



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

Τεχνική Αναφορά Ευφυείς Πράκτορες Σεπτέμβριος 2025	
<i>Ομάδα</i>	
Δημήτριος Φουσέκης	Π21181
Μάριο Πέτο	Π21138
Πολυχρονόπουλος Παρασκευάς	Π21141
Νικόλαος Κονιδάρης	Π21063

Περιγραφή συστήματος

Το σύστημα που αναπτύχθηκε αποσκοπεί στην πρόβλεψη αποτελεσμάτων ποδοσφαιρικών αγώνων μέσω της ολοκληρωμένης αξιοποίησης στατιστικών δεδομένων, μηχανικής μάθησης και πολυπρακτορικής επικοινωνίας (multi-agent communication). Η βασική του καινοτομία έγκειται στον συνδυασμό πολλαπλών πηγών δεδομένων με σύγχρονες μεθοδολογίες ανάλυσης και μοντελοποίησης, ώστε να παραχθούν όσο το δυνατόν πιο αξιόπιστες εκτιμήσεις για τις πιθανότητες νίκης, ισοπαλίας και ήττας σε κάθε παιχνίδι.

Σε πρώτο στάδιο, το σύστημα αντλεί ένα ευρύ φάσμα στατιστικών στοιχείων που σχετίζονται με τις ομάδες και τα πρόσφατα αποτελέσματά τους. Τα δεδομένα αυτά μπορεί να περιλαμβάνουν την επιθετική και αμυντική απόδοση, τον αριθμό τερμάτων υπέρ και κατά, τη φόρμα των παικτών, αλλά και πιο προχωρημένους δείκτες όπως το Expected Goals (xG). Οι πληροφορίες συλλέγονται από διαφορετικές πηγές και συγκεντρώνονται σε ενιαία μορφή για περαιτέρω επεξεργασία.

Προτού αξιοποιηθούν για ανάλυση, τα δεδομένα υφίστανται κανονικοποίηση, ώστε να εξαλείφονται οι ασυμβατότητες και να διασφαλίζεται η συγκρισιμότητα μεταξύ διαφορετικών σετ αγώνων και ομάδων. Έτσι, η βάση δεδομένων του συστήματος παρέχει σταθερό υπόβαθρο για την εκπαίδευση και την αξιολόγηση των μοντέλων μηχανικής μάθησης.

Κεντρικό ρόλο στο σύστημα κατέχει η χρήση αλγορίθμων μηχανικής μάθησης, και συγκεκριμένα του XGBoost classifier, ο οποίος έχει εκπαιδευτεί με βάση τα κανονικοποιημένα στατιστικά στοιχεία.

Το μοντέλο XGBoost, το οποίο εξειδικεύεται στην εκτίμηση πιθανοτήτων για κάθε πιθανή έκβαση (νίκη, ισοπαλία, ήττα) δημιουργεί την πρώτη εκδοχή πιθανοτήτων που χρησιμοποιεί το σύστημα.

Για την περαιτέρω βελτίωση των αποτελεσμάτων και την αποτελεσματική διαχείριση της ροής δεδομένων, το σύστημα αξιοποιεί μοντέλα Claude της Anthropic, τα οποία εκτελούνται στο cloud. Συγκεκριμένα:

Δύο μοντέλα Claude Haiku χρησιμοποιούνται για απλές διαδικασίες κλήσης μοντέλων (model calling), εξασφαλίζοντας ταχύτητα και χαμηλή υπολογιστική επιβάρυνση.

Δύο μοντέλα Claude Sonnet χρησιμοποιούνται σε διαδικασίες που σχετίζονται με την είσοδο και έξοδο δεδομένων (I/O), εξασφαλίζοντας ακρίβεια στην παρουσίαση και ανάλυση των αποτελεσμάτων.

Η συνεργασία αυτών των μοντέλων επιτρέπει στο σύστημα να συνδυάζει τον αμιγώς στατιστικό υπολογισμό των XGBoost classifiers με την επεξηγηματική και επαγωγική ικανότητα των LLMs.

Η ολοκλήρωση της πολυπρακτορικής αρχιτεκτονικής επιτυγχάνεται με την αξιοποίηση του in-memory graph state του LangGraph, το οποίο λειτουργεί ως μέσο επικοινωνίας μεταξύ των διαφορετικών agents. Κάθε agent μπορεί να αναλάβει συγκεκριμένο ρόλο στη διαδικασία: ένας agent συλλέγει δεδομένα, άλλος τα κανονικοποιεί, ένας τρίτος επικοινωνεί με τα XGBoost μοντέλα, ενώ τα Claude μοντέλα συνθέτουν και παρουσιάζουν τα αποτελέσματα.

Το in-memory graph state επιτρέπει την αποδοτική ανταλλαγή πληροφορίας χωρίς την ανάγκη βαριάς αποθήκευσης σε μόνιμη βάση δεδομένων για κάθε βήμα, καθιστώντας το σύστημα πιο ευέλικτο και ταχύτερο. Έτσι, η όλη διαδικασία μπορεί να ολοκληρωθεί σχεδόν σε πραγματικό χρόνο, ακόμη και όταν ο όγκος δεδομένων είναι σημαντικός.

Στο τελικό στάδιο, τα κανονικοποιημένα στατιστικά δεδομένα, οι εκτιμήσεις των XGBoost classifiers και οι πληροφορίες από τα Claude μοντέλα ενοποιούνται. Το σύστημα καταλήγει σε τελικές πιθανότητες για νίκη, ισοπαλία ή ήττα, λαμβάνοντας υπόψη τόσο τη στατιστική βάση όσο και τις αποκλίσεις που εντοπίζονται σε δεδομένα αποδόσεων από εξωτερικά APIs και συγκρίνοντας τα με τα αρχικά δεδομένα.

Με τον τρόπο αυτόν, οι τελικές προβλέψεις δεν βασίζονται αποκλειστικά σε μία μόνο μέθοδο, αλλά αποτελούν συνδυαστική εκτίμηση που ενισχύει την ακρίβεια και την αξιοπιστία τους και μπορούν να επιστρέφουν με μεγαλύτερη ακρίβεια πιθανές αγορές για τον χρήστη που θα έχουν μεγάλη πιθανότητα εκπλήρωσης.

Τεχνολογίες και αρχιτεκτονική

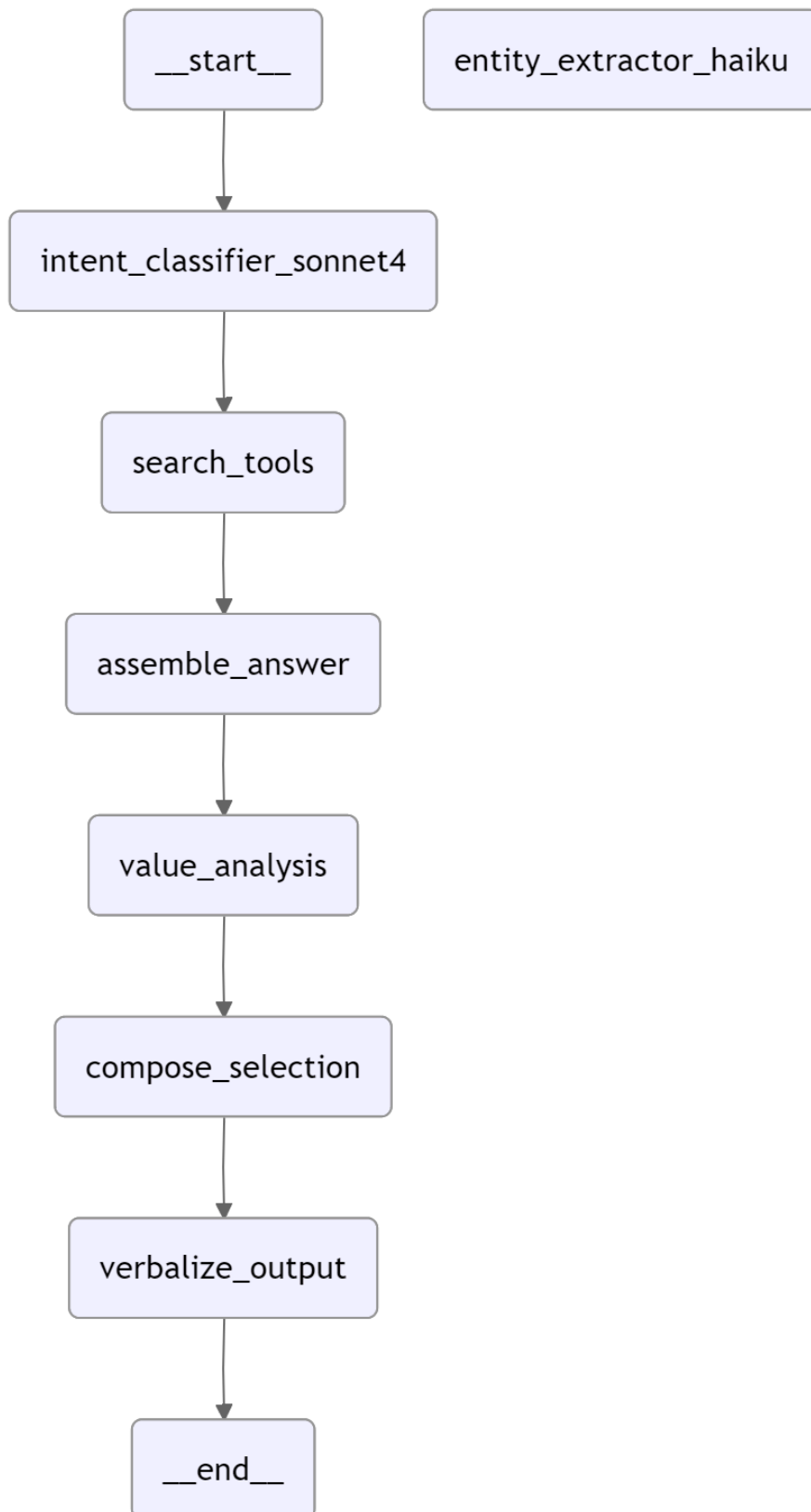
Για την υλοποίηση της εφαρμογής χρησιμοποιήθηκε LangGraph και python. Σε αυτή κάνουμε κλήσεις στο API-FOOTBALL με την βασική βιβλιοθήκη της python για http request. Επιπλέον χρησιμοποιήθηκε η πλατφόρμα του kaggle στην οποία με scikit learn υλοποιήθηκε ένας XGBoost classifier για την αξιολόγηση των αποδόσεων των εξεταζόμενων αγορών.

Στην αρχιτεκτονική εντάχθηκαν και γλωσσικά μοντέλα της Anthropic, τα οποία αξιοποιούνται μέσω cloud. Συγκεκριμένα, δύο μοντέλα Claude Haiku χρησιμοποιούνται για απλές κλήσεις μοντέλων (model calling), εξασφαλίζοντας γρήγορη απόκριση και χαμηλό υπολογιστικό κόστος. Παράλληλα, δύο μοντέλα Claude Sonnet διαχειρίζονται διεργασίες που σχετίζονται με την είσοδο και έξοδο δεδομένων (I/O), ώστε να διασφαλίζεται η ομαλή ροή πληροφοριών και η αξιόπιστη επικοινωνία με τα υπόλοιπα υποσυστήματα.

Η επικοινωνία μεταξύ των agents πραγματοποιείται με τη χρήση του in-memory graph state του LangGraph, που λειτουργεί ως κεντρικός μηχανισμός ανταλλαγής δεδομένων. Με αυτόν τον τρόπο οι agents συνεργάζονται σε πραγματικό χρόνο, αποφεύγοντας την ανάγκη συνεχούς αποθήκευσης σε μόνιμη βάση δεδομένων. Αυτό καθιστά το σύστημα πιο ευέλικτο, αποδοτικό και κατάλληλο για δυναμικές ροές δεδομένων.

Συνολικά, η ροή δεδομένων ξεκινά με την άντληση στατιστικών από το API-FOOTBALL, συνεχίζεται με την κανονικοποίηση και ανάλυση τους από τον XGBoost classifier, και ολοκληρώνεται με τη συνδυαστική αξιολόγηση και παρουσίαση των αποτελεσμάτων μέσω των Claude μοντέλων. Ο LangGraph λειτουργεί ως συνδετικός κρίκος όλων των επιμέρους τμημάτων, επιτρέποντας στο σύστημα να λειτουργεί με συνέπεια και να παράγει αξιόπιστες προβλέψεις.

Διαγράμματα Ροής



Προβλήματα που αντιμετωπίστηκαν

Τα προβλήματα που αντιμετωπίστηκαν αφορούν κυρίως στην συλλογή δεδομένων, αρχικά για από την πλευρά του μοντέλου ώστε να μπορεί να χρησιμοποιήσει τα εργαλεία και να εξαγάγει τα καλύτερα αποτελέσματα, και στην συνέχεια στην επιλογή επισημασμένων δεδομένων παρόμοιων με αυτά που συλλέγονται από τους πράκτορες ώστε να εκπαιδευτεί ο ταξινομητής ο οποίος είναι υπεύθυνος για την αξιολόγηση των αγώνων από τα μοντέλα.

Το πρώτο πρόβλημα οφείλεται στην μη ύπαρξη δωρεάν API με δεδομένα για ποδοσφαιρικούς αγώνες που χρειαζόμασταν. Στον αρχικό μας σχεδιασμό θέλαμε να συμπεριλάβουμε advanced metrics όπως xG(Expected Goals), xGa(Expected goals against) ή χρονικές κατανομές στις οποίες μία ομάδα σκοράρει ή δέχεται γκολ. Αυτά τα δεδομένα υπήρχαν σε μερικά APIs αλλά πίσω από απαγορευτικές τιμές, εν τέλει καταλήξαμε στο API-FOOTBALL το οποίο είχε την καλύτερη σχέση ποιότητας/τιμής που μπορούσαμε να πληρώσουμε. Σε αυτό το σημείο αξίζει να αναφέρουμε ότι πολλά δωρεάν πλάνα για access API ήταν πολύ απαγορευτικά αφού μερικά είχαν όρια μικρότερα από 100 request το μήνα, τα οποία δεν φτάνουν ούτε στο ελάχιστο καν για εκσφαλμάτωση.

Το δεύτερο πρόβλημα της έλλειψης δεδομένων έγκειται στο ότι μόνοι μας φτιάξαμε ένα feature schema με τα διαθέσιμα δεδομένα από το API-FOOTBALL το οποίο δεν ακολουθείται από πουθενά αλλού. Αυτό το πρόβλημα προσπεράστηκε σε αυτό το notebook, κάνοντας χειροκίνητα την μετατροπή δεδομένων από ένα, πολύ ποιοτικό, σετ δεδομένων της Πρέμιερ Λιγκ(σε αυτό το notebook του Kaggle:

<https://www.kaggle.com/code/fousekhs/notebook036a180296>).

Ένα μικρότερο πρόβλημα που αντιμετωπίζουμε ήταν η έλλειψη υπολογιστική δύναμης στους υπολογιστές μας, την οποία προσπεράσαμε με την χρήση του cloud της anthropic.

Αξιολόγηση/δοκιμή με διαφορετικά user inputs

```
PS C:\Users\dimit\Documents\intelligent_agents> python main.py "Chelsea v Brighton on saturday"
C:\Users\dimit\AppData\Local\Programs\Python\Python310\lib\site-packages\pydantic\main.py:519: UserWarning: PydanticSerializationUnexpectedValue(Expected `dict[str, any]` - serialized value may not be as expected)
  return self.__pydantic_serializer__.to_json(
C:\Users\dimit\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:442: InconsistentVersionWarning: InconsistentVersionWarning: The persistence format used to save this model was changed in 1.2. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
  warnings.warn(
✅ Match packet ready.
- Fixture resolved: yes
- Pre-match odds: available
- H2H size: 10
- Lineups: 0

Value analysis complete for 1 fixtures.

Chelsea Win - Delta: 0.107, Odds: 1.94 (1xBet)

Finding the best value market with target achieved.

• Chelsea vs Brighton – H ( $\Delta=+0.107$ ) @ 1.94 1xBet

Combined odds: 1.94
```

Όταν ο χρήστης δίνει έγκυρο input, όπως στην περίπτωση του αγώνα *“Chelsea v Brighton on Saturday”*, το σύστημα επεξεργάζεται το αίτημα, αντλεί τα αντίστοιχα δεδομένα από το API και επιστρέφει το πιο πιθανό αποτέλεσμα: νίκη, ισοπαλία ή ήττα με αποδοχή για αγορά στοιχηματισμού. Με αυτόν τον τρόπο αποδεικνύεται ότι η ροή δεδομένων και η κανονικοποίηση λειτουργούν κανονικά και το σύστημα παρέχει έγκυρες προβλέψεις για συγκεκριμένα παιχνίδια.

```
PS C:\Users\dimit\Documents\intelligent_agents> python main.py "give me x3 acca from this saturdays premier league"
C:\Users\dimit\AppData\Local\Programs\Python\Python310\lib\site-packages\pydantic\main.py:519: UserWarning: PydanticSerializationUnexpectedValue(Expected `dict[str, any]` - serialized value may not be as expected [input_value=...]
  return self.__pydantic_serializer__.to_json(
C:\Users\dimit\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:442: InconsistentVersionWarning: InconsistentVersionWarning: The persistence format used to save this model was changed in 1.2. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
  warnings.warn(
✅ Packets ready: 8 (including 7 matches)

Value analysis complete for 7 fixtures.

• Crystal Palace vs Liverpool: Draw ( $\Delta=+0.164$ , 3.85)
• Nottingham Forest vs Sunderland: Away ( $\Delta=+0.145$ , 4.82)
• Chelsea vs Brighton: Home ( $\Delta=+0.107$ , 1.94)

Product: 36.00

3x accumulator target achieved.

• Crystal Palace vs Liverpool – D ( $\Delta=+0.164$ ) @ 3.85 188Bet
• Nottingham Forest vs Sunderland – A ( $\Delta=+0.145$ ) @ 4.82 Pinnacle
• Chelsea vs Brighton – H ( $\Delta=+0.107$ ) @ 1.94 1xBet

Combined odds: 36.00
```

Το σύστημα ανταποκρίνεται επίσης σε πιο σύνθετα αιτήματα, όπως στην περίπτωση όπου ο χρήστης ζητά *“give me x3 acca from this saturdays premier league”*. Εδώ, οι agents συνεργάζονται, αξιοποιώντας τόσο τα κανονικοποιημένα στατιστικά όσο και τα αποτελέσματα των μοντέλων, ώστε να συνθέσουν και να επιστρέψουν τις πιο πιθανές τριάδες αγώνων. Αυτό δείχνει τη δυνατότητα της αρχιτεκτονικής να υποστηρίζει πρακτικά σενάρια χρήσης και πιο πολύπλοκες απαιτήσεις εισόδου.

```
PS C:\Users\dimit\Documents\intelligent_agents> python main.py "What colour is chelsea wearing?"
C:\Users\dimit\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:442: InconsistentVersionWarning: Trying to unpickle estimator SimpleImputer from version 1.2.2 when using version 1.7.2. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
  warnings.warn(
C:\Users\dimit\AppData\Local\Programs\Python\Python310\lib\site-packages\xgboost\core.py:729: UserWarning: [22:51:32] WARNING: C:\actions-runner\_work\xgboost\xgboost\src\common\error_msg.cc:52: Empty dataset at worker: 0
  return func(**kwargs)
Unfortunately I can't help with that. I only handle best value markets or xN bet requests scoped to a league, group of leagues, batch, or a single match.

Value analysis complete for 0 fixtures.

No value bets identified in current market scan.
```

Το σύστημα διαθέτει ενσωματωμένους μηχανισμούς ασφαλείας (guardrails), οι οποίοι ενεργοποιούνται όταν ο χρήστης εισάγει μη σχετικές, άσχετες ή ακατάλληλες πληροφορίες. Στην περίπτωση αυτή, όπως φαίνεται στην εικόνα, το σύστημα αναγνωρίζει το τυχαίο input *“What colour is chelsea wearing?”* και επιστρέφει μήνυμα ότι δεν μπορεί να βοηθήσει με αυτό, προστατεύοντας έτσι τόσο τη λειτουργικότητα όσο και την αξιοπιστία της εφαρμογής.