

Μηχανική Μάθηση & Μεγάλα Δεδομένα στον Επιχειρηματικό Αθλητισμό

Η περίπτωση της Αγγλικής Premier League



Τόγιας Παναγιώτης
Θεολόγος Κωνσταντίνος

14^ο Φοιτητικό Συνέδριο Διοικητικής Επιστήμης & Τεχνολογίας
Οικονομικό Πανεπιστήμιο Αθηνών

14^ο Φοιτητικό Συνέδριο
Διοικητικής Επιστήμης & Τεχνολογίας

Outline



- Ιστορικά στοιχεία Premier League
- Η αναλυτική στο χώρο του ποδοσφαίρου
- Τεχνικές Μηχανικής Μάθησης
- Αποτελέσματα

Ιστορικά στοιχεία Premier League

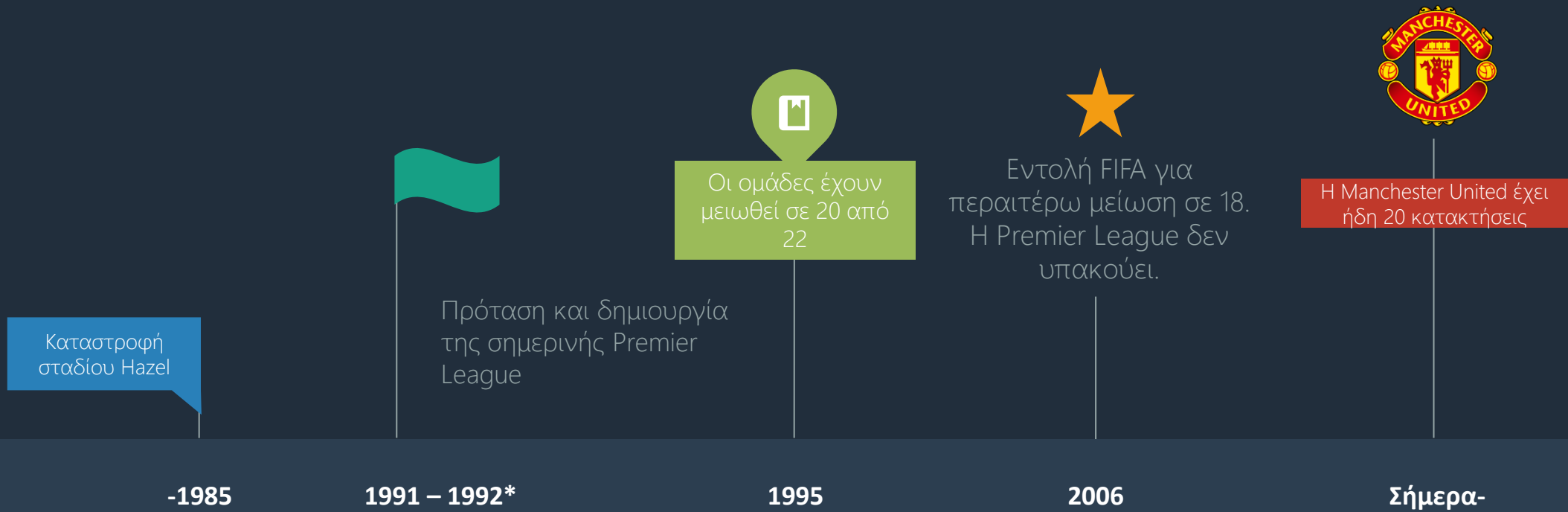


Στο τέλος της αγωνιστικής περιόδου του 1991, υπήρξε η πρόταση για τη δημιουργία ενός νέου πρωταθλήματος με την αιτιολογία ότι θα αποφέρει περισσότερα χρήματα στο παιχνίδι συνολικά. Η ιδρυτική συμφωνία των μελών, η οποία υπεγράφη στις 17 Ιουλίου του 1991 από τους κορυφαίους συλλόγους της κατηγορίας, έθεσε τις βασικές αρχές για τη σύσταση της FA Premier League.



*Ο αριθμός των συλλόγων
μειώθηκε σε 20 το 1995,
όταν τέσσερις ομάδες
είχαν υποβιβαστεί από το
πρωτάθλημα και μόνο δύο
ομάδες προβιβάστηκαν*

Σημαντικότερα Γεγονότα



**Η νεοσυσταθείσα κορυφαία κατηγορία είχε πλέον εμπορική ανεξαρτησία από την Ποδοσφαιρική Ομοσπονδία και τη Football League, καθώς είχε τη δυνατότητα να διαπραγματευτεί το πρόγραμμα τηλεοπτικής κάλυψης των αγώνων και τις συμφωνίες χορηγίας. Το επιχείρημα που διατυπώθηκε κατά την εποχή εκείνη ήταν ότι το επιπλέον εισόδημα θα επέτρεπε τους αγγλικούς συλλόγους να ανταγωνιστούν τις μεγάλες ποδοσφαιρικές ομάδες της Ευρώπης*



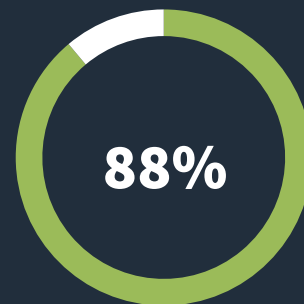
2

Διοργανώσεις



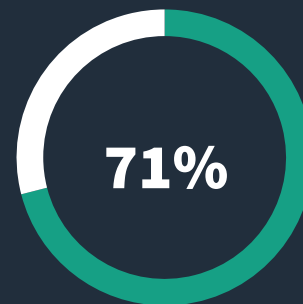
20

Ομάδες



11.7εκατ

Συνολική προσέλευση στα
γήπεδα (Σεζόν 2016/2017)



\$5.136δισ

Προβλεπόμενα κέρδη από
τηλεοπτικά δικαιώματα τη
περίοδο 2016-2019

Ανάλυση

Κατά μέσο όρο, σε κάθε αγώνα της Premier League υπάρχει πληρότητα 88% στις θέσεις των θεατών, ενώ η αύξηση η οποία υπάρχει, στο χρηματικό πόσο από έσοδα τηλεοπτικών δικαιωμάτων, από τη σεζόν 2013 είναι της τάξης του 71%

Η αναλυτική στο χώρο του ποδοσφαίρου

Στο σύγχρονο παιχνίδι η χρήση των δεδομένων έχει να κάνει με την εύρεση της λεπτομέρειας. Αυτού του 1% που μπορεί να εκμεταλλευτούμε για να ανακαλύψουμε τις αδυναμίες του αντιπάλου και κάνει την διαφορά μεταξύ νίκης και ήττας. Είναι κάτι περισσότερο από απλή εφαρμογή δεδομένων σε κομμάτια τακτικής. Η αντικειμενική πληροφόρηση χρησιμοποιείται από όλα τα σωματεία (football clubs) την ενίσχυση της αποτελεσματικότητας και την ανάπτυξη των διαδικασιών που επιτρέπουν στον οργανισμό να είναι όσο το δυνατόν προετοιμασμένος από το γήπεδο μέχρι την αίθουσα συνεδριάσεων.

Η αναλυτική στο χώρο του ποδοσφαίρου



Αξιοποίηση Απόδοσης
Παίκτη



- ♦ **8-10** ψηφιακές κάμερες παρακολουθούν κάθε παίκτη μέσα στον αγωνιστικό χώρο
- ♦ **10** σημεία δεδομένων συλλέγονται κάθε δευτερόλεπτο



- ♦ **1,4** εκατομμύρια σημεία δεδομένων ανά παιχνίδι
- ♦ **3X** πιο αξιόπιστα δεδομένα από άλλες μεθόδους

Οι αναλυτές έπειτα θα αποκωδικοποιήσουν τα δεδομένα για κάθε τάκλιν, σουτ ή πάσα και θα τα προωθήσουν στους προπονητές και τους αναλυτές απόδοσης, προκειμένου να αποκτήσουν γνώση για το τι ακριβώς συνέβη στην διάρκεια του αγώνα με ή χωρίς την μπάλα.

Τεχνικές Μηχανικής Μάθησης



Μοντέλα Ενδυνάμωσης

Προκύπτουν από ένα σύνολο πιο απλών εκτιμητών πρόβλεψης χρησιμοποιώντας πολλαπλά δέντρα κατηγοριοποίησης



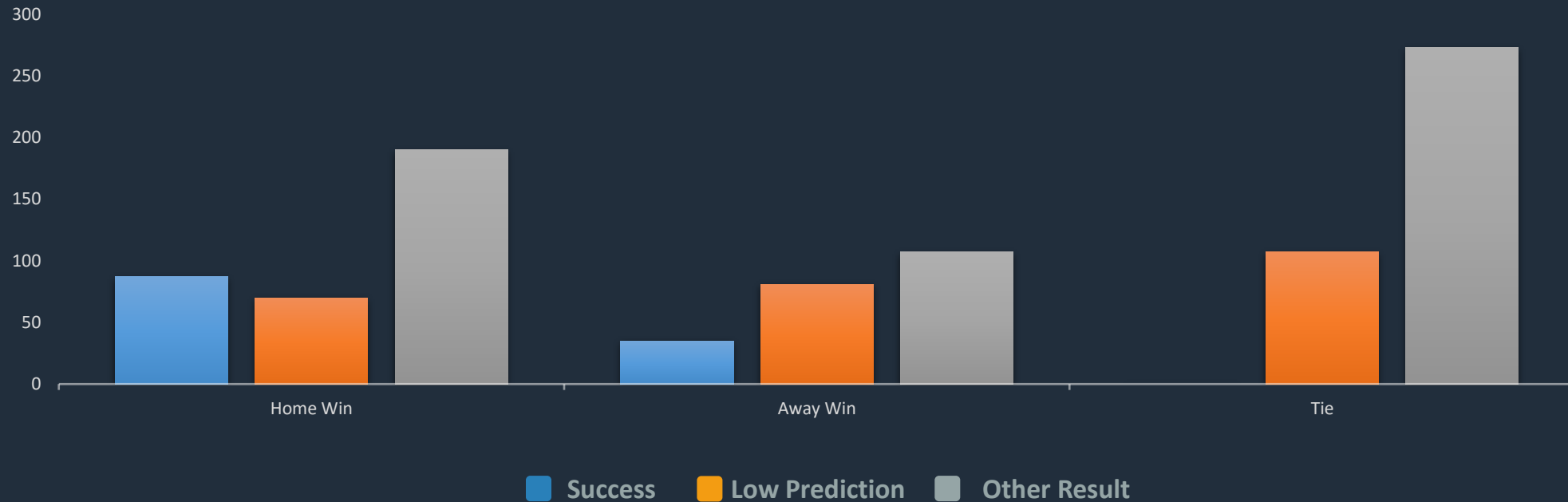
Κανόνες a priori

Ο αλγόριθμος a priori είναι ένας κλασικός αλγόριθμος για την εκμάθηση κανόνων συσχέτισης με σκοπό να αποκαλύπτουν ενδιαφέρουσες σχέσεις μεταξύ δεδομένων

Σκοπός είναι η δοκιμή και η εφαρμογή μεθόδων μηχανικής μάθησης και εξόρυξης δεδομένων για τον εντοπισμό κρυφών σχέσεων ανάμεσα σε δεδομένα όσο και για τη δημιουργία υποθέσεων στις οποίες αναζητούμε τους παράγοντες αυτούς που έχουν μεγαλύτερη σημαντικότητα για τη διαμόρφωση του τελικού αποτελέσματος ενός αγώνα ποδοσφαίρου της Αγγλικής Premier League τη σεζόν 2016.

Αποτελέσματα

Δοκιμή Υποθέσεων



Η απόδοση του κάθε μοντέλου για κάθε υπόθεση εκλέχθηκε με τη δημιουργία τριών κατηγοριών και βάση εξαγόμενης ποσοστιαίας πιθανότητας. Αναλυτικά:

- Για προβλέψεις με πιθανότητα μεγαλύτερη ή ίση από 0.7 θεωρήσαμε ότι το μοντέλο έχει προβλέψει, σε ανεκτά όρια, το τελικό αποτέλεσμα του αγώνα. (Success)
- Για προβλέψεις, με πιθανότητα μικρότερη από 0.7, θεωρήσαμε ότι το μοντέλο απέτυχε να προβλέψει σωστά το τελικό αποτέλεσμα του αγώνα. (Low Prediction)
- Εκτός από τις δύο κατηγορίες υπάρχει και αυτή που στην οποία το μοντέλο απέτυχε στο να αναγνωρίσει την νίκη είτε με μεγάλο είτε με μικρό ποσοστό. (Other Result)

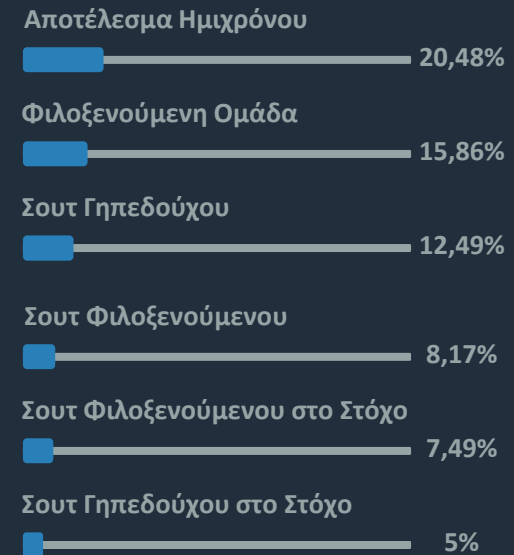
Αποτελέσματα

Παράγοντες που οδηγούν στη νίκη της γηπεδούχου, φιλοξενούμενης και ισοπαλία



Νίκη Γηπεδούχου

Τελικό Αποτέλεσμα / Σημαντικότεροι Παράγοντες



55,4% | Θετική Απόδοση Μοντέλου

Νίκη Φιλοξενούμενου

Τελικό Αποτέλεσμα / Σημαντικότεροι Παράγοντες

Αποτέλεσμα Ημιχρόνου



Γηπεδούχος Ομάδα



Φιλοξενούμενη ομάδα



Διαιτητής



Σουτ Γηπεδούχου στο Στόχο



Σουτ Φιλοξενούμενου στο Στόχο



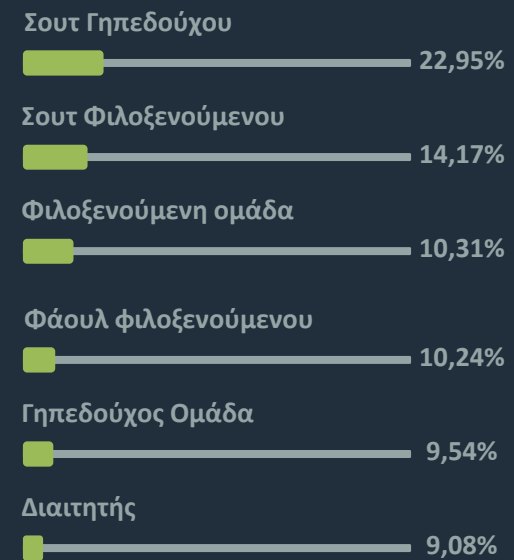
30,1% | Θετική Απόδοση Μοντέλου





Ισοπαλία

Τελικό Αποτέλεσμα / Σημαντικότεροι Παράγοντες



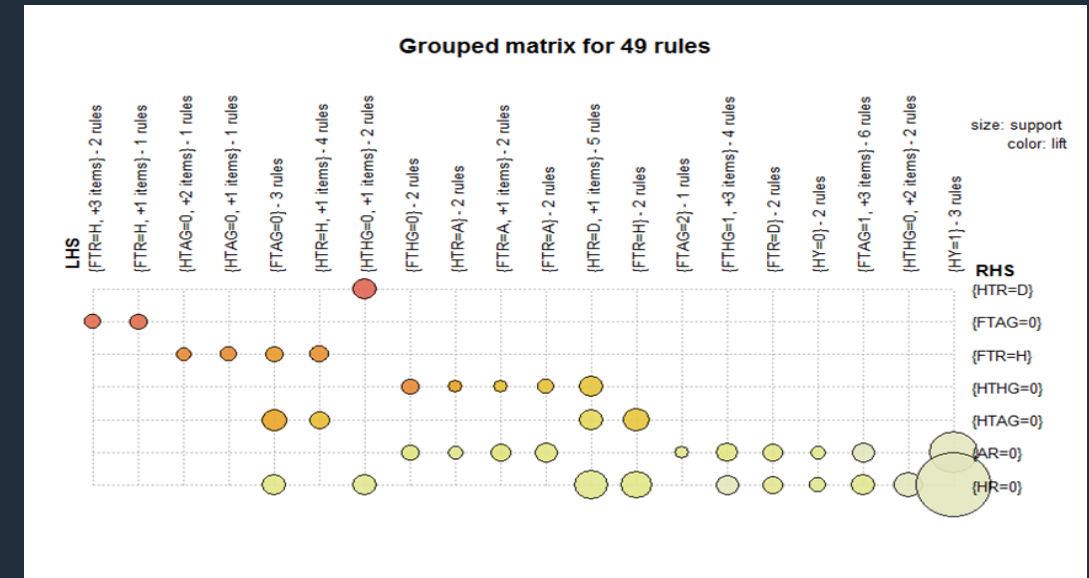
Το μοντέλο δεν κατάφερε να επιστρέψει θετικό ποσοστό επιτυχίας

Κανόνες a priori

Ενδεικτικά, μερικοί από τους εξαγόμενους κανόνες είναι οι εξής:

Rule		Support	Confidence	Lift
RHS	LHS			
1. {HTHG=0,HTAG=0}	{HTR=D}	0.3157895	1	2.261905
2. {HTR=H,HR=0}	{FTR=H}	0.2394737	0.7913043	1.915259
3. {FTR=H,HTAG=0,HR=0}	{FTAG=0}	0.2315789	0.7154472	2.210325
4. {HTAG=0,HTR=H}	{FTR=H}	0.2210526	0.7924528	1.918039
5. {HTAG=0,HTR=H,HR=0}	{FTR=H}	0.2105263	0.8	1.936306
6. {FTR=H,HTAG=0,AR=0}	{FTAG=0}	0.2052632	0.7155963	2.210785

Κανόνες a priori



Σχηματικά, το σύνολο των κανόνων μπορεί να εκφραστεί με διαγράμματα παράλληλων συντεταγμένων και με διαχωρισμό βάση συσταδοποίησης.

Υπόθεση ελέγχου διαιτητών

Συμπεράσματα

01

Μεγαλύτερο ποσοστό επιτυχίας πρόβλεψης της περίπτωσης νίκης γηπεδούχου από κάθε άλλη περίπτωση

02

Το αποτέλεσμα ημιχρόνου παίζει καταλυτικό ρόλο (20,5%) όπως επίσης και τα συνολικά σουτ του γηπεδούχου (15,8%) που είναι δυνατότητα απειλής για τον αντίπαλο

03

Τα συμπεράσματα από τους εξαγόμενους κανόνες τις περισσότερες φορές οδήγησε να είναι ο δείκτης των γκολ της γηπεδούχου στο ημίχρονο να είναι μηδέν (0) και τα γκολ της γηπεδούχου στο τελικό αποτέλεσμα του αγώνα να είναι ένα (1)



04

Εμφανίζεται μεγάλη δυνατότητα παρακολούθησης από τις ομάδες της Premier League καθώς η εξισορρόπηση του παθητικού των γκολ σε μηδέν από τον γηπεδούχο δίνει εμφανές προβάδισμα

05

Είναι ευδιάκριτο ότι οι ομάδες πρέπει να προσαρμόσουν και βελτιώσουν την άμυνά τους στο πρώτο τέταρτο του γηπέδου και να εξαλείψουν τους κινδύνους για την εστία τους

06

Το 31% των διαιτητών εμφανίζεται να καταλογίζει συγκεκριμένες παραβιάσεις στα παιχνίδια τους, με κάθε μία από αυτές να συνδυάζονται μεταξύ τους. Μόλις το 2% από αυτές, είναι μεμονωμένα περιστατικά



Ευχαριστούμε