

# Predviđanje vremena dostave hrane

Mateja Vuradin

David Iveković

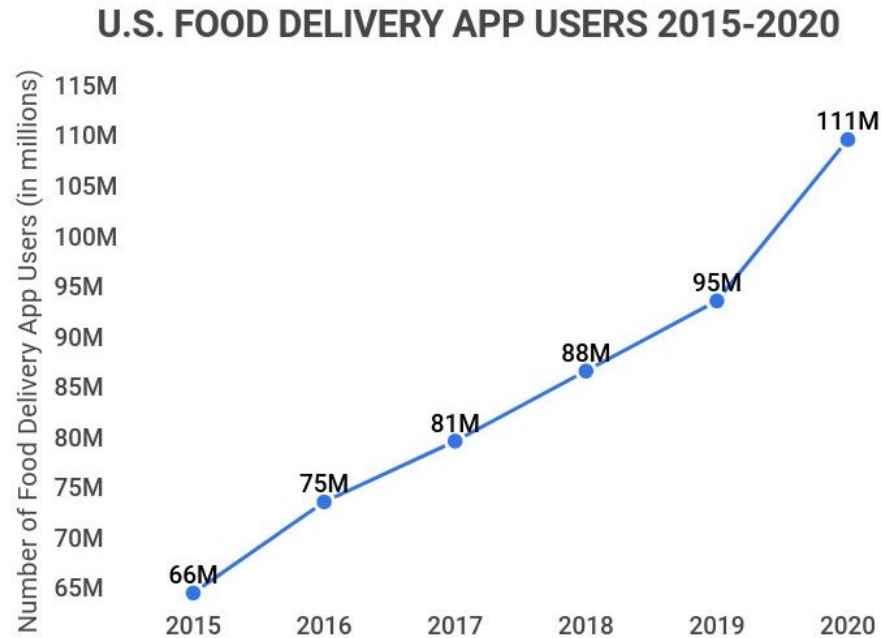
Ivo Veverec

Matej Zubić

Fran Ostroški



# Motivacija



Izvor: <https://www.zippia.com/advice/food-delivery-industry-statistics/>

- Dostava hrane sve je popularnija među korisnicima
- Zadovoljstvo korisnika uvelike ovisi o kvaliteti dostave
- Procjena troškova

# Skup podataka

## Kaggle

- 45 453 podatka

## Atributi:

- 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
- Weather conditions
- Road traffic density
- Vehicle condition
- Multiple deliveries
- Festival
- City
- Time taken(min)

Vizualizacija i  
analiza  
podataka



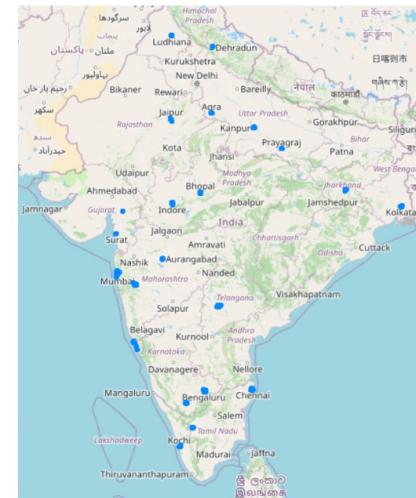
Čišćenje  
podatkovnog  
skupa



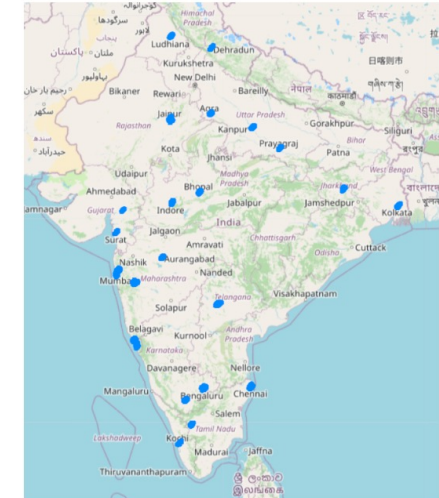
Obogaćivanje  
skupa sa novim  
podacima

# Analiza podataka i čišćenje

- Uklanjanje nedostajućih podataka
- Uočene negativne vrijednosti za geografsku širinu ili dužinu -> podaci izvan željene lokacije
- Nije uočena velika linearna zavisnost vremena dostave i tipa vozila, dobi dostavljača, tipu narudžbe ili udaljenosti.



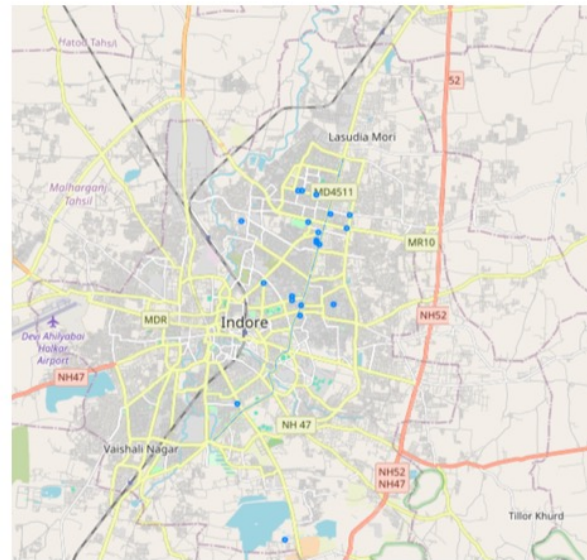
(a) Restorani



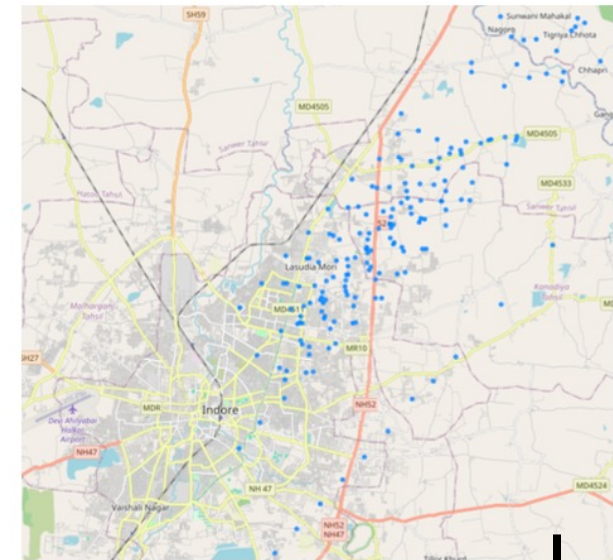
(b) Mjesta dostave

# Obogaćivanje podatkovnog skupa

- Udaljenost između restorana i mjesta dostave
- Iz identifikatora dostavljača izdvojen kod grada dostave
- Izračunate su granice dostave (bounding box) za svaki grad

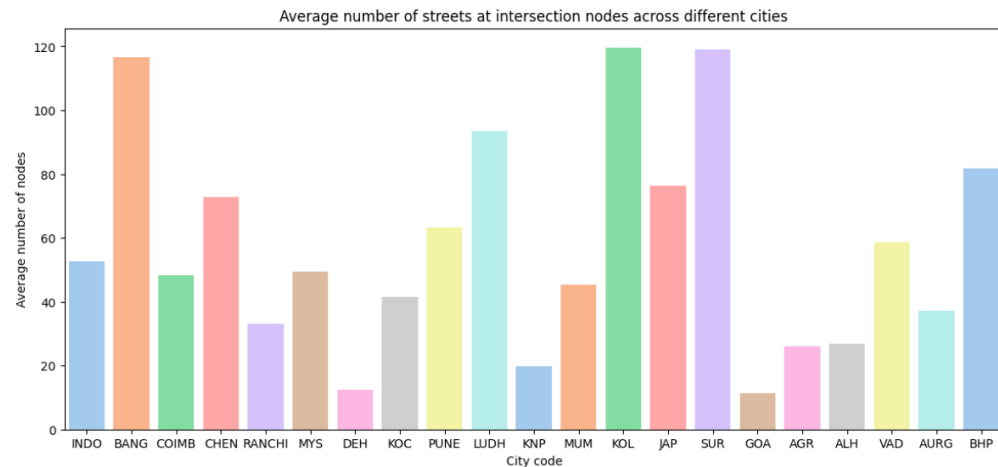


(a) Restorani



(b) Mjesta dostave

# Obogaćivanje podatkovnog skupa



Dohvaćanje informacija o gradu

Duljina ceste

Gustoća cesta

Ukupan broj cesta

Prosječan broj cesta

Zamjena kategoričnih varijabli za heuristike

Veće vrijednosti za

Tip narudžbe kojoj treba duža priprema

Brža vozila


Lošije vremenske uvjete

Gušći promet

Gradskim područjima

# Trening i evaluacija

Treniranje podataka provedeno je na 80% podataka, a testiranje na ostalih 20%



20% podataka za treniranje odvojeno je za validaciju



# Linearna regresija sa polinomijalnim značajkama

```
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)
```

---

- Adam optimizator s koeficijentom učenja 0.01
- srednja kvadratna pogreška kao funkcijom gubitka
- 300 epoha
- Batch veličine 1024

# Neuronska mreža

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)

---

- Ulazni sloj, tri skrivena sloja i izlazni sloj
- Skriveni slojevi veličina 64, 32, 16 i aktivacijskih funkcija ReLU, sigmoid, ReLU
- Adam optimizator 0.001
- Funkcija gubitka srednje kvadratne pogreške
- 300 epoha
- Batch size 1024

# Neuronska mreža sa polinomijalnim značajkama

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)		

# Ostali modeli i pristupi

---

- Regularizirana regresija
- Stablo odluke
- Nasumične šume
- SVR
- Elastic Net
- Gradient Boosting Regression
- Upotreba jednostavne mreže uz 8 foldnu unakrsnu validaciju

# Rezultati – neuronske mreže

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

# Rezultati – metrike alternativnih pristupa bez polinomijalnih značajka

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

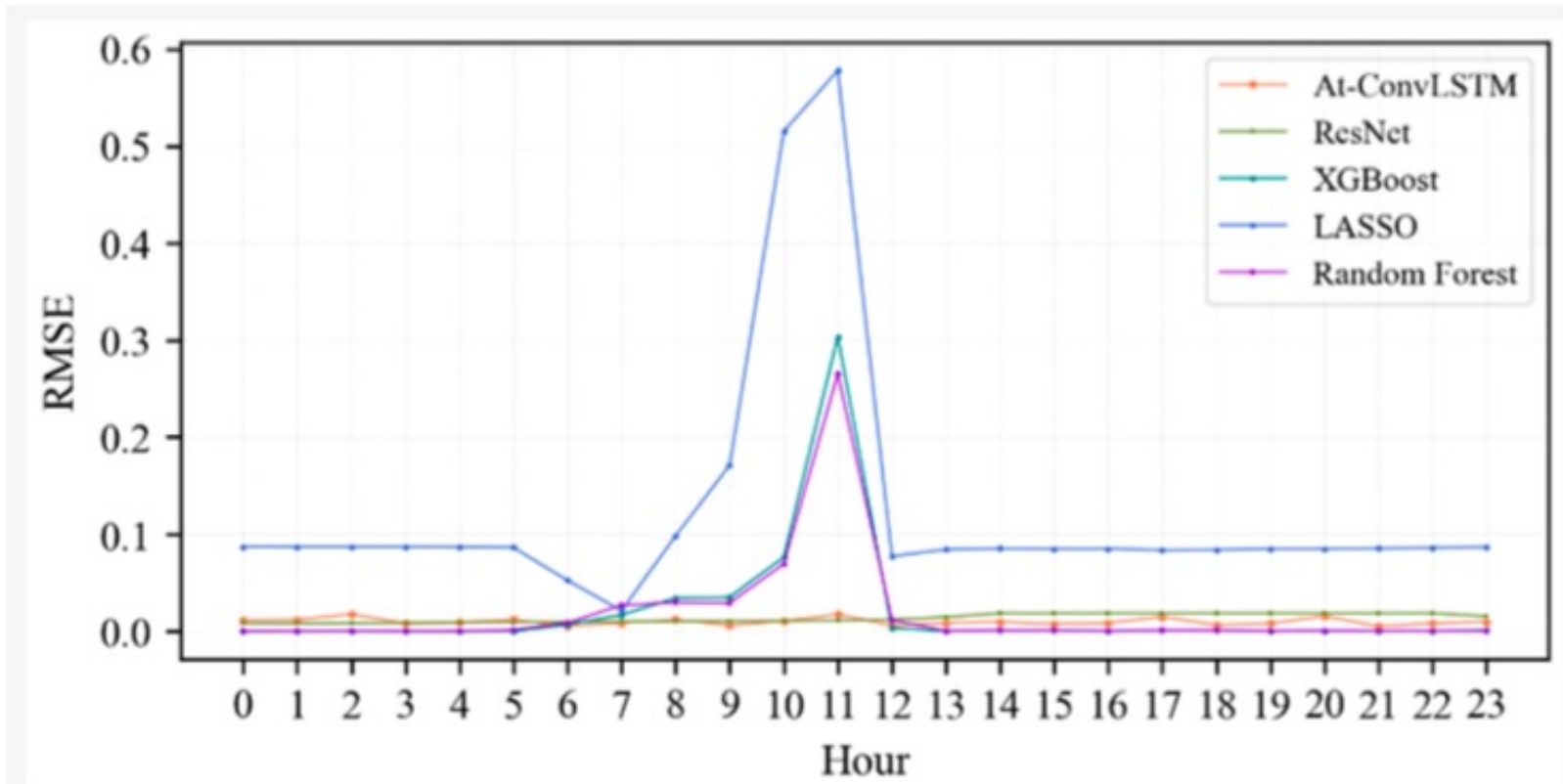
Model	Test set		
	MSE	Avg. rel. diff.	R2 score
Lin. reg.	36.4439	0.2706	0.5798
Ridge	36.4599	0.2065	0.5729
Lasso	36.4160	0.2065	0.5735
ElasticNet	36.4168	0.2065	0.5734
SVR	26.1213	0.1708	0.6940
Stablo odluke	23.4520	0.1602	0.7253
Nasumična šuma	14.5215	0.1288	0.8299
GBR	14.2210	0.1290	0.8334

# Rezultati – metrike alternativnih pristupa sa polinomijalnim značajkama

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

# Usporedba s postojećim rješenjima





Hvala na pažnji!