

Univerzitet u Novom Sadu
Fakultet tehničkih nauka
Departman za industrijsko inženjerstvo i menadžment

METODE ISTRAŽIVANJA I EKSPLOATACIJE PODATAKA

KREIRANJE MODELA MAŠINSKOG UČENJA ZA PREDVIĐANJE TEMPERATURE
VAZDUHA OD PODATAKA PRIKUPLJENIH PREKO INFORMACIONOG SISTEMA *eFARM*

Profesor: dr Dubravko Ćulibrk
Asistent: Veljko Ilić

Student:
Jovan Milošević IT47/2018

Sadržaj:

1. Uvod	3
2. Tok izrade projekta.....	4
3. Rezultat analize	6
4. Zaključak.....	7

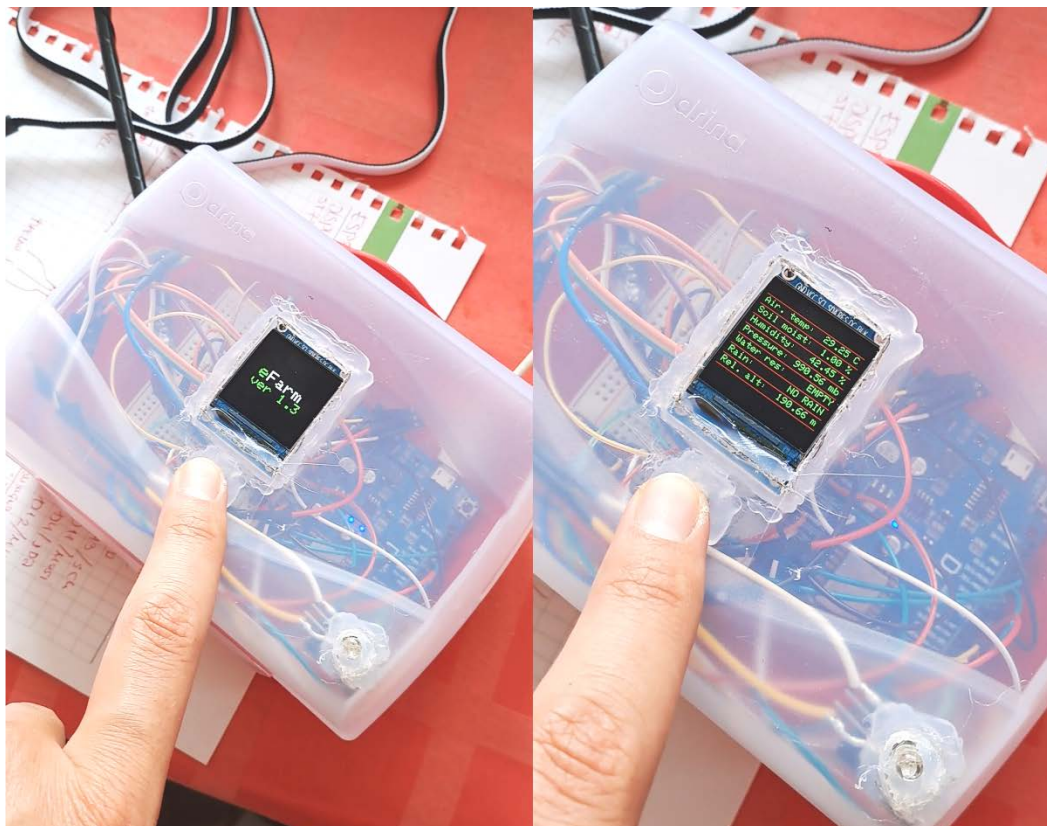
1. Uvod

Projekat iz predmeta „Metode istraživanja i eksploatacije podataka“ kreiran je sa ciljem da se implementira i obučni model za predviđanje temperature vazduha na osnovu drugih atmosferskih parametara koji će biti opisani u nastavku.

Inicijalni projekat koji je uslovio ideju za kreiranje modela mašinskog učenja je **eFarm** – informacioni sistem razvijen duže od dve godine, koji ima za cilj prikupljanje i obradu atmosferskih i drugih podataka u realnom vremenu na više lokacija u Republici Srbiji. Uređaji koji se koriste u pomenutom informacionom sistemu su bazirani na **ESP8266** i **ESP32** mikročipovima, tako da se podaci prikupljeni za različitih senzora u realnom vremenu šalju na udaljeni server korišćenjem veb servisa napisanog u **.NET ASP MVC** tehnologiji. Veb aplikacija za pristup podacima dostupna je na adresi <http://135.181.33.9:4200/>.

Parametri koji se šalju serveru sa uređaja su: vlažnost zemljišta, temperatura vazduha, vlažnost vazduha, pritisak vazduha, relativna nadmorska visina, stanje pumpe za navodnjavanje, i drugi, a sve u zavisnosti od tipa uređaja. Ideja je da različiti tipovi uređaja sa različitim senzorima budu deo istog informacionog sistema i da međusobno razmenjuju podatke u realnom vremenu.

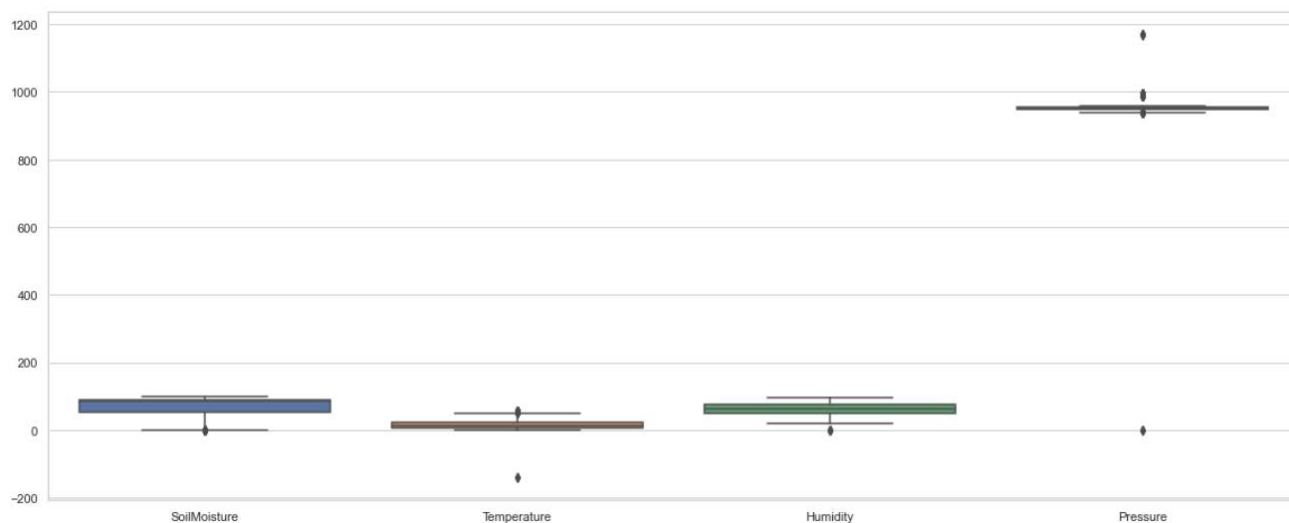
Za kreiranje projekta iz ovog predmeta korišćen je skup podataka prikupljenih sa dve geografske lokacije u Srbiji (Beograd i Užice) u periodu od 20. novembra 2021. godine do 27. juna 2022. godine. Prilikom demonstracije modela biće importovan najažurniji set podataka radi testiranja i provere funkcionalnosti modela. Bitno je napomenuti da podaci nisu prikupljeni kontinuirano svaki dan, već je bilo prekida zbog tehničkog održavanja uređaja, sistema, kao i dopuna eksternih baterija za napajanje uređaja.



Slika 1. Jedan tip uređaja korišćen za prikupljanje podataka baziran na ESP8266 mikročip

2. Tok izrade projekta

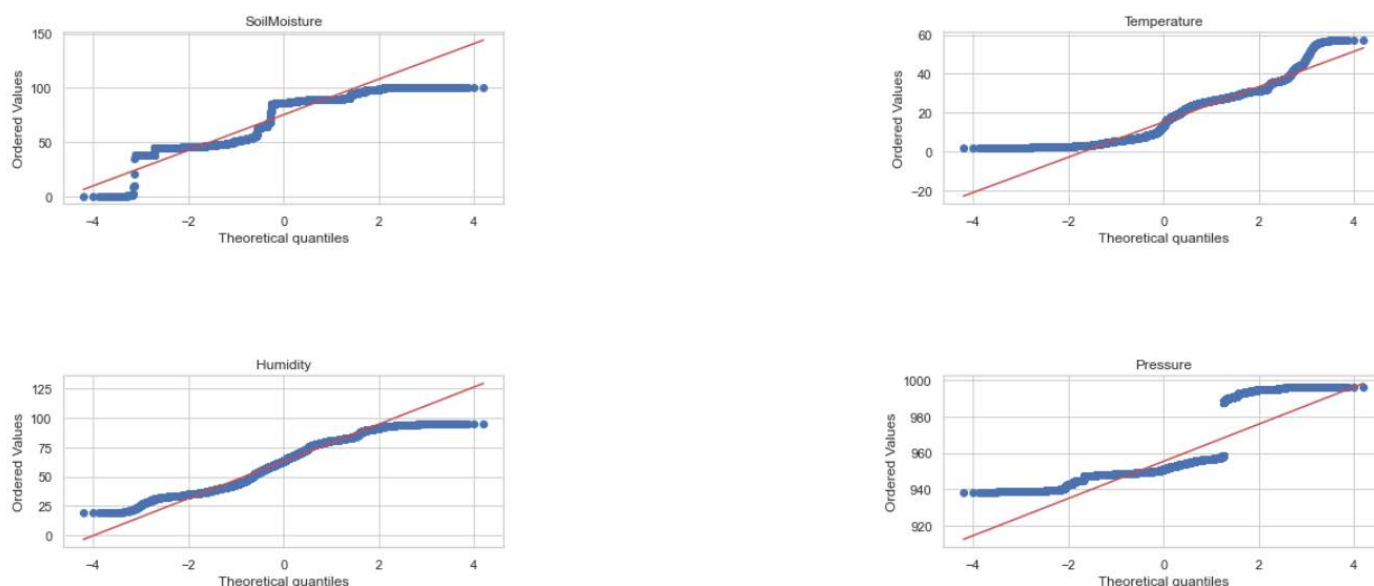
Set podataka korišćen za izradu projekta eksportovan je iz **Microsoft SQL Server** baze podataka, bez ikakve obrade, tako da sadrži i neadekvatne zapise koji su nastali kao proizvod loših očitavanja senzorike u složenim meteo uslovima. Da bi se uočile nepravilnosti u setu podataka, nacrtan je **Box plot** i zatim su obrisane neadekvatne vrednosti. Pre toga su podaci sortirani po datumu i vremenu, a zatim je ta kolona obrisana jer je označena kao nepotrebna. *Data frejm* je zatim reindeksiran.



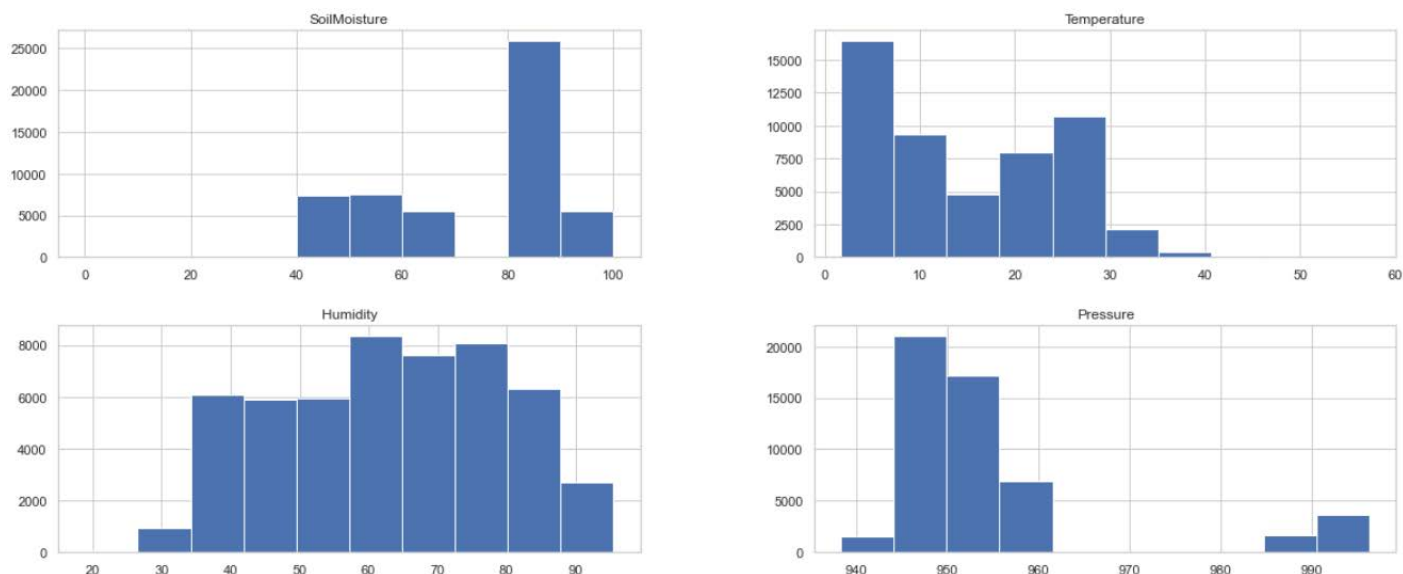
Slika 2. Box plot

Na **slici 2.** se vidi da postoje neprimerena očitavanja temperature, vlažnosti i pritiska vazduha.

Nakon **Box plot**-a nacrtani su **Q-Q plot** (**slika 3.**) zbog grafičkog upoređivanja između distribucija verovatnoće i histogram da bi se uočila relativna reprezentacija distribucije numeričkih podataka (**slika 4.**).



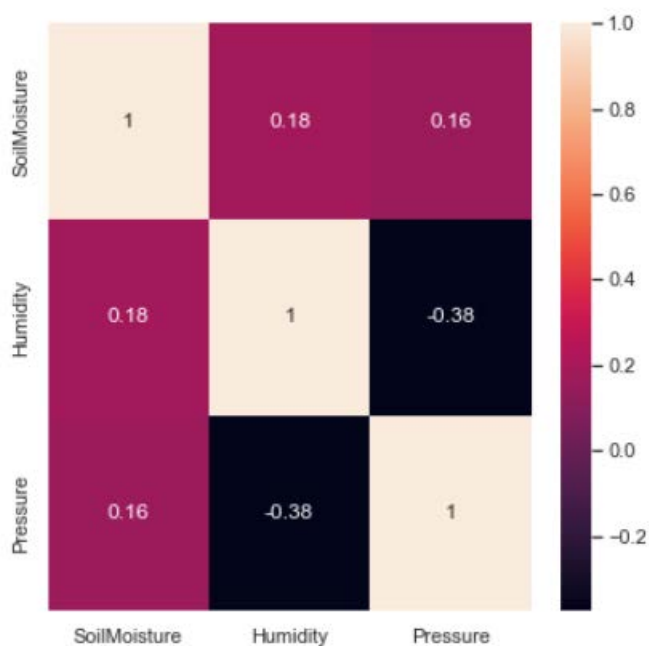
Slika 3. Q-Q plot



Slika 4. Histogram

Bitno je napomenuti da nisu vršene transformacije nad podacima radi kreiranja boljeg modela mašinskog učenja. Podaci korišćeni za predikciju temperature su: vlažnost zemljišta, vlažnost vazduha i pritisak vazduha.

U sledećim koracima podaci su podeljeni u testne i trening setove. 20% podataka je dodeljeno setu za testiranje. Izvršeno je skaliranje podataka i predstavljena matrica korelacije (slika 5.) trening seta podataka da bi se eventualno identifikovale značajne razlike.



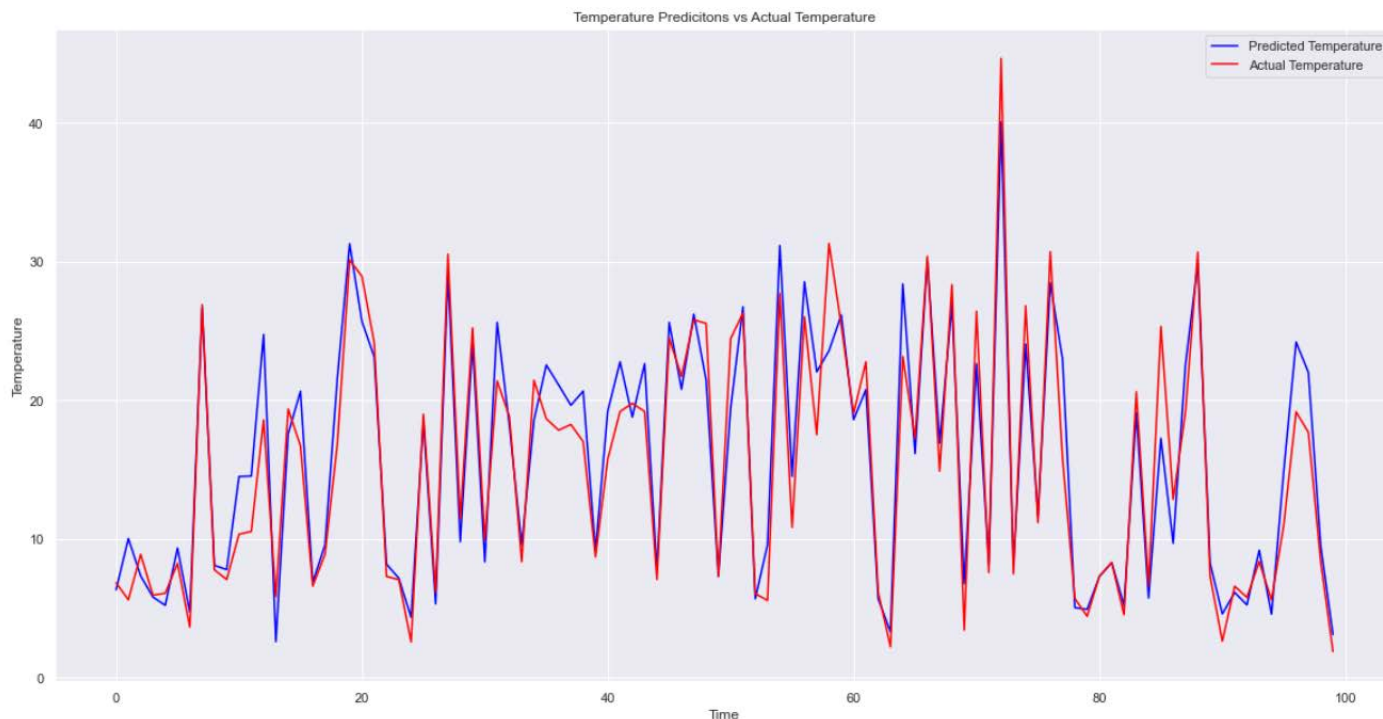
Slika 5. Matrica korelacije

Zatim su trening podaci „ubačeni“ u **PCA (Principal Component Analysis)** komponentu da bi se dobio niz numeričkih vrednosti ***explained_variance_ratio_***. U konkretnom primeru iskorišćene su sve tri komponente niza.

Nakon izvršenog treninga i predikcije, urađena je evaluacija trening modela (sa rezultatom preciznosti od **90,2098%**) i evaluacija testnog modela (sa rezultatom preciznosti od **89,9765%**)

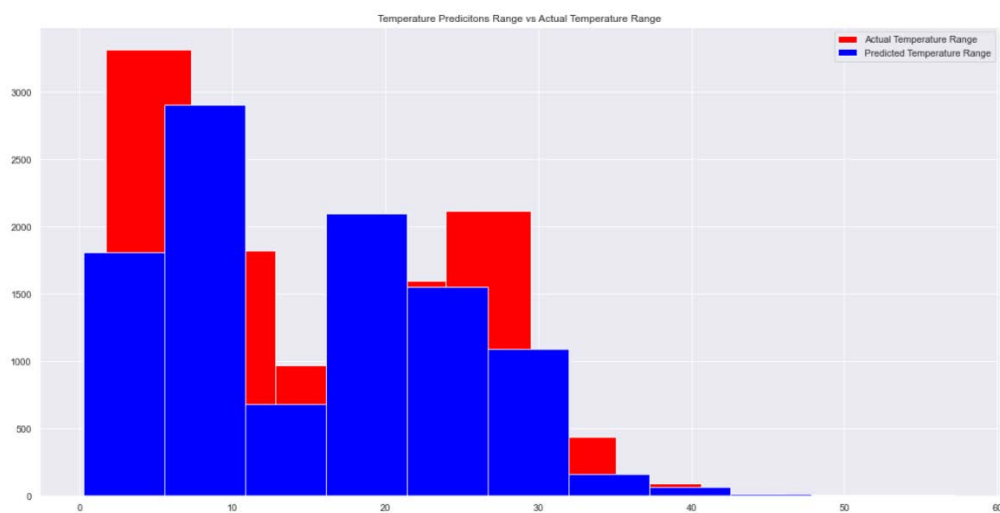
3. Rezultat analize

Kao rezultat analize kreiran je grafik koji uporedo prikazuje prave vrednosti temperature i vrednosti temperature koje model predviđa (**slika 6.**). Kao što se primećuje, iako ima odstupanja, model je prilično dobro istreniran. Uz korišćenje optimalnijeg seta podataka, sa još većeg broja uređaja, i sa više različitih lokacija, model bi verovatno efikasnije i efektivnije funkcionisao.



Slika 6. Uporedni prikaz stvarne temperature vazduha i temperature predviđene od strane modela

Na **slici 7.** prikazan je plot sa upoređivanjem temperaturnog domena stvarnih očitavanja i predviđanja koje je izvršio model.



Slika 7. Uporedni plot sa temperaturnim domenima stvarnih očitavanja i predviđanja modela

4. Zaključak

Kreiranje modela mašinskog učenja za predikciju određenih vrednosti kao deo eFarm sistema je od ranije bila želja autora. Proučavanjem alata i tehnika za izradu ovog projekta stečena su relevantna znanja da se model mašinskog učenja dalje unapređuje i usavršava, kako bi eventualno postao standardan deo eFarm informacionog sistema. Iako je set podataka sadržao preko 50 hiljada redova, za ozbiljniju analizu neophodno je više uređaja na različitim geografskim lokacijama koji bi duži vremenski period (minimum godinu dana) u realnom vremenu slali podatke na server.

Jovan Milošević,
IT47/2018

jovanmilosevicue@gmail.com
milosevic.jovan@uns.ac.rs

Jun, 2022.
