

Dubinska analiza podataka za pametno upravljanje hladnim lancem

1st Tomislav Ivanovski
Odjel za informatiku
Sveučilište u Rijeci
Rijeka, Hrvatska
tivanovski00@gmail.com

I. UVOD

Tokom transporta hrane od proizvođača do krajnjeg potrošača postoji potreba za analizom podataka pomoću metoda strojnog učenja s ciljem otkrivanja neispravnih proizvoda [2] i nadziranja transporta brzo kvarljive hrane u hladnom lancu [3,4,5,6,7]. Koristeći RFID tehnologiju moguće je odrediti lokacije proizvoda tokom njegova transporta, dok pomoću mreže bežičnih senzora su dostupni podatci o temperaturi, vlažnosti i udjelu određenih plinova u kontejneru. Prikupljeni podatci najčešće su bežičnim putem poslani do centralne lokacije na kojoj se obrađuju pomoću nekih od metoda strojnog učenja [11,12], što rezultira uštedom resursa [13,14,15]. Kvar jednog ili više senzora ili gubitak komunikacije moguć je scenarij tokom transporta. Dubinskom analizom senzorskih podataka moguće je predvidjeti i nadomjestiti podatke koji nedostaju usljed opisanih događaja [17]. Računarstvo postaje jedan od ključnih čimbenika u razvoju agrokulture [15], gdje metode strojnog učenja sve više nalaze primjenu. Neke od primjena su detekcija bolesti biljaka ili kontrola korova [1]. Jedan od načina određivanja kvalitete hrane je pomoću računalnog vida [8]. Dubinska analiza senzorskih podataka osnovna je metoda klasifikacije i identifikacije proizvoda kod zaštite njihova porijekla [9] i povećanja prinosa usjeva [10]. Nakon berbe plodova obavezno je njihovo kvalitetno skladištenje radi očuvanja kvalitete. Plod nad kojim su izvođena mjerenja i analize opisane u ovom radu je plod breskve. Karakteristika ploda breskve je velika brzina sazrijevanja uspoređujući ga s plodovima drugih biljaka [18]. Da bi se smanjio broj otpisanih plodova usljed nezadovoljavajućih karakteristika potrebno je predvidjeti optimalno vrijeme branja ploda. Optimizacija transportnih uvjeta do krajnjeg konzumenta također je potrebna da bi se osigurala kvaliteta. Pomoću mjerenja karakteristika ploda breskve bez njena oštećenja [16] i stvaranjem modela strojnog učenja nad prikupljenim podacima moguće je predvidjeti optimalno vrijeme branja, što za rezultat smanjuje troškove proizvođača i osigurava potrebnu kvalitetu proizvoda kranjem potrošaču.

II. PREDVIĐANJE ZRELOSTI PLODA POMOĆU MODELA LINEARNE REGRESIJE

U ovom eksperimentu analiziraju se dva skupa podataka prikupljenih u dva uzastopna dana. Podatci unutar skupova mjerene su karakteristike plodova breskve. Čvrstoća ploda

dobar je pokazatelj zrelosti i potrebno je ispitati korelaciju između čvrstoće i ostalih izmjerenih karakteristika. U slučaju postojanja korelacije, prepoznavanje zrelosti breskve na temelju prikupljenih podataka je moguće. U poglavlju II-A opisana su dva modela linearne regresije s jednim prediktorom, dok su modeli linearne regresije s više prediktora opisani u poglavlju II-B. Skupovi podataka nad kojima su modeli trenirani se sastoje od sto observacija i dvadeset tri varijable. Neke od varijabli se koriste za izračun drugih, primjerice gustoća breskve se dobije kao kvocjent mase i volumena breskve. Na temelju ovisnosti između varijabli moguće je filtrirati one varijable čiji se podatci mogu naći u nekoj od drugih varijabli. Konačno, čvrstoća breskve kao pokazatelj njene zrelosti odabrana je kao odziv, dok su gustoća, impedancija i SSC/TA odabrani kao prediktori. SSC (engl. *soluble solids content*) označava udio šećera unutar uzorka soka ploda breskve, dok TA (engl. *titratable acidity*) označava aproksimaciju kiselosti otopine. Na Slici. 1 prikazan je *scatterplot* prvog skupa podataka. Moguće je primijetiti povezanost prediktora impedancije i varijable odziva. Prikupljene podatke

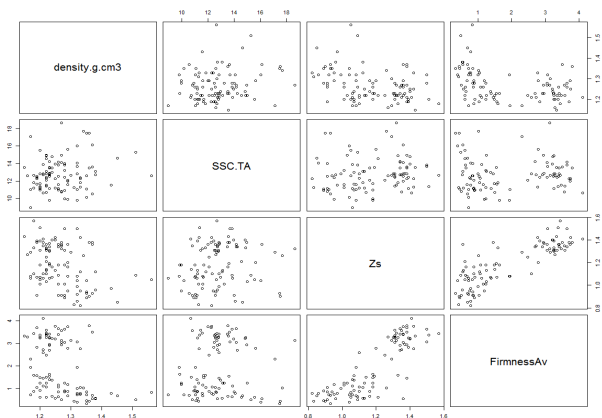


Fig. 1. Scatterplot prvog skupa podataka

potrebno je obraditi. Svaki skup podataka podijeljen je u dva pod-skupa, za treniranje i testiranje. Skup podataka za treniranje iznosi osamdeset posto originalnog skupa, dok skup za testiranje iznosi preostalih dvadeset posto. Min-max tehnika skaliranja je primjenjena nad skupovima. *Scatterplot* drugog skupa podataka prikazan je na Slici 2. Skup čini sto obser-

vacija mjerenih karakteristika istih plodova u razmaku od 24 sata u odnosu na prvi skup. Budući da postoje dva skupa

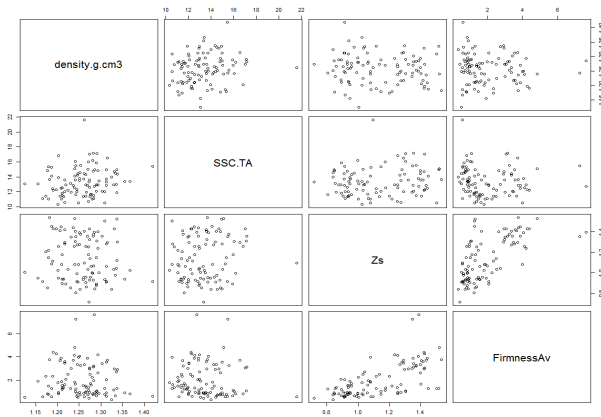


Fig. 2. Scatterplot drugog skupa podataka

podataka na kojima se izvodi analiza potrebno je izračunati varijance prediktora u svakom od skupova. Na slici. 3 prikazan je *box plot* varijabli impedancije, gdje crna točka predstavlja srednju vrijednost. Sa slike je moguće zaključiti da je varijanca varijable impedancije veća u drugom skupu podataka. Varijabla impedance je odabrana za prikaz jer daje snažna predviđanja zrelosti breskve, odnosno postoji čvrsta korelacija između impedancije i čvrstoće ploda breskve. Razlog zbog

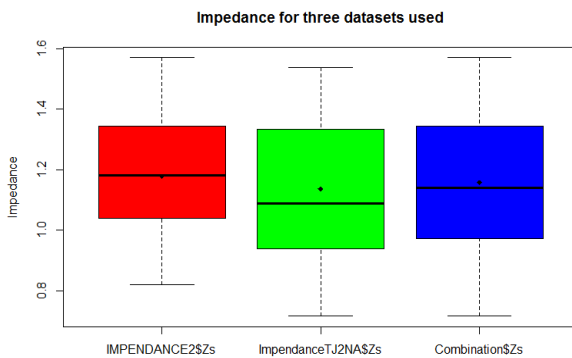


Fig. 3. Box plot varijabli impedancije

kojeg je veći iznos varijance varijable impedancije u drugom skupu podataka je sazrijevanje plodova. Kao što je prije navedeno drugi skup podataka dobiven je mjerenjem karakteristika istih plodova kao i u prvom skupu, ali sa razlikom od 24 sata. Plodovi breskve sazrijevaju brzo i kao rezultat varijanca varijabli drugog skupa podataka je veća. Pomoću ANOVA postupka moguće je utvrditi postojanje statističke razlike između prediktora i ishoda u različitim skupovima podataka. Primjenom ANOVA postupka s impedancijom kao prediktorom i čvrstoćom kao ishodom dobije se vrijednost *F-statistic* koja je tri puta veća za prvi skup podataka od drugog. Na temelju dobivenog, moguće je zaključiti da se skupovi podataka ne mogu kombinirati i da će model treniran nad

prvim skupom dati bolje rezultate od modela koji je treniran nad drugim.

A. Modeli linearne regresije s jednim prediktorom

U ovom su poglavlju opisana dva modela linearne regresije s jednim prediktorom koji su trenirani na dva različita skupa podataka. Kod treniranja modela impedancija je odabrana kao prediktor, a čvrstoća kao varijabla ishoda. Razlog odabira impedancije kao varijable prediktora je taj što ima najveću korelaciju u usporedbi s ostalim prediktorima. *RSE* (engl. *Residual Standard Error*) je korišten za indicaciju preciznosti modela [19].

$$RSE = \sqrt{\frac{RSS}{n - p - 1}} \quad (1)$$

Gdje je:

- *RSS*: Residual standard error
- *n*: Broj observacija
- *p*: Broj prediktora

Prvi model linearne regresije s jednim prediktorom treniran je nad prvim skupom podataka. Za svaki se model mjere vrijednosti *RSE*, R^2 i *F-statistic*. *F-statistic* ukazuje na ovisnost između varijabli prediktora i ishoda, dok R^2 daje informaciju o postotku varijance varijable ishoda koja je objašnjena prediktorom. Mala vrijednost *RSE* je poželjna jer ukazuje na veliku preciznost modela. Sl. 4 prikazuje model linearne regresije s jednim prediktorom koji je treniran nad prvim skupom podataka.

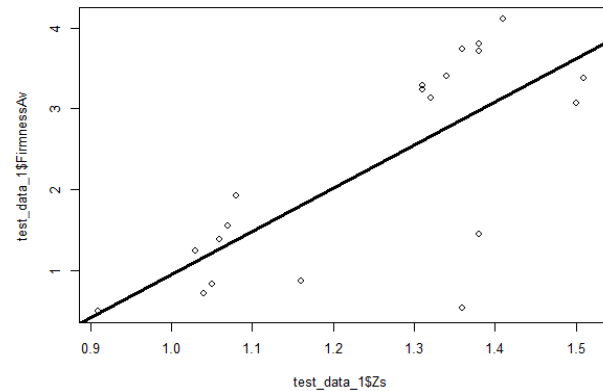


Fig. 4. Model linearne regresije s jednim prediktorom treniran nad prvim skupom podataka

Drugi model linearne regresije treniran je nad drugim skupom podataka. Dobivena vrijednost *F-statistic* i R^2 manjeg je iznosa od one dobivene kod modela treniranog nad prvim skupom podataka, dok je vrijednost *RSE* veća. Uspoređujući dobivene vrijednosti moguće je zaključiti da model treniran nad drugim skupom podataka daje lošije rezultate od modela treniranog nad prvim skupom. Varijanca varijabli prvog skupa podataka manja je od varijance drugog skupa, samim time model linearne regresije daje bolje rezultate na modelu čija je varijanca manja. Isti je zaključak dobiven u poglavlju II.

B. Modeli linearne regresije s dva ili više prediktora

U ovom su poglavlju opisana dva modela linearne regresije s više prediktora. Za varijable prediktora su odabrane impedancija, SSC/TA i gustoća, dok je varijabla odziva čvrstoća ploda breskve. Svaki je model treniran na posebnom skupu podataka, kao i u poglavlju II-A. Usporedbom izračunatih vrijednosti za RSE , R^2 i F -statistic dvaju modela moguće je doći do istih zaključaka kao i u prethodnom poglavlju. Model treniran nad skupom podataka s manjom varijancom varijabli daje bolje rezultate od modela koji je treniran nad drugim skupom. Također moguće je zaključiti da modeli linearne regresije koji koriste više varijabli prediktora ne daju značajno bolje rezultate od modela s jednim prediktorom. Razlog je taj što postoji velika korelacija između varijabli impedancije i čvrstoće, dok je za ostale prediktore korelacija manja.

III. PREDVIĐANJE ZRELOSTI PLODA KORIŠTENJEM NEURALNIH MREŽA

Dva modela umjetnih neuralnih mreža su trenirana nad dva skupa podataka, kao što je to slučaj kod modela linearne regresije. Čvrstoća ploda breskve dobar je pokazatelj njene zrelosti, stoga je potrebno predvidjeti vrijednost čvrstoće na temelju ostalih mjerenih karakteristika ploda. Rezultate je potrebno usporediti s rezultatima dobivnim u poglavljima II-A i II-B. U poglavlju III-A opisan je proces treniranja i testiranja modela neuralnih mreža nakon čega su modeli uspoređeni

A. Opis modela neuralnih mreža

Dva modela neuralnih mreža su trenirana nad dva različita skupa podataka koji su podijeljeni u omjeru 80:20 na skup za treniranje i skup za testiranje. Skupovi podataka su skalirani korištenjem min-max normalizacije i nakon treniranja su de-normalizirani s ciljem izračuna kvadratne pogreške. Oba modela sadrže dva neurona u skrivenom sloju. MSE (eng. *means square error*) daje informaciju o preciznosti metode treniranja modela. MSE izračunat nad test skupom podataka je korišten kao indikacija preciznosti metode treniranja neuronske mreže. MSE i RSE su korišteni za indikaciju točnosti modela, međusobnu usporedbu kao i za usporedbu s modelima linearne regresije. Slijedi jednadžba za MSE [19].

$$MSE = \frac{RSS}{n} \quad (2)$$

Gdje je:

- RSS : Residual sum of squares
- n : Broj observacija unutar skupa podataka

B. Model neuronske mreže treniran nad prvim skupom podataka

Prediktori su impedancija, SSC/TA i gustoća, dok je varijabla odziva čvrstoća ploda kao kod modela linearne regresije. Model neuronske mreže treniran nad prvim skupom podataka prikazan je na Slici 5. Skup podataka za treniranje čini osamdeset observacija, dok dvadeset observacija čini skup podataka za testiranje. Na slici. 6 prikazan je utjecaj svakog od prediktora na odziv. Sa slike je moguće vidjeti da varijabla

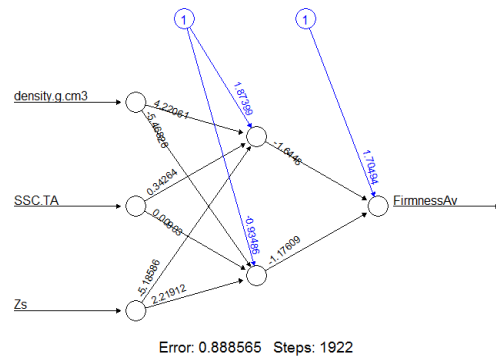


Fig. 5. Model neuronske mreže treniran nad prvim skupom podataka

impedancije ima najveći utjecaj na odziv što vrijedi za oba skupa podataka. RSE iznosi 0.61 što ukazuje da model dobro

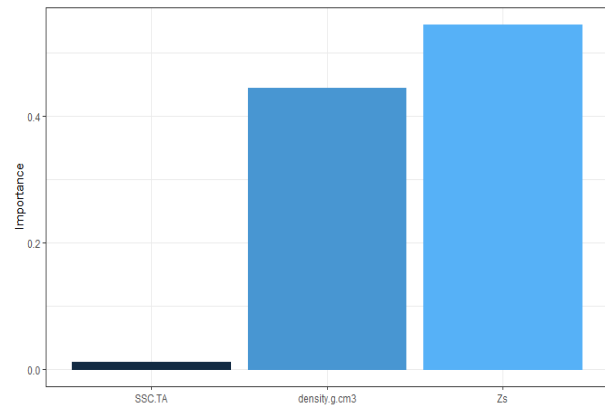


Fig. 6. Prikaz utjecaja prediktora na odziv

opisuje podatka nad kojima je treniran.

C. Model neuronske mreže treniran nad drugim skupom podataka

Struktura drugog modela neuronske mreže ista je kao kod prvog modela s dva neurona u skrivenom sloju. Model je treniran nad osamdeset observacija i testiran nad dvadeset. Dobiveni RSE ima vrijednost 1.072. RSE je izračunat kao srednja vrijednost deset iteracija treniranja modela. Razlike između izračunatih vrijednosti RSE unutar svake iteracije su male kod prethodnog modela, dok su znatno veća kod modela treniranog nad drugim skupom podataka. Moguće je zaključiti da model neuronske mreže bolje opisuje prvi skup podataka zbog manje varijance unutar samog skupa. Isti zaključak izveden je kod usporedbe modela linearne regresije. Prvi skup podataka trebao bi se koristiti za potrebe predviđanja, bilo pomoću ANN modela ili modela linearne regresije. Valja primijetiti da modeli neuralnih mreža daju bolje rezultate od modela linearne regresije, što je i očekivano.

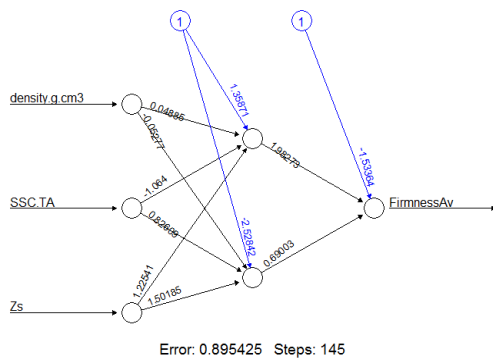


Fig. 7. Model neuronske mreže treniran nad drugim skupom podataka

IV. ZAKLJUČAK

Izvedeni eksperimenti pokazuju mogućnost predviđanja zrelosti plodova na temelju njihovih karakteristika dobivenih mjerenjem. Kao daljnji rad očekuje se nastavak prikupljanja podataka u suradnji sa ekspertima iz područja primjene. Time će se stvoriti bolji temelj za učenje kvalitetnijih modela, posebno neuronskih mreža, čija će se dodatna optimizacija izvršiti na temelju smjernica danih u [20].

REFERENCES

- [1] Nanyang Zhu, Xu Liu, Ziqian Liu, et al. (2018) Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. *Int J Agric & Biol Eng*, 2018; 11(4): 32–44.
- [2] Matindoust, Samaneh, et al. (2016) Food quality and safety monitoring using gas sensor array in intelligent packaging. *Sensor Review* 36.2: 169–183.
- [3] Zhang, J., Liu, L., Mu, W., Moga, L. M., & Zhang, X. (2009). Development of temperature-managed traceability system for frozen and chilled food during storage and transportation. *Journal of Food, Agriculture & Environment*, 7(3&4), 28-31.
- [4] Xiao, X., Qi, L., Fu, Z., & Zhang, X. (2013). Monitoring method for cold chain logistics of table grape based on compressive sensing. *Transactions of the Chinese Society of Agricultural Engineering*, 29(22), 259-266.
- [5] Wang, Xiang; He, Qile; Matetić, Maja; Jemrić, Tomislav; Zhang, Xiaoshuan, (2017) Development and evaluation on a Wireless Multi- Gas-Sensors System for improving traceability and transparency of table grape cold chain // *Computers and electronics in agriculture*, 135
- [6] Wang, X., Matetić, M., Zhou, H., Zhang, X., & Jemrić, T. (2017). Postharvest quality monitoring and variance analysis of peach and nectarine cold chain with multi-sensors technology. *Applied Sciences*, 7(2), 133.
- [7] Zhang, Yongjun; Wang, Wensheng; Yan, Liu; Glamuzina, Branko; Zhang, Xiaoshuan, (2019) Development and evaluation of an intelligent traceability system for waterless live fish transportation. // *Food Control*, 95, 283-297
- [8] Sun, Da-Wen, ed. *Computer vision technology for food quality evaluation*. Academic Press, 2016.
- [9] Granitto, PABLO M., et al. (2007) Modern data mining tools in descriptive sensory analysis: A case study with a Random forest approach. *Food quality and preference* 18.4 (2007): 681-689.
- [10] Majumdar, Jharna, Sneha Naraseeyappa, and Shilpa Ankalaki. (2017) Analysis of agriculture data using data mining techniques: application of big data. *Journal of Big Data* 4.1: 20.
- [11] R. Badia-Melis, U. Mc Carthy, L. Ruiz-Garcia, J. Garcia-Hierro, J.I. Robla Villalba. (2018) New trends in cold chain monitoring applications - A review. *Food Control* 86 (2018) 170-182
- [12] Kamilaris, Andreas, Andreas Kartakoullis, and Francesc X. Prenafeta-Boldú. (2017) A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture* 143: 23-37.
- [13] Sonka, Steve. (2016) Big data: fueling the next evolution of agricultural innovation. *Journal of Innovation Management* 4.1: 114.
- [14] Xiao, Xinqing; Li, Zhigang; Matetić, Maja; Brkić Bakarić, Marija; Xiaoshuan Zhang, (2017) Energy-efficient sensing method for table grapes cold chain management // *Journal of cleaner production*, 152.
- [15] Channe, Hemlata, Sukhesh Kothari, and Dipali Kadam. (2015) Multi-disciplinary model for smart agriculture using internet-of-things (IoT), sensors, cloud-computing, mobile-computing & big-data analysis. *Int. J. Computer Technology & Applications* 6.3: 374-382.
- [16] Satyam Srivastava, Shashikant Sadistap. (2017) Non-destructive sensing methods for quality assessment of on-tree fruits: a review. *Springer Science+Business Media, LLC* 2017, *Food Measure* DOI 10.1007/s11694-017-9663-6
- [17] Ganjar Alfian, Jongtae Rhee, Hyejung Ahn, Jaeho Lee, Umar Farooq, Muhammad Fazal Ijaz, M. Alex Syaekhoni. (2017) Integration of RFID, wireless sensor networks, and data mining in an e-pedigree food traceability system. Article in *Journal of Food Engineering*, May 2017 DOI: 10.1016/j.jfoodeng.2017.05.008
- [18] Shabir Ahmad Mir, Manzoor Ahmad Shah, Mohammad Maqbool Mir. *Postharvest Biology and Technology of Temperate Fruits*. Springer International Publishing AG, part of Springer Nature 2018
- [19] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer Science+Business Media New York 2013 (Corrected at th printing 2017)
- [20] Andrew Ng. (2019) *Machine Learning Yearning Technical Strategy for AI Engineers, In the Era of Deep Learning*. 2018 Andrew Ng. All Rights Reserved.