



TEKNILLINEN TIEDEKUNTA

**Datapohjaiset mallit teräksen valmistuksessa,
erityisesti konvertterissa**

Janne Määttä

Prosessitekniikka

Kandidaatintyö

Toukokuu 2022

TIIVISTELMÄ

Datapohjaiset mallit teräksen valmistuksessa, erityisesti konvertterissa

Janne Määttä

Oulun yliopisto, Prosessitekniiikan tutkinto-ohjelma

Kandidaatintyö 2022, 26 s.

Työn ohjaajat yliopistolla: Dos. Jari Ruuska & Dos. Aki Sorsa

Työn tavoitteena on tutkia datapohjaisten mallien hyödyntämistä hiiliteräksen valmistuksessa sekä analysoida ja mallintaa kirjallisuudesta löytyvä, terästehtaalta kerätty data. Tarkastelu on rajattu vain teräskonvertteriprosessiin. Työssä perehdytään konvertteriprosessiin sekä datapohjaisiin malleihin, joita prosessissa on käytetty. Tämän lisäksi työssä tutustutaan yleisesti datapohjaisen mallinnuksen eri vaiheisiin, kuten datan esikäsittelyyn, tärkeiden muuttujien valintaan, datan jakamiseen opetus-, testaus-, ja validointiaineistoon sekä mallin luomiseen. Mallin tarkkuuden ja hyvyyden arviointimenetelmiä on myös tarkasteltu. Työssä luodaan lineaarinen monimuuttujaregressiomalli teräksen saannon ennustamiseksi terästehtaan mittausdatan perusteella. Terästehtaan mittausdataa analysoidaan eksploratiivisen data-analyysin avulla ja mallin hyvyttä arvioidaan eri kriteerein. Työssä luodun datapohjaisen mallin suhteellisen virheen keskiarvoksi on saatu 0,578 %, joten mallin tarkkuus on hyvä. Mallin ennustamat saannot sekä todelliset mittaukset antavat lähes saman lopputuloksen. Luotua mallia on verrattu myös samasta datasta luotuun neuroverkkomalliin ja eri arviointikriteerien perusteella voidaan sanoa työssä luodun lineaarisen monimuuttujaregressiomallin olevan tarkempi kuin neuroverkkomallin.

Asiasanat: teräskonvertteri, datapohjainen malli, lineaarinen monimuuttujaregressiomalli

SISÄLLYSLUETTELO

TIIVISTELMÄ

SISÄLLYSLUETTELO

1 Johdanto	4
2 Konvertteri	5
3 Datapohjaiset mallit	7
3.1 Datapohjaiset mallit teräskonvertterissa	8
4 Kokeellinen osuus	15
4.1 Eksploratiivinen data-analyysi	15
4.2 Lineaarinen regressiomalli	18
4.3 Mallin hyvyyden arviointi.....	19
5 Pohdinnat ja johtopäätökset	22
6 Yhteenveto	24

LÄHDELUETTELO

1 JOHDANTO

Työn aiheena on datapohjaiset mallit teräksen valmistuksessa, erityisesti konvertterissa. Työ koostuu kahdesta osasta, kirjallisuustutkielmasta ja kokeellisesta osuudesta. Työn tavoitteena on tutustua datapohjaisten mallien hyödyntämiseen teräskonvertteriprosessissa, suorittaa kirjallisuudatalle eksploraatiivinen data-analyysi ja muodostaa regressiomalli kirjallisuusaineistolle. Datapohjaisten mallien tarkastelu on rajattu vain hiiliteräksen valmistuksen teräskonvertteriprosessiin ja siinä käytettyihin malleihin.

Kirjallisuustutkielmassa on tarkasteltu konvertterin toimintaperiaatetta ja konvertteriprosessiin luotuja datapohjaisia malleja. Datapohjaisten mallien avulla ennustetaan esimerkiksi puhalluksen jälkeistä teräksen loppulämpötilaa sekä loppupitoisuuksia. Lisäksi työssä tarkastellaan datapohjaisten mallien luomisen vaiheita, kuten datan jakamista opetus-, testaus-, ja validointiaineistoon. Merkittävimpien muuttujien valintaa ja luodun mallin hyvyyden arvioinnin erilaisia menetelmiä sekä kriteereitä on tarkasteltu.

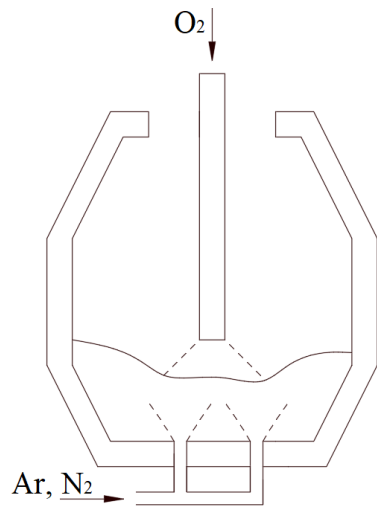
Kokeellisessa osassa analysoidaan terästehtaan mittausdataa eksploraatiivisen data-analyysin avulla ja luodaan lineaarinen monimuuttujaregressiomalli teräksen loppusaannon ennustamiseksi. Data-analyysi sisältää esimerkiksi muuttujien välisten riippuvuuksien tarkasteluja. Mallin hyvyyttä tarkastellaan erilaisten kriteerien avulla, jotka ovat esitetty työssä. Tässä työssä luotua lineaarista monimuuttujaregressiomallia on verrattu saman datan pohjalta luotuun neuroverkkomalliin. Erilaisten kriteerien ja arvioiden perusteella voidaan sanoa työssä luodun monimuuttujaregressiomallin olevan parempi kuin neuroverkkomalli.

2 KONVERTTERI

Hiiliteräksen valmistuksessa konvertteriprosessin kaksi päätehtävää ovat masuunilta tulevan raakaraudan hiilipitoisuuden alentaminen sekä sen lämpötilan nostaminen. Konvertterissa raakarauda muuttuu raakateräkseksi, kun raakaraudan hiilipitoisuus laskee 4 %:sta alle 0,08 %:iin. Sulan loppulämpötila on noin 1650 °C. Konvertteriin panostetaan raakaraudan lisäksi kierrätysterästä. Panostuksen jälkeen aloitetaan puhallusvaihe, jossa sulan pinnalle puhalletaan puhdasta happea. Happi puhalletaan happilanssilla, joka laskeutuu yläpuolelta konvertterin sisälle. Prosessin tarkoituksena on poistaa sulasta hiilen lisäksi epäpuhtauksia, kuten piitä, mangaania sekä fosforia. Lämpötilan nostaminen on puhalluksen avulla mahdollista, koska lämpöä syntyy hapetusreaktiosta. Reaktiossa hiili hapettuu hiilimonoksidiksi sekä hiilidioksidiksi ja ne poistetaan prosessista. Puhalluksen aikana konvertteriin lisätään kalkkia ja/tai dolomiittikalkkia kuonan muodostamiseksi. Hapettuneet epäpuhtaudet, pii, mangaani ja fosfori, muodostavat yhdessä kalkin sekä dolomiittikalkin kanssa kuonan, joka on teräksen pinnalla konvertterissa. Kun vaadittu hiilipitoisuus ja lämpötila on saavutettu, nestemäinen teräs kaadetaan konvertterista kaatoreiän kautta senkkaan. Koko prosessi on kestoltaan noin puoli tuntia. (Wang et al. 2010)

Lähteen (Jalkanen & Holappa 2014) mukaan puhalluksen aikana konvertteriin puhalletaan hapen lisäksi myös inerttiä kaasua (yleensä typpeä tai argonia) pohjasuuttimien avulla. Inertin kaasun puhaltamisen pääasiallinen tarkoitus on sulan sekoitus. Samalla tapahtuu useita monimutkaisia kemiallisia reaktioita, jotka etenevät samanaikaisesti. Samaan aikaan tapahtuu myös voimakasta lämmöntuotantoa sekä kulutusta. Prosessin monimutkaisuudesta johtuen puhalluksen jälkeisen teräksen kemiallisen koostumuksen sekä lämpötilan ennustaminen on haastavaa. Lämpötilalla on merkittävä vaikutus koko prosessiketjuun, sillä teräksen valuprosessin onnistuminen on riippuvainen siitä. Loppulämpötilan sekä kemiallisen koostumuksen tarkka arviointi ovat ratkaisevia tekijöitä teräksen laadun ylläpitämiseksi. (Sala et al. 2019)

Konvertterissa tapahtuva hapen puhallus happilanssin avulla sekä inertin kaasun puhallus pohjan kautta on esitetty kuvassa 1.



Kuva 1. Teräskonvertteri. Mukailen lähteestä (Hüsken et al. 2011).

3 DATAPOHJAISET MALLIT

Datan kerääminen ja tallennus teollisista prosesseista on muuttunut yhä helpommaksi hajautettujen ohjausjärjestelmien sekä uusien mittausteknologioiden ansiosta. Samalla kehittyneet datan käsittelymenetelmät luovat mahdollisuuksia datapohjaisten mallien muodostamiselle ja hyödyntämiselle teollisissa prosesseissa. Näiden seurauksena datapohjaiset sovellukset prosessien monitoroinnissa, säädössä ja optimoinnissa ovat yleistyneet. Datapohjaiset menetelmät ovat sopeutuvia monimutkaisiin teollisiin prosesseihin. (Ge 2017)

Datapohjaisessa mallinnuksessa datan esikäsittely on kriittinen vaihe ennen varsinaisen mallin luomista. Esikäsittelyn tarkoituksena on parantaa datan laatua ja muokata alkuperäinen data sopivaan muotoon. Poikkeavat mittaustulokset poistetaan datasta ja puuttuva data täytetään. Käytössä olevasta datasta määritetään opetus-, validointi- ja testausdata. Tämän jälkeen voidaan muodostaa datapohjainen malli opetusdatalle. Validointidataa käytetään mallin muodostamisessa antamaan indikaatiota soveltuvimmista valinnoista. Näitä valintoja ovat esimerkiksi sopivan mallirakenteen valinta. Testausdataa käytetään arvioimaan mallin yleistyskykyä. Mallin muodostaminen on erittäin riippuvainen datan ominaisuuksista sekä eri prosessimuuttujien välisistä suhteista. Tästä johtuen datan tunteminen on tarpeellista ennen sopivan mallin rakenteen valitsemista. Datapohjainen malli voidaan luoda esimerkiksi vaikeasti mitattavien suureiden ennustamista varten. Malli voi olla esimerkiksi lineaarinen regressiomalli. (Ge 2017)

Lähteen (Runkler 2020) mukaan korrelaatio määrittelee riippuvuuden muuttujien välillä. Se ei kuitenkaan tarkoita syy-seuraussuhdetta. Korrelaatioanalyysin tarkoituksena on löytää nämä riippuvuudet, jotta havaitut ilmiöt voidaan selittää tai vastaavasti halutut lopputulokset saavutetaan. Lineaarisen korrelaation menetelmät ovat laskennallisesti tehokkaita, mutta ne havaitsevat vain lineaariset riippuvuudet. Epälineaarista riippuvuutta on mahdollista tutkia epälineaarisilla korrelaatiomenetelmillä, mutta tällöin parametrit täytyy valita tarkkaan. Vaikka lineaarinen korrelaatio on kahden muuttujan välillä pientä, epälineaarinen korrelaatio voi olla hyvinkin vahva. Yksi mahdollinen tapa tutkia epälineaarista korrelaatiota on esimerkiksi χ^2 -testi. Korrelaatioanalyysin avulla voidaan esimerkiksi parantaa tuotantolaitoksen saannon laatua muokkaamalla systemaattisesti vain merkittävimpiä muuttujia.

Edellä mainitut korrelaatiomenetelmät ilmaisevat vuorovaikutuksen suuruuden ja suunnan muuttujien välillä, kun taas regression avulla arvioidaan todellista toiminnallista riippuvuutta. Korrelaatioanalyysistä saadaan siis selville ne muuttujat, jotka vaikuttavat lähtöarvoon. Regressiomallin avulla voidaan esimerkiksi kuvata, mihin arvoihin muuttujat tulisi asettaa, jotta lähtöarvona saadaan halutut arvot. Muuttujien valinta mahdollistaa relevanttien muuttujien huomioon ottamisen regressiomalleissa, jolloin mallit ovat tarkempia. (Runkler 2020)

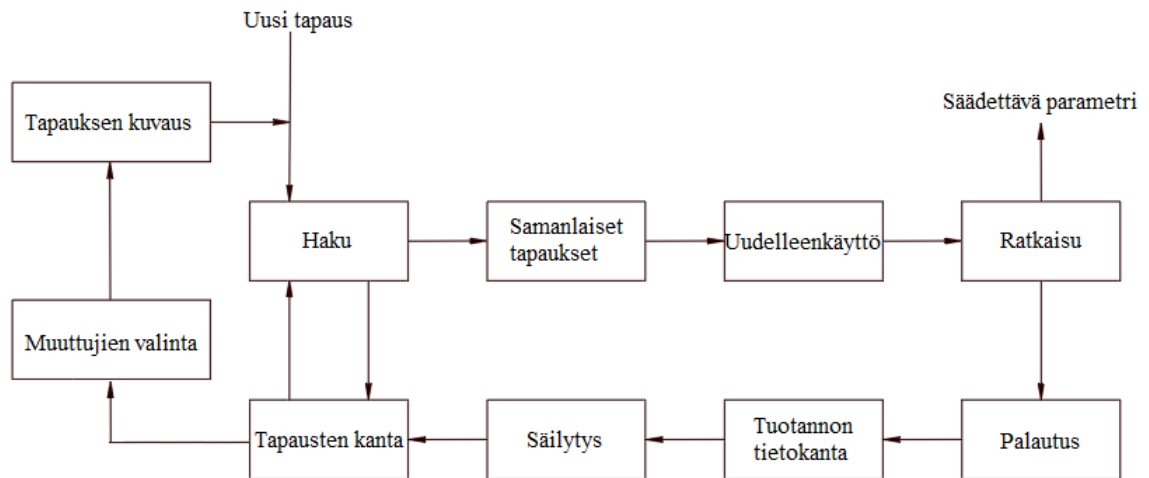
Lineaarisia regressiomalleja voidaan tehokkaasti laskea kovarianssien avulla, mutta ne rajoittuvat lineaarisiin riippuvuuksiin. Epälineaarissa regressiomallissa muuttujat, kuten vuorovaikutukset ja neliölliset vaikutukset, ovat epälineaarissa muodossa. Esimerkiksi neuroverkot ovat epälineaarisia regressiomalleja. (Runkler 2020)

3.1 Datapohjaiset mallit teräskonvertterissa

Lähteessä (Wang et al. 2017) esitellään menetelmä hiilipitoisuuden reaaliaikaiseen seurantaan prosessin toisessa puhallusvaiheessa. Samalla menetelmällä voidaan arvioida myös loppuhiilipitoisuutta. Toinen puhallusvaihe tarkoittaa vaihetta ensimmäisen mittauksen jälkeen. Menetelmässä aikapohjainen eksponentiaalinen hiilenpoistomalli on muutettu happipohjaiseksi. Tämän jälkeen sopivaksi datapohjaiseksi menetelmäksi on valittu tapauskohtainen päättely (case-based reasoning), jonka avulla on selvitetty hiilenpoiston avainparametri. Mallin tarkkuuden parantamiseksi on käytetty keskinäisinformaatioon (mutual information) perustuvaa algoritmia, jonka perusteella on valittu tärkeät muuttujat tapauskohtaiseen päättelyyn. Mallin hyvyys on arvioitu oikealla prosessidatalla; osa datasta on käytetty alkuperäisenä tapauksena ja jäljellä olevalla datalla on testattu menetelmää. Mallin hyvyttä on arvioitu ennustetun ja todellisen hiilipitoisuuden osumistarkkuudella sekä virheen keskineliöllä (mse, mean squared error).

Tapauskohtainen päättely on datapohjainen menetelmä, jossa opitaan menneistä tapauksista. Menetelmän perusidea on se, että samanlaisilla ongelmilla on usein samanlaiset ratkaisut. Menetelmä sisältää seuraavat vaiheet: tapauksen kuvaus, tapauksen haku, tapauksen uudelleenkäyttö, tapauksen tarkastaminen sekä tapauksen palautus. Tapauksen kuvauksen perusteella haetaan samanlaisia tapauksia, jonka jälkeen näiden tapauksien ratkaisujen avulla luodaan uusi ratkaisu nykyiselle tapaukselle. (Wang et al.

2017). Tapauskohtaisen päättelyn vaiheet säädettävän parametrin ratkaisemiseksi on esitetty kuvassa 2.



Kuva 2. Tapauskohtaisen päättelyn vaiheet säädettävän parametrin ratkaisemiseksi. Mukailen lähteestä (Wang et al. 2017).

Wang et al. (2010) esittelevät tutkimuksessaan painotetun tukivektorikoneen konvertterin puhalluksen jälkeisen teräksen hiilipitoisuuden sekä lämpötilan ennustamiselle. Mallissa on sovellettu kahta muuttujien valintavaihetta. Ensiksi konvertterissa tapahtuvaa reaktiota ja sen mekanismia on analysoitu ja määritetty jokaisen muuttujan merkittävyys, jotta mallin tuloarvojen määrä on saatu mahdollisimman pieneksi. Toiseksi on laskettu keskinäisinformaatiot tulo- sekä lähtömuuttujien välille. Keskinäisinformaatio on valittu käytettäväksi, sillä sen avulla on mahdollista arvioida yleistä riippuvuutta kahden muuttujan välillä. Laskettujen keskinäisinformaatioiden arvojen avulla määritetään painotetut kertoimet, jotta tärkeimpien tulomuuttujien vaikutus kasvaa lähtöarvossa. Ennen mallinnusta data on normalisoitu välille 0–1.

Konvertteriprosessin merkittäviä tulomuuttujia ovat mittauslanssilla mitattu hiilipitoisuus ja lämpötila sekä tämän lisäksi puhallettu hapen määrä että lisättyjen jäähdysaineiden määrä. Loppuhiilipitoisuuden ennustamista varten tulomuuttujiksi on valittu mitattu hiilipitoisuus ja puhalletun hapen määrä. Loppulämpötilan ennustamismalliin on valittu kaikki mainitut muuttujat. Kun sopivat muuttujat on valittu, tukivektorikonemenetelmää voidaan käyttää loppupistettä ennustavan mallin rakentamiseen. (Wang et al. 2010)

Tukivektorikoneen avulla kartoitetaan tulomuuttujien sekä lähtömuuttujien välisiä riippuvuuksia moniulotteisessa datassa. Tässä muuttujien avaruudessa suoritetaan lineaarinen regressio, jolloin saadaan mallille yhtälö. Tukivektorikoneen parametrit

saadaan selville datan ristivalidoinnin avulla. Mallinnukseen on käytetty 35 datapistettä ja mallin testaukseen 25 datapistettä. Mallin hyvyden arviointi on tehty laskemalla osumistarkkuus mitatun sekä ennustetun lämpötilan sekä hiilipitoisuuden välillä ja mallin tarkkuutta on arvioitu virheen keskineliön neliöjuurella (rmse, root mean squared error). (Wang et al. 2010)

Konvertterin puhalluksen jälkeistä teräksen lämpötilaa sekä hiilipitoisuutta on tutkittu myös lähteessä (Li et al. 2013). Lähteen mukaan lineaarisen regressioanalyysin perusteella luodaan ennustava lineaarinen regressiomalli. Tämän jälkeen vastepintamenetelmän (Response Surface Method, RSM) avulla luodaan epälineaarinen ennustava malli. Lineaarisen regressioanalyysin teorian perusteella luodaan lineaariset mallit lämpötilan sekä hiilipitoisuuden ennustamiselle. Lineaaristen mallien tuntemattomat parametrit ratkaistaan lähteessä esitetyn algoritmin avulla. Vastepintamenetelmän tarkoituksena on muodostaa vasteen ja tulomuuttujien riippuvuutta kuvaava matemaattinen funktio ja sen avulla minimoida tai maksimoida tutkittava vaste. Menetelmä sisältää tärkeiden tekijöiden valitsemisen niin, että kokeet voidaan suorittaa tehokkaasti ja suoritettavien kokeiden lukumäärä on mahdollisimman vähäinen. Kokeista saatu data mallinnetaan ja analysoidaan ja vastemuuttuja optimoidaan. Vastepinta määritetään ensimmäisen ja toisen kertaluvun malleilla. Mallit validoidaan vertaamalla testausdatan perusteella saatuja tuloksia mitattuihin arvoihin. Tulosten perusteella toisen kertaluvun mallilla saadaan tarkempia tuloksia kuin lineaarisella regressiomallilla. Tuloksia on vertailtu virheen keskineliön avulla.

Sala et al. (2019) ovat lähteessään luoneet datapohjaisen mallin moniulotteisen aikasarjan avulla ennustaakseen konvertterin puhalluksen jälkeistä teräksen lämpötilaa sekä kemiallista koostumusta. Mallin avulla voidaan ennustaa fosforin, mangaanin, rikin sekä hiilen pitoisuutta teräksessä puhallusvaiheen jälkeen.

Ennustus perustuu kahteen erilaiseen datajoukkoon, joista molemmista luodaan mallit. Ensimmäinen data sisältää staattisia muuttujia, kun taas toinen sisältää aikasarjadataa. Aikasarjadatan avulla luodaan tarkempi ennuste loppupisteen arvoista ja tämän lisäksi sen avulla voidaan säätää tarvittaessa prosessin parametreja. (Sala et al. 2019)

Staattista mallia varten muuttujat valitaan mallipohjaisella analyysillä. Dynaamisen mallin tekoa varten kerättyä dataa täydennetään staattisen mallin tärkeimmillä muuttujilla. Näin saadun datan pohjalta muodostetulla ennustavalla mallilla voidaan

ennustaa loppupisteen arvoja ennen mittaamista, jolloin on mahdollista laskea esimerkiksi tarkemmat arviot lämpötilan säädölle. Toinen ennustava malli luodaan konvertterin happipuhalluksen aikana mitatun aikasarjadataan perusteella. Molemmat ennustavat mallit on tehty erilaisilla mallirakenteilla, kuten harjanneregressiolla (ridge regression), satunnaismetsillä (random forest) sekä gradient boost-päätöksentekopuuregressiolla (gradient boosted regression trees). Harjanneregressio on lineaarinen malli, ja satunnaismetsä sekä päätöksentekopuuregressio ovat epälineaarisia monimuuttujamalleja. Mallien hyvyyden arviontiin on käytetty virheen keskineliön neliöjuurta. (Sala et al. 2019)

Lähteessä (Ruuska et al. 2003) on luotu malli hyödyntäen datapohjaisia menetelmiä puhalluksen jälkeisen teräksen lämpötilan ennustamiseksi. Tavoitteena on luoda mahdollisimman yksinkertainen malli, jotta laskennat pysyvät kohtuullisen helppoina. Malliksi on valittu lineaarinen malli, jonka kulmakertoimen määritetään alkuarvojen perusteella, sillä lämpötilan muutoksen laskemiseen käytettävissä oleva aika on lyhyt sekä laskelmat ovat yksinkertaisempaa toteuttaa. Kulmakertoimen laskentaan käytetään mitattua loppulämpötilaa sekä pudotussondilla mitattua lämpötilaa. Näiden lisäksi laskennassa huomioidaan koko puhallukseen kulunut aika sekä aika, joka on kulunut pudotussondin käyttöön mennessä.

Datan esikäsittelyä varten on luotu neljä rakenteellisesti samanlaista alustavaa mallia. Data on suodatettu kolmessa vaiheessa eri perustein. Esimerkiksi ensimmäisessä vaiheessa on jätetty pois kaikki konvertterin puhallukset, joissa vaadittavia mittauksia ei ole suoritettu ja puhallukset, joissa pudotussondin avulla on mitattu lämpötila vain puhalluksen jälkeen. Kahdessa muussa vaiheessa datasta on poistettu lisää tapauksia, esimerkiksi ne, joissa kulmakertoimen numeroarvo on yli yhden. (Ruuska et al. 2003)

Mallien hyvyyttä on tutkittu tilastollisten parametrien avulla ja on huomattu, että kulmakertoimen keskihajonta laskee mitä parempi malli on kyseessä. Datan esikäsittely on validoitu kahdella erillisellä datasetillä. Datasettien avulla on arvioitu myös alustavien mallien tehokkuutta. Konvertterin panoskoko vaihtelee sekä tämän lisäksi loppulämpötila ja hiilipitoisuus eivät aina ole samat. Tämän vuoksi dataa on jaettu eri mallinnusryhmiin. (Ruuska et al. 2013)

Wang et al. (2014) ovat kehittäneet datapohjaisen mallin teräksen fosforipitoisuuden ennustamiselle. Malli on luotu yhdistämällä monimuuttujaregressioanalyysi sekä

rekursiivisuus toisiinsa, jolloin on saatu monitasoinen rekursiivinen regressiomalli. Merkittävimpiä tekijöitä fosforin poistossa on tutkittu prosessidatasta tehtyjen korrelaatioanalyysien avulla. Korrelaatiota on tutkittu fosforin poistosuhteen ja teräksen loppulämpötilan, kuonan rautaoksidipitoisuuden sekä kuonan emäksisyyden välillä.

Usean selittäjän regressiomalli on luotu monimuuttujaregressioanalyysin perusteella. Regressiomallin tarkkuus on rajallinen, kun käytetään ajan suhteen vakioita malliparametreja. Jos ajan suhteen muuttuvat parametrit yksinkertaistetaan olettamalla niiden olevan vakioita, virheellisten ennusteiden määrä kasvaa huomattavasti. Tästä johtuen ennustuksen tarkkuuden parantamiseksi yhdistetään analyysiin myös rekursiivisuus. Tällöin saadaan malli, joka ottaa huomioon ajan suhteen muuttuvat parametrit. Mallia laatiessa on hyödynnetty prosessidataa noin 1450 sulatuksesta. Data on jaettu kahteen osaan: toinen osa mallin luomista varten ja toinen sen hyvyden testaamiselle. Mallia on testattu myös kahdelta muulta terässulatolta mitatun datan perusteella ja arviointikriteerinä on käytetty mitatun ja ennustetun pitoisuuden osumistarkkuutta. (Wang et al. 2014)

Lähteessä (Barui et al. 2019) on esitelty monimuuttujaregressiomalli fosforipitoisuuden ennustamiseksi ja sen toimintaa on tutkittu useista näkökulmista. Jotta malli toimii, täytyy data muokata siihen muotoon, että regressiomallin teorian vaatimukset täyttyvät. Vaatimusten mukaan esimerkiksi virheen keskiarvo täytyy olla nolla sekä virheiden varianssin on oltava ajan suhteen vakio. Virheiden normaalisuutta on tutkittu normaalikvantiilikuvaajalla (QQ-plot) sekä Shapiro-Wilkin testillä. Jos virheet ovat normaaleja, QQ-kuvaajan mukaan pisteet olisivat suoralla viivalla. Muiden oletusten validointiin on käytetty erilaisia menetelmiä. Esimerkiksi mallin lineaarisuutta on validoitu standardoidulla residuaalikuvaajalla. Tarpeettomat selittävät muuttujat on eliminoitu vaiheittaisella regressiolla (stepwise regression).

Malli on validoitu k-fold-ristiinvalidointimenetelmällä, jossa datasetti jaetaan k yhtä suureen osaan satunnaisesti. Menetelmässä lasketaan keskineliövirheen neliöjuuria (RMSE) eri k:n arvoilla. Pienet RMSE:n arvot kuvaavat siitä, että malli on toiminut kohtuullisesti. (Barui et al. 2019)

Tuloksia on tutkittu kuvaavilla tilastoilla, joiden perusteella voidaan tehdä erilaisia tulkintoja. Tilastot sisältävät kaikkien muuttujien tietoja, kuten näytteiden määrän, keskiarvon, keskihajonnan sekä mediaanin. Esimerkiksi fosforipitoisuuden arvoista

piirretyn histogrammin avulla voidaan arvioida virheiden normaalisuutta. Regressiomallien sopivuutta on tutkittu myös selityksasteen eli R^2 -suureen avulla. (Barui et al. 2019)

Cox et al. (2002) ovat tutkineet keinotekkoisten neuroverkkojen (artificial neural network, ANN) hyödyntämistä teräskonvertterin puhalluksen loppuvaiheessa. Neuroverkon avulla ennustetaan puhallettavan hapen ja jäähdytysaineiden määrää prosessin loppuvaiheelle, jotta haluttu lämpötila sekä hiilipitoisuus saavutetaan.

Keinotekoinen neuroverkko on mallinnusmenetelmä, jossa ongelman ratkaisu opetetaan verkolle esimerkkiaineiston avulla. Yleisin käytössä oleva keinotekoinen neuroverkko on eteenpäin kytketty neuroverkko, jonka opetuksessa on käytetty backpropagation-menetelmää. Se on verkkorakenne, joka oppii approksimoimaan esimerkkiaineistoa. Konvertterin tapauksessa aineiston riippuvuudet ovat epälineaarisia, mutta se ei tuota ongelmia käytetylle neuroverkolle. (Cox et al. 2002)

Neuroverkko koostuu neuroneista, jotka ottavat vastaan useita tuloarvoja ja laskevat niiden painotetun summan. Sen jälkeen neuronit käyttävät aktivaatiofunktioita ja antavat lähtöarvona lopputuloksen. Neuronit muodostavat yhdessä kokonaisia kerroksia. Keinotekoinen neuroverkko sisältää tulokerroksen, lähtökerroksen sekä vähintään yhden piilokerroksen. Neuroverkkoon tuodaan epälineaarisuus käyttämällä aktivaatiofunktioita tuloarvojen summaan. Aktivaatiofunktiona on käytetty sigmoid-funktioita, joka skaalaa numeroarvot välille 0–1. (Cox et al. 2002)

Neuroverkko opetetaan opetusdatan avulla. Opetusdatan tuloarvot syötetään verkkoon, jolloin saadaan joukko lähtöarvoja. Verkon laskemaa lähtöarvoa verrataan datan lähtöarvoon, ja niiden väliset virheet syötetään takaisin verkoston läpi. Neuroverkkoa ei kuitenkaan saa opettaa liikaa, sillä silloin se oppii vain opetusdatan yksityiskohdat tulo- ja lähtöarvojen kartoituksen sijasta. (Cox et al. 2002)

Mallia varten data kerättiin tehtaan tietokannasta ja prosessiparametrit valittiin prosessin tuntemuksen perusteella. Neuroverkkomalli luotiin ennustamaan jäähdytysaineen lisäystarvetta. Jos mallin ennusteen mukaan jäähdytysainetta täytyy lisätä, ohjautui tieto uuteen neuroverkkomalliin, joka ennusti opetetun datan perusteella lisättävän jäähdytysaineen määrän. Neuroverkon perusteella saatu hapenmäärä seurasi todellista

tarvetta. Mallin hyvyyttä arvioitiin vertailemalla neuroverkon antamia tuloksia käytössä olevaan malliin. (Cox et al. 2002)

4 KOKEELLINEN OSUUS

Lähteessä (Laha, 2013) on tutkittu terästehtaan teräksen saantoa ja siihen vaikuttavia tekijöitä. Saanto on määritelty valetun teräksen rautapitoisuuden ja raakaraudan, romuraudan sekä rautamalmin yhteenlasketun rautapitoisuuden suhteena. Tarkasteltava teräksenvalmistusprosessi on ahjomellotus (open hearth furnace). Muuttujia on kerätty yhteensä kymmenen, mutta tärkeimmät teräksen saantoon vaikuttavat muuttujat on valittu regressioanalyysin (the best subset regression analysis) avulla. Analyysistä on saatu tärkeimmiksi teräksen saantoon vaikuttaviksi muuttujiksi raakaraudan määrä (hot metal rate, HMRT), sulatuksen koko (heat size, HSZ), malmin määrä (ore rate, ORRT) sekä panostetun romun määrä (scrap charged, SCH). Nämä muuttujat sisältävä data on jaettu kolmeen osaan: opetus-, testaus- ja validointiaineistoon. Opetusaineisto sisältää 38 mittausta, testausaineisto 8 mittausta ja validointiaineisto 8 mittausta.

Teräksen saantoon vaikuttavien muuttujien valintaa voidaan tutkia eri data-analyysimenetelmien lisäksi myös prosessituntemuksen avulla. Tässä tapauksessa prosessiin panostettavan raakaraudan määrällä on merkitystä teräksen lopulliseen saantoon, sillä raakaraudan lisääminen johtaa siihen, että myös teräksen määrä lisääntyy kuonassa epäpuhtauksien ohella. Myös panostettavan romun määrällä on vaikutusta teräksen saantoon, sillä suurempi määrä romua vaatii pidemmän sulatusajan. Pidemmän ajan kuluessa on mahdollista, että teräs hapettuu kuonan sekaan, jolloin teräksen lopullinen saanto pienenee. Saantoon vaikuttaa myös sulatuksen koko, sillä se määräytyy panostetun romun ja raakaraudan perusteella. Prosessiin lisätään myös rautamalmia, jolla on vaikutus teräksen saantoon. Kun rautamalmia lisätään, prosessissa kuluu enemmän aikaa, sillä kaikki muut reaktiot hidastuvat. Tällöin terästä häviää kuonaan tai erilaisina kaasuina, jolloin sillä on vaikutusta saantoon.

Seuraavaksi tässä työssä perehdytään artikkelissa esitettyyn dataan. Dataa analysoidaan ja siitä luodaan lineaarinen malli, jonka jälkeen mallin hyvyttä arvioidaan.

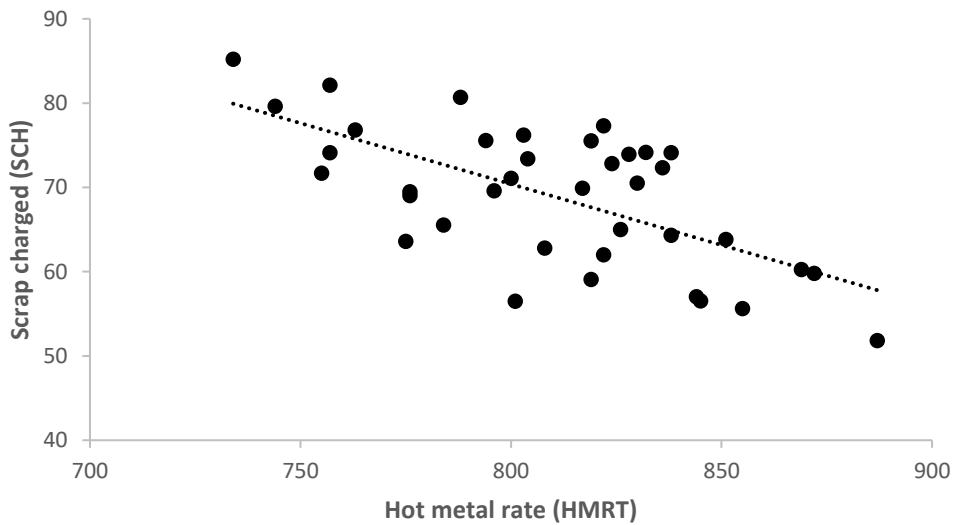
4.1 Eksploratiivinen data-analyysi

Opetusaineiston muuttujista on luotu korrelaatiomatriisi, josta on helposti nähtävissä muuttujien välisiä korrelaatioita. Korrelaatiomatriisi on esitetty taulukossa 1. Matriisista huomataan, että panostetun romun ja raakaraudan välillä on merkittävää korrelaatiota.

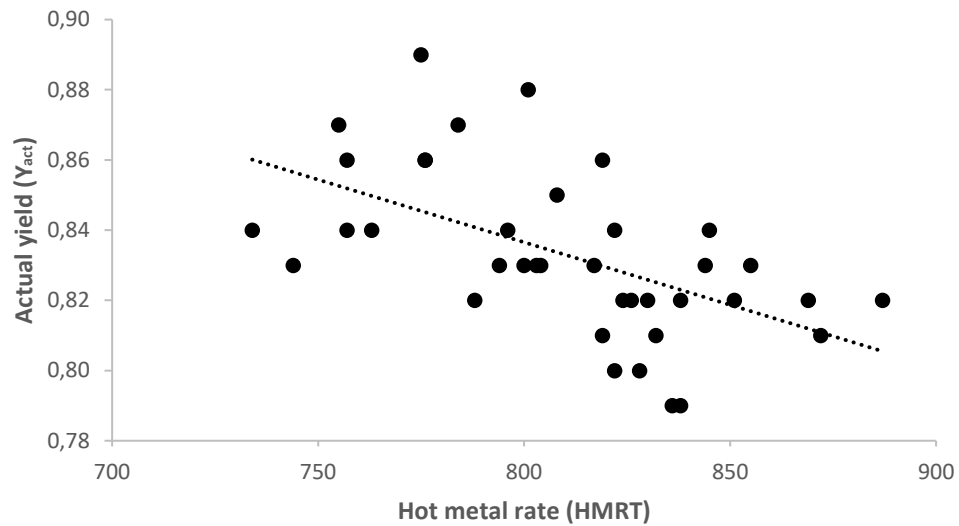
Matriisista huomataan myös, että teräksen saannon sekä panostetun romun välillä on vähäistä korrelaatiota. Muiden muuttujaparien välillä korrelaatio on vähäistä tai heikkoa. Nämä kertovat vain kuitenkin mahdolliset lineaaristen riippuvuuksien suuruudet muuttujien välillä. Kaksi suurinta muuttujien välistä korrelaatiota ovat esitetty xy-kuvaajina kuvissa 3 ja 4.

Taulukko 1. Korrelaatiomatriisi. HMRT on raakaraudan määrä, HSZ on sulatuksen koko, ORRT on malmin määrä, SCH on panostetun romun määrä ja Y_{act} on teräksen saanto.

	HMRT	HSZ	ORRT	SCH	Y_{act}
HMRT	1,00				
HSZ	-0,36	1,00			
ORRT	0,28	0,30	1,00		
SCH	-0,65	0,17	-0,32	1,00	
Y_{act}	-0,56	0,25	-0,25	-0,22	1,00

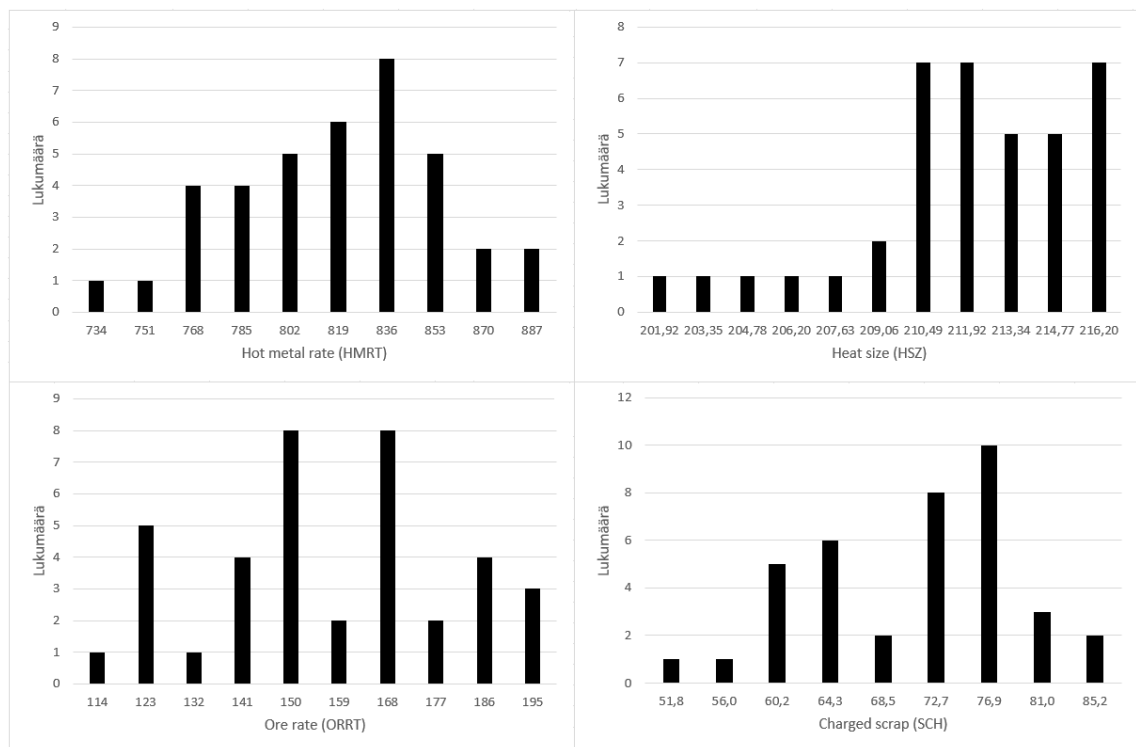


Kuva 3. Panostetun romun sekä raakaraudan välinen riippuvuus.

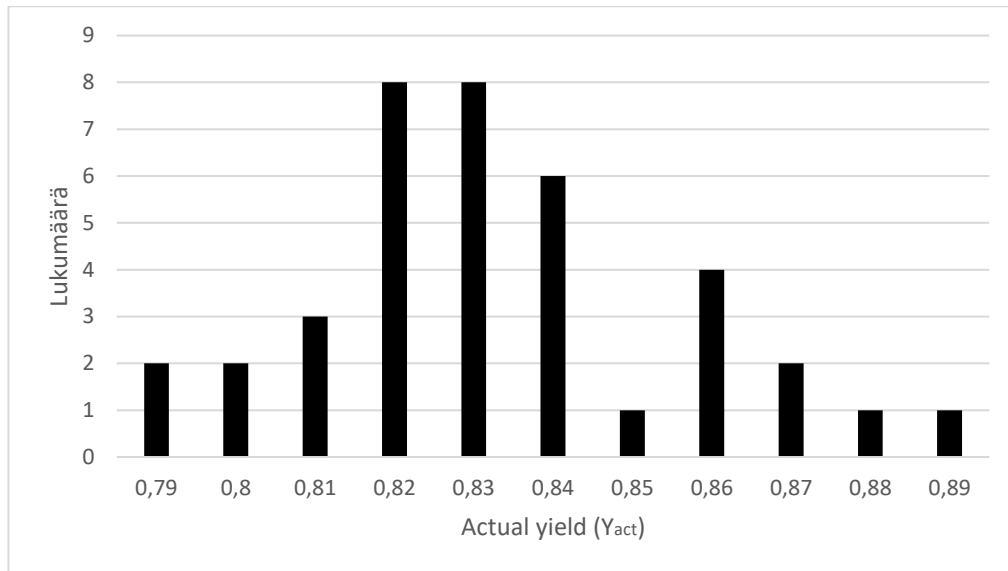


Kuva 4. Teräksen saannon ja panostetun raakaraudan välinen riippuvuus.

Tärkeimmät saantoon vaikuttavat muuttujat ja niiden mittausdatat on esitetty histogrammeina kuvassa 5. Histogrammeista nähdään esimerkiksi se, että miten mittaustulokset ovat jakautuneet. Vastaavasti teräksen mitatut saannot on esitelty histogrammissa kuvassa 6.



Kuva 5. Teräksen saantoon vaikuttavat tärkeimmät muuttujat.



Kuva 6. Teräksen mitattu saanto.

4.2 Lineaarinen regressiomalli

Luodaan regressioanalyysin avulla lineaarinen regressiomalli teräksen saannon ennustamiseksi. Ennustettava muuttuja on teräksen saanto ja selittäviä muuttujia on neljä: panostetun raakaraudan määrä (x_1), sulatuksen kokonaiskoko (x_2), sulatukseen lisätyn rautamalmin määrä (x_3) sekä romun määrä sulatuksessa (x_4). Sovitetaan mitattuun dataan lineaarinen malli, joka on esitetty yhtälössä 1. Käytetään mallina monimuuttujamallia, jonka muodoksi valitaan

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4, \quad (1)$$

missä \hat{y} on mallilla ennustettu saannon arvo eli selitettävä muuttuja,
 b_n ovat malliin sovitetut regressiokertoimet,
 x_1 on panostetun raakaraudan määrä / [kg per tonni terästä],
 x_2 on sulatuksen kokonaiskoko / [tonni per sulatus],
 x_3 on lisätyn rautamalmin määrä / [kg per tonni terästä] ja
 x_4 on panostettu romun määrä / [kg per tonni terästä].

Mallin tuntemattomat parametrit eli regressiokertoimet b_n ratkaistaan pienimmän neliösumman mielessä. Valittu malli voidaan kirjoittaa matriisimuodossa:

$$y = Ab. \quad (2)$$

Yllä y on pystyvektori, joka sisältää lähtömuuttujan mitatut arvot eli teräksen saannon todelliset mittaukset. A on kerroinmatriisi, johon on sijoitettu jokaisen selitettävän muuttujan yksittäisen mittauksen arvot. Vastaavasti pystyvektori b sisältää malliin sovitetut regressiokertoimet. Pystyvektori b voidaan ratkaista pseudoinverssin avulla. Pseudoinverssi on:

$$\hat{b} = (A^T A)^{-1} A^T y. \quad (3)$$

Vektorista \hat{b} saadaan estimaatit malliparametreille, jotka voidaan sijoittaa malliin kertoimien paikalle, jolloin saadaan yhtälö 4.

Lopulliseksi malliksi teräksen saannolle työssä käytetyllä datalla saadaan:

$$\hat{y} = 1,498945 - 0,00072x_1 + 0,000844x_2 - 0,00034x_3 - 0,00308x_4 \quad (4)$$

missä \hat{y} on mallin estimaatti lopulliselle teräksen saannolle / [%].

4.3 Mallin hyvyyden arviointi

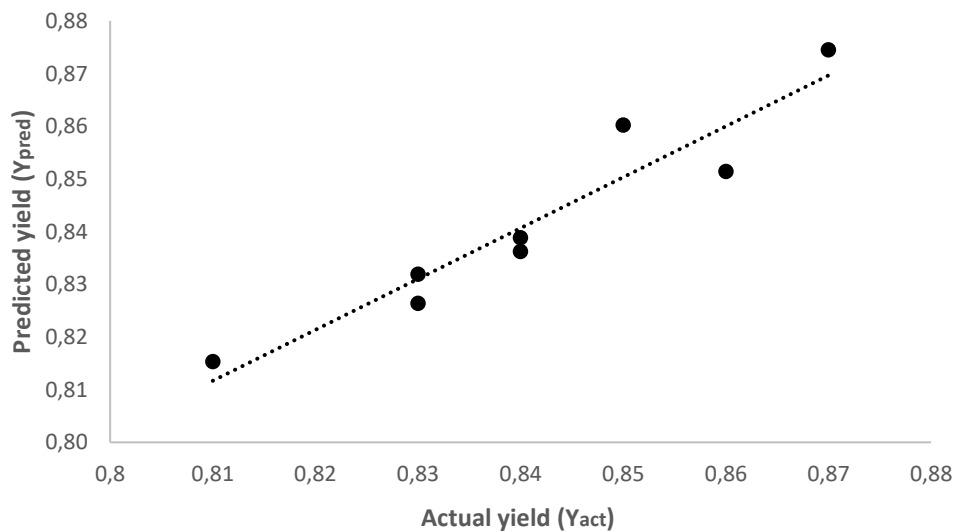
Saadun mallin tarkkuutta voidaan arvioida mallin selitysasteen eli R^2 -arvon avulla. Mallille laskettu selitysaste saa arvon 0,97, joten sen perusteella tarkkuus on hyvä. Kun mallille lasketaan korjattu R^2 -arvo, on se 0,967. Tämä tukee myös mallin tarkkuuden hyvyyttä eikä myöskään ylisovituksen vaaraa tämän perusteella ole.

Testausaineiston data sijoitetaan saatuun mallirakenteeseen ja lasketaan mallin estimaatit lopulliselle teräksen saannolle. Mallin avulla lasketut teräksen saannot Y_{pred} sekä todelliset mitatut saannot Y_{act} on esitetty taulukossa 2. Tuloksista huomataan, että mallin avulla ennustetut teräksen loppusaannot ovat lähes samat, mitä todellisuudessa on mitattu.

Taulukko 2. Mitattu teräksen saanto (Y_{act}) ja mallin ennustama teräksen saanto (Y_{pred}).

Y_{act}	Y_{pred}
0,87	0,875
0,83	0,832
0,84	0,839
0,81	0,815
0,84	0,836
0,86	0,851
0,85	0,860
0,83	0,826

Mallin tarkkuutta voidaan tutkia myös xy-kuvaajalla, jossa ennustetut sekä mitatut saannot esitetään. Kuvaaja on esitetty kuvassa 7. Kuvaajassa x-akselilla on esitetty mitattu saanto ja y-akselilla mallin ennustama saanto. Kuvaajasta nähdään, kuinka pisteisiin sovitettu suora kulkee läpi kuvaajan diagonaalin. Mitatulla ja ennustetulla saannolla on vahvaa lineaarista riippuvuutta. Lineaarinen riippuvuus huomataan myös, kun lasketaan korrelaatiokerroin saantojen välille. Korrelaatiokertoimeksi saadaan 0,95, joka kuvaa vahvaa korrelaatiota saantojen välillä.



Kuva 7. Mallin ennustaman teräksen saannon sekä mitatun saannon välinen riippuvuus.

Mallin tarkkuutta voidaan tarkastella laskemalla suhteelliset virheet (mean absolute percentage error, MAPE) loppusaannoille. Kun suhteelliset virheet lasketaan testausaineistosta todellisten mittausten sekä mallin estimaattien avulla ja otetaan niistä keskiarvo, keskiarvoksi saadaan 0,578 %. Suhteellisten virheiden keskiarvon perusteella ennustettujen arvojen sekä mitattujen arvojen välillä ei ole suurta virhettä, joka

huomataan myös vertailemalla mallilla saatuja saantoja todellisten mitattujen saantojen kanssa.

Tutkitaan artikkelissa luodun neuroverkkomallin hyvyttä myös samoilla tunnusluvuilla. Jos neuroverkkomallilla lasketaan saantoja saman testausaineiston pohjalta, saadaan mallin suhteellisten virheiden keskiarvoksi 0,751 %. Vastaavasti neuroverkolla estimoitujen saantojen ja todellisten saantojen välille laskettu korrelaatiokerroin on 0,93. Tämän perusteella voidaan sanoa, että työssä luotu lineaarinen monimuuttujaregressiomalli on tarkasteltavassa prosessissa tarkempi kuin neuroverkkomalli.

Neuroverkkomallin heikompaa tarkkuutta saannon ennustamiselle voi selittää esimerkiksi se, että käytetyssä validointiaineistossa on käytössä vain kahdeksan eri mittausta ja arvoa muuttujille. Keinotekoinen neuroverkko on myös helposti yliopettavissa, jolloin neuroverkko oppii vain tietyt yksityiskohdat datasta. Tällöin malli ei toimi oikein eikä kykene ennustamaan prosessin todellista saantoa, kun siihen tuodaan uutta dataa.

5 POHDINNAT JA JOHTOPÄÄTÖKSET

Konvertteriprosessin aikana tapahtuu useita monimutkaisia ilmiöitä ja prosessiolosuhteet ovat paikoin erittäin haastavia. Tämän vuoksi datapohjaiset mallinnusmenetelmät ovat tärkeässä osassa konvertteriprosessin ohjauksessa ja teräksen laadun ylläpitämisessä, sillä ilmiöpohjainen mallinnus on haastavaa. Terästehtaiden kehittyneet mittausjärjestelmät ja datan tallentaminen ovat mahdollistaneet datapohjaisten mallien hyödyntämisen teräksen valmistuksessa.

Datapohjaisessa mallinnuksessa on useita tärkeitä vaiheita, jotka vaikuttavat lopullisen mallin tarkkuuteen. Käytössä oleva mittausdata täytyy jakaa kolmeen eri osaan: opetusaineistoon, testausaineistoon sekä validointiaineistoon. Jaettuja aineistoja hyödynnetään mallin luomisen eri vaiheissa. Luodun mallin tarkkuutta on mahdollista arvioida erilaisten kriteerien avulla, kuten esimerkiksi selitysasteen eli R^2 -arvon avulla tai mitatun ja ennustetun suureen välisellä suhteellisella virheellä.

Kappaleessa 3 tutkittujen datapohjaisten mallien perusteella voidaan sanoa, että yleisimmät konvertteriprosessissa ennustettavat lähtömuuttujat ovat puhalluksen jälkeinen teräksen lämpötila ja loppupitoisuudet. Mallia luotaessa on tunnettava tärkeimmät tulomuuttujat, jotka vaikuttavat lähtöarvoon. Tässä työssä esitetyissä malleissa tärkeimmät lähtöarvoon vaikuttavat tulomuuttujat on tutkittu esimerkiksi laskettujen keskinäisinformaatioiden perusteella. Tämän lisäksi tärkeimpiä muuttujia voidaan valita prosessituntemuksen avulla. Suurin osa esitetyistä malleista perustuu regressioanalyysiin.

Tutkittujen datapohjaisten mallien luonnissa (kappale 3) toistuvat samat vaiheet. Ensiksi valitaan sopiva mallirakenne, joka kuvaa mallinnettavaa ilmiötä. Seuraavaksi dataa käsitellään ja se jaetaan osiin. Tämän lisäksi tutkitaan tärkeimmät muuttujat, joilla on vaikutusta ennustettavaan muuttujaan ja otetaan ne mukaan valittuun mallirakenteeseen. Lopulta luodun mallin hyvyttä arvioidaan eri menetelmin.

Mallirakenteen valinnassa on otettava huomioon erilaisia tekijöitä. Joissakin tapauksissa mallin on oltava mahdollisimman yksinkertainen, jotta laskennat voidaan suorittaa riittävän nopeassa ajassa. Toisaalta mallin on oltava riittävän tarkka, jotta sitä voidaan käyttää hyödyksi teräksen laadun ylläpitämisessä. Tämän työn kappaleessa 4 on luotu yksinkertainen lineaarinen regressiomalli. Vaikka malli on yksinkertainen, se ei tarkoita

sitä, etteikö se olisi tarkka. Tämä huomataan, kun työssä luotua mallia verrataan samasta datasta luotuun neuroverkkomalliin. Neuroverkko ei ole paras mahdollinen valinta, sillä datapisteitä on niukasti sen luomiseen, joka selittää neuroverkon heikomman tarkkuuden.

Teräskonvertterin datapohjaisessa mallinnuksessa on myös omat haasteensa. Kyseessä on panosprosessi, jossa valmistettavan teräksen laatu ja ominaisuudet vaihtelevat. Tämän vuoksi prosessissa on paljon muuttujia, eikä kaikkia voida ottaa mallissa huomioon. Lisäksi haasteita tuovat epänormaalit prosessiolosuhteet, joita malli ei kykene tunnistamaan.

6 YHTEENVETO

Työssä on tarkasteltu datapohjaisten mallinnusmenetelmien hyödyntämistä terästeollisuudessa, erityisesti teräskonvertterissa. Tarkasteluun on rajattu vain teräskonvertteri, jonka toimintaperiaatetta sekä merkitystä koko teräksen valmistusketjulle on tarkasteltu käytettyjen datapohjaisten mallien lisäksi. Konvertteriprosessin mallintaminen on haastavaa, sillä prosessin aikana tapahtuu useita samanaikaisia kemiallisia reaktioita sekä korkeaa lämmöntuotantoa. Konvertteri on kuitenkin tärkeä vaihe teräksen laadun ylläpitämiseksi, joten datapohjaisten mallinnusmenetelmien hyödyntäminen on tarpeellista. Datapohjaisten mallien avulla on mahdollista ennustaa esimerkiksi konvertteriprosessin jälkeistä sulan teräksen lämpötilaa sekä hiilipitoisuutta, jotka ovat teräksen valmistuksen tärkeimpien muuttujien joukossa.

Työssä on hyödynnetty eräässä artikkelissa esitettyä mittausdataa terästehtaalta. Dataa ja merkittävimpiä muuttujia on tutkittu eksploraatiivisen data-analyysin avulla ja datan pohjalta on luotu lineaarinen monimuuttujaregressiomalli teräksen loppusaannon ennustamiseksi. Mallin hyvyttä on arvioitu eri tunnuslukujen avulla ja mallia on verrattu myös artikkelissa luotuun keinotekoiseen neuroverkkoon. Molempia malleja on vertailtu esimerkiksi suhteellisen virheen pohjalta. Mallien arvioinnin jälkeen on huomattu, että tässä työssä luotu lineaarinen monimuuttujaregressiomalli on parempi kuin artikkelissa luotu keinotekoinen neuroverkko. Neuroverkon heikompa tarkkuutta voi mahdollisesti selittää esimerkiksi testiaineiston mittausten kohtalaisen vähäinen lukumäärä, sillä mittauksia on vain kahdeksan. Neuroverkon opettamiseen olisi hyvä käyttää useampaa mittausta. Tämän lisäksi neuroverkkoa luodessa on mahdollista yliopettaa neuroverkko, jolloin ennustamisen tarkkuus kärsii. Yliopetettu neuroverkko oppii vain opetusdatan ulkoa, eikä riippuvuuksia lähtömuuttujan ja tulomuuttujan välille.

LÄHDELUETTELO

Barui, S., Mukherjee, S., Srivastava, A., & Chattopadhyay, K. (2019). Understanding Dephosphorization in Basic Oxygen Furnaces (BOFs) Using Data Driven Modeling Techniques. *Metals* 2019, Vol. 9, Page 955, 9(9), 955. <https://doi.org/10.3390/MET9090955>

Cox, I. J., Lewis, R. W., Ransing, R. S., Laszczewski, H., & Berni, G. (2002). Application of neural computing in basic oxygen steelmaking. *Journal of Materials Processing Technology*, 120(1–3), 310–315. [https://doi.org/10.1016/S0924-0136\(01\)01136-0](https://doi.org/10.1016/S0924-0136(01)01136-0)

Ge, Z. (2017). Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemometrics and Intelligent Laboratory Systems*, 171, 16–25. <https://doi.org/10.1016/J.CHEMOLAB.2017.09.021>

Hüsken, R., Fechner, R., & Cappel, J. (2011). Use of hot metal with high phosphorous content in combined blowing BOF converters. *Iron & Steel Technology*, 8(11), 46-58.

Jalkanen, H., & Holappa, L. (2014). Chapter 1.4 - Converter Steelmaking. In S. Seetharaman (Ed.), *Treatise on Process Metallurgy* (pp. 223–270). Elsevier. <https://doi.org/10.1016/B978-0-08-096988-6.00014-6>

Laha, D. (2013). ANN modeling of a steelmaking process. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 8298 LNCS (Issue PART 2)*. https://doi.org/10.1007/978-3-319-03756-1_28

Li, J. N., Li, Y., Liu, F., Bie, S. P., Tang, C. C., & Zhang, X. L. (2013). Modeling of BOF Steelmaking Based on the Data-Driven Approach. *Advanced Materials Research*, 818, 92–97. <https://doi.org/10.4028/WWW.SCIENTIFIC.NET/AMR.818.92>

Runkler, T. A. (2020). *Data Analytics*. <https://doi.org/10.1007/978-3-658-29779-4>

Ruuska, J., Ollila, S., & Leiviskä, K. (2003). Temperature Model for LD-KG Converter. *IFAC Proceedings Volumes*, 36(24), 71–76. [https://doi.org/10.1016/S1474-6670\(17\)37606-1](https://doi.org/10.1016/S1474-6670(17)37606-1)

Sala, D.A., Jalalvand, A., Van Yperen-De Deyne, A., & Mannens, E. (2019). Multivariate Time Series for Data-Driven Endpoint Prediction in the Basic Oxygen Furnace. *Proceedings – 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1419-1426. <https://doi.org/10.1109/ICMLA.2018.00231>

Wang, X., Han, M., & Wang, J. (2010). Applying input variables selection technique on input weighted support vector machine modeling for BOF endpoint prediction. *Engineering Applications of Artificial Intelligence*, 23(6), 1012–1018. <https://doi.org/10.1016/J.ENGAPPAI.2009.12.007>

Wang, X., Xing, J., Dong, J., & Wang, Z. (2017). *Data driven based endpoint carbon content real time prediction for BOF steelmaking*. <https://doi.org/10.23919/ChiCC.2017.8028905>

Wang, Z., Xie, F., Wang, B., Liu, Q., Lu, X., Hu, L., & Cai, F. (2014). The control and prediction of end-point phosphorus content during BOF steelmaking process. *Steel Research International*, 85(4), 599–606. <https://doi.org/10.1002/srin.201300194>