

# Predviđanje pojave boli u lumbarnom dijelu kralježnice strojnim učenjem

Grupa Sarsorama – Dario Bogović, Fran Borić, Iva Tutiš

## 1. Uvod

Istražujemo korespondenciju tjelesnih mjera u donjem dijelu kralježnice i deformacija s pojavom boli. Promatramo mjere kao što su nagib kralježnice, radijus zdjelice i slično. Skup podataka na kojemu istražujemo je „Lower Back Pain Symptoms Dataset“ preuzet sa stranice [www.kaggle.com](http://www.kaggle.com).

Bol može biti uzrokovana brojnim deformacijama koje uključuju: uklještenje velikih i malih živaca, pretjerano ukrućeni ili rastegnuti mišići, hernija diska, oštećenja na kostima i ligamentima.

Veliki dio populacija se susreće s (kroničnim) bolovima u donjem dijelu kralježnice barem u nekom trenutku u životu. Naš cilj je pokušati identificirati ljude koji imaju potencijal razvitka boli u lumbalnom dijelu kralježnice obzirom na tjelesne mjere.

## 2. Cilj i hipoteze

Cilj je nadziranim učenjem pronaći funkciju koja će moći predvidjeti hoće li osoba razviti bol u donjem dijelu kralježnice.

Naš skup podataka se sastoji od podataka prikupljenih od 310 različitih ljudi. Osim tjelesnih mjera zabilježen je podatak „State“ koji poprima dvije vrijednosti: „Normal“ i „Abnormal“. Konkretno, naš algoritam će na osnovu preostalih podataka moći donijeti

zaključak o podatku „State“ i time predvidjeti pojavu boli u donjem dijelu kralježnice.

Omjer „Abnormalnih“ naspram „Normalnih“ state-ova u bazi podataka je 210/100 što je, premda pesimistički i deprimirajuće, realni omjer ljudi budući da 40% svjetske populacije u nekom trenutku života ima bolove u donjem dijelu kralježnice, a taj broj raste i do 80% u „razvijenom svijetu“.

## 3. Pregled dosadašnjih istraživanja

U drugim istraživanjima veliki fokus je bio stavljen i na grupiranje podataka („clustering“).

Mi ga nećemo raditi jer nam takav način obrade podataka daje odgovor na pitanja: „Koji su najveći skupovi deformacija koji dovode do boli u leđima?“ i „Koji su najčešći skupovi deformacija koji dovode do boli u leđima?“, a pitanje na koje mi tražimo odgovor je direktnije i jednostavnije: „Hoće li osoba razviti bol u donjem dijelu leđa?“ Iz tog razloga radimo samo regresiju podataka.

## 4. Materijali, metodologija i plan istraživanja

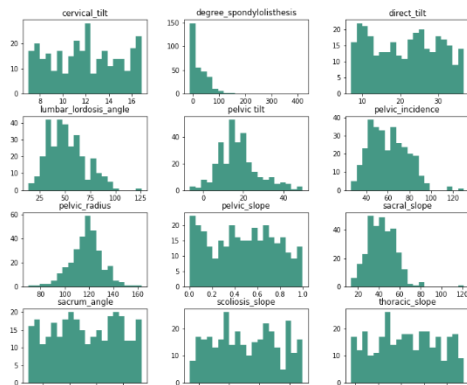
U projektu ćemo samo koristiti podatke iz ranije spomenute baze podataka gdje će nasumično određenih 70-80% biti ostavljeno za proces

učenja, a preostalih 20-30% za testiranje funkcije.

Time dobivamo dvije (iste) funkcije koje su trenirane na različitim bazama koje su različitih veličina.

Analizom podataka smo već utvrdili da podaci u gotovo svim dimenzijama nisu normalno distribuirani te da imamo veliki broj outlier-a u svakoj dimenziji. Osim toga, provjeravamo korelaciju među dimenzijama.

U bazi se nalazi 13 dimenzija numeričke vrijednosti i jedna dihotomna dimenzija.

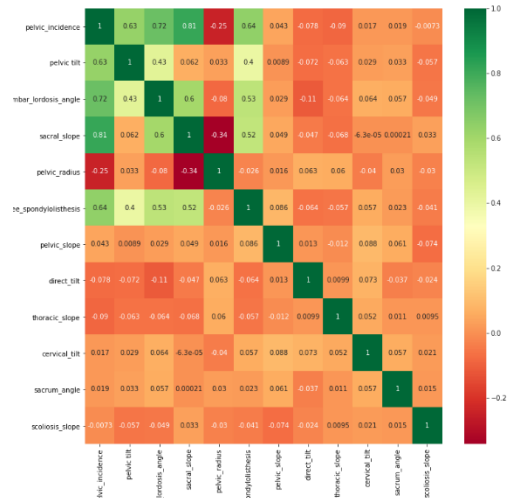


Slika 1. Histogrami podataka unutar svake dimenzije

Nakon eksploratorne analize podataka primjenjujemo logičku regresiju nadziranim učenjem budući da je rezultat funkcije binaran: „normal“ ili „abnormal“. To jest, radi se o binarnoj klasifikaciji.

Uspješnost rezultata ocjenjujemo na testnom dijelu baze podataka. Ukoliko će funkcija moći procijeniti normalnost („state“) kralježnice osobe na osnovu samih podataka i dobiti rezultat koji je upisan u tablici, utoliko ćemo taj rezultat smatrati pozitivnim i ispravnim.

Prikupljene rezultate jedne i druge funkcije ćemo statistički obraditi i usporediti te na



osnovi rezultata donijeti zaključak o potrebnoj veličini skupa podataka za testiranje i zaključak o tome koja je funkcija bila točnija.

Slika 2. Korelacija podataka u svakom paru dimenzija

Za kraj, ako ćemo biti u mogućnosti pristupiti podacima (obzirom na COVID-19 situaciju), testirat ćemo funkciju na mjerama očitanih s 3D skena kolege Bogovića (koji se skenirao za potrebe istraživanja provedenih na Tekstilno tehnološkom fakultetu u Zagrebu).

## 5. Očekivani rezultati predloženog projekta

Očekujemo da će ona funkcija koja je imala veći skup podataka za treniranje imati veću točnost budući da radimo na relativno malom skupu podataka.

Slični radovi koji su koristili logičku regresiju su, u najboljem slučaju, trenirali funkciju koja je bila točna u 87% slučajeva.

Nadamo se da ćemo uspjeti trenirati funkciju s točnosti od 80%.

## 6. Literatura

<https://web.math.pmf.unizg.hr/nastava/su/>

<https://www.fer.unizg.hr/predmet/su>

<https://www.kaggle.com/sammy123/lower-back-pain-symptoms-dataset>

<https://www.kaggle.com/meldadede/logistic-regression-for-lower-back-pain-symptoms>

[https://rstudio-pubs-static.s3.amazonaws.com/270798\\_46c7a5a8fb7d4980af1b942443271a6f.html](https://rstudio-pubs-static.s3.amazonaws.com/270798_46c7a5a8fb7d4980af1b942443271a6f.html)

<https://www.kaggle.com/nasirislamsujan/exploratory-data-analysis-lower-back-pain>

[https://en.wikipedia.org/wiki/Low\\_back\\_pain](https://en.wikipedia.org/wiki/Low_back_pain)

<https://www.geeksforgeeks.org/regression-classification-supervised-machine-learning/>

<https://www.coursera.org/lecture/machine-learning/classification-wlPeP>