Analýza využiteľnej technológie a testovanie zariadení

Life Defender - Ochranca života

Táto publikácia vznikla vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt *Life Defender – Ochranca života*, ktorý je spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

sféra
Grafické informačné systémy

SFÉRA, a.s. • Karadžičova 2 • 811 08 Bratislava
tel.: +421 2 502 13 142

www.sfera.sk

EURÓPSKA ÚNIA
Európsky fond regionálneho rozvoja
OP Integrovaná infraštruktúra 2014 – 2020

MINISTERSTVO
DOPRAVY
SLOVENSKEJ REPUBLIKY

MINISTERSTVO
ŠKOLSTVA, VEDY,
VÝSKUMU A ŠPORTU
SLOVENSKEJ REPUBLIKY

Táto publikácia je dielom kolektívu autorov:
Kováč Marek, Minárik Michal, Galajda Miroslav, Weissensteiner Anton

Ostatní autori:
Novotný Jozef, Moško Daniel, Kalužay Jozef, Kaňuk Martin

OBSAH

1 ÚVOD

Cieľom dokumentu je popísať vývoj softvérovej časti projektu a jeho nadväznosť na hardvérovú časť, ktorého činnosti sa realizovali v termíne 1.1.2023 - 30.06.2023. Dokument je výstupom projektu*Life Defender – Ochranca života*, kód projektu v ITMS: 313011ASQ6 v rámci míľnika č. 4 - *Analýza využiteľnej technológie a testovanie zariadení* pre aktivitu H1 - Riešenie SW platformy na integrovanie evidencie návštevníkov, zberov dát z existujúceho HW, ako i prototypov nového HW do jednotného informačného systému Life Defender – Ochranca života - prototyp (EV80) a aktivitu H2 - Riešenie SW platformy na integrovanie evidencie návštevníkov, zberov dát z existujúceho HW, ako i prototypov nového HW do jednotného informačného systému Life Defender – Ochranca života - prototyp (EV80) – flexibilita 15 %.

Vo výstupe projektu k jeho tretiemu míľniku sme popisovali vytvorené softvérové prototypy pre zariadenia, ktoré sú výstupom projektu. Súčasťou každej časti je podrobný popis architektúry riešenia, dátového modelu a webového rozhrania pre oblasť Domácej karantény, Automatickej testovacej bunky, Dezinfekčného robota v prepojení na Mobilnú aplikáciu a Model na zbieranie, analýzu a vyhodnocovanie symptómov.

V jednotlivých kapitolách aktuálneho míľnika postupne predstavujeme stav riešenia vzhľadom na ďalšie využitie vytvorenej platformy a venujeme sa témam ako možné prínosy a využitie platformy počas pandémie a mimo pandémie, registrácii a prihláseniu používateľov do systému, vykonávanie testov, poskytnutiu dát tretím stranám či v prípade dezinfekčného robota aj ovládaniu nastavovania jednotlivých módov robota. Výsledkom je aj synergický efekt ktorý sa týka väčšej celkovej účinnosti a hodnoty, ktorú môžu poskytnúť kombinované použitie softvéru pre domácu karanténu, automatickú testovaciu bunku a dezinfekčného robota.

Téma zavedenia a podpory platformy v budúcej prevádzke tvorí ďalšiu časť dokumentu. Pre jednotlivé pracovné balíky hardvérovej časti popisujeme základné funkčnosti, architektúru riešenia, integráciu na jednotné cloudové úložisko, ďalšie typy integrácií; taktiež problematiku bezpečnosti, autentifikácie, autorizácie, monitorovania a podpory prevádzky systémov.

Zámerom analýzy pre tému prototypu dátového modelu na zbieranie, analýzu a vyhodnocovanie symptómov pomocou analytických nástrojov s využitím umelej inteligencie bolo poskytnúť prehľad o otázkach bezpečnosti dát v systéme Life Defender – ochranca života. Zameriavame sa na zber, správu, analýzu, ukladanie a distribúciu údajov, ako aj na testovanie rôznych prístupov k anonymizácii údajov. Ponúkneme podrobnú analýzu prehľadu štatistických algoritmov pre analýzu a predikciu zdravotníckych dát vrátane témy metód strojového učenia a ich efektivity v kontexte modelovania zdravotníckych dát. Ďalšou témou bude zdravotný monitoring a metodologický prístup k modelovaniu zdravotných dát. Technické podrobnosti procesu anonymizácie boli zložité, ale neboli zahrnuté v tejto správe, ktorá sa zameriava na kľúčové zistenia, odporúčania a špecifikácie. Zaoberáme sa rovnako aj aspektmi, ako je zabezpečenie zberu údajov, generovanie súhrnných štatistík na monitorovanie a výskum a anonymizácia údajov, pričom sme zvážili overiteľnosť a prepojenie na konkrétne subjekty. Skúmame uchovávanie údajov a správnu manipuláciu, pričom zdôrazňujeme potrebu nenávratného zničenia pôvodných údajov a udržiavanie bezpečných intervalov údajov počas distribúcie. Analyzujeme minimálne charakteristiky vzorky pre efektívnu agregáciu údajov a minimalizáciu bezpečnostných rizík. Ďalšou oblasťou je vytváranie robustného bezpečnostného systému. Všetky zainteresované strany musia pochopiť rozhodnutie o anonymizácii údajov a jeho vzťah k ochrane údajov používateľov. Výber a implementácia metód anonymizácie zohľadňuje špecifický kontext systému, typy údajov, aplikácie a právne/regulačné požiadavky. Vzdelávanie a informovanosť medzi zainteresovanými stranami sú nevyhnutné, aby sa zabezpečila znalosť používateľov o procesoch anonymizácie, opatreniach na ochranu súkromia a možných dôsledkoch neoprávneného zdieľania údajov. Vývojári a poskytovatelia Školenie a informovanosť zohrávajú kľúčovú úlohu pri odstraňovaní bezpečnostných rizík.

2 VYUŽITIE PLATFORMY

Využitie platformy domácej karantény sleduje hlavný cieľ v tom, aby bol navrhnutý a implementovaný SW nastroj, ktorý počas pandémie, ale aj mimo nej pomôže modernizovať a zefektívniť zdravotníctvo a zvýši ochranu pacientov a zamestnancov zdravotného sektora.

2.1 Prototyp pre Domácu karanténu

Všetci používatelia platformy domácej karantény budú mať k dispozícii mobilnú aplikáciu, prostredníctvom ktorej budú vykonávať monitorovanie zdravotného stavu a webovú platformu s mobilným úložiskom, ktorá bude slúžiť na globálny monitoring správu a uchovávanie zozbieraných údajov.

Platforma je navrhnutá tak, aby prinášala hodnotu v rôznych situáciách a zohrávala kľúčovú úlohu v zlepšovaní zdravotnej starostlivosti a kontrole šírenia infekcií.

2.1.1 Možné prínosy a využitie platformy počas pandémie a mimo pandémie

Počas pandémie zlepšenie sledovania a kontroly šírenia vírusu. Umožnenie lepšieho sledovania a kontroly šírenia vírusu vďaka možnosti vykonávania testov na diaľku, sledovaniu zdravotných parametrov a upozorneniam na potenciálne zdravotné problémy. Týmto spôsobom môže platforma prispieť k rýchlej identifikácii a izolácii prípadov, čím sa znižuje riziko prenosu vírusu na ostatných.

Platforma tiež podporuje zachovanie kapacity zdravotníckych zariadení tým, že umožňuje pacientom vykonávať testy a sledovať svoje zdravie z pohodlia svojho domova. Tým sa znižuje zaťaženie nemocníc a zdravotníckych zariadení, umožňuje lepšiu koordináciu medzi zdravotníckymi pracovníkmi a poskytuje lepšiu starostlivosť o pacientov. Chráni zdravie zdravotného personálu a napomáha eliminovať stav, ku ktorému počas pandémii bežne dochádza, a to je nedostatok zdravotného personálu z dôvodu ochorenia.

Okrem využitia platformy počas pandémie je možné ju využiť aj v iných oblastiach, ako sú diagnostika, telemedicína a sledovanie chronických ochorení. Platforma môže byť prispôsobená a rozšírená tak, aby zohrávala úlohu v poskytovaní zdravotnej starostlivosti na diaľku a zlepšovala prístup k zdravotníckym službám.

2.1.2 Registrácia a prihlásenie používateľov do systému

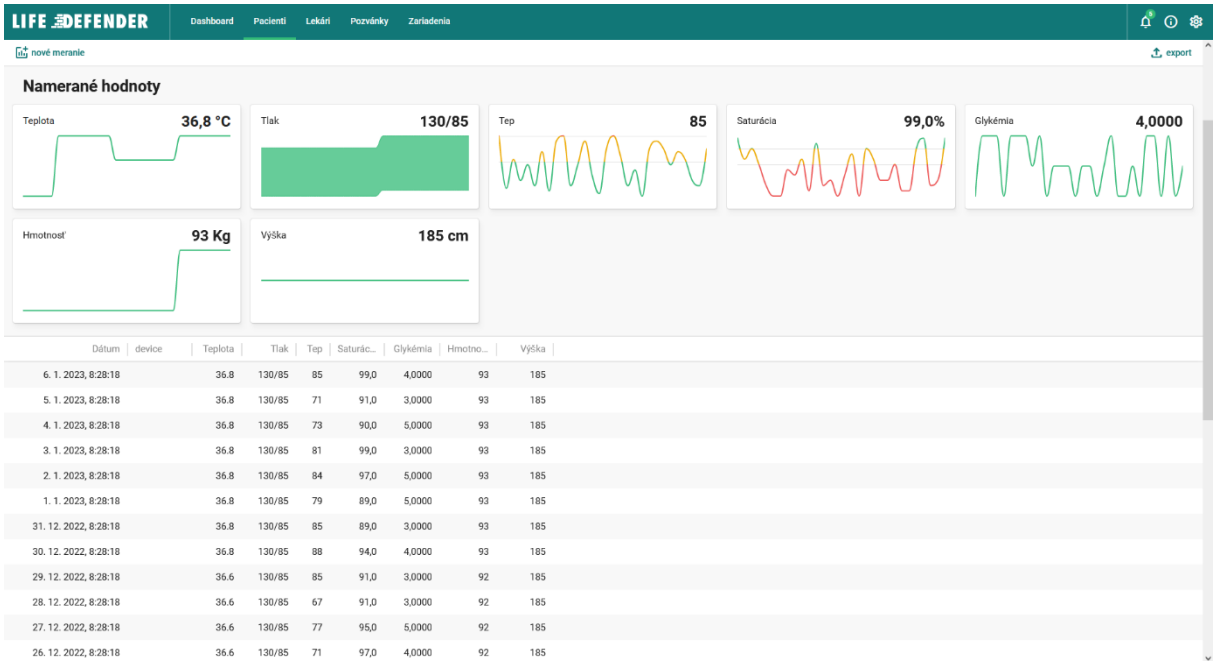
Registrácia do platformy prebieha na základe pozvánky, ktorú obdrží pacient. Systém vygeneruje jedinečný odkaz alebo kód pre každého pacienta, ktorý mu bude zaslaný prostredníctvom e-mailu. Týmto spôsobom je zabezpečené, že iba oprávnení používatelia majú prístup k platforme. Po kliknutí na odkaz alebo zadanie kódu z pozvánky sa pacient dostane na registračnú stránku, kde musí poskytnúť základné informácie, ako sú meno, priezvisko, e-mailová adresa a heslo.

Systém zbiera iba tie údaje, ktoré sú nevyhnutné pre poskytovanie služieb a účel platformy. Tým sa zabezpečuje, že sa spracováva minimálny počet osobných údajov potrebných pre plnenie jeho účelu.

Platforma zabezpečuje pomocou svojho back end systému, aby pacienti mohli uplatniť svoje práva podľa GDPR, ako je právo na prístup k svojim údajom, právo na opravu, právo na vymazanie údajov („právo byť zabudnutý“), právo na obmedzenie spracovania údajov, právo na prenosnosť údajov a právo namietať proti spracovaniu údajov. Súhlas pacientov so spracovaním ich osobných údajov je základným prvkom GDPR. Platforma informuje pacientov o účeloch spracovania údajov a o právach, ktoré majú v súvislosti so spracovaním ich osobných údajov.

2.1.3 Vykonávanie testov

Pacient vykoná test podľa poskytnutých inštrukcií. Testovanie môže zahŕňať meranie teploty, tepu, tlaku alebo iných relevantných parametrov. Testovacie zariadenie je napojené na mobilnú aplikáciu, prostredníctvom ktorej sa údaje automaticky upladuju na webovú platformu, ktorá zbiera a zaznamenáva údaje počas testovania. Po dokončení testu platforma automaticky analyzuje získané údaje a poskytuje výsledky pacientovi. Platforma umožňuje pacientom sledovať svoje výsledky testov v čase, čo im pomáha získať prehľad o svojom zdravotnom stave a sledovať pokrok v liečbe alebo zlepšenie zdravia. Výsledky testov sú okamžite zdieľané s lekármi alebo inými zdravotníckymi pracovníkmi prostredníctvom centralizovanej webovej platformy na konzultáciu a ďalšie vyhodnotenie. Platforma umožňuje bezpečné zdieľanie údajov z testov s lekármi a zdravotníckymi pracovníkmi, čo umožňuje lepšiu spoluprácu, rýchlejšiu diagnostiku a efektívnejšiu liečbu.



Dátum	device	Teplota	Tlak	Tep	Saturác...	Glykémia	Hmotno...	Výška
6. 1. 2023, 8:28:18		36.8	130/85	85	99,0	4,0000	93	185
5. 1. 2023, 8:28:18		36.8	130/85	71	91,0	3,0000	93	185
4. 1. 2023, 8:28:18		36.8	130/85	73	90,0	5,0000	93	185
3. 1. 2023, 8:28:18		36.8	130/85	81	99,0	3,0000	93	185
2. 1. 2023, 8:28:18		36.8	130/85	84	97,0	5,0000	93	185
1. 1. 2023, 8:28:18		36.8	130/85	79	89,0	5,0000	93	185
31. 12. 2022, 8:28:18		36.8	130/85	85	89,0	3,0000	93	185
30. 12. 2022, 8:28:18		36.8	130/85	88	94,0	4,0000	93	185
29. 12. 2022, 8:28:18		36.6	130/85	85	91,0	3,0000	92	185
28. 12. 2022, 8:28:18		36.6	130/85	67	91,0	3,0000	92	185
27. 12. 2022, 8:28:18		36.6	130/85	77	95,0	5,0000	92	185
26. 12. 2022, 8:28:18		36.6	130/85	71	97,0	4,0000	92	185

Obrázok 1 Web platforma – prehľad nameraných údajov

2.1.4 Poskytnutie dát tretím stranám

V projekte bolo brané na zreteľ, že v prípade epidémii budú takéto zdravotnícke informácie potrebne aj pre ďalšie subjekty štátu a zdravotníctvo preto, aby bolo možné situáciu vyhodnocovať a monitorovať. Pred poskytnutím údajov tretím stranám je dôležité získať súhlas pacienta. Pacient má právo kontrolovať, ktoré údaje sú zdieľané a s kým. Platforma bude umožňovať pacientom nastaviť svoje preferencie týkajúce sa zdieľania údajov a súhlas s poskytnutím údajov tretím stranám.

Údaje pacientov môžu byť poskytnuté tretím stranám v niektorých prípadoch, napríklad:

- lekárom alebo zdravotníckym pracovníkom pre účely diagnostiky a liečby,
- výskumným inštitúciám alebo akademickým organizáciám pre účely výskumu a štúdia,
- štátnym alebo verejným zdravotníckym organizáciám pre účely sledovania pandémií, alebo epidémií,
- poisťovniam alebo iným finančným inštitúciám pre účely plnenia poistných záväzkov.

V prípade, ak by malo dôjsť k poskytovaniu údajov tretím stranám bude musieť byt implementovaná aj čiastočná alebo úplná anonymizácia, alebo pseudonymizácia údajov pred poskytnutím tretím stranám. Tým sa minimalizuje riziko porušenia súkromia pacientov a zabezpečuje ochrana ich osobných údajov.

Anonymizácia znamená odstránenie všetkých osobne identifikovateľných informácií, zatiaľ čo pseudonymizácia zahŕňa nahradenie týchto informácií náhodnými identifikátormi.

2.2 Prototyp pre Automatickú testovaciu bunku

Automatická testovacia bunka je z pohľadu SW vybavenia vybavená dvoma dotykovými obrazovkami, prostredníctvom ktorých je riadený proces registrácie a identifikácie testovanej osoby ako vykonanie jednotlivých krokov vykonania testu a oznámenia výsledkov testovanej osobe.

2.2.1 Možné prínosy a využitie platformy počas pandémie a mimo pandémie

Počas pandémie môže byť Automatická testovacia bunka využitá na efektívne a rýchle testovanie veľkého počtu ľudí. Softvérová časť zabezpečuje správne riadenie testovacieho procesu, analýzu dát a poskytovanie výsledkov v reálnom čase. Platforma môže byť integrovaná s existujúcimi zdravotníckymi systémami pre lepšiu koordináciu a komunikáciu medzi zdravotníckymi pracovníkmi.

Mimo pandémie môže byť automatická testovacia bunka upravená a použitá na testovanie iných infekčných chorôb alebo zdravotných stavov. Softvérová platforma môže byť rozšírená o nové moduly a funkcie pre rôzne účely, ako sú diagnostika, sledovanie a liečba. Taktiež môže byť využitá v rámci preventívnych programov a zdravotnej starostlivosti.

Hardvérové zariadenia môžu byť tiež prispôsobené pre iné účely, napríklad testovanie alergií, cukrovky alebo kardiovaskulárnych chorôb. Senzory a meracie zariadenia môžu byť vymenené alebo pridané podľa potreby, čo umožňuje širokú škálu možností a prispôsobenie rôznym zdravotníckym požiadavkám.

2.2.2 Registrácia a prihlásenie používateľov do systému

Registrácia používateľa prebieha na dotykovej obrazovke, ktorá umožňuje pacientovi zadávať svoje osobné údaje. Proces môže zahŕňať nasledujúce kroky:

1. Zadanie osobných údajov: Pacient zadáva svoje meno, priezvisko, dátum narodenia a iné požadované informácie.
2. Zadanie kontaktnej informácie: Pacient zadáva svoj e-mail alebo telefónne číslo pre prípadnú komunikáciu a zaslanie výsledkov.
3. Do systému je možné pridať aj osobu bez dokladu totožnosti, resp. dieťa, ale jeho registráciu musí zabezpečiť zaregistrovaná osoba pod svojím kontom.

Autorizácia a overenie totožnosti:

1. Porovnanie obrazu tváre s dokladom totožnosti: Softvér porovná fotografie tváre a dokladu totožnosti pacienta, aby zistil či sa zhodujú.
2. Biometrická autentifikácia: Systém použije biometrické údaje z rozpoznávanie tváre, na ďalšie overenie totožnosti pacienta.
3. Výsledok overenia: Ak sa obraz tváre zhoduje s obrazom na doklade totožnosti, ktorý pacient naskenuje a prejde biometrickými kontrolami na porovnanie tvare a identifikačného dokladu, pacient je úspešne overený a autorizovaný na používanie automatickej testovacej bunky.

Obrázok 2 Verifikačný proces

2.2.3 Vykonávanie testov

Softvér automatickej testovacej bunky poskytuje jednoduchý a prehľadný wizard, ktorý sprevádza pacienta procesom vykonávania testov na samostatnej obrazovke, ktorá slúži len k vykonávaniu testov. Tento wizard zahŕňa nasledujúce kroky:

1. Úvod: Pacient je uvítaný na úvodnej obrazovke, kde je stručne informovaný o krokoch, ktoré ma vykonať.
2. Inštrukcie pre testovanie: Pacient dostane podrobné inštrukcie, ako vykonať test, vrátane použitia príslušného testovacieho zariadenia a správneho postupu na odber vzorky.
3. Vykonanie testu: Pacient vykonáva test podľa poskytnutých inštrukcií.
4. Analýza výsledkov: Softvér automaticky analyzuje výsledky testu a generuje správu o výsledkoch.
5. Informácia o zaslaní výsledkov: Pacient je informovaný, že výsledky testu budú zaslané na jeho e-mailovú adresu.

Obrázok 3 Inštrukcie pre vykonanie testu

2.3 Prototyp pre Dezinfekčného robota

Pre použitie dezinfekčného robota bol navrhnutý prototyp mobilnej aplikácie slúžiacej na ovládanie robota, ktorá umožni jednak priame ovládanie robota, ale aj naplánovanie jednotlivých funkčností, ktoré majú byt vykonane v rôznom čase, ako napríklad dezinfekcie len vybraných miestnosti alebo detekcie vírusov statická alebo pohybová.

2.3.1 Možné prínosy a využitie platformy počas pandémie a mimo pandémie

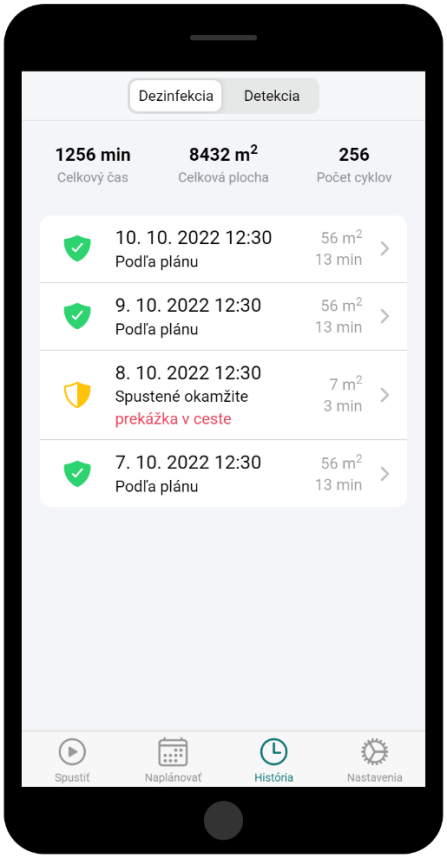
Počas epidémie mobilná aplikácia umožňuje rýchle a efektívne ovládanie dezinfekčného robota, ktorý môže byť nasadený v zdravotníckych zariadeniach, verejných priestoroch, školách a iných miestach, kde je potrebné minimalizovať šírenie vírusu. Dezinfekčný robot znižuje riziko pre zdravotnícky personál tým, že automatizuje dezinfekčný proces a znižuje potrebu manuálnej práce. Mobilná aplikácia umožňuje plánovanie dezinfekčných prác a sledovanie priebehu dezinfekcie, čo umožňuje efektívne rozdelenie zdrojov a kontrolu nad procesom detekcie vírusov a dezinfekcie.

Mimo pandémie môže byť dezinfekčný robot využitý aj na prevenciu šírenia iných infekčných chorôb a zlepšenie celkovej úrovne hygieny v rôznych zariadeniach a verejných priestoroch. Mobilná aplikácia môže byť rozšírená a prispôsobená pre rôzne účely a prostredia, ako sú priemyselné zariadenia, kancelárie, hotely alebo dopravné prostriedky. Môže byť integrovaná s existujúcimi systémami pre správu budov, zabezpečenie alebo zdravotnú starostlivosť, čo umožňuje koordináciu a komunikáciu medzi rôznymi zariadeniami a službami.

2.3.2 Ovládanie a nastavovanie jednotlivých modov robota

Používatelia môžu prostredníctvom mobilnej aplikácie naplánovať jednotlivé módy detekcii vírusov a dezinfekcií:

1. Výber módu: Používatelia môžu zvoliť či chcú použiť robot na detekciu vírusov, dezinfekciu, alebo oboje.
2. Nastavenie časového plánu: Používatelia môžu nastaviť časový plán pre detekciu a dezinfekciu, napríklad denne, týždenne alebo v určitých intervaloch.
3. Voľba dezinfekčných prostriedkov: Používatelia môžu zvoliť typ dezinfekčných prostriedkov, ktoré chcú robot používať a nastaviť ich koncentráciu a množstvo.
4. Vytvorenie mapy: Používatelia môžu vytvoriť mapu priestoru, kde má robot pôsobiť a označiť oblasti, ktoré vyžadujú detekciu a/alebo dezinfekciu.
5. Nastavenie trasy: Používatelia môžu nastaviť trasu robota tak, aby zabezpečil účinnú detekciu a dezinfekciu vybraných oblastí. Trasu je možné prispôsobiť podľa veľkosti priestoru, množstva ľudí v danom priestore a iných faktorov.
6. Uloženie a história: Systém automaticky ukladá jednotlivé vykonane trasy a zbiera údaje z detekcie vírusov a vykonaných dezinfekcií



Obrázok 4 História vykonaných dezinfekcií

2.4 Synergický efekt

Synergický efekt v tejto súvislosti sa týka väčšej celkovej účinnosti a hodnoty, ktorú môžu poskytnúť kombinované použitie softvéru pre domácu karanténu, automatickú testovaciu bunku a dezinfekčného robota. Opisujeme nasledujúce hlavné oblasti, kde môže dochádzať k synergickým efektom:

Zdieľanie údajov

Integrácia rôznych systémov a zdieľanie údajov medzi nimi umožňuje efektívne monitorovanie, sledovanie a reagovanie na pandemickú situáciu. Táto integrácia môže zahŕňať:

1. Výmenu údajov o testovaných osobách medzi domácim karanténnym systémom a automatickou testovacou bunkou.
2. Synchronizáciu údajov o dezinfekcii medzi dezinfekčným robotom a domácim karanténnym systémom alebo automatickou testovacou bunkou.
3. Spoluprácu medzi systémami na plánovanie a vykonávanie opatrení, ako sú testovanie, dezinfekcia a karanténa.

Zlepšenie efektívnosti

Synergické efekty medzi systémami môžu viesť k zlepšeniu efektívnosti v rámci celého procesu riadenia pandémie:

1. Domáca karanténa môže byť účinnejšie monitorovaná a riadená vďaka informáciám získaným z automatickej testovacej bunky a dezinfekčného robota.

2. Automatická testovacia bunka môže lepšie sledovať a očakávať potrebu testovania na základe údajov z domácej karantény a dezinfekčného robota.
3. Dezinfekčný robot môže byť presnejšie nasmerovaný na oblasti s vyšším rizikom šírenia vírusu vďaka údajom z automatickej testovacej bunky.

3 ZAVEDENIE A PODPORA PLATFORMY V PREVÁDZKE

3.1 Prototyp pre Domácu karanténu

3.1.1 Opis funkčnosti

Informačný systém pre domácu karanténu sa skladá z dvoch kľúčových častí:

1. cloudová služba,
2. mobilná aplikácia.

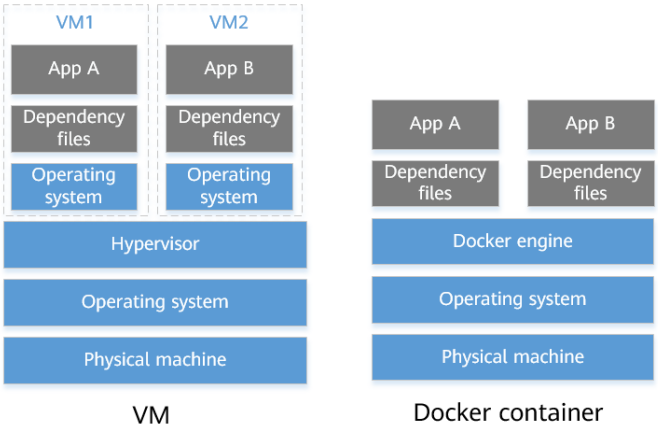
Cloudová služba – platforma pre ukladanie nameraných údajov v cloude poskytuje používateľom webové rozhranie a rozhranie pre automatizovanú komunikáciu cez REST API. Tieto rozhrania umožňujú získavať údaje z rôznych zdrojov a synchronizovať ich s mobilnou aplikáciou.

Mobilná aplikácia – umožňuje používateľom získať a odoslať namerané údaje do cloudovej platformy. Údaje z meracieho zariadenia (Checkme Pro) sa ukladajú do internej databázy aplikácie a následne sa synchronizujú s cloudovou platformou. Pre vytvorenie mobilnej aplikácie bola zvolená platforma Xamarin, ktorá umožňuje vytvárať mobilné aplikácie pre obidve platformy - Android aj iOS - súčasne.

Všetky zvolené technológie a frameworky sú open source.

3.1.2 Opis architektúry nasadenia

Cloudová služba pre domácu karanténu bola vytvorená s cieľom umožniť prevádzku v docker kontajneroch, čo predstavuje najmodernejší spôsob prevádzkovania informačných systémov v súčasnosti. Projekt sa zameriaval na testovanie tejto metódy a využitie jej výhod v porovnaní so štandardnou prevádzkou na virtuálnych serveroch (virtualizácia).



Obrázok 5 Docker kontajner vs Virtuálne servery (VM)

Na obrázku je vidieť porovnanie dvoch najpoužívanejších možnosti prevádzkovania informačných systémov, a to:

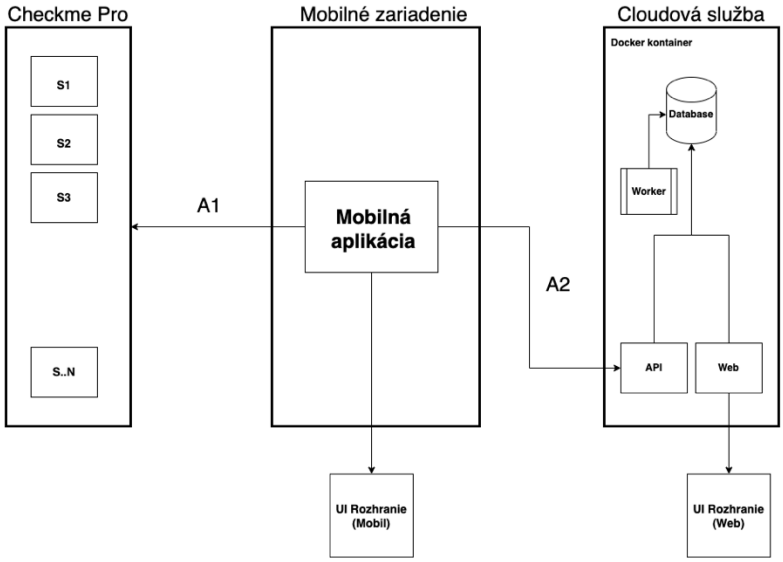
- VM - Virtualizácia,
- Docker container – Kontajnerizácia.

Virtualizácia – je technológia, ktorá umožňuje prevádzkovať viacero nezávislých virtuálnych serverov na jednom fyzickom serveri. Prevádzka informačných systémov v tomto virtuálnom prostredí má množstvo výhod oproti prevádzke na fyzickom prostredí, najmä z hľadiska bezpečnosti a ekonomiky.

Kontajnerizácia – je forma virtualizácie, ktorá využíva menšie množstvo systémových zdrojov. V tejto metóde sú aplikácie prevádzkované v docker kontajneroch, ktoré obsahujú iba minimálne systémové prostriedky potrebné pre prevádzku aplikácie. Prevádzka pomocou docker kontajnerov prináša množstvo výhod, ako napríklad zjednodušenie nasadenia aplikácií, minimalizáciu závislostí aplikácií na operačnom systéme a spravovanie aktualizácií aplikácií a prostredia.

3.1.3 Integrácia na jednotne cloudové úložisko

Nasledujúci obrázok zobrazuje základnú logickú architektúru informačného systému pre domácu karanténu.



Obrázok 6 Logická architektúra systému

Prostredníctvom technológie Bluetooth (integrácia A1 na obrázku) mobilná aplikácia synchronizuje dáta zo zariadenia Checkme Pro do svojej internej databázy. Následne tieto získané dáta sa synchronizujú prostredníctvom Rest API služieb (integrácia A2 na obrázku) do cloudovej služby. Mobilná aplikácia poskytuje používateľské rozhranie pre ovládanie synchronizácie dát. Cloudová služba umožňuje prístup k nameraným dátam prostredníctvom webového rozhrania. Dáta sú ukladané do relačnej databázy v cloudovej službe. Cloudová služba obsahuje komponentu worker, ktorá predstavuje sadu aplikačných služieb pre vykonávanie asynchrónnych operácií.

3.1.3.1 Typy integrácií

Systém zahŕňa dve dôležité integrácie: jednu medzi mobilnou aplikáciou a cloudovou službou a druhú medzi mobilnou aplikáciou a meracím zariadením Checkme Pro.

Tabuľka 1 Zoznam integrácií

Integrácia	Systém	Popis
A1	Mobilná ap. -> Checkme Pro	Bluetooth slúži na synchronizáciu
A2	Mobilná ap. -> cloudová služba	HTTP Rest API slúži na synchronizáciu

3.1.4 Zabezpečenie osobných dát

Zabezpečenie osobných dát v informačnom systéme vychádza zo všeobecného nariadenia o ochrane údajov – GDPR:

Šifrovanie údajov - použitie šifrovania na ochranu osobných údajov v databáze. Šifrovanie zaručuje, že údaje sú nečitateľné pre neoprávnené osoby, ak by sa dostali do neoprávneného prístupu k údajom.

Prístupové práva a kontrola oprávnení: Pravidelná kontrola prístupových práv a oprávnení používateľov k osobným údajom. To zahŕňa pridelenie oprávnení len tým osobám, ktoré majú oprávnenie pristupovať k osobným údajom a upravovať ich, sledovanie a správu prístupových práv.

Bezpečnostné zálohy: Pravidelné zálohovanie osobných údajov a implementácia opatrení na ich zabezpečené uloženie a obnovu.

Audity a monitorovanie: Sledovanie a monitorovanie prístupov k osobným údajom v informačnom systéme, vrátane záznamov o prístupoch a aktivitách s osobnými údajmi.

Politiky a postupy ochrany údajov: Vypracovanie a implementácia jasných politík a postupov týkajúcich sa ochrany údajov v rámci informačného systému.

Bezpečnostné aktualizácie a patche: Pravidelná implementácia bezpečnostných aktualizácií a patchov.

3.1.5 Bezpečnosť

Bezpečnosť systému je zabezpečená na viacerých úrovniach, vrátane infraštruktúry a aplikačnej vrstvy. Infraštruktúra je navrhnutá tak, aby bola vysoko dostupná a minimalizovala riziko straty dát. Všetky používateľské aj automatizované rozhrania sú chránené pred neautorizovaným prístupom a zabezpečené pomocou SSL protokolu.

3.1.5.1 Autentifikácia

Vstup používateľov do informačného systému pre domácu karanténu sa zabezpečuje autentifikáciou pomocou autentifikačného modulu, ktorý je založený na štandardoch OpenID Connect/OAuth 2.0. Pri autentifikácii používateľ zadá svoje prihlasovacie meno a heslo, ktoré autentifikačný modul overí. Ak autentifikácia prebehne úspešne, autorizačný modul pridelí používateľovi príslušné oprávnenia na základe jeho používateľských rolí. V prípade neúspešnej autentifikácie sa používateľovi zakáže vstup do systému.

Po úspešnej autentifikácii používateľ získa autentifikačný token s preddefinovanou dĺžkou životnosti, ktorý mu umožní pristupovať k jednotlivým častiam systému na základe jeho pridelených používateľských rolí. Ak autentifikačný token exspiruje, používateľ sa musí znovu prihlásiť. Autentifikácia a autentifikačný token sú zabezpečené proti neautorizovanému vstupu a všetky rozhrania systému sú zabezpečené pomocou SSL protokolu.

3.1.5.2 Autorizácia

V systéme pre domácu karanténu sa autorizácia týka riadenia prístupu používateľa k funkčnostiam a dátam informačného systému. Tento proces bude založený na rolách RBAC (Role-based access control), čo znamená, že každá používateľská rola bude mať priradené určité oprávnenia. Administrátor bude zodpovedný za správu používateľských rolí a priradenie ich k jednotlivým používateľom.

3.1.6 Monitorovanie systému

Monitoring informačného systému je rozdelený do troch základných úrovní, a to:

1. Systémová - systémová úroveň monitoringu je sledovanie dostupnosti informačného systému z pohľadu prostredia, kde je nainštalovaný (infraštruktúra, podporné doménové služby).

2. Aplikačná - aplikačná úroveň monitoringu je sledovanie dostupnosti komponentu informačného systému a niektorých ich vlastností.
3. Procesná (biznis) - procesná úroveň monitoringu je sledovanie stavu dôležitých biznis procesov informačného systému.

Na monitorovanie informačného systému bude použitý nástroj Nagios, ktorý je open-source softvér určený na monitorovanie IT infraštruktúry. Slúži na sledovanie stavu a výkonnosti hardvéru, softvéru, sietí a aplikácií, čím umožňuje prevádzkovateľom a správcom systémov rýchle odhaľovanie problémov a ich následné riešenie. Nagios dokáže monitorovať množstvo rôznych zariadení a služieb, ako sú napríklad servery, routery, switche, aplikácie a databázy. V prípade detekcie nejakého problému môže Nagios upozorniť správcov systémov rôznymi spôsobmi, ako sú e-mail, SMS, hlasové hovory, správy na displeji a podobne. Nagios teda zabezpečuje neustále sledovanie výkonu IT infraštruktúry a pomáha predchádzať výpadkom alebo iným problémom, ktoré by mohli ohroziť prevádzku informačného systému.

3.1.7 Podpora prevádzky systému

Prevádzka systému bude zabezpečená bežnými dostupnými možnosťami:

- informáciami na stránke – dokumentácia a FAQ,
- emailom,
- telefonickým kontaktom.

Pri kontaktovaní podpory prevádzky prostredníctvom emailu alebo telefonicky bude vytvorený tiket v tiketovacom nástroji OTRS (Open-source Ticket Request System), ktorý slúži ako softvérový nástroj na správu a sledovanie požiadaviek zákazníkov a problémov v oblasti IT služieb, a podpory zákazníkov. Je to open-source systém, ktorý je vyvinutý v jazyku Perl a beží na webovom serveri. Nástroj poskytuje funkcie, ktoré umožňujú organizáciám vytvoriť centrálny bod pre správu požiadaviek a dotazov od zákazníkov. Systém umožňuje zákazníkom posielať požiadavky prostredníctvom e-mailu alebo webového rozhrania a sledovať stav svojich požiadaviek. Správcovia systému potom môžu tieto požiadavky prijímať, prideliť ich príslušným zamestnancom a sledovať ich riešenie. Nástroj obsahuje tiež ďalšie užitočné funkcie, ako je správa znalostnej bázy, prehľady a štatistiky, automatická eskalácia, sledovanie časov reakcie a ďalšie. Systém je navrhnutý tak, aby bol ľahko konfigurovateľný a prispôsobiteľný potrebám rôznych organizácií.

Personálne obsadenie podpory prevádzky je rozdelené do troch základných úrovní: L1, L2 a L3:

- **Úroveň L1** – odborníci sa zaoberajú základnými spotrebiteľskými problémami a majú všeobecné znalosti o produkte a službách. Získavajú informácie o zákazníkoch, analyzujú príznaky a identifikujú základné problémy.
- **Úroveň L2** – odborníci majú rozsiahlejšie skúsenosti a znalosti a môžu pomôcť odborníkom na úrovni L1 pri riešení základných technických problémov. Skúmajú existujúce problémy a hľadajú známe riešenia pre zložitejšie problémy.
- **Úroveň L3** – špecialisti sa zaoberajú najnáročnejšími problémami a sú odborníkmi vo svojom odbore. Niekedy pomáhajú odborníkom na úrovni L1 aj L2. Okrem toho skúmajú a vyvíjajú riešenia pre nové alebo neznáme problémy.

3.2 Prototyp pre Automatickú testovaciu bunku

3.2.1 Opis funkčnosti

Informačný systém pre dezinfekčného robota sa skladá z troch kľúčových častí:

1. Cloudová služba.
2. Interný systém pre ovládanie dezinfekčného robota.
3. Mobilná aplikácia pre ovládanie dezinfekčného robota.

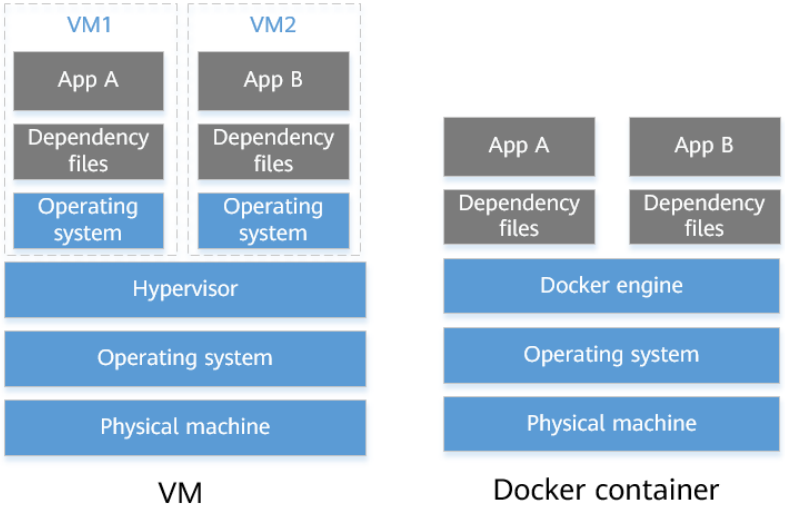
Cloudová služba je vytvorená ako platforma pre uloženie meraných údajov získaných z rôznych zdrojov. Tieto údaje sa ukladajú do cloudovej služby, ktorá ponúka používateľom webové rozhranie a rozhranie pre automatizovanú komunikáciu pomocou Rest API. Rozhranie Rest API je určené na synchronizáciu údajov so systémom pre ovládanie testovacej bunky.

Interný systém pre ovládanie dezinfekčného robota je vložený do automatickej testovacej bunky a slúži na ovládanie robota a zasielanie dát do cloudovej služby.

Mobilná aplikácia pre ovládanie dezinfekčného robota slúži na vzdialené ovládanie dezinfekčného robota.

3.2.2 Opis architektúry nasadenia

Informačný systém pre automatickú testovaciu bunku je vytvorený tak, aby sa dal prevádzkovať v docker kontajneroch (kontajnerizácia). V dnešnej dobe je to najmodernejší spôsob prevádzkovania informačných systémov. V rámci tohto projektu bol jeden z cieľov vyskúšať tento spôsob, aby sme využili jeho výhody voči štandardnej prevádzke vo virtuálnych serveroch (virtualizácia).



Obrázok 7 Docker kontajner vs Virtuálne servery (VM)

Na obrázku je vidieť porovnanie dvoch najpoužívanejších možnosti prevádzkovania informačných systémov, a to:

1. VM - Virtualizácia,
2. Docker container – Kontajnerizácia.

Virtualizácia – je technológia, ktorá umožňuje prevádzkovať viacero nezávislých virtuálnych serverov na jednom fyzickom serveri. Prevádzka informačných systémov v tomto virtuálnom prostredí má množstvo výhod oproti prevádzke na fyzickom prostredí, najmä z hľadiska bezpečnosti a ekonomiky.

Kontajnerizácia – je forma virtualizácie, ktorá využíva menšie množstvo systémových zdrojov. V tejto metóde sú aplikácie prevádzkované v docker kontajneroch, ktoré obsahujú iba minimálne systémové.

Tabuľka 2 Zoznam kontajnerov

Názov kontajnera	Image	Popis
ld-web	nginx:latest	Kontajner pre webový server
ld-api	mcr.microsoft.com/dotnet/aspnet:5.0	Kontajner pre API služby
ld-worker	mcr.microsoft.com/dotnet/runtime:6.0	Kontajner pre aplikačný server
ld-pgadmin	dpage/pgadmin4:6	Kontajner pre webové rozhrania databázového servera
ld-postgres	postgres:13.4	Kontajner pre databázový server
ld-dis	docker:5000/ld-dis:latest	Kontajner pre digitál identity service

3.2.2.1.1 Prostredia

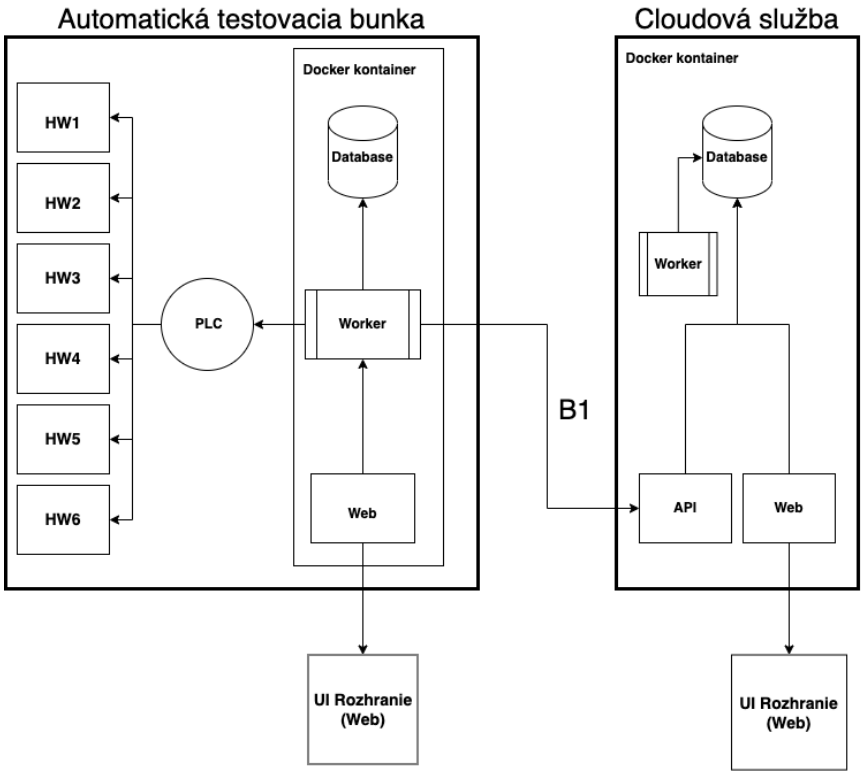
Cloudová služba pre automatickú testovaciu bunku je prevádzkovaná v jednom vývojovom prostredí na infraštruktúre spoločnosti sféra, a.s.

Tabuľka 3 Zoznam prostredí

Názov	Doména	Infraštruktúra
Vývojové prostredie	lifedefender.sk	HW spoločnosti sféra, a.s.

3.2.3 Integrácia na jednotné cloudové úložisko

Na nasledujúcom obrázku je vidieť základnú logickú architektúru systému.



Obrázok 8 Logická architektúra systému

Cloudová služba sa skladná z nasledujúcich logických komponentov:

- Web,
- API,
- Worker,
- Databáza.

Cloudová služba poskytuje Rest API rozhranie pre synchronizáciu dát z automatickej testovacej bunky. Synchronizované dáta sa uložia do relačnej databázy. Dáta sú sprístupnené pre používateľov pomocou cloudovej služby prostredníctvom webového rozhrania. Cloudová služba obsahuje aplikačný server, ktorý vykonáva asynchronne úlohy, ako je notifikovanie používateľov alebo príprava dát pre export.

(Vnorený systém) softvérová časť automatickej testovacej bunky sa skladná z nasledujúcich logických komponentov:

- Web,
- Worker,
- Databáza.

Systém pre ovládanie automatizovanej testovacej bunky je vnorený do docker kontajnerov na internom serveri bunky a pozostáva z troch základných komponentov: webového rozhrania, workeru a databázy.

Webové rozhranie poskytuje používateľské rozhranie na ovládanie testovacej bunky, a je prístupné prostredníctvom dotykového displeja. Komunikuje s aplikačným serverom a databázou pomocou interného API rozhrania.

Worker obsahuje aplikačné služby na ovládanie hardvéru testovacej bunky prostredníctvom PLC rozhrania. Služba DIS slúži na autentifikáciu používateľa pomocou biometrických údajov a na synchronizáciu spracovaných dát do cloudovej služby.

Databáza slúži ako dočasné úložisko pre nesynchronizované dáta.

3.2.3.1 Typy integrácií

Primárna integrácia je medzi automatickou testovacou bunkou a cloudovou službou. Pomocou tejto integrácie sa synchronizujú nameraná dáta smerom do cloudovej služby.

Tabuľka 4 Zoznam integrácií

Integrácia	Systém	Popis
B1	Testovacia bunka -> cloudová služba	HTTP Rest API, slúži na synchronizáciu dát

3.2.3.2 Komponenty systému

- **Web (UI rozhranie)** - komponent, ktorý zabezpečuje používateľské rozhranie prostredníctvom webového rozhrania.
- **API** - komponent, ktorý zabezpečuje REST API služby pre synchronizáciu dát medzi automatickou testovacou bunkou a cloudovou službou.
- **Worker** - komponent, ktorý reprezentuje sadu aplikačných služieb.
- **Database** - komponent, ktorý zabezpečuje perzistentné alebo dočasné úložisko pre namerané dáta – relačná databáza.
- **PLC** - komponent, ktorý zabezpečuje komunikáciu s HW časťami testovacej bunky.

- **Digital Identity Service (DIS)** - softvérový produkt slovenskej spoločnosti Innovatrics, líder trhu s biometrickým softvérom, slúži na overenie identity osoby automatizovane a na diaľku bez potreby obslužného personálu. Overenie identity funguje tak, že osoba na vyzvanie systému naskenuje doklad totožnosti a umiestni svoju tvár pred kameru. Systém extrahuje z dokladu totožnosti všetky dostupné informácie a overí pravosť a fyzickú prítomnosť dokumentu pri skenovaní. Sken tváre zabezpečí kontrolu živosti osoby pred kamerou. Následne sú porovnané údaje dokumentu s výsledkami analýzy fotografie osoby. Porovnáva sa tvár, odhadovaný vek, pohlavie a ďalšie parametre. Výsledkom množstva kontrol je skóre autentickosti osoby, ktoré je spolu so všetkými zozbieranými dátami poskytnuté klientskému systému.

  Služba Digital Identity Service je lokálne hosťovaná u klienta buď v cloude, alebo na serveri, ktorý je súčasťou lokálnej infraštruktúry. Jej rozhraním je sa REST API endpointov, ktoré klientský systém volá v rôznych krokoch overovania identity osoby. Služba neuchováva žiadne osobné údaje, perzistencia dát je kompletne v réžii klientského systému. V automatizovanej testovacej bunke je služba DIS nasadená ako súčasť lokálnej infraštruktúry danej bunky. Cieľom je umožniť autentifikáciu, a tým dočasný samostatný chod aj v prípade krátkych výpadkov internetového pripojenia. Konfiguračne je však možné smerovať systém automatizovanej bunky aj na cloudové nasadenie služby. DIS je použitá vo fáze overenia identity pacienta, kedy analyzuje doklad z čítačky dokladov a snímky kamery umiestnenej nad displejom testovacej bunky. Použitím tejto služby je docielené, že potvrdenie o vykonaní testu bude vystavené naozaj na testovanú osobu.

3.2.4 Zabezpečenie osobných dát

Zabezpečenie osobných dát v informačnom systéme vychádza zo všeobecného nariadenia o ochrane údajov – GDPR:

Šifrovanie údajov - použitie šifrovania na ochranu osobných údajov v databáze. Šifrovanie zaručuje, že údaje sú nečitateľné pre neoprávnené osoby, ak by sa dostali do neoprávneného prístupu k údajom.

Prístupové práva a kontrola oprávnení - pravidelná kontrola prístupových práv a oprávnení používateľov k osobným údajom. To zahŕňa pridelenie oprávnení len tým osobám, ktoré majú oprávnenie pristupovať k osobným údajom a upravovať ich, a sledovanie a správu prístupových práv.

Bezpečnostné zálohy - pravidelné zálohovanie osobných údajov a implementácia opatrení na ich zabezpečené uloženie a obnovu.

Audity a monitorovanie - sledovanie a monitorovanie prístupov k osobným údajom v informačnom systéme, vrátane záznamov o prístupoch a aktivitách s osobnými údajmi.

Politiky a postupy ochrany údajov - vypracovanie a implementácia jasných politík a postupov týkajúcich sa ochrany údajov v rámci informačného systému.

Bezpečnostné aktualizácie a patche - pravidelná implementácia bezpečnostných aktualizácií a patchov.

3.2.5 Bezpečnosť

Bezpečnosť systému sa zabezpečuje na viacerých úrovniach, vrátane infraštruktúry aj aplikačnej vrstvy. Infraštruktúra, kde je systém nainštalovaný, je navrhnutá s cieľom dosiahnuť vysokú dostupnosť a minimalizovať riziko straty dát. Všetky rozhrania systému, vrátane používateľských a automatizovaných sú chránené pred neautorizovaným prístupom, a sú zabezpečené pomocou protokolu SSL.

3.2.5.1 Autentifikácia

Autentifikácia v informačnom systéme pre domácu karanténu zahŕňa proces identifikácie a overenia identity používateľa pri vstupe do systému. Na tento účel je využitý autentifikačný modul, ktorý používa štandardy OpenID Connect/OAuth 2.0. Všetky rozhrania systému sú zabezpečené SSL protokolom a autentifikácia používateľov je realizovaná prostredníctvom jednofaktorovej autentifikácie, kde sa používa prihlasovacie meno a heslo. Po úspešnej autentifikácii, autorizačný modul pridelí oprávnenia používateľovi na základe jeho používateľských rolí, ktoré určujú, na ktoré časti systému má používateľ prístup. Počas autentifikácie používateľ získa autentifikačný token s preddefinovanou životnosťou, ktorý umožní prístup k systému. Exspirácia autentifikačného tokenu je konfigurovateľná a závisí od bezpečnostných požiadaviek aplikácie. Pre automatickú testovaciu bunku zabezpečuje proces autentifikácie softvér DIS od spoločnosti Innovatrics, ktorý overuje používateľa pomocou biometrických údajov získaných kamerou bunky a informácií v občianskom preukaze.

3.2.5.2 Autorizácia

Autorizácia v systéme pre domácu karanténu je proces kontroly prístupu používateľa k funkcionalitám a dátam informačného systému. Pretože bezpečnosť a ochrana dát sú kľúčové pre tento systém, autorizácia bude založená na princípe riadenia prístupu na základe rolí (RBAC - Role-based access control). To znamená, že používateľ bude mať prístup len k tým funkciám a dátam, ktoré sú mu pridelené na základe jeho používateľskej role. Administrátor systému bude mať možnosť spravovať katalóg rolí a prideľovať ich používateľom podľa ich potrieb a oprávnení.

3.2.6 Monitorovanie systému

Monitoring informačného systému je rozdelený do troch základných úrovní, a to:

1. Systémová - systémová úroveň monitoringu je sledovanie dostupnosti informačného systému z pohľadu prostredia, kde je nainštalovaný (infraštruktúra, podporné doménové služby).
2. Aplikačná - aplikačná úroveň monitoringu je sledovanie dostupnosti komponentu informačného systému a niektorých ich vlastností.
3. Procesná (biznis) - procesná úroveň monitoringu je sledovanie stavu dôležitých biznis procesov informačného systému.

Na monitorovanie informačného systému bude použitý nástroj Nagios, ktorý je open-source softvér určený na monitorovanie IT infraštruktúry. Slúži na sledovanie stavu a výkonnosti hardvéru, softvéru, sietí a aplikácií, čím umožňuje prevádzkovateľom a správcom systémov rýchle odhaľovanie problémov a ich následné riešenie. Nagios dokáže monitorovať množstvo rôznych zariadení a služieb, ako sú napríklad servery, routery, switche, aplikácie a databázy. V prípade detekcie nejakého problému môže Nagios upozorniť správcov systémov rôznymi spôsobmi, ako sú e-mail, SMS, hlasové hovory, správy na displeji a podobne. Nagios teda zabezpečuje neustále sledovanie výkonu IT infraštruktúry a pomáha predchádzať výpadkom alebo iným problémom, ktoré by mohli ohroziť prevádzku informačného systému.

3.2.7 Podpora prevádzky systému

Prevádzka systému bude zabezpečená bežnými dostupnými možnosťami:

- informáciami na stránke – dokumentácia a FAQ,
- emailom,
- telefonickým kontaktom.

Pri kontaktovaní podpory prevádzky prostredníctvom emailu alebo telefonicky bude vytvorený tiket v tiketovacom nástroji OTRS. OTRS (Open-source Ticket Request System), ktorý slúži ako softvérový nástroj na správu a sledovanie požiadaviek zákazníkov a problémov v oblasti IT služieb, a podpory zákazníkov. Je to open-source systém, ktorý je vyvinutý v jazyku Perl a beží na webovom serveri. Nástroj poskytuje funkcie, ktoré umožňujú organizáciám vytvoriť centrálny bod pre správu požiadaviek a dotazov od zákazníkov. Systém umožňuje zákazníkom posielať požiadavky prostredníctvom e-mailu alebo webového rozhrania a sledovať stav svojich požiadaviek. Správcovia systému potom môžu tieto požiadavky prijímať, prideliť ich príslušným zamestnancom a sledovať ich riešenie. Nástroj obsahuje tiež ďalšie užitočné funkcie, ako je správa znalostnej bázy, prehľady a štatistiky, automatická eskalácia, sledovanie časov reakcie a ďalšie. Systém je navrhnutý tak, aby bol ľahko konfigurovateľný a prispôsobiteľný potrebám rôznych organizácií.

Personálne obsadenie podpory prevádzky je rozdelené do troch základných úrovní: L1, L2 a L3:

- **Úroveň L1** – odborníci sa zaoberajú základnými spotrebiteľskými problémami a majú všeobecné znalosti o produkte a službách. Získavajú informácie o zákazníkoch, analyzujú príznaky a identifikujú základné problémy.
- **Úroveň L2** – odborníci majú rozsiahlejšie skúsenosti a znalosti a môžu pomôcť odborníkom na úrovni L1 pri riešení základných technických problémov. Skúmajú existujúce problémy a hľadajú známe riešenia pre zložitejšie problémy.
- **Úroveň L3** – špecialisti sa zaoberajú najnáročnejšími problémami a sú odborníkmi vo svojom odbore. Niekedy pomáhajú odborníkom na úrovni L1 aj L2. Okrem toho skúmajú a vyvíjajú riešenia pre nové alebo neznáme problémy.

3.3 Prototyp pre Dezinfekčného robota

3.3.1 Opis funkčnosti

Informačný systém pre dezinfekčného robota sa skladá z troch kľúčových častí:

1. Cloudová služba.
2. Interný systém pre ovládanie dezinfekčného robota.
3. Mobilná aplikácia pre ovládanie dezinfekčného robota.

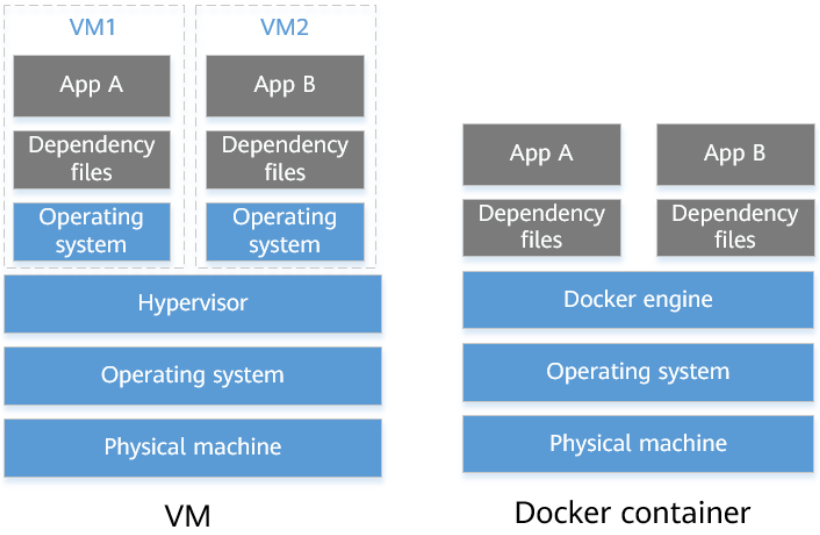
Cloudová služba je budovaná ako platforma pre ukladanie nameraných údajov, ktoré môžu byť získané z rôznych zdrojov. Cloudová služba má rozhranie pre používateľov prostredníctvom webového rozhrania a rozhranie pre automatizovanú komunikáciu prostredníctvom Rest API – určené pre synchronizáciu so systémom pre ovládanie testovacej bunky.

Interný systém pre ovládanie dezinfekčného robota je vnorený systém prevádzkovaný priamo v automatickej testovacej bunke. Primárnou úlohou systému je ovládanie dezinfekčného robota a synchronizácia nameraných dát do cloudovej služby.

Mobilná aplikácia pre ovládanie dezinfekčného robota je mobilná aplikácia, pomocou ktorej je možné vzdialene ovládať dezinfekčného robota.

3.3.2 Opis architektúry nasadenia

Cloudová služba pre dezinfekčného robota bola navrhnutá tak, aby bola prevádzkovaná v docker kontajneroch, čo predstavuje najmodernejší spôsob prevádzkovania informačných systémov v dnešnej dobe. V rámci tohto projektu bol jeden z cieľov vyskúšať tento spôsob prevádzky, aby sme využili jeho výhody oproti tradičnej prevádzke vo virtuálnych serveroch založenej na virtualizácii.



Obrázok 9 Docker kontajner vs Virtuálne servery (VM)

Na obrázku je vidieť porovnanie dvoch najpoužívanejších možnosti prevádzkovania informačných systémov, a to:

1. VM - Virtualizácia.
2. Docker container – Kontajnerizácia.

Virtualizácia umožňuje prevádzkovať viac nezávislých virtuálnych serverov na jednom fyzickom serveri. Prevádzka informačných systémov vo virtuálnom prostredí je výhodná z bezpečnostného a ekonomického hľadiska v porovnaní s prevádzkou na fyzickom prostredí.

Kontajnerizácia je forma virtualizácie, ktorá využíva menšie množstvo systémových zdrojov. Aplikácie sú prevádzkované v docker kontajneroch, ktoré obsahujú minimálne systémové prostriedky potrebné pre ich prevádzku. Prevádzka pomocou docker kontajnerov je výhodná najmä z hľadiska jednoduchšieho nasadenia aplikácií, závislostí aplikácií na operačnom systéme a aktualizácií aplikácií a prostredia.

Tabuľka 5 Zoznam kontajnerov

Názov kontajnera	Image	Popis
ld-web	nginx:latest	Kontajner pre webový server
ld-api	mcr.microsoft.com/dotnet/aspnet:5.0	Kontajner pre API
ld-worker	mcr.microsoft.com/dotnet/runtime:6.0	Kontajner pre aplikačný server
ld-pgadmin	dpage/pgadmin4:6	Kontajner pre webové rozhrania databázového servera
ld-postgres	postgres:13.4	Kontajner pre databázový server

3.3.3 Prostredia

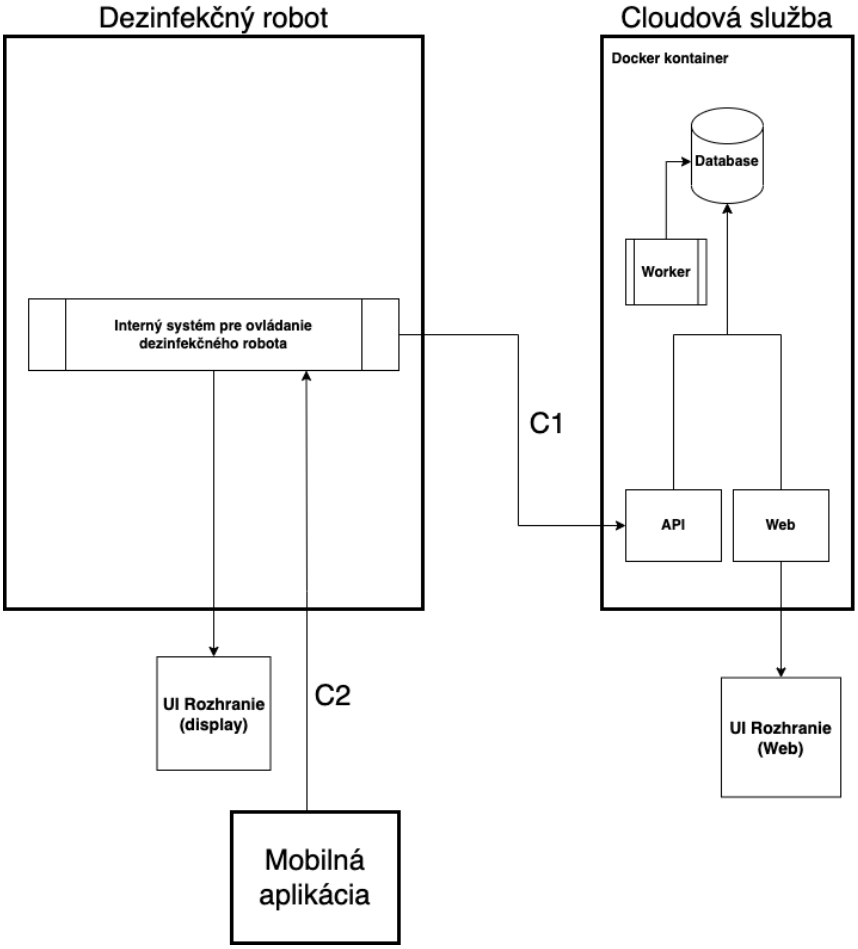
Cloudová služba pre dezinfekčného robota je prevádzkovaná v jednom vývojovom prostredí na infraštruktúre spoločnosti sféra, a.s.

Tabuľka 6 Zoznam prostredí

Názov	Doména	Infraštruktúra
Vývojové prostredie	lifedefender.sk	HW spoločnosti sféra, a.s.

3.3.4 Integrácia na jednotne cloudové úložisko

Na nasledujúcom obrázku je vidieť základnú logickú architektúru systému.



Obrázok 10 Logická architektúra systému

Cloudová služba sa skladná z nasledujúcich logických komponentov:

- Web,
- API,
- Worker,
- Databáza.

3.3.4.1 Typy integrácií

Dezinfekčný robot a cloudová služba sú primárne integrované pre synchronizáciu nameraných dát. Okrem toho je robot tiež integrovaný s mobilnou aplikáciou.

Tabuľka 7 Zoznam integrácií

Integrácia	Systém	Popis
C1	Dezinfekčný robot -> cloudová služba	HTTP Rest API, slúži na synchronizáciu dát
C2	Mobilná aplikácia -> cloudová služba	Bluetooth, služba pre ovládanie robota

3.3.4.2 Komponenty systému

- **Web (UI rozhranie)** - komponent, ktorý zabezpečuje používateľské rozhranie prostredníctvom webového rozhrania.
- **API** - komponent, ktorý zabezpečuje REST API služby pre synchronizáciu dát medzi automatickou testovacou bunkou a cloudovou službou.
- **Worker** - komponent, ktorý zabezpečuje vykonávanie plánovaných úloh.
- **Database** - komponent, ktorý zabezpečuje perzistentné úložisko dát – relačná databáza.

3.3.5 Bezpečnosť

Bezpečnosť systému (cloudová služba) je riešená na úrovni infraštruktúry, ale aj na úrovni aplikačnej vrstvy. Infraštruktúra, do ktorej je systém inštalovaný, je budovaný s ohľadom na vysokú dostupnosť a s požiadavkou na minimalizáciu rizika straty dát. Všetky používateľské aj automatizované rozhrania systému sú zabezpečené voči neautorizovanému vstupu, a sú publikované prostredníctvom SSL protokolu.

3.3.5.1 Autentifikácia

Autentifikácia je proces identifikácie a overenia identity používateľa pri vstupe do informačného systému. V systéme pre domácu karanténu je autentifikácia používateľov zabezpečená pomocou autentifikačného modulu, ktorý používa štandardy OpenID Connect/OAuth 2.0. Voči tomuto autentifikačnému modulu sa budú overovať používatelia pri vstupe do všetkých rozhraní systému. Autentifikácia je jednofaktorová s použitím prihlasovacieho mena a hesla.

Pri vstupe používateľa do systému autentifikačný overí zadané meno a heslo používateľa. Po úspešnom overení autorizačný modul pridelí oprávnenia používateľovi na základe jeho používateľských rolí, v opačnom prípade autentifikačný modul zabráni používateľovi vstup do systému. Ak je používateľ úspešne overený, môže pristupovať k jednotlivým častiam systému na základe pridelených používateľských rolí.

Počas autentifikácie používateľ systému získa autentifikačný token, ktorý má preddefinovanú životnosť. Po exspirácii autentifikačného tokenu (dĺžka exspirácie je konfigurovateľná, jej dĺžka závisí od bezpečnostných požiadaviek aplikácie) sa používateľ musí opätovne prihlásiť.

3.3.5.2 Autorizácia

Autorizácia je proces riadenia prístupu používateľa k funkčnostiam a dátam informačného systému. V systéme pre domácu karanténu bude proces riadenia prístupov založený na rolách RBAC (Role-based access control), čo znamená, že používateľ získa oprávnenia prostredníctvom používateľskej roly. Administrátor pre správu používateľov bude mať možnosť spravovať používateľské roly (katalóg rolí) a prideľovať ich používateľom.

3.3.6 Monitorovanie systému

Monitoring informačného systému je rozdelený do troch základných úrovní, a to:

1. Systémová - systémová úroveň monitoringu je sledovanie dostupnosti informačného systému z pohľadu prostredia, kde je nainštalovaný (infraštruktúra, podporné doménové služby).
2. Aplikačná - aplikačná úroveň monitoringu je sledovanie dostupnosti komponentu informačného systému a niektorých ich vlastností.
3. Procesná (biznis) - procesná úroveň monitoringu je sledovanie stavu dôležitých biznis procesov informačného systému.

Na monitorovanie informačného systému bude použitý nástroj Nagios, ktorý je open-source softvér určený na monitorovanie IT infraštruktúry. Slúži na sledovanie stavu a výkonnosti hardvéru, softvéru, sietí a aplikácií, čím umožňuje prevádzkovateľom a správcom systémov rýchle odhaľovanie problémov a ich následné riešenie. Nagios dokáže monitorovať množstvo rôznych zariadení a služieb, ako sú napríklad servery, routery, switche, aplikácie a databázy. V prípade detekcie nejakého problému môže Nagios upozorniť správcov systémov rôznymi spôsobmi, ako sú e-mail, SMS, hlasové hovory, správy na displeji a podobne. Nagios teda zabezpečuje neustále sledovanie výkonu IT infraštruktúry a pomáha predchádzať výpadkom alebo iným problémom, ktoré by mohli ohroziť prevádzku informačného systému.

3.3.7 Podpora prevádzky systému

Prevádzka systému bude zabezpečená bežnými dostupnými možnosťami:

- informáciami na stránke – dokumentácia a FAQ,
- emailom,
- telefonickým kontaktom.

Pri kontaktovaní podpory prevádzky prostredníctvom emailu alebo telefonicky bude vytvorený tiket v tiketovacom nástroji OTRS. OTRS (Open-source Ticket Request System), ktorý slúži ako softvérový nástroj na správu a sledovanie požiadaviek zákazníkov a problémov v oblasti IT služieb a podpory zákazníkov. Je to open-source systém, ktorý je vyvinutý v jazyku Perl a beží na webovom serveri. Nástroj poskytuje funkcie, ktoré umožňujú organizáciám vytvoriť centrálny bod pre správu požiadaviek a dotazov od zákazníkov. Systém umožňuje zákazníkom posielať požiadavky prostredníctvom e-mailu alebo webového rozhrania a sledovať stav svojich požiadaviek. Správcovia systému potom môžu tieto požiadavky prijímať, prideliť ich príslušným zamestnancom a sledovať ich riešenie. Nástroj obsahuje tiež ďalšie užitočné funkcie, ako je správa znalostnej bázy, prehľady a štatistiky, automatická eskalácia, sledovanie časov reakcie a ďalšie. Systém je navrhnutý tak, aby bol ľahko konfigurovateľný a prispôsobiteľný potrebám rôznych organizácií.

Personálne obsadenie podpory prevádzky je rozdelené do troch základných úrovní: L1, L2 a L3:

- **Úroveň L1** – odborníci sa zaoberajú základnými spotrebiteľskými problémami a majú všeobecné znalosti o produkte a službách. Získavajú informácie o zákazníkoch, analyzujú príznaky a identifikujú základné problémy.
- **Úroveň L2** – odborníci majú rozsiahlejšie skúsenosti a znalosti a môžu pomôcť odborníkom na úrovni L1 pri riešení základných technických problémov. Skúmajú existujúce problémy a hľadajú známe riešenia pre zložitejšie problémy.
- **Úroveň L3** – špecialisti sa zaoberajú najnáročnejšími problémami a sú odborníkmi vo svojom odbore. Niekedy pomáhajú odborníkom na úrovni L1 aj L2. Okrem toho skúmajú a vyvíjajú riešenia pre nové alebo neznáme problémy.

4 EXPERIMENTÁLNY VÝVOJ PROTOTYPU MODULU POKROČILEJ ANALÝZY A VIZUALIZÁCIE DÁT

4.1 Prototyp modulu vizualizácie dát

4.1.1 Kibana

Kibana je vizualizačný a analytický nástroj, ktorý je súčasťou Elastic Stack (dôraznej platformy na spracovanie a analýzu údajov). Je navrhnutý na vizualizáciu, vyhľadávanie a analyzovanie údajov zo zdrojov, ako je Elasticsearch. Kibana poskytuje používateľské rozhranie, cez ktoré môžete vytvárať interaktívne grafy, tabuľky, dashboardy a správy, aby ste vizuálne preskúmali a porozumeli svojim údajom. Jeho funkcie zahŕňajú:

Vizualizácie: Kibana umožňuje vytvárať rôzne typy vizualizácií, vrátane stĺpcových a čiarových grafov, koláčových diagramov, histogramov, mapových vizualizácií a ďalších. Tieto vizualizácie pomáhajú objaviť vzorce, trendy a závislosti vo vašich údajoch.

Dashboardy: Môžete vytvárať vlastné dashboardy, kde kombinujete rôzne vizualizácie a správy na prehľadné zobrazenie dôležitých údajov na jednom mieste. Tieto dashboardy môžete ľahko prispôsobiť a zdieľať s ostatnými používateľmi.

Hľadanie a filtrovanie: Kibana poskytuje možnosti vyhľadávania a filtrovania údajov v reálnom čase. Môžete definovať komplexné dotazy na vyhľadávanie a zúžiť zobrazené údaje podľa rôznych kritérií.

Monitorovanie a upozornenia: Kibana umožňuje sledovať rôzne metriky a ukazovatele v reálnom čase a nastaviť upozornenia na základe definovaných podmienok. Tieto upozornenia vám umožňujú reagovať na kritické udalosti a problémy v systéme.

Integrácia s Elasticsearch: Kibana je úzko integrovaná s Elasticsearch, čo umožňuje rýchle spracovanie a vyhľadávanie veľkého množstva údajov. Môžete používať Elasticsearch ako zdroj údajov pre Kibanu a využívať jeho silné vyhľadávacie a analytické schopnosti.

Nástroj Kibana spolu s databázou Elasticsearch je prevádzkovaná v Dockery.

Nami preferovaná kategória vizualizácií predstavuje vizualizáciu relačných dát, teda dát, medzi ktorými existuje nejaký vzťah. Úlohou vizualizácie je prezentovať vzťah a jeho vlastnosti medzi týmito dátami. Vzťah však iba ťažko vizualizovať pomocou už spomínaných vizualizácií. Na zobrazenie vzťahu sú vhodné diagramy ako relačný diagram, tetivový diagram alebo mriežka, ktorá prezentuje existenciu vzťahu medzi dvomi hodnotami.

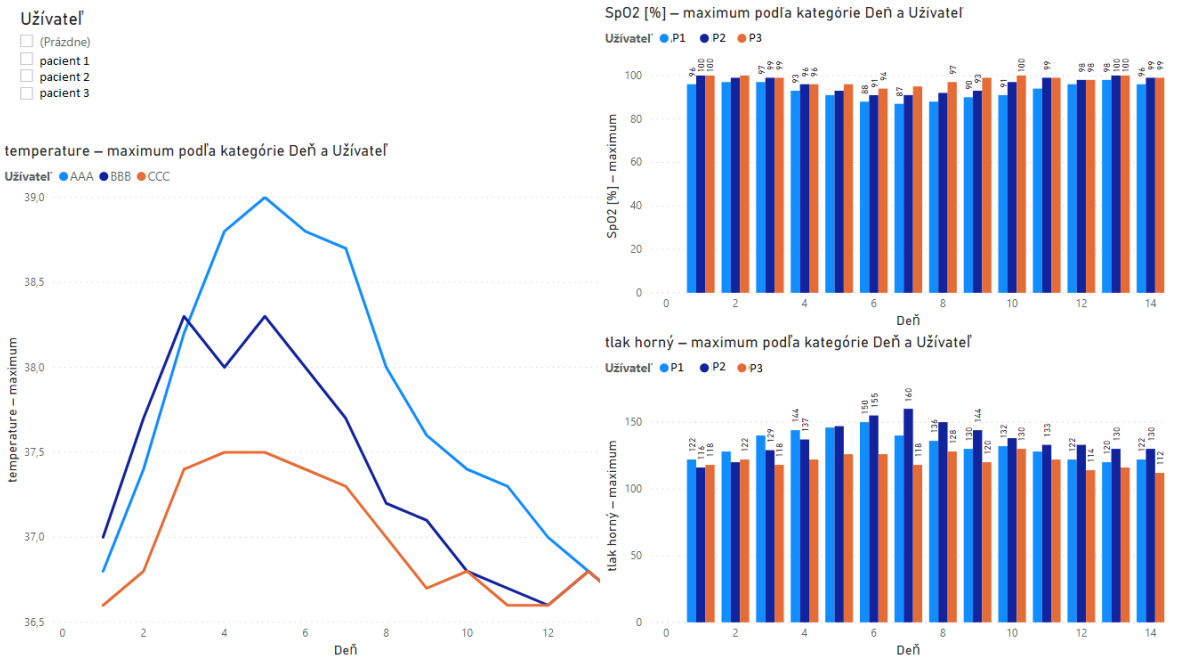
Najznámejším a najrozsiahlejším rozšírením Kibany je Kibi[1] . V tomto prípade nejde v pravom zmysle slova o plugin nástroja Kibana ale o open-source nástroj zameraný na dátovú inteligenciu, ktorý je postavený na Kibane. Okrem štandardných vizualizácií nástroja Kibana poskytuje aj nové druhy vizualizácií ako napríklad word cloud alebo tzv. radarový diagram. Rovnako prináša možnosť definovať vzťahy medzi dátami uloženými naprieč viacerými indexami a podobne.[2]

4.1.1.1 Možnosti vizualizácie relačných dát

Vzťahové alebo relačné dáta sú dáta, ktoré nesú informáciu o určitom vzťahu. Tieto vzťahy medzi dátami môžu byť dvojakého typu – hierarchické a rovnocenné. Hierarchické dáta popisujú vzťah, v

[1] Kibi User Guide: Introduction. Siren Solutions [online]. [cit. 2016-03-30]. Dostupné z: http://siren.solutions/kibi/docs/current/introduction.html
[2] Kibi User Guide: Introduction. Siren Solutions [online]. [cit. 2016-03-30]. Dostupné z: http://siren.solutions/kibi/docs/current/introduction.html

ktorom je jeden dokument (záznam, objekt) nadradený nad iný. Môže ísť napríklad o dáta nesúce informáciu o organizačnej štruktúre podniku, rodokmeni, pacientov patriacich pod určitého lekára/poisťovňu a podobne. Na vizualizáciu takéhoto druhu relačných dát je najvhodnejší stromový diagram zložený z uzlov a hrán medzi nimi. Takýto graf má hlavný uzol - koreň, prípadne viac koreňov stromu, ktorý predstavuje vrchol organizačnej štruktúry a čím nižšie sa uzol v grafe nachádza, tým je nižšie aj v danej organizačnej štruktúre. Hrany medzi uzlami predstavujú priamy vzťah, napríklad otec-syn. Primárne sa zaoberáme druhou kategóriou relačných dát a tou sú dáta nesúce informáciu o rovnocennom vzťahu dvoch objektov. S obrovským počtom evidovaných pacientov a s rastúcim množstvom dát, ktoré jednotlivé organizácie uchovávajú vzrastá aj potreba takéto dáta vizualizovať. Takýto rovnocenný vzťah medzi objektami sa dá najjednoduchšie predstaviť ako sieť zložená, rovnako ako stromový diagram, z uzlov a hrán. Uzly sú ale rovnocenné a hrany zobrazujú vzťah medzi nimi. O uchovávaní relačných dát v Elasticsearch sme pojednávali vo výstupnom dokumente míľnika č. 3 projektu Life Defender – Ochranca života. Takéto dáta sa dajú vizualizovať viacerými druhmi vizualizácií.[3]



Obrázok 11 Vizualizácia nameraných dát

4.2 Problematika zberu, analýzy a vyhodnocovania symptómov pomocou analytických nástrojov s využitím umelej inteligencie

4.2.1 Prehľad štatistických algoritmov pre analýzu a predikciu zdravotníckych dát

V predchádzajúcom míľniku analýzy dát sme identifikovali vhodné algoritmy pre analýzu a predikciu zdravotníckych dát. Tento proces vyústil do formalizovaného výstupu, ktorý určil najlepšie

[3] https://is.muni.cz/th/l7oyt/visualization_thesis_racek.pdf

prognostické metódy a štatistické nástroje pre predikovanie vývoja viacerých časových radov. Tieto boli zvolené na základe ich dostupnosti v dátových štruktúrach súčasného monitorovacieho systému zdravotníckych ukazovateľov. Cieľom tohto procesu bolo nájsť metodologicky vhodné prístupy pre podporu inteligentných zdravotných systémov, ktoré by mohli posilniť alebo nahradiť analytické funkcie súčasných systémov.

V rámci procesu analýzy zdravotníckych dát bola potrebná úvaha nad niekoľkými kľúčovými aspektmi. Prvým z nich boli charakteristiky súčasného systému poskytovania, evidencie a využitia zdravotníckych údajov. Tento faktor bol potrebné zohľadniť, aby bolo možné analyzovať a predikovať vývoj zdravotného stavu pacientov s ohľadom na ich záznamy v súčasnom systéme.

Ďalším aspektom bola hodnota metadát získaných z analýzy. Bolo dôležité zohľadniť, akú hodnotu a aký vplyv budú mať tieto metadáta na výsledky analýz a predikcií.

Bezpečnosť z pohľadu anonymity dát bola tiež veľmi dôležitým aspektom. Bolo potrebné zabezpečiť, aby dáta pacientov boli chránené pred neoprávneným prístupom a aby ich anonymity nebola ohrozená.

Ďalším dôležitým faktorom bolo zabezpečiť vhodnú interpretáciu dát. To znamená, že výstupy analýz a predikcií museli byť interpretovateľné pre zdravotníckych pracovníkov a mali by byť schopní použiť tieto výsledky na zlepšenie zdravotnej starostlivosti.

Jednoduchosť implementácie do súčasných systémov bola ďalším dôležitým aspektom. Aby bola táto metóda použiteľná v praxi, bolo potrebné, aby bola jednoduchá na implementáciu a aby sa mohla integrovať do súčasných systémov.

Posledným faktorom, ktorý bol braný do úvahy, bola presnosť algoritmov pre potreby analýz a predikcií. Pri analýze vhodnosti jednotlivých algoritmov boli brané do úvahy dostupné diagnostické zdravotnícke parametre, ako napríklad tlak, vek a BMI. Tieto parametre boli potrebné zohľadniť, aby bolo možné určiť presnosť algoritmov a ich vhodnosť pre potreby analýz a predikcií.

4.2.1.1 Predspracovanie dát

Pred analýzou zdravotníckych údajov je potrebné dáta zo systému skontrolovať a upraviť do správneho formátu, ktorý uľahčí ich interpretáciu a manipuláciu. Pri testovaní rôznych metód predikcie sa zistilo, že najvhodnejším vstupným profilom pre tieto metódy sú časové rady s dlhodobo stabilným tokom dát, ktoré majú štandardné hodnoty a bez signifikantných posunov. Profily s nepravidelným charakterom hodnôt a s výraznými zmenami alebo skokmi v hodnotách spomaľujú tréning systému a môžu dokonca znemožniť štatisticky robustnú predikciu. Preto je dôležité dáta správne upraviť a vybrať vhodné profily pre tréning a predikciu zdravotníckych údajov.

Pri tvorbe algoritmických modelov je nevyhnutné mať na pamäti primárnu kontrolu kvality získaných dát. Tento proces zahŕňa detekciu a odstránenie akýchkoľvek evidentných chýb, ako napríklad nezvyčajne vysokú telesnú teplotu (100 °C), podozrivo vysoké hodnoty krvného tlaku alebo BMI. Existujú dva spôsoby ako takéto chyby ošetriť. Ak sa model stavia na citlivých dátach, kde minimálna chybovosť je nevyhnutná, alebo na malých vzorkách dát, je potrebné tieto anomálie vyhľadať a jednotlivo ošetriť. V prípade, že model je postavený na dostatočne veľkých vzorkách dát, je možné takéto extrémne hodnoty ošetriť stanovením hodnotovej hranice pomocou deskriptívnych štatistík, ako sú priemer, medián, maximálna a minimálna hodnota a štandardná odchýlka. Pre filtrovanie neštandardných zmien tvaru dát sa používa validačná funkcia.

Pre účely kvalitnej analýzy dát je kritické, aby vstupné dáta mali podobný rozsah. To znamená, že vstupné hodnoty by nemali byť príliš rozptýlené, aby bolo možné ich ľahšie porovnávať a analyzovať. Štandardným postupom je preto preškálovanie každého vstupu na rozsah od 0 do 1, alebo tak, aby mal priemer 0 a štandardnú odchýlku 1. Tento proces sa nazýva normalizácia údajov a je bežne používaný pri predspracovaní údajov.

Ďalšou dôležitou úpravou dát je definovanie rozdielu medzi nulovou hodnotou a prázdnej hodnotou. Nulová hodnota znamená meranie hodnoty, ktorá je rovná 0, zatiaľ čo prázdna hodnota znamená neplatnú hodnotu, napríklad keď došlo k výpadku merania.

Dôležitou podmienkou pre validitu a spoľahlivosť záverov je dostatočná veľkosť vzorky. To znamená, že musí byť dostatočný počet dátových profilov, ich dĺžka a frekvencia merania bodových hodnôt. V niektorých prípadoch môže byť nutné extrapolovať získané dáta, ak je vzorka príliš malá na získanie štatisticky významných prípadov. Takto rozšírená dátová vzorka však musí spĺňať prísne požiadavky na robustnosť, internú konzistenciu a dynamiku v čase. Veľkosť vzorky ovplyvňuje silu testovania štatistických hypotéz.

Tréning systémov pre odhad a predikcie zdravotného stavu pacientov môže byť realizovaný buď off-line alebo on-line. Off-line tréning sa vykonáva na tzv. off-line vzorke dát, ktorá predstavuje dátovú sadu, na ktorej sa tréningový model vyvíja a testuje. Tento prístup umožňuje intenzívnejšie testovanie modelu, ale existuje riziko, že systém nezahrnie všetky signály ovplyvňujúce predikciu v reálnom dátovom toku, pretože tieto signály môžu chýbať v off-line vzorke dát.

On-line tréning sa vykonáva na tréningových dátach z aktívneho dátového toku, čo znamená, že systém zahrnie všetky signály ovplyvňujúce predikciu v reálnom čase. Tento prístup je intenzívnejší a zabezpečuje, že všetky existujúce signály v aktívne bežiacich dátových profiloch sa zahrnú do tvorby modelu. Avšak, prvotná chybovosť modelu môže mať priamy dopad na bežiaci systém a množstvo irelevantných informácií v dátovom toku a profily s dlhším časom vyhodnotenia môžu spomaliť výpočet predikcie.

V zdravotníckom kontexte sa preto používa dvojkroková analýza, ktorá zabezpečuje, že tréning predikčných modelov z on-line dátových profilov sa vykonáva na ich uložených verziách a bodových meraniach. Tieto hodnoty sa potom aktualizujú v potrebnom časovom horizonte pred výpočtom predikcií (napr. denne). Tento prístup je kompromisom medzi intenzitou tréningu a rizikom dopadu chýb na bežiaci systém.

V nasledujúcej časti sa budú popisovať vybrané predikčné modely v rámci projektu odhadu a predikcií zdravotného stavu pacientov.

4.2.1.2 Validácia modelu

Validácia modelu je kľúčový proces pri štatistickom modelovaní a strojovom učení. Je to proces overovania, či vytvorený model dobre zodpovedá skutočnosti a či ho možno použiť na predpovedanie budúcich výsledkov. Validácia modelu má veľký význam v oblasti zdravotníctva, finančníctva, marketingu, predpovedaní počasia a mnohých ďalších oblastiach.

Existuje niekoľko spôsobov, ako overiť, či je model validný. Jedným z najjednoduchších spôsobov je použitie metódy oddeľovania dát, ktorá spočíva v rozdelení dát na trénovacie a testovacie dáta. Trénovacie dáta sa používajú na trénovanie modelu a testovacie dáta na overenie, či model dobre funguje na nových dátach. Tento postup pomáha minimalizovať efekt overenia na trénovacie dáta a zaručuje, že model je schopný predpovedať na nových dátach.

Ďalším spôsobom overenia modelu je použitie krížovej validácie. Táto metóda rozdeľuje dáta na k-foliové oddiely a postupne používa jeden oddiel na testovanie a zvyšné oddiely na trénovanie. Tento proces sa opakuje pre všetky oddiely a výsledky sa kombinujú na vytvorenie konečného modelu. Krížová validácia je dôležitá pre overenie, či model je stabilný a či je výkonný na rôznych dátových vzorkách.

Ďalší spôsob validácie je použitie výberu atribútov. Táto metóda sa používa na odstránenie nezmyselných alebo nepodstatných atribútov, ktoré by mohli narušiť výkonnosť modelu. Výber atribútov sa robí pomocou rôznych algoritmov, ktoré určujú, ktoré atribúty sú najdôležitejšie pre predpovedanie.

Validácia modelu pri štatistickom modelovaní a strojovom učení je dôležitá pre dosiahnutie presných a spoľahlivých predpovedí. Použitie viacerých metód a techník pomáha minimalizovať chyby a zaručuje, že model je validný a schopný predpovedať na nových dátach.

Validácia modelu je dôležitý krok v procese modelovania a učenia strojov. Jedná sa o overenie, ako dobre model predikuje na nových, neznámych dátach, a to bez toho, aby sme upravovali jeho parametre na základe výsledkov na trénovacích dátach. Ak model výrazne zaostáva v predikcii na nových dátach, hovoríme o pretrénovaní.

Aby sme validovali model, musíme oddeliť trénovacie dáta od testovacích dát. Trénovacie dáta používame na trénovanie modelu a testovacie dáta používame na overenie jeho výkonnosti. Častým spôsobom validácie modelu je použitie tzv. K-fold cross-validácie. V tomto prípade rozdelíme trénovaciu vzorku na K častí a v každom kroku jednu z častí použijeme ako testovaciu vzorku a ostatné časti ako trénovacie. Model natrénujeme na trénovacích dátach a overíme jeho výkonnosť na testovacích dátach. Tento proces opakujeme K-krát a výsledný odhad výkonnosti modelu je priemer z K-overení.

Je dôležité, aby sme pri validácii modelu používali nezávislé testovacie dáta, aby sme získali správny odhad chybovosti modelu na nových dátach. Preto je nutné, aby sme si vyčlenili testovaciu vzorku z nezávislých dát, ktoré nie sú súčasťou trénovacích dát. Ak model prejde testovacou fázou, môžeme ho použiť na predikciu na nových dátach s väčšou istotou, že jeho výsledky budú spoľahlivé.

4.2.2 Metódy strojového učenia a ich efektivita v kontexte modelovania zdravotníckych dát

Metódy strojového učenia (machine learning) sa v súčasnosti stávajú neoddeliteľnou súčasťou výskumu a modelovania zdravotníckych dát. Jednou z hlavných výhod strojového učenia je jeho schopnosť naučiť sa rozpoznávať vzory a vytvárať z nich modely, ktoré môžu byť použité na predikciu, klasifikáciu a ďalšie úlohy. Tieto modely môžu byť využité pre zlepšenie diagnostiky, prognózovania a rôznych iných aplikácií v oblasti zdravotníctva.

Existuje mnoho rôznych metód strojového učenia, ktoré sa využívajú v modelovaní zdravotníckych dát. Medzi najpoužívanejšie patria:

- Učenie s učiteľom - táto metóda sa používa pri trénovaní modelu s označenými dátami, kde každá vzorka obsahuje označenie výstupu (label). Model sa potom učí identifikovať vzťah medzi vstupom a označením výstupu a potom môže byť použitý na predikciu výstupu pre nové vstupy.
- Učenie bez učiteľa - táto metóda sa používa na trénovanie modelov, ktoré nemajú označené dáta. Model sa snaží nájsť vzory a skryté zákonitosti v dátach a potom môže byť použitý na klastrovanie dát a identifikáciu podobných vzorov.
- Spätnoväzobné učenie - táto metóda sa používa na trénovanie modelov, ktoré sú schopné učiť sa na základe interakcie s prostredím. Model sa učí prostredníctvom pokusov a omylov, ktoré sú odmenené alebo trestané v závislosti od výsledku.
- Hlboké učenie - táto metóda používa viacvrstvové neurónové siete na učenie zložitejších vzťahov medzi vstupom a výstupom. Táto metóda sa používa najmä v oblastiach ako je spracovanie obrazu, rozpoznávanie reči a analýza textu.
- Strojové učenie je analytický prístup, pri ktorom sa vytvára matematický model na predikciu cieľovej premennej na základe iných vstupných premenných. V zdravotníctve sa tento prístup používa na analýzu a predikciu zdravotného stavu pacientov.

Na začiatku procesu strojového učenia sa zhromaždia dostupné zdravotnícke údaje, ktoré sa organizujú vo forme databázy premenných alebo faktorov. Tieto údaje sú vstupom do trénovania modelu, ktorý sa snaží predikovať cieľovú premennú na základe nezávislých premenných. Napríklad by

sme mohli použiť trénovacie dáta na vytvorenie modelu, ktorý predikuje, ktorí pacienti majú zvýšené riziko výskytu cukrovky na základe ich krvného tlaku, veku, BMI a úrovne inzulínu v krvi.

Proces trénovania modelu zahŕňa automatizované a dynamické nastavenie parametrov modelu tak, aby najlepšie zodpovedali profilom zdravotných údajov. Vytvorený model potom umožňuje vykonávať predikcie založené na regresných alebo klasifikačných algoritmoch. Regresné algoritmy sa používajú na predikciu spojitej premenné, napríklad rizika vývoja cukrovky u pacientov. Klasifikačné algoritmy sa používajú na predikciu diskrétnej premenné, napríklad diagnostikovanie ochorenia.

Celkovým cieľom strojového učenia v zdravotníctve je poskytnúť zdravotníckym pracovníkom nástroje na lepšie predikovanie rizík a diagnostikovanie ochorení na základe dostupných zdravotníckych údajov. Týmto spôsobom môžu byť rýchlejšie a presnejšie stanovené diagnózy, zlepšené prognózy a lepšie rozhodnutia o liečbe pacientov.

4.2.2.1 Lineárna regresia

Pri modeloch založených na lineárnej regresii sa predikuje výstupná premenná (tiež nazývaná cieľová premenná) ako lineárna kombinácia viacerých vstupných premenných. Tieto vstupné premenné môžu zahŕňať predchádzajúce hodnoty predikovanej veličiny, vonkajšie premenné, iné známe časové rady a ďalšie relevantné faktory.

Pri tejto metóde sú identifikované modely neznámych parametrov pomocou lineárnych funkcií. Tieto parametre sa nazývajú vstupné atribúty alebo charakteristiky (anglicky features) a ich úlohou je predikovať sériu cieľových premenných v priestore všetkých možných funkcií. Cieľom modelovania je nájsť a popísať takú hypotézu, ktorá najlepšie predikuje pohyb cieľovej premennej.

Lineárna regresia sa snaží nájsť lineárnu vzťah medzi vstupnými premennými a cieľovou premennou. Model je definovaný pomocou váhových koeficientov, ktoré sa priradzujú každej vstupnej premennej a určujú jej prínos k predikcii výstupnej premennej. Cieľom je minimalizovať chybovú funkciu, ktorá meria rozdiel medzi skutočnými hodnotami cieľovej premennej a hodnotami predikovanými modelom.

Po trénovaní lineárneho regresného modelu je možné použiť ho na predikciu hodnôt cieľovej premennej pre nové vstupy. Model môže poskytnúť odhadované hodnoty, predikcie budúcich udalostí alebo identifikovať vzory a trendy v dátach.

Lineárna regresia je jednou z najbežnejších a najjednoduchších metód strojového učenia. Je široko využívaná v rôznych oblastiach, vrátane ekonómie, financií, medicíny, sociálnej vedy a mnohých ďalších. Je vhodná, ak existuje lineárny vzťah medzi vstupnými a cieľovými premennými, a poskytuje interpretovateľné výsledky, ktoré môžu pomôcť porozumieť a analyzovať dáta.

Pri hodnotení efektivity vybranej hypotézy v strojovom učení sa často využívajú chybové funkcie. Tieto funkcie merajú rozdiel medzi skutočnými hodnotami a hodnotami predikovanými modelom. Jednou z najjednoduchších chybových funkcií je napríklad štvorec rozdielu medzi reálnou hodnotou z trénovacích dát a predikovanou hodnotou. Cieľom je nájsť taký model, ktorý generuje čo najnižšiu sumu funkčných hodnôt chybovej funkcie na vzorke všetkých vstupných atribútov.

Existuje aj ďalšia alternatíva v podobe chybových funkcií založených na lineárnom programovaní. Tieto funkcie sú síce komputačne náročnejšie, ale majú výhodu v prípade, keď distribúcia dát nevyplýva z normálneho rozdelenia. Lineárne programovanie umožňuje definovať optimalizačný problém so zadanými lineárnymi obmedzeniami a cieľovou funkciou. V prípade hodnotenia modelov na základe lineárneho programovania je cieľom minimalizovať chybu predikcie pri zachovaní daných obmedzení.

Výber správnej chybovej funkcie závisí od povahy dát a konkrétneho problému, ktorý sa rieši. Ak je distribúcia dát normálna, jednoduché chybové funkcie, ako je štvorec rozdielu, môžu byť efektívne a ľahko interpretovateľné. Avšak ak distribúcia dát nie je normálna, môže byť vhodnejšie použiť chybové funkcie založené na lineárnom programovaní, ktoré sa dokážu prispôsobiť rôznym distribúciám a

obmedzeniam. Výber správnej chybovej funkcie je dôležitým krokom pri modelovaní a hodnotení efektivity modelu. Správna chybová funkcia umožňuje kvantifikovať rozdiel medzi skutočnými a predikovanými hodnotami a slúži ako meradlo úspešnosti modelu.

Lineárna regresia je populárnou metódou modelovania, pretože je časovo efektívna a umožňuje distribuované trénovanie, čo je výhodné pri veľkých množstvách dát. Avšak často sa stretávame s problémom, že vzťahy medzi vstupmi a výstupom nie sú lineárne. To znamená, že priamočiara lineárna závislosť medzi premennými nemusí byť dostatočne presná pre popis skutočného vzťahu.

Na riešenie tohto problému sa používajú rôzne postupy. Jednou z alternatív je logaritmické ošetrenie hodnôt, čo vedie k vyššej mierke linearity. Tento postup transformuje hodnoty pomocou logaritmu a pomáha znížiť rozptyl hodnôt, čím je možné modelovať aj nelineárne závislosti medzi premennými. Logaritmická transformácia je užitočná, keďže môže "vyrovnávať" veľké rozdiely v rozptyloch hodnôt a spraví vzťahy medzi premennými lineárnejšími.

Ďalším validným prístupom je rozdelenie dát do skupín, ktoré majú približne lineárnu závislosť. Tento postup sa nazýva rozdelenie na podsady alebo rozdeľovanie na segmenty. Týmto spôsobom je možné identifikovať rôzne lineárne závislosti v rôznych oblastiach dátového priestoru. Každá skupina alebo segment dát môže byť potom modelovaný pomocou lineárnej regresie, čo umožňuje zachytiť komplexnejšie vzťahy a nelineárne štruktúry.

Pri výbere vhodného prístupu je dôležité zvážiť charakter dát a štruktúru vzťahov medzi premennými. Lineárna regresia je jednou z klasických metód s relatívne jednoduchými, ale stabilnými modelmi. Jej hlavnou nevýhodou je neschopnosť presne zachytiť nelineárne závislosti medzi premennými. Tento nedostatok je možné čiastočne obísť pridaním kombinácií atribútov alebo použitím iných pokročilejších metód, ktoré dokážu modelovať komplexnejšie vzťahy medzi premennými.

Metóda generalizovanej lineárnej regresie je alternatívou k štandardnej lineárnej regresii. Jej hlavným rozdielom je predspracovanie množiny atribútov pomocou bázových funkcií, čo umožňuje cieľovým premenným mať inú distribúciu chybovosti než normálnu distribúciu. Týmto spôsobom sa dá prispôsobiť modelovanie rôznym typom dát a ich špecifickým charakteristikám. Bázové funkcie môžu byť napríklad polynomiálne funkcie, Fourierove transformácie alebo iné transformačné metódy.

Ďalšou alternatívou s veľkým potenciálom je metóda lokálnej váhovanej aproximácie. Táto metóda identifikuje kľúčový stav v trénovacích dátach a uplatňuje váhovanie ostatných stavov v závislosti od ich vzdialenosti od kľúčového stavu. Tento proces zabezpečuje, že blízke stavy majú väčší vplyv na model ako vzdialenejšie stavy. Jednou zo známych funkcií, ktoré sa používajú pri váhovaní, je bázová funkcia Kernel. Táto funkcia priraďuje váhy bodom v závislosti od ich vzdialenosti od kľúčového stavu. V grafickej interpretácii sa klasická lineárna regresia snaží optimalizovať priamku prechádzajúcu cez všetky trénovacie stavy, bez ohľadu na ich váhu. Metóda lokálnej váhovanej aproximácie však zohľadňuje váhy a snaží sa lepšie prispôsobiť okolitému prostrediu jednotlivých stavov.

Nevýhodou lokálnej váhovanej aproximácie je, že pre každú novú sadu trénovacích dát je potrebné definovať špecifickú sústavu kľúčových stavov. To znamená, že táto metóda je citlivá na zmeny v dátach a vyžaduje prispôsobenie pre každý nový trénovací súbor. Avšak táto metóda je užitočná pri modelovaní nelineárnych vzťahov a v situáciách, kde nie je vhodné použiť globálny prístup, ako je štandardná lineárna regresia.

Optimálne zadefinovanie kľúčových stavov v trénovacích dátach je dôležitou otázkou pri predikcii medicínskych dát. Existuje niekoľko prístupov, ktoré sa môžu použiť na výber týchto kľúčových stavov.

V prípade, že sú dáta rozvrhnuté rovnomerne, je možné kľúčové stavy rozmiestniť v štruktúre pravidelnej mriežky. Tento prístup je vhodný, keď máme dostatočný veľký počet dátových profilov a predpokladáme, že rovnomerné pokrytie priestoru dátovými stavmi je dostačujúce pre modelovanie vzťahov.

Alternatívne je možné vybrať kľúčové stavy náhodne. Tento prístup môže byť vhodný, keď nemáme konkrétnu predstavu o distribúcii dátových stavov a chceme získať reprezentatívny vzorkový priestor. Výhodou je, že náhodný výber kľúčových stavov je jednoduchý a nemusíme poznať štruktúru dát.

Ďalším prístupom je dynamicky vyberať optimálnu polohu kľúčových stavov pomocou metód reinforcement learning a analýzy zhlukov. Tento prístup je náročnejší na výpočet, ale môže byť veľmi efektívny, najmä pri menších a nerovnomerne rozložených dátových vzorkách. Metódy reinforcement learning sa používajú na adaptívne optimalizovanie polohy kľúčových stavov na základe spätnoväzby zo systému. Analýza zhlukov môže identifikovať podobné vzory v dátach a pomôže nám rozdeliť priestor na relevantné segmenty.

Výber optimálneho spôsobu definovania kľúčových stavov závisí od dostupných dát, ich distribúcie a od konkrétnej úlohy predikcie medicínskych dát. Každý prístup má svoje výhody a nevýhody, a preto je dôležité testovať a porovnávať ich účinnosť na konkrétnych dátových vzorkách a úlohách predikcie.

4.2.2.2 Umelé neurónové siete

Neurónové siete sú komplexné výpočtové systémy, ktoré pozostávajú z viacerých jednoduchých, efektívne prepojených prvkov nazývaných neuróny. Tieto siete spracovávajú informácie na základe dynamických odpovedí neurónov na externé vstupy, pričom ich správanie závisí od ich aktuálneho stavu. Neurónové siete sú obvykle tvorené viacerými vzájomne prepojenými rozhodovacími uzlami, ktoré majú aktivačnú funkciu.

Uzly v neurónových sieťach sú organizované do vrstiev. Vstupné dáta sú vložené do vstupnej vrstvy, ktorá komunikuje s viacerými skrytými vrstvami, kde dochádza k spracovaniu dát. Neurónové siete majú niekoľko špecifík. Nie sú klasicky sekvenčné a nemusia byť deterministické. V týchto sieťach chýba centrálny procesor a je nahradený paralelnými odpoveďami neurónov na vstupy. Neexistujú oddelené pamäťové adresy na ukladanie dát počas analýzy, informácia je reprezentovaná celkovým aktivačným nastavením všetkých uzlov v sieti. Nové poznatky sú teda reprezentované stavom siete ako celku, ktorý má väčšiu interpretatívnu silu ako súčet stavov jednotlivých uzlov.

Spojenia medzi uzlami v neurónovej sieti sú vážené a výsledok spracovania je prezentovaný na výstupnej vrstve uzlov. Váhy medzi uzlami sú nastavované podľa predom zadaného pravidla učenia. Existuje niekoľko pravidiel učenia, pričom jedným z najčastejšie používaných je delta pravidlo. Toto pravidlo umožňuje spätnú propagáciu chyby z jednej vrstvy do uzlov vrstvy predchádzajúcej. Pomocou pravidiel učenia sa umelá neurónová sieť postupne posúva na učacej krivke od náhodného typovania až k expertnému výberu.

Aby sa zvýšila efektivita neurónových sietí, často sa používajú sigmoidálne aktivačné funkcie, ktoré polarizujú aktivitu siete a prispievajú k jej stabilizácii. Tieto aktivačné funkcie zabezpečujú, že aktivity neurónov sú obmedzené na určitý rozsah, čo môže pomôc

Umelá neurónová sieť je metóda univerzálnej aproximácie, čo znamená, že dokáže modelovať rôzne systémy s dostatočnou toleranciou k počiatočnej chybovosti. Je vhodná na zachytávanie asociácií a identifikáciu pravidelností v množine dátových profilov. Taktiež je schopná analyzovať dátové štruktúry s veľkým počtom premenných alebo veľkým objemom dát.

Po úspešnom natrénovaní neurónovej siete na požadovanú úroveň sa môže použiť ako analytický nástroj na iných vzorkách dát. V tejto fáze už nie je potrebné špecifikovať trénovanie siete, ale iba spustiť vpred propagáciu dát cez sieť. Výstupom tejto propagácie (známej ako "forward propagation run") je predikčný model pre dané dáta. Tento model môže byť následne použitý v ďalšej komplexnej analýze alebo interpretácii dát.

Umelé neurónové siete majú teda široké spektrum aplikácií a výhody. Sú schopné naučiť sa vzťahy medzi premennými, aj keď sú neúplne popísané, a zvládajú pracovať s veľkým počtom premenných a rozsiahlymi dátovými sadami. Ich schopnosť univerzálnej aproximácie im umožňuje modelovať rôzne

systémy a identifikovať vzory a vzťahy v dátach. Týmto spôsobom môžu byť významným nástrojom pri analýze a využití komplexných dátových sád.

Okrem všeobecných metodologických limitácií, ktoré sme spomínali pri popise metód lineárnej regresie, neurónové siete založené na princípe spätnej propagácie a aj iné typy sietí sa v podstate javia ako "čierne skrinky". To znamená, že okrem definovania primárnej architektúry siete a prvých testov s náhodnými číslami nemá administrátor inú možnosť, ako dodávať vstupné dáta do siete a spracovávať jej výstupy. Niektoré nástroje, ako napríklad NevProp alebo Mactivation, umožňujú získavať vzorky postupu siete v definovaných intervaloch, ale učiaca krivka siete sa vyvíja samostatne.

Výsledkom tohto procesu je samotná sieť, nie však konkrétna rovnica alebo séria koeficientov, ktoré by popisovali vzťahy medzi premennými mimo architektúry siete, ako je tomu pri lineárnej regresii. Existuje potenciálna limitácia, že neurónové siete je možné pretrénovať, čo znamená, že sieť dosahuje dobré výsledky na trénovacích dátach, ale jej predikčné schopnosti na nových dátach sú slabé.

Na vyriešenie tohto problému sa používajú rôzne postupy. Medzi ne patria predčasné ukončenie trénovania, regulácia (ktorá zahŕňa pridanie vhodne vážených parametrov siete do chybovej funkcie a ich minimalizáciu), dropout (ktorý náhodne vypína niektoré neuróny počas trénovania) a dropconnect (ktorý náhodne vypína niektoré spojenia medzi neurónmi počas trénovania). Výber týchto postupov závisí od konkrétnej situácie a úlohy.

Neurónové siete a ich modely sú veľmi citlivé na voľbu parametrov, pričom vo vedeckej komunite neexistujú striktné pravidlá pre ich výber. Preto je jedným z dôležitých cieľov nášho projektu testovať viaceré parametrické nastavenia neurónových sietí, ktoré budú slúžiť ako model pre predikciu faktorov v zdravotníckych dátach.

Výzvou v rámci tohto projektu je správny výber vstupných parametrov a ich počet, ako aj vhodná veľkosť a typy skrytých vrstiev v sieti.

Hlavnou výhodou neurónových sietí je ich schopnosť modelovať nelineárne a dynamické závislosti, čo ich robí vhodnými na riešenie komplexných problémov. V porovnaní s metódami lineárnej regresie sú neurónové siete presnejšie a majú rýchlejšiu trénovaciu schopnosť. Ich hlavnou silou je schopnosť zachytiť rôznorodé vzťahy medzi premennými, čo umožňuje modelovať aj veľmi komplexné fenomény, ktoré by bolo obtiažne, ak nie nemožné, modelovať pomocou tradičných štatistických nástrojov.

Od roku 2006 sa významne presadzujú neurónové siete s veľkým počtom vrstiev, nazývané aj deep networks. Tieto siete majú viacero skrytých vrstiev a hoci presný spôsob, akým tieto vrstvy fungujú, nie je úplne známy, majú potenciál dosahovať presnejšie výsledky ako klasické neurónové siete. Predbežné výsledky naznačujú, že pri dostatočnom množstve dostupných dát majú deep networks vysokú efektivitu a sú účinné v širokom spektre oblastí.

Avšak existuje aj určitý nedostatok, ktorým je možnosť, že neurónové siete sa jednoducho nedokážu správne natrénovať. Trénovanie neurónovej siete je proces, ktorý vyžaduje správnu konfiguráciu a dostatok dát. Ak nie sú dostupné dostatočné množstvo reprezentatívnych dát, môže sa stať, že sieť nebude schopná naučiť sa relevantné vzorce a nevyprodukujú sa presné predikcie na nových dátach. Preto je dôležité zabezpečiť dostatočné množstvo kvalitných trénovacích dát pre úspešné natrénovanie neurónových sietí.

4.2.2.3 Support vector machines (podporné vektory)

Podporné vektory (anglicky Support Vector Machines) sú metódou, ktorá identifikuje dôležité vzorky v trénovacích dátach a následne využíva podobnosti medzi vstupom a týmito vzorkami na predikciu výsledku. Táto metóda konštruuje nadroviny vo veľkodimenzionálnom alebo dokonca nekonečnom priestore. Tieto nadroviny sa potom využívajú na riešenie problémov regresie alebo klasifikácie. Support vector machines identifikujú funkčné hranice, ktoré majú najväčšiu vzdialenosť od dátových bodov inej triedy. Čím väčšia je táto vzdialenosť, tým menšia je generalizačná chyba

klasifikátora. Pomocou jadrovej metódy je možné vytvoriť nelineárne klasifikátory, keďže nadroviny s maximálnym lemom nie sú obmedzené na lineárne hranice.

Existujú aj viactriedové support vector machines, ktoré priraďujú inštanciám vlastnosti vybrané z konečného počtu preddefinovaných prvkov. Zaujímavé sú najmä modely, ktoré sa pozerajú na problém viac tried ako na jeden optimalizačný problém a neriešia ho ako sériu binárnych klasifikačných problémov (napríklad Crammerova a Singerova metóda). Jedinečnou vlastnosťou support vector machines je, že maximalizácia geometrického lemu medzi triedami a minimalizácia klasifikačnej chyby sa vykonávajú súčasne.

Nevýhodou tejto metódy je jej vysoká časová náročnosť pri trénovaní, ktorá je väčšinou kvadratická vzhľadom na počet vzoriek. Týmto spôsobom nie je vhodná pre veľké dátové sady. V praxi sa však support vector machines úspešne využívajú napríklad pri klasifikácii hypertextov, kategorizácii obrázkov, rozpoznávaní písaného textu a klasifikácii proteínov v medicíne.

4.2.3 Zdravotný monitoring a metodologický prístup k modelovaniu zdravotných dát

Pri diskusii o zbere zdravotných dát sa stretávame s metodologickými špecifikami, na ktoré je potrebné zareagovať. Vzhľadom na to, že zberové systémy akumulujú obrovské množstvo dát, nie je možné vložiť všetky merania do jedného algoritmu ako jednorazový vstup. Preto je vhodné najprv klastrovať zberné jednotky (napríklad pacientov) a následne používať agregované dáta zhlukov dátových profilov na vytváranie modelov. Tento prístup má navyše pridanú hodnotu v tom, že v rámci klastrov môžeme deskriptívne sledovať rôzne trendy, ako napríklad stabilitu, intenzitu rastu alebo poklesu, ktoré môžu ovplyvniť vývoj v budúcnosti. Príkladom môže byť sledovanie vývoja geografických oblastí a predikcia ich budúceho vývoja. V tejto podkapitole sa zameriame na odpoveď na otázku, akým spôsobom je možné vytvárať dátové klastre najoptimalizovanejším spôsobom. Existuje viacero štatistických nástrojov pre efektívne klastrovanie dát, a v krátkosti predstavíme tie, ktoré sú relevantné v kontexte profilov dátových sád, s ktorými počítame po nástupe inteligentného merania.

4.2.3.1 K-means

Metóda K-means je algoritmus používaný na klastrovanie dát, ktorý pôvodne pochádza z oblasti spracovania signálov v šume. Jeho princíp spočíva v rozdelení dátových bodov do k počtu klastrov podľa ich podobnosti.

Pri k-means algoritme sa najprv vyberie počiatočné k bodov, ktoré slúžia ako centroidy pre jednotlivé klastre. Následne sa iteratívne opakuje nasledovné: každý dátový bod je priradený k najbližšiemu centroidu na základe ich euklidovskej vzdialenosti. Potom sa vypočítajú nové pozície centroidov ako priemerné hodnoty všetkých bodov priradených k danému centroidu. Tento proces sa opakuje, až kým sa centroidy prestanú meniť alebo dosiahne maximálny počet iterácií.

K-means je rýchla a dobre paralelizovateľná metóda, ktorá je často využívaná pri klastrovaní veľkých sád dát. Výhodou je jej jednoduchosť a efektivita v porovnaní s inými algoritmami klastrovania. Avšak má aj niekoľko významných nevýhod.

Jednou z nevýhod je citlivosť na škálovanie vstupných premenných. To znamená, že ak sú jednotlivé premenné rôzne škálované, napríklad jedna premenná je v desiatkach a druhá v jednotkách, tak môže mať táto nerovnosť v škále vplyv na výsledky klastrovania. Preto je dôležité pred samotným k-means algoritmom premenné normalizovať alebo škálovať, aby boli v rovnakej škále.

Ďalšou nevýhodou je citlivosť na korelované vstupné dáta. Ak sú niektoré premenné vo vstupných dátach silne korelované, môže to viesť k nepresným výsledkom klastrovania. K-means algoritmus totiž predpokladá, že jednotlivé premenné sú nezávislé a rovnako významné. Ak však existuje silná korelácia medzi premennými, môže to viesť k neadekvátnemu priradeniu bodov do klastrov.

Preto je dôležité pri použití metódy K-means mať na pamäti tieto nevýhody a vhodne upraviť vstupné dáta pred samotným klastrovaním.

4.2.3.2 Sparse coding

Metóda sparse coding sa zameriava na hľadanie vzorov v dátových setoch, ktoré sú schopné vysvetliť tieto dáta. Jej cieľom je nájsť optimálnu sadu vzorov, ktorá dokáže vysvetliť každý dátový bod pomocou iba niekoľkých vybraných vzorov. Tento prístup je známy aj ako sparse coding, pretože sa snaží minimalizovať počet použitých vzorov na vysvetlenie dát.

Proces učenia sparse coding prebieha autonómne, bez potreby externého učiteľa. Týmto spôsobom je možné získať hlbší vhľad do správania meraných veličín. Metóda sparse coding sa často využíva na tvorbu profilov rôznych kategórií pacientov a na detekciu aktivít jednotlivých parametrov počas dňa a spánku. Tieto zistenia nielenže prispievajú k celkovému predikčnému modelu, ale majú aj samostatnú interpretatívnu silu, pokiaľ ide o sledované vzťahy.

Dôležitým faktorom pri využití sparse coding je frekvencia merania. Pri nízkej frekvencii merania je možné sledovať len zmeny v celkovom charaktere dát. Tradične sa dátový tok rozlišuje len podľa prípadov s abnormalitami a zariadenia sú klasifikované do spotrebných kategórií.

Existuje niekoľko komplexnejších modelov, ako napríklad faktorizované skryté Markovove modely, ktoré môžu zvýšiť efektivitu finálneho modelu. V tejto fáze projektu však nie je potrebné podrobne opisovať tieto modely.

4.2.4 Optimalizačné nástroje

Projekt zahŕňa nielen predikciu a modelovanie, ale aj ďalšie dôležité aspekty. Výstupné dáta z týchto modelov nám poskytujú hodnotné interpretačné informácie, ktoré môžeme využiť pri poskytovaní poradenstva a optimalizácii zdravotného stavu pacientov. Využívame širokú škálu štatistických nástrojov, ktoré nám umožňujú automatizovať tento proces a dosiahnuť vysokú efektivitu.

Existuje niekoľko prístupov, ktoré sa v rámci projektu testujú a vyvíjajú. Prvým z nich je optimalizácia procesov, ktorá sa zameriava na hľadanie najlepších stratégií a postupov pre diagnostiku a liečbu pacientov. Využívame pokročilé algoritmy a metódy na identifikáciu najefektívnejších postupov a prispôsobenie ich individuálnym potrebám pacientov.

Druhým prístupom je optimalizácia systémov, ktorá sa zaoberá vylepšovaním celkového fungovania zdravotných systémov. Analyzujeme a modelujeme rôzne aspekty, ako napríklad čakacie doby, efektivitu procesov a distribúciu zdrojov. Cieľom je nájsť optimálne riešenia, ktoré zlepšia prístup k zdravotnej starostlivosti a zabezpečia vysokú kvalitu poskytovaných služieb.

Tretím prístupom je využitie štatistických nástrojov na interpretáciu a analýzu výstupných dát. Pomocou pokročilých štatistických metód identifikujeme vzťahy, trendy a dôležité faktory, ktoré ovplyvňujú zdravotný stav pacientov. Tieto poznatky nám umožňujú poskytovať presnejšie a personalizovanejšie poradenstvo, čo prispieva k zlepšeniu výsledkov liečby a celkovej starostlivosti o pacientov.

Projekt zameraný na predikciu, modelovanie a optimalizáciu v zdravotníctve prináša veľký potenciál pre zlepšenie diagnostiky, liečby a správy zdravotného stavu pacientov. Kombinácia pokročilých analytických nástrojov a využitie veľkých objemov dát nám umožňuje posúvať sa smerom k personalizovanejšej, efektívnejšej a výkonnejšej zdravotnej starostlivosti.

4.2.4.1 Interger linear programming (celočíselné lineárne programovanie)

Celočíselné lineárne programovanie (celočíselné LP) je mocný matematický nástroj, ktorý umožňuje riešiť optimalizačné problémy so zreteľom na celočíselné hodnoty premenných. Táto technika

je široko využívaná v rôznych oblastiach, vrátane logistiky, plánovania výroby, dopravy, financií a mnohých ďalších.

V celočíselnom lineárnom programovaní je hlavným cieľom nájsť hodnoty premenných, ktoré minimalizujú alebo maximalizujú lineárnu funkciu týchto premenných. Tieto premenné sú obmedzené lineárnymi podmienkami, ktoré zohľadňujú obmedzenia a požiadavky daného problému. Okrem toho sa špecifikuje, že niektoré premenné musia byť celočíselné, čo dodáva problému diskrétny charakter.

Celočíselné LP je schopné modelovať a riešiť rôznorodé problémy. Jeho sila spočíva v tom, že dokáže zachytiť aj nelineárne závislosti medzi premennými pomocou vhodných transformácií. Napriek tomu, že sa lineárne programovanie zameriava na lineárne vzťahy, celočíselné LP poskytuje flexibilitu a rozširuje možnosti modelovania.

Existuje niekoľko metód na riešenie celočíselného LP. Niektoré z nich sú presné metódy, ktoré skúmajú celý priestor možných riešení, zatiaľ čo iné sú heuristiky alebo metaheuristiky, ktoré poskytujú približné riešenia s určitou mierou optimalizácie. Výber metódy závisí od špecifík problému, veľkosti dát a časových obmedzení.

Celočíselné lineárne programovanie je veľmi užitočným nástrojom pri riešení optimalizačných problémov, ktoré vyžadujú diskrétny charakter premenných. Jeho využitie prispieva k zlepšeniu výkonu, efektivity a rozhodovacieho procesu v mnohých odvetviach a aplikáciách. S vhodným modelovaním a riešením celočíselného LP môžeme dosiahnuť významné zlepšenia a nájsť optimálne stratégie v mnohých oblastiach podnikania a inžinierstva.

4.2.4.2 Heuristické metódy

Heuristické metódy sú výpočtové postupy, ktoré sa používajú na hľadanie riešenia problémov, kde exaktné metódy by boli príliš náročné alebo nemožné. Tieto metódy sa zameriavajú na prehľadávanie stavového priestoru riešení a postupne vylepšujú aktuálne riešenie pomocou lokálnych optimalizácií. Heuristiky sú založené na rôznych princípoch, ktoré sa snažia nájsť riešenia, ktoré sú dostatočne dobré, ale nemusia byť nutne najlepšie z hľadiska optimalizácie v zložitých problémoch.

Heuristické riešenia sú často tvorené sériou jednoduchých procedúr, ktoré sa opakujú až kým nie je dosiahnuté určené kritérium zastavenia. Tieto procedúry môžu zahŕňať generovanie náhodných riešení, aplikáciu lokálnych optimalizácií, pravidlá selekcie a mutácie, a ďalšie. Cieľom je nájsť riešenie, ktoré spĺňa požadované kritériá a je prijateľné v kontexte daného problému. Heuristiky často poskytujú rýchle a praktické riešenia, aj keď nie sú exaktne optimálne.

Jednou z najpoužívanejších skupín heuristík sú genetické algoritmy. Tieto algoritmy sú inšpirované biologickou evolúciou a využívajú princípy ako iteratívne testovanie, mutácia, pravidlá selekcie a obmedzenie počtu agentov, parametrov alebo vstupov. Genetické algoritmy pracujú s populáciou riešení, kde každé riešenie je reprezentované ako genetický kód. Prostredníctvom operátorov kríženia, mutácie a selekcie sa generujú nové potomstvo, ktoré postupne konverguje k lepším riešeniam. Genetické algoritmy majú schopnosť optimalizovať veľký počet parametrov a sú flexibilné pre rôzne problémy.

Je dôležité si uvedomiť, že heuristické riešenia sú často navrhnuté a prispôsobené pre konkrétny problém. Ich úspešnosť závisí od správneho nastavenia parametrov a vhodného modelovania problému.

4.2.4.3 Reinforcement Learning

Metódy Reinforcement Learning (RL) predstavujú súbor algoritmov a modulov, ktoré sa zameriavajú na optimalizáciu rozhodovacích procesov agenta v problémovom priestore. Tento priestor je definovaný množinou akcií, ktoré môže agent vykonávať, a stavov, v ktorých sa môže nachádzať.

Hlavnou motiváciou agenta je minimalizovať tresty a/alebo maximalizovať odmeny, čo mu pomáha identifikovať optimálne riešenie pre daný problém.

Reinforcement Learning umožňuje rozloženie komplexného problému na jednoduchšie podproblémy. Agent sa snaží nájsť optimálnu stratégiu rozhodovania, ktorá mu pomôže maximalizovať očakávané dlhodobé odmeny. Prostredníctvom procesu učenia sa agent učí priradiť hodnoty jednotlivým akciám a stavom na základe skúseností a interakcie s prostredím. Týmto spôsobom sa vytvára model problémového priestoru, ktorý agentovi pomáha rozhodovať sa efektívne a adaptovať sa na zmeny prostredia alebo parametrov.

Výhodou Reinforcement Learningu je jeho relatívne jednoduchá štruktúra a možnosť efektívneho riadenia procesu učenia. Na rozdiel od niektorých iných metód, ako napríklad neurónové siete, celý proces RL je transparentný a výsledky sú interpretovateľné. Navyše, do procesu učenia je možné kedykoľvek zasiahnuť supervízorom, čo umožňuje ľahšie zohľadňovanie expertného vedomia.

Systémy založené na metódach Reinforcement Learningu sú schopné dynamicky sa prispôsobovať zmenám v problémovom priestore alebo parametrizácii. Ich výstupom nie je jediná finálna stratégia, ale skôr zoznam alternatívnych riešení spolu s hodnotami, ktoré vyjadrujú ich efektivitu. Okrem toho tieto systémy poskytujú popis problémového priestoru a jeho komponentov a ich vplyvu na stratégiu rozhodovania.

Nevýhodou metód Reinforcement Learningu je potreba iteratívneho procesu učenia, čo vyžaduje časovú náročnosť. S komplexnosťou problémového stavu narastá aj počet iterácií potrebných na nájdenie optimálnej stratégie.

4.2.5 Kvalita spracovaných zdravotníckych dát

Podoba odvetvia zdravotnej starostlivosti sa úplne zmenila vďaka technologickému pokroku. Softvérové riešenia, ako napríklad elektronické zdravotné záznamy, uľahčujú organizáciám prístup k zdravotným údajom a histórii pacientov prostredníctvom integrácie údajov, čo prináša niekoľko výhod.

V súčasnosti sa technologický pokrok stal nevyhnutným prvkom pre vývoj zdravotnej starostlivosti. Elektronické zdravotné záznamy a ďalšie softvérové riešenia, ktoré umožňujú prístup k zdravotným údajom pacientov, výrazne zjednodušujú a zlepšujú procesy starostlivosti o pacientov. Vďaka integrácii údajov z rôznych zdrojov je možné získať presný prehľad o zdravotnom stave pacientov a optimalizovať ich liečbu pomocou presných údajov.

Avšak, pri získavaní a spracovaní týchto údajov je dôležité byť opatrný a zodpovedný. Nie vždy je možné zabezpečiť kvalitu a spoľahlivosť údajov, a preto je potrebné zohľadniť nadväzujúce prevádzkové a administratívne problémy, ktoré môžu vzniknúť. Starostlivosť o zdravie pacientov by sa mala riadiť etickými a právnymi normami, aby boli zabezpečené ich práva a ochrana osobných údajov. Zodpovedné spravovanie údajov a technologických riešení v zdravotníctve je nevyhnutné pre zabezpečenie kvalitnej starostlivosti o pacientov.

Kvalita údajov v zdravotníctve má veľký význam pre správne rozhodnutia pri starostlivosti o pacientov. Ak sú údaje neúplné, nepresné alebo neaktuálne, môže to viesť k chybám v diagnostike a liečbe pacientov. Preto je dôležité, aby bol zabezpečený prístup k kvalitným údajom a aby boli tieto údaje správne spracované.

Jednou z výhod zdieľania informácií o pacientoch je, že umožňuje lekárom a iným poskytovateľom zdravotnej starostlivosti prístup k dôležitým údajom, ktoré by inak mohli byť nedostupné. Avšak, citlivá povaha zdravotníckych údajov si vyžaduje, aby tieto údaje boli chránené prísnymi bezpečnostnými opatreniami a aby boli spracované v súlade s príslušnými predpismi a nariadeniami.

Koncepcia riadenia kvality údajov v zdravotníctve sa podobá koncepcii kontroly kvality, ktorá sa používa v iných odvetviach. V oblasti zdravotníctva sa používajú rôzne technológie, nástroje a postupy

na overenie zdrojov získaných údajov a na zabezpečenie ich kvality. To zahŕňa overenie toho, či sú údaje presné, úplné a aktuálne, a ak nie, tak identifikácia a oprava chýb.

Hlavným cieľom riadenia kvality údajov v zdravotníctve je chrániť pacientov a zabezpečiť, aby boli informácie o ich zdravotnom stave presné a spoľahlivé. To zahŕňa integráciu spoľahlivých zdrojov údajov a prenos týchto údajov spoľahlivým príjemcom. V konečnom dôsledku sa zabezpečenie kvality údajov v zdravotníctve snaží zabezpečiť to, aby pacienti dostávali najlepšiu možnú starostlivosť o svoje zdravie.

Dôvera je jednou z najdôležitejších zložiek úspechu v odvetví zdravotnej starostlivosti a bez presných údajov, ktoré slúžia ako základ pre presné diagnostikovanie a liečbu, nie je možné dôveru vytvoriť. Kvalita údajov je preto kritickým faktorom pre úspešné fungovanie systému zdravotnej starostlivosti. Problémy s kvalitou údajov môžu viesť k nesprávnym diagnózam a liečbe, čo môže mať zničujúce následky pre pacientov.

Preto je veľmi dôležité riešiť problémy s kvalitou údajov v zdravotníctve. Rôzne technológie, nástroje a postupy sú používané na kontrolu kvality údajov. Tieto postupy zahŕňajú overenie zdrojov získaných údajov, aby sa zabezpečila ich spoľahlivosť, a integráciu údajov zo spoľahlivých zdrojov. Hlavným cieľom týchto postupov je chrániť pacientov tým, že overujú informácie a prenášajú ich spoľahlivým príjemcom.

Vylepšenie kvality údajov má preto priamy vplyv na výsledky v zdravotníctve, vrátane lepšej starostlivosti o pacientov a efektívnejšej starostlivosti o nich. Neustále zisťovanie chýb v údajoch a ich odstránenie sú preto kritické pre poskytovanie presných údajov, čo zase vedie k väčšej dôvere v odvetví zdravotnej starostlivosti a úspešnému zlepšeniu výsledkov liečby pacientov.

4.2.5.1 Výhody kvality údajov

Vysokokvalitné údaje majú kľúčový vplyv na informované rozhodovanie v rámci zdravotníckeho odvetvia. Čím presnejšie údaje sú k dispozícii, tým lepšie je možné prijímať rozhodnutia, ktoré budú mať zásadný vplyv na pacienta a jeho liečbu. Presné informácie sa prejavia v rôznych oblastiach, od diagnostiky a liečby až po sledovanie stavu pacienta. Navyše, kvalitné údaje zvyšujú dôveru v rozhodnutia, ktoré sú prijímané a znižujú riziká spojené s chybami v údajoch.

Okrem toho, vysoká kvalita údajov umožňuje dôsledné zlepšovanie výsledkov. Analyzovaním presných údajov možno zistiť slabé miesta v liečbe a navrhnúť zlepšenia, ktoré zlepšia celkové výsledky. Tento proces zlepšovania sa môže ďalej rozšíriť na celé odvetvie, čo povedie k lepšej starostlivosti o pacientov a efektívnejšiemu poskytovaniu zdravotnej starostlivosti. Preto je dôležité klásť dôraz na zabezpečenie vysokokvalitných údajov v rámci zdravotníckeho odvetvia.

Presné a presné dáta zvyšujú efektivitu a úspešnosť marketingových a reklamných kampaní v zdravotníctve. Zameriavanie sa na konkrétnu skupinu pacientov, ktorí najpravdepodobnejšie budú potrebovať určité druhy zdravotnej starostlivosti, znižuje náklady na marketing a zvyšuje pravdepodobnosť úspechu. Z toho vyplýva, že zlepšená kvalita údajov v zdravotníctve pomáha aj pri vytváraní cielenejších marketingových kampaní, zvyšuje záujem potenciálnych pacientov o produkty a služby a pomáha dosahovať ciele v oblasti zdravotníckeho marketingu.

Kvalitné a presné údaje môžu prispieť k zlepšeniu vzťahov medzi pacientmi a lekármi a zvýšiť úspešnosť zdravotníckych zariadení. Keďže údaje sú dôležitým prvkom starostlivosti o pacientov, ich zber a zhromažďovanie môže pomôcť lekárom pochopiť ich pacientov lepšie a vytvoriť osobný prístup ku každému z nich.

Presné údaje umožňujú lekárom a zdravotníckym pracovníkom pochopiť potreby pacientov a prispôsobiť svoje služby podľa nich. Zber údajov o pacientoch môže pomôcť poskytovateľom zdravotnej starostlivosti identifikovať rizikové faktory a riešiť ich predtým, ako sa zhoršia a vyžadujú si väčšiu liečbu.

Vytváranie dôvery medzi pacientmi a zdravotníckymi pracovníkmi je kľúčové pre poskytovanie kvalitnej zdravotnej starostlivosti. Kvalitné údaje môžu pomôcť lekárom a zdravotníckym pracovníkom vytvárať dôverné vzťahy s pacientmi a zlepšiť komunikáciu medzi nimi. To môže viesť k vyššej spokojnosti pacientov a podpore ich zdravotného stavu.

Zároveň môžu údaje o pacientoch pomôcť zdravotníckym zariadeniam zlepšiť svoje služby a lepšie cieliť svoje kampane. Zhromažďovanie údajov o pacientoch umožňuje zdravotníckym pracovníkom zistiť, kto sú ich potenciálni pacienti, a zamerať sa na nich s cieľavedomými kampaniami. Toto môže pomôcť zvýšiť povedomie o zdravotných službách a zvýšiť počet pacientov, ktorí vyhľadávajú zdravotnú starostlivosť. Celkovo teda zlepšenie kvality údajov môže prispieť k lepšej starostlivosti o pacientov a efektívnejšej prevencii a liečbe zdravotných problémov.

Jednoduchšia implementácia vysokokvalitných údajov prináša niekoľko výhod. Predovšetkým, vysokokvalitné údaje sú spravidla priamo použiteľné a nevyžadujú rozsiahle úpravy. To znamená, že sa môžu rýchlo a jednoducho implementovať do existujúcich procesov, čím sa zvyšuje efektívnosť. Na rozdiel od toho, oprava nekonzistentných alebo neúplných údajov je zložitá a zvyčajne zaberá veľa času. To môže spomaliť procesy a predĺžiť čas potrebný na implementáciu nových poznatkov z analýzy údajov. S vysokokvalitnými údajmi je tak možné rýchlo a jednoducho zlepšiť výkon a efektívnosť zdravotníckych zariadení a zvýšiť kvalitu poskytovanej starostlivosti.

Vysoká kvalita údajov má priamy vplyv na ziskovosť zdravotníckych zariadení. V prvom rade vedie k lepším vzťahom s pacientmi, čo môže mať za následok zvýšenú vernosť a spokojnosť pacientov, ktorí sa budú radi vracať a odporúčať zariadenie ďalším. Okrem toho, presné a spoľahlivé údaje umožňujú zdravotníckym zariadeniam lepšie a informovanejšie rozhodovanie, čo vedie k lepším výsledkom liečby a optimalizácii zdrojov.

Efektívny zber údajov znižuje množstvo zbytočne plytvaných zdrojov a času, ktorý sa môže následne venovať zlepšovaniu procesov zdravotnej starostlivosti. Na druhej strane, nekvalitné údaje môžu viesť k početným úzkym miestam, ktoré zvyšujú náklady a zníženie efektívnosti.

Dobrá kvalita údajov tiež umožňuje zdravotníckym zariadeniam lepšie cielenie zdravotných kampaní a aktivít, čo môže viesť k zvýšeniu počtu pacientov, ktorí využívajú zariadenie. To v konečnom dôsledku prispieva k zvýšeniu ziskovosti a konkurencieschopnosti. Preto je dôležité, aby sa zdravotnícke zariadenia venovali správnej starostlivosti o získavanie, ukladanie a spracovávanie údajov s cieľom dosiahnuť čo najvyššiu kvalitu.

4.2.5.2 Ukazovatele

Presnosť a precíznosť sú jednými z kľúčových aspektov dobrej kvality údajov. Presnosť znamená, že údaje sú správne a zodpovedajú realite, zatiaľ čo precíznosť sa týka toho, ako presne sú údaje zaznamenané a merané. Ak sú údaje presné a precízne, môžu byť použité na správne rozhodnutia a na riešenie problémov.

Je dôležité, aby boli údaje v zdravotníctve presné a správne zaznamenané, pretože zavádzajúce alebo nesprávne údaje môžu viesť k zlej diagnóze, nesprávnemu liečeniu a dokonca k zdravotným komplikáciám alebo smrti pacienta. Preto je dôležité investovať do špičkového systému riadenia kvality údajov v zdravotníctve.

Správne a precízne údaje tiež umožňujú lekárom a zdravotným pracovníkom vytvárať účinné liečebné plány a zabezpečovať lepšiu zdravotnú starostlivosť. V neposlednom rade, správne zaznamenané údaje umožňujú aj efektívnejšie hospodárenie s finančnými prostriedkami, znižujú náklady na zdravotnú starostlivosť a zvyšujú výnosy z investícií.

Platnosť údajov je dôležitým faktorom určujúcim kvalitu získaných údajov. V zdravotníctve sa často môžu vyskytnúť dáta, ktoré sú dôležité, ale nie sú oficiálne zaradené do analýz a preto sa môžu považovať za neplatné. Preto je kľúčové určiť, ktoré údaje sú relevantné a ktoré nie. Každý zdravotnícky subjekt má svoje vlastné súbor pravidiel pre spracovanie dát, ktoré umožňujú posúdiť platnosť údajov.

Tieto pravidlá sa zvyčajne týkajú napríklad typu zdroja údajov, úplnosti údajov, dátumu získania údajov a ďalších faktorov, ktoré môžu ovplyvniť platnosť dát. Dôležité je, aby každý zdravotnícky subjekt bol schopný preukázať platnosť svojich údajov a dodržiaval pravidlá stanovené pre správne spracovanie dát. To zabezpečuje, že údaje sú správne interpretované a môžu sa použiť na správne rozhodovanie.

Spoľahlivosť dát je kľúčovým faktorom pre zabezpečenie kvality údajov. Dáta môžu byť získané z rôznych zdrojov a systémov, a preto je dôležité zabezpečiť, aby všetky tieto dáta boli zhromaždené a konsolidované bez chýb a nezrovnalostí. To zabezpečuje, že údaje s rovnakými charakteristikami nebudú odlišovať sa v závislosti na zdroji, z ktorého boli získané. Preto je potrebné zabezpečiť, aby bol mechanizmus zhromažďovania a ukladania údajov stabilný a algoritmicky zabezpečený, aby sa minimalizovali možné odchýlky v údajoch a aby sa zabezpečila spoľahlivosť údajov.

Súdržnosť dát sa týka ich úplnosti a zhodnosti. Údaje musia byť kompletné a obsahovať všetky potrebné informácie pre ich správne vyhodnotenie a použitie. Ak sa v dátach niečo chýba, môže to viesť k neefektívnym rozhodnutiam alebo k nekonzistentným výsledkom. Preto je dôležité, aby sa údaje zbierali a ukladali v súlade s predom definovanými pravidlami a štandardmi, ktoré zabezpečujú ich súdržnosť.

Okrem toho je tiež dôležité, aby dáta boli zhodné, čiže aby boli v rovnakom formáte a štruktúre bez ohľadu na to, z akého zdroja pochádzajú. Ak sa dáta zhromažďujú z viacerých zdrojov, môže to viesť k rozdielnym štruktúram a formátom, ktoré môžu byť ťažko zosúladené. To môže viesť k nekonzistentným výsledkom alebo k chybám pri vyhodnocovaní dát. Preto je dôležité zabezpečiť súdržnosť dát nielen v rámci jednej organizácie, ale aj v rámci celého odvetvia.

Pri zhromažďovaní údajov je veľmi dôležité venovať pozornosť detailom, aby sa zabezpečila ich jedinečnosť. To znamená, že každý údaj by mal byť unikátny a neopakovať sa v súbore údajov. Ak je súbor údajov usporiadaný a detailný, pomáha to poskytnúť celkový obraz, ktorý sa zhoduje s očakávaniami. Zároveň to znižuje riziko nesprávnych rozhodnutí a neistoty v rámci celej organizácie.

Správna úroveň granularity dát je tiež kľúčová pre správne fungovanie operácií. Týka sa to najmä časového rozlíšenia, ktoré by malo byť dostatočne detailné, aby umožnilo sledovanie zmien v čase a zároveň dostatočne široké, aby poskytlo celkový obraz. Pri spracovaní údajov je preto dôležité mať na pamäti, že jedinečnosť a správna úroveň granularity údajov sú rozhodujúce pre bezproblémový chod operácií.

Riadenie kvality údajov v zdravotníctve je kritické pre poskytovateľov zdravotnej starostlivosti, ktorí sú zodpovední za správu a interpretáciu zdravotných záznamov pacientov. Tieto údaje sú veľmi citlivé a majú dôležitý vplyv na život jednotlivca. Preto musia poskytovatelia zdravotnej starostlivosti dodržiavať prísne predpisy a ochranné opatrenia, aby zabezpečili bezpečnosť a dôvernosť týchto údajov.

Navyše, aby poskytovatelia zdravotnej starostlivosti pochopili potrebu riadenia kvality údajov, musia sa uvedomiť aj dôsledky nekvalitných údajov. Nesprávne rozhodnutia a neefektívne procesy môžu mať vážne následky na zdravie pacientov a celkovú kvalitu zdravotnej starostlivosti. Preto je dôležité, aby poskytovatelia zdravotnej starostlivosti a iné príslušné subjekty pochopili, že riadenie kvality údajov je dôležitý prvok zabezpečenia bezpečnej a účinnej zdravotnej starostlivosti.

Nakoniec, riadenie kvality údajov v zdravotníctve má aj širší vplyv na spoločnosť ako celok. Správne spravované zdravotné údaje môžu poskytnúť cenné informácie pre výskum a vývoj nových liekov a terapií. Tieto údaje môžu byť tiež použité na vytvorenie programov a politík zameraných na zlepšenie verejného zdravia. Preto je dôležité, aby sa riadenie kvality údajov v zdravotníctve bralo vážne a zabezpečilo sa, aby tieto údaje boli presné, spoľahlivé, súdržné a jedinečné.

Kvalita údajov v zdravotníctve má veľký vplyv na jednotlivca, ktorý je pod liečbou, pretože informácie získané o pacientovi, vrátane správ a záznamov, ovplyvňujú liečbu, ktorú mu poskytujú poskytovatelia zdravotnej starostlivosti. Preto je dôležité, aby organizácie a subjekty, ktoré pracujú s lekárskymi údajmi, mali kvalitu týchto údajov na zreteli a dodržiavali prísne predpisy a ochranné opatrenia, aby chránili citlivé informácie pacientov a zlepšovali výsledky ich liečby.

Jedným z nástrojov na zlepšenie kvality údajov v zdravotníctve je integrovaná analýza údajov, ktorá umožňuje minimalizovať chyby, zlepšiť správu údajov a celkový dátový proces. Tento proces sa delí na tri fázy. Prvá fáza zahŕňa zachytávanie a zber údajov. Druhá fáza zahŕňa štruktúrovanie údajov, ktoré zahŕňa formátovanie a triedenie prijatých dát. Posledná fáza, fáza prenosu, zahŕňa prenos údajov z vopred určeného úložiska do koncovej databázy.

Okrem integrovanej analýzy údajov sa dajú použiť aj ďalšie metódy na zlepšenie kvality údajov v zdravotníctve, napríklad kvantifikácia a kvalifikácia údajov. Je veľmi dôležité, aby organizácie a subjekty, ktoré pracujú s lekárskymi údajmi, dodržiavali prísne predpisy v oblasti zdravotníctva a aby údaje boli v správnom formáte.

4.2.5.3 Používanie nástrojov na kvantifikáciu a kvalifikáciu údajov

Štatistické vyhodnotenie kolekcie dát je dôležité pre zabezpečenie kvality údajov v zdravotníctve. To zahŕňa výber správnych nástrojov a metrík, ktoré majú byť použité na získanie potrebnej informácie z dát v optimálnej miere. Preto je kvalifikovaný výber štatistických nástrojov rozhodujúci pre celkovú štruktúru spôsobu využitia údajov. Kvalita údajov v zdravotníctve je taktiež naviazaná na ich konzistentnosť, presnosť, správnosť a ich časovú disponibilitu. Preto je dôležité poskytovať údaje načas a vo vhodnom formáte, aby boli jednoducho použiteľné a spracovateľné organizáciou. Existuje niekoľko ďalších metód zlepšovania a významu kvality údajov v zdravotníctve. Jednou z nich je zabezpečenie toho, aby sa záznamy údajov vykonávali v súlade so zásadami ochrany údajov, ale zároveň sa nezanedbávali ciele, ktoré sú vopred definované. Ďalšou metódou je kategorizácia údajov, čo zabezpečí, že potreby pacientov budú splnené včas a bez prekážok. Okrem toho je dôležité, aby správne údaje boli vkladané do správneho poľa, aby boli vždy usporiadané a v prípade potreby ľahšie prehľadávateľné. Všetky tieto metódy môžu prispieť k zlepšeniu a zabezpečeniu kvality údajov v zdravotníctve.

Organizácie sa líšia v štruktúrach a metódach zabezpečenia kvality údajov v zdravotníctve, pretože každá organizácia je jedinečná a má svoje vlastné opatrenia. Avšak existujú niektoré základné tipy a triky, ktoré platia pre všetkých. Jedným z prístupov je kontrola zdravotných záznamov počas období mimo pracovnej náplne a vyčlenenie špecializovaných kapacít, ktoré majú lekárske znalosti na udržiavanie kvality zdravotných záznamov.

Je dôležité overiť údaje získané od pacienta a vyhodnotiť ich zdroje. Napríklad všetky demografické údaje získané od pacienta sa musia overiť pred konzultáciou a aj po nej. Pacientovi by sa mal odovzdať výtlačok daných informácií na overenie, aby sa zistilo, či sú údaje presné alebo nie. Pravidelné audity zdravotnej dokumentácie sú tiež dôležité pre overenie kvality údajov.

Udržanie aktualizácie údajov je kľúčové pre zachovanie ich kvality. Jedným z prístupov je archivácia záznamov pacientov, ktorí sú buď neaktívni alebo zomreli. Zabezpečenie toho, aby sa záznamy údajov vykonávali v súlade so zásadami ochrany údajov, ale zároveň sa nezanedbávali ciele, ktoré sú vopred definované, je ďalším spôsobom zabezpečenia kvality údajov v zdravotníctve. Kategorizácia údajov je tiež dôležitá, aby sa zabezpečilo, že potreby pacientov budú splnené včas a bez prekážok, a aby sa správne údaje zadávali do správnych polí a zabezpečila sa ich usporiadanosť a prehľadnosť.

Nástroje kvality v zdravotníctve pozostávajú z procesov a plánov, ktoré umožňujú zdravotníckym subjektom pochopiť a analyzovať ich úsilie o udržanie kvality dát. Medzi nástroje na zlepšovanie kvality patrí zameranie na klienta, pretože hlavným cieľom zdravotníckych zariadení je uspokojovať potreby svojich klientov. To znamená, že kroky, ktoré nevedú k spokojnosti klienta, sa musia eliminovať a efektívnosť sa musí zvýšiť na plný potenciál. Okrem toho je dôležité sa zamerať na systémové procesy, pretože riadenie kvality závisí od systémov, ktoré usporadúvajú ľudí, nástroje a postupy. Ďalším nástrojom je zameranie sa na tímovú prácu, pretože aby organizácia dosiahla úspech, musí každý tím organizácie spolupracovať. Vzájomne závislé systémy, ktoré udržiavajú kvalitu údajov, sa dajú lepšie prevádzkovať, keď medzi tímami v organizácii prebieha neprerušovaná komunikácia.

4.2.5.4 Dôsledky nízkej kvality údajov v zdravotníctve

Zlá kvalita údajov má vážne dôsledky na fungovanie zdravotníckej organizácie a všetkých zainteresovaných subjektov. Negatívne ovplyvňuje každé oddelenie a aspekt činnosti. Nedostatky v údajoch môžu viesť k chybným rozhodnutiam, nesprávnym diagnózam, nesprávnemu liečeniu pacientov a zvýšenému riziku chýb. Správne riadenie kvality údajov je nevyhnutné na zabezpečenie spoľahlivých, presných a aktuálnych informácií. To znamená, že je potrebné mať účinné procesy a postupy na zber, spracovanie, uchovávanie a zdieľanie údajov. Súčasne je potrebné mať aj kvalifikovaný personál, ktorý je schopný správne interpretovať a používať údaje. Len tak je možné dosiahnuť efektívnu a bezpečnú zdravotnú starostlivosť pre pacientov. Organizácie musia preto venovať primeranú pozornosť riadeniu kvality údajov a investovať do potrebných zdrojov a technológií na zabezpečenie vysokých štandardov údajov a informačnej bezpečnosti.

Nesprávne zaobchádzanie s pacientmi je jedným z dôsledkov zlej kvality údajov v zdravotníctve. Po vložení nepresných údajov do elektronického zdravotného záznamu je vykonávanie úprav zložité a zdĺhavé. V praxi to znamená, že pracovník musí manuálne zasiahnuť a riešiť nezrovnalosti v zázname. Ak sú v zázname nepresné informácie, existuje riziko nesprávneho ošetrenia a nesprávnej diagnózy, čo vedie k negatívnym skúsenostiam pacientov a ich nedôvere v systém zdravotnej starostlivosti.

Zlá kvalita údajov v zdravotníctve má za následok zvýšený počet úzkych miest. Manuálne zásahy, ktoré sa vykonávajú na opravu týchto zdravotných záznamov, vedú k väčšej pravdepodobnosti vzniku ďalších chýb. Okrem toho sú tieto manuálne zásahy časovo náročné, čo spomaľuje celkovú prevádzku a vedie k neefektívnosti. Avšak existujú systematickejšie prístupy, ktoré umožňujú predchádzať týmto nedôsledným operáciám a minimalizovať ich vplyv.

Zlá kvalita údajov v zdravotníctve vedie k nesprávnym rozhodnutiam. S narastajúcou závislosťou poskytovateľov zdravotnej starostlivosti na elektronických zdravotných záznamoch je kritické, aby údaje boli presné a spoľahlivé. Akékoľvek nezrovnalosti alebo nepresnosti v údajoch môžu mať vážne následky, pretože poskytovateľom zdravotnej starostlivosti bránia v prijímaní informovaných rozhodnutí. Tieto nepresnosti môžu viesť k nesprávnym diagnózam, nesprávnemu liečeniu alebo nevhodnému manažmentu pacientov, čo negatívne ovplyvňuje celkovú kvalitu starostlivosti a bezpečnosť pacientov. Dôsledkom je nedôvera pacientov voči poskytovateľom zdravotnej starostlivosti a ich systémom, čo má vplyv na vzťah medzi poskytovateľmi a pacientmi.

Frustrácia zamestnancov v zdravotníctve sa zvyšuje v dôsledku potreby častého zásahu pri oprave nepresných údajov. Tento proces vyžaduje čas a úsilie zamestnancov, čo vedie k ich frustrácii a nespokojnosti. Neustále opravy nepresných údajov narušujú dôveru zamestnancov v informácie poskytované prostredníctvom IT systému. Zamestnanci začínajú strácať vieru v presnosť a spoľahlivosť systému a preto sa viac spoliehajú na manuálne vykazovanie údajov. Táto situácia vytvára priestor pre chyby a nepresnosti, čo ďalej zvyšuje neefektívnosť a zhoršuje kvalitu práce zamestnancov v zdravotníctve. Celkovým výsledkom je nezdravé pracovné prostredie a znížená výkonnosť, čo negatívne ovplyvňuje poskytovanie zdravotnej starostlivosti a pacientov.

4.2.6 Ochrana citlivých údajov v kontexte zdravotných dát

Ochrana osobných údajov je zásadnou prioritou pre všetky systémy, ktoré zhromažďujú, spracovávajú a uchovávajú informácie o jednotlivcoch. Tieto systémy majú za úlohu zabezpečiť, že tieto informácie sú dostupné len pre oprávnené osoby a že ich identifikovateľnosť zo strany neoprávnených subjektov je minimalizovaná. Ochrana osobných údajov zahŕňa opatrenia na zabezpečenie dôvernosti, integrity a dostupnosti týchto údajov. Systémy musia byť navrhnuté a implementované takým spôsobom, aby minimalizovali riziko neoprávneného prístupu, úniku alebo zneužitia osobných údajov. Patria sem technické a organizačné opatrenia, ako je šifrovanie dát, silné autentifikácie, prístupové práva, bezpečnostné audity a dôkladné riadenie prístupov. Ochrana osobných údajov je nevyhnutná na ochranu súkromia jednotlivcov a zabezpečenie dôvery v systémy, ktoré s týmito údajmi pracujú.

Moderné systémy, ktoré podporujú súčasné štatistické modely pre regresné a klasifikačné modelovanie zdravotných údajov, majú v dnešnej dobe veľký potenciál. Avšak získanie týchto dát a metadát o pacientoch a klientoch, ktoré tieto systémy zhromažďujú, predstavuje potenciálny zdroj peňazí pre jednotlivcov aj organizovaných kybernetických zločincov. Bohužiaľ, ochrana týchto údajov často zaostáva. Legislatívne opatrenia sú nedostatočné, vrátane definície pojmu osobných údajov a ich uplatňovania v praxi. Operatívne opatrenia týkajúce sa spracovania a uchovávania údajov, ich anonymizácie a fyzického prístupu k nim sú tiež nedostatočné. V posledných rokoch sme boli svedkami prudkého nárastu kybernetických útokov na vládne databázy európskych štátov. Tento trend môže byť spôsobený zvyšujúcou sa počítačovou gramotnosťou ľudí, poklesom ceny výpočtovej techniky a zvýšením jej výkonu, ako aj uvoľňovaním štátnych údajov prostredníctvom iniciatív na voľný prístup k informáciám (OpenData) bez dostatočného anonymizácie. Niektoré krajiny, vrátane Veľkej Británie, Ruska, Nórska a Nemecka, uznali vážnosť situácie a zaviedli otvorenú diskusiu o ochrane osobných údajov a financujú vnútorné projekty na modernizáciu systémov ochrany štátnych databáz.

V nasledujúcich stranách sa budeme venovať problematike ochrany osobných údajov najprv z teoretického hľadiska. Skúmame základné princípy a koncepty spojené s anonymizáciou údajov a identifikovateľnosťou jednotlivcov. Sústreďujeme sa na limity anonymizácie a vymedzujeme rôzne prístupy k riešeniu tohto problému.

Ďalej sa zameriavame na aplikáciu týchto princípov a prístupov v kontexte systémov zdravotných dát. Analyzujeme, ako je problematika ochrany osobných údajov špecifická v oblasti zdravotníctva, kde sa zhromažďujú citlivé informácie o pacientoch a ich zdravotnom stave. Skúmame výzvy a riziká, ktoré sú spojené s ochranou týchto údajov v zdravotníckom prostredí.

Venujeme sa aj možným riešeniam a opatreniam, ktoré sa používajú na ochranu osobných údajov v systémoch zdravotných dát. Preskúmavame technologické metódy, ako je šifrovanie, anonymizácia, pseudonymizácia a prístupové kontroly. Diskutujeme o ich výhodách, obmedzeniach a možných výzvach pri ich implementácii.

Na záver sa zaoberáme aj etickými a právnymi aspektmi ochrany osobných údajov v systémoch zdravotných dát. Skúmame súčasnú legislatívu a regulácie, ktoré majú za cieľ zabezpečiť ochranu súkromia a bezpečnosť údajov. Rovnako analyzujeme dôležitosť etického a zodpovedného správania sa v rámci spracovania a využívania zdravotných údajov.

Celkovo si kladieme za cieľ poskytnúť komplexný pohľad na problematiku ochrany osobných údajov v kontexte systémov zdravotných dát. Skúmame jej teoretické základy, aplikáciu v praxi a etické aspekty, s dôrazom na výzvy a riešenia v tejto špecifické oblasti.

4.2.6.1 Osobné údaje

Pojem "osobný údaj" je v slovenskej legislatíve definovaný zákonom číslo 122/2013 Z. z. o ochrane osobných údajov. Tento zákon zabezpečuje súlad so Smernicou 95/46/ES Európskeho parlamentu a Rady Európskej únie, ktorá sa týka ochrany jednotlivcov pri spracovaní osobných údajov a voľnom pohybe týchto údajov.

Legislatívne opatrenia a smernice, ako je Smernica 95/46/ES, majú za cieľ poskytnúť právnu ochranu a reguláciu týkajúcu sa spracovania osobných údajov. Ich zavedenie je dôležité z hľadiska zabezpečenia súkromia, bezpečnosti a dôveryhodnosti údajov jednotlivcov. Slovenský zákon o ochrane osobných údajov je v súlade s touto smernicou, čo znamená, že uplatňuje a dodržiava štandardy a požiadavky týkajúce sa ochrany osobných údajov v rámci Európskej únie.

Tento zákon definuje, aké typy informácií sa považujú za osobné údaje a stanovuje práva jednotlivcov v súvislosti s ich spracovaním. Tiež určuje povinnosti subjektov, ktoré spracúvajú osobné údaje, a vyžaduje dodržiavanie určitých bezpečnostných a technických opatrení na ochranu týchto údajov.

Cieľom zákona o ochrane osobných údajov je zabezpečiť, aby sa osobné údaje spracovávali zákonným a transparentným spôsobom, s dodržiavaním práv jednotlivcov a zároveň zabezpečením ochrany osobných údajov pred neoprávneným prístupom, zneužitím a neprimeraným spracovaním. Kompatibilita s európskou smernicou zabezpečuje, že slovenská legislatíva je v súlade so štandardmi ochrany osobných údajov v celej Európskej únii.

Definícia pojmu "osobný údaj" v plnom znení je nasledovná: Pod pojmom "osobné údaje" sa rozumie údaje, ktoré sa vzťahujú k určenej alebo určiteľnej fyzickej osobe. Fyzickou osobou sa rozumie osoba, ktorú je možné priamo alebo nepriamo identifikovať, najmä na základe všeobecne používaného identifikátora alebo na základe jednej či viacerých charakteristík či znakov, ktoré tvoria jej fyzickú, fyziologickú, psychickú, mentálnu, ekonomickú, kultúrnu alebo sociálnu identitu.

Táto definícia poskytuje širší obraz o tom, aké typy údajov sa považujú za osobné údaje a aké aspekty identifikácie jednotlivca sú braté do úvahy. Osobné údaje môžu zahŕňať rôzne informácie o jednotlivcoch, ako napríklad ich mená, adresy, dátumy narodenia, identifikačné čísla, fyzické charakteristiky, zdravotné údaje, hospodárske informácie, kultúrnu príslušnosť a sociálne aspekty ich identity.

Definícia je v súlade s ochranou osobných údajov a zohľadňuje širokú škálu informácií, ktoré môžu byť považované za osobné údaje. Jej cieľom je zabezpečiť, že subjekty, ktoré zhromažďujú a spracovávajú osobné údaje, budú dodržiavať zákonné povinnosti a zabezpečia primeranú ochranu týchto údajov, aby sa minimalizoval riziko neoprávneného použitia alebo zneužitia osobných informácií o jednotlivcoch.

Definícia údaja znamená jednotlivý fakt alebo informáciu, zatiaľ čo osobný údaj je zloženým súborom týchto faktov, ktoré umožňujú identifikáciu alebo určenie osoby. Teda, osobné údaje sú skladané z rôznych údajov, ktoré dohromady umožňujú identifikáciu konkrétnej osoby. Je dôležité si uvedomiť, že akýkoľvek náhodný súbor údajov môže byť považovaný za osobné údaje, ak je možné na základe týchto údajov priamo alebo nepriamo identifikovať jednotlivca.

Táto charakteristika osobných údajov nám ukazuje, že osobný údaj nie je len jednotlivým faktom, ale súborom údajov, ktoré majú potenciál identifikovať konkrétnu osobu. V praxi to znamená, že osobné údaje môžu pozostávať z rôznych informácií a sú relevantné vtedy, keď spolu umožňujú jednoducho alebo s dostatočnou istotou určiť identitu jednotlivca.

Definícia je dôležitá pre ochranu osobných údajov, pretože nám pomáha rozlíšiť medzi obyčajnými údajmi a osobnými údajmi, a tým určiť, ktoré informácie je potrebné chrániť z hľadiska súkromia a bezpečnosti. Identifikovateľnosť a určiteľnosť jednotlivca na základe údajov je kľúčovým aspektom pri rozhodovaní o tom, či ide o osobné údaje a akým spôsobom je potrebné s nimi zaobchádzať z hľadiska ich spracovania a ochrany.

4.2.6.2 Náročná aplikácia legislatívnych procesov v praxi

Spomínaná definícia osobných údajov prináša dôležitý rámec pre identifikáciu údajov, ktoré môžu byť považované za osobné. Avšak, v praxi je potrebné si uvedomiť, že charakterizácia údajov ako osobných závisí od ich kontextu. To znamená, že rovnaký údaj môže byť považovaný za osobný alebo nepersonálny v závislosti od situácie, v ktorej je uvedený.

Napríklad, ak sa údaj týka náboženského presvedčenia a nachádza sa v vzorke, kde väčšina ľudí vyznáva tú istú vieru, takýto údaj by v tomto kontexte malú identifikačnú hodnotu, pretože by nebolo možné jednoznačne identifikovať jednotlivca. Avšak, ak by tento údaj patril k jednej z mála osôb v vzorke s týmto náboženským presvedčením, mohol by sa stať dôležitým nástrojom na identifikáciu tejto osoby. V takom prípade by bol považovaný za osobný údaj podľa legislatívnej definície.

Tento príklad poukazuje na to, že samotná definícia osobných údajov nie je absolútna a ich klasifikácia závisí od kontextu a okolností. Údaje, ktoré by sme pôvodne nepovažovali za osobné, môžu sa stať osobnými, ak ich kombinujeme s inými údajmi alebo ak sú v určitom kontexte vzácne a unikátne.

Preto je dôležité pri spracovaní a zdieľaní údajov dbať na ich kontext a zohľadniť potenciálne riziká identifikácie jednotlivca. Bez ohľadu na to, či ide o zdanlivo nevýznamné informácie, je potrebné zabezpečiť primeranú ochranu údajov a dodržiavať legislatívne požiadavky na ochranu osobných údajov, aby sme minimalizovali riziko ich zneužitia alebo nesprávneho použitia.

Je dôležité si uvedomiť, že osobné údaje nie sú obmedzené len na štandardné identifikačné informácie, ako sú mená, priezviská, rodné čísla, adresy alebo čísla poistencov. Tieto údaje sa často považujú za osobné, pretože sa v súčasnom systéme často používajú ako identifikátory. Avšak v správnom kontexte môže akýkoľvek dostupný údaj o skupine ľudí slúžiť ako identifikátor.

Problémom je, že určité údaje, ako napríklad číslo občianskeho preukazu alebo rodné číslo, samy osebe poskytujú ďalšie informácie o jednotlivcoch, ktoré v danom kontexte nemusia byť nevyhnutné, ale umožňujú tretej strane identifikovať konkrétnu osobu. Tieto údaje môžu obsahovať informácie o narodení, pohlaví, občianstve alebo iné osobné charakteristiky, ktoré môžu byť citlivé alebo súkromné.

Preto je dôležité zabezpečiť, aby sa tieto identifikačné údaje riadne chránili a spracovávali s ohľadom na ich potenciálnu identifikačnú hodnotu. Je nevyhnutné zabezpečiť, aby sa tieto údaje spracovávali iba v nevyhnutnom rozsahu a aby sa minimalizovala možnosť ich zneužitia alebo nesprávneho použitia treťou stranou.

V rámci ochrany osobných údajov je potrebné prihliadať na princíp minimálneho spracovania, čo znamená, že sa majú spracovávať len tie údaje, ktoré sú nevyhnutné pre daný účel, a zbytočne sa nemajú získavať alebo uchovávať nadbytočné informácie. Taktiež je dôležité implementovať bezpečnostné opatrenia, ako je šifrovanie údajov, prístupové práva a kontroly prístupu, aby sa minimalizovalo riziko neoprávneného prístupu k týmto údajom.

Zavedenie nového štandardu v systémoch štátnej správy by mohlo poskytnúť riešenie problému krížovej validácie dát a zvýšiť ochranu osobných údajov. Tento nový štandard by mal byť implementovaný prostredníctvom spolupráce viacerých ministerstiev Slovenskej republiky a mal by slúžiť ako identifikátor pre všetky evidované osoby, napríklad v registri obyvateľov.

Ideálne by sa jednalo o alfanumerickú sadu znakov, ktorá by bola priradená každej osobe a neobsahovala žiadne informácie popisujúce jej charakteristiky. Na rozdiel od súčasných systémových identifikátorov, ako je rodné číslo, by tento nový identifikátor neumožňoval odvodiť informácie o pohlaví, veku, rase alebo iných osobných charakteristikách.

Takýto nový identifikátor by poskytoval menej príležitostí pre zneužitie pri krížovej validácii a bol by vhodnejší pre proces anonymizácie dát. Využitie tohto nového identifikátora by zabezpečovalo, že osobné údaje by boli spracovávané s minimálnym rizikom ich nesprávneho pripojenia alebo odhalenia citlivých informácií.

Implementácia tohto nového štandardu by vyžadovala úzku spoluprácu medzi ministerstvami a dôkladné plánovanie, aby sa zabezpečila jeho účinná implementácia a dodržiavanie legislatívnych noriem ochrany osobných údajov. Zavedenie takéhoto nového identifikátora by predstavovalo významný krok v zlepšovaní ochrany osobných údajov a minimalizovaní rizika ich zneužitia v systémoch štátnej správy.

4.2.7 Proces anonymizovania dát

Existencia kvalitného procesu anonymizácie dát v komunikačnom systéme umožňuje bezpečný prenos dôležitých informácií aj cez miesta s nižšou úrovňou informačnej bezpečnosti, ako napríklad medzi organizáciami. Proces anonymizácie zabezpečuje, že prenesené údaje sú odstránené alebo upravené takým spôsobom, že sa minimalizuje riziko zverejnenia citlivých informácií.

Anonymizovaný prenos dát prináša relatívne nízke riziko úniku citlivých informácií, pretože identifikačné údaje a iné osobné charakteristiky sú odstránené alebo nahradené tak, aby nebolo možné

jednoznačne identifikovať jednotlivcov. Týmto spôsobom je zachované súkromie a bezpečnosť údajov, čo je veľmi dôležité pri prenose informácií, najmä medzi organizáciami.

Anonymizovaný prenos umožňuje následnú analýzu a hodnotenie prenesených informácií. Aj keď osobné údaje sú anonymizované, zostávajú dostatočné informácie pre ďalšiu spracovateľskú činnosť, ktorá môže poskytnúť cenné poznatky a pohľady na dáta. Tieto analýzy môžu byť použité na výskum, plánovanie a rozhodovanie, pričom citlivé informácie zostávajú chránené.

Dôležitým aspektom procesu anonymizácie je zabezpečiť, že identifikovateľné informácie sú dostatočne zašifrované alebo nahradené nereverzibilnými metódami, ktoré neumožňujú ich obnovenie do pôvodnej formy. To zaisťuje, že aj keď by sa dáta dostali do nesprávnych rúk, je nemožné získať pôvodné osobné informácie.

Spoločnosti a organizácie, ktoré využívajú komunikačné systémy s kvalitným procesom anonymizácie dát, môžu mať istotu, že prenesené informácie sú bezpečné a chránené pred neoprávneným prístupom. Tento proces poskytuje rovnováhu medzi ochranou osobných údajov a potrebou využívať dáta pre analytické a hodnotiace účely.

Anonymizácia dát je všeobecne proces, ktorým sa transformujú čisté alebo surové dáta do sofistikovanej podoby, ktorá neumožňuje reverznou manipuláciou získať pôvodné surové dáta ani identifikovať konkrétnu entitu, ako je napríklad osoba alebo skupina osôb. Hlavným cieľom anonymizácie je zachovať súkromie a bezpečnosť údajov, zatiaľ čo umožňuje ich použitie pre rôzne účely, ako je výskum, štatistika a analýza.

Proces anonymizácie môže zahŕňať rôzne metódy a techniky, ako je odstraňovanie priamo identifikujúcich informácií, nahradenie identifikačných údajov generickými hodnotami alebo pseudonymizácia, pri ktorej sa priraďuje náhodný identifikátor k jednotlivcom. Ďalšie postupy zahŕňajú agregáciu údajov, kedy sa kombinujú informácie o viacerých jednotlivcoch do skupín alebo kategórií, a generalizáciu, kedy sa presne hodnoty nahradzujú rozsahmi alebo intervalmi.

Cieľom týchto techník je znemožniť identifikáciu jednotlivcov z anonymizovaných dát. Napríklad, ak by sme mali súbor údajov obsahujúci mená a adresy ľudí, anonymizácia by mohla spočívať v odstránení mien a presných adries a ich nahradení generickými identifikátormi. Týmto spôsobom sa údaje stávajú anonymnými, pretože nemôžu byť priamo priradené k špecifickým osobám.

Dôležité je tiež zabezpečiť, aby anonymizované dáta zostali dostatočne relevantné a užitočné pre ich zamýšľané použitie. Je potrebné vyvážiť úroveň anonymizácie tak, aby sa minimalizovalo riziko identifikácie, zatiaľ čo sa zároveň zachovala hodnota a kvalita dát pre analýzu a výskum.

Anonymizácia dát je neoddeliteľnou súčasťou ochrany osobných údajov a zabezpečuje dodržiavanie súkromia jednotlivcov.

Je potrebné doplniť, že napriek anonymizácii dát podľa definícií môže existovať možnosť identifikácie entít, ak je to potrebné alebo ak existujú dodatočné informácie, ktoré sa nedostatočne zohľadnili pri anonymizácii. To znamená, že aj keď sa anonymizácia zameriava na ochranu konkrétnej entity, môže stále existovať riziko identifikácie iných entít, ak sú získané dodatočné informácie alebo ak sú použité ďalšie techniky na kombinovanie údajov.

Pri anonymizácii dát sa používa rôznorodé spektrum techník, ktoré majú za cieľ zachovať štruktúru dátového poľa, ako je veľkosť, pozícia, typ a frekvencia údajov. Dôležité je, aby aj po anonymizácii dáta pôsobili realisticky a mohli byť použité v testovacích prostrediach alebo analytických procesoch. Tým sa zabezpečuje, že anonymizované dáta sú stále užitočné pre rôzne účely, ako je napríklad testovanie softvéru, výskum a analýza.

Treba však zdôrazniť, že absolútne anonymizované dáta, kde sú všetky hodnoty vynulované, majú veľmi obmedzenú analytickú hodnotu. Anonymizácia obmedzuje hĺbku a detaily údajov, čo znižuje možnosti ich analýzy a význam pre rôzne aplikácie. Preto je dôležité nájsť kompromis medzi bezpečnosťou a úžitkom dát pri procese anonymizácie.

4.2.7.1 Základné metódy anonymizovania dát

Metóda vytvorenia náhradných symbolov je jednou z intuitívnych metód anonymizácie. Pri tejto metóde sa čísla alebo identifikačné údaje nahradia náhodne vygenerovanými symbolmi, čo výrazne komplikuje úlohu útočníka pri identifikácii pôvodných dát. Napríklad nahradenie čísel dátumu narodenia náhodnými znakmi robí anonymizované údaje prakticky neprekonateľnými.

Treba však zdôrazniť, že táto metóda anonymizácie má svoje obmedzenia. Nahradenie dát náhodnými symbolmi vedie k výraznému zníženiu informačnej hodnoty dát. Analytici a výskumníci budú mať obmedzené možnosti využitia takýchto anonymizovaných dát pre analýzu a výskum.

Dôležité je rozlišovať medzi vytvorením náhradných dát pomocou kľúča a náhodným priraďovaním symbolov. V prípade vytvorenia náhradných dát podľa kľúča, kde je k dispozícii špeciálny kľúč, je proces anonymizácie skôr považovaný za pseudo-anonymizáciu. Pretože ak útočník získa prístup k tomuto kľúču, môže získať pôvodné dáta. Existencia kľúča zvyšuje pravdepodobnosť úspešného útoku na anonymizované dáta a zároveň zvyšuje informačný potenciál týchto dát.

Je dôležité brať do úvahy aj bezpečnostné aspekty a riziká spojené s existenciou kľúča a prístupu k pôvodným dátam. Je potrebné zabezpečiť, aby takýto kľúč bol správne chránený a aby sa minimalizovala pravdepodobnosť jeho zneužitia.

Metóda nulovania dát je druhou metódou anonymizácie, ktorá predstavuje radikálny prístup. Pri tejto metóde sa citlivým dátam priradí hodnota N/A alebo NULL, čím sa prakticky vymažú informácie obsiahnuté v týchto údajoch. Týmto spôsobom sa zabezpečuje maximálna ochrana proti nežiaducemu odhaleniu citlivých informácií. Avšak v dôsledku nulovania dát sa informačná hodnota dát výrazne znižuje, čo môže mať negatívny vplyv na ich analyzovateľnosť a využiteľnosť.

Je dôležité si uvedomiť, že metóda nulovania dát nie je vhodná pre prípady, kde dáta majú byť podrobené komplexnejšej post-anonymizačnej analýze. Z takto anonymizovaných dát je možné vyčítať iba základné štatistiky, ako je frekvencia výskytu v čase a ich celkový objem. Podrobné informácie a vzťahy medzi dátami sú v rámci tejto metódy prakticky nedostupné.

Ďalším drastickým prístupom k anonymizácii je odovzdávanie dát vo forme agregovaných údajov. Pri tejto metóde sa surové dáta spracujú a zmenšia ich granularitu, napríklad pomocou priemerovania alebo iných agregačných funkcií. Tieto agregované údaje sa potom odovzdávajú tretej strane, pričom tá nemá podrobné informácie o jednotlivých záznamoch. Tento prístup je obzvlášť vhodný v prípadoch, keď nemáme kontrolu nad tým, ako tretia strana bude dáta používať a akým spôsobom ich bude analyzovať.

Pri anonymizácii je vždy potrebné zvážiť kompromis medzi bezpečnosťou a analyzovateľnosťou dát. Rôzne metódy anonymizácie majú svoje výhody a obmedzenia, a preto je dôležité vybrať vhodný prístup v závislosti od konkrétneho prípadu a požiadaviek na ochranu citlivých informácií.

Metóda tvorby virtuálnych dátových vzoriek je tretou metódou anonymizácie, ktorá sa využíva na zabezpečenie ochrany citlivých informácií. Virtuálne dátové vzorky sú vytvorené takým spôsobom, že majú rovnaké charakteristiky ako pôvodná vzorka dát, ako napríklad poradie, typ, frekvencia a rozloženie údajov. Avšak v rámci tejto metódy sú hodnoty jednotlivých atribútov nahradené inými hodnotami, ktoré nemajú žiadne spojenie s pôvodnými skutočnými dátami.

Využitie virtuálnych dátových vzoriek vyžaduje vytvorenie knižnice, z ktorej sa budú tieto virtuálne údaje vyberať a priraďovať k jednotlivým reálnym dátam. Taktiež je potrebné definovať logické pravidlá, podľa ktorých budú virtuálne hodnoty priraďované jednotlivým stĺpcom. Týmto spôsobom je možné dosiahnuť, že pôvodné dáta budú anonymizované a zachovajú si svoju informačnú hodnotu.

Anonymizovanie dát pomocou virtuálnych vzoriek poskytuje dobrý kompromis medzi zachovaním informačnej hodnoty dát a zabezpečením ich bezpečnosti. Keďže virtuálne dáta sú navrhnuté tak, aby zachovávali štatistické charakteristiky a vzťahy medzi dátami, poskytujú dostatočne reprezentatívne

údaje pre analýzu a spracovanie. Zároveň však zabezpečujú, že pôvodné identifikovateľné informácie sú nahradené fiktívnymi hodnotami, čím sa minimalizuje riziko odhalenia citlivých údajov.

Využitie metódy virtuálnych dátových vzoriek vyžaduje starostlivé navrhovanie a implementáciu, aby sa zabezpečila primeraná úroveň anonymizácie a ochrany dát. Správne použitie tejto metódy v kombinácii s ďalšími opatreniami môže prispieť k efektívnemu a bezpečnému spracovaniu citlivých informácií.

Pri aplikácii metódy virtuálnych vzoriek je dôležité vyhnúť sa klasickým chybám, ako je data overflow. K tomu dochádza, keď nedodržíme charakteristiky pôvodnej vzorky pri vytváraní virtuálnej vzorky. To môže zahŕňať zmeny v počte riadkov a stĺpcov alebo v distribúcii údajov. Ak chceme úspešne virtualizovať naše dáta, je potrebné najskôr popísať ich vlastnosti, ktoré chceme zachovať. Napríklad, ak sú rodné čísla v našich dátach deliteľné číslom 11, pri virtualizácii a použití týchto čísel ako identifikátorov v systéme, ktorý vyžaduje deliteľnosť jedenástimi, je dôležité zachovať túto charakteristiku aj vo virtuálnych dátach.

Z hľadiska bezpečnosti je výhodné virtualizovať dáta tak, aby boli na prvý pohľad nerozoznateľné od pôvodných dát. Pre číselné hodnoty môžeme použiť podobnú sústavu nahradenia, pre geografické názvy môžeme zachovať podobné geografické názvy namiesto náhodných kombinácií písmen. Dôležité je, aby virtuálne dáta zachovávali realistický vzhľad, čo znamená, že by mali mať podobné rozloženie a charakteristiky ako pôvodné dáta.

Pri tvorbe virtuálnych vzoriek je preto potrebné mať na pamäti nielen bezpečnostné aspekty, ale aj zachovanie dôležitých charakteristík a štruktúry pôvodných dát. Správne použitie tejto metódy umožňuje dosiahnuť vysokú úroveň anonymizácie a ochrany citlivých informácií, zatiaľ čo zároveň poskytuje dostatočne reprezentatívne dáta pre analytické účely.

Metóda shufflingu predstavuje štvrtú metódu anonymizácie. Spočíva v preskupení pôvodnej vzorky dát podľa preddefinovaných osí a konkrétnych premenných alebo dátových reťazcov. Táto metóda vyžaduje predchádzajúcu prípravu a zváženie jej vhodnosti pre konkrétny prípad. Aby bola efektívna, je potrebné, aby preskupovanie dát bolo dostatočne zložité. Je dôležité, aby podobné premenné neboli prehádzané, ak je v dátach iba málo takýchto premenných, pretože to môže uľahčiť de-anonymizáciu. Metódu shufflingu neodporúčame používať pri vzorkách s malým počtom premenných. Pri dostatočne veľkých vzorkách, kde dáta zahŕňajú názvy, mená alebo iné nominálne premenné, poskytuje anonymizácia pomocou shufflingu lepší výsledok z hľadiska informačnej hodnoty dát v porovnaní s anonymizáciou pomocou virtuálnych vzoriek.

Pri anonymizácii numerických intervalových hodnôt a niektorých ordinálnych premenných je možné využiť metódu posunutia. Táto metóda spočíva v posunutí hodnôt o preddefinovaný interval alebo náhodný interval. Manipuláciou s veľkosťou intervalu môžeme ovplyvniť zachovanie štatistickej hodnoty dát. Týmto spôsobom je možné dosiahnuť anonymizáciu týchto typov premenných a zároveň zachovať ich analytickú hodnotu.

Pri použití metódy shufflingu a posunutia je dôležité dbať na vyvážený prístup medzi anonymizáciou a zachovaním informačnej hodnoty dát. Cieľom je minimalizovať riziko zneužitia citlivých informácií, zároveň však umožniť analytické využitie dát v bezpečnom prostredí.

Pri anonymizácii dát, a najmä pri použití metódy shufflingu, je dôležité venovať pozornosť hodnotám, ktoré výrazne odchádzajú od priemeru a nezodpovedajú štandardnému rozloženiu dát vo vzorke. Tieto hodnoty môžu byť unikátne a nereprezentatívne pre bežné dátové rozdelenie. Je tu pravidlo, že čím viac sú dáta v vzorke monotónne, tým jednoduchšie je vzorku anonymizovať. Naopak, ak v vzorke existuje viac štatistických výnimiek alebo odchýlok, je pre útočníka jednoduchšie extrahovať z vzorky čiastočné dáta.

Táto skutočnosť poukazuje na dôležitosť správnej analýzy a pochopenia charakteru dát pred začatím anonymizácie. Identifikácia a odlíšenie významných odchýlok od priemeru môže byť kritická pri

rozhodovaní, ako postupovať pri anonymizácii a aké techniky použiť. Je potrebné nájsť rovnováhu medzi zabezpečením citlivých informácií a zachovaním integrovaných štatistických vlastností dát.

Z toho vyplýva, že pri anonymizácii dát je potrebné mať na pamäti nielen samotné techniky anonymizácie, ale aj charakteristiku samotných dát. Dôkladná analýza dát a ich vlastností je dôležitá pre správny výber metód anonymizácie a dosiahnutie optimálneho kompromisu medzi bezpečnosťou a informačnou hodnotou dát.

Pri implementácii základných techník anonymizácie je vhodné kombinovať jednotlivé prístupy. Týmto spôsobom môžeme využiť výhody rôznych metód a dosiahnuť značnú úroveň bezpečnosti dát, pričom stále zachovávame ich interpretatívnu hodnotu v post-anonymizačnej analýze.

Dôležité je pristupovať k návrhu a aplikácii systému anonymizácie dát systematicky. To znamená mať jasno v požiadavkách na anonymizáciu a zabezpečenie osobných informácií, ako aj v analytickom potenciáli, ktorý chceme zachovať. Každý prístup k anonymizácii má svoje výhody a obmedzenia, a preto je potrebné vybrať vhodnú kombináciu techník a metód, aby sme dosiahli žiadaný výsledok.

Pri navrhovaní systému anonymizácie je dôležité zvážiť nielen samotné techniky, ale aj procesy a postupy. To zahŕňa definovanie pravidiel anonymizácie, výber vhodných nástrojov a implementáciu týchto pravidiel do procesov manipulácie s dátami. Systém anonymizácie by mal byť flexibilný, aby mohol byť prispôsobený konkrétnym požiadavkám a charakteristikám dátového súboru.

Celkovým cieľom je nájsť kompromis medzi ochranou osobných informácií a zachovaním analytického potenciálu dát. Anonymizácia dát je iteratívny proces, ktorý vyžaduje starostlivé plánovanie, implementáciu a overovanie. Správne zvolená kombinácia techník a systematický prístup nám umožňuje dosiahnuť bezpečnosť dát a súčasne uchovať ich hodnotu pre ďalšie analytické účely.

Neriadené alebo prehnané anonymizovanie dát často nie je vhodné a môže byť dokonca kontraproduktívne z nasledujúcich dôvodov. Prvým dôvodom je, že väčšina útočníkov nemá k dispozícii všetky potrebné informácie na systematickú extrakciu údajov z anonymizovaných dát. Nie je teda nevyhnutné anonymizovať všetky údaje, keďže nie všetky informácie v databáze sú citlivé z hľadiska cieľa zberu týchto údajov. Okrem toho nie všetky dáta majú rovnakú mieru citlivosti v kontexte analýzy a interpretácie, pre ktoré sú zbierané. Niektoré vzťahy medzi premennými nemusia byť predmetom anonymizácie, napríklad vzťahy medzi používateľmi v on-line komunikačných systémoch sú často verejné, ale vzťahy používateľa k službe, ako napríklad prihlásenie do služby alebo frekvencia používania, sa považujú za senzitívne. Preto je dôležité mať od začiatku systém anonymizácie, ktorý je založený na vedomostiach o tom, ktoré informácie v dátach sú senzitívne a ktoré nie. Je potrebné presne definovať, čo sa považuje za osobný údaj a čo nie. Taktiež je dôležité dopredu plánovať pomer medzi anonymizáciou dát a stratou ich informačnej hodnoty a riadiť sa týmto plánom počas vývoja anonymizačného systému.

Na anonymizáciu dát sa využívajú aj pokročilé prístupy a nástroje, ako napríklad hashing, syntaktické bezpečnostné modely (ako K-anonymita, l-diverzita, t-bližina, e-diferenciálna súkromnosť) a systém crowd blending privacy. Každý z týchto prístupov má svoje výhody a nevýhody, často spojené s ich komplexitou.

Metóda crowd blending privacy je relatívne jednoduchá na použitie a umožňuje dynamickú anonymizáciu dát. Avšak, nemá silné záruky bezpečnosti a môže byť náchylná na útoky, ak nie sú dodržané prísne bezpečnostné opatrenia. Na druhej strane, algoritmy založené na diferenciálnej súkromnosti sú náročnejšie na návrh, najmä ak majú podporovať komplexné analytické úlohy po anonymizácii. Každá úloha vyžaduje vlastné optimalizačné metódy, aby bola zachovaná dostatočná úroveň použiteľnosti dát.

Syntaktické bezpečnostné modely, ako napríklad k-anonymita, l-diverzita a e-diferenciálna súkromnosť, tiež ponúkajú prístup k anonymizácii dát. Avšak, aj tu existuje problém s nedostatočnou úžitkovosťou dát. Každý model vyžaduje špecifickú optimalizáciu pre každú úlohu, čo môže byť náročné a časovo náročné.

Jednou z často používaných metód je hashing, avšak táto metóda môže mať svoje riziká. Proces de-anonymizácie dát môže byť relatívne jednoduchý, ak sú použité slabé algoritmy hashovania alebo ak útočník disponuje dostatočnou výpočtovou silou. Preto je dôležité venovať pozornosť výberu správnych a bezpečných hashovacích algoritmov, aby sa minimalizovalo riziko de-anonymizácie.

Pri navrhovaní anonymizačných systémov je dôležité zvážiť tieto rôzne prístupy a nástroje a zvoliť kombináciu, ktorá najlepšie vyhovuje požiadavkám na bezpečnosť a zachovanie analytického potenciálu dát.

4.2.7.2 Pseudoanonymizovanie dát

Anonymizácia dát sa často zamieňa s pseudo-anonymizáciou, čo je dôležité rozlíšiť. Pseudo-anonymizácia je proces, ktorý zabezpečuje, že osobné údaje v dátach nie sú priamo identifikovateľné, avšak je stále možné ich zosledovať alebo odhaliť ich identitu s dostatočnými vedomosťami a prístupom k ďalším informáciám.

Pri pseudo-anonymizácii sa z pôvodných surových dát odstráni priamo identifikovateľné informácie, ako sú mená, rodné čísla alebo e-mailové adresy. Namiesto toho sa použijú náhodne generované identifikátory alebo skrátené verzie pôvodných údajov. Cieľom je znemožniť priamy spoj medzi osobami a ich údajmi v dátach.

Je však dôležité si uvedomiť, že pseudo-anonymizácia nie je úplne nezvratná. Existujú techniky, ktoré umožňujú re-identifikáciu jednotlivcov alebo priblíženie k ich pôvodnej identite. Ak útočník disponuje ďalšími informáciami, napríklad z iných zdrojov, môže skombinovať tieto informácie s pseudo-anonymizovanými dátami a obnoviť pôvodnú identitu jednotlivcov.

Preto je dôležité mať na pamäti, že pseudo-anonymizácia neposkytuje úplnú anonymitu a je nevyhnutné dodržiavať ďalšie bezpečnostné opatrenia, aby sa minimalizovalo riziko de-anonymizácie. Príkladom takýchto opatrení je obmedzenie prístupu k citlivým informáciám, implementácia dodatočných bezpečnostných vrstiev alebo vytvorenie štatistických metód, ktoré zabraňujú de-anonymizácii na základe kombinácie údajov.

Celkovo je dôležité si uvedomiť, že pseudo-anonymizácia je len jedným z nástrojov v rámci širšieho procesu anonymizácie dát. Pri navrhovaní anonymizačných systémov je potrebné brať do úvahy rôzne faktory a zohľadňovať najnovšie techniky a postupy na zabezpečenie ochrany osobných údajov a zachovanie analytického potenciálu dát.

Pseudo-anonymizácia dát sa často používa ako postup na dosiahnutie rovnováhy medzi bezpečnosťou a analytickým potenciálom. V uzamknutom stave, keď sú dáta pseudo-anonymizované, majú relatívne dobrú úroveň bezpečnosti. V tomto stave je hlavným rizikom získanie kódovacieho kľúča útočníkom, čo by im umožnilo reverznú konverziu dát späť na pôvodný stav.

Po odomknutí alebo de-pseudo-anonymizácii dát je možné získať plný štatistický potenciál týchto údajov. To znamená, že dáta sú analyticky využiteľné a poskytujú cenné informácie pre rôzne analýzy, štatistické výpočty a iné úlohy. To je dôležité najmä v prípadoch, keď anonymizácia nie je nevyhnutná alebo kde pseudo-anonymizácia poskytuje dostatočnú ochranu vzhľadom na kontext a požiadavky.

Avšak, existuje nebezpečenstvo, keď ľudia predpokladajú, že ich dáta sú plne anonymizované, keď v skutočnosti sú len pseudo-anonymizované. To môže viesť k situáciám, keď tieto dáta poskytnú tretej strane, pričom sa domnievajú, že osobné údaje sú úplne skryté a ich identifikácia je nemožná. Toto predstavuje riziko pre ochranu osobných údajov a môže viesť k narušeniu súkromia jednotlivcov.

Na druhej strane, plné anonymizovanie dát je v kontexte, kde je pseudo-anonymizácia dostatočná alebo ideálna, zbytočným krokom. Plná anonymita by znamenala, že údaje sú úplne odstránené zo vzťahu k jednotlivcom a neexistuje spôsob, ako ich identifikovať. To by obmedzilo analytický potenciál dát a znemožnilo by ich využitie pri ďalších výskumoch, analýzach alebo zdieľaní s oprávnenými stranami.

Preto je dôležité mať presnú predstavu o rozdieli medzi anonymizáciou a pseudo-anonymizáciou dát a vedieť, kedy a prečo použiť jeden alebo druhý prístup.

4.2.7.3 Proces de-anonymizovania a nedostatočné anonymizovanie dát

De-anonymizácia dát je proces, ktorý prináša hrozby pre systémy, kde je anonymizácia považovaná za nevyhnutnú. Aj keď podľa definícií by malo byť anonymizovanie nezvratné, v skutočných aplikáciách existujú spôsoby, ako de-anonymizovať anonymizované dáta. Tieto metódy predstavujú reálne riziká a ohrozenie pre bezpečnosť takýchto systémov.

De-anonymizácia sa vykonáva pomocou krížovej referencie a validácie viacerých anonymizovaných databáz. Týmto spôsobom je možné extrahovať aspoň časť citlivých údajov o používateľoch systému. Existuje viacero faktorov, ktoré umožňujú de-anonymizáciu dát. Jedným z nich je narastajúci počet dát, ktoré sú považované za osobné údaje a majú charakter citlivých informácií. Ďalším faktorom je zvyšujúci sa počet verejných zdrojov dát, ktoré je možné skrížiť s anonymizovanými databázami.

Aby sme si to lepšie predstavili, uveďme príklad. Napríklad, pri krížovej referencii dát z verejných databáz USA, kde každá z nich je anonymizovaná v rámci svojho kontextu, no necháva niektoré stopy (viditeľné údaje), ktoré v tomto kontexte nie sú považované za citlivé údaje, je možné identifikovať až 87 % obyvateľov USA na základe ich PSČ, veku a pohlavia. Toto ukazuje, že len malé množstvo dostupných informácií môže byť postačujúce na identifikáciu veľkého percenta jednotlivcov.

Legislatívne rozširovanie definície "osobných údajov" nemusí byť účinným riešením na zvýšenie bezpečnosti systémov. Aj keď sa definícia rozširuje, stále existujú riziká, že anonymizované dáta môžu byť de-anonymizované pomocou dostupných informácií z iných zdrojov. Preto je dôležité vyvíjať a implementovať efektívne metódy anonymizácie, ktoré berú do úvahy tieto riziká a poskytujú dostatočnú úroveň bezpečnosti pre používateľov.

Ochrana pred de-anonymizáciou dát vyžaduje dodržanie niekoľkých podmienok. Prvým krokom je uvedomenie si a akceptovanie skutočnosti, že riziko de-anonymizácie je reálne a prítomné. Ďalšou dôležitou podmienkou je vnímanie tvorby bezpečnostného systému ako kontinuálnej snahy o minimalizáciu rizika, s dôrazom na dynamické prispôsobenie sa novým hrozbám. Komunikácia s inštitúciami a zúčastnenými stranami je nevyhnutná, aj keď to môže byť náročné, pretože požadované absolútne riešenia nie vždy sú možné alebo zmysluplné. Avšak čím viac zainteresovaných strán akceptuje a podporuje bezpečnostné opatrenia, tým jednoduchšie je vyvíjať a implementovať potrebné opatrenia. Dôležité je aj racionálne obmedziť množstvo detailov dát, aby sa minimalizovala možnosť identifikácie jednotlivcov. Počas vývoja je nevyhnutné opakovane revidovať dáta, ktoré majú byť posielané tretím stranám, a analyzovať potenciálne možnosti identifikácie týchto dát zo strany tretích strán. Tieto opatrenia prispievajú k zvýšeniu bezpečnosti a ochrane dát pred de-anonymizáciou.

Pri riešení de-anonymizácie dát je dôležité zohľadniť, či existujú segmenty dát, ktoré sú citlivejšie a obsahujú viac informácií, než je potrebné pre spracovanie a interpretáciu údajov. Ak áno, znamená to, že tieto dáta nie sú dostatočne chránené a predstavujú riziko pre anonymitu jednotlivcov. Je nevyhnutné identifikovať tieto senzitívne segmenty a prijať opatrenia na ich správne anonymizovanie a ochranu.

Okrem toho, ak je možné agregovať jednotlivcov v dátach do stabilných kohort alebo skupín, ktoré reprezentujú priemerné správanie a charakteristiky v dátach, takéto dáta sú náchylnejšie k de-anonymizácii. Ak útočník disponuje dodatočnými dátami alebo zdrojmi informácií, môže pomocou krížovej referencie a validácie identifikovať jednotlivcov v anonymizovaných dátach na základe ich unikátnych kombinácií a vzťahov voči týmto stabilným kohortám. Preto je dôležité analyzovať, aké informácie a vzťahy sú zachované v anonymizovaných dátach a zvážiť opatrenia, ktoré minimalizujú riziko de-anonymizácie, najmä pokiaľ ide o agregované údaje.

Zabezpečenie správnej ochrany týchto senzitívnych segmentov a zváženie úrovne agregácie v dátach je dôležitým aspektom pri návrhu anonymizačného procesu. Tým sa minimalizuje možnosť identifikácie jednotlivcov a zvyšuje sa celková bezpečnosť dát.

Metóda hashing sa často používa na anonymizáciu dát, no v praxi je často nedostatočne efektívna a môže viesť k problémom pri ochrane citlivých informácií. Ak si predstavíme funkciu hashing ako nástroj, ktorý prevedie vstupný údaj (napríklad rodné číslo) na unikátny identifikátor (ako napríklad "k9089jthg876fe6534"), môže sa zdať, že týmto spôsobom sa údaje úplne anonymizujú.

Avšak, existujú určité problémy spojené s touto metódou. Po prvé, hashing je deterministický proces, čo znamená, že rovnaký vstup vždy vygeneruje rovnaký výstup. To znamená, že ak máme viacero údajov s rovnakým vstupom (napríklad viacero rovnakých rodných čísel), budú mať všetky tieto údaje rovnaký anonymizovaný identifikátor. Ak teda niekto získa prístup k niektorým nepozmeneným údajom a vie, akým spôsobom je na ne aplikovaná anonymizácia, môže tieto informácie skombinovať a de-anonymizovať časť dát.

Ďalším problémom je, že v prípade použitia slabého hasovacieho algoritmu alebo krátkych identifikátorov je možné použiť tzv. "útok hrubej sily" (brute-force attack) a vytvoriť mapu vstupov a ich anonymizovaných identifikátorov. Týmto spôsobom je možné získanie pôvodných údajov na základe anonymizovaných identifikátorov.

Takisto je dôležité si uvedomiť, že hashing sám o sebe neposkytuje žiadnu kryptografickú ochranu. Anonymizované identifikátory môžu byť ľahko dešifrovateľné, ak útočník získa prístup k pôvodným údajom alebo ak pozná algoritmus hashingu a použité vstupy.

Preto je dôležité starostlivo zvoliť vhodný hasovací algoritmus a riadiť sa osvedčenými postupmi pri implementácii anonymizácie pomocou hashingu. Prípadne je lepšie kombinovať viacero metód anonymizácie a dodržiavať najnovšie bezpečnostné štandardy a odporúčania pri spracovaní a uchovávaní citlivých údajov.

Hashovacie funkcie majú určité vlastnosti, ktoré zabezpečujú ich účinnosť a bezpečnosť v kontexte anonymizácie dát. Tieto vlastnosti zahŕňajú:

a) Jednoznačnosť výstupu: Pre rovnaký vstup vždy poskytujú hashovacie funkcie rovnaký výstup. To znamená, že ak zadáme rovnaké rodné číslo do hashovacej funkcie, dostaneme vždy rovnaký anonymizovaný identifikátor. Táto vlastnosť umožňuje neskôr identifikovať rovnaké údaje bez potreby ukladať ich pôvodné hodnoty.

b) Jednosmernosť: Hashovacie funkcie sú navrhnuté tak, aby bolo nemožné z výstupného anonymizovaného identifikátora získať pôvodný vstup. To znamená, že nie je možné reverzne prepojiť anonymizované identifikátory na pôvodné údaje bez znalosti špecifického algoritmu hashovania.

c) Jedinečnosť: Hashovacie funkcie by mali mať minimálnu pravdepodobnosť kolízie, čo znamená, že pre rôzne vstupy by mali generovať rôzne anonymizované identifikátory. Tým sa minimalizuje možnosť, že dva rôzne údaje budú mať rovnaký anonymizovaný identifikátor.

Aj keď tieto vlastnosti sú užitočné pri anonymizácii dát, existujú určité hrozby, ktorým je hashing náchylný. Skúsení útočníci môžu využiť skutočnosť, že majú prístup k hashovacej funkcii a poznajú pôvodné údaje, ktoré boli hashované. Na základe tejto znalosti môžu vytvoriť "mapu" medzi anonymizovanými identifikátormi a pôvodnými údajmi. Takýmto spôsobom môžu skúsení útočníci de-anonymizovať niektoré časti dát a získať prístup k citlivým informáciám, ako je napríklad rodné číslo.

Preto je dôležité si byť vedomý týchto hrozieb a zvoliť vhodné opatrenia na ochranu dát. Jedným z prístupov môže byť použitie silnejších hashovacích algoritmov, ktoré majú nižšiu pravdepodobnosť kolízie a sú ťažšie obídateľné. Ďalším krokom môže byť kombinácia hashovania s ďalšími met

Je pravda, že skúsení útočníci môžu využiť špecifickú štruktúru rodných čísel a skúsiť de-anonymizovať dáta pomocou útokov hrubou silou. Tieto útoky spočívajú v generovaní všetkých

možných kombinácií rodných čísel a ich následnom hashovaní, aby sa porovnali s anonymizovanými identifikátormi.

Avšak dôležité je si uvedomiť, že úspešnosť tohto typu útokov závisí od rôznych faktorov, ako je dĺžka rodného čísla, dostupné zdroje výpočtového výkonu a časové obmedzenia. Ak je rodné číslo dostatočne dlhé a hashovacia funkcia je dostatočne silná, prehľadávanie všetkých možných kombinácií by mohlo trvať veľmi dlho alebo by bolo nerealistické z hľadiska výpočtových nárokov.

Napriek tomu je dôležité, aby tvorcovia a správcovia systémov, ktoré obsahujú citlivé dáta, venovali pozornosť tejto hrozbe a prijali opatrenia na minimalizáciu rizika de-anonymizácie. Jedným z prístupov môže byť použitie dlhších a komplexnejších identifikátorov (napr. UUID), ktoré neodrážajú žiadne štruktúry pôvodných údajov. Taktiež je dôležité výberom silnej hashovacej funkcie, ktorá je odolná voči útokom hrubou silou a kolíziám.

Dodržiavanie správnych bezpečnostných postupov, ako je ochrana prístupu k hashovacej funkcii a implementácia ďalších ochranných opatrení, môže ďalej zvýšiť bezpečnosť a minimalizovať riziko úspešného de-anonymizovania dát. Je dôležité si uvedomiť, že žiadna metóda anonymizácie nie je stopercentne bezpečná a je neustále potrebné aktualizovať a zlepšovať opatrenia s cieľom minimalizovať riziko úniku citlivých informácií.

4.2.8 Dizajn bezpečnostného systému

Pri dizajne a vývoji systému, ktorý spracováva dátový tok, je dôležité dodržiavať niekoľko zásad a princípov. Jedným z týchto princípov je interoperabilita existujúceho a vyvíjaného softvéru. Systémy musia byť schopné reagovať na rôzne typy požiadaviek, čo znamená, že existujúci softvér aj softvér vo vývoji by mali byť schopné pracovať s anonymizovanými a pseudo-anonymizovanými dátami.

Ďalším dôležitým aspektom je extenzibilita existujúceho softvéru. Je potrebné zabezpečiť, aby systém bol schopný generovať agregované štatistiky, čo môže zahŕňať rôzne metódy a techniky pre spracovanie dát. To umožňuje prispôsobenie systému na vyhovovanie konkrétnym požiadavkám a potrebám rôznych inštitúcií.

Kľúčovým aspektom je škálovateľnosť systému. Anonymizované dáta môžu mať podobnú veľkosť ako pôvodná vzorka dát, a tak počet agregátov a metadát požadovaných jednotlivými časťami systému môže dramaticky narásť. Preto je dôležité, aby systém bol schopný spracovať a spravovať veľké množstvo dát a prispôsobiť sa rastúcim požiadavkám na spracovanie a uchovávanie dát.

Ďalším dôležitým princípom je kontinuálne zlepšovanie využiteľnosti dátového toku. Dáta nie sú používané v izolovanom systéme, ale môžu slúžiť ako zdroj informácií pre rôzne inštitúcie. Pri vytváraní agregovaných dát pre tieto inštitúcie je dôležité brať do úvahy rozmanitosť potenciálnych zdrojov dát, ktoré môžu byť použité pri procese de-anonymizácie krížovou validáciou. Preto by služby poskytujúce agregované dáta mali byť koordinované a sledovať najnovšie metódy a postupy na ochranu dát a minimalizáciu rizika de-anonymizácie.

Celkovo je dôležité, aby pri dizajne a vývoji systému, ktorý spracováva anonymizované dáta, boli dodržané tieto princípy interoperability a extenzibility.

Pri tvorbe anonymizačného systému je nevyhnutné začať aktívnu diskusiu s vlastníkom dát, ktoré budú v tomto systéme spracovávané, ako aj s budúcimi administrátormi tohto systému. Cieľom tejto diskusie je dosiahnuť porozumenie a zhodu všetkých zúčastnených strán v súvislosti s dôležitým vzťahom medzi mierou anonymizácie dát (bezpečnosťou) a použiteľnosťou dát v analytických procesoch.

Je potrebné, aby všetky strany mali jasnú predstavu o tom, že vyššia miera anonymizácie môže mať za následok nižšiu použiteľnosť dát pre analytické účely. V rámci diskusie by sa mali spoločne definovať parametre tohto vzťahu, ktoré vytvoria akýsi rámec pre rozhodovanie o úrovni anonymizácie v kontexte konkrétnych modelov alebo analýz.

Diskusia má viesť k určeniu konkrétnej miery pomeru medzi bezpečnosťou a použiteľnosťou dát. Napríklad v prípade modelov, ktoré slúžia na odhad zdravotných parametrov, môže skupina stanoviť, aká miera anonymizácie je akceptovateľná a v akej miere sú ochotní prijať určitý stupeň rizika v súvislosti s konkrétnymi dátovými parametrami.

Výsledkom tejto diskusie by mali byť aj definované metriky bezpečnosti dát a metriky ich využiteľnosti. Metrika bezpečnosti dát by slúžila na hodnotenie úrovne anonymizácie a zabezpečenia dát v systéme, zatiaľ čo metrika utility by sa zameriavala na hodnotenie použiteľnosti a kvality spracovávaných dát. Tieto metriky by poskytovali objektívne merateľné ukazovatele, ktoré by slúžili na hodnotenie a monitorovanie systému anonymizácie.

Celkovo je dôležité, aby diskusia a vzájomná spolupráca medzi vlastníkmi dát, administrátormi systému a ostatnými zainteresovanými stranami viedla k jasnému stanoveniu cieľov, parametrov a metrík, ktoré by zabezpečili vyvážený pomer medzi bezpečnosťou a použiteľnosťou dát v anonymizačnom systéme.

Pri tvorbe anonymizačného systému je nevyhnutné začať aktívnu diskusiu s vlastníkom dát, ktoré budú v tomto systéme spracovávané, ako aj s budúcimi administrátormi tohto systému. Cieľom tejto diskusie je dosiahnuť porozumenie a zhodu všetkých zúčastnených strán v súvislosti s dôležitým vzťahom medzi mierou anonymizácie dát (bezpečnosťou) a použiteľnosťou dát v analytických procesoch.

Systém musí tiež byť pripravený na potreby a požiadavky jeho používateľov. Používateľov by sa mala informovať, že dáta, ktoré dostanú z databázy, sú buď anonymizované alebo agregované s menšou úrovňou detailov. Obmedzenia a štandardný formát, v ktorom budú poskytované dáta tretím stranám, musia byť jasne definované, aby sa predišlo nedorozumeniam pri vývoji nástrojov, ktoré budú automaticky generovať a poskytovať tieto upravené dátové vzorky.

Ideálne by mal byť vytvorený systém na mapovanie dotazov a požiadaviek na databázu, ktorý by pomohol v ich riadení. Tento systém by poskytoval informácie o zaťažení databázy požiadavkami a umožňoval detekciu podozrivých požiadaviek, napríklad častých alebo neautorizovaných. S takýmto systémom by bolo možné vytvoriť interaktívny prístup k databázam, ktorý by poskytoval agregované dáta na základe požiadaviek používateľov (ak by tieto požiadavky spĺňali bezpečnostné podmienky). Namiesto poskytovania rozsiahlych anonymizovaných dátových sád, nad ktorými by systém stratil kontrolu po ich odovzdaní, by tento systém umožňoval dynamické poskytovanie dát podľa potrieb používateľov.

Takýto interaktívny prístup by zabezpečil efektívnejšie využívanie dát a minimalizoval by riziko straty kontroly nad anonymizovanými dátami. Používatelia by mohli získať potrebné agregované informácie, pričom by sa zabezpečili bezpečnostné opatrenia a obmedzenia prístupu.

4.2.8.1 Podmienky ochrany citlivých údajov

Jedným z cieľov výskumného projektu je podpora správcovských inštitúcií, používateľov, dodávateľov a tretích strán pri vývoji hardvérových a softvérových riešení. Predpokladá sa, že zvýšená frekvencia zbierania dát spolu s výkonnými analytickými nástrojmi, interpretáciou a integrovaním vnútorných komunikačných štruktúr výsledkov analýz, prinesie systému výhody pre všetky zainteresované strany. Tieto výhody umožnia lepšie rozhodovanie, lepšie pochopenie potrieb a požiadaviek klientov a distribučných partnerov, čo má pozitívny dopad na efektivitu celého projektu.

Vyššia frekvencia zbierania dát znamená, že sa bude mať k dispozícii čerstvý a aktuálny pohľad na dianie v systéme. Spolu so silnejšími analytickými nástrojmi a interpretáciou týchto dát budú správcovské inštitúcie, používatelia, dodávatelia a tretie strany schopné lepšie porozumieť trendom a vzorcom v dátach. Týmto sa zlepší schopnosť robiť informované rozhodnutia založené na presných a aktuálnych informáciách.

Integrácia vnútroštruktúrnej komunikácie záverov analýz umožní, aby tieto poznatky boli dostupné a využiteľné pre všetky zúčastnené strany. To znamená, že správcovské inštitúcie, používatelia, dodávatelia a tretie strany budú mať prístup k relevantným informáciám, ktoré im umožnia lepšie pochopiť potreby a požiadavky svojich klientov a distribučných partnerov. Táto spolupráca a výmena informácií povedie k zlepšeniu výkonu projektu a efektivity celého systému.

Výhodou tohto prístupu je, že umožňuje zapojenie všetkých relevantných strán do vývoja a zlepšovania systému. Správcovské inštitúcie, používatelia, dodávatelia a tretie strany budú mať možnosť aktívne prispievať svojimi schopnosťami a možnosťami k vývoju hardvérových a softvérových riešení.

Prirodzene, pri takomto projekte vznikajú závažné otázky týkajúce sa ochrany osobných údajov koncových používateľov. Zbieranie zdravotných údajov s vysokou frekvenciou a granularitou sekundárnych premenných vyžaduje zvláštnu starostlivosť. Monitorovanie zdravotného stavu zahŕňa sledovanie telesných funkcií aj na sekundovej úrovni. Tieto dáta majú významnú hodnotu pre farmaceutické spoločnosti, zdravotnícke zariadenia a výskumné organizácie zamerané na rôzne choroby. Ak by tieto informácie boli zneužité alebo de-anonymizované, vážne by to zasiahlo do základných ľudských práv pacienta (klienta).

Preto je nevyhnutné, aby vývojový tím venoval maximálnu pozornosť vytvoreniu kvalitného bezpečnostného systému pre zber, správu, analýzu a distribúciu dát o zdravotných parametroch. Tento systém by mal byť navrhnutý tak, aby poskytoval maximálnu ochranu osobných údajov a minimalizoval riziko ich zneužitia alebo de-anonymizácie.

Implementácia bezpečnostných opatrení by mala zahŕňať silné metódy anonymizácie, šifrovanie prenosu a uloženia dát, prísne prístupové práva a kontroly prístupu, monitorovanie podozrivých aktivít a riadenie rizík. Dôležité je tiež zabezpečiť dodržiavanie príslušných legislatívnych a regulačných požiadaviek týkajúcich sa ochrany osobných údajov, ako napríklad Všeobecnej nariadenia o ochrane údajov (GDPR) v Európskej únii.

Okrem technických opatrení je dôležité zabezpečiť aj správne vzdelávanie a informovanie všetkých zainteresovaných strán o zásadách ochrany osobných údajov a dôležitosti ich dodržiavania. To zahŕňa školenie pracovníkov, dodávateľov a používateľov systému, aby boli informovaní o správnych postupoch a bezpečnostných opatreniach.

Vytvorenie kvalitného bezpečnostného systému na zber, manažment, analýzu a distribúciu dát o zdravotných parametroch je nevyhnutné pre zachovanie dôveryhodnosti projektu.

4.2.8.2 Podmienky na vývoj bezpečnostného systému v kontexte zdravotníckych dát

Pri tvorbe tohto projektu je kľúčové nájsť rovnováhu medzi pripisovaním dát konkrétnym pacientom a ochranou osobných údajov fyzických a právnických osôb, ktoré tieto dáta generujú svojím správaním. Cieľom je zabezpečiť, že dáta budú priradené k správnym subjektom a zároveň chrániť súkromie jednotlivcov.

Aby sa dosiahol tento cieľ, je nevyhnutné implementovať vhodné bezpečnostné opatrenia a postupy. Napríklad sa môže využiť anonymizácia dát, ktorá odstraňuje priame identifikátory a zabezpečuje, že dáta sú agregované alebo transformované takým spôsobom, že je nemožné jednoznačne identifikovať jednotlivých pacientov. Týmto spôsobom je možné zachovať úroveň anonymnosti, zatiaľ čo sa dáta stále priraďujú k príslušným pacientom na účely analýzy a výskumu.

Je však dôležité zaistiť, že tieto bezpečnostné opatrenia nebudú mať negatívny vplyv na kontinuálny tok odmerných dát vo vysokej frekvencii. Systém musí byť navrhnutý tak, aby umožňoval efektívne zhromažďovanie, spracovanie a prenos dát bez výrazného oneskorenia alebo prerušenia. To zabezpečí, že kontinuálny prúd dát bude neustále dostupný pre analýzu a využitie.

Okrem toho je tiež dôležité minimalizovať negatívny dopad na prebiehajúce operácie v sieti. Bezpečnostné opatrenia a procesy spracovania dát by nemali obmedzovať alebo rušiť bežné fungovanie siete a systému. Musí sa zabezpečiť, že dáta budú správne spracované a distribuované bez toho, aby to spôsobilo prerušenie alebo degradáciu sieťovej infraštruktúry.

Zachovanie pripisovania dát konkrétnym pacientom a ochrana osobných údajov sú dôležitými aspektmi tohto projektu. Ich úspešná implementácia a dodržiavanie zabezpečia, že projekt bude účinný, spoľahlivý a zákonný v súlade s príslušnými legislatívnymi požiadavkami na ochranu údajov.

V kontexte budúceho systému sú definované nasledujúce špecifikácie. Prvým kritériom je bezpečné priradenie zdravotných dát ku konkrétnym klientom alebo pacientom pre zdravotné analýzy. Pre monitorovanie postačuje generovanie agregovaných štatistík s frekvenciou na úrovni minút alebo hodín. Pre účely výskumu nie je potrebné priradzovať merané dáta konkrétnym osobám či inštitúciám, stačí ich anonymizácia za predpokladu, že sú overiteľné a je možné ich priradiť ku konkrétnemu subjektu. Anonymizované dáta môžu byť zbierané a vyhodnocované aj v reálnom čase. Najmenšou akceptovateľnou jednotkou anonymizovaných dát o pacientoch pre štatistické modelovanie je dátový agregát na úrovni každého monitorovaného pacienta. Systém by mal byť schopný spravovať meracie prvky a ich funkcie aj pre anonymných pacientov podľa požiadaviek manažmentu monitoringu. Dôležitou podmienkou pre bezpečný systém je dôverný vzťah medzi organizáciami, pretože vlastníctvo a prístup k dátam generovaným zdravotníckymi jednotkami nie je jednoznačne definované a závisí od legislatívnych predpisov a vzťahov medzi jednotlivými stranami. Kvalitná definícia týchto vzťahov je kľúčová pre vývoj bezpečnostného systému, ktorý je úzko viazaný na daný kontext a prostredie.

V rámci týchto špecifikácií je možné stanoviť pravidlá pre manažment a manipuláciu s dátami v takomto systéme. Zber dát by mal byť striktne obmedzený na dáta, ktoré sú nevyhnutné pre konkrétne operácie, za ktoré každá inštitúcia zodpovedá v rámci systému monitorovania zdravotných údajov. Tieto operácie môžu zahŕňať plánovanie a manažment, testovanie efektivity distribúcie, výskum a ďalšie. Týmto spôsobom je zabezpečené, že iba potrebné dáta sú zbierané a manipulované, čo prispieva k ochrane osobných údajov a minimalizácii rizika zneužitia.

V rámci tohto systému je nevyhnutné prísne obmedziť uchovávanie dát, z ktorých je možné získať osobné údaje alebo informácie o správaní pacientov. Toto obmedzenie by malo zahŕňať každú časť systému a mala by sa uchovávať iba doba potrebná na splnenie schválených aktivít, o ktorých bol koncový používateľ informovaný a súhlasil s nimi. Napríklad, pre tvorbu predikcií môže byť potrebné získať veľké množstvo dát, ale táto doba uchovávania by mala byť jasne stanovená a obmedzená. Po vykonaní všetkých schválených aktivít s dátami majú inštitúcie povinnosť nenávratne zničiť pôvodné dáta, ako aj všetky ich kópie a metadáta, ktoré z týchto dát vznikli. Týmto spôsobom sa zabezpečuje, že osobné údaje a citlivé informácie sú bezpečne odstránené po splnení stanovených úloh a minimalizuje sa riziko neoprávneného prístupu k týmto údajom.

Distribúcia dát v rámci tohto systému vyžaduje zavedenie a udržanie bezpečných dátových intervalov. Je dôležité, aby sme boli schopní správne implementovať princíp minimalizácie zbieraných dát a preto potrebujeme porozumieť, aké informácie môžeme odvodiť z týchto dát v závislosti na rôznych intervaloch zberu. Je nevyhnutné skúmať vzťah medzi veľkosťou týchto intervalov, množstvom informácií, ktoré môžeme z takýchto intervalov získať, a zároveň zohľadňovať bezpečnosť v kontexte siete Smart Grid. Toto vyžaduje ďalší výskum a štúdium, aby sme lepšie porozumeli, ako tieto faktory ovplyvňujú distribúciu dát a zabezpečili, že využitie intervalov zberu je v súlade s potrebami bezpečnosti a ochrany údajov.

Okrem intuitívnej agregácie dát v čase, je vhodné tiež vytvárať agregáciu dát na základe pacientov. Týmto spôsobom môžeme značne obmedziť bezpečnostné riziká spojené s meraním zdravotného správania používateľov. Systém by mal vytvárať agregáty na základe blízkosti alebo podobnosti používateľov, napríklad podľa diagnózy a ďalších faktorov, a zároveň anonymizovať individuálne dáta. Pri tomto postupe je dôležité definovať, do akej miery treba dáta agregovať a aký je správny pomer medzi bezpečnosťou a užitočnosťou týchto agregátov v konkrétnych prípadoch. Je však dôležité mať na pamäti, že agregácia sama osebe nedokáže úplne chrániť jednotlivca, ak majú všetci v agregovanej

vzorke podobné hodnoty vo viacerých premenných. V takých prípadoch je potrebné kombinovať túto metódu s inými typmi anonymizácie, aby sa zabezpečila ochrana osobných údajov.

Minimálne charakteristiky vzorky potrebné pre efektívnu agregáciu dát sú nasledovné: Vzorka by mala obsahovať minimálne 15 používateľov alebo rozlíšiteľné skupiny používateľov. V každej rozlíšiteľnej skupine používateľov by nemala existovať jediná premenná, ktorá by umožňovala identifikáciu iba jedného používateľa. Zároveň by žiadny používateľ nemal generovať viac ako 15 % zo všetkých dát, ktoré majú byť agregované spolu. Ak pravidlo 15/15 nie je dodržané, odporúča sa nedistribuovať takéto dáta tretím stranám, nad ktorými nemáme kontrolu. Je vhodné uplatniť princíp prístupu k dátam s určitou latenciou pre všetky dáta, ktoré sú poskytované priamo používateľom, fyzickým osobám, komerčným a nekomerčným právnickým osobám mimo systému. Jednou z obáv zneužitia zdravotných dát je ich potenciálne využitie ako nástroja na vydieranie. Aplikáciou časového oneskorenia pri poskytovaní dát tretím stranám a úpravou časových záznamov by sa tento problém mohol minimalizovať. Tieto charakteristiky sú dôležité pre zabezpečenie bezpečnosti a ochrany súkromia pri agregácii a distribúcii zdravotných dát. Ich dodržiavanie umožní efektívne využitie agregovaných informácií, pričom zároveň minimalizuje riziko identifikácie jednotlivých používateľov a možnosť zneužitia dát. Prínosom je aj aplikácia latencie pri poskytovaní dát tretím stranám, čo zvyšuje bezpečnosť a zníženie rizika potenciálneho zneužitia.

4.2.9 Zhodnotenie

V tejto časti analýzy sme mali za cieľ poskytnúť prehľad o problematike bezpečnosti zberu, manažmentu, analýzy, uchovávania a distribúcie dát v systéme Life Defender - Ochranca života. Ďalej sme sa zameriavali na testovanie rôznych prístupov k anonymizácii týchto dát na sprostredkovanej vzorke. Technický popis tohto procesu je však príliš komplexný na zahrnutie do tohto reportu, ktorý slúži ako výstup projektu. Preto sme sa zamerali na hlavné závery, odporúčania a špecifikácie, ktoré sme získali počas vykonávania práce.

V rámci analýzy sme identifikovali rôzne aspekty súvisiace s bezpečnosťou dát v systéme Life Defender. Zaoberali sme sa otázkou zabezpečenia zberu dát, aby boli dáta priradené ku konkrétnym klientom alebo pacientom bezpečne a spoľahlivo. Ďalej sme sa venovali generovaniu agregovaných štatistík na rôznych časových intervaloch pre účely monitorovania a výskumu. Dôležitou súčasťou analýzy bolo aj skúmanie možností anonymizácie dát, pričom sme zohľadnili požiadavky na overiteľnosť a možnosť priradenia anonymizovaných dát ku konkrétnym subjektom.

Okrem toho sme sa zaoberali aj otázkou uchovávania dát a ich správnou manipuláciou. Definovali sme potrebu nenávratného zničenia pôvodných dát, ich kópií a metadát po vykonaní schválených aktivít. Taktiež sme zdôraznili význam zachovania bezpečných dátových intervalov pri distribúcii dát. Identifikovali sme minimálne charakteristiky vzorky, ktoré by mali byť splnené pre efektívnu agregáciu dát a minimalizáciu bezpečnostných rizík.

V závere sme poukázali na skutočnosť, že technický popis procesu anonymizácie je príliš rozsiahly pre tento report, ktorý slúži ako súhrn výsledkov projektu. Preto sme sa zamerali na extrahovanie hlavných záverov, odporúčaní a špecifikácií, ktoré budú slúžiť ako usmernenie pre ďalší vývoj a implementáciu systému Life Defender - Ochranca života.

Anonymizačné metódy majú v kontexte bezpečnosti a správy osobných údajov veľký význam a sú nástrojmi, ktoré môžu byť veľmi užitočné. Avšak, ako pri akomkoľvek aspekte týkajúcom sa bezpečnosti, kľúčom k vytvoreniu odolného systému je porozumenie kontextu, v ktorom sa bezpečnostný systém použije.

Všetky zložky, ktoré sú aktívne zapojené do systému, musia úplne porozumieť rozhodnutiu anonymizovať svoje dáta a musia mať jasno v tom, aký vzťah má daný proces anonymizácie k ochrane osobných údajov používateľov. Je dôležité, aby každá inštitúcia alebo organizácia, ktorá pracuje so systémom a spracováva osobné údaje, mala jasný prehľad o anonymizačných postupoch, ktoré používa, a pochopila ich dôsledky a obmedzenia.

Pri výbere a implementácii anonymizačných metód je nevyhnutné zvážiť konkrétny kontext a potreby daného systému. Niektoré metódy anonymizácie môžu byť viac vhodné pre určité typy dát alebo aplikácií, zatiaľ čo iné metódy môžu byť účinnejšie v iných prípadoch. Rovnako je potrebné zvážiť aj právne a regulačné požiadavky, ktoré sa vzťahujú na anonymizáciu osobných údajov v konkrétnom sektore alebo jurisdikcii.

Okrem samotných technických aspektov anonymizácie je dôležité mať na pamäti aj educáciu a informovanosť všetkých zainteresovaných strán. Používatelia systému a inštitúcie musia mať dostatočné znalosti o anonymizačných procesoch a ich význame pre ochranu osobných údajov. Rovnako je dôležité zabezpečiť, aby boli dodržané všetky práva a povinnosti v súlade s relevantnými zákonmi a reguláciami týkajúcimi sa ochrany osobných údajov.

V konečnom dôsledku je dôležité, aby všetky strany zapojené do systému Life Defender alebo akéhokoľvek iného systému, ktorý spravuje a spracováva osobné údaje, mali jasné a dobre premyslené postupy anonymizácie.

V rámci realizácie projektu Life Defender - Ochranca života je dôležité identifikovať, ktoré údaje z tohto projektu potrebujú jednotlivé inštitúcie, akým cieľom a pre koho budú tieto údaje sprostredkované. Táto prvá fáza je nevyhnutná pre vyjasnenie povahy dátových požiadaviek a vzájomnej interakcie medzi inštitúciami.

Po úspešnom zmapovaní požiadaviek jednotlivých inštitúcií je možné určiť optimálny pomer medzi mierou anonymizácie a zachovaním štatistického potenciálu dát pre každú z týchto inštitúcií a pre ich vzájomnú dátovú interakciu. Je potrebné nájsť rovnováhu medzi ochranou osobných údajov a uchovávaním dostatočného množstva informácií pre účely výskumu, analýzy a plánovania v rámci jednotlivých inštitúcií.

Aby sme vytvorili kvalitný bezpečnostný systém, je nevyhnutné presne pochopiť potreby a očakávania jednotlivých inštitúcií v rámci systému Life Defender. To zahŕňa jasnú definíciu dátových požiadaviek, požadovaných úrovni ochrany osobných údajov a spôsobu dátového zdieľania medzi inštitúciami. Toto porozumenie umožní vytvorenie efektívneho a bezpečného systému, ktorý bude slúžiť potrebám všetkých zainteresovaných strán.

Pri vývoji softvéru je dôležité spojiť teoretické poznatky o anonymizácii dát s praktickým využitím v zdravotníctve. Aj keď to môže byť náročný a dlhodobý proces, nemôžeme prehliadnuť jeho konečný cieľ. Analyzujúc podobné bezpečnostné systémy, ukazuje sa, že jedným z ich najzraniteľnejších miest je samotný používateľ. Používateľ požiada o prístup k svojim dátam o zdravotnom správaní (čo je jeho oprávnenie), avšak potom tieto dáta poskytne tretej strane, ktorá má potenciál ich zneužiť.

Vzdelávanie koncových používateľov o bezpečnom správaní sa so svojimi dátami má veľký potenciál eliminovať množstvo takýchto útokov. Je dôležité, aby používatelia boli informovaní o rizikách a opatreniach, ktoré môžu prijať na ochranu svojich osobných údajov. Tieto informácie by mali zahŕňať správne používanie systému, zásady zdieľania údajov a upozornenia na možné dôsledky ich neoprávneného poskytovania tretej strane.

Vzdelávacie programy a kampane by mali zdôrazňovať význam ochrany súkromia, bezpečného správania s dátami a identifikáciu príznakov neoprávnenej žiadosti o dáta. Používateľom by sa mali poskytnúť nástroje na správu svojich súkromných informácií, napríklad možnosti nastavenia úrovne prístupu k svojim dátam alebo sledovania aktivity v ich účte. Týmto spôsobom by sa zvýšila povedomosť a zodpovednosť používateľov, čo by významne prispelo k ochrane ich osobných údajov a zníženiu rizika útokov a zneužitia dát.

V konečnom dôsledku je významné, aby vývojári a poskytovatelia systému Life Defender venovali primeranú pozornosť vzdelávaniu používateľov a zabezpečili, aby boli informovaní o správnom a bezpečnom používaní systému. Vhodné školenie a osveta môžu zohrať kľúčovú úlohu pri eliminácii bezpečnostných rizík.

5 ZÁVER – ZHRNUTIE PROJEKTU

V projekte Life Defender – Ochranca života, ktorý bol realizovaný v období 9.2.2021 – 30.6.2023 sme postupne predkladali výstupné správy, v ktorých bol zhrnutý výskum projektových tém. Projekt bol za žiadateľa rozdelený na dve aktivity:

- Aktivita H1 – Riešenie SW platformy na integrovanie evidencie návštevníkov, zberov dát z existujúceho HW ako i prototypov nového HW do jednotného informačného systému Life Defender – ochranca života – prototyp
- Aktivita H2 - Riešenie SW platformy na integrovanie evidencie návštevníkov, zberov dát z existujúceho HW ako i prototypov nového HW do jednotného informačného systému Life Defender – ochranca života – prototyp – flexibilita 15%

Projekt bol členený do štyroch míľnikov v nasledujúcich časových etapách:

- Míľnik č. 1 - Analýza potrieb spoločnosti (9.2.2021 – 31.5.2021)
- Míľnik č. 2 - Návrh prototypu SW platformy – spôsob evidencie, zberu dát a komunikácie (1.6.2021 – 30.6.2022)
- Míľnik č. 3 - Návrhy postupov pre použitie platformy pre vybrané procesy (SW + HW), (1.7.2022 – 31.1.2022)
- Míľnik č. 4 - Analýza využiteľnej technológie a testovanie zariadení (1.1.2023 – 30.6.2023)

Tématicky boli aktivity H1 a H2 rozdelené do troch pracovných balíkov:

- Experimentálny vývoj prototypu SW platformy a cloudového úložiska
- Experimentálny vývoj prototypu mobilnej aplikácie
- Experimentálny vývoj prototypu modulu pokročilej analýzy a vizualizácie dát

V rámci prvého projektového míľnika *Analýza potrieb spoločnosti* (9.2.2021 – 31.5.2021) sme sa zaoberali analýzou pracovných balíkov projektu, pričom boli identifikované všetky potrebné vstupy. Pre pracovný balík Experimentálny vývoj prototypu SW platformy a cloudového úložiska sme analzyovali dostupné prostredia, nástroje a platformy vhodné pre zber dát, ich spracovanie, uskladnenie, vizualizáciu a analýzu. Tiež sme riešili a popisovali návrh procesov a štruktúr a ponúkli zoznam nefunkčných požiadaviek či návrh architektúry vrátane požiadaviek na poskytovateľov hardvéru. V pracovnom balíku vývoja prototypu mobilnej aplikácie rozoberáme prepojenie projektových cieľov s témami domáceho monitorovacieho systému, automatickej testovacej stanice a mobility obyvateľstva. Analyzujeme aktuálnu situáciu na trhu s dostupnými aplikáciami. Pracovný balík Experimentálny vývoj prototypu modulu pokročilej analýzy a vizualizácie dát sa zaoberá prototypom modulu vizualizácie dát a tiež prototypom dátového modelu na zbieranie, analýzu a vyhodnocovanie symptómov pomocou analytických nástrojov s využitím prvkov umelej inteligencie. V kapitole venovanej prototypu modulu vizualizácie dát analyzujeme aplikácie, ktoré by mohli byť vhodné pre potreby riešenia problematiky projektu. Ponúkame opis funkčných a nefunkčných požiadaviek a rozoberáme problematiku umelej inteligencie z pohľadu analýzy dát. Ďalej navrhujeme postup pri výbere modelu a hodnotiacej metriky. Na základe vyzbieraných, upravených, vyčistených a transformovaných pilotných dát sme trénovali a testovali modely strojového učenia.

Druhý projektový míľnik - *Návrh prototypu SW platformy – spôsob evidencie, zberu dát a komunikácie* bol realizovaný v období 1.6.2021 – 30.6.2022. V roku 2021 prebiehali konfiguračné a programovacie práce na vytvorení prototypov SW podpory v nadväznosti na aktivitu partnera H3. Do konca míľnika boli návrhy prototypov vypracované nad všetkými tromi pracovnými balíkmi. V časti Experimentálny vývoj prototypu SW platformy a cloudového úložiska sme popisovali vytvorené

prototypy pre zariadenia, ktoré sa stali výstupom projektu. Súčasťou každej časti bol podrobný popis architektúry riešenia, dátového modelu a webového rozhrania. Pre oblasť Domácej karantény bolo cieľom vývoja prototypu vytvorenie nástroja pre lekárov a pacientov, ktorý umožňuje vzdialené monitorovanie pacientov a vyhodnocovanie nazbieraných údajov z jednotného cloudového úložiska. Pre oblasť Automatickej testovacej bunky bolo cieľom vývoja prototypu vyvinúť riešenie, ktoré ukladá a poskytuje dáta o otestovaných subjektoch do jednotného cloudového úložiska dát, a taktiež poskytnúť ovládacie prostredie pre dotykové obrazovky na automatickej testovacej bunke. Pre oblasť Dezinfekčného robota sme v rámci míľnika č. 2 vyvinuli riešenie, ktoré ukladá a poskytuje dáta o stave ovzdušia (hustota zamorenia konkrétnych častíc) v ľubovoľnom priestore, ako sú kancelárie, obchodné priestory, chodby, letiskové haly, továrne a podobne. V časti Experimentálny vývoj prototypu mobilnej aplikácie sme popisovali vytvorené prototypy pre zariadenia, ktoré budú výstupom projektu. Súčasťou každej časti je podrobný popis architektúry riešenia, dátového modelu a webového rozhrania. V časti Experimentálny vývoj prototypu modulu pokročilej analýzy a vizualizácie dát sme popisovali výskum v oblasti Prototypu modulu vizualizácie dát a Prototypu dátového modelu na zbieranie, analýzu a vyhodnocovanie symptómov pomocou analytických nástrojov s využitím umelej inteligencie. Pre oblasť Prototypu modulu vizualizácie dát sme navrhli a popísali možnosť použitia vizuálneho nástroja Kibana, ktorým je možné interaktívnym spôsobom prezerať dáta uložené v Elasticsearch. V tejto kapitole sa venujeme aj zoznamu grafických metód, zobrazeniu spojitých premenných a ich charakteristikám. Pre oblasť Prototypu dátového modelu na zbieranie, analýzu a vyhodnocovanie symptómov pomocou analytických nástrojov s využitím umelej inteligencie sme popísali viaceré metódy pre analýzu zdravotníckych dát.

Tretí projektový míľnik - *Návrhy postupov pre použitie platformy pre vybrané procesy (SW + HW)* sme realizovali v období 1.7.2022 – 31.12.2022. Vo výstupnom dokumente sme sa zaoberali témou Metodiky testovania platformy, kde sme popísali zvolené metódy testovania, ktoré sme identifikovali ako vhodné pre jednotlivé oblasti riešenia – či už pre vytvorené prototypy domácej karantény, automatickej testovacej bunky alebo pre dezinfekčného robota; a tiež samotné testy – UC´s – ktorými sme preukázali vhodnosť navrhnutého riešenia. Pre navrhnutý prototyp mobilnej aplikácie sme otestovali viaceré procesné situácie a vylepšili sme užívateľské rozhranie. V časti Experimentálny vývoj prototypu modulu pokročilej analýzy a vizualizácie dát sme pre oblasť Prototypu modulu vizualizácie dát navrhli a popísali možnosť použitia vizuálneho nástroja Kibana, ktorým je možné interaktívnym spôsobom prezerať dáta uložené v Elasticsearch. Pre Prototyp dátového modelu na zbieranie, analýzu a vyhodnocovanie symptómov pomocou analytických nástrojov s využitím umelej inteligencie sme sa zamerali na prehľad štatistických algoritmov pre analýzu a predikciu zdravotníckych dát, metódy strojového učenia a ich efektivita v kontexte modelovania zdravotníckych dát, zdravotný monitoring a metodologický prístup k modelovaniu zdravotných dát, optimalizačné nástroje, kvalitu spracovaných zdravotníckych dát, tému ochrany citlivých údajov v kontexte zdravotných dát, proces anonymizovania dát a dizajn bezpečnostného systému.

Posledný, štvrtý projektový míľnik - *Analýza využiteľnej technológie a testovanie zariadení* (1.1.2023 – 30.6.2023) sme zdokumentovali vo výstupe, ktorý práve držíte v rukách. V jednotlivých kapitolách štvrtého míľnika postupne predstavujeme stav riešenia vzhľadom na ďalšie využitie vytvorenej platformy a venujeme sa témam ako možné prínosy a využitie platformy počas pandémie a mimo pandémie, registrácii a prihláseniu používateľov do systému, vykonávanie testov, poskytnutiu dát tretím stranám či v prípade dezinfekčného robota aj ovládaniu nastavovania jednotlivých módov robota. Výsledkom je aj synergický efekt ktorý sa týka väčšej celkovej účinnosti a hodnoty, ktorú môžu poskytnúť kombinované použitie softvéru pre domácu karanténu, automatickú testovaciu bunku a dezinfekčného robota. Problematika zavedenia a podpory platformy v budúcej prevádzke tvorí ďalšiu časť dokumentu. Pre jednotlivé pracovné balíky hardvérovej časti popisujeme základné funkčnosti, architektúru riešenia, integráciu na jednotné cloudové úložisko, ďalšie typy integrácií; taktiež problematiku bezpečnosti, autentifikácie, autorizácie, monitorovania a podpory prevádzky systémov. Zámerom analýzy pre pracovný balík prototypu dátového modelu na zbieranie, analýzu a vyhodnocovanie symptómov

pomocou analytických nástrojov s využitím umelej inteligencie bolo poskytnúť prehľad o otázkach bezpečnosti dát v systéme Life Defender – ochranca života. Zameriavame sa na zber, správu, analýzu, ukladanie a distribúciu údajov, ako aj na testovanie rôznych prístupov k anonymizácii údajov. Ponúkame podrobnú analýzu prehľadu štatistických algoritmov pre analýzu a predikciu zdravotníckych dát vrátane témy metód strojového učenia a ich efektivity v kontexte modelovania zdravotníckych dát. Ďalšou témou bol zdravotný monitoring a metodologický prístup k modelovaniu zdravotných dát. Zaoberáme sa rovnako aj aspektmi, ako je zabezpečenie zberu údajov, generovanie súhrnných štatistík na monitorovanie a výskum a anonymizácia údajov, pričom sme zvážili overiteľnosť a prepojenie na konkrétne subjekty. Skúmame uchovávanie údajov a správnu manipuláciu, pričom zdôrazňujeme potrebu nenávratného zničenia pôvodných údajov a udržiavanie bezpečných intervalov údajov počas distribúcie. Analyzujeme minimálne charakteristiky vzorky pre efektívnu agregáciu údajov a minimalizáciu bezpečnostných rizík. Ďalšou oblasťou je vytváranie robustného bezpečnostného systému. Výber a implementácia metód anonymizácie zohľadňuje špecifický kontext systému, typy údajov, aplikácie a právne/regulačné požiadavky.

Výstupmi projektu sú vytvorené SW postupy pre prototypy HW časti, ktorej sa vo svojich pracovných balíkoch venoval partner v projekte, spoločnosť Bizzcom, s.r.o. Jednotlivé demá postupov ovládania zariadení domácej karantény, automatickej testovacej bunky, či dezinfekčného robota, sú dostupné na projektovej webovej stránke https://lifedefender.sk/

- SW ovládanie pre automatickú testovacu bunku:
  https://touch1.lifedefender.sk/home
  https://touch2.lifedefender.sk/home
- SW ovládanie pre domácu karanténu:
  https://app.lifedefender.sk/login
- SW ovládanie pre automatickú dezinfekciu:
  https://lifedefender.sk/prototyp-mobilna-aplikacia/

Projekt Life Defender – Ochranca života, v rámci splnenia merateľného ukazovateľa „P0762 Počet publikácií vytvorených v rámci projektu“ vytvoril nasledujúce publikácie:

- Analýza SW platformy pre ukladanie, manažment, zdieľanie a zber zdravotníckych dát
- Návrh prototypu SW platformy – spôsob evidencie, zberu dát a komunikácie
- Analýza HW platformy na prevádzku jednotného informačného systému Life Defender
- Návrh prototypov zariadení HW platformy
- Vývoj prototypu a aplikovanie SW platformy
- Radar-Based Volumetric Precipitation Nowcasting: A 3D Convolutional Neural Network with U-Net Architecture
- WarpSTR: Determining tandem repeat lengths using raw nanopore signals

Publikácie sú dostupné napríklad na webovej stránke: https://lifedefender.sk/projektova-dokumentacia/

6 ZDROJE

[1,2] Kibi User Guide: Introduction. Siren Solutions [online]. [cit. 2016-03-30]. Dostupné z: http://siren.solutions/kibi/docs/current/introduction.html

[3] Vizualizácia dát pomocou nástroja Kibana. Rácek, T. [online]. [2016]. Dostupné z: https://is.muni.cz/th/l7oyt/visualization_thesis_racek.pdf

7 ZOZNAM OBRÁZKOV

8 ZOZNAM TABULIEK