



Aalto-yliopisto
Perustieteiden
korkeakoulu

On cluster structures of NHL players (aihe-esittely)

Elmeri Lähevirta

15.01.2018

Ohjaaja: TkT *Lauri Viitasaari*

Valvoja: Prof. *Pauliina Ilmonen*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Tausta

- NHL-pelaajien tilastoja on aloitettu keräämään tarkemmin ja tilastanalytiikkaan on alettu panostaa enemmän 2010-luvulla.
 - NHL lisäsi vuonna 2015 omille sivuilleen kaikkien saataville 45 kehittyneempää ja monipuolisempaa joukkue- ja pelaajatilastoa.
- Elokuva Moneyball (2011) mainittu monessa lähteessä urheilun tilastanalytiikan vallankumouksen käännekohtana.

Tavoitteet

- Tehdä klusterointianalyysi k-means-algoritmilla NHL:n runkosarjan 2016-17 pelaajatilastoista.
 - Tutkia, miten hyvin k-means-algoritmi pystyy erottamaan puolustajat hyökkääjistä ja pystyykö se mahdollisesti erottamaan jopa eri pelaajatyyppejä toisistaan.

Rajaukset

- Maalivahdit jätetään todennäköisesti klusterointianalyysin ulkopuolelle ja analyysissä keskitytään vain kenttäpelaajiin (hyökkääjät ja puolustajat).

Menetelmät

- K-means-algoritmi
 - K-means++-alustus

Työvaiheet

1. NHL:n runkosarjan 2016-17 pelaajatilastoiden hankinta ja niihin tutustuminen.
2. Datan rajaaminen ja muokkaaminen ja sopivien muuttujien valinta.
3. Klusterointimenetelmien teoriaan perehtyminen.
4. Klusterointianalyysin toteuttaminen ja tulosten tulkinta.

Työkalut

- RStudio
 - Tilastolliseen analyysiin, kuvaajien piirtämiseen ja klusterointianalyysin toteuttamiseen.

Tietolähteet ja aineistot

Tietolähteet

- Maimon, Oded Z., and Lior Rokach, *Data Mining and Knowledge Discovery Handbook. 2nd ed.*, New York; London: Springer, 2010.
- Han J, Kamber M, Pei J., *Data mining (third edition)*. Boston: Morgan Kaufmann, (2012).

Aineistot

- NHL:n runkosarjan 2016-2017 pelaajakohtaiset tilastot

Aikataulu

- Aihehakemus ja työnaloitus 4.12.2017
- Aineiston hankinta ja siihen tutustuminen 12/2017
- Teoriaan tutustuminen 12/2017
- Aiheen esittely 15.01.2018
- Klusterointianalyysin tekeminen 01/2018
- Tulosten pohdinta ja työn viimeistely 01/2018
- Valmiin työn esittely 20.02.2018