



Co-funded by
the European Union

FAAI: The Future is in Applied Artificial Intelligence
Budúcnosť je v aplikovanej umelej inteligencii Projekt
Erasmus+ 2022-1-PL01-KA220-HED-000088359

01.09.2022 – 31.08.2024

Výskum 7: Zhromažďovanie IT špecifikácií osvedčených postupov v AI: Analýza pre WP2





**Co-funded by
the European Union**

Výroba tohto dokumentu bola možná vďaka podpore projektu ERASMUS+: Budúcnosť je v aplikovanej umelej inteligencii (2022-1-PL01-KA220-HED-000088359)

Financované Európskou úniou. Vyjadrené názory a názory sú však len názormi autora (autorov) a nemusia nevyhnutne odrážať názory a názory Európskej únie alebo národnej agentúry (NA). Európska únia ani NA za ne nezodpovedajú.



Dátum

15.09.2021

Miesta vývoja výsledku

Univerzita Bielsko-Biala, Bielsko-Biala, Poľsko

Univerzita knižničných štúdií a informačných technológií, Sofia, Bulharsko

Univerzita v Niši, Srbsko

Univerzita sv. Cyrila a Metoda v Trnave na Slovensku

Univerzita v Čiernej Hore, Čierna Hora

Zhrnutie: 10 riadkov

Kľúčové slová: menej ako 5 kľúčových slov

1. Úvod

30 riadkov

2. Zber a analýza údajov

Údaje získali vedci z piatich partnerských inštitúcií. V aktuálnom výskume bolo získaných 8 dotazníkov z UBB, 6 dotazníkov z ULSIT, 5 z UNi, 5 z UoM a 1 z USCM. Spolu 11 výskumníkov zozbieralo 25 dotazníkov.

3. Výsledky

3.1. Krajina, v ktorej prebieha školenie

Prvá otázka prieskumu, po údajoch pre názov organizácie a výskumníka, sa pýta, kde prebieha školenie. Výsledky sú uvedené nižšie.



Popis údajov:

Podľa údajov z výskumu sa školenia väčšinou konajú v Srbsku, po ktorom nasleduje Bulharsko. Jeden respondent odpovedal, že na školenia využívajú len Courseru alebo O'Reilly. V ostatných krajinách prieskumu prebieha jedno až tri školenia.

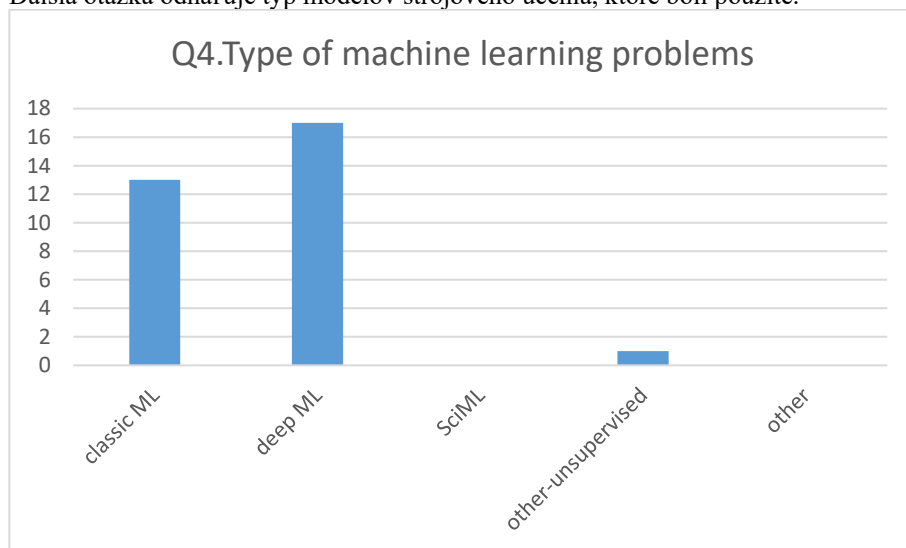
Diskusia:

Prieskum ukazuje, že je potrebné viac školení pre IT špecialistov, aby bolo možné pozorovať viac osvedčených postupov v oblasti AI. Hlavné závery:

- V krajinách EÚ je potrebných viac školení.
- Najviac školení prebieha v Srbsku a Bulharsku.

3.2. Typ problémov strojového učenia

Ďalšia otázka odhaľuje typ modelov strojového učenia, ktoré boli použité.



Popis údajov:

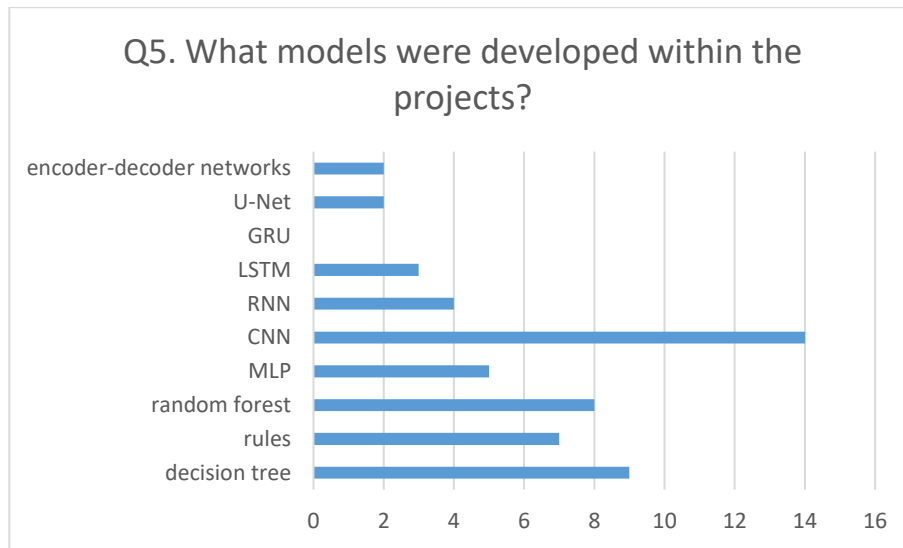
Dáta ukazujú, že najpoužívanejším modelom strojového učenia je hlboké ML, po ktorom nasleduje klasické ML. V posledných rokoch sa hlboké učenie stalo hlavným nástrojom pri riešení rôznych problémov AI. Preukázal lepší výkon pri rôznych úlohách vrátane spracovania prirodzeného jazyka, vízie, virtuálnych asistentov, chatbotov, zdravotnej starostlivosti atď.

Diskusia:

- Najpoužívanejším modelom na základe prieskumu je hlboké ML.
- SciML alebo iné modely strojového učenia nie sú medzi kontrolovanými projektmi bežné.

3.3. Aké modely boli vyvinuté (študované) v rámci projektov?

Nasledujúca otázka sa pýta, ktoré modely boli konkrétne študované. Výsledky sú uvedené v tabuľke nižšie.

**Popis údajov:**

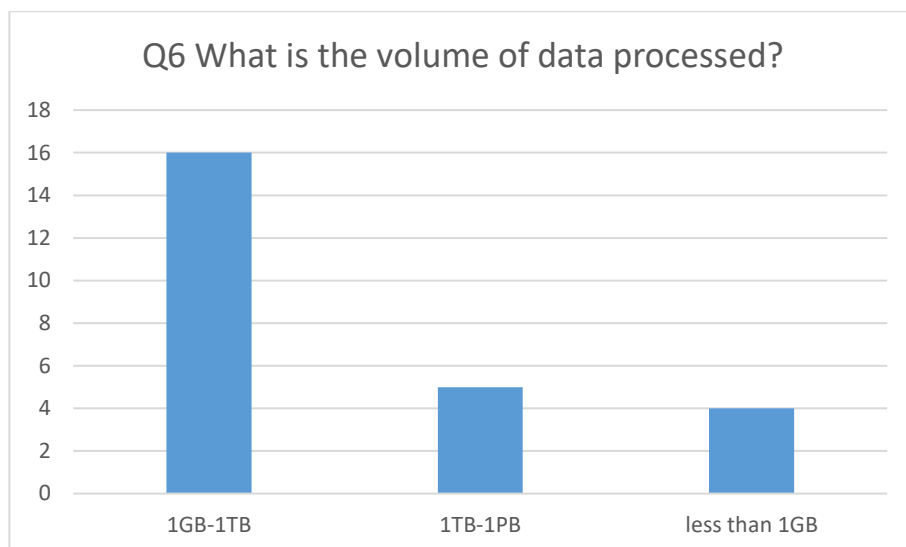
Podľa výsledkov boli v projektoch najčastejšie používané modely konvolučných neurónových sietí. Je to pochopiteľné, pretože tieto modely majú rôzne aplikácie – klasifikácia obrazu, detekcia objektov, rozpoznávanie tváre, analýza medicínskych snímok, NLP, autonómne vozidlá alebo technológie. CNN sú obzvlášť vhodné pre úlohy rozpoznávania obrázkov, pretože sa dokážu automaticky naučiť detegovať zložité prvky, ako sú hrany, rohy a textúry. Rozhodovací strom spolu s náhodným lesom sú ďalšie modely, ktoré boli vyvinuté pre rôzne problémy AI. Mnoho aplikácií rozhodovacích stromov a náhodných lesov sa široko používa v odvetviach, ako sú financie, zdravotníctvo, výroba a monitorovanie životného prostredia. Rozhodovacie stromy a náhodné lesy sú populárne algoritmy strojového učenia, pretože sa ľahko používajú, chápu a interpretujú, dokážu spracovať numerické aj kategorické údaje a dokážu spracovať veľké a zložité súbory údajov.

Diskusia:

- Konvolučné neurónové siete sú najbežnejšími modelmi v projektoch.
- GRU nebola vyvinutá v rámci projektov.

3.4. Aké modely boli vyvinuté (študované) v rámci projektov?

Ďalšia otázka poukazuje na objem údajov, ktoré boli spracované. Možné odpovede boli menej ako 1 GB, 1 GB – 1 TB, 1 TB – 1 PB a viac ako 1 PB.

**Popis údajov:**

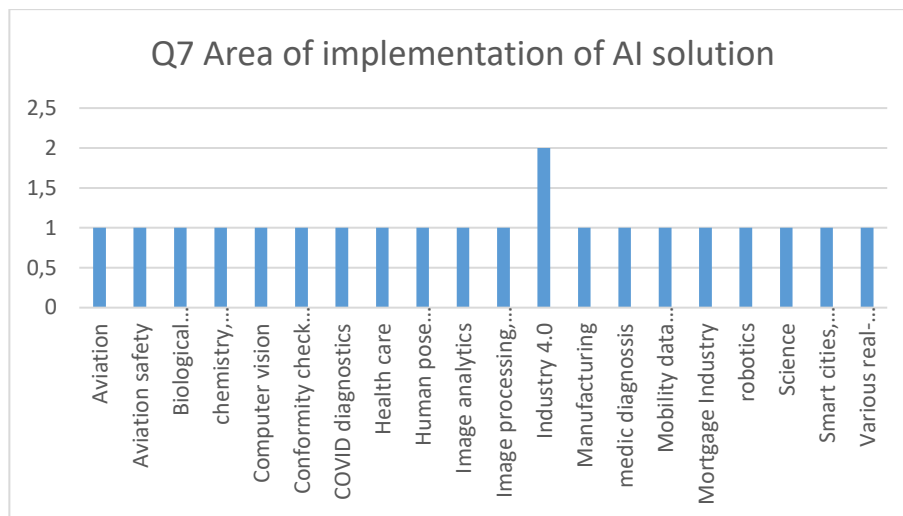
Ako vidno z grafu, objem spracovaných dát sa pohyboval väčšinou medzi 1 GB a 1 TB. Tento objem predstavuje relatívne veľké množstvo dát, s ktorými sa bežne stretávame v mnohých rôznych typoch projektov, no stále sú zvládnuteľné z hľadiska úložných a spracovateľských schopností najmodernejších počítačov. Objem spracovaných dát väčší ako 1 TB nie je až taký bežný, pravdepodobne kvôli dostupnosti a cene. Správa a spracovanie údajov v rozsahu petabajtov si vyžaduje špecializovanú infraštruktúru a zdroje, ktoré nie sú pre väčšinu organizácií ľahko dostupné. Ukladanie a spracovanie veľkého množstva dát môže byť drahé, a to z hľadiska hardvéru aj prevádzkových nákladov. Na druhej strane objem menší ako 1 GB nie je bežný ani preto, že sa považuje za dostatočne malý na to, aby sa dal efektívne spracovať na jednom stroji bez potreby špecializovaného hardvéru alebo distribuovaných výpočtových systémov.

Diskusia:

- Najviac sa využíva objem spracovaných dát medzi 1 GB a 1 TB.
- Objem väčší ako 1 TB alebo menší ako 1 GB nie je takto použiteľný.

3.5. Aká je oblasť implementácie riešenia AI?

Ďalšia otázka smeruje do oblastí implementácie riešení AI. Odpovede zahŕňajú vlastný vstup respondentov na základe každého prípadu. Výsledky sú uvedené graficky nižšie.

**Popis údajov:**

Treba poznamenať, že výskumníci hľadali projekty v rôznych oblastiach, ktorými sú: poľnohospodárstvo, AI v medicíne, chirurgia, riadenie leteckej dopravy, letectvo, bezpečnosť letectva, analýza biologickej sekvencie, chémia, robotika, zdravie, počítačové videnie, zhoda kontrola v leteckom priemysle, diagnostika COVID, zdravotná starostlivosť, odhad ľudskej polohy, analýza obrazu, spracovanie obrazu, predikcia cien, priemysel 4.0, výroba, lekárska diagnostika, veda a analýza údajov o mobilite, hypotekárny priemysel, veda, inteligentné mestá, monitorovanie dopravy, rôzne prípady z reálneho sveta. Každý respondent skúmal jednu alebo dve z týchto oblastí implementácie riešení AI.

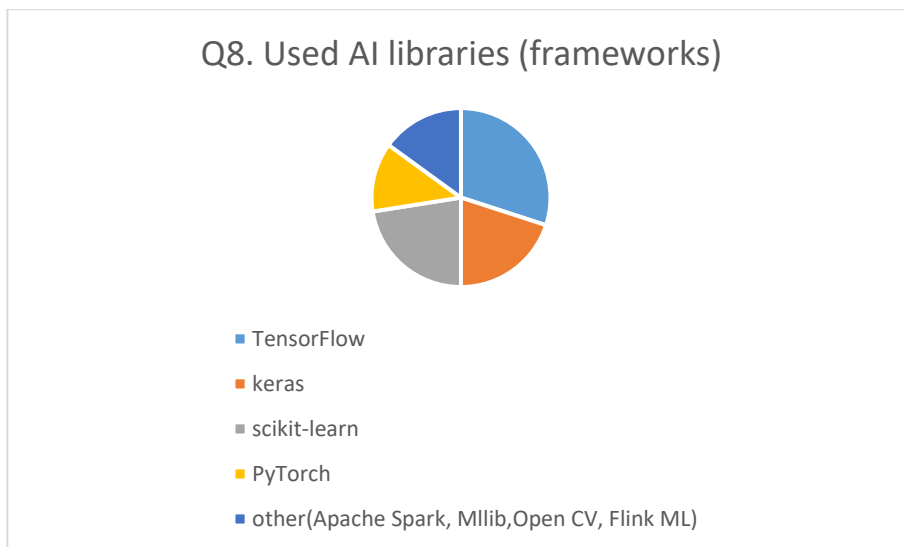
Diskusia:

- Každý respondent mal za úlohu preskúmať konkrétnu oblasť, aby bolo možné preskúmať čo najviac oblastí.

3.6. Použité knižnice AI (frameworks).

Otázka 8 navrhuje rôzne knižnice AI, ktoré sa zvyčajne používajú v oblasti umelej inteligencie. Sú zobrazené v koláčovom grafe nižšie.

Q8. Used AI libraries (frameworks)



Popis údajov:

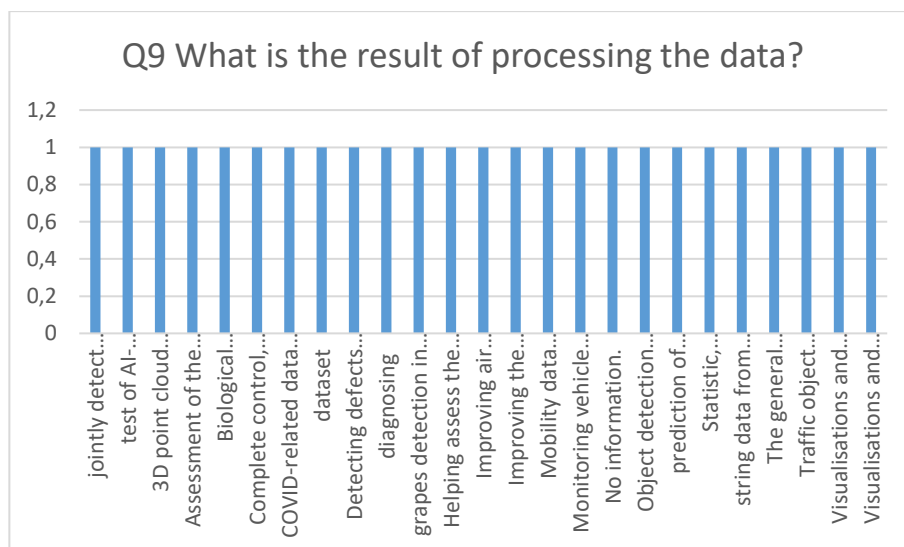
Výsledky naznačujú, že najpoužívanejšou knižnicou AI je TensorFlow. TensorFlow je jednou z najpopulárnejších a najpoužívanejších knižníc s otvoreným zdrojovým kódom na vytváranie a nasadzovanie modelov strojového učenia a hlbokého učenia. K popularite knižnice vďaka aj jej početnej a aktívnej komunite, ktorá ju podporuje a prispieva k jej rozvoju. Medzi ďalšie populárne knižnice AI patria Keras a Scikit-learn. Ďalšie knižnice používané pri problémoch s AI zahŕňajú PyTorch, Apache Spark, MLib, Open CV a Flink ML.

Diskusia:

- Výber knižnice na použitie závisí od rôznych faktorov vrátane špecifických potrieb projektu, osobných preferencií a úrovne odbornosti vývojára.

3.7. Aký je výsledok spracovania údajov?

Táto otázka zahŕňa otvorené odpovede respondentov v závislosti od každého prípadu. Preto bolo pre každý projekt dvadsaťpäť rôznych výsledkov.



Popis údajov:

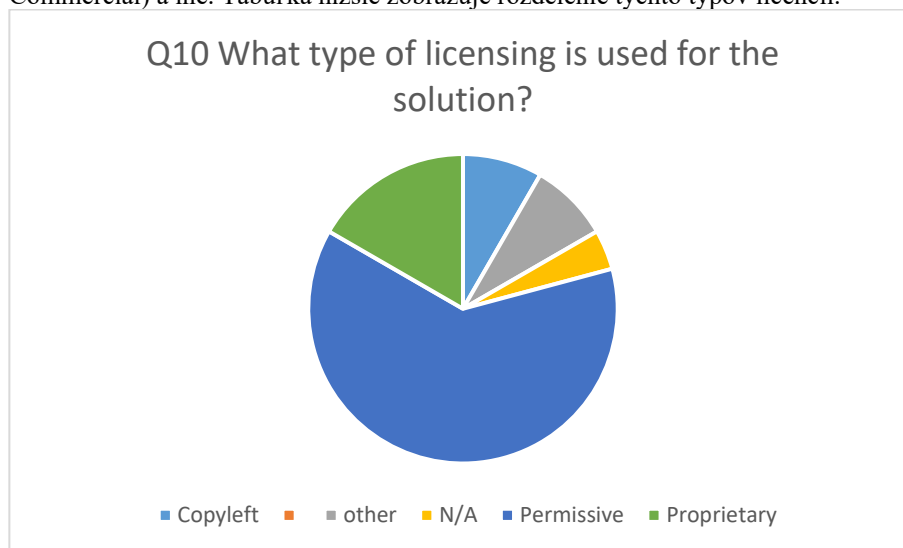
V prvom projekte bola výsledkom spoločná detekcia kľúčových bodov ľudského tela, tváre a chodidiel. V druhom projekte bol výsledkom test metód AI-TWILIGHT v automobilovom priemysle, záhradníctve a pouličnom osvetlení. Tretím výsledkom bola 3D klasifikácia a segmentácia mračna bodov. Štvrtým výsledkom bolo hodnotenie pravdepodobnosti a závažnosti infekcie Covid-19. Piatym výsledkom bola klasifikácia biologickej sekvencie. Šiestym výsledkom bola úplná kontrola, plná artikulácia a inteligentná spätná väzba. Siedmym výsledkom bola analýza údajov súvisiaca s COVID. Ôsmym výsledkom bolo vytvorenie súboru údajov. Deviaty výsledok bol zameraný na zisťovanie chýb na výrobkoch vo výrobe. Desiaty výsledok bol zameraný na implementáciu do diagnostiky. Jedenástym výsledkom bola detekcia hrozna vo vinohrade. Dvanástym výsledkom bola pomoc pri hodnotení stavu strojov a celkového vybavenia závodu. Trinástym výsledkom bolo zlepšenie riadenia letovej prevádzky prostredníctvom spolupráce strojového učenia na súkromných súboroch údajov. Štrnástym výsledkom bolo zlepšenie rozhodovacieho procesu v scenároch obletov, čo má veľký význam pre bezpečnosť leteckých spoločností aj poskytovateľov leteckých navigačných služieb v ATM. Pätnástym výsledkom bola analýza údajov o mobilite. Šestnástym výsledkom bolo monitorovanie vozidla, ktoré premáva po mestách, počítaním áut zo snímkov získaných z inteligentných kamier. Sedemnástym výsledkom bola detekcia a rozpoznávanie objektov. Osemnástym výsledkom bola predikcia aktivít chemických zlúčenín, automatické riadenie robotov, analýza ľudských pohybov z nositeľných senzorov. Devätnástym výsledkom bola štatistika, predpisovanie liekov a vymenovanie liečby. Dvadsiatym výsledkom bolo získavanie údajov z obrázkov a objekt z obrázkov. Dvadsiaty prvý výsledok bol pokus o vývoj prototypu globálneho systému monitorovania viacerých rizík a systému včasného varovania. Dvadsiatym druhým výsledkom bola detekcia a rozpoznávanie dopravných objektov. Dvadsiaty tretí výsledok prezentoval vizualizácie a informácie o budúcich cenách. Dvadsiaty štvrtý výsledok predstavoval vizualizácie a modely ML. Bol jeden projekt, v ktorom neboli zdokumentované žiadne výsledky.

Diskusia:

- Každý respondent mal po spracovaní údajov nájsť výsledky v závislosti od každého projektu. Týmto spôsobom boli objavené rôzne výsledky v doménach AI.

3.8. Aký typ licencovania sa používa na riešenie?

Nasledujúca otázka sa pýta, aký typ licencovania je v riešení použitý a zahŕňa štyri alternatívy – permissívne (BSD, MIT), copyleft (GPL, LGPL), proprietárne (Bespoke, Commercial) a iné. Tabuľka nižšie zobrazuje rozdelenie týchto typov licencií.

**Popis údajov:**

Výsledky ukazujú, že najpoužívanejším typom licencovania je permissívny (v 15 projektoch), nasledovaný proprietárnym (v 4 projektoch), copyleft (v 2 projektoch). V jednom projekte bola použitá voľne dostupná originálna licencia OpenPose na nekomerčné použitie. Permissívne licencie sú obľúbené, pretože používateľom a vývojárom ponúkajú vysoký stupeň slobody. Tieto typy licencií zvyčajne umožňujú používateľom upravovať a redistribuovať softvér bez toho, aby sa vyžadovalo, aby boli akékoľvek zmeny alebo vylepšenia vydané v rámci rovnakej licencie. Toto je v protiklade s copyleft licenciami, ktoré vyžadujú, aby všetky odvodené diela boli licencované za rovnakých podmienok ako originál. Permissívne licencie si často vyberajú jednotlivci a organizácie, ktoré chcú podporiť spoluprácu a inovácie, pričom zároveň umožňujú maximálnu flexibilitu v spôsobe používania a distribúcie softvéru. Vlastné licencovanie je na druhom mieste, hoci sa vyskytuje len v štyroch projektoch. Tento typ licencovania je bežný v softvérovom priemysle, ale často obmedzuje spôsoby, akými môžu používatelia softvér používať a distribuovať. Niektoré spoločnosti sa stále rozhodujú používať proprietárne licenčné modely ako spôsob ochrany svojho duševného vlastníctva a udržiavania kontroly nad svojim softvérom.

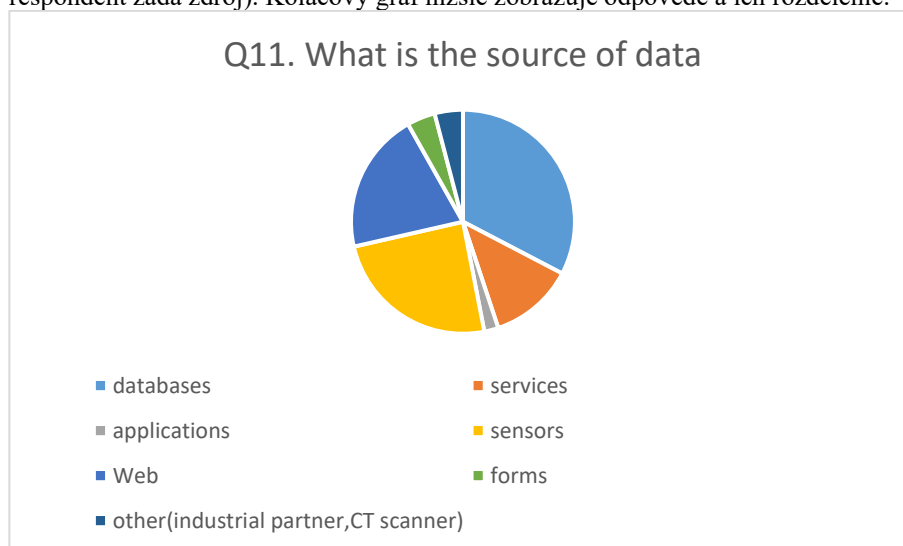
Copyleftové licencie nie sú také bežné ako permissívne alebo proprietárne licencie, pretože ukladajú viac obmedzení na spôsob použitia a distribúcie softvéru.

Diskusia:

- Permissívne licencie sú obľúbené, pretože ponúkajú rovnováhu slobody a flexibility, ktorá umožňuje spoluprácu a inovácie a zároveň minimalizuje prekážky vstupu a prijatia.
- Copyleftové licencie sú menej bežné ako permissívne alebo proprietárne licencie, pretože ukladajú viac obmedzení, čo nemusí byť žiaduce pre všetkých používateľov a vývojárov.

3.9 Aký je zdroj údajov?

Otázka 11 obsahuje odpovede týkajúce sa zdroja údajov. Možnosti sa líšia od databáz, služieb, aplikácií, senzorov, webu, formulárov alebo iných (ak je odpoveď iná, respondent zadá zdroj). Koláčový graf nižšie zobrazuje odpovede a ich rozdelenie.

**Popis údajov:**

Výsledky ukazujú, že základným zdrojom údajov sú databázy (v 16 správach), nasledujú senzory (v 12 správach) a web (v 10 správach). Najmenej častým zdrojom sú formuláre (v 2 hláseniach), iné, napr. priemyselný partner alebo CT skener (v 2 hláseniach) a žiadosti (v 1 hlásení). Databázy sú preferovanými zdrojmi, pretože údaje v nich sú štruktúrované a dajú sa efektívne získať. Okrem toho sú údaje konzistentné a presné a databázy sa dajú ľahko integrovať s inými systémami, čo uľahčuje zdieľanie informácií medzi rôznymi aplikáciami a platformami. Na druhej strane žiadosti a formuláre nie sú bežným zdrojom informácií, pretože údaje sú často rozptýlené a ich kvalita je nejednotná. Aplikácie alebo formuláre nemusia byť navrhnuté na spracovanie veľkého množstva údajov a prístup môže byť obmedzený z dôvodu bezpečnosti alebo ochrany súkromia. Okrem toho môžu formuláre používať

rôzne formáty údajov a protokoly, čo môže sťažiť integráciu údajov medzi rôznymi systémami a aplikáciami.

Diskusia:

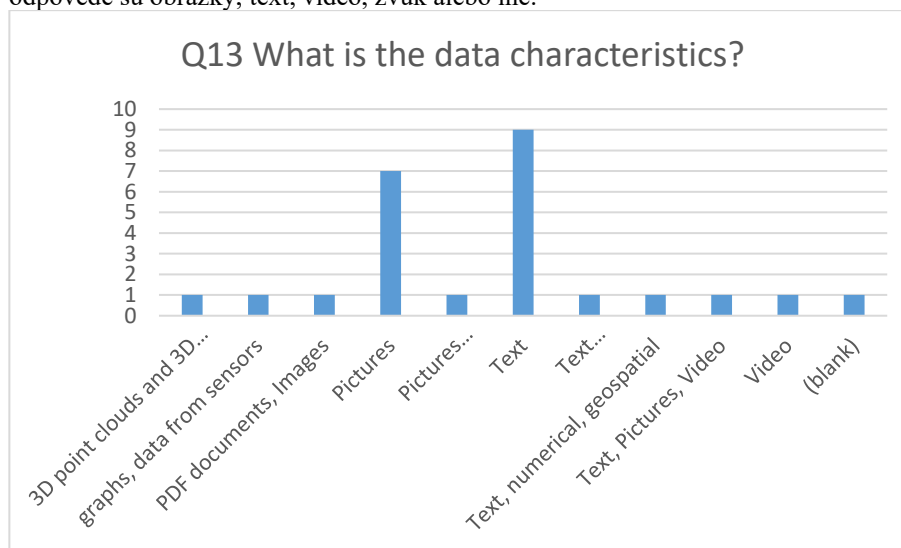
- Databázy sú preferovaným zdrojom údajov, pretože ponúkajú spoľahlivý, efektívny a bezpečný spôsob ukladania a správy veľkého množstva údajov konzistentným a štruktúrovaným spôsobom.
- Aplikácie a formuláre môžu byť zdrojmi údajov, ale ich extrahovanie a efektívne využitie si často vyžadujú značné úsilie a zdroje.

3.10. Reprezentácia údajov.

Táto otázka poskytuje informácie o type údajov. V projektoch sú zahrnuté nasledujúce typy - grafy, dáta zo senzorov, text, obrázky, csv, avro, parkety, JSON API, MySQL server, FTP server, 3D bodový cloud, CT hrudníka, medicínsky výskum.

3.11. Aké sú charakteristiky údajov?

Nasledujúca otázka poskytuje informácie o charakteristikách údajov. Možné odpovede sú obrázky, text, video, zvuk alebo iné.



Popis údajov:

Ako je možné vidieť v grafe, hlavnými charakteristikami údajov sú texty a obrázky alebo obrázky s textom alebo videom. Obrázky a texty sú ľahko pochopiteľné a interpretovateľné. Môžu byť generované vo veľkých objemoch a môžu byť použité v rôznych kontextoch, od marketingu a reklamy až po vedecký výskum a analýzu údajov. Obrázky a texty môžu byť ľahko dostupné a zdieľané na rôznych platformách.

Diskusia:

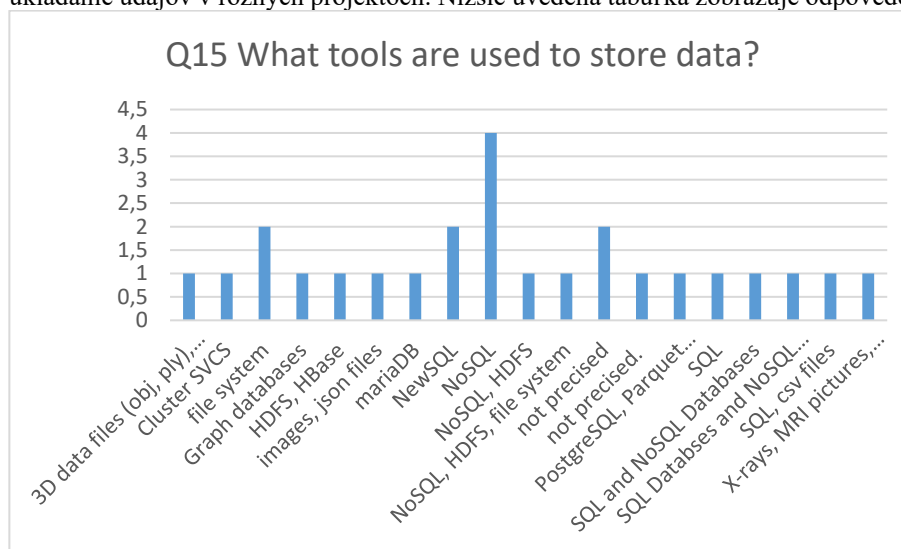
- Obrázky a texty sú bežné dátové charakteristiky, pretože sú ľahko zrozumiteľné, generujú sa vo veľkých objemoch, sú všestranné a prístupné. Pokrok v technológii tiež uľahčil analýzu a extrahovanie poznatkov z neštruktúrovaných údajov, vďaka čomu sú cenným zdrojom informácií pre podniky aj výskumníkov.

3.12. Spracovanie a kvalita údajov.

Otázka 14 poskytuje informácie o kvalite údajov a spôsobe spracovania údajov. V závislosti od prípadu sa použili rôzne metódy. V niektorých projektoch sa odborné rozhodnutia prijímali na základe lekárskej expertízy. V iných projektoch boli súbory prístupné a formátované cez Python. V niektorých projektoch boli údaje vyčistené a vizualizované. V iných projektoch bola použitá klasifikácia a regresia. V niektorých prípadoch sa použilo škálovanie, označovanie, spracovanie zvuku alebo videa. Zistilo sa, že tieto techniky sú najbežnejšie pri spracovaní údajov.

3.13. Aké nástroje sa používajú na ukladanie údajov?

Táto otázka je otvorená a poskytuje informácie o nástrojoch, ktoré sa používajú na ukladanie údajov v rôznych projektoch. Nižšie uvedená tabuľka zobrazuje odpovede.



Popis údajov:

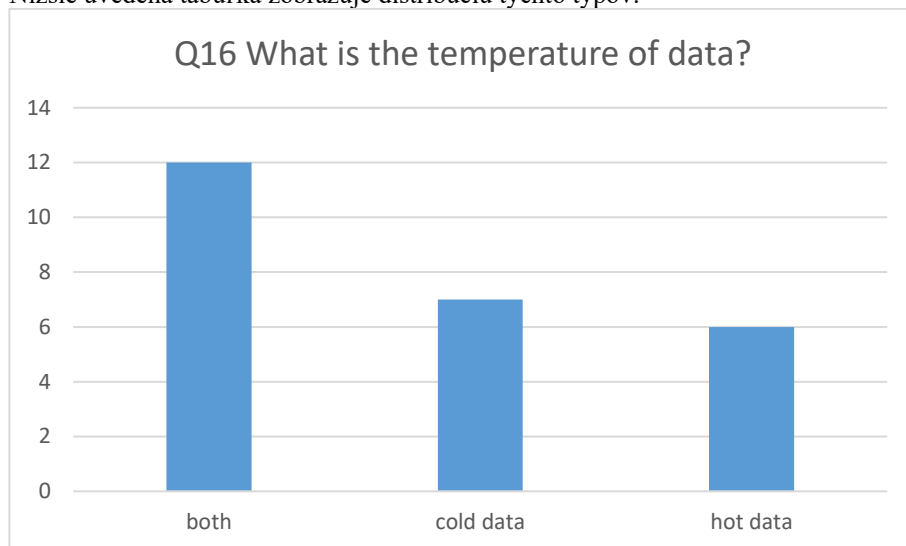
Najbežnejšími nástrojmi sú databázy NoSQL. Tieto databázy sú navrhnuté tak, aby spracovávali neštruktúrované alebo pološtruktúrované údaje, čo z nich robí lepšiu voľbu ako tradičné relačné databázy pre aplikácie, ktoré vyžadujú flexibilné modelovanie údajov. To umožňuje vývojárom ukladať a získavať dáta spôsobom, ktorý lepšie vyhovuje potrebám ich aplikácie. Mnohé databázy NoSQL sú open source, čo z nich robí cenovo výhodnú možnosť pre vývojárov a organizácie. Ostatné nástroje sú takmer rovnomerne rozdelené, takže možno konštatovať, že nástroj, ktorý sa použije, závisí od prípadu a problému IT.

Diskusia:

- Celkovo NoSQL databázy ponúkajú niekoľko výhod oproti tradičným relačným databázam, a preto sú čoraz bežnejšie ako nástroje na ukladanie údajov. Ponúkajú väčšiu flexibilitu, škálovateľnosť, dostupnosť a výkon, čo z nich robí ideálnu voľbu pre moderné aplikácie, ktoré tieto funkcie vyžadujú.

3.14. Aká je teplota dát?

Nasledujúce otázky odhalia, či sú bežnejšie údaje o chlade, horúčave alebo oboch. Nižšie uvedená tabuľka zobrazuje distribúciu týchto typov.

**Popis údajov:**

Je možné jasne pozorovať, že studené aj horúce údaje. Ukladanie studených údajov je zvyčajne lacnejšie, pretože sa často ukladajú na pomalších a lacnejších úložných zariadeniach, ako sú pásy alebo pevné disky. Vďaka tomu je ideálnym riešením na ukladanie historických údajov, záloh a archívov, ku ktorým možno netreba často pristupovať, ale je potrebné ich uchovávať z dôvodu dodržiavania predpisov alebo právnych dôvodov. Na druhej strane horúce dáta sa zvyčajne ukladajú na rýchlejšie a drahšie úložné zariadenia. K tomuto typu údajov často pristupujú aplikácie, používatelia alebo služby a musia byť rýchlo dostupné, aby mohli podporovať operácie, transakcie alebo analýzy v reálnom čase. Vo väčšine organizácií sú pre obchodné operácie potrebné horúce aj studené údaje.

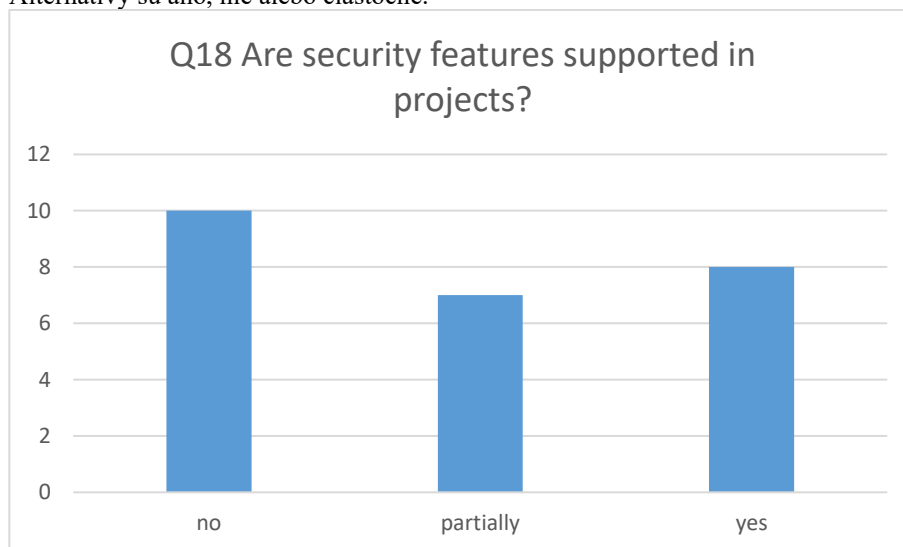
Diskusia:

- Studené aj horúce údaje sú spoločné rovnako, pretože slúžia na rôzne účely a sú potrebné v rôznych časoch. Studené údaje sa týkajú údajov, ku ktorým

sa pristupuje zriedkavo alebo vôbec, zatiaľ čo horúce údaje sa týkajú údajov, ku ktorým sa často pristupuje alebo sa aktívne používajú.

3.15. Sú v projektoch podporované bezpečnostné prvky?

Ďalšia otázka sa pýta, či sú v projektoch podporované bezpečnostné prvky. Alternatívy sú áno, nie alebo čiastočne.



Popis údajov:

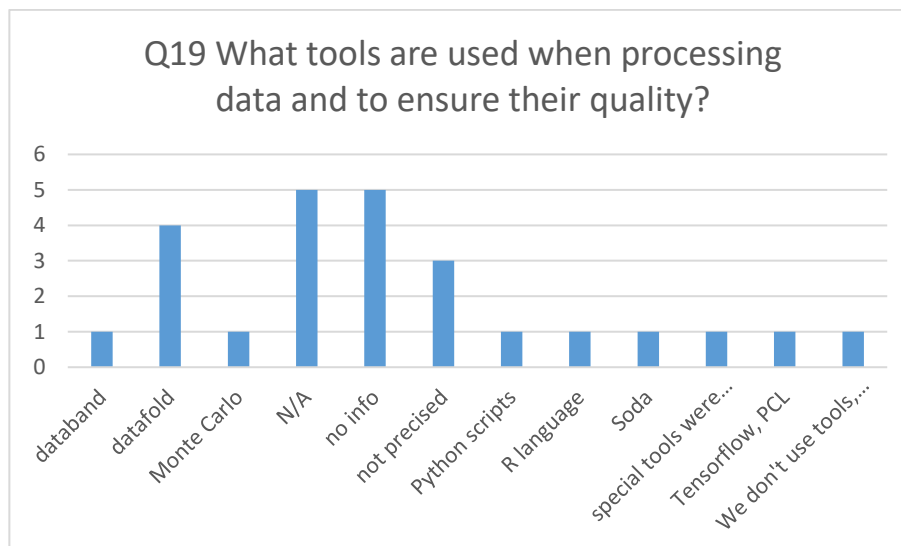
Výsledky naznačujú, že bezpečnostné prvky nie sú podporované o niečo častejšie ako naopak. Možným vysvetlením by mohlo byť, že implementácia robustných bezpečnostných funkcií môže vyžadovať čas a zdroje a môže vyžadovať značné investície do hardvéru, softvéru a personálu. V niektorých prípadoch môžu organizácie uprednostniť iné funkcie alebo funkcie pred bezpečnosťou z dôvodu rozpočtových alebo časových obmedzení. Ďalším pravdepodobným dôvodom môže byť nedostatok odborných znalostí, pretože bezpečnosť môže byť zložitá a špecializovaná oblasť a nie všetci vývojári alebo organizácie môžu mať zdroje alebo znalosti na správne riešenie bezpečnostných problémov. Rozdiely sú však malé, takže nemožno zovšeobecňovať, že bezpečnostné prvky nie sú podporované, iba v niektorých prípadoch.

Diskusia:

- Funkcie zabezpečenia nie sú vo väčšine projektov podporované.
- Pre organizácie a vývojárov je dôležité, aby uprednostňovali bezpečnosť a podnikli kroky na implementáciu robustných bezpečnostných opatrení na ochranu svojich používateľov a ich údajov.

3.16. Aké nástroje sa používajú pri spracovaní údajov a na zabezpečenie ich kvality?

Tieto otázky naznačujú rôzne možnosti týkajúce sa nástrojov pri spracovaní údajov. Možnosti sú Talend, Toro, Soda, Datafold, Databand, Precisely, Monte Carlo alebo iné. Rozdelenie výsledkov je uvedené v tabuľke nižšie.



Popis údajov:

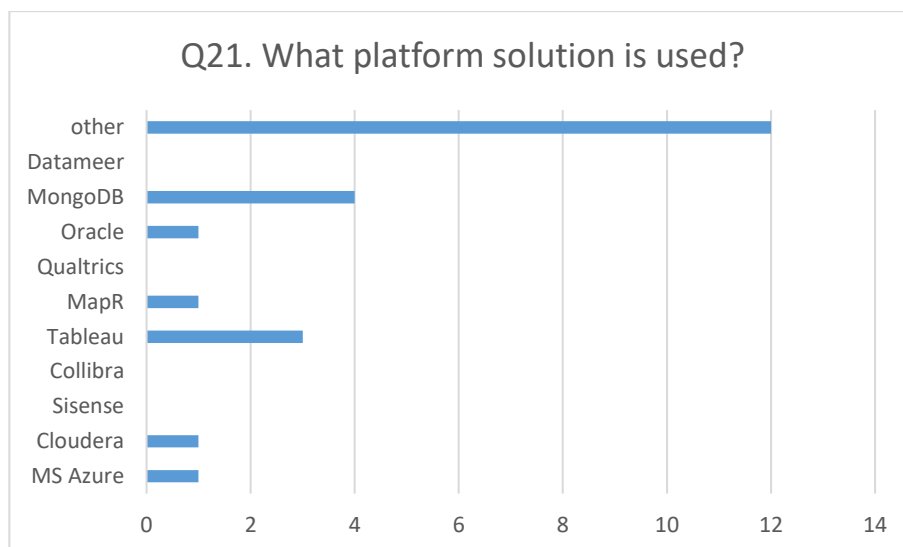
Výsledky sú prevažne „N/A“ alebo „žiadne informácie“, čo naznačuje, že v analyzovaných štúdiách vedci nespomenuli nástroje, ktoré použili pri spracovaní údajov. Datafold je na treťom mieste. Je čoraz viac používaný a považuje sa za cenný nástroj pre pracovné toky spracovania údajov.

Diskusia:

- O nástrojoch, ktoré sa používajú pri spracovaní údajov, sa často nehovorí.
- Datafold je relatívne nový nástroj na spracovanie údajov, ktorý si získal obľubu medzi dátovými inžiniermi a dátovými vedcami.

3.17. Aké platformové riešenie sa používa?

Ďalšia otázka prezrádza, aké platformové riešenie sa používa. Možné odpovede sú MS Azure, Cloudera, Sisense, Collibra, Tableau, MapR, Qualtrics, Oracle, MongoDB, Datameer alebo iné, ktoré zadá respondent. Výsledky sú uvedené v tabuľke.

**Popis údajov:**

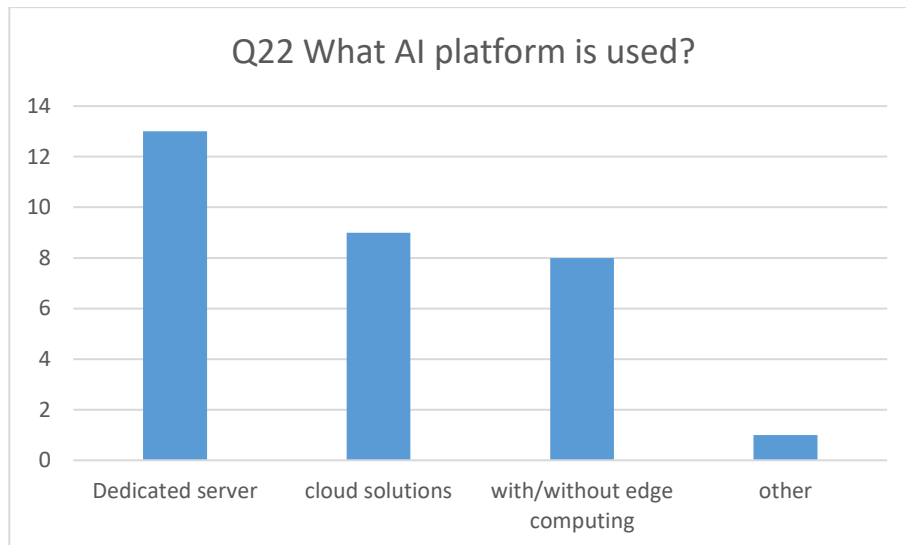
Platformové riešenia používané v projektoch sú prevažne iné. Zahŕňajú R studio, NVIDIA CUDA, Clara, CUDA-X, TensorFlow/TensorRT, Anaconda, Google Colab, Apache Spark, Apache Flink, PySpark, AWS, Bonseyes, špecializovaný rádiologický informačný systém, Amazon Web Services S3 Data Lake, FedML, OpenMMLab , gRPC.

Diskusia:

- Platformové riešenia, ktoré sú navrhované ako odpovede, nie sú také populárne. MongoDB a Tableau medzi nimi zaujímajú druhú a tretiu pozíciu.
- Vo viacerých prípadoch sa využívajú iné platformy ako Anaconda, Apache, NVIDIA, R studio, TensorFlow atď.

3.18. Aký typ platformy AI sa používa (napr. serverové, cloudové riešenia, s/bez podpory edge computingu alebo iné)?

Nasledujúca otázka smeruje k typu platformy AI. Možnými alternatívami sú dedikovaný server, cloudové riešenia s/bez podpory edge computingu alebo iné typy, ktoré zadali respondenti. Výsledky sú uvedené v tabuľke.

**Popis údajov:**

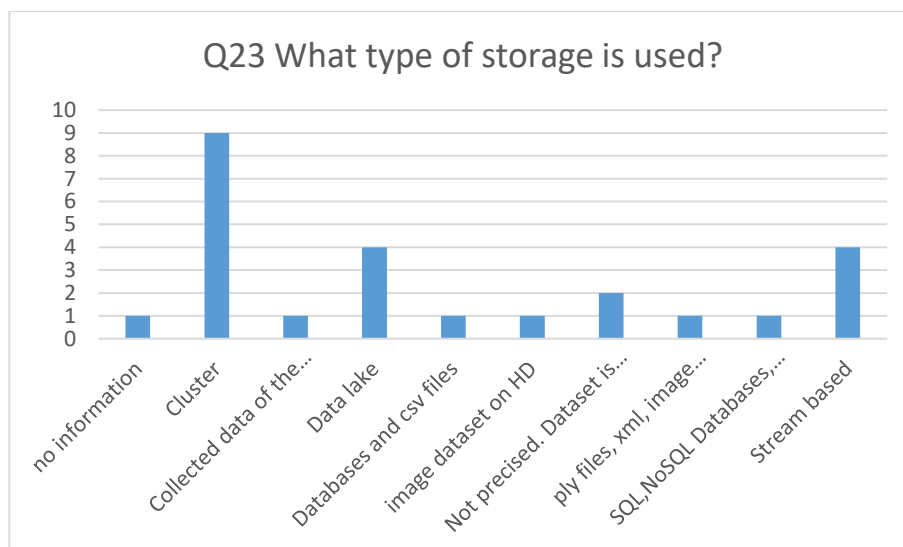
Výsledky naznačujú, že najpoužívanejším typom platformy AI je dedikovaný server. Dedikované servery sa často používajú ako typy platform AI, pretože ponúkajú množstvo výhod, vďaka ktorým sú vhodné na spúšťanie úloh AI. Tieto servery je možné prispôsobiť, ich výkon je vysoký a často sa považujú za bezpečnejšie ako riešenia zdieľaného hostingu. Okrem toho je možné dedikované servery zväčšiť alebo zmenšiť a ponúkajú väčšiu kontrolu nad serverovým prostredím ako riešenia zdieľaného hostingu.

Diskusia:

- Vysoký výkon, prispôbitelnosť, bezpečnosť, škálovateľnosť a ovládanie, ktoré ponúkajú dedikované servery, z nich robí obľúbenú voľbu pre spustenie pracovných záťaží AI.

3.19. Aký typ úložiska sa používa?

Nasledujúce otázky poskytujú podrobnosti o type úložiska. Navrhované možnosti sú klastrové, prúdové, dátové jazero alebo iné. Rozloženie odpovedí je znázornené v tabuľke.

**Popis údajov:**

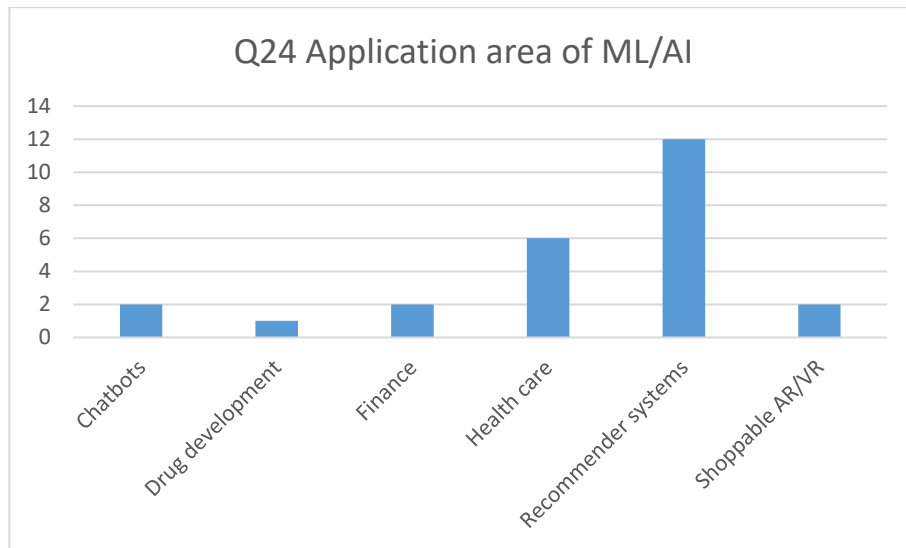
Výsledky naznačujú, že klastre je medzi výskumníkmi najobľúbenejším typom úložiska. Možným vysvetlením by mohlo byť, že klastre sú navrhnuté tak, aby poskytovali vysokú dostupnosť údajov. Klastre sú vysoko škálovateľné a odolné voči chybám. Klastre môžu poskytovať vysokú úroveň výkonu a môžu byť nákladovo efektívnym riešením ukladania dát, keďže ich možno zostaviť pomocou bežného hardvéru.

Diskusia:

- Klastre sú obľúbené ako typ úložiska, pretože ponúkajú vysokú dostupnosť, škálovateľnosť, odolnosť voči chybám, výkon a nákladovú efektívnosť. Tieto výhody z nich robia ideálnu voľbu pre podniky a aplikácie, ktoré vyžadujú spoľahlivé a škálovateľné úložné riešenia.

3.20. Oblasť použitia ML/AI?

Nasledujúce otázky odhaľujú oblasti použitia strojového učenia/umelej inteligencie. Otázka navrhuje osem možností – systémy odporúčaní, chatboti, A/B testy, obchodovateľná rozšírená realita-AR/VR, zdravotná starostlivosť, vývoj liekov, financie, kybernetická bezpečnosť. Výsledky sú uvedené v tabuľke nižšie.

**Popis údajov:**

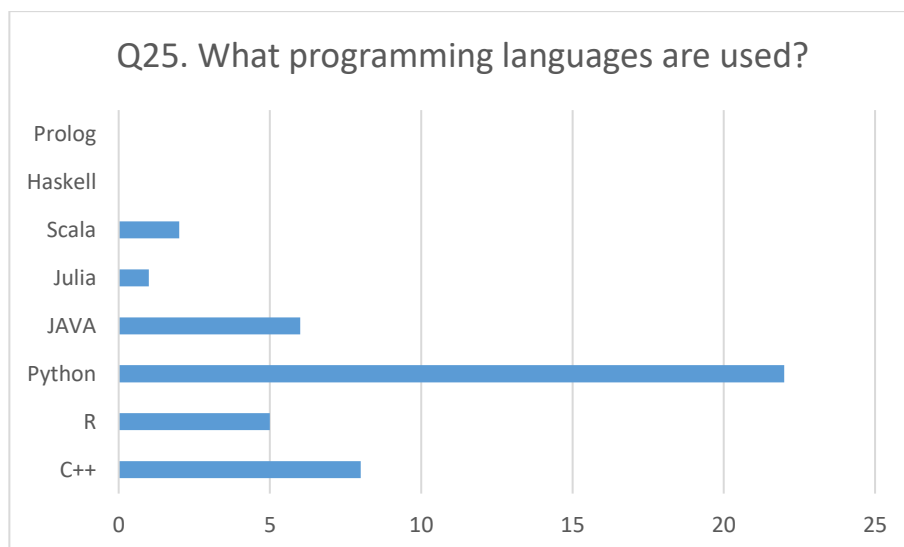
Údaje z kontrolovaných projektov ukazujú, že najbežnejšou oblasťou, v ktorej sa uplatňuje ML a AI, sú systémy odporúčaní. Systémy odporúčaní sú navrhnuté tak, aby používateľom poskytovali prispôbené odporúčania na základe ich preferencií. Systémy odporúčaní zvyčajne pracujú s veľkými údajmi, ako sú napríklad správanie používateľov a informácie o produktoch. Od mnohých systémov odporúčaní sa vyžaduje, aby poskytovali odporúčania v reálnom čase, napríklad pre stránky elektronického obchodu alebo streamingové platformy. Systémy odporúčaní môžu poskytnúť významné obchodné výhody, ako je zvýšený predaj, zapojenie zákazníkov a lojalita.

Diskusia:

- Systémy odporúčaní sú bežnou oblasťou použitia umelej inteligencie, pretože vyžadujú analýzu zložitých súborov údajov, aby poskytli personalizované odporúčania v reálnom čase. Algoritmy AI môžu pomôcť zautomatizovať a zlepšiť tento proces, čo môže viesť k lepším obchodným výsledkom a lepším užívateľským skúsenostiam.

3.21. Aké programovacie jazyky sa používajú?

Táto otázka poukazuje na programovacie jazyky, ktoré sa používajú pri problémoch s AI. Výsledky sú uvedené v tabuľke.

**Popis údajov:**

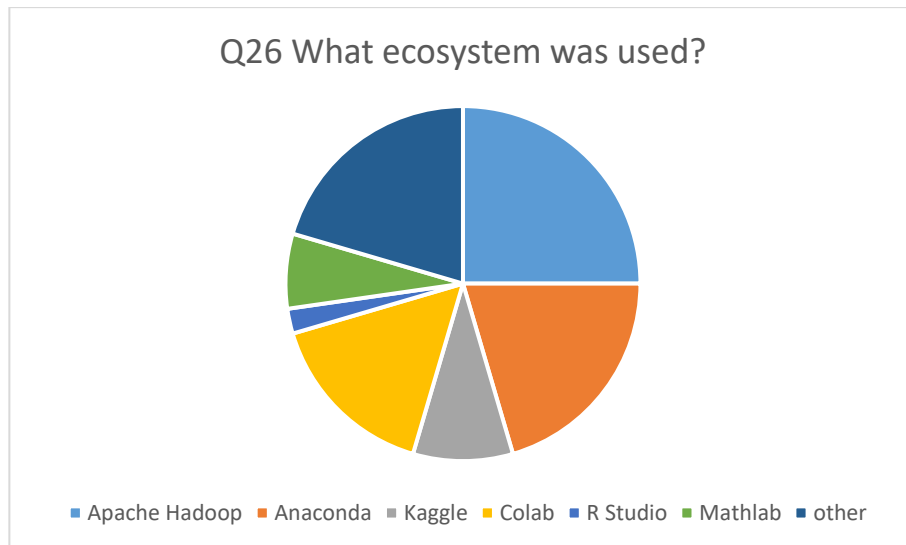
Výsledky jasne ukazujú, že najpoužívanejším programovacím jazykom je Python. Python má veľkú a aktívnu komunitu vývojárov, ktorí vytvorili množstvo knižníc a rámcov pre vývoj AI. Python je flexibilný jazyk, ktorý možno použiť pre širokú škálu úloh AI vrátane spracovania údajov, strojového učenia a spracovania prirodzeného jazyka. Okrem toho je Python kompatibilný so širokou škálou platforiem a systémov vrátane Windows, Mac a Linux. Aj keď Python nie je najrýchlejší programovací jazyk, je dostatočne rýchly pre väčšinu úloh AI.

Diskusia:

- Python je populárny programovací jazyk pre umelú inteligenciu, pretože sa jednoducho používa, má veľkú komunitu, flexibilitu, kompatibilitu a výkon. Tieto faktory z neho robia ideálnu voľbu pre vytváranie a nasadzovanie aplikácií AI.

3.22. Aký ekosystém bol použitý?

Ďalšia otázka poskytuje informácie o ekosystéme, ktorý bol použitý. Toto sú možné alternatívy - Apache Hadoop, Anaconda, Kaggle, Colab, R studio, Matlab alebo iné. Rozdelenie odpovedí je uvedené v tabuľke nižšie.

**Popis údajov:**

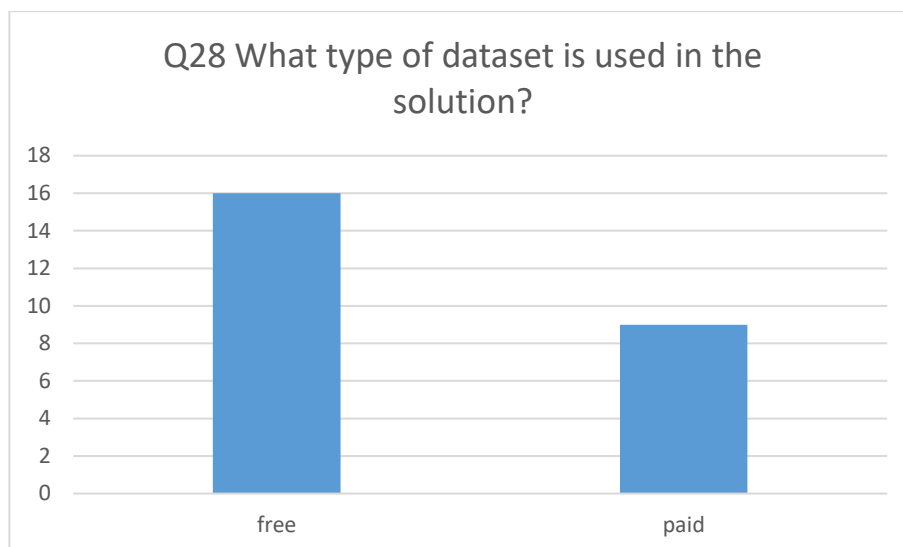
Výsledky naznačujú, že najviac sa využíval Apache Hadoop s miernou prevahou nad Anacondou a inými systémami. Možnosť „iné“ zahŕňa ekosystémy, ako napríklad NVIDIA CUDA, TensorFlow, lokálne Python IDE, Apache Spark, DataBricks.

Diskusia:

- Ekosystém Apache Hadoop je široko používaný na spracovanie a analýzu veľkých dát vďaka svojej škálovateľnosti, distribuovanému spracovaniu, odolnosti voči chybám, open source povahe, veľkému ekosystému a prijatiu v priemysle. Tieto faktory z neho robia ideálne riešenie pre organizácie, ktoré potrebujú spracovať a analyzovať veľké množstvo údajov.
- Anaconda a ďalšie ekosystémy súvisiace s jazykom Python sú tiež ďalšími populárnymi ekosystémami. Anaconda je distribúcia programovacích jazykov Python a R spolu so zbierkou open-source knižníc, nástrojov a rámcov pre vedu o údajoch a vedecké výpočty.

3.23. Aký typ súboru údajov sa používa v riešení?

Ďalšia otázka poskytuje informácie o type súboru údajov – či je bezplatný alebo platený. Výsledky sú prezentované v koláčovom grafe.

**Popis údajov:**

Výsledky ukazujú prevahu voľných datasetov, ktoré sú použité v riešeniach. Bezplatné súbory údajov sú prístupnejšie širšiemu okruhu používateľov vrátane študentov, výskumníkov a vývojárov. Mnoho bezplatných súborov údajov je otvorených údajov, čo znamená, že sú voľne dostupné komukoľvek na použitie, úpravu a distribúciu. Platené súbory údajov často prichádzajú s obmedzeniami, pokiaľ ide o ich použitie, čo môže obmedziť ich užitočnosť pre určité typy analýz alebo aplikácií.

Diskusia:

- Dostupnosť bezplatných súborov údajov prispela k rastu a rozvoju oblastí vedy o údajoch a strojového učenia podporou dostupnosti, otvorenosti, spolupráce a inovácií.

LITERATÚRA

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/databse>
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. Systematický prehľad a perspektíva výskumu odporúčacích systémov. 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, SP Mishra, Pooja Asopa, Konceptuálne chápanie konvolučnej neurónovej siete – prístup hlbokého učenia, Procedia Computer Science, zväzok 132, 2018, strany 679-688, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.05.069>.