



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

Ημιτελική Φάση του Παγκοσμίου Κυπέλου 2022

L. Egidi, Β. Παλάσκα, 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 Βασιλείου 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.

Μπήκαμε στην τελική ευθεία για την ανακήρυξη του φετινού νικητή του Παγκόσμιου κυπέλλου 2022. Μια διοργάνωση γεμάτη εκπλήξεις και με την Κροατία και (κυρίως) το Μαρόκο να φτάνουν στα ημιτελικά. Θα τα καταφέρουν να φτάσουν στον τελικό ή ακόμα καλύτερα να κερδίσουν τον τίτλο;

Οι Προβλέψεις του Μοντέλου για την Φάση των ημιτελικών

Οι ενημερωμένες προβλέψεις για τους 2 ημιτελικούς με τη μορφή πιθανοτήτων δίνονται στον Πίνακα 1. Ως φαβορί δηλώνεται η ομάδα με τη μεγαλύτερη πιθανότητα νίκης.

Συγκεκριμένα από τον Πίνακα 1 βλέπουμε ότι

- Η **Αργεντινή** έχει προβάδισμα επι της Κροατίας με πιθανότητα νίκης 58% όμως η Κροατία έχει 42% πιθανότητα να πάει τον αγώνα στην παράταση ή να κερδίσει. Αυτό πρακτικά σημαίνει ότι αν οι δύο ομάδες παίξουν 10 φορές περιμένουμε περίπου τις 6 να κερδίσει η Αργεντινή και στους 4 να οδηγηθεί ο αγώνας στην παράταση ή να κερδίσει η Κροατία.
- Στον δεύτερο αγώνα η Γαλλία επίσης έχει προβάδισμα επί του Μαρόκο με πιθανότητα επικράτησης 50%. Αυτό σημαίνει ότι να οι δύο ομάδες παίξουν 10 φορές στους 5 αγώνες θα κερδίσει η Γαλλία και στους άλλους 5 ο αγώνας θα πάει παράταση ή θα κερδίσει το Μαρόκο (που σημαίνει ότι το Μαρόκο έχει καλές πιθανότητες πρόκρισης). Να σημειώσουμε ότι το Μαρόκο ήδη έχει κερδίσει Ισπανία και Πορτογαλία όπου

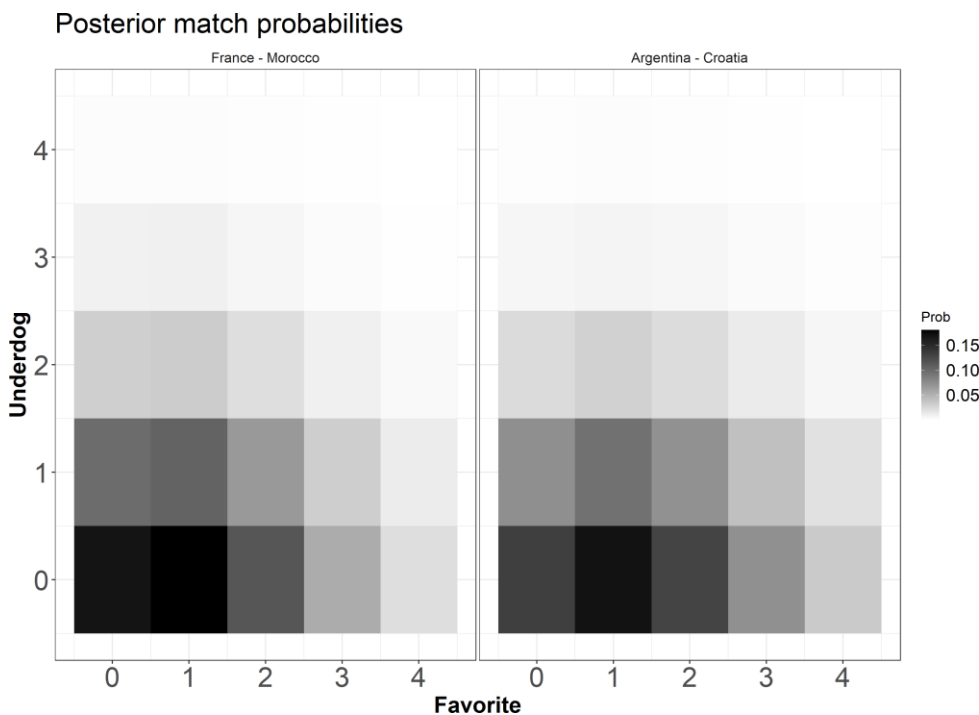
οι πιθανότητες ήταν 56-44% και 54-46% αντίστοιχα. Συνεπώς εδώ βλέπουμε ότι το Μαρόκο έχει αυξήσει τις πιθανότητες του σε σχέση με τους προηγούμενους αγώνες.

Πίνακας 1: Πιθανότητες Αποτελεσμάτων για την Ημιτελική φάση 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	Argentina	Croatia	0.580	0.263	0.157
2	France	Morocco	0.503	0.302	0.195

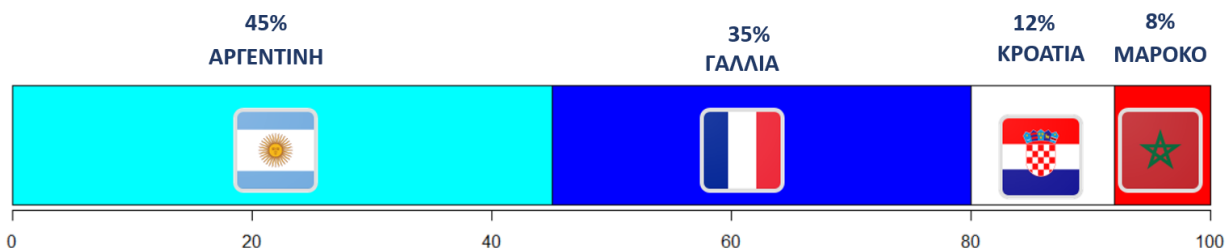
Στο Διάγραμμα 1 μπορείτε να δείτε τις πιθανότητες για το κάθε σκορ για καθένα από τους 2 αγώνες της Ημιτελικής φάσης.

Διάγραμμα 1: Διάγραμμα Πιθανοτήτων πιθανών σκορ για τους Αγώνες της Ημιτελικής Φάσης του Παγκοσμίου Κυπέλλου 2022



Πιθανότητες κατάκτησης του Παγκοσμίου Κυπέλλου

Στη συνέχεια χρησιμοποιήθηκε το μοντέλο για να προσομοιώσουμε πολλά εναλλακτικά σενάρια με βάση τις πιθανότητες που υπολογίστηκαν. Με βάση αυτές τις πιθανότητες η πιθανότητες κατάκτησης του Παγκοσμίου Κυπέλλου 2022 είναι οι ακόλουθες



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

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

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

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

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

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

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

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

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

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

- 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 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 εμπνευση που μας έδωσε ο κος Kesenne.