

## Kauzalni modeli obrade podataka o kakvoći hrane pomoću umjetne inteligencije

### SAŽETAK

*Pozadina istraživanja.* Svrha je ovog istraživanja bila naglasiti važnost korištenja umjetne inteligencije (AI) i modeliranja uzročnosti kakvoće hrane, te analize velike količine podataka. Umjetna inteligencija sa strukturnim uzročnim modeliranjem (SCM), temeljena na Bayesovim mrežama i dubokom učenju, omogućuje integraciju teorijskog znanja iz područja prehrambene tehnologije s podacima o proizvodnom procesu, fizikalno-kemijskim svojstvima te organoleptičkim ocjenama proizvoda. Prehrambeni proizvodi su složene naravi, a podaci vrlo dimenzionalni, sa zamršenim međuodnosima (korelacijama), koje je teško povezati sa osjetilnom percepcijom potrošača o kakvoći hrane. Standardne tehnike regresijskog modeliranja, kao što su višestruki obični najmanji kvadrati (OLS) i parcijalni najmanji kvadrati (PLS), učinkovito se primijenjuju za predviđanje učinaka pomoću linearnih interpolacija opaženih podataka pod stacionarnim uvjetima presjeka. Nadogradnja modela linearne regresije pomoću strojnog učenja (ML) uzima u obzir nelinearne odnose i otkriva funkcionalne obrasce, ali je sklona zbunjujućim i neuspjelim predviđanjima u neopaženim nestacionarnim uvjetima. Interferencija varijabli podataka glavna je prepreka primjeni regresijskih modela u prehrambenim inovacijama pod prethodno neuvježbanim uvjetima. Stoga se ovaj rad usredotočuje na primjenu kauzalnih grafičkih modela s Bayesovim mrežama za zaključivanje uzročno-posljedičnih odnosa i učinaka intervencije između procesnih varijabli i senzorske procjene kakvoće hrane.

*Eksperimentalni pristup.* Ovo se istraživanje temelji na podacima dostupnim u literaturi, o procesu pečenja kruha od pšeničnog brašna, potrošačkim ocjenama senzorske kakvoće fermentiranih mliječnih proizvoda, te rezultatima stručnog kušanja vina. Podaci o kakvoći pečenja pšenice regulirani su operatorom najmanjeg apsolutnog skupljanja i odabira (LASSO elastična mreža). Bayesova statistika primijenjena je za procjenu zajedničke funkcije vjerojatnosti modela za zaključivanje mrežne strukture i parametara. Dobiveni strukturni kauzalni modeli prikazani su kao usmjereni aciklički grafovi (DAG). Kriteriji D-odvajanja primijenjeni su za blokiranje interferirajućih učinaka pri procjeni izravnih i ukupnih uzročnih učinaka procesnih varijabli i percepcije potrošača o kakvoći hrane. Distribucije vjerojatnosti uzročnih učinaka intervencije pojedinih procesnih varijabli na kakvoću prikazane su kao dijagrami djelomične ovisnosti, određeni Bayesovim neuronskim mrežama. U slučaju uzročnosti kakvoće vina, ukupni uzročni učinci utvrđeni pomoću SCM potvrđeni su algoritmom dvostrukog strojnog učenja (DML).

*Rezultati i zaključci.* Ispitan je skup podataka od 45 kontinuiranih varijabli koje odgovaraju različitim varijablama kemijskih, fizikalnih i biokemijskih svojstava sedam hrvatskih kultivara pšenice prikupljenim tijekom dvije godine kontroliranog uzgoja. LASSO regulacija skupa podataka dala je deset ključnih prediktora, koji obuhvaćaju 98 % varijance podataka o kakvoći pečenja. Na osnovi ključnih varijabli izveden je prediktivni model slučajne šume sa 75 % točnosti unakrsne provjere. Uzročna analiza između kakvoće i ključnih prediktora temeljila se na Bayesovom modelu prikazanom kao DAG. Udjel proteina imao je najveći izravni uzročni učinak s koeficijentom puta od 0,71; udjel ukupnih podjedinica glutenina velike molekularne mase bio je neizravni uzrok s koeficijentom puta od 0,42; dok je prosječni uzročni učinak (ACE) ukupnog udjela proteina bio 0,65. Veliki skup podataka o kakvoći fermentiranih mliječnih proizvoda uključivao je binarne senzorske podatke (okus, miris, zamućenost), kontinuirane fizikalne varijable (temperatura, masnoća, pH, boja) i tri stupnja potrošačke ocjene proizvoda. Model slučajnih šuma izveden je radi predviđanja klasifikacije kakvoće s „out-of-bag“ (OOB) pogreškom od 0,28 %. Bayesov mrežni model predviđa da na klasifikaciju okusa izravno utječu temperatura, boja i udjel masti, dok na klasifikaciju kakvoće izravno utječu temperatura, zamućenost, miris i udjel masti. Procijenjeni su ključni ACE od  $-0,04$  stupnja kakvoće/°C i 0,3 stupnja kakvoće/udjelu masti. Ovisnost ACE o temperaturi pokazuje nelinearni tip kao negativno zasićenje s točkom „prijeloma“ na 60 °C, dok je ACE udjela masti imao pozitivan linearni trend. Uzročna analiza kakvoće crnog i bijelog vina temeljila se na velikom skupu podataka od jedanaest kontinuiranih varijabli fizikalnih i kemijskih svojstava i procjena kakvoće razvrstanih u deset klasa, od 1 do 10. Svaku je klasifikaciju u tri ponavljanja proveo panel profesionalnih kušača vina. Za procjenu ACE ukupne kakvoće primijenjen je algoritam nestrukturalnog dvostrukog strojnog učenja (DML). Udjel alkohola u crnom i bijelom vinu imao je ključni pozitivni ACE relativni faktor od 0,35 kakvoće/udjelu alkohola, dok je hlapljiva kiselost imala ključni negativni ACE od  $-0,2$  kakvoće/kiselosti. Dobivena predviđanja ACE nestrukturiranim algoritmom DML uvelike odgovaraju onima dobivenim strukturnim SCM-om.

*Novost i znanstveni doprinos.* Prikazane su nove metodologije i rezultati primjene kauzalnih modela umjetne inteligencije u analizi potrošačke procjene kakvoće prehrambenih proizvoda. Primjena Bayesovih mrežnih strukturno kauzalnih modela (SCM) omogućuje d-odvajanje izraženih učinaka konfuzije između parametara u nekauzalnim regresijskim modelima. Na temelju SCM-a, zaključivanje ACE-a daje potkrijepljene i potvrđene istraživačke hipoteze za nove proizvode i podršku za odluke o mogućim intervencijama u svrhu poboljšanja dizajna proizvoda, uvođenja novih procesa, kontrolu procesa, upravljanja i marketinga.

**Ključne riječi:** Bayesova mreža; AI uzročnost; učinci intervencije; prosječni uzročni učinak (ACE); kakvoća hrane