



Προβλέψεις για το ΕΥΡΩΠΑΪΚΟ ΠΡΩΤΑΘΛΗΜΑ ΠΟΔΟΣΦΑΙΡΟΥ 2024 βασισμένες σε Στατιστικά Μοντέλα Αναλυτικής Ποδοσφαίρου

Ερευνητική ομάδα AUEB & Trieste Sports Analytics,

Οικονομικό Πανεπιστήμιο Αθηνών και Πανεπιστήμιο της Τεργέστης



Το παρόν άρθρο επιμελήθηκαν και συνέγραψαν ο Ιωάννης Ντζούφρας, Καθηγητής Στατιστικής του ΟΠΑ, και η Αργυρώ Δαμουλάκη, Υποψήφια Διδάκτωρ στο ίδιο τμήμα. Το άρθρο είναι βασισμένο στην ανάλυση της συνεργαζόμενης ομάδας της Τεργέστης (καθηγητή Leonardo Egidi και τους υποψήφιους διδάκτορες Roberto Macri Demartino και Giulio Fantuzzi) με τη συνδρομή των Β. Παλάσκα (OpenBet, ανάπτυξη εφαρμογής) Δ. Καρλή (Στατιστική ΟΠΑ, σύμβουλος ανάλυσης). Το τελικό αποτέλεσμα είναι συνεργασία των ερευνητικών ομάδων των δύο πανεπιστημίων σε θέματα Αναλυτικής του Αθλητισμού.

Η πρώτη αγωνιστική του Ευρωπαϊκού πρωταθλήματος ήταν απολαυστικά πλήρης: κάποιες εύκολες νίκες (Γερμανία, Ελβετία, Ισπανία και Ρουμανία), ανατροπές αποτελεσμάτων (η Πορτογαλία επικράτησε στις καθυστερήσεις της Τσεχίας), εκπλήξεις (με την Σλοβακία να κερδίζει το Βέλγιο), όμορφα γκολ (η Τουρκία σκόραρε εντυπωσιακά απέναντι στη Γεωργία) και κάποια φαβορί που κατάφεραν να επικρατήσουν «ολοκληρώνοντας απλά την δουλειά» (Ιταλία, Ολλανδία, Αγγλία και Γαλλία). Αναμένοντας με ανυπομονησία την συνέχεια, θα κάνουμε έναν σύντομο απολογισμό των προβλέψεων μας για την πρώτη αγωνιστική και θα παρουσιάσουμε τα «προγνωστικά» μας για την δεύτερη αγωνιστική της διοργάνωσης.

Υπενθύμιση για τους φίλους της Στατιστικής

Η χρήση στατιστικών τεχνικών για την πρόβλεψη αγώνων ποδοσφαίρου πρώτο-εμφανίστηκε στην επιστημονική βιβλιογραφία το 1968 με την πρωτοπόρα επιστημονική δημοσίευση των Reep & Benjamin. Οι επόμενες πραγματικές καινοτομίες εμφανίζονται στη δεκαετία του 80 (με την εργασία του Michael Maher) και τη δεκαετία του 90 (με την εργασία του Lee το 1997). Οι πρώτες όμως σημαντικές δημοσιεύσεις στο χώρο, εισάγοντας μοντέλα στα οποία βασίζονται και μοντέλα που χρησιμοποιούμε ακόμα και σήμερα, ήταν οι εργασίες των Dixon & Coles το 1997 και το διμεταβλητό μοντέλο Poisson των Καρλή και Ντζούφρα το 2003 (δύο από τους συγγραφείς της συγκεκριμένης ανάλυσης). Τα δύο αυτά μοντέλα έθεσαν τη βάση των συγχρόνων μοντέλων πρόβλεψης των αποτελεσμάτων αγώνων ποδοσφαίρου.

Σε αυτή την ανάλυση χρησιμοποιούμε το μοντέλο των Καρλή και Ντζούφρα μέσω του πακέτου “footbayer” στη στατιστική γλώσσα προγραμματισμού R που έχουν αναπτύξει ο Καθηγητής Leonardo Egidi από το Πανεπιστήμιο της Τεργέστης με τη συνδρομή του Βασίλη Παλάσκα (Αναλυτής στην Open Bet και ενεργό μέλος του AUFB Sports Analytics Group). Το μοντέλο επίσης συμπεριλαμβάνει την εκτίμηση παραμέτρων που εκτιμούν την απόδοση κάθε ομάδας που αλλάζουν στον χρόνο. Για την εκμάθηση του μοντέλου χρησιμοποιήθηκαν όλοι οι διεθνείς αγώνες της περιόδου 2020-2024. Κύρια επεξηγηματική μεταβλητή είναι η διαφορά μεταξύ των δύο ομάδων στο δείκτη Coca-Cola/FIFA ranking. Το μοντέλο, που προτάθηκε για πρώτη φορά από τους Καρλή & Ντζούφρα το 2003, επεκτείνει το συνηθισμένο διμεταβλητό μοντέλο Poisson. Λεπτομέρειες για το μοντέλο στατιστικής και μηχανικής μάθησης που χρησιμοποιήθηκε θα βρείτε στο τέλος αυτού του άρθρου.

Απολογισμός 1ης αγωνιστικής

Στον Πίνακα 1 θα βρείτε τις προβλέψεις των πρώτων 12 αγώνων (1η αγωνιστική) και τα τελικά αποτελέσματα. Όπως βλέπετε το μοντέλο κατάφερε να προβλέψει σωστά (με βάση την πιθανότητα των πιθανών εκβάσεων του αγώνα δλδ Νίκη, ισοπαλία, ήττα) αρκετούς αγώνες (75%) μεταξύ των οποίων ξεχωρίζουμε τον αγώνα μεταξύ Ουγγαρίας και Ελβετίας. Η αναμέτρηση αυτή ήταν η πιο αμφίρροπη με βάση το μοντέλο μας δίνοντας ένα ελαφρύ προβάδισμα στην Ελβετία η οποία και τελικά επικράτησε. Επίσης, σε 2 ακόμα αγώνες το μοντέλο έδειχνε το τελικό αποτέλεσμα να είναι αρκετά πιθανό (αλλά όχι το πιθανότερο ενδεχόμενο). Πιο συγκεκριμένα στον αγώνα Σλοβενία-Δανία έχουμε αξιοσημείωτη πιθανότητα της ισοπαλίας (27%) ενώ στον αγώνα Ρουμανία-Ουκρανία η πιθανότητα νίκης ήταν αξιοσημείωτη τόσο για την Ουκρανία (47%) όσο και για την Ρουμανία (25%). Όσον αφορά την μεγάλη έκπληξη της Σλοβακίας, το μοντέλο έδινε μόλις 8% πιθανότητα νίκης όπως και κάθε λογικό μοντέλο που θα μπορούσαμε να φτιάξουμε βασισμένο σε δεδομένα. Να σημειώσουμε εδώ ότι ένα λογικό μοντέλο στατιστικής και μηχανικής μάθησης σε καμία περίπτωση δε θα μπορέσει να πιάσει εκπλήξεις σαν και αυτές οι οποίες μπορεί να συμβούν λόγω τυχαιότητα ή συγκεκριμένων καταστάσεων που δεν λαμβάνονται υπόψη από το μοντέλο και από τα δεδομένα στα οποία έχουν εκπαιδευτεί.

Πίνακας 1: Πίνακας με τις πιθανότητες έκβασης των αγώνων για την 1^η αγωνιστική του Ευρωπαϊκού Πρωταθλήματος 2024.

Αντίπαλες ομάδες (A-B)	Πιθανότητες			Επικρατέστερο Αποτέλεσμα (Πιθανότητα)	Τελικό Αποτέλεσμα
	Νίκη A ομάδας	Ισοπαλία	Νίκη B ομάδας		
Γερμανία Σκοτία	0.579	0.243	0.178	1-0 (0.143)	5-1
Ουγγαρία Ελβετία	0.326	0.329	0.345	0-0 (0.176)	1-3
Ισπανία Κροατία	0.455	0.289	0.256	1-0 (0.140)	3-0
Ιταλία Αλβανία	0.723	0.190	0.088	2-0 (0.148)	2-1
Πολωνία Ολλανδία	0.156	0.214	0.630	0-2 (0.113)	1-2
Σλοβενία Δανία	0.186	0.270	0.543	0-1 (0.167)	1-1
Σερβία Αγγλία	0.107	0.212	0.681	0-1 (0.150)	0-1
Ρουμανία Ουκρανία	0.254	0.277	0.469	0-1 (0.137)	3-0
Βέλγιο Σλοβακία	0.729	0.190	0.081	2-0 (0.158)	0-1
Αυστρία Γαλλία	0.170	0.243	0.588	0-1 (0.145)	0-1
Τουρκία Γεωργία	0.491	0.240	0.269	1-0 (0.097)	3-1
Πορτογαλία Τσεχία	0.693	0.196	0.111	2-0 (0.134)	2-1

Προβλέψεις 2ης αγωνιστικής

Προχωράμε με αισιοδοξία για την δεύτερη αγωνιστική, λοιπόν, με τις προβλέψεις του μοντέλου να παρουσιάζονται στον Πίνακα 2.

Από τον πίνακα αυτό ξεχωρίζουμε τον αγώνα

- Σλοβακία - Ουκρανία

ως τον πιο αμφίτροπο.

Ως φαβορί ξεχωρίζουν

1. Πορτογαλία με πιθανότητα νίκης 69% έναντι της Τουρκία
2. Βέλγιο με πιθανότητα νίκης 65% έναντι της Ρουμανίας
3. Κροατία με πιθανότητα νίκης 62% έναντι της Αλβανία
4. Αγγλία με πιθανότητα νίκης 58% έναντι της Δανίας
5. Ελβετία με πιθανότητα νίκης 57% έναντι της Σκωτία
6. Τσεχία με πιθανότητα νίκης 53% έναντι της Γεωργίας

Τέλος έχουμε άλλους πέντε αγώνες που είναι σχετικά αμφίτροποι αλλά με ελαφρό προβάδισμα της μίας από τις δύο ομάδες. Σε αυτούς τους αγώνες θεωρούμε ότι οι ομάδες είναι σχετικά κοντά και μπορεί να έρθουν και ισοπαλία λόγω τακτικής και στρατηγικής. Συγκεκριμένα, έχουμε

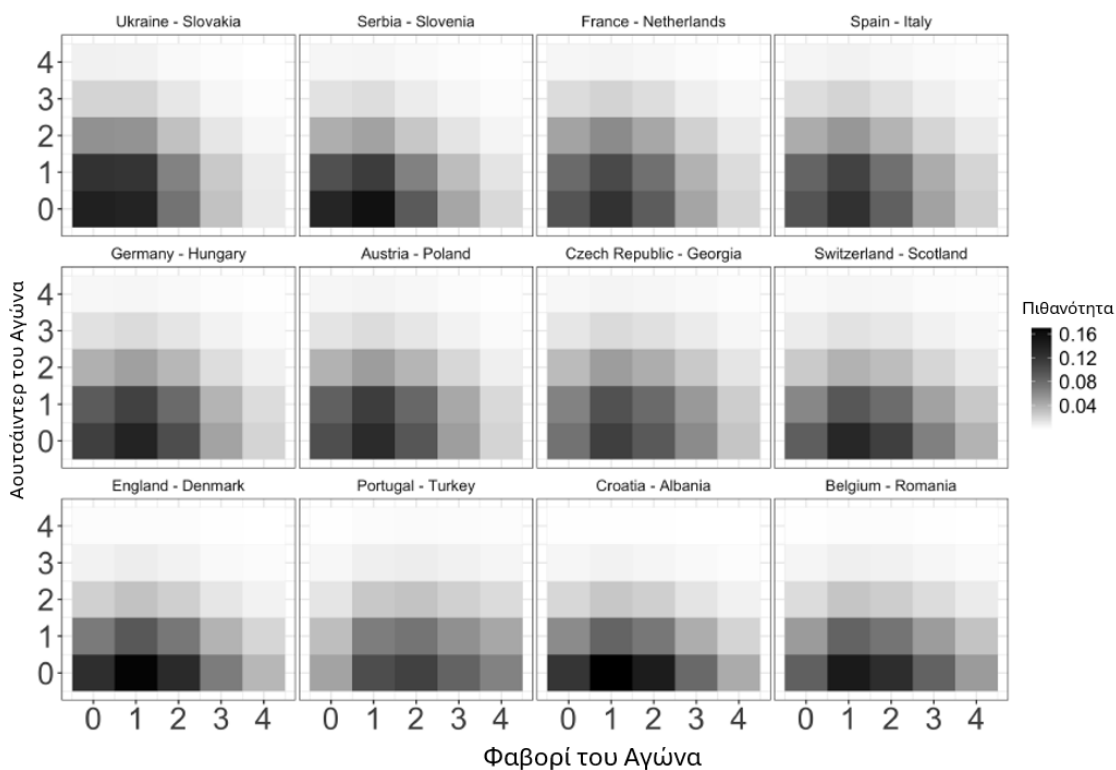
1. Αυστρία (49%) να επικρατεί της Πολωνίας (24%)
2. Γερμανία (48%) να επικρατεί της Ουγγαρίας (24%)
3. Ισπανία (47,5%) να επικρατεί της Ιταλίας (26%)
4. Γαλλία (47%) να επικρατεί της Ολλανδίας (27%)
5. Σερβία (46%) να επικρατεί της Σλοβενίας (45%)

Πίνακας 2: Πίνακας με τις πιθανότητες έκβασης των αγώνων για την 2^η αγωνιστική του Ευρωπαϊκού Πρωταθλήματος 2024.

Αντίπαλες ομάδες (A-B)	Πιθανότητες			Επικρατέστερο Αποτέλεσμα (Πιθανότητα)
	Νίκη A ομάδας	Ισοπαλία	Νίκη B ομάδας	
Κροατία Αλβανία	0.624	0.243	0.132	1-0 (0.170)
Γερμανία Ουγγαρία	0.482	0.274	0.244	1-0 (0.137)
Σκωτία Ελβετία	0.187	0.239	0.574	0-1 (0.133)
Σλοβενία Σερβία	0.248	0.292	0.460	0-1 (0.158)
Δανία Αγγλία	0.157	0.260	0.583	0-1 (0.169)
Ισπανία Ιταλία	0.475	0.264	0.261	1-0 (0.124)
Σλοβακία Ουκρανία	0.314	0.304	0.382	0-0 (0.140)
Πολωνία Αυστρία	0.244	0.268	0.489	0-1 (0.129)
Ολλανδία Γαλλία	0.269	0.264	0.467	0-1 (0.123)
Γεωργία Τσεχία	0.227	0.242	0.530	0-1 (0.112)
Τουρκία Πορτογαλία	0.118	0.183	0.699	0-2 (0.110)
Βέλγιο Ρουμανία	0.649	0.221	0.130	1-0 (0.147)

Το Διάγραμμα 1 δίνει με πιο πολύ λεπτομέρεια τις πιθανότητες για το κάθε σκορ για καθένα από τους 12 αγώνες της 2^{ης} αγωνιστικής.

Πιθανότητα αποτελεσμάτων αγώνα (αφού λάβουμε υπόψη μας τα δεδομένα)



Διάγραμμα 1: Διάγραμμα Πιθανοτήτων πιθανών σκορ για τους Αγώνες της 2^{ης} αγωνιστικής του Ευρωπαϊκού Πρωταθλήματος 2024.

Βιβλιογραφία για διαβαστερούς φιλάθλους

- Dixon, M.J. and Coles, S.G. (1997), Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **46**, 265-280.
- Karlis, D. and Ntzoufras, I. (2003), Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **52**, 381-393.
- Lee A.J. (1997). Modeling Scores in the Premier League: Is Manchester United Really the Best? *Chance*, **10**, 15-19.
- Maher, M.J. (1982), Modelling association football scores. *Statistica Neerlandica*, **36**, 109-118.
- Reep, C., & Benjamin, B. (1968). Skill and Chance in Association Football. *Journal of the Royal Statistical Society. Series A (General)*, **131**, 581-585.

Οι Μαγικές Εξισώσεις του στατιστικού μοντέλου

$$(X_i, Y_i) \sim \begin{cases} (1-p)\text{BP}(x_i, y_i | \lambda_1, \lambda_2, \lambda_3) & \text{if } x \neq y \\ (1-p)\text{BP}(x_i, y_i | \lambda_1, \lambda_2, \lambda_3) + pD(x, \eta) & \text{if } x = y, \end{cases} \quad (1)$$

$$\log(\lambda_{1i}) = \text{att}_{h_i, t} + \text{def}_{a_i, t} + \frac{\gamma}{2}(\text{ranking}_{h_i} - \text{ranking}_{a_i}) \quad (2)$$

$$\log(\lambda_{2i}) = \text{att}_{a_i, t} + \text{def}_{h_i, t} - \frac{\gamma}{2}(\text{ranking}_{h_i} - \text{ranking}_{a_i}), \quad i = 1, \dots, n \text{ (matches)}, \quad (3)$$

$$\log(\lambda_{3i}) = \rho, \quad (4)$$

$$\text{att}_{k, t} \sim \mathcal{N}(\text{att}_{k, t-1}, \sigma^2), \quad (5)$$

$$\text{def}_{k, t} \sim \mathcal{N}(\text{def}_{k, t-1}, \sigma^2), \quad (6)$$

$$\rho, \gamma \sim \mathcal{N}(0, 1) \quad (7)$$

$$p \sim \text{Uniform}(0, 1) \quad (8)$$

$$\sum_{k=1}^{n_t} \text{att}_k = 0, \quad \sum_{k=1}^{n_t} \text{def}_k = 0, \quad k = 1, \dots, n_t \text{ (teams)}, \quad t = 1, \dots, T \text{ (times)}. \quad (9)$$

- i είναι ο δείκτης του αγώνα
- X_i και Y_i είναι ο αριθμός των γκολ μεταξύ της 1^{ης} και της 2^{ης} ομάδας στον αγώνα i
- h_i και a_i είναι η 1^η και 2^η ομάδα αντίστοιχα (ή η εντός και εκτός έδρα ομάδα – όπου ισχύει) για τον i αγώνα.
- $\text{att}_{k, t}$ και $\text{def}_{k, t}$ οι παράμετροι που εκτιμούν της επιθετική και αμυντική δυναμικότητα/ ικανότητα της ομάδας k την χρονική στιγμή t (δυναμικές παράμετροι που αλλάζουν στο χρόνο)
- ranking_k δείκτης Coca-Cola FIFA ranking την 4^η Απριλίου 2024 για την ομάδα k .

Λίγα λόγια για τους Συγγραφείς

AUEB Sports Analytics Group



Ο **Ιωάννης Ντζουφρας** είναι καθηγητής Στατιστικής και πρόεδρος στο Τμήμα Στατιστικής του Οικονομικού Πανεπιστημίου Αθηνών. Είναι ιδρυτικό μέλος της ερευνητικής ομάδας AUEB Sports Analytics Group μαζί με τον Δημήτρη Καρλή. Έχει αναγνωρισμένη επιστημονική δραστηριότητα σε τομείς όπως η Μπευζιανή στατιστική μεθοδολογία, υπολογιστική στατιστική, Βιοστατιστική, ψυχομετρία και αναλυτική των σπορ.



Ο **Δημήτρης Καρλής** είναι καθηγητής Στατιστικής και αναπληρωτής πρόεδρος στο Τμήμα Στατιστικής του Οικονομικού Πανεπιστημίου Αθηνών. Είναι ιδρυτικό μέλος της ερευνητικής ομάδας AUEB Sports Analytics Group μαζί με τον Ιωάννη Ντζούφρα. Έχει αναγνωρισμένη επιστημονική δραστηριότητα σε τομείς όπως η στατιστική μεθοδολογία, υπολογιστική στατιστική, Βιοστατιστική, και αναλυτική των σπορ.



Η **Αργυρώ Δαμουλάκη** είναι υποψήφια διδάκτωρ στο Τμήμα Στατιστικής του ΟΠΑ. Η τρέχουσα έρευνά της επικεντρώνεται σε μοντέλα αξιολόγησης απόδοσης με τη χρήση της Μπευζιανής στατιστικής. Ασχολείται με αναλυτική στα σπορ και είναι ενεργό μέλος AUEB Sports Analytics Group από το 2022.



Ο **Βασίλης Παλάσκας** είναι Στατιστικός Αναλυτής και Επιστήμονας Δεδομένων στην Open Bet. Στο παρελθόν δούλεψε ως επικεφαλής της ανάλυσης και πρόβλεψης στην Fantasy Sports Interactive (FSI). Είναι ενεργό μέλος της ερευνητικής ομάδας AUEB Sports Analytics από το 2019 όπου τελείωσε το M.Sc. in Statistics του Οικονομικού Πανεπιστημίου Αθηνών.

Η ερευνητική ομάδα του Πανεπιστημίου της Τεργέστης



Ο **Leonardo Egidì** είναι επίκουρος καθηγητής Στατιστικής στο Πανεπιστήμιο της Τεργέστης στην Ιταλία και μέλος της ερευνητικής ομάδας του Οικονομικού Πανεπιστημίου Αθηνών AUEB Sports Analytics Group. Έχει διδακτορικό στην μοντελοποίηση και αναλυτική ποδοσφαίρου και έντονη ερευνητική δραστηριότητα στη Μπευζιανή Στατιστική μεθοδολογία.



Ο **Nicola Torelli** είναι Καθηγητής Στατιστικής στο Πανεπιστήμιο της Τεργέστης. Τα κύρια ερευνητικά του πεδία αφορούν τη μοντέλα ταξινόμησης, την ανάλυση κατά συστάδες και τη Μπευζιανή μοντελοποίηση. Είναι ο πρώην πρόεδρος της Ιταλική Στατιστικής Επιστημονικής Εταιρείας.



Ο **Roberto Macri Demartino** είναι Υποψήφιος Διδάκτωρ στο Πανεπιστήμιο της Πάντοβα. Τα κύρια ερευνητικά του ενδιαφέροντα επικεντρώνονται στην Μπευζιανή στατιστική μεθοδολογία με εφαρμογές σε ποικίλους τομείς, όπως αναλυτική των σπορ.



Ο **Giulio Fantuzzi** έλαβε πτυχίο στη Στατιστική και είναι τώρα μεταπτυχιακός φοιτητής στην Επιστήμη Δεδομένων και Τεχνητή Νοημοσύνη στο Πανεπιστήμιο της Τεργέστης. Είναι πολύ δραστήριος στον τομέα της μοντελοποίησης των σπορ κυρίως με χρήση R και Python.

Ενεργές Συνεργασίες των Συγγραφέων

Οι τρεις συγγραφείς (L. Egidì, I. Ντζούφρας και Δ. Καρλής) του άρθρου αυτή τη στιγμή συνεργάζονται για τη συγγραφή ενός βιβλίου σε **Football Analytics** σε διεθνή επιστημονικό οίκο ενώ στο τελευταίο workshop της ομάδας έδωσαν ένα σεμιναριακό μάθημα σε Football analytics.

Τα δύο Πανεπιστήμια επίσης ετοιμάζουν ένα μεταπτυχιακό πρόγραμμα στην Αναλυτική των Σπορ με σχεδιασμό να ξεκινήσει τον Οκτώβριο του 2025.

Ο **L. Egidi** και **B. Παλάσκας** συνεργάζονται στην ανάπτυξη του λογισμικού “footbayes” (βιβλιοθήκη της στατιστικής γλώσσας προγραμματισμού R).

Ο **L. Egidi**, **I. Ντζούφρας** και **B. Παλάσκας** συνεργάζονται στην συγγραφή ενός επιστημονικού άρθρου αξιολόγησης παικτών στο Βόλεϊ.

Ο **B. Παλάσκας** είναι απόφοιτος του μεταπτυχιακού του ΟΠΑ M.Sc. in Sports Analytics, ενεργό μέλος του AUEB Sports Analytics ενώ έχει συνεπιβλέψει μια διπλωματική εργασία στα πλαίσια του M.Sc. in Statistics του ΟΠΑ και της προηγούμενης συνεργασίας του με την FSI (Fantasy Sports Interactive).

Η Ομάδα AUEB Sports Analytics



Η ερευνητική ομάδα του Οικονομικού Πανεπιστημίου Αθηνών **AUEB Sports Analytics Group** ιδρύθηκε το 2015 από τους καθηγητές Ιωάννη Ντζούφρα και Δημήτρη Καρλή. Μέλη του είναι σημαντικά μέλη της κοινότητας της αναλυτικής των σπορ όπως οι Leonardo Egidi (Πανεπιστήμιο Trieste), Ιωάννης Κοσμίδης (Warwick), Κωνσταντίνος Πελεχρίνης (Pittsburg), Nial Friel (UCD) και Gianluca Baio (UCL) καθώς επίσης και ο πρώην προπονητής της εθνικής Ελλάδας Βόλεϊ, Σωτήρης Δρίκος και ο scouter των Σακραμέντο Κινγκς, Χρήστος Μαρμαρινός. Η ερευνητική ομάδα είναι υπεύθυνη για της σειρά ετήσιων συνεδρίων με το όνομα AUEB Sports Analytics Workshop (6 συνολικά) ενώ το 2019 διοργάνωσε το διεθνές συνέδριο MathSport 2019 με 200 συμμετέχοντες επιστήμονες από όλο τον κόσμο. Η ομάδα έχει μια σειρά από σημαντικές επιστημονικές δημοσιεύσεις στο χώρο της αναλυτικής των σπορ. Τέλος θα θέλαμε να αναφέρουμε ότι η ομάδα ιδρύθηκε το 2015 λόγω της επίσκεψης του καθηγητή Stefan Kesenne (Πανεπιστήμιο Antwerp & Leuven), σπουδαιού Οικονομολόγου του Αθλητισμού που έπαιξε και ενεργό ρόλο στην υπόθεση Bosman. Ο Stefan Kesenne στήριξε ενεργά την ομάδα μέχρι και το 2021 όπου ξαφνικά απεβίωσε. Η ύπαρξη της ομάδας AUEB Sports Analytics Group οφείλεται σε μεγάλο ποσοστό στη συνδρομή και την έμπνευση που μας έδωσε ο κος Kesenne.

Website του group <https://aueb-analytics.wixsite.com/sports>

Website του AUEB Sports Analytics Workshop <https://aueb-analytics.wixsite.com/saw2024>

Ακούστε το podcast για Basketball Analytics από τον Ιωάννη Ντζουφρα, Αργυρώ Δαμουλάκη και Χρήστο Μαρμαρινό [εδώ](#).