

SUURANDMETE JA MASINÖPPE TULEVANE ROLL TÖÖOHUTUSE JA TÖÖTERVISHOIU JÄRELEVALVE TÕHUSTAMISEL

Sissejuhatus

Töenäoliselt kõige olulisem poliitikavahend, mida riiklikud tööinspeksioonid kasutavad, tagamaks, et ettevõtted rakendavad vajalikke meetmeid töötervishoiu ja tööohutuse eeskirjade järgimiseks, on kontrollid. Ent kontrollide mõju sõltub mitmest eri tegurist. Üks põhitegur on kontrollitavate ettevõtete või töökohtade valimise protsess. Põhimõtteliselt on valimiseks vähemalt kolm eri lähenemisviisi. Esimese lähenemisviisi korral inspekteeritakse kõiki ettevõtteid sõltumata võimalikust riskist, ettevõtte suurusest, tegevusvaldkonnast või muudest kriteeriumidest. Teise lähenemisviisi korral valitakse ettevõtted juhusliku valimi põhjal ning igal ettevõttel on sõltumata omadustest võrdne võimalus osutada valituks. Mis puutub ennetavat ja majanduslikku mõju, peetakse mõlemat meetodit tavaliselt ebatõhusaks (Blanc, 2013). Seega valib enamik tööinspeksioone kontrollitavaid ettevõtteid ja töökohad kolmanda ehk riskipõhise lähenemisviisi alusel. Lühidalt öeldes hõlmab riskipõhine lähenemisviis inspekteerimisobjektide valimist riskitaseme alusel.

Ehkki riskipõhine lähenemisviis on enamiku nüüdisaegsete tööinspeksioonide jaoks oluline põhimõte, esineb selle praktikas rakendamisel olulisi probleeme. Seda peamiselt seepärast, et puuduvad piisavalt täpsed riskianalüüsi meetodid (Mischke *et al.*, 2013). Riskipõhist prioriseerimist võimaldavate sobivate meetodite puudumisel tekib risk, et riskipõhine lähenemisviis muutub valitsuse poliitikaavalduseks ilma tegelike praktiliste tulemusteta. Seepärast on vaja välja töötada meetodid, mis võimaldavad välja valida suure riskiga ettevõtteid (Weil, 2008).

Enamik tööinspeksioone kogub ja säilitab suures koguses andmeid kontrollitavate ettevõtete ja järelevalvetevõtmise kohta. Seega on inspeksioonide valduses potentsiaalselt suured ja kiiresti kasvavad andmemahud, millele viidatakse tänapäeval terminiga „suurandmed“. Suurandmeid koos masinõppe tehnoloogiaga kasutatakse üha enam prognoosimisel, uurides andmete varjatud arengusuundumusi. Suurandmete ja masinõppe meetodite väärtust prognooside tegemisel testitakse väga erinevates valdkondades, nagu vähktõve prognoos ja patsientide ravitulemused, pankroti prognoosimine, naftahinna prognoosimine, maksupettuste avastamine, kuritegevuse prognoosimine ja aktsiaturu prognoosid. Selles dokumendis käsitletav põhiküsimus on, kas suurandmete ja masinõppe tehnoloogia kasutamine suure riskiga inspekteerimisobjektide valikuks võiks olla tööinspeksioonide jaoks paljutõotav võimalus.

Riskipõhine valimine

Majanduskoostöö ja Arengu Organisatsiooni kavandatud õiguspoliitika parimate tavade põhimõtete kohaselt (OECD, 2014) peaks tööinspeksioonide inspekteerimisobjektide valik põhinema riskianalüüsil ja -hindamisel. See tähendab, et ettevõtted tuleks kontrollimiseks valida, hinnates selliste riskitegurite nagu õnnetused, kahjulik kokkupuude ja ebaseaduslikud töötingimused tõenäosust ja tagajärgi. Riskipõhise valimise alus on teadvustamine, et piiratud inspekteerimisvahendite tõttu ei ole võimalik kontrollida kõiki riskivaldkondi ja -objekte. Tööinspeksioonide tööohutuse ja töötervishoiu järelevalve kontekstis tähendab see, et mõnda probleemset valdkonda tuleb võrreldes teistega olulisemaks seada. Lisaks tuleb mõne ettevõtte inspekteerimist prioriseerida teistega võrreldes.

Riskipõhise valiku põhimõtte ei ole uus. Ligi 50 aastat tagasi hinnati Robensi komitee poolt läbi viidud Ühendkuningriigi tööhutuse ja -tervishoiu järelevalvesüsteemi analüüsile tuginedes riskipõhist lähenemisviisi (koos eneseregulatsiooniga) kui ideaalset põhimõtet järelevalve ajakohastamiseks (Robens, 1972). Vahendite kulutõhusa kasutamise tagamiseks soovitati Robensi aruandes järelevalveasutusel keskendada oma vahendid valikuliselt kõige tõsisematele probleemsetele valdkondadele ning prioriseerida ettevõtted ja probleemid, mis tuvastati kõigi tööhutuse- ja töötervishoiuandmete, näiteks õnnetuste statistika, tehniline teave ja tööinspeksioonide kohalikud teadmised, süstemaatilisel analüüsil.

Tööinspeksioonid rahvusvahelisel tasandil ja ELi liikmesriikides on Robensi aruandes esitatud soovitusel laialdaselt kasutusele võtnud. Riskipõhise lähenemisviisi levik tähendab, et enamik nüüdisaegseid tööinspeksioone pooldab ideed vähendada tähelepanu väikse riskiga objektidele ja suunata rohkem vahendeid suurima riskiga objektide kontrollile. Et see oleks võimalik, on vaja mingit tüüpi andmeanalüüsi. Suure riskiga valdkondade ja riskiga kokkupuutuvate töötajate rühmade tuvastamise analüüsimeetodid on hästi välja töötatud. Sellised riskipõhised analüüsid põhinevad tavaliselt riiklikul statistikal seoses näiteks kutsehaiguste, tööõnnetuste ja ohuteguritega kokkupuutega töökeskkonnas. Analüüsid on järelevalvekampaaniate, strateegiakavade ning riiklike ja isegi rahvusvaheliste prioriteetsete valdkondade paikapanemise aluseks.

Palju vähem kui laialdased riskipõhised meetodid on levinud meetodid, mis võimaldavad prioriteetsete ettevõtete väljavalimist valdkonnasiselt. Tööinspeksioonide seas on levinud riskiettevõtete valimisel tuginev inspektorite kohalikele teadmistele. Mõnede riikide tööinspeksioonid, näiteks Taanis ja Rootsis, on uurinud aditiivse skaala põhiste riskijärjestussüsteemide kasutatavust. Aditiivse skaala alusel saab iga ettevõtte riskihinde, mis põhineb ettevõtte mitmel omadusel (nt suurus, tegevusvaldkond ja registreeritud õnnetuste arv), mis liidetakse liithinde saamiseks, ning kõrgeima liithindega ettevõtete kontrollimine seatakse prioriteediks. Sellise aditiivse skaala kasutamise probleem on, et sel on suhteliselt väike prognoostiline valiidsus, st hinde alusel ei ole eriti võimalik eristada suure riskiga ettevõtteid väikse riskiga ettevõtetest.

Suurandmed ja masinõpe

Kõiki ettevõtteid hõlmav prioriseerimise protsess on võrreldav nõela leidmisega heinakuhjast. See heinakuhi koosneb potentsiaalselt sadadest tuhandetest võimalikest inspekteerimisobjektidest, kuid üksnes teatud arv neist objektidest on nõelad, st neil esineb lubamatu riskitase. Nõeltele leidmine heinakuhjast on suures osas see, mida suurandmed ja masinõppe teevad.

Masinõppe algoritmide peamine eesmärk on leida statistiline mudel, mida saab kasutada prognooside, klassifikatsioonide, hinnangute või sarnaste ülesannete jaoks. Näiteks vähktõve prognoosimise valdkonnas on teadlased juba rohkem kui kolm aastakümnet rakendanud masinõppe algoritme vähile vastuvõtlikkuse, vähi taastekke ja vähipatsientide elulemuse prognoosimiseks. Temaatilisel on vähktõve prognoosimine kaugel kontrollitavate ettevõtete riskipõhisest valimisest. Mõlemad on siiski prognoosimisprobleemide ehk piltlikult öeldes nõela heinakuhjast otsimise probleemide näited.

Kaks masinõppe algoritmide põhitüüpi on juhendatud õpe ja juhendamata õpe. Juhendatud õppe korral koosneb algoritm sõltuvast muutujast (nt riskitase), mida tuleb prognoosida sõltumatute muutujate kogumi põhjal. Täpsed prognoosid eeldavad loomulikult suurt korrelatsiooni sõltumatute muutujate ja sõltuva muutuja vahel. Juhendamata õppe korral ei ole prognoosimiseks sõltuvat muutujat, vaid algoritmi eesmärk on jaotada andmed rühmadesse (nt erinevad riskirühmad) sarnasuse alusel. Vastupidi aditiivsetele skaaladele, nagu need, mida kasutasid Rootsi ja Taani tööinspeksioonid, parandavad masinõppe algoritmid peamiselt katse-eksituse meetodil järkjärgult oma prognoose. See tähendab, et masin õpib varasemate õnnestumiste (õiged prognoosid) ja vigade (väärad prognoosid) põhjal ning proovib neid teadmisi rakendada, et muuta prognoosid saadud tagasiside põhjal täpsemaks.

Suurandmete ja masinõppe kasutamine kontrollitavate ettevõtete ja töökohtade valimisel

Juhendatud ja juhendamata õppe algoritmid vajavad piisavat kogust vaatluseid ja muutujaid, mida tavaliselt nimetatakse tunnusteks. Nagu eespool märgitud, kogub ja säilitab enamik tööinspeksioone väga palju andmeid kontrollitavate ettevõtete ja -järelvalvetegevuste kohta. Kättesaadavad andmed on tavaliselt seotud ettevõttele omaste tunnustega, näiteks töötajate arv, ettevõtte vanus, tegevusvaldkond, varasemate kontrollide arv, varasemate kontrollide tulemused ja teated õnnetuste kohta. Lisaks suureneb andmekogus uute inspekteerimistulemuste lisamisega iga päev. Põhimõtteliselt peaks seega suure riskiga ettevõtete valimine suurandmete abil vähemalt esmapilgul sobima hästi masinõppe algoritmidega. Sellegipoolest on seda vähe katsetatud. Mõned märkimisväärsed erandid on siiski olemas ning illustreerivad asjaolu, et suurandmed ja masinõpe võivad olla väga olulised tööinspeksioonide jaoks, et lahendada suure riskiga ettevõtete valimise probleemi.

Esimene näide on teadusuuring, milles vaadeldi masinõppe meetodikate sobivust tööõnnetuste, konkreetsemalt samal tasapinnal libastumise prognoosimiseks (Matías *et al.*, 2008). Vaatamata suhteliselt täpsetele prognoosidele on selle uuringu puudus, et algoritmides sisalduvad tunnused ei ole seda tüüpi andmed, mis on tööinspeksioonidel tavaliselt olemas (nt isikukaitsevahendite kasutamine ja ruumide koristamise rutiinid). Lisaks on samal tasapinnal libastumise juhtumid üksnes väike osa töökohta riskidest, mida tööinspeksioonid käsitlevad.

Teine näide on samuti teadusuuring (Hajakbari ja Minaei-Bidgoli, 2014). Selles uuringus töötati välja tööõnnetuste riski prognoosimise punktsüsteem. Uuringus jõuti järeldusele, et eri tüüpi tööõnnetuste riski on võimalik ettevõtte mõne üldtunnuse (ettevõtte põhitegevus, sooline jaotus, töötajate arv jne) põhjal suhteliselt täpselt prognoosida. Lisaks järeldati uuringus, et algoritmi saab kasutada nende töökohtade tuvastamiseks, kus on vaja teha perioodilisi tervise- ja ohutuskontrolle. Selles uuringus kasutati andmeid tööinspeksiooni andmebaasist. Uuringu puudus on siiski, et tööõnnetused on ainult üks paljudest töökohta riskidest, millega tööinspeksioonid tegelevad. Lisaks on vigastuste statistikale tuginemisel probleemiks, et tööõnnetused võivad sageli jääda registreerimata.

Kolmas näide on Norra tööinspeksiooni (NLIA) välja töötatud töövahend inspektorite abistamiseks ettevõtete valimisel seoses riskidega (Dahl *et al.*, 2018). Töövahend, mida nimetatakse riskirühma prognoosimise vahendiks (Risk Group Prediction Tool, RGPT), jaotab ettevõtted prognoositava riski alusel nelja rühma: väikseima riskiga, väikse riskiga, suure riskiga ja suurima riskiga ettevõtted. Mida suurema riskiga on asjaomase ettevõtte riskirühm, seda suurem on tõenäosus, et selle ettevõtte tulevase kontrolli käigus tuvastatakse tõsiseid kõrvalekaldeid tööhutuse- ja tervishoiueeskirjadest. Rühm, kuhu ettevõtte on määratud, on nähtav inspektoritele NLIA siseveebipõhise kasutajaliidese kaudu. Seega teavad inspektorid ettevõtete valimisel kontrollimiseks ettevõtete riskirühmi ja saavad teha riskiteabel põhinevaid valikuid.

RGPT töötati välja prognoosiva modelleerimise põhjal masinõppe algoritmi abil, kasutades nn binaarset logistilist regressioonianalüüsi. Regressioonimudeli põhjal jagatakse kõik Norra ettevõtted (ligikaudu 230 000) ühte neljast riskirühmast. Seda tehakse kahes etapis. Esimeses etapis prognoositakse regressioonimudeliga tõenäosust, et tulevane kontroll tuvastab tõsised kõrvalekalded tööhutuse- ja -tervishoiueeskirjadest. Teises etapis kasutab mudel prognoositud tõenäosusväärtust, et määrata ettevõtte riskirühma.

Algselt töötati see töövahend välja ligikaudu 35 000 NLIA tehtud tööhutuse ja tervishoiu kontrolli registrikande põhjal. Samas muutuvad vahendi tehtud prognoosid kontrollide arvu suurendes järk-järgult ja automaatselt täpsemaks. See tähendab, et algoritm kohandab ennast tagasiside põhjal (õiged või väärad prognoosid), mis ta saab, kui toimuvad uued kontrollid ja need registreeritakse NLIA andmebaasis.

RGPT kuulub juhendatud õppe algoritmide kategooriasse, kus tõsiseid kõrvalekaldeid tööhutus- ja -tervishoiueeskirjadest (sõltuv muutuja) prognoositakse ettevõtte omaduste (tunnused) kogumi alusel. RGPT kasutatavad tunnused on ettevõtte üldised omadused, näiteks ettevõtte suurus, tegevusvaldkond, varasemate kontrollide arv, varasemate kontrollide tulemused, ettevõtte vanus, geograafiline asukoht ja õnnetustest arv. Töövahendi prognooside paikapidavust kontrollitakse iga kuu ja senine kogemus (pärast umbes 18-kuulist testimist) näitab, et algoritm suudab väga täpselt leida suure riskiga ettevõtteid. See tähendab, et valepositiivseid ja valenegatiivseid tulemusi on väga vähe, st väikseima riskiga rühmas on vähe kontrole, mille tulemusena tuvastatakse tõsiseid kõrvalekaldeid, samas suurima riskiga rühmas tuvastatakse valdava enamiku kontrollide tulemusena tõsiseid kõrvalekaldeid. Väikse ja suure riskiga rühmad jäävad kahe äärmuse vahele.

NLIA välja töötatud töövahendi kasutamise tulemused näitavad, et inspekteerimisobjekte on võimalik leida suurandmete ja masinõppe abil. Sarnaseid masinõppe lähenemisviise on paljutöötavate tulemustega katsetanud veel vähemalt kaks Euroopa tööinspeksiooni: Rootsi töökeskkonna amet (Ridemar, 2018) ja Madalmaade inspeksioon SZW (Jacobusse ja Veenman, 2016). Norra töövahendit ei saa siiski tingimata teistele tööinspeksioonidele üle kanda, sest selle kasutamine sõltub andmete säilitamise viisist, andmete kvaliteedist, andmetele juurdepääsust ja andmebaasi struktuurist. Lisaks hõlmab töövahendi abil ettevõtete valimine riski algoritmide määratlemise ja rakendamise viisi aktsepteerimist. Nagu eespool kirjeldatud, põhineb töövahend riski määratlusel, mis tähendab, et mida suurema riskiga on asjaomase ettevõtte riskirühm, seda suurem on tõenäosus, et selle ettevõtte kontrollimisel tuvastatakse tõsiseid kõrvalekaldeid tööhutus- ja -tervishoiueeskirjadest. See tähendab, et see töövahend on seotud peamiselt nn juhtimis- ja kontrolliriskide, mitte olemuslike riskidega. Kui juhtimis- ja kontrolliriskid tulenevad ettevõtte suutlikkusest ja tahtest riske juhtida (nt vastavate eeskirjade järgimisega), tulenevad olemuslikud riskid ettevõtte tegevuse laadist (nt kõrgelt kukkumine, kokkupuude kemikaalidega ja luu- ja lihaskonna koormus).

Praktikas on juhtimis- ja kontrolliriskid ning olemuslikud riskid seotud. See ei tähenda siiski, et kahte tüüpi riskid on tingimata tugeva empiirilise korrelatsiooniga. Seega võib pimesi tuginemine töövahenditele, millega valitakse ettevõtte üht tüüpi riskide põhjal, põhjustada teist tüüpi riskide märkamata jätmist. Norra õiguskorra raames lahendatakse see probleem, keskendudes olemuslikele riskidele prioriteetsete valdkondade, riskidega kokkupuutuvate töötajate klastrite ja suure riskiga valdkondade määramisel, samas kui juhtimis- ja kontrolliriskid on kesksel kohal konkreetsete kontrollitavate ettevõtete valimisel.

Probleemid

Tõsiasi, et juhtimis- ja kontrolliriskid ning olemuslikud riskid ei ole omavahel tingimata empiirilises korrelatsioonis, võib tuleneda teine, tõenäoliselt veel suurem probleem suurandmete ja masinõppe algoritmide rakendamisel riskipõhise valimise. Eespool esitatud kolm masinõppevahendite näidet on kõik ühemõõtmelise valiku näited, st sellise valiku näited, mis põhineb riski ühel konkreetsetel määratlusel ja rakendusel. Tööelu riskid ei ole siiski ainult üht konkreetset tüüpi. Sellest tulenevalt tegelevad järelevalveasutused mitut tüüpi riskidega, nt õnnetused, kokkupuude kemikaalidega, kokkupuude bioloogiliste ohuteguritega, psühhosotsiaalsed ohud, luu- ja lihaskonna riskitegurid ning sotsiaalne dumping. Seda tüüpi riskidel on omakorda veel palju rohkem alatüüpe. Selliste riskimudelite väljatöötamine, mis suudavad hõlmata kogu mitmekesisust, on väga keeruline, sest eri tüüpi riskid ei ole tingimata korrelatsioonis. Seega erineb selle mitmekesisusega arvestamine üsna palju ühe konkreetset tüüpi riski tõenäosuse prognoosimisest (Dahl *et al.*, 2018).

Teine, kuid seotud probleem muudab riskipõhise valimise ülesande veelgi keerukamaks. See on nn poliitiline lõks (Black, 2010). Ehkki masinõppe algoritmide on dunaamilised selles mõttes, et need suudavad õppida õnnestumisest ja vigadest ning nendega kohaneda, ei arvesta need erinevate poliitiliste seisukohtadega. Esiteks on poliitiline kontekst muutlik. Seega ei pruugi praegu

prioriseerimist vajavad riskitüübid olla prioriteetsed homme. Teiseks on poliitiline kontekst mitmetahuline. Seega on eri sidusrühmad, näiteks poliitikud, tööandjad, töötajad, meedia ja üldsus eri seisukohal, mis riske tasub prioriseerida. See illustreerib, et risk töömaailmas ei ole tingimata objektiivne näitaja, vaid sotsiaalne konstruktsioon.

Kolmas probleem, millega tasub arvestada, on seotud asjaoluga, et kuigi tööinspeksioonidel on väga palju andmeid inspekteerimisobjektide kohta, on need andmed tavaliselt seotud ettevõtte tasandiga ning ettevõtte tasandi andmed ei ole tingimata kõige asjakohasemad andmed, mida arvesse võtta (vt Gunningham ja Sinclair, 2007). Andmebaasis antakse konkreetsele ettevõttele tavaliselt kordumatu identifikaator, näiteks registrikood. Masinõppe algoritmi suutlikkus määrata etteantud prognoositud riskiväärtus konkreetsele ettevõttele sõltub kordumatutest identifikaatoritest. Kõiki võimalikke kontrolliobjekte ei ole siiski võimalik automaatselt identifitseerida kordumatu identifikaatori abil. Näiteks ehitusvaldkonnas ei valita inspekteerimiseks tingimata konkreetset ettevõtet vaid ajutine ehitusplats. Sellise ajutise olemusega on seotud vähemalt kaks probleemi. Esiteks ei pruugi ehitusplatsid ja muud ajutised töökohad olla tuvastatavad kordumatute identifikaatorite abil. Teiseks, isegi kui need oleksid tuvastatavad, tähendab nende ajutisus, et masinõppe algoritm ei pruugi saada võimalust õppida eelnevatest õnnestumistest ja vigadest enne, kui ehitusplatsi tegevus lakkab ja platsil tegutsenud ettevõtte on liikunud uuele ehitusplatsile.

Lõppmärkused

Eespool kirjeldatud probleemid tõendavad, et suure riskiga inspekteerimisobjektide valimisel esineb suuri raskusi seoses suurandmete ja masinõppe meetodite kasutamisega. Samas ei tühista need probleemid kuidagi selliste meetodite kasulikkust riskipõhise lähenemisviisi raames. Pigem tõendavad probleemid, et riskipõhise valimise seisukohast ei ole tõenäoliselt kasulik tugineda täielikult masinõppe algoritmidele. Seda näitab eespool esitatud Norra näide. Selle asemel, et lubada algoritmil objekte otse leida ja valida, lubatakse inspektoritel teha riskipõhiseid otsuseid vastavalt algoritmi tehtud prognoosidele. See hõlmab tehise- ja inimintellekti kombinatsiooni, milles need täiendavad üksteist. Keerukate sotsiaalsete sündmuste üldiseks prognoosimiseks on ilmselt vaja kasutada mõlemat tüüpi intellekti koos.

Viited

- Black, J., 2010. Risk-based regulation: Choices, practices and lessons being learnt. OECD 2010 (ed.), Risk and regulatory policy: Improving the governance of risk. OECD, Pariis.
- Blanc, F., 2013. Inspection reforms: why, how, and with what results. OECD, Pariis.
- Dahl, Ø., Kilskar, S. S., Skarholt, K., Rosness, R., 2018. Risikobasert tilsyn i de nordiske arbeidstilsynene. [Riskipõhised tööinspekteerimised Põhjamaades.] Põhjamaade Ministrite Nõukogu, Kopenhaagen.
- Gunningham, N., Sinclair, D., 2007. Factors impinging on the effectiveness of the Mines Inspectorate. The Australian National University, Canberra.
- Hajakbari, M. S., Minaei-Bidgoli, B., 2014. A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data. Journal of Loss Prevention in the Process Industries, 32(1):443–453.
- Jacobusse, G., Veenman, C., 2016. On selection bias with imbalanced classes. International Conference on Discovery Science, Bari, lk 325–340.
- Matías, J. M., Rivas, T., Martín, J. E., Taboada, J., 2008. A machine learning methodology for the analysis of workplace accidents. International Journal of Computer Mathematics 85, lk 559–578.
- Mischke, C., Verbeek, J. H., Job, J., Morata, T. C., Alvesalo-Kuusi, A., Neuvonen, K., Clarke, S., Pedlow, R. I., 2013. Occupational safety and health enforcement tools for preventing occupational diseases and injuries. Cochrane Database of Systematic Reviews 2013, 8. väljaanne, art. nr CD010183.
- OECD, 2014. OECD Best practice principles for regulatory policy: Regulatory enforcement and inspections. Majanduskoostöö ja Arengu Organisatsioon, Pariis.
- Ridemar, A., 2018. Decision support for SWEA inspections. Kuninglik Tehnoloogiainstituut (KTH), Stockholm.
- Robens, L., 1972. Safety and Health at Work: Report of the Committee 1970–1972. Her Majesty's Stationery Office, London.
- Weil, D., 2008. A strategic approach to labour inspection. International Labour Review 147, lk 349–375.

*Autorid: Øyvind Dahl, vanemteadur, SINTEF Digital – Safety and Reliability, Trondheim, Norra
Annick Starren, Euroopa Tööhutuse ja Töötervishoiu Agentuur (EU-OSHA)*