



# Primena metoda mašinskog učenja za predviđanje potražnje automobilskih rezervnih delova

master rad

mentor: dr Aleksandar Kartelj

Andđelka Milovanović  
1033/2020

Matematički fakultet  
Univerzitet u Beogradu

# Sadržaj

1. Uvod
2. Korišćene metode i metrike
  - 2.1 Korišćene metode
  - 2.2 Metrike i kategorički atributi
3. Podaci
4. Prikaz rezultata
5. Diskusija i zaključak
6. Pitanja

## Motivacija

- Problem predviđanja potražnje (eng. *demand forecasting*) je značajan i široko zastupljen problem u mnogim industrijama;
- Davanje najbolje procene potražnje za nekim proizvodom u budućnosti, pod zadatim ulaznim parametrima;
- Ušteda energije, ušteda materijala, smanjenje troškova, povećanje efikasnosti...
- Industrija popravki automobila: količine potrebnih rezervnih delova, isporuka delova, broj sati raspoloživih za popravku, broj dostupnih automehaničarskih radnji.

# Opis problema

## ① Problem kao vremenska serija:

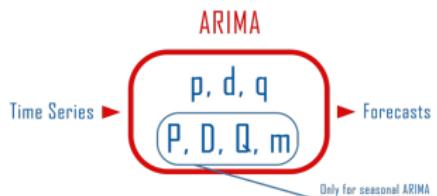
- dnevno uzorkovanje promenljive koja predstavlja potražnju;
- fokus na predviđanju 1 korak unapred;
- predviđanje se temelji samo na prethodnim vrednostima posmatrane promenljive.

## ② Problem posmatran kroz attribute:

- regresioni problem koji zavisi od nekoliko atributa;
- rešavan je ansambl metodom ekstremnog gradijentnog pojačavanja (XGBoost).

## Korišćene metode

- Autoregresivni model pokretnih proseka za integrisane vremenske serije (ARIMA)
- Prophet<sup>1</sup> (sezonalnost, praznici)
- XGBoost



(a) Arima

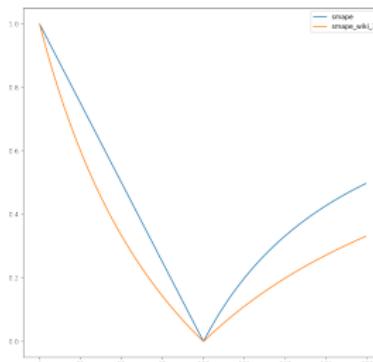


(b) Prophet

<sup>1</sup>Razvijen je od strane Facebook zajednice.

## Metrike i kategorički atributi

- Standardne metrike i izmenjena SMAPE metrika

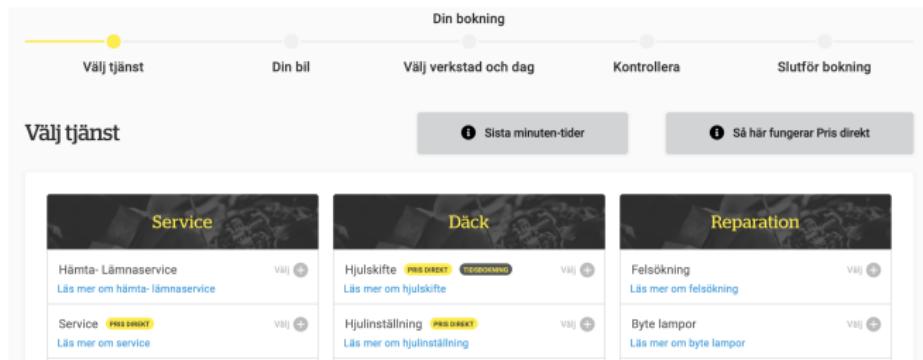


Slika: Izmenjena SMAPE metrika (smape) i podrazumevana SMAPE metrika (smape\_wiki3)

- Kodiranje uticajem (eng. *Impact encoding*)

## Podaci

- Švedska kompanija;
- Rezervisanje popravki automobila putem interneta (eng. *booking portal*);
- Početak rada krajem 2019. godine;



Slika: Primer internet portala za rezervacije

# Podaci

	vehicle_make	garage	reachable_population	year	week_of_year	number_of_competitors	x_unit_cost	x_units	demand_value	is_training_data
0	AE	CU	0.557521	2020	48	0.218750	0.741525	2.100000	0.666667	True
1	AE	CX	0.701217	2020	39	0.203125	0.595870	1.966667	0.666667	True
2	AE	DC	1.835690	2020	38	1.578125	1.324153	0.366667	0.666667	True
3	AE	DD	1.089601	2020	45	1.375000	1.101695	2.133333	1.333333	True
4	AE	DT	1.821892	2019	52	1.609375	1.191737	3.033333	0.666667	True
...	...	...	...	...	...	...	...	...	...	...
25799	BI	CS	0.701217	2021	8	0.203125	0.825212	6.600000	3.333333	True
25800	BI	CS	0.701217	2021	9	0.203125	1.006356	2.333333	2.000000	True
25801	BI	CS	0.701217	2021	10	0.203125	1.006356	1.666667	1.333333	True
25802	BI	CS	0.701217	2021	11	0.203125	1.006356	1.666667	1.333333	True
25803	BI	CS	0.701217	2021	12	0.203125	1.006356	1.000000	0.666667	False

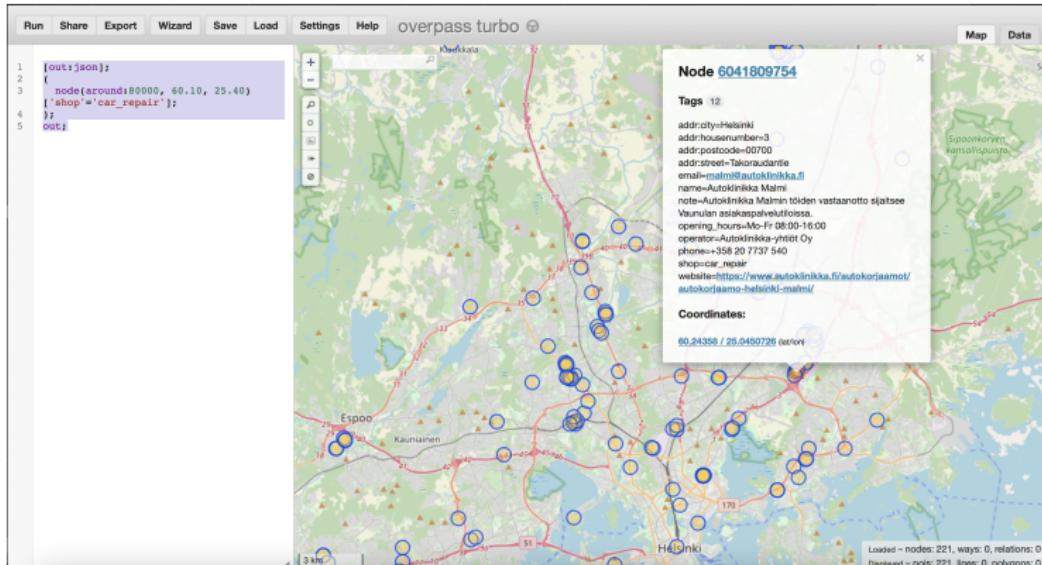
25804 rows × 10 columns

**Slika:** Primer skupa atributa i ciljne promenljive

## Obogaćivanje eksternim podacima

- Svaka automehaničarska radnja ima svoju geografsku širinu i geografsku dužinu;
- *reachable\_population*: suma populacije velikih regiona u opsegu koji predstavlja rastojanje pređeno za 40 minuta vožnje, brzinom 60 km/h, od lokacije radnje;
- *number\_of\_competitors*: broj drugih radnji u opsegu od 80km od automehaničarske radnje.

# Open Street Map



Slika: Open Street Map, Overpass API



## Podaci - vremenske serije

- jednostavne vremenske serije (univarijantne)

**Tabela:** Primer korišćenih podataka za dnevnu vremensku seriju (prvih 7 vrednosti)

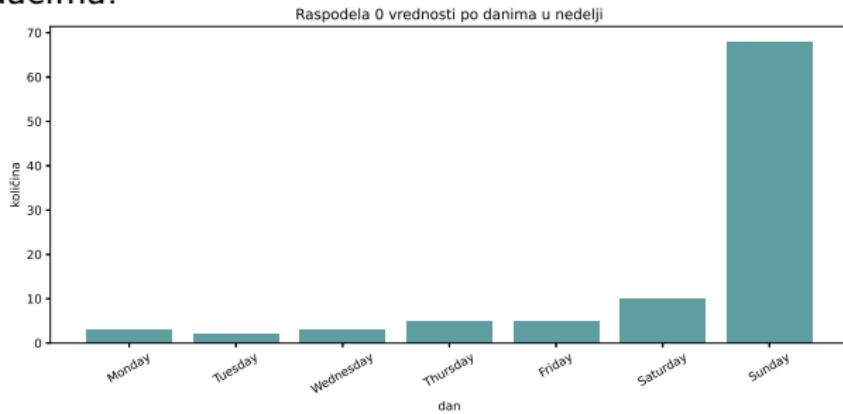
date	demand_value
2019-12-18	0.098
2019-12-19	0.245
2019-12-20	0.402
2019-12-23	0.490
2019-12-27	0.520
2019-12-28	0.020
2019-12-30	0.500

**Tabela:** Primer korišćenih podataka za nedeljnu vremensku seriju (prvih 7 vrednosti)

year_week	demand_value
201951	0.144
201952	0.295
202001	0.140
202002	0.333
202003	0.790
202004	0.921
202005	0.915

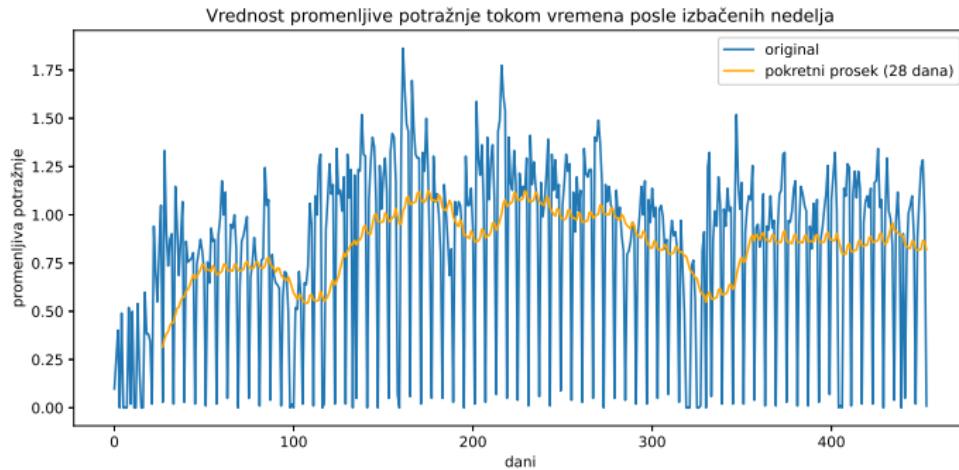
## Dnevni nivo

- podaci su grupisani na nivou cele države (Švedske);
- vremenski period: 18.12.2019 - 29.05.2021;
- izbačene su sve nedelje, nakon čega su ostala 454 dana u podacima.



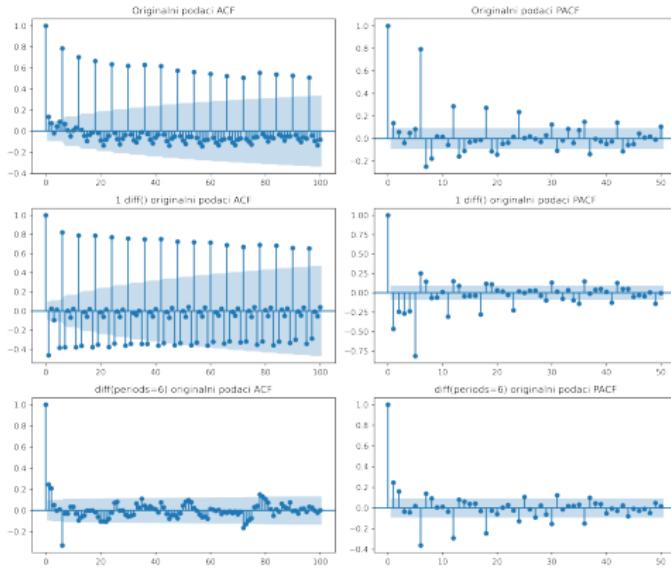
**Slika:** Raspodela nula vrednosti promenljive potražnje po danima u nedelji

# Dnevni nivo



Slika: Vremenska serija bez nedelja

# Dnevni nivo - ARIMA (stacionarnost, p, q)



**Slika:** ACF i PACF grafici. Na x-osi su pomeraji, a na y-osi vrednost korelacije

# Dnevni nivo - ARIMA

Tabela: SARIMA/ARIMA vrednosti AIC metrike

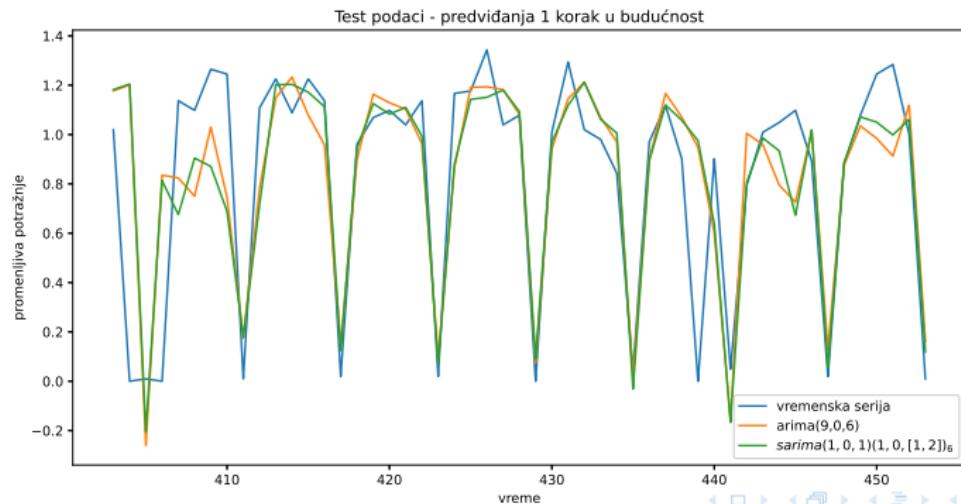
MODEL	AIC
ARIMA(1, 0, 1)	567.39
ARIMA(1, 0, 0)	598.30
ARIMA(6, 0, 6)	14.31
ARIMA(9, 0, 6)	14.63
ARIMA(1, 0, 1)(1, 0, [1,2]) <sub>6</sub>	3.81
ARIMA(2, 0, 0)(2, 0, 1) <sub>6</sub>	8.57

Tabela: SARIMA/ARIMA metrike evaluacije

MODEL	MAE	MSE	RMSE	SMAPE
ARIMA(1, 0, 1)	0.390	0.230	0.480	0.379
ARIMA(1, 0, 0)	0.383	0.216	0.465	0.376
ARIMA(6, 0, 6)	0.564	0.094	0.307	0.213
ARIMA(9, 0, 6)	0.209	0.095	0.309	0.224
ARIMA(1, 0, 1)(1, 0, [1,2]) <sub>6</sub>	0.192	0.095	0.308	0.208
ARIMA(2, 0, 0)(2, 0, 1) <sub>6</sub>	0.200	0.092	0.304	0.216

## Dnevni nivo - ARIMA

- Trening skup: svi dani pre 01.03.2021.
- Predviđanje 1 po 1 dan u budućnost

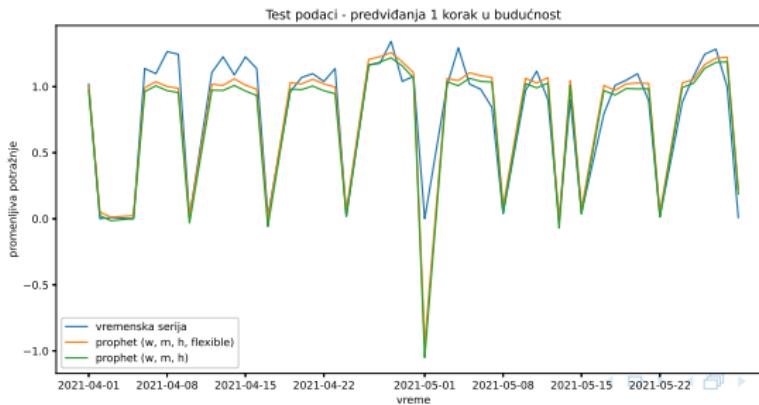


# Dnevni nivo - Prophet

- Uračunavanje praznika u model

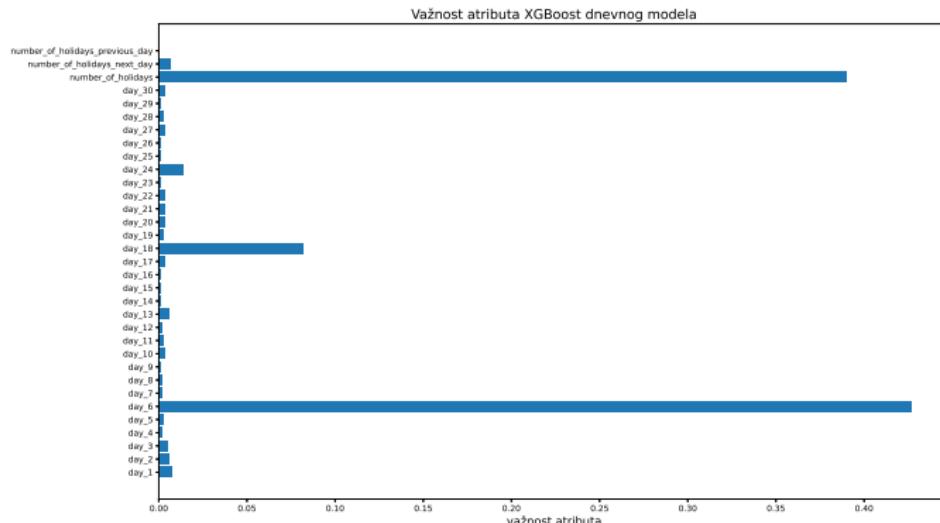
Tabela: Prophet metrike evaluacije

MODEL	MAE	MSE	RMSE	SMAPE
prophet (w, m, h)	0.121	0.038	0.196	0.141
prophet (w, m, h, flexible)	0.113	0.032	0.180	0.130
prophet (d, w, m, h)	0.119	0.035	0.188	0.139
prophet (d, w, m, h, flexible)	0.114	0.032	0.180	0.130



## Dnevni nivo - XGBoost

- Atributi su vrednosti promenljive potražnje iz prethodnih nekoliko dana
- 30 vrednosti iz prošlosti + 3 atributa za praznike



Slika: Važnost atributa kod XGBoost dnevnog modela

## Nedeljni nivo

- 76 dostupnih nedelja;

Tabela: Evaluacione metrike nad test podacima nedeljne vremenske serije

MODEL	MAE	MSE	RMSE	SMAPE
prophet (w, m, h)	0.180	0.052	0.227	0.164
prophet (w, m, h, flexible)	0.145	0.038	0.195	0.144
ARIMA(1, 0, 0)	0.155	0.041	0.203	0.151
ARIMA(1, 1, 1)	0.162	0.046	0.216	0.155
ARIMA(1, 0, 2)	0.159	0.042	0.205	0.154
ARIMA(12, 1, 0)	0.163	0.046	0.215	0.157
ARIMA(13, 0, 2)	0.162	0.044	0.210	0.158

Tabela: Metrike evaluacije nad test podacima nedeljnog XGBoost modela

MAE	MSE	RMSE	SMAPE
0.146	0.046	0.215	0.139

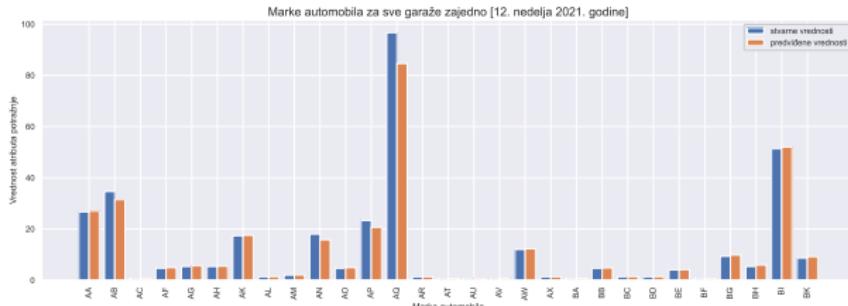
# XGBoost - veća granularnost

- Sitniji nivo predviđanja: nivo automehaničarskih radnji i marki automobila;
- Isprobana su 4 modela;

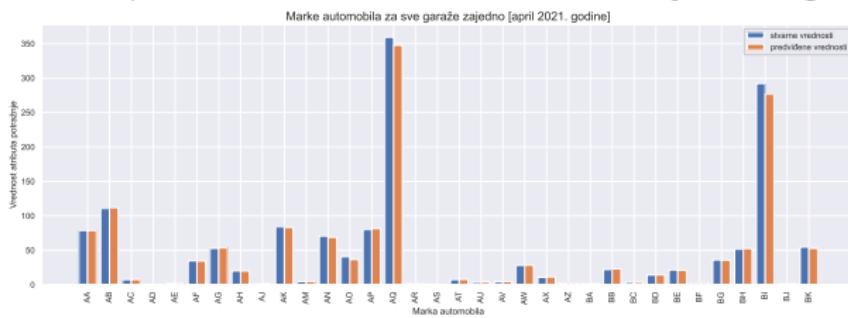
**Tabela:** XGBoost metrike evaluacije za 4 modela

MODEL	MAE	MSE	RMSE	SMAPE
year_week_garage_category	0.3044	0.257	0.507	0.315
year_month_garage_make	0.175	0.112	0.334	0.201
year_week_garage_make	0.183	0.129	0.359	0.207
year_month_garage_category	0.331	0.278	0.528	0.316

# XGBoost - veća granularnost



Slika: Raspodela marki automobila za 12. nedelju 2021. godine



Slika: Raspodela marki automobila za april mesec 2021. godine

# Diskusija i zaključak

- Nedovoljna količina podataka;
- Potencijalno nestandardni podaci: pandemija Kovid19 virusa;
- Problem sa malim radnjama;
- Kombinovanje zaključaka dobijenih ARIMA metodom sa XGBoost metodom;
- Obogaćivanje eksternim podacima (vremenska prognoza, procenjena cena automobila);
- Klasifikacija + regresija?

Hvala na pažnji.  
Pitanja?