

Primjena strojnog učenja u kemijskom inženjerstvu

Nenad Bolj¹, Željka Ujević Andrijić¹

¹Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije, Trg Marka Marulića 19, 10 000 Zagreb

Sažetak: *Strojno učenje u kemijskom inženjerstvu nalazi primjenu u procjeni teško mjerljivih varijabli stanja i svojstava proizvoda u stvarnom vremenu, optimiranju procesnih uvjeta, analizi mjernih informacija, kao i u prediktivnoj dijagnostici i razvoju digitalnih blizanaca procesa. Primjenjuje se za donošenje brzih i pouzdanih odluka temeljenih na podacima. Digitalizacijom i automatizacijom postrojenja i laboratorija nastaju goleme količine podataka, a metode strojnog učenja izdvajaju ključne informacije i pružaju potporu pri donošenju odluka. U ovom je radu prikazano strojno učenje iz perspektive kemijskog inženjerstva kroz niz istraživačkih i praktičnih primjera koje su proveli suradnici Zavoda za mjerenja i automatsko vođenje procesa Sveučilišta u Zagrebu Fakulteta kemijskog inženjerstva i tehnologije. Prikazani su primjeri softverskih senzora za kontinuiranu procjenu svojstava proizvoda, modeli za praćenje učinkovitosti izmjenjivača topline, modeli za procjenu koncentracije u procesu kristalizacije farmaceutskih djelatnih tvari primjenom procesne analitičke tehnologije, kao i procjenu raspodjele veličine čestica. Dan je i osvrt na razvoj klasifikacijskog modela temeljenog na molekulskim deskriptorima za predviđanje nastanka kokristala. Također je prikazan primjer predviđanja koncentracije lebdećih čestica i dušikovih oksida u zraku.*

Prikazane studije pokazuju kako strojno učenje, u kombinaciji s domenskim znanjem, postaje vrijedan alat za optimiranje tehnoloških procesa, smanjenje troškova i ubrzanje razvojnog ciklusa u kemijskom inženjerstvu.

Ključne riječi: *strojno učenje, kemijsko inženjerstvo, softverski senzori, procesna analitička tehnologija*

1. Uvod

Digitalizacija industrijskih procesa i laboratorija, kao i razvoj metoda umjetne inteligencije posljednjih su godina značajno proširili mogućnosti analize, nadzora i vođenja složenih procesa u kemijskom inženjerstvu. Suvremena procesna postrojenja i laboratorijski sustavi stvaraju goleme količine podataka, no sama dostupnost podataka ne

znači automatski i korisne informacije. Upravo zato raste važnost modeliranja, statističke analize podataka i strojnog učenja, koje donose zaključke iz mjerenja, procjenjuju teško mjerljive, a bitne varijable, te donose pouzdanije odluke u realnom vremenu.

Postoji više razloga zbog kojih bi kemijski inženjeri i istraživači trebali u svom radu razumjeti i primjenjivati metode strojnog učenja. Tijekom posljednja dva desetljeća, a osobito u posljednjem, strojno učenje nalazi široku primjenu u kemijskom inženjerstvu, posebice u optimizaciji i naprednom vođenju procesa, analizi velikih skupova podataka iz procesnih postrojenja i laboratorija, preventivnoj i prediktivnoj dijagnostici procesa i procesne opreme, modeliranju kemijskih reakcija, predviđanju svojstava materijala, kao i upravljanju kvalitetom proizvoda. Metode strojnog učenja pružaju alate za učinkovitu analizu mjernih podataka, prepoznavanje složenih obrazaca i izdvajanje vrijednih informacija koje nije jednostavno uočiti klasičnim pristupom.

Kemijsko inženjerstvo po svojoj je prirodi multidisciplinarno područje stoga zahtijeva suradnju s drugim poljima znanosti, osobito s računalnim znanostima. Poznavanje metoda strojnog učenja doprinosi povezivanju domenskog znanja s naprednim računalnim alatima. Prednost kemijskog inženjera u odnosu na stručnjaka računalnog usmjerenja u ovom kontekstu jest sposobnost postavljanja i rješavanja zadataka, kritičke analize podataka i smislenog tumačenja dobivenih rezultata u skladu s fizikalno-kemijskim zakonitostima.

Posebna vrijednost kemijskog inženjerskog pristupa jest u tome što gotovo uvijek ima praktičnu primjenu. Zbog toga umjetna inteligencija pridonosi razvoju rješenja koja se izravno primjenjuju na konkretne zadatke, bilo u ranom razvoju u laboratoriju, pilot postrojenju, bilo u primjeni na postrojenju.

Industrijski kemijske procese najčešće karakterizira veliki broj međusobno povezanih varijabli, nelinearni odnosi među njima, vremenska promjenjivost i dinamičke promjene. Zbog toga je razvoj pouzdanih modela koji se temelje isključivo na fundamentalnim zakonima veoma zahtjevan, a u pojedinim slučajevima i nedostatan za dobar opis stvarnog vladanja postrojenja, ponajprije zbog složenosti procesa i ograničene dostupnosti procesnih parametara. S druge strane, pristupi modeliranju temeljeni na podacima omogućuju razvoj empirijskih modela koji opisuju vladanje procesa na temelju reprezentativnih i kvalitetno obrađenih skupova podataka.

U mnogim slučajevima ključne procesne varijable ili svojstva proizvoda nije moguće kontinuirano mjeriti u realnom vremenu. Laboratorijske analize često su spore i diskontinuirane, dok su *on-line* procesni analizatori skupi, zahtjevni za održavanje i podložni kvarovima. Zbog toga se razvijaju softverski senzori za procjenu varijabli stanja. Softverski senzori mogu se temeljiti na fundamentalnim modelima ili, češće, empirijskim, na podacima na kojima se primjenjuju metode identifikacije procesa [1, 2]. Paralelno s time, metode strojnog učenja nalaze sve širu primjenu i u drugim

područjima kemijskog inženjerstva, primjerice u nadzoru procesa, procjeni kvalitete proizvoda, dijagnostici procesne opreme [3, 4], optimiranju radnih uvjeta, analizi spektroskopskih podataka, praćenju procesa [5, 6] i razvoju novih farmaceutskih formulacija i spojeva [7, 8], kao i zaštiti okoliša [9, 10]. Dodatni zamah dao je razvoj procesne analitičke tehnologije (PAT-a), naprednih senzorskih sustava, dostupnost komercijalnih i otvorenih sofisticiranih programskih alata za obradu podataka, kao i računalnih metoda za razvoj i vrednovanje modela.

Područje kemijskog inženjerstva kao i ostala tehnička područja već neko vrijeme prolazi kroz značajnu transformaciju potaknutu integracijom umjetne inteligencije (UI). Noviji pregledni radovi ističu promjenu paradigme od tradicionalnih pristupa temeljenih na pravilima i modelima zasnovanim na teorijskim načelima prema pristupima usmjerenima na podatke i modelima koji uče iz podataka [11, 12]. Ta je promjena osobito vidljiva u području razvoja novih materijala i molekula, pri čemu modeli umjetne inteligencije povezuju strukture i svojstva te ubrzavaju postupke probira, čime se značajno skraćuje vrijeme razvoja [13, 14]. Nadalje, primjena UI olakšava optimiranje procesa u stvarnom vremenu, prediktivno održavanje procesne opreme i razvoj digitalnih blizanaca dijelova ili cjelokupnih postrojenja [14, 15].

Osim povećanja učinkovitosti procesa, metode strojnog učenja sve se više primjenjuju i na složene proizvodne sustave, uključujući zelene procese poput valorizacije biomase i optimiranja sustava kemijskog kruženja (engl. *chemical looping*) [16, 17]. Također njihova se primjena sve više usmjerava prema načelima zelene kemije, s ciljem razvoja održivijih proizvodnih procesa optimiranjem iskorištenja resursa i racionalizacijom potrošnje energije [17].

Unatoč napretku i dalje postoje brojni izazovi. Stručnjaci iz industrije posebno ističu potrebu za reprezentativnim podacima iz različitih radnih uvjeta, većom interpretabilnošću modela i učinkovitim upravljanjem životnog ciklusa (razvoj, validacija, primjena i održavanje). To uključuje kontinuirano praćenje točnosti i pouzdanosti modela, njegovo ažuriranje s novim podacima i prilagodbu promjenama radnih uvjeta, kako bi se osigurala pouzdanost i sigurnost u industrijskom okruženju [14, 16]. Dodatna ograničenja su poteškoće u generalizaciji modela na različite radne uvjete, izazovi integracije modela u postojeće sustave za vođenje procesa, kao i regulatorni zahtjevi i sigurnost podataka. Osobito je to izraženo u farmaceutskoj proizvodnji u kojoj je primjena softverskih alata dodatno ograničena zahtjevima validacije sustava. U daljnjem razvoju ovog područja očekuje se da će razvoj hibridnih modela koji povezuju pristupe temeljene na fizikalnim zakonitostima i dubokom učenju, pomoći premostiti jaz između algoritama “crne kutije” i strogih zahtjeva kemijskog procesnog inženjerstva [15, 17]. Kako bi se povećalo povjerenje stručnjaka u primjenu modela, očekuje se i veća “objašnjivost” modela strojnog učenja primjenom metoda objašnjive umjetne inteligencije poput SHAP (engl. *SHapley Additive exPlanations*), LIME (engl. *Local Interpretable Model-agnostic Explanations*), analize važnosti značajki modela [18, 19].

U nastavku ovog rada dan je pregled metoda strojnog učenja koje su autori primijenili na vlastitim istraživačkim i suradničkim projektima na više područja istraživanja:

- razvoj softverskih senzora za procjenu teško mjerljivih procesnih varijabli,
- prediktivna dijagnostika industrijskih izmjenjivača topline,
- primjena procesne analitičke tehnologije u procesu kristalizacije,
- predviđanje raspodjele veličine čestica i koncentracije u suspenzijama,
- primjena klasifikacijskih modela u predviđanju nastanka kokristala i
- modeliranje koncentracije onečišćujućih tvari u zraku.

Svrha je prikazati primjenu tih metoda kroz primjere iz istraživačke i industrijske prakse kako bi se pokazalo da integracija procesnog znanja, eksperimentalnih podataka i metoda strojnog učenja doprinosi boljem razumijevanju, vođenju i optimiranju kemijskih procesa.

2. (Pred)obrada podataka, modeli strojnog učenja i softverski alati

2.1 Predobrada podataka

Razvoj modela strojnog učenja u kemijskom inženjerstvu uvelike ovisi o kvaliteti i reprezentativnosti dostupnih podataka. Budući da se modeli temelje na mjernim podacima dobivenim laboratorijskim eksperimentima ili iz industrijskih postrojenja, prvi korak u razvoju empirijskih modela predstavlja kritička analiza i predobrada podataka. Cilj predobrade jest uklanjanje pogrešaka i neželjenih utjecaja iz mjernih signala te priprema podataka u obliku prikladnom za razvoj modela.

Jedan od prvih koraka u analizi podataka jest primjena deskriptivne statistike radi osnovnog uvida u distribuciju i karakteristike mjerenih veličina. Grafički prikazi poput histograma, *box-plot* dijagrama i dijagrama raspršenja prikladne su za vizualni pregled distribucije podataka, identifikaciju odstupanja i analizu korelacija između varijabli.

U bazama procesnih podataka česte nedostaju neke vrijednosti (engl. *missing values*) zbog kvara senzora, problema u prijenosu signala ili održavanja mjerne opreme. Takve vrijednosti potrebno je za daljnju obradu podataka nadomjestiti, primjerice interpoliranim vrijednostima. Dodatni koraci obuhvaćaju uklanjanje šuma, trendova i klizanja karakteristike senzora primjenom metoda filtriranja i izgladivanja, primjerice filtara pomičnog prosjeka, Savitzky-Golay ili LOESS filtra [2].

Posebnu pozornost potrebno je posvetiti otkrivanju (uočavanju) ekstremnih vrijednosti (engl. *outliers*) koje nastaju zbog pogrešaka u mjerenju, kvara mjernog pretvornika, poremećaja u radu procesa ili odstupanja od uobičajenih radnih uvjeta. Za identifikaciju često se primjenjuje pravilo 3σ , analiza reziduala i metode temeljene na multivarijatnim udaljenostima.

Prije razvoja modela provodi se i skaliranje podataka jer procesne varijable mogu imati različite redove veličina i fizikalne jedinice. U tu svrhu primjenjuju se postupci normalizacije ili standardizacije. Važno je naglasiti da se predobrada podataka, odabir značajki i razvoj modela treba provoditi isključivo na skupu podataka za treniranje kako bi se izbjegao problem “curenja” podataka (engl. *data leakage*).

Predobrada spektralnih podataka

Kod procesne analitičke tehnologije (engl. *process analytical technology*, PAT) za praćenje kemijskog sastava i svojstava u stvarnom vremenu u primjeni su metode poput Ramanove i infracrvene spektroskopije. Spektralni podaci dobiveni tim metodama često sadrže različite neželjene doprinose signalu koji mogu otežati interpretaciju i smanjiti pouzdanost modela. Cilj predobrade spektralnih podataka jest uklanjanje ili smanjenje utjecaja pojava poput fluorescencije, raspršenja svjetlosti, šuma detektora, pogrešaka umjeravanja, kozmičkih zraka i fluktuacija snage izvora zračenja. Budući da i male razlike u spektrima mogu sadržavati važne informacije, potrebno je izbjeći pretjeranu obradu podataka kako ne bi došlo do izobličenja spektra ili gubitka relevantnih informacija.

Uobičajene metode predobrade obuhvaćaju korekciju raspršenosti i derivacijske metode [20]. U korekciju raspršenosti ubrajaju se multiplikativna korekcija signala, standardna normalna varijacija, normalizacija spektra i korekcija bazne linije. Derivacijske metode primjenjuju se za razdvajanje preklapljenih vrpca i smanjenje utjecaja promjena bazne linije, a uobičajeno se kombiniraju s postupcima izgladivanja, poput Savitzky-Golay filtriranja. Postupke predobrade potrebno je prilagoditi spektroskopskoj metodi i karakteristikama analiziranog sustava.

2.2 Modeli strojnog učenja

Razvoj metoda strojnog učenja donio je primjenu naprednih računalnih modela u analizi i optimiranju kemijskih procesa. U kemijskom inženjerstvu metode strojnog učenja najčešće se koriste za razvoj softverskih senzora, analizu spektroskopskih podataka, optimiranje procesa, predviđanje svojstava materijala te modeliranju u području zaštite okoliša. Ovisno o prirodi zadatka, metode strojnog učenja dijele se na metode regresije, klasifikacije i modele za analizu vremenskih nizova. Pregled standardnih metoda dan je u tablici 1.

Tablica 1: Pregled standardnih metoda strojnog učenja u kemijskom inženjerstvu

Skupina metoda	Algoritmi	Tip problema	Tipične primjene u kemijskom inženjerstvu
Linearni i multivarijantni modeli	MLR, PCA, PLS	regresija	kemometrija, analiza spektara, kalibracijski modeli u PAT sustavima, softverski senzori
Metode stabala odlučivanja	stabla odlučivanja (engl. <i>decision tree</i>), slučajna (nasumična) šuma (engl. <i>random forest</i>), XGBoost	regresija, klasifikacija	predviđanje kvalitete proizvoda, procesni podaci, analiza okolišnih podataka (npr. koncentracije onečišćujućih tvari), optimiranje procesa
Metode potpornih vektora	SVM, SVR	klasifikacija, regresija	klasifikacija kemijskih spojeva i materijala, analiza spektroskopskih podataka, predviđanje fizikalno-kemijskih svojstava tvari
Umjetne neuronske mreže	MLP, konvolucijske mreže (CNN)	regresija, klasifikacija	modeliranje nelinearnih kemijskih procesa, analiza spektroskopskih i slikovnih podataka, predviđanje svojstava materijala
Rekurentne mreže	RNN, LSTM, GRU	vremenski nizovi	dinamika procesa, predviđanje vladanja procesa, analiza vremenskih nizova iz postrojenja, praćenja onečišćivala u okolišu
Ansambel metode	slučajna šuma, <i>gradient boosting</i>	regresija, klasifikacija	prediktivno modeliranje složenih procesa, analiza velikih skupova podataka, razvoj robusnih modela za optimiranje i nadzor procesa

Metode linearne i multivarijantne regresije

Jedan od standardnih pristupa u kemijskom inženjerstvu je multivarijantna regresijska analiza koja (funkcijski) opisuje odnose između više ulaznih varijabli i jednog ili više izlaza sustava. Ove metode posebno su važne u analizi spektroskopskih podataka i u razvoju jednostavnijih struktura softverskih senzora.

Između ostalih metoda ističu se:

- višestruka linearna regresija (engl. *multilinear regression*, MLR)
- regresija parcijalnih najmanjih kvadrata (engl. *partial least squares*, PLS)
- analiza glavnih komponenti (engl. *principal component analysis*, PCA)

PCA metoda primjenjuje se za smanjenje dimenzionalnosti podataka i otkrivanje skrivenih obrazaca u velikim skupovima podataka, dok se PLS-om modelira odnos između ulaznih i izlaznih varijabli, čak i kada su ulazne varijable međusobno značajno korelirane. Zbog toga se ove metode široko primjenjuju u kemometriji i analizi spektroskopskih podataka.

Metode temeljene na stablima odlučivanja

Metode temeljene na stablima odlučivanja predstavljaju skup algoritama za regresijske i klasifikacijske zadatke. Stabla odlučivanja temelje se na hijerarhijskoj strukturi čvorova u kojoj se skup podataka postupno dijeli prema vrijednostima ulaznih varijabli kako bi se dobile što homogenije skupine podataka s obzirom na ciljnu varijablu. Prednost im je relativno jednostavna interpretacija rezultata i rad s nelinearnim odnosima između varijabli.

U ovoj skupini ističu se: stabla odlučivanja, slučajna (nasumična) šuma, *gradient boosting* i *extreme gradient boosting* (tzv. XGBoost).

Metoda slučajne šume i XGBoost posebno su popularne u kemijskom inženjerstvu zbog visoke pouzdanosti pri predviđanju i otpornosti na smetnje u podacima [21]. Te metode primjenjuju se u predviđanju kvalitete proizvoda, modeliranju podataka koncentracija onečišćujućih tvari u okolišu (npr. plinova i lebdećih čestica) i optimiranju industrijskih procesa.

Metode potpornih vektora

Metoda potpornih vektora (engl. *support vector machines*, SVM) razvijena je za rješavanje klasifikacijskih i regresijskih problema u složenim i nelinearnim sustavima. Ova metoda traži optimalnu granicu razdvajanja između različitih klasa podataka ili funkciju koja najbolje opisuje odnos između ulaznih i izlaznih varijabli. Pri tome se nelinearni odnosi modeliraju primjenom različitih *kernel* funkcija. U kemijskom inženjerstvu SVM se često primjenjuje za klasifikaciju kemijskih spojeva, predviđanje svojstava materijala, analizu spektroskopskih podataka i razvoj softverskih senzora.

Umjetne neuronske mreže

Umjetne neuronske mreže najučestalija je metoda strojnog učenja u kemijskom inženjerstvu. Temelje se na međusobno povezanim umjetnim neuronima organiziranim u slojeve koji tijekom procesa učenja prilagođavaju težinske koeficijente veza kako bi aproksimirali odnos između ulaznih i izlaznih podataka. Služe za modeliranje složenih nelinearnih odnosa između procesnih varijabli. Primjenjuje se za modeliranje kemijskih reakcija, predviđanje svojstava i kvalitete proizvoda, optimiranje procesa i analizu spektroskopskih podataka.

Uobičajene mrežne arhitekture su:

- višeslojne perceptronske (MLP) mreže
- konvolucijske neuronske mreže (CNN)
- rekurentne neuronske mreže (RNN)
- *Long Short-Term Memory* (LSTM) i *Gated Recurrent Unit* (GRU) mreže
- transformerske neuronske mreže (engl. *Transformers*).

Rekurentne neuronske mreže kao i LSTM mreže [22] posebno su pogodne za analizu vremenskih nizova i modeliranje dinamičkog vladanja procesa jer uzimaju u obzir vremensku ovisnost podataka. Transformerske arhitekture temelje se na mehanizmu pažnje

(engl. *attention*), a primjenjuju se u analizi sekvencijskih podataka, modeliranju molekularnih struktura i predviđanju fizikalno-kemijskih svojstava molekula i materijala.

Ansambl metode i duboko učenje

U novijim istraživanjima sve se češće primjenjuju ansambl metode koje kombiniraju rezultate više pojedinačnih modela kako bi se postigla veća točnost i robusnost predviđanja. Takvi pristupi uključuju metode poput slučajne šume i *gradient boostinga* ([23]).

Paralelno s time, razvoj dubokog učenja (neuronskih mreža s većim brojem skrivenih slojeva) omogućuje primjenu kompleksnih modela koji automatski uče značajke iz velikih skupova podataka. Za razliku od tradicionalnih metoda strojnog učenja, duboke neuronske mreže automatski izdvajaju relevantne značajke iz sirovih podataka bez potrebe za opsežnim ručnim odabirom ulaznih varijabli. Duboke neuronske mreže danas se primjenjuju u: analizi spektroskopskih i slikovnih podataka, modeliranju molekularnih struktura, predviđanju svojstava materijala i razvoju digitalnih blizanaca kemijskih procesa.

2.3 Softverski alati za primjenu strojnog učenja u kemijskom inženjerstvu

Primjena strojnog učenja uvelike je potaknuta razvojem računalnih alata za obradu velikih skupova podataka, modeliranje, kao i integraciju modela u sustave za nadzor i vođenje procesa. U praksi je na raspolaganju velik broj softverskih alata.

U području analize podataka i strojnog učenja ističe se Python kojim se, zahvaljujući brojnim bibliotekama, razvijaju različiti modeli strojnog i dubokog učenja. Ističu se biblioteke NumPy i Pandas za obradu podataka, Scikit-learn za strojno učenje, kao i TensorFlow i PyTorch za razvoj dubokih neuronskih mreža.

U kemijskom i procesnom inženjerstvu široko se primjenjuje i MATLAB, koji nudi brojne alate za numeričku analizu, identifikaciju procesa i razvoj modela strojnog učenja. Posebno su važni *Statistics and Machine Learning Toolbox* i *Deep Learning Toolbox*.

Za statističku analizu i obradu podataka postoji programski jezik R koji sadrži velik broj paketa za statističko modeliranje, vizualizaciju podataka i primjenu metoda strojnog učenja.

Osim programskih jezika, postoje i platforme za analizu podataka s grafičkim sučeljem, poput KNIME, RapidMiner, Quasar i Orange, u kojima se programira vizualno. U kemijskoj industriji i procesnoj analitici važnu ulogu imaju i specijalizirani kemometrijski softveri poput SIMCA, Unscrambler i Pirouette, koji se primjenjuju za multivarijantnu analizu.

Razvoj i dostupnost ovih alata značajno su doprinijeli širenju primjene strojnog učenja u kemijskom inženjerstvu te omogućili integraciju naprednih analitičkih metoda u istraživačke i industrijske procese.

3. Primjeri primjene modela strojnog učenja

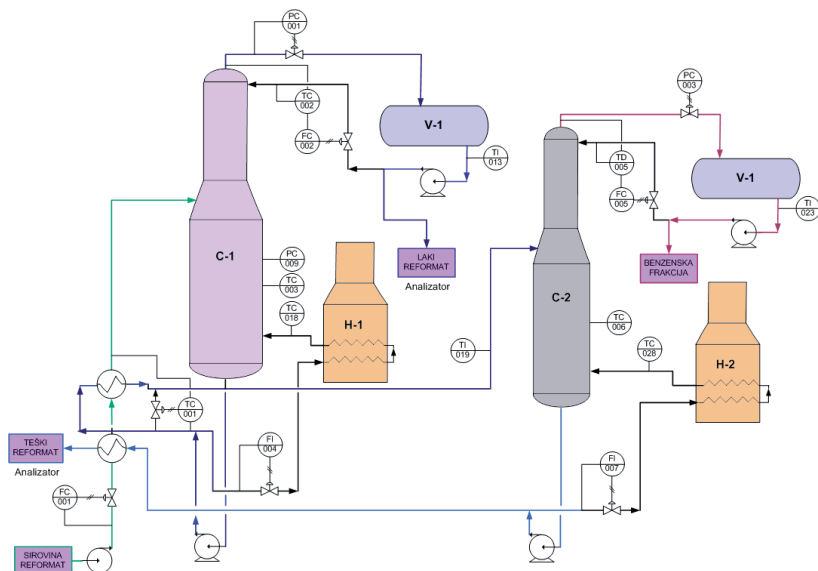
3.1 Softverski senzori u kemijskoj procesnoj industriji

Softverski senzori predstavljaju matematičke modele s kojima se procjenjuju procesne varijable ili svojstva proizvoda koja se ne mogu izravno mjeriti ili čije je mjerenje otežano. Temelje se na dostupnim procesnim mjerenjima poput temperature, tlaka ili protoka. U industrijskoj praksi softverski senzori služe kao zamjena za skupe ili nepouzdanе analitičke instrumente ili kao pomoćni sustavi za procjenu kvalitete proizvoda u realnom vremenu.

Softverski senzori se razvijaju linearnim i nelinearnim modelima (engl. *system identification*), kao i metodama strojnog učenja. U procesnoj industriji primjenjuju se dinamičke linearne i nelinearne strukture parametarskih modela poput FIR, ARX, OE i Hammerstein-Wiener modela za opisivanje dinamičkog vladanja procesa.

3.1.1 Primjer: Procjena sadržaja benzena u lakom reformatu

Jedan od primjera softverskih senzora razvijenih na Zavodu za mjerenja i automatsko vođenje procesa odnosi se na procjenu sadržaja benzena u katalitičkom reformatu u rafinerijskoj proizvodnji[24]. Proces katalitičkog reformiranja obuhvaća frakcioniranje reformata u sustavu kolona (slika 1), pri čemu se benzen koncentrira u lakom reformatu koji se dalje obrađuje kako bi se zadovoljili zahtjevi za kvalitetom goriva. Kontinuirano praćenje sadržaja benzena u reformatu važno je zbog ekoloških regulativa i zahtjeva za kvalitetom goriva. Međutim, *on-line* analizatori za mjerenje sastava benzena često nisu dostupni ili zahtijevaju složeno i učestalo održavanje, a laboratorijske analize su prerijetke s obzirom na moguće promjene u procesu. U tom slučaju razvijen je softverski senzor temeljen na procesnim podacima dobivenim iz distribuiranog sustava za vođenje procesa (DCS).



Slika 1: Pojednostavljena shema frakcioniranja reformata i procesne varijable korištene za razvoj softverskog senzora [24]

Modeli su razvijeni primjenom linearnih i nelinearnih metoda identifikacije procesa, uključujući FIR, *Output Error* (OE) i Hammerstein-Wiener (HW) modele u programskom jeziku Matlab. FIR i OE modeli pripadaju skupini linearnih dinamičkih modela koji opisuju odnos između ulaznih i izlaznih varijabli procesa, dok HW model predstavlja nelinearnu strukturu koja kombinira linearni dinamički blok s nelinearnim transformacijama ulaza i izlaza. Ulazne varijable su kontinuirano mjerene procesne veličine poput temperature, ulaznog protoka u kolonu, temperature dna kolone, temperature u koloni, tlaka u koloni i protoka refluksa u frakcionacijsku kolonu. Podaci su prikupljeni tijekom nekoliko tjedana neprekidnog rada postrojenja, uz odgovarajuću predobradu podataka, filtriranje i uklanjanje ekstremnih vrijednosti. Modeli su optimirani primjenom genetičkih algoritama kako bi se odredili optimalni redovi i parametri modela. Usporedba razvijenih modela pokazala je da je OE model ostvario najveću točnost predviđanja ($FIT \approx 90\%$) na validacijskom skupu podataka, dok je HW model pokazao nešto slabije, ali zadovoljavajuće rezultate ($FIT \approx 88\%$), a FIR model, kao strukturno najjednostavniji model, postigao je najnižu točnost ($FIT \approx 79\%$).

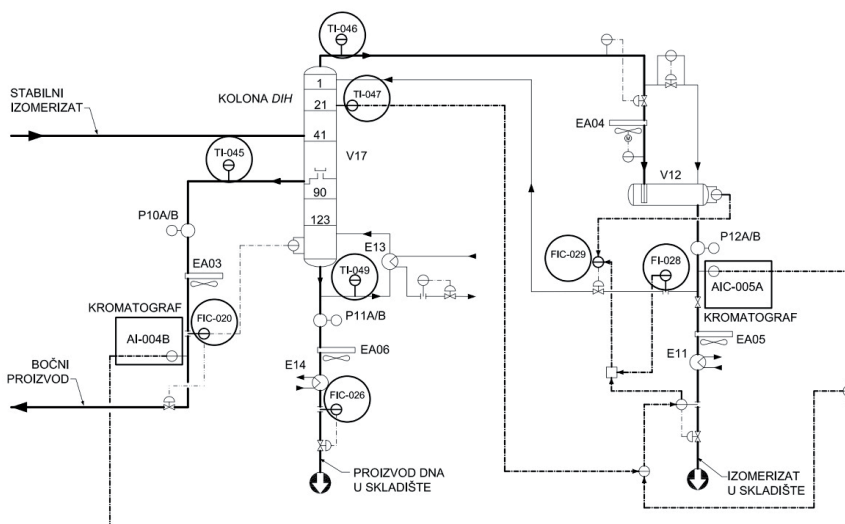
3.1.2 Primjer procjene sadržaja sumpora u procesu hidrodosulfurizacije

Sličan pristup modeliranju može se primijeniti i u procesu hidrodosulfurizacije. U ovom procesu se kontinuirano prati sadržaj sumpora u proizvodu, što je iznimno važno zbog zahtjeva za kvalitetom goriva i zaštitom okoliša. Budući da su laboratorijske analize rijetke, a procesni analizatori mogu biti u kvaru ili vremenski zahtjevni, razvi-

jaju se softverski senzori za procjenu sadržaja sumpora na temelju kontinuirano mje-
renih procesnih varijabli kao što su temperatura i tlak reaktora, ulazni protok, omjer
 H_2 / nafta, temperatura nakon izmjenjivača i tlak separatora [25]. U istraživanju [25]
razvijeni su modeli temeljeni na metodama identifikacije procesa (FIR, ARX, OE,
NARX i HW) i modeli strojnog učenja poput umjetnih neuronskih mreža, LSTM i
GRU mreža. Usporedba modela pokazala je da su modeli temeljeni na neuronskim
mrežama postigli najveću točnost predviđanja sadržaja sumpora, dok su OE i HW
modeli pokazali vrlo dobre rezultate uz jednostavniju i interpretabilniju strukturu.

3.1.3 Primjer procjene kvalitete proizvoda u procesu izomerizacije

U radu Hercega i sur. [26] razvijeni su modeli za procjenu sastava proizvoda u procesu
izomerizacije primjenom metoda strojnog učenja. Na slici 2 prikazana je sekcija deizo-
heksanizacijske kolone u kojoj se odvaja izomerat čiji sastav značajno utječe na oktanski
broj benzina. U razvijenim modelima izlazna varijabla bila je procesni parametar
povezan s oktanskim brojem benzina izražen udjelom spojeva 2,3-dimetilbutana (2,3-
DMB) i 3-metilpentana (3-MP). Poseban naglasak stavljen je na primjenu dinamičkih
neuronskih mreža LSTM arhitekture za modeliranje vremenski ovisnih procesa.



Slika 2: Sekcija deizoheksanizacijske kolone u procesu izomerizacije [26]

Rezultati su uspoređeni s drugim pristupima modeliranju temeljenim na podacima,
uključujući višeslojni perceptron, metode potpornih vektora i dinamičke polinomske
modele. Dobiveni rezultati pokazali su da LSTM mreže mogu uspješno opisati dina-
mičko vladanje industrijskih procesa za procjenu teško mjerljivih svojstava proizvo-
da u realnom vremenu.

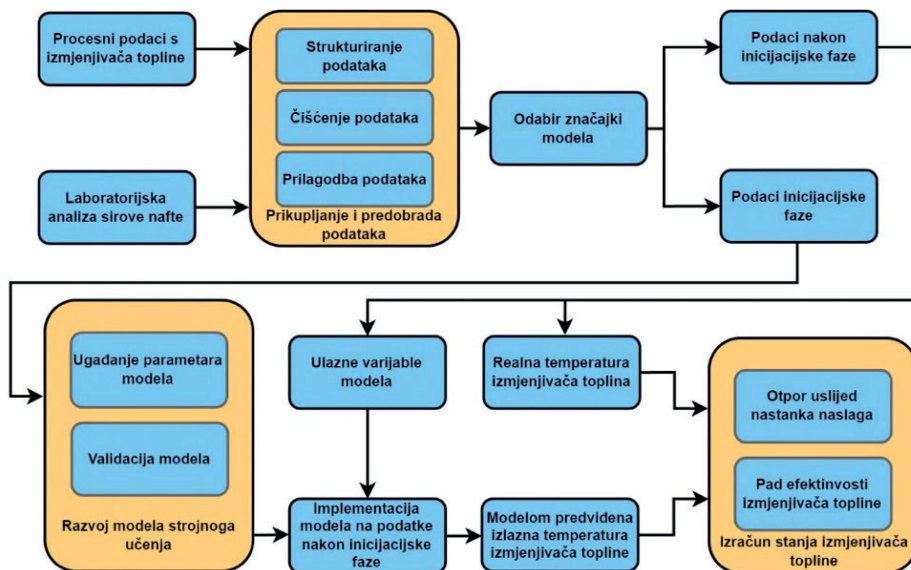
3.2 Praćenje učinkovitosti izmjenjivača topline

Učinkovitost industrijskih izmjenjivača topline s vremenom pada zbog taloženja naslaga na površinama za prijenos topline (engl. *fouling*), promjena radnih uvjeta ili degradacije opreme. Smanjenje učinkovitosti izmjenjivača topline izravno utječe na potrošnju energije, ekonomičnost procesa i emisije štetnih plinova u okoliš. Praćenje učinkovitosti izmjenjivača topline temelji se na analizi toplinskih parametara kao što su koeficijent prijenosa topline, temperaturne razlike i protoci radnih fluida. Razvojem modela temeljenih na procesnim podacima kontinuirano se procjenjuje stanje izmjenjivača topline u realnom vremenu i pravovremeno detektiraju promjene.

Jedan od primjera je razvoj modela za detekciju nastajanja naslaga u izmjenjivačima topline u rafinerijskom sustavu za predgrijavanje nafte u sklopu postrojenja za preradu sirove nafte [27]. U radu su uzeti podaci iz industrijskog postrojenja procesnih varijabli poput ulaznih i izlaznih temperatura toplog i hladnog toka, protoka fluida i drugih parametara vezanih za izmjenu topline.

Modeli su razvijeni primjenom LSTM neuronske mreže i XGBoost algoritma. Uspoređeni su s poluempirijskim pristupom temeljenim na ϵ -NTU metodi. Modeli su uvježbani na podacima iz početne faze rada izmjenjivača kada se sustav smatra čistim. Cilj je bio predvidjeti izlazne temperature izmjenjivača topline u tom stanju (slika 3).

Kako se tijekom rada na površini za prijenos topline postupno razvijaju naslage, pojavljuje se odstupanje između predviđenih i stvarno izmjerenih izlaznih temperatura. Ta odstupanja temelj su za procjenu koeficijenta prolaza topline i izračun otpora naslaga (engl. *fouling resistance*). Na taj način kontinuirano se prati učinkovitost izmjenjivača topline i pravovremeno detektira *fouling*. U radu je također analiziran utjecaj sastava i fizikalno-kemijskih svojstava sirove nafte na nastajanje *foulinga*. Provedena je detaljna analiza značajki (engl. *feature analysis*) i interpretacija modela za bolje razumijevanje doprinosa pojedinih procesnih varijabli i fizikalno-kemijskih parametara pojavi naslaga u izmjenjivačima topline.



Slika 3: Postupak procjene otpora uslijed nastanka naslaga i učinkovitosti izmjenjivača topline primjenom LSTM i XGBoost modela [27]

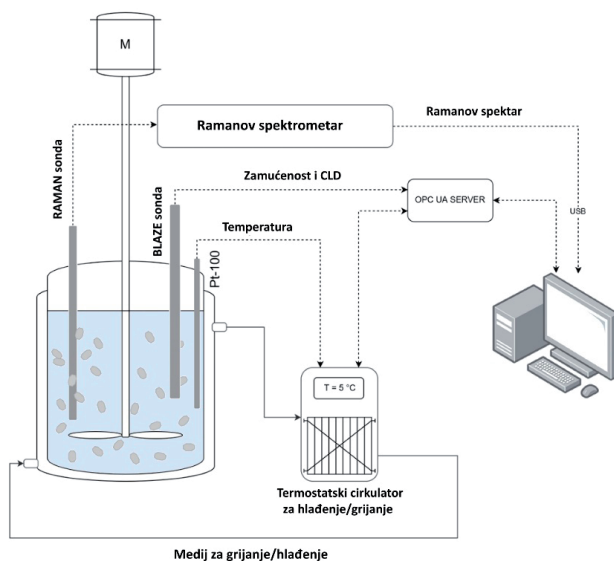
U radu Ujević Andrijić i sur. [28] razvijen je sličan sustav za detekciju nastajanja naslaga u industrijskom izmjenjivaču topline u rafinerijskom postrojenju hidrokrekiranja primjenom modela temeljenih na podacima i poluempirijskih pristupa. Stanje izmjenjivača procjenjuje se modelom temeljenom na metodi broja prijenosnih jedinica (ϵ -NTU), umjetnom neuronskom mrežom i nelinearnim dinamičkim modelima tipa NFIR. Modeli su razvijeni na temelju kontinuirano mjerenih temperatura i protoka radnih fluida. Predviđaju izlazne temperature izmjenjivača topline. Odstupanje između predviđenih i stvarnih izlaznih temperatura indikator je smanjenja učinkovitosti prijenosa topline i razvoja naslaga na površinama za prijenos topline. Rezultati su pokazali da neuronske mreže kao i NFIR modeli mogu pouzdano detektirati nastanak naslaga i kvantificirati promjene u učinkovitosti izmjenjivača topline.

3.3 Procesna analitička tehnologija u kristalizacijskim procesima

Kristalizacija je jedan od ključnih procesa u farmaceutskoj industriji jer određuje fizikalna svojstva proizvoda poput veličine čestica, polimorfnog oblika i topljivosti. Kontinuirano praćenje kristalizacijskog procesa ostvaruje se primjenom alata procesne analitičke tehnologije (PAT). PAT obuhvaća primjenu naprednih analitičkih metoda za mjerenje kritičnih procesnih varijabli u realnom vremenu, čime se postiže bolje vođenje procesa i kvaliteta proizvoda. U kristalizacijskim procesima često se rabe tehnike poput Raman spektroskopije, FTIR spektroskopije, *in-line* mikroskopije i mjerenja zamućenosti. Ramanova spektroskopija primjenjuje se za procjenu kon-

centracije otopljene tvari i praćenje promjena u kristalnoj strukturi tijekom procesa kristalizacije. Zbog složenosti signala često je potrebno razviti multivarijatne kalibracijske modele koji uzimaju u obzir promjene u veličini i koncentraciji kristala.

Primjer takvog pristupa prikazan je u razvoju kalibracijskog modela za procjenu koncentracije otopljene djelatne farmaceutske tvari tijekom kristalizacije ceritiniba [29]. Pokazano je da uključivanje varijabli povezanih s veličinom kristala i koncentracijom čvrste faze značajno poboljšava procjenu koncentracije u dvofaznom sustavu. U istraživanju su primijenjene Ramanova spektroskopija, *in-line* procesna mikroskopija i mjerenje zamućenosti suspenzije, kako je prikazano na Slici 4.



Slika 4: Eksperimentalni postav procesa kristalizacije [29]

Ramanovom spektroskopijom prati se koncentracija otopljene tvari i identificira kristalne strukture tijekom provedbe kristalizacije. *In-line* procesnom mikroskopijom procjenjuje se veličina kristala putem distribucije duljine tetiva (engl. *chord length distribution*), pri čemu su kriteriji percentili D10, D50 i D90. Zamućenost suspenzije indikator je koncentracije čvrste faze.

Prije razvoja kalibracijskog modela provedena je analiza i obrada prikupljenih Ramanovih spektara. Primijenjene su različite metode predobrade spektralnih podataka. Za procjenu robusnosti modela primijenjena je metoda unakrsne validacije, pri čemu su pojedini skupovi podataka s različitim koncentracijama otopljene tvari uzimani naizmjenično za treniranje i validaciju modela. Za procjenu koncentracije otopljene djelatne tvari razvijeni su multivarijatni kalibracijski modeli temeljeni na parcijalnoj regresiji najmanjih kvadrata i umjetnim neuronskim mrežama. Ulazne varijable modela uključivale su Ramanove spektre, temperaturu procesa, zamućenost suspenzije

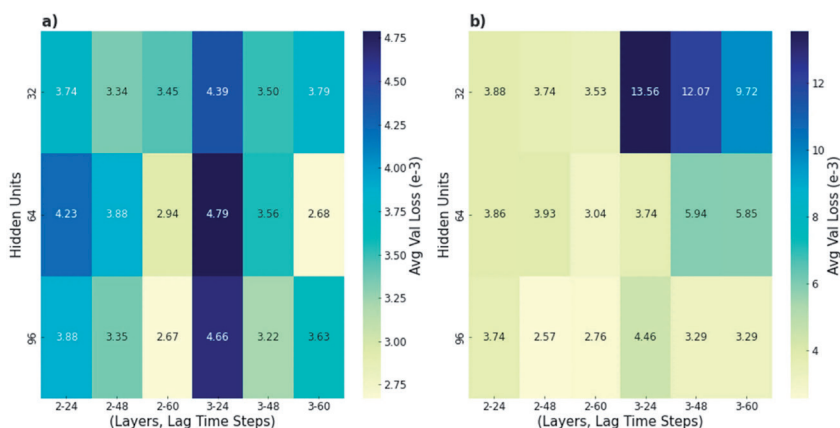
i parametre raspodjele veličine kristala. Rezultati su pokazali da modeli neuronskih mreža dobro procjenjuju koncentraciju otopljene tvari tijekom kristalizacije, pri čemu pogreška predviđanja nije prelazila približno 1 % nominalne koncentracije.

3.4 Predviđanje raspodjele veličine čestica tijekom kristalizacije

Raspodjela veličine čestica (engl. *particle size distribution*, PSD) predstavlja jedno od ključnih svojstava kristalnih proizvoda u farmaceutskoj i kemijskoj industriji. Ona izravno utječe na svojstva poput topljivosti, brzine otapanja, filtrabilnosti i daljnje obradivosti proizvoda. Za procjenu raspodjele veličine čestica primjenjuju se *in-line* mikroskopija, FBRM metode i optičke metode analize slike. Razvojem matematičkih modela i metoda strojnog učenja povezuju se procesne varijable s promjenama u raspodjeli veličine čestica.

U radu Vrbana i sur. [30] razvijen je model temeljen na strojnom učenju za predviđanje parametara veličine kristala tijekom procesa kristalizacije hlađenjem. Za analizu vremenskih nizova procesnih podataka primijenjena je LSTM neuronska mreža. Podaci su prikupljeni *in-situ* mikroskopijom, mjerenjem temperature, brzine hlađenja i količine cjepiva kristala. Cilj je bio predvidjeti metrike veličine kristala - D10, D50 i D90 koji karakteriziraju raspodjelu veličina čestica. Kako bi poboljšali predikciju modela proveden je postupak inženjerstva značajki (engl. *feature engineering*) pri čemu su uvedeni dodatni procesni deskriptori, poput derivacija i integrala temperature. Rezultati pokazuju da se LSTM modelom dobro predviđa raspodjela veličina kristala bez izravnog mjerenja prezasićenosti (Slika 5).

Takav pristup omogućuje bolje razumijevanje utjecaja procesnih uvjeta na razvoj kristala, a istovremeno pomaže pri optimiranju kristalizacijskih procesa i prijenosu procesa s laboratorijske na poluindustrijsku i industrijsku razinu.



Slika 5: Usporedba Avg Val Loss za LSTM mreže: (a) uz inženjerstvo značajki i (b) bez inženjerstva značajki [30]

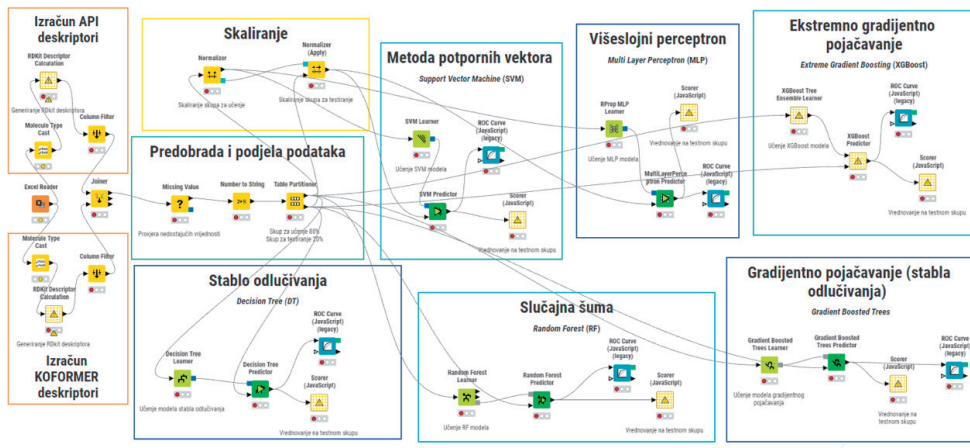
3.5 Strojno učenje u predviđanju nastanka kokristala

Kokristali predstavljaju kristalne sustave sastavljene od djelatne farmaceutske tvari (engl. *active pharmaceutical ingredient* - API) i koformera u određenom molarnom omjeru. Primjenjuju se za poboljšanje fizikalno-kemijskih svojstava lijekova poput topljivosti, stabilnosti i brzine otapanja. Tradicionalni postupci ispitivanja nastanka kokristala temelje se na eksperimentalnom probiru (engl. *screening*) velikog broja kombinacija API-ja i koformera, što je vremenski i financijski zahtjevno. Stoga se sve češće primjenjuju metode strojnog učenja za predviđanje mogućnosti formiranja kokristala na temelju molekulskih deskriptora. Modeli strojnog učenja koriste različite molekulske deskriptore poput broja donora i akceptora vodikovih veza, $\log P$ vrijednosti, topološke polarne površine i molekulske mase, broja rotacijskih veza i drugih strukturnih značajki molekula. Takvim deskriptorima kvantitativno se opisuju molekulska svojstva te povezuju strukture molekula s potencijalnim formiranjem kokristala.

3.5.1 Primjer klasifikacijskog modela za predviđanje kokristala

Primjer klasifikacijskog modela za predviđanje nastanka kokristala temelji se na uređenim parovima API-ja i koformera te skupu molekulskih deskriptora izračunatih iz strukturnih zapisa molekula. Podaci o poznatim kokristalima prikupljaju se iz baza kristalnih struktura ili iz relevantnih znanstvenih radova, nakon čega se iz molekulskih struktura izračunavaju odgovarajući deskriptori. Na temelju tih podataka razvijen je klasifikacijski model neuronske mreže koji predviđa vjerojatnost nastanka kokristala za određeni par molekula. Kako bi se procijenila sposobnost generalizacije model je uvježban i validiran primjenom postupaka unakrsne validacije na novim kombinacijama API-ja i koformera. Među brojnim strukturama modela (Slika 6) najbolje rezultate postigao je XGBoost model s prosječnim slaganjem od 80 % na validacijskom skupu.

Takvi modeli služe kao alat u preliminarnom odabiru perspektivnih kombinacija za eksperimentalna ispitivanja, čime se značajno smanjuje broj laboratorijskih eksperimenata i, time, ubrzava razvoj novih farmaceutskih formulacija.

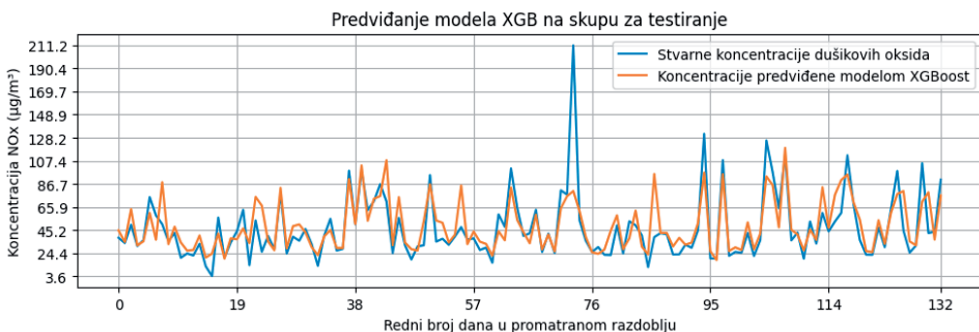


Slika 6: Dijagram toka razvoja modela u KNIME platformi [31]

3.6 Modeliranje koncentracije onečišćujućih tvari u zraku

3.6.1 Primjer razvoja modela za predviđanje koncentracije dušičnih oksida

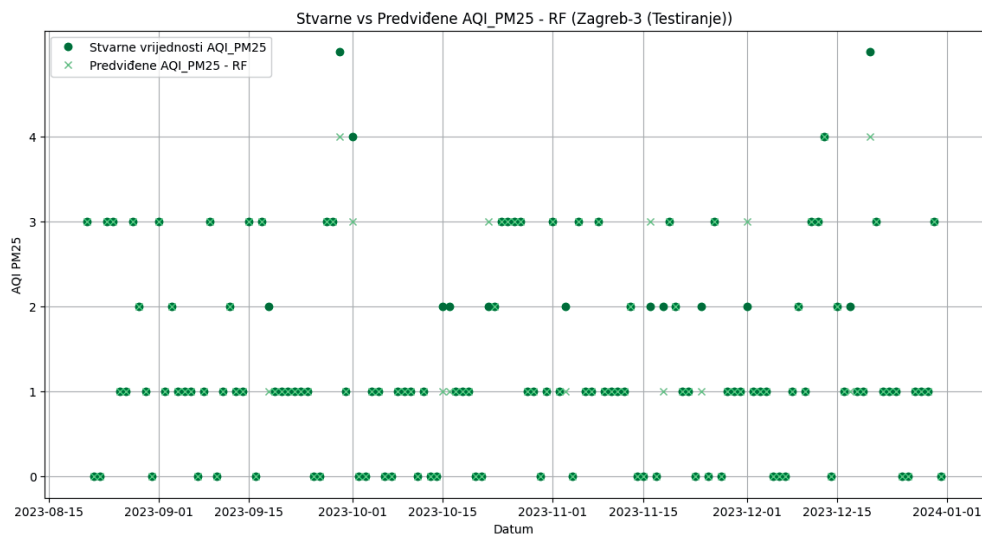
U radu od Zaninović [32] razvijeni su modeli strojnog učenja za predviđanje koncentracije dušikovih oksida (NO_x) u zraku na području Zagreba. Modeli su razvijeni na temelju podataka s mjernih postaja za praćenje kvalitete zraka i meteoroloških podataka: temperature zraka, brzine i smjera vjetera, relativne vlažnosti i atmosferskog tlaka. Za razvoj modela primijenjene su metode slučajne šume, XGBoost i umjetne neuronske mreže, a razvoj i evaluacija modela provedeni su u programskom jeziku Python. Dobiveni rezultati (slika 7) pokazuju da modeli uspješno opisuju promjene koncentracije NO_x ($R^2 \approx 0,70$; $\text{RMSE} \approx 28,7 \mu\text{g}/\text{m}^3$) i prepoznaju trendove onečišćenja zraka, što potvrđuje prikladnost za primjenu u sustavima za praćenje kvalitete zraka i potporu pri donošenju preventivnih mjera.



Slika 7: Usporedba vrijednosti koncentracije dušikovih oksida modelom XGBoost za mjernu postaju ZG3 sa stvarnim vrijednostima koncentracije dušikovih oksida na skupu za validaciju [32]

3.6.2 Primjer razvoja modela za predviđanje indeksa kvalitete zraka

U diplomskom radu Klonkay [33] razvijeni su modeli strojnog učenja za procjenu indeksa kvalitete zraka (AQI) povezanog s koncentracijama lebdećih čestica PM10 i PM2,5 na području Zagreba. Modeli su razvijeni na temelju podataka prikupljenih na mjernim postajama državne mreže za praćenje kvalitete zraka te meteoroloških podataka. Za razvoj klasifikacijskih modela primijenjene su metoda slučajne šume i metoda potpornih vektora (SVM) u programskom jeziku Python. Rezultati su pokazali da model slučajne šume ostvaruje veću točnost (97 %) u odnosu na SVM model (87 %) na sve tri lokacije što se može pripisati boljoj sposobnosti modela slučajne šume da opiše složene nelinearne odnose između varijabli i većoj otpornosti na šum u podacima. Na Slici 8 prikazana je dobra podudarnost između stvarnih i modelom slučajne šume predviđenih vrijednosti AQI za testni skup podataka za svih šest razina AQI.



Slika 8: Prikaz stvarnih i RF modelom predviđenih vrijednosti AQI za testni skup podataka [33]

4. Zaključak

Ovim pregledom nastojalo se prikazati kako se metode modeliranja i strojnog učenja primjenjuju u različitim granama kemijskog procesnog inženjerstva.

Integracija matematičkih modela, naprednih analitičkih metoda i strojnog učenja dala je značajan impuls za razvoj digitalnih alata za praćenje, dijagnostiku i optimiranje industrijskih procesa. Očekuje se da će se daljnji razvoj zbivati kroz integraciju velikih skupova podataka, naprednih metoda optimizacije, hibridnih modela i digitalnih blizanaca procesa.

Razvoj metoda strojnog učenja i primjena softverskih senzora imaju značajnu perspektivu u kemijskom inženjerstvu, osobito u području prediktivnog vođenja procesa i optimiranja rada u stvarnom vremenu, kao i u praćenju stanja sustava i okoliša. Posebno se ističe primjena u prediktivnom održavanju za pravovremeno predviđanje kvarova i problema u radu postrojenja čime se smanjuje vrijeme zastoja i planira održavanja opreme i postrojenja.

U skoroj budućnosti očekuje se sve veća primjena u razvoju digitalnih blizanaca postrojenja koji kombiniraju modele temeljene na fizikalnim zakonitostima i one razvijene metodama strojnog učenja. Takvi sustavi služe za simulaciju različitih scenarija rada bez rizika za stvarno postrojenje. Također, pružaju podršku pri optimiranju procesa i donošenju operativnih odluka.

Unatoč navedenim prednostima, primjena ovih metoda i dalje je ograničena izazovima poput nedostatka relevantnih podataka, nedovoljne interpretabilnosti modela i složene integracije u postojeće sustave.

U kontekstu primjene strojnog učenja u industrijskim procesima sve se više naglašava koncept “*human-in-the-loop*” sustava, u kojima modeli strojnog učenja pružaju potporu odlučivanju, dok nezamjenjivu ulogu i dalje ima inženjer koji detaljno poznaje proces, podatke i ograničenja sustava pa, na temelju toga, smisleno tumači i primjenjuje rezultate analize.

Kemijsko inženjerstvo nalazi se u fazi intenzivne digitalne transformacije, pri čemu integracija procesnog znanja s metodama podatkovne znanosti i strojnog učenja predstavlja ključan smjer budućeg razvoja koji otvara nove mogućnosti za unapređenje industrijskih procesa.

5. Literatura

- [1] Kadlec, P.; Gabrys, B., Strandt, S.: Data-driven soft sensors in the process industry, *Computers and Chemical Engineering*, **33** (2009) 795-814, <https://doi.org/10.1016/j.compchemeng.2008.12.012>
- [2] Fortuna, L.; Graziani, S., Rizzo, A., Xibilia, M.G.: *Soft Sensors for Monitoring and Control of Industrial Processes (Advances in Industrial Control)*; Springer: London, UK (2007)
- [3] Wang, J.; Sun, L., Li, H., Ding, R., Chen, N.: Prediction Model of Fouling Thickness of Heat Exchanger Based on TA-LSTM Structure, *Processes*, **11** (2023) 845, <https://doi.org/10.3390/pr11092594>.
- [4] Al Shabaan, M. A.; Nemer, Z. N.: Oil and Gas Production Forecasting Using Decision Trees, Random Forest, and XGBoost, *Journal of Al-Qadisiyah Computer Science and Mathematics*, **16** (2024) 864, <https://doi.org/10.29304/jqcs.2024.16.11431>.

- [5] Lin, M.; Wu, Y., Rohani, S.: Simultaneous Measurement of Solution Concentration and Slurry Density by Raman Spectroscopy with Artificial Neural Network, *Crystal Growth & Design*, **20** (2020) 1752-1759
- [6] Wu, Y.; Gao, Z., Rohani, S.: Deep Learning-Based Oriented Object Detection for in Situ Image Monitoring and Analysis: A Process Analytical Technology (PAT) Application for Taurine Crystallization, *Chemical Engineering Research and Design*, **170** (2021) 444-455
- [7] Birolo, R.; Bravetti, F., Alladio, E., Priola, E., Bianchini, G., Novelli, R., Aramini, A., Gobetto, R., Chierotti, M. R.: Speeding Up the Cocrystallization Process: Machine Learning-Combined Methods for the Prediction of Multicomponent Systems, Supporting Materials, *Crystal Growth & Design*, **23** (2023) 2-12
- [8] Deng, J.; Ye, Z., Zheng, W., Chen, J., Gao, H., Wu, Z., Chan, G., Wang, Y., Cao, D., Wang, Y.; Lee, S. M.: Machine learning in accelerating microsphere formulation development, *Drug Delivery and Translational Research*, **13** (2022) 966–982 <https://doi.org/10.1007/s13346-022-01253-z>.
- [9] Méndez, M.; Merayo, M. G., Núñez, M.: Machine learning algorithms to forecast air quality: A survey, *Artificial Intelligence Review*, **56** (2023) 10031–10066
- [10] Altamirano-Astorga, J.; Gutierrez-Garcia, J.O., Roman-Rangel, E.: Forecasting Indoor Air Quality in Mexico City Using Deep Learning Architectures, *Atmosphere (Basel)*, **15** (2024) 1529, <https://doi.org/10.3390/atmos15121529>.
- [11] Namdeti, R.: Artificial Intelligence in Chemical Engineering: Past, Present, and Future Perspectives, *Journal of Chemical Health Risks*, **13** (2023) 2051-2061, <https://doi.org/10.52783/jchr.v13.i6.2058>
- [12] Al Sharah, A.; Owida, H. A., Alnaimat, F., Hassan, M., Abuowaida, S., Alhaj, M., Sharadqeh, A.: Application of machine learning in chemical engineering: outlook and perspectives, *IAES International Journal of Artificial Intelligence*, **13** (2024) 619-630, <https://doi.org/10.11591/ijai.v13.i1>.
- [13] Ding, C.; Gui, X., Jiang, J.: Advancing chemical engineering technology with artificial intelligence, *Clean Energy*, **9** (2025) 55–74, <https://doi.org/10.1093/ce/zkaf036>
- [14] Chiang, L. H. T.; Braun, B., Wang, Z., Castillo, I.: Towards AI at Scale in the Chemical Industry, *AIChE Journal*, **68** (2022) <https://doi.org/10.1002/aic.17644>
- [15] He, C.; Zhang, C., Bian, T., Jiao, K., Su, W., Wu, K.-J., Su, A.: A Review on Artificial Intelligence Enabled Design, Synthesis, and Process Optimization of Chemical Products for Industry 4.0, *Processes*, **11** (2023) 330, <https://doi.org/10.3390/pr11020330>
- [16] Mafat, I. H.; Surya, D.V., Sharma, S.K., Rao, C.S.: Exploring Machine Learning Applications in Chemical Production through Valorization of Biomass, Plastics, and Petroleum Resources: A Comprehensive Review, *Journal of Analytical and Applied Pyrolysis*, **180** (2024), 106512, <https://doi.org/10.1016/j.jaap.2024.106512>
- [17] Selvakumar, P.; Preethi, C., Nehru, P., Saravanan, A., Das, A.: *AI in Green Chemistry Sustainable Manufacturing Processes*, IGI Global Scientific Publishing, ISBN 979-8-3693-7483-2, Hershey, PA, USA, (2025)

- [18] Carter, A.; Imtiaz, S.; Naterer, G.F.: Review of interpretable machine learning for process industries, *Process Safety and Environmental Protection*, **170** (2023) 647-659, <https://doi.org/10.1016/j.psep.2022.12.018>
- [19] Di Bonito, L.P.; Campanile, L., Di Natale, F., Mastroianni, M., Iacono, M., eXplainable Artificial Intelligence in Process Engineering: Promises, Facts, and Current Limitations, *Applied System Innovation*, **7** (2024), 121, <https://doi.org/10.3390/asi7060121>
- [20] Brereton, R.: *Chemometrics: Data Driven Extraction for Science*, John Wiley & Sons, Hoboken, NJ, USA, (2018), 121-181
- [21] Liang, H.; Jiang, K., Yan, T.-A., Chen, G.-H.: XGBoost: An Optimal Machine Learning Model with Just Structural Features to Discover MOF Adsorbents of Xe/Kr, *ACS Omega*, **6** (2021) 13, 9066-9074, <https://doi.org/10.1021/acsomega.1c00471>
- [22] Hochreiter, S.; Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, **9** (1997) 8, 1735-1780, <https://doi.org/10.1162/neco.1997.9.8.1735>
- [23] Ahn, J. M.; Kim, J., Kim, K.: Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting, *Toxins (Basel)*, **15** (2023) 10, 608, <https://doi.org/10.3390/toxins15100608>
- [24] Ujević Andrijić, Ž.; Rolich, T., Bolf, N.: Soft sensor development for the estimation of benzene content in catalytic reformat, *Industrial & engineering chemistry research*, **51** (2012) 3007-3014, <https://doi.org/10.1021/ie202362d>
- [25] Ujević Andrijić, Ž.; Herceg, S., Šimić, M., Bolf, N.: Intelligent Soft Sensors for Inferential Monitoring of Hydrodesulfurization Process Analyzers, *Actuators*, **14** (2025) 1-26, <https://doi.org/10.3390/act14080410>
- [26] Herceg, S.; Ujević Andrijić, Ž., Rimac, N., Bolf, N.: Development of mathematical models for industrial processes using dynamic neural networks, *Mathematics*, **11** (2023) 4518, <https://doi.org/10.3390/math11214518>
- [27] Ujević Andrijić, Ž.; Rimac, N.: Data-driven fouling detection in refinery preheat train heat exchangers using neural networks and gradient boosting, *Sensors*, **25** (2025) 4936, <https://doi.org/10.3390/s25164936>
- [28] Ujević Andrijić, Ž.; Bolf, N., Rimac, N., Brzović, A.: Fouling Detection in Industrial Heat Exchanger Using Number of Transfer Units Method, Neural Network, and Nonlinear Finite Impulse Response Models, *Heat Transfer Engineering*, **43** (2022) 1852-1866, <https://doi.org/10.1080/01457632.2021.2016149>
- [29] Gavran, M.; Andrijić, Ž.U., Bolf, N., Rimac, N., Sacher, J., Šahnić, D.: Development of a Calibration Model for Real-Time Solute Concentration Monitoring during Crystallization of Ceritinib Using Raman Spectroscopy and In-Line Process Microscopy, *Processes*, **11** (2023) 3439, <https://doi.org/10.3390/pr1133439>
- [30] Vrbanić, I.; Bolf, N., Budimir Sacher, J.: Data-Driven Prediction of Crystal Size Metrics Using LSTM Networks and In Situ Microscopy in Seeded Cooling Crystallization, *Processes*, **13** (2025) 1860, <https://doi.org/10.3390/pr13061860>

- [31] Vretenar, F.; Hartmann, H., Ujević Andrijić, Ž.: *In silico* ispitivanje nastajanja kokristala u farmaceutskom razvoju, Knjiga sažetaka - XVI. susret mladih kemijskih inženjera, Zagreb (2026).
- [32] Zanimović, V., Diplomski rad: Predviđanje koncentracija dušikovih oksida (NO_x) u zraku u gradu Zagrebu primjenom metoda strojnog učenja, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije, 2025
- [33] Klonekay, P., Diplomski rad: Procjena indeksa kvalitete zraka za lebdeće čestice PM₁₀ i PM_{2,5} grada Zagreba primjenom metoda strojnog učenja, Sveučilište u Zagrebu Fakultet kemijskog inženjerstva i tehnologije, 2024.

Zahvala

Ova istraživanja i analiza provedeni su u okviru projekta *Digi-ChemEng-Lab* (112113), financiranog sredstvima NextGenerationEU fonda Europske unije iz izvora 581 – Mehanizam za oporavak i otpornost, u sklopu Programa financiranja javnih visokih učilišta i javnih znanstvenih instituta. Dio istraživanja rezultat je projekta *CrystAPC - Napredno vođenje procesa kristalizacije* sufinanciranog iz Europskog fonda za regionalni razvoj.