



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

ΦΑΣΗ των 16

L. Egidi, B. Παλάσκας, I. 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 Βασιλείου Παλάσκα (Fantasy Sports Interactive) με τις συμβουλευτικές συνδρομές των I. 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 των ομίλων του μουντιάλ με συναρπαστικού αγώνες και μη αναμενόμενα αποτελέσματα με μεγαλύτερη έκπληξη τον αποκλεισμό της Γερμανίας και την πρόκριση της μαχητικής Ιαπωνίας.

Οπότε ήρθε ώρα για τις προβλέψεις της φάσης των 16. Τη στιγμή που γραφόταν αυτό το κείμενο, ήδη είχαν περάσει η Ολλανδία (επί των ΗΠΑ) και η Αργεντινή επί της Αυστραλίας (έστω και δύσκολα). Αποτελέσματα αναμενόμενα με βάση την κοινή λογική αλλά και βάση του μοντέλου το οποίο έδινε 47% για νίκη της Ολλανδίας και 66,5% για νίκη της Αργεντινής.

Να υπενθυμίσουμε ότι σε αυτή την ανάλυση χρησιμοποιούμε το μοντέλο των Καρλή και Ντζούφρα μέσω του πακέτου "footbayes" 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 Φάση των 16

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 τα ακόλουθα φαβορί με πιθανότητες νίκης μεγαλύτερης του 50%:

- Η Βραζιλία με πιθανότητα νίκης 74% έναντι της Νότιας Κορέας.
- Η Αργεντινή με πιθανότητα νίκης 66,5% έναντι της Αυστραλίας (στον αγώνα ήδη κέρδισε έστω και δύσκολα η Αργεντινή)
- Η Αγγλία με πιθανότητα νίκης 63% έναντι της Σενεγάλης
- Η Γαλλία με πιθανότητα νίκης 62% έναντι της Πολωνίας
- Η Ισπανία με πιθανότητα νίκης 56% έναντι του Μαρόκο

Από τα παραπάνω βλέπουμε ότι η Βραζιλία φαίνεται να έχει τη μεγαλύτερη πιθανότητα νίκης. Η Ισπανία φαίνεται να έχει οριακή πιθανότητα πάνω από το 50% έναντι του Μαρόκο (γεγονός που οφείλεται στα αποτελέσματα της 3ης αγωνιστικής του μουντιάλ).

Μετά έχουμε 2 αγώνες όπου το φαβορί έχει προβάδισμα με πιθανότητα νίκης κάτω από 50% αλλά μεγαλύτερη του 40%. Αυτές οι δύο ομάδες είναι η **Πορτογαλία** με πιθανότητα επικράτησης 48,6% έναντι της Ελβετίας (που έχει καλές πιθανότητες να πάει τον αγώνα παράταση ή να κερδίσει) και η **Ολλανδία** έναντι των ΗΠΑ (με πιθανότητα 47%). Για τον τελευταίο αγώνα ήδη ξέρουμε ότι επικράτησε η Ολλανδία όπως αναμενόταν.

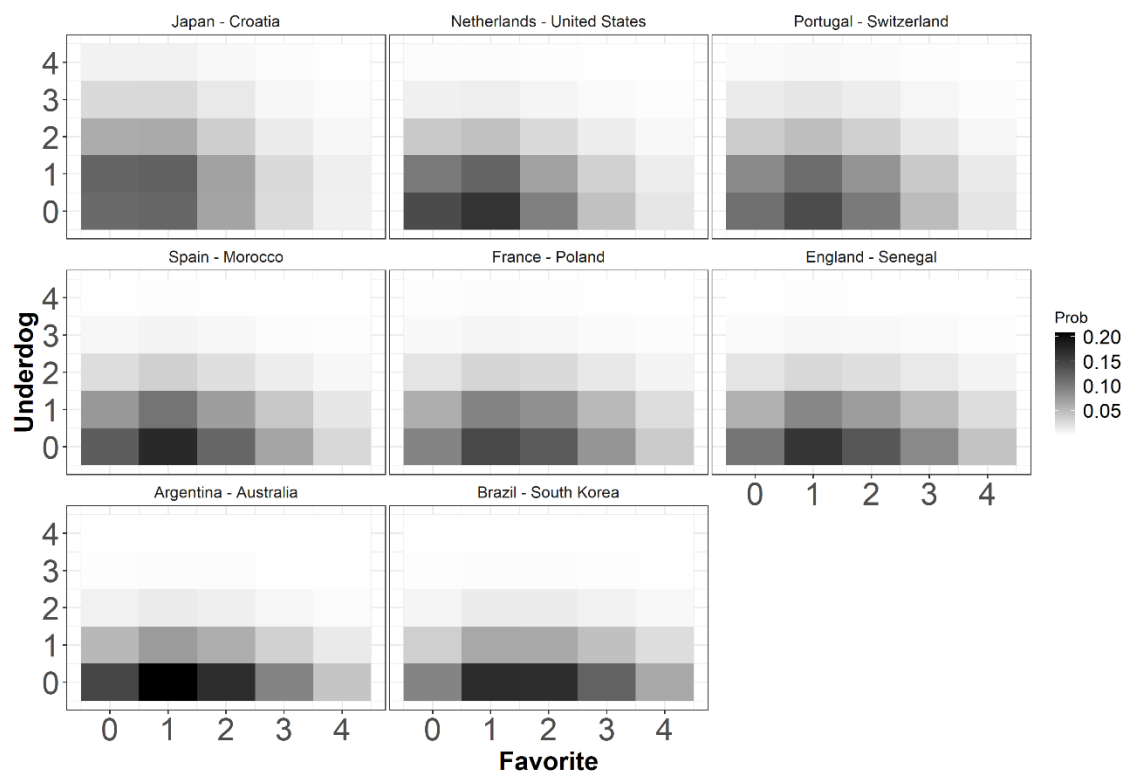
Τέλος, το μοντέλο δίνει ότι όλα τα αποτελέσματα έχουν παρόμοια πιθανότητα στον αγώνα της **Ιαπωνίας με την Κροατία**. Η Ιαπωνία έχει ελαφρύ προβάδισμα με 36,3% λόγω των εξαιρετικών αγώνων στον όμιλο της ενώ η Κροατία έχει λίγο μικρότερη πιθανότητα νίκης (35%). Ανυπομονούμε να δούμε τι θα γίνει λοιπόν σε αυτόν το αγώνα.

Πίνακας 1: Πιθανότητες Αποτελεσμάτων για τη φάση των 16 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	Netherlands	United States	0.471	0.298	0.231
2	Argentina	Australia	0.665	0.246	0.089
3	France	Poland	0.620	0.235	0.100
4	England	Senegal	0.632	0.235	0.132
5	Japan	Croatia	0.363	0.288	0.349
6	Brazil	South Korea	0.740	0.190	0.070
7	Spain	Morocco	0.562	0.270	0.168
8	Portugal	Switzerland	0.486	0.273	0.241

Στο Διάγραμμα 1 μπορείτε να δείτε τις πιθανότητες για το κάθε σκορ για καθένα από τους 8 αγώνες της φάσης των 16.

Posterior match probabilities



Διάγραμμα 1: Διάγραμμα Πιθανοτήτων πιθανών σκορ για τους Αγώνες της Φάσης των 16 του Παγκοσμίου Κυπέλλου 2022.

Βιβλιογραφία για διαβαστέρους 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 οι αριθμοί 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 6^η Οκτωβρίου 2022 για την ομάδα k .

Λίγα λόγια για τους Συγγραφείς



Ο **Leonardo Egidi** 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. Egidi, 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. 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 **I. Ντζούφρας** και **B. Παλάσκας** 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.