

**TO**

Want to receive updates from THE STEP?

MONDAY 19 DECEMBER

**TO BHMA**

Login - Register

NO

YES I WANT

## World Cup 2022: Predictions for the quarter-final stage



Predictions are based on Statistical Models of Analytical Soccer. In the round of 16, the model "caught" 7 out of 8 matches.

*The BHIMA Team*

09.12.2022, 11:15

Cookies

**TO**

Want to receive updates from THE STEP?

MONDAY 19 DECEMBER

**TO BHMA**

The 16-team phase ended with the results being as expected from the model and from logic with the exception perhaps of the match between Spain and Morocco where the latter passed and was considered a surprise. [The model caught 7/8 races \(87.5%\) and "missed" Morocco's qualification.](#)

Note that here Morocco had a fairly significant chance of bringing the match to extra time (27%) while the probability of the match going to extra time or Morocco qualifying (cumulatively) was around 44%. This means that if the two teams played 4 times, in about 1 match we would have extra time while if the two teams played 7 matches then in 3 we would either have extra time or Morocco would win. Therefore the model told us that Spain is the favorite but the chances of Morocco succeeding were not negligible.

One of the model's impressive results in this round was it "predicted" a tie between Japan and Croatia, whose match was decided on penalties.

---

[\*Messi calls on... God: "Let's go Diego, from heaven"\*](#)

---

## The Model's Predictions for the Quarter-Final Stage

Updated predictions in the form of odds are given in Table 1. The team with the highest probability of victory is declared as favourite.

**Specifically from Table 1 only**

TO

Want to receive updates from THE STEP?

TO BHMA

MONDAY 19 DECEMBER

the match. So a p  
Morocco beat th

here is: Can

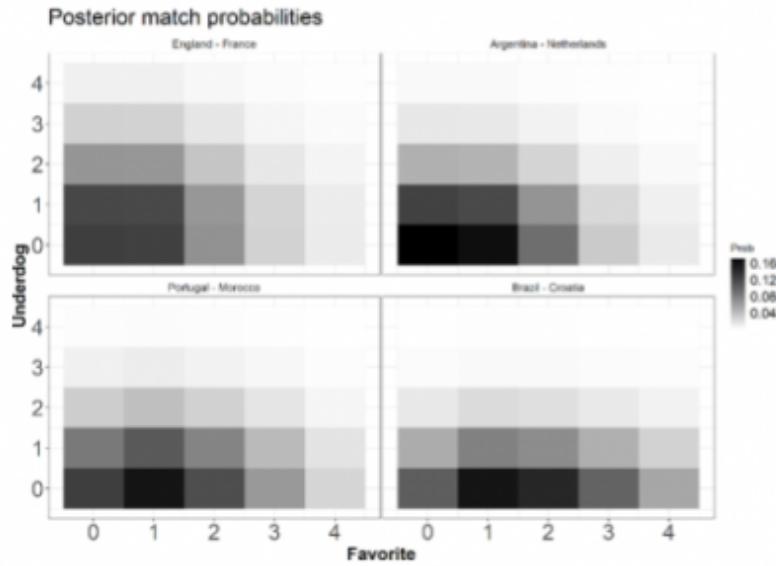
- **Argentina** has a slight lead against the Netherlands with a 41% chance of victory compared to 32% for the second.
- Finally the England-France match is fully balanced and all results are possible.

**Πίνακας 1:** Πιθανότητες Αποτελεσμάτων για την Προημπελική φάση με βάση το Μοντέλο Μπευζανής Στατιστικής Μηχανικής  
Μάθησης της Ερευνητικής ομάδας AUEB Sports Analytics

Αγώνας	Φαβορί	Αουτσάντερ	Νίκη για το φαβορί	Ισοπαλία	Νίκη για το Αουτσάντερ
1	Brazil	Croatia	0.676	0.220	0.104
2	Argentina	Netherlands	0.414	0.264	0.322
3	Portugal	Morocco	0.536	0.191	0.273
4	England	France	0.361	0.292	0.348

Στο Διάγραμμα 1 μπορείτε να δείτε τις πιθανότητες για το κάθε σκορ για καθένα από τους 4 αγώνες της Προημπελικής φάσης.

**Διάγραμμα 1:** Διάγραμμα Πιθανοτήτων πιθανών σκορ για τους Αγώνες της Προημπελική Φάση του Παγκοσμίου Κυπέλου 2022



## Predictive Analysis and Model Comparison (Strictly for beginners only)

In this section, we will present an analysis of the predictive performance of three proposed models, which focus on modeling the goals scored by both opposing teams, across all match days held up to

**TO**

Want to receive updates from THE STEP?

**TO BHMA**

MONDAY 19 DECEMBER

1. Diagonal Infla

2. Bivariate Poiss

3. Simple Poisson (Double Poisson)

Για την προσαρμογή αυτών των μοντέλων, χρησιμοποιήσαμε τα ίδια δεδομένα για την εκπαίδευσή τους, καθώς επίσης και το ίδιο επύπεδο πληροφορίας πριν δούμε τα δεδομένα. Το σύνολο δεδομένων για την αξιολόγηση των επιδόσεων αποτελείται από 56 αγώνες.

Για τη σύγκριση των μοντέλων χρησιμοποιήθηκε ο Μπενζιανός δείκτης LOOIC (Vehtari et al., 2017). Επιγραμματικά, αναφέρουμε ότι το LOOIC του Μοντέλου 1 ήταν το χειρότερο σε κάθε αγωνιστική ημέρα, ενώ το καλύτερο ήταν το τρίτο μοντέλο, αλλά με πολύ μικρή διαφορά σε σχέση με το Μοντέλο 2. Ως επόμενο βήμα, σκεφτήκαμε να συγκρίνουμε την προγνωστική απόδοση των προσαρμοσμένων μοντέλων μας, χρησιμοποιώντας την εκ των υστέρων κατανομή πρόβλεψης της διαφοράς τερμάτων, σε σχέση με τις παρατηρούμενες διαφορές τερμάτων.



**ΤΟ**

Want to receive updates from THE STEP?

**ΤΟ ΒΗΜΑ**

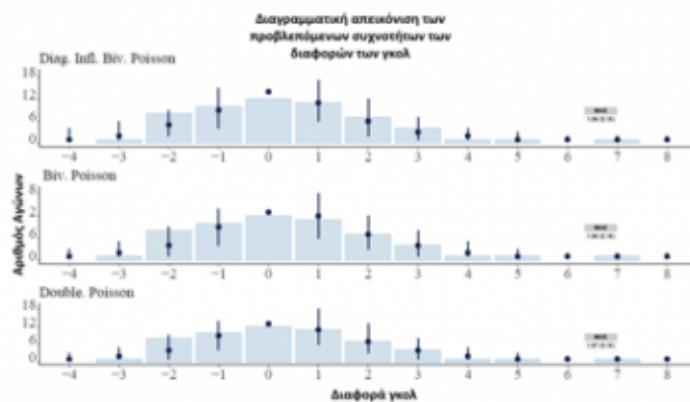
MONDAY 19 DECEMBER

Αυτια τα οιαγραμματων υστέρων διασπατηρούμενες.  
Απόλυτου Σφάλμα συχνοτήτων των προβλέψεων από τις παρατηρούμενες συχνότητες για την κάθε τιμή των διαφορών των γκολ.

ειναι τα γεγονο εκ τις αντίστοιχες «Μέσου δη απόσταση των

Με βάση το Διάγραμμα 2, παρατηρούμε ότι όλα τα προσαρμοσμένα μοντέλα παρουσιάζουν παρόμοιες επιδόσεις τόσο όσον αφορά τη γραφική αναπαράσταση όσο και τη τιμή σφάλματος του ΜΑΕ. Η ομοιότητα μεταξύ των προβλεπτικών επιδόσεων όλων των μοντέλων είναι προφανής από τα αντίστοιχα διαγράμματα, όπου σε όλα αυτά, οι διάμεσοι των προβλέψεων (σκούρα σημεία) είναι κοντά στις παρατηρούμενες συχνότητες. Μια προφανής διαφορά μεταξύ των διαφορετικών προσεγγίσεων είναι ότι το Μοντέλο 1 προβλέπει με αυξημένη πιθανότητα την ισοπαλία (δλδ τη διαφορά τερμάτων ίση με μηδέν) σε σύγκριση με τα Μοντέλα 2 και 3.

**Διάγραμμα 2:** Σύγκριση παρατηρούμενων και αναμενόμενων αριθμών αγώνων για την κάθε τιμή διαφοράς γκολ (για τους Αγώνες μέχρι τη Φάσης των 16 του Παγκοσμίου Κυπέλου 2022).



Οι γαλάζιοι ράθοι απεικονίζουν τον παρατηρούμενο αριθμό αγώνων για κάθε τιμή της διαφοράς των γκολ. Τα μπλε σκούρα σημεία απεικονίζουν την (εκ-των-υστέρων) διάμεσο πρόβλεψης ενώ οι συνοδευτικές μπλε μπάρες σφαλμάτων απεικονίζουν το 95% διάστημα των προβλέψεων.

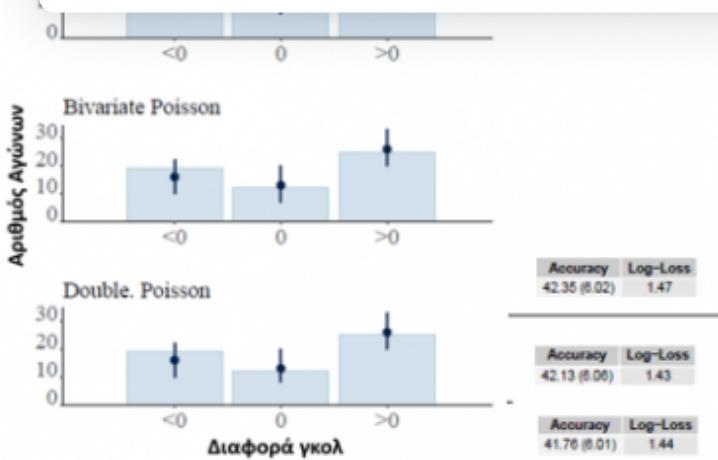
Γενικά, τα αποτελέσματα του Διαγράμματος 2 δείχνουν ότι και τα τρία μοντέλα δίνουν παρόμοιες προβλέψεις με μικρό προβάδισμα στα Μοντέλα 1 και 2. Για το λόγο αυτό επικεντρωθήκαμε

**ΤΟ**

Want to receive updates from THE STEP?

**ΤΟ ΒΗΜΑ**

MONDAY 19 DECEMBER



Οι γαλάζιοι ράβδοι απεικονίζουν τον παρατηρούμενο αριθμό αγώνων για κάθε τιμή της διαφοράς των γκολ. Τα μπλε σκουρά σημειώνουν την (εκ-των-υστέρων) διάμεσο πρόβλεψης ενώ οι συνοδευτικές μπλε μπάρες σφαλμάτων απεικονίζουν το 95% διάστημα των προβλέψεων.

Πιο συγκεκριμένα, οι μικρές διαφορές των αποτελεσμάτων LOO και MAE στο Διάγραμμα 2 μας οδήγησαν να συγκρίνουμε τα μοντέλα μας με βάση την προγνωστική τους απόδοση μόνο για το τελικό αποτέλεσμα του αγώνα χωρίς να λαμβάνουμε υπόψη την ακριβή διαφορά τερμάτων των ομάδων (βλ. Διάγραμμα 3). Δεδομένου ότι τώρα το ενδιαφέρον μας επικεντρώνεται σε τρία πιθανά αποτελέσματα (νίκη της γηπεδούχου ομάδας, ισοπαλία, νίκη της φιλοξενούμενης ομάδας), θα μετρήσουμε την προβλεπτική απόδοση των μοντέλων με βάση δύο δείκτες: (α) την «Ακρίβεια» (accuracy) και (β) τον δείκτη multi-class log-loss (ή αλλιώς multi-class cross entropy – για περισσότερες λεπτομέρειες σχετικά με με αυτό το μέτρο παραπέμπουμε στην ερευνητική εργασία των Grandini, M., et. al, 2020). Αυτό το μέτρο είναι το καταλληλότερο για μοντέλα πρόβλεψης κατηγορικών αποτελεσμάτων, δεδομένου ότι ποσοτικοποιεί πόσο κοντά είναι οι πιθανοθεωρητικές προβλέψεις μας σε σχέση με τα παρατηρούμενα τελικά αποτελέσματα των αγώνων. Τα αποτελέσματα και των δύο αυτών δεικτών δίνονται στο Διάγραμμα 3 μαζί με το αντίστοιχο διάγραμμα των συχνοτήτων των αποτελεσμάτων. Σε αντίθεση με το Διάγραμμα 2, εδώ τα δύο μέτρα υποδεικνύουν ότι το απλούστερο μοντέλο της διμεταβλητής Poisson (δηλαδή το Μοντέλο 2) είναι ελαφρώς καλύτερο από το Μοντέλο 1.

Συμπερασματικά, οι προβλέψεις μας για τους υπόλοιπους αγώνες του Παγκοσμίου Κυπέλλου θα βασίζονται πλέον στο Μοντέλο 2 (Bivariate Poisson). Η συγκεκριμένη αλλαγή στη στρατηγική μοντελοποίησης μας βασίζεται στους ακόλουθους λόγους:

**ΤΟ**

Want to receive updates from THE STEP?

**ΤΟ ΒΗΜΑ**

MONDAY 19 DECEMBER

- Το απλό διμεταβολικό μέτρο που σημειώνει την καθοριστική απόδοση των προσαρμοσμένων μοντέλων

Παρ' όλα αυτά, ένας αναλυτής θα λαμβάνει μεριμνή αποτελεσμάτων χρησιμοποιώντας υποιαδήποτε από τα δύο πρώτα μοντέλα για την πρόβλεψη των τερμάτων που σημείωσαν δύο αντίπαλες ομάδες καθώς και για τη διαφορά τερμάτων σε κάθε αγώνα.

Μέχρι το τέλος του τουρνουά, καθώς θα έχουμε όλο και περισσότερους αγώνες στο δείγμα μας, θα προχωρήσουμε σε μια πιο λεπτομερή αξιολόγηση της προβλεπτικής απόδοσης των προσαρμοσμένων μοντέλων χρησιμοποιώντας πρόσθετα προγνωστικά μέτρα.

### **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.

**ΤΟ**

Want to receive updates from THE STEP?

**ΤΟ ΒΗΜΑ**

MONDAY 19 DECEMBER

$$\log(\lambda_{1i}) = \alpha$$

$$\log(\lambda_{2i}) = \beta$$

$$\log(\lambda_{3i}) = \rho_i$$

$$\text{att}_{k,t} \sim \mathcal{N}$$

$$\text{def}_{k,t} \sim \mathcal{N}_{\lambda_{3i} h_i t = 1, \sigma^2},$$

$$\rho_i, \gamma \sim \mathcal{N}(0, 1)$$

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

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

- $i$  είναι ο δείκτης του αγώνα
- $X_i$  και  $\bar{Y}_i$  είναι ο αριθμός των γκολ μεταξύ της 1<sup>η</sup> και της 2<sup>η</sup> ομάδας στον αγώνα  $i$ .
- $h_i$  και  $a_i$  είναι η 1<sup>η</sup> και 2<sup>η</sup> 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)
- $\text{ranking}_k$  δείκτης Coca-Cola FIFA ranking την 6<sup>η</sup> 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 μοντελοποιηση και αναλυτική ποδοσφαίρου και έντονη ερευνητική δραστηριότητα στη Μπευζιανή Στατιστική μεθοδολογία.

Ο **Basileios Palloskas** 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 **Iωάννης 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 Ντζουφρα. Έχει αναγνωρισμένη επιστημονική δραστηριότητα σε τομείς όπως η στατιστική μεθοδολογία, υπολογιστική στατιστική, Βιοστατιστική, και αναλυτική των σπορ.

**ΤΟ**

Want to receive updates from THE STEP?

**ΤΟ ΒΗΜΑ**

MONDAY 19 DECEMBER

workshop της ομάδας

**Ο L. Egidi και B. Paláskas** 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.

**\* 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 Βασίλειου Παλάσκα (Fantasy Sports Interactive) με τις συμβουλευτικές συνδρομές των I. Ντζούφρα και Δ. Καρλή. Και οι

Cookies

**TO**

Want to receive updates from THE STEP?

**TO BHMA**

MONDAY 19 DECEMBER



... ΤΙ ΕΙΔΗΣΕΙΣ ΕΙΝΑΙ ΤΩΡΑ...

**Οικονομία**

Τέλη κυκλοφορίας 2023: Δίμηνη παράταση προανήγγειλε ο Χρήστος Σταϊκούρας

Ακολουθήστε **TO BHMA** στο **Google News** και μάθετε πρώτοι όλες τις ειδήσεις

Δείτε όλες τις τελευταίες **Ειδήσεις** από την Ελλάδα και τον Κόσμο, από **TO BHMA**

Δείτε επίσης **MOYNTIAΛ 2022**

**MUST R E A D**PRESENTED BY **TO BHMA**

**Εύα Καϊλή:** Η άνοδος, ο πειρασμός και η έκπτωση



**Σβήνουν τις γκρίζες ζώνες νότια της Κρήτης - Οι ελληνικ...**



**Εκλογές: Μετρούν κάστρα και μαύρες τρύπες**



**Ανεργία: Οι «πρωταθλήτριες» περιοχές**

**TO BHMA****in****TA NEA****OT**

**ΤΟ**

MONDAY 19 DECEMBER

CTN

Want to receive updates from THE STEP?

**TO BHMA****TO BHMA**

## Το παρασκήνιο της απόφασης Μπαστάκη για τη Σπυράκη

19.12.2022 - 07:00



Φριτέζα Αέρος Xiaomi Mi  
Αποσπώμενο Κάδο 3.5lt 1  
Λευκό

**89.39 €**

99.90 €

ΑΓΟΡΑΣΕ ΤΟ

**TO**

MONDAY 19 DECEMBER

**TO BHMA**

Want to receive updates from THE STEP?

Cookies