

Tero Karjalainen

Random forest

Joutui armas aika ja suvi suloinen. Pitkä odotus, paljon työtä, onnistumisia, epäonnistumisia, ehkä verta, hikeä ja kyyneleitäkin on takana, mutta nyt se on vihdoinkin täällä – jalkapallon MM-kisat. Joillekin meistä hyvinkin uskonnolliseen kokemukseen rinnastettava tapahtuma, jonka aiheuttama tunneskaala vetää hyvinkin vertoja koulujen päättäjäisjuhlissa kokemiimme. Jalkapallon MM-kisat ovat myös hyvin paljon tiedettä, tai ainakin tutkittua tietoa ja siihen perustuvia toimenpiteitä. Jokaisessa ottelussa jokaisen pelaajan jokainen suoritus tilastoidaan ja analysoidaan. Tulevan ottelun pelitaparatkaisut ja pelaajavalinnat perustuvat ainakin osittain niin omien pelaajien tilastoihin kuin myös vastustajan tekemisiin aiemmissa otteluissa. Valmennus- ja tilastointititimit joukkueilla ovatkin kooltaan jotain aivan muuta kuin kotoisten alasarjojen penkkien takana on nähtävissä.

Myös tekoälytutkijat ovat olleet aktiivisia MM-kisojen suhteen. Tieteentekijät Dortmundista, Ghentistä ja Münchenistä kasasivat ison data-aineiston sisältäen kisoihin osallistuvien maiden kansantaloudellista dataa, vedonlyöntitoimiston kertoimia, FIFA-sijoitusta, maanosa ja maanosaliittotiedon, tietoja joukkueiden koostumuksista – kuinka monta seurajoukkuekaveria joukkueessa on, keski-ikä, kuinka monta pelaajaa pelasi Mestareiden liigassa jne. Tiedot kerättiin edellisistä neljistä MM-kisoista sekä vuodelta 2018. Tutkijat käyttivät kolmea eri simulaatiomenetelmää ennustaakseen jokaisen lopputurnausottelun lopputuloksen ja sen jälkeen parasta

mallia hyödyntäen ennustivat todennäköisyydet turnausvoittajaksi. Kunkin mallin kyky ennustaa testattiin neljien edellisten lopputurnausten datalla ja parhaaksi ennustemalliksi valikoitui tilastollisen mallintamisen ja koneoppimisen sekamalli ”random forest”.

Toivon kaikille lukijoille voimia antavaa kesäaikaa. Latautuu akkusi parhaiten sitten riippumatossa MM-kisoja katsellen, merellä purjehtien, marjametsässä liikkuen, perheen kanssa matkailien tai tutkimustyöstä ilman keskeyttäviä häiriöitä nauttien, niin toivottavasti kesäsi sisältää riittävästi aikaa akkujesi latautumiseen.

Ps. Heinäkuun puolivälissä sitten selviää voitaako Saksa finaalissa Brasilian, kuten tekoäly ennustaa, vai viekö Brasilia mestaruuden kuten ihmisasiantuntijat laajalti ennustavat vai juhliiko päätoimittaja perheveikkauksen voittajana (jälleen) ja Messi vihdoinkin onnistuu osoittamaan pelaajasuuruutensa myös maajoukkueessa.

Iwish all our readers a summer time with enough time for activities that boost your energy levels. Whether it means sailing, berry picking, travelling with family, doing research without unnecessary interruptions or watching FIFA World Cup lying in the hammock. Artificial Intelligence (random forests model) predicts that Germany will beat Brazil in the final. Human experts hold Brazil as a favorite and the editor in chief is a happy man if Messi finally shows his greatness as a football player also in the National team.