



*Università degli studi di Brescia*  
*Corso di laurea magistrale in Management*  
*Curriculum in marketing*

# Big Data e Machine Learning: nuovi orizzonti per l'analisi dei dati nel basket

Relatore:

Chiar.ma Prof.ssa Paola Zuccolotto

Correlatore:

Chiar.ma Prof.ssa Marica Manisera

Laureando:

Davide Verzeletti

Matricola n. 84520



## Indice:

- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket
- Le probabilità di fare canestro: l'albero di classificazione
- Nuovi indici di tiro

- **Gli obiettivi dell'analisi**
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- Nuovi indici di tiro

## Obiettivo principale:

analizzare la performance di tiro delle squadre tenendo conto di alcuni fattori che possono generare situazioni di gioco di stress o di forte pressione psicologica.

## Alcune considerazioni da tenere presente:

1. È necessario un numero elevato di dati provenienti dai play-by-play per ottenere stime robuste → Big Data
2. Complesse relazioni ed interazioni tra le variabili coinvolte caratterizzano il fenomeno analizzato → Tecniche di analisi multivariate e di machine learning (in particolare, l'algoritmo CART)
3. Differenti reazioni alle situazioni di stress da parte dei giocatori, che possono rendere difficile individuare situazioni di stress

- Gli obiettivi dell'analisi
- **Il dataset**
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- Nuovi indici di tiro

- Play-by-play di tutte le partite della stagione regolare 2015/2016 del campionato Italiano di basket serie A2 → web scraping

Tempo	Alma-Agenzia per il Lavoro Trieste	Punti	Margine	Tezenis Verona
1 00:00	Fine partita	39-68	-29	
2 00:09		39-68	-29	Marco Spanghero, rimbalzo difensivo
3 00:10	team, tiro da tre punti sbagliato	39-68	-29	
4 00:32		39-68	-29	team, tiro da due punti realizzato
5 00:38	Lorenzo Baldasso, palla persa	39-66	-27	
6 00:39		39-66	-27	team, recupero
7 00:47		39-66	-27	Marco Spanghero, palla persa
8 00:48	Lorenzo Baldasso, recupero	39-66	-27	
9 00:52	team, palla persa trattamento di palla	39-66	-27	
10 00:53		39-66	-27	Marco Spanghero, recupero

- Centinaia di istruzioni su R per migliorare la struttura e aggiungere alcune variabili (ad esempio: la posizione in classifica, la differenza di punteggio in classifica, la sequenza di vittorie/sconfitte ecc.)

- Gli obiettivi dell'analisi
- **Il dataset**
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- Nuovi indici di tiro

<u>Dataset</u>	Serie A2ITA
Competizione	Campionato – Stagione regolare
Periodo	04/10/2015 – 23/04/2016
Numero di partite	480
Numero di squadre	32
Numero di giocatori	438
Numero di tiri	69688 (50,6 % realizzati)
Numero di tiri da 2 punti	33682 (48,3 % - 50,9 % realizzati)
Numero di tiri da 3 punti	21163 (30,4 % - 34,1 % realizzati)
Numero di tiri liberi	14843 (21,3 % - 73,5 % realizzati)

- Analisi su squadra e giocatori di Tezenis Verona dopo ogni giornata di campionato, sotto la supervisione del coach Marco Crespi
- Stesse analisi anche per l'avversario della giornata successiva

- Gli obiettivi dell'analisi
- Il dataset
- **Le situazioni di stress nel basket**

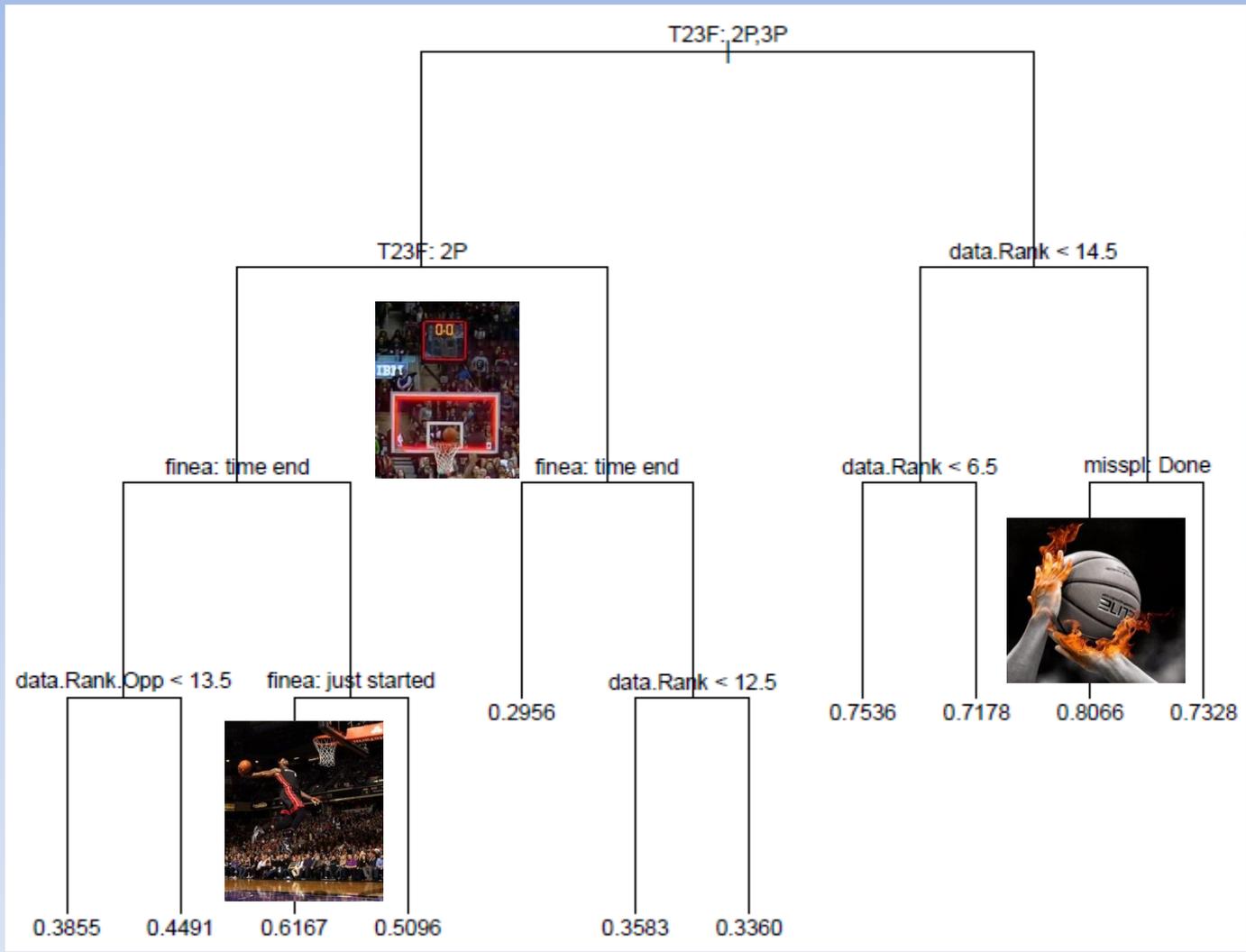
- Le probabilità di fare canestro: l'albero di classificazione
- Nuovi indici di tiro

Le situazioni di stress individuate:

- quando il cronometro dei 24 secondi sta scadendo (FINEA)
- quando il divario di punteggio è piccolo (FORBICE)
- quando la squadra non sta mantenendo una performance globale buona, fino al momento considerato (MISSTEAMP)
- quando il giocatore ha sbagliato il tiro precedente (MISSPL)
- quando il divario di punteggio è piccolo e ci si trova negli ultimi 5 minuti di partita (FINEMFORB)

- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- **Le probabilità di fare canestro: l'albero di classificazione**
- Nuovi indici di tiro



- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**

Sulla base dell'albero di classificazione, sono stati elaborati nuovi indici di tiro:

### • PERFORMANCE DI TIRO

$$P_i(T) = \text{av}_{j \in J_T} (x_{ij} - \pi_{ij})$$

### • DIFFERENZIALE DI PERFORMANCE

$$D_i^{(S)}(T) = \text{av}_{j \in J_T \cap J_S} (x_{ij} - \pi_{ij}) - \text{av}_{j \in J_T} (x_{ij} - \pi_{ij}) = P_i^{(S)}(T) - P_i(T)$$

### • PROPENSIONE AL TIRO

negli ultimi 2 secondi

$$R_i^{(S)}(T) = \left( \frac{\sum_{j \in J_T \cap J_S} x_{ij}}{\sum_{j \in J_T} x_{ij}} \right) \cdot 100$$

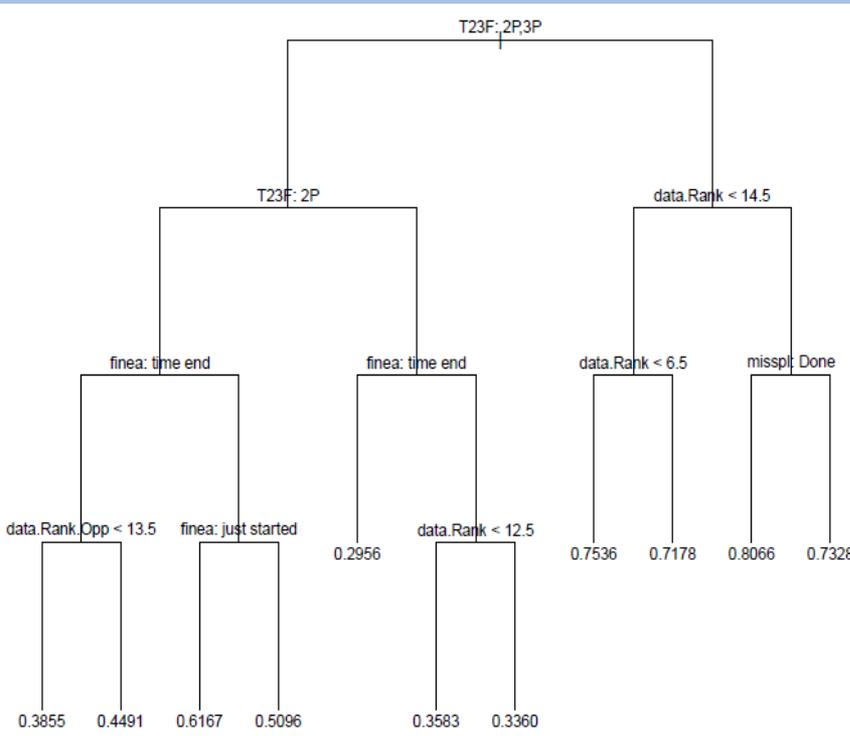
quando il punteggio è all'interno della forbice [-4; 4]

$$R_i^{(S)}(T) = \left[ \frac{\frac{\sum_{j \in J_T \cap J_S} x_{ij}}{\sum_{j \in J_T} x_{ij}}}{\frac{T_i^{(S)}}{T_i}} - 1 \right] \cdot 100 = \left[ \frac{\sum_{j \in J_T \cap J_S} x_{ij}}{\sum_{j \in J_T} x_{ij}} \cdot \frac{T_i}{T_i^{(S)}} - 1 \right] \cdot 100$$

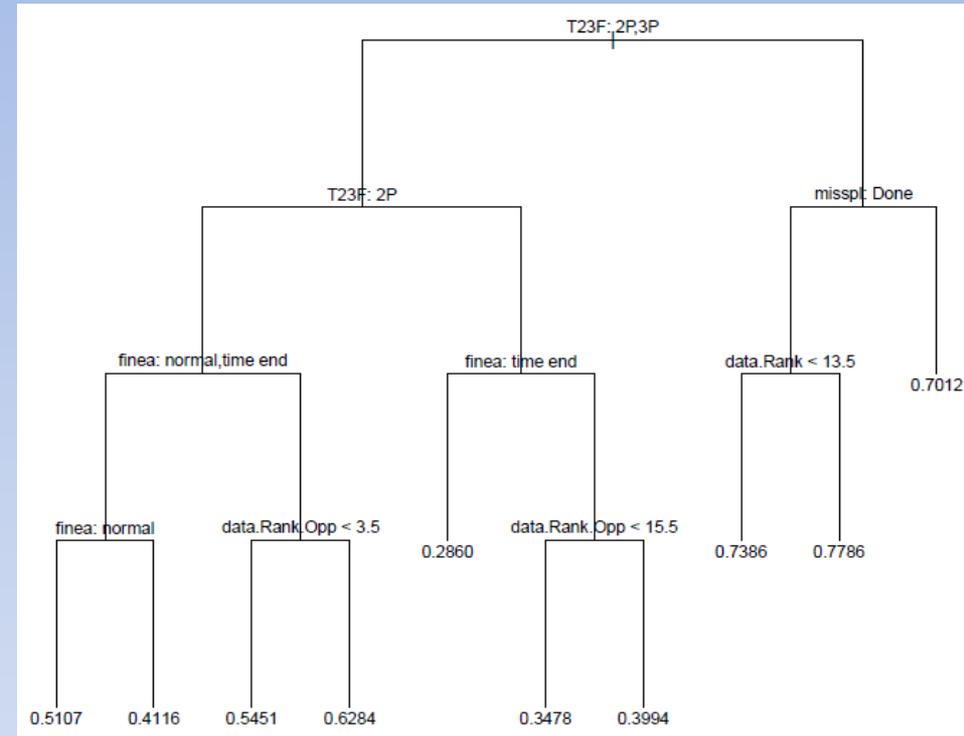
- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**

Girone di andata



Girone di ritorno

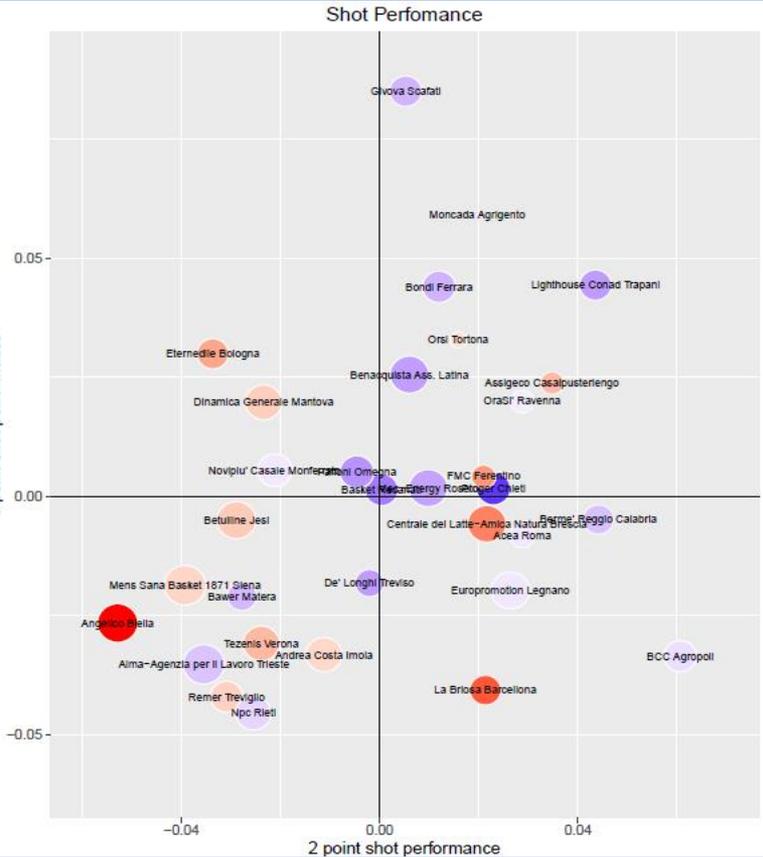


*Nonostante i dati siano completamente diversi si notano lievi differenze, ma la struttura è molto simile: si tratta di un successo perché gli alberi di classificazione sono molto "instabili"!*

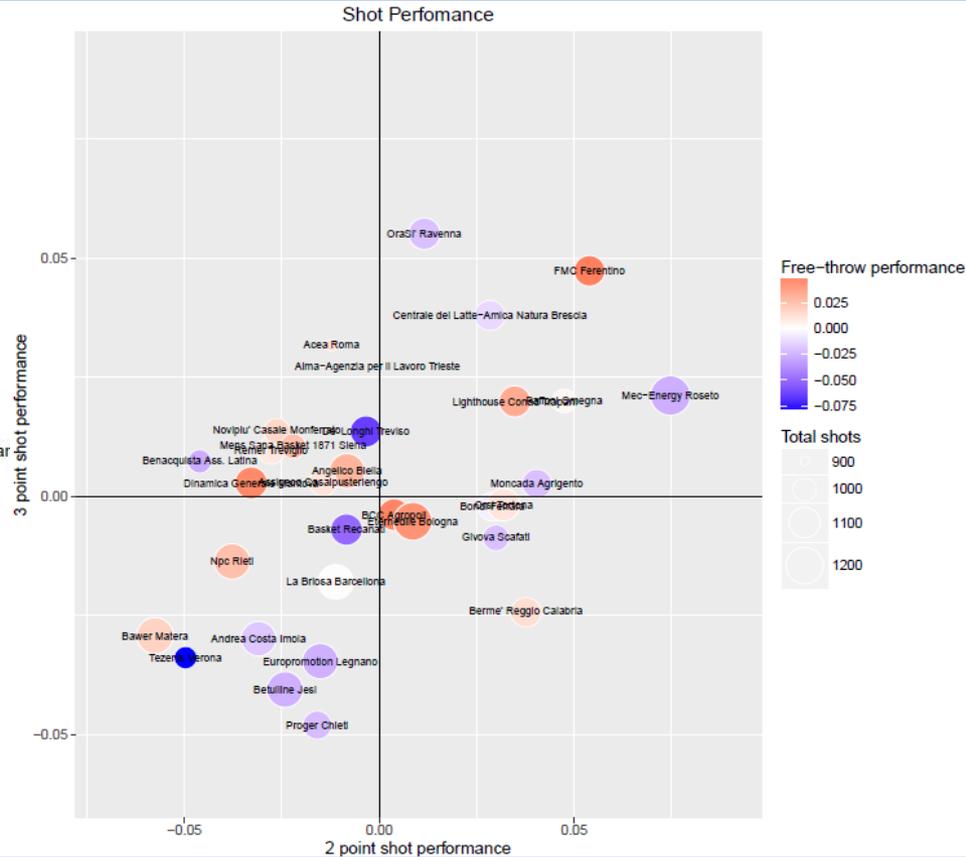
- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**

### Girone di andata

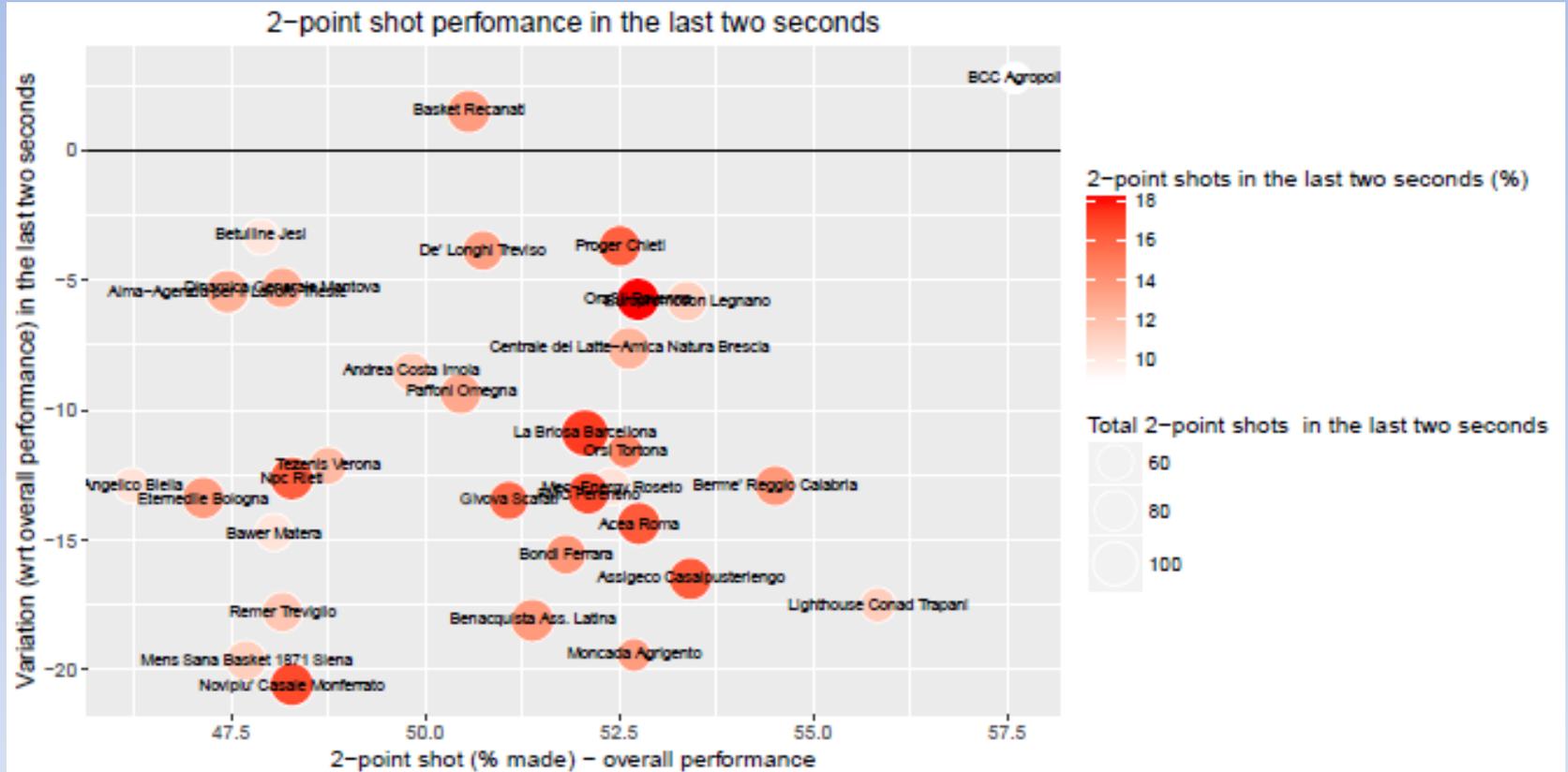


### Girone di ritorno



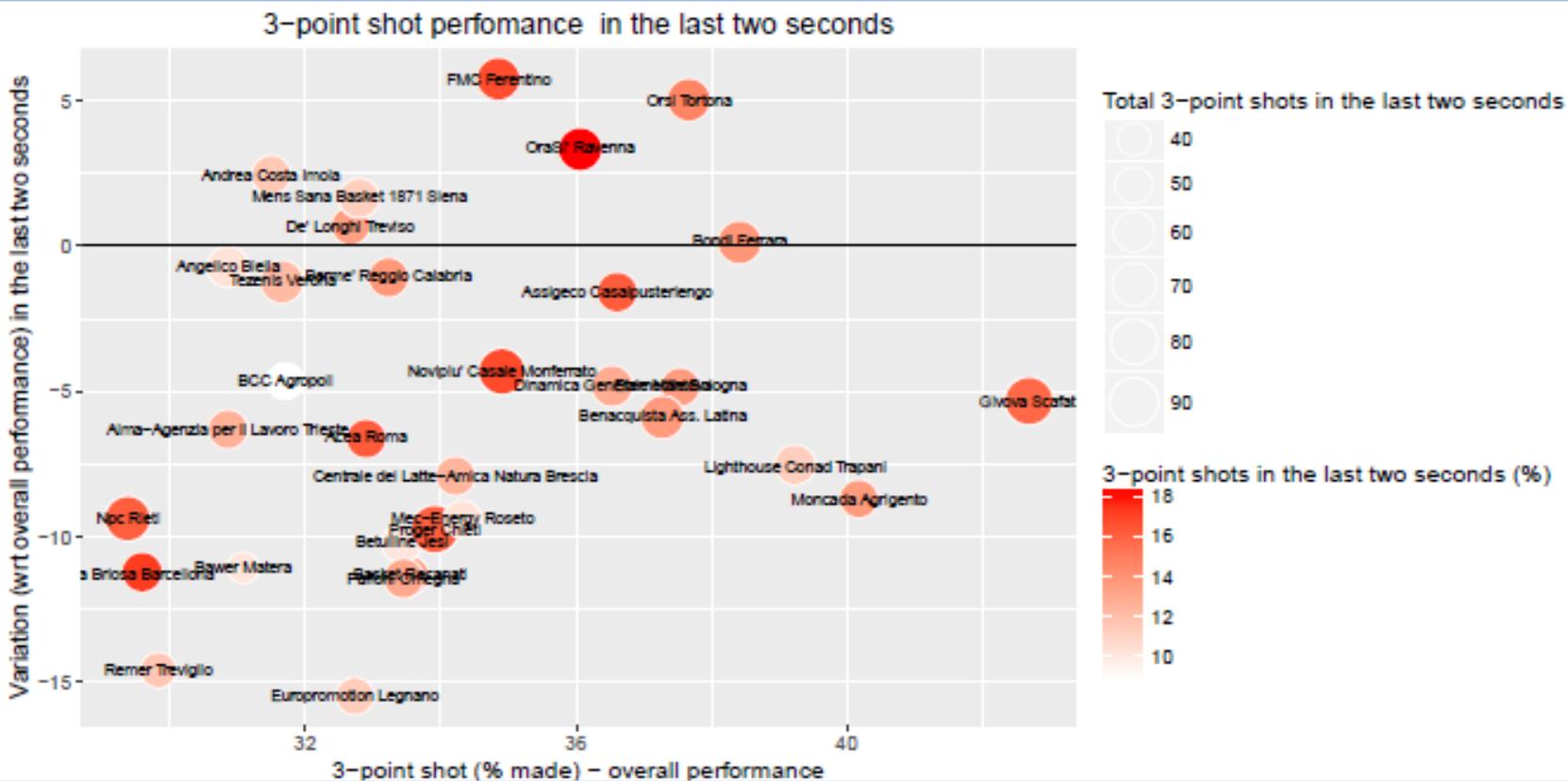
- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**



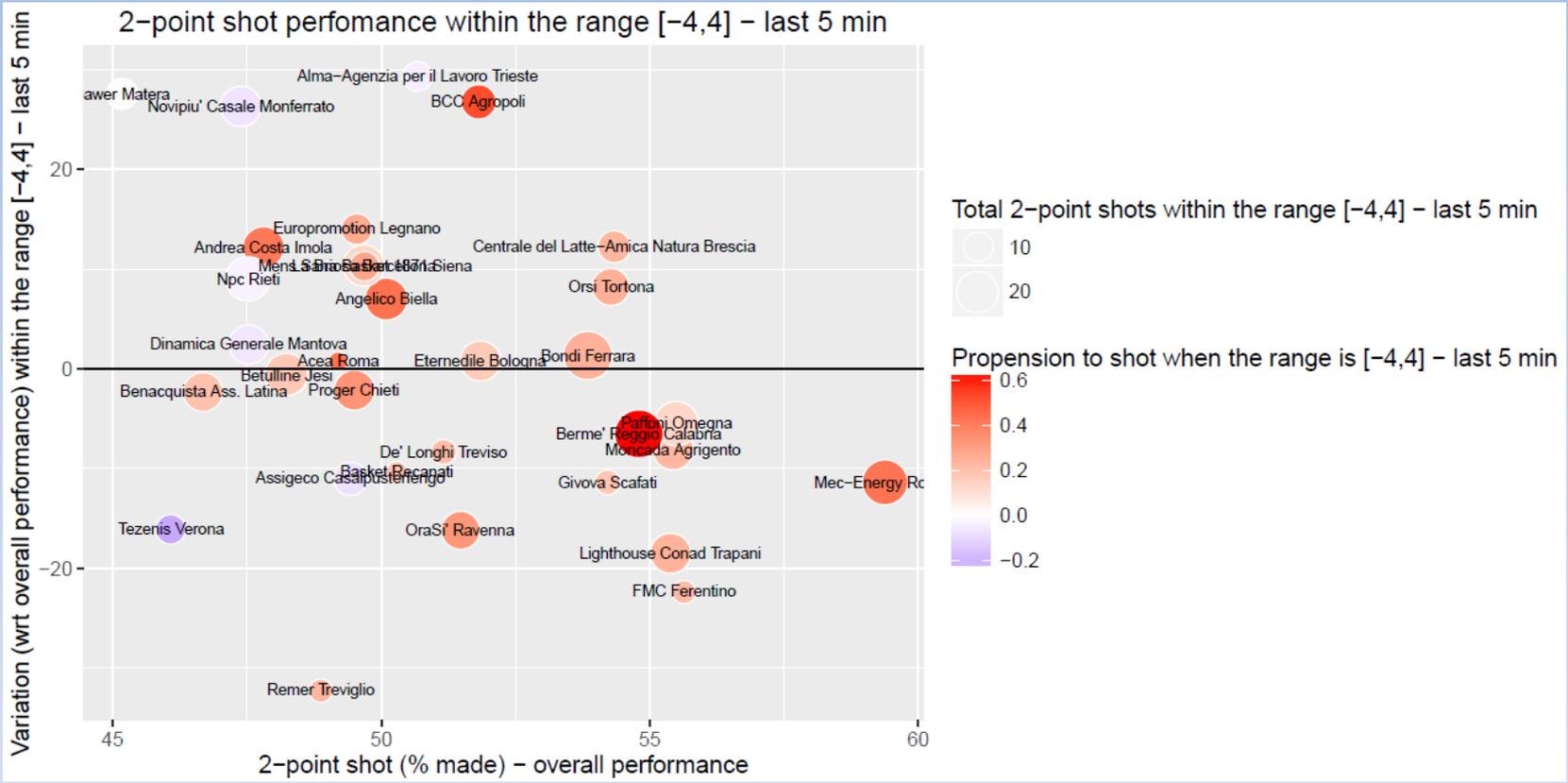
- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**



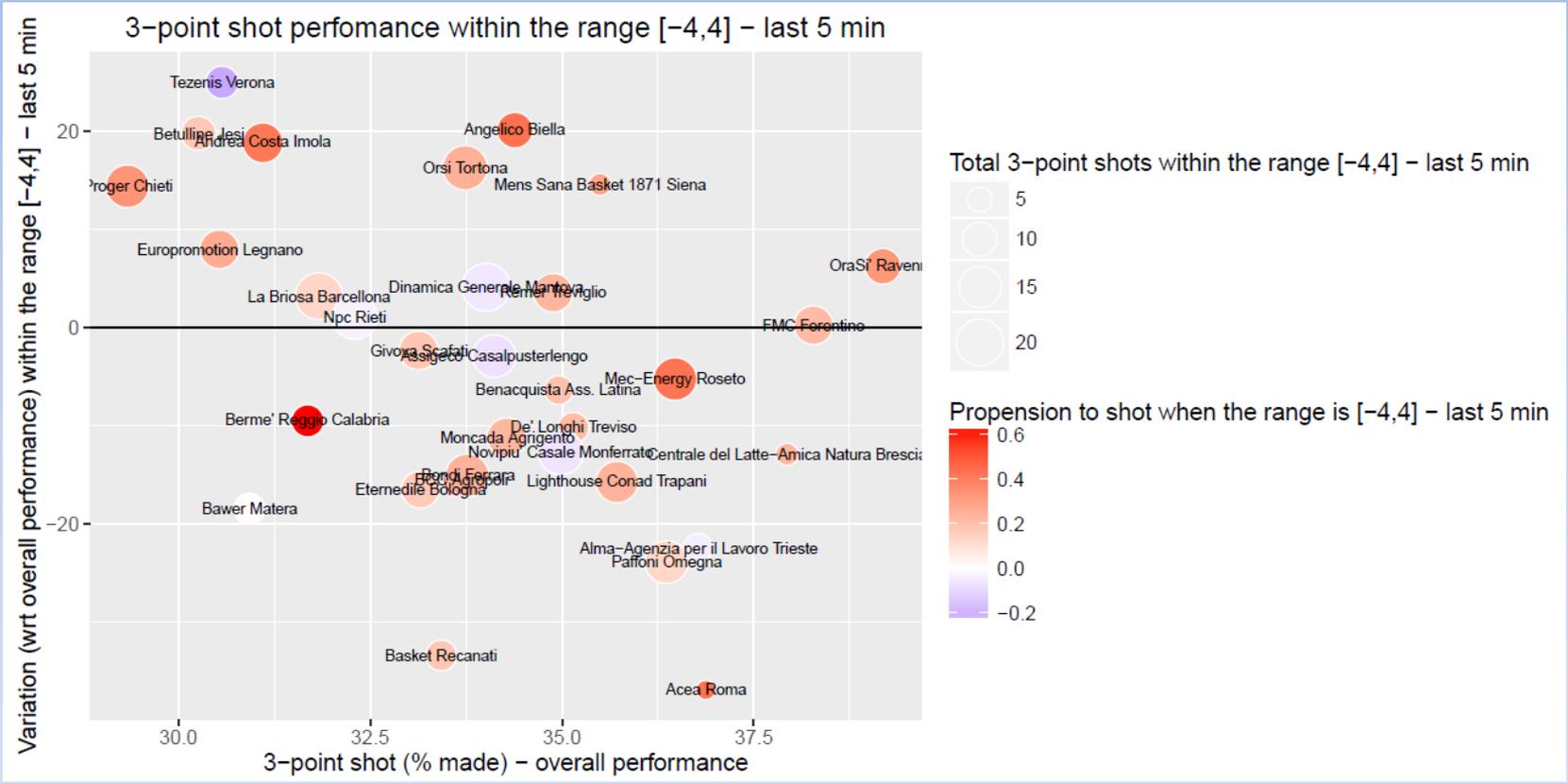
- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**



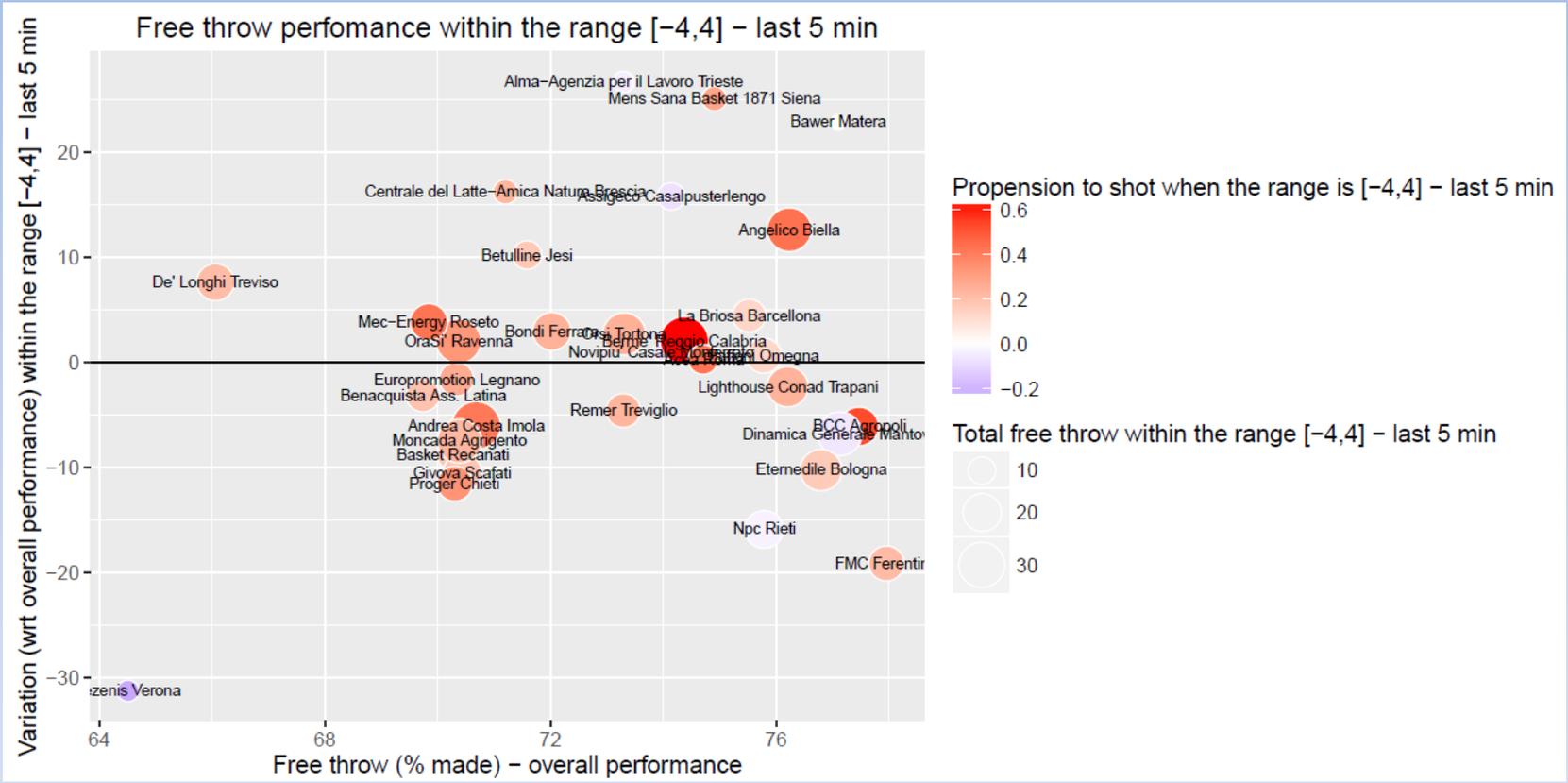
- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**



- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**



- Gli obiettivi dell'analisi
- Il dataset
- Le situazioni di stress nel basket

- Le probabilità di fare canestro: l'albero di classificazione
- **Nuovi indici di tiro**

## Conclusioni:

- i nuovi indici e le relative rappresentazioni grafiche elaborati consentono di quantificare il miglioramento o peggioramento della performance di tiro nelle specifiche situazioni (in particolare, la situazione di Verona);
- le variazioni delle performance di tiro spiegano alcune variazioni di posizione in classifica, ma non tutte → è solo una parte di un gioco molto complesso, altri aspetti (ad esempio: la difesa) sono altrettanto importanti;
- approfondire tali aspetti e integrare le statistiche di base con analisi di questo genere, basate su big data e tecniche di machine learning, potrebbe aprire nuovi orizzonti per l'analisi dei dati nel basket!

# Grazie per l'attenzione!

[bodai.unibs.it/BDSports](http://bodai.unibs.it/BDSports)