


Turvallisuusjohtamisen kehittäminen tiedon louhinnalla

AI SAFETY -HANKKEEN LOPPURAPORTTI



Lantto Eero
Tiikkaja Maria
Haavisto Olli
Heikkilä Tarja
Kannisto Henriikka
Kivimäki Ilkka
Kärkimaa Jukka
Nurmi Akseli
Puro Vuokko
Räsänen Tuula
Virtanen Simo

Turvallisuusjohtamisen kehittäminen tiedon louhinnalla

AI SAFETY -hankkeen loppuraportti

Lantto Eero
Tiikkaja Maria
Haavisto Olli
Heikkilä Tarja
Kannisto Henriikka
Kivimäki Ilkka
Kärkimaa Jukka
Nurmi Akseli
Puro Vuokko
Räsänen Tuula
Virtanen Simo

Työterveyslaitos

www.ttl.fi

PL 40

00251 Helsinki

www.ttl.fi

© 2023 Työterveyslaitos ja kirjoittajat

Hanke on toteutettu Työsuojelurahaston tuella.

Tämän teoksen osittainenkin kopiointi on tekijänoikeuslain (404/61, siihen myöhemmin tehtyine muutoksineen) mukaisesti kielletty ilman asianmukaista lupaa.

ISBN 978-952-391-135-2 (PDF)

Tiivistelmä

Turvallisuusjohtamisen kehittäminen tiedonlouhinnalla (AI Safety)- tutkimushanke tarkasteli koneoppimismenetelmien hyödyntämismahdollisuuksia turvallisuusjohtamista tukevinä työkaluina. Koneoppimismenetelmien ohella hanke tarkasteli myös organisaatioiden erilaisten datojen hyödyntämispotentiaalia eri analyyseissä. Keskeistä hankkeen tutkimusasetelmassa oli näiden erilaisten aineistojen yhdistäminen analyysikohtaisiksi aineistokokonaisuuksiksi. Tämän tarkoituksena oli etsiä sellaisia yhteyksiä datojen väliltä, jotka eivät ole ilmeisiä, tai jotka ovat huonosti tunnettuja. Koneoppimismenetelmiin on vuosien saatossa liitetty paljon odotuksia esimerkiksi ennakoivan analytiikan suhteen. Yksi tämän hankkeen tavoitteista olikin juuri sen selvittäminen, mitä ennakoivan analytiikan mahdollisuuksia organisaatioissa nykyisin olemassa oleva data mahdollistaa. Toisaalta hankkeessa myös pyrittiin yleisesti ottaen selvittämään monipuolisesti eri menetelmillä, voiko koneoppimismenetelmillä tuottaa turvallisuusjohdolle uutta tietoa vuosien saatossa organisaation normaalista toiminnasta kertyneestä datasta. Hankkeen aineisto tuli alueilta työturvallisuus, työympäristö, henkilöstö ja tuotanto ja se kerättiin hankkeen neljästä osallistujaorganisaatiosta, jotka ovat Suomessa toimivia teollisuusyrityksiä. Tutkimusaineistona käytettiin myös Tapaturmavakuutuskeskuksen vahinkokuvaustekstejä. Aineistoa analysoitiin monipuolisesti erilaisilla koneoppimismenetelmillä, ja osaa aineistoa analysoitiin myös asiantuntija-analyysillä. Asiantuntija-analyysillä tutkittiin, miten monipuolisesti organisaatiot ovat tunnistaneet tapaturmiin liittyneitä inhimillisiä ja ihmisen toiminnan taustalla vaikuttaneita tekijöitä.

Hanke syvensi ymmärrystä osallistujaorganisaatioiden datojen mahdollisuuksista ja puutteista koneoppimismenetelmiin perustuvassa analyysissä, ja tuotti tietoa rakenteisen ja rakenteettoman datan hyödyntämismahdollisuuksista koneoppimismenetelmillä. Tulosten perusteella organisaatioiden keräämässä turvallisuusdatassa on kehitettävää ja tietoaaukkoja, jotka vaikeuttavat tekoälyn hyödyntämistä turvallisuuteen liittyvässä analytiikassa ja siten datalähtöisessä turvallisuusjohtamisessa. Kehitettävää havaittiin tekstimuotoisen datan laadussa ja kattavuudessa – esimerkiksi turvallisuustapahtumien taustalla vaikuttavien tekijöiden tunnistamisessa – rakenteisen datan määrien ja muuttujien riittävydessä sekä mahdollisuuksissa yhdistää rakenteisia dataja toisiinsa. Turvallisuusdatan laadun ja muun organisaatiodatan (kontekstidatan) saatavuuden ja yhdisteltävyyden kehittäminen onkin keskeistä, mikäli tekoälyyn perustuvaa analytiikkaa halutaan hyödyntää turvallisuusjohtamisessa. Dataan liittyvistä haasteista huolimatta tutkimuksessa käytetyt koneoppimismallit tunnistivat kuitenkin merkkejä siitä, että esimerkiksi turvallisuuskoulutuksilla ja perehdytyksillä pystytään vaikuttamaan

työturvallisuuteen, ja että korkeat tuotanto- ja tilausmäärät näyttävät yleisellä tasolla olevan yhteydessä työturvallisuustason heikentymiseen. Analyysimenetelmillä havaittiin myös, että kohonneet turvallisuushavaintomäärät indikoivat kasvua työturvallisuustapahtumien määrässä ja vakavuudessa, mikä viittaa siihen, että vakavampia turvallisuustapahtumia kuten työtapaturmia voitaisiin estää, mikäli työturvallisuuden korjaavia toimenpiteitä pystytään tekemään aktiivisesti ja ennakoiden turvallisuushavainnoilla saatavan tiedon perusteella.

Abstract

The development of safety management through data mining (AI Safety) research project examined the potential use of machine learning methods as supportive tools for safety management. In addition to machine learning methods, the project also explored the potential of utilising various types of organisational data in different analyses. A key aspect of the research framework of this project was the combination of these various datasets into analysis-specific datasets. The purpose was to explore connections between the data that may not be obvious or well-known. Over the years, machine learning methods have been associated with high expectations, especially in predictive analytics. One of the goals of this project was to determine the possibilities of predictive analytics that existing data in organizations currently enables. On the other hand, the project also aimed to comprehensively investigate whether machine learning methods can provide safety management with new insights from data accumulated over the years from normal organizational operations. The data for the project came from the areas of occupational safety, work environment, personnel, and production and was collected from four participating industrial organisations in Finland. The data was analysed using various machine learning methods, and some of the data was also analysed through expert analysis. Expert analysis examined how comprehensively organisations have identified human factors and factors influencing human actions related to accidents.

The project deepened the understanding of the potential and limitations of participant organisations' data in machine learning-based analysis and provided insight to the utilisation possibilities of structured and unstructured data with machine learning methods. Based on the results, there are deficiencies and data gaps in organisations' safety data that hinder the utilisation of machine learning in safety management analytics and, thus, data-driven safety management. Deficiencies were observed in the quality and coverage of textual data, such as identifying factors behind safety events, as well as in the adequacy of structured data quantities and variables and the ability to combine structured data with each other. Therefore, improving the quality of safety data and the availability and compatibility of other organisational data (contextual data) is crucial if machine learning-based analytics is to be used in safety management. Despite data-related challenges, the machine learning models used in the study identified signs that safety training and orientation, for example, can have an impact on workplace safety, and that high production and order quantities appear to be generally associated with a decrease in safety levels. The analysis methods also revealed that increased safety observation counts indicate an increase in the number and severity of safety incidents, suggesting that more serious safety events, such as workplace

accidents, could be prevented if corrective measures are actively taken based on the information obtained from safety observations.

Kiitokset

Turvallisuusjohtamisen kehittäminen tiedon louhinnalla – AI Safety -tutkimushanke toteutettiin vuosien 2020–2023 aikana. Haluamme kiittää Työsuojelurahastoa AI Safety -tutkimushankkeen rahoituksesta. Hankkeen suunnitteluun ja toteuttamiseen osallistui Työterveyslaitokselta kirjoittajien lisäksi Anssi Smedlund, Riku Louhimo, Teemu Okkonen ja Kimmo Sirén – heille lämpimät kiitokset!

Kiitokset hankkeeseen osallistuneille organisaatioille arvokkaasta ajallisesta ja taloudellisesta panostuksesta hankkeen toteutukseen. Erityiskiitokset näiden organisaatioiden yhteyshenkilöille hyvin onnistuneesta yhteistyöstä! Tutkimushankkeessa käytettiin organisaatioiden datan lisäksi Tapaturmavakuutuskeskuksen (TVK) suomalaisilta työpaikoilta keräämiä vahinkokuvaustekstejä. Haluamme kiittää hyvää yhteistyöstä TVK:ta ja erityisesti tietokanta-analyytikko Janne Sysi-Ahoa. Kiitokset myös hankkeen ohjausryhmälle!

Helsingissä 29.9.2023

AI Safety -tutkimusryhmä

Sisällys

Tiivistelmä.....	3
Abstract.....	5
Kiitokset.....	7
Sisällys.....	8
1 Johdanto.....	10
1.1 Systeemilähtöinen turvallisuusajattelu.....	10
1.2 Koneoppiminen turvallisuusjohtamisen tukena.....	11
1.3 Tutkimuksen lähtökohta ja tavoite.....	12
2 Tiedonhallinta ja tietosuoja hankkeessa.....	14
3 Aineisto ja menetelmät.....	16
3.1 Aineistojen valinta ja datankeruun valmistelu.....	16
3.2 Aineisto.....	17
3.3 Analyysien valmistelu.....	18
3.4 Korrelaatioanalyysi.....	18
3.5 Ennustemallinnus.....	19
3.6 Klusterointi.....	19
3.7 Aihemallinnus.....	20
3.8 Inhimilliset tekijät -tekstianalyysi koneoppimismetodilla.....	20
3.8.1 Opetusdatan muodostaminen ja tarkentaminen analyysien edetessä.....	21
3.8.2 Neuroverkkomallin kouluttaminen.....	24
3.9 Työpaja-aineisto.....	24
3.10 Työtaturmatutkintojen inhimillisten tekijöiden asiantuntija-analyysi.....	25
4 Tulokset ja tulosten tarkastelu.....	30
4.1 Ennustemallit ja korrelaatiot.....	30
4.2 Klusterointi.....	33
4.3 Aihemallinnus.....	35
4.4 Inhimillisten tekijöiden koneoppimismalli.....	41

4.4.1	Havaintoja HF-4 analyysin tuloksista.....	46
4.5	Työpajojen tulokset.....	47
4.6	Tapaturmatutkintojen HF-asiantuntija-analyysien tulokset.....	53
4.6.1	Organisaatio- ja ryhmätason tekijöitä tunnistetaan harvemmin.....	53
4.6.2	Huolimattomuuden ja inhimillisen virheen syitä ei selvitetä	59
4.6.3	Positiivisten tekijöiden ja onnistumisten analysointi tapaturmatutkinnassa.....	60
4.6.4	Kokonaisarviot raporteista ja niiden kielestä	61
4.6.5	Keskeisimmät havainnot HF-asiantuntija-analyyseistä	62
5	Johtopäätökset ja suositukset.....	64
5.1	Johtopäätökset.....	64
5.1.1	Organisaatioiden keräämän datan hyödynnettävyys turvallisuuden johtamisessa.....	64
5.1.2	Työturvallisuuden haasteita ja kehittämistarpeita	66
5.1.3	Turvallisuustietojen hyödyntäminen päätöksentekotilanteissa.....	67
5.2	Toimenpidesuosituksia	68
5.3	Jatkotutkimusehdotukset.....	71
	Lähteet.....	72

1 Johdanto

1.1 Systeemilähtöinen turvallisuusajattelu

Turvallisuusajattelussa on tapahtunut merkittäviä muutoksia viimeisten vuosikymmenten aikana. Siinä missä niin kutsuttu perinteinen turvallisuusajattelu (Safety I) suhtautui ihmisiin turvallisuusongelmien ja virheiden lähteinä (Hollnagel, 2014; Hollnagel ym., 2013), nykyaikainen turvallisuusajattelu (Safety II) korostaa uutta positiivista lähestymistapaa turvallisuuden hallintaan: se näkee ihmiset turvallisuuden toteuttajina ja ongelmien ratkaisijoina. Vaikka ihmisen toiminnan merkitys turvallisuuspoikkeamissa on tiedostettu pitkään (Reason, 2008; Dekker, 2002; Reason, 1997), poikkeamatutkinnoissa korostuvat tekniset asiat ja toisaalta yksilötason virhetekijät (Teperi, 2012; Hollnagel, 2009; Kirwan, 2003). Erytisesti turvallisuuskriittisillä aloilla turvallisuusajattelun näkökulma on vähitellen laajentunut teknisistä tekijöistä monipuolisesti erilaisiin inhimillisiin ja organisatorisiin tekijöihin (Reiman ja Oedewald, 2009; Hollnagel ym., 2006; Hale ja Hovden, 1998).

Turvallisuusajattelun kehittyessä myös käsitys inhimillisistä tekijöistä (human factors, HF) on muotoutunut yksilö- ja virhelähtöisestä tarkastelusta kokonaisvaltaiseksi lähestymistavaksi, jossa inhimilliset tekijät käsitetään laajasti eri tasoiksi ihmisen toimintaan vaikuttaviksi tekijöiksi kattaen yksilötason tekijöiden lisäksi myös työ-, organisaatio- ja ryhmätasojen tekijät (Wilson, 2014; Teperi, 2012; Wickens, 2008). Kokonaisvaltaisessa, systeemilähtöisessä turvallisuusajattelussa (esim. Leveson, 2011; Dekker, 2011; Leveson, 2020) turvallisuus nähdään koko sosioteknisen järjestelmän tuotoksena ja ihminen yhtenä osana monimutkaista järjestelmää. Johtamisen näkökulmasta systeemilähtöisyys edellyttää ymmärrystä ihmisen toiminnan edellytyksistä ja rajoitteista, ja samalla ymmärrystä sosioteknisestä järjestelmästä ja sen osien kytköksistä ja dynamiikasta kokonaisuutena.

Ihmisen toiminnalla on vaikutusta useimmissa työtapaturmissa (Ford ja Tetrick, 2008), mutta systeemilähtöisen ajattelun mukaan ihmisen toimintaa tulee tarkastella osana sosioteknistä järjestelmää ja siinä ajan, paikan ja olosuhteiden kontekstissa, jossa toiminta kullakin hetkellä tapahtuu. Kokonaisvaltaista ymmärrystä ihmisen toiminnan taustalla vaikuttavista tekijöistä tarvitaan paitsi poikkeamatutkinnoissa myös ennakoivassa turvallisuustoiminnassa (riskienarvioinnit, turvallisuuskoulutus, turvallisuushavainnot). Aiempien ydinvoimatuotantoon ja lennonvarmistukseen keskittyvien tutkimusten (Teperi ym., 2017b; Teperi ym., 2015) mukaan kokonaisvaltainen ymmärrys inhimillisistä tekijöistä vaikuttaa myönteisesti

turvallisuusosaamiseen sekä kykyyn tunnistaa ja analysoida omassa työssä vaikuttavia eri tason taustatekijöitä.

1.2 Koneoppiminen turvallisuusjohtamisen tukena

Perinteisesti turvallisuusjohtamisessa on käytetty reagoivia menetelmiä (Leveson, 2015). Systeemilähtöisessä turvallisuusjohtamisessa turvallisuutta tarkastellaan suhteessa muuhun organisaation toimintaan, ja tarkoituksenmukainen päätöksenteko edellyttää kattavaa tilannekuvaa organisaation turvallisuudesta ja sen taustalla vaikuttavista tekijöistä. Tunnistamalla olennaisia turvallisuustilanteita ennakoivia tietolähteitä ja hyödyntämällä koneoppimismenetelmiä, voi tiedonkeruuta ja analysointia kohdentaa olennaisiin asioihin ja kustannustehokkaammin.

Perinteistä työturvallisuusdataa kuten työtaturma- ja vaaratilanneraportteja aineistona käyttäen koneoppimismallit pystyvät ennustamaan esimerkiksi työtaturmien määrää ja vakavuutta (Poh ym., 2018) ja tämä onkin ollut viime vuosina melko yleinen tapa kokeilla koneoppimismenetelmiä turvallisuustutkimuksessa. Tiedonloughinnan ja koneoppimisen soveltaminen turvallisuustutkimuksessa on nopeasti kehittyvä aihealue, ja uusia julkaisuja aihepiiristä on tullut kasvavaan tahtiin. Suurin osa julkaisuista on keskittynyt työtaturmiin, mutta myös työterveyshaittoja on tutkittu. Tutkimuskohteista yleisimmät ovat olleet rakennusteollisuus ja terästeollisuus (Lavezo dos Reis ym., 2021, Rathor ym., 2022, Liu ym., 2022).

Aikaisemmissa tutkimuksissa aineistona on käytetty esimerkiksi tavanomaista työtaturmadataa (esim. Ciarapica ja Giacchetta, 2009), tavanomaista työtaturmadataa rajattuna tiettyihin tapaturmiin (esim. Nenonen, 2013; Sarkar ym., 2019; Ruso ja Stojanovic, 2012) ja dataa työperäisistä sairauksista (esim. Krishna ym., 2015). Muita aineistoja koneoppimista tai tiedonloughintaa metodina käyttäneissä työturvallisuusaiheisissa tutkimuksissa on ollut sairauspoissaoloihin liittyvät etuushakemukset (esim. Bertke ym., 2012), työntekijöiden arviointiin käytetty sensoridata (esim. Xie ja Chang, 2018) ja valokuva- sekä videokuvamateriaali (esim. Shein ym., 2015; Siddula ym., 2016). Edellä mainittuja tutkimuksia yhdistää se, että niiden aineisto koostuu yhdestä tietolajista.

Työturvallisuusriskien ennakointiin on kehitetty myös malleja, joiden aineisto koostuu perinteisen työturvallisuusdatan lisäksi esimerkiksi kuvamateriaalista ja työpaikan talousasioista kertovasta datasta (Ajayi ym., 2019). Tällaisen *monimuotoisen aineiston* hyödyntäminen on tehokkainta juuri koneoppimismenetelmillä (Huang ym., 2019). Tiedonloughinnan yksi hyödyistä on se, että samalla kertaa voi analysoida useaa eri

aineistoa ja etsiä aineistojen välisiä yhteyksiä (Wang ym., 2018). AI Safety -hankkeen suunnitteluvaiheessa tutkimusaukoksi tunnistettiin erityisesti yllä mainitun monimuotoisen aineiston hyödyntäminen koneoppimismenetelmiä soveltavassa turvallisuustutkimuksessa. Niinpä AI Safetyssä kerättiin useammasta erilaisesta datasta koostuva aineisto neljästä Suomessa toimivasta teollisuusorganisaatiosta.

1.3 Tutkimuksen lähtökohta ja tavoite

Nolla tapaturmaa -ajattelun mukaan hyvä työturvallisuuskulttuuri ja -johtaminen sisältää turvallisuuden jatkuvan kehittämisen ja pyrkimyksen nolnaan työtapaturmaan ja vaaratilanteeseen (Zwetsloot ym., 2017). Faktoihin perustuvan turvallisuusjohtamisen tueksi organisaatioissa halutaan enenevässä määrin hyödyntää niille kertyvää työturvallisuusdataa. Monien erilaisten tietojärjestelmien ja raportointikäytäntöjen myötä turvallisuusjohdolla voi kuitenkin olla saatavilla niin suuri määrä dataa, että sen läpikäyminen perinteisin menetelmin olisi erittäin työlästä. Tämä koskee etenkin tekstiaineistoja, joita työpaikoille syntyy työtapaturma- ja vaaratilanneilmoituksista tai työtapaturmatutkinnoista. Työpaikat, jotka pyrkivät jatkuvasti kehittämään työturvallisuuttaan, etsivät siihen ymmärrettävästi apuja modernista analytiikasta. Tämän vuoksi tiedonlouhinta ja koneoppimismenetelmät ovat relevatti tutkimuskohde myös turvallisuusjohtamisen näkökulmasta (Goel ym., 2017). Koneoppimismenetelmien sovelluskohteita työturvallisuuden kehittämisessä ovat muun muassa ennakoiva huolto, turvallisuusanalytiikka, riskinarviointi ja henkilönsuojaimet (Xu ja Saleh, 2021).

Koneoppimismenetelmät – kuten tekstinlouhinta – mahdollistavat suurten aineistojen nopean analysoinnin. Perinteisen työturvallisuusdatan lisäksi työpaikan turvallisuustilannetta voi pyrkiä analysoimaan myös kontekstidatan eli sellaisen datan avulla, joka ei suoraan liity työturvallisuuteen, mutta tuo ymmärrystä siitä toimintaympäristöstä ja kontekstista, jossa turvallisuus – tai sen haasteet – kullakin hetkellä rakentuu. Koneoppimismenetelmät helpottavat piilossa olevan informaation löytämistä suurista datoista (Mohammadpoor, 2018), jonka vuoksi myös turvallisuusjohto voi hyötyä niistä.

AI Safety -tutkimushankkeessa tutkimuskohteenä oli monipuoliseen dataan perustuva turvallisuusanalytiikka koneoppimismenetelmillä sekä hyödyntämismahdollisuudet turvallisuusjohtamisessa. AI Safety -hankkeen tarkoituksena oli tutkia, miten koneoppimismenetelmiä voi hyödyntää turvallisuusjohtamisessa työkaluina, jotka syventävät ymmärrystä datasta ja organisaation työturvallisuudesta sekä tarjoavat ennakoivia mahdollistavaa analytiikkaa. Perinteisen työturvallisuusaineiston lisäksi hankkeessa pyrittiin hyödyntämään monipuolisesti myös organisaatioiden muita

aineistoja, jotka eivät suoraan kerro työturvallisuudesta. AI Safetyssä näitä muita dataja haettiin henkilöstöhallinnosta ja tuotannosta. Lisäksi hankkeessa yhdistettiin inhimillisten tekijöiden viitekehystä (Teperi, 2023) datatieteen menetelmiin ja tutkittiin, miten inhimillisten tekijöiden tunnistaminen ja kirjaaminen toteutuvat kerätyn organisaatiodatan valossa.

Tutkimuksen päätavoitteina oli:

- 1) Tuottaa tietoa tietoaineistojen luovutuksesta ja käytöstä tutkimuksessa
- 2) Tietoaineistojen yhdistämisen mahdollistama tapaturmien ja turvallisuustilanteiden parempi ymmärtäminen
- 3) Tunnistaa tietoaukkoja tapaturmaraporttien vapaamuotoisissa tekstikuvauksissa ja muussa tekstimuotoisessa aineistossa (esim. johdon ja turvallisuustiimin kokouspöytäkirjat)
- 4) Edistää tietoon perustuvaa työtapaturmien seurantaa ja ennakointia sekä tuottaa tietoa päätöksenteon (kuten turvallisuusinvestointien ja muiden toimenpiteiden) tueksi

Taulukossa 1 on esitetty tutkimuksen päätavoitteisiin liittyvät tutkimuskysymykset. Tutkimuskysymykset ovat ohjanneet datan keruuta, käsittelyä ja analysointia. Tämä loppuraportti pyrkii mahdollisimman hyvin kertomaan, miten tutkimuskysymyksiin vastattiin.

Taulukko 1: Tutkimustavoitteisiin liittyvät tutkimuskysymykset.

Tutkimuskysymykset
1.1) Miten organisaatiot voivat luovuttaa työntekijöidensä henkilötunnisteista dataa tutkimuskäyttöön tietoturvallisesti ja eettisesti?
1.2) Miten eettiset- ja tietosuojakysymykset vaikuttavat tietoaineistojen yhdistämiseen?
2.1) Mitkä muut saatavilla olevat tietoaineistot selittävät tapaturmia ja turvallisuustilanteita?
2.2) Millaisia ryhmiä (klustereita) erilaisista tapaturmatyypeistä voidaan muodostaa ja mitä ja miksi jää niiden ulkopuolelle?
2.3) Mitä turvallisuuden heikkoja signaaleja voidaan tunnistaa datalähtöisesti?
3.1) Toistuvatko raporteissa saman tyyppiset tapaturmat?
3.2) Saadaanko työpaikoilla riittävästi tietoa tapaturmien syistä niiden toistumisen estämiseksi?
4.1) Mitä ennakoivia toimintamalleja voidaan kehittää tiedonlouhinnalla?
4.2) Voidaanko tapaturmia luotettavasti ennakoida turvallisuusdatan ja muun datan yhdistelmällä?
4.3) Millaisia tietojohdamisen prosesseja ja järjestelmiä tapaturmien jatkuvasti päivittyvään ennakointiin kannattaa rakentaa?

2 Tiedonhallinta ja tietosuoja hankkeessa

Datan luovuttaminen työpaikoilta sen ulkopuoliseen tutkimuskäyttöön on aina huoellisuutta vaativa prosessi, johon voi kulua myös yllättävän paljon aikaa. Jos data sisältää henkilötietoja, on lainsäädännön mukaisesti käsittelylle osoitettava *käsittelyperuste*, joka tutkimuksessa on usein EU:n yleisen tietosuoja-asetuksen ja Suomen tietosuojalain mahdollistama *yleisen edun mukainen* tieteellinen tutkimus (GDPR; Tietosuojalaki).

Kun työpaikka päättää luovuttaa dataansa tutkimuskäyttöön, on otettava huomioon datassa olevat henkilötiedot, ja siihen liittyen datassa olevien henkilöiden eli rekisteröityjen oikeudet tietojensa käytöstä. Tässä projektissa henkilötietojen käsittelystä tehtiin tietosuoja-asetuksen mukainen vaikutustenarviointi sen varmistamiseksi, että henkilötietojen käsittely tutkimuksessa oli yhteensopivaa ja oikeasuhtaista tietojen alkuperäisen käyttötarkoituksen kanssa ja että henkilötietojen suojatoimet olivat riittävät.

Yleisen edun mukaiseen henkilötietojen käsittelyperusteeseen liittyvä rekisteröidyn oikeus vastustaa tietojensa käsittelyä, mutta siitä on tietyissä olosuhteissa mahdollisuus poiketa. AI Safety -hankkeessa työntekijöille pyrittiin tiedottamaan tietojen käyttämisestä ja heidän oikeuksistaan eri tavoilla ennen tietojen luovuttamista ja päätettiin, että vastustamisoikeudesta ei poiketa.

Hankkeessa kerättiin tutkimusdataa neljältä eri organisaatiolta useista eri lähteistä. Data oli hyvin monimuotoista ja osittain vaikeasti hyödynnettävää. Helpoiten pystyttiin hyödyntämään rakenteellista eli taulukkomuotoista dataa, esimerkiksi turvallisuushavaintojen ja -tapahtumien lukumäärä- ja luokittelutietoja tai työvuorotietoja. Vaikeammin hyödynnettävää oli rakenteeton data eli esimerkiksi käsin täytetyt, keskenään hieman erimuotoiset Excel-taulukot tai täysin vapaamuotoinen teksti, kuten turvallisuushavaintojen ja -tapahtumien vapaamuotoiset kuvaukset.

Rakenteettoman datan hyödyntämistä rajoitti sekä sen vaikea systemaattinen hyödyntäminen ennustemalleissa että riittävän tietosuojan takaaminen. Siinä, missä taulukkomuotoinen data on helppo pseudonymisoida ja poistaa sieltä ylimääräinen tai liian tarkka tieto henkilöistä, rakenteettoman datan pseudonymisointi on työläämpää. Onneksi tässä hankkeessa pystyttiin hyödyntämään tähän tehtävään myös tekoälyä. Onnistuneen pseudonymisoinnin jälkeen esimerkiksi turvallisuushavaintojen ja -tapahtumien kuvausten hyödyntäminen oli suoraviivaista.

Pseudonymisoinnissa suorat henkilötunnisteet muunnetaan pseudonyymeiksi, joita ei voi suoraan liittää henkilöön ilman lisätietoa. Tässä hankkeessa tutkimusaineisto pseudonymisoitiin osittain paikan päällä organisaatioissa ennen varsinaista luovutusta tutkimuskäyttöön tai sitten eristetyllä pseudonymisointipalvelimella Työterveyslaitoksen IT-ympäristössä. Joka tapauksessa pseudonymisointi tehtiin ennen varsinaisen tutkimuksen aloittamista kullekin aineistolle. Tekoälyä pystyttiin hyödyntämään esimerkiksi vapaiden tekstien pseudonymisoinnissa, jossa tekoäly tunnisti lauseista henkilöön viittaavia sanoja myös lauserakenteen ja tekstin kontekstin perusteella. PDF muodossa olevien tapaturmatutkintojen pseudonymisointia tehtiin myös käsin poistamalla henkilöihin viittaavat tiedot.

Kaiken kaikkiaan tässä hankkeessa kerätyn kokonaisvaltaisen työturvallisuuteen tavalla tai toisella linkittyvän tiedon hyödyntäminen organisaatioissa mahdollistuisi tietovarastoinnin avulla. Myös tiedonlouhintaa hyödyntävät jatkotutkimukset helpottuisivat merkittävästi. Tietovarastoinnissa kuvatus kaltainen, monista eri lähteistä eri muodoissa kerätty tieto mallinnetaan helposti hyödynnettävään muotoon yhtenäiseen tietokantaan. Tiedon mallinnuksessa teknisistä järjestelmistä haettu tieto nimetään yhdenmukaisesti työturvallisuuden käsitteistöllä. Tämä helpottaa merkittävästi esimerkiksi data-analyttikoiden ja työturvallisuusasiantuntijoiden välistä vuoropuhelua. Hankkeen organisaatioissa ei laajamittaista yksittäistä tietovarastoa työturvallisuuden osalta ollut saatavilla vaan tiedot haettiin monista eri lähteistä, kuten yllä kuvattiin.

3 Aineisto ja menetelmät

3.1 Aineistojen valinta ja datankeruun valmistelu

Hankkeen alussa toteutettiin alkukartoituskysely, haastattelut ja alkutyöpaja. Alkukartoituskyselyllä kartoitettiin yleiskuvaa hankkeeseen osallistuneiden organisaatioiden työturvallisuustoiminnasta ja -aineistoista. Alkukartoituskysely kohdennettiin organisaatioiden yhteyshenkilöille. Näitä kyselytuloksia hyödynnettiin organisaatiokohtaisten alkutyöpajojen toteutuksen suunnittelussa.

Alkutyöpajat ja -haastattelut toteutettiin kaikissa neljässä organisaatiossa. organisaatioista osallistui 11 johtajaa viiteen haastattelutilanteeseen ja 29 henkilöä neljään alkutyöpajaan. Litteroitua aineistoa näistä haastatteluista ja työpajoista kertyi yhteensä 367 sivua.

Alkutyöpajat toteutettiin etäyhteyksin sisältäen alustuksia, ryhmätöitä sekä yhteisen tilannekuvan työstöä. Alkutyöpajoissa aiheina olivat organisaatioiden nykyinen työturvallisuustilanne, turvallisuustietojen kerääminen ja niiden hyödyllisyys. Lisäksi keskusteltiin siitä, miten turvallisuusdataa hyödynnetään työturvallisuuden kehittämisessä, miten työturvallisuustilannetta organisaatioissa ennakoidaan, millainen olisi ideaalitalanne ja mitkä ovat kehittämistarpeet. Alkutyöpajoissa keskusteltiin erityisesti organisaatioiden turvallisuusjohtamisessa käytetyistä datoista sekä niiden koetusta hyödyllisyydestä. Datat jaoteltiin työturvallisuusaiheiseen dataan, työympäristö-, henkilöstö- ja tuotantodataan. Työpajoissa käsiteltiin noin 30 dataa/organisaatio.

Alkukartoitus, organisaatioiden kanssa käydyt lukuisat palaverit ja työpajat sekä johtajien haastattelut olivat tärkeässä roolissa tutkimusaineistona käytettävän datan valitsemisessa. Valintaprosessissa selvitettiin, minkälaista dataa organisaatioissa on olemassa, sen määrää ja ajanjaksoja sekä mitä dataa on mahdollista saada tutkimushankkeen käyttöön. Samalla selvitettiin myös datainfrastruktuuriin ja datan tietoturvalliseen siirtoon liittyviä seikkoja. Näin saatiin koottua organisaatiokohtaiset aineistot erilaisia datakokonaisuuksia työturvallisuudesta, tuotannosta, henkilöstöhallinnosta ja työympäristöstä. Luvuissa 3.4-3.10 on kuvattu datan käyttöä menetelmä-/analyysikohtaisesti.

3.2 Aineisto

Tutkimusaineistona oli monipuolisesti erilaista rakenteista dataa organisaatioista, sekä vapaata tekstiä työtapaturmaraporttien, vaaratilanneraporttien, työympäristöhavaintojen ja tapaturmatutkintojen muodossa. Tutkimusaineistoon sisältyi myös Tapaturmavakuutuskeskuksen (TVK) keräämiä tekstimuotoisia vahinkokuvauksia. Kaikilta organisaatioilta saadun aineiston aikaväli kattoi vuodet 2015-2021. TVK:lta saadun aineiston aikaväli kattoi vuodet 2003-2019. Datakokonaisuus oli jokaisen organisaation kohdalla omanlaisensa, mutta siinä mielessä samankaltainen, että data tuli alueilta henkilöstöhallinto, tuotanto, työympäristö ja työturvallisuus. Dataa pyrittiin saamaan hankkeen käyttöön mahdollisimman monipuolisesti. Analyyseissä hyödynnetyn datan määrä vaihteli 0,1–2,6 miljoonan yksittäisen arvon välillä organisaatiokohtaisesti.

Aikasarjamuodossa kerätty rakenteellinen data koostui työturvallisuustapahtumista (raportit työtapaturmista, vaaratilanteista ja työympäristön havainnoista) sekä organisaatiosta riippuen työaikatiedoista, sairauspoissaoloista, työntekijöille järjestettyjen koulutusten tiedoista, työntekijöiden ikäjakaumasta, työn määrää kuvaavista muuttujista, kuten tilaus- ja tuotantomäärät sekä kunnossapitotyön määrä. Näitä tietoja analysoitiin yhdessä aluksi korrelaatioanalyysillä ja hyödynnettiin sitten tapaturmariskin ennustamisessa (ennustemallit) sekä työturvallisuustapahtumien ryhmittelyssä klusteroimalla.

Aihemallinnuksessa käytettiin tekstiaineistoa sekä hankkeen osallistujaorganisaatioiden että TVK:n datasta. Tekstiaineisto koostui työtapaturmien ja vaaratilanteiden kuvauksista. TVK:n datassa työtapaturmien kuvaukset ovat siinä muodossa, jossa ne on lähetetty vakuutusyhtiölle työtapaturmien korvaushakemuksia varten. Valtaosa aihemallinnuksen työtapaturmadatasta koostui TVK:n datasta, jota saatiin hankkeen käyttöön yhteensä n. 1,67 miljoonaa riviä. TVK:n aineisto rajattiin aihemallinnusta varten teollisuusyrityksissä raportoituihin työtapaturmiin ja huomioimalla vain tekstit, joiden pituus oli yli 200 sanaa, jolloin jäljelle jäi 18 677 tekstiä. Tähän aineistoon yhdistettiin täydentävänä elementtinä osallistujaorganisaatioiden työtapaturma- ja vaaratilanneraportit, minkä jälkeen aineistossa oli 20 410 tekstiä.

Inhimillisten tekijöiden koneoppimismalli-analyysissä käytettiin neljän yhteistyöorganisaation läheltä piti- ja työtapaturmadataa. Koneoppimismalli koulutettiin datalla, joka sisälsi edellä mainittujen tapahtumatyyppien lisäksi turvallisuushavaintoja ja turvallisuuskeskusteluja. Työturvallisuusasiantuntijat koostivat organisaatiodatasta esimerkkejä kullekin HF-kehyksen inhimilliselle tekijälle, ja suuri, esikoulutettu kielimalli hienosäädettiin (jatkokoulutettiin) näillä asiantuntijavoimin

laadituilla koulutusesimerkeillä luokittelemaan vapaata suomenkielistä tekstiä inhimillisten tekijöiden neljän ylätasoa mukaan. Läheltä piti ja tapaturmatekstien analyysijä verrattiin organisaatiokohtaisesti keskenään. Tapaturmavakuutuskeskuksen teollisuusyrityksissä sattuneista tapaturmista koostettua aineistoa verrattiin kaikkien organisaatioiden yhdistelmätulokseen.

Asiantuntija-analyysi inhimillisten tekijöiden viitekehysessä tehtiin kahden organisaation työtaturmien tutkintaraportteille. Kaksi neljästä organisaatiosta toimitti analysoitavaksi yhteensä 67 tapaturman tutkintaraporttia vuosilta 2015–2021. Tapaturmien tutkintatapa sekä tutkintaraporttien muoto ja laajuus vaihtelivat ja aineisto koostui organisaation turvallisuusjohtamisjärjestelmän PDF tulosteista, juurisyyanalyysistä 5 X Miksi tai kalanruotokaavio -menetelmillä sekä organisaation omista tapaturman ilmoitus- ja tutkintalomakkeista. 52 tapauksista oli tutkittu 5 X Miksi menetelmällä, 14 kalanruotokaaviolla ja yhdestä tutkintaraportista ei ilmennyt mitään juurisyyanalyysimenetelmää oli käytetty/oliko juurisyyanalyysimenetelmää käytetty. Raportit sisälsivät vapaata tekstiä sekä valittavissa olevia luokitteluja.

3.3 Analyysien valmistelu

Ennen varsinaista analyysiä aikasarjamuodossa kerättyä rakenteellista dataa esikäsiteltiin. Tärkeimpänä menetelmänä käytettiin keskiarvoistusta ajan suhteen, jotta esimerkiksi päivä- ja viikkotasolla kerätyt muuttujat saatiin yhteismitallisiksi. Pääosin analyysissä hyödynnettiin kuukausikeskiarvoja, mutta osin myös viikkokeskiarvoja. Osaa muuttujista myös suodatettiin ajan suhteen, jolloin pidemmän aikavälin ilmiöt pystyttiin erottamaan nopeammasta vaihtelusta. Suurin osa muuttujista myös normalisoitiin käytetyn työajan suhteen, eli jaettiin muuttujan kuukausikeskiarvot samassa kuussa tehdyillä kokonaistyötunneilla. Normalisoinnilla pyrittiin poistamaan mm. loma-aikojen vaikutusta muuttujien välisiin riippuvuuksiin (esimerkiksi kesäkuukausien aikana työtunnit ja tuotanto saattavat molemmat laskea merkittävästi, mikä näkyy normalisoimattomien muuttujien välisenä voimakkaana mutta itsestään selvänä korrelaationa).

3.4 Korrelaatioanalyysi

Aikasarjamuotoisen rakenteellisen datan riippuvuuksia hahmotettiin ensin korrelaatioanalyysien avulla. Korrelaatioanalyysillä voidaan selvittää kahden muuttujan välistä riippuvuutta, jonka suuruutta kuvataan korrelaatiokertoimella. Korrelaatiokerroin saa arvoja välillä -1 ja 1. Mitä enemmän sen arvo poikkeaa nolasta, sitä suurempi

muuttujien riippuvuus on. Korrelaatioanalyysi rajattiin tässä tapauksessa lineaaristen riippuvuuksien tutkimiseen. Koska kyseessä olivat laajat, monien muuttujien yhteisvaikutuksista koostuvat kokonaisuudet, voitiin korrelaatioanalyysien tuloksia tarkastella vain suuntaa antavina.

3.5 Ennustemallinnus

Ennustemallinnus tehtiin XGBoost-mallinnuksella (eXtreme Gradient Boosting) (Chen ja He 2014, Chen ja Guestrin 2016), jossa turvallisuustapahtumien esiintymistä viikko- tai kuukausitasolla ennustettiin muiden organisaatiosta kerättyjen muuttujien avulla. XGBoost-algoritmi perustuu yksittäisten yksinkertaisten päätöspuiden yhdistämiseen siten, että seuraava päätöspuu ennustaa aina edellisten puiden jättämää virhettä (Boosting). Puiden lisäämisessä kokonaismallinnusvirheen kustannusfunktiota minimoidaan gradienttimenetelmällä. XGBoost-mallien rakennetta ja muuttujien merkittävyyttä niissä tulkittiin Shapley-arvoilla (Shapley, 1953), jotka perustuvat peliteoriaan.

Ennustemallinnusta varten laskettiin organisaatiokohtaisesti läheltä piti -tapausten ja mahdollisten tapaturmien määrää ja vakavuutta kuvaava riski-indeksi (jatkossa "riski"), jota pyrittiin ennustamaan muilla organisaation rakenteellisessa muodossa olevilla muuttujilla eli selittäjillä. Näitä olivat mm. tuotantoon, työaikoihin, sairauspoissaoloihin, koulutukseen ja työntekijöiden ikäjakaumaan liittyvät tiedot. XGBoost-mallinnusmenetelmän hyvänä puolena on se, että mukaan voidaan ottaa suuri määrä muuttujia, ja menetelmä itse nostaa esiin ne, jotka mallinnuksen kannalta ovat merkittäviä.

3.6 Klusterointi

Klusteroinnissa turvallisuustapahtumia luokiteltiin datapohjaisesti. Analyysin tavoitteena oli selvittää, minkälaisia turvallisuustapahtumaluokkia organisaatioilta kerätyissä aineistossa esiintyy ja minkä tyyppiset tapahtumat mahdollisesti jäävät luokkien ulkopuolelle. Luokittelu tehtiin turvallisuustapahtumista rakenteellisessa muodossa kerättyjen tietojen sekä ennustemallien esiin nostamien tärkeimpien rakenteellisten muuttujien avulla. Luokittelumenetelmänä käytettiin DBSCAN-algoritmia (Density-based spatial clustering of applications with noise) (Ester ym., 1996), joka perustuu näytepisteiden tiheyden tarkasteluun; lähellä toisiaan sijaitsevat näytteet luokitellaan samaan luokkaan, kun taas erillään sijaitsevat näytteet tulkitaan kohinaksi.

3.7 Aihemallinnus

Aihemallinnuksella voitiin tutkia hankkeessa kerättyjä tekstiaineistoja ohjaamattomasti. Tähän käytettiin Latent Dirichlet Allocation (LDA) -menetelmää (Blei ym., 2003; Blei 2012). LDA on generatiivinen bayesiläinen menetelmä, jonka keskeinen käsite, *aihe*, määrittellään todennäköisyysjakaumana sanaston yli. Esimerkiksi yhdessä aiheessa voi esiintyä todennäköisimpinä sanoina "käsi", "kone" ja "haava", jolloin aiheen voidaan tulkita viittaavan käsikonetyötapaturmiin. LDA:lle annetaan syötteenä kokoelma tekstidokumentteja, ja sen perusoletus on, että kokoelma on generoitu eli tuotettu niin, että jokaisen dokumentin taustalla on todennäköisyysjakauma aiheiden joukon yli, ja että dokumentissa esiintyvät sanat on tuon aihejakauman perusteella generoituja (muistaen, että aiheet ovat todennäköisyysjakaumia sanaston yli). Bayesiläisen päättelyn avulla voidaan kouluttaa eli laskea malli (siis määrittää sellainen aiheiden joukko), joka noudattaa yllä olevaa perusoletusta. Hankkeessa koulutettiin LDA-malli käyttäen organisaatioilta kerättyjä tapahtumakuvaustekstejä tapaturma- ja vaaratilanteista sekä Tapaturmavakuutuskeskukselta saatuja teollisuuden alan vahinkokuvaustekstejä. Mallin avulla voitiin tutkia, mitkä aiheet kunkin organisaation vahinkokuvausteksteissä esiintyivät usein sekä vertailla aiheiden esiintymistä eri organisaatioiden, eri tapahtumatyyppien sekä eri ajanjaksojen välillä.

3.8 Inhimilliset tekijät -tekstianalyysi koneoppimismetodilla

HF-viitekehukseen perustuvassa tekstianalyysissä oli tavoitteena opettaa neuroverkkoa tunnistamaan tekstimassasta eri inhimillisiin tekijöihin liittyviä tekijöitä. Nämä tekijät malli luokitteli neljään ylätasoon ryhmään, jotka ovat yksilön toiminta ja piirteet, työtoiminta ja työn piirteet, ryhmätason tekijät ja organisaatiotason tekijät. Tavoitteena oli tätä kautta yrittää tunnistaa tietoaaukkoja aineistossa ja tarkastella koskeeko raportointi turvallisuutta edistävää vai heikentävää asiaa vai molempia tai turvallisuuteen liittyvää toimenpidettä.

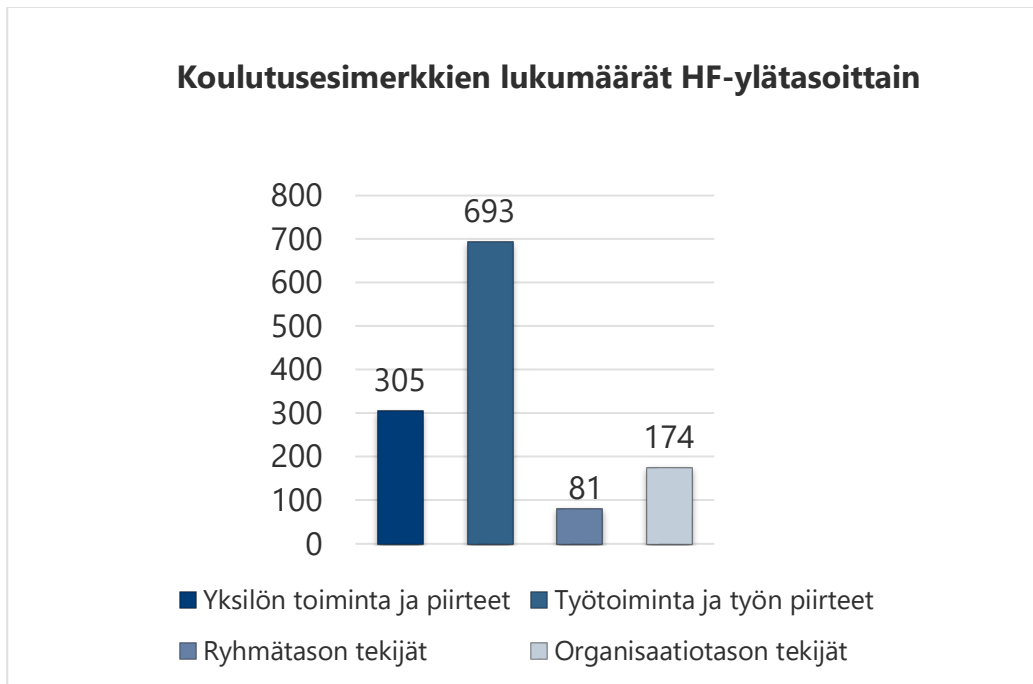
Inhimillisten tekijöiden tekstianalyysissä menetelmänä toimii suuri, esiopetettu kielimalli (Large Language Model, LLM) jonka päätöksenteko luokitteluille perustuu sekä ohjaamattomasta esiopetuksesta (pre-training), että ohjatusta hienosäädöstä (fine-tuning) (Devlin ym., 2018). Suuri kielimalli on esiopetuksen aikana oppinut sanojen konteksteja lauseparien yhteenkuuluvuutta ja peitettyjä sanoja ennustamalla (mukaelma cloze-testistä). Esioppimista on vahvistettu asiantuntijatyönä koostetulla hienosäätödatalla, joka on koostettu organisaatioiden turvallisuusdatasta tutkimusryhmän työturvallisuusasiantuntijoiden toimesta. Suoritettu neuroverkkomallin hienosäätö vie vain joitakin kymmeniä minuutteja, kun taas esiopetus on kestänyt

useita kymmeniä tunteja. Käytetyn esiopetetun mallin on esikouluttanut Turun yliopiston luonnollisten kielten tutkimusryhmä (Virtanen ym., 2019).

Hienosäätö vahvistaa esikoulutetun mallin neuroverkon yhteyksiä niin, että malli luokittelee sille syötetyn vapaan tekstin koulutusdatan perusteella siihen parhaiten sopivaan inhimillisten tekijöiden ylätasoon. Koulutusdatassa esiintyvien tekstien sanat, ja sanojen kontekstit vaikuttavat siihen, miten neuroverkon painoja muutetaan esikoulutetun mallin hienosäädössä. Näin toimivien neuroverkkomallien arkkitehtuuri nojaa nk. huomiomekanismiin (engl. attention) (Vaswani ym., 2017). Tällä tavalla, ihannetilanteessa, samat sanat voivat esiintyä eri HF-4 ylätasossa, mutta sanojen konteksti ja lauseiden merkitys vaikuttaa eniten luokitteluun. Käytetyn datan sotkuisuus, katkonaisuus (noise) ja epäjohdonmukaiset, puutteelliset tekstit vaikeuttavat hienosäädön onnistumista sekä heikentävät analyysin tarkkuutta.

3.8.1 Opetusdatan muodostaminen ja tarkentaminen analyysien edetessä

Neuroverkon opettamiseen laadittiin opetusdata, joka sisältäisi vähintään 10 aineistokatkelmaa jokaiseen HF Tool™ -työkalun (Teperi, 2023: s. 57) kohtaan (ks. Kuva 1). Kaikkiaan yli tuhannesta koulutusesimerkistä yli puolet kuului Työtoiminta ja työn piirteet -ylätasoon. Tämän epätasapainon vaikutusta vähennettiin painottamalla vähiten esiintyviä inhimillisiä tekijöitä. Tämä tarkoittaa sitä, että neuroverkkoa hienosäätäessä, vähemmän esiintyvän inhimillisen tekijän piirteet korostuvat enemmän, kuin lukumääräisempien tekijöiden vastaavat piirteet.



Kuva 1: Neuroverkkomallin hienosäädössä käytettyjen tekstikuvauspatkien HF-4 ylätasojakauma.

Inhimillinen tekijä saattoi esiintyä kuvauksessa

- turvallisuutta edistävässä (HF+, positiivinen) valossa, esimerkiksi HF1+, *työntekijällä oli työhön tarvittava ammattitaito ja osaaminen*
- turvallisuutta heikentävässä (HF1-, negatiivinen) valossa, esimerkiksi HF3-, *työntekijä ei noudattanut turvallisuusohjeita*
- molemmissa sekä positiivisessa että negatiivisessa valossa
- toimenpiteen muodossa, esimerkiksi HF29 *”johdot siivottava kulkutieltä”*

Opetusdatan laatimisen tavoitteena oli saada jokaiseen HF Toolin™ kohtaan (1-12, 20-29, 30-36, 40-47) vähintään 10 suomenkielistä aineistokatkelmaa opetusdataksi.

Opetusdatan laadinta toteutettiin seuraavasti:

1. Käytiin läpi neljän organisaation turvallisuushavaintojärjestelmän dataa, jotka oli pilkottu yhden virkkeen mittaisiksi. Organisaatiokohtaista tiedostoa käytiin läpi järjestyksessä luokitellen jokainen virke mahdollisuuksien mukaan yhteen tai useampaan inhimilliseen tekijään. Organisaatiokohtaisesti käytiin läpi 80-

180 havaintoa, kunnes huomattiin että yritysten turvallisuushavaintodata tarjoaa aineistokatkelmia vain osaan inhimillisiin tekijöihin (erityisesti HF23 Laitejärjestelmien/tekniikan toimivuus ja HF29 Fyysinen työympäristö, työolosuhteet).

2. Kahden yhteistyöorganisaation tapaturmat oli analysoitu aiemmin samalla viitekehyksellä, joten opetusdataan sisällytettiin tapaturmaraporteista soveltuvia tekstipätkiä niihin kohtiin, joissa aineistokatkelmia ei muuten ollut vielä riittävästi.
3. Yhteistyöorganisaatioiden turvallisuushavainto- ja tapaturmatutkintadatalta ei saatu suurimpaan osaan kohdista riittävää määrää aineistokatkelmia (vähintään 10 tekstipätkää/kohta). Puuttuviin kohtiin etsittiin otoksesta TVK:n vahinkokuvausdataa tietyillä hakusanoilla ja niiden osilla lisää osumia. Esimerkiksi kohtaan HF7. Vireystila, väsymysoireet löydettiin seuraavasti osumia hakusanoilla: "väsy" 8 osumaa, "huonosti nuk" 1 osuma, "nukku" 1 osuma, kun taas hakusanat "vireys", "valvo" ja "nukah" eivät tuottaneet osumia.
4. TVK:n vahinkokuvausdatasta hakemallakaan ei löytynyt kaikkiin kohtiin riittävästi opetusdataa, joten luettiin julkisia kuolemaan johtaneiden tapaturmien TOT-tutkintaraportteja ja löydettiin niistä mm. organisaatiotason tekijöihin aineistokatkelmia.
5. Niihin kohtiin, joissa ei ollut vielä riittävästi opetusdataa, muotoiltiin kahden inhimillisten tekijöiden asiantuntijan toimesta koulutusmateriaaliin pohjautuen lyhyitä tekstikatkelmia.
6. Mallin suorituskyvyn arvioimisen tueksi ja menetelmän osuvuuden arvioimiseksi, TVK:n vahinkokuvausdatan Teollisuus-toimialalta koostettiin kaikista vähintään 250 merkkiä pitkistä kuvauksista otos, josta inhimillisten tekijöiden asiantuntija jatkokehitti sopivan kuvaavia ja eheitä tekstikokonaisuuksia.
7. Tarvittaessa tekstejä muokattiin kevyesti vastaamaan paremmin organisaatioiden käyttämää kieltä ja yhtenäistettiin termejä kuten NN-> työntekijä tai henkilö.

Taulukko 2: Opetuslauseiden lukumäärä HF-työkalun alakohtien mukaan.

Yksilön toiminta	HF 1	HF 2	HF 3	HF 4	HF 5	HF 6	HF 7	HF 8	HF 9	HF 10	HF 11	HF 12
lkm	17	57	88	16	31	12	12	13	13	17	11	28
Työtoiminta ja työn piirteet	HF 20	HF 21	HF 22	HF 23	HF 24	HF 25	HF 26	HF 27	HF 28	HF 29		
lkm	10	26	14	261	67	16	10	15	12	262		
Ryhmä	HF 30	HF 31	HF 32	HF 33	HF 34	HF 35	HF 36					
lkm	14	9	10	10	14	15	9					
Organisaatio	HF 40	HF 41	HF 42	HF 43	HF 44	HF 45	HF 46	HF 47				
lkm	31	34	11	16	14	15	43	10				

3.8.2 Neuroverkkomallin kouluttaminen

Turun Yliopiston luonnollisten kielten tutkimusryhmä on esikouluttanut tässä hankkeessa käytetyn neuroverkkomallin [FinBERT:n](#). Esikoulutusdata, datan siistimisen ja esikäsitteilyn jälkeen sisälsi 234 miljoonaa lausetta. Nämä muodostuivat 36 miljoonasta Ylen ja STT:n uutistekstistä ja -uutissähkeestä, 118 miljoonasta suomi24 keskustelupalstan arkistoista poimitusta lauseesta ja loput, 79 miljoonaa lausetta, .fi maatumuksen perusteella valittujen internet-sivujen läpikäynnistä kerätyistä tekstistä. Kaikkiaan lauseet muodostivat 21 miljoonaa eri tekstiä käsittävän kokonaisuuden. Nämä lauseet pilkottiin neuroverkon esiopetusvaiheessa yli kolmeen miljardiin (3 300 000) osamerkkijonoon (token).

Esikäsitteilyssä on poistettu toisteisia ja ei-suomenkielisiä arvioituja tekstejä.

3.9 Työpaja-aineisto

Toiset organisaatiokohtaiset työpajat toteutettiin alustavien organisaatiokohtaisten tulosten valmistuttua. Tulostyöpajoihin osallistui organisaatioiden työturvallisuuden asiantuntijoita, työsuojeluvaltuutettuja, yritysjohtoa, asiantuntijoita ja esihenkilöitä, yhteensä 27 osallistujaa. Näissä työpajoissa esiteltiin osallistujille organisaatioiden data-analyyysien sekä tapaturmatutkintojen HF-analyyysien alustavia tuloksia ja havaintoja, keskusteltiin työturvallisuustietojen tietoaukoista sekä tunnistettiin uusia tai kehitettäviä työturvallisuuden toimintatapoja asiantuntijaehdotusten perusteella.

organisaatioiden yhteisiä tulostyöpajoja järjestettiin kaksi syyskuussa 2023. Niihin osallistui yhteensä 10 organisaatioiden edustajaa, kaikki työturvallisuuden ja työsuojelun asiantuntijoita. Organisaatioiden tulostyöpajoissa käsiteltiin toteutettujen organisaatiokohtaisten analyysien tuloksia ja analyysiprosessia sekä niihin vaikuttaneita tekijöitä. Samalla tunnistettiin mahdollisia teknisiä tai toiminnallisia tietoaukkoja sekä toimintatapojen kehittämiskohteita.

Organisaatioiden yhteisissä päätöstyöpajoissa aiheina olivat työturvallisuus- ja organisaatiodatan hyödyntäminen turvallisuusjohtamisessa, työturvallisuusdatan laadun kehittäminen, työturvallisuusdatan tulkinta suhteessa organisaatiodataan sekä työturvallisuusdataan perustuva oppiminen.

Ryhmätöissä pohdittiin mikä pitää muuttua tai mitä kehittää, että heikommin hyödynnettävät datat olisivat paremmin hyödynnettävissä työturvallisuuden tilannekuvan luomisessa. Esiin nousi monipuolisesti aiheita eri datalähteisiin liittyen.

3.10 Työtaturmatutkintojen inhimillisten tekijöiden asiantuntija-analyysi

Tutkimuksen organisaatioiden työtaturmaraporttien inhimillisten tekijöiden asiantuntija-analyysin (HF-asiantuntija-analyysi) tavoitteena oli tunnistaa tietoaukkoja raporttien vapaamuotoisissa tekstikuvauksissa ja arvioida, saadaanko työpaikoilla riittävästi tietoa tapaturmien syistä niiden toistumisen estämiseksi. Lisäksi tarkasteltiin turvallisuustilanteita kuvaavissa tekstiaineistoissa käytettyä kieltä ja rajauksia inhimillisten tekijöiden viitekehyksessä: esimerkiksi sitä kuinka yksilökeskeistä tai vaihtoehtoisesti laajasti eri tasojen tekijöitä huomioivaa kuvauksissa käytetty kieli ja kerronta on, ja miten syyllistävästi tai syyllistämättömästi tapauksia kuvataan raporteissa.

HF-asiantuntija-analyysin välineenä käytettiin Teperin (2012; 2023) kehittämä inhimillisten tekijöiden työkalua (HF Tool™), joka auttaa tunnistamaan turvallisuustilanteisiin vaikuttavia tekijöitä eri tasoilla; yksilö-, ryhmä- ja organisaatiotasolla sekä työtoiminnassa ja työn piirteissä. Työkalu on alkujaan kehitetty lennonjohdon työhön, ja sitä on myöhemmin sovellettu muilla turvallisuuskriittisillä aloilla, kuten ydinvoimassa, ilmailunhuollossa ja merenkulussa (Puro ym., 2021; Teperi ym., 2018a; Teperi ym., 2017a; Teperi ym., 2017b; Teperi ym., 2015; Teperi, 2012). Rakennusalalle on muokattu HF Tool™ -työkalusta rakennusalan työturvallisuusoppimiseen soveltuva versio (Tiikkaja ym., 2020). Lisäksi HF Tool™ -

työkalua on viime vuosina hyödynnetty mm. esittävien taiteiden turvallisuusjohtamisen ja -kulttuurin kehittämiseen ja tutkimiseen (Visuri ym., 2020).

Inhimillisten tekijöiden asiantuntijat tutkimusryhmästä muodostivat Exceliin analyysikehikon, joka rakentui HF Tool™ -työkalun erittelemistä inhimillisistä tekijöistä yksilön- ja työn piirteissä sekä ryhmä- ja organisaatiotasolla (yhteensä 37 kohtaa). Alkuperäistä, lennonjohdon työhön kehitettyä HF Tool™ -työkalua (Teperi, 2012, ks. kuva 1) muokattiin yleisemmälle, eri toimialoille soveltuvalla tasolla poistamalla siitä ilmaisupesifejä termejä ja toimijoita. Esimerkiksi ryhmätason tekijä numero 34 yksinkertaistui alkuperäisestä muotoilusta *"34. Communication; ATCO pilot, between ATCos (GND-TWR-APP-ACC etc), ATC maintenance; RTF"* muotoon *"34. Kommunikaatio eri yhteistyötahojen kesken"*.

Kaikki tapaturmiin liittyvä aineisto käytiin analyysikehikon kanssa läpi: raporttien sisältämät tapahtumien kulkuun vaikuttaneet syytekijät luokiteltiin soveltuvien inhimillisten tekijöiden alle (HF-). Lisäksi tarkasteltiin, missä määrin raporteissa esitettiin turvallisuutta edistäviä tekijöitä ja onnistumisia osana tapahtumakuvausta tai tapahtuman analyysia luokitellen ne myös analyysikehikkoon (HF+). Tutkintaraporteissa esitetyt toimenpiteet tapahtuman toistumisen ehkäisemiseksi luokiteltiin myös samalla tavalla.



Kuva 2: HF Tool™ (Teperi, 2023: 57).

Taulukko 3: HF Tool™ -työkalun rakenne ja sisältö (Teperi, 2023).

Osa-alueet	Osa-alueiden sisältö
1. Yksilön toiminta ja piirteet	<ol style="list-style-type: none"> 1. Ammattitaito, työn hallinnan taso 2. Tilannetietoisuus (tarkkaavaisuus, muisti, päätöksenteko, reagointi) 3. Normien ja sovitujen toimintatapojen noudattaminen 4. Kokonaisuuden ymmärtäminen 5. Tilanteiden ennakointi; oletukset ja varmistaminen 6. Yli- tai alikuormitus ja niiden hallintakeinot 7. Vireystila, väsymysoireet 8. Elämäntilanne, huolet, yleinen stressitaso 9. Ikä, työkokemuksen määrä ja laatu 10. Yleinen terveystilanne 11. Motivaatio, asenteet 12. Tunnereaktiot, mieliala
2. Työtoiminta ja työn piirteet	<ol style="list-style-type: none"> 20. Työn laatu ja sisältö; työtilanteen vaativuus 21. Työn määrä, aikapaine, kiire 22. Työnjako, tehtäväkuvaukset, töiden organisointi, selkeys 23. Laittejärjestelmien/tekniikan toimivuus ja käytettävyys 24. Työmenetelmät ja ohjeet; miten kirjattu, toimivuus 25. Vaikutusmahdollisuudet omaan työhön ja työoloihin 26. Työstä saatu palaute, ammatillinen arvostus 27. Mahdollisuus/ kyky arvioida ja kehittää omia työprosesseja 28. Koulutus: sisältö, vaikuttavuus, mahdollisuudet järjestää 29. Fyysinen työympäristö, työolosuhteet, työhygieeniset tekijät (ilmastointi, valaistus, lämpötila; layout)
3. Ryhmätason tekijät	<ol style="list-style-type: none"> 30. Yhtenäinen kuva tilanteesta kaikilla jäsenillä 31. Ryhmän kaikkien jäsenten tietämyksen hyödyntäminen 32. Väärinkäsitykset, -tulkinnat, -kuulemiset sekä näiden korjaaminen (otetaanko puheeksi) 33. Ryhmän rakenne ja kiinteys, muu ryhmädynamiikka (sosiaaliset suhteet, ilmapiiri, keskinäinen tuki) 34. Kommunikaatio eri yhteistyötahojen kesken 35. Tiedonkulku (käytännöt mm. vuoronvaihto) 36. Päätöksenteko ryhmässä
4. Organisaatiotason tekijät	<ol style="list-style-type: none"> 40. Johtamistapa ja -järjestelmä 41. Organisaatio-/toimintakulttuuri 42. Eri organisaatiotasojen ja -tahojen välinen yhteistyö (alue, yksiköt, konsernihallinto) 43. Kokonaishallinta, keskinäinen ymmärrys toistensa töistä 44. Tehdyt päätökset (mm. resurssit, henkilöstö, kalusto) 45. Muutoksen hallinta (henkilövaihdokset, järjestelmät) 46. Yhteistyökumppanit; toimintatapaerot, keskinäinen yhteistyö 47. Konsernipalveluiden tuki yksikölle (HR, talous)

HF Tool™ -viitekehyksen mukaisen luokittelun jälkeen analysointiin osallistuneet HF-asiiantuntijat tekivät kokonaisarviot siitä,

- kuinka yksilölähtöinen vs. laajasti eri tasoilla tekijöitä huomioiva raportti on (arviointiasteikko 1=hyvin yksilölähtöinen...5=hyvin systeemitasoinen).
- kuinka syyllistävää vs. syyllistämätöntä tapahtuman kuvaus on (arviointiasteikko 1=syyllistävää... 3=toteavaa, neutraalia... 5=syyllistämätön, reilua kuvausta).
- tunnistetaanko raportissa kattavasti erilaisia vaikuttavia tekijöitä ja tapahtuman juurisyitä vai jääkö paljon avoimia kysymyksiä (arviointiasteikko 1= paljon avoimia kysymyksiä ... 5=tunnistetaan kattavasti vaikuttavia tekijöitä ja juurisyitä).
- saadaanko työpaikoilla riittävästi tietoa tapaturmien syistä niiden toistumisen estämiseksi (arviointiasteikko 1=ei ollenkaan... 5=riittävästi).

4 Tulokset ja tulosten tarkastelu

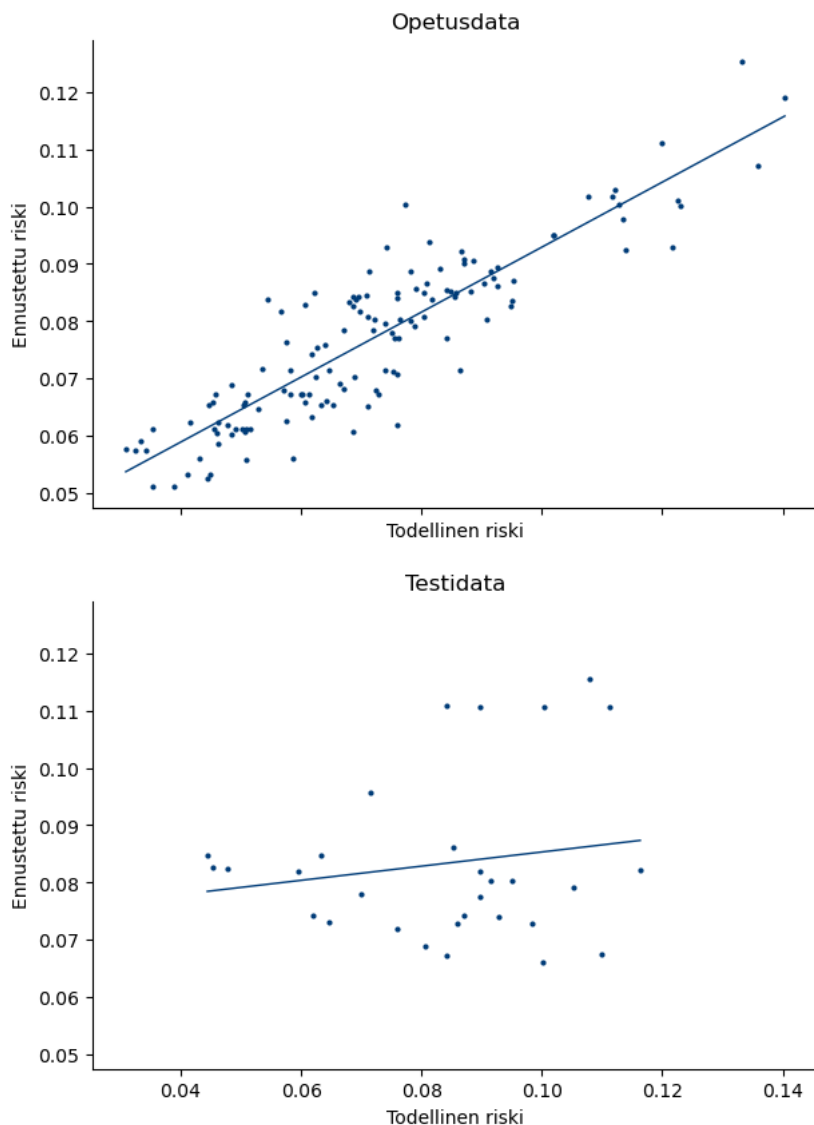
4.1 Ennustemallit ja korrelaatiot

Korrelaatiotarkastelulla (ks. 3.4) selvitettiin yksittäisten, organisaatioon liittyvien muuttujien vaikutusta turvallisuushavaintojen ja -tapahtumien määrään ja vakavuuteen. Korrelaatiot olivat yleisesti melko matalia, mutta esiin nousi muuttujia liittyen tuotannon tai tilausten määrään, sisäisen työn (esim. kunnossapito) määrään, työturvallisuuskoulutuksen määrään tai voimassaoloon sekä turvallisuuskierrosten havaintomäärään. Yleisiä havaintoja eri organisaatioiden korrelaatiotarkasteluista olivat muun muassa tuotannon määrän positiivinen korrelaatio turvallisuustapahtumien määrän ja vakavuuden kanssa. Tämä selittynee sillä, että lisääntynyt toiminta ja mahdollinen kiire lisäävät myös turvallisuustapahtumien määrää.

Turvallisuuskoulutuksen yhteys turvallisuustapahtumiin vaihteli organisaatiokohtaisesti. Toisaalta koulutuksen lisääminen vähentää läheltä piti -tapauksia ja tapaturmia, mutta toisaalta saattaa myös madaltaa kynnystä esimerkiksi läheltä piti -tapausten raportointiin, jolloin niiden määrä lisääntyy.

Yleisellä tasolla ennustemallit (ks. 3.5) eivät pystyneet ennustamaan tulevaa riskitasoa kovin luotettavasti. Tämä johtunee osin mallinnetun ilmiön laajuudesta ja pitkästä aikavälistä (ennustetaan koko organisaation riskiä samalla mallilla käyttäen dataa useamman vuoden ajanjaksolta, jonka aikana esimerkiksi muutokset organisaatiossa voivat vaikuttaa kerättävään dataan) sekä osin siitä, että kaikkia riskiin vaikuttavia tekijöitä ei ollut saatavilla rakenteellisena datana. Yhden organisaation kohdalla turvallisuustapahtumia oli kertynyt uuteen järjestelmään siirtymisen myötä niin vähän, että luotettavaa ennustemallia ei saatu rakennettua. Kolmessa muussa organisaatiossa mallinnus onnistui ainakin kohtuullisesti.

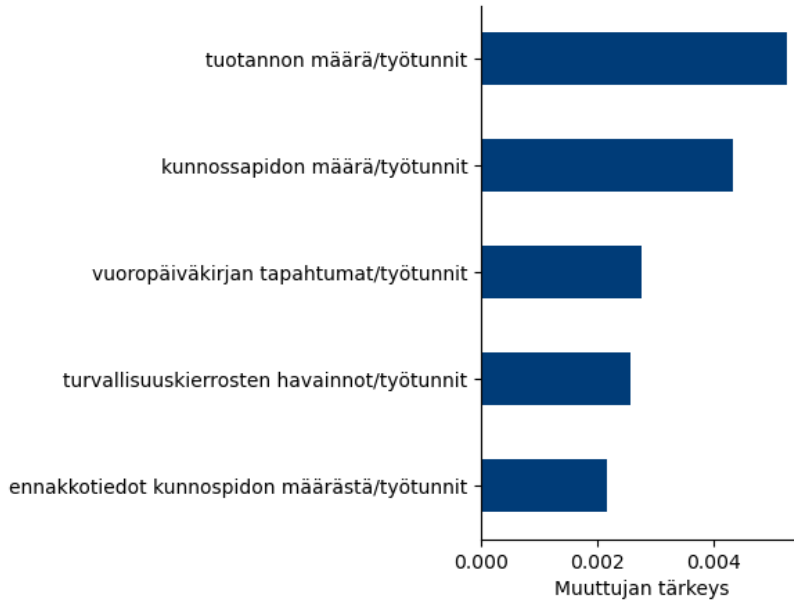
Kuva 3 esittää ennustemallin suoriutumista opetusdatalla ja testidatalla yhdessä osallistujaorganisaatiossa. Kuvaajasta nähdään, että malli ennustaa opetuksesta riippumattomalla testidatalla yleisesti pienempiä riskejä kuin pitäisi. Toisaalta mallin ennustaessa isompaa riskitasoa se on yleensä oikeassa, eli virheet ovat pääosin vääriä negatiivisia.



Kuva 3: Hankkeessa kehitetyn riskitasoa ennustavan mallin suorituskyky opetus- ja testidatalla yhden organisaation osalta.

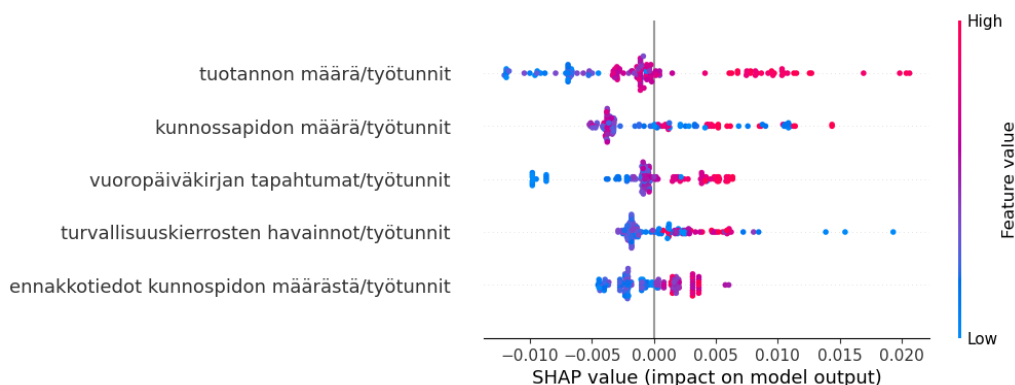
Ennustemallien analyysi Shapley-arvoilla kertoo siitä, mitkä selittävät muuttujat ovat mallin mielestä merkittävimpiä ennustetta muodostettaessa. Kuva 4 esittää esimerkkiorganisaation mallin mukaisesti eri muuttujien suhteellisen tärkeyden riskitason kannalta. Tässä tarkastelussa mallin tärkeimmät muuttujat ovat niitä, joihin

vaikuttaminen tai joiden arvoihin reagoiminen saattaa auttaa ehkäisemään turvallisuusriskien kasvamista.



Kuva 4: Ennustemallin muuttujien tärkeys riskitason kannalta esimerkkiorganisaatiossa.

Shapley-arvojen avulla voidaan myös analysoida tarkemmin muuttujien vaikutusmekanismeja mallissa. Koska kyseessä on epälineaarinen malli, vaikutuksia ei voi kuvata yhdellä muuttujakohtaisella kertoimella, vaan riippuvuudet voivat olla monimutkaisempia. Kuva 5 esittää esimerkkimallin muuttujien eri arvojen vaikutusta riskitasoon: värikoodaus kuvaa muuttujan saamaa arvoa, ja pisteen sijainti vaaka-akselilla kuvaa sen vaikutusta riskitasoa ennustettaessa. Enemmän oikealla sijaitsevat pisteet aiheuttavat riskitason nousua ja vasemmalla sijaitsevat laskevat riskiä. Yleisellä tasolla kuvaajasta nähdään, että korkea tuotannon määrä suhteutettuna työtunteihin nostaa riskiä, kun taas matalampi tuotantomäärä laskee sitä. Riskin riippuvuus sisäisestä kunnossapidosta puolestaan on mallin mukaan vähemmän suoraviivainen.



Kuva 5: Shapley-arvot kertovat selittävien muuttujien merkityksen mallin ennusteen kannalta. Punaiset arvot kuvaavat muuttujan suurempia arvoja ja siniset pienempiä. Vaaka-akselilla näkyy muuttujan vaikutus ennustettavaan suureeseen eli keskimääräiseen riskiin.

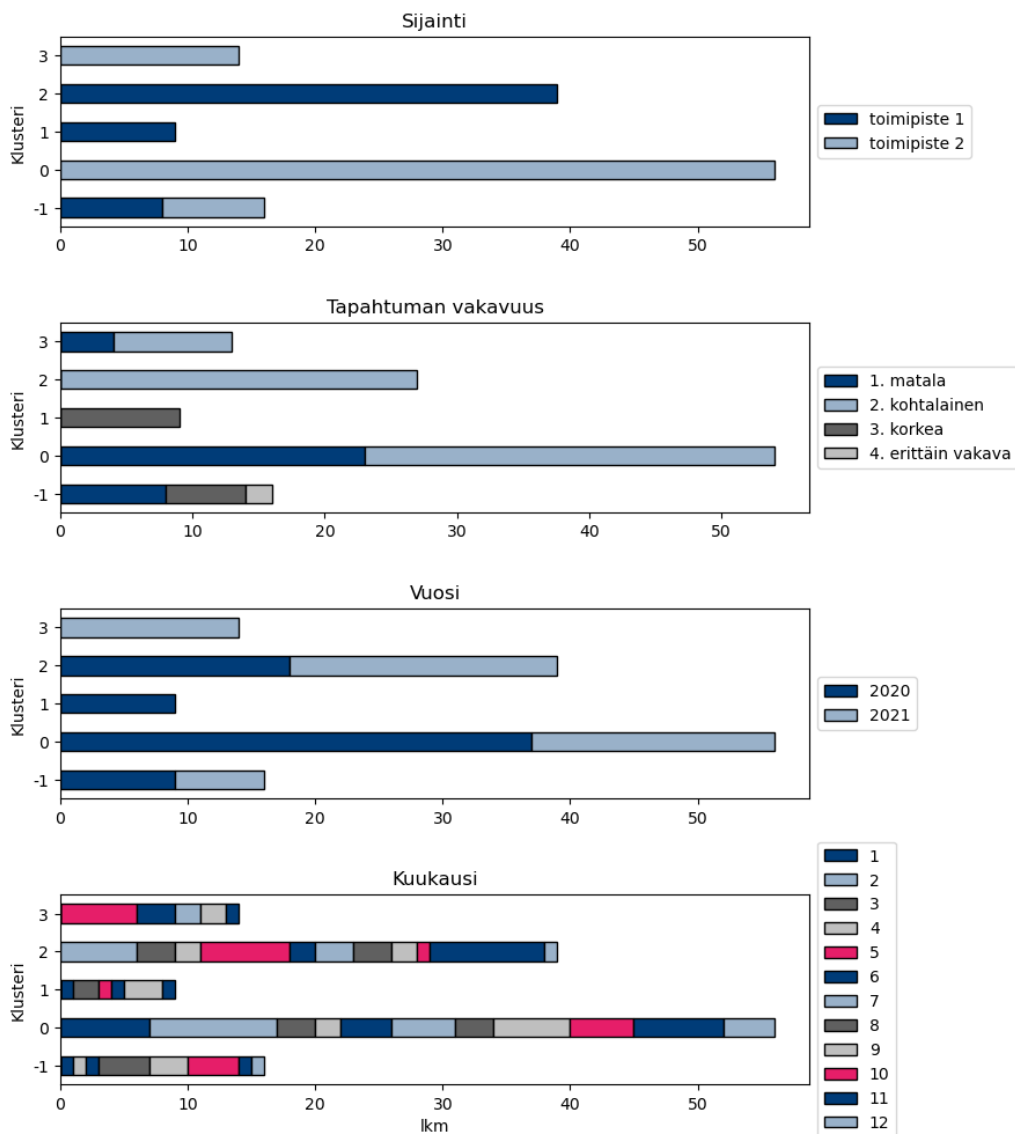
Yleisesti ennustemallinnuksen perusteella saadut tulokset olivat organisaatiokohtaisia, koska eri organisaatioiden datoissa oli vain osin samoja suureita mittaavia muuttujia. Turvallisuusperehdytyksillä ja -koulutuksilla havaittiin olevan pääasiassa negatiivinen vaikutus riskiin, eli ne vähensivät ennustettua riskiä, kuten voidaankin olettaa. Toisaalta turvallisuushavaintojen määrä sekä turvallisuuskierrosten tai turvatuokioiden havaintomäärä ja osallistumisaktiivisuus vaikuttivat riskiin kohottavasti, mikä selittynee sillä, että kun turvallisuuspuutteita havaitaan enemmän, myös läheltä piti -tilanteiden ja tapaturmien määrä ja vakavuus kasvavat. Korrelaatioanalyysistä vastaavasti myös ennustemallit tunnistivat suuremman tuotannon ja tilausten vaikuttavan riskiä kohottavasti.

4.2 Klusterointi

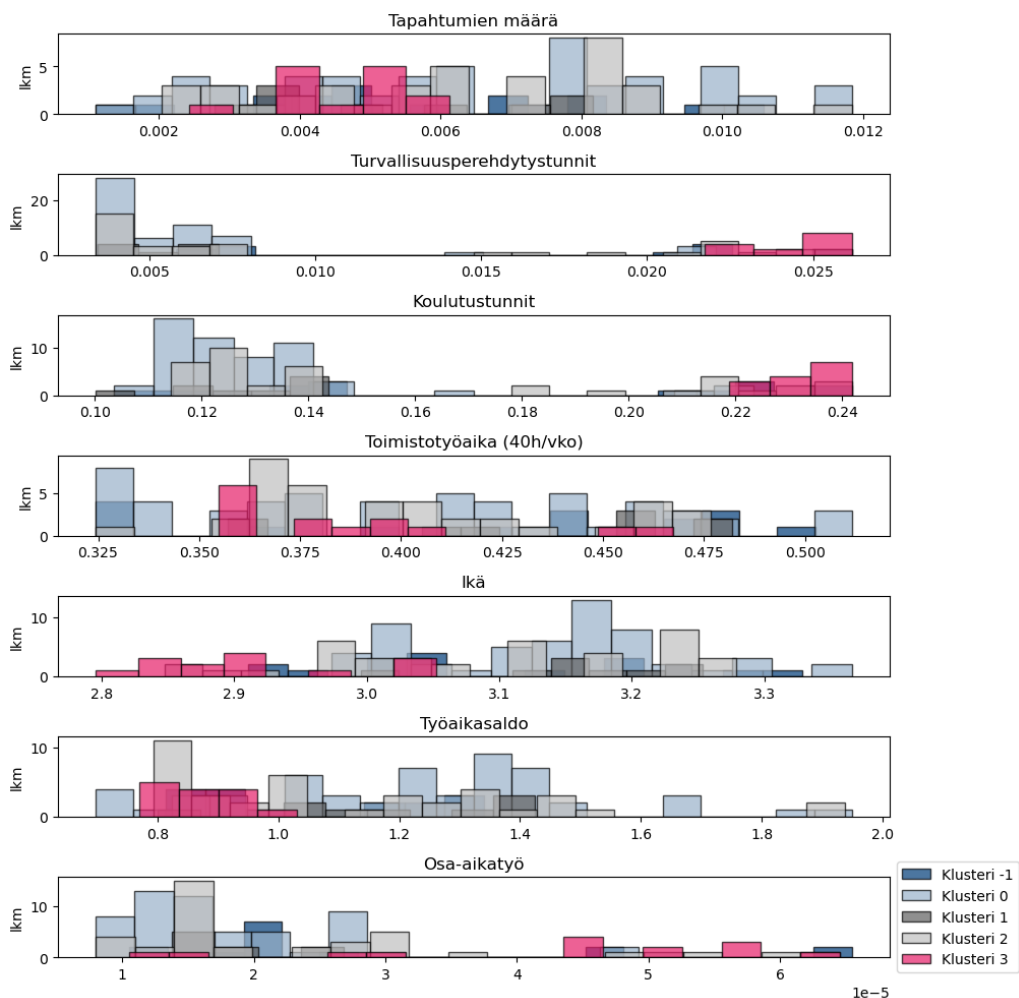
Klusteroinnissa (ks. 3.6) tavoitteena oli löytää datan perusteella erilaisia turvallisuustapahtumaryhmiä sekä tunnistaa, minkälaiset tapahtumat jäävät luokittelun ulkopuolelle. Klusterointi onnistui eri organisaatioissa vaihtelevasti riippuen siitä, kuinka kattavasti eri muuttujia pystyttiin datassa yhdistämään turvallisuustapahtumiin ja mitä dataa oli saatavilla. Yksi hyvin toiminut klusterointi on esitetty kuvissa 6 ja 7. Siinä tapahtumat eri toimipisteistä on saatu eroteltua omiin ryhmiinsä (toimipiste 1: klusterit 1 ja 2; toimipiste 2: klusterit 0 ja 3). Klusteri -1 kuvaa algoritmin tunnistamaa kohinaa eli tapahtumia, jotka eivät sovi mihinkään muodostettuun ryhmään.

Tarkemmalla tarkastelulla voidaan havaita esimerkiksi, että klusteri 3 edustanee turvallisuustapahtumia, jotka liittyvät kesäharjoittelijoiden työskentelyyn (tapahtuvat

kesäkuukausina, niiden aikoihin on tehty turvallisuuskoulutusta ja -perehdytystä enemmän kuin keskimäärin, työntekijöiden keski-ikä on ollut matalampi ja osa-aikatyö yleisempää). Turvallisuustapahtumiin liittyvien kuvausten tarkastelu vahvisti osaltaan tätä tulkintaa. Toimipisteen 2 osalta klusterit 1 ja 2 erottuvat toisistaan tapahtumien vakavuuden osalta, sillä klusterissa 1 kaikki tapahtumat on luokiteltu ”korkea”-vakavuusluokkaan, kun klusterissa 2 ne ovat luokassa ”kohtalainen”.



Kuva 6: Turvallisuustapahtumien luokittelu kategoristen muuttujien osalta esimerkkiorganisaatiossa.



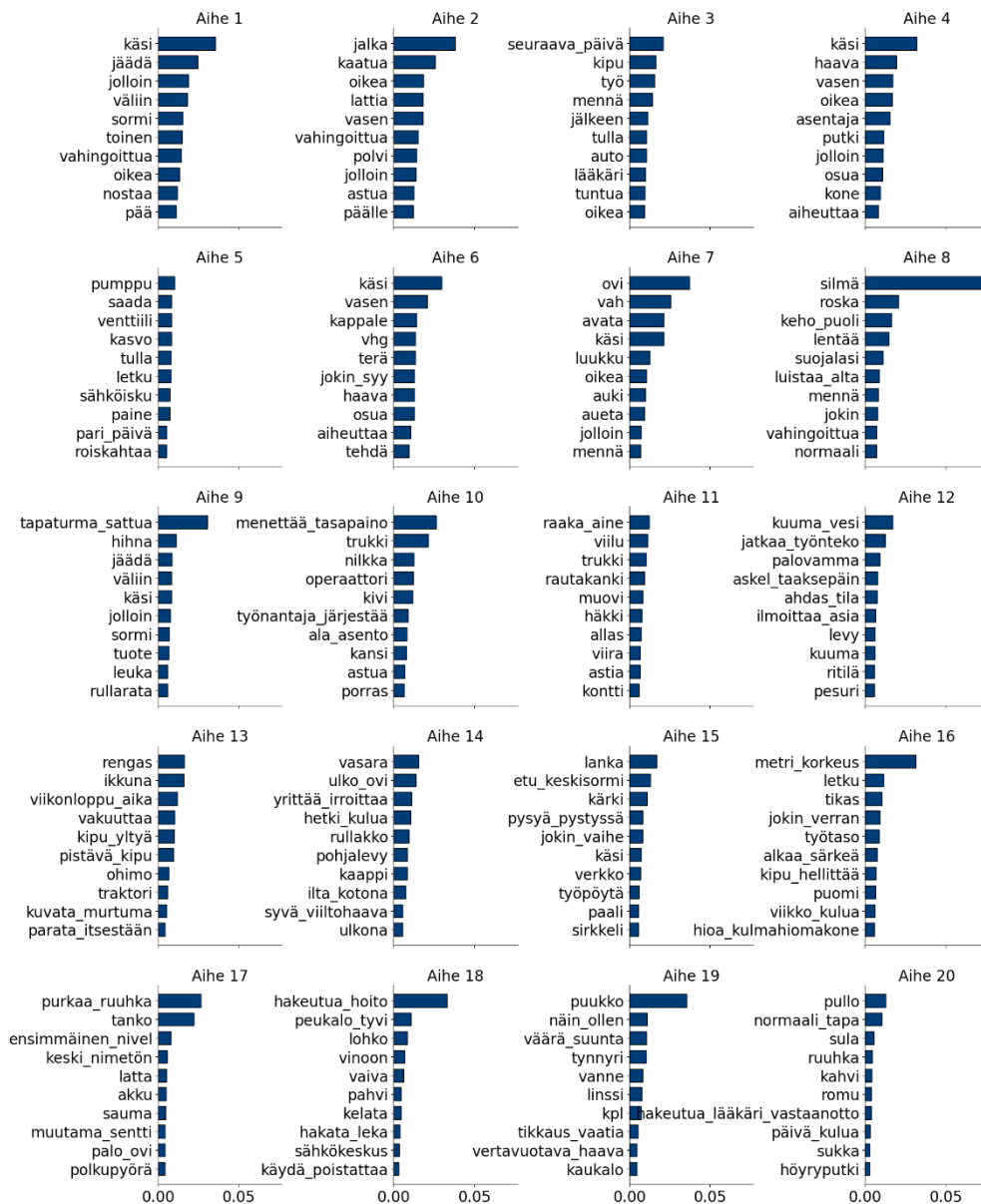
Kuva 7: Turvallisuu tapahtumien luokittelu jatkuva-arvoisten muuttujien osalta esimerkkiorganisaatiossa.

4.3 Aihemallinnus

Aihemallinnuksessa (ks. 3.7) käytetyille LDA-aihemalleille täytyy syöteaineiston lisäksi kertoa haluttu aiheiden lukumäärä. Aiheiden määrän lisäksi mallille voi määrittää myös muita parametreja, jotka vaikuttavat aiheiden sanajakauman sekä dokumenttien aihejakaumien muotoon. Aihemallinnus suoritettiin Pythonin (versio 3.8) Gensim-paketilla (versio 4.3.1.; Řehůřek ja Sojka 2010).

Kartoitimme vaihtoehtoja laskemalla eri parametrivalinnoilla eri LDA-malleja käyttäen osaa (80 %) aineistosta koulutusaineistona, ja laskemalla mallin *hämmentyneisyysarvon* (perplexity value) käyttäen loppua (20 %) aineistosta testiaineistona. Mallin hämmentyneisyysarvo on monotoninen sen uskottavuuden (likelihood) kanssa, ja se kvantifioi, kuinka hyvin koulutettu malli selittää testiaineiston rakennetta. Lopulta valitsimme hämmentyneisyysarvojen perusteella mielekkäimpien parametrivalintavaihtoehtojen tuottamien mallien joukosta silmämääräisesti parhaiten tulkittavissa olevan mallin. Mallissa on 20 aiheetta, ja sen aiheiden todennäköisimmät sanat on esitetty Kuvassa 8.

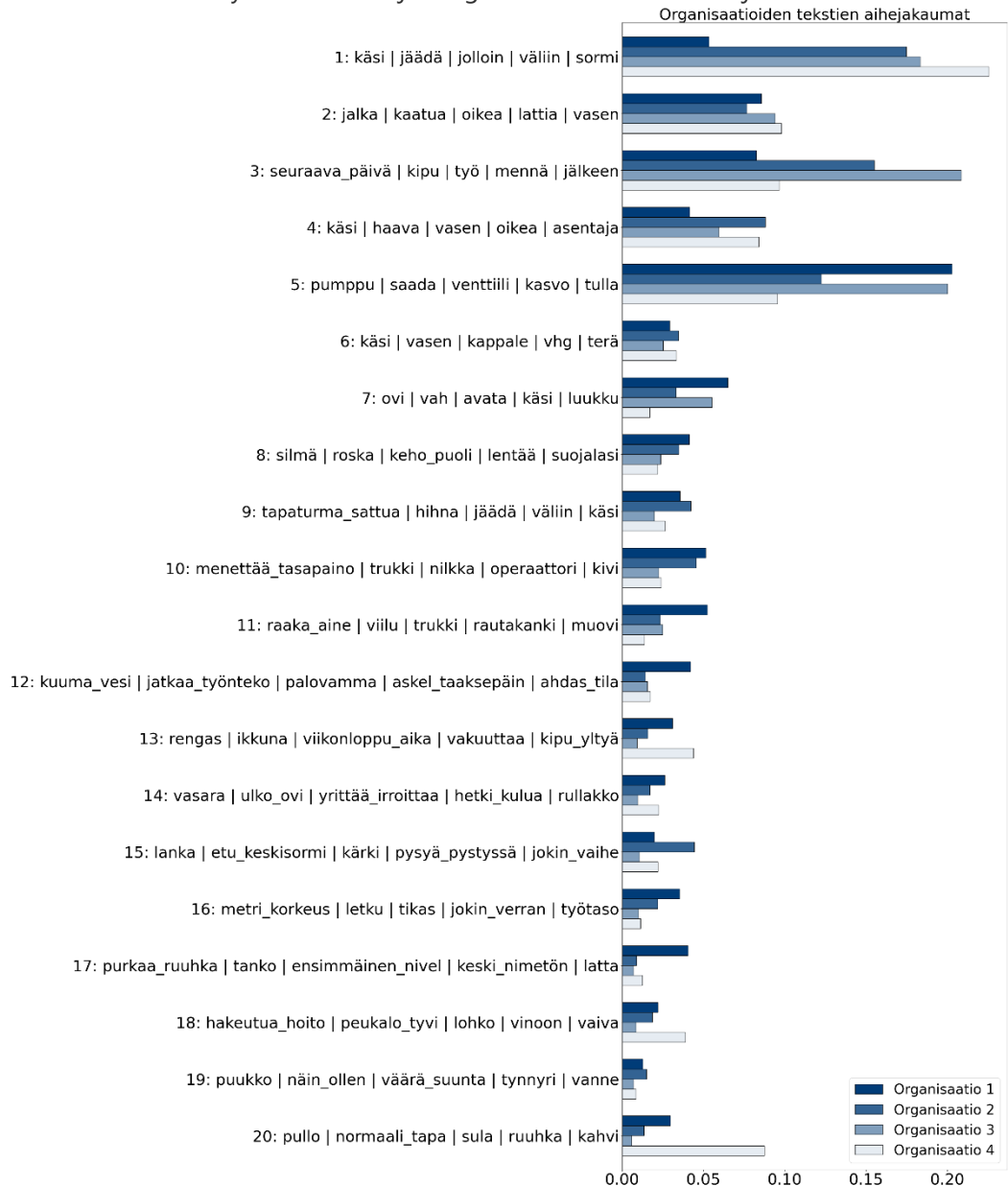
LDA-mallin aiheiden todennäköisimmät sanat



Kuva 8: LDA-aihemallin aiheiden todennäköisimmät sanat ja niiden todennäköisyydet.

Mallin koulutuksen tuloksena organisaatioiden vahinkokuvausteksteille muodostui aihejakaumat, joiden avulla voidaan vertailla eri organisaatioiden tekstiaineistoja.

Kuvassa 9 on esitetty hankkeen neljän organisaation tekstien aihejakaumat.

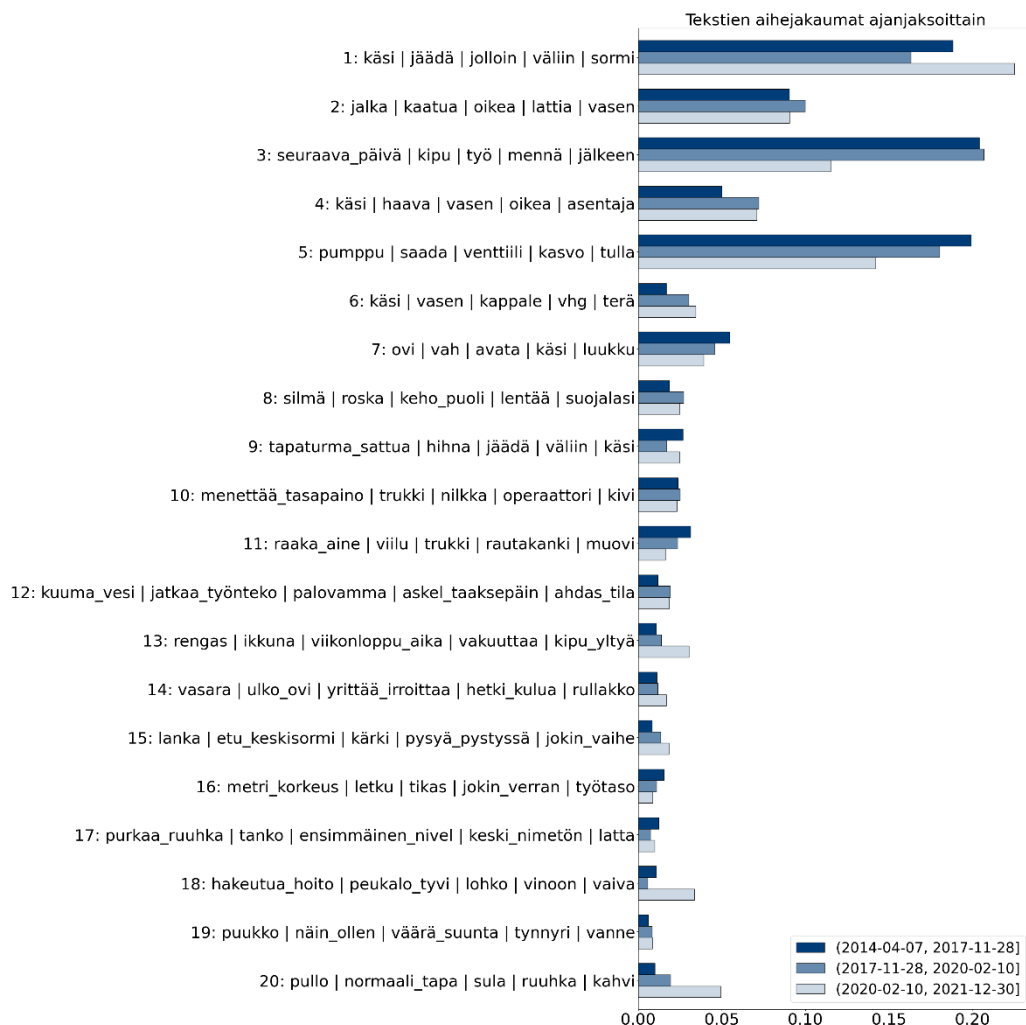


Kuva 9: Vahinkokuvaustekstien aihejakaumat organisaatioittain.

Kuvassa pylvään pituus kertoo, millä todennäköisyydellä kunkin organisaation tekstiaineistosta satunnaisesti valittu sana on mallin mukaan tietyn aiheen generoima. Aihe 1 (*käsi-jäädä-väliin-sormi*) korostui erityisesti yrityksissä 2, 3 ja 4, aihe 3

(seuraava_päivä-kipu-työ-mennä-jälkeen) organisaatioissa 2 ja 3, ja aihe 5 (*pumppu-haava-vasen-oikea-asentaja*) organisaatioissa 1 ja 3. Tämä perusteella näyttää siltä, että aihemalli pystyi havaitsemaan organisaatioiden välisiä eroja eri aiheiden esiintyvyydessä.

Aiheiden esiintyvyyttä organisaatioiden vahinkokuvausteksteissä voidaan tarkastella myös ajan yli. Tätä varten ajanjakso, jolta organisaatioiden kuvaustekstejä hankkeessa kertyi, jaettiin kolmeen ajanjaksoon niin, että kultakin kolmelta ajanjaksolta oli kertynyt yhtä monta tekstiä. Kuvassa 10 on esitetty aiheiden esiintyvyydet näiden kolmen ajanjakson kuvausteksteissä siten, että neljän yhteistyöorganisaation aihe-esiintyvyyksistä on otettu keskiarvo. Pylvään pituuden voi siis tulkita Kuvassa 10 niin, että jos neljästä organisaatiosta valitaan satunnaisesti yksi, ja kolmesta ajanjaksosta valitaan satunnaisesti yksi, niin mikä on sen todennäköisyys, että tuolta organisaatiolta tuolta ajanjaksolta kertyneistä teksteistä satunnaisesti valittu sana on tietyn aiheen generoima.

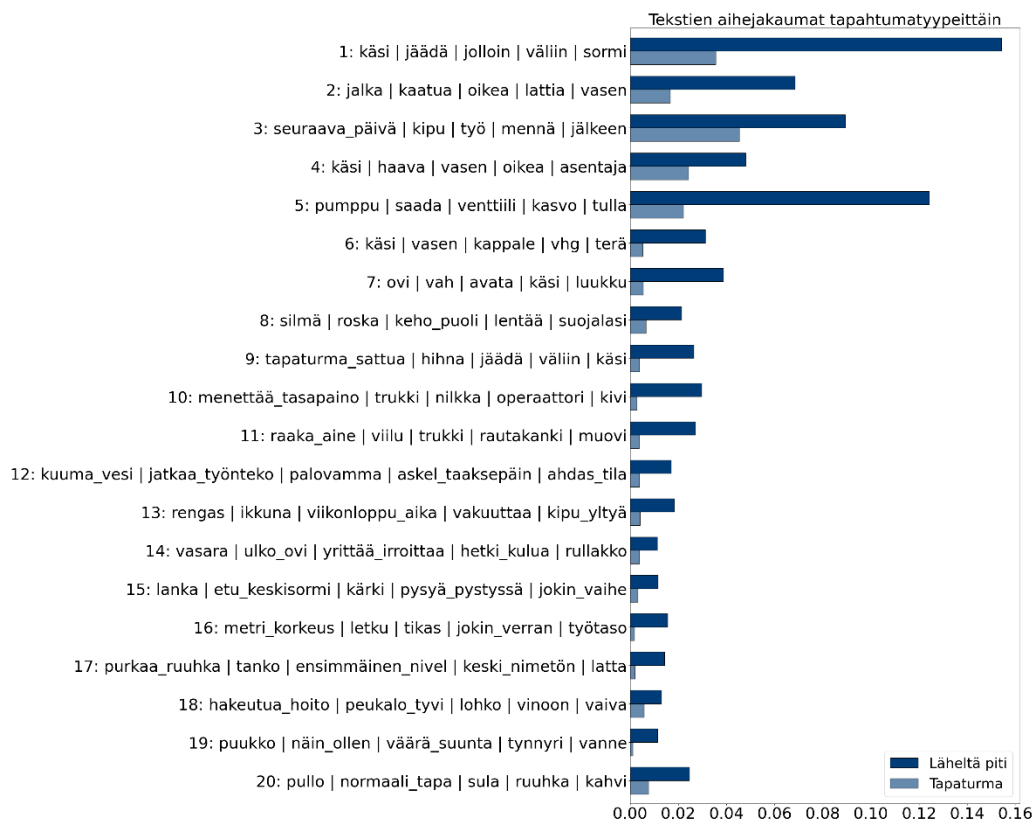


Kuva 10: Kaikilta organisaatioilta koottujen vahinkokuvaustekstien aihejakaumat ajanjaksoittain. Aika, jolta tekstejä on aineistossa, on jaettu kolmeen ajanjaksoon niin, että kultakin ajanjaksolta on kertynyt yhtä monta tekstiä.

Ajanjaksoittain tarkasteltuna nähdään, että aiheiden 3 ja 5 esiintyvyys on vähentynyt uusinta ajanjaksoa kohden, kun taas aiheissa 1, 13, 18 ja 20 esiintyvyys näyttäisi puolestaan kasvaneen.

Lisäksi eri tapahtumatyyppien – tarkastelussa siis vaara- (eli läheltä piti) -tilanteiden sekä tapaturmien – vahinkotekstien aiheita voidaan vertailla keskenään. Kuvassa 11 esitetään näiden kahden tapahtumatyyppin vahinkokuvaustekstien aihejakaumat, jälleen organisaatioiden keskiarvoina. Pylvään pituus kertoo siis, millä todennäköisyydellä

organisaation teksteistä satunnaisesti valittu sana on tietyn aiheen generoima ja tietyn tapahtumatyyppin tekstistä, kun ensin on valittu organisaatio satunnaisesti.



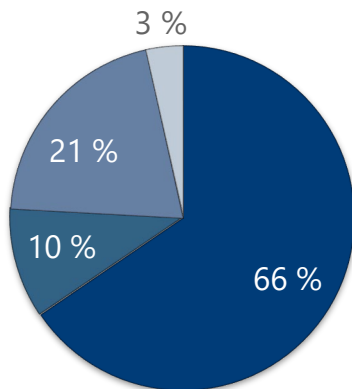
Kuva 11: Kaikilta organisaatioilta koottujen vahinkokuvaustekstien aihejakaumat jaettuina vaaratilanne- ("läheltä piti") ja tapaturmatilanteittain.

Kuten kuvasta 11 nähdään, läheltä piti -tilanteissa suurimmat esiintyvyydet ovat aiheilla 1 ja 5, kun taas tapaturmien kohdalla suurimmat esiintyvyydet ovat aiheilla 3 ja 1.

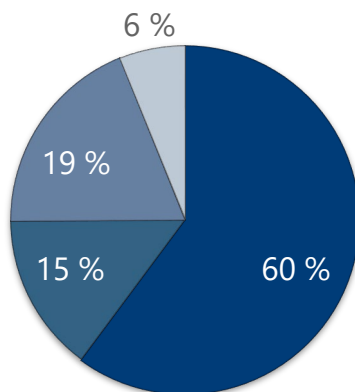
4.4 Inhimillisten tekijöiden koneoppimismalli

Kun suuri kielimalli oli jatkokoulutettu edellä esitellyllä koulutusmateriaalilla, ajettiin kunkin yrityksen turvallisuusdatat mallin 'läpi'. Mallin ajaminen on graafisella prosessointiyksiköllä (GPU) verraten nopeaa, eräänkin yrityksen tuhansia rivejä sisältävä aineisto käsiteltiin alle varttitunnissa. Tässä kappaleessa esitellyt kuvaajat tuovat näkyviin eroja yritysten tapahtumien, sekä raportoinnissa käytetyn kielen osalta.

Organisaatio A



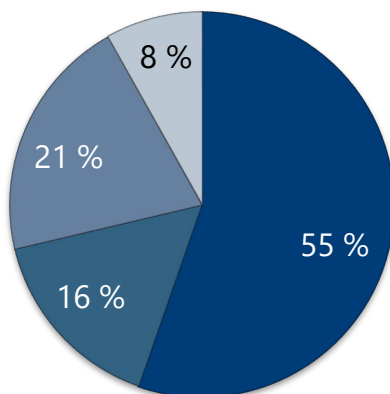
Organisaatio B



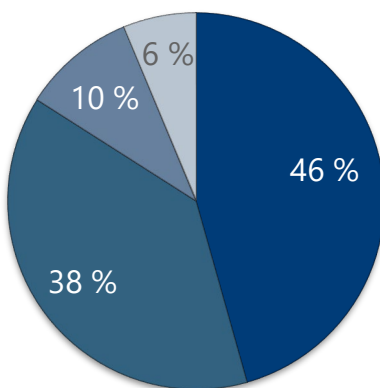
- Yksilön toiminta ja piirteet
- Työtoiminta ja työn piirteet
- Ryhätason tekijät
- Organisaatiotason tekijät

Kuva 12: HF-4 koneoppimismallilla analysoitujen organisaatiodatojen organisaatiokohtaiset inhimillisten tekijöiden jakaumat, organisaatio A ja B

Organisaatio C



Organisaatio D



- Yksilön toiminta ja piirteet
- Työtoiminta ja työn piirteet
- Ryhmätason tekijät
- Organisaatiotason tekijät

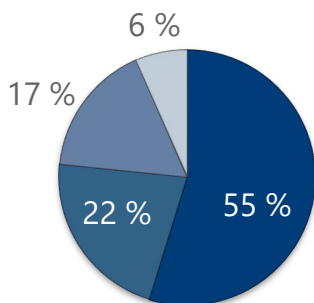
Kuva 13: HF-4 koneoppimismallilla analysoitujen organisaatiodatoiden organisaatiokohtaiset inhimillisten tekijöiden jakaumat, organisaatio C ja D

Inhimillisiä tekijöitä analysoitiin koneellisesti (ks. 3.8) neljältä yhteistyöorganisaatiolta ja Tapaturmavakuutuskeskuksen datasta. Kuvassa 13 olevissa kuvaajissa esitetään organisaatiokohtaisesti koneoppimismallin tunnistamat tapaturma- ja läheltä piti - tapahtumien jakaumat inhimillisten tekijöiden neljällä ylätasolla (yksilön toiminta ja piirteet, työtoiminta ja työn piirteet, ryhmätason tekijät sekä organisaatiotason tekijät).

Organisaatio A:n jakauma ei kuvaa organisaation tapaturma- ja läheltä piti -tilanteita kovin luotettavasti, sillä muiden organisaatioiden kanssa vertailukelpoista dataa oli melko vähän. Kaikkien yritysten kohdalla koneoppimismalli havaitsi vähiten organisaatiotason tekijöitä. Tämä viittaa siihen, että organisaatiotason tekijöitä ei aina tunnisteta ja/tai raportoida, koska hyvin usein muiden tasojen (yksilö, työtoiminta, ryhmä) tekijöiden taustalla olisi tunnistettavissa organisaatiotason taustatekijä tai -tekijöitä.

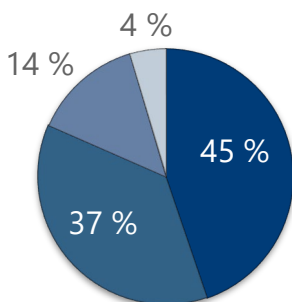
Organisaatio B:n ja Organisaatio C:n jakaumat ovat likipitään samat. Organisaatio D:n jakauma näyttää eroavan jonkin verran muista organisaatioista, ja siinä malli on tunnistanut teksteistä enemmän työtoimintaan ja työn piirteisiin liittyviä tekijöitä kuin muissa organisaatioissa. Tähän saattaa vaikuttaa Organisaatio D:n tapa raportoida tai kirjata kuvaustekstejä turvallisuushavainnoista, mikä voi olla erilaista muihin organisaatioihin verrattuna tai esimerkiksi poikkeamantutkimamalli saattaa ohjata tunnistamaan korostuneesti tietyn tyyppisiä tekijöitä. Toisaalta kyse voi olla myös käytetystä sanastosta: koneoppimismalli voi tulkita tietyt sanat työtoimintaan ja työn piirteisiin liittyviksi, vaikka todellisuudessa ne kuuluisivat jollekin toiselle tasolle.

Läheltä piti + Tapaturmat: Organisaatiot



- Yksilön toiminta ja piirteet
- Työtoiminta ja työn piirteet
- Ryhmätason tekijät
- Organisaatiotason tekijät

Läheltä piti + Tapaturmat: TVK (Teollisuus)



- Yksilön toiminta ja piirteet
- Työtoiminta ja työn piirteet
- Ryhmätason tekijät
- Organisaatiotason tekijät

Kuva 14: HF-4 koneoppimismallilla analysoitujen neljän yrityksen yhdistelmä jakauma ja Tapaturmavakuutuskeskuksen datan inhimillisten tekijöiden jakauma.

Kuvassa 14 on esitettyä ylemmässä kuvaajassa kaikkien yhteistyöorganisaatioiden yhdistelmäjakama ja alemman kuvaajan osuudet kuvaavat Tapaturmavakuutuskeskuksen tapaturmadatan vahinkokuvausten HF-4 kuvaajaa rajattuna teollisuusalan yrityksiin.

TVK:n datasta malli on tunnistanut selvästi enemmän Työtoiminnan ja työn piirteitä kuin tutkimukseen osallistuneiden organisaatioiden datassa. Organisaatioiden datassa puolestaan on enemmän yksilötasolle kirjattuja tekijöitä. Koska TVK:n dataa on satoja tuhansia rivejä, analyysi ihmisvoimin on likipitään mahdotonta. Arvioimme, että TVK:n datan jakaumassa korostuva Työtoiminnan ja työn piirteiden suurempi esiintyvyys liittyy TVK:n datan luonteeseen. Tapaturmia ylös kirjanneet tahot ovat puhuneet hyvin tapaturmamekanismikeskeisesti, ja ulkoiset seikat (ympäristö, käytetyt välineet ja koneet, jne.) tulevat selvästi esiin teksteissä.

Koska TVK:n data eroaa organisaatiodatasta luonteeltaan, keräystarkoitukseltaan, kirjoitustyyliltään ja monilta muilta piirteiltään, ei HF-4 malli luokitellut dataa aivan samoin kuin organisaatiodataa. Tästä huolimatta, neuroverkkomallin luokittelemia HF-4 ylätasoa käsitteitä kannattaa lähestyä kunkin ilmiön tunnusmerkkien kautta:

Yksilön toiminta ja piirteet käsittelee usein hyvin yksilökeskeisiä asioita, puhutaan henkilöistä, heidän toimistaan ja turvallisuusnäkökulmasta hyvin hienojakoisen tason asioista.

Työtoiminta ja työn piirteet sisältää usein ympäristöä, laitteita, välineitä, olosuhteita ja muita vastaavia kuvaavia tekstejä.

Ryhmätason ja Organisaatiotason tekijät käsittelevät yksilöön ja työtoimintaan liittymättömiä, tai niistä selvästi irrallisia käsitteitä, kuten kulttuuria ja johtamista, tiedonkulkua, organisatorisia muutoksia, tiimien toimintaa ja niin edelleen.

4.4.1 Havainnot HF-4 analyysin tuloksista

Tutkimuksessa kehitetty koneoppimismalli pystyi havaitsemaan eroja eri organisaatioiden välillä. Kaikkien organisaatioiden yhteistulos ei vastaa täysin asiantuntija-analyysissä tunnistettujen inhimillisten tekijöiden jakaumia, mutta TVK:n datassa ollaan huomattavasti lähempänä. Tämä tukee sitä havaintoa, että organisaatioiden poikkeamien tutkinnoissa ei aina tunnisteta tai raportoinnissa ei kirjata kattavasti eri tasojen tekijöitä. Vaikka TVK:n datassa havaittiin samoja ongelmia tekstien laadun, ja toisaalta kuvattujen ilmiöiden tarkkuuden suhteen, on TVK:n dataa lukumääräisesti niin paljon, että jakauma osoittaa tarkemmin tapaturmien ylätasojen jakaumaa ja keskimäärin kuvaukset olivat kuitenkin laajempia kuin organisaatiodatassa.

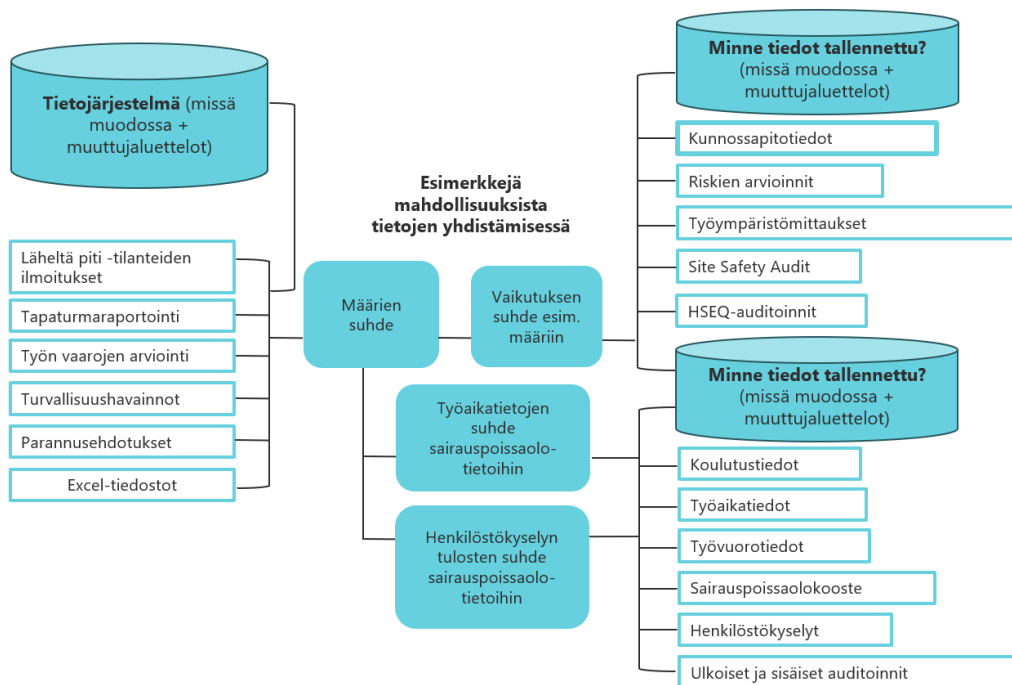
Arvoituamme mallin tarkkuutta käsin tehdyillä testiesimerkeillä, havaitsimme, että malli kiinnittää selvästi enemmän huomiota tekstin muotoiluun, kuin itse asiaan. "Minä, me, hän" -tyyppinen teksti luokitellaan siis herkemmin Yksilön toiminta ja piirteet -ylätasoon kuuluvaksi, kuin "Organisaation, yrityksen, tiimin" -kuvaus, vaikka tekstissä olisikin kyse organisaatio- tai ryhmätason tekijöistä. Olemme osin myötävaikuttaneet tähän, sillä laatiessamme koulutusdataa, olemme saattaneet ymmärtää kontekstin paremmin, kuin mitä pelkästä kuvaustekstistä on tulkittavissa ei-asiantuntijan tai koneen toimesta.

Mallin ennustukset tarjoavat mielenkiintoisen näkökulman organisaation turvallisuusdataan; Kuinka eri ylätasojen tekijät korostuvat, etenkin Yksilö-, Työ-, Ryhmä-, ja Organisaatio-tasolla. Mallimme arvioikin parhaiten tekstin painotusta edellä mainitun neljän tason näkökulmasta. Organisaatioiden datan laadun kehittäminen kattavammaksi ja monipuolisemmin eri tasojen tekijöitä huomioivaksi tuottaisi täsmällisempää opetusdataa ja mahdollistaisi tarkemman luokittelun.

4.5 Työpajojen tulokset

Turvallisuusjohtamisessa hyödynnettävät datat ovat pitkälti työturvallisuuden suoraan liittyviä. Datojen hyödyntämisessä on haasteena se, että niihin perehtymiseen ja analyysiin ei ole riittävästi aikaa tai osaavia tekijöitä organisaatioissa ja siten työturvallisuustilanteen järjestelmällinen ennakointi on vaikeaa. Hyödyntäminen on pitkälti tapauskohtaista analyysiä, kuvailevan analytiikan hyödyntämistä, erilaista seuranta- ja historiatiedon analyysiä. Toisaalta tietoa hyödynnetään myös painopisteiden tunnistamiseen, jolloin toimintaan saadaan lisää vaikuttavuutta.

Turvallisuuden vaikuttavien datojen yhdistäminen rikastuttaa työturvallisuuden tilannekuvaa merkittävästi (Kuva 15). Turvallisuusdataa käsitellään organisaatioissa järjestelmällisesti ja tilannetta seurattiin päivittäin työturvallisuuden tilannekuvan päivittämiseksi. Kerättyä dataa hyödynnetään myös muun muassa investointipäätösten kohdistamisessa, investointibudjettien laadinnassa ja turvallisuusinvestointien perusteluissa. Työturvallisuusdataa käytetään organisaatioiden tavoitteiden toteutumisen seurannassa, toimenpiteiden ja turvallisuusviestinnän suunnittelussa ja sisällössä, tapahtumien tapaustutkimuksissa ja juurisyyanalyysissä, historiatiedon analyysissä sekä aktiivisuuden ja muutostrendien tarkkailussa.



Kuva 15: Tiedonlouhinta erilaisilla datoilla

Turvallisuusjohtamisen datankäytön ja tiedolla johtamisen ideaalitalanne kuvaa sellaista datankäytön toivetilaa, jossa kaikki tekniset ja taloudelliset rajoitteet datoihin liittyen on poistettu.

Tiedolla johtamisen ideaalitalanteessa

- Tiedot olisivat mahdollisimman reaaliaikaista ja aina luotettavia.
- Tiedot kerääntyisivät mahdollisimman automaattisesti tai vähällä vaivalla suoraan muista datajärjestelmistä tai esimerkiksi anturien tai kameroiden välityksellä.
- Tietojärjestelmät olisivat helppokäyttöisiä ja eri tietojärjestelmät keskustelisivat keskenään.
- Ihmisten käyttäytymisestä ja toiminnasta saataisiin enemmän tietoa.
- Mobiililiittymän käyttö pakottaisi tekemään ja kuittaamaan viimehetken riskienarvioinnin ennen työn aloittamista.

- Tietojärjestelmä kokoaisi kaikki työn suorittamisen kannalta olennaiset tiedot yhteen lähtötiedoiksi työn aloittamiselle tai esimerkiksi seisokin suunnitteluun.
- Käytössä olisi kokoava käyttöliittymä, joka koostaisi kaiken olennaisen tiedon visuaaliseen, havainnolliseen ja helposti hyödynnettävään muotoon, mutta jonka näkymän kautta yksittäinenkin havainto tai tilanne olisi jäljitettävissä ja tutustuttavissa
- Tietojärjestelmä yhdistäisi eri datoja ja pystyisi ennustamaan tulevaisuuden turvallisuustilannetta
- Data-analytiikka antaisi valmiit ratkaisuehdotukset havaittuun ongelmaan.

Alkutyöpajojen ja -haastattelujen aineistoa tarkasteltiin myös siitä näkökulmasta, miten kohdeyritysten johto ja työturvallisuustoimijat tällä hetkellä tuovat turvallisuuteen liittyvässä puheessaan esiin inhimillisiä tekijöitä ja niihin liittyviä kehittämistarpeita. Inhimillisiä tekijöitä käsiteltiin laajana käsitteenä – sisältäen yksilöön, työhön, ryhmään ja organisaatioon liittyviä inhimillisiä tekijöitä. Työpajoissa ja johdon haastatteluissa ei suoraan kysytty asiaa, mutta tilaisuuksissa puhuttiin mm. turvallisuustilanteesta, sen kehittämistarpeista.

Joissakin organisaatioissa inhimillisten tekijöiden hallinta oli tunnistettu ja kuvattiin nykyiseksi tai seuraavaksi turvallisuuden johtamisen kehittämiskohteeksi. Tiedolla johtamisen ideaalitulanteessa useat johtajat ja turvallisuuden kehittäjät sanoittivat tavoitteeksi ”päästä ihmisen pään sisään” tai ymmärtää ”miksi yksilöt tekevät sellaisia valintoja tai päätöksiä työssään kuin he tekevät”. Tai miksi jollekin ryhmälle saattaa sattua muita useammin tietyn tyyppisiä poikkeamia. Tavoitteena olisi tunnistaa sitä kautta esimerkiksi tuen tai koulutuksen tarpeita erilaisille työntekijäryhmille.

Toisaalta havaittiin, että monilla näistä inhimillisten tekijöiden osa-alueilla oli tehty kehittämistä (positiivista kehittämispuhetta, toteutettuja hankkeita), joka myös tunnistettiin ja sanoitettiin.

Haastatteluaineistoa tarkasteltiin siitä näkökulmasta minkälaisista inhimillisistä tekijöistä ihmiset puhuvat ja havainnollistettiin mihin puhe keskittyi muutaman esimerkin kautta. Eniten lukumääräisesti esiin nousi organisaatiotason (45 %) HF-tekijöitä. Puhuttiin turvallisuuskulttuurista, organisaation toiminnasta ja toimintakulttuurista sekä kokonaisuusien ymmärtämisestä. Seuraavaksi eniten tunnistettiin työn piirteitä (27 %), kuten työn sisältöä ja määrää, kiirettä, töiden orgnaisointia, ohjeita ja fyysisen

työympäristön tekijöitä. Kolmanneksi eniten esiin nousi yksilötason tekijöitä (16 %). Näistä esiin nousivat ammattitaito, tilannetietoisuus ja motivaatio sekä asenteet. Ryhmän toimintaan, ryhmdynamiikkaan tms. liittyvää puhetta nousi esiin selkeästi vähemmän kuin muiden tasojen asioita osallistujien puhuessa oman työpaikkansa turvallisuudesta ja sen kehittämisestä. Ryhmätason tekijöitä ei välttämättä osata vielä tunnistaa ihmisen toimintaan vaikuttavina tekijöinä ja sitä kautta turvallisuuteen vaikuttavina tekijöinä.

Alkuhaastattelu- ja alkutyöpaja-aineiston valossa näyttää siltä, että organisaatioissa on tavalla tai toisella jo hyvinkin tiedostettu, ainakin turvallisuusroolissa olevien ja johtajien osalta, inhimillisten tekijöiden vaikutus turvallisuuteen.

Organisaatioiden tulostyöpajoissa tutkijat esittivät tutkimuksessa havaittujen asioiden pohjalta ehdotuksia työturvallisuustoiminnan kehittämiseksi yhteisen keskustelun pohjaksi.

- Työturvallisuustoimintaa edistäisi, jos positiiviset havainnot ja onnistumiset saataisiin esiin entistä paremmin ja tapahtumatutkinnoissa analysoitaisiin vahvemmin inhimilliset tekijät.
- Myös työturvallisuuskoulutusten entistä tehokkaampi hyödyntäminen ja sisältöjen kehittäminen toisi toimintaan lisää vaikuttavuutta.
- Tuotannon muutosten ja organisaatiomuutosten tietojen kerääminen ja niiden käyttäminen antaisi hyvän historiatiedon aikajanelle yhdistettäväksi turvallisuusdataan.
- Nyt muutostieto ei ole joko saatavilla lainkaan, se on sirpaleista tai pelkästään hiljaista tietoa.
 - o Organisaatioiden tulee kehittää hyvät käytännöt ja tekniset ratkaisut muutostietojen keräämiseen.
- Tapahtumista oppimisen prosessin analysointi ja prosessin hyödyntäminen oppimisen näkökulmasta voisi toimia kehittämisen pohjana.
- Hyvinvointinäkökulmaan liittyvän datan saaminen data-analyysiin mukaan vaatii kehittämistä.
- Tietosuojarajoitusten tuomien datankäyttöesteiden ratkaisemiseen pitäisi kehittää hyviä käytäntöjä.

Tulostyöpajoissa nousi esiin myös tietojärjestelmien ominaisuus, jossa toimenpiteet ovat joskus pitkän aikaa "käsittelyssä"-tilassa eli toimenpiteitä ei saada valmiiksi.

Eryteisesti toimenpiteet, jotka edellyttävät toteutuakseen investointeja ja hankintoja, ovat pitkään tietojärjestelmissä avoinna eikä niiden vaiheittaista enenemistä yleensä pysty kirjaamaan tietojärjestelmään.

Tulostyöpajoissa luotiin näkökulmia työturvallisuuden tilannekuvan luomiseen datapohjaa laajentamalla:

- Riskienarviointidatan ajantasaisuus ja päivitykset tietoihin saatavilla useammin kuin vain kerran vuodessa.
- Riskienarviointidatan ja turvallisuushavaintodatan yhdistäminen paikkatasolla ja samojen avainsanojen käyttäminen. Yhdistävänä tekijänä esimerkiksi paikkatieto ja tarkalla tasolla työsolut.
- Käyttäytymiseen perustuvat riskit ja riskienhallinnan tason muutokset saatava näkyviin.
- Käyttäytymisen havainnointia tehtävä objektiivisesti.
- Kunnossapitodatoissa näkyvissä sekä ennakoiva että korjaava huolto, toteutettu muutos ja tarkka sijaintipaikka.
- Määräaikaistarkastuksista saatava datana tilannekuvaus, toimenpiteet ja varmistettava tiedonkulku.
- Muutostietojen saaminen koskien organisaation, tuotannon, työympäristön muutosten tunnistamista, yhteistä systemaattista dokumentaatiota ja seuranta. Esimerkiksi avainhenkilöt/roolimutokset, järjestelmä-, vaatimus-, tavoitemuutokset, bonusvaikutukset, tuotannon/työympäristön/koneiden muutokset, loma-ajat ja fuusiot. Tätä voidaan käyttää työturvallisuustilanteen ennustamiseen. Muutostietojen tiedonkulku eri muodoissa, tarvitaan sekä "puskaradio"-efektiä että intraa.
- Yhteistyökumppanitietojen käyttämöhdöllisuus, esimerkiksi määrät, muutokset ja palaute sekä verkostojen työturvallisuusdatan käyttö, esimerkiksi vartiointi.
- Henkilöstöhallinnon datan aito hyödyntäminen tavoitteena. Sisältäen muun muassa ylityö-, koulutus-, osaamistiedot ja henkilömäärät.
- Tapahtumahetkellä voidaan kerätä henkilöstöhallinnon tietoa ja taustoja, mutta vaatii erillisdokumentointia.
- Toimenpiteisiin merkittävässä "valmis"-tila vasta, kun työsolu on kirjattu.

- Turvallisuuskoulutusdatan yhdistäminen henkilöstöhallinnon dataan. Myös esimiesten käyttöön.

Työpajassa mietittiin myös yhdessä seuraavia kehittämisenäkökulmia liittyen tekstidatan laadun, tietojärjestelmien käytön sekä tekniikan kehittämiseen.

Tekstidatan laadun kehittämisessä nähtiin merkitystä olevan hyvällä opastuksella ja koulutuksella, hyvillä esimerkeillä, motivoinnilla ja palkitsemisella. Laatua voidaan kehittää myös käyttämällä kirjaamiseen sanelua tai toisaalta valokuvia täydennettynä lyhyillä teksteillä. Suuria toiveita asetettiin myös tekoälyn mahdollisuuksiin olla reaaliaikaisesti täyttäjän apuna sekä tekoälyn tekemä tekstidatan sisältötarkistus. Valokuvien analysointi koneoppimisen menetelmin vaatii vielä tekniikan kehittämistä. Toisaalta tekstidatan laadun kehittämiseen nähtiin yhtenä askeleena HF Tool - tutkintamenetelmän käyttämisen tutkinnoissa.

Tietojärjestelmien käytön kehittämisessä koettiin tärkeäksi käyttökokemuksen parantaminen. Kehittäminen tulee tehdä käyttäjänäkökulmasta, myös peruskäyttäjien tarpeet tulee tunnistaa ja heitä tulee tukea. Tavoitteena on järjestelmän käyttäminen itsenäisesti, mutta se ei saa vaatia liikaa osaamista. Tietojärjestelmien käyttöoikeudet mahdollistavat ja rajaavat. On otettava huomioon omat ja muut käyttäjät sekä erilaiset käyttötavat – kuten mobiilijärjestelmät ja QR-koodien käyttö – että niihin liittyvät käytännön ominaisuudet. Myös paikkatiedon (alue- ja solujaot) ja taustatietojen, kuten ikä, työkokemus, työaika, perehdytys, koulutus ja työtehtäväkokemuksen keräämistä halutaan kehittää.

Tietojärjestelmien tekninen kehittäminen edellyttää niiden välistä tiedonsiirtoa ja toimivia liityntäpintoja. Teknistä kehittämistä on myös käyttäjäoikeuksien roolitus aina dokumentoinnista raportointiin. Tietojärjestelmien teknisessä kehittämisessä on keskiössä kielen ja määritelmien ymmärrettävyys, yksinkertaisuus ja helppous sekä tiedon löydettävyys eli mahdollisimman lyhyet polut, vähän tasoja ja klikkauksia. Tietojärjestelmien tekniikkaan kaivattiin myös helpommin tehtäviä muutoksia ja automaattista tiedon linkitystä eri järjestelmiin. Raportointiin halutaan lisää fiksua automatisointia, visuaalisuutta, käytettävyyttä ja muokattavuutta sekä mahdollisuuksia omiin muokkauksiin. Automaattiset ja visuaaliset raporttipohjat ovat haluttuja. Koettuun käytettävyyteen vaikuttaa myös järjestelmän luomis-/täyttönopeus eri käyttötilanteissa sekä tietojen ajantasaisuus. Tekniset ominaisuudet vaikuttavat myös tiedon hyödynnettävyyteen.

Kaikkien motivointi työturvallisuuteen on perusta kaikelle työturvallisuustekemiselle ja -toiminnalle. Muuttuva maailma ja uudet ikäluokat töissä täytyy ottaa huomioon. Uusilla

ikäluokilla on erilaiset odotukset myös työturvallisuustoiminnalle. Heillä on erilaista osaamista, josta esimerkkeinä erilaiset käden taidot, tietotekninen osaaminen, kokemukset ja kokemattomuus sekä sitoutumisaika yhteen työpaikkaan. Uusia ikäluokkia tulee osata motivoida heille sopivilla tavoilla ja samalla oppia yhdessä uudenlaisia toimintatapoja.

Työturvallisuusdatalla johtamisessa yksi näkökulma on miten tapaturmien tutkinta ja riskinarviointi on toteutettu. Tavoitteena on selkeyttää vastuukysymyksiä ja parantaa läpinäkyvyyttä tutkintojen ja toimenpiteiden sekä riskienarvioinnin osalta tietojärjestelmissä. Läpinäkyvyys mahdollistaa tilanteen seurannan, lisäresursoinnin ja valvonnan johtamisen näkökulmasta. Datan pitäisi olla niin laadukasta, että turhia palaverieita ei tarvita vaan dataa pystytään aidosti hyödyntämään johtamisessa kaikkialla organisaatiossa. Seuranta helpottaa, kun prosessin toimenpiteiden tilannekuvaus on riittävän kuvaava. Nykyisin jaottelu on yleisesti ”aloittamatta, kesken ja valmis”, missä kesken-vaihe monesti venähtää pitkäksi eikä sellaisenaan ole riittävän kuvaava kertomaan toiminnasta. Työturvallisuusdatalla johtamista helpottaisi, jos eri järjestelmät olisivat linkittyneet toisiinsa niin tiedonsiirron, toimintakuittausten kuin merkittävyyksien tunnistamisen osalta.

4.6 Tapaturmatutkintojen HF-asiantuntija-analyyseiden tulokset

4.6.1 Organisaatio- ja ryhmätason tekijöitä tunnistetaan harvemmin

Tapaturmatutkintojen HF-analyyseissä (ks. 3.10) yksilön toimintaan ja piirteisiin liittyviä tekijöitä (Taulukko 4) tunnistettiin toiseksi eniten heti työtoiminnan ja työn piirteiden jälkeen. Niitä oli reilu kolmannes (36 %), ja yleisimpiä olivat tilannetietoisuus (HF2), tilanteiden ennakointi (oletukset ja varmistaminen) (HF5) sekä ohjeiden ja sovittujen toimintatapojen noudattaminen (HF3). Yksilön toimintaan ja piirteisiin liittyen määriteltiin kuitenkin vain vähän toimenpiteitä (13 kpl, eli noin 6 % kaikista toimenpiteistä).

Taulukko 4: Poikkeamaraporteissa tunnistetut yksilötason inhimilliset tekijät.

Inhimillinen tekijä	Lkm vaikuttavana tekijänä	Lkm toimenpiteitä	Esimerkkejä aineistokatkelmista (muokattu tunnistamattomiksi)
HF 1. Ammattitaito, työn hallinnan taso	10	2	"Työntekijä ei tiennyt, miten tunnistaa viiltosuojakäsineet." (HF-) "Työntekijä oli kokenut ja ammattitaitoinen." (HF+)
HF 2. Tilannetietoisuus	32	0	"Työntekijän ajatukset eivät ollut suoritettavassa tehtävässä, hän suunnitteli seuraavaa tehtävää." (HF-) "Pidetään ajatus kirkaana liikkuessa paikasta toiseen. Kohteessa tasoero lattiassa jonka kyllä huomaa jos ollaan huolellisia liikuttaessa." (toimenpide)
HF 3. Normien ja sovittujen toimintatapojen noudattaminen	18	8	"Työntekijä ei noudattanut hyviä työkäytäntöjä tai turvallisuusohjeita" (HF-) "Viiltosuojahanskojen käyttö jatkuvasti, vaikka terien kanssa ei sillä hetkellä työskenneltäisiin." (toimenpide)
HF 4. Kokonaisuuden ymmärtäminen	0	0	-
HF 5. Tilanteiden ennakointi	23	3	"Hän otti suuren riskin, koska ei pysähtynyt ajattelemaan mitä tekee, eikä täyttänyt vaarojen arviointivihkoa." (HF-) "Jos tulee ongelmia, pitää pysähtyä ja miettiä turvallinen tapa jatkaa." (toimenpide)
HF 6. Yli- tai alikuormitus ja niiden hallintakeinot	1	0	"[Työtehtävän suorittamisessa] oli ollut ongelmia." (HF-)
HF 7. Vireystila, väsymysoireet	0	0	-
HF 8. Elämäntilanne, huolet, yleinen stressitaso	0	0	-
HF 9. Ikä, työkokemuksen määrä ja laatu	13	0	"Työntekijä oli hyvin kokenut ja tunsi pakkaustyön hyvin." (HF+) "Työryhmän kokemattomuus saattoi vaikuttaa työtapaan ja olosuhteisiin." (HF-)
HF 10. Yleinen terveystilanne	2	0	"Varhaisen välittämisen malli" oli käytössä ko. henkilöllä." (HF-) "Henkilö oli tuntenut selässään kipua jo aiemmin." (HF-)
HF 11. Motivaatio, asenteet	8	0	"Oma asenne – henkilön suhtautuminen turvallisuuteen ei ole riittävällä tasolla." (HF-) "Henkilön aiemmat kokemukset viiltosuojakäsineistä, että eivät suojaa käsiä." (HF-)
HF 12. Tunnereaktiot, mieliala	1	0	"Kiireen tuntu johtuen itseasetetusta määräajasta projektille, yksilön halu saada projekti valmiiksi." (HF-)
YHTEENSÄ YKSILÖTASON TEKIJÖITÄ	108	13	

Eniten noin 46 % kaikista tapaturmaraporteissa tunnistetuista inhimillisistä tekijöistä liittyi työtoimintaan ja työn piirteisiin (ks. Taulukko 5). Yleisimmin tunnistettuja työtason tekijöitä olivat työmenetelmät ja ohjeet (HF24), laitejärjestelmien ja tekniikan toimivuus ja käytettävyys (HF23) sekä fyysinen työympäristö ja työolosuhteet (HF29).

Taulukko 5: Poikkeamaraporteissa tunnistetut työtoimintaan ja työn piirteisiin liittyvät inhimilliset tekijät.

Inhimillinen tekijä	Lkm vaikuttavana tekijänä	Lkm toimenpiteitä	Esimerkkejä aineistokatkelmista (muokattu tunnistamattomiksi)
HF 20. Työn laatu ja sisältö	2	0	"Kaksi työntekijää olivat täysin työllistettyjä käynnissä olevassa siirto-operaatioissa." (HF-) "Joutui käyttämään kohtuuttomasti voimaa." (HF-)
HF 21. Työn määrä, aikapaine, kiire	11	2	"Työntekijä ei tee pääsääntöisesti kyseistä työtä ja hänet oltiin siirretty kyseiseen tehtävään/tiimiin resurssien vuoksi." (HF-) "Pidetään huolta siitä, että ei anneta tuotantopaineen vaikuttaa turvallisuuskeskusteluihin työryhmässä ja esimiehen kanssa." (toimenpide)
HF 22. Työnjako, tehtävänkuvaukset, töiden organisointi, selkeys	15	8	"Työhön ei ollut saatavissa paria, koska työpari työskenteli muualla." (HF-) "Työparien muodostaminen siten, että toinen on kokenut ja toinen aloittelija" (toimenpide)
HF 23. Laittejärjestelmien/tekniikan toimivuus ja käytettävyys	31	21	"Naulapistoolissa ei ole ominaisuutta, joka estäisi sen laukeamisen kahdesti." (HF-) "Työkalut, kulmahiomakone ja talja, kunnossa ja tarkastetut." (HF+)
HF 24. Työmenetelmät ja ohjeet	38	36	"[Työpaikalla] ei ole ohjeita ja menettelyä kuljettajien turvallisuusperehdytykseen." (HF-) "Työn aloituksen vaarojen arviointivihko täytetty ja tarkastettu." (HF+)
HF 25. Vaikutusmahdollisuudet omaan työhön ja työoloihin	1	7	"Liukuestepohjalliset rikki, ei ollut saanut uusia." (HF-) "Mietitään yhdessä korjaavia toimenpiteitä, että vastaavaa ei jatkossa satu. Keskustellaan tavaroiden kantamisesta rappusissa ja mietitään vaihtoehtoja kantamiselle." (toimenpide)
HF 26. Työstä saatu palaute, ammatillinen arvostus	0	0	-

HF 27. Mahdollisuus/kyky arvioida ja kehittää omia työprosesseja	1	1	<i>"Keskusteluissa olisi hyvä tuoda esiin se, että kannattaa pysähtyä hetkeksi ja arvioida riskit ennen töiden aloittamista. Lisäksi havainnoista ja kokemuksista olisi syytä tehdä ilmoitukset ennen kuin mitään sattuu. Tapauksesta ei ollut mitään kirjauksia, mutta kaikilla näytti olevan kosolti kokemuksia työkalun lipsahduksista ja holkin äkillisistä aukeamisista. Kukaan ei ollut noteerannut em. mitenkään tärkeäksi / miksikään turvallisuusasiaksi ennen LTI:tä." (HF- ja toimenpide)</i>
HF 28. Koulutus	8	9	<i>"Työntekijä on ollut vuosia [yrityksen] palveluksessa ja hän on sisäistänyt työturvallisuus asiat kurssien ja käytännön kokemuksella." (HF+) "Tutkinnan aikana huomattiin, että työturvallisuuskortti oli vanhentunut." (HF-)</i>
HF 29. Fyysinen työympäristö, työolosuhteet, työhygieeniset tekijät	31	18	<i>"Tilanahtauden vuoksi tuotetta ei voinut tuoda omalle paikalleen trukilla." (HF-) "Lattia oli tasainen, puhdas ja kuiva, eikä siinä ollut tavaroita." (HF+)</i>
YHTEENSÄ TYÖTASON TEKIJÖITÄ	138	102	

Ryhmätason tekijöitä tunnistettiin poikkeamaraporteissa kaikkein vähiten, alle 7 prosenttia kaikista inhimillisistä tekijöistä liittyi ryhmätasoon (Taulukko 6). Kuitenkin 17 % poikkeamaraporteissa määritellyistä toimenpiteistä liittyi ryhmätason tekijöihin – lähinnä tiedonkulkuun (HF35). Toimenpiteet jäivät kuitenkin suurimmaksi osaksi hyvin ylätasolle, esimerkiksi "käydään tapaus ja juurisyysanalyysin tulokset läpi kaikissa vuoroissa" tai tehdään tiedotusmateriaalia info-TV:hen.

Taulukko 6: Poikkeamaraporteissa tunnistetut ryhmätason inhimilliset tekijät.

Inhimillinen tekijä	Lkm vaikuttavana tekijänä	Lkm toimenpiteitä	Esimerkkejä aineistokatkelmista (muokattu tunnistamattomiksi)
HF 30. Yhtenäinen kuva tilanteesta kaikilla jäsenillä	4	0	"Työnjohto ei käynyt kohteessa, eikä asia tullut ilmi asentajien ja operaattorien keskustelussa." (HF-) "Kommunikaatio nosturinkuljettajan ja käsittelijän välillä tärkeää; nosturinkuljettaja ohjaa nosturia näytettyjen merkkien mukaisesti." (HF+)
HF 31. Ryhmän kaikkien jäsenten tietämyksen hyödyntäminen	2	6	"Parasta käytettävissä olevaa ammattitaitoa/kohdetuntemusta ei hyödynnetty" (HF-) "Jos jokin ihmetyttää, keskusteltava kollegoiden kanssa." (toimenpide)
HF 32. Väärinkäsitykset, -tulkinnat, -kuulemiset sekä näiden korjaaminen	1	1	"Parannetaan kommunikaatiota ja varmistetaan, että työryhmän jäsenien tehtävät ovat selvät, erityisesti merkinantajan tehtävät ja varmistukset ennen toimintamerkkiä." (HF- ja toimenpide)
HF 33. Ryhmän rakenne ja kiinteys, muu ryhmädynamiikka	6	1	"Epäselvät roolit ja vastuut." (HF-) "Kukaan paikallaolija ei "noteerannut" riskinottoa, eikä puuttunut siihen työn pysäyttäen." (HF-)
HF 34. Kommunikaatio eri yhteistyötahojen kesken	0	2	"Tapaus raportoidaan [toimijalle] ja siitä keskustellaan [tiimissä]." (toimenpide) "Havaittujen vaarojen merkitseminen ja viestintä. Ei jätetä toisille ansoja." (toimenpide)
HF 35. Tiedonkulku	4	27	"[Henkilöt] eivät olleet tietoisia suoritettavasta kunnossapitotyön sisällöstä, eikä täten huomioinut tyhjennystarvetta." (HF-) "Turvatuokio extra -materiaali ruokalan infotelevisioihin." (toimenpide)
HF 36. Päätöksenteko ryhmässä	3	0	"Työn suunnittelussa ja työparin kommunikoinnissa oli parannettavaa, koska [laitteen] päätä ei olisi kannattanut vetää yksin." (HF-) "Käytetty työtapo arvioitiin helpoimmaksi ja nopeimmaksi tavaksi." (HF-)
YHTEENSÄ RYHMÄTASON TEKIJÖITÄ	20	37	

Organisaatiotason tekijöitä tunnistettiin poikkeamaraporteissa noin 11 % kaikista tekijöistä (Taulukko 7). Yleisimmin organisaatiotason tekijät liittyivät kokonaishallintaan ja keskinäiseen ymmärrykseen toistensa töistä (HF43), organisaatiotasolla tehtyihin päätöksiin (HF44) ja yhteistyöhön yhteistyökumppaneiden kanssa (HF46). Toimenpiteitä organisaatiotasolle määriteltiin kuitenkin 80 kappaletta, joka on jopa 35 % kaikista toimenpiteistä. Useat organisaatiotasolle määritellyt toimenpiteet jäivät kuitenkin hyvin ylätasolle, kuten useassa raportissa toistuvat ”järjestetään juurisyytutkinta” ja ”pidetään tehtaanjohtajan turvallisuuspalaveri”.

Taulukko 7: Poikkeamaraporteissa tunnistetut organisaatiotason inhimilliset tekijät.

Inhimillinen tekijä	Lkm vaikuttavana tekijänä	Lkm toimenpiteitä	Esimerkkejä aineistokatkelmista (muokattu tunnistamattomiksi)
HF 40 Johtamistapa ja -järjestelmä	3	9	<i>”Tehtäväksiannon yhteydessä oli puhuttu liukkaudesta, pimeydestä ja ahtaudesta.” (HF+)</i> <i>”Tapaturmatutkinta ja töiden keskeyttäminen. Henkilö työkieltoon [tapahtuman] käsittelyn ajaksi.” (toimenpide)</i>
HF 41. Organisaatio-/toimintakulttuuri	5	28	<i>” ”Aina tehty näin” ”Aiemmin ei ole sattunut tapaturmaa” (HF-)</i> <i>”Tehdään juurisyyanalyysi” (toimenpide)</i>
HF 42. Eri organisaatiotasojen ja -tahojen välinen yhteistyö	0	5	<i>” Selvitetään muiden tehtaiden työmenetelmiä vastaaviin tilanteisiin.” (toimenpide)</i> <i>”Käydään tapaus läpi [turvallisuusryhmän] seuraavassa kokouksessa keskeisine havaintoineen ja toimenpiteineen. (toimenpide)</i>
HF 43. Kokonaishallinta, keskinäinen ymmärrys toistensa töistä	10	13	<i>”Turvallisuusnäkökohtia ei suunniteltu työtilaukselle.” (HF-)</i> <i>”Riskien uudelleen arviointi.” (toimenpide)</i>
HF 44. Tehdyt päätökset	7	3	<i>”Edellisenä päivänä päätettiin, että [laite] tyhjennetään yöllä, jolloin alueen eristäminen on helpompaa.” (HF+)</i> <i>”Erityistöiden tunnistaminen osana seisokkien suunnittelua.” (toimenpide)</i>
HF 45. Muutoksen hallinta	1	1	<i>”Ns. [aineita] tulee käyttöön osittain hallitsemattomasti siten, että niiden vaaratekijöitä ei arvioida kunnolla etukäteen. Niitä käytetään, vaikka niiden vaaratekijät eivät ole aina tiedossa.” (HF-)</i> <i>”Selvitetään kaikkien ns. [aineiden] käyttötarkoitukset ja valitaan asianmukaiset aineet, joista laaditaan hyväksytyjen aineiden luettelo.” (toimenpide)</i>

HF 46. Yhteistyökumppanit	7	21	<i>"Letkun tyyppi/materiaali on saattanut muuttua vuosien kuluessa, eikä siitä ole ilmoitettu toimittajalle." (HF-) "Neuvotaan ja ohjeistetaan ostoa ilmoittamaan toimittajalle pakkausvaatimuksesta."(toimenpide)</i>
HF 47. Konsernipalveluiden tuki yksikölle	0	0	-
YHTEENSÄ ORGANISAATIO- TASON TEKIJÖITÄ	33	80	

Kuten yllä olevista taulukoistakin on nähtävissä, poikkeamaraporttiaineistosta puuttuu kokonaan maininnat seuraavista inhimillisistä tekijöistä; yksilötasolla vireystila (HF7) tai elämäntilanne, huolet, yleinen stressitaso (HF8), työtasolla työstä saatu palaute, ammatillinen arvostus (HF26) ja organisaatiotasolla konsernipalveluiden tuki liiketoiminnoille (HF47).

4.6.2 Huolimattomuuden ja inhimillisen virheen syitä ei selvitetä

43 %:ssa poikkeamaraporttien tekstikuvauksista oli mainittu tapauksen syynä huolimattomuus ja varomattomuus, esimerkiksi seuraavasti:

"Huolimattomuus - kuljettaja ajoi tolppaan jostain syystä."

"Tapaturma sattui henkilön tehdessä työtä harkitsemattomasti ja liian suurta voimaa käyttäen."

"Nilkka nyrjähti huolimattomasti valvomosta poistuttaessa."

"Ajattelemattomuus."; "Henkilö oli ajattelematon."

"Huolimattomuus, inhimillinen tekijä." "Huomaamattomuus."

"Tutkintaryhmä ei löytänyt muuta korjaavaa toimenpidettä kuin huolellisuus ja varovaisuus."

"Inhimillinen erehdys"

"Käyttäytyminen."

"Ei kiinnitetty riittävästi huomiota."

Toisen organisaation tapaturmien tutkintalomake ohjasi valitsemaan yhden tai useamman tapauksen juurisyyn valmiista vaihtoehtoista tai kirjaamaan "muu juurisyys" kenttään muun juurisyyn. Yhtenä juurisyysvaihtoehtoista oli käyttäytyminen, ja se oli juurisyysvaihtoehtoista kaikkein yleisimmin valittu. 72 %:ssa tapauksista oli valittu käyttäytyminen juurisyysiksi. Muita tutkintalomakkeen valmiita vaihtoehtoja juurisyysiksi olivat työmenettelyt ja ohjeet (56 %:ssa valittu), työkalut ja välineet (32 %:ssa valittu) ja koulutus (24 %).

Yritysten tapaturmaraportit herättivät paljon kysymyksiä "huolimattomuuden", "varomattomuuden" tai "inhimillisen virheen" taustalla vaikuttaneista tekijöistä, joihin raportit eivät tarjonneet vastausta. Kirjallisesta raportista ei käynyt ilmi, oliko asiaa selvitetty tai pyritty selvittämään tutkintatilanteessa tarkemmin (esim. vireystilaa, työkuormitusta, ryhmätason tekijöitä) mutta jätetty vain kirjaamatta valmiiseen raporttiin esimerkiksi tapauksessa, jossa selittävää tekijää näistä ei löytynyt.

Hankkeen työpajoissa selvisi, että tutkintatilanteen keskusteluissa käsitellään tarkemmin moniakin asioita, joita ei kuitenkaan kirjata tutkintaraporttiin. Tyypillisesti nämä käsittelevät yksilön vireyttä, työkuormitusta ja koettua stressiä. Erityisesti, kun nämä asiat ovat kunnossa, ei niitä koeta tarpeelliseksi kirjata. Joissain tutkintatilanteissa koetaan myös hankalaksi esittää tällaisia henkilökohtaisia kysymyksiä tai tietoa ei haluta antaa eikä raportoida eteenpäin.

4.6.3 Positiivisten tekijöiden ja onnistumisten analysointi tapaturmatutkinnassa

Valtaosa (noin 87 %) raporteissa kuvatuista ja edellä esitellyistä inhimillisistä tekijöistä oli tilanteen aiheuttaneita/tilannetta pahentaneita negatiivisia tekijöitä (HF-). Positiivisia tekijöitä tai hyvin hoidettuja asioita (HF+) oli tunnistettu ja kirjattu osassa raporteissa. Raporteissa mainitut positiiviset inhimilliset tekijät liittyivät mm.

- yksilötason tekijöistä ammattitaitoon ja työn hallinnan tasoon (HF1), ohjeiden noudattamiseen (HF3), tilanteiden ennakointiin (HF5) ja työkokemukseen (HF9).
- työtason tekijöistä työnjakoon, tehtäväkuvauksiin ja töiden organisointiin (HF22), laitejärjestelmien ja tekniikan toimivuuteen (HF23), työmenetelmiin ja ohjeisiin (HF24), koulutukseen (HF28) ja fyysiseen työympäristöön/olosuhteisiin (HF29).
- ryhmätason tekijöissä mainittiin vain yksi positiivinen inhimillinen tekijä, joka liittyi yhtenäiseen tilannekuvaan (HF30).

- organisaatiotason tekijöistä johtamistapaan/-järjestelmään (HF40), kokonaisuuden hallintaan, turvallisuuden ymmärrykseen (HF43) ja organisaation päätöksentekoon (HF44).

Kummankaan organisaation tapaturmatutkintalomakkeet/turvallisuuden tietojärjestelmän tekstikentät eivät ohjeistaneet tai rohkaisseet systemaattisesti hyvin hoidettujen asioiden tai onnistumisten esille nostamiseen tai ylös kirjaamiseen. Kunnossa olevia asioita tuli systemaattisesti kirjattua vain toisen organisaation tutkintalomakkeessa, jossa pyydettiin arvioimaan riskienarviointiin ja valvontaan liittyvien toimenpiteiden toteutumista Kyllä/Ei/Ei sovellu -asteikolla.

4.6.4 Kokonaisarviot raporteista ja niiden kielestä

Tarkemman inhimillisten tekijöiden luokittelujen jälkeen laaditut kokonaisarviot tutkintaraporteissa käytetystä kielestä ja kuvauksista (asteikolla 1-5) on tiivistetty Taulukkoon 8. Kuten taulukosta nähdään, tutkintaraportit vaihtelivat huomattavasti kaikkien arvioitujen kysymysten suhteen eikä yritys selitä vaihtelua, vaan molempien yritysten tutkinnat saivat arvioita asteikon ylä- ja alalaidasta. Tutkinnan laadun kannalta parhaan arvon, nro 5, saaneita raporteja on kaikilla mittareilla arvioituna vähän. Merkittävä osa raporteista saa arvoja 1-2, mikä kertoo että kyseiset raportit on kirjoitettu hyvin yksilölähtöisesti, niiden kerronta on syllistävää ja tutkintaraportin lukemisen jälkeen jää paljon avoimia kysymyksiä.

Taulukko 8: Kokonaisarviot tutkintaraporttien tekstikuvauksista.

	Systeemi- vai yksilölähtöinen? 1=erittäin yksilölähtöinen, 5=systeemilähtöinen (erittäin hyvin eri tasojen tekijöitä huomioon ottava)	Kuvauksen tyyli 1=syyllystävää, 3=toteavaa, neutraalia, 5=syyllystämätöntä, reilua, pyrkii ymmärtämään ihmisen toimintaa	Tapahtumaan johtaneiden syiden tunnistaminen 1= Paljon avoimia kysymyksiä, 5 = Syitä tunnistettu kattavasti	Saadaanko työpaikoilla riittävästi tietoa tapaturmien syistä niiden toistumisen estämiseksi? 1=ei ollenkaan, 5=riittävästi
Arvon 1 saaneiden osuus (%)	18 %	15 %	7 %	6 %
Arvon 2 saaneiden osuus (%)	16 %	21 %	30 %	31 %
Arvon 3 saaneiden osuus (%)	24 %	34 %	18 %	24 %
Arvon 4 saaneiden osuus (%)	36 %	21 %	39 %	34 %
Arvon 5 saaneiden osuus (%)	6 %	9 %	6 %	4 %
Keskiarvo	2,96	2,88	3,06	3,00

4.6.5 Keskeisimmät havainnot HF-asiantuntija-analyyseistä

Tapaturmatutkintojen laatu sekä tapaturmaan johtaneiden syiden ja ihmisen toiminnan taustalla vaikuttavien tekijöiden tunnistaminen vaihtelevat paljon organisaatioiden välillä mutta myös organisaation sisällä eri tutkintojen välillä. Tapaturmatutkinta-aineistojen analysointi osoittaa, että työpaikoilla tarvitaan lisää osaamista ja työkaluja ihmisen toiminnan taustalla vaikuttavien tekijöiden parempaan tunnistamiseen ja ymmärtämiseen, kuten myös aiemmat turvallisuuskriittisten toimialojen tutkimukset (esim. Teperi ym., 2018b; Teperi ym., 2017a) ovat osoittaneet. Yksilön ja työn piirteet ylikorostuvat tapaturmatutkintojen syytekijöinä, kun taas ryhmätason tekijöitä ei tunnisteta nykyisillä tutkintamenetelmillä ja -työkaluilla. Useat HF-työkaluun sisältyvät

inhimilliset tekijät eivät nouse ollenkaan esiin analysoiduissa tapaturmaraporteissa (esim. vireystila, työstä saatu palaute tai konsernipalveluiden tuki liiketoiminnoille).

Organisaatioiden nykyiset tutkintakäytännöt eivät myöskään nosta systemaattisesti esiin onnistumisten ja hyvin hoidettujen asioiden kirjaamista tapahtumaketjun eri vaiheissa, vaikka modernin turvallisuusajattelun (Hollnagel, 2018, 2014; Hollnagel ym., 2013) mukaan myös onnistumisista olisi tärkeää oppia. Onnistumisten kirjaaminen saattaisi joissakin tilanteissa selventää ihmisen toiminnan taustoja ja annettujen tavoitteiden välisiä ristiriitoja (esim. saatu tuote lähtemään tavoiteajassa asiakkaalle, mutta siihen on päästy turvallisuustoimia ohittamalla), ja olla signaali laajemmista ongelmista organisaation turvallisuus- ja arvoviestinnässä (Ratilainen ym., 2018). Onnistumisten huomioiminen rakentaa aidosti reilua ja oikeudenmukaista kulttuuria tutkintaan, auttaa tutkinnan vastuuhenkilöitä ymmärtämään paremmin poikkeaman osallisten toimintaa ja kehittää tutkinnan laatua (Dekker, 2007; Teperi ym., 2017a).

Turvallisuuden jatkuvan kehittämisen vuoksi on tärkeää, että tapaturmatutkintatilanne opitaan hyödyntämään yksittäistä tapaturmaa laajempänä oppimisikkunana organisaation toimintaan. Tutkimusaineisossa tällaista laajempaa oppimista on havaittavissa joissakin yksittäisissä tapaturmaraporteissa.

Kuten kokonaisarviot, myös tapaturmatutkinnan kirjausten laatu ja syvyys sekä tapahtumakuvausten reiluus ja oikeudenmukaisuus vaihtelee huomattavasti samankin organisaation eri tapaturmien osalta. Se kuvanee sitä, että osassa tapauksista ihmisen toiminnan taustalla vaikuttavat tekijät on helpompi tunnistaa, jolloin raportista saadaan laajemmin eri tekijöitä huomioiva ja syyllistämätön – toisaalta näissä tapauksissa on monesti kyse työympäristöön tai -välineisiin liittyvistä tekijöistä, eikä tapahtuma- tai päätöksentekoketjua välttämättä enää selvitetä eteenpäin (esim. hankintavaiheessa valittu kone, jonka huollettavuus on vaikeaa tai kunnossapidon resurssien vähentämisen takia huoltoväli on tullut liian pitkäksi). Nykyiset tutkintakäytännöt ja -menettelyt eivät ohjaa systemaattisesti huomioimaan ja kirjaamaan ylös esimerkiksi ryhmätason tekijöitä tai haasta selvittämään ihmisen toiminnan taustalla vaikuttavia tekijöitä, vaan raporteissa tyydytään helposti selitykseen huolimattomuudesta ja varomattomuudesta.

5 Johtopäätökset ja suositukset

Viime vuosien tekoälynoste ja merkittävät koneälyavusteiset teknologiset saavutukset ovat saaneet niin yritykset kuin tutkijatkin kiinnostumaan siitä, millaisia ratkaisuja tekoäly voisi tarjota liiketoiminnan ja työelämän haasteisiin. Tässä tutkimushankkeessa tarkasteltiin tekoälyn, tarkemmin ottaen koneoppimismallien ja syvien neuroverkkojen käyttömahdollisuuksia organisaatiodatan analysoinnissa ja siten työturvallisuuden kehittämisessä. Tekoälyinnostuksen pyörteissä saattaa kuitenkin hämärtyä, mitä tekoäly oikeastaan tekee ja voi tehdä – ja mitä ei. Siksi tässä tutkimuksessa haluttiin selvittää, *miten ja millä edellytyksillä* koneoppimismenetelmiä voidaan hyödyntää turvallisuusjohtamisen työkaluna sekä turvallisuustutkimuksessa, kun käytävissä on sitä organisaatiodataa, joita organisaatiot tyypillisesti kerryttävät turvallisuudesta ja muusta toiminnastaan. Lisäksi tutkittiin koneoppimismalleja ja asiantuntija-analyysejä hyödyntämällä, *minkälaisia näkökulmia* tekoälyä hyödyntävät analyysimenetelmät ja eri tietoaaineistojen yhdistäminen tuovat työturvallisuuden tilannekuvaan ja työturvallisuuden kehittämistarpeisiin. Tässä luvussa on esitetty tutkimuksen keskeiset johtopäätökset ja toimenpidesuositukset sekä esitetty tarpeita jatkotutkimukselle.

5.1 Johtopäätökset

5.1.1 Organisaatioiden keräämän datan hyödynnettävyys turvallisuuden johtamisessa

Koneoppiminen ja tekoäly soveltuvat lähtökohtaisesti suurille datoilta hyvin, sillä ne pystyvät helposti käsittelemään suuria määriä rakenteellista dataa. Myös vapaan tekstin analysointi on koneoppimismenetelmille hyvin otollinen sovelluskohde. Datan pohjalta rakennettu tekoälymalli soveltuu hyvin juuri kyseisen datan analysointiin, mutta ei tyypillisesti pysty ennustamaan tai kuvaamaan uusia, alkuperäisen datan ulkopuolelle jääviä ja siitä poikkeavia ilmiöitä. Ihmisen tekemään analyysiin verrattuna koneoppimismalli on myös objektiivinen; se perustuu täysin samaansa dataan, eikä ole altis ennakkoluuloille tai subjektiivisille näkemyksille, jotka saattavat vääristää ihmisen tekemiä johtopäätöksiä. Toisaalta, mikäli koulutusdata on hyvin painottunutta, tai kuvaa ilmiötä tai asiaa vääristyneesti, malli oppii nämä vinoumat. Mallin suorituskyvyn arviointi alkuperäisen koulutusdatan ulkopuolisella datalla onkin tärkeää. Kymmenien tuhansien tekstikuvausten läpikäynti pelkästään asiantuntijavoimin on hidasta, ja monotonista. Koneoppimismallin analyysin luotettavuus ei kärsi työn toisteisuudesta tai määrästä, ja siksi uudet menetelmät vakiintunevat maailmassa, jossa digitaalisen datan määrä kasvaa vuosittain merkittävästi.

Kun tekoälyn käyttömahdollisuuksia tutkittiin turvallisuusjohtamisen näkökulmasta, keskeiseksi haasteeksi nousi organisaatioiden keräämän datan kattavuus ja laatu. Koneoppimismenetelmiä sovellettaessa mahdollisimman suuri määrä dataa on ihanteellinen lähtökohta, mutta määrä ei automaattisesti korvaa laatua, vaan analysoinnin pohjaksi tarvitaan riittävä määrä riittävän laadukasta dataa (Ramasamy ja Chowdhury, 2020): sen täytyy olla systemaattista, kattavaa, tarkkaa ja luotettavaa, ja sen täytyy antaa todenmukainen ja ajantasainen kuva. Tässäkin hankkeessa havaittiin, että organisaatioiden työturvallisuusdatassa on puutteita tai tietoaukkoja, jotka hankaloittavat datan hyödyntämistä tekoälyyn perustuvissa analyyseissä. Esimerkiksi organisaatiokohtaiset poikkeamatutkinnat saattavat löytyä vain skannattuina dokumentteina ja koneluettavassa datassa sanalliset kuvaukset ovat niin tiiviitä, ettei niiden perusteella välttämättä saa käsitystä tapahtumasta tai sen taustalla vaikuttaneista tekijöistä. Taustalla voi vaikuttaa se, että koneluettavaan muotoon saatetaan kirjata vain niitä asioita, joita aiotaan kommunikoida vakuutusyhtiölle. Tekstipohjaiselle analysoinnille lisähaasteensa tuo esimerkiksi se, että kirjauksissa käytetty kieli ja termit ovat epäyhtenäisiä, ja kirjauksia saatetaan tehdä kiireessä tai huonoilla käyttöliittymillä, jolloin tekstidata sisältää myös paljon kirjoitusvirheitä.

Eri tietojärjestelmien epäyhtenäisyys ja ajalliseen yhdistämiseen tarvittavien aikaleimojen puuttuminen vaikeuttavat eri datalähteiden yhdistämistä ja yhteistarkastelua. Turvallisuus ei kuitenkaan synny tyhjiössä vaan aina kiinteästi linkittyneenä organisaation muuhun toimintaan ja tapahtumiin. Siksi turvallisuusdataa ei voi tarkastella irrallaan muusta organisaatiodatasta. Kuitenkin nykytilanne useissa yrityksissä, kuten tähän tutkimukseen osallistuneissa organisaatioissakin on se, että turvallisuusdataa kerätään erillisiin järjestelmiin, järjestelmät eivät kommunikoi keskenään, tai koko raakadataa ei edes saada ulos järjestelmästä, ja muun organisaatiodatan yhdistäminen turvallisuusdataan on vähintäänkin haastavaa ja monesti mahdotonta. Oman haasteensa tuo se, että ulkoisten järjestelmätoimittajat eivät yleensä tarjoa käyttäjäorganisaatiolle pääsyä raakadataan, jota tarvitaan datalähtöisessä ja koneoppimismalleihin perustuvassa data-analytiikassa. Jos raakadataa olisikin mahdollista saada ulos koneluettavassa muodossa käyttäjäorganisaatiolle, järjestelmän peruseraportteista poikkeavista datapyynnöistä laskutetaan käyttäjäorganisaatiota yleensä erikseen. Tämä voi heikentää organisaatioiden halukkuutta hankkia analyysikäyttöön riittävän monipuolisesti datalähteitä, mikä puolestaan vaikuttaa data-analyysien laatuun ja edustavuuteen. Sen seurauksena turvallisuuteen liittyvät tapahtumat ja muutokset jäävät irrallisiksi, eikä turvallisuustapahtumia ja niiden taustalla vaikuttaneiden tekijöiden verkostoa pystytä näkemään ja tulkitsemaan osana laajempaa kokonaisuutta.

5.1.2 Työturvallisuuden haasteita ja kehittämistarpeita

Hankkeen analyyseissä havaittiin, että organisaation turvallisuustapahtumien määrää ja vakavuutta voidaan yleisellä tasolla kytkeä organisaation muihin rakenteellisiin muuttujiin. Mitattuun työturvallisuuteen vaikuttaa kuitenkin tiedossa olevien muuttujien lisäksi paljon tekijöitä, joita ei joko tässä tutkimuksessa ollut saatavilla tai ei edes välttämättä pystytä suoraan numeerisesti arvioimaan. Lisäksi riippuvuussuhteet voivat vaihdella vuosien kuluessa esimerkiksi organisaatorakenteen tai ympäröivien olosuhteiden muuttuessa. Siksi mallinnuksissa työturvallisuutta ei pystytty kattavasti ennustamaan tiedossa olevilla muuttujilla. Analyysien tulokset antavat kuitenkin viitteitä siitä, että esimerkiksi turvallisuuskoulutuksilla ja perehdytyksillä pystytään vaikuttamaan työturvallisuuteen. Vastaavasti korkeat tuotanto- ja tilausmäärät näyttävät yleisellä tasolla olevan yhteydessä työturvallisuustason heikentymiseen. Kohonneet turvallisuushavaintomäärät indikoivat kasvua työturvallisuustapahtumien määrässä ja vakavuudessa, mikä antaa viitteitä siitä, että vakavampia turvallisuustapahtumia kuten työtapaturmia voitaisiin estää, mikäli työturvallisuuden korjaavia toimenpiteitä pystytään tekemään aktiivisesti ja ennakoiden turvallisuushavainnoilla saatavan tiedon perusteella.

Työturvallisuustapahtumien datapohjaisen ryhmittelyn tulosten havaittiin olevan voimakkaasti riippuvaisia käytettävien muuttujien valikoimasta, laadusta ja määrästä. Parhaassa tapauksessa luokittelussa päästiin erottelemaan eri toimipisteiden ja vakavuusasteiden tapahtumia sekä havaitsemaan yksittäisiä työntekijäryhmiä ja ajanjaksoja, joille tiettyjä turvallisuustapahtumia vaikuttaa kerääntyvän muita enemmän.

Faktoihin perustuvan turvallisuuden johtamisen näkökulmasta tapaturma- ja poikkeamatutkinnoissa sekä tutkinnan tulosten dokumentoinnissa on vielä kehitettävää. Organisaatioiden tapaturmatutkintojen laatu sekä tapahtuman ja ihmisen toiminnan taustalla vaikuttavien tekijöiden tunnistaminen vaihteli paljon sekä organisaatioiden välillä että organisaation sisällä. Jos organisaatioissa ei ole käytössä menetelmää, joka auttaa tunnistamaan tapaturmien ja poikkeamien taustalla vaikuttavia tekijöitä laajemmin sekä ajallisesti että eri tasojen tekijöitä tunnistuen, kerättävä tieto voi olla vaillinaista ja johtaa huonosti vaikuttaviin toimenpiteisiin. Työpaikoilla tarvitaan edelleen lisää osaamista ja työkaluja ihmisen toiminnan taustalla vaikuttavien tekijöiden parempaan tunnistamiseen ja ymmärtämiseen sekä niiden kirjaamiseen tutkintaraportteihin. Yksilön ja työn piirteet ylikorostuvat tapahtumien taustalla vaikuttaneiden tekijöiden tunnistamisessa, ja varsinkin ryhmätason tekijöitä ei tunnisteta nykyisillä tutkintamenetelmillä ja -työkaluilla. Vastaava havainto tehtiin Turvaa tiimistä! – Työryhmät turvallisuuden tekijöinä -tutkimuksessa (Kannisto ym., 2022), jossa todettiin, että ryhmätason tekijöitä ei tunnisteta tapaturmatutkinnoissa eikä

niistä tehdä juurikaan turvallisuushavaintoja. Poikkeamiin liittyneitä onnistumisia ja hyvin hoidettuja asioita ei juurikaan tunnisteta, vaikka myös onnistumisista tulisi oppia ja onnistumisten tunnistamisella ja tunnustamisella on merkitystä myös reilun ja oikeudenmukaisen kulttuurin kehittämisessä.

5.1.3 Turvallisuustietojen hyödyntäminen päätöksentekotilanteissa

Työturvallisuuteen liittyvien tietojen hyödyntäminen erilaisissa päätöksentekotilanteissa organisaatioissa on hyvä suunnitella osaksi kokonaisvaltaista tuotannon tietojen hyödyntämistä. Ohessa joitakin esimerkkejä tietojen keräämisestä ja hyödyntämisestä:

- Riskinarviointi: Tietojen kerääminen ja analysointi työtapaturmista, sairauspoissaoloista ja muista työturvallisuuteen liittyvistä tapahtumista auttaa organisaatiota tunnistamaan riskialttiit alueet tuotannossa. Näiden tietojen perusteella voidaan suunnitella välittömiä toimia riskien minimoimiseksi.
- Seuranta ja mittarit: työturvallisuuteen liittyviä tietoja kerätään myös työympäristöä havainnoimalla käyttäen apuna validoituja mittareita. Havainnoinneista kertynyttä aineistoa voidaan yhdistää esimerkiksi tapaturmien tai sairauspoissaolopäivien määrään suhteutettuna tehtyihin työtunteihin. Näin voidaan seurata työturvallisuuteen tehtyjen investointien vaikuttavuutta tietyllä ajanjaksolla. Tämä tieto on tärkeä johdolle investointipäätöksiä tehtäessä.
- Tietoja tulee kerätä myös organisaation ja ryhmätason tekijöistä sekä yksilön toiminnan ja työtoiminnan piirteistä. Näiden tietojen perusteella saadaan kokonaisvaltaisempi kuva riskien ja niistä mahdollisesti aiheutuvien tapaturmien syy-yhteyksistä.
- Koulutus ja tietoisuus: erilaisten organisaation toimintaan liittyvien tietojen yhdistäminen ja analyysitietojen tulkinta työturvallisuuden kannalta lisää henkilöstön tietoisuutta tapahtumien taustalla olevista tekijöistä. Tapahtumien ennakointi ja keinot, joilla niihin voidaan vaikuttaa, tulee suunnitella yhdessä johdon ja muun henkilöstön kesken.
- Resurssien kohdentaminen: Tiedot voivat auttaa organisaatiota määrittelemään tarvittavat resurssit työturvallisuustoimenpiteisiin. Tämä voi sisältää investoinnit henkilöstön koulutukseen, laitteisiin ja muihin turvallisuutta parantaviin toimenpiteisiin. Nämä investoinnit henkilöstön terveyteen ja turvallisuuteen voivat lisätä tuottavuutta ja säästää kustannuksia pitkällä aikavälillä.
- Lainsäädännön noudattaminen: tietojen suunnitelmallinen hyödyntäminen auttaa organisaatioita pysymään lain määräämissä vaatimuksissa ja välttämään näin mahdollisia oikeudellisia seuraamuksia.

- Sidosryhmäviestintä: organisaatiot voivat käyttää turvallisuuteen liittyviä tietoja myös viestiäkseen sidosryhmille, kuten asiakkaille ja sijoittajille, sitoutumisestaan turvalliseen työympäristöön ja turvallisuusjohtamiseen laajemmin.
- Jatkuva parantaminen: tietojen kerääminen ja analysointi niin, että niiden avulla pystytään ennakoimaan turvallisuustilanteen kehitystä organisaatiossa tietyllä ajanjaksolla, tarjoaa niille mahdollisuuden varautua ajoissa erilaisiin muutoksiin. Näitä muutosajureita syntyy organisaation ulkopuolella ja niillä on vaikutusta organisaation toiminnan muotoutumiseen ja jatkuvuuteen.

Työturvallisuuden kehityksen ennakoiminen organisaatiossa edellyttää systemaattista tiedonkeruuta erilaisista organisaation tietolähteistä ja tietojen huolellista analysointia sekä raportointia. Organisaation on myös oltava sitoutunut turvallisuuskulttuuriin, joka edellyttää ajantasaiseen tietoon perustuvaa jatkuvaa kehittämistä.

5.2 Toimenpidesuosituksia

1. Tekstinlouhinnan kehittäminen organisaatiokohtaisesti tukemaan juuri kyseisen organisaation turvallisuusjohtamisen tarpeita. Jos organisaatio tuottaa laadukasta tekstidataa, voi tekstinlouhinnalla käsitellä isoja tekstiaineistoja eri tavoin turvallisuusjohdon tiedonhallinnan tukena.
2. Laadukkaiden havaintokohtaisten aikamerkintöjen sisällyttäminen kaikkeen relevanttiin dataan, mahdollistaa syy-seuraus-suhteita tarkastelevan monimuotoista aineistoa hyödyntävän tiedonlouhinnan.
3. Organisaatiot tarvitsevat tapaturman tutkinnan kehittämistä inhimillisten tekijöiden näkökulmasta sekä parempaa osaamista ja ymmärrystä inhimillisistä tekijöistä, mutta myös tutkintamalleja ja -käytäntöjä, jotka tukevat tutkinnan tekemistä. Suosituksena esimerkiksi HF Tool™.
4. Reilun ja oikeudenmukaisen kulttuurin luominen poikkeamatutkintatilanteeseen ja viestintään tutkinnan opeista.
 - a. Esimerkiksi selvittämällä systemaattisesti tapahtumaketjun eri vaiheiden onnistumisia ja hyvin hoidettuja asioita
 - b. Huolehtimalla, että tutkinnassa päästään yksilöihin liittyviä välittömiä syytekijöitä pidemmälle ja löydetään huolellisuudesta ja varovaisuudesta muistuttamista vaikuttavampia toimenpiteitä vastaavien tapaturmien ennaltaehkäisemiseksi.

5. Poikkeamatutkinnat tulisi nähdä laajempina oppimisikkunoina organisaatioon ja sen toimintaan - mahdollisuutena keskustella ja kehittää turvallisuutta ja työn sujuvuutta.
6. Jotta koko organisaation oppiminen poikkeamista on mahdollista, tutkinnan tulokset ja korjaavat toimenpiteet tulee kirjata systemaattisesti turvallisuusjohtamisen tietojärjestelmään tai yhteisiin tietokantoihin.
7. Erityisesti kokoluokaltaan pienemmissä organisaatioissa, joissa tapaturmien määrä on usein vain vähäinen, on tärkeä opetella oppimaan ja kehittämään turvallisuutta tutkimalla poikkeama- ja vaaratilanteita ja turvallisuushavaintoja niiden seurausten vakavuudesta riippumatta. Tai jopa onnistumisia ja normaalitoimintaa.
8. Turvallisuushavaintojen tietojärjestelmien kehittäminen niin, että tutkinnan tulokset, määritellyt korjaavat toimenpiteet yms. ovat helposti tarkasteltavissa, koostettavissa ja hyödynnettävissä.

Kuvassa 16 on esitetty hankkeen tulosten ja havaintojen pohjalta keskeisimmät askeleet datalähtöisen turvallisuusjohtamisen kehittämiseen kolmesta olennaisesta näkökulmasta. Kuvan avulla voi myös arvioida, miten omassa organisaatiossa toteutuvat datalähtöisen turvallisuusjohtamisen keskeiset edellytykset.



Kuva 16: Keskeiset askelet datalähtöisen turvallisuusjohtamisen kehittämiseen.

5.3 Jatkotutkimusehdotukset

AI Safety'n aineisto oli monimuotoinen kokoelma erilaista dataa. Tämä tarkoittaa tietenkin sitä, että vastaavissa tutkimuksissa aineisto voisi olla myös hyvin erilainen. Jos aineisto muodostuisi työturvallisuusdatan lisäksi esimerkiksi sensoridatasta tai videokuvasta, tutkimuskysymyksetkin voisivat olla hyvin erilaisia. Niinpä monimuotoista dataa analysoiva metodologialtaan koneoppimispainotteinen tutkimus on vielä pitkään relevantti tutkimusaihe, sillä pelkästään aineiston osalta on monia mahdollisuuksia, joita ei ole vielä juurikaan tutkittu. Oma lukunsa ovat itse metodit, jotka kehittyvät jatkuvasti.

Tässä hankkeessa kehitettyä HF-tekstianalysimallia voi kehittää pidemmälle, niin että se tunnistaisi tarkemmin HF-toolin 37 alakohtaa. Toki neuroverkkoa voi opettaa tunnistamaan vastaavalla tavalla muitakin – erilaisiin teoreettisiin viitekehyksiin perustuvia – rakenteita samankaltaisesta tekstiaineistosta, kuin mitä tässä hankkeessa käytettiin.

Lähteet

- Ajayi, A., Lukumon, O., Davila Delgado, J. (2019). Big data platform for health and safety accident prediction, *World Journal of Science, Technology and Sustainable Development*, Vol. 16 Issue: 1, pp.2-21.
- Bertke, S. J., Meyers, A. R., Wurzelbacher, S. J., Bell, J., Lampl, M. L., & Robins, D. (2012). Development and evaluation of a Naïve Bayesian model for coding causation of workers' compensation claims. *Journal of Safety Research*, 43(5-6), 327-332.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M.; Ng, A. Y., Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of machine Learning research* 3. Jan: 993-1022.
- Chen, T. & He, T. (2014). Higgs boson discovery with boosted trees. HEPML'14: Proceedings of the 2014 international conference on high-energy physics and machine learning, 42, 69-80.
- Chen, T. & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. KDD '16: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 785-794
- Ciarapica, F. E., & Giacchetta, G. (2009). Classification and prediction of occupational injury risk using soft computing techniques: An Italian study. *Safety Science*, 47(1), 36-49.
- Dekker, S. W. A. (2002). Reconstructing the human contribution to accidents: The new view of human error and performance. *Journal of Safety Research*. Vol. 33:3. S. 371-385.
- Dekker, S. (2011). *Drift into failure: From hunting broken components to understanding complex systems*. Farnham: Ashgate Publishing Co.
- Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. (2018). "BERT: Pre-training of deep bidirectional transformers for language understanding.
- Ester, M., Kriegel H.-P., Sander, J. & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the second international conference on knowledge discovery and data mining (KDD-96). AAAI Press. pp. 226-231.

- Ford, M. T., & Tetrick, L. E. (2008). Safety motivation and human resource management in North America. *The International Journal of Human Resource Management*. Vol. 19:8. S. 1472–1485.
- GDPR. EU:n yleinen tietosuoja-asetus. Saatavilla: <https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=CELEX:32016R0679>
- Goel, P., Datta, A., Mannan, S. (2017). Application of big data analytics in process safety and risk management. 2017 IEEE. International conference on big data.
- Hale, A. R. & Hovden, J. (1998). Management and culture: the third age of safety. A review of approaches to organizational aspects of safety, health and environment. In: Feyer, A., Williamson, A. (Eds.), *Occupational Injury: Risk, Prevention and Intervention*. London: Taylor & Francis. S. 129–265.
- Hollnagel, E., Woods, D., & Leveson, N. (2006). *Resilience engineering: Concepts and precepts*. Hampshire, UK: Ashgate Publishing Ltd.
- Hollnagel, E. (2009). ETTO-principle, efficiency-thoroughness trade-off. Why things that go right sometimes go wrong. Ashgate Publishing Ltd.
- Hollnagel, E. (2014). Is safety a subject for science? *Safety Science*. Vol. 67. S. 21–24.
- Hollnagel, E., Leonhardt, J., Licu, T. & Shorrock, S. (2013). *From Safety-I to Safety-II: A White Paper*. Brussels, Belgium: EUROCONTROL. [Viitattu 19.5.2020]. Saatavilla: <http://skybrary.aero/bookshelf/books/2437.pdf>
- Hollnagel, E. (2018). *Safety-II in practice: Developing the resilience potentials*. Abingdon: Taylor & Francis.
- Huang, L., Wu, C., Wang, B. (2019). Challenges, opportunities and paradigm of applying big data to production safety management: From a theoretical perspective. *Journal of Cleaner Production* 231 (2019) 592e599.
- Kannisto, H., Asikainen, I., Puro, V., Ansio, H., Koskela, I., Valtonen, T., Paajanen, T., & Toivio, P. (2022). Turvaa tiimistä! Työryhmät turvallisuuden tekijöinä.
- Kirwan, B. (2003). An overview of a nuclear processing plant human factors programme. *Applied Ergonomics*. Vol 34:5. S. 441–452.
- Krishna, O. B., Maiti, J., Ray, P. K., & Mandal, S. (2015). Assessment of risk of musculoskeletal disorders among crane operators in a steel plant: a data mining-based analysis. *Human Factors and Ergonomics in Manufacturing*, 25(5), 559-572.

- Lavezo dos Reis, B., Francisco da Rosa, A., de Araujo Machado, A. (2021). Data mining in occupational safety and health: a systematic mapping and roadmap. *Production*, 31, e20210048.
- Leveson N. (2015). A systems approach to risk management through leading safety indicators. *Reliability Engineering & System Safety*. Vol. 136, 17-34.
- Leveson, N. (2020). *Safety III: A systems approach to safety and resilience*.
- Liu J.; Luo H.; Liu H. (2022) Deep learning-based data analytics for safety in construction. *Automation in Construction* 140
- Mohammadpoor, M., Torabi, F. (2018). Big data analytics in oil and gas industry: An emerging trend. *Petroleum*, December 2018.
- Nenonen, N. (2013). Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and diseases statistics database. *Applied Ergonomics*, 44(2), 215-224.
- Poh, C., Ubeinarayana, C., Goh, Y. (2018). Safety leading indicators for construction sites: A machine learning approach. *Automation in Construction*, Volume 93, 1-402.
- Puro, V., Kannisto, H., Lantto, E. (2021). Learning from operational events on organizational level—findings from Finnish nuclear power industry. In *Human factors in the nuclear industry* (pp. 163-190). Woodhead Publishing.
- Ramasamy, A., & Chowdhury, S. (2020). Big data quality dimensions: A systematic literature review. *Journal of Information Systems and Technology Management*, 17, e202017003.
- Rathor K., Lenka S., Pandya K. A. Gokulakrishna, B. S., Ananthan S. S. and Khan Z. T. (2022). A Detailed View on industrial Safety and Health Analytics using Machine Learning Hybrid Ensemble Techniques. 2022 International conference on edge computing and applications (ICECAA), Tamilnadu, India, 2022, pp. 1166-1169.
- Ratilainen, H., Salminen, S., Kalakoski, V., Zwetsloot, G., Perttula, P., Starren, A., Ukkonen, A., Steijn, W., Puro, V., & Pahkin, K. (2018). *Value of Safety (VALOSA)*. Finnish Institute of Occupational Health.
- Reason, J. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate.

- Reason, J. (2008). The human contribution: unsafe acts, accidents and heroic recoveries. Cornwall, UK: Ashgate.
- Řehůřek, R. & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks. Valletta, Malta. p. 46-50.
- Reiman, T. & Oedewald, P. (2009). Evaluating safety critical organizations. Focus on the nuclear industry. Swedish radiation safety authority, research report 2009:12.
- Ruso, J., & Stojanović, V. (2012). Occupational health and safety using data mining. International Journal of Qualitative Research, 6(4), 168-194.
- Sarkar, S., Lodhi, V., & Maiti, J. (2019). Text-clustering based deep neural network for prediction of occupational accident risk: a case study. In 2018 International Joint Symposium on artificial intelligence and natural language processing (iSAI-NLP) (pp. 1-6). New York.
- Shapley, L. S. (1953). A value for n-person games. Contributions to the theory of games 2.28: 307-317.
- Shein, M. M., Hamilton-Wright, A., Black, N., Samson, M., & Lecanelier, M. (2015). Assessing ergonomic and postural data for pain and fatigue markers using machine learning techniques. In International conference and workshop on computing and communication (IEMCON) (pp. 1-6). New York.
- Siddula, M., Dai, F., Ye, Y., & Fan, J. (2016). Classifying construction site photos for roof detection. Construction Innovation, 16(3), 368-389.
- Teperi, AM. (2012). Improving the mastery of human factors in a safety critical ATM organization. Cognitive Science, Institute of Behavioural Sciences, Faculty of Behavioural Sciences, University of Helsinki, Finland. Doctoral dissertation.
- Teperi, AM., Asikainen I., Ala-Laurinaho T., Valtonen, T. & Paajanen, T. (2018a). Modeling safety criticality in aviation maintenance operations to support mastery of human factors. International Conference on Applied Human Factors and Ergonomics 791. S. 331-341.
- Teperi, AM., Leppänen, A & Norros, L. (2015). Application of new human factors tool in an air traffic management organization. Safety Science. Vol. 73. S. 23-33.
- Teperi, AM., Puro V. & Lappalainen J. (2017a). Promoting positive safety culture in the maritime industry by applying the Safety-II perspective. In: Bernatik, A., Kocurkova,

- L., Jørgensen, K., eds. Prevention of accidents at work: proceedings of the 9th International Conference on the prevention of accidents at work 2017 Prague. CRC Press, 2017.
- Teperi, AM., Puro V., Tiikkaja M. & Ratilainen H. (2018b). Developing and implementing a human factors (HF) tool to improve safety management in the nuclear industry. Helsinki: Finnish Institute of Occupational Health, 2018.
- Teperi, AM., Puro, V & Ratilainen, H. (2017b). Applying a new human factor tool in the nuclear energy industry. Safety Science. Vol. 95. S. 125–139.
- Teperi, A-M. (2023). Ihminen turvallisuuden tekijänä. Gaudeamus, Tallinna. 413 s.
- Tietosuojalaki. Saatavilla: <https://www.finlex.fi/fi/laki/alkup/2018/20181050>
- Tiikkaja, M., Puro, V., Heikkilä, T., Kannisto, H., Lantto, E., Lukander, K., Nykänen, M., Räsänen, T., Simpura, F. & Uusitalo, J. (2020). Modernia turvallisuusoppimista rakennusalalle (MoSaC). Tutkimushankkeen loppuraportti. Työterveyslaitos, Helsinki. PunaMusta Oy, Tampere. 58 s. (tämä linkki jää siis pois [MoSaC-tutkimuksen loppuraportti \(julkari.fi\)](#))
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N; Kaiser L.; Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30, 5998-6008.
- Virtanen, A.; Kanerva, J.; Ilo, R.; Luoma, J.; Luotolahti, J.; Salakoski, T.; Ginter F.; Pyysalo, S. (2019). Multilingual is not enough: BERT for Finnish. *arXiv preprint arXiv:1912.07076*.
- Visuri, S., Ansio, H., Puro, V., Kannisto, H., Heiniö, M., Lantto, E., Hirvonen, M., Perttula P. & Teperi, A-M. (2020). Floor is Yours! Turvallisuusjohtamisen ja -kulttuurin kehittäminen esittävässä taiteissa. Työterveyslaitos, Helsinki. PunaMusta Oy, Tampere. 91 s. (tämä linkki jää siis pois [Floor is Yours! Turvallisuusjohtamisen ja -kulttuurin kehittäminen esittävässä taiteissa \(julkari.fi\)](#))
- Wang, B., Wu, C., Huang, L., Kang, L. (2018). Using data-driven safety decision-making to realize smart safety management in the era of big data: A theoretical perspective on basic questions and their answers. Journal of Cleaner Production 210 (2019) 1595-1604.
- Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on SA theory and measurement. Human Factors. Vol 50:3. S. 397–403.

- Wilson, JR. (2014). Fundamentals of systems ergonomics/human factors. *Applied Ergonomics*. Vol 45:1. S. 5–13.
- Xie, X., & Chang, Z. (2018). Intelligent wearable occupational health safety assurance system of power operation. *Journal of Medical Systems*, 43(1), 16.
- Xu, Z.; Saleh, H. J. (2021). Machine learning for reliability engineering and safety applications: Review of current status and future opportunities. *Reliability Engineering & System Safety*, Vol 211.
- Zwetsloot, G.; Leka, S.; Kines, P. (2007). Vision zero: From accident prevention to the promotion of health, safety and well-being at work. *Policy Pract. Health Saf.*, 15, 88–100.



Työsuojelurahasto
Arbetskyddsfonden
The Finnish Work Environment Fund

Työterveyslaitos
Arbetshälsoinstitutet
Finnish Institute of Occupational Health

PL 40, 00032 Työterveyslaitos

www.ttl.fi

I
ISBN 978-952-391-135-2 (PDF)

