MMORPG-pelaajan pelin lopettamisen ennustaminen koneoppimisella

Pro gradu -tutkielma
Turun yliopisto
Tietotekniikan laitos
Tietojenkäsittelytieteet
2021
Emma Lepistö

Tarkastajat:
Markus Viljanen
Tapio Pahikkala

TURUN YLIOPISTO
Tietotekniikan laitos

EMMA LEPISTÖ: MMORPG-pelaajan pelin lopettamisen ennustaminen koneoppimisella

Pro gradu -tutkielma, 78 s.
Tietojenkäsittelytieteet
2021

Massiiviset monen pelaajan verkkoroolipelit eli MMORPG-pelit (eng. Massively Multiplayer Online Role-playing Game) ovat suosittuja verkossa pelattavia pelejä, joiden tunnusmerkkejä ovat fantasiapainotteinen roolipelaaminen sekä jaetussa pelimaailmassa pelaaminen. Verkkopelaamisen harrastajamäärät ovat jatkuvassa kasvussa, ja suosituilla MMORPG-peleillä on miljoonia pelaajia. Peliyhtiöt kilpailevat pelaajien ajasta ja sitoutumisesta, ja ovat valmiita muokkaamaan peliä potentiaalisia pelaajia houkuttelevaksi.

Tässä tutkimuksessa ennustetaan koneoppimistekniikoita käyttämällä suositun MMORPG-pelin potentiaalisista pelaajista ne, jotka tulevat lopettamaan pelin pelaamisen tulevaisuudessa. Peliyhtiöille on tärkeää pystyä tunnistamaan pelaajia, joiden kiinnostus peliä kohtaan on laskemassa, jo ennen kuin pelaaja varsinaisesti lopettaa pelaamisen. Näin peliyhtiöt voivat pyrkiä pitämään pelaajaa pelin parissa tarjoamalla pelaajalle esimerkiksi houkuttimia tai helpotusta pelaamiseen. Lopettavien pelaajien tunnistaminen auttaa myös peliyhtiöitä pelin kehittämisessä ja peliyhtiöt voivat yrittää poistaa peleistään sellaisia ominaisuuksia, jotka nostavat pelaajien pelin lopettamisen todennäköisyyttä. Pelkkä tieto siitä, ketkä tulevat lopettamaan pelin pelaamisen, ei siis riitä. Peliyhtiöitä kiinnostaa myös se, millä tavalla pelin lopettavat pelaajat eroavat pelaajista, joiden motivaatio peliä kohtaan on säilynyt.

Tutkimuksen data on peräisin IEEE:n 2017 isännöimästä pelindatanlouhintakilpailusta, ja tutkimuksessa tutkitaan kilpailussa menestyneiden joukkueiden kilpailutöitä. Tutkimuksessa pyritään parantamaan Turun yliopiston (UTU) kilpailujoukkueen kilpailutyön ennustustarkkuutta lisäämällä malliin uusia piirteitä. Älykkäät piirteet vähentämät malliin tarvittavien piirteiden määrää. Tutkimuksessa tutkittiinkin yli sataa potentiaalista piirrettä, joista 40 valittiin uuteen malliin sovitettavaksi. Uusien piirteiden, sekä tiimi UTU:n mittaamien piirteiden, toimivuutta mitattiin usealla tekniikalla, joista parhaimman ristiinvalidointitarkkuuden saavuttivat harjanneluokittelija, lineaarinen tukivektorikone ja logistinen regressio. Testidatojen validoinnissa logistinen regressio onnistui parantamaan tiimin kilpailuratkaisua eniten.

Parhaiten menestyneessä mallissa oli vain yksitoista piirrettä, joista viisi oli uusia piirteitä ja kuusi sisältyi myös tiimi UTU:n kilpailutyöhön. Sekä tämän tutkimuksen, että varsinaisessa kilpailussa toiseksi päätyneen tiimi UTU:n ratkaisu, poikkeavat merkittävästi kilpailun voittajajoukkueen Yokozuna Data:n mallista. Voittajajoukkue käytti mallissaan jopa 500 piirrettä ja monimutkaisia tekniikoita, kuten syväoppimista ja satunnaistettuja päätöspuita. Koska molemmat lineaarista mallia käyttävät ratkaisut päätyivät melkein samaan tulokseen kuin voittajajoukkueen malli, tutkimuksesta käy ilmi, että juuri älykäs piirteiden valinta on avainasemassa MMORPG-pelin pelaajien lopettamisen ennustamisessa ja että lopettavat pelaajat voi ennustaa hyvin pienellä määrällä piirteitä.

Asiasanat: MMORPG, lineaarinen mallinnus, pelidatanlouhinta, IEEE GIC 2017

Sisältö

Luku 1

Johdanto

Massiiviset monen pelaajan verkkoroolipelit (engl. massive multiplayer online role-playing games, MMOPRG) ovat suosittuja ympäri maailmaa, mutta erityisesti Aasiasta löytyy paljon sekä peliyhtiöitä että pelaajia. Verkkopeleissä pelaajat pelaavat peliä yhteisessä peliympäristössä verkossa. Pelaajat kommunikoivat pelimaailmassa keskenään, ja voivat pelistä riippuen pelata joko toisiaan vastaan tai yhdessä peliä vastaan. MMORPG eroaa muista monen pelaajan verkkopeleistä sillä, että sen erikoisuutena on juuri fantasiapainotteinen roolipelaaminen.

Tässä tutkimuksessa käytetään NCsoft -nimisen [1] tunnetun Korealaisen peliyhtiön dataa. NCsoft on tuottanut useita suosittuja MMORPG-pelejä, joista suuren suosion ovat saavuttaneet ainakin Aion, Lineage, sekä tässä tutkimuksessa tutkittu Blade & Soul. NCsoft on merkittävä MMORPG-pelien tuottaja, ja yhtiöllä on satoja miljoonia asiakkaita.

Tämän tutkimuksen tutkimuskysymys on, kuinka tunnistaa koneoppimismenetelmiä käyttämällä pelaajat, jotka ovat aikeissa lopettaa Blade & Soul -pelin pelaamisen. Kysymys on tärkeä peliyhtiölle, sillä ensinnäkin tunnistamalla etukäteen pelaajat, jotka ovat aikeissa lopettaa pelin pelaamisen, peliyhtiö voi yrittää tarjota pelaajille apua tai houkuttimia pelaajan pitämiseksi pelin parissa. Toisaalta tunnistamalla sellaiset pelaajatyypit, jotka ovat

muita pelaajatyyppejä uskollisimpia pelaajia, peliyhtiö voi kohdentaa pelin mainontaa ja kehittää peliään vastaamaan uskollisimpien asiakkaiden toiveita. Vastaavasti tunnistamalla syitä miksi pelaajat lopettavat pelaamisen, peliyhtiö voivat yrittää kehittää peliään ja laajentaa sen yleisöä.

Tutkimuksessa käytetään tuhansien Blade & Soul -pelin pelaajien palvelinlokidataa. Data on siinä mielessä poikkeuksellista, että siitä on pyritty poistamaan sekä botit, että kaikista epäaktiivisimmat pelaajat, jolloin jäljelle on jäänyt vain potentiaalisia aitoja pelaajia [2]. Tutkimuskysymys on siis aavistuksen vaativampi, kuin tavanomainen pelin lopettamisen ennustaminen (eng. churn prediction), sillä tutkimuksessa tutkitaan kuinka löytää kaikista potentiaalisimpien MMORPG-pelin pelaajien joukosta ne pelaajat, jotka ovat lopettamassa pelin pelaamisen. Tutkimusdatan julkaisijoiden mukaan aktiivisien pelaajien joukosta onkin vaikeampaa ennustaa pelin lopettavat pelaajat, kuin epäaktiivisien pelaajien joukosta [2].

Toinen tutkimusta haastava seikka on se, että data on valmiiksi jaettu train- ja testisetteihin, ja tutkimustehtävänä on ennustaa testisettien lopettavat pelaajat train-datan avulla. Train-datasetti ja testisetit ovat kerätty eri ajanjaksoina, joten sekä kilpailutehtävänä että tutkimustehtävänä on löytää yleinen malli, joka toimii eri aikoina mitatun datan ennustamiseen. Myös pelin maksustrategia on muuttunut datan testisettien välillä kuukausittaisesta tilauspohjaisesta mallista maksuttomaksi markkinointimalliksi. Train-data on ensimmäisen testisetin tavoin mitattu ajanjaksona, jolloin pelaajat ovat maksaneet pelin pelaamisesta tilausperusteista maksua.[2]

Tutkimuksen motivaattorina toimii IEEE Game Data Mining competition 2017 -niminen kilpailu, johon peliyhtiö NCsoft on tarjonnut tutkimuksessa käytetyn pelidatansa. Tässä tutkimuksessa siis tutkitaan peliyhtiön julkiseen tutkimukseen tarjotun datan synnyttä-

miä tutkimustuloksia ja todistetaan, että yksinkertaisella regressiomallilla ja muutamalla älykkäästi valitulla piirteellä voidaan saavuttaa sama tulos, kuin kilpailun voittajajoukkue saavutti monimutkaisella mallilla ja suurella dataan perustuvalla piirrevalikoimalla.

Suuressa roolissa tutkimuksessa on myös Turun yliopiston kilpailujoukkue UTU:n kilpailuratkaisu, jota käsitellään luvussa 4.1. Tutkimuksessa pyritään jäljentämään kilpailun olosuhteita ja kilpailutehtävän sääntöjä mahdollisimman tarkasti, jotta ne vastaisivat puitteiltaan tiimi UTU:n ja muiden kilpailuun osallistuneiden joukkueiden ratkaisuja. Tarkoituksena on parantaa tiimi UTU:n ratkaisua sekä hiomalla mallin piirteitä että pienentämällä mallin piirteiden määrää toimivammin rakennettujen piirteiden avulla. Parantamalla tiimi UTU:n tulosta piirteiden valinnalla, tämä tutkimus todistaa, että piirteiden valinta on avainasemassa pelin pelaajien pelimotivaation ennustamisessa.

MMORPG-peleistä on tehty melko vähän tutkimusta siihen nähden kuinka suosittuja ne ovat, kuinka paljon peleissä liikkuu rahaa ja kuinka paljon pelityylillä on harrastajia. Yksi ongelma on tutkimukseen käytettävissä olevan datan niukkuus, sillä peliyhtiöt ovat taloudelliseen voittoon pyrkiviä yhtiöitä, ja datan julkaisemisessa on aina omat riskinsä. Tällaisia riskejä voivat olla esimerkiksi se, että kilpailijat hyötyvät datasta, tai että datan julkaiseminen helpottaa vilpillistä pelaamista. Datan julkaisemin julkiseen tutkimukseen ei välttämättä ole peliyhtiöille tarpeeksi kannattavaa riskeihin nähden. Voi myös olla, että julkinen tutkimus ei tarjoa peliyhtiöille hyödyllisiä ratkaisuja, vaan tutkimuksessa käytetään esimerkiksi liian monimutkaisia malleja, joilla ei välttämättä ole yhtymäkohtaa todellisten ratkaisujen kanssa.

Ongelma on siis usein se, että tutkimuslaitokset, datatiedettä tekevät yhtiöt ja instituutiot tarjoavat tutkimuksissaan pelimuodolle toinen toistaan monimutkaisempi malleja. Osa tutkimuksista, kuten esimerkiksi kilpailun voittajajoukkueen tutkimus, käyttävät mo-

nimutkaisia tekniikoita ja valtavaa määrää statistisia piirteitä. Tämä tutkimus kuitenkin osoittaa, että yksinkertaisilla malleilla voi saavuttaa hyvin samanlaisia tuloksia. Toisin sanoen syväoppiminen ja muut monimutkaiset mallit eivät välttämättä ole perusteltuja tämän tutkimuksen kaltaiseen tutkimuskohteeseen.

Koska peliyhtiöt ovat itse suunnitelleet pelinsä, ja jopa osittain valinneet pelaajansa kohdentamalla mainontaa tai valitsemalla missä maissa tai maanosissa peli julkaistaan, peliyhtiöt omistavat valtavasti tietoa pelistä ja pelaajista. Näin ollen peliyhtiöt pystyvät olemassa olevan tiedon pohjalta valitsemaan pelistä ja pelaajista sellaisia piirteitä, jotka mittaavat pelaajan innostusta peliä kohtaan. Tällaista tietoon tai pelimotivaatioteoriaan pohjautuvaa datan louhintaa on tutkittu jonkin verran MMORPG-pelien yhteydessä. Esimerkiksi Borbora et al. tulivat 2011 julkaisemassa tutkimuksessa tulokseen, että muutamat pelimotivaatioteorian mukaan koostetut piirteet antoivat yhdessä vain aavistuksen heikomman tarkkuuden, kuin lukuisat datasta suoraan valitut piirteet [3].

Tässä tutkimuksessa tullaan hyvin samalaiseen tulokseen Borbora et al. tutkimuksen kanssa, eli vaikuttaisi siltä, että piirteiden valitseminen on tärkeää MMORPG-pelien pelaajien pelin lopettamisen ennustamisessa. Tässä tutkimuksessa saavutettiin muutamalla hyvin rakennetulla piirteellä sama tai miltei sama tarkkuus, kuin voittajajoukkue saavutti 500:lla statistisesti valitulla piirteellä. Tietoon perustuvien piirteiden käyttö on järkevää, sillä peliyhtiöillä on ennalta tietoa pelaajista ja pelistä. Puhtaasti lokidataan perustuva datanlouhinta voisi olla kannattavaa siinä tapauksessa, jos pelistä ei ole saatavilla tarpeeksi tietoa.

Yksinkertaisilla tekniikoilla on lukuisia hyviä puolia monimutkaisempiin tekniikoihin verrattuna. Yksinkertaisemmat tekniikat mahdollistavat tutkimuksen tekemisen pienemmillä resursseilla, sekä säästävät aikaa ja rahaa. Joskus yksinkertaiset mallit voivat myös tarjota monimutkaisia malleja paremman selityksen sille, miksi pelaajat lopettavat pe-

lin pelaamisen. Satoja piirteitä sisältävät monimutkaiset mallit toimivat mustan laatikon tavoin, eivätkä itse ennusteen lisäksi selitä mitkä muuttujat ennustavat pelaajien sitoutumista peliin. Toisaalta monimutkaiset mallit pystyvät mallintamaan yksinkertaisia malleja monimutkaisempia ilmiöitä, ja saattavat löytää datasta sellaisia piirteitä, joiden olemassaoloa ei ole tiedetty, tai joita ei ole osattu etsiä.

Tässä tutkimuksessa tutkitaan myös sitä, miten ilmaisten pelien datanlouhinta poikkeaa maksullisten pelien louhinnasta, ja voiko maksullisille ja ilmaisille peleille löytää toimivan yhteisen mallin. Pelaajille ilmaiset pelit ovat kaupallisessa mielessä hyvin erityinen peliryhmä. Ilmaisia pelejä kutsutaan lyhenteellä F2P, joka tulee englanninkielisestä termistä "free-to-play". F2P-pelien tuotto koostuu esimerkiksi mainostuloista, tai ostoksista, joita pelaajat voivat vapaaehtoisesti tehdä pelissä [4].

Ilmaisissa peleissä erityisesti aloittavan pelaajan eliniän ennustaminen on poikkeuksellista, sillä uusien pelaajien kynnys kokeilla ilmaista peliä on huomattavasti pienempi kuin maksullista peliä. Näin ollen F2P-pelin saattaa usein aloittaa sellainen pelaaja, joka ei lähtökohtaisestikaan ole potentiaalinen asiakas. Pelaaja ei esimerkiksi tykkää MMOPRG-peleistä tai hänellä ei ole riittävää nettiyhteyttä tai sopivaa tietokonetta pelin pelaamiselle. Toisaalta F2P-pelit ovat myös alttiimpia vilpilliselle pelaamiselle, bottien käyttämiselle sekä taloudelliseen voittoon pyrkivälle pelaamiselle, sillä pelitilien tekeminen on ilmaista. Tutkimuksen data on arvokas ilmaisten pelien datan louhinnalle, sillä se tarjoaa mahdollisuuden vertailla miten hinnoittelustrategiat vaikuttavat datan louhintaan saman pelin datalla.

Luku 2

Kirjallisuus

Verkkopeleistä on suhteellisen helppo saada dataa ja peliyhtiöillä on hyvin resursseja tutkimuksen tekemiseen. Siitä huolimatta pelit eivät ole olleet erityisen paljon julkisen tutkimuksen kohteena. Tämä voi johtua siitä, että pelimaailmaa ei pidetä niin elintärkeäni tutkimuksen kohteena kuin monia muita aihepiirejä. Toisaalta ne tahot, jotka hyötyvät eniten pelien tietolouhinnasta, ovat peliyhtiö itse, ja näin ollen suurin osa tutkimuksesta tehdään suljetusti peliyhtiöiden sisällä. Pelidatan julkaiseminen julkiseen tutkimukseen voi olla suuri riski peliyhtiöille, sillä kilpailijoiden lisäksi datasta voivat hyötyä esimerkiksi botteja tekevät pelaajat.

MMORPG-pelit ovat saaneet alkunsa jo 90-luvun loppupuolella, mutta tutkimusta on pääsääntöisesti tehty huomattavasti myöhemmin. Suurin osa MMORPG-pelien tutkimuksesta sijoittuu 2010-luvulle, kuten Google Scholar hakusanatutkimuksesta voi päätellä (kuva 2.1). Kuvasta huomaa, että suurin nousu MMORPG-hakusanan sisältävissä artikkeleissa on kuitenkin tapahtunut jo 2000–2009. Tämän jälkeen artikkeleiden määrä on pysynyt vakiona.

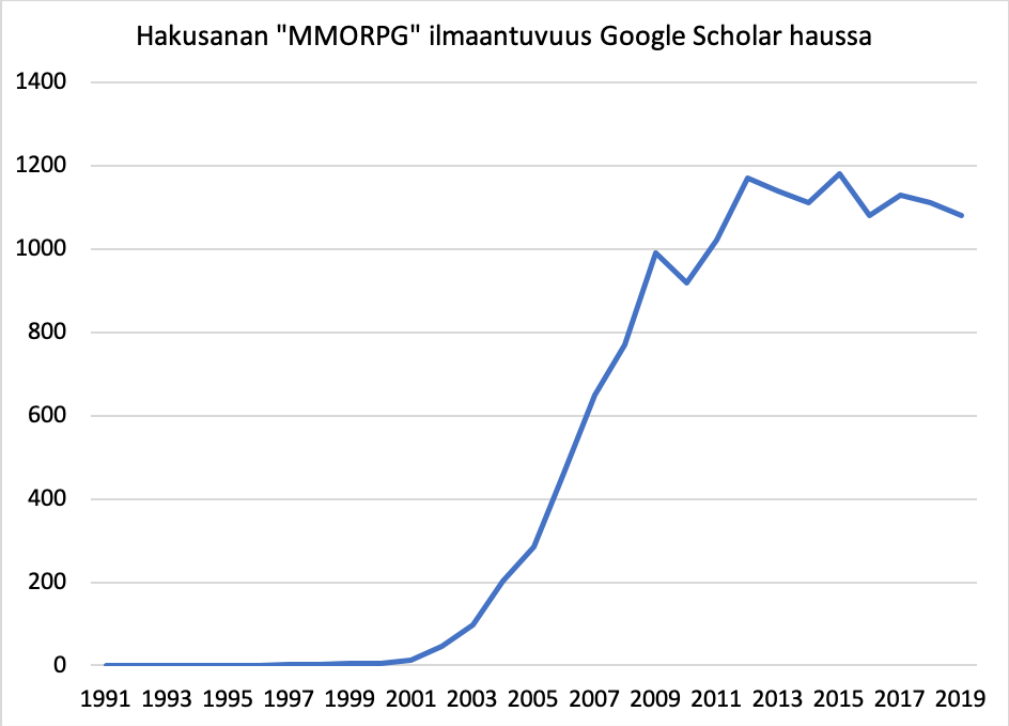
Hyvä esimerkki 2000-luvun alun MMORPG-pelin datan louhintaa käsitelleet tutkimuksesta on Larry Shy ja Weiyun Huang 2004 julkaisema tutkimus "Apply Social Network

Tekijät	Julkaisun nimi	Vuosi	Peli	Lähde
W. C. Feng, D. Brandt ja D. Saha	A long-term study of a popular MMORPG	2007	EVE Online	[5]
J. Kawale, A. Pal ja J. Srivastava	Churn Prediction in MMORPGs: A Social Influence Based Approach	2009	EverQuest II	[6]
Z. Borbora, J. Srivastava, K. W. Hsu ja D. Williams	Churn Prediction in MMORPGs using player motivation theories and an ensemble approach	2011	EverQuest II	[3]
Z. Borbora, J. Srivastava	User Behavior Modelling Approach for Churn Prediction in Online Games	2012	EverQuest II	[7]
L. Martins Kummer and J. Cesar Nievola and E. Paraiso	Applying Commitment to Churn and Remaining Players Lifetime Prediction	2018	Blade & Souls	[8]
E. Lee, B. Kim, S. Kang, B. Kang, Y. Jang and H. K. Kim	Profit Optimizing Churn Prediction for Long-Term Loyal Customers in Online Games	2020	Aion	[9]

Taulukko 2.1: MMORPG-pelien pelaajien selviytymistä ja pelaajauskollisuudesta tehdyt tutkimukset.

Analysis and Data Mining to Dynamic Task Synthesis for Persistent MMORPG Virtual World"[10], jossa tutkittiin, miten pelaajan pitkäkestoista sitoutumista peliin voitaisiin kasvattaa datanlouhintatekniikkoja hyödyntämällä.

Pelidatanlouhinta on mainittu kirjallisuudessa ensimmäisen kerran 2013 teoksessa "Game Analytics"[11]. Pelidatanlouhintaan lukeutuu useita erilaisia tutkimuskohteita, kuten esimerkiksi pelaajien pelissä käyttämä aika, pelaajien uskollisuus pelille, pelaajien käyttäytyminen pelissä, pelin sosiaaliset aspektit sekä pelien heikkoudet [11]. Kyseisiä aiheita on kuitenkin tutkittu jo kauan ennen kuin pelidatanlouhinta on terminä esitelty. Pelaamisen lopettamisen ennustamista (eng. Churn prediction) pidetään luonnollisesti kiinnostavana tutkimuskohteena peliyhtiöissä. Tunnistamalla etukäteen lähtöuhan alla olevat pelaajat peliyhtiöillä on vielä mahdollisuus tarjota pelaajalle houkuttimia jotka estävät lähdön.



Kuva 2.1: Hakusanan "MMORPG"tulokset Google Scholar haussa. Haku on tehty vain englanninkielisistä julkaisuista, eikä sisällä patentteja tai sitaatteja.

Taulukossa 2.1 on lueteltu sellaisia MMORPG-pelien pelaajista tehtyjä tutkimuksia, joissa tutkitaan ja ennustetaan jatkavatko pelaajat pelaamista. Feng et al. julkaisivat 2007 pitkäaikaistutkimuksen EVE Online –nimisen MMORPG-pelin pelaajista. Tutkimuksen data sisälsi jopa 100 000 pelaajaa, ja keräysjakso kesti vuoden 2003 toukokuusta aina vuoden 2006 maaliskuuhun asti. Tutkimuksen mukaan pelaajan peliaikahistoria paljastaa pelaajan kiinnostuksen vähenemisen, ja esimerkiksi ajan myötä lyhenevät pelisessioiden pituudet ennustavat pelaajan lopettamista [5].

Tutkijat J. Srivasta ja Z. Borbora ovat julkaisseet erilaisien tutkimusryhmien kanssa useita tutkimuksia EverQuest II -nimisestä pelistä vuosien 2009 ja 2012 välillä. Srivasta julkaisi 2009 "Churn Prediction in MMORPGs: A Social Influence Based Approach"–nimisen tutkimuksen yhdessä Jaya Kawale ja Aditya Paul kanssa. Tutkimusryhmä nimisesi kaksi tutkimaansa tekijää pelin lopettamisen ennustamiseen: pelaajan sitoutuminen ja pelaajan

sosiaalinen verkosto. Pelaajan sitoutumista mitattiin pelaajan pelissä viettämästä ajasta eli pelisessioista ja sessioiden pituudesta. Sosiaalinen verkosto taas on koostettu pelaajista ja heidän keskinäisestä vuorovaikutuksesta, joka tutkimuksessa tarkoittaa pelaamista yhdessä. Verkostoon on lisätty side jokaisen sellaisen kahden pelaajan välille, jotka ovat pelanneet yhdessä, ja siteelle on annettu sitä enemmän painoarvoa mitä enemmän pelaajat ovat saavuttaneet yhdessä pisteitä pelissä. [6]

Borbora ja Srivasta julkaisivat samasta pelistä uuden tutkimuksen 2011 Hsun ja Williamsin kanssa. Tällä kertaa tutkimuksessa tutkittiin pelimotivaatioteorian (eng. player motivation theory) sopeutumista pelin lopettamisen ennustamiseen. Pelimotivaatioteoria on Nick Yeen vuonna 2007 julkaisemassa tutkimuksessa "Motivations for Play in Online Games"syntynyt teoria pelimotivaatiotekijöistä [12]. Borboran et al. tutkimuksessa tutkittiin pelimotivaatioteorian lisäksi sitä miten teoriapohjaiset ja datapohjaiset piirteet suoriutuvat tutkimuksessa. Tutkimusryhmä tuli siihen lopputulokseen, että datapohjaisia piirteitä käyttävä tutkimus oli vain hieman tarkempi ennustamaan pelin lopettamista, kuin teoriapohjaisia piirteitä käyttävä tukimus, ja että teoriapohjainen tutkimus tarjoaa peliasiantuntijoille todennäköisesti tulkittavampia piirteitä pelin lopettamisesta kuin datapohjainen.[3]

Kummerin et al. julkaisema tutkimus "Applying Commitment to Churn and Remaining Players Lifetime Prediction"on tämän tutkimuksen valossa erityisen kiinnostava lähde, sillä tutkimus on tehty samasta datasta ja samoilla kilpailutehtävillä kuin kilpailuun osallistuneet työt. Tutkimustiimi onnistui löytämään kummastakin kilpailun kilpailutehtävästä ratkaisun, joka näyttäisi olevan parempi kuin yksikään kilpailuun osallistunut työ. Tutkimusryhmän saavuttamat tulokset eivät kuitenkaan ole linjassa kilpailuun osallistuneisiin töihin nähden. Kummer et al. käyttivät piirteinä 54:ää pelitoimintoa, pelitoiminnon lokiviestityypistä laskettua keskiarvoa sekä lokiviestien viikoittaisesta ilmaantuvuudesta laskettua tendenssiä. Pelitoimintojen lisäksi tiimin mallissa oli kahdeksan muuta piirrettä,

joihin lukeutui muun muassa sessioiden pituus, pelaajan taso (eng. level) ja pelattujen päivien määrä. Kaiken kaikkiaan tiimillä oli kummassakin kilpailutehtävässä 122 piirrettä. Tutkimuksessa tutkittiin muutamia erilaisia koneoppimistekniikoita, mutta tutkimustiimi päätyi molemmissa kilpailutehtävissä päätöspuumallin variaatioon.[8]

Hadiji et al. 2014 julkaisema tutkimus "Predicting Player Churn in the Wild"on merkityksellinen tutkimus ilmaisten pelien pelaajien elinaika-analyysistä. Tutkimuksessa käytettiin vain yhtätoista erilaista piirrettä, joista suurin osa liittyi peliaikaan liittyviin ominaisuuksiin, kuten pelisession kestoon tai pelattujen päivien määrään. Hadji et al. tulivat tutkimuksessa johtopäätökseen, että juuri ajalliset määreet olivat piirteistä merkityksellisimpiä [13].

On myös huomattava, että F2P-peleistä on tehty huomattavan vähän datanlouhintaa, ja vaikka tutkittava kohde olisikin F2P-peli, ilmaisuutta ei olla nähty erityisen merkitsevänä ominaisuutena. Tuoreempi 2020 Rothmeierin et al. julkaisema tutkimus "Prediction of Player Churn and Disengagement Based on User Activity Data of a Freemium Online Strategy Game"keskittyi ajallisiin aktiivisuutta mittaaviin ominaisuuksiin [14]. Tutkimusryhmä onnistui ennustamaan yllättävän tarkasti onko pelaaja palaamassa takaisin peliin kahden viikon aikana satunnaistettu metsä –nimisen (eng. Random Forest) päätöspuutekniikan avulla [14].

F2P-pelejä on kuitenkin tutkittu muulla tapaa, joka hyödyttää tutkimusta. Esimerkiksi Holin Lin and Chuen-Tsai Sun ovat tuoneet 2011 tehdyssä tutkimuksessaan "Cash Trade in Free-to-Play Online Games"esiin pelikokemuksellisia eroavaisuuksia, joita F2P-peleillä usein on muihin peleihin verrattuna [4]. Artikkeli ei kosketa datanlouhintaa, mutta se antaa viitteitä siitä, että F2P-peleissä on pelikokemuksellista eroa muihin pelityyppeihin verrattuna.

Luku 3

Data ja sen alkuperä

Tutkimuksessa käytetyn datan on kerännyt peliyhtiö NCsoft ja yhtiö on tarjonnut datan IEEE Game Data Mining competition 2017 -nimisen kilpailun kilpailudataksi. Data on peräisin Blade & Soul -nimisestä MMORPG-pelistä, ja koostuu 10 000:nnen pelaajan pelilokitiedoista. Pelaajan data on kerätty palvelimelle kirjatuista lokitiedoista, ja näin ollen data on hyvin vähän prosessoitua ja sisältää valtavan määrän tietoa pelaajan suorittamista pelitoiminnoista. Pelilokiin on kirjattu jokaisen pelitoiminnon aikaleima ja toiminnon tunnistuskoodi, mutta myös muuta lisäinformaatiota, kuten pelihahmon tunnus, jolla toiminto on suoritettu, pelisession, pelialueen ja palvelin tunnus, sekä mahdollinen joukkue, johon pelaaja on toiminnon tehdessä kuulunut.[2]

Kilpailun tehtävänä oli ennustaa koneoppimistekniikoita käyttämällä pelaajat, jotka tulevat lopettamaan pelin pelaamisen ja pelaajat, jotka tulevat jatkamaan pelaamista. Kilpailu koostui kahdesta eri kilpailutehtävästä, joista ensimmäisen tehtävänä oli ennustaa, ketkä pelaajista tulevat lopettamaan pelin pelaamisen, ja toisena kilpailutehtävänä oli ennustaa, kuinka monta päivää samaiset pelaajat tulevat viettämään pelissä ennalta määrätyn mittausajanjakson jälkeen. Kilpailijat palauttivat ratkaisunsa joko molempiin tai vain toiseen kilpailukysymykseen.

seq	time	logid	session	actor_code	actor_id	actor_account_id	actor_zone_id	actor_zone_channel_id	actor_party_id	
69462583	20.4.2016 23:14	1003	4295000751	10	19411560	00C172F0	0	0	0	...
69553202	20.4.2016 23:15	1005	4295000751	10	19411560	00C172F0	104	1	0	...
69658177	20.4.2016 23:16	1022	4295000751	10	19411560	00C172F0	104	1	0	...
69685085	20.4.2016 23:16	1022	4295000751	10	19411560	00C172F0	104	1	0	...
69832217	20.4.2016 23:17	1022	4295000751	10	19411560	00C172F0	104	1	0	...
69832218	20.4.2016 23:17	1022	4295000751	10	19411560	00C172F0	104	1	0	...
69832219	20.4.2016 23:17	1022	4295000751	10	19411560	00C172F0	104	1	0	...
69841258	20.4.2016 23:17	1023	4295000751	10	19411560	00C172F0	104	1	0	...
69841259	20.4.2016 23:17	2109	4295000751	10	19411560	00C172F0	104	1	0	...

Kuva 3.1: Esimerkkikuva pelaajan palvelinlokidatasta. Logid on tässä tutkimuksessa nimetty lokiviestitoiminnoksi.

Kilpailun voitti joukkue, jonka rakentama malli onnistui tarjoamaan parhaan ennustustarkkuuden kilpailukysymyksiin. Kilpailu oli avoin kaikille halukkaille osanottajille, ja näin ollen myös kilpailuun tarjottu data oli avoimesti kaikkien saatavilla. Osallistuneet tiimit olivat sekä yksittäisiä kilpailijoita, että useasta jäsenestä koostuvia joukkueita. Kilpailuun osallistui paljon akateemisia tutkijoita, sekä datatiedettä tekeviä yrityksiä ja instituutioita. [2]

3.1 Peli

Blade & Soul on NCsoft:n valmistama MMORPG-peli ja peli on julkaistu vuonna 2012 [15]. Pelin juoni sijoittuu kuvitteelliseen fantasiamaailmaan, joka on jakautunut kahteen suureen allianssiin eli liittoumaan. Pelaaja saa itse valita liittyykö Ceruelin kiltaan (eng. The Cerulean Order) vai Crimsonin legioonaan (eng. The Crimson Legion) [16]. Liittoutumien lisäksi pelimaailma on jakautunut pelialueisiin, joista jokainen sisältää vielä lukuisia pienempiä pelialueita, kuten luolastoja ja kaupunkeja. Kaikki alueet eivät ole heti pelaajan käytössä, vaan avautuvat pelaajalle juonen edetessä.

Juonen lisäksi taistelulajit ovat tärkeä osa peliä. Pelaaja voi taistella pelissä sekä pelimaailmaa eli tietokonetta vastaan (eng. player versus environment, PvE), että oikeita pelaajia vastaan (eng. player versus player, PvP). Fantasiaroolipeleille ominaisesti pelaaja voi kustomoida pelihahmoaan hyvin vapaasti: pelaaja voi valita hahmonsa neljän rodun

ja yhdeksän luokan väliltä, sekä valita hahmolleen sukupuolen, ulkonäön ja vaatetuksen [17]. Mahdollisia kombinaatioita on tuhansia. Myös itse peli ja pelimaailma ovat pelaajalle hyvin vapaita, sillä pelaaja voi joko keskittyä pelin juonen suorittamiseen, ratkaista päivittäisiä ja viikoittaisia haasteita, haastaa pelaajia PvP-taisteluihin areenassa tai tutkia pelimaailmaa. Pelissä on myös mahdollista kommunikoida muiden pelaajien kanssa viestittelemällä pelin keskustelupalstalla tai pelata muiden pelaajien kanssa muodostamalla tiimejä, joukkueita tai liittymällä kiltoihin. [18]

Kuva 3.2: Kuvankaappaus pelistä

3.2 Kilpailu

Kilpailun sääntöjen mukaan kilpailun ratkaisu piti saavuttaa koneoppimistekniikalla ja ensimmäisen tehtävän mallien suoritus piti arvioida F1:sen kanssa (1) ja toisen tehtävän mallit RMSLE:llä (eng. Root Mean Squared Log Error)(2) [2]. Tiimin lopullinen tarkkuus laskettiin kummankin testisetin saavuttamien tarkkuuksien keskiarvona. Datan testisettien

tulos tuli ennustaa train-datasetin pelaajien avulla. Kilpailijoilla ei ollut käytössä testisettien labeleita, vaan ne julkaistiin vasta kilpailun jälkeen.

$$F1 = 2 \cdot \frac{1}{\frac{1}{recall} + \frac{1}{precision}} = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{3.1}$$

$$\epsilon = \sqrt{\frac{1}{2}\sum_{i=1}^{n}(\log(p_i + 1)) - (\log(a_i + 1))^2} \tag{3.2}$$

Kummassakin kilpailun tehtävässä parhaimman tarkkuuden saavutti tiimi Yokozuna Data (YD) Silicon Studiosta Japanista. Tiimi käytti ratkaisussaan sekä päätöspuita ratkaisun ennustamiseen, (eng. Extra Trees Classifier) että syväoppimista piirteiden valintaan. Kilpailussa tuli toiseksi tiimi UTU Turun yliopistosta (ensimmäisestä kilpailutehtävästä toinen sija ja toisesta kilpailutehtävästä kolmas sija) ja merkille pantavaa on se, että tiimi käytti ratkaisuissaan hyvin yksinkertaisia malleja, kuten lineaarista ja logistista regressiomallia. Ensimmäisen kierroksen kolmannen sijan saavutti tiimi TripleS, ja toisen kierroksen toisen sijan tiimi IISLABSKKU. Kummatkin tiimeistä käyttivät kilpailuratkaisussaan päätöspuita.

Kuten taulukosta 3.1 voi huomata, ensimmäisen kilpailutehtävän paras ennustustarkkuus oli 0,63 (tiimi YD, testisetti 2). Viiden parhaimman joukkueen tulokset ovat hyvin lähellä toisiaan, ja kuudennesta joukkueesta eteenpäin tulokset ovat alle 0,5. Kilpailussa on merkittävää se, että tulokset ovat hyvin lähellä toisiaan siitäkin huolimatta, että tiimit käyttivät erilaisia malleja, piirteitä ja piirteiden määrää.

Tiimi	**Testisetti 1 F1-tulos**	**Testisetti 2 F1-tulos**	**Malli**
YD	0.61	0.63	Syväoppiminen, päätöspuut
UTU	0.60	0.60	Regressio
TripleS	0.57	0.62	Päätöspuut
TheCowKing	0.59	0.60	Päätöspuut
goedle.io	0.57	0.60	Syväoppiminen, päätöspuut

Taulukko 3.1: Kilpailun ykköstehtävän viisi parasta tulosta.

Tiimi	**Testisetti 1 tulos**	**Testisetti 2 tulos**	**Malli**
YD	0.88	0.61	Päätöspuut
IISLABSKKU	1.03	0.67	Päätöspuut
UTU	0.92	0.89	Regressio
TripleS	0.95	0.89	Päätöspuut
TeDTND	1.03	0.93	?

Taulukko 3.2: Kilpailun kakkostehtävän kaikki kilpailutulokset

Kilpailun datan tarjonnut NCsoft julkaisi kilpailun jälkeen oman tutkimuksensa kilpailun tuloksista. EunJon et al. kirjoittama julkaisu "Game Data Mining Competition on Churn Prediction and Survival Analysis using Commercial Game Log Data"käsittelee muun muassa dataa, kilpailun tuloksia ja yksittäisten osallistujajoukkueiden ratkaisuja [2]. Tämän lisäksi kilpailun voittajajoukkue YD julkaisi oman ratkaisunsa kilpailun jälkeen [19].

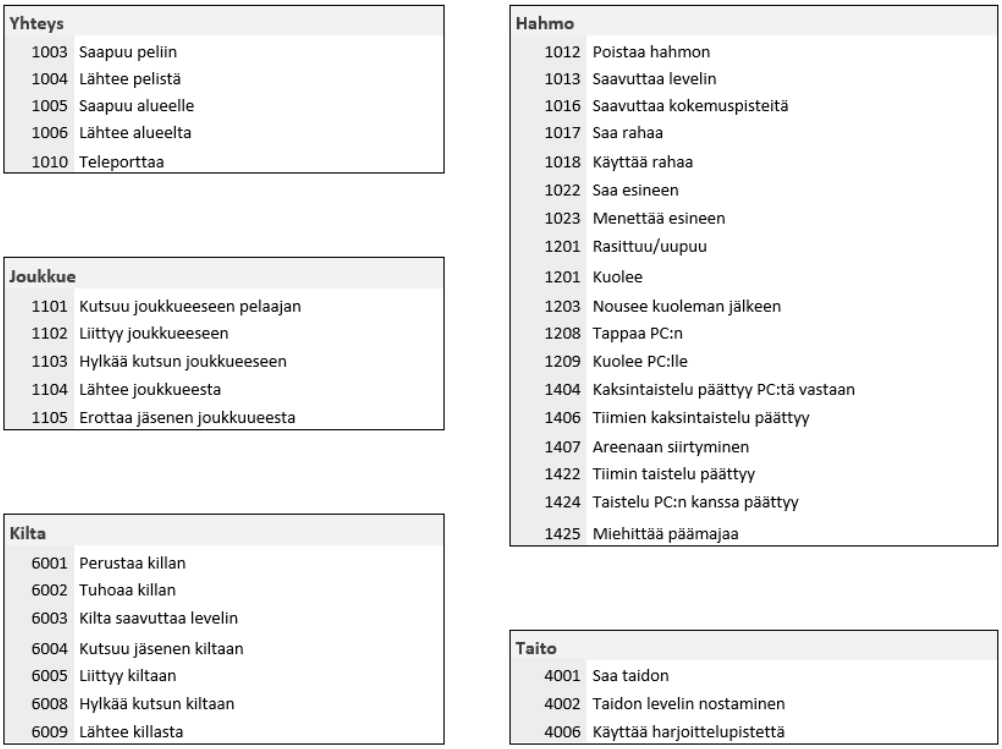
3.3 Data

Pelidata on kaupallista lokidataa ja koostuu pelaajista, jotka ovat lojaaleja tai aktiivisista pelaajia, eivätkä osoita merkkejä vilpillisestä pelaamisesta. Pelistä on mitattu vuoden sisällä kolmena ajankohtana dataa ja datasettien keräys on kestänyt 6-8 viikkoa. Keräysjakson jälkeen datan labelit $churn\_yn$ (boolean) ja $survival\_time$ (integer) on kerätty noin kolmen viikkoa keräysajan loppumisen jälkeen.

Ajanjaksoa, jonka aikana dataa ei ole kerätty, kutsutaan tässä tutkimuksessa mittaustauoksi. Mittaustauko on kolmen viikon pituinen ja sen tarkoitus on toimia rajanvetona siihen, ketkä pelaajista ovat lopettaneet pelaamisen. Kaikki pelaajat, jotka ovat jatkaneet pelaamista seuraavan viiden viikon aikana mittaustauon jälkeen, ovat merkittyinä jatkaneiksi pelaajaksi. Jäljitysajalta mitataan sitä, kuinka monen päivän päästä mittaustauosta pelaaja on vielä pelannut peliä. Jokaisen pelaajan palvelinlokidata on pakattu omaan CSV-tiedostoonsa. [2]

Kilpailun ensimmäisessä tehtävässä kilpailijoiden tuli mitataan sitä, kuinka tarkasti keräysajan datasta voidaan ennustaa pelaajat, jotka ovat pelanneet peliä jäljitysajan aikana. Kaikille pelaajille, jotka ovat pelanneet peliä jäljitysaikana, on $churn\_yn$ merkitty arvolla "0". Vastaavasti pelaajille, jotka eivät ole pelanneet peliä jäljitysaikana, $churn\_yn$ on merkitty arvolla "1", eli heidän katsotaan lopettaneen pelaamisen.

Kilpailun toisessa tehtävässä kilpailijoiden tuli mitataan sitä, kuinka monena päivänä pelaaja pelaa mittaustauon jälkeen. Pelaajille on merkitty pelattujen päivien määräksi eli $survival\_time$:ksi se määrä päiviä, joka on kulunut jäljitysajan alkamisesta siihen päivään, kun käyttäjä on viimeisen kerran pelannut peliä. Jokaisen pelaajan pelattujen päivien määrä on näin ollen väliltä 0–320+ päivää.



Yhteys	
1003	Saapuu peliin
1004	Lähtee pelistä
1005	Saapuu alueelle
1006	Lähtee alueelta
1010	Teleporttaa

Joukkue	
1101	Kutsuu joukkueeseen pelaajan
1102	Liittyy joukkueeseen
1103	Hylkää kutsun joukkueeseen
1104	Lähtee joukkueesta
1105	Erottaa jäsenen joukkuueesta

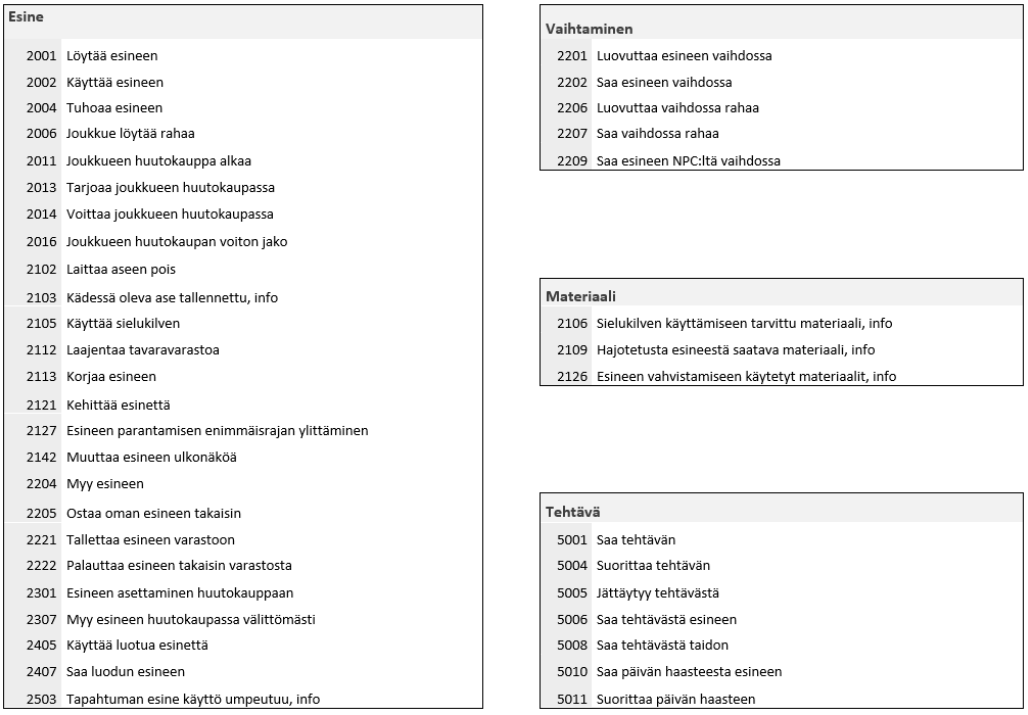
Kilta	
6001	Perustaa killan
6002	Tuhoaa killan
6003	Kilta saavuttaa levelin
6004	Kutsuu jäsenen kiltaan
6005	Liittyy kiltaan
6008	Hylkää kutsun kiltaan
6009	Lähtee killasta

Hahmo	
1012	Poistaa hahmon
1013	Saavuttaa levelin
1016	Saavuttaa kokemuspisteitä
1017	Saa rahaa
1018	Käyttää rahaa
1022	Saa esineen
1023	Menettää esineen
1201	Rasittuu/uupuu
1201	Kuolee
1203	Nousee kuoleman jälkeen
1208	Tappaa PC:n
1209	Kuolee PC:lle
1404	Kaksintaistelu päättyy PC:tä vastaan
1406	Tiimien kaksintaistelu päättyy
1407	Areenaan siirtyminen
1422	Tiimin taistelu päättyy
1424	Taistelu PC:n kanssa päättyy
1425	Miehittää päämajaa

Taito	
4001	Saa taidon
4002	Taidon levelin nostaminen
4006	Käyttää harjoittelupistettä

Kuva 3.3: Peliyhteyteen, pelihahmoon, joukkueisiin, kiltoihin ja pelihahmon taitoihin liittyvät lokiviestit ja niiden tunnistuskoodit palvelinlokidatasta.

Kilpailun data on valmiiksi jaettu testi- ja train-datasetteihin. Molemmat testisetit sisältävät 3000:nnen pelaajan palvelinlokidatan, kun train-data taas on koottu 4000:nnen pelaajan datasta. Jokainen datasetti on mitattu eri ajankohtana vuosien 2016-2017 aikana. Testisettien välissä on tapahtunut myös merkittävä maksustrategiamuutos: ensimmäisen testisetin aikana peli on ollut maksullinen ja perustunut kuukausittaisiin maksuihin, kun taas toisen testisetin aikana peli on ollut ilmainen. Train-data on mitattu ennen kumpaakaan testisettiä, eli se on myös kerätty aikana jolloin peli on ollut pelaajille maksullinen. On myös huomionarvoista, että pelin lopettaneiden määrä on huomattavasti pienempi kuin peliä jatkaneiden pelaajien määrä. Kaikkien datasettien kohdalla 30 % pelaajista on lopettanut pelin.[2]



Esine	
2001	Löytää esineen
2002	Käyttää esineen
2004	Tuhoaa esineen
2006	Joukkue löytää rahaa
2011	Joukkueen huutokauppa alkaa
2013	Tarjoaa joukkueen huutokaupassa
2014	Voittaa joukkueen huutokaupassa
2016	Joukkueen huutokaupan voiton jako
2102	Laittaa aseen pois
2103	Kädessä oleva ase tallennettu, info
2105	Käyttää sielukilven
2112	Laajentaa tavaravarastoa
2113	Korjaa esineen
2121	Kehittää esinettä
2127	Esineen parantamisen enimmäisrajan ylittäminen
2142	Muuttaa esineen ulkonäköä
2204	Myy esineen
2205	Ostaa oman esineen takaisin
2221	Tallettaa esineen varastoon
2222	Palauttaa esineen takaisin varastosta
2301	Esineen asettaminen huutokauppaan
2307	Myy esineen huutokaupassa välittömästi
2405	Käyttää luotua esinettä
2407	Saa luodun esineen
2503	Tapahtuman esine käyttö umpeutuu, info

Vaihtaminen	
2201	Luovuttaa esineen vaihdossa
2202	Saa esineen vaihdossa
2206	Luovuttaa vaihdossa rahaa
2207	Saa vaihdossa rahaa
2209	Saa esineen NPC:ltä vaihdossa

Materiaali	
2106	Sielukilven käyttämiseen tarvittu materiaali, info
2109	Hajotetusta esineestä saatava materiaali, info
2126	Esineen vahvistamiseen käytetyt materiaalit, info

Tehtävä	
5001	Saa tehtävän
5004	Suorittaa tehtävän
5005	Jättäytyy tehtävästä
5006	Saa tehtävästä esineen
5008	Saa tehtävästä taidon
5010	Saa päivän haasteesta esineen
5011	Suorittaa päivän haasteen

Kuva 3.4: Esineisiin, materiaaleihin, vaihtamiseen sekä pelin tehtäviin liityvät lokiviestit ja niiden tunnistuskoodit palvelinlokidatasta.

Pelin data sisältää suuren määrän erilaisia mitattavia pelitoimintoja, kuten esimerkiksi pelirahan määrän, kuinka monta kertaa pelihahmo kuolee joko ympäristön tai muiden pelaajien toimesta, kiltoihin liittymisen, käytetyt tai löydetyt esineet sekä siirtyminen pelialueelta toiselle. Lokidatassa on yhteensä 81 erilaista pelitoimintoa, ja jokaisella toiminnolla oma palvelinlokikoodinsa eli $logid\_id$. Erilaisia kategorioita ovat pelitila, hahmo, esine, taito, tehtävä ja kilta.

Tässä tutkimuksessa esine-kategoriasta on eritelty omaksi ryhmäkseen vaihtamiseen liittyvät pelitoiminnot, eli toiminnot joissa pelaaja vaihtaa pelirahaa esineisiin tai esineitä pelirahaksi muiden pelaajien tai pelintarjoajan kanssa. Esine-toiminnoista on myös eroteltu materiaali-infot, jotka sisältävät tietoa esimerkiksi aseen luomiseen käytetyistä materiaaleista. Tämän lisäksi omaksi kategoriaksi on lajiteltu kaikki joukkueeseen liittyvät pelitoiminnot. Pelidatan rakenne on tarkemmin esitetty kuvissa 3.2 ja 3.3.

Datasetti	**Ajanjakso**	**Pituus**	**Pelaajien määrä**	**Pelin maksustrategia**
Train-data	1.4.-11.5.2016	40 päivää	4000 pelaajaa	kuukausittainen maksu
Testidata 1	13.7.-7.9.2016	56 päivää	3000 pelaajaa	kuukausittainen maksu
Testidata 2	14.12.2016 - 8.2.2017	56 päivää	3000 pelaajaa	ilmainen

Taulukko 3.3: Kilpailun datasettien keräysaikataulu, datasetin pelaajien määrä ja pelin maksustrategia datasetin keräyksen aikana.

Jokaiselle lokitoiminnolle on asetettu myös 77 erilaista yksityiskohtaisempaa saraketta, jotka kuvaavat erilaisia lisätietoja, kuten määrän muuttumista tai kohdetta. Nämä tietokentät muodostavat neljä erilaista kategoriaa: yleistiedot, toimija, objekti ja kohde. Lokiviestin tyypistä riippuen kaikki kentät eivät ole aina relevantteja. Yleiskategoria sisältää pelitietoja, kuten milloin toiminto on ajallisesti tehty. Toimijaan on sisällytetty pelihahmoon kohdistuvaa dataa, kuten pelaajan taso, hahmon tunnuskoodi, rotu ja luokka. Kolmas kategoria on objekti. Objekti tarjoaa lisätietoa esineisiin kohdistuvista toiminnoista, kuten onko esine pelaajan luoma ja minkä tyyppinen esine on. Kohde-kategoriassa on määritelty toiminnon kohde ja se voi olla esimerkiksi toinen pelaaja.

Luku 4

Kilpailumallit ja ajallisten piirteiden merkitys kilpailussa

4.1 Tiimi UTU:n kilpailumalli

Tiimi UTU:n kilpailumallin on luonut Turun yliopiston kilpailujoukkue. Tässä kappaleessa esitelty kilpailumalli on tiimin työtä. Tiimi menestyi kilpailussa hyvin, ja saavutti ensimmäisessä kilpailutehtävässä toisen sijan ja toisessa kilpailutehtävässä kolmannen sijan. Tiimin käyttämä lineaarinen malli oli yksinkertaisuudessaan hyvin poikkeuksellinen muihin kilpailussa menestyneihin kilpailutöihin nähden.

4.1.1 Piirteiden prosessointi

UTU-tiimi keräsi yhteensä 102 piirrettä, joista 18 päätyi lopulliseen malliin. Tiimin työssä onkin merkityksellisintä se, että lopullisen mallin piirteiden määrä on huomattavan pieni. Kerättyjä piirteitä oli kolmenlaisia: ajallisia piirteitä, pelihahmoon liittyviä piirteitä sekä lokiviestipiirteitä eli pelitoiminnoista kerättyjä statistisia piirteitä.

Tiimi normalisoi kaikki käyttämänsä piirteet $Scikit-learn$ -nimisen kirjaston standardiskaalaajan (eng. standardScalar) $z = (x-u)/s$ avulla, jossa x on näyte (eng. sample), u on näytteen keskiarvo ja s on kaikkien näytteiden keskihajonta [20]. Tämän lisäksi tiimi teki joitakin piirrekohtaisia ratkaisuja, kuten esimerkiksi pienensi piirteen arvoja kymmenkantaisen logaritmin avulla. Pienennettyjä arvoja olivat esimerkiksi erilaiset kokemuspisteiden määrää mittaavat piirteet, jotka olivat lähes poikkeuksetta todella suuria.

Tiimi huomasi myös, että piirteitä valitessa ja prosessoidessa on tärkeä huomioida, että kilpailun data on valmiiksi jaettu datasetteihin. Tämä tarkoittaa sitä, että datasetit ovat pituudeltaan erilaisia ja mitattu eri ajankohtina. Näin ollen malliin tuli valita sellaisia piirteitä, jotka pysyvät melko samanlaisina testisettien välillä. UTU-tiimi mittasi piirteiden yhtenevyyttä datasettien välillä esimerkiksi keskiarvon, mediaanin ja tiheyden perusteella. Yleiseen malliin toimivat piirteet ovat todennäköisemmin sellaisia, jossa datasettien keskiarvot ovat lähellä toisiaan ja jakauma on samanlainen. Kuvassa 4.1 esitetään tiimin tekemää tutkimustyötä aktiivisuuspiirteiden eri arvojen jakaumasta testisetin pelaajien välillä. Kuvassa on myös ilmoitettu keskiarvo ja mediaani.

Hahmopiirteiden prosessointi

Tiimi keräsi pelaajan pelihahmoista tietoa, kuten suurimman ja pienimmän tason (eng. level), suurimman ja pienimmän kokemuspistemäärän tai kuinka moneen kiltaan, tiimiin tai joukkueeseen hahmo on kuulunut. Tämän lisäksi kerättyjä piirteitä olivat muun muassa hahmon rahatilanne datan alussa ja lopussa, onko hahmo poistettu ja mikä on hahmon työ, rooli ja sukupuoli. Tiimi mittasi myös kuinka kauan pelaaja on pelannut kutakin hahmoa milläkin pelin alueella ja kumpaan pelin ryhmittymään eli allienssiin pelihahmo kuuluu.

Kerätyistä hahmotiedoista tiimi prosessoi muun muassa pelaajan kokonaispeliajan (UTU-H1), pelattujen ja poistettujen hahmojen määrän (UTU-H2,UTU-H3), kaikkien hahmojen kokemuspisteiden kokonaismäärän (UTU-H5), pelaajan rating-menestyksen kaksintaistelussa (UTU-H12) ja pelaajan joukkueiden, kiltojen ja tiimien määrä (UTU-H15, UTU-H16 ja UTU-H17).

$(UTU-H1)$ **actors_time** $\sum$: *Kaikkien pelattujen hahmojen peliaikojen summa.*

$(UTU-H2)$ **actors_played** $\sum$: *Pelattujen hahmojen määrä.*

$(UTU-H3)$ **actors_deleted** $\sum$: *Poistettujen hahmojen määrä.*

$(UTU-H4)$ **actors_diversity** $jakauma$: *Kuinka tasaisesti pelaaja on pelannut kaikkia pelihahmojaan.*

$(UTU-H5)$ **actors_totalexp** $\sum$: *Kaikkien pelattujen hahmojen kokemuspisteiden summa.*

$(UTU-H6)$ **actors_maxexp** $\max$: *Käyttäjän suurin kokemuspistemäärä pelihahmolla.*

$(UTU-H7)$ **actors_avgexp** avg: *Pelaajan pelihahmojen maksimaalisten kokemuspisteiden keskiarvo.*

$(UTU-H8)$ **actors_play_avgexp** $\sum$: *Pelaajan kaikkien pelihahmojen keskimääräisten kokemuspistemäärien summa. Hahmon keskimääräinen kokemuspistemäärä on hahmon suurimman ja pienimmän kokemuspistemäärän keskiarvo. Pelaajan kaikkien pe-*

lihahmojen keskimääräistä kokemuspistemääräsummaa laskiessa kunkin hahmon arvoa painotetaan sen mukaan, kuinka suuren osan pelaaja on pelannut pelistä kyseisellä hahmolla.

$(UTU-H9)$ **actors_add_totalexp** $\sum$: *Pelaajan kaikkien pelihahmojen suurimman ja pienimmän kokemuspistemäärän erotusten summa.*

$(UTU-H10)$ **actors_add_maxexp** $erotus$: *Pelaajan suurimman ja pienimmän kokemuspistemäärän erotus.*

$(UTU-H11)$ **actors_start_newbie** $boolean$: *Pelaajan pienin level on 1 (uusi pelaaja).*

$(UTU-H12)$ **actor_rating_avg** avg: *Pelaajan keskivertoinen menestyminen (eng. rating) kaksintaisteluissa.*

$(UTU-H13)$ **actors_money_avg** avg: *Pelaajan keskivertoinen pelirahamäärä.*

$(UTU-H14)$ **actors_hasparties** $boolean$: *Onko pelaaja pelannut joukkueessa.*

$(UTU-H15)$ **actors_nparties** $\sum$: *Joukkueiden määrä.*

$(UTU-H16)$ **actors_nguilds** $\sum$: *Kiltojen määrä.*

$(UTU-H17)$ **actors_nteams** $\sum$: *Tiimien määrä.*

$(UTU-H18)$ **actors_start_maxexp** $\max$: *Pelaajan pienimmän levelin kokemuspistemäärä (eli kuinka paljon pelaajalla on kokemuspisteitä ennen datan keräämistä).*

Aktiivisuuspiirteiden prosessointi

Aktiivisuuspiirteet mittaavat käyttäjän peliaikaa. Tiimi mittasi käyttäjän pelissä viettämää aikaa muun muassa kokonaisaikana, sessioiden määrään ja pituuden mukaan, pelattujen päivien määrästä sekä lokiviestien päivittäisestä määrästä. Tiimi prosessoi kerättyjä peliaikoja kahdella eri ajanjaksolla: koko datan keräysjaksolta (UTU-A1,UTU-A4,UTU-A7) ja viimeisen viikon ajalta (UTU-A2,UTU-A5,UTU-A8).

$(UTU-A1)$ **day_availability** $\%$: *Keskiarvoinen prosentuaalinen aika pelitilassa.*

Piirre on laskettu päivittäisen pelitilassa vietetyn prosentuaalisen osuuden keskiarvosta. Yhdeksi pelitilaksi on laskettu sellainen aika, jossa peräkkäisten lokiviestin välillä on kulunut alle 15 minuuttia. Piirre kuvaa sitä, kuinka monta prosenttia kaikista neljästäkymmenestä päivästä pelaaja on ollut pelitilassa.

$(UTU-A2)$ **end_availability** $\%$: *Keskiarvoinen prosentuaalinen aika pelitilassa viimeisen 10 päivän ajalta.*

$(UTU-A3)$ **inc_availability** inc: *Pelitilassa vietetyn ajan muuttuminen pelattujen päivien myötä.*

Piirre on laskettu lineaarisen regression avulla: x on päivä ja y on kokonaisaika pelitilassa jaettuna päivien määrällä.

$(UTU-A4)$ **day_probability** $\%$: *Kuinka monena päivänä kaikista päivistä pelaaja on pelannut peliä.*

$(UTU-A5)$ **end_probability** $\%$: *Kuinka monena päivänä pelaaja on pelannut peliä*

viimeisen viikon aikana.

$(UTU-A6)$ **inc_probability** : *Päivittäisen pelaamisen muutos.*

Piirre on laskettu lineaarisen regression avulla: x on päivä ja y on kokonaispelaamistodennäköisyys jaettuna päivien määrällä.

$(UTU-A7)$ **day_msgs** $\%$: *Lokiviestien määrä päivässä.*

$(UTU-A8)$ **end_msgs** avg: *Lokiviestien määrä päivässä viimeisen viikon aikana.*

$(UTU-A9)$ **inc_msgs** inc: *Lokiviestien päivittäisen määrän muutos.*

Piirre on laskettu lineaarisen regression avulla: x on päivä ja y on lokiviestien kokonaismäärä jaettuna päivien määrällä.

$(UTU-A10)$ **session_length** avg: *Pelisession pituus.*

$(UTU-A11)$ **first_day** $\%$: *Keskiarvoinen prosentuaalinen aika pelitilassa datan ensimmäisenä päivänä.*

$(UTU-A12)$ **last_day** $\sum$: *Kuinka monena päivänä pelaaja pelasi viimeisen viikon aikana.*

Lokiviestipiirteiden prosessointi

Lokidatassa on yhteensä 81 erilaista lokiviestitunnusta pelin pelitoiminnoille. Lokiviestit ovat esitelty luvussa 3 kuvissa 3.3 ja 3.4. UTU-tiimi prosessoi pelaajan kaikki lokiviestit, ja keräsi lokiviesteistä sekä jokaisen lokiviestityypin prosentuaalisen määrän kaikista pelaajan lokiviesteistä (UTU-L1) että prosentuaalisesti pelatun ajan pelaajan koko peliajasta (UTU-L2).

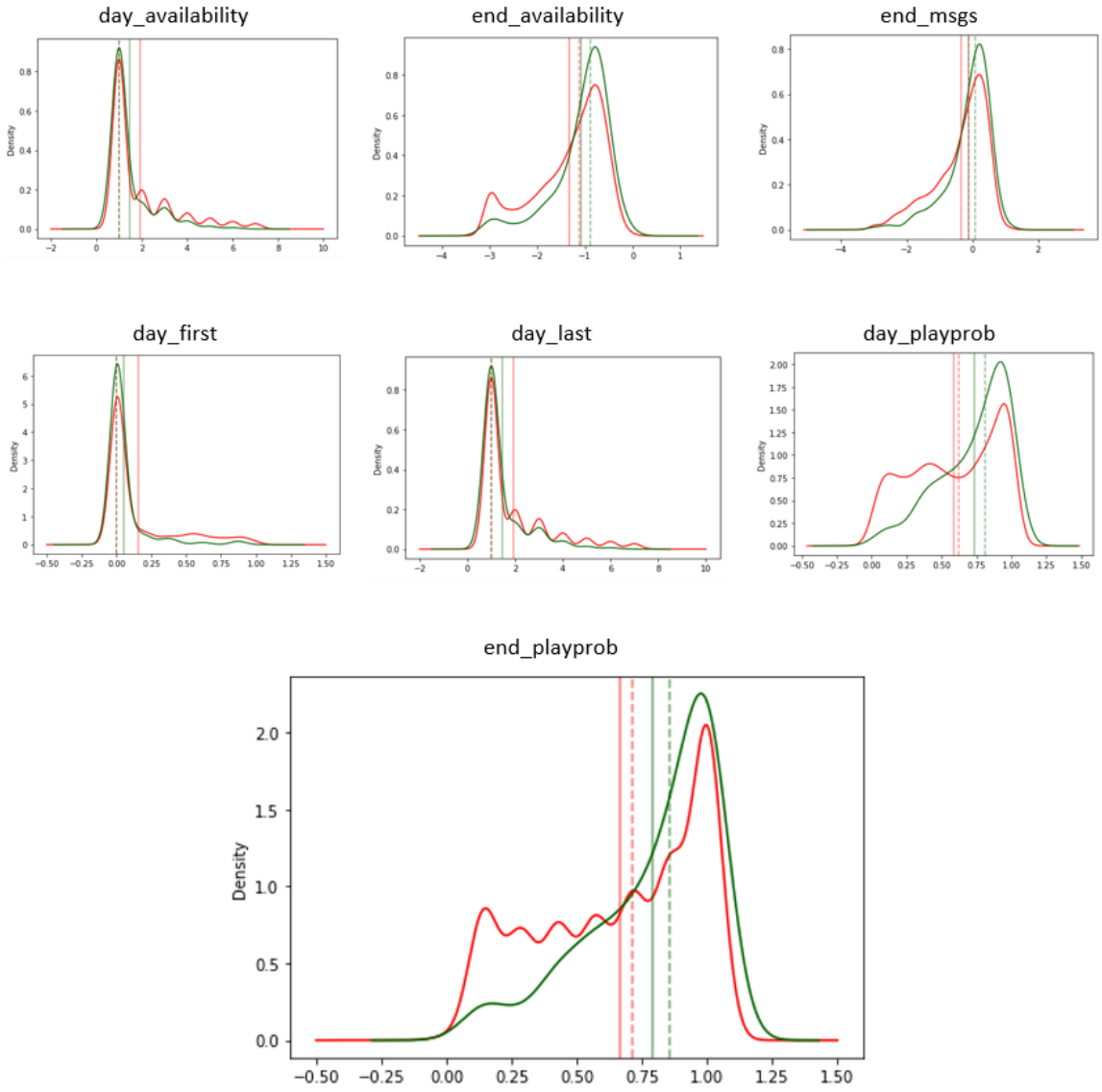
$(UTU-L1)$ **[id]_count** $\%$: *Lokiviestityypin määrä jaettuna kaikkien lokiviestien kokonaismäärällä.*

$(UTU-L2)$ **[id]_time** $\%$: *Lokiviestityypin pelaamiseen mennyt aika jaettuna pelaajan kokonaispeliajalla.*

Kokonaisaika on laskettu pelisessioiden pituuksien summana. Yhdeksi pelisessioksi on laskettu sellainen yhtäjaksoinen pelaaminen, jonka peräkkäisten lokimerkintöjen välillä on kulunut alle 15 minuuttia.

Piirteiden valinta

UTU-tiimi käytti piirteiden valintaan useita erilaisia tekniikoita, kuten pääkomponenttianalyysiä (eng. principal component analysis, PCA), $Scikit-learn$ -kirjaston rekursiivista piirteiden eliminointia (eng. recursive feature elimination, RFE) [21] ja samaisen kirjaston $SelectKBest$-nimistä univariaattia piirteiden valintaa [22]. Tämän lisäksi tiimi käytti piirteiden valinnassa erilaisia statistisia mittareita, kuten keskiarvoa, mediaani ja jakaumaa (kuva 4.1). Näillä mittareilla tiimi varmisti sen, että piirteet ovat tarpeeksi samanlaisia datasettien välillä. Tiimi päätyi lopullisessa kilpailumallissaan sellaiseen joukkoon piirteitä, joiden jakaumat olivat samankaltaisia eri dataseteissä ja jotka antoivat parhaan tuloksen rekursiivisessa piirteiden eliminoissa.



Kuva 4.1: Kuvassa on kaikki malliin valitut aktiivisuuspiirteet ja niiden arvojen jakauma (tiheys) eri dataseteissä. Punainen kuvaa train-data ja vihreä testidata 1:stä. Katkoviiva kuvaa esiintymien keskiarvoja ja yhtenäinen viiva kuvaa esiintymän mediaania. UTU-tiimi käytti mallissaan vain sellaisia piirteitä joiden esiintymien tiheyden jakautuminen on samankaltainen kaikissa testiseteissä.

Lopullisessa mallissa oli 18 piirrettä, joista neljä oli lokiviestipiirteitä (kuva 4.4). Valitut lokiviestipiirteet mittasivat kuinka usein pelaaja osallistui joukkueen huutokaupan voiton jakoon, sai tai käytti luotua esinettä tai suoritti päivän haasteen. Lokiviestipiirteet ovat varsin yllättäviä, mutta kertovat todennäköisesti siitä, että pelistä motivoituneet pelaajat käyttävät harvinaisempia pelitoimintoja, kuten huutokauppaa tai luo omia esineitä. Tällaisten harvinaisempien toimintojen käyttäminen viestii siitä, että pelaaja tietää miten peli toimii ja pelaaja on utelias kokeilemaan kaikkia toimintoja. Lokiviestipiirteistä vähinten yllättävä oli päivittäisten haasteiden suorittaminen, joka luonnollisesti mittaa myös sitä kuinka monena päivänä pelaaja on pelannut peliä.

Aktiivisuuspiirteitä malliin valittiin seitsemän kappaletta (4.3). Neljä aktiivisuuspiirrettä kuvaavat datan keräämisen loppuvaiheen aktiivisuutta, kuten sitä kuinka paljon pelaajalla on lokiviestejä viimeiseltä viikolta, tai kuinka monena päivänä pelaaja on pelannut datan viimeisellä viikolla. Loput malliin valituista aktiivisuuspiirteistä vastaavasti mittaavat pelaajan aktiivisuutta datan alkuaikoina tai koko datan ajalta. Mittaamalla aktiivisuutta datan alusta ja lopusta tiimi onnistui löytämään pelaajat, joiden kiinnostus peliä kohtaan on kasvanut tai hiipunut.

Kuusi lopulliseen malliin valittua hahmopiirrettä (kuva 4.4) mittaavat pelaajan kokemuspisteitä, pelirahan määrää ja menestymistä. Ainoa malliin päätynyt sosiaalinen piirre oli pelaajan kiltojen määrä. Pelaaja voi kuulua vain yhteen kiltaan kerrallaan, joten piirre mittaa sitä, kuinka aktiivisesti pelaaja on eronnut killasta ja liittynyt uuteen kiltaan. Pelaajan kokemuspistemäärä ennen dataa taas mittaa sitä, kuinka kokenut pelaaja on ollut ennen kuin dataa on alettu keräämään. Tämä mittari antaa hyvää vertailupohjaa muille pelaajasta mitatuille kokemuspistemäärille, kuten esimerkiksi pelaajan kaikkien pelihahmojen kokemuspisteiden summalle.

Aktiivisuuspiirre	Nimi mallissa (englanniksi)	Määrittely
Päivittäinen pelitila	*day_availability*	Keskiarvoinen prosentuaalinen aika pelitilassa
Viimeisen viikon pelitila	*end_availability*	Keskiarvoinen prosentuaalinen aika pelitilassa datan viimeisellä viikolla
Pelattujen päivien osuus	*day_playprob*	Kuinka monena päivänä kaikista pelipäivistä pelaaja on pelannut
Pelattujen päivien osuus viimeisestä viikosta	*end_playprob*	Kuinka monena päivänä kaikista viikon pelipäivistä pelaaja on pelannut datan viimeisellä viikolla
Viimeisen viikon lokiviestien määrä	*end_msgs*	Pelaajan lokiviestien kokonaismäärä datan viimeiseltä viikolta
Ensimmäinen pelipäivä	*day_first*	Keskiarvoinen prosentuaalinen aika pelitilassa datan ensimmäisenä päivänä
Viimeinen peliviikko	*day_last*	Kuinka monena päivänä pelaaja on pelannut datan viimeisellä viikolla

Kuva 4.2: Lopulliseen malliin päätyneet seitsemän aktiivisuuspiirrettä.

4.1.2 Kilpailutehtävä: Logistinen regressiomalli

Kilpailun tehtävänä oli luokitella pelaajat pelaajiin, jotka lopettavat pelin ja pelaajiin, jotka jatkavat pelaamista. Team UTU käytti ratkaisumallina binääristä logistista regressiomallia lasso-tappiofunktiolla eli l1:llä. Logistinen regressiomalli on luonnollinen valinta kilpailutyön ensimmäiseen tehtävään, sillä pelaajien pelin lopettaminen on datassa merkitty binäärisesti ja logistinen regressio käyttää todennäköisyyksien logaritmia [23]. Usean muuttujan x lineaarinen regressiomalli voidaan ilmaista kaavalla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \epsilon \tag{4.1}$$

yhtälössä β_p regressiokerroin p, y on ja ϵ on mallin residuaali. Monen muuttujan logistinen regressio voidaan esittää yhtälöllä:

$$ln\frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p \tag{4.2}$$

Hahmopiirre	Nimi mallissa (englanniksi)	Määrittely
Pelaajan kokemuspisteet	*actor_totalexp*	Pelaajan kaikkien pelihahmojen suurimpien kokemuspisteiden summa
Pelaajan kokemusmäärä ennen dataa	*actor_start_maxexp*	Pelaajan pienimpään leveliin tarvittava kokemuspistemäärä (eli kuinka kokenut pelaaja on, kun dataa on alettu keräämään)
Pelaajan keskiverto kokemuspistemäärä	*actor_play_avgexp*	Pelaajan pelihahmojen keskiverto kokemuspistemäärä (suurimman ja pienimmän kokemuspistemäärän keskiarvo) painotettuna pelihahmon pelimäärällä
Pelaajan menestyminen	*actor_rating_avg*	Pelaajan pelihahmojen keskimääräinen menetyspisteytys painotettuna pelihahmon pelimäärällä
Pelaajan peliraha	*actor_money_avg*	Pelaajan pelihahmojen keskimääräinen pelirahamäärä painotettuna pelihahmon pelimäärällä
Kiltojen määrä	*actor_nguilds*	Kuinka moneen kiltaan pelaaja on liittynyt

Kuva 4.3: Lopulliseen malliin päätyneet kuusi hahmopiirrettä.

jossa $\frac{\pi(x)}{1-\pi(x)}$ on jonkin tapahtuman x tapahtumisen todennäköisyys verrattuna siihen, että tapahtuma ei tapahdu, eli $\frac{P(Y=1|x)}{P(Y=0|x)}$ [24].Tutkimuksessa tappiofunktioksi valittiin lassofunktio eli pienimmän neliösumman menetelmää käyttävä rangaistusfunktio, joka ehkäisee mallin train-datan ylisovittamista. Lasso sakottaa regressiokertoimia β seuraavan yhtälön mukaisesti [25]:

$$\beta^{lasso} = argmin\left\{\frac{1}{2}\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{4.3}$$

Tiimi paransi mallinsa ennustustarkkuutta painoitusmuutoksen avulla. Muutos on merkittävä, sillä ilman muutosta train-datan ristiinvalidoinnin testidatan ennustetarkkuus oli alle 0,6. Tiimi muutti painoarvoa niin, että malli suosi enemmän pelin lopettaneita pelaajia, kuin peliä jatkaneita pelaajia. Toisin sanoen malli nosti jokaisen pelaajan kohdalla todennäköisyyttä, että pelaaja kuuluisi ennemmin lopettaneisiin pelaajiin kuin jatkaneisiin.

Lokipiirre	Nimi mallissa (englanniksi)	Määrittely
Joukkueen huutokaupan voiton jako	*msg_2016*	Kuinka monta kertaa pelaaja on osallistunut joukkueen huutokaupan voiton jakoon verrattuna pelaajan lokiviestien kokonaismäärään
Pelaaja käyttää luotua esinettä	*msg_2405*	Kuinka monta kertaa pelaaja on käyttänyt luotua esinettä verrattuna pelaajan lokiviestien kokonaismäärään
Pelaaja saa luodun esineen	*msg_2407*	Kuinka monta kertaa pelaaja on saanut luodun esineen verrattuna pelaajan lokiviestien kokonaismäärään
Pelaaja suorittaa päivän haasteen	*msg_5011*	Kuinka monta kertaa pelaaja on suorittanut päivän haasteen verrattuna pelaajan lokiviestien kokonaismäärään

Kuva 4.4: Lopulliseen malliin päätyneet neljä lokiviestipiirrettä.

```
# UTU-tiimin painonmuutosalgoritmi (paino on 0.3)

Input: y-arvo (y_prob) ennen painon muutosta
Output: y-arvo (y_prob_weighted) painon muutoksen jalkeen

1: input: y_prob
2: let: weight <- 0.3
3: y_True <- y_prob * 0.5/weight
4: y_False <- (1 - y_prob) * 0.5/(1 - weight)
5: y_prob_weighted <- y_True / (y_True + y_False)
6: output: y_prob_weighted

#
```

Kilpailuun osallistuneet tiimit eivät saaneet testisettien label-arvoja ja kilpailutehtävänä olikin ennustaa testisettien pelaajien pelin jatkaminen train-datasetin avulla. Tiimi

UTU käytti train-datan ristiinvalidoinnissa $Scikit-learn$ -kirjaston $cross\_val\_score$ -tekniikkaa [26], jossa ristiinvalidointistrategiaksi on asetettu kymmeneen kerrokseen ositettu ristiinvalidointi eli $StratifiedKFold$ ($k = 10$) [27].

4.2 Kilpailussa menestyneet kilpailumallit

Kilpailun voittajajoukkue Yokozuna Data (YD) käytti mallissaan hieman alle 500:aa piirrettä, jotka tiimi valitsi tutkimastaan 3000:sta piirteestä neuroverkkotekniikka käyttävän autoenkoodaajan (engl. autoencoder) avulla [19]. Piirteiden määrä tuntuu todella suurelta, jos ottaa huomioon, että kilpailun ensimmäisessä tehtävässä tiimi UTU jäi voittajatiimistä ennustustarkkuudessa vain noin 1,8 % ja heidän piirremääränsä oli alle 20. Myös viidenneksi päätynyt goedle.io käytti mallissaan useita satoja piirteitä. [2]

Voittajatiimi otti myös huomioon testisettien eroavaisuudet, ja mallin piirteiksi valittiin vain sellaisia piirteitä, joiden jakauma testisettien välillä ei ollut liian suuri. Tiimi huomasi, että testisettien välillä oli selkeä ero pelaajakäyttäytymisessä [19].Yhteisiä piirteitä YD:llä ja UTU:lla olivat esimerkiksi pelattu kokonaisaika ja sessioiden kesto. Tämän lisäksi tiimeillä oli jonkin verran yhteistä esimerkiksi siinä, miten tiimit vangitsivat piirteeksi erilaisten toimintojen kehittymisen ajan mukana. Selkein eroavaisuus tiimien välillä on piirteiden määrä, mutta myös se, että tiimi YD sisällytti piirteisiin huomattavasti enemmän pelitapahtumiin liittyvää tietoa kuin tiimi UTU [19]. Tiimi UTU myös rakensi enemmän laskennallisia piirteitä, jotka koostuivat erilaisista pelitapahtumista ja ajallisista piirteistä, kun tiimi YD taas luotti mallissa piirteiden määrään.

Voittajatiimi YD käytti mallin ennustamisessa äärimmäisen satunnaistettuja päätöspuita (eng. extremely randomized trees). Päätöspuutekniikat olivat yleisestikin käytetyin tekniikka kilpailussa menestyneissä ratkaisuissa, kun taas lineaarisia malleja käyttävistä jouk-

kueista ainoastaan UTU-tiimi ylsi kärkiviisikkoon. Kolmanneksi, neljänneksi ja viidenneksi ensimmäisessä kilpailutehtävässä sijoittuneet joukkueet TripleS, TheCowKing ja goedleio käyttivät kaikki päätöspuumallia. Kolmannelle sijalle jäänyt TripleS:n malli käytti satunnaistettua päätösmetsää (eng. randomized forest), neljänneksi sijoittunut TheCowKing LightGBM-tekniikkaa ja viidenneksi sijoittunut goedleio yhdisteli voittajajoukkueen tavoin syväoppimista ja päätöspuita. Tiimi UTU:n tavoin logistista regressiota käyttäneet joukkueet YK ja GoAlone jäivät kauas sijoille kymmenen ja yksitoista [2].

Eunjo Lee al. tutkivat artikkelissaan "Game Data Mining Competition on Churn Prediction and Survival Analysis using Commercial Game Log Data"kilpailussa käytettyjä malleja. Tutkimusryhmä lajitteli kilpailunjoukkueiden piirteet ryhmiin sen perusteella, oliko piirre mitattu päivittäisessä tai viikoittaisessa aikavälissä, tai oliko piirre statistiikkaan perustuva. Tutkimusryhmän koostamasta taulukosta selviää, että kaikki neljä kilpailussa parhaiten pärjännyttä joukkuetta käyttivät statistiikkaan perustuvia piirteitä. Kahden parhaan, sekä kolmannelle ja neljännelle sijalle jääneiden välillä on kuitenkin se ero, että kaksi parhaiten pärjännyttä joukkuetta käyttivät mallissaan myös aika-painotteisia piirteitä [2].

Tiimi UTU:n ja muiden tehtävässä menestyneiden joukkueiden mallien välillä oli myös se ero, että muiden tiimien malleissa korostui lokiviestipiirteet. TripleS ja TheCowKing keräsivät lokiviestipiirteiden lisäksi tietoa pelaajan pelitasosta (eng. level), ja tämän lisäksi TripleS:n mallissa oli mukana pelaajan peliaika ja kokemuspisteet [2]. Lokiviestipiirteet olivat vahvasti mukana myös voittajajoukkueen mallissa, ja joukkue mittasi lokiviesteistä useita erilaisia statistisia piirteitä, kuten keskiarvon, mediaanin ja keskihajonnan [19]. Kaikki muut joukkueet käyttivät train-datan ristiinvalidoinnissa viisinkertaista ositusta, kun taas tiimi UTU ositti train-datan jopa kymmeneen osaan.

Luku 5

Pelilliset piirteet

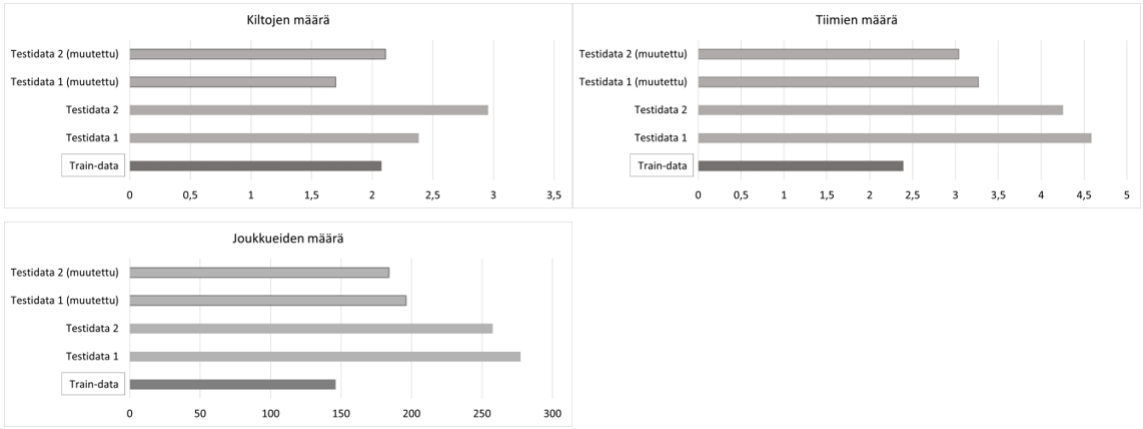
Tässä kappaleessa käsitellään pelillisiä piirteitä, eli mitä pelaaja tekee pelissä, miten pelaaja pelaa peliä muiden pelaajien kanssa, millaisen hahmon pelaaja valitsee ja kuinka hyvin pelaaja menestyy pelissä. Tutkimuksen pelillisiksi piirteiksi on valittu kolme pääkategoriaa: sosiaalinen pelaaminen, pelihahmoon ja pelin alueisiin liittyvät pelivalinnat, sekä menestyminen pelissä. Kategorioiden valitsemiseen on käytetty Yeen pelimotivaatioteoriaa, jossa korostuivat pelin sosiaaliset piirteet, menestymisen tunne ja roolipelaaminen [12]. Tutkituista piirteistä muodostettiin lopulta 29 uuttaa piirrettä, joista 12 oli sosiaalisia piirteitä ja 20 pelialueeseen liittyviä piirteitä. Pelihahmon ja menestymisen osalta malliin ei liitetty uusia piirteitä, mutta kappaleessa tutkitaan tiimi UTU:n pelihahmon valintaa ja menestystä mittaavia piirteitä.

5.1 Sosiaaliset piirteet

Pelin pelaaminen muiden pelaajien kanssa on luonnollisesti suuressa roolissa verkkomoninpeleissä. Sosiaalinen pelaaminen voi tarkoittaa joko pelaamista tiimissä muiden pelaajien kanssa tai pelaamista muita pelaajia vastaan. Tämän lisäksi peleissä voi usein kommunikoida esimerkiksi pelin chattikanavilla. Pelaaminen voi olla ajan viettämistä yhdessä ystävien kanssa, tuen saamista muista kanssapelaajista tai työkalu uusien tuttavien löytä-

miseen. Pelaaminen muiden pelaajien kanssa, kuten esimerkiksi kiltoihin liittyminen, tuo usein jotakin helpotusta pelaamiseen. Pelaamisen sosiaalinen aspekti mahdollistaa myös sen, että peliä saattaa pelata myös sellainen henkilöt, jotka eivät pelaisi peliä yksin.

Joissakin tilanteissa pelaaja pelaa peliä vain sen takia, että hänen sosiaalinen verkostonsa sattuu pelaamaan kyseistä peliä. Kahn et al. tulivat tutkimuksessaan [28] siihen lopputulokseen, että on olemassa tietyntyyppinen ”sosialisoijat” -pelaajatyyppi, joka nauttii erityisen paljon pelaamisen sosiaalisesta aspektista. Tutkimusryhmä päätteli, että sellaiset tähän ryhmään kuuluvat pelaajat, joilla oli ystäviä pelaamassa peliä, pelasivat peliä todennäköisemmin juuri siksi, että heidän ystävänsä pelasivat peliä [28].

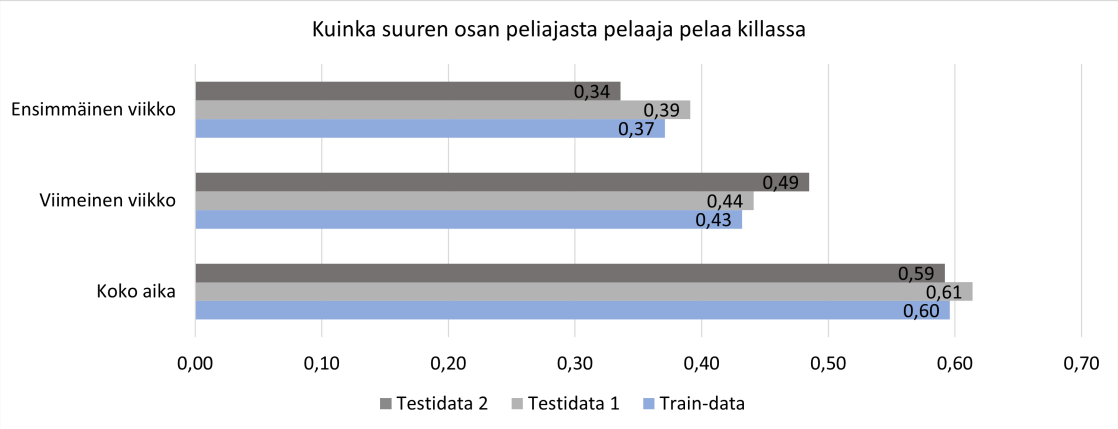


Kuva 5.1: Taulukossa on kuvattu neljäntuhannen pelaajan (train-data) tiimien, joukkueiden ja kiltojen lukumäärä

Train-datasetti ja testidatat eivät ole täysin vertailukelpoisia, sillä testidatoissa pelaajat ovat pelanneet peliä 56 päivää, kun taas train-datassa vain 40 päivää. Näin ollen esimerkiksi pelaajan joukkueiden määrä on luonnollisesti suurempi testidatassa kuin train-datassa. Jotta piirteen p luvuista saa vertailukelpoisia, kahden datasetin m ja n lukuja pitää käsitellä seuraavanlaisesti:

$$m > n : p_n = p_m * (m/n) \tag{5.1}$$

Toisin sanoen pienemmän datasetin piirre p_n kerrotaan datasettien kokoeron mukaan. Suuremman datasetin piirre p_m pysyy sellaisenaan. Kuvasta 5.1 voi huomaa, että muokatut testisetit ovat jokaisessa kuvan tapauksessa (kiltojen, joukkueiden ja tiimien määrässä) lähempänä train-datan vastaavia lukuja kuin muokkaamattomat arvot. Tämä johtuu siitä, että kiltojen, joukkueiden ja tiimien määrä on ajallisesti riippuvaista, eli mitä pidempään pelaaja on pelannut, sitä suurempia luvut yleensä ovat.



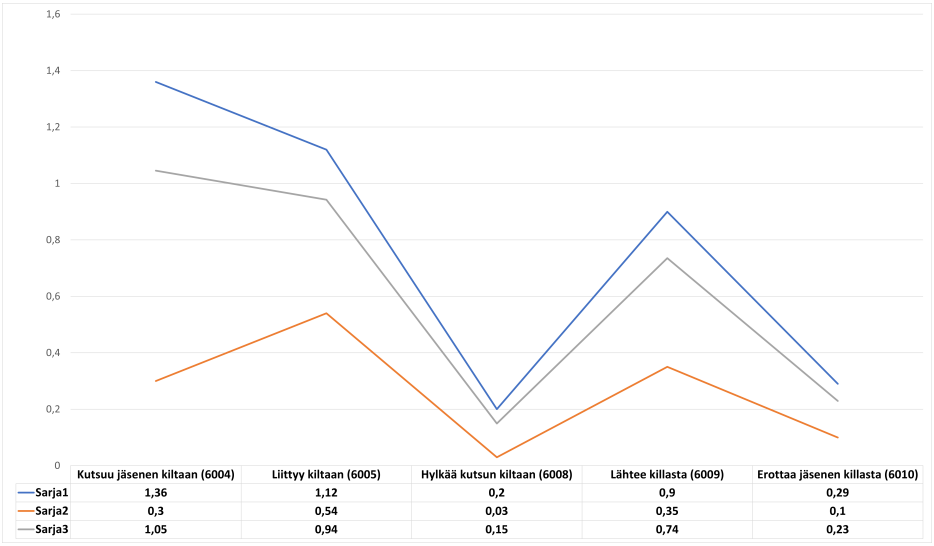
Kuva 5.2: Eri datasettien pelaajien killassa pelaaminen piirteet.

$(S1-3)$ **guild/team/party _availability** $\%$: *Pelaaminen killassa, tiimissä tai joukkueessa suhteessa koko peliaikaan.*

$(S4-6)$ **end_guild/team/party _availability** $\%$: *Pelaaminen killassa, tiimissä tai joukkueessa suhteessa koko peliaikaan (vain viimeinen viikko).*

$(S7-9)$ **first_guild/team/party _availability** $\%$: *Pelaaminen killassa, tiimissä tai joukkueessa suhteessa koko peliaikaan (vain ensimmäinen viikko).*

$(S9-11)$ **first_end_guild/team/party _availability_diff** : *Killassa, tiimissä tai joukkueessa pelaamisen ero ensimmäisen ja viimeisen viikon välillä).*



	Kutsuu jäsenen kiltaan (6004)	Liittyy kiltaan (6005)	Hylkää kutsun kiltaan (6008)	Lähtee killasta (6009)	Erottaa jäsenen killasta (6010)
Sarja1	1,36	1,12	0,2	0,9	0,29
Sarja2	0,3	0,54	0,03	0,35	0,1
Sarja3	1,05	0,94	0,15	0,74	0,23

Kuva 5.3: Taulukon data on muodostettu 4000:sta pelaajasta (train-data). Sarja 1 (sininen) on koottu pelaajista, jotka ovat jatkaneet pelaamista, sarja 2 (oranssi) pelaajista, jotka ovat lopettaneet pelaamisen. Sarja 3 (harmaa) on koottu kaikista pelaajista.

5.2 Pelihahmo, pelialueet ja roolipelaaminen

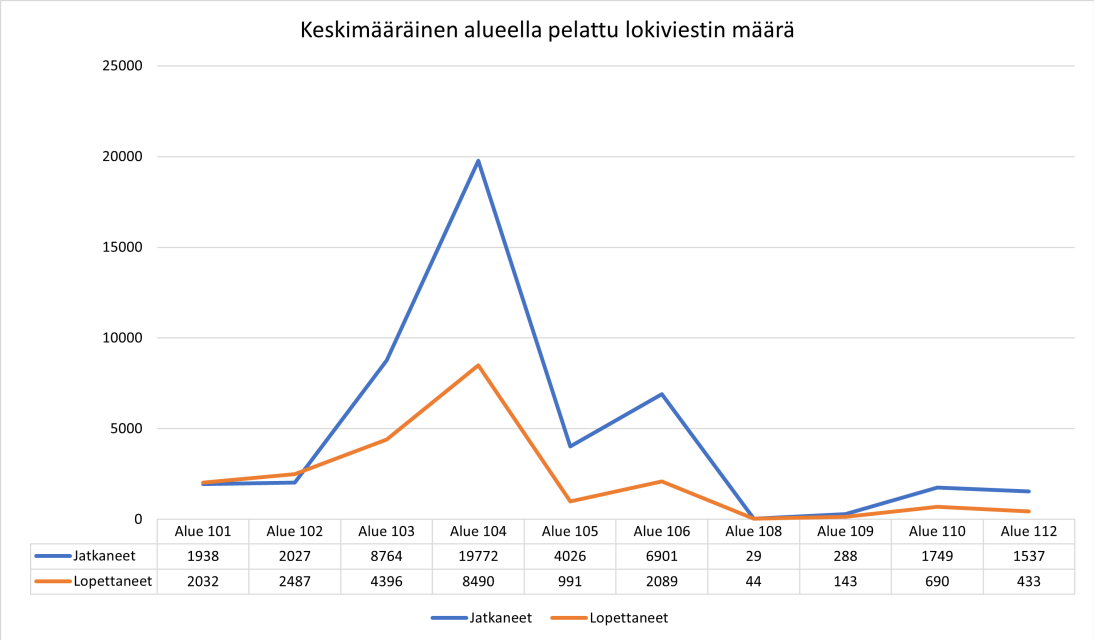
Blade & Soul:ssa on kymmenen erilaista aluetta, joista neljä on erikoisalueita. Pelin tutoriaali sijoittuu Heaven's Reach -nimiselle alueella. Kolme muuta erikoisaluetta ovat harjoittelualue (Hongmoon Training Room), kaksintaisteluareena (Dueling Grounds) ja palvelinten välinen alue (Cross Server Dungeon) [29]. Datasta käy ilmi, että kaikilla pelialueilla ei pelata yhtä paljon, ja pelissä on selkeästi suosittuja ja epäsuosittuja pelialueita (kuvat 7.1 ja 7.2).

Pelaajien lokiviestien keskivertomäärä on kaikissa dataseteissä korkein alueella 104. Lopettaneet pelaajat ovat pelanneet kaikilla muilla alueilla jatkaneita pelaajia vähemmän, paitsi alueella 102. Testisettien välillä on huomattavan vähän eroa ja suurimmillaan ero on viimeisen alueen 112 kohdalla, jossa F2P-pelin datasetin pelaajat ovat pelanneet mak-

sullisen datasetin pelaajia vähemmän (7.2).

$(AL-1)$ **msgs_zone_x** : *Lokiviestien määrä pelialueella x.*

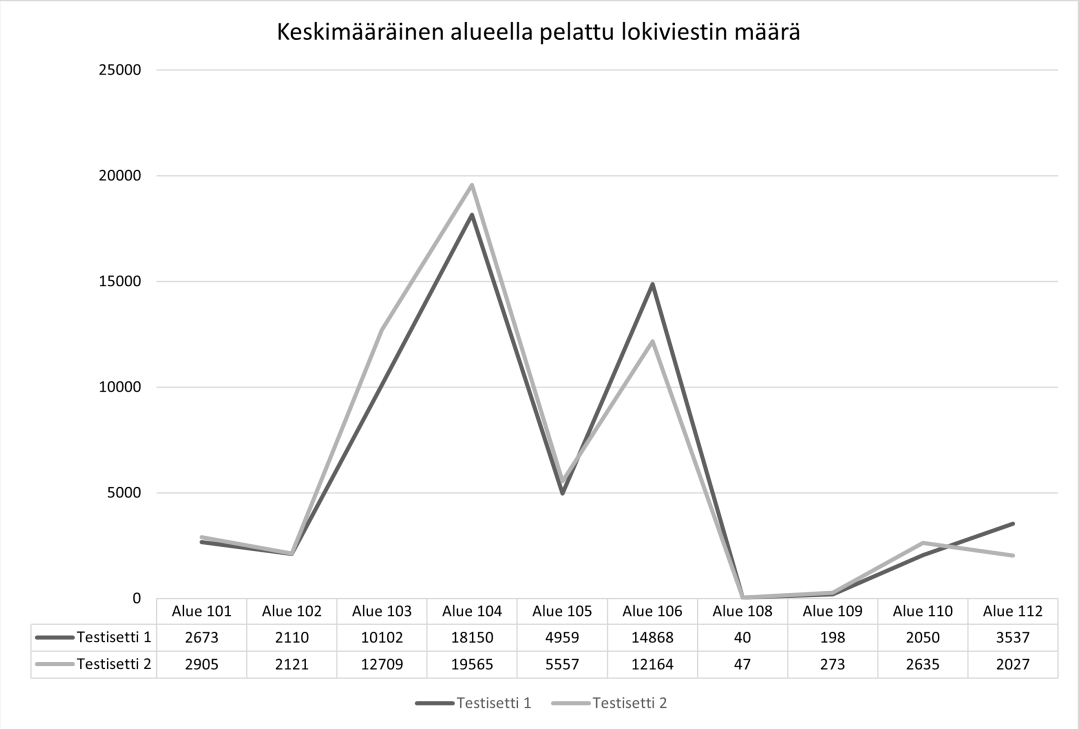
$(AL-2)$ **msgs_zone_x_%** $\%$: *Prosentuaalinen peliaika alueella x pelaajan kokonaispeliajasta.*



	Alue 101	Alue 102	Alue 103	Alue 104	Alue 105	Alue 106	Alue 108	Alue 109	Alue 110	Alue 112
Jatkaneet	1938	2027	8764	19772	4026	6901	29	288	1749	1537
Lopettaneet	2032	2487	4396	8490	991	2089	44	143	690	433

Kuva 5.4: Train-datan pelaajien keskimääräinen lokiviestimäärä kullakin pelin kymmenestä pelialueesta. Data on jaettu lopettaneisiin ja jatkaneisiin pelaajiin.

Blade & Soul:ssa pelaaja luo itselleen pelaamista varten vähintään yhden pelihahmon. Pelaaja valitsee hahmolleen sukupuolen, rodun, luokan ja ulkonäön (kuva 6.1). Tutkimuksessa selvisi, että pelaajat ovat luoneet enemmän pelihahmoja silloin, kun peli on ollut pelaajille ilmainen. Myös train-datan lopettaneiden ja jatkaneiden pelaajien hahmojen lukumäärän välillä on selvä ero, sillä jatkaneet pelaajat ovat luoneet keskimäärin yhden tai kaksi hahmoa enemmän kuin pelin lopettaneet pelaajat.



Kuva 5.5: Pelaajien keskimääräinen lokiviestimäärä kullakin pelin kymmenestä pelialueesta. Kuvassa on kuvattu kummankin testisetin pelaajien pelialueilla pelaaminen.

$(H-1)$ **deleted_actors_%** : poistettujen hahmojen osuus pelaajan hahmoista (kun pelihahmojen määrä > 1)

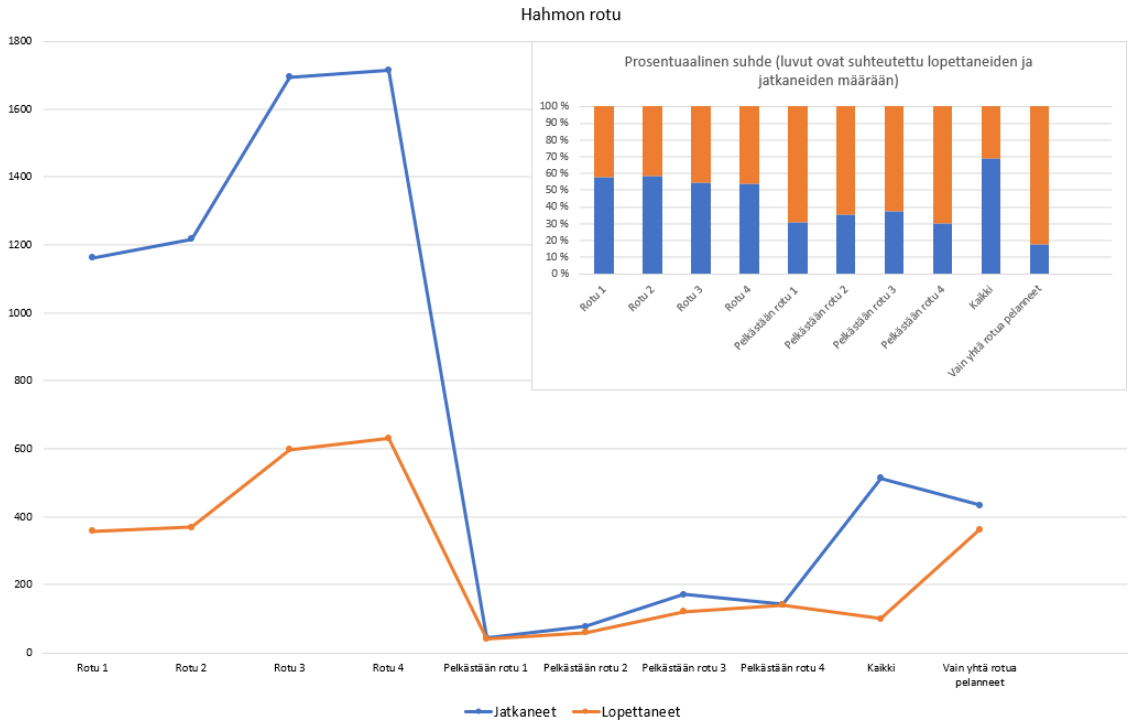
$(H-2)$ **num_actors** $\sum$: Pelihahmojen lukumäärä

Hahmon sukupuoli on pelissä joko nainen tai mies, mutta lokidatassa sukupuoli on merkitty ”sukupuoli 1” ja ”sukupuoli 2” ja datasta ei selviä kumpi sukupuolista on nainen ja kumpi mies. Kolmelle rodulle neljästä (Gon, Lyn ja Jin) käytössä on molemmat sukupuolet, mutta Yun-rotua edustavat hahmot voivat olla vain naispuolisia [17]. Odotetusti pelihahmon sukupuolella ei näyttäisi olevan eroa lopettaneiden ja jatkaneiden pelaajien välillä. Datan mukaan 38% pelaajista on pelannut molempia sukupuolia. Tämä tarkoittaa sitä, että pelaajalla on ollut useampi hahmo, ja hahmot ovat olleet keskenään eri suku-

Kuva 5.6: Kuvankaappaus pelin hahmovalinnasta. Pelaaja voi valita hahmonsa rodun, luokan ja sukupuolen. Hahmon vaikeusaste riippuu pelaajan valinnoista.

puolta. Loput 62% prosenttia pelaajista ovat joko pelanneet vain yhtä hahmoa, tai useaa hahmoa, mutta valinneet hahmoilleen aina saman sukupuolen. Jälkimmäisen joukon pelaajista peräti 84% on pelannut hahmoa tai hahmoja joiden sukupuoli on datassa merkitty "sukupuoli 2".

Pelaaja voi valita hahmonsa neljästä eri rodusta: Gon, Lyn, Yun ja Jin. Roduilla on erilaisia ominaisuuksia, vahvuuksia, heikkouksia, mutta myös arvioitu vaikeusaste [17]. Hahmojen eroavaisuuksista johtuen on luonnollista, että jotkin hahmot ovat suositumpia kuin toiset. Jatkaneet pelaajat ovat luonnollisesti pelanneet todennäköisemmin kaikkia rotuja kuin lopettaneet pelaajat (kuva 5.7). Vastaavasti lopettaneet pelaajat ovat useammin pelanneet vain yhtä rotua. Sukupuolen tavoin rodut on merkitty lokidataan numerokoodilla, ja näin ollen datasta ei suoraan ilmene mikä pelin roduista on pelatuin.

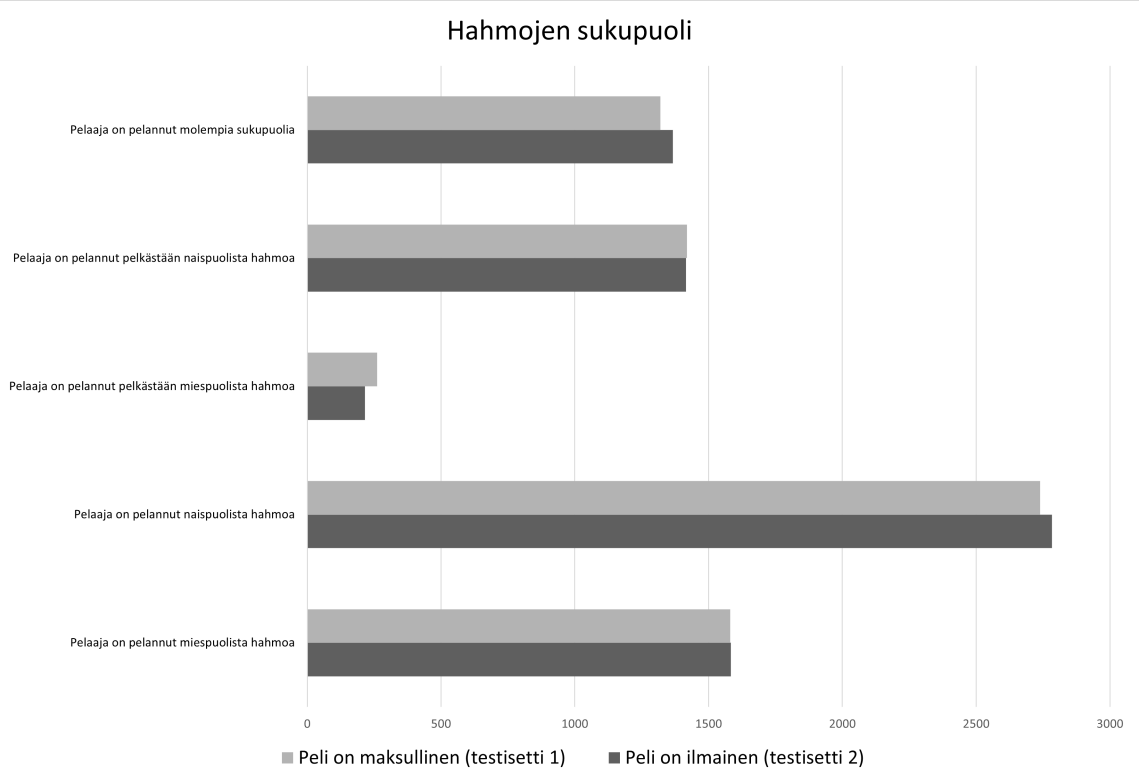


Kuva 5.7: Train-datasetin pelaajien pelihahmojen valinta pelihahmon rodun osalta. Jatkaneet pelaajat ovat merkittyinä sinisellä ja lopettaneet oranssilla.

5.3 Saavutukset, menestys ja kilpaileminen

Blade & Souls -pelissä menestystä ei voi mitata pelaajan pelihahmon kuolemista, sillä kuten suurinta osaa muistakin pelitoiminnoista, peliä jatkaneet pelaajat ovat myös kuolleet pelissä enemmän kuin pelin lopettaneet pelaajat. Pelaaja voi menestyä lukuisissa eri osa-alueilla pelissä. Tällaisia osa-alueita ovat esimerkiksi kuinka paljon leveleitä pelaajalla on, kuinka paljon pelaaja on suorittanut pelin tehtäviä, mikä on pelaajan areena-kaksintaistelujen voittoprosentti ja kuinka monta harvinaista esinettä pelaaja omistaa.

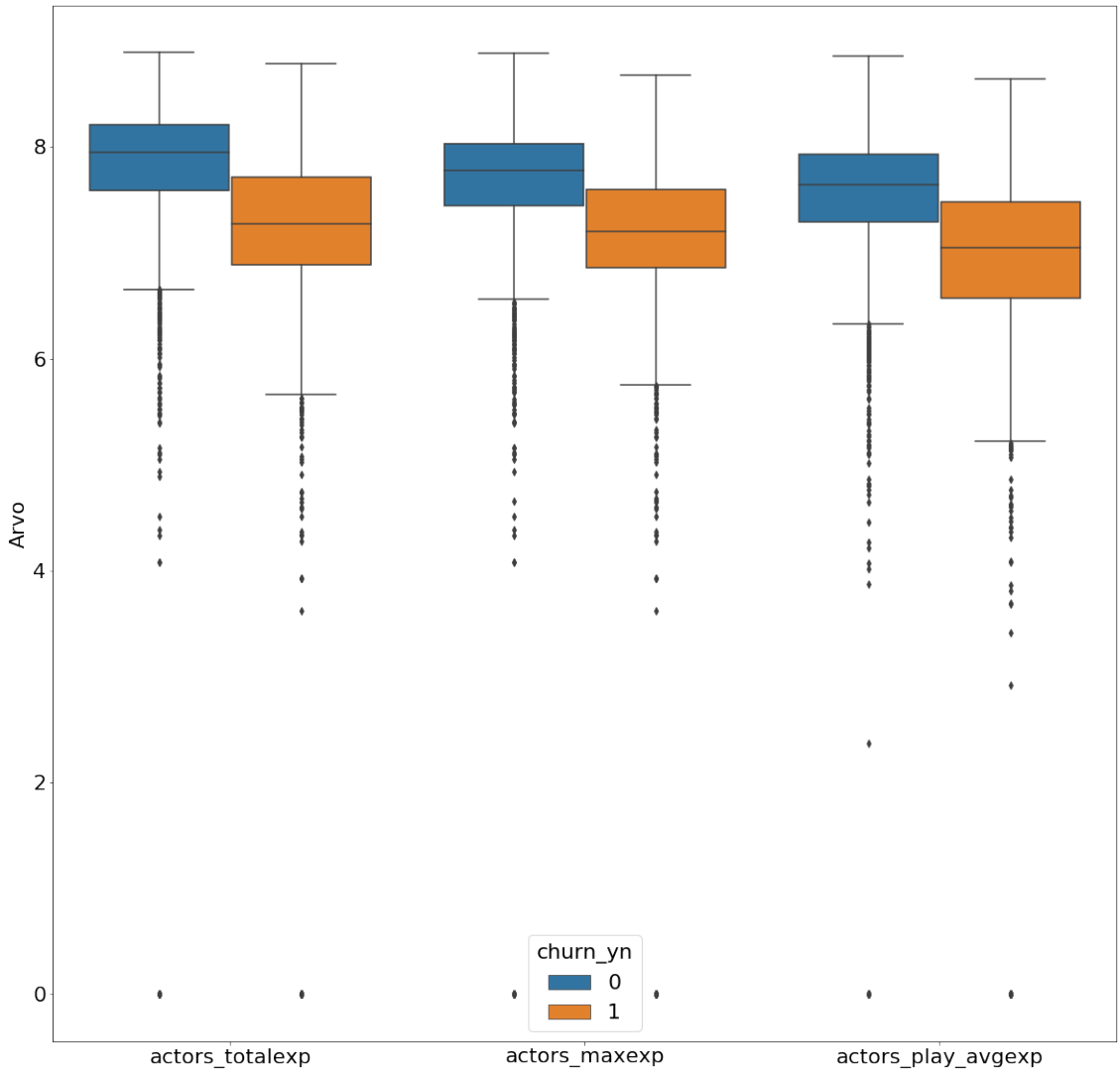
UTU-tiimi mittasi kilpailumallissaan pelaajien menestystä muun muassa kokemuspisteiden avulla (kuva 5.10). Kokemuspisteet (eng. experience points) ovat pisteitä, joita pelaajat saavat pelissä saavutusten myötä. Kokemuspisteet mittaavat epäsuorasti saavutusten lisäksi pelissä vietettyä aikaa. Kokemuspisteiden lisäksi UTU-tiimi mittasi pelaajien kak-



Kuva 5.8: Testisettien pelaajien pelihahmojen sukupuoli.

sintaistelumenestystä (eng. ranking) ja pelirahan määrää, sekä niiden muuttumista keräysajanjakson aikana. Pelirahan lisäksi tutkimuksen kannalta olisi kiinnostava tietää, kuinka paljon pelaajat vaihtavat rahaa pelin tarjoamissa vaihto-toiminnoissa.

Raha on yksi hyvä mittari pelaajan menestymisen mittaamiseen, sillä pelaaja saa pelirahaa esimerkiksi päihittämällä muita pelaajia, tai suorittamalla pelin tehtäviä. Omistamalla pelin rahaa, pelaaja voi ostaa esimerkiksi parempia aseita tai vahvempia suojuksia pelihahmolleen. Joskus pelaajat saattavat kerätä pelirahaa myös siksi, että hyötyvät siitä taloudellisesti. Pelaaminen voi olla taloudellisesti hyödyllistä silloin, kun pelissä on oikean rahan arvoisia virtuaaliesineitä tai pelivaluuttaa esimerkiksi palkintoina, löytöinä tai arpavoittoina.



Kuva 5.9: Pelaajan pelihahmojen suurin kokemuspistemäärä, kaikkien hahmojen kokemuspisteiden summa ja hahmojen kokemuspisteiden keskiarvo. Kokemuspisteiden määrä on normalisoitu kymmenkantaisella logaritmilla.

Kauppatoiminto onkin kiinnostava siitä syystä, että pelaajat, jotka myyvät esineitä vaihtotoiminnon kautta, pelaavat peliä ainakin jollakin tasolla kaupallisen hyödyn takia. Pelaamista, jonka tavoitteena on maksimaalistaa oikeaksi rahaksi vaihdettavia peliesineiden tai pelirahan kerääminen, kutsutaan kullanviljelyksi (eng. gold-farming) [30]. Taloudelliseen hyötyyn pyrkivä pelaaminen on yleistä erityisesti F2P-peleissä, joihin käyttäjä voi luoda useita tilejä maksutta. Maksullisissa peleissä kate jää tällöin luonnollisesti huonommaksi.

Luku 6

Peliajalliset piirteet

Pelaajan pelin pelaamiseen käyttämää kokonaisaikaa voi mitata kerätyn datan ajalta, eli kuinka monta minuuttia (A-1), tuntia (A-2) tai päivää (A-3, A-4) pelaaja on ollut pelissä datan aikaleimojen mukaan. Datasetit ovat kuitenkin pituudeltaan hieman erikokoisia, ja koska kilpailutehtävänä on ennustaa testisettien pelaajat train-datan perusteella, mallin yleistämiseksi aikaerot tulee tasoittaa. Kaikki sellaiset piirteet, joihin vaikuttaa pelatun ajan pituus, tulisi suhteuttaa train-datan keräyspituuden mukaan samalla tavalla kuin kappaleessa 5.1 on esitetty. Tämä koskee siis sekä ajallisia että pelillisiä piirteitä. Taulukossa 6.1 on esitetty piirteiden muunnettuja arvoja.

$(A-1)$ **avg_min_played** avg: *Päivässä pelattujen minuuttien keskiarvo.*

$(A-2)$ **hours_played** sum: *Pelattujen tuntien määrä.*

$(A-3)$ **days_played** avg: *Pelattujen päivien kokonaisarvon keskiarvo.*

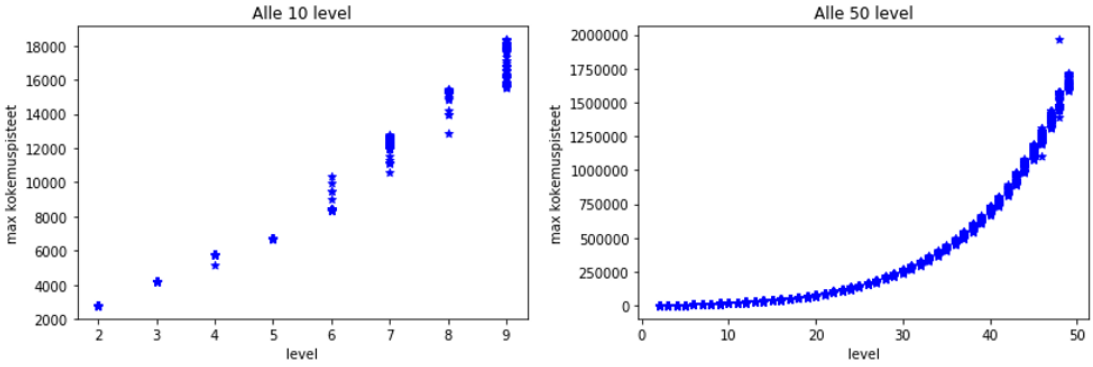
$(A-4)$ **days_unplayed** avg: *Pelaamattomien päivien kokonaisarvon keskiarvo.*

Train-datasetin pelaajat ovat pelanneet keskimäärin 24:nä päivänä keräysajasta. Prosentuaalisesti tämä tarkoittaa, että pelaajat ovat pelanneet 43 % kaikista pelipäivistä. Jokaista pelattua pelipäivää kohden pelaaja on pelannut keskimäärin 148 minuuttia. Pelipäivä on hajautettu peräti kymmeneen erilliseen pelisessioon, mikä viittaa siihen, että pelaajat lähtevät pois pelimaailmasta ja palaavat takaisin useaan kertaan yhden pelikerran aikana. Lopettaneiden ja jatkaneiden pelaajien välillä on huomattava ero erityisesti kokonaispeliajan ja viimeisen viikon peliajan osalta (kuva 6.1).

	Train-data	Testidata 1	Testidata 1 uusi	Testidata 2	Testidata 2 uusi
Päivien määrä datassa	**40 päivää**	56 päivää	**päivien määrä suhteutettu train-dataan**	56 päivää	**päivien määrä suhteutettu train-dataan**
Pelattujen päivien määrä	**24 päivää**	33 päivää	**24 päivää**	37 päivää	**26 päivää**
Pelaamattomien päivien määrä	**16 päivää**	23 päivää	**17 päivää**	19 päivää	**14 päivää**
Kokonaispeliaika	**99 tuntia**	156 tuntia	**111 tuntia**	158 tuntia	**113 tuntia**

Kuva 6.1: Datasettien ajallisia peruspiirteitä. "Testidata 1 uusi"ja "testidata 2 uusi"ovat mukautettuja train-datan pituuden kanssa.

Blade & Soul:ssa pelaaja saa tasoja eli leveleitä keräämiensä kokemuspisteiden mukaan. Pelissä saavutettava maksimilevel on 50 ja tavallisen levelien lisäksi pelaaja voi saavuttaa vielä 25 ylimääräistä Hongmoonin leveliä [31]. Lähes kaikki train-datan pelaajat ovat datan mukaan jollakin hahmollaan levelillä 50, joten level ei itsessään toimi ennustamaan pelaajia, jotka ovat jatkamassa pelaamista. Datasta voi kuitenkin saada jonkinlaista käsityksen pelaajan kokemuspistemäärästä ennen datan keräämistä laskemalla kuinka monta kokemuspistettä pelaaja tarvitsee ansaitakseen yhden levelin (kuva 6.2).



Kuva 6.2: Pelaajien pelihahmojen level ja kokemuspisteiden määrä.

6.1 Yhtäjaksoinen pelaaminen

Yhtäjaksoisesti pelattua aikaa voi niin ikään mitata monella tavalla. Yhtäjaksoisesti pelattua aikaa voi mitata sessioiden kestosta, eli kuinka pitkään pelaaja pelaa peliä yhtäjaksoisesti keskeyttämättä. Toisaalta peliajan yhtäjaksoisuutta voi mitata myös kuinka monena peräkkäisenä päivänä pelaaja on pelannut peliä. Peräkkäisten päivien jaksoa kutsutaan putkeksi (eng. Streak). Rothmeier al. ennustivat menestyksekkäästi tutkimuksessaan [14] ketkä pelaajista tulevat lopettamaan ilmaisen verkkostrategiapelin pelaamisen. Tutkimustiimi huomasi, että peräkkäisten päivien putki oli tärkeä piirre useassa kokeilemassaan mallissa.

Pelaajan suurin peliputki on maksimaalinen peräkkäisten pelipäivien määrä, eli kuinka monena peräkkäisenä päivänä pelaaja on pelannut peliä. Vastaavasti pelaajalta voidaan myös laskea maksimaalinen pelaamaton putki, eli kuinka monena peräkkäisenä päivänä pelaaja ei ole pelannut peliä ollenkaan. Peliputkien määrä on laskettu kuvan 9.2 mukaisesti. Datan jokainen päivä on merkitty joko 1 (pelattu) tai 0 (ei pelattu) sen mukaan, onko pelaajalla lokiviestimerkintöjä päivän aikaleimalla vai ei. Putki on laskettu peräkkäisten

samojen arvojen määrän kumulatiivisena summana ja näin laskettuna suurin arvo on pelaajan pisin putki. Pisin putki voi olla pelaamattomien tai pelattujen päivien putki.

	Pelattu	Putki
1.4.2016	0	1
2.4.2016	0	2
3.4.2016	1	1
4.4.2016	1	2
5.4.2016	1	3
6.4.2016	1	4
7.4.2016	0	1
8.4.2016	1	1
9.4.2016	1	2
10.4.2016	0	1
11.4.2016	0	2
12.4.2016	0	3
13.4.2016	1	1
14.4.2016	1	2
15.4.2016	0	1
16.4.2016	1	1
17.4.2016	0	1
18.4.2016	1	1

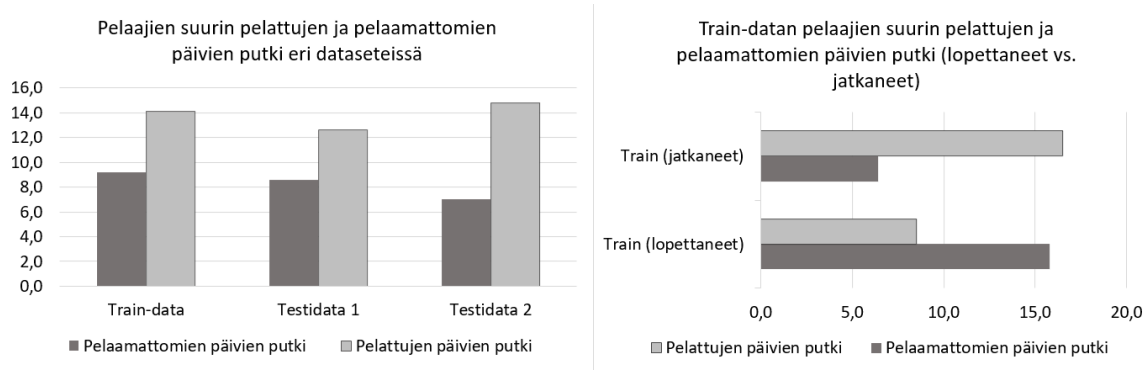
Kuva 6.3: Esimerkkipelaajan pelatut päivät sekä pelattujen ja pelaamattomien päivien putket. Vasemmanpuoleiseen sarakkeeseen on merkitty, onko pelaaja pelannut peliä kyseisenä päivänä vai ei. Punaisella merkitty ”0” tarkoittaa, että pelaaja ei ole pelannut, ja vihreä ”1” tarkoittaa, että pelaaja on pelannut peliä. Vasemmassa sarakkeessa on peräkkäisten samojen arvojen lukumäärän kumulatiivinen summa. Esimerkkipelaajan pisin pelattujen päivien putki on neljä ja pisin pelaamattomien päivien putki on kolme.

$(P-1)$ **1_Streak** max: *Suurin peräkkäisten päivien määrän kumulatiivinen summa, kun kaikkina peräkkäisinä päivinä on pelattu.*

$(P-2)$ **0_Streak** max: *Suurin peräkkäisten päivien määrän kumulatiivinen summa, kun yhtenäkään peräkkäisinä päivänä ei ole pelattu.*

Kaikista kolmesta datasetistä train-datan pelaajien keskimääräinen peliputki ja pelaamattomien päivien putki näyttäisivät olevan pienemmät kuin testisettien vastaavat keskiarvot (kuva 9.3). Testidata 2, joka on kerätty ajalta, jolloin peli on ollut muista dataseteistä poiketen ilmainen, näyttäisi saavan korkeamman keskiarvon pelattujen päivien putkessa. Vastaavasti pelaamattomien päivien putki on testidatassa pienempi.

Jos taas vertaa sellaisia Train-datasetin pelaajia, jotka ovat lopettaneet tai jatkaneet pelaamista, eron näkee melko selkeästi erityisesti pelattujen ja pelaamattomien päivien putkista (kuva 9.4). Pelaajien, jotka ovat lopettaneet pelaamisen, suurin peräkkäisten pelipäivien putki on keskimäärin 9, kun jatkaneiden vastaava arvo on 16. Pelaamattomien päivien putkessa on vielä suurempi ero, sillä lopettaneet ovat pitäneet keskimäärin 16 päivän tauon pelistä, kun peliä jatkaneilla pisin tauko on keskimäärin 6 päivää.



Kuva 6.4: Vasemmanpuoleisessa kuvassa on vertailtu eri datasettien pelaajien pelaamattomien ja pelattujen peliputkien keskiarvoja. Vasemmanpuoleisen kuvan testidatat on muokattu vastaamaan ajallisesti train-datan pituista aikaväliä. Oikeanpuoleisessa kuvassa on mitattu Train-datan lopettaneiden ja jatkaneiden pelaajien peliputkia.

Toinen tapa mitata yhtäjaksoista pelaamista on session pituus. Yhdellä pelisessioilla tarkoitetaan yleensä yhtä pelikertaa, eli aikaväliä pelin käynnistämisen ja sammuttamisen välillä. Session pituuden voi laskea vielä tarkemmin rajaamalla pois ajan, jolloin pelaaja on ollut epäaktiivinen, vaikka peli on ollut päällä. Pelaaja voi suorittaa useita pelisessioita vuorokauden aikana. Session pituus kertoo, kuinka paljon aikaa pelaaja on yhtäjaksoisesti viettänyt pelin parissa.

Kilpailuun mitatussa datassa pelisessiot ovat merkittyinä lokidataan yksilöidyllä tunnisteella, joten niiden laskeminen on datasta vaivatonta. Datan mukaan uusi sessio alkaa siitä, kun pelaaja siirtyy pelialueelle. Näin ollen, mikäli pelaaja pelaamisensa aikana poistuu pelialueelta alkuvalikkoon ja siirtyy sieltä takaisin peliin, hänelle lasketaan alkaneen uusi pelisessio. Pelisessioita voi laskea siis kahdella tavalla: yhtäjaksoisesti pelattuna aikana ottamatta huomioon sitä mitä pelaaja tekee pelissä peliajan aikana tai yhtäjaksoisesti pelissä pelattuna aikana niin, että alkuvalikkoa tai muuta pelialueen ulkopuolista toimintoa ei lasketa mukaan.

Tässä tutkimuksessa on käytetty jälkimmäistä tapaa, eli pelialueella pelaamista, mutta tutkimuksen kannalta olisi ollut kiinnostavaa tutkia kahden pelisessiotavan eroja. Hypoteettisesti voisi ajatella, että pelaajat, jotka ovat kiinnostuneempia pelistä, kävisivät muita pelaajia enemmän alkuvalikossa säätämässä esimerkiksi hahmon asetuksia, ja näin ollen heidän todellinen pelisessionsa jakautuisi useampaan pelialuesessioon.

6.2 Vuorokaudenaika

Pelaajan suosimat vuorokauden ajat luonnollisesti riippuvat paljon pelaajan vapaa-ajan määrästä ja sijoittumisesta viikkorytmiin. Esimerkiksi pelkästään iltapäivisin ja viikonloppuisin pelaaminen voi viitata siihen, että pelaaja käy päivätöissä tai koulussa. Toisaal-

ta vaihteleva vuoronkauden aika voi viitata myös vuorotöihin tai siihen ettei pelaaja käy töissä tai koulussa. Tutkimuksen mukaan se, mihin aikaan pelaaja pelaa peliä, on sidoksissa siihen, kuinka vahvasti pelaaja on addiktoitunut pelaamiseen. Tiberti al.2018 julkaiseman tutkimuksen mukaan erityisesti viikonloppuaamuina pelaaminen korreloi addiktion kanssa [32]. Tässä tutkimuksessa vuorokausi on jaettu neljään aikaan:

aamu	klo. 06.00 - 11.59
päivä	klo. 12.00 - 17.59
ilta	klo. 18.00 - 23.59
yö	klo. 00.00 - 05.59

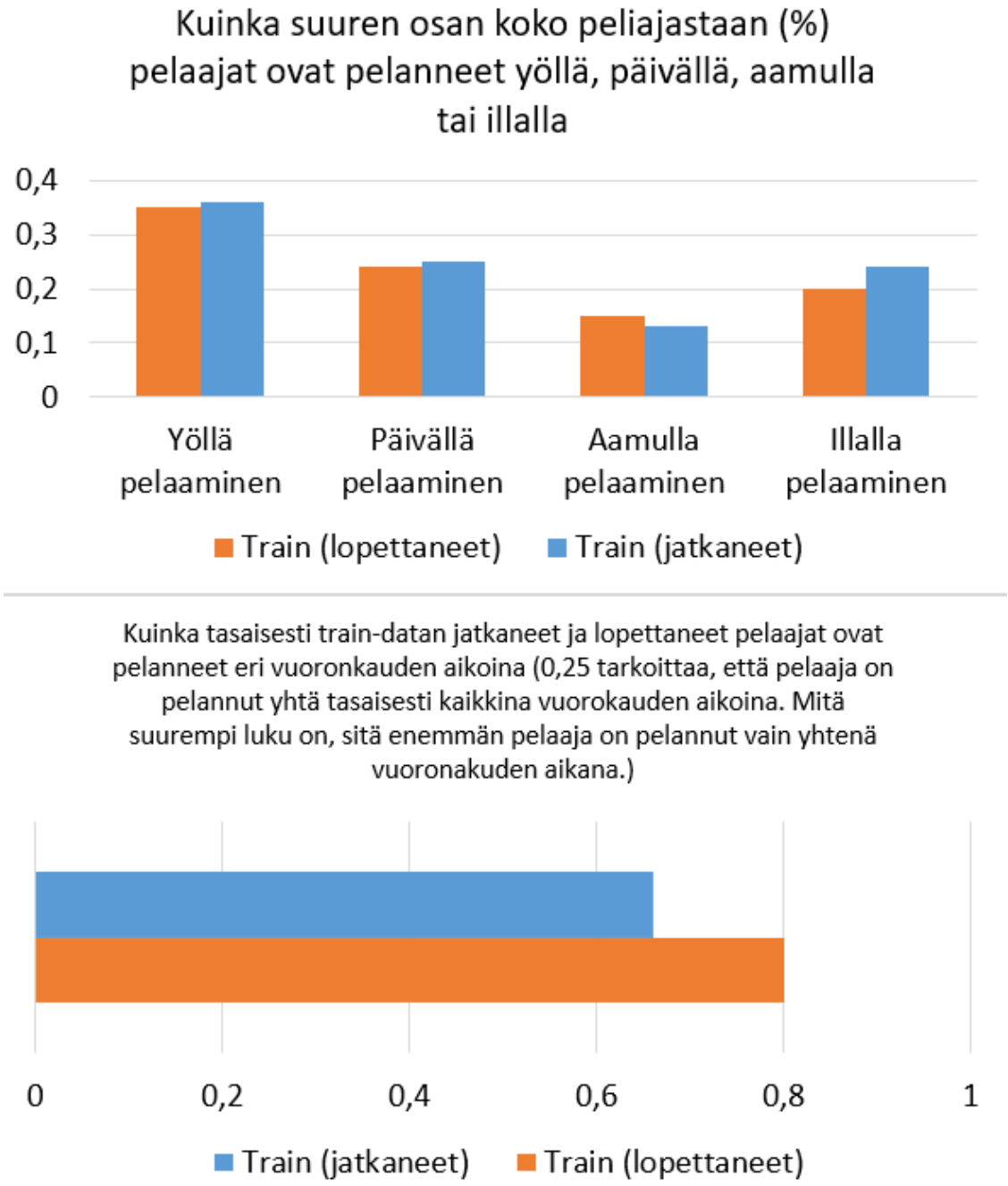
Pelaajilta on laskettu kokonaislokiviestimäärä jokaiselta vuorokaudenajalta ja näin saatu viestimäärä on jaettu pelaajan kokonaisviestimäärällä, jolloin jäljelle jäävä prosentti kuvaa kuinka suuren osan kokonaispeliajasta pelaaja on pelannut tiettynä vuorokaudenaikana. Vuorokauden ajat saattavat kuitenkin olla datassa vääristyneitä, sillä pelaaja saattavat todellisuudessa pelata peliä eri aikavyöhykkeellä kuin millä pelin palvelin sijaitsee.

$(V-1)$ **morning_availability** $\%$:*Kuinka suuren osan koko peliajastaan pelaaja on pelannut aamulla.*

$(V-2)$ **day_availability** $\%$: *Kuinka suuren osan koko peliajastaan pelaaja on pelannut päivällä.*

$(V-3)$ **evening_availability** $\%$:*Kuinka suuren osan koko peliajastaan pelaaja on pelannut illalla.*

$(V-4)$ **night_availability** $\%$: *Kuinka suuren osan koko peliajastaan pelaaja on pelannut yöllä.*



Kuva 6.5: Yläpuolella olevassa kuvassa on kuvattu train-datasetin lopettaneiden ja jatkaneiden pelaajien pelaamista eri vuorokaudenaikoina. Luku on prosentuaalinen osuus koko peliajasta. Alemmassa kuvassa taas on esitetty kuinka tasaisesti pelaajat pelaavat eri vuorokaudenaikoina.

$(V-5)$ **morning_day_evening_night_diff** $\%$: *Kuinka tasaisesti pelaaja on pelannut eri vuorokauden aikoina.*

Laskemalla jokaisen vuorokaudenajan prosentuaalisen peliajan erotuksen 0,25:sta $(1/4)$, ja summaamalla kaikkien erotusten itseisarvot yhteen, saadusta luvusta saadaan selville kuinka tasaisesti pelaaja on pelannut kaikkina neljänä vuorokauden aikana. Luku 0 tarkoittaa sitä, että pelaaja on pelannut peliajastaan tasan 25 % jokaisena neljänä vuorokaudenaikana. Vastaavasti luku 1,5 tarkoittaa sitä, että pelaaja on pelannut koko peliaikansa yhtenä vuorokaudenaikana.

$$x_{morning\_day\_evening\_night\_diff} = |x_{morning\_availability} - 0,25| \\ +|x_{dayh\_availability} - 0,25| + |x_{evening\_availability} - 0,25| + |x_{night\_availability} - 0,25| \tag{6.1}$$

Traindatan jatkaneiden pelaajien vastaavan luvun keskiarvo on 0,66 ja lopettaneiden pelaajien 0,80. Tästä voidaan päätellä, että jatkaneet pelaajat pelaavat tasaisemmin eri vuorokauden aikoina kuin lopettaneet pelaajat. Maksullista ja ilmaista peliä pelanneiden pelaajien peliajassa ei näyttäisi olevan juurikaan eroa (kuva 6.8).

6.3 Peliajan muutos

Peliajan muutoksella pyritään mittaamaan sitä, kuinka pelaajan peliaika muuttuu sen mukaan, mitä pidempään pelaaja on pelannut. Muutosta voi mitata esimerkiksi viikkotasolla seuraamalla kasvaako vai laskeeko pelaajan viikoittainen peliaika. Toisaalta peliajan muutosta voi seurata esimerkiksi sessioiden pituuden muutoksena tai verrata peliaikaa datan mittausajan alusta ja lopusta. Peliajan negatiivinen muutos voi ilmaista esimerkiksi sitä, että pelaajan kiinnostus peliä kohtaan on hiipumassa tai pelaajalla on vähemmän aikaa pelata peliä.

Peliaika voi myös muuttua tiettyjen pelitoiminnallisuuksien osalta. Tällainen toiminnallisuus voi olla esimerkiksi pelaaminen muiden pelaajien kanssa. Mikäli pelaajan yksin pelissä viettämä aika alkaa kasvaa verrattuna tiimiläisten kanssa pelattuun aikaan, tämä saattaa viitata esimerkiksi siihen, että pelaajan pelikaverit ovat lopettaneet pelin pelaamisen. Toisaalta tämä voi tarkoittaa myös sitä, että pelaaja on peliseuraansa enemmän viehättynyt pelistä, joten muutosta ei voi pitää yksiselitteisenä. Samankaltaista muutosta voi mitata myös esimerkiksi kaksintaisteluareenassa vietetystä ajasta tai harjoittelupisteellä vietetystä ajasta.

Wu-chang Feng et al. julkaisivat 2007 tutkimuksen EVE Online-nimisestä MMORPG-pelistä, jossa he löysivät yhteyden pelin lopettamisen ja session pituuden muuttumisen välillä [5]. Tutkimuksen mukaan pelisessioiden pituuden lyheneminen ajan saatossa korreloi sen kanssa, kuinka todennäköisesti pelaaja lopettaa pelin pelaamisen. Vastaavasti myös pelisessioiden välisen ajan piteneminen ajan myötä viittaa siihen, että pelaaja on aikeissa lopettaa pelin pelaamisen. Pelaajien käyttäytymistä on kuitenkin seurattu EVE Online -tutkimuksessa vuosia, joten tutkimus ei ole tämän tutkimuksen kanssa täysin vertailukelpoinen.

Peliajan hajonnalla taas tarkoitetaan sitä, kuinka järjestyksenmukaisesti pelaaja pelaa aina tiettyyn aikaan. Järjestelmällinen peliaika voi tarkoittaa esimerkiksi sitä, että pelaaja pelaa peliä viikon sisällä tiettyinä viikonpäiväni, tai hänen taukonsa pelistä ovat aina yhtä pitkiä. Pelaajan pelisessiot ovat suurin piirtein saman pituisia ja sijoittuvat samoihin vuoronkaudenaikoihin. Yang et al. 2020 julkaiseman tutkimuksen mukaan pelin lopettaneilla pelaajilla on alhaisempia entropia ja enemmän epätasaisuutta peliajassa, kuin pelaajilla, jotka ovat palanneet takaisin pelin pariin [33].

Kummer et al. julkaisivat 2018 tutkimuksen [8], jossa he käyttivät samaa dataa ja samaa ennustustehtävää, kuin mitä tässä tutkimuksessa on käytetty. Tutkimusryhmä käytti merkitsevänä piirteenä lokiviestipiirteiden tendenssiä, jonka he laskivat seuraavan kaavan avulla:

$$Tendenssi = \frac{(((\sum_{i=1}^{n} -1S_1 - S_{n+1}) * -1) + 1)}{2} \tag{6.2}$$

Yhtälössä S_n on piirteen S esiintymien määrä viikolla n. Tässä tutkimuksessa käytetään tendenssin laskemiseen samaa kaavaa, kuin minkä tutkimusryhmä esitti julkaisussaan, mutta viikon sijaan S_n on piirteen S esiintymien määrä päivänä n. Näin saadaan muodostettua tendenssi pelatuille päiville, päivittäisille lokiviesteille ja päivittäisten sessioiden määrälle.

$(T-1)$ **day_played_tendency** : *Pelattujen päivien tendenssi.*

$(T-2)$ **day_msgs_tendency** : *Lokiviestien päivittäisen määrän tendenssi.*

$(T-3)$ **day_sessions_tendency** : *Pelatun päivän sessioiden määrän tendenssi.*

Tämän lisäksi sessioiden pituuden tendenssi laskettiin niin, että n on yksi pelisessio. Piirteen tarkoituksena on tutkia, muuttuuko pelaajan pelisessioiden tendenssi.

$(T-4)$ **session_length_tendency** : *Sessioiden pituuden tendenssi.*

6.4 Ajallisesti poikkeuksellinen pelaajakäyttäytyminen

Uusille pelaajille avautuu pelissä ensimmäisenä tutoriaali, jonka tarkoitus on tutustuttaa pelaaja juoneen ja pelimekaniikkaan. Pelaajaa pyydetään tutoriaalissa toistamaan pelillisiä toimintoja, joiden tavoitteena on tutustuttaa pelaaja pelin toimintaan. Tällaisia toimintoja ovat esimerkiksi aseiden käyttö, kartan avaaminen ja pelihahmon liikuttaminen. Tutoriaali kestää suurin piirtein ensimmäisen kymmenen levelin verran ja sijoittuu Heavens' Ra- nimiselle harjoittelualueelle. Alueelle voi palata koska tahansa pelin aikana, mutta se on tarkoitettu nimenomaan harjoittelemista varten. Tutoriaalipelaajat ovat muihin pelaajiin verrattuna poikkeuksellisia pelaajia, sillä he eivät varsinaisesti itse johda omaa pelaamistaan. Lokidatasta katsottuna kaikki tutoriaalipelaajat ovat siis hyvin samalaisia, ja heidän välillään on vaikea tehdä eroa siitä kuinka viehättyneitä he ovat peliin.

Peliyhtiöt yrittävät yleensä järjestelmällisesti poistaa pelaajia, jotka pelaavat vilpillisesti. Tällaista vilpillistä toimintaa edustaa esimerkiksi botit, jotka ovat pelaajien luomia tietokoneohjelmia, ja jotka hyödyntävät esimerkiksi keinotekoista älykkyyttä pelien pelaamiseen. Botit ovat siis pelissä ilmaantuvia pelihahmoja, joita ei pelaa pelkästään ihminen, vaan joko täysin tietokone, tai ajoittain tietokone ja ajoittain oikea ihminen. Botit ovat usein ajallisesti poikkeuksellisia pelaajia, sillä botit voivat pelata pitkät aikoja keskeyttämättä ja niiden peliaktiivisuus voi olla epäinhimillisen korkea. [34]

Lee EunJon et. al. onnistuivat tutkimuksessaan löytämään joitakin ominaisuuksia, jotka ennustavat bottipelaajia Blade & Souls pelistä. Kolme korkeimman rekressiokertoimen saanutta ominaisuutta olivat lokiviestien kokonaismäärä, lokiviestien itsesimilaarisuus eli samankaltaisuus itsensä kanssa sekä nollakosinisimilaaristen lokiviestien määrä [34]. Kyseisen tutkimuksen avulla on poistettu tämän tutkimuksen datasta sellaiset pelaajat, jotka yllä mainittujen ominaisuuksien valossa näyttäisivät olevan botteja [2]. Tutkimuksen datan pitäisi siis ainakin teoriassa olla bottivapaata.

Toinen poikkeuksellinen vilpillinen pelaajaryhmä ovat kullanlouhijat. Kullanlouhiminen saattaa näyttäytyä ajallisesti hyvin epätavallisena pelaamisena, sillä pelaajan motivaationa toimii menestymisen tai viihteen sijasta taloudellinen voitto. Pelaaja, jonka tavoitteena on maksimoida pelistä saatavan rahan määrä, toimivat pelissä usein hyvin monotonisesti, ja keskittyvät toistamaan muutamaa sellaista pelitoimintoa, joista saa helpoiten kerättyä pelissä rahaa. Tällaisia tekemisiä saattaa olla pelistä riippuen esimerkiksi yksinkertaisten materiaalien ja esineiden kerääminen, ja helppojen vihollisten päihittämistä.

Luku 7

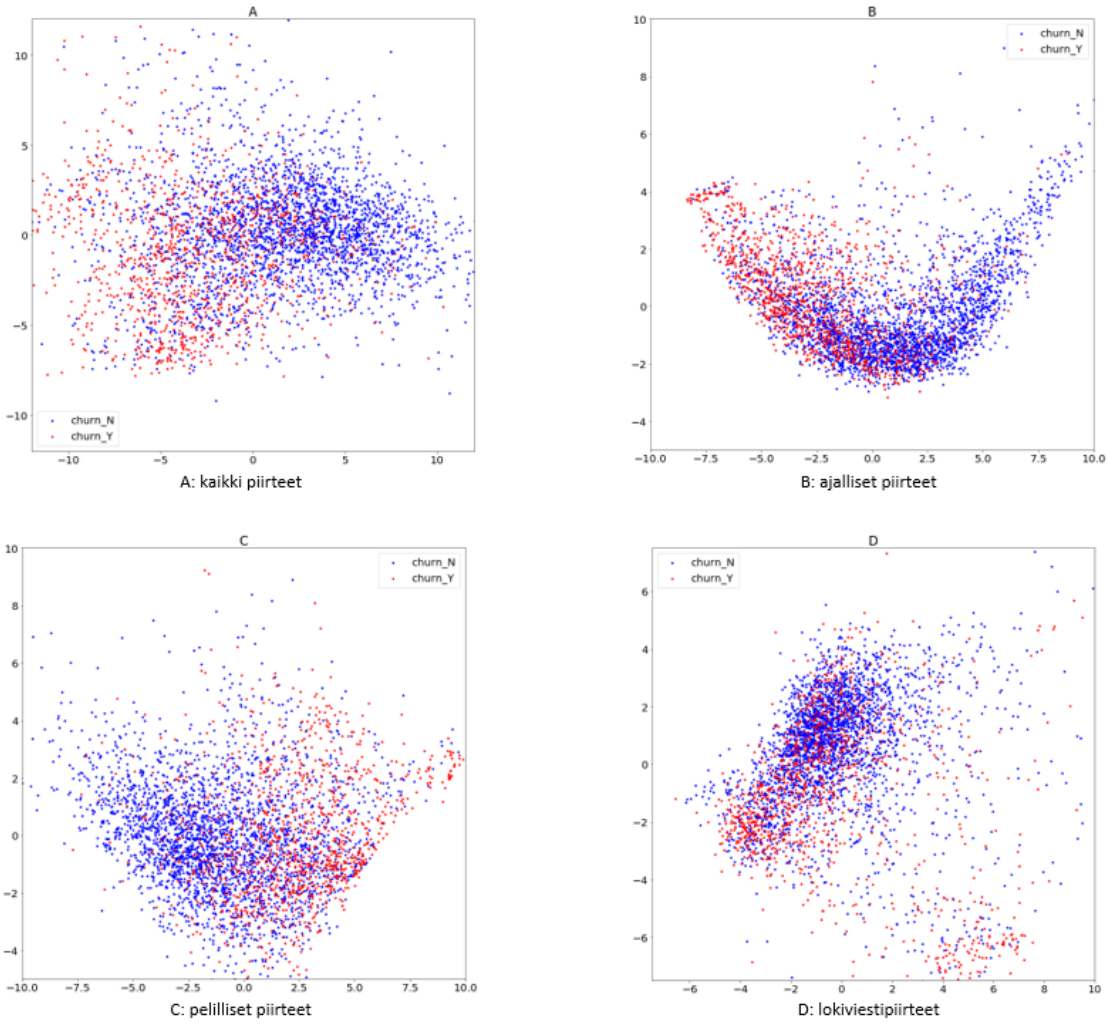
Pelin lopettamisen ennustaminen

7.1 Piirteiden valitseminen

Piirteiden analysoimisessa käytettiin pääkomponenttianalyysia (eng. principal component analys, PCA). Pääkomponenttianalyysia varten piirteet lajiteltiin neljään ryhmään: ryhmä A koostuu kaikista tutkituista piirteistä, ryhmässä B:ssä on vain ajalliset piirteet, ryhmä C koostuu kaikista pelillisistä piirteistä, ja ryhmässä D on lokiviestipiirteet. Tämän lisäksi pelillisten piirteiden ryhmä C on jaettu kolmeen alaryhmään, joista C1 koostuu pelialue-piirteistä, C2 sosiaalisista piirteistä, ja C3 koostuu pelihahmon valintaan ja menestymiseen liittyvistä piirteistä.

A	Kaikki pelaajat	170 piirrettä
B	Ajalliset piirteet	32
C	Pelilliset piirteet	58
C1	Pelialueet	19
C2	Sosiaaliset piirteet	20
C3	Pelihahmo ja menestyminen	20
D	Lokiviestipiirteet	79

Taulukko 7.1: Piirreryhmät ja piirteiden määrä piirreryhmässä.



Kuva 7.1: Traindatan pelaajista kootut pääkomponenttianalyysit piirreryhmistä A, B, C ja D. Lopettaneet pelaajat on merkitty punaisella ja jatkaneet pelaajat on merkitty sinisellä. Piirteet on normalisoitu standardiskaalaajan avulla.

Pääkomponenttianaalyysi on matemaattinen työkalu, jolla korkeadimensionaalin data voidaan esittää kaksi- tai kolmeulotteisena [35]. Kaikki piirreryhmät vaikuttaisivat jakavan train-datan pelaajat PCA:n avulla suhteellisen selkeästi kahteen eri ryhmään eli lopettaneisiin ja jatkaneisiin pelaajiin, mutta piirreryhmä D eli lokiviestipiirteet on kaikista piirreryhmistä selvästi heikoin. Lokiviestipiirteet on kuitenkin määrällisesti suurin piirreryhmä (taulukko 7.1). Kaikki tutkimuksessa tutkitut piirteet ovat esitelty tarkemmin liittessä 1.

Piirteiden valinnassa täytyy ottaa huomioon, että kilpailun tehtävänä on ennustaa testisettien pelin lopettaneet ja peliä jatkaneet pelaajat train-datasetin avulla. Haasteena on erityisesti se, että datasettien mittauksen välillä on kulunut jonkin verran aikaa, ja peli on hieman muuttunut tässä välissä. Datasetit ovat myös mittausaikataulultaan erilaisia ja esimerkiksi testisettejä on kerätty pidemmältä aikaväliltä kuin train-dataa. Mallin riskinä on siis myös piirteiden ylisovitaminen (eng. overfitting) train-dataan, eli malli saattaa sovittaa ratkaisun liian tarkasti train-dataan, eikä onnistu yleistämään mallia testidatojen ennustamiseen.

Univariaatti piirteiden valinta perustuu piirteiden valitsemisella niiden univariaattien statististen arvojen perusteella. $Scikit-learn$ -kirjaston $SelectKBest$ valitsee k parasta piirrettä malliin ja tässä tutkimuksessa parhaiten toimi $f\_classif$, jossa $SelectKBest$ käyttää k parhaan piirteen valitsemiseen ANOVA F-arvoa [22]. $SelectKBest$ -tekniikalla kymmenen parasta train-datan piirrettä ovat: pelitilassa vietetty aika koko ajasta (%), viimeisen viikon pelitilassa vietetty aika koko ajasta (%), päivittäinen pelaamistodennäköisyys, päivittäinen lokiviestimäärä, ensimmäisen datan mittauspäivän pelaaminen, päivittäisen pelaamisen tendenssi, pelattujen päivien määrä, pelaamattomien päivien määrä, suurin pelaamattomien päivien putki ja pelaajan pelihahmojen yhteenlaskettu kokemuspistemäärä. $SelectKBest$ -tekniikalla korostuvat erityisesti ajalliset piirteet.

Univariaattien piirteidenvalintatekniikoiden heikkous on se, että ne eivät mittaa piirteiden välisiä suhteita. Suhteiden mittaamiseen toimii paremmin esimerkiksi rekursiivinen piirteiden eliminointi, joka valitsee piirteet rekursiivisesti karsimalla kaikista piirteistä pois vähiten merkitykselliset piirteet, kunnes jäljellä jää vain toivotun kokoinen osajoukko piirteitä [21].

7.2 Harjanneluokittelija

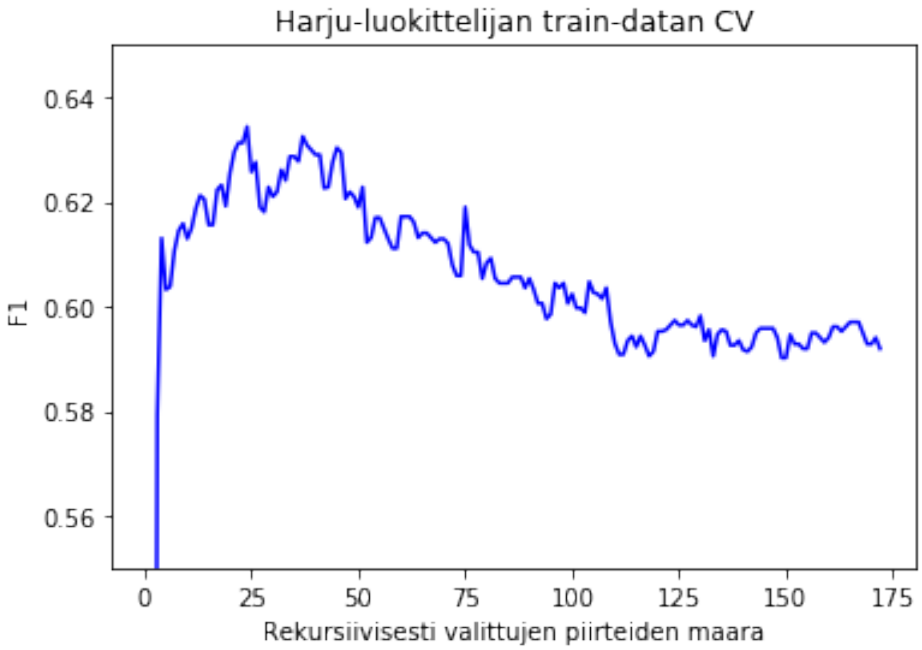
Lineaarisessa regressiomallissa malli sovitetaan pienimmän neliösumman avulla. Toisin sanoen malliksi ennustetaan sellainen datapisteiden kautta kulkeva suora, joka minimoi suoran etäisyyden kaikkiin mallin datapisteisiin [36]. Usean muuttujan lineaarinen regressio voidaan ilmaista:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \epsilon \tag{7.1}$$

jossa β_p regressiokerroin p ja ϵ on mallin residuaali. Harjanneregressio (eng. ridge regressio) on regressiomalli, jossa tappiofunktio on lineaarisen regression tavoin pienimmän neliösumman menetelmä, mutta ylisovittamisen estämiseksi säännönmukaistamiseen käytetään harjuregression sakkoa, eli l2:sta (7.2). Tappiofunktio sakottaa eli pienentää regressiokertoimia β seuraavan yhtälön mukaan [25]:

$$\beta^{ridge} = argmin\left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \tag{7.2}$$

Tässä tutkimuksessa käytetään harjanneluokittelijaa, joka toimii muuten samoin kuin regressiomalli, mutta muuttaa kohdearvot luokittelutehtävään sopiviksi [37]. Luokittelijan parametrien valinnassa käytettiin $Scikit-learn$ -kirjaston $GridSearchCV$-tekniikkaa [38] k-kertaisella ristiinvalidoinnilla ($k = 10$). harjanneluokittelijan parametreiksi valittiin



Kuva 7.2: Train-datan ristiinvalidoinnin F1-tulokset rekursiivisella piirteiden valinnalla (RFECV). Parhaan tuloksen antoi 24 piirrettä.

ne parametrit, jotka antoivat parhaan tuloksen (F1) train-datan ristiinvalidoinnissa niin, että traind-datasetin testisetti on satunnaisesti valittu 25 % alkuperäisestä train-datasta. Parhaan tuloksen antoi harjanneluokittelija, jonka alpha-arvo eli säännönmukaistamiseen käytettävä arvo oli 10, eli luokittelija käyttää vahvaa säännönmukaistusta. Ratkaisijaksi (eng. solver) valikoitui konjugoivaa gradianttia käyttävä $sparse\_cg$ [37]. Luokkapaino on niin ikään balansoitu.

Vastaavasti piirteitä valitessa parhaan train-datan ristiinvalidointituloksen antoi rekursiivisella piirteidenvalinnalla valitut 24 piirrettä. Tässä tutkimuksessa luotuja piirteitä olivat suurin pelattujen ja pelaamattomien päivien putki, päivittäisten lokiviestimäärien tendenssi, pelituntien kokonaismäärä, viimeisen viikon päivittäisten peliminuuttien määrä, joukkueessa ja tiimissä pelaamisen tendenssi, joukkueessa pelaamisen muuttuminen ensimmäisen ja viimeisen viikon välillä, alueella 102 pelattujen lokiviestien määrä ja alu-

eella 108 pelattujen lokiviestien osuus kaikista lokiviesteistä. Train-datan yksinkertaisesta ristiinvalidoinnissa (testidata on 25 % alkuperäisestä train-datasetistä) saadaan F1-menetelmällä laskettu tarkkuus 0,634 (kuva 7.1).

7.3 Logistinen regressio

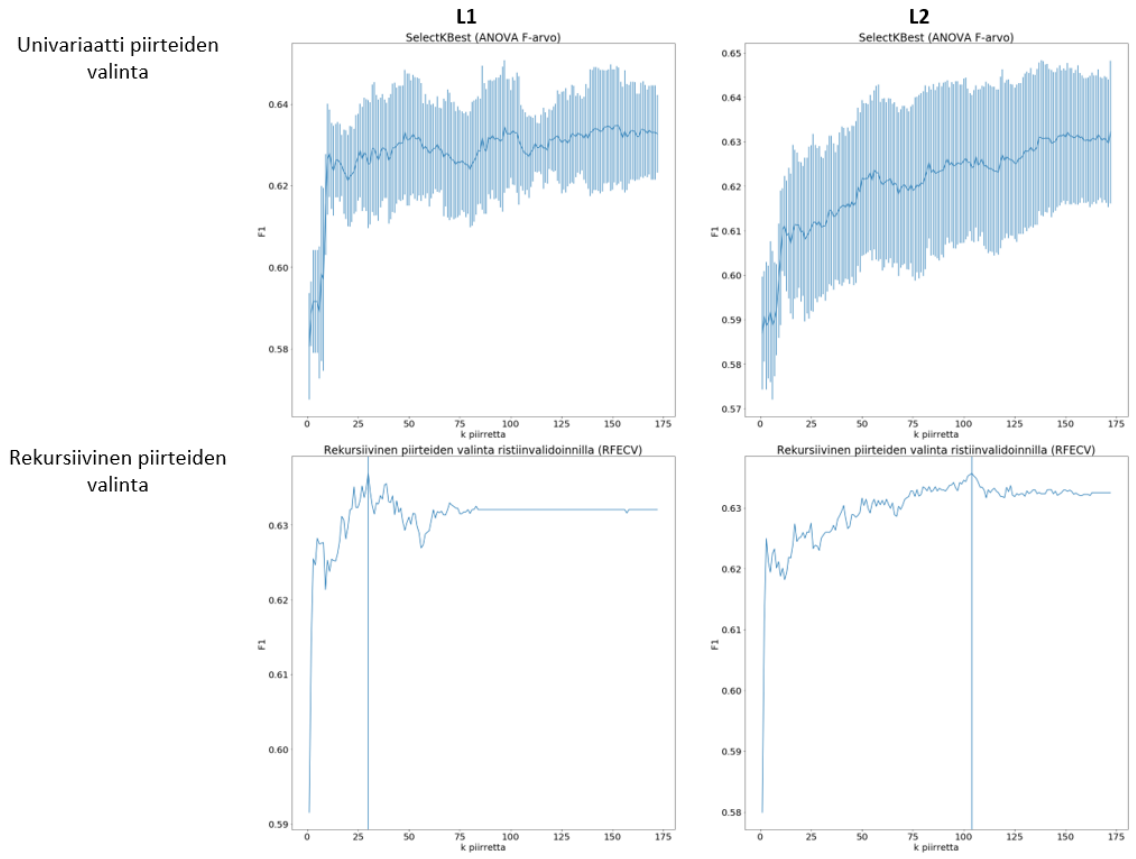
UTU-tiimin käyttämä logistinen regressio on esitelty luvussa 4.1.2 ja tässä tutkimuksessa käytetty regressiomalli mukailee tiimin kilpailutyötä. Yksinkertaisella logistisella regressiolla saavutettiin hieman parempia train-datan ristiinvalidoinnin tuloksia kuin UTU-kilpailutiimi saavutti kilpailutyöllään.

UTU-tiimin käyttämä painonsiirto paransi ristiinvalidointitarkkuutta myös tässä tutkimuksessa. Mallissa käytetään $Scikit-learn$ -kirjaston valmista lineaarista mallia logistiselle regressiolla [39]. Yksinkertaisella ristiinvalidoinnilla logistinen regressio (l1-sakolla) antaa train-datan testisetille F1-arvon 0,629 vain yhdellätoista piirteellä. Piirteet on valittu niin ikään $Scikit-learn$ -kirjaston rekursiivisella piirteiden valinnalla (RFE).

	Piirteiden määrä	**Train-data CV F1**
Logistinen regressio (l1)	11	0.629
Logistinen regressio (l2)	26	0.628

Taulukko 7.2: Logististen regressiomallien tarkkuudet train-datasetin ristiinvalidoinnissa (yksinkertainen ristiinvalidointi). Mallissa kokeiltiin logistista regressiota l1-sakolla ja l2-sakolla.

Valitut piirteet ovat päivittäisten lokiviestien tendenssi, alueilla 102 ja 105 pelatut lokiviestit, peliputkien tendenssi, suurin pelattujen päivien putki, keskiarvoinen prosentuaalinen aika pelitilassa datan viimeisen peliviikon aikana, keskiarvoinen prosentuaalinen aika pelitilassa datan ensimmäisenä päivänä, pelaajan kaikkien pelihahmojen kokemuspis-



Kuva 7.3: Logististen regressiomallien tarkkuudet train-datasetin ristiinvalidoinnissa (kymmeneen osaan ositettu ristiinvalidointi). Rekursiivinen piirteiden valinta on toteutettu RFECV:llä.

teiden summa, pelaajan pelihahmojen kokemuspisteiden keskiarvo, pelaajan pelihahmojen keskiverto kokemuspistemäärä painotettuna pelihahmon pelimäärällä ja kuinka monta kertaa pelaaja on käyttänyt luotua esinettä pelissä verrattuna pelaajan lokiviestien kokonaismäärään. Piirteistä viisi oli uusia tässä tutkimuksessa käytettyjä piirteitä.

7.4 Tukivektorikone

Tukivektorikone (eng. Support Vector Machine) perustuu parhaan korkeaulotteisen tason (eng. hyperplane) löytämiseen vektoreina esitettujen datapisteiden välille. Datapisteiden luokittelussa paras sovitus on siis sellaisen erottelun löytäminen, jossa korkeaulotteisen

tason marginaalin leveys eli paino on mahdollisimman suuri. Marginaaliksi voi työssä ajatella tasoa, joka erottelee mahdollisimman hyvin train-datan lopettaneet ja jatkaneet pelaajat eri puolille marginaalia.

Tutkimuksen mukaan lineaarinen tukivektorikone näyttäisi toimivan datan kanssa parhaiten. Lineaarisen tukivektorikoneen ydin (eng. kernel) on lineaarinen, mikä tarkoittaa sitä, että datapisteet jakava marginaali on suora kulmakertoimella x. Lineaarinen tukivektorikone saa tutkimuksessa parhaimman F1-arvon 0,623 train-datan yksinkertaisessa ristiinvalidoinnissa vain kahdeksalla RFE:llä valitulla piirteellä. Tutkimuksen lineaarinen tukivektorikone käyttää l1-sakkoa.

Parhaan mallin saavuttamat piirteet ovat lokiviestien tendenssi, suurin pelattujen päivien putki, keskiarvoinen prosentuaalinen aika pelitilassa datan viimeisen peliviikon aikana, keskiarvoinen prosentuaalinen aika pelitilassa datan ensimmäisenä päivänä, pelaajan kaikkien pelihahmojen kokemuspisteiden summa, killassa pelaaminen viimeisen viikon aikana ja kuinka monta kertaa pelaaja on käyttänyt luotua esinettä pelissä verrattuna pelaajan lokiviestien kokonaismäärään. Tutkimuksessa ei onnistuttu parantamaan lineaarisen tukivektorikoneen hyperparametrejä, joten tutkimuksessa käytetään tiimi UTU:n tutkimaa mallia.

7.5 Validointi ja tulokset

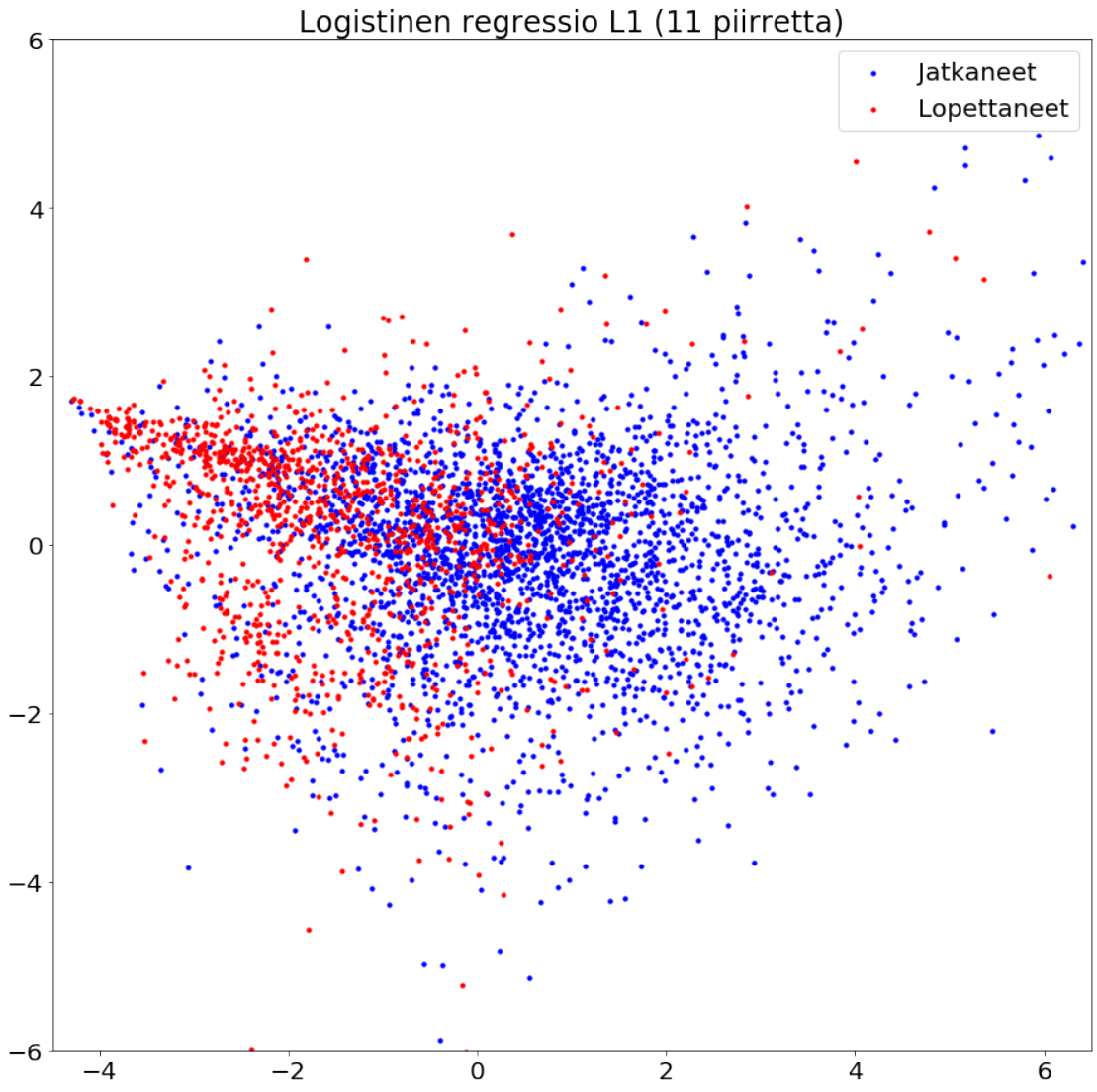
Kilpailun sääntöjen mukaan validointi eli varmistaminen tulee tehdä kilpailussa valmiiksi asetettujen testisettien kanssa. Testisetti 2 poikkeaa train-datasetistä ja ensimmäisestä datasetistä sillä, että se on mitattu ajalta, jolloin peli on ollut pelaajille ilmainen eli F2P-peli. Koneoppimismallien validoidut tulokset löytyvät taulukosta 7.3.

Malli	**Train-datasetti CV (F1)**	**Testisetti 1 (F1)**	**Testisetti 2 (F1)**	**Keskiarvo (F1)**	**Piirteet**
Logistinen regressio (L1)	0.629	0.608	0.617	0.613	11
Tukivektorikone, lineaarinen (L1)	0.623	0.6	0.618	0.609	8
Harjanneluokittelija	0.634	0.6	0.61	0.605	24

Taulukko 7.3: Lopullisten mallien validointi testiseteillä. Taulukossa on ilmoitettu kunkin tekniikan ristiinvalidointitulos ja testidatojen validointitulokset. Keskiarvo on laskettu testisettien tuloksista. Viimeisessä sarakkeessa on ilmoitettu mallin piirremäärä.

Train-datan ristiinvalidoinnissa parhaimman tarkkuuden saanut harjanneluokittelija pärjäsi varsinaisessa validoinnissa muita malleja heikommin, ja tästä voidaan päätellä, että malli ylisovittaa train-dataa kahta muuta mallia enemmän. Parhaiten validaatiossa pärjäsi logistinen regressio, joka saavutti testisettien validoinnissa keskiarvon 0,613. Malli onnistui tukivektorikonetta paremmin yleistämään mallin kumpaakin testisettiin sopivaksi, vaikkakin tukivektorikone sai paremman tarkkuuden F2P-testidatan mallinnuksessa.

Tutkimuksen mukaan pelaajien pelin lopettamisen voi siis ennustaa hyvin pienestä määrästä piirteitä. Valituista piirteistä viisi olivat tässä tutkimuksessa lisättyjä piirteitä: kaksi pelialueeseen liittyvää piirrettä ja kolme ajallista piirrettä, joista kaksi liittyivät peliputkiin ja yksi päivittäisten lokiviestimäärien tendenssiin. Jäljelle jäävistä kuudesta piirteestä kaikki muut olivat tiimi UTU:n mallissa, lukuun ottamatta $actors\_avgexp$ -nimistä piirrettä, joka mittaa pelaajan pelihahmojen kokemuspisteiden keskiarvoa.



Kuva 7.4: Parhaiten pärjänneiden 11 piirteen pääkomponenttianalyysi train-datasetistä. Piirteet ovat logistisen regressiomallin rekursiivisesti valitut piirteet. Piirteet ovat normalisoitu.

Lopullisessa mallissa yhdestätoista piirteestä viisi kuului luokkaan B eli ajallisiin piirteisiin, ja niin ikään viisi luokkaan C, eli pelillisiin piirteisiin. Yksi malliin päätynyt luokan D eli lokiviestipiirreluokan edustajan voi myös laskea pelilliseksi piirteeksi, sillä se luonnollisesti mittaa sitä mitä pelaaja pelissä tekee. Näin ollen näyttäisi siltä, että paras piirreryhmä sisältää suurin piirtein yhtä paljon ajallisia piirteitä ja pelillisiä piirteitä. Huomionarvoista on myös se, että malli ei sisällä yhtään ryhmän C2 piirteitä, eli sosiaalisia piirteitä.

Luku 8

Tulokset

Tässä tutkimuksessa saavutettiin parempi tulos, kun kilpailussa toiseksi tullut kilpailujoukkue UTU saavutti kilpailussa. Tutkimuksen mallin pohja oli sama kuin tiimin kilpailuratkaisu, mutta tarkkuus parani muutamalla lisätyllä älykkäällä piirteellä. Siinä missä UTU-tiimin ratkaisussa oli 18 piirrettä, tämän tutkimuksen paras malli sisälsi vain 11 piirrettä. Oikein valituilla piirteillä vaikuttaisi olevan suuri merkitys mallin tarkkuuteen. Tutkimuksessa tutkittiin peliajallisia piirteitä, pelitoiminnollisia piirteitä, sekä lokiviestien tyypin mukaan mitattuja statistisia piirteitä. Tutkimuksen mukaan paras tarkkuus saavutettiin piirresetillä, joka sisälsi tasaisesti sekä ajallisia piirteitä, että pelillisiä piirteitä.

Ajalliset piirteet eivät ota kantaa siihen millainen peli on, tai mitä pelissä tehdään, ja näin ollen piirteet ovat vähemmän pelikeskeisiä kuin piirteet, jotka sisältävät toimintoja pelimaailmasta. Ajalliset piirteet voisivat kuitenkin sopia hyvin yleiseen malliin. Yleisessä mallissa on se etu, että sen avulla peliyhtiöt voisivat vertailla eri pelien pelaajien peliajan ennustetta toisiinsa. Samoin pelien maksustrategiamuutoksia, kuten Blade & Soul -pelin muuttuminen maksuttomasta maksulliseksi, voisi tutkia yleisellä mallilla. Pienemmätkin muutokset pelissä ja datassa saattavat muokata pelillisten piirteiden käyttöä, ja näin ollen pelkästään yhden peliominaisuuden poistuminen pelistä voi tehdä pelillisiä ominaisuuksia sisältävän mallin käyttökelvottomaksi.

Ajallisten piirteiden lisäksi tutkimuksessa tutkittiin pelillisiä piirteitä, kuten millaisen pelihahmon pelaaja luo, kuinka sosiaalista hänen pelaamisensa on ja millaisilla pelialueilla pelaaja viihtyy pelissä. Pelillisistä piirteistä lopulliseen malliin selvisi kaksi pelialuepiirrettä ja pelaajan pelihahmojen kokemuspisteiden summa ja keskiarvo. Tämän lisäksi malli sisälsi pelaajan pelihahmojen keskivertoisen kokemuspistemäärän painotettuna kunkin pelihahmon pelimäärällä. Kokemuspisteisiin perustuvat piirteet toimivat hyvin yleiseen malliin, sillä kokemuksen kerryttäminen ja mittaaminen on hyvin tyypillinen piirre peleille. Sen sijaan pelialueet, sekä ainoa lopulliseen malliin selvinnyt lokiviestipiirre, joka mittaa sitä kuinka usein pelaaja on käyttänyt pelissä luotua esinettä, ovat hyvin pelikeskeisiä piirteitä.

Tämän tutkimuksen malli saavutti ensimmäisen testisetin kanssa saman tarkkuuden kuin kilpailun voittanut joukkue saavutti kilpailussa. Sen sijaan tutkimus jää aavistuksen F2P-testisetin eli testisetti 2:sen validoinnissa. Tiimi UTU:n malli onnistui ennustamaan kummatkin testisetit yhtä tarkasti, kun taas tämä tutkimus ja voittajatiimin malli onnistuivat F2P-testisetin kanssa paremmin kuin testisetti 1:sen kanssa. Tulos on hieman odottamatonta, sillä train-datasetti on mitattu ajankohtana, jolloin pelin hinnoittelustrategia on ollut yhtenevä testidata 1:sen kanssa, mutta ei ilmaisen testidata 2:sen kanssa.

Tutkimus antaa myös viitteitä siitä, että paremmalla piirrevalikoimalla voitaisiin saavuttaa vielä parempia tuloksia. Tutkimuksessa ei tutkittu kaikkia mahdollisia piirteitä, ja näin ollen piirteiden tutkimusta voisi jatkokehittää. Jatkotutkimuksella tiimi UTU:n mallilla voitaisiin saavuttaa kilpailun voittajajoukkuetta parempi tarkkuus. Tutkimuksessa ei kuitenkaan onnistuttu parantamaan tiimi UTU:n logistista regressiota, tai löytämään toista haastajaa mallille. Tutkimuksessa käsiteltiin kolmea parhaiten menestynyttä tekniikkaa, jotka olivat logistista regressio, harjanneluokittelija ja lineaarinen tukivektorikone. Lisäksi tutkimuksessa tutkittiin erilaisia lasso- ja harjuregressioita sekä päätöspuutekniikoita.

Tiimi	**Tekniikka**	**Testisetti 1 (F1)**	**Testisetti 2 (F1)**	**Piirteiden määrä**
Yokozuna Data	Syväoppiminen ja päätöspuutekniikka	0.61	0.63	500
Tämä tutkimus	Logistinen regressio	0.61	0.62	11
UTU	Logistinen regressio	0.6	0.6	18
TripleS	Päätöspuutekniikka	0.57	0.62	?
TheCowKing	Päätöspuutekniikka	0.59	0.6	?
goedleio	Syväoppiminen ja päätöspuutekniikka	0.57	0.6	600

Taulukko 8.1: Kilpailun ensimmäisen tehtävän tulokset ja tutkimuksen sijoittuminen kilpailun tuloksiin nähden.

Kaikki kolme lopulliseen tutkimukseen päätynyttä mallia ovat melko samanlaisia. Toisaalta tämä tutkimus myös paljastaa, että koska tutkimuksessa saavutettiin hyvin lähelle sama tulos, kuin kilpailun voittajajoukkue saavutti kilpailussa, voittajatiimin Yokozuna Datan käyttämä syväoppimismalli ei ole paras ratkaisu kilpailukysymykseen. Tämä tutkimus näyttäisikin vievän pohjan Yokozuna Datan kilpailuratkaisulta ja ratkaisusta kirjoitetusta artikkelista [19], jossa käytetään syväoppimista piirteiden valintaan ja äärimmäisen satunnaistettua päätösmetsää pelaajien lopettamisen ennustamiseen. Koska tämä tutkimus ja tiimi UTU:n regressiomalli osoittautuivat lähes yhtä toimiviksi kuin voittajatiimin monimutkainen malli, vaikuttaisi siltä, että syväoppimismallit tai satunnaistetut päätösmetsät eivät tuo lisäarvoa tutkimukselle.

Monimutkaisissa malleissa on myös se heikkous, että ne toimivat usein mustan laatikon tavoin, eivätkä tarjoa tietoa siitä miksi pelaajat ovat lopettaneet pelaamisen. Sen sijaan hyvin valitut piirteet saattavat tarjota asiantuntijoille tärkeää tietoa siitä miten lopettavien

ja jatkavien pelaajien pelaaminen eroaa toisistaan. Näin ollen mallit antavat pelkän ennustamisen lisäksi myös osviittaa syistä, joiden takia pelaajien kiinnostus peliä kohtaan on laskenut. Tämä on arvokasta peliyhtiöille, jotka voivat tiedon avulla kehittää peliään tai kohdentaa pelin mainontaa.

MMORPG-pelien datanlouhinnassa onkin kiinnostavaa pohtia, onko datapohjaista tutkimusta järkevää tehdä, mikäli tietopohjainen tutkimus on mahdollista. Tietopohjaisessa tutkimuksessa keskitytään etsimään kysymystä ennustavia piirteitä, kun datapohjaisessa tutkimuksessa tutkitaan mahdollisimman suurtaa raakaa datapohjaa. Puhtaassa datapohjaisessa datan louhinnassa on se etu, että piirteiden prosessointi on nopeampaa, eikä vaadi aiheen asiantuntijuutta. Datapohjainen tutkimus ei ole niin altista inhimilliselle virheelle kuin tietoon pohjautuva piirteiden prosessointi. Datapohjaisen koneoppimisen huono puoli on taas se, että se toimii mustan laatikon tavoin.

Tässä tutkimuksessa päädytään melko samanlaiseen lopputulokseen kuin Borboran et al. päätyivät tutkimuksessa "Churn Prediction in MMORPGs using player motivation theories and an ensemble approach". Tutkimus on esitelty kappaleessa 2. Tutkimusryhmän mukaan malli, joka oli rakennettu useasta niin sanotusti suoraan datasta otetusta piirteestä, oli vain hieman tarkempi kuin malli, joka sisälsi vain muutaman tarkasti tiedon pohjalta valitun piirteen. oletettavasti tahoilla, jotka ovat kiinnostuneita ennustamaan tämän tutkimuksen kaltaisia kysymyksiä, omistavat melko paljon tiedollista pääomaa aiheesta. Näin ollen dataan perustuva tutkimus ei välttämättä ole tämän kaltaisille tahoille optimaalinen ratkaisu.

Viitteet

[1] NCSOFT West. https://us.ncsoft.com/en-us. Katsottu: 2020-02-08.

[2] Eunjo Lee, Yoonjae Jang, Du-Mim Yoon, Jihoon Jeon, Seong-il Yang, Sang-Kwang Lee, Dae-Wook Kim, Pei Pei Chen, Anna Guitart, Paul Bertens ja et al. Game Data Mining Competition on Churn Prediction and Survival Analysis Using Commercial Game Log Data. *IEEE Transactions on Games*, 11(3):215–226, Sep 2019. ISSN 2475-1510. URL http://dx.doi.org/10.1109/TG.2018.2888863.

[3] Z. Borbora, J. Srivastava, K. Hsu ja D. Williams. Churn Prediction in MMORPGs Using Player Motivation Theories and an Ensemble Approach. Teoksessa *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, ss. 157–164. 2011.

[4] Holin Lin ja Chuen-Tsai Sun. Cash Trade in Free-to-Play Online Games. *Games and Culture*, 6(3):270–287, 2011.

[5] Wu-chang Feng, David Brandt ja Debanjan Saha. A Long-Term Study of a Popular MMORPG. Teoksessa *Proceedings of the 6th ACM SIGCOMM Workshop on Network and System Support for Games*, NetGames '07, s. 19–24. Association for Computing Machinery, New York, NY, USA, 2007. ISBN 9780980446005. URL https://doi.org/10.1145/1326257.1326261.

[6] J. Kawale, A. Pal ja J. Srivastava. Churn Prediction in MMORPGs: A Social Influence Based Approach. Teoksessa *2009 International Conference on Computational Science and Engineering*, osa 4, ss. 423–428. 2009.

[7] Z. H. Borbora ja J. Srivastava. User Behavior Modelling Approach for Churn Prediction in Online Games. Teoksessa *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, ss. 51–60. 2012.

[8] L. B. Martins Kummer, J. Cesar Nievola ja E. C. Paraiso. Applying Commitment to Churn and Remaining Players Lifetime Prediction. Teoksessa *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, ss. 1–8. 2018.

[9] E. Lee, B. Kim, S. Kang, B. Kang, Y. Jang ja H. K. Kim. Profit Optimizing Churn Prediction for Long-Term Loyal Customers in Online Games. *IEEE Transactions on Games*, 12(1):41–53, 2020.

[10] L. Shi ja W. Huang. Apply social network analysis and data mining to dynamic task synthesis for persistent MMORPG virtual world. *Lecture Notes in Computer Science*, 3166:204–215, 2004.

[11] Togelius J. Yannakakis G.N. Bauckhage C Drachen A., Thurau C. *Game Data Mining*. 2013.

[12] Nick Yee. Motivations for Play in Online Games. *CyberPsychology Behavior*, 9(6):772–775, 2007.

[13] Fabian Hadiji, Rafet Sifa, Anders Drachen, Christian Thurau, Kristian Kersting ja Christian Bauckhage. Predicting player churn in the wild. Teoksessa *2014 IEEE Conference on Computational Intelligence and Games*, ss. 1–8. IEEE, 2014. ISBN 9781479935475. ISSN 2325-4270.

[14] K. Rothmeier, N. Pflanzl, J. Hüllmann ja M. Preuss. Prediction of Player Churn and Disengagement Based on User Activity Data of a Freemium Online Strategy Game. *IEEE Transactions on Games*, ss. 1–1, 2020.

[15] Announcing Blade Soul: NCsoft to Introduce Western Gamers to Asian Martial Arts Fantasy with Blade Soul®. https://web.archive.org/web/20120917072030/http://us.bladeandsoul.com/en/news/. Katsottu: 2020-03-13.

[16] Blade & Souls: Factions. https://www.bladeandsoul.com/en/game/factions. Katsottu: 2020-12-30.

[17] Blade & Souls: Races. https://www.bladeandsoul.com/en/game/races. Katsottu: 2020-12-30.

[18] Blade & Soul wiki: The World. Haettu osoitteesta: https://bladeandsoul.gamepedia.com/World_map. Katsottu: 2019-10-07.

[19] Anna Guitart, Pei Pei Chen ja África Periáñez. The Winning Solution to the IEEE CIG 2017 Game Data Mining Competition. *Machine Learning and Knowledge Extraction*, 1(1):252–264, 2019. ISSN 2504-4990. URL https://www.mdpi.com/2504-4990/1/1/16.

[20] Scikit-learn: StandardScaler. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html. Katsottu: 2020-02-04.

[21] Scikit-learn: RFE. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html#sklearn.feature_selection.RFE. Katsottu: 2020-01-30.

[22] Scikit-learn: SelectKBest. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.

SelectKBest.html#sklearn.feature_selection.SelectKBest. Katsottu: 2020-01-04.

[23] J.M. Hilbe. *Logistic Regression Models*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2009. ISBN 9781420075779. URL https://books.google.fi/books?id=tmHMBQAAQBAJ.

[24] Xin Yan ja Xiaogang Su. Linear Regression Analysis: Theory And Computing. Singapore: World Scientific Publishing Company; 2009.

[25] Tibshirani R Hastie T ja Friedman J. The Elements of Statistical Learning : Data Mining, Inference, and Prediction, Second Edition : Data Mining, Inference, and Prediction, Second Edition. New York, NY: Springer New York; 2017.

[26] Scikit-learn: cross_val_score. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_val_score.html. Katsottu: 2020-01-31.

[27] Scikit-learn: StratifiedKFold. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html. Katsottu: 2020-01-16.

[28] Adam S. Kahn, Cuihua Shen, Li Lu, Rabindra A. Ratan, Sean Coary, Jinghui Hou, Jingbo Meng, Joseph Osborn ja Dmitri Williams. The Trojan Player Typology: A cross-genre, cross-cultural, behaviorally validated scale of video game play motivations. *Computers in Human Behavior*, 49:354–361, 2015. ISSN 0747-5632. URL https://www.sciencedirect.com/science/article/pii/S0747563215002046.

[29] Blade & Souls Wiki: World map. https://bladeandsoul.gamepedia.com/World_map. Katsottu: 2020-12-30.

[30] R. Heeks. Current Analysis and Future Research Agenda on “Gold Farming”: Real-World Production in Developing Countries for the Virtual Economies of Online Games. (32), 2008.

[31] Blade & Souls Wiki: Training System. https://bladeandsoul.gamepedia.com/Training_System. Katsottu: 2020-02-17.

[32] Stefano Triberti, Luca Milani, Daniela Villani, Serena Grumi, Sara Peracchia, Giuseppe Curcio ja Giuseppe Riva. What matters is when you play: Investigating the relationship between online video games addiction and time spent playing over specific day phases. *Addictive Behaviors Reports*, 8:185 – 188, 2018. ISSN 2352-8532. URL http://www.sciencedirect.com/science/article/pii/S235285321830035X.

[33] W. Yang, T. Huang, J. Zeng, L. Chen, S. Mishra ja Y. Liu. Utilizing Players' Playtime Records for Churn Prediction: Mining Playtime Regularity. *IEEE Transactions on Games*, ss. 1–1, 2020.

[34] Eunjo Lee, Jiyoung Woo, Hyoungshick Kim, Aziz Mohaisen ja Huy Kang Kim. You are a Game Bot!: Uncovering Game Bots in MMORPGs via Self-similarity in the Wild. 10.14722/ndss.2016.23436, 01/ 2016.

[35] Markus Ringnér. What is principal component analysis? *Nature Biotechnology*, 26:303–304, 2008. ISSN 1546-1696.

[36] D.C. Montgomery, E.A. Peck ja G.G. Vining. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 2012. ISBN 9780470542811. URL https://books.google.fi/books?id=0yR4KUL4VDkC.

[37] Scikit-learn: RidgeClassifier. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeClassifier.

html#sklearn.linear_model.RidgeClassifier. Katsottu: 2020-01-30.

[38] Scikit-learn: GridSearchCV. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. Katsottu: 2020-01-30.

[39] Scikit-learn: Logistic Regression. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html. Katsottu: 2020-012-10.

Piirreluokka	*Piirteen nimi*	*Piirteen selitys*	*Onko piirre uusi vai tiimi-UTU:n tekemä?*	*Oliko piirre tämän tutkimuksen malliss?*	*Oliko piirre tiimi UTU:n mallissa?*
B	actors_time	Kaikkien pelattujen hahmojen peliaikojen summa.	UTU		
C3	actors_played	Pelattujen hahmojen määrä.	UTU		
C3	actors_deleted	Poistettujen hahmojen määrä.	UTU		
B	actors_diversity	Kuinka tasaisesti pelaaja on pelannut kaikkia pelihahmojaan.	UTU		
B	actors_totalexp	Kaikkien pelattujen hahmojen kokemuspisteidensumma.	UTU	X	X
B	actors_maxexp	Käyttäjän suurin kokemuspistemäärä pelihahmolla.	UTU		
B	actors_avgexp	Pelaajan pelihahmojen maksimaalisten kokemuspisteiden keskiarvo.	UTU		
B	actors_play_ avgexp	Pelaajan kaikkien pelihahmojen keskimääräisten kokemuspistemäärien summa. Hahmon keskimääräinen kokemuspistemäärä on hahmon suurimman ja pienimmän kokemuspistemäärän keskiarvo. Pelaajan kaikkien pelihahmojen keskimääräistä kokemuspistemääräsummaa laskiessa kunkin hahmon arvoa painotetaan sen mukaan, kuinka suuren osan pelaaja on pelannut pelistä kyseisellä hahmolla.	UTU		X
B	actors_add_ totalexp	Pelaajan kaikkien pelihahmojen suurimman ja pienimmän kokemuspistemäärän erotusten summa.	UTU		
B	actors_add_ maxexp	Pelaajan suurimman ja pienimmän kokemuspistemäärän erotus.	UTU		
B	actors_start _newbie	Pelaajan pienin level on 1 (uusi pelaaja).	UTU		
C3	actors_rating _avg	Pelaajan keskivertoinen menestyminen (eng.rating) kaksintaisteluissa.	UTU		X
C3	actors_money_ avg	Pelaajan keskivertoinen pelirahamäärä.	UTU		X
C3	actors_has_ party	Onko pelaaja pelannut joukkueessa.	UTU		
C3	actors_n_ guild/team/party	Joukkueiden/tiimien/kiltojen määrä.	UTU		X (1 kpl)
B	actors_start_ maxexp	Pelaajan pienimmän levelin kokemuspistemäärä (eli kuinka paljon pelaajalla on kokemuspisteitä ennen datan keräämistä.	UTU		X
B	day_availability	Keskiarvoinen prosentuaalinen aika pelitilassa.	UTU		X
B	end_availability	Keskiarvoinen prosentuaalinen aika pelitilassa viimeisen 10 päivän ajalta.	UTU	X	X
B	inc_availability	Pelitilassa vietetyn ajan muuttuminen pelattujenpäivien myötä.	UTU		X
B	day_probability	Kuinka monena päivänä kaikista päivistä pelaajan pelannut peliä.	UTU		X

B	**end_probability**	**Kuinka monena päivänä pelaaja on pelannut peliä viimeisen viikon aikana.**	**UTU**		**X**
B	inc_probability	Päivittäisen pelaamisen muutos.	UTU		
B	day_msgs	Lokiviestien määrä päivässä.	UTU		
B	end_msgs	Lokiviestien määrä päivässä viimeisen viikon aikana.	UTU		X
B	inc_msgs	Lokiviestien päivittäisen määrän muutos.	UTU		
B	session_length	Pelisession pituus.	UTU		
B	first_day	Keskiarvoinen prosentuaalinen aika pelitilassa datan ensimmäisenä päivänä.	UTU		X
B	last_day	Kuinka monena päivänä pelaaja pelasi viimeisen viikonaikana.	UTU		X
D	log[id]_time	Lokiviestityypin pelaamiseen mennyt aika jaettuna pelaajankokonaispeliajalla (79 piirrettä).	UTU	X (3 kpl)	X (4 kpl)
B	avg_min_played	Päivässä pelattujen minuuttien keskiarvo (min).	uusi		
B	hours_played	Pelattujen tuntien määrä.	uusi		
B	days_played	Pelattujen päivien kokonaisarvon keskiarvo.	uusi		
B	days_unplayed	Pelaamattomien päivien kokonaisarvon keskiarvo.	uusi		
B	played_Streak	Suurin peräkkäisten päivien määrän kumulatiivinen summa,kun kaikkina peräkkäisinä päivinä on pelattu.	uusi	X	
B	unplayed_Streak	Suurin peräkkäisten päivien määrän kumulatiivinen summa,kun yhtenäkään peräkkäisinä päivänä ei ole pelattu.	uusi		
B	morning/daytime/ evening/night _availability	Kuinka suuren osan koko peliajastaan pelaaja onp elannut aamulla/päivällä/iltapäivällä/yöllä (4 piirrettä).	uusi		
B	playingtime_diff	Kuinka tasaisesti pelaaja on pelannut eri vuorokauden aikoina.	uusi		
B	days_tendency	Pelattujen päivien tendenssi.	uusi		
B	msgs_tendency	Lokiviestien päivittäisen määrän tendenssi.	uusi	X	
B	session_tendency	Pelatun päivän sessioiden määrän tendenssi.	uusi		
B	streak_tendency	Sessioiden pituuden tendenssi.	uusi	X	
B	session_length	Session keskimääräinen pituus.	uusi		
C1	msgs_zone_x	Lokiviestien määrä pelialueella x (10 piirrettä).	uusi	X (2 kpl)	
C1	msgs_zone_x_%	Prosentuaalinen peliaika alueella x pelaajan kokonaispeliajasta (10 piirrettä).	uusi		
C2	guild/team/party _availability	Pelaaminen killassa, tiimissä tai joukkueessa suhteessa koko peliaikaan.	uusi		
C2	end_guild/team/party _availability	Pelaaminen killassa, tiimissä tai joukkueessa suhteessa koko peliaikaan (vain viimeinen viikko).	uusi		
C2	first_guild/team/party _availability	Killassa, tiimissä tai jouk-kueessa pelaamisen ero ensimmäisen ja viimeisen viikon välillä).	uusi		
C3	deleted_actors_%	Poistettujen hahmojen osuus pelaajan hahmoista (kunpelihahmojen määrä > 1).	uusi		
C3	num_actors	Pelihahmojen lukumäärä.	uusi		