

# Mogućnosti primene mašinskog učenja u oplemenjivanju mlečnih goveda

Laslo Tarjan, Ivana Šenk

Katedra za mehatroniku, robotiku i automatizaciju  
Univerzitet u Novom Sadu, Fakultet tehničkih nauka  
Novi Sad, Srbija  
laci@uns.ac.rs, ivanas@uns.ac.rs

Doni Pracner

Univerzitet u Novom Sadu, Prirodno-matematički fakultet  
Novi Sad, Srbija  
doni.pracner@dmi.uns.ac.rs

Dušan Rajković, Ljuba Štrbac

Univerzitet u Novom Sadu, Poljoprivredni fakultet  
Novi Sad, Srbija  
rajkovicdusan@gmail.com, ljuba.strbac@stocarstvo.edu.rs

**Sažetak**— Razvoj tehnika mašinskog učenja i prisutnost različitih modela mašinskog učenja u poslednjoj deceniji otvorio je put za njihovu implementaciju u svim oblastima istraživanja. Metode dubokog učenja izdvajaju se među najefikasnijim alatima, čemu je doprineo eksplozivni rast količine prikupljenih i dostupnih podataka na svim poljima, kao i značajan tehnološki napredak grafičkih procesora i nauke u ovoj oblasti. Cilj ove studije je da istraži potencijal primene mašinskog učenja u oplemenjivanju mlečnih goveda, na sličnim primerima iz oblasti, i da razmotri mogući odabir skupa podataka za obuku modela u cilju iznalaženja kvantitativnih opisa odabranih parametara u oplemenjivačkim zadacima.

**Ključne riječi**- mašinsko učenje, deep learning, oplemenjivanje mlečnih goveda; (key words)

## I. UVOD

Prema procenama FAO (eng. Food and Agriculture Organization of the United Nations) i drugih organizacija svetska populacija ljudi će do 2050. godine porasti na broj od 9 do 10 milijardi [1], [2]. Po izveštaju FAO iz 2012. godine [3] u svetu ima dovoljno hrane za oko 7 milijardi ljudi, ali skoro milijarda je gladna i neuhranjena. Da bi se nahranilo blizu deset milijardi ljudi do 2050. godine poljoprivredna proizvodnja bi trebala da se udvostruči, što predstavlja izazov [2], [4].

Rad na oplemenjivanju životinja je dao prilično dobre rezultate poslednjih 150 godina, te je do danas stočarska proizvodnja višestruko porasla. Farmeri su uvideli da pojedinac, sam, može vrlo malo da uradi u oplemenjivanju, pa su tako na zapadu nastale odgajivačke organizacije koje su bile rezultat težnje da se u zajedničkom radu dođe do napretka (prva matična knjiga je uvedena 1822. godine za tovnja goveda, 1875. za mlečna goveda u Holandiji, a prva asocijacija farmera mlečnih goveda osnovana je 1895. u Danskoj). Još su tada uvideli koliko je važno imati, kako tačne i pouzdane podatke o svim važnim aspektima proizvodnje, tako i što više

objedinjenih podataka kako bi ocene bile tačnije. Ipak, ulaganja u poljoprivredu, u većini zemalja u razvoju, tokom poslednjih decenija su u padu, što je rezultiralo manjim ulaganjima u istraživanja i razvoj u poređenju sa razvijenim zemljama, a to je dovelo i do niže produktivnosti i stagnacije proizvodnje [5]. Da bi se poboljšala situacija nadležna zakonodavna tela (država) svoju politiku pored velikih treba da usmeri i prema malim poljoprivrednicima i da kreira programe podrške koji bi pomogli u stvaranju povoljne investicione klime za poljoprivredu [5].

Sprovođenje odgajivačkih programa u stočarstvu na teritoriji Republike Srbije je organizovano na više nivoa (od odgajivača preko odgajivačkih i organizacija sa posebnim ovlašćenjima do nadležnih pokrajinskih i republičkih resornih ministarstava). Odgajivačke organizacije su organizovane na tri nivoa: osnovne, regionalne i glavne odgajivačke organizacije. Na teritoriji AP Vojvodine glavna odgajivačka organizacija (GOO) u govedarstvu je Departman za stočarstvo, Poljoprivrednog fakulteta u Novom Sadu [6]. Posao ove organizacije je izrada i sprovođenje glavnih odgajivačkih programa po rasama (holštajn-frizijsku, simentalsku, braon svis, tovnje i autohtone rase goveda). Između ostalog ova organizacija vodi glavnu matičnu evidenciju i izdaje različita zootehnička dokumenta (npr. Pedigree). Međutim, jedan od najvažnijih poslova jeste izračunavanje procena oplemenjivačkih vrednosti (eng. estimated breeding values (EBV)) i rangiranje grla na nivou populacije.

Kontinualno praznjenje sela dovodi do manjka radnika koji rade u poljoprivredi u uzgoju životinja, što se prvo ogleda u problemima oko ishrane i održavanju životinja, ali i u prikupljanju podataka. Automatizacija ovih poslova ili barem dela posla može biti od izuzetne pomoći [7]. Da bi se moglo razmišljati o razvoju, neophodno je ljude rasteretiti od posla i dati značaj prikupljanju podataka. Jedan od zadataka odgajivačkih organizacija je da prikupljaju podatke od odgajivača i da ih savetuje kako da maksimalizuju svoju proizvodnju pre svega putem genetskog unapređenja. Pored

---

Rad je objavljen na engleskom jeziku i nalazi se na IEEE Xplore na adresi <https://ieeexplore.ieee.org/document/9400672> za citiranje molimo vas koristite sledeći oblik:

L. Tarjan, I. Šenk, D. Pracner, D. Rajković and L. Štrbac, "Possibilities for applying machine learning in dairy cattle breeding," 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH), 2021, pp. 1-6, doi: 10.1109/INFOTEH51037.2021.9400672.

uobičajenih statističkih i analitičkih metoda za izračunavanje oplemenjivačkih vrednosti, algoritmi se sve više oslanjaju i na veštačku inteligenciju. Razvojem metoda mašinskog učenja, koji se u poslednjoj deceniji vrlo uspešno primenjuje u različitim servisima baziranim na velikim količinama podataka, postavlja se logično razmišljanje zašto to ne bi moglo da se iskoristi i u zadacima oplemenjivanja životinja, u konkretnom slučaju mlečnih goveda, gde bi se na osnovu dosadašnjih podataka obučila mreža koja bi bila u stanju da predloži buduće parove roditelja. Brojne studije, rađene u protekloj deceniji, pokazuju da se algoritmi mašinskog učenja efikasno mogu koristiti i u procesu planiranja odgoja mlečnih goveda, kako za odabir rasa, roditeljskih parova, načina ishrane, tako i za prognozu ekonomske isplativosti ulaganja u određenu rasu u zavisnosti od željenih izlaznih proizvoda [8]–[11].

Cilj ovog rada je da na osnovu dosadašnjih radova iz oblasti analizira opravdanost implementacije alata na bazi modela dobijenim mašinskim učenjem, kako za pomoć odgajivačima, tako i za pomoć odgajivačkim organizacijama u Srbiji, a pre svega na teritoriji AP Vojvodini, za koju se trenutno podaci obrađuju, kako bi se modeli i iskustva iz razvijenih zemalja primenili i u našoj zemlji.

Alati (servisi) razvijeni na bazi modela dobijenim mašinskim učenjem, koje bi mogli koristiti kako odgajivači tako i odgajivačke organizacije, bi doprinelo u podizanju svesti o potrebi i značaju beleženja i deljenja podataka jer sami odgajivači sa tim podacima ne mogu uraditi puno, ali zajednički na nivou države i regije itekako mogu. Ovi alati bi bili od velike pomoći i za samog odgajivača, ali i za državu u donošenju odluka i mera.

Rad je organizovan u četiri poglavlja. Nakon uvodnog dela u drugom poglavlju diskutuje se potreba za oplemenjivanjem u cilju povećanja produktivnosti, kao i baza podataka o trenutnoj populaciji mlečnih goveda u Srbiji koju čine podaci skupljeni iz više izvora i koje je potrebno povezati i validirati pre upotrebe. U trećem delu je dat pregled radova dugih autora, i prikazuje koje metode mašinskog učenja su se dobro pokazale u praksi drugih zemalja, i odakle ima smisla kretati. U četvrtom delu su dati dosadašnji zaključci i pravci daljeg rada.

## II. OPLEMENJIVANJE – ZNAČAJ I POTREBA

Oplemenjivanje mlečnih goveda je kontinuirani ljudski rad na poboljšanju rasa mlečnih goveda zarad ostvarivanja boljih proizvodnih rezultata (veće mlečnosti, boljeg hemijskog sastava mleka, dugovečnosti i dr.) kroz primenu saznanja o principima nasleđivanja osobina. U tom radu se koriste dva osnovna alata: selekcija i parenje [12]. Selekcija podrazumeva odabir roditelja naredne generacije i kao takva predstavlja najvažniji faktor promene frekvencije gena u populaciji. Dakle, ne dozvoljava se svakoj junici i svakom biku da postane roditelj, već se vrši odabir najboljih grla. S obzirom da se krave u intenzivnoj proizvodnji tele u proseku od 2,8 do 3,2 puta u toku životu [13], a upotrebom seksiranog semena oko 1,9 puta [14] gotovo sve junice ostaju u priplodu, pa je selekcijski pritisak sa ženske strane mali. Sa druge strane, pojavom i razvojem metoda veštačkog osemenjavanja otvorila se mogućnost da genetski superiorni bikovi imaju ogroman broj potomaka, pa je selekcijski pritisak sa muške strane veoma

velik. Dakle, veći deo genetskog napretka u govedarstvu, dolazi kroz selekciju (odabir) bikova. Parenje sa druge strane, predstavlja korišćenje u reprodukciji odabranih parova, najčešće veštačko osemenjavanje odabranih krava ili junica semenom određenog bika. Za razliku od npr. živinarstva, gde se u Srbiji proizvodnja zasniva na gajenju hibrida, u govedarstvu se oplemenjivanje odvija sprovođenjem selekcije na nivou rasa. Da bi se procenilo kolika je oplemenjivačka vrednost nekog bika koriste se svi dostupni podaci, i to: proizvodni rezultati njegovih predaka, sestara i potomaka; kao i svi ostali dostupni podaci koji ulaze u modele za procenu oplemenjivačkih vrednosti. Kao što je to već u uvodu rečeno na zapadu su pre više od sto godina uočili da u oplemenjivanju goveda pojedinačni farmer sam ne može puno da uradi, ali da se zajedničkim naporima može doći do željenih rezultata. Odgajivačke organizacije su nastale kao rezultat težnje da se zajedničkim radom dođe do napretka, da se obezbede tačni i pouzdani podaci o svim važnim aspektima proizvodnje, koji objedinjuju sve podatke kako bi ocene bile tačnije. Kako bi se poboljšao selekcijski rad 1983. godine je osnovan Interbull-centar [15] koji zahvaljujući međunarodnoj razmeni informacija, pomaže zemljama članicama da razvijaju efikasne metode za genetsko vrednovanje goveda, što je neophodno za genetski napredak i povećanje ekonomičnosti proizvodnje.

### A. Baza prikupljenih podataka

Posedovanje baze podataka o trenutnoj populaciji mlečnih goveda u Srbiji je od ključnog značaja za dalji napredak u oplemenjivačkom radu. Saradnja sa okruženjem je vrlo bitna, ali prvi korak je stvaranje jedinstvene baze podataka o poreklu i proizvodnim rezultatima mlečnih goveda na teritoriji Srbije. U Srbiji odgajivačke organizacije definišu i prikupljaju od odgajivača podatke značajane za dalji oplemenjivački rad. S obzirom na to da se podaci od odgajivača do sada nisu direktno postavljali u centralizovanu bazu podataka već više različitih osnovnih odgajivačkih organizacija vode posebne evidencije za sebe, a prema glavnoj odgajivačkoj organizaciji ove podatke dostavljaju preko obrazaca u vidu tabela različitih formata, u načinu zapisivanja podataka postoje određena odstupanja što otežava krajnji rad sa istima. Pored ovih problema postoji i problem sa nepostojanjem validacije podataka koje u svoju evidenciju unose osnovne odgajivačke organizacije, tako da korisnik može da napravi određene greške što kasnije predstavlja veliki problem u radu.

Budući da više različitih osnovnih odgajivačkih organizacija vodi sopstvene evidencije, prvi korak u dobijanju jedinstvene baze je sređivanje već postojećih podataka u jedinstvenu bazu, uvezanu tako da se lako mogu izlistati povezani podaci. Ovo je neophodno kako zbog statističkih analiza tako i zbog mogućeg ukrštanja baze sa drugim bazama iz okruženja, kao i mogućnošću korišćenja baze za obuku modela kod mašinskog učenja. Zadatak *BioITGenoSelect* projekta, finansiran od strane Fonda za nauku Republike Srbije, koji je pokrenut u julu 2019. godine, čiji je i ovaj rad deo, je da

1) pripomogne u formiranju jedinstvene baze podataka o poreklu grla i proizvodnim rezultatima mlečnih goveda u Srbiji,

---

Rad je objavljen na engleskom jeziku i nalazi se na IEEE Xplore na adresi <https://ieeexplore.ieee.org/document/9400672> za citiranje molimo vas koristite sledeći oblik:

L. Tarjan, I. Šenk, D. Pracner, D. Rajković and L. Štrbac, "Possibilities for applying machine learning in dairy cattle breeding," 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH), 2021, pp. 1-6, doi: 10.1109/INFOTEH51037.2021.9400672.

2) da se izvrši genotipizacija određenog broja životinja sa ciljem da se započne sa formiranjem referentne populacije mlečnih goveda u Srbiji,

3) razviju softverski alati za olakšanu manipulaciju bazama podataka,

4) razviju odgovarajući statistički modeli i izračunaju genomskih oplemenjivačkih vrednosti,

5) izvrši edukaciju odgajivača i stručnjaka u odgajivačkim organizacijama za tumačenje različitih izvora informacija.

U cilju kreiranja jedinstvene baze, dosadašnji podaci dostupni u glavnoj odgajivačkoj organizaciji su uveženi u bazu podataka koja objedinjuje informacije o poreklu i proizvodnim osobinama mlečnih goveda. Prvo su objedinjeni podaci o poreklu grla upisanih u glavnu matičnu evidenciju (najmanje dve generacije predaka) sa njihovim podacima o važnijim osobinama mlečnosti (količina mleka, sadržaj mlečne masti i mlečnih proteina).

Rad na centralnoj aplikaciji za skladištenje podataka o životinjama je obuhvatao kreiranje modela za postojeće podatke, kao i kreiranje adekvatnih sistema za konverziju i uvoz podataka iz postojećih baza (excel tabela). Važan momenat u ovim sistemima je i provera postojećih podataka, pošto su oni skupljani iz više izvora i moguća je određena nekonzistentnost, kao i greške u unosu podataka. U bazi se nalazi preko 120 hiljada grla od kojih postoje podaci o proizvodnji mleka za preko 70 hiljada grla.

Drugi bitan aspekt aplikacije je veb interfejs, koji će omogućiti zainteresovanima da pristupe podacima. Veb interfejs je baziran na Django (Python) web frejmworku, koji treba da omogući brzi razvoj aplikacija. Pored uvoza, započet je rad i na mogućnostima za izvoz podataka u formatima koji su neophodni za obradu u različitim specijalizovanim softverima za kvantitativno genetske analize (analiza pedigreea, ocena genetskih parametara i procene oplemenjivačkih vrednosti).

### III. PRIMENA MAŠINSKOG UČENJA

Primena algoritama mašinskog učenja u modernim softverskim alatima je budućnost koja treba da se koristi i u oplemenjivanju životinja u Srbiji, kao što se to radi i u razvijenom svetu. Mnogobrojne studije [8]–[11], rađene u protekloj deceniji, mogu biti primeri dobre prakse za implementaciju u Srbiji.

Studija istraživača Mwanga i dr. [10], koja je rađena na prikupljenim podacima sa više od 16.000 farmi sa teritorije Tanzanije, Kenijem Etiopije i Ugande, u periodu jun 2015. – jun 2016. pokazuje uspešnost metoda mašinskog učenja kao što su: Random Forest (RF), Artificial Neural Network (ANN), logistic, K-nearest neighbor (KNN), decision tree (DT), i Gaussian Mix Model (GMM). Prikupljeni podaci sadrže karakteristike farme i farmera, demografske podatke, troškove korišćenih institucija, troškove implementacije usvojenih tehnologija kao i prihode i rashode farme. Set podataka je podeljen u dva dela: 70% je korišćeno za trening modela, a 30% za testiranje.

U zajedničkom radu grupe istraživača sa Univerziteta iz Teherana i Univerziteta iz Viskonsina istraživana je mogućnost predviđanja oplemenjivačkih vrednosti za mlečna goveda korišćenjem veštačke neuronske mreže i neuro-fazi sistema [8]. Cilj rada je bio da istraži potencijal pomenutih metoda učenja, kako bi se dobile procene EBV iranske populacije mlečnih goveda. Podaci su uzeti od Iranskog centra za uzgoj životinja (ABCI, Teheran), a sadržali su 119.899 podataka o laktacijama Holštajn krava prvog pariteta koje su se telile između 22 i 36 meseca starosti u periodu od 1990. do 2005. Praćeni podaci se odnose na prinos mleka, udeo mlečne masti, i uslove okoline kao što su temperatura okoline, vlažnost vazduha i dužina fotoperioda.

Artificial Neural Network (ANN), - Kod primene ANN tokom faze učenja podaci prolaze mrežom i neuronska mreža prilagođava svoje parametre kako bi modelirala složene odnose između ulaznih i izlaznih promenljivih. Obuka ANN-a često olakšava otkrivanje ranije nepoznatih odnosa između ulaznih i izlaznih promenljivih, koji se mogu koristiti i u procesima klasifikacije i predviđanja [16]. Ova sposobnost prepoznavanja obrazaca mreža može se poboljšati različitim tehnikama kao što su: prethodni odabir ili predobrada podataka [17], pronalaženje optimalne mrežne arhitekture, određivanje odgovarajućeg broja ciklusa obuke, variranje kombinacija ulaznih promenljivih [18] i prilagođavanje vrednosti parametara učenja [19].

U istraživanju autora Shahinfar i dr. [8] korišćen je feed forward backpropagation multilayer perceptron koji sadrži 1 ulazni, 2 skrivena i 1 izlazni sloj. Čvorovi u ulaznom sloju odgovaraju po jednoj *explanatory* promenljivoj. Čvorovi u skrivenom sloju koriste tengens hiperbolikus kao funkciju aktivacije [20] i uzimaju ponderisani zbir svih ulaznih promenljivih, a izlazni čvorovi uzimaju ponderisani zbir svih čvorova u drugom skrivenom sloju i koriste istu funkciju aktivacije za izračunavanje izlazne vrednosti. Funkcija tangens hipebolikus daje vrednosti u intervalu od  $\pm 1$  i diferencirabilna je, što je pogodno u korišćenju u backpropagation algoritmu. Ova funkcija je pogodna i u proceni oplemenjivačkih vrednosti koje su uvek u nekom "±" intervalu, jer se računaju na osnovu odstupanja određene individue od proseka.

*Fazi logika* je oblik viševrednosne logike koja se bavi približnim (umesto preciznim) rezonovanjem i višestrukim vrednostima istine [8]. U ovoj logici se upotrebljavaju fazi skupovi, koji se sastoje od različitih kategorija, koje kvalitativno izražavaju delimičnu pripadnost elementa. Stepenn pripadnosti elementa određenom skupu definisan je funkcijom pripadnosti. Na primer, u radu autora Shahinfar i dr. [8] navodi se evidencija o proizvodnji mleka koja bi mogla da klasifikuje količinu mleka u pet kategorija kao: vrlo nisku, nisku, srednju, visoku ili vrlo visoku. Ovih pet kategorija bi bila predstavljena sa pet fazi skupova, a zapis za određenu kravu delimično pripada svakom od dva susedna skupa.

*Neuro-fazi sistemi* (eng. Neuro-Fuzzy Systems - NFS) su hibridni inteligentni sistemi koji kombinuju subjektivne karakteristike fazi logike sa strukturnim učenjem neuronskih mreža [8]. NFS su dizajnirani delom na osnovu stručnog znanja, a delom naučeni iz podataka.

---

Rad je objavljen na engleskom jeziku i nalazi se na IEEE Xplore na adresi <https://ieeexplore.ieee.org/document/9400672> za citiranje molimo vas koristite sledeći oblik:

L. Tarjan, I. Šenk, D. Pracner, D. Rajković and L. Štrbac, "Possibilities for applying machine learning in dairy cattle breeding," 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH), 2021, pp. 1-6, doi: 10.1109/INFOTEH51037.2021.9400672.



Ukupni model u radu autora Shahinfar i dr. [8] koristi NFS sa jednim skrivenim slojem i linearnim neuronom u izlaznom sloju koji jednostavno izračunava ponderisani zbir izlaza lokalno linearnih neurona. Optimizacija parametara vrši se metodom najmanjih kvadrata [21], a inkrementalni algoritam učenja zasnovan je na stablu lokalnih linearnih modela (eng. local linear model tree - LOLIMOT) jer pruža optimalan model sa maksimalnom generalizacijom i dobro se izvodi u aplikacijama predviđanja. U sprovedenom istraživanju [8] EBV su prvo procenjene korišćenjem BLUP modela (eng. conventional best linear unbiased prediction - BLUP) životinje za više osobina. Nakon ustanovljenog modela, za izradu ANN korišćen je višeslojni perceptron za predviđanje EBV na osnovu podataka o učinku jedinki kandidata za selekciju. Nakon toga korišćena je fazi logika za formiranje NFS-a. Evaluirano je ukupno 20 mreža za svaki *multilayer perceptron* i algoritam *stabla lokalnih linearnih modela*. U radu [8] je pokazano da se korišćenjem ANN i NFS pristupa, može jednoznačno prognozirati prinos mleka i masti u EBV. Za prinos mleka, korelacije između EBV izračunate upotrebom BLUP animal modela i predviđene preko ANN i NFS u radu [8] bile su 0,92, odnosno 0,93 a odgovarajuće korelacije za prinos masti bile su 0,93, odnosno 0,93. Ovo pokazuje da se sa velikom efikasnošću može proceniti približan EBV koristeći računarski efikasni algoritama iz oblasti veštačke inteligencije i mašinskog učenja.

Ovu prethodnu tvrdnju potkrepljuje i sledeći rad grupe istraživača Univerziteta iz Viskonsina [9], koji je u sklopu projekta podržanog od Nacionalnog instituta za hranu i poljoprivredu USDA (eng. USDA National Institute of Food and Agriculture, Washington, DC) istraživala mogućnost primene alternativnih algoritama mašinskog učenja za predviđanje ishoda osemenjavanja kod Holštajn goveda. Cilj je bio da se razvije jednostavan alat (aplikacija) koji će odgajivačima pomoći u donošenju odluka o upravljanju reprodukcijom. Implementirano je nekoliko različitih algoritama mašinskog učenja za predviđanje ishoda osemenjavanja pojedinih krava na osnovu fenotipskih i genotipskih podataka sa 26 farmi mleka. Konkretno, testirano je 5 tipova algoritama, i to:

- Naive *Bayes* (NB) koji je jedan od efikasnijih induktivnih algoritama mašinskog učenja i rudarenje podataka, koji je zasnovan na Bajesovom pravilu (eng. Bayes rule) [22] i predstavlja najjednostavniji oblik Bajesove mreže [9].
- *Bayesian network* (BN) predstavlja zajedničku raspodelu verovatnoće skupa promenljivih kao diskretni aciklični graf i skup uslovnih raspodela verovatnoće koje odgovaraju određenim karakteristikama i koje su vrlo pogodne za male i nepotpune podatke jer i u takvim situacijama mogu postići dobru tačnost predviđanja [9], [23]. Za razliku od NB, BN nema snažnu pretpostavku nezavisnosti između karakteristika [9]. Ove mreže su posebno korisne, jer se pomoću Bajesove teoreme lako može izračunati raspodela verovatnoće potomaka u odnosu na vrednosti njihovih roditelja, kao i obrnuto, tj. raspodela verovatnoće roditelja u odnosu na željene vrednosti njihovih potomaka [9]. Odnosno, BN može

da se kreće ne samo od uzroka do posledice, već i od posledica do uzroka [9], [24].

- Decision trees (DT) spadaju među jednostavne, intuitivne i lako razumljive algoritme mašinskog učenja koji su u širokoj upotrebi. To su modeli u obliku stabla sa proverama u unutrašnjim ;vorovima izvedenim na vrednostima karakteristika i oznakama klasa u listovima [9], [25]. Nove instance se klasifikuju prolaskom kroz čvorove i odgovarajuće provere unutar stabla, a oznakom završnog lista do kojeg dolazi svaka instanca predviđena je klasa za taj primer [9].
- *Bootstrap agregacija* (eng. Bootstrap Aggregation) odnosno (Bagging - BG) je zajednička metoda u kojoj će se generisati više verzija prediktora za pokretanje agregiranog (sumiranog) prediktora. BG u osnovi poboljšava performanse predviđanja gradeći nekoliko modela i puštajući ih da glasaju. Može se koristiti sa bilo kojom vrstom klasifikatora, a lako je primenjivo i računski izvedivo. Koristi nestabilnost modela za poboljšanje predvidljivosti.
- *Random Forest* (RF) je skup metoda, u kojima se obuka višestrukih klasifikatora pomoću  $n$  uzoraka bootstrapa iz skupa podataka za obuku kombinuje sa slučajnim odabirom podskupa karakteristika za generisanje svakog od tih klasifikatora [27], [28]. RF algoritam je vrlo sličan BG algoritmu, osim što u svakoj iteraciji izgradnje stabla bira nasumični podskup obeležja i deli instance na osnovu najinformativnije karakteristike. Budući da je zadatak ograničen na mali podskup svojstava i instanci, RF je računski efikasna tehnika koja se može koristiti sa visoko dimenzionalnim skupovima podataka. Jedna od najvećih prednosti RF-a i glavni razlog zbog kojeg je odabran u ovoj studiji je taj što je RF vrlo efikasan za procenu vrednosti koje nedostaju i može da održi visoku tačnost kada nedostaje veliki deo podataka; ovo je česta situacija kada se analiziraju zdravstveni podaci mlečnih goveda koje prijavljuju proizvođači.

Skup podataka, u radu autora Shahinfar i dr. [9], obuhvata desetogodišnji period od 2000. do 2010. sa preko 120 hiljada zapisa o uzgoju prvotelki i preko 195 hiljada zapisa o uzgoju Holštajn krava viših pariteta. NB, BN i DT u ovom radu pokazali su nešto slabije klasifikacione performanse. Postupak selekcije varijabli zasnovan na informacijama identifikovao je podatke kao što su: prosečna stopa začeca broj prethodnih (neuspelih) osemenjavanja, dužina laktacije kao i učestalost ketoze i mastitisa.

#### IV. ZAKLJUČAK

Na donošenje odluke farmera o tome da li će uzgajati određenu kravu ili ne, pozitivno bi uticalo postojanje informacije o očekivanom ishodu, jer bi na osnovu nje mogao da proceni profitabilnost odgajivačkog programa i neto prihod farme. U radu su analizirani rezultati ranijih istraživanja u ovoj oblasti, koja pokazuju mogućnost korišćenja algoritama veštačke inteligencije i mašinskog učenja za procenu oplemenjivačkih vrednosti sa velikom efikasnošću. Analizirani

su algoritmi zasnovani na veštačkoj neuronskoj mreži, fazi odnosno neuro-fazi logike, kao i algoritmi Naïve Bayes, Bayesian network, Decision trees, Bootstrap agregacija, K-nearest neighbor, Gaussian Mix Model i Random Forest koji pokazuju primenljivost i veliku efikasnost, veću od 90%.

Na ishod svakog uzgoja mogu uticati mnogobrojni faktori koje se razlikuju između farmi. Nijedna populacija nije potpuno zatvorena, već postoje emigracije i imigracije. Imigracija semena stranih bikova u Srbiji je veoma velika što otežava oplemenjivački rad. Zato je svaka razmena podataka sa okruženjem i usaglašavanje metoda prikupljanja istih na obostranu korist obe strane. Bilo bi stoga dobro za oplemenjivanje, da se pored saradnje i udruživanja odgajivača na nivou Srbije, u budućnosti to proširi i na region, pa i čitavu Evropu, kako bi uzgajivači bili konkurentni u odnosu na ostatak sveta. U cilju kreiranja jedinstvene baze u Srbiji, dosadašnji podaci su uveženi u bazu podataka koja objedinjuje informacije o poreklu i proizvodnim osobinama mlečnih goveda. Prvo su objedinjeni podaci iz pedigrea grla upisanih u glavnu matičnu evidenciju sa njihovim podacima o važnijim osobinama mlečnosti. U narednom periodu će se raditi i na objedinjavanju sa podacima o osobinama tipa i pojedinim reproduktivnim osobinama.

Sposobnost algoritama mašinskog učenja da se prilagode složenim odnosima u podacima i mogućnost procene vrednosti koje nedostaju čini ove algoritme vrlo pogodnim za ispitivanje reproduktivnih performansi kod mlečnih goveda. Postojanje alata, koje bi odgajivači i naučne institucije mogle da koriste za procenu oplemenjivačkih vrednosti bi dalo smisla za pravilno evidentiranje i vođenje svih bitnih podataka, pa i eventualnom automatizacijom nadzora i merenja. Dobar servis koji je koristan za sve bi mogao da obezbedi i pravilne (proverljive) podatke koji bi uticali i na poboljšavanje inteligencije aplikacije (servisa). Budući rad će biti fokusiran na testiranje ovih algoritama na postojećim podacima u Srbiji, kako za obuku nove mašinske inteligencije tako i za validaciju podataka, pomoću mreža koje su testirane sa podacima iz regiona.

#### ZAHVALNICA

Rad je podržan sredstvima Fonda za nauku Republike Srbije, kroz PROMIS program, za projekat sa akronimom BioITGenoselect, oznake: 6066512. Za sadržinu ove publikacije isključivo su odgovorni autori i sadržina ne izražava stavove Fonda za nauku Republike Srbije.

#### LITERATURA

- [1] FAO, "The future of food and agriculture – Trends and challenges," *The future of food and agriculture: trends and challenges*, 2017. [Online]. Available: <http://www.fao.org/3/a-i6583e.pdf>. [Accessed: 15-Jan-2015].
- [2] EnvironmentReports, "Is There Enough Food\_ for the Future?," *Food Matters*. [Online]. Available: <http://www.environmentreports.com/enough-food-for-the-future/>. [Accessed: 15-Jan-2021].
- [3] N. Alexandratos and J. Bruinsma, "World agriculture towards 2030/2050: the 2012 revision," *FAO Agricultural Development Economics Division*, 2012. [Online]. Available: <http://www.fao.org/3/ap106e/ap106e.pdf>. [Accessed: 15-Jan-2015].
- [4] UN, "World Population Prospects 2019," *Population Dynamics*, 2019. [Online]. Available: <https://population.un.org/wpp/Graphs/Probabilistic/POP/TOT/900>. [Accessed: 15-Jan-2021].
- [5] M. Elferink and F. Schierhorn, "Global Demand for Food Is Rising. Can We Meet It?," *Harvard Business Review*, 2016. [Online]. Available: <https://hbr.org/2016/04/global-demand-for-food-is-rising-can-we-meet-it>. [Accessed: 15-Jan-2021].
- [6] GOO, "Glavna odgajivačka organizacija u govedarstvu na teritoriji AP Vojvodine." [Online]. Available: <https://www.og.stocarstvo.edu.rs>. [Accessed: 15-Jan-2021].
- [7] G. Ostojić, S. Morača, M. Stanojević, M. Plavsiš, and D. Kučević, "Monitoring of Pasture in Grazing," *Journal of Mechatronics, Automation and Identification Technology*, vol. 5, no. 3, pp. 1–4, 2020.
- [8] S. Shahinfar, H. Mehrabani-Yeganeh, C. Lucas, A. Kalhor, M. Kazemian, and K. A. Weigel, "Prediction of Breeding Values for Dairy Cattle Using Artificial Neural Networks and Neuro-Fuzzy Systems," *Computational and Mathematical Methods in Medicine*, vol. 2012, p. 127130, 2012.
- [9] S. Shahinfar, D. Page, J. Guenther, V. Cabrera, P. Fricke, and K. Weigel, "Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms," *Journal of Dairy Science*, vol. 97, no. 2, pp. 731–742, 2014.
- [10] G. Mwanga, S. Lockwood, D. F. N. Mujibi, Z. Yonah, and M. G. G. Chagunda, "Machine learning models for predicting the use of different animal breeding services in smallholder dairy farms in Sub-Saharan Africa," *Tropical Animal Health and Production*, vol. 52, no. 3, pp. 1081–1091, 2020.
- [11] M. Cockburn, "Review: Application and Prospective Discussion of Machine Learning for the Management of Dairy Farms," *Animals*, vol. 10, no. 9. MDPI, 2020.
- [12] R. M. Bourdon, *Understanding animal breeding (2nd Edition)*. Prentice Hall, 2000.
- [13] E. Hare, H. D. Norman, and J. R. Wright, "Survival Rates and Productive Herd Life of Dairy Cattle in the United States," *Journal of Dairy Science*, vol. 89, no. 9, pp. 3713–3720, 2006.
- [14] M. Yamaguchi, M. Takayama, T. Nishisouzu, H. López, and O. Dochi, "155 Fertility of sexed semen in Holstein heifers and cows," *Reproduction, Fertility and Development*, vol. 32, no. 2, p. 204, 2020.
- [15] "International Bull Evaluation Service." [Online]. Available: <https://interbull.org/index>. [Accessed: 01-Mar-2021].
- [16] M. S. Sayeed, A. D. Whittaker, and N. D. Kehtarnavaz, "Snack Quality Evaluation Method Based on Image Features and Neural Network Prediction," *Transactions of the ASAE*, vol. 38, no. 4, pp. 1239–1245, 1995.
- [17] R. Ruan, S. Almaer, C. Zou, and P. L. Chen, "SPECTRUM ANALYSIS OF MIXING POWER CURVES FOR NEURAL

Rad je objavljen na engleskom jeziku i nalazi se na IEEE Xplore na adresi <https://ieeexplore.ieee.org/document/9400672> za citiranje molimo vas koristite sledeći oblik:

L. Tarjan, I. Šenk, D. Pracner, D. Rajković and L. Štrbac, "Possibilities for applying machine learning in dairy cattle breeding," 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH), 2021, pp. 1-6, doi: 10.1109/INFOTEH51037.2021.9400672.

- NETWORK PREDICTION OF DOUGH RHEOLOGICAL PROPERTIES," *Transactions of the ASAE*, vol. 40, no. 3, pp. 677–681, 1997.
- [18] R. S. Parmar, R. W. McClendon, G. Hoogenboom, P. D. Blankenship, R. J. Cole, and J. W. Dorner, "Estimation of aflatoxin contamination in preharvest peanuts using neural networks," *AGRIIS*, vol. 40, no. 3, pp. 809–811.
- [19] F. Salehi, R. Lacroix, and K. M. Wade, "EFFECTS OF LEARNING PARAMETERS AND DATA PRESENTATION ON THE PERFORMANCE OF BACKPROPAGATION NETWORKS FOR MILK YIELD PREDICTION," *Transactions of the ASAE*, vol. 41, no. 1, pp. 253–259, 1998.
- [20] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesús, *Neural Network Design (2nd Edition)*. 2014.
- [21] O. Nelles, *Nonlinear System Identification : From Classical Approaches to Neural Networks and Fuzzy Models*. Springer-Verlag Berlin and Heidelberg GmbH & Co. KG, 2000.
- [22] P. Domingos and M. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Machine Learning*," *Machine Learning*, vol. 29, pp. 103–130, 1997.
- [23] P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri, and P. Grunwald, "Comparing Predictive Inference Methods for Discrete Domains," in *In Proceedings of the sixth international workshop on artificial intelligence and statistics*, 1997, pp. 311–318.
- [24] L. Uusitalo, "Advantages and challenges of Bayesian networks in environmental modelling," *Ecological Modelling*, vol. 203, no. 3, pp. 312–318, 2007.
- [25] Y. Dong, Z. Fu, S. Stankovski, S. Wang, and X. Li, "Nutritional Quality and Safety Traceability System for China's Leafy Vegetable Supply Chain Based on Fault Tree Analysis and QR Code," *IEEE Access*, vol. 8, pp. 161261–161275, 2020.
- [26] L. Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [27] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [28] T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995, vol. 1, pp. 278–282 vol.1.

#### ABSTRACT

The development of machine learning techniques and the omnipresence of machine learning models in the last decade has opened diverse possibilities for their implementation in all research areas. Deep learning methods stand out among the most efficient tools, driven by the extensive growth of the acquired and available data in all areas, as well as the significant technological advances in graphical processors and the research in this field. This study aims to explore the potential of applying machine learning in dairy cattle breeding, by surveying similar examples in the field, and to consider the possible selection of a dataset for training the model with the objective of finding the quantitative description of the key parameters in breeding tasks.

#### Possibilities for applying machine learning in dairy cattle breeding

Laslo Tarjan, Ivana Šenk, Doni Pracner, Dušan Rajković, Ljuba Štrbac