



Προβλέψεις για το ΠΑΓΚΟΣΜΙΟ ΚΥΠΕΛΛΟ ΠΟΔΟΣΦΑΙΡΟΥ 2022 βασισμένες σε Στατιστικά Μοντέλα Αναλυτικής Ποδοσφαίρου

L. Egidi, B. Παλάσκας, Ι. Ντζούφρας & Δ. Καρλής

Ερευνητική ομάδα AUEB Sports Analytics,

Οικονομικό Πανεπιστήμιο Αθηνών, Πανεπιστήμιο της Τεργέστης & Fantasy Sports Interactive

Συγγραφέας του παρόντος άρθρου είναι ο Ιωάννης Ντζούφρας. Το άρθρο είναι βασισμένο στην ανάλυση των Leonardo Egidi (Πανεπιστήμιο της Τεργέστης) και του Βασίλειου Παλάσκα (Fantasy Sports Interactive) με τις συμβουλευτικές συνδρομές των Ι. Ντζούφρα και Δ. Καρλή. Και οι τέσσερις συγγραφείς είναι ενεργά μέλη της ερευνητικής ομάδας του Οικονομικού Πανεπιστημίου Αθηνών: AUEB Sports Analytics Group.

Φτάσαμε στο τέλος της 2^{ης} αγωνιστικής της φάσης των ομίλων του μουντιάλ. Ας σου δούμε που βρισκόμαστε μέχρι στιγμής με βάση τα αποτελέσματα και το μοντέλο της ερευνητικής μας ομάδας.

Να υπενθυμίσουμε ότι σε αυτή την ανάλυση χρησιμοποιούμε το μοντέλο των Καρλή και Ντζούφρα μέσω του πακέτου “footbaves” στη στατιστική γλώσσα προγραμματισμού R που έχουν αναπτύξει οι 2 πρώτοι συγγραφείς αυτού του άρθρου και της ανάλυσης. Το μοντέλο επίσης συμπεριλαμβάνει την εκτίμηση παραμέτρων που εκτιμούν την απόδοση κάθε ομάδας που αλλάζουν στον χρόνο. Για την εκμάθηση του μοντέλου χρησιμοποιήθηκαν περισσότερα από 3000 διεθνείς αγώνες της περιόδου 2018-2022. 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.

Οι Προβλέψεις του Μοντέλου για την 3^η Αγωνιστική

Η 2^η αγωνιστική είχε πολλές εκπλήξεις και μη αναμενόμενα αποτελέσματα. Οι ενημερωμένες προβλέψεις με τη μορφή πιθανοτήτων δίνονται στον Πίνακα 1. Ως φαβορί δηλώνεται η ομάδα με τη μεγαλύτερη πιθανότητα νίκης.

Συγκεκριμένα από τον Πίνακα 1 έχουμε 3 αγώνες που οι δύο ομάδες είναι κοντά και στην ουσία δεν είναι εύκολο να αποφασίσουμε τον νικητή. Συγκεκριμένα αυτοί οι αγώνες είναι

- Εκουαδόρ-Σενεγάλη
- Ελβετία-Σερβία
- ΗΠΑ-Ιράν

Σύμφωνα με τα αποτελέσματα έχουμε 10 αγώνες που έχουμε το νικητή με πιθανότητα μεγαλύτερη του 50%. Συγκεκριμένα,

1. Η Βραζιλία με πιθανότητα νίκης 80% έναντι του Καμερούν
2. Η Ολλανδία με πιθανότητα νίκης 77% έναντι του Κατάρ
3. Η Ουρουγουάη με πιθανότητα νίκης 68% έναντι της Γκάνα
4. Η Αργεντινή με πιθανότητα νίκης 66% έναντι της Πολωνίας
5. Η Γερμανία με πιθανότητα νίκης 65% έναντι της Κόστα Ρίκα
6. Η Γαλλία με πιθανότητα νίκης 64% έναντι της Τυνησίας
7. Η Αγγλία με πιθανότητα νίκης 62% έναντι της Ουαλίας
8. Η Πορτογαλία με πιθανότητα νίκης 61% έναντι της Νότιας Κορέας
9. Η Ισπανία με πιθανότητα νίκης 57.5% έναντι της Ιαπωνίας, και
10. Η Δανία με πιθανότητα νίκης 55% έναντι της Αυστραλίας

Να σημειώσουμε ότι το μοντέλο δεν λαμβάνει υπόψη του τυχόν σκοπιμότητες ή επιπλέον ενδιαφέρον που έχουν κάποιες ομάδες για πρόκριση στον επόμενο γύρο.

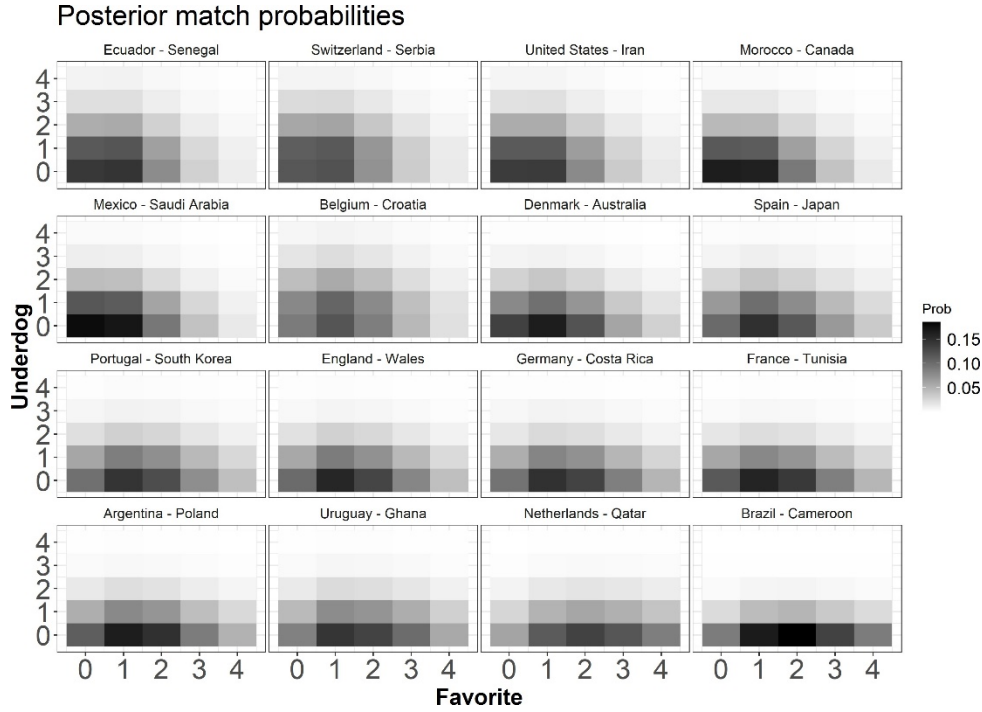
Τέλος έχουμε άλλους τρεις αγώνες όπου το φαβορί έχει ελαφρύ προβάδισμα με πιθανότητα νίκης μεταξύ 40% και 50%. Συγκεκριμένα έχουμε τους αγώνες

1. Βέλγιο-Κροατία (43% vs. 26%) με την ισοπαλία επίσης αρκετά πιθανή.
2. Μεξικό-Σαουδική Αραβία (44% vs 24,3%) με την ισοπαλία επίσης αρκετά πιθανή, και
3. Βέλγιο-Κροατία (47 vs 27%)

Πίνακας 1: Πιθανότητες Αποτελεσμάτων για τη 3^η Αγωνιστική με βάση το Μοντέλο Μπευζιανής Στατιστικής Μηχανικής Μάθησης της Ερευνητικής ομάδας AUEB Sports Analytics

Αγώνας	Φαβορί	Αουτσάιντερ	Νίκη για το φαβορί	Ισοπαλία	Νίκη για το Αουτσάιντερ
1	Ecuador	Senegal	0.389	0.302	0.309
2	Netherlands	Qatar	0.767	0.160	0.073
3	United States	Iran	0.402	0.297	0.301
4	England	Wales	0.624	0.236	0.140
5	France	Tunisia	0.643	0.236	0.122
6	Denmark	Australia	0.546	0.274	0.180
7	Argentina	Poland	0.658	0.229	0.112
8	Mexico	Saudi Arabia	0.436	0.322	0.243
9	Belgium	Croatia	0.470	0.260	0.270
10	Morocco	Canada	0.428	0.313	0.259
11	Spain	Japan	0.575	0.251	0.175
12	Germany	Costa Rica	0.646	0.224	0.130
13	Portugal	South Korea	0.611	0.236	0.153
14	Uruguay	Ghana	0.681	0.208	0.111
15	Switzerland	Serbia	0.393	0.284	0.322
16	Brazil	Cameroon	0.796	0.162	0.041

Στο Διάγραμμα 1 μπορείτε να δείτε τις πιθανότητες για το κάθε σκορ για καθένα από τους 16 αγώνες της 3^{ης} αγωνιστικής.



Διάγραμμα 1: Διάγραμμα Πιθανοτήτων πιθανών σκορ για τους Αγώνες της 2^{ης} αγωνιστικής του Παγκοσμίου Κυπέλλου 2022.

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

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

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

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

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

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

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

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

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

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

- 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.
- $att_{k,t}$ και $def_{k,t}$ 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 6^η Οκτωβρίου 2022 για την ομάδα k .

Λίγα λόγια για τους Συγγραφείς



Ο **Leonardo Egidio** 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 **Βασίλης Παλάσκακας** είναι Στατιστικός Αναλυτής και Επιστήμονας Δεδομένων 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.



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.

Ενεργές Συνεργασίες των Συγγραφέων

Οι τρεις συγγραφείς (L. Egidio, 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.

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

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

Ο **I. Ντζούφρας** και **B. Παλάσκακας** συνεπιβλέπουν μια διπλωματική εργασία στα πλαίσια του M.Sc. in Statistics του ΟΠΑ και της συνεργασίας 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 νυν προπονητής της Εθνικής ομάδας Μπάσκετ του Κοσόβου, 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), σπουδαίου 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.