

# Programma

Il corso è articolato su **7 giornate non consecutive di 8 ore ciascuna per un totale di 56 ore** di lezioni frontali, in cui si alterneranno insegnamenti teorici ed esercitazioni pratiche, project work e laboratori; i partecipanti saranno guidati da docenti di alto livello all'apprendimento di tecniche utili all'organizzazione e comprensione dei big data e in particolare all'elaborazione di modelli per la match analysis e lo scouting nel calcio.

Il programma didattico è articolato intorno alle tematiche esposte di seguito:

## Prima Parte

**7, 8 e 9 ottobre 2019 - Formazione erogata da TOP-IX**

### **GIORNO 1. I linguaggi e gli strumenti per estrarre valore dai dati: teoria e pratica**

- Introduzione a Python e alle funzionalità di Jupyter Notebook (code e markdown)
- Operazioni e sintassi di base
- Stringhe, tipi numerici, liste, tuple, set e dizionari
- Accesso a valori e slicing su liste
- Control flow: if statement, ciclo while e for
- List/dict/set comprehension
- Funzioni e lambda

### **GIORNO 2. Come passare da un piccolo foglio Excel all'analisi dei Big Data**

- Intro a numpy array e matrici
- Calcolo vettoriale e operazioni statistiche
- Intro a Pandas e Pandas DataFrame
- Importare/esportare file Excel e csv
- Manipolazione dei Pandas DataFrame
- Slicing e creazioni di nuove colonne con lambda
- Operazioni statistiche
- Gestione di valori nulli e outliers
- Esplorazione di variabili continue e categoriche
- Raggruppamenti
- Estrazione di nuove conoscenze e processo di Data Science

### **GIORNO 3. Come prendere i dati da Wyscout, analizzarli e renderli graficamente accattivanti: esercitazione guidata**

- Cosa sono le API di Wyscout e come funzionano
- WEB API: protocollo HTTP, CRUD e formato JSON
- Utilizzo libreria *requests* per raccolta e parsing di risorse remote
- Esplorazione di un csv o Excel con librerie Python grafiche (Matplotlib/Seaborn o Plotly)
- Ragionamento su quali tipi di grafico è opportuno usare a seconda del caso
- Uso di Pandas per sistemare DataFrame
- Estrazioni di informazioni da rappresentazioni grafiche
- Interpretazioni e presentazione delle conclusioni finali

**Seconda Parte**

**11,12,13 e 14 Novembre 2019 - Formazione erogata da *PlayeRank***

### **GIORNO 4. Introduzione ai dati di Wyscout**

Conoscere il dato su cui si lavora è il requisito principale per utilizzarlo al meglio. E' la prima fase nei processi di Data Science, chiamata Data Understanding. I partecipanti avranno quindi una panoramica sui dati forniti da wyscout e sul come filtrarli, aggregarli, analizzarne le distribuzioni. La lezione sarà interattiva: ai partecipanti saranno forniti i Jupyter notebook con cui seguire, passo-passo, le operazioni illustrate.

Argomenti affrontati:

- Parte 1
  - formato dei documenti *competition, player, team*
  - formato di un documento *match*
  - formato di un documento *event*
- Parte 2
  - aggregazioni di base su documenti event (numero passaggi, accuratezza, ecc.)
  - statistiche e visualizzazioni di base (posizione degli eventi, heatmap, ecc.)

### **GIORNO 5. Match data analysis**

Nella match analysis esistono già metriche consolidate, fornite dai principali software ed utilizzate già dagli analisti nel loro lavoro quotidiano. Tuttavia, l'analista e lo staff tecnico di cui fa parte hanno anche la loro particolare visione dei principi di sviluppo del gioco da andare a verificare. Ai partecipanti verrà illustrato come, partendo dai dati "grezzi", costruire le metriche principali e più conosciute.

Argomenti affrontati:

- definizione, ricostruzione e analisi della rete di passaggi
- metriche di performance basate sulla rete di passaggi
- calcolo pericolosità e indice di invasione
- calcolo “expected goals”

## **GIORNO 6. Player scouting**

Lo scouting è uno degli utilizzi principali dei soccer logs, ovvero i dati relativi ad ogni evento sulla palla registrato in campo. La peculiarità principale di questo tipo di dati è infatti la sua estensione: non solo i principali campionati italiani, ma anche le seconde divisioni, i campionati primavera ed i campionati “minori” vengono coperti dalla raccolta dati, rendendo questo tipo di dato ottimo per la ricerca dei calciatori. La performance di un calciatore, nel nostro ambito, è un insieme di eventi osservati durante una partita disputata dal calciatore. Ai partecipanti verrà illustrato l’algoritmo Playerank, che permette di sfruttare tecniche di intelligenza artificiale per confrontare le performance dei calciatori, espresse sotto forma di eventi registrati sul campo. Playerank fornisce inoltre all’analista la possibilità di personalizzare l’algoritmo e di modellare e validare la propria ipotesi di valutazione della performance.

Argomenti affrontati

- valutazione delle performance con l’algoritmo PlayeRank
- ricerca nel database dei calciatori e confronto dell’evoluzione della performance
- personalizzazione della valutazione attraverso cambio dei parametri dell’algoritmo
- calcolo della similarità tra calciatori

**14 Novembre 2019 - Formazione erogata da *The Information Lab***

## **GIORNATA 7: Data Visualization con Tableau**

- Introduzione dati Wyscout & Setting Up Connections and Data Sources
- Area di lavoro
- Analisi temporali
- Correlazioni fra dati
- Mappare dati sul campo (Shoot analysis, Passing Path)
- Calcoli
- % di Contributo
- Creazione di cruscotti interattivi