



UNIVERZITET U NIŠU
ELEKTRONSKI FAKULTET
Katedra za računarstvo



Sistem za preporuku android aplikacije

-SEMINARSKI RAD-

Mentor: Doc. dr Miloš Bogdanović

Student: Predrag Nikolić 653

Niš, 2021.

Sadržaj

1	Uvod	1
2	Sistemi za preporuke	2
2.1	Filtriranje bazirano na sadržaju	2
2.2	Filtriranje bazirano na interakciji drugih korisnika	3
2.3	Hibridni pristup	3
2.4	Metoda za određivanje preporuke	4
3	Implementacija	5
3.1	Korišćeni alati	5
3.2	Analiza i obrada podataka	6
3.3	Implementacija sistema za preporuku	13
4	Rezultat	15
5	Zaključak	19
	Literatura	20

1 Uvod

Milioni korisnika svakodnevno koriste *Google Play* prodavnicu da bi instalirali android aplikacije na svojim uređajima. Zbog toga što korisnici obično ne znaju šta konkretno hoće da instaliraju, oni provode vreme u traženju novih aplikacije, koje će možda da im se svide. Kako na *Google Play* servisu postoje milioni aplikacija, pretraživanje aplikacija da bi se našla odgovarajuća može da bude iscrpljujuća za korisnike. Da bi se korisnicima ovaj postupak olakšao *Google Play* prodavnica može da da preporuku, baziranu na tome šta je korisnik već instalirao, kako bi se korisniku pružio manji podskup aplikacija koje treba da pregleda i kako bi mu se dala ideja da instalira aplikaciju, za koju ne zna.

Ovim problemom se bave sistemi za preporuku. Sistemi za preporuku predstavljaju skup tehnika i metoda za predikciju preference koju bi korisnik imao prema nekoj stavci. Preferenca korisnika se određuje raznim metodama koje su bazirane na podacima o tim stavkama i podacima o tome koje je stavke korisnik preferirao u prošlosti ili preferira trenutno.

Ovaj projekat se bavi sistemom preporuke za android aplikacije sa *Google Play* prodavnice. Sistem preporuke je baziran na podacima o aplikacijama sa *Google Play* prodavnice, koji su dati u okviru skupa podataka. U ovom radu je predstavljen postupak analize i pripreme podataka. Sam sistem preporuke je implementiran kao aplikacija koja na osnovu skupa podataka i metoda određivanja sličnosti vraća preporuke android aplikacija.

U sledećem poglavlju je dat teoretski osvrt na sisteme za preporuku. U trećem poglavlju su dati detalji implementacije kao i postupak analize i obrade skupa podataka, a četvrto poglavlje se bavi analizom rezultata.

2 Sistemi za preporuke

Sistemi za preporuku su podskup sistema za filtriranje informacija koji pokušavaju da predvide rejting ili preferencu koju korisnik ima prema nekoj aplikaciji [1]. Sistemi za preporuku se koriste u raznim oblastima, a oni generišu preporuke za razne servise kao što su muzički i video servisi, veb prodavnice, socijalne mreže itd. Zadatak sistema za preporuku je da da predlog korisniku, kako bi olakšao njegovo korišćenje nekog servisa. Na primer, na osnovu toga šta neki korisnik koristi od mobilnih aplikacija, sistem za preporuku može da preporuči korisniku niz aplikacija koje do sada nije koristio, a koje mogu da mu budu od koristi.

Sistemi za preporuku mogu da koriste razne tehnike i metode da bi pronašli preporuku. Tehnike i metode za pronalaženje preporuka su bazirane na podacima o korisnicima nekog sistema. Sistemi za preporuku koriste takve podatke da bi pronašli neku korelaciju između onoga što korisnik traži i onoga što se nalazi u tom sistemu.

Postoje dva glavna pristupa koji se koriste prilikom implementiranja sistema za preporuku, koji mogu da se koriste nezavisno ili zajedno da bi se odredila preporuka. Jedan je filtriranje bazirano na sadržaju (eng. *content-based filtering*), drugi je filtriranje bazirano na interakciji drugih korisnika (eng. *collaborative filtering*).

2.1 Filtriranje bazirano na sadržaju

Filtriranje bazirano na sadržaju (eng. *content-based filtering*) je jedan od osnovnih pristupa za implementaciju sistema za preporuku. Ove metode i tehnike su bazirane na informacijama vezanim za stavku koju korisnik traži i na informacijama o tome koje preference korisnik ima. Ovaj pristup je koristan kada imamo podatke o stavkama, a ne o korisniku. Obzirom na prirodu podataka koje imamo dostupne za android aplikacije, u ovom projektu je korišćen ovaj pristup za implementaciju sistema za preporuku android aplikacija.

Sistemi bazirani na ovom pristupu tretiraju preporuku kao klasifikaciju specifičnu za korisnika, koja klasifikuje šta je korisniku potrebno ili ne, bazirano na podacima o stavkama. Ovi sistemi pokušavaju da preporuče stavke koje su slične kao stavke koje je korisnik preferirao u prošlosti ili trenutno. Da bi se odredilo šta korisnik preferira, koriste se informacije o tome koje je stavke korisnik preferirao u prošlosti, ali se koriste i podaci o interakciji korisnika sa sistemom za preporuku. Na osnovu ovih stavki se kasnije kreira preporuka.

Da bi podaci o stavkama mogli da se koriste u procesu klasifikacije potrebno je da se oni pretvore u oblik koji odgovara takvim tehnikama i metodama. Podaci koji opisuju stavku se, na osnovu analize podataka, pretvaraju u vektore koji određenim vrednostima opisuju stavku. Tako vektorizovani podaci se kasnije koriste za određivanje preporuke. Tehnike mašinskog učenja koje mogu da se koriste za generisanje preporuke su Bajesovi klasifikatori, analize klastera, stabla odluke i neuronske mreže.

Glavni nedostatak ovog pristupa se ogleda u tome što je sistem za preporuku limitiran da daje preporuke o stavkama koje su istog tipa kao stavke koje korisnik već koristi. Ovi sistemi su limitirani na izvor sadržaja, koji korisnik pretražuje i ne mogu da daju preporuku za stavke koje potiču iz drugačijeg izvora. Da bi se ovo prevazišlo u realnim primenama se koriste hibridni sistemi, koji kombinuju ovaj pristup sa drugim tehnikama i metodama.

2.2 Filtriranje bazirano na interakciji drugih korisnika

Filtriranje bazirano na sadržaju interakcije korisnika (eng. *collaborative filtering*) je pristup za kreiranje sistema za preporuku, koji je baziran na navikama drugih korisnika. Ovaj pristup je baziran na tome da ako dva korisnika preferiraju sličan skup stavki, onda preporuka može da bude neka stavka sličnog korisnika koju korisnik koji trenutno pretražuje nije preferirao do sada. Sistem kreira preporuke na osnovu preferencija korisnika prema stavkama, tako što lokalizuje i povezuje korisnike, koji imaju slične preference. Ove metode se dele na metode bazirane na memoriji i metode bazirane na modelu. Glavna prednost ovakvog pristupa se ogleda u tome što sistem može da kreira preporuku bez eksplicitnih podataka o stavkama koje preporučuje.

Da bi se pronašli slični korisnici, potrebno je da se kreiraju profili korisnika koji se kasnije vektorizuju, da bi se nad vektorima kasnije primenjivale razne metode za određivanje sličnosti. Za određivanje sličnosti dva korisnika mogu da se koriste različite mere i koeficijenti, kao što je Pirsonova korelacija, a mogu da se koriste i metode mašinskog učenja kao što je *k-nearest neighbor*.

Jedan od nedostataka ovog pristupa je pronalaženje sličnog korisnika za novog korisnika, jer novi korisnik nema informacije o korišćenim aplikacijama. Takođe problem može da se javi, ako sistem sadrži veliki broj stavki, jer korisnici obično koriste mali podskup stavki, koji može da bude dosta različit za korisnike. Veliki broj stavki može da utiče i na skalabilnost sistema, jer je u tom slučaju potrebno više vremena i resursa za izvršavanje metoda za pronalaženje sličnih korisnika.

2.3 Hibridni pristup

Obzirom da prethodno pomenuti pristupi imaju svoje nedostatke, prilikom implementacije sistema za preporuku se obično koristi kombinovani pristup, koji uključuje tehnike i metode oba pristupa. Pored mogu da se koriste i druge tehnike i metode, da bi se upotpunili nedostaci oba pristupa. Hibridni pristup može da se realizuje na nekoliko načina:

- Kombinovanjem predikcija koje su dobijene primenom oba pristupa zasebno.
- Dodavanjem mogućnosti jednog pristupa u drugi.
- Kombinovanjem oba pristupa u jedan zajednički model.

2.4 Metoda za određivanje preporuke

Kao što smo videli u prethodnim odeljcima, pristup koji se koristi za kreiranje sistema za preporuku zavisi od toga koji su podaci dostupni, kao i od toga kako i na koji način želimo da tretiramo podatke. Ono što je glavno za dva osnovna pristupa je to da traže neku sličnost u podacima. Sličnost u podacima može da se nađe preko raznih mera sličnosti, ali može da se odredi preko raznih relacija među podacima, koje mogu da se dobiju korišćenjem tehnika i metoda mašinskog učenja. U ovom projektu su za pronalaženje preporuka iskorišćena metoda *k-nearest neighbor* i kosinusna sličnost, pa će oni biti detaljnije objašnjeni u nastavku.

Metoda *k* najbližijih "komšija" (eng. *k-nearest neighbor*) je metoda mašinskog učenja, koja pretpostavlja da se slične stvari mogu naći "blizu" jedna drugoj. Koliko su dve stvari "blizu" može da se odredi na osnovu neke mere distance ili sličnosti, kao na primer euklidska distanca ili kosinusna sličnost. Ova metoda pronalazi *k* sličnih podataka u odnosu na podatak za koji tražimo slične podatke. Podaci se pretražuju u skupu svih podataka. U ovom projektu je za kriterijum pronalaženja *k* najbližijih podataka iskorišćena euklidska distanca. Metoda *k* najbližijih "komšija" pronalazi *k* podataka koji imaju najmanju Euklidsku distancu u odnosu na zadati primer podatka. Euklidska distanca je mera distance između dve tačke odnosno dva vektora u prostoru koja se izračunava kao norma razlike ta dva vektora:

$$d(\vec{p}, \vec{q}) = |\vec{p} - \vec{q}|, \quad (2.1)$$

Nakon pronalaženja *k* najbližijih podataka korišćenjem metode *k-nearest neighbor*, ti podaci se sortiraju u odnosu na meru kosinusne sličnosti sa podatkom za koji tražimo slične podatke. Oni se sortiraju u opadajućem redosledu u odnosu na vrednost kosinusne sličnosti (maksimalna vrednost je 1). Kosinusna sličnost je mera sličnosti između dva vektora koja se izračunava kao količnik skalarnog proizvoda i proizvoda dužina vektora:

$$\cos(\theta) = \frac{\vec{p} \cdot \vec{q}}{||\vec{p}|| * ||\vec{q}||}, \quad (2.2)$$

3 Implementacija

Za implementaciju je korišćen programski jezik *Python* i biblioteka *ScikitLearn*, kao i neke druge standardne biblioteke koje se koriste u oblasti mašinskog učenja. Skup podataka na osnovu koga je napravljen ovaj sistem za preporuke sadrži informacije o aplikacijama sa *Google Play* prodavnice. Ovo je javno dostupni skup podataka koji se može naći na sajtu *Kaggle* (<https://www.kaggle.com/lava18/google-play-store-apps>).

Implementacija je bazirana na pristupu koji je obrađen u radu [8]. Implementacija je podeljena u dva dela. Prvi deo obuhvata pripremu i analizu podataka koja je prikazana u okviru *Jupyter notebook*-a. Kao rezultat ovog dela koda dobijaju se obrađeni podaci koji se kasnije koriste za obučavanje modela. Drugi deo implementacije predstavlja aplikaciju koja za datu aplikaciju vraća preporuke.

3.1 Korišćeni alati

Aplikacija je implementirana u programskom jeziku *Python*, verziji 3. Biblioteke i alati koji su korišćeni u kodu su:

- *NumPy* [2]
- *Pandas* [3]
- *Jupyter notebook* [4]
- *Matplotlib* [5]
- *Seaborn* [6]
- *SciKit learn* [7]

Numpy je *Python* biblioteka koja pruža implementaciju multidimenzionalnog vektora kao objekta. Sam *Python* ne pruža mogućnost za manipulaciju sa listama, pa se *numpy* uglavnom uvek koristi u okviru ovakvih aplikacija. Osnovni objekat je *ndarray*, koji pruža napredne funkcije nad nizom podataka. Takođe ova biblioteka pruža mogućnost izvršavanja operacija nad dva ili više *ndarray* objekata. Ova biblioteka je napisana u programskom jeziku *C*, tako da pruža visoke performanse prilikom obavljanja operacija.

Biblioteka *pandas* pruža objekte i strukture dizajnirane za rad sa označenim i relacionim podacima. Kao takva se uglavnom koristi u projektima koji uključuju analizu, obradu i pripremu podataka. Ova biblioteka je jedna od najkorisnijih u oblasti mašinskog učenja. Dva objekta koja implementira ova biblioteka su *Series* i *DataFrame*. *DataFrame* je objekat koji pruža funkcionalnosti manipulacije i obrade tabelarnih podataka i učitavanja podataka iz raznih formata fajlova i baza podataka. Ova biblioteka je napisana u *Python*-u i koristi objekat *ndarray*, tako da je saglasna sa bibliotekom *Numpy*. Zbog toga što je

implementirana u *Python*-u, nema visoke performanse, pa treba obratiti pažnju u radu sa njenim objektima.

Jupyter je *open-source* projekat koji pruža alate za interaktivan rad u oblastima obrade podataka i mašinskog učenja. Alat koji je korišćen u okviru ovog projekta je *jupyter notebook*. Ovaj alat omogućava pisanje dokumenata koji može da sadrži izvršivi kod, tekst, jednačine i vizualizaciju podataka, pa je kao takav zgodan za analizu podatka i prikazivanje toka obrade i pripreme podataka, treniranja modela itd. U ovom projektu je preko ovog alata prikazan proces analize i pripreme podataka.

Biblioteke *matplotlib* i *seaborn* su biblioteke koje služe za vizualizaciju podataka. Pružaju mogućnost crtanja raznih grafikona, tabela, statistika itd. Ove biblioteke se koriste u okviru alata *jupyter notebook*, da bi omogućile vizualizaciju podataka.

SciKit learn je *open-source* projekat implementiran u programskom jeziku *Python* i bibliotekama *numpy*, *scipy* i *matplotlib*. Ovaj projekat pruža efikasne alate za oblast mašinskog učenja. Ova biblioteka pruža funkcionalnosti za pripremu podataka, kao i veliki broj modela koji implementiraju neke od tehnika mašinskog učenja kao što su klasifikacija, regresija i klasterovanje. Ova biblioteka sadrži veliki broj modela koji mogu lako da se podese, treniraju i testiraju nad podacima. Takođe, u ovoj biblioteci su sadržane i metode za redukciju dimenzionalnosti skupa podataka kao i metode za odabir parametara prilikom treniranja modela.

3.2 Analiza i obrada podataka

Skup podataka predstavlja podatke o aplikacijama koji su sačuvani u okviru jednog *csv* fajla. Svaka aplikacija opisana je sa 13 atributa. Skup podataka sadrži informacije za 10841 aplikacija sa *Google Play* prodavnice.

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159 19M	10,000+	Free	0	Everyone	Art & Design		7-Jan-18 1.0.0		4.0.3 and up
Coloring book moana	ART_AND_DESIGN	3.9	967 14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play		15-Jan-18 2.0.0		4.0.3 and up
U Launcher Lite â€” FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510 8.7M	5,000,000+	Free	0	Everyone	Art & Design		1-Aug-18 1.2.4		4.0.3 and up
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644 25M	50,000,000+	Free	0	Teen	Art & Design		8-Jun-18 Varies with device		4.2 and up
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967 2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity		20-Jun-18		1.1 4.4 and up
Paper flowers instructions	ART_AND_DESIGN	4.4	167 5.6M	50,000+	Free	0	Everyone	Art & Design		26-Mar-17		1 2.3 and up
Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178 19M	50,000+	Free	0	Everyone	Art & Design		26-Apr-18		1.1 4.0.3 and up
Infinite Painter	ART_AND_DESIGN	4.1	36815 29M	1,000,000+	Free	0	Everyone	Art & Design		14-Jun-18 6.1.61.1		4.2 and up
Garden Coloring Book	ART_AND_DESIGN	4.4	13791 33M	1,000,000+	Free	0	Everyone	Art & Design		20-Sep-17 2.9.2		3.0 and up
Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121 3.1M	10,000+	Free	0	Everyone	Art & Design;Creativity		3-Jul-18		2.8 4.0.3 and up
Text on Photo - Fonttee	ART_AND_DESIGN	4.4	13880 28M	1,000,000+	Free	0	Everyone	Art & Design		27-Oct-17 1.0.4		4.1 and up
Name Art Photo Editor - Focus n Filters	ART_AND_DESIGN	4.4	8788 12M	1,000,000+	Free	0	Everyone	Art & Design		31-Jul-18 1.0.15		4.0 and up
Tattoo Name On My Photo Editor	ART_AND_DESIGN	4.2	44829 20M	10,000,000+	Free	0	Teen	Art & Design		2-Apr-18		3.8 4.1 and up
Mandala Coloring Book	ART_AND_DESIGN	4.6	4326 21M	100,000+	Free	0	Everyone	Art & Design		26-Jun-18 1.0.4		4.4 and up
3D Color Pixel by Number - Sandbox Art Coloring	ART_AND_DESIGN	4.4	1518 37M	100,000+	Free	0	Everyone	Art & Design		3-Aug-18 1.2.3		2.3 and up
Learn To Draw Kawaii Characters	ART_AND_DESIGN	3.2	55 2.7M	5,000+	Free	0	Everyone	Art & Design		6-Jun-18 NaN		4.2 and up
Photo Designer - Write your name with shapes	ART_AND_DESIGN	4.7	3632 5.5M	500,000+	Free	0	Everyone	Art & Design		31-Jul-18		3.1 4.1 and up
350 Diy Room Decor Ideas	ART_AND_DESIGN	4.5	27 17M	10,000+	Free	0	Everyone	Art & Design		7-Nov-17		1 2.3 and up
FlipaClip - Cartoon animation	ART_AND_DESIGN	4.3	194216 39M	5,000,000+	Free	0	Everyone	Art & Design		3-Aug-18 2.2.5		4.0.3 and up

Slika 3.1: Izgled skupa podataka.

Za svaku aplikaciju su sačuvani sledeći podaci:

- *App* - ime aplikacije
- *Category* - kategorija aplikacije

- *Rating* - rejting aplikacije koji je izračunat na osnovu ocena korisnika
- *Reviews* - broj korisnika koji su ocenili aplikaciju
- *Size* - veličina aplikacije na disku
- *Installs* - broj instalacija aplikacije
- *Type* - tip aplikacije (besplatna/kupuje se)
- *Price* - cena aplikacije
- *Content Rating* - preporuka za uzrast
- *Genres* - žanr aplikacije
- *Last updated* - poslednji *update* aplikacije
- *Current Ver* - trenutna verzija aplikacije
- *Android Ver* - potrebna verzija androida da bi aplikacija radila

Da bi ovi podaci mogli da se koriste za treniranje modela na osnovu koga se generišu preporuke potrebno je pripremiti podatke u odgovarajućem obliku. Model i metode koje se koriste za generisanje preporuke su bazirani na merama sličnosti kao što su Euklidska i kosinusna sličnost. Obzirom da ove mere mogu da se primenjuju na vektorima koji sadrže brojeve vrednosti, potrebno je da se podaci, iz skupa podataka o android aplikacijama, konvertuju u takve vektore. Skup podataka sadrži brojeve i kategoričke vrednosti. Neki od podataka su prikazani kombinacijom brojeva i znakova, koji prikazuju npr. mernu jedinicu. Ovakve podatke je potrebno konvertovati da budu prikazani samo brojevima, npr. u slučaju merne jedinice treba konvertovati sve redove u istu mernu jedinicu. Kategorički podaci sa druge strane mogu da se vektorizuju i budu prikazani preko vektora koji sadrže 0 i 1. Priprema podataka za ovaj sistem je urađena na osnovu analize podataka. Analiza podataka kao i njihova obrada prikazana je u nastavku.

Skup podataka o android aplikacijama sadrži polja koja kao vrednost imaju upisanu vrednost *NaN*. Ova vrednost označava da podatak za određeni atribut nedostaje u opisu aplikacije. Pošto modelu sa kojim radimo ne možemo da prosledimo ovakvu vrednost potrebno je nekako obraditi ove vrednosti. Jedan od načina je da obrišemo redove koji sadrže ove vrednosti ili da ih zamenimo odgovarajućim vrednostima (standardna vrednost, prosečna vrednost, središnja vrednost itd.).

Kao što možemo da vidimo na slici 3.2, atribut *Rating* ima veći broj vrednosti koje nedostaju tako da je potrebno zameniti ih nekom vrednošću koja odgovara vrednostima za rejting drugih aplikacija. Što se tiče drugih atributa sa vrednostima *NaN*, oni ne predstavljaju veliki procenat u odnosu na sve podatke, tako da njih uklanjamo nakon obrade svih kolona, ako i dalje budu sadržani u skupu podataka.

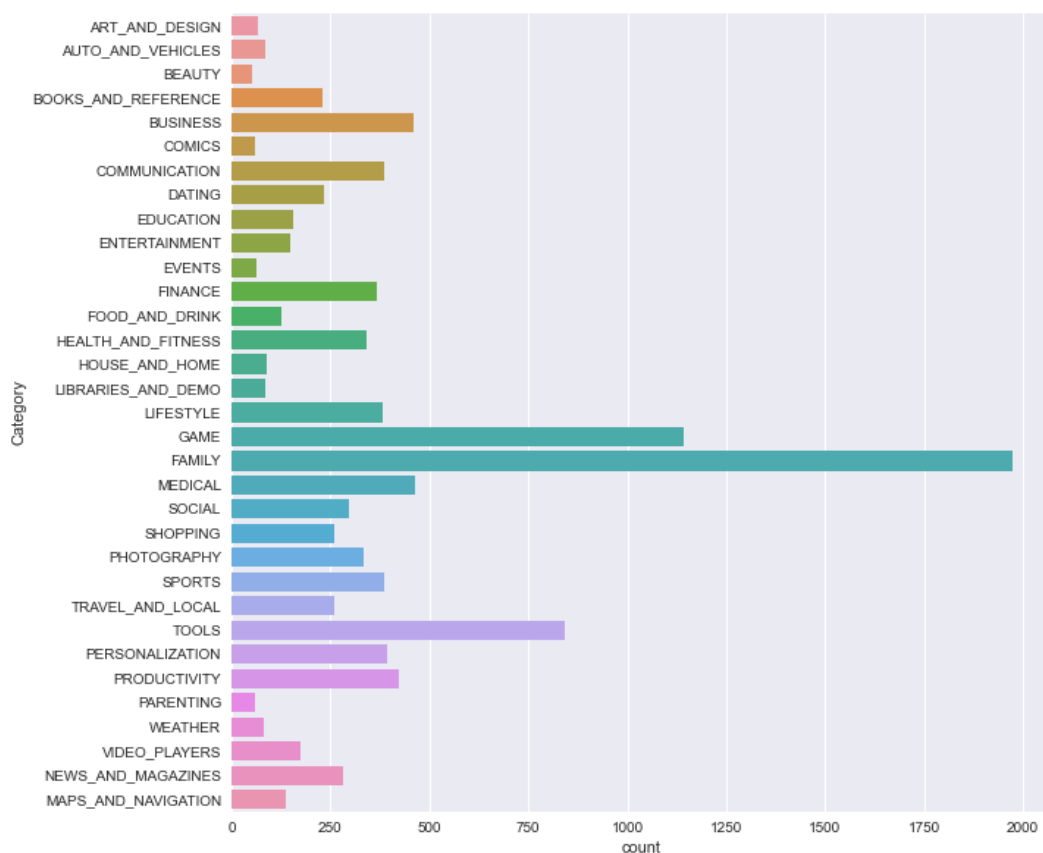
Atribut *App* je tipa string i predstavlja ime aplikacije, koje ne daje nikakvu informaciju o sličnosti sa drugim aplikacijama tako da ga za potrebe treniranja modela ne koristimo. Ova kolona iz skupa podataka se na kraju čuva u zasebni fajl kako bi se koristila za

App	0
Category	0
Rating	1474
Reviews	0
Size	0
Installs	0
Type	1
Price	0
Content Rating	1
Genres	0
Last Updated	0
Current Ver	8
Android Ver	3

Slika 3.2: Atributi skupa podataka i broj *NaN* vrednosti za svaki atribut.

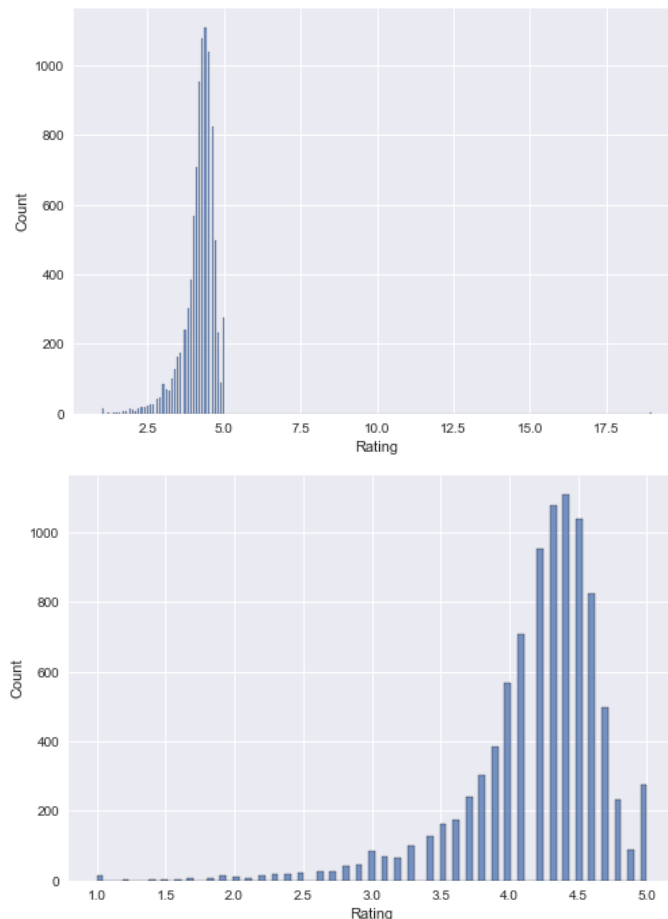
pretragu aplikacija i davanje preporuka po imenu. Svaka aplikacije se vezuje za odgovarajući vektor iz obrađenog skupa podataka. Vektor iz obrađenog skupa podataka se koristi za pronalaženje sličnih aplikacija.

Category predstavlja kategoriju aplikacije, koja je zapisana u obliku stringa. Ovaj atribut je kategorički, pa kao takav odgovara modelu koji se koristi, pa nema potrebe za promenom. Na slici 3.3 su prikazane kategorije kao i broj aplikacija po svakoj kategoriji.



Slika 3.3: Kategorije aplikacija.

Rating predstavlja prosečnu ocenu koju su korisnici davali aplikaciji. To je brojeva vrednost između 0 i 5, pa kao takva odgovara. Potrebno je utvrditi da su sve vrednosti razlomljeni brojevi između 0 i 5. Analizom ovog atributa utvrđeno je da postoje ocene koje nisu u traženom opsegu (slika 3.4). S obzirom da je u pitanju samo jedna aplikacija koja ima ocenu van opsega, tu aplikaciju brišemo iz skupa podataka.

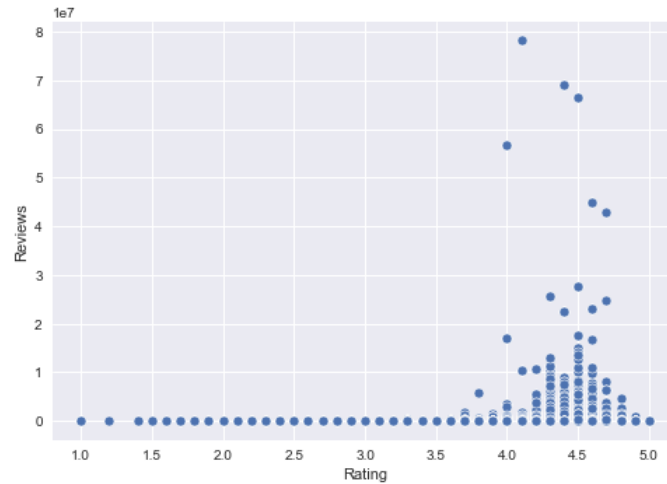


Slika 3.4: Opseg atributa rejting pre i posle uklanjanja ne validnih vrednosti.

Takođe u skupu podataka postoji 1474 aplikacija za koje nedostaje podatak o rejtingu. Pošto te aplikacije čine 10% skupa podataka, one se ne brišu, već se svim tim aplikacijama dodeljuje vrednost 2.5. Ova vrednost je određena u odnosu na atribut *Reviews*. Analizom korelacije između ova dva atributa (slika 3.5), utvrđeno je da porastom broja ocena raste i sam rejting. Zbog toga što maksimalna i srednja vrednost atributa *Reviews*, za aplikacije kojima nedostaje ocena, spadaju u opseg vrednosti atributa *Reviews* za aplikacije koje imaju ocene između 2 i 3, aplikacijama koje nemaju vrednost rejtinga daje se rejting 2.5.

Reviews kao što je već rečeno, predstavlja broj korisnika koji su ocenili neku aplikaciju. Ovaj atribut je predstavljen kao celobrojna vrednost kod svih aplikacija i kao takav odgovara za proces treniranja modela.

Size predstavlja veličinu aplikacije na disku. Ovaj atribut je u skupu podataka upisan kao string vrednost. String je sačinjen od broja koji predstavlja veličinu i slova koje predstavlja mernu jedinicu (K - kilobajt, M - megabajt). Takođe neke aplikacije imaju specijalnu vrednost *"Varies with device"*. Ovaj atribut se za potrebe treniranja modela



Slika 3.5: Korelacija između rejtinga i broja datih ocena za aplikaciju.

konvertuje u brojevnu vrednost, tako što se brišu slova koja označavaju mernu jedinicu, kilobajti se konvertuju u megabajte, a specijalna vrednost se menja sa -1.

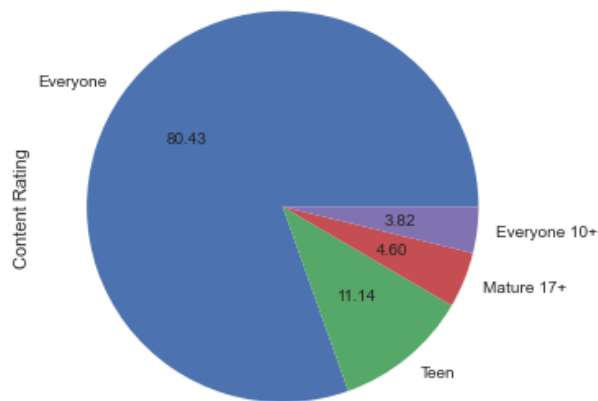
Broj instalacija (*Installs*) je predstavljen brojem, koji predstavlja donju granicu broja instalacija, i znakom +, koji označava da je broj instalacija veći od donje granice. Za potrebe treniranja modela uklanjamo + i dobijamo brojevnu vrednost. Tip aplikacije (*Type*) je označen sa dva stringa "Paid" i "Free", pa možemo da ga tretiramo kao nominalnu vrednost. Atribut *Price* koji predstavlja cenu aplikacije je brojevna vrednost i nema nevalidnih podataka ili aplikacija za koje ova vrednost nedostaje, tako da je kao takva odgovarajuća za treniranje modela.

Atribut *Content Rating* je tipa string, koji ima nekoliko mogućih vrednosti (slika 3.5). Ovaj atribut možemo da tretiramo kao nominalnu vrednost. S obzirom da je broj aplikacija koje za ovaj atribut imaju vrednosti "Adults only 18+" i "Unrated", možemo da kažemo da su ove vrednosti slabo predstavljene u ovom skupu podataka.

Everyone	8714
Teen	1208
Mature 17+	499
Everyone 10+	414
Adults only 18+	3
Unrated	2

Slika 3.6: Vrednosti atributa *Content Rating* i broj aplikacija za svaku od tih vrednosti.

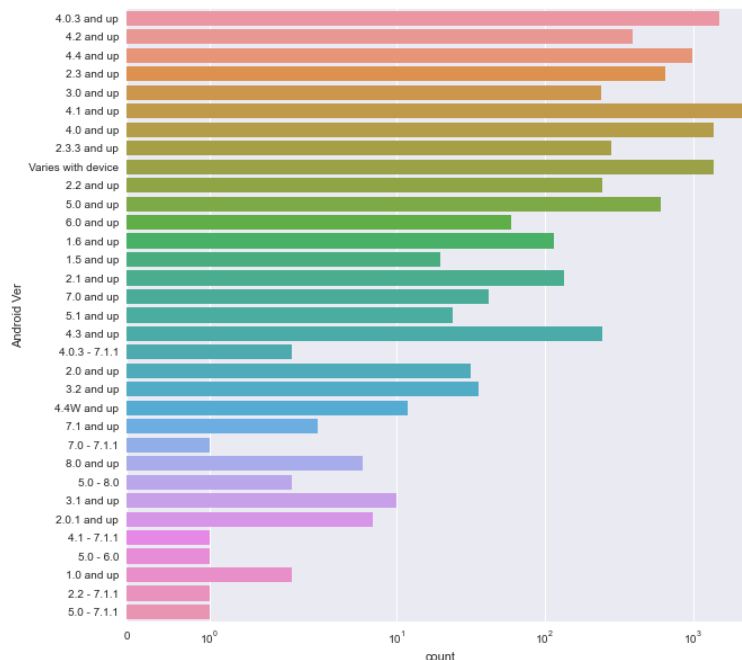
Da bi smanjili veličinu vektora koji se koristi kao ulaz za treniranje modela, ove dve vrednosti možemo da promenimo u vrednost "everyone", koja predstavlja opštu preporuku. Na ovaj način smo smanjili broj vrednosti kojima se predstavlja ovaj atribut.



Slika 3.7: Raspodela nakon obrade atributa *Content Rating*.

Genres predstavlja žanr aplikacije upisan kao string u skupu podataka. Iz skupa podataka možemo da vidimo da je žanr aplikacije isti kao kategorija sa dodatkom nekih drugi odrednica, koje bolje opisuju žanr aplikacije. Kao takav nam daje iste informacije kao atribut *Category*, pa ga odbacujemo kao nepotrebnu informaciju za treniranje modela.

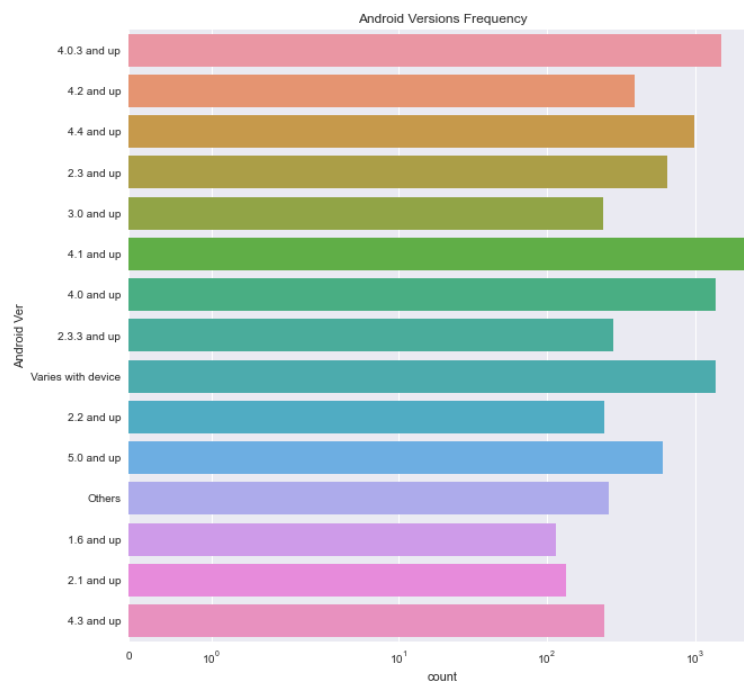
Atributi *Last updated* i *Current Ver* ne daju nikakve informacije o sličnosti između aplikacija, pa ih kao takve ne koristimo za treniranje modela. Sa druge strane atribut *Android Ver* daje korisnu informaciju o aplikaciji. Vrednost ovog atributa u skupu podataka može da bude verzija androida, "Others" i "Varies with device" (slika 3.8).



Slika 3.8: Android verzije i broj aplikacija za svaku verziju.

Možemo da vidimo da su neke od verzija malo zastupljene među aplikacijama. Da bi smanjili ulazni vektor za treniranje modela u slučaju ovog atributa sve vrednosti verzije

androida koje nemaju više od 100 aplikacija u skupu podataka menjamo u vrednost "Others", čime dobijamo bolje predstavljanje svake vrednosti ovog atributa i smanjujemo veličinu ulaznog vektora za treniranje modela (slika 3.9).



Slika 3.9: Android verzije i broj aplikacija za svaku verziju nakon pripajanja malo zastupljenih vrednosti drugim vrednostima.

Nakon analize, obrade i odbacivanja atributa dobijamo skup podataka sa 9 atributa i 10837 redova (slika 3.10). Imamo 5 atributa koji su prikazani brojevima (*Rating*, *Reviews*, *Size*, *Installs* i *Price*) i 4 kategorička atributa prikazana stringovima (*Category*, *Type*, *Content Rating*, *Android Ver*).

	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Android Ver
App									
Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19.0	10000	Free	0.0	Everyone	4.0.3 and up
Coloring book moana	ART_AND_DESIGN	3.9	967	14.0	500000	Free	0.0	Everyone	4.0.3 and up
U Launcher Lite – FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7	5000000	Free	0.0	Everyone	4.0.3 and up
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25.0	50000000	Free	0.0	Teen	4.2 and up
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8	100000	Free	0.0	Everyone	4.4 and up

Slika 3.10: Izgled skupa podataka nakon obrade.

Da bi treniranje modela bilo moguće potrebno je da svi atributi budu predstavljeni brojevnim vrednostima. Za pretvaranje kategoričke vrednosti u numeričku koristimo metodu kodiranja u vektor sa vrednostima 0 i 1 (eng. *one hot encoding*). Svaka vrednost se pretvara u vektor sa vrednostima 0 i 1. Vektor je dužine n , gde n predstavlja broj različitih vrednosti koja kategorička vrednost može da ima. Vektor za neku vrednost ima sve 0 i jednu 1 na odgovarajućoj poziciji. Na ovaj način se jedna kolona u skupu podataka menja sa n kolona koje imaju vrednosti 0 ili 1. Primenom ove metode na sve kategoričke attribute dobijamo skup podataka sa 60 kolona, koje imaju brojevne vrednosti.

Da bi treniranje modela bilo stabilno potrebno je da svi atributi budu u istom opsegu. S obzirom da različiti atributi predstavljaju različite osobine aplikacije, opseg vrednosti im se razlikuje. Vrednosti različitog opsega mogu loše da utiču na treniranje modela, jer će atributi sa većim vrednostima više da utiču na proces treniranja, pa je potrebno da se svi atributi svedu na isti opseg. Da bi se ovo postiglo koristi se metoda *MinMax* skaliranja. Ova metoda mapira opseg između minimalne vrednosti maksimalne vrednosti atributa u opseg između 0 i 1.

Na kraju kada su svi podaci numerički i u opsegu $[0, 1]$, postoji mogućnost da postoje dve aplikacije koje se preslikavaju u isti vektor. Ovakve aplikacije su identične u odnosu na meru sličnosti, pa će mera sličnosti za takve dve aplikacije biti maksimalna, što će rezultovati da se jedna aplikacija izlazi kao preporuka drugoj. Metode koje koriste ovakve mere ne mogu da razlikuju dve aplikacije koje se preslikavaju u isti vektor, ali se predpostavlja da te aplikacije i jesu slične jedna drugoj ako se preslikavaju u isti vektor.

Ovako obrađeni skup podataka pamtimo u *csv* fajlu da ga kasnije koristili za sistem preporuke.

3.3 Implementacija sistema za preporuku

Sistem za preporuku android aplikacije, napravljen je kao aplikacija koja na osnovu imena aplikacije vraća listu od n aplikacija koje su sortirane opadajuće po sličnosti, gde je n broj koji korisnik unosi na početku. Takođe su kao rezultat pretrage date i informacije o aplikacijama koje su izbačene kao rezultat pretrage.

Aplikacije koje su rezultat preporuke koje aplikacija daje se dobijaju korišćenjem modela *k-Nearest-Neighbor*. Ovaj model je istreniran, metodom nenadgledanog učenja, na sređenom skupu podataka, koji je dobijen kao rezultat faze analize i pripreme podataka. Za meru sličnosti *k-Nearest-Neighbor* koristi euklidsku distancu.

Inicijalizacija i treniranje modela:

```
model = NearestNeighbors(metric='euclidean')
model.fit(data_df)
```

Ulazni vektor se dobija učitavanjem odgovarajućeg vektora, za aplikaciju čije je ime korisnik uneo. Model, na osnovu ulaznog vektora, za definisani broj k , vraća k najbližiji vektora. Rezultat se sortira na osnovu sličnosti sa ulaznim vektorom. Za meru sličnosti koristi se kosinusna sličnost. Za rezultujuće vektore se nalaze odgovarajuća imena u skupu imena i prikazuju se korisniku na izlazu.

Metoda za dobijanje preporuka aplikacija:

```
def getRecommendedApps(appname, xdf, app_names, model, recommend_apps=5):  
  
    appID = app_names.index[app_names['App'] == appname].tolist()  
    if (len(appID) != 1):  
        print('Invalid app name. Exit.')  
        return []  
  
    _, neighbors = model.kneighbors(xdf.loc[appID], n_neighbors=recommend_apps+1)  
  
    similar_apps = []  
    for neighbor in neighbors[0][1:]:  
        similar_apps.append(app_names.loc[neighbor][0])  
  
    similarity = []  
    for app in similar_apps:  
        simAppID = app_names.index[app_names['App'] == app].tolist()  
  
        sim = cosine_similarity(xdf.loc[appID], xdf.loc[simAppID]).flatten()[0]  
        similarity.append(sim*100)  
  
    sim_df = pd.DataFrame({'App':similar_apps, 'Similarity':similarity})  
    sim_df.sort_values(by='Similarity', ascending=False)  
  
    return sim_df
```


4 Rezultat

Sistem za preporuku android aplikacija sa *Google Play* prodavnice, implementiran je kao *python* scripta. Korisnik nakon pokretanja aplikacije unosi broj željenih preporuka i ime aplikacije za koju traži preporuku. Aplikacija zatim vraća sortiranu listu preporuka za android aplikaciju.

Korisnik unosi željeni broj preporuka i ime aplikacije, na osnovu koje sistem traži slične aplikacije:

```
Number of similar apps to display: 5
Find similar apps for: VMate
App      Category  Rating Reviews Size  Installs  Type Price Content Rating  Genres  Last Updated  Current Ver  Android Ver
4673 VMate VIDEO_PLAYERS 4.2 192677 17M 50,000,000+ Free 0 Teen Video Players & Editors August 3, 2018 1.78 4.0.3 and up
```

Aplikacija kao rezultat vraća:

```
Similar apps for app VMate are:

App      Similarity
0        Vote for 100.000000
1        Free TV series 99.725480
2        Vigo Video 99.650416
3        Carnivores: Dinosaur Hunter 78.574825
4        LokLok: Draw on a Lock Screen 78.468841

App      Category  Rating Reviews Size  Installs  Type Price Content Rating  Genres  Last Updated  Current Ver  Android Ver
3677 Vote for VIDEO_PLAYERS 4.2 193381 17M 50,000,000+ Free 0 Teen Video Players & Editors August 3, 2018 1.78 4.0.3 and up

App      Category  Rating Reviews Size  Installs  Type Price Content Rating  Genres  Last Updated  Current Ver  Android Ver
3673 Free TV series VIDEO_PLAYERS 3.7 400 7.2M 100,000+ Free 0 Teen Video Players & Editors May 28, 2018 v1.0.3 4.0.3 and up

App      Category  Rating Reviews Size  Installs  Type Price Content Rating  Genres  Last Updated  Current Ver  Android Ver
3686 Vigo Video VIDEO_PLAYERS 4.3 1615596 Varies with device 50,000,000+ Free 0 Teen Video Players & Editors August 3, 2018 Varies with device 4.0.3 and up
4686 Vigo Video VIDEO_PLAYERS 4.3 1615418 Varies with device 50,000,000+ Free 0 Teen Video Players & Editors August 3, 2018 Varies with device 4.0.3 and up

App      Category  Rating Reviews Size  Installs  Type Price Content Rating  Genres  Last Updated  Current Ver  Android Ver
5421 Carnivores: Dinosaur Hunter GAME 4.2 62636 17M 1,000,000+ Free 0 Teen Action July 2, 2018 1.8.0 4.0.3 and up

App      Category  Rating Reviews Size  Installs  Type Price Content Rating  Genres  Last Updated  Current Ver  Android Ver
3899 LokLok: Draw on a Lock Screen COMMUNICATION 4.1 14766 16M 500,000+ Free 0 Teen Communication September 20, 2017 0.81 beta 4.0.3 and up
```

Vidimo da aplikacija vraća listu od pet aplikacija koje su najslićnije unetoj aplikaciji. Takođe možemo da vidimo meru sličnosti pored imena svake aplikacije. Možemo da vidimo da su prve tri aplikacije veoma slične sa aplikacijom koju je korisnik predložio. Iz podataka možemo da vidimo da te aplikacije pripadaju istoj kategoriji i da imaju sličan rejting. Poslednje dve aplikacije nisu baš slične kao aplikacija koju je korisnik predložio, ali se to i ogleda u tome što je vrednost sličnosti manja u odnosu na prve tri.

U sledećem primeru aplikacija vraća:

Find similar apps for: Dictionary - WordWeb														
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
4748 Dictionary - WordWeb	BOOKS_AND_REFERENCE	4.6	124970	Varies with device	5,000,000+	Free	0	Teen	Books & Reference	May 26, 2018	Varies with device	Varies with device		
Similar apps for app Dictionary - WordWeb are:														
App	Similarity													
0 NOOK: Read eBooks & Magazines	99.994280													
1 Audiobooks from Audible	99.899634													
2 Spanish English Translator	99.910068													
3 Mattpad ■ Free Books	99.893138													
4 Bible	99.892910													
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
146 NOOK: Read eBooks & Magazines	BOOKS_AND_REFERENCE	4.5	155446	Varies with device	10,000,000+	Free	0	Teen	Books & Reference	April 25, 2018	Varies with device	Varies with device		
6497 NOOK: Read eBooks & Magazines	BOOKS_AND_REFERENCE	4.5	155466	Varies with device	10,000,000+	Free	0	Teen	Books & Reference	April 25, 2018	Varies with device	Varies with device		
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
5651 Audiobooks from Audible	BOOKS_AND_REFERENCE	4.5	568922	Varies with device	100,000,000+	Free	0	Teen	Books & Reference	August 1, 2018	Varies with device	Varies with device		
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
157 Spanish English Translator	BOOKS_AND_REFERENCE	4.2	87873	Varies with device	10,000,000+	Free	0	Teen	Books & Reference	May 28, 2018	Varies with device	Varies with device		
9570 Spanish English Translator	BOOKS_AND_REFERENCE	4.2	87919	Varies with device	10,000,000+	Free	0	Teen	Books & Reference	May 28, 2018	Varies with device	Varies with device		
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
139 Mattpad ■ Free Books	BOOKS_AND_REFERENCE	4.6	2914724	Varies with device	100,000,000+	Free	0	Teen	Books & Reference	August 1, 2018	Varies with device	Varies with device		
4715 Mattpad ■ Free Books	BOOKS_AND_REFERENCE	4.6	2915189	Varies with device	100,000,000+	Free	0	Teen	Books & Reference	August 1, 2018	Varies with device	Varies with device		
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
3941 Bible	BOOKS_AND_REFERENCE	4.7	2440695	Varies with device	100,000,000+	Free	0	Teen	Books & Reference	August 2, 2018	Varies with device	Varies with device		

U ovom primeru možemo da vidimo da je svih pet predloženih aplikacija veoma slično sa predloženom aplikacijom, jer svih pet aplikacija ima visok koeficijent sličnosti sa predloženom aplikacijom. Možemo da primetimo da sve predložene aplikacije pripadaju istoj kategoriji i da imaju sličan rejting, a pri tome su sve besplatne.

Find similar apps for: Art Drawing Ideas														
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
31 Art Drawing Ideas	ART_AND_DESIGN	4.1	227	5.2M	50,000+	Free	0	Everyone	Art & Design	May 31, 2018	1.2	2.3 and up		
Similar apps for app Art Drawing Ideas are:														
App	Similarity													
0 Easy Realistic Drawing Tutorial	99.998936													
1 Pencil Sketch Drawing	99.975674													
2 How To Draw Food	99.960491													
3 Paper flowers instructions	99.948146													
4 350 Diy Room Decor Ideas	99.769478													
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
29 Easy Realistic Drawing Tutorial	ART_AND_DESIGN	4.1	223	4.2M	100,000+	Free	0	Everyone	Art & Design	August 22, 2017	1.0	2.3 and up		
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
28 Pencil Sketch Drawing	ART_AND_DESIGN	3.9	136	4.6M	10,000+	Free	0	Everyone	Art & Design	July 12, 2018	6.0	2.3 and up		
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
48 How To Draw Food	ART_AND_DESIGN	4.3	845	7.9M	100,000+	Free	0	Everyone	Art & Design	May 28, 2018	1.0	2.3 and up		
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
5 Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up		
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver	
17 350 Diy Room Decor Ideas	ART_AND_DESIGN	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	November 7, 2017	1.0	2.3 and up		

U ovom primeru vidimo da svi rezultati imaju visoku meru sličnosti, i da prva tri rezultata kao i zadata aplikacija služe za crtanje. Dok ostale aplikacije imaju visok skor, ali nisu skroz slične, međutim pripadaju istoj kategoriji i imaju druge slične karakteristike.

```

Find similar apps for: SMS Park
App      Category Rating Reviews Size  Installs Type Price Content Rating Genres Last Updated Current Ver Android Ver
90 SMS Park AUTO_AND_VEHICLES 4.5 3617 7.9M 100,000+ Free 0 Everyone Auto & Vehicles May 16, 2018 2.8.2 4.1 and up

Similar apps for app SMS Park are:

App      Category Rating Reviews Size  Installs Type Price Content Rating Genres Last Updated Current Ver Android Ver
0 Police Lights, Sirens & Follow Me 99.998975
1 Best Car Wallpapers 99.974334
2 ezETC (ETC balance inquiry, meter trial, real-... 99.976977
3 ZUL - Rotativo Digital BH 99.867784
4 FJ Drive: Mercedes-Benz Lease 99.630203

App      Category Rating Reviews Size  Installs Type Price Content Rating Genres Last Updated Current Ver Android Ver
76 Police Lights, Sirens & Follow Me AUTO_AND_VEHICLES 4.5 737 8.9M 100,000+ Free 0 Everyone Auto & Vehicles June 8, 2018 2.3.5.1 4.1 and up
App      Category Rating Reviews Size  Installs Type Price Content Rating Genres Last Updated Current Ver Android Ver
78 Best Car Wallpapers AUTO_AND_VEHICLES 4.5 994 2.9M 100,000+ Free 0 Everyone Auto & Vehicles March 27, 2018 1.6 4.1 and up
62 ezETC (ETC balance inquiry, meter trial, real-... AUTO_AND_VEHICLES 4.3 38846 8.6M ... Everyone Auto & Vehicles July 28, 2018 2.20 Build 02 4.1 and up

[1 rows x 13 columns]

App      Category Rating Reviews Size  Installs Type Price Content Rating Genres Last Updated Current Ver Android Ver
6245 ZUL - Rotativo Digital BH AUTO_AND_VEHICLES 4.2 29 Varies with device 10,000+ Free 0 Everyone Auto & Vehicles August 1, 2018 1.6.5 4.1 and up
App      Category Rating Reviews Size  Installs Type Price Content Rating Genres Last Updated Current Ver Android Ver
10486 FJ Drive: Mercedes-Benz Lease AUTO_AND_VEHICLES 4.6 107 27M 10,000+ Free 0 Everyone Auto & Vehicles November 6, 2017 2.0.0 4.1 and up

```

Ovde vidimo da je za zadatu aplikaciju koja služi za plaćanje parkinga, sistem izbacio preporuke za aplikacije koje su vezane za automobile na neki način. Pošto u skupu podataka verovatno ne postoje druge aplikacije za plaćanje parkinga, sistem za sugestije je izbacio aplikacije koje su u širem kontekstu povezane sa zadatom aplikacijom.

Ako korisnik nije zadovoljan rezultatima koje sistem za preporuku vraća možemo da pokušamo da povećamo broj preporuka koje sistem vraća kako bi korisnik došao do zanimljive aplikacije.

```

Similar apps for app SMS Park are:

App      Similarity
0 Police Lights, Sirens & Follow Me 99.998975
1 Best Car Wallpapers 99.974334
2 ezETC (ETC balance inquiry, meter trial, real-... 99.976977
3 ZUL - Rotativo Digital BH 99.867784
4 FJ Drive: Mercedes-Benz Lease 99.630203
5 Offroad Oil Tanker Driver Transport Truck 2019 99.328573
6 American Muscle Car Race 99.005644
7 Tickets + PDA 2018 Exam 99.053199
8 EV Trip Optimizer for Tesla 98.965045
9 Extreme Rally Championship 97.810306

```

Find similar apps for: Marvel Unlimited

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver
317	Marvel Unlimited	COMICS	3.7	22551	11M	1,000,000+	Free	0		Teen Comics	July 24, 2018	3.16.0	4.4 and up

Similar apps for app Marvel Unlimited are:

	App	Similarity
0	Tapas - Comics, Novels, and Stories	99.314639
1	MTV	77.716857
2	TappyToon Comics & Webtoons	77.430683
3	Comic Es - Shoyo manga / love comics free of c...	77.768959
4	DM for IG 📷 - Image & Video Saver for Instagram	77.538936

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver
318	Tapas - Comics, Novels, and Stories	COMICS	4.5	29839	29M	1,000,000+	Free	0		Teen Comics	July 31, 2018	4.1.0	4.4 and up
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver
949	MTV ENTERTAINMENT	3.8	35279	15M	1,000,000+	Free	0			Teen Entertainment	July 8, 2018	11.45.0	4.4 and up
980	MTV ENTERTAINMENT	3.8	35279	15M	1,000,000+	Free	0			Teen Entertainment	July 8, 2018	11.45.0	4.4 and up
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver
6414	TappyToon Comics & Webtoons	COMICS	3.5	4205	10M	100,000+	Free	0		Teen Comics	July 13, 2018	1.77	4.0.3 and up
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver
310	Comic Es - Shoyo manga / love comics free of c...	COMICS	3.9	2181	10M	100,000+	Free	0		Teen Comics	March 5, 2018	1.2.12	4.0.3 and up
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content	Rating	Genres	Last Updated	Current Ver	Android Ver
8545	DM for IG 📷 - Image & Video Saver for Instagram	SOCIAL	3.7	55	3.7M	5,000+	Free	0		Teen Social	March 3, 2018	3.20.3	4.4 and up

Ovaj primer prikazuje slučaj u kome samo jedna aplikacija ima visoku meru sličnosti sa zadatom aplikacijom. Možemo da vidimo da ta aplikacija najviše odgovara tipu aplikacije koja je zadata i služi za čitanje stripova.

Similar apps for app 8 Ball Pool are:

	App	Similarity
0	Rolling Sky	99.696041
1	Block Craft 3D: Building Simulator Games For Free	99.669156
2	Red Ball 4	99.659479
3	Solitaire TriPeaks	99.615139
4	Beach Buggy Blitz	99.614784
5	Jewels Legend - Match 3 Puzzle	99.597885
6	Word Crossy - A crossword game	99.595839
7	Bubble Shooter	99.568564
8	Banana Kong	99.572601
9	Beauty Idol: Fashion Queen	99.557356

Na predhodnoj slici imamo rezultat sugestija za igru 8 Ball Pool. Ova igra simulira igru bilijara. Ono što možemo da primetimo da su sve sugestije iz kategorije igara, međutim igrice koje je sistem izbacio kao preporuku ne pripadaju istom žanru kao ponuđena igrice, a imaju visoku meru sličnosti sa zadatom aplikacijom. Ovo je rezultat nedovoljnog broja informacija o aplikacijama, što je rezultat obrade skupa podataka. Naime informacije o žanrovima aplikacija su odbačeni jer u dosta slučajeva imaju istu vrednost kao i kategorija. Drugi razlog izbacivanja ove kolone se ogleda u tome što ona ima mnogo različitih vrednosti pa bi prilikom konverzije u vektor 0 i 1 taj vektor imao previše podataka, što je limitacija metoda i pristupa koji se koriste u ovom radu.

Prikazana aplikacija predstavlja dokaz koncepta sistema preporuke baziranog na sadržaju. Sistem preporuke android aplikacija sa *Google Play* prodavnice, treba da ima pristup podacima korisnika ove veb aplikacije, na osnovu kojih bi kreirao preporuke za tog korisnika. Sam sistem bi na osnovu aplikacija koje je korisnik koristio ili koristi, kreirao preporuke kao prikazana aplikacija. Za preporuku bi koristio isti model, a korisniku bi vraćao samo aplikacije koje imaju visok skor sličnosti.

Ovaj projekat predstavlja primer sistema preporuke, kao i metode za pronalaženje preporuka. Takođe kroz ovaj projekat je prikazano kako se podaci koriste i koji su potrebni koraci da bi se dobila funkcionalnost jednog sistema za preporuke.

5 Zaključak

Da bi se korisnicima olakšalo pretraživanje android aplikacija i na taj način povećao broj instalacija android aplikacija, *Google* dosta ulaže u sistem za preporuke android aplikacija sa *Google Play* prodavnice. Takođe sistemi za preporuku su neizostavni deo raznih servisa za strimovanje audio i video fajlova, veb prodavnica i raznih drugih servisa. Sistemi za preporuku su ključni za razvoj *ecommerce*-a.

Ovaj projekat se bavi implementacijom jednog takvog sistema. U okviru ovog projekta je implementiran sistem za preporuku android aplikacija. Bitno je napomenuti da u okviru projekta nije implementiran kompletan sistem za preporuku, jer bi to zahtevalo povezivanje sa *Google Play* servisom. U okviru ovog projekta su prikazane tehnike i metode koje se koriste u realizaciji jednog ovakvog sistema. Takođe je prikazan postupak analize i pripreme podataka. Ova faza je veoma bitna, jer od pripreme podataka zavisi tačnost tehnika i metoda koje se kasnije koriste za kreiranje preporuke. Napravljena je aplikacija koja može da vrati preporuku na osnovu zadate aplikacije.

Sa obzirom da skup podataka sadrži samo deo aplikacija koje mogu da se pronađu u okviru *Google Play* prodavnice, tehnike i metode koje su korišćene u ovom projektu ne bi mogle da se skaliraju u pravom sistemu. Međutim, osnovne ideje, metode kao i postupak pripreme podataka mogu da se primene pri pravljenju bilo kog sistema za preporuku. Ovaj projekat je primer koji pokazuje osnovne funkcije jednog sistema za preporuku, ali i način na koji on može da se realizuje.

Literatura

- [1] Francesco Ricci and Lior Rokach and Bracha Shapira, "Introduction to Recommender Systems Handbook", Recommender Systems Handbook, Springer, pp. 1-35, 2011.
- [2] NumPy API documentation, <https://numpy.org/doc/stable/>, 2021.
- [3] Pandas API documentation, <https://pandas.pydata.org/docs/>, 2021.
- [4] Jupyter documentation, <https://jupyter.org/documentation>, 2021.
- [5] Matplotlib API documentation, <https://matplotlib.org/stable/contents.html>, 2021.
- [6] Seaborn API documentation, <https://seaborn.pydata.org/>, 2021.
- [7] Scikit Learn API documentation, <https://scikit-learn.org/stable/>, 2021.
- [8] Ahlam Fuad, Sahar Bayoumi, Hessah Al-Yahya "A Recommender System for Mobile Applications of Google Play Store", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 9, 2020.