

# Data mining u sportu

# Sadržaj

- 1. Uvod I izvorišne ideje
- 2. Potrebe
- 3. Košarka
- 4. Nogomet
- 5. Američki nogomet
- 6. Nove mogućnosti
- 7. Zaključak
- 8. Literatura

# 1. Uvod I izvorišne ideje

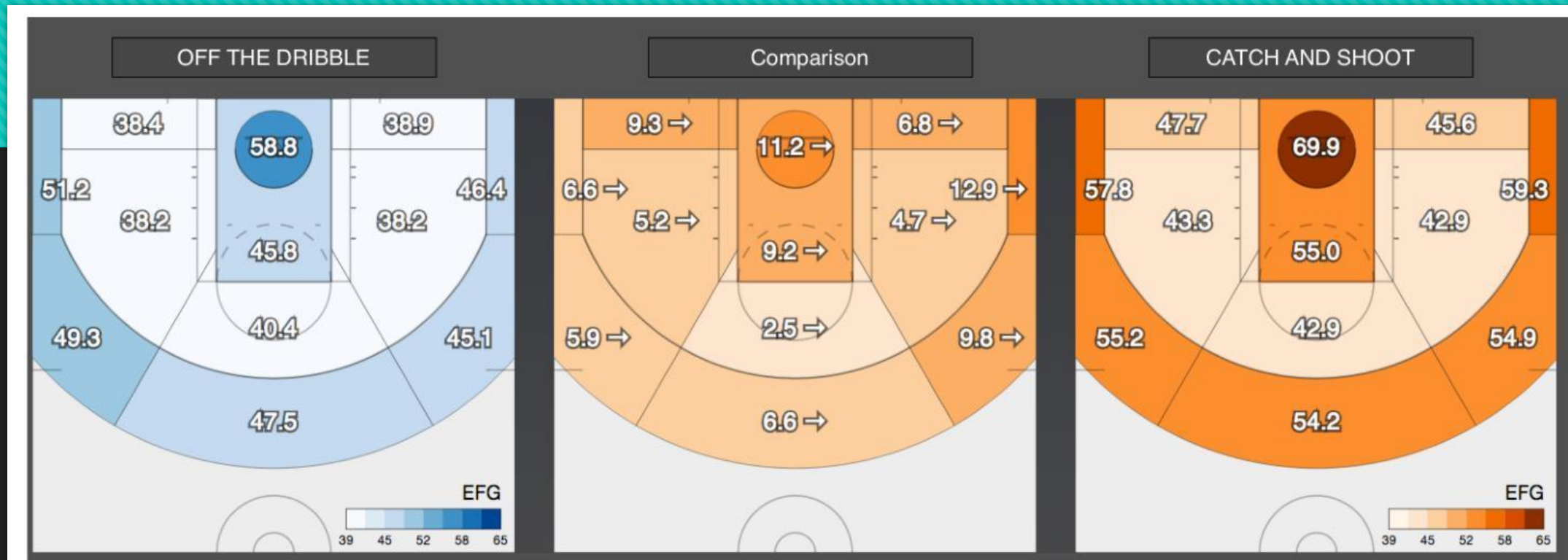
- Sport je postao vrlo profitabilan biznis
  - Rizik sportskih klubova je veći nego ikada
  - Kladionice rade s ogromnim iznosima novca
- Data mining – otkrivanje novog znanja u podacima
  - Prediktivni
  - Deskriptivni
- Važnost statistike u baseballu – sport s turbulentnom prošlošću
  - -> SABR (Society for American Baseball Research)
  - Billy Bean uveo Moneyball, Theo Epstein postigao velike uspjehe

## 2. Potrebe

- Deskriptivne
  - Raspodjela resursa – sportski i financijski
  - Analiza kvalitete i stila
  - Optimiranje strategije
  - Analiza protivnika
  - Procjena slučajnosti
- Prediktivne
  - Predviđanje rezultata
  - Kvaliteta igrača kroz godine + rizik ozljeda
  - Simulacija taktike

# 3. Košarka

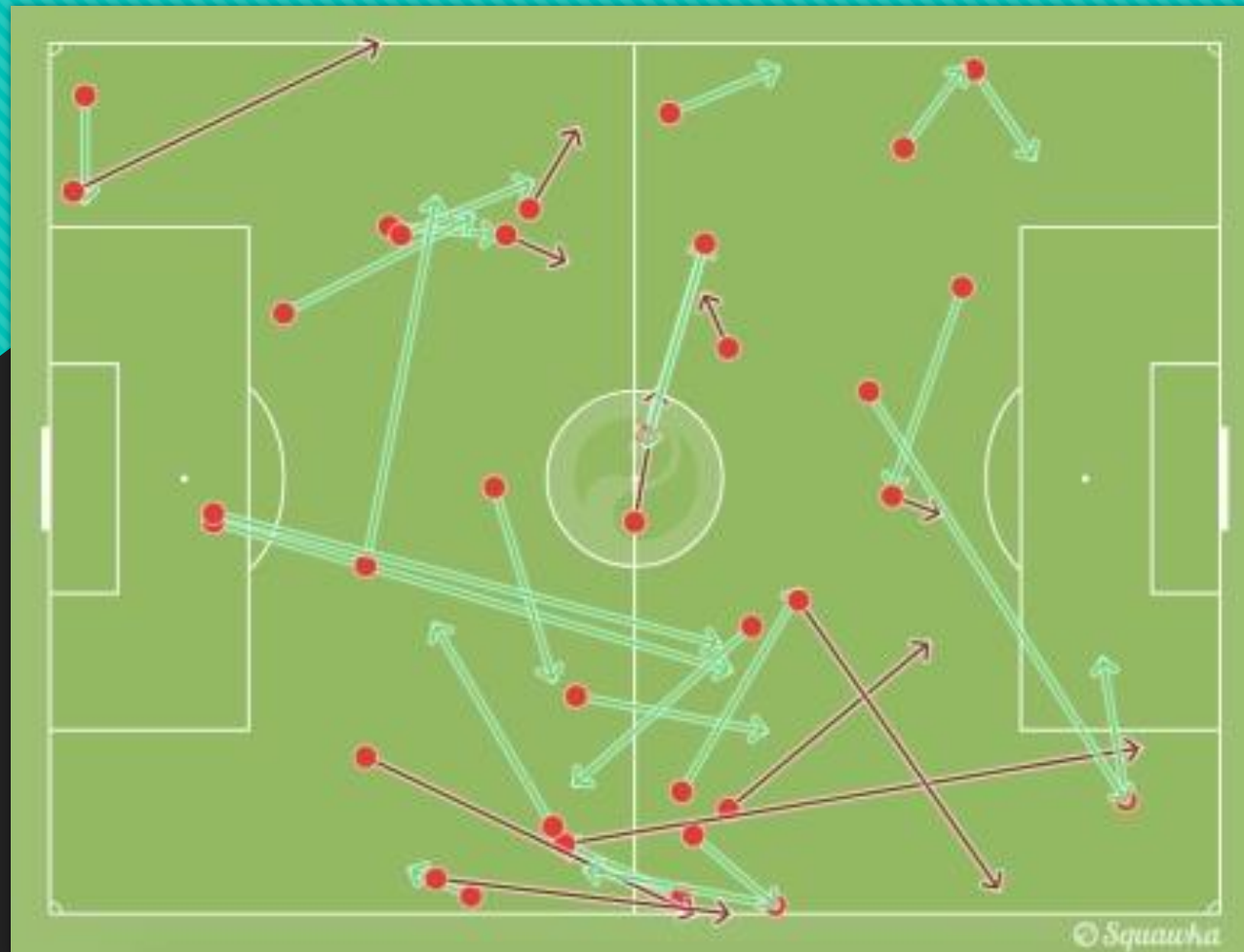
- Analiza kvalitete šuta
  - Efikasnost igrača (EFG, TS%)
  - Efikasnost zona na terenu
    - Trica iz kornera s linije bolja nego sa sredine
    - Duge dvice užasne - slična preciznost kao trice, a nose samo 2 poena
      - $3 * 0.35 = 2 * x \Rightarrow x = 52.5\%$  za prosječnog tricaša za efikasan šut duge dvice
- Analiza na draftu
  - Velika važnost svakog izbora
  - Stotine modele različitih entuzijasta



Slika 1: Prikaz kvalitete šuta iz rada Maheswaran, et al. 2014.

# 4. Nogomet

- Globalni sport -> velik broj analitičara i pristupa
- Taktički (klasični) pristup uz statistiku, npr. [Spielverlagerung.de](https://www.spielverlagerung.de)
- Praćenje igrača tijekom utakmica
  - Dodavanja, trčanje, oduzete lopte
  - Squaka, Opta, ... - velik broj opcija, čak neki HNL klubovi koriste te metode
- Korištenje Football Manager baze podataka
  - Stotine novinara rade za lokalne igrače -> institucijsko znanje koje stranci nemaju
  - Podaci zapisani numerički -> vrlo lijep input u metode strojnog/dubokog učenja

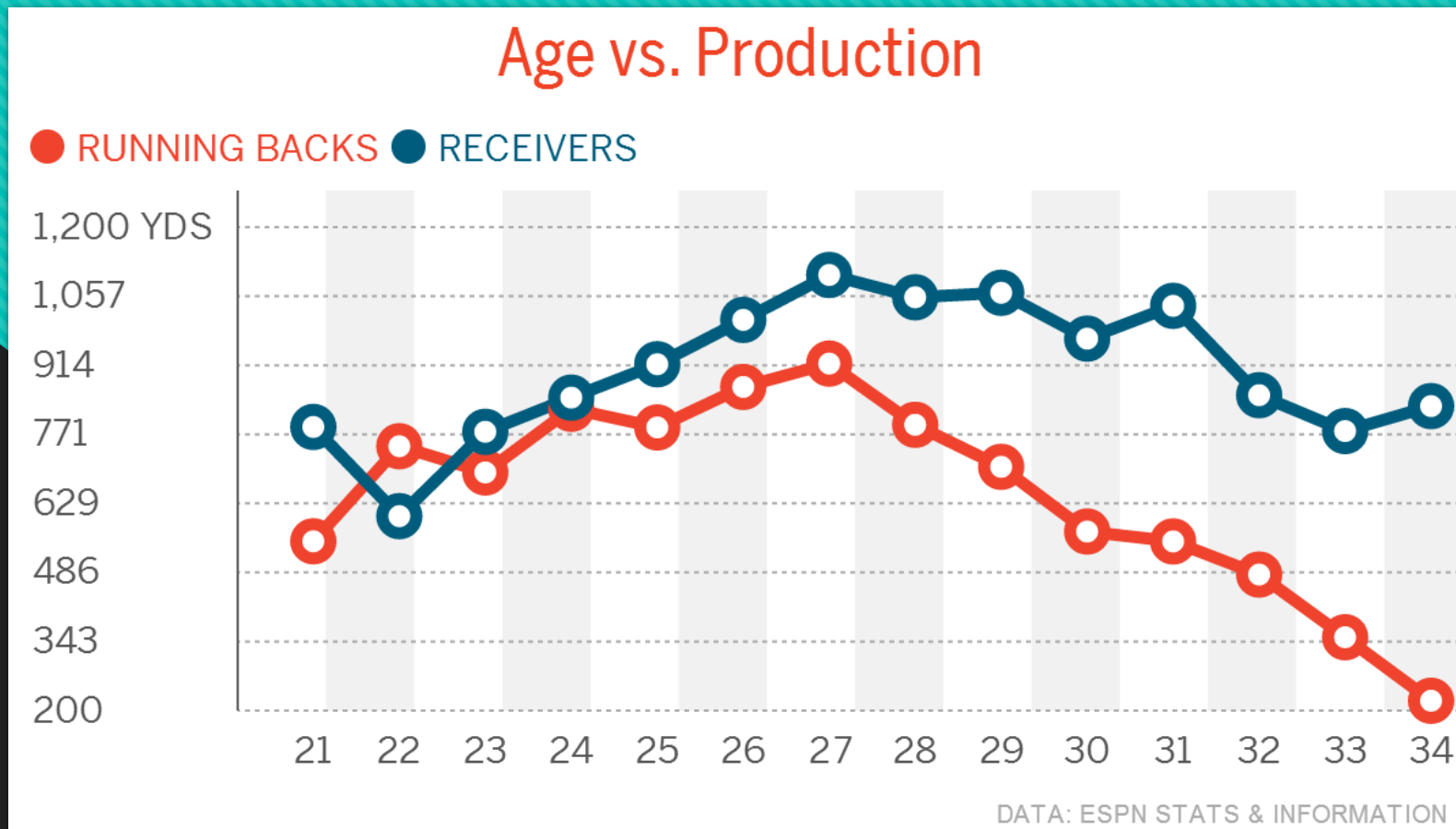


Slika 2: Prikaz u sustavu Squacka Footballa



# 5. Američki nogomet

- Određena razina skrivenosti podataka od javnosti
- Mogu se pratiti očita poboljšanja sustavnih grešaka
  - Držanje osrednjih veterana -> rizik s mladim igračima
  - Visok odabir trkača na draftu
  - Favoriziranje trčanja
- Novi principi
  - Shvaćanje vrijednosti ofenzivne linije
  - Izbacivanje dugih dodavanja



Slika 3: Broj osvojenih jardi trkača i hvatača prema životnoj dobi, ESPN.com

# 6. Nove mogućnosti

- Simulacija utakmica u cilju pronalaženja novih podataka ili vjerojatnosti
  - Monte-Carlo metoda
  - Igrice, web stranice nisu baš točne
- *Text mining* izvještaja skauta
  - 2017. znanstveni rad za NHL
  - Problem mijenjanja ljudi kroz vrijeme

# 7. Zaključak

- Različiti sportovi, različiti pristupi
- Postoje neistražene opcije
- Sigurno postoji potreba s obzirom na vrijednost klubova i novce kladionica
- Ispunjeni ciljevi:
  - Upoznavanje s data miningom u sportu
  - (Vjerojatno) pronađena tema diplomskog rada

# Pitanja



# Literatura

- Jeff Su Sheldon Kwok Tal Levy Adam Wexler Kevin Squire Yu-Han Chang, Rajiv Maheswaran. Quantifying shot quality in the NBA, 2014. [Preuzeto: 11.05.2017.]
- ESPN staff. Clubs to use football manager data, 2014. [Preuzeto: 11.05.2017.]
- Mikeie Reiland. Why aren't soccer analytics a bigger deal?, 2017. [Preuzeto: 11.05.2017.]
- Mike Rovito Timo Seppa, Michael E. Schuckers. Text mining of scouting reports as a novel data source for improving NHL draft analytics, 2017. [Preuzeto: 11.05.2017.]
- te drugi radovi navedeni u tekstu seminara

Hvala na pažnji!



THANKS

A 3D rendered graphic of the word "THANKS" in a playful, blocky font. The letters are colored as follows: 'T' is purple, 'H' is yellow-green, 'A' is red, 'N' is light purple, 'K' is red, and 'S' is green. The letters are arranged in a slightly staggered, horizontal line. The background is a dark blue gradient with a subtle green glow behind the letters.