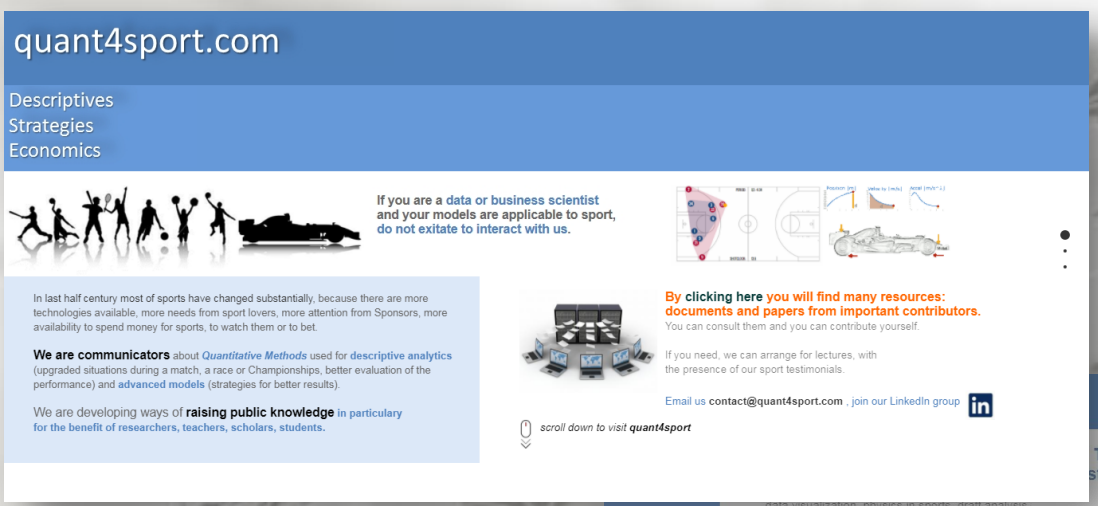
SPORT E STATISTICA: UN CONNUBIO VINCENTE

quant4sport.com

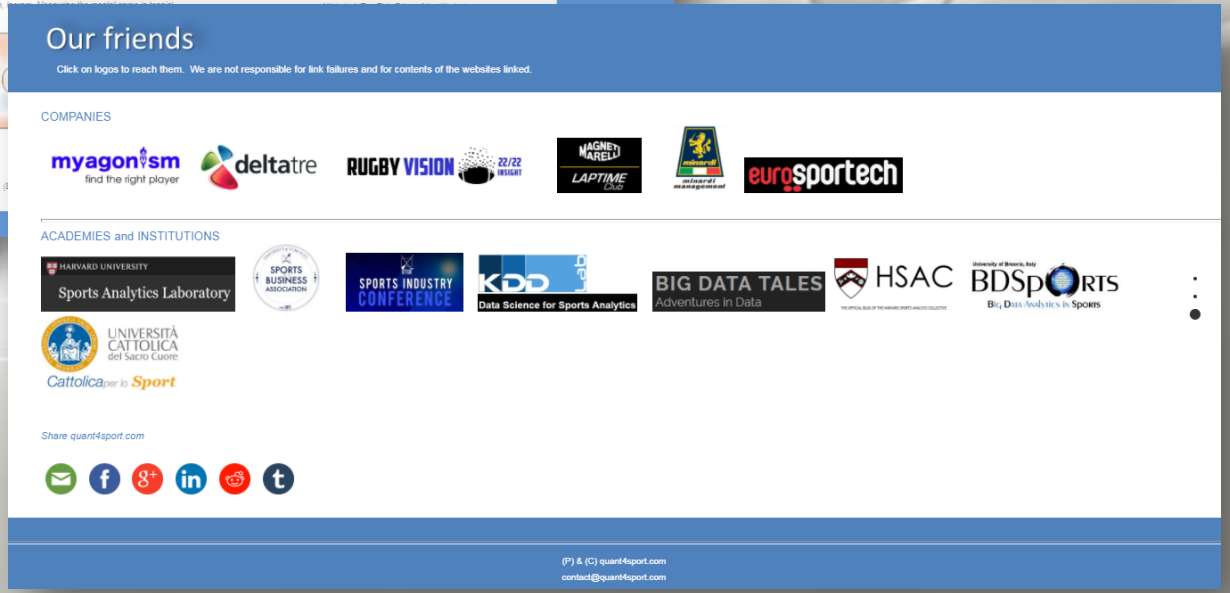
1 dicembre 2017

UNIVERSITÀ DEGLI STUDI DI TORINO
DIPARTIMENTO DI MANAGEMENT



quant4sport è un progetto di condivisione delle conoscenze in ambito economico e matematico riferite allo sport. Si è voluto da subito dare anche una connotazione molto pratica al progetto.

E' per questo motivo che è nato quant4sport.com che è una piattaforma che raccoglie principalmente documenti accademici relativi allo sport declinato nelle prospettive economiche e matematico/statistiche.



Quant4sport ha rapporti di collaborazione e di condivisione con una serie di “friends” che sono tutti raggiungibili direttamente dalla pagina dedicata presente su quant4sport.com

HARVARD UNIVERSITY
Sports Analytics Laboratory

University of Brescia, Italy
BDSports
Big Data Analytics in Sports

HSAC
THE OFFICIAL BLOG OF THE HARVARD SPORTS ANALYSIS COLLECTIVE

UNIVERSITY of TORONTO
SPORTS BUSINESS ASSOCIATION
est 2011

MIT Sloan
SportsAnalyticsConference

KDD Lab
Data Science for Sports Analytics

BIG DATA TALES
Adventures in Data

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

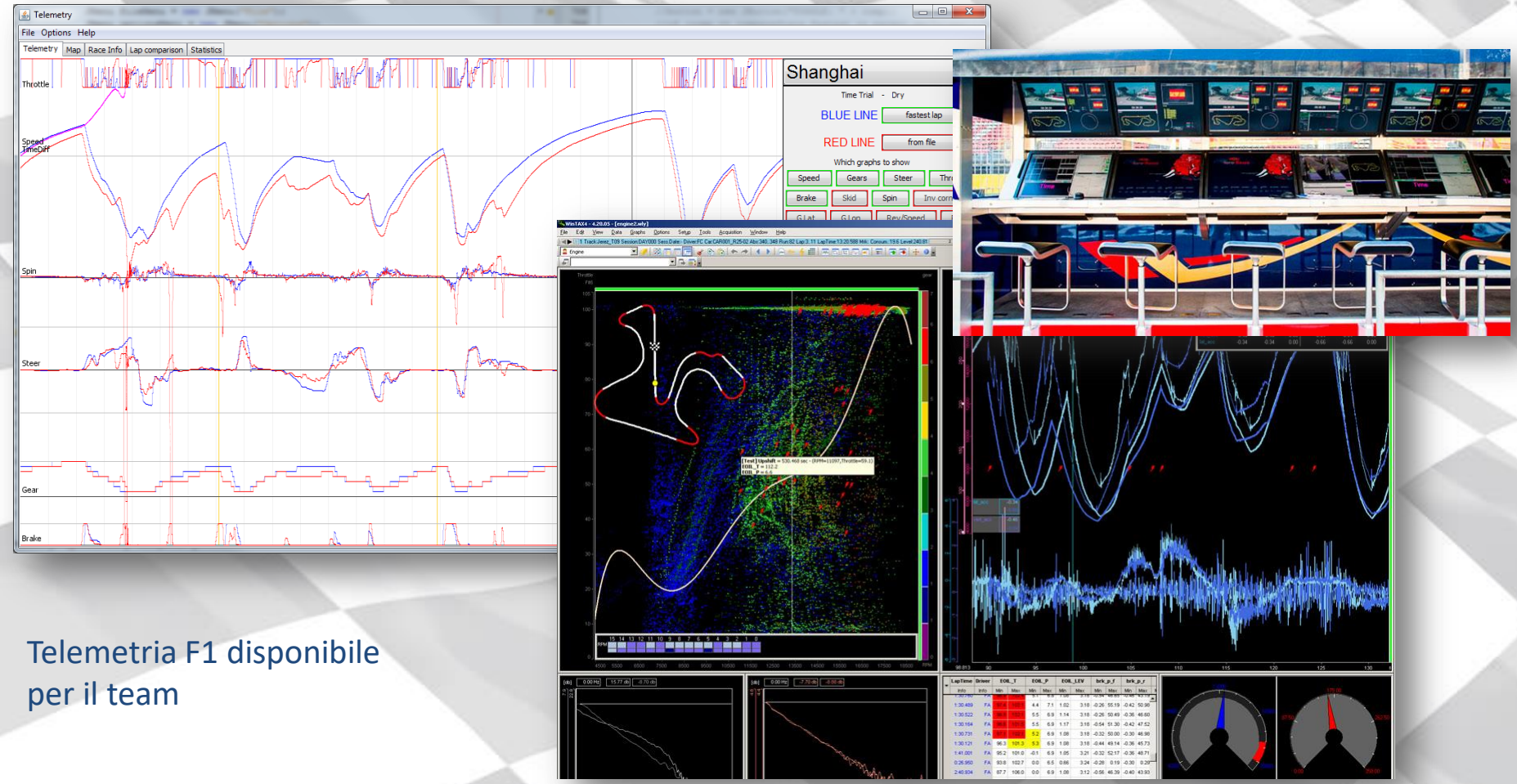
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
DEPARTMENT OF STATISTICS

Sport Research

UTSB
SPORTS INDUSTRY CONFERENCE

Analisi
Descrittiva
Per il miglioramento della performance
Pubblico, scommesse
Teams
Sponsors
Impatti economici
Sponsorizzazioni, biglietteria, merchandising

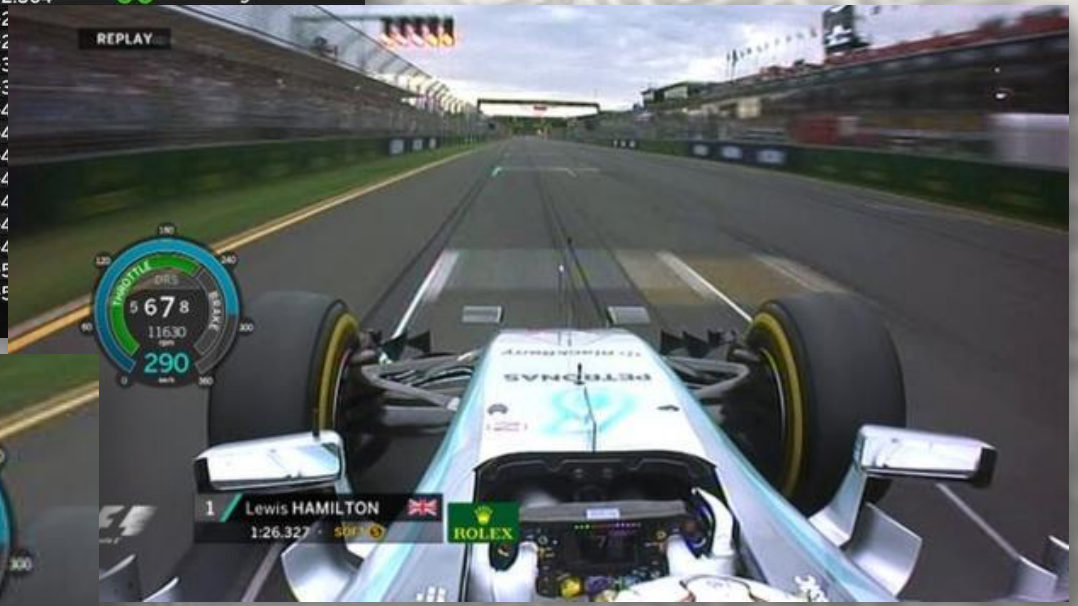


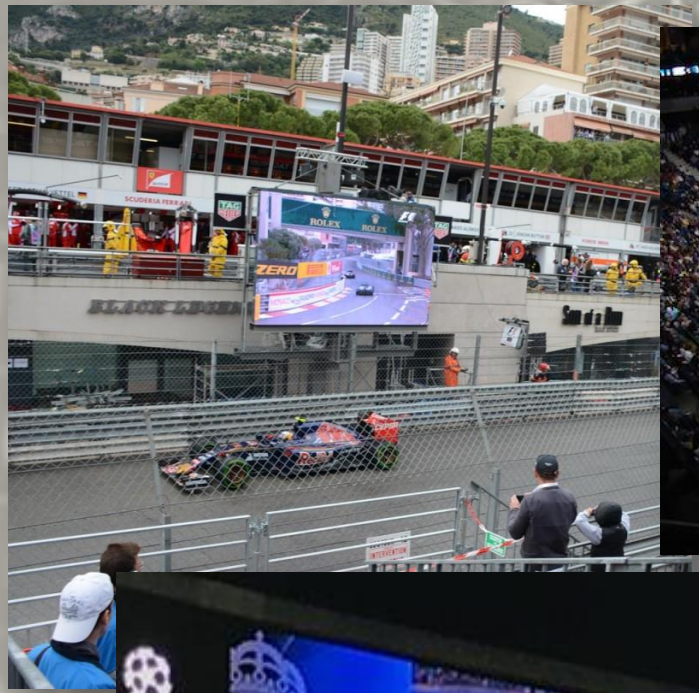
Telemetria F1 disponibile per il team

Telemetria F1 disponibile per il pubblico

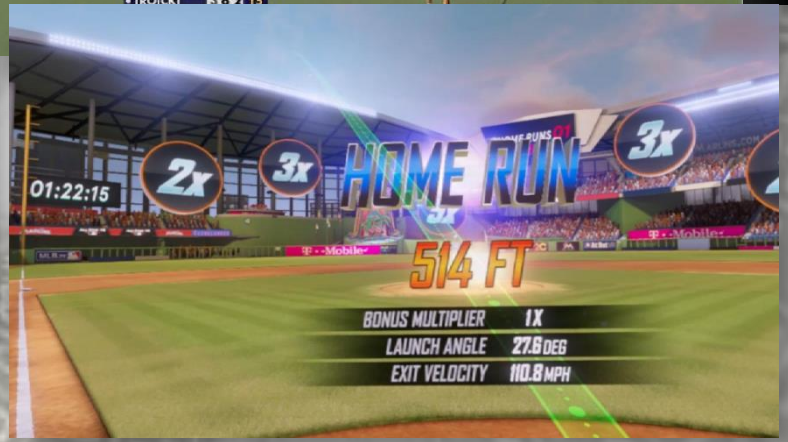
F1® QUALIFYING 1 CLASSIFICATION — SILVERSTONE — #BritishGP

			Q1 Lap Time	Gap	Stint	Laps
1	Fernando Alonso	McLaren	1:37.598		I I S	9
2	Max Verstappen	Red Bull Racing	1:38.912	+1.314	I I	8
3	Lewis Hamilton	Mercedes	1:39.069	+1.471	S I I	10
4	Nico Hulkenberg	Renault	1:39.201	+1.603	I I	7
5	Valtteri Bottas	Mercedes	1:39.698	+2.100	S I I	10
6	Esteban Ocon	Force India	1:39.738	+2.140	S I I S	10
7	Sebastian Vettel	Ferrari	1:39.962	+2.364	I I	9
8	Stoffel Vandoorne	McLaren	1:40.011	+2		
9	Kimi Raikkonen	Ferrari	1:40.455	+2		
10	Carlos Sainz	Toro Rosso	1:41.114	+3		
11	Jolyon Palmer	Renault	1:41.404	+3		
12	Daniil Kvyat	Toro Rosso	1:41.726	+4		
13	Felipe Massa	Williams	1:41.874	+4		
14	Sergio Perez	Force India	1:42.009	+4		
15	Romain Grosjean	Haas	1:42.042	+4		
16	Lance Stroll	Williams	1:42.573	+4		
17	Kevin Magnussen	Haas	1:42.577	+4		
18	Pascal Wehrlein	Sauber	1:42.593	+4		
19	Marcus Ericsson	Sauber	1:42.633	+5		
20	Daniel Ricciardo	Red Bull Racing	1:42.966	+5		

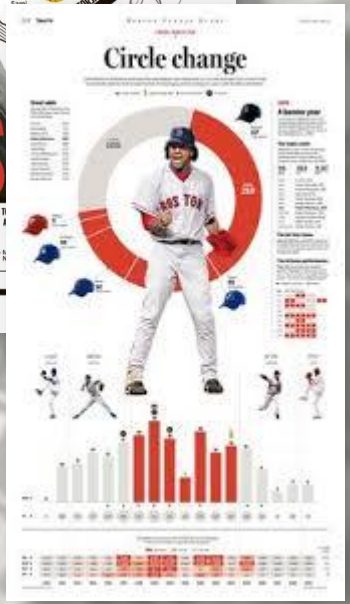




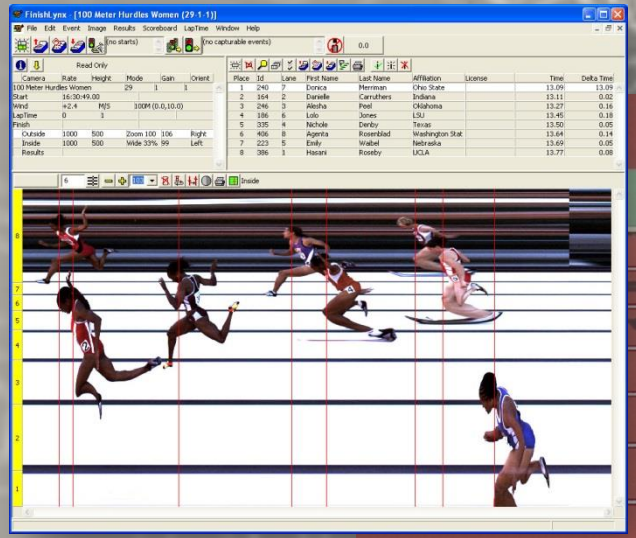
Informazioni disponibili per il pubblico sono presenti in ogni sport, da fruire direttamente sul luogo di gara…

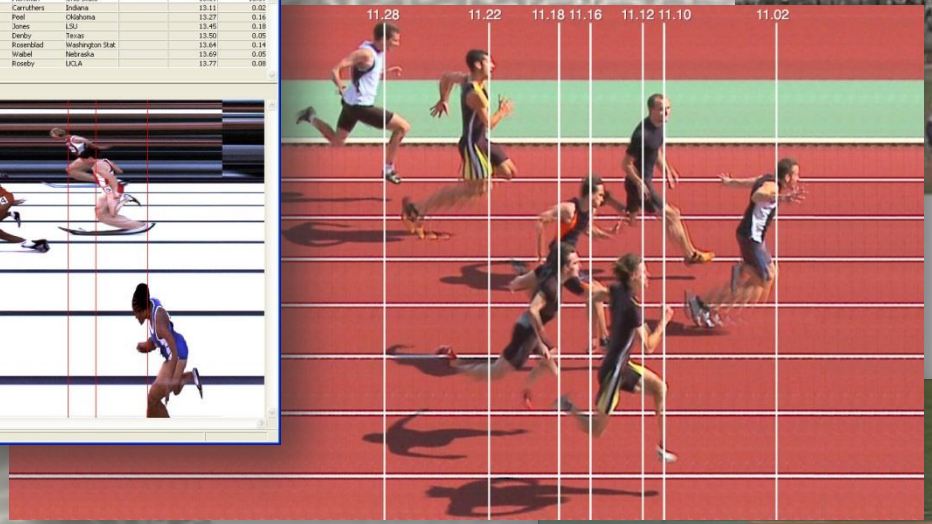


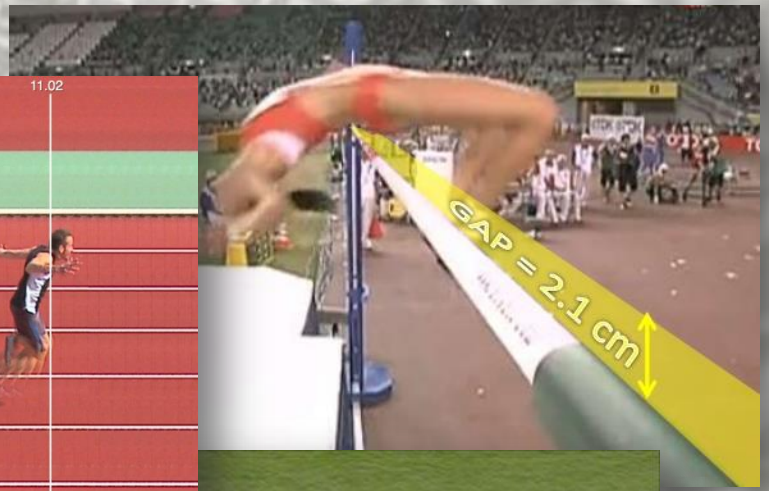
…oppure in remoto su televisioni o device di varia natura.



Le informazioni statistiche sono utilizzate per le infografiche di riviste e siti web







I medesimi sensori che forniscono i dati per i team o per il pubblico sono un aiuto essenziale per le analisi di svolgimento della gara, anche per sport che non hanno basi tecnologiche avanzate (come la F1).

ANALYTICS con tracking fisico

Alcuni applicativi forniscono agli allenatori prodotti e servizi innovativi per tracciare ed analizzare statistiche di livello avanzato e dati relativi ai loro giocatori, in maniera semplice ed affidabile.

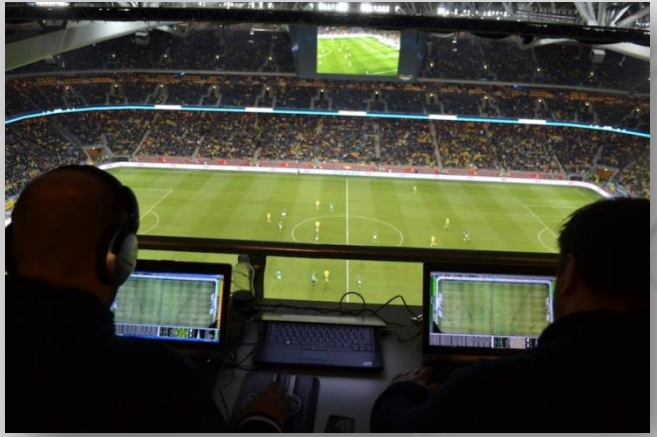
Con dei «device» è possibile tracciare i movimenti dei giocatori e della palla, ottenendo immagini 2D o 3D.

Ad esempio nel basket:

- Velocità e accelerazione;
- Numero degli sprint, clusterizzati (<5m, <10m, >10m);
- Elevazioni;
- Movimento in campo;
- …

Agonism Your Scoreboard Tracker
Find The Right Player

ANALYTICS con tracking ottico



Il Tracking Ottico usa le immagini per determinare le posizioni di ogni punto di interesse (giocatori, arbitri e guardalinee) e fornisce, tramite una scansione a 25 frames al secondo, dati in tempo reale utilizzando le coordinate spaziali di tre dimensioni e calcolando la posizione, il movimento, la velocità…

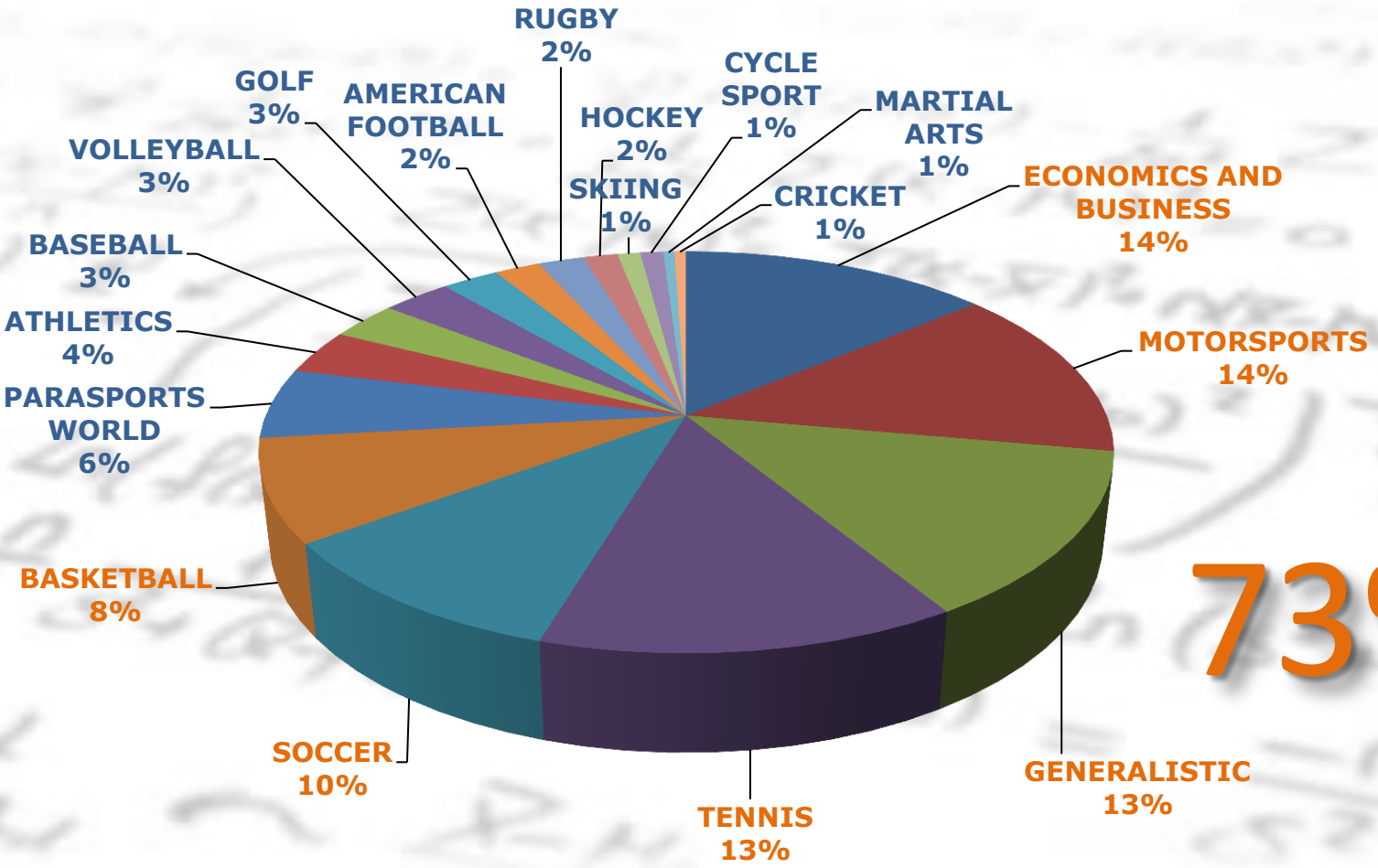
I dati vengono generati in un formato adatto per una moltitudine di piattaforme: sistemi grafici broadcast, web, piattoforme mobile. Il Tracking Ottico fornisce statistiche riguardo la performance, utile per le analisi.

AREE COPERTE

(dicembre 2017)

- Economics and business
- Soccer
- American football
- Baseball
- Basketball
- Motorsports
- Rugby
- Athletics
- Tennis
- Skiing
- Cycle sport
- Martial arts
- Hockey
- Golf
- Volleyball
- Cricket
- Parasport world

I documenti raccolti da quant4sport nei primi mesi di attività sono circa duecento e coprono una grande quantità di sport. A fianco di alcuni documenti di carattere più ampio, riferiti al contesto sportivo in generale, la maggioranza dei paper si concentrano nel calcio, negli sport motoristici, nel tennis e nel basket.
E' presente una sezione dedicata alla documentazione relativa agli sport per gli atleti diversamente abili.

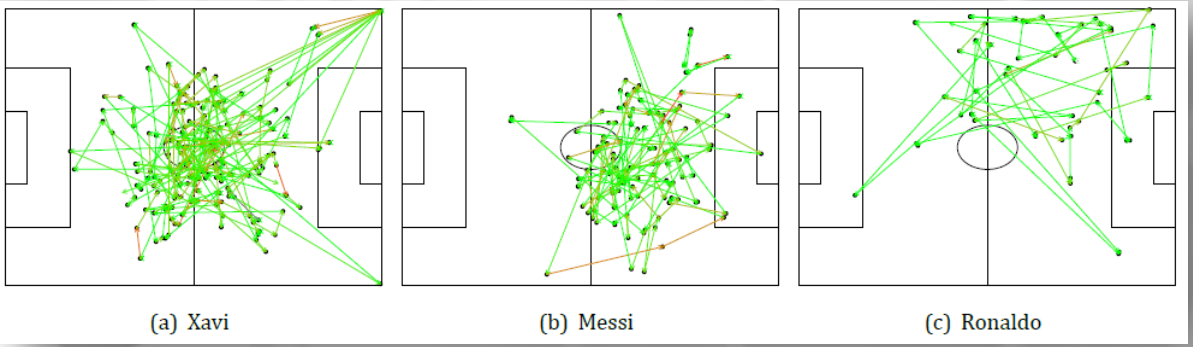

RUGBY 2%
CYCLE SPORT 1%
MARTIAL ARTS 1%
ECONOMICS AND BUSINESS 14%
CRICKET 1%
SKIING 1%
HOCKEY 2%
AMERICAN FOOTBALL 2%
GOLF 3%
VOLLEYBALL 3%
BASEBALL 3%
ATHLETICS 4%
PARASPORTS WORLD 6%
MOTORSPORTS 14%
BASKETBALL 8%
SOCCER 10%
TENNIS 13%
GENERALISTIC 13%

73%

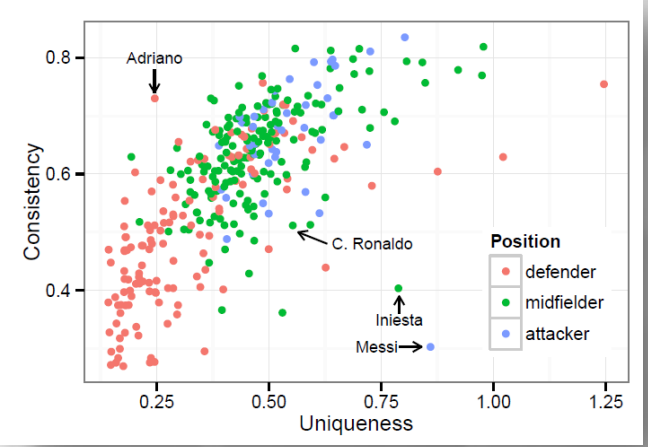
CALCIO

PROFILING MOVEMENTS

Tramite un dataset generato da OPTA (2021/13) sono stati rilevati il punto di inizio e di fine del movimento del giocatore (sugli assi x e y), la velocità del vento ed il possesso palla a quel momento. Questo modello è stato ritenuto utile per verificare le similitudini delle giocate. La sofisticazione del modello include la velocità della corsa, elevata disegnata in verde, lenta disegnata in rosso.



L'unicità e la consistenza dei giocatori nel campionato. Il colore dei punti indica la posizione dei giocatori. I valori anomali delle posizioni sono rispettivamente Adriano, Iniesta e Messi.

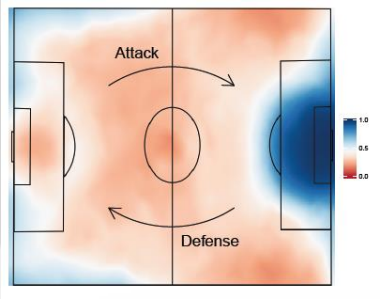


Analyzing In-Game Movements of Soccer Players at Scale
László Gyarmati, Mohamed Hefeeda
Qatar Computing Research Institute, HBKU

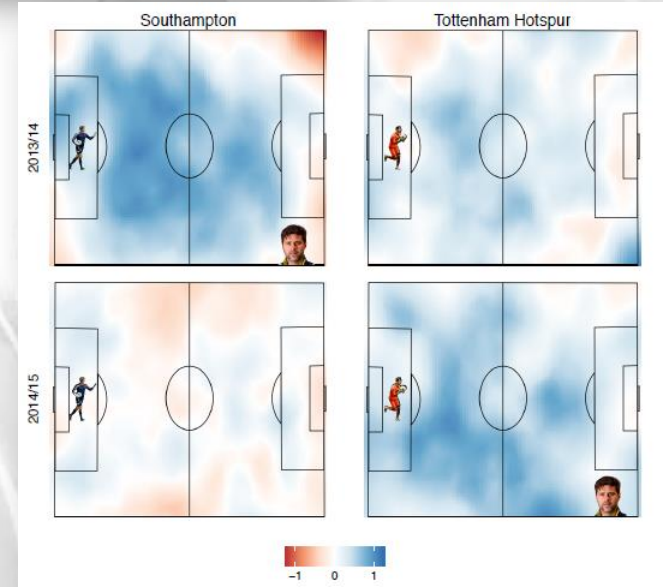
CALCIO

CARTOGRAFIA

I passaggi costituiscono un ruolo chiave nel gioco del calcio, perché autiano la squadra a mantenere il possesso palla. Ovviamente le strategie sono opposte a seconda che si veda da una prospettiva di difesa piuttosto che di attacco. L'interruzione del possesso palla può essere analizzata mediante una cartografia appropriata.



L'immagine rappresenta la superficie media di interruzione per la Premier League inglese 2014/15. Il blu indica un'alta probabilità di un'interruzione mentre il rosso indica una bassa probabilità. I dati sono girati in modo tale che la squadra con la palla tira da sinistra a destra e la squadra in difesa sta giocando da destra a sinistra.



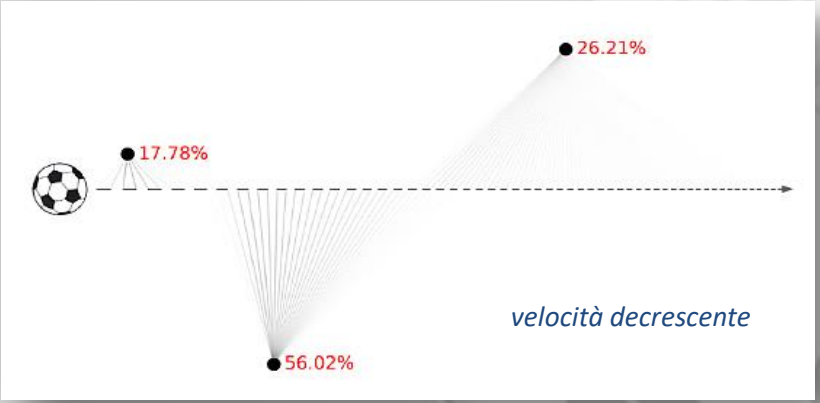
L'effetto Pochettino.
Superfici di interruzione media per Southampton e Tottenham Hotspur durante il 2013/14 e Premier League inglese 2014/15. L'immagine di Mauricio Pochettino nell'angolo in basso a destra indica quale squadra lui stava allenando. Il rosso indica sotto la media, il bianco indica la media e il blu indicati sopra i valori medi.
L'allenatore e i suoi metodi hanno fatto la differenza per il Tottenham Hotspur nella stagione 2014/15.

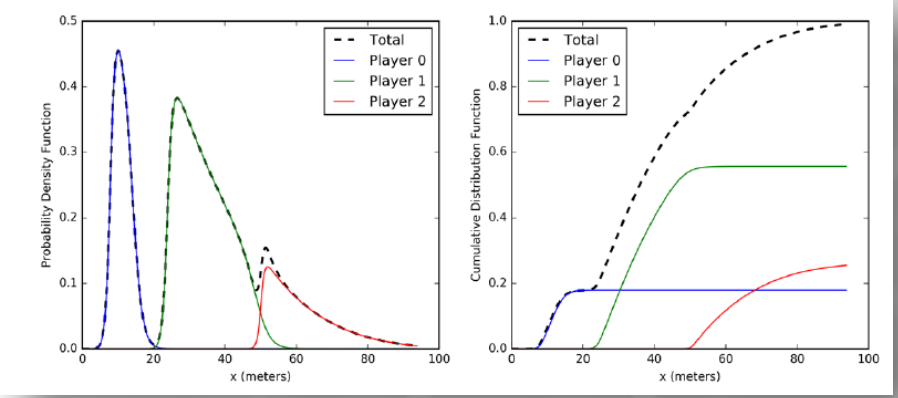
The Pressing Game: Optimal Defensive Disruption in Soccer
Iavor Bojinov and Luke Bornn

CALCIO

PROBABILITA’

Sempre in merito ai passaggi di palla, i modelli probabilistici sono utilizzati per modellizzare tattiche opportune da applicare in campo.





Questa figura mostra la traiettoria del pallone da calcio e le possibili traiettorie di intercettazione per tre giocatori, con le relative probabilità di successo.
Il primo dei giocatori dovrebbe spostarsi molto rapidamente per intercettare il pallone; l'ultimo giocatore ha molte possibili traiettorie di intercettazione, ma da allora è probabile che i primi due giocatori abbiano già intercettato la palla.

a) La funzione di densità di probabilità;
b) La funzione di distribuzione cumulativa.

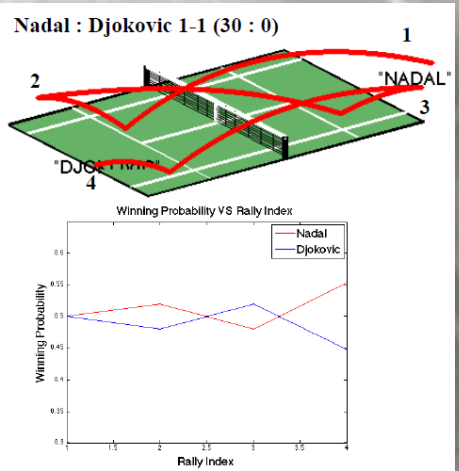
Ogni giocatore è rappresentato separatamente dalle linee colorate mentre la somma per i tre giocatori è rappresentata dalla linea nera tratteggiata.

Physics-Based Modeling of Pass Probabilities in Soccer
William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop

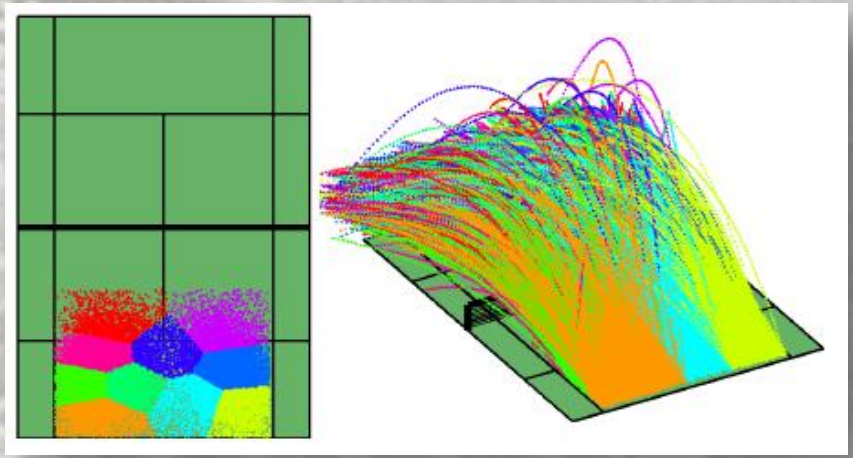
TENNIS

TRAJECTORIES MOVEMENTS

Mediante l'utilizzo di tre anni di Hawk-Eye degli Australian Open maschili sono stati analizzati 37727 battute. In questo caso il sistema Hawk-Eye utilizza tre posizioni della pallina: x, y e z, quindi anche l'altezza da terra.



Il modello consente di stimare la probabilità che la battuta porti ad un punto vincente.
Fra le analisi possibili, la clusterizzazione delle traiettorie dei tiri rappresenta un buon sistema.

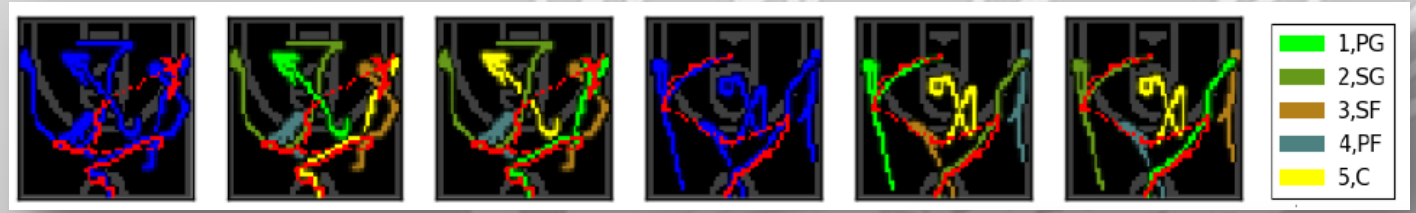


"The Thin Edge of the Wedge": Accurately Predicting Shot Outcomes in Tennis using Style and Context Priors
Xinyu Wei1, Patrick Lucey2, Stuart Morgan3, Machar Reid4 and Sridha Sridharan1
1Queensland University of Technology, 2STATS LLC, 3Australian Institute of Sport, 4Tennis Australia

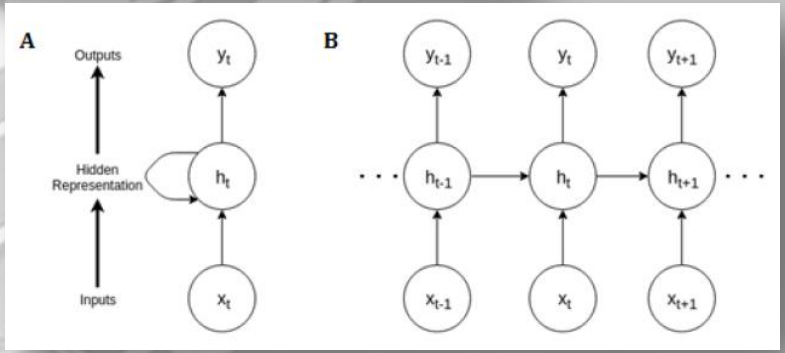
BASKET

NEURAL NETWORKS

Applicando tecniche di machine learning per elaborare direttamente i dati di tracciamento di SportVU, in particolare le varianti di reti neurali, è stato possibile etichettare correttamente un grande numero di sequenze, circa il doppio dei dati rispetto ad un intervento umano e con alta precisione, utilizzando solo una frazione del tempo.



La traccia rossa rappresenta la palla. Dopo aver eseguito la traccia grafica è stato ultizzato un sistema classico di rete neurale. Gli input corrispondono alle coordinate del giocatore di SportVU, e gli output sono le etichette di classe per le diverse strategie offensive. Il diagramma A mostra come appare realmente il modello in cui le connessioni nascoste si alimentano da sé. Il diagramma B illustra come appare la rete nel tempo.

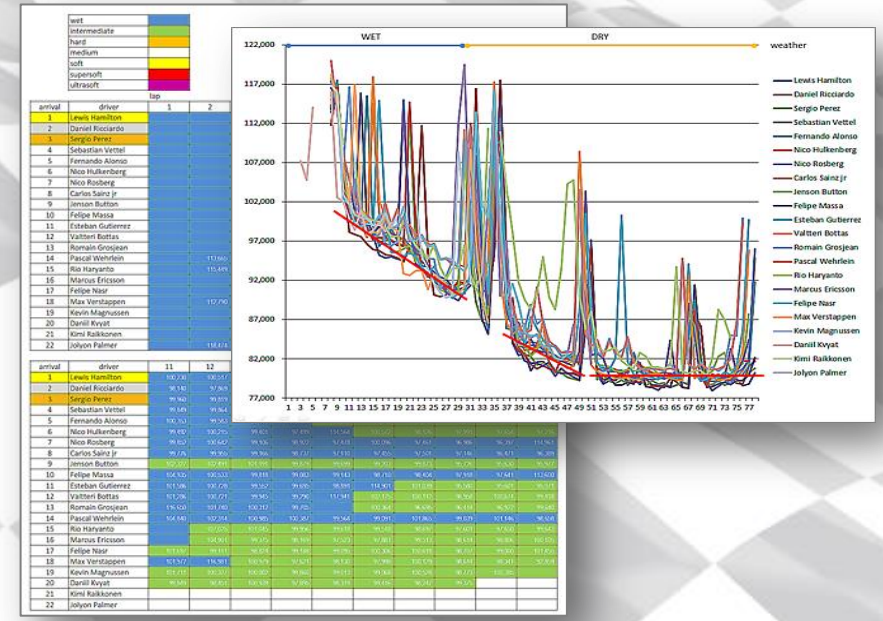


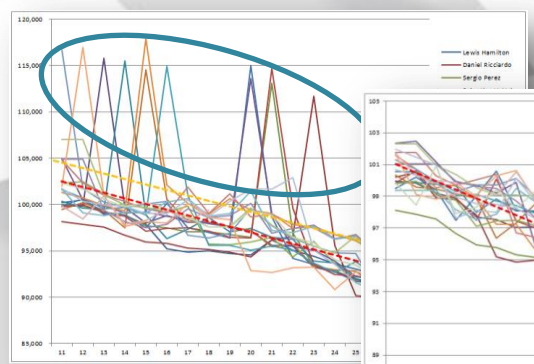
Classifying NBA Offensive Plays Using Neural Networks
Kuan-Chieh Wang, Richard Zemel
University of Toronto

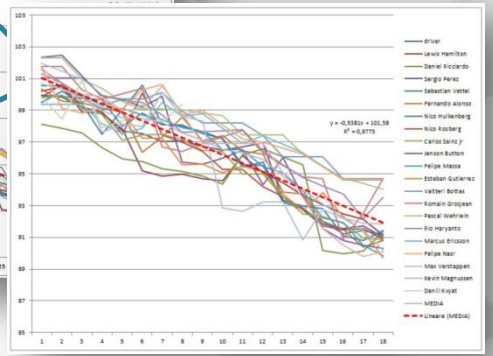
MOTORSPORTS

ANALISI DI REGRESSIONE – POT Peaks over Threshold

Mediante la telemetria dei team sono stati raccolti i tempi sul giro, con particolare riferimento ai tempi durante i pit-stops. Pirelli ha fornito i dati relativi agli pneumatici utilizzati. L'analisi di regressione abbinata al metodo POT (Peaks over Threshold) ha consentito di affinare l'utilizzo dei dati raccolti e di trarre opportune considerazioni.







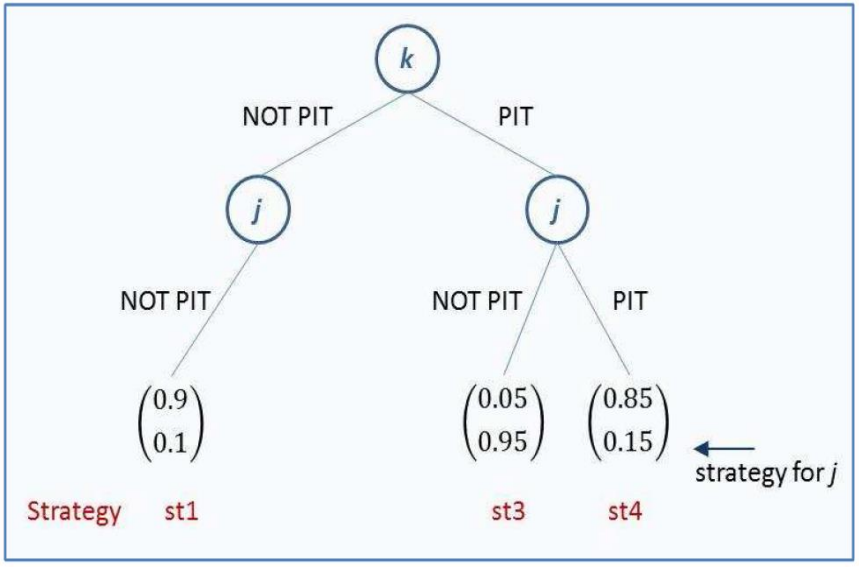
Sono state determinate le regole che hanno determinato l'ordine di arrivo, in considerazione della forte variabilità del clima, della grossa quantità di soste e di relativi cambi gomme e della lunghezza dei singoli stint.

Statistical Analysis of F1 Monaco Grand Prix 2016. Relations Between Weather, Tyre Type and Race Stints.
Gianluca Rosso, Andrea Filippo Rosso.

MOTORSPORTS

ALBERI DECISIONALI

Altro caso quello che riguarda la modellizzazione nel corso della gara. Durante lo svolgimento della competizione si verificano costantemente situazioni che impongono a decisioni immediate. Le gare sono una continua gestione della situazione del momento, a prescindere dalla preparazione effettuate ante gara, e nell'ottica di ottenere il miglior risultato possibile. Ciò che in alcuni casi viene definito solamente con il termine di gara o partita, in molti altri casi prende più propriamente il significato di gioco. Per questo la teoria dei giochi viene incontro per aiutare l'analista a definire le situazioni in evoluzione. Ed in particolare alcuni strumenti appartenenti alla teoria dei giochi diventano utili per scegliere le strategie migliori. Fra questi gli Alberi Decisionali. Questo Albero Decisionale illustrato è riferito ad una simulazione per i Pit Stop per la Formula1.

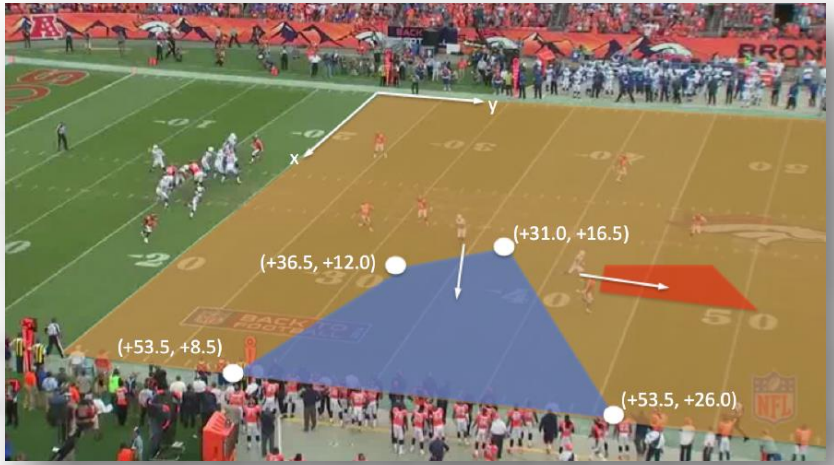


Assess the risk of pit stopping, using probability and Game Theory with sequential Game Decision Making.
Gianluca Rosso

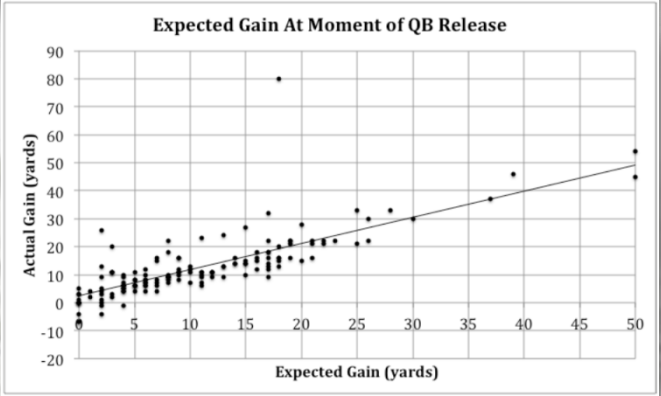
FOOTBALL AMERICANO

ANALISI QUANTITATIVA GEOSPAZIALE

Identificazione mediante radiofrequenza (RFID) è in uso a NFL fin dal 2014.
Il sistema consente di analizzare i dati relativamente alla performance dei giocatori ed alle loro capacità di decision-making, specialmente per il ruolo del Quarterback.
In particolare viene analizzato il movimento dei singoli giocatori in rapporto alle loro zone di maggiore influenza.



L'output dell'analisi geospaziale viene utilizzata con modelli statistici più tradizionali, come ad esempio lo scatter-plot, per le valutazioni del caso, come in questo caso il confronto fra il guadagno atteso in yards per ogni completamento di passi (148 in totale) contro il guadagno attuale detrminato in assenza del modello statistico.

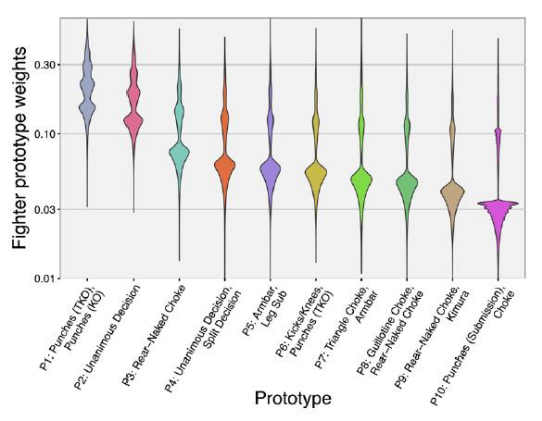


Finding the Open Receiver: A Quantitative Geospatial Analysis of Quarterback Decision---Making
Jeremy Hochstedler

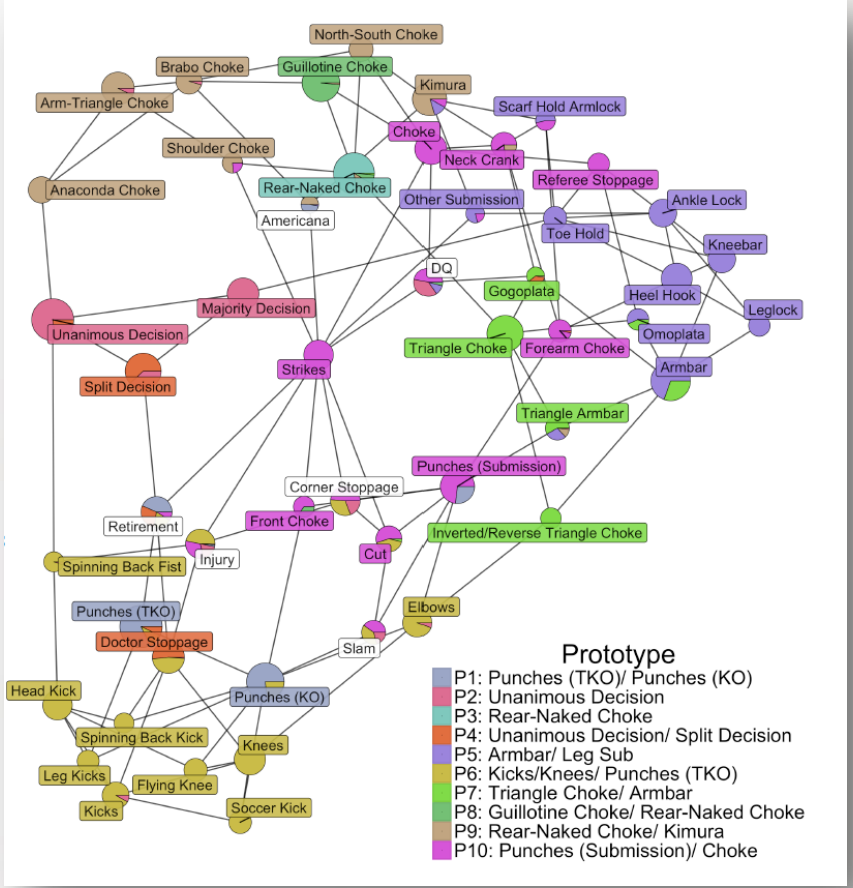
ARTI MARZIALI

GRAPPLING & NETWORK

Grazie alle strutture a rete e a grappolo, mediante una serie di opportuni algoritmi, è stato possibile studiare l'efficacia delle mosse finali nelle arti marziali, utilizzando un considerevole database ed eseguendo simulazioni ed ipotesi.



Il cosiddetto «diagramma a violino» evidenzia le distribuzioni delle concentrazione delle mosse dei dei lottatori.

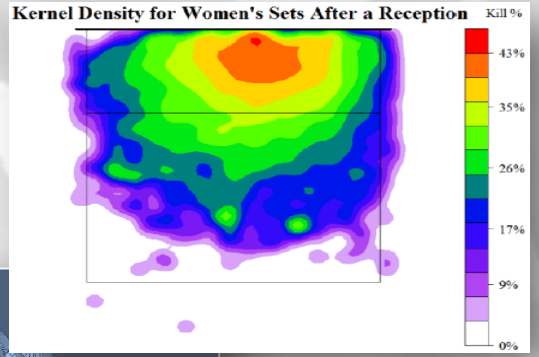


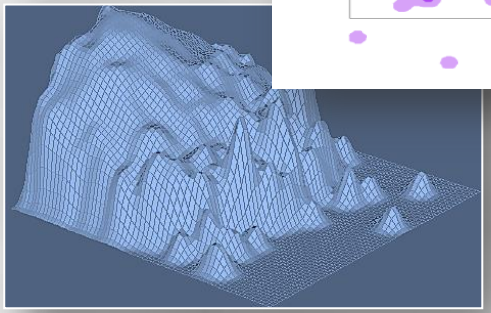
Mixed Membership Martial Arts: Data-Driven Analysis of Winning Martial Arts Styles
Sean R. Hackett and John D. Storey

PALLAVOLO

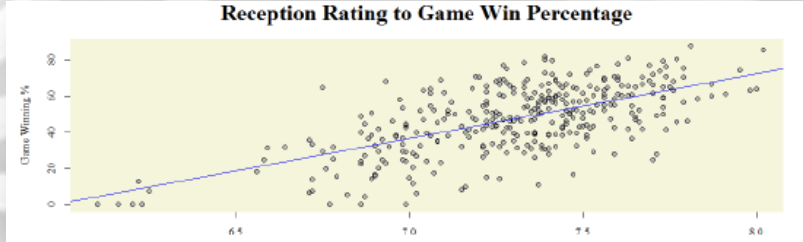
ANALISI DI DENSITA' 2D – 3D

VolleyMetric è un sistema in uso da diversi anni in grado di rilevare diversi parametri durante il gioco. Tramite l'analisi delle coordinate di gioco, in particolare nel momento della battuta/ricezione, e tramite l'ausilio di R è possibile ottenere elaborazioni accurate.





I diagrammi possono indicare la posizione preferenziale di gioco al momento della ricezione, e possono essere disegnati bidimensionalmente o tridimensionalmente.



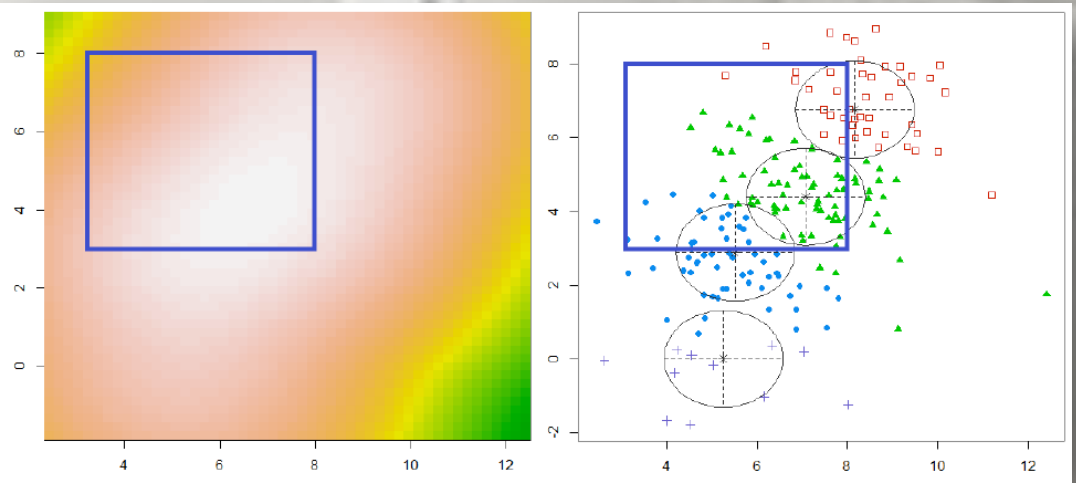
Semplici grafici a dispersione consentono di monitorare le relazioni fra i parametri, come ad esempio il rapporto fra ricezione e game vincente.

Bump, Set, Spike: Using Analytics to Rate Volleyball Teams and Players
Corley Bagley, Brett Ware

BASEBALL

ZONE RATING

Pecota, ZiPS and Steamer sono alcuni fra i «projection systems» utilizzati nel campo del baseball.
La tecnica «zone rating» risulta utile per analizzare il lancio in relazione alla zona di strike.
L'abbinamento fra un modello complesso, la tecnica di regressione e l'analisi di covarianza genera mappe indispensabili per analizzare le accuratezze dei lanci.



Arsenal/Zone Rating: A PitchF/X based pitcher projection system
Pei Zhe Shu

http://www.quant4sport.com

in Home quant4sport

YouTube quant4sport

contact@quant4sport.com
quant4sport@gmail.com

GRAZIE

quant4sport.com