

Alexi Viitaharju

**DATA-ANALYTIIKAN HYÖDYNTÄMINEN
AMERIKKALAISEN JALKAPALLON PE-
LIN SISÄISESSÄ VALMENNUKSESSA**
Kandidaatintyö

Kandidaatintyö
Tietojohtaminen
Tarkastaja: Jussi Myllärniemi
4/2023

TIIVISTELMÄ

Alexi Viitaharju: Data-analytiikan hyödyntäminen amerikkalaisen jalkapallon pelin sisäisessä valmennuksessa

Utilization of data analytics in American football's in-game coaching

Kandidaatintyö

Tampereen yliopisto

Tietojohtaminen

Huhtikuu 2023

Data-analytiikan tuomat mahdollisuudet ovat lisääntyneet teknologian kehityksen ja big datan yleistymisen seurauksena. Myös urheilussa on mahdollista hyödyntää data-analytiikkaa ja saavuttaa sen avulla kilpailullista etua. Tässä työssä tutkitaan data-analytiikan nykytilaa ja sen tuomia mahdollisuuksia amerikkalaisen jalkapallon pelin sisäiseen päätöksentekoon. Tätä aihetta tutkitaan alatutkimuskysymyksien avulla, jotka käsittelevät data-analytiikan hyödyntämisen tapoja amerikkalaisessa jalkapallossa sekä näiden tapojen käytön tuomaa kilpailullista etua. Tämän lisäksi alatutkimuskysymyksillä tutkitaan saatavilla olevaa dataa amerikkalaisessa jalkapallossa sekä data-analytiikan käytön ongelmia ja haasteita.

Tutkimus on toteutettu kirjallisuuskatsauksena käyttäen hyväksi kirjallisuustutkimuksen processimallia. Tutkimusaineisto käsittelee pääasiassa data-analytiikan käyttöä amerikkalaisen jalkapallossa, jonka lisäksi aineisto käsittelee yleisellä tasolla urheiluanalytiikkaa, big dataa ja dataohjautuvaa päätöksentekoa. Aineistoa data-analytiikan käytöstä amerikkalaisessa jalkapallossa on hyvin, mutta tämänkaltaista työtä, jossa olisi tutkittu data-analytiikan hyödyntämistä pelin sisäisissä päätöksentekotilanteissa laajemmin, ei ole vielä tehty, mikä loi tarpeen tämän tutkimuksen tekemiselle.

Tutkimuksen tuloksena saatiin hyvä kokonaiskuva data-analytiikan hyödyntämisestä amerikkalaisen jalkapallon pelin sisäisissä päätöksentekotilanteissa sekä sen tuomista mahdollisuuksista. Hyödyntämällä data-analytiikkaa ennustamaan hyökkäysvuorossa olevan vastustajan taktisia valintoja sekä tehdessä hyökkäysvuorossa valintoja neljänsillä yrityksillä ja touchdownin jälkeisillä lisäpisteyrityksillä voidaan saavuttaa merkittävääkin kilpailullista etua. Siitä huolimatta data-analytiikan käyttö ei ole vielä kovin suuressa suosiossa. Tähän syynä on erilaisia organisaatiollisia sekä dataan liittyviä teknisiä haasteita. Monissa joukkueissa omistajaportaalla on kielteinen näkemys data-analytiikan käyttöä kohtaan, mikä saa valmentajat tekemään voiton todennäköisyyden kannalta heikkoja ja data-analytiikan vastaisia ratkaisuja etenkin, jos valmentaja tuntee työpaikkansa olevan vaarassa.

Data-analytiikan vähäistä hyödyntämistä selittää myös Ellsbergin paradoksi, jonka mukaan ihmiset useimmiten valitsevat mieluiten vaihtoehdon, jossa tiedetään todennäköisyys mieluummin kuin vaihtoehdon, jonka todennäköisyyttä ei tiedetä, vaikka se olisi odotusarvoltaan suurempi. Tämä siis selittää data-analytiikan ratkaisuja konservatiivisempien ratkaisujen suosimista. Pelien vähäinen määrä ja sen seurauksena vähäisen datan määrän katsotaan myös rajoittavan data-analytiikan käyttöä. Tähän liittyy myös peliin vaikuttavien muuttujien suuri määrä. Tällöin siis monet katsovat, että samat muuttujat sisältävien tilanteiden määrä ei ole tarpeeksi suuri johtopäätöksien tekemiseksi.

Tutkimuksen pohjalta voidaan siis todeta, että dataa tulee kerätä lisää, jotta malleista tulee entistä luotettavampia. Kuitenkin nykyiselläänkin niillä voidaan saavuttaa merkittävää kilpailullista etua, mutta data-analytiikan hyödyntämistä haittaavat erilaiset organisaatiolliset haasteet.

Avainsanat: Data-analytiikka, Urheiluanalytiikka, Dataohjautuva päätöksenteko, Big data, Amerikkalainen Jalkapallo, NFL

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

ALKUSANAT

Tämän kandidaatintyön aihe valittiin amerikkalaisen jalkapallon data-analytiikan käytön ajankohtaisuuden ja henkilökohtaisen mielenkiintoni vuoksi. Data-analytiikan hyödyntämistä voidaan tässä kontekstissa pitää hyödyllisenä, mutta sen käyttö on siitä huolimatta suhteellisen vähäistä.

Haluan kiittää työn ohjaajaa Jussi Myllärniemeä hyvästä ja rakentavasta palautteesta sekä kannustuksesta työn edetessä. Kiitos myös oponojilleni sekä muulle kandidaatin työn ryhmäläisille loistavasta palautteesta ja tuesta. Lopuksi vielä kiitän perhettäni ja ystäviäni saamastani tuesta työn tekemisen varrella.

Tampereella, 3.5.2023

Aleksi Viitaharju

SISÄLLYSLUETTELO

1. JOHDANTO	1
1.1 Tutkimuksen rajaus	2
1.2 Tutkimusongelma ja tutkimuksen tavoitteet	2
1.3 Työn rakenne	3
2. TUTKIMUSMENETELMÄ JA -AINEISTO	5
2.1 Tutkimusmenetelmä	5
2.2 Tutkimusaineisto	5
3. DATAOHJAUTUVA PÄÄTÖKSENTEKO	8
4. URHEILUANALYTIIKAN KÄYTTÖ AMERIKKALAISEN JALKAPALLON VALMENNUKSESSA	10
4.1 Urheiluanalytiikka	10
4.2 Amerikkalainen jalkapallo	10
4.3 Saatavilla oleva data NFL:ssä	11
4.4 Urheiluanalytiikan käyttö amerikkalaisen jalkapallon valmennuksessa	12
4.4.1 Vastustajan hyökkäyspäätösten ennustaminen	12
4.4.2 Neljäs yritys	13
4.4.3 Lisäpiste	14
4.5 Käytettyjen tapojen vaikutus kilpailullisen edun luomiseen	14
5. HAASTEET URHEILUANALYTIIKAN HYÖDYNTÄMISESSÄ AMERIKKALAISESSA JALKAPALLOSSA	16
5.1 Organisaatiolliset haasteet	16
5.2 Urheiluanalytiikan puutteet ja ongelmakohtat	17
6. YHTEENVETO	19
6.1 Tutkimuksen tulokset	19
6.2 Tutkimuksen arviointi ja jatkotutkimustarpeet	20
LÄHTEET	22

1. JOHDANTO

Kiinnostus data-analytiikkaa kohtaan urheilumaailmassa on viime vuosien aikana lisääntynyt tasaisesti relevantin datan laajan saatavuuden seurauksena (Roumani 2022). Data-analytiikan käyttö urheilussa, eli urheiluanalytiikka on siis hyvin ajankohtainen aihe. Monet ovat varmasti urheilulähetystyksiä katsoessaan huomanneet erilaisia data-analytiikkaan pohjautuvia tilastoja, kuten esimerkiksi maaliodottama ja voiton todennäköisyydet, joita on alkanut ilmestyä viime vuosien aikana katsojien saataville useissa urheilulajeissa.

Amerikkalaisen jalkapallon NFL-liiga (National Football League) on tulojen perusteella maailman suurin urheilusarja, mutta tästä huolimatta data-analytiikan hyödyntäminen laahaa siinä muita lajeja perässä (Roumani 2022). Fergusonin (2018) mukaan enemmistö amerikkalaisen jalkapallon valmentajista tekee pelin sisäisiä päätöksiä, jotka ovat ristiriidassa data-analytiikan suosittamien ratkaisujen kanssa, eli selkeästi data-analytiikan käyttö ei ole vielä suuressa suosiossa. Clarkin (2018) mukaan amerikkalaisessa jalkapallossa data-analytiikan käyttö on moniin muihin lajeihin verrattuna ollut pienemässä roolissa, mihin pidetään syynä alhaista pelien määrää ja sitä kautta liian pientä saatavilla olevaa datamäärää sekä tilanteisiin vaikuttavien muuttujien liian suurta määrää. Joukkueet ovat kuitenkin alkaneet 2010-luvun lopulla panostamaan selkeämmin analytiikan käyttöön. (Clark 2018) Aihetta voidaan siis pitää ajankohtaisena etenkin amerikkalaisessa jalkapallossa.

Kokonaisuutena amerikkalaisen jalkapallon data-analytiikan käyttöä pelin sisäisessä päätöksenteossa ei ole tutkittu, vaan pääosa tutkimuksista keskittyy johonkin tiettyyn päätöksentekotilanteeseen. Tämä tutkimus kokoaa nämä päätöksentekotilanteet sekä työssä pohditaan analytiikan käyttöä sekä sen haasteita pelin sisäisessä valmennuksessa kokonaisuutena.

Urheiluanalytiikka on hyvin monessa lähteessä terminä käytössä data-analytiikan käytölle urheilussa. Urheiluanalytiikalla tarkoitetaan prosessia, jossa urheiluun liittyvää dataa käytetään ja mallinnetaan helpottamaan päätösten tekemistä (Stolbunov 2014). Aina erona voidaan siis yleiseen data-analytiikkaan verrattuna pitää sitä, että analysoitava data on urheiluun liittyvää. Urheiluanalytiikassa käytettävä data on pitkälti big dataa (Abeza et al. 2022). Big datalla tarkoitetaan Janssenin et al. (2016) mukaan dataa, joka

on todella monipuolista sekä sitä tuotetaan ja käsitellään hyvin nopeasti. Kyvystä hyödyntää kaikkea saatavilla olevaa tietoa on tullut kriittinen kyky organisaation menestymisen kannalta. (Janssen et al. 2016)

1.1 Tutkimuksen rajaus

Tutkimuksessa käsitellään urheiluanalytiikan käyttöä valmennuksessa. Pyritään siis tarkastelemaan sitä, kuinka joukkue voisi saavuttaa urheiluanalytiikan avulla kilpailullista etua valmennuksen näkökulmasta pelistrategian ja pelin sisäisten päätösten tekemisessä. Tutkimuksessa rajataan siis tästä aiheesta pois esimerkiksi joukkueen strategiat markkinoinnissa ja taloudellisissa asioissa, jotta kokonaisuus pysyy yhtenäisenä. Myöskään pelaajiin liittyviä valintoja ei käsitellä tässä tutkimuksessa.

Tässä tutkimuksessa keskitytään amerikkalaiseen jalkapalloon, sillä siinä on paljon pelin sisäistä taktista päätöksentekoa ja nämä päätökset ovat todella tärkeitä pelin lopputuloksien kannalta. Näistä päätöksentekotilanteista löytyy myös paljon dataa, jota voi käyttää hyödyksi. Amerikkalainen jalkapallo on hyvin suosittu laji Yhdysvalloissa ja sen pääsarja NFL on maailman suurin urheiluliiga siitä saatavien tulojen perusteella selvällä erolla toisena tulevaan baseballin pääsarja MLB:iin (Roumani 2022). Tämän vuoksi tutkimuksia amerikkalaisesta jalkapallosta löytyy paljon ja niiden perusteella voidaan etsiä ratkaisuja valittuun tutkimusongelmaan ja vastauksia tutkimuskysymyksiin.

1.2 Tutkimusongelma ja tutkimuksen tavoitteet

Tutkimusongelmaksi on valittu analytiikan tuomien mahdollisuuksien vähäinen hyödyntäminen amerikkalaisen jalkapallon valmennuksessa. Kuten jo johdannon alussa mainittiin, valmentajien analytiikan käyttö on vielä suhteellisen vähäistä, varsinkin otettaessa huomioon data-analytiikan kehityksen ja sen tuoman mahdollisen lisäarvo. Työssä pyritään tutkimaan, millaisilla analytiikan käyttötavoilla voidaan luoda joukkueelle kilpailullista etua ja mitä haasteita data-analytiikan hyödyntäminen sisältää. Pyritään myös tarkastelemaan saatavilla olevaa dataa ja sen riittävyttä vastaamaan valmennuksen tietotarpeisiin.

Tutkimuksen päätutkimuskysymyksenä on

- Mikä on data-analytiikan käytön nykytila ja mahdollisuudet amerikkalaisessa jalkapallossa pelin sisäisissä päätöksissä?

Tutkimusongelmaa ja päätutkimuskysymystä ratkotaan osissa alatutkimuskysymyksiä avulla, joita ovat seuraavat:

- Millaista dataa on kerätty amerikkalaisesta jalkapallosta?
- Miten data-analytiikkaa hyödynnetään amerikkalaisessa jalkapallossa ja millaista kilpailuetua sen avulla voidaan saavuttaa?
- Mitä haasteita ja ongelmia data-analytiikan käytölle on amerikkalaisessa jalkapallossa?

Ensimmäisenä alatutkimuskysymyksenä on, millaista dataa on kerätty amerikkalaisessa jalkapallossa. Tämän pohjalta voidaan siis analysoida, onko jo kerätty data ottanut huomioon tarpeeksi erilaisia tekijöitä, ja täten voidaan tunnistaa uusia datan keruutarpeita. Tätä käytetään hyödyksi myös haasteiden löytämisessä.

Toisen alatutkimuskysymyksen pohjalta tutkitaan analytiikan käyttöä amerikkalaisessa jalkapallossa ja etenkin sen suurimmassa sarjassa NFL:ssä. Tutkitaan siis tapoja data-analytiikan hyödyntämiselle sekä millaista kilpailuetua data-analytiikan hyödyntämisellä voidaan saavuttaa. Tämän kysymyksen pohjalta on hyvä edetä tutkimaan aihetta yksityiskohtaisemmin sekä haasteita ja ongelmia data-analytiikan hyödyntämiselle.

Kolmantena alatutkimuskysymyksenä tässä työssä on, mitä haasteita ja ongelmia data-analytiikan käyttö urheilussa aiheuttaa. Tämän voidaan ajatella olevan hyvä alatutkimuskysymys, koska urheilussa ja etenkin amerikkalaisessa jalkapallossa on hyvin monia suoritukseen vaikuttavia tekijöitä. Tällä pyritään siis selvittämään, mitä kaikkia tekijöitä on otettava huomioon päätöksiä tehtäessä. Tällaisia tekijöitä voivat olla esimerkiksi yksittäisten pelaajien taitotasot, jotka lisäävät muuttujien määrää huomattavasti (Vinue & Epifanio 2017).

Tutkimuksen tavoitteena on siis saada kokonaiskuva data-analytiikan hyödyntämisen nykytilasta amerikkalaisen jalkapallon pelin sisäisessä valmennuksessa. Tavoitteeseen pyritään pääsemään jakamalla aihe osiin ja tutkimalla eri osia yksitellen sekä toisiaan hyödyntäen.

1.3 Työn rakenne

Työtä pohjustetaan johdannon lisäksi tutkimusmenetelmän -ja aineiston esittelyllä. Työn kolmas luku esittelee teoriaa tutkimuksen taustalla käsittelemällä dataohjautuvaa päätöksentekoa. Neljännessä luvussa esitellään yleisesti urheiluanalytiikkaa sekä tutkitaan sen käyttöä amerikkalaisessa jalkapallossa käyttötapojen ja niillä saavutettavan kilpailuedun näkökulmasta. Viidennessä luvussa pohditaan haasteita ja ongelmia data-analytiikan hyödyntämiselle amerikkalaiselle jalkapallossa. Luvussa pohditaan siis syitä sille,

minkä vuoksi analytiikan hyödyntäminen on vähäistä suhteessa sen tuomiin mahdollisuuksiin ja hyötyihin. Näiden osioiden lisäksi tutkimuksen tulokset kootaan yhteenvetossa ja pohditaan mahdollisia jatkotutkimustarpeita.

2. TUTKIMUSMENETELMÄ JA -AINEISTO

Tässä osiossa esitellään tutkimuksessa käytettävä menetelmä sekä käydään läpi tutkimusprosessia.

2.1 Tutkimusmenetelmä

Tutkimus toteutettiin kirjallisuuskatsauksena hyödyntäen Finkin (2014) seitsemän kohdan prosessimallia. Mallin kohdat ovat seuraavat:

1. Tutkimuskysymysten valinta
2. Lähdetietokantojen valinta
3. Hakusanojen valinta
4. Hakukriteerien valinta
5. Metodologian soveltaminen
6. Kirjallisuuskatsauksen tekeminen
7. Tulosten kokoaminen (Fink 2014)

Työssä käytettiin lähdetietokantoina Googlen Scholar-palvelua ja Tampereen yliopiston Andoria. Näistä kahdesta työssä suosittiin Andoria, sillä kaikki aineisto siellä on ilmaiseksi saatavilla Tampereen Yliopiston opiskelijoille Googlen Scholar-palvelussa monien artikkelien ollessa maksullisia. Hakupalveluiden lisäksi tutkittiin Journal of Sports Analytics -julkaisua, jossa nimensä mukaisesti oli urheiluanalytiikkaa koskevia julkaisuja, joita voidaan pitää laadukkaina. Aineistosta etsittiin tietoa yleisesti data-analytiikasta, data-analytiikan käytöstä urheilusta, joiden lisäksi data-analytiikasta amerikkalaisessa jalkapallossa, mikä tässä työssä onkin päätutkimuskohde ja sitä suurin osa lähteistä käsittelee. Tutkimusaineistoa tutkittaessa big data nousi melko isoon rooliin tutkimusta tehdessä, joten myös siitä etsittiin tietoa etenkin työn teoriaosiota varten.

2.2 Tutkimusaineisto

Alla olevassa taulukossa on esiteltynä valitut hakusanat sekä niiden tuottamat hakutulosten lukumäärät Andor- ja Scholar-hakupalveluissa.

Hakutermit	Tulosten määrä, Andor	Tulosten määrä, Scholar
"Data analytics" AND Sport	1 638	33 000
Sport* analytics	36 340	185 000
"Data driven" AND Decision	29 696	1 450 000
Big data AND Sport	17 481	2 150 000
Analytics AND ("American football" OR NFL)	845	25 000

Taulukko 1: Hakutermit ja niiden hakutulosten määrä

Kuten taulukosta nähdään, näillä hakusanoilla löydettiin hyvin aihetta koskevia artikkeleita. Hieman haasteita aiheutti amerikkalaisen jalkapallon nimitys, sillä Yhdysvalloissa siitä käytetään usein nimitystä "Football" nimityksen "American football" sijaan. Ajoittain oli haastavaa erottaa eurooppalaista ja amerikkalaista jalkapalloa toisistaan hakutuloksia tutkiessa, sillä "Football" nimitys antaa tuloksia molempien maanosien jalkapalloista. Tästä syystä tietoa haettiin myös amerikkalaisen jalkapallon johtavan liigan NFL:n nimellä lajin nimen rinnalla hakuoperaattoria hyödyntäen.

Lähteistä pyrittiin löytämään poimintoja, jotka auttavat ratkaisemaan valittua tutkimusongelmaa ja vastaamaan tutkimuskysymyksiin. Tavoitteena oli käyttää tutkimuksessa pääosin melko tuoreita lähteitä, jotta tiedot olivat relevantteja ja ajantasaisia. Esimerkiksi kymmenen vuotta sitten data-analytiikan käyttö urheilussa on ollut hyvin paljon yksinkertaisempaa ja vähemmän suosittua kuin viimeisen muutaman vuoden aikana. Joitakin vanhempiakin lähteitä tutkimuksessa on käytössä, mutta niitä voidaan pitää kuitenkin soveltuvina tutkimukseen, sillä ne eivät suoranaisesti käsittele data-analytiikkaa ja sen käytön nykytilaa amerikkalaisessa jalkapallossa.

Tekijä(t)	Keskeinen sisältö
Abeza et al. (2022)	Tutkimuksessa on haastateltu Yhdysvaltojen eri urheilulajien pääsarjojen joukkueiden data-analytiikan parissa työskenteleviä henkilöitä. Tutkimus sisältää hyvin tietoa analytiikan hyödyntämisen tilasta sekä haasteista yleisesti urheilussa sekä myös erikseen NFL:ssä.
Clark (2018)	Artikkeli kertoo analytiikan käytön kehityksestä sekä haasteista NFL:ssä.
Pelechrinis (2016)	Tutkimuksessa käsitellään valmentajan tekemiä päätöksiä pelin sisäisissä tilanteissa sekä niiden vaikutusta joukkueen menestymisen kannalta.
Gandhi et al. (2022)	Kirjassa kerrotaan dataohjautuvasta päätöksenteosta sekä big datasta.

Taulukko 2: Keskeinen aineisto

Yllä olevassa taulukossa on esiteltyä tässä työssä käytetty keskeisin aineisto. Näistä suurimassa roolissa oli ensimmäisenä taulukossa esitelty tutkimus, josta löytyi paljon relevanttia tietoa tätä tutkimusta varten, sillä siinä haastateltiin tämän työn käsittelemän aiheen parissa työskenteleviä henkilöitä.

3. DATAOHJAUTUVA PÄÄTÖKSENTEKO

Datan ja analytiikan tuomat mahdollisuudet ovat merkittävästi lisänneet erilaisten organisaatioiden kiinnostusta liiketoimintatietoa ja analytiikkaa kohtaan niiden tuodessa apua organisaatioiden päätöksentekoon analysoimalla tärkeää liiketoimintaan liittyvää dataa (Chen et al. 2012). Datan määrä kasvaa Gandhin et al. (2022) mukaan joka päivä ja big datasta tulee yksi mielenkiintoisimmista mahdollisuuksista seuraavan muutaman vuoden aikana. Erilaisten tekniikoiden kuten pilvipalveluiden avulla dataa pystytään analysoimaan ja hallitsemaan pienemmillä kustannuksilla ja vaivattomammin (Gandhi et al. 2022). Organisaatioiden on siten helpompi käyttää dataa hyödyksi päätöksenteossa ja sen avulla luoda lisäarvoa toimintoihinsa. Tämä pätee myös urheiluorganisaatioiden toimintaan, vaikka urheilussa tullaankin yleisellä tasolla tässä asiassa hieman myöhässä muihin aloihin verrattuna (Abeza et al. 2022). Dataohjautuvalla päätöksenteolla tarkoitetaan Gandhin et al. (2022) mukaan päätösten tekemistä data-analytiikkaan ja sen tulkitsemiseen pohjautuen.

Tietotarpeet ovat keskeinen osa dataohjautuvaa päätöksentekoa, kun tutkitaan saatavilla olevaa dataa ja sen riittävyyttä. Tietotarpeilla tarkoitetaan nykyisen tiedon ja päätöksentekoon tarvittavan tiedon väliin jäävää aukkoa eli tarvittavaa tietoa, jota ei vielä ole (Laihonen et al. 2013). Tietotarpeiden määrittely on hyvin tärkeä vaihe tiedonhallinnassa, jotta pystytään selvittämään esimerkiksi, millaista dataa tarvitsee kerätä tarvittavan tiedon luomiseksi, jotta sitä pystytään käyttämään hyödyksi organisaation päätöksentekotilanteissa. Tietotarpeet ovat hyvin merkittävä osa tätä tutkimusta, kun puhutaan urheiluanalytiikan haasteista amerikkalaisessa jalkapallossa. Pyritään siis tutkimaan valmentajan tietotarpeita sekä niihin vastaamista koskien pelin sisäistä taktista päätöksentekoa.

Big datan yleistyminen tuo kuitenkin mukanaan haasteita saatavilla olevan datan valttavan määrän seurauksena. Datan suuren määrän vuoksi sen tehokas analysoiminen muuttuu haastavaksi. (Merendino et al. 2018) Tämän vuoksi voidaankin siis pitää hyvin tärkeänä edellisessä kappaleessa esiteltyä tietotarpeiden määrittelyä. Kun tietotarpeet on määritelty hyvin ja tarkasti, on helpompaa päätellä, onko data ja sen analysointi tarpeellista vai vain ajan ja resurssien hukkaamista. Janssenin et al. (2017) mukaan big datan käyttö ja sen analysointi ei suoranaisesti tarkoita parempaa päätöksentekoa, sillä on erittäin paljon tekijöitä, jotka vaikuttavat big datan ja sen analysoimisen tuomiin hyötyihin.

Organisaatioiden täytyy kehittää monipuolisia prosesseja, jotta ne voivat ymmärtää kontekstin ja tarkan merkityksen käytettävälle big datalle, jotta sitä voidaan käyttää hyödyksi päätöksenteossa. Big data voi tarjota uusia ja yllättäviäkin näkökulmia, joilla voidaan parantaa päätöksentekoa. Tämä kuitenkin vaatii hyvän yleiskuvan ja ymmärryksen datan lähteistä ja niiden luotettavuudesta sekä siitä, miten dataa on prosessoitu. Näin voidaan ymmärtää kerätyn datan merkitys ja luotettavuus tuloksia käsitellessä. Näissä big dataa käsittelevissä prosesseissa voidaan usein tarvita organisaation ulkopuolista apua big data -asiantuntijoilta. (Janssen et al. 2017) Kaikilla NFL-joukkueilla on omat analytiikkatiimit, jotka pyrkivät auttamaan omaa organisaatiotaan päätöksentekotilanteissa (Walder 2022).

Dataohjautuvaa päätöksentekoa on viime vuosina alettu tehostamaan käyttämällä apuna koneoppimista ja tekoälyä. Tekoälyllä tarkoitetaan tietokoneen kyvykkyyksiä toimia ihmisen tavoin, eli esimerkiksi sen kykenemistä ihmismäisiin ajatusprosesseihin kuten oppimiseen, järkeilyyn ja virheistä oppimiseen (Kok et al. 2009). Koneoppiminen on yksi aktiivisimmin tutkittavista tekoälyn osa-alueista (Wang et al. 2022). Koneoppimisen avulla pyritään luomaan malleja ja kaavoja sekä löytämään yhteyksiä ja korrelaatioita suurista määristä dataa (Esposito & Esposito 2020). Hyödyntämällä koneoppimisen algoritmeja organisaatioiden ei tarvitse enää vain luottaa ihmisten osaamiseen monimutkaisissa datan käsittely- ja analysointiprosesseissa, vaan voidaan säästää tärkeitä resursseja ja aikaa sekä saavuttaa parempaa reagointikykyä muutoksiin (Fernández-Revueleta Pérez & Romero Blasco 2022). Monissa luoduissa malleissa, joita esitellään seuraavassa osiossa, on käytetty hyödyksi koneoppimista.

4. URHEILUANALYTIIKAN KÄYTTÖ AMERIKKALAISEN JALKAPALLON VALMENNUKSESSA

Tässä osiossa käsitellään data-analytiikan käyttöä amerikkalaisen jalkapallon valmennuksessa. Tutkitaan siis tapoja data-analytiikan käytölle, jotka ovat käytössä sekä niiden tapojen vaikutusta kilpailullisen edun luomiseen.

4.1 Urheiluanalytiikka

Datan hyödyntäminen urheilussa on selkeästi yleistymässä. Kuitenkin Abezan et al. (2022) haastattelemien Yhdysvaltojen eri lajien pääsarjajoukkueiden datan parissa työskentelevien henkilöiden mukaan urheilussa tullaan datan hyödyntämisen kanssa jäljessä muita aloja kuten pankki-, lento- ja verkkokaupankäyntialaa. Kaikkien artikkelin 21:n haastatellun mukaan big dataa käytetään sekä pelin ulkopuoliseen että pelin sisäiseen analysoimiseen. Haastateltavat myös totesivat big datan hyödyntämisen vaihtelevan suuresti eri lajien ja sarjojen välillä. (Abeza et al. 2022) Kuten jo johdannossa mainittiin, Clarkin (2018) mukaan amerikkalainen jalkapallo on urheiluanalytiikan hyödyntämisessä muita lajeja jäljessä.

Chun & Wangin (2019) baseballin data-analytiikan käyttöön keskittyneen empiirisen tutkimuksen mukaan urheiluanalytiikan käytöllä on selkeä korrelaatio joukkueen menestymisen kanssa. Tutkimuksessa selvitettiin joukkueiden menestymistä baseballissa verraten sitä joukkueen data-analytiikan käyttöön uskomiseen. Tutkimuksen mukaan data-analytiikkaan uskovat ja sitä käyttävät joukkueet pärjäsivät muita selkeästi paremmin, kun tutkittiin joukkueiden pääsemistä pudotuspeleihin. (Chu & Wang 2019) Urheiluanalytiikalla voidaan siis todeta olevan positiivista korrelaatiota menestymisen kannalta ainakin baseballin saralla.

4.2 Amerikkalainen jalkapallo

Amerikkalainen jalkapallo on peli, jota pelataan 100 jaardin eli n. 91 metrin pituisella kentällä. Alla esitetyt säännöt ovat käytössä NFL:ssä. Molemmilla joukkueilla on kentällä 11 pelaajaa. Joukkueet yrittävät saada vietyä pallon vastustajan maalialueelle, jota kutsutaan touchdowniksi. Tästä saa kuusi pistettä sekä mahdollisuuden yrittää lisäpistettä potkaisemalla potkumaali 15 jaardin päästä tai kahta lisäpistettä etenemällä kahden jaardin päästä maalialueelle. (Alder 2019)

Eteneminen kentällä tapahtuu joko juoksemalla pallon kanssa tai heittämällä sitä. Joukkueella on neljä yritystä edetä kymmenen jaardia, jonka edetessään joukkue saa uudet neljä yritystä. Mikäli joukkue ei etene kymmentä jaardia neljällä yrityksellä, saa vastustaja pallon siitä kohdasta, mihin hyökkäävä joukkue jää. Neljänellä yrityksellä yhtenä vaihtoehtona on myös puntata pallo eli potkaista se vastustajan päättyyn, jolloin vastustaja jatkaa sieltä, mihin pallo päätyy. Toisena vaihtoehto on potkaista potkumaali vastustajan maalialueen takana olevaan haarukkaan, mistä hyökkäävä joukkue saa kolme pistettä. Peli-aika on 60 minuuttia, joka on jaettu kahteen 30 minuutin puoliaikaan ja neljään 15 minuutin neljännekseen. Ottelut kestävät kokonaisuudessaan keskimäärin noin kolme tuntia. (Alder 2019) Yhdessä ottelussa tehdään NFL:ssä keskimäärin noin 45 pistettä (Epstein 2022).

4.3 Saatavilla oleva data NFL:ssä

NFL on viimeisen vuosikymmenen aikana ottanut askelia edistyneemmän datan keräämisen suuntaan. Vuonna 2014 kaikkien pelaajien varusteisiin asennettiin sirut, joilla pystyttiin mittaamaan pelaajan sijaintia kentällä, nopeutta, ja pelaajan hartialinjan suuntaa, eli mihin suuntaan pelaaja liikkuu tai katsoo. Vuonna 2017 tämä samanlainen tekniikka lisättiin myös palloihin, jotta niidenkin liikettä pystyttiin analysoimaan paremmin. Vasta vuonna 2018 NFL julkaisi kerätyn datan joukkueille, mutta ei vielä silloinkaan tehnyt datasta kokonaan julkista. Kuitenkin vuonna 2019 tätä dataa alettiin julkaisemaan myös yleisön käyttöön. (Ennis 2020)

Vuodesta 2019 lähtien NFL on alkanut järjestämään vuosittain avointa kilpailua nimeltään Big Data Bowl, jossa yleisölle annetaan mahdollisuuksia luoda omia malleja hyödyntäen esimerkiksi kehittyneitä koneoppimisen tekniikoita edellisessä kappaleessa esitettyyn dataan. Etenkin laitahyökkääjien, eli heitettyjen pallojen kiinniottajien dataa hyödynnettiin paljon Big Data Bowl -kilpailussa, koska datasta pystyttiin vertaamaan heidän sijaintiaan puolustajiin. Mitä kauempana he ovat lähimmästä puolustajasta, sitä helpompaa ja turvallisempaa heille on heittää palloa. Myös pallon kanssa juokseville pelaajille luotiin malli, jonka avulla pystyttiin arvioimaan kuinka monta jaardia he pystyvät juoksemaan saadessaan pallon käteensä muiden pelaajien sijaintien perusteella. (NFL Operations 2023a) Melko yksinkertaisestakin datastakin on siis pystytty luomaan hyvin monipuolisia ja hyödyllisiä malleja.

NFL:n keräämän datan lisäksi otteluvideoilta on pystytty keräämään erilaista dataa. Tätä on etenkin edesauttanut Clarkin (2018) mukaan niin sanotun "all-22"-kuvakulman tuominen julkiseksi. Tuolla kuvakulmalla tarkoitetaan sitä, että kuvassa näkyy kaikki kentällä

olevat 22 pelaajaa. Siitä pystytään keräämään monipuolisesti erilaista dataa yleisön toimesta ja se on luonut paljon edellytyksiä yleisön omien mallien luomiselle sekä helpottanut sitä. (Clark 2018)

4.4 Urheiluanalytiikan käyttö amerikkalaisen jalkapallon valmennuksessa

Tapoja ja käyttötarkoituksia urheiluanalytiikalle amerikkalaisen jalkapallon valmennuksessa on monia. Analytiikkaa voidaan käyttää esimerkiksi ennustamaan vastustajan pelitaktiikoita sekä arvioimaan vaihtoehtoja omissa päätöksentekotilanteissa. Näitä tilanteita avataan tässä osiossa.

4.4.1 Vastustajan hyökkäyspäätösten ennustaminen

Yksi analytiikan käyttökohteista amerikkalaisessa jalkapallossa on vastustajan päätösten ennustaminen ja tätä voidaan pitää valmentajalle hyvin merkittävänä tietotarpeena.

Joashin et al. (2020) amerikkalaista jalkapalloa käsittelevässä tutkimusartikkelissa esitellyllä koneoppisen avulla luodulla mallilla pystyttiin arvioimaan 75,3 % todennäköisyydellä, että juokseeko vai heittääkö hyökkäämässä oleva joukkue. Tällä tiedolla puolustava joukkue voi saada merkittävää kilpailullista etua valitsemalla oikeita taktiikoita puolustukseen hyökkäystä. Artikkelissa esiteltiin myös erilliset mallit jokaiselle joukkueelle ja niiden tarkkuus vaihteli n. 65 %:n ja 82,5 %:n välillä. (Joash et al. 2020)

Joashin et al. (2020) mallissa otettiin hyvin huomioon erilaisia tekijöitä. Siinä käytettiin esimerkiksi suosittua amerikkalaisen jalkapallon videopelin, Maddenin, antamia arvioita pelaajien taitotasosta. Tämän avulla pystyttiin arvioimaan joukkueiden voimasuhteita kokonaisuudessaan sekä pelin eri osa-alueilla. Esimerkiksi hyvillä laitahyökkääjillä, eli hyökkäävillä pelaajilla, jotka ottavat heitettyjä palloja kiinni, varustettu joukkue oli herkempi heittämään palloa juoksemisen sijaan ja tämä otettiin mallissa huomioon. (Joash et al. 2020)

Valmentajan tietotarpeisiin pyritään vastaamaan malleilla, mutta ne eivät luonnollisesti-kaan voi antaa jokaisella kerralla oikeaa ratkaisua. Tätä mallia voidaan kuitenkin käyttää apuna päätöksentekotilanteissa sen ollessa oikeassa melko hyvällä tarkkuudella eli kolmella kerralla neljästä. Tämä mallin toimimisen todennäköisyys on kuitenkin hyvä pitää mielessä päätöksiä tehdessä, jotta osataan varautua myös tilanteisiin, jossa malli ei anna oikeaa ratkaisua.

4.4.2 Neljäs yritys

Mielenkiintoinen käyttökohde analytiikalle on myös hyökkävän joukkueen päätöksenteko neljännen yrityksen tilanteessa. Nämä päätökset herättävät hyvin usein kiivasta keskustelua sosiaalisessa mediassa. Kuten kappaleessa 4.2 kerrottiin, amerikkalaisessa jalkapallossa hyökkävällä joukkueella on neljä yritystä edetä yhteensä 10 jaardia (hieman yli yhdeksän metriä). Mikäli neljällä yrityksellä ei pääse etenemään tuota 10 jaardia, saa vastustaja pallon siitä kohtaa mihin neljäs yritys jäi. Vaihtoehtoisesti hyökkävä joukkue voi neljännellä yrityksellä puntata eli potkaista pallon vastustajan maalia kohti, jolloin puolustava joukkue saa pallon siitä kohtaa, mihin se jää puntauksen jälkeen. Kolmas vaihtoehto on potkaista potkumaali, jolla joukkue tekee kolme pistettä. Potkumaa- leja yritetään usein neljännellä yrityksellä, kun hyökkävä joukkue on alle 40 jaardin päässä vastustajan maalista. Näiden vaihtoehtojen välillä tehtyjä päätöksiä voidaan hyvin paljon sosiaalisessa mediassa, sillä ne ovat pelin tuloksen kannalta usein hyvinkin ratkaisevassa asemassa. (Pelechris 2016) Monet valmentajat tekevät näissä tilanteissa Williamsin et al. (2022) mukaan usein konservatiivisempia ratkaisuja kuin perusteella olisi järkevää tehdä, mutta viime vuosina rohkeammat ratkaisut ovat alkaneet yleistymään analytiikan suosion selkeän kasvun ansiosta.

Tätä käyttötarkoitusta varten on vuonna 2013 rakennettu New York Timesin toimesta 4th Down Bot, joka kertoo, mitä missäkin tilanteessa, eli kenttäaseman ja uusiin yrityksiin tarvittavien jaardien määrän perusteella, kannattaa tehdä. Tätä varten on kerätty dataa vuosilta 2000–2013, eli dataa on siis hyvin laajalti. Tämä työkalu ei kuitenkaan ota huomioon esimerkiksi joukkueiden voimasuhteita, joilla on kuitenkin merkittäviä vaikutuksia eri pelitapahtumien onnistumisen kannalta. (Burke & Quealy 2013) Tähän työkaluun ei siis kannattaisi joukkueiden voimasuhteiden huomioimatta jättämisen vuoksi sokeasti luottaa, vaan valmentajan voisi olla esimerkiksi järkevää katsoa, mitä työkalu näyttää ja sitten suhteuttaa sen tulokseen oma arvionsa joukkueiden voimasuhteista. Tällöin voisi saada realistisemmän kuvan kyseisestä päätöksentekotilanteesta. Vuonna 2015 tästä Botista esiteltiin paranneltu versio, joka pystyi ottamaan huomioon joukkueen potkaisijan ja sääolosuhteiden vaikutukset päätöksentekotilanteissa (Causey et al. 2015).

Paranneltu versio antoi siis hieman paremman kuvan tilanteesta, mutta sekään ei vielä huomioinut esimerkiksi joukkueiden voimasuhteita uusien downien eli yritysten yrittämisen onnistumisen todennäköisyyttä arvioidessa. Valmentajan tietotarpeena tässä tilanteessa on siis se, mikä olisi järkevin ratkaisu kussakin tilanteessa. Siihen ei vielä pystytty täydellisesti mallien avulla vastaamaan, mutta niitä voidaan pitää kuitenkin hyvinä ja käyttökelpoisina apuvälineinä päätöksenteossa.

4.4.3 Lisäpiste

Hieman samanlainen päätöksentekotilanne edelliseen tilanteeseen verrattuna on, kun joukkue tekee kuuden pisteen arvoisen touchdownin, eli saa vietyä pallon vastustajan maalialueelle juoksemalla tai heittämällä. Tämän jälkeen joukkue voi potkaista pallon maalihaarukkaan, josta saa yhden pisteen, tai joukkue voi yrittää juosta tai heittää pallon vastustajan maalialueelle kahden jaardin päästä maalialueesta. Tämän onnistuessa joukkue saa kaksi lisäpistettä.

Potkun onnistumisen todennäköisyys on n. 95 % ja kahden pisteen yrityksen n. 51 %. Tällöin siis matemaattisesti olisi järkevämpää yrittää tehdä kahta pistettä pois lukien tilanteet, joissa yhden ja kahden pisteen tuoma lisähyöty on sama. Näin on esimerkiksi ottelun lopulla pelin tasoittuessa touchdownilla, jolloin ei ole käytännössä merkitystä voiton todennäköisyyteen, että onko johto lisäpisteen tai lisäpisteiden teon jälkeen yhden vai kahden pisteen verran.

Tästä kahden lisäpisteen yrittämisen paremmasta odotusarvosta huolimatta valmentajat lähes poikkeuksetta valitsevat potkaista yhden pisteen arvoisen potkun, vaikka sillä ansaitaan keskimäärin laskennallisesti vähemmän pisteitä. (Pelechrinis 2016) Tässäkin tilanteessa valmentajan tietotarpeena on tietää, mikä on järkevin ratkaisu. Kuitenkaan valmentajat eivät tunnu noudattavan tätä matemaattisesti järkevintä ratkaisua.

4.5 Käytettyjen tapojen vaikutus kilpailullisen edun luomiseen

Vastustajan hyökkäyspäätösten ennustamisella voidaan Leen et al. (2016) mukaan sanoa olevan erittäin merkittäviä vaikutuksia kilpailullisen edun luomiseen. Joukkueet käyttävät lukemattomia tunteja ja resursseja saadakseen tietoja vastustajajoukkueen taipumuksista ja mieltymyksistä kentällä. Puolustavalle joukkueelle pienikin tieto vastustajan pelivalinnasta auttaa paljon puolustuksen pelikutsun ja kentälle laitettavien pelaajien valinnassa ja sijoittamisessa. Näillä valinnoilla on merkittävä vaikutus ottelun lopputulokseen. (Lee et al. 2016)

Kappaleessa 4.4.2 esitettyä neljännen yrityksen päätöksentekoa on tutkittu Yamin & Lopezin (2019) työssä luomalla malli käyttäen dataa, joka on kerätty 13 vuoden aikana NFL:ssä. Tämä malli laskee, mitä olisi käynyt, jos valmentajat olisivat toimineet osiossa 3.2. esitetyn 4th Down Bot-työkalun antamien ratkaisujen mukaan. Mallilla arvioitiin, että paremmalla päätöksenteolla yksittäinen joukkue olisi keskimäärin saanut 0,4 voittoa enemmän kauden aikana. (Yam & Lopez 2019) Tuo lukema voi kuulostaa pieneltä, mutta todellisuudessa se on hyvin suuri, sillä NFL:ssä pelattiin vielä viime kauteen (2022–2023) asti kauden aikana vain 16 runkosarjapeliä, minkä jälkeen pelien määrä

nousi yhdellä. Tämän mallin perusteella voidaan siis todeta, että analytiikan avulla voidaan luoda erittäin merkittävää kilpailullista etua joukkueelle.

Analytiikkaa hyödyntämällä pelin sisäisissä taktisissa päätöksissä voidaan siis saada aikaan hyvin merkittävääkin kilpailullista etua monissa erilaisissa tilanteissa. Analytiikan käytölle on kuitenkin paljon erilaisia haasteita, jotka ovat syynä sille, että analytiikan käyttö on vielä melko vähäistä, vaikka sen käytön on todettukin tuovat joukkueelle kilpailullista etua. Näitä haasteita käsitellään seuraavassa osiossa.

5. HAASTEET URHEILUANALYTIIKAN HYÖDYNTÄMISESSÄ AMERIKKALAISESSA JALKAPALLOSSA

Tässä osiossa käsitellään haasteita urheiluanalytiikan hyödyntämisessä amerikkalaisessa jalkapallossa. Löydetyt haasteet voidaan jakaa kahteen osaan, jotka ovat ihmisten urheiluanalytiikkaan kohdistuvat negatiiviset ja epäilevät asenteet, joita voidaan kutsua organisaatiollisiksi haasteiksi sekä urheiluanalytiikan puutteet ja ongelmakohdat, eli analytiikan käyttöön liittyvät teknisemmät ongelmat.

5.1 Organisaatiolliset haasteet

Yhtenä haasteena on Abezan et al. (2022) suorittamien haastatteluiden mukaan organisaation oppimisen kulttuuriin liittyvät haasteet. Artikkelissa mainittiin ongelmana, että organisaation eri osastojen välillä pitäisi olla yhteinen ymmärrys big datan ja analytiikan arvosta ja tämän puuttuminen tekee data-analytiikkaan pohjautuvasta päätöksenteosta haastavaa. (Abeza et al. 2022) Clarkin (2018) mukaan tähän on syynä ainakin osittain omistajien negatiivinen suhtautuminen pelillisiin päätöksiin, jotka ovat rohkeampia, kuin mihin on vuosien saatossa totuttu. Tämä siitäkkin huolimatta, että data-analytiikan avulla voiton todennäköisyys kasvaa. Täten siis jotkut valmentajat tekevät työpaikkansa menettämisen pelon vuoksi konservatiivisempia ratkaisuja kuin olisi analytiikan mukaan järkevämpää tehdä. (Clark 2018)

Tähän aiheeseen liittyen on tehty tutkimusta yliopistojalkapallon puolelta. Owensin ja Roachin (2018) mukaan Yhdysvaltojen yliopistojalkapallojoukkueiden (amerikkalainen jalkapallo) päävalmentajien ratkaisut neljänsillä yrityksillä muuttuivat konservatiivisemmiksi, kun heidän asemansa päävalmentajana heikkeni ja potkujen todennäköisyyttä pidettiin suurempana. Työpaikkansa ollessa vaarassa päävalmentajat siis tekivät ratkaisuja, jotka eivät olleet optimaalisia joukkueen voiton todennäköisyyden kannalta. Sen sijaan valmentajat, joilla oli vankka asema, tekivät riskialttiimpia ratkaisuja täten maksimoimien joukkueen todennäköisyyden voitolle. (Owens & Roach 2018)

Haasteita urheiluanalytiikan käytölle aiheuttaa myös se, että analytiikan antama tulos voi olla ristiriidassa valmentajien sekä muun johtoportaan omiin mielipiteisiin. Abezan et al. (2022) mukaan analytiikan tuoman ratkaisun ollessa ristiriidassa henkilön ”mututuntuman” kanssa monet jättävät kokonaan huomioimatta analytiikan tuoman ratkaisun. Puo-

let artikkelissa haastatelluista analytiikan parissa Yhdysvaltojen pääsarjoissa työskentelevistä mainitsivat tämän ongelman. (Abeza et al. 2022) Tätä voidaan siis pitää hyvin merkittävä haasteena ja sen perusteella voitaisiin päätellä, että analytiikkaa käytetään vain vahvistamaan omia mielipiteitä eikä yleisellä tasolla olla valmiita muutoksille omista päätöksentekotilanteissa. Tämä osittain selittää sen, miksi monet valmentajat toimivat analytiikan antamien ratkaisujen vastaisesti.

Haasteena voidaan arvostetun taloustieteen opettajan Romerin (2006) mukaan pitää myös klassista 1960-luvulla esiteltyä Ellsbergin paradoksia. Sen mukaan useimmiten ihmiset valitsevat mieluiten vaihtoehdon, jossa tiedetään todennäköisyys mielummin kuin vaihtoehdon, jonka todennäköisyyttä ei tiedetä, vaikka se olisi odotusarvoltaan suurempi. Ihmiset pyrkivät siis useimmiten valitsemaan vähiten riskialttiin ratkaisun, vaikka se ei matemaattisesti olisikaan paras ratkaisu. Tämä selittää valmentajien tekemiä päätöksiä neljänsillä yrityksillä sekä lisäpistetilanteissa valmentajien usein valitessa heikomman odotusarvon ratkaisuja riskien välttämiseksi, kun odotettu lopputulos on jo melko suurella varmuudella tiedossa. (Romer 2006)

5.2 Urheiluanalytiikan puutteet ja ongelmakohdat

Usein puhuttaessa syistä, jotka johtavat analytiikan sivuuttamiseen amerikkalaisen jalkapallon yhteydessä, nousee esiin datan määrän vähäisyys (Clark 2018). Yhden kauden aikana pelataan NFL:ssä yhteensä 285 ottelua (NFL 2023), joten muihin lajeihin verrattuna ottelumäärä on hyvin alhainen. Alhainen datan määrä johtuu siis osittain tästä pelien vähäisyydestä. Toinen syy tähän on pelitilanteiden erilaisuus. Pelitilanteeseen vaikuttaa erittäin monia tekijöitä. Joukkueiden ja pelaajien voimasuhteiden lisäksi sääolosuhteet, ottelun tilanne, pallon sijainti pelikentällä sekä monesko yritys on kyseessä ja kuinka monta jaardia sillä on edettävä uusien yritysten saavuttamiseksi. Tekijöitä on siis hyvin paljon, joten samanlaisista tilanteista on melko pieni määrä dataa saatavilla. Osa NFL:n general managereista sanoo, että kestää vielä vuosia ennen kuin dataa on kerätty tarvittava määrä sen hyödyntämiseen ja sen pohjalta päätelmien tekemiseen (Clark 2018).

Dataa on kerätty paljon myös amerikkalaisen jalkapallon yliopistosarjoista, mutta sitä dataa ei voida pitää vertailukelpoisena pääsarja NFL:ään verrattuna. Tämä johtuu siitä, että jotkin säännöt ovat osittain erilaisia näiden sarjojen välillä. Myös pelaajien taitotasissa on selkeitä eroja, mikä vaikuttaa myös joukkueiden pelitaktiikoihin. Esimerkiksi NFL:n potkaisijat onnistuvat potkumaaliyrityksissä selkeästi useammin ja pidemmiltä

etäisyyksiltä kuin yliopistojoukkueiden potkaisijat. Tällä on esimerkiksi vaikutusta neljän-
nen yrityksen päätöksentekoon sekä hieman myös lisäpistetilanteisiin. (Foster et al.
2021)

Tilanteisiin vaikuttavien muuttujien suuri määrä nousee esiin monissa tutkituissa mal-
leissa. Hyvin usein malleista jää pois paljon muuttujia, joilla olisi todellisuudessa vaiku-
tusta tilanteisiin ja siten myös mallin toimintaan ja oikeellisuuteen. Monien malleja esitte-
levien artikkelien lopussa selitetään jatkotutkimustarpeiksi useampien muuttujien sisällyt-
tämistä malliin tarkkuuden lisäämiseksi.

Haasteita data-analytiikan käytölle tuovat myös NFL:n säännöt. Valmentajilla ei ole sallit-
tua olla käytössä apuvälineenä kentän laidalla muuta kuin kuvia ja tekstiä (NFL Operati-
ons 2023b). Ei siis ole mahdollista syöttää tiettyjä parametrejä johonkin ohjelmaan, joka
antaisi esimerkiksi todennäköisyydet jollekin tapahtumalle. Tämä luonnollisesti luo mer-
kittävän haasteen erilaisten mallien käytölle pelin aikana.

6. YHTEENVETO

Työn päätutkimuskysymyksen eli amerikkalaisen jalkapallon pelin sisäisen data-analytiikan käytön nykytilan ja mahdollisuuksien selvittämisen tutkiminen tehtiin jakamalla päätutkimuskysymys alatutkimuskysymyksiin, jotka yhdistettynä kattavat vastauksen päätutkimuskysymykseen. Alatutkimuskysymyksillä käsiteltiin tapoja, joilla data-analytiikkaa hyödynnetään pelin sisäisessä päätöksenteossa sekä sen vaikutusta kilpailullisen edun luomiseen. Alatutkimuskysymyksillä tutkittiin myös, millaista dataa on kerätty aiheesta sekä millaisia haasteita data-analytiikan käyttöön liittyy tässä kontekstissa. Alatutkimuskysymyksiin saatiin kattavat vastaukset ja täten myös päätutkimuskysymyksen vastauksesta saatiin kattava.

6.1 Tutkimuksen tulokset

NFL:ssä kerätty data on melko kattavaa, mutta sen jakaminen joukkueille ja yleisölle tapahtui vasta vuosia sen keräämisen aloittamisen jälkeen, mikä osaltaan on hidastanut data-analytiikan käytön kehitystä amerikkalaisessa jalkapallossa. Kuitenkin viime vuosina datan ollessa yhä paremmin saatavilla, on luotu paljon malleja erilaisista tilanteista, joita voidaan pitää hyödyllisinä valmentajan tekemissä pelin sisäisissä päätöksissä.

Data-analytiikkaa voidaan hyödyntää pelin sisäisessä päätöksenteossa esimerkiksi enustamaan vastustajan päätöksiä hyökkäysvuorolla. Tämän lisäksi sitä voidaan hyödyntää omalla hyökkäysvuorolla neljännen yrityksen sekä touchdownin jälkeisen lisäpisteen päätöksentekotilanteissa. Näillä data-analytiikan hyödyntämisen tavoilla voidaan saavuttaa merkittävääkin kilpailullista etua. Siitä huolimatta monet valmentajat toimivat näissä tilanteissa päinvastaisesti data-analytiikan tuomien ratkaisuiden kanssa erilaisten haasteiden ja ongelmien vuoksi.

Data-analytiikan hyödyntämisen haasteista monet liittyvät organisaation kulttuuriin. Usein joukkueiden eri osastojen välillä ei vallitse yhteisymmärrystä big datan ja analytiikan arvosta, mikä vaikeuttaa niiden hyödyntämistä päätöksentekotilanteissa. Etenkin joukkueiden omistajat ovat usein rohkeampia ratkaisuja vastaan, vaikka ne olisivatkin data-analytiikan suosittamia. Tämä johtaa siihen, että varsinkin valmentajat, jotka tuntevat olevan vaarassa menettää työpaikkansa, päätyvät tekemään konservatiivisempia ratkaisuja, vaikka ne analytiikan mukaan eivät optimoi joukkueen voiton todennäköisyyttä.

Monet valmentajat ja johtoportaan henkilöt käyttävät analytiikkaa vain vahvistamaan omia mielipiteitään eivätkä ole valmiita muuttamaan mielipiteitään, vaikka data-analytiikka niin suosittelisikin. Tämä aiheuttaa myös haasteita analytiikan käytölle. Myös Ellsbergin paradoksi selittää data-analytiikan suhteellisen vähäistä käyttöä. Sen mukaan useimmiten ihmiset valitsevat mieluiten vaihtoehdon, jossa tiedetään todennäköisyys mielummin kuin vaihtoehdon, jonka todennäköisyyttä ei tiedetä, vaikka se olisi odotusarvoltaan suurempi. Tämä paradoksi selittää konservatiivisten ratkaisujen suosiota.

Pelien vähäinen määrä ja sen seurauksena vähäisen datan määrän katsotaan rajoittavan data-analytiikan käyttöä. Tähän liittyy myös peliin vaikuttavien muuttujien suuri määrä. Tällöin siis monet katsovat, että samat muuttujat sisältävien tilanteiden määrä ei ole tarpeeksi suuri johtopäätöksien tekemiseksi. Monissa tutkimuksissa nousi esiin, että mallit eivät sisällä tarpeeksi muuttujia tarkkojen tulosten saamiseksi. NFL:ssä ei myöskään ole sallittua, että valmentajalla olisi kentän laidalla muita apuvälineitä kuin tekstin ja kuvien muodossa olevia, mikä luo myös haasteita erilaisten mallien käytölle pelin aikana.

Tutkimuksen pohjalta voidaan siis todeta, että dataa tarvitsee kerätä lisää, jotta malleista tulee entistä luotettavampia. Kuitenkin nykyiselläänkin niillä voidaan saavuttaa merkittävää kilpailullista etua, mutta data-analytiikan hyödyntämistä haittaavat erilaisten organisaatiolliset haasteet.

6.2 Tutkimuksen arviointi ja jatkotutkimustarpeet

Tutkimuksen tulosten avulla saadaan hyvä kokonaiskuva amerikkalaisen jalkapallon pääsarja NFL:n data-analytiikan hyödyntämisen ja dataohjautuvan päätöksenteon nykytilasta sekä niiden tuomista mahdollisuuksista. Täten siis työn tavoitteeseen päästiin. Ennestään ei ollut tämän työn kaltaista data-analytiikan käyttöä pelin sisäisessä päätöksenteossa amerikkalaisessa jalkapallossa kokoavaa tutkimusta.

Finkin prosessimallia käyttäen kirjallisuuskatsauksen tekeminen oli systemaattista ja siinä onnistuttiin hyvin vaihe vaiheelta edeten. Soveltuvia lähteitä löytyi tarpeeksi kattavasti työn laajuuteen nähden, ja ne olivat melko tuoreita sekä pääosin vertaisarvioituja. Niitä ei kuitenkaan ollut niin paljoa, että aineistojen vertailu ja vuoropuhelu olisi ollut mahdollista. Lähes kaikki aineisto oli toisiaan tukevia, eli ristiriidassa olevia väitteitä ja tuloksia ei löydetty. Niiden avulla tutkimukseen olisi voitu saada hieman enemmän pohdintaa aineistoon liittyen.

Työssä ei käsitelty juurikaan kehitysehdotuksia haasteiden ja ongelmien korjaamiseksi. Näitä olisi hyvä tutkia lisää, jotta data-analytiikan käyttöä voitaisiin kehittää amerikkalaisen jalkapallon pelin sisäisessä päätöksenteossa. Monet etenkin organisaatiollisista haasteista ovat soveltuvia myös muihin urheilulajeihin sekä urheilun ulkopuolellekin, joten kehitysehdotuksia voitaisiin soveltaa myös amerikkalaisen jalkapallon ulkopuolelle. Tämä tutkimus luo hyvän pohjan mahdollisille jatkotutkimuksille tästä aiheesta.

LÄHTEET

- Abeza, G., O'Reilly, N., Nadeau, J. & Abdourazakou, Y. 2022. Big data in professional sport: the perspective of practitioners in the NFL, MLB, NBA, and NHL. *Journal of Strategic Marketing*, ahead-of-print 1–21. Saatavilla: <https://doi.org/10.1080/0965254X.2022.2108881> (16.2.2023)
- Alder, J. 2019. The Basic Rules of Football. Saatavilla: <https://www.liveabout.com/football-101-the-basics-of-football-1333784> (15.4.2023)
- Burke B. & Quealy K. 2013. How Coaches and the NYT 4th Down Bot Compare. Saatavilla: <http://www.nytimes.com/newsgraphics/2013/11/28/fourth-downs/post.html> (30.3.2023)
- Causey, T., Katz, J. & Quealy, K. 2015. A Better 4th Down Bot: Giving Analysis Before the Play. *New York Times (Online)*. Saatavilla: <https://www.nytimes.com/2015/10/02/upshot/a-better-4th-down-bot-giving-analysis-before-the-play.html> (4.4.2023)
- Chen, H., Chiang, R. H. L. & Storey, V. C. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188. Saatavilla: <https://doi.org/10.2307/41703503> (12.4.2023)
- Chu, D. & Wang, C. 2019. Empirical study on relationship between sports analytics and success in regular season and postseason in Major League Baseball. Saatavilla: <https://content.iospress.com/articles/journal-of-sports-analytics/jsa190269> (3.2.2023)
- Clark, K. 2018. The nfl's analytics revolution has arrived. Saatavilla: <https://www.theringer.com/nfl/2018/12/19/18148153/nfl-analytics-revolution>. (30.3.2023)
- Ennis, A. (2020). An Introduction to Modeling NFL Tracking Data (Doctoral dissertation, New York, NY: Yeshiva College. Yeshiva University.). Saatavilla: <https://repository.yu.edu/handle/20.500.12202/5659> (15.4.2023)
- Epstein, J. 2022. NFL scoring is down in 2022. What are the causes? Yahoo sports. Saatavilla: <https://sports.yahoo.com/nfl-scoring-is-down-in-2022-what-are-the-causes-013232184.html> (15.4.2023)
- Esposito, D. & Esposito, F. 2020. *Introducing machine learning* (1st edition). Published with the authorization of Microsoft Corporation by Pearson Education. Saatavilla: https://andor.tuni.fi/permalink/358FIN_TAMPO/1j3mh4m/alma9911130590205973 (13.4.2023)
- Ferguson, D. 2018. Utilizing Analytics in American Football to Improve Decision Making on Fourth Down. *Graduate Theses - Engineering Management*. 7. Saatavilla: https://scholar.rose-hulman.edu/engineering_management_grad_theses/7 (2.2.2023)
- Fernández-Revuelta Pérez, L. & Romero Blasco, Á. 2022. A Data Science Approach to Cost Estimation Decision Making - Big Data and Machine Learning. *Revista de Contabilidad*, 25(1), 45–57. Saatavilla: <https://doi.org/10.6018/rcsar.401331> (14.4.2023)
- Fink, A. 2014. *Conducting research literature reviews : from the Internet to paper* (4th ed.). Sage

Foster, G., O'Reilly, N. & Naidu, Z. 2021. Playing-Side Analytics in Team Sports: Multiple Directions, Opportunities, and Challenges. *Frontiers in Sports and Active Living*, 3, 671601–671601. <https://doi.org/10.3389/fspor.2021.671601> (14.4.2023)

Gandhi, P., Bhatia, S. & Dev, K. 2022. Data driven decision making using analytics (P. Gandhi, S. Bhatia, & Kapil Dev, Eds.; First edition.). CRC Press. p. 107–108. Saatavilla: <https://www-taylorfrancis-com.libproxy.tuni.fi/books/edit/10.1201/9781003199403/data-driven-decision-making-using-analytics-parul-gandhi-surbhi-bhatia-kapal-dev> (16.2.2023)

Janssen, M., van der Voort, H. & Wahyudi, A. 2017. Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338–345. Saatavilla: <https://doi.org/10.1016/j.jbusres.2016.08.007> (13.4.2023)

Joash Fernandes, C., Yakubov, R., Li, Y., Prasad, A. & Chan, T. 2020. Predicting plays in the National Football League. *Journal of Sports Analytics*, vol. 6, no. 1, pp. 35-43. Saatavilla: <https://content.iospress.com/articles/journal-of-sports-analytics/jsa190348> (9.3.2023)

Kok, J. N., Boers, E. J., Kusters, W. A., Van der Putten, P. & Poel, M. 2009. Artificial intelligence: definition, trends, techniques, and cases. *Artificial intelligence*, 1, 270-299. Saatavilla: <https://www.eolss.net/Sample-Chapters/C15/E6-44.pdf> (13.4.2023)

Laihonon, H., Hannula, M., Helander, N., Ilvonen, I., Jussila, J., Kukko, M., Kärkkäinen, H., Lönnqvist, A., Myllärniemi, J., Pekkola, S., Virtanen, P., Vuori, V. & Yliniemi, T. 2013. Tietojohdaminen. Tampereen teknillinen yliopisto, Tietojohdamisen tutkimuskeskus Novi. Saatavilla: <https://trepo.tuni.fi/bitstream/handle/10024/128357/tietojohdaminen.pdf?sequence=1&isAllowed=y> (16.2.2023)

Lee, P., Chen, R. & Lakshman, V. 2016. Predicting offensive play types in the National Football League. Saatavilla: <https://www.semanticscholar.org/paper/Predicting-Offensive-Play-Types-in-the-National-Lee-Chen/37c5268dc039d4f17bb8cbb66881bf1bf8187dba?p2df> (14.4.2023)

Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L. & Canhoto, A. 2018. Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, 93, 67–78. Saatavilla: <https://doi.org/10.1016/j.jbusres.2018.08.029> (13.4.2023)

NFL. 2023. Schedule. Saatavilla: <https://www.nfl.com/schedules/> (5.4.2023)

NFL Operations. 2023a. Big Data Bowl. Saatavilla: <https://operations.nfl.com/the-game/big-databowl/> (14.4.2023)

NFL Operations. 2023b. Technology and the game. Saatavilla: <https://operations.nfl.com/gameday/technology/technology-and-the-game/> (18.4.2023)

Owens, M. F. & Roach, M. A. 2018. Decision-making on the hot seat and the short list: Evidence from college football fourth down decisions. *Journal of Economic Behavior & Organization*, 148, 301–314. Saatavilla: <https://doi.org/10.1016/j.jebo.2018.02.023> (4.4.2023)

- Pelechrinis, K. 2016. Decision Making in American Football: Evidence from 7 Years of NFL Data. Saatavilla: <https://www.semanticscholar.org/paper/Decision-Making-in-American-Football%3A-Evidence-from-Pelechrinis/804574339524ee52221edbdadbcf776f3da3fc38> (11.3.2023)
- Romer, D. 2006. Do Firms Maximize? Evidence from Professional Football. *The Journal of Political Economy*, 114(2), 340–365. Saatavilla: <https://doi.org/10.1086/501171> (17.4.2023)
- Roumani, Y. F. 2022. Sports analytics in the NFL: classifying the winner of the superbowl. *Annals of Operations Research*, 1–16. Saatavilla: <https://doi.org/10.1007/s10479-022-05063-x> (16.2.2023)
- Stolbunov, V. 2014. Introduction to Sport Analytics. University of Toronto Sports Analytics Student Group. Saatavilla: <https://sportsanalytics.sa.utoronto.ca/2014/12/11/introduction-to-sports-analytics/> (4.2.2023)
- Vinué, G. & Epifanio, I. 2017. Archetypoid analysis for sports analytics. Saatavilla: https://www.researchgate.net/publication/317336036_Archetypoid_analysis_for_sports_analytics (4.2.2023)
- Walder, S. 2022. NFL analytics survey 2022: Teams that use advanced metrics most, least. ESPN. Saatavilla: https://www.espn.com/nfl/story/_/id/35189929/2022-nfl-analytics-survey-most-least-analytically-inclined-teams-overrated-underrated-players-more (20.4.2023)
- Wang, P., Liu, J. & Liao, B. 2022. Prediction of Sports Performance and Analysis of Influencing Factors Based on Machine Learning and Big Data Statistics. *Journal of Sensors*, 2022, 1–9. Saatavilla: <https://doi.org/10.1155/2022/3276576> (13.4.2023)
- Williams, B., Palmquist, W. & Elmore, R. Simulation-based decision making in the NFL using NFLSimulatoR. *Ann Oper Res* 2022. Saatavilla: <https://doi-org.libproxy.tuni.fi/10.1007/s10479-022-04524-7> (9.3.2023)
- Yam, D. R. & Lopez, M. J. 2019. What was lost? A causal estimate of fourth down behavior in the National Football League. *Journal of Sports Analytics*, 5(3), 153–167. <https://doi.org/10.3233/JSA-190294>