



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Predmet:

Veb majning

**Arhitektura i metodologije razvoja savremenih sistema za
preporuku: Pristup zasnovan na mašinskom učenju**

-seminarski rad-

Profesor: Marija D. Blagojević

**Student: Ognjen Tomić
Broj indeksa: 606/2023**

Čačak, 2024. godina



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Uvod

Definicija sistem za preporuku

Sistem za preporuku predstavlja sofisticirani informacijski sistem čiji je osnovni cilj personalizovano predlaganje entiteta, poput proizvoda, usluga ili sadržaja, korisnicima na osnovu njihovih individualnih preferencija, istorije ponašanja i demografskih karakteristika. Ovaj sistem se oslanja na kompleksne algoritme mašinskog učenja kako bi analizirao i interpretirao velike skupove podataka, identifikujući skrivene obrasce i veze između korisničkih akcija¹.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Značaj personalizovanih preporuka

Značaj personalizovanih preporuka u savremenom digitalnom okruženju je od suštinskog značaja, pružajući niz koristi kako korisnicima, tako i pružaocima usluga ili proizvoda. Personalizovane preporuke prilagođavaju se individualnim preferencijama i potrebama korisnika, što rezultira unapređenim korisničkim iskustvom. Korisnici doživljavaju platformu ili uslugu kao prilagođenu, što povećava njihovo zadovoljstvo i vernost. U digitalnom dobu, gde su opcije često preplavljajuće, personalizovane preporuke pomažu korisnicima da brže donose informisane odluke. Smanjuje se vreme potrebno za pretraživanje i filtriranje informacija, a korisnici su izloženi relevantnim sadržajima. Kroz pružanje preporuka koje su direktno relevantne za korisnika, povećava se nivo angažovanja. Korisnici su skloniji interakciji s platformom ili uslugom koja demonstrira razumevanje njihovih potreba.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Pregled pristupa razvoju

Pregled pristupa razvoju sistema za preporuku obuhvata različite metodologije i tehnike koje se koriste u konstrukciji ovih sistema. Razvoj sistema za preporuku je kompleksan proces koji zahteva pažljivu analizu podataka, implementaciju odgovarajućih algoritama mašinskog učenja i kontinuirano usklađivanje sa promenama u korisničkim preferencama.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Pregled pristupa razvoju

Filter kolaborativnog filtriranja- Ovaj pristup se zasniva na sličnosti između korisnika. Modeli predviđanja se konstruišu koristeći informacije o ponašanju sličnih korisnika kako bi se identifikovale preferencije ciljanog korisnika. Postoje dve osnovne vrste filtera kolaborativnog filtriranja

Filter zasnovan na sadržaju-Ovaj pristup analizira karakteristike proizvoda ili sadržaja kako bi se pravile preporuke. Sistem identifikuje sličnosti između proizvoda na osnovu njihovih karakteristika ili korisničkih preferencija, a zatim preporučuje proizvode koji su slični onima koje je korisnik već koristio ili ocenio pozitivno.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Sakupljanje i analiza podataka

Važnost relevantnih podataka

U kontekstu razvoja sistema za preporuku, važnost relevantnih podataka se ne može prenaglasiti, budući da su podaci osnovni građevinski blokovi koji omogućavaju modelima da pruže tačne i korisne preporuke korisnicima. Ključni aspekti koji naglašavaju važnost relevantnih podataka u ovom kontekstu uključuju: Kvalitet podataka direktno utiče na preciznost sistema za preporuku. Ako su podaci tačni, obimni i relevantni, modeli će biti u stanju bolje identifikovati obrasce i pružiti pouzdane preporuke korisnicima.⁴ Relevantni podaci omogućavaju sistemima za preporuku da stvore personalizovane preporuke koje odražavaju stvarne preferencije i interese korisnika. Što su podaci sveobuhvatniji, to je personalizacija efikasnija.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Tipovi podataka za treniranje modela

Pravilno odabrani i pažljivo pripremljeni podaci su ključni za efikasan rad sistema za preporuku. Raznolikost podataka pruža modelima dobar osnov za učenje i generalizaciju. Različite vrste podataka koriste se za treniranje modela, obezbeđujući širinu informacija i obuhvatnost. Neki od ključnih tipova podataka u treniranju modela sistema za preporuku obuhvataju: Informacije o aktivnostima korisnika, uključujući pretrage, pregledane proizvode, ocene, kupovine i druge interakcije. Ovi podaci o ponašanju korisnika pružaju uvid u njihove preferencije i interese. Demografski podaci uključuju informacije o korisnicima kao što su godine, pol, geografsko poreklo i druge socio-demografske karakteristike.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Etički aspekti sakupljanja podataka

Etički aspekti sakupljanja podataka postaju sve važniji kako tehnologija napreduje, posebno u oblasti sistema za preporuku. Ova pitanja obuhvataju pitanja privatnosti, transparentnosti i odnosa prema korisnicima. Pri prikupljanju podataka o korisnicima, posebna pažnja treba posvetiti očuvanju njihove privatnosti. Osetljive informacije poput ličnih podataka, adresa, i finansijskih informacija zahtevaju posebne mere opreza i transparentnosti u obradi. Korisnici bi trebalo jasno razumeti koje informacije se prikupljaju, kako se koriste i kako utiču na preporuke. Otvorenost i transparentnost u vezi sa procesima sakupljanja podataka gradi poverenje između korisnika i sistema za preporuku. Sakupljanje minimalno neophodnih podataka predstavlja etički pristup.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Podela podataka i evaluacija performansi

Razdvajanje na skup za obuku i skup za testiranje

Razdvajanje podataka na skup za obuku i skup za testiranje predstavlja ključnu praksu u razvoju sistema za preporuku zasnovanih na mašinskom učenju. Ovaj korak omogućava ocenu performansi modela, pružajući pouzdanu procenu njegove sposobnosti generalizacije na neviđene podatke. Ključni aspekti ovog procesa uključuju: Osnovni cilj razdvajanja podataka na skup za obuku i skup za testiranje je omogućiti modelu da nauči određene obrasce iz podataka za obuku i potom proveri svoju efikasnost na podacima koje nije video tokom treniranja. Ovo pomaže u proceni generalizacije modela. Podaci se obično dele na skupove za obuku i testiranje u određenom odnosu, na primer, 80% podataka za obuku i 20% za testiranje. Ovaj odnos može varirati u zavisnosti od veličine skupa podataka i specifičnosti problema.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Metrike za evaluaciju sistema

Evaluacija sistema za preporuku uključuje korišćenje različitih metrika kako bi se procenila tačnost, relevantnost i performanse modela. Preciznost se meri kao odnos tačno preporučenih entiteta (npr. proizvoda) u odnosu na ukupan broj preporučenih entiteta. Formula za preciznost je:

Preciznost = broj tačnih preporuka / ukupan broj preporučenih entiteta



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Metrike za evaluaciju sistema

Recall meri sposobnost sistema da identifikuje sve relevantne entitete među svim stvarnim relevantnim entitetima. Formula za recall je:

recall = broj tačnih preporuka / ukupan broj stvarno relevantnih entiteta



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Metrike za evaluaciju sistema

F1-score je harmonična sredina između preciznosti i odaziva i pruža izbalansiran pristup evaluaciji modela. Formula za F1-score je:

$$F1 = 2 \times \text{preciznost} \times \text{recall} / (\text{preciznost} + \text{recall})$$



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Praćenje i poboljšanje performansi

Praćenje i poboljšanje performansi sistema za preporuku ključni su koraci u održavanju efikasnog i korisnički orijentisanog sistema tokom vremena. Aktivno prikupljanje povratnih informacija od korisnika predstavlja dragocen izvor informacija o zadovoljstvu korisnika i njihovim potrebama. Korišćenje anketa, ocena, ili komentara može pomoći u identifikaciji oblasti za unapređenje. Praćenje metrika performansi, kao što su preciznost, odaziv, i F1-score, pomaže u oceni tačnosti sistema. Redovna analiza ovih metrika omogućava identifikaciju slabih tačaka i usmeravanje resursa na unapređenje tih aspekata.



FAKULTET TEHNIČKIH NAUKA U ČAČKU
UNIVERZITET U KRAGUJEVCU

Algoritmi mašinskog učenja u sistemima za preporuku

Filter kolaborativnog filtriranja

Filter kolaborativnog filtriranja je tehnika koja se često koristi u sistemima za preporuku kako bi se predvideli interesovanja korisnika na osnovu sličnosti između njih i drugih korisnika. Ovaj pristup se oslanja na ideju da će korisnici koji imaju slične ukuse ili ponašanje imati slične preferencije u vezi sa proizvodima ili sadržajem. U filteru kolaborativnog filtriranja, ne analiziraju se eksplicitne karakteristike proizvoda ili korisnika, već se fokus stavlja na obrasce ponašanja zajednice korisnika. Postoje dve osnovne vrste filtera kolaborativnog filtriranja: user-based (bazirano na korisnicima) i item-based (bazirano na proizvodima).



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Filter kolaborativnog filtriranja

User-Based Collaborative Filtering (UCF)

User-Based Collaborative Filtering (UCF): U filteru baziranom na korisnicima, preporuke se prave na osnovu sličnosti između ciljnog korisnika i drugih korisnika. Ako dva korisnika imaju slične preferencije u prošlosti, pretpostavlja se da će imati slične preferencije i u budućnosti. Na primer, ako korisnik A i korisnik B često biraju iste proizvode ili ocenjuju proizvode slično, sistem će preporučiti proizvode koje korisnik B voli, a koje korisnik A još nije vidio.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Filter kolaborativnog filtriranja

Item-Based Collaborative Filtering (ICF)

U filteru baziranom na proizvodima, fokus je na sličnosti između proizvoda. Ako korisnik pokazuje interesovanje za određene proizvode, sistem će preporučiti druge proizvode koji su slični onima koje je korisnik već ocenio ili kupio. Ova metoda se često koristi kada je broj korisnika veliki, a broj proizvoda manji. Jedna od ključnih prednosti kolaborativnog filtriranja je što ne zahteva eksplicitno znanje o proizvodima ili korisnicima. Osim toga, može efikasno raditi u situacijama gde postoje skriveni ili nepoznati faktori koji utiču na korisničke preference.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Filter zasnovan na sadržaju

Filter zasnovan na sadržaju je još jedan ključni pristup u razvoju sistema za preporuku. Ovaj metod se fokusira na analizu karakteristika i osobina samih proizvoda ili sadržaja, a ne na sličnosti između korisnika. Ideja je prilagoditi preporuke korisnicima na osnovu njihovih prethodnih preferencija, uzimajući u obzir relevantne karakteristike proizvoda. Filter zasnovan na sadržaju analizira karakteristike proizvoda kako bi stvorio profil svakog proizvoda. Ove karakteristike mogu uključivati ključne reči, žanrove, autore, ili druge relevantne attribute. Sadržaj se često vektorizuje kako bi se predstavio numerički.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Hibridni pristupi

Hibridni pristupi u sistemima za preporuku kombinuju više različitih tehnika i metoda kako bi unapredili tačnost, pouzdanost i personalizaciju preporuka. Ova kombinacija omogućava prevazilaženje nedostataka pojedinačnih metoda i pruža sveobuhvatniji pristup zadatku preporuke. Ovi pristupi koriste različite tehnike preporuke i kombinuju ih u jedinstvenu funkcionalnost.



FAKULTET TEHNIČKIH NAUKA U ČAČKU UNIVERZITET U KRAGUJEVCU

Zaključak

Sistemski razvoj preporučenih sistema u poslednjim decenijama snažno je podstaknut napretkom u oblasti mašinskog učenja. Algoritmi za preporuke postali su sve sofisticiraniji, omogućavajući sistemima da pruže personalizovane preporuke sa visokom tačnošću. Personalizovane preporuke su postale ključne u poboljšanju korisničkog iskustva na različitim platformama. Sistemi za preporuke pružaju korisnicima sadržaj i proizvode koji su relevantni za njihove individualne preferencije, čime se povećava angažovanje i zadovoljstvo korisnika. Sakupljanje i obrada podataka igra ključnu ulogu u efikasnom funkcionisanju sistema za preporuke. Izazovi kao što su problem hladnog početka, pristrasnost podataka, i etički aspekti sakupljanja podataka zahtevaju pažljivu analizu i implementaciju odgovarajućih rešenja.