



Basketball Analytics: the use of Data Science to describe and predict the performance of an NBA team

Alberto Ferrario

Relatore: **Prof. M. Pelagatti**

Correlatore: **Prof.ssa P. Zuccolotto**

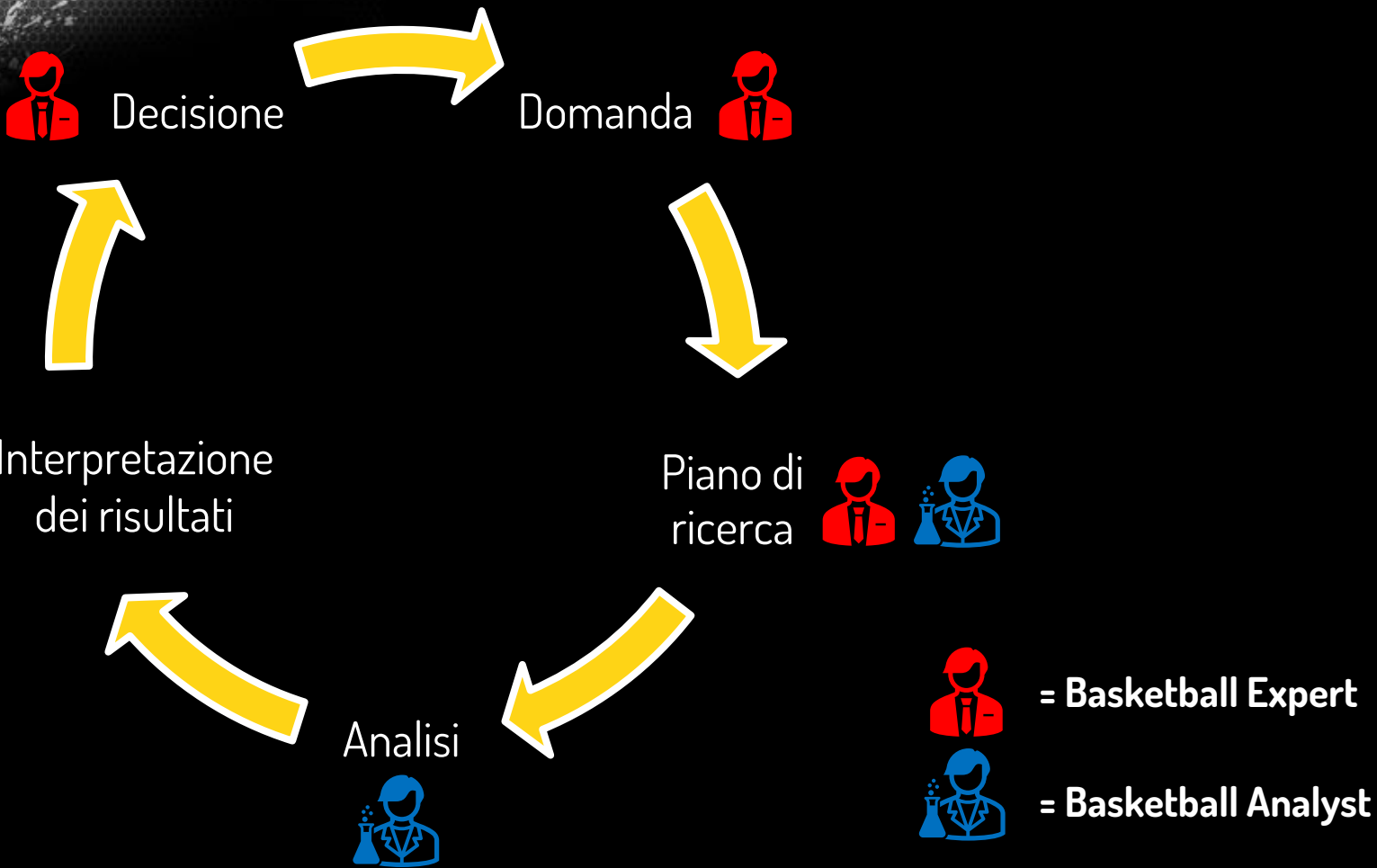


National Basketball Association (NBA)

- Nasce nel 1949 dalla fusione tra NBL e BAA
- David Stern: NBA commissioner 1984-2014
- 1° lega professionistica di basket al mondo, trasmessa in 215 Stati e 49 lingue
- 3° lega sportiva al mondo per volume di entrate (8.76 miliardi di dollari) alla fine della stagione 2018-19
- I giocatori dell'NBA sono i più pagati al mondo: stipendio medio di 8.32 milioni di dollari l'anno nella stagione 2019-20



Basketball Analytics





Dataset

- 2 dataset: uno contenente le statistiche riguardanti 7379 partite di regular season disputate dal 2012 al 2018, l'altro contenente informazioni relative al posizionamento in classifica delle 30 squadre per ogni giornata in cui è stata giocata almeno una partita
- Pulizia dati e missing imputation
- Feature engineering – ricalcolo e definizione nuove metriche
- 7379 righe per 208 colonne, no NA



Elo Rating

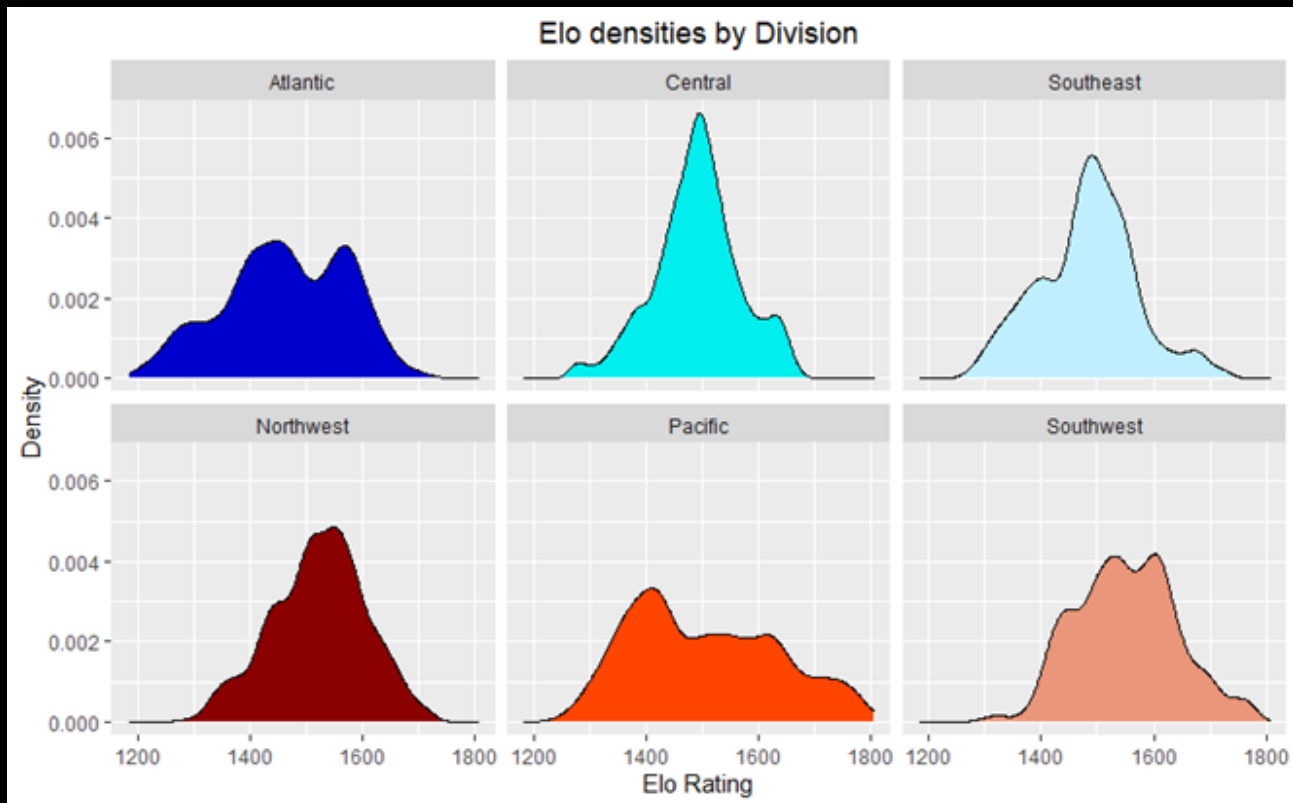
Sistema per valutare le performance delle squadre di NBA.

Caratteristiche principali:

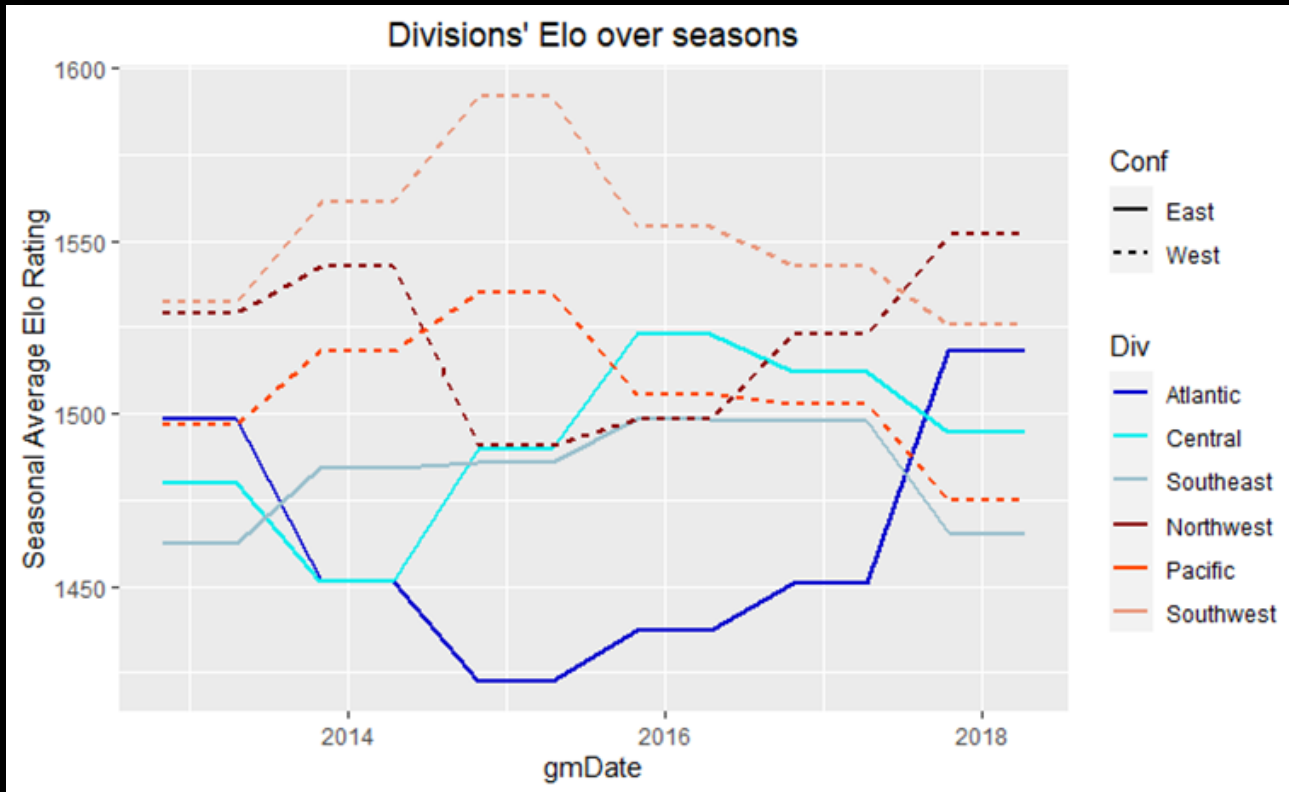
- Tutte le squadre partono da 1500 *punti Elo*
- Il punteggio si aggiorna dopo ogni partita e dipende esclusivamente da:
 - Risultato finale della partita
 - *Elo Rating* delle squadre in campo
 - Dove è stata giocata la partita
- È un sistema a somma zero: dopo la fine di una partita, la squadra vincente guadagna un certo numero di *punti Elo*, la squadra sconfitta perde lo stesso numero di *punti Elo*

Scopo: tracciare l'evoluzione delle performance di una squadra nel corso delle stagioni

Elo Rating



Elo Rating





Value of Ball Possession Differential

$$VBPdiff = ORtg - VBP$$

con:

$$ORtg = \frac{PTS}{POSS} \cdot 100$$

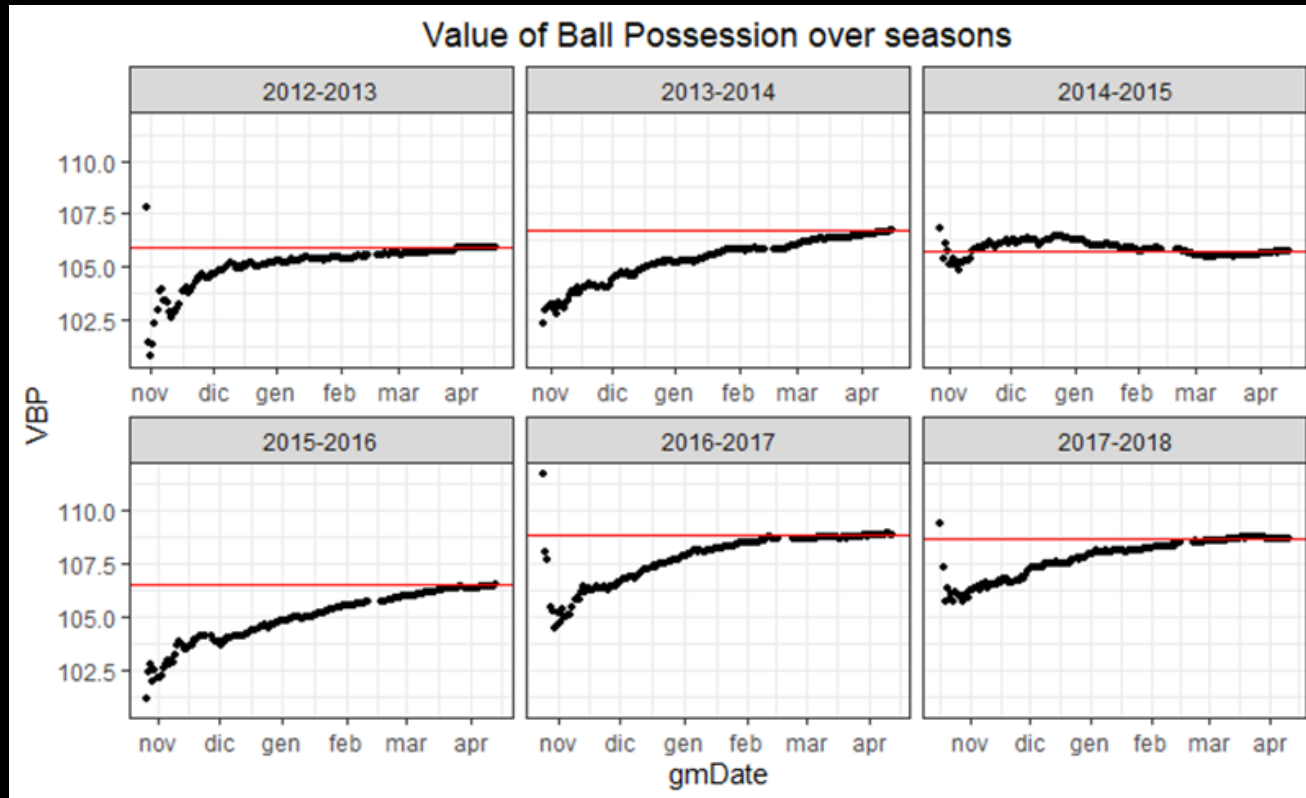
e

VBP = ORtg medio di lega calcolato su base giornaliera cumulata

Scopo: valutare se la performance offensiva di una squadra in una determinata partita è superiore alla performance media di lega fino a quel punto della stagione

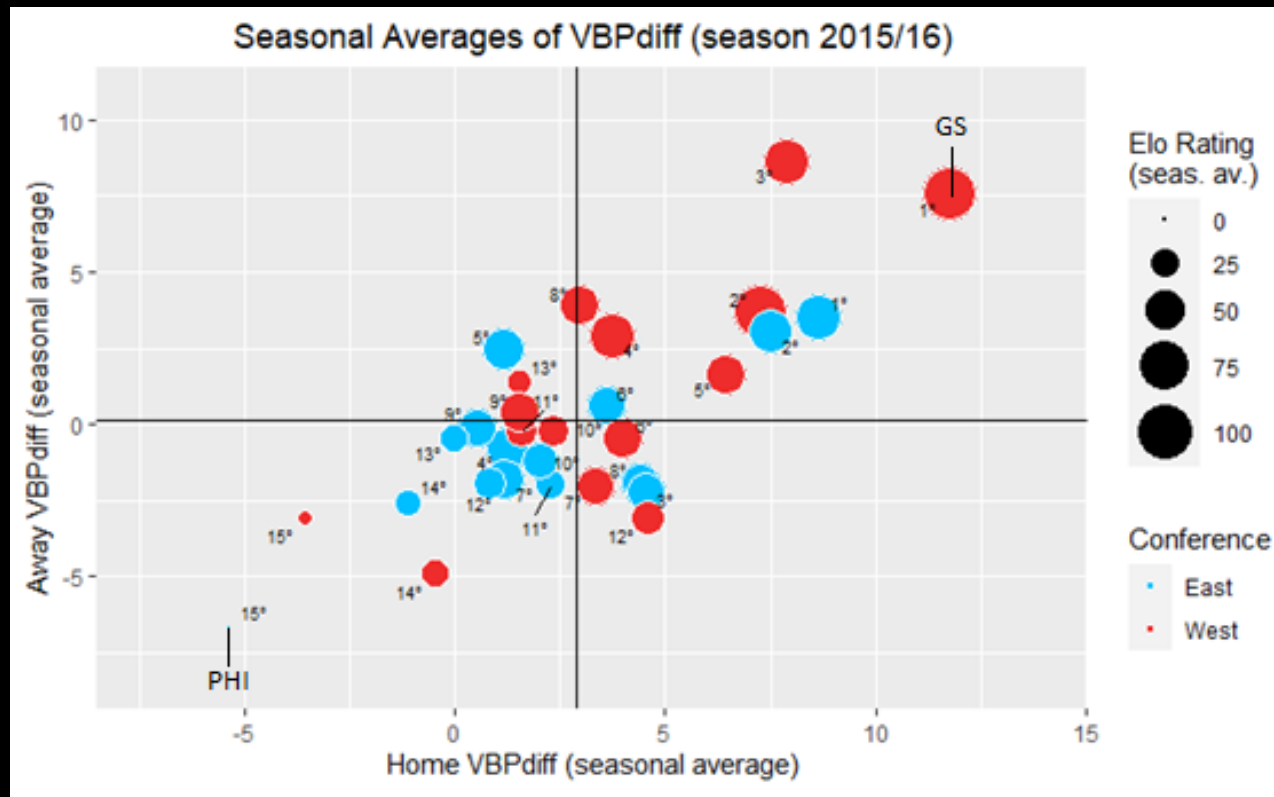


Value of Ball Possession





Value of Ball Possession Differential



A silhouette of a basketball player in a white jersey, dribbling a basketball, set against a background of orange and black hexagonal patterns.

Previsione

Preparazione dei dati in 2 fasi:

- a. Ottenimento di dati utilizzabili
 - Lag(1)
 - Performance recenti – SMA(7) e EWA
- b. Eliminazione di:
 - Prime partite della stagione
 - Variabili *near-zero variance*
 - Combinazioni lineari
 - Variabili altamente correlate (>0.85)

7276 righe per 88 colonne (85% training – 15% test)

Previsione



Training di più di 50 modelli – metodo di ricampionamento: *10-fold cv ripetuta 3 volte*

Il modello finale è un ensemble di 3 modelli base: *Penalized Discriminant Analysis, Boosted Logistic Regression e C5.0*

Simple Average Ensemble Confusion Matrix

		Actual	
		loss	win
Predicted	loss	271	153
	win	181	486

Accuracy = 0.6939



Grazie per l'attenzione

