

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακή Διατριβή **στα Πληροφοριακά και Επικοινωνιακά Συστήματα**



**Ανάλυση Στατιστικών Δεδομένων και Πρόβλεψη Μελλοντικών
Αποτελεσμάτων με Χρησιμοποίηση του Αλγόριθμου Τεχνητής
Νοημοσύνης A-star**

Σταύρος Χατζηβέης

**Επιβλέπων Καθηγητής
Δρ Δημήτριος Κυριάκου**

Μάιος 2015

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

**Ανάλυση Στατιστικών Δεδομένων και Πρόβλεψη Μελλοντικών
Αποτελεσμάτων με Χρησιμοποίηση Αλγόριθμου Τεχνητής
Νοημοσύνης A-star**

Σταύρος Χατζηβέης

**Επιβλέπων Καθηγητής
Δρ Δημήτριος Κυριάκου**

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε
προς μερική εκπλήρωση των απαιτήσεων για απόκτηση

μεταπτυχιακού τίτλου σπουδών
στα Πληροφοριακά Συστήματα

από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών
του Ανοικτού Πανεπιστημίου Κύπρου

Μάιος 2015

Περίληψη

Η μεταπτυχιακή διατριβή αφορά στη συλλογή ενός μεγάλου όγκου στατιστικών δεδομένων των ποδοσφαιρικών αγώνων της διασυλλογικής διοργάνωσης Champions League 2014-15 και στην κατάλληλη επεξεργασία τους με σκοπό την πρόβλεψη του τελικού νικητή.

Για την υλοποίηση της μεταπτυχιακής διατριβής επελέγη ο αλγόριθμος τεχνητής νοημοσύνης A-star. Προκειμένου να υπολογιστούν τα κόστη μετάβασης από κόμβο σε κόμβο του αλγορίθμου χρησιμοποιήθηκε και επεξεργάστηκε κατάλληλα μία μεγάλη γκάμα στατιστικών στοιχείων από τους ποδοσφαιρικούς αγώνες του Champions League 2014-15. Ο πηγαίος κώδικας με τη χρήση κατάλληλων ερωτημάτων στη βάση δεδομένων που υλοποιήθηκε, υπολογίζει τον κάθε κόμβο που επισκέπτεται ο αλγόριθμος μέχρι να φτάσει στην τελική κατάσταση που ισοδυναμεί με την πρόβλεψη του νικητή μίας ποδοσφαιρικής αναμέτρησης. Εκτός από την πρόβλεψη του τελικού νικητή ο αλγόριθμος υπολογίζει ασφαλώς και τη διαδρομή με το ελάχιστο κόστος, καθώς και εάν υπάρχει εναλλακτική διαδρομή, προφανώς με μεγαλύτερο κόστος, μέσω της οποίας να προκύπτει ενδεχομένως επικράτηση της φαινομενικά υποδεέστερης ομάδας. Χρησιμοποιήθηκε η μεθοδολογία σταδιακής ανάπτυξης (Incremental Build Model) κατά την οποία το μοντέλο σχεδιάζεται, υλοποιείται και ελέγχεται κάθε φορά αυξητικά, έως ότου ικανοποιηθούν όλες οι απαιτήσεις.

Ο αλγόριθμος εφαρμόστηκε για πρώτη φορά στους αγώνες της φάσης των 16 και προέβλεψε σωστά τον τελικό νικητή στις 6 από τις 8 αναμετρήσεις. Εν συνεχεία και στην προσπάθεια περαιτέρω βελτίωσής του διερευνήθηκαν επιπλέον παράγοντες όπως η προϊστορία των ομάδων στη διοργάνωση, καθώς και η δυναμικότητα του πρωταθλήματος από το οποίο προέρχονται. Μοντελοποιήθηκαν έτσι δύο νέες μετρικές, οι οποίες συνδυάστηκαν με τα αποθηκευμένα στατιστικά της βάσης δεδομένων προκειμένου να δώσουν ακριβέστερα αποτελέσματα. Τελικός στόχος είναι η χρήση του αλγορίθμου για μία αρκούντως τεκμηριωμένη πρόβλεψη του νικητή του τελικού της 6ης Ιουνίου 2015.

Summary

The master thesis concerns the collection of a large volume of statistics of football matches of the club competition Champions League 2014-15 and their appropriate treatment in order to predict the final winner.

For the implementation of the master thesis, artificial intelligence algorithm A-star was chosen. In order to calculate the costs of transition from node to node of the algorithm, a wide range of statistics of football matches of Champions League 2014-15 was used and properly treated. The source code using appropriate queries to the implemented database, calculates each node that the algorithm visits until it reaches the final state which equals to the prediction of the winner of a football match. Besides predicting the winner, the algorithm surely computes the route with the minimum cost and whether there is an alternative route, certainly at a higher cost, through which possible prevalence of the seemingly inferior team arises. The method of software development used is the Incremental Build Model in which the model is designed, implemented and tested incrementally until all requirements are fulfilled.

The algorithm run for the first time in the round of 16 knock-out games and correctly predicted the winner in six of the eight games. Subsequently, and trying to further improve its outcomes, additional factors were investigated such as history of teams in the tournament and the capacity of their domestic leagues. As a result two new metrics were modeled and combined with the statistics stored in the database in order to provide more accurate results. The ultimate objective is to use the algorithm for a sufficiently documented prediction of the winner of the final, which will be held on June 6, 2015.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή Δρ Δημήτριο Κυριάκου για την καθοδήγηση που μου παρείχε καθόλη τη διάρκεια της εκπόνησης της μεταπτυχιακής διατριβής. Ευχαριστώ επίσης τους γονείς μου Παναγιώτη και Δήμητρα Χατζηβέη για την αμέριστη υποστήριξή τους όλα αυτά τα χρόνια.

Αφιερώνω τη μεταπτυχιακή διατριβή στη μέλλουσα σύζυγό μου Χαριτίνη Μπλιάτσου.

Περιεχόμενα

1	Εισαγωγή.....	1
1.1	Σκοπός της Μεταπτυχιακής Διατριβής.....	1
1.2	Στάδια Εκπόνησης της Μεταπτυχιακής Διατριβής.....	2
1.3	Υφιστάμενες Υπηρεσίες Ανάλυσης Στατιστικών Δεδομένων	3
1.3.1	Το Παιχνίδι Football Manager	4
1.3.2	Τυχερά Παιχνίδια	4
1.3.3	Αλγόριθμοι κατάταξης παικτών και ομάδων βάσει επεξεργασμένων στατιστικών στοιχείων	5
1.4	Κύρια Ερευνητικά Ερωτήματα.....	6
2.	Αλγόριθμοι Τεχνητής Νοημοσύνης.....	8
2.1	Χρήσιμες Έννοιες Προβλημάτων	8
2.2	Αλγόριθμοι Αναζήτησης.....	10
2.2.1	Ορισμοί.....	10
2.2.2	Χαρακτηριστικά Αλγορίθμων	10
2.2.3	Διαδικασία Επιλογής Αλγορίθμου Αναζήτησης	11
2.2.4	Γενικός Αλγόριθμος Αναζήτησης	12
2.2.5	Αλγόριθμοι Τυφλής Αναζήτησης.....	13

2.2.6	Αναζήτηση Πρώτα σε Βάθος (Depth-First Search).....	13
2.2.7	Αναζήτηση Πρώτα σε Πλάτος (Breadth-First Search)	16
2.2.8	Αλγόριθμοι Ευριστικής Αναζήτησης.....	18
2.2.9	Ο Αλγόριθμος Greedy Best-First Search.....	18
2.2.10	Ο Αλγόριθμος A* (Άλφα-Άστρο).....	20
2.2.11	Σύγκριση και Επιλογή Αλγορίθμου.....	22
3.	Σχεδιασμός και Υλοποίηση Βάσης Δεδομένων	26
3.1	Αναζήτηση Δεδομένων.....	26
3.2	Σχεδιασμός της Βάσης Δεδομένων.....	30
3.3	Υλοποίηση της Βάσης Δεδομένων.....	38
4.	Επεξεργασία και Καθορισμός Βαρύτητας Στατιστικών Στοιχείων.....	39
4.1	Επεξεργασία Στατιστικών Στοιχείων.....	39
4.2	Καθορισμός Βαρύτητας Στατιστικών Στοιχείων	42
5.	Σχεδιασμός και Υλοποίηση Αλγορίθμου	47
5.1	Σχεδιασμός Αλγορίθμου	47
5.2	Υπολογισμός κόστους μονοπατιών	49
5.3	Υλοποίηση και Αποτελέσματα Αλγορίθμου	61
6.	Βελτίωση Αλγορίθμου	65
6.1	Εισαγωγή Νέων Μετρικών.....	65

6.2	Εκτέλεση Βελτιωμένου Αλγορίθμου.....	68
7.	Επίλογος.....	72
	Βιβλιογραφία	74

Κεφάλαιο 1

Εισαγωγή

1.1 Σκοπός της Μεταπτυχιακής Διατριβής

Η μεταπτυχιακή διατριβή αφορά στη συλλογή ενός μεγάλου όγκου στατιστικών δεδομένων των ποδοσφαιρικών αγώνων της διασυλλογικής διοργάνωσης Champions League 2014-15 και στην κατάλληλη επεξεργασία τους με σκοπό την πρόβλεψη του τελικού νικητή.

Η εν λόγω μεταπτυχιακή διατριβή σκοπεύει να καταδείξει τον τρόπο αποτελεσματικής χρησιμοποίησης μίας μεγάλης συλλογής δεδομένων με στόχο την πρόβλεψη του τελικού αποτελέσματος ενός γεγονότος παγκόσμιου ενδιαφέροντος. Θα επιχειρήσει να καταδείξει επίσης τις δυσκολίες διατύπωσης μίας απόλυτα ακριβούς πρόβλεψης και να εξηγήσει τις πιθανές αιτίες μίας ενδεχόμενης απόκλισης από το πραγματικό τελικό αποτέλεσμα.

Η παρούσα μεταπτυχιακή διατριβή άπτεται του επιστημονικού πεδίου της τεχνητής νοημοσύνης και κύριοι άξονές της είναι:

-Η ανάλυση μιας μεγάλης συλλογής δεδομένων

-Η χρησιμοποίηση αλγορίθμου τεχνητής νοημοσύνης για επεξεργασία των δεδομένων

-Η εξαγωγή όσο το δυνατόν ακριβέστερων προβλέψεων για μελλοντικά αποτελέσματα

1.2 Στάδια Εκπόνησης της Μεταπτυχιακής Διατριβής

Στο πρώτο στάδιο της εκπόνησης της μεταπτυχιακής διατριβής μελετήθηκε η σχετική με τους αλγορίθμους τεχνητής νοημοσύνης βιβλιογραφία. Συγκεκριμένα μελετήθηκαν οι αλγόριθμοι «Αναζήτηση πρώτα σε πλάτος», «Αναζήτηση πρώτα σε βάθος», «Greedy Best First Search» και «A-star». Για την υλοποίηση της μεταπτυχιακής διατριβής επελέγη ως καταλληλότερος ο αλγόριθμος A-star, κυρίως λόγω της ιδιότητάς του να χρησιμοποιεί, προκειμένου να υπολογίσει τον κάθε κόμβο που επισκέπτεται, τόσο το κόστος που έχει ήδη δαπανηθεί για να φτάσει στον κόμβο αυτό όσο και το εκτιμώμενο κόστος μέχρι να φτάσει στην τελική κατάσταση. Αναλυτικά το στάδιο της βιβλιογραφικής αναζήτησης παρατίθεται στο Κεφάλαιο 2.

Προκειμένου να υπολογιστούν τα κόστη μετάβασης από κόμβο σε κόμβο του αλγορίθμου, χρησιμοποιήθηκε και επεξεργάστηκε κατάλληλα μία μεγάλη γκάμα στατιστικών στοιχείων από τους ποδοσφαιρικούς αγώνες του Champions League 2014-15. Συνεπώς, το δεύτερο στάδιο εκπόνησης της διατριβής αφορούσε το σχεδιασμό και την υλοποίηση της κατάλληλης βάσης δεδομένων με χρήση του πακέτου λογισμικού Microsoft SQL Server 2012. Η καταχώριση των δεδομένων στους πίνακες της βάσης δεδομένων διήρκεσε καθόλη τη διάρκεια της χρονιάς και για την πραγματοποίησή της αντλήθηκαν δεδομένα από το δικτυακό τόπο <http://www.whoscored.com/>, στον οποίο εντοπίστηκαν συγκεντρωμένα τα πιο αναλυτικά στατιστικά στοιχεία ανά αγώνα της διοργάνωσης. Αναλυτικά ο σχεδιασμός και η υλοποίηση της βάσης δεδομένων περιγράφονται στο Κεφάλαιο 3. Παράλληλα, κατά το δεύτερο αυτό στάδιο εκπόνησης της μεταπτυχιακής διατριβής αναζητήθηκαν και καταγράφηκαν διαθέσιμες υφιστάμενες υπηρεσίες ανάλυσης στατιστικών δεδομένων, οι οποίες και παρατίθενται στο τμήμα 1.3 του παρόντος Κεφαλαίου.

Το τρίτο και τελευταίο στάδιο εκπόνησης της μεταπτυχιακής διατριβής αφορά στην υλοποίηση μίας εκδοχής του αλγορίθμου A-star με χρήση του περιβάλλοντος Visual Studio 2012 και της γλώσσας προγραμματισμού C#. Ο πηγαίος κώδικας είναι συνδεδεμένος με τη βάση δεδομένων που υλοποιήθηκε και με τη χρήση κατάλληλων ερωτημάτων σε αυτήν υπολογίζει τον κάθε κόμβο που επισκέπτεται ο αλγόριθμος μέχρι να φτάσει στην τελική κατάσταση που ισοδυναμεί με την πρόβλεψη του νικητή μίας ποδοσφαιρικής αναμέτρησης. Για τη σύνταξη των ερωτημάτων αυτών αξιοποιήθηκε και η εμπειρία από το στάδιο της συμπλήρωσης των πινάκων

της βάσης δεδομένων, κατά το οποίο φανερώθηκαν πτυχές των ποδοσφαιρικών αγώνων που περισσότερο ή λιγότερο συχνά συνηγορούν υπέρ της επικράτησης ενός εκ των αντιπάλων. Αναλυτικά η επεξεργασία των στατιστικών και ο καθορισμός της βαρύτητάς τους περιγράφονται στο Κεφάλαιο 4. Εκτός από την πρόβλεψη του τελικού νικητή ο αλγόριθμος υπολογίζει ασφαλώς και τη διαδρομή με το ελάχιστο κόστος, καθώς και εάν υπάρχει εναλλακτική διαδρομή, προφανώς με μεγαλύτερο κόστος, μέσω της οποίας να προκύπτει ενδεχομένως επικράτηση της φαινομενικά υποδεέστερης ομάδας.

Ο αλγόριθμος εφαρμόστηκε για πρώτη φορά στους αγώνες της φάσης των 16 και προέβλεψε σωστά τον τελικό νικητή στις 6 από τις 8 αναμετρήσεις. Εν συνεχεία και στην προσπάθεια περαιτέρω βελτίωσής του διερευνήθηκαν επιπλέον παράγοντες όπως η προϊστορία των ομάδων στη διοργάνωση, καθώς και η δυναμικότητα του πρωταθλήματος από το οποίο προέρχονται. Μοντελοποιήθηκαν έτσι δύο νέες μετρικές, οι οποίες συνδυάστηκαν με τα αποθηκευμένα στατιστικά της βάσης δεδομένων προκειμένου να δώσουν ακριβέστερα αποτελέσματα. Παράλληλα, σε κάθε επόμενη φάση της διοργάνωσης ο αλγόριθμος εκτελείται με ανανεωμένη έκδοση της βάσης δεδομένων, στην οποία έχουν συμπεριληφθεί και τα στατιστικά των πιο πρόσφατων αγώνων των ομάδων στη διοργάνωση. Οι λεπτομέρειες του σχεδιασμού και της υλοποίησης του αλγορίθμου αποτυπώνονται στα κεφάλαια 5 και 6. Τελικός στόχος είναι η χρήση του αλγορίθμου για μία αρκούντως τεκμηριωμένη πρόβλεψη του νικητή του τελικού της 6ης Ιουνίου 2015.

Στον επίλογο της μεταπτυχιακής διατριβής αποτυπώνονται αντικειμενικές δυσκολίες που καθιστούν αδύνατη μία απόλυτα επιτυχημένη πρόβλεψη του αλγορίθμου που αναπτύχθηκε, ενώ παράλληλα διατυπώνονται προτάσεις για την περαιτέρω βελτίωσή του.

1.3 Υφιστάμενες Υπηρεσίες Ανάλυσης Στατιστικών Δεδομένων

Η χρησιμοποίηση αλγορίθμων τεχνητής νοημοσύνης για τη διεξαγωγή προβλέψεων ή εκτιμήσεων, αξιοποιώντας διαθέσιμα στατιστικά στοιχεία, απαντάται σε διάφορες υπηρεσίες και εφαρμογές. Παρακάτω παρατίθενται τρία παραδείγματα τέτοιων υπηρεσιών και εφαρμογών.

1.3.1 Το Παιχνίδι Football Manager

Το παιχνίδι Football Manager είναι ένα παιχνίδι προσομοίωσης ποδοσφαιρικών αγώνων. Αναπτύχθηκε από την Sports Interactive (αρχικά ως Championship Manager) και κυκλοφορεί από τη Sega [14]. Αρκετές ποδοσφαιρικές ομάδες χρησιμοποιούν τη βάση δεδομένων του για αναζήτηση νέων παικτών (scouting).

Η Τεχνητή Νοημοσύνη παίζει κυρίαρχο ρόλο στο παιχνίδι. Τεχνητή Νοημοσύνη είναι ο σχεδιασμός και η μελέτη προγραμμάτων υπολογιστών που συμπεριφέρονται «έξυπνα» [01]. Η Τεχνητή Νοημοσύνη χρησιμοποιεί τη μάθηση ως μέσο για την αποδοτική επίλυση ενός προβλήματος [05]. Τόσο η τακτική όσο και οι ενέργειες των παικτών είναι πραγματικές. Οι ενέργειες των παικτών και οι υπολογισμοί κατά τη διάρκεια του παιχνιδιού βασίζονται σε πιθανότητες. Οι παίκτες ενσωματώνουν ευφυΐα, ώστε να σκέπτονται, να αναλύουν τις επιλογές τους και να επιλέγουν την καλύτερη. Συμπεριφέρονται δηλαδή όπως και στην πραγματικότητα, π.χ. αν κάποιος παίκτης δεχτεί την μπάλα και η καλύτερη επιλογή του είναι να κάνει σέντρα, θα λειτουργήσει κατά αυτόν τον τρόπο. Εάν προς το τέλος του αγώνα ένας παίκτης έχει κουραστεί δε θα κυνηγάει την μπάλα, κοκ. Ο παίκτης του Football Manager μαθαίνει την ομάδα να παίζει σύμφωνα με τη δική του λογική. Εάν π.χ. θέλει η ομάδα του να παίζει με αντεπιθέσεις, προπονεί αντίστοιχα τους παίκτες ώστε να είναι γρήγοροι και ακριβείς [02]. Εάν θέλει να παίζει παιχνίδι κατοχής, όπως η Μπαρτσελόνα, προπονεί τους παίκτες ώστε να βελτιώσουν την τεχνική τους και να δίνουν πάσες με ακρίβεια.

Ένα άλλο χαρακτηριστικό Τεχνητής Νοημοσύνης που ενσωματώνει το Football Manager είναι η ύπαρξη θετικών ή αρνητικών συνεπειών εφαρμόζοντας μία συγκεκριμένη τακτική, έχοντας στην ενδεκάδα συγκεκριμένους παίκτες. Για παράδειγμα εάν κάποιος έχει σχεδιάσει την ομάδα του να παίζει παιχνίδι κατοχής και στον αγώνα όμως αποφασίσει να την παρατάξει με 5-4-1 και να παίζει με μακρινές μπαλιές, η ομάδα δε θα αποδώσει και θα παίζει χειρότερα από ότι συνήθως [02].

1.3.2 Τυχερά Παιχνίδια

Η τεχνητή νοημοσύνη συχνά χρησιμοποιείται για να προσομοιώσει τον ανθρώπινο νου. Οι επιστήμονες θέλουν να γνωρίζουν πώς οι άνθρωποι οδηγούνται στις επόμενες κινήσεις τους. Η πρόβλεψη συμπεριφοράς είναι πολύ συνηθισμένη στην Τεχνητή Νοημοσύνη.

Πρόσφατα οι επιστήμονες δημιούργησαν ένα σύστημα που προβλέπει τα πονταρίσματα των παικτών τυχερών παιχνιδιών με εξαιρετικά μεγάλη ακρίβεια. Χρησιμοποίησαν δεδομένα από ένα σύνολο 675 online παιχνιδιών 6 παικτών του Texas Holdem [06]. Δημιούργησαν αρχικά ένα νευρωνικό δίκτυο βασισμένο στα πρώτα πονταρίσματα των παικτών. Το δίκτυο αυτό στη συνέχεια βελτιώθηκε μέσω της μάθησης που αποκτήθηκε από τις αρχικές προβλέψεις που έκανε και αποδείχτηκαν ορθές ή λανθασμένες.

Το τελικό μοντέλο που παρουσίασαν οι επιστήμονες προέβλεπε με ακρίβεια τριών δεκαδικών ψηφίων τα ποσά πονταρίσματος των 6 αυτών παικτών. Προέβλεπε επίσης με παρόμοια ακρίβεια τις συνολικές φορές που κέρδισαν ή έχασαν. Συνεπώς, βασισμένο στα αρχικά τους παιχνίδια, το μοντέλο κατάφερε να προβλέπει με συνέπεια τη μελλοντική συμπεριφορά των παικτών, τη στρατηγική τους και τελικά τις επιτυχίες και τις αποτυχίες τους στο παιχνίδι, αποδεικνύοντας ότι ακόμα και ο τζόγος μπορεί να μοντελοποιηθεί [06].

1.3.3 Αλγόριθμοι κατάταξης παικτών και ομάδων βάσει επεξεργασμένων στατιστικών στοιχείων

Σε πολλούς δικτυακούς τόπους αθλητικού ενδιαφέροντος απαντώνται αλγόριθμοι κατάταξης παικτών και ομάδων, οι οποίοι βασίζονται στη συλλογή και στην κατάλληλη επεξεργασία στατιστικών στοιχείων. Ο δικτυακός τόπος <http://www.whoscored.com> θεωρείται ότι δημοσιεύει τους πιο ακριβείς δείκτες απόδοσης στο χώρο του ποδοσφαίρου και οι δείκτες αυτοί χρησιμοποιούνται από Μέσα Μαζικής Ενημέρωσης, τις εταιρίες στοιχημάτων, αλλά και τις ίδιες τις ποδοσφαιρικές ομάδες.

Οι παίκτες και οι ομάδες βαθμολογούνται βάσει στατιστικού αλγορίθμου με υπολογισμούς που εκτελούνται κατά τη διάρκεια των αγώνων. Στους υπολογισμούς υπεισέρχονται πάνω από 200 στατιστικά στοιχεία, τα οποία και αξιολογούνται με βαρύτητα ανάλογη της επίδρασής τους στη διαμόρφωση του αποτελέσματος. Κάθε σημαντικό γεγονός λαμβάνεται υπόψη επιδρώντας είτε θετικά είτε αρνητικά στη βαθμολογία του παίκτη και κατ' επέκταση της ομάδας. Για παράδειγμα μία επιτυχημένη ντρίμπλα και μάλιστα στην περιοχή της αντίπαλης ομάδας θα έχει ως αποτέλεσμα την αύξηση της ατομικής βαθμολογίας του παίκτη που την επιχείρησε. Η κλίμακα της βαθμολογίας κυμαίνεται μεταξύ 6 και 10 και η βαθμολογία κάθε παίκτη (από την οποία προκύπτει και η βαθμολογία της ομάδας) ανανεώνεται κάθε 30 δευτερόλεπτα κατά τη διάρκεια ενός αγώνα [13]. Ένα στιγμιότυπο της διαδικασίας αυτής αποτυπώνεται στην Εικόνα 1.1.

Μετά τη λήξη του αγώνα και την οριστικοποίηση των στατιστικών στοιχείων ανακηρύσσεται ο καλύτερος παίκτης του αγώνα. Στοιχεία όπως η θέση που παίζει ο παίκτης, καθώς και ο χρόνος συμμετοχής του συνυπολογίζονται στον αλγόριθμο, διαμορφώνοντας την τελική βαθμολογία. Παρόμοιες μεθόδους χρησιμοποιούν οι εταιρίες στοιχημάτων προκειμένου να διαμορφώσουν τις αποδόσεις των προσφερόμενων κατηγοριών στοιχηματισμού.



Εικόνα 1.1: Διαφοροποίηση βαθμολογίας παικτών και ομάδων κατά τη διάρκεια ενός αγώνα

1.4 Κύρια Ερευνητικά Ερωτήματα

Αφού εξετάστηκαν οι διάφορες υφιστάμενες υπηρεσίες αξιοποίησης στατιστικών στοιχείων με τη χρήση αλγορίθμων τεχνητής νοημοσύνης, τέθηκαν τα κύρια ερωτήματα του τρόπου επεξεργασίας των στατιστικών στοιχείων στα πλαίσια της παρούσας μεταπτυχιακής διατριβής, καθώς και της επιλογής του κατάλληλου αλγόριθμου τεχνητής νοημοσύνης. Οι απαντήσεις στα δύο αυτά ερωτήματα δίνονται στις παραγράφους 4.1 και 2.2.11 αντίστοιχα. Ένα επιπλέον ερευνητικό ερώτημα που τέθηκε ήταν ο τρόπος με τον οποίο θα αποφασιστεί η βαρύτητα του κάθε στατιστικού στοιχείου, ερώτημα το οποίο απαντάται στην παράγραφο 4.2.

Κατά τη διάρκεια της μεταπτυχιακής διατριβής και αφού υλοποιήθηκε η πρώτη έκδοση του αλγορίθμου τέθηκε το ερώτημα κατά πόσον οι υπολογισμοί που εκτελούνται για τον υπολογισμό της ελάχιστης διαδρομής για την πρόβλεψη του τελικού νικητή είναι δυνατό να

βελτιωθούν με αναζήτηση στατιστικών στοιχείων και άλλων δεδομένων, τα οποία δεν αποτυπώνονται στους πίνακες των στατιστικών των ποδοσφαιρικών αγώνων. Αναλυτικά το ερώτημα αυτό και ο τρόπος προσέγγισής του αποτυπώνεται στην παράγραφο 6.1.

Κεφάλαιο 2

Αλγόριθμοι Τεχνητής Νοημοσύνης

Το παρόν κεφάλαιο αφού καταγράψει χρήσιμες έννοιες των προβλημάτων και των αλγορίθμων αναζήτησης, περιγράφει και συγκρίνει τέσσερις δημοφιλείς αλγορίθμους τεχνητής νοημοσύνης: Αναζήτηση Πρώτα σε Βάθος (Depth-First Search - DFS), Αναζήτηση Πρώτα σε Πλάτος (Breadth First Search - BFS), Greedy Best First Search και Άλφα-Άστρο (A^*).

2.1 Χρήσιμες Έννοιες Προβλημάτων

- **Χώρος Καταστάσεων**

Χώρος καταστάσεων (state space ή domain space) ενός προβλήματος ονομάζεται το σύνολο όλων των έγκυρων καταστάσεων [16].

- **Περιγραφή Προβλημάτων με Χώρο Καταστάσεων**

Ο κόσμος ενός προβλήματος αποτελείται από τα αντικείμενα, τις ιδιότητες των αντικειμένων και τις σχέσεις που τα συνδέουν.

Κατάσταση ενός κόσμου είναι ένα στιγμιότυπο (instance) ή φωτογραφία (snapshot) μίας συγκεκριμένης χρονικής στιγμής της εξέλιξης του κόσμου [16].

- **Τελεστές Μετάβασης**

Τελεστής μετάβασης (transition operator) ή ενέργεια (action) είναι μια αντιστοίχιση μίας κατάστασης του κόσμου σε νέες καταστάσεις [16].

- **Αρχικές και Τελικές Καταστάσεις**

Η αρχική (initial state) και τελική (final ή goal state) ή τελικές καταστάσεις εκφράζουν το δεδομένο και το ζητούμενο αντίστοιχα [16].

- **Ορισμός Προβλήματος**

Ένα πρόβλημα (Problem) ορίζεται ως η τετράδα $P = (I, G, T, S)$ όπου:

I είναι η αρχική κατάσταση, $I \in S$

G είναι το σύνολο των τελικών καταστάσεων, $G \subseteq S$

T είναι το σύνολο των τελεστών μετάβασης, $T: S \leftrightarrow S$

S είναι ο χώρος καταστάσεων.

- **Λύση Προβλήματος**

Λύση (Solution) σε ένα πρόβλημα (I, G, T, S) , είναι μία ακολουθία από τελεστές μετάβασης $t_1, t_2, \dots, t_n \in T$ με την ιδιότητα $g = t_n(\dots(t_2(t_1(I))))$, όπου $g \in G$.

- **Χώρος Αναζήτησης**

Δοθέντος ενός προβλήματος (I, G, T, S) , χώρος αναζήτησης (search space) SP είναι το σύνολο όλων των καταστάσεων που είναι προσβάσιμες από την αρχική κατάσταση [16].

Ο χώρος αναζήτησης είναι υποσύνολο του χώρου καταστάσεων, δηλαδή $SP \subseteq S$.

2.2 Αλγόριθμοι Αναζήτησης

2.2.1 Ορισμοί

Ένας αλγόριθμος είναι μία αυστηρά καθορισμένη ακολουθία βημάτων-εντολών που επιδιώκει να λύσει ένα πρόβλημα. Οι αλγόριθμοι που αναζητούν τη λύση σε ένα πρόβλημα ονομάζονται αλγόριθμοι αναζήτησης (search algorithms).

Η επιλογή ενός αλγορίθμου αναζήτησης για ένα συγκεκριμένο πρόβλημα είναι σημαντική, διότι οι αλγόριθμοι αυτοί διαφέρουν μεταξύ τους σε αρκετά χαρακτηριστικά. Ένας αλγόριθμος αναζήτησης δεν μειώνει τον χώρο αναζήτησης (που είναι δεδομένος), αλλά καθορίζει τον αριθμό των καταστάσεων που επισκέπτεται [16].

2.2.2 Χαρακτηριστικά Αλγορίθμων

Δοθέντος ενός προβλήματος $P=(I,G,T,S)$ και μετά την εφαρμογή κάποιου αλγορίθμου στο χώρο αναζήτησής του, προκύπτει το επιλυμένο πρόβλημα (solved problem), το οποίο ορίζεται ως μία τετράδα $P_s=(V,A,F,G_s)$, όπου:

V είναι το σύνολο των καταστάσεων που εξέτασε ο αλγόριθμος αναζήτησης,

A είναι ο αλγόριθμος που χρησιμοποιήθηκε,

F είναι το σύνολο των λύσεων που βρέθηκαν,

G_s είναι το σύνολο των τελικών καταστάσεων που εξετάστηκαν.

Ο αριθμός των καταστάσεων που περιέχει το V και η σχέση του με το χώρο καταστάσεων S ενός προβλήματος και τον χώρο αναζήτησης SP , είναι ένα από τα χαρακτηριστικά της αποδοτικότητας του αλγορίθμου. Ένας αλγόριθμος ονομάζεται εξαντλητικός (exhaustive) όταν το σύνολο των καταστάσεων που εξετάζει ο αλγόριθμος για να βρει τις απαιτούμενες λύσεις είναι ίσο με το χώρο αναζήτησης, δηλαδή $V=SP$. Κλάδεμα ή αποκοπή καταστάσεων (pruning)

του χώρου αναζήτησης είναι η διαδικασία κατά την οποία ο αλγόριθμος απορρίπτει, κάτω από ορισμένες συνθήκες, κάποιες καταστάσεις και μαζί με αυτές όλο το υποδένδρο που εκτυλίσσεται κάτω από τις καταστάσεις αυτές [16].

Ένας αλγόριθμος δεν λύνει πάντα κάποιο πρόβλημα, έστω και αν υπάρχει κάποια λύση. Τότε τα σύνολα G_s και F είναι κενά. Ένας αλγόριθμος αναζήτησης ονομάζεται πλήρης (complete) αν εγγυάται ότι θα βρει μία λύση για οποιαδήποτε τελική κατάσταση, αν τέτοια λύση υπάρχει. Σε αντίθετη περίπτωση, ο αλγόριθμος ονομάζεται μη-πλήρης (incomplete) [16].

Μία λύση ονομάζεται βέλτιστη (optimal) αν οδηγεί στην καλύτερη, σύμφωνα με τη διάταξη, τελική κατάσταση. Όταν δεν υπάρχει διάταξη, μία λύση ονομάζεται βέλτιστη αν είναι η συντομότερη (shortest). Ένας αλγόριθμος αναζήτησης καλείται αποδεκτός (admissible) αν εγγυάται ότι θα βρει τη βέλτιστη λύση, αν μια τέτοια λύση υπάρχει [16].

2.2.3 Διαδικασία Επιλογής Αλγορίθμου Αναζήτησης

Η επιλογή ενός αλγορίθμου βασίζεται στα εξής κριτήρια [16]:

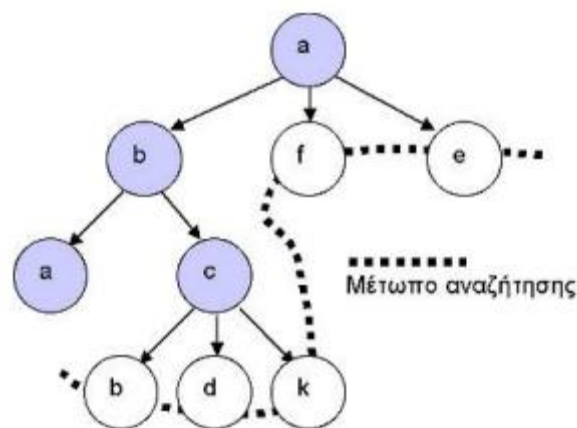
- αριθμός των καταστάσεων που αυτός επισκέπτεται
- δυνατότητα εύρεσης λύσεων εφόσον αυτές υπάρχουν
- αριθμός των λύσεων
- ποιότητα των λύσεων
- αποδοτικότητά του σε χρόνο
- αποδοτικότητά του σε χώρο (μνήμη)
- ευκολία υλοποίησής του

2.2.4 Γενικός Αλγόριθμος Αναζήτησης

Οι αλγόριθμοι αναζήτησης χρησιμοποιούν κοινές δομές δεδομένων:

Μέτωπο της αναζήτησης (search frontier) ενός αλγορίθμου είναι η λίστα των καταστάσεων που ο αλγόριθμος έχει ήδη επισκεφτεί, αλλά δεν έχουν ακόμη επεκταθεί (βλ. Εικόνα 2.1) [16].

Κλειστό σύνολο (closed set) ενός αλγορίθμου αναζήτησης είναι το σύνολο όλων των καταστάσεων που έχουν ήδη επεκταθεί [16], [04].



Εικόνα 2.1: Μέτωπο Αναζήτησης ενός Αλγορίθμου Αναζήτησης

Ένας Γενικός Αλγόριθμος Αναζήτησης περιλαμβάνει τα ακόλουθα βήματα:

1. Βάλε την αρχική κατάσταση στο μέτωπο της αναζήτησης.
2. Αν το μέτωπο αναζήτησης είναι άδειο τότε σταμάτησε.
3. Πάρε την πρώτη σε σειρά κατάσταση του μετώπου της αναζήτησης.
4. Αν η κατάσταση αυτή είναι μέρος του κλειστού συνόλου τότε πήγαινε στο βήμα 2.
5. Εάν η κατάσταση αυτή είναι τελική κατάσταση τότε τύπωσε τη λύση και πήγαινε στο βήμα 2.

6. Εφάρμοσε τους τελεστές μετάβασης για να παράγεις τις καταστάσεις-παιδιά.
7. Βάλε τις νέες καταστάσεις-παιδιά στο μέτωπο της αναζήτησης.
8. Κλάδειψε τις καταστάσεις που δε χρειάζονται (σύμφωνα με κάποιο κριτήριο), βγάζοντάς τις από το μέτωπο της αναζήτησης.
9. Κάνε αναδιάταξη στο μέτωπο της αναζήτησης (σύμφωνα με κάποιο κριτήριο).
10. Βάλε την κατάσταση-γονέα στο κλειστό σύνολο.
11. Πήγαινε στο βήμα 2.

2.2.5 Αλγόριθμοι Τυφλής Αναζήτησης

Οι αλγόριθμοι τυφλής αναζήτησης (blind search algorithms) εφαρμόζονται σε προβλήματα στα οποία δεν υπάρχει επιπρόσθετη πληροφορία που να επιτρέπει αξιολόγηση των καταστάσεων, πέραν αυτής που παρέχεται στον ορισμό του προβλήματος [16].

Στους αλγορίθμους τυφλής αναζήτησης έχει σημασία η χρονική σειρά με την οποία παράγονται οι καταστάσεις από το μηχανισμό επέκτασης [16]. Οι αλγόριθμοι τυφλής αναζήτησης μπορούν μόνο να παράγουν νέες καταστάσεις-παιδιά και να διακρίνουν αν μία κατάσταση είναι τελική ή όχι [10].

Θα μελετηθούν δύο αλγόριθμοι τυφλής αναζήτησης:

- Αναζήτηση Πρώτα σε Βάθος (Depth-First Search - DFS)
- Αναζήτηση Πρώτα σε Πλάτος (Breadth First Search - BFS)

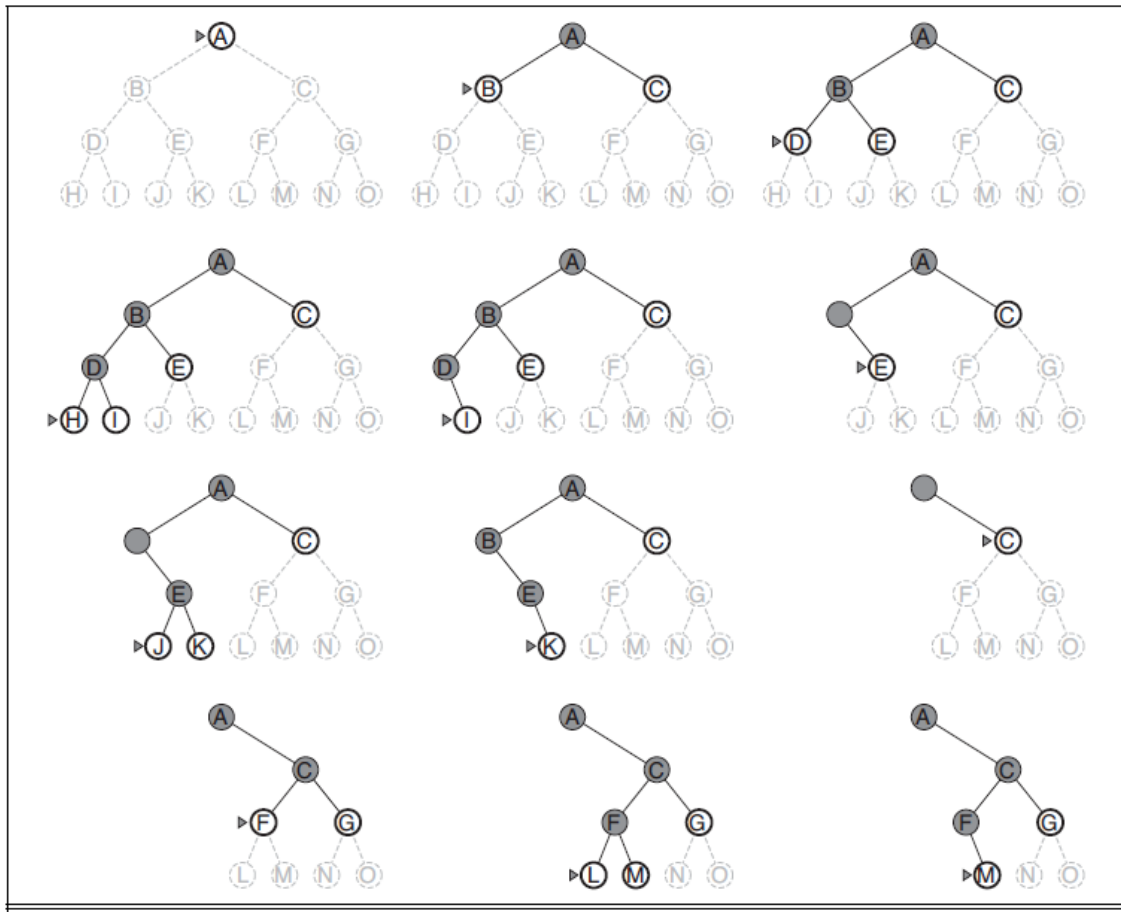
2.2.6 Αναζήτηση Πρώτα σε Βάθος (Depth-First Search)

Ο αλγόριθμος πρώτα σε βάθος (Depth-First Search - DFS) επιλέγει προς επέκταση την κατάσταση που βρίσκεται πιο βαθιά στο δένδρο [16].

Ο αλγόριθμος DFS ακολουθεί τα εξής βήματα:

1. Βάλε την αρχική κατάσταση στο μέτωπο της αναζήτησης.
2. Αν το μέτωπο της αναζήτησης είναι κενό τότε σταμάτησε.
3. Βγάλε την πρώτη κατάσταση από το μέτωπο της αναζήτησης.
4. Αν η κατάσταση ανήκει στο κλειστό σύνολο τότε πήγαινε στο βήμα 2.
5. Αν η κατάσταση είναι μία από τις τελικές, τότε ανέφερε τη λύση.
6. Αν θέλεις και άλλες λύσεις πήγαινε στο βήμα 2. Αλλιώς σταμάτησε.
7. Εφάρμοσε τους τελεστές μετάβασης για να βρεις τις καταστάσεις-παιδιά.
8. Βάλε τις καταστάσεις-παιδιά στην αρχή του μετώπου της αναζήτησης.
9. Βάλε την κατάσταση-γονέα στο κλειστό σύνολο.
10. Πήγαινε στο βήμα 2.

Το παρακάτω σχήμα (Εικόνα 2.2) [10] απεικονίζει τη σειρά με την οποία επισκέπτεται τους κόμβους ο αλγόριθμος DFS με τη μορφή δυαδικού δέντρου [10].



Εικόνα 2.2: Αλγόριθμος DFS με τη μορφή δυαδικού δέντρου

Το μέτωπο της αναζήτησης στην περίπτωση του αλγορίθμου DFS είναι μία δομή στοίβας (Stack LIFO). Η εξέταση αμέσως προηγούμενων (χρονικά) καταστάσεων ονομάζεται χρονική οπισθοδρόμηση (chronological backtracking) [16].

Πολυπλοκότητα Χρόνου

Ο DFS με τη μορφή δέντρου μπορεί να παράγει όλους τους κόμβους του δέντρου αναζήτησης σε χρόνο $O(b^m)$, όπου b είναι ο συντελεστής διακλάδωσης και m είναι το μέγιστο βάθος οποιουδήποτε κόμβου [10]. Το m μπορεί να είναι από μόνο του πολύ μεγαλύτερο από το d (το βάθος της λιγότερο «βαθιάς» λύσης). Συμπερασματικά, ο αλγόριθμος DFS δεν είναι αποδοτικός από άποψη χρόνου.

Πολυπλοκότητα Χώρου

Ο DFS με τη μορφή δέντρου χρειάζεται να αποθηκεύσει ένα μόνο μονοπάτι από τη ρίζα μέχρι τον κόμβο-φύλλο μαζί με τους υπόλοιπους κόμβους-αδέλφια που δεν έχουν ακόμα επεκταθεί, για κάθε κόμβο στο μονοπάτι. Από τη στιγμή που ένας κόμβος έχει επεκταθεί, μπορεί να αφαιρεθεί μόλις εξερευνηθούν πλήρως όλοι οι κόμβοι-παιδιά του. Για ένα χώρο καταστάσεων με συντελεστή διακλάδωσης b και μέγιστο βάθος m , ο DFS απαιτεί αποθηκευτικό χώρο για μόνο $O(bm)$ κόμβους και από αυτήν την άποψη θεωρείται αποδοτικός [10].

Συμπερασματικά, ο αλγόριθμος DFS έχει το πλεονέκτημα ότι το μέτωπο της αναζήτησης δε μεγαλώνει πάρα πολύ. Στον αντίποδα τα κύρια μειονεκτήματά του είναι ότι δεν εγγυάται ότι η πρώτη λύση που θα βρεθεί είναι η βέλτιστη και θεωρείται μη-πλήρης αλγόριθμος. Ωστόσο, όταν ο χώρος αναζήτησης είναι πεπερασμένος και χρησιμοποιείται κλειστό σύνολο, ο DFS θα βρει λύση, εάν μία τέτοια υπάρχει.

2.2.7 Αναζήτηση Πρώτα σε Πλάτος (Breadth-First Search)

Ο αλγόριθμος αναζήτησης πρώτα σε πλάτος (Breadth First Search - BFS) εξετάζει πρώτα όλες τις καταστάσεις που βρίσκονται στο ίδιο βάθος και μετά συνεχίζει στην επέκταση καταστάσεων στο αμέσως επόμενο επίπεδο [16].

Ο αλγόριθμος BFS ακολουθεί τα εξής βήματα::

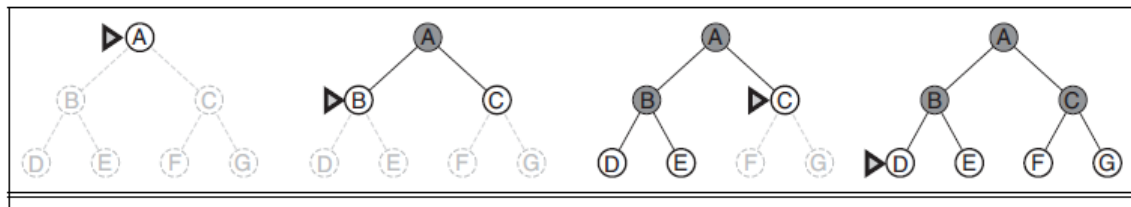
1. Βάλτε την αρχική κατάσταση στο μέτωπο της αναζήτησης.
2. Αν το μέτωπο της αναζήτησης είναι κενό τότε σταμάτησε.
3. Βγάλε την πρώτη κατάσταση από το μέτωπο της αναζήτησης.
4. Αν είναι η κατάσταση ανήκει στο κλειστό σύνολο τότε πήγαινε στο βήμα 2.
5. Αν η κατάσταση είναι μία τελική τότε ανέφερε τη λύση.
6. Αν θέλεις και άλλες λύσεις πήγαινε στο βήμα 2. Αλλιώς σταμάτησε.
7. Εφάρμοσε τους τελεστές μετάβασης για να βρεις τις καταστάσεις-παιδιά.

8. Βάλε τις καταστάσεις-παιδιά στο τέλος του μετώπου της αναζήτησης.

9. Βάλε την κατάσταση-γονέα στο κλειστό σύνολο.

10. Πήγαινε στο βήμα 2.

Το παρακάτω σχήμα (Εικόνα 2.3) [10] απεικονίζει τη σειρά με την οποία επισκέπτεται τους κόμβους ο αλγόριθμος BFS με τη μορφή δυαδικού δέντρου [10].



Εικόνα 2.3: Αλγόριθμος BFS με τη μορφή δυαδικού δέντρου

Το μέτωπο της αναζήτησης στην περίπτωση του αλγορίθμου BFS είναι μία δομή ουράς (Queue FIFO) [03]. Τα κύρια πλεονεκτήματα του BFS είναι ότι βρίσκει πάντα την καλύτερη λύση (μικρότερη σε μήκος) και ότι είναι πλήρης αλγόριθμος. Το κύριο μειονέκτημά του είναι ότι το μέτωπο της αναζήτησης μεγαλώνει πολύ σε μέγεθος [16].

- **Πολυπλοκότητα Χρόνου**

Έστω ότι ο BFS με τη μορφή δέντρου έχει b καταστάσεις-παιδιά για κάθε κόμβο. Αν η λύση (τελική κατάσταση) βρίσκεται σε βάθος d και μάλιστα είναι ο τελευταίος κόμβος που παράγεται σε αυτό το επίπεδο τότε ο συνολικός αριθμός παραγόμενων κόμβων είναι $O(b^d)$ [10].

- **Πολυπλοκότητα Χώρου**

Στον BFS με τη μορφή γράφου κάθε παραγόμενος κόμβος παραμένει στη μνήμη. Έτσι έχουμε $O(b^{d-1})$ κόμβους στο εξερευνημένο σύνολο και $O(b^d)$ κόμβους στο μέτωπο αναζήτησης [10].

Συνεπώς και σε ό,τι αφορά το χρόνο αλλά κυρίως σε ό,τι αφορά το χώρο, ο BFS δεν είναι αποδοτικός.

Γενικά, οι αλγόριθμοι τυφλής αναζήτησης δε χρησιμοποιούνται σε προβλήματα εκθετικής πολυπλοκότητας παρά μόνο εάν είναι σίγουρο ότι το βάθος d είναι μικρό.

2.2.8 Αλγόριθμοι Ευριστικής Αναζήτησης

Ο χώρος αναζήτησης ενός προβλήματος συνήθως αυξάνεται εκθετικά. Απαιτείται συνεπώς πληροφορία για αξιολόγηση των καταστάσεων (ευριστικός μηχανισμός). Οι αλγόριθμοι που εκμεταλλεύονται τέτοια πληροφορία ονομάζονται Αλγόριθμοι Ευριστικής Αναζήτησης. Αν δεν υπήρχαν ευριστικοί μηχανισμοί, τότε τα προβλήματα αυτά θα λύνονταν πολύ δύσκολα, διότι οι συνδυασμοί που πρέπει να γίνουν είναι πάρα πολλοί. Παράδειγμα ευριστικής αναζήτησης είναι η συναρμολόγηση ενός puzzle [16].

Ευριστικός (heuristic) μηχανισμός είναι μία στρατηγική, βασισμένη στη γνώση για το συγκεκριμένο πρόβλημα, η οποία χρησιμοποιείται ως βοήθημα στη γρήγορη επίλυσή του. Ο ευριστικός μηχανισμός υλοποιείται με ευριστική συνάρτηση (heuristic function). Ένας ευριστικός μηχανισμός συνήθως βελτιώνει την ποιότητα των διαδρομών που εξερευνώνται κατά την αναζήτηση [09].

Ευριστική τιμή (heuristic value) είναι η τιμή της ευριστικής συνάρτησης και εκφράζει το πόσο κοντά βρίσκεται μία κατάσταση σε μία τελική [08], [16]. Η ευριστική τιμή δεν είναι η πραγματική τιμή της απόστασης από μία τερματική κατάσταση, αλλά μία εκτίμηση που πολλές φορές μπορεί να είναι και λανθασμένη [16].

2.2.9 Ο Αλγόριθμος Greedy Best-First Search

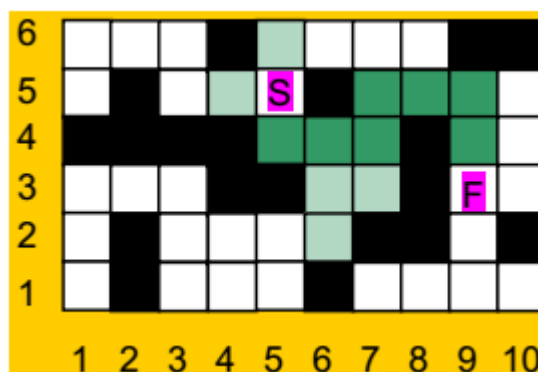
Ο αλγόριθμος Greedy BestFS κρατά όλες τις καταστάσεις στο μέτωπο αναζήτησης [16].

Ο αλγόριθμος Greedy BestFS ακολουθεί τα εξής βήματα:

1. Βάλτε την αρχική κατάσταση στο μέτωπο αναζήτησης.
2. Αν το μέτωπο αναζήτησης είναι κενό τότε σταμάτησε.
3. Πάρε την πρώτη σε σειρά κατάσταση από το μέτωπο αναζήτησης.

4. Αν η κατάσταση είναι μέλος του κλειστού συνόλου τότε πήγαινε στο 2.
5. Αν η κατάσταση είναι μία τελική τότε ανάφερε τη λύση και σταμάτα.
6. Εφάρμοσε τους τελεστές μετάβασης για να παράγεις τις καταστάσεις-παιδιά.
7. Εφάρμοσε σε κάθε παιδί την ευριστική συνάρτηση που υπολογίζει την απόσταση από την τελική κατάσταση.
8. Βάλε τις καταστάσεις-παιδιά στο μέτωπο αναζήτησης.
9. Αναδιάταξε το μέτωπο αναζήτησης, έτσι ώστε η κατάσταση με την καλύτερη (μικρότερη) ευριστική τιμή να είναι πρώτη.
10. Βάλε τη κατάσταση-γονέα στο κλειστό σύνολο.
11. Πήγαινε στο βήμα 2.

Έστω ο λαβύρινθος της Εικόνας 2.4 [16], στον οποίο μπορούμε να μετακινηθούμε κάθε φορά μόνο κατά μία θέση (πάνω, κάτω, δεξιά ή αριστερά) σε σχέση με την τρέχουσα. Ο Greedy BestFS επιχειρεί να λύσει το πρόβλημα της μετάβασης από την κατάσταση S στην κατάσταση F, χρησιμοποιώντας ως ευριστική συνάρτηση την απόσταση της κάθε κατάστασης από την τελική κατάσταση F [16].



Μέτωπο Αναζήτησης	Κλειστό Σύνολο	Κατάσταση	Παιδιά
<5-5>	<>	5-5	5-4 ⁵ ,5-6 ⁷ ,4-5 ⁷
<5-4 ⁵ ,5-6 ⁷ ,4-5 ⁷ >	<5-5>	5-4	5-5 ⁶ ,6-4 ⁴
<6-4 ⁴ ,5-5 ⁶ ,5-6 ⁷ ,4-5 ⁷ >	<5-5,5-4>	6-4	5-4 ⁷ ,6-3 ³ ,7-4 ³
<6-3 ³ ,7-4 ³ ,5-5 ⁶ ,5-6 ⁷ ,...>	<5-5,5-4,6-4>	6-3	6-4 ⁴ ,6-2 ³ ,7-3 ²
<7-3 ² ,6-2 ³ ,7-4 ³ ,6-4 ⁴ ,5-5 ⁶ ,...>	<5-5,5-4,...>	7-3	6-3 ³ ,6-4 ⁴
<6-3 ³ ,6-2 ³ ,7-4 ³ ,6-4 ⁴ ,5-5 ⁶ ,...>	<...6-3,...>	6-3	Βρόχος
<6-2 ³ ,7-4 ³ ,6-4 ⁴ ,5-5 ⁶ ,5-6 ⁷ ,...>	<...>	6-2	5-2 ⁵ ,6-3 ³
<7-4 ³ ,6-4 ⁴ ,5-2 ⁵ ,...>	<...>	7-4	7-5 ⁴ ,6-4 ⁴ ,7-3 ²
<7-3 ² ,7-5 ⁴ ,6-4 ⁴ ,5-2 ⁵ ,...>	<...7-3,...>	7-3	Βρόχος
<4-5 ⁴ ,6-4 ⁴ ,5-2 ⁵ ,...>	<...>	7-5	7-4 ³ ,8-5 ³ ,7-6 ⁵
<8-5 ³ ,7-4 ³ ,6-4 ⁴ ,...>	<...>	8-5	8-6 ⁴ ,7-5 ⁴ ,9-5 ²
<9-5 ² ,7-4 ³ ,6-4 ⁴ ,8-6 ⁴ ,...>	<...>	9-5	8-5 ³ ,9-4 ¹
<9-4 ¹ ,8-5 ³ ,7-4 ³ ,...>	<...>	9-4	9-3 ⁰ ,9-5 ² ,10-4 ²
<9-3 ⁰ ,9-5 ² ,10-4 ² ,...>	<...>	9-3	ΤΕΛΙΚΗ ΚΑΤΑΣΤΑΣΗ
		ΤΕΛΟΣ	

Εικόνα 2.4: Παράδειγμα Αλγόριθμου Greedy Best-First Search

Τα κύρια πλεονεκτήματα του Greedy BestFS είναι ότι είναι πλήρης (με τη μορφή γράφου και για πεπερασμένο αριθμό καταστάσεων) και ότι προσπαθεί, ανάλογα και με τον ευριστικό μηχανισμό που χρησιμοποιεί, να δώσει μια γρήγορη λύση σε κάποιο πρόβλημα. Το σημαντικότερο μειονέκτημα του αλγορίθμου είναι ότι το μέτωπο αναζήτησης μεγαλώνει με υψηλό ρυθμό και μαζί του μεγαλώνει και ο χώρος που χρειάζεται για την αποθήκευσή του. Επιπλέον ο αλγόριθμος Greedy BestFS δεν εγγυάται ότι η λύση που θα βρεθεί είναι η βέλτιστη [16].

Η πολυπλοκότητα χρόνου και χώρου στη χειρότερη περίπτωση είναι $O(b^m)$, όπου m είναι το μέγιστο βάθος του χώρου αναζήτησης και b είναι ο συντελεστής διακλάδωσης. Ωστόσο, με μία καλή ευριστική συνάρτηση, η πολυπλοκότητα μπορεί να μειωθεί σημαντικά [10].

2.2.10 Ο Αλγόριθμος A* (Άλφα-Άστρο)

Ο αλγόριθμος A* (Άλφα Άστρο) είναι κατά βάσει BestFS, αλλά με ευριστική συνάρτηση:

$$F(S) = g(S) + h(S)$$

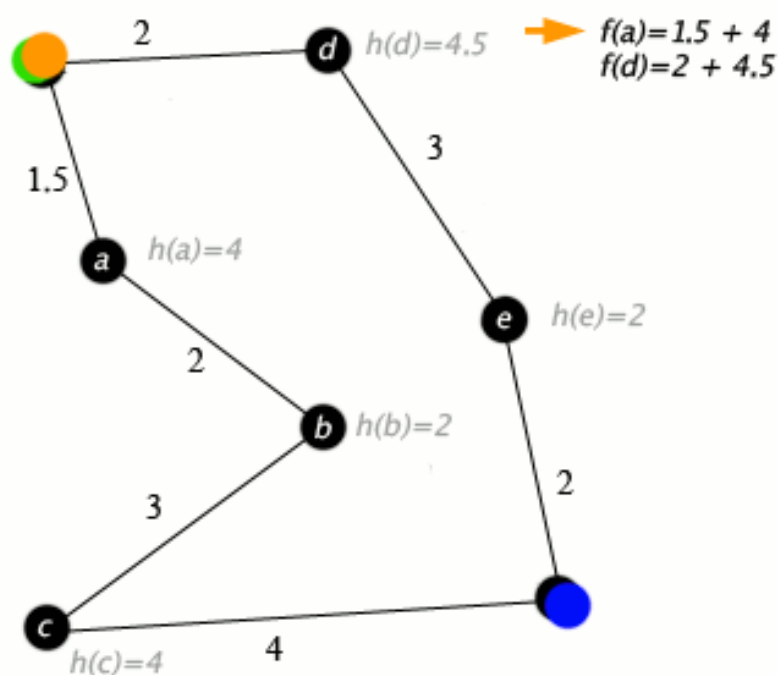
όπου:

η $g(S)$ δίνει την απόσταση της S από την αρχική κατάσταση, η οποία είναι πραγματική και γνωστή,

η $h(S)$ δίνει την εκτίμηση της απόστασης της S από την τελική κατάσταση μέσω μιας ευριστικής συνάρτησης, όπως ακριβώς στον BestFS [16].

Αν για κάθε κατάσταση η τιμή $h(S)$ είναι μικρότερη ή το πολύ ίση με την πραγματική απόσταση της S από την τελική κατάσταση, τότε ο A^* βρίσκει πάντα τη βέλτιστη λύση [16]. Συνεπώς είναι και πλήρης και βέλτιστος αλγόριθμος [07].

Στο παρακάτω σχήμα (Εικόνα 2.5) [15] δίνεται ένα παράδειγμα του αλγορίθμου A^*



Εικόνα 2.5: Παράδειγμα Αλγορίθμου A^*

Η πολυπλοκότητα χρόνου του αλγορίθμου A^* εξαρτάται από την ευριστική συνάρτηση. Στη χειρότερη περίπτωση κατά τη διάρκεια επίλυσης του προβλήματος, οι κόμβοι αυξάνονται εκθετικά. Στην περίπτωση όμως του αλγορίθμου με τη μορφή δέντρου, η πολυπλοκότητα είναι πολυωνυμική όταν υπάρχει μία μόνο τελική κατάσταση και η ευριστική συνάρτηση h έχει την παρακάτω ιδιότητα:

$$|h(x) - h^*(x)| = O(\log h^*(x))$$

όπου h^* είναι η βέλτιστη ευριστική, δηλαδή το ακριβές κόστος για τη μετάβαση από την κατάσταση x στην τελική [15].

Ωστόσο, το κύριο μειονέκτημα του αλγορίθμου δεν είναι ο υπολογιστικός χρόνος, αλλά ο αποθηκευτικός χώρος, από τη στιγμή που αποθηκεύονται στη μνήμη όλοι οι κόμβοι. Για το λόγο αυτό, ο A^* δεν είναι πρακτικός για προβλήματα μεγάλης κλίμακας.

2.2.11 Σύγκριση και Επιλογή Αλγορίθμου

Στόχος της περιγραφής των τεσσάρων αλγορίθμων τεχνητής νοημοσύνης ήταν να καταδειχθούν τα συγκριτικά τους πλεονεκτήματα και μειονεκτήματα, ώστε να επιλεγεί ο καταλληλότερος για χρήση στο πρόβλημα της πρόβλεψης του νικητή του Champions League 2014-15.

Ο Πίνακας 2.1 συνοψίζει τις διαφορές των αλγορίθμων τυφλής αναζήτησης DFS και BFS.

	Πλεονεκτήματα	Μειονεκτήματα	Πολυπλοκότητα χρόνου	Πολυπλοκότητα χώρου
DFS	Το μέτωπο της αναζήτησης δε μεγαλώνει πάρα πολύ	Δεν εγγυάται ότι η πρώτη λύση που θα βρεθεί είναι η βέλτιστη. Θεωρείται μη-πλήρης.	$O(b^m)$	$O(bm)$

BFS	Βρίσκει πάντα την καλύτερη λύση (μικρότερη σε μήκος). Είναι πλήρης.	Το μέτωπο της αναζήτησης μεγαλώνει πολύ σε μέγεθος.	$O(b^d)$	$O(b^d)$
------------	--	---	----------	----------

Πίνακας 2.1: Σύγκριση Αλγορίθμων DFS και BFS

Πέραν των γενικών χαρακτηριστικών και ιδιοτήτων, η επιλογή μεταξύ των διαθέσιμων αλγορίθμων εξαρτάται εν πολλοίς και από τη φύση του προβλήματος στο οποίο επιχειρούν να δώσουν λύση. Για παράδειγμα, αν θέλουμε να εντοπίσουμε σε ένα γενεαλογικό δέντρο έναν συγκεκριμένο συγγενή που ζει ακόμα, θα προτιμήσουμε ασφαλώς την αναζήτηση σε βάθος (DFS). Αντίθετα, αν θέλουμε να αναζητήσουμε έναν συγκεκριμένο συγγενή που έχει πεθάνει εδώ και πολλά χρόνια θα προτιμήσουμε την αναζήτηση σε πλάτος (BFS).

Όπως ήδη αναφέρθηκε, οι αλγόριθμοι τυφλής αναζήτησης μπορούν μόνο να παράγουν νέες καταστάσεις-παιδιά και να διακρίνουν αν μία κατάσταση είναι τελική ή όχι. Ακριβώς επειδή οι αλγόριθμοι τυφλής αναζήτησης δεν εφαρμόζονται σε προβλήματα που υπάρχει επιπρόσθετη πληροφορία που να επιτρέπει αξιολόγηση των καταστάσεων, δεν κρίνονται κατάλληλοι για χρήση σε ένα πρόβλημα, όπως η πρόβλεψη του νικητή του Champions League. Στο δεδομένο πρόβλημα η αναζήτηση δε θα πρέπει να είναι τυφλή αλλά να καθοδηγείται από την αποδοτική επεξεργασία της πληθώρας των στατιστικών στοιχείων που είναι διαθέσιμα προς αξιοποίηση.

Επομένως, δεδομένης της επιλογής ενός αποδοτικού ευριστικού μηχανισμού, πιο πρόσφοροι για το πρόβλημα της πρόβλεψης του νικητή μίας ποδοσφαιρικής διοργάνωσης κρίνονται οι ευριστικοί αλγόριθμοι.

Ο Πίνακας 2.2 συνοψίζει τις διαφορές των ευριστικών αλγορίθμων Greedy BestFs και A*.

Ο αλγόριθμος A^* υπερτερεί ως προς την ποιότητα των λύσεων που δίνει δεδομένου ότι ουσιαστικά αποτελεί μία βελτίωση του Greedy-BestFS. Πρόκειται για μία ευριστική μέθοδο αναζήτησης που χρησιμοποιείται για να ελαχιστοποιήσει το κόστος αναζήτησης σε ένα δεδομένο πρόβλημα. Σκοπός του αλγορίθμου είναι να βρει το μονοπάτι με τον ελάχιστο κόστος από μία συγκεκριμένη αρχική κατάσταση σε μία συγκεκριμένη τελική κατάσταση.

Ο GreedyBest first-search προσπαθεί επίσης να βρει μία λύση που να ελαχιστοποιεί το συνολικό κόστος του μονοπατιού, ωστόσο η διαφορά έγκειται στο ότι ο A^* λαμβάνει επιπλέον υπόψη του το κόστος από την αρχική κατάσταση και όχι μόνο το επιμέρους κόστος από την τρέχουσα κατάσταση μέχρι την τελική.

	Πλεονεκτήματα	Μειονεκτήματα	Πολυπλοκότητα χρόνου	Πολυπλοκότητα χώρου
Greedy BFS	<p>Προσπαθεί να δώσει μια γρήγορη λύση σε κάποιο πρόβλημα. Εξαρτάται πολύ από τον ευριστικό μηχανισμό.</p> <p>Είναι πλήρης (με τη μορφή γράφου και για πεπερασμένο αριθμό καταστάσεων)</p>	<p>Το μέτωπο αναζήτησης μεγαλώνει με υψηλό ρυθμό και μαζί του ο χώρος που χρειάζεται για την αποθήκευσή του.</p> <p>Δεν εγγυάται ότι η λύση που θα βρεθεί είναι η βέλτιστη.</p>	<p>$O(b^m)$</p> <p>στη χειρότερη περίπτωση</p> <p>Ωστόσο, με μία καλή ευριστική συνάρτηση, η πολυπλοκότητα μπορεί να μειωθεί σημαντικά.</p>	<p>$O(b^m)$</p> <p>στη χειρότερη περίπτωση</p> <p>Ωστόσο, με μία καλή ευριστική συνάρτηση, η πολυπλοκότητα μπορεί να μειωθεί σημαντικά.</p>

A*	Είναι και πλήρης και βέλτιστος αλγόριθμος.	Απαιτεί μεγάλο αποθηκευτικό χώρο, από τη στιγμή που αποθηκεύονται στη μνήμη όλοι οι κόμβοι.	Πολυωνυμική υπό προϋποθέσεις	$O(b^d)$
-----------	--	---	------------------------------	----------

Πίνακας 2.2: Σύγκριση Αλγορίθμων Greedy BestFS και A*

Συνεπώς, ο Greedy BestFS δεν μπορεί να εγγυηθεί ότι θα επιλέξει το συντομότερο μονοπάτι προς την τελική κατάσταση. Για παράδειγμα, αν υπάρχουν δύο υποψήφια καταστάσεις, εκ των οποίων η μία απέχει πολύ από την αρχική κατάσταση, αλλά η εκτίμηση για την απόστασή της από την τελική κατάσταση είναι ελαφρώς ευνοϊκότερη σε σχέση με τη δεύτερη υποψήφια κατάσταση, ο Greedy BestFS θα επιλέξει αυτήν για επόμενη κατάσταση, ακόμα και αν η δεύτερη υποψήφια κατάσταση απέχει πολύ λιγότερο από την αρχική σε σχέση με την πρώτη. Αυτό ακριβώς το μειονέκτημα διορθώνεται με τη χρήση του αλγορίθμου A*.

Επομένως, υπό προϋποθέσεις ο A* βρίσκει πάντα τη βέλτιστη λύση, σε αντίθεση με τον Greedy-BestFS, ο οποίος δε βρίσκει πάντα τη βέλτιστη λύση. Από τα παραπάνω προκύπτει, ότι μεταξύ των δύο αυτών ευριστικών αλγορίθμων είναι προτιμότερο να χρησιμοποιηθεί ο A*.

Κεφάλαιο 3

Σχεδιασμός και Υλοποίηση Βάσης Δεδομένων

3.1 Αναζήτηση Δεδομένων

Για την υλοποίηση της εκδοχής του αλγορίθμου A^* που θα προβλέπει το νικητή της ποδοσφαιρικής διοργάνωσης Champions League 2014-15 ήταν απαραίτητη η συλλογή μίας πληθώρας στατιστικών στοιχείων από τους αγώνες της τρέχουσας περιόδου. Η κατάλληλη επεξεργασία των στατιστικών αυτών αποτέλεσε εν συνεχεία τη βάση για τον καθορισμό του κόστους μετάβασης από κόμβο σε κόμβο του αλγορίθμου.

Η διαδικτυακή αναζήτηση κατέδειξε τον ιστότοπο <http://www.whoscored.com/> ως τον πλέον κατάλληλο για την άντληση αναλυτικών στατιστικών στοιχείων από τους ποδοσφαιρικούς αγώνες τη διοργάνωσης. Ο συγκεκριμένος ιστότοπος παρουσιάζει κατηγοριοποιημένα λεπτομερή στατιστικά, η αξιοπιστία των οποίων εξασφαλίζεται από το γεγονός ότι αντλούνται από την, κυρίαρχη στο χώρο των αθλητικών στατιστικών παγκοσμίως, εταιρία Opta [12]. Η

αξιοπιστία των στατιστικών επαληθεύτηκε και από σύγκρισή τους με τα πιο συνοπτικά στατιστικά που παρέχει στο δικτυακό της τόπο η Ευρωπαϊκή Ομοσπονδία Ποδοσφαίρου UEFA (Union of European Football Associations) [11].

Ενδεικτικά, στον ιστότοπο <http://www.whoscored.com/>, αναρτώνται δεδομένα όπως:

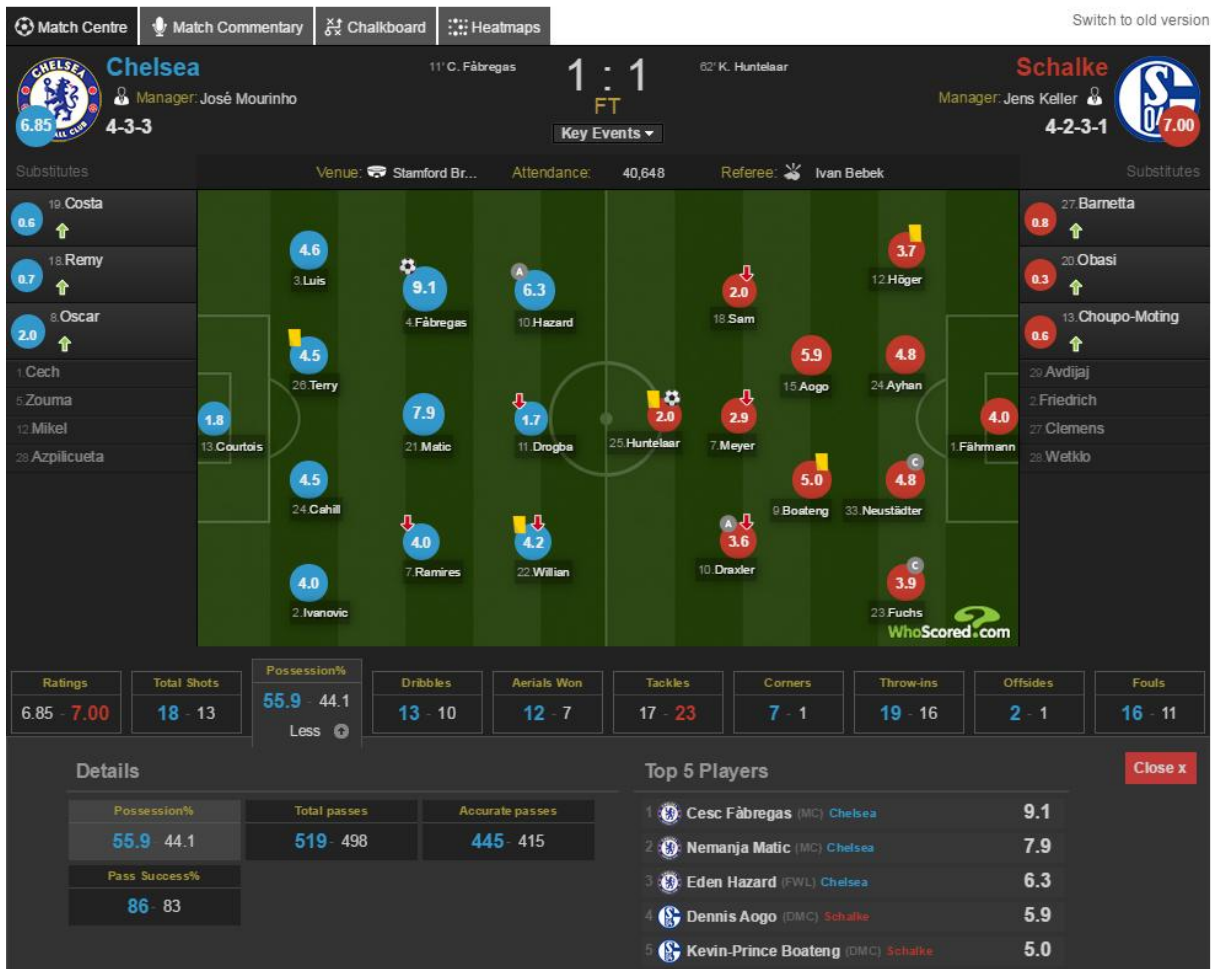
- Κερδισμένα κόρνερ
- Ποσοστό επιθέσεων ανά πλευρά του γηπέδου (αριστερά, δεξιά, κέντρο)
- Ντρίμπλες
- Οφσάιντ
- Πάσες σε παίκτη που φεύγει στον κενό χώρο
- Γεμίσματα
- Γεμίσματα που μπλοκαρίστηκαν από τον αντίπαλο
- Πάσες-κλειδί (key passes)
- Χρόνος κάθε γκολ
- Γκολ στην κανονική ροή (όχι στημένα ή αντεπιθέσεις)
- Γκολ από αντεπιθέσεις
- Γκολ από πέναλτι
- Γκολ από αυτογκόλ του αντιπάλου
- Γκολ από στημένες φάσεις
- Γκολ με το δεξί πόδι

- Γκολ με το αριστερό πόδι
- Γκολ με το κεφάλι
- Σουτ που δέχεται κάθε ομάδα
- Ποσοστό σουτ που δέχτηκαν από κάθε πλευρά του γηπέδου (αριστερά, δεξιά, κέντρο)
- Σουτ που δέχτηκαν από κάθε απόσταση (εντός μεγάλης περιοχής, εντός μικρής περιοχής, εκτός περιοχής)
- Τάκλιν (μαρκαρίσματα)
- Αποτυχημένα Τάκλιν (μαρκαρίσματα)
- Κλειψίματα
- Αποκρούσεις σε σουτ εντός μικρής περιοχής
- Αποκρούσεις σε σουτ εντός μεγάλης περιοχής
- Αποκρούσεις σε σουτ εκτός περιοχής
- Αποσοβήσεις (clearances)
- Κατοχή της μπάλας
- Ποσοστό κατοχής της μπάλας ανά περιοχή αγωνιστικού χώρου (στο δική τους πλευρά, στο κέντρο του γηπέδου, στην πλευρά του αντιπάλου)
- Αριθμός πασών
- Σωστές πάσες
- Λανθασμένες πάσες

- Πάσες που μπλοκαρίστηκαν από τον αντίπαλο
- Χαμένες κατοχές της μπάλας από λάθος control
- Χαμένες κατοχές της μπάλας από επιτυχημένο μαρκάρισμα του αντιπάλου
- Κερδισμένες εναέριες μονομαχίες
- Χαμένες εναέριες μονομαχίες
- Φάουλ που υπέπεσε
- Κερδισμένα Φάουλ
- Αριθμός κίτρινων καρτών
- Αριθμός κόκκινων καρτών
- Αριθμός καρτών για φάουλ
- Αριθμός καρτών για αντιεπαγγελματική συμπεριφορά
- Αριθμός καρτών για θέατρο
- Αριθμός καρτών για άλλους λόγους

και πολλά ακόμη.

Η Εικόνα 3.1 παρουσιάζει τον τρόπο διάρθρωσης ενός μέρους των διαθέσιμων κατηγοριών στατιστικών στον ιστότοπο <http://www.whoscored.com/>



Εικόνα 3.1: Κατηγορίες στατιστικών στον ιστότοπο <http://www.whoscored.com>

3.2 Σχεδιασμός της Βάσης Δεδομένων

Με βάση τα διαθέσιμα στατιστικά στοιχεία σχεδιάστηκαν οι πίνακες της βάσης δεδομένων και οι συσχετίσεις τους με σκοπό αφενός να περιλαμβάνουν τα απαραίτητα πεδία για την τήρηση των στατιστικών στοιχείων των αγώνων της διοργάνωσης, αφετέρου να μπορούν να εκτελεστούν με ευκολία κατάλληλα ερωτήματα για την επεξεργασία και την αποδοτικότερη αξιοποίηση των δεδομένων.

Παρακάτω παρατίθενται αναλυτικά οι πίνακες από τους οποίους αποτελείται η βάση δεδομένων, με μία σύντομη επεξήγηση του κάθε πεδίου που περιλαμβάνουν.

Πίνακες και Πεδία	Περιγραφή
<code>dbo.tbl_awayteams</code>	Ομάδες εκτός έδρας
<code>match_id</code>	Κωδικός Αγώνα

team_id	Κωδικός Ομάδας
dbo.tbl_cards	Κάρτες
card_id	Κωδικός Κάρτας
card_type	Τύπος Κάρτας (Κίτρινη ή Κόκκινη)
dbo.tbl_clearances	Αποσοβήσεις
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
total	Αριθμός Αποσοβήσεων
dbo.tbl_corners	Κόρνερ
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
total	Αριθμός κερδισμένων κόρνερ
dbo.tbl_countries	Χώρες
country_id	Κωδικός Χώρας
country_name	Όνομα Χώρας
dbo.tbl_defensives	Άμυνα
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
crossballs_blocked	Αριθμός γεμισμάτων που μπλοκαρίστηκαν
interceptions	Αριθμός Κλεψιμάτων
tackles	Αριθμός Τάκλιν (μαρκαρισμάτων)
tackles_failure	Αριθμός αποτυχημένων Τάκλιν (μαρκαρισμάτων)
corners_conceded	Αριθμός Κόρνερ που παραχωρήθηκαν
fouls_conceded	Αριθμός Φάουλ στα οποία υπέπεσε η ομάδα
penalties_conceded	Αριθμός Πέναλτι στα οποία υπέπεσε η ομάδα
goals_against	Αριθμός Γκολ τα οποία δέχθηκε η ομάδα
shots_against	Αριθμός Σουτ τα οποία δέχθηκε η ομάδα
dbo.tbl_discipline	Πειθαρχία
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
yellows	Αριθμός Κίτρινων Καρτών τις οποίες δέχθηκε η ομάδα
reds	Αριθμός Κόκκινων Καρτών τις οποίες δέχθηκε η ομάδα
totalcards_foul	Αριθμός Καρτών τις οποίες δέχθηκε η ομάδα για φάουλ
totalcards_pretends	Αριθμός Καρτών τις οποίες δέχθηκε η ομάδα για θέατρο

totalcards_unprofessional	Αριθμός Καρτών τις οποίες δέχθηκε η ομάδα για αντιεπαγγελματική συμπεριφορά των παικτών της
totalcards_other	Αριθμός Καρτών τις οποίες δέχθηκε η ομάδα για άλλους λόγους
totalcards	Συνολικός Αριθμός Καρτών που δέχθηκε η ομάδα
dbo.tbl_duels	Μονομαχίες
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
aerials_won	Αριθμός εναέριων μονομαχιών που κέρδισε η ομάδα
aerials_lost	Αριθμός εναέριων μονομαχιών που κέχασε η ομάδα
dribbles_success	Αριθμός επιτυχημένων ντριμπλών της ομάδας
dribbles_failure	Αριθμός αποτυχημένων ντριμπλών της ομάδας
dribbles_total	Συνολικός Αριθμός ντριμπλών της ομάδας
dbo.tbl_foulwins	Κερδισμένα Φάουλ
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
total	Σύνολο Κερδισμένων Φάουλ
dbo.tbl_gkallowedgoals	Γκολ εις βάρος
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
inside6yardsbox	Γκολ εις βάρος εντός μικρής περιοχής
inside18yardsbox	Γκολ εις βάρος εντός μεγάλης περιοχής
outsidebox	Γκολ εις βάρος εκτός περιοχής
dbo.tbl_gkallowedshots	Σουτ που δέχθηκε η ομάδα
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
inside6yardsbox	Σουτ που δέχθηκε η ομάδα εντός μικρής περιοχής
inside18yardsbox	Σουτ που δέχθηκε η ομάδα εντός μεγάλης περιοχής
outsidebox	Σουτ που δέχθηκε η ομάδα εκτός περιοχής
dbo.tbl_gksaves	Επεμβάσεις Τερματοφύλακα
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
inside6yardsbox	Επεμβάσεις Τερματοφύλακα εντός μικρής περιοχής
inside18yardsbox	Επεμβάσεις Τερματοφύλακα εντός μεγάλης περιοχής
outsidebox	Επεμβάσεις Τερματοφύλακα εκτός περιοχής

dbo.tbl_goalbodyparts	Γκολ από μέρη του σώματος
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
headed	Αριθμός Γκολ με το κεφάλι
leftfoot	Αριθμός Γκολ με το αριστερό πόδι
rightfoot	Αριθμός Γκολ με το δεξί πόδι
other	Αριθμός Γκολ με άλλο μέρος του σώματος
dbo.tbl_goals	Γκολ
goal_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
gls_time	Λεπτό στο οποίο σημειώθηκε το γκολ
dbo.tbl_goaltotals	Σύνολο Γκολ
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
freekick	Αριθμός Γκολ από στημένη φάση
openplay	Αριθμός Γκολ στην κανονική ροή του αγώνα (χωρίς στημένα/αντεπιθέσεις)
counterattacks	Αριθμός Γκολ από αντεπιθέσεις
penalties	Αριθμός Γκολ από Πέναλτι
opponents_own_goal	Αριθμός Γκολ από αυτογκόλ του αντιπάλου
total	Συνολικός Αριθμός Γκολ
dbo.tbl_hometeams	Ομάδες εντός έδρας
match_id	Κωδικός Αγώνα
team_id	Κωδικός Ομάδας
dbo.tbl_lineups	Ομάδες σε Αγώνες (χρησιμοποιείται για συσχέτιση με τους πίνακες που περιέχουν τα στατιστικά των διάφορων κατηγοριών)
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
match_id	Κωδικός Αγώνα
team_id	Κωδικός Ομάδας
season_id	Κωδικός Σεζόν
result	Αποτέλεσμα Ομάδας στον Αγώνα (Νίκη/Ισοπαλία/Ήττα)
dbo.tbl_matches	Αγώνες
match_id	Κωδικός Αγώνα
match_date	Ημερομηνία Αγώνα
match_attendance	Αριθμός Θεατών

round_id	Κωδικός Φάσης της Διοργάνωσης (π.χ Φάση των 16)
dbo.tbl_offenses	Παραβάσεις που οδηγούν σε παρατήρηση με κάρτα
offense_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
card_id	Κωδικός Κάρτας
ofns_time	Λεπτό στο οποίο σημειώθηκε η παράβαση
dbo.tbl_offensives	Επίθεση
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
attacks_leftpart	Ποσοστό Επιθέσεων από την αριστερή πλευρά του γηπέδου
attacks_rightpart	Ποσοστό Επιθέσεων από τη δεξιά πλευρά του γηπέδου
attacks_middle	Ποσοστό Επιθέσεων από το κέντρο του γηπέδου
offsides	Αριθμός οφσάιντ
possession_loss_badcontrol	Αριθμός λανθασμένων κοντρόλ που οδήγησαν σε απώλεια της κατοχής της μπάλας
ballpossession	Κατοχή της μπάλας
ballpossession_owthird	Ποσοστό παιχνιδιού του παιχνιδιού στο ένα τρίτο του γηπέδου που αμύνεται η ομάδα
ballpossession_opponentsthird	Ποσοστό παιχνιδιού του παιχνιδιού στο ένα τρίτο του γηπέδου που επιτίθεται η ομάδα
ballpossession_middlthird	Ποσοστό παιχνιδιού του παιχνιδιού στο ένα τρίτο του γηπέδου που βρίσκεται στο κέντρο
dbo.tbl_passes	Πάσες
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
total_success	Αριθμός Επιτυχημένων Πασών
total_failure	Αριθμός Αποτυχημένων Πασών
total	Συνολικός Αριθμός Πασών
keypasses	Αριθμός Πασών-Κλειδί (πάσες που οδηγούν σε μεγάλη ευκαιρία για γκολ)
through_balls	Αριθμός Πασών σε παίκτη που φεύγει στον κενό χώρο
cross_balls	Αριθμός Γεμισμάτων
dbo.tbl_penalties	Πέναλτι
penalty_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
penoutcome_id	Κωδικός Αποτελέσματος Εκτέλεσης Πέναλτι
pen_time	Λεπτό στο οποίο σημειώθηκε το πέναλτι

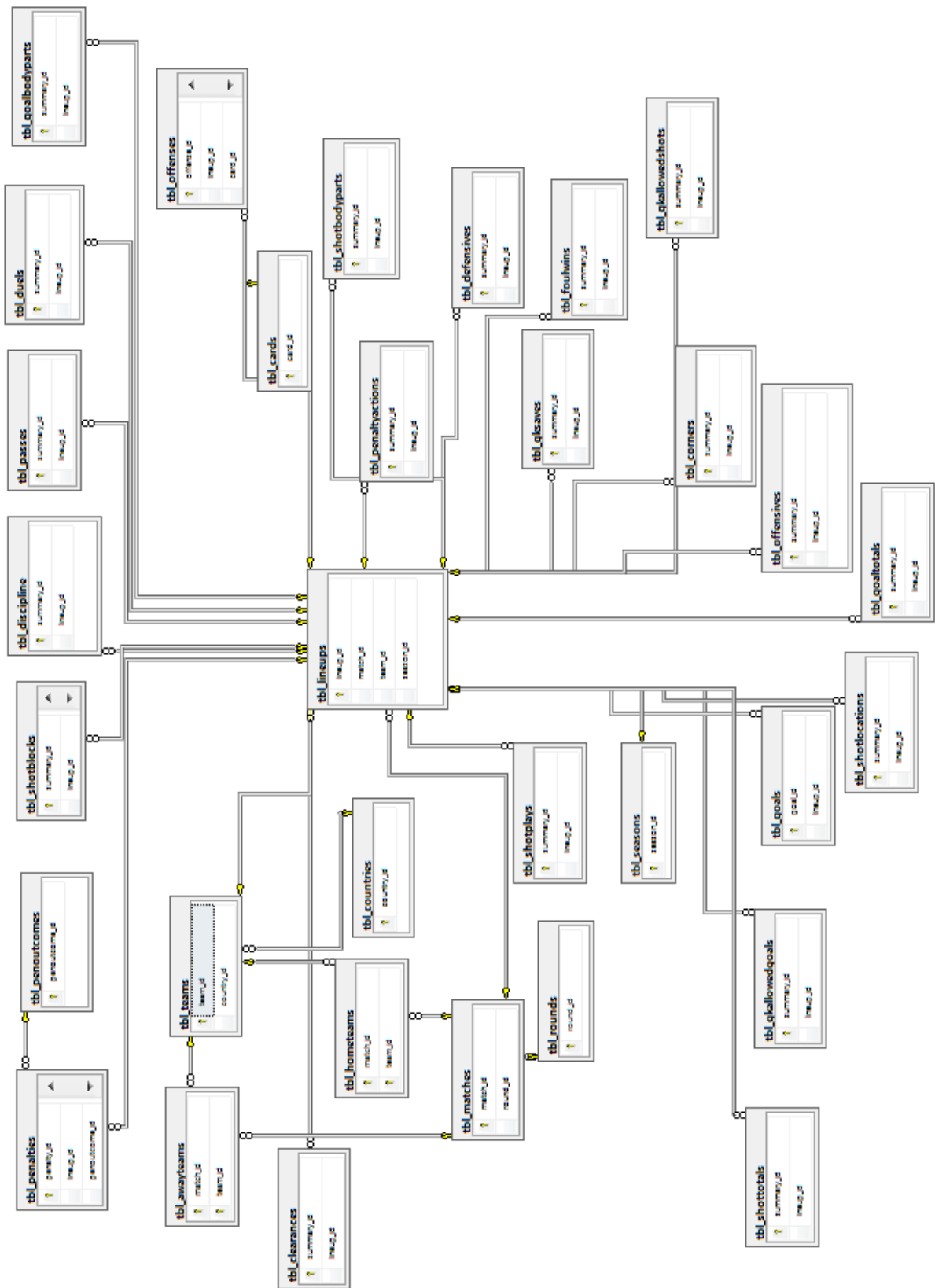
dbo.tbl_penaltyactions	Ενέργειες Πέναλι
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
taken	Αριθμός Κερδισμένων Πέναλι
saved	Αριθμός Πέναλι που αποκρούστηκαν
offtarget	Αριθμός άστοχων Πέναλι
ontarget	Αριθμός εύστοχων Πέναλι
dbo.tbl_penoutcomes	Αποτελέσματα Εκτέλεσης Πέναλι
penoutcome_id	Κωδικός Αποτελέσματος Εκτέλεσης Πέναλι
po_desc	Περιγραφή Αποτελέσματος Εκτέλεσης Πέναλι (π.χ εύστοχο/άστοχο)
dbo.tbl_rounds	Γύροι
round_id	Κωδικός Φάσης της Διοργάνωσης
round_desc	Περιγραφή Φάσης της Διοργάνωσης (π.χ φάση των 16)
dbo.tbl_seasons	Σεζόν
season_id	Κωδικός Σεζόν
season_name	Όνομα Σεζόν (π.χ. 2014-15)
dbo.tbl_shotblocks	Σουτ που μπλοκαρίστηκαν από την αντίπαλη ομάδα
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
total	Αριθμός Σουτ που μπλοκαρίστηκαν από την αντίπαλη ομάδα
dbo.tbl_shotbodyparts	Σουτ από μέρη του σώματος
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
head	Αριθμός Σουτ με το κεφάλι
leftfoot	Αριθμός Σουτ με το αριστερό πόδι
rightfoot	Αριθμός Σουτ με το δεξί πόδι
other	Αριθμός Σουτ με άλλο μέρος του σώματος
dbo.tbl_shotlocations	Σουτ από μέρη του γηπέδου
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
inside18yard	Αριθμός Σουτ εντός μεγάλης περιοχής

inside6yard	Αριθμός Σουτ εντός μικρής περιοχής
outsidebox	Αριθμός Σουτ εκτός περιοχής
leftpart	Ποσοστό σουτ από την αριστερή πλευρά του γηπέδου
rightpart	Ποσοστό σουτ από τη δεξιά πλευρά του γηπέδου
middle	Ποσοστό σουτ από το κεντρικό μέρος του γηπέδου
dbo.tbl_shotplays	Σουτ ανά τρόπο παιχνιδιού
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
openplay	Αριθμός σουτ στην κανονική ροή του αγώνα
setplay	Αριθμός σουτ που ξεκίνησαν από στημένη φάση
counterattacks	Αριθμός Σουτ σε αντεπίθεση
total	Συνολικός Αριθμός Σουτ
dbo.tbl_shottotals	Συνολικά Σουτ
summary_id	Αύξων Αριθμός
lineup_id	Κωδικός που αντιστοιχεί σε συγκεκριμένη ομάδα σε συγκεκριμένο αγώνα
ontarget	Αριθμός εύστοχων Σουτ
offtarget	Αριθμός άστοχων Σουτ
goalpost	Αριθμός Σουτ στο δοκάρι
blocked	Αριθμός Σουτ που μπλοκαρίστηκαν από την αντίπαλη ομάδα
dbo.tbl_teams	Ομάδες
team_id	Κωδικός Ομάδας
country_id	Κωδικός Χώρας
team_name	Όνομα Ομάδας
team_experience_factor	Παράγοντας Εμπειρίας στη Διοργάνωση (εξηγείται στο Κεφάλαιο 6)
country_factor	Παράγοντας Δυναμικότητας Πρωταθλήματος από το οποίο προέρχεται η ομάδα (εξηγείται στο Κεφάλαιο 6)

Πίνακας 3.1: Πίνακες και Πεδία της Βάσης Δεδομένων

Η Εικόνα 3.2 παρουσιάζει τη διαγραμματική απεικόνιση των πινάκων της βάσης δεδομένων και των μεταξύ τους συσχετίσεων.

Από την απεικόνιση αυτή καθίσταται σαφές ότι κομβικό ρόλο στη βάση δεδομένων έχει ο πίνακας tbl_lineups, κάθε γραμμή του οποίου αντιστοιχεί σε μία ομάδα σε συγκεκριμένο αγώνα. Ο πίνακας αυτός χρησιμοποιείται για συσχέτιση με τους πίνακες που περιέχουν τα στατιστικά των διάφορων κατηγοριών.



Εικόνα 3.2: Διαγραμματική απεικόνιση των πινάκων της βάσης δεδομένων και των συσχετίσεών τους

3.3 Υλοποίηση της Βάσης Δεδομένων

Η υλοποίηση του παραπάνω σχήματος της βάσης δεδομένων έγινε με χρήση του πακέτου λογισμικού Microsoft SQL Server 2012. Οι πίνακες συμπληρώθηκαν με τα στατιστικά των αγώνων (από τη φάση των ομίλων και έπειτα) των ομάδων οι οποίες προκρίθηκαν στις δεκαέξι καλύτερες της διοργάνωσης. Πιο συγκεκριμένα, περιλαμβάνονται τα στατιστικά των 80 συνολικά αγώνων των ομάδων αυτών στη φάση των ομίλων, καθώς και όλοι οι αγώνες της φάσης των 16, των προημιτελικών και των ημιτελικών. Στις νοκ άουτ φάσεις της διοργάνωσης, για λόγους δικαιοσύνης μεταξύ των ομάδων, σε περίπτωση αναμετρήσεων που κρίθηκαν στην παράταση, συμπληρώθηκαν μόνο τα στατιστικά της κανονικής διάρκειας των αγώνων.

Κεφάλαιο 4

Επεξεργασία και Καθορισμός Βαρύτητας Στατιστικών Στοιχείων

4.1 Επεξεργασία Στατιστικών Στοιχείων

Τα στατιστικά στοιχεία των ποδοσφαιρικών αναμετρήσεων με τα οποία συμπληρώθηκαν οι πίνακες της βάσης δεδομένων επεξεργάστηκαν και συνδυάστηκαν κατάλληλα, ώστε αφενός να απαλειφθούν τυχόν επαναλήψεις που υποκρύπτονται μέσα σε αυτά και αφετέρου να οδηγήσουν στην εξαγωγή πιο χρήσιμων συμπερασμάτων σχετικά με τη δυναμικότητα των ομάδων. Για την επεξεργασία αυτή των στατιστικών βοήθησε και η εμπειρία που αποκτήθηκε από την παρατήρηση κατά το στάδιο της συμπλήρωσης των πινάκων της βάσης δεδομένων.

Για παράδειγμα, παρατηρήθηκε ότι αν το ποσοστό των σουτ που επιχειρεί μία ομάδα εκτός περιοχής σε σχέση με το συνολικό αριθμό των σουτ είναι χαμηλότερο από ότι της αντίπαλης

ομάδας, οι πιθανότητες επικράτησής της είναι αυξημένες. Ως εκ τούτου, έχει μεγαλύτερη αξία να χρησιμοποιηθεί αυτό το επεξεργασμένο στατιστικό σε σύγκριση π.χ. με το πρωτόλειο δεδομένο του αριθμού των σουτ εκτός περιοχής που επιχείρησε η κάθε ομάδα. Αντιστοίχως, δεν έχει μεγάλη αξία η σύγκριση δύο ομάδων ανάλογα με τον αριθμό των χαμένων κατοχών της μπάλας, αν δε λαμβάνεται υπόψη και το συνολικό ποσοστό κατοχής της μπάλας της κάθε ομάδας. Επιπλέον, μία σύγκριση των ομάδων π.χ. με βάση τον αριθμό των καρτών που δέχονται συνιστά πλεονασμό αν έχουν ήδη συγκριθεί χωριστά στον αριθμό των καρτών που δέχονται για φάουλ, για θέατρο, για αντιεπαγγελματική συμπεριφορά και για άλλους λόγους. Με βάση τα παραπάνω, διακρίθηκαν οι ακόλουθες πρωτόλειες και επεξεργασμένες κατηγορίες στατιστικών με βάση τις οποίες θα συγκρίνεται το κάθε ζεύγος αντιπάλων.

goaltotals	Αριθμός γκολ
clearances	Αριθμός αποσοβήσεων
corners	Αριθμός κερδισμένων κόρνερ
corners_conceded	Αριθμός παραχωρημένων κόρνερ
crossballs_blocked	Αριθμός γεμισμάτων που μπλοκαρίστηκαν
fouls_conceded	Αριθμός φάουλ που υπέπεσε
interceptions	Αριθμός κλεψιμάτων
shots_against	Αριθμός σουτ που δέχτηκε
goals_against	Αριθμός γκολ που δέχτηκε
penalties_conceded	Αριθμός πέναλτι που παραχώρησε
tackles_failure	Λόγος των αποτυχημένων τάκλιν (μαρκαρίσματα) προς τα συνολικά τάκλιν
aerials_lost	Λόγος των χαμένων εναέριων μονομαχιών προς τις κερδισμένες
totalcards_foul	Αριθμός καρτών για φάουλ
gksavesinside6yardsbox	Αριθμός επεμβάσεων τερματοφύλακα εντός μικρής περιοχής
gksavesinside18yardsbox	Αριθμός επεμβάσεων τερματοφύλακα εντός μεγάλης περιοχής
gkallowedshotsinside6yardsbox	Αριθμός σουτ που δέχτηκε εντός μικρής περιοχής
gkallowedshotsinside18yardsbox	Αριθμός σουτ που δέχτηκε εντός μεγάλης περιοχής
gkallowedshotsoutsidebox	Λόγος των σουτ που δέχτηκε εκτός περιοχής προς τα σουτ που δέχτηκε εντός της μεγάλης και της μικρής περιοχής
shotblocks	Αριθμός σουτ που μπλοκαρίστηκαν
fairplay	Αριθμός καρτών για αντιεπαγγελματική συμπεριφορά, θέατρο και άλλους λόγους πλην φάουλ
reds	Αριθμός κόκκινων καρτών
ballpossession	Ποσοστό κατοχής της μπάλας
ballpossession_ownthird	Ποσοστό παιχνιδιού του παιχνιδιού στο ένα τρίτο του γηπέδου που αμύνεται η ομάδα

ballpossession_opponentsthird	Ποσοστό παιξίματος του παιχνιδιού στο ένα τρίτο του γηπέδου που επιτίθεται η ομάδα
keypasses	Αριθμός Πασών-Κλειδί (πάσες που οδηγούν σε μεγάλη ευκαιρία για γκολ)
through_balls	Αριθμός Πασών σε παίκτη που φεύγει στον κενό χώρο
cross_balls	Αριθμός Γεμισμάτων
total_success	Αριθμός επιτυχημένων πασών
total_failure	Λόγος αποτυχημένων πασών προς τις συνολικές πάσες
dribbles_failure	Λόγος αποτυχημένων ντριμπλών προς τις συνολικές ντρίμπλες
dribbles_success	Αριθμός επιτυχημένων ντριμπλών
badcontrol	Λόγος λανθασμένων κοντρόλ που οδήγησαν σε απώλεια της κατοχής της μπάλας προς την κατοχή της μπάλας
foulwins	Αριθμός κερδισμένων φάουλ
offsides	Αριθμός οφσάιντ
shottotals	Αριθμός σουτ
ontarget	Λόγος σουτ εκτός εστίας προς σουτ εντός εστίας
goalpost	Αριθμός σουτ που κατέληξαν στο δοκάρι
penaltiestaken	Αριθμός κερδισμένων πέναλτι
shotlocations	Λόγος σουτ εκτός περιοχής προς σουτ εντός μικρής και μεγάλης περιοχής
shotplays	Λόγος σουτ που ξεκίνησαν από στημένη φάση προς τα σουτ στην κανονική ροή του αγώνα
counterattacks	Αριθμός σουτ σε αντεπιθέσεις
shotsmiddlepart	Αριθμός σουτ από το κεντρικό μέρος του γηπέδου
attacksrightpart	Αριθμός επιθέσεων που εκδηλώθηκαν από τα δεξιά
wayofscoring	Λόγος γκολ σε αντεπιθέσεις προς γκολ που ξεκίνησαν από στημένη φάση
shotbodyparts	Λόγος κεφαλιών προς σουτ
goalbodyparts	Λόγος γκολ από κεφαλιές προς γκολ με σουτ

Πίνακας 4.1: Πρωτόλειες και Επεξεργασμένες Κατηγορίες Στατιστικών

4.2 Καθορισμός Βαρύτητας Στατιστικών Στοιχείων

Για τον καθορισμό της βαρύτητας καθενός από τα παραπάνω στατιστικά στοιχεία υπολογίζεται αρχικά ο λόγος:

**(Αριθμός αγώνων στη βάση δεδομένων που η νικήτρια ομάδα υπερείχε στο στατιστικό)
/ (Συνολικός αριθμός αγώνων με νικητή στη βάση δεδομένων)**

Επειδή ένας λόγος κοντά στο 0,5 προφανώς φανερώνει ένα σχετικά αδιάφορο στατιστικό στοιχείο, από τον λόγο αυτό αποφασίστηκε να αφαιρεθεί ένας συντελεστής 0,4 ώστε να πριμοδοτούνται περισσότερο οι κατηγορίες στατιστικών που συμβάλλουν πραγματικά στη διαμόρφωση ενός αποτελέσματος. Παραδείγματος χάρη η υπεροχή στα γκολ που αρχικά θα έδινε συντελεστή βαρύτητας 1 τώρα δίνει συντελεστή 0,6, θεωρείται όμως πλέον όχι 2 αλλά 6 φορές σπουδαιότερη από ένα σχετικά ασήμαντο στατιστικό που αρχικά θα είχε συντελεστή 0,5 και τώρα έχει 0,1. Ο συντελεστής αυτός αποφασίστηκε να είναι 0,4 και όχι 0,5 δεδομένου ότι υπάρχει και η όχι σπάνια περίπτωση της απόλυτης ισορροπίας δύο αντιπάλων στο συγκεκριμένο στατιστικό στοιχείο.

Επιπλέον υπάρχει μία ομάδα στατιστικών στοιχείων όπως ο αριθμός των κόκκινων καρτών ή ο αριθμός των κερδισμένων πέναλτι τα οποία ενώ είναι σημαντικά όταν συμβαίνουν, δεν απαντώνται ιδιαίτερα συχνά στις αναμετρήσεις παρά μόνο σε ένα 20-30% εξ'αυτών. Για αυτά τα στατιστικά που δεν απαντώνται σε κάθε παιχνίδι, όπως συμβαίνει π.χ. με άλλα στατιστικά όπως τα κερδισμένα φάουλ, προκειμένου να μπορέσουν να υπολογιστούν με πιο δίκαιο τρόπο οι βαρύτητές τους αφαιρείται από τον παραπάνω λόγο αντί του 0,4 ένας συντελεστής 0,15.

Ο πίνακας 4.1 παρουσιάζει τις βαρύτητες των στατιστικών στοιχείων όπως υπολογίστηκαν από τον παραπάνω τύπο μετά το τέλος της φάσης των 16. Στον πίνακα αυτό σημειώνεται επίσης για ποια στατιστικά στοιχεία υπεροχή μίας ομάδας υπάρχει όταν εμφανίζει μεγαλύτερη τιμή από την αντίπαλό της και για ποια μικρότερη τιμή. Σημειώνονται επίσης τα στατιστικά στοιχεία που δεν εμφανίζονται πάντα στις ποδοσφαιρικές αναμετρήσεις και για τα οποία η βαρύτητα έχει υπολογιστεί αφαιρώντας το συντελεστή 0,15 αντί του 0,4.

Εξυπακούεται ότι οι βαρύτητες αυτές διαφοροποιούνται ελαφρώς καθώς προχωράει η διοργάνωση και ενσωματώνονται περισσότεροι αγώνες στη βάση δεδομένων.

Όνομα Στατιστικού	Περιγραφή Στατιστικού	Βαρύτητα	Η μεγαλύτερη τιμή δείχνει υπεροχή;	Το στατιστικό απαντάται σε όλες τις αναμετρήσεις;
goaltotals	Αριθμός γκολ	0,6	ΝΑΙ	ΝΑΙ
clearances	Αριθμός αποσοβήσεων	0,13	ΟΧΙ	ΝΑΙ
corners	Αριθμός κερδισμένων κόρνερ	0,18	ΝΑΙ	ΝΑΙ
corners_conceded	Αριθμός παραχωρημένων κόρνερ	0,18	ΟΧΙ	ΝΑΙ
crossalls_blocked	Αριθμός γεμισμάτων που μπλοκαρίστηκαν	0,09	ΟΧΙ	ΝΑΙ
fouls_conceded	Αριθμός φάουλ που υπέπεσε	0,18	ΟΧΙ	ΝΑΙ
interceptions	Αριθμός κλεψιμάτων	0,05	ΟΧΙ	ΝΑΙ
shots_against	Αριθμός σουτ που δέχτηκε	0,38	ΟΧΙ	ΝΑΙ
goals_against	Αριθμός γκολ που δέχτηκε	0,6	ΟΧΙ	ΝΑΙ
penalties_conceded	Αριθμός πέναλτι που παραχώρησε	0,05	ΟΧΙ	ΟΧΙ
tackles_failure	Λόγος των αποτυχημένων τάκλιν (μαρκαρίσματα) προς τα συνολικά τάκλιν	0,26	ΟΧΙ	ΝΑΙ
aerials_lost	Λόγος των χαμένων εναέριων μονομαχιών προς τις κερδισμένες	0,26	ΟΧΙ	ΝΑΙ
totalcards_foul	Αριθμός καρτών για φάουλ	0,21	ΟΧΙ	ΝΑΙ
gksavesinside6yardsbox	Αριθμός επεμβάσεων τερματοφύλακα εντός μικρής	0,21	ΟΧΙ	ΝΑΙ

	περιοχής			
gksavesinside18yardsbox	Αριθμός επεμβάσεων τερματοφύλακα εντός μεγάλης περιοχής	0,13	OXI	OXI
gkallowedshotsinside6yardsbox	Αριθμός σουτ που δέχτηκε εντός μικρής περιοχής	0,1	OXI	NAI
gkallowedshotsinside18yardsbox	Αριθμός σουτ που δέχτηκε εντός μεγάλης περιοχής	0,34	OXI	NAI
gkallowedshotsoutsidebox	Λόγος των σουτ που δέχτηκε εκτός περιοχής προς τα σουτ που δέχτηκε εντός της μεγάλης και της μικρής περιοχής	0,17	OXI	NAI
shotblocks	Αριθμός σουτ που μπλοκαρίστηκαν	0,02	OXI	NAI
fairplay	Αριθμός καρτών για αντιεπαγγελματική συμπεριφορά, θέατρο και άλλους λόγους πλην φάουλ	0,19	OXI	OXI
reds	Αριθμός κόκκινων καρτών	0,01	OXI	OXI
ballpossession	Ποσοστό κατοχής της μπάλας	0,26	NAI	NAI
ballpossession_owthird	Ποσοστό παιξίματος του παιχνιδιού στο ένα τρίτο του γηπέδου που αμύνεται η ομάδα	0,31	OXI	NAI
ballpossession_opponents third	Ποσοστό παιξίματος του	0,31	NAI	NAI

	παιχνιδιού στο ένα τρίτο του γηπέδου που επιτίθεται η ομάδα			
keypasses	Αριθμός Πασών-Κλειδί (πάσες που οδηγούν σε μεγάλη ευκαιρία για γκολ)	0,34	NAI	NAI
through_balls	Αριθμός Πασών σε παίκτη που φεύγει στον κενό χώρο	0,22	NAI	NAI
cross_balls	Αριθμός Γεμισμάτων	0,17	NAI	NAI
total_success	Αριθμός επιτυχημένων πασών	0,28	NAI	NAI
total_failure	Λόγος αποτυχημένων πασών προς τις συνολικές πάσες	0,22	OXI	NAI
dribbles_failure	Λόγος αποτυχημένων ντριμπλών προς τις συνολικές ντριμπλές	0,27	OXI	NAI
dribbles_success	Αριθμός επιτυχημένων ντριμπλών	0,21	NAI	NAI
badcontrol	Λόγος λανθασμένων κοντρόλ που οδήγησαν σε απώλεια της κατοχής της μπάλας προς την κατοχή της μπάλας	0,14	OXI	NAI
foulwins	Αριθμός κερδισμένων φάουλ	0,19	NAI	NAI
offsides	Αριθμός οφσάιντ	0,09	NAI	NAI
shottotals	Αριθμός σουτ	0,38	NAI	NAI

ontarget	Λόγος σουτ εκτός εστίας προς σουτ εντός εστίας	0,22	OXI	NAI
goalpost	Αριθμός σουτ που κατέληξαν στο δοκάρι	0,14	NAI	OXI
penaltiestaken	Αριθμός κερδισμένων πέναλτι	0,05	NAI	OXI
shotlocations	Λόγος σουτ εκτός περιοχής προς σουτ εντός μικρής και μεγάλης περιοχής	0,17	OXI	NAI
shotplays	Λόγος σουτ που ξεκίνησαν από στημένη φάση προς τα σουτ στην κανονική ροή του αγώνα	0,11	NAI	NAI
counterattacks	Αριθμός σουτ σε αντεπιθέσεις	0,22	NAI	OXI
shotsmiddlepart	Αριθμός σουτ από το κεντρικό μέρος του γηπέδου	0,21	NAI	NAI
attacksrightpart	Αριθμός επιθέσεων που εκδηλώθηκαν από τα δεξιά	0,1	OXI	NAI
wayofscoring	Λόγος γκολ σε αντεπιθέσεις προς γκολ που ξεκίνησαν από στημένη φάση	0,02	NAI	OXI
shotbodyparts	Λόγος κεφαλιών προς σουτ	0,22	NAI	NAI
goalbodyparts	Λόγος γκολ από κεφαλές προς γκολ με σουτ	0,07	NAI	OXI

Πίνακας 4.2: Βαρύτητα Στατιστικών Στοιχείων

Κεφάλαιο 5

Σχεδιασμός και Υλοποίηση

Αλγορίθμου

5.1 Σχεδιασμός Αλγορίθμου

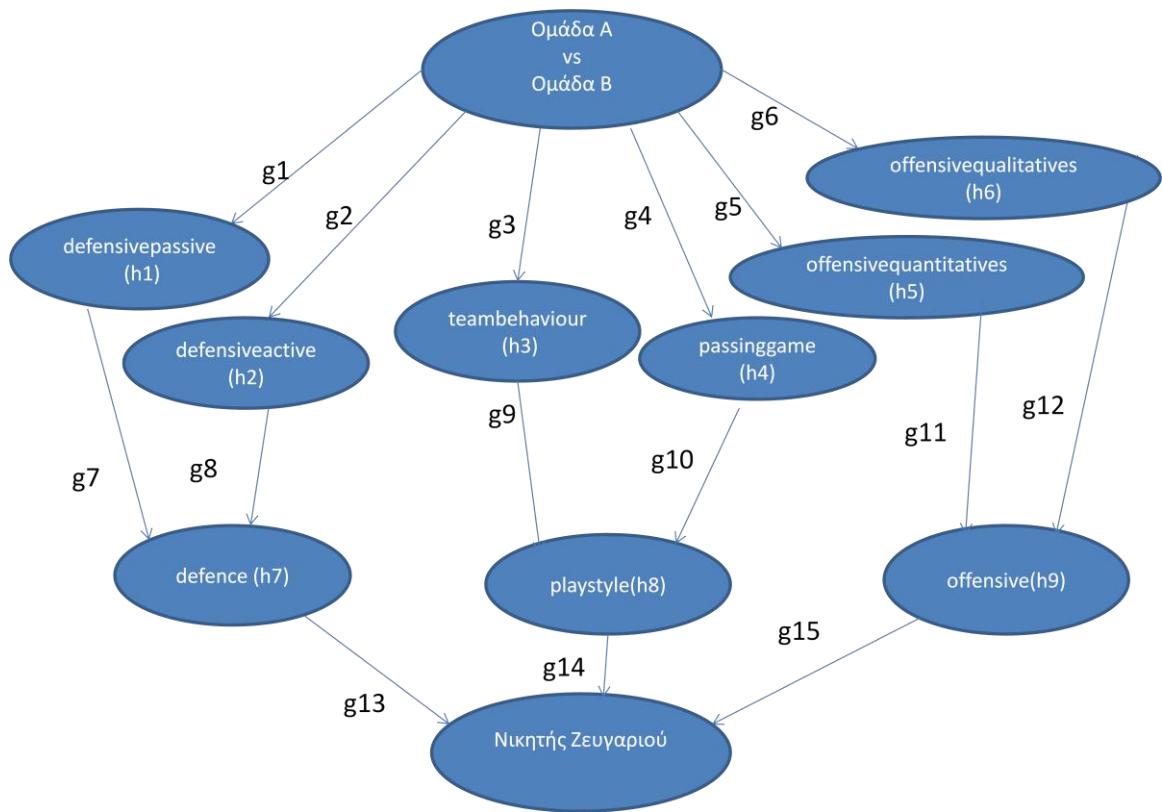
Μετά τη συγκέντρωση των στατιστικών στοιχείων και τον καθορισμό της βαρύτητάς τους, σχεδιάστηκε ο τρόπος με τον οποίο αυτά θα ομαδοποιηθούν, ώστε να σχηματίσουν τους κόμβους του αλγορίθμου που θα χρησιμοποιηθεί για την πρόβλεψη του νικητή μίας ποδοσφαιρικής αναμέτρησης. Τα στατιστικά στοιχεία που παρουσιάστηκαν στο κεφάλαιο 4 ομαδοποιήθηκαν σε 6 μεγάλες κατηγορίες οι οποίες αποτελούν το πρώτο επίπεδο κόμβων του αλγορίθμου. Ο αλγόριθμος αμέσως μετά την αρχική κατάσταση, η οποία ορίζεται ως το ζευγάρι των δύο αντίπαλων ομάδων, θα επισκεφθεί έναν από τους 6 αυτούς κόμβους. Στον πίνακα 5.1 δίνονται οι ονομασίες των 6 αυτών κόμβων και τα στατιστικά από τα οποία αποτελούνται.

Όνομα Κόμβου	Στατιστικά από τα οποία αποτελείται
defensivepassive	shots_against, goals_against, tackles_failure, aerials_lost, gkallowedshotsinside18yardsbox, gkallowedshotsinside6yardsbox, gkallowedshotsoutsidebox
defensiveactive	clearances , corners_conceded , penalties_conceded, shotblocks , crossballs_blocked , interceptions , gksavesinside18yardsbox , gksavesinside6yardsbox
teambehaviour	reds, fairplay, totalcards_foul, dribbles_failure, dribbles_success, badcontrol, fouls_conceded, foulwins
passinggame	ballpossession, ballpossession_opponentsthird, ballpossession_ownthird, keypasses, cross_balls, through_balls, total_failure, total_success
offensivequantitatives	penaltiestaken, goalpost, offsides, corners ,goaltotals, shottotals
offensivequalitatives	goalbodyparts, shotbodyparts, wayofscoring, attacksrightpart, shotsmiddlepart, counterattacks, ontarget, shotlocations, shotplays

Πίνακας 5.1: Κόμβοι Πρώτου Επιπέδου

Το δεύτερο επίπεδο του αλγορίθμου περιλαμβάνει 3 κόμβους τους: defence, playstyle και offensive. Στον κόμβο defence μπορεί να φτάσει ο αλγόριθμος από τους κόμβους defensivepassive και defensiveactive. Στον κόμβο playstyle μπορεί να φτάσει ο αλγόριθμος από τους κόμβους teambehaviour και passinggame και στον κόμβο offensive από τους κόμβους offensivequantitatives και offensivequalitatives.

Στην τελική κατάσταση, που αντιστοιχεί στον νικητή της αναμέτρησης, μπορεί να φτάσει ο αλγόριθμος από έναν εκ των τριών κόμβων του δεύτερου επιπέδου (defence, playstyle, offensive). Η Εικόνα 5.1 παρουσιάζει τους κόμβους του αλγορίθμου και τον τρόπο με τον οποίο συνδέονται μεταξύ τους.



Εικόνα 5.1: Δομή του αλγορίθμου

Οι τιμές g στην Εικόνα 5.1 εκφράζουν το κόστος μετάβασης από τον έναν κόμβο στον άλλο. Οι τιμές h εκφράζουν την εκτίμηση του κόστους για τη μετάβαση από τον τρέχοντα κόμβο μέχρι την τελική κατάσταση. Ο τρόπος υπολογισμού των τιμών g και h για την εκδοχή του αλγορίθμου A^* που αναπτύχθηκε αποτελούν το αντικείμενο της επόμενης παραγράφου 5.2.

5.2 Υπολογισμός κόστους μονοπατιών

Για τον υπολογισμό του κόστους της διαδρομής από την αρχική κατάσταση σε κάθε έναν από τους 6 κόμβους του πρώτου επιπέδου θα πρέπει να συγκριθούν οι δύο ομάδες του εξεταζόμενου ζευγαριού σε κάθε ένα από τα στατιστικά που συνθέτουν τον κάθε κόμβο. Θα υπολογιστεί άρα για κάθε στατιστικό ο συντελεστής υπεροχής V της μίας ομάδας σε σχέση με την άλλη. Το αλγεβρικό άθροισμα των συντελεστών υπεροχής όλων των στατιστικών που συνθέτουν έναν κόμβο εκφράζει την υπεροχή της μίας ομάδας έναντι της άλλης στην ομάδα αυτή των στατιστικών. Προφανώς, όσο μεγαλύτερη είναι η τιμή του συντελεστή υπεροχής τόσο περισσότερο υπερέρχει η μία ομάδα έναντι της άλλης.

Αναλυτικά οι υπολογισμοί που εκτελούνται για να υπολογιστούν οι ομαδοποιημένοι συντελεστές υπεροχής είναι οι εξής:

Ομαδοποιημένος συντελεστής υπεροχής παθητικών στατιστικών άμυνας:

$$\begin{aligned} \mathbf{Vdefensivepassive} = & \mathbf{Vshots_against} + \mathbf{Vgoals_against} + \mathbf{Vtackles_failure} + \mathbf{Vaerials_lost} \\ & + \mathbf{Vgkallowedshotsinside18yardsbox} + \mathbf{Vgkallowedshotsinside6yardsbox} \\ & + \mathbf{Vgkallowedshotsoutsidebox} \end{aligned}$$

Όπου

$$\mathbf{Vshots_against} = \frac{\mathbf{shots_against2} - \mathbf{shots_against1}}{\mathbf{shots_against1} + \mathbf{shots_against2}} * \mathbf{Bshots_against}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός σουτ που δέχτηκε η ομάδα

$$\mathbf{Vgoals_against} = \frac{\mathbf{goals_against2} - \mathbf{goals_against1}}{\mathbf{goals_against1} + \mathbf{goals_against2}} * \mathbf{Bgoals_against}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός γκολ που δέχτηκε η ομάδα

$$\mathbf{Vtackles_failure} = \frac{\mathbf{tackles_failure2} - \mathbf{tackles_failure1}}{\mathbf{tackles_failure1} + \mathbf{tackles_failure2}} * \mathbf{Btackles_failure}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος των αποτυχημένων τάκλιν (μαρκαρίσματα) προς τα συνολικά τάκλιν της ομάδας

$$\mathbf{Vaerials_lost} = \frac{\mathbf{aerials_lost2} - \mathbf{aerials_lost1}}{\mathbf{aerials_lost1} + \mathbf{aerials_lost2}} * \mathbf{Baerials_lost}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος των χαμένων εναέριων μονομαχιών που είχε η ομάδα προς τις κερδισμένες

$$\mathbf{Vgkallowedshotsinside6yardsbox} = \frac{\mathbf{gkallowedshotsinside6yardsbox2} - \mathbf{gkallowedshotsinside6yardsbox1}}{\mathbf{gkallowedshotsinside6yardsbox1} + \mathbf{gkallowedshotsinside6yardsbox2}} *$$

$$\mathbf{Bgkallowedshotsinside6yardsbox}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός σουτ που δέχτηκε η ομάδα εντός της μικρής περιοχής

$$Vgkallowedshotsinside18yardsbox = \frac{gkallowedshotsinside18yardsbox2 - gkallowedshotsinside18yardsbox1}{gkallowedshotsinside18yardsbox1 + gkallowedshotsinside18yardsbox2} *$$

Bgkallowedshotsinside18yardsbox

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός σουτ που δέχτηκε η ομάδα εντός της μεγάλης περιοχής

$$Vgkallowedshotsoutsidebox = \frac{gkallowedshotsoutsidebox2 - gkallowedshotsoutsidebox1}{gkallowedshotsoutsidebox1 + gkallowedshotsoutsidebox2} *$$

Bgkallowedshotsoutsidebox

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος των σουτ που δέχτηκε η ομάδα εκτός περιοχής προς τα σουτ που δέχτηκε εντός της μεγάλης και της μικρής περιοχής

Ομαδοποιημένος συντελεστής υπεροχής ενεργητικών στατιστικών άμυνας:

Vdefensiveactive = *Vclearances* + *Vcorners_conceded* + *Vpenalties_conceded* + *Vshotblocks* + *Vcrossballs_blocked* + *Vinterceptions* + *Vgksavesinside18yardsbox* + *Vgksavesinside6yardsbox*

Όπου

$$Vclearances = \frac{clearances2 - clearances1}{clearances1 + clearances2} * Bclearances$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός αποσοβήσεων που πέτυχε η ομάδα

$$Vcorners_conceded = \frac{corners_conceded2 - corners_conceded1}{corners_conceded1 + corners_conceded2} * Bcorners_conceded$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός κόρνερ που παραχώρησε η ομάδα

$$Vpenalties_conceded = \frac{penalties_conceded2 - penalties_conceded1}{penalties_conceded1 + penalties_conceded2} * Bpenalties_conceded$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός πέναλτι που παραχώρησε η ομάδα

$$Vshotblocks = \frac{shotblocks2 - shotblocks1}{shotblocks1 + shotblocks2} * Bshotblocks$$

$$V_{crossballs_blocked} = \frac{crossballs_blocked2 - crossballs_blocked1}{crossballs_blocked1 + crossballs_blocked2} * B_{crossballs_blocked}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός γεμισμάτων που μπλοκαρίστηκαν από την ομάδα

$$V_{interceptions} = \frac{interceptions2 - interceptions1}{interceptions1 + interceptions2} * B_{interceptions}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός κλεψιμάτων που είχε η ομάδα

$$V_{gksavesinside6yardsbox} = \frac{gksavesinside6yardsbox2 - gksavesinside6yardsbox1}{gksavesinside6yardsbox1 + gksavesinside6yardsbox2} *$$

$B_{gksavesinside6yardsbox}$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός επεμβάσεων τερματοφύλακα της ομάδας εντός μικρής περιοχής

$$V_{gksavesinside18yardsbox} = \frac{gksavesinside18yardsbox2 - gksavesinside18yardsbox1}{gksavesinside18yardsbox1 + gksavesinside18yardsbox2} *$$

$B_{gksavesinside18yardsbox}$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός επεμβάσεων τερματοφύλακα της ομάδας εντός μεγάλης περιοχής

Ομαδοποιημένος συντελεστής υπεροχής στατιστικών σχετικών με την αγωνιστική συμπεριφορά:

$$V_{teambehaviour} = V_{reds} + V_{fairplay} + V_{totalcards_foul} + V_{dribbles_failure} + V_{dribbles_success} + V_{badcontrol} + V_{fouls_conceded} + V_{foulwins}$$

Όπου

$$V_{reds} = \frac{reds2 - reds1}{reds1 + reds2} * B_{reds}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός κόκκινων καρτών που δέχτηκε η ομάδα

$$V_{fairplay} = \frac{fairplay2 - fairplay1}{fairplay1 + fairplay2} * B_{fairplay}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός καρτών που δέχτηκε η ομάδα για αντιεπαγγελματική συμπεριφορά, θέατρο και άλλους λόγους πλην φάουλ

$$V_{totalcards_foul} = \frac{totalcards_foul2 - totalcards_foul1}{totalcards_foul1 + totalcards_foul2} * B_{totalcards_foul}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός καρτών που δέχτηκε η ομάδα για φάουλ

$$V_{dribbles_failure} = \frac{dribbles_failure2 - dribbles_failure1}{dribbles_failure1 + dribbles_failure2} * B_{dribbles_failure}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος αποτυχημένων ντριμπλών προς τις συνολικές ντριμπλές που επιχείρησε η ομάδα

$$V_{dribbles_success} = \frac{dribbles_success1 - dribbles_success2}{dribbles_success1 + dribbles_success2} * B_{dribbles_success}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός επιτυχημένων ντριμπλών που είχε η ομάδα

$$V_{badcontrol} = \frac{badcontrol2 - badcontrol1}{badcontrol1 + badcontrol2} * B_{badcontrol}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος λανθασμένων κοντρόλ που οδήγησαν σε απώλεια της κατοχής της μπάλας προς τη συνολική κατοχή της μπάλας

$$V_{fouls_conceded} = \frac{fouls_conceded2 - fouls_conceded1}{fouls_conceded1 + fouls_conceded2} * B_{fouls_conceded}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός φάουλ στα οποία υπέπεσε η ομάδα

$$V_{foulwins} = \frac{foulwins1 - foulwins2}{foulwins1 + foulwins2} * B_{foulwins}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός φάουλ που κέρδισε η ομάδα

Ομαδοποιημένος συντελεστής υπεροχής στατιστικών σχετικών με το passing game και την κατοχή της μπάλας:

$$V_{\text{passinggame}} = V_{\text{ballpossession}} + V_{\text{ballpossession_opponentsthand}} + V_{\text{ballpossession_ownhand}} + V_{\text{keypasses}} + V_{\text{cross_balls}} + V_{\text{through_balls}} + V_{\text{total_failure}} + V_{\text{total_success}}$$

Όπου

$$V_{\text{ballpossession}} = \frac{\text{ballpossession1} - \text{ballpossession2}}{\text{ballpossession1} + \text{ballpossession2}} * B_{\text{ballpossession}}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Κατοχή της μπάλας που είχε η ομάδα

$$V_{\text{ballpossession_opponentsthand}} = \frac{\text{ballpossession_opponentsthand1} - \text{ballpossession_opponentsthand2}}{\text{ballpossession_opponentsthand1} + \text{ballpossession_opponentsthand2}} *$$

$B_{\text{ballpossession_opponentsthand}}$

είναι ο συντελεστής υπεροχής του στατιστικού: Ποσοστό παιχνιδιού του παιχνιδιού στο ένα τρίτο του γηπέδου που επιτίθεται η ομάδα

$$V_{\text{ballpossession_ownhand}} = \frac{\text{ballpossession_ownhand2} - \text{ballpossession_ownhand1}}{\text{ballpossession_ownhand1} + \text{ballpossession_ownhand2}} *$$

$B_{\text{ballpossession_ownhand}}$

είναι ο συντελεστής υπεροχής του στατιστικού: Ποσοστό παιχνιδιού του παιχνιδιού στο ένα τρίτο του γηπέδου που αμύνεται η ομάδα

$$V_{\text{keypasses}} = \frac{\text{keypasses1} - \text{keypasses2}}{\text{keypasses1} + \text{keypasses2}} * B_{\text{keypasses}}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός πασών-κλειδί που είχε η ομάδα (πάσες που οδηγούν σε μεγάλη ευκαιρία για γκολ)

$$V_{\text{cross_balls}} = \frac{\text{cross_balls1} - \text{cross_balls2}}{\text{cross_balls1} + \text{cross_balls2}} * B_{\text{cross_balls}}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός γεμισμάτων που είχε η ομάδα

$$V_{\text{through_balls}} = \frac{\text{through_balls1} - \text{through_balls2}}{\text{through_balls1} + \text{through_balls2}} * B_{\text{through_balls}}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός πασών σε παίκτη που φεύγει στον κενό χώρο

$$V_{total_failure} = \frac{total_failure2 - total_failure1}{total_failure1 + total_failure2} * B_{total_failure}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος αποτυχημένων πασών που είχε η ομάδα προς τις συνολικές πάσες

$$V_{total_success} = \frac{total_success1 - total_success2}{total_success1 + total_success2} * B_{total_success}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός επιτυχημένων πασών που είχε η ομάδα

Ομαδοποιημένος συντελεστής υπεροχής ποσοτικών στατιστικών επίθεσης:

$$V_{offensivequantitatives} = V_{penaltiestaken} + V_{goalpost} + V_{offsides} + V_{corners} + V_{goaltotals} + V_{shottotals}$$

Όπου

$$V_{penaltiestaken} = \frac{penaltiestaken1 - penaltiestaken2}{penaltiestaken1 + penaltiestaken2} * B_{penaltiestaken}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός κερδισμένων πέναλτι που είχε η ομάδα

$$V_{goalpost} = \frac{goalpost1 - goalpost2}{goalpost1 + goalpost2} * B_{goalpost}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός σουτ της ομάδας που κατέληξαν στο δοκάρι

$$V_{offsides} = \frac{offsides1 - offsides2}{offsides1 + offsides2} * B_{offsides}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός φορών που η ομάδα υποδείχθηκε σε θέση οφσάιντ

$$V_{corners} = \frac{corners1 - corners2}{corners1 + corners2} * B_{corners}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός κόρνερ που κέρδισε η ομάδα

$$V_{goaltotals} = \frac{goaltotals1 - goaltotals2}{goaltotals1 + goaltotals2} * B_{goaltotals}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός γκολ που πέτυχε η ομάδα

$$V_{shottotals} = \frac{shottotals1 - shottotals2}{shottotals1 + shottotals2} * shottotals$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός σουτ που επιχείρησε η ομάδα

Ομαδοποιημένος συντελεστής υπεροχής ποιοτικών στατιστικών επίθεσης:

$$V_{offensivequalitatives} = V_{goalbodyparts} + V_{shotbodyparts} + V_{wayofscoring} + V_{attacksrightpart} + V_{shotsmiddlepart} + V_{counterattacks} + V_{ontarget} + V_{shotlocations} + V_{shotplays}$$

Όπου

$$V_{goalbodyparts} = \frac{goalbodyparts1 - goalbodyparts2}{goalbodyparts1 + goalbodyparts2} * B_{goalbodyparts}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος των γκολ από κεφαλιές προς τα γκολ με σουτ που πέτυχε η ομάδα

$$V_{shotbodyparts} = \frac{shotbodyparts1 - shotbodyparts2}{shotbodyparts1 + shotbodyparts2} * B_{shotbodyparts}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος των κεφαλιών προς τα σουτ που επιχείρησε η ομάδα

$$V_{wayofscoring} = \frac{wayofscoring1 - wayofscoring2}{wayofscoring1 + wayofscoring2} * B_{wayofscoring}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος των γκολ της ομάδας σε αντεπιθέσεις προς τα γκολ που ξεκίνησαν από στημένη φάση

$$V_{attacksrightpart} = \frac{attacksrightpart2 - attacksrightpart1}{attacksrightpart1 + attacksrightpart2} * B_{attacksrightpart}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός επιθέσεων που εκδήλωσε η ομάδα από τα δεξιά

$$V_{shotsmiddlepart} = \frac{shotsmiddlepart1 - shotsmiddlepart2}{shotsmiddlepart1 + shotsmiddlepart2} * B_{shotsmiddlepart}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός σουτ που επιχείρησε η ομάδα από το κεντρικό μέρος του γηπέδου

$$V_{counterattacks} = \frac{counterattacks1 - counterattacks2}{counterattacks1 + counterattacks2} * B_{counterattacks}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Αριθμός σουτ που επιχείρησε η ομάδα σε αντεπιθέσεις

$$V_{ontarget} = \frac{ontarget2 - ontarget1}{ontarget1 + ontarget1} * B_{ontarget}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος σουτ της ομάδας που κατέληξαν εκτός εστίας προς τα σουτ εντός εστίας

$$V_{shotlocations} = \frac{shotlocations2 - shotlocations1}{shotlocations1 + shotlocations2} * B_{shotlocations}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος σουτ που επιχείρησε η ομάδα εκτός περιοχής προς τα σουτ εντός μικρής και μεγάλης περιοχής

$$V_{shotplays} = \frac{shotplays1 - shotplays2}{shotplays1 + shotplays2} * B_{shotplays}$$

είναι ο συντελεστής υπεροχής του στατιστικού: Λόγος σουτ της ομάδας που ξεκίνησαν από στημένη φάση προς τα σουτ στην κανονική ροή του αγώνα

Σε όλους τους παραπάνω υπολογισμούς, για κάθε στατιστικό i οι τιμές $i1$ και $i2$ εκφράζουν τις επιδόσεις της ομάδας 1 και της ομάδας 2 στο συγκεκριμένο στατιστικό στους αγώνες που έχουν ήδη δώσει και έχουν καταγραφεί στη βάση δεδομένων. Η τιμή B_i εκφράζει τη βαρύτητα του συγκεκριμένου στατιστικού.

Μία θετική τιμή ενός συντελεστή υπεροχής V_i δείχνει την υπεροχή της ομάδας 1 και μία αρνητική τιμή την υπεροχή της ομάδας 2. Για το λόγο αυτό, σε περιπτώσεις στατιστικών στις οποίες μία ομάδα υπερέχει όταν έχει επίδοση με μικρότερη τιμή από την αντίπαλη ομάδα (π.χ κόρνερ που παραχώρησε), ο αριθμητής στον παραπάνω τύπο γράφεται i_2-i_1 αντί για i_1-i_2 , ώστε το πρόσημο του συντελεστή υπεροχής να δηλώνει πάντα σωστά την ομάδα που υπερέχει.

Επειδή εξ'ορισμού ο αλγόριθμος A^* επισκέπτεται αρχικά τον κόμβο με τη μικρότερη ευριστική τιμή, ορίζεται ως κόστος μίας διαδρομής του πρώτου επιπέδου η απόλυτη τιμή του αντίστροφου του αντίστοιχου συντελεστή υπεροχής.

Τα κόστη $g_1, g_2, g_3, g_4, g_5, g_6$ του πρώτου επιπέδου (βλ. Εικόνα 5.1) υπολογίζονται συνεπώς ως εξής:

$$g_1 = |1/V_{\text{defensivepassive}}|$$

$$g_2 = |1/V_{\text{defensiveactive}}|$$

$$g_3 = |1/V_{\text{teambehaviour}}|$$

$$g_4 = |1/V_{\text{passinggame}}|$$

$$g_5 = |1/V_{\text{offensivequantitatives}}|$$

$$g_6 = |1/V_{\text{offensivequalitatives}}|$$

Για να υπολογιστούν τα κόστη $g_7, g_8, g_9, g_{10}, g_{11}, g_{12}$ του δεύτερου επιπέδου υπολογίζονται πρώτα οι βαρύτητες των ομαδοποιημένων στατιστικών.

$$\begin{aligned} \mathbf{B}_{\text{defensivepassive}} = & B_{\text{shots_against}} + B_{\text{goals_against}} + B_{\text{tackles_failure}} + B_{\text{aerials_lost}} \\ & + B_{\text{gkallowedshotsinside18yardsbox}} + B_{\text{gkallowedshotsinside6yardsbox}} \\ & + B_{\text{gkallowedshotsoutsidebox}} \end{aligned}$$

$$\begin{aligned} \mathbf{B}_{\text{defensiveactive}} = & B_{\text{clearances}} + B_{\text{corners_conceded}} + B_{\text{penalties_conceded}} + B_{\text{shotblocks}} + \\ & B_{\text{crossballs_blocked}} + B_{\text{interceptions}} + B_{\text{gksavesinside18yardsbox}} + B_{\text{gksavesinside6yardsbox}} \end{aligned}$$

Bteambehaviour=Breds+Bfairplay+Btotalcards_foul+Bdribbles_failure+Bdribbles_success+Bbadcontrol+Bfouls_conceded+Bfoulwins

Bpassinggame=Bballpossession+Bballpossession_opponentsthird+Bballpossession_ownthird+Bkeypasses+Bcross_balls+Bthrough_balls+Btotal_failure+Btotal_success

Boffensivequantitatives=Bpenaltiestaken+Bgoalpost+Boffsides+Bcorners+Bgoaltotals+Bshottotals

Boffensivequalitatives=Bgoalbodyparts+Bshotbodyparts+Bwayofscoring+Battacksrightpart+Bshotsmiddlepart+ Bcounterattacks+Bontarget+Bshotlocations+Bshotplays

Η βαρύτητα των ομαδοποιημένων στατιστικών που συνθέτουν τον κόμβο defence είναι:

Bdefence=Bdefensivepassive+Bdefensiveactive

Ο λόγος Bdefensivepassive/Bdefence δείχνει τη σπουδαιότητα των στατιστικών που συνθέτουν τον κόμβο defensivepassive σε σχέση με τα συνολικά στατιστικά της άμυνας. Αντίστοιχα, ο λόγος Bdefensiveactive/Bdefence δείχνει τη σπουδαιότητα των στατιστικών που συνθέτουν τον κόμβο defensiveactive σε σχέση με τα συνολικά στατιστικά της άμυνας. Παρόμοια με όσα αναφέρθηκαν προηγουμένως για τον αλγόριθμο A* , τα κόστη g7 και g8 προκύπτουν ως το αντίστροφο των παραπάνω λόγων αντίστοιχα, δηλαδή:

$g7 = Bdefence / Bdefensivepassive$

$g8 = Bdefence / Bdefensiveactive$

Η βαρύτητα των ομαδοποιημένων στατιστικών που συνθέτουν τον κόμβο playstyle είναι:

Bplaystyle=Bteambehaviour+Bpassinggame

Ο λόγος Bteambehaviour/Bplaystyle δείχνει τη σπουδαιότητα των στατιστικών που συνθέτουν τον κόμβο teambehaviour σε σχέση με τα συνολικά στατιστικά του τρόπου παιχνιδιού. Αντίστοιχα, ο λόγος Bpassinggame/Bplaystyle δείχνει τη σπουδαιότητα των στατιστικών που συνθέτουν τον κόμβο passinggame σε σχέση με τα συνολικά στατιστικά του τρόπου παιχνιδιού. Τα κόστη g9 και g10 προκύπτουν ως το αντίστροφο των παραπάνω λόγων αντίστοιχα, δηλαδή:

$$g9 = \text{Bplaystyle} / \text{Bteambbehaviour}$$

$$g10 = \text{Bplaystyle} / \text{Bpassinggame}$$

Τέλος, η βαρύτητα των ομαδοποιημένων στατιστικών που συνθέτουν τον κόμβο offensive είναι:

$$\mathbf{Boffensive} = \text{Boffensivequantitatives} + \text{Boffensivequalitatives}$$

Ο λόγος $\text{Boffensivequantitatives} / \text{Boffensive}$ δείχνει τη σπουδαιότητα των στατιστικών που συνθέτουν τον κόμβο $\text{offensivequantitatives}$ σε σχέση με τα συνολικά στατιστικά της επίθεσης. Αντίστοιχα, ο λόγος $\text{Boffensivequalitatives} / \text{Boffensive}$ δείχνει τη σπουδαιότητα των στατιστικών που συνθέτουν τον κόμβο $\text{offensivequalitatives}$ σε σχέση με τα συνολικά στατιστικά της επίθεσης. Τα κόστη $g11$ και $g12$ προκύπτουν ως το αντίστροφο των παραπάνω λόγων αντίστοιχα, δηλαδή:

$$g11 = \text{Boffensive} / \text{Boffensivequantitatives}$$

$$g12 = \text{Boffensive} / \text{offensivequalitatives}$$

Το άθροισμα των συντελεστών βαρύτητας όλων των στατιστικών είναι:

$$\mathbf{Btotal} = \text{Bdefence} + \text{Bplaystyle} + \text{Boffensive}$$

Παρόμοια με το προηγούμενο επίπεδο τα κόστη $g13$, $g14$, $g15$ (βλ. Εικόνα 5.1) προκύπτουν ως οι λογοί:

$$g13 = \text{Btotal} / \text{Bdefence}$$

$$g14 = \text{Btotal} / \text{Bplaystyle}$$

$$g15 = \text{Btotal} / \text{Boffensive}$$

Οι τιμές $h1$, $h2$, $h3$, $h4$, $h5$, $h6$, $h7$, $h8$, $h9$ (βλ. Εικόνα 5.1) εκφράζουν μία εκτίμηση της απόστασης από την τελική κατάσταση και υπολογίζονται ως:

$$h1 = g7 + g13$$

$$h2= g8+g13$$

$$h3= g9+g14$$

$$h4= g10+g14$$

$$h5= g11+g15$$

$$h6= g12+g15$$

$$h7=g13$$

$$h8=g14$$

$$h9=g15$$

5.3 Υλοποίηση και Αποτελέσματα Αλγορίθμου

Η υλοποίηση του αλγορίθμου έγινε με χρήση του πακέτου λογισμικού Microsoft Visual Studio 2012 και της γλώσσας προγραμματισμού C#. Χρησιμοποιήθηκε η μεθοδολογία σταδιακής ανάπτυξης (Incremental Build Model) κατά την οποία το μοντέλο σχεδιάζεται, υλοποιείται και ελέγχεται κάθε φορά αυξητικά, έως ότου ικανοποιηθούν όλες οι απαιτήσεις. Αρχικά ο αλγόριθμος υλοποιήθηκε ώστε να υπολογίζει τον πρώτο κόμβο που επισκέπτεται. Αφού ελέγχθηκε η ορθή λειτουργία του, επεκτάθηκε ώστε να προβλέπει την ομάδα που θα επικρατήσει και τη διαδρομή με το ελάχιστο κόστος. Τέλος, προστέθηκε η λειτουργικότητα του εντοπισμού πιθανού εναλλακτικού μονοπατιού, προφανώς με μεγαλύτερο κόστος, μέσω του οποίου προκύπτει επικράτηση της άλλης ομάδας.

Η είσοδος του προγράμματος είναι το ζευγάρι των ομάδων που θα τεθεί αντιμέτωπο. Για την εκτέλεση των απαραίτητων υπολογισμών έχουν συμπεριληφθεί στον πηγαίο κώδικα όλα τα απαιτούμενα ερωτήματα στη βάση που αντλούν τα κατάλληλα δεδομένα για τον υπολογισμό της βαρύτητας του κάθε στατιστικού και για την επίδοση της κάθε ομάδας σε κάθε στατιστική κατηγορία. Ο αλγόριθμος παράγει τα αποτελέσματα αφού πρώτα υπολογίσει τα κόστη όλων των διαδρομών με τον τρόπο που περιγράφηκε στην παράγραφο 5.2.

Ο αλγόριθμος εκτελέστηκε πρώτη φορά μετά τη φάση των ομίλων με σκοπό να προβλέψει το νικητή στις 8 αναμετρήσεις της φάσης των 16. Στον πίνακα 5.2 παρατίθενται τα αποτελέσματα του αλγορίθμου. Συγκεκριμένα, καταγράφεται η πρόβλεψη του νικητή, το κόστος της ελάχιστης διαδρομής και η διαδρομή αυτή, το κόστος της εναλλακτικής διαδρομής (εάν υπάρχει), μέσω της οποίας προκύπτει επικράτηση της άλλης ομάδας και ποια είναι η εναλλακτική αυτή διαδρομή. Σε περίπτωση αποτυχίας εντοπισμού εναλλακτικής διαδρομής, ο αλγόριθμος πρακτικά δεν εντοπίζει τρόπο επικράτησης της υποδεέστερης ομάδας. Τέλος, σημειώνεται ο πραγματικός νικητής της αναμέτρησης, ώστε να είναι δυνατή η σύγκριση με τα αποτελέσματα της πρόβλεψης.

Αγώνας	Πρόβλεψη Νικητή	Κόστος Διαδρομής	Διαδρομή	Εναλλακτική Διαδρομή (επικράτηση της άλλης ομάδας)	Κόστος Εναλλακτικής Διαδρομής	Πραγματικός Νικητής
Ρεάλ-Σάλκε	Ρεάλ	6,20	defensivepassive→defence→nikitis	-	-	Ρεάλ
Άρσεναλ-Μονακό	Μονακό	6,42	defensivepassive→defence→nikitis	offensivequantitatives→offensive→nikitis	7,44	Μονακό
Λεβερκούζεν-Ατλέτικο Μ.	Ατλέτικο Μ.	9,32	offensivequantitatives→offensive→nikitis	defensiveactive→defence→nikitis	30,73	Ατλέτικο Μ.
Μπάγερν-Σαχτάρ	Μπάγερν	6,15	passinggame→playstyle→nikitis	-	-	Μπάγερν
Πόρτο-Βασιλεία	Πόρτο	6,28	defensivepassive→defence→nikitis	-	-	Πόρτο
Ντόρτμουντ-Γιουβέντους	Ντόρτμουντ	8,39	teambehaviour→playstyle→nikitis	passinggame→playstyle→nikitis	17,96	Γιουβέντους
Τσέλισι-Παρί ΣΖ	Τσέλισι	7,51	defensivepassive→defence→nikitis	defensiveactive→defence→nikitis	18,83	Παρί ΣΖ
Μπαρτσελόνα-Μάντσεστερ Σ.	Μπαρτσελόνα	7,56	passinggame→playstyle→nikitis	offensivequalitatives→offensive→nikitis	18,35	Μπαρτσελόνα

Πίνακας 5.2: Εκτέλεση του αλγορίθμου στη φάση των 16

Από την παράθεση των αποτελεσμάτων του Πίνακα 5.2 προκύπτει ότι ο αλγόριθμος προέβλεψε ορθά τον τελικό νικητή στις 6 από τις 8 αναμετρήσεις (επιτυχία 75%). Ωστόσο και στις δύο

αναμετρήσεις που απέτυχε να προβλέψει σωστά τον τελικό νικητή, κατάφερε να υπολογίσει διαδρομή με μεγαλύτερο κόστος, μέσω της οποίας προκύπτει επικράτηση της ομάδας που τελικά προκρίθηκε από το ζευγάρι. Αξιοσημείωτο είναι επίσης το γεγονός ότι από τα 8 παιχνίδια, η διαδρομή με το μεγαλύτερο κόστος (9,32) υπολογίστηκε στην αναμέτρηση που αποδείχτηκε η πιο αμφίροπη (Λεβερκούζεν-Ατλέτικο Μ.) και κρίθηκε στη διαδικασία των πέναλτι.

Ακολουθώντας, και αφού ενσωματώθηκαν στη βάση δεδομένων και τα στατιστικά όλων των αγώνων της φάσης των 16, ο αλγόριθμος εκτελέστηκε για να προβλέψει το νικητή στις τέσσερις αναμετρήσεις της προημιτελικής φάσης. Τα αποτελέσματα της δεύτερης εκτέλεσης του αλγορίθμου παρουσιάζονται στον Πίνακα 5.3.

Αγώνας	Πρόβλεψη Νικητή	Κόστος Διαδρομής	Διαδρομή	Εναλλακτική Διαδρομή (επικράτηση της άλλης ομάδας)	Κόστος Εναλλακτικής Διαδρομής	Πραγματικός Νικητής
Ρεάλ-Ατλέτικο Μ.	Ρεάλ	6,66	passinggame →playstyle→ nikitis	defensivepassive→def ence→nikitis	8,77	Ρεάλ
Μπαρτσελόνα-Παρί ΣΖ	Μπαρτσελόνα	6,8	passinggame →playstyle→ nikitis	offensivequalitatives →offensive→nikitis	8,77	Μπαρτσελόνα
Μπάγερν-Πόρτο	Μπάγερν	6,09	passinggame →playstyle→ nikitis	defensivepassive→def ence→nikitis	16,37	Μπάγερν
Γιουβέντους-Μονάκο	Γιουβέντους	6,9	passinggame →playstyle→ nikitis	-	-	Γιουβέντους

Πίνακας 5.3: Εκτέλεση του αλγορίθμου στην προημιτελική φάση

Όπως προκύπτει από τα αποτελέσματα, ο αλγόριθμος κατάφερε να προβλέψει σωστά τον τελικό νικητή και στις τέσσερις αναμετρήσεις, σημείωσε δηλαδή επιτυχία 100%.

Αφού ενσωματώθηκαν και τα στατιστικά των αγώνων της προημιτελικής φάσης στη βάση δεδομένων, ο αλγόριθμος εκτελέστηκε εκ νέου προκειμένου να προβλέψει τους νικητές των δύο ημιτελικών. Τα αποτελέσματα παρουσιάζονται στον πίνακα 5.4.

Αγώνας	Πρόβλεψη Νικητή	Κόστος Διαδρομής	Διαδρομή	Εναλλακτική Διαδρομή (επικράτηση της άλλης ομάδας)	Κόστος Εναλλακτικής Διαδρομής	Πραγματικός Νικητής
Ρεάλ-Γιουβέντους	Ρεάλ	7,65	passinggame →playstyle→ nikitis	defensivepassive→def ence→nikitis	11,64	Γιουβέντους
Μπαρτσελόνα-Μπάγερν	Μπαρτσελόνα	10,41	teambehavior→playstyle →nikitis	defensivepassive→def ence→nikitis	13,41	Μπαρτσελόνα

Πίνακας 5.4: Εκτέλεση του αλγορίθμου στην ημιτελική φάση

Ο αλγόριθμος προέβλεψε σωστά την έκβαση της αναμέτρησης Μπαρτσελόνα-Μπάγερν, αλλά όχι και αυτήν της Ρεάλ με τη Γιουβέντους. Συνολικά και στις τρεις φάσεις, ο αλγόριθμος προέβλεψε σωστά την ομάδα που επικράτησε σε 11 από τις 14 συνολικά αναμετρήσεις (ποσοστό επιτυχίας 79%).

Παρακάτω παρατίθεται η πρόβλεψη του αλγορίθμου για τον τελικό της 6ης Ιουνίου, έχοντας συνηπολογίσει και τις επιδόσεις των ομάδων στην ημιτελική φάση.

Αγώνας	Πρόβλεψη Νικητή	Κόστος Διαδρομής	Διαδρομή	Εναλλακτική Διαδρομή (επικράτηση της άλλης ομάδας)	Κόστος Εναλλακτικής Διαδρομής	Πραγματικός Νικητής
Μπαρτσελόνα-Γιουβέντους	Μπαρτσελόνα	7,48	passinggame →playstyle→ nikitis	defensivepassive→def ence→nikitis	13,43	-

Πίνακας 5.5: Εκτέλεση του αλγορίθμου για πρόβλεψη του τελικού νικητή

Κεφάλαιο 6

Βελτίωση Αλγορίθμου

6.1 Εισαγωγή Νέων Μετρικών

Στην προσπάθεια περαιτέρω βελτίωσης των αποτελεσμάτων του αλγορίθμου αναζητήθηκαν παράγοντες που ενδεχομένως επηρεάζουν την έκβαση του αποτελέσματος μίας ποδοσφαιρικής αναμέτρησης, αλλά δεν αποτυπώνονται στους πίνακες των στατιστικών. Ένας από τους παράγοντες αυτούς είναι η δύναμη της έδρας, ο οποίος όμως προτιμήθηκε να μη ληφθεί τελικά υπόψη, δεδομένου ότι αφενός ο τελικός της διοργάνωσης διεξάγεται σε ουδέτερο γήπεδο, αφετέρου όλες οι διπλές νοκ άουτ αναμετρήσεις αντιμετωπίζονται στο πλαίσιο της μεταπτυχιακής αυτής διατριβής ως ένας αγώνας με την έννοια ότι ο αλγόριθμος ενδιαφέρεται για την ομάδα που θα προκριθεί.

Ένας άλλος παράγοντας που φαίνεται να επηρεάζει την έκβαση κρίσιμων ποδοσφαιρικών αναμετρήσεων είναι η εμπειρία της κάθε ομάδας στη διοργάνωση. Στη διαχείριση ειδικών συνθηκών και καταστάσεων σε ένα σημαντικό αγώνα, κατά κανόνα ανταποκρίνονται καλύτερα ομάδες που έχουν αρκετές φορές στο παρελθόν βρεθεί σε παρόμοια θέση. Ως εκ τούτου, κρίθηκε σκόπιμο να υπολογιστεί για κάθε ομάδα ένας συντελεστής, ενδεικτικός της εμπειρίας της και του «βάρους» του ονόματός της στη διοργάνωση. Προκειμένου να επιτευχθεί ο υπολογισμός αυτός αναζητήθηκαν και συγκεντρώθηκαν σε πίνακα οι επιδόσεις των 16 καλύτερων φετινών ομάδων

στη διοργάνωση από το 1992-93 και έπειτα, οπότε και ξεκίνησε η διοργάνωση Champions League με τη μορφή που έχει μέχρι και σήμερα. Για να μοντελοποιηθεί ο παράγοντας της εμπειρίας εφαρμόστηκε ο κανόνας:

- 1 βαθμός για κάθε σεζόν που η ομάδα έφτασε μέχρι τη φάση των ομίλων
- 2 βαθμοί για κάθε σεζόν που η ομάδα έφτασε μέχρι τη φάση των 16
- 3 βαθμοί για κάθε σεζόν που η ομάδα έφτασε μέχρι τα προημιτελικά
- 4 βαθμοί για κάθε σεζόν που η ομάδα έφτασε μέχρι τα ημιτελικά
- 5 βαθμοί για κάθε σεζόν που η ομάδα ήταν φιναλίστ
- 6 βαθμοί για κάθε κατάκτηση του τροπαίου

Στη συνέχεια οι βαθμοί αυτοί για κάθε ομάδα αθροίστηκαν και ο συντελεστής εμπειρίας κάθε ομάδας ορίστηκε ως το πηλίκο των συνολικών της βαθμών προς τους συνολικούς βαθμούς όλων των ομάδων. Ο Πίνακας 6.1 που ακολουθεί δείχνει το συντελεστή εμπειρίας των ομάδων που προκρίθηκαν τη σεζόν 2014-15 στις 16 καλύτερες της διοργάνωσης, βάσει των παραπάνω υπολογισμών.

Ένας ακόμη παράγοντας που κρίθηκε ότι πρέπει να ληφθεί υπόψη είναι η δυναμικότητα του πρωταθλήματος από το οποίο προέρχονται οι ομάδες. Η συμπλήρωση των πινάκων της βάσης δεδομένων με τα στατιστικά των αγώνων της φετινής διοργάνωσης και η πρόβλεψη βάσει της επεξεργασίας τους εγκυμονεί τον κίνδυνο υπέρμετρης προμοδότησης ομάδων που είχαν την τύχη να έχουν μία ευνοϊκή κλήρωση, ενώ στην πραγματικότητα δεν είναι και τόσο ισχυρές. Αντίθετα, ελλοχεύει ο κίνδυνος να υποτιμηθούν ισχυρές ομάδες, των οποίων τα στατιστικά στους προηγούμενους αγώνες τους ενδεχομένως να μην ήταν ιδιαίτερα εντυπωσιακά, επειδή έτυχε να αντιμετωπίσουν εξίσου ή και περισσότερο ισχυρούς αντιπάλους. Το πρόβλημα αυτό επιχειρεί να διορθώσει, σε συνδυασμό και με το συντελεστή εμπειρίας, ένας ακόμα συντελεστής που εκτιμά τη δυναμικότητα του πρωταθλήματος της χώρας από την οποία προέρχεται η κάθε ομάδα. Για τον υπολογισμό του συντελεστή αυτού χρησιμοποιήθηκε η βαθμολογία της UEFA κάθε χώρας σε συλλογικό επίπεδο την τελευταία πενταετία. Ακολουθώντας, ο συντελεστής χώρας υπολογίστηκε ως το πηλίκο του αθροίσματος της βαθμολογίας της χώρας τα τελευταία 5 έτη προς τη συνολική βαθμολογία των 9 χωρών με εκπρόσωπο στη φάση των 16 του Champions League 2014-15.

Ο Πίνακας 6.2 που ακολουθεί δείχνει το συντελεστή κάθε μίας από τις 9 αυτές χώρες, βάσει του παραπάνω υπολογισμού.

Team	Experience Factor
Barcelona	0,138
Real M.	0,143
Atletico M.	0,024
Arsenal	0,077
Chelsea	0,092
Manchester City	0,009
Juventus	0,090
Bayern M.	0,134
Dortmund	0,053
Leverkusen	0,035
Schalke	0,029
Paris Saint-Germain	0,026
Monaco	0,035
Porto	0,079
Shakhtar Donetsk	0,026
Basel	0,011

Πίνακας 6.1: Συντελεστής Εμπειρίας Ομάδων

Country	Country Factor
Spain	0,170
England	0,143
Germany	0,139
Italy	0,118
Portugal	0,109
France	0,093

Russia	0,089
Ukraine	0,077
Switzerland	0,062

Πίνακας 6.2: Συντελεστής Δυναμικότητας Πρωταθλήματος Χώρας

Οι δύο παραπάνω συντελεστές υπεισέρχονται στους υπολογισμούς της επίδοσης των ομάδων σε κάθε κατηγορία στατιστικών, πολλαπλασιάζοντας την τιμή της επίδοσης στην περίπτωση στατιστικού που μία ομάδα υπερέχει όταν έχει μεγαλύτερη τιμή από την αντίπαλό της και διαιρώντας την τιμή της επίδοσης στην περίπτωση στατιστικού που μία ομάδα υπερέχει όταν έχει μικρότερη τιμή από την αντίπαλό της.

Έτσι για παράδειγμα όταν συγκρίνονται δύο ομάδες στο στατιστικό «κερδισμένα κόρνερ», υπολογίζεται η τιμή

$$(\text{Σύνολο Κόρνερ}) * (1 + \text{Experience_Factor} + \text{Country_Factor})$$

Αντίστοιχα, όταν συγκρίνονται δύο ομάδες στο στατιστικό «φάουλ που υπέπεσαν», υπολογίζεται η τιμή

$$(\text{Σύνολο Φάουλ}) / (1 + \text{Experience_Factor} + \text{Country_Factor})$$

6.2 Εκτέλεση Βελτιωμένου Αλγορίθμου

Αφού υπεισλήθαν στους υπολογισμούς των επιδόσεων των ομάδων οι δύο συντελεστές που αναφέρθηκαν στην παράγραφο 6.1, ο αλγόριθμος εκτελέστηκε εκ νέου για τους αγώνες της φάσης των 16.

Στον πίνακα 6.3 παρατίθενται τα αποτελέσματα της νέας εκτέλεσης του αλγορίθμου. Καταγράφονται εκ νέου η πρόβλεψη του νικητή, το κόστος της ελάχιστης διαδρομής και η διαδρομή αυτή, το κόστος της εναλλακτικής διαδρομής (εάν υπάρχει), μέσω της οποίας προκύπτει επικράτηση της άλλης ομάδας και ποια είναι η εναλλακτική αυτή διαδρομή. Τέλος,

σημειώνεται ο πραγματικός νικητής της αναμέτρησης, ώστε να είναι δυνατή η σύγκριση με τα αποτελέσματα της πρόβλεψης.

Αγώνας	Πρόβλεψη Νικητή	Κόστος Διαδρομής	Διαδρομή	Εναλλακτική Διαδρομή (επικράτηση της άλλης ομάδας)	Κόστος Εναλλακτικής Διαδρομής	Πραγματικός Νικητής
Ρεάλ-Σάλκε	Ρεάλ	5,97	defensivepassive→defence→nikitis	-	-	Ρεάλ
Άρσεναλ-Μονακό	Μονακό	6,69	defensivepassive→defence→nikitis	offensivequantitatives→offensive→nikitis	7,31	Μονακό
Λεβερκούζεν-Ατλέτικο Μ.	Ατλέτικο Μ.	9,17	offensivequantitatives→offensive→nikitis	defensiveactive→defence→nikitis	33,16	Ατλέτικο Μ.
Μπάγερν-Σαχτάρ	Μπάγερν	5,71	passinggame→playstyle→nikitis	-	-	Μπάγερν
Πόρτο-Βασιλεία	Πόρτο	6,03	defensivepassive→defence→nikitis	-	-	Πόρτο
Ντόρτμουντ-Γιουβέντους	Ντόρτμουντ	8,48	teambehaviour→playstyle→nikitis	passinggame→playstyle→nikitis	15,73	Γιουβέντους
Τσέλσι-Παρί ΣΖ	Τσέλσι	6,85	defensivepassive→defence→nikitis	defensiveactive→defence→nikitis	22,62	Παρί ΣΖ
Μπαρτσελόνα-Μάντσεστερ Σ.	Μπαρτσελόνα	6,5	passinggame→playstyle→nikitis	offensivequalitatives→offensive→nikitis	52,08	Μπαρτσελόνα

Πίνακας 6.3: Εκτέλεση του νέου αλγορίθμου στη φάση των 16

Ο αλγόριθμος καταφέρνει και πάλι να προβλέψει σωστά το αποτέλεσμα στις 6 από τις 8 αναμετρήσεις. Ωστόσο, καταφέρνει να υπολογίσει μικρότερο κόστος διαδρομής για τους νικητές των πέντε από τις έξι αυτές αναμετρήσεις, δηλαδή εκφράζει μεγαλύτερη σιγουριά για την πρόβλεψη που επιχειρεί. Θετικό κρίνεται επίσης το γεγονός ότι η έκτη αναμέτρηση, της οποίας την έκβαση προβλέπει ορθά, και στην οποία υπολογίζει μεγαλύτερο κόστος για να φτάσουμε στον τελικό νικητή σε σχέση με πριν είναι η αναμέτρηση Άρσεναλ-Μονακό, η οποία ήταν

εξαιρετικά αμφίροπη. Επιπλέον, ο αλγόριθμος καταφέρνει τώρα να υπολογίσει και ένα μικρότερο κόστος εναλλακτικής διαδρομής (επικράτηση Γιουβέντους) στην αναμέτρηση Ντόρτμουντ-Γιουβέντους, την οποία απέτυχε να εκτιμήσει σωστά ως προς τον τελικό νικητή. Ο αλγόριθμος αποτυγχάνει να εκτιμήσει σωστά την έκβαση του αγώνα Τσέλσι-Παρί Σεν Ζερμέν, ακόμα περισσότερο σε σχέση με την πρωτότυπη έκδοσή του, σε μία αναμέτρηση που κατά γενική ομολογία ωστόσο η επικράτηση της γαλλικής ομάδας αποτέλεσε έκπληξη.

Ακολουθώντας, και αφού ενσωματώθηκαν στη βάση δεδομένων και τα στατιστικά όλων των αγώνων της φάσης των 16, η νέα έκδοση του αλγορίθμου εκτελέστηκε για να προβλέψει το νικητή στις τέσσερις αναμετρήσεις της προημιτελικής φάσης. Τα αποτελέσματα της δεύτερης εκτέλεσης της νέας έκδοσης του αλγορίθμου παρουσιάζονται στον Πίνακα 6.4.

Αγώνας	Πρόβλεψη Νικητή	Κόστος Διαδρομής	Διαδρομή	Εναλλακτική Διαδρομή (επικράτηση της άλλης ομάδας)	Κόστος Εναλλακτικής Διαδρομής	Πραγματικός Νικητής
Ρεάλ-Ατλέτικο Μ.	Ρεάλ	6,19	passinggame →playstyle→ nikitis	defensivepassive→defence→nikitis	10,02	Ρεάλ
Μπαρτσελόνα-Παρί ΣΖ	Μπαρτσελόνα	6,02	passinggame →playstyle→ nikitis	offensivequalitatives →offensive→nikitis	9,46	Μπαρτσελόνα
Μπάγερν-Πόρτο	Μπάγερν	5,87	passinggame →playstyle→ nikitis	defensivepassive→defence→nikitis	46,52	Μπάγερν
Γιουβέντους-Μονάκο	Γιουβέντους	6,45	passinggame →playstyle→ nikitis	-	-	Γιουβέντους

Πίνακας 6.4: Εκτέλεση του νέου αλγορίθμου στην προημιτελική φάση

Όπως προκύπτει από τα αποτελέσματα, ο αλγόριθμος κατάφερε να προβλέψει σωστά τον τελικό νικητή και στις τέσσερις αναμετρήσεις, σημείωσε δηλαδή επιτυχία 100% και μάλιστα υπολογίζοντας μικρότερα κόστη για την ελάχιστη διαδρομή που οδηγεί στο νικητή και μεγαλύτερα κόστη για την εναλλακτική διαδρομή, άρα εκφράζοντας μεγαλύτερη σιγουριά σε σχέση με την αρχική έκδοση του αλγορίθμου.

Στη συνέχεια και αφού ενσωματώθηκαν και τα στατιστικά των αγώνων της προημιτελικής φάσης στη βάση δεδομένων, ο αλγόριθμος εκτελέστηκε εκ νέου προκειμένου να προβλέψει τους νικητές των δύο ημιτελικών. Τα αποτελέσματα παρουσιάζονται στον πίνακα 6.5.

Αγώνας	Πρόβλεψη Νικητή	Κόστος Διαδρομής	Διαδρομή	Εναλλακτική Διαδρομή (επικράτηση της άλλης ομάδας)	Κόστος Εναλλακτικής Διαδρομής	Πραγματικός Νικητής
Ρεάλ-Γιουβέντους	Ρεάλ	6,76	passinggame →playstyle→ nikitis	defensivepassive→def ence→nikitis	21,38	Γιουβέντους
Μπαρτσελόνα-Μπάγερν	Μπαρτσελόνα	9,93	teambehavior→playstyle →nikitis	offensivequalitatives →offensive→nikitis	14,66	Μπαρτσελόνα

Πίνακας 6.5: Εκτέλεση του νέου αλγορίθμου στην ημιτελική φάση

Ο αλγόριθμος προέβλεψε σωστά την έκβαση της αναμέτρησης Μπαρτσελόνα-Μπάγερν, αλλά όχι και αυτήν της Ρεάλ με τη Γιουβέντους. Συνολικά και στις τρεις φάσεις, ο νέος αλγόριθμος προέβλεψε σωστά την ομάδα που επικράτησε σε 11 από τις 14 συνολικά αναμετρήσεις (ποσοστό επιτυχίας 79%).

Παρακάτω παρατίθεται η πρόβλεψη του νέου αλγορίθμου για τον τελικό της 6ης Ιουνίου, έχοντας συνυπολογίσει και τις επιδόσεις των ομάδων στην ημιτελική φάση.

Αγώνας	Πρόβλεψη Νικητή	Κόστος Διαδρομής	Διαδρομή	Εναλλακτική Διαδρομή (επικράτηση της άλλης ομάδας)	Κόστος Εναλλακτικής Διαδρομής	Πραγματικός Νικητής
Μπαρτσελόνα-Γιουβέντους	Μπαρτσελόνα	6,74	passinggame →playstyle→ nikitis	defensivepassive→def ence→nikitis	31,79	-

Πίνακας 6.6: Εκτέλεση του νέου αλγορίθμου για πρόβλεψη του τελικού νικητή

Κεφάλαιο 7

Επίλογος

Στην παρούσα μεταπτυχιακή διατριβή έγινε συλλογή και επεξεργασία ενός μεγάλου όγκου στατιστικών δεδομένων των ποδοσφαιρικών αγώνων της διασυλλογικής διοργάνωσης Champions League 2014-15 με σκοπό την πρόβλεψη του τελικού νικητή. Για την υλοποίηση της μεταπτυχιακής διατριβής επελέγη ο αλγόριθμος τεχνητής νοημοσύνης A-star. Ο πηγαίος κώδικας με τη χρήση κατάλληλων ερωτημάτων στη βάση δεδομένων που υλοποιήθηκε, υπολογίζει τον κάθε κόμβο που επισκέπτεται ο αλγόριθμος μέχρι να φτάσει στην τελική κατάσταση που ισοδυναμεί με την πρόβλεψη του νικητή μίας ποδοσφαιρικής αναμέτρησης. Εκτός από την πρόβλεψη του τελικού νικητή ο αλγόριθμος υπολογίζει ασφαλώς και τη διαδρομή με το ελάχιστο κόστος, καθώς και εάν υπάρχει εναλλακτική διαδρομή, προφανώς με μεγαλύτερο κόστος, μέσω της οποίας να προκύπτει ενδεχομένως επικράτηση της φαινομενικά υποδεέστερης ομάδας..

Ο αλγόριθμος εφαρμόστηκε για πρώτη φορά στους αγώνες της φάσης των 16 και προέβλεψε σωστά τον τελικό νικητή στις 6 από τις 8 αναμετρήσεις. Είχε προηγηθεί η διαδικασία καθορισμού της βαρύτητας κάθε κατηγορίας στατιστικών, βάσει των στοιχείων που είχαν αποθηκευτεί στη βάση δεδομένων, η οποία αρχικά συγκέντρωνε την πληροφορία από τους αγώνες της φάσης των ομίλων. Εν συνεχεία και στην προσπάθεια περαιτέρω βελτίωσης του

αλγόριθμοι διερευνήθηκαν επιπλέον παράγοντες όπως η προϊστορία των ομάδων στη διοργάνωση, καθώς και η δυναμικότητα του πρωταθλήματος από το οποίο προέρχονται. Μοντελοποιήθηκαν έτσι δύο νέες μετρικές, οι οποίες συνδυάστηκαν με τα αποθηκευμένα στατιστικά της βάσης δεδομένων προκειμένου να δώσουν ακριβέστερα αποτελέσματα. Ο αλγόριθμος κατάφερε να προβλέψει σωστά την ομάδα που επικράτησε και στις τέσσερις αναμετρήσεις της προημιτελικής φάσης, ενώ η βάση δεδομένων που χρησιμοποιήθηκε ενσωμάτωσε πλέον και τα στατιστικά στοιχεία από τους αγώνες των ομάδων στη φάση των 16. Τελικός στόχος είναι η χρήση του αλγόριθμου για μία ακριβέστερη τεκμηριωμένη πρόβλεψη του νικητή του τελικού της 6ης Ιουνίου 2015.

Ο αλγόριθμος που αναπτύχθηκε αν και εκτίμησε σε πολύ μεγάλο ποσοστό ορθά την έκβαση των ποδοσφαιρικών αναμετρήσεων στις οποίες δοκιμάστηκε, δεν μπορεί να θεωρηθεί ως ένα απόλυτα ασφαλές εργαλείο πρόβλεψης. Πριν ή κατά τη διάρκεια μίας ποδοσφαιρικής αναμέτρησης συντελούνται αρκετά συχνά γεγονότα, τα οποία είναι εξαιρετικά δύσκολο να μοντελοποιηθούν στα πλαίσια του αλγόριθμου, όπως π.χ οι δυσμενείς καιρικές συνθήκες την ώρα του αγώνα, ένα εξωαγωνιστικό συμβάν που μπορεί να επηρεάσει την ψυχολογία κάποιων παικτών, μία κόκκινη κάρτα στο ξεκίνημα της αναμέτρησης που αλλάζει τις ισορροπίες, μία κακή διατροφή, η αιφνίδια απόλυση ή παραίτηση ενός προπονητή, κλπ. Μία ακόμα δυσκολία που απαντάται είναι η εξαιρετικά χρονοβόρα διαδικασία της συγκέντρωσης των στατιστικών στοιχείων των ποδοσφαιρικών αναμετρήσεων σε μεγάλη λεπτομέρεια. Είναι βέβαιο ότι η ενσωμάτωση στατιστικών στοιχείων από περισσότερους αγώνες των ομάδων, π.χ. από τους αγώνες που δίνουν στα εθνικά τους πρωταθλήματα ή από αγώνες τους στο Champions League τις προηγούμενες χρονιές, θα οδηγούσε σε ακριβέστερους υπολογισμούς και κατ'επέκταση στη δυνατότητα διατύπωσης ασφαλέστερης πρόβλεψης για την έκβαση μίας ποδοσφαιρικής αναμέτρησης. Συνεπώς, ως συνέχιση της μεταπτυχιακής διατριβής στην προσπάθεια περαιτέρω βελτιστοποίησης των αποτελεσμάτων του αλγόριθμου, προτείνεται η ενσωμάτωση στατιστικών στοιχείων από περισσότερους αγώνες των ομάδων, καθώς και ο εντοπισμός επιπλέον κατηγοριών στατιστικών στοιχείων που θα μπορούσαν να χρησιμοποιηθούν στους υπολογισμούς. Τέλος, πρόκληση αποτελεί η επιτυχής προσπάθεια ποσοτικοποίησης κάποιων από τους απρόβλεπτους παραγόντες που προαναφέρθηκαν, ώστε να μπορούν να συνεισφέρουν στους υπολογισμούς με αποδοτικό τρόπο, ώστε να καταστεί εφικτή μία ασφαλέστερη πρόβλεψη του νικητή μίας ποδοσφαιρικής διοργάνωσης.

Βιβλιογραφία

- [01] T. Dean, J. Allen, Y. Aloimonos, "Artificial Intelligence: Theory and Practice", The Benjamin/Cummings Publishing Company, 1995
- [02] FootballManagerEvolution, "Artificial Intelligence", Τελευταία Πρόσβαση στις 10/5/2015: <http://www.footballmanagerevolution.com/en/guide/artificial-intelligence>
- [03] M. Ginsberg, "Essentials of Artificial Intelligence", Morgan Kaufmann, 1993
- [04] G. Luger, W. Stubblefield, "AI Algorithms, Data Structures, and Idioms in Prolog, Lisp, and Java, Addison Wesley, 2009
- [05] T. Mitchell, "Machine Learning", McGraw-Hill, 1997
- [06] C. Nicholson, "Artificial Intelligence Predicts Gambling Behaviour", Τελευταία Πρόσβαση στις 10/5/2015: <http://www.scientificamerican.com/podcast/episode/artificial-intelligence-predicts-ga-09-07-21/>
- [07] N. Nilsson, "Artificial Intelligence: A New Synthesis", Morgan Kaufmann, 1998
- [08] J. Pearl, "Heuristics: Intelligent Search Strategies for Computer Problem Solving", Addison-Wesley, 1984
- [09] E. Rich, K. Knight, "Artificial Intelligence", McGraw-Hill, 2nd edition, 1992
- [10] S. Russel, P. Norvig, "Artificial Intelligence: A Modern Approach", 3rd edition, Prentice Hall, 2009
- [11] Uefa, "Uefa Champions League Match Statistics", Τελευταία Πρόσβαση στις 10/5/2015: <http://www.uefa.com/uefachampionsleague/season=2015/statistics/round=2000548/matches/index.html>

- [12] Whoscored, "Uefa Champions League", Τελευταία Πρόσβαση στις 10/5/2015:
<http://www.whoscored.com/Regions/250/Tournaments/12/Europe-UEFA-Champions-League>
- [13] Whoscored, "Whoscored Ratings Explained", Τελευταία Πρόσβαση στις 10/5/2015:
<http://www.whoscored.com/Explanations>
- [14] Wikipedia, "Artificial Intelligence", Τελευταία Πρόσβαση στις 10/5/2015:
http://en.wikipedia.org/wiki/Football_Manager
- [15] Wikipedia, "A* Search Algorithm ", Τελευταία Πρόσβαση στις 10/5/2015:
http://en.wikipedia.org/wiki/A*_search_algorithm
- [16] Ι. Βλαχάβας, Π. Κεφάλας, Ν. Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου, "Τεχνητή Νοημοσύνη", Γ' Έκδοση, Εκδόσεις Πανεπιστημίου Μακεδονίας, Θεσσαλονίκη, 2011