



**ΣΧΟΛΗ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

*Forecasting the outcome of Greek football games using
Mathematical Models and Power Rankings*

ΔΗΜΗΤΡΑ ΠΑΛΙΑΤΣΑ

**ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:
ΠΟΛΥΧΡΟΝΗΣ ΚΟΥΤΣΑΚΗΣ**

ΧΑΝΙΑ, ΣΕΠΤΕΜΒΡΙΟΣ 2014

Περίληψη

Το αντικείμενο της παρούσας διπλωματικής εργασίας είναι η πρόβλεψη των αποτελεσμάτων των αγώνων ποδοσφαίρου και η εκτίμηση της τελικής βαθμολογίας των ομάδων. Ειδικότερα, μας ενδιαφέρει η πρόβλεψη των αποτελεσμάτων των αγώνων του ελληνικού πρωταθλήματος της Α' κατηγορίας ανδρών για τις σεζόν 2007-2008 έως 2013-2014. Για τον σκοπό αυτό κρίθηκε απαραίτητη η καταγραφή των αναλυτικών αποτελεσμάτων ανά αγωνιστική για όλες τις ομάδες (διαφορά τερμάτων, τρέχουσα βαθμολογία, αριθμός νικών, ισοπαλιών, ηττών κ.α). Στη συνέχεια χρησιμοποιώντας τα δεδομένα που συγκεντρώθηκαν και με τη χρήση διαφόρων τεχνικών προσπαθούμε να προβλέψουμε τα τελικά αποτελέσματα των αγώνων. Η πρώτη προσέγγισή μας εντάσσεται στα συστήματα κατάταξης (sport rating systems), όπου κατατάσσουμε τις ομάδες βάσει της δύναμης που κατέχουν (power ranking). Η δεύτερη προσέγγισή μας στηρίζεται στην συσταδοποίηση και υλοποιείται με την εφαρμογή του αλγορίθμου k-means clustering. Στην τρίτη προσέγγισή μας γίνεται χρήση δύο ανεξάρτητων κατανομών Poisson για την εκτίμηση των γκολ που σημειώνει η γηπεδούχος και η φιλοξενούμενη ομάδα. Επιπλέον, χρησιμοποιούμε δύο απλοϊκές μεθόδους για την εξαγωγή baseline αποτελεσμάτων, όπου στην πρώτη θεωρούμε πως θα κερδίζει πάντα η γηπεδούχος ομάδα και στη δεύτερη θεωρούμε πως θα επαναληφθούν τα αποτελέσματα της προηγούμενης χρονιάς για τις ομάδες που αναμετρώνται για δύο συνεχόμενα έτη. Τέλος, όσον αφορά στην πρόβλεψη της τελικής βαθμολογίας των ομάδων, εκτός από τις μεθόδους που αναφέρθηκαν, χρησιμοποιούμε και τη μέθοδο του κινητού μέσου, του σταθμισμένου κινητού μέσου και της απλής εκθετικής εξομάλυνσης. Τα αποτελέσματα που προκύπτουν είναι αρκετά ελπιδοφόρα, γεγονός που αποδεικνύει τη χρησιμότητα των υπό μελέτη μεθόδων και καθιστά την εξέταση του συγκεκριμένου προβλήματος ενδιαφέρουσα για μελλοντικές βελτιώσεις.

Abstract

The subject of this thesis is the prediction of the results of football matches and the estimation of the final points gathered by the teams at the end of the football year. In particular, we are interested in the prediction of the outcome of the Greek Superleague games for the season 2007-2008 to 2013-2014. For this purpose it was necessary to record the relevant data for all teams after each game (goal difference, current points, number of wins, number of draws, number of defeats, etc.). Then, using the data gathered and utilizing various techniques we try to predict the final outcome of the matches. Our first approach falls into the sport rating systems category, where we rank all the teams based on their power scores (power ranking procedure). Our second approach is based on clustering and is implemented using the k-means algorithm. In the third approach we use two independent Poisson distributions for describing the goals scored by the home and the away team, respectively. Moreover, we use two naive methods for extracting baseline results: in

the first we assume that the home team is always the winner and in the second we assume that the last year's result will be repeated for all teams that compete in two consecutive years. Finally, we use the moving average, the weighted moving average and the exponentially weighted moving average to estimate the final points that each Superleague team will gather at the end of the football year. The results of our study are promising, demonstrating the usefulness of the proposed models and making the further study of this problem interesting for future improvements.

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα της διπλωματικής μου εργασίας κ. Πολυχρόνη Κουτσάκη για την εμπιστοσύνη και την αμέριστη συμπαράστασή του καθ' όλη τη διάρκεια της εκπόνησής της. Επίσης, θα ήθελα να ευχαριστήσω την οικογένειά μου και ειδικότερα τους γονείς μου Δαμιανό και Γεωργία και τον αδερφό μου Παναγιώτη, καθώς επίσης και τις φίλες μου για όλη την συμπαράσταση και την υποστήριξη που μου δίνουν όλα αυτά τα χρόνια.

Περιεχόμενα

Περίληψη	3
Abstract	3
Ευχαριστίες	5
Λίστα Πινάκων	8
Λίστα Σχημάτων	10
1. Εισαγωγή.....	16
1.1 Περιγραφή του προβλήματος	17
1.2 Στόχοι της εργασίας	18
1.3 Δεδομένα και παραδοχές.....	18
1.4 Κριτήρια σύγκρισης μεθόδων	18
1.5 Οργάνωση της εργασίας	20
2. Συστήματα Κατάταξης.....	21
2.1 Βασικές έννοιες και εφαρμογές.....	21
2.2 Τα κυριότερα συστήματα κατάταξης.....	21
2.2.1 Το επίσημο σύστημα κατάταξης της FIFA (<i>The official FIFA ranking</i>)	22
2.2.2 Το σύστημα κατάταξης Elo (<i>The Elo rating system</i>)	23
2.2.3 Το επίσημο σύστημα κατάταξης της FIFA για το παγκόσμιο πρωτάθλημα γυναικών (<i>FIFA Women's World Rankings</i>)	25
2.2.4 EloRatings.net	26
2.2.5 Jeff Sagarin.....	26
2.2.6 RPI.....	27
2.3 Περιγραφή της μεθόδου υλοποίησης.....	28
2.4 Αποτελέσματα	28
2.5 Αποτελέσματα βάσει της διαφοράς θέσεων κατάταξης των ομάδων	41
2.6 Πλεονεκτήματα/Μειονεκτήματα των συστημάτων κατάταξης.....	48
3. Συσταδοποίηση	49
3.1 Εισαγωγή στην Συσταδοποίηση.....	49
3.2 Βήματα Συσταδοποίησης	50
3.3 Διαμεριστική Συσταδοποίηση (Partitional clustering)	52
3.4 Ο αλγόριθμος k-means.....	52
3.4.1 Επιλογή των χαρακτηριστικών	53
3.4.2 Επιλογή του μέτρου ομοιότητας και του αριθμού των ομάδων	54
3.5 Περιγραφή της μεθόδου υλοποίησης.....	55

3.5.1 Εφαρμογή για k=3 clusters.....	55
3.5.2 Εφαρμογή για k=4 clusters.....	56
3.5.3 Εφαρμογή για k=5 clusters.....	57
3.5.4 Εφαρμογή για k=6 clusters.....	58
3.6 Αποτελέσματα	60
4. Στατιστική εκτίμηση του Σκορ.....	62
4.1 Κατανομή Poisson και Συντελεστής συσχέτισης Pearson.....	62
4.2 Περιγραφή της μεθόδου υλοποίησης.....	63
4.3 Αποτελέσματα	64
5. Απλά μοντέλα πρόβλεψης για Baseline αποτελέσματα	69
5.1 Το πλεονέκτημα της έδρας.....	69
5.1.1 Η επίδραση των φιλάθλων	69
5.1.2 Η επίδραση του ταξιδιού.....	69
5.1.3 Η εξοικείωση	69
5.1.4 Η προκατάληψη του διαιτητή	70
5.1.5 Η ειδική τακτική	70
5.1.6 Οι ψυχολογικοί παράγοντες	70
5.1.7 Άλλες αιτίες	70
5.2 Προβλέψεις βάσει της έδρας (εντός/εκτός)	71
5.3 Προβλέψεις βάσει των αποτελεσμάτων της προηγούμενης σεζόν	71
5.4 Αποτελέσματα	73
6. Μέθοδοι εξομάλυνσης.....	75
6.1 Απλός Κινητός Μέσος (MA).....	75
6.2 Σταθμισμένος Κινητός Μέσος (WMA)	78
6.3 Απλή Εκθετική Εξομάλυνση (EWMA).....	79
6.4 Αποτελέσματα	82
7. Συμπεράσματα και επεκτάσεις.....	84
7.1 Συμπεράσματα	84
7.2 Επεκτάσεις.....	85
8. Παράρτημα	86
Βιβλιογραφία.....	135

Λίστα Πινάκων

Πίνακας 2-1 : Η βαθμολογία s_A που δίνεται στην ηττημένη ομάδα A	25
Πίνακας 2-2 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας	29
Πίνακας 2-3 : Ποσοστό ευστοχίας	29
Πίνακας 2-4: Ποσοστό σφάλματος ισοπαλιών	29
Πίνακας 2-5: Ποσοστό σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη	29
Πίνακας 2-6 : Ποσοστό ευστοχίας στις προβλέψεις μας βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2007-8 έως 2010-11)	41
Πίνακας 2-7 : Ποσοστό ευστοχίας στις προβλέψεις μας βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2011-12 έως 2013-14)	42
Πίνακας 2-8 : Ποσοστό αστοχίας στις ισοπαλίες βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2007-8 έως 2010-11)	43
Πίνακας 2-9 : Ποσοστό αστοχίας στις ισοπαλίες βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2011-12 έως 2013-14)	43
Πίνακας 2-10 : Ποσοστό αστοχίας στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη, βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2007-8 έως 2010-11)	44
Πίνακας 2-11 : Ποσοστό αστοχίας στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη, βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2011-12 έως 2013-14)	45
Πίνακας 3-1 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας	60
Πίνακας 3-2 : Ποσοστό ευστοχίας	60
Πίνακας 3-3: Ποσοστό σφάλματος ισοπαλιών	61
Πίνακας 3-4: Ποσοστό σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη	61
Πίνακας 4-1 : Συντελεστής συσχέτισης Pearson	63
Πίνακας 4-2 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας	65
Πίνακας 4-3 : Ποσοστό ευστοχίας	65
Πίνακας 4-4 : Ποσοστό σφάλματος ισοπαλιών	65
Πίνακας 4-5 : Ποσοστό σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη	66
Πίνακας 4-6 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2007-2008)	66
Πίνακας 4-7 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2008-2009)	66

Πίνακας 4-8 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2009-2010).....	67
Πίνακας 4-9 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2010-2011).....	67
Πίνακας 4-10 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2011-2012).....	67
Πίνακας 4-11 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2012-2013).....	68
Πίνακας 4-12 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2013-2014).....	68
Πίνακας 5-1 : Ομάδες που συμμετέχουν στην Superleague τις σεζόν 2007-2008 έως 2010-2011	72
Πίνακας 5-2 : Ομάδες που συμμετέχουν στην Superleague τις σεζόν 2011-2012 έως 2013-2014	72
Πίνακας 5-3 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας	73
Πίνακας 5-4 : Ποσοστό ευστοχίας.....	73
Πίνακας 5-5 : Ποσοστό σφάλματος ισοπαλιών	73
Πίνακας 5-6 : Ποσοστό σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη	73
Πίνακας 6-1 : Μέθοδος Απλού Κινητού Μέσου	77
Πίνακας 6-2 : Μέθοδος Σταθμισμένου Κινητού Μέσου	79
Πίνακας 6-3 : Μέθοδος Απλής Εκθετικής Εξομάλυνσης.....	81
Πίνακας 6-4 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας	82
Πίνακας 6-5 : Ομάδες που συμμετέχουν στην Superleague τις σεζόν 2007-2008 έως 2010-2011	83
Πίνακας 6-6 : Ομάδες που συμμετέχουν στην Superleague τις σεζόν 2011-2012 έως 2013-2014	83

Λίστα Σχημάτων

Εικόνα 2-1 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική	30
Εικόνα 2-2 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική	31
Εικόνα 2-3 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική.....	31
Εικόνα 2-4 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική ..	32
Εικόνα 2-5 : Οι τιμές του RPI Weighted της Λάρισας και των αντιπάλων της ανά αγωνιστική	33
Εικόνα 2-6 : Οι τιμές του RPI Weighted της ΑΕΚ και των αντιπάλων της ανά αγωνιστική	33
Εικόνα 2-7 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική	34
Εικόνα 2-8 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική	35
Εικόνα 2-9 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική.....	35
Εικόνα 2-10 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική	36
Εικόνα 2-11 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική.....	37
Εικόνα 2-12 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική	37
Εικόνα 2-13 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική.....	38
Εικόνα 2-14 : Οι τιμές του RPI Weighted του Ηρακλή και των αντιπάλων του ανά αγωνιστική ..	39
Εικόνα 2-15 : Οι τιμές του RPI Weighted της Βέροιας και των αντιπάλων της ανά αγωνιστική ...	39
Εικόνα 2-16 : Οι τιμές του RPI Weighted του Απόλλωνα Καλαμαριάς και των αντιπάλων του ανά αγωνιστική.....	40
Εικόνα 2-17 : Μέσο ποσοστό ευστοχίας στη διάρκεια των 7 εξεταζόμενων ετών	46
Εικόνα 2-18 : Μέσο ποσοστό αστοχίας, στους αγώνες όπου η εκτιμώμενη νικήτρια ομάδα ήταν η ηττημένη, στη διάρκεια των 7 εξεταζόμενων ετών	47
Εικόνα 3-1 : Ομαδοποίηση δεδομένων	50
Εικόνα 3-2 : Βήματα εκτέλεσης αλγορίθμου K-means	53
Εικόνα 8-1 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική	86
Εικόνα 8-2 : Οι τιμές του RPI Weighted της Λάρισας και των αντιπάλων της ανά αγωνιστική	86
Εικόνα 8-3 : Οι τιμές του RPI Weighted της ΑΕΚ και των αντιπάλων της ανά αγωνιστική	87

Εικόνα 8-4 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική.....	87
Εικόνα 8-5 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική..	88
Εικόνα 8-6 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική	88
Εικόνα 8-7 : Οι τιμές του RPI Weighted του Πανσερραϊκού και των αντιπάλων του ανά αγωνιστική.....	889
Εικόνα 8-8 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική.....	889
Εικόνα 8-9 : Οι τιμές του RPI Weighted του Ηρακλή και των αντιπάλων του ανά αγωνιστική	90
Εικόνα 8-10 : Οι τιμές του RPI Weighted του Αστήρα Τρίπολης και των αντιπάλων του ανά αγωνιστική.....	90
Εικόνα 8-11 : Οι τιμές του RPI Weighted του Πανθρακικού και των αντιπάλων του ανά αγωνιστική.....	91
Εικόνα 8-12 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική.....	91
Εικόνα 8-13 : Οι τιμές του RPI Weighted του Θρασύβουλου και των αντιπάλων του ανά αγωνιστική.....	92
Εικόνα 8-14 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική.....	92
Εικόνα 8-15 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική.....	93
Εικόνα 8-16 : Οι τιμές του RPI Weighted του Πανιώνιου και των αντιπάλων του ανά αγωνιστική	93
Εικόνα 8-17 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική.....	94
Εικόνα 8-18 : Οι τιμές του RPI Weighted του Ηρακλή και των αντιπάλων του ανά αγωνιστική..	94
Εικόνα 8-19 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική.....	95
Εικόνα 8-20 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική.....	95
Εικόνα 8-21 : Οι τιμές του RPI Weighted της Καβάλας και των αντιπάλων της ανά αγωνιστική .	96
Εικόνα 8-22 : Οι τιμές του RPI Weighted του Πας Γιάννενα και των αντιπάλων του ανά αγωνιστική.....	96
Εικόνα 8-23 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική.....	97
Εικόνα 8-24 : Οι τιμές του RPI Weighted της Λάρισας και των αντιπάλων της ανά αγωνιστική .	97
Εικόνα 8-25 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική	98

Εικόνα 8-26 : Οι τιμές του RPI Weighted του Πανιώνιου και των αντιπάλων του ανά αγωνιστική	98
Εικόνα 8-27 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική	99
Εικόνα 8-28 : Οι τιμές του RPI Weighted της ΑΕΚ και των αντιπάλων της ανά αγωνιστική	99
Εικόνα 8-29 : Οι τιμές του RPI Weighted του Πανθρακικού και των αντιπάλων του ανά αγωνιστική	100
Εικόνα 8-30 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική	100
Εικόνα 8-31 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική	101
Εικόνα 8-32 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική	101
Εικόνα 8-33 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική ...	102
Εικόνα 8-34 : Οι τιμές του RPI Weighted του Ηρακλή και των αντιπάλων του ανά αγωνιστική	102
Εικόνα 8-35 : Οι τιμές του RPI Weighted της Κέρκυρας και των αντιπάλων της ανά αγωνιστική	103
Εικόνα 8-36 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική	103
Εικόνα 8-37 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική	104
Εικόνα 8-38 : Οι τιμές του RPI Weighted του Ολυμπιακού Βόλου και των αντιπάλων του ανά αγωνιστική	104
Εικόνα 8-39 : Οι τιμές του RPI Weighted της Λάρισας και των αντιπάλων της ανά αγωνιστική	105
Εικόνα 8-40 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική	105
Εικόνα 8-41 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική	106
Εικόνα 8-42 : Οι τιμές του RPI Weighted του Πανσερραϊκού και των αντιπάλων του ανά αγωνιστική	106
Εικόνα 8-43 : Οι τιμές του RPI Weighted της ΑΕΚ και των αντιπάλων της ανά αγωνιστική	107
Εικόνα 8-44 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική	107
Εικόνα 8-45 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική	108
Εικόνα 8-46 : Οι τιμές του RPI Weighted της Καβάλας και των αντιπάλων της ανά αγωνιστική	108

Εικόνα 8-47 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική	109
Εικόνα 8-48 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική	109
Εικόνα 8-49 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική.....	110
Εικόνα 8-50 : Οι τιμές του RPI Weighted της ΑΕΚ και των αντιπάλων της ανά αγωνιστική	110
Εικόνα 8-51 : Οι τιμές του RPI Weighted του Παναιτωλικού και των αντιπάλων του ανά αγωνιστική.....	111
Εικόνα 8-52 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική ...	111
Εικόνα 8-53 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική	112
Εικόνα 8-54 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική	112
Εικόνα 8-55 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική	113
Εικόνα 8-56 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική.....	113
Εικόνα 8-57 : Οι τιμές του RPI Weighted του Πας Γιάννενα και των αντιπάλων του ανά αγωνιστική.....	114
Εικόνα 8-58 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική.....	114
Εικόνα 8-59 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική.....	115
Εικόνα 8-60 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική.....	115
Εικόνα 8-61 : Οι τιμές του RPI Weighted της Δόξας Δράμας και των αντιπάλων της ανά αγωνιστική.....	116
Εικόνα 8-62 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική.....	116
Εικόνα 8-63 : Οι τιμές του RPI Weighted της Κέρκυρας και των αντιπάλων της ανά αγωνιστική	117
Εικόνα 8-64 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική.....	117
Εικόνα 8-65 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική.....	118
Εικόνα 8-66 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική	118

Εικόνα 8-67 : Οι τιμές του RPI Weighted του Πανιώνιου και των αντιπάλων του ανά αγωνιστική	119
Εικόνα 8-68 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική ...	119
Εικόνα 8-69 : Οι τιμές του RPI Weighted του Αστέρρα Τρίπολης και των αντιπάλων του ανά αγωνιστική.....	120
Εικόνα 8-70 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική.....	120
Εικόνα 8-71 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική.....	121
Εικόνα 8-72 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική.....	121
Εικόνα 8-73 : Οι τιμές του RPI Weighted του Πλατανιά Χανίων και των αντιπάλων του ανά αγωνιστική.....	122
Εικόνα 8-74 : Οι τιμές του RPI Weighted του Πας Γιάννενα και των αντιπάλων του ανά αγωνιστική.....	122
Εικόνα 8-75 : Οι τιμές του RPI Weighted της Βέροιας και των αντιπάλων της ανά αγωνιστική	123
Εικόνα 8-76 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική.....	123
Εικόνα 8-77 : Οι τιμές του RPI Weighted του Πανθρακικού και των αντιπάλων του ανά αγωνιστική.....	124
Εικόνα 8-78 : Οι τιμές του RPI Weighted της Κέρκυρας και των αντιπάλων της ανά αγωνιστική	124
Εικόνα 8-79 : Οι τιμές του RPI Weighted της ΑΕΚ και των αντιπάλων της ανά αγωνιστική	125
Εικόνα 8-80 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική.....	125
Εικόνα 8-81 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική ...	126
Εικόνα 8-82 : Οι τιμές του RPI Weighted του Απόλλωνα Σμύρνης και των αντιπάλων του ανά αγωνιστική.....	126
Εικόνα 8-83 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική.....	127
Εικόνα 8-84 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική.....	127
Εικόνα 8-85 : Οι τιμές του RPI Weighted του Πας Γιάννενα και των αντιπάλων του ανά αγωνιστική.....	128
Εικόνα 8-86: Οι τιμές του RPI Weighted του Αστέρρα Τρίπολης και των αντιπάλων του ανά αγωνιστική.....	128
Εικόνα 8-87 : Οι τιμές του RPI Weighted του Πλατανιά Χανίων και των αντιπάλων του ανά αγωνιστική.....	129

Εικόνα 8-88 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική	129
Εικόνα 8-89 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική.....	130
Εικόνα 8-90 : Οι τιμές του RPI Weighted του Πανθρακικού και των αντιπάλων του ανά αγωνιστική.....	130
Εικόνα 8-91 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική	131
Εικόνα 8-92 : Οι τιμές του RPI Weighted της Βέροιας και των αντιπάλων της ανά αγωνιστική	131
Εικόνα 8-93 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική	132
Εικόνα 8-94 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική.....	132
Εικόνα 8-95 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική.....	133
Εικόνα 8-96 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική	133
Εικόνα 8-97 : Οι τιμές του RPI Weighted του Παναιτωλικού και των αντιπάλων του ανά αγωνιστική.....	134
Εικόνα 8-98 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική.....	134

1. Εισαγωγή

Ο άνθρωπος από τη φύση του επιθυμεί να μαθαίνει. Ένα από τα κυριότερα χαρακτηριστικά του ανθρώπου, το οποίο έχει παραμείνει άτρωτο στο χρόνο είναι η ανάγκη του για γνώση. Μέσα στο πλαίσιο αυτό εντάσσεται και η ακόρεστη περιέργεια και προσπάθειά του για πρόβλεψη μελλοντικών φαινομένων και συνθηκών. Ήδη από την αρχαία εποχή, ο άνθρωπος μέσω των ουράνιων καταγραφών και των κοσμολογικών παρατηρήσεων προσπαθούσε να βγάλει συμπεράσματα για τις μελλοντικές πιθανές καταστροφές και τις καιρικές συνθήκες που θα επικρατούσαν, γνώση που του ήταν απαραίτητη για τα μακρινά εμπορικά και εξερευνητικά ταξίδια. Επιπλέον, στην αρχαιότητα ήταν δημοφιλή και σεβαστά από όλους τα μαντεία, τα οποία επηρέαζαν το λαό στην καθημερινή του ζωή και τις πόλεις στις ιστορικές τους τύχες. Κατά την κλασική αρχαιότητα, το μαντείο ήταν ο επίσημος, θρησκευτικού χαρακτήρα θεσμός σε συγκεκριμένο τόπο, όπου οι μάντεις εμπνέονταν σοφές συμβουλές, έδιναν χρησμούς, ή προέβλεπαν το μέλλον ως ενδιάμεσοι με τους θεούς. Μάλιστα, ακόμα και οι ηγέτες της εποχής τα συμβουλευόνταν πριν την λήψη μεγάλων αποφάσεων, όπως η έναρξη ενός πολέμου.

Σήμερα, στον αιώνα των ραγδαίων τεχνολογικών και επιστημονικών εξελίξεων και επιτευγμάτων η ανάγκη του ανθρώπου για γνώση και για πρόβλεψη μελλοντικών καταστάσεων παραμένει αναλλοίωτη. Οι κυριότεροι λόγοι αφορούν τη διαβίωσή του και ειδικότερα την αποφυγή καταστροφικών φαινομένων, όπως σεισμοί και πλημμύρες, μέσω της πρόβλεψης καιρικών συνθηκών και την πρόληψη ενάντια σε ασθένειες με την εκτίμηση της πιθανότητας εμφάνισης τους σε κάποιον ασθενή. Επίσης, οι λόγοι είναι και οικονομικοί και αφορούν την εμφάνιση οικονομικού κέρδους μέσω της πρόβλεψης των τιμών του χρηματιστηρίου ή της κατάστασης της αγοράς και της έκβασης των αγώνων σε πλήθος αθλημάτων.

Έχοντας λοιπόν απομυθοποιήσει το θεσμό των μαντείων και των χρησμών και μέσω της παρατήρησης και της μελέτης των παρελθοντικών συνθηκών και αποτελεσμάτων ο άνθρωπος προσπαθεί να κάνει εκτιμήσεις για το μέλλον. Τα περισσότερα πραγματικά φαινόμενα είναι αρκετά σύνθετα και πολύπλοκα, ώστε να μπορέσουν να μοντελοποιηθούν και να περιγραφούν από έναν απλό μαθηματικό τύπο. Παρ' όλα αυτά, όσο τυχαία και χαώδη κι αν φαίνονται, παρατηρούνται σε αυτά κάποια μοτίβα συμπεριφορών, μέσω των οποίων μπορούν να γίνουν εύστοχες εκτιμήσεις.

Στην παρούσα εργασία, χρησιμοποιούμε κάποιες μεθόδους, οι οποίες αναλύονται στα επόμενα κεφάλαια, μέσω των οποίων κάνουμε εκτιμήσεις για την τελική έκβαση ενός αγώνα ποδοσφαίρου.

1.1 Περιγραφή του προβλήματος

Στις μέρες μας, ο αθλητισμός είναι μία από τις μεγαλύτερες επιχειρήσεις. Ιδίως, στα δημοφιλή αθλήματα (μπάσκετ, ποδόσφαιρο) δαπανώνται κάθε χρόνο υπέρογκα ποσά για μεταγραφές παιχτών και προπονητών. Δε θα μπορούσαν λοιπόν να μην ενταχτούν στο χώρο των στοιχημάτων. Μεγαλύτερο μάλιστα ενδιαφέρον από άποψη στοιχημάτων φαίνεται να συγκεντρώνει το λαοφιλέστερο άθλημα, το ποδόσφαιρο. Δεν είναι τυχαίο ότι χαρακτηρίζεται «ο βασιλιάς των αθλημάτων» και έχει κερδίσει την αγάπη όλων των κοινωνικών στρωμάτων και ηλικιών. Σύμφωνα με πρόσφατες έρευνες ασχολούνται ενεργά περισσότεροι από 250 εκατομμύρια αθλητές σε περισσότερα από 200 κράτη σε όλο το κόσμο.

Όσο περίεργο κι αν ακούγεται, αυτό που εξιτάρει περισσότερο τους παίχτες στοιχημάτων είναι η δυσκολία σωστής πρόβλεψης, η οποία στο ποδόσφαιρο είναι μεγαλύτερη. Ένας λόγος για αυτό είναι το γεγονός ότι στο ποδόσφαιρο υπάρχουν τρία πιθανά αποτελέσματα, σε αντίθεση με αθλήματα όπως το μπάσκετ και το βόλεϊ, για παράδειγμα, όπου υπάρχουν μόνο δύο. Επιπλέον, στο μπάσκετ και στο βόλεϊ ο ισχυρότερος κατά κανόνα κερδίζει. Βλέποντας, π.χ., μια κάρτα στατιστικών ενός αγώνα μπάσκετ, θα βρούμε στατιστικά που αφορούν τους πόντους που έχουν επιτύχει οι ομάδες, τα επιθετικά και αμυντικά φάουλ των παιχτών, τα ποσοστά ευστοχίας τους και άλλες πληροφορίες μέσω των οποίων μπορούμε να βγάλουμε ασφαλή συμπεράσματα. Για παράδειγμα μια ομάδα που διαθέτει καλύτερη αμυντική και επιθετική λειτουργία από την αντίπαλό της είναι σχεδόν βέβαιο ότι θα την κερδίσει, έστω και με μικρή διαφορά πόντων. Επιπροσθέτως, ο αριθμός των φάουλ των παιχτών παίζει σημαντικό ρόλο τόσο για την ομάδα όσο και για τον ίδιο τον παίκτη, σε αντίθεση με το ποδόσφαιρο όπου δεν υπάρχει κάποιος περιορισμός στον αριθμό φάουλ που μπορούν να κάνουν οι παίχτες. Επομένως, για το ποδόσφαιρο το στοιχείο αυτό μας είναι περιττό. Ακόμα, ένας ποδοσφαιρικός αγώνας σε πολλές περιπτώσεις κρίνεται από τη διαφορά ενός τέρματος. Ένα λάθος δηλαδή της άμυνας ή του τερματοφύλακα είναι ικανό να κρίνει το τελικό αποτέλεσμα, παρ' όλο που είναι πιθανό η συνολική εικόνα της ηττημένης ομάδας να είναι καλύτερη. Συν τοις άλλοις, στις μέρες μας κορυφαίοι προπονητές και ευρωπαϊκοί σύλλογοι θεωρούν επιτυχημένη στρατηγική το να δίνεται έμφαση στην άμυνα και να προσπαθεί η ομάδα μέσω αντεπιθέσεων να επιτύχει κάποιο γκολ. Μάλιστα, όταν το σημειώνει η εντολή που δίνεται στους παίχτες είναι να ροκανίζουν το χρόνο. Έτσι, μια ομάδα με μικρό αριθμό τελικών προσπαθειών και ατομικών ενεργειών είναι δυνατόν να σημειώσει κάποιο θετικό αποτέλεσμα (νίκη ή ισοπαλία). Αντίθετα, στο μπάσκετ δίνεται σε κάθε ομάδα περιορισμένος χρόνος για την επιθετική της προσπάθεια, κατά τη διάρκεια του οποίου καλείται να σημειώσει κάποιο καλάθι. Αν αυτή δε το κάνει και θελήσει απλά να κυλήσει ο χρόνος η μόνη κερδισμένη από αυτό θα είναι η αντίπαλός της. Βάσει λοιπόν όσων αναφέρθηκαν κατανοούμε το μέγεθος της πολυπλοκότητας και της τυχαιότητας που επικρατεί στο χώρο του ποδοσφαίρου.

Όπως έχουμε ήδη πει, στην παρούσα εργασία θα ασχοληθούμε με τις προβλέψεις ποδοσφαιρικών αγώνων. Κάποιος θα μπορούσε να ενδιαφερθεί για τον αριθμό των γκολ που θα σημειώσει η κάθε ομάδα σε κάθε ημίχρονο ή συνολικά, για το ποιο θα είναι το αποτέλεσμα του ημιχρόνου ή το τελικό, αν ο αγώνας θα λήξει με λιγότερα ή περισσότερα από 3 γκολ (under/over) και να κάνει προβλέψεις για αυτά που αναφέρθηκαν. Εμείς, όπως έχουμε ήδη τονίσει κάνουμε εκτιμήσεις για τα τελικά αποτελέσματα των ομάδων. Προσπαθούμε δηλαδή να προβλέψουμε αν ο κάθε αγώνας θα λήξει με νίκη της γηπεδούχου, ισοπαλία ή νίκη της φιλοξενούμενης

ομάδας (1,x,2).Τέλος, βάσει των αποτελεσμάτων εκτιμούμε ποιά θα είναι η συνολική βαθμολογία των ομάδων.

1.2 Στόχοι της εργασίας

Σκοπός της εργασίας είναι η εκτίμηση των τελικών αποτελεσμάτων ανά αγωνιστική για κάθε ομάδα και ο υπολογισμός της τελικής βαθμολογίας των ομάδων για τις σεζόν 2007-2008 έως 2013-2014 για το ελληνικό πρωτάθλημα ποδοσφαίρου αντρών Α' κατηγορίας.

1.3 Δεδομένα και παραδοχές

Η Superleague είναι η πρώτη ποδοσφαιρική επαγγελματική κατηγορία της Ελλάδας. Ο αριθμός των ομάδων που συμμετέχουν σε αυτή για τις σεζόν 2007-2008 έως 2012-2013 είναι 16 και για τη σεζόν 2013-2014 ο αριθμός αυτός αυξήθηκε σε 18. Αυτό σημαίνει πως κάθε χρονιά, πλην της πιο πρόσφατης, λαμβάνουν χώρα 240 αγώνες σε 30 αγωνιστικές, δηλαδή 8 ποδοσφαιρικοί αγώνες ανά αγωνιστική και για τη σεζόν 2013-2014 διεξάγονται 306 αγώνες σε 34 αγωνιστικές, δηλαδή 9 παιχνίδια ανά αγωνιστική. Δεδομένης της μεταβλητής αγωνιστικής συμπεριφοράς των ομάδων στις πρώτες κυρίως αγωνιστικές επιλέξαμε να ξεκινήσουμε τις εκτιμήσεις μας από την 6^η αγωνιστική έως το πέρας της τελευταίας. Επομένως, για τη σεζόν 2013-2014 οι εκτιμήσεις μας αφορούν συνολικά 261 αγώνες, ενώ για όλες τις υπόλοιπες χρονιές οι εκτιμήσεις μας γίνονται για συνολικά 200 αγώνες για κάθε σεζόν.

Η συλλογή των δεδομένων έγινε από τις πηγές [1,2,3,4] σε αρχεία της μορφής .xlsx. Η υλοποίηση των μεθόδων που περιγράφονται στα κεφάλαια που ακολουθούν έγινε στο περιβάλλον της Matlab.

1.4 Κριτήρια σύγκρισης μεθόδων

Ως κριτήρια για τη σύγκριση των μεθόδων χρησιμοποιήσαμε: α) το ποσοστό ακρίβειας των προβλέψεών μας, β) το σφάλμα στις ισοπαλίες, δηλαδή το σφάλμα στους αγώνες που είτε η εκτίμησή μας ήταν ισοπαλία και το αποτέλεσμα ήταν διαφορετικό, είτε το αποτέλεσμα ήταν ισοπαλία και η πρόβλεψή μας για την συγκεκριμένη αναμέτρηση ήταν νίκη της μιας ομάδας και ήττα της άλλης, γ) το σφάλμα στους αγώνες που η εκτίμησή μας ήταν νίκη της μιας ομάδας και το πραγματικό αποτέλεσμα ήταν ήττα αυτής, και δ) το μέσο απόλυτο ποσοστιαίο σφάλμα βαθμολογίας.

Το γενικό σφάλμα πρόβλεψης περιλαμβάνει τις λάθος εκτιμήσεις όλων των αγώνων και δίνεται από τη σχέση :

$$\frac{\sum_{i=6}^k \text{error}(i)}{\sum_{i=6}^k i}$$

Η ακρίβεια (ή αλλιώς ευστοχία, και οι δύο όροι χρησιμοποιούνται ισοδύναμα στην διπλωματική) στις προβλέψεις μας αντιστοιχεί στο ποσοστό των σωστών προβλέψεων και δίνεται από τη σχέση:

$$1 - \frac{\sum_{i=6}^k \text{error}(i)}{\sum_{i=6}^k i}$$

Το σφάλμα στις ισοπαλίες περιγράφεται από τη σχέση :

$$\frac{\sum_{i=6}^k \text{err_draw}(i)}{\sum_{i=6}^k i}$$

Η σχέση που δίνει το σφάλμα στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη είναι:

$$\frac{\sum_{i=6}^k \text{overtured_match}(i)}{\sum_{i=6}^k i}$$

Προφανώς, ισχύει ότι :

$$\frac{\sum_{i=6}^k \text{err_draw}(i)}{\sum_{i=6}^k i} + \frac{\sum_{i=6}^k \text{overtured_match}(i)}{\sum_{i=6}^k i} = \frac{\sum_{i=6}^k \text{error}(i)}{\sum_{i=6}^k i}$$

Η μεταβλητή k αντιστοιχεί στον συνολικό αριθμό αγωνιστικών κάθε χρονιάς. Για τη χρονιά 2013-2014 ισχύει ότι k=34, ενώ για τις υπόλοιπες k=30.

Το μέσο απόλυτο ποσοστιαίο σφάλμα εξετάζει τη συμπεριφορά της απόλυτης τιμής του σφάλματος της πρόβλεψης σε σχέση με την πραγματική τιμή και ορίζεται ως :

$$\frac{1}{N} \sum_{k=1}^N \frac{|F_k - A_k|}{A_k}$$

όπου ο όρος N αντιπροσωπεύει το πλήθος των ομάδων, ο όρος F_k είναι η εκτιμώμενη τιμή βαθμολογίας και ο όρος A_k αντιστοιχεί στην πραγματική τιμή βαθμολογίας.

1.5 Οργάνωση της εργασίας

Στο πρώτο κεφάλαιο διατυπώθηκε το πρόβλημα που μελετήσαμε και οι στόχοι της εργασίας μας. Επίσης, αναφερθήκαμε εν συντομία στις μεθόδους που χρησιμοποιήθηκαν και οι οποίες αναλύονται εκτενώς στα κεφάλαια που ακολουθούν. Τέλος, επισημάνθηκαν τα κριτήρια που χρησιμοποιούμε για την σύγκριση των μεθόδων.

Στο δεύτερο κεφάλαιο παρουσιάζεται μια κατηγορία συστημάτων που χρησιμοποιείται ευρέως για την κατάταξη των ομάδων σε διάφορα αθλήματα. Στο κεφάλαιο αυτό περιγράφουμε ορισμένες από τις δημοφιλέστερες μεθόδους κατάταξης και καταγράφουμε τα κυριότερα πλεονεκτήματα και μειονεκτήματα της χρήσης των εν λόγω συστημάτων. Μέσα σε αυτή την κατηγορία εντάσσεται και η πρώτη προσέγγισή μας για την επίλυση του προβλήματος της πρόβλεψης των αγώνων. Έπειτα, αναλύεται η πρώτη μέθοδος που εφαρμόσαμε και παρουσιάζουμε τα εξεγόμενα αποτελέσματα.

Το τρίτο κεφάλαιο περιλαμβάνει την κατηγορία της συσταδοποίησης. Αρχικά, γίνεται μια σύντομη αναφορά στη διαμεριστική συσταδοποίηση και κατόπιν αναλύεται διεξοδικά ο αλγόριθμος που επιλέξαμε να χρησιμοποιήσουμε (k-means) δίνοντας την περιγραφή του και τον τρόπο που τον εφαρμόσαμε στο πρόβλημά μας. Τέλος, εμφανίζουμε τα αποτελέσματα που προκύπτουν από τη χρήση της συγκεκριμένης μεθόδου.

Στο τέταρτο κεφάλαιο αναφερόμαστε στην κατανομή Poisson, όπου αρχικά παρέχουμε τον ορισμό της και έπειτα αναλύουμε πως αυτή μπορεί να χρησιμοποιηθεί για την επίλυση του προβλήματος μας. Τέλος, καταγράφουμε τα αποτελέσματα που προέκυψαν από τη χρήση της.

Στο πέμπτο κεφάλαιο παρουσιάζουμε δύο απλοϊκές μεθόδους πρόβλεψης, όπου στη μια χρησιμοποιούμε το στοιχείο της έδρας, ενώ στην άλλη χρησιμοποιούμε τα αποτελέσματα της προηγούμενης πριν την προβλεπόμενη σεζόν. Οι μέθοδοι αυτές χρησιμεύουν για την εξαγωγή baseline αποτελεσμάτων με τα οποία συγκρίνουμε τα αποτελέσματα των δικών μας μεθόδων.

Το έκτο κεφάλαιο περιέχει τις κυριότερες μεθόδους εξομάλυνσης βάσει των οποίων προβλέπουμε τη τελική βαθμολογία των ομάδων. Με τις συγκεκριμένες μεθόδους δεν κάνουμε εκτίμηση για την τελική έκβαση ενός αγώνα.

Στο έβδομο κεφάλαιο παρουσιάζονται τα κυριότερα συμπεράσματα που προέκυψαν, καθώς και κάποιες ιδέες για πιθανές μελλοντικές επεκτάσεις.

Στο όγδοο και τελευταίο κεφάλαιο (Παράρτημα) γίνεται παράθεση των γραφικών παραστάσεων που προκύπτουν βάσει της μεθόδου που αναλύεται στο δεύτερο κεφάλαιο για τις σεζόν 2008-2009 έως 2013-2014.

2. Συστήματα Κατάταξης

2.1 Βασικές έννοιες και εφαρμογές

Η στατιστική πρόβλεψη αποτελεσμάτων αγώνων ποδοσφαίρου είναι μια μέθοδος που χρησιμοποιείται στα στοιχήματα ποδοσφαίρου για να προβλέψει την έκβαση των αγώνων με τη βοήθεια στατιστικών εργαλείων. Οι bookmakers τα χρησιμοποιούν για να υπολογίσουν τις πιθανότητες (odds) των αποτελεσμάτων των ποδοσφαιρικών αγώνων. Μια ευρέως χρησιμοποιούμενη στατιστική προσέγγιση για την πρόβλεψη είναι η «κατάταξη».

Οι κατατάξεις παιχτών ή ομάδων στον αθλητισμό έχουν πολλές σημαντικές εφαρμογές. Κύριος στόχος τους είναι να παρέχουν αντικειμενικές ενδείξεις της ισχύος των παιχτών ή των ομάδων, με βάση τις προηγούμενες επιδόσεις τους. Με τον τρόπο αυτό, προσφέρουν πληροφορίες σχετικά με το πραγματικό επίπεδο και την τρέχουσα εξέλιξη των ανταγωνιστικών παιχτών ή ομάδων.

Με τον όρο σύστημα κατάταξης (sport rating system) αναφερόμαστε σε ένα σύστημα, το οποίο αναλύει τα αποτελέσματα των αθλητικών διοργανώσεων, ώστε να παρέχει αντικειμενικές αξιολογήσεις για κάθε ομάδα. Στη συνέχεια, η κατάταξη κάθε ομάδας προέρχεται από την αξιολόγησή της. Το σύστημα κατάταξης αθλημάτων χρησιμοποιεί τη μέθοδο power ranking για τις αξιολογήσεις των ομάδων. Η συγκεκριμένη μέθοδος στηρίζεται στον υπολογισμό της δύναμης (power score) της κάθε ομάδας σε σχέση με τις υπόλοιπες. Στη πρώτη θέση της κατάταξης βρίσκεται η ομάδα με το υψηλότερο σκορ (power score), την οποία θεωρούμε και ισχυρότερη όλων και ακολουθούν οι υπόλοιπες βάσει του power score που διαθέτουν.

2.2 Τα κυριότερα συστήματα κατάταξης

Η χρήση των συστημάτων κατάταξης ξεκίνησε ήδη από τη δεκαετία του 1930 (Dunkel Index, [5]) και η χρήση τους παρουσιάζει διαρκώς αυξητική πορεία με την πάροδο του χρόνου. Μάλιστα, η εν λόγω κατηγορία συστημάτων χρησιμοποιείται σε πολλά αθλήματα, όπως το τένις (ATP rankings, WTA rankings), το ράγκμπι (IRB World Rankings), το ποδόσφαιρο (FIFA World Rankings, FIFA Women's World Rankings), το μπόξ (WPBF Computerized World Rankings), το κρίκετ (ICC Team Rankings, ICC Player Rankings), το γκολφ (Official World Golf Rankings, Women's World Golf Rankings), το μπάσκετ (FIBA Basketball World Rankings), το βόλεϊ (FIVB Volleyball World Rankings) και το μπέιζμπολ (IBAF Baseball World Rankings).

Στις υποενότητες που ακολουθούν παρουσιάζουμε τα πιο δημοφιλή και αξιόπιστα συστήματα κατάταξης.

2.2.1 Το επίσημο σύστημα κατάταξης της FIFA (*The official FIFA ranking*)

Η τρέχουσα μεθοδολογία κατάταξης της FIFA εισήχθη μετά τη λήξη του Παγκοσμίου Κυπέλλου που διεξήχθη στη Γερμανία το 2006 [6]. Η περιγραφή του αλγορίθμου μπορεί να βρεθεί στην επίσημη ιστοσελίδα της FIFA (FIFA.com, [7]). Για τον υπολογισμό των πόντων κατάταξης για τις ομάδες, προσμετρούνται τα αποτελέσματα των τελευταίων τεσσάρων ετών. Για κάθε ομάδα ο τύπος για τον υπολογισμό των πόντων (P) υπολογίζεται μετά από κάθε παιχνίδι ως εξής :

$$P=M \times I \times T \times C$$

όπου :

- M: οι πόντοι που αντιστοιχούν στο αποτέλεσμα του παιχνιδιού
- I: η σημασία του παιχνιδιού
- T: η δύναμη της αντίπαλης ομάδας
- C: ο μέσος όρος δύναμης των συνομοσπονδιών των συμμετεχουσών ομάδων.

Για την καταγραφή των πόντων M εφαρμόζεται μια τυποποιημένη σύμβαση. Αν το αποτέλεσμα μιας αναμέτρησης είναι νίκη, τότε απονέμονται στην ομάδα 3 βαθμοί, αν είναι ισοπαλία 1 βαθμός και στην ήττα δίνονται 0 βαθμοί. Οι αγώνες που έληξαν ισόπαλοι και ο νικητής προέκυψε ύστερα από την διαδικασία των πέναλτι αντιμετωπίζονται με διαφορετικό τρόπο και ειδικότερα η νικήτρια ομάδα λαμβάνει 2 βαθμούς, ενώ η ηττημένη παίρνει 1 βαθμό.

Ο όρος I, ο οποίος προσμετρά τη σημασία ενός παιχνιδιού λαμβάνει τιμές από 1 έως 4. Οι αγώνες του Παγκοσμίου Κυπέλλου θεωρούνται οι πιο σημαντικοί (I=4), ενώ τα αποτελέσματα από τα φιλικά παιχνίδια είναι τα λιγότερο σημαντικά (I=1).

Ο μέσος όρος δύναμης των συνομοσπονδιών των συμμετεχουσών ομάδων C λαμβάνει τιμές από 0.85 έως 1. Η ευρωπαϊκή συνομοσπονδία ποδοσφαίρου (UEFA) και η αντίστοιχη της Νότιας Αμερικής (CONMEBOL) διαθέτουν τη μέγιστη τιμή ενώ τη χαμηλότερη τιμή έχει η Ποδοσφαιρική Συνομοσπονδία Ωκεανίας (OFC). Όταν δύο ομάδες παίζουν, ο όρος C υπολογίζεται ως ο μέσος όρος της δύναμης των συνομοσπονδιών στην οποία ανήκουν.

Τέλος, η δύναμη της αντίπαλης ομάδας T υπολογίζεται ως 200 μείον τη θέση της αντίπαλης ομάδας. Ως εξαίρεση, η ομάδα που βρίσκεται στην κορυφή της παγκόσμιας κατάταξης λαμβάνει τη μέγιστη ισχύ (T=200), ενώ στις ομάδες που κατατάσσονται από την 150^η θέση και κάτω δίνεται η ελάχιστη ισχύ (T=50).

Μόλις έχουμε υπολογίσει τους πόντους P που μια ομάδα κέρδισε, στη συνέχεια εφαρμόζουμε ένα σταθμισμένο μέσο όρο των πόντων κατά τον ακόλουθο τρόπο. Για κάθε χρονιά ο συνολικός αριθμός πόντων αντιστοιχεί στο άθροισμα των πόντων που προκύπτουν μετά από κάθε παιχνίδι προς το πλήθος των παιχνιδιών. Σε περίπτωση που μια ομάδα έπαιξε λιγότερο από πέντε παιχνίδια σε ένα χρόνο διαιρούμε το συνολικό αριθμό πόντων με πέντε. Στη συνέχεια, δίνεται βαρύτητα 0.2,

0.3, 0.5 και 1 στις τέσσερις ετήσιες τιμές P, όπου στα πιο πρόσφατα αποτελέσματα εκχωρείται υψηλότερη βαρύτητα.

2.2.2 Το σύστημα κατάταξης Elo (*The Elo rating system*)

Το σύστημα Elo [8] είναι μια μέθοδος που αρχικά δημιουργήθηκε για την κατάταξη παιχτών σκάκι. Δημιουργός του είναι ο Arpad Elo, ένας ουγγρικής καταγωγής καθηγητής φυσικής και επαγγελματίας παίχτης σκάκι. Το σύστημα αυτό αποτελεί ένα zero-sum σύστημα, δηλαδή όσους πόντους κερδίζει ένας παίχτης, τους χάνει ο άλλος. Το κλειστό αυτό σύστημα ανταλλαγής πόντων δεν έχει εφαρμογή μόνο στο σκάκι, αλλά και σε αθλήματα όπως το ποδόσφαιρο, το ράγκμπι, το μπάσκετ, το μπίιζμπολ, ακόμη και σε video games, όπως το Counter Strike, το World of Warcraft και το League of Legends. Η διαφορά στην κατάταξη δύο παιχτών (ή ομάδων) χρησιμοποιείται ως προγνωστικός παράγοντας για την έκβαση ενός αγώνα. Δύο παίχτες (ή ομάδες) με ίδια κατάταξη θεωρείται ότι θα σημειώσουν ίσο αριθμό νικών. Η βαθμολογία ενός παίκτη (ή μιας ομάδας) αντιπροσωπεύεται από έναν αριθμό που αυξάνεται ή μειώνεται με βάση την έκβαση των αποτελεσμάτων που σημειώνει. Αν ο ισχυρότερος παίχτης (ή ομάδα) κερδίσει έναν κατώτερο αντίπαλό του λαμβάνει λιγότερους πόντους στην κατάταξη, σε σχέση με την περίπτωση που συμβεί το αντίθετο. Επίσης, σε περίπτωση ισοπαλίας πάλι ο κατώτερος αντίπαλος λαμβάνει περισσότερους πόντους. Στο σύστημα αυτό δηλαδή, ένα θετικό αποτέλεσμα ενός ανίσχυρου παίκτη (ή ομάδας) του προσφέρει μεγαλύτερη ισχυροποίηση, σε σχέση με κάποιον ισχυρό που σημειώνει το ίδιο αποτέλεσμα.

Κεντρική παραδοχή του συστήματος Elo είναι ότι η απόδοση του κάθε παίκτη (ή ομάδας) σε κάθε παιχνίδι είναι μια κανονική κατανομημένη τυχαία μεταβλητή. Παρά το γεγονός ότι η απόδοση ενός παίκτη (ή μιας ομάδας) αλλάζει από το ένα παιχνίδι στο άλλο, ο Elo υποστηρίζει ότι η μέση τιμή των αποδόσεων κάθε παίκτη (ή ομάδας) αλλάζει αργά με την πάροδο του χρόνου. Θεωρεί δηλαδή ότι η πραγματική ικανότητα ενός παίκτη (ή μιας ομάδας) είναι ο μέσος όρος των αποδόσεων του συγκεκριμένου παίκτη (ή ομάδας).

Οι εξισώσεις που δίνουν τις πιθανότητες νίκης του παίκτη (ή της ομάδας) a και του παίκτη (ή της ομάδας) b είναι αντίστοιχα :

$$E_a = \frac{1}{1+10^{\frac{R_b-R_a}{400}}} \text{ και } E_b = \frac{1}{1+10^{\frac{R_a-R_b}{400}}}$$

όπου :

- Ο όρος E_x αντιστοιχεί στην αναμενόμενη πιθανότητα ότι ο παίχτης (ή η ομάδα) X θα κερδίσει τον αγώνα. Προφανώς, ισχύει ότι : $E_a + E_b = 1$.
- Ο όρος R_x είναι η βαθμολογία του παίκτη (ή της ομάδας) X.

Η βαθμολογία R_x , η οποία αντιστοιχεί στους πόντους κατάταξης του παίχτη (ή της ομάδας) X ενημερώνεται επαναληπτικά μετά από κάθε αγώνα. Ο τύπος υπολογισμού είναι ο ακόλουθος :

$$R'_x = R_x + K \times (s_x - p_x)$$

όπου :

- R'_x : η νέα τιμή βαθμολογίας του παίχτη (ή της ομάδας) X
- R_x : η παλιά τιμή βαθμολογίας του παίχτη (ή της ομάδας) X
- s_x : το πραγματικό αποτέλεσμα του αγώνα που σημείωσε ο παίχτης (ή η ομάδα) X
- p_x : το αναμενόμενο αποτέλεσμα του παίχτη (ή της ομάδας) X
- K : μια θετική σταθερά, η οποία λαμβάνει τιμή 32 για ανίσχυρους παίκτες (ή ομάδες), τιμή 24 για παίκτες (ή ομάδες) μεσαίας δυναμικότητας και τιμή 16 για ισχυρούς παίκτες (ή ομάδες).

Το πραγματικό αποτέλεσμα του αγώνα s_x παίρνει τιμές 0 ή 0.5 ή 1 στην περίπτωση που ο παίχτης (ή η ομάδα) X χάσει, φέρει ισοπαλία ή κερδίσει το παιχνίδι αντίστοιχα.

Όπως αναφέραμε και προηγουμένως, το μοντέλο Elo υποθέτει πως η απόδοση του παίχτη (ή της ομάδας) X ακολουθεί μια κανονική κατανομή σε έναν αγώνα γύρω από τη μέση τιμή της R_x . Μια απλουστευμένη παραδοχή είναι ότι η διακύμανση είναι ομοιογενής ανάμεσα σε όλους τους παίκτες (ή τις ομάδες), δηλαδή $\sigma_x = \sigma$ για κάθε παίκτη (ή ομάδα) X . Όταν δύο παίκτες (ή ομάδες), A και B , πρόκειται να αναμετρηθούν συγκρίνουμε τις δύο κατανομές της απόδοσής τους P_A και P_B με $P_A \sim N(R_A, \sigma^2)$ και $P_B \sim N(R_B, \sigma^2)$. Η πιθανότητα ο παίκτης (ή η ομάδα) A να κερδίσει το παιχνίδι είναι ίση με την πιθανότητα του γεγονότος ότι κατέχει υψηλότερη τιμή P_A από την αντίστοιχη του αντιπάλου του. Από τις ιδιότητες της κανονικής κατανομής έχουμε ότι $P_A - P_B \sim N(R_A - R_B, 2 \times \sigma^2)$, σύμφωνα με την παραδοχή της ανεξαρτησίας. Έτσι, υπολογίζουμε το αναμενόμενο αποτέλεσμα του παιχνιδιού από την πλευρά του παίκτη (ή της ομάδας) A :

$$p_A = P(P_A > P_B) = \Phi\left(\frac{r_A - r_B}{\sigma\sqrt{2}}\right)$$

όπου Φ υποδηλώνει την αθροιστική συνάρτηση κατανομής μιας τυποποιημένης κανονικής κατανομής $N \sim (0,1)$. Εάν η υπολογιζόμενη τιμή της p_A είναι περίπου 0.5, τότε περιμένουμε ισοπαλία. Πιθανές επεκτάσεις στο μοντέλο πρόβλεψης Elo για την πιθανότητα ισοπαλίας συζητούνται από τους Rao και Kupper (1967) [9] και τον Davidson (1969) [10]. Σε πολλές εφαρμογές, είναι σύνηθες να χρησιμοποιούν την λογιστική κατανομή για την απόδοση ενός παίχτη, η οποία υπολογίζεται ως εξής :

$$p_A = \frac{1}{1 + e^{-\alpha \times (r_A - r_B)}}$$

όπου a είναι ο κατάλληλος συντελεστής κλιμάκωσης. Ο τύπος αυτός που προέρχεται από την λογιστική κατανομή φαίνεται να είναι πιο προσιτός. Η κύρια ιδέα πίσω από το μοντέλο που περιγράφηκε είναι ότι με βάση τις τρέχουσες αξιολογήσεις κάνουμε πρόβλεψη για το μέλλον του παιχνιδιού. Μάλιστα εάν ένας παίχτης (ή μια ομάδα) απέδωσε καλύτερα από ό, τι αναμέναμε ενάντια στον αντίπαλό του B , δηλαδή αν $s_A > p_A$, τότε θα αυξηθούν σημαντικά οι πόντοι του και θα βελτιωθεί η κατάταξή του στο σύστημα.

2.2.3 Το επίσημο σύστημα κατάταξης της FIFA για το παγκόσμιο πρωτάθλημα γυναικών (*FIFA Women's World Rankings*)

Η FIFA χρησιμοποιεί διαφορετικό αλγόριθμο για την κατάταξη και την αξιολόγηση των εθνικών ομάδων ποδοσφαίρου των γυναικών. Στην πραγματικότητα, μπορούμε να αναγνωρίσουμε μια εκδοχή του μοντέλου Elo πίσω από αυτή τη μέθοδο κατάταξης. Η περιγραφή του αλγορίθμου μπορεί να βρεθεί στην επίσημη ιστοσελίδα της FIFA (FIFA.com,[11]). Στο σύστημα αυτό οι βαθμολογίες ενημερώνονται σύμφωνα με τον τύπο :

$$R'_x = R_x + K \times I \times (s_x - p_x)$$

Ο βασικός συντελεστής K έχει οριστεί με σταθερή τιμή 15. Επιπλέον, υπάρχει ένας ακόμη όρος που σχετίζεται με τη σημασία του παιχνιδιού. Αντίστοιχα με το σύστημα κατάταξης των ανδρών της FIFA, οι τιμές του I δεν είναι σταθερές και κυμαίνονται από 1(φιλικοί αγώνες) έως 4 (αγώνες παγκοσμίου κυπέλου). Μια τροποποίηση του ανωτέρω συστήματος συμβαίνει στις τιμές του s_x . Ειδικότερα, αν στο σύστημα κατάταξης των ανδρών η ομάδα A έχανε, τότε $s_A=0$. Στο σύστημα όμως των γυναικών, αν η ομάδα A χάσει 1-0, τότε $s_A=0.15$ και η αντίστοιχη τιμή της αντιπάλου της είναι 0.85. Μάλιστα όσο μικρότερη είναι η διαφορά της ήττας και όσο περισσότερα είναι τα τέρματα που σημειώνει η ηττημένη ομάδα A , τόσο μεγαλύτερη θα είναι η τιμή του s_A που θα διαθέτει. Στον πίνακα που ακολουθεί παρουσιάζονται ακριβώς οι τιμές του s_A που δίνονται σε κάθε ηττημένη ομάδα A . Η αντίπαλός της θα λαμβάνει τιμές $s_{\text{opponent}} = 1 - s_A$.

Πίνακας 2-1 : Η βαθμολογία s_A που δίνεται στην ηττημένη ομάδα A

Γκολ Υπέρ	Διαφορά Τερμάτων						
	0	1	2	3	4	5	6+
0	0.5	0.15	0.08	0.04	0.03	0.02	0.01
1	0.5	0.16	0.089	0.048	0.037	0.026	0.015
2	0.5	0.17	0.098	0.056	0.044	0.032	0.02
3	0.5	0.18	0.107	0.064	0.051	0.038	0.025
4	0.5	0.19	0.116	0.072	0.058	0.044	0.03
5+	0.5	0.2	0.125	0.08	0.065	0.05	0.035

Τέλος, το αναμενόμενο αποτέλεσμα του παιχνιδιού υπολογίζεται με τη χρήση της συνάρτησης της λογιστικής κατανομής. Επιπλέον, γίνεται μία διόρθωση για να ενσωματωθεί το πλεονέκτημα της έδρας. Συγκεκριμένα, η πιθανότητα η ομάδα Α να κερδίσει το παιχνίδι υπολογίζεται ως εξής :

$$P_A = \frac{1}{1 + 10^{-(100 + r_A - r_B)}}$$

όπου η σταθερά 100 πιστώνεται στον οικοδεσπότη του παιχνιδιού, που στην συγκεκριμένη περίπτωση είναι η ομάδα Α.

2.2.4 EloRatings.net

Σε αυτή την υποενότητα περιγράφουμε μια άλλη εκδοχή του μοντέλου Elo που περιγράφεται στην ιστοσελίδα EloRatings.net [12]. Ο τύπος υπολογισμού του και η πιθανότητα νίκης μιας ομάδας Α δίνονται αντίστοιχα από τις σχέσεις :

$$R'_x = R_x + K \times G \times (s_x - p_x)$$

$$P_A = \frac{1}{1 + 10^{-(100 + r_A - r_B)}}$$

Οι πιθανές τιμές του s_A είναι 0 (ήττα), 0.5 (ισοπαλία) και 1 (νίκη). Οι τιμές του G , προκύπτουν από τη διαφορά τερμάτων N των δύο ομάδων και συγκεκριμένα :

- $G=1$, αν $N \leq 1$
- $G=1.5$, αν $N=2$
- $G = \frac{N+11}{8}$, αν $N \geq 3$
- Η σταθερά K , λαμβάνει τιμές από 20(φιλική αναμέτρηση) έως 60(αγώνας παγκοσμίου κυπέλου), ανάλογα με τη σημασία του αγώνα.

2.2.5 Jeff Sagarin

Ο Jeff Sagarin είναι ένας Αμερικανός στατιστικολόγος γνωστός για την ανάπτυξη του συστήματός του [13], [14] για την κατάταξη και την αξιολόγηση ομάδων σε διάφορα αθλήματα και κυρίως στο μπάσκετ (NBA, NCAA Men's Division | Basketball Championship), στο ποδόσφαιρο (Major League Soccer), στο ράγκμπι (National Football League) και στο μπέιζμπολ (Major League Baseball). Δεν είναι ακόμα γνωστή η μεθοδολογία του συστήματός του. Αυτό όμως που γνωρίζουμε είναι ότι αποτελείται από δύο μεθόδους κατάταξης. Η μία είναι το «σύστημα κατάταξης Elo», που περιγράφηκε στην ενότητα 2.2.2. Αυτό το σύστημα χρησιμοποιεί μόνο νίκες και ήττες χωρίς να προσμετρά το μέγεθος της νίκης ή της ήττας. Η δεύτερη μέθοδος είναι το «Pure Points», η οποία λαμβάνει υπ' όψιν το μέγεθος αυτό. Έτσι, μια ομάδα που κερδίζει με μικρή διαφορά ανταμείβεται λιγότερο από ότι μια ομάδα που κερδίζει τον ίδιο αντίπαλο με μεγάλη διαφορά πόντων. Τέλος,

όπως και στην περίπτωση του συστήματος Elo, οι ομάδες ισχυροποιούνται και λαμβάνουν υψηλότερη κατάταξη κερδίζοντας ισχυρότερους αντιπάλους.

2.2.6 RPI

Το Rating Percentage Index, κοινώς γνωστό ως RPI, χρησιμοποιείται [15] για την κατάταξη των ομάδων. Βασίζεται στις νίκες και στις ήττες μιας ομάδας, καθώς επίσης και στο πρόγραμμά της, δηλαδή στη δύναμη των ομάδων που αντιμετωπίζει. Το σύστημα αυτό χρησιμοποιείται από το 1981 σε πλήθος αθλημάτων, όπως για παράδειγμα στο μπάσκετ (NBA, NCAA Men's Division I Basketball Championship, NCAA Women's Division I Basketball Championship), στο μπέιζμπολ (Major League Baseball) και στο ράγκμπι (National Football League). Δωρεάν προβλέψεις που προκύπτουν βάσει του RPI παρέχονται στις πηγές [16], [17], [18], [19], [20].

Η σχέση που ισχύει για την πρόβλεψη αγώνων μπάσκετ είναι:

$$\text{RPI} = 0.25 \times \text{WP} + 0.5 \times \text{OWP} + 0.25 \times \text{OOWP}$$

όπου :

- RPI : Rating Percentage Index
- WP : Winning Percentage
- OWP : Opponents' Winning Percentage
- OOWP : Opponents' Opponents' Winning Percentage

Ο όρος WP υπολογίζεται ως ο αριθμός των νικών που σημειώνει κάθε ομάδα προς το πλήθος των παιχνιδιών που δίνει. Ο όρος OWP υπολογίζεται λαμβάνοντας το μέσο όρο των WP κάθε αντίπαλης ομάδας, αφαιρώντας του αγώνες που έχουν δοθεί εναντίον της ομάδας της οποίας θέλουμε να υπολογίσουμε το RPI. Τέλος, ο όρος OOWP υπολογίζεται ως ο μέσος όρος των OWP κάθε αντίπαλης ομάδας.

Από τη σεζόν 2004-2005 άρχισε να χρησιμοποιείται μια τροποποίηση του RPI, σύμφωνα με την οποία δίνεται διαφορετική βαρύτητα για την εντός έδρας νίκη και ήττα από ότι για την εκτός έδρας. Συγκεκριμένα, δίνεται βαρύτητα 0.6 για την εντός έδρας νίκη και για την εκτός έδρας ήττα και βαρύτητα 1.4 για την εντός έδρας ήττα και την εκτός έδρας νίκη. Η τροποποίηση αυτή αναφέρεται μόνο στον υπολογισμό του όρου WP. Οι όροι OWP και OOWP υπολογίζονται όπως και στην περίπτωση του κλασικού RPI. Τη συγκεκριμένη μέθοδο την ονομάζουμε RPI Weighted, ο τροποποιημένος όρος WP αναφέρεται ως WP Weighted και ισχύουν οι ακόλουθες σχέσεις :

$$\text{RPI Weighted} = 0.25 \times \text{WP Weighted} + 0.5 \times \text{OWP} + 0.25 \times \text{OOWP}$$

$$\text{WP Weighted} = \frac{(0.6 \times \text{HomeWin} + 1.4 \times \text{AwayWin})}{(0.6 \times \text{HomeWin} + 1.4 \times \text{AwayWin} + 0.6 \times \text{AwayLoss} + 1.4 \times \text{HomeLoss})}$$

Ο όρος Home Win αντιστοιχεί στην εντός έδρας νίκη, ο όρος Away Win στην εκτός έδρας νίκη, ο όρος Home Loss στην εντός έδρας ήττα και ο όρος Away Loss αναφέρεται στην εκτός έδρας ήττα.

Η σχέση που ισχύει για την πρόβλεψη αγώνων μπέιζμπολ είναι η ίδια που περιγράφηκε παραπάνω με τη διαφορά ότι δίνεται βαρύτητα 0.7 για την εντός έδρας νίκη και για την εκτός έδρας ήττα και βαρύτητα 1.3 για την εκτός έδρας νίκη και την εντός έδρας ήττα.

2.3 Περιγραφή της μεθόδου υλοποίησης

Λόγω της συχνής χρήσης των συστημάτων κατάταξης και ιδίως του RPI σε πλήθος διαφορετικών αθλητικών διοργανώσεων σκεφτήκαμε να το εφαρμόσουμε και στην περίπτωση του ποδοσφαίρου. Απ' όσο γνωρίζουμε, είναι η πρώτη φορά που το RPI χρησιμοποιείται στην σχετική βιβλιογραφία για εκτίμηση αποτελεσμάτων αγώνων ποδοσφαίρου.

Έτσι, για τις σεζόν 2007-2008 έως 2013-2014 υπολογίζουμε το RPI κάθε ομάδας ανά αγωνιστική εφαρμόζοντας τη φόρμουλα που χρησιμοποιείται στο μπάσκετ. Στη συνέχεια, βάσει του RPI κατατάσσουμε τις ομάδες και ελέγχουμε κάθε φορά τις τιμές του RPI των δύο αντίπαλων ομάδων. Οι πιθανές τιμές του RPI και του RPI Weighted κυμαίνονται από 0 έως 1. Για τη σύγκριση των τιμών δύο ομάδων λαμβάνουμε υπ' όψιν μας ακρίβεια δύο δεκαδικών ψηφίων. Αν το RPI μιας ομάδας είναι μεγαλύτερο από της άλλης, τότε θεωρούμε νίκη υπέρ της και ήττα της αντιπάλου της, ενώ αν έχουν ίδια τιμή, τότε θεωρούμε πως το αποτέλεσμα της αναμέτρησης θα είναι ισοπαλία. Την ίδια ακριβώς λογική ακολουθούμε και για το RPI Weighted. Έπειτα, βάσει των αποτελεσμάτων που έχουμε εκτιμήσει για κάθε ομάδα υπολογίζουμε την τελική της βαθμολογία.

2.4 Αποτελέσματα

Στους πίνακες που ακολουθούν παρουσιάζονται το μέσο ποσοστιαίο σφάλμα βαθμολογίας, το ποσοστό ευστοχίας στις προβλέψεις μας, το σφάλμα που έγινε στους αγώνες που είτε η εκτίμησή μας ήταν ισοπαλία και ο αγώνας τελείωσε με νικητή και ηττημένο, είτε το πραγματικό αποτέλεσμα ήταν ισοπαλία και η εκτίμησή μας ήταν διαφορετική και το σφάλμα στους αγώνες, όπου η ομάδα που εκτιμήσαμε ότι θα κερδίσει ήταν τελικά η ηττημένη.

Πίνακας 2-2 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
RPI	34.6	35.62	34.54	28.14	37.73	28.19	32.41	33.03
RPI Weighted	28.28	33.44	34.72	26.85	33.69	29.38	31.81	31.17

Πίνακας 2-3 : Ποσοστό ευστοχίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
RPI	48.5	47.5	46.5	43.5	47	46	44.8	46.26
RPI Weighted	52.5	47.5	48	47.5	50	49	49.1	49.09

Πίνακας 2-4: Ποσοστό σφάλματος ισοπαλιών

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
RPI	26	33	28.5	31	27	28	24.1	28.23
RPI Weighted	23.5	33	25.5	26.5	25.5	26	23	26.14

Πίνακας 2-5: Ποσοστό σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη

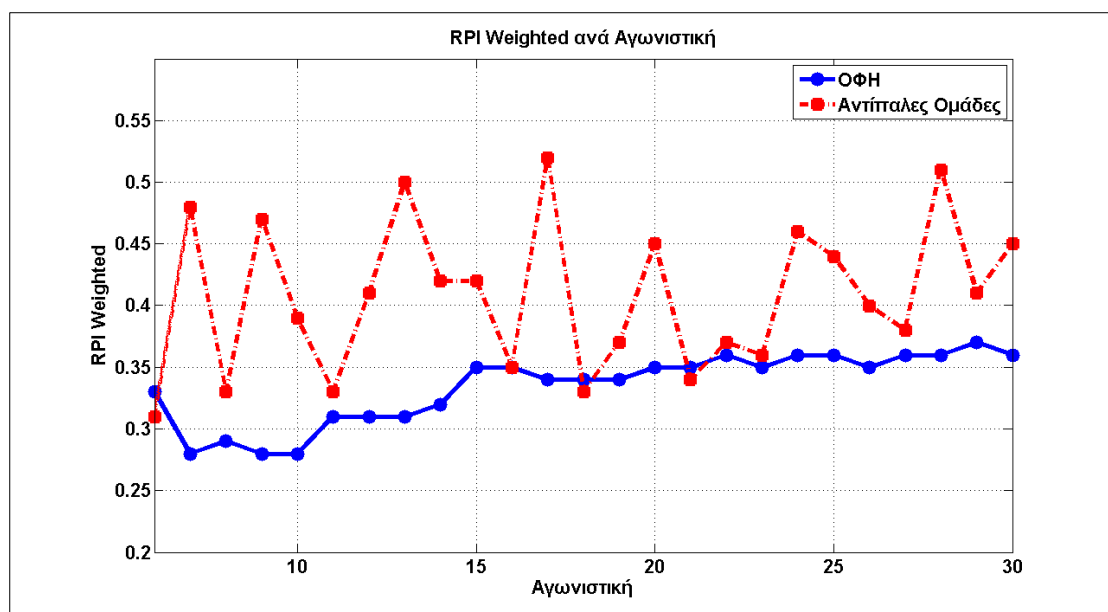
	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
RPI	25.5	19.5	25	25.5	26	26	31	25.5
RPI Weighted	24	19.5	26.5	26	24.5	25	27.9	24.77

Από τα εξαγόμενα αποτελέσματα παρατηρούμε πως η μέθοδος RPI Weighted μας προσφέρει καλύτερα αποτελέσματα από τη μέθοδο RPI. Ιδίως το ποσοστό ευστοχίας με τη χρήση του RPI Weighted είναι μεγαλύτερο για όλες τις χρονιές, με εξαίρεση τη σεζόν 2008-2009, όπου και οι δύο μέθοδοι είναι εξίσου εύστοχες. Για τη συγκεκριμένη όμως σεζόν το μέσο ποσοστιαίο σφάλμα βαθμολογίας που προσφέρει

το RPI Weighted είναι μικρότερο από το αντίστοιχο του RPI. Άρα υπερτερεί έναντι του RPI και για τη συγκεκριμένη σεζόν. Παράλληλα, παρατηρούμε πως το RPI Weighted μας προσφέρει μικρότερο σφάλμα ισοπαλιών για κάθε χρονιά, πλην της σεζόν 2008-2009, όπου το σφάλμα είναι ίδιο και για τις δύο μεθόδους. Τέλος, για όλες σχεδόν τις σεζόν διακρίνουμε πως το ποσοστό «απόλυτης ανατροπής» της πρόβλεψής μας, δηλαδή των περιπτώσεων που η εκτιμώμενη νικήτρια ομάδα ηττάται, ήταν περίπου το ίδιο και για τις δύο μεθόδους. Μόνο τη σεζόν 2013-2014 παρατηρούμε μεγαλύτερη ευστοχία και άρα μικρότερο ποσοστό σφάλματος του RPI Weighted συγκρινόμενο με το αντίστοιχο του RPI. Επομένως, η ύπαρξη διαφορετικής βαρύτητας, που περιέχει η μέθοδος RPI Weighted, στο αποτέλεσμα (νίκη/ήττα) μιας αναμέτρησης και στην έδρα (εντός/εκτός), όπου διεξάγεται ο αγώνας, κρίνεται αποδοτική και εύστοχη.

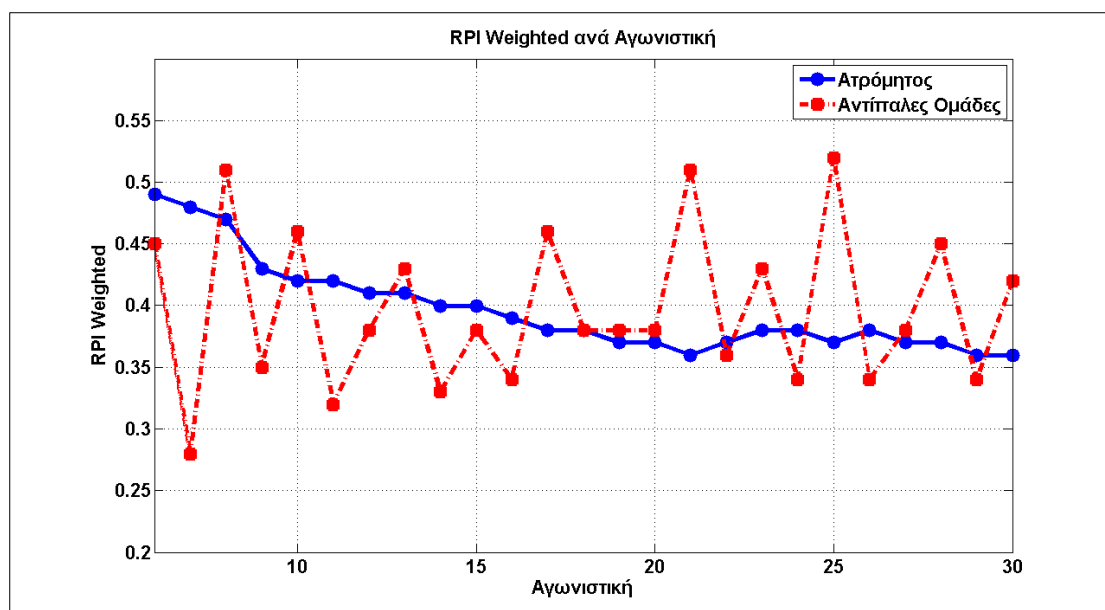
Στα σχήματα που ακολουθούν παρουσιάζεται η διακύμανση του RPI Weighted για κάθε ομάδα (μπλε γραμμή) και οι τιμές του RPI Weighted των αντιπάλων της κάθε μιας (κόκκινη γραμμή) την αγωνιστική που τις αντιμετωπίζει.

- Για τη σεζόν 2007-2008:



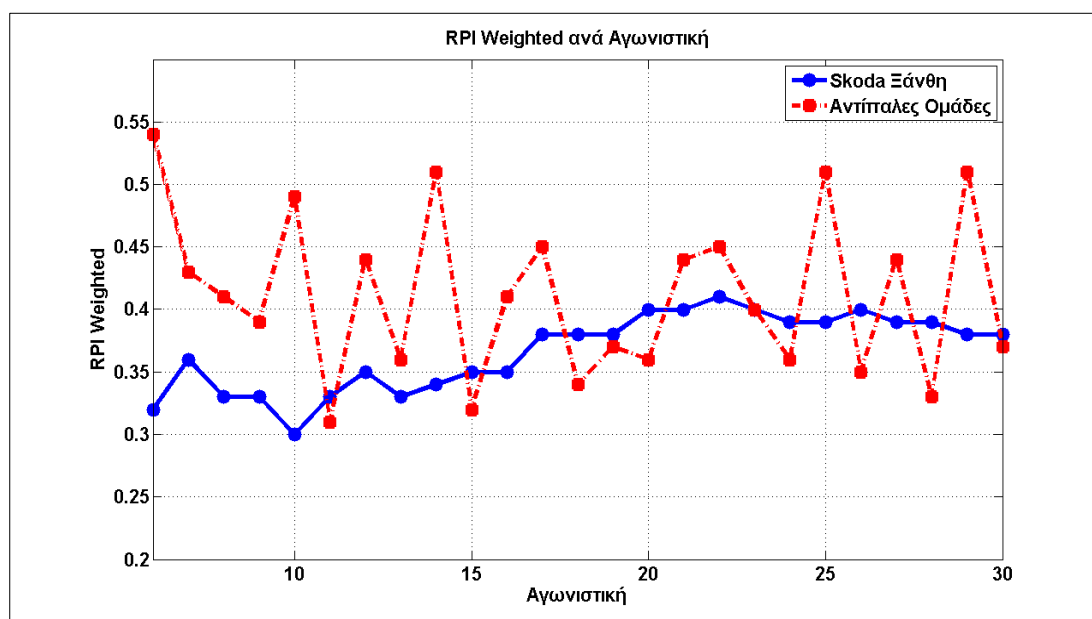
Εικόνα 2-1 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-1 παρατηρούμε ότι από την 6^η αγωνιστική έως την 15^η η τιμή του RPI Weighted του ΟΦΗ κυμαίνεται από 0.28 έως 0.35, ενώ από την 15^η και έπειτα η τιμή του RPI Weighted που διαθέτει είναι σχεδόν σταθερή και κυμαίνεται από 0.34 έως 0.37. Επιπλέον, διακρίνουμε ότι στην 6^η, 11^η, 16^η, 18^η, 21^η, 22^η, 23^η και 27^η οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον ΟΦΗ, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν από αυτόν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



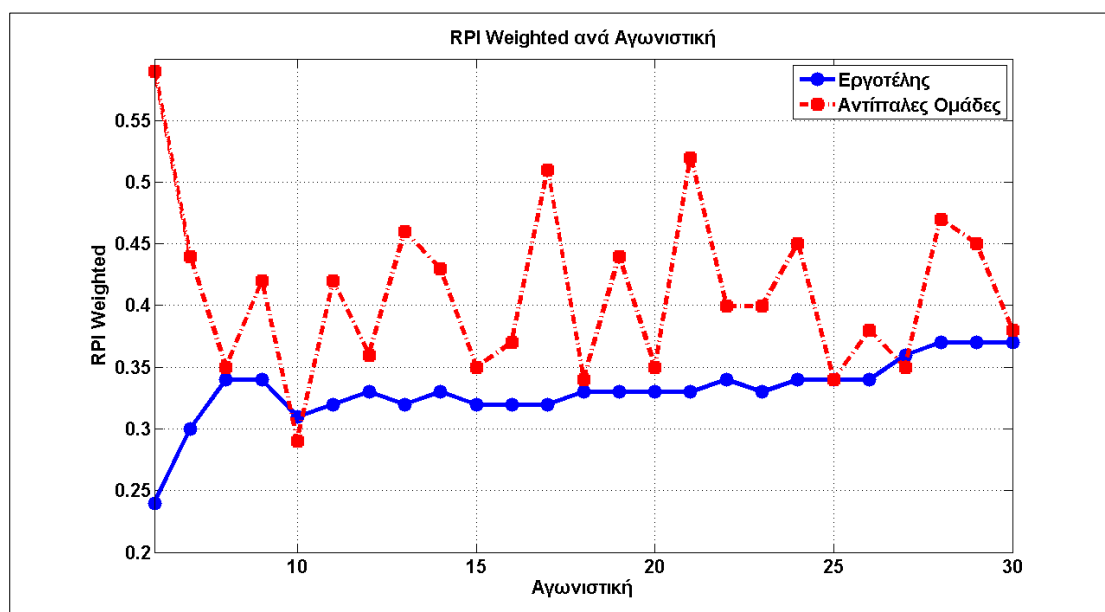
Εικόνα 2-2 : Οι τιμές του RPI Weighted του Ατρόμητου και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-2 παρατηρούμε ότι από την 6^η αγωνιστική έως το τέλος της 30^{ης} η τιμή του RPI Weighted του Ατρόμητου παρουσιάζει πτωτική πορεία. Ειδικότερα, ξεκινά στην 6^η αγωνιστική με τιμή λίγο μικρότερη από 0.5 και στο τέλος της 30^{ης} η τιμή του φτάνει το 0.36. Επιπλέον, από την 15^η αγωνιστική έως το τέλος της 30^{ης} η τιμή του RPI Weighted που διαθέτει κυμαίνεται από 0.35 έως 0.4. Επιπρόσθετα, διακρίνουμε ότι στην 13^η, 15^η, 18^η, 19^η, 20^η, 22^η, 27^η και 29^η οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον Ατρόμητο, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν ή υστερούν από αυτόν.



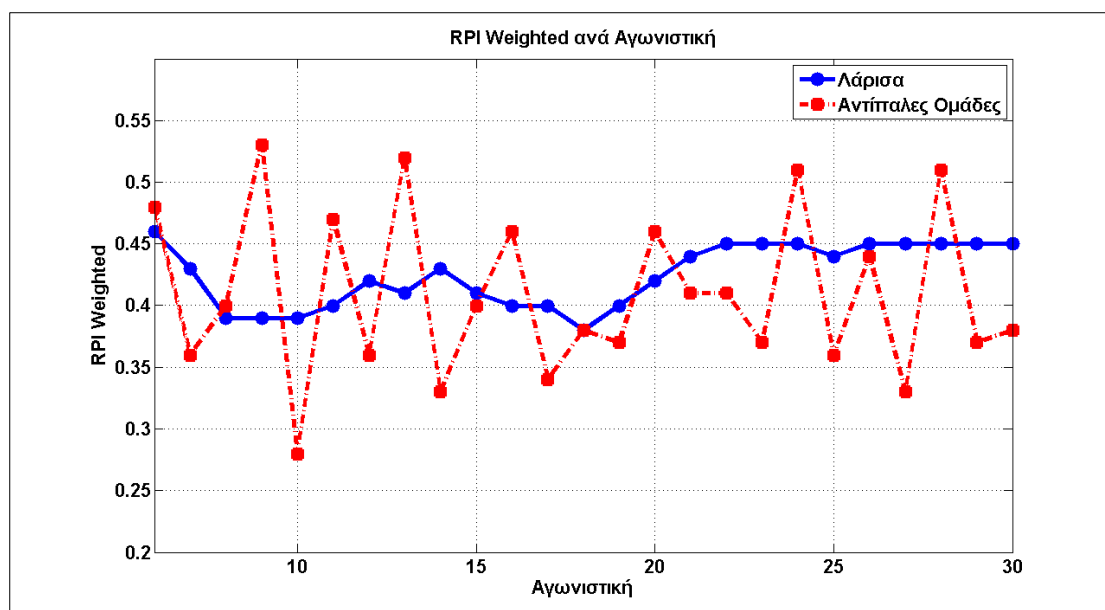
Εικόνα 2-3 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική

Από την εικόνα 2-3 παρατηρούμε ότι από την 6^η αγωνιστική έως τη 15^η η τιμή του RPI Weighted της Skoda Ξάνθης μεταβάλλεται και πιο συγκεκριμένα κυμαίνεται από 0.3 έως 0.36. Στη συνέχεια, από την 15^η αγωνιστική έως το τέλος της 30^{ης} η τιμή του RPI Weighted της Skoda Ξάνθης αυξάνεται και ειδικότερα κυμαίνεται από 0.35 έως 0.41. Παράλληλα, διακρίνουμε ότι στην 11^η, 15^η, 19^η, 23^η, 24^η και 30^η οι αντίπαλοί της είναι παρόμοιας δυναμικής με την Skoda Ξάνθη, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί της ξεκάθαρα υπερέχουν ή υστερούν από αυτήν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



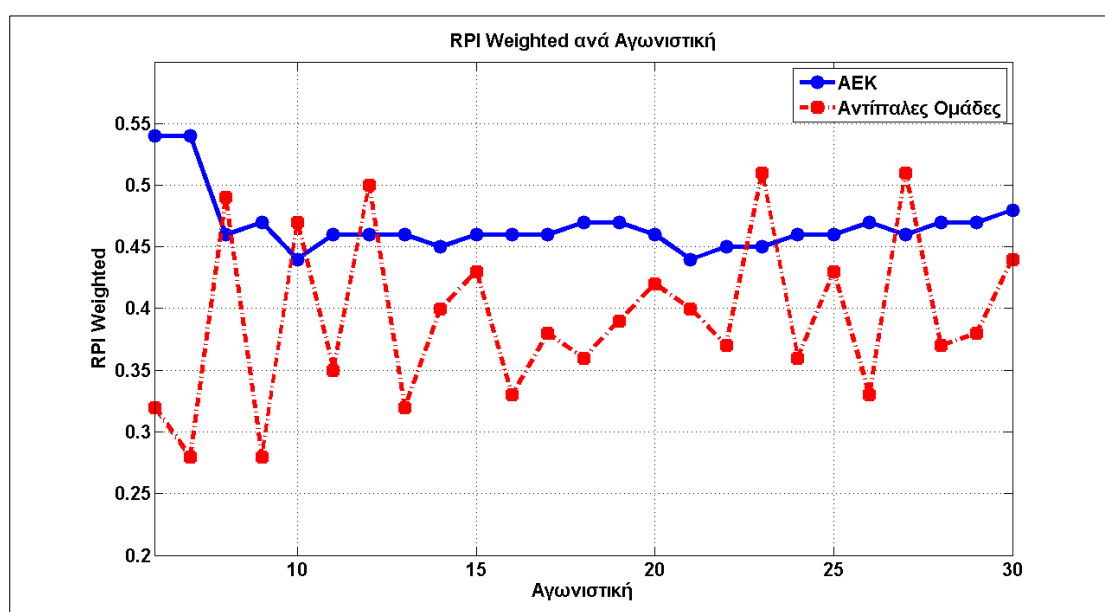
Εικόνα 2-4 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-4 παρατηρούμε ότι από την 6^η αγωνιστική έως το τέλος της 30^{ης} η τιμή του RPI Weighted παρουσιάζει ανοδική πορεία. Πιο συγκεκριμένα, ξεκινά με τιμή λίγο μικρότερη από 0.25, ενώ στο τέλος της 30^{ης} φτάνει το 0.37. Ιδίως από την 10^η αγωνιστική έως την 26^η η τιμή του RPI Weighted του Εργοτέλη παρουσιάζει μικρή μεταβλητότητα και λαμβάνει τιμή μεγαλύτερη από 0.3 και μικρότερη από 0.35. Παράλληλα, διακρίνουμε ότι στην 8^η, 10^η, 18^η, 20^η, 25^η, 27^η και 30^η οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον Εργοτέλη, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν από αυτόν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



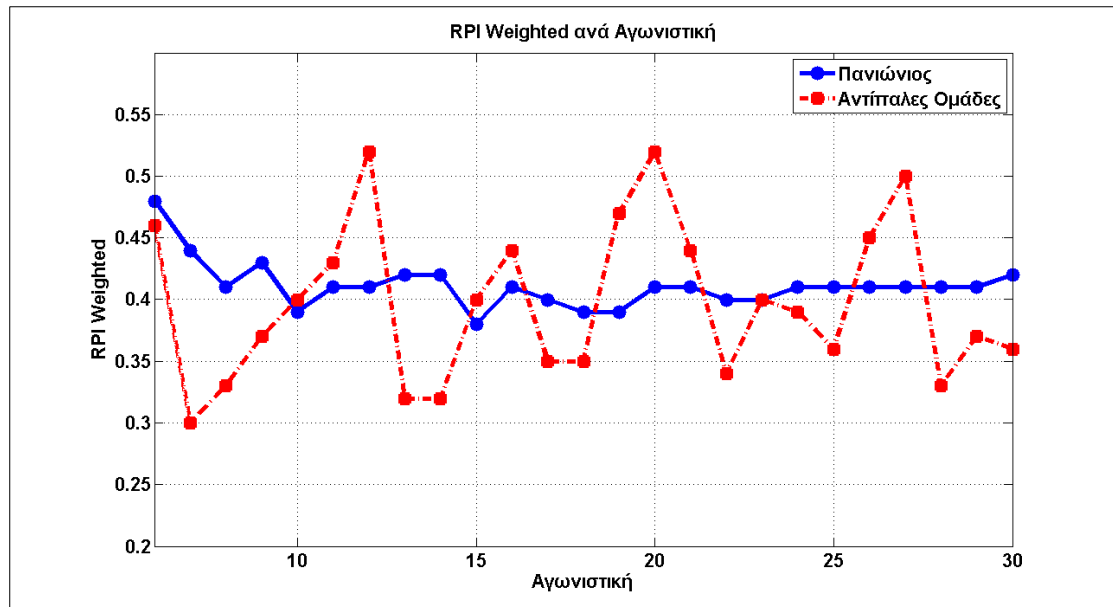
Εικόνα 2-5 : Οι τιμές του RPI Weighted της Λάρισας και των αντιπάλων της ανά αγωνιστική

Από την εικόνα 2-5 παρατηρούμε ότι από την 22^η αγωνιστική έως το τέλος της 30^{ης} η τιμή του RPI Weighted της Λάρισας είναι σχεδόν σταθερή και ίση με 0.45, ενώ για τις υπόλοιπες αγωνιστικές η τιμή του RPI Weighted κυμαίνεται από 0.38 έως 0.46. Επιπλέον, διακρίνουμε ότι στην 6^η, 8^η, 15^η, 18^η και 26^η οι αντίπαλοί της είναι παρόμοιας δυναμικής με την Λάρισα, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί της ξεκάθαρα υπερέχουν ή υστερούν από αυτήν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



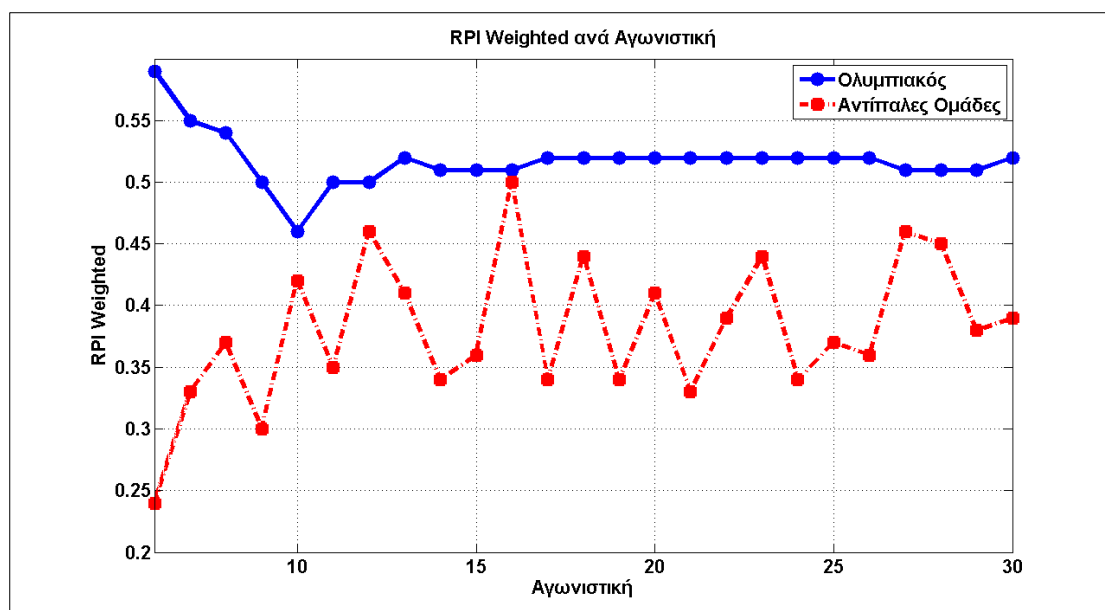
Εικόνα 2-6 : Οι τιμές του RPI Weighted της ΑΕΚ και των αντιπάλων της ανά αγωνιστική

Από την εικόνα 2-6 παρατηρούμε ότι η μεγαλύτερη τιμή του RPI Weighted της ΑΕΚ σημειώνεται στην 6^η και 7^η αγωνιστική, με τιμή λίγο μικρότερη από 0.55. Στις υπόλοιπες αγωνιστικές η τιμή του RPI Weighted κυμαίνεται από 0.44 έως 0.48. Επιπλέον, διακρίνουμε ότι στην 8^η και 10^η αγωνιστική οι αντίπαλοί της είναι παρόμοιας δυναμικής με την ΑΕΚ, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί της ξεκάθαρα υπερέχουν ή υστερούν από αυτήν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



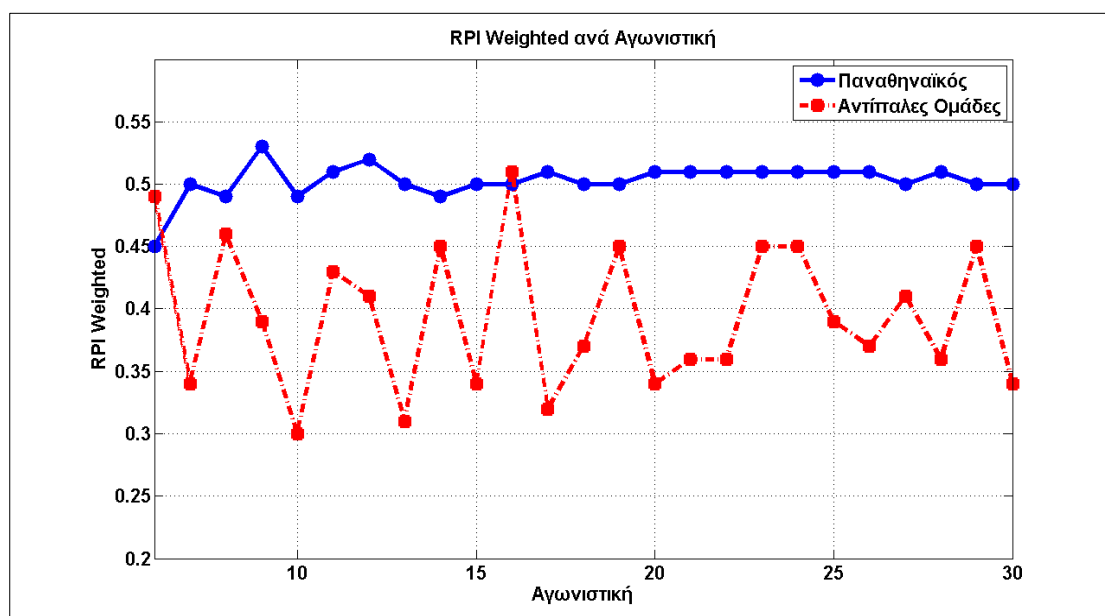
Εικόνα 2-7 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-7 παρατηρούμε ότι η μεγαλύτερη τιμή του RPI Weighted του Πανιωνίου σημειώνεται στην 6^η, με τιμή 0.48. Επιπλέον, από την 16^η αγωνιστική και έπειτα παρουσιάζει σχεδόν σταθερή τιμή RPI Weighted που κυμαίνεται από 0.39 έως 0.41. Για τις υπόλοιπες αγωνιστικές το εύρος της τιμής του RPI Weighted είναι από 0.38 έως 0.44. Επιπλέον, διακρίνουμε ότι στην 6^η, 10^η, 11^η, 15^η, 16^η, 23^η και 24^η αγωνιστική οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον Πανιώνιο, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν ή υστερούν από αυτόν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



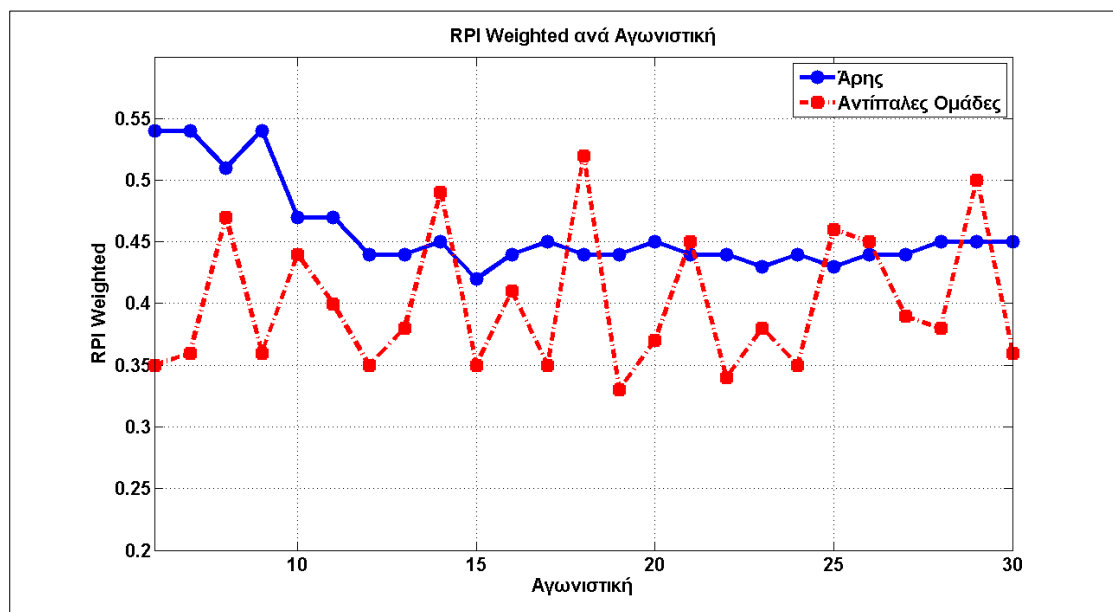
Εικόνα 2-8 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-8 παρατηρούμε ότι η μεγαλύτερη τιμή του RPI Weighted του Ολυμπιακού σημειώνεται στην 6^η, με τιμή λίγο μικρότερη από 0.6. Η τιμή αυτή είναι η μεγαλύτερη που σημειώθηκε για τη σεζόν 2007-2008. Επιπλέον, από την 11^η αγωνιστική και έπειτα παρουσιάζει σχεδόν σταθερή τιμή RPI Weighted που κυμαίνεται από 0.5 έως 0.52. Για τις υπόλοιπες αγωνιστικές το εύρος της τιμής του RPI Weighted είναι από 0.46 έως 0.55. Επιπλέον, διακρίνουμε ότι μόνο στην 16^η αγωνιστική η αντίπαλός του είναι παρόμοιας δυναμικής με τον Ολυμπιακό, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υστερούν από αυτόν.



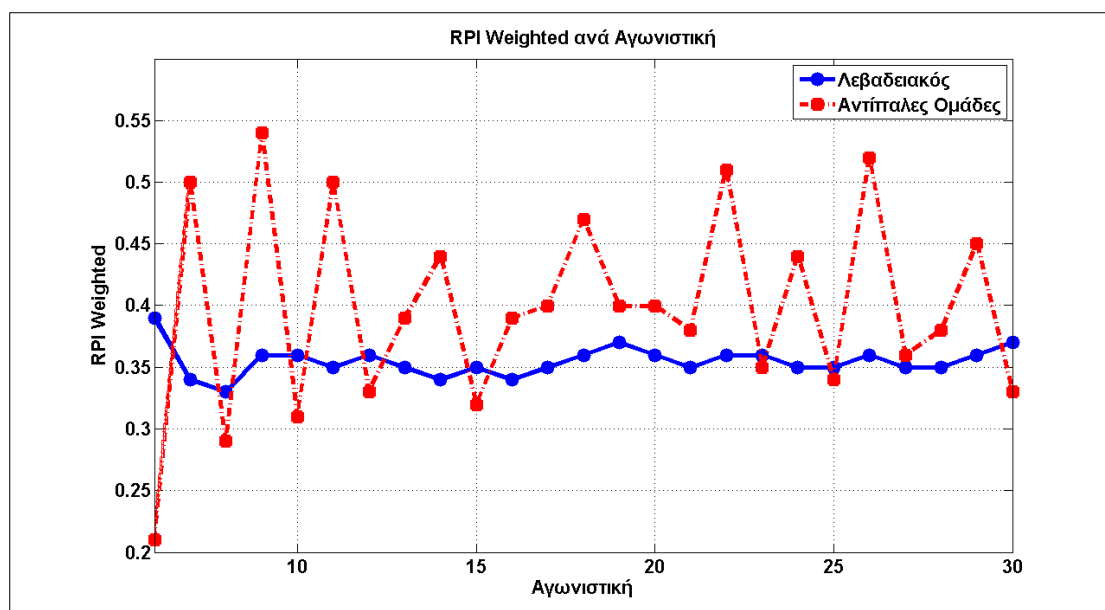
Εικόνα 2-9 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-9 παρατηρούμε ότι από την 13^η αγωνιστική και έπειτα η τιμή του RPI Weighted του Παναθηναϊκού είναι σχεδόν σταθερή και κυμαίνεται από 0.49 έως 0.51. Για τις υπόλοιπες αγωνιστικές το εύρος της τιμής του RPI Weighted είναι από 0.45 έως 0.53. Επιπλέον, διακρίνουμε ότι μόνο στην 16^η αγωνιστική η αντίπαλός του είναι παρόμοιας δυναμικής με τον Παναθηναϊκό, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υστερούν από αυτόν, με εξαίρεση την 6^η αγωνιστική, όπου υπερτερεί η αντίπαλός του, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



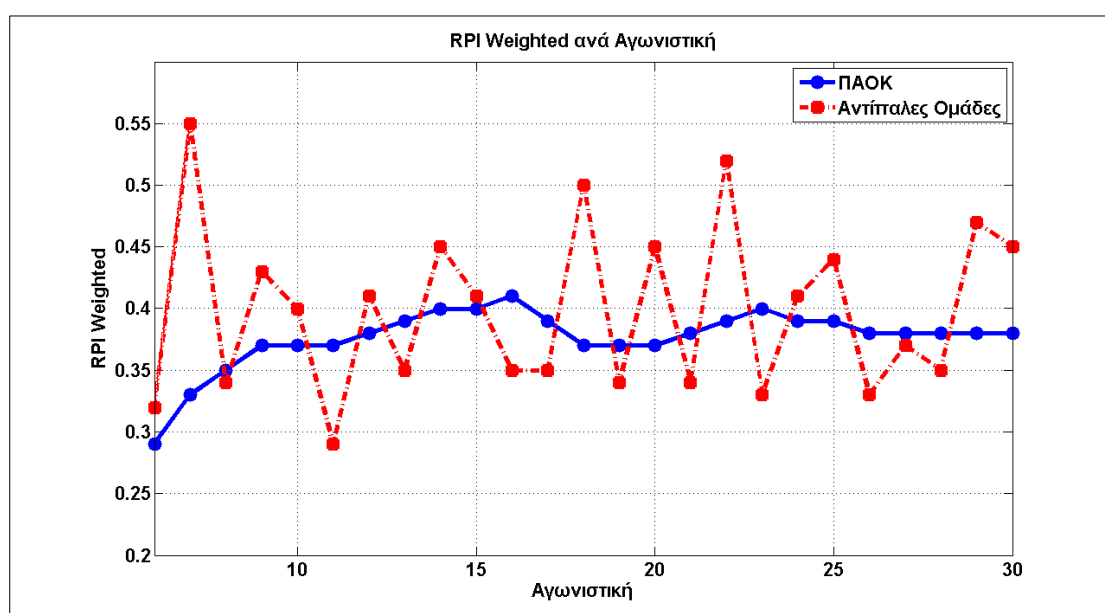
Εικόνα 2-10 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-10 παρατηρούμε ότι η μεγαλύτερη τιμή του RPI Weighted του Άρη σημειώνεται στην 6^η, 7^η και 9^η αγωνιστική, με τιμή λίγο μικρότερη από 0.55. Επιπλέον, από την 12^η αγωνιστική και έπειτα παρουσιάζει σχεδόν σταθερή τιμή RPI Weighted που κυμαίνεται από 0.43 έως 0.45. Για τις υπόλοιπες αγωνιστικές το εύρος της τιμής του RPI Weighted είναι από 0.46 έως 0.51. Επιπλέον, διακρίνουμε ότι στην 21^η και 26^η αγωνιστική οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον Άρη, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν ή υστερούν από αυτόν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



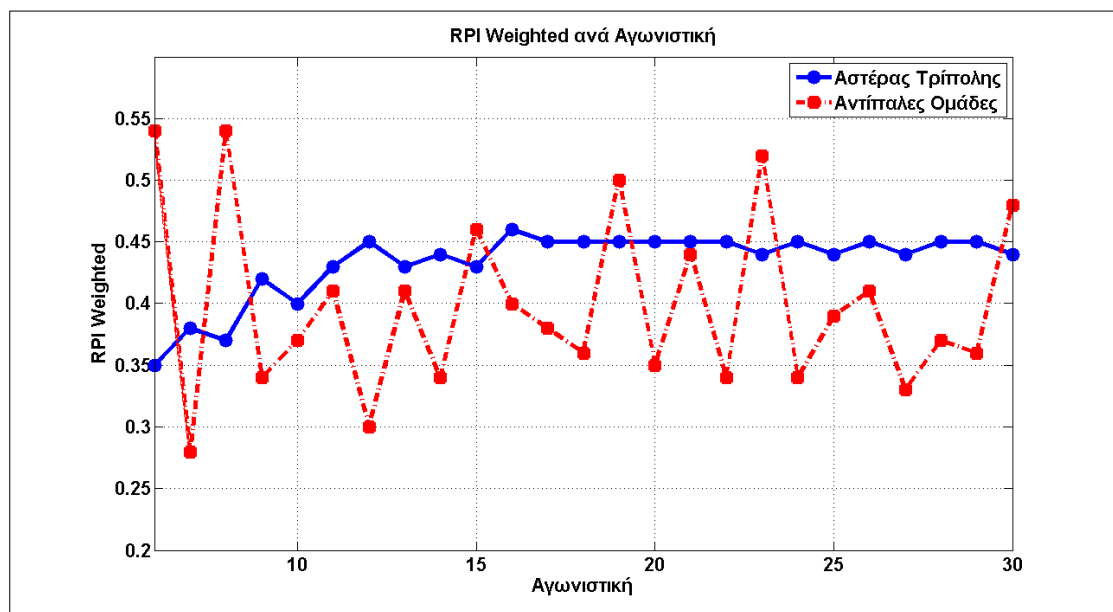
Εικόνα 2-11 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-11 παρατηρούμε ότι η μεγαλύτερη τιμή του RPI Weighted του Λεβαδειακού σημειώνεται στην 6^η αγωνιστική, με τιμή λίγο μικρότερη από 0.4. Επιπλέον, από την 7^η αγωνιστική και έπειτα παρουσιάζει σχεδόν σταθερή τιμή RPI Weighted που κυμαίνεται από 0.34 έως 0.37. Ακόμα, διακρίνουμε ότι στην 23^η, 25^η και 27^η αγωνιστική οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον Λεβαδειακό, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν ή υστερούν από αυτόν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



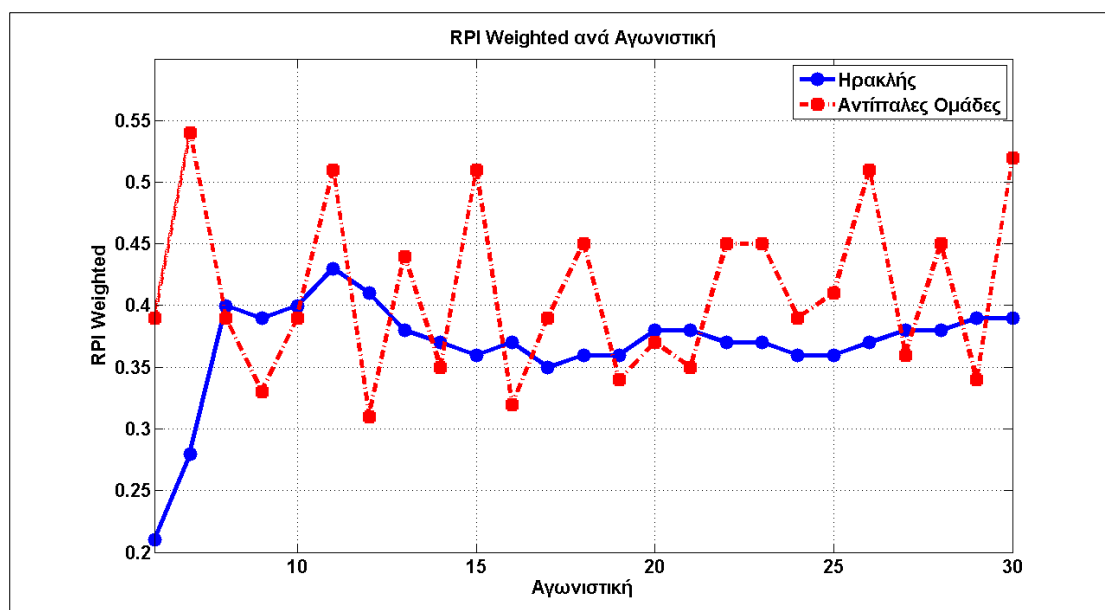
Εικόνα 2-12 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-12 παρατηρούμε ότι η μικρότερη τιμή του RPI Weighted του ΠΑΟΚ σημειώνεται στην 6^η αγωνιστική, με τιμή λίγο μικρότερη από 0.3 και η μέγιστη λαμβάνει χώρα στην 16^η αγωνιστική με τιμή περίπου 0.41. Για όλες τις υπόλοιπες αγωνιστικές η τιμή του RPI Weighted κυμαίνεται από 0.33 έως 0.4. Ακόμα, διακρίνουμε ότι στην 8^η, 15^η και 27^η αγωνιστική οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον ΠΑΟΚ, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν ή υστερούν από αυτόν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



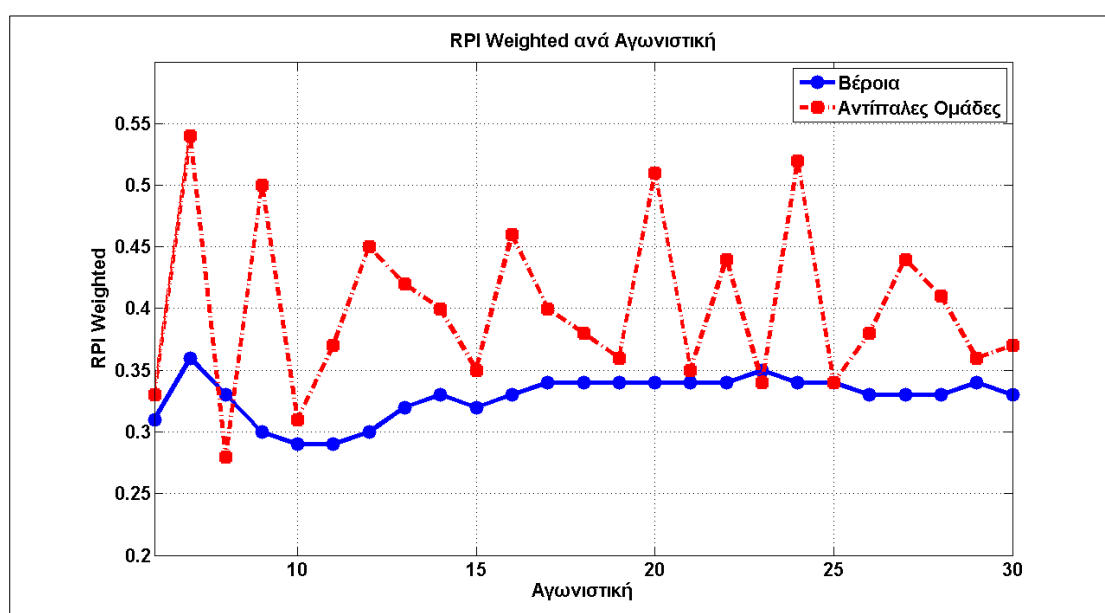
Εικόνα 2-13 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-13 παρατηρούμε ότι από την 16^η αγωνιστική και έπειτα η τιμή του RPI Weighted του Αστέρα Τρίπολης είναι σχεδόν σταθερή και ίση με 0.45. Για όλες τις υπόλοιπες αγωνιστικές η τιμή του RPI Weighted κυμαίνεται από 0.35 έως 0.46. Ακόμα, διακρίνουμε ότι στην 11^η, 13^η και 21^η αγωνιστική οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον Αστέρα Τρίπολης, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν ή υστερούν από αυτόν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν.



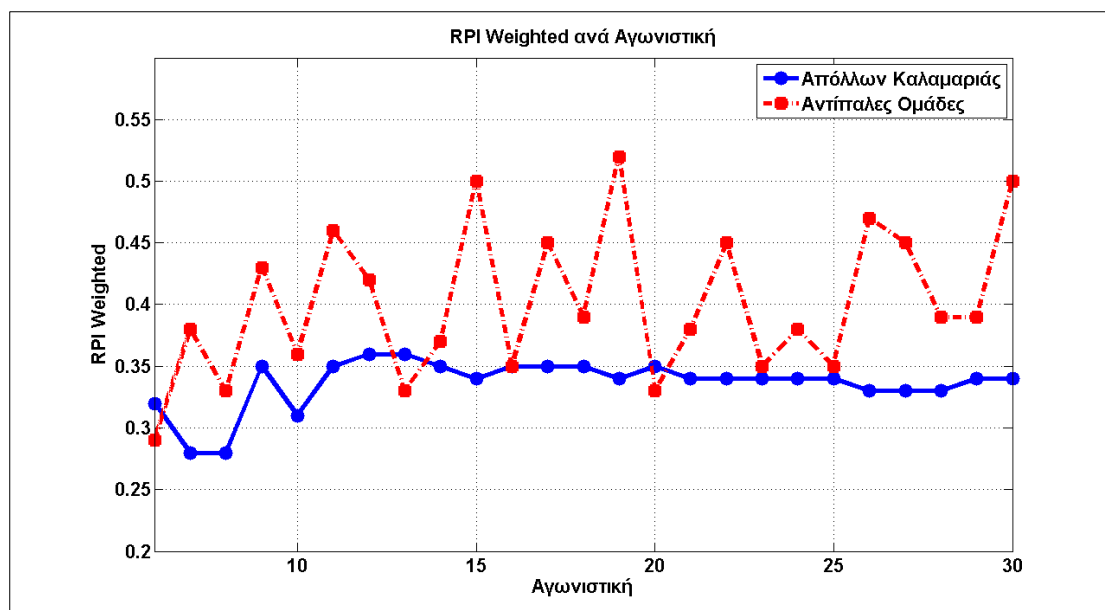
Εικόνα 2-14 : Οι τιμές του RPI Weighted του Ηρακλή και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-14 παρατηρούμε ότι η μικρότερη τιμή του RPI Weighted του Ηρακλή σημειώνεται στην 6^η αγωνιστική και είναι λίγο μεγαλύτερη από 0.2. Στη συνέχεια, μέχρι την 11^η αγωνιστική η πορεία της τιμής του είναι ανοδική και φθάνει έως την τιμή του 0.43. Για όλες τις υπόλοιπες αγωνιστικές η τιμή του RPI Weighted κυμαίνεται από 0.35 έως 0.41. Ακόμα, διακρίνουμε ότι στην 8^η, 10^η, 14^η, 19^η, 20^η και 27^η αγωνιστική οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον Ηρακλή, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν ή υστερούν από αυτόν.



Εικόνα 2-15 : Οι τιμές του RPI Weighted της Βέροιας και των αντιπάλων της ανά αγωνιστική

Από την εικόνα 2-15 παρατηρούμε ότι από την 17^η αγωνιστική και έπειτα η τιμή του RPI Weighted της Βέροιας είναι σχεδόν σταθερή με τιμή που κυμαίνεται από 0.33 έως 0.35. Ακόμα, διακρίνουμε ότι στην 6^η, 10^η, 21^η, 23^η, 25^η και 29^η αγωνιστική οι αντίπαλοί της είναι παρόμοιας δυναμικής με τη Βέροια, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί της ξεκάθαρα υπερτερούν από αυτήν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν, με εξαίρεση την 8^η αγωνιστική, όπου η αντίπαλός της υστερεί.



Εικόνα 2-16 : Οι τιμές του RPI Weighted του Απόλλωνα Καλαμαριάς και των αντιπάλων του ανά αγωνιστική

Από την εικόνα 2-16 παρατηρούμε ότι από την 15^η αγωνιστική και έπειτα η τιμή του RPI Weighted του Απόλλωνα Καλαμαριάς είναι σχεδόν σταθερή με τιμή που κυμαίνεται από 0.33 έως 0.35, ενώ για τις υπόλοιπες η τιμή του κυμαίνεται από 0.28 έως 0.36. Ακόμα, διακρίνουμε ότι στην 6^η, 14^η, 16^η, 20^η, 23^η και 25^η αγωνιστική οι αντίπαλοί του είναι παρόμοιας δυναμικής με τον Απόλλωνα Καλαμαριάς, ενώ στις υπόλοιπες αγωνιστικές οι αντίπαλοί του ξεκάθαρα υπερέχουν από αυτόν, όσον αφορά την τιμή του RPI Weighted που διαθέτουν, με εξαίρεση τη 13^η αγωνιστική, όπου η αντίπαλός του υστερεί από αυτόν.

Οι γραφικές παραστάσεις του RPI Weighted των ομάδων για τις σεζόν 2008-2009 έως 2013-2014 βρίσκονται στο κεφάλαιο 8.

2.5 Αποτελέσματα βάσει της διαφοράς θέσεων κατάταξης των ομάδων

Θέλοντας να εμβαθύνουμε και να αναλύσουμε περισσότερο τα αποτελέσματα σκεφτήκαμε να υπολογίσουμε το ποσοστό ευστοχίας στις προβλέψεις μας και τα επιμέρους ποσοστά σφαλμάτων που υπολογίστηκαν πιο πάνω βάσει της διαφοράς των θέσεων των ομάδων στην κατάταξη που προκύπτει από τη χρήση του RPI Weighted. Αντίστοιχη ανάλυση γίνεται στο [21] για το πρωτάθλημα ράγκμπι των ΗΠΑ. Η διαφορά των θέσεων υπολογίζεται ως :

$$|RPI\ Weighted_{(ομάδας\ i)} - RPI\ Weighted_{(αντιπάλου\ ομάδας\ i)}|$$

Η μικρότερη πιθανή τιμή ισούται με 0, σε περίπτωση που οι δύο ομάδες διαθέτουν ίση τιμή και η μέγιστη ισούται με το πλήθος των ομάδων-1, σε περίπτωση που η μία ομάδα είναι πρώτη στην κατάταξη και η αντίπαλός της τελευταία.

Πίνακας 2-6 : Ποσοστό ευστοχίας στις προβλέψεις μας βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2007-8 έως 2010-11)

Διαφορά Θέσεων RPI Weighted	2007-2008	2008-2009	2009-2010	2010-2011
0	(1/3)	(2/7)	(2/10)	(5/7)
1	(9/21)	(7/21)	(8/22)	(9/20)
2	(10/23)	(2/19)	(6/11)	(12/22)
3	(10/21)	(6/16)	(8/13)	(8/19)
4	(9/16)	(12/26)	(7/24)	(8/18)
5	(5/16)	(5/11)	(11/23)	(5/13)
6	(12/19)	(14/20)	(11/24)	(8/22)
7	(12/18)	(10/21)	(8/13)	(8/17)
8	(9/20)	(7/12)	(9/16)	(6/10)
9	(6/12)	(5/9)	(1/6)	(5/11)
10	(2/4)	(3/7)	(5/9)	(7/17)
11	(7/8)	(8/11)	(6/7)	(3/6)
12	(6/10)	(5/7)	(6/10)	(3/6)
13	(3/4)	(2/3)	(4/7)	(4/5)
14	(2/3)	(4/5)	(0/1)	(2/4)
15	(2/2)	(3/5)	(4/4)	(2/3)

Πίνακας 2-7 : Ποσοστό ευστοχίας στις προβλέψεις μας βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2011-12 έως 2013-14)

Διαφορά Θέσεων RPI Weighted	2011-2012	2012-2013	2013-2014
0	(1/7)	(4/7)	(1/7)
1	(7/17)	(8/18)	(9/27)
2	(15/32)	(5/15)	(14/26)
3	(9/23)	(10/27)	(9/21)
4	(6/11)	(7/16)	(12/25)
5	(8/21)	(14/25)	(12/23)
6	(10/20)	(11/18)	(9/18)
7	(10/13)	(7/13)	(11/23)
8	(7/11)	(5/12)	(8/14)
9	(5/11)	(7/12)	(7/14)
10	(4/7)	(6/11)	(5/8)
11	(8/8)	(5/13)	(6/9)
12	(5/10)	(1/2)	(4/10)
13	(3/4)	(4/6)	(9/17)
14	(1/2)	(2/3)	(2/4)
15	(1/3)	(2/2)	(5/7)
16	-	-	(2/5)
17	-	-	(3/3)

Όσον αφορά την ευστοχία της μεθόδου παρατηρούμε πως για διαφορά θέσεων κατάταξης από 0 έως 7 το ποσοστό ευστοχίας είναι 48.08% για τη σεζόν 2007-2008, 39.89% για τη σεζόν 2008-2009, 44.6% για τη σεζόν 2009-2010, 47.43% για τη σεζόν 2010-2011, 45.13% για τη σεζόν 2011-2012, 48.33% για τη σεζόν 2012-2013 και 42.79% για τη σεζόν 2013-2014. Συνεπώς, η ευστοχία είναι μικρότερη από 50% για κάθε χρονιά για διαφορά θέσεων από 0 έως 7. Αντίστοιχα, για διαφορά θέσεων κατάταξης από 8 έως 15 και από 8 έως 17 για τη σεζόν 2013-2014 το ποσοστό ευστοχίας είναι 66.77% για τη σεζόν 2007-2008, 63.45% για τη σεζόν 2008-2009, 53.92% για τη σεζόν 2009-2010, 55.41% για τη σεζόν 2010-2011, 59.32% για τη σεζόν 2011-2012, 59.54% για τη σεζόν 2012-2013 και 59.07% για τη σεζόν 2013-2014. Επομένως, η ευστοχία είναι μεγαλύτερη από 50% για κάθε χρονιά και για κάποιες μάλιστα υπερβαίνει το 60% για διαφορά θέσεων από 8 έως 15 και από 8 έως 17 για τη σεζόν 2013-2014. Άρα, μπορούμε να πούμε πως, όπως ήταν αναμενόμενο, όσο μεγαλύτερη είναι η διαφορά δυναμικότητας των ομάδων βάσει της μεθόδου μας, τόσο καλύτερα, κατά μέσο όρο, αποτελέσματα ακρίβειας έχουμε στις προβλέψεις μας.

Πίνακας 2-8 : Ποσοστό αστοχίας στις ισοπαλίες βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2007-8 έως 2010-11)

Διαφορά Θέσεων RPI Weighted	2007-2008	2008-2009	2009-2010	2010-2011
0	(2/3)	(5/7)	(8/10)	(2/7)
1	(8/21)	(8/21)	(7/22)	(6/20)
2	(5/23)	(9/19)	(2/11)	(6/22)
3	(5/21)	(4/16)	(4/13)	(5/19)
4	(2/16)	(10/26)	(5/24)	(3/18)
5	(4/16)	(5/11)	(5/23)	(4/13)
6	(2/19)	(3/20)	(6/24)	(7/22)
7	(6/18)	(7/21)	(2/13)	(5/17)
8	(7/20)	(3/12)	(3/16)	(3/10)
9	(2/12)	(2/9)	(2/6)	(1/11)
10	(0/4)	(3/7)	(1/9)	(6/17)
11	(1/8)	(2/11)	(1/7)	(1/6)
12	(1/10)	(1/7)	(1/10)	(2/6)
13	(1/4)	(1/3)	(3/7)	(1/5)
14	(1/3)	(1/5)	(1/1)	(1/4)
15	(0/2)	(2/5)	(0/4)	(0/3)

Πίνακας 2-9 : Ποσοστό αστοχίας στις ισοπαλίες βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2011-12 έως 2013-14)

Διαφορά Θέσεων RPI Weighted	2011-2012	2012-2013	2013-2014
0	(6/7)	(3/7)	(6/7)
1	(6/17)	(7/18)	(10/27)
2	(6/32)	(4/15)	(7/26)
3	(6/23)	(10/27)	(4/21)
4	(2/11)	(5/16)	(1/25)
5	(7/21)	(3/25)	(3/23)
6	(3/20)	(3/18)	(3/18)
7	(3/13)	(3/13)	(5/23)
8	(1/11)	(1/12)	(2/14)
9	(3/11)	(3/12)	(2/14)
10	(1/7)	(2/11)	(1/8)
11	(0/8)	(5/13)	(3/9)
12	(5/10)	(0/2)	(2/10)
13	(1/4)	(2/6)	(5/17)
14	(0/2)	(1/3)	(2/4)
15	(1/3)	(0/2)	(1/7)
16	-	-	(3/5)
17	-	-	(0/3)

Όσον αφορά στην αστοχία των προβλέψεών μας στην περίπτωση ισοπαλίας, παρατηρούμε πως για διαφορά θέσεων κατάταξης από 0 έως 7 το ποσοστό αστοχίας είναι 28.96% για τη σεζόν 2007-2008, 39.27% για τη σεζόν 2008-2009, 30.47% για τη σεζόν 2009-2010, 27.6% για τη σεζόν 2010-2011, 31.93% για τη σεζόν 2011-2012, 28.56% για τη σεζόν 2012-2013 και 28.02% για τη σεζόν 2013-2014. Συνεπώς, η αστοχία είναι μικρότερη από 40% για κάθε χρονιά και ειδικότερα για τις σεζόν 2007-2008, 2010-2011, 2012-2013 και 2013-2014 το ποσοστό αυτό είναι μικρότερο από 30% για διαφορά θέσεων από 0 έως 7. Αντίστοιχα, για διαφορά θέσεων κατάταξης από 8 έως 15 και από 8 έως 17 για τη σεζόν 2013-2014, το ποσοστό αστοχίας είναι 16.56% για τη σεζόν 2007-2008, 26.99% για τη σεζόν 2008-2009, 28.79% για τη σεζόν 2009-2010, 21.17% για τη σεζόν 2010-2011, 19.87% για τη σεζόν 2011-2012, 19.87% για τη σεζόν 2012-2013 και 24.81% για τη σεζόν 2013-2014. Επομένως, η αστοχία είναι μικρότερη από 30% για κάθε χρονιά και μάλιστα για τις χρονιές 2007-2008, 2011-2012 και 2012-2013 το ποσοστό αυτό είναι μικρότερο από 20% για διαφορά θέσεων από 8 έως 15. Άρα, μπορούμε να πούμε πως η αύξηση της διαφοράς των θέσεων μας παρέχει κατά μέσο όρο μικρότερο ποσοστό αστοχίας στις ισοπαλίες.

Πίνακας 2-10 : Ποσοστό αστοχίας στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη, βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2007-8 έως 2010-11)

Διαφορά Θέσεων RPI Weighted	2007-2008	2008-2009	2009-2010	2010-2011
1	(4/21)	(6/21)	(7/22)	(5/20)
2	(8/23)	(8/19)	(3/11)	(4/22)
3	(6/21)	(6/16)	(1/13)	(6/19)
4	(5/16)	(4/26)	(12/24)	(7/18)
5	(7/16)	(1/11)	(7/23)	(4/13)
6	(5/19)	(3/20)	(7/24)	(7/22)
7	(0/18)	(4/21)	(3/13)	(4/17)
8	(4/20)	(2/12)	(4/16)	(1/10)
9	(4/12)	(2/9)	(3/6)	(5/11)
10	(2/4)	(1/7)	(3/9)	(4/17)
11	(0/8)	(1/11)	(0/7)	(2/6)
12	(3/10)	(1/7)	(3/10)	(1/6)
13	(0/4)	(0/3)	(0/7)	(0/5)
14	(0/3)	(0/5)	(0/1)	(1/4)
15	(0/2)	(0/5)	(0/4)	(1/3)

Πίνακας 2-11 : Ποσοστό αστοχίας στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη, βάσει της διαφοράς των θέσεων κατάταξης των ομάδων (για τις σεζόν 2011-12 έως 2013-14)

Διαφορά Θέσεων RPI Weighted	2011-2012	2012-2013	2013-2014
1	(4/17)	(3/18)	(8/27)
2	(11/32)	(6/15)	(5/26)
3	(8/23)	(7/27)	(8/21)
4	(3/11)	(4/16)	(12/25)
5	(6/21)	(8/25)	(8/23)
6	(7/20)	(4/18)	(6/18)
7	(0/13)	(3/13)	(7/23)
8	(3/11)	(6/12)	(4/14)
9	(3/11)	(2/12)	(5/14)
10	(2/7)	(3/11)	(2/8)
11	(0/8)	(3/13)	(0/9)
12	(0/10)	(1/2)	(4/10)
13	(0/4)	(0/6)	(3/17)
14	(1/2)	(0/3)	(0/4)
15	(1/3)	(0/2)	(1/7)
16	-	-	(0/5)
17	-	-	(0/3)

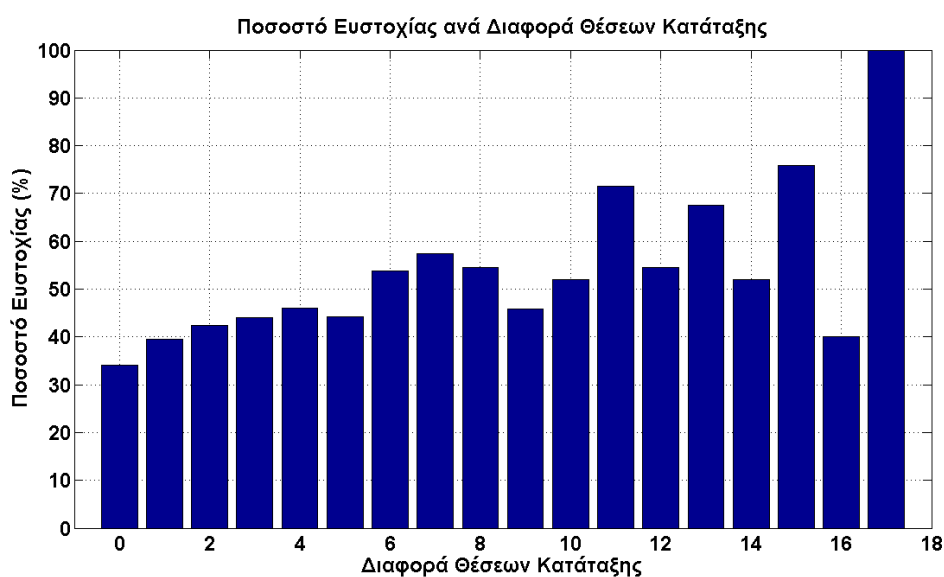
Όσον αφορά την αστοχία στους αγώνες που η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη παρατηρούμε πως για διαφορά θέσεων κατάταξης από 1 έως 8 το ποσοστό αστοχίας είναι 25.46% για τη σεζόν 2007-2008, 22.92% για τη σεζόν 2008-2009, 28.06% για τη σεζόν 2009-2010, 26.22% για τη σεζόν 2010-2011, 26.35% για τη σεζόν 2011-2012, 29.36% για τη σεζόν 2012-2013 και 32.76% για τη σεζόν 2013-2014. Συνεπώς, η αστοχία είναι μικρότερη από 30% για κάθε χρονιά, με εξαίρεση τη σεζόν 2013-2014, όπου το ποσοστό αυτό είναι 32.76% για διαφορά θέσεων από 1 έως 8. Αντίστοιχα, για διαφορά θέσεων κατάταξης από 9 έως 15 και από 9 έως 17 για τη σεζόν 2013-2014, το ποσοστό αστοχίας είναι 16.19% για τη σεζόν 2007-2008, 8.55% για τη σεζόν 2008-2009, 16.19% για τη σεζόν 2009-2010, 25.33% για τη σεζόν 2010-2011, 19.88% για τη σεζόν 2011-2012, 16.72% για τη σεζόν 2012-2013 και 14.74% για τη σεζόν 2013-2014. Επομένως, η αστοχία είναι μικρότερη από 20% για κάθε χρονιά, με εξαίρεση τη σεζόν 2010-2011, όπου το ποσοστό είναι 25.33%. Μάλιστα για τη χρονιά 2008-2009 το ποσοστό αυτό είναι μικρότερο από 9% για διαφορά θέσεων από 9 έως 15. Άρα, μπορούμε να πούμε πως η αύξηση της διαφοράς των θέσεων παρέχει κατά μέσο όρο μικρότερο ποσοστό αστοχίας.

Τα παραπάνω αποτελέσματα επιβεβαιώνουν ένα σημαντικό εύρημα της μεθόδου μας, που είχε διαφανεί και από τα αποτελέσματα του Πίνακα 2-5, ότι δηλαδή η μέθοδός μας μπορεί να προβλέψει με πολύ σημαντική ακρίβεια ποια ομάδα δεν θα χάσει (δηλαδή θα νικήσει ή θα φέρει ισοπαλία). Θεωρητικά, επιλέγοντας απολύτως τυχαία το τελικό αποτέλεσμα του αγώνα, το ενδεχόμενο σωστής πρόβλεψης για τα ποια ομάδα δεν θα χάσει θα ήταν ίσο με $2/3=66.6\%$. Όμως από τον Πίνακα 2-5

φαίνεται ότι η υλοποίηση της μεθόδου μας οδηγεί σε αντίστοιχη ακρίβεια 75.23%.

Από τα αποτελέσματα που παρουσιάζονται στους πίνακες 2-6 έως 2-11 παρατηρούμε πως οι περισσότεροι αγώνες λαμβάνουν χώρα μεταξύ των ομάδων που έχουν διαφορά θέσεων από 1 έως 8. Ειδικότερα, για τη σεζόν 2007-2008 σημειώνονται 154 αγώνες με διαφορά των ομάδων από 1 έως 8 θέσεις, τη σεζόν 2008-2009 διεξάγονται αγώνες 146 , τη χρονιά 2009-2010 γίνονται 146 παιχνίδια, τη χρονιά 2010-2011 141 παιχνίδια, τη σεζόν 2011-2012 148 αγώνες, τη σεζόν 2012-2013 144 παιχνίδια και τέλος τη σεζόν 2013-2014 σημειώνονται 177 αγώνες. Οι αγώνες μεταξύ ομάδων με διαφορά θέσεων μεγαλύτερη των 8, παρ' ότι αποτελούν το ήμισυ του συνόλου των πιθανών διαφορών (και λίγο μεγαλύτερο από το ήμισυ για την αγωνιστική σεζόν 2013-2014), είναι στην πράξη πολύ λιγότεροι. Ειδικότερα, διεξήχθησαν 46 αγώνες για το έτος 2007-2008, 54 αγώνες για το έτος 2008-2009, 54 παιχνίδια για τη χρονιά 2009-2010, 59 παιχνίδια για τη χρονιά 2010-2011, 52 αγώνες για τη χρονιά 2011-2012, 56 παιχνίδια για τη χρονιά 2012-2013 και τέλος 84 αγώνες για τη χρονιά 2013-2014.

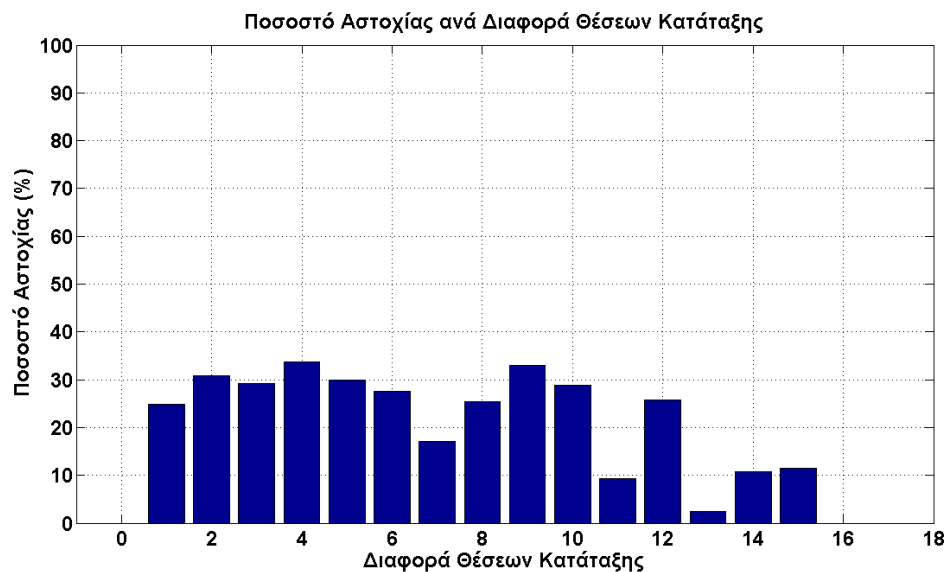
Στις γραφικές παραστάσεις που ακολουθούν καταγράφηκε κατά μέσο όρο στη διάρκεια των 7 εξεταζόμενων ετών το ποσοστό ευστοχίας και το ποσοστό αστοχίας στους αγώνες που η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη βάσει της διαφοράς θέσεων κατάταξής τους.



Εικόνα 2-17 : Μέσο ποσοστό ευστοχίας στη διάρκεια των 7 εξεταζόμενων ετών

Από την εικόνα 2-17 παρατηρούμε πως η ευστοχία είναι μεγαλύτερη από 40% για όλες τις διαφορές θέσεων κατάταξης, με εξαίρεση τη διαφορά θέσεων 0 (ίδια θέση ομάδων), όπου η ευστοχία μας είναι κοντά στο 35%. Μάλιστα για διαφορά

θέσεων 6,7,8,10,11,12,13,14,15 η ακρίβεια στις προβλέψεις μας είναι μεγαλύτερη από 50% και για τη μέγιστη διαφορά 17 θέσεων η ακρίβεια είναι 100%.



Εικόνα 2-18 : Μέσο ποσοστό αστοχίας, στους αγώνες όπου η εκτιμώμενη νικήτρια ομάδα ήταν η ηττημένη, στη διάρκεια των 7 εξεταζόμενων ετών

Από την εικόνα 2-18 παρατηρούμε πως για διαφορές θέσεων 16-17 η μέθοδός μας δεν αστοχεί, δηλαδή ποτέ δεν κερδίζει η σημαντικά ασθενέστερη ομάδα. Επιπλέον, για διαφορά θέσεων κατάταξης 7,11,13,14,15 το ποσοστό αστοχίας είναι μικρότερο από 18% και ιδίως για διαφορά 11 θέσεων το ποσοστό αυτό είναι μικρότερο από 5%. Τέλος για όλες τις υπόλοιπες διαφορές θέσεων το ποσοστό αστοχίας κυμαίνεται από 25% έως 35%.

2.6 Πλεονεκτήματα/Μειονεκτήματα των συστημάτων κατάταξης

Πλεονεκτήματα

1. Οι αξιολογήσεις των ομάδων γίνονται αντικειμενικά, χωρίς προκαταλήψεις υπέρ ή κατά κάποιων ομάδων ή παιχτών.
2. Οι αξιολογήσεις «έχουν μνήμη», λαμβάνουν δηλαδή υπ' όψιν τους και τους παλιότερους αγώνες της κάθε ομάδας και όχι μόνο τα αποτελέσματα που σημείωσε η κάθε μια στην προηγούμενη αγωνιστική.
3. Οι αξιολογήσεις στην n- οστή αγωνιστική στηρίζονται στα αποτελέσματα που έχουν σημειώσει οι ομάδες μέχρι και την (n-1)-οστή αγωνιστική, οπότε με την πάροδο των αγωνιστικών περιλαμβάνουν μια πλήρη εικόνα των ομάδων.

Μειονεκτήματα

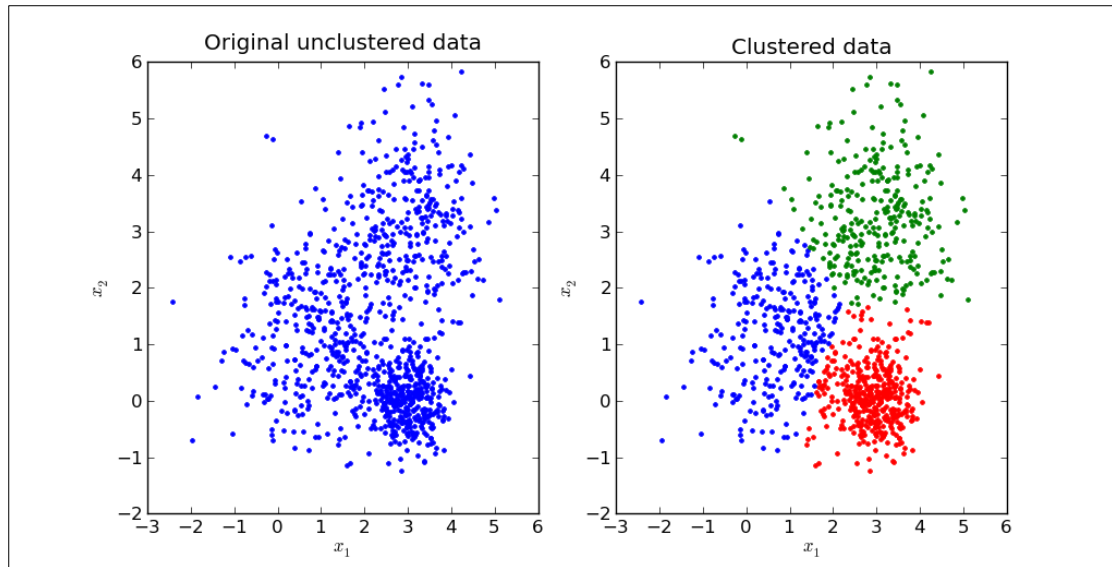
1. Κάθε μοντέλο περιλαμβάνει διαφορετικά κριτήρια, τα οποία επιλέγονται υποκειμενικά. Κάποιες περιλαμβάνουν μόνο τα αποτελέσματα των ομάδων, άλλες περιλαμβάνουν και επιπλέον κριτήρια, όπως το ακριβές σκορ και το εύρος της νίκης.
2. Οι περισσότερες μέθοδοι αγνοούν ποιοτικά κριτήρια, όπως οι καιρικές συνθήκες, οι τραυματισμοί ή οι αλλαγές παιχτών, οι ατομικές προσπάθειες κ.α.
3. Τα συστήματα αξιολόγησης δε μπορούν να ξεχωρίσουν τους αγώνες, ανάλογα με την ύπαρξη ή μη κινήτρου δίνοντας για παράδειγμα μικρότερη βαρύτητα σε μια ομάδα που έχει εξασφαλίσει το πρωτάθλημα και επομένως δεν την ενδιαφέρουν ιδιαίτερα τα επόμενα παιχνίδια και αντίστοιχα μεγαλύτερη σε κάποια που την αντιμετωπίζει και διεκδικεί την είσοδό της στου αγώνες των playoffs ή την παραμονή της στο πρωτάθλημα.

3. Συσταδοποίηση

3.1 Εισαγωγή στην Συσταδοποίηση

Με τον όρο ομαδοποίηση ή συσταδοποίηση (clustering) αναφερόμαστε στη διαδικασία που οργανώνει πρότυπα, παρατηρήσεις, δεδομένα ή διανύσματα χαρακτηριστικών σε ομάδες ή συστάδες (clusters), όπου τα μέλη μιας ομάδας είναι παρόμοια μεταξύ τους σύμφωνα με κάποιο κριτήριο. Σκοπός είναι να προσδιοριστούν οι ομάδες στις οποίες ανήκουν διάφορες ποσότητες δεδομένων, με βάση κάποια κριτήρια ομοιογένειας [22]. Η ποικιλία των εφαρμογών της ομαδοποίησης είναι μεγάλη σε πολλά επιστημονικά πεδία, καθώς συντελεί στην εξαγωγή χρήσιμων συμπερασμάτων. Μερικές βασικές εφαρμογές της αφορούν στο χώρο της βιολογίας και ειδικότερα στην ομαδοποίηση γονιδίων και πρωτεϊνών που έχουν ίδια λειτουργία ή φυτών και ζώων με παρόμοια χαρακτηριστικά, στο χώρο της οικονομίας και συγκεκριμένα στην ομαδοποίηση μετοχών με παρόμοια διακύμανση τιμών, στο κλάδο της ιατρικής και κυρίως στην ομαδοποίηση χαρακτηριστικών διαφόρων ασθενειών, στο χώρο της πληροφορικής και πιο συγκεκριμένα στην ομαδοποίηση weblog για την εύρεση παρόμοιων προτύπων προσπέλασης και στο κλάδο του marketing και πιο ειδικά στην ομαδοποίηση πελατών με παρόμοια συμπεριφορά. Είναι μια κοινή τεχνική για την ανάλυση των στατιστικών στοιχείων, που χρησιμοποιούνται σε πολλούς τομείς, συμπεριλαμβανομένης της μηχανικής μάθησης, της αναγνώρισης προτύπων, της ανάλυσης εικόνων, της ανάκτησης πληροφοριών και της βιοπληροφορικής. Η τεχνική της ομαδοποίησης υπάγεται στην ευρύτερη κατηγορία των τεχνικών μάθησης χωρίς επίβλεψη. Η διαφορά της ομαδοποίησης δεδομένων (data clustering) από την ταξινόμηση δεδομένων (data classification) είναι ότι, στην ταξινόμηση οι ομάδες στις οποίες θα τοποθετηθούν τα δεδομένα είναι προκαθορισμένες. Αυτό σημαίνει, ότι είναι εκ των προτέρων γνωστός ο αριθμός των ομάδων, τα ονόματα και οι ταυτότητες τους. Είναι και αυτό ένα σύστημα μάθησης μιας και οι ετικέτες που δίνονται από τα διαθέσιμα πρότυπα χρησιμοποιούνται, ώστε να μάθει το σύστημα ταξινόμησης την περιγραφή κάθε κλάσης και να είναι σε θέση να ταξινομήσει ένα νέο πρότυπο. Αντίθετα, στην ομαδοποίηση δεδομένων τονίζεται ιδιαίτερα ότι οι ομάδες δεν προϋπάρχουν αλλά αποφασίζονται από τον αλγόριθμο κατά δυναμικό τρόπο. Στην ομαδοποίηση δεδομένων δηλαδή, υπάρχει ένα σύνολο δεδομένων το οποίο πρέπει να διαχειριστούμε ώστε από αυτό να προκύψουν δυναμικά οι ομάδες. Σκοπός είναι να δημιουργηθούν ομάδες, που η καθεμία από αυτές θα συγκεντρώνει ομοιογενή στοιχεία. Κάθε μία από αυτές τις ομάδες διατηρεί ένα κέντρο, συνήθως το πιο κεντρικό στοιχείο της. Μια συσταδοποίηση θεωρείται καλή, όταν παράγει συστάδες καλής ποιότητας, δηλαδή όταν υπάρχει μεγάλη ομοιότητα εντός της συστάδας μεταξύ των αντικειμένων και μικρή ομοιότητα ανάμεσα στις συστάδες [23].

Ένα παράδειγμα ομαδοποίησης δίνεται στην παρακάτω εικόνα, όπου αριστερά παρουσιάζεται το αρχικό σύνολο των στοιχείων πριν την ομαδοποίηση και δεξιά η καταχώρηση των στοιχείων σε 3 ομάδες.



Εικόνα 3-1 : Ομαδοποίηση δεδομένων

3.2 Βήματα Συσταδοποίησης

Για την επίλυση ενός προβλήματος ομαδοποίησης δεδομένων συνήθως ακολουθούνται τα παρακάτω βήματα [22,24]:

- Αναπαράσταση των προτύπων
- Καθορισμός μιας μετρικής, ενδεικτικής της γειννίαςης των προτύπων, ανάλογα με τον τύπο δεδομένων
- Τεχνική ομαδοποίησης των δεδομένων
- Αφαίρεση δεδομένων -αν χρειαστεί-
- Αξιολόγηση του τελικού αποτελέσματος

Ένα χαρακτηριστικό της ομαδοποίησης είναι η επανατροφοδότηση, δηλαδή το αποτέλεσμα, το οποίο προκύπτει από την ομαδοποίηση επανατροφοδοτείται στο σύστημα, το οποίο συνδυάζοντας το αποτέλεσμα αυτό με τις υπόλοιπες εισόδους, προχωράει στην εξαγωγή χαρακτηριστικών και στους υπολογισμούς των σχέσεων ομοιότητας, με στόχο την τελική εξαγωγή των ομάδων.

Η αναπαράσταση των προτύπων αναφέρεται στο πλήθος των κλάσεων, το πλήθος των διαθέσιμων προτύπων και το πλήθος, τον τύπο και την κλίμακα των χαρακτηριστικών που είναι διαθέσιμα στον συγκεκριμένο αλγόριθμο ομαδοποίησης. Ωστόσο, μερικά από τα προηγούμενα δεν είναι πάντα άμεσα διαθέσιμα. Ενδιαφέρον παρουσιάζει η διαδικασία της επιλογής χαρακτηριστικών (features selection), κατά την οποία επιλέγονται τα πιο σημαντικά χαρακτηριστικά των στοιχείων τα οποία θα χρησιμοποιηθούν στη συσταδοποίηση. Η διαδικασία της εξαγωγής χαρακτηριστικών (features extraction) χρησιμοποιεί έναν ή περισσότερους μετασχηματισμούς των χαρακτηριστικών εισόδου, για την παραγωγή άλλων νέων, τα οποία πιθανόν να είναι πιο ενδιαφέροντα. Οποιαδήποτε από τις τεχνικές αυτές, μπορεί να χρησιμοποιηθεί για τη δημιουργία ενός συνόλου με τα πιο κατάλληλα χαρακτηριστικά, που θα

χρησιμοποιηθεί για την αναπαράσταση των στοιχείων που προορίζονται για ομαδοποίηση.

Η γειτνίαση των προτύπων συνήθως μετριέται με βάση μία συνάρτηση απόστασης που ορίζεται για ζεύγη προτύπων. Συνήθως, χρησιμοποιείται η Ευκλείδεια απόσταση. Η συνάρτηση απόστασης, η οποία επιλέγεται, αποτελεί κάθε φορά το μέτρο της ομοιότητας μεταξύ των προτύπων. Με βάση αυτό το μέτρο γίνεται η καταχώρηση τους στην ίδια ή σε διαφορετικές ομάδες. Το πόσο επιτυχημένο θεωρείται το αποτέλεσμα της ομαδοποίησης δεδομένων [25,26], εξαρτάται από τα κριτήρια που θα χρησιμοποιηθούν για τον διαχωρισμό των στοιχείων σε ομάδες. Η σωστή επιλογή των κριτηρίων αυτών είναι ένα πολύ σημαντικό ζήτημα.

Η τεχνική της συσταδοποίησης μπορεί να υλοποιηθεί με πολλούς τρόπους. Υπάρχουν πολλοί αλγόριθμοι ομαδοποίησης, καθένας από τους οποίους μπορεί να έχει διαφορετικό αποτέλεσμα είτε αυστηρό, είτε ασαφές. Το στάδιο αυτό, η επιλογή δηλαδή του κατάλληλου αλγορίθμου, αποτελεί και το κυρίως μέρος της όλης διαδικασίας της συσταδοποίησης. Φυσικά το μέτρο ομοιότητας που επιλέχτηκε παραπάνω θα χρησιμοποιηθεί από τον αλγόριθμο που θα επιλεγεί. Οι αλγόριθμοι που επιλύουν προβλήματα ομαδοποίησης είναι πολλοί και στηρίζονται σε διαφορετικές τεχνικές. Η επιλογή εξαρτάται από τη μορφή των δεδομένων και από τον χρήστη.

Η αφαίρεση δεδομένων, είναι η διαδικασία η οποία έχει σαν αποτέλεσμα μια απλή και συμπαγή αναπαράσταση του συνόλου των δεδομένων. Ο όρος απλή αναπαράσταση μπορεί να εξηγηθεί είτε από την οπτική γωνία της αυτοματοποιημένης ανάλυσης είτε από την οπτική γωνία του ανθρώπου. Στην πρώτη περίπτωση, το επιθυμητό για τα δεδομένα είναι να αναπαρίστανται με τέτοιο σαφή και απλό τρόπο, ώστε μια περαιτέρω υπολογιστική επεξεργασία να είναι εξίσου εφικτή. Στη δεύτερη περίπτωση, η απλή αναπαράσταση των δεδομένων τα κάνει πιο κατανοητά στους ειδικούς που πρόκειται να τα επεξεργαστούν και να εξάγουν συμπεράσματα. Συνήθως, η αφαίρεση δεδομένων στο clustering είναι μια συνοπτική αναπαράσταση κάθε ομάδας μέσω κάποιου αντιπροσώπου-πρωτότυπου στοιχείου το οποίο καλείται centroid (κεντροειδής).

Τέλος, στην αξιολόγηση του αποτελέσματος ελέγχεται η εγκυρότητα των ομάδων. Εδώ εξετάζεται εάν το τελικό αποτέλεσμα του αλγορίθμου είναι επιτυχές οπότε ο αλγόριθμος θεωρείται αξιόπιστος. Πρακτικά εξετάζεται αν οι ομάδες είναι αντιπροσωπευτικές σε σχέση με τα σημεία που έπρεπε να ομαδοποιηθούν, αν τα σημεία τελικά τοποθετήθηκαν στις κατάλληλες ομάδες κ.ο.κ. Η αξιολόγηση συνήθως γίνεται συγκρίνοντας τη ληφθείσα δομή με μια δεδομένη εκ των προτέρων δομή. Για την ανάλυση που γίνεται σε αυτό το στάδιο, χρησιμοποιείται ένα συγκεκριμένο κριτήριο βελτιστοποίησης, ανάλογα με το πρόβλημα που αντιμετωπίζεται κάθε φορά.

3.3 Διαμεριστική Συσταδοποίηση (Partitional clustering)

Στην διαμεριστική ομαδοποίηση, ορίζονται εξ αρχής οι ομάδες και τοποθετούνται τα σημεία του συνόλου δεδομένων που είναι διαθέσιμα στις ομάδες αυτές. Στη συνέχεια επαναπροσδιορίζονται αυτές οι αναθέσεις έως ότου κάποιο κριτήριο τερματισμού εκπληρωθεί. Ουσιαστικά αν υπάρχει ένα σύνολο δεδομένων πλήθους N , κατασκευάζονται k ομάδες και στη συνέχεια βελτιστοποιούνται. Το κυρίως πρόβλημα αυτών των τεχνικών είναι η σωστή επιλογή των ομάδων και του αριθμού k . Όπως και προηγουμένως αφού έχουν πραγματοποιηθεί όλες οι αναθέσεις των σημείων στις ομάδες, επαναπροσδιορίζονται οι αναθέσεις αυτές. Αυτό μπορεί να συμβεί αν σε κάθε ομάδα αποδοθεί μια τιμή και σιγά σιγά ελαχιστοποιείται. Η τιμή αυτή μπορεί να αποδοθεί με το άθροισμα των τετραγωνικών αποστάσεων από το μέσο όρο στην κάθε ομάδα. Οι πιο δημοφιλείς αλγόριθμοι στην κατηγορία αυτή είναι οι Centroids αλγόριθμοι, όπου η κάθε ομάδα αναπαρίσταται με το κέντρο μάζας της.

Ο πιο κλασσικός αλγόριθμος της κατηγορίας αυτής είναι ο k -means [27]. Ο αλγόριθμος αυτός αναλύεται διεξοδικά στο κεφάλαιο 3.4 και είναι αυτός που υλοποιήθηκε στην παρούσα διπλωματική εργασία, για την συσταδοποίηση των ομάδων ποδοσφαίρου της Superleague με βάση την δυναμικότητά τους.

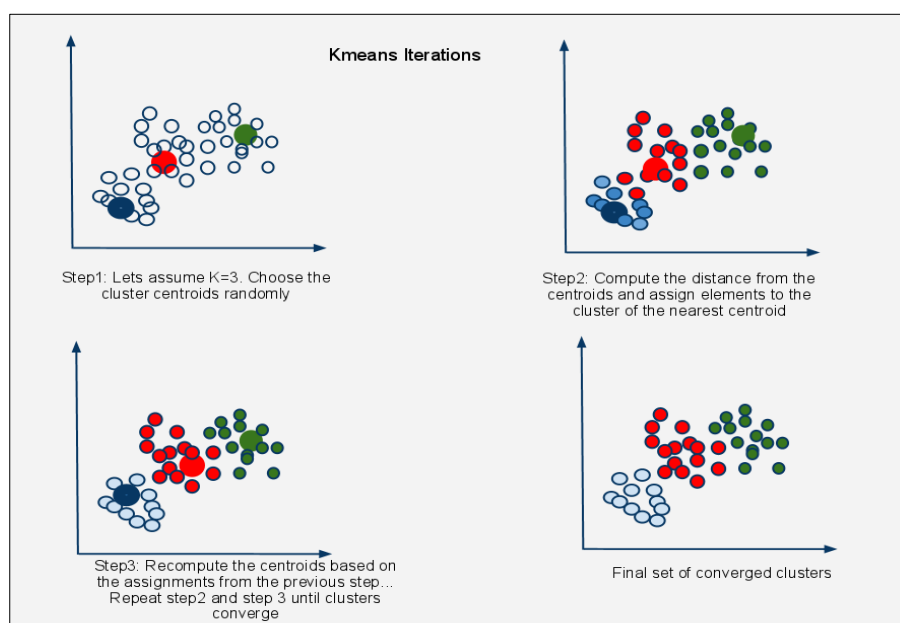
3.4 Ο αλγόριθμος k -means

Ο διαμεριστικός αλγόριθμος k -means [22, 28] είναι ένας από τους πιο απλούς και δημοφιλέστερους αλγορίθμους ομαδοποίησης που ανήκουν στην ευρύτερη κατηγορία των τεχνικών μάθησης χωρίς επίβλεψη. Ο αλγόριθμος αυτός είναι δημοφιλής εξαιτίας της απλότητας της υλοποίησης του και της γραμμικής πολυπλοκότητας του η οποία είναι της τάξης n ($O(n)$), όπου n το σύνολο των στοιχείων. Η διαδικασία της ομαδοποίησης ενός συνόλου δεδομένων με βάση τον k -means είναι εύκολη, αρκεί να είναι εκ των προτέρων καθορισμένος ο αριθμός k των clusters (ομάδων) που θα προκύψουν. Η κύρια ιδέα είναι να προσδιοριστούν αρχικά k centroids (κεντροειδή), ένα για κάθε cluster. Αυτά τα αρχικά centroids πρέπει να επιλεγούν κατάλληλα, γιατί διαφορετικές αρχικές θέσεις για τα centroids δίνουν διαφορετικά αποτελέσματα. Δηλαδή, η αρχική θέση των centroids επηρεάζει το αποτέλεσμα που θα δώσει ο αλγόριθμος. Έτσι, συχνά θεωρείται καλύτερη η επιλογή εκείνων των centroids ώστε να απέχουν μεταξύ τους όσο περισσότερο γίνεται. Το επόμενο βήμα είναι επιλογή κάθε στοιχείου από το σύνολο δεδομένων και η συσχέτιση του με το κοντινότερο σε αυτό centroid. Όταν αυτό γίνει για όλα τα στοιχεία του συνόλου δεδομένων, το πρώτο βήμα έχει ολοκληρωθεί και μία πρώτη και «πρόχειρη» ομαδοποίηση έχει ήδη προκύψει. Στη συνέχεια, απαιτείται να υπολογιστούν ξανά k νέα centroids, τα οποία θα αποτελούν το κέντρο βάρους για κάθε ένα cluster που προέκυψε από το προηγούμενο βήμα. Αφού λοιπόν οριστούν τα νέα k centroids, ακολουθεί και πάλι η ίδια διαδικασία ανάθεσης καθενός από τα στοιχεία του συνόλου δεδομένων στο κοντινότερο με αυτό, νέο πλέον, centroid. Έτσι, γίνεται μια επανάληψη της ίδιας διαδικασίας. Αποτέλεσμα αυτής της επανάληψης είναι ότι σε κάθε βήμα τα centroids αλλάζουν θέση (ορίζονται νέα) και τα στοιχεία ανατίθενται στο κατάλληλο cluster κάθε φορά με βάση το κοντινότερο centroid. Η επαναληπτική διαδικασία αυτή τερματίζει από τη στιγμή που θα ικανοποιηθεί ένα συγκεκριμένο κριτήριο. Ο αλγόριθμος υποθέτει ότι τα χαρακτηριστικά του αντικειμένου δημιουργούν ένα χώρο διανυσμάτων και ο σκοπός του είναι να

ελαχιστοποιήσει τη συνολική διακύμανση της ομάδας ή τη συνάρτηση τετραγωνικού σφάλματος.

Συνοπτικά, τα βασικά βήματα του αλγορίθμου είναι τα εξής:

1. Αρχικά γίνεται η επιλογή K σημείων στο πεδίο των δεδομένων έτσι ώστε τα σημεία αυτά να αποτελούν τα κέντρα των αρχικών ομάδων.
2. Ανάθεσε κάθε δεδομένο σε μία ομάδα για το οποίο η απόσταση του από το κέντρο της ομάδας να είναι η μικρότερη από κάθε άλλη από τα κέντρα των υπολοίπων ομάδων.
3. Όταν όλα τα δεδομένα έχουν ανατεθεί στις ομάδες υπολόγισε ξανά τα κέντρα των ομάδων, παίρνοντας τον μέσο όρο των δεδομένων κάθε ομάδας.
4. Επανάλαβε τα βήματα 2,3 μέχρι να επαληθευτεί κάποιο κριτήριο σύγκλισης.



Εικόνα 3-2 : Βήματα εκτέλεσης αλγορίθμου K-means

3.4.1 Επιλογή των χαρακτηριστικών

Όπως σε πολλά ζητήματα που αφορούν το clustering, δεν υπάρχουν σαφείς οδηγίες, που να οδηγούν στην βέλτιστη αναπαράσταση των στοιχείων και στην καλύτερη επιλογή χαρακτηριστικών. Ο ρόλος του ενδιαφερόμενου, στην αναπαράσταση των στοιχείων προς ομαδοποίηση, είναι να συγκεντρώσει όσο το δυνατόν περισσότερα στοιχεία (γνώσεις, συμπεράσματα, υποθέσεις) σχετικά με τα δεδομένα. Στην συνέχεια, με βάση την γνώση για τα δεδομένα, την κατάλληλη επιλογή και παραγωγή γνωρισμάτων, θα είναι σε θέση να πραγματοποιήσει την βέλτιστη αναπαράσταση για τα στοιχεία που θα επεξεργαστεί ο αλγόριθμος. Επίσης, μια πολύ προσεκτική ματιά στα σημαντικότερα χαρακτηριστικά μπορούν να οδηγήσουν σε καλύτερα αποτελέσματα. Μια καλή αναπαράσταση των στοιχείων συνήθως οδηγεί σε μια απλή και κατανοητή ομαδοποίηση. Αντίθετα, μια όχι και τόσο προσεκτική αναπαράσταση, μπορεί να οδηγήσει σε μια πολύπλοκη ομαδοποίηση με

δομές δύσκολα αντιληπτές και γενικά δυσδιάκριτες. Στην παρούσα εργασία επιλέξαμε 4 ποσοτικά χαρακτηριστικά {τον αριθμό νικών της κάθε ομάδας, τη διαφορά τερμάτων της, τη τρέχουσα βαθμολογία της και το RPI Weighted που διαθέτει}, τα οποία λαμβάνουν διακριτές τιμές και βάσει αυτών των χαρακτηριστικών εκτελείται η διαδικασία της συσταδοποίησης. Με τον όρο «διαφορά τερμάτων» εννοούμε το αποτέλεσμα που προκύπτει, είτε θετικό είτε αρνητικό, αν από τα γκολ που έχει επιτύχει η ομάδα αφαιρέσουμε τα γκολ που έχει δεχθεί.

3.4.2 Επιλογή του μέτρου ομοιότητας και του αριθμού των ομάδων

Το πόσο επιτυχημένο θεωρείται το αποτέλεσμα της ομαδοποίησης δεδομένων με τη χρήση του k-means εξαρτάται από τα κριτήρια που θα χρησιμοποιηθούν για τον διαχωρισμό των στοιχείων σε ομάδες. Το πιο σύνθηες κριτήριο που χρησιμοποιείται είναι η απόσταση μεταξύ των στοιχείων του συνόλου δεδομένων. Τα στοιχεία που ανήκουν σε κάθε cluster παρουσιάζουν ομοιότητα μεταξύ τους. Αυτή η ιδιότητα εξάλλου, είναι αναγκαία προκειμένου να ορισθεί ένα νέο ξεχωριστό cluster κατά την διαδικασία του clustering. Έτσι, για όλες τις τεχνικές clustering είναι σημαντικό να ορίζεται ένα μέτρο ομοιότητας μεταξύ δύο στοιχείων από το χώρο δεδομένων. Σε πολλές περιπτώσεις μάλιστα, με το μέτρο ομοιότητας, αυτό που συνήθως μετράται δεν είναι η ομοιότητα, αλλά η διαφορετικότητα δύο τυχαίων στοιχείων. Το μέτρο ομοιότητας καλείται και απόσταση και συμβολίζεται με D και ικανοποιεί τα παρακάτω για δύο στοιχεία x, y :

- $D(x, y) \geq 0$ (μη αρνητικό)
- $D(x, y) = D(y, x)$ (συμμετρία)
- $D(x, x) = 0$ (ένδειξη ταυτοποίησης)
- $D(x, y) = 0$ εάν $x = y$ (απολυτότητα)
- $D(x, y) \leq D(x, z) + D(z, y)$ (τριγωνική ανισότητα)

Το πιο γνωστό μέτρο ομοιότητας που χρησιμοποιείται και στην παρούσα διπλωματική εργασία είναι η Ευκλείδεια απόσταση η οποία ορίζεται ως εξής:

Έστω $x = (x_1, x_2, \dots, x_d)$ και $y = (y_1, y_2, \dots, y_d)$ δύο διανύσματα του χώρου R^d , τότε η Ευκλείδεια απόσταση είναι :

$$D(x, y) = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}$$

Προφανώς όσο μεγαλύτερο είναι το $D(x, y)$, τόσο αυξάνει η πιθανότητα τα δύο διανύσματα x και y να τοποθετηθούν σε διαφορετικές ομάδες (clusters).

Άλλος τρόπος για να υπολογιστεί η απόσταση (ανομοιότητα) μεταξύ δύο στοιχείων είναι η απόσταση Manhattan:

$$D(x, y) = \sum_{i=1}^d |x_i - y_i|$$

Ένας τρίτος τρόπος που χρησιμοποιείται στην βιβλιογραφία είναι το μέγιστο της διαφοράς σε κάθε διάσταση για τον υπολογισμό της απόστασης μεταξύ δύο στοιχείων:

$$D(x,y) = \max_{i=1}^d |x_i - y_i|$$

Στην παρούσα διπλωματική εργασία δοκιμάσαμε τις περιπτώσεις των $k=3,4,5,6$ clusters και η επιλογή του κατάλληλου k έγινε βάσει των εξαγόμενων αποτελεσμάτων.

3.5 Περιγραφή της μεθόδου υλοποίησης

3.5.1 Εφαρμογή για $k=3$ clusters

Χρησιμοποιούμε τα χαρακτηριστικά που αναφέραμε στην ενότητα 3.4.1 ξεκινώντας από αυτά που διαθέτουν οι ομάδες μετά το τέλος της 5^{ης} αγωνιστικής και έπειτα και ειδικότερα χρησιμοποιούμε 4 κριτήρια: τον αριθμό νικών της κάθε ομάδας, τη διαφορά τερμάτων της, τη τρέχουσα βαθμολογία της και το *RPI Weighted* που διαθέτει. Ο αριθμός των clusters που επιλέξαμε σε αυτή την υλοποίηση είναι $k=3$. Για την επιλογή των αρχικών centroids ακολουθήσαμε τη διαδικασία που προτείνεται στο [28] κι έτσι για κάθε ένα κριτήριο επιλέξαμε αντίστοιχα:

- $M1=[\mu_1-\sigma_1 \mu_1 \mu_1+\sigma_1]$
- $M2=[\mu_2-\sigma_2 \mu_2 \mu_2+\sigma_2]$
- $M3=[\mu_3-\sigma_3 \mu_3 \mu_3+\sigma_3]$
- $M4=[\mu_4-\sigma_4 \mu_4 \mu_4+\sigma_4]$

όπου $\mu_1, \mu_2, \mu_3, \mu_4$ είναι οι μέσες τιμές και $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ είναι οι τυπικές αποκλίσεις των στοιχείων κάθε κριτηρίου. Έπειτα, ακολουθούμε τα βήματα που αναφέρθηκαν στο κεφάλαιο 3.4, έως ότου δεν είναι δυνατή η μετακίνηση ομάδων σε διαφορετικά clusters.

Για κάθε ένα κριτήριο εφαρμόζουμε τον k-means clustering και βάσει αυτού προβλέπουμε τα αποτελέσματα της επόμενης αγωνιστικής. Ειδικότερα, για κάθε ένα από τα 4 κριτήρια ξεχωριστά, ομαδοποιούμε τις ομάδες σε κάποιο από τα 3 γκρουπ (clusters). Στη συνέχεια, εισάγουμε ένα σύστημα βαθμολόγησης των ομάδων και ειδικότερα για τις ομάδες που βρίσκονται στο γκρουπ Excellent δίνουμε βαρύτητα 3, για όσες βρίσκονται στο γκρουπ Good 2 και όσες βρίσκονται στο γκρουπ Bad έχουν βαρύτητα 1.

- Έτσι, αν μια ομάδα βρίσκεται στο γκρουπ Excellent και για τα τέσσερα κριτήρια, τότε διαθέτει τη μέγιστη βαθμολόγηση: $\frac{3+3+3+3}{4} = 3$.
- Αν βρίσκεται και για τα 4 κριτήρια στη Bad ομάδα, τότε διαθέτει τη μικρότερη βαθμολόγηση: $\frac{1+1+1+1}{4} = 1$.
- Οι πιθανές τιμές βαθμολόγησης που μπορεί να έχει κάποια ομάδα κυμαίνονται από 1 έως 3 και είναι συγκεκριμένα οι παρακάτω :

[1	1.25	1.5	1.75	2	2.25	2.5	2.75	3]
Bad			Good			Excellent		

Στη συνέχεια, βάσει της βαθμολόγησης ταξινομούμε τις ομάδες στα 3 γκρουπ και θεωρούμε πως οι ομάδες που βρίσκονται στο ίδιο γκρουπ θα έρθουν ισόπαλες στα μεταξύ τους παιχνίδια, θα κερδίσουν τις αντιπάλους τους που βρίσκονται σε κατώτερο γκρουπ και θα χάσουν από εκείνες που βρίσκονται σε ανώτερο.

3.5.2 Εφαρμογή για k=4 clusters

Χρησιμοποιούμε τα χαρακτηριστικά που αναφέραμε στην ενότητα 3.4.1 ξεκινώντας από αυτά που διαθέτουν οι ομάδες μετά το τέλος της 5^{ης} αγωνιστικής και έπειτα και ειδικότερα χρησιμοποιούμε 4 κριτήρια: τον αριθμό νικών της κάθε ομάδας, τη διαφορά τερμάτων της, τη τρέχουσα βαθμολογία της και το *RPI Weighted* που διαθέτει. Ο αριθμός των clusters που επιλέξαμε σε αυτή την υλοποίηση είναι k=4. Για την επιλογή των αρχικών centroids ακολουθήσαμε τη διαδικασία της αριθμητικής προόδου, η οποία διαθέτει δύο ισοδύναμους τύπους:

1. Ο γενικός τύπος, όπου ορίζεται ο νιοστός όρος συναρτήσει του πρώτου όρου και της διαφοράς :

$$a_n = a_1 + (n-1) \times \omega$$

2. Ο αναδρομικός τύπος, όπου ορίζεται ο n-οστός όρος συναρτήσει του προηγούμενου όρου και της διαφοράς :

$$a_n = a_{n-1} + \omega$$

Επιπλέον, για την αριθμητική πρόοδο a_n με $n \in \mathbb{N}^*$ και διαφορά ω ισχύει ότι :

- Αν $\omega > 0$, η πρόοδος a_n είναι γνησίως αύξουσα, δηλαδή ισχύει ότι $a_{n+1} > a_n$
- Αν $\omega < 0$, η πρόοδος a_n είναι γνησίως φθίνουσα, δηλαδή ισχύει ότι $a_{n+1} < a_n$
- Αν $\omega = 0$, η πρόοδος a_n είναι σταθερή

Εμείς εφαρμόσαμε τον αναδρομικό τύπο με $\omega = \sigma_1$. Έτσι, ως αρχικά centroids έχουμε επιλέξει για κάθε ένα κριτήριο αντίστοιχα :

- $M = [Ma1 \ Ma2 \ Ma3 \ Ma4]$
- $M2 = [Mb1 \ Mb2 \ Mb3 \ Mb4]$
- $M3 = [Mc1 \ Mc2 \ Mc3 \ Mc4]$
- $M4 = [Md1 \ Md2 \ Md3 \ Md4]$

όπου ο όρος Ma4 αντιστοιχεί στο μέγιστο αριθμό νικών που έχει επιτύχει κάποια ομάδα, ο όρος Mb4 είναι η μεγαλύτερη θετική διαφορά τερμάτων που έχει κάποια ομάδα, το Mc4 αντιστοιχεί στη μεγαλύτερη βαθμολογία που διαθέτει μια ομάδα και το Md4 είναι η μεγαλύτερη τιμή RPI Weighted που έχει κάποια ομάδα. Οι υπόλοιποι όροι προκύπτουν βάσει του αναδρομικού τύπου. Έπειτα, ακολουθούμε τα βήματα που αναφέρθηκαν στο κεφάλαιο 3.4, έως ότου δεν είναι δυνατή η μετακίνηση ομάδων σε διαφορετικά clusters.

Ακολουθώντας, για κάθε ένα κριτήριο εφαρμόζουμε τον k-means clustering και βάσει αυτού προβλέπουμε τα αποτελέσματα της επόμενης αγωνιστικής. Ειδικότερα, για κάθε ένα από τα 4 κριτήρια ξεχωριστά, ομαδοποιούμε τις ομάδες σε κάποιο από τα 4 γκρουπ (clusters). Στη συνέχεια, εισάγουμε ένα σύστημα βαθμολόγησης των ομάδων και ειδικότερα για τις ομάδες που βρίσκονται στο γκρουπ Excellent δίνουμε βαρύτητα 4, για όσες βρίσκονται στο γκρουπ Very Good 3, για όσες βρίσκονται στο γκρουπ Good 2 και όσες βρίσκονται στο γκρουπ Bad έχουν βαρύτητα 1.

- Έτσι, αν μια ομάδα βρίσκεται στο γκρουπ Excellent και για τα 4 κριτήρια, τότε διαθέτει τη μέγιστη βαθμολόγηση: $\frac{4+4+4+4}{4} = 4$.
- Αν βρίσκεται και για τα 4 κριτήρια στη Bad ομάδα, τότε διαθέτει τη μικρότερη βαθμολόγηση: $\frac{1+1+1+1}{4} = 1$.
- Οι πιθανές τιμές βαθμολόγησης που μπορεί να έχει κάποια ομάδα κυμαίνονται από 1 έως 4 και είναι συγκεκριμένα οι παρακάτω :

[1	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.25	3.5	3.75	4]
Bad			Good			Very Good			Excellent			

Στη συνέχεια, βάσει της βαθμολόγησης ταξινομούμε τις ομάδες στα 4 γκρουπ και θεωρούμε πως οι ομάδες που βρίσκονται στο ίδιο γκρουπ θα έρθουν ισόπαλες στα μεταξύ τους παιχνίδια, θα κερδίσουν τις αντιπάλους τους που βρίσκονται σε κατώτερο γκρουπ και θα χάσουν από εκείνες που βρίσκονται σε ανώτερο.

3.5.3 Εφαρμογή για k=5 clusters

Χρησιμοποιούμε τα χαρακτηριστικά που αναφέραμε στην ενότητα 3.4.1 ξεκινώντας από αυτά που διαθέτουν οι ομάδες μετά το τέλος της 5^{ης} αγωνιστικής και έπειτα και ειδικότερα χρησιμοποιούμε 4 κριτήρια: τον αριθμό νικών της κάθε ομάδας, τη διαφορά τερμάτων της, τη τρέχουσα βαθμολογία της και το RPI Weighted που διαθέτει. Ο αριθμός των clusters που επιλέξαμε σε αυτή την υλοποίηση είναι k=5. Για την επιλογή των αρχικών centroids εφαρμόσαμε τη διαδικασία της αριθμητικής

προόδου κάνοντας χρήση του αναδρομικού τύπου με $\omega = \frac{\sigma_1}{2}$. Έτσι, ως αρχικά centroids έχουμε επιλέξει για κάθε ένα κριτήριο αντίστοιχα:

- $M=[Ma1 \ Ma2 \ Ma3 \ Ma4 \ Ma5]$
- $M2=[Mb1 \ Mb2 \ Mb3 \ Mb4 \ Mb5]$
- $M3=[Mc1 \ Mc2 \ Mc3 \ Mc4 \ Mc5]$
- $M4=[Md1 \ Md2 \ Md3 \ Md4 \ Md5]$

όπου το Ma1 αντιστοιχεί στο μικρότερο αριθμό νικών που έχει επιτύχει κάποια ομάδα, το Mb1 αντιστοιχεί στη μεγαλύτερη αρνητική τιμή διαφοράς τερμάτων που έχει κάποια ομάδα, το Mc1 είναι η μικρότερη βαθμολογία κάποιας ομάδας και το Md1 είναι ο μικρότερος αριθμός RPI Weighted που έχει κάποια ομάδα. Οι υπόλοιποι όροι προκύπτουν βάσει του αναδρομικού τύπου. Έπειτα, ακολουθούμε τα βήματα που αναφέρθηκαν στο κεφάλαιο 3.4, έως ότου δεν είναι δυνατή η αλλαγή των ομάδων σε clusters.

Ακολούθως, για κάθε ένα κριτήριο εφαρμόζουμε τον k-means clustering και βάσει αυτού προβλέπουμε τα αποτελέσματα της επόμενης αγωνιστικής. Πιο συγκεκριμένα, για κάθε ένα από τα 4 κριτήρια ξεχωριστά, ομαδοποιούμε τις ομάδες σε κάποιο από τα 5 γκρουπ (clusters). Στη συνέχεια, εισάγουμε ένα σύστημα βαθμολόγησης των ομάδων και ειδικότερα για τις ομάδες που βρίσκονται στο 5^ο γκρουπ δίνουμε βαρύτητα 5, στο 4^ο γκρουπ δίνουμε βαρύτητα 4, για όσες βρίσκονται στο 3^ο γκρουπ 3, για όσες βρίσκονται στο 2^ο γκρουπ 2 και όσες βρίσκονται στο 1^ο γκρουπ έχουν βαρύτητα 1.

- Έτσι, αν μια ομάδα βρίσκεται στο 5^ο γκρουπ και για τα 4 κριτήρια, τότε διαθέτει τη μέγιστη βαθμολόγηση: $\frac{5+5+5+5}{4} = 5$.
- Αν βρίσκεται και για τα 4 κριτήρια στο 1^ο γκρουπ, τότε διαθέτει τη μικρότερη βαθμολόγηση: $\frac{1+1+1+1}{4} = 1$.
- Οι πιθανές τιμές βαθμολόγησης που μπορεί να έχει κάποια ομάδα κυμαίνονται από 1 έως 5 και είναι συγκεκριμένα οι παρακάτω :

[1	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	4.75	5]
1 st group			2 nd group			3 rd group			4 th group			5 th group				

Στη συνέχεια, βάσει της βαθμολόγησης ταξινομούμε τις ομάδες στα 5 γκρουπ και θεωρούμε πως οι ομάδες που βρίσκονται στο ίδιο γκρουπ θα έρθουν ισόπαλες στα μεταξύ τους παιχνίδια, θα κερδίσουν τις αντιπάλους τους που βρίσκονται σε κατώτερο γκρουπ και θα χάσουν από εκείνες που βρίσκονται σε ανώτερο.

3.5.4 Εφαρμογή για k=6 clusters

Χρησιμοποιούμε τα χαρακτηριστικά που αναφέραμε στην ενότητα 3.4.1 ξεκινώντας από αυτά που διαθέτουν οι ομάδες μετά το τέλος της 5^{ης} αγωνιστικής και

έπειτα και ειδικότερα χρησιμοποιούμε 4 κριτήρια: τον αριθμό νικών της κάθε ομάδας, τη διαφορά τερμάτων της, τη τρέχουσα βαθμολογία της και το *RPI Weighted* που διαθέτει. Ο αριθμός των clusters που επιλέξαμε σε αυτή την υλοποίηση είναι $k=6$. Για την επιλογή των αρχικών centroids ακολουθήσαμε τη διαδικασία της αριθμητικής προόδου εφαρμόζοντας τον αναδρομικό τύπο με $\omega = \frac{\sigma_1}{2}$. Έτσι, ως αρχικά centroids έχουμε επιλέξει για κάθε ένα κριτήριο αντίστοιχα :

- $M=[Ma1 \ Ma2 \ Ma3 \ Ma4 \ Ma5 \ Ma6]$
- $M2=[Mb1 \ Mb2 \ Mb3 \ Mb4 \ Mb5 \ Mb6]$
- $M3=[Mc1 \ Mc2 \ Mc3 \ Mc4 \ Mc5 \ Mc6]$
- $M4=[Md1 \ Md2 \ Md3 \ Md4 \ Md5 \ Md6]$

όπου το $Ma1$ αντιστοιχεί στο μικρότερο αριθμό νικών που έχει επιτύχει κάποια ομάδα, το $Mb1$ αντιστοιχεί στη μεγαλύτερη αρνητική τιμή διαφοράς τερμάτων που έχει κάποια ομάδα, το $Mc1$ είναι η μικρότερη βαθμολογία κάποιας ομάδας και το $Md1$ είναι ο μικρότερος αριθμός *RPI Weighted* που έχει κάποια ομάδα. Οι υπόλοιποι όροι προκύπτουν βάσει του αναδρομικού τύπου. Έπειτα, ακολουθούμε τα βήματα που αναφέρθηκαν στο κεφάλαιο 3.4, έως ότου δεν είναι δυνατή η αλλαγή των ομάδων σε clusters.

Ακολούθως, για κάθε ένα κριτήριο εφαρμόζουμε τον k-means clustering και βάσει αυτού προβλέπουμε τα αποτελέσματα της επόμενης αγωνιστικής. Πιο συγκεκριμένα, για κάθε ένα από τα 4 κριτήρια ξεχωριστά, ομαδοποιούμε τις ομάδες σε κάποιο από τα 6 γκρουπ (clusters). Στη συνέχεια, εισάγουμε ένα σύστημα βαθμολόγησης των ομάδων και ειδικότερα για τις ομάδες που βρίσκονται στο 6^ο γκρουπ δίνουμε βαρύτητα 6, για τις ομάδες που είναι στο 5^ο γκρουπ δίνουμε βαρύτητα 5, για όσες βρίσκονται στο 4^ο γκρουπ 4, για όσες βρίσκονται στο 3^ο γκρουπ 3, για όσες βρίσκονται στο 2^ο γκρουπ 2 και όσες βρίσκονται στο 1^ο γκρουπ έχουν βαρύτητα 1.

- Έτσι, αν μια ομάδα βρίσκεται στο 6^ο γκρουπ και για τα 4 κριτήρια, τότε διαθέτει τη μέγιστη βαθμολόγηση: $\frac{6+6+6+6}{4} = 6$.
- Αν βρίσκεται και για τα 4 κριτήρια στο 1^ο γκρουπ, τότε διαθέτει τη μικρότερη βαθμολόγηση: $\frac{1+1+1+1}{4} = 1$.
- Οι πιθανές τιμές βαθμολόγησης που μπορεί να έχει κάποια ομάδα κυμαίνονται από 1 έως 6 και είναι συγκεκριμένα οι παρακάτω :

[1	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	4.75	5	5.25	5.5	5.75	6]
1 st group			2 nd group			3 rd group			4 th group			5 th group			6 th group					

Στη συνέχεια, βάσει της βαθμολόγησης ταξινομούμε τις ομάδες στα 6 γκρουπ και θεωρούμε πως οι ομάδες που βρίσκονται στο ίδιο γκρουπ θα έρθουν ισόπαλες στα

μεταξύ τους παιχνίδια, θα κερδίσουν τις αντιπάλους τους που βρίσκονται σε κατώτερο γκρουπ και θα χάσουν από εκείνες που βρίσκονται σε ανώτερο.

3.6 Αποτελέσματα

Στους πίνακες που ακολουθούν παρουσιάζονται το μέσο ποσοστιαίο σφάλμα βαθμολογίας, το ποσοστό ευστοχίας στις προβλέψεις μας, το σφάλμα που έγινε στους αγώνες που είτε η εκτίμησή μας ήταν ισοπαλία και ο αγώνας τελείωσε με νικητή και ηττημένο, είτε το πραγματικό αποτέλεσμα ήταν ισοπαλία και η εκτίμησή μας ήταν διαφορετική και το σφάλμα στους αγώνες, όπου η ομάδα που εκτιμήσαμε ότι θα κερδίσει ήταν τελικά η ηττημένη.

Πίνακας 3-1 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
3 clusters	28.27	25.27	26.72	21.9	26.81	27.29	25.75	26
4 clusters	31.46	23.06	26.77	19.94	27.45	29.59	26.94	26.46
5 clusters	30.75	25.35	29.7	26.39	28.91	27.16	30.16	28.35
6 clusters	31.86	26.02	30.6	25.64	29.77	28.04	30.67	28.94

Πίνακας 3-2 : Ποσοστό ευστοχίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
3 clusters	49	45	42	41.5	42.5	48.5	41.8	44.33
4 clusters	46	51.5	42.5	43.5	50	45	40.6	45.59
5 clusters	50	47.5	44	43	48.5	46.5	46.7	46.6
6 clusters	50.5	47	47	43	49.5	47	46.4	47.2

Πίνακας 3-3: Ποσοστό σφάλματος ισοπαλιών

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
3 clusters	37	41.5	39.5	41.5	43	36	40.6	39.87
4 clusters	34.5	36.5	37	39.5	32	38.5	41.4	37.06
5 clusters	28	38	36	35.5	31	31.5	30.2	32.89
6 clusters	27	38.5	30.5	34.5	29.5	31	30.2	31.6

Πίνακας 3-4: Ποσοστό σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
3 clusters	14	13.5	18.5	17	14.5	15.5	17.6	15.8
4 clusters	19.5	12	20.5	17	18	17	18	17.43
5 clusters	22	14.5	20	21.5	20.5	22	23	20.5
6 clusters	22.5	14.5	22.5	22.5	21	22	23.4	21.2

Από τα εξαγόμενα αποτελέσματα παρατηρούμε πως η αύξηση του αριθμού των clusters οδηγεί σε αναμενόμενα αποτελέσματα. Από τη μία συμβάλει στη αύξηση του ποσοστού ευστοχίας και στη μείωση του ποσοστού αστοχίας στις ισοπαλίες, αφού η διαίρεση των ομάδων σε περισσότερα clusters δημιουργεί μεγαλύτερες «αποστάσεις ασφαλείας» μεταξύ τους, ώστε να κρίνουμε καλύτερα την δυναμικότητά τους. Ταυτόχρονα, προκαλεί αύξηση του ποσοστού σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη, αφού όσο λιγότερα είναι τα clusters, τόσο μικρότερη είναι η πιθανότητα να συμβεί ανατροπή στο τελικό αποτέλεσμα. Αυτές οι ανατροπές οδηγούν και στην μικρή αύξηση του σφάλματος υπολογισμού της τελικής βαθμολογίας όταν αυξάνεται ο αριθμός των clusters.

4. Στατιστική εκτίμηση του Σκορ

4.1 Κατανομή Poisson και Συντελεστής συσχέτισης Pearson

Μια συνηθισμένη εφαρμογή της κατανομής Poisson είναι η πρόβλεψη του αριθμού των γεγονότων που παρατηρούνται, υποθέτοντας βέβαια ότι οι συνθήκες του πειράματος παραμένουν σταθερές με τη πάροδο του χρόνου.

Μία διακριτή τυχαία μεταβλητή X ακολουθεί κατανομή Poisson με παράμετρο $\lambda > 0$, αν, για $k = 0, 1, 2, \dots$, η συνάρτηση πυκνότητας πιθανότητας του X δίνεται από τη σχέση:

$$f(k; \lambda) = \Pr(X=k) = \frac{\lambda^k \times e^{-\lambda}}{k!}$$

όπου e είναι ο αριθμός του Euler ($e = 2.718$) και $k!$ είναι το παραγοντικό του k . Ο θετικός πραγματικός αριθμός λ είναι ίσος με την αναμενόμενη τιμή του X και επίσης με την διακύμανση:

$$\lambda = E(X) = \text{Var}(X)$$

Στη στατιστική, ο συντελεστής συσχέτισης Pearson, ο οποίος αναφέρεται και ως PPMCC ή PCC ή Pearson's r , είναι ένα μέτρο γραμμικής συσχέτισης (εξάρτησης) μεταξύ δύο μεταβλητών X και Y . Αναπτύχθηκε από τον Karl Pearson [29], αν και η αρχική ιδέα εισήχθη από τον Francis Galton [30]. Ορίζεται με βάση ένα δείγμα n ζευγών παρατηρήσεων (x_i, y_i) με $i=1, 2, \dots, n$, συμβολίζεται με $r(X, Y)$ ή απλά με r και δίνεται από τον τύπο:

$$r = \frac{\sum_{i=1}^n (x_i - \mu_x) \times (y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \times \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}}$$

Από τον ορισμό του r παρατηρούμε ότι για μεγάλες τιμές x_i της X και y_i της Y (μεγαλύτερες από τη μέση τιμή τους) οι διαφορές $(x_i - \mu_x)$ και $(y_i - \mu_y)$ είναι θετικές, οπότε το γινόμενό τους είναι θετικό. Όμοια για μικρές τιμές x_i και y_i , οι διαφορές $(x_i - \mu_x)$ και $(y_i - \mu_y)$ είναι αρνητικές, οπότε το γινόμενό τους είναι πάλι θετικό. Επομένως, όταν σε μεγάλες τιμές της μεταβλητής X αντιστοιχούν και μεγάλες τιμές της Y , ή σε μικρές τιμές της X αντιστοιχούν μικρές τιμές της Y τότε ο r συντελεστής είναι θετικός και λέμε ότι οι X, Y είναι θετικά συσχετισμένες. Ανάλογα μπορούμε να δούμε ότι ο r παίρνει αρνητικές τιμές όταν σε μεγάλες τιμές της μιας μεταβλητής αντιστοιχούν μικρές τιμές της άλλης, οπότε λέμε ότι οι μεταβλητές αυτές είναι αρνητικά συσχετισμένες.

Ο συντελεστής συσχέτισης είναι καθαρός αριθμός, δηλαδή δεν εκφράζεται σε συγκεκριμένες μονάδες μέτρησης. Άρα, είναι ανεξάρτητος των χρησιμοποιούμενων μονάδων μέτρησης των μεταβλητών X και Y . Επιπλέον, ισχύει πάντοτε ότι:

$$-1 \leq r \leq +1$$

Πιο συγκεκριμένα, διακρίνονται οι παρακάτω περιπτώσεις :

- $0 < r < +1$: τότε οι X, Y είναι **θετικά** γραμμικά συσχετισμένες
- $-1 < r < 0$: τότε οι X, Y είναι **αρνητικά** γραμμικά συσχετισμένες
- $r=+1$: τότε έχουμε **τέλεια θετική γραμμική συσχέτιση** και όλα τα σημεία βρίσκονται πάνω σε μια ευθεία με θετική κλίση
- $r=-1$: τότε έχουμε **τέλεια αρνητική γραμμική συσχέτιση** και όλα τα σημεία βρίσκονται πάνω σε μια ευθεία με αρνητική κλίση
- $r=0$: τότε **δεν υπάρχει γραμμική συσχέτιση** μεταξύ των μεταβλητών. Οι μεταβλητές δηλαδή X, Y είναι **γραμμικά ασυσχέτιστες**

4.2 Περιγραφή της μεθόδου υλοποίησης

Για κάθε μια σεζόν υπολογίσαμε το συντελεστή συσχέτισης Pearson μεταξύ του συνολικού αριθμού γκολ των γηπεδούχων και του αντίστοιχου αριθμού των φιλοξενούμενων ομάδων [31]. Στον πίνακα που ακολουθεί φαίνεται η συσχέτισή τους για κάθε έτος :

Πίνακας 4-1 : Συντελεστής συσχέτισης Pearson

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14
συσχέτιση Pearson(r)	-0.055	+0.019	-0.028	-0.075	-0.059	-0.133	-0.128

Για όλες λοιπόν τις χρονιές, η συσχέτιση μεταξύ των γκολ των γηπεδούχων και των φιλοξενούμενων ομάδων είναι μικρή και μάλιστα αρνητική, άρα οι μεταβλητές X (γκολ γηπεδούχων) και Y (γκολ φιλοξενούμενων) είναι αρνητικά γραμμικά συσχετισμένες, με εξαίρεση τη χρονιά 2008-2009, όπου παρατηρείται μικρή και θετική συσχέτιση και κατά συνέπεια οι μεταβλητές X και Y είναι θετικά γραμμικά συσχετισμένες. Επομένως, από τον υπολογισμό του συντελεστή συσχέτισης Pearson παρατηρήσαμε πολύ μικρή συσχέτιση (σχεδόν μηδενική) μεταξύ των μεταβλητών X και Y.

Σύμφωνα με τον Maher [32] τα γκολ που σημειώνει η γηπεδούχος ομάδα i και η φιλοξενούμενη ομάδα j σε κάθε αγώνα (x_{ij}, y_{ij}) είναι δύο ανεξάρτητες κατανομές Poisson με μέση τιμή $a_i b_j$ και $c_i d_j$ αντίστοιχα, όπου το a_i αντιπροσωπεύει την επιθετική ικανότητα της γηπεδούχου ομάδας, όταν αγωνίζεται στην έδρα της, το b_j αντιστοιχεί στην αδυναμία της αμυντικής λειτουργίας της φιλοξενούμενης ομάδας, όταν αγωνίζεται εκτός έδρας, το c_i φανερώνει την αδυναμία της αμυντικής λειτουργίας της γηπεδούχου ομάδας, όταν αγωνίζεται εντός έδρας και το d_j αντιπροσωπεύει την επιθετική ικανότητα της φιλοξενούμενης ομάδας, όταν αγωνίζεται εκτός έδρας.

Εφαρμόζοντας λοιπόν παρόμοια λογική και θεωρώντας ως μέσο όρο για κάθε ομάδα (μ και λ , αντίστοιχα) τον μέσο όρο των γκολ που έχει σημειώσει μέχρι εκείνη την αγωνιστική ως γηπεδούχος (μ) και ως φιλοξενούμενη (λ), χρησιμοποιούμε την αντίστοιχη κατανομή Poisson για να υπολογίσουμε την πιθανότητα να σκοράρει η κάθε μια 0, 1, 2, 3, 4, 5 ή και περισσότερα από 5 γκολ.

Έτσι, για $k=0, 1, 2, 3, 4, 5$ οι πιθανότητες υπολογίζονται από την ακόλουθη σχέση:

$$\Pr(X=k) = \frac{\lambda^k \times e^{-\lambda}}{k!}$$

Για $k>5$ ισχύει ότι:

$$\Pr(X>k) = 1 - \sum_{i=0}^k \Pr(X = i)$$

Έπειτα, αφού έχουμε υπολογίσει τις πιθανότητες αυτές για κάθε ομάδα τις κατανέμουμε σε ένα εύρος τιμών από 0 έως 1 ως εξής :

[0,Pr(X=0)]
(Pr(X=0), Pr(X=0)+Pr(X=1)]
(Pr(X=0)+Pr(X=1), Pr(X=0)+Pr(X=1)+Pr(X=2)]
(Pr(X=0)+Pr(X=1)+Pr(X=2), Pr(X=0)+Pr(X=1)+Pr(X=2)+Pr(X=3)]
(Pr(X=0)+Pr(X=1)+Pr(X=2)+Pr(X=3), Pr(X=0)+Pr(X=1)+Pr(X=2)+Pr(X=3)+Pr(X=4)]
(Pr(X=0)+Pr(X=1)+Pr(X=2)+Pr(X=3)+Pr(X=4), Pr(X=0)+Pr(X=1)+Pr(X=2)+Pr(X=3)+Pr(X=4)+Pr(X=5)]
(Pr(X=0)+Pr(X=1)+Pr(X=2)+Pr(X=3)+Pr(X=4)+Pr(X=5), 1]

όπου το πρώτο εύρος καλύπτει την περίπτωση των 0 γκολ, το δεύτερο εύρος τη περίπτωση του ενός γκολ, το τρίτο εύρος την περίπτωση των δύο γκολ κ.ο.κ. Στη συνέχεια, χρησιμοποιούμε τη συνάρτηση *rand*, η οποία διαλέγει τυχαία ένα αριθμό από το 0 ως το 1 και ανάλογα με το εύρος στο οποίο ανήκει αυτός ο αριθμός θεωρούμε ότι θα σημειωθούν και τα αντίστοιχα γκολ. Ίδια διαδικασία ακολουθούμε για όλες τις ομάδες. Ακολουθώντας, βάσει των γκολ που εκτιμούμε ότι θα σημειωθούν σε κάθε αγώνα, εκτιμούμε ποια ομάδα θα κερδίσει και ποια θα χάσει ή αν θα λήξει ο αγώνας ισόπαλος.

4.3 Αποτελέσματα

Λόγω του γεγονότος ότι τα γκολ που σκοράρουν οι ομάδες και κατ' επέκταση τα τελικά αποτελέσματα των αγώνων προκύπτουν από τη συνάρτηση *rand*, κάθε

φορά που τρέχουμε το πρόγραμμα παίρνουμε διαφορετικά αποτελέσματα. Επιλέξαμε λοιπόν να παρουσιάσουμε τα αποτελέσματα που προκύπτουν από 3 τυχαίες εκτελέσεις του προγράμματος.

Στους πίνακες που ακολουθούν παρουσιάζονται το μέσο ποσοστιαίο σφάλμα βαθμολογίας, το ποσοστό ευστοχίας στις προβλέψεις μας, το σφάλμα που έγινε στους αγώνες που είτε η εκτίμησή μας ήταν ισοπαλία και ο αγώνας τελείωσε με νικητή και ηττημένο, είτε το πραγματικό αποτέλεσμα ήταν ισοπαλία και η εκτίμησή μας ήταν διαφορετική και το σφάλμα στους αγώνες, όπου η ομάδα που εκτιμήσαμε ότι θα κερδίσει ήταν τελικά η ηττημένη.

Πίνακας 4-2 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
1 ^η εκτέλεση	20.79	31.58	24.85	23.31	17.39	19.22	21.03	22.6
2 ^η εκτέλεση	27.74	28.63	25.26	22.74	13.25	15.57	18.8	21.71
3 ^η εκτέλεση	23.08	28.44	18.99	30.19	25.49	17.45	18.83	23.21

Πίνακας 4-3 : Ποσοστό ευστοχίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
1 ^η εκτέλεση	43.5	36	34.5	35	42.5	36.5	42.1	38.59
2 ^η εκτέλεση	36.5	36	42.5	45	47	40	43.7	41.53
3 ^η εκτέλεση	36	37.5	38	36	39.5	39	42.9	38.41

Πίνακας 4-4 : Ποσοστό σφάλματος ισοπαλιών

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
1 ^η εκτέλεση	33.5	38.5	41	42.5	33	42	32.9	37.63
2 ^η εκτέλεση	39.5	43	34.5	31.5	37.5	36	32.9	36.41
3 ^η εκτέλεση	37	41	42	39.5	36	38.5	37.1	38.73

Πίνακας 4-5 : Ποσοστό σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
1 ^η εκτέλεση	23	25.5	24.5	22.5	24.5	21.5	24.9	23.77
2 ^η εκτέλεση	24	21	23	23.5	15.5	24	23.3	22.04
3 ^η εκτέλεση	27	21.5	20	24.5	24.5	22.5	19.9	22.84

Από τα αποτελέσματα που μόλις αναφέρθηκαν παρατηρούμε πως το μέσο ποσοστιαίο σφάλμα βαθμολογίας κυμαίνεται από 21.71 έως 23.21. Αντίστοιχα, ο μέσος όρος του ποσοστού ευστοχίας λαμβάνει τιμές 38.41 από έως 41.53. Ο μέσος όρος σφάλματος ισοπαλιών κυμαίνεται από 36.41 έως 38.73, ενώ ο αντίστοιχος του ποσοστού σφάλματος στους αγώνες, που η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη από 22.04 έως 23.77.

Τέλος, υπολογίσαμε την αναμενόμενη (expected) συχνότητα εμφάνισης 0,1,2,3,4,5 και περισσότερων από 5 γκολ στους εντός και εκτός έδρας αγώνες και τη συγκρίναμε με την πραγματική (observed).

Πίνακας 4-6 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2007-2008)

No of Goals	1 ^η εκτέλεση		2 ^η εκτέλεση		3 ^η εκτέλεση		Observed	
	Expected		Expected		Expected			
	Home	Away	Home	Away	Home	Away	Home	Away
0	69	98	54	82	61	75	44	89
1	64	48	77	79	60	75	79	66
2	32	37	37	26	37	31	47	26
3	19	14	20	10	22	10	16	13
4	10	3	7	3	15	5	9	6
5	4	0	3	0	3	2	4	0
>5	2	0	2	0	2	2	1	0

Πίνακας 4-7 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2008-2009)

No of Goals	1 ^η εκτέλεση		2 ^η εκτέλεση		3 ^η εκτέλεση		Observed	
	Expected		Expected		Expected			
	Home	Away	Home	Away	Home	Away	Home	Away
0	71	90	75	92	71	85	60	83
1	64	62	64	57	71	69	69	70
2	31	37	35	40	42	31	41	31
3	23	7	13	7	10	10	19	13
4	8	3	9	4	5	4	10	1
5	1	1	4	0	1	1	1	2
≥5	2	0	0	0	0	0	0	0

Πίνακας 4-8 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2009-2010)

No of Goals	1 ^η εκτέλεση		2 ^η εκτέλεση		3 ^η εκτέλεση		Observed	
	Expected		Expected		Expected			
	Home	Away	Home	Away	Home	Away	Home	Away
0	56	84	64	97	53	97	49	76
1	71	65	61	51	70	53	65	69
2	41	34	42	34	49	28	50	34
3	21	14	25	12	18	14	26	18
4	8	3	6	3	6	6	9	3
5	2	0	2	3	2	2	1	0
>5	1	0	0	0	2	0	0	0

Πίνακας 4-9 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2010-2011)

No of Goals	1 ^η εκτέλεση		2 ^η εκτέλεση		3 ^η εκτέλεση		Observed	
	Expected		Expected		Expected			
	Home	Away	Home	Away	Home	Away	Home	Away
0	67	88	66	100	78	87	56	78
1	71	64	78	59	73	76	75	73
2	34	31	32	29	32	26	37	35
3	17	13	16	8	12	7	23	10
4	6	4	7	2	4	4	6	4
5	2	0	1	2	1	0	1	0
>5	3	0	0	0	0	0	2	0

Πίνακας 4-10 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2011-2012)

No of Goals	1 ^η εκτέλεση		2 ^η εκτέλεση		3 ^η εκτέλεση		<i>Observed</i>	
	Expected		Expected		Expected			
	Home	Away	Home	Away	Home	Away	Home	Away
0	68	92	64	89	64	99	60	87
1	57	72	65	64	62	55	67	61
2	42	19	39	27	46	23	44	38
3	23	9	16	13	18	16	20	11
4	7	4	8	4	5	6	7	3
5	0	3	5	1	3	1	1	0
>5	3	1	3	2	2	0	1	0

Πίνακας 4-11 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2012-2013)

No of Goals	1 ^η εκτέλεση		2 ^η εκτέλεση		3 ^η εκτέλεση			
	Expected		Expected		Expected		<i>Observed</i>	
	Home	Away	Home	Away	Home	Away	Home	Away
0	70	100	66	86	77	85	59	86
1	65	70	69	68	58	73	62	65
2	42	22	40	29	34	28	55	35
3	17	6	13	12	20	10	16	8
4	4	1	6	4	6	4	6	6
5	2	1	3	1	4	0	2	0
>5	0	0	3	0	1	0	0	0

Πίνακας 4-12 : Παρατηρούμενη και αναμενόμενη συχνότητα εμφάνισης γκολ (2013-2014)

No of Goals	1 ^η εκτέλεση		2 ^η εκτέλεση		3 ^η εκτέλεση			
	Expected		Expected		Expected		<i>Observed</i>	
	Home	Away	Home	Away	Home	Away	Home	Away
0	73	113	74	115	67	114	60	104
1	67	79	77	88	79	80	70	90
2	75	44	52	34	60	45	74	41
3	29	20	28	16	34	16	39	16
4	11	4	18	5	11	4	14	6
5	5	1	8	2	4	0	2	4
>5	1	0	4	1	6	2	2	0

5. Απλά μοντέλα πρόβλεψης για *Baseline* αποτελέσματα

5.1 Το πλεονέκτημα της έδρας

Το πλεονέκτημα της έδρας στο ποδόσφαιρο είναι ένας σημαντικός παράγοντας στον καθορισμό του αποτελέσματος του παιχνιδιού. Παρά τη μελέτη πολλών χρόνων, οι ακριβείς αιτίες του πλεονεκτήματος της έδρας και ο τρόπος με τον οποίο επηρεάζουν το αποτέλεσμα δεν είναι πλήρως κατανοητές. Γεγονός πάντως είναι ότι η συμβολή της έδρας τείνει να είναι μεγαλύτερη στο ποδόσφαιρο από ό, τι σε άλλα ομαδικά αθλήματα.

Η πρώτη αναφορά και ποσοτική περιγραφή του πλεονεκτήματος της έδρας στο ποδόσφαιρο ήταν από τον Morris το 1981 [33] και ακολούθησε ο Dowie [34] και ο Pollard [35,36], ο οποίος περιέγραψε τα κύρια πιθανολογούμενα αίτια, τα οποία παρουσιάζονται παρακάτω.

5.1.1 Η επίδραση των φιλάθλων

Αυτός είναι ο πιο προφανής παράγοντας που εμπλέκεται με το πλεονέκτημα της έδρας [37,38]. Μάλιστα, ακόμη και οι ομάδες που είναι λιγότερο δημοφιλείς και που διαθέτουν μικρό αριθμό φιλάθλων φαίνεται να αποδίδουν καλύτερα στους εντός έδρας αγώνες [36,39], διότι εκτός από το μέγεθος του πλήθους, η πυκνότητά του, η έντασή του και η υποστήριξη που προσφέρει στους παίκτες είναι παράγοντες ιδιαίτερα σημαντικοί [40,41].

5.1.2 Η επίδραση του ταξιδιού

Όπως γίνεται εύκολα κατανοητό η μετακίνηση των παιχτών φέρνει σε μειονεκτική θέση την φιλοξενούμενη ομάδα, καθώς η κούραση των παιχτών της εντείνεται. Μάλιστα, φαίνεται ότι το πλεονέκτημα έδρας είναι μεγαλύτερο σε αγώνες μεταξύ ομάδων διαφορετικών χωρών και μειώνεται στους τοπικούς αγώνες, όπου τα ταξίδια είναι μικρότερα [35,42,43,44].

5.1.3 Η εξοικείωση

Μελέτες σχετικά με την επίδραση της έδρας υποδεικνύουν ότι η εξοικείωση είναι ένας παράγοντας που επιδρά θετικά στην απόδοση των παιχτών. Ειδικότερα, όταν μια ομάδα αγωνίζεται στην έδρα της, όπου οι παίκτες είναι εξοικειωμένοι με το γήπεδο και τις τοπικές κλιματολογικές συνθήκες [42,44], αποδίδει καλύτερα.

5.1.4 Η προκατάληψη του διαιτητή

Υπάρχουν πλέον συντριπτικά στοιχεία ότι οι αποφάσεις του διαιτητή ευνοούν συνήθως τη γηπεδούχο ομάδα. Η θέση αυτή υποστηρίχτηκε ύστερα από την ανάλυση της συχνότητας των καρτών και των υπολοίπων αποφάσεων του διαιτητή από αρκετές πηγές [40, 41, 45-52]. Ο λόγος για την προφανή μεροληψία του διαιτητή πιστεύεται ότι είναι η επίδραση των φιλάθλων της γηπεδούχου ομάδας.

5.1.5 Η ειδική τακτική

Τις περισσότερες φορές η γηπεδούχος και η φιλοξενούμενη ομάδα προσεγγίζουν διαφορετικά τον αγώνα από άποψη τακτικής. Ως επί το πλείστον η φιλοξενούμενη ομάδα δίνει μεγαλύτερη προσοχή στην αμυντική της λειτουργία, με αποτέλεσμα να παραδίδει ένα εδαφικό και ψυχολογικό πλεονέκτημα στη γηπεδούχο ομάδα [34,36].

5.1.6 Οι ψυχολογικοί παράγοντες

Δεδομένου ότι οι παίκτες και οι προπονητές γνωρίζουν καλά την ύπαρξη του πλεονεκτήματος της έδρας, είναι λογικό η ψυχική τους κατάσταση πριν και κατά τη διάρκεια μιας αναμέτρησης να επηρεάζεται. Η σπουδαιότητα του συγκεκριμένου παράγοντα αναφέρεται στις πηγές [35,36]. Μια εκτενής αναφορά του πλεονεκτήματος της έδρας στο ποδόσφαιρο από ψυχολογικής άποψης έχει πραγματοποιηθεί από τους Neave και Wolfson [53].

5.1.7 Άλλες αιτίες

Η ικανότητα των ομάδων φαίνεται να είναι ένας παράγοντας που επηρεάζει το μέγεθος του πλεονεκτήματος της έδρας και πρέπει να λαμβάνεται υπ' όψιν [42,54]. Ιδίως, στην περίπτωση που αναμετρώνται δύο ομάδες παρόμοιας δυναμικότητας η έδρα επιδρά σημαντικά υπέρ της γηπεδούχου. Επιπλέον, η σημασία του παιχνιδιού είναι ένας ακόμη παράγοντας που επηρεάζει το μέγεθος του πλεονεκτήματος της έδρας [55]. Προφανώς, για μια ομάδα η διεξαγωγή των κρίσιμων παιχνιδιών στην έδρα της λειτουργεί ως ευνοϊκός παράγοντας.

5.2 Προβλέψεις βάσει της έδρας (εντός/εκτός)

Μια μέθοδος που χρησιμοποιείται αρκετά στη βιβλιογραφία και παρουσιάζεται ως αφελής (naïve model) είναι να θεωρούμε πως η γηπεδούχος ομάδα θα είναι πάντα η νικήτρια, ανεξάρτητα από τη «δύναμη» των δύο αντίπαλων ομάδων. Στο ποδόσφαιρο η μέθοδος αυτή προσφέρει αξιολογικά αποτελέσματα, γεγονός που φανερώνει ότι η έδρα ως παράγοντας κατέχει σημαντικό ρόλο. Για τα έτη που εξετάζουμε ο μέσος όρος εντός έδρας νικών που σημειώνεται επί του συνόλου των αγώνων που διεξάγονται (κανονική περίοδος και play-offs) στην Α' κατηγορία στο ελληνικό πρωτάθλημα είναι 49.31%, στο ισπανικό 49.36%, στο αγγλικό 46.5%, στο ιταλικό 47.49% και στο γερμανικό 45.21%. Από τα ποσοστά αυτά, διακρίνουμε την ευστοχία της μεθόδου κατά 45% με 50% σε πλήθος διαφορετικών πρωταθλημάτων. Τα αποτελέσματα αυτά μπορούν να χρησιμοποιηθούν ως baseline αποτελέσματα ώστε να συγκριθούν με εκείνα των πιο «έξυπνων» μεθόδων που εξετάσαμε.

Βασικό μειονέκτημα της συγκεκριμένης μεθόδου είναι ότι δεν κάνουμε καμία εκτίμηση για τις ισοπαλίες. Θωρούμε δηλαδή πως κάθε αγώνας θα τελειώνει με νικητή (γηπεδούχος ομάδα) και ηττημένο (φιλοξενούμενη ομάδα).

5.3 Προβλέψεις βάσει των αποτελεσμάτων της προηγούμενης σεζόν

Στη συγκεκριμένη μέθοδο θεωρούμε πως θα επαναληφθούν τα αποτελέσματα που σημειώθηκαν στους εντός και εκτός έδρας αγώνες της προηγούμενης σεζόν -πριν την προβλεπόμενη- για κάθε ομάδα. Ένα μειονέκτημα της συγκεκριμένης μεθόδου είναι ότι δεν κάνουμε εκτιμήσεις για τις ομάδες που προβιβάστηκαν στη Superleague την προβλεπόμενη χρονιά και επομένως δε διαθέτουν την απαιτούμενη προϊστορία με τους αντιπάλους τους.

Στον πίνακα που ακολουθεί έχουμε καταγράψει τις ομάδες που συμμετείχαν κάθε αγωνιστική σεζόν στην Superleague με τη σειρά κατάταξής τους βάσει της τελικής τους βαθμολογίας. Με κόκκινο χρώμα έχουν σημειωθεί οι ομάδες για τις οποίες δεν κάνουμε εκτιμήσεις.

Πίνακας 5-1 : Ομάδες που συμμετέχουν στην Superleague τις σεζόν 2007-2008 έως 2010-2011

2007-2008	2008-2009	2009-2010	2010-2011
Ολυμπιακός	Ολυμπιακός	Παναθηναϊκός	Ολυμπιακός
ΑΕΚ	ΠΑΟΚ	Ολυμπιακός	Παναθηναϊκός
Παναθηναϊκός	Παναθηναϊκός	ΠΑΟΚ	ΑΕΚ
Άρης	ΑΕΚ	ΑΕΚ	ΠΑΟΚ
Πανιώνιος	Λάρισα	Άρης	Ολυμπιακός Βόλου
Λάρισα	Άρης	Καβάλα	Άρης
Αστέρας Τρίπολης	Skoda Ξάνθη	Ατρόμητος	Καβάλα
Skoda Ξάνθη	Πανιώνιος	Λάρισα	Εργοτέλης
ΠΑΟΚ	Εργοτέλης	Πανιώνιος	Skoda Ξάνθη
Ηρακλής	Ηρακλής	Ηρακλής	Πανιώνιος
Λεβαδειακός	Πανθρακικός	Εργοτέλης	Ατρόμητος
ΟΦΗ	Αστέρας Τρίπολης	Αστέρας Τρίπολης	Κέρκυρα
Εργοτέλης	Λεβαδειακός	Skoda Ξάνθη	Αστέρας Τρίπολης
Ατρόμητος	ΟΦΗ	Λεβαδειακός	Ηρακλής
Βέροια	Πανσερραϊκός	Πας Γιάννενα	Λάρισα
Απόλλων Σμύρνης	Θρασύβουλος	Πανθρακικός	Πανσερραϊκός

Πίνακας 5-2 : Ομάδες που συμμετέχουν στην Superleague τις σεζόν 2011-2012 έως 2013-2014

2011-2012	2012-2013	2013-2014
Ολυμπιακός	Ολυμπιακός	Ολυμπιακός
Παναθηναϊκός	ΠΑΟΚ	ΠΑΟΚ
ΠΑΟΚ	Αστέρας Τρίπολης	Ατρόμητος
Ατρόμητος	Ατρόμητος	Παναθηναϊκός
ΑΕΚ	Πας Γιάννενα	Αστέρας Τρίπολης
Αστέρας Τρίπολης	Παναθηναϊκός	ΟΦΗ
Λεβαδειακός	Skoda Ξάνθη	Εργοτέλης
Πας Γιάννενα	Πανιώνιος	Παναιτωλικός
Άρης	Πλατανιάς Χανίων	Λεβαδειακός
ΟΦΗ	Πανθρακικός	Πανθρακικός
Skoda Ξάνθη	Λεβαδειακός	Πας Γιάννενα
Πανιώνιος	Βέροια	ΑΕΛ Καλλονής
Κέρκυρα	Άρης	Πανιώνιος
Εργοτέλης	ΟΦΗ	Πλατανιάς Χανίων
Παναιτωλικός	ΑΕΚ	Βέροια
Δόξα Δράμας	Κέρκυρα	Skoda Ξάνθη
-	-	Απόλλων Σμύρνης
-	-	Άρης

5.4 Αποτελέσματα

Πίνακας 5-3 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
Home Team	29.06	39.62	32.82	26.48	27.74	25.52	28.33	29.94
Last year score	30.99	26	15.86	20.34	21.93	26.62	20.98	23.25

Πίνακας 5-4 : Ποσοστό ευστοχίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
Home Team	53	42	52	45	49	46.5	56.3	49.11
Last year score	35.2	40.5	50.1	36.3	38.7	41.6	49.3	41.67

Πίνακας 5-5 : Ποσοστό σφάλματος ισοπαλιών

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
Home Team	22.5	30	22	27	23.5	26.5	20.3	24.54
Last year score	41.7	40.9	33.7	25.7	25	38.1	33.1	34.03

Πίνακας 5-6 : Ποσοστό σφάλματος στους αγώνες, όπου η εκτιμώμενη νικήτρια ομάδα ήταν τελικά η ηττημένη

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
Home Team	24.5	28	26	28	27.5	27	23.4	26.34
Last year score	23.1	18.5	16.1	38	36.3	20.3	17.7	24.29

Συγκρίνοντας τις δύο μεθόδους παρατηρούμε πως σχεδόν για κάθε χρονιά η μέθοδος Last Year Score προσφέρει καλύτερα αποτελέσματα όσον αφορά τη τελική βαθμολογία των ομάδων. Ιδίως για τη σεζόν 2009-2010 το μέσο ποσοστιαίο σφάλμα της μεθόδου Home Team είναι διπλάσιο από το αντίστοιχο της μεθόδου Last Year Score. Επιπλέον, πολύ μεγάλη διαφορά σημειώνεται και για τη σεζόν 2008-2009, όπου το μέσο ποσοστιαίο σφάλμα βαθμολογίας της μεθόδου Last Year Score είναι μικρότερο κατά 13.62% από το αντίστοιχο της μεθόδου Home Team, κατά 6.14% για τη σεζόν 2010-2011, κατά 5.81% για τη σεζόν 2011-2012 και κατά 7.35% για τη σεζόν 2013-2014. Επίσης, η μέθοδος Last Year Score εμφανίζει μικρότερο σφάλμα «πλήρους ανατροπής» της πρόβλεψής μας, όπως φαίνεται από τον Πίνακα 5-6.

Η μέθοδος όμως Last Year Score είναι λιγότερη εύστοχη όσον αφορά τα τελικά αποτελέσματα των αγώνων. Για κάθε έτος το ποσοστό ευστοχίας που παρέχει είναι μικρότερο από το αντίστοιχο της μεθόδου Home Team. Ειδικότερα, για τη σεζόν 2007-2008 η ευστοχία της είναι κατά 17.8% μικρότερη από την αντίστοιχη της μεθόδου Home Team, κατά 1.5% για το έτος 2008-2009, κατά 1.9% για το έτος 2009-2010, κατά 8.7% για τη χρονιά 2010-2011, κατά 10.3% για τη χρονιά 2011-2012, κατά 4.9% για τη χρονιά 2012-2013 και τέλος κατά 7% για τη σεζόν 2013-2014. Κύριος λόγος για αυτό είναι το μεγάλο ποσοστό σφάλματος ισοπαλιών που παρέχει, που για κάποιες χρονιές ξεπερνά το 40%. Αναλυτικότερα, για τη σεζόν 2007-2008 η αστοχία στις ισοπαλίες που παρέχει είναι κατά 19.2% μεγαλύτερη από την αντίστοιχη της μεθόδου Home Team, κατά 10.9% για το έτος 2008-2009, κατά 11.7% για το έτος 2009-2010, κατά 1.5% για τη χρονιά 2011-2012, κατά 11.6% για τη χρονιά 2012-2013 και τέλος κατά 12.8% για τη σεζόν 2013-2014.

6. Μέθοδοι εξομάλυνσης

Οι μέθοδοι εξομάλυνσης (smoothing methods) είναι τεχνικές με τις οποίες προσδιορίζονται οι μελλοντικές τιμές μιας μεταβλητής με βάση τον τρόπο εφαρμογής τους [56]. Αποσκοπούν δηλαδή στην πρόβλεψη του μέλλοντος με τη χρήση δεδομένων του παρελθόντος. Για παράδειγμα, μια εταιρεία μπορεί να χρησιμοποιήσει στοιχεία των ετήσιων πωλήσεων της για τα τελευταία πέντε χρόνια προκειμένου να προβλέψει τις μελλοντικές πωλήσεις της τον επόμενο χρόνο. Οι συγκεκριμένες μέθοδοι υποθέτουν ότι η υπό εξέταση μεταβλητή θα ακολουθήσει το ίδιο υπόδειγμα κίνησης όπως και στο παρελθόν. Οι τεχνικές αυτές ονομάζονται μέθοδοι εξομάλυνσης, διότι η δημιουργία των προβλέψεων προέρχεται από την εξομάλυνση της διαχρονικής εξέλιξης των τιμών της μεταβλητής, ώστε να αναγνωριστεί καλύτερα ο τρόπος συμπεριφοράς της. Κύριο χαρακτηριστικό της εφαρμογής αυτών των μεθόδων είναι ότι μπορούμε, χωρίς μεγάλο υπολογιστικό βαθμό δυσκολίας, να διαμορφώσουμε εύκολα και σχετικά γρήγορα προβλέψεις για μια μεταβλητή, που προέρχονται από τη διαχρονική μελέτη του τρόπου δημιουργίας των τιμών της συγκεκριμένης μεταβλητής [57]. Οι μέθοδοι αυτές εφαρμόζονται για μεταβλητές, τα δεδομένα των οποίων προέρχονται αποκλειστικά από χρονοσειρές, δηλαδή από παρατηρήσεις που έχουν καταγραφεί κατά τη διάρκεια ίσων διαδοχικών χρονικών περιόδων, ενώ ο αριθμός των διαθέσιμων παρατηρήσεων της μεταβλητής δεν αποτελεί κατά κανόνα ανασταλτικό παράγοντα για την εφαρμογή τους. Ορισμένες από αυτές τις μεθόδους μπορούν να χρησιμοποιηθούν ακόμα και για δείγματα παρατηρήσεων μικρού μεγέθους.

Αρχικά θα αναπτύξουμε τη μέθοδο του απλού κινητού μέσου m -περιόδων, σύμφωνα με την οποία η πρόβλεψη είναι ο αριθμητικός μέσος όρος των m πιο πρόσφατων παρατηρήσεων της χρονοσειράς. Έπειτα, θα αναφερθούμε στη μέθοδο του σταθμισμένου κινητού μέσου, όπου δίνεται διαφορετική βαρύτητα στις m παρατηρήσεις. Στη συνέχεια θα παρουσιάσουμε τη μέθοδο της απλής εκθετικής εξομάλυνσης σύμφωνα με την οποία η πρόβλεψη είναι ένας σταθμικός μέσος όρος των παρατηρήσεων της χρονοσειράς, οι οποίες όσο πιο πρόσφατες είναι έχουν και μεγαλύτερη βαρύτητα. Στόχος όλων αυτών των μεθόδων εξομάλυνσης είναι η αναγνώριση του τρόπου δημιουργίας των παρατηρήσεων της χρονοσειράς, ώστε να αποκτηθούν οι καλύτερες δυνατές προβλέψεις.

6.1 Απλός Κινητός Μέσος (MA)

Η μέθοδος του απλού κινητού μέσου m -περιόδων (simple moving average) είναι μία πολύ απλή μέθοδος προβλέψεων που χρησιμοποιεί ως πρόβλεψη την τιμή του αριθμητικού μέσου όρου των m πλέον πρόσφατων παρατηρήσεων της χρονοσειράς [58]. Αυτό συμβαίνει διότι οι πλέον πρόσφατες παρατηρήσεις της χρονοσειράς θεωρούνται περισσότερο αντιπροσωπευτικές για τη δημιουργία προβλέψεων από ότι οι πιο απομακρυσμένες στο παρελθόν. Ο μέσος όρος αυτός ονομάζεται κινητός, επειδή η τιμή του δεν είναι σταθερή, αλλά αναπροσαρμόζεται κάθε φορά που μια νέα παρατήρηση της χρονοσειράς γίνεται διαθέσιμη και υπολογίζεται ένας νέος μέσος απορρίπτοντας την παλαιότερη παρατήρηση του δείγματος από τον μέσο και συμπεριλαμβάνοντας τη νεότερη. Καθώς υπολογίζεται κάθε νέος μέσος, γίνεται πρόβλεψη για την επόμενη περίοδο. Ο αριθμός των

παρατηρήσεων του δείγματος που συμπεριλαμβάνονται στον υπολογισμό του μέσου προσδιορίζεται στην αρχή της διαδικασίας πρόβλεψης. Συνεπώς, κάθε πρόβλεψη χρησιμοποιεί τον ίδιο αριθμό παρατηρήσεων δείγματος από τη χρονοσειρά και περιλαμβάνει μόνο την πλέον πρόσφατη παρατήρηση.

Η μέθοδος του κινητού μέσου όρου, σε αλγεβρική μορφή, διατυπώνεται ως εξής :

$$Z_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-n+1}}{n} = \frac{1}{n} (\sum_{i=t-n+1}^t Y_i)$$

όπου t είναι η πιο πρόσφατη παρατήρηση και $t+1$ είναι η επόμενη περίοδος. Ο τύπος αυτός απαιτεί ο προβλέπων να διαθέτει τις τιμές των παρατηρήσεων του παρελθόντος. Με την προσθήκη μιας νέας παρατήρησης και την εξάλειψη της παλαιότερης, μπορούμε να επαναδιατυπώσουμε τον τύπο ως εξής:

$$Z_{t+1} = \frac{1}{n} (\sum_{i=t-n}^{t-1} Y_i) + \frac{1}{n} (Y_t - Y_{t-n}) = Z_t + \frac{Y_t}{n} - \frac{Y_{t-n}}{n}$$

Ο τύπος του κινητού μέσου όρου εκφράζει απλώς μια προσαρμογή στην πρόβλεψη (Z_t) που έγινε την προηγούμενη περίοδο. Ο απλός κινητός μέσος είναι πολύ χρήσιμος για να απομακρύνει την τυχαία μεταβλητότητα στην πρόβλεψη, όταν η ζήτηση δεν έχει τάση και εποχικότητα. Είναι πολύ σημαντικό να επιλεγεί το κατάλληλο διάστημα για τον κινητό μέσο. Όσο μεγαλύτερο είναι το επιλεγμένο διάστημα τόσο περισσότερο εξομαλύνεται το τυχαίο στοιχείο στην πρόβλεψη. Όταν όμως υπάρχει τάση στην ζήτηση, δηλαδή αυξάνεται ή μειώνεται σε συνάρτηση με το χρόνο, ο κινητός μέσος ενός μεγάλου διαστήματος θα εξομαλύνει και την τάση. Επομένως, ένα μικρότερο χρονικό διάστημα αν και θα παρουσιάζει μεγαλύτερη διακύμανση ακολουθεί με μεγαλύτερη ακρίβεια την τάση της ζήτησης.

Οι κυριότερες αντιρρήσεις που έχουν εκφραστεί αναφορικά με τη χρησιμοποίηση της μεθόδου του απλού κινητού μέσου είναι οι εξής :

1. Η ίση στάθμιση των δεδομένων κάθε μιας εκ των προηγούμενων περιόδων μπορεί να μην εκφράζει ικανοποιητικά την παρούσα κατάσταση.
2. Η αύξηση του αριθμού των καλυπτόμενων χρονικών περιόδων n , έχει ως αποτέλεσμα την μεγαλύτερη εξομάλυνση των διακυμάνσεων, αλλά κάνει τη μέθοδο λιγότερο ευαίσθητη στις πραγματικές αλλαγές των δεδομένων.
3. Ο απλός κινητός μέσος παρουσιάζει μία χρονική υστέρηση σε αλλαγές των πραγματικών τιμών των δεδομένων λόγω μακροχρόνιας τάσεως ή άλλων αιτιών.
4. Η μέθοδος του απλού κινητού μέσου απαιτεί εκτεταμένα αρχεία για δεδομένα παρελθουσών περιόδων.
5. Με τη μέθοδο του απλού κινητού μέσου είναι δυνατή η πρόβλεψη μόνο για την επόμενη χρονική περίοδο.

Σκεφτήκαμε λοιπόν να εφαρμόσουμε τη μέθοδο του απλού κινητού μέσου στο πρόβλημά μας. Στόχος μας είναι ο υπολογισμός της τελικής βαθμολογίας των ομάδων. Για το σκοπό αυτό για κάθε σεζόν ξεκινώντας από το 2004-2005 υπολογίσαμε τη διαφορά μεταξύ του ποσοστού νικών, ισοπαλιών και ηττών κάθε ομάδας της τελευταίας αγωνιστικής και του αντίστοιχου ποσοστού μετά το τέλος της μισής αγωνιστικής σεζόν. Έτσι, χρησιμοποιώντας το ποσοστό που είχε η κάθε ομάδα

στο ήμισυ της τρέχουσας αγωνιστικής περιόδου και το μέσο όρο της αντίστοιχης υπολογιζόμενης διαφοράς υπολογίσαμε τα τελικά ποσοστά νικών, ισοπαλιών και ηττών των ομάδων.

Ο πίνακας που ακολουθεί παρουσιάζει τον τρόπο υπολογισμού του απλού κινητού μέσου για την περίπτωση των νικών, λαμβάνοντας ως διάστημα τις διαφορές τριών ετών (n=3). Ανάλογη διαδικασία ακολουθήθηκε για τον υπολογισμό των τελικών ποσοστών ισοπαλιών και ηττών.

Πίνακας 6-1 : Μέθοδος Απλού Κινητού Μέσου

Αγωνιστική Σεζόν	Παρατηρήσεις	Εκτίμηση Διαφοράς	Εκτίμηση Τελικού Ποσοστού
2004-2005	$Y_1 = [WP(30)-WP(15)]_{2004-5}$	-	-
2005-2006	$Y_2 = [WP(30)-WP(15)]_{2005-6}$	-	-
2006-2007	$Y_3 = [WP(30)-WP(15)]_{2006-7}$	-	-
2007-2008	$Y_4 = [WP(30)-WP(15)]_{2007-8}$	$Z_1 = \frac{(Y_1+Y_2+Y_3)}{3}$	$W_1=WP(15)_{2007-8}+Z_1$
2008-2009	$Y_5 = [WP(30)-WP(15)]_{2008-9}$	$Z_2 = \frac{(Y_2+Y_3+Y_4)}{3}$	$W_2=WP(15)_{2008-9}+Z_2$
2009-2010	$Y_6 = [WP(30)-WP(15)]_{2009-10}$	$Z_3 = \frac{(Y_3+Y_4+Y_5)}{3}$	$W_3=WP(15)_{2009-10}+Z_3$
2010-2011	$Y_7 = [WP(30)-WP(15)]_{2010-11}$	$Z_4 = \frac{(Y_4+Y_5+Y_6)}{3}$	$W_4=WP(15)_{2010-11}+Z_4$
2011-2012	$Y_8 = [WP(30)-WP(15)]_{2011-12}$	$Z_5 = \frac{(Y_5+Y_6+Y_7)}{3}$	$W_5=WP(15)_{2011-12}+Z_5$
2012-2013	$Y_9 = [WP(30)-WP(15)]_{2012-13}$	$Z_6 = \frac{(Y_6+Y_7+Y_8)}{3}$	$W_6=WP(15)_{2012-13}+Z_6$
2013-2014	-	$Z_7 = \frac{(Y_7+Y_8+Y_9)}{3}$	$W_7=WP(15)_{2013-14}+Z_7$

Έτσι, έχουμε εκτιμήσει τα τελικά ποσοστά νικών, ισοπαλιών και ηττών των ομάδων. Στη συνέχεια, μετατρέπουμε τα ποσοστά αυτά σε πλήθος παιχνιδιών :

$$\text{Number_}W_i = W_i \times k, \text{ για } i=1,..7$$

όπου $k=34$ για τη σεζόν 2013-2014 και $k=30$ για όλες τις υπόλοιπες χρονιές.

Αντίστοιχα, υπολογίζουμε το πλήθος των ισοπαλιών (Number_ D_i) και των ηττών (Number_ L_i) της κάθε ομάδας. Τέλος, η τελική εκτιμώμενη βαθμολογία της κάθε μιας προκύπτει ως :

$$\text{Points} = 3 \times \text{Number_}W_i + 1 \times \text{Number_}D_i + 0 \times \text{Number_}L_i$$

6.2 Σταθμισμένος Κινητός Μέσος (WMA)

Η μέθοδος του σταθμισμένου κινητού μέσου είναι μια απλή μέθοδος προβλέψεων. Μια από τις αντιρρήσεις που έχουν εκφραστεί για τον απλό κινητό μέσο είναι η ίση στάθμιση των δεδομένων κάθε μιας εκ των προηγούμενων περιόδων, η οποία μπορεί να μην εκφράζει ικανοποιητικά την παρούσα κατάσταση. Το μειονέκτημα αυτό μπορούμε να το εξουδετερώσουμε χρησιμοποιώντας το σταθμισμένο κινητό μέσο. Η λογική και η μεθοδολογία της μεθόδου αυτής είναι ακριβώς η ίδια με αυτή του απλού κινητού μέσου με τη μόνη διαφορά ότι εδώ σταθμίζουμε τα δεδομένα των περιόδων που εξετάζουμε, χρησιμοποιώντας διαφορετικούς συντελεστές βαρύτητας-στάθμισης ανάλογα με το χρονικό ορίζοντα απόκτησης των δεδομένων. Συνήθως, τα δεδομένα των πιο πρόσφατων περιόδων τα σταθμίζουμε χρησιμοποιώντας συντελεστές βαρύτητας που είναι μεγαλύτεροι από αυτούς που χρησιμοποιούνται για τη στάθμιση δεδομένων πιο μακρινών περιόδων[58]. Η επιλογή των συντελεστών βαρύτητας εξαρτάται από την κρίση και την εμπειρία του εκάστοτε αναλυτή. Γενικά θα μπορούσαμε να πούμε ότι αν το πρόσφατο, σε σχέση με το απώτερο, παρελθόν θεωρείται καλύτερος δείκτης πρόβλεψης του μέλλοντος, τότε μεγαλύτεροι συντελεστές βαρύτητας θα πρέπει να δίνονται στις πιο πρόσφατες παρατηρήσεις. Ωστόσο, αν η χρονολογική σειρά που εξετάζουμε διακρίνεται από μεγάλη μεταβλητότητα, τότε ίσως είναι ορθότερο να χρησιμοποιούμε ίσους συντελεστές βαρύτητας για κάθε παρατήρηση.

Ο κινητός μέσος με βάρη υπολογίζεται σύμφωνα με τον τύπο:

$$Z_t = \frac{w_1 \times Y_{t-1} + w_2 \times Y_{t-2} + \dots + w_n \times Y_{t-n}}{w_1 + w_2 + \dots + w_n} = \frac{\sum_{i=1}^n (w_i \times Y_{t-i})}{\sum_{i=1}^n w_i}$$

Αξίζει να σημειωθεί ότι τα μειονεκτήματα που αναφέραμε στην προηγούμενη ενότητα και που αφορούν τον απλό κινητό μέσο ισχύουν και για τον σταθμισμένο κινητό μέσο, με εξαίρεση αυτό της ίσης στάθμισης.

Εφαρμόσαμε λοιπόν τη μέθοδο του σταθμισμένου κινητού μέσου στο πρόβλημά μας, ακολουθώντας παρόμοια λογική με αυτή του απλού. Στόχος μας είναι ο υπολογισμός της τελικής βαθμολογίας των ομάδων. Για το σκοπό αυτό για κάθε σεζόν ξεκινώντας από το 2004-2005 υπολογίσαμε τη διαφορά μεταξύ του ποσοστού νικών, ισοπαλιών και ηττών κάθε ομάδας της τελευταίας αγωνιστικής και του αντίστοιχου ποσοστού μετά το τέλος της μισής αγωνιστικής σεζόν. Έτσι, χρησιμοποιώντας το ποσοστό που είχε η κάθε ομάδα στο ήμισυ της τρέχουσας αγωνιστικής περιόδου και το μέσο όρο της αντίστοιχης υπολογιζόμενης διαφοράς υπολογίσαμε τα τελικά ποσοστά νικών, ισοπαλιών και ηττών των ομάδων. Η μεγαλύτερη βαρύτητα επιλέξαμε να δοθεί στην πιο πρόσφατη αναμέτρηση.

Ο πίνακας που ακολουθεί παρουσιάζει τον τρόπο υπολογισμού του σταθμισμένου κινητού μέσου για την περίπτωση των νικών, λαμβάνοντας ως διάστημα τις διαφορές τριών ετών ($n=3$) και με βαρύτητες : $w_1=1$, $w_2=2$ και $w_3=3$. Ανάλογη διαδικασία ακολουθήθηκε για τις ισοπαλίες και τις ήττες.

Πίνακας 6-2 : Μέθοδος Σταθμισμένου Κινητού Μέσου

Αγωνιστική Σεζόν	Παρατηρήσεις	Εκτίμηση Διαφοράς	Εκτίμηση Τελικού Ποσοστού
2004-2005	$Y_1 = [WP(30)-WP(15)]_{2004-5}$	-	-
2005-2006	$Y_2 = [WP(30)-WP(15)]_{2005-6}$	-	-
2006-2007	$Y_3 = [WP(30)-WP(15)]_{2006-7}$	-	-
2007-2008	$Y_4 = [WP(30)-WP(15)]_{2007-8}$	$Z_1 = \frac{(w1 \times Y1 + w2 \times Y2 + w3 \times Y3)}{\sum_{i=1}^3 w_i}$	$W_1 = WP(15)_{2007-8} + Z_1$
2008-2009	$Y_5 = [WP(30)-WP(15)]_{2008-9}$	$Z_2 = \frac{(w1 \times Y2 + w2 \times Y3 + w3 \times Y4)}{\sum_{i=1}^3 w_i}$	$W_2 = WP(15)_{2008-9} + Z_2$
2009-2010	$Y_6 = [WP(30)-WP(15)]_{2009-10}$	$Z_3 = \frac{(w1 \times Y3 + w2 \times Y4 + w3 \times Y5)}{\sum_{i=1}^3 w_i}$	$W_3 = WP(15)_{2009-10} + Z_3$
2010-2011	$Y_7 = [WP(30)-WP(15)]_{2010-11}$	$Z_4 = \frac{(w1 \times Y4 + w2 \times Y5 + w3 \times Y6)}{\sum_{i=1}^3 w_i}$	$W_4 = WP(15)_{2010-11} + Z_4$
2011-2012	$Y_8 = [WP(30)-WP(15)]_{2011-12}$	$Z_5 = \frac{(w1 \times Y5 + w2 \times Y6 + w3 \times Y7)}{\sum_{i=1}^3 w_i}$	$W_5 = WP(15)_{2011-12} + Z_5$
2012-2013	$Y_9 = [WP(30)-WP(15)]_{2012-13}$	$Z_6 = \frac{(w1 \times Y6 + w2 \times Y7 + w3 \times Y8)}{\sum_{i=1}^3 w_i}$	$W_6 = WP(15)_{2012-13} + Z_6$
2013-2014	-	$Z_7 = \frac{(w1 \times Y7 + w2 \times Y8 + w3 \times Y9)}{\sum_{i=1}^3 w_i}$	$W_7 = WP(15)_{2013-14} + Z_7$

Έτσι, έχουμε εκτιμήσει τα τελικά ποσοστά νικών, ισοπαλιών και ηττών των ομάδων. Στη συνέχεια, μετατρέπουμε τα ποσοστά αυτά σε πλήθος παιχνιδιών :

$$\text{Number_}W_i = W_i \times k, \text{ για } i=1,..7$$

όπου $k=34$ για τη σεζόν 2013-2014 και $k=30$ για όλες τις υπόλοιπες χρονιές.

Αντίστοιχα, υπολογίζουμε το πλήθος των ισοπαλιών (Number_D_i) και των ηττών (Number_L_i) της κάθε ομάδας. Τέλος, η τελική εκτιμώμενη βαθμολογία της κάθε μιας προκύπτει ως :

$$\text{Points} = 3 \times \text{Number_}W_i + 1 \times \text{Number_}D_i + 0 \times \text{Number_}L_i$$

6.3 Απλή Εκθετική Εξομάλυνση (EWMA)

Στα πλαίσια της παρουσίασης του απλού κινητού μέσου, που αποτελεί μια πολύ απλή μέθοδο προβλέψεων, διαπιστώσαμε ότι ένα από τα μειονεκτήματα της μεθόδου είναι η ίση βαρύτητα που δίνει, κατά τον υπολογισμό των προβλέψεων, σε κάθε παρατήρηση ανεξάρτητα από τόσο κοντά ή μακριά βρίσκεται σε σχέση με την προβλεπόμενη περίοδο. Ένας τρόπος για την αντιμετώπιση αυτού του προβλήματος

είναι η χρήση της μεθόδου του σταθμισμένου κινητού μέσου. Μία άλλη μέθοδος που μπορεί να αντιμετωπίσει το ανωτέρω πρόβλημα και που τυγχάνει ευρείας αποδοχής σε διάφορες περιπτώσεις που απαιτούνται προβλέψεις οικονομικών μεγεθών, είναι η μέθοδος της απλής εκθετικής εξομάλυνσης [58]. Σύμφωνα με τη μέθοδο αυτή οι προβλέψεις δημιουργούνται βάσει κάποιου σταθμικού μέσου όρου, έτσι ώστε να δίνεται διαφορετική βαρύτητα σε κάθε παρατήρηση. Αναλυτικότερα, με τη μέθοδο αυτή δίνεται μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις, σε σχέση πάντοτε με την προβλεπόμενη περίοδο, από αυτή που δίνεται στις πιο απομακρυσμένες.

Για να μπορέσουμε να κατανοήσουμε τον τρόπο λειτουργίας της μεθόδου αυτής θα θεωρήσουμε ότι οι προβλέψεις της χρονοσειράς που εξετάζουμε δημιουργούνται ως εξής :

$$Z_{t+1} = a \times Y_t + a \times (1-a) \times Y_{t-1} + a \times (1-a)^2 \times Y_{t-2} + \dots + a \times (1-a)^n \times Z_{t-n}$$

Η παράμετρος a της σχέσης αυτής ονομάζεται σταθερά εξομάλυνσης (smoothing constant) και λαμβάνει τιμές από μηδέν έως και ένα, δηλαδή $0 \leq a \leq 1$. Στην πράξη, μελέτες που έγιναν, δείχνουν ότι οι τιμές από 0.05 έως 0.30 ταιριάζουν πολύ καλά στα πρότυπα της εκθετικής εξομάλυνσης. Τιμές της a που είναι μεγαλύτερες από 0.30 συνήθως δείχνουν ότι κάποιο εναλλακτικό μοντέλο πρόβλεψης θα είναι περισσότερο κατάλληλο.

Από τη σχέση που μόλις αναφέρθηκε γίνεται αντιληπτό ότι η πρόβλεψη Z_{t+1} προκύπτει ως ένας σταθμικός μέσος όρος των παρατηρήσεων της χρονοσειράς, καθώς το άθροισμα όλων των συντελεστών βαρύτητας ισούται με τη μονάδα. Επιπροσθέτως, διαπιστώνουμε ότι όσο αυξάνεται ο αριθμός των χρονικών περιόδων μεταξύ της πρόβλεψης και της πραγματικής τιμής της χρονοσειράς τόσο οι συντελεστές βαρύτητας μειώνονται και μάλιστα εκθετικά. Γι' αυτό και η υπό εξέταση μέθοδος καλείται απλή «εκθετική» εξομάλυνση. Γίνεται λοιπόν κατανοητό, ότι όσο μεγαλύτερη είναι η τιμή της σταθεράς εξομάλυνσης a τόσο πιο μεγάλη είναι η βαρύτητα που δίνεται στις πιο πρόσφατες παρατηρήσεις και τόσο πιο μικρή είναι η βαρύτητα που δίνεται στις πιο απομακρυσμένες παρατηρήσεις. Εναλλακτικά, η πρόβλεψη για την περίοδο t , Z_t , που γίνεται στην αρχή της περιόδου αυτής, θα είναι :

$$Z_t = a \times Y_{t-1} + a \times (1-a) \times Y_{t-2} + \dots + a \times (1-a)^n \times Z_{t-n+1}$$

Από τις σχέσεις που αναφέρθηκαν και που δίνουν το Z_{t+1} και Z_t προκύπτει ότι:

$$Z_{t+1} = a \times Y_t + (1-a) \times Z_t$$

όπου t είναι η τρέχουσα περίοδος, τα Z_{t+1} και Z_t , είναι τιμές πρόβλεψης για την επόμενη και την τρέχουσα περίοδο και Y_t είναι η τιμή που παρατηρήθηκε την τρέχουσα περίοδο. Αφού η παραπάνω εξίσωση περιλαμβάνει μόνο μία σταθερά, το μοντέλο αυτό είναι μοντέλο εκθετικής εξομάλυνσης μιας παραμέτρου. Η ανωτέρω σχέση αποτελεί την μαθηματική έκφραση της μεθόδου της απλής εκθετικής εξομάλυνσης για $t = 2, 3, \dots, n$ και με αρχική συνθήκη $Z_2 = Y_1$.

Εφαρμόσαμε λοιπόν τη μέθοδο της απλής εκθετικής εξομάλυνσης στο πρόβλημά μας για $a=0.3$, ακολουθώντας τη λογική που μόλις περιγράφηκε. Στόχος μας είναι ο υπολογισμός της τελικής βαθμολογίας των ομάδων. Για το σκοπό αυτό

για κάθε σεζόν ξεκινώντας από το 2004-2005 υπολογίσαμε τη διαφορά μεταξύ του ποσοστού νικών, ισοπαλιών και ηττών κάθε ομάδας της τελευταίας αγωνιστικής και του αντίστοιχου ποσοστού μετά το τέλος της μισής αγωνιστικής σεζόν. Έτσι, χρησιμοποιώντας το ποσοστό που είχε η κάθε ομάδα στο ήμισυ της τρέχουσας αγωνιστικής περιόδου και το μέσο όρο της αντίστοιχης υπολογιζόμενης διαφοράς υπολογίσαμε τα τελικά ποσοστά νικών, ισοπαλιών και ηττών των ομάδων.

Ο πίνακας που ακολουθεί παρουσιάζει τον τρόπο υπολογισμού της απλής εκθετικής εξομάλυνσης για την περίπτωση των νικών. Ανάλογη διαδικασία ακολουθήθηκε για τις ισοπαλίες και τις ήττες.

Πίνακας 6-3 : Μέθοδος Απλής Εκθετικής Εξομάλυνσης

Αγωνιστική Σεζόν	Παρατηρήσεις	Εκτίμηση Διαφοράς	Εκτίμηση Τελικού Ποσοστού
2004-2005	$Y_1 = [WP(30)-WP(15)]_{2004-5}$	-	-
2005-2006	$Y_2 = [WP(30)-WP(15)]_{2005-6}$	$Z_2=Y_1$	-
2006-2007	$Y_3 = [WP(30)-WP(15)]_{2006-7}$	$Z_3=a \times Y_2 + (1-a) \times Z_2$	-
2007-2008	$Y_4 = [WP(30)-WP(15)]_{2007-8}$	$Z_4=a \times Y_3 + (1-a) \times Z_3$	$W_1=WP(15)_{2007-8}+Z_4$
2008-2009	$Y_5 = [WP(30)-WP(15)]_{2008-9}$	$Z_5=a \times Y_4 + (1-a) \times Z_4$	$W_2=WP(15)_{2008-9}+Z_5$
2009-2010	$Y_6 = [WP(30)-WP(15)]_{2009-10}$	$Z_6=a \times Y_5 + (1-a) \times Z_5$	$W_3=WP(15)_{2009-10}+Z_6$
2010-2011	$Y_7 = [WP(30)-WP(15)]_{2010-11}$	$Z_7=a \times Y_6 + (1-a) \times Z_6$	$W_4=WP(15)_{2010-11}+Z_7$
2011-2012	$Y_8 = [WP(30)-WP(15)]_{2011-12}$	$Z_8=a \times Y_7 + (1-a) \times Z_7$	$W_5=WP(15)_{2011-12}+Z_8$
2012-2013	$Y_9 = [WP(30)-WP(15)]_{2012-13}$	$Z_9=a \times Y_8 + (1-a) \times Z_8$	$W_6=WP(15)_{2012-13}+Z_9$
2013-2014	-	$Z_{10}=a \times Y_9 + (1-a) \times Z_9$	$W_7=WP(15)_{2013-14}+Z_{10}$

Έτσι, έχουμε εκτιμήσει τα τελικά ποσοστά νικών, ισοπαλιών και ηττών των ομάδων. Στη συνέχεια, μετατρέπουμε τα ποσοστά αυτά σε πλήθος παιχνιδιών :

$$\text{Number_}W_i = W_i \times k, \text{ για } i=1,..7$$

όπου $k=34$ για τη σεζόν 2013-2014 και $k=30$ για όλες τις υπόλοιπες χρονιές.

Αντίστοιχα, υπολογίζουμε το πλήθος των ισοπαλιών (Number_D_i) και των ηττών (Number_L_i) της κάθε ομάδας. Τέλος, η τελική εκτιμώμενη βαθμολογία της κάθε μιας προκύπτει ως :

$$\text{Points} = 3 \times \text{Number_}W_i + 1 \times \text{Number_}D_i + 0 \times \text{Number_}L_i$$

6.4 Αποτελέσματα

Από την εφαρμογή των μεθόδων που αναφέρθηκαν προέκυψαν τα ακόλουθα αποτελέσματα :

Πίνακας 6-4 : Μέσο ποσοστιαίο σφάλμα βαθμολογίας

	2007-8	2008-9	2009-10	2010-11	2011-12	2012-13	2013-14	M.O
MA	15.87	13.86	15.67	8.41	16.68	11.63	14.44	13.79
WMA	16.45	13.84	15.05	8.17	18.41	11.39	13.50	13.83
EWMA	15.24	10.25	14.32	7.13	20.97	12.87	13.66	13.49

Από τα αποτελέσματα που προκύπτουν παρατηρούμε πως και οι τρεις μέθοδοι είναι αρκετά εύστοχες και το σφάλμα που παρέχουν είναι μικρό. Μάλιστα, για όλες τις χρονιές το σφάλμα που σημειώνεται είναι μικρότερο από 16% με εξαίρεση τη σεζόν 2011-2012 που είναι ελαφρώς μεγαλύτερο για τη μέθοδο του απλού κινητού μέσου, αγγίζει το 18.41% για τη μέθοδο του σταθμισμένου κινητού μέσου και φθάνει το 20.97% για τη μέθοδο της απλής εκθετικής εξομάλυνσης. Ιδίως το έτος 2010-2011 διακρίνεται το μικρότερο ποσοστό σφάλματος με τιμή μικρότερη από 8.5% και για τις τρεις μεθόδους. Συνολικά, ελαφρώς καλύτερα αποτελέσματα προσφέρει η μέθοδος της απλής εκθετικής εξομάλυνσης με $a=0.3$. Πιο αναλυτικά, για τις σεζόν 2007-2008, 2009-2010, 2012-2013 και 2013-2014 το σφάλμα που προκύπτει από την εφαρμογή των τριών μεθόδων είναι παρεμφερές. Για τη χρονιά 2008-2009 παρατηρούμε υπεροχή της απλής εκθετικής εξομάλυνσης κατά 3% περίπου και για τη χρονιά 2010-2011 κατά 1% περίπου. Τέλος, για το έτος 2011-2012 φαίνεται πως υπερτερεί η μέθοδος του απλού κινητού μέσου κατά 2% περίπου έναντι του σταθμισμένου κινητού μέσου και κατά 4% περίπου έναντι της απλής εκθετικής εξομάλυνσης.

Το βασικό όμως μειονέκτημα των μεθόδων εξομάλυνσης είναι ότι οι ομάδες θα πρέπει να βρίσκονται στη Superleague για 3 συνεχόμενα έτη πριν την προβλεπόμενη σεζόν. Στον πίνακα που ακολουθεί παρουσιάζονται οι ομάδες βάσει της πραγματικής, τελικής τους κατάταξης. Με κόκκινο χρώμα σημειώνονται οι ομάδες, οι οποίες δε διαθέτουν την απαιτούμενη προϋστορία και για τις οποίες δε γίνεται κάποια εκτίμηση:

Πίνακας 6-5 : Ομάδες που συμμετέχουν στην Superleague τις σεζόν 2007-2008 έως 2010-2011

2007-2008	2008-2009	2009-2010	2010-2011
Ολυμπιακός	Ολυμπιακός	Παναθηναϊκός	Ολυμπιακός
ΑΕΚ	ΠΑΟΚ	Ολυμπιακός	Παναθηναϊκός
Παναθηναϊκός	Παναθηναϊκός	ΠΑΟΚ	ΑΕΚ
Άρης	ΑΕΚ	ΑΕΚ	ΠΑΟΚ
Πανιώνιος	Λάρισα	Άρης	Ολυμπιακός Βόλου
Λάρισα	Άρης	Καβάλα	Άρης
Αστέρας Τρίπολης	Skoda Ξάνθη	Ατρόμητος	Καβάλα
Skoda Ξάνθη	Πανιώνιος	Λάρισα	Εργοτέλης
ΠΑΟΚ	Εργοτέλης	Πανιώνιος	Skoda Ξάνθη
Ηρακλής	Ηρακλής	Ηρακλής	Πανιώνιος
Λεβαδειακός	Πανθρακικός	Εργοτέλης	Ατρόμητος
ΟΦΗ	Αστέρας Τρίπολης	Αστέρας Τρίπολης	Κέρκυρα
Εργοτέλης	Λεβαδειακός	Skoda Ξάνθη	Αστέρας Τρίπολης
Ατρόμητος	ΟΦΗ	Λεβαδειακός	Ηρακλής
Βέροια	Πανσερραϊκός	Πας Γιάννενα	Λάρισα
Απόλλων Σμύρνης	Θρασύβουλος	Πανθρακικός	Πανσερραϊκός

Πίνακας 6-6 : Ομάδες που συμμετέχουν στην Superleague τις σεζόν 2011-2012 έως 2013-2014

2011-2012	2012-2013	2013-2014
Ολυμπιακός	Ολυμπιακός	Ολυμπιακός
Παναθηναϊκός	ΠΑΟΚ	ΠΑΟΚ
ΠΑΟΚ	Αστέρας Τρίπολης	Ατρόμητος
Ατρόμητος	Ατρόμητος	Παναθηναϊκός
ΑΕΚ	Πας Γιάννενα	Αστέρας Τρίπολης
Αστέρας Τρίπολης	Παναθηναϊκός	ΟΦΗ
Λεβαδειακός	Skoda Ξάνθη	Εργοτέλης
Πας Γιάννενα	Πανιώνιος	Παναιτωλικός
Άρης	Πλατανιάς Χανίων	Λεβαδειακός
ΟΦΗ	Πανθρακικός	Πανθρακικός
Skoda Ξάνθη	Λεβαδειακός	Πας Γιάννενα
Πανιώνιος	Βέροια	ΑΕΛ Καλλονής
Κέρκυρα	Άρης	Πανιώνιος
Εργοτέλης	ΟΦΗ	Πλατανιάς Χανίων
Παναιτωλικός	ΑΕΚ	Βέροια
Δόξα Δράμας	Κέρκυρα	Skoda Ξάνθη
-	-	Απόλλων Σμύρνης
-	-	Άρης

7. Συμπεράσματα και επεκτάσεις

7.1 Συμπεράσματα

Στην παρούσα διπλωματική εργασία μελετήθηκε το πρόβλημα της πρόβλεψης των αποτελεσμάτων αγώνων ποδοσφαίρου. Παρά τη μεγάλη δυσκολία του συγκεκριμένου προβλήματος, λόγω της τυχειότητας που υπάρχει, καθώς επίσης και των πολλών παραγόντων που επιδρούν σε αυτό, καταφέραμε με τη χρήση διαφόρων τεχνικών να λάβουμε αξιόλογα αποτελέσματα. Από τις μεθόδους που μελετήσαμε και αναλύσαμε στα προηγούμενα κεφάλαια διακρίνεται ξεκάθαρα πως τα καλύτερα αποτελέσματα, όσον αφορά στο μέσο ποσοστιαίο σφάλμα βαθμολογίας τα παρέχουν οι μέθοδοι της εξομάλυνσης και κυρίως η μέθοδος της απλής εκθετικής εξομάλυνσης. Διαπιστώσαμε δηλαδή πως χρησιμοποιώντας την τελική βαθμολογία των προηγούμενων ετών μιας ομάδας, την αντίστοιχη τωρινή έως τα μέσα της σεζόν και με τη χρήση των μεθόδων εξομάλυνσης εξάγουμε σημαντικά αποτελέσματα για την πρόβλεψη της τελικής βαθμολογίας.

Όσον αφορά στα υπόλοιπα κριτήρια σύγκρισης διακρίνουμε υπεροχή των μεθόδων Home Team και RPI Weighted. Διακρίναμε δηλαδή πόσο καθοριστική είναι η έδρα στη διεξαγωγή ενός ποδοσφαιρικού αγώνα και στην έκβαση του τελικού αποτελέσματος και πως βάσει ενός απλοϊκού μοντέλου είναι δυνατόν να παραχθούν σημαντικά αποτελέσματα. Επιπλέον, διαπιστώσαμε πως η μέθοδος RPI Weighted, η οποία χρησιμοποιείται ευρέως σε αθλήματα όπως το μπάσκετ και το μπίτζμπολ μπορεί να εφαρμοστεί και στην περίπτωση του ποδοσφαίρου παρέχοντας σημαντικά αποτελέσματα. Η RPI Weighted και η Home Team έχουν ίσο ποσοστό ακρίβειας πρόβλεψης (49.1%), με την RPI Weighted να έχει μικρότερο σφάλμα στην περίπτωση πλήρους ανατροπής της πρόβλεψης και την Home Team να έχει μικρότερο σφάλμα στην εκτίμηση των ισοπαλιών και της τελικής βαθμολογίας.

Ακόμα, συμπεράναμε πως η μέθοδος του clustering και ειδικότερα η εφαρμογή του k-means προσέφερε το μικρότερο ποσοστό αστοχίας από όλες τις μεθόδους, όσον αφορά στους αγώνες όπου η εκτιμώμενη νικήτρια ομάδα, ήταν τελικά η ηττημένη, όμως αύξησε σημαντικά το ποσοστό σφάλματος ισοπαλιών.

Επιπρόσθετα, με τη χρήση της μεθόδου Last Year Score λάβαμε μικρό ποσοστό ευστοχίας. Άρα, η χρήση των αποτελεσμάτων της προηγούμενης χρονιάς δε συντελεί ιδιαίτερα στην πρόβλεψη των αποτελεσμάτων της επόμενης. Τέλος, η χρήση δύο ανεξάρτητων κατανομών Poisson για τη περιγραφή των γκολ της γηπεδούχου και της φιλοξενούμενης ομάδας προσέφερε θετικά αποτελέσματα όσον αφορά την αναμενόμενη συχνότητα εμφάνισης γκολ της γηπεδούχου και της φιλοξενούμενης ομάδας, όμως το ποσοστό ευστοχίας προβλέψεων που παρείχε δεν είναι ιδιαίτερα σημαντικό.

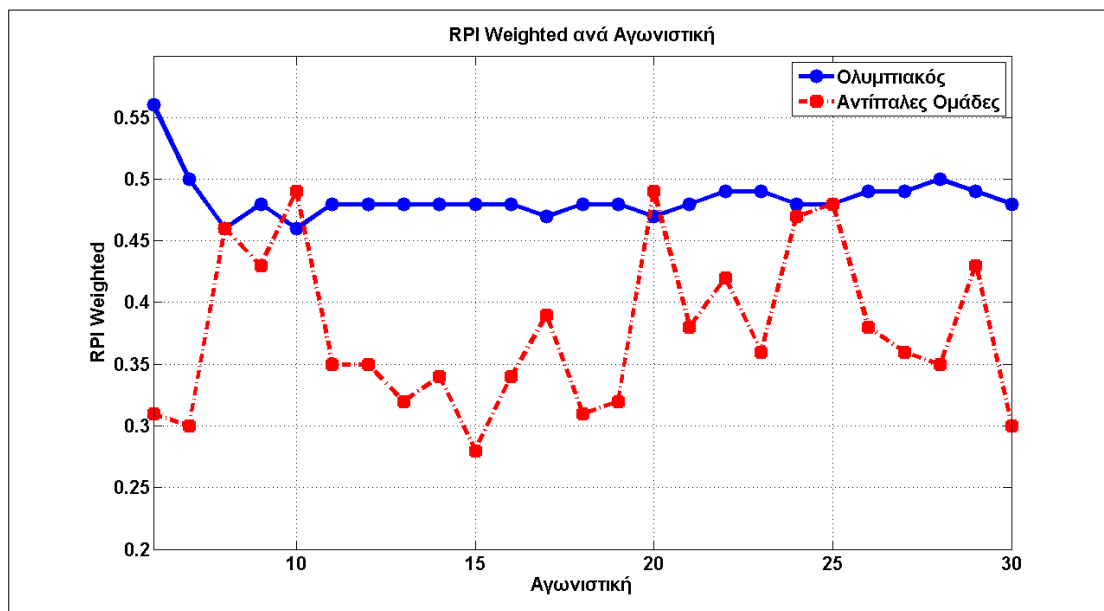
7.2 Επεκτάσεις

Βελτιώσεις και επεκτάσεις στα μοντέλα και τις μεθόδους που εφαρμόστηκαν στην παρούσα εργασία θα μπορούσαν να περιλαμβάνουν:

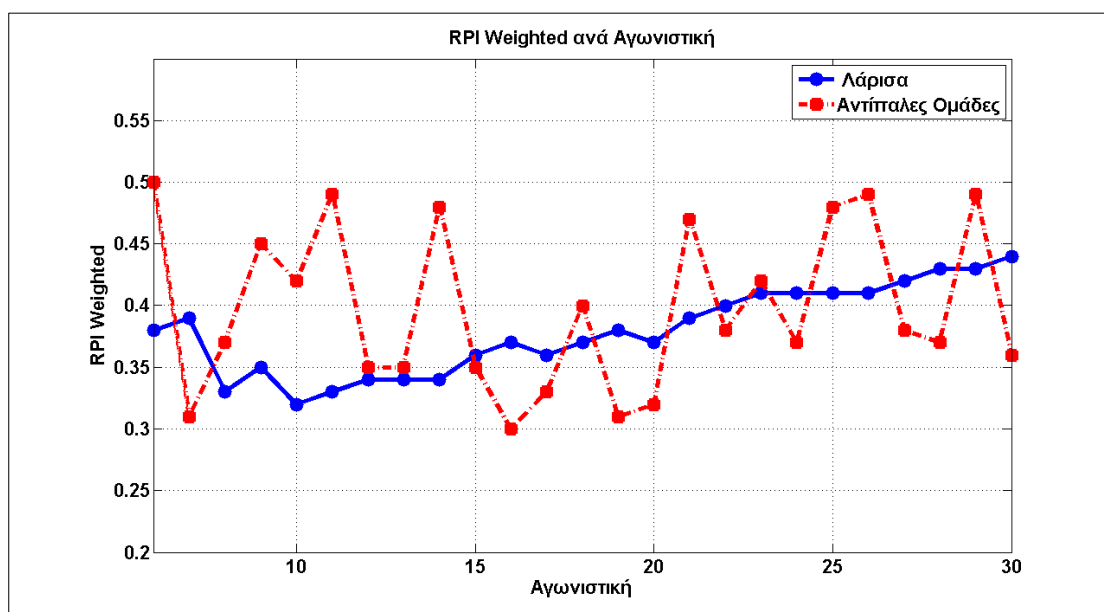
- Την περαιτέρω μελέτη των δεδομένων, ώστε να εξαχθούν από αυτά περισσότερα χαρακτηριστικά και την ένταξη όσων κρίνονται ουσιώδη στις μεθόδους που αναλύθηκαν.
- Την μεμονωμένη καταγραφή και μελέτη των ισοπαλιών, ώστε να προκύψουν συμπεράσματα για το ποιοί παράγοντες την ευνοούν.
- Την εφαρμογή των μεθόδων και σε άλλα πρωταθλήματα, ώστε να διαπιστωθεί αν η ευστοχία τους είναι μεμονωμένη για το ελληνικό πρωτάθλημα ή καθολική.
- Την δημιουργία ενός μοντέλου που να περιλαμβάνει την από κοινού χρήση των μεθόδων Home Team και RPI Weighted.

8. Παράρτημα

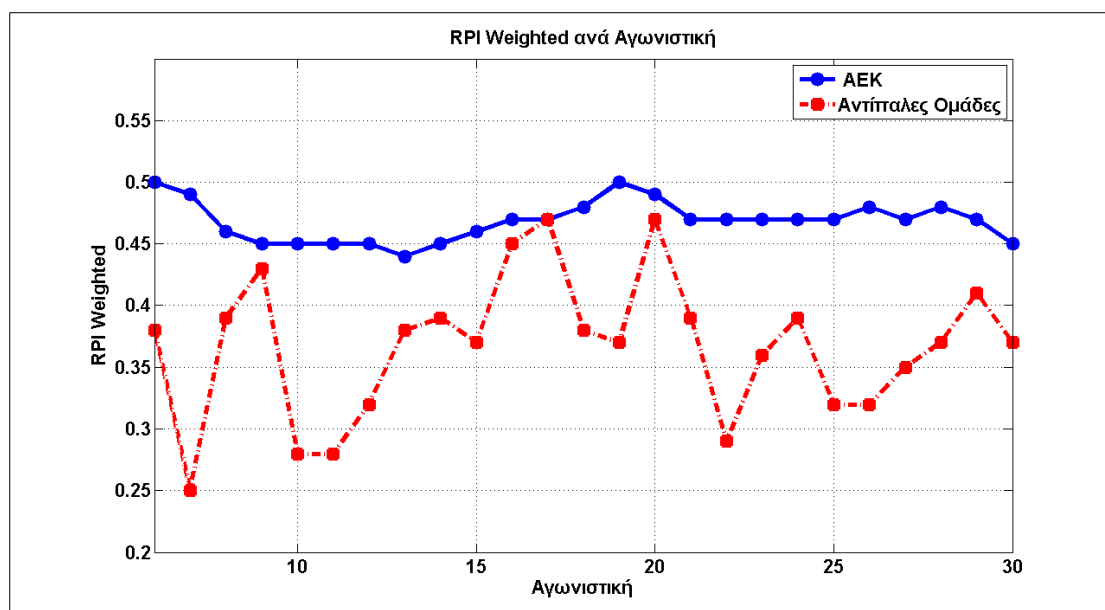
- Για τη σεζόν 2008-2009:



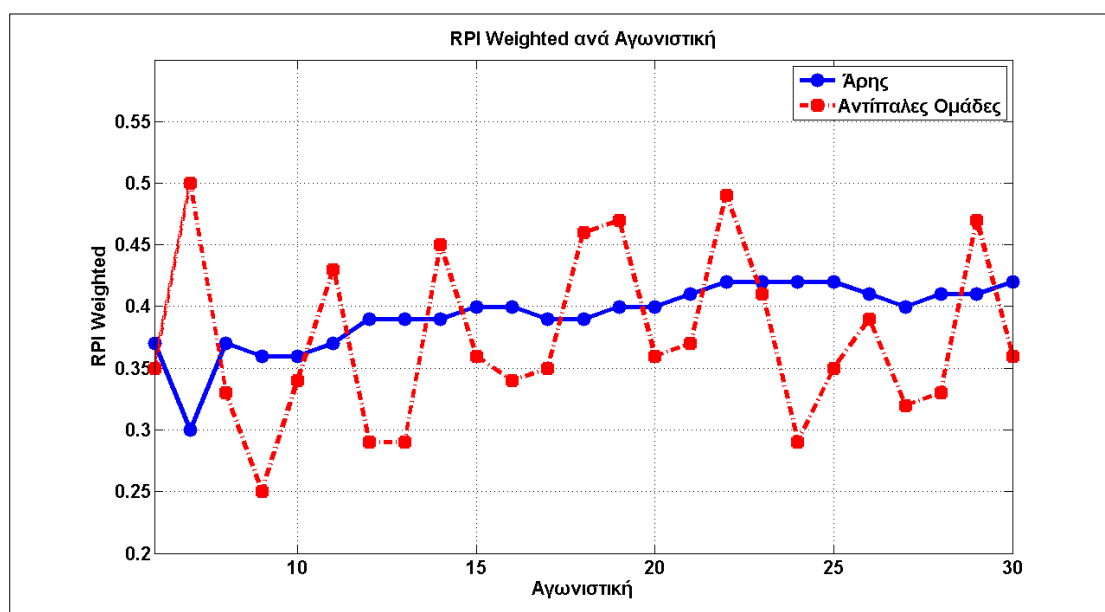
Εικόνα 8-1 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική



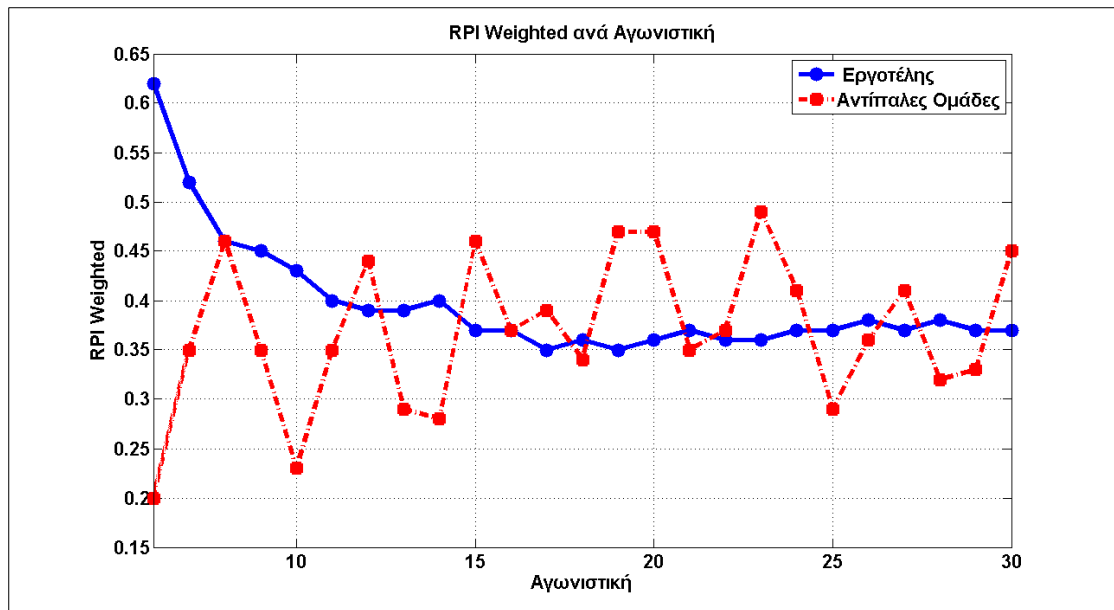
Εικόνα 8-2 : Οι τιμές του RPI Weighted της Λάρισας και των αντιπάλων της ανά αγωνιστική



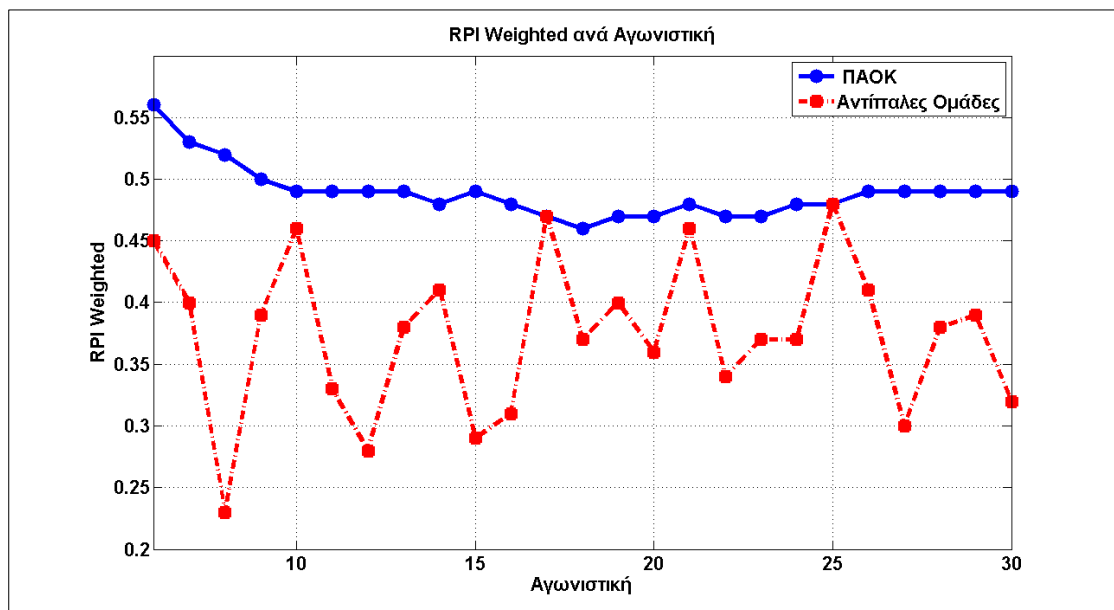
Εικόνα 8-3 : Οι τιμές του RPI Weighted της AEK και των αντιπάλων της ανά αγωνιστική



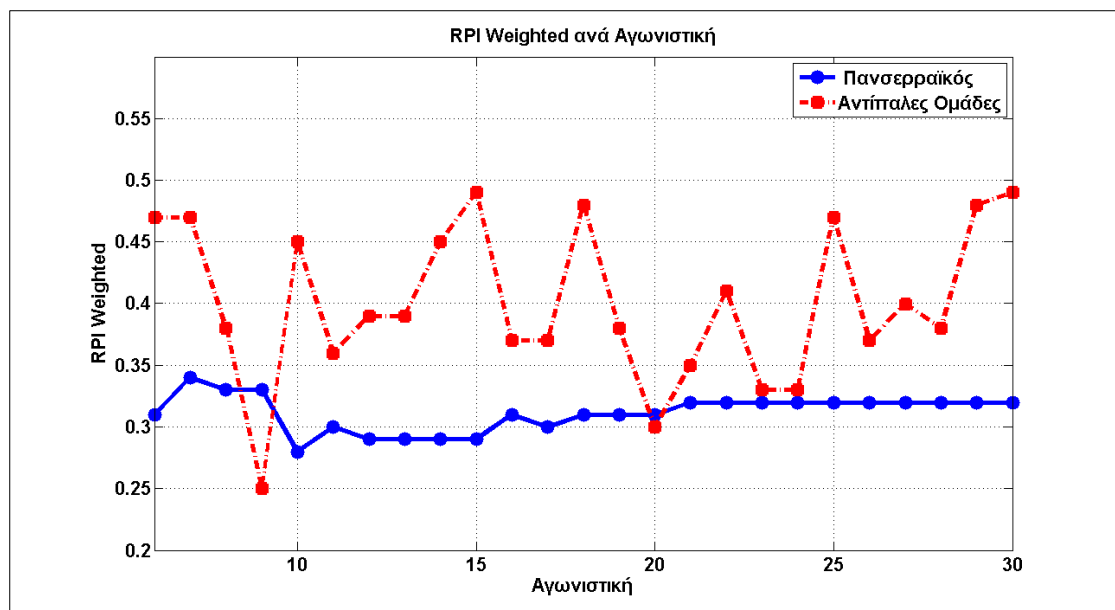
Εικόνα 8-4 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική



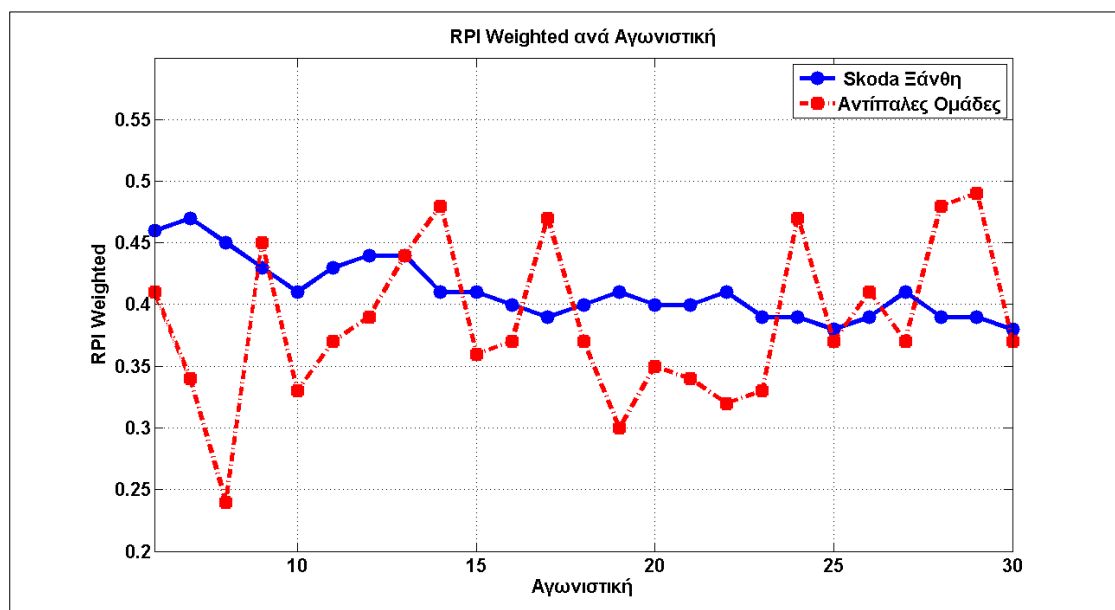
Εικόνα 8-5 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική



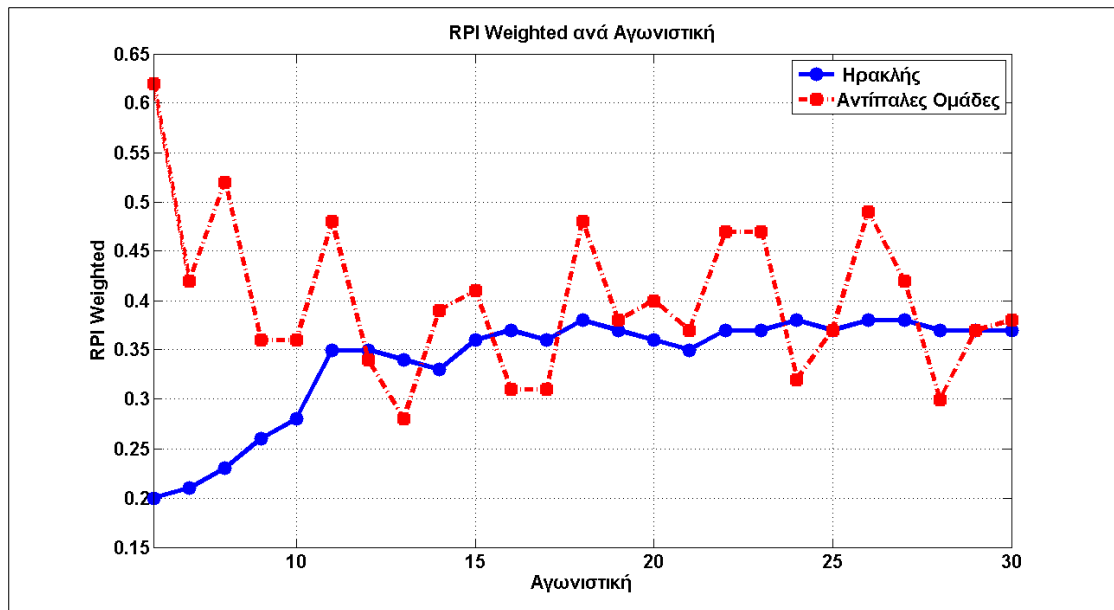
Εικόνα 8-6 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική



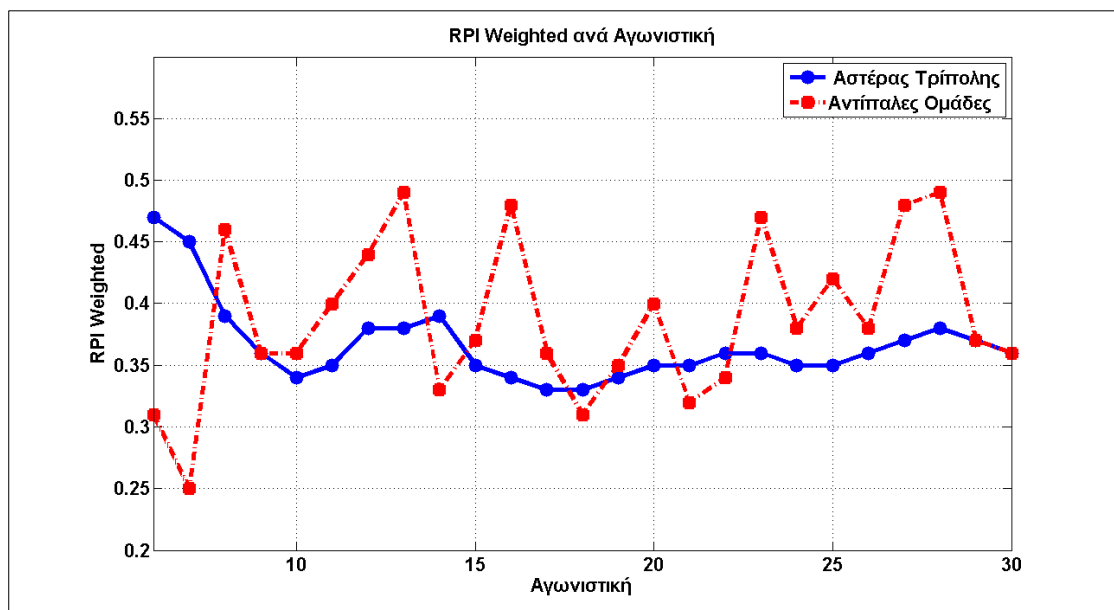
Εικόνα 8-7 : Οι τιμές του RPI Weighted του Πανσερραϊκού και των αντιπάλων του ανά αγωνιστική



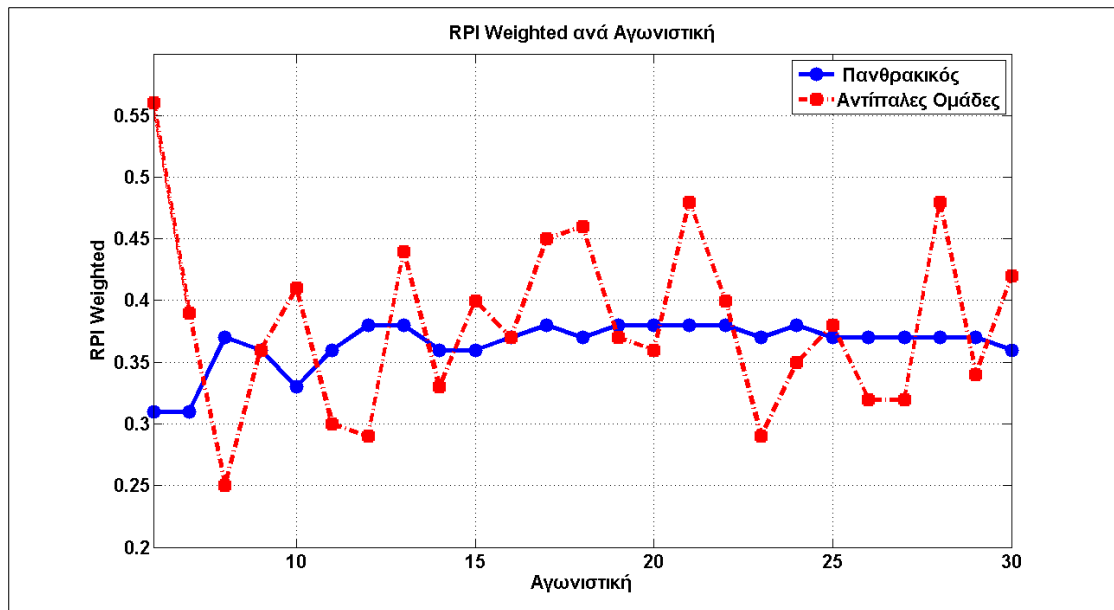
Εικόνα 8-8 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική



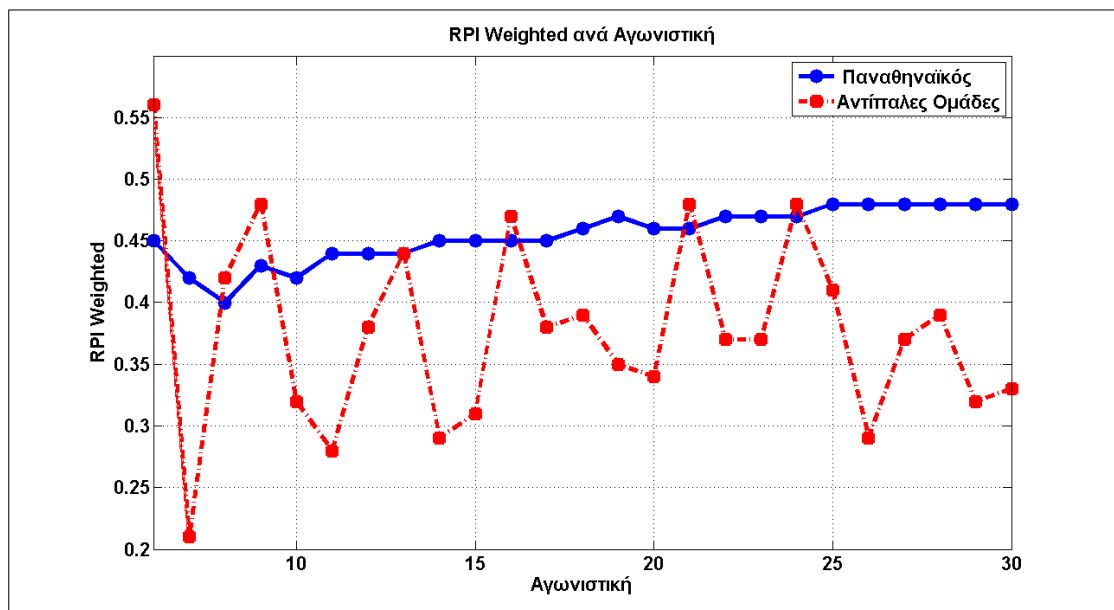
Εικόνα 8-9 : Οι τιμές του RPI Weighted του Ηρακλή και των αντιπάλων του ανά αγωνιστική



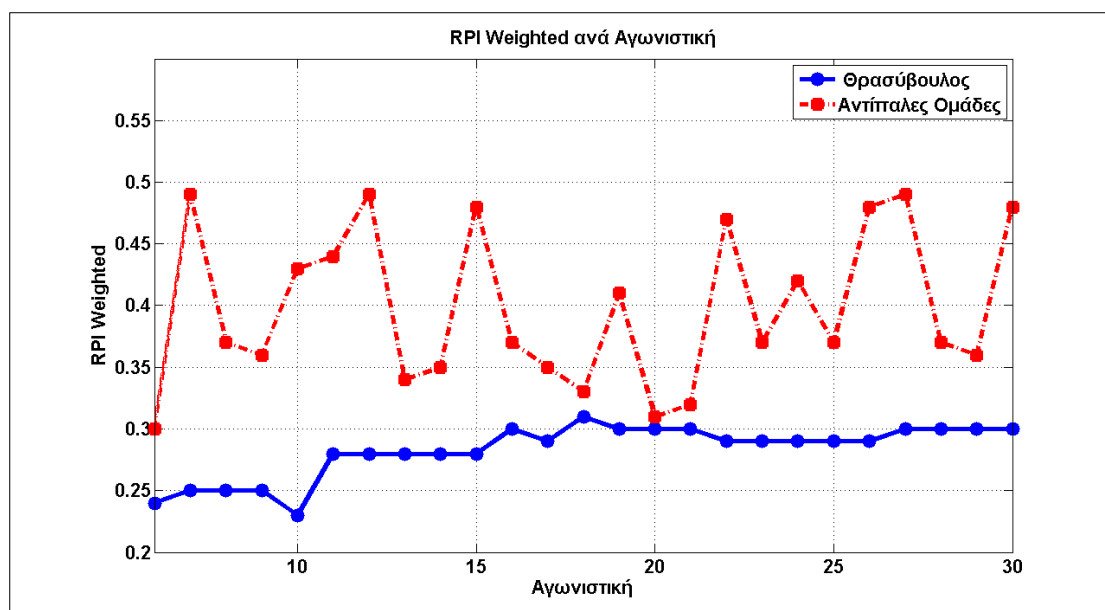
Εικόνα 8-10 : Οι τιμές του RPI Weighted του Αστήρα Τρίπολης και των αντιπάλων του ανά αγωνιστική



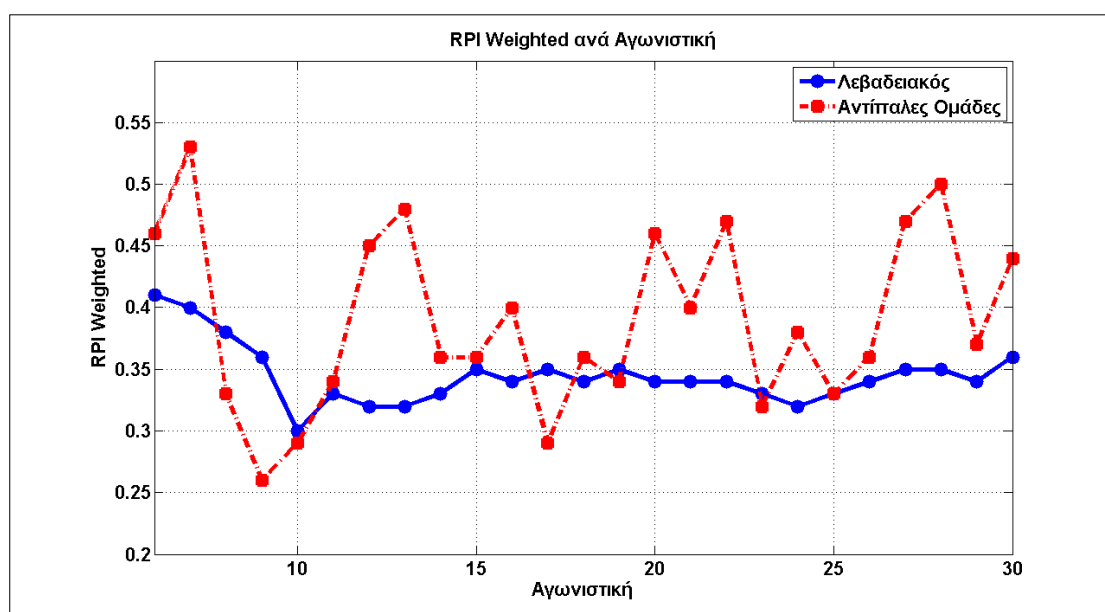
Εικόνα 8-11 : Οι τιμές του RPI Weighted του Πανθρακικού και των αντιπάλων του ανά αγωνιστική



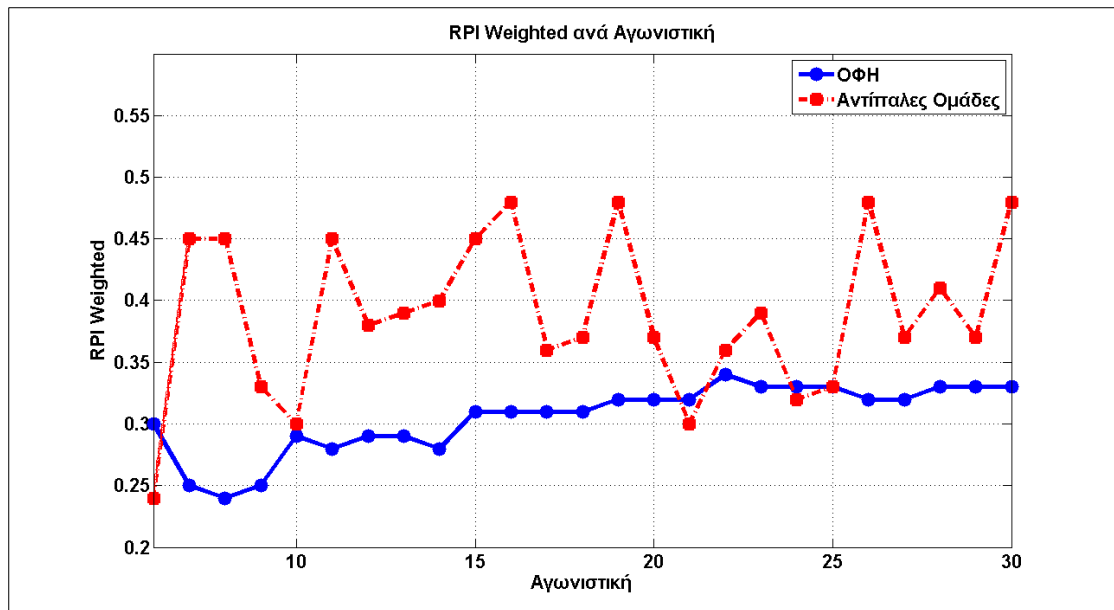
Εικόνα 8-12 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική



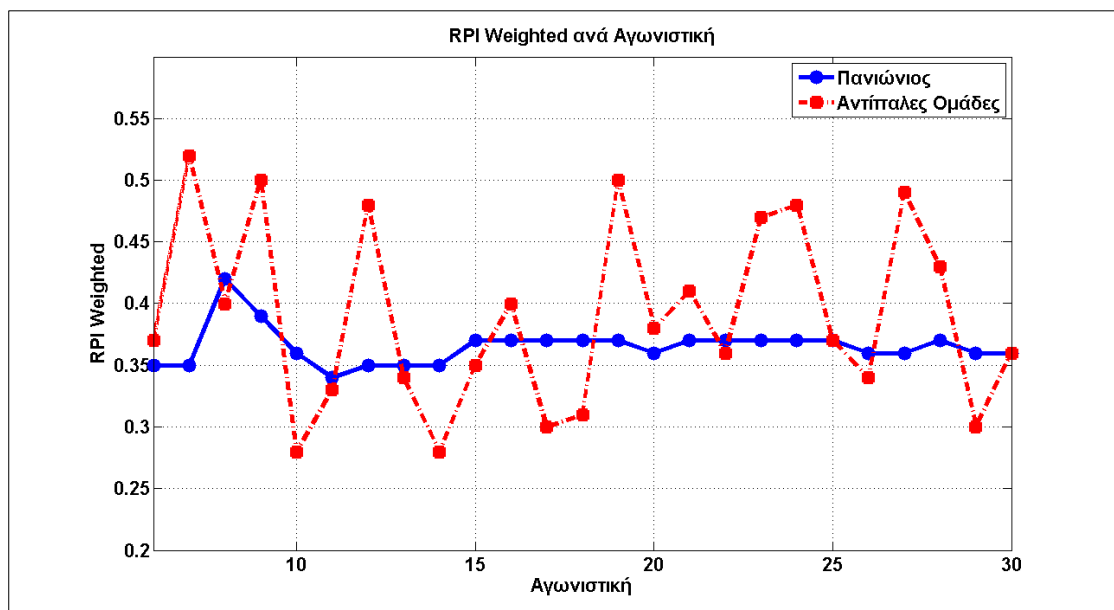
Εικόνα 8-13 : Οι τιμές του RPI Weighted του Θρασύβουλου και των αντιπάλων του ανά αγωνιστική



Εικόνα 8-14 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική

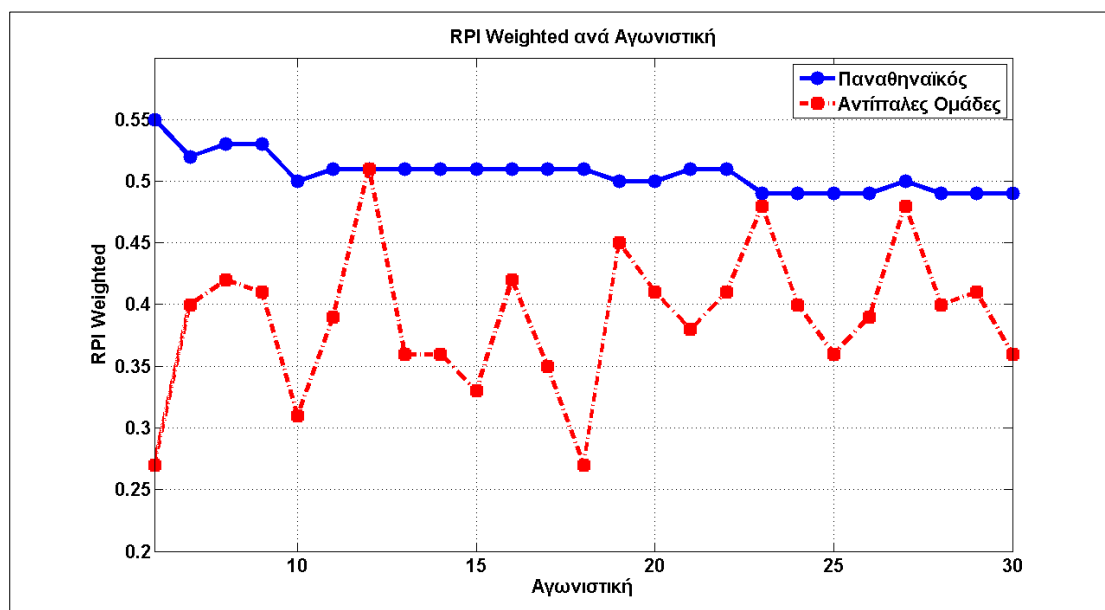


Εικόνα 8-15 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική

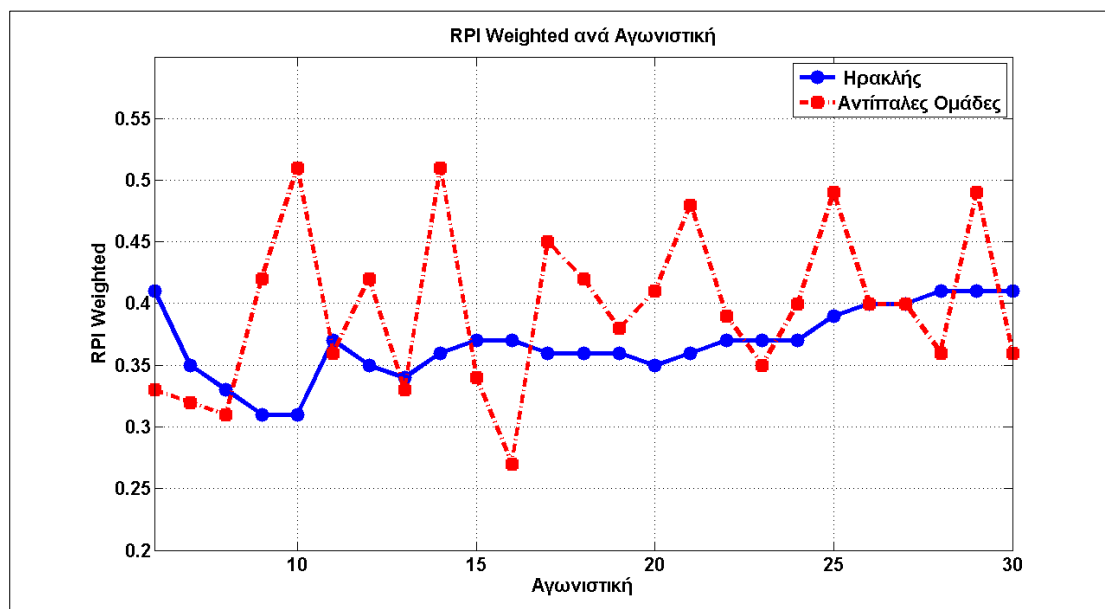


Εικόνα 8-16 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική

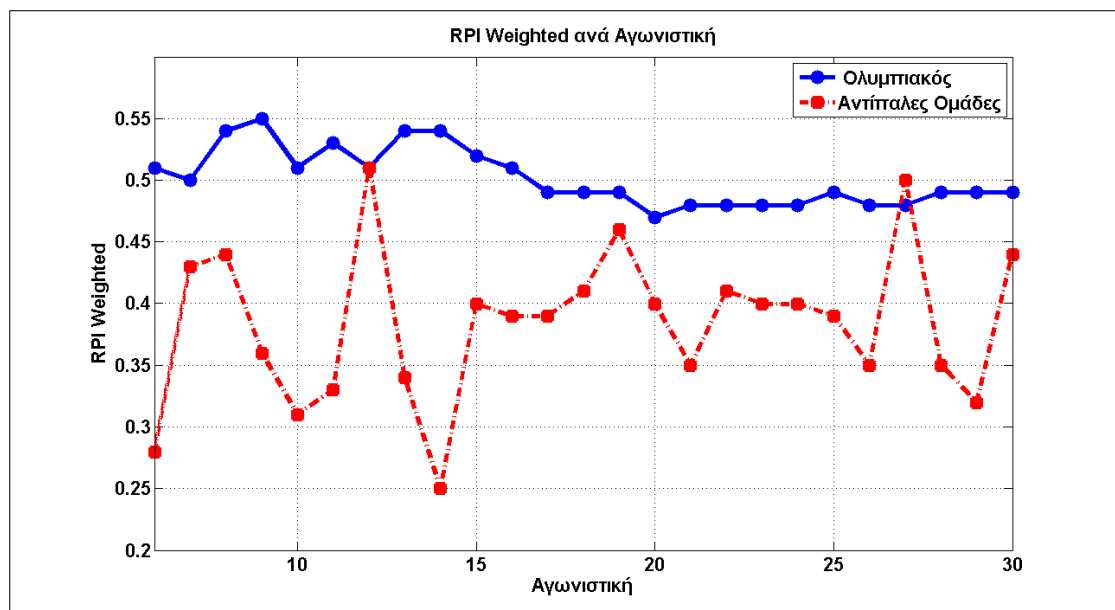
- Για τη σεζόν 2009-2010:



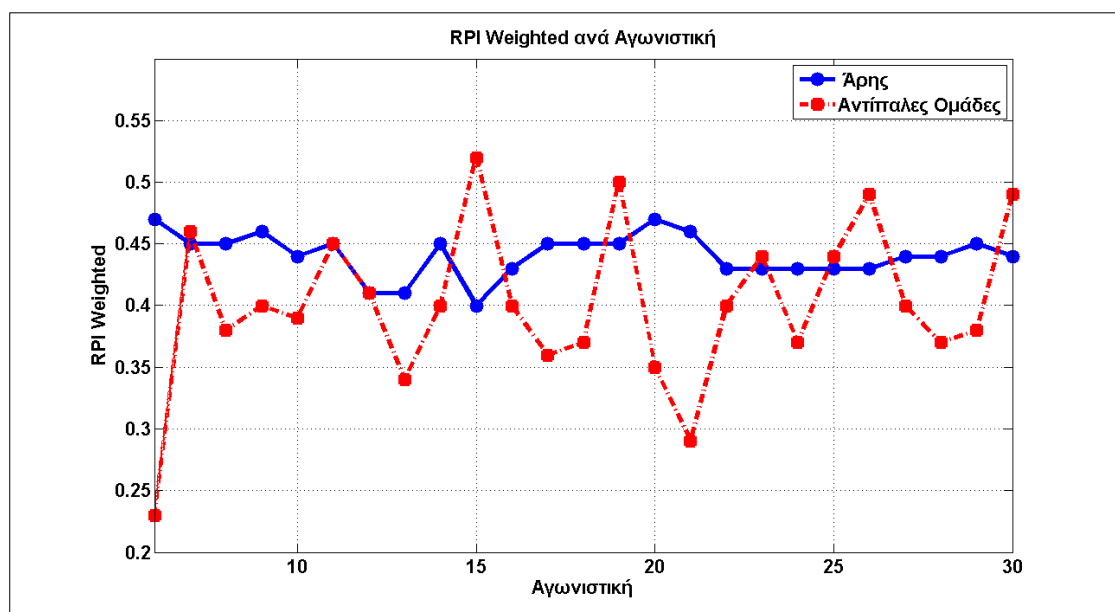
Εικόνα 8-17 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική



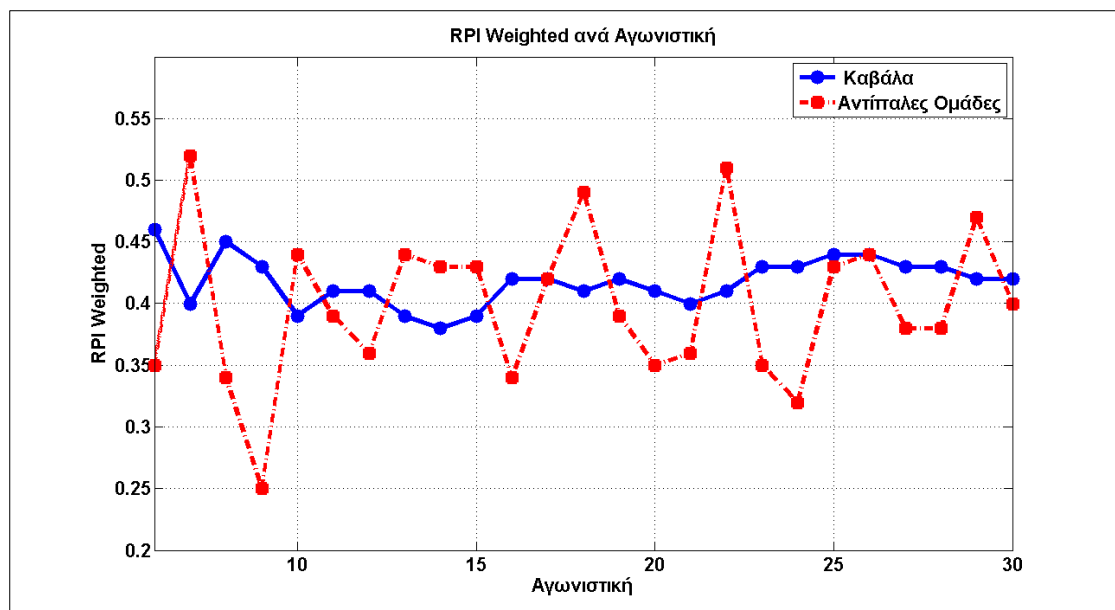
Εικόνα 8-18 : Οι τιμές του RPI Weighted του Ηρακλή και των αντιπάλων του ανά αγωνιστική



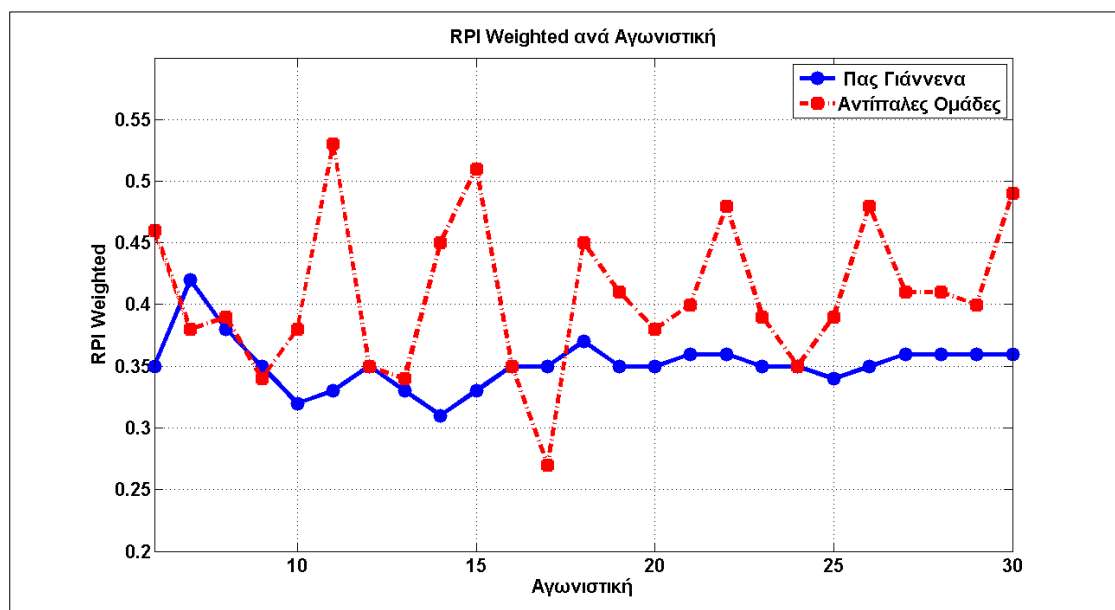
Εικόνα 8-19 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική



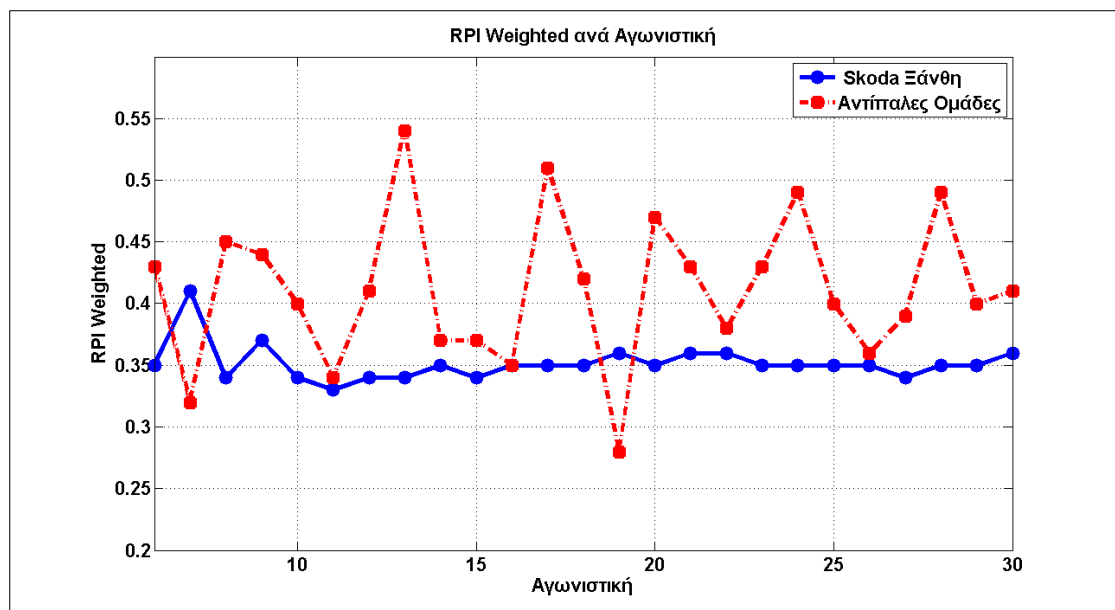
Εικόνα 8-20 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική



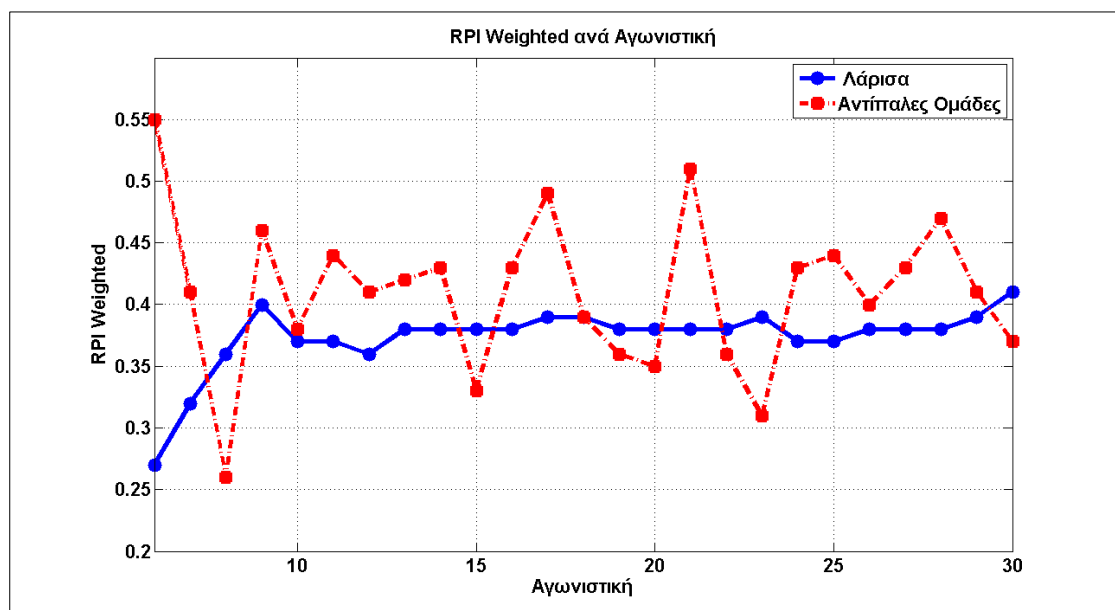
Εικόνα 8-21 : Οι τιμές του RPI Weighted της Καβάλας και των αντιπάλων της ανά αγωνιστική



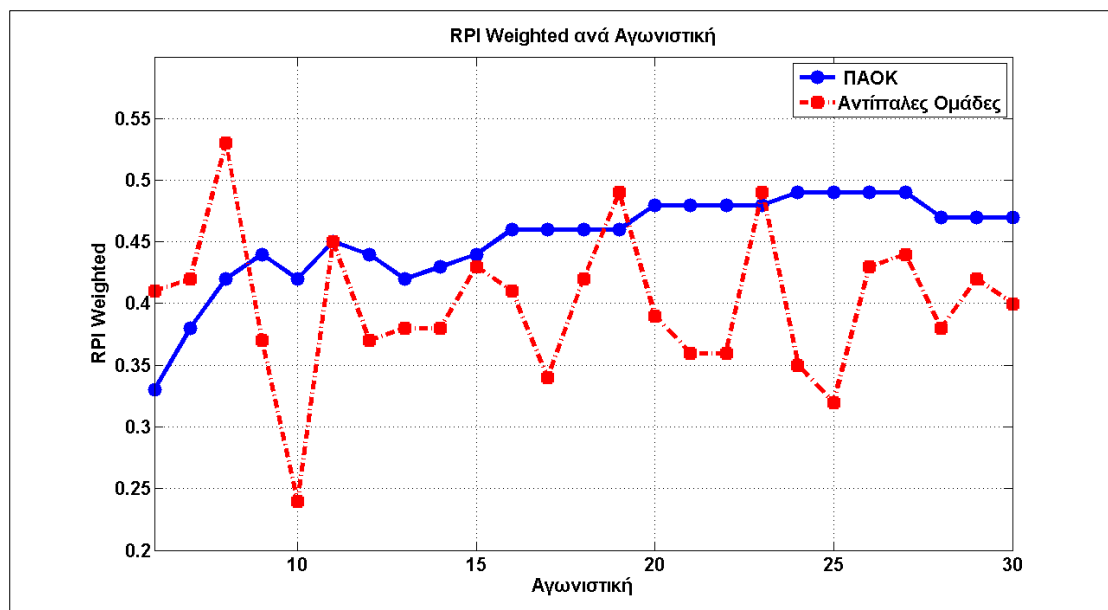
Εικόνα 8-22 : Οι τιμές του RPI Weighted του Πας Γιάννενα και των αντιπάλων του ανά αγωνιστική



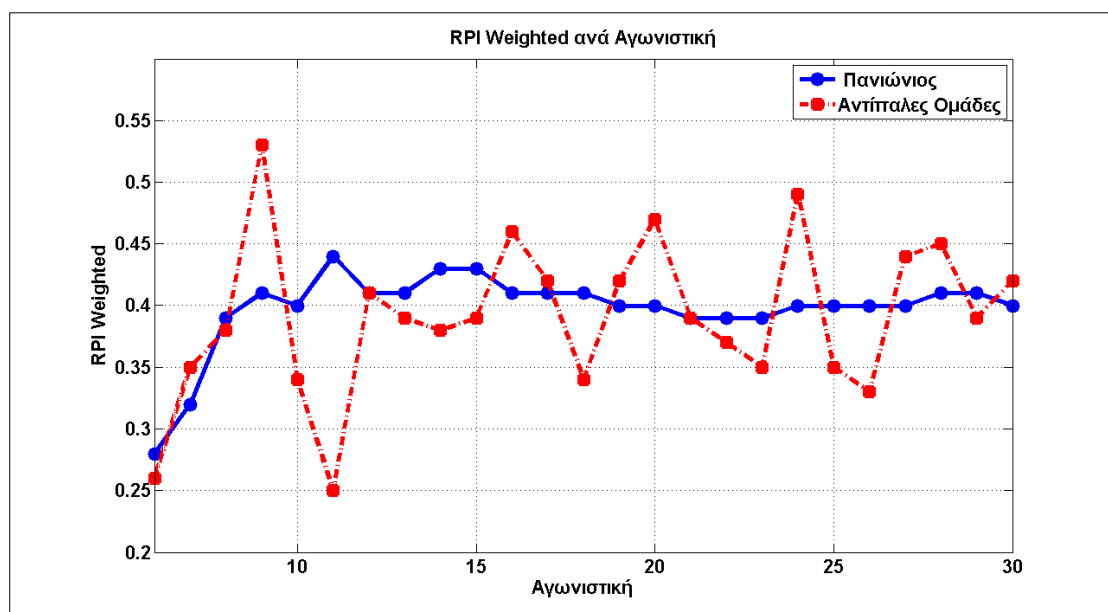
Εικόνα 8-23 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική



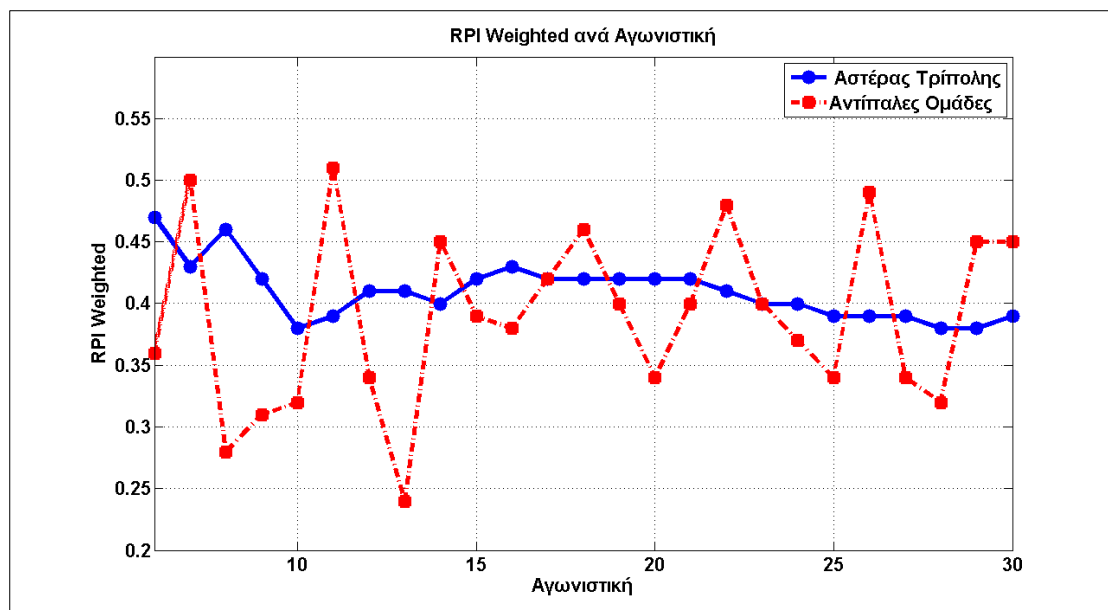
Εικόνα 8-24 : Οι τιμές του RPI Weighted της Λάρισας και των αντιπάλων της ανά αγωνιστική



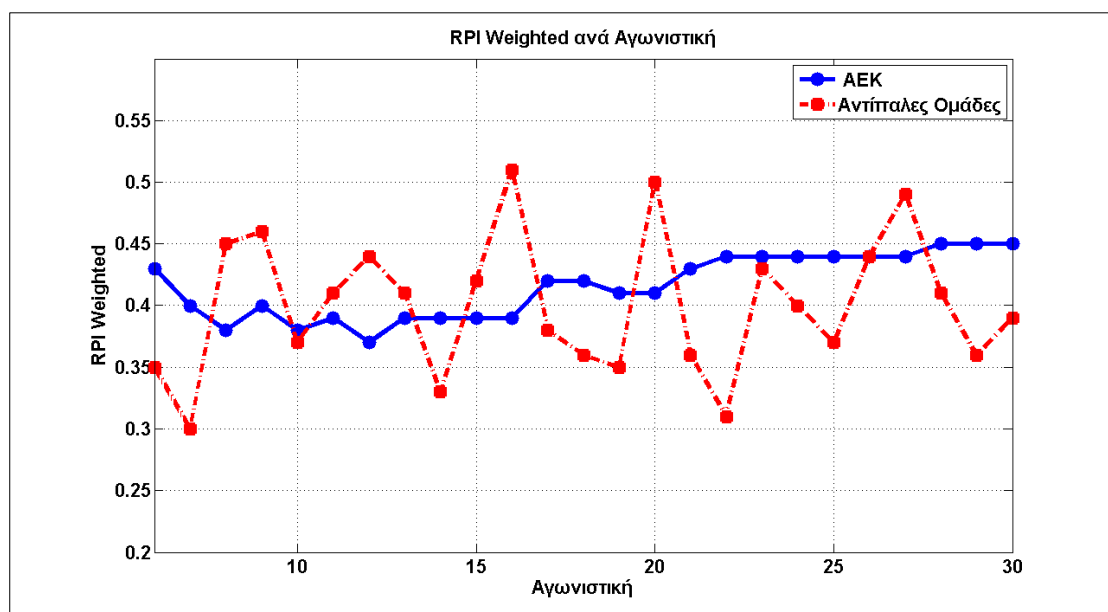
Εικόνα 8-25 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική



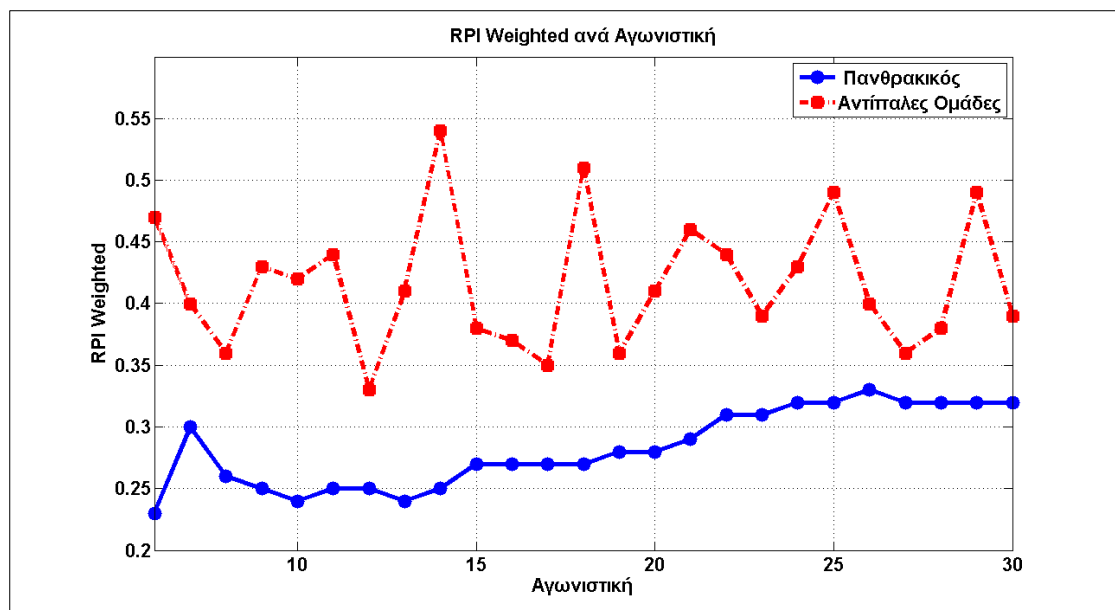
Εικόνα 8-26 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική



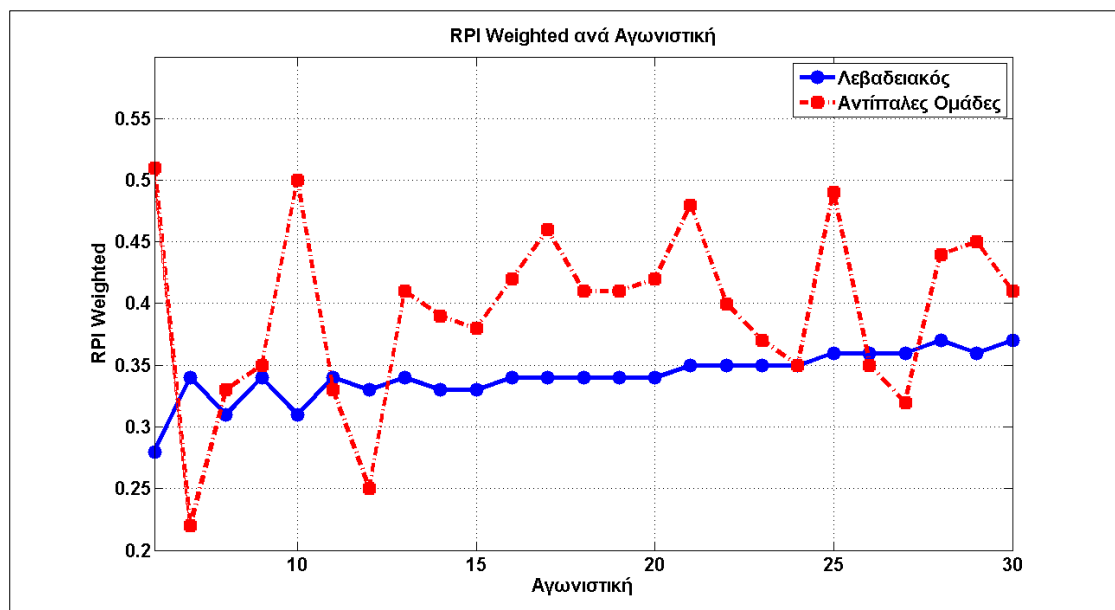
Εικόνα 8-27 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική



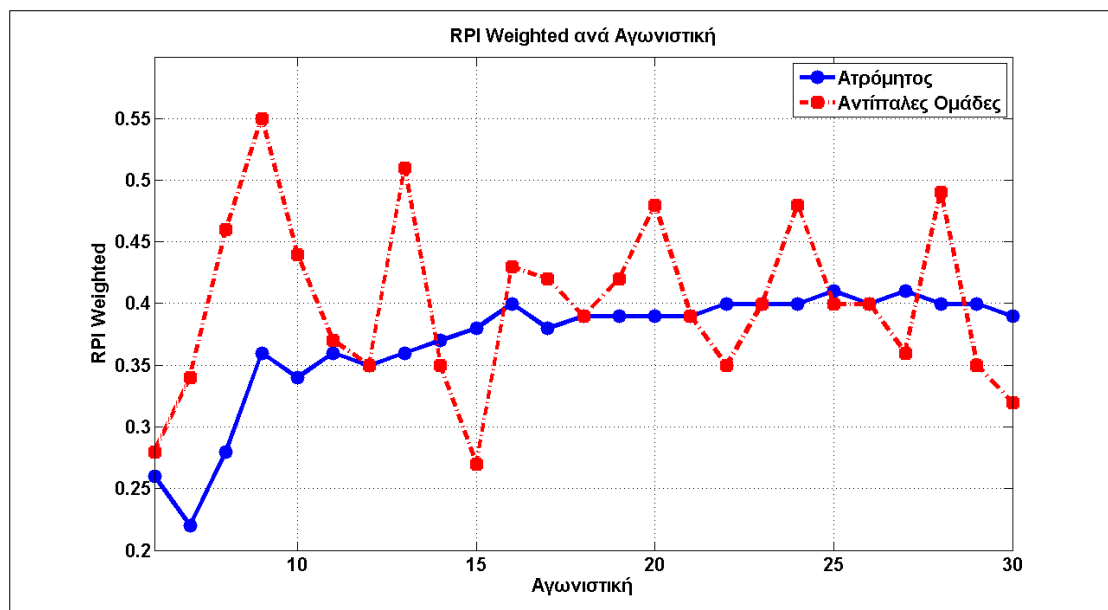
Εικόνα 8-28 : Οι τιμές του RPI Weighted της ΑΕΚ και των αντιπάλων της ανά αγωνιστική



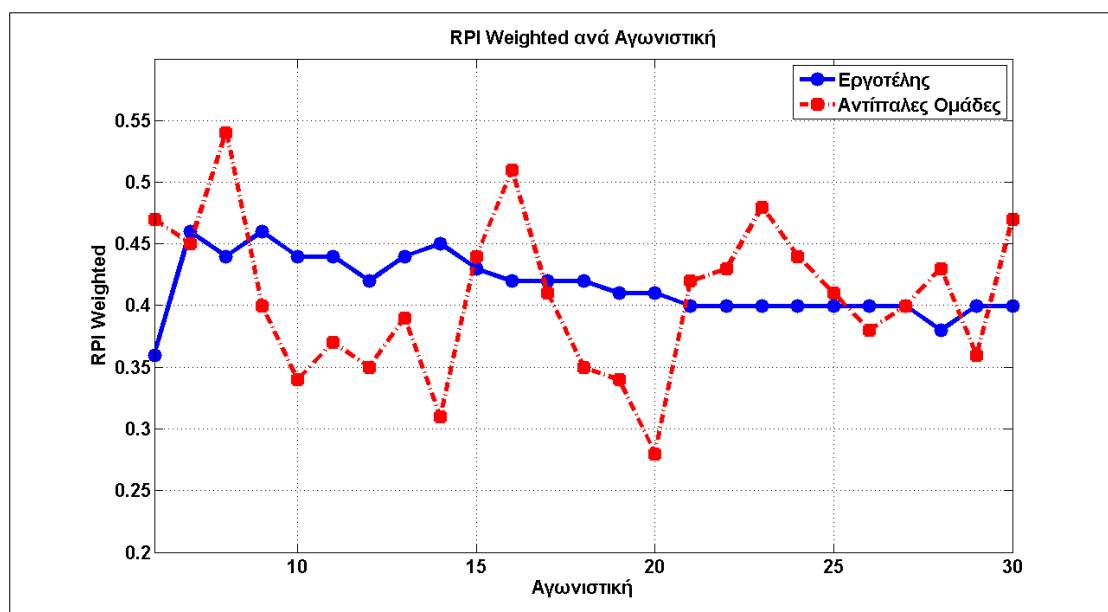
Εικόνα 8-29 : Οι τιμές του RPI Weighted του Πανθρακικού και των αντιπάλων του ανά αγωνιστική



Εικόνα 8-30 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική

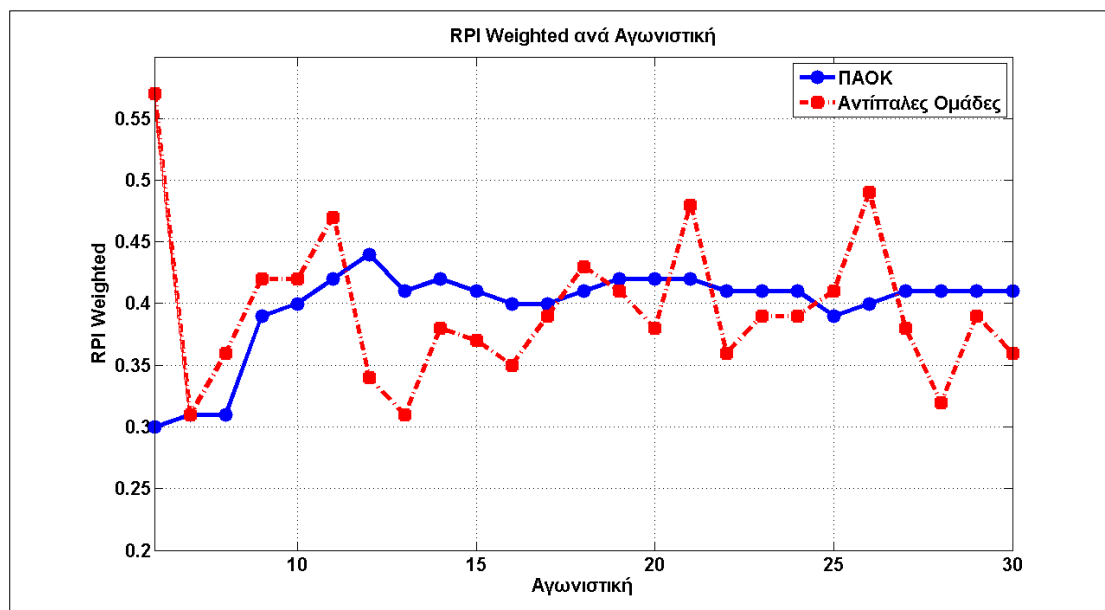


Εικόνα 8-31 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική

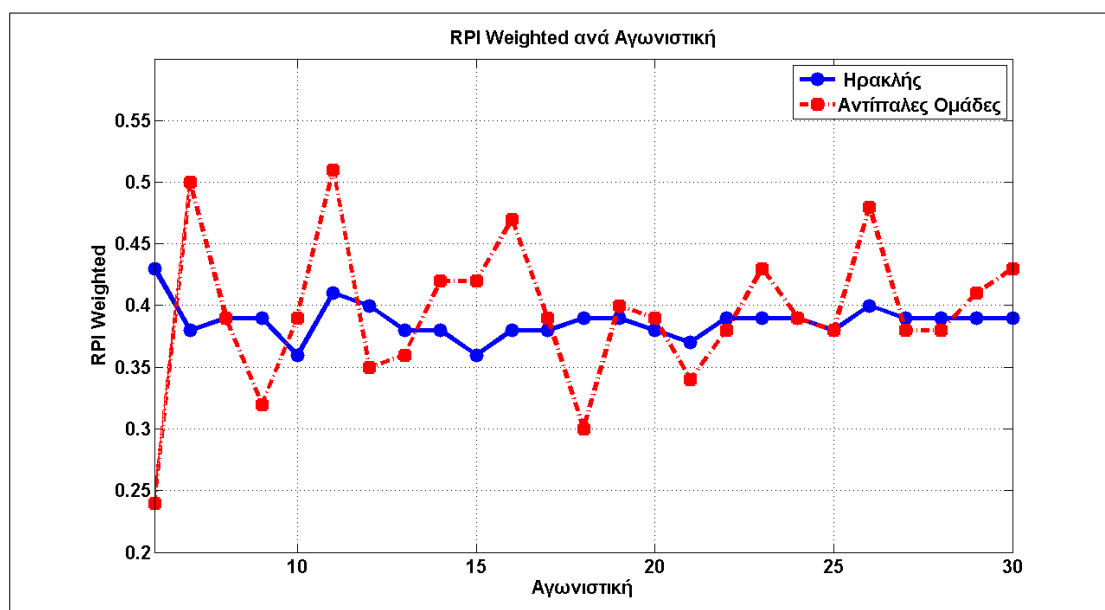


Εικόνα 8-32 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική

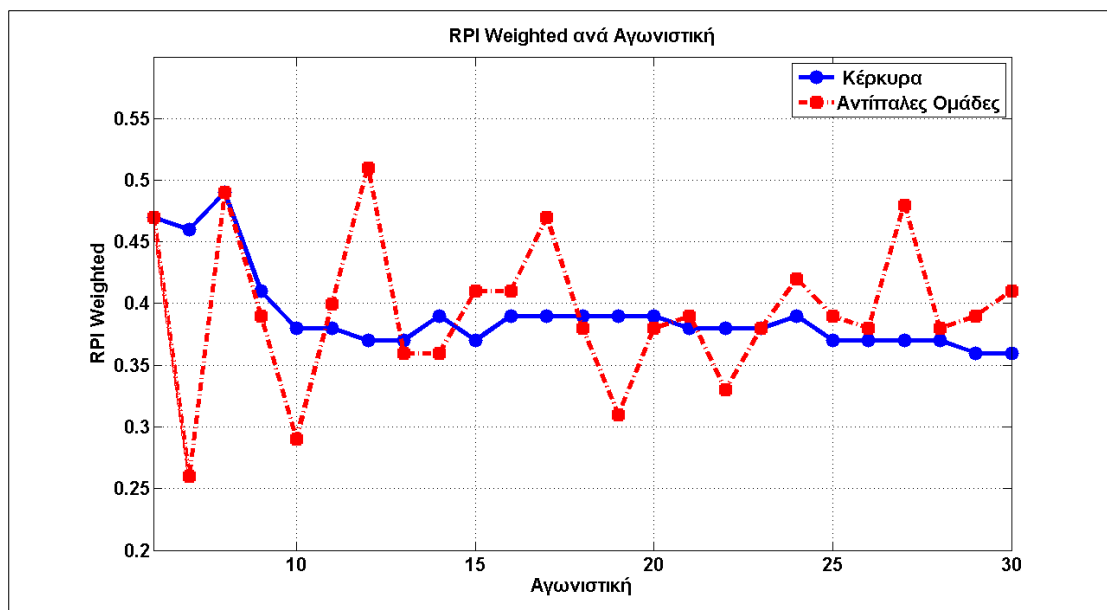
- Για τη σεζόν 2010-2011 :



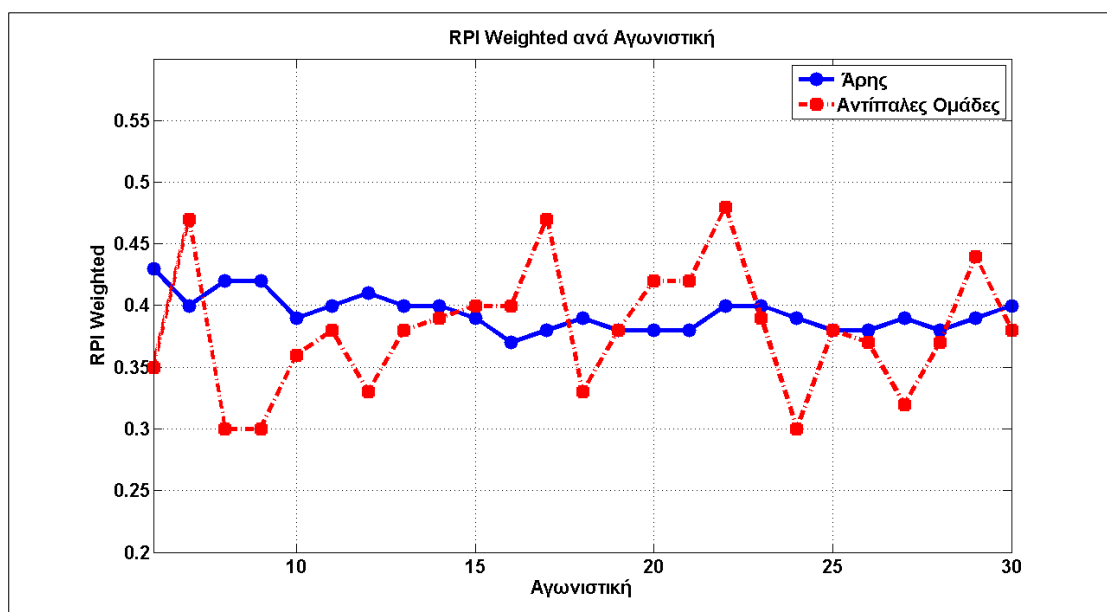
Εικόνα 8-33 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική



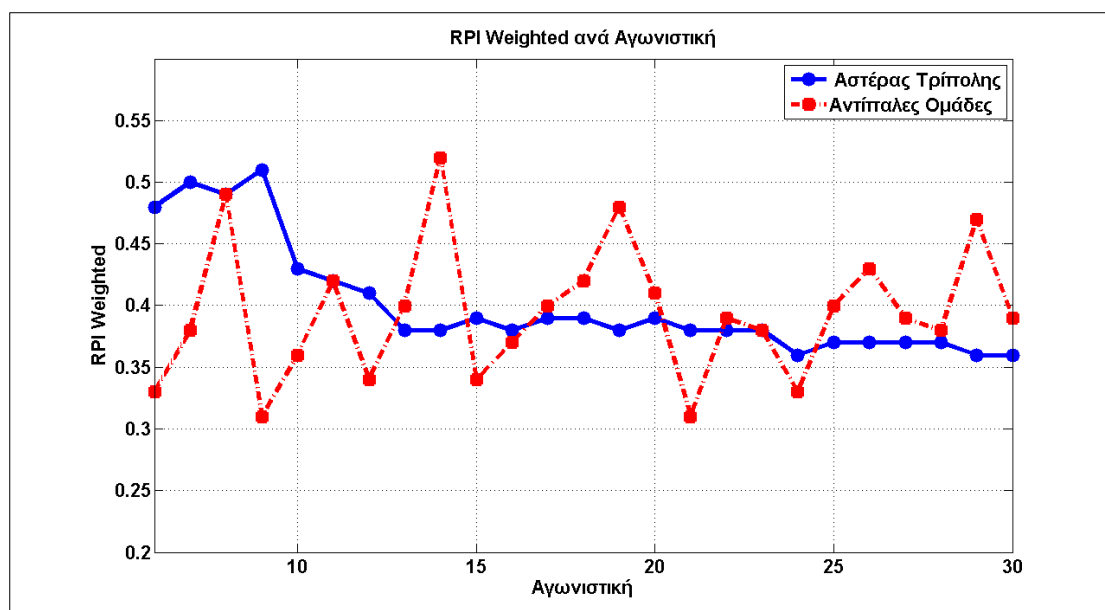
Εικόνα 8-34 : Οι τιμές του RPI Weighted του Ηρακλή και των αντιπάλων του ανά αγωνιστική



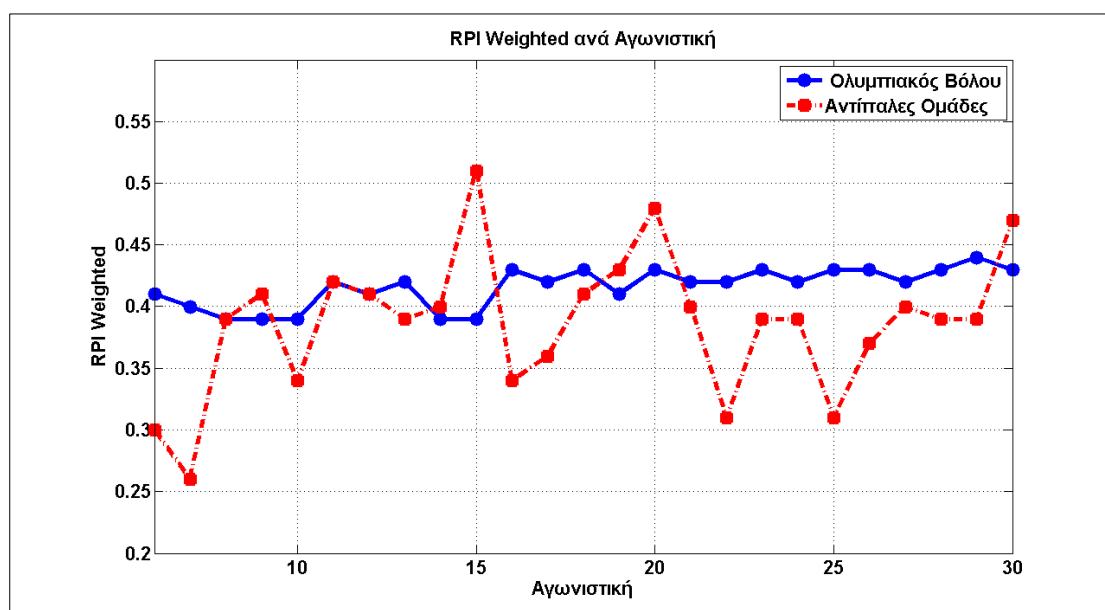
Εικόνα 8-35 : Οι τιμές του RPI Weighted της Κέρκυρας και των αντιπάλων της ανά αγωνιστική



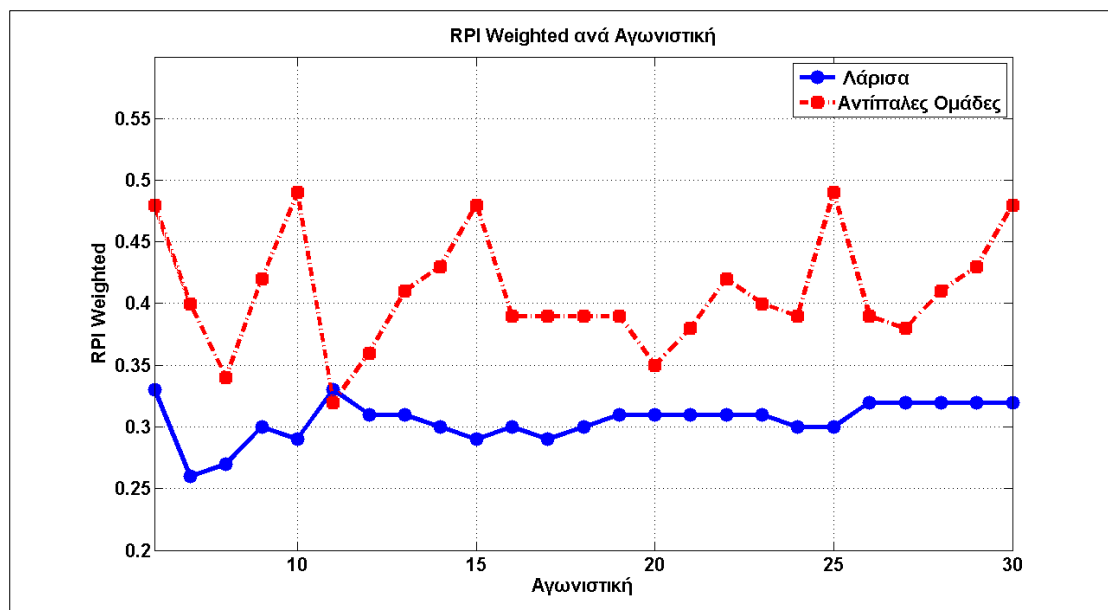
Εικόνα 8-36 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική



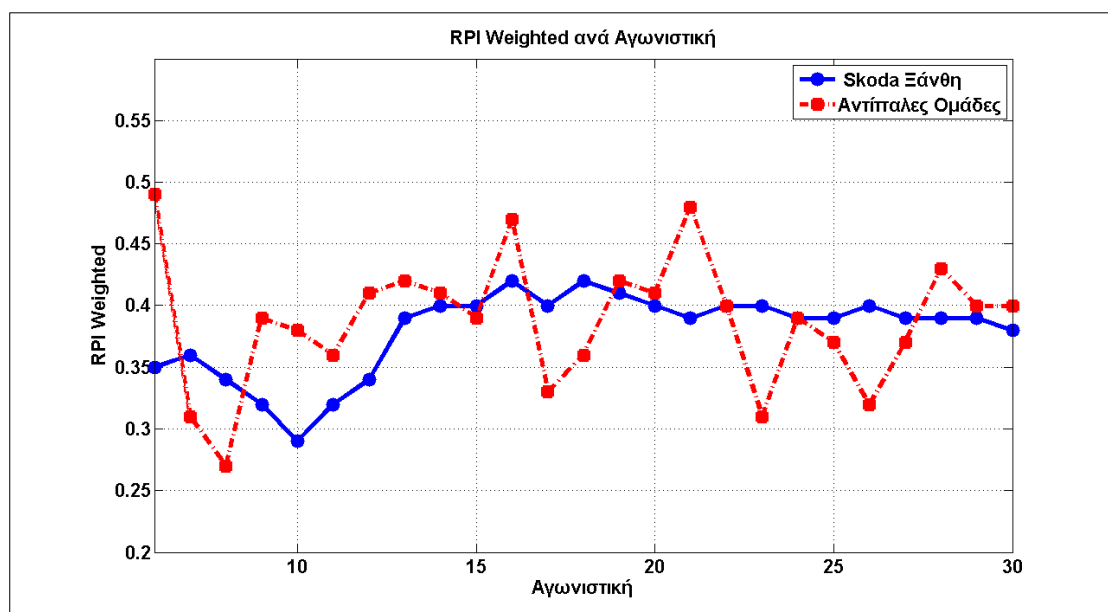
Εικόνα 8-37 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική



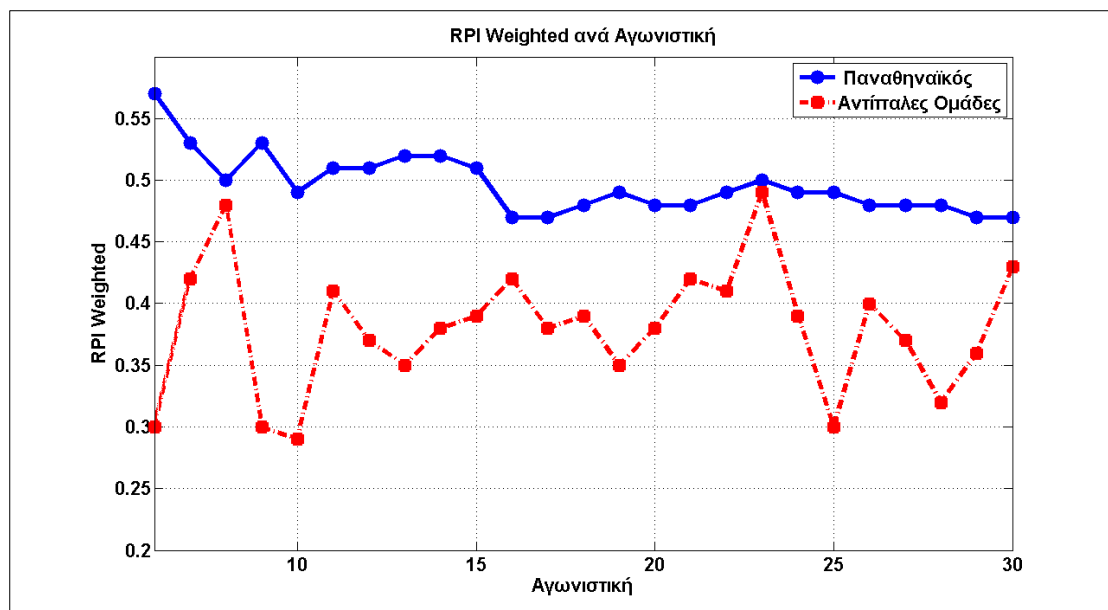
Εικόνα 8-38 : Οι τιμές του RPI Weighted του Ολυμπιακού Βόλου και των αντιπάλων του ανά αγωνιστική



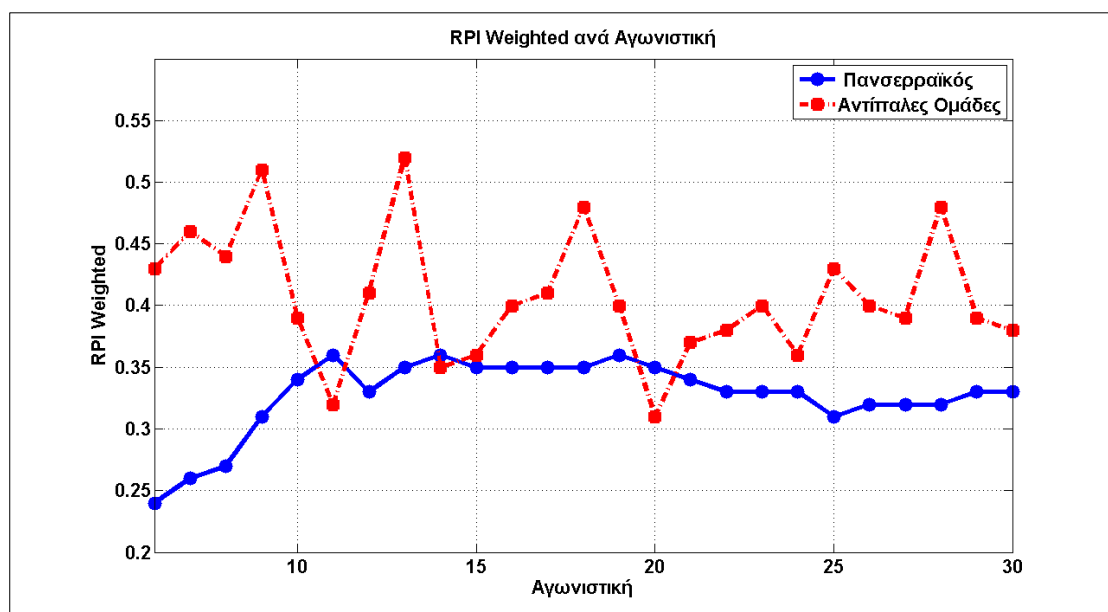
Εικόνα 8-39 : Οι τιμές του RPI Weighted της Λάρισας και των αντιπάλων της ανά αγωνιστική



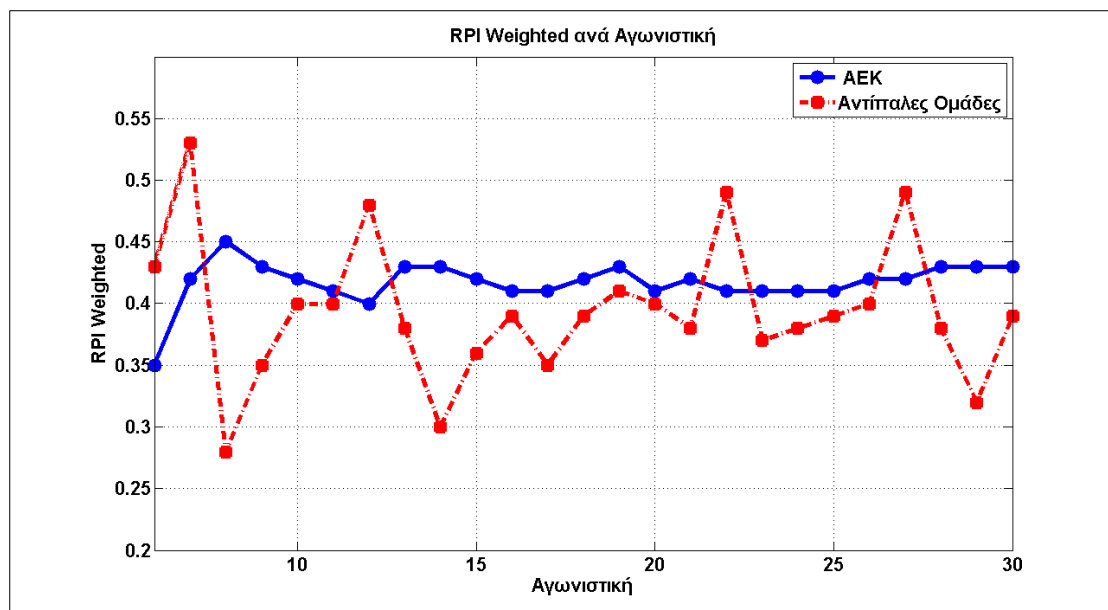
Εικόνα 8-40 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική



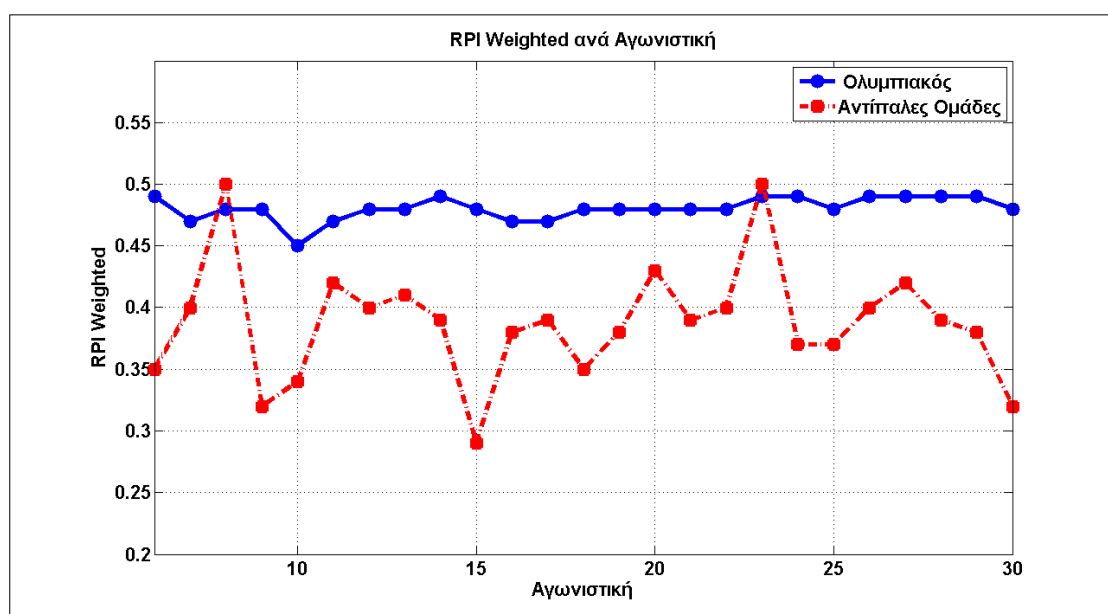
Εικόνα 8-41 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική



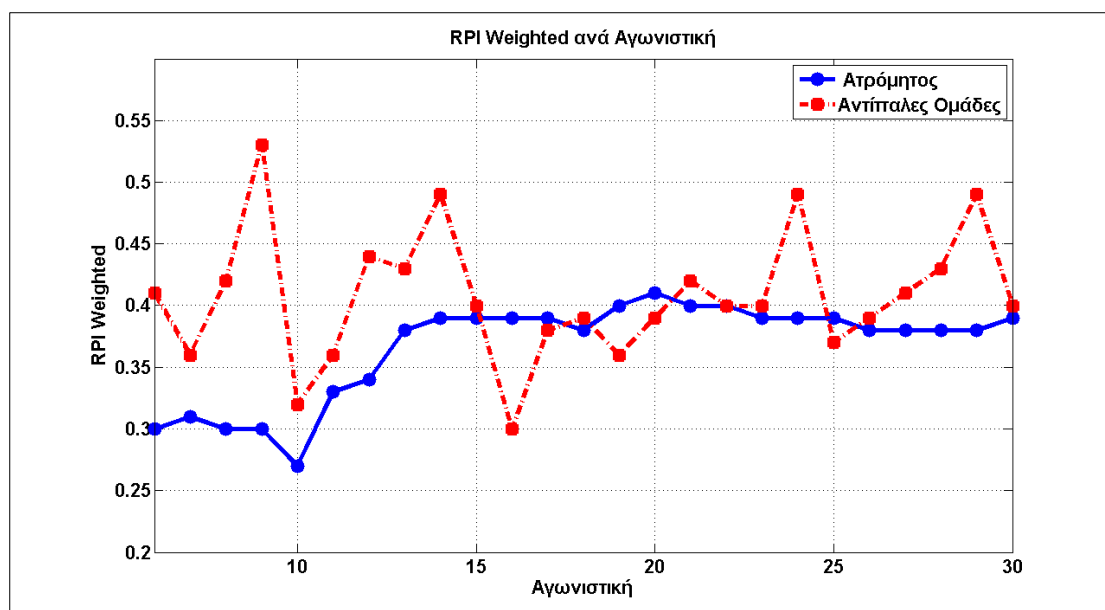
Εικόνα 8-42 : Οι τιμές του RPI Weighted του Πανσερραϊκού και των αντιπάλων του ανά αγωνιστική



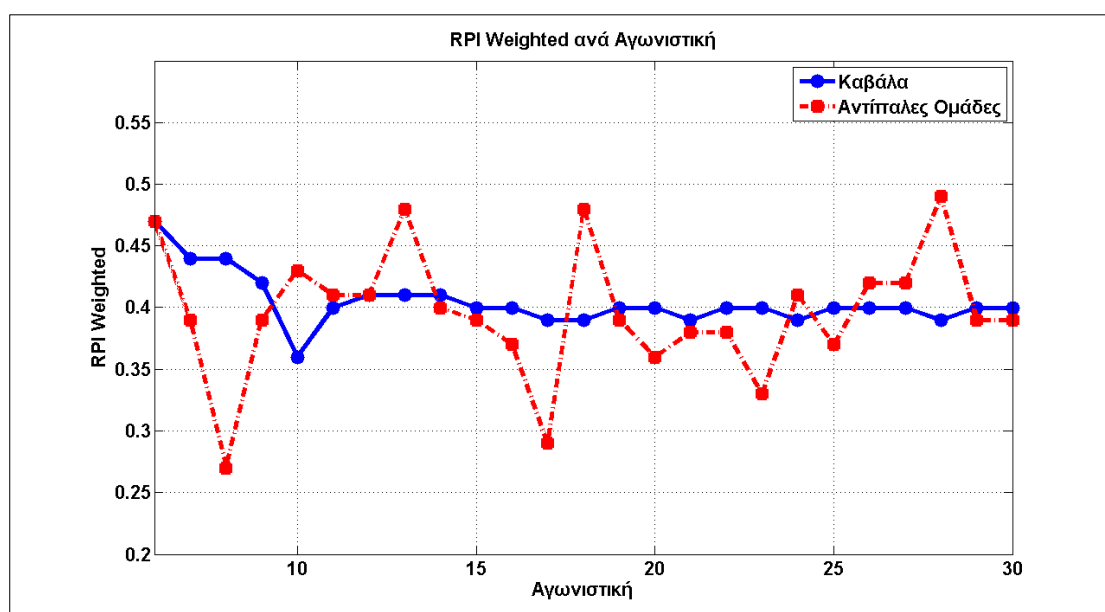
Εικόνα 8-43 : Οι τιμές του RPI Weighted της AEK και των αντιπάλων της ανά αγωνιστική



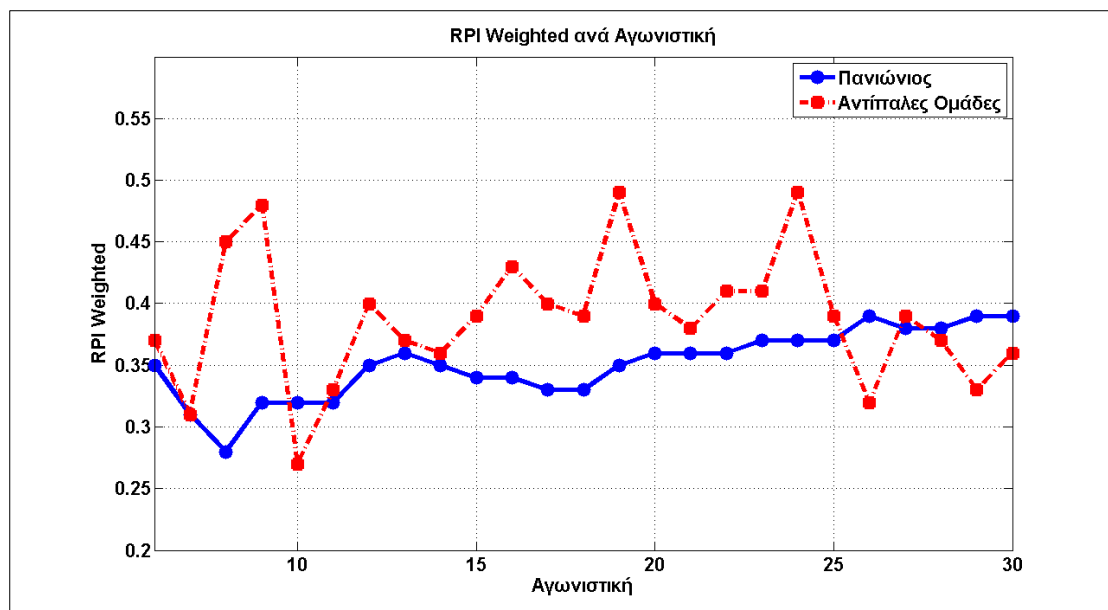
Εικόνα 8-44 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική



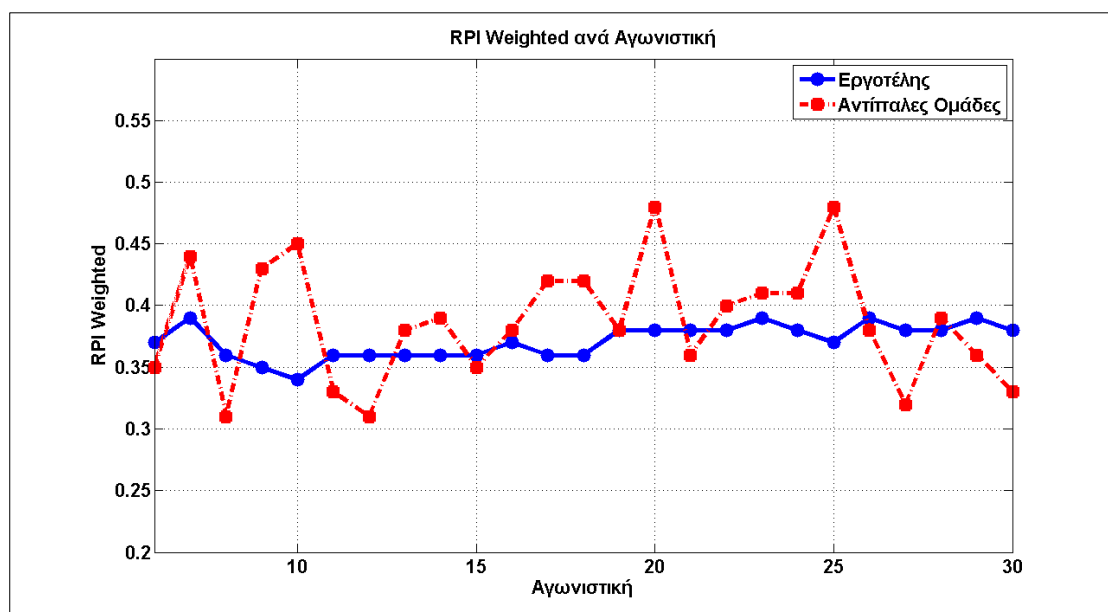
Εικόνα 8-45 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική



Εικόνα 8-46 : Οι τιμές του RPI Weighted της Καβάλας και των αντιπάλων της ανά αγωνιστική

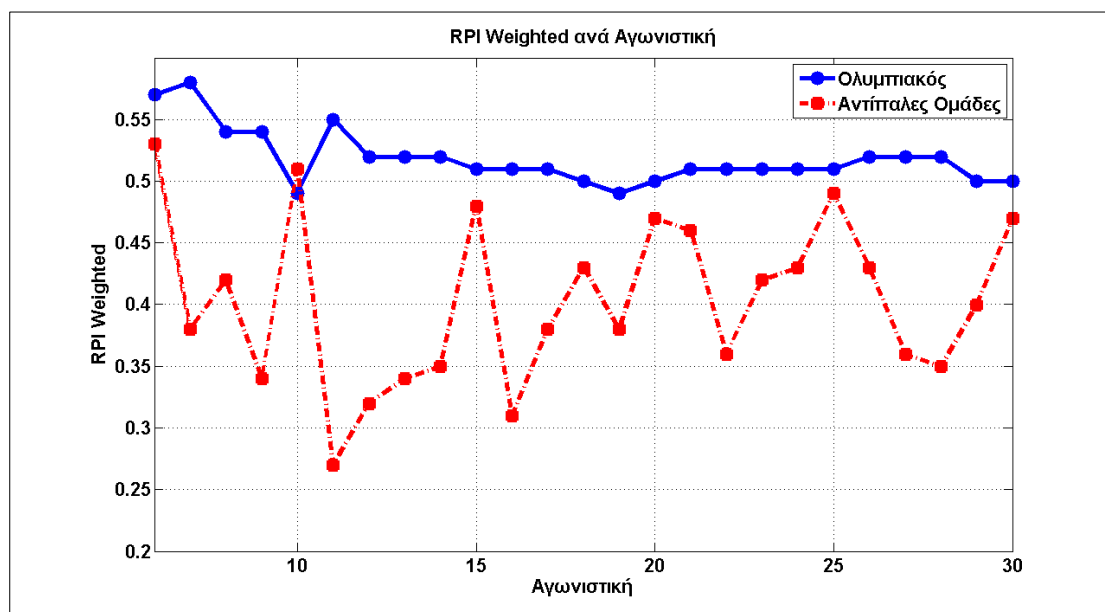


Εικόνα 8-47 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική

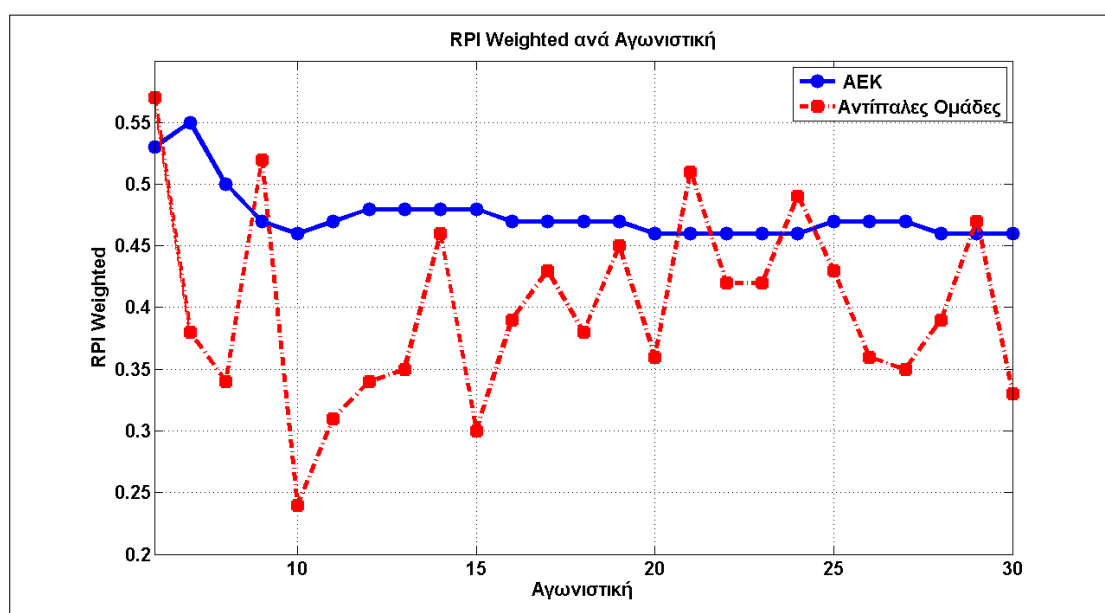


Εικόνα 8-48 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική

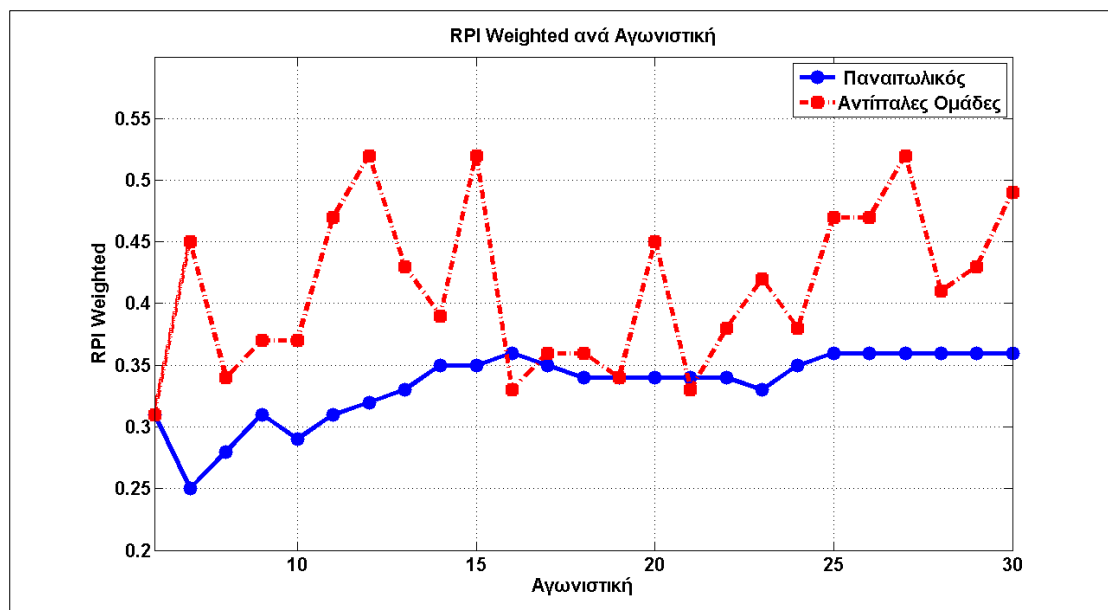
- Για τη σεζόν 2011-2012 :



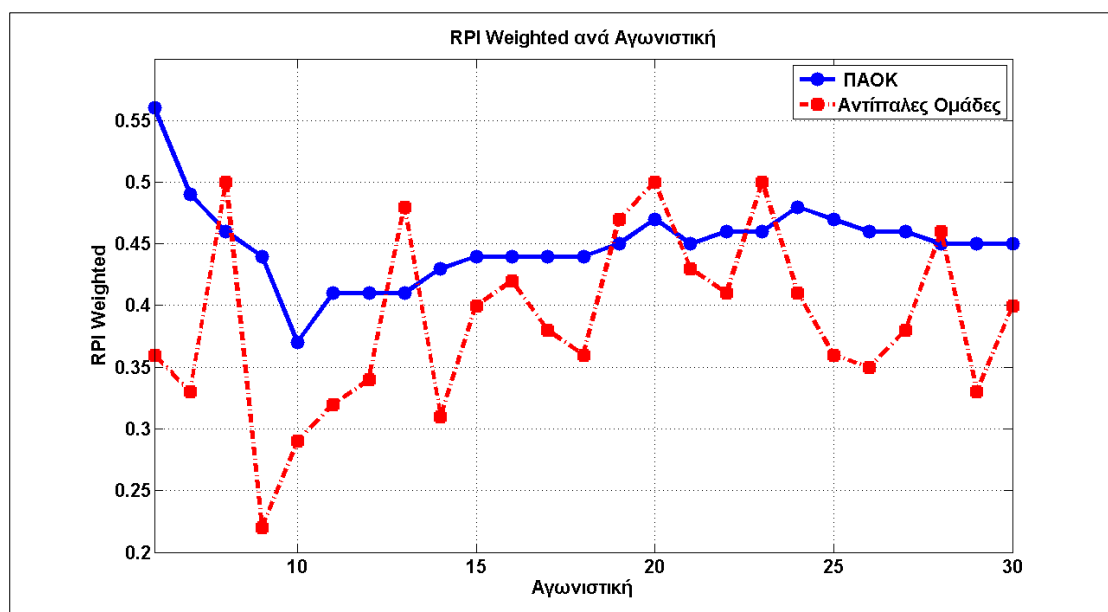
Εικόνα 8-49 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική



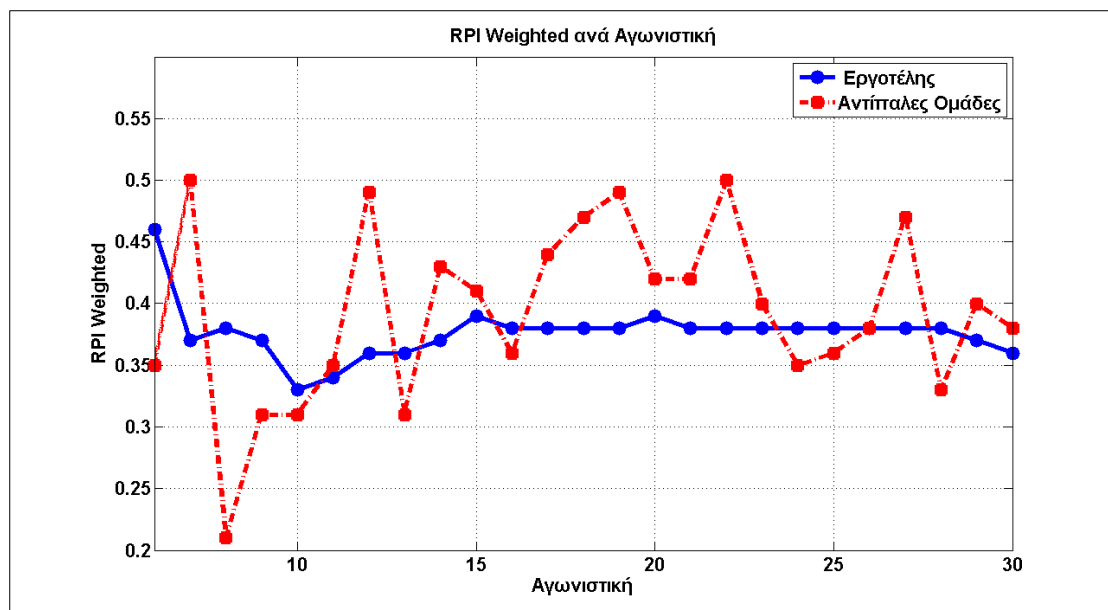
Εικόνα 8-50 : Οι τιμές του RPI Weighted της AEK και των αντιπάλων της ανά αγωνιστική



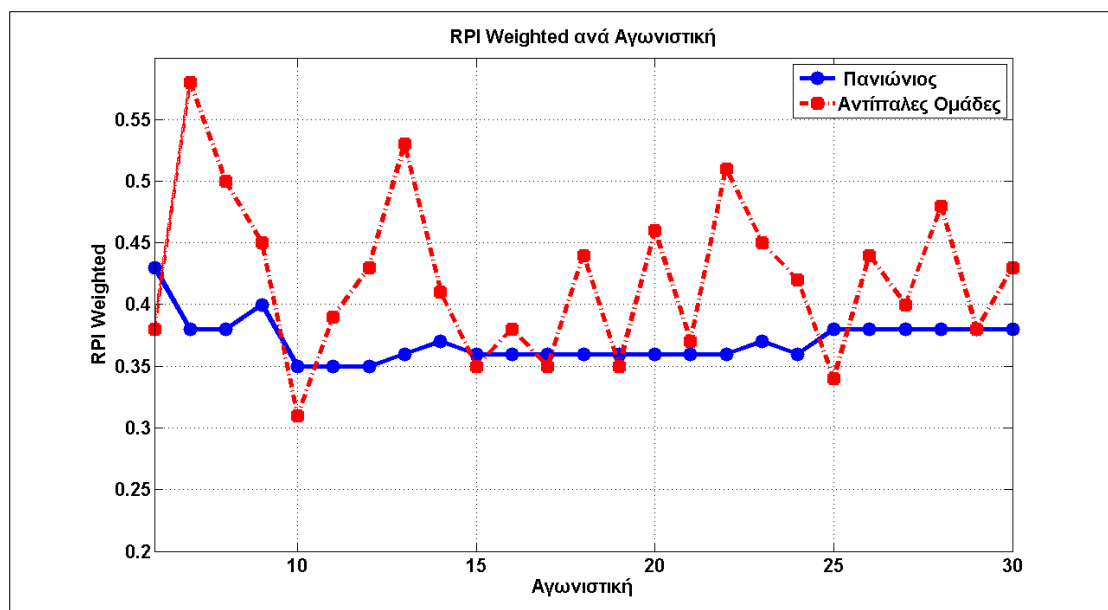
Εικόνα 8-51 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική



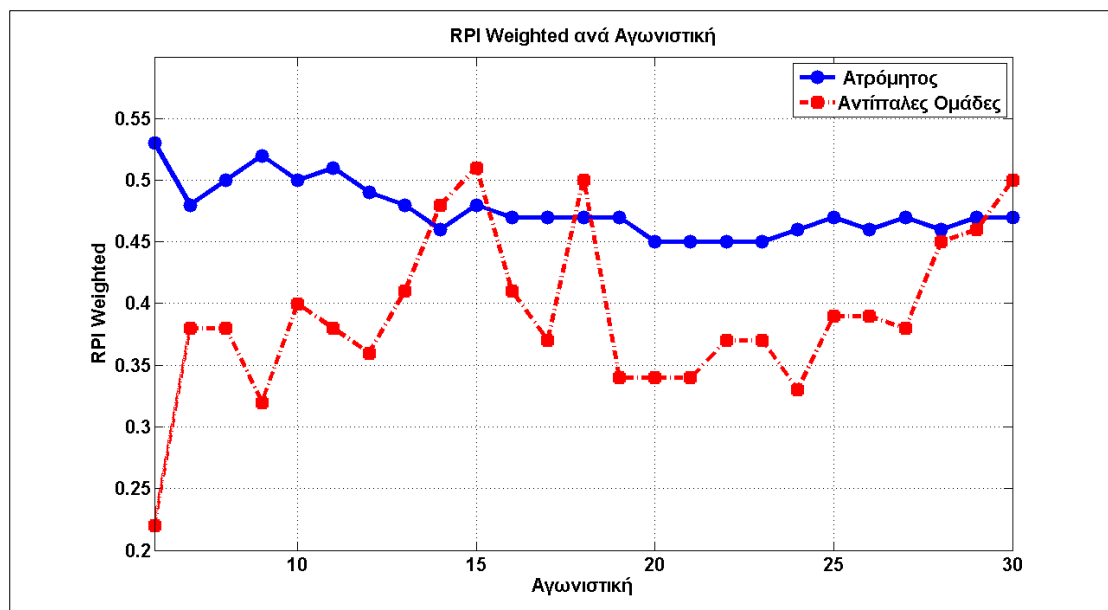
Εικόνα 8-52 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική



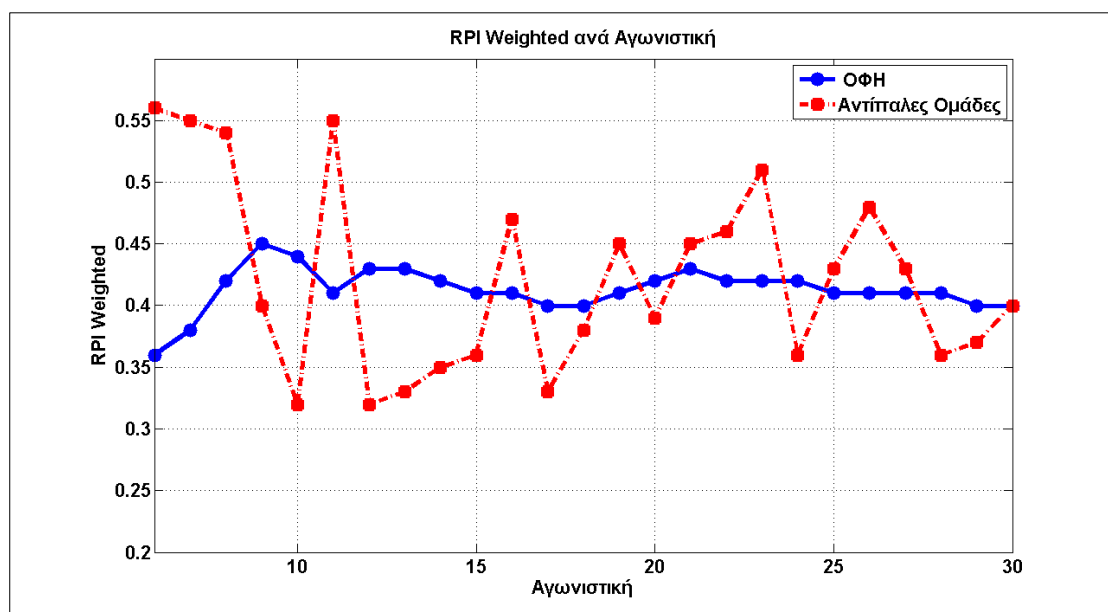
Εικόνα 8-53 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική



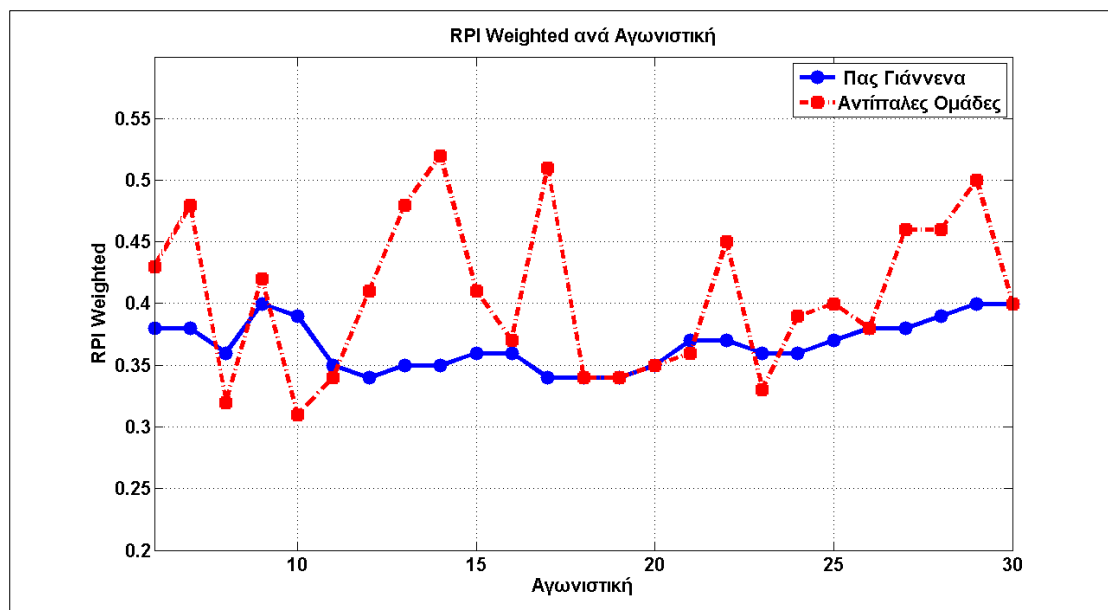
Εικόνα 8-54 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική



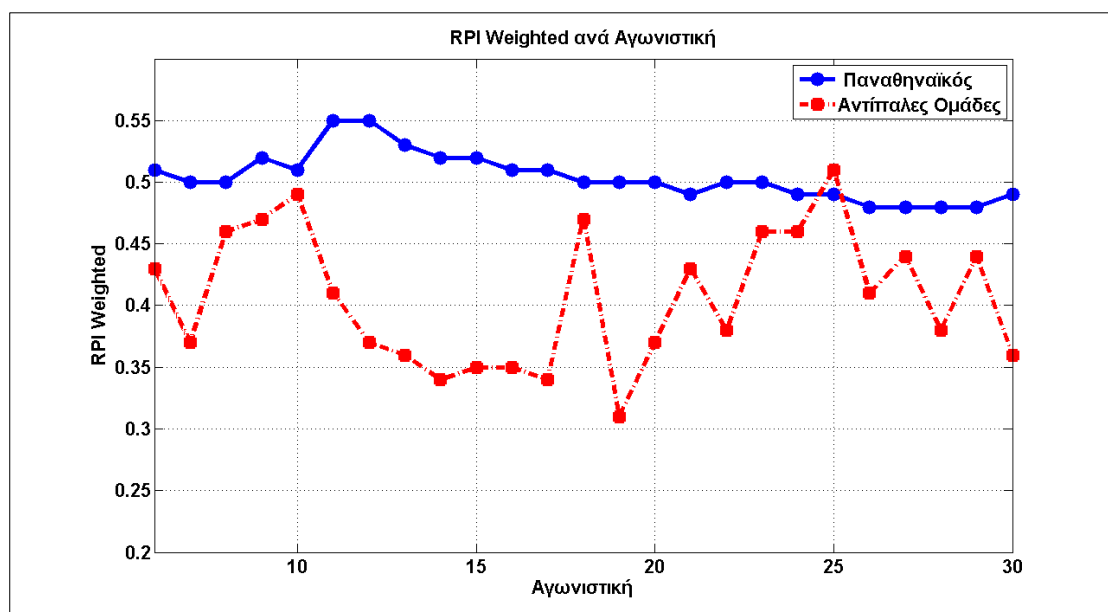
Εικόνα 8-55 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική



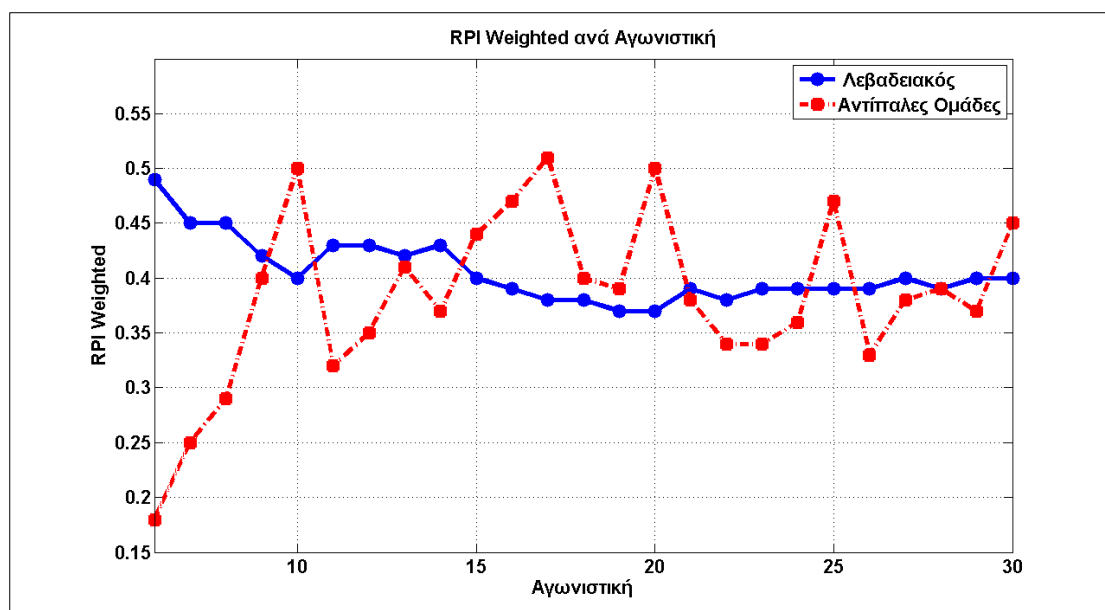
Εικόνα 8-56 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική



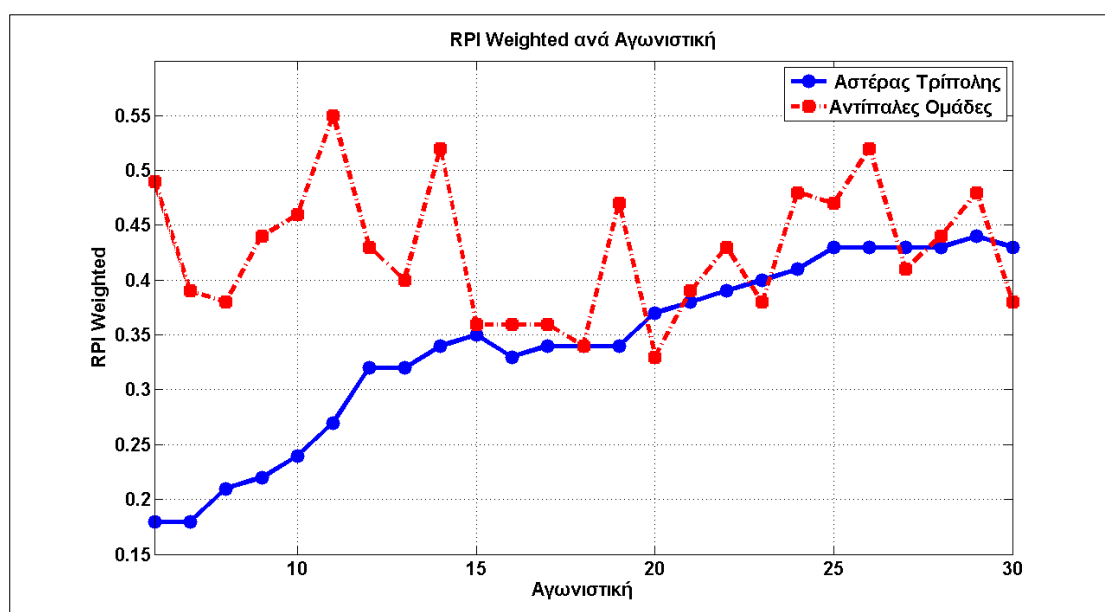
Εικόνα 8-57 : Οι τιμές του RPI Weighted του Πας Γιάννενα και των αντιπάλων του ανά αγωνιστική



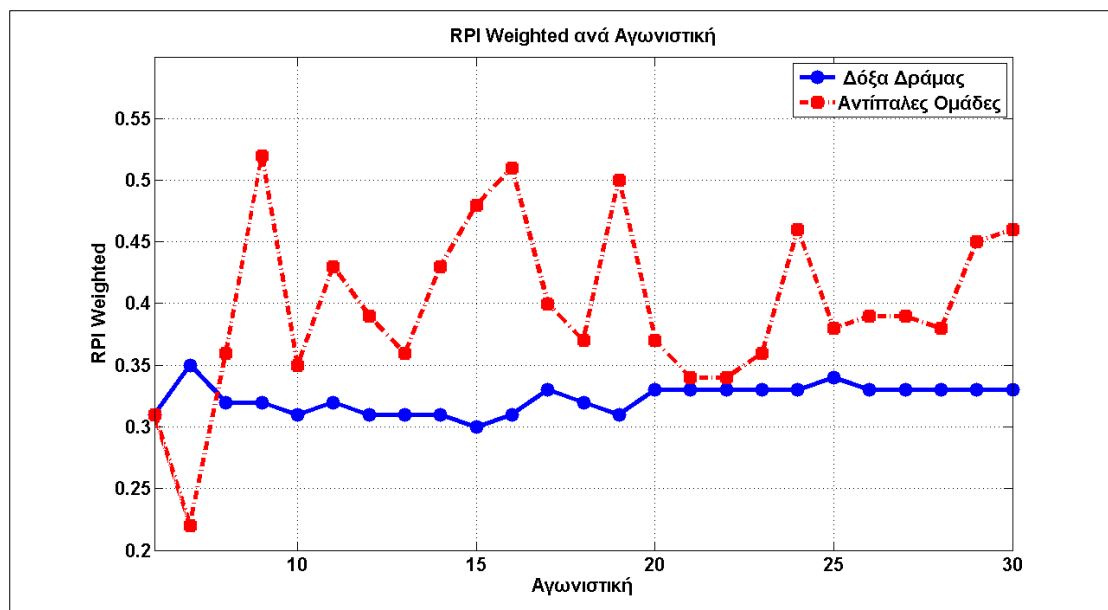
Εικόνα 8-58 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική



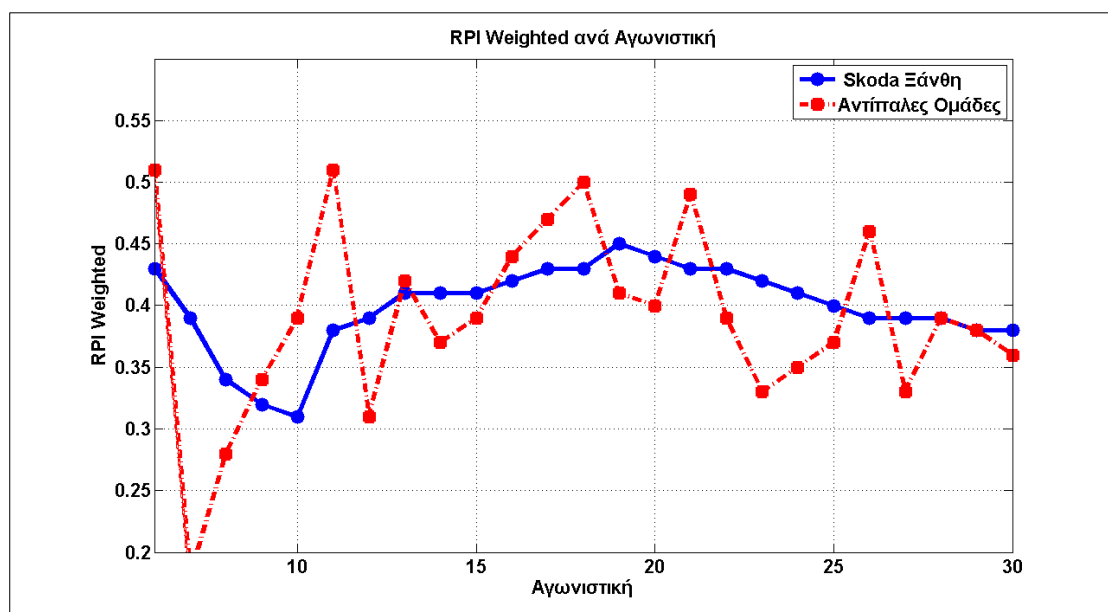
Εικόνα 8-59 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική



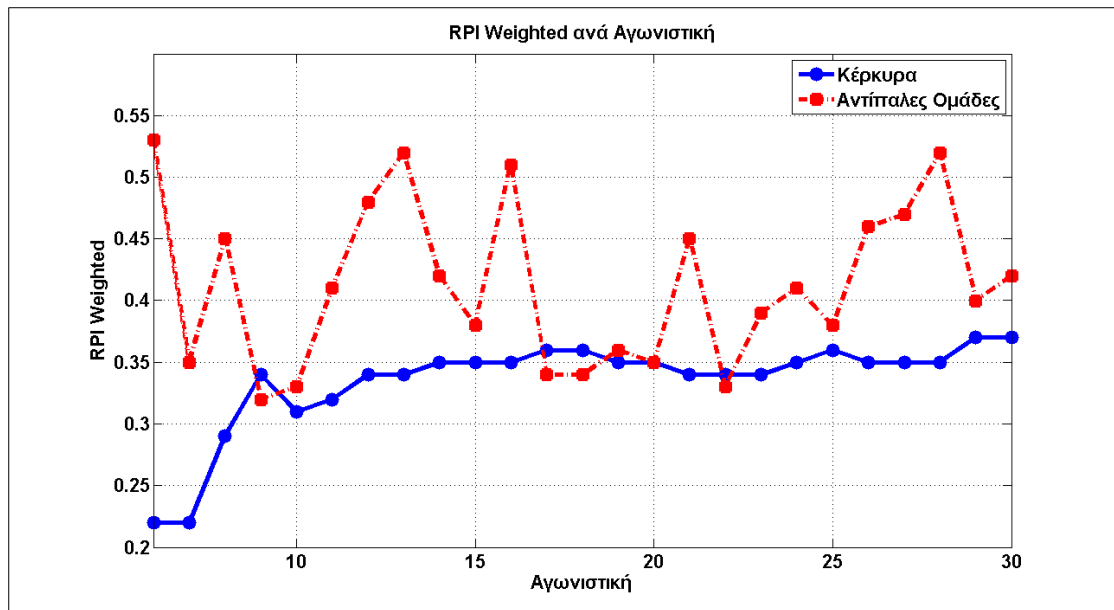
Εικόνα 8-60 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική



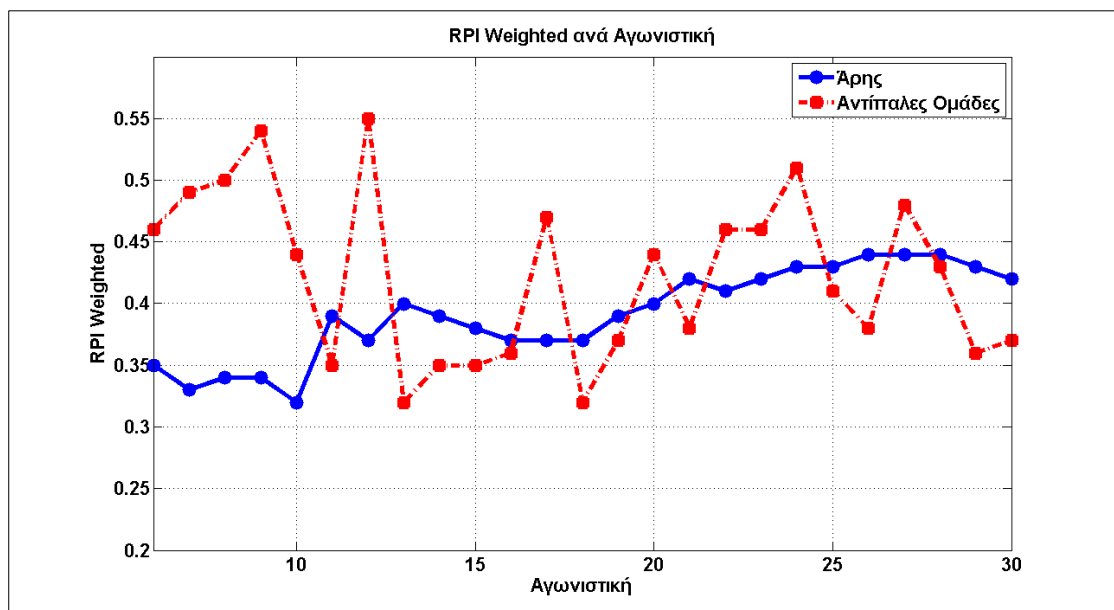
Εικόνα 8-61 : Οι τιμές του RPI Weighted της Δόξας Δράμας και των αντιπάλων της ανά αγωνιστική



Εικόνα 8-62 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική

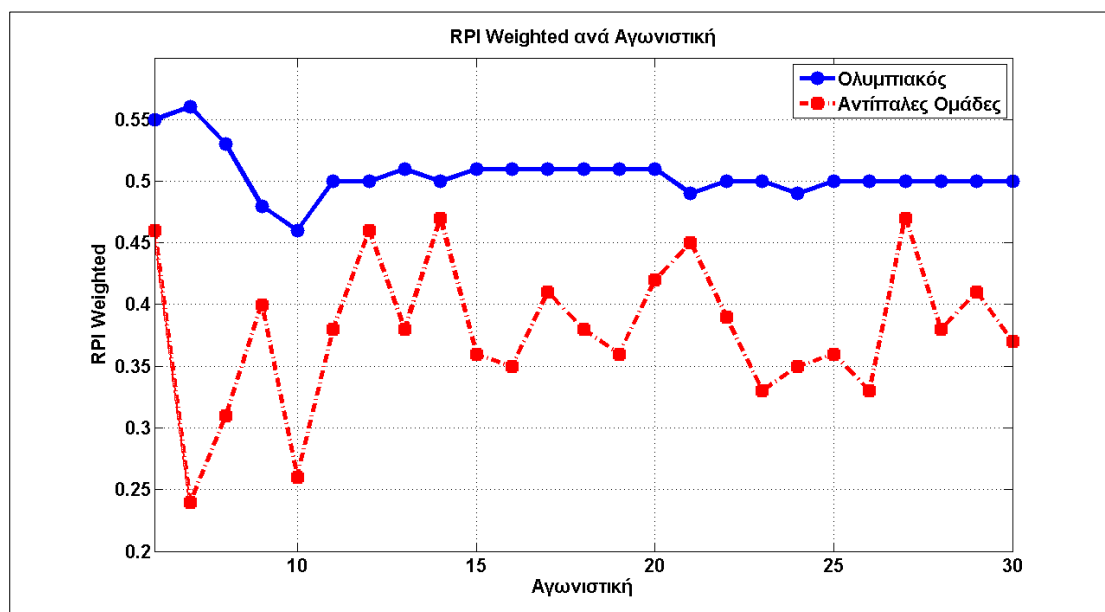


Εικόνα 8-63 : Οι τιμές του RPI Weighted της Κέρκυρας και των αντιπάλων της ανά αγωνιστική

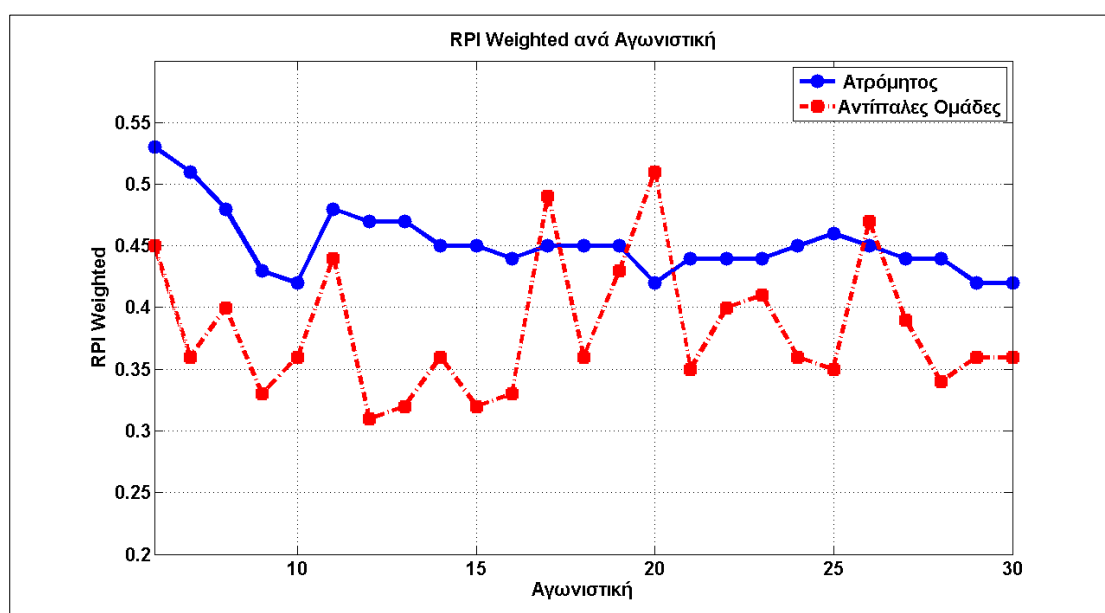


Εικόνα 8-64 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική

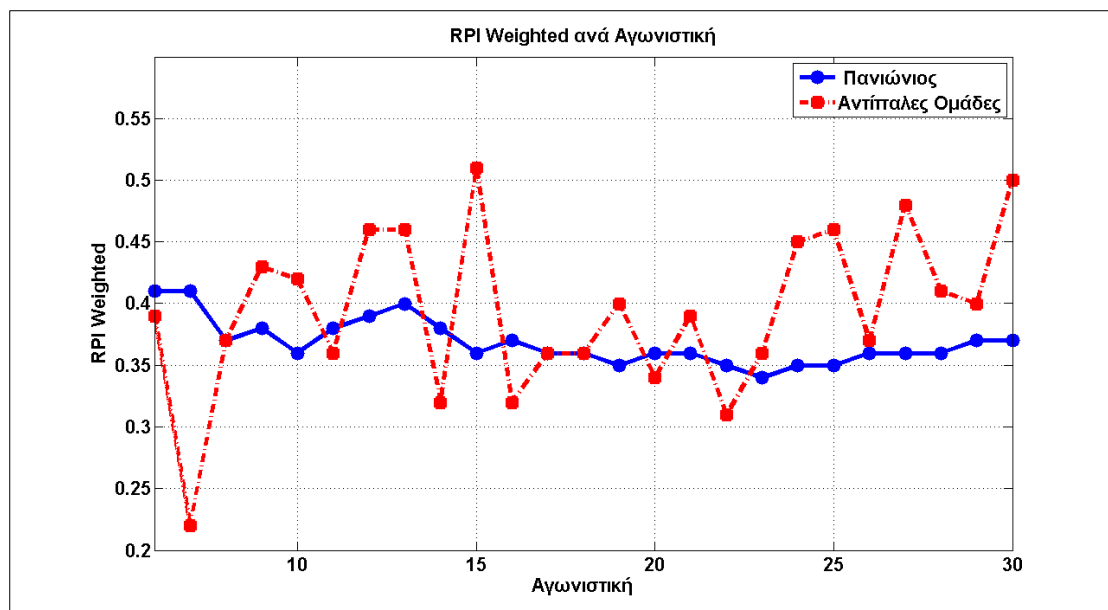
- Για τη σεζόν 2012-2013 :



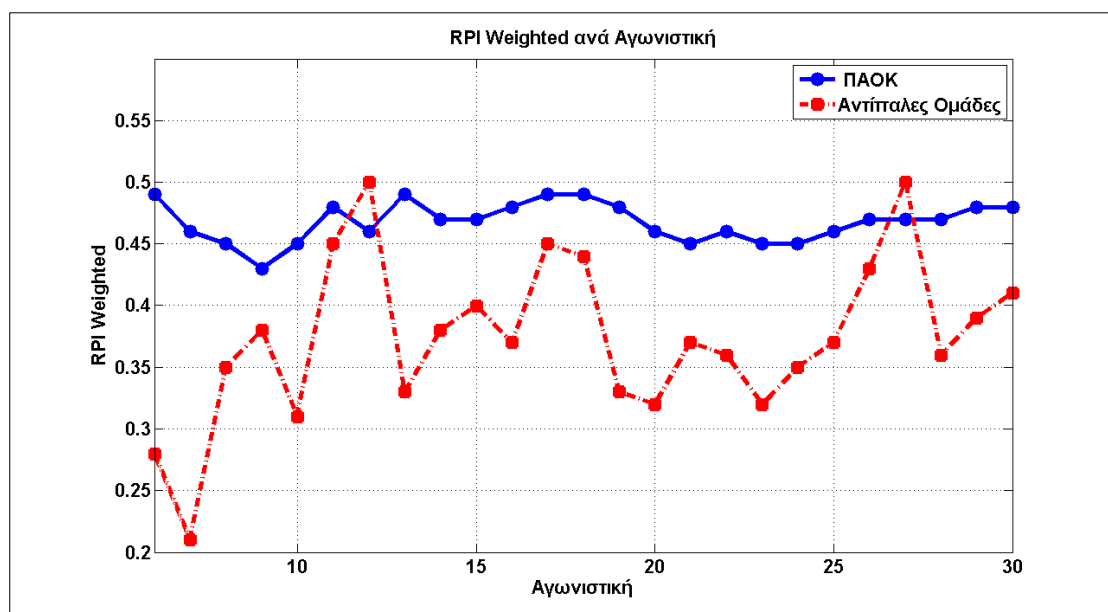
Εικόνα 8-65 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική



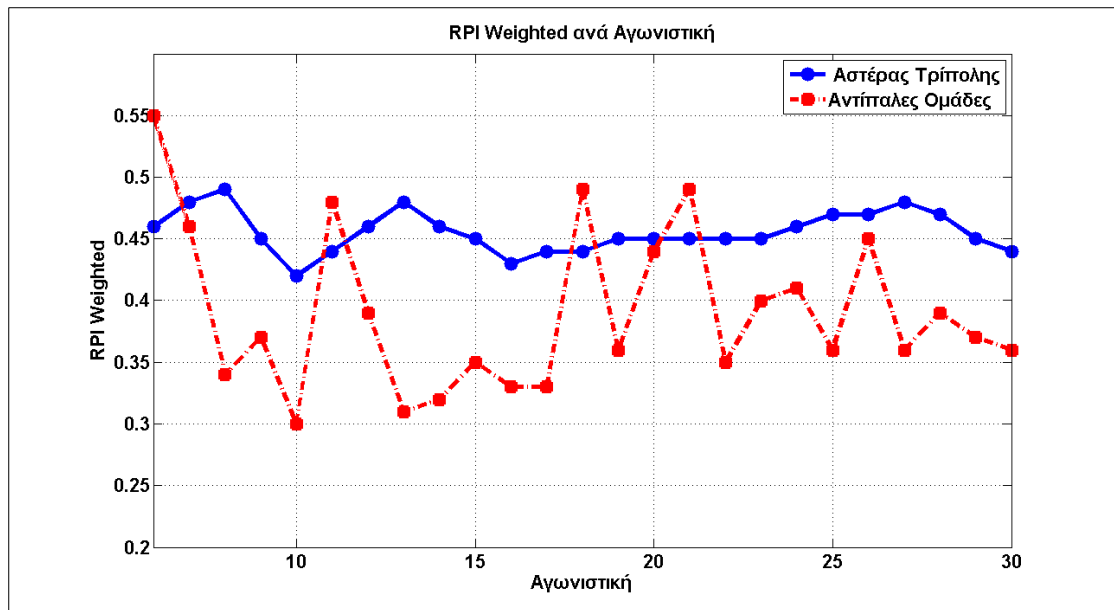
Εικόνα 8-66 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική



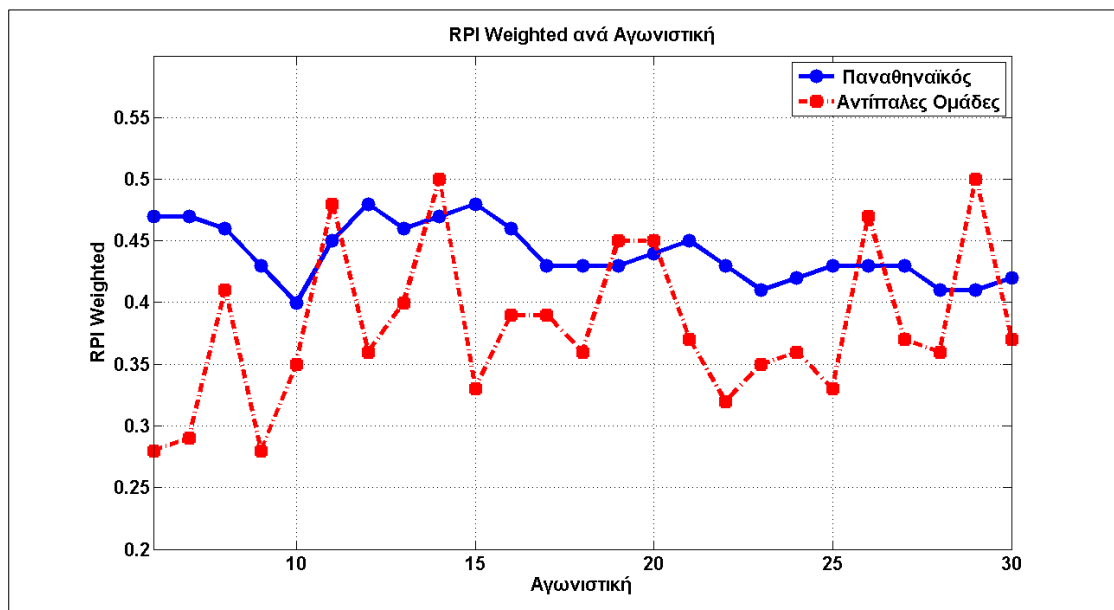
Εικόνα 8-67 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική



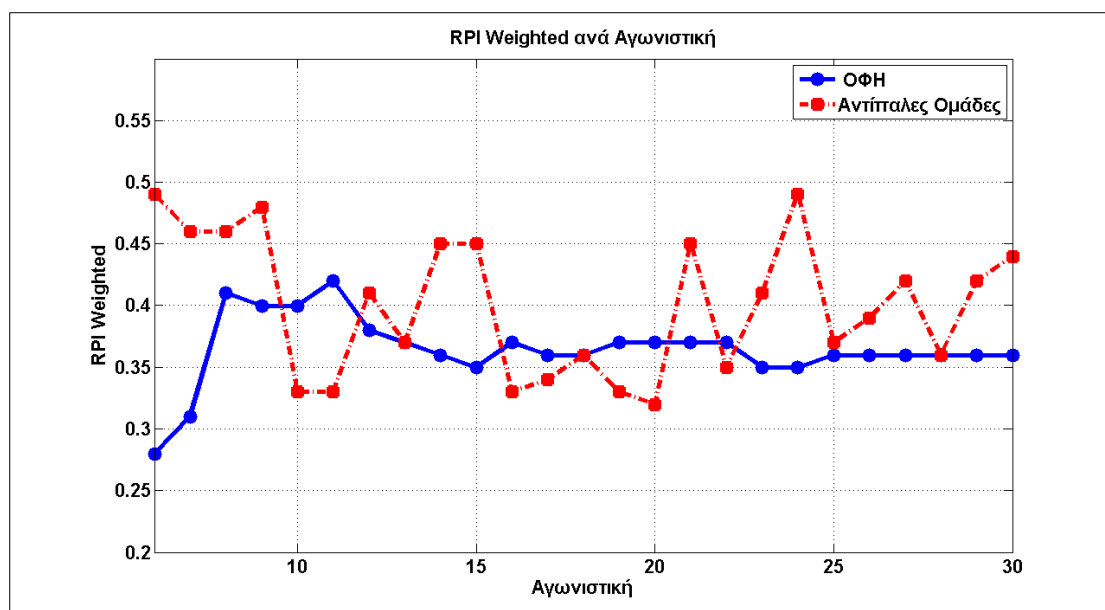
Εικόνα 8-68 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική



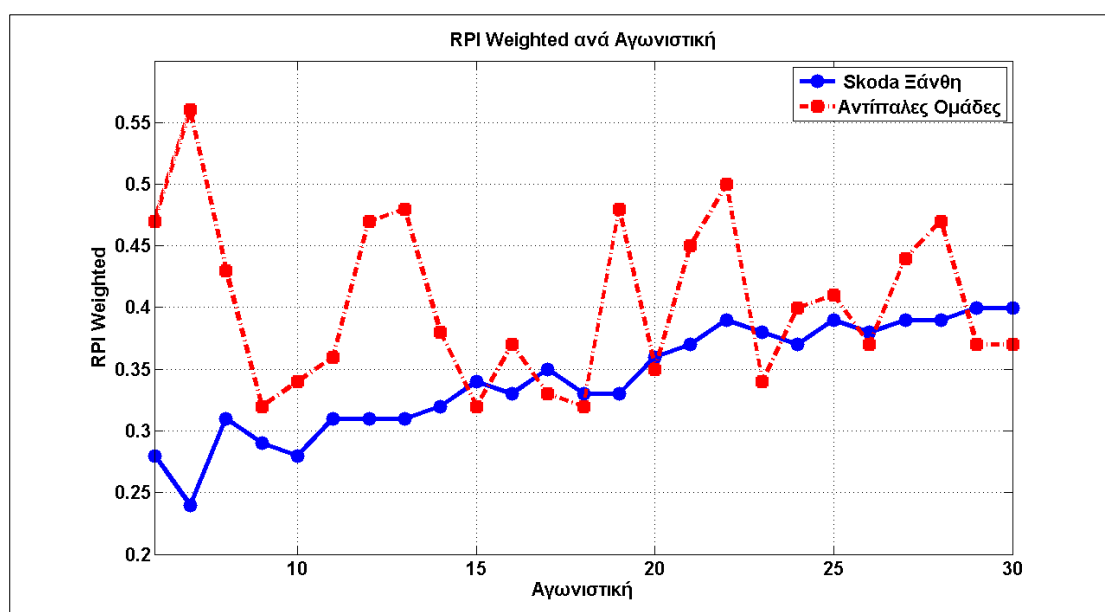
Εικόνα 8-69 : Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική



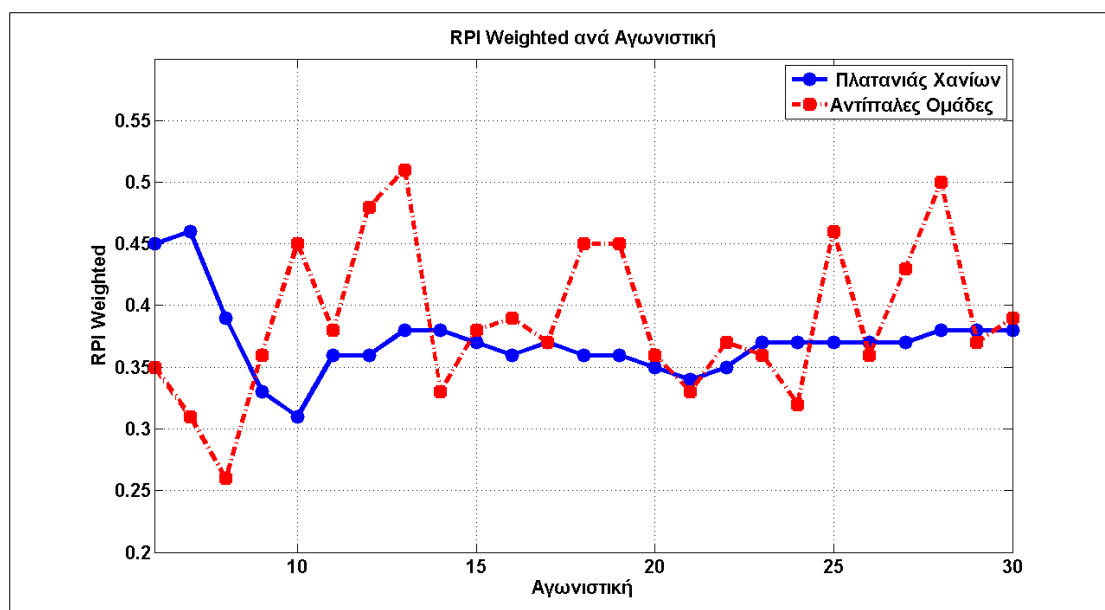
Εικόνα 8-70 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική



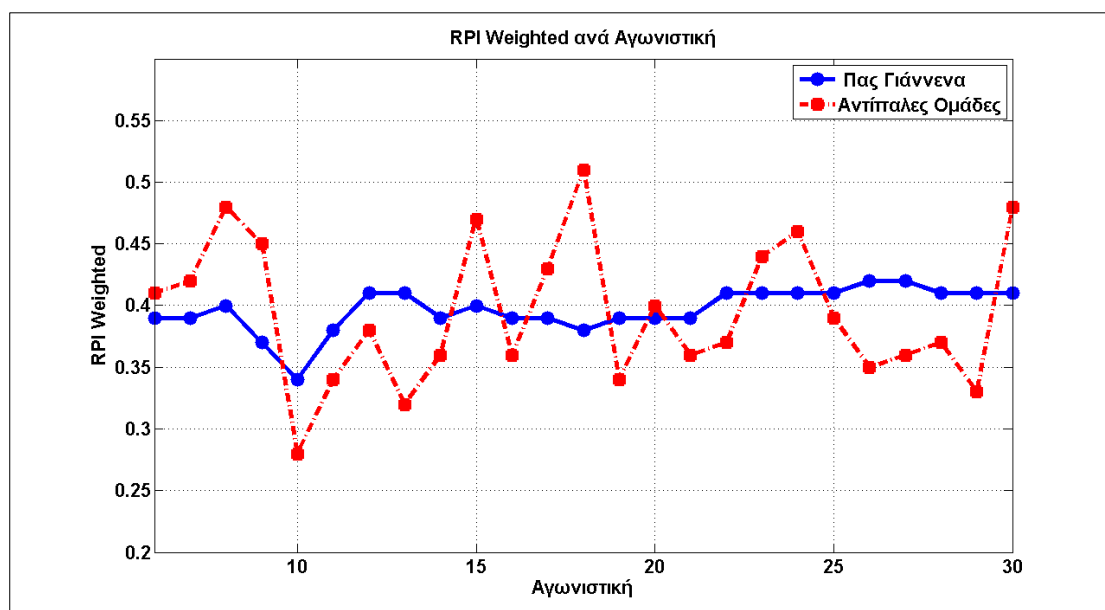
Εικόνα 8-71 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική



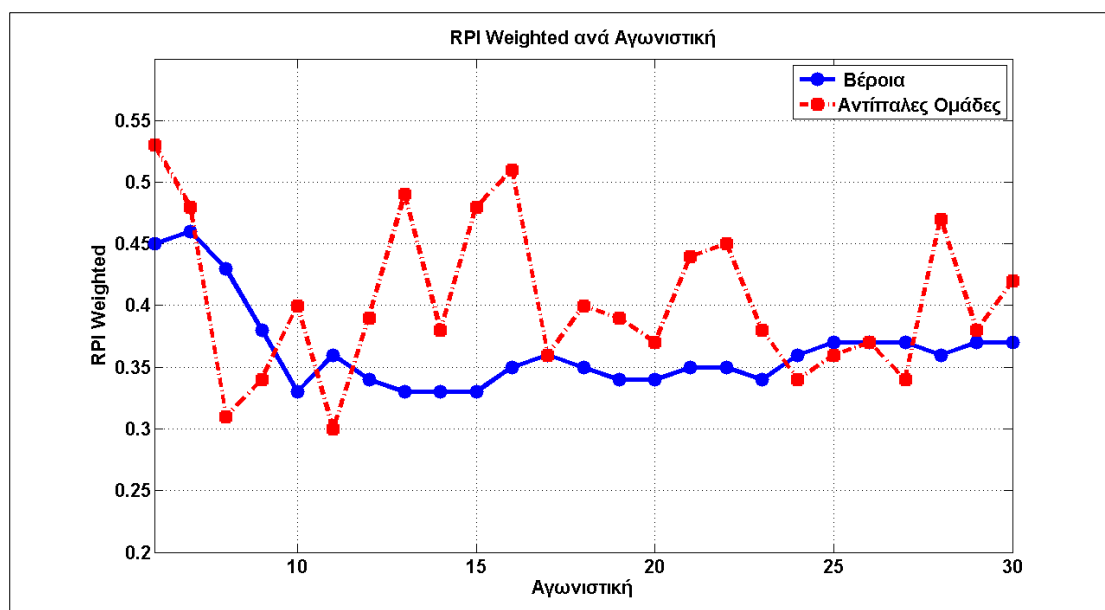
Εικόνα 8-72 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική



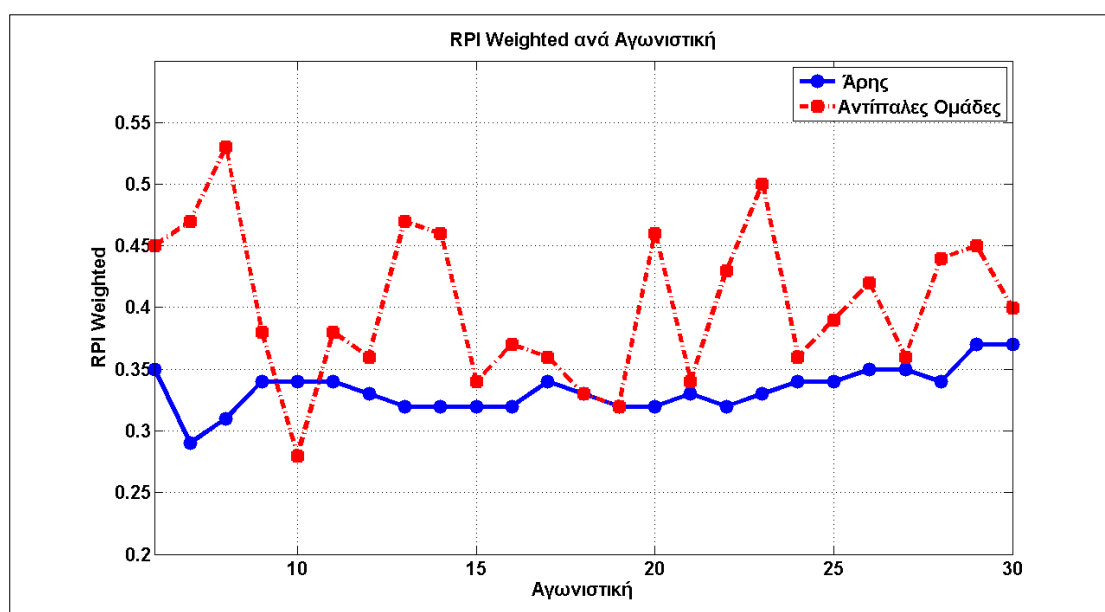
Εικόνα 8-73 : Οι τιμές του RPI Weighted του Πλατανιά Χανίων και των αντιπάλων του ανά αγωνιστική



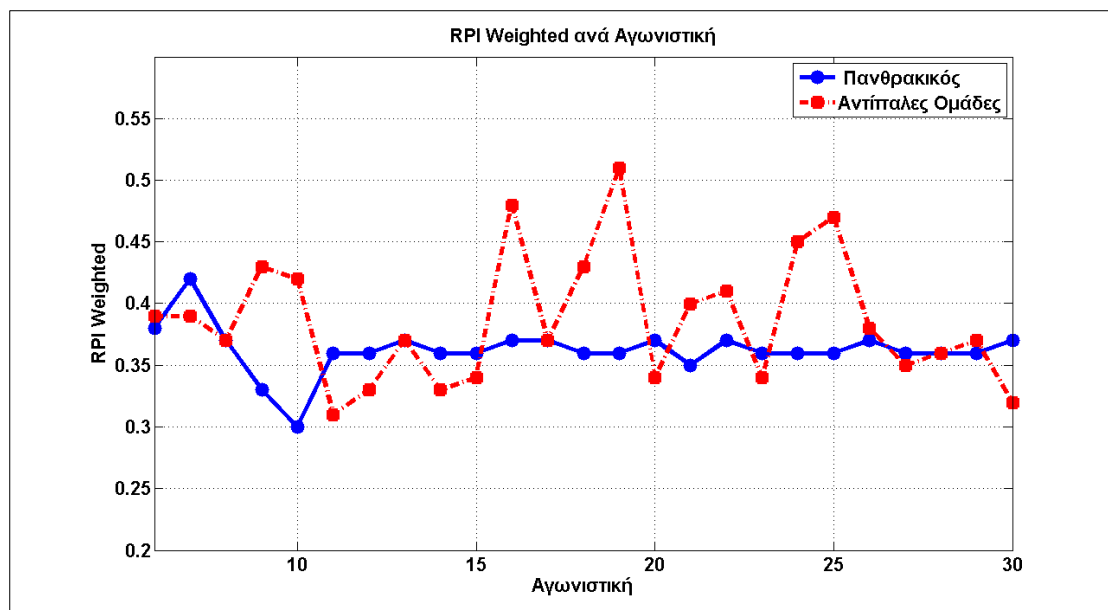
Εικόνα 8-74 : Οι τιμές του RPI Weighted του Πας Γιάννενα και των αντιπάλων του ανά αγωνιστική



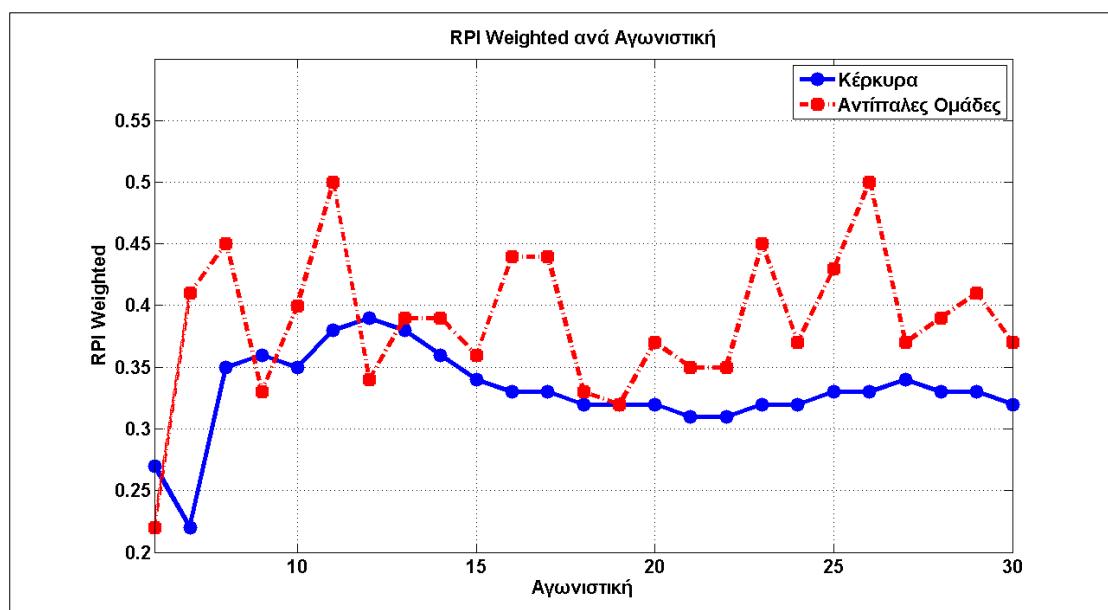
Εικόνα 8-75 : Οι τιμές του RPI Weighted της Βέροιας και των αντιπάλων της ανά αγωνιστική



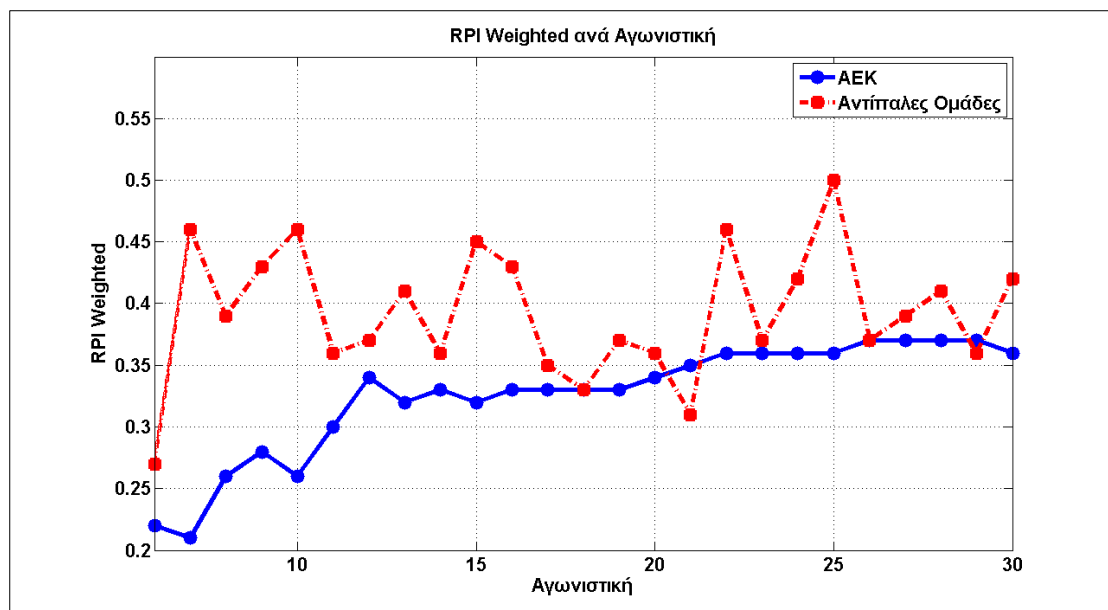
Εικόνα 8-76 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική



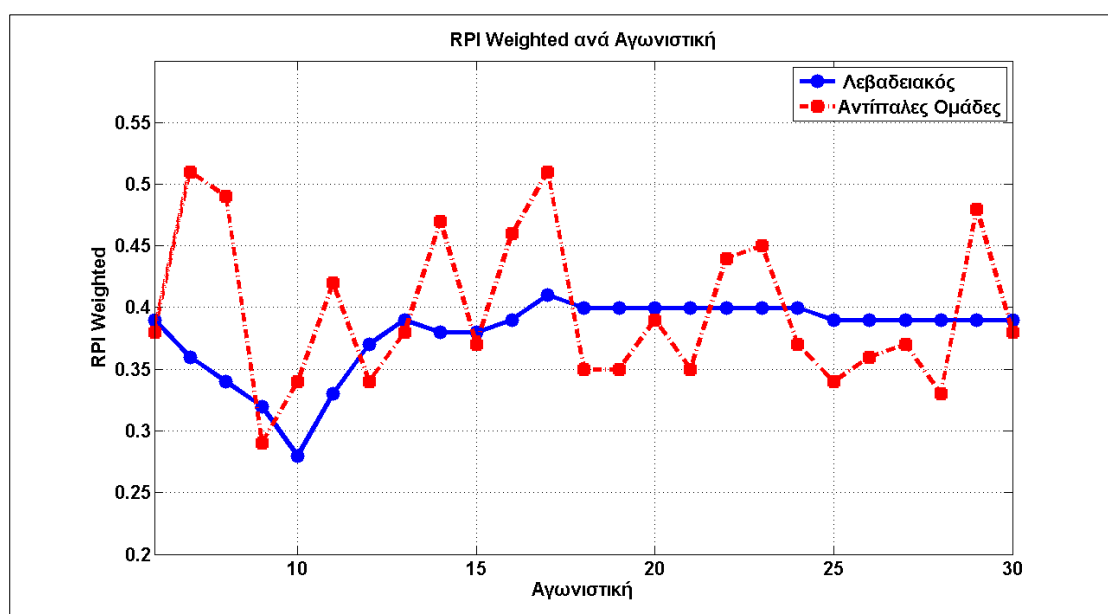
Εικόνα 8-77 : Οι τιμές του RPI Weighted του Πανθρακικού και των αντιπάλων του ανά αγωνιστική



Εικόνα 8-78 : Οι τιμές του RPI Weighted της Κέρκυρας και των αντιπάλων της ανά αγωνιστική

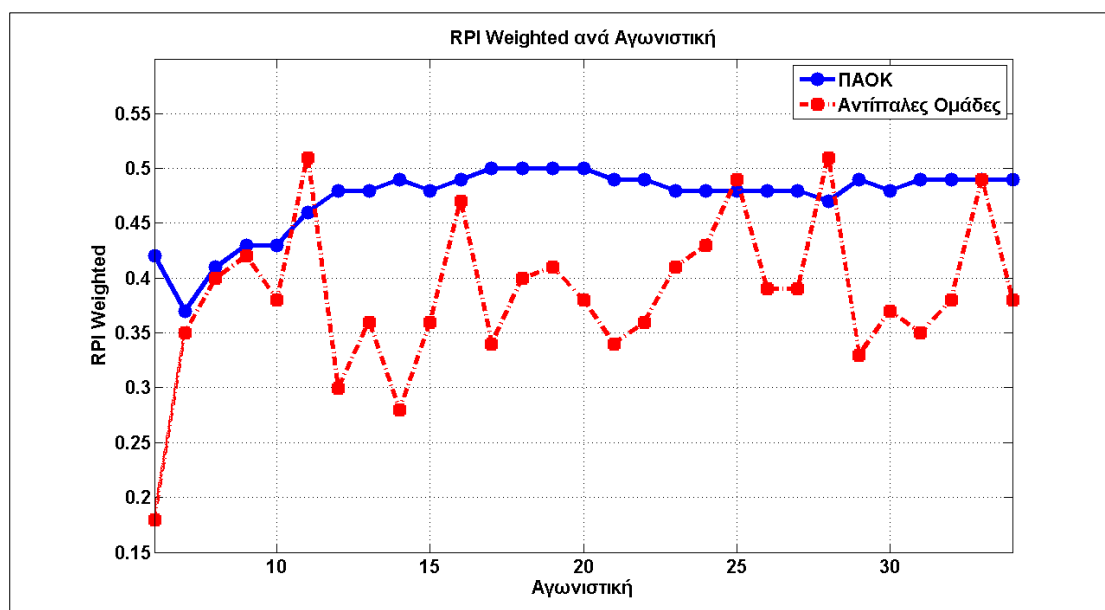


Εικόνα 8-79 : Οι τιμές του RPI Weighted της AEK και των αντιπάλων της ανά αγωνιστική

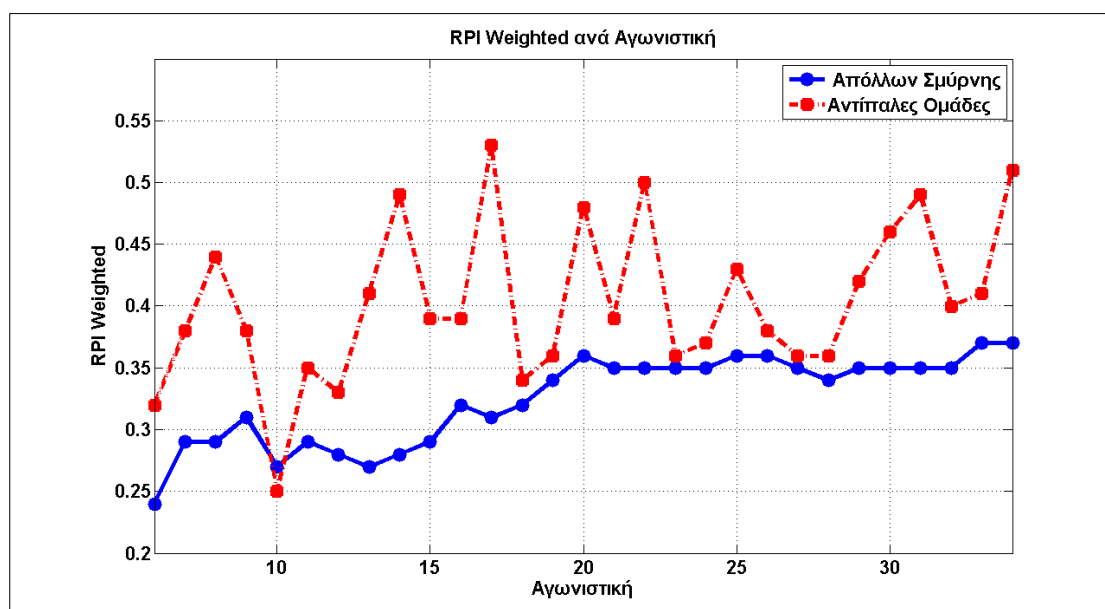


Εικόνα 8-80 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική

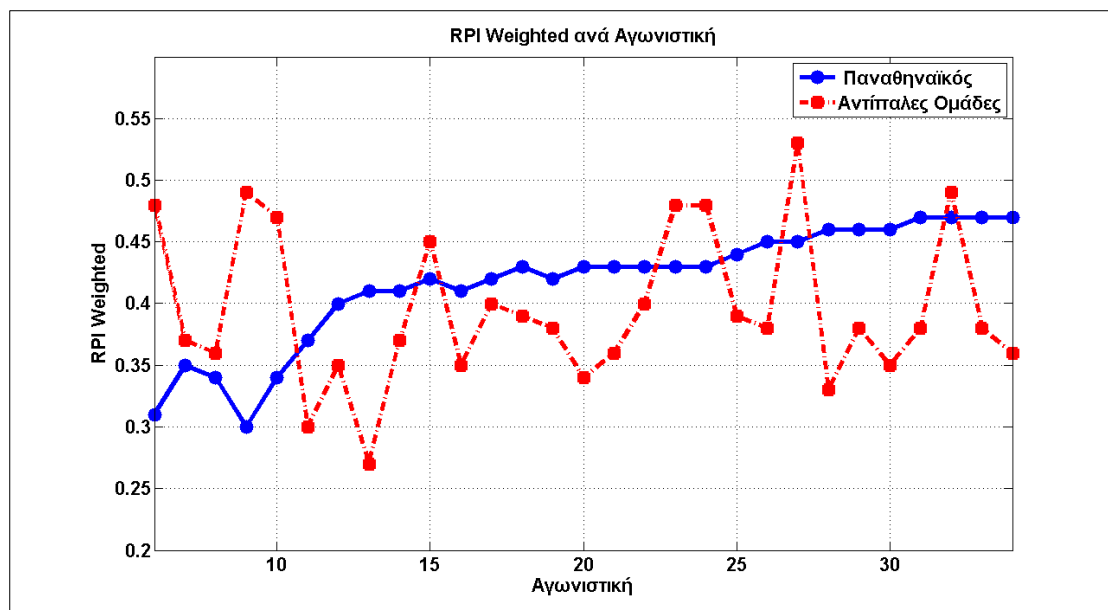
- Για τη σεζόν 2013-2014 :



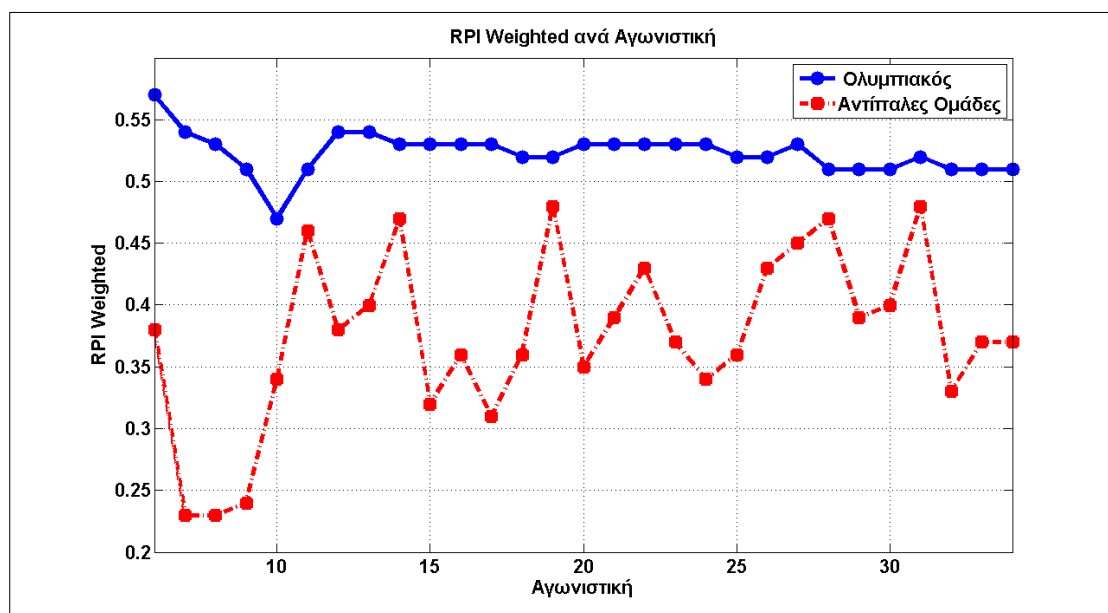
Εικόνα 8-81 : Οι τιμές του RPI Weighted του ΠΑΟΚ και των αντιπάλων του ανά αγωνιστική



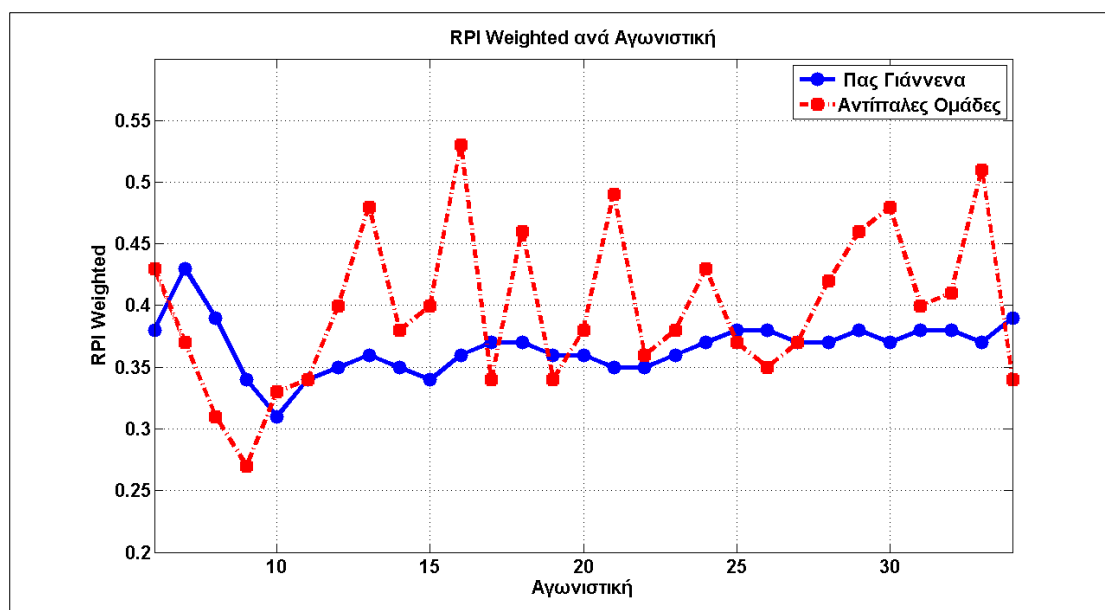
Εικόνα 8-82 : Οι τιμές του RPI Weighted του Απόλλωνα Σμύρνης και των αντιπάλων του ανά αγωνιστική



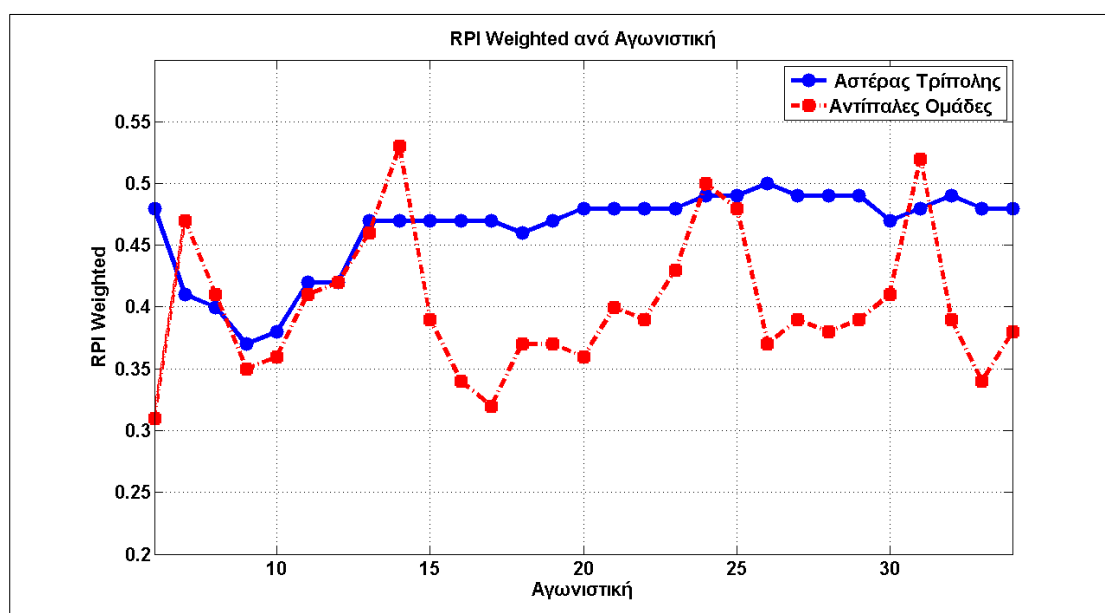
Εικόνα 8-83 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική



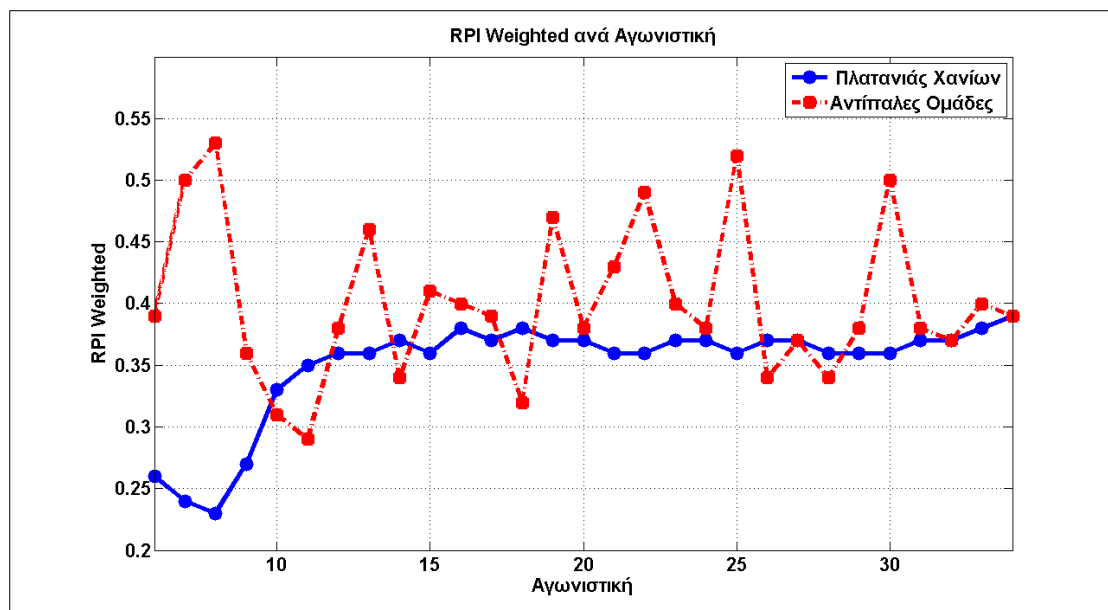
Εικόνα 8-84 : Οι τιμές του RPI Weighted του Ολυμπιακού και των αντιπάλων του ανά αγωνιστική



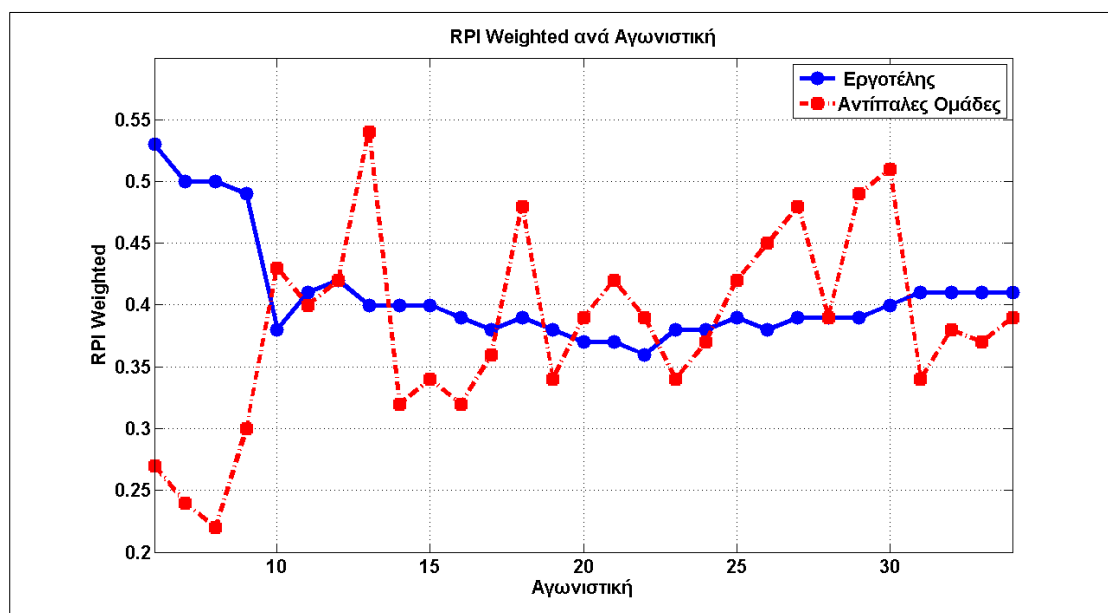
Εικόνα 8-85 : Οι τιμές του RPI Weighted του Πας Γιάννενα και των αντιπάλων του ανά αγωνιστική



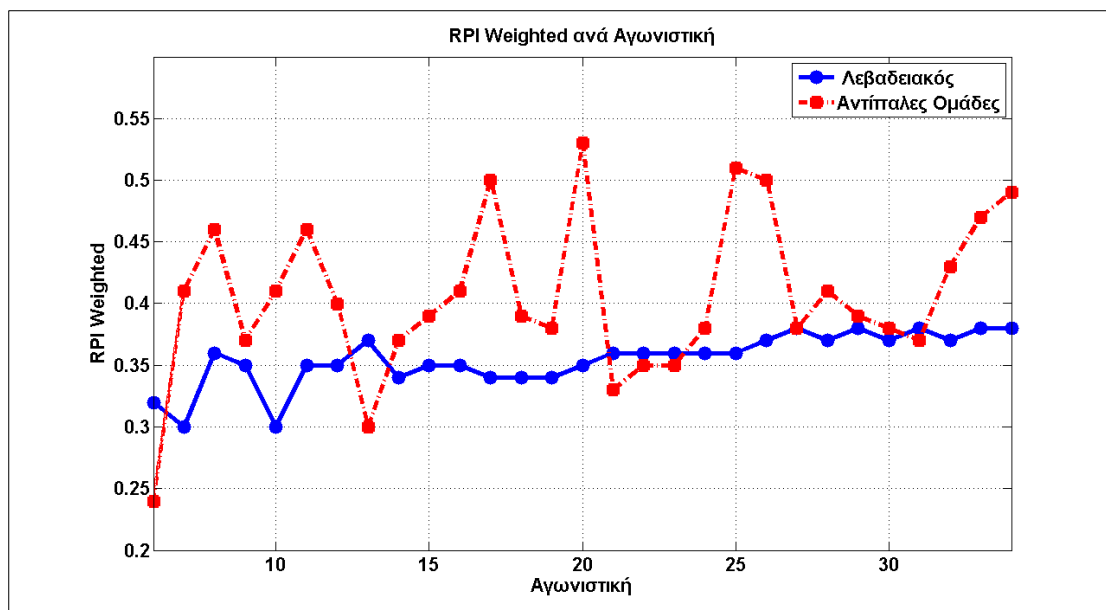
Εικόνα 8-86: Οι τιμές του RPI Weighted του Αστέρα Τρίπολης και των αντιπάλων του ανά αγωνιστική



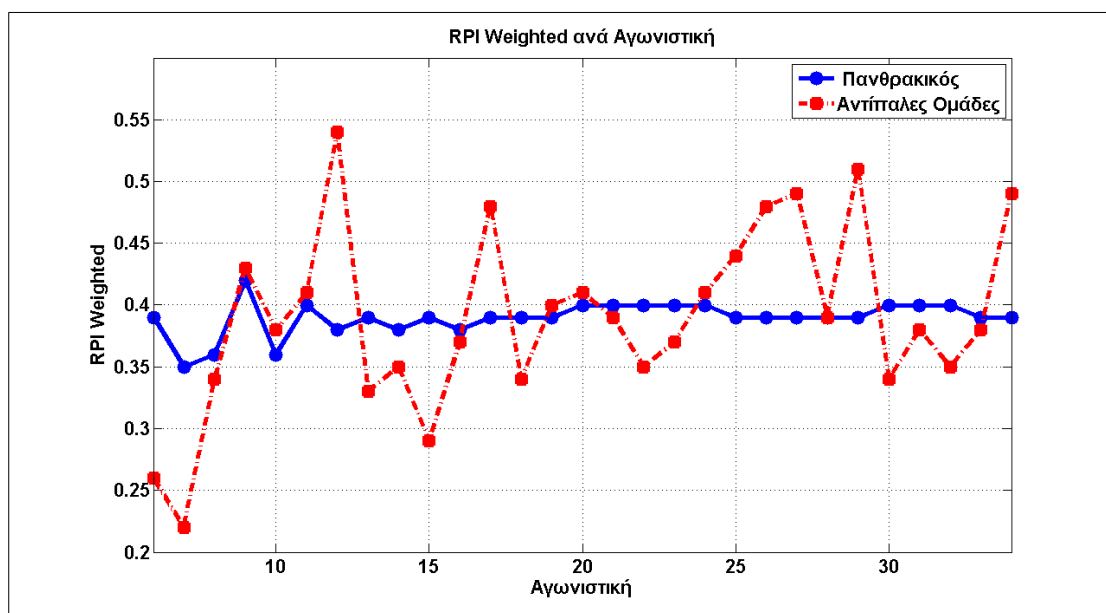
Εικόνα 8-87 : Οι τιμές του RPI Weighted του Πλατανιά Χανίων και των αντιπάλων του ανά αγωνιστική



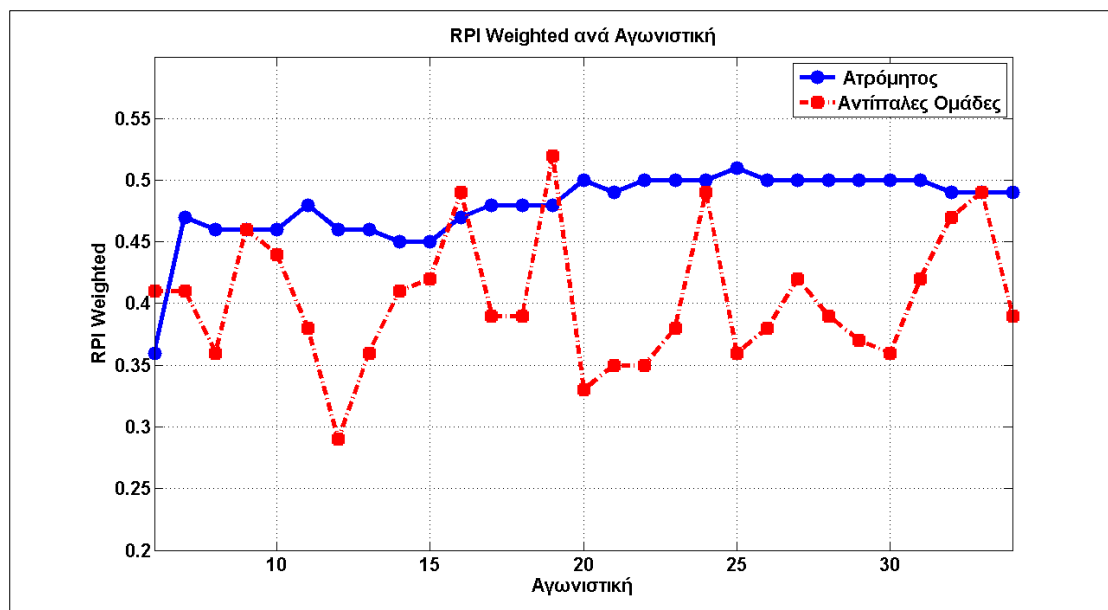
Εικόνα 8-88 : Οι τιμές του RPI Weighted του Εργοτέλη και των αντιπάλων του ανά αγωνιστική



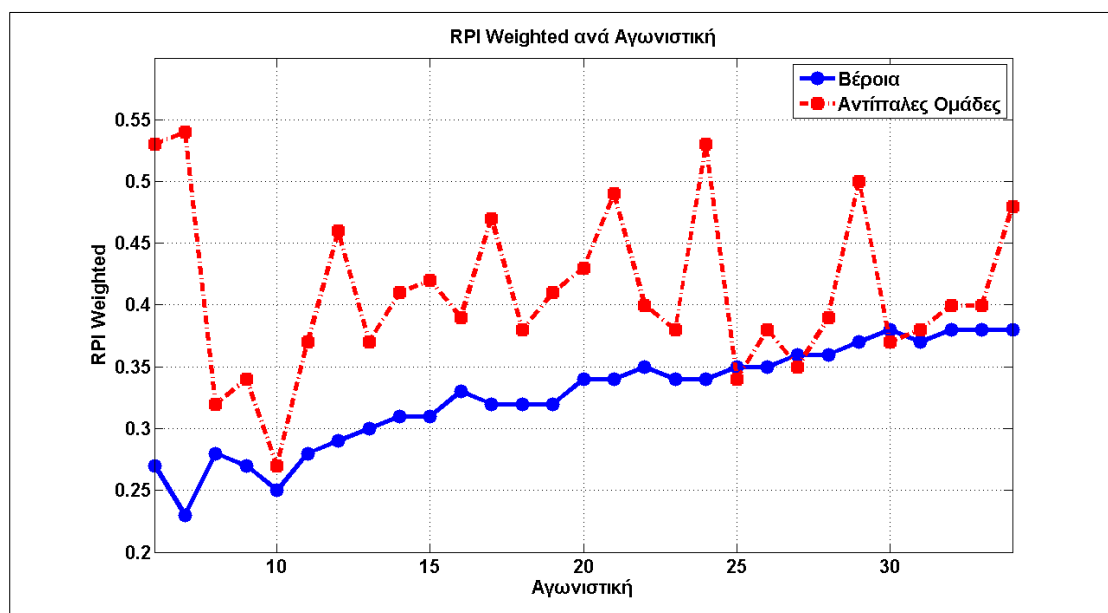
Εικόνα 8-89 : Οι τιμές του RPI Weighted του Λεβαδειακού και των αντιπάλων του ανά αγωνιστική



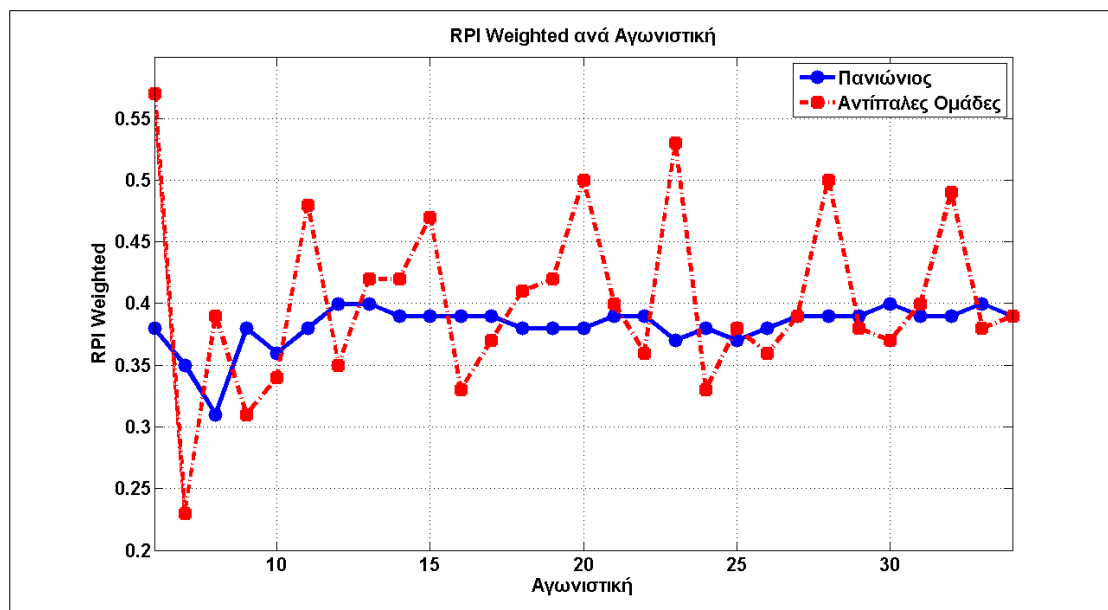
Εικόνα 8-90 : Οι τιμές του RPI Weighted του Πανθρακικού και των αντιπάλων του ανά αγωνιστική



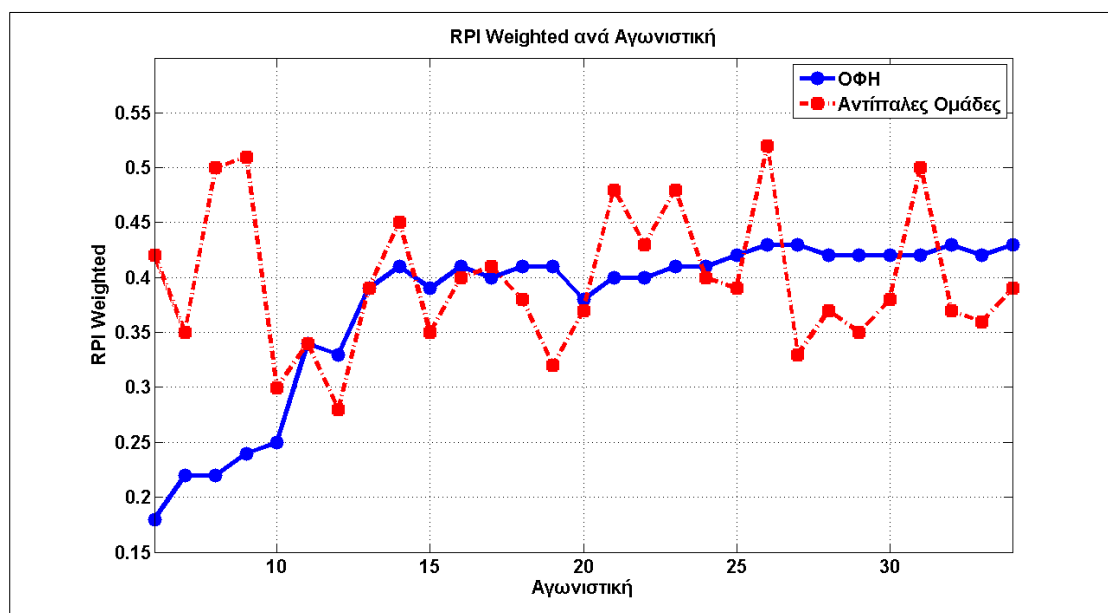
Εικόνα 8-91 : Οι τιμές του RPI Weighted του Ατρομήτου και των αντιπάλων του ανά αγωνιστική



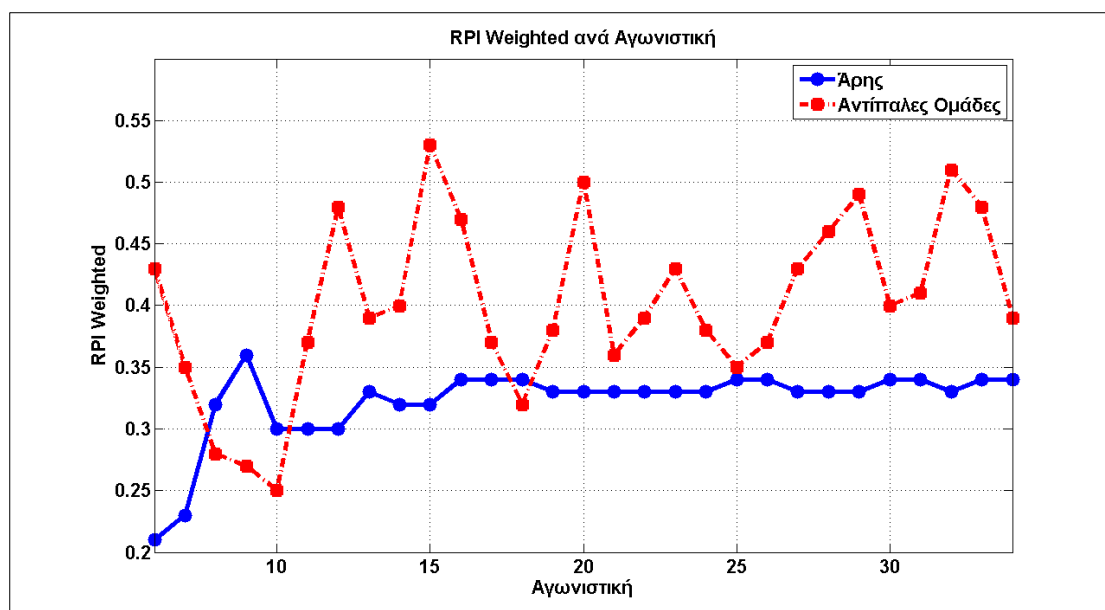
Εικόνα 8-92 : Οι τιμές του RPI Weighted της Βέροιας και των αντιπάλων της ανά αγωνιστική



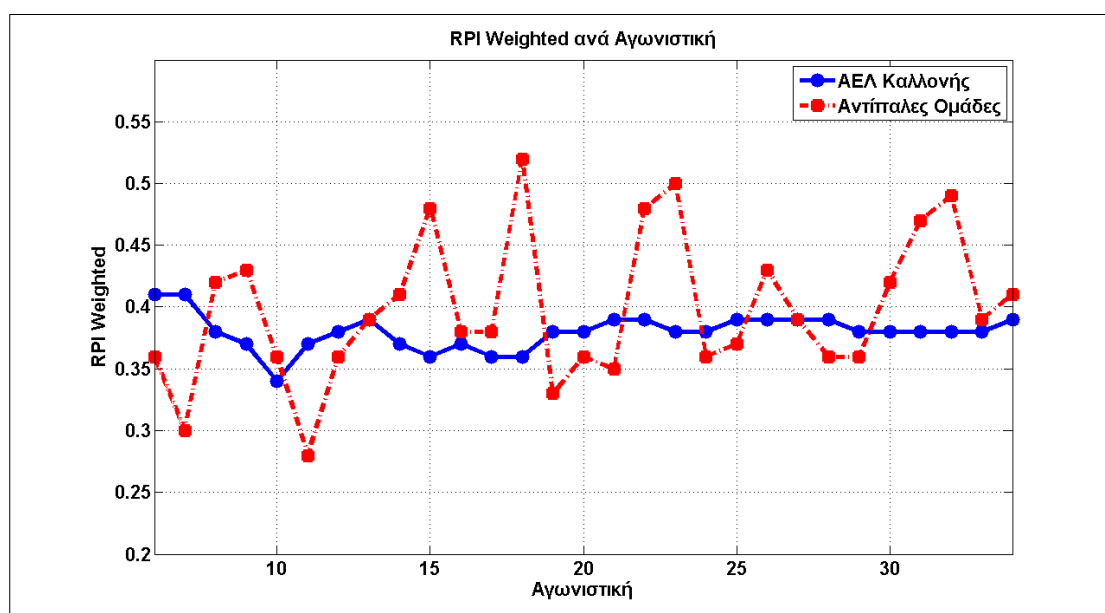
Εικόνα 8-93 : Οι τιμές του RPI Weighted του Πανιωνίου και των αντιπάλων του ανά αγωνιστική



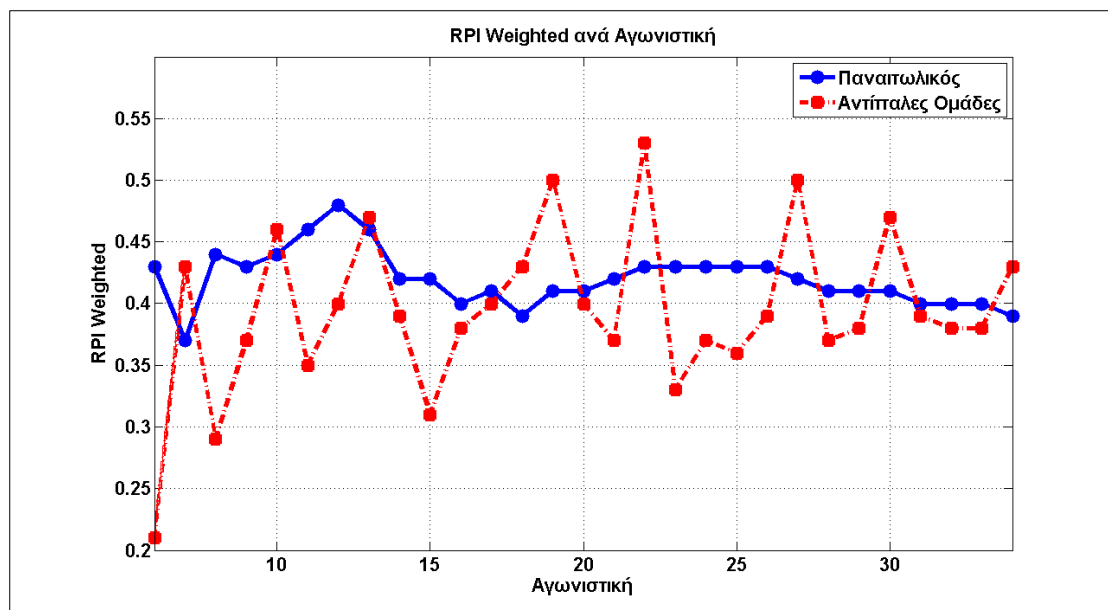
Εικόνα 8-94 : Οι τιμές του RPI Weighted του ΟΦΗ και των αντιπάλων του ανά αγωνιστική



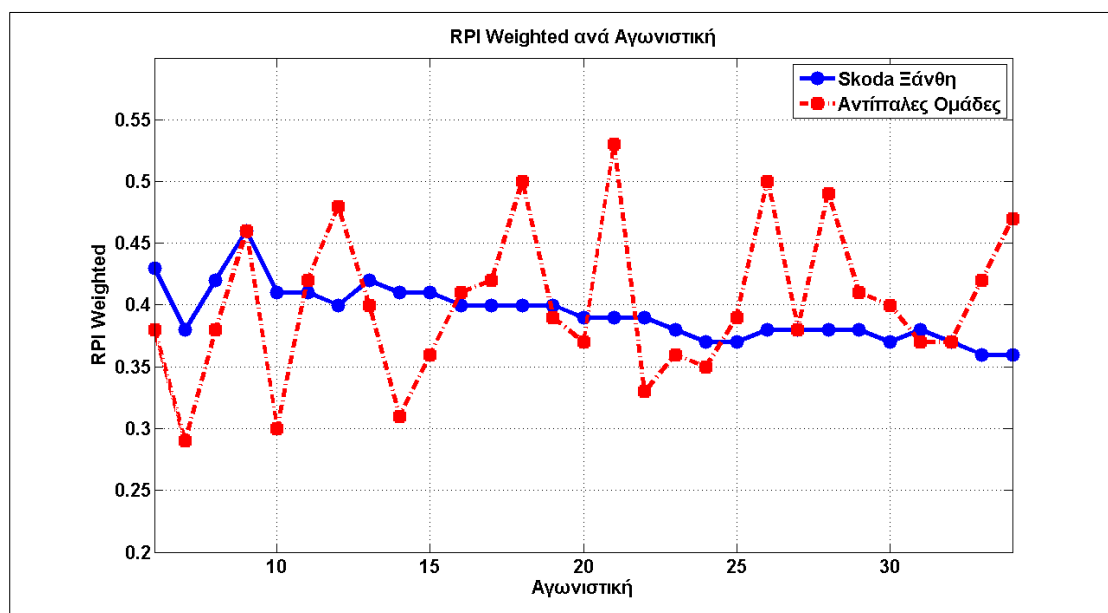
Εικόνα 8-95 : Οι τιμές του RPI Weighted του Άρη και των αντιπάλων του ανά αγωνιστική



Εικόνα 8-96 : Οι τιμές του RPI Weighted του Πανωνίου και των αντιπάλων του ανά αγωνιστική



Εικόνα 8-97 : Οι τιμές του RPI Weighted του Παναθηναϊκού και των αντιπάλων του ανά αγωνιστική



Εικόνα 8-98 : Οι τιμές του RPI Weighted της Skoda Ξάνθης και των αντιπάλων της ανά αγωνιστική

Βιβλιογραφία

- [1] [Online] <http://www.sport24.gr>
- [2] [Online] <http://www.finalscore.gr>
- [3] [Online] <http://www.greek-sportbook.gr/proistoria.php>
- [4] [Online] <http://www.sportdata.gr>
- [5] [Online]: http://www.dunkelindex.com/aboutindex_copy%281%29.asp
- [6] Suzuki, K. and Ohmori, K. (2008), “Effectiveness of FIFA/Coca-Cola World Ranking in predicting the results of FIFA World Cup Finals”, *Football Science* 5, 18–25.
- [7] [Online] “FIFA/Coca-Cola World Ranking Procedure”.
<http://www.fifa.com/worldranking/procedureandschedule/menprocedure/index.html>.
- [8] [Online] “Football-rankings.info :Elo Ratings”.
<http://www.football-rankings.info.html>.
- [9] Rao, P. and Kupper, L. (1967), “Ties in paired-comparison experiments: A generalization of the Bradley-Terry model”, *Journal of the American Statistical Association* 62(317), 194–204.
- [10] Davidson, R. (1969), “On extending the Bradley-Terry model to accommodate ties in paired comparison experiments”, Technical report, FSU Statistics Report M169 ONR Technical Report No. 37.
- [11] [Online] “FIFA Women’s World Ranking Methodology”.
<http://www.fifa.com/worldranking/procedureandschedule/womenprocedure/index.html>.
- [12] [Online] “The World Football Elo Rating System”.
<http://www.eloratings.net/system.html>.
- [13] [Online] “Jeff Sagarin’s Computer Ratings”. <http://www.sagarin.com> .
- [14] [Online] “Jeff Sagarin’s Computer Ratings”.
<http://www.usatoday/sports/sagarin.htm> .
- [15] Brady T. West, “A Simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament”, *Journal of Quantitative Analysis in Sports*, 2006
- [16] [Online] “Rating Percentage Index”. <http://www.rpiratings.com> .
- [17] [Online] “Rating Percentage Index”. <http://www.RealTimeRPI.com>

- [18] [Online] “Rating Percentage Index”. <http://www.Live-RPI.com> .
- [19] [Online] “Rating Percentage Index”. <http://www.RPIForecast.com> .
- [20] [Online] “Rating Percentage Index”. <http://www.TeamRankings.com>
- [21] Bryan L. Boulier, H.O Stekler, “Predicting the outcomes of National Football League games”, International Journal of Forecasting 19(2003), 257-270.
- [22] Ακακιάδου Γεωργία, “Μελέτη του αλγορίθμου ομαδοποίησης k-means σε δεδομένα του παγκόσμιου ιστού”, διπλωματική εργασία, Τμήμα Πληροφορικής Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, 2007.
- [23] P.-N. Tan, M.Steinbach, V. Kumar, “Introduction to Data Mining”, Addison Wesley, 2006.
- [24] A.K. Jain, R.C. DUBES, “Algorithms for Clustering Data”, Prentice-Hall advanced reference series, Prentice-Hall, Inc., Upper Saddle River, NJ, 1998.
- [25] B. Mirkin, “Mathematical Classification and Clustering”, Kluwer Academic Publishers Group, 1996.
- [26] S. K. Halgamuge, W. Lipo, “Classification and Clustering for Knowledge Discovery”, Springer-Verlag Berlin and Heidelberg GmbH & Co, 2005.
- [27] J.B. McQueen, “Some Methods for Classification and Analysis of Multivariate Observations”, Proc. of the 5th Berkley Symposium on Mathematical Statistics and Probability, pp. 281-297, 1994.
- [28] Ορέστης – Σταύρος Λοϊζίδης “How good is your work? A New Conference Classification Approach for Judging the Quality of a Publication”, διπλωματική εργασία, Σχολή ΗΜΜΥ Πολυτεχνείου Κρήτης, 2012.
- [29] Karl Pearson (June 20, 1895) “Notes on regression and inheritance in the case of two parents”, Proceedings of the Royal Society of London, 58 : 240–242.
- [30] Galton, F. (1886) “Regression towards mediocrity in hereditary stature”, Journal of the Anthropological Institute of Great Britain and Ireland, 15 : 246–263.
- [31] [Online] <http://www.socscistatistics.com>
- [32] M.J Maher, “Modelling association football scores”, Statistica Neerlandica 36(1982), nr.3.
- [33] Morris D. “The Soccer Tribe. London: Cape 1981”.
- [34] Dowie J. “Why Spain Should Win the World Cup”. New Sci 1982;94: 693-95.
- [35] Pollard R. “Home advantage in soccer: a retrospective analysis”. J Sports Sci 1986; 4: 237-48.

- [36] Pollard R. "Home advantage in soccer: variations in its magnitude and a literature review of the inter-related factors associated with its existence". *J Sport Behav* 2006; 29: 169-89.
- [37] Wolfson S, Wakelin D, Lewis M. "Football supporters' perceptions of their role in the home advantage". *J Sports Sci* 2005; 23: 365-74.
- [38] Lewis M, Goltsi V. "Perceptions of contributions to the home advantage by English and Greek football fans". In: Theodorakis Y, Goudas M, Papaioannou A, Eds. *Book of long papers, 12th European Congress of Sport Psychology*; 2007: Halkidiki, Greece: FEPSAC 2007; pp. 61-64.
- [39] Pollard R, Pollard G. "Long-term trends in home advantage in professional team sports in North America and England (1876-2003)". *J Sports Sci* 2005; 23: 337-50.
- [40] Nevill AM, Newell SM, Gale S. "Factors associated with home advantage in English and Scottish soccer matches". *J Sports Sci* 1996; 14: 181-86.
- [41] Boyko RH, Boyko AR, Boyko MG. "Referee bias contributes to home advantage in English Premiership football". *J Sports Sci* 2007;25: 1185-94.
- [42] Pollard R, da Silva CD, Nísio CM. "Home advantage in football in Brazil: differences between teams and the effects of distance traveled". *Braz J Soccer Sci* 2008; Forthcoming.
- [43] Pollard R, Seckin A. "Why is home advantage in South-east Europe the highest in the world?" In: Theodorakis Y, Goudas M, Papaioannou A, Eds. *Book of long papers, 12th European Congress of Sport Psychology*; 2007: Halkidiki, Greece: FEPSAC 2007; pp. 53-56.
- [44] Seckin A, Pollard R. "Home advantage in Turkish professional soccer". *J Sports Sci Med* 2007; Suppl.10: 203-04.
- [45] Glamser FD. "Contest location, player misconduct, and race: a case from English soccer". *J Sport Behav* 1990; 13: 41-49.
- [46] Lefebvre LM, Passer MW. "The effects of game location and importance on aggression in team sport". *Int J Sport Psych* 1974; 5: 102-10.
- [47] Garciano L, Palacios-Huerta I, Prendergast C. "Favoritism under social pressure". Working Paper 8376. Cambridge: National Bureau of Economic Research 2001.
- [48] Nevill A, Balmer N, Williams M. "Crowd influence on decisions in association football". *Lancet* 1999; 353: 1416.
- [49] Nevill AM, Balmer NJ, Williams AM. "The influence of crowd noise and experience upon refereeing decisions in football". *Psychol Sport Exerc* 2002; 3: 261-72.

[50] Sutter M, Kocher MG. "Favoritism of agents - The case of referees' home bias". J Econ Psychol 2004; 25: 461-69. [30] Dohmen TJ. Social pressure influences decisions of individuals: evidence from the behavior of football referees. Discussion Paper No. 1595. Bonn: IZA 2005.

[51] Buraimo B, Forrest D, Simmons R. "The twelfth man? Refereeing bias in English and German soccer". Working Paper Series, Paper No. 07-07. Limoges, France: International Association of Sports Economists 2007.

[52] Dawson P, Dobson S, Goddard J, Wilson J. "Are football referees really biased and inconsistent?: evidence on the incidence of disciplinary sanction in the English Premier League". J R Statist Soc A 2007; 170: 231-50.

[53] Neave N, Wolfson S. "The home advantage: psychological and physiological factors in soccer". In: Lavalle D, Thatcher J, Jones MV, Eds. Coping and Emotion in Sport. New York, Nova Science Publishers 2004; 127-44.

[54] Bray SR, Law J, Foyle, J. "Team quality and game location effects in English professional soccer". J Sport Behav 2003; 26: 319-34.

[55] Page L, Page K. "The second leg home advantage: evidence from European football cup competitions". J Sports Sci 2007; 25: 1547-56.

[56] Ευθύμιος Ι. Νικολάου, "Συγκριτική Ανάλυση και Εφαρμογή Γραμμικών, Μη-γραμμικών και Νευρο-ασαφών Μεθόδων, για τη βραχυπρόθεσμη πρόβλεψη παραγωγής ενέργειας από αιολικά πάρκα", διπλωματική εργασία, Σχολή ΗΜΜΥ Πολυτεχνείου Κρήτης, 2007.

[57] Γεωργία Μαργιά, "Ανάλυση και Πρόβλεψη Χρονοσειρών", διπλωματική εργασία, Τμήμα Φυσικής Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης, 2009.

[58] Γεώργιος Σ. Παναγιώτου, "Προβλέψεις Πωλήσεων των Ι.Χ αυτοκινήτων σε δεκαπέντε χώρες-μέλη της Ευρωπαϊκής Ένωσης", διπλωματική εργασία, Τμήμα Οργάνωσης και Διοίκησης Επιχειρήσεων Πανεπιστημίου Πειραιώς, 2005.