

# Ekspertski sistem za preporuku algoritama mašinskog učenja

Stefan Kandić, Damjan Manojlović

## 1 Motivacija

Početak druge decenije 21. veka došlo je do naglog rasta popularnosti tehnika mašinskog učenja, koje su ranije bile prevashodno poznate samo u akademskim krugovima. Danas, sistemi zasnovani na mašinskom učenju se koriste u praktično svim granama industrije, od preporuke sadržaja na internetu i vožnje autonomnih automobila do određivanja zemlje pogodne za poljoprivredne delatnosti na osnovu satelitskih snimaka.

Intenzivna istraživanja su dovela do stvaranja velikog broja grana u samom polju, što je dodatno otežalo pristup ljudima koji ranije nisu bili upoznati sa mašinskim učenjem. Naš cilj je da napravimo alat pomoću kog će programeri koji nisu eksperti u ovom polju moći lako da se usmere ka rešenju problema koji rešavaju.

## 2 Pregled problema

Ideja je da korisnik sistema ima problem koji želi da reši, kao i podatke potrebne za treniranje algoritma. Uloga sistema je da mu da način na koji to može da izvede (algoritam ili familiju algoritama) ili da mu da koristan feedback ukoliko ne može da mu preporuči algoritam (npr. potrebno je pribaviti još podataka).

Što se tiče postojećih rešenja, nismo uspeli da ih nađemo. Postoje određene neformalne heuristike za rešavanje nekih tipskih problema, ali ništa što obuhvata više različitih grana mašinskog učenja.

## 3 Metodologija rada

### 3.1 Ulazi u sistem

Ulaz u sistem će biti opis problema koji se rešava. Neke od karakteristika problema bitnih za izbor algoritma su:

- šta korisnik tačno želi od algoritma (da predvidi kategoriju ili vrednost podataka, njihovu strukturu, generiše nove podatke ili možda da trenira agenta u specifičnom okruženju)
- struktura podataka nad kojim će se algoritam trenirati (slika, tekst, tabela itd) i informacija da li su podaci označeni
- broj uzoraka u skupu podataka
- specifične stvari koje zavise od prethodnih tačaka

### 3.2 Izlazi iz sistema

Izlaz iz sistema će predstavljati naziv algoritma koji bi se najbolje pokazao na datom problemu i eventualno dodatna literatura koja može pomoći korisniku da se bolje upozna sa njegovim radom.

### 3.3 Baza znanja

Bazu znanja će činiti pristupi koji su predstavljaju state-of-the-art za uobičajene probleme mašinskog učenja (klasifikacija slika - *CNN*<sup>1</sup>, tekst - *RNN*<sup>2</sup> itd). Njih ćemo identifikovati analizom naučnih radova i druge relevantne literature (časopisi, blogovi vezani za trendove u oblasti mašinskog učenja).

## 4 Pravila

### 4.1 Prvi nivo rezonovanja

U svrhu preporučivanja algoritma potrebno je odrediti opšte potrebe korisnika, odnosno opštu namenu algoritma, tip podataka sa kojim se rukuje, kao i količinu podataka. Ova faza predstavlja osnovu za kasniju detaljniju paramtrizaciju. U zavisnosti da li je korisnik izabrao da radi sa teksutalnim podacima ili slikama, kao i da li želi da radi klasifikaciju, da predvidi vrednost, ili nešto drugo, doći će do različitih zaključaka.

### 4.2 Drugi nivo rezonovanja

Drugi nivo rezonovanja je čvrsto povezan sa prvim i on služi za detaljniju parametrizaciju. Takođe, on se potenjano može satojati iz više podanivoa, sve u zavisnosti od prethodnih parametara.

### 4.3 Alarmi

Ukoliko izlaz iz sitema preporučuje neki zahtevniji algoritam, a broj podataka je veći od broja predviđenog za regularan rad tog algoritama, sistem će obavesitit korisnika o potencijalnim dužim čekanjima i problemima.

## 5 Primer rezonovanja

1. Korisnik unosi da želi da vrši predikciju kategorija
2. Dalje, korisnik unosi da su podaci sa kojima će se raditi u tekstualnom obliku i da poseduje 1 500 uzoraka
3. Na pitanje da li su podaci označeni korisnik odgovara da nisu
4. Na osnovu svih datih informacija, ekspertski sistem preporučuje algoritam K-Means za rešavanje datog problema

---

<sup>1</sup>**CNN** - Convolutional Neural Network

<sup>2</sup>**RNN** - Recurrent Neural Network