

# *Synechron*

AI/ML zadatak - Sistem za preporuke u osiguranju

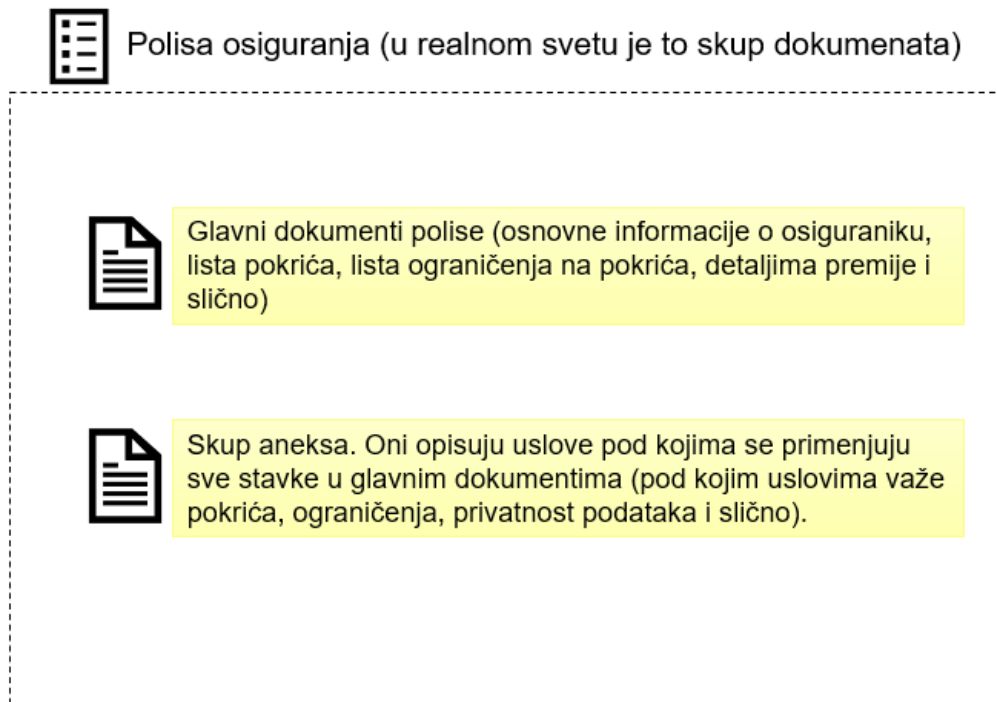
barKod v2 - Games of Code

## Sadržaj

1. Objašnjenje domena problema.....	2
2. Opis zadatka, skupa podataka i početne pretpostavke.....	3
3. Tehnička ograničenja .....	4
4. Metrike za merenje performansi modela.....	4

## 1. Objašnjenje domena problema

Da bi osiguravajuća kuća završila izdavanje polise osiguranja, postoji niz deklaracija i niz unapred definisanih obrazaca ili aneksa (engl. forms) koje je potrebno priložiti uz samu polisu. Ove forme/aneksi moraju pokrivati glavne odredbe politike polise (engl. policy provisions), rasporede, odobrenja, isključenja i druge aspekte politike same polise koju klijent uzima.



Slika 1 Sastav polise osiguranja

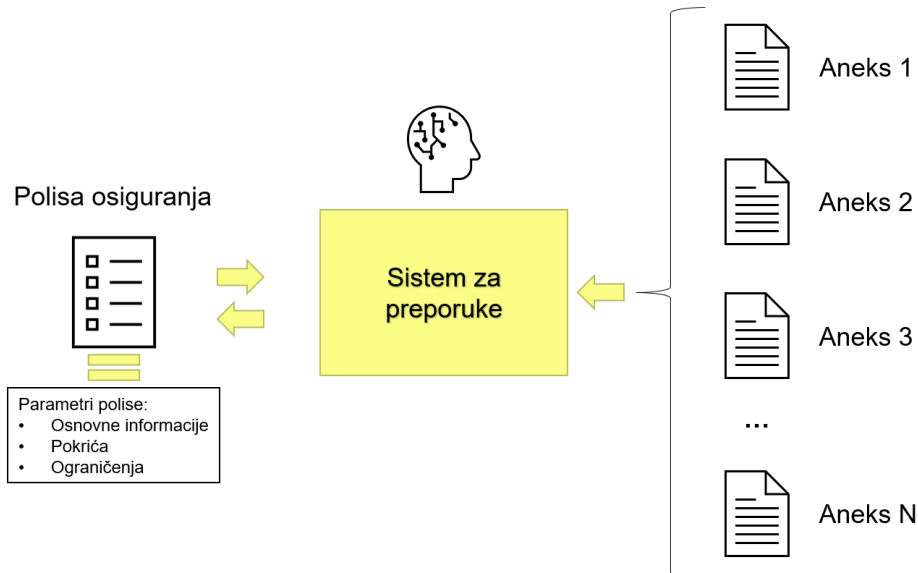
Povezivanje ovih aneksa sa polisom je zakonski zahtev koji se razlikuje od države do države i razlikuje se u zavisnosti od tipa polise, što dovodi do velike baze podataka potencijalnih aneksa koji će možda morati da se pridruže bilo kojoj polisi i još većeg broja kombinacija u zavisnosti od delatnosti i države na koje utiče politika.

Ako osiguravajuća kuća ne priloži bilo koji od ovih aneksa, ne samo da će uzrokovati finansijski gubitak osiguravajućem društvu, već će ih učiniti i pravno neusklađenim. U nekim slučajevima, to takođe može dovesti do akcije protiv osiguravajućeg društva od strane državnog ili saveznog regulatornog tela za osiguranje.

**Aneksi mogu biti vezani za bilo koju specifičnu informaciju iz same polise.** Jedini aneksi koji će uvek biti prikačeni uz polisu su aneksi vezani za privatnost podataka koje klijent daje prilikom uzimanja polise i njih ćete takođe videti u skupu podataka jer se pojavljuju u 100% slučajeva. Ostali aneksi zavise od toga šta je klijent sve uzeo od pokrića i pod kojim uslovima.

## 2. Opis zadatka, skupa podataka i početne pretpostavke

Implementirati sistem za preporuke aneksa za polise osiguranja na osnovu istorijskog skupa podataka koji sadrži listu realnih polisa osiguranja, zajedno sa aneksima koje su uz njih priključili ljudi u osiguravajućem društvu.



Slika 2 Osnovni zadatak sistema

U realnom svetu, osiguravajuće kuće često definišu skup pravila (engl. rules) pod kojim se aneksi priključuju uz polisu osiguranja (pravila tipa implikacije, recimo "država=TX  $\wedge$  šifra\_detalnosti=12345 => Aneks1") i ova pravila zavise od zakona u toj državi i direktno ih prate. Međutim problem koji mi rešavamo sadrži više od 100 mogućih aneksa pa je ručno pisanje ovakvih pravila nezgodno i komplikovano, a nije ni održivo jer se zakoni često menjaju.

Napraviti sistem za preporuke koji će biti sposoban da modeluje preporučivanje aneksa za zadatu polisu osiguranja na ulazu, korišćenjem tehnika mašinskog učenja, rovarjenja po podacima (engl. data mining) i/ili statističkih mera.

Skup podataka koji se koristi u ovom zadatku se odnosi samo na polise komercijalnog osiguranja (kada kompanija uzima osiguranje za svoju imovinu, usluge, ljude i slično) a ne na pojedinačna osiguranja (životno, putno, ...). Zbog toga ćete u skupu podataka videti da se polisa odnosi na kompaniju, a ne na pojedinačne osobe.

Opis namene kolona u skupu podataka:

- **Žute kolone** – Žute kolone sadrže osnovne podatke iz polise koje se nalaze u svakoj polisi. Tu se nalaze osnovne informacije o kompaniji koja uzima osiguranje, o njenoj šifri delatnosti, osnovnim geografskim informacijama i slično.
- **Plave kolone** – Plave kolone se odnose na informacije koje su specifične za tu pojedinačnu polisu osiguranja. U ovim kolonama se nalaze informacije o tome koja pokrića je klijent uzeo, pod kojim uslovima, da li postoje limiti (ograničenja) i slično. Ovi atributi utiču na neke anekse i definišu opšte informacije i uslove pokrića, ograničenja i slično koji se nalaze u tim aneksima.

- Sive kolone – Aneksi. Kada je aneks označen sa 1, to znači da je on bio prikazan uz tu pojedinačnu polisu osiguranja. Ako je označen sa 0, znači da je taj aneks bio izostavljen iz date polise. Ovo su naše target varijable (labele).

Ovaj sistem radi sa preko 100 mogućih aneksa koje može preporučiti (u praksi taj broj može ići i na preko 800, kad se uključi više vrsta osiguranja – individualna kao putno, životno, i slično, kao i neka druga komercijalna osiguranja). ***Ideja je da sistem čoveku preporuči najverovatnije anekse koje bi trebalo prikazati uz datu polisu, da čovek ne bi morao prolaziti kroz sve anekse i ručno odlučivati koji prikazati a koji ne. U prevodu, nije važno ako sistem preporuči više aneksa nego što bi trebalo, jer konačnu odluku ipak donosi čovek, ali ne bi smeo izostaviti nešto što bi trebalo prikazati.***

### 3. Tehnička ograničenja

Za implementaciju rešenja možete koristiti bilo koji programski jezik, bilo koji skup biblioteka i bilo koju platformu. Ocenjuje se kreativnost rešenja i AI/ML pristup problemu, a ne čistoća koda, izbor programskog jezika, platforme i slično. Jedino tehničko ograničenje jeste da rešenje mora da radi i da daje preporuke aneksa.

### 4. Metrike za merenje performansi modela

Pošto je u pitanju implementacija sistema za preporuke, konačnu odluku o tome da li će izvršiti prikazivanje aneksa uz polisu donosi čovek, tako da sistem predstavlja dodatak čoveku, odnosno proširenje ljudskog odlučivanja.

Ali pošto sistem ne sme izostaviti nijedan aneks koji bi trebao biti prikazan uz polisu (postoje legalna ograničenja i osiguravajuća kuća može dobiti tužbu ukoliko neki izostavi), važno je da sistem ne sme izostaviti nijednu relevantnu formu prilikom preporučivanja zaposlenom u osiguravajućoj kući.

**Zbog toga definišemo da recall mera sistema za preporuke mora biti jednaka 100% (sistem treba da preporuči sve što je relevantno).** Precision mera ne mora biti 100%, ali je cilj da bude što viša, jer će samim tim sistem davati manje nerelevantnih preporuka i zaposlenom koji koristi ovaj sistem će biti lakše da zna šta bi trebalo da bude prikazano. Što manje nerelevantnih stvari sistem preporuči, to će precision mera biti viša i to je takođe cilj (idealno 100%, ali ne mora).

Kriterijum ocenjivanja rešenja:

1. Kreativnost u samoj izradi rešenja (u modelima, pripremi podataka, objašnjivost modela, i slično).
2. Inovativnost i istraživački pristup u rešavanju problema.
3. Recall i precision kao mere performansi modela. Pošto skup podataka nije veliki i teško je obučiti model koji će raditi jako tačno, ovaj kriterijum možete posmatrati samo kao dodatni indikator da li ste izabrali dobar model na osnovu hipoteze koju ste postavili.
4. Objašnjivost modela (engl. model explainability)

