

Predviđanje vremena dostave hrane

1st Mateja Vuradin 2nd Matej Zubić 3rd David Iveković 4th Fran Ostroški 5th Ivo Veverec
mateja.vuradin@fer.hr matej.zubic@fer.hr david.ivekovic@fer.hr fran.ostroski@fer.hr ivo.veverec@fer.hr

Sažetak—Približno vrijeme dostave hrane je vrlo korisna informacija kupcima, ali i onima od kojih naručuju. U ovom radu istraženo je obogaćivanje podataka o dostavi hrane u najvećim gradovima Indije pomoću OSMnx paketa te njihova upotreba u raznim modelima strojnog i dubokog učenja s ciljem predikcije vremena dostave. Implementirane su neuronske mreže, (regularizirana) regresija, ElasticNet, SVR, stablo odluke, nasumična šuma i GBR. Uz MSE od ≈ 14.5 , kao najkorisniji modeli dobiveni su nasumična šuma i GBR. Slijedi ih jednostavna troslojna neuronska mreža s MSE od ≈ 17.51 .

Index Terms—dostava hrane, OSMnx paket, neuronske mreže, nasumične šume, XGB

I. UVOD I MOTIVACIJA

Vrijeme pandemije COVID-19 virusa značajno je utjecalo na živote ljudi, ali i na gospodarske i poslužiteljske djelatnosti. Brojni restorani završili su u dugovima, a neki su bili prisiljeni i zatvoriti se. Rješenje za mnoge poslodavce bilo je uvođenje mogućnosti online narudžbe i dostave hrane na kućnu adresu [1]. S druge strane, korisnici su mogli bezbrižno uživati u toplom obroku kod kuće. U ožujku 2020. procijenjeno je da ljudi troše čak 70% više na dostavu hrane i obroka, nego u ožujku godinu ranije, prije izbijanja pandemije [2].

Indija, kao poprilično naseljena zemlja, bila je teško pogođena pandemijom te je dostava hrane postala popularna [3]. Zadovoljstvo korisnika ovisilo je o nekoliko faktora. Neki od njih bili su kvaliteta hrane i kvaliteta dostave, npr. sigurnosne mjere protiv širenja zaraze, ponašanje dostavljača i vrijeme dostave hrane [4]. Pokazalo se da postoji određena količina dodatnih troškova (napojnica, naplata dostave...) i vremena koji su korisnici voljni čekati za dostavu pojedine hrane (uz potrošenu specifičnu svotu novaca). Tako npr. većina stanovnika SAD-a ne želi imati više od 10\$ dodatnih troškova ili čekati duže od 40 minuta [5]. Ovo upućuje kako je predviđanje vremena čekanja i dostave hrane u aplikacijama za dostavu hrane vrlo važno kako bi uopće došlo do stvaranja narudžaba.

Motiviran ovim problemima, ovaj rad nastoji razviti model za što točnije predviđanje vremena dostave hrane korisnicima u Indiji kako bi se omogućilo što bolje iskustvo i zadovoljstvo naručitelja.

Pregled postojećih istraživanja dan je u odjeljku II, dok su u odjeljku III prikazani korišteni materijali i metode. Odjeljak IV prikazuje dobivene rezultate koji su onda uspoređeni s postojećim rezultatima u odjeljku V.

II. PREGLED POSTOJEĆIH PRISTUPA

Online naručivanje hrane i njena dostava je široko istraživana tema. Od aplikacija za naručivanje hrane i njihovog

dizajna, preko faktora koji poboljšavaju zadovoljstvo korisnika do predikcije vremena dostave.

Ata Amrullah [6] je u svom radu koristila isti skup podataka. Uz samo dvije iskorištene značajke (Delivery_person_Age i Delivery_person_Ratings), dodana je i udaljenost restorana i mjesta dostave izračunata Haversinovom formulom. Treniranje i predikcija provedeni su uz LSTM neuronsku mrežu. LSTM, tj. duga kratkoročna memorija je vrsta rekurzivnih mreža koje omogućuju učenje dugoročnih ovisnosti među podacima, a rješavaju neke od postojećih problema RNN-a, npr. iščezavanje gradijenta.

Konvolucijske LSTM mreže koristili su i Xinlian Yu i drugi [7] u svom radu o predikciji broja narudžaba u određeno doba dana.

U svom članku, Saloni Jhalani [8] napravila je detaljnu analizu podataka, ekstrakciju značajki i isprobala brojne algoritme za treniranje i predikciju. Pristupila je problemu uz pomoć linearne regresije, stabla odluke, nasumične šume i XGB regresora.

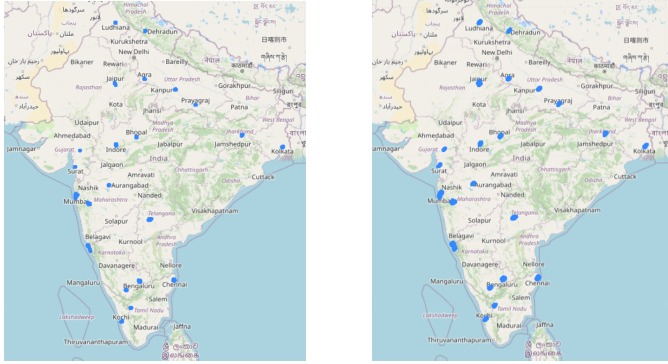
Jye Sawtell-Rickson je u svom članku [9] proučavao drugačiji skup podataka, ali je i sagledao problem dostave s novog gledišta. Naime, Sawtell-Rickson nije samo predviđao vrijeme dostave, već je i klasificirao količinu kašnjenja u dvije skupine: narudžbe s kašnjenjem 5 do 10 minuta koje korisnici toleriraju i one s kašnjenjem većim od 10 minuta. Za poboljšanje skupa podataka korišten je Google Maps API te je dobiveno "point-to-point" vrijeme putovanja. Korišteni algoritmi bili su nasumične šume, linearna regresija i ansambl.

III. MATERIJALI I METODE

A. Skup podataka

Korišteni podaci preuzeti su iz javno dostupnog skupa podataka na Kaggle stranici [10]. Skup se sastoji od 45453 podatka, od kojih svaki ima sljedeće attribute: ID, Delivery_person_ID, Delivery_person_Age, Delivery_person_Ratings, Restaurant_latitude, Restaurant_longitude, Delivery_location_latitude, Delivery_location_longitude, Type_of_order, Type_of_vehicle, Order_Date, Time_Orderd, Time_Order_picked, Weatherconditions, Road_traffic_density, Vehicle_condition, multiple_deliveries, Festival, City, Time_taken(min). Provedena je analiza i čišćenje podataka. Nedostajući podaci su uklonjeni, redundantni podaci nisu postojali, ali su uočene negativne vrijednosti za geografsku širinu ili dužinu što bi smještalo narudžbe izvan Indije. Ispravljanjem ovih podataka (zamjena s apsolutnim vrijednostima) i izbacivanjem onih izvan Indije, dobiveno je podudaranje lokacije s gradom koji je bilo moguće očitati iz identifikatora dostavljača. Ove lokacije prikazane su na Slici

1. Nije uočena velika linearna zavisnost vremena dostave i tipa vozila, dobi dostavljača ili tipa narudžbe.

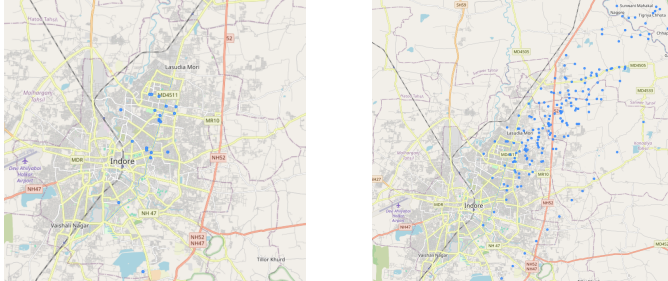


(a) Restorani

(b) Mjesta dostave

Slika 1: Lokacije nakon čišćenja

Provedeno je obogaćivanje podataka. Svakom podatku u skupu dodana je udaljenost između restorana i mjesta dostave. Nakon toga je iz identifikatora dostavljača izdvojen kod grada dostave i izračunate su granice dostave (*bounding box*) za svaki grad. One su se koristile kako bi se za svaki grad uz pomoć Python OSMnx paketa dohvatila i spremila mreža cesta. Na kraju se na temelju koordinata restorana, odnosno mjesta dostave, svakom podatku u skupu pridružio odgovarajući OSM (*OpenStreetMap*) čvor u grafu cesta (prikazano sa Slici 2).



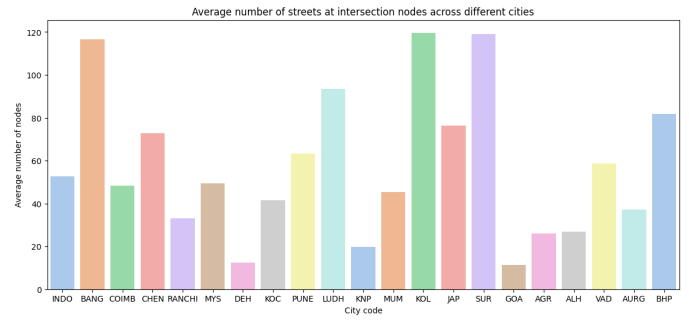
(a) Restorani

(b) Mjesta dostave

Slika 2: Čvorovi u gradu Indore

Također, na temelju ove mreže, početnog i završnog čvorova za pojedinačnu narudžbu pronađen je i izračunat najkraći put kroz grad. OSMnx paket iskorišten je i za dohvaćanje osnovnih informacija o svakom gradu, poput duljine cesta, gustoće cesta, ukupnog i prosječnog broja cesta, gustoće raskrižja prikazane na Slici 3 itd.

Nakon obogaćivanja, provedeno je ponovno čišćenje podataka te su podaci pripremljeni za korištenje u neuronskim mrežama. Kategorijske varijable zamijenjene su heuristički numeričkim vrijednostima. Tipu narudžbe kojoj je potrebna duža priprema (npr. meso), kao i bržim vozilima, lošijim vremenskim uvjetima, gušćem prometu i gradskim područjima pridružene su veće vrijednosti. Odabrane su samo najznačajnije značajke i podaci su standardizirani prije treniranja.



Slika 3: Gustoća raskrižja u gradovima

B. Neuronska mreža i ostali modeli

Za početak je isprobana jednostavna neuronska mreža s ulaznim, nula skrivenih slojeva i jednim izlaznim slojem, tj. obična linearna regresija. Kod ove mreže, kao i svih ostalih, korišten je Adam optimizator s promijenjivim koeficijentom učenja od 10^{-2} do 10^{-5} , ranim zaustavljanjem u slučaju konvergencije i srednjom kvadratnom pogreškom kao funkcijom gubitka. Broj epoha ograničen je na 300, a batch veličina na 1024.

Linearna regresija isprobana je i s polinomijalno transformiranim značajkama te se time povećao broj parametara (vidi Sliku 4).

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 1)	121
Total params: 121 (484.00 Byte)		
Trainable params: 121 (484.00 Byte)		
Non-trainable params: 0 (0.00 Byte)		

Slika 4: Linearna regresija s polinomijalnim značajkama

Sljedeća mreža je imala ulazni sloj, tri skrivena sloja i izlazni sloj koji je služio za predikciju vremena dostave. Skriveni slojevi veličina 64, 32, 16 sadržavali su sljedeće aktivacijske funkcije: ReLU, sigmoidalnu, ReLU. Korišteni optimizator bio je Adam optimizator s promijenjivim koeficijentom učenja od 10^{-3} do 10^{-5} , ranim zaustavljanjem u slučaju konvergencije i funkcijom gubitka srednjom kvadratnom pogreškom. Korišteno je 300 epoha i batch veličina 1024. Implementirana mreža prikazana je na Slici 5.

Mreža istih svojstava (vidi Sliku 6), ali s polinomijalnom matricom dizajna također je bila isprobana.

Isprobani su i brojni drugi načini predikcije. To su bili: linearna regresija, regularizirana regresija, Lasso, stablo odluke, nasumične šume, SVR, Elastic Net, Gradient Boosting Regressor. Za svaki algoritam provedena je i optimizacija hiperparametara.

Također, ponovljena je i upotreba jednostavne troslojne mreže, ali korištena je i osam foldna unakrsna validacija. Ovaj postupak koristio se s ciljem otkrivanja skrivenih značajki,

tj. otkrivanja hoće li treniranje na različitim podskupovima uzrokovati drugačiju kvalitetu predikcije.

Layer (type)	Output Shape	Param #
dense_2 (Dense)	(None, 64)	2560
dense_3 (Dense)	(None, 32)	2080
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 1)	17
Total params: 5185 (20.25 KB)		
Trainable params: 5185 (20.25 KB)		
Non-trainable params: 0 (0.00 Byte)		

Slika 5: Implementirana troslojna neuronska mreža

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 64)	7744
dense_7 (Dense)	(None, 32)	2080
dense_8 (Dense)	(None, 16)	528
dense_9 (Dense)	(None, 1)	17
Total params: 10369 (40.50 KB)		
Trainable params: 10369 (40.50 KB)		
Non-trainable params: 0 (0.00 Byte)		

Slika 6: Implementirana troslojna neuronska mreža s polinomijalnim značajkama

IV. REZULTATI

Treniranje podataka provedeno je na 80% podataka, a testiranje na ostalih 20%. 20% podataka za treniranje odvojeno je za validaciju. U tablici I su prikazani rezultati dobiveni za implementirane neuronske mreže.

Tablica I: Rezultati - algoritmi s neuronskim mrežama

Model	Test set		
	MSE	Avg. rel. diff.	R2 score
NN (0 skr. sloja)	36.5536	0.2048	0.5718
NN (0 skr. sloja + poli.)	29.6047	0.1807	0.6532
NN (3 skr. sloja)	17.5102	0.1399	0.7949
NN (3 skr. sloja + poli.)	18.6684	0.1445	0.7813

Iz Tablice I može se iščitati da se najmanja srednja kvadratna pogreška(MSE) na skupu za testiranje ostvaruje uporabom neuronske mreže s tri skrivena sloja (bez polinomijalne transformacije) i iznosi 17.5102. Najveća MSE dobivena je uporabom neuronske mreže bez skrivenih slojeva, tj. običnom linearnom regresijom te iznosi 36.5536.

Model	Train set		
	MSE	Avg. rel. diff.	R2 score
Lin. reg.	36.9451	0.2117	0.5790
Ridge	36.9593	0.2118	0.5789
Lasso	36.9911	0.2119	0.5785
ElasticNet	36.9907	0.2119	0.5785
SVR	21.6370	0.1503	0.7534
Stablo odluke	23.2166	0.1619	0.7354
Nasumična šuma	1.9916	0.0480	0.9773
GBR	10.6344	0.1145	0.8788

Model	Test set		
	MSE	Avg. rel. diff.	R2 score
Lin. reg.	36.4439	0.2706	0.5798
Ridge	36.4599	0.2065	0.5729
Lasso	36.4160	0.2065	0.5735
ElasticNet	36.4168	0.2065	0.5734
SVR	26.1213	0.1708	0.6940
Stablo odluke	23.4520	0.1602	0.7253
Nasumična šuma	14.5215	0.1288	0.8299
GBR	14.2210	0.1290	0.8334

Tablica II: Metrike alternativnih pristupa (bez poli.)

Model	Train set		
	MSE	Avg. rel. diff.	R2 score
Lin. reg. (poli.)	29.1780	0.1881	0.6675
Ridge (poli.)	29.1776	0.1881	0.6675
Lasso (poli.)	29.7107	0.1896	0.6614
ElasticNet (poli.)	29.8006	0.1900	0.6604
SVR (poli.)	19.6553	0.1449	0.7760
Stablo odluke (poli.)	17.5634	0.1401	0.7999
Nasumična šuma (poli.)	2.2528	0.0501	0.9743
GBR (poli.)	13.6556	0.1289	0.8444

Model	Test set		
	MSE	Avg. rel. diff.	R2 score
Lin. reg. (poli.)	29.0034	0.1821	0.6603
Ridge (poli.)	28.9943	0.1821	0.6604
Lasso (poli.)	29.2429	0.1829	0.6575
ElasticNet (poli.)	29.3392	0.1833	0.6563
SVR (poli.)	21.8629	0.1536	0.7439
Stablo odluke (poli.)	18.5982	0.1415	0.7821
Nasumična šuma (poli.)	16.4311	0.1349	0.8075
GBR (poli.)	16.1193	0.1359	0.8112

Tablica III: Metrike alternativnih pristupa (poli.)

Osim ovih algoritama, isprobani su i alternativni pristupi problemu, čije metrike su prikazane u Tablicama II i III.

Među alternativnim pristupima (bez polinomijalne transformacije) najmanja MSE na skupu za testiranje ostvaruje se uporabom GBR algoritma i iznosi 14.2210.

Nakon provedbe polinomijalne transformacije nad značajkama najbolji pristup je ponovo algoritam GBR, ovaj put uz nešto veću MSE vrijednost na skupu za testiranje, iznosa 16.1193. Najveća vrijednost MSE, postignuta za regulariziranu regresiju, smanjila se s prijašnjih 36.45 na 29.33, sada pri uporabi algoritma ElasticNet.

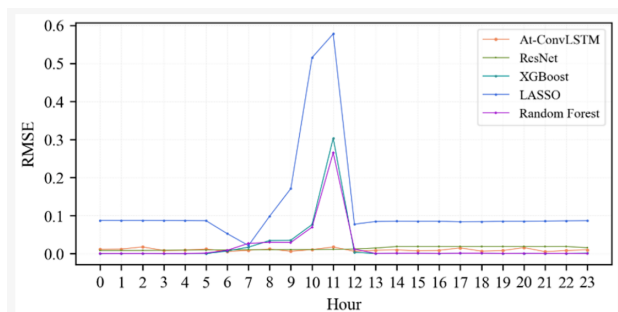
Za kraj, unakrsnom provjerom na troslojnoj mreži dobivena je vrijednost MSE 16.9230, uz standardnu devijaciju 0.7097.

Ovo dokazuje da nije bilo skrivenih varijabli koje bi utjecale na konačnu standardnu devijaciju.

V. USPOREDBA S POSTOJEĆIM RJEŠENJIMA

Spomenuti radovi većinom su imali drugačiji problem ili pristup te su ovdje zato samo prikazani ukratko njihovi rezultati.

U istraživanju [7] za predviđanje vremena naručivanja pokazano je da su neuronske mreže (At-ConvLSTM i ResNet) puno pouzdanije za predikciju nego jednostavni algoritmi poput nasumične šume ili XGB-a (vidi Sliku 7). U ovom radu, to je također bio slučaj, tj. već jednostavna neuronska mreža sa samo tri skrivena sloja je imala puno bolje predikcije od većine jednostavnih algoritama strojnog učenja.



Slika 7: Dobiveni RMSE po satu za različite modele - preuzeto iz [7]

Jhalani S. [8] je koristila isti skup podataka kao ovaj rad te je isprobala linearnu regresiju, stablo odluke, nasumičnu šumu i XGBoost. Unakrsnom provjerom uspostavljeno je da je XGBoost najbolji za predikciju vremena dostave hrane te je evaluacija na testnim podacima rezultirala srednjom kvadratnom pogreškom od 15.65. Ovaj rezultat je bolji od predloženih jednostavnih mreža, ali lošiji od nasumične šume i GBR-a predloženih u ovom radu.

U istraživanju [9] fokus je stavljen na točne i zakašnjele narudžbe, ali se u procesu računalo i vrijeme dostave. Pokazano je kako je najbolje korsitit ansambl u kombinaciji s informacijama o prometu iz Google Maps API-a (pogreška predviđanja oko 21 min), slično kao što je u ovom radu iskorišten OSMNx paket.

VI. ZAKLJUČAK

Istraživanje predikcije vremena dostave hrane pokazalo se vrlo korisno, kako za restorane, tako i za same korisnike. Konačni rezultati ukazuju na neuronsku mrežu s tri skrivena sloja bez polinomijalne transformacije kao najbolji model neuronske mreže, te GBR i nasumičnu šumu kao najbolje alternativne metode za predviđanje vremena dostave hrane. Također, istaknuta je važnost pravilne predobradbe i obogaćivanja skupa podataka koji se koriste kako bi se izbjegla pojava skrivenih značajki.

Neuronske mreže, imaju mnoštvo varijabilnih svojstava, od dubine i aktivacijskih funkcija do veličine batch-a. U ovom

radu je značajno poboljšanje u odnosu na većinu algoritama strojnog učenja postignuto već s jednostavnim plitkim mrežama. Ovo otvara mogućnosti za budući rad gdje bi se mogla istražiti predikcija nešto dubljim mrežama, slojevima izbacivanja itd. Osim toga, moguće su dodatne dopune skupa podataka, npr. sa prosječnom količinom sna dostavljača ili s podacima vezanim za ekonomiju regije u kojoj se obavljaaju dostave.

LITERATURA

- [1] Brewer P, Sebbby AG. The effect of online restaurant menus on consumers' purchase intentions during the COVID-19 pandemic. *Int J Hosp Manag.* 2021 Apr;94
- [2] Chen McCain S.L., Lolli J., Liu E., Lin L.C. An analysis of a third-party food delivery app during the COVID-19 pandemic. *British food journal* 2022 doi: 10.1108/BFJ-03-2021-0332.
- [3] Meena P, Kumar G. Online food delivery companies' performance and consumers expectations during Covid-19: An investigation using machine learning approach. *Journal of Retailing and Consumer Services.* 2022 Sep;68:103052. doi: 10.1016/j.jretconser.2022.103052. Epub 2022 Jun 15. PMID: PMC9355939.
- [4] Li Y, Yao P, Osman S, Zainudin N, Sabri MF. A Thematic Review on Using Food Delivery Services during the Pandemic: Insights for the Post-COVID-19 Era. *International Journal of Environmental Research and Public Health.* 2022; 19(22):15267. <https://doi.org/10.3390/ijerph192215267>
- [5] Pourrahmani E, Jaller M, Fitch-Polse D.T. Modeling the online food delivery pricing and waiting time: Evidence from Davis, Sacramento, and San Francisco. *Transportation Research Interdisciplinary Perspectives.* Volume 21 2023
- [6] Amrullah A. Food Delivery Time Prediction with LSTM Neural Network. *Analytics Vidhya.* 2023
- [7] Yu X, Lan A, Mao H. Short-Term Demand Prediction for On-Demand Food Delivery with Attention-Based Convolutional LSTM. *Systems.* 2023; 11(10):485. doi: <https://doi.org/10.3390/systems11100485>
- [8] Jhalani S. Food Delivery Time Prediction Model — Hands on Project. 2023. <https://medium.com/@salonijhalani.sj/food-delivery-time-prediction-model-77200d394f2b>
- [9] Sawtell-Rickson J. Food Delivery Time Prediction: How Does It Work? 2022. <https://builtin.com/machine-learning/food-delivery-time-prediction>. <https://doi.org/10.3390/systems11100485>
- [10] <https://www.kaggle.com/datasets/gauravmalik26/food-delivery-dataset?select=train.csv>