# ANNO ACCADEMICO 2016-2017

# PROGETTAZIONE E IMPLEMENTAZIONE DI UN SISTEMA DI RACCOMANDAZIONE PER IL GIOCO DEL FANTACALCIO: VIRTUAL COACH

PROGETTO PER IL CORSO SISTEMI INTELLIGENTI PER INTERNET

BIAGIO CALDARARO – IGOR MASTROLORENZO

426748 - 433041 Università Roma Tre

# **INDICE**

1	INTRODUZIONE	2
	1.1 COSA È IL FANTACALCIO	2
	1.2 OBIETTIVI DEL PROGETTO	2
2	TECNOLOGIE UTILIZZATE	3
3	RECUPERO DEI DATI	4
	3.1 PROBABILI FORMAZIONI	4
	3.2 QUOTAZIONE GIOCATORI	5
	3.3 STATISTICHE GIOCATORI	6
	3.4 CLASSIFICA	6
	3.5 RISULTATI PRECEDENTI	7
	3.6 CALENDARIO	8
4	ARCHIVIAZIONE DEI DATI	9
	4.1 FRESCHEZZA	9
	ALGORITMO DI RANKING	
	5.1 FORMULA	
	5.2 GESTIONE DEI MODULI	
6	FUNZIONALITÀ	12
	6.1 INSERIMENTO/ELIMINAZIONE GIOCATORE	12
	6.2 GESTIONE DELLA ROSA	13
	6.3 VISUALIZZAZIONE STATISTICHE GIOCATORE	13
	6.4 SELEZIONE DEL MODULO	13
	6.5 SCHIERAMENTO DELLA FORMAZIONE	14
	6.6 TREQUARTISTI	15
7	TEST E RISULTATI	16
	CONCLUSIONI E SVILUPPI FUTURI	
9	RIFERIMENTI	18

#### 1 INTRODUZIONE

VirtualCoach è un'applicazione web in grado di supportare i partecipanti al gioco del fantacalcio nella scelta degli 11 giocatori che, settimana per settimana, andranno a comporre la formazione che "scenderà in campo".

### 1.1 COSA È IL FANTACALCIO

Il fantacalcio è un gioco fantasy sport sul calcio, consistente nell'organizzare e gestire squadre virtuali formate da calciatori reali, scelti fra quelli che giocano il torneo cui il gioco si riferisce (Serie A, UEFA Champions League, Mondiale, Europeo).

Inizia con un'asta in cui i partecipanti creano la loro squadra formata, di norma, da 25 giocatori: 3 portieri, 8 difensori, 8 centrocampisti e 6 attaccanti.

Durante ogni giornata di campionato ogni membro deve scegliere gli 11 giocatori da schierare dalla sua rosa secondo dei moduli prestabiliti (343, 442, 541...). La formazione cosi schierata si sfiderà con la formazione dell'avversario.

Per decretare il vincitore si sommano i voti che la redazione sportiva di riferimento ha assegnato ai giocatori sommando eventuali bonus (gol, assist, rigore parato) o malus (ammonizione, espulsione, autogol...).



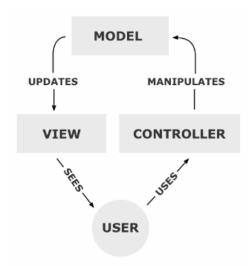
#### 1.2 OBIETTIVI DEL PROGETTO

L'obiettivo di VirtualCoach è stato quello di creare un sistema di raccomandazione che, partendo dalla rosa dei 25 giocatori, sceglie i migliori 11 secondo il miglior modulo possibile massimizzando così le possibilità di vittoria.

#### 2 TECNOLOGIE UTILIZZATE

Lo sviluppo di VirtualCoach ha richiesto l'utilizzo di svariate tecnologie software: in particolare il back-end dell'applicazione è scritta in JAVA e ciò ha permesso l'utilizzo di strumenti di sviluppo quali l'IDE Eclipse nella sua versione EE e Apache Tomcat come application server.

VirtualCoach segue le specifiche del pattern architetturale Model-View-Controller (MVC) in grado di separare la logica di presentazione dei dati dalla logica di business.



Per quanto riguarda il tema della persistenza abbiamo scelto di usare come DBMS MongoDB, il quale, essendo Document Oriented ci ha facilitato non poco le operazioni di archiviazione (vedi capitolo 4), mentre i dati di interesse sono stati reperiti sul web combinando codice scritto in Python e la shell di Unix (l'intera applicazione è stata sviluppata sul sistema operativo Ubuntu).

La parte front-end infine è stata realizzata con l'aiuto del framework Bootstrap.



#### 3 RECUPERO DEI DATI

La parte più importante per un sistema di raccomandazione sta nel recupero delle informazioni che poi verranno elaborate e utilizzate.

Per quanto riguarda il nostro scopo abbiamo utilizzato dati relativi al campionato di Serie A: singoli giocatori, squadre in cui militano e andamento generale della competizione.

Le principali sorgenti informative sono state le probabili formazioni, dalle quali è possibile dedurre se un giocatore sarà titolare o meno, le quotazioni che vengono assegnate a ciascun giocatore, le relative statistiche, la classifica della Serie A e il calendario delle partite.

Per facilitare il recupero delle informazioni dalle pagine web abbiamo utilizzato la libreria Python BeautifulSoup che permette di analizzare la struttura di una pagina html mettendo a disposizione degli elementi utili per estrarre le informazioni di interesse.

#### 3.1 PROBABILI FORMAZIONI

Per quanto riguarda le probabili formazioni abbiamo estratto le informazioni che ci servivano dal sito di Fantagazzetta.

Di seguito viene mostrata la struttura della pagina:

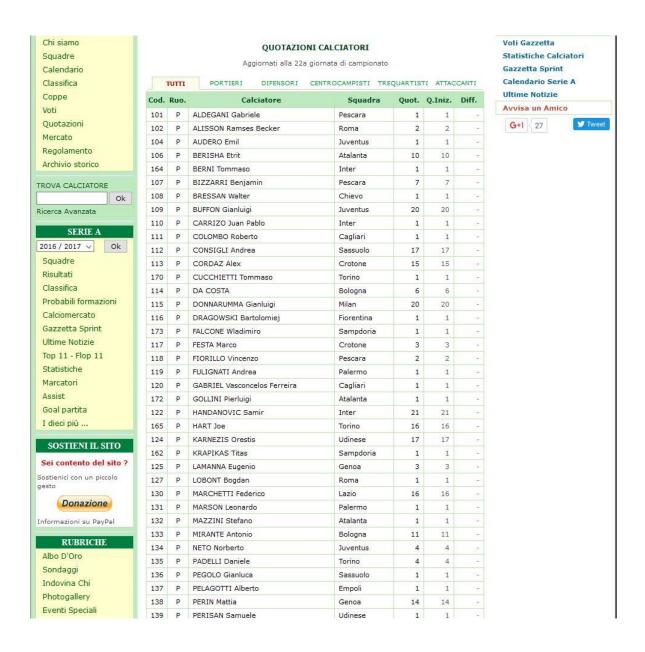


Le informazioni di interesse sono il nome del giocatore, la percentuale che rappresenta la possibilità di giocare o di entrare dalla panchina e se risulta titolare o meno.

#### 3.2 QUOTAZIONE GIOCATORI

Le quotazioni dei giocatori sono state estratte dalla pagina Fantagiaveno.

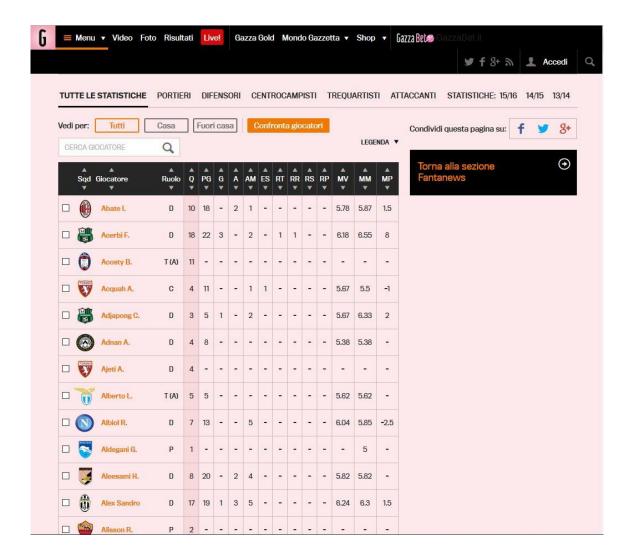
Qui sono presenti le quotazioni che le redazioni sportive hanno assegnato ai giocatori a inizio campionato (quotazione iniziale) e quella aggiornata all'ultima giornata di campionato (quotazione attuale); ecco la struttura della pagina:



In questo caso, trattandosi di una tabella, è possibile andare a recuperare le informazioni di interesse andando ad esaminare le varie righe.

#### 3.3 STATISTICHE GIOCATORI

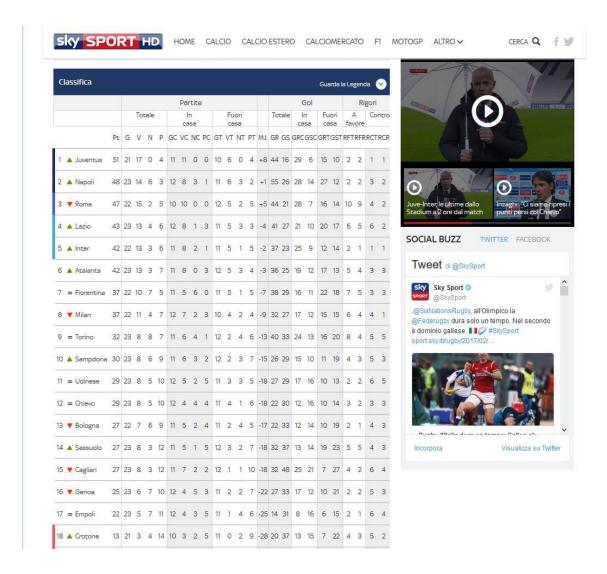
Per le statistiche dei giocatori abbiamo usato come sorgente informativa la pagina web del La Gazzetta dello sport dalla quale abbiamo recuperato importanti informazioni relative alle partite giocate quali gol, assist, media voto e fantamedia (media dei voti con annessi bonus/malus).



Anche in questo caso, trattandosi di una tabella, è stato possibile recuperare le informazioni analizzando le righe del codice html.

#### 3.4 CLASSIFICA

Per la classifica abbiamo recuperato le informazioni dalla pagina di Sky Sport; abbiamo utilizzato questa classifica perché, come si può vedere sotto, contiene molte informazioni utili oltre ai punti; ad esempio i gol fatti e subiti in casa e fuori casa.

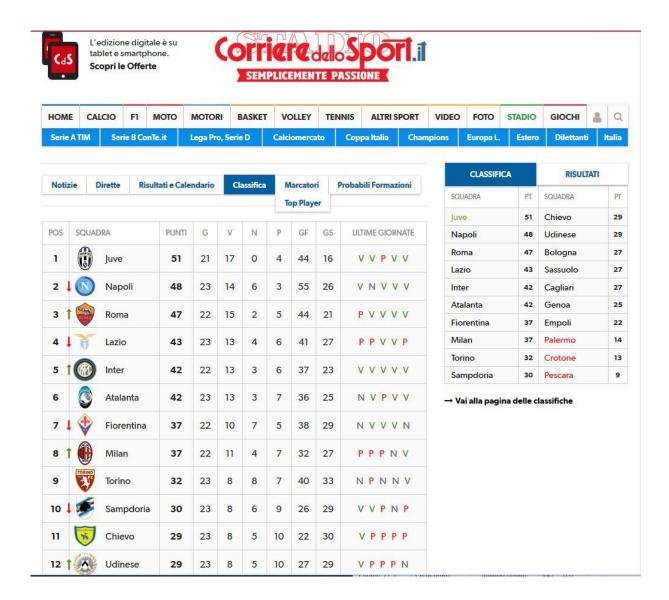


Anche in questo caso trattandosi di una tabella abbiamo operato come fatto in precedenza.

#### 3.5 RISULTATI PRECEDENTI

Per ottimizzare il sistema di raccomandazione abbiamo preso in considerazione anche l'andamento della squadra in cui milita il calciatore. A questo scopo abbiamo preso in considerazione i risultati delle cinque precedenti partire.

Questa informazione l'abbiamo recuperata dalla seguente pagina de Il Corriere dello Sport:



#### 3.6 CALENDARIO

Per quanto riguarda il calendario, trattandosi di una informazione "statica", non è necessario recuperarla per ogni giornata ma basta inserirla solo una volta nel database.

#### 4 ARCHIVIAZIONE DEI DATI

Per quanto riguarda l'archiviazione dei dati abbiamo optato per l'utilizzo di MongoDB. MongoDB è un DBMS non relazionale, orientato ai documenti. Classificato come un database di tipo NoSQL, MongoDB si allontana dalla struttura tradizionale basata su tabelle dei database relazionali in favore di documenti in stile JSON con schema dinamico (MongoDB chiama il formato BSON), rendendo l'integrazione di dati di alcuni tipi di applicazioni più facile e veloce.

Abbiamo deciso di utilizzare MongoDB in primo luogo in quanto un DBMS NoSQL orientato ai documenti rappresenta una delle migliori opzioni per l'archiviazione di informazioni reperite dal web, come in questo caso; in secondo luogo in quanto risulta un'alternativa di facile utilizzo.

Come è possibile notare nel codice delle applicazioni utilizzate per il recupero delle informazioni, è possibile gestire meglio le informazioni e strutturarle in modo da poter essere inserite immediatamente e più facilmente nel DB.

A questo scopo, ogni applicazione per il recupero delle informazioni genera un file formato JSON contenente tutti i dati di interesse; ogni riga rappresenta un documento in formato BSON pronto per essere aggiunto al DB.

Altra caratteristica che ci ha fatto optare per questa tecnologia è la grande scalabilità che essa offre. Permettendoci così di poter in futuro ampliare la nostra applicazione senza dover andare a ristrutturare il tutto.

#### 4.1 FRESCHEZZA

Uno degli elementi di maggiore importanza per un sistema si raccomandazione è quello di poter contare su dati sempre aggiornati in quanto, altrimenti, il sistema, facendo riferimento a dati vecchi, perde di precisione e di utilità.

Per garantire ciò, sul server che gestisce l'applicazione web, viene lanciato periodicamente uno script che va ad aggiornare tutte le informazioni presenti nel database.

#### 5 ALGORITMO DI RANKING

La parte centrale del nostro lavoro è stata senza dubbio lo sviluppo dell'algoritmo di ranking.

Tale algoritmo si occupa di generare i valori numerici da associare ai giocatori; più sono alti tali valori, più è consigliabile inserire i giocatori associati nella "formazione titolare".

#### 5.1 FORMULA

L'algoritmo di Ranking in questione si basa su una formula che tiene conto di tutte le informazioni trattate nel capitolo 3, ciascuna pesata in modo da avere la propria rilevanza nel risultato finale.

Dopo vari studi e test appropriati la formula su cui attualmente si basa VirtualCoach può essere descritta in questo modo:

$$RANK_p = \begin{cases} -1 & \text{se il giocatore } p \text{ \`e indisponibile}^1 \\ (P+r)RANK_{parz} & \text{altrimenti} \end{cases}$$

$$RANK_{parz} = Q_C + \left(\frac{Q_C - Q_I}{2}\right) + \left(\frac{S - S_0}{2}\right) + h + F + (R - R_0)$$

#### Dove:

•  $Q_C$  è la quotazione attuale del giocatore

- $Q_I$  è la quotazione iniziale del giocatore
- S sono i punti totalizzati dalla squadra in cui milita il giocatore
- ullet S<sub>0</sub> sono i punti totalizzati dalla squadra prossima avversaria della squadra in cui milita il giocatore
- h dipende se la squadra in cui milita il giocatore gioca in casa o meno
- ullet F è la fantamedia relativa al giocatore
- R dipende dagli ultimi 5 risultati della squadra in cui milita il giocatore
- ullet  $R_0$  dipende dagli ultimi 5 risultati della squadra prossima avversaria della squadra in cui milita il giocatore
- P è la probabilità del giocatore di giocare dall'inizio della partita o di subentrare a partita in corso
- r dipende se il giocatore gioca dall'inizio o subentra (dipende da P)

<sup>1</sup> I giocatori che sono infortunati, squalificati o in generale non convocati sono considerati indisponibili

#### 5.2 GESTIONE DEI MODULI

Una volta ordinati i giocatori della rosa in base ai propri valori di ranking, per schierare la formazione abbiamo dovuto considerare i vari moduli.

Il modulo è la disposizione in campo dei giocatori di una squadra, in base al numero di difensori, centrocampisti e attaccanti.

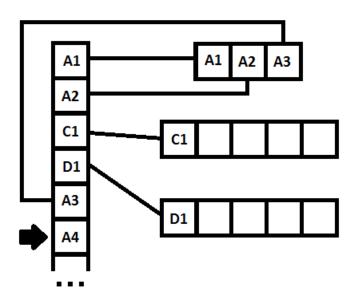
L'allenatore schiera in campo una formazione composta di 11 giocatori, tra cui un portiere non incluso nella numerazione del modulo.

Tale numerazione consiste nel conteggio dei giocatori schierati in ogni reparto (difesa, centrocampo e attacco).

Per esempio, nell'accezione odierna un 3-5-2 è interpretato come uno schieramento di tre difensori, cinque centrocampisti e due attaccanti.

Nel gioco del fantacalcio i moduli disponibili comprendono un minimo di 3 e un massimo di 5 elementi per difesa e centrocampo e un minimo di un solo elemento e un massimo di 3 per l'attacco.

VirtualCoach, scandendo la rosa dei giocatori, seleziona automaticamente il miglior portiere mentre per quanto riguarda gli altri reparti essi vengono "riempiti" fino alla loro massima capacità o al raggiungimento dei 10 elementi complessivi; come si può vedere in figura l'attaccante A4 non può essere inserito in formazione poiché non esiste un modulo che prevede più di 3 attaccanti.



Gli altri elementi vanno a comporre la panchina, utile qualora la predizione di qualche titolare si rivelasse errata o, per qualche ragione, la redazione calcistica di riferimento non dovesse attribuirgli il voto.

# 6 FUNZIONALITÀ

Terminata la fase di sviluppo VirtualCoach si presenta in questo modo:



# 6.1 INSERIMENTO/ELIMINAZIONE GIOCATORE

La prima operazione che l'utente deve fare per poter interagire con VirtualCoach è l'inserimento dei giocatori che andranno a comporre la rosa; questa operazione può essere effettuata scrivendo il nome del giocatore nella barra in alto a sinistra e schiacciando il pulsante Inserisci.

È disponibile un menu a tendina che guida l'utente nell'inserimento.



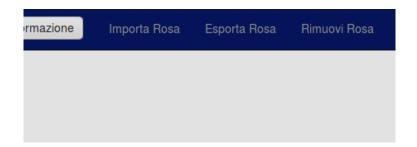
Una volta inserito, il giocatore comparirà nella rosa (nella colonna a sinistra): vengono visualizzate altre info di base quali il ruolo e squadra.

È possibile eliminare il giocatore dalla rosa schiacciando sul pulsante con la croce rossa presente affianco.



#### **6.2 GESTIONE DELLA ROSA**

Per rendere più agevoli le operazioni sulla rosa sono presenti le funzionalità di importazione, esportazione e rimozione rosa: una volta inserita la rosa giocatore per giocatore conviene esportarla (la rosa viene esportata in un file di testo) per poi importarla di nuovo alla prossima esecuzione.



#### 6.3 VISUALIZZAZIONE STATISTICHE GIOCATORE

Cliccando sul nome del giocatore è possibile visualizzarne le statistiche nella parte destra dello schermo.

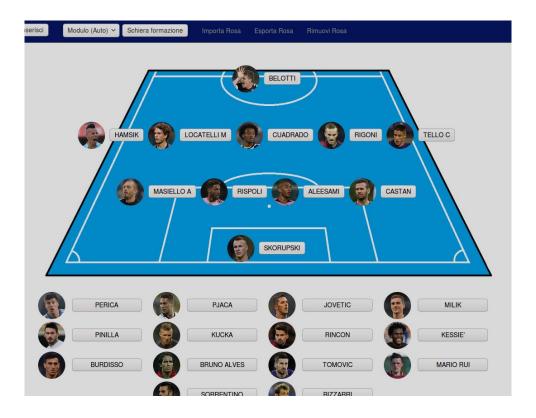
#### 6.4 SELEZIONE DEL MODULO

VirtualCoach seleziona automaticamente il modulo, tuttavia l'utente può comunque scegliere il modulo che preferisce usare.



#### 6.5 SCHIERAMENTO DELLA FORMAZIONE

Dopo aver inserito la rosa e selezionato il modulo (facoltativo) l'utente può schierare la formazione schiacciando sull'apposito tasto.



I giocatori "non presenti in campo" formano la panchina; sono comunque ordinati in base a ruolo e ranking.

# **6.6 TREQUARTISTI**

Alcuni giocatori, a seconda della redazione sportiva, possono essere inseriti nella lista dei centrocampisti o nella lista degli attaccanti.

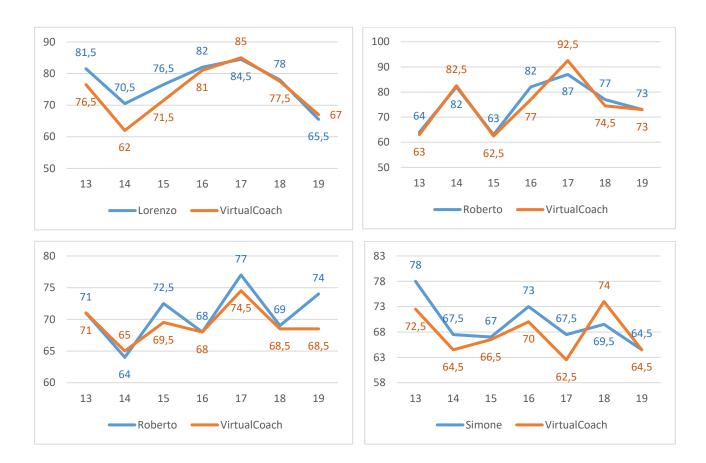
È possibile cambiare il ruolo a questi giocatori, riconoscibili poiché presentano l'etichetta del ruolo di colore blu, schiacciando sul loro ruolo nella colonna della rosa.



#### 7 TEST E RISULTATI

VirtualCoach è stato testato su un intervallo del campionato di Serie A (precisamente dalla 13° alla 19° giornata) e su quattro rose diverse.

Sono state comparate le formazioni generate da VirtualCoach con le formazioni schierate dai rispettivi giocatori di fantacalcio e i risultati sono stati i seguenti:



Come si può evincere da i grafici i risultati risultano assolutamente comparabili, raramente VirtualCoach fa meglio del giocatore ma ciò dipende da fattori puramente imprevedibili statisticamente (giocatori che fino a quel momento avevano fatto cattive prestazioni ne hanno fatte ottime in determinate giornate, oppure giocatori di squadre negli ultimi posti della classifica fanno ottime prestazione contro squadre molto più in alto) e dal fatto che l'algoritmo non è perfettamente ottimizzato (servirebbero test su interi campionati per migliorarlo).

#### **8 CONCLUSIONI E SVILUPPI FUTURI**

In questo documento è stata descritta la progettazione e l'implementazione di VirtualCoach: un sistema di raccomandazione per il gioco del fantacalcio.

Abbiamo visto come semplici strumenti software possono funzionare tra loro in modo egregio e come i risultati, ovvero le formazioni schierate da VirtualCoach, sono assolutamente in accordo con quelli sperati, anche se, con ampi margini di miglioramento.

VirtualCoach tuttavia non deve essere visto come un lavoro a sé stante ed archiviato come prodotto finito ma come un punto di inizio per un sistema più grande, più efficiente e più versatile.

Per esempio si potrebbe estendere la copertura dati a livello internazionale coprendo campionati esteri quali la Premier League inglese e la Liga spagnola, si potrebbero estendere le funzionalità alla creazione automatica della rosa (anche se questo è tutto un altro paio di maniche) e addirittura lanciarsi su altri sport quali il Football Americano.

Sicuramente si può, e diciamo si deve, migliorare la formula dell'algoritmo di ranking aggiungendo nuove variabili come ad esempio un coefficiente di rischio, per evitare di giocare con meno uomini, un bonus tifoso che predilige giocatori di una certa squadra a cui l'utente è legato (anche se può rivelarsi una mossa azzardata) oppure una serie di dati a fronte di Sentiment Analysis relative agli, ormai di uso comune, social networks.

#### 9 RIFERIMENTI

- 1. Fantagazzetta: <a href="https://www.fantagazzetta.com/probabili-formazioni-serie-a">https://www.fantagazzetta.com/probabili-formazioni-serie-a</a>
- 2. Fantagiaveno: <a href="http://www.fantagiaveno.it/quotazioni-calciatori.asp">http://www.fantagiaveno.it/quotazioni-calciatori.asp</a>
- 3. La Gazzetta dello Sport: http://www.gazzetta.it/calcio/fantanews/statistiche/serie-a-2016-17/
- 4. Sky Sport: <a href="http://sport.sky.it/statistiche/calcio/serie-a/classifiche.shtml">http://sport.sky.it/statistiche/calcio/serie-a/classifiche.shtml</a>
- 5. Corriere dello Sport: <a href="http://www.corrieredellosport.it/live/classifica-serie-a.html?cookieAccept">http://www.corrieredellosport.it/live/classifica-serie-a.html?cookieAccept</a>
- 6. VirtualCoach (applicazione): <a href="https://github.com/biagiocaldararo/VirtualCoach">https://github.com/biagiocaldararo/VirtualCoach</a>