



ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ
ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΕΥΦΥΕΙΣ ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΔΙΚΤΥΟΥ - WEBINTELLIGENCE

**ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΕΠΙΣΤΗΜΟΜΕΤΡΙΑΣ
ΑΚΑΔΗΜΑΪΚΟΥ ΠΡΟΣΩΠΙΚΟΥ ΤΜΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΕΛΛΑΔΑ ΚΑΙ ΕΞΟΡΥΞΗ
ΓΝΩΣΗΣ**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΗΛΙΑΝΑΣ ΜΑΤΖΙΑΡΗ

Επιβλέπων : Στέφανος Ουγιάρογλου
Επίκουρος Καθηγητής

Θεσσαλονίκη, Μάρτιος 2022



ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΕΥΦΥΕΙΣ ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΔΙΚΤΥΟΥ – WEB
INTELLIGENCE

ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΕΠΙΣΤΗΜΟΜΕΤΡΙΑΣ ΑΚΑΔΗΜΑΪΚΟΥ ΠΡΟΣΩΠΙΚΟΥ ΤΜΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΕΛΛΑΔΑ ΚΑΙ ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΗΛΙΑΝΑΣ ΜΑΤΖΙΑΡΗ

Επιβλέπων : Στέφανος Ουγιάρογλου
Επίκουρος Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις Choose a date.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Όνομα Επώνυμο

Choose an item. ΔΙ.ΠΑ.Ε.

.....
Όνομα Επώνυμο

Choose an item. ΔΙ.ΠΑ.Ε.

.....
Όνομα Επώνυμο

Choose an item. ΔΙ.ΠΑ.Ε.

Θεσσαλονίκη, Μάρτιος 2022

(Υπογραφή)

.....

Click here to enter text.

Click here to enter text.

© Choose a date– Allrightsreserve

Ευχαριστίες

Κατά τη διάρκεια εκπόνησης της διπλωματικής εργασίας μου, θα ήθελα να ευχαριστώ τον επιβλέποντα καθηγητή μου κ. Ουγιάρογλου Στέφανο, για την άριστη συνεργασία που είχαμε, την καθοδήγηση, τις πολύτιμες συμβουλές του και την άμεση βοήθεια που μου παρείχε όποτε χρειάστηκε.

Η ολοκλήρωση όμως αυτής της εργασίας, δεν θα μπορούσε να υλοποιηθεί αν δεν υπήρχε η υποστήριξη του συζύγου και του παιδιού μου, η κατανόηση των γονιών μου και η παρότρυνση της αδερφής του συζύγου μου.

Περίληψη

Η αξιολόγηση των αποτελεσμάτων της έρευνας που διεξάγεται από επιστήμονες αποτελεί ένα πεδίο που απασχολεί όλο και περισσότερο τους ερευνητικούς και κρατικούς φορείς, καθώς η παραγωγικότητα και απήχηση του επιστημονικού έργου καταδεικνύει τις τάσεις και τον προσανατολισμό της έρευνας και διαδραματίζει σημαντικό ρόλο στις στρατηγικές λήψεις αποφάσεων και στη λήψη χρηματοδοτήσεων. Αν και η αξιολόγηση από ομότιμους κριτές (peer review) συνεχίζει να αποτελεί μια διαδεδομένη ποιοτική μέθοδο αποτίμησης του ερευνητικού έργου, ωστόσο, οι δυνατότητες και τα εργαλεία που προσφέρονται μέσω του διαδικτύου και της Επιστήμης της Πληροφορίας, παρέχουν εναλλακτικές μεθόδους αξιολόγησης. Η επιστημομετρική ανάλυση αποτελεί μια μέθοδο για την ποσοτική έκφραση του ερευνητικού έργου που παράγεται, καταγράφοντας και μετρώντας τις επιδόσεις των ερευνητών και τη συμβολή του έργου τους στην επιστημονική κοινότητα.

Η παρούσα διπλωματική εργασία επικεντρώνεται στην ανάλυση των δεδομένων επιστημομετρίας του ακαδημαϊκού προσωπικού των τμημάτων Πληροφορικής στην Ελλάδα και στην εξόρυξη γνώσης. Το πρώτο μέρος της εργασίας εστιάζει στην αξιολόγηση του ερευνητικού έργου των 33 τμημάτων πληροφορικής που λειτουργούν πανελλαδικά, μέσω στατιστικής ανάλυσης. Κύριος στόχος της μελέτης αποτέλεσε η καταγραφή και ανάλυση των επιστημομετρικών δεδομένων του ακαδημαϊκού προσωπικού των εν λόγω τμημάτων και η αξιολόγηση των επιδόσεών τους στην έρευνα. Τα δεδομένα συλλέχθηκαν μέσω της μηχανής αναζήτησης Google Scholar και η ανάλυση βασίστηκε στη χρήση επιστημομετρικών δεικτών, που εξετάζουν την παραγωγικότητα των ερευνητών και τη συνολική απήχηση του επιστημονικού τους έργου, ενώ ταυτόχρονα εξετάστηκε η δυνατότητα εξέλιξής τους σε επόμενη βαθμίδα.

Το δεύτερο μέρος της εργασίας επικεντρώνεται στην εξόρυξη γνώσης από τα επιστημομετρικά δεδομένα του ακαδημαϊκού προσωπικού που καταγράφηκε. Για την εφαρμογή της εξόρυξης γνώσης χρησιμοποιήθηκε το λογισμικό Weka και εκτελέστηκαν αλγόριθμοι κατηγοριοποίησης και συσταδοποίησης. Στόχος μας είναι η αναζήτηση κοινών ιδιοτήτων και ο εντοπισμός ομάδων με κοινά χαρακτηριστικά, που θα προσφέρουν επιπρόσθετες πληροφορίες για τα μέλη Δ.Ε.Π. των πανεπιστημιακών τμημάτων πληροφορικής και θα διαμορφώσουν ένα μοντέλο πρόβλεψης για την ακαδημαϊκή τους βαθμίδα.

Λέξεις Κλειδιά: Επιστημομετρία, αξιολόγηση, τμήματα πληροφορικής, εξόρυξη γνώσης, κατηγοριοποίηση, συσταδοποίηση, αλγόριθμοι, Weka

Abstract

The evaluation of scientific work is a field that increasingly concerns research and government agencies, since the productivity and impact of scientific work demonstrates research trends and orientation, and plays an important role in strategic decision-making and funding as well. Although peer review continues to be a widespread qualitative method of evaluating scientific work, the possibilities and tools offered through the internet and Information Science provide alternative evaluation methods. Scientometric analysis is a method for the quantitative expression of the research produced, which records and measures the performance of researchers and the contribution of their work to the scientific community.

The present postgraduate thesis focuses on the analysis of scientometric data of the academic members of the departments of Informatics in Greece and on data mining. The first part of the thesis focuses on the evaluation of scientific work performed at 33 IT departments operating nationwide, through statistical analysis. The main objective is to record and analyze the scientometric data of the academic members and to evaluate their performance in research. Data were collected through the Google Scholar search engine and the analysis was based on the use of scientometric indicators, which examine the productivity of researchers and the overall impact of their scientific work, while examining the possibility of their development to the next academic level.

The second part of the thesis focuses on data mining, resulting from scientometric data of the recorded academic members. Weka software was used to perform data mining, including the application of categorization and clustering algorithms. The main goal is to search for common properties and to identify groups with common characteristics, which will offer additional and valuable information concerning the academic members of the IT departments, and create a predictive model for their academic position.

Keywords: Scientometrics, evaluation, IT departments, data mining, categorization, clustering, algorithms, Weka

Πίνακας περιεχομένων

Ευχαριστίες.....	i
Περίληψη	ii
Abstract	iii
Πίνακας περιεχομένων	iv
Κατάλογος Πινάκων.....	vi
Κατάλογος Εικόνων.....	vii
Κατάλογος Γραφημάτων	viii
ΚΕΦΑΛΑΙΟ	
ΚΕΦΑΛΑΙΟ 1: Εισαγωγή.....	1
1.1 Επιστημομετρία (Scientometrics)	1
1.2 Διαδικτυακοί τόποι συλλογής δεδομένων επιστημομετρίας.....	4
1.3 Μέλη ακαδημαϊκού προσωπικού πανεπιστημιακών τμημάτων.....	6
1.4 Κίνητρο	7
1.5 Συνεισφορά	8
1.6 Οργάνωση της διπλωματικής εργασίας	9
ΚΕΦΑΛΑΙΟ 2: Θεωρητικό υπόβαθρο	11
2.1 Εξόρυξη γνώσης	11
2.2 Τεχνικές εξόρυξης γνώσης.....	13
2.3 Αλγόριθμοι κατηγοριοποίησης	15
2.3.1 Μπεϋζιανός ταξινομητής	15
2.3.2 Δέντρα απόφασης	17
2.3.3 Κατηγοριοποίηση πλησιέστερου γείτονα (K-nearest neighbor).....	20
2.4 Αλγόριθμοι συσταδοποίησης.....	22
2.3.1 Αλγόριθμος K-Means	22
ΚΕΦΑΛΑΙΟ 3: Το σύνολο Δεδομένων	25
ΚΕΦΑΛΑΙΟ 4: Στατιστική Ανάλυση των δεδομένων.....	28
4.1 Συνολικά επιστημομετρικά δεδομένα τμημάτων Πανελλαδικά	29
4.2 Αξιολόγηση μελών Δ.Ε.Π. Πανελλαδικά	39
4.3 Αξιολόγηση μελών Δ.Ε.Π. ανά Περιφερειακή Ενότητα.....	46
ΚΕΦΑΛΑΙΟ 5: Εξόρυξη Γνώσης	68
5.1 Το λογισμικό Weka.....	73
5.2 Εφαρμογή εξόρυξης γνώσης από τα δεδομένα επιστημομετρίας στο Weka	74
5.3 Αλγόριθμοι ατηγοριοποίησης.....	74

5.3.1	Μπεϋζιανός ταξινομητής	72
5.3.2	Δέντρα απόφασης	75
5.3.3	Αλγόριθμος IBk (K-nearest neighbor).....	78
5.4	Αλγόριθμος συσταδοποίησης K-Means.....	80
ΚΕΦΑΛΑΙΟ 6: Συμπεράσματα - Μελλοντική έρευνα.....		81
6.1	Συμπεράσματα	81
6.2	Μελλοντική Έρευνα.....	85
ΒΙΒΛΙΟΓΡΑΦΙΑ		86

Κατάλογος πινάκων

Πίνακας 4.13. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Ήπειρο.....	55
Πίνακας 4.14. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στο Ιόνιο.....	57
Πίνακας 4.15. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Πελοπόννησο.....	58
Πίνακας 4.16. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στη Δυτική Ελλάδα.....	59
Πίνακας 4.17. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Κρήτη.....	61
Πίνακας 4.18. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στο Βόρειο Αιγαίο.....	64
Πίνακας 4.19. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Αττική.....	64
Πίνακας 5.1. Τιμές SSE ανά k	79
Πίνακας 5.2. Τιμές SSE ανά k (2 επανάληψη του αλγόριθμου).....	86

Κατάλογος εικόνων

Εικόνα 2.1. Παράδειγμα εφαρμογής αλγόριθμου KNN.....	21
Εικόνα 4.1. Χάρτης της Ελλάδας που παρουσιάζει τις περιφερειακές ενότητες.....	46
Εικόνα 5.1. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο Naïve Bayes.....	72
Εικόνα 5.2. Το σύνολο των δεδομένων πριν την εφαρμογή της SMOTE.....	73
Εικόνα 5.3. Το σύνολο των δεδομένων μετά την εφαρμογή της SMOTE.....	74
Εικόνα 5.4. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο Naïve Bayes και τη χρήση της μεθόδου SMOTE.....	75
Εικόνα 5.5. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. με τον αλγόριθμο J48.....	76
Εικόνα 5.6. Απεικόνιση του Δέντρου Απόφασης	77
Εικόνα 5.7. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο J48 και τη χρήση της μεθόδου SMOTE.....	77
Εικόνα 5.8. Απεικόνιση του Δέντρου Απόφασης με τη χρήση της μεθόδου SMOTE.....	78
Εικόνα 5.9. Προσδιορισμός k-πλησιέστερου μέσω WEKA.....	79
Εικόνα 5.10. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο IBk.....	79
Εικόνα 5.11. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο IBk με τη μέθοδο SMOTE.....	80
Εικόνα 5.12. Αποτελέσματα συσταδοποίησης των μελών Δ.Ε.Π. με τον αλγόριθμο K-means.....	82
Εικόνα 5.13. Αποτελέσματα συσταδοποίησης με K-means ανά τμήμα.....	83
Εικόνα 5.14. Αποτελέσματα συσταδοποίησης με K-means ανά σύνολο δημοσιεύσεων.....	83
Εικόνα 5.15. Αποτελέσματα συσταδοποίησης με K-means ανά σύνολο αναφορών.....	84
Εικόνα 5.16. Αποτελέσματα συσταδοποίησης με K-means ανά σύνολο ετών δημοσιεύσεων.....	84
Εικόνα 5.17 Διαγραφή χαρακτηριστικών.....	85
Εικόνα 5.18. Αποτελέσματα συσταδοποίησης των μελών Δ.Ε.Π. με τον αλγόριθμο K-means (2 ^η εκτέλεση).....	86

Κατάλογος γραφημάτων

Γράφημα 4.1. Σύνολο δημοσιεύσεων ανά τμήμα πληροφορικής.....	30
Γράφημα 4.2. Σύνολο αναφορών ανά τμήμα πληροφορικής.....	30
Γράφημα 4.3. Μέσος όρος δημοσιεύσεων ανά μέλος Δ.Ε.Π.....	33
Γράφημα 4.4. Μέσος όρος αναφορών ανά μέλος Δ.Ε.Π.....	33
Γράφημα 4.5. Μέσος όρος δείκτη h (h-index) ανά μέλος Δ.Ε.Π.....	36
Γράφημα 4.6. Τυπική απόκλιση δημοσιεύσεων ανά τμήμα πληροφορικής.....	38
Γράφημα 4.7. Τυπική απόκλιση αναφορών ανά τμήμα πληροφορικής.....	38
Γράφημα 4.8. Συνολικά στοιχεία μ.ο. δημοσιεύσεων και αναφορών ανά βαθμίδα Δ.Ε.Π.	39
Γράφημα 4.9. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή	40
Γράφημα 4.10. Μέσος όρος δημοσιεύσεων και αναφορών της βαθμίδας του Λέκτορα πανελλαδικά.....	41
Γράφημα 4.11. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.....	41
Γράφημα 4.12. Μέσος όρος δημοσιεύσεων και αναφορών της βαθμίδας του Επίκουρου Καθηγητή πανελλαδικά.....	42
Γράφημα 4.13. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή (δημοσιεύσεις).....	42
Γράφημα 4.14. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή (αναφορές)...	43
Γράφημα . 4.15. Μέσος όρος δημοσιεύσεων και αναφορών της βαθμίδας του Αναπληρωτή Καθηγητή πανελλαδικά.....	44
Γράφημα . 4.16. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή (δημοσιεύσεις).....	44
Γράφημα 4.17. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή (αναφορές).....	45
Γράφημα 4.18. Συνολικά στοιχεία των μελών Δ.Ε.Π. ανά βαθμίδα στην Ανατολική Μακεδονία.....	47
Γράφημα 4.19. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.....	48
Γράφημα 4.20. Συνολικά στοιχεία των μελών Δ.Ε.Π. ανά βαθμίδα στην Κεντρική Μακεδονία.....	49
Γράφημα 4.21. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.....	49
Γράφημα 4.22. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.....	50
Γράφημα 4.23. Στατιστικά στοιχεία για τη βαθμίδα του Επίκουρου Καθηγητή.....	50
Γράφημα 4.24. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	51
Γράφημα 4.25. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.....	52
Γράφημα 4.26. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	53

Γράφημα 4.27. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	54
Γράφημα 4.28. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	55
Γράφημα 4.29. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.....	56
Γράφημα 4.30. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	56
Γράφημα 4.31. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.....	57
Γράφημα 4.32. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	58
Γράφημα 4.33. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	59
Γράφημα 4.34. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.....	60
Γράφημα 4.35. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.....	60
Γράφημα 4.36. Άτομα για εξέλιξη στη βαθμίδα του Λέκτορα.....	60
Γράφημα 4.37. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.....	62
Γράφημα 4.38. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	63
Γράφημα 4.39- 4.40. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.....	65
Γράφημα 4.41. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.....	66
Γράφημα 4.42. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.....	67
Γράφημα 5.1. Μέθοδος αγκώνα για την επιλογή του k	79
Γράφημα 5.2. Μέθοδος αγκώνα για την επιλογή του k	86

Κεφάλαιο 1

Εισαγωγή

1.1. Επιστημομετρία (Scientometrics)

Η αξιολόγηση των αποτελεσμάτων της έρευνας αποτελεί ένα πεδίο που έχει αναπτυχθεί ευρέως τις τελευταίες δεκαετίες. Τόσο τα πανεπιστήμια και τα ερευνητικά κέντρα, όσο και κρατικοί φορείς προσανατολίζονται στη χρήση μεθόδων για την ποιοτική και ποσοτική αξιολόγηση των επιστημονικών δεδομένων. Η ποιοτική αξιολόγηση του ερευνητικού έργου, με κυριότερη την «αξιολόγηση από ομότιμους κριτές» (peer review), συνεχίζει ν' αποτελεί μια μέθοδο με ευρεία αποδοχή από την επιστημονική κοινότητα. Ωστόσο, η μέθοδος αυτή είναι χρονοβόρα, απαιτεί σημαντικούς ανθρώπινους πόρους, ενώ δε λείπει και το υποκειμενικό στοιχείο που μπορεί να επηρεάζεται από προκαταλήψεις και τη συγκεκριμένη επιστημονική στόχευση των κριτών [1] [2].

Από την άλλη πλευρά, η αποτίμηση του ερευνητικού έργου που βασίζεται στην ποσοτική ανάλυση μέσω των επιστημομετρικών μεθόδων, αποτελεί μια όλο και πιο διαδεδομένη μέθοδο αξιολόγησης των επιδόσεων στην έρευνα. Η Επιστημομετρία χρησιμοποιώντας διάφορες μετρήσεις και αναλύσεις, μπορεί να προσφέρει μια ανασκόπηση των εργασιών στον εκάστοτε τομέα έρευνας, μελετώντας το σύνολο των δημοσιεύσεων και παραπομπών του κάθε ερευνητή. Η μεγάλη απήχηση της μεθόδου οφείλεται στα πιο αντικειμενικά κριτήρια αξιολόγησης (βάσει των ανεξάρτητων μετρήσεων), παρουσιάζει πολλαπλά επίπεδα ανάλυσης, ενώ ταυτόχρονα πρόκειται για μια οικονομική και ταχύτατη διαδικασία [3].

Η Επιστημομετρία (Scientometrics) θεμελιώθηκε ως επιστήμη από τον Derek de Solla Prince στις αρχές της δεκαετίας του '60, και εξελίχθηκε τις επόμενες δεκαετίες. Ωστόσο, ο όρος Επιστημομετρία επινοήθηκε το 1969 από τους Nalimov και Mulchenko και ορίζεται ως η «ποσοτική μελέτη της επιστήμης και της τεχνολογίας» [4]. Πρόκειται ουσιαστικά για μια επιστήμη που μελετά και αξιολογεί την ανάπτυξη και την εξέλιξη της επιστημονικής δραστηριότητας και των τεχνολογικών επιτευγμάτων χρησιμοποιώντας μαθηματικά και στατιστικά μοντέλα [5].

Η Επιστημομετρία αποτελεί πεδίο του κλάδου της Επιστήμης της Πληροφορίας. Επικαλύπτει και επικαλύπτεται από διάφορα παρεμφερή επιστημονικά πεδία που χρησιμοποιούν

μαθηματικές και στατιστικές τεχνικές. Ένα από τα πιο διαδεδομένα πεδία είναι η Βιβλιομετρία ή βιβλιομετρική ανάλυση που αναπτύχθηκε τον 20^ο αι. από το ενδιαφέρον των ερευνητών να παρακολουθήσουν τις εξελίξεις των επιστημών [6]. Ο όρος βιβλιομετρία προτάθηκε το 1969 από τον Alan Pritchard και ορίστηκε ως «η εφαρμογή των μαθηματικών και στατιστικών μεθόδων στα βιβλία και τα άλλα μέσα δημοσίευσης» [7]. Συχνά, η Βιβλιομετρία συγγέεται με την Επιστημομετρία, καθώς ως πεδία παρουσιάζουν αρκετές ομοιότητες στις μεθόδους και τα εργαλεία που χρησιμοποιούν. Διαφοροποιούνται, ωστόσο, καθώς η βιβλιομετρία χρησιμοποιείται για να περιγράψει την ποσοτική μελέτη όλων των πτυχών της βιβλιογραφίας της επιστήμης και της τεχνολογίας, ενώ η δεύτερη περιλαμβάνει μετρήσεις και εργαλεία που δε σχετίζονται μόνο με τη βιβλιογραφία (π.χ. οι πρακτικές των επιστημόνων, η διοίκηση έρευνας και ανάπτυξης, κτλ) [8].

Από τα τέλη της δεκαετίας του '70 ξεκίνησε η έκδοση επιστημονικών περιοδικών εξειδικευμένων στα επιστημομετρικά και βιβλιομετρικά θέματα και την ερευνητική αξιολόγηση (Scientometrics, Research Evaluation, κτλ.), δημοσιεύτηκαν αρκετά σχετικά συγγράμματα, ενώ διοργανώθηκαν και τα πρώτα διεθνή συνέδρια [6]. Η ανάπτυξη της επιστήμης των υπολογιστών και της τεχνολογίας συνέβαλλε στη γρήγορη εξέλιξη της Βιβλιομετρίας και της Επιστημομετρίας. Μάλιστα, τις τελευταίες δεκαετίες έχουν αναπτυχθεί και διάφορα άλλα παραπλήσια πεδία της Επιστήμης της Πληροφορίας που χρησιμοποιούν μαθηματικές και στατιστικές τεχνικές. Έτσι, κάποια εστιάζουν στην ποσοτική μέτρηση της πληροφορίας οποιασδήποτε μορφής στο διαδίκτυο, χωρίς να περιορίζεται μόνο στον τομέα της επιστήμης και της τεχνολογίας (πληροφοριομετρία - Informetrics), στην ανάλυση του επιστημονικού περιεχομένου του διαδικτύου (ιστομετρία - Webometrics), ή την ποσοτικοποίηση των ιστοσελίδων και τη μέτρηση των ηλεκτρονικών πηγών πληροφοριών (Κυβερνομετρία - Cybermetrics) [5].

Η Επιστημομετρία χρησιμοποιείται συχνά για την αξιολόγηση της παραγόμενης επιστημονικής έρευνας, μέσω ποσοτικών μελετών των δημοσιεύσεων. Ουσιαστικά πρόκειται για την ποσοτική έκφραση ερευνητικού έργου, το οποίο έχει δημοσιευθεί στη διεθνή ακαδημαϊκή κοινότητα. Το πεδίο εφαρμογής της είναι πολύ μεγάλο και καλύπτει όλους τους επιστημονικούς κλάδους. Παρέχει τη γενική εικόνα ενός συγκεκριμένου ερευνητικού πεδίου, χρησιμοποιώντας ως εργαλεία τις μεγάλες βάσεις δεδομένων που είναι προσβάσιμες στο διαδίκτυο.

Η σημασία της Επιστημομετρίας έγκειται στην παροχή επιστημονικών πληροφοριών για το εκάστοτε επιστημονικό πεδίο και την παρακολούθηση του προσανατολισμού της έρευνας, των εξελίξεων και των νέων τάσεων. Σε ευρύτερο πλαίσιο, η αξιολόγηση του ερευνητικού

έργου που συνήθως εφαρμόζεται από πανεπιστήμια, ερευνητικούς και κρατικούς φορείς επηρεάζει τόσο τη χρηματοδότηση της έρευνας, όσο και τις στρατηγικές λήψης αποφάσεων.

Επιστημονομετρικές μέθοδοι και δείκτες μέτρησης της έρευνας

Η πιο διαδεδομένη μέθοδος που εφαρμόζει σήμερα η Επιστημομετρία είναι η ανάλυση αναφορών (Citation Analysis). Παρουσιάστηκε πρώτη φορά από τον Eugene Garfield, ο οποίος περιέγραψε την ιδέα ενός δείκτη αναφορών των επιστημονικών δημοσιεύσεων [9]. Η μέθοδος αυτή αποτελεί πηγή δεδομένων για τη διεξαγωγή πολλών ειδών επιστημονομετρικών αναλύσεων. Με τη χρήση των αναφορών μπορεί να μετρηθεί η επίδραση ενός συγγραφέα, ενός άρθρου ή μιας δημοσίευσης με βάση τον αριθμό των αναφορών του. Έτσι, τα αποτελέσματα της μεθόδου μας βοηθούν να αξιολογήσουμε την επιρροή του έργου ενός ερευνητή και τη σημασία του, καθώς και τις διασυνδέσεις μεταξύ συγγραφέων παρόμοιων επιστημονικών πεδίων.

Για την ανάλυση των στοιχείων που αντλούνται και τις μετρήσεις των τεκμηρίων η Επιστημομετρία χρησιμοποιεί διάφορους δείκτες. Πρόκειται για την ποσοτική έκφραση της επιστημονομετρικής ανάλυσης. Οι δείκτες αυτοί σχετίζονται με μια επιστημονική περιοχή και κατασκευάζονται από μετρήσιμες ποσότητες, όπως για παράδειγμα, ο αριθμός των επιστημονικών δημοσιεύσεων, ο οποίος αποτελεί τον απλούστερο δείκτη για την καταγραφή της παραγωγής επιστημονικών εργασιών και, κατά συνέπεια, του ερευνητικού έργου ανά επιστήμονα, οργανισμό, επιστημονικό κλάδο ή χώρα.

Οι βασικοί δείκτες ανάλυσης που συνήθως χρησιμοποιούνται στην επιστημομετρία δε διαφέρουν πολύ από αυτούς της βιβλιομετρικής ανάλυσης. Έτσι, οι δείκτες ανάλυσης μετρούν την *παραγωγικότητα (productivity)*, η οποία αφορά τον αριθμό των δημοσιεύσεων σε επιστημονικά περιοδικά κατά τη διάρκεια ενός συγκεκριμένου χρονικού πλαισίου, την *απήχηση (impract)*, που σχετίζεται με τον αριθμό των αναφορών μιας δημοσίευσης και επομένως με την επιρροή ενός ερευνητικού έργου, γ) την *αποδοτικότητα (efficiency)*, που μετρά τον αριθμό των αναφορών ανά δημοσίευση σε ποσοστιαίες μονάδες, ενώ δ) υπάρχουν και οι λεγόμενοι *υβριδικοί δείκτες (hybrid)*, που εξετάζουν τη συνεργασία αναφορών και την ανάλυση συνδυασμού λέξεων [6].

Οι υβριδικοί δείκτες (δείκτες συν- αναφοράς) μπορούν να συνδυαστούν με τη δημοσίευση και την αναφορά για την κατασκευή μιας πολύπλευρης αναπαράστασης των ερευνητικών πεδίων, της ανάπτυξης δεσμών μεταξύ τους και των συμμετεχόντων που τους διαμορφώνουν. Ο πλέον δημοφιλής και ευρύτερα χρησιμοποιούμενος δείκτης αυτής της κατηγορίας είναι ο δείκτης *h* που προτάθηκε από τον J.E. Hirsch και ορίζει ότι: «ένας επιστήμονας έχει *h*-index

ίσο με h , εάν h από σύνολο N_p των δημοσιεύσεών του έχει τουλάχιστον h αναφορές η κάθε μία, και οι υπόλοιπες δημοσιεύσεις (N_p-h) έχουν το μέγιστο h αναφορές η κάθε μία» [10]. Ουσιαστικά, ο δείκτης h , αναπτύχθηκε για να λύσει το πρόβλημα των μετρήσεων στις αναφορές, όταν για παράδειγμα, ένας ερευνητής πραγματοποιήσει μία επιτυχημένη αναφορά, ενώ οι υπόλοιπες δεν έχουν την ίδια απήχηση, ή ακόμη και στην περίπτωση συλλογικών άρθρων και αναφορών που πιστώνονται όλοι οι συγγραφείς, αλλά δεν είναι ξεκάθαρη η συνεισφορά του καθενός. Αποτελεί, δηλαδή, μια προσπάθεια συνδυασμού ποιοτικής και ποσοτικής αποτίμησης του αξιολογούμενου ερευνητικού έργου. Συνιστά μία απλή και αποδοτική προσέγγιση για την ερευνητική αξιολόγηση ενός ερευνητή.

Ο δείκτης h μπορεί να υπολογιστεί είτε σε ατομικό επίπεδο είτε ως μέσος όρος σε επίπεδο τμήματος ή και ευρύτερης ομάδας. Ωστόσο, και εδώ υπάρχουν προβλήματα. Στην περίπτωση των νέων ερευνητών δεν παρουσιάζει ικανοποιητικά αποτελέσματα, καθώς τα επίπεδα δημοσιεύσεων και αναφορών τους είναι ακόμη χαμηλά, σε σχέση με ερευνητές που δραστηριοποιούνται δεκαετίες. Επιπλέον, ο δείκτης h μπορεί να αυξηθεί ακόμη και αν δεν υπάρχουν νέες δημοσιεύσεις, αλλά αυξάνονται οι αναφορές σε παλαιότερα έργα του. Ακόμη, κρίνεται αναγκαία η σύγκριση μεταξύ ερευνητών του ίδιου γνωστικού αντικείμενου και επιστημονικού πεδίου, καθώς οι πρακτικές δημοσιεύσεων σε κάθε επιστημονικό πεδίο διαφοροποιούνται [10] [11] [12].

Τέλος, πρέπει να τονιστεί ότι οι δείκτες βασίζονται σε μια συγκριτική προσέγγιση, καθώς προσλαμβάνουν την πλήρη σημασία τους μόνο σε σύγκριση με εκείνες των άλλων ομάδων, ενώ οι απόλυτες τιμές από μόνες τους δεν είναι ενδεικτικές. Επιπλέον, η ανάλυση πρέπει να ενσωματώνει όσο το δυνατόν μεγαλύτερο όγκο δεδομένων, ώστε να επιτρέπει τη μείωση του στατιστικού σφάλματος. Η συνδυαστική χρήση ενός πλήθους επιστημομετρικών δεικτών με σκοπό τη σύγκρισή τους μπορεί να προσφέρει όσο το δυνατόν πιο αντικειμενικά αποτελέσματα και να μειώσει το σφάλμα [2] [6].

1.2. Διαδικτυακοί τόποι συλλογής δεδομένων

επιστημομετρίας

Η άντληση των επιστημομετρικών δεδομένων γίνεται συνήθως από μεγάλες μηχανές αναζήτησης στο διαδίκτυο. Οι διάφορες βάσεις που ιδρύθηκαν είτε από επιχειρήσεις, είτε από δημόσιους ή ιδιωτικούς φορείς χρησιμοποιούνται για να παρουσιάζουν τα αποτελέσματα της

επιστημονικής και τεχνολογικής δραστηριότητας. Για τους σκοπούς της επιστημονομετρίας μπορούν να αξιοποιηθούν δεδομένα είτε από μία μόνο μηχανή αναζήτησης, είτε και από περισσότερες. Ωστόσο, πρέπει να σημειωθεί ότι εφόσον κάθε μηχανή αναζήτησης θέτει διαφορετικά κριτήρια εισαγωγής των αναφορών, τα αποτελέσματα πιθανώς να διαφοροποιούνται ποσοτικά. Σημαντικό είναι το γεγονός ότι, εκτός από τις βιβλιογραφικές εγγραφές των ερευνητών, οι μηχανές αναζήτησης περιέχουν και στοιχεία για τις αναφορές των δημοσιεύσεων. Επιπλέον, είναι εύχρηστες και κάποιες από αυτές παρέχουν δωρεάν πρόσβαση.

Από την άλλη, οι μηχανές αναζήτησης εμφανίζουν και κάποιες αδυναμίες. Αρχικά, τα δεδομένα που συλλέγονται από το διαδίκτυο δεν μπορούν να θεωρηθούν ολοκληρωμένα. Επιπλέον, ευρετηριάζεται ένας μικρός αριθμός από το σύνολο των επιστημονικών περιοδικών, παρουσιάζοντας πολλές φορές χρονολογικά κενά, ενώ εμφανίζουν διαφορετική κάλυψη μεταξύ των επιστημονικών πεδίων, μεταξύ των γλωσσών και του τύπου των δημοσιεύσεων, καθώς και των χωρών προέλευσης των συγγραφέων. Ακόμη, η ύπαρξη λανθασμένων αναφορών, διπλοεγγραφών ή και αυτό-αναφορών μπορούν να επηρεάσουν αρνητικά και να οδηγήσουν σε σφάλμα κατά την ανάλυση. Σημαντικό είναι, επίσης, το γεγονός ότι εμφανίζονται λάθη που οφείλονται στην έλλειψη ομοιομορφίας τύπων για την εισαγωγή του ονόματος ενός συγγραφέα, πανεπιστημιακού τμήματος, κτλ.

Από τις πιο διαδεδομένες βάσεις αναζήτησης παγκοσμίου εμβέλειας, στις οποίες περιλαμβάνονται βιβλιογραφικές εγγραφές από το χώρο της επιστημονικής κοινότητας, καθώς και στοιχεία για τις μεταξύ τους αναφορές είναι η Web of Science, η Scopus και η Google Scholar. Εκτός από αυτές υπάρχει φυσικά πλήθος άλλων μηχανών αναζήτησης, κάποιες από τις οποίες εξειδικεύονται σε συγκεκριμένα επιστημονικά πεδία [6]. Ωστόσο, κάθε μηχανή αναζήτησης κατά τη χρήση της στην αναζήτηση αναφορών, εκτός από τα πλεονεκτήματα, παρουσιάζει και μειονεκτήματα. Για παράδειγμα, η βάση δεδομένων Scopus παρουσιάζει ευρύτερη κάλυψη από τη Web of Science, αλλά η κάλυψη των δημοσιεύσεων του επιστημονικού πεδίου των κοινωνικών επιστημών είναι σχετικά μικρή. Επιπλέον, οι παλαιότερες εγγραφές τοποθετούνται χρονικά μετά το 1966, τη στιγμή που η Web of Science παρέχει βιβλιογραφικές εγγραφές και αναφορές από το 1900. Από την άλλη, η βάση Web of Science χαρακτηρίζεται από άνισο τρόπο κάλυψης τόσο θεματικά, όσο και γεωγραφικά, ενώ καλύπτει κυρίως άρθρα περιοδικών, με αποτέλεσμα να μην περιλαμβάνονται μονογραφίες και συλλογικοί τόμοι. Όσον αφορά τη βάση Google Scholar, αυτή καλύπτει έναν τεράστιο όγκο πηγών, όπως άρθρα περιοδικών, πρακτικά συνεδρίων, μονογραφίες, διπλωματικές εργασίες, διδακτορικές διατριβές, ανάτυπα και τεχνικές αναφορές που διατίθενται από εμπορικούς οίκους, ιδρυματικά αποθετήρια και πολλούς άλλους τύπους υλικού. Η πρόσβαση είναι ελεύθερη και καλύπτει ένα μεγάλο εύρος επιστημονικών τεκμηρίων. Ωστόσο, η κάλυψη

πηγών θεωρείται αδιαφανής, η ανάκτηση πολλαπλών εγγραφών για την ίδια δημοσίευση (διπλοεγγραφές) αποτελεί σημαντικό πρόβλημα, ενώ η κάλυψη μη ακαδημαϊκών και μη αξιολογημένων από ομότιμους δημοσιεύσεων, απαρτίζει μεγάλο ποσοστό του υλικού.

Διαπιστώνεται, λοιπόν, ότι η κάθε μηχανή αναζήτησης εμφανίζει αρκετά πλεονεκτήματα, αλλά και μειονεκτήματα, για την άντληση επιστημονικών δεδομένων. Σε κάποιες περιπτώσεις, η επιστημονική έρευνα μπορεί να απαιτεί τη συνδυαστική χρήση διαφορετικών βάσεων δεδομένων. Γενικότερα, οι χρήστες των επιστημονικών δεικτών θα πρέπει να επιλέγουν τη βάση δεδομένων που ανταποκρίνεται καλύτερα στις ιδιαίτερες ανάγκες τους και στους στόχους της εκάστοτε έρευνας.

1.3. Μέλη ακαδημαϊκού προσωπικού πανεπιστημιακών τμημάτων

Τα μέλη που σήμερα απαρτίζουν το διδακτικό προσωπικό ενός πανεπιστημιακού τμήματος έχουν θεσπιστεί από την Ελληνική Κυβέρνηση σύμφωνα με το Νόμο 1268 για τη δομή και λειτουργία των Ανώτατων Εκπαιδευτικών Ιδρυμάτων (Ν.1268/1982- ΦΕΚ 87/Τεύχος Α/ 16-7-1982).

Οι βαθμίδες των μελών Δ.Ε.Π. ενός πανεπιστημιακού ιδρύματος είναι οι ακόλουθες: Ομότιμος Καθηγητής (Emeritus), Καθηγητής, Αναπληρωτής Καθηγητής και Επίκουρος ΚαθηγητήςΚαθηγητής. Οι Καθηγητές και οι Αναπληρωτές Καθηγητές είναι μόνιμοι, ενώ οι Επίκουροι εκλέγονται για τριετή θητεία, με δυνατότητα ανανέωσης ύστερα από κρίση. Τα μέλη Δ.Ε.Π. αποχωρούν υποχρεωτικά στο τέλος του ακαδημαϊκού έτους, στο οποίο έχουν συμπληρώσει το εξηκοστό έβδομο έτος της ηλικίας τους. Οι Καθηγητές πρώτης βαθμίδας που έχουν διακριθεί διεθνώς ως επιστήμονες, λαμβάνουν τον τίτλο του Ομότιμου Καθηγητή, κατόπιν απόφασης της Συγκλήτου. Επιπλέον, πρέπει να υπογραμμιστεί ότι η βαθμίδα του Λέκτορα που θεσπίστηκε με το Ν.1268/1982, έχει καταργηθεί από το 2011 (Νόμος 4009/2011- ΦΕΚ 195/Τεύχος Α/6-9-2011).

Σύμφωνα με τη νομοθεσία, όλα τα μέλη Δ.Ε.Π. κρίνονται με βάση το διδακτικό και ερευνητικό τους έργο, προκειμένου να εκλεγούν στην επόμενη βαθμίδα. Για την εξέλιξή τους, οι Επίκουροι και οι Αναπληρωτές Καθηγητές έχουν το δικαίωμα να ζητήσουν την προκήρυξη θέσης στην επόμενη βαθμίδα έχοντας υπηρετήσει στη βαθμίδα που κατέχουν το απαιτούμενο χρονικό διάστημα.

Η εκλογή και εξέλιξη γίνεται με ανοιχτή διαδικασία. Καθώς βασικό κριτήριο επιλογής για την εξέλιξη στην επόμενη βαθμίδα αποτελεί το διδακτικό και ερευνητικό έργο του μέλους Δ.Ε.Π., διαπιστώνεται ότι ο αριθμός των πρωτότυπων δημοσιεύσεων σε επιστημονικά περιοδικά και, επομένως, η αποδεδειγμένη παραγωγικότητα του εκάστοτε μέλους Δ.Ε.Π. αποτελεί καθοριστικό παράγοντα για την επιτυχή εξέλιξή του.

Λαμβάνοντας υπόψη τα παραπάνω, σχετικά με τις βαθμίδες των μελών ακαδημαϊκού προσωπικού των πανεπιστημιακών τμημάτων, στην παρούσα εργασία συμπεριλαμβάνονται και αξιολογούνται οι βαθμίδες του Καθηγητή, του Αναπληρωτή Καθηγητή και του Επίκουρου Καθηγητή. Επιπλέον, αποφασίστηκε να εξεταστεί και η βαθμίδα του Λέκτορα, καθώς κάποια μέλη του ακαδημαϊκού προσωπικού δεν εξελίχθηκαν ποτέ στις επόμενες βαθμίδες, ενώ παράλληλα οι καθηγητές εφαρμογών των τμημάτων πρώην Τ.Ε.Ι. μετεξελίχθηκαν σε Λέκτορες. Επιπρόσθετα, στην εργασία έχουν συμπεριληφθεί και τα μέλη Εργαστηριακού Διδακτικού Προσωπικού (ΕΔΠ) των τμημάτων πληροφορικής, τα οποία αποτελούν το λοιπό επιστημονικό προσωπικό ενός ιδρύματος.

1.4. Κίνητρο

Αφορμή για την υλοποίηση της παρούσας εργασίας είναι η συνεχώς αυξανόμενη ανάγκη για την αξιολόγηση του επιστημονικού έργου των ερευνητικών φορέων και των ερευνητών στα διάφορα επιστημονικά και τεχνολογικά πεδία. Λαμβάνοντας υπόψη το γεγονός ότι η παραγωγικότητα και απήχηση του επιστημονικού έργου καταδεικνύει τις τάσεις και τον προσανατολισμό της έρευνας, ότι διαδραματίζει σημαντικό ρόλο στη λήψη χρηματοδοτήσεων από τα ακαδημαϊκά ιδρύματα και τους ερευνητικούς και κρατικούς φορείς και, επιπλέον, επηρεάζει τις στρατηγικές λήψεις αποφάσεων, γίνεται αντιληπτό ότι η χρήση των μεθόδων και των εργαλείων που προσφέρει η Επιστημομετρία είναι ιδιαίτερα σημαντικές.

Πιο συγκεκριμένα, η έρευνα επικεντρώνεται στα 33 τμήματα πληροφορικής που λειτουργούν στην Ελλάδα (βλ. πίν. 1.1), με στόχο την αξιολόγηση του ερευνητικού τους έργου, τόσο σε επίπεδο τμήματος, όσο και σε επίπεδο ερευνητή (μέλους Δ.Ε.Π.), ώστε να μετρηθεί η παραγωγικότητα, αλλά και η απήχηση του έργου στην επιστημονική κοινότητα. Τα περισσότερα μέλη Δ.Ε.Π. των τμημάτων πληροφορικής εμφανίζουν ένα προφίλ ως ερευνητές/ μελετητές στη μηχανή αναζήτησης της Google Scholar, όπου και παρουσιάζουν το επιστημονικό τους έργο. Η μηχανή αναζήτησης Google Scholar παρέχει διάφορα δεδομένα επιστημομετρίας για το ακαδημαϊκό προσωπικό πληροφορικής και των συναφών ειδικοτήτων

των πανεπιστημιακών τμημάτων της Ελλάδας. Χρησιμοποιώντας τη μέθοδο και τα εργαλεία της επιστήμης της Επιστημομετρίας, επιχειρείται η αξιολόγηση και η σύγκριση των τμημάτων και των μελών Δ.Ε.Π., με βάση την παραγωγικότητα και την απήχηση του ερευνητικού τους έργου. Η χρήση των εργαλείων αυτών δε στοχεύει μόνο στη μέτρηση των ποσοτικών παραμέτρων και τη στατιστική ανάλυση, αλλά λαμβάνει υπόψη και ποιοτικά στοιχεία που προκύπτουν από το συνδυασμό των μετρήσιμων δεικτών. Επιπλέον, εξετάζονται και άλλοι παράγοντες ή περιορισμοί που θα μπορούσαν να επηρεάσουν τα δεδομένα (π.χ. αριθμός μελών Δ.Ε.Π. κάθε τμήματος, έτος ίδρυσης του κάθε τμήματος, κτλ.), παρέχοντας ταυτόχρονα μια κριτική ματιά στην εφαρμογή της μεθόδου. Σε δεύτερο επίπεδο, εφόσον βασικό κριτήριο για την εξέλιξη σε επόμενη βαθμίδα των μελών Δ.Ε.Π. αποτελεί το διδακτικό και ερευνητικό έργο τους, τα μετρήσιμα δεδομένα μπορούν να χρησιμοποιηθούν, ώστε να διαπιστωθεί η δυνατότητα εξέλιξης του εκάστοτε μέλους.

Παράλληλα, σημαντικό ζητούμενο αποτελεί η δυνατότητα εξέτασης επιπρόσθετων χαρακτηριστικών και συσχετισμών, οι οποίοι δεν είναι ορατοί στη στατιστική ανάλυση, αλλά θα μπορούσαν να διαφανούν μέσα από την εφαρμογή της εξόρυξης γνώσης. Ουσιαστικά, διερευνάται η δυνατότητα να κατηγοριοποιηθούν και να ομαδοποιηθούν τα μέλη Δ.Ε.Π. των τμημάτων πληροφορικής, με τη χρήση τεχνικών εξόρυξης γνώσης (κατηγοριοποίηση, συσταδοποίηση), ώστε να προκύψουν ενδεχόμενες επιπρόσθετες πληροφορίες και νέα γνώση, αλλά και να προσφέρουν ένα μοντέλο πρόβλεψης για την ακαδημαϊκή βαθμίδα.

1.5. Συνεισφορά

Η παρούσα διπλωματική εργασία προσφέρει μια αξιολόγηση του ερευνητικού έργου των τμημάτων πληροφορικής που λειτουργούν στην Ελλάδα με βάση τα δεδομένα επιστημομετρίας που είναι διαθέσιμα στο Google Scholar. Η ανάλυση των επιστημομετρικών δεδομένων και τα αποτελέσματα μιας τέτοιας εφαρμογής προσφέρουν πολύ χρήσιμες πληροφορίες σχετικά με το έργο των τμημάτων, την παραγωγικότητα μέσω των δημοσιεύσεων, αλλά και την απήχηση αυτού του έργου στην επιστημονική κοινότητα. Κατ' επέκταση, οι επιδόσεις των τμημάτων, θα μπορούσαν να υποδείξουν σε ποια τμήματα παράγεται μεγαλύτερο ερευνητικό έργο, ποια μέλη Δ.Ε.Π. ξεχωρίζουν για την έντονη δραστηριότητά τους, ενώ εντοπίζονται και οι παράγοντες που φαίνεται να επηρεάζουν τα στοιχεία αυτά. Έτσι, προσφέρεται όχι μόνο μια αποτίμηση των επιδόσεων των πανεπιστημιακών τμημάτων και των μελών Δ.Ε.Π., αλλά και μια κριτική θεώρηση των παραγόντων που επηρεάζουν την παραγωγικότητα και την απήχηση, που δεν περιορίζεται μόνο στα μετρήσιμα στοιχεία, αλλά και σε άλλες παραμέτρους που προφανώς επιδρούν είτε θετικά, είτε αρνητικά. Επιπλέον, η εξόρυξη γνώσης μέσω της εκτέλεσης των επιλεγμένων

αλγορίθμων, μπορεί να προσφέρει επιπρόσθετες πληροφορίες, οδηγώντας μας σε ενδιαφέρουσες παρατηρήσεις και συμπεράσματα, αλλά και στην μελλοντική πρόβλεψη ένταξης σε ακαδημαϊκή βαθμίδα των μελών Δ.Ε.Π.

1.6. Οργάνωση της διπλωματικής εργασίας

Το 2^ο Κεφάλαιο προσφέρει το θεωρητικό υπόβαθρο σχετικά με την εφαρμογή της εξόρυξης γνώσης. Δίνεται ο ορισμός και παρουσιάζονται τα στάδια που ακολουθούνται κατά τη διαδικασία, καθώς και οι τεχνικές που χρησιμοποιούνται για την εξόρυξη γνώσης από τα δεδομένα. Στη συνέχεια, περιγράφονται οι αλγόριθμοι κατηγοριοποίησης και συσταδοποίησης που θα χρησιμοποιηθούν στην εργασία και δίνονται τα αντίστοιχα παραδείγματα.

Το Κεφάλαιο 3 αναφέρεται στα δεδομένα που χρησιμοποιήθηκαν για την υλοποίηση της παρούσας έρευνας. Παρουσιάζεται το σύνολο των επιστημονικών δεδομένων που αντλήθηκαν από τη βάση της Google Scholar και χρησιμοποιήθηκαν για τη στατιστική ανάλυση και την εξόρυξη γνώσης.

Στο 4^ο Κεφάλαιο εφαρμόζεται η στατιστική ανάλυση των επιστημονικών δεδομένων και παρουσιάζονται τα αποτελέσματά της, με τη βοήθεια πινάκων και γραφημάτων. Αρχικά, η ανάλυση αφορά τις επιδόσεις των πανεπιστημιακών τμημάτων, με βάση τους επιστημονικούς δείκτες που χρησιμοποιήθηκαν, ενώ στη συνέχεια αξιολογούνται τα μέλη Δ.Ε.Π., τόσο σε πανελλαδικό επίπεδο, όσο και σε επίπεδο περιφέρειας, και εξετάζεται η δυνατότητα εξέλιξής τους σε επόμενη βαθμίδα.

Το 5^ο Κεφάλαιο εστιάζει στην εξόρυξη γνώσης μέσω της χρήσης του λογισμικού εξόρυξης γνώσης και μηχανικής μάθησης Weka. Χρησιμοποιούνται οι τεχνικές εξόρυξης γνώσης, όπως η κατηγοριοποίηση και η συσταδοποίηση, και εκτελούνται οι επιλεγμένοι αλγόριθμοι, ώστε να προκύψουν μοντέλα κατηγοριοποίησης και συσταδοποίησης, αλλά και μοντέλα μελλοντικής πρόβλεψης.

Τέλος, στο Κεφάλαιο 6 παρουσιάζονται τα αποτελέσματα και τα συμπεράσματα της στατιστικής ανάλυσης και της εξόρυξης γνώσης των επιστημονικών δεδομένων, ενώ γίνεται λόγος και για μελλοντικές προεκτάσεις της έρευνας.

	Τμήμα πληροφορικής*	Έδρα	Έτος Ίδρυσης
ai@uom	Τμήμα Εφαρμοσμένης Πληροφορικής - Πανεπιστήμιο Μακεδονίας	Θεσσαλονίκη	2013
ceid@upatras	Τμήμα Μηχανικών Η/Υ & Πληροφορικής - Πανεπιστήμιο Πατρών	Πάτρα	1979
cs@aueb	Οικονομικό Πανεπιστήμιο Αθηνών- Τμήμα Πληροφορικής	Αθήνα	1984
cs@ihu	Τμήμα Πληροφορικής - Διεθνές Πανεπιστήμιο Ελλάδος ΔΙ.ΠΑ.Ε.*	Καβάλα	2019
cs@ionio	Τμήμα Πληροφορικής- Ιόνιο Πανεπιστήμιο	Κέρκυρα	2004
cs@unipi	Τμήμα Πληροφορικής- Πανεπιστήμιο Πειραιώς	Αθήνα	1992
cs@uoc	Τμήμα Επιστήμης Υπολογιστών- Πανεπιστήμιο Κρήτης	Ηράκλειο	1984
cs@uoi	Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής- Πολυτεχνική Σχολή- Πανεπιστήμιο Ιωαννίνων	Ιωάννινα	1990
cs@uown	Τμήμα Πληροφορικής- Πανεπιστήμιο Δυτικής Μακεδονίας*	Καστοριά	2019
cs@uth	Τμήμα Πληροφορικής και Τηλεπικοινωνιών- Πανεπιστήμιο Θεσσαλίας	Λαμία	2013
csd@auth	Τμήμα Πληροφορικής - Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, ΑΠΘ	Θεσσαλονίκη	1991
di@uoa	Τμήμα Πληροφορικής και Τηλεπικοινωνιών - ΕΚΠΑ	Αθήνα	1989
dib@uth	Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική - Πανεπιστήμιο Θεσσαλίας	Λαμία	2004
dit@hua	Τμήμα Πληροφορικής και Τηλεματικής - Χαροκόπειο Πανεπιστήμιο	Αθήνα	2006
dit@uoi	Τμήμα Πληροφορικής και Τηλεπικοινωνιών - Πανεπιστήμιο Ιωαννίνων*	Άρτα	2019
dit@uop	Τμήμα Πληροφορικής και Τηλεπικοινωνιών της Σχολής Οικονομίας και Τεχνολογίας του – Πανεπιστήμιο Πελοποννήσου	Τρίπολη	2013
ds@uop	Τμήμα Ψηφιακών Συστημάτων-Σχολής Οικονομίας και Τεχνολογίας- Πανεπιστήμιο Πελοποννήσου*	Σπάρτη	2019
ds@uth	Τμήμα Ψηφιακών Συστημάτων-Πανεπιστήμιο Θεσσαλίας*	Λάρισα	2019
ece@uth	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών - Πανεπιστήμιο Θεσσαλίας	Βόλος	2000
ece@hmu	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών – Ελληνικό Μεσογειακό Πανεπιστήμιο*	Ηράκλειο	2019
ece@ntua	Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών – Εθνικό Μετσόβιο Πολυτεχνείο	Αθήνα	1992
ece@tuc	Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών – Πολυτεχνείο Κρήτης	Χανιά	1990
ece@uop	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών- Πανεπιστήμιο Πελοποννήσου*	Πάτρα	2019
ece@uowm	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών- Πανεπιστήμιο Δυτικής Μακεδονίας-Πολυτεχνική Σχολή	Κοζάνη	2005
ece@upatras	Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών- Πανεπιστήμιο Πατρών	Πάτρα	1983
ee@auth	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης	Θεσσαλονίκη	1982
ee@duth	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Δημοκρίτειο Πανεπιστήμιο Θράκης	Ξάνθη	1975
ee@hmu	Τμήμα Ηλεκτρονικών Μηχανικών- Ελληνικό Μεσογειακό Πανεπιστήμιο*	Χανιά	2019
eee@uniwa	Τμήμα Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών -Σχολή Μηχανικών - Πανεπιστήμιο Δυτικής Αττικής*	Αθήνα	2018
ice@uniwa	Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών- Σχολή Μηχανικών- Πανεπιστήμιο Δυτικής Αττικής*	Αθήνα	2018
icsd@aegean	Τμήμα Μηχανικών Πληροφορικών και Επικοινωνιακών Συστημάτων- Πανεπιστήμιο Αιγαίου	Σάμος	1997
ict@ihu	Τμήμα Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών, Διεθνές Πανεπιστήμιο Ελλάδος, ΔΙ.ΠΑ.Ε.*	Σέρρες	2019
iee@ihu	Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων , Διεθνές Πανεπιστήμιο Ελλάδος, ΔΙ.ΠΑ.Ε.*	Θεσσαλονίκη	2019

Πίν. 1.1. Πανεπιστημιακά Ιδρύματα τμημάτων Πληροφορικής στην Ελλάδα.

(με * δηλώνονται τα τμήματα πρώην Τ.Ε.Ι. που μετεξελίχθηκαν σε πανεπιστημιακά τμήματα και ως έτος ίδρυσης αναφέρεται το έτος μετεξέλιξης)

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

2.1 Εξόρυξη γνώσης

Η ανάπτυξη των νέων τεχνολογιών και οι εφαρμογές του διαδικτύου, προσφέρουν τη δυνατότητα πρόσβασης σε μεγάλες ποσότητες δεδομένων και πλήθος πληροφοριών. Ωστόσο, μεγάλο μέρος αυτών των πληροφοριών παραμένει ανεκμετάλλευτο, καθώς τα δεδομένα βρίσκονται σε ακατέργαστη μορφή και δεν έχουν επεξεργαστεί ώστε να προσφέρουν χρήσιμες και ουσιαστικές πληροφορίες.

Η εξόρυξη γνώσης από δεδομένα (data mining) αποτελεί μια τεχνική που στοχεύει στην εξαγωγή χρήσιμων πληροφοριών και προτύπων, εστιάζοντας στη σημαντική πληροφορία που βρίσκεται κρυμμένη μέσα σε σύνολα δεδομένων. Ένας σύντομος ορισμός της Εξόρυξης Γνώσης είναι: «η εξόρυξη χρήσιμων πληροφοριών από μεγάλα σύνολα δεδομένων» [13]. Πιο αναλυτικά, η Εξόρυξη Γνώσης από δεδομένα ορίζεται ως: «η διαδικασία εξαγωγής υπονοούμενης γνώσης, άγνωστης αλλά ενδεχομένως χρήσιμης γνώσης υπό τη μορφή συσχετίσεων, προτύπων και τάσεων, μέσω διαδικασίας εξέτασης, ανάλυσης και επεξεργασίας βάσεων δεδομένων, με συνδυασμό και χρήση τεχνικών Μηχανικής Μάθησης, την Αναγνώριση Προτύπων, τη Στατιστική, τις Βάσεις Δεδομένων και την Οπτικοποίηση» [14].

Συνεπώς, η εξόρυξη γνώσης αποτελεί σημαντικό εργαλείο, καθώς συμβάλλει στην ανακάλυψη της κρυμμένης γνώσης, κρατώντας τις χρήσιμες και δομημένες πληροφορίες που μπορούν να αναδείξουν τους μη ορατούς συσχετισμούς μεταξύ των εξεταζόμενων δεδομένων. Η εξόρυξη γνώσης χρησιμοποιείται για την αναζήτηση και ανάλυση των στοιχείων γνώσης, με στόχο το φιλτράρισμα βάσει περιεχομένου και τη δημιουργία συνδέσμων και σχέσεων μεταξύ των στοιχείων, που θα οδηγήσουν στη νέα γνώση.

Ουσιαστικά, η εξόρυξη γνώσης αποτελεί ένα μόνο από τα στάδια της διαδικασίας της Ανακάλυψης Γνώσης σε Βάσεις Δεδομένων (Knowledge Discovery in Databases – KDD), το σύνολο των οποίων οδηγεί σε νέα και πιθανώς χρήσιμα στοιχεία και κατανοητά πρότυπα [15] [16]. Ο όρος Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων είναι δανεισμένος από το πεδίο της Τεχνητής Νοημοσύνης και μπορεί να οριστεί ως: «μία ντετερμινιστική διαδικασία αναγνώρισης καινοτόμων, πρωτότυπων, έγκυρων και -ενδεχομένως- χρήσιμων και κατανοητών προτύπων από τα δεδομένα» [14]. Ο ορισμός αυτός αναγνωρίζει την Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων ως μια διαδικασία, η οποία ξεκινά με την

εισαγωγή των πηγαιών δεδομένων και καταλήγει στην εξαγωγή της παραγόμενης γνώσης, μέσα από μια αλληλουχία διαδοχικών σταδίων. Τα στάδια αυτής της διαδικασίας είναι τα ακόλουθα [16]:

Επιλογή των δεδομένων: Πρόκειται για το πρώτο στάδιο της Ανακάλυψης Γνώσης που αφορά στην επιλογή του συνόλου των υπό εξέταση δεδομένων από την αποθήκη δεδομένων, τα οποία θα χρησιμοποιηθούν στη διαδικασία της εξόρυξης γνώσης. Έτσι, εντοπίζονται τα δεδομένα και ενσωματώνονται σε ένα σύνολο που θα αποτελέσει τον πυρήνα των εξεταζόμενων στοιχείων.

Καθαρισμός και Προεπεξεργασία των δεδομένων: Είναι ένα σημαντικό βήμα της διαδικασίας που συμβάλλει στην ποιότητα και την αξιοπιστία των δεδομένων. Σκοπός είναι να μετασχηματίσει τα ακατέργαστα δεδομένα σε κατάλληλη μορφή για περαιτέρω ανάλυση. Ουσιαστικά, αφορά τον καθαρισμό των δεδομένων (data cleaning), δηλαδή τον εντοπισμό και την απομάκρυνση του θορύβου, τη διαχείριση ελλιπών τιμών, ή ακραίων παρατηρήσεων (outliers).

Μετασχηματισμός των δεδομένων: Στο στάδιο αυτό περιλαμβάνονται οι τεχνικές που έχουν ως στόχο να τροποποιήσουν/ μετασχηματίσουν το σύνολο των δεδομένων σε κατάλληλη μορφή, ώστε να χρησιμοποιηθούν στην εξόρυξη γνώσης.

Εξόρυξη γνώσης: Είναι το βασικό στάδιο της Ανακάλυψης Γνώσης και περιλαμβάνει την επιλογή των κατάλληλων μεθόδων και τεχνικών που θα χρησιμοποιηθούν, ώστε το σύνολο των δεδομένων να οδηγήσει σε χρήσιμα και ενδιαφέροντα πρότυπα γνώσης.

Αξιολόγηση: Σε αυτό το στάδιο γίνεται η αξιολόγηση και ερμηνεία των αποτελεσμάτων που προέκυψαν από την εξόρυξη γνώσης. Εδώ μπορούν να χρησιμοποιηθούν διάφορες τεχνικές, ώστε να διαπιστωθεί εάν τα νέα πρότυπα που προέκυψαν είναι χρήσιμα και κατανοητά, ή εάν χρειάζεται η προσθήκη νέων μεταβλητών και η επανάληψη των προηγούμενων βημάτων. Στόχος είναι να διατηρηθούν τα νέα πρότυπα που ανταποκρίνονται στις ανάγκες της εκάστοτε έρευνας και παρουσιάζουν ενδιαφέρον για να χρησιμοποιηθούν πλέον ως νέα γνώση.

Απεικόνιση και παρουσίαση της γνώσης: Πρόκειται για το τελευταίο βήμα της διαδικασίας, όπου η εξορυγμένη γνώση απεικονίζεται και παρουσιάζεται στο χρήστη, με τη βοήθεια τεχνικών αντιπροσώπευσης και απεικόνισης.

Στο σημείο αυτό πρέπει να υπογραμμιστεί ότι η διαδικασία για την Ανακάλυψη Γνώσης είναι επαναληπτική. Αυτό σημαίνει ότι πολλά από τα στάδια που μεσολαβούν δύνανται να επαναληφθούν, καθώς ο χρήστης μπορεί να επανέλθει σε προηγούμενο στάδιο, να

τροποποιήσει τις μεθόδους και τις τεχνικές και να επαναλάβει τη διαδικασία, έως ότου οδηγηθεί στη νέα γνώση που είναι χρήσιμη και ενδιαφέρουσα για την εκάστοτε έρευνα.

2.2 Τεχνικές εξόρυξης γνώσης

Η εφαρμογή της εξόρυξης γνώσης γίνεται με υπολογιστικά προγράμματα που επιτρέπουν τη χρήση στατιστικών και μαθηματικών αποτελεσμάτων, με στόχο την εξαγωγή συμπερασμάτων. Τα προγράμματα αυτά χρησιμοποιούν τα αποτελέσματα των στατιστικών, ώστε να εντοπίσουν πιθανούς συσχετισμούς, πρότυπα ή και αποκλίσεις, οι οποίες δεν είναι ορατές με την απλή στατιστική ανάλυση, αλλά προκύπτουν από το φιλτράρισμα ενός μεγάλου όγκου δεδομένων.

Οι τεχνικές που χρησιμοποιούνται για την εξόρυξη γνώσης περιλαμβάνουν την εφαρμογή αλγορίθμων και την κατασκευή μοντέλων, τα οποία συμβάλλουν στην ανακάλυψη και την εξαγωγή προτύπων. Τα μοντέλα που χρησιμοποιούνται στοχεύουν είτε στην πρόβλεψη συμπεριφοράς των στοιχείων (Προβλεπτικό μοντέλο), είτε στον εντοπισμό προτύπων (Περιγραφικό μοντέλο). Το προβλεπτικό μοντέλο, βάσει των ήδη γνωστών κατηγοριοποιήσεων, εκπαιδεύεται ώστε να παρέχει τη δυνατότητα διατύπωσης προβλέψεων και εκτιμήσεων για το μέλλον, δηλαδή για άγνωστες ή μελλοντικές τιμές. Από την άλλη, το περιγραφικό μοντέλο καταδεικνύει νέες ομαδοποιήσεις των δεδομένων και περιγράφει τις ιδιότητες των ήδη διαθέσιμων τιμών, χωρίς ωστόσο, να επιδιώκει τη διατύπωση μελλοντικών προβλέψεων [17].

Για την εξόρυξη γνώσης χρησιμοποιούνται διάφορες τεχνικές ανάλογα με τη στόχευση της κάθε έρευνας. Στην παρούσα εργασία χρησιμοποιήθηκαν δύο από τις βασικές τεχνικές της εξόρυξης γνώσης, η κατηγοριοποίηση και η συσταδοποίηση [18] [19].

Η κατηγοριοποίηση δεδομένων (data classification) αποτελεί την τεχνική όπου, με την εφαρμογή των σχετικών αλγορίθμων, τα δεδομένα κατηγοριοποιούνται σύμφωνα με τις κοινές ιδιότητες και χαρακτηριστικά που εντοπίζονται, βασισμένα σε προκαθορισμένες και ήδη γνωστές κλάσεις. Πρόκειται για μάθηση με επίβλεψη (supervised learning), καθώς οι κατηγορίες των τιμών είναι ήδη γνωστές και προϋπάρχουν στη βάση δεδομένων [20]. Η τεχνική αυτή χρησιμοποιείται για το προβλεπτικό μοντέλο, καθώς δίνεται η δυνατότητα να κατηγοριοποιηθούν νέα στοιχεία με άγνωστες τιμές, βάσει των προκαθορισμένων κλάσεων. Ωστόσο, η κατηγοριοποίηση μπορεί να χρησιμοποιηθεί και για το περιγραφικό μοντέλο, ως επεξηγηματικό εργαλείο.

Για την εφαρμογή της τεχνικής χρησιμοποιείται ένας αλγόριθμος, που ονομάζεται κατηγοριοποιητής, ο οποίος εκπαιδεύεται με μία συνάρτηση για να κατηγοριοποιεί τα δεδομένα με όσο το δυνατόν μεγαλύτερη ακρίβεια. Έπειτα από την κατασκευή του μοντέλου υπάρχει η δυνατότητα αξιολόγησης της ποιότητας και της ακρίβειας που αυτό πετυχαίνει. Η έμφαση δίνεται περισσότερο στην ικανότητα πρόβλεψης του μοντέλου παρά στην αποδοτικότητά του (ταχύτητα κατασκευής του μοντέλου, ταξινόμησης εγγραφών, κτλ.). Έτσι, αναπτύσσονται μέθοδοι για την εκτίμηση της απόδοσης του μοντέλου (π.χ., σφάλματα, ακρίβεια, κτλ.), αλλά και για τη σύγκριση δύο ή και περισσότερων ανταγωνιστικών μοντέλων.

Αξιολογούνται, λοιπόν, ο ρυθμός αληθώς θετικών πλειάδων (TP rate), ο ρυθμός ψευδώς θετικών πλειάδων (FP rate), η ακρίβεια (precision), η ανάκληση (recall), το μέτρο (P measure) και η περιοχή ROC (ROC area). Ένα άλλο μέγεθος που δείχνει την ακρίβεια της λύσης σε ένα πρόβλημα κατηγοριοποίησης είναι η μήτρα σύγχυσης (confusion matrix). Σε αυτήν την περίπτωση οι τιμές που έχουν κατηγοριοποιηθεί σωστά εντοπίζονται στη διαγώνιο του πίνακα, που αποτελεί και τον κορμό της κατηγοριοποίησης, ενώ ιδανικά οι μηδενικές τιμές εκτός της διαγωνίου δίνουν και τις καλύτερες λύσεις στην κατηγοριοποίηση.

Μία ακόμη διαδεδομένη τεχνική για την εξόρυξη γνώσης είναι η συσταδοποίηση δεδομένων (data clustering). Η συσταδοποίηση αφορά στον εντοπισμό υποομάδων μέσα στο σύνολο, οι οποίες εμφανίζουν όμοια χαρακτηριστικά μεταξύ τους. Οι αλγόριθμοι ταξινομούν ετερογενή στοιχεία σε συστάδες, στα οποία εντοπίζονται ομοιότητες, αλλά διαφοροποιούνται σαφώς από τα χαρακτηριστικά των υπόλοιπων συστάδων. Στόχος είναι να επιλεγούν τα καταλληλότερα γνωρίσματα στα οποία πρόκειται να εφαρμοστεί η συσταδοποίηση, ώστε να επιτυγχάνεται η βέλτιστη ομοιογένεια σε κάθε συστάδα. Έτσι η προεπεξεργασία των δεδομένων πριν την εφαρμογή της διαδικασίας συσταδοποίησης κρίνεται απαραίτητη. Η τεχνική της συσταδοποίησης προσφέρει μεθόδους για την αξιολόγηση των αποτελεσμάτων του αλγόριθμου και την ορθότητα της συσταδοποίησης. Αυτή περιλαμβάνει τη σύγκριση των αποτελεσμάτων της ανάλυσης με κάποια ήδη γνωστά αποτελέσματα ή τη σύγκριση των αποτελεσμάτων δύο διαφορετικών συσταδοποιήσεων.

Η τεχνική της συσταδοποίησης ανήκει στη μάθηση χωρίς επίβλεψη (unsupervised learning), καθώς δε βασίζεται σε προκαθορισμένες κατηγορίες. Το συγκεκριμένο μοντέλο είναι περιγραφικό, καθώς παρουσιάζει και περιγράφει νέες ομάδες δεδομένων, δηλαδή πρότυπα ή σχέσεις που υπάρχουν στα δεδομένα και αρχικά φαίνονταν άσχετες μεταξύ τους.

2.3 Αλγόριθμοι Κατηγοριοποίησης

2.3.1 Μπεϋζιανός ταξινομητής (Naïve Bayes)

Μία από τις πιο διαδεδομένες μεθόδους κατηγοριοποίησης είναι ο αφελής Μπεϋζιανός ταξινομητής (Naïve Bayes). Αφορά στην εξέταση κατανομών πιθανότητας, μετρώντας τη συχνότητα και τους συνδυασμούς τιμών σε ένα σύνολο δεδομένων, ενώ κάθε γνώρισμα είναι ανεξάρτητο από τα υπόλοιπα και συνεισφέρει εξίσου στην κατηγοριοποίηση.

Ο Naïve Bayes βασίζεται στη στατιστική θεωρία κατηγοριοποίησης του Bayes. Στόχος είναι να κατηγοριοποιηθεί ένα στιγμιότυπο X σε μια από τις δεδομένες κατηγορίες C_1, C_2, \dots, C_n χρησιμοποιώντας ένα μοντέλο πιθανότητας που ορίζεται σύμφωνα με τη θεωρία του Bayes [16]:

$$P(A|B) = P(B|A) * P(A) / P(B)$$

Το θεώρημα του Bayes είναι χρήσιμο επειδή επιτρέπει να εκφραστεί η εκ των υστέρων πιθανότητα σε σχέση με την εκ των προτέρων πιθανότητα. Έτσι, το θεώρημα εκφράζει την εξής παραδοχή: «Η πιθανότητα να συμβεί το ενδεχόμενο A , δεδομένου του B , ισούται με την πιθανότητα να συμβεί το B δεδομένου του A , επί την πιθανότητα να συμβεί το A , διά την πιθανότητα να συμβεί το B ».

Ο Naïve Bayes είναι ένας πρόθυμος κατηγοριοποιητής που προσφέρει υψηλή ακρίβεια και ενδείκνυται για μεγάλες βάσεις δεδομένων. Εμφανίζει μεγάλη ταχύτητα στη διαδικασία εκπαίδευσης και κατηγοριοποίησης, καθώς απαιτεί μόνο μία ανάγνωση των δεδομένων εκπαίδευσης, ενώ το μοντέλο που προκύπτει είναι απλό και κατανοητό. Επιπλέον, έχει καλή ανοχή στο “θόρυβο” και διαχειρίζεται με επιτυχία τις ελλιπείς τιμές. Ωστόσο, ο κατηγοριοποιητής δε χειρίζεται τα αριθμητικά (συνεχή) γνωρίσματα και είναι ιδιαίτερα αποτελεσματικός κυρίως στην κατηγοριοποίηση κειμένου. Ακόμη, η ακρίβεια της πρόβλεψης επηρεάζεται αρνητικά από πιθανές εξαρτήσεις μεταξύ των γνωρισμάτων.

Παράδειγμα αλγόριθμου Naïve Bayes

Τα δεδομένα εκπαίδευσης που θα χρησιμοποιήσουμε γι' αυτό το παράδειγμα «καλού καιρού για τένις» παρουσιάζονται στον πίνακα 2.1. Ακολουθεί η δοκιμή σε ποια κατηγορία (ναι / όχι) θα κατηγοριοποιηθεί η πλειάδα <Ηλιόλουστος, Μικρή, Υψηλή, Ισχυρός>

Ημέρα	Καιρός	Θερμοκρασία	Υγρασία	Άνεμος	Τένις
1	Ηλιόλουστος	Μεγάλη	Υψηλή	Αδύναμος	Όχι
2	Ηλιόλουστος	Μεγάλη	Υψηλή	Ισχυρός	Όχι
3	Συννεφιασμένος	Μεγάλη	Υψηλή	Αδύναμος	Ναι
4	Βροχερός	Μεσαία	Υψηλή	Αδύναμος	Ναι
5	Βροχερός	Μικρή	Κανονική	Αδύναμος	Ναι
6	Βροχερός	Μικρή	Κανονική	Ισχυρός	Όχι
7	Συννεφιασμένος	Μικρή	Κανονική	Ισχυρός	Ναι
8	Ηλιόλουστος	Μεσαία	Υψηλή	Αδύναμος	Όχι
9	Ηλιόλουστος	Μικρή	Κανονική	Αδύναμος	Ναι
10	Βροχερός	Μεσαία	Κανονική	Αδύναμος	Ναι
11	Ηλιόλουστος	Μεσαία	Κανονική	Ισχυρός	Ναι
12	Συννεφιασμένος	Μεσαία	Υψηλή	Ισχυρός	Ναι
13	Συννεφιασμένος	Μεγάλη	Κανονική	Αδύναμος	Ναι
14	Βροχερός	Μεσαία	Υψηλή	Ισχυρός	Όχι

Πίνακας 2.1. Δεδομένα εκπαίδευσης αλγόριθμου Naïve Bayes

Καιρός:	$P(\text{Ηλιόλουστος} \mid \text{όχι}) = 3/5,$ $P(\text{Συννεφιασμένος} \mid \text{όχι}) = 0/5,$ $P(\text{Βροχερός} \mid \text{όχι}) = 2/5,$	$P(\text{Ηλιόλουστος} \mid \text{ναι}) = 2/9,$ $P(\text{Συννεφιασμένος} \mid \text{ναι}) = 4/9,$ $P(\text{Βροχερός} \mid \text{ναι}) = 3/9$
Θερμοκρασία:	$P(\text{Μεγάλη} \mid \text{όχι}) = 2/5,$ $P(\text{Μεσαία} \mid \text{όχι}) = 2/5,$ $P(\text{Μικρή} \mid \text{όχι}) = 1/5,$	$P(\text{Μεγάλη} \mid \text{ναι}) = 2/9,$ $P(\text{Μεσαία} \mid \text{ναι}) = 4/9,$ $P(\text{Μικρή} \mid \text{ναι}) = 3/9$
Υγρασία:	$P(\text{Υψηλή} \mid \text{όχι}) = 4/5,$ $P(\text{Κανονική} \mid \text{όχι}) = 1/5,$	$P(\text{Υψηλή} \mid \text{ναι}) = 3/9,$ $P(\text{Κανονική} \mid \text{ναι}) = 6/9$
Άνεμος:	$P(\text{Αδύναμος} \mid \text{όχι}) = 3/5$ $P(\text{Ισχυρός} \mid \text{όχι}) = 2/5$	$P(\text{Αδύναμος} \mid \text{ναι}) = 7/9$ $P(\text{Αδύναμος} \mid \text{ναι}) = 2/9$

Κατηγορία ΝΑΙ:

$$P(\text{ναι}) * P(\text{Ηλιόλουστος} \mid \text{ναι}) * P(\text{Μικρή} \mid \text{ναι}) * P(\text{Υψηλή} \mid \text{ναι}) * P(\text{Ισχυρός} \mid \text{ναι}) = 9/14 * 2/9 * 3/9 * 3/9 * 2/9 = 324/91854$$

Διαπιστώνουμε ότι η απόδοση του αλγόριθμου είναι μεγαλύτερη αν κατηγοριοποιήσουμε τη νέα πλειάδα στην κατηγορία όχι.

2.3.2 Δέντρα Απόφασης (Decision Trees).

Πρόκειται ίσως για το πιο δημοφιλές μοντέλο κατηγοριοποίησης για τη μάθηση με επίβλεψη. Τα δέντρα απόφασης περιλαμβάνουν σχηματικά έναν αρχικό κόμβο (ρίζες), εσωτερικούς κόμβους και εξωτερικούς κόμβους (φύλλα), στα οποία γίνεται ο διαχωρισμός/

κατηγοριοποίηση των δεδομένων. Η ταξινόμηση ξεκινάει από τη ρίζα του δέντρου και στη συνέχεια διαμοιράζεται σε κάποιο κόμβο, καταλήγοντας με μεγαλύτερη ακρίβεια στα φύλλα. Έτσι, η πολυπλοκότητα του δέντρου υπολογίζεται από το συνολικό αριθμό των κόμβων, το συνολικό αριθμό των φύλλων, το βάθος του δέντρου και τον αριθμό των χαρακτηριστικών που χρησιμοποιούνται [21] [22].

Όσο περισσότερους κόμβους και φύλλα αποκτά το δέντρο, τόσο περισσότερη ακρίβεια εμφανίζει στην πρόβλεψη του μοντέλου, καθώς σε μία τέτοια περίπτωση έχει χρησιμοποιηθεί πλήθος χαρακτηριστικών που κατηγοριοποιούνται. Από την άλλη, τις περισσότερες φορές προτιμάται ένα δέντρο απόφασης που δεν είναι ιδιαίτερα περίπλοκο, καθώς είναι πιο εύκολο να κατανοηθεί. Ένα δέντρο απόφασης μπορεί να σταματήσει την ανάπτυξή του όσο ακόμα είναι κατανοητό, εφόσον δίνεται η δυνατότητα ελέγχου για τη διακοπή της διαδικασίας με διάφορες τεχνικές κλαδέματος. Γενικότερα, τα δέντρα απόφασης με λιγότερα επίπεδα (μικρότερο βάθος) είναι προτιμότερα, καθώς είναι πιο ισορροπημένα και κατανοητά.

Η εκτέλεση του αλγόριθμου για τη δημιουργία ενός δέντρου απόφασης είναι η ακόλουθη:

Είσοδος: D // Δεδομένα εκπαίδευσης

Έξοδος: T // Δένδρο απόφασης

Αλγόριθμος **DTBuild**

$T \leftarrow \emptyset$;

Καθόρισε το καλύτερο κριτήριο διάσπασης;

$T \leftarrow$ Δημιούργησε τον κόμβο ρίζα και ονόμασέ τον με το όνομά του χαρακτηριστικού διάσπασης;

$T \leftarrow$ Πρόσθεσε τόσα τόξα από τον κόμβο ρίζα όσα και τα κατηγορήματα διάσπασης (τιμές χαρακτηριστικού) και ονόμασε τα

Για κάθε τόξο **επανάλαβε**

$D \leftarrow$ Δεδομένα εκπαίδευσης που παραμένουν εφαρμόζοντας το κατηγορήμα διάσπασης στο ρ

Αν ικανοποιείται το κριτήριο τερματισμού για αυτό το μονοπάτι **τότε**

$T' \leftarrow$ Δημιούργησε ένα κόμβο φύλλο και ονόμασε τον με το όνομα της κλάσης

Αλλιώς

$T' \leftarrow$ DTBuild (D);

Τέλος **_αν**

$T =$ πρόσθεσε τόξο στο T'

Τέλος **_επανάληψης**

Ο αλγόριθμος ID3

Ο αλγόριθμος ID3 εκτελείται για την κατασκευή ενός δέντρου απόφασης [23]. Χρησιμοποιεί σαν κριτήριο για τον προσδιορισμό του «καλύτερου χαρακτηριστικού διάσπασης» το «κέρδος πληροφορίας» (information gain), δηλαδή προσπαθεί να προσδιορίσει την παράμετρο που του δίνει τη μεγαλύτερη πληροφορία κέρδους:

$$\text{InformationGain}(S, A) = \text{Info}(S) - \text{Info}(S, A)$$

Το «κέρδος πληροφορίας» μετριέται ποσοτικά με την εντροπία (entropy). Δεδομένων των πιθανοτήτων p_1, p_2, \dots, p_k , όπου $\sum p_i = 1$, η εντροπία E ορίζεται ως εξής:

$$E(S) = E(p_1, p_2, \dots, p_k) = -\sum (p_i \log(1/p_i))$$

Η εντροπία αφορά ουσιαστικά το μέγεθος της ανομοιογένειας ενός συνόλου δεδομένων. Επομένως, όσο μεγαλύτερη ομοιογένεια παρουσιάζουν τα δεδομένα της κατηγοριοποίησης, τόσο μικρότερη θα είναι η εντροπία, ενώ αυτή θα είναι μηδενική όταν όλα τα δεδομένα αντιστοιχούν σε μία κλάση. Εφόσον, ο στόχος των δέντρων απόφασης είναι ο όσο το δυνατόν πιο ακριβής διαχωρισμός των δεδομένων, τότε γίνεται κατανοητό ότι επιδιώκονται μηδενικές τιμές εντροπίας σε κάθε κλάση.

Η δομή του δέντρου απόφασης εξαρτάται από τα δεδομένα εκπαίδευσης. Έτσι, ένα μικρό σύνολο εκπαίδευσης οδηγεί σε ένα μικρό και όχι ιδιαίτερα λεπτομερές δέντρο απόφασης για την ταξινόμηση των γενικότερων δεδομένων. Αντίθετα, ένα μεγάλο δέντρο απόφασης μπορεί να υπερπροσαρμόζεται (overfits) και ο υπολογισμός του να γίνει πολύπλοκος. Κατά τη διαδικασία δίνεται η δυνατότητα τροποποιήσεων, ώστε να βελτιωθεί η απόδοσή του με διάφορες τεχνικές, όπως π.χ. το κλάδεμα (pruning).

Η κατασκευή ενός δέντρου απόφασης θεωρείται ολοκληρωμένη όταν κατηγοριοποιηθούν όλα τα δεδομένα σε κλάσεις. Ωστόσο, υπάρχουν περιπτώσεις που αυτό μπορεί να γίνει σε πρωιμότερο στάδιο, ώστε να αποφευχθεί η κατασκευή ενός μεγάλου δέντρου, ή να αποφύγουμε το φαινόμενο της υπερπροσαρμογής (overfitting). Το στάδιο ολοκλήρωσης εξαρτάται από τη συναλλαγή (trade-off) μεταξύ ακρίβειας (accuracy) και απόδοσης (performance) του αλγόριθμου. Από την άλλη πλευρά, η κατασκευή ενός μεγάλου δέντρου μπορεί να εξυπηρετήσει τις περιπτώσεις όπου εντοπίζονται κατηγορίες δεδομένων που δεν αντιπροσωπεύονται στο σύνολο της εκπαίδευσης.

Παράδειγμα αλγόριθμου ID3

Στο παράδειγμα του πίνακα 2.2 εκτελείται ο αλγόριθμος ID3. Διαπιστώνεται ότι το κέρδος πληροφορίας είναι μεγαλύτερο όταν η διάσπαση γίνει στην κατηγορία height, ενώ στην κατηγορία gender η απόδοση του αλγόριθμου, και επομένως το κέρδος πληροφορίας, είναι μικρότερα.

Name	Gender	Height	Output1
Kristina	F	1.6m	Short
Jim	M	2m	Tall
Maggie	F	1.9m	Medium
Martha	F	1.88m	Medium
Stephanie	F	1.7m	Short
Bob	M	1.85m	Medium
Kathy	F	1.6m	Short
Dave	M	1.7m	Short
Worth	M	2.2m	Tall
Steven	M	2.1m	Tall
Debbie	F	1.8m	Medium
Todd	M	1.95m	Medium
Kim	F	1.9m	Medium
Amy	F	1.8m	Medium
Wynette	F	1.75m	Medium

Πίνακας 2.2. Δεδομένα εκπαίδευσης αλγόριθμου ID3

Αρχική κατάσταση εντροπίας: $H(D) = 4/15 \log(15/4) + 8/15 \log(15/8) + 3/15 \log(15/3) = 0.4384$

Κέρδος αν γίνει διάσπαση στο **gender**:

Gender='F': $3/9 \log(9/3) + 6/9 \log(9/6) = 0.2764$

Gender='M': $1/6 \log(6/1) + 2/6 \log(6/2) + 3/6 \log(6/3) = 0.4392$

Weighted sum: $(9/15)(0.2764) + (6/15)(0.4392) = 0.3415$

Gain: $0.4384 - 0.3415 = \mathbf{0.0969}$

Κέρδος αν γίνει διάσπαση στο **height**:

Weighted sum: $0+0+\dots + (2/15)(0.301) = 0.0401$

Gain: $0.4384 - 0.0401 = \mathbf{0.3983}$

Κατηγορήματα διάσπ.: $(0, 1.6], (1.6, 1.7], (1.7, 1.8], (1.8, 1.9], (1.9, 2.0], (2.0, \infty)$

Επιλέγουμε **height**

2.3.3 Κατηγοριοποιητές πλησιέστερου γείτονα (*k*-Nearest Neighbor, *KNN*)

Η κατηγοριοποίηση πλησιέστερου γείτονα (*k*-Nearest Neighbor), αφορά σε αλγόριθμο επιβλεπόμενης μάθησης. Βασίζεται στην απόσταση των *k*-κοντινών δεδομένων εκπαίδευσης για την ένταξη και κατηγοριοποίηση ενός νέου παραδείγματος. Ουσιαστικά, εντοπίζει όμοια προγενέστερα παραδείγματα με βάση τον υπολογισμό των αποστάσεων και την μεταξύ τους παρεμβολή. Επομένως, ως μοντέλο πρόβλεψης, κατηγοριοποιεί τα νέα δεδομένα σε κλάσεις με βάση την πλειοψηφία των γειτόνων (majority voting) [24] [25].

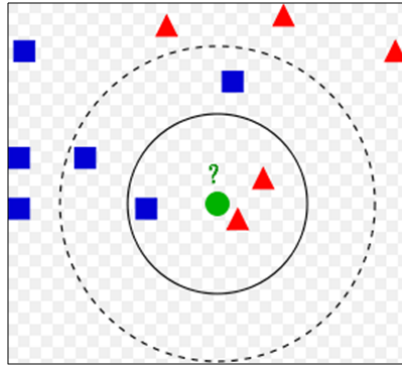
Ιδιαίτερα σημαντικός είναι ο προσδιορισμός της παραμέτρου *k*, δηλαδή της απόστασης από τον *k* πλησιέστερο γείτονα (μέτρο απόστασης), κάτι που εξαρτάται από τα δεδομένα. Ο αριθμός των γειτόνων *k* είναι ένας θετικός ακέραιος αριθμός, όπου αν *k*=1 τότε το αντικείμενο απλά αποδίδεται στην κατηγορία που ανήκει ο πλησιέστερος γείτονας. Ένα νέο δεδομένο, λαμβάνει τη θέση του με βάση την απόσταση που έχει από τα προγενέστερα δεδομένα, αναγνωρίζει τους *k* πλησιέστερους γείτονες και κατηγοριοποιείται σύμφωνα με αυτούς. Έχει διαπιστωθεί ότι όσο μεγαλύτερη είναι η τιμή του *k*, τόσο περισσότερο μειώνεται ο “θόρυβος” στην ταξινόμηση, αλλά τα όρια μεταξύ των κλάσεων γίνονται λιγότερο διακριτά. Υπάρχουν διάφοροι αλγόριθμοι βελτιστοποίησης, ώστε να επιλεγεί η καλύτερη δυνατή τιμή του *k*, ενώ στην περίπτωση που ο αριθμός *k* επιλέγεται να ισούται με τη μονάδα, τότε ο αλγόριθμος ονομάζεται πλησιέστερου γείτονα.

Ο αλγόριθμος πλησιέστερου γείτονα είναι από τους πιο απλούς αλγόριθμους υπολογιστικής μάθησης και το μόνο που απαιτείται είναι η επιλογή του αριθμού των γειτόνων *k*. Ωστόσο, δε χρησιμοποιεί τα δεδομένα εκπαίδευσης για να κάνει οποιαδήποτε γενίκευση, αλλά κρατά όλα τα δεδομένα για εκπαίδευση σε όλες τις φάσεις δοκιμών. Επομένως, πρόκειται για έναν αλγόριθμο δαπανηρό από άποψη υπολογιστικής ισχύος, καθώς αποθηκεύει και χρησιμοποιεί το σύνολο των δεδομένων για τη λήψη απόφασης.

Παράδειγμα αλγόριθμου KNN

Στο παράδειγμα που ακολουθεί γίνεται αντιληπτό πως ο πράσινος κύκλος πρέπει να ταξινομηθεί είτε στην κατηγορία μπλε τετράγωνο, είτε στην κατηγορία κόκκινο τρίγωνο. Αν επιλεγεί ο αριθμός γειτόνων να είναι τρεις, *k*=3 (με τη συνεχόμενη γραμμή), τότε θα επιλεγεί η

κατηγορία κόκκινο τρίγωνο, αφού μέσα στον κύκλο υπάρχουν 2 κόκκινα τρίγωνα και 1 μπλε τετράγωνο. Αντίθετα, αν επιλεγεί ο αριθμός γειτόνων να είναι πέντε, $k=5$ (με τη διακεκομμένη γραμμή), τότε θα επιλεγεί η κατηγορία μπλε τετράγωνο αφού στο όριο υπάρχουν 3 μπλε τετράγωνα και 2 κόκκινα τρίγωνα.



Εικόνα 2.1. Παράδειγμα εφαρμογής αλγόριθμου KNN

Είσοδος: T // Σύνολο δεδομένων εκπαίδευσης
 K // Αριθμός κοντινότερων γειτόνων
 T // πλειάδα προς κατηγοριοποίηση
Έξοδος: γ // Κλάση όπου θα κατηγοριοποιηθεί η t

Αλγόριθμος $_K_$ **Κοντινότερων** $_$ **Γειτόνων**

$N = \emptyset$

Για κάθε $pe \in T$ επανέλαβε

Αν $|N| \leq K$ τότε

$N = N \cup \{d\};$

Αλλιώς

Αν $\exists u \in N$ τέτοιο ώστε $\text{dist}(t, u) \leq \text{απόσταση}(t, d)$, τότε

$N = N - \{u\};$

$N = N \cup \{d\};$

Τέλος $_$ αν

Τέλος $_$ επανάληψης

ντο = κλάση όπου τα περισσότερα $u \in N$ κατηγοριοποιούνται

Τέλος αλγορίθμου

Εκτέλεση του αλγόριθμου Knn

2.4 Αλγόριθμοι Συσταδοποίησης

2.4.1 Αλγόριθμος συσταδοποίησης K-means

Από τους πιο διαδεδομένους αλγόριθμους συσταδοποίησης είναι ο αλγόριθμος K-means. Ο αλγόριθμος βασίζεται στην αρχή της διαχωριστικής συσταδοποίησης, δηλαδή στο διαχωρισμό των δεδομένων σε συστάδες (k clusters), όπου το καθένα ανήκει σε ένα υποσύνολο και δεν επικαλύπτεται από άλλη συστάδα. Ο αλγόριθμος θεωρεί ότι ο αριθμός των συστάδων που θα προκύψουν είναι γνωστός εκ των προτέρων. Γι' αυτό είναι ιδιαίτερα σημαντική η επιλογή του πλήθους των συστάδων. Τα σημεία (δεδομένα) που ορίζουν την κάθε συστάδα θα πρέπει να εμφανίζουν μεγάλη ομοιότητα μεταξύ τους και επομένως μικρή απόσταση. Από την άλλη, οι συστάδες θα πρέπει να εμφανίζουν μικρή ομοιότητα και μεγάλη απόσταση μεταξύ τους [16] [26] [27].

Ο αλγόριθμος K-Means περιλαμβάνει έναν ήδη γνωστό αριθμό συστάδων στον οποίο διαχωρίζονται τα δεδομένα. Ωστόσο, δίνεται η δυνατότητα να επιστρέψουμε στην εκτέλεσή του ορίζοντας και νέες συστάδες. Πρόκειται ουσιαστικά για μια επαναληπτική διαδικασία, καθώς η αρχική επιλογή των k clusters είναι τυχαία, ενώ στη συνέχεια γίνονται πιο στοχευμένες επιλογές έως ότου φτάσουμε στην βέλτιστη επιλογή συστάδων.

Ο αλγόριθμος χρησιμοποιεί την έννοια του κέντρου της συστάδας, δηλαδή τη μέση τιμή για κάθε μεταβλητή της ομάδας, και έπειτα κατατάσσει τα δεδομένα ανάλογα με την απόσταση που έχουν από τα κέντρα όλων των ομάδων. Στη συνέχεια κατατάσσεται κάθε παρατήρηση που είναι πιο κοντά υπολογίζοντας την ευκλείδεια απόστασή της (Euclidean distance) από τα κέντρα των ομάδων. Ωστόσο, ο ορισμός του μέσου της ομάδας σε μη συνεχή δεδομένα είναι προβληματικός. Έτσι, σε κατηγορικά δεδομένα με κατάταξη μπορούμε να χρησιμοποιήσουμε το διάνυσμα των διαμέσων, ενώ σε ονομαστικά δεδομένα μπορούμε να χρησιμοποιήσουμε την κορυφή (επικρατούσα τιμή). Στα δεδομένα μικτού τύπου το κέντρο μπορεί να αποτελείται από τις κορυφές των κατηγορικών μεταβλητών και τους μέσους των συνεχών.

Για την εκτίμηση της ποιότητας του αλγόριθμου η πιο συνηθισμένη μέτρηση είναι το άθροισμα των τετραγωνικών σφαλμάτων (Sum of Squared Error - SSE). Για κάθε σημείο, το σφάλμα είναι η απόστασή του από την κοντινότερη συστάδα. Για να υπολογίσουμε το SSE, παίρνουμε το τετράγωνο αυτών των σφαλμάτων και τα προσθέτουμε όπου dist η Ευκλείδεια απόσταση, x είναι ένα σημείο στη συστάδα C_i και m_i είναι το κεντρικό σημείο της συστάδας C_i :

$$SSE = \sum_{i=0}^k \sum_{i=0}^k dist^2(mi, x)$$

Στόχος είναι να εντοπιστεί η τιμή στην οποία ελαχιστοποιείται το άθροισμα των τετραγωνικών αποστάσεων των παρατηρήσεων από τα κέντρα των ομάδων που ανήκουν. Η ολοκλήρωση της διαδικασίας αντιπροσωπεύεται με ομάδες που έχουν τον ίδιο αριθμό παρατηρήσεων.

Ο αλγόριθμος K-means είναι ιδιαίτερα γρήγορος, καθώς με λίγες επαναλήψεις πλησιάζει κοντά στην τελική λύση. Επιπλέον, δε χρειάζεται μεγάλη υπολογιστική ισχύ και ενδείκνυται για μεγάλες βάσεις δεδομένων.

Παράδειγμα αλγόριθμου K-means

Στον πίνακα 2.3 έχουμε 12 σημεία, όπου εκτελούμε τον αλγόριθμο k-means με k=2 για τη συσταδοποίηση τους. Έστω ότι Σ1 και Σ2 υποδηλώνουν τις συστάδες με τυχαία αρχικά κέντρα τα σημεία 1(7,4) και 3(5,6) αντίστοιχα. Υπολογίζονται οι ευκλείδειες αποστάσεις κάθε σημείου από τα κέντρα των συστάδων και τα σημεία τοποθετούνται στη συστάδα με την όποια έχουν μικρότερη απόσταση από το κέντρο της.

Σημείο	X	Y	Απόσταση από Σ1	Απόσταση από Σ2	Συστάδα	Σημεία
1	7	4	0	2,8284	Σ1	1,2,4,5,6,9,1
2	6	4	1	2,2361	Σ1	
3	5	6	2,8284	0	Σ2	3,7,8,10,11
4	4	2	3,6056	4,1231	Σ1	
5	6	3	1,4142	3,1623	Σ1	
6	5	2	2,8284	4	Σ1	
7	3	3	4,1231	3,6056	Σ2	
8	4	5	3,1623	1,4142	Σ2	
9	6	5	1,4142	1,4142	Σ1	
10	3	6	4,4721	2	Σ2	
11	4	4	3	2,2361	Σ2	
12	8	2	2,2361	5	Σ1	

Πίνακας 2.3. Εκτέλεση αλγόριθμου k-means σε 12 σημεία

Έτσι προκύπτουν οι συστάδες $\Sigma_1=\{1,2,4,5,6,9,12\}$ και $\Sigma_2=\{3,7,8,10,11\}$. Στη συνέχεια, βρίσκουμε τα νέα κέντρα των συστάδων τα οποία προκύπτουν από το μέσο όρο των συντεταγμένων των σημείων της κάθε συστάδας, όπου:

X1	Y1
6	3.1429
X2	Y2
3.8	4.8

Πίνακας 2.4. Υπολογισμός νέων κέντρων συστάδων

Έπειτα από τον υπολογισμό των νέων κέντρων (πίν. 2.4), υπολογίζουμε και πάλι τις ευκλείδειες αποστάσεις κάθε σημείου από τα νέα κέντρα των συστάδων και τα σημεία τοποθετούνται στη συστάδα με την όποια έχουν μικρότερη απόσταση από το κέντρο της. Έτσι, σύμφωνα με τον πίνακα 2.5 προκύπτουν οι συστάδες $\Sigma_1=\{1,2,4,5,6,9,12\}$ και $\Sigma_2=\{3,7,8,10,11\}$.

Σημείο	X	Y	Απόσταση	Απόσταση	Συστάδα	Σημεία
1	7	4	1,3171	3,2985	Σ_1	1,2,4,5,6,9,12
2	6	4	0,8571	2,3409	Σ_1	
3	5	6	3,0271	1,6971	Σ_2	3,7,8,10,11
4	4	2	2,3035	2,8071	Σ_1	
5	6	3	0,1429	2,8425	Σ_1	
6	5	2	1,5186	3,0463	Σ_1	
7	3	3	3,0034	1,9698	Σ_2	
8	4	5	2,7293	0,2828	Σ_2	
9	6	5	1,8571	2,2091	Σ_1	
10	3	6	4,1429	1,4422	Σ_2	
11	4	4	2,1759	0,8246	Σ_2	
12	8	2	2,3035	5,0478	Σ_1	

Πίνακας 2.5. Εκτέλεση αλγόριθμου k-means με νέα κέντρα συστάδων

Σύμφωνα με τον πίνακα 2.5, διαπιστώνουμε ότι δεν υπάρχει αλλαγή στις συστάδες σε σχέση με το προηγούμενο βήμα, οπότε ο αλγόριθμος συγκλίνει και σταματάει εδώ. Το αποτέλεσμα του αλγόριθμου k-Means είναι η δημιουργία δύο συστάδων, η συστάδα $\Sigma_1=\{1,2,3,5,6,9,12\}$ και η συστάδα $\Sigma_2=\{3,7,8,10,11\}$.

Κεφάλαιο 3

Το σύνολο Δεδομένων

Στο Κεφάλαιο αυτό παρουσιάζεται το σύνολο των δεδομένων που επιλέχθηκε και εξετάζεται για τους σκοπούς της παρούσας έρευνας, ενώ γίνεται λόγος για τη μεθοδολογία που ακολουθήθηκε. Έτσι, παρουσιάζονται αναλυτικά τα δεδομένα που ανακτήθηκαν από τη μηχανή αναζήτησης Google Scholar και γίνεται η περιγραφή του Συνόλου Δεδομένων που χρησιμοποιήθηκαν τόσο για τη στατιστική ανάλυση, όσο και για την εξόρυξη γνώσης μέσω του λογισμικού Weka.

Η έρευνα βασίστηκε στην πρωτογενή βιβλιογραφία, κυρίως ακαδημαϊκά άρθρα, αναπροσαρμοσμένη στις μεγάλες βιβλιογραφικές βάσεις δεδομένων. Για την έρευνα έγινε χρήση της μηχανής αναζήτησης Google Scholar ως πηγής άντλησης των επιστημονικών δεδομένων. Οι λόγοι επιλογής της συγκεκριμένης βάσης δεδομένων είναι η ελεύθερη πρόσβαση σε αυτήν, το εύρος των επιστημονικών πεδίων και η κάλυψη επιστημονικών εκδόσεων, αλλά και το εύρος των τύπων δημοσιεύσεων (κάλυψη μεγάλου αριθμού διεθνών συνεδρίων), όπως και η παροχή εργαλείων που προσφέρουν επιστημονικά δεδομένα.

Για κάθε Τμήμα που σχετίζεται με την επιστήμη της πληροφορικής καταγράφηκαν από τις επίσημες ιστοσελίδες τους τα μέλη Δ.Ε.Π. που διαθέτουν λογαριασμό στη Google Scholar. Επαναλαμβάνεται ότι οι βαθμίδες που αξιολογήθηκαν περιλαμβάνουν αυτή του Καθηγητή, του Αναπληρωτή καθηγητή, του Επίκουρου Καθηγητή, του Λέκτορα και του Ε.Δ.Ι.Π. Τα στοιχεία που συλλέχθηκαν στην πρώτη αυτή φάση για κάθε μέλος Δ.Ε.Π. είναι:

- Ονοματεπώνυμο
- Google Scholar id, που αντιστοιχεί στο μοναδικό αριθμό που διαθέτει κάθε ερευνητής
- Τμήμα που ανήκει
- Η ακαδημαϊκή βαθμίδα που κατέχει το μέλος Δ.Ε.Π.

Από τα 33 τμήματα πληροφορικής που λειτουργούν στην Ελλάδα, καταγράφηκαν συνολικά 901 ενεργά μέλη Δ.Ε.Π., με βάση τα στοιχεία που συλλέχθηκαν έως το Δεκέμβριο 2021.

Στη συνέχεια, τα στοιχεία αυτά εισήχθησαν σε μια εφαρμογή που αναπτύχθηκε από τον επιβλέποντα καθηγητή της παρούσας εργασίας, απ' όπου και αντλήθηκαν τα επιστημονικά δεδομένα των μελών Δ.Ε.Π. από το Google Scholar. Η εφαρμογή αυτή

αναπτύχθηκε σε python. Για κάθε μέλος Δ.Ε.Π., συνδέεται στο Google Scholar και χρησιμοποιώντας τεχνικές web scraping ανακτά το δεδομένα επιστημομετρίας που αφορούν το συγκεκριμένο μέλος Δ.Ε.Π. Η ανάλυση των δεδομένων που προέκυψε για τα τμήματα πληροφορικής και τα μέλη Δ.Ε.Π. που εργάζονται σε αυτά βασίστηκε στην συλλογή στοιχείων που το πρόγραμμα αντλούσε σε εβδομαδιαία βάση.

Μέσω των παραπάνω ενεργειών συλλέχθηκαν και συμπληρώθηκαν για κάθε μέλος Δ.Ε.Π. τα παρακάτω στοιχεία:

- Ο αριθμός δημοσιεύσεων
- Ο αριθμός αναφορών
- Ο δείκτης h (h-index)
- Ο αριθμός δημοσιεύσεων ανά έτος
- Ο αριθμός αναφορών ανά έτος
- Ο αριθμός δημοσιεύσεων των τελευταίων 5 ετών
- Ο αριθμός αναφορών των τελευταίων 5 ετών
- Ο μέσος όρος του h-index των τελευταίων 5 ετών
- Ο μέσος όρος ετών δημοσιεύσεων ανά μέλος Δ.Ε.Π. ανά έτος δημοσίευσης
- Τυπική απόκλιση δημοσιεύσεων ανά τμήμα
 - Τυπική απόκλιση αναφορών ανά τμήμα

Τα στοιχεία αυτά καταγράφηκαν σε δύο αρχεία EXCEL, όπου το πρώτο χρησιμοποιήθηκε για τις ανάγκες της στατιστικής ανάλυσης, ενώ το δεύτερο για την εξόρυξη γνώσης μέσω του λογισμικού Weka. Παρακάτω αναφέρονται αναλυτικά τα δεδομένα που χρησιμοποιήθηκαν για το πρώτο (στατιστική ανάλυση) και δεύτερο μέρος (εξόρυξη γνώσης) της εργασίας.

Οι δείκτες που χρησιμοποιήθηκαν για τη στατιστική ανάλυση περιλαμβάνουν τον αριθμό δημοσιεύσεων (παραγωγικότητα) και τον αριθμό αναφορών (δείκτης απήχησης), τόσο σε επίπεδο τμήματος, όσο και σε επίπεδο ερευνητή, ενώ ο δείκτης h (h-index) χρησιμοποιήθηκε για την αξιολόγηση σε επίπεδο τμήματος. Η χρήση των παραπάνω δεικτών επιτρέπει την αξιολόγηση πολλαπλών παραμέτρων και δεν απαιτείται χρήση αυξημένων πόρων και εξειδικευμένων γνώσεων. Επιπλέον, η συνδυαστική χρήση των παραπάνω δεικτών οδηγεί σε μια, όσο το δυνατόν, ολοκληρωμένη ανάλυση.

Με βάση τις παραπάνω πληροφορίες, για την αξιολόγηση σε επίπεδο τμημάτων, υπολογίστηκε το σύνολο των δημοσιεύσεων και αναφορών του κάθε τμήματος, ο μέσος όρος (μ.ο.) δημοσιεύσεων και αναφορών που αντιστοιχεί σε κάθε μέλος Δ.Ε.Π. ανά τμήμα

πληροφορικής, ο μέγιστος και ελάχιστος δείκτης h ανά τμήμα πληροφορικής και η τυπική απόκλιση δημοσιεύσεων και αναφορών.

Στη συνέχεια, εξετάστηκε η δυνατότητα του κάθε μέλους Δ.Ε.Π. να μεταπηδήσει σε ανώτερη βαθμίδα (εξέλιξη) σε πανελλαδικό επίπεδο, σύμφωνα με τα ποσοτικά δεδομένα που προέκυψαν από το μ.ο. των δημοσιεύσεων και αναφορών ανά βαθμίδα. Τέλος, εξετάστηκε η δυνατότητα των μελών Δ.Ε.Π. να εξελιχθούν στην επόμενη βαθμίδα και σε επίπεδο περιφερειακής ενότητας, σύμφωνα με το μ.ο. δημοσιεύσεων και αναφορών ανά βαθμίδα και ανά περιφέρεια.

Στο δεύτερο μέρος της εργασίας, το σύνολο των δεδομένων που χρησιμοποιήθηκε για την εφαρμογή της εξόρυξης γνώσης μέσω του λογισμικού Weka περιλαμβάνει τον αριθμό δημοσιεύσεων και αναφορών κάθε μέλους Δ.Ε.Π., το μέσο όρο h -index, τις δημοσιεύσεις και αναφορές των μελών Δ.Ε.Π. τα τελευταία 5 έτη, το μέσο όρο του h -index τα τελευταία 5 έτη και το μέσο όρο δημοσιεύσεων ανά μέλος Δ.Ε.Π. ανά έτος δημοσίευσης.

Τα στοιχεία αυτά χρησιμοποιήθηκαν για την εκτέλεση των αλγόριθμων κατηγοριοποίησης και συσταδοποίησης. Για την εφαρμογή της τεχνικής της κατηγοριοποίησης και της συσταδοποίησης εξετάστηκε η ακαδημαϊκή βαθμίδα του κάθε μέλους Δ.Ε.Π..

Στο σημείο αυτό είναι χρήσιμο να αναφερθούμε σε κάποιους περιορισμούς που προέκυψαν κατά τη διαδικασία συλλογής των δεδομένων. Αρχικά, πρέπει να ληφθεί υπόψη το γεγονός ότι κάποια μέλη Δ.Ε.Π. τμημάτων πληροφορικής δε διαθέτουν λογαριασμό στη μηχανή αναζήτησης Google Scholar, οπότε δεν συμπεριλήφθηκαν στην ανάλυση. Επιπρόσθετα, κατά την αναζήτηση των μελών Δ.Ε.Π. στη Google Scholar, εντοπίστηκαν κάποιες περιπτώσεις εμφάνισης πολλαπλών ονομάτων λόγω παραλλαγών ονόματος ή ιδρύματος. Να σημειωθεί ότι σε αυτές τις περιπτώσεις τα στοιχεία που συλλέχθηκαν ελέγχθηκαν ως προς τη συνωνυμία, ενώ για την ανάλυση επιλέχθηκαν μόνο αυτά που αφορούν το αντικείμενο της παρούσας έρευνας.

Κεφάλαιο 4

Στατιστική Ανάλυση των δεδομένων

Στο Κεφάλαιο αυτό παρουσιάζονται και αναλύονται τα αποτελέσματα της στατιστικής ανάλυσης των επιστημονομετρικών δεδομένων που συλλέχθηκαν από το Google Scholar. Τα 33 τμήματα κλάδου Πληροφορικής που εντάχθηκαν στην έρευνα παρουσιάζονται στον πίνακα 1.1., όπου αναφέρεται η πόλη που εδρεύει το κάθε τμήμα και το έτος ίδρυσης ή μετατροπής του σε Α.Ε.Ι. (για τα πρώην τμήματα Τ.Ε.Ι). Για την ανάλυση χρησιμοποιήθηκαν πίνακες και γραφήματα για την σύγκριση των τμημάτων Πληροφορικής που λειτουργούν στη χώρα και εξετάζονται τα στοιχεία που προέκυψαν για τα μέλη Δ.Ε.Π., σύμφωνα με τη μεθοδολογία που ακολουθήθηκε. Στον πίνακα 4.1 αποτυπώνονται τα στοιχεία του συνόλου των μελών Δ.Ε.Π. πανελλαδικά, λαμβάνοντας υπόψη την παραγωγικότητα, το δείκτη απήχησης και το δείκτη h.

Μέλη	901
Δημοσιεύσεις	106.374
Αναφορές	1.944.963
Μέσος όρος h index	18,6
Δημοσιεύσεις μελών Δ.Ε.Π. τα τελευταία 5 έτη	26.397
Αναφορές τα τελευταία 5 έτη	863.932
Μέσος όρος του h index τα τελευταία 5 έτη	12,5
Μέσος όρος δημοσιεύσεων ανά μέλος Δ.Ε.Π.	117,4
Μέσος όρος αναφορών ανά μέλος Δ.Ε.Π.	2.146,8
Μέσος όρος ετών δημοσίευσης	24,4
Μέσος όρος δημοσιεύσεων ανά μέλος Δ.Ε.Π. ανά έτος δημοσίευσης	4,7

Πίνακας 4.1. Συγκεντρωτικά στοιχεία μελών Δ.Ε.Π. Πληροφορικής στην Ελλάδα.

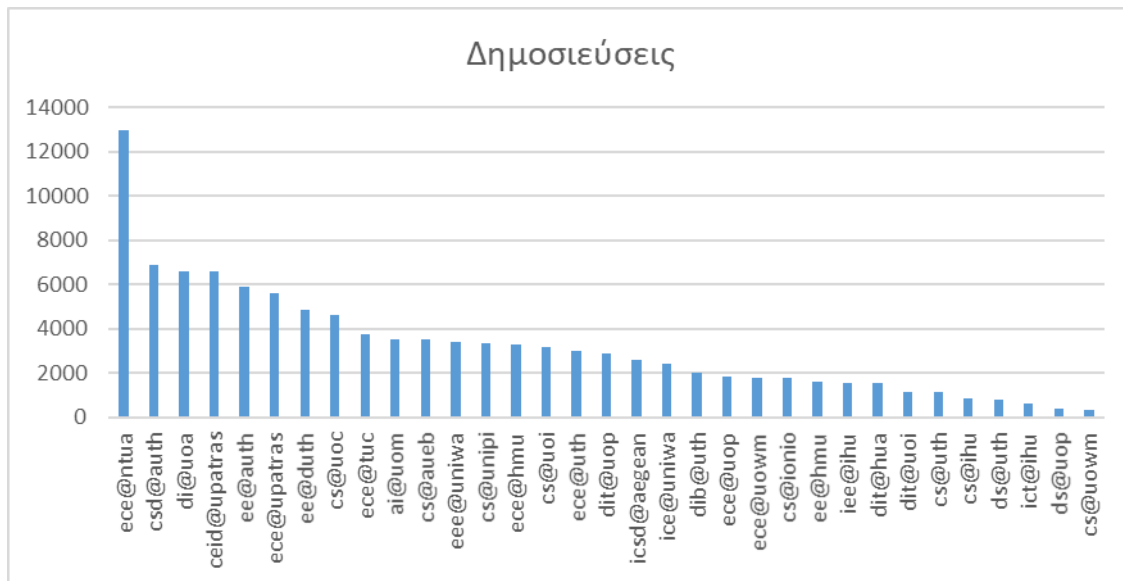
Στις ενότητες που ακολουθούν, αρχικά παρουσιάζονται και αξιολογούνται τα επιστημονομετρικά δεδομένα ανά τμήμα πληροφορικής σε πανελλαδικό επίπεδο. Στη δεύτερη ενότητα ακολουθεί η συγκριτική εξέταση των μελών Δ.Ε.Π. σε πανελλαδικό επίπεδο, ενώ στην τρίτη ενότητα η σύγκριση των μελών Δ.Ε.Π. γίνεται σε επίπεδο περιφερειακής ενότητας. Σε κάθε περίπτωση, εξετάζεται η δυνατότητα εξέλιξης των μελών Δ.Ε.Π. στην επόμενη βαθμίδα, σύμφωνα με την αξιολόγηση του επιστημονικού τους έργου.

4.1 Συνολικά επιστημονικά δεδομένα τμημάτων Πανελλαδικά

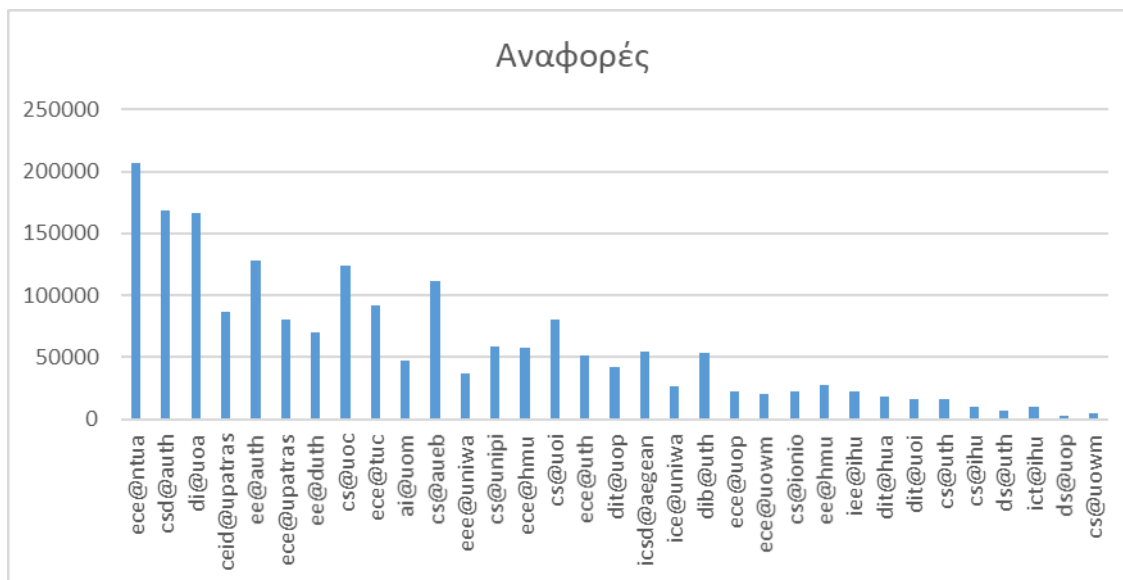
Στον Πίνακα 4.2 και στα Γραφήματα 4.1 και 4.2, παρουσιάζεται το σύνολο των δημοσιεύσεων και των αναφορών ανά τμήμα πληροφορικής στην Ελλάδα.

Πανεπιστήμια	Δημοσιεύσεις	Αναφορές
ece@ntua	12966	206925
csd@auth	6876	168189
di@uoa	6582	166525
ceid@upatras	6575	86499
ee@auth	5876	127981
ece@upatras	5609	80129
ee@duth	4860	70419
cs@uoc	4616	123869
ece@tuc	3735	92399
ai@uom	3549	47645
cs@aueb	3514	111123
eee@uniwa	3414	37100
cs@unipi	3341	58768
ece@hmu	3321	57631
cs@uoi	3151	80314
ece@uth	2984	51089
dit@uop	2876	41909
icsd@aegean	2616	55123
ice@uniwa	2403	26516
dib@uth	2030	53188
ece@uop	1817	22217
ece@uowm	1795	20073
cs@ionio	1778	22514
ee@hmu	1627	27982
iee@ihu	1573	22849
dit@hua	1559	18705
dit@uoi	1175	16693
cs@uth	1171	16558
cs@ihu	857	9647
ds@uth	810	6592
ict@ihu	606	9869
ds@uop	386	3058
cs@uowm	326	4865

Πίνακας 4.2. Συνολικός αριθμός δημοσιεύσεων και αναφορών ανά τμήμα πληροφορικής.



Γράφημα 4.1. Σύνολο δημοσιεύσεων ανά τμήμα πληροφορικής.



Γράφημα 4.2. Σύνολο αναφορών ανά τμήμα πληροφορικής.

Όπως γίνεται αντιληπτό, το μεγαλύτερο όγκο δημοσιεύσεων, αλλά και αναφορών, παρουσιάζει η Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου

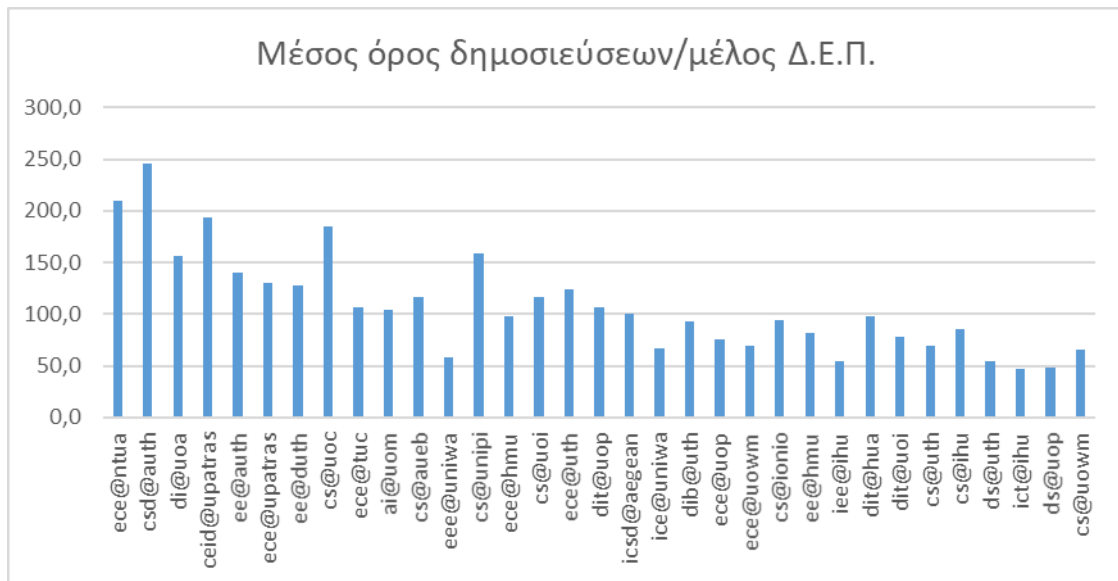
Πολυτεχνείου (ΕΜΠ). Ακολουθούν το Τμήμα Πληροφορικής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης (ΑΠΘ), το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού Καποδιστριακού Πανεπιστημίου Αθηνών (ΕΚΠΑ) και το Τμήμα Μηχανικών Η/Υ & Πληροφορικής του Πανεπιστημίου Πατρών, τα οποία εμφανίζουν σχεδόν τις μισές δημοσιεύσεις σε σχέση με τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ. Προφανώς θα πρέπει κανείς να λάβει υπόψη ότι, το ΕΜΠ διαθέτει το μεγαλύτερο αριθμό ακαδημαϊκών σε σχέση με όλα τα υπόλοιπα τμήματα (βλ. πίν. 4.3.), επομένως είναι αναμενόμενος ο μεγάλος αριθμός δημοσιεύσεων και αναφορών του τμήματος. Ωστόσο, εξετάζοντας τον αριθμό των αναφορών του Τμήματος Πληροφορικής του ΑΠΘ και του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ, διαπιστώνει κανείς ότι αν και έχουν τις μισές δημοσιεύσεις σε σχέση με τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ, ο αριθμός αναφορών -και επομένως ο δείκτης απήχησης του έργου τους- δεν απέχει πολύ από το πρώτο.

Παρομοίως, υψηλό δείκτη απήχησης (αρ. αναφορών > 100,000) εμφανίζουν το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ, το Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης και το Τμήμα Πληροφορικής του Οικονομικού Πανεπιστημίου Αθηνών, αν και ο αριθμός δημοσιεύσεων είναι σχετικά μικρός (> 6,000). Αντίθετα, η μικρότερη παραγωγικότητα, που συνοδεύεται και από χαμηλό δείκτη απήχησης, εντοπίζεται στο Τμήμα Πληροφορικής του Πανεπιστημίου Δυτικής Μακεδονίας και στο Τμήμα Ψηφιακών Συστημάτων της Σχολής Οικονομίας και Τεχνολογίας του Πανεπιστημίου Πελοποννήσου (με αρ. δημοσιεύσεων < 400 και αρ. αναφορών < 4000).

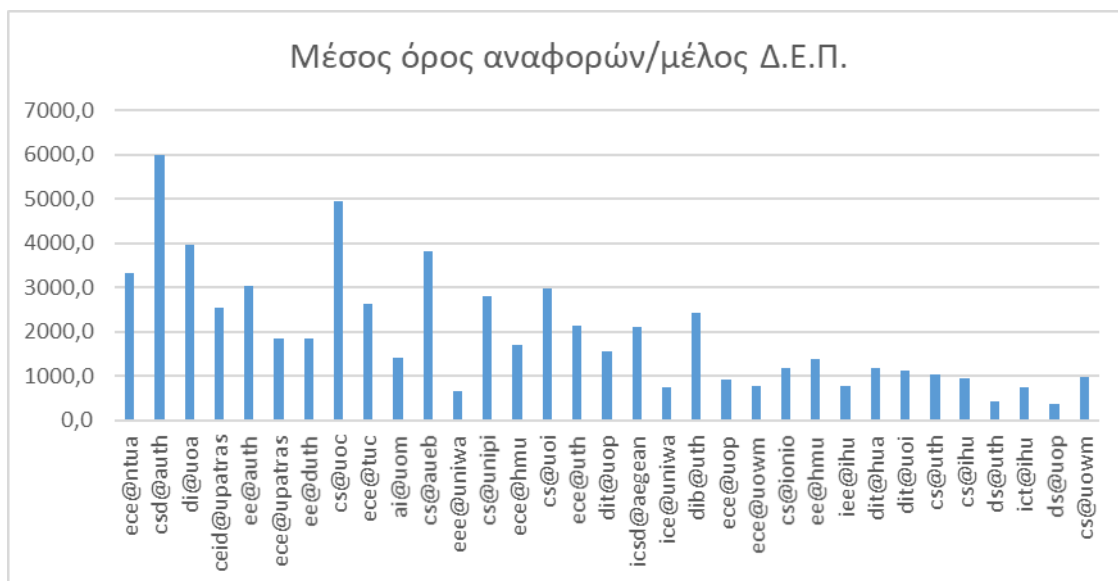
Στον πίνακα 4.3 και στα Γραφήματα 4.3 και 4.4 παρουσιάζονται στοιχεία για το μέσο όρο των δημοσιεύσεων και των αναφορών που αντιστοιχούν ανά μέλος Δ.Ε.Π. σε κάθε τμήμα. Επιπλέον, ο πίνακας 4.3 παρέχει αριθμητικά στοιχεία για το σύνολο των μελών Δ.Ε.Π. που διαθέτει το κάθε τμήμα πληροφορικής.

Πανεπιστήμια	Μέλη Δ.Ε.Π.	Μέσος όρος δημοσιεύσεων/μέλος Δ.Ε.Π.	Μέσος όρος αναφορών/μέλος Δ.Ε.Π.
ece@ntua	62	209,1	3337,5
csd@auth	28	245,6	6006,8
di@uoa	42	156,7	3964,9
ceid@upatras	34	193,4	2544,1
ee@auth	42	139,9	3047,2
ece@upatras	43	130,4	1863,5
ee@duth	38	127,9	1853,1
cs@uoc	25	184,6	4954,8
ece@tuc	35	106,7	2640,0
ai@uom	34	104,4	1401,3
cs@aueb	30	117,1	3831,8
eee@uniwa	59	57,9	650,9
cs@unipi	21	159,1	2798,5
ece@hmu	34	97,7	1695,0
cs@uoi	27	116,7	2974,6
ece@uth	24	124,3	2128,7
dit@uop	27	106,5	1552,2
icsd@aegean	26	100,6	2120,1
ice@uniwa	36	66,8	757,6
dib@uth	22	92,3	2417,6
ece@uop	24	75,7	925,7
ece@uowm	26	69,0	772,0
cs@ionio	19	93,6	1184,9
ee@hmu	20	81,4	1399,1
iee@ihu	29	54,2	787,9
dit@hua	16	97,4	1169,1
dit@uoi	15	78,3	1112,9
cs@uth	17	68,9	1034,9
cs@ihu	10	85,7	964,7
ds@uth	15	54,0	439,5
ict@ihu	13	46,6	759,2
ds@uop	8	48,3	382,3
cs@uowm	5	65,2	973,0

Πίν. 4.3 Μέσος όρος δημοσιεύσεων και αναφορών ανά μέλος Δ.Ε.Π..



Γράφημα 4.3. Μέσος όρος δημοσιεύσεων ανά μέλος Δ.Ε.Π..



Γράφημα 4.4. Μέσος όρος αναφορών ανά μέλος Δ.Ε.Π..

Αρχικά, διαπιστώνουμε ότι η Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ και το Τμήμα Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών του Πανεπιστημίου Δυτικής Αττικής διαθέτουν το μεγαλύτερο αριθμό μελών Δ.Ε.Π. (62 και 59 μέλη Δ.Ε.Π. αντίστοιχα). Ο μικρότερος αριθμός μελών Δ.Ε.Π. εντοπίζεται στο Τμήμα Πληροφορικής του Πανεπιστημίου Δυτικής Μακεδονίας (5 μέλη Δ.Ε.Π.) και στο Τμήμα Ψηφιακών Συστημάτων του Πανεπιστημίου Πελοποννήσου (8 μέλη Δ.Ε.Π.).

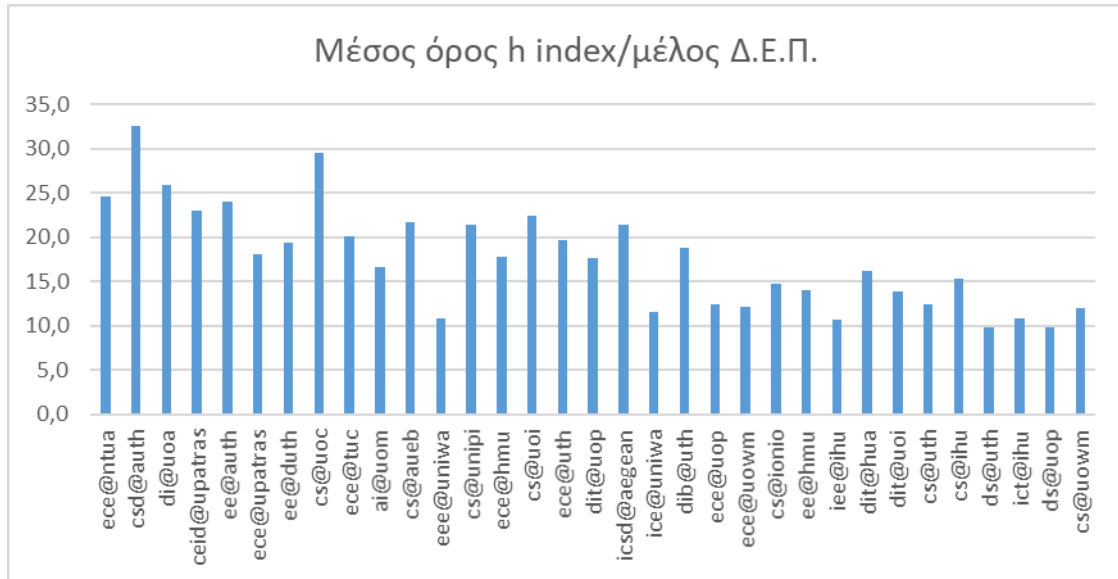
Στη συνέχεια, εξετάζοντας το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π., διαπιστώνεται ότι τη μεγαλύτερη παραγωγικότητα εμφανίζουν τα μέλη Δ.Ε.Π. του Τμήματος Πληροφορικής του ΑΠΘ με μ.ο. 246 δημοσιεύσεις ανά μέλος, παρουσιάζοντας ταυτόχρονα και το μεγαλύτερο δείκτη απήχησης με μ.ο. 6006 αναφορές ανά μέλος Δ.Ε.Π.. Ωστόσο, σε αυτό συμβάλλει ως ένα βαθμό το γεγονός ότι το πανεπιστήμιο διαθέτει μόλις 28 μέλη Δ.Ε.Π.. Αντίθετα, στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ τα μέλη Δ.Ε.Π. εμφανίζουν μικρότερη παραγωγικότητα (209 δημοσιεύσεις ανά μέλος Δ.Ε.Π.) και αρκετά μικρότερο δείκτη απήχησης (3337 αναφορές ανά μέλος Δ.Ε.Π.), τη στιγμή που το τμήμα στο σύνολό του είχε την πρωτοκαθεδρία σε δημοσιεύσεις και αναφορές πανελλαδικά. Υπενθυμίζεται, όμως, ότι το τμήμα διαθέτει συνολικά 62 μέλη Δ.Ε.Π. και αυτό φαίνεται να επηρεάζει το μ.ο. που αντιστοιχεί σε κάθε μέλος Δ.Ε.Π..

Από την άλλη πλευρά, ο χαμηλότερος μ.ο. δημοσιεύσεων και αναφορών συνεχίζει να εντοπίζεται στο Τμήμα Πληροφορικής του Πανεπιστημίου Δυτικής Μακεδονίας και στο Τμήμα Ψηφιακών Συστημάτων της Σχολής Οικονομίας και Τεχνολογίας του Πανεπιστημίου Πελοποννήσου, όπου ο συνολικός αριθμός μελών Δ.Ε.Π. είναι πολύ μικρός (<10). Επιπλέον, χαμηλός είναι και ο μ.ο. των δημοσιεύσεων και αναφορών μελών Δ.Ε.Π. του Τμήματος Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών του Διεθνούς Πανεπιστημίου Ελλάδος (ΔΙ.ΠΑ.Ε.).

Εκτός από τα ποσοτικά δεδομένα που παρατέθηκαν παραπάνω, μια πιο ποιοτική διερεύνηση του παραγόμενου έργου των τμημάτων πληροφορικής εστιάζει στο συνδυασμό των δημοσιεύσεων και αναφορών του κάθε τμήματος, με βάση το δείκτη h (h -index). Στον Πίνακα 4.4 και το Γράφημα 4.5 παρουσιάζονται στοιχεία σχετικά με το μέγιστο και ελάχιστο δείκτη h (h -index), καθώς και το μ.ο. ανά τμήμα πληροφορικής.

Πανεπιστήμια	Μέσος όρος h index	Μέγιστο h index	Ελάχιστο h index
ece@ntua	24,7	61	3
csd@auth	32,5	87	11
di@uoa	26,0	73	2
ceid@upatras	23,0	49	3
ee@auth	24,1	86	3
ece@upatras	18,0	51	2
ee@duth	19,3	48	3
cs@uoc	29,5	51	8
ece@tuc	20,1	66	2
ai@uom	16,7	36	2
cs@aueb	21,7	55	0
eee@uniwa	10,9	27	0
cs@unipi	21,4	51	2
ece@hmu	17,8	52	4
cs@uoi	22,4	59	8
ece@uth	19,7	62	6
dit@uop	17,7	37	3
icsd@aegean	21,4	46	6
ice@uniwa	11,5	32	0
dib@uth	18,9	53	6
ece@uop	12,5	40	1
ece@uowm	12,1	28	2
cs@ionio	14,7	27	6
ee@hmu	14,1	44	2
iee@ihu	10,7	34	2
dit@hua	16,3	25	5
dit@uoi	13,9	39	4
cs@uth	12,4	40	0
cs@ihu	15,3	29	7
ds@uth	9,9	17	2
ict@ihu	10,8	20	4
ds@uop	9,9	18	3
cs@uowm	12,0	26	2

Πίνακας 4.4. Μέγιστος, ελάχιστος και μ. ο. δείκτη h (h-index) ανά τμήμα πληροφορικής.



Γράφημα 4.5. Μέσος όρος δείκτη h (h-index) ανά μέλος Δ.Ε.Π.

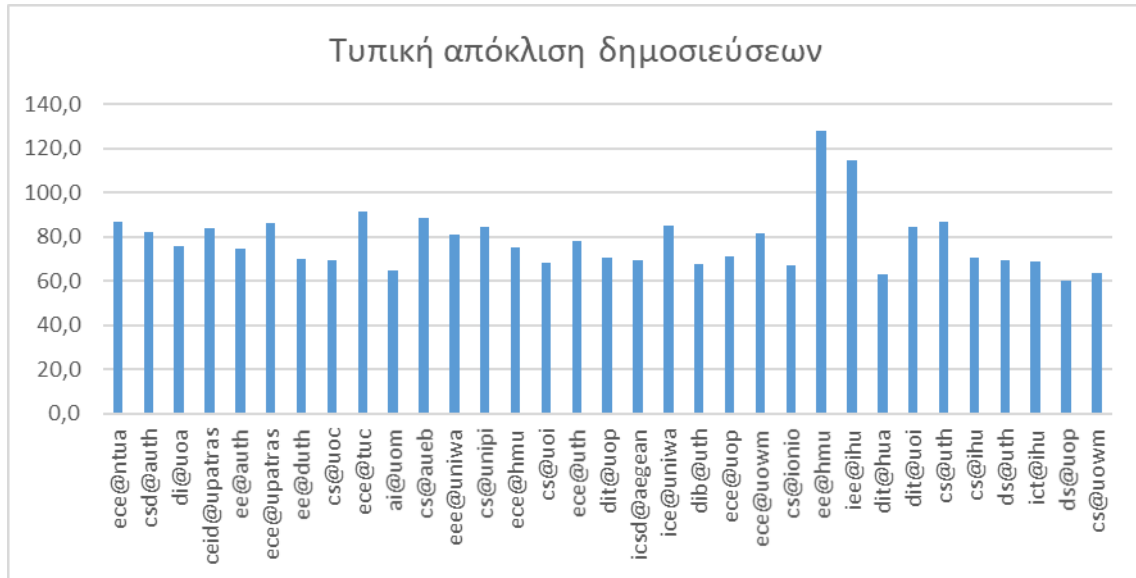
Από τα παραπάνω στοιχεία προκύπτει ότι το Τμήμα Πληροφορικής του ΑΠΘ εμφανίζει την υψηλότερη παραγωγικότητα και απήχηση των δημοσιεύσεων των μελών Δ.Ε.Π.. Ακολουθούν το Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης, το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ, η Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ, το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ και το Τμήμα Μηχανικών Η/Υ & Πληροφορικής του Πανεπιστημίου Πατρών.

Αντίθετα, τη χαμηλότερη παραγωγικότητα και απήχηση των δημοσιεύσεων των μελών Δ.Ε.Π. εμφανίζουν το Τμήμα Ψηφιακών Συστημάτων του Πανεπιστημίου Πελοποννήσου, το Τμήμα Ψηφιακών Συστημάτων του Πανεπιστημίου Θεσσαλίας, το Τμήμα Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών του ΔΙ.ΠΑ.Ε. καθώς και το Τμήμα Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών του Πανεπιστημίου Δυτικής Αττικής.

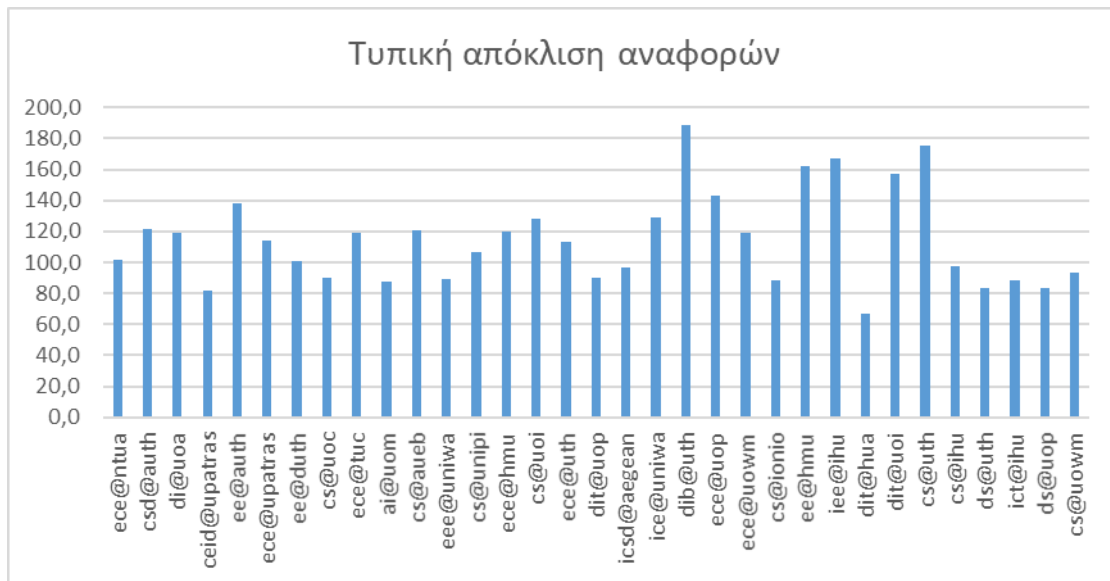
Για να ελέγξουμε εάν το δείγμα μας αντιπροσωπεύει ικανοποιητικά τα δεδομένα, εξετάστηκε και ο παράγοντας της τυπικής απόκλισης των μετρήσεων για τις δημοσιεύσεις και τις αναφορές ανά τμήμα (Πίν. 4.4).

Πανεπιστήμια	Τυπική απόκλιση δημοσιεύσεων	Τυπική απόκλιση αναφορών
ece@ntua	86,7	101,8
csd@auth	82,4	121,6
di@uoa	75,6	119,2
ceid@upatras	83,9	82,2
ee@auth	74,8	138,4
ece@upatras	86,0	114,0
ee@duth	70,0	101,0
cs@uoc	69,6	89,7
ece@tuc	91,7	119,2
ai@uom	64,6	87,8
cs@aueb	88,6	120,5
eee@uniwa	81,0	89,0
cs@unipi	84,4	106,4
ece@hmu	75,0	119,8
cs@uoi	68,3	128,5
ece@uth	78,3	113,0
dit@uop	70,5	90,5
icsd@aegean	69,4	96,5
ice@uniwa	84,9	128,7
dib@uth	67,9	188,8
ece@uop	71,2	143,0
ece@uowm	81,9	119,4
cs@ionio	67,1	88,1
ee@hmu	127,8	162,4
iee@ihu	114,9	167,0
dit@hua	63,1	66,7
dit@uoi	84,5	156,8
cs@uth	86,9	175,6
cs@ihu	70,7	97,2
ds@uth	69,3	83,8
ict@ihu	68,7	88,3
ds@uop	60,0	83,3
cs@uowm	63,5	93,1

Πίνακας 4.5. Τυπική απόκλιση δημοσιεύσεων και αναφορών ανά τμήμα πληροφορικής.



Γράφημα 4.6. Τυπική απόκλιση δημοσιεύσεων ανά τμήμα πληροφορικής.



Γράφημα 4.7. Τυπική απόκλιση αναφορών ανά τμήμα πληροφορικής.

Σύμφωνα με τα δεδομένα που προέκυψαν τη μεγαλύτερη τυπική απόκλιση δημοσιεύσεων εμφανίζει το Τμήμα Ηλεκτρονικών Μηχανικών του Ελληνικού Μεσογειακού Πανεπιστημίου, ενώ ακολουθεί το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε. Όσον αφορά την τυπική απόκλιση αναφορών, οι μεγαλύτερες τιμές εντοπίζονται στο Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική και στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Θεσσαλίας.

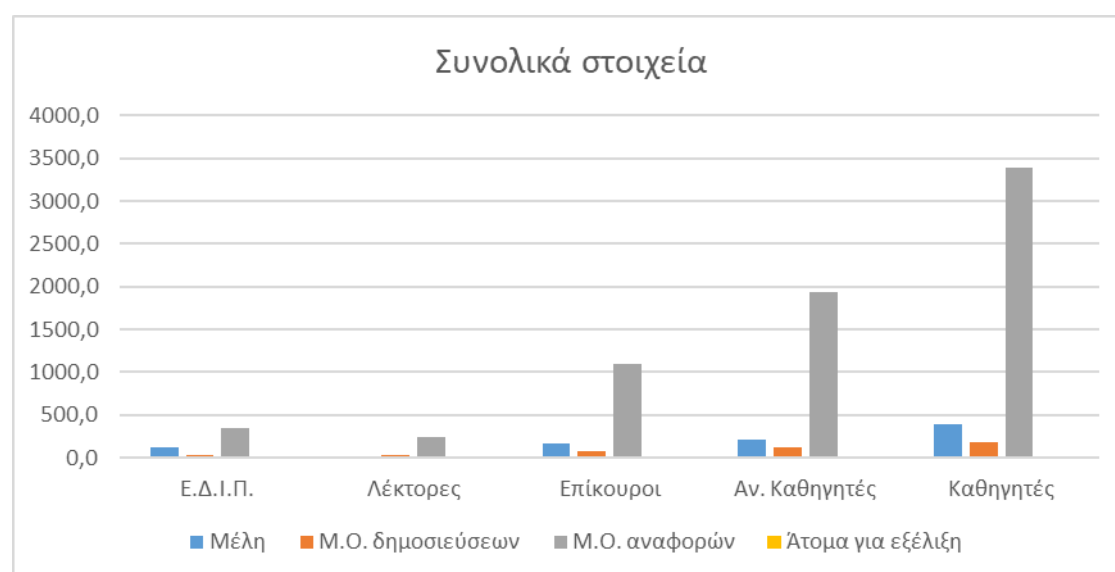
4.2 Αξιολόγηση μελών Δ.Ε.Π. Πανελλαδικά

Η ενότητα αυτή παρουσιάζει στοιχεία για τα μέλη Δ.Ε.Π. των τμημάτων πληροφορικής πανελλαδικά. Εξετάζονται τα μέλη Δ.Ε.Π. που μπορούν να εξελιχθούν στην επόμενη βαθμίδα, λαμβάνοντας υπόψη την παραγωγικότητα (δημοσιεύσεις) και την απήχηση του επιστημονικού τους έργου (αναφορές), συγκριτικά με τα στοιχεία που συγκεντρώθηκαν για κάθε βαθμίδα μελών Δ.Ε.Π. σε πανελλαδικό επίπεδο.

Στο σημείο αυτό πρέπει να διευκρινιστεί ότι, στην παρούσα εργασία θεωρήθηκαν ως ώριμα για εξέλιξη τα μέλη της προηγούμενης βαθμίδας που διαθέτουν περισσότερες δημοσιεύσεις και περισσότερες αναφορές από τον μ.ο. της επόμενης βαθμίδας.

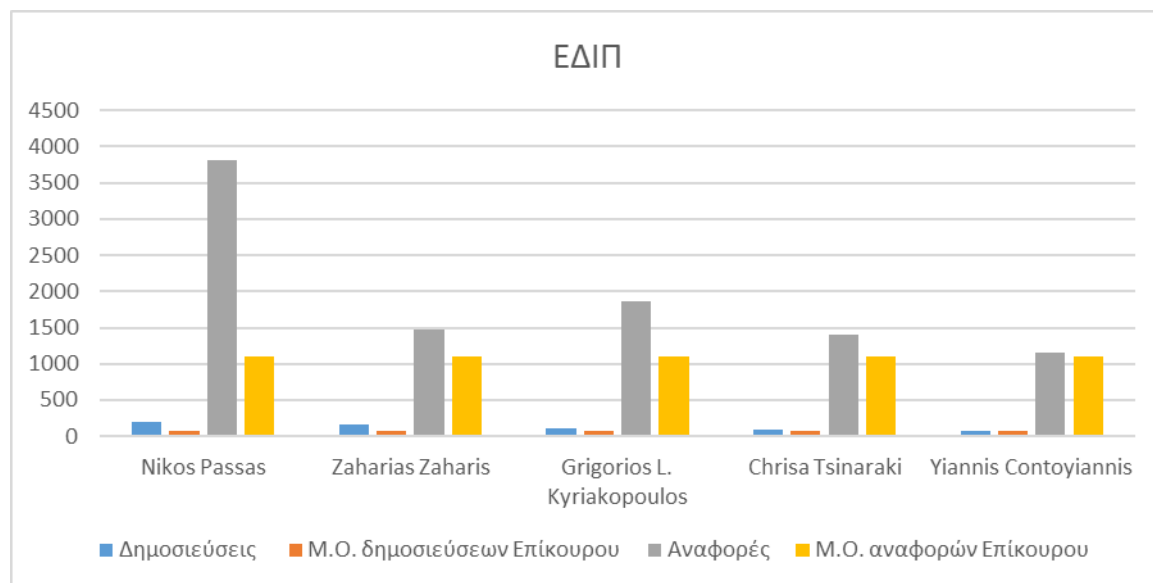
	Μέλη	Μ.Ο. δημοσιεύσεων	Μ.Ο. αναφορών	Άτομα για εξέλιξη
Ε.Δ.Ι.Π.	120	31,1	351,0	5
Λέκτορες	21	23,6	246,3	2
Επίκουροι Καθηγητές	171	67,9	1096,5	9
Αν. Καθηγητές	205	113,1	1931	15
Καθηγητές	389	173,4	3391,1	

Πίνακας 4.6. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. σε πανελλαδικό επίπεδο ανά βαθμίδα Δ.Ε.Π..



Γράφημα 4.8. Συνολικά στοιχεία μ.ο. δημοσιεύσεων και αναφορών ανά βαθμίδα Δ.Ε.Π..

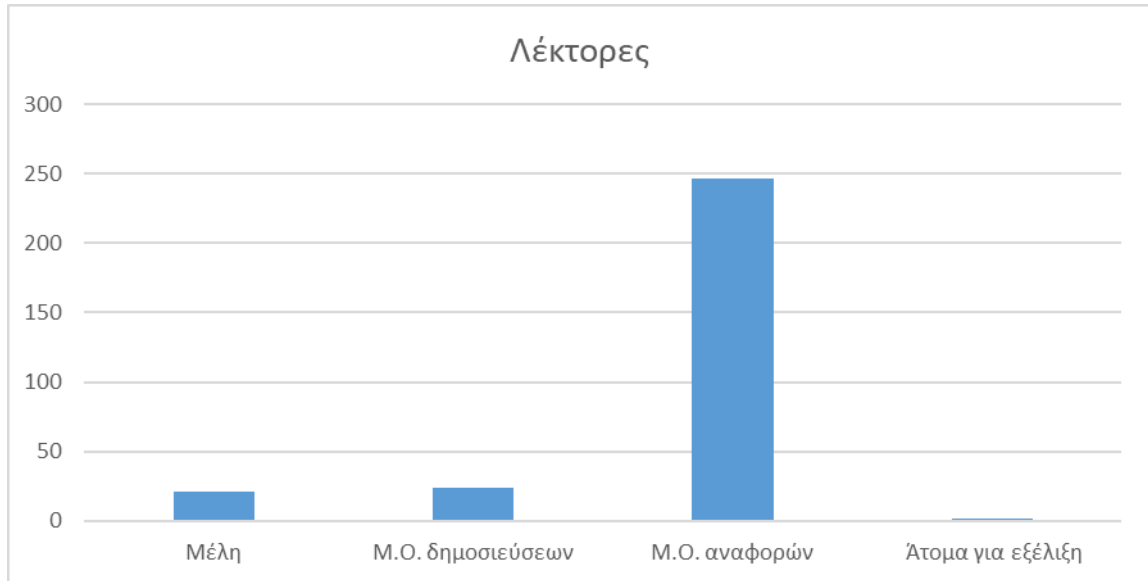
Λαμβάνοντας υπόψη τα στοιχεία που συλλέχθηκαν για τα μέλη ΕΔΙΠ των τμημάτων πληροφορικής σε πανελλαδικό επίπεδο, διαμορφώθηκε το ακόλουθο γράφημα, στο οποίο παρουσιάζονται οι περιπτώσεις που ξεχώρισαν.



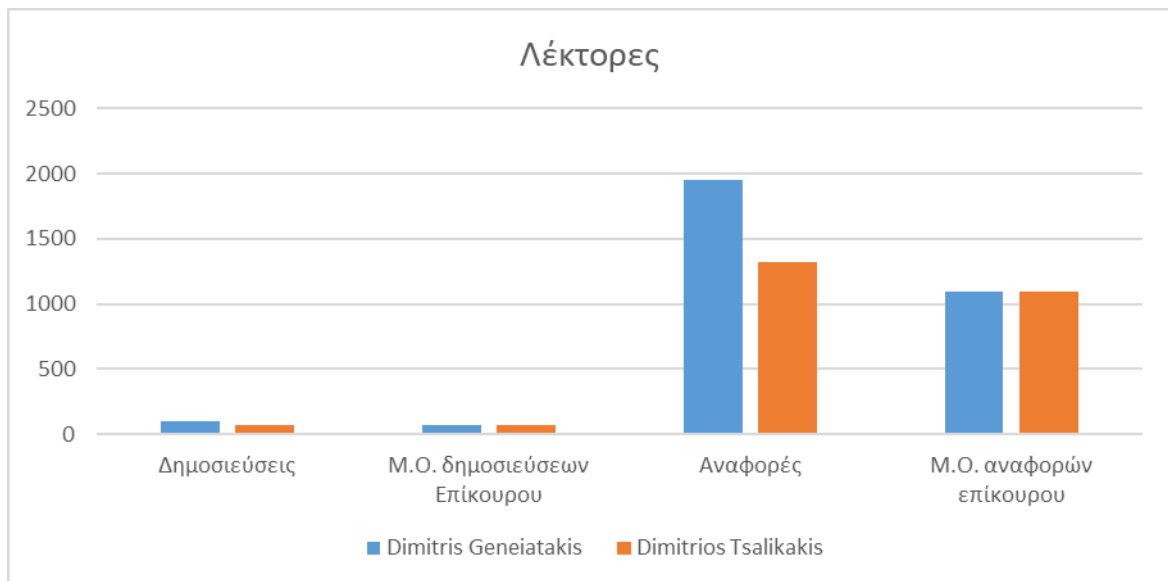
Γράφημα 4.9. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή .

Σύμφωνα, λοιπόν, με το Γράφημα 9, παρατηρούμε ότι πέντε άτομα Ειδικό Επιστημονικό Προσωπικό εμφανίζουν μεγαλύτερο αριθμό δημοσιεύσεων από το μ.ο. δημοσιεύσεων (67,9) και μ.ο. αναφορών (1096,5) της βαθμίδας του Επίκουρου Καθηγητή. Συγκεκριμένα, οι Nikos Passas από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών ΕΚΠΑ, ο Zaharias Zaharis από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ, ο Grigorios L. Kyriakopoulos από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ, η Chrisa Tsinaraki από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Πελοποννήσου και ο Yannis Contoyiannis από το Τμήμα Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών του Πανεπιστημίου Δυτικής Αττικής. Τα ανωτέρω μέλη Δ.Ε.Π. θα μπορούσαν να εξελιχθούν στη βαθμίδα του Επίκουρου Καθηγητή.

Συνεχίζοντας με τη βαθμίδα του Λέκτορα, ο μ.ο. των δημοσιεύσεων σε πανελλαδικό επίπεδο είναι 23,6 και ο μ.ο. αναφορών 246,3 (βλ. Γράφ. 4.10).



Γράφημα 4.10. Μέσος όρος δημοσιεύσεων και αναφορών της βαθμίδας του Λέκτορα πανελλαδικά.



Γράφημα 4.11. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.

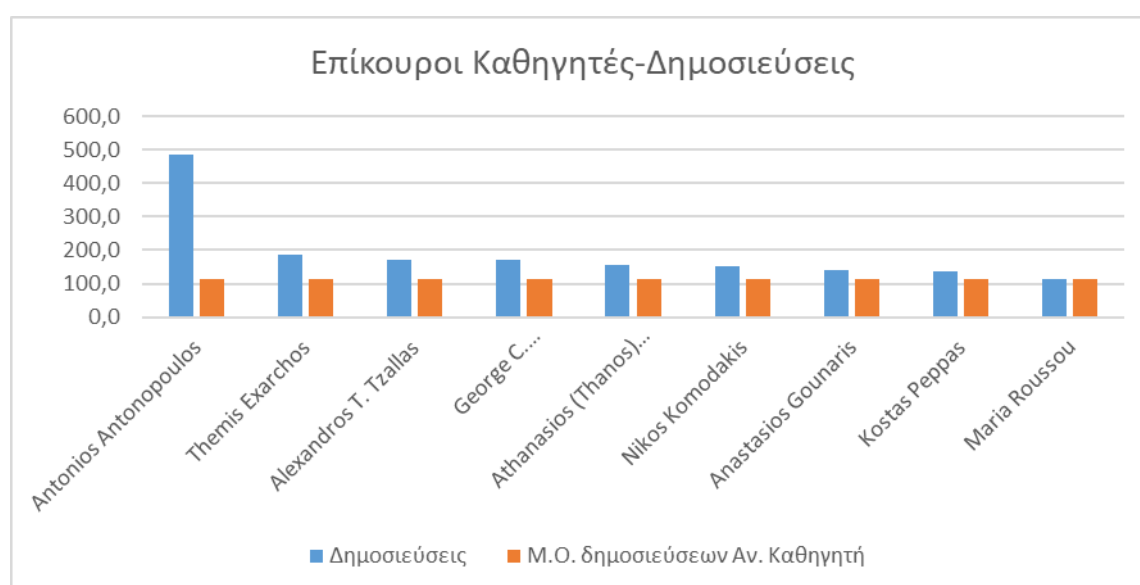
Από τα στοιχεία που συλλέχθηκαν προκύπτει ότι δύο μέλη Δ.Ε.Π. που βρίσκονται στη βαθμίδα του Λέκτορα θα μπορούσαν να εξελιχθούν σε Επίκουροι Καθηγητές, καθώς ο αριθμός των δημοσιεύσεων και αναφορών τους ξεπερνάει το μ.ο. της βαθμίδας του Επίκουρου Καθηγητή (Γράφ. 4.11). Πρόκειται για τον Dimitris Geneiatakis από το Τμήμα

Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ με 99 δημοσιεύσεις και 1948 αναφορές και τον Dimitrios Tsalikakis από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Δυτ. Μακεδονίας με 73 δημοσιεύσεις και 1320 αναφορές.

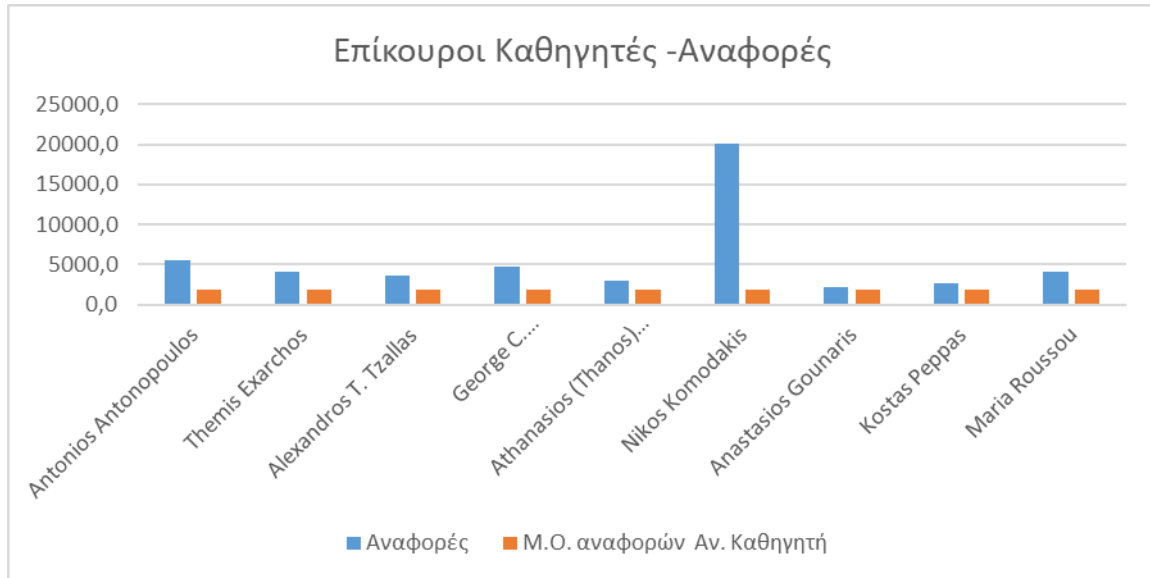
Στη βαθμίδα των Επίκουρων Καθηγητών τα μέλη Δ.Ε.Π. πανελλαδικά εμφανίζουν μ.ο. 67,9 δημοσιεύσεις και μ.ο. 1096,5 αναφορές (Γραφ. 4.12).



Γράφημα 4.12. Μέσος όρος δημοσιεύσεων και αναφορών της βαθμίδας του Επίκουρου Καθηγητή πανελλαδικά.



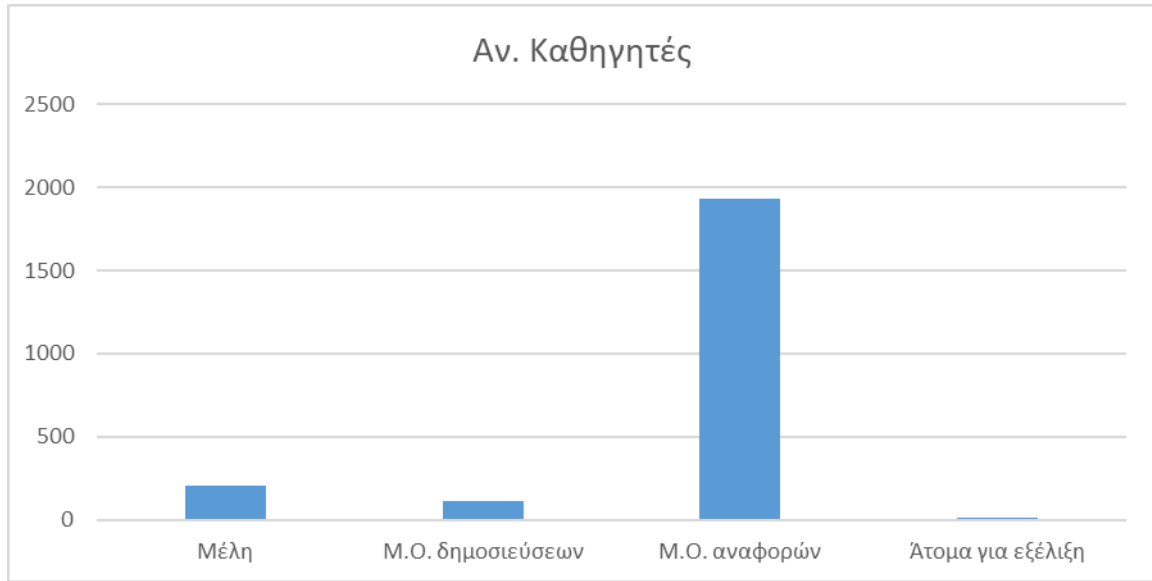
Γράφημα 4.13. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή (δημοσιεύσεις).



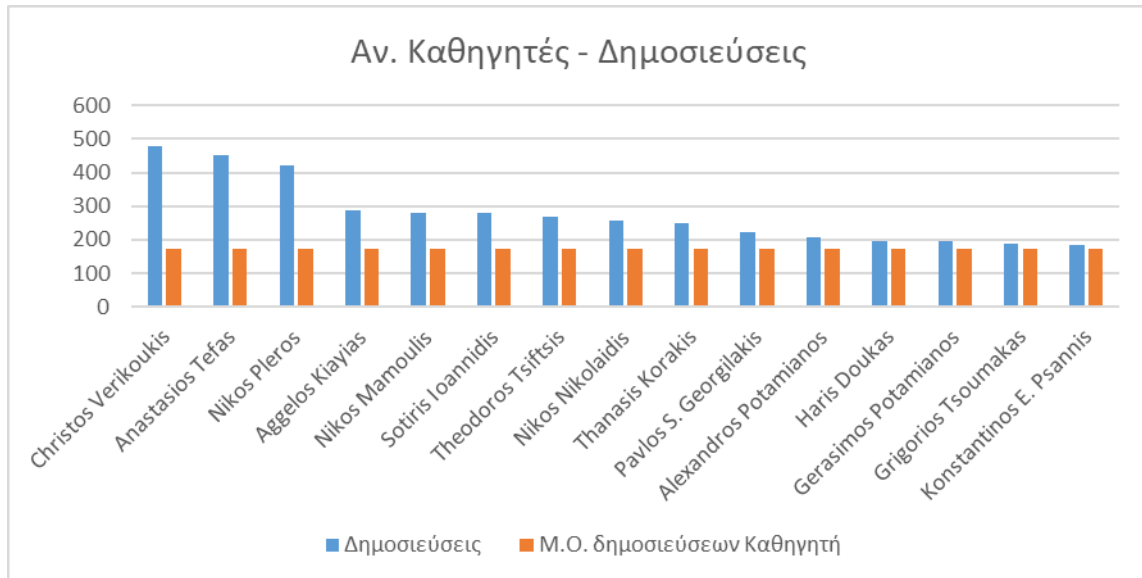
Γράφημα 4.14. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή (αναφορές).

Στην περίπτωση των Επίκουρων Καθηγητών, διαπιστώνεται ότι εννέα μέλη Δ.Ε.Π. είναι ώριμα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή (Γράφ. 13-14). Πρόκειται για τους: Antonios Antonopoulos της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ, Themis Exarchos του Τμήματος Πληροφορικής του Ιονίου Πανεπιστημίου, Alexandros T. Tzallas του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων, George C. Alexandropoulos του Τμήματος Πληροφορικής και Τηλεπικοινωνιών ΕΚΠΑ, Athanasios (Thanos) Voulodimos του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών του Πανεπιστημίου Δυτικής Αττικής, Nikos Komodakis του Τμήματος Επιστήμης Υπολογιστών από το Πανεπιστήμιο Κρήτης, Anastasios Gounaris από το Τμήμα Πληροφορικής του ΑΠΘ, Kostas Peppas από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών της Σχολής Οικονομίας και Τεχνολογίας του Πανεπιστημίου Πελοποννήσου και Maria Roussou του Τμήματος Πληροφορικής και Τηλεπικοινωνιών ΕΚΠΑ. Τα μέλη αυτά εμφανίζουν μ.ο. δημοσιεύσεων πάνω από 113,1 και μ.ο. αναφορών πάνω από 1931, ξεπερνώντας τον αντίστοιχο μ.ο. της βαθμίδας του Επίκουρου Καθηγητή.

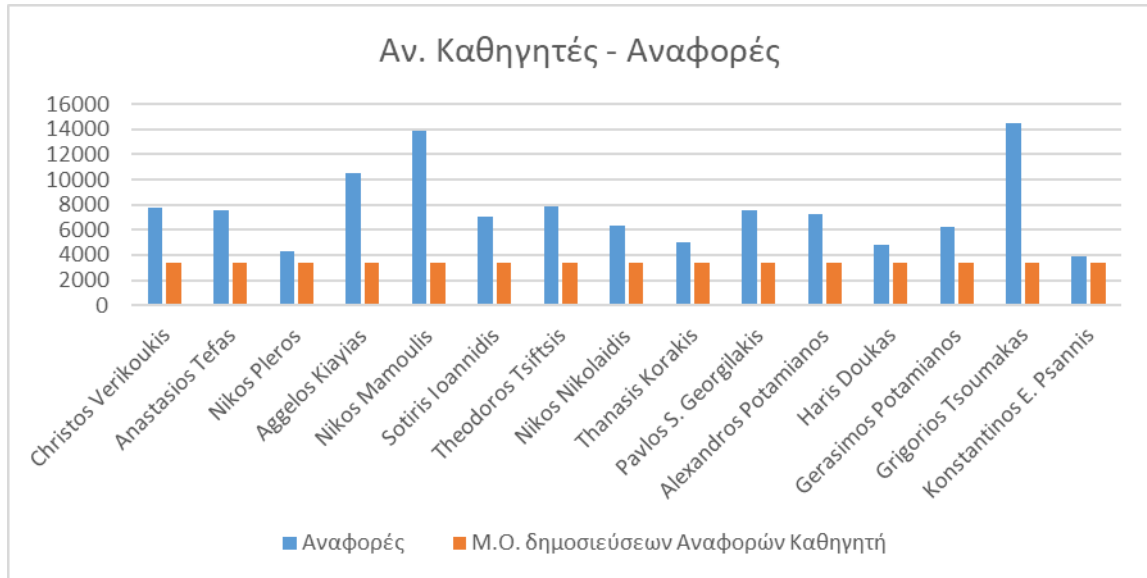
Στην περίπτωση των Αναπληρωτών Καθηγητών, τα δεδομένα που προέκυψαν είναι μέσος όρος 113,1 δημοσιεύσεις και 1931 αναφορές ανά μέλος Δ.Ε.Π..



Γράφημα 4.15. Μέσος όρος δημοσιεύσεων και αναφορών της βαθμίδας του Αναπληρωτή Καθηγητή πανελλαδικά.



Γράφημα 4.16. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή (δημοσιεύσεις).



Γράφημα 4.17. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή (αναφορές).

Εξετάζοντας τους Αναπληρωτές καθηγητές, παρατηρείται ότι 15 μέλη Δ.Ε.Π. έχουν μ.ο. δημοσιεύσεων πάνω από 173,4 και μ.ο. αναφορών πάνω από 3391,2, ξεπερνώντας τον αντίστοιχο μ.ο. της βαθμίδας του Καθηγητή και θα μπορούσαν να εξελιχθούν σε αυτήν. Συγκεκριμένα, πρόκειται για τους: Konstantinos E. Psannis από το τμήμα Εφαρμοσμένης πληροφορικής του ΠΑΜΑΚ, Nikos Mamoulis από το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Ιωαννίνων, Theodoros Tsiftsis από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Θεσσαλίας, Anastasios Tefas, Nikos Pleros, Nikos Nikolaidis, Grigorios Tsoumakas από το Τμήμα Πληροφορικής του ΑΠΘ, Aggelos Kiaγias του Τμήματος Πληροφορικής και Τηλεπικοινωνιών ΕΚΠΑ, Pavlos S. Georgilakis, Alexandros Potamianos, Haris Doukas από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ, Sotiris Ioannidis από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης, και Thanasis Korakis και Gerasimos Potamianos από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ.

4.3 Αξιολόγηση μελών Δ.Ε.Π. ανά Περιφερειακή Ενότητα

Στην ενότητα αυτή παρουσιάζονται στοιχεία για τα μέλη Δ.Ε.Π. των τμημάτων πληροφορικής ανά περιφερειακή ενότητα. Πιο συγκεκριμένα, περιλαμβάνονται τα τμήματα από τις περιφερειακές ενότητες: Ανατολικής Μακεδονίας, Κεντρικής Μακεδονίας, Δυτικής Μακεδονίας, Θεσσαλίας, Στερεάς Ελλάδας, Ηπείρου, Ιονίων νήσων, Αττικής, Πελοποννήσου, Δυτικής Ελλάδας, Κρήτης και Βορείου Αιγαίου. Σε κάθε περιφερειακή ενότητα αξιολογούνται τα μέλη Δ.Ε.Π. λαμβάνοντας υπόψη την παραγωγικότητα και την απήχηση του ερευνητικού τους έργου και εξετάζεται η δυνατότητα να εξελιχθούν στην επόμενη βαθμίδα.



Εικόνα 4.1. Χάρτης της Ελλάδας που παρουσιάζει τις περιφερειακές ενότητες.

Ανατολική Μακεδονία	ee@duth	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Δημοκρίτειο Πανεπιστήμιο Θράκης	Ξάνθη
	cs@ihu	Τμήμα Πληροφορικής, Διεθνές Πανεπιστήμιο Ελλάδος ΔΙ.ΠΑ.Ε.	Καβάλα
Κεντρική Μακεδονία	ai@uom	Πανεπιστήμιο Μακεδονίας – Σχολή Επιστημών Πληροφορίας	Θεσσαλονίκη
	csd@auth	Τμήμα Πληροφορικής, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, ΑΠΘ	Θεσσαλονίκη
	ee@auth	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης	Θεσσαλονίκη
	ict@ihu	Τμήμα Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών, Διεθνές Πανεπιστήμιο Ελλάδος, ΔΙ.ΠΑ.Ε.	Σέρρες
	iee@ihu	Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων, Διεθνές Πανεπιστήμιο Ελλάδος, ΔΙ.ΠΑ.Ε.	Θεσσαλονίκη
Δυτική Μακεδονία	cs@uowm	Τμήμα Πληροφορικής- Πανεπιστήμιο Δυτικής Μακεδονίας	Καστοριά
	ece@uowm	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Πανεπιστήμιο Δυτικής Μακεδονίας-Πολυτεχνική Σχολή	Κοζάνη
Θεσσαλία	ds@uth	Τμήμα Ψηφιακών Συστημάτων-Πανεπιστήμιο Θεσσαλίας	Λάρισα
	ece@uth	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών -Πανεπιστήμιο Θεσσαλίας	Βόλος
Στερεά Ελλάδα	cs@uth	Τμήμα Πληροφορικής και Τηλεπικοινωνιών- Πανεπιστήμιο Θεσσαλίας	Λαμία
	dib@uth	Τμήμα Πληροφορικής με Εφαρμογές στη Βιοιατρική του Πανεπιστημίου Θεσσαλίας	Λαμία
Ήπειρος	cs@uoi	Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής-Πολυτεχνική Σχολή- Πανεπιστήμιο Ιωαννίνων	Ιωάννινα
	dit@uoi	Τμήμα Πληροφορικής και Τηλεπικοινωνιών – Πανεπιστήμιο Ιωαννίνων	Ιωάννινα
Ιόνιο	cs@ionio	Τμήμα Πληροφορικής- Ιόνιο Πανεπιστήμιο	Κέρκυρα
	cs@aeub	Οικονομικό Πανεπιστήμιο Αθηνών- Τμήμα Πληροφορικής	Αθήνα
Αττική	cs@unipi	Τμήμα Πληροφορικής- Πανεπιστήμιο Πειραιώς	Πειραιάς
	di@uoa	Τμήμα Πληροφορικής και Τηλεπικοινωνιών ΕΚΠΑ	Αθήνα
	dit@hua	Τμήμα Πληροφορικής και Τηλεματικής του Χαροκόπειου Πανεπιστημίου	Αθήνα
	ece@ntua	Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών – Εθνικό Μετσόβιο Πολυτεχνείο	Αθήνα
	eee@uniwa	Τμήμα Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών -Σχολή Μηχανικών -Πανεπιστήμιο Δυτικής Αττικής	Αθήνα
Πελοπόννησος	ice@uniwa	Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών- Σχολή Μηχανικών-Πανεπιστήμιο Δυτικής Αττικής	Αθήνα
	dit@uop	Τμήμα Πληροφορικής και Τηλεπικοινωνιών της Σχολής Οικονομίας και Τεχνολογίας του – Πανεπιστήμιο Πελοποννήσου	Τρίπολη
	ds@uop	Τμήμα Ψηφιακών Συστημάτων-Σχολής Οικονομίας και Τεχνολογίας-Πανεπιστήμιο Πελοποννήσου	Σπάρτη
Δυτική Ελλάδα	ceid@upatras	Τμήμα Μηχανικών Η/Υ & Πληροφορικής-Πανεπιστήμιο Πατρών	Πάτρα
	ece@uop	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Πανεπιστήμιο Πελοποννήσου	Πάτρα
	ece@upatras	Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών-Πανεπιστήμιο Πατρών	Πάτρα
Κρήτη	cs@uoc	Τμήμα Επιστήμης Υπολογιστών- Πανεπιστήμιο Κρήτης	Ηράκλειο
	ece@hmu	Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών – Ελληνικό Μεσογειακό Πανεπιστήμιο	Ηράκλειο
	ece@tuc	Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών – Πολυτεχνείο Κρήτης	Χανιά
	ee@hmu	Τμήμα Ηλεκτρονικών Μηχανικών- Ελληνικό Μεσογειακό Πανεπιστήμιο	Χανιά
Βόρειο Αιγαίο	icsd@aegean	Τμήμα Μηχανικών Πληροφορικών και Επικοινωνιακών Συστημάτων-Πανεπιστήμιο Αιγαίου	Σάμος

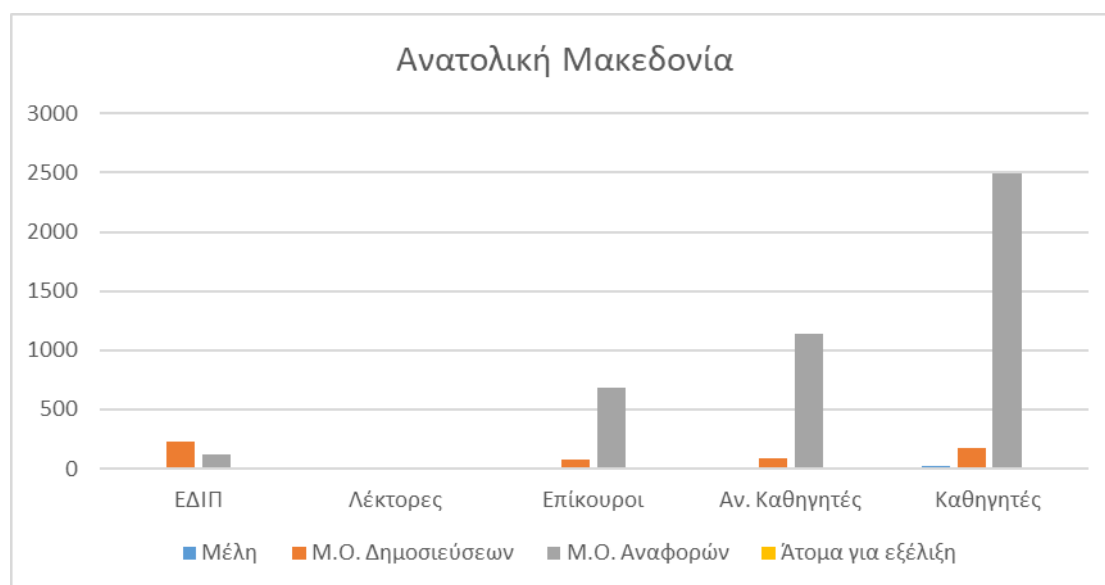
Πίνακας 4.7. Οι περιφερειακές ενότητες της Ελλάδας και τα τμήματα πληροφορικής που αντιστοιχούν σε αυτές.

ΑΝΑΤΟΛΙΚΗ ΜΑΚΕΔΟΝΙΑ

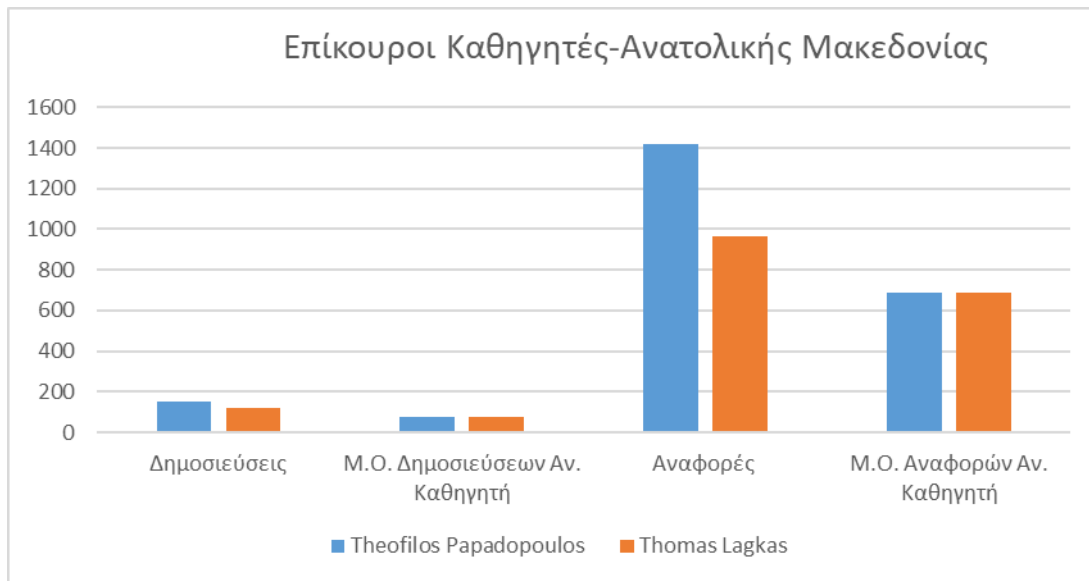
Στον πίνακα 4.8. και στο Γράφημα 4.18 αποτυπώνονται ανά βαθμίδα τα μέλη Δ.Ε.Π. των τμημάτων πληροφορικής της Ανατολικής Μακεδονίας (συνολικά 48 μέλη Δ.Ε.Π.). Σε αυτήν ανήκουν το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Δημοκρίτειου Πανεπιστημίου Θράκης (ΔΠΘ) και το τμήμα Πληροφορικής του Διεθνούς Πανεπιστημίου Ελλάδος (ΔΙ.ΠΑ.Ε.) που εδρεύει στην Καβάλα.

	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΙΠ	4	225,5	121,3	-
Λέκτορες	-	-	-	-
Επίκουροι Καθηγητές	6	77,2	685,3	2
Αν. Καθηγητές	14	81,9	1135,9	-
Καθηγητές	24	168,2	2498,6	-

Πίνακας 4.8. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Ανατολική Μακεδονία.



Γράφημα 4.18. Συνολικά στοιχεία των μελών Δ.Ε.Π. ανά βαθμίδα στην Ανατολική Μακεδονία.



Γράφημα 4.19. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.

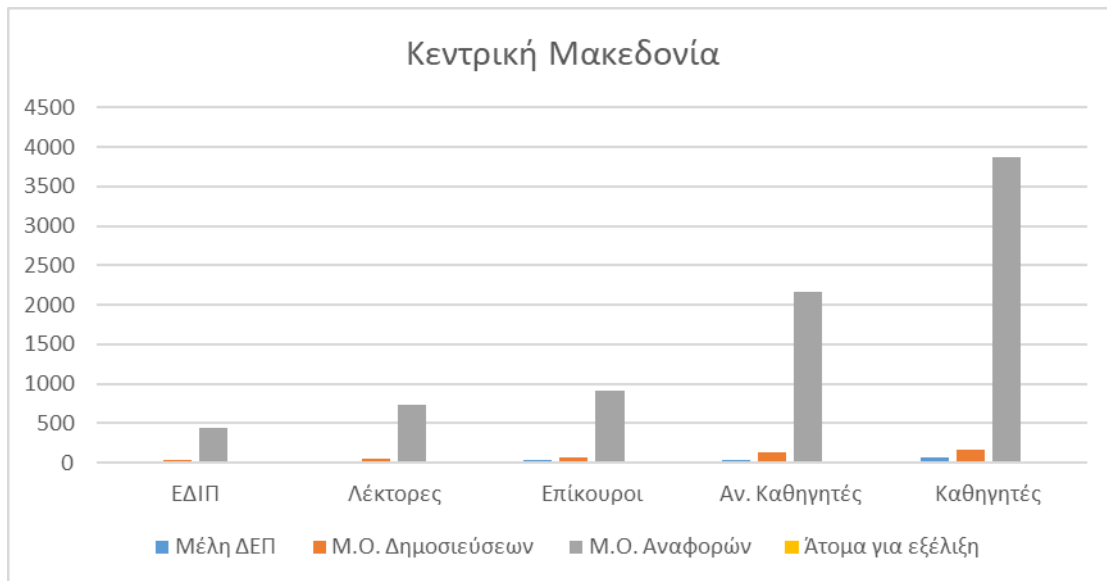
Στο παραπάνω γράφημα αναφέρονται οι Επίκουροι Καθηγητές των Πανεπιστημίων της Ανατολικής Μακεδονίας. Παρατηρούμε στην 1η θέση τον Theofilos Papadopoulos από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΔΠΘ που έχει 152 δημοσιεύσεις, αριθμός αρκετά μεγαλύτερος από το μ.ο. δημοσιεύσεων του Αναπληρωτή Καθηγητή που είναι 77,2. Επιπλέον, έχει 1420 αναφορές ενώ ο μ.ο. της βαθμίδας του Αν. Καθηγητή είναι 685,3. Παρομοίως, στη 2^η θέση, ο Thomas Lagkas από το Τμήμα Πληροφορικής του ΔΙ.ΠΑ.Ε. έχει 123 δημοσιεύσεις και 964 αναφορές. Και τα δύο μέλη Δ.Ε.Π. θα μπορούσαν να εξελιχτούν στη βαθμίδα του Αν. Καθηγητή.

ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ

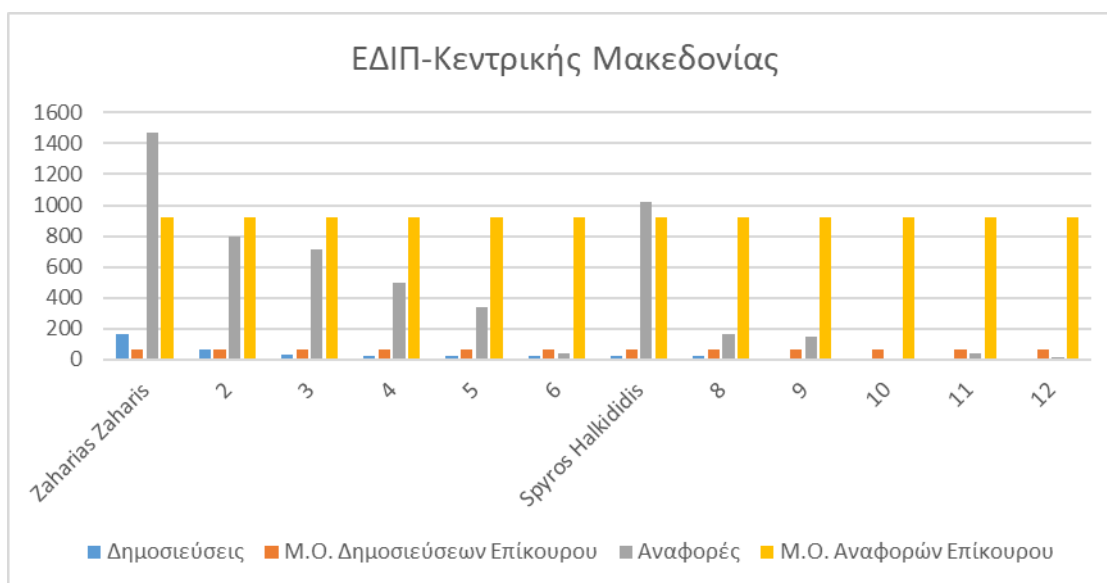
Στον πίνακα 4.9. και στο Γράφημα 4.20 αποτυπώνονται ανά βαθμίδα τα 146 μέλη Δ.Ε.Π. των τμημάτων πληροφορικής της Κεντρικής Μακεδονίας. Πρόκειται για το Τμήμα Εφαρμοσμένης Πληροφορικής του ΠΑΜΑΚ, το Τμήμα Πληροφορικής του ΑΠΘ, το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ, το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε. που εδρεύουν στη Θεσσαλονίκη και το Τμήμα Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών του ΔΙ.ΠΑ.Ε. στις Σέρρες.

	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για
ΕΔΙΠ	12	33,8	438,4	1
Λέκτορες	3	46,7	728	1
Επίκουροι Καθηγητές	25	65,4	918,1	-
Αν. Καθηγητές	37	133,1	2165,4	5
Καθηγητές	69	165,1	3868,1	-

Πίνακας 4.9. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Κεντρική Μακεδονία.

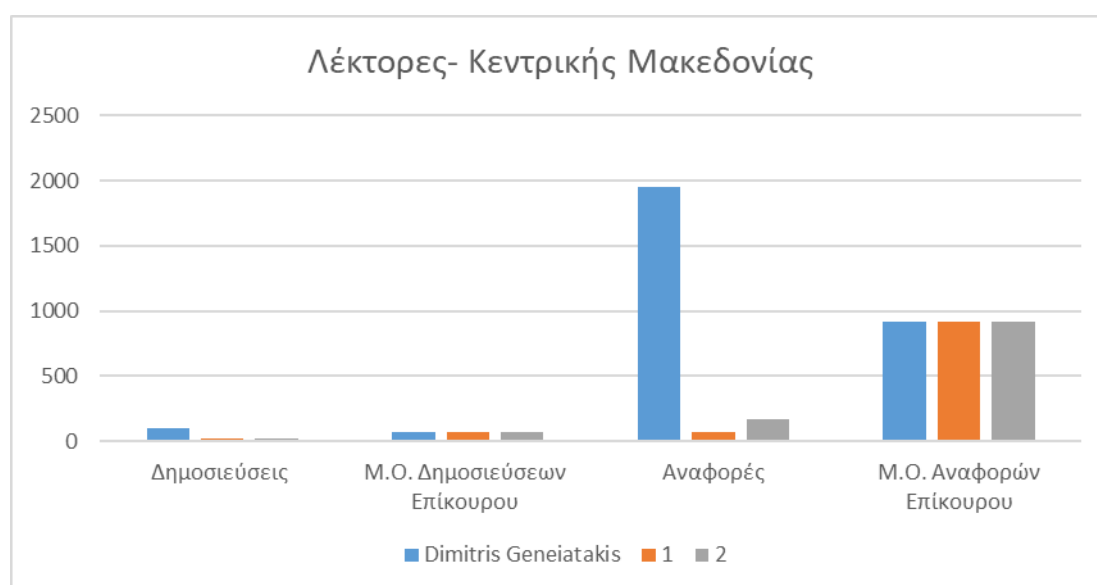


Γράφημα 4.20. Συνολικά στοιχεία των μελών Δ.Ε.Π. ανά βαθμίδα στην Κεντρική Μακεδονία.



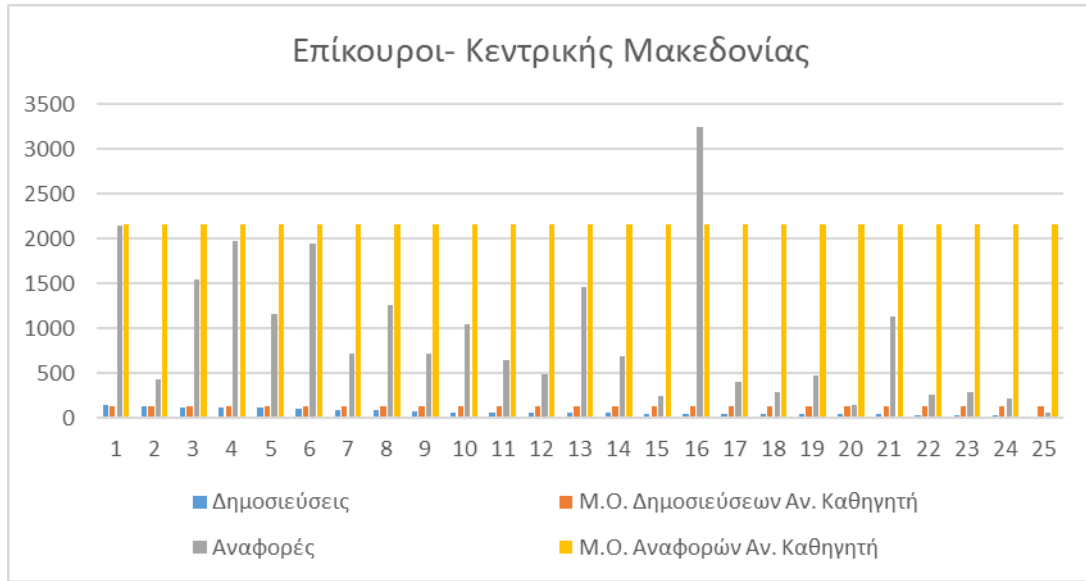
Γράφημα 4.21. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.

Στο Γράφημα 4.21 απεικονίζονται οι δημοσιεύσεις και αναφορές των μελών ΕΔΙΠ των Πανεπιστημίων της Κεντρικής Μακεδονίας. Όπως παρατηρούμε στη 1^η θέση είναι ο Zaharias Zaharis από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ με 163 δημοσιεύσεις, ξεπερνώντας κατά πολύ το μ.ο. δημοσιεύσεων του Επίκουρου Καθηγητή που είναι 65,4. Επίσης, έχει 1473 αναφορές του ερευνητικού του έργου, ενώ ο μ.ο. αναφορών της βαθμίδας του Επίκουρου Καθηγητή είναι 918,1. Συνεπώς, θα μπορούσε να εξελιχθεί στη βαθμίδα του Επίκουρου Καθηγητή. Αξίζει επίσης να σημειωθεί ότι ο Spyros Halkidis από το Τμήμα Εφαρμοσμένης Πληροφορικής του ΠΑΜΑΚ σε 22 δημοσιεύσεις άρθρων του έχει λάβει 1024 αναφορές.



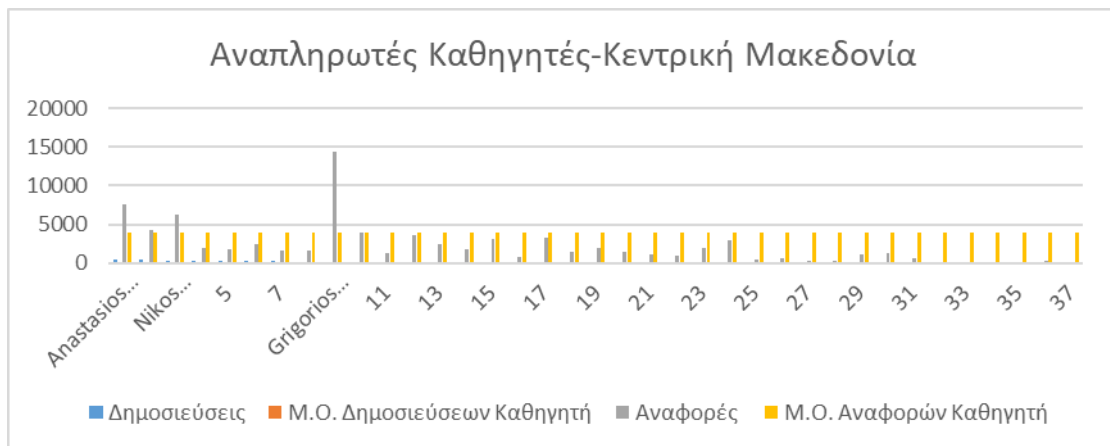
Γράφημα 4.22. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή

Εξετάζοντας τις δημοσιεύσεις και αναφορές των μελών Δ.Ε.Π. που ανήκουν στη βαθμίδα του Λέκτορα (Γράφ. 22), ξεχωρίζει ο Dimitris Geneiatakis από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ, ο οποίος έχει 99 δημοσιεύσεις έναντι 65,4 του μ.ο. δημοσιεύσεων του Επίκουρου Καθηγητή και 1948 αναφορές έναντι 1473 που είναι ο μ.ο. αναφορών του Επίκουρου Καθηγητή. Επομένως, και σε αυτήν την περίπτωση, ο συγκεκριμένος ερευνητής θα μπορούσε να εξελιχθεί από τη βαθμίδα του Λέκτορα σε αυτήν του Επίκουρου Καθηγητή.



Γράφημα 4.23. Στατιστικά στοιχεία για τη βαθμίδα του Επίκουρου Καθηγητή.

Στην περίπτωση των μελών Δ.Ε.Π. που βρίσκονται στη βαθμίδα του Επίκουρου Καθηγητή δεν υπάρχει κάποιο μέλος που να ξεπερνά το μ.ο. των δημοσιεύσεων (2165,4) και των αναφορών (133,1) προκειμένου να εξελιχθεί (Γράφ. 4.23). Ωστόσο, αξίζει να σημειωθεί ότι ο Μίνas Alexiadis (θέση 16) από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ έχει λάβει για 43 δημοσιεύσεις του 3246 αναφορές, έναν πολύ υψηλό αριθμό που υποδηλώνει και την απήχηση του ερευνητικού έργου του.



Γράφημα 4.24. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.

Τέλος, στη βαθμίδα των Αναπληρωτών Καθηγητών ξεχωρίζουν πέντε μέλη Δ.Ε.Π. που είναι ώριμα προς εξέλιξη στη βαθμίδα του Καθηγητή. Αναλυτικά, στη θέση 1 είναι ο Anastasios

Tefas με 453 δημοσιεύσεις και 7571 αναφορές, στη θέση 2 ο Nikos Pleros με 420 δημοσιεύσεις και 4300 αναφορές, στη θέση 3 ο Nikos Nikolaidis με 257 δημοσιεύσεις και 6344 αναφορές, στη θέση 9 ο Grigorios Tsoumakas με 188 δημοσιεύσεις και 14481 αναφορές -όλοι τους μέλη του Τμήματος Πληροφορικής του ΑΠΘ. Στη θέση 11 βρίσκεται ο Konstantinos E. Psannis από το Τμήμα Εφαρμοσμένης Πληροφορικής του ΠΑΜΑΚ με 184 δημοσιεύσεις και 3900 αναφορές, όταν ο μέσος όρος της επόμενης βαθμίδας είναι 165,1 δημοσιεύσεις και 3868,1 αναφορές.

ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ

Στον πίνακα 4.10. παρουσιάζονται ανά βαθμίδα τα 31 μέλη Δ.Ε.Π. των τμημάτων πληροφορικής της Δυτικής Μακεδονίας. Πρόκειται για το Τμήμα Πληροφορικής και το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Δυτικής Μακεδονίας με έδρα την Κοζάνη.

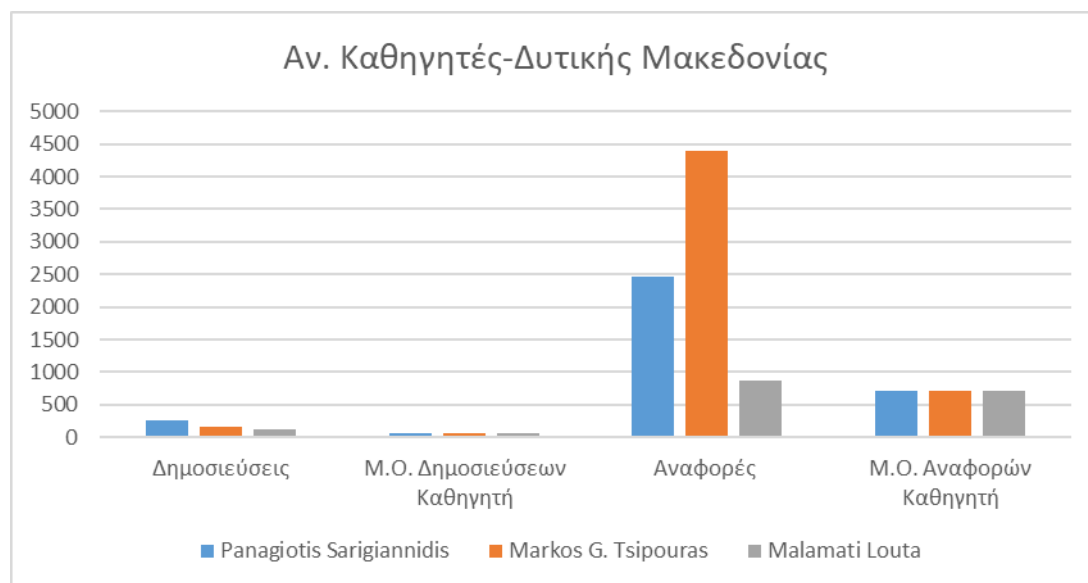
	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΠ	3	4,7	62	-
Λέκτορες	2	39	691,5	1
Επίκουροι Καθηγητές	9	54,3	691,7	-
Αν. Καθηγητές	5	138,4	1765,4	3
Καθηγητές	12	71,1	703,3	-

Πίνακας 4.10. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Δυτική Μακεδονία.



Γράφημα 4.25. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.

Σύμφωνα με το Γράφημα 25, διαπιστώνεται ότι ο Dimitrios Tsalikakis του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ έχει τις προϋποθέσεις να εξελιχθεί από τη βαθμίδα του Λέκτορα σε αυτήν του Επίκουρου Καθηγητή, καθώς έχει 73 δημοσιεύσεις και 1320 αναφορές 1320.



Γράφημα 4.26. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.

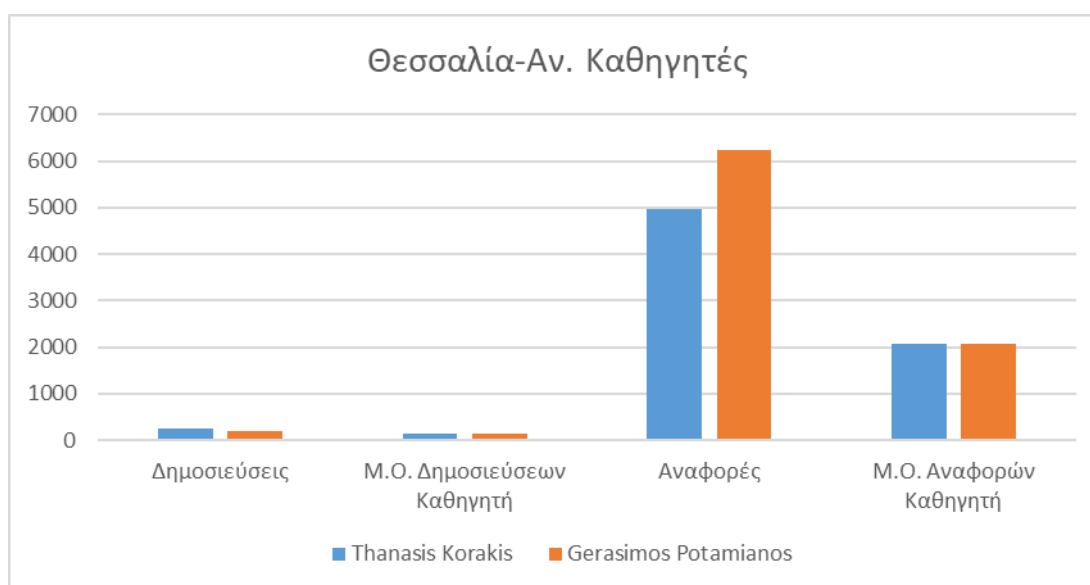
Αντίστοιχα, από τη βαθμίδα του Αναπληρωτή Καθηγητή μπορούν να εξελιχθούν σε αυτήν του Καθηγητή ο Panagiotis Sarigiannidis με 262 δημοσιεύσεις και 2457 αναφορές, υπερβαίνοντας κατά πολύ το μ.ο. δημοσιεύσεων (71,1) και αναφορών (703,3) των τμημάτων. Παρομοίως, ο Markos G. Tsiouras με 161 δημοσιεύσεις και 4386 αναφορές, καθώς και η Malamati Louta με 118 δημοσιεύσεις και 865 αναφορές, θα μπορούσαν να εξελιχθούν στη βαθμίδα του Καθηγητή. Όλοι τους είναι μέλη Δ.Ε.Π. του Τμήματος Ηλεκτρολόγων Μηχανικών του Πανεπιστημίου Δυτικής Μακεδονίας.

ΘΕΣΣΑΛΙΑ

Στο Πανεπιστήμιο Θεσσαλίας ανήκει το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ με έδρα το Βόλο και το Τμήμα Ψηφιακών Συστημάτων με έδρα τη Λάρισα. Ο συνολικός αριθμός μελών Δ.Ε.Π. ανέρχεται σε 39 μέλη.

	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο Αναφορών	Άτομα για εξέλιξη
ΕΔΠΙ	3	42,7	484	-
Λέκτορες	2	13,5	47	-
Επίκουροι Καθηγητές	5	48,4	431	-
Αν. Καθηγητές	17	94,6	1723,5	2
Καθηγητές	12	149,7	2069,7	-

Πίνακας 4.11. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Θεσσαλία.



Γράφημα 4.27. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.

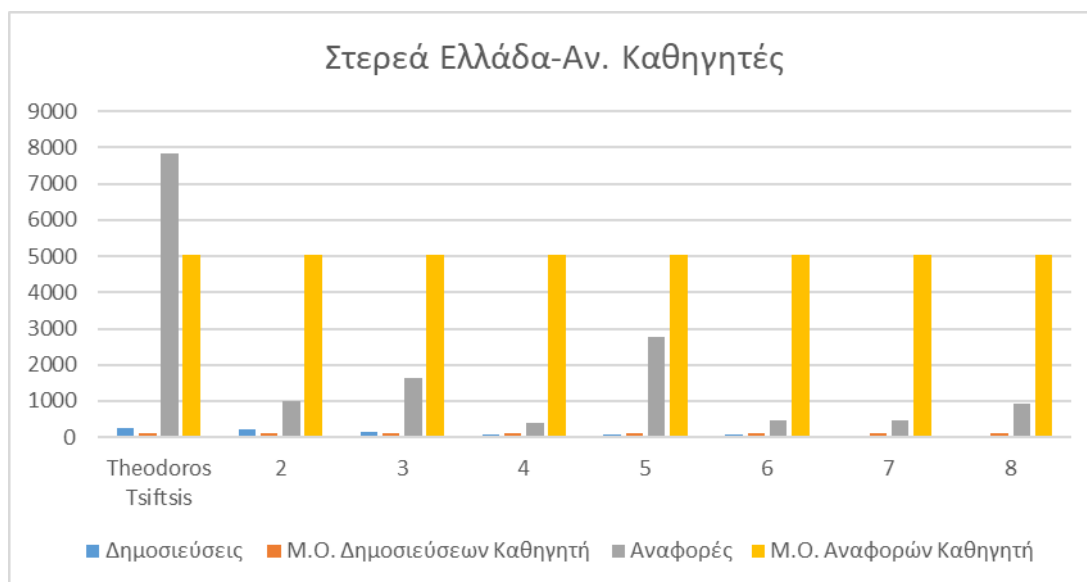
Σύμφωνα με τα στοιχεία που συλλέχθηκαν, για εξέλιξη από τη βαθμίδα του Αναπληρωτή Καθηγητή σε αυτήν του Καθηγητή, είναι ώριμοι οι Thanasis Korakis από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ του Βόλου με 248 δημοσιεύσεις και 4970 αναφορές, τη στιγμή που ο μ.ο. δημοσιεύσεων της επόμενης βαθμίδας είναι 149,7 και ο μ.ο. αναφορών είναι 2069,7. Επίσης, ο Gerasimos Potamianos από το ίδιο τμήμα θα μπορούσε να εξελιχθεί σε Καθηγητής, με 195 δημοσιεύσεις και 6230 αναφορές στο ερευνητικό του έργο.

ΣΤΕΡΕΑ ΕΛΛΑΔΑ

Στο γεωγραφικό διαμέρισμα Στερεάς Ελλάδος υπάγονται το Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική και το Τμήμα Πληροφορικής και Τηλεπικοινωνιών που εδρεύουν στη Λαμία. Ο συνολικός αριθμός μελών Δ.Ε.Π. είναι 39.

	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΠ	4	21,8	254	-
Λέκτορες	-	-	-	-
Επίκουροι Καθηγητές	19	63,5	689,8	-
Αν. Καθηγητές	8	118,4	1937,6	1
Καθηγητές	8	120,3	5036,9	-

Πίνακας 4.12. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στη Στερεά Ελλάδα.



Γράφημα 4.28. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.

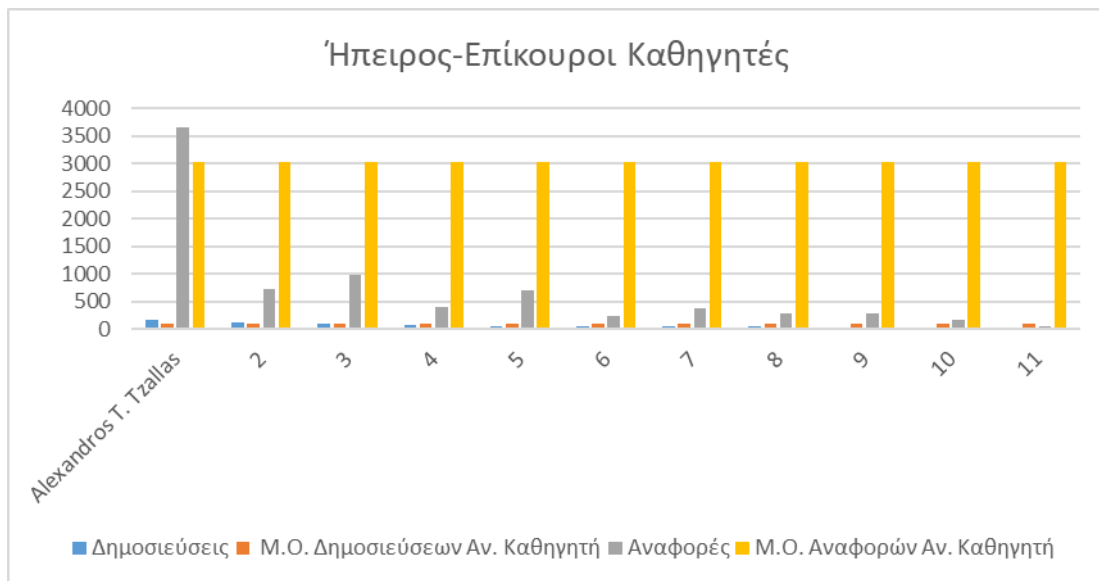
Από τη βαθμίδα του Αναπληρωτή Καθηγητή ξεχωρίζει ο Theodoros Tsiftis του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Θεσσαλίας (με έδρα τη Λαμία), καθώς έχει 267 δημοσιεύσεις και 7848 αναφορές, όταν ο μ.ο. δημοσιεύσεων του Καθηγητή είναι 120,3 και ο αριθμός αναφορών 5036,9.

ΗΠΕΙΡΟΣ

Η γεωγραφική περιφέρεια Ηπείρου περιλαμβάνει το Τμήμα Πληροφορικής και Τηλεπικοινωνιών, καθώς και το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής της Πολυτεχνικής Σχολής του Πανεπιστημίου Ιωαννίνων. Ο συνολικός αριθμός μελών Δ.Ε.Π. είναι 42.

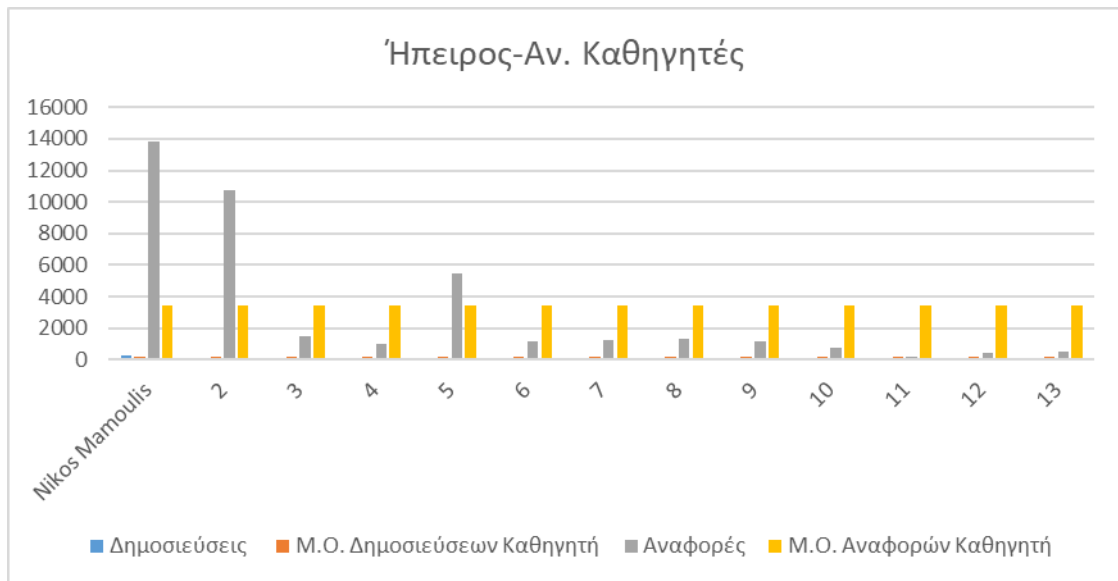
	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΠ	4	25,5	420,8	-
Λέκτορες	-	-	-	-
Επίκουροι Καθηγητές	11	62,2	715,7	1
Αν. Καθηγητές	13	101,6	3027,3	1
Καθηγητές	14	158,7	3448,2	

Πίνακας 4.13. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Ήπειρο.



Γράφημα 4.29. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.

Από τη βαθμίδα του Επίκουρου Καθηγητή είναι ώριμος για εξέλιξη σε αυτήν του Αναπληρωτή Καθηγητή ο Alexandros T. Tzallas από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών, ο οποίος έχει 173 δημοσιεύσεις και 3665 αναφορές. Ο μ.ο. δημοσιεύσεων του Αν. Καθηγητή είναι 101,6 και ο μ.ο. αναφορών 3027,3.



Γράφημα 4.30. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.

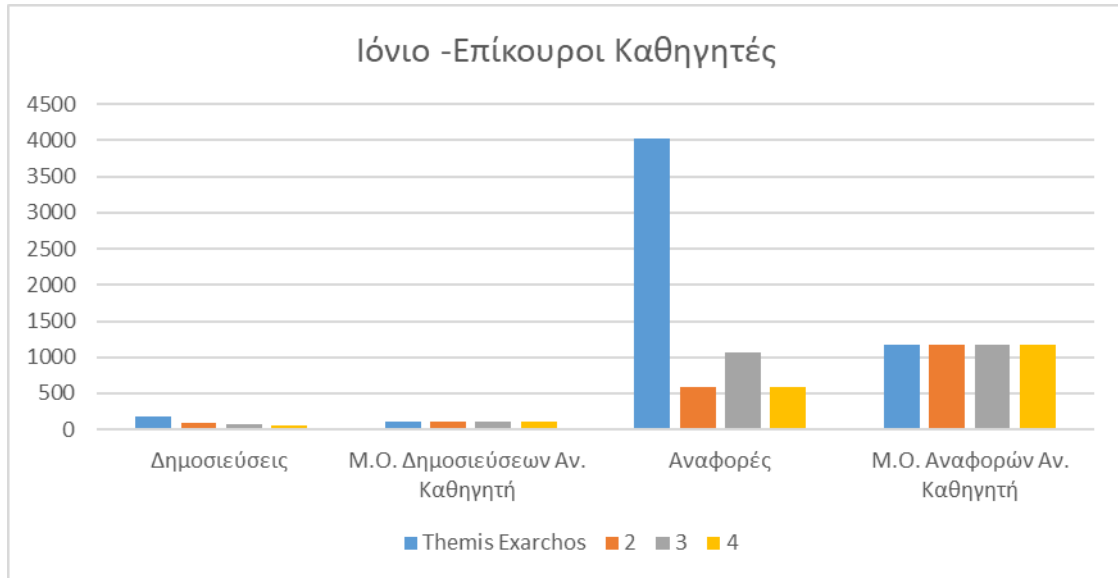
Παρομοίως, ώριμος για εξέλιξη από τη βαθμίδα του Αν. Καθηγητή σε αυτήν του Καθηγητή, είναι ο Nikos Mamoulis από το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής της Πολυτεχνικής Σχολής του Πανεπιστημίου Ιωαννίνων, με 281 δημοσιεύσεις και 13843 αναφορές, όταν ο μ.ο. δημοσιεύσεων του Καθηγητή είναι 158,7 και ο μ.ο. αναφορών 3448,2.

ΙΟΝΙΟ

Στην περιφέρεια Ιονίων νήσων ανήκει το Τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου που εδρεύει στην Κέρκυρα. Ο συνολικός αριθμός μελών Δ.Ε.Π. είναι 19 (Πίν. 4.14.).

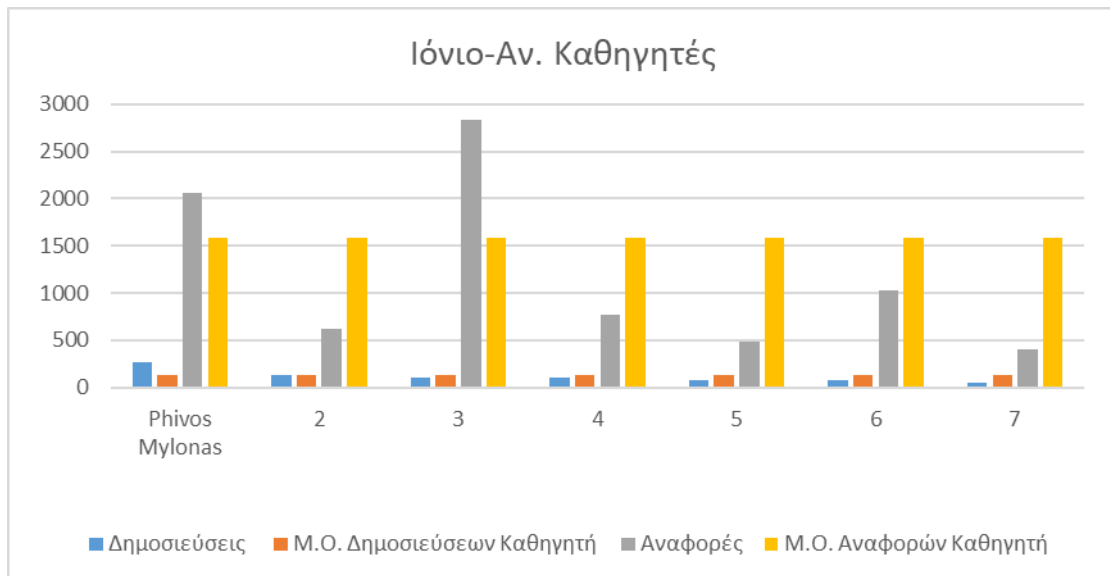
	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΙΠ	4	21,5	458,5	-
Λέκτορες		-		-
Επίκουροι Καθηγητές	4	99,3	1566,3	1
Αν. Καθηγητές	7	112,9	1170,1	1
Καθηγητές	4	128	1583,8	-

Πίνακας 4.14. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στο Ιόνιο.



Γράφημα 4.31. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.

Από το τμήμα ξεχωρίζει ο Themis Exarchos με 187 δημοσιεύσεις και 4031 αναφορές, και μπορεί να εξελιχθεί από Επίκουρος ΚαθηγητήςΚαθηγητής σε Αν. Καθηγητής, καθώς ο μ.ο. των δημοσιεύσεων της επόμενης βαθμίδας είναι 112,9 και ο μ.ο. αναφορών 1170,1.



Γράφημα 4.32. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.

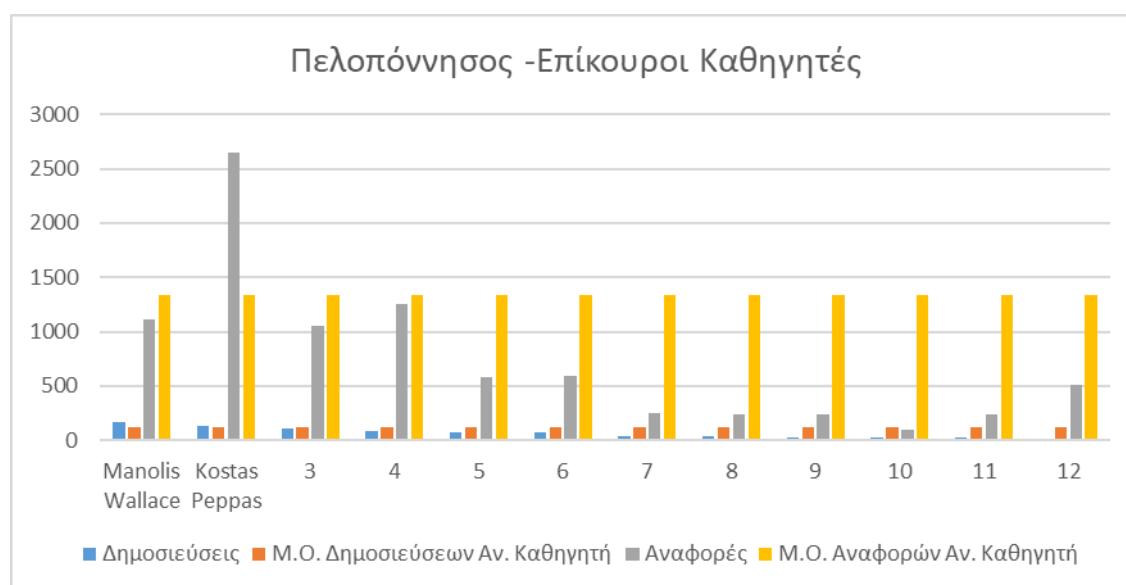
Επιπλέον, από τη βαθμίδα του Αναπληρωτή ο Ρhivos Mylonas είναι ώριμος για εξέλιξη στη βαθμίδα του Καθηγητή, με 266 δημοσιεύσεις και 2064 αναφορές. Ο μ.ο. των δημοσιεύσεων της ανώτερης βαθμίδας είναι 128 και ο μ.ο. αναφορών 1583,8.

ΠΕΛΟΠΟΝΝΗΣΟΣ

Το πανεπιστήμιο Πελοποννήσου περιλαμβάνει το Τμήμα Πληροφορικής και Τηλεπικοινωνιών της Σχολής Οικονομίας και Τεχνολογίας που εδρεύει στην Τρίπολη και το Τμήμα Ψηφιακών Συστημάτων της Σχολής Οικονομίας και Τεχνολογίας που εδρεύει στη Σπάρτη. Ο συνολικός αριθμός μελών Δ.Ε.Π. είναι 35 (Πίν. 4.15).

	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΙΠ	7	26,6	272	-
Λέκτορες	-	-	-	-
Επίκουροι Καθηγητές	12	66,8	733,3	2
Αν. Καθηγητές	7	120,4	13423,7	-
Καθηγητές	9	159,1	2771,9	-

Πίνακας 4.15. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Πελοπόννησο.



Γράφημα 4.33. Άτομα για εξέλιξη στη βαθμίδα του Αν. Καθηγητή.

Από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών της Σχολής Οικονομίας και Τεχνολογίας που εδρεύει στη Σπάρτη, οι Manolis Wallace με 168 δημοσιεύσεις και 1113 αναφορές και

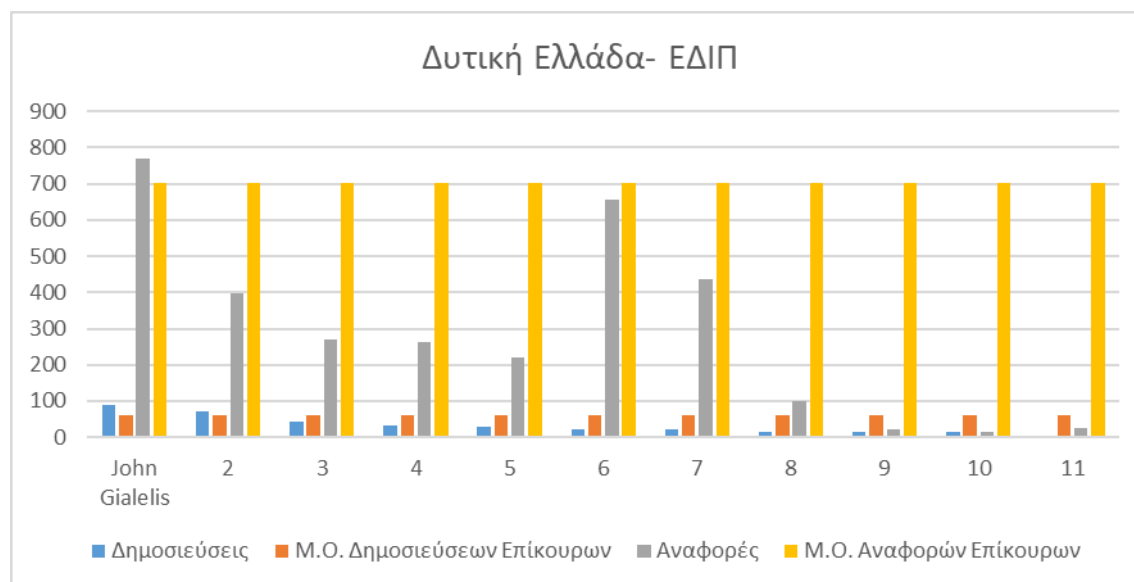
Kostas Peppas με 135 δημοσιεύσεις και 2646 αναφορές είναι ώριμοι για εξέλιξη στη βαθμίδα του Αν. Καθηγητή, τη στιγμή που ο μ.ο. αναφορών της επόμενης βαθμίδας είναι 120,4 και ο μ.ο. δημοσιεύσεων είναι 1342,7.

ΔΥΤΙΚΗ ΕΛΛΑΔΑ

Στην περιφέρεια Δυτικής Ελλάδος ανήκουν το Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών-Πανεπιστήμιο Πατρών και το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Πανεπιστήμιο Πελοποννήσου. Ο συνολικός αριθμός μελών ΔΕΠ είναι 101.

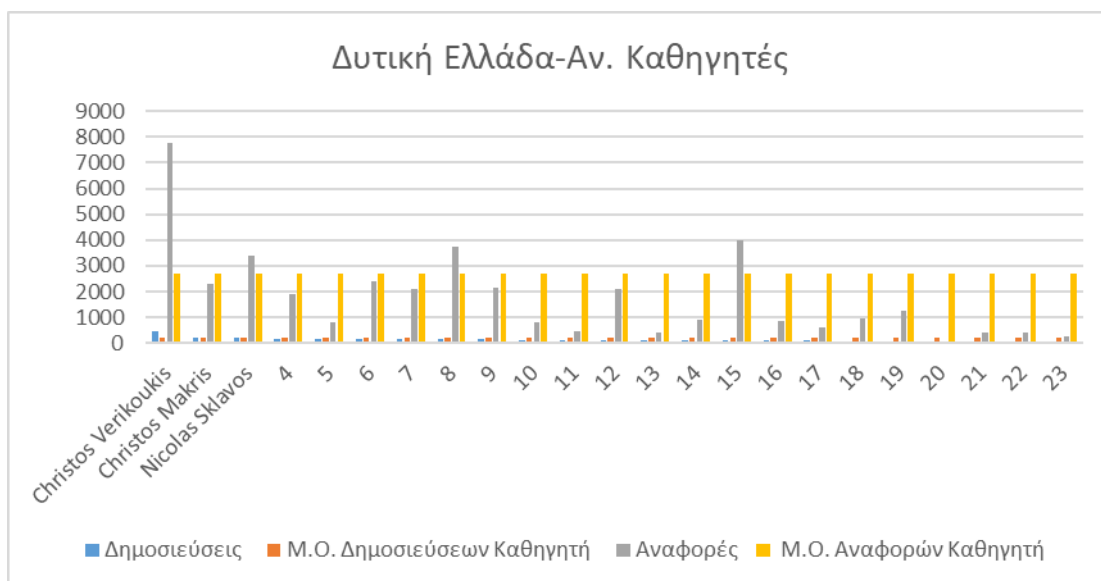
	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΙΠ	11	32,5	288,9	1
Λέκτορες		-	-	-
Επίκουροι Καθηγητές	17	60,4	702,9	-
Αν. Καθηγητές	23	133,1	1742,8	2
Καθηγητές	50	191,3	2680,7	-

Πίνακας 4.16. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στη Δυτική Ελλάδα.



Γράφημα 4.34. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή καθηγητή.

Ο John Gialelis από το Τμήμα Μηχανικών Η/Υ & Πληροφορικής θα μπορούσε να εξελιχθεί στη βαθμίδα του Επίκουρου Καθηγητή, καθώς έχει 88 δημοσιεύσεις και 769 αναφορές. Ο μ.ο. δημοσιεύσεων της βαθμίδας του Επίκουρου Καθηγητή είναι 60,4 και ο μ.ο. αναφορών 702,9.



Γράφημα 4.35. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.

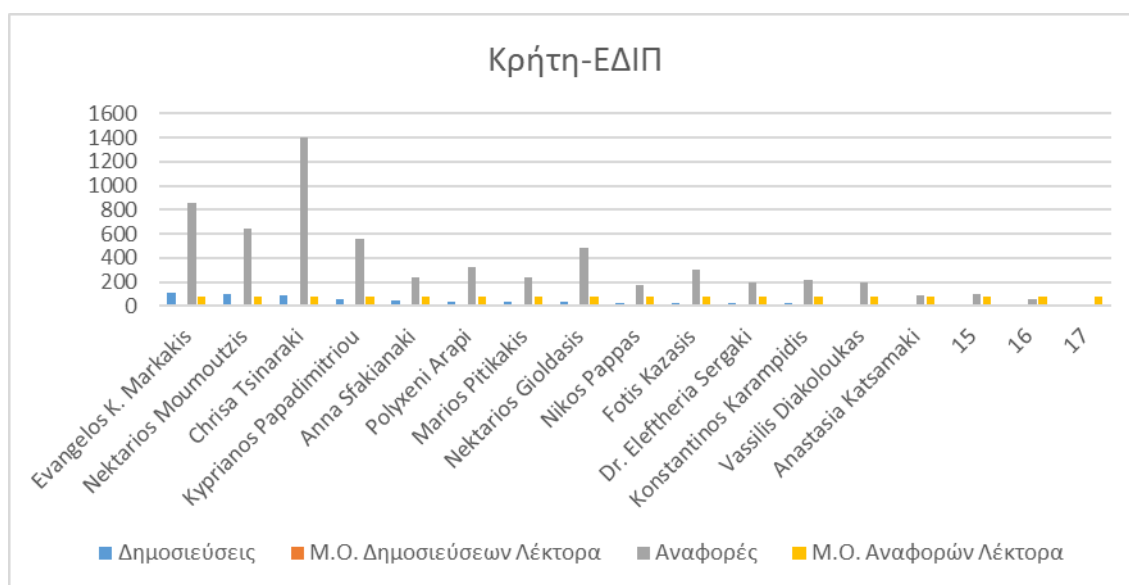
Από τη βαθμίδα του Επίκουρου Καθηγητή οι Christos Verikoukis και Nicolas Sklavos του Τμήματος Μηχανικών Η/Υ & Πληροφορικής είναι ώριμοι για εξέλιξη στην βαθμίδα του Καθηγητή. Οι δημοσιεύσεις τους είναι 480 και 208 και οι αναφορές τους 7773 και 3376, αντίστοιχα. Σημειώνεται ότι ο μ.ο. των δημοσιεύσεων του Καθηγητή είναι 1921,3 και ο μ.ο. των αναφορών είναι 1680,7.

ΚΡΗΤΗ

Η περιφέρεια της Κρήτης περιλαμβάνει το Τμήμα Επιστήμης Υπολογιστών (Πανεπιστήμιο Κρήτης) και το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών (Ελληνικό Μεσογειακό Πανεπιστήμιο) που εδρεύουν στο Ηράκλειο, τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών (Πολυτεχνείο Κρήτης) και το Τμήμα Ηλεκτρονικών Μηχανικών (Ελληνικό Μεσογειακό Πανεπιστήμιο) που εδρεύουν στα Χανιά. Ο συνολικός αριθμός μελών Δ.Ε.Π. είναι 114.

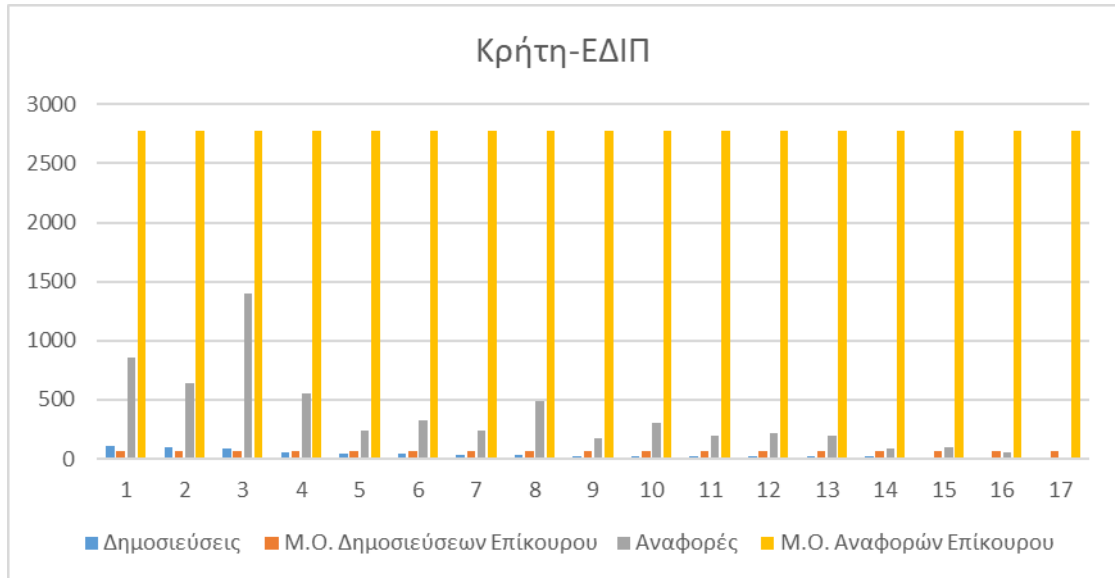
	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΙΠ	17	38,5	356,5	14
Λέκτορες	2	18	81	-
Επίκουροι Καθηγητές	14	65,1	2773,2	2
Αν. Καθηγητές	32	90,5	1713,8	1
Καθηγητές	49	179,9	4140,3	-

Πίνακας 4.17. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Κρήτη.

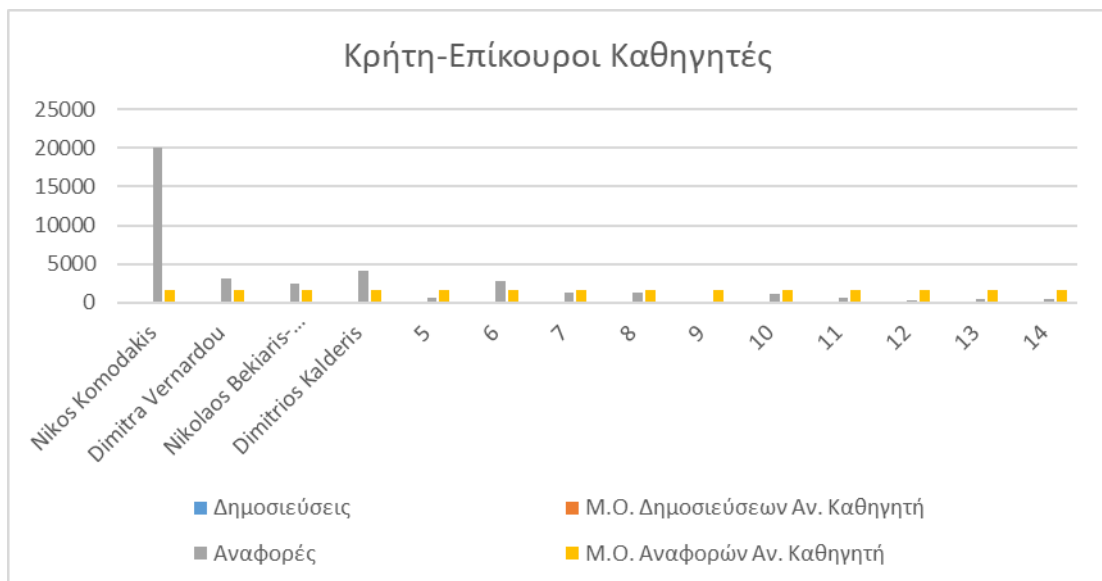


Γράφημα 4.36. Άτομα για εξέλιξη στη βαθμίδα του Λέκτορα.

Σύμφωνα με τα στοιχεία που συλλέχθηκαν, προκύπτει ότι τα 14 πρώτα μέλη ΕΔΙΠ από τα τμήματα της Κρήτης έχουν δημοσιεύσεις που υπερβαίνουν το μ.ο. του Λέκτορα (> 18) και το μ.ο. των αναφορών (> 81). Επομένως θα μπορούσαν να εξελιχθούν στην επόμενη βαθμίδα.



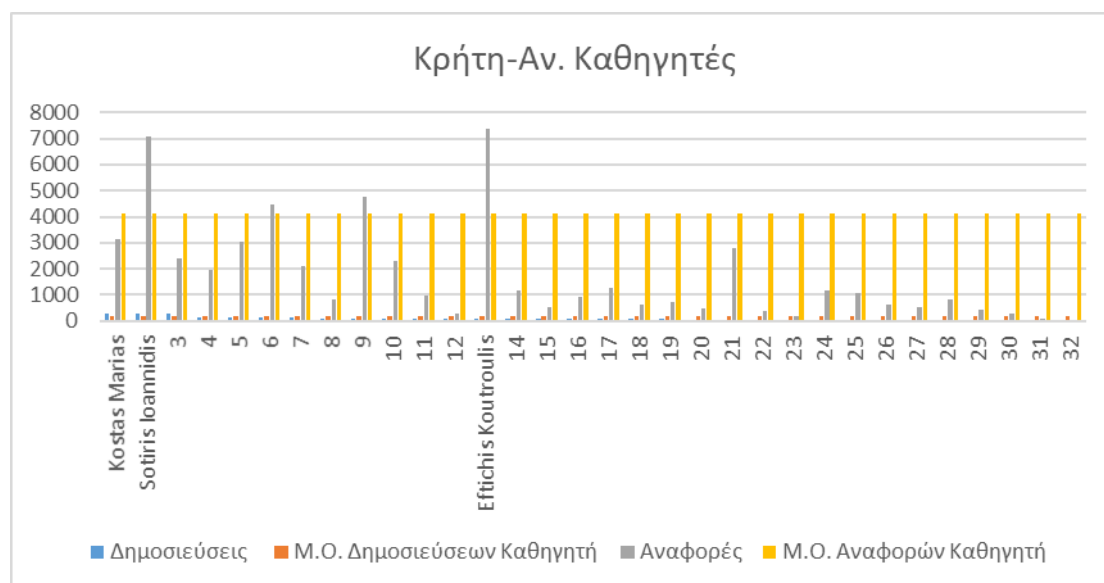
Παρατηρώντας το παραπάνω γράφημα διαπιστώνουμε ότι ενώ οι τρεις πρώτοι στην κατάταξη έχουν δημοσιεύσεις πάνω από το μ.ο. της βαθμίδας του Επίκουρου Καθηγητή που είναι 65,1 με ποσοστό 89, 103,105, ωστόσο οι αναφορές που έχουν λάβει τα άρθρα τους είναι χαμηλότερες από το μ.ο. των αναφορών του Επίκουρου Καθηγητή που είναι 2773,2.



Γράφημα 4.37. Ατομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.

Από το παραπάνω γράφημα παρατηρούμε ότι ο Νίκος Κομοδάκης από το Τμήμα Επιστήμης Υπολογιστών με 152 δημοσιεύσεις και 2025 αναφορές και η Dimitra Vernardou από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών με 105 δημοσιεύσεις και 3136 αναφορές, είναι ώριμοι για εξέλιξη στην επόμενη βαθμίδα. Ο μ.ο. δημοσιεύσεων της

επόμενης βαθμίδας (Αναπληρωτή Καθηγητή) είναι 90,5 και ο μ.ο. αναφορών είναι 1713,8. Ωστόσο, αξίζει να σημειωθεί ότι ο Nikolaos Bekiaris – Liberis και ο Dimitrios Kalderis έχουν υψηλό ποσοστό αναφορών.



Γράφημα 4.38. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.

Όσον αφορά τη βαθμίδα των Αναπληρωτών Καθηγητών (Γραφ. 38), παρατηρούμε ότι ο Sotiris Ioannidis από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης έχει 279 δημοσιεύσεις και 7080 αναφορές και είναι ώριμος για εξέλιξη στη βαθμίδα του Καθηγητή. Ο μ.ο. δημοσιεύσεων της επόμενης βαθμίδας είναι 179,9 και ο μ.ο. αναφορών είναι 4140,3. Ωστόσο, ο Kostas Marias από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών (Ελληνικό Μεσογειακό Πανεπιστήμιο) έχει 308 δημοσιεύσεις, αλλά μόνο 3150 αναφορές, ενώ αντίστροφα ο Eftichis Koutroulis από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης έχει 7395 αναφορές σε 90 δημοσιεύσεις του.

ΒΟΡΕΙΟ ΑΙΓΑΙΟ

Στο Βόρειο Αιγαίο λειτουργεί το Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων του Πανεπιστημίου Αιγαίου που εδρεύει στη Σάμο. Ο συνολικός αριθμός μελών Δ.Ε.Π. είναι 26.

	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΠ	-	-	-	-
Λέκτορες	-	-	-	-
Επίκουροι Καθηγητές	7	48,1	564,6	-
Αν. Καθηγητές	9	75,7	1579,3	-
Καθηγητές	10	160,1	3710,4	-

Πίνακας 4.18. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στο Βόρειο Αιγαίο.

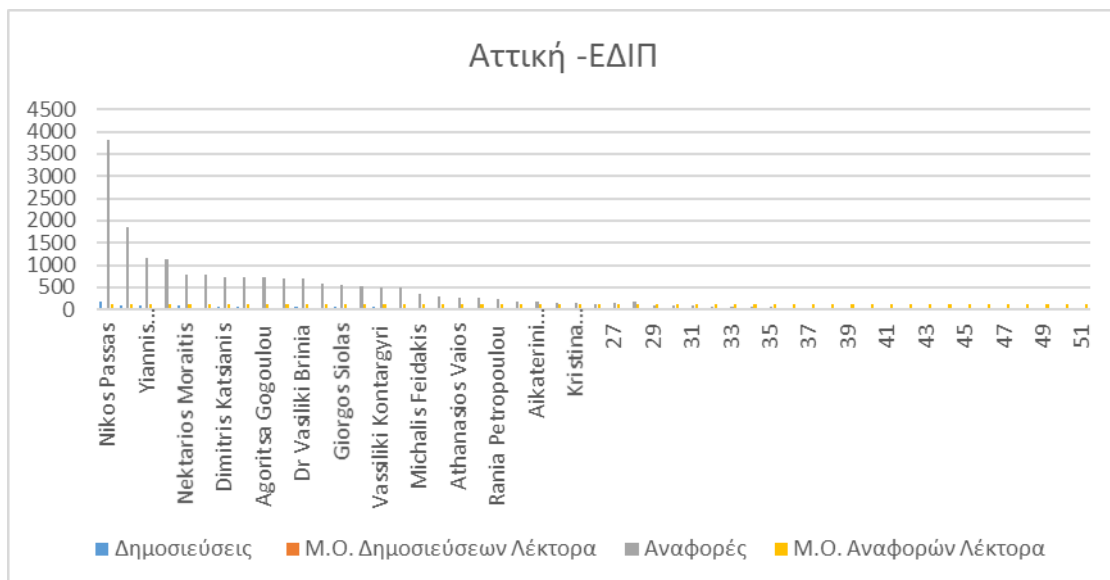
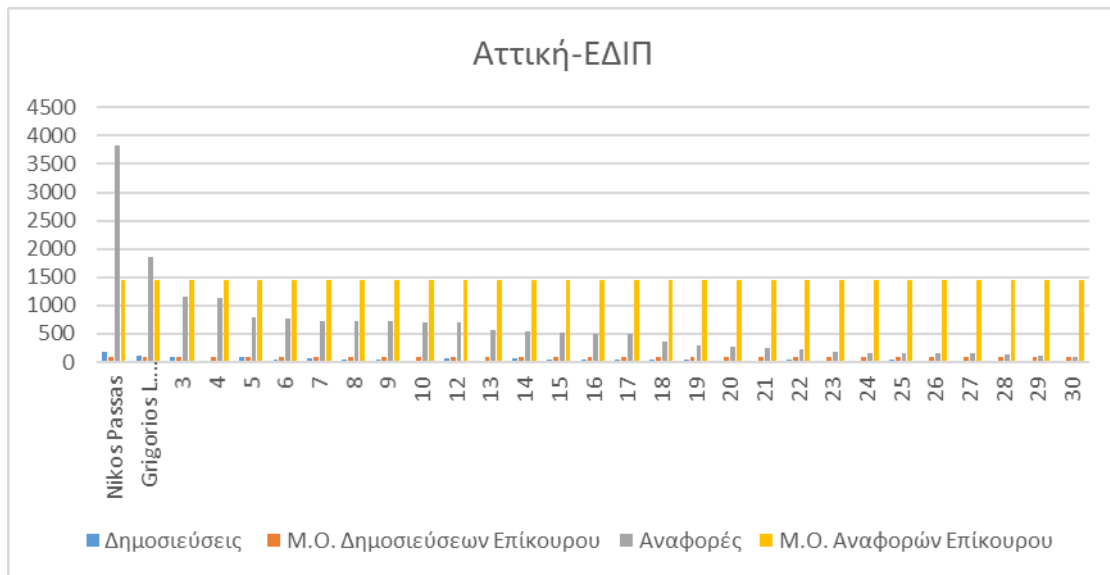
Σύμφωνα με τα στοιχεία του πίνακα 4.18, διαπιστώνεται ότι στο Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων δεν υπάρχουν καθόλου μέλη ΕΔΠ, καθώς επίσης και καθηγητές στη βαθμίδα του Λέκτορα. Επιπλέον, δεν υπάρχουν μέλη Δ.Ε.Π. από τις υπόλοιπες βαθμίδες, των οποίων η παραγωγικότητα και απήχηση του επιστημονικού έργου τους (δημοσιεύσεις και αναφορές) να δικαιολογεί την εξέλιξή τους σε επόμενη βαθμίδα.

ΑΤΤΙΚΗ

Στην περιφέρεια Αττικής λειτουργούν τα περισσότερα τμήματα πληροφορικής στην Ελλάδα. Πρόκειται για το Τμήμα Πληροφορικής του Οικονομικού Πανεπιστημίου Αθηνών, το Τμήμα Πληροφορικής του Πανεπιστημίου Πειραιώς, το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ, το Τμήμα Πληροφορικής και Τηλεματικής του Χαροκόπειου Πανεπιστημίου, η Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου, το Τμήμα Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, και το Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής. Ο συνολικός αριθμός μελών Δ.Ε.Π. είναι 266.

	Μέλη	Μ.Ο. Δημοσιεύσεων	Μ.Ο. Αναφορών	Άτομα για εξέλιξη
ΕΔΠ	51	31,8	373,7	26
Λέκτορες	12	17,9	112,4	-
Επίκουροι Καθηγητές	42	81,3	1459,7	3
Αν. Καθηγητές	33	129,8	2429	3
Καθηγητές	128	189,6	3635,2	-

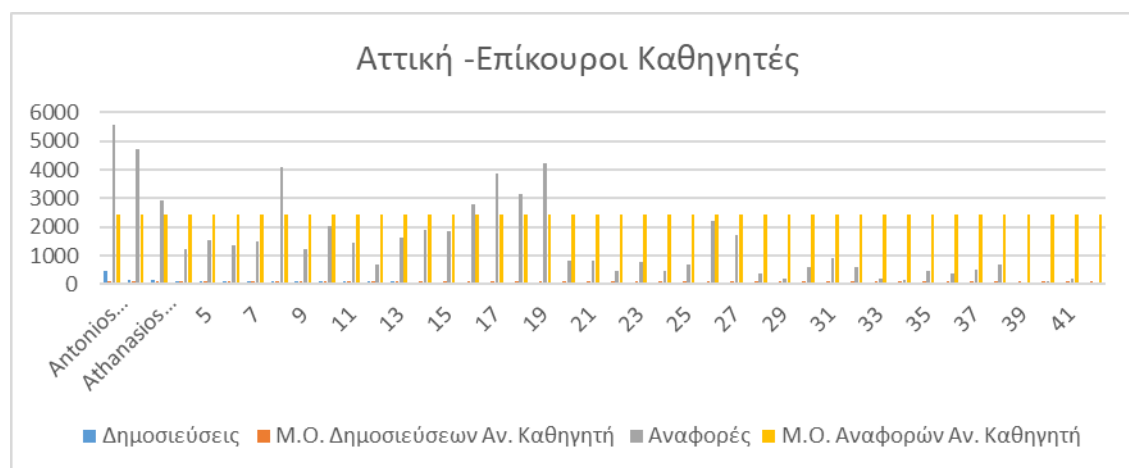
Πίνακας 4.19. Στοιχεία για το μ.ο. δημοσιεύσεων και αναφορών των μελών Δ.Ε.Π. ανά βαθμίδα στην Αττική.



Γράφημα 4.39- 4.40. Άτομα για εξέλιξη στη βαθμίδα του Επίκουρου Καθηγητή.

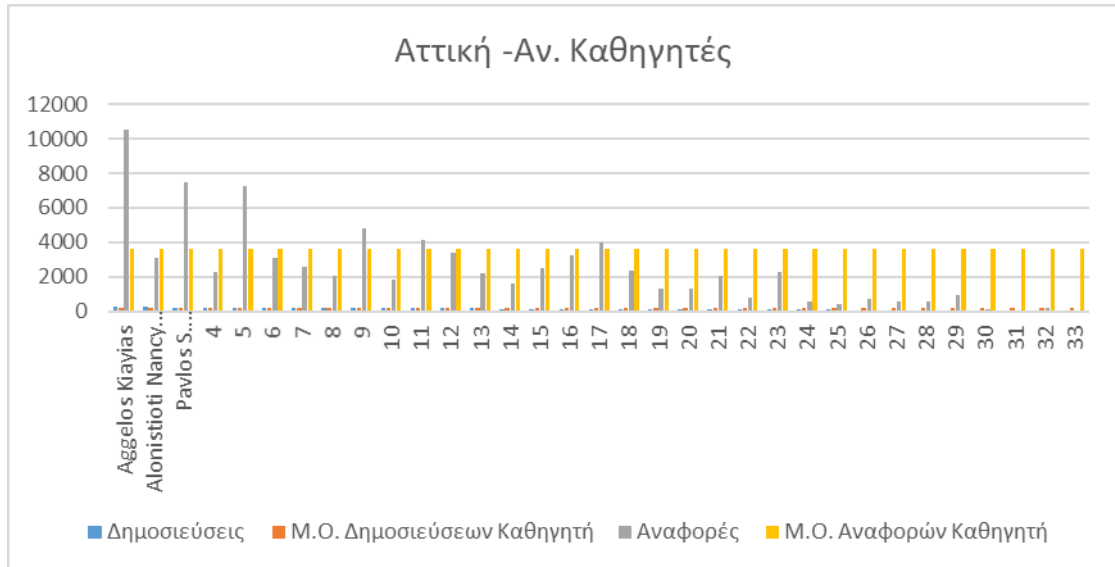
Εξετάζοντας αρχικά τη βαθμίδα ΕΔΙΠ (Γραφ. 4.39), παρατηρούμε ότι τα δύο πρώτα μέλη, ο Nikos Passas από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών ΕΚΠΑ έχει 190 δημοσιεύσεις και 3819 αναφορές και ο Grigorios L. Kyriakoroulos από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου έχει 104 δημοσιεύσεις και 1864 αναφορές. Και οι δύο θα μπορούσαν να εξελιχθούν σε Επίκουροι Καθηγητές, καθώς ο μ.ο. των δημοσιεύσεων του Επίκουρου

Καθηγητή είναι 81,3 και ο μ.ο. των αναφορών είναι 1459,7. Επιπλέον, από το γράφημα 4.40. προκύπτει ότι 24 ακόμη μέλη ΕΔΙΠ από τα τμήματα της Αττικής έχουν δημοσιεύσεις που υπερβαίνουν το μ.ο. του Λέκτορα (> 17,9) και το μ.ο. των αναφορών (> 112,4), επομένως θα μπορούσαν να εξελιχθούν στην επόμενη βαθμίδα του Λέκτορα, εφόσον αυτή συνέχιζε να υφίσταται. Ωστόσο, δεν διαθέτουν τον απαιτούμενο αριθμό δημοσιεύσεων και αναφορών για τη θέση του Επίκουρου Καθηγητή.



Γράφημα 4.41. Άτομα για εξέλιξη στη βαθμίδα του Αναπληρωτή Καθηγητή.

Από τη βαθμίδα του Επίκουρου Καθηγητή ξεχωρίζουν: Ο Antonios Antonopoulos από Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ με 486 δημοσιεύσεις και 5554 αναφορές, Ο George C. Alexandropoulos από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ με 170 δημοσιεύσεις και 4712 αναφορές και ο Athanasios Voulodimos από το Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής με 154 δημοσιεύσεις και 2947 αναφορές. Και οι τρεις θα μπορούσαν να εξελιχθούν στη βαθμίδα του Αναπληρωτή Καθηγητή, καθώς ο μ.ο. δημοσιεύσεων αυτής της βαθμίδας είναι 129,8 και ο μ.ο. αναφορών είναι 2429. Επιπλέον, αξίζει να σημειωθεί ότι και η Maria Roussou από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ για 114 δημοσιεύσεις της έχει 4089 αναφορές.



Γράφημα 4.42. Άτομα για εξέλιξη στη βαθμίδα του Καθηγητή.

Τέλος, τρεις Αναπληρωτές Καθηγητές εμφανίζουν μεγαλύτερο αριθμό δημοσιεύσεων και αναφορών από το μ.ο. της βαθμίδας του Καθηγητή (μ.ο. δημοσιεύσεων: 189,6, μ.ο. αναφορών: 3635,2) και θα μπορούσαν να εξελιχθούν σε Καθηγητές. Πρόκειται για τον Aggelos Kiaias από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ με 289 δημοσιεύσεις και 10524 αναφορές, την Alonistioti Nancy (Athanasia) από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ με 276 δημοσιεύσεις και 3115 αναφορές και τον Pavlos S. Georgilakis από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ με 222 δημοσιεύσεις και 7511 αναφορές.

Κεφάλαιο 5

Εξόρυξη γνώσης από τα δεδομένα επιστημομετρίας

5.1 Το λογισμικό Weka

Για την εφαρμογή της εξόρυξης γνώσης απαιτείται η χρήση του κατάλληλου λογισμικού που θα προσφέρει τις μεθόδους και τα εργαλεία για τους σκοπούς της εκάστοτε έρευνας. Στην παρούσα εργασία αποφασίστηκε η χρήση του λογισμικού WEKA (Waikato Environment for Knowledge Analysis). Πρόκειται για ένα λογισμικό μηχανικής μάθησης γραμμένο σε Java, που αναπτύχθηκε στο Πανεπιστήμιο του Waikato, στη Νέα Ζηλανδία. Το λογισμικό χρησιμοποιείται για την άντληση δεδομένων, την εξόρυξη γνώσης και την ανάπτυξη μοντέλων, ενώ είναι ελεύθερα προσβάσιμο και ιδιαίτερα εύχρηστο [28].

Το Weka παρέχει πολλά εργαλεία και τεχνικές για την ανάλυση δεδομένων και την εξόρυξη γνώσης. Έτσι, εκτός από τα εργαλεία οπτικοποίησης και τους αλγόριθμους, παρέχει στο χρήστη τη δυνατότητα προεπεξεργασίας των δεδομένων. Επιπλέον, οι τεχνικές εξόρυξης γνώσης που μπορούν να εκτελεστούν περιλαμβάνουν τόσο την κατηγοριοποίηση, όσο και τη συσταδοποίηση, τεχνικές που χρησιμοποιήθηκαν για τους σκοπούς της παρούσας εργασίας.

Για την εκτέλεση του Weka στην αρχική σελίδα του προγράμματος δίνεται η δυνατότητα στο χρήστη να επιλέξει τα παρακάτω περιβάλλοντα εργασίας:

Explorer (εξερευνητής): Είναι το βασικότερο και πιο εύχρηστο γραφικό περιβάλλον για το χρήστη και παρέχει τη δυνατότητα πρόσβασης σε όλες τις κύριες τεχνικές της εξόρυξης γνώσης. Έτσι, διαθέτει εργαλεία και αλγόριθμους για την προεπεξεργασία και το μετασχηματισμό των δεδομένων, αλγόριθμους για την εκτέλεση της κατηγοριοποίησης και της συσταδοποίησης, για την εύρεση κανόνων συσχετίσεων και για την επιλογή των πιο σχετικών χαρακτηριστικών μέσα από το σύνολο δεδομένων. Επιπλέον, προσφέρει τη δυνατότητα για την οπτικοποίηση των μοντέλων με γραφήματα και άλλα μέσα. Πιο συγκεκριμένα, στον explorer εντοπίζονται οι ακόλουθες κατηγορίες λειτουργιών:

- **Preprocess:** Παρέχει εργαλεία και αλγόριθμους για την προεπεξεργασία και το μετασχηματισμό του συνόλου δεδομένων στην επιθυμητή μορφή.
- **Classify:** Περιέχει αλγόριθμους κατηγοριοποίησης και παλινδρόμησης για την εφαρμογή τους στα προεπεξεργασμένα δεδομένα μας.

- Cluster: Περιέχει αλγόριθμους συσταδοποίησης για την εφαρμογή τους στα προεπεξεργασμένα δεδομένα μας.
- Associate: Περιέχει αλγόριθμους κατάλληλους για την εύρεση κανόνων συσχέτισης για την εξαγωγή γνώσης απ' τα δεδομένα μας.
- Select Attributes: Περιέχει αλγόριθμους κατάλληλους για την επιλογή των πιο σχετικών χαρακτηριστικών, αποκλείοντας γνωρίσματα από το σύνολο δεδομένων.
- Visualize: Παρέχει εργαλεία για τη οπτικοποίηση των δεδομένων ή των μοντέλων μέσω γραφημάτων, διαγραμμάτων, κτλ.

Experimenter (Πειραματιστής): Το περιβάλλον του Experimenter δίνει τη δυνατότητα στο χρήστη να προχωρήσει σε πειράματα και στατιστικές δοκιμές, εφαρμόζοντας διάφορα μαθησιακά σχήματα. Με αυτόν τον τρόπο αξιολογούνται οι διάφορες τεχνικές εξόρυξης γνώσης και συγκρίνονται οι επιδόσεις διαφορετικών μοντέλων.

Knowledge Flow (Ροή γνώσης): Πρόκειται για ένα περιβάλλον που μοιάζει πολύ με τον Explorer, καθώς ο χρήστης μπορεί να εκτελέσει και εδώ τις ίδιες εργασίες. Ωστόσο, ενδείκνυται για χρήστες που ενδιαφέρονται να παρακολουθήσουν λεπτομερώς όλη τη διαδικασία ανάλυσης και επεξεργασίας των δεδομένων, μέσω της οπτικής σχεδίασης της Ανακάλυψης Γνώσης.

Workbench (Θέση εργασίας): Το περιβάλλον αυτό συνδυάζει τις δυνατότητες και τα εργαλεία που προσφέρουν τα τρία προηγούμενα περιβάλλοντα του Weka, με την επιλογή καθορισμού των επιθυμητών κάθε φορά εφαρμογών στις ρυθμίσεις.

Command Line Interface (Διεπαφή Γραμμής Εντολών): Το τελευταίο περιβάλλον εργασίας που εμφανίζεται στο Weka είναι το Command Line Interface, που απευθύνεται σε πιο προχωρημένους χρήστες. Το περιβάλλον αυτό δε διαθέτει γραφικά βοηθήματα και δίνει τη δυνατότητα εισαγωγής και εκτέλεσης εντολών απευθείας από το λειτουργικό σύστημα.

Η μορφή των αρχείων εισόδου για την ανάλυση δεδομένων που δέχεται το Weka είναι τα αρχεία με την κατάληξη ARFF (Attribute Relation File Format). Πρόκειται για αρχεία κειμένου χαρακτήρων ASCII (ASCII text file), όπου οι τιμές διαχωρίζονται με κόμμα (CSV - Comma Separated Values). Τα αρχεία ARFF περιέχουν μια σειρά από στιγμιότυπα (instances) που περιγράφονται από κάποια γνωρίσματα (attributes). Οι τύποι δεδομένων γνωρισμάτων που υποστηρίζει το Weka περιλαμβάνουν:

- Αριθμητικά γνωρίσματα (numeric attributes): Τα αριθμητικά γνωρίσματα αφορούν σε μετρήσιμες ποσότητες και μπορεί να είναι είτε πραγματικοί είτε ακέραιοι αριθμοί.

- Ονομαστικά γνωρίσματα (nominal attributes): Τα ονομαστικά γνωρίσματα δέχονται τιμές με τη μορφή συμβόλων ή ονομάτων αντικειμένων. Οι δυνατές τιμές δηλώνονται μέσα σε αγκύλες.
- Αλφαριθμητικά γνωρίσματα (string attributes): Τα αλφαριθμητικά γνωρίσματα επιτρέπουν τη δημιουργία αυθαίρετων αλφαριθμητικών δομών, κάτι το οποίο αφορά κυρίως σε εφαρμογές εξόρυξης κειμένου (text-mining).
- Ημερομηνίες (date attributes): Η προεπιλεγμένη μορφή δέχεται τη συνδυασμένη μορφή ημερομηνίας και ώρας.

5.2 Εφαρμογή εξόρυξης γνώσης από τα δεδομένα επιστημομετρίας στο Weka

Η έκδοση του Weka που χρησιμοποιήθηκε για την εξόρυξη γνώσης από τα δεδομένα επιστημομετρίας είναι η 3.8.5. Όπως προαναφέρθηκε, η προεπιλεγμένη μορφή αρχείων εισόδου που χρησιμοποιείται στο λογισμικό Weka είναι τα αρχεία ARFF. Στην παρούσα εργασία προχωρήσαμε στη δημιουργία δικών μας αρχείων ARFF από ένα αρχείο EXCEL. Το αρχείο EXCEL με τα δεδομένα μας μετατράπηκε σε αρχείο CSV και στη συνέχεια προχωρήσαμε στις απαραίτητες διορθώσεις, όπου αντικαταστάθηκαν τα ερωτηματικά με κόμματα για να γίνει εφικτή η αναγνώριση των γνωρισμάτων από το Weka. Τέλος, το αρχείο CSV εισήχθη στο Weka και μετατράπηκε σε αρχείο ARFF.

Ακολούθησε η προεπεξεργασία του συνόλου των δεδομένων στο αρχείο ARFF, ώστε τα δεδομένα που θα χρησιμοποιηθούν να είναι αξιόπιστα. Έτσι για την εφαρμογή κάποιων από τις τεχνικές της εξόρυξης γνώσης πραγματοποιήθηκε περαιτέρω διακριτοποίηση των αριθμητικών γνωρισμάτων (numeric attributes) σε ονομαστικά γνωρίσματα (nominal attributes), με τη χρήση της λειτουργίας preprocess που προσφέρει ο Explorer. Ταυτόχρονα, πραγματοποιήθηκε η διαχείριση ελλিপών τιμών σε όλο το σύνολο των δεδομένων μας, οι οποίες συμπληρώθηκαν με το αγγλικό ερωτηματικό (?). Οι ελλειπείς αυτές τιμές εντοπίστηκαν σε πέντε περιπτώσεις μελών Δ.Ε.Π. (Λέκτορες), όπου εμφανίζονται εγγραφές μόνο για ελάχιστες δημοσιεύσεις.

Έπειτα από την προεπεξεργασία του συνόλου των δεδομένων, ακολούθησε η επιλογή των κατάλληλων αλγορίθμων για την εξόρυξη γνώσης. Έτσι, με βάση τα δεδομένα και τους στόχους της εργασίας μας, αποφασίστηκε η χρήση του αλγόριθμου J48 για τη δημιουργία δέντρου απόφασης, η χρήση του Μπεϋζιανού ταξινομητή (Naïve Bayes) και του IBk (k-

Nearest Neighbor) για την κατηγοριοποίηση των δεδομένων. Επιπλέον, για την τεχνική της συσταδοποίησης χρησιμοποιήθηκε ο αλγόριθμος K-Means.

Για τη μέτρηση της ακρίβειας των αλγορίθμων κατηγοριοποίησης που εκτελέστηκαν χρησιμοποιήθηκε η επιλογή 10 fold Cross-Validation. Με τη μέθοδο Cross-Validation (διασταυρωμένη επικύρωση) το σύνολο διαιρείται σε 10 υποσύνολα και η επιλογή των συνόλων γίνεται τυχαία. Ένα από τα υποσύνολα χρησιμοποιείται ως σύνολο επικύρωσης (testing set) και τα υπόλοιπα εννέα συνενώνονται και δημιουργούν το σύνολο εκπαίδευσης (training test). Το μοντέλο εκπαιδεύεται χρησιμοποιώντας το σύνολο εκπαίδευσης και δοκιμάζεται έναντι του συνόλου επικύρωσης. Η διαδικασία επαναλαμβάνεται 10 φορές, κάθε φορά χρησιμοποιώντας διαφορετικό σύνολο επικύρωσης και τα υπόλοιπα εννέα ως σύνολο εκπαίδευσης. Στο τέλος υπολογίζεται η μέση απόδοση του μοντέλου.

Στις επόμενες ενότητες παρουσιάζεται αναλυτικά η εκτέλεση των επιλεγμένων αλγορίθμων κατηγοριοποίησης και συσταδοποίησης για την εξόρυξη γνώσης. Σε κάθε περίπτωση, ακολουθεί η αξιολόγηση και η ερμηνεία της εξαγόμενης γνώσης.

5.3 Αλγόριθμοι κατηγοριοποίησης

5.3.1 Μπεϋζιανός ταξινομητής (Naïve Bayes)

Για την εκτέλεση του αλγορίθμου Naïve Bayes έγινε προεπεξεργασία των δεδομένων που χρησιμοποιήθηκαν. Αυτή περιλαμβάνει τη μετατροπή των αριθμητικών γνωρισμάτων (numeric) σε περιγραφικά (nominal), με τη χρήση των φίλτρων του Weka και ακολουθώντας τη διαδρομή `weka.filters.unsupervised.attribute.NumericToNominal`.

Τα αριθμητικά γνωρίσματα που μετατράπηκαν σε περιγραφικά είναι τα ακόλουθα: Δημοσιεύσεις (P), Αναφορές (C), Μέσος όρος h index (H), Δημοσιεύσεις μελών Δ.Ε.Π. τα τελευταία 5 έτη (P5), Αναφορές τα τελευταία 5 έτη (C5), Μέσος όρος του h index τα τελευταία 5 έτη (H5), Μέσος όρος δημοσιεύσεων ανά μέλος Δ.Ε.Π. ανά έτος δημοσίευσης (age).

Έπειτα από την προεπεξεργασία των δεδομένων, προχωρήσαμε στην εκτέλεση του αλγορίθμου Naïve Bayes. Το ποιοτικό χαρακτηριστικό που εξετάστηκαν για την κατηγοριοποίηση είναι η ακαδημαϊκή βαθμίδα του μέλους ΔΕΠ (P). Για τη μέτρηση της ακρίβειας του μοντέλου χρησιμοποιήθηκε η επιλογή 10 fold Cross-Validation.

Στην εικόνα 5.1 εμφανίζονται τα αποτελέσματα σχετικά με την κατηγοριοποίηση των μελών Δ.Ε.Π. ανά βαθμίδα, όπου διαπιστώνεται ότι κατηγοριοποιήθηκαν σωστά 454 μέλη από το σύνολο των 901 μελών Δ.Ε.Π. Από τα αποτελέσματα του αλγόριθμου γίνεται αντιληπτό ότι δεν έγινε πολύ καλή πρόβλεψη, καθώς το ποσοστό επιτυχίας ανέρχεται στο 50.3885%. Διαπιστώνεται ότι την καλύτερη πρόβλεψη παρουσίασε η βαθμίδα του Καθηγητή καθώς πρόβλεψε τους 231 από τους 382. Επιπλέον είναι σαφές ότι στη βαθμίδα του Αναπληρωτή Καθηγητή και του Επίκουρου Καθηγητή η πρόβλεψη είναι πολύ χαμηλή καθώς φαίνεται ότι ο αλγόριθμος δεν τους κατηγοριοποιεί σωστά. Αυτό επιβεβαιώνεται και από τη μήτρα σύγχυσης (Confusion Matrix), όπου στη διαγώνιο του πίνακα παρουσιάζονται τα λίγα στιγμιότυπα που έχουν κατηγοριοποιηθεί σωστά.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   454      50.3885 %
Incorrectly Classified Instances 447      49.6115 %
Kappa statistic                  0.3217
Mean absolute error              0.2172
Root mean squared error          0.3659
Relative absolute error          76.1062 %
Root relative squared error      96.8923 %
Total Number of Instances       901

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      0,605  0,164  0,731  0,605  0,662  0,457  0,824  0,760  Professor
      0,379  0,160  0,363  0,379  0,371  0,216  0,733  0,431  Assistant Professor
      0,395  0,193  0,377  0,395  0,386  0,199  0,683  0,352  Associate Professor
      0,633  0,131  0,427  0,633  0,510  0,429  0,851  0,505  Lab Lecturer
      0,000  0,011  0,000  0,000  0,000  -0,016  0,607  0,032  Lecturer
Weighted Avg.  0,504  0,162  0,523  0,504  0,508  0,337  0,773  0,554

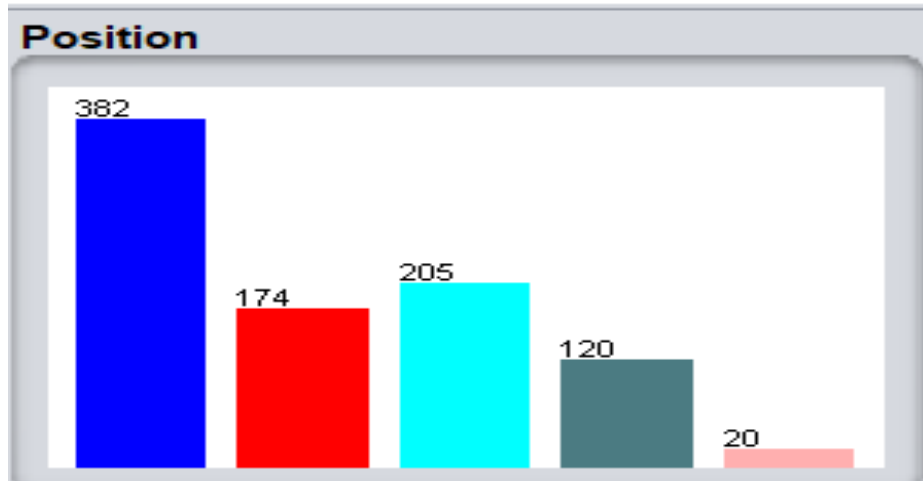
=== Confusion Matrix ===
 a  b  c  d  e <-- classified as
231 41 73 32  5 | a = Professor
 23 66 50 34  1 | b = Assistant Professor
 54 47 81 20  3 | c = Associate Professor
  8 26  9 76  1 | d = Lab Lecturer
  0  2  2 16  0 | e = Lecturer

```

Εικόνα 5.1. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο Naïve Bayes

Επειδή το σύνολο των δεδομένων είναι μη ισορροπημένα (imbalanced) δηλαδή η πλειοψηφία των στιγμιότυπων ανήκει στην κλάση professor εκτελέσαμε oversampling (υπερπροσαρμογή) μέσω της μεθόδου SMOTE ώστε να ενισχύσουμε τις άλλες κλάσεις με «τεχνητά» στιγμιότυπα με στόχο την αύξηση της ακρίβειας. Η μέθοδος SMOTE είναι

προσβάσιμη στο WEKA από το tab-preprocess ->filter->filters->supervised->instances. Αυξάνοντας τα στιγμιότυπα των κλάσεων όλες οι κλάσεις περιέχουν τον ίδιο αριθμό στιγμιότυπων (εικ. 5.2, 5.3). Για τη δημιουργία ενός συνθετικού δείγματος επιλέγεται τυχαία ένας από τους k-πλησιέστερους γείτονες και πολλαπλασιάζεται η διαφορά διανυσμάτων με έναν τυχαίο αριθμό μεταξύ [0,1] και προκύπτει το επαναληπτικό σχήμα.



Εικόνα 5.2. Το σύνολο των δεδομένων πριν την εφαρμογή της SMOTE



Εικόνα 5.3. Το σύνολο των δεδομένων μετά την εφαρμογή της SMOTE

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   1078      56.4398 %
Incorrectly Classified Instances  832      43.5602 %
Kappa statistic                 0.4555
Mean absolute error             0.1845
Root mean squared error        0.3738
Relative absolute error        57.6583 %
Root relative squared error    93.4455 %
Total Number of Instances      1910

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      0,469   0,037   0,762   0,469   0,580   0,526   0,894   0,667   Professor
      0,380   0,110   0,463   0,380   0,417   0,291   0,780   0,482   Assistant Professor
      0,351   0,088   0,500   0,351   0,412   0,303   0,784   0,449   Associate Professor
      0,730   0,239   0,433   0,730   0,544   0,416   0,812   0,483   Lab Lecturer
      0,893   0,071   0,758   0,893   0,820   0,774   0,945   0,859   Lecturer
Weighted Avg. 0,564 0,109 0,583 0,564 0,555 0,462 0,843 0,588

=== Confusion Matrix ===

  a  b  c  d  e  <-- classified as
179 38 88 70 7  | a = Professor
 13 145 36 164 24 | b = Assistant Professor
 42 87 134 115 4 | c = Associate Professor
 1 18 10 279 74 | d = Lab Lecturer
 0 25 0 16 341 | e = Lecturer

```

Εικόνα 5.4. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο Naïve Bayes και τη χρήση της μεθόδου SMOTE

Όπως διαπιστώνεται, με τη μέθοδο SMOTE, βελτιώθηκε ελάχιστα το ποσοστό πρόβλεψης του αλγόριθμου σε 56.4398%, αλλά κατηγοριοποιήθηκαν καλύτερα η βαθμίδα του Λέκτορα και οι ΕΔΠΠ. Αυτό φαίνεται και από το Precision που αρχικά ήταν 0,427 και έγινε 0,433, ενώ το Recall από 0,633 έγινε 0,730 για τους ΕΔΠΠ. Παρομοίως, στην περίπτωση του Λέκτορα από 0,000 το Precision έγινε 0,758 και το Recall 0,893.

5.3.2 Δέντρα Απόφασης (Decision Trees)

Προχωρήσαμε στην εκτέλεση του αλγόριθμου J48, που είναι ο πιο κλασικός αλγόριθμος για την παραγωγή δέντρου απόφασης και την κατηγοριοποίηση των δεδομένων. Για τον έλεγχο του δέντρου χρησιμοποιήθηκε η τεχνική 10 fold Cross Validation. Για την εκτέλεση του αλγόριθμου με το αρχικό σύνολο δεδομένων ορίστηκε η παράμετρος minNumObj με την τιμή 6 και η παράμετρος unpruned ως False.

Όπως διαπιστώνεται από τα στοιχεία της εικόνας 5.5, ο αλγόριθμος προέβλεψε σωστά 510 στιγμιότυπα από το σύνολο των 901. Γίνεται λοιπόν αντιληπτό ότι το ποσοστό επιτυχίας του

αλγόριθμοι δεν είναι πολύ υψηλό, καθώς κυμαίνεται στο 56.6038 %. Όσον αφορά την πρόβλεψη για την βαθμίδα του Καθηγητή, έχει προβλέψει 330 από τους 382 όπως διαπιστώνεται και από το Precision που είναι 0,651 και το Recall 0,864. Επιπλέον κατηγοριοποίησε τους 75 από τους 120 ΕΔΙΠ, με Precision 0,503 και Recall 0,625.

```

Number of Leaves : 4
Size of the tree : 7

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 510 56.6038 %
Incorrectly Classified Instances 391 43.3962 %
Kappa statistic 0.366
Mean absolute error 0.2308
Root mean squared error 0.3436
Relative absolute error 80.8544 %
Root relative squared error 90.9891 %
Total Number of Instances 901

=== Detailed Accuracy By Class ===

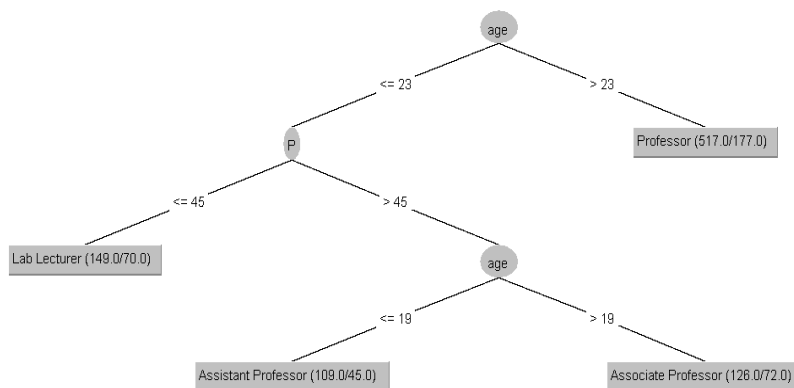
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0,864 0,341 0,651 0,864 0,742 0,521 0,786 0,657 Professor
0,460 0,139 0,442 0,460 0,451 0,316 0,725 0,362 Assistant Professor
0,122 0,056 0,391 0,122 0,186 0,108 0,612 0,298 Associate Professor
0,625 0,095 0,503 0,625 0,558 0,485 0,751 0,359 Lab Lecturer
0,000 0,000 ? 0,000 ? ? 0,820 0,103 Lecturer
Weighted Avg. 0,566 0,197 ? 0,566 ? ? 0,731 0,466

=== Confusion Matrix ===
a b c d e <-- classified as
330 26 18 8 0 | a = Professor
40 80 15 39 0 | b = Assistant Professor
109 60 25 11 0 | c = Associate Professor
27 13 5 75 0 | d = Lab Lecturer
1 2 1 16 0 | e = Lecturer

```

Εικόνα 5.5. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. με τον αλγόριθμο J48

Στην εικόνα 5.6 απεικονίζεται το δέντρο απόφασης, έπειτα από την κατηγοριοποίηση με τον αλγόριθμο J48. Όπως γίνεται αντιληπτό πρόκειται για ένα σχετικά μικρό διάγραμμα δέντρου.



Εικόνα 5.6. Απεικόνιση του Δέντρου Απόφασης

Στη συνέχεια, αποφασίστηκε να γίνει νέα δοκιμή, ώστε να βελτιωθεί το ποσοστό πρόβλεψης του αλγόριθμου με τη χρήση της μεθόδου SMOTE. Έτσι, εκτελέσαμε τον αλγόριθμο με το αρχικό σύνολο δεδομένων ορίζοντας, όμως, ως παράμετρο `minNumObj` την τιμή 6 και την παράμετρο `unpruned` ως `False`. Όπως διαπιστώνεται από τα στοιχεία της εικόνας 5.7, ο αλγόριθμος αύξησε το ποσοστό επιτυχίας στο 71.2565%. Αυτό γίνεται, λοιπόν, αντιληπτό και από το Precision και το Recall, όπου οι τιμές όλων των κλάσεων βρίσκονται πάνω από το 0,5.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   1361      71.2565 %
Incorrectly Classified Instances  549      28.7435 %
Kappa statistic                 0.6407
Mean absolute error             0.1424
Root mean squared               0.2946
Relative absolute error         44.5124 %
Root relative squared           73.6395 %
Total Number of Instances      1910

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      0,738   0,099   0,650   0,738   0,691   0,610   0,883   0,629   Professor
      0,652   0,096   0,629   0,652   0,640   0,548   0,839   0,607   Assistant Professor
      0,497   0,094   0,569   0,497   0,531   0,425   0,815   0,535   Associate Professor
      0,733   0,046   0,798   0,733   0,764   0,709   0,899   0,756   Lab Lecturer
      0,942   0,023   0,911   0,942   0,927   0,908   0,975   0,942   Lecturer
Weighted Avg.   0,713   0,072   0,711   0,711   0,713   0,711   0,640   0,882   0,694

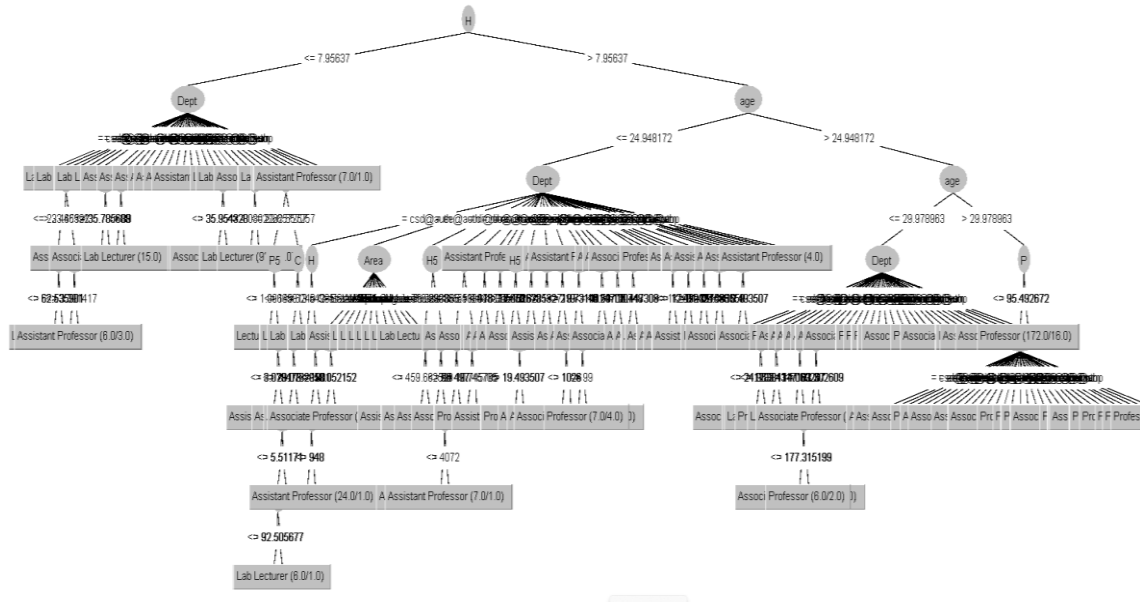
=== Confusion Matrix ===

 a  b  c  d  e  <-- classified as
282 29 60 9 2  | a = Professor
43 249 56 26 8  | b = Assistant Professor
90 73 190 22 7  | c = Associate Professor
17 40 27 280 18  | d = Lab Lecturer
2 5 1 14 360  | e = Lecturer

```

Εικόνα 5.7. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο J48 και τη χρήση της μεθόδου SMOTE

Στην εικόνα 5.8 απεικονίζεται το δέντρο απόφασης μετά τη χρήση της μεθόδου SMOTE. Διαπιστώνεται ότι πρόκειται για ένα αρκετά εκτενές διάγραμμα δέντρου, καθώς λόγω της χρήσης της μεθόδου αυξήθηκε το σύνολο των δεδομένων.

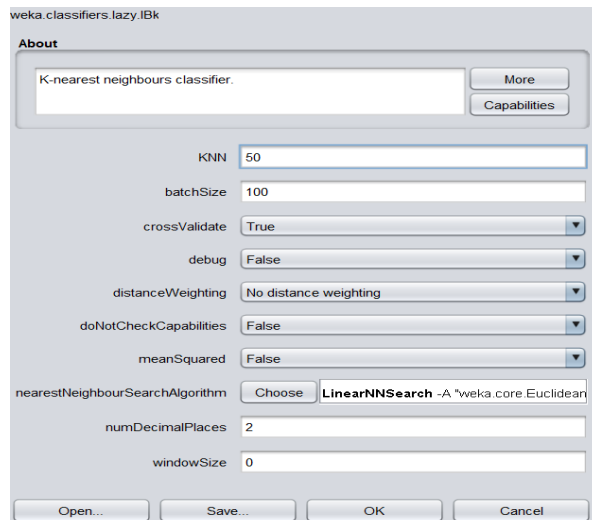


Εικόνα 5.8. Απεικόνιση του Δέντρου Απόφασης με τη χρήση της μέθοδου SMOTE

5.3.3 Αλγόριθμος IBk (K-nearest neighbor)

Για την κατηγοριοποίηση των δεδομένων μας χρησιμοποιήθηκε και ο αλγόριθμος IBk (K-nearest neighbor). Το ποιοτικό χαρακτηριστικό που εξετάστηκε είναι και πάλι η ακαδημαϊκή βαθμίδα των μελών ΔΕΠ (P).

Στην περίπτωση του αλγόριθμου IBk είναι ιδιαίτερα σημαντικός ο προσδιορισμός της παραμέτρου k, δηλαδή της απόστασης από τον k πλησιέστερο γείτονα (μέτρο απόστασης). Μέσα από το WEKA μπορέσαμε να εκτελέσουμε τον ibk με την καλύτερη δυνατή τιμή για το k, η οποία αποδεικνύεται βάσει της μεθόδου cross-validation (χρήση validation set). Αυτό γίνεται εφόσον επιλέξαμε “cross-validate” που επιτρέπει τον καθορισμό πολλαπλών μετρήσεων για αξιολόγηση από το παράθυρο παραμέτρων του κατηγοριοποιητή, και του δώσαμε την τιμή True. Σε αυτή την περίπτωση το WEKA έκανε τόσες δοκιμές όσες η τιμή που βάλαμε στο textbox “KNN”, δηλαδή 50 και μας εμφάνισε ότι το καλύτερο k=4 (εικ. 5.9)



Εικόνα 5.9. Προσδιορισμός k-πλησιέστερου μέσω WEKA

Σύμφωνα με τα αποτελέσματα της εκτέλεσης του αλγόριθμου, από τα 901 στιγμιότυπα ταξινομήθηκαν σωστά τα 522 στιγμιότυπα, που αντιστοιχούν σε ποσοστό επιτυχίας 57,9%. Από τους Καθηγητές ταξινομήθηκαν με μεγάλη ακρίβεια οι 322 από τους 380, ενώ καλή κατηγοριοποίηση προέκυψε και για τους Λέκτορες (62 από τους 120) (εικ. 5.10).

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      522      57.9356 %
Incorrectly Classified Instances    379      42.0644 %
Kappa statistic                    0.3873
Mean absolute error                 0.1992
Root mean squared error            0.3472
Relative absolute error            69.7928 %
Root relative squared error        91.9568 %
Total Number of Instances          901

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      0,843   0,304   0,671   0,843   0,747   0,533   0,846   0,746   Professor
      0,454   0,116   0,485   0,454   0,469   0,347   0,728   0,405   Assistant Professor
      0,278   0,131   0,385   0,278   0,323   0,167   0,644   0,322   Associate Professor
      0,517   0,051   0,608   0,517   0,559   0,499   0,805   0,500   Lab Lecturer
      0,100   0,007   0,250   0,100   0,143   0,146   0,671   0,113   Lecturer
Weighted Avg.  0,579   0,188   0,552   0,579   0,558   0,401   0,768   0,537

=== Confusion Matrix ===

 a  b  c  d  e  <-- classified as
322 17 38  5  0 | a = Professor
 42 79 38 13  2 | b = Assistant Professor
100 36 57 11  1 | c = Associate Professor
 15 27 13 62  3 | d = Lab Lecturer
  1  4  2 11  2 | e = Lecturer

```

Εικόνα 5.10. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο IBk

Επιπλέον, πραγματοποιήθηκαν δοκιμές στα δεδομένα μας με βάση τη μέθοδο SMOTE. Σε αυτήν την περίπτωση, παρατηρήθηκε πολύ καλή απόδοση πρόβλεψης του αλγόριθμου με ποσοστό επιτυχίας 72,1466% (εικ. 5.10). Όπως διαπιστώνεται από τα στοιχεία της εικόνας 5.10, οι τιμές του Precision και του Recall, είναι >0,5 για όλες τις βαθμίδες και μάλιστα σε κάποιες περιπτώσεις πλησιάζουν το 1 (εικ. 5.11.).

```

IB1 instance-based classifier
using 4 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   1378       72.1466 %
Incorrectly Classified Instances  532       27.8534 %
Kappa statistic                 0.6518
Mean absolute error             0.1333
Root mean squared error         0.2851
Relative absolute error         41.649 %
Root relative squared error     71.2829 %
Total Number of Instances      1910

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      0,754   0,098   0,659   0,754   0,703   0,625   0,900   0,692   Professor
      0,691   0,098   0,638   0,691   0,663   0,576   0,866   0,618   Assistant Professor
      0,450   0,089   0,558   0,450   0,499   0,393   0,818   0,487   Associate Professor
      0,759   0,041   0,822   0,759   0,789   0,740   0,917   0,784   Lab Lecturer
      0,953   0,022   0,915   0,953   0,933   0,917   0,982   0,936   Lecturer
Weighted Avg.  0,721   0,070   0,718   0,721   0,718   0,650   0,896   0,703

=== Confusion Matrix ===

  a  b  c  d  e  <-- classified as
288 22 61  6  5 | a = Professor
 33 264 51 28  6 | b = Assistant Professor
102 84 172 20  4 | c = Associate Professor
 13 39 21 290 19 | d = Lab Lecturer
  1  5  3  9 364 | e = Lecturer

```

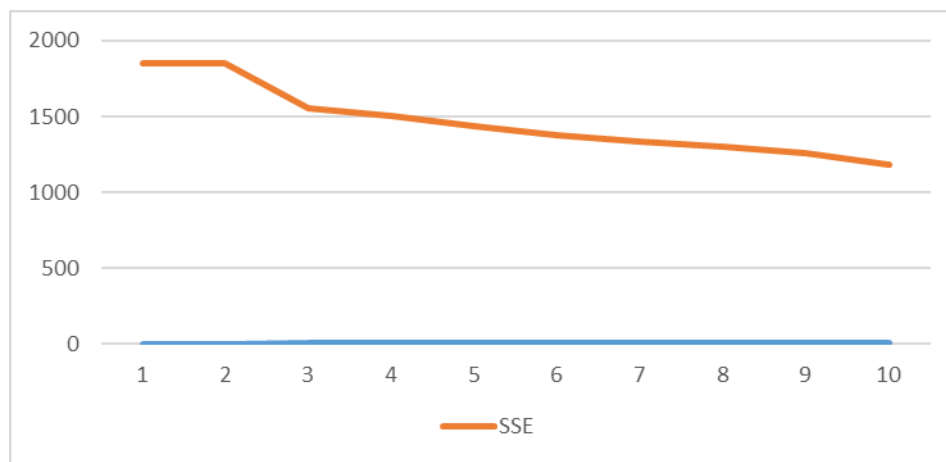
Εικόνα 5.11. Αποτελέσματα κατηγοριοποίησης μελών Δ.Ε.Π. ανά βαθμίδα με τον αλγόριθμο IBk με τη μέθοδο SMOTE

5.4 Αλγόριθμος συσταδοποίησης K-Means

Για την τεχνική της συσταδοποίησης των δεδομένων χρησιμοποιήθηκε ο αλγόριθμος K-Means, με στόχο την ομαδοποίηση των μελών Δ.Ε.Π. ανα ακαδημαϊκή βαθμίδα. Η επιλογή του αριθμού των συστάδων βασίστηκε στη μέθοδο του αγκώνα (Elbow method) [25]. Έτσι, ο αλγόριθμος K-means εκτελέστηκε για διάφορες τιμές του k, καταγράφοντας το άθροισμα των τετραγώνων των σφαλμάτων (SSE- Sum of Squared Error). Σύμφωνα με τις τιμές που προέκυψαν, διαπιστώθηκε ότι η τιμή k=3 είναι η πιο ικανοποιητική για την ομαδοποίηση των μελών ΔΕΠ (βλ. πίν. 5.1, γράφημα 5.1).

k	2	3	4	5	6	7	8	9	10	11
SSE	1855,29	1848,89	1551,92	1505,93	1433,81	1381,50	1333,29	1302,47	1258,73	1187,37

Πίνακας 5.1. Τιμές SSE ανά k



Γράφημα 5.1. Μέθοδος αγκώνα για την επιλογή του k

Στη συνέχεια προχωρήσαμε στην εκτέλεση του αλγόριθμου K-means για k=3 με τη επιλογή Use training set. Για τον υπολογισμό της απόστασης χρησιμοποιήθηκε η ευκλείδεια μετρική και με don't normalize= false. Το αποτέλεσμα του αλγόριθμου παρουσιάζεται στην εικόνα 5.12.

kMeans

=====

Number of iterations: 8

Within cluster sum of squared errors: 1648.894201558836

Initial starting points (random):

Cluster 0: Professor,ece@ntua,Attiki,404,3380,27,12,637,13,33

Cluster 1: 'Associate Professor',cs@aueb,Attiki,21,191,8,6,73,5,20

Cluster 2: 'Assistant Professor',dib@uth,'Sterea Ellada',23,543,10,10,320,8,24

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (901.0)	Cluster#		
		0 (362.0)	1 (280.0)	2 (259.0)

=====

Position	Professor	Professor Associate	Professor Assistant	Professor
Dept	ece@ntua	ece@ntua	eee@uniwa	ai@uom
Area	Attiki	Attiki	Attiki Central	Makedonia
P	118.3862	181.5801	91.4643	59.166
C	2217.8616	3695.0193	1547.8123	877.6404
H	18.9445	25.9116	16.0929	12.2896
P5	25.4406	31.7873	23.9286	18.2046
C5	868.6116	1299.3646	681.3516	468.9985
H5	11.9877	15.1851	10.7498	8.8571
age	25.3485	30.7514	23.475	19.8224

Time taken to build model (full training data) : 0.1 seconds

=== Model and evaluation on training set ===

Clustered Instances

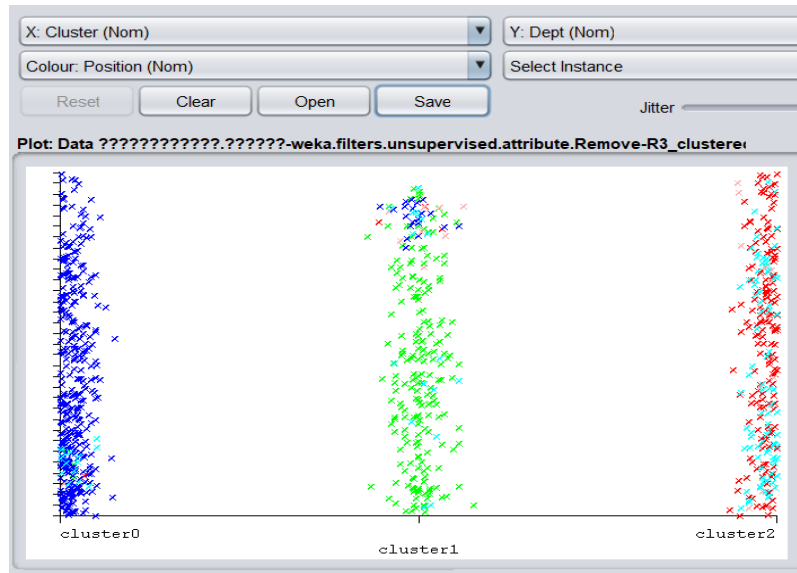
0	362 (40%)
1	280 (31%)
2	259 (29%)

Εικόνα 5.12. Αποτελέσματα συσταδοποίησης των μελών Δ.Ε.Π. με τον αλγόριθμο K-means

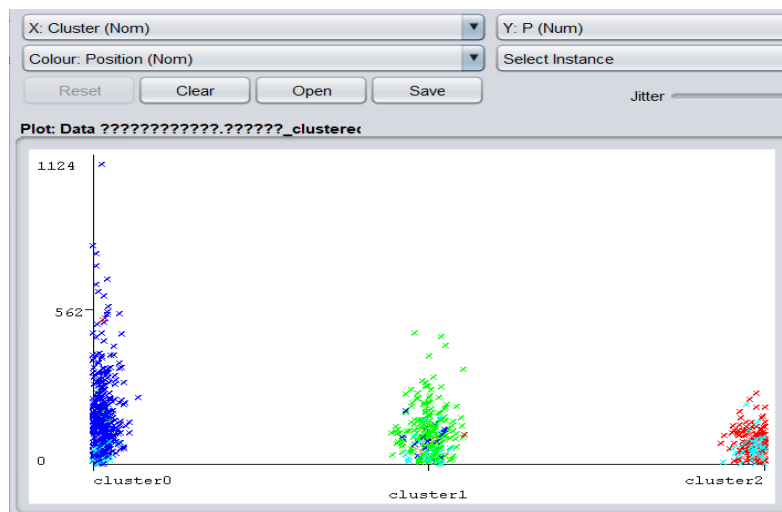
Σύμφωνα με τα αποτελέσματα της συσταδοποίησης διαχωρίστηκαν τρεις συστάδες. Στην πρώτη συστάδα (cluster 0) ομαδοποιήθηκαν 362 στιγμιότυπα, καλύπτοντας το 40% του συνόλου. Η δεύτερη συστάδα (cluster 1) αντιστοιχεί σε 280 στιγμιότυπα με ποσοστό 31%. Τέλος, η Τρίτη συστάδα (cluster 2) περιλαμβάνει 259 στιγμιότυπα με ποσοστό 29% επί του συνόλου.

Στη συνέχεια ακολούθησε η διερεύνηση των τριών συστάδων σύμφωνα με τα αποτελέσματα του αλγόριθμου. Για το σκοπό αυτό οπτικοποιήσαμε τα αποτελέσματα επιλέγοντας Visualize Cluster Assignments. Στις επόμενες εικόνες παρατηρούμε τις τρεις συστάδες ανά τμήμα (εικ. 5.13), ανά σύνολο δημοσιεύσεων (εικ. 5.14), ανά σύνολο αναφορών (εικ. 5.15) και ανά σύνολο ετών δημοσιεύσεων (εικ. 5.16). Με μπλε χρώμα απεικονίζονται οι Καθηγητές, με

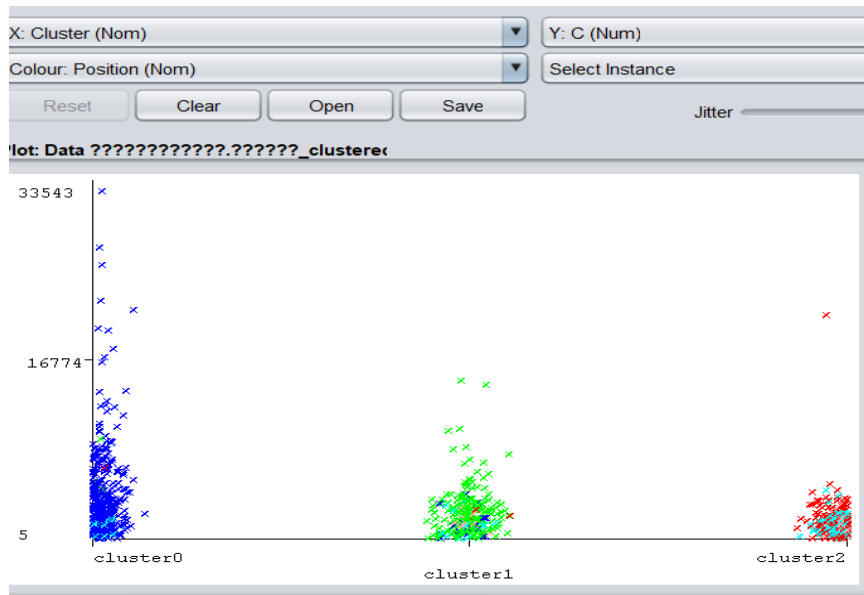
πράσινο οι Αναπληρωτές Καθηγητές και με κόκκινο οι Επικουροι καθηγητές, με ανοιχτό γαλάζιο οι ΕΔΙΠ και με ανοιχτό κόκκινο οι Λέκτορες.



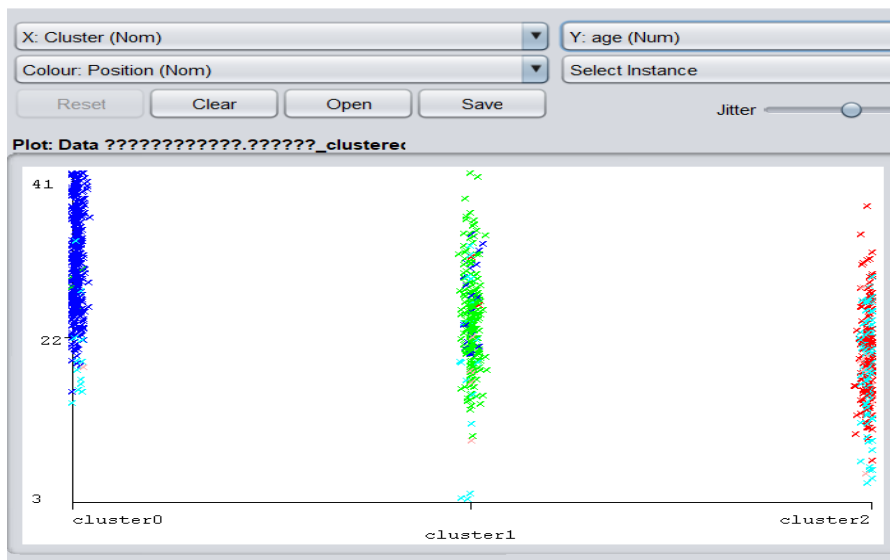
Εικόνα 5.13. Αποτελέσματα συσταδοποίησης με K-means ανά τμήμα



Εικόνα 5.14. Αποτελέσματα συσταδοποίησης με K-means ανά σύνολο δημοσιεύσεων



Εικόνα 5.15. Αποτελέσματα συσταδοποίησης με K-means ανά σύνολο αναφορών



Εικόνα 5.16. Αποτελέσματα συσταδοποίησης με K-means ανά σύνολο ετών δημοσιεύσεων

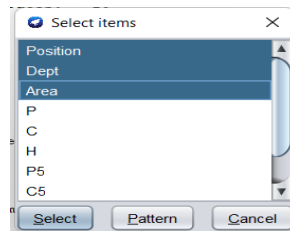
Σύμφωνα με τα αποτελέσματα του αλγόριθμου και τις παραπάνω εικόνες, οι τρεις συστάδες που προέκυψαν αντιστοιχούν σε ακαδημαϊκές βαθμίδες ανάλογα με τα χαρακτηριστικά της.

Η πρώτη συστάδα (Cluster 0) περιλαμβάνει σε μεγάλο βαθμό Καθηγητές και άλλα μέλη ΔΕΠ από το Εθνικό Μετσόβιο Πολυτεχνείο. Στη δεύτερη συστάδα (Cluster 1) ταξινομήθηκαν κατά κύριο λόγο Αναπληρωτές Καθηγητές ενώ ελάχιστες περιπτώσεις αντιστοιχούν σε μέλη ΕΔΙΠ του Πανεπιστημίου Δυτικής Αττικής καθώς και μέλη ΕΔΙΠ από άλλα πανεπιστήμια με

υψηλές ερευνητικές επιδόσεις. Τέλος η τρίτη συστάδα περιλαμβάνει κυρίως Επίκουρους Καθηγητές και λίγα μέλη ΕΔΙΠ.

Αλγόριθμος συσταδοποίησης K-Means (2η εκτέλεση)

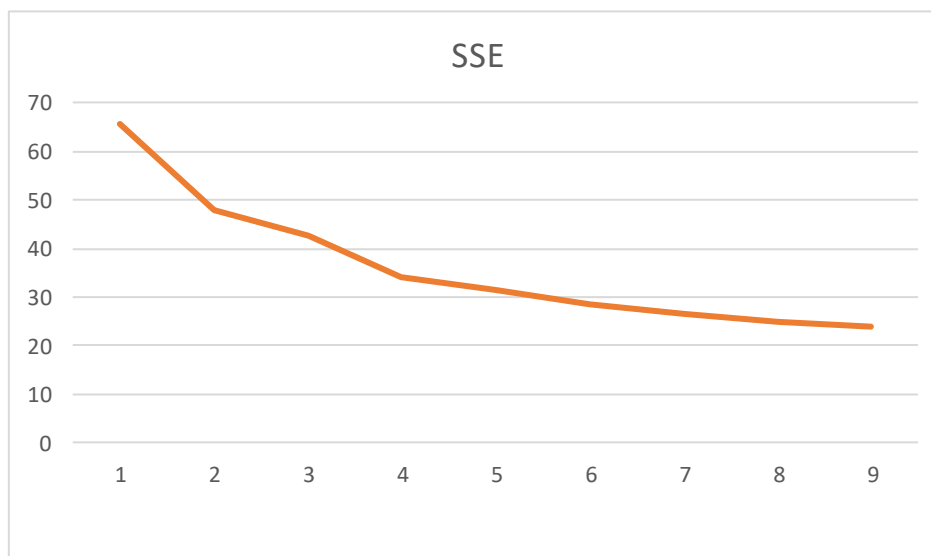
Στη συνέχεια εκτελέσαμε τον αλγόριθμο K-means, με στόχο την ομαδοποίηση των μελών Δ.Ε.Π. με βάση των ερευνητικών τους επιδόσεων. Αφαιρέσαμε τα χαρακτηριστικά βαθμίδα (position), τμήμα (dept), περιοχή (area) (εικ. 5.17). Η επιλογή του αριθμού των συστάδων βασίστηκε στη μέθοδο του αγκώνα (Elbow method). Έτσι, ο αλγόριθμος K-means εκτελέστηκε για διάφορες τιμές του k, καταγράφοντας το άθροισμα των τετραγώνων των σφαλμάτων (SSE- Sum of Squared Error). Σύμφωνα με τις τιμές που προέκυψαν, διαπιστώθηκε ότι η τιμή k=4 είναι η πιο ικανοποιητική για την ομαδοποίηση των μελών ΔΕΠ



Εικόνα 5.17 Διαγραφή χαρακτηριστικών

K	2	3	4	5	6	7	8	9	10
SSE	65,69	48,06	42,51	34,18	31,59	28,47	26,41	24,96	23,75

Πίνακας 5.2. Τιμές SSE ανά k



Γράφημα 5.2. Μέθοδος αγκώνα για την επιλογή του k (2^η εκτέλεση του αλγόριθμου)

Στη συνέχεια προχωρήσαμε στην εκτέλεση του αλγόριθμου K-means για k=5 με τη επιλογή Use training set. Για τον υπολογισμό της απόστασης χρησιμοποιήθηκε η ευκλείδεια μετρική και με don't normalize= false. Το αποτέλεσμα του αλγόριθμου παρουσιάζεται στην εικόνα 5.18.

```

kMeans
=====

Number of iterations: 24

Within cluster sum of squared errors: 34.18402809236463

Initial starting points (random):

Cluster 0: 404,3380,27,12,637,13,33
Cluster 1: 21,191,8,6,73,5,20
Cluster 2: 23,543,10,10,320,8,24
Cluster 3: 144,1431,21,7,330,10,29
Cluster 4: 40,253,7,13,150,6,16

Attribute Full Data    0    1    2    3    4
          (901.0) (79.0) (243.0) (248.0) (144.0) (187.0)
=====
P      118.3862 368.6835 43.6379 62.8911 146.9931 161.3476
C      2217.8616 9959.0506 562.3963 741.0968 2361.4375 2946.6684
H      18.9445 45.7975 9.9424 12.2177 22.2361 25.6845
P5     25.4406 74.9241 13.3786 10.1331 18.6042 45.7754
C5     868.6116 3855.3797 279.2636 244.0806 671.3819 1352.7914
H5     11.9877 28.3038 7.4689 7.1129 11.8611 17.5294
age    25.3485 31.3038 17.1152 27.3226 35.9236 22.7701

Time taken to build model (full training data) : 0.16 seconds
=== Model and evaluation on training set ===

Clustered Instances

0      79 ( 9%)
1      243 ( 27%)
2      248 ( 28%)
3      144 ( 16%)
4      187 ( 21%)

```

Εικόνα 5.18 Αποτελέσματα συσταδοποίησης των μελών Δ.Ε.Π. με τον αλγόριθμο K-means (2^η εκτέλεση του αλγόριθμου)

Σύμφωνα με τα αποτελέσματα της συσταδοποίησης διαχωρίστηκαν πέντε συστάδες. Η πρώτη συστάδα (cluster 0) περιλαμβάνει 79 στιγμιότυπα, καλύπτοντας το 9% του συνόλου. Η δεύτερη συστάδα (cluster 1) αντιστοιχεί σε 243 στιγμιότυπα με ποσοστό 27%. Η τρίτη συστάδα (cluster 2) περιλαμβάνει 248 στιγμιότυπα με ποσοστό 28%, η τέταρτη συστάδα (cluster 3) 144 στιγμιότυπα με ποσοστό 16% και η πέμπτη συστάδα (cluster 4) 187 στιγμιότυπα με ποσοστό 21%.

Η πρώτη συστάδα (Cluster 0) περιλαμβάνει κυρίως Καθηγητές και λίγους Αν. Καθηγητές με πολύ υψηλές ερευνητικές επιδόσεις, $h\ index > 30$ και 3736 αναφορές. Στη δεύτερη συστάδα (Cluster 1) περιλαμβάνει κυρίως Αναπληρωτές Καθηγητές και Επίκουρους Καθηγητές καθώς και Λέκτορες, ΕΔΙΠ με λίγα χρόνια δημοσιεύσεων, που έχουν αξιόλογο έργο τα τελευταία 5 χρόνια, και έχουν μέχρι 2700 αναφορές. Ακόμη περιλαμβάνει και λίγους Καθηγητές από πρώην τμήματα ΤΕΙ. Η τρίτη συστάδα (cluster 2) περιλαμβάνει κυρίως Καθηγητές διαφόρων ηλικιών που έχουν μέχρι 3500 αναφορές αλλά την τελευταία πενταετία δεν έχουν αναφορές και δημοσιεύσεις. Επιπλέον υπάρχουν και πολλοί Αναπληρωτές Καθηγητές. Η τέταρτη συστάδα (cluster 3) έχει κατά κύριο λόγο Καθηγητές και λίγους Αναπληρωτές Καθηγητές με πολλά χρόνια δημοσιεύσεων, πάνω από 30 χρόνια. Η πέμπτη συστάδα (cluster 4) είναι καθηγητές όλων των βαθμίδων, νέοι, με πολλές δημοσιεύσεις τα τελευταία 5 χρόνια αλλά δεν έχουν λάβει πολλές αναφορές.

Κεφάλαιο 6

Συμπεράσματα - Μελλοντική έρευνα

6.1 Συμπεράσματα

Στο πρώτο μέρος της παρούσας εργασίας αξιολογήθηκαν τα τμήματα πληροφορικής που λειτουργούν στην Ελλάδα, καθώς και το επιστημονικό έργο των μελών Δ.Ε.Π., με την εφαρμογή στατιστικής ανάλυσης στα επιστημονομετρικά δεδομένα. Χρησιμοποιήθηκε μια σειρά επιστημονομετρικών δεικτών, για την αξιολόγηση της ερευνητικής δραστηριότητας (παραγωγικότητας) και της απήχησης που αυτή έχει στη διεθνή επιστημονική κοινότητα. Η άντληση των στοιχείων έγινε από τη βάση δεδομένων Google Scholar.

Σύμφωνα, λοιπόν, με τη στατιστική ανάλυση, διαπιστώθηκε ότι η Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ, εμφανίζει το μεγαλύτερο όγκο ερευνητικού έργου, σε σχέση με τα υπόλοιπα πανεπιστημιακά τμήματα πληροφορικής. Ωστόσο, όταν εξετάζεται ο μ.ο. των δημοσιεύσεων και αναφορών ανά μέλος Δ.Ε.Π., προηγείται το Τμήμα Πληροφορικής του ΑΠΘ, το οποίο όμως αριθμεί μόνο 28 μέλη Δ.Ε.Π.. Αντίθετα, η Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ, διαθέτει συνολικά 62 μέλη Δ.Ε.Π., που προφανώς επηρεάζουν και το μ.ο. που αντιστοιχεί σε δημοσιεύσεις και αναφορές. Από την άλλη, η εξέταση του παραγόμενου έργου των τμημάτων πληροφορικής με βάση το συνδυασμό των δημοσιεύσεων και αναφορών του κάθε τμήματος (με βάση το δείκτη h), υποδεικνύει ότι το Τμήμα Πληροφορικής του ΑΠΘ εμφανίζει την υψηλότερη παραγωγικότητα και απήχηση στις δημοσιεύσεις των μελών Δ.Ε.Π..

Οι χαμηλότερες επιδόσεις σε παραγωγικότητα και απήχηση του επιστημονικού έργου εντοπίζονται στο Τμήμα Πληροφορικής του Πανεπιστημίου Δυτικής Μακεδονίας και στο Τμήμα Ψηφιακών Συστημάτων της Σχολής Οικονομίας και Τεχνολογίας του Πανεπιστημίου Πελοποννήσου, κάτι που επιβεβαιώνεται και από τους χαμηλούς μ.ο. δημοσιεύσεων και αναφορών ανά μέλος Δ.Ε.Π.. Μάλιστα, η εξέταση του δείκτη h , επαληθεύει τη χαμηλή παραγωγικότητα και απήχηση, κυρίως του Τμήματος Ψηφιακών Συστημάτων της Σχολής Οικονομίας και Τεχνολογίας του Πανεπιστημίου Πελοποννήσου.

Όσον αφορά την αξιολόγηση των μελών Δ.Ε.Π. σε πανελλαδικό επίπεδο, διαπιστώθηκε ότι ο μεγαλύτερος αριθμός μελών Δ.Ε.Π. προς εξέλιξη στην επόμενη βαθμίδα αφορά σε Αναπληρωτές Καθηγητές (15 μέλη). Από τις υπόλοιπες βαθμίδες είναι λίγα τα μέλη Δ.Ε.Π.

που θα μπορούσαν να εξελιχθούν σύμφωνα με την παραγωγικότητα και την απήχηση του έργου τους. Τα περισσότερα μέλη Δ.Ε.Π. που παρουσιάζουν έντονο επιστημονικό έργο και έχουν τις προϋποθέσεις για εξέλιξη ανήκουν στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ (5 μέλη), στο Τμήμα Πληροφορικής του ΑΠΘ (5 μέλη), στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών του ΕΚΠΑ (4 μέλη) και στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΑΠΘ (4 μέλη). Τα στοιχεία αυτά συμβαδίζουν, επίσης, με τη θέση που κατέχουν τα τμήματα αυτά ως προς τη συνολικά υψηλή παραγωγικότητα και απήχηση του επιστημονικού έργου τους.

Επιπλέον, η αξιολόγηση των μελών ανά περιφερειακή ενότητα, έδειξε ότι το μεγαλύτερο αριθμό Δ.Ε.Π. που μπορεί να εξελιχθεί στην επόμενη βαθμίδα διαθέτουν τα τμήματα πληροφορικής της Αττικής (32 μέλη). Αυτό αφορά κυρίως σε μέλη ΕΔΙΠ, που θα μπορούσαν να μεταπηδήσουν στη βαθμίδα του Λέκτορα. Ακολουθούν τα τμήματα πληροφορικής στις περιφέρειες Κρήτης και της Κεντρικής Μακεδονίας, με 17 και επτά μέλη όλων των βαθμίδων Δ.Ε.Π., που θα μπορούσαν να εξελιχθούν στην επόμενη βαθμίδα.

Στο σημείο αυτό αξίζει να σχολιάσουμε κάποιους παράγοντες που έχουν συμβάλει στους υψηλότερους δείκτες κάποιων τμημάτων πληροφορικής. Αρχικά, ο αριθμός των μελών Δ.Ε.Π. που διαθέτει κάθε τμήμα, επηρεάζει τόσο το σύνολο των δημοσιεύσεων και αναφορών, όσο και το μ.ο. που αντιστοιχεί σε κάθε μέλος. Έτσι, τμήματα με πολυάριθμο προσωπικό Δ.Ε.Π. φαίνεται ότι παρουσιάζουν μεγαλύτερο όγκο δημοσιεύσεων και αναφορών. Στον υπολογισμό, όμως, του μ.ο. που αντιστοιχεί για κάθε μέλος Δ.Ε.Π., το αποτέλεσμα διαφοροποιείται ελαφρώς, ευνοώντας τα τμήματα με μικρότερο αριθμό μελών Δ.Ε.Π. και υψηλούς αριθμούς δημοσιεύσεων και αναφορών.

Ένα δεύτερο σημαντικό στοιχείο είναι η αρχαιότητα του κάθε τμήματος, καθώς τα παλαιότερα χρονολογικά τμήματα είναι αναμενόμενο να παρουσιάζουν μεγαλύτερο όγκο δημοσιεύσεων και πιθανώς αναφορών. Δεν είναι τυχαίο ότι στις πρώτες θέσεις τμημάτων με μεγαλύτερο όγκο δημοσιεύσεων βρίσκονται πανεπιστημιακά τμήματα της Θεσσαλονίκης, Αθήνας, Πάτρας και Κρήτης που ιδρύθηκαν κατά τη δεκαετία του '80 και '90 (Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΕΜΠ, Τμήμα Πληροφορικής ΑΠΘ, Τμήμα Πληροφορικής και Τηλεπικοινωνιών ΕΚΠΑ, Τμήμα Μηχανικών Η/Υ & Πληροφορικής- Πανεπιστήμιο Πατρών, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΑΠΘ, Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών- Πανεπιστήμιο Πατρών, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΔΠΘ, Τμήμα Επιστήμης Υπολογιστών- Πανεπιστήμιο Κρήτης). Μάλιστα, αυτά τα τμήματα διαθέτουν και αρκετά μεγάλο αριθμό μελών Δ.Ε.Π.. Από την άλλη, τμήματα που ιδρύθηκαν τις τελευταίες δεκαετίες, διαθέτουν μικρό αριθμό δημοσιεύσεων και μικρό αριθμό μελών

Δ.Ε.Π., όπως για παράδειγμα τα τμήματα που σχετικά πρόσφατα μετατράπηκαν από Τ.Ε.Ι. σε Α.Ε.Ι., στις Σέρρες, την Καστοριά, τη Λάρισα, τη Λαμία, την Καβάλα και την Τρίπολη (Τμήμα Πληροφορικής και Τηλεπικοινωνιών- Πανεπιστήμιο Θεσσαλίας, Τμήμα Πληροφορικής ΔΙ.ΠΑ.Ε., Τμήμα Ψηφιακών Συστημάτων-Πανεπιστήμιο Θεσσαλίας, Τμήμα Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών, ΔΙ.ΠΑ.Ε., Τμήμα Πληροφορικής και Τηλεπικοινωνιών-Πανεπιστήμιο Πελοποννήσου, Τμήμα Πληροφορικής- Πανεπιστήμιο Δυτικής Μακεδονίας).

Σε αυτές τις περιπτώσεις ο δείκτης h μπορεί να προσφέρει μια πιο αντικειμενική εικόνα. Έτσι, διαπιστώνεται ότι τα τμήματα με μ.ο. δείκτη $h > 20$ αφορούν κατά βάση παλιά τμήματα που λειτουργούν δεκαετίες και διαθέτουν μεγάλο προσωπικό Δ.Ε.Π. (Θεσσαλονίκη, Αθήνα, Κρήτη, Ιωάννινα, Πάτρα). Όπως προαναφέρθηκε, το Τμήμα Πληροφορικής του ΑΠΘ έχει την πρωτοκαθεδρία, αν και διαθέτει μικρότερο αριθμό μελών Δ.Ε.Π. από άλλα μεγάλα τμήματα. Όσο για τα τμήματα με το χαμηλότερο δείκτη h (μ.ο. δείκτη $h < 11$), διαπιστώνουμε ότι αυτά αφορούν τμήματα που ιδρύθηκαν την τελευταία δεκαετία και είναι λογικό να παρουσιάζουν τη χαμηλότερη παραγωγή επιστημονικού έργου, λόγω του μικρού χρονικού διαστήματος λειτουργίας τους ως Α.Ε.Ι. Τέλος, τα περισσότερα τμήματα (18) εμφανίζουν μ.ο. του δείκτη h που κυμαίνεται στο 12-20, περιλαμβάνοντας όχι μόνο τμήματα που μετατράπηκαν πρόσφατα από Τ.Ε.Ι. σε Α.Ε.Ι., αλλά και τμήματα που ιδρύθηκαν από το 2000 και διαθέτουν αρκετό διδακτικό προσωπικό. Εδώ πρέπει να σημειωθεί η περίπτωση του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων, το οποίο αν και λειτουργεί ήδη από το 1998, εμφανίζει πολύ χαμηλό αριθμό δημοσιεύσεων και αναφορών, αλλά και χαμηλό μ.ο. δείκτη h .

Συμπερασματικά, τα στοιχεία της στατιστικής ανάλυσης υποδηλώνουν ότι τα τμήματα που βρίσκονται στις υψηλότερες θέσεις στον πίνακα κατάταξης είναι περισσότερο προσανατολισμένα στην παραγωγή ερευνητικού έργου με αντίκτυπο στην επιστημονική κοινότητα, και αφορούν κυρίως τμήματα που ιδρύθηκαν παλαιότερα χρονολογικά, διαθέτουν σημαντικό αριθμό μελών Δ.Ε.Π. και βρίσκονται σε μεγάλες πόλεις της Ελλάδας. Αντίθετα, τα τμήματα με τη χαμηλότερη παραγωγή και απήχηση του ερευνητικού τους έργου, αφορούν τμήματα που μετατράπηκαν πρόσφατα από Τ.Ε.Ι. σε Α.Ε.Ι, βρίσκονται σε περιφερειακές πόλεις της Ελλάδας και διαθέτουν συνήθως μικρό αριθμό διδακτικού προσωπικού. Ωστόσο, θα πρέπει να τονιστεί ότι σημαντικό παράγοντα παραγωγής ερευνητικού έργου των πανεπιστημίων αποτελεί η διάθεση κονδυλίων και η λήψη χρηματοδοτήσεων για ερευνητικούς σκοπούς. Σε αυτές τις περιπτώσεις, ίσως να ευνοούνται περισσότερο τα παλαιότερα χρονολογικά τμήματα και τα μεγαλύτερα πανεπιστήμια της χώρας, καθώς εμφανίζουν μεγαλύτερο κύρος και διαθέτουν περισσότερο προσωπικό, που ενδεχομένως

προσανατολίζεται περισσότερο σε τέτοιου είδους χρηματοδοτήσεις και στη δημοσίευση άρθρων σε έγκριτα διεθνή επιστημονικά περιοδικά.

Στο δεύτερο μέρος της έρευνας, υλοποιήθηκε η εξόρυξη γνώσης, με την εφαρμογή αλγορίθμων κατηγοριοποίησης και συσταδοποίησης. Χρησιμοποιήθηκε το λογισμικό Weka και η εκτέλεση των εργασιών έγινε στο περιβάλλον του Explorer. Πριν την εκτέλεση των αλγορίθμων τα αρχεία μετατράπηκαν σε αρχεία ARFF και ακολούθησε η προεπεξεργασία των δεδομένων, ώστε αυτά να είναι όσο το δυνατόν πιο αξιόπιστα. Στη συνέχεια εκτελέστηκαν αλγόριθμοι κατηγοριοποίησης, που περιλαμβάνουν τον αλγόριθμο J48 για τη δημιουργία δέντρου απόφασης, τον Naïve Bayes και τον IBk για την κατηγοριοποίηση του πλησιέστερου γείτονα. Επιπλέον, για την τεχνική της συσταδοποίησης εκτελέστηκε ο αλγόριθμος K-Means. Καθώς κατά την εκτέλεση των αλγορίθμων κατηγοριοποίησης το σύνολο των δεδομένων δεν ήταν ισορροπημένο, προχωρήσαμε στη χρήση της μεθόδου SMOTE ώστε αυτά να ισοσταθμιστούν.

Η εκτέλεση του κατηγοριοποιητή Naïve Bayes έγινε με στόχο την πρόβλεψη των μελών Δ.Ε.Π. σε ακαδημαϊκή βαθμίδα. Ο αλγόριθμος παρουσίασε χαμηλή απόδοση με ποσοστό περίπου 50%. Με τη χρήση της μεθόδου SMOTE το ποσοστό επιτυχίας δεν μεταβλήθηκε πολύ και έφτασε στο 56%. Ωστόσο, αξίζει να σημειωθεί ότι η κατηγοριοποίηση της βαθμίδας του Λέκτορα και των ΕΔΙΠ βελτιώθηκε αρκετά.

Ο αλγόριθμος J48 εμφάνισε μέτρια απόδοση στην πρόβλεψη των μελών Δ.Ε.Π. περιλαμβάνοντας όλες τις μεταβλητές (56,6%), και είχε ως αποτέλεσμα την απεικόνιση ενός μικρού δέντρου απόφασης. Το μοντέλο βελτιώθηκε κατά πολύ με τη χρήση της μεθόδου SMOTE όπου το ποσοστό επιτυχούς κατηγοριοποίησης έφτασε το 71,2%. Επιπλέον παρουσίασαν μεγάλη βελτίωση οι τιμές Precision και Recall όλων των κλάσεων.

Για τον αλγόριθμο IBk, ορίστηκε ως καλύτερη τιμή πλησιέστερων γειτόνων $k=4$. Και σε αυτήν την περίπτωση εξετάστηκε η πρόβλεψη των μελών Δ.Ε.Π. σε ακαδημαϊκή βαθμίδα. Αρχικά ο αλγόριθμος είχε χαμηλή απόδοση (58%), ενώ με την εφαρμογή της μεθόδου SMOTE έδωσε αρκετά υψηλό ποσοστό πρόβλεψης (72%). Και σ' αυτή την περίπτωση οι τιμές Precision και Recall όλων των κλάσεων παρουσίασαν μεγάλη βελτίωση.

Για την ομαδοποίηση των δεδομένων με τον αλγόριθμο K-means, η επιλογή του αριθμού των συστάδων βασίστηκε στη μέθοδο του αγκώνα και η πιο ικανοποιητική τιμή ορίστηκε για $k=3$. Ο υπολογισμός της απόστασης έγινε με την ευκλείδεια μετρική και ως αποτέλεσμα διαχωρίστηκαν τρεις συστάδες. Αυτές περιλαμβάνουν τις βαθμίδες του Καθηγητή (43%), του Αναπληρωτή καθηγητή (28%) και του Επίκουρου Καθηγητή (30%).

Σύμφωνα με τα παραπάνω αποτελέσματα γίνεται αντιληπτό ότι οι αλγόριθμοι κατηγοριοποίησης έχουν χαμηλή απόδοση πρόβλεψης για τις κλάσεις των μελών Δ.Ε.Π. ανά ακαδημαϊκή βαθμίδα (P), με τον αλγόριθμο IBk να δίνει την καλύτερη πρόβλεψη -αν και χαμηλή- σε σχέση με τους υπόλοιπους. Ωστόσο, η εφαρμογή της μεθόδου SMOTE βελτιώνει κατά πολύ την επίδοση πρόβλεψης των αλγορίθμων με εξαίρεση τον αλγόριθμο Naïvy Bayes που κυμαίνεται στα ίδια επίπεδα. Όσον αφορά την ομαδοποίηση των δεδομένων, διαπιστώνεται ότι κατά την εκτέλεση του αλγόριθμου την πρώτη φορά προέκυψαν συστάδες με κοινά χαρακτηριστικά ως προς το ίδρυμα. Για το λόγο αυτό προχωρήσαμε σε δεύτερη εκτέλεση του αλγόριθμου αφαιρώντας κάποια χαρακτηριστικά όπως τη βαθμίδα (position), το ίδρυμα (dept) και την περιοχή (area) προκειμένου να κρατήσουμε τα ποιοτικά χαρακτηριστικά. Οι συστάδες που προέκυψαν ομαδοποίησαν τα μέλη ΔΕΠ σύμφωνα με τις υψηλές ερευνητικές τους επιδόσεις, το έργο της τελευταίας πενταετίας, τα χρόνια δημοσιεύσεων τους, τις αναφορές τους και τον h index.

6.2 Μελλοντική Έρευνα

Η παρούσα έρευνα κινήθηκε σε συγκεκριμένο μεθοδολογικό πλαίσιο και έκανε χρήση των ευρέως χρησιμοποιούμενων μοντέλων για τη στατιστική ανάλυση και την εξόρυξη γνώσης, όπως αυτά υλοποιήθηκαν στο πλαίσιο της μεταπτυχιακής εργασίας. Στο σημείο αυτό προτείνονται κάποιες μελλοντικές προεκτάσεις, που θα μπορούσαν να συμβάλλουν σε μια πιο ολοκληρωμένη εικόνα των δεδομένων και της γνώσης που μπορεί να εξαχθεί. Έτσι, ξεκινώντας από το κομμάτι της επιλογής των επιστημονομετρικών δεικτών, αυτοί μελλοντικά θα μπορούσαν να εμπλουτιστούν, ώστε να μας προσφέρουν μεγαλύτερο εύρος συγκρίσεων μεταξύ των ακαδημαϊκών τμημάτων και των μελών Δ.Ε.Π. Λαμβάνοντας υπόψη περισσότερα στοιχεία και παραμέτρους, θα μπορούσαμε να οδηγηθούμε σε πιο ακριβή και πολύπλευρα συμπεράσματα. Επιπλέον, όσον αφορά την εξόρυξη γνώσης, ο εμπλουτισμός και η χρήση του συνόλου των δεδομένων στην εκτέλεση των αλγορίθμων, ενδεχομένως να προσφέρει καλύτερα και πιο ακριβή αποτελέσματα. Ταυτόχρονα, η χρήση και άλλων τεχνικών εξόρυξης γνώσης, αλλά και η εφαρμογή περισσότερων και πιο πολύπλοκων αλγορίθμων, θα μπορούσε να συμβάλει στην ανάπτυξη πιο επιτυχημένων μοντέλων πρόβλεψης και ομαδοποίησης.

Βιβλιογραφία

- [1] Holmes, A. & Oppenheim, C. (2001). Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61: Library and information management. *Information Research*, 6 (2). Available at: <http://InformationR.net/ir/6-2/paper103.html>
- [2] Martin, B.R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, 36, 343-362.
- [3] Wallin, J.A. (2005). Bibliometric methods: Pitfalls and possibilities. *Basic and Clinical Pharmacology and Toxicology*, 97, 261-275.
- [4] Nalimov, V.V. and Mulchenko, Z.M. (1969). *Naukometriya. Izuchenie razvitiya nauki kak informacionnogo processa (The study of the development of science as information process)*. Moscow.
- [5] Σίτας, Α., Κώνστα, Ο., & Ντούλια, Π. (2007). *Βιβλιομετρική - Πληροφοριομετρική - Ιστομετρική*. Διαθέσιμο στο: http://eprints.rclis.org/14652/1/Sitas-Konsta-Doulia_Bibliometriki.pdf.
- [6] Παπαβλασόπουλος Σ. (2015). *Βιβλιομετρία*. Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα. Διαθέσιμο στο: www.kallipos.gr
- [7] Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentarion*, 25, 348-349.
- [8] Wilson, C. S. (1999). Informetrics. *Annual Review of Information Science and Technology*, 34. Available at: <https://doi.org/10.29085/9781783300761.015>
- [9] Garfield, E. (1955). Citation Indexes to Science: A New Dimension in Documentation through Association of Ideas. *Science*, 122, 108-111.
- [10] Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, USA*, 102(46), 16569–16572.
- [11] Glänzel, W. (2006). On the opportunities and limitations of the H-index. *Science Focus*, 1(1), 10–11.
- [12] Τσέλιος, Ν. και Αλτανοπούλου, Π. (2011). Αξιολόγηση Ελληνικών Πανεπιστημιακών τμημάτων Εκπαίδευσης με τη χρήση του Google Scholar και του δείκτη h. *2ο Πανελλήνιο συνέδριο «Ένταξη και χρήση των ΤΠΕ στην εκπαιδευτική διαδικασία»*, Πάτρα, 28-30 Απριλίου 2011, 867-876.
- [13] Hand, D., Mannila, H. and Smyth, P. (2001). *Principles of data mining*. The MIT press.

- [14] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *Artificial Intelligence Magazine*, 17(3), 37-54. doi: 10.1609/aimag.v17i3.1230
- [15] Maimon, O., & Rokach, L. (2005). *The Data Mining and Knowledge Discovery Handbook*. New York, NY: Springer Science, Business Media Inc.
- [16] Han, J., Pei, J. and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- [17] Κύρκος, Ε. 2015. Επιχειρηματική ευφυΐα και εξόρυξη δεδομένων. Αθήνα: Κάλλιπος, Ανοιχτές Ακαδημαϊκές Εκδόσεις. Διαθέσιμο στο: <http://hdl.handle.net/11419/1226>
- [18] Tan, P.N., Steinbach, M. and Kumar, V. (2017). *Εισαγωγή στην εξόρυξη δεδομένων*. Εκδόσεις Τζιόλας.
- [19] Olson, D. and Delen, D. (2008). *Advanced Data Mining Techniques*. Springer Verlag [17] Altman, N.S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*, 46 (3), 175-185.
- [20] Kotsiantis, S. B., Zaharakis, I., Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160 (1), 3–24.
- [21] Νανόπουλος, Α., και Μανωλόπουλος, Ι. (2008). *Εισαγωγή στην Εξόρυξη και τις Αποθήκες Δεδομένων*. Αθήνα, Εκδόσεις Νέων Τεχνολογιών.
- [22] Χαλκίδη, Μ., και Βεζυργιάννης, Μ. (2005). *Εξόρυξη Γνώσης από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό*. Αθήνα, Τυπωθήτω.
- [23] Quinlan, J.R. (1986). Induction of Decision Trees, *Machine Learning*, 1, 81-106.
- [24] Altman, N. S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*, 46 (3), 175-185.
- [25] Thirumuruganathan, S. (2010). A Detailed Introduction to K-Nearest Neighbor (KNN) Algorithm. Available at: <https://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/>
- [26] Kodinariya, T. M. and Makwana, P. R. (2013). “Review on Determining Number of Cluster in K-Means Clustering”, *International Journal of Advance Research in Computer Science and Management Studies*, Vol. 1, No. 6. ISSN 2321 - 7782.
- [27] Jain, A. K. (2009). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31 (8), 651-666.

[28] Kirkby, R. and Frank, E. (2006). WEKA Explorer User Guide for Version 3-4. University of Waikato. Available at: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.169.8280&rep=rep1&type=pdf>