



II° Corso di Football Data Intelligence

Passing Style Analyser

Progettazione ed implementazione di un'applicazione web interattiva per lo studio dello stile di gioco delle squadre della Serie A 2021/22 attraverso l'analisi dettagliata delle reti di passaggi

Autore

Erasmo Purificato

30 Novembre 2022

Indice

1	Introduzione	3
2	Risorse e concetti preliminari	8
2.1	Lavori di ispirazione	8
2.2	Pass Network	9
2.3	Field Tilt	10
2.4	Normalizzazione	11
3	Passing Style Analyser	12
3.1	Supporto tecnologico	12
3.2	Preparazione del dataset	13
3.2.1	Estrazione di informazioni dal campo <i>qualifier</i>	13
3.2.2	<i>dec_time</i>	14
3.2.3	<i>pass_recipient</i>	14
3.2.4	<i>team_ids.json</i>	15
3.2.5	<i>player_start_pos</i> e <i>pass_recip_start_pos</i>	16
3.2.6	<i>match_status</i>	16
3.2.7	<i>teams_lineups.json</i> e <i>teams_match_status.json</i>	16
3.2.8	Filtraggio del dataset	17
3.3	Componenti dell’interfaccia grafica	17
3.3.1	Home Page	17
3.3.2	Full Pitch & Field Tilt Networks	18
3.3.3	Match Status Networks	19

3.4	Visualizzazione dei risultati	21
3.4.1	Full Pitch & Field Tilt Networks	21
3.4.2	Match Status Network	27
4	Conclusione e Sviluppi Futuri	31
A	Formations Explained	32

Capitolo 1

Introduzione

Il possesso palla, negli ultimi anni, ha acquisito una notevole importanza nel modello di gioco offensivo di molte squadre di calcio. Questa circostanza è stata favorita, tra i tanti fattori, soprattutto dal successo di squadre come il Barcellona, il Manchester City, il Bayern Monaco o nazionali come la Spagna e il Portogallo, che basano il proprio modello offensivo sul controllo del gioco, attraverso appunto il possesso palla.

I progressi tecnologici delle ultime decadi, e la conseguente possibilità di avere a disposizione una mole di dati considerevole circa le performance individuali e di squadra, hanno contribuito, nel recente periodo, alla rapida e continua evoluzione e diffusione della *Football Analytics*. In questo contesto, uno degli aspetti più studiati è proprio il possesso palla, ed il modo in cui le singole squadre organizzano questa fase di gioco, sempre più predominante.

Tra il 2010 e il 2020, la maggior parte delle analisi è stata incentrata sulla ricerca di una potenziale correlazione tra volume di possesso palla e risultati sportivi. Molti studi hanno dimostrato che la percentuale di possesso palla delle squadre vincenti era superiore a quella delle squadre non vincenti. Ad esem-

pio, durante il torneo della Coppa del Mondo 2010, la nazionale spagnola ha vinto il campionato con la più alta percentuale di possesso palla medio (66,3%), così come la nazionale tedesca ha vinto la Coppa del Mondo 2014 facendo registrare la più alta percentuale di possesso palle medio del 56,7%. Tuttavia, il ruolo della percentuale di possesso nell’analisi degli indicatori tecnici si è gradualmente indebolito, complice il fatto che lo sviluppo di questo specifico stile di gioco (o quantomeno il tentativo) è diventato piuttosto comune, a tutti i livelli. Diversi studi hanno infatti sottolineato che tale indicatore non riflette la situazione reale del gioco, e valori maggiori dello stesso non sono correlati con maggiori opportunità di tiro e di punteggio. Nella Coppa del Mondo 2018, ad esempio, la nazionale francese ha vinto la Coppa del Mondo con una percentuale di possesso palla inferiore al 50% per partita.

Le analisi correnti cercano invece di stabilire la relazione tra le caratteristiche del possesso palla e le prestazioni della squadra. Alcuni studi hanno indagato la relazione tra il possesso palla e gli indicatori tecnici (ad esempio, passaggi, tiri e gol) e fisici (ad esempio, distanza totale di corsa, distanza di corsa ad alta intensità, distanza di corsa con e senza palla). Altri sfruttano invece le enormi potenzialità della *data visualisation* per provare ad analizzare lo stile di gioco di una determinata squadra attraverso il modo in cui la fase di possesso palla viene organizzata. A tal proposito, uno degli strumenti maggiormente utilizzati è quello delle *pass network* (reti di passaggi).

Le reti di passaggi possono essere considerate tra le tipologie di visualizzazioni di dati più familiari nel contesto dell’analisi dati nel calcio. Questo strumento (definito anche come metrica in alcuni casi) mostra i punti (ovvero

le posizioni in campo) in cui la squadra effettua il maggior numero di passaggi, i calciatori coinvolti in tali passaggi, quali sono i calciatori che effettuano il maggior numero di passaggi ed in quale posizione. Diversi noti provider (come ad esempio Soccerment¹) o blogger indipendenti rendono disponibili pubblicamente le reti di passaggi di moltissime partite. Come accade, però, in diverse situazioni in questo dominio, nonostante la lettura e la comprensione delle reti di passaggi è piuttosto intuitiva, la loro realizzazione, o semplicemente la possibilità di modificarne i parametri per poter visualizzare aspetti differenti, risulta particolarmente complesso per un utente medio o addetto ai lavori senza competenze informatiche.

Dato lo scenario descritto, come progetto finale per il II° Corso di Football Data Intelligence, viene presentata **Passing Style Analyser**, un'applicazione web interattiva progettata per lo studio dello stile di gioco delle squadre attraverso l'analisi dettagliata delle reti di passaggi. Obiettivo principale del presente lavoro è quello di fornire uno strumento fruibile da tutti gli utenti, attraverso il quale poter personalizzare in tempo reale la visualizzazione delle reti di passaggi per diverse tipologie di analisi.

In particolare, i contributi innovativi della applicazione web pubblica interattiva per la visualizzazione delle reti di passaggi per la Serie A 2021/22 presentata in questo elaborato, possono essere riassunti come segue:

- visualizzazione delle reti di passaggi secondo diverse viste comparative: singola squadra, due squadre o tutte le squadre del campionato per una singola giornata o un intervallo di giornate (consecutive);

¹<https://twitter.com/SoccermentBlog/status/1581924042680209408>

- possibilità di analisi delle formazioni (iniziali) maggiormente utilizzate da ogni squadra nel corso di tutto il campionato o per un intervallo specifico di giornate (consecutive);
- possibilità di visualizzazione delle reti di passaggio a campo intero o “*field tilt*” (la definizione precisa verrà data nel prossimo capitolo);
- rappresentazione della fase di possesso palla di ogni squadra per ogni condizione di partita (vittoria, pareggio, sconfitta) per una singola giornata o un intervallo di giornate (consecutive);
- calcolo del baricentro di ogni squadra e per ogni condizione di partita, con la possibilità di includere o meno la posizione del portiere nel calcolo (e successiva visualizzazione);
- visualizzazione delle reti di passaggi in differenti stili grafici, per mostrare diverse potenzialità dello strumento, in vista di futuri perfezionamenti ed ulteriori personalizzazioni.

Il contenuto dell’elaborato si articherà secondo la seguente struttura: nel secondo capitolo verranno descritti dei lavori da cui è stata presa ispirazione e definiti alcuni dei concetti fondamentali alla base della applicazione sviluppata. Il Capitolo 3 riguarderà la descrizione dell’applicazione Passing Style Analyser, partendo dalla preparazione del dataset fino allo sviluppo delle due funzionalità principali. Nel quarto capitolo verranno esaminati alcuni dei risultati preliminare derivanti da un caso di studio condotto personalmente, incluse limitazioni e problematiche riscontrate nell’utilizzo del dataset fornito. Infine,

nel Capitolo 5 verranno discusse le conclusioni e illustrati alcuni spunti per potenziali sviluppi futuri.

Passing Style Analyser è accessibile al seguente link:

<https://erasmopurif-passing-style-analysis--home-xkeymq.streamlit.app/>.

È necessario, comunque, specificare che si tratta di un prototipo sviluppato come lavoro finale per il II° Corso di Football Data Intelligence, senza alcuna presunzione di essere scevro da errori. Essendo il primo progetto open source di questo genere, può costituire una solida base per potenziali future analisi che includono l'analisi delle reti di passaggi.

Capitolo 2

Risorse e concetti preliminari

In questo capitolo verranno presentati risorse che hanno costituito fonte di ispirazione per questo lavoro e definiti diversi concetti fondamentali per la comprensione del lavoro sviluppato.

2.1 Lavori di ispirazione

La principale fonte di ispirazione del progetto presentato in questo elaborato è rappresentato dall'articolo pubblicato da *The Athletic* il 7 Giugno 2022 dal titolo “*How each Premier League team pass*”². Il punto focale della pubblicazione, che mostra le reti di passaggi di tutte le squadre della Premier League 2021/22, risiede nell'analisi e discussione della necessità ed utilità nel rappresentare reti di passaggi “cumulative”, ovvero che permettono la visualizzazione di questa particolare metrica tenendo in conto più giornate di campionato e non una singola, in modo da poter avere un'idea generale di come ogni squadra (o allenatore) gestisce la propria fase di possesso palla in un periodo prolungato di

²<https://theathletic.com/3346644/2022/06/07/premier-league-pass-networks/>

tempo, al fine di evidenziare eventuali pattern ricorrenti, che è uno degli scopi principali di ogni attività di analytics.

Da ulteriori approfondimenti, non sono emerse evidenze di analisi simili effettuate di recente per il campionato di Serie A, ed in aggiunta alla possibilità di mettere a disposizione uno strumento interattivo utilizzabile da tutti gli utenti, indipendentemente dalle conoscenze tecniche, questi elementi hanno costituito il motore per lo sviluppo di **Passing Style Analyser**.

2.2 Pass Network

Come brevemente menzionato nel capitolo introduttivo, le reti di passaggi sono ad oggi una delle tipologie di visualizzazioni più comuni nell'ambito della *Football Analytics*. Costituiscono uno strumento prezioso per fornire una panoramica dell'approccio di una squadra in fase di possesso: la sua forma media, l'importanza relativa dei diversi giocatori, i fulcri del gioco e la mole di passaggi sviluppata, rappresentando di fatto una versione idealizzata dello stile di gioco espresso in campo da una squadra.

Più nello specifico, una rete di passaggi è formata da:

- Circonferenze (o nodi) all'interno del campo da gioco, raffiguranti i calciatori posizionati nella posizione media rispetto alle proprie coordinate dei passaggi completati (ovvero effettuati con successo). La dimensione di tali circonferenze rappresenta il valore di una particolare metrica scelta, come ad esempio numero di passaggi effettuati da quel calciatore (caso standard), o numero di tocchi.

- Linee (o archi) tra coppie di calciatori, che rappresentano i passaggi effettuati tra i due calciatori collegati. Per ogni coppia di calciatori, ci può essere una singola linea bidirezionale, definita sommando i passaggi in entrambe le direzioni, o due linee separate, per distinguere quanti passaggi sono stati effettuati e ricevuto da ognuno dei calciatori della coppia analizzata. Lo spessore della linea, in questo caso, corrisponde logicamente al numero di passaggi completati in quella direzione.
- Infine, solitamente le circonferenze rappresentanti i calciatori sono raffigurate con uno specifico colore che può indicare un’ulteriore metrica (semplice o avanzata) legata al singolo calciatore, come ad esempio *possession value* o *expected assist* (xA).

2.3 Field Tilt

Il *field tilt* è un concetto proposto per misurare la percentuale di possesso palla di una squadra in termini di tocchi o passaggi effettuati unicamente nel terzo di campo offensivo. Costituisce una misura utile per valutare quanto una squadra ha avuto iniziativa offensiva durante una gara (o una serie di gare), piuttosto che considerare soltanto quanto tempo ha mantenuto il pallone in generale.

Nell’applicazione illustrata in questo elaborato, questo concetto è utilizzato per presentare una nuova tipologia di rappresentazione delle reti di passaggi, considerando solo i passaggi avvenuti nel terzo di campo offensivo. In questo modo, non solo si potrà capire “*quanto*” una squadra ha tenuto il possesso nella zona offensiva (attraverso la percentuale di *field tilt* pubblicata da tutti i provi-

der), ma anche “*in che modo*” è stata gestita in quale porzione di campo, quindi analizzando quali calciatori sono maggiormente coinvolti e in che direzione i passaggi sono maggiormente effettuati.

2.4 Normalizzazione

Uno degli aspetti fondamentali da considerare nell’ambito della *data analytics*, non solo in ambito calcistico, è il *contesto*. Un valore numerico, di per sé, non costituisce un’informazione rilevante se non contestualizzata, ovvero correlata ai fattori “ambientali” che lo hanno determinato.

Uno dei modi principali per contestualizzare un dato numerico e permettere la comparazione è quello di utilizzare la *normalizzazione* (o *standardizzazione*), ovvero fare in modo che due valori diversi vengano valutati su una scala comune. Le normalizzazioni più popolari nella *Football Analytics* sono quelle “*per minuti giocati*” (rappresentato con P , solitamente $P90$ ad indicare un dato valutato per ogni 90 minuti giocati dal calciatore) e “*per tocchi effettuati*” (rappresentato con T , solitamente $T50$ o $T100$ ad indicare un dato valutato per ogni 50 o 100 tocchi effettuati dal calciatore). Le statistiche normalizzate per minuti giocati sono spesso utilizzate per valutare performance e stati di forma, mentre la normalizzazione per tocchi effettuati risulta molto utile per comparare statistiche derivate da azioni con la palla (ad esempio un tiro o un passaggio).

Nello specifico, la formula per il calcolo di un valore normalizzato $T50$ è la seguente (similmente si calcola in tutti gli altri casi):

$$stat_T50 = \frac{stat_misurata}{numero_tocchi} \cdot 50 \quad (2.1)$$

Capitolo 3

Passing Style Analyser

Il seguente capitolo costituisce il cuore dell'intero elaborato, dove verrà descritta in dettaglio la progettazione e l'implementazione dell'applicazione **Passing Style Analyser**, illustrando il supporto tecnologico, la preparazione del dataset e tutti i singoli componenti sviluppati.

3.1 Supporto tecnologico

Per la realizzazione dell'applicazione, il linguaggio di programmazione utilizzato è stato *Python 3.9.13*, con l'ausilio delle seguenti librerie:

- *Pandas*: libreria open source per l'analisi e manipolazione dei dati;
- *NumPy*: libreria che aggiunge il supporto per vettori e matrici multidimensionali di grandi dimensioni, insieme a una vasta collezione di funzioni matematiche di alto livello per operare su vettori;
- *Matplotlib*: libreria completa per la creazione di visualizzazioni statiche, animate ed interattive;

- *mplsoccer*: libreria utilizzata per disegnare grafici (basati su *Matplotlib*) su un campo da calcio;
- *Streamlit*: framework open-source per la realizzazione, in linguaggio Python, di applicazioni web in modo semplice e veloce, particolarmente utilizzata in ambito data science.

3.2 Preparazione del dataset

Tra i diversi dataset messi a disposizione per la realizzazione del progetto finale, quello utilizzato per lo sviluppo dell'applicazione presentata nel presente elaborato è stato *soccerment_serieA_2021-22_allEvents*, contenente tutti gli eventi registrati (formazioni ufficiali, modulo di gioco, passaggi, tiri, cartellini, etc.) in ogni singola partita del campionato di Serie A 2021/22.

La preparazione del dataset è stata effettuata eseguendo le procedure elencate nelle sezioni seguenti.

3.2.1 Estrazione di informazioni dal campo *qualifier*

Dato che il dataset fornito (in formato *CSV*) è stato ricavato da un file originale in formato *XML*, alcune tipologie di eventi (solitamente non comune per analisi standard) non sono state esplicitamente estratte e inserite in colonne separate, ma rimaste a far parte di un campo denominato *qualifier*. La prima procedura eseguita è stata quella di definire dei metodi per l'estrazione ad-hoc degli attributi da tale campo. Gli attributi estratti dal campo *qualifier* sono:

- *team_formation*: modulo schierato ad inizio partita (codificato con un numero da 1 a 25);
- *team_player_pos*: vettore contenente le posizioni dei calciatori secondo il modulo di partenza (questo vettore è sempre formato dai numeri da 1 a 11, in ordine crescente, nelle prime 11 posizioni, e poi tanti 0 quanti il numero dei calciatori in panchina per una specifica squadra e partita);
- *jersey_number*: vettore contenente i numeri di maglia dei calciatori in ordine di posizione (dalla posizione 1 alla 11 in corrispondenza della relativa posizione nel modulo di partenza schierato e poi i sostituti);
- *player_list*: analogo all'attributo precedente, ma sostituendo i numeri di maglia agli ID dei singoli calciatori;
- *yellow_card*, *2nd_yellow_card*, *red_card*: valori booleani (True/False) indicanti la tipologia di cartellino preso dallo specifico calciatore.

3.2.2 *dec_time*

La colonna di valori rappresentante il tempo in forma decimale (partendo dalle colonne *min* e *sec*) è aggiunta per ottimizzare i calcoli di statistiche temporizzate, come ad esempio la somma dei minuti giocati da una squadra con uno specifico modulo iniziale.

3.2.3 *pass_recipient*

Il dataset originale fornito non contiene, per gli eventi di tipo passaggio, l'indicazione del calciatore ricevente, sia in caso di passaggio completato che non

completato. Al fine di creare il nuovo campo *pass_recipient*, sono stati effettuate due operazioni:

1. Il dataset è stato ordinato seguendo le indicazioni presenti nella documentazione *StatsPerform*, fornita a corredo del dataset stesso. In particolare, la lista di valori per i quali ordinare il dataset è la seguente: *matchday*, *team_id*, *period_id*, *min*, *sec*, *timestamp*.
2. Dopo aver ordinato il dataset per eventi consecutivi temporalmente, possiamo assumere che in caso di passaggio completato, l'evento successivo al passaggio, sia esso un tiro, un passaggio o un semplice tocco di palla, sarà eseguito dal calciatore che ha ricevuto il passaggio. Le colonne *pass_recipient* e *pass_recipient_id* sono quindi create traslando di una riga in alto, rispettivamente, i valori presenti nelle colonne *player* e *player_id*, rappresentati il calciatore che ha effettuato l'evento successivo al passaggio. In caso di eventi non corrispondenti a passaggi, o in caso di valori uguali per *player* e *pass_recipient*, indice di un passaggio non completato con successo, tali celle sono svuotate.

3.2.4 *team_ids.json*

Invece di interrogare il dataset di volta in volta per associare il nome di una squadra al proprio ID, è stato generato e memorizzato un file di tipo *JSON* denominato *team_ids.json* contenente tale associazione.

3.2.5 *player_start_pos* e *pass_recip_start_pos*

Il lavoro presentato mira anche all’analisi dello stile di gioco delle squadre della Serie A 2021/22 considerato più partite consecutive. Dato che la formazione iniziale può non essere sempre la stessa (in termini di calciatori schierati), le colonne *player_start_pos* e *pass_recip_start_pos* sono create per associare ogni calciatore titolare alla relativa posizione nello schema tattico (o modulo) di partenza, secondo le indicazioni sulle posizioni per ognuno dei moduli possibili fornite all’interno della documentazione aggiuntiva di *StatsPerform* denominata “*Appendix 5 - Formations Explained*”. Per una migliore e più semplice lettura, le posizioni espresse nella documentazione menzionata sono riportate in Appendice A, in un’immagine creata apposito a tale scopo.

3.2.6 *match_status*

Per la visualizzazione dello sviluppo di gioco nelle tre diverse condizioni della partita, la colonna *match_status* è aggiunta al dataset e valorizzata per ogni evento valutando opportunamente i valori contenuti nelle colonne *team_id*, *home_team_id*, *away_team_id*, *home_score* e *away_score*.

3.2.7 *teams_lineups.json* e *teams_match_status.json*

Come nel caso descritto in precedenza, per evitare di interrogare costantemente il dataset per recuperare delle informazioni che richiederebbero un grande tempo di computazione, sono create due ulteriori risorse di supporto:

- *teams_lineups.json*, contenente i minuti giocati da ogni squadra per ogni giornata con la formazione iniziale, ovvero fino al momento della prima

sostituzione o espulsione. Si tratta sostanzialmente del modo più comune di visualizzazione delle reti di passaggi.

- *teams_match_status.json*, contenente i valori di tempo speso da ogni squadra durante tutte le partite per ognuna delle tre condizioni di risultato.

3.2.8 Filtraggio del dataset

Per ridurre la dimensione del dataset da utilizzare nell'applicazione web, al termine delle procedure di processamento e preparazione, sono state rimosse tutte le colonne non utili ai fini dell'esecuzione.

3.3 Componenti dell'interfaccia grafica

In questa sezione verranno descritti e mostrati tutti i componenti dell'interfaccia grafica dell'applicazione **Passing Style Analyser**.

Per ragioni di complessità e di lunghezza dell'elaborato, non verranno spiegate in dettaglio le procedure implementate in linguaggio Python per la realizzazione dell'applicazione. Tutto il codice dell'applicazione è, in ogni caso, pubblicamente disponibile e consultabile al seguente link:

<https://github.com/erasmopurif/passing-style-analysis>.

3.3.1 Home Page

La pagina iniziale dell'applicazione (Fig. 3.1), oltre ai loghi delle aziende promotori del II° Corso di Football Data Intelligence e il nome dell'autore, fornisce una descrizione delle due componenti di analisi principali. Sulla parte sini-

stra, nella cosiddetta *sidebar*, sono visibili i link di accesso alle pagine delle due funzionalità di analisi disponibili.

Full Pitch & Field Tilt Networks

After selected a single matchday or a range of matchdays, it is possible to analyze:

- a single team,
- two teams, or
- all the 20 Serie A teams,

by viewing their passing network in the most used formation for the selected matchdays, either visualizing the full-pitch or field-tilt passing network.

Additionally, the user can select the number of minimum passes to be displayed in the passing networks (Note that for multiple matchdays, this value is multiplied by the number of selected matchdays).

Match Status Networks

After selected a single matchday or a range of matchdays, and a team, it is possible to analyze the full-pitch passing network for each match status (i.e. winning, drawing, losing) and view their barycenter in the "passing" scenario.

Additionally, the user can decide whether or not to keep the goalkeeper in the barycenter computation, and select the number of minimum passes to be displayed in the passing networks (Note that in this case, for multiple matchdays, this value is NOT multiplied by the number of selected matchdays).

Figura 3.1: Passing Style Analyser - Home Page

3.3.2 Full Pitch & Field Tilt Networks

La prima componente sviluppata, denominata *Full Pitch & Field Tilt Networks*, permette ad un utente di visualizzare la rete di passaggi sia a campo intero (*full pitch network*) che nel terzo di campo offensivo (*field tilt network*), con la possibilità di personalizzare la visualizzazione, come mostrato in Fig. 3.2, attraverso la scelta dei valori per ognuno degli attributi disponibili:

- singola giornata o intervallo di giornate (consecutive);
- singola squadra, due squadre o tutte le squadre della Serie A 2021/22;
- sete di passaggi a campo intero o terzo offensivo;

- numero di passaggi minimi tra due calciatori da considerare, tenendo in conto che per un intervallo di giornate, questo valore è moltiplicato per il numero di giornate selezionate.

Full Pitch & Field Tilt Networks

Select matchday(s)

Matchday visualization

Single matchday
 Range of matchdays

Matchday

1

Select team(s)

Team visualization

Single team
 Two teams
 All teams

Teams

Juventus Internazionale

Select the area of the pitch for visualizing the passing network

Area of the pitch to visualize

Complete pitch (Full passing network)
 Final third (Field-tilt passing network)

Select the number of minimum passes

Number of minimum passes

3

Run analysis

Figura 3.2: Passing Style Analyser - Full Pitch & Field Tilt Networks

3.3.3 Match Status Networks

La seconda componente implementata, denominata *Match Status Networks*, ha lo scopo di mostrare la rete di passaggi per una singola squadra nelle tre condizioni di partita (vittoria, pareggio, sconfitta), mostrando anche il baricentro della squadra per ognuna delle condizioni. Interagendo con questa area dell'applicazione, si possono personalizzare i seguenti parametri (Fig. 3.3):

- singola giornata o intervallo di giornate (consecutive);
 - squadra da analizzare;
 - numero di passaggi minimi tra due calciatori da considerare, uguale per tutte e tre le condizioni di partita, ma tenendo in conto questa volta che per un intervallo di giornate, questo valore non è moltiplicato per il numero di giornate selezionate;
 - considerare o meno la posizione del portiere nel calcolo e visualizzazione del baricentro della squadra.

Match Status Networks

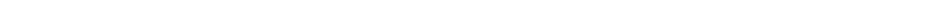
Select matchday(s)

Matchday visualization

Single matchday

Range of match

Range of matchdays



1 38
1 38

Select team

Team

Atlanta

Select the number of minimum passes to visualize

Number of minimum passes

?

5

1

Select whether or not to keep the goalkeeper in barycenter computation

Exclude GK

Run analysis

Figura 3.3: Passing Style Analyser - Match Status Networks

3.4 Visualizzazione dei risultati

Di seguito verranno presentate brevemente (quasi) tutte le tipologie di visualizzazione che un utente può ricevere interagendo con l'applicazione presentata, dividendo la sezione per componente e descrivendo, per ogni configurazione, il modo in cui quella particolare tipologia di visualizzazione è stata realizzata e le differenze con le precedenti. In casi specifici, verranno mostrate le problematiche riscontrate nel dataset utilizzato.

3.4.1 Full Pitch & Field Tilt Networks

Per quanto riguarda la componente di visualizzazione delle linee di passaggio a campo intero o terzo offensivo, per tutte le tipologie, la grandezza dei punti rappresentanti i singoli calciatori è determinata dal loro valore di *pass_count_T50*, ovvero il numero di passaggi effettuati per ogni 50 tocchi di palla; mentre il colore dei singoli calciatori è determinato dal valore della metrica *xA_T50*, ovvero di *expected assist* normalizzato per ogni 50 tocchi di palla.

Le linee di passaggio, invece, evidenziano entrambe le direzioni.

Single matchday - Single team

Le legende di *xA_T50* e *pass_count_T50* sono mostrate lateralmente e, per distinguere i calciatori, viene utilizzato il nome, non il numero. In alcune circostanze, il nome può costituire una limitazione, in quanto facilmente sovrapponibile con un altro, in caso di calciatori con posizioni medie in campo molto simili.

In Fig. 3.4 e 3.5 sono mostrati degli esempi di questa configurazione con 3 passaggi minimi, rispettivamente, nel caso *full pitch* e *field tilt*.

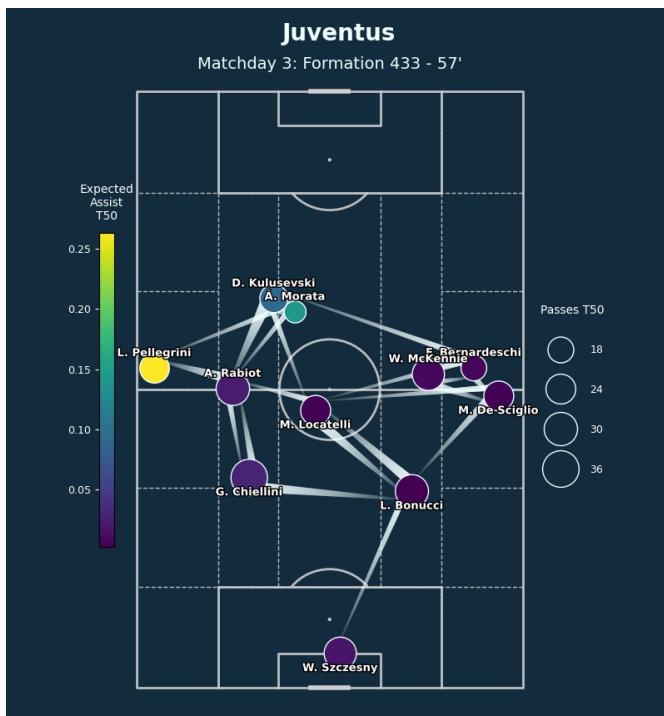


Figura 3.4: Full Pitch Network - Single matchday/Single team

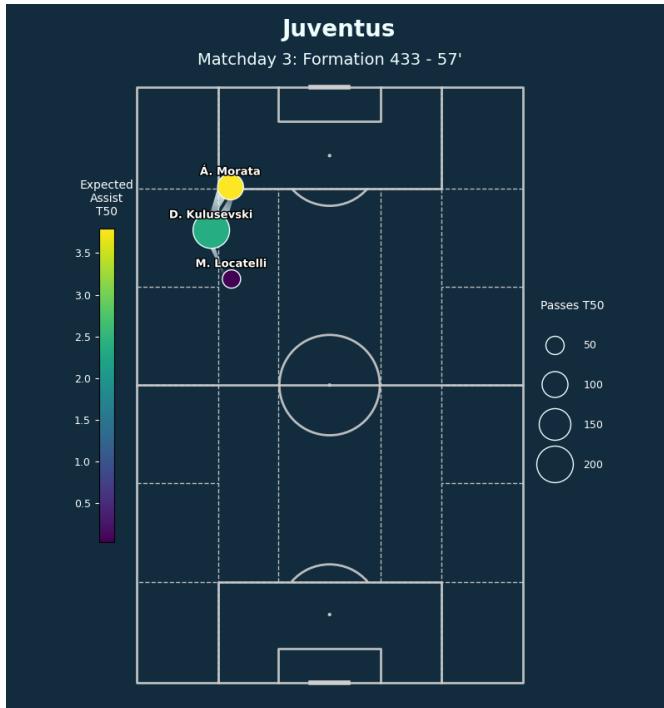


Figura 3.5: Field Tilt Network - Single matchday/Single team

Multiple matchdays - Single team

Le differenze con la precedente visualizzazione sono che viene visualizzata la rete di passaggi relativa al modulo più utilizzato nell'intervallo di partite selezionato e i calciatori sono etichettati con il numero relativo alla posizione in campo di quello specifico modulo, secondo lo schema in Appendice A. Il grafico a barre riportante il minutaggio per ognuno dei moduli di partenza utilizzati, viene mostrato al di sotto del campo (in verde la barra relativa alla formazione più utilizzata). In Fig. 3.6 è riportato un esempio di visualizzazione a campo intero con 2 passaggi minimi (analogo per il terzo offensivo).

Single matchday - Two teams

La differenza rispetto all'analogia visualizzazione a singola squadra è che le legende sono poste in alto in modo da poter affiancare le due squadre per un miglior confronto.

L'esempio *full pitch* mostrato con 3 passaggi minimi in Fig. 3.7 non è casuale, ma riguarda la prima giornata del campionato analizzato. Nel caso specifico della gara Napoli-Venezia, all'interno del dataset sono stati registrati dei valori anomali per quanto riguarda gli xA (soprattutto del Venezia, come visibile). La maggior dei passaggi completati dal Venezia in quella partita è stata registrato con valore di xA pari a 0.99, decisamente irrealistico.



Figura 3.6: Full Pitch Network - Multiple matchdays/Single team

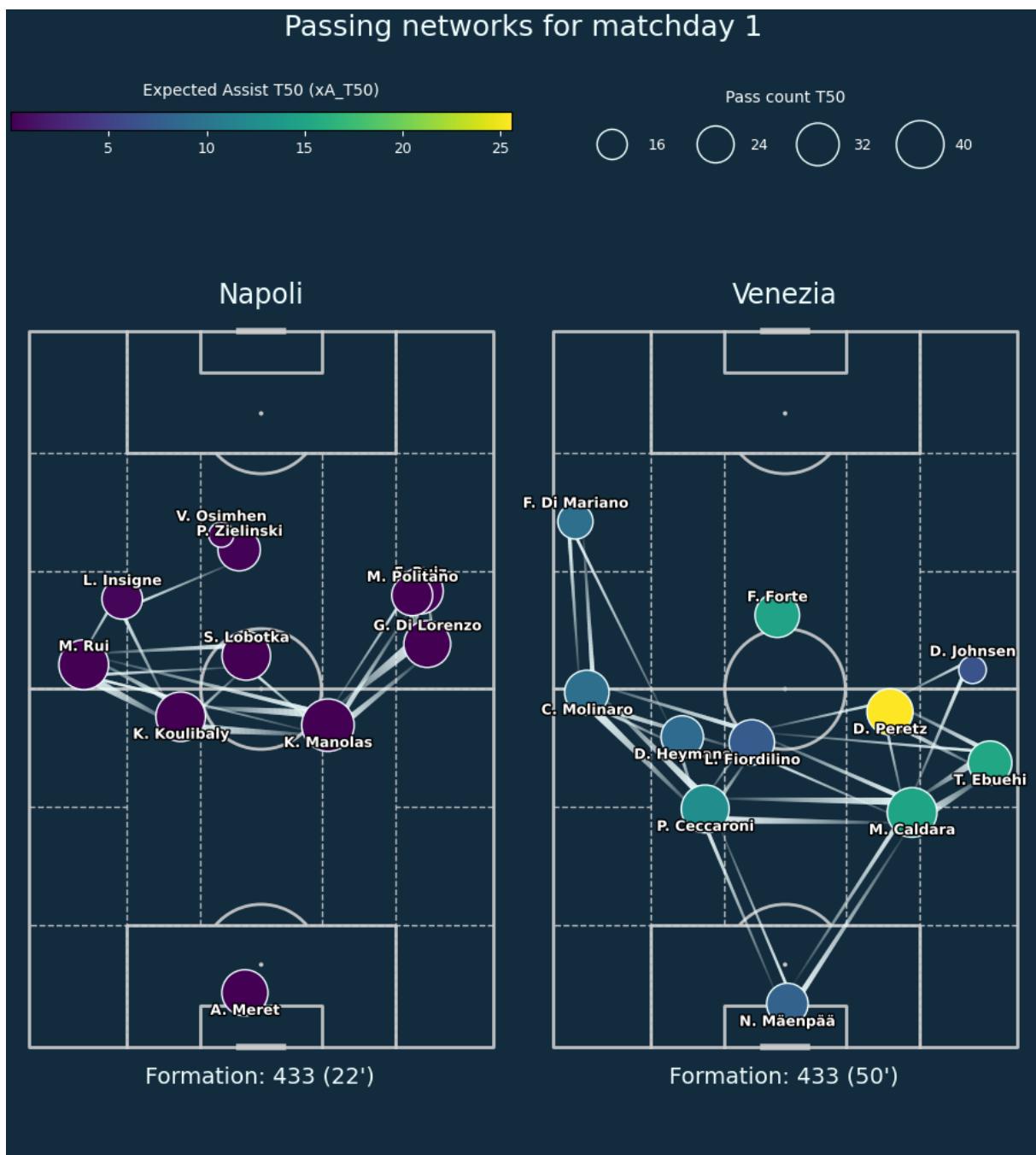


Figura 3.7: Full Pitch Network - Single matchday/Two teams

Multiple matchdays - Two teams

Come nel caso singola squadra, la visualizzazione è corredata dal grafico a barre riportante i moduli utilizzati nelle giornate selezionate, mostrando poi

le linee di passaggio per il modulo con il maggior minutaggio. In Fig. 3.8 è riportato un esempio a campo intero con 3 passaggi minimi su 37 giornate.

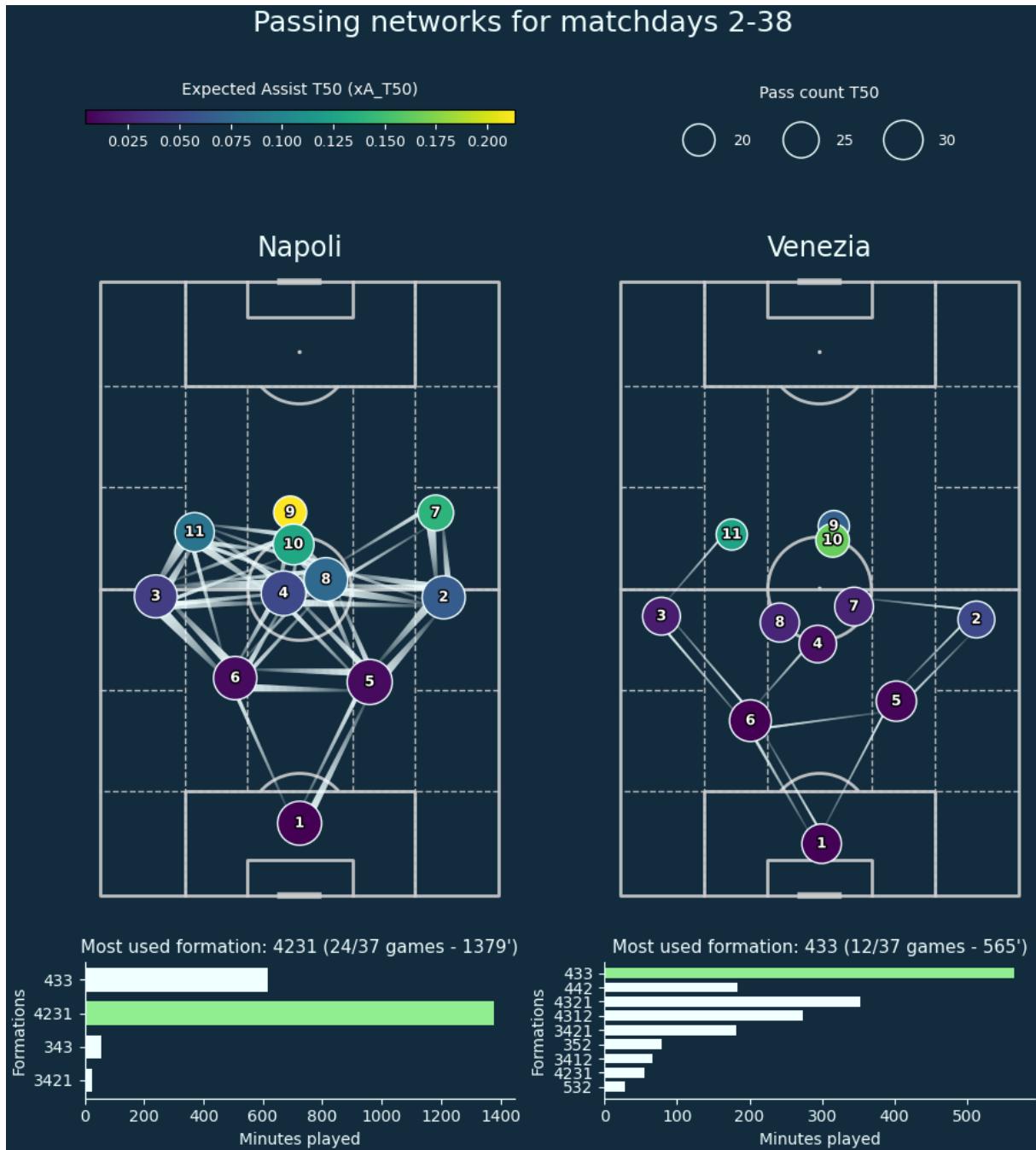


Figura 3.8: Full Pitch Network - Multiple matchdays/Two teams

Multiple matchdays - All teams

In questa configurazione, i calciatori non vengono etichettati per questioni di spazio e per evitare troppa confusione nella lettura del grafico e, anche in caso di intervallo di gare, per gli stessi motivi, la formazione più utilizzata è fornita in forma testuale, senza aggiungere un ulteriore grafico. Essendo pressoché visualizzazioni analoghe, come esempio, in Fig. 3.9 è mostrato il caso su più gare a campo intero con 3 come numero minimo di passaggi.

3.4.2 Match Status Network

Per la visualizzazione delle linee di passaggio nelle tre condizioni di risultato, vengono apportate alcune differenze grafiche:

- il campo è mostrato con sfondo “naturale”, in erba;
- non sono presenti le legende di xA_T50 e $pass_count_T50$;
- i calciatori, hanno un unico colore rappresentativo;
- nelle situazioni multi-partita, vengono considerate tutte le formazioni di partenza, non solo quella con il minutaggio maggiore (quindi, in questo caso, i numeri delle posizioni dei calciatori non hanno la stessa significatività della precedente componente, e potrebbero essere evitati, ma vengono aggiunti comunque per aiutare l’analisi);
- sono mostrati due grafici a barre, riportanti il minutaggio speso dalla squadra per ognuna delle tre condizioni di risultato e per ognuna delle formazioni iniziali utilizzate per la/le giornata/e considerata/e.



Figura 3.9: Full Pitch Network - Multiple matchdays/All teams

Multiple matchdays - Exclude GK

Nell'esempio mostrato in Fig. 3.10-3.11 in cui sono stati considerati 8 passaggi minimi ed escluso il portiere dal calcolo del baricentro, si può notare come il Napoli, nelle giornate selezionate (10-30), abbia avuto un baricentro più basso in situazioni di vittoria e molto più alto quando in svantaggio. Ma riprendendo in considerazione il discorso del contesto, il dato sulle situazioni di svantaggio potrebbe non essere significativo, a causa del basso minutaggio in quella situazione di punteggio (riscontrabile anche dalla mole di passaggi presente).

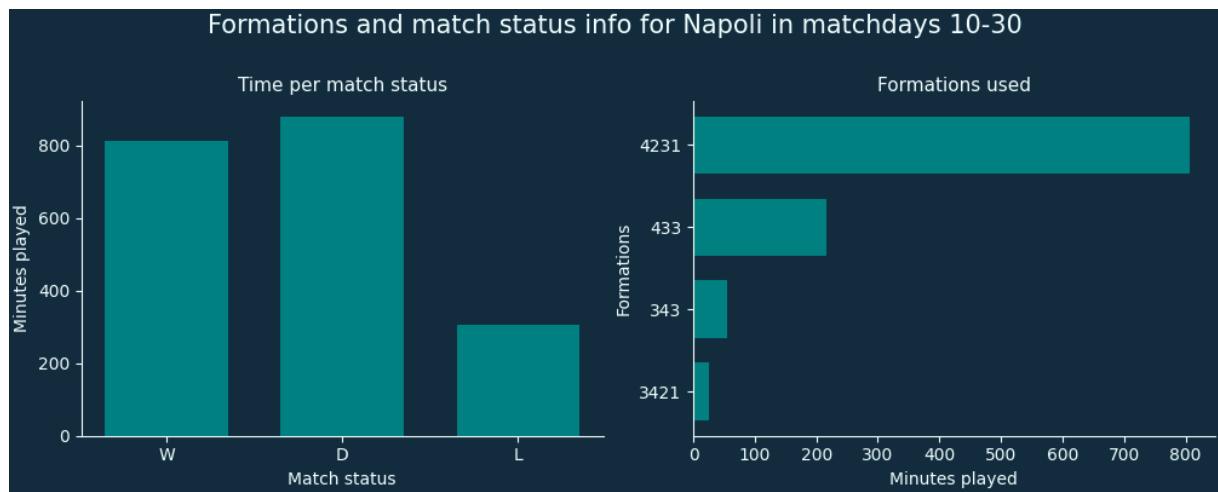


Figura 3.10: Match Status Network - Multiple matchdays - Team info

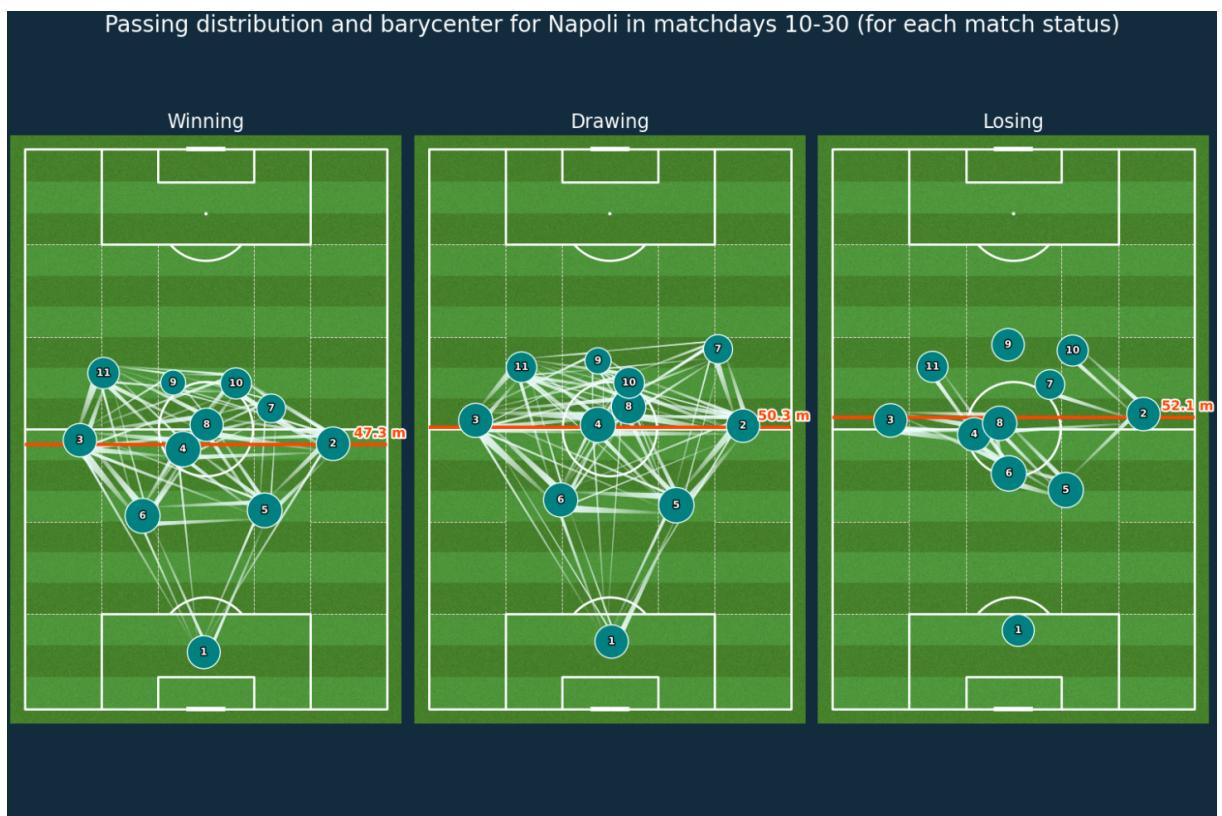


Figura 3.11: Match Status Network - Multiple matchdays - Pitch

Capitolo 4

Conclusione e Sviluppi Futuri

Nel presente elaborato, è stata presentata **Passing Style Analyser**, un'applicazione web interattiva per lo studio dello stile di gioco delle squadre della Serie A 2021/22 attraverso l'analisi dettagliata delle reti di passaggi. Attraverso la descrizione dettagliata di tutte le fasi del progetto, dalla preparazione del dataset fino alla visualizzazione dei risultati, si è fornita una panoramica completa sulle funzionalità che ogni utente può utilizzare.

Per quanto riguarda sviluppi futuri, l'intenzione è di rendere l'applicazione sempre più *user-friendly* e personalizzabile, oltre che migliorarne l'efficienza a livello computazionale.

Appendice A

Formations Explained

Vengono di seguito riportate, con una visualizzazione più comprensibile e compatta rispetto alla documentazione fornita, le formazioni disponibili con l'indicazione delle relative posizioni in campo, necessarie a comprendere il posizionamento dei calciatori nei casi di analisi multi-partita.

Le posizioni in campo, per le 20 formazioni selezionate sono riportate in Fig. A.1 della pagina successiva.



Figura A.1: Formations explained