



**Co-funded by
the European Union**

FAAI:

The Future is in Applied Artificial Intelligence
(Budućnost je u Primijenjenoj Vještačkoj Inteligenciji)
Erasmus+ project 2022-1-PL01-KA220-HED-000088359

01.09.2022 – 31.08.2024

Istraživanje 7: Prikupljanje IT specifikacija dobrih praksi u Vještačkoj inteligenciji: analiza stanja u oblasti za radni paket 2 (WP2)





**Co-funded by
the European Union**

Izrada ovog dokumenta bila je moguća zahvaljujući podršci ERASMUS+ projekta: Budućnost je u Primijenjenoj Vještačkoj Inteligenciji (2022-1-PL01-KA220-HED-000088359)

Finansiran od strane Europske Unije. Izraženi stavovi i mišljenja su, ipak, samo autorova(-a) i ne odražavaju nužno stavove i mišljenja Europske Unije ili Nacionalne Agencije (NA). Za njih se ne mogu smatrati odgovornima ni Europska unija ni NA.



Datum

15.09.2021

Mjesta na kojima su rezultati dobijeni

Univerzitet Bielsko-Biala (UBB), Poljska

Univerzitet za Biliotekarstvo i Informacione Tehnologije (ULSIT), Sofija, Bugarska

Univerzitet u Nišu (UNi), Srbija

Univerzitet Svetih Ćirila i Metodija (USCM), Trnava, Slovačka

Univerzitet Crne Gore (UCG), Crna Gora

Rezime: 10 redova

Ključne riječi: manje od 5 ključnih riječi

1. Uvod

30 redova

2. Prikupljanje i analiza podataka

Podatke su prikupili istraživači sa 5 partnerskih naučnih institucija (univerziteta). U ovom istraživanju 8 upitnika dobijeno je od UBB, 6 upitnika od ULSIT, 5 od UNi, 5 od UCG i 1 od USCM. Ukupno je 11 istraživača prikupilo 25 upitnika.

3. Rezultati

3.1. Država u kojoj se obuka održava

Prvo pitanje upitnika, nakon podataka o nazivu organizacije i samog istraživača, glasilo je: Gdje se odvija obuka? Rezultati su prikazani u nastavku.



Prevod kategorija na grafiku: Bugarska, Njemačka, Italija, Crna Gora, Holandija, Poljska, Portugal, Srbija, Slovačka, Španija.

Opis rezultata:

Prema rezultatima istraživanja, obuke se uglavnom odvijaju u Srbiji, a zatim u Bugarskoj. Jedan je ispitanik odgovorio da je za treninge koristio samo Coursera ili O'Reilly. Između jedne i tri obuke održavaju se u drugim državama uključenim u istraživanje.

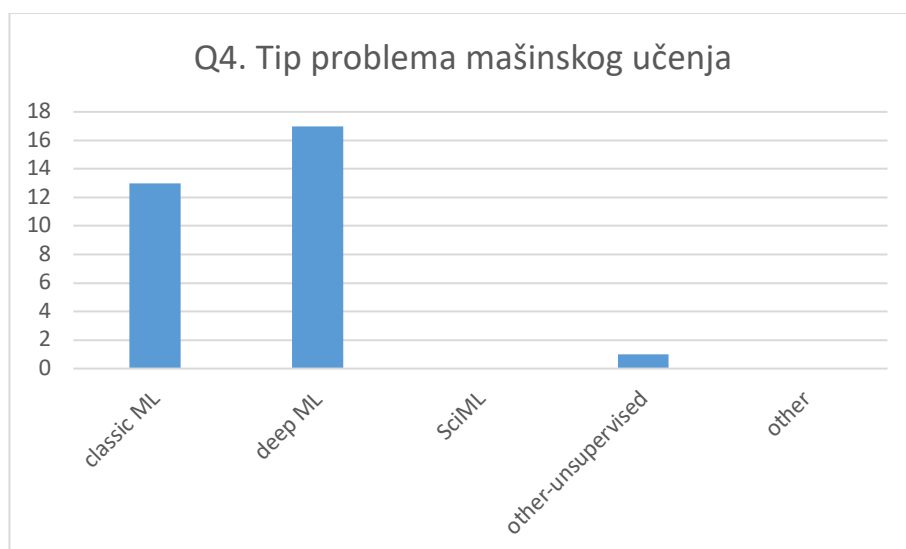
Diskusija:

Istraživanje pokazuje da postoji potreba za više obuka za IT stručnjake kako bi se moglo uočiti više dobrih praksi u domenu vještačke inteligencije. Glavni zaključci:

- Postoji potreba za više obuka u zemljama EU.
- Najviše treninga ima u Srbiji i Bugarskoj.

3.2. Tipovi problema mašinskog učenja

Sljedeće pitanje otkriva tipove problema mašinskog učenja koji se koriste.



Prevod kategorija na grafiku: klasično mašinsko učenje, duboko mašinsko učenje, SciML, drugi-nenadgledano, drugi.

Opis rezultata:

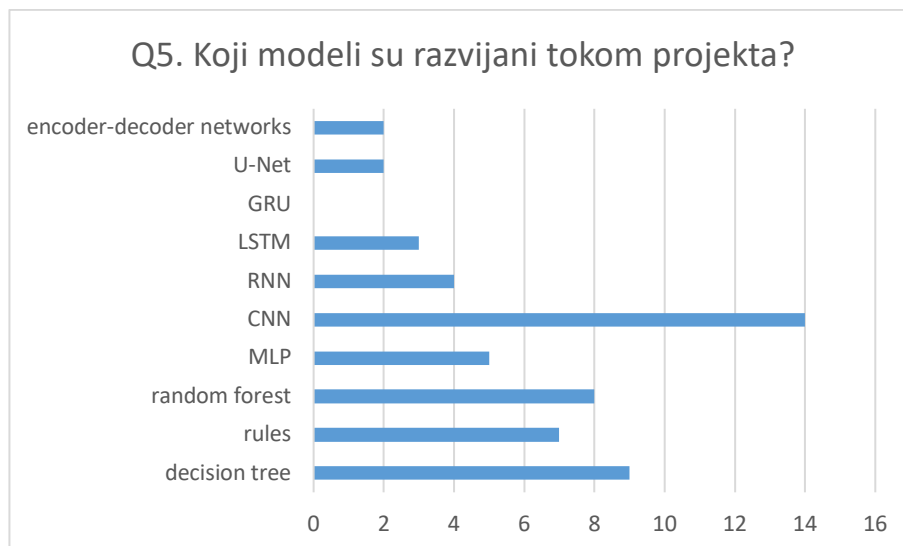
Rezultati pokazuju da je najviše korišćeni tip mašinskog učenja duboko učenje, a potom klasično mašinsko učenje. Tokom proteklih godina duboko učenje postalo je glavni alat za razne probleme vještačke inteligencije. Pokazalo je bolje performanse na različitim zadacima, uključujući obradu prirodnih jezika, kompjuterski vid, virtuelne pomoćnike, chatbotove, primjene u zdravstvu itd.

Diskusija:

- Najčešće korišćeni tip, na osnovu ankete, je duboko mašinsko učenje.
- SciML i drugi tipovi mašinskog učenja nisu uobičajeni u obukama.

3.3. Koji modeli su razvijani i obrađivani tokom projekta?

Sljedeće pitanje se odnosi na modele koji su razvijani tokom projekata. Rezultati slijede.



Prevod kategorija na grafiku: encoder-dekoder mreže, U-Net, GRU, LSTM, RNN, CNN, MLP, slučajne šume, drvo odlučivanja.

Opis rezultata:

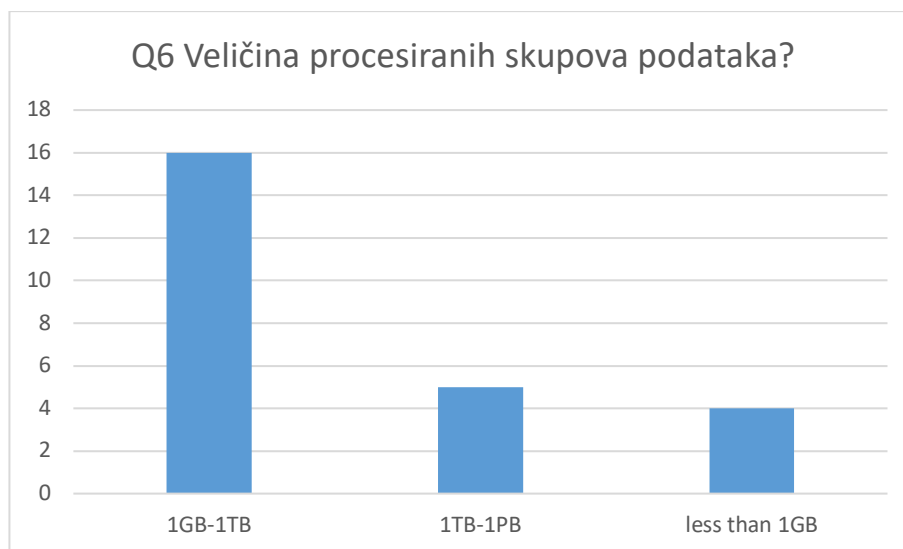
Prema rezultatima ankete, modeli koji su se najviše koristili u projektima su konvolucione neuronske mreže. To je i razumljivo jer ti modeli imaju različite primjene – klasifikacija slika, detekcija objekata, prepoznavanje lica, analiza medicinskih slika, obrada prirodnih jezika, autonomna vozila. Ovi algoritmi su posebno prikladni za zadatke prepoznavanja slika jer mogu automatski naučiti da detektuju složene karakteristike kao što su ivice, uglovi i teksture. Stablo odlučivanja, zajedno sa Slučajnom šumom, drugi su modeli koji su razvijeni za različite probleme vještačke inteligencije. Mnoge primjene Stabala odlučivanja i Slučajnih šuma široko se koriste u industrijama kao što su finansije, zdravstvo, proizvodnja i monitoring okoline. Stabla odlučivanja i Slučajne šume popularni su algoritmi mašinskog učenja jer su jednostavni za upotrebu, razumijevanje i tumačenje, mogu obrađivati i numeričke i kategoričke podatke, te velike i složene skupove podataka.

Diskusija:

- Konvolucione neuronske mreže su najčešće korišćeni modeli u projektima
- GRU (Gated recurrent unit) modeli se nisu razvijali u projektima

3.4. Veličina procesiranih skupova podataka?

Sljedeće pitanje se odnosi na veličinu procesiranih skupova podataka. Mogući odgovori: manje od 1GB, 1GB – 1TB, 1TB – 1PB i preko 1 PB. Rezultati su dati u nastavku.



Prevod kategorija na grafiku: 1GB-1TB, 1TB-1PB, manje od 1GB.

Opis rezultata:

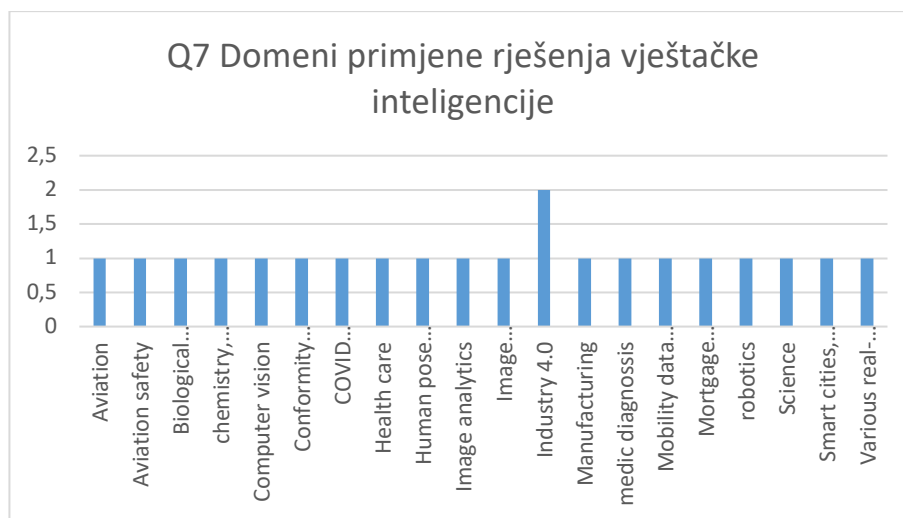
Kao što se može vidjeti na grafikonu, količina obrađenih podataka bila je uglavnom između 1 GB i 1 TB. Ta veličina predstavlja relativno velike količine podataka koji se obično sreću u mnogim različitim vrstama projekata, ali su još uvijek podnošljivi u smislu mogućnosti pohranjivanja i obrade najsavremenijih računara. Količina obrađenih podataka veća od 1 TB nije tako česta, vjerovatno zbog pristupačnosti i cijene. Upravljanje i obrada skupova podataka na nivou petabajta zahtijeva specijalizovanu infrastrukturu i resurse koji nisu lako dostupni većini organizacija. Pohranjivanje i obrada velikih količina podataka mogu biti skupi, kako u smislu hardverskih tako i operativnih troškova. S druge strane, ni skupovi podataka veličine manje od 1 GB nijesu uobičajeni jer se smatraju suviše malim da se mogu učinkovito obraditi na jednoj mašini, bez potrebe za specijalizovanim hardverom ili distribuiranim računarskim sistemom.

Diskusija:

- Najčešća veličina procesiranog skupa podataka je između 1 GB and 1 TB.
- Skupovi podataka veći od 1 TB ili manji od 1 GB nisu primjenjivi.

3.5. Koji su domeni primjene rješenja vještačke inteligencije?

Sljedeće pitanje se odnosi na domene primjene rješenja vještačke inteligencije. Odgovori uključuju i direktne slobodne unose ispitanika, za svaki pojedinačni slučaj. Rezultati su prikazani grafički ispod.



Prevod kategorija na grafiku: Avijacija, bezbjednost u avijaciji, biološke sekvence, hemija, kompjuterski vid, vizuelna provjera usklađenosti mašina sa CAD modelom, dijagnostika COVID-a, zdravstvena njega, procjena položaja ljudskog tijela, analitika slika, obrada slike, Industrija 4.0, proizvodnja, medicinska dijagnoza, mobilnost, industrija hipoteka, robotika, nauka, pametni gradovi, razni slučajevi iz stvarnog svijeta.

Opis rezultata:

Treba napomenuti da su istraživači tražili projekte u različitim poljima, a to su: poljoprivreda, medicina, hirurgija, upravljanje avionskim saobraćajem, vazduhoplovstvo, sigurnost avionskog saobraćaja, analiza bioloških sekvenci, hemija, robotika, zdravstvo, računarski vid, vizuelna provjera usklađenosti mašina sa CAD modelom, dijagnostika COVID-a, procjena položaja ljudskog tijela, analitika slika, obrada slike, predviđanje cijena, Industrija 4.0, proizvodnja, medicinska dijagnoza, mobilnost, industrija hipoteka, pametni gradovi, praćenje saobraćaja, razni slučajevi iz stvarnog svijeta. Svaki je ispitanik istražio jedno ili dva od ovih domena implementacije rješenja vještačke inteligencije.

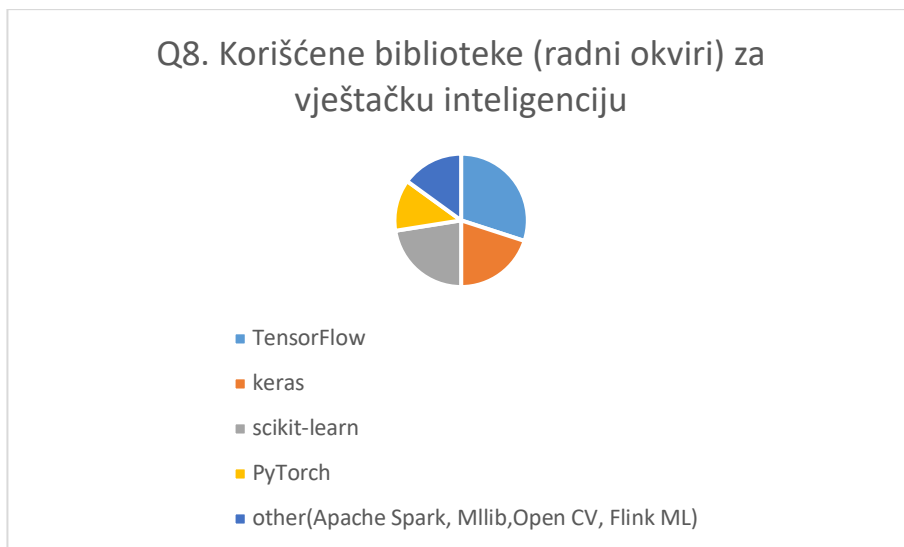
Diskusija:

- Svaki je ispitanik je dobio zadatak da istraži određeni domen kako bi se dobio pregled što više domena.

3.6. Korišćene biblioteke (radni okviri) za vještačku inteligenciju

Pitanje 8 se odnosi na najčešće korišćene biblioteke i radne okvire u domenu vještačke inteligencije. Rezultati su prikazani na grafikonu u nastavku.

Q8. Korišćene biblioteke (radni okviri) za vještačku inteligenciju



Prevod kategorija na grafiku: Tensorflow, keras, scikit-learn, PyTorch, drugi (Apache Spark, MLib, OpenCV, Flink ML)

Opis rezultata:

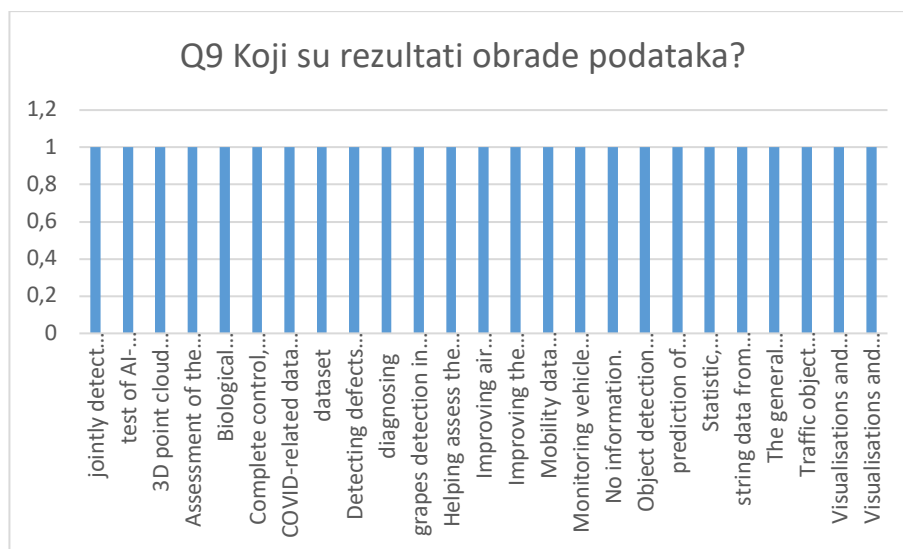
Rezultati pokazuju da je najkorištenija AI biblioteka TensorFlow. TensorFlow jedna je od najpopularnijih i naširoko korišćenih biblioteka otvorenog koda za razvoj i implementaciju modela mašinskog učenja i dubokog učenja. Popularnost te biblioteke je tolika zahvaljujući i velikoj i aktivnoj zajednici koja pruža podršku i doprinosi njenom razvoju. Druge popularne AI biblioteke su Keras i Scikit-learn. Ostale biblioteke koje se koriste u problemima vještačke inteligencije uključuju PyTorch, Apache Spark, MLib, Open CV i Flink ML.

Diskusija:

- Izbor biblioteke koja će se koristiti zavisi od različitih faktora, uključujući specifične potrebe projekta, lične preferencije i nivo ekspertize programera.

3.7. Koji su rezultati obrade podataka?

Ovo pitanje podrazumijevalo je otvorene odgovore za svaki pojedinačni slučaj. Stoga je bilo 25 različitih rezultata za svaki projekat.



Prevod kategorija na grafiku: zajednička detekcija ključnih tačaka ljudskog tijela, lica i stopala, test AI-TWILIGHT metoda u domenama automobilizma, hortikulture i ulične rasvjete, klasifikacija i segmentacija 3D oblaka točaka, procjena vjerojatnoće i težine infekcije Covid-19, klasifikacija bioloških sekvenci, potpuna kontrola puna artikulacija i inteligentna povratna informacija, analitika podataka u vezi s COVID-om, izrada skupa podataka, otkrivanje nedostataka na proizvodima u proizvodnji, dijagnostika, detekcija grožđa u vinogradu, pomoć u procjeni stanja mašina i ukupne opreme unutar fabrike, poboljšanje upravljanja avionskim saobraćajem kroz saradnju mašinskog učenja na privatnim skupovima podataka, poboljšanje procesa donošenja odluka u scenarijima prelaska, analiza podataka o mobilnosti, praćenje vozila koja se kreću u gradovima brojanjem automobila sa slika dobijenih pametnim kamerama, bez informacije, otkrivanje i prepoznavanje predmeta, predviđanje aktivnosti hemijskih spojeva, automatsko upravljanje robotima, analiza ljudskih pokreta pomoću nosivih senzora; statistika, propisivanje lijekova i liječenja, dobijanje podataka iz slika i objekta iz slika, pokušaj razvoja prototipa globalnog sistema za praćenje više opasnosti i ranog upozoravanja, detekcija i prepoznavanje saobraćajnih objekata, vizualizacije i informacije o budućim cijenama, vizualizacije i modele mašinskog učenja.

Opis rezultata:

U prvom projektu rezultat je bila zajednička detekcija ključnih tačaka ljudskog tijela, lica i stopala. U drugom projektu rezultat je bio test AI-TWILIGHT metoda u domenama automobilizma, hortikulture i ulične rasvjete. Treći rezultat bila je klasifikacija i segmentacija 3D oblaka točaka. Četvrti rezultat bila je procjena vjerojatnoće i težine infekcije Covid-19. Peti rezultat bila je klasifikacija bioloških sekvenci. Šesti rezultat bila je potpuna kontrola, puna artikulacija i inteligentna povratna informacija. Sedmi rezultat bila je analitika podataka u vezi s COVID-om. Osmi rezultat bila je izrada skupa podataka. Deveti rezultat bio je usmjeren na otkrivanje nedostataka na proizvodima u proizvodnji. Deseti rezultat bio je usmjeren

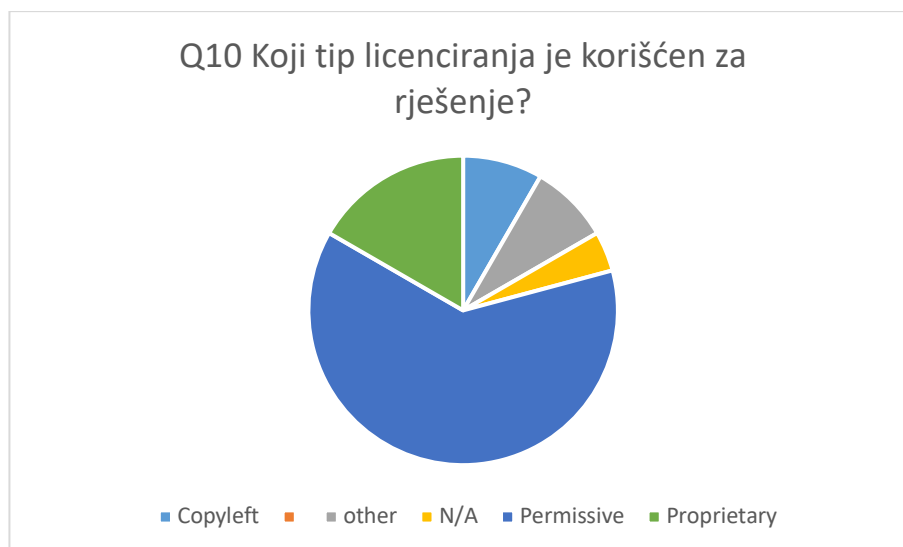
na implementaciju u dijagnostici. Jedanaesti rezultat je detekcija grožđa u vinogradu. Dvanaesti rezultat bila je pomoć u procjeni stanja mašina i ukupne opreme unutar fabrike. Trinaesti rezultat bilo je poboljšanje upravljanja avionskim saobraćajem kroz saradnju mašinskog učenja na privatnim skupovima podataka. Četrnaesti rezultat bilo je poboljšanje procesa donošenja odluka u scenarijima prelaska, što je od velike važnosti za sigurnost avio prijevoznika i pružatelja usluga u vazduhoplovstvu u ATM-u. Petnaesti rezultat bila je analiza podataka o mobilnosti. Šesnaesti rezultat bio je praćenje vozila koja se kreću u gradovima brojanjem automobila sa slika dobijenih pametnim kamerama. Sedamnaesti rezultat bilo je otkrivanje i prepoznavanje predmeta. Osamnaesti rezultat bilo je predviđanje aktivnosti hemijskih spojeva, automatsko upravljanje robotima, analiza ljudskih pokreta pomoću nosivih senzora. Devetnaesti rezultat bila je statistika, propisivanje lijekova i liječenja. Dvadeseti rezultat bio je dobijanje podataka iz slika i objekta iz slika. Dvadeset i prvi rezultat bio je pokušaj razvoja prototipa globalnog sistema za praćenje više opasnosti i ranog upozoravanja. Dvadeset i drugi rezultat je detekcija i prepoznavanje saobraćajnih objekata. Dvadeset i treći rezultat predstavlja vizualizacije i informacije o budućim cijenama. Dvadeset i četvrti rezultat predstavlja vizualizacije i modele mašinskog učenja. Postojao je jedan projekt u kojem nisu dokumentovani rezultati.

Diskusija:

- Svaki je ispitanik morao pronaći rezultate nakon obrade podataka, zavisno od pojedinačnog projekta. Na taj su način otkriveni različiti ishodi u domenama primjene vještačke inteligencije.

3.8. Koji tip licenciranja je korišćen za rješenje?

Sljedeće pitanje se odnosi na to koja se vrsta licenciranja koristi u rješenju i uključuje četiri alternative – permisivno (BSD, MIT), copyleft (GPL, LGPL), vlasničko (Bespoke, Commercial) i ostalo. Grafikon u nastavku prikazuje distribuciju ovih vrsta licenciranja.



Prevod kategorija na grafiku: copyleft, drugi, N/A, permisivna, vlasnička.

Opis rezultata:

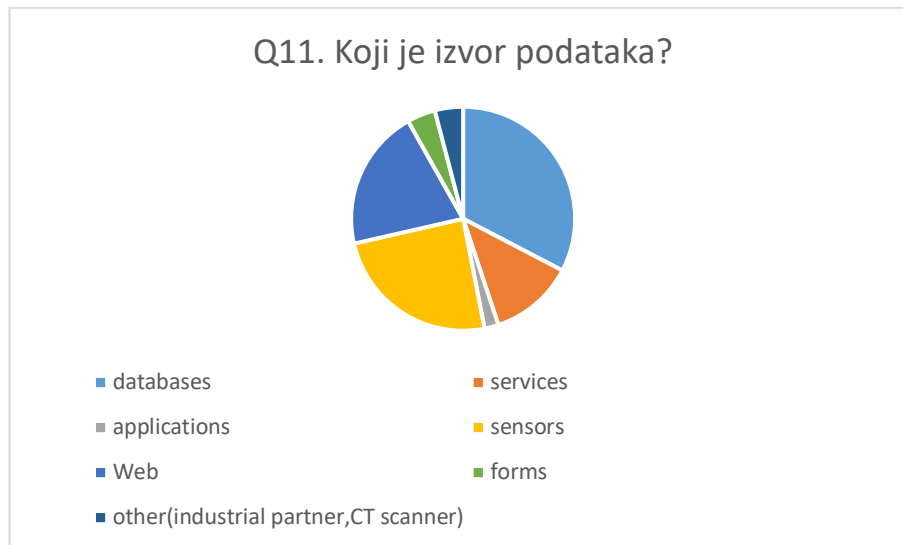
Rezultati pokazuju da je najčešće korišćena vrsta licenciranja permisivna (u 15 projekata), zatim vlasnička (u 4 projekta) a potom copyleft (u 2 projekta). U jednom projektu korišćena je izvorna Licenca OpenPose (slobodno dostupna za nekomercijalnu upotrebu). Permisivne licence popularne su jer nude visok stupanj slobode korisnicima i programerima. Ove vrste licenci obično dopuštaju korisnicima da mijenjaju i redistribuiraju softver bez potrebe da se bilo kakve promjene ili poboljšanja izdaju pod istom licencom. Ovo je u suprotnosti s copyleft licencama, koje zahtijevaju da svi izvedeni radovi budu licencirani pod istim uslovima kao i original. Permisivne licence često biraju pojedinci i organizacije koji žele podstaći saradnju i inovacije, dok u isto vrijeme omogućavaju maksimalnu fleksibilnost u načinu na koji se softver koristi i distribuira. Vlasničko licenciranje zauzima drugo mjesto iako se nalazi u samo četiri projekta. Ta je vrsta licenciranja uobičajena u softverskoj industriji, ali često ograničava načine na koje korisnici mogu koristiti i distribuirati softver. Neke kompanije još uvijek odlučuju da koriste vlasničke modele licenciranja kao način zaštite svog intelektualnog vlasništva i održavanja kontrole nad svojim softverom. Licence Copyleft nisu tako uobičajene kao permisivne ili vlasničke licence jer nameću veća ograničenja na način na koji se softver može koristiti i distribuirati.

Diskusija:

- Permisivne licence su popularne jer nude ravnotežu slobode i fleksibilnosti koja omogućuje saradnju i inovacije dok minimizira prepreke ulasku i usvajanju.
- Copyleft licence rjeđe su od permisivnih ili vlasničkih licenci jer nameću više ograničenja, što možda nije poželjno za sve korisnike i programere.

- **3.9 Koji je izvor podataka?**

Pitanje 11 daje odgovore u vezi s izvorima podataka. Opcije idu od baza podataka, usluga, servisa, aplikacija, do senzora, Weba, obrazaca (ako je odgovor “Ništa od navedenog”, ispitanik unosi izvor slobodnim unosom). “Pita” grafikom u nastavku prikazuje odgovore i njihovu distribuciju.



Prevod kategorija na grafiku: baze podataka, aplikacije, web, drugi (industrijski partner, CT skener), servisi, senzori, obrasci.

Opis rezultata:

Rezultati pokazuju da su osnovni izvor podataka baze podataka (u 16 upitnika), zatim senzori (u 12 upitnika) i Web (u 10 upitnika). Najrjeđi izvor su obrasci (u 2 upitnika), ostali: poput industrijskog partnera ili CT skenera (u 2 upitnika) i aplikacije (u 1 upitniku). Baze podataka su preferirani izvori jer su podaci tamo strukturirani i mogu se efikasno učítavati. Osim toga, podaci su konzistentni i tačni, a baze podataka se mogu lako integrisati s drugim sistemima, što olakšava razmjenu informacija između različitih aplikacija i platformi. S druge strane prijave i obrasci nisu čest izvor informacija jer su podaci često raštrkani i njihov kvalitet nekonzistentan. Aplikacije ili obrasci možda nisu dizajnirani za rukovanje velikim količinama podataka i pristup može biti ograničen zbog zabrinutosti za sigurnost ili privatnost. Osim toga, obrasci mogu imati različite formate podataka i protokole, što može otežati integraciju podataka u različite sisteme i aplikacije.

Diskusija:

- Baze podataka su preferirani izvor podataka jer nude pouzdan, učinkovit i siguran način za pohranu i upravljanje velikim količinama podataka na dosljedan i strukturiran način.

- Aplikacije i obrasci mogu biti izvori podataka, ali često zahtijevaju značajan napor i resurse da bi se podaci izdvojili i učinkovito koristili.

3.10. Prikaz podataka

Ovo pitanje daje informacije o vrsti podataka. Sljedeće vrste su uključene u projekte - grafikoni, podaci sa senzora, tekst, slike, csv, avro, parquet, JSON API, MySQL server, FTP server, 3D oblak tačaka, CT grudnog koša, medicinska istraživanja.

3.11. Koje su karakteristike podataka?

Sljedeće pitanje daje informacije o karakteristikama podataka. Mogući odgovori su slike, tekst, video, audio ili drugi.



Prevod kategorija na grafiku: 3D oblaci tačaka, grafici i podaci sa senzora, PDF dokumenti i slike, slike, slike, tekst, tekst, tekst, tekst, tekst, tekst, tekst, tekst, tekst, tekst, video, video.

Opis rezultata:

Kao što se može vidjeti na grafikonu, glavne karakteristike podataka su tekstovi i slike ili slike s tekстом ili video. Slike i tekstovi su lako razumljivi i interpretirani. Mogu se generisati u velikim količinama i mogu se koristiti u različitim kontekstima, od marketinga i oglašavanja do naučnih istraživanja i analize podataka. Slikama i tekstovima može se lako pristupiti i mogu se dijeliti na različitim platformama.

Diskusija:

Slike i tekstovi uobičajene su karakteristike podataka jer su laki za razumijevanje, generišu se u velikim količinama, svestrani su i dostupni. Napredak tehnologije takođe

je olakšao analizu i izvlačenje zaključaka iz nestrukturiranih podataka, čineći ih vrijednim izvorom informacija za kompanije i istraživače.

3.12. Obrada i kvaliteta podataka

Pitanje 14 daje informacije o kvalitetu podataka i načinu na koji su podaci obrađeni. Korišćene su različite metode zavisno od slučaja. U nekim projektima donesene su stručne odluke na temelju medicinskog vještačenja. U drugim projektima datotekama se pristupalo i formatirale su se putem Pythona. U nekim projektima podaci su očišćeni i vizualizovani. U drugim projektima korišćena je klasifikacija i regresija. U nekim slučajevima primijenjeno je skaliranje, označavanje, audio ili video obrada. Utvrđeno je da su ove tehnike najčešće u obradi podataka.

3.13. Koji se alati koriste za pohranu podataka?

Ovo pitanje je otvoreno i daje informacije o alatima koji se koriste za pohranu podataka u različitim projektima. Grafikon u nastavku prikazuje odgovore.



Prevod kategorija na grafiku: 3D fajlovi, klaster SVCS, fajl system, Graph baze podataka, HDFS, HBase, slike I json fajlovi, mariaDB, NewSQL, NoSQL, NoSQL I HDFS, nije precizirano, nije precizirano, PostgreSQL Parquet, SQL, SQL I NoSQLbaze podataka, SQL baze podataka i NoSQL, SQL csv fajlovi, X zruci MRI slike.

Opis rezultata:

Najčešći alati su NoSQL baze podataka. Te su baze podataka dizajnirane za rukovanje nestrukturiranim ili polustrukturiranim podacima, što ih čini boljim izborom od

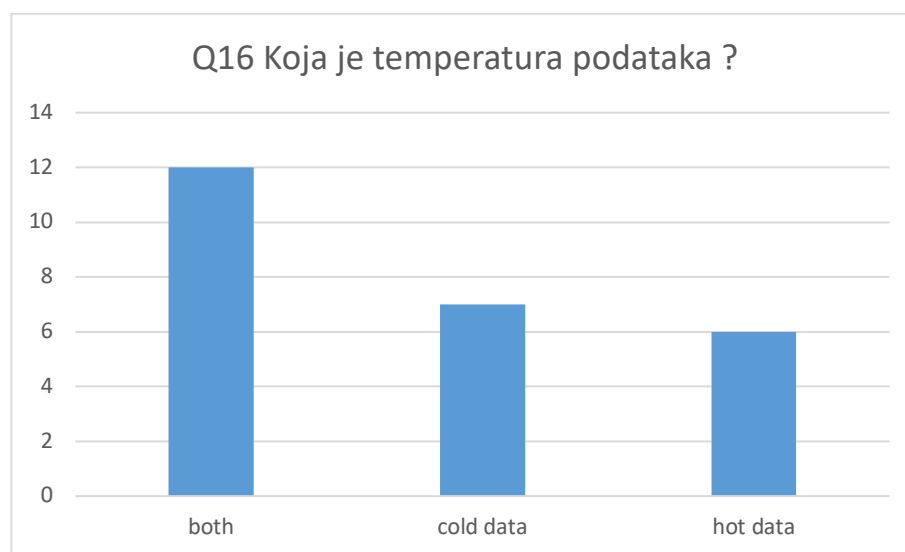
tradicionalnih relacionih baza podataka za aplikacije koje zahtijevaju fleksibilno modeliranje podataka. To programerima omogućuje pohranjivanje i dohvaćanje podataka na način koji bolje odgovara potrebama njihove aplikacije. Mnoge NoSQL baze podataka su otvorenog koda, što ih čini isplativom opcijom za programere i organizacije. Ostali alati su gotovo jednako raspoređeni, pa se može zaključiti da alat koji će se koristiti zavisi od slučaja i IT problema.

Diskusija:

Uopšteno, NoSQL baze podataka nude nekoliko prednosti u odnosu na tradicionalne relacione baze podataka i zato postaju sve češće kao alati za pohranu podataka. Oni nude više fleksibilnosti, skalabilnosti, dostupnosti i performansi, što ih čini idealnim izborom za moderne aplikacije koje zahtijevaju te karakteristike.

3.14. Koja je temperatura podataka?

Sljedeće pitanje otkriva jesu li najčešće korišćeni podaci hladni, vrući ili oboje. Grafikon u nastavku prikazuje distribuciju ovih vrsta.



Prevod kategorija na grafiku: oboje, hladni podaci, vrući podaci.

Opis rezultata:

Jasno se može uočiti da se koriste i hladni i vrući podaci. Hladni podaci obično su jeftiniji za pohranjivanje, jer se često pohranjuju na sporijim, jeftinijim uređajima za pohranjivanje poput trake ili tvrdih diskova. To ih čini idealnim rješenjem za pohranu istorijskih podataka, sigurnosnih kopija i arhiva, kojima se možda neće morati često pristupati, ali ih je potrebno pohraniti zbog usklađenosti ili pravnih razloga. Vrući podaci, s druge strane, obično se pohranjuju na bržim i skupljim uređajima za

pohranu. Ovoj vrsti podataka često pristupaju aplikacije, korisnici ili usluge i moraju biti brzo dostupni kako bi podržali operacije, transakcije ili analitiku u realnom vremenu. U većini organizacija za poslovanje su potrebni i topli i hladni podaci.

Diskusija:

I hladni i vrući podaci podjednako su uobičajeni jer služe različitim svrhama i potrebni su u različito vrijeme. Hladni podaci odnose se na podatke kojima se pristupa rijetko ili se uopšte ne pristupa, dok se vrući podaci odnose na podatke kojima se često pristupa ili se aktivno koriste.

3.15. Jesu li sigurnosne karakteristike podržane u projektima?

Sljedeće pitanje postavlja pitanje jesu li sigurnosne karakteristike podržane u projektima. Alternative su da, ne ili djelomično.



Prevod kategorija na grafiku: ne, djelimično, da.

Opis rezultata:

Rezultati pokazuju da nešto češće sigurnosne karakteristike nisu podržane nego obrnuto. Moguće objašnjenje moglo bi biti da implementacija robusnih sigurnosnih karakteristika može oduzeti vrijeme i resurse te može zahtijevati značajna ulaganja u hardver, softver i osoblje. U nekim slučajevima organizacije mogu dati prednost drugim karakteristikama ili funkcijama u odnosu na sigurnost zbog budžetskih ili vremenskih ograničenja. Drugi vjerovatni razlog može biti nedostatak dovoljnog stručnog znanja jer sigurnost može biti složeno i specijalizovano područje i nemaju svi programeri ili organizacije resurse ili znanje za pravilno rješavanje sigurnosnih

problema. Međutim, razlike su male, pa se ne može generalizovati da sigurnosne karakteristike nisu podržane, samo u nekim slučajevima.

Diskusija:

- Sigurnosne karakteristike nisu podržane u većini projekata.
- Važno je da organizacije i programeri daju prioritet sigurnosti i preduzmu korake za implementaciju snažnih sigurnosnih mjera za zaštitu svojih korisnika i njihovih podataka.

3.16. Koji se alati koriste pri obradi podataka i osiguravanju njihovog kvaliteta?

Ovo pitanje sugerise različite opcije, u pogledu alata prilikom obrade podataka. Opcije su Talend, Toro, Soda, Datafold, Databand, Precisely, Monte Carlo ili druge. Distribucija rezultata prikazana je u grafikonu ispod.



Prevod kategorija na grafiku: databand, datafold, Monte Carlo, N/A, nema informacije, nije precizirano, Python skripte, R jezik, Soda, specijalni alati, Tensorflow PCL, ne koristimo alate.

Opis rezultata:

Rezultati su uglavnom "N/A" ili "nema informacija" što sugerise da u analiziranim studijama istraživači nisu spomenuli alate koje su koristili prilikom obrade podataka. Datafold zauzima treće mjesto. Sve se više koristi i smatra se vrijednim alatom za radne tokove obrade podataka.

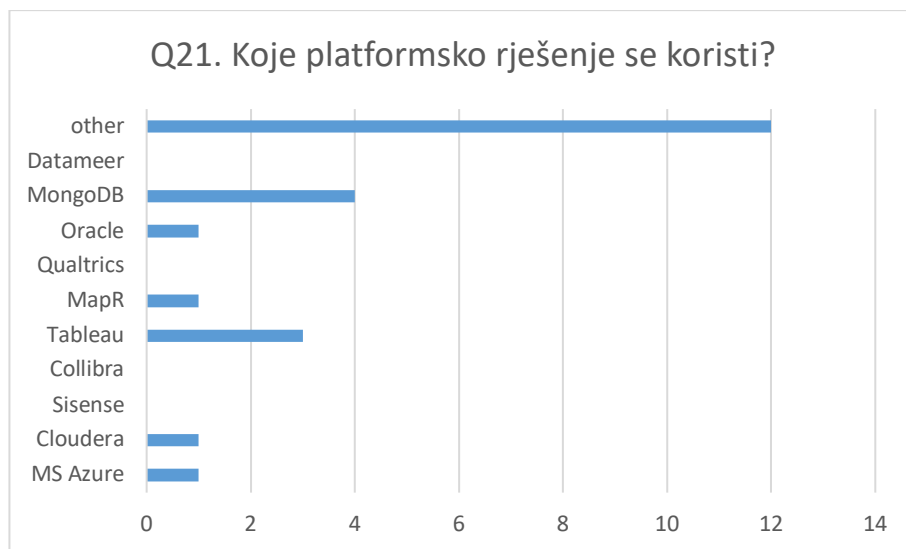
Diskusija:

- Alati koji se koriste pri obradi podataka rijetko se spominju.

- Datafold je relativno novi alat za obradu podataka koji je stekao popularnost među data inženjerima i data istraživačima.

3.17. Koje platformsko rješenje se koristi?

Sljedeće pitanje otkriva koje se platformsko rješenje koristi. Mogući odgovori su MS Azure, Cloudera, Sisense, Collibra, Tableau, MapR, Qualtrics, Oracle, MongoDB, Datameer ili drugi koje upisuje ispitanik. Rezultati su prikazani u grafikonu.



Prevod kategorija na grafiku: drugi, Datameer, MongoDB, Oracle, Qualtrics, MapR, Tableau, Collibra, Sisense, Cloudera, MS Azure.

Opis rezultata:

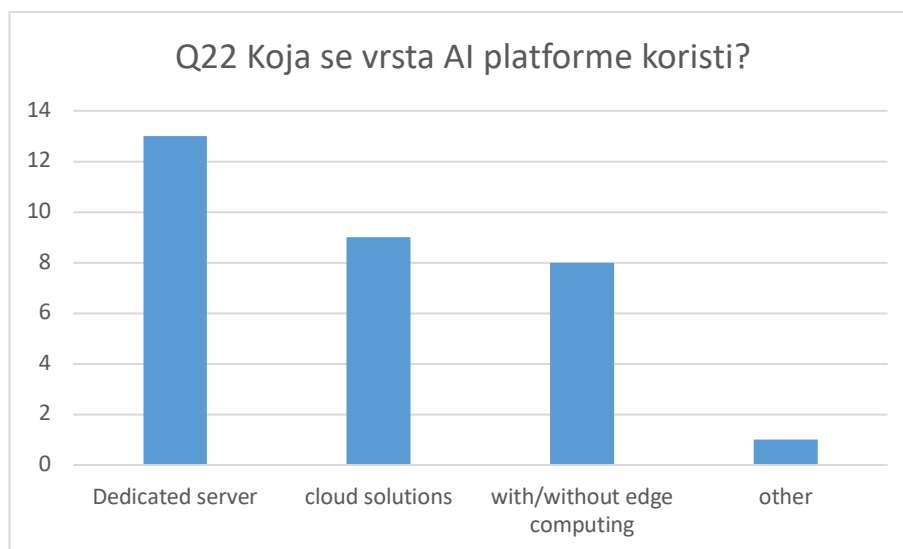
Platformska rješenja koja se koriste u projektima su pretežno drugačija u odnosu na ponudjena u anketi. Oni uključuju R studio, NVIDIA CUDA, Clara, CUDA-X, TensorFlow/TensorRT, Anaconda, Google Colab, Apache Spark, Apache Flink, PySpark, AWS, Bonseyes, specijalizovani radiološki informacijski sistem, Amazon Web Services S3 Data Lake, FedML, OpenMMLab, gRPC.

Diskusija:

- Platformska rješenja koja se predlažu kao odgovori u upitniku nisu toliko popularna. MongoDB i Tableau među ponuđenim odgovorima zauzimaju drugo i treće mjesto.
- Druge platforme, kao što su Anaconda, Apache, NVIDIA, R studio, TensorFlow itd. koriste se u više slučajeva.

3.18. Koja se vrsta AI platforme koristi (npr. serverska, u oblaku, s/bez podrške za ugnježdjeno računarstvo ili drugo)?

Sljedeće pitanje odnosi se na vrstu AI platforme. Moguće alternative su namjenski serveri, rješenja u oblaku, s/bez podrške za ugnježdjeno računarstvo ili druge vrste, koje su unijeli ispitanici. Rezultati su prikazani u grafikonu.



Prevod kategorija na grafiku: namjenski server, rješenje u oblaku, sa/bez podrške za ugnježdjeno računarstvo, drugi.

Opis rezultata:

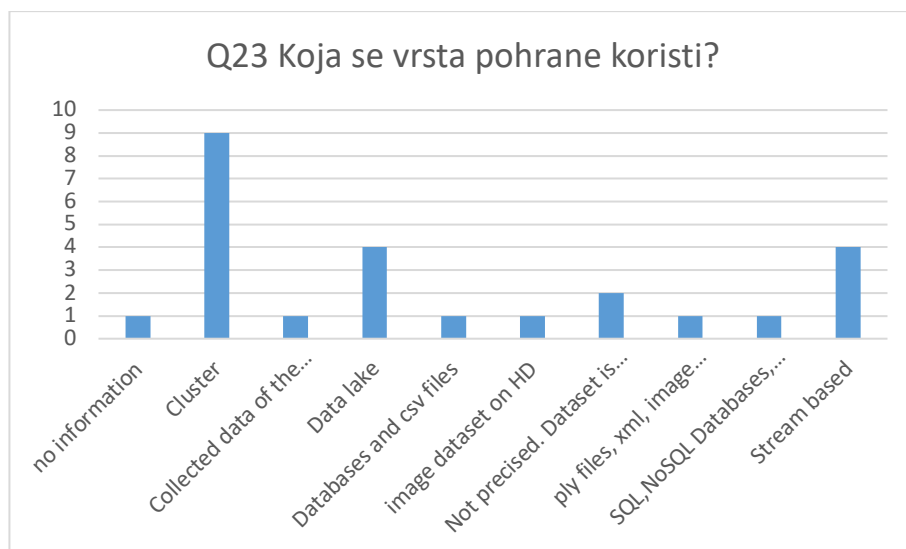
Rezultati pokazuju da je najčešće korištena vrsta AI platforme namjenski server. Namjenski serveri često se koriste kao tipovi AI platformi jer nude brojne prednosti koje ih čine prikladnima za pokretanje AI radnih opterećenja. Ti se serveri mogu prilagoditi, njihove performanse su dobre i često se smatraju sigurnijima od rješenja za dijeljeni hosting. Osim toga, namjenski serveri mogu se povećati ili smanjiti i ponuditi veću kontrolu nad okruženjem servera nego rješenja za dijeljeni hosting.

Diskusija:

Visoke performanse, prilagodljivost, sigurnost, skalabilnost i kontrola koje nude namjenski serveri čine ih popularnim izborom za pokretanje AI radnih opterećenja.

3.19. Koja se vrsta pohrane koristi?

Sljedeće pitanje daje detalje o vrsti pohrane. Predložene opcije su cluster, stream-based, data lake ili druge. Distribucija odgovora prikazana je na grafikonu.



Prevod kategorija na grafiku: nema informacije, klaster, prikupljeni podaci, Data lake, baze podataka i csv fajlovi, skup slika na hard disku, Nije precizirano, ply fajlovi xml slika, SQL NoSQL baze podataka, tok podataka.

Opis rezultata:

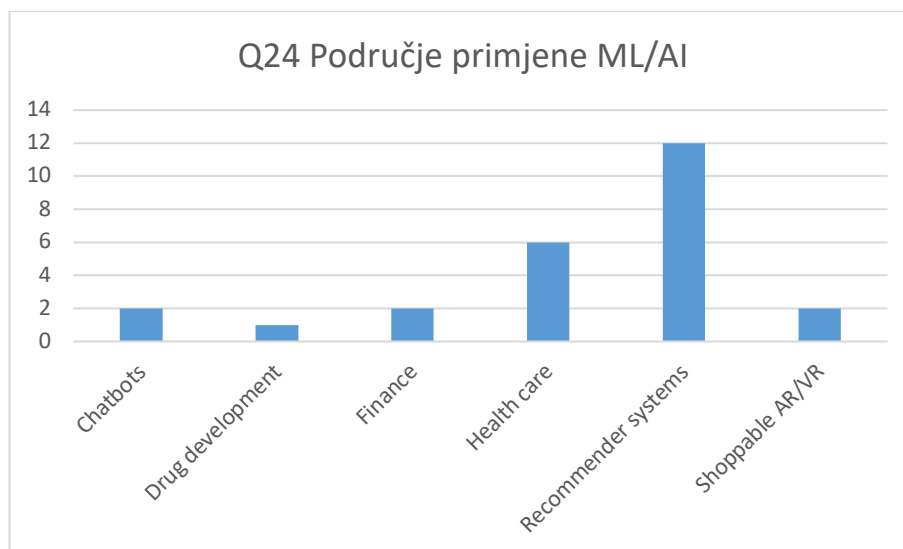
Rezultati pokazuju da je klaster najpopularnija vrsta pohrane među istraživačima. Moguće objašnjenje moglo bi biti da su klasteri dizajnirani za obezbjeđivanje visoke dostupnosti podataka. Klasteri su visoko skalabilni i otporni na greške. Klasteri mogu pružiti visok nivo performansi i mogu biti finansijski povoljno rješenje za pohranu, budući da se mogu izgraditi korišćenjem uobičajenog hardvera.

Diskusija:

Klasteri su popularni kao vrsta pohrane jer nude visoku dostupnost, skalabilnost, toleranciju na greške, performanse i isplativost. Ove ih prednosti čine idealnim izborom za kompanije i aplikacije koje zahtijevaju pouzdana i skalabilna rješenja za pohranu podataka.

3.20. Područje primjene ML/AI?

Sljedeće pitanje otkriva područja primjene mašinskog učenja/vještačke inteligencije. Pitanje predlaže osam opcija – sistemi preporuka, chatbotovi, A/B testovi, proširena stvarnost-AR/VR, zdravstvena njega, razvoj lijekova, finansije, sajber sigurnost. Rezultati su prikazani u grafikonu ispod.



Prevod kategorija na grafiku: chatbotovi, razvoj lijekova, finansije, zdravstvena njega, sistemi preporuka, proširena stvarnost-AR/VR.

Opis rezultata:

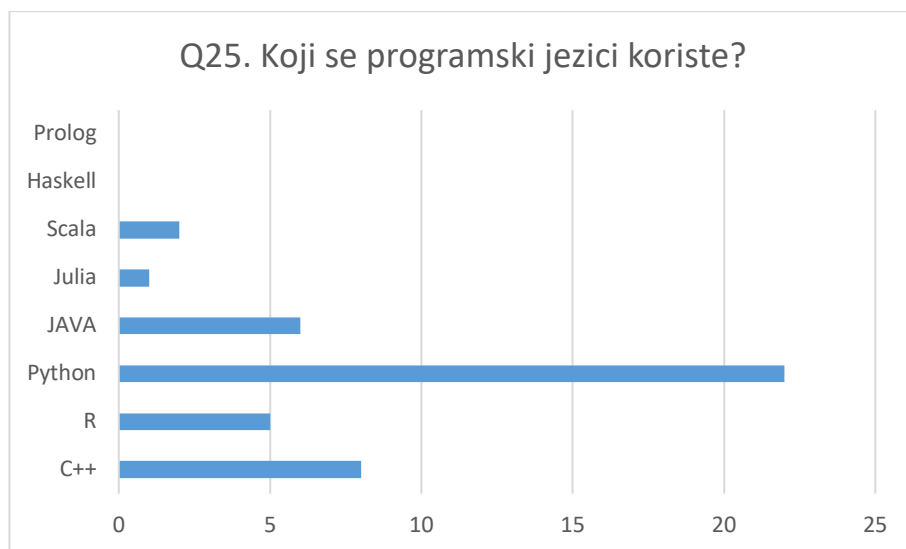
Podaci iz pregledanih projekata pokazuju da su najčešće područje u kojem se primjenjuju ML i AI sistemi preporuka. Sistemi preporuka dizajnirani su za pružanje personalizovanih preporuka korisnicima na temelju njihovih preferencija. Sistemi preporuka obično rade s velikim podacima, kao što su ponašanje korisnika i informacije o proizvodu. Mnogi sistemi za preporuku potrebni su za pružanje preporuka u realnom vremenu, kao što su veb-sajtovi za e-trgovinu ili platforme za streaming. Sistemi preporuka mogu pružiti značajne poslovne prednosti, kao što su povećana prodaja, angažman kupaca i lojalnost.

Diskusija:

Sistemi preporuka uobičajeno su područje primjene vještačke inteligencije jer zahtijevaju analizu složenih skupova podataka za pružanje personalizovanih preporuka u realnom vremenu. AI algoritmi mogu pomoći u automatizaciji i poboljšanju ovog procesa, što može dovesti do boljih poslovnih rezultata i poboljšanog korisničkog iskustva.

3.21. Koji se programski jezici koriste?

Ovo pitanje ukazuje na programske jezike koji se koriste u problemima vještačke inteligencije. Rezultati su prikazani u grafikonu.

**Opis rezultata:**

Rezultati jasno pokazuju da je najkorišteniji programski jezik Python. Python ima veliku i aktivnu zajednicu programera koji su stvorili brojne biblioteke i radne okvire za razvoj vještačke inteligencije. Python je fleksibilan jezik koji se može koristiti za širok raspon zadataka vještačke inteligencije, uključujući obradu podataka, mašinsko učenje i obradu prirodnih jezika. Osim toga, Python je kompatibilan sa širokim rasponom platformi i sistema, uključujući Windows, Mac i Linux. Iako Python nije najbrži programski jezik, dovoljno je brz za većinu AI zadataka.

Diskusija:

Python je popularan programski jezik za vještačku inteligenciju zbog svoje jednostavne upotrebe, velike zajednice, fleksibilnosti, kompatibilnosti i performansi. Ovi ga faktori čine idealnim izborom za razvoj i implementaciju AI aplikacija.

3.22. Koji je ekosistem korišćen?

Sljedeće pitanje daje informacije o ekosistemu koji je korišćen. Ovo su moguće alternative - Apache Hadoop, Anaconda, Kaggle, Colab, R studio, Matlab ili druge. Distribucija odgovora prikazana je u grafikonu ispod.



Prevod kategorija na grafiku: Apache Hadoop, Anaconda, Kaggle, Colab, R studio, Matlab, drugi.

Opis rezultata:

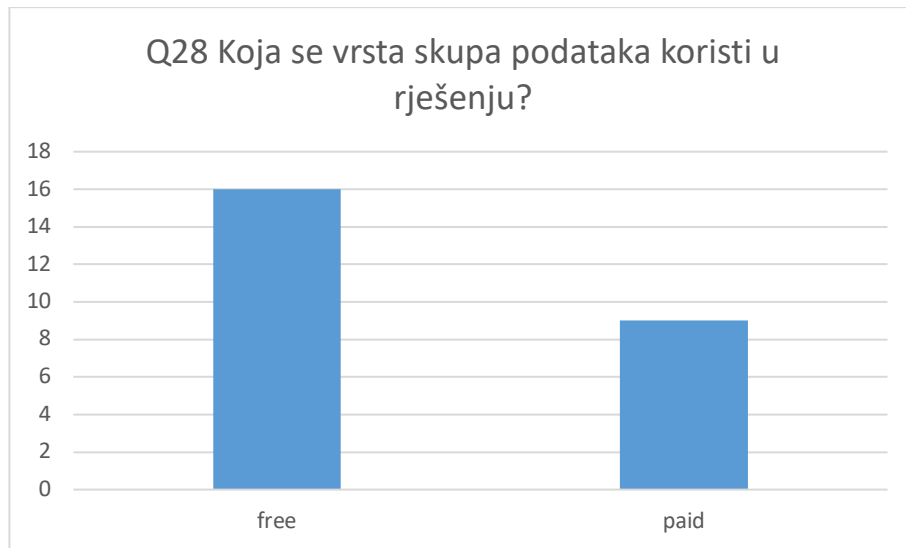
Rezultati pokazuju da se najviše koristio Apache Hadoop, s blagom predominacijom nad Anacondom i ostalim sistemima. Opcija "ostalo" uključuje ekosisteme kao što su NVIDIA CUDA, TensorFlow, lokalni Python IDE, Apache Spark, DataBricks.

Diskusija:

- Ekosistem Apache Hadoop naširoko se koristi za obradu i analizu velikih količina podataka zbog svoje skalabilnosti, distribuirane obrade, tolerancije na greške, prirode otvorenog koda, velikog ekosistema i prihvatanja u industriji. Ovi ga faktori čine idealnim rješenjem za organizacije koje treba da obrade i analiziraju velike količine podataka.
- Anaconda i drugi ekosistemi koji se uglavnom odnose na jezik Python, takođe su popularni. Anaconda je distribucija programskih jezika Python i R, zajedno sa kolekcijom biblioteka otvorenog koda, alata i radnih okvira za nauku o podacima i naučno računarstvo.

3.23. Koja se vrsta skupa podataka koristi u rješenju?

Sljedeće pitanje pruža informacije o vrsti skupa podataka – je li besplatan ili plaćen. Rezultati su prikazani u kružnom dijagramu.



Prevod kategorija na grafiku: besplatan, plaćen.

Opis rezultata:

Rezultati pokazuju prevladavanje besplatnih skupova podataka koji se koriste u rješenjima. Besplatni skupovi podataka dostupniji su širem krugu korisnika, uključujući studente, istraživače i programere. Mnogi besplatni skupovi podataka otvoreni su podaci, što znači da su besplatno dostupni svima za korištenje, izmjenu i distribuciju. Plaćeni skupovi podataka često dolaze s ograničenjima kako se mogu koristiti, što može ograničiti njihovu korisnost za određene vrste analiza ili aplikacija.

Diskusija:

Dostupnost besplatnih skupova podataka doprinijela je rastu i razvoju nauke o podacima i mašinskog učenja promovisanjem pristupačnosti, otvorenosti, saradnje i inovacija.

REFERENCE

1. <https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063>
2. <https://www.ibm.com/topics/convolutional-neural-networks>
3. <https://hub.packtpub.com/tensorflow-always-tops-machine-learning-artificial-intelligence-tool-surveys/>
4. <https://blog.ipleaders.in/permissive-license-copyleft-possible-distinctions/>

5. <https://www.techtarget.com/searchdatamanagement/definition/database>
6. <https://www.simplilearn.com/rise-of-nosql-and-why-it-should-matter-to-you-article>
7. <https://www.dataversity.net/cold-vs-hot-data-storage-whats-the-difference/>
8. <https://datalogistics.lt/en/dedicated-servers-are-an-increasingly-popular-hosting-service/>
9. <https://www.techtarget.com/searchstorage/magazineContent/The-benefits-of-clustered-storage>
10. Roy, D., Dutta, M. A systematic review and research perspective on recommender systems. J Big Data 9, 59 (2022). <https://doi.org/10.1186/s40537-022-00592-5>
11. <https://www.pulumi.com/why-is-python-so-popular/>
12. <https://www.projectpro.io/article/apache-hadoop-turns-10-the-rise-and-glory-of-hadoop/211>
13. <https://towardsdatascience.com/an-overview-of-the-anaconda-distribution-9479ff1859e6>
14. Sakshi Indolia, Anil Kumar Goswami, S.P. Mishra, Pooja Asopa, Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach, Procedia Computer Science, Volume 132, 2018, Pages 679-688, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.05.069>.