

Prosesseja voi optimoida datan pohjalta – jos data sen mahdollistaa

17.2.2023

SeAMKin TehoData-hankkeen työpaketti 2:ssa toteutettiin kaksi pilottia teollisuusprosessien tehostamiseksi prosessidatan pohjalta. Pilotit toteutettiin yhteistyössä koneistusalihankintayrityksen ja toisaalta prosessiteollisuuden yrityksen kanssa. Koneistamon tapauksessa pyrittiin löytämään tunnistetuista prosessivaiheista (työstösoluista ym.) tuotteiden läpimenoon vaikuttavia pullonkauloja prosessilouhinnan avulla. Prosessiteollisuuden yrityksen kohdalla taas etsittiin juurisyitä prosessin lopputulosta kuvaavien parametrien vaihteluun prosessidatan pohjalta. Pilottien tavoitteena oli osoittaa toiminnasta kerätyn datan hyödynnettävyys toiminnan kehittämisen tukena: Koneistamossa toimitusvarmuuden kasvattamiseen ja prosessissa lopputuloksen vakioimiseen optimaaliselle tasolle.

Prosessilouhinta paljastaa tapahtumien kulun datan pohjalta

Prosessi on hallittu toimenpiteiden sarja, jonka kautta tehdään organisaatiossa tavoiteltuja asioita (Object Management Group, 2023). Prosessin eräs keskeinen piirre on sen toistuvuus: prosessin pitäisi toistua samanlaisena suorituskerrasta toiseen. Organisaatioissa pyritään kuvaamaan toimintaa prosesseina, joita johdetaan ja kehitetään tietoisesti, jättäen satunnaiselle vaihtelulle ja virheille mahdollisimman vähän tilaa. Varsin usein dokumentoidut prosessikuvaukset ja tapahtumien tosiasiallinen kulku esimerkiksi tuotantotoiminnassa eivät kuitenkaan vastaa täsmälleen toisiaan, vaan käytännön olosuhteet johtavat poikkeamiin ja improvisointiin (Matias Kokko, 2022). Prosessilouhinnalla tarkoitetaan karkeasti datalähtöistä prosessien kuvausta virallisten dokumenttien ulkopuolelta: prosessien vuot kuvautuvat sellaisiksi kuin ne

esimerkiksi toiminnanohjausjärjestelmän merkintöjen perusteella todellisuudessa ovat olleet. Tämä auttaa tunnistamaan toiminnan pullonkauloja esimerkiksi resurssien riittävyyden ja prosessin osien (tai eri prosessien) yhteensopivuuden näkökulmasta ja toisaalta tunnistamaan suunniteltujen ja toteutuneiden prosessien eroavaisuudet.

Käytännössä prosessilouhintaa tehdään siihen erikoistuneilla ohjelmistoilla ja ohjelmointikirjastoilla. Tässä pilotissa tarkasteltiin Python-kielen PM4PY-kirjastoa (Fraunhofer FIT, 2023) ja Apromore-ohjelmistoa (Apromore, 2023). Edellinen osoittautui pilotin tarpeiden kannalta käyttökelpoiseksi. Pilotissa tutkittiin, olisiko toimitusvarmuutta mahdollista parantaa louhimalla datamassasta sellaisia tilauksia tai töitä, jotka kuvastaisivat “tyypillistä” työtä tai näkemystä jostain työvaiheesta, johon kuluu rutkasti aikaa suhteessa muihin. Louhittava data saatiin pilottiyrityksen toiminnanohjausjärjestelmästä vuosilta 2007 – 2020, joten sitä oli runsaasti, enemmän kuin tarpeeksi. Toteutuneita prosessikulkuja etsittäessä kohdattiin muutamia ongelmia:

1. Niitä tuotteita, joihin liittyi useita työvaiheita, oltiin tehty vain yksittäisiä kappaleita tai hyvin pieniä sarjoja, jolloin prosessin määritelmä toistuvuudesta ei toteudu (vaan ohjaus hoituu pikemminkin projektimaisesti).
2. Runsaiden tuotantovolyymien tuotteisiin liittyi datan mukaan vain vaihe tai pari, mikä ei vastaa prosessin tarkoittamaa toimenpiteiden sarjaa. Datan ulkopuolelta voi arvuutella, onko tuotteita työstetty useammassa vaiheessa tai työpisteessä mutta tiedot jääneet kirjaamatta vai onko kyseessä jotkin hyvin yksinkertaiset tuotteet.
3. Datan laatu ja olemus vaihtuivat ajan myötä. Tämä voi viitata siihen, että datasarjan alkupäässä tuolloin tuoretta ERPiä vasta totuteltiin käyttämään ja kirjaamistottumukset muuttuivat vuosien varrella.
4. Prosessilouhinta tuotti lopputuloksena silmukkamaisen kuvaajan, josta näkee lähinnä sen itsestäänselvyyden, että tuotteille tehdään joko yksittäisiä tai useita toimenpiteitä.

Pilotin kokemusten pohjalta opittiin, että prosessilouhinnassa käytettävän datan tulee olla sisäisesti yhteneväistä ja *kaiken olennaisen* kattavaa. Tämä teoreettisesti kevyen tuntuinen vaatimus osoittautuu käytännön toiminnassa haastavaksi toteuttaa, sillä etenkin manuaalisesti kerätty data on altis laadun vaihtelulle. Jälkikäteen arvioiden, datan etukäteissiivoamiselle olisi tullut varata resurssia huomattavasti toteutunutta enemmän: Datan sisällön merkitys olisi kannattanut käydä huolellisesti läpi pilottikohteen edustajien kanssa ja jättää sieltä ainakin datasarjan alkupään vuodet kokonaan pois. Kaikkiaan pilotin havainto on, että prosessilouhinta tuottaa juuri sellaista tietoa, kuin sille syötetty data edellyttää. Laadukkaasta datasta syntyy prosessikuvia, joiden pohjalta toimintaa on mahdollista tehokkaasti havainnoida ja kehittää.

Data paljastaa tuotantoparametrien keskinäisiä riippuvuuksia

Yksinkertaisissa prosesseissa lopputuloksen ja siihen johtaneiden työvaiheiden parametrien syy-seuraussuhteiden havaitseminen on ihmiselle luonnostaan helpohkoa työtä. Kun prosessin vaiheiden ja sitä kautta parametrien vapausasteiden lukumäärä kasvaa, yhteyksien luonnollinen havaitseminen muuttuu satunnaisemmaksi. Erityisen vaikeaa on havaita sellaisia vaikutusketjuja, joissa lopputuloksen ja parametrin välinen yhteys on epälineaarinen. Eräs esimerkki tästä ilmiöstä on prosessiteollisuuden lopputuotteiden saannot, jotka voivat riippua paitsi prosessiin syötettyjen raaka-aineiden määrästä, myös raaka-aine-erien

sisäisestä laadunvaihtelusta tai vaikkapa prosessin vaiheissa käytetyistä lämpötiloista tai käsittelyajoista.

Pilotissa pyrittiin todentamaan tilastollisen datankäsittelyn hyötyjä tällaisessa juurisyyn etsimisessä prosessissa havaitulle lopputuloksen vaihtelulle.

Pilotissa toteutettiin kaksi tarkastelua: Ensimmäisessä esimerkissä tehtiin erilaisia analyysejä datalle ja koitetaan löytää sieltä lopputulosta ennustavia parametrien keskinäisiä riippuvuussuhteita. Toisessa esimerkissä mittaustulosten perusteella luokiteltiin prosessin lopputuloksen laatua. Pilottiyrityksen tapauksessa tuotantodataa oli prosessikohtaisesti varsin hyvin saatavilla (n. 30 mittapistettä), mutta prosessin toistoja turhan niukasti (n. 300). Datan siivoaminen tehtiin Knime-työkalulla (Knime, 2023), joka mahdollistaa erilaisten datankäsittelyputkien luomisen low-code tyylisesti. Eräs haaste datan hyödynnettävyydelle syntyi datankeruuprosessista: datapisteet oli mahdollisesti tallennettu sähköiseen muotoon vasta varsinaisen tapahtumahetken jälkeen, sillä merkinnöissä oli laatuvaihtelua.

Pilotin tarkasteluissa todettiin, että varsinaisia lopputulosta selittäviä muuttujapareja ei löydetty datasta, eli saannon ja mitattujen parametrien väliltä ei löydetty selkeitä korrelaatioita. Joko dataa oli liian vähän tai tulokseen vaikuttavaa parametria tai parametrejä ei yksinkertaisesti oltu mitattu prosessista. Jälkimmäisessä tapauksessa kyse olisi sikäli merkittävästä havainnosta, että prosessin kehittämisponnistukset tulisi suunnata johonkin muualle kuin nyt kontrolloituihin parametreihin.

Datan käsittelyyn tulee varata riittävästi resursseja ennen varsinaisia analyysejä

Molempien pilottien keskeinen havainto oli, että tuotantojärjestelmistä kerättävään dataan tulee kiinnittää huomiota ennen analyysien tekemistä: ylipäänsä datan analysoimiseksi sen esikäsittelemiseen täytyy varata riittävästi resursseja. Toisaalta, datan semantiikka suhteessa reaaliprosessiin on ymmärrettävä riittävän syvällisesti, datan keruutavat tulee olla tiedossa ja analyysiin vietävät datarivit tulee tuntea riittävän hyvin, jotta niiden keskinäisestä koherenssista voidaan olla varmoja. Lopultakin yleiset datan hyödynnettävyyteen liittyvät vaatimukset ovat voimassa: jotta tuotantodatan pohjalta voidaan tehdä tietoisia toiminnan kehittämispäätöksiä, datan tulee olla

1. Virheetöntä. Datan täytyy olla tarkkaa ja virheetöntä ollakseen käyttökelpoista.
2. Täydellistä. Jotta datasta saadaan vastauksia, sen täytyy olla aukotonta ja kuvata kaikki tapahtuneet seikat.
3. Vanhentunut data ei hyödytä liiketoimintaa.
4. Dataan täytyy päästä käsiksi, jotta sitä voidaan käyttää.

Tommi Ylimäki
lehtori
SeAMK, TKI

Juha-Matti Arola
projektipäällikkö
SeAMK, TKI

Lähteet:

Apromore. (2023). https://apromore.com

Fraunhofer FIT. (2023). State-of-the-art-process mining in Python. https://pm4py.fit.fraunhofer.de/

Knime. (2023). End to End Data Science. https://www.knime.com/

Matias Kokko, M. V. (20.12 2022). Mikä on prosessilouhinta. https://blogit.lab.fi/labfocus/mika-prosessilouhinta/

Object Management Group. (7.2.2023). Business process model and notation (BPMN Glossary). https://www.bpmn.org