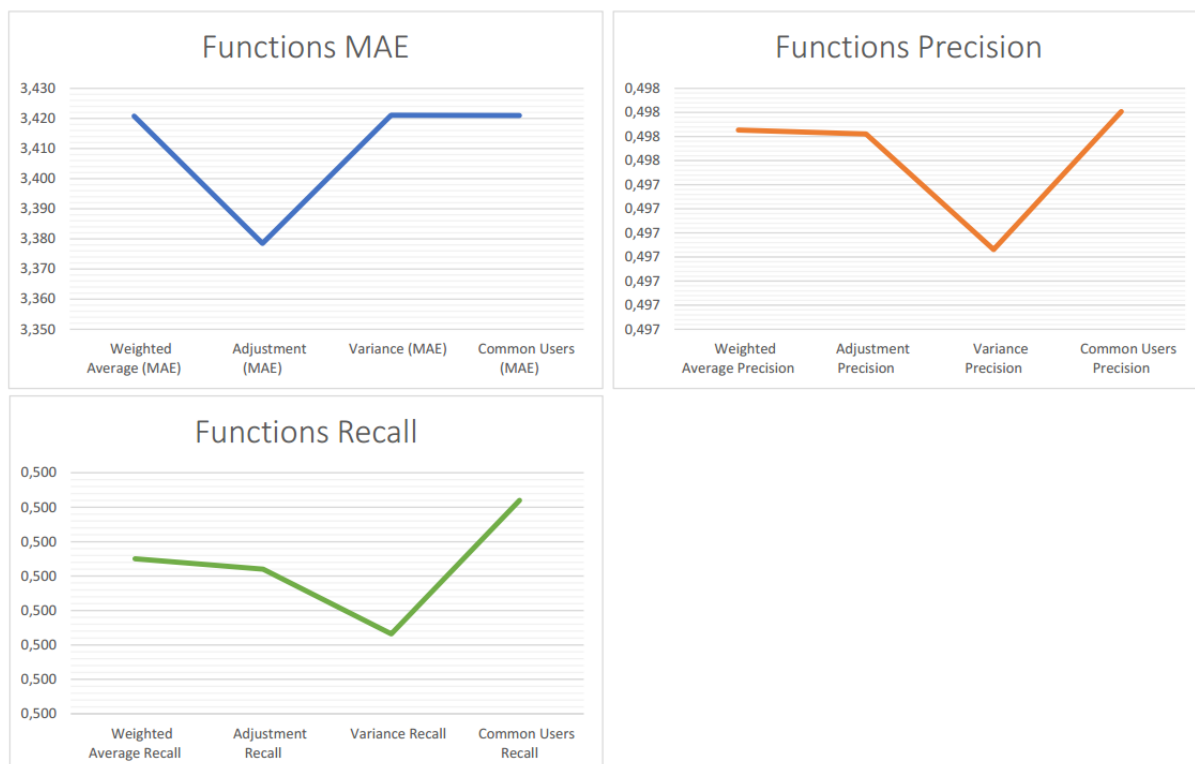
Functions	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Total Average
Weighted Average (MAE)	3,441	3,360	3,440	3,422	3,441	3,421
Weighted Average Precision	0,493	0,493	0,497	0,513	0,493	0,498
Weighted Average Recall	0,499	0,498	0,500	0,502	0,499	0,500
Adjustment (MAE)	3,398	3,314	3,394	3,389	3,398	3,379
Adjustment Precision	0,493	0,493	0,496	0,513	0,493	0,498
Adjustment Recall	0,499	0,498	0,500	0,502	0,499	0,500
Variance (MAE)	3,441	3,361	3,440	3,422	3,441	3,421
Variance Precision	0,492	0,491	0,497	0,513	0,492	0,497
Variance Recall	0,499	0,498	0,500	0,502	0,499	0,500
Common Users (MAE)	3,441	3,361	3,440	3,422	3,441	3,421
Common Users Precision	0,493	0,493	0,497	0,513	0,493	0,498
Common Users Recall	0,499	0,499	0,500	0,502	0,499	0,500

Total Average MAE	3,410
Total Average Precision	0,498
Total Average Recall	0,500

Functions/Values	Weighted Average	Adjustment	Variance	Comom Users
MAE	3,421	3,379	3,421	3,421
Precision	0,498	0,498	0,497	0,498
Recall	0,500	0,500	0,500	0,500

Τα σχετικά γραφήματα που παρουσιάζουν το σύνολο των αποτελεσμάτων για τον μέσο όρο των 5 πειραμάτων:





### Παρατηρήσεις:

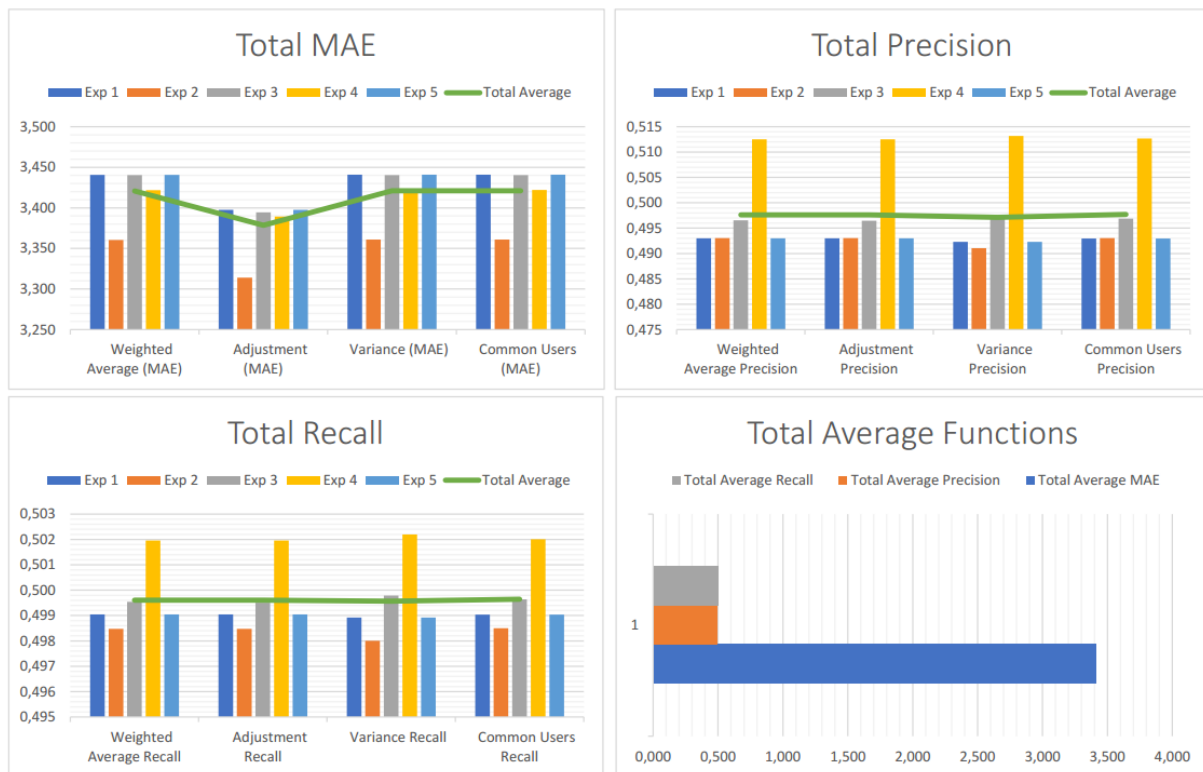
Στα παραπάνω διαγράμματα διακρίνονται οι μεταβολές των μέσων τιμών για τα MAE, Precision και Recall. Οι μεταβολές αυτές είναι αμελητέες καθώς η καμπύλη που προκύπτει δεν μεταβάλλεται σε καμία περίπτωση περισσότερο από μερικές δεκαδικές μονάδες. Συγκεκριμένα στο MAE η μεταβολή είναι περίπου στο 0,05 ενώ για τις άλλες δυο μετρικές είναι συγκριτικά πολύ μικρότερες έως μηδενικές.

Συμπερασματικά θα μπορούσαμε να καταλήξουμε στις εξής τιμές:

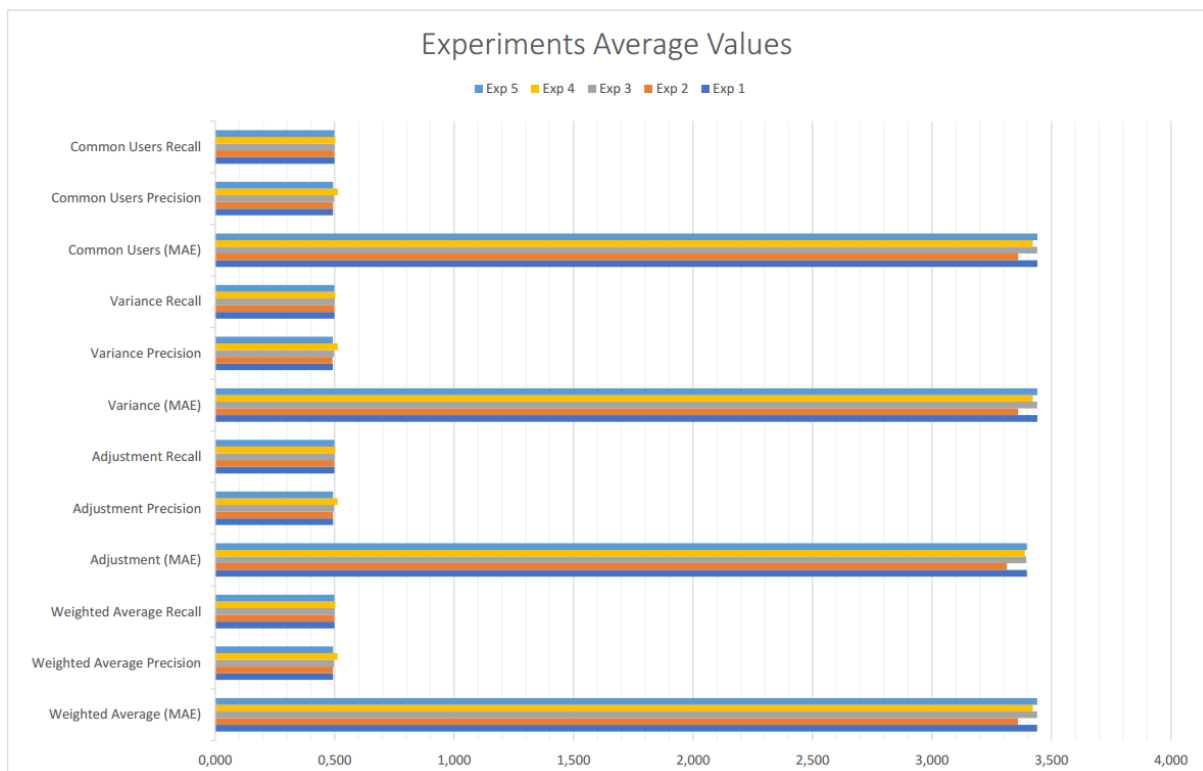
**MAE = 3,40**

**Precision = 0,49**

**Recall = 0,50**



Τελος, παρουσιάζονται οι συνολικές μέσες τιμές για κάθε συνάρτηση καθώς και κάθε διαφορετική περίπτωση πειράματος. Μικρή παρατήρηση σε αυτά τα πειράματα θα μπορούσε να είναι η σχετικά υψηλότερη τιμή για το πείραμα 4. Η οποία και πάλι όμως δεν αποκλίνει σημαντικά από τον μέσο όρο καθώς η διαφορά αυτή εντοπίζεται σε επίπεδο μερικών δεκαδικών ψηφίων.



## Αρχεία

Στο αρχείο του τελικού παραδοτέου περιέχονται:

Ο κώδικας του πειράματος στα αρχεία *filter-ratings.py* και *recommender.py* στον φάκελο scripts.

Τα αποτελέσματα σε μορφή φύλλου excel και σε αρχεία pdf στο φάκελο results.

Το αρχικό και το μειωμένο dataset *ratings.csv* και *ratings-reduced.csv*.